A DNA Aptamer That Binds Adenosine and ATP†

David E. Huizenga and Jack W. Szostak*

Department of Molecular Biology, Massachusetts General Hospital, Boston, Massachusetts 02114, and Program in Biological and Biomedical Sciences, Harvard Medical School, Boston, Massachusetts

Received August 29, 1994; Revised Manuscript Received October 31, 1994®

ABSTRACT: We have used *in vitro* selection to isolate adenosine/ATP-binding DNA sequences from a pool of $\approx 2 \times 10^{14}$ different random-sequence single-stranded DNA molecules. One of these aptamers has been characterized and binds adenosine in solution with a dissociation constant of 6 ± 3 μM. Experiments with ATP analogs indicate that functional groups on both the base and the sugar of ATP are involved in the ligand/aptamer interaction. The binding domain of this aptamer was localized to a 42 base sequence by deletion analysis. A pool of mutagenized versions of this sequence was then synthesized and screened for functional adenosine binding sequences; comparison of the selected variants revealed two highly conserved guanosine-rich regions, two invariant adenosine residues, and two regions of predominantly Watson-Crick covariation. This data led us to propose a model of the ATP-binding DNA structure which is based on a stable framework composed of two stacked G-quartets. The two highly conserved adenosine residues may stack between the top G-quartet and the two short stems, forming a pocket in which the adenosine or ATP ligand binds. Site-directed mutagenesis, base analog substitution studies, and the design of highly divergent but functional sequences provide support for this model.

Repeated cycles of enrichment and enzymatic amplification allow rare molecules with specific properties to be isolated from large populations ($\approx 10^{15}$) of random RNA or DNA sequences [for review see Green et al. (1991) Szostak (1992), and Famulok and Szostak (1992a)]. Through *in vitro* selection, a number of RNA and single-stranded DNA molecules with specific ligand binding activities (aptamers) have been recovered from pools of random-sequence oligonucleotides. These aptamers have been selected for their ability to bind a variety of small molecules such as organic dyes (Ellington & Szostak, 1990, 1992), amino acids (Famulok & Szostak, 1992b; Connell et al., 1993; Famulok, 1994), bases and nucleotides (Sassanfar & Szostak, 1993; Jenison et al., 1994), cofactors (Lorsch & Szostak, 1994a), and proteins (Tuerk et al., 1990, 1992; Bock et al., 1992).

ATP is an important substrate in biological reactions, and adenosine is a component of many biological cofactors. To explore the potential ability of RNA to catalyze reactions that involve ATP as a substrate or that use adenosine linked cofactors, an RNA aptamer for adenosine/ATP was isolated by selection for binding to an ATP-agarose column (Sassanfar & Szostak, 1993). This aptamer has a stem-bulge-stem secondary structure and has an equilibrium dissociation constant (K_d)[1] for ATP of 0.7-8 μM, depending upon the specific salt and Mg^{2+} concentrations.

Both DNA and RNA aptamers that bind specific organic dyes have been isolated (Ellington & Szostak, 1990, 1992). In all cases, the RNA versions of DNA aptamers were inactive, and the DNA versions of RNA aptamers were inactive. In line with this trend, it was known that the DNA sequence of the RNA aptamer for ATP was inactive (Sassanfar & Szostak, 1993). We were therefore interested in whether or not a DNA sequence could be found which would bind ATP. If such a DNA sequence existed, it seemed likely that it would form an entirely different structure than that of the RNA aptamer.

In this paper we describe the isolation and characterization of a DNA aptamer that binds adenosine and ATP in solution. Covariation, nucleotide analog substitution, and redesigned aptamers indicate that the structure of this aptamer in solution consists of two small Watson-Crick helices and two G-quartets.

Because of the importance of G-quartet structures to the analysis of this aptamer, we briefly review this field below. The G-quartet structure is very stable and highly symmetric. Four guanosines assemble by hydrogen bonding between the Hoogsteen and Watson-Crick faces of adjacent guanosines. Two to four G-quartets can stack on each other to form aggregates that are stabilized by monovalent cations such as Na^+ and K^+ which are thought to interact with the carbonyl oxygens of the guanosines [for review see Saenger (1984) and Sen and Gilbert (1992)] (Figure 1A).

A number of variants of this basic structure exist, which differ in symmetry, the orientation of adjacent strands, and the glycosidic configuration about the quartets. In the simplest form, four different strands of DNA (Sen & Gilbert, 1988, 1990) or RNA (Kim et al., 1991) each contribute a single guanosine to each quartet. In this configuration the four strands of DNA are parallel, and the guanosines that comprise the quartet are all *anti* in configuration.

G-quartets can also arise from two strands of DNA that assemble such that each strand contributes two guanosines to each quartet (Sundquist & Klug, 1989). This dimerization can yield a variety of different structures. One such structure, which can be viewed as a dimer of hairpin loops, has been characterized by X-ray crystallography; in this structure all

† This work was supported by a grant from Hoechst A.G. to the Department of Molecular Biology, Massachusetts General Hospital. D.E.H. was supported by an NIH Training Grant for one year.

® Abstract published in *Advance ACS Abstracts*, December 15, 1994.

[1] Abbreviations: K_d, equilibrium dissociation constant; PCR, polymerase chain reaction; nt, nucleotide; NMR, nuclear magnetic resonance.

0006-2960/95/0434-656$09.00/0

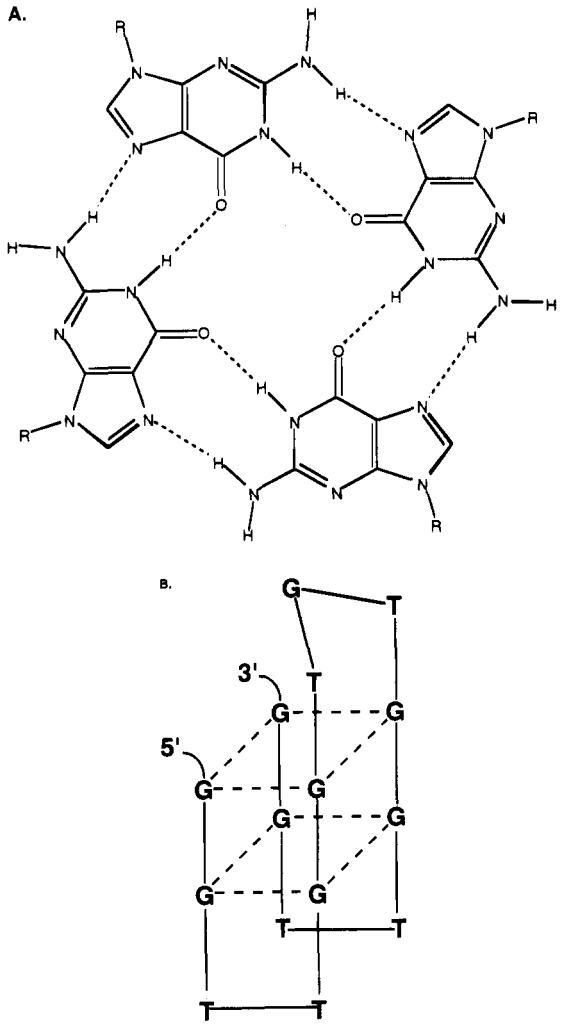

FIGURE 1: Representation of G-quartets. (A) The hydrogen-bonding relationship between the four guanosines of a quartet. (B) Representation of the thrombin aptamer, illustrating the two stacked G-quartets.

adjacent strands are antiparellel, and the guanosines are *syn-anti-syn-anti* around the quartet (Kang et al., 1992). This type of hairpin interaction can also take place in an intrastrand association (Williamson et al., 1989). In a second form of dimer, adjacent strands are from different DNA molecules and are alternately parallel and antiparallel (Smith & Feigon, 1992, 1993). In this structure the guanosines are *syn-syn-anti-anti* around the quartet, and the stack of quartets has minor, intermediate, and major grooves.

A DNA aptamer for thrombin was isolated by Bock et al. (1992). The thrombin aptamer consists of the 15 base sequence 5′-GGTTGGTGTGGTTGG-3′, and the structure of this aptamer was solved (Wang et al., 1993a,b; Macaya et al., 1993) and found to consist of two stacked G-quartets, with all adjacent strands antiparallel.

