Functional Requirements for Specific Ligand Recognition by a Biotin-Binding RNA Pseudoknot†

Charles Wilson,* Jay Nix, and Jack Szostak‡

Department of Biology and Center for the Molecular Biology of RNA University of California at Santa Cruz, Santa Cruz, California 95064

Received June 10, 1998; Revised Manuscript Received August 13, 1998

ABSTRACT: Ligand-binding RNAs and DNAs (aptamers) isolated by in vitro selection from random sequence pools provide convenient model systems for understanding the basic relationships between RNA structure and function. We describe a series of experiments that define the functional requirements for an RNA motif that specifies high-affinity binding to the carboxylation cofactor biotin. A simple pseudoknot containing an adenosine-rich loop accounts for binding in all independently derived aptamers selected to bind biotin, suggesting that it alone represents a global optimum for recognition of this particular nonaromatic, electrostatically neutral ligand. In contrast to virtually all previously identified aptamers, unpaired nucleotides make up a small fraction of the binding motif. Instead, the identity of 14 nucleotides involved in base pairing is highly conserved among functional clones and their substitution by nonidentical base pairs significantly reduces or eliminates binding. Chemical probing is consistent with the predicted pseudoknot motif and indicates that relatively little change in structure accompanies ligand binding, a strong contrast with results for other aptamers. Competition experiments suggest that the aptamer recognizes all parts of the biotin ligand, including its thiophane ring and fatty acid tail. Two alternative modes of binding are suggested by a three-dimensional model of the pseudoknot, both of which entail significant interactions with base-paired nucleotides.

Recent experiments have shown that it is possible to obtain RNA receptors for a variety of small molecule ligands from large (10^{13}−10^{15}) pools of random sequence molecules (*1*−*7*). RNAs that specifically recognize ATP, tryptophan, arginine, guanosine, cyanocobalamin, and an assortment of dye molecules have been isolated. The binding domains of these RNAs typically span up to 30 nucleotides, and their structures generally consist of loops and short bulges presented in specific secondary structural contexts. Binding is generally quite tight (K_Ds range from 30 nM to >20 μM) and highly specific (receptors being capable of discriminating between ligands that differ by a single atom). The successful isolation of these aptamers has bolstered RNA origin of life theories by suggesting that RNA is broadly capable of molecular recognition and thus primitive ribozymes may have specifically interacted with a wide range of substrates.

Structural studies of RNAs isolated by in vitro selection have begun to address the basic mechanisms by which RNAs in general fold into well-defined conformations and tightly interact with specific target molecules. High-resolution NMR structures of aptamers selected for citrulline, arginine, FMN, ATP, aminoglycoside antibiotics, and theophylline have recently been determined and they reveal a wide range of specific tertiary interactions and previously unforeseen distortions in backbone topology (*8*−*13*). In all of these cases, RNA helices are used to present either an internal bulge or a hairpin loop that is principally responsible for binding. Helical residues in these structures do not typically make base-specific contacts with either the aptamer ligand or the conserved unpaired nucleotides and may thus be altered as long as base pairing is preserved. With the exception of citrulline, the ligands for these aptamers contain either planar aromatic ring systems or multiple positively charged functional groups. As such, relatively generic nucleotide base-stacking interactions and simple electrostatic interactions with the charged RNA phosphate backbone offer simple mechanisms for achieving high-affinity binding. Indeed, NMR structures have shown both energetic effects to play a dominant role in binding. The FMN, ATP, and theophylline aptamers all fold in a manner such that their respective ligands neatly sit between aptamer nucleotides to maintain base stacking in a pseudo-helical arrangement (*9*, *14*, *15*). In the contrasting case of the tobramycin−aptamer complex, phosphates from the RNA backbone surround all sides of the trisaccharide antibiotic, positioned to neutralize its charged amines (*12*).

In previous experiments to evolve a self-alkylating ribozyme, we obtained a single biotin-binding aptamer from a pool of 5×10^{14} sequences (*16*). In several respects, this aptamer and its ligand differ significantly from those characterized previously. Mutant analysis suggests that 31 nucleotides define the functional core of the aptamer and that it folds as a pseudoknot to closely resemble retroviral

† This work was supported by grants from the NIH (GM52707) and the Packard Foundation (C.W.) and by a grant from Hoechst AG to the Department of Molecular Biology, MGH (J.S.). J.N. was supported by a GAANN fellowship awarded to UCSC.

* To whom correspondence should be addressed.

‡ Current address: Department of Molecular Biology, Massachusetts General Hospital, Boston, MA 02114.

S0006-2960(98)01371-3 CCC: $15.00
Published on Web 09/19/1998

frame-shifting elements (*17*). Of the conserved nucleotides, 28 are virtually invariant, an unusually large number, especially considering that the aptamer lacks the internal bulges or hairpin loops that are generally used to specify binding. In contrast to most other aptamers, the bulk of these residues correspond to nucleotides predicted to form standard Watson−Crick base pairs. Given their conservation, it would seem that some fraction of nucleotides involved in secondary structure would also play a direct role in tertiary structure formation or in ligand binding. As such, the mechanisms by which this aptamer and its ligand specifically associate may more closely resemble those utilized by complex-forming biological RNAs such as the ribosome or the SRP in which nucleotides involved in Watson−Crick base pairs are known to make specific direct interactions with other molecules.

In addition to these differences with respect to the RNA, the biotin ligand itself is unlike that recognized by most aptamers. Biotin lacks both the planar aromatic rings and charged functional groups known to direct binding in other aptamer−ligand complexes. Instead, the two aliphatic rings that define the headgroup of biotin are locked at approximate right angles, making it impossible for biotin to intercalate between nucleotide bases. These features of biotin clearly present no problem for the protein-based biotin binders, avidin and streptavidin, which bind their ligand with some of the highest affinities measured for any macromolecule−ligand complex. By developing a model of the way biotin binding is achieved by an aptamer and then comparing it to that obtained for avidin or streptavidin (characterized at atomic resolution through X-ray crystallography), we may potentially draw general conclusions about the mechanisms utilized by RNAs and proteins to interact with specific molecular ligands. In the current report, we present detailed characterization of the biotin aptamer, focusing in particular on the basic structure of the aptamer and how this structure may facilitate ligand binding.

MATERIALS AND METHODS

In Vitro Selection. Selection for biotin binding RNAs has been described previously (*16*). Briefly, a pool of $\sim 5 \times 10^{14}$ random RNA molecules generated by in vitro transcription of a synthesized DNA template was resuspended in selection buffer [0.1 M KCl, 5 mM $MgCl_2$, and 10 mM Na-HEPES (pH 7.4)] and allowed to equilibrate at room temperature. After application to a 500 μL equilibrated biotin agarose column, nonbinding RNAs were removed with 15×500 μL selection buffer washes. Biotin-binding RNAs were obtained by eluting with 5×500 μL elution buffer (selection buffer with 5 mM D-biotin, pH 7.4). Affinity-eluted RNAs were pooled and concentrated and then subjected to enzymatic amplification [reverse transcription, PCR, and transcription (*16*)]. The enriched pool of RNAs was gel purified and then subjected to another round of selection and amplification. Biotin-binding RNAs were detected in the fourth round of selection and accounted for >50% of sequences present in the seventh round.

