Supporting Methods

Sampling error

Treating the per-position nucleotide frequencies (F_i) as though they were actual probabilities (P_i) results in a sampling error in the information content calculation. Schneider et al.²¹ demonstrated a method for calculating the standard error by computing the weighted uncertainties for all possible combinations of bases for a given number of sequences in an alignment. To account for the reduction in uncertainty due to a finite number of sequences, we subtract a correction factor from the information content calculated at each position. The correction factor is the difference between the known maximum uncertainty (2 bits/position for RNA) and the calculated expectation value for the maximum uncertainty in an alignment of a certain size. To compute the values for the expected maximum uncertainty, error correction and variance per position, we implemented a version of CALHNB by Schneider et al.²¹ in Perl. We modified their algorithm to adjust the frequency distributions obtained from the sequence alignments to compute the information content required to specify the active structures in completely random RNA sequence space. The values were computed for all of the alignments with prior base probabilities of 0.79, 0.07, 0.07, 0.07. The chart below gives a sense of the magnitude of these values.

Number of sequences	Expectation value (bits)	Error correction (bits)	Variance per position
30	1.74	0.26	0.060
40	1.82	0.18	0.034
50	1.86	0.14	0.020
60	1.89	0.11	0.013
70	1.91	0.09	0.009
80	1.92	0.08	0.006

Normalizing and adjusting base frequencies

We computed the information content of each aptamer using the selection sequence alignments. The sequence pools were synthesized with 21% rate of mutagenesis, where each of the non-wild-type bases was equally likely. Because we know the prior probability of each base type *i* at every position, we can use the frequency distribution of nucleotides from the alignment to compute the information content of the aptamers in relation to random ($P_A=P_U=P_C=P_G$) sequence space.

As an illustration (see below), consider an aptamer that requires a purine (A or G) in a certain position to be functional. For simplicity imagine that we synthesize a pool of 100 aptamer sequences mutated 21% at that position with A as the original base. In the idealized case we would have 79 sequences with an A, 7 with a U, 7 with a G, and 7 with a C. After selection, only the A- and G-containing sequences remain. The total number of sequences in the pool is reduced to 86, where 79 are A and 7 are G. The relative frequency of A and G after the selection reflects the fact that A and G are equally compatible with a functional molecule and that A was present much more often in the starting pool than G. To normalize each F_i for the skewed starting frequencies we divide each F_i after the selection by the F_i before the selection. Dividing each normalized F_i by the sum of the normalized F_i s gives the adjusted frequencies of each base type relative to the others, In the current example, the normalized and adjusted F_A =0.5 and F_G =0.5; the information content of this position is 1 bit, which is the information content needed to specify a two-base varying position in an RNA structure.

Before Selection

After Selection

Normalized

base	# seqs	frequency
A	79	0.790
U	7	0.070
G	7	0.070
С	7	0.070
Total	100	1.000

base	# seqs	frequency
A	79	0.919
U	0	0.000
G	7	0.081
С	0	0.000
Total	86	1.000

-0

2	base	# seqs	frequency
ñ	A	79	1.160
3	U	0	0.000
3	G	7	1.160
ŝ	C	0	0.000
ŝ	Total	86	2.320

Ξ.

Adjusted

base	# seqs	frequency
A	79	0.500
U	0	0.000
 G	7	0.500
C	0	0.000
Total	86	1.000

Supporting Figure 1A-K. GTP aptamer sequence alignments. DNA sequence alignments from each of the selections are shown in order from highest affinity for GTP to lowest. The top line in each alignment contains the minimized sequence (corresponding to the length of the optimized sequence). This is the sequence used as the template for synthesizing the library of mutants. Immediately below that is the secondary structure model of the optimized aptamer shown in bracket form where the 5' -most open bracket "(" base pairs with the 3'-most closed bracket ")". Paired strands are shaded with same color. Point mutations in the loops that were assayed are indicated on the same line as the secondary structure. Blue bases were tested and found not to improve binding activity. Red bases improved the activity of the aptamer relative to the minimized construct and were incorporated into the optimized version.

Mutations in the stem regions of the alignment are shaded to indicate:

Red = Watson-Crick covariation Orange = new Watson-Crick base pair Green = new wobble (A-C or G-U) pair Black = broken pairing

Mutations in the loop are shaded purple

Figure 1A. 9 – 4.

Figure 1B. Class V.



Figure 1C. 10 – 10.





Figure 1E. 10 – 59.