Stacked G-quartets form a very stable structure that is apparently able to tolerate a wide variation in the details of backbone structure, including strand orientations, glycosidic conformations, and connecting loop lengths. This suggests that G-quartets may be an ideal framework upon which to assemble many different structures, including aptamers of very different specificities.

MATERIALS AND METHODS

RNA Pools. The pool of random-sequence RNA molecules that was used in this work has been previously described (Bartel & Szostak, 1993); the particular pool we used contained $\approx 2 \times 10^{14}$ different molecules with 72 random nucleotides flanked by defined primer binding sites.

The mutagenized pool used in the secondary selection was based on the sequence 5′-GTGCTTGGGGGAGTATTGCG-GAGGAAAGCGGCCCTGCTGAAG-3′, flanked by the same primer binding sites as in the original random-sequence pool. During the synthesis the central domain was mutagenized to an extent of ≈30%, i.e., each nucleotide had a 10% chance of mutating to each other base (Bartel & Szostak, 1993; Green et al., 1991). A total of 480 μg of DNA was recovered after deprotection and purification by polyacrylamide gel electrophoresis; a primer extension reaction using a limiting amount of the 3′ primer indicated that 19% of the DNA could be copied into a complementary strand, implying a yield of 92 μg or 2×10^{15} molecules of amplifiable DNA.

PCR. PCR reactions contained 200 μM of each dNTP, 10 mM Tris-HCl, pH 8.4, 50 mM KCl, 2.5 mM $MgCl_2$, and 2.5 units of Taq polymerase per 100 μL reaction. Primer concentrations were 1 μM crude oligonucleotide or 0.5 μM if gel purified. 5′ primers were: DH22.23, 5′-GGAACAC-TATCCGACTGGCACC-3′ and D33.70; the 3′ primer was D20.99; the latter two primers are described in Bartel and Szostak (1993). Thermal cycling was 94 °C for 45 s, 52−54 °C for 70 s, and 72 °C for 45 s.

Single-stranded DNA was generated by removing 1/10 of a PCR reaction for use as a template in a second PCR reaction with no added 3′ primer. One micromolar of the oligonucleotide DH20.117 5′-CCTTGGTCATTAGGATCC-3′, the complement of the 3′ primer, was added to inactivate the remaining 3′ primer and maximize the fraction of DNA synthesis resulting from extension of the 5′ primer. Radioactively labeled DNA was synthesized by including [α-^{32}P]-dATP in the primer extension reaction. Fifteeen cycles of primer extension were used for single-stranded DNA synthesis.

Amplification of the synthetic DNA pool requires a large PCR volume (10−500 mL) to ensure efficient amplification of all molecules, so that pool complexity is maintained. Thermal cycling of large volumes was done manually in water baths. The DNA was initially melted at 96 °C for 8 min. Cycles of PCR were 94 °C for 4 min, 42 °C for 5−6 min, and 72 °C for 7 min. The final extension was for 20 min. For the first round of the original selection, a 21 mL single-strand synthesis reaction was done for 6 cycles with 60 μg of double-stranded DNA as template. Single-stranded DNA from the mutagenized pool was generated in a 48 mL reaction with 9.2 μg of synthetic DNA as template. This template was amplified for two cycles with both 5′ and 3′ primers, after which the 3'-complement DH20.117 was added to 1 μM, followed by eight cycles of primer extension. After PCR the DNA was phenol/chloroform extracted and gel purified (Sambrook et al., 1989).

Selections. ATP−agarose (Sigma) columns of 1 mL bed volume (1−3 mM ATP) (Figure 2A) were pre-equilibrated with approximately 25 mL of column buffer (300 mM NaCl, 5 mM $MgCl_2$, 20 mM Tris, pH 7.6). Single-stranded DNA (150 μg in the first round of selection and thereafter 1−5 μg, 0.03−0.13 nmol, labeled with [α-^{32}P]dATP as described above) in the same buffer was heated to 75 °C for 5 min, cooled to room temperature for 20−30 min, and loaded onto the column. After equilibrating for ≈5 min, the column was washed with 5 mL of column buffer (10 mL during selection

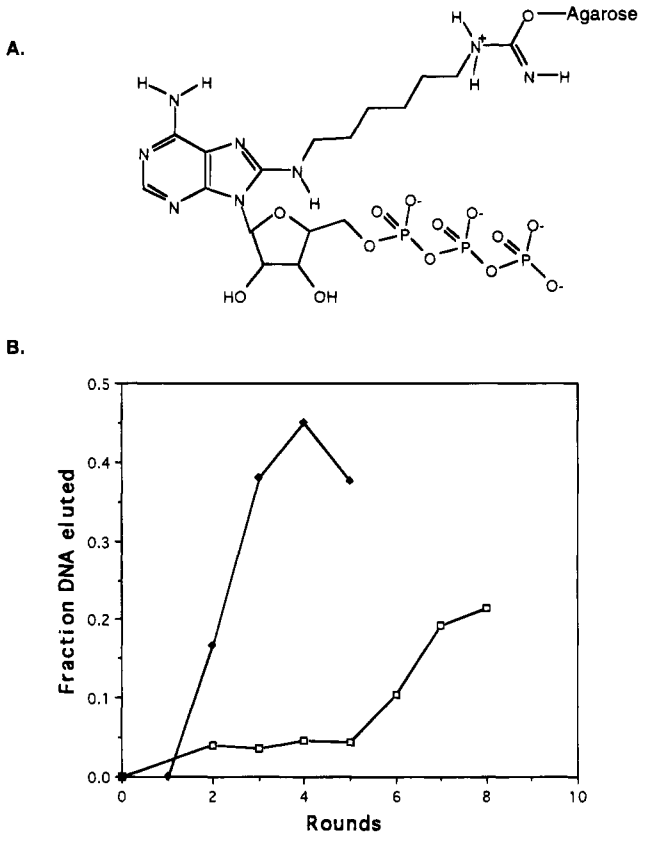

FIGURE 2: (A) Matrix used for affinity selection. The ATP is linked at the C8 of the purine ring through a 9 atom spacer to agarose. (B) Fraction of loaded DNA specifically eluted from the column vs selection round. Open boxes: original selection from the random-sequence pool. Fraction of DNA remaining on the ATP column after washing with 5 column volumes of buffer, and eluted with 3 column volumes of buffer plus ATP. Data point at 8th round represents polyacrylamide gel-purified DNA; non-gel-purified DNA showed ≈17% retention on the column. Closed diamonds: secondary selection from the mutagenized sequence pool. Fraction of DNA remaining on the ATP column after washing with 10 column volumes of buffer, and eluted with 3 column volumes of buffer plus ATP.

from the mutagenized pool). The remaining DNA was specifically eluted with 3 mL of column buffer containing 3 mM ATP. Each 1 mL fraction was then analyzed by Cerenkov counting. The first two fractions (three in the secondary selection) after starting ligand elution were collected and ethanol precipitated, and residual ATP was removed by Sephadex spin-column chromatography. This DNA was then used as a template for PCR amplification.

Preparation of Synthetic Oligonucleotides for Binding Assays. DNA was synthesized by solid phase β-cyanoethyl phosphoramidite chemistry on a Milligen/Biosearch Model 8750 or Millipore Expedite automated synthesizer. Nucleoside phosphoramidite analogs (Glen Research) were coupled manually to conserve material. Deprotected synthetic DNA was extracted with 1-butanol (Sawadago & Van Dyke, 1991) and purified by polyacrylamide gel electrophoresis. DNA (50−75 μg, 6−9 nmol) was 5′ phosphorylated with ^{32}P using T4 polynucleotide kinase (New England Biolabs) and then affinity purified on an ATP−agarose affinity column. This purification step was necessary because of variation in the quality and purity of synthetic DNA. The ATP eluted DNA was ethanol precipitated and resuspended in H_2O, and residual ATP was removed by Sephadex spin-column chromatography.

The complex of oligonucleotides DH20.155 and DH21.82 was assembled by heating 14 μM ^{32}P-labeled DH20.155 and 270 μM DH21.82 to 75 °C for 5 min followed by cooling to room temperature.

