Molecular Diversity, **4:** 75–89, 1999.
KLUWER/ESCOM

The limits of specificity: An experimental analysis with RNA aptamers to MS2 coat protein variants

Ichiro Hirao[a,*], Marc Spingola[b], David Peabody[b] & Andrew D. Ellington[a,†]
[a] *Department of Chemistry, ICMB A4800, University of Texas at Austin, Austin, TX 78712, U.S.A.*
[b] *Department of Molecular Genetics and Microbiology, University of New Mexico School of Medicine, Albuquerque, NM 87131, U.S.A.*

Received 27 May 1998; Accepted 13 January 1999

Key words: aptamer, coat protein, in vitro selection, molecular recognition, MS2, SELEX

Summary

It has been hypothesized that selections for aptamers with high affinity for a given target molecule will of necessity identify aptamers that have high specificity for that target. We have attempted to assess this hypothesis by selecting aptamers that can bind to MS2 coat protein or to single- or double-substitution variants of the coat protein. Some aptamers selected to bind MS2 coat protein or its variants were mildly specific for their cognate targets, discriminating by two- to fourfold against closely related proteins. Specificity determinants on both the coat proteins and the aptamers could be identified. However, many aptamers could readily bind to each of the different coat proteins. The identification of such aptamer 'generalists' belies the proposed relationship between the affinities and specificities of selected RNA ligands. These results imply that, while aptamers may not finely discriminate between closely related targets, neither will their binding be negated by mutations in targets. Aptamer pharmaceuticals may therefore better resist the evolution of resistance.

Introduction

Biopolymer ligands that bind tightly and specifically to target molecules can be selected from random sequence populations. For example, peptide ligands have been isolated from peptide libraries generated synthetically or displayed on the surface of phage, while nucleic acid ligands (aptamers) have been isolated from large, random sequence nucleic acid populations. In vitro selection experiments have been used to map relationships between the sequence, structure, and function of natural biopolymers, and to identify novel biopolymer ligands that can serve as research reagents or drug leads.

Selection experiments may also serve to elucidate general principles that govern the evolution of molecular recognition. For example, Eaton and co-workers [1] have hypothesized that the natural or artificial selection of ligands that have high affinities for their targets will in general beget ligands with high specificities for their targets. In their words,

> *We have come to the conclusion that a sufficiently high-affinity ligand can be confidently expected to be highly specific for its target .... Selecting for even tighter binding [should], we believe, eventually give selective binding even when the competitor is closely related.*

This hypothesis is based on the model that as ligands with progressively higher affinities for a target evolve, they will form more or tighter bonds or steric interactions with a target and will therefore meld more precisely to the surface of the target. A schematized version of this hypothesis is shown in Figure 1a. In support of this hypothesis, Eaton et al. [1] have argued that RNA molecules, selected to bind basic fibroblast

* Present address: Yokoyama CytoLogic Project, ERATO, Japan Science and Technology Corporation (JST), c/o The Institute of Physical and Chemical Research (RIKEN), Hirosawa, Wako-shi, Saitama 351-01, Japan.
† To whom correspondence should be addressed.

growth factor, had much greater affinity for their cognate target than for other, related fibroblast growth factors. Similarly, DNA molecules selected to bind thrombin did not bind other serine proteases [2], while RNA molecules selected to bind one protein kinase C isozyme bound with reduced affinity to highly related isozymes [3].

While the hypothesis presented by Eaton et al. [1] is plausible and empirically supported, it is unclear whether it will of necessity be true in all instances. Complementarity between a selected ligand and surface features on a cognate target could potentially extend to similar surface features on related but non-cognate targets (compare Figures 1a and 1b). Since natural and unnatural evolutionary processes are blind, there is no reason to assume that selected ligands will of necessity be drawn to similar surface features on similar targets while eschewing dissimilar surface features. In addition, the monomer sets used in the construction of biopolymer ligands may not be sufficiently 'granular' to distinguish between surface features or epitopes on closely related target molecules (compare Figures 1a and 1c). A simplistic but practical example of such a limitation would be the inability of a single-stranded nucleic acid to discriminate between its complement and the same complement containing 2,6-diaminopurine in place of adenine: 2,6-diaminopurine can pair with uracil as well as or better than adenine can. Similar but equally inherent limitations on binding specificity may exist for nucleic acid:protein pairs.

In order to better test the generalization that selection for binding affinity begets binding specificity, we have carried out multiple selection experiments that targeted a series of related proteins. The coat protein of bacteriophage MS2 regulates production of the viral replicase by binding to a short hairpin structure in the MS2 genome [4,5]. The MS2 coat protein binds specifically to its cognate RNA operator, and does not productively interact with RNA operators of related bacteriophages, such as bacteriophage Qβ [5,6]. The crystal structure of the complex between the MS2 coat protein dimer and its cognate RNA hairpin has been solved [7]. Coat protein variants that contained amino acid substitutions at positions known to contact the MS2 operator were found to have altered binding specificities, and could interact with the Qβ operator [8–11]. We selected RNA molecules that bound to these single- and double-substitution variants of the MS2 coat protein, and assayed the aptamers for their ability to discriminate between closely related protein targets. Despite the fact that optimal binding species were identified for each target, many of the aptamers could readily cross-recognize related, non-cognate targets.

Materials and methods

Coat proteins

Wild-type MS2 and Qβ bacteriophage coat proteins and variants of the MS2 coat protein were prepared as previously described [11,12]. In short, plasmid pCT1 contains an expression cassette consisting of the lac promoter and a coat protein gene. Plasmids were transformed into *E. coli* strain CSH41 (lac$^-$, pro$^-$, galE, thi$^-$ [13]) and single colonies were grown in LB medium (500 ml) to saturation. Cells were pelleted by centrifugation and resuspended in 50 ml lysis buffer (100 mM NaCl, 50 mM Tris·Cl, pH 8.5, 10 mM EDTA, 2 mg/ml lysozyme). After 60 min on ice, sodium deoxycholate was added to a final concentration of 0.05%. The mixture was kept on ice for another 60 min and then sonicated to reduce viscosity. Polyethyleneimine was then added to a final concentration of 0.2% and the lysate was incubated on ice for another 60 min. Following centrifugation to remove cellular debris and precipitants, ammonium sulfate was added to the supernatant to 50% of saturation. The ammonium sulfate precipitant was collected by centrifugation, dissolved in 100 mM NaCl, 10 mM Tris·Cl, pH 7.5, 0.1 mM $MgSO_4$, 0.01 mM EDTA, and applied to a Sepharose CL4B column as previously described by Peabody [12]. Fractions containing coat proteins were identified by SDS-PAGE, pooled, and concentrated in Centricon (Amicon, Beverly, MA) centrifugal concentrators. Bacteriophage capsids were disaggregated in 50% acetic acid and dialyzed against 10 mM acetic acid. This procedure yields coat protein at purities estimated by gel electrophoresis to exceed 95%. Purified coat proteins were stored at 4 °C in 1 mM acetic acid and 10 mM DTT [14,15] and were generally used within two weeks to ensure that activity was not lost.