				10						12	20						J3	0						140					1:	50
TG ((G T ((A A ((G T ((TC ((CCA	A	A	GG	C	GG	G	TI	G	G	AA	G	A G	A 1	A	ТС	Α	T	A G	GA))	5	CT))	T ()	10	G T	
TG	GT	AA	GT	TC	000	A	A	GGG	C	666	GGC		660	GI	AA	GG	IG	AC	A	TOTO	A	T	AGAC	AA	A	T	TG	TO	AT	
CG	GC	A A A A	GT	TC	000	A	A	GGC	c	GGG	GG	††	G	GG	A	G	AG	AI	A	AC	A	Ť	AG	GAG	G	T	TG	TO	GG	
T G	GC	C A A A	GT	TC	CA	A	A	GG	č	GG	G	i i T i	G	GI	AA	G	AG	AT	A	TO	A	T	AG	G A G A	G	T	GGA	TO	GT	
TG	G G T	A A A A	G A G A	T C T C	CCC	AA	A	GGG	ĉ	GGG	G	AT	G	G	A A A A	G	A G A G	AA	AA	TC	A	T	A G A G	G A G A	GA	T	TG	TO	AT	
TG	GT	A A A A	GT	T A T C	CA	A	A	GG	Ċ	GGG	G		G	G	AA	G	AG	AA	A	TO	A	T.	A G	GA	G	C T	GG	CC	AT	
G	GT	AA	GI	TC	CA	A	AI	GG	ĉ	GG	G	TI	GG	GI	AA	G	AG	AI	A	TO	A	Ţ	AG	GA	G		TA	000	GI	
G	GŤ	AA	GC	TC	000	A	A	GG	č	GGG	G		G	GI	A A A A	ţ	AG	AI	A	TO	A	Ţ	AG	GA	A	T	TG	CO	GI	
T G	GT	A A A A	GT	ŤĊ	cc	A	A	GG	č	GGG	G	ij	GC	G	T A	G	AG	AI	A	İ	A	Ì.	AG	GAG	G	T	TG	TO	GT	
G	GT	AAAA	ĞŤ	ŤČ	cc	A	A	GGG	T	GGG	GG	İİ	G	GI	T A	G	AG	AI	A	İČ	A	İ	AG	GAG	G	T	İÏ	CO	GT	
G	GC	AAA	GT	TGTC	CC	A	A	GG	C	GG	G	GT	G	G	A A A A	G	A G A G	AT	A	TO	A	Ť	GAG	C G A	A	T	TG	TO	AT	
G	G T G T	C A A A	GT	T C T C	CCC	A	A	GG	C C	GGG	G		G	GI	A A A A	G	AG	AA	A	TO	A	T	A G A G	G A G A	G	T	TG	TO	GG	
G	GT	A A A A	GT	TC	CCC	A	A	GG	ĉ	GGG	G	T T	G	G	A A A A	G	AG	AT	A	TO	A	T	AG	GA	G	C T	TG	TO	GI	
G	GT	AA	GT	TC	000	A	A	GGG	C C	GGG	GG		G	GI	A A	GG	AG	AI	A	TO	A	Ţ	AG	GA	G	Ť	TA	TO	GT	
G	GT	AA	GT	TC	000	A	A	GG	C	GGG	G		GG	GI	AA	G	AG	ATAT	A	to	A	Ţ	AG	GA	G	T	TG	çç	GG	
G	GC	AA	GT	TC	000	A	A	GG	č	GGG	G	ij	GG	G	F A	G	AG	AI	A	ŤČ	A	Ţ	AG	GA	G	T	TG	TO	ĊŢ	
G	ČŤ	AA	Ğ Ť G Ť	ŤĊ	cc	A	A	GGG	č	GGG	G	t i	G	G	AA	G	AG	AI	A	İ	A	Ť	AG	GA	G	Ť	TG	TO	AT	
G	GT	AA	GT	ŤČ	ĉ	A	A	GG	TA	GGG	G	ĨI	G	G	AA	G	AG	AI	A	İC	A	İ	AG	GAG	G	T	TG	TO	GC	
G	GT	A A A A	GT	TC	CC	A	A	GGG	ċ	ġ Ġ	G	TT	G	Ġ	AA	G	AG	ATAT	A	TO	A	T	A G A G	G A A A	G	T	TG	TO	GGG	
G	G T C T	C A A A	GT	TA	CCC	A	AI	GG	C	GG	G	TC	G	GI	A A A A	G	A G A G	AAA	AA	TO	A	T	A G A G	GA	G	C T	TG	ACAC	GT	
G	AT	AA	AT	TC	CC	A	A	GG	c	GG	G		G	GI	AA	G	AG	A	A	TO	A	T	AG	GA	G	C T	TG	TO	GI	
G	GC	G A A	GG	TC	CC	A	A	GGG	c	66	G		G	GI	A A	GG	AG	AI	A	TO	A	ţ	CT	GA	T	TT	TG	IC	GI	
Ğ	GT	AA	GT	GC	CA	G	A	GGG	č	GG	G		G	GI	r A	G	AG	AT	G	TO	A	T	AG	GC	G	È T	TG	TO	Ğ	
G	GT	AA	GT	TC	CCC	Â	A	GGG	č	GG	G	İ.	G	GI	AA	G	AG	AT	A	TO	A	T	AG	GA	Ğ	T	TA	AC	ŤĠ	
G	GT	AA	GT	ŤĊ	CA	G	A	GG	č	GG	Ğ	İİ	G	G	AA	G	GG	AA	A	İ	A	Ť	AG	GAGA	G	T	CG	T	GT	
AG	GT	AAAA	GT	GC	CC	A	A	GG	C	GG	G	T T T T	G	G	A T A A	G	AG	ATAT	AA	TO	A	Ť	AG	GA	G	T	TA	G	GI	
G	GT	A A A A	CTGC	CCTC	000	A	A	GG	ĉ	GGG	G	TT	G	GA	A A A A	G	A G A G	AT	A	TC	A	T	GG	G A G A	G	T	TG	TO	G	
G	GT	A A A A	GT	T C T C	CCC	A	A	GG	C C	GGG	G		G	G	G A	G	A G A G	AI	A	TO	A	T	A G	G A G A	G	T	TG	TO	AT	
GG	GT	AGAA	GC	TC	AC	A	A	GG	ĉ	GGG	Ġ	TA	G	G	AA	AG	CG	AI	A	TO	A	Ţ	AG	GAGA	G		TG	CC	GT	
GG	GT	AA	GC	TC	000	A	A	GG	C	GGG	GG		GG	G	AA	GG	AG	AI	A	IC	A	T	AG	GA	G	T	AA	CC	GI	
G	GG	AA	GT	CCA	100	A	A	GGG	C	GGGG	GGC		G	GI	AA	G	AG	AI	A	TO	A	T	AG	GGG	G	Ť	TG	TO	GT	
G	TT	A A A A	GT	TC	CC	A	A	GG	C	GGGG	Ğ	İİ	GC	GI	AA	G	TG	AI	A	TO	A	Ť	AG	GA	G	T	TG	AC	GI	
r G A G	GT	AAAA	GT	TC	000	A	A	GGG	ČG	GGG	GG	İİ	G	GI	AA	GG	AG	AT	A	TO	A	Ť	AG	GAG	G	TA	TG	TO	CI	
F G F G	GT	A A A A	GT	TC	CCC	A	A	GG	T	GGG	GG	TT	G	G	AA	G	AG	ATAT	TA	TO	A	T	A G A G	G A G A	G	T	TG	TO	T I G T	