Mapping the Functional Domain by Alkaline Hydrolysis. The 5′- and 3′-boundaries for the biotin-binding domain within the BB8-5 RNA were identified in separate experiments using 3′- and 5′-end-labeled fragments generated by alkaline hydrolysis. ^{32}P labels were introduced into BB8-5 RNA at the 5′-terminus by polynucleotide kinase (using γ-[^{32}P]ATP as a substrate) or at the 3′-terminus by RNA ligase (using cytidine-5′-[^{32}P]-3′-diphosphate as a substrate). A complete set of end-labeled fragments was generated by incubating labeled BB8-5 RNA with 0.1 M $NaHCO_3$, pH 9, for 10 min at 85 °C. After ethanol precipitation and resuspension in selection buffer together with carrier non-hydrolyzed, nonlabeled BB8-5 RNA, these fragments were applied to a biotin agarose column equilibrated in selection buffer. Nonbinding fragments were eluted by extensive washing while binding fragments were eluted by washing with biotin-containing selection buffer. Approximately equal concentrations of fragment RNA from both the flow-through and specific elution fractions were then analyzed by gel electrophoresis. Limited digestion with RNAse T1 (Boehringer Mannheim) was used to generate a sequencing ladder to map the fragments.

Chemical Probing. Reactivity of individual nucleotides to modification reagents was measured in three different buffers. Buffer A contained 50 mM sodium cacodylate, pH 7.4. Buffer B was identical to buffer A but also contained 0.1 M KCl and 5 mM $MgCl_2$. Buffer C was identical to buffer B but also contained 5 mM biotin. Reaction with modification reagents was detected by either primer extension with reverse transcriptase or by aniline-induced cleavage as described below.

Primer Extension Modification Reactions. Biotin aptamer RNA in a volume of 100 μL was incubated with 2 μL DMS (1:1 in 20% ethanol) for 15 min at room temperature. Reactions were terminated by the addition of 13 μL of stop solution (1 μL of tRNA (10 mg/mL) + 12 μL of 5 M ammonium acetate]. Modification reagent was removed by extraction with 200 μL of phenol followed by 96% chloroform in 4% isoamyl alcohol. Samples were then ethanol precipitated and lyophilized. After resuspension in water, a 5′-^{32}P-end-labeled primer (corresponding to the 3′-primer used for SELEX) was annealed with the modified RNA by heating to 65 °C, followed by slow cooling to room temperature. AMV reverse transcriptase (Promega) was used to direct primer extension. Parallel reactions containing dideoxynucleotide chain terminators were used to generate a sequencing ladder for comparison. Reactions were resolved by 10% denaturing gel electrophoresis and visualized by autoradiography.

Cleavage Modification Reactions. Trace, kinased RNA was mixed with 10 μg of carrier tRNA in a total of 5 μL/reaction. Samples (100 μL) containing the RNA mixture and 95 μL buffer A, B, or C were incubated for 10 min at RT and separated into 2×50 μL aliquots. (1.25 μL) DEPC[1] was added to one aliquot and incubated for 15 min at room temperature. A 1:50 dilution of DMS (1.25 μL) was added to the other aliquot and incubated for 8 min. at room temperature. Denaturing controls for both modification reagents were carried out using buffer A with incubation for 1 min at 80 °C. Reactions were quenched by the addition of 1 μL of 10 mg/mL tRNA, 5 μL of 3 M sodium acetate (pH 4.5), and ethanol precipitated. DMS reactions only were resuspended in 10 μL of 1 M Tris-HCl (pH 8.2) and reduced

[1] Abbreviations: DEPC, diethylpyrocarbonate; DMS, dimethyl sulfate.

with 10 μL of freshly prepared 0.2 M sodium borohydride. After incubation on ice in the dark for 30 min, samples were quenched by the addition of 200 μL of cold stop solution (0.6 M sodium acetate and 25 μg/mL tRNA) and precipitated by the addition of 600 μL of ethanol. Both DMS and DEPC reactions were cleaved by resuspension in 20 μL of 1 M aniline acetate (prepared by mixing 9.1 mL of aniline, 10 mL of acetic acid in a final volume of 100 mL), followed by incubation at 60 °C for 20 min in the dark. Cleavage was terminated by freezing the reactions on dry ice and lyophilizing twice. Cleavage fragments were resolved on a 12% denaturing gel and visualized by autoradiography.

K_D Determination. The affinity of the aptamer for biotin was determined by measuring the fraction of RNA bound to biotin agarose at a range of immobilized ligand concentrations. The biotin concentration in the binding experiment was adjusted by simply diluting the biotin agarose with selection buffer to final fractional volumes in the range of 1:2 to 1:40 (volume packed agarose:volume suspension). To minimize nonspecific binding, each diluted matrix suspension was presaturated with selection buffer containing 100 μg/mL tRNA. The concentration of biotin available for binding was determined by saturating biotin agarose with excess labeled aptamer and measuring the amount of RNA specifically eluted following washing. Assuming a single binding site per biotin molecule, the available biotin concentration on the biotin agarose was estimated at 348 nmol/mL undiluted matrix.

For binding studies, 20 μL of diluted matrix and trace amounts (10 μL) of kinased aptamer RNA were added to 0.2 μ spin filters (Costar) and diluted to a final volume ranging 50−1000 μL. Binding was allowed to equilibrate on a rocker for 30 min at room temperature. Following equilibration, the binding reactions were spun in a microcentrifuge for 1 min and immediately washed with 150 μL of selection buffer (total wash time less than 20 s/reaction; control experiments suggest that <3% of specifically bound RNA is removed during the wash). Specifically bound RNA was eluted from the matrix by equilibrating with 500 μL of elution buffer for 30 min, followed by a 150 μL wash with additional elution buffer. These eluates were combined and counted in a scintillation counter. The fraction of RNA specifically eluted as a function of immobilized biotin concentration was plotted and fitted by nonlinear regression to a function of the form

$$f = f_0[\text{biotin}]/[\text{biotin}] + K_D \quad (1)$$

where f is the fraction of input RNA bound to the matrix, K_D is the apparent dissociation constant, and f_0 is the fraction of input RNA capable of folding and binding to the ligand.

Competitive Elution. The binding of ligand analogues to the biotin aptamer was assessed by competitive elution. Radioactively labeled biotin aptamer RNA was passed over a biotin agarose column and washed with 15 column volumes of selection buffer (effectively removing misfolded, nonfunctional molecules). The biotin-immobilized RNA was then washed with 5 mM biotin analogue in selection buffer. Radiolabeled RNA in the collected fractions was quantitated using a scintillation counter and normalized to the amount of RNA eluted with 5 mM biotin.

RESULTS

In Vitro Selection. A pool of efficient biotin binding RNAs was obtained after seven rounds of in vitro selection and amplification as described previously (*16*). To characterize these molecules, corresponding PCR template DNA was cloned into a T-vector and the insert sequences were determined. Of the first 12 clones sequenced, 10 were clearly derived from a single precursor in the original pool. One of these (BB8-5) was chosen as the focus for extensive further study (each of the other related clones contained a few randomly distributed point mutations introduced in the course of in vitro selection). The complete sequence for BB8-5 is shown in Figure 1.