Random sequence pools

The N30 pool used for coat protein selection experiments has previously been described [16]. The chemically synthesized, single-stranded N30 DNA pool (approximately 1 μg $\approx 2 \times 10^{13}$ sequences) was amplified in a polymerase chain reaction (PCR) with the primers 24.30 and 41.30 [17]. The double-stranded

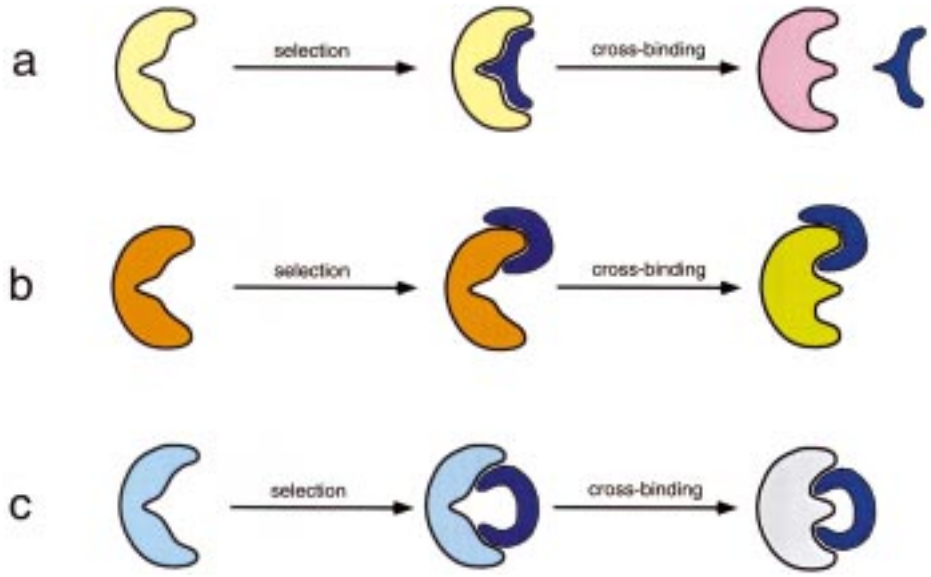


Figure 1. Interactions between selected biopolymers and related targets. (a) Aptamers (or other biopolymers) selected to bind to a given target may 'fit' that target so tightly that any perturbation of the interface in a related target will disrupt binding. This is essentially the model of Eaton et al. [1]. (b) Aptamers selected to bind a given target may bind to related targets because at least some surface features are similar between the targets. (c) Aptamers selected to bind a given target may 'fit' that target relatively loosely, forming multiple weak or non-directed interactions, and thus may be relatively impervious to the presence of structural differences on the surfaces of related targets.

DNA pool was purified by ethanol precipitation in the presence of 1 M ammonium acetate, pH 7.4. An RNA pool was transcribed from the amplified DNA pool using an Ampliscribe kit (Epicentre Technologies, Madison, WI). The kit was used according to the manufacturer's directions, except that 20 nmol of α-^{32}P UTP (3000 Ci/mmol) was included in the reaction. Transcripts were purified on 10% denaturing polyacrylamide gels.

In vitro selection

The purified RNA pool was further prepared for selection experiments. The pool was ethanol precipitated, collected by centrifugation, and re-dissolved in 1× binding buffer (100 mM Tris·Cl, pH 8.3, 80 mM KCl, 10 mM magnesium acetate). For selections involving Qβ coat protein the binding buffer was adjusted to pH 6.0. To promote the thermal equilibration of RNA conformers the pool was denatured at 75 °C for 3 min prior to each round of selection and allowed to reanneal at 4 °C for 3 min. The thermally equilibrated pool (200 μl) was passed through a modified cellulose filter (HAWP filters, Millipore, Bedford, MA) that had been pre-wetted with binding buffer; this pre-filtration step was repeated from one to three times. While 5 μg of RNA was used in the 1st through 3rd, and 10th through 12th rounds of selection, 0.5 μg of RNA was used in other rounds. There were approximately eight copies of each sequence in the first round of selection, and no binding species should have been lost in subsequent rounds due to population bottlenecks.

The prepared pool was mixed with protein targets in binding reactions. The RNA pool was mixed with 20 pmol of coat proteins in the 1st to 6th, and 10th to 12th rounds of selection, and 5 pmol coat proteins in the 7th to 9th rounds of selection. In summary, the RNA:protein ratio varied from 10:1 in the 1st through 3rd rounds, to 1:1 in the 4th through 6th rounds to 40:1 in the 7th through 9th rounds. The binding reactions were then equilibrated at 4 °C for 60 min. In each round of selection binding species were separated from unbound RNAs by passing the binding reaction through a modified cellulose filter under pressure (≈5 in Hg). Weakly or non-specifically bound species were removed by washing the filter three times with 200 μl of binding buffer. Bound RNA molecules were eluted by incubating the filter twice with 200 μl of 7 M urea, 0.1 M sodium citrate, pH 5.0, and 3 mM EDTA at 100 °C for 5 min. The eluted RNAs (400 μl total volume) were precipitated with isopropyl alcohol. Selected RNA species were amplified by reverse transcription, PCR, and in vitro transcription as previously described [18]. In the 9th through 12th rounds of selection RNA populations were passed through a modified

cellulose filter prior to amplification. This procedure was found to be extremely effective at eliminating a small subpopulation of nucleic acids that bound exclusively to the filter in the absence of protein.

Binding assays

A filter binding assay was used to monitor the efficiencies of RNA:coat protein interactions during the course of the selection. RNA samples (20 pmol) were heat denatured and annealed at 4 °C in binding buffer (100 μl). Coat proteins (20 pmol) were added and the binding reaction was incubated at 4 °C for 60 min. The binding reaction was filtered through a modified cellulose filter and washed three times with 200 μl of binding buffer (600 μl total). Retained counts were quantitated using a Phosphorimager (Molecular Dynamics, Sunnyvale, CA) and compared to applied counts.

A competition assay was used to measure the relative affinities of aptamers for coat protein variants. Competitor RNAs were either a labeled, wild type MS2 hairpin containing a U to C substitution in the loop that was known to improve binding affinity ([19]; 5′-GGGGCAAACAUGAGGAUCACCCAUGU), a wild type Qβ hairpin (5′-GGGAAAUGCAUGUCUAAGACAGCAU), or an aptamer from Family 15 containing a nonamer adenylate tail (5′-GGGAAUGGAUCCACAUCUACGAAUUCUCAAGCUGGCAGUCGCGAGCAUCAGCCGCAUUCACUGCAGACUUAAAAAAAAAA). Competitor RNAs (400 nM final concentration) were thermally equilibrated and mixed with labeled aptamers (400 nM) and limiting amounts of coat proteins (200 nM) in binding buffer (100 μl). Binding reactions were incubated at 4 °C for 60 min, passed through a modified cellulose filter, and washed three times with 200 μl of binding buffer. Bound RNAs were eluted with 100 μl of loading dye (7 M urea, 0.1 M sodium citrate, pH 5.0, 3 mM EDTA, 0.05% bromophenol blue) at 100 °C for 10 min. Two-fifths (40 μl) of the eluate was used for gel electrophoresis. Competitor RNAs and aptamers were separated on either 10 or 15% denaturing polyacrylamide gels. Relative band intensities were determined using a Phosphorimager. The relative binding ratio was calculated based on the following formula: [(counts filtered, aptamer)/(counts unfiltered, aptamer)]/[(counts filtered, competitor)/(counts unfiltered, competitor)]. The binding ratio should represent the equilibrium dissociation constant of an aptamer relative to the equilibrium dissociation constant of a competitor RNA for a given protein target.

Sequence analysis

Aptamers from the 12th round of selection were converted into double-stranded DNA and cloned into a TA cloning vector (Invitrogen, San Diego, CA). Sequences were derived from individual plasmid DNAs using standard dideoxy sequencing protocols. Multiple sequences were aligned and compared using the MEGALIGN package (DNA*, Madison, WI), which is based in part on the CLUSTAL algorithm of Higgins and Sharp [20]. Aptamer secondary structures were predicted using the program MULFOLD [21].

Results and discussion

In vitro selection of RNA aptamers that bind MS2 and Qβ coat proteins

An initial series of selections with the wild type coat proteins from bacteriophages MS2 and Qβ were carried out to establish a baseline for correlating the affinities and specificities of RNA ligands. The MS2 and Qβ coat proteins likely descend from a common ancestor (Figure 2), but do not cross-recognize one another's RNA ligands. Previous in vitro selection experiments that targeted the coat protein from another, related bacteriophage, R17, returned aptamers whose sequences and structures were similar to those of the wild type RNA ligand [14]. It was therefore expected that aptamers selected to bind the MS2 or Qβ coat proteins would be similar in sequence to their respective wild type RNA ligands, and thus would be able to easily discriminate between the coat proteins. In essence, these initial experiments were a positive control for the question we eventually wished to answer: does selection for binding affinity lead to binding specificity?