Figure 1F. 10 – 24.

•	10	J -	- 2	24	•																																									
1							1	11)				 	 12	0						13	0									ŀ	40									1	50	2			
					A CAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA	TCTTTTTTTTTTTTTTTTTTTTTTTTTTTT		1 T . T . T . T . T . T . T . T . T . T		TATTTTTTTTTTAATTTT				2 T TTTTTCCTTTTTTTTTTAAGA	T TTTTTTATTTGTTTTTTT	C TUUUUUUUUUUUUUUUUUUUU	000×0000000000000000000000000000000000	A O O O O O O O O O O O O O O O O O			3 A AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA	0 A AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA		C TOCAAACCGCGCCCCACCCC	T AAATTTTTTTTTTATTTTT	CT · T CTCCAATTTTTTCCCCC	Ċ	T TTTTTTTTTTTTTTATCTT	A AAATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA	CACCCCAACCCCCCCACCAAC			A			· · · · · · · · · · · · · · · · · · ·	0.000000000000000000000000000000000000	CTTTCTTACCCTTTTTTCCCCTCT	G) GOT GOGGG A GOGGGGGGGGGGGGGGGGGGGGGGGGG	ADAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA				CONTROCTOR	00000000000000000000000000000000000000	
JOG G G G G G G G G G G G G G G G G G G				CCCCTCCCC	AGAATAAAA	TTTGTAT T		TTTTTG	AAAAAAAGAA	TTTTTA				CTTTTTC		LOCOCACCOC	STOP-SSTOP	CUUUUAAUUUU	GCCTTCA		AAAAAAAAAA	AAAAGGAA	LOCOCOCACO	COTACCECTO	TTTTTTAT	CTCACCCTC		ATCATGTTT	AAAAAAAAAAA	LOCACCGCCA		AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA	AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA		T	cc c	CCCCCACCCC	TOTGCACOO		AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA				COTOTTTOO	0000000000000	CONTROCOCCO

Figure 1G. 9 – 12.