Mapping the Functional Domain. Nucleotides responsible for biotin binding in the BB8-5 clone were identified in two separate experiments. A minimal functional domain was identified by measuring the efficiency of binding of end-labeled RNA fragments generated by alkaline hydrolysis. As indicated schematically in Figure 1, fragments lacking nucleotides downstream of approximately G-15 and upstream of C-75 are incapable of binding efficiently to the biotin agarose, suggesting that the intervening region 15−75 defines a minimal binding domain.

We further characterized the biotin-binding domain by generating a pool of heavily mutagenized sequences based on the original biotin aptamer sequence and then reselecting for RNAs capable of binding to biotin despite the introduction of random mutations (*16*). Functional biotin aptamer mutants isolated after four rounds of reselection were sequenced and analyzed to identify nucleotides that were highly conserved (thus involved in binding in some way) or which covaried with other sites (to identify secondary structural elements). Thirty-one nucleotides were shown to be conserved within the minimal domain, and of these, 28 were completely conserved or substituted in only one mutant sequence. Figure 1 summarizes this analysis.

As shown in Figure 1, the nucleotides in this sequence can be modeled as an RNA pseudoknot although there is little covariational evidence to support it. A proposed A·G mismatch at the end of helix 1 is frequently replaced by either an A:U pair or a C:G pair, but no mutations are observed at any of the other sites in the pseudoknot model. To test this structure, we synthesized a series of aptamer minimal domain mutants in which a proposed Watson−Crick base pair was disrupted by substitution of one nucleotide and then restored by a compensatory substitution at the proposed pairing partner. As shown in Figure 2, single base substitutions that disrupt Watson−Crick base pairs within the pseudoknot helices eliminate biotin binding while the restoration of base pairing (by effectively substituting one base pair for another) reconstitutes binding, albeit at levels lower than for the wild-type sequence. These findings provide strong support for the pseudoknot model. The reduced binding observed with the doubly substituted mutants may result from two different effects: (1) mutant sequences may tend to misfold or unfold to a greater extent than the wild-type, or (2) the helical core may be directly involved in binding or in making tertiary interactions that define the biotin-binding site. As noted below, we have searched for independently derived biotin aptamers to obtain a clearer sense of the sequence constraints on the pseudoknot.

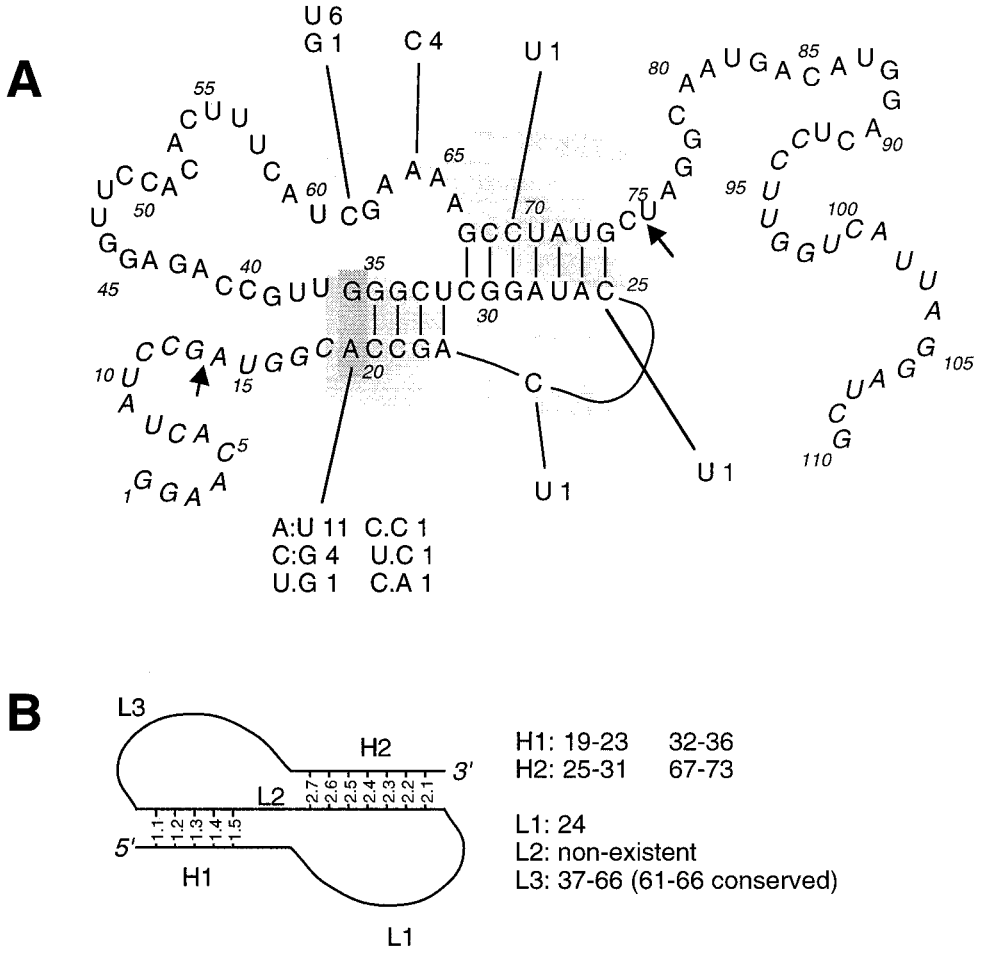

FIGURE 1: (A) Sequence of the major biotin aptamer clone folded as a pseudoknot. The pseudoknot structure was deduced from phylogenetic analysis of 28 clones isolated from a reselected pool as described previously (*16*). All nucleotides within the large shaded box are absolutely conserved in all sequences except as noted in the figure. Covariation is indicated for the boxed nucleotides. Arrowheads indicate the boundaries of the minimal functional fragment defined by alkaline hydrolysis. (B) A generic pseudoknot representation showing the labeling of structural elements used throughout the text.

A distinguishing feature of the biotin aptamer is the 24 nucleotide nonconserved loop lying between the two domains that make up the pseudoknot. Progressive deletion of this loop causes a reduction in the efficiency of binding although a single nucleotide loop is still functional (data not shown). No binding is observed when the loop is effectively cut by transcribing each half of the pseudoknot separately and then mixing the products together (data not shown).

The Density of Biotin Aptamers in RNA Sequence Space and Analysis of Minor Clones. Previous selection experiments have typically yielded a large number of ligand-binding species in the final enriched pool of aptamers (*1, 6, 18*). Extrapolating from these results, it has been argued that, with a probability of $10^{-10}-10^{-11}$, a random RNA sequence is capable of binding an arbitrary ligand with approximately micromolar affinity (*1*). More than 90% of the molecules making up the final selected set of biotin aptamers are derived from a single precursor sequence present in the original pool of 5×10^{14} random RNAs. The probability of finding biotin-binding sequences thus appears to be several orders of magnitude lower than that found in most other cases. Two different hypotheses can account for this observation: (1) high-affinity binding to biotin requires a larger, more highly defined molecular surface than that required for most other aptamer ligands, or (2) early in this particular selection experiment, an artifactual bottleneck significantly reduced the complexity of the selected pool. To distinguish between these possibilities, we have isolated minor sequence species in the selected aptamer pool and used them to estimate the informational complexity of their common structure. As described below, these results argue for a relatively large binding domain that occurs only a handful of times in the sparse random sampling of sequence space in the original pool.