A random sequence RNA pool that spanned 30 positions (N30) was mixed with the coat protein targets and binding species were iteratively selected by filtration. The binding buffers that were used for selection had previously been used to assay interactions between the MS2 or Qβ coat proteins and their cognate RNAs [15]. The numbers of different RNA species ($\approx 2 \times 10^{13}$) that were present in the initial binding reactions were sufficient to span all sequence motifs of length 22 or less, a length that was similar to the known length of the wild type RNA operators. Therefore, it was likely that aptamers returned from the selection experiments would contain optimal sequence and structural motifs of similar complexity to the wild type ligands.

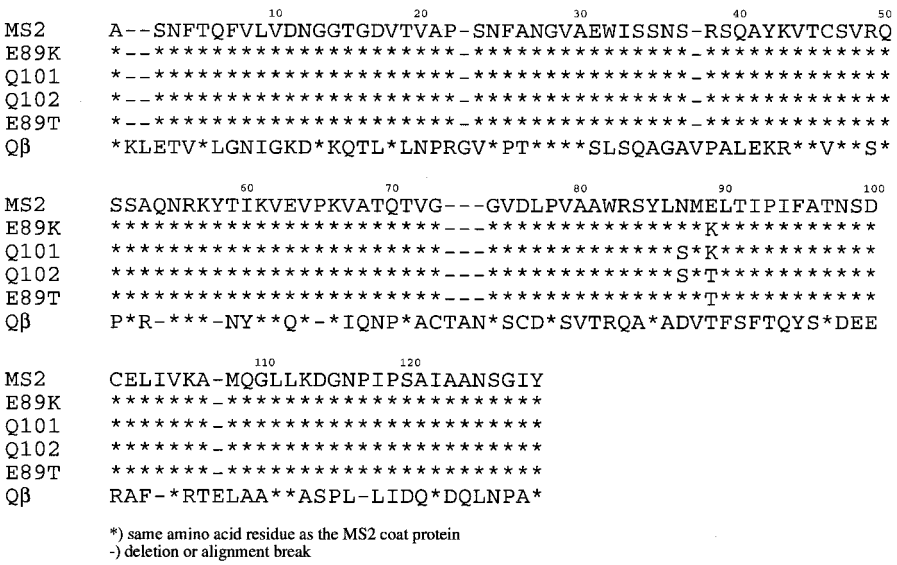


Figure 2. Amino acid sequences of protein targets. The sequences of the various coat proteins used in selection experiments. The names of the coat proteins and variants follow standard conventions, except for: Q101 = N87S,E89K and Q102 = N87S, E89T.

Although protein-dependent RNA binding species predominated in the population after six rounds of selection and amplification, the selection was continued an additional six rounds to promote competition between binding species and to ensure the selection of the highest affinity species. At the conclusion of the selection, 53% of aptamers selected to bind to MS2 coat protein could be captured in a standard filtration assay, and 53% of aptamers selected to bind Qβ coat protein could be captured. While some filter-binding species were present in earlier rounds of selection, these species had been eliminated by the 12th round. All of the individual aptamers characterized in these studies showed little or no background binding to modified cellulose filters (<1% in a standard filtration assay).

At the conclusion of the selection experiments aptamers from each pool were cloned and their sequences determined (28 aptamers for the MS2 coat protein, 20 for the Qβ coat protein; Table 1). The sequences were compared with one another using the program MEGALIGN (DNA*) and most of the aptamers could be grouped into families based on sequence similarities. In some cases variants within a family were clearly the result of mutagenesis during amplification (e.g., Family 1), while in other cases it was likely that variants were independently derived (e.g., Family 8). The repetition of sequences or motifs within families indicated that the selection had likely winnowed the population to those few species that had the highest affinities for their targets. The program MULFOLD was used to predict the secondary structures of individual aptamers [21]. In general, all of the sequences within a family were predicted to form similar secondary structures.

As expected, several of the families (Families 1–7) showed a distinct resemblance to wild type ligands. For example, aptamers selected to bind the MS2 coat protein were predicted to fold into hairpins that contained an AUCA tetraloop and a bulged A residue (Figure 3). These structural features have previously been shown to be critical for recognition of the wild type MS2 RNA [7,22,23]. The AUCA tetraloop differed slightly from the wild type AUUA, but the U to C substitution had previously been shown to improve interactions with the MS2 coat protein [19]. Two aptamer families were similar to the wild type MS2 operator but also contained unique sequence or structural features. Aptamers that were members of Family 6 were predicted to form a trinucleotide rather than a tetranucleotide loop, and contained a bulged adenosine that was displaced one base down the stem. Aptamers that were members of Family 7 contained no bulged adenosine and sometimes had an AACA tetraloop.

Aptamers selected to bind the Qβ coat protein were predicted to form hairpins topped by an AAA loop, similar to the UAA loop found in the wild type (Figure 3). The 3′-most adenosine in the loop had previously been shown to be critical for recognition of Qβ

Table 1. Selected sequences.