1	10			20			130	40	
ACTCTA	GCAGTI	CAGGT	AA	CCAC.	GT	AAGA	TACG	GTGCTAGA	A
ACTCTA	GCACTI	CAGGT	AA	CCAC.	ĠŤ	ÀÀĠĂ	GÀCĠ	GTGCTAGA	G
ACTCTA	GCAGTT	CAGGI	AA	CCAC.	GT	AAGA	TACG	G T G C T A G A G T G C T A G A	A
ACTCTA	GCAGTI	CAGGT	AA	CCAC.	GT	AAGA	TACG	GTGCTAGA	G
ACTCTA	GCAGTI	CAGGT	AA	CCAC.	ĞŢ	GAGA	TACG	GTGCTAGA	A
ACTCTA	TCAGTT	CAGGT	AA	CCACA.	Ġ Ŧ ĺ	AAGA	TACG	GTGCTAGAG	A
ACTCTA	GCAGTI	CAGGT	AA	CCAC.	GT	AAGA	TACG	G T G C T A G T G T G C T A T A	A
ACTCTA	GCAGTT	CAGGT	AA	CCAC.	GT	AAGA	AGCG	GTGCTAGA	G
ACTCTA	GCAGTI	CAGGT	AA	CCACA	GT	AAGA	TACG	GTGCTAGA	A
A C T C T A G T G C T A	GATTTI	TAGGT	AA	CCAC.	GT	ACGA	TACG	G T T C T A G G G A G C T A G C	G
ACTCTA	GCTGTT	CAGGT	AA	CCAC.	GT	AAGA	TACG	GTGCTGGG	A
ACTCTA	GCAATI	CAGGT	AA	CCAC.	ĞT	AAGA	TACG	GTGCTAGG	T
ACTCTA	TCAGTT	TAGGT	AA	CCAC.	GT	A A G A A A G A	TACG	G T C C T A G A G T G A T A G A	A
ATTCTT	GCAGTT	CAGGT	AA	CCAC.	GT	AAGA	TACG	G T G C G A G A	A
ACTCTA	GCAGTI	CAGGT	AA	CCACA	GI	AAGA	TACG	GTGCTAGA	A
ATTGTA	GCAGTT	CAGGT	AA	CCAC.	GT	A	AACG	GTGCTACA	A
ACCCTA	GCAGTT	TAGGT	AA	CCAC.	GT	AAGA	TACG	G T G C T A G G G T G A T A G A	A
ACTCTA	GCAGTI	CAGGT	AA	CCAC.	GT	AAGA	TACG	GTGCTAGT	A
ACCCTA	GCTGTT	CAGGT	AA	CCAC.	GT	AAGA	TACG	GTGCTAGT	G
A C T C T A G C T C T A	GCTGTT GCAATT	CAGGT	AA	CCAC.	GT	A A G A A A G A	TACG	G T G C T G G G G T G C T A G A	A
GCTCTA	GCAATT	CAGGT	AA	CCAC.	GT	AAGA	TACG	GTGCTAGA	TO
ACTCTA	GCAGTI	CAGGT	AA	CCAC.	GT	AAGA	GACG	GTGCTAGA	T
ACTCTA	GCAGTI	CAGGT	AA	CCAC.	ĞŤ	AAGA	AACG	GTGCTAGA	G
ACTCTA	GTAGTT	TAGGT	AA	CCAC.	GT	AAGA AAGA	TACG	G T A C T A G A G T A C T A G A	A
ACTCTA	GGAATI	CAGGT	AA	CCAC.	GT	AAGA	TACG	G T C C T A G A G C C C T A C A	A
ACTCTA	GCAGTI	CAGGT	AA	CCAC.	GT	AAGA	GACG	GTGCTAGA	G
ACTCTA	GCTGTT	CAGGT	AA	CCAC.	GT	AAGA	AACG	GTGCTAGA	G
A C T C T A G C T C T A	GCAGTT	CAGGT	AA	CCAC.	GT	AAGA	AGCG	G	G
ACTCTA	GCAGTT	TAGGT	AA	CCAC.	GT	AAGA	TACG	GTGCTAGA	A
ACTCTA	GCAGTI	CAGGT	AA	CCAC.	GT	AAGA	TACG	GTGCTAGA	A
ACTCTA	GCAATT	CAGGT	AA	CCAC.	GT	AAGA	TACG	GTGCTAGG	T
ACTCTA	GCAATT	CAGGT	AA	CCAC.	GT	AAGA	TACG	G C G C T A G A G C G C T A G A	A
ACTCTA	GCAATT	CAGGT	AA	CCAC.	GI	AAGA	TACG	GTGCTCGA	A
ACTCTA	GCTATT	CAGGT	AA	CCAC.	GT	AAGA	TACG	GTGCTAGA	A
ATTCTA	G C A G T T	TAGGT	AA	CCAC.	GT	AAGA	AACG	G T G C T A G A	A
AGTCTA	GCAGTT	TAGGT	AA	CCAC.	GT	AAGA	TACG	G T G C T A G A G T G C G A G A	CA
ACTCCA	GCAGTI	TAGGT	AA	CCAC.	GT	AAGA	GACG	GTGCTGGA	A
ACTCT	GCAATT	CAGTT	AA	CCACA	T	AAGA	AACG	GTGCGGAG	TO
ACTCTA	TCGTTT	TAGGA	AA	CCAC.	GT	AAGA	TACG	G G G C T A C A G T G A T A G A	TO
GCTGTA	GCAATT	CAGGT	AA	CCAC.	IT	AAGA	AACG	GAGCTATG	60
ACTGTA	GTAATT	TAGGT	AA	CCAC.	GT	AAGA	AACG	GTACTGTA	G

Figure 1H. 10 – 6.



Figure 11. Class II.