To efficiently characterize minor species in the selected aptamer pool, we eliminated the major clone by completely digesting the final selected pool in its dsDNA form with the restriction enzyme *Eco*NI. The recognition site for this enzyme appears fortuitously in the major clone, only partially overlapping with the conserved region. Remaining full-length molecules were amplified by five cycles of PCR and then ligated into a T-vector. Of 14 clones sequenced from this BB8-5-depleted pool, two correspond to major clone variants in which point mutation has eliminated the *Eco*NI site. The remaining 12 clones consist of three new, independently derived sequences. Within each new sequence, a pseudoknot structure containing an adenosine-rich loop can be readily identified (predicted foldings are shown schematically in Figure 3). To test whether these hypothetical pseudoknots actually define the biotin-binding site of each

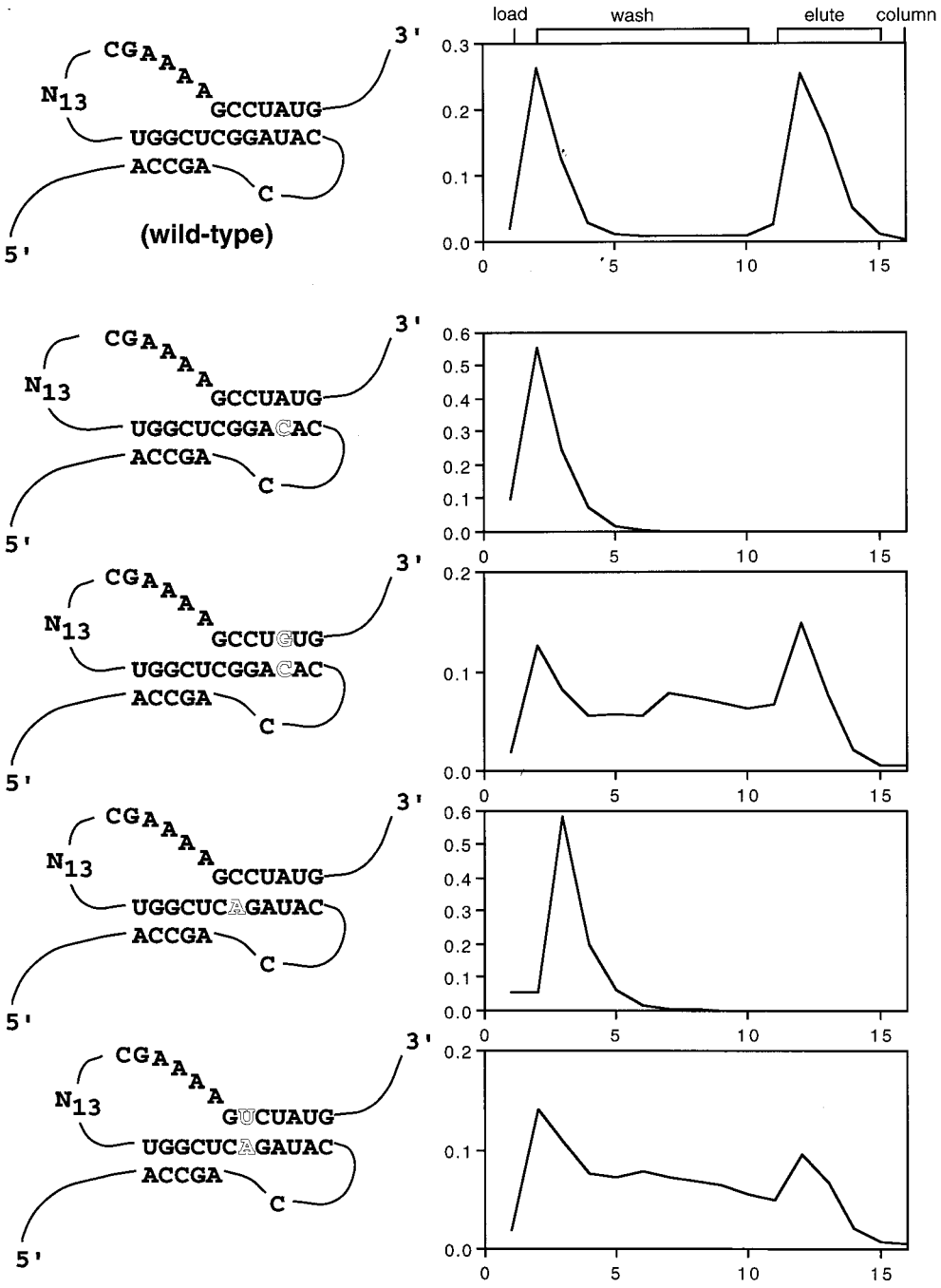

FIGURE 2: Testing the pseudoknot model by site-directed mutagenesis. Variants of the original biotin aptamer at base pairs 2.3 and 2.6 were prepared by in vitro transcription of synthetic DNA templates containing one or two site-specific mutations as indicated. Binding to the biotin agarose column was measured as described in the Materials and Methods.

aptamer, we prepared minimal domains by in vitro transcription and measured their binding to biotin agarose. All of the minimal domains efficiently bound. The behavior of minor clone 3, predicted to carry a U·U mismatch at the end of helix 1, was especially interesting in that it appeared to bind more efficiently to the biotin column than the major clone. The fact that it makes up a relatively small fraction of the selected pool may be explained by its tendency to elute slowly upon washing with biotin. Clearly, successful enrichment of aptamers by affinity elution requires that molecules display not only high-affinity binding but also exhibit relatively fast binding kinetics. As such, potentially tight binders may ultimately be selected against.

The presence of the pseudoknot in all four independently derived clones suggests that this motif is the optimal solution for biotin recognition within the sequence space spanned by the original random pool (corresponding approximately to the set of all contiguous 28-mer sequences). In contrast to the major clone, all of the minor clones have a short connecting loop between the 3′-side of helix 1 and the adenosine-rich bulge. Several specific features of the pseudoknot structure are conserved between the independent isolates, helping to define the attributes responsible for biotin binding. In each case, the sequence immediately upstream of the 3′-side of helix 2 (L3) consists of four adenosines preceded by another purine. L1 always starts with a cytidine,