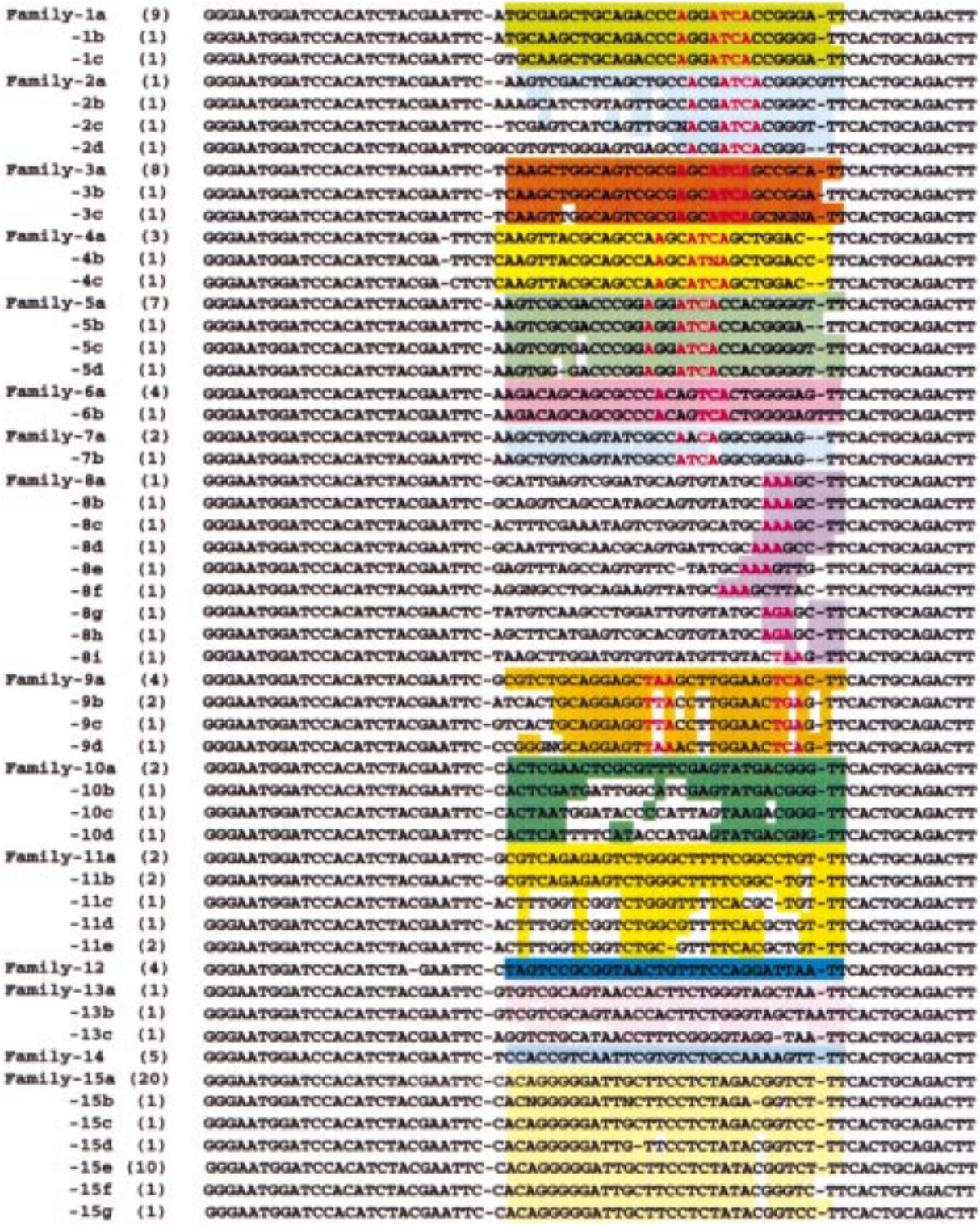
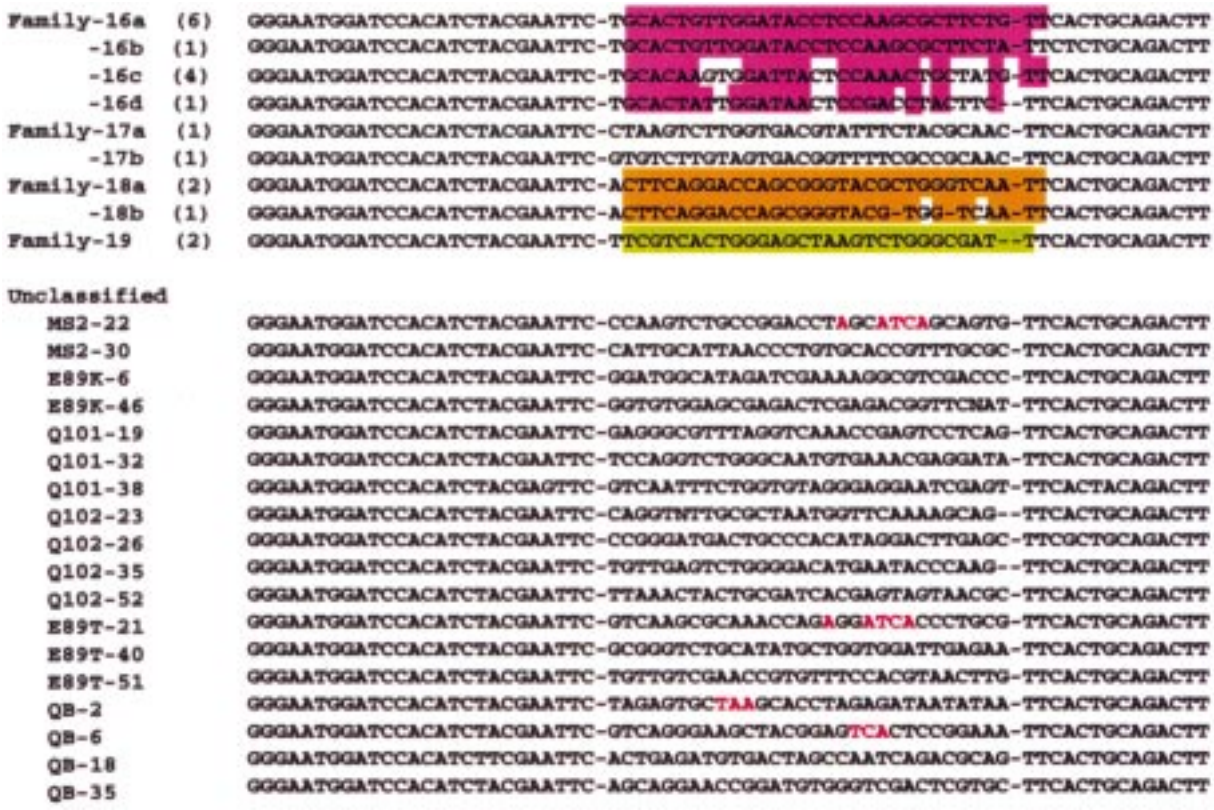


Table 1. Continued.



Families were grouped together based on sequence or structural similarities. Sequence comparisons were carried out using the program MEGALIGN (DNA*), and secondary structures were predicted using the program MULFOLD [21]. The number of times a particular sequence variant was isolated is shown in parentheses. Sequence similarities are highlighted by colored blocks. In the case of Family 8, sequence diversity is extensive and only the terminal portion of a predicted stem-loop structure is highlighted. Predicted loop or bulge residues that are similar to those seen in the wild-type MS2 operator (Families 1–7, 9, and some unclassified sequences) or in the wild-type Qβ operator (Families 8–9 and some unclassified sequences) are shown in red. 'N' denotes a residue whose identity is uncertain. The first sequence shown in a family was the aptamer used for binding experiments.

RNA [5,15]. The predicted secondary structure was supported by sequence covariations. For example, aptamer 8(i) contained a predicted A:U pairing in place of a predicted G:C pairing. Interestingly, many of the Qβ coat protein ligands also contained a new structural feature, a U-U mismatch two base pairs down from the loop. This mismatch appeared to have been independently fixed in a number of aptamers (see the diverse set of Family 8 sequences), and was therefore likely important for binding function. Although the constant region played a role in establishing the U-U mismatch in many of the Family 8 members, at least one aptamer, 8(f), was predicted to form this structure independent of the constant region.

Aptamers chosen from the different families were assayed for their ability to bind the wild type coat proteins. Individual aptamers were transcribed and mixed with either wild type MS2 or Qβ operators in the presence of limiting amounts of either MS2 or Qβ coat proteins. RNA:protein complexes were isolated by filtration, bound aptamers and wild type RNAs were separated from one another by denaturing polyacrylamide gel electrophoresis, and the amounts of bound RNAs were quantitated. The fraction of an RNA ligand (selected or natural) that was bound by a coat protein was determined by comparison to unbound RNA samples. The ratio of the fraction of bound aptamer to the fraction of bound wild type RNA was the 'relative binding ratio' (see also section Materials and