1									20	10	Ę.									20	0								1	13
	Ą	Ģ	ç	C	G	G	10	A	A	G	A	G	C	A	C	G	Т	A	T	A	C	G	С	A	G	G	G	ç	Ţ	0
	G	ĉ	ĉ	A	AG	À	-	Ā	Ġ	Ť	Å	Ġ	ċ	Å	ċ	A.	Ť	Å	A	Ā	ċ	Ġ	ċ	Ā	AA)	G	G	C	0
	Ā	G	G	Ţ	G	G	+	A	A	G	A	G	C	A	C	G	A	A	Ţ	A	C	G	C	A	G	G	C	Ċ	T	9
	A	G	6	C	G	G	÷.	A	A	G	A	6	ĉ	A	ĉ	6	A	A	÷	A	ĉ	G	C	A	G	G	C	č	÷	2
į	A	Ġ	č	č	A	Ğ		A	A	Ğ	A	G	č	A	č	G	Ť	A	ċ	A	č	Ğ	č	A	A	G	A	č	Ť	(
	A	Ţ	Ţ	C	A	G	-	A	A	G	A	G	C	A	C	G	T	A	T	A	ç	G	C	A	A	G	T	T	T	(
	A	G	č	è	A	G		A	A	G	A	G	č	A	č	Ğ	Ť	A	A	A	č	G	č	A	A	Ğ	G	Ť	Ť	1
-	C	G	C	C	G	G		A	A	G	A	G	C	A	C	G	Ţ	A	T	A	ç	G	C	A	G	G	G	C	G	ł,
	A	G	T	A	A	G	100	A	A	G	A	G	č	A	è	G	÷	A	A	A	è	G	č	A	A	T	G	C	c	1
	A	G	Ċ	C	A	G	-	А	A	G	A	G	C	A	C	G	Ċ	A	T	A	Ĉ	G	C	A	A	G	G	Ċ	T	1
	A	G	C	T	G	G	Ť	A	A	G	A	G	C	A	C	G	A	A	4	A	ĉ	G	C	A	G	A	G	C	A	-
	A	G	č	č	Ğ	G		Â	Ã	G	Â	G	č	Â	č	A	Ť	Â	Â	A	č	G	č	A	Ğ	G	G	Ť	Ŧ	1
	A	G	C	C	Ģ	G	2	A	A	G	A	G	ç	A	ç	G	T	A	Ţ	A	ç	G	C	A	G	G	G	ç	C	1
	A	A	č	ĉ	G	G		A	A	G	G	G	č	A	č	G	A	A	÷	A	č	Ğ	č	A	G	G	G	t	÷	1
	A	A	C	C	G	G		A	A	G	A	G	C	A	C	G	T	A	Ť	A	C	G	C	A	G	A	G	T	T	1
	A	A	ĉ	C	G	G	-	A	A	G	A	G	C	A	C	G	T	A	C	A	ç	G	C	A	G	G	G	A	T	-
	A	t	č	č	G	Ğ	-	A	A	G	A	G	č	A	č	G	ċ	A	Ť	A	č	G	č	A	Ğ	Ğ	G	G	Ť	1
	A	Ţ	C	C	G	G	i.	A	A	G	A	G	C	A	C	G	T	A	Ţ	A	C	G	C	A	G	G	G	A	Ţ	
	A	t	t	è	A	G	-	A	A	G	A	G	č	A	č	G	÷	A	÷	A	č	G	č	A	G	G	G	A	÷	-
	A	Ţ	C	C	G	G	+	A	A	G	A	G	C	A	C	G	T	A	Ţ	A	C	G	C	A	G	G	G	A	T	
	A A	ł	c	è	G	G	÷.	A	A	G	A	G	č	A	c	G	÷	A	A	A	č	G	C	A	G	G	G	G	÷	-
	A	Ť	č	A	G	G		А	A	Ğ	A	G	č	A	č	G	Ť	A	Ť	A	č	Ğ	č	A	G	T	G	A	Ť	-
	A	G	A	C	G	G	-	A	A	G	A	G	C	A	ç	G	T	A	Ţ	A	ê	G	ç	A	G	G	T	ç	T	
	A	Ğ	A	è	G	Ğ		A	A	G	A	G	č	A	č	Ğ	ċ	Â	Ť	Â	č	Ğ	č	Â	Ğ	G	t	è	Ť	1
	A	G	A	C	A	G		A	A	G	A	G	C	A	ç	G	T	A	T	A	ç	G	C	A	A	G	Ţ	ç	T	
	A	G	A	č	G	G	1	A	A	G	A	G	č	A	č	G	T	A	ċ	A	č	G	č	A	G	G	ť	č	÷	ş
	A	Ģ	A	Ç	A	G	4	A	A	G	A	G	C	A	C	G	T	A	C	A	C	G	C	A	A	G	ç	C	Ţ	
	A	G	A	C	G	G		A	A	G	A	G	č	A	ĉ	G	Ť	A	A	A	č	G	ĉ	A	G	G	t	ĉ	Ť	
	A	G	A	C	G	G		A	A	G	A	G	C	A	Ĉ	G	Ť	A	Å	A	C	G	Ċ	A	G	G	Ť	Č	Ť	1
	A	G	AT	T	GA	G	÷	A	A	G	A	G	C	A	C	G	T	A	T	A	C	G	C	A	G	A	T	C	T	
	A	G	c	č	A	G	-	A	A	G	A	G	č	A	č	G	Ť	A	Ť	A	č	G	č	A	A	G	G	č	Ť	1
10	A	G	T	C	A	G	0.0	A	A	G	A	G	C	A	C	G	Ţ	A	C	A	C	G	C	A	A	G	A	C	G	1
i.	A	2	A	F	A	6	-	A	A	6	A	6	F	A	c	6	A	A	+	A	F	6	T	A	A	6	T	5		

Figure 1J. Class IV.