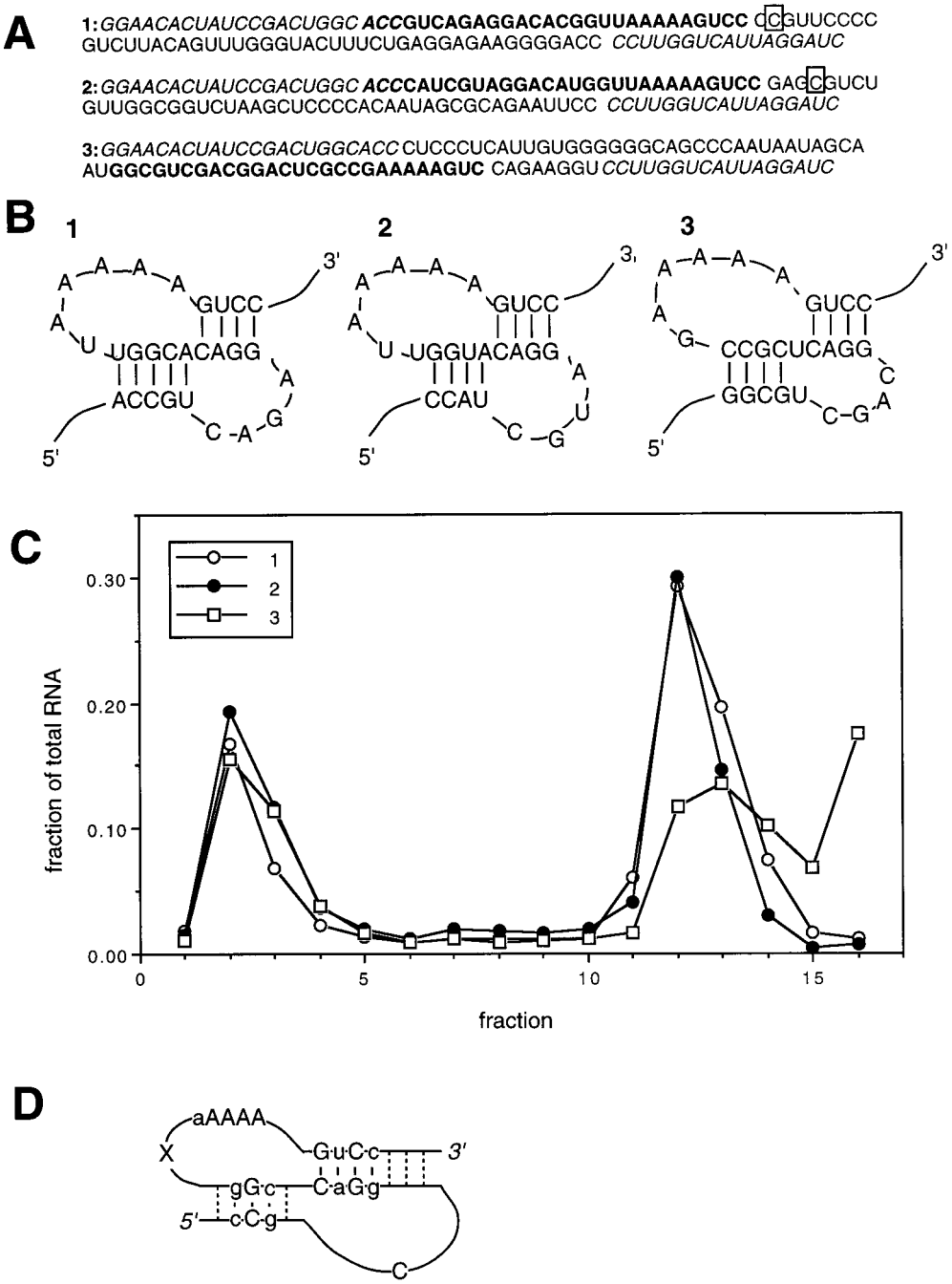

FIGURE 3: Minor clones present in the selected pool. (A) Complete sequences of the minor species in the selected pool with putative binding motifs underlined. Boxed nucleotides indicate 3′-boundaries for the functional aptamer determined by alkaline hydrolysis as described in the Materials and Methods. (B) Predicted folded motifs identified in the minor clone sequences. (C) Binding of the minor clones to biotin agarose. (D) Consensus motif deduced from comparison of all biotin-binding aptamers. Upper case letters are absolutely conserved nucleotides while lower case letters are conserved in three out of four aptamers. Dashed lines indicate the presence of either mismatches or A:U base pairs. X corresponds to an intervening loop, most often a single nucleotide.

and the total connecting loop between the two sides of helix 1 (made up by L1 and the 5′-side of helix 2) is exactly eight nucleotides. The fraction of this loop used to form helix 2 varies from all but one nucleotide in the major clone to only half of the nucleotides in the minor clones. The addition of 5′-UCU-3′ to the 3′-end of minor clone 1 extends its helix 2 to seven base pairs and also increases the fraction of RNA retained on a biotin agarose column to approximately 90% (data not shown), suggesting that stabilization of the pseudoknot core is essential for high-efficiency binding. It is interesting to note that the length of the connecting loop L3 and the length of helix 2 are correlated in all of the clones. One possibility is that additional base pairing in helix 2 is able to compensate for an elongated, unconstrained L3 to maintain efficient pseudoknot folding.

Comparison of the predicted base pairing in the pseudoknot helices of the different clones shows absolute conservation at a handful of positions, although general patterns are strongly preferred everywhere but the first and last base pair of helix 1. On the basis of this comparison, we propose the consensus sequence motif shown in Figure 3D. In summary, helix 1 consists of three central G/C base pairs, flanked on