Binding Assays. Affinity purified DNA (10−500 ng, 1.2−60 pmol) was loaded onto a 1/4 mL bed volume ATP−agarose column pre-equilibrated with column buffer. After 2−5 min, the column was washed with 10 column volumes of buffer. The remaining bound DNA was eluted with 3 column volumes of 3 mM ATP in the same buffer. Each fraction was then analyzed by Cerenkov counting. The number of counts in the specifically eluted fractions was then divided by the total loaded onto the column to give the fraction eluted.

Determination of the Equilibrium Dissociation Constants. The K_d for adenosine was measured by analytical ultrafiltration (Jenison et al., 1994). In this technique bound and unbound ligands partition across a membrane in an ultrafiltration device. DNA (0.06−30 μM) was incubated for 5 min at room temperature in 200 μL of column buffer containing 5.49 nM [2,8,5′-^{3}H]adenosine. The reaction mixture was loaded onto a Microcon 10 microconcentrator (Amicon) and spun at 14 000 rpm in an Eppendorf 5415 centrifuge for 4 min, resulting in 100 μL of ultrafiltrate. This method could not be used with ATP because only 30% of ATP passed through the filter in 50% of the volume. Even with adenosine, only 40−45% passed through in 50% of the volume. The K_d was estimated from the concentration of DNA calculated to decrease the amount of adenosine passing through the filter by 50%. During the course of the experiment the DNA concentration increases by a factor of 2 since the DNA is retained above the filter; therefore, using the initial DNA concentration would give an artificially low K_d, and using the final DNA concentration would give an artificially high K_d. The K_d reported is the average of these two estimates. At 12 μM DNA the fraction of adenosine retained was independent of adenosine concentration over a range from 1.65 to 165 nM.

Cloning and Sequencing. PCR-amplified DNA from the round 8 pool of the primary selection was cloned by restriction enzyme cleavage at the *Eco*RI (added by PCR with primer DH31.41 5′ GCG*GAATTC*GGAACACTATC-CGACTGGCACC-3′) and *Bam*HI sites in the primer regions, followed by ligation with an *Eco*RI−*Bam*HI-digested vector. PCR amplified DNA from round 4 of the secondary selection was cloned by the T/A cloning method using a kit manufactured by Invitrogen. Sequencing was by the dideoxy method using Sequenase from United States Biochemical Co. using their protocols.

RESULTS

Selection of Single-Stranded DNA Molecules That Bind ATP. DNA sequences that bind ATP were isolated by repeated rounds of enrichment by affinity chromatography on ATP−agarose followed by PCR amplification and single-strand synthesis. The initial population of single-stranded random-sequence DNA molecules was generated by six cycles of 5′ primer extension on a pool of ≈2 × 10^{14} PCR-amplified dsDNA molecules (Bartel and Szostak, 1993) consisting of 72 random nucleotides flanked by 20 nucleotide primer binding sites. Approximately 150 μg of the single-stranded DNA pool was used for the first round of selection, corresponding to an average of 10 copies of each sequence.

The DNA, after equilibration in column buffer, was loaded onto a 1 mL column of cross-linked agarose containing 1−3 mM ATP linked at C8 to the matrix *via* a linker (Figure 2A). Unbound or weakly bound DNA was washed off of the column with 5 column volumes of buffer. DNA that specifically bound ATP in solution was eluted by washing the column with 3 column volumes of buffer containing 3 mM ATP. The DNA in the first 2 column volumes was concentrated and amplified by PCR, and a new enriched population of single-stranded DNAs was generated by primer extension as above. After 7 rounds of selection, ≈18% of the input DNA bound to the column and was specifically eluted with ATP (Figure 2B). This fraction increased to 21% at the 8th round. It is possible that this apparent plateau at ≈20% binding is due to contamination of the single-stranded DNA with a substantial fraction of double-stranded DNA, or misfolded single-stranded DNA.

The round 8 DNA was cloned, and 17 clones were sequenced. All of these clones had different sequences and appeared to be of independent origin. Single-stranded DNA was generated from two clones and was found to bind to the ATP−agarose column as well as the round 8 pool (≈20%) (Figure 3A). One of these clones (clone 4) was chosen for further analysis. The low fraction bound does not seem to be due to a large fraction of misfolded DNA, as very little flowthrough DNA (≈3%) binds upon retesting, whereas specifically eluted DNA does bind efficiently (≈60%) when retested (Figure 3B). The presence of complementary strand DNA may explain these results. A series of synthetic oligonucleotides were made, covering progressively shorter portions of the clone 4 sequence. These oligonucleotides were assayed for ATP binding, and the ATP-binding domain was localized to a 42 nucleotide sequence contained completely within the originally random region of the clone 4 sequence. The 42 nucleotide sequence bound at least as well as the full length clone.

Conserved and Invariant Bases within ATP-Binding Domain. In order to further define the bases that were critical for ATP-binding activity, we prepared a highly degenerate pool of sequences based upon the 42 nt active sequence, selected active sequence variants from the pool in a secondary selection, and then aligned and compared the sequences of these variants. A synthetic oligonucleotide consisting of the 42 nucleotide sequence mutagenized to an extent of ≈30%, flanked by the same primer binding sites as used before, was used as template for PCR amplification and subsequent single-stranded DNA synthesis. After 4 rounds of selection and amplification carried out as described above, ≈45% of the DNA bound to the ATP−agarose column and was specifically eluted with ATP (Figure 2B). No further enrichment was obtained with additional rounds of selection. PCR-amplified DNA from round 4 was cloned, and 45 clones were sequenced (Figure 4A). Of the 45 sequences, 44 were unique, and one, clone 16, was present twice.

Alignment of these sequences (Figure 4A) revealed two distinct regions of conservation (Figure 4B), flanked by apparently nonconserved sequences. Each conserved region was 6 bases long, was highly guanosine-rich, and ended in a single invariant adenosine residue. We also observed two areas of covariation in the aligned sequences: one immediately inside the conserved G-rich segments, and a second immediately flanking the G-rich regions (Figure 4A, Table 1).

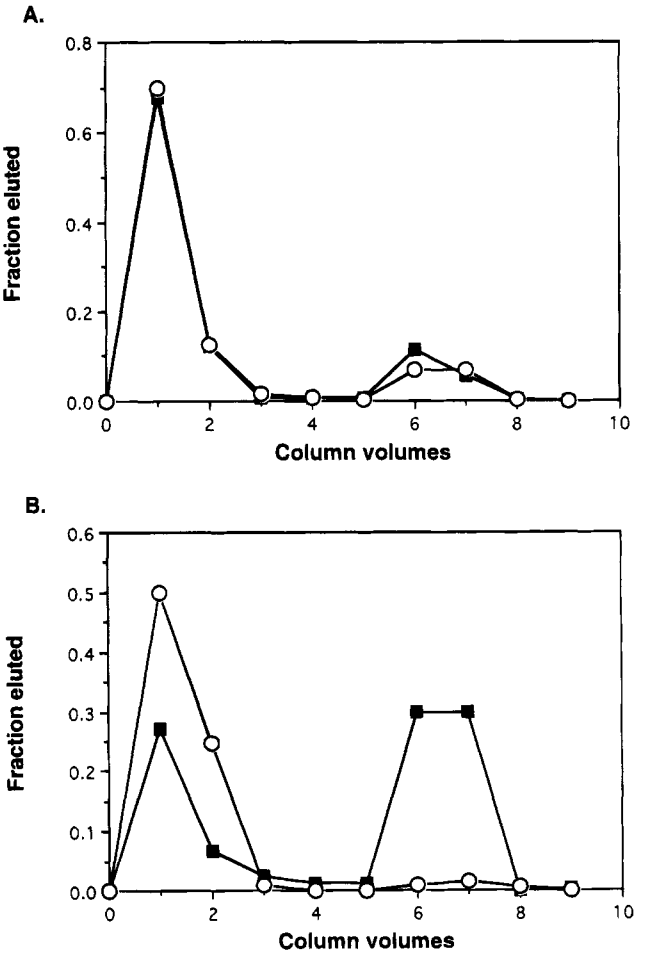

FIGURE 3: Elution profiles of selected DNAs. (A) Column binding profile of the round 8 pool and clone 4. Fraction of loaded DNA eluted is plotted as a function of column volumes of buffer wash; ATP was added to the wash buffer after 5 column volumes. Closed boxes: 8th round pool; open circles: clone 4. (B) Column binding profile of flowthrough and specifically eluted fractions of clone 4 DNA. Clone 4 DNA was chromatographed as above, and the 1st wash fraction and the pooled ATP eluate fractions were precipitated and desalted, and residual ATP was removed by Sephadex spin-column chromatography, heat denatured, and renatured in column buffer. This DNA was then placed on a new ATP column, and the profile is shown. Open circles: clone 4 flowthrough; closed boxes: clone 4 ATP eluate. In each case the data point of the 9th column volume is that which was remaining on the column after all washes.