11									11	0									12	0									13	0									4	-0		
G	G	A	G	C	A	С	A	A	A	A	A	G	G	A	G	G	A	A	G	Т	G	G	T	A	A	G	A	С	A	Т	С	A	Т	G	T	G	Т	G	C	Т	С	C
((((((((C	Т			•				•								•								•		A))))))))
G	G	A	A	C	A	G	A	A	A	A	A	G	G	A	G	G	A	A	G	Т	G	G	Τ	A	A	G	A	C	А	Т	C	A	T	A	Т	C	Т	G	G	T	T	C
G	G	A	A	C	A	T	А	А	A	A	A	G	G	А	G	G	A	A	G	Т	G	G	Т	A	A	G	A	C	A	Т	C	A	Т	A	T	G	T	G	C	Т	C	C
Ģ	G	A	G	C	A	C	G	A	A	A	A	G	G	A	G	G	A	A	G	Т	G	G	Т	A	A	G	A	Α	A	Т	C	A	Т	A	T	G	C	G	Т	Т	C	C
G	G	A	G	C	G	Ċ	A	C	T	A	A	G	G	A	G	G	A	A	G	T	G	Ģ	T	A	A	G	A	C	A	Т	C	A	Т	G	Т	G	T	A	C	Т	C	Ċ
G	G	A	G	C	A	C	A	T	Ţ	A	A	G	G	A	G	G	Ą	Ą	G	Ţ	G	G	Ţ	A	Ą	G	A	C	Ą	Τ	Ç	A	Ţ	G	Ţ	G	T	Ţ	T	T.	Ċ.	C
G	G	A	G	I	A	C	A	G	C	Ą	A	G	G	A	G	6	A	Ą	G	Ţ	G	G	Ţ	A	A	G	A	C	A	Ţ	ç	A	T	C	Ţ	G	1	A	C	I	C	C
G	G	C	G	1	G	C	A	A	A	A	A	G	G	G	G	G	A	A	G	1	6	G	1	A	A	6	A	Č.	A	1	ç	A	1	6	1	G	1	6	ç	G	5	C
G	6	ć	6	5	A	1	A	A	A	A	A	6	6	A	6	6	A	A	6	1	6	G.	1	A	A	Ģ.	A	5	A	1	ç	A	1	6	1	6	1	6	÷	6	ć	C.
6	6	A	6	5	A	6	A	A	A	A	A	6	6	A	6	6	A	A	6	A	6	6	÷	A	A	9	A	5	A	÷	ç	A	÷	1	÷	6	÷	6	1	Į.	5	2
6	6	A	6	5	A	5	A	A	A	A	A	6	6	A	6	6	A	A	6	A	6	5	4	A	A	6	A	2	A	÷	5	A	+	5	÷	6	+	5	5	Ļ	è	2
6	G	A	10	2	A	2	A	C	C	A	A	4	5	A	5	5	A	A	5	+	5	5	+	A	A	2	A	2	A	+	2	A	÷	2	+	4	÷	5	2	+	2	2
6	4	A	5	2	A	2	A	5	A	A	A	5	6	A	6	6	A	A	5	+	6	5	+	A	A	2	A	2	A	÷	2	A	+	5	+	5	÷	4	2	÷	2	2
2	6	A	2	2	A	2	A	G.	5	A	A	2	6	A	2	6	A	A	2	÷	6	4	÷	A	A	5	A	2	A	4	2	A	÷	2	÷	2	÷	5	2	÷	2	2
6	6	A	6	2	A	2	A	A	G	A	A	6	G	A	6	6	A	A	6	+	6	5	÷	A	A	5	A	2	A	÷	2	A	+	5	÷	2	÷	5	2	÷	2	2
6	C	A	C	F	A	è	A	A	C	A	A	2	6	A	6	6	A	A	6	÷	2	2	+	A	A	2	A	7	A	÷	č	A	÷	200	+	6	÷	2	2	+	Ť	2
T	2	A	2	F	A	č	A	A	A	A	A	E	e c	A	C	E	A	A	C	r	C C	č	+	A	A	č	A	F	A	÷	2	A	÷	č	÷	ê	÷	č	č	÷	-	č
ċ	č	A	č	è	A	č	A	A	A	A	A	č	C.	A	č	č	A	A	č	č	C	č	÷	A	A	C.	A	è	A	÷	č	Â	÷	č	÷	C.	÷	C	č	÷	è	č
C	c	A	č	è	A	č	A	A	Ť	Â	A	č	C.	A	č	č	A	Ä	č	T	C.	č	÷	A	A	č	A	č	A	÷	è	A	÷	č	÷	č	÷	č	è	÷	è	č
C	č	Â	č	è	Δ	č	A	Δ	Δ	Â	A	č	č	Â	č	č	A	Â	č	Ť	č	č	÷	Ť	Δ	č	A	č	Δ	÷	č	A	÷	A	Å	č	÷	č	č	Ť	č	č
č	Ğ	A	G	è	A	č	A	A	T	A	A	G.	Ğ	c	G	č	A	A	G	Ť	G	G	Ť	÷	A	G	A	è	A	÷.	č	A	Ť	G	Ŧ	G.	÷	G	č	Ť	č	č
T	A	A	G	č	A	C	A	A	ċ	A	A	G	Ğ	Ā	Ğ	Ğ	A	A	Ğ	Ť	G	G	Ť	Å	A	G	A	č	A	Ť	č	A	Ť	G	Ť	G	Ť	G	č	Ť	T	T
G	G	A	G	Ċ	A	C	A	G	A	A	A	G	G	A	G	G	A	A	G	Ť	G	G	Ċ	A	A	G	A	C	A	Ť.	Ċ	A	Ť	A	Ť	G	Ť	G	Ċ	Ť	ċ	Ċ
Ġ	G	G	G	Ċ	A	C	A	A	A	A	A	G	G	А	G	G	A	A	G	Ť	G	G	Ť	A	A	G	A	C	A	Ť	Ċ	A	Ť	A	Ť	G	A	G	Ċ	C	C	C
G	G	A	G	C	A	C	Т	A	Т	A	A	G	G	A	G	G	A	A	G	Т	G	G	Т	A	A	G	A	C	A	Т	C	A	Т	T	T	G	T	G	C	T	Т	Ċ
G	G	A	G	C	A	C	C	C	G	A	A	G	G	G	G	G	A	A	G	A	G	G	Т	A	Α	G	A	C	Α	T	C	A	T	G	A	G	Т	G	C	Т	C	G
G	A	A	G	C	A	C	A	C	A	A	A	G	G	G	G	G	A	A	G	Т	G	G	Т	A	A	G	A	C	A	Т	C	A	Т	A	T	G	Т	G	С	Т	T	C
G	G	G	C	A	C	A	A	G	T	A	A	G	G	А	G	G	А	А	G	Т	G	G	T	А	A	G	A	C	А	Т	C	А	Т	G	T	G	т	G	C	T	C	C
A	G	A	C	C	G	T	A	C	A	A	A	G	G	A	G	G	А	A	G	T	G	G	A	А	A	G	A	C	А	Т	C	A	Т	G	C	G	Т	A	С	T	C	C
G	G	C	G	C	A	T	G	A	А	A	A	G	G	A	C	Т	А	А	С	T	G	G	T	A	А	G	A	C	А	Т	C	A	C	G	Т	G	C	G	C	Т	C	C
C	G	G	G	C	A	G	A	A	A	T	А	G	G	A	G	G	A	A	C	T	A	G	T	A	A	A	A	A	A	Т	C	A	G	C	T	G	Т	G	С	Т	C	C
G	G	A	A	C	A	C	A	A	А	A	A	A	G	А	G	G	А	T	A	T	G	G	Т	А	А	A	A	C	A	Т	C	А	G	G	C	A	Т	A	A	T	C	C
G	C	A	G	C	A	C	A	A	A	A	A	G	G	A	Т	G	A	G	G	Т	G	A	Т	A	A	G	A	G	C	G	C	A	Т	C	G	G	Т	T	Т	A	С	C
G	G	A	G	A	C	C	A	Т	A	A	A	A	G	A	G	G	T	A	G	A	C	G	T	Т	А	G	A	С	A	A	C	A	A	G	Т	G	С	G	С	Т	G	Т