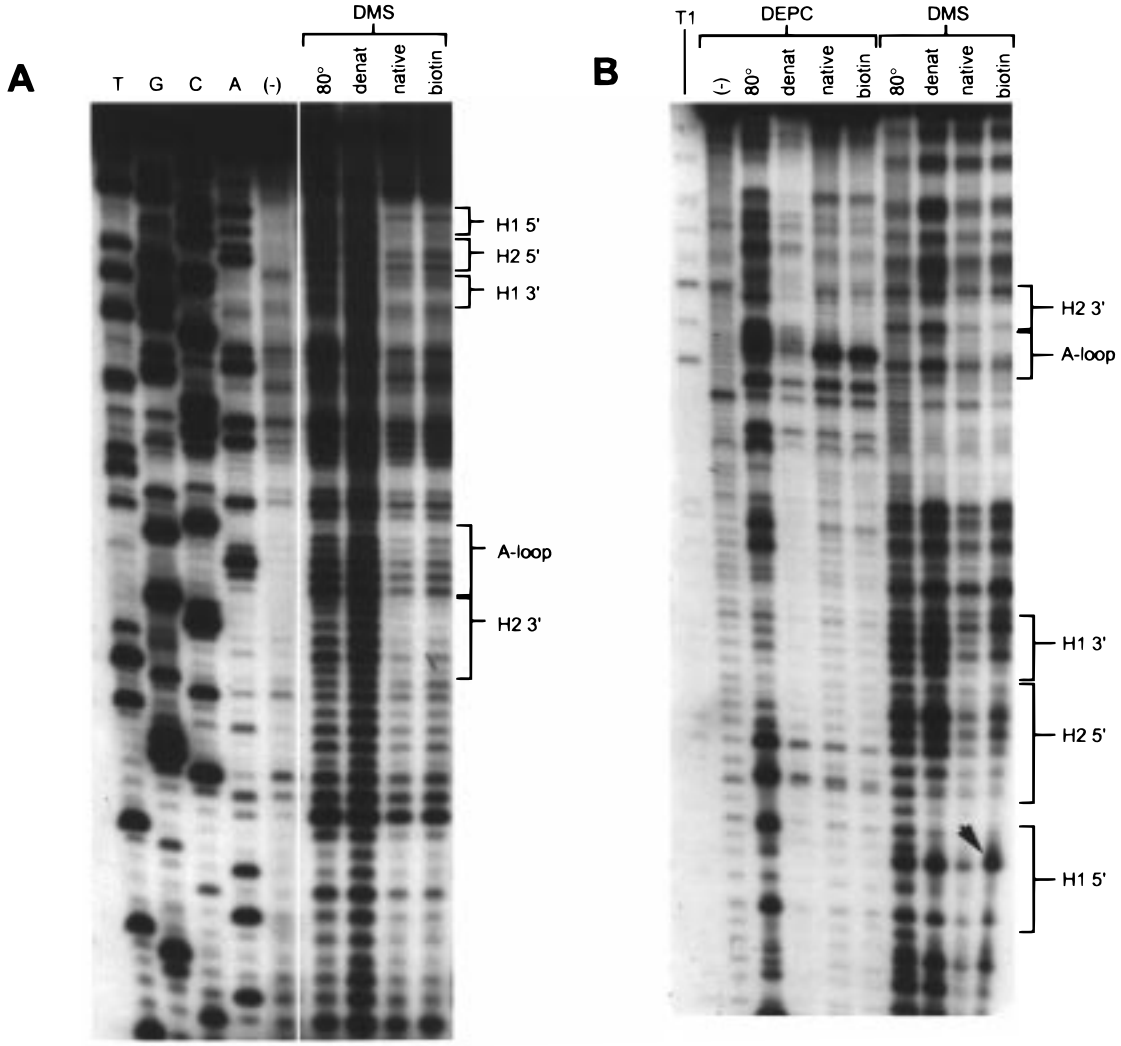

FIGURE 4: Chemical modification. BB8-5 RNA was subjected to modification with either DMS or DEPC and analyzed as described in the Materials and Methods. (A) Results for analysis by reverse transcriptase-catalyzed extension of 5′-end-labeled primer. T,G,C,A: sequencing lanes generated by reverse transcription in the presence of ddATP, ddCTP, ddGTP, and dTTP, respectively; (—) extension of an unmodified RNA substrate, subjected to the workup described for DMS-treated samples. The following lanes correspond to reaction with DMS under four different regimes: 80 °C, samples were incubated in buffer alone at 80 °C; denat, samples were incubated in the absence of salt and magnesium at room temperature; native, samples were incubated in selection buffer at room temperature; biotin, samples were incubated in selection buffer containing 5 mM biotin. (B) Modification reactions analyzed by aniline-induced cleavage (probing N7 accessibility). T1, sequencing ladder generated by limited RNAse T1 digestion. (—) Mock control was not incubated with modification reagent but was worked up as described for DMS reactions. 80°, denat, native, biotin: modification reactions were carried out under four different regimes as described for panel A.

either side by G/C-poor base pairs or mismatches. Base pair 1.3 is C:G in every case while 1.4 is consistently purine: pyrimidine. Helix 2 very closely fits the consensus 5′-GGAC-′3 with positions 2.5 and 2.7 being absolutely conserved and 2.4 and 2.6 accepting purines and pyrimidines with an absolutely conserved pattern. The original pseudoknot mutations shown in Figure 2 do not test this consensus, since base pair 2.3 does not exist in the minor clones and mutation of pair 2.6 leads to its replacement with a base pair known to function well in the minor clones (C:G → U:A). As a simple test of this proposed consensus, we prepared a mutant in which the central C:G base pair of helix 1 was replaced by U:A and assayed its ability to bind biotin agarose. In contrast to the results obtained for base pair substitutions at positions 2.3 and 2.6, mutation at position 1.3 completely abolishes binding with no evidence of even transient retention on the biotin column (data not shown).

Chemical Modification. To obtain information about the conformation of each nucleotide in the pseudoknot and its role in ligand binding, we carried out chemical modification studies on the major clone with the alkyating agents DMS and DEPC (Figure 4). Modification of the N1 position of adenosines and the N3 position of cytidines is probed by reverse transcriptase-catalyzed primer extension of DMS-treated RNA. Modification of the N7 positions of adenosines and guanosines is probed by DEPC and DMS, respectively, when analyzed by aniline-induced cleavage. Protections observed upon addition of salt and magnesium are consistent with the pseudoknot model. The pattern of protections for the A-loop are interesting in that they indicate a highly exposed conformation for the first conserved adenosine (A-63) but a relatively packed and protected conformation at the other sites. Relatively few changes in nucleotide accessibility appear to be induced by ligand binding, a strong

contrast to the results inferred by chemical probing of the ATP and vitamin B12 aptamers (*2*, *6*). The N7 positions of several adenosines in the A-rich loop (A-64−A-66) are somewhat protected in the absence of ligand but become completely protected upon the addition of biotin. The only site that displays ligand-induced hypermodification is the N7-position of guanosine-22 (arrowhead in Figure 4B). Interestingly, this site corresponds to a highly conserved base pair at the core of the pseudoknot. The N3 of cytidine-33 (to which guanosine-22 pairs) is largely protected from modification both in the presence and absence of biotin, suggesting that binding may not disrupt the base pair. As mentioned previously, substitution of the adjacent 1.3 base pair with a nonnative base pair yields a nonfunctional aptamer. Together, these results point to this site within the aptamer as critical for binding of biotin.

The MCSYM algorithm was used to construct a structural model for the biotin aptamer, using a library of standard RNA conformations, and using the secondary structure inferred from phylogenetic analysis as an additional set of constraints (*19*). This algorithm generates all feasible structures that satisfy basic geometric constraints (e.g., loops must be closed and specific van der Waals contacts must be avoided). The sequence modeled corresponds to that for BB8-5 but with the A·G mismatch substituted by the preferred A:U and with the nonconserved nucleotides of L3 absent (closely mimicking then the secondary structure of minor clone 1 with the helix 2 extension). While a large family of structures are possible, most closely resemble that shown in Figure 5.

Several conserved features of the biotin aptamers may be explained in light of this model structure. Helixes 1 and 2 are ideally stacked in a coaxial arrangement such that the single nucleotide in L1 exactly matches the distance required to span from H1 to H2. This arrangement is optimized by the fact that the sum of the lengths of L1 and H2 (spanning from one side of helix 1 to the other) is 8 nucleotides, a conserved feature of all biotin aptamers. This constraint on the length of L1 + H2 has also been noted previously by Du and Hoffman in their studies of retroviral frame-shifting pseudoknots (*20*). NMR experiments with such pseudoknots provide direct physical evidence for coaxial stacking between the constituent helices. Superimposing the chemical modification results upon the structure (color coded in Figure 5) shows that base protections are clustered around helix 1. In the original MCSYM-generated structure, the A-loop nucleotides are perfectly stacked upon each other and run down the minor groove of helix 1. The pattern of protections observed upon modification with DEPC is consistent with this result for the last three nucleotides in the run of adenosines (protection of N1, slight exposure of N7) but not for the first. This discrepancy suggests the alternate structure shown in Figure 5 in which A-63 is swung out to simultaneously protect the N7 position while exposing the N1 position. Guanosine-22, the only nucleotide whose accessibility increases upon addition of biotin, lies in the center of an otherwise highly protected core.