To see if the conserved and covarying sequences were sufficient for ATP binding, a 25 base oligonucleotide (DH25.42; Figure 5A) that contained only these regions was synthesized. When this oligonucleotide was assayed for ATP binding, greater then 90% of the DNA remained on the ATP−agarose column after washing with 10 column volumes of buffer and was specifically eluted by ATP.

Specificity of the Ligand/Aptamer Interaction. This minimal ATP aptamer was used to obtain a qualitative view of the specificity of the ligand−aptamer interaction by assessing the ability of various ATP analogs to elute bound aptamer from an ATP−agarose column (Figures 9 and 10). These experiments indicate that N7, N6, and N1 of the adenine moiety of ATP are important for interaction with the aptamer. Substitution or methylation of these positions prevents elution in 3 column volumes of 1.5 mM analog. In contrast, substitutions at C8 did not affect binding, consistent with the fact that the initial selection was for binding to ATP linked to the matrix through C8. Methylation of either C2

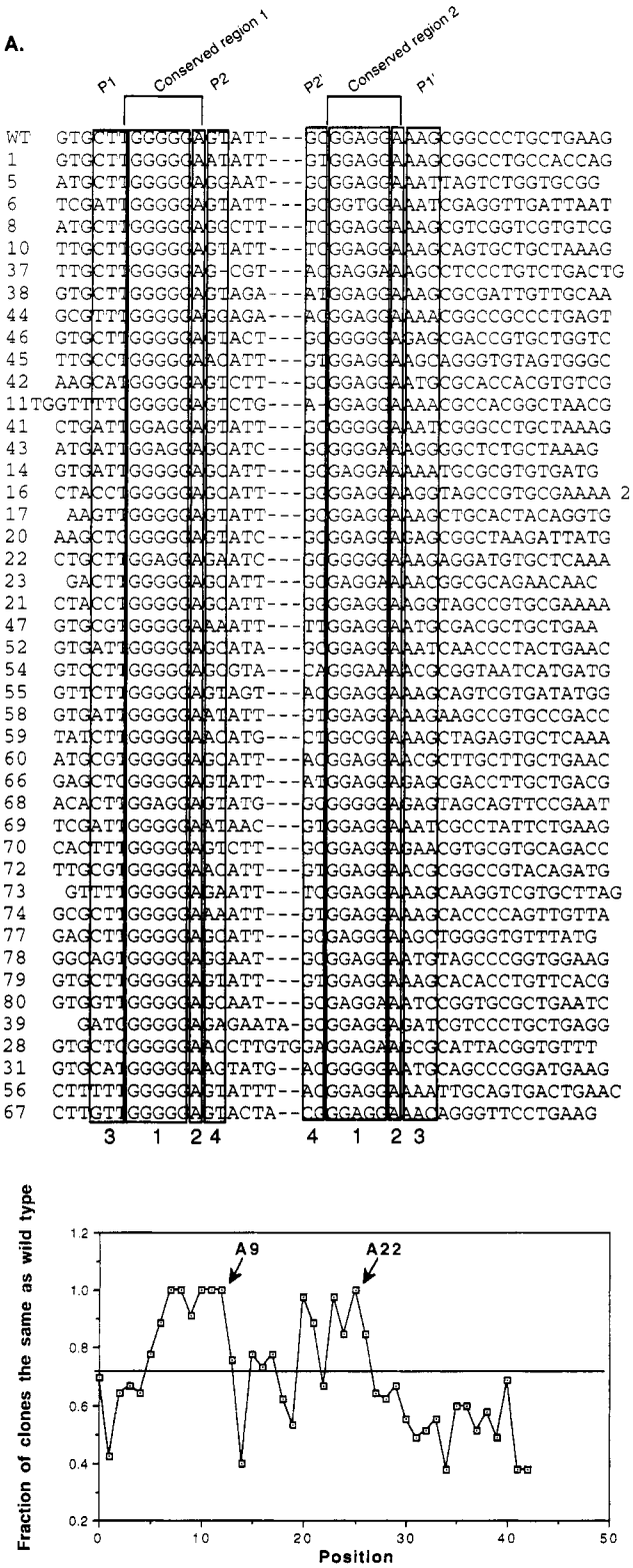

FIGURE 4: Selection of functional ATP-binding sequences from a highly mutagenized sequence pool based upon the DH42.79 sequence. (A) Sequence of 45 clones from the 4th round selection. Box 1: conserved guanosine-rich areas. Box 2: invariant adenosine residues. P1 and P1′ (box 3) and P2 and P2′ (box 4) are complementary regions that show extensive covariation. (B) Graphic representation of sequence conservation after selection from the mutagenized pool. Conservation is based on 45 clones shown above.

or N3 did not prevent aptamer binding. Similar experiments provide evidence for interactions between the aptamer and the ribose moiety of ATP. Removal of the 3′-hydroxyl decreased binding, whereas loss of the 2′-hydroxyl had no effect. However, deletion of both the 2′- and the 5′-hydroxyls of adenosine resulted in a partial loss of binding, suggesting the possibility of an interaction with O5′. The ligand−aptamer interaction is stereoselective with respect to the glycosidic bond, since 9-α-ribofuranosyladenine did not elute bound aptamer from the ATP−agarose column.

Table 1: Mutations in Proposed Stems Conserving Base Pairing[a]

	wild-type	Watson−Crick	G-T	other
C_1−C_{25}	24	11	1	9
C_2−G_{24}	29	8	5	3
T_3−A_{23}	37	4	3	1
G_{10}−C_{16}	28	8	3	6
T_{11}−G_{15}	12	19	1	13

[a] Covariation at positions proposed to base pair. The number of mutant clones out of 45 in each class is shown. In many of the clones the P1 stem was longer than three base pairs.

It is difficult to obtain quantitative binding data from column elution experiments. An estimate of adenosine affinity in solution was obtained by ultrafiltration experiments (Jenison et al., 1994). These experiments indicated that the K_d for adenosine in solution was 6 ± 3 μM.

Proposed Structure. A reasonable hypothesis for the structure of the DNA−ATP aptamer can be made based on the sequence comparison data (Figure 7). The covariation data indicates that bases C1−G25, C2−G24, and T3−A23 are base-paired and are therefore likely to form a short region of double helix (P1). Similarly, G10−C16 and T11−G15 are base-paired and form another short duplex (P2) closed by a three base loop (L2). The two G-rich regions can be viewed as four GG doublets and are consistent with the formation of two stacked G-quartets, as seen in the thrombin aptamer (Bock et al., 1992). However, in our model the loops below the bottom G-quartet (L1 and L3) consist of a single nucleotide (A or G) rather then two T's as in the thrombin aptamer. The highly conserved adenosines, A9 and A22, are stacked between the top G-quartet and the stems P1 and P2.

Length and Sequence Effects in the Base-Paired Stems. A series of oligonucleotides was synthesized (Figure 5B) to determine the minimal and optimal length of each base-paired stem. These experiments indicated that a P1 stem of one base pair was sufficient to detect ligand interaction on the ATP column, but that three base pairs or more were required for optimal binding. A single base pair P2 stem retains some binding activity, but two base pairs are optimal for binding. Increasing the length of stem P2 to three base pairs results in an ≈2-fold decrease in column binding.

Experiments in which the sequences in the P1 and P2 regions were changed, but base-pairing potential was retained, strongly support the hypothesis that these regions of the aptamer are interacting to form small double helix regions. There is no clear indication that a strong sequence preference exists in either of these regions.