Figure 1K. Class III.



Supporting Chart 1. Original, minimized, and optimized sequences. The chart contains DNA sequences for each aptamer included in our analysis. The top in each group, marked 'ori', is a clone that was identified from the original GTP aptamer selection of Davis and Szostak¹⁹. Aptamers referred to by number indicate that only one isolate was identified. Aptamers named Class I-V originated multiple times as judged by the presence of diverse flanking sequences - these suggested 5' and 3' boundaries for the functional regions ('ori K_d construct'). For aptamers with multiple isolates (Class I-V), the specific original sequence chosen to derive the optimized versions is displayed. The 'min K_d construct' is the minimized functional sequence determined by deletion end mapping. 'min' is the minimized sequence flanked by constant primer regions and was employed as the template for selecting active sequence variants. The last sequence, 'opt', is the optimized version of each aptamer. K_ds are shown in parentheses. The constant primer-binding sites included for RT-PCR are shown in blue, where applicable. Sequence changes, relative to the original isolate, are shown in red. Occasionally unpaired G's were appended to the 5' ends of the 'min K_d construct' and 'opt' constructs to increase transcription yields; these are colored green.

The optimized sequences will be deposited in GenBank. All sequences, (including selection alignments) are available in FASTA format on our website:

http://genetics.mgh.harvard.edu/szostakweb/publications/extra/informationalcomplexitypaper2003.html

9-4

ori (25 nM)

5'-GGAGGCGCCAACTGAATGAAGTTGCCAGCTGCGAGCACGTGAATAGACTGCTTCGGCAGTGTCTCG ACGTGTGTAGGGGAAAGTATCCTCCGTAACTAGTCGCGTCAC-3'

min (Kd construct 25 nM)

5'-GGAGGCGCCAACTGAATGAAGTTGCCAGCTGCGAGCACGTGAATAGACTGCTTCGGCAGTGTCTCG ACGTGTGTAGGGGAAAGTATCCTCCGTAACTAGTCGCGTCAC-3'

min

opt (9 nM)

5'-GGGACGAGCACGTGAATCGACTGCTTCGGCAGTGTCTCGACGTGTGTAGGGGGAAAGTATCCCCCGTCC C-3 '

Class V

ori (4000 nM)

min (K_d construct 15000 nM)

5'-**GGG**GGGCATTTTGGTAGGTCGGTCGCTGCTTCGGCAGTAAGGGGTAGGCATTGCTGGCCTAGGGTC CGTAAC-3'

min

opt (17 nM)

5'-**GG**GGGCATTTTGGTAGGTCGGTCGCTGCTTCGGCAGT**G**AGGGGTAGGCATTGCTGGCCTAGGGTCC **CC**-3'

10-10

ori (30 nM)

5'-GGAGGCGCCAACTGAATGAATTGCTCAACAGTAGCCAACACGAGTACTGCTTCGGCAGTGGAACCAAC GTAGTATGTTTAGCATTCCGTAACTAGTCGCGTCAC-3'

min (K_d construct 30 nM)

5'-GGTTGCTCAACAGTAGCCAACACGAGTACTGCTTCGGCAGTGGAACCAACGTAGTATGTTTAGCAT-3'

min

5'-GGAGCACGAACTCGGTATCCTTGCTCAACAGTAGCCAACACGAGTACTGCTTCGGCAGTGGAACCAAC GTAGTATGTTTAGCATGCGAACCATTCGGAACATCG-3'

opt (30 nM)

Class I

ori

5'-GGAGGCGCCAACTGAATGAAUUGCUUCGAGTCTTGAAGTGGTTGGGCTGCTTCGGCAGTGTGAAAATG AGGCTTTTAAGGGTCCGTAACTAGTCGCGTCAC-3'

ori (K_d construct 160 nM)

5'-GGGAGTCTTGAAGTGGTTGGGCTGCTTCGGCAGTGTGAAAATGAGGCTTTT-3'

min (K_d construct 76 nM)

5'-ggga**cg**aagtggttgggcgcttcggcgtgtgaaaa<mark>cgtctc-</mark>3'

min

5'-GGAGCACGAACTCGGTATCCGGGGA**CG**AAGTGGTTGGGCGCGTTCGGCGTGTGAAAACCGTCTCGCGAACCATCGGAACATCG-3'

opt (76 nM)

5'-GGGA**CG**AAGTGGTTGGGCGCGTTCGGCGTGTGAAAA<mark>CGTCTC-</mark>3'

10-59

ori (285 nM)