Strength and Specificity of Binding. The strength of aptamer binding to immobilized ligand was quantified by measuring the dissociation constant as described in the Methods and Materials. Briefly, biotin agarose was diluted 1:2, 1:10, 1:20, and 1:40 with selection buffer. Equal volumes of each diluted matrix were then brought up to

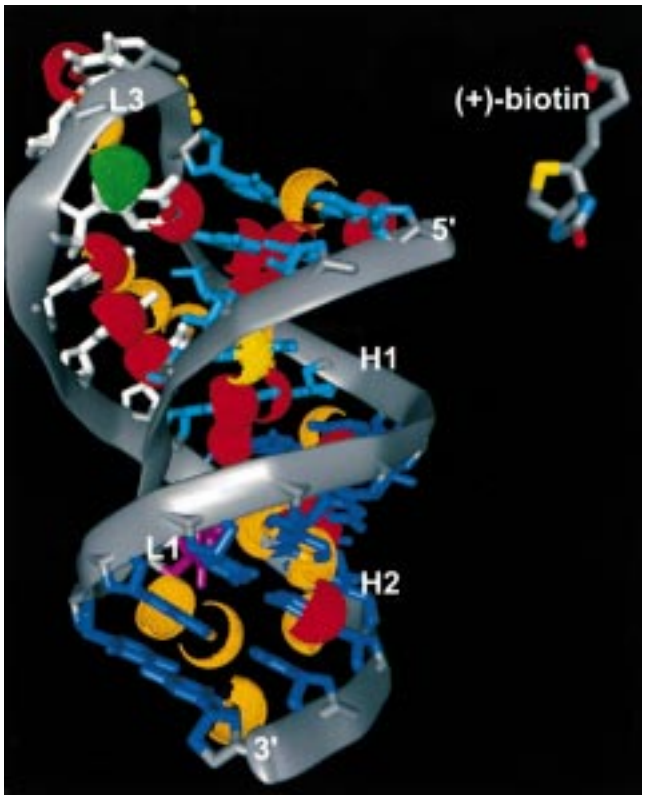

FIGURE 5: Structural model of BB8-5 generated by the MCSYM algorithm (*19*) and color coded as follows. Helix 1, cyan; helix 2, blue; loop 1, magenta; loop 3, white. Sites whose accessibility may be inferred by chemical modification are color coded according to the degree of protection as follows (results correspond to those obtained in selection buffer with biotin bound): highly protected, red; weakly protected, orange; not protected, yellow; hyper-exposed, green.

different final volumes to obtain a range of biotin concentrations. Trace amounts of 5′-end-labeled RNA were allowed to equilibrate with the matrix, nonspecifically bound species were washed off, and specifically bound RNA was competed off the matrix with elution buffer. The fraction of RNA specifically eluted at each biotin concentration yields a binding saturation curve from which the K_D was determined by least-squares analysis (Figure 6). Approximately 60% of the total aptamer tested in the binding reactions is able to fold correctly and bind immobilized ligand. The data from each dilution fit the composite curve well with $r^2 = 0.99$. The overall dissociation constant (using all ligand dilutions) is ~6 μM. While clearly not the strongest interaction between an RNA aptamer and ligand possible [e.g., a cyanocobalamin aptamer binds with $K_D \approx 90$ nM (*2*)], the association is remarkably strong given the relatively limited set of molecular interactions that seem available. By comparison, aptamers selected for ATP and L-citrulline bind their ligands with dissociation constants of 14 and 68 μM, respectively (*6*). The relatively strong binding suggests the existence of extensive complementarity between the ligand and the RNA.

In an effort to identify specific functional groups on biotin required for binding, the ability of various biotin analogues to elute RNA bound to a biotin column was measured (Figure 7). Radioactively labeled BB8-5 RNA was applied to a biotin agarose column, and unbound RNA was removed by washing with 10 column volumes of binding buffer. Biotin

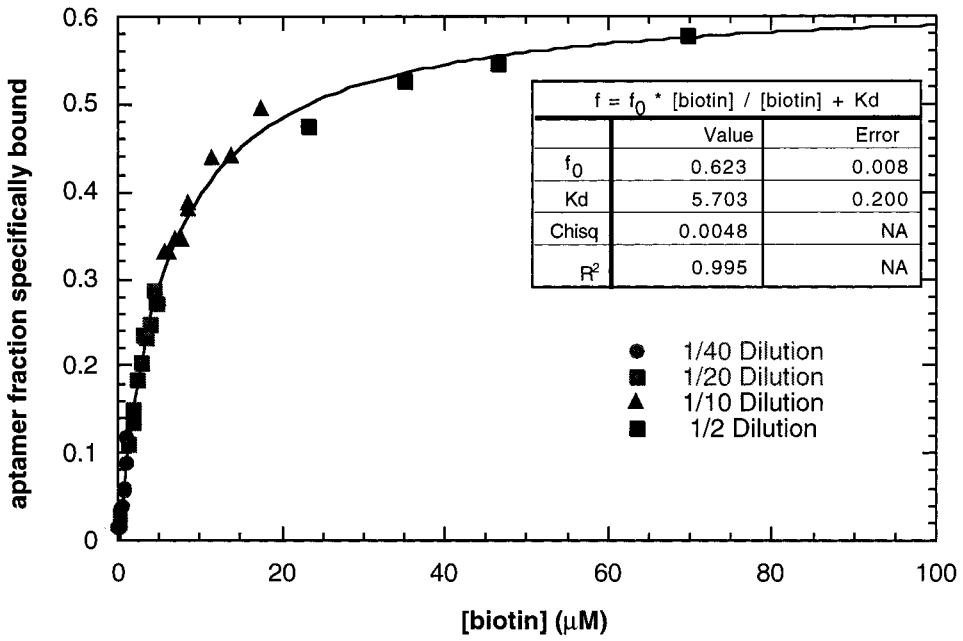

FIGURE 6: Saturation curve for the binding of the biotin aptamer to immobilized ligand as described in the Materials and Methods. (●) 1/40 dilution of matrix; (■) 1/20 dilution; (▲) 1/10 dilution; (■) 1/2 dilution. Nonlinear regression was carried out using Kaleidograph software, fitting to eq 1.