Substitutions in the G-Quartet Regions. In our model, eight highly conserved Gs (four GG pairs) assemble to form two stacked G-quartets. To examine in greater detail the contribution of each of these eight residues to the functional aptamer structure, 32 oligonucleotides (Figure 5C) were made with single substitutions of 2′-deoxyadenosine, 2′-deoxyinosine, 7-deaza-2′-deoxyguanosine, and O^6-methyl-2′-deoxyguanosine at each of these positions (Figure 6). All

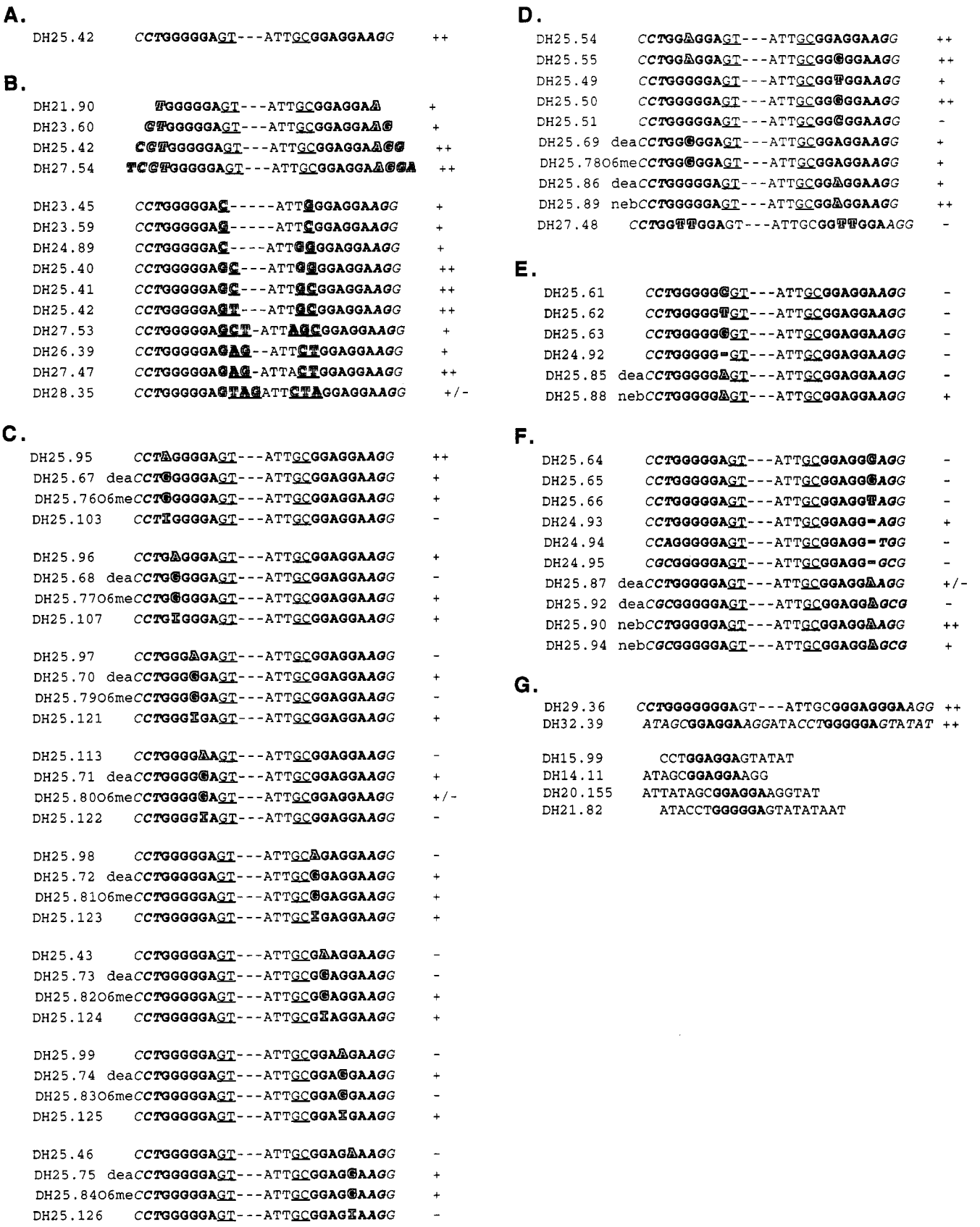

FIGURE 5: Sequences synthesized to test importance of specific bases or functional groups of the aptamer. In all cases the outlined base(s) represents the change being made. Binding efficiency of the aptamer is denoted as a ++ = 95−80% retention on the column, + = 79−20% retention on the column, +/− = 19−1% retention on the column, and − = no column retention. (A) Parental or reference sequence DH25.42 was derived from the sequence of clone 16, the only clone represented twice in the selected pool. (B) Sequences tested for optimal stem lengths. (C) Sequences with changes in highly conserved guanosine pairs. In sequences preceded by "dea" or "O6me" the substitutions were 7-deaza-2′-deoxyguanosine and O^6-methyl-2′-deoxyguanosine, respectively; I represents inosine. (D) Changes in L1 and L3, the single bases bridging the proposed minor grooves of the G-quartet. (E) Changes of invariant adenosine A9. Sequences preceded by "dea" or "neb" are 7-deaza-2′-deoxyadenosine or 2′-deoxynebularine (purine riboside) substitutions, respectively. (F) Changes of invariant adenosine A22. (G) Sequence of engineered aptamers and oligonucleotides used in the heterodimer assembly studies.

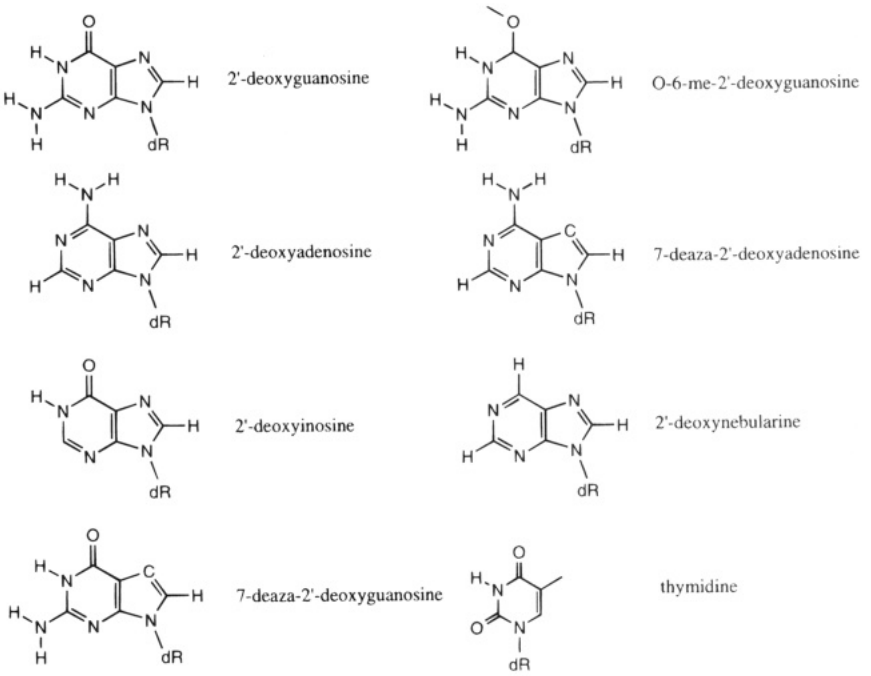

FIGURE 6: The analogs of deoxyadenosine and deoxyguanosine that were used for the substitutions at each of the conserved adenosine residues and the conserved guanosine residues.

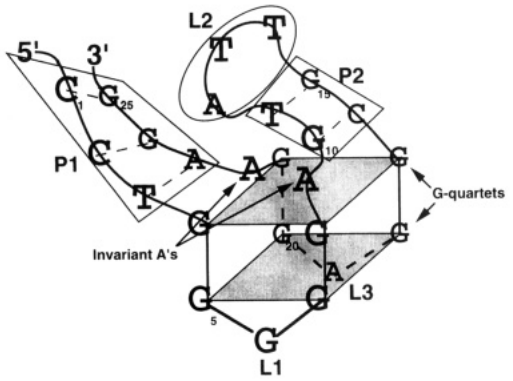

FIGURE 7: Model of structure of ATP aptamer. Base-paired stems are labeled P1 and P2; single base lower loops are L1 and L3; loop closing stem P2 is L2. Two stacked G-quartets act as a platform for two invariant adenosine residues proposed to stack in between upper quartet and stems P1 and P2. The proposed binding site for ATP is above the upper quartet and between the invariant adenosines. This representation is not meant to suggest the exact orientation of any residue.

of the substitutions disrupt hydrogen bonds known to exist in G-quartets.