5'-GGAGGCGCCAACTGAATGAACAGGATGGTAAGTTCCCCAAGGCGGGTTGGAAGAGATATCATAGGAGCT TGTCGTTCTGGTCATCCTCCGTAACTAGTCGCGTCAC-3'

min (K_d construct 285 nM)

5'-GGATGGTAAGTTCCCAAGGCGGGTTGGAAGAGATATCATAGGAGCTTGTCGT-3'

min

5'-GGAGCACGAACTCGGTATCCATGGTAAGTTCCCAAGGCGGGTTGGAAGAGATATCATAGGAGCTTGTC GTGCGAACCATTCGGAACATCG-3'

opt (250 nM)

5'-GGGATGGTAAGTTCCCAAGGCGGGTTGGAAGAGATATCATAGGAGCTTGTCGTCCC-3'

10-24

ori (500 nM)

5'-GGAGGCGCCAACTGAATGAATGATCCTTTGGAGGGGGCATCTATATACTGCTTCGGCAGGGAACTCTAC TAAGCACCGATGTCACTCCGTAACTAGTCGCGTCAC-3'

min (K_d construct 300 nM)

5'-GCGGGGCATCTATATACTGCTTCGGCAGGGAACCTCTACTAAGCACCGATGTCCGC-3'

min

5'-GGGCATGTCCTTTTTCCTAATGCGGGGCATCTATATACTGCTTCGGCAGGGAACCTCTACTAAGCACCG ATGTCCGCGTAATAGCAGCATACTCGGA-3'

opt (300 nM)

5'- GCGGGCATCTATATACTGCTTCGGCAGGGAACCTCTAATAAGCACCGATGTCCGC-3'

9-12

ori (300 nM)

5'-GGAGGCGCCAACTGAATGAACAAAGTTGTTAACACCCACTCTAGCAGTTCAGGTAACCACGTAAGATA CGGTGCTAGAATGGGATCCGTAACTAGTCGCGTCAC-3'

min (K_d construct 300 nM)

5'-GGGACTCTAGCAGTTCAGGTAACCACGTAAGATACGGTGCTAGAAT-3'

min

5'-GGAGCACGAACTCGGTATCCACTCTAGCAGTTCAGGTAACCACGTAAGATACGGTGCTAGAATGCGAA CCATTCGGAACATCG-3'

opt (300 nM)

5'- GGGACTCTAGCAGTTCAGGTAACCACGTAAGATACGGTGCTAGAGT-3'

10-6

ori (500 nM)

5'-GGAGGCGCCAACTGAATGAAGATAGTCGGGACCACTTCCCTGAGTTCTGCTTCGGCAGATAACTCCAA GGAGAAAGGTCGTCATTCCGTAACTAGTCGCGTCAC-3'

min (K_d construct 500 nM)

5'-GCGAATGAAGATAGTCGGGACCACTTCCCTGAGTTCTGCTTCGGCAGATAACTCCAAGGAGAAAGGTC GTCATTCGC-3'

min

5'-GGAGCACGAACTCGGTATCCGCGAAGATAGTCGGGACCACTTCCCTGAGTTCTGCTTCGGCAGA TAACTCCAAGGAGAAAGGTCGTCATTCGCGCGCGAACCATTCGGAACATCG-3'

opt (300 nM)

5'-GCGAATGAAGATAGTCGGGACCACTTCGCTTCGGCGAGAAAGGTCGTCATTCGC-3'

Class II

ori

5'-GGAGGCGCCAACTGAATGAAGCTGCTAGCGCGGTAGAAAACCGAGCCGGAAGAGCACGTATACGCAGG GCTCAACTACATTCCGTAACTAGTCGCGTCAC-3'

ori (K_d construct 4600 nM)

5'-GGGACGGCTGCTAGCGCGGTAGAAAACCGAGCCGGAAGAGCACGTATACGCAGGGCTCAACTACA-3' min (K_d construct 2500 nM)

5'-**GGGACGCTG**AACGAGCCGGAAGAGCACGTATACGCAGGGCTCAACTA**G**-3'

min

5'-GGAGCCGACTAATGATTAAT**CTG**AACGAGCCGGAAGAGCACGTATACGCAGGGCTCAACTAGCGAGAG TCCGCTAACATCGC-3'

opt (400 nM)

5'-GGGAGCCAGAAGAGCACGTATACGCAAGGCTC-3'

Class IV

ori (2400 nM)

min (K_d construct 2400 nM)

5'-GGGTTGGAGCACAAAAAGGAGGAAGTGGTAAGACATCATGTGCTCCCGT-3'

min

opt (900 nM)

5'- GGGAGCACAAAAAGGAGGAAGTGGTAAGACATCATATGTGCTCC-3'

Class III

ori

5'-GGAGGCGCCAACTGAATGAACACACNCCTAAAAGGATACCCATGAACTGCTTCNGCAGTTTGCTAAA AACCACTCGTGGGTACNTTCCGTAACTAAGTCGCGTCAC-3'

ori (K_d construct 112000 nM)

min (K_d construct 112000 nM)

5'-GGGACACCCCTAAAAGGATACCCATGAACTGCTTCGCAGTTTGCTAAAAACCACTCGTGGGTACCTT-3'

min

5'-GGAGCCGACTAATGATTAATACACACCCTAAAAGGATACCCATGAACTGCTTCGCAGTTTGCTAAAAA CCACTCGTGGGTACCTTCCGTCCTACATCGGGCATTC-3'

opt (8000 nM)

5'- GGGATGATCGTCTTCGGACGTTGCTAAAAACCAGTCATCCC-3'