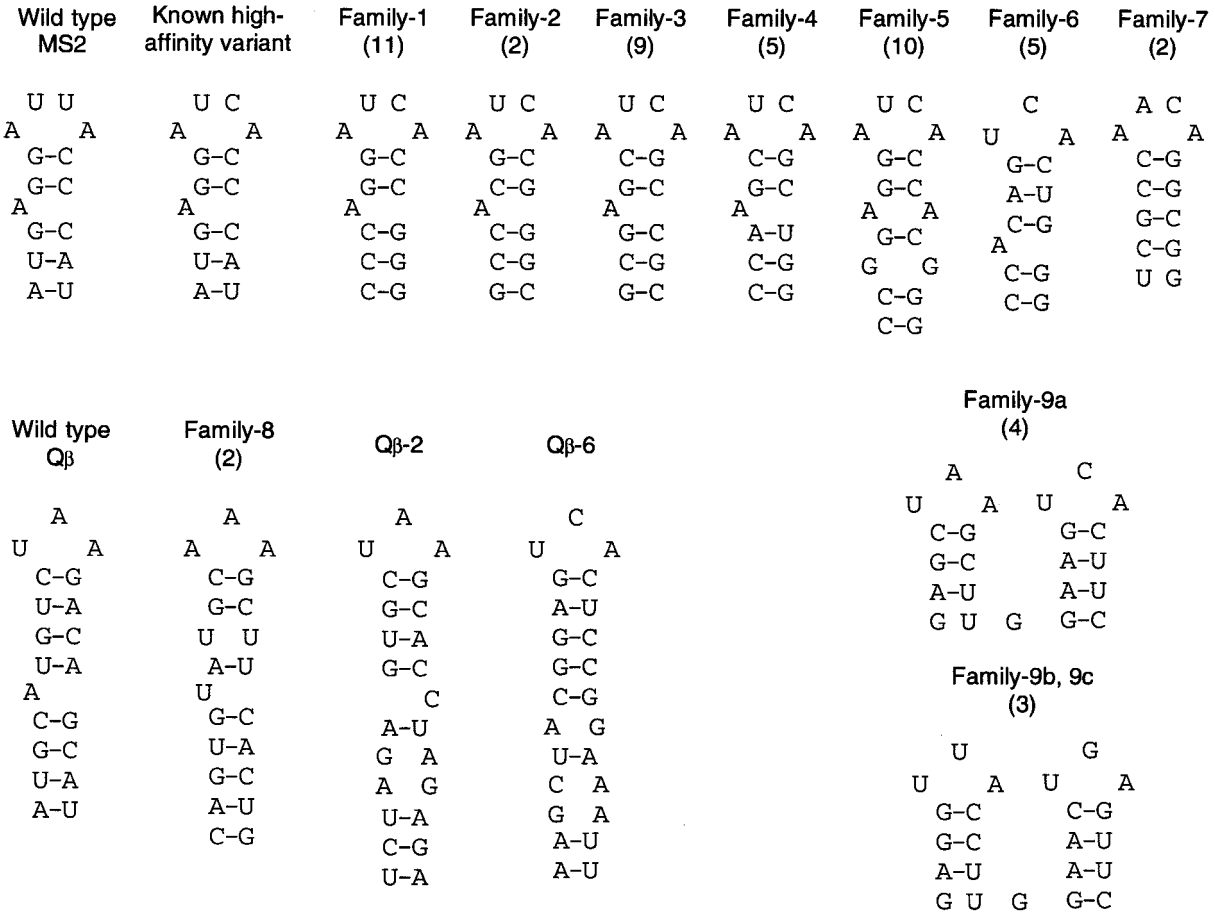
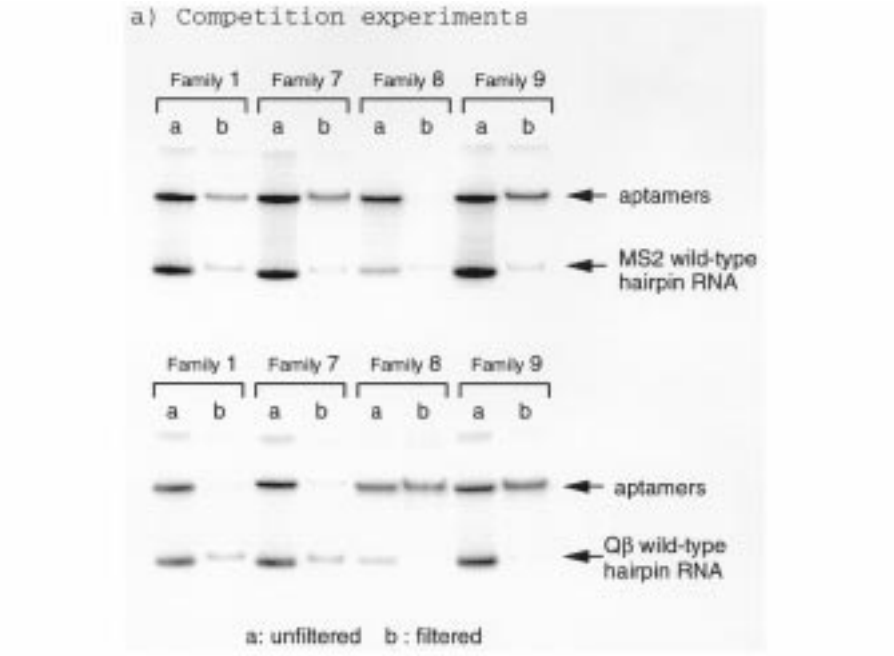


Figure 3. Operator-like aptamers. The wild type MS2 operator and a high-affinity variant that had previously been identified [17] are shown at the top left; the Qβ operator is shown at the bottom left. Aptamers selected to bind MS2 or Qβ coat proteins could be folded into structures that resembled their respective wild type operators and are shown alongside the wild type operators. The names of the aptamer families are shown above the predicted structures; the numbers in parentheses indicate the number of times that sequence was found. In general, the number of aptamers is equivalent to the size of a family, except for Family 8, which contains several related but divergent aptamers. For purposes of comparison, the same numbers of base pairs are shown in the stems of operators and operator-like aptamers (five base pairs for MS2-like operators, eight base pairs for Qβ-like operators). The predicted aptamer structures may contain additional base pairs. The dual-specificity aptamers (Family 9) are shown as side-by-side hairpins, but these hairpins could of course be oriented in a variety of ways relative to one another.

methods) and should have reflected the relative dissociation constants of the RNA ligands for a coat protein target. The relative binding ratio was independent of the specific activities of the RNA and protein samples and the error observed between triplicate assays was $\pm 5\%$. While this competitive binding assay did not yield individual rate or equilibrium constants for RNA:coat protein interactions, it most closely resembled the competition for binding that occurred during the selection itself and thus was most germane to determining whether selection for binding affinity led to binding specificity.

Consistent with the hypothesis that selection for binding affinity begets binding specificity, the aptamers in general bound from two- to fivefold better than the wild type operators to their cognate coat proteins, but did not bind well to non-cognate coat proteins (Figure 4). One caveat to these results is that since the pH optima for the MS2 and Qβ coat proteins were different, the binding reactions for these coat proteins were carried out at different pHs. Nonetheless, alteration of the pH in the competitive binding assay did not appear to significantly alter the interactions of selected ligands with their cognate proteins. The amount of selected RNA captured in a standard



b) Relative binding ratios

	Family 1	Family 3-A_x	Family 6	Family 7	Family 8	Family 9
MS2 coat protein	2.4	5.0	3.4	5.0	0.20	7.9
Qβ coat protein	0.29	0.69	0.51	0.29	4.1	9.5

Figure 4. Affinities of operator-like aptamers. (a) Data from competition experiments. Radiolabeled aptamers were incubated with wild type coat proteins and wild type operators, complexes were captured by filtration, and the captured RNAs were eluted and separated by electrophoresis on a denaturing polyacrylamide gel. A portion of the binding reaction prior to filtration was loaded in parallel. The number of counts in each band was determined using a Phosphorimager. (b) Relative binding ratios, calculated as described in Materials and methods, for the experiment shown in (a) and for other experiments.

filtration assay by the MS2 coat protein was the same at pH 6.0 as it was at pH 8.3, while the amount of selected RNA captured by the Qβ coat protein was similar (53% at pH 6.0, 38% at pH 8.3). These results were consistent with those previously observed by Witherell and Uhlenbeck [15].

There was one exception to the empirical rule that selection for affinity begets specificity. Aptamers from one family (Family 9) could apparently bind to both the MS2 and Qβ coat proteins, and were in fact found in both selections. Aptamers from Family 9 were predicted to form two adjacent hairpin structures, one of which was similar to the native hairpin loop found in Qβ RNA, and one of which was similar to aptamers from Family 6 that contained a variant of the hairpin loop found in the MS2 operator. Since the MS2 operator binds to a coat protein dimer, it was possible that the bi-lobal aptamer was symmetrically positioned to bind both monomers simultaneously. Of the aptamers that were selected Family 9 had the highest affinity for both the MS2 and Qβ coat proteins (Figure 4b), consistent with the formation of additional contacts to coat protein dimers.

While it was possible that the dual-specificity aptamer was a result of serial cross-contamination between the selections that targeted MS2 and Qβ coat proteins, it appeared unlikely: first, since both selections returned multiple, different variants of the MS2 operator-like or Qβ operator-like aptamers, there did not appear to have been any unforeseen population bottlenecks, such as the amplification of small amounts of nucleic acids in aerosols. Second, and

more importantly, there were two, independently isolated aptamers in Family 9. While it was possible that low levels of one aptamer could have been passed back and forth between selections, it was unlikely that two independent aptamers would have passed through the same population bottlenecks in the same ways. The independent isolates contained sequence covariations (C:G to G:C, and G:C to C:G base pair changes) that firmly established the proposed bi-lobal structure (Figure 3).

In vitro selection of RNA aptamers that bind variants of the MS2 coat protein

Having first demonstrated that RNA ligands selected to bind distantly related protein targets were specific for those targets, we next chose to look at more closely related targets. Variants of the MS2 coat protein that had either lost the ability to repress the MS2 operator or that had gained the ability to repress the Qβ operator had previously been isolated using a genetic selection [8–11]. Two of the selected protein variants were single amino acid substitutions, E89K and E89T, while two of the protein variants were double substitutions that also included N87S (N87S,E89K = Q101, N87S, E89T = Q102). These amino acid substitutions were known to lie either within or adjacent to the RNA binding domain of the MS2 coat protein [7,23]. The protein variants did not lose their ability to bind to the MS2 RNA hairpin in vitro, but gained the ability to bind to the Qβ RNA hairpin. The close sequence relationships and large specificity differences between these coat protein variants rendered them ideal targets for determining whether selection experiments could in general return RNA ligands with both high affinities and specificities.