Substitution of 2′-deoxyadenosine would be expected to lead to loss of at least two hydrogen bonds from a quartet. At 6 of the 8 positions, a G to A change completely eliminated binding in the column assay. At the remaining two positions, corresponding to the first GG pair, a G to A substitution in the putative top quartet G had little effect and a G to A substitution in the putative bottom quartet G decreased but did not eliminate binding. G to T changes at these positions eliminated binding to the ATP column. These results are consistent with the highly conserved nature of the GG pairs seen in the secondary selection from the mutagenized pool.

A G to 2′-deoxyinosine substitution should be less disruptive, resulting in loss of one hydrogen bond, with minimal structural distortion. 2′-Deoxyinosine substitution reduced but did not eliminate binding at five positions and eliminated binding to the column at the remaining three positions (Figure 5C). These three critical positions were all in the putative top quartet. It is possible that the N2 amino groups of G4, G8, and G21 play a role above and beyond G-quartet formation, either directly interacting with ligand or stabilizing the aptamer structure through some tertiary interaction.

Substitution with 7-deaza-2′-deoxyguanosine should also lead to loss of one hydrogen-bond, but with the potential for greater distortion due to a hydrogen-hydrogen steric clash. O^6-Methyl-2′-deoxyguanosine substitution would also disrupt one hydrogen bond in a G-quartet, but would cause even greater distortion due to the bulky methyl group being introduced into the central region of the quartet. Both of these substitutions reduced binding at six positions and eliminated binding at two positions on the lower quartet.

Taken together, these data suggest that N7, O6, and N2 of the conserved guanosine residues play an important role in aptamer structure and/or ligand binding, and are consistent with the assembly of these G residues into G-quartets.

Tests of Proposed Secondary Structure. Two molecules which greatly altered the sequence of the aptamer were synthesized (Figure 5G; Figure 8A,B) to provide independent tests of the proposed secondary structure. In the first test, the sequence was changed so that it would fold into a structure with three G-quartets instead of the original two quartets. To do this, each GG pair was changed to a GGG triplet. This altered sequence binds ATP−agarose as well as the parental two-quartet sequence, providing strong evidence in favor of a folded G-quartet structure.

In a second test, the P1 stem was changed into a stem−loop and the P2 stem−loop was changed into a stem. This results in a permuted sequence in which the original 5′ and 3′ ends are joined with a three base loop and the proposed loop closing P2 is opened to create new 5′ and 3′ ends. In addition, P2 was extended by 3 additional base pairs following a single unpaired adenosine (A12). This molecule has a very different primary sequence, but can be folded into nearly the same secondary structure as the parental aptamer; the altered sequence binds ATP−agarose as well as the

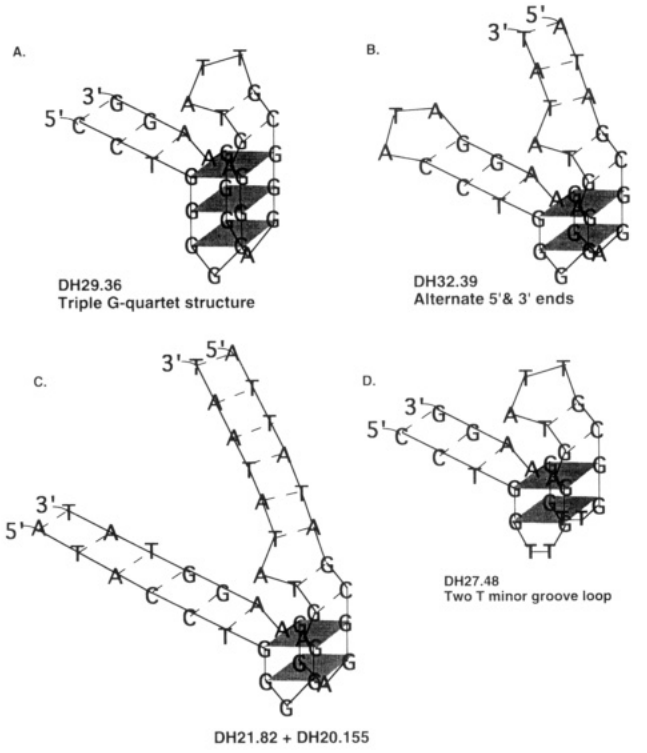

FIGURE 8: Schematic diagrams of modified aptamer structures. Primary sequences are from Figure 5. (A) Predicted structure of DH29.36, with three stacked G-quartets. (B) Predicted structure of DH32.39. Original loop L2 has been deleted, creating new 5′ to 3′ ends, while original 5′ and 3′ ends have been joined by a new loop. (C) Predicted structure formed by assembly of oligonucleotides DH21.82 and DH20.155. (D) Oligonucleotide DH27.48, showing change of each single base minor groove bridge to the TT bridge employed by thrombin aptamer.

parental sequence. Interestingly, removal of the extrahelical adenosine residue results in a partial loss of binding activity, supporting a role for A12 in aptamer structure and/or ligand binding (see below).

A third test of the proposed secondary structure involved the assembly of the complete structure from two oligonucleotides (Figure 5G; Figure 8C). The aptamer structure was divided in two by simply removing L2, the loop that closes the P2 stem. To compensate for the resulting loss in stability, the P1 pairing was extended to 6 base pairs, and P2 was extended to 2 base pairs plus an additional 6 base pairs after the unpaired adenosine. Neither of these oligonucleotides alone bound ATP, but when annealed together, the resulting heterodimer was able to bind to the ATP−agarose column and was specifically eluted with ATP. Oligonucleotides that would pair to make shorter P1 and P2 stems bound more weakly.

Connecting Loops. The first and second GG pairs, and the third and fourth GG pairs, are connected by single purine residues. These single base "loops", referred to as L1 and L3, lie below the bottom quartet and, by analogy with the thrombin aptamer structure, would span the minor groove sides of this quartet. Molecular modeling indicates that a single nucleotide could potentially span the width of the minor grooves (P. Schultze and J. Feigon, personal communication). In the secondary selection from the mutagenized pool, L1 was predominantly a G, and L3 was always a purine. The corresponding loops in the thrombin aptamer are TT doublets. It was therefore of interest to see if changing L1 and L3 to TT would be consistent with ATP binding activity. The corresponding oligonucleotide was synthesized (Figures 5D and 8D), but did not show detectable binding to the ATP−agarose column, suggesting that the

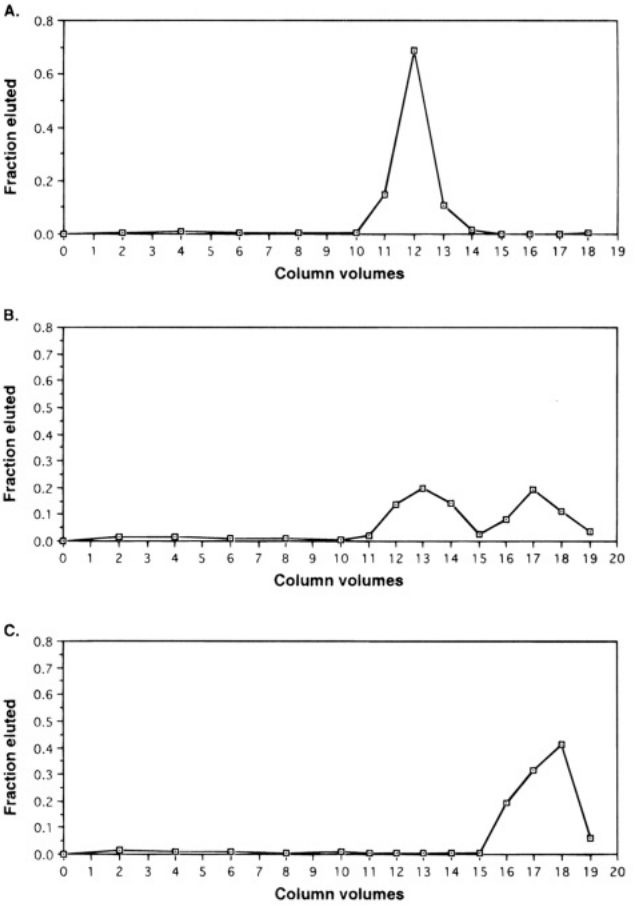

FIGURE 9: Examples of elution profiles with analogs of ATP. DNA was loaded onto an ATP−agarose affinity column and washed with 10 column volumes of wash buffer. Bound DNA was then eluted with 3 column volumes of wash buffer containing 1.5 mM analog, and then washed with 2 more column volumes of wash buffer; the remaining DNA was then eluted in 3 column volumes of wash buffer containing 1.5 mM ATP. Data point at column volume 19 is that remaining on the column after all washes. Elution profiles with (A) adenosine, (B) 3-methyladenine, and (C) purine riboside.