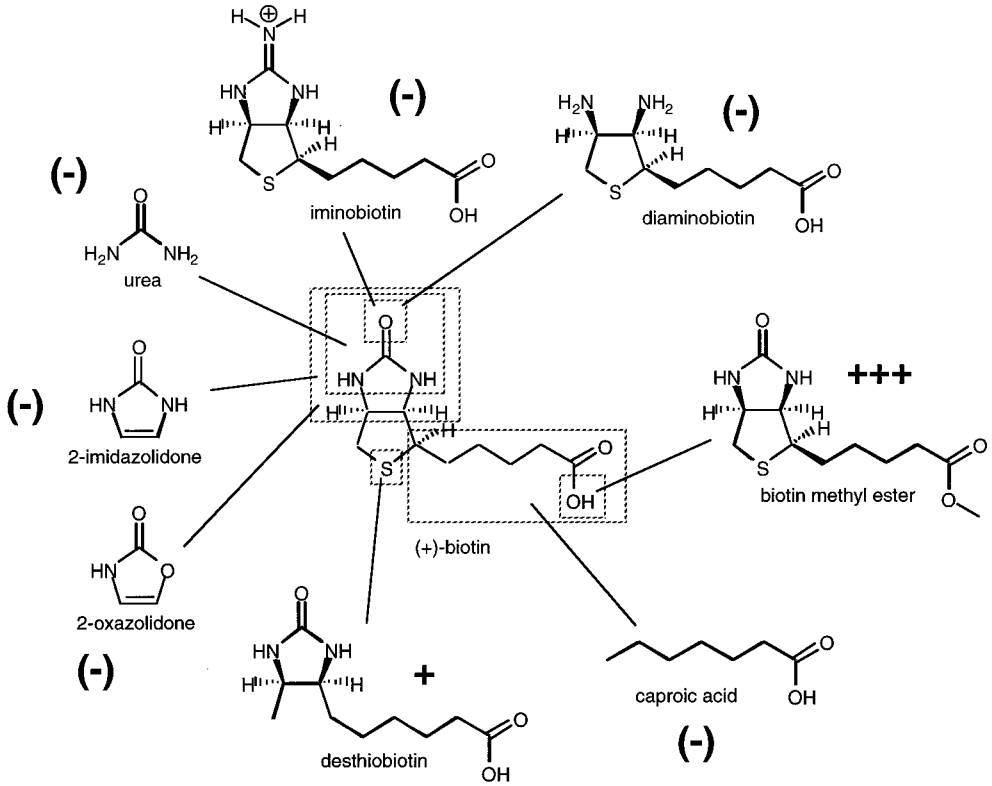

FIGURE 7: Ligand specificity. Biotin analogues were tested for the ability to compete for binding to BB8-5 RNA bound to biotin agarose. Analogues were prepared at 5 mM concentration in selection buffer and used to elute biotin agarose-immobilized RNA. Counts eluted after five column washes with analog-containing buffer were measured by scintillation counting. (+++) Highly efficient; (++) Efficient elution; (−) no effect. Normalized percentage of RNA remaining on the column after elution: biotin, 19.2%; biotin methyl ester, 16.3%; desthiobiotin, 26.8%; diaminobiotin, 89.9%; iminobiotin, 88.2%; imidazolidone, 91.9%; 2-oxazolidone, 90.0%; urea, 88.6%; caproic acid, 90.7%; none (selection buffer), 89.6%.

methyl ester was somewhat more effective than biotin as an affinity eluant, suggesting that the binder interacts in some way with the fatty acid tail and disfavors the negatively charged carboxylate form. Desthiobiotin, a less conformationally constrained analogue lacking the sulfur in the thiophane ring, eluted BB8-5 from the column but was less effective than biotin. Smaller molecules representing fragments of the cofactor were completely ineffective in eluting the bound aptamer. Biotin can be conceptually considered as the overlapping union of 2-oxazolidone, thiophene, and caproic acid. Individually, none of these molecules efficiently compete with biotin for binding (Figure 7). From

these results, we conclude that simultaneous interactions with several functionalities dispersed throughout the biotin molecule are required for binding.

DISCUSSION

The current results have shown that a simple pseudoknot defined by the general consensus shown in Figure 3 specifies high-affinity biotin binding. In many respects, this aptamer differs from those previously characterized in that it relies largely upon nucleotides involved in Watson−Crick base pairing for function. Recognition of biotin, a potentially difficult task given that it lacks functional group characteristics common to other aptamer ligands, appears to rely on numerous dispersed interactions, suggesting that the binding site may be large and complex. A three-dimensional model may be generated which is consistent with a number of both phylogenetic and biochemical constraints.

Given this model (Figure 5), two alternative modes of binding seem feasible. Biotin may lie in the minor groove of helix 1, simultaneously pulling the A-rich loop and the 5′-side of helix 1 together to make specific interactions with the ligand. The chemical modification data would be explained if biotin physically blocks access of DEPC to the A-rich loop and if, in closing in on the ligand, the other side of helix 1 becomes more accessible to reaction. This model, however, fails to account for the conservation of nucleotides in helix 2 since binding is well removed from this helix. In an alternative model, biotin may bind in the major groove between helices 1 and 2, forcing a conformational change in G-22. Conservation of base pairs in both helices would be explained if biotin interacted directly with each. Increased protection of A-loop N7 atoms would be accounted for if binding stabilizes the pseudoknot structure as a whole and prevents occasional unfolding by the connecting loop. However, the fact that base pair 1.5 lies at the junction between the two helices near the proposed binding site yet it is poorly conserved would tend to argue against this model. Given the shortcomings of each proposed mode of binding, we cannot definitively argue for one or the other. We have recently succeeded in obtaining diffracting crystals of the minor clone 1 aptamer complexed with biotin (*21*). Ultimately, by solving its structure, we should unambiguously determine how biotin is recognized by the pseudoknot defined by these experiments.

REFERENCES

1. Ellington, A. D., and Szostak, J. W. (1990) *Nature 346*, 818−822.
2. Lorsch, J. R., and Szostak, J. W. (1994) *Biochemistry 33*, 973−982.
3. Bock, L. C., Griffin, L. C., Latham, J. A., Vermaas, E. H., and Toole, J. J. (1992) *Nature 355*, 564−566.
4. Ciesiolka, J., Gorski, J., and Yarus, M. (1995) *RNA 1*, 538−550.
5. Geiger, A., Burgstaller, P., von der Eltz, H., Roeder, A., and Famulok, M. (1996) *Nucleic Acids Res 24*, 1029−1036.
6. Sassanfar, M., and Szostak, J. W. (1993) *Nature 364*, 550−553.
7. Tuerk, C., and Gold, L. (1990) *Science 249*, 505−510.
8. Yang, Y., Kochoyan, M., Burgstaller, P., Westhof, E., and Famulok, M. (1996) *Science 272*, 1343−1347.
9. Fan, P., Suri, A. K., Fiala, R., Live, D., and Patel, D. J. (1996) *J. Mol. Biol. 258*, 480−500.
10. Jiang, F., Kumar, R. A., Jones, R. A., and Patel, D. J. (1996) *Nature 382*, 183−186.
11. Jiang, F., Fiala, R., Live, D., Kumar, R. A., and Patel, D. J. (1996) *Biochemistry 35*, 13250−13266.
12. Jiang, L., Suri, A. K., Fiala, R., and Patel, D. J. (1997) *Chem. Biol. 4*, 35−50.
13. Zimmermann, G., Jenison, R., Wick, C., and Simorre, J. (1997) *Nature Struct. Biol. 4*, 644−649.
14. Dieckmann, T., Suzuki, E., Nakamura, G. K., and Feigon, J. (1996) *RNA 2*, 628−640.
15. Schmitz, U., Freymann, D. M., James, T. L., Keenan, R. J., Vinayak, R., and Walter, P. (1996) *RNA 2*, 1213−1227.
16. Wilson, C., and Szostak, J. W. (1995) *Nature 374*, 777−782.
17. Chen, X., Chamorro, M., Lee, S. I., Shen, L. X., Hines, J. V., Tinoco, I., Jr., and Varmus, H. E. (1995) *EMBO J. 14*, 842−852.
18. Burke, D. H., and Gold, L. (1997) *Nucleic Acids Res. 25*, 2020−2024.
19. Major, F., Turcotte, M., Gautheret, D., Lapalme, G., Fillion, E., and Cedergren, R. (1991) *Science 253*, 1255−1260.
20. Du, Z., and Hoffman, D. W. (1997) *Nucleic Acids Res. 25*, 1130−1135.
21. Nix, J. C., Newhoff, A. R., and Wilson, C. (1998) *Acta Crystallogr., Sect. D* (in press).

BI981371J