Aptamers that bound to the variant coat proteins were again selected over 12 rounds from the same N30 pool. The selection again appeared successful after six rounds, but was continued another six rounds to fix the best aptamers in the population. From 20 to 40 aptamers from each pool were cloned and sequenced. While some of the aptamers were similar to those derived from the selection that targeted the MS2 coat protein, the majority of aptamers were different and could be grouped into several new families based on sequence similarities and predicted secondary structures.

Although all of the aptamers listed in Table 1 were selected in the presence of a single target, some of the aptamer families (e.g., Family 15) contained 'generalists' that were selected by several different targets and other aptamer families (e.g., Family 11) contained 'specialists' that were selected by only one or two targets. The distinction between generalists and specialists can be most clearly seen by examining the number of members of each RNA family that were found in each selection experiment (Figure 5). For example, aptamers that fit the Family 15 consensus were extracted from the random sequence pool 35 times, 14 times each from selections that targeted E89K and Q101, 6 times from the selection that targeted Q102, and once from the selection that targeted E89T. Similarly, members of the MS2 operator-like Family 1 were isolated from one to four times in the selections that targeted the wild type MS2 coat protein and each of its substitution variants. In contrast, aptamers that fit the Family 11 consensus were isolated multiple times only in the selection that targeted E89T, and Family 16 was isolated only in the selection that targeted Q102.

The most likely explanation for the identification of identical or related sequences in different selection experiments was that family members were independently extracted from the same random sequence pool. For example, members of the MS2 operator-like Family 2 were isolated three times in the selection that targeted Q102 and once in the selection that targeted E89K. While the four individual aptamers resembled one another, it was also apparent that all four were independent of one another and could not have been derived by a process of cross-contamination and point mutation. Further evidence for the independence of the selection experiments was found by examining the 10 sequences that comprised the MS2 operator-like Family 3. Only 1 of the 10 sequences was found in the selection that targeted E89K, yet this sequence differed from those that were found in selections that targeted the wild type MS2 coat protein or the coat protein variants Q101 and Q102.

The relative binding ratios for a number of the aptamer families were determined for the entire range of protein targets (Figure 6). In order to ensure that competitive binding assays with multiple, different aptamer families and multiple, different protein targets could be readily compared with one another, a binding 'generalist' from Family 3 was used as the competitor. The 'generalist' contained a nine-residue 'tail' that facilitated electrophoretic separation but did not affect binding (see also Figure 4). As a control, several competitive binding assays were also carried out with the MS2 coat protein and the MS2 operator as the competitor. The relative values and rank order of the relative binding ratios for different aptamer fam-

Families	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
MS2 pool	4	-	5	3	6	2	3	-	1	-	-	-	-	-	-	-	-	-	2
E89K pool	1	1	1	-	1	-	-	-	-	4	-	-	-	-	14	-	-	-	-
Q101 pool	3	-	1	1	-	1	-	-	-	1	-	-	-	-	14	-	-	-	-
Q102 pool	2	3	2	-	2	2	-	-	-	-	-	2	-	-	6	12	-	3	-
E89T pool	1	-	1	1	1	-	-	-	-	-	8	2	3	5	1	-	2	-	-
Qβ pool	-	-	-	-	-	-	-	9	7	-	-	-	-	-	-	-	-	-	-

Figure 5. Number of family members isolated in individual selections. The same or similar aptamers were sometimes isolated in the six individual selections that were carried out against MS2 coat protein, E89K, Q101, Q102, E89T, and the Qβ coat protein. The number of aptamers from a given family that were isolated in a given selection is shown. These values accord with those shown in Table 1, where the numbers of times individual aptamer sequences were isolated are listed.

Families	1	6	7	8	9	10	11	12	13	14	15	16
MS2 coat protein	0.96	1.1	1.1	0.28	1.9	0.26	0.50	0.59	0.43	0.57	0.75	0.22
E89K	0.65	0.68	1.1	0.45	1.7	2.1	0.84	0.71	0.79	0.74	2.3	3.2
Q101	0.56	0.61	0.64	0.36	0.87	1.2	0.59	0.62	0.60	0.53	1.5	1.3
Q102	0.68	0.72	0.57	0.32	0.76	0.67	1.8	2.4	2.5	2.2	1.4	2.9
E89T	0.61	0.62	0.86	0.26	0.64	0.31	1.8	2.0	1.9	1.7	1.1	1.9
Qβ coat protein	0.29	0.51	0.29	4.1	9.5	0.54	0.54	0.34	0.22	0.29	0.24	0.12
	MS2 operator-like			Qβ operator-like	Dual specificity	E89K-specific	E89T-specific				generalist	avoids E89

Figure 6. Relative binding ratios for aptamers. The values shown were derived via competition assays similar to those previously described in Figure 4, except that more aptamer families were included in this analysis and a more stringent competitor (Family 3) was used. The values shown for the Qβ coat protein were derived in competition with the wild type Qβ operator. The classification of different aptamer families (shown below the binding ratios) is based on affinity profiles for the different targets, and mirrors the descriptions provided in the text.

ilies were similar, irrespective of whether the Family 3 generalist or the MS2 operator was used as a competitor. The correlation coefficient between the two independent data sets was 0.92.

Our results confirmed the finding that residue 89 in the MS2 coat protein was a specificity determinant for RNA ligands. Substitutions at position 89 have previously been shown to impart the ability to bind the Qβ operator [8,11], and we similarly found that substitutions at position 89 appeared to control which targets recognized which aptamers. The aptamer families could be conveniently grouped into four different specificity classes, depending on how they interacted with different substitutions at position 89 (Figure 6). First, aptamers from Families 11–14 were selected by coat proteins containing the E89T substitution, and discriminated by roughly fourfold against the wild type MS2 coat protein and against the E89K variants. Second, Family 10 aptamers were selected by coat proteins containing the E89K substitution, and discriminated by threefold or more against the wild type MS2 coat protein and against the E89T variants. Third, Families 7 and 9 bound best to proteins that contained either glutamate (wild type) or lysine at position 89, and discriminated slightly against proteins with other or additional substitutions. Both Families 7 and 9 had MS2 operator-like loop sequences but lacked a bulged adenosine residue. Finally, Family 16 bound very well to all of the coat protein variants but discriminated against the wild type MS2 coat protein. In contrast to the above results, aptamers that could preferentially recognize position 87 were not identified by selection: none of the families recognized Q101 (E89K, N87S) or Q102 (E89T, N87S) and discriminated against wild

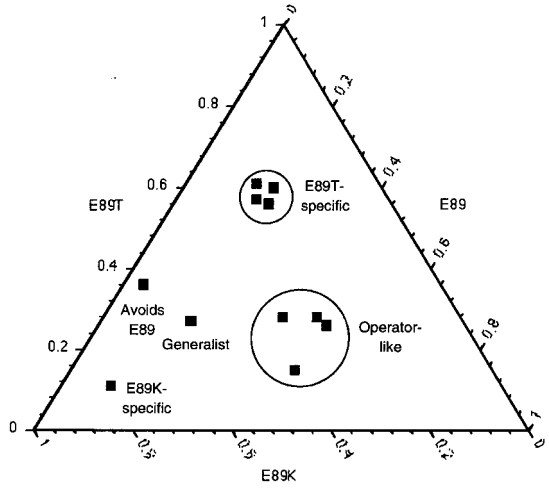


Figure 7. Graphical representation of specificity classes. Axes on this triangular graph represent the relative contribution of a given aptamer's binding ratio to the total affinity profile of the aptamer. The three values that place an aptamer on this graph were derived from Figure 6 as follows: A given relative binding ratio (for either wild type MS2, E89K, or E89T) was divided by the sum of these three relative binding ratios, and the result was normalized to 1. Amongst the operator-like aptamers, the outlier is the dual-specificity aptamer.