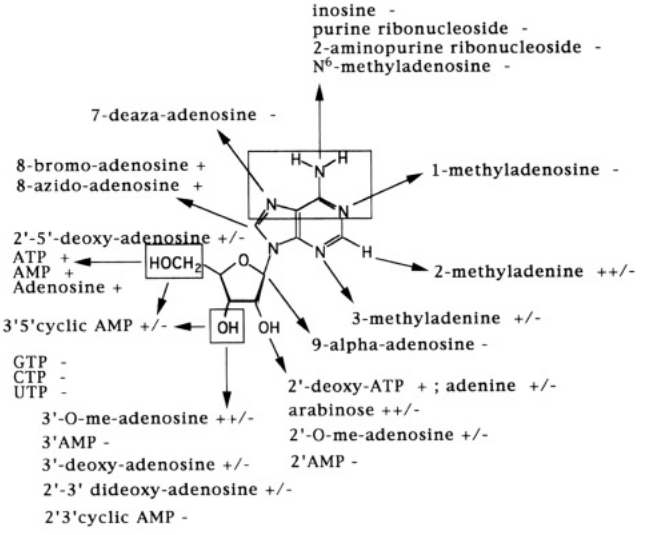

FIGURE 10: Specificity of the adenosine/ATP aptamer, as determined by analog elution experiments. ++, 85−100% of DNA eluted in 3 column volumes of 1.5 mM ligand; +, 70−84% of DNA eluted in 3 column volumes of 1.5 mM ligand; +/−, 6−69% of DNA eluted in 3 column volumes of 1.5 mM ligand; −, <5% of DNA eluted in 3 column volumes of 1.5 mM ligand. The regions most important for binding are highlighted by the shaded boxes.

single base loops may cause a distortion of the G-quartets that is essential to binding.

In contrast to the length and/or sequence specificity of L1 and L3, neither the sequence nor the length of L2, the loop that closes the P2 stem, is critical. This region is highly variable in the secondary selection.

Conserved Adenosine Residues. The invariant nature of A9 and A22 in the selection from the mutagenized pool indicated that these residues were critical for aptamer structure, ligand interaction, or both. Substitution of 2′-deoxycytidine, 2′-deoxyguanosine, or thymidine for A9, or deletion of A9, results in complete loss of binding (Figure 5E; Figure 6). Substitution of 7-deaza-2′-deoxyadenosine (Figure 6) at this position also eliminates binding to the ATP−agarose column, and substitution of 2′-deoxynebularine (purine riboside) (Figure 6) at this position reduces binding.

The corresponding substitutions at A22 gave a more complex pattern. Binding to the ATP−agarose column was eliminated by 2′-deoxycytidine, a 2′-deoxyguanosine, or a thymidine substitution. Remarkably, deletion of A22 reduced but did not eliminate binding; in addition, 7-deaza-2′-deoxyadenosine substitution did not eliminate binding, and the 2′-deoxynebularine substitution had no effect (Figure 5F; Figure 6). A closer inspection of the primary sequence provided a possible explanation for this inconsistency, in that A22 is followed immediately by another A residue in the DH25.42 sequence. Although A23 is normally paired with T3 in the P1 stem, it seemed possible that, in the absence of A22, A23 could replace A22, with P1 being reconstituted as a two base pair stem (C2•G25, T3•G24). To test this hypothesis, a new set of oligonucleotides were synthesized (Figure 5F) which changed the T3•A23 base pair to C3•G23. In this context, substitution of 2′-deoxycytidine, 2′-deoxyguanosine, or thymidine for A22, as well as deletion of A22, eliminated binding. Furthermore, in this context, 7-deaza-2′-deoxyadenosine at position 22 eliminated binding, and 2′-deoxynebularine substitution reduced binding. These results support the involvement of N7 and N6 of both A9 and A22, either directly or indirectly, in ligand binding.

DISCUSSION

We have isolated a single-stranded DNA aptamer which binds adenosine and ATP in a highly specific manner. It is of interest to compare this molecule with the previously isolated RNA aptamer for ATP, with other previously isolated DNA aptamers, and more generally, to compare RNA and DNA aptamers.

The RNA and DNA aptamers for ATP are of similar size (minimal functional structures of 32 and 25 nucleotides, respectively) and similar affinity and specificity for ATP. The similar affinities (1−10 μM) probably reflect the particular selection conditions: sequences with weaker binding would have been lost from the population over several rounds of selection. Both aptamers bind ATP, AMP, and adenosine with similar affinity, suggesting that the negatively charged triphosphate does not interact with either aptamer and may be oriented away from the negatively charged RNA or DNA backbone. Adenosine recognition by both aptamers appears to involve interactions with functional groups on both the base and the sugar, although the details differ (for example, the 2′-hydroxyl of adenosine is important for recognition by the RNA aptamer, while the 3′-hydroxyl is more important for recognition by the DNA aptamer, and N7 of the adenine moiety is important for recognition by the DNA aptamer but is not required for recognition by the RNA aptamer).

The RNA and DNA aptamers are of similar informational content. The information required to specify the RNA structure can be estimated by considering the number of invariant bases (7 invariant bases in the 11 nt loop, and one invariant unpaired G), the number of semiconserved bases (4 in the 11 nt loop), and the number of non-sequence specific base pairs required (8 base pairs to form two 4 base pair stems). The chance of finding such a sequence in a random 32 nucleotide sequence is $4^{-8} \times 2^{-4} \times 4^{-8}$, or 4^{-18}. Similarly, the DNA aptamer has 10 invariant bases (2 G-quartets and 2 invariant A's), 2 semiconserved bases, and 5 base pairs, for a probability of occurrence in a random 25-mer of $4^{-10} \times 2^{-2} \times 4^{-5}$, or 4^{-16}. These estimates provide only a rough comparison, because of the numerous possibilities for variable length stems and loops. However, it appears that RNA and DNA do not differ grossly in the complexity of the structures that are required to form a functional ATP-binding site. These results are consistent with earlier results of Ellington and Szostak (1992), who compared RNA and DNA selections for binding to several organic dyes and found that in some cases DNA aptamers were more common than RNA aptamers, while in other cases the reverse was true.

There are also interesting differences between the RNA and DNA aptamers for ATP. Although RNA and DNA aptamers differ only in the presence or absence of 2′-hydroxyls, a DNA version of the RNA aptamer is nonfunctional and an RNA version of the DNA aptamer is nonfunctional. This small chemical difference must therefore be compensated for by the formation of completely different secondary and tertiary structures: the RNA aptamer has a stem−bulge−stem secondary structure, while the DNA aptamer appears to be formed around a framework of two G-quartets. The ATP-binding site in the RNA aptamer is likely to be formed from the folded structure of the 11 nt loop, whereas in the DNA aptamer, the binding site is likely to be in a cleft between two stems stacked on top of a G-quartet. Although the number of examples is still small, it appears that, overall, the 2′-hydroxyl of RNA does not make it either easier or harder to form stable, folded ligand-binding structures.