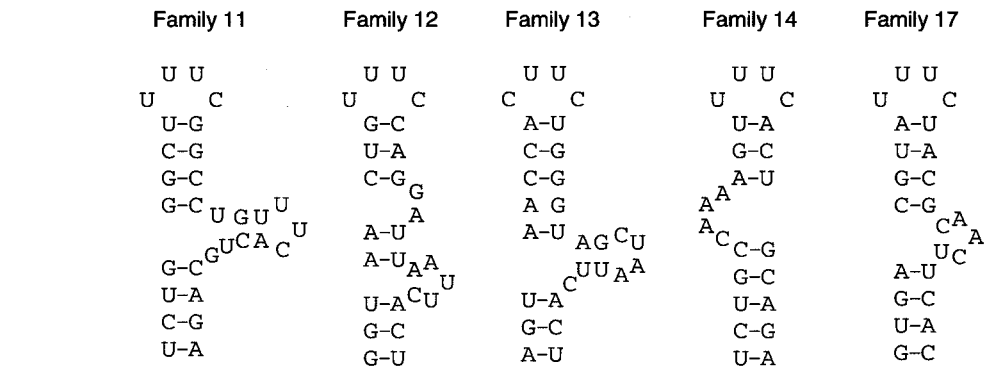
type or singly substituted MS2 coat proteins. In this regard, it is interesting to note that residue 87 directly contacts the MS2 operator, while residue 89 is adjacent to the RNA binding site [7].

A more graphical representation of the relative binding ratio data is shown in Figure 7. The axes on this triangular graph correspond to relative binding ratio data for individual protein variants. For example, the largest relative binding ratio for Family 10 was for the protein variant E89K, so Family 10 mapped along a 60° diagonal to the E89K axis near to the 1 value. Conversely, the smallest relative binding ratio for Family 10 was for the wild type MS2 coat protein, so Family 10 mapped along a 60° diagonal to the E89 axis that was near to the 0 value. Family 10 also had a small relative binding ratio for E89T, and this final value fixed the position of Family 10 on the triangular graph. By representing the data in this fashion it was immediately apparent which aptamers had similar specificities. Several families that contained MS2 operator-like sequences and structures (Families 1, 6, 7, and 9) clustered together, while several E89T-specific families (Families 11–14) also formed a tight cluster. In this representation the preference of the Family 15 'generalist' for non-wild type variants became more apparent. However, it should still be recalled that the Family 15 'generalist' bound the wild type MS2 coat protein approximately threefold better than did the wild type MS2 operator.

Our results suggested that the loop regions of RNA hairpins might be specificity determinants for bacteriophage coat proteins. As noted above, the hairpins selected to bind wild type MS2 and Qβ operators contained loop sequences already known to be critical for recognition [14,15]. Thus, it might naively be expected that RNA ligands selected to bind coat protein variants would also be predicted to form hairpin structures, and that the loop sequences of these hairpins would be found to covary with the protein sequence. A closer examination of the aptamer families provided a potential example of such a covariation. Families 11–14 had a common preference for E89T and were predicted to form stable hairpin structures. These hairpins also had in common the loop sequence YUUC (Figure 8). Following up on this insight, we examined the sequences and structures of other aptamer families that were not originally assayed for binding. Family 17 was also predicted to form a hairpin structure containing the loop sequence YUUC. Like Families 11–14, Family 17 was selected only by coat proteins containing the E89T substitution.

Surprisingly, the results of the binding experiments mimicked the selection results: Family 15 contained generalists that could bind to MS2 coat protein and all of its variants as well as or better than the Family 3 generalists; Family 15 was isolated in all selections that targeted the MS2 coat protein or its variants. Family 10 showed high affinity for variants that contained the E89K substitution; Family 10 was isolated only in selections that targeted variants that contained the E89K substitution. Family 8 showed poor affinity for the MS2 coat protein and all of its variants; Family 8 was not isolated in selections that targeted the MS2 coat protein or its variants, but was isolated only in selections that targeted the Qβ coat protein. In fact, only aptamer families originally isolated in selections that targeted the Qβ coat protein were found to compete effectively with the Qβ operator. None of the newly selected families showed better than wild type binding to the Qβ coat protein in our assay, despite the fact that these MS2 coat protein variants had previously been shown to display enhanced affinities for the Qβ operator.

While the correspondence between affinity and sequence data was striking, there were also several exceptions to this rule. For example, while only the coat protein variant E89T extracted Family 13 from a random sequence pool, Family 13 in fact bound



Preference for	Family 11	Family 12	Family 13	Family 14
Q102	3.5	4.1	5.7	3.9
E89T	3.6	3.4	4.4	3.1

Figure 8. E89T-specific aptamers. The predicted structures of aptamer families that are specific for E89T (as described in Figures 6 and 7) are shown. All of the aptamers were predicted to form hairpins topped by a YUUC tetraloop. No additional sequence or structural similarities could be discerned between these disparate aptamers. It is unknown whether the side bulges or adjacent stems that are shown also play a role in E89T recognition. Family 17 is included because it resembles Families 11–14 and was also isolated in selections that targeted E89T. The preferences for either the double substitution variant Q102 (N87S,E89T) or the single substitution variant E89T are shown below the structures. The preference value = {[relative binding ratio, MS2 coat protein variant]/[relative binding ratio, wild type MS2 coat protein]}.

slightly better to Q102. Most of the exceptions that were noted were similar to this example: seeming omissions from the selection results. In some cases, this seeming omission may have been due to the small number of family members that were cloned (there were only three Family 13 aptamers). In other cases, though, it appeared as though an aptamer may have enjoyed a selective advantage in one population relative to another: Family 11 (8 members) bound equally well to E89T and Q102, but was only recovered from selections that targeted E89T; Family 16 (12 members) bound equally well to E89K and Q102, but was only recovered from selections that targeted Q102. The fact that selection results did not precisely follow measured affinities may have been the result of stochastic binding events ('jackpots') that occurred in early, stringent rounds of selection.

Conclusions

It can be argued that the selected aptamers are optimal binding sequences. The sizes of the selected motifs and structures were consistent with a thorough search of all sequences of length 22 or shorter (the calculated complexity of the original pool). For example, if we assume that the dual-specificity aptamers of Family 9 (Figure 3) arose independently of one another, then at least 19 residues must have become fixed during their evolution. Other aptamers that clearly arose independently of one another, such as different members of Families 2 or 11, may have fixed 20 or more residues during their evolution. Even if the aptamers are not considered to be optimal binding sequences, the multiple aptamer families can nonetheless serve as an ideal test set of RNA ligands for assessing whether selection for high-affinity binding also leads to highly specific binding.

While both natural and unnatural RNA ligands could discriminate between distantly related proteins (MS2 and Qβ coat proteins), the selected RNA ligands did not always discriminate between closely related proteins (the MS2 coat protein and its variants). Instead, aptamers that had high (and perhaps optimal) affinity for one target sometimes bound many of the closely related targets equally well (e.g., Family 15 had high but similar affinities for all the MS2 coat protein variants). The aptamer generalists that were recovered may have interacted with surface features that were shared in common between the targets (as in Figure 1b). We have previously observed such generalists during the selection of aptamers that bound to closely related protein kinase C isozymes [3,24]. While aptamer specialists were found, they showed very modest specificity for their cognate target relative to closely related targets (e.g., while Families 11–14 favored proteins containing the E89T substitution, they discriminated against the wild type MS2 coat pro-

tein containing glutamate at position 89 by only four- to fivefold; Figure 8). Overall, these results emphasize that selections for affinity do not of necessity beget specificity.

The finding that there is not always a correlation between selected binding affinity and binding specificity extends and refines analysis originally carried out by Eaton et al. [1]. Any correlations that may exist between binding affinity and binding specificity can best be observed when the distances between the sequences and structures of related protein targets are large, as Eaton et al. [1] also point out. Thus, ligands selected to bind different fibroblast growth factors or divergent coat proteins (MS2 and Qβ) do not bind well to non-cognate targets. Conversely, correlations between binding affinity and binding specificity may be lost when the distances between related protein targets grow so small that few or no RNA ligands can adequately make distinctions. Thus, ligands selected to bind MS2 coat protein or its closely related variants can bind well to non-cognate targets.

Our results can best be understood in terms of structural hypotheses and models. Both nucleic acids and proteins are constructed from monomers that have discrete chemistries and structures. There are no canonical nucleotides that are 'intermediate' between the purines adenosine and guanosine, nor canonical amino acids that fill the gap between the basic side chains of lysine and arginine. If nucleotides, amino acids, and ultimately their polymers are chemically and structurally quantized or granular (that is, capable of displaying only a limited range of chemical moieties on a limited range of shapes), rather than chemically and structurally continuous (capable of forming virtually any surface), then it should not be surprising that complexes between these polymers are similarly quantized or granular, and that complementarity between ligands and receptors may not always be unique.

To bolster the hypothesis that the specificities of selected interactions may be limited by the range of chemistries and structures that biopolymers can assume, we have begun to examine the structures of coat protein:aptamer complexes. As an initial effort, MS2 operator-like aptamer with a trinucleotide rather than a tetranucleotide loop was soaked into MS2 crystals and the co-crystal structure was solved [25]. The aptamer bound almost identically to the wild type MS2 operator. In order to achieve this, interactions between the stem of the aptamer and the protein binding site underwent a rearrangement that allowed the major specificity determinants to retain the same orientations relative to one another. These results are congruent with the hypothesis that binding interactions between biopolymers are limited to a discrete range of chemistries of structures. Thus, it should not be surprising that the binding specificities that can be achieved by biopolymers may also be inherently limited.

An interesting corollary to this analysis is that if all 'fits' of molecular interfaces are not possible, then the best possible 'fits' for related interfaces may differ significantly in sequence. In other words, if a target molecule changes even slightly, its best cognate ligand may have to change drastically in order to realign bonds and steric contacts and maintain the tightest possible interaction. In support of this hypothesis, we have observed that small changes in the sequence of a protein target were compensated for by large changes in the sequences and predicted structures of cognate RNA ligands. While aptamers that were mildly specific for the E89T variant of the MS2 coat protein could be selected, the sequence changes that were required to effect this recognition were relatively large. At the least, the single amino acid change in the protein elicited a three-base change (AUCA $\rightarrow$ UUUC) in the loop and the deletion of a bulged base in the MS2 operator, and perhaps required more involved sequence and structural changes as well (Figure 8).

The notion that searches through the vast sequence and structural spaces available to biopolymers do not always yield specific 'fits' to targets has important consequences for drug discovery. If selected biopolymers such as aptamers were perfectly specific and could meld themselves precisely to a target, then virtually any amino acid substitution at a molecular interface might sterically inhibit binding. However, if selected biopolymers such as aptamers are not perfectly specific, then at least some RNA binding species may be robust to amino acid substitutions. For example, the Family 15 generalists bound tightly to the original coat protein and were not deterred by substitutions at either position 87 or 89. Counterintuitively, because selected biopolymers such as aptamers bind over a relatively large surface area and cannot always achieve the tightest possible 'fit', they may in the end prove to be better drugs against targets that can mutate than small organic molecules that must rely on fewer, more precise, and sterically constrained interactions.

References

1. Eaton, B.E., Gold, L. and Zichi, D.A., *Let's get specific: The relationship between specificity and affinity*, Chem. Biol., 2 (1995) 633–638.
2. Bock, L.C., Griffin, L.C., Latham, J.A., Vermaas, E.H. and Toole, J.J., *Selection of single-stranded DNA molecules that bind and inhibit human thrombin*, Nature, 355 (1992) 564–566.
3. Conrad, R., Keranen, L.M., Ellington, A.D. and Newton, A.C., *Isozyme-specific inhibition of protein kinase C by an RNA aptamer*, J. Biol. Chem., 269 (1994) 32051–32054.
4. Bernardi, A. and Spahr, P., *Nucleotide sequence at the binding site for coat protein on RNA of bacteriophage R17*, Proc. Natl. Acad. Sci. USA, 69 (1972) 3033–3037.
5. Witherell, G.W., Gott, J.M. and Uhlenbeck, O.C., *Specific interaction between RNA phage coat proteins and RNA*, Prog. Nucleic Acid Res. Mol. Biol., 40 (1991) 185–220.
6. Spahr, P.F., Farber, M. and Gesteland, R.F., *Binding site on R17 RNA for coat protein*, Nature, 222 (1969) 455–458.
7. Valegard, K., Murray, J.B., Stockley, P.G., Stonehouse, N.J. and Liljas, L., *Crystal structure of an RNA bacteriophage coat protein-operator complex*, Nature, 371 (1994) 623–626.
8. Peabody, D.S., *The RNA binding site of bacteriophage MS2 coat protein*, EMBO J., 12 (1993) 595–600.
9. Lim, F., Spingola, M. and Peabody, D.S., *Altering the RNA binding specificity of a translational repressor*, J. Biol. Chem., 269 (1994) 9006–9010.
10. Lim, F., Spingola, M. and Peabody, D.S., *The RNA-binding site of bacteriophage Qβ coat protein*, J. Biol. Chem., 271 (1996) 31839–31845.
11. Spingola, M. and Peabody, D.S., *MS2 coat protein mutants which bind Qβ RNA*, Submitted.
12. Peabody, D.S., *Translational repression by bacteriophage MS2 coat protein expressed from a plasmid*, J. Biol. Chem., 265 (1990) 5684–5689.
13. Miller, J.H., Experiments in Molecular Genetics, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1972.
14. Schneider, D., Tuerk, C. and Gold, L., *Selection of high affinity RNA ligands to the bacteriophage R17 coat protein*, J. Mol. Biol., 228 (1992) 862–869.
15. Witherell, G.W. and Uhlenbeck, O.C., *Specific RNA binding by Qβ coat protein*, Biochemistry, 28 (1989) 71–76.
16. Conrad, R.C., Giver, L., Tian, Y. and Ellington, A.D., *In vitro selection of nucleic acid aptamers that bind proteins*, Methods Enzymol., 267 (1996) 336–367.
17. Lato, S.M., Boles, A.R. and Ellington, A.D., *In vitro selection of RNA lectins: Using combinatorial chemistry to interpret ribozyme evolution*, Chem. Biol., 2 (1995) 291–303.
18. Giver, L., Bartel, D., Zapp, M., Pawul, A., Green, M. and Ellington, A.D., *Selective optimization of the Rev-binding element of HIV-1*, Nucleic Acids Res., 21 (1993) 5509–5516.
19. Talbot, S.J., Goodman, S., Bates, S.R.E., Fishwick, C.W.G. and Stockley, P.G., *Use of synthetic oligoribonucleotides to probe RNA–protein interactions in the MS2 translational operator complex*, Nucleic Acids Res., 18 (1990) 3521–3528.
20. Higgins, D.G. and Sharp, P.M., *CLUSTAL: A package for performing multiple sequence alignment on a microcomputer*, Gene, 73 (1988) 237–244.
21. Jaeger, J.A., Turner, D.H. and Zuker, M., *Predicting optimal and suboptimal secondary structure for RNA*, Methods Enzymol., 183 (1989) 281–306.
22. Beckett, D., Wu, H. and Uhlenbeck, O.C., *Roles of operator and non-operator RNA sequences in bacteriophage R17 capsid assembly*, J. Mol. Biol., 204 (1988) 939–947.
23. Stockley, P.G., Stonehouse, N.J., Murray, J.B., Goodman, S.T.S., Talbot, S.J., Adams, C.J., Liljas, L. and Valegard, K., *Probing sequence-specific RNA recognition by the bacteriophage MS2 coat protein*, Nucleic Acids Res., 23 (1995) 2512–2518.
24. Conrad, R. and Ellington, A.D., *Detecting immobilized protein kinase C isozymes with RNA aptamers*, Anal. Biochem., 242 (1996) 261–265.
25. Convery, M.A., Rowsell, S., Stonehouse, N.J., Ellington, A.D., Hirao, I., Murray, J.B., Peabody, D.S., Phillips, S.E.V. and Stockley, P.G., *Crystal structure of an RNA aptamer-protein complex at 2.8 Å resolution*, Nat. Struct. Biol., 5 (1998) 133–139.