The structure of the DNA aptamer for ATP appears to be remarkably similar to that of the DNA aptamer for thrombin (Bock et al., 1992). The solution structure of the thrombin aptamer has been solved by NMR (Wang et al., 1993a,b; Macaya et al., 1993). These studies indicate that the 15 base sequence 5′-GGTTGGTGTGGTTGG-3′ forms two stacked G-quartets, with all adjacent strands antiparellel and a *syn-anti* alternation of G's around each quartet and along each strand. This structure has two minor and two major grooves, arranged such that the minor grooves are spanned by TT loops below the lower quartet, while the TGT loop spans the major groove above the top quartet. Our model for the structure of the DNA aptamer for ATP invokes the same stacked G-quartets, which are formed in the same way by the folding of the DNA strand. The differences lie only in the connecting loops, with single bases below the bottom quartet spanning the minor grooves, and the major grooves being connected by short base-paired stems above the upper quartet. However, the structural similarity may not translate

into similarity in the mechanism of molecular recognition. Thrombin is thought to recognize its aptamer largely *via* the arrangement of phosphates on the outside of the aptamer (Paborsky et al., 1993), while we have proposed that the ATP aptamer recognizes its ligand by hydrogen-bonding interactions with conserved bases (A9 and A22) located in the interior of a cleft formed between the two stems stacked above the top quartet. In both cases, the stacked G-quartets act as a stable structural framework upon which important functional groups can be precisely arrayed. This motif may be able to act as a scaffold upon which many different binding sites could be assembled.

Our model for the structure of the ATP aptamer is based largely upon sequence conservation, covariation, and the effects of mutations and functional group substitutions. The base-paired stems P1 and P2 are strongly supported by the functional equivalence of different sequences that retain complementarity. The G-quartet structure is supported by the strong conservation of the 4 GG doublets, and by the fact that they can be replaced by four GGG triplets. We propose that the invariant A9 and A22 residues are stacked in between the top quartet and the P1 and P2 stems, just as unpaired purine residues stack into the B form double helix rather then sticking out into solvent (Patel et al., 1982; Hare et al., 1986; Woodson & Crothers, 1988; Nikonowicz et al., 1989, 1990; Rosen et al., 1992). This location for the invariant A residues suggests that they may be ideally positioned to form the actual adenosine-binding site. The fact that the N7 and N6 of both adenosine residues are necessary for ligand−aptamer interaction could then be explained as the result of a direct interaction of these functional groups with the bound adenosine. The resulting A-A-A triplet would be similar to a single triplet from a U-A-U or I-A-I triple helix (Blake et al., 1967; Massoulie, 1968; Arnott et al., 1976). The proposed interaction is also similar to that observed between two adenosines across from each other in a double helix (Joshua-Tor et al., 1992). These adenosines stack into the helix and interact in a reverse Hoogsteen manner, with N7 and N6 of one adenosine hydrogen bonding to N1 and N6 of the other adenosine.

The specificity of ATP binding exhibited by the DNA aptamer raises the possibility that some DNA sequences might be able to stabilize reaction transition states with respect to ground state structures and, thus, exhibit catalytic function. Furthermore, the RNA aptamer for ATP has been used as a starting point for the selection of ribozymes with polynucleotide kinase activity (Lorsch & Szostak, 1994b). Now that a DNA aptamer for ATP has been isolated, similar selections for catalytic DNAs can also be attempted.

ACKNOWLEDGMENT

We thank D. Bartel for the random sequence pool and R. Roberts and D. Bartel for helpful comments on the manuscript.

REFERENCES

Arnott, S., Bond, P. J., Selsing, E., & Smith, P. J. C. (1976) *Nucleic Acids Res. 3*, 2459−2470.
Bartel, D. P., & Szostak, J. W. (1993) *Science 261*, 1411.
Blake, R. D., Massoulie, J., & Fresco, J. R. (1967) *J. Mol. Biol. 30*, 291−308.
Bock, L. C., Griffin, L. C., Latham, J. A., Vermaas, E. H., & Toole, J. J. (1992) *Nature 355*, 564−566.
Connell, G. J., Illangesekare, M., & Yarus, M. (1993) *Biochemistry 32*, 5497−5502.
Ellington, A. D., & Szostak, J. W. (1990) *Nature 346*, 818−822.
Ellington, A. D., & Szostak, J. W. (1992) *Nature 355*, 850−852.
Famulok, M. (1994) *J. Am. Chem. Soc. 116*, 1698−1706.
Famulok, M., & Szostak, J. W. (1992a) *Angew. Chem., Int. Ed. Engl. 31*, 979−988.
Famulok, M., & Szostak, J. W. (1992b) *J. Am. Chem. Soc. 114*, 3990−3991.
Green, R., Ellington, A. D., Bartel, D. P., & Szostak, J. W. (1991) *Methods: A Companion to Methods in Enzymology*, Vol. 2, No. 1, pp 75−86, Academic Press, New York.
Hare, D., Shapiro, L., & Patel, D. J. (1986) *Biochemistry 25*, 7456−7464.
Jenison, R. D., Gill, S. C., Pardi, A., & Polisky, B. (1994) *Science 263*, 1425−1429.
Joshua-Tor, L., Frolow, F., Appella, E., Hope, H. D. R., & Sussman, J. L. (1992) *J. Mol. Biol. 225*, 397−431.
Kang, C., Zhang, X., Ratliff, R., Moyzis, R., & Rich, A. (1992) *Nature 356*, 126−131.
Kim, J., Cheong, C., & Moore, P. B. (1991) *Nature 351*, 331−332.
Lorsch, J. R., & Szostak, J. W. (1994a) *Biochemistry 33*, 973−982.
Lorsch, J. R., & Szostak, J. W. (1994b) *Nature 371*, 31−36.
Macaya, R. F., Schultze, P., Smith, F. W., Roe, J. A., & Feigon, J. (1993) *Proc. Natl. Acad. Sci. U.S.A. 90*, 3745−3749.
Massoulie, J. (1968) *Eur. J. Biochem. 3*, 439−447.
Nikonowicz, E., Roongta, V., Jones, C. R., & Gorenstein, D. G. (1989) *Biochemistry 28*, 8714−8725.
Nikonowicz, E. P., Meadows, R. P., & Gorenstein, D. G. (1990) *Biochemsitry 29*, 4193−4202.
Paborsky, L. R., McCurdy, S. N., Griffin, L. C., Toole, J. J., & Leung, L. L. K. (1993) *J. Biol. Chem. 268*, 20808−20811.
Patel, D. J., Kozlowski, S. A., Marky, L. A., Rice, J. A., Broka, C., Itakura, K., & Breslaur, K. J. (1982) *Biochemistry 21*, 445−451.
Rosen, M. A., Live, D., & Patel, D. J. (1992) *Biochemistry 31*, 4004−4014.
Saenger, W. (1984) *Principles of Nucleic Acid Structure*, Springer-Verlag, New York.
Sambrook, J., Fritsch, E. F., & Maniatis, T. (1989) *Molecular Cloning*, 2nd ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
Sassanfar, M., & Szostak, J. W. (1993) *Nature 364*, 550−553.
Sawadogo, M., & Van Dyke, M. W. (1991) *Nucleic Acids Res. 19*, 674.
Sen, D., & Gilbert, W. (1988) *Nature 334*, 364−366.
Sen, D., & Gilbert, W. (1990) *Nature 344*, 410−414.
Sen, D., & Gilbert, W. (1992) *Methods Enzymol. 211*, 191−199.
Smith, F. W., & Feigon, J. (1992) *Nature 356*, 164−168.
Smith, F. W., & Feigon, J. (1993) *Biochemistry 32*, 8682−8692.
Sundquist, W. I., & Klug, A. (1989) *Nature 342*, 825−829.
Szostak, J. W. (1992) *Trends Biochem. Sci. 17*, 89−93.
Tuerk, C., & Gold, L. (1990) *Science 249*, 505−510.
Tuerk, C., MacDougal, M., & Gold, L. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89*, 6988−6992.
Wang, K. Y., Krawczyk, S. H., Bischofberger, N., Swaminathan, S., & Bolton, P. H. (1993a) *Biochemistry 32*, 11285−11292.
Wang, K. Y., McCurdy, S., Shea, R. G., Swaminthan, S., & Bolton, P. H. (1993b) *Biochemistry 32*, 1899−1904.
Williamson, J. R., Raghuaman, M. K., & Cech, T. R. (1989) *Cell 59*, 871−880.
Woodson, S. A., & Crothers, D. M. (1988) *Biochemistry 27*, 3130−3141.

BI942033V