Supporting Information

Direct Selection of Fluorescence-Enhancing RNA Aptamers

Michael Gotrik, Gurpreet Sekhon, Saumya Saurabh, Margaret Nakamoto, Michael Eisenstein, H. Tom Soh

Supplementary Figures

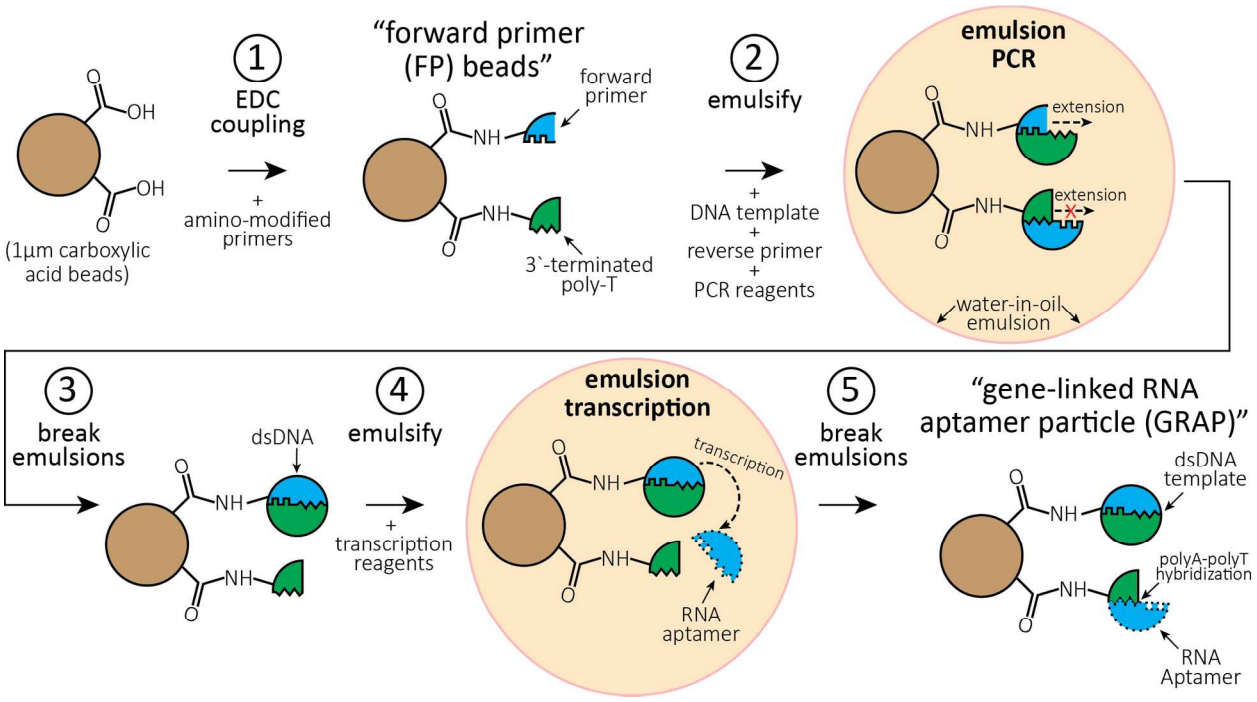


Figure S1: Assembly of Gene-linked RNA Aptamer Particles (GRAPs). 1μm paramagnetic, polystyrene carboxylic acid beads undergo EDC conjugation chemistry (**Step 1**) in the presence of a 1:9 ratio of 5'-amino modified forward primer (FP) and 5'-amino modified, 3'-terminated poly-T primer. The resulting FP beads are then emulsified in the presence of reverse primer and PCR reagents (**Step 2**) such that each water-in-oil emulsion contains, on average, zero or one DNA template. Emulsion PCR is performed, the emulsions are broken (Step 3), and excess reagents are washed away to yield particles that each express multiple identical copies of the starting DNA template. Because the poly-T is terminated with a 3'-inverted deoxythymidine, it is not extended during PCR. These beads then undergo a second emulsion reaction (Step 4) in the presence of T7 transcription mix. The DNA template, which contains a 5' T7 promoter and a 3' poly-A tail, is transcribed to yield a single-stranded RNA sequence that pairs with the expressed poly-T primer through complementary base hybridization. The emulsions are broken (Step 5) and excess reagents are washed away to yield gene-linked RNA aptamer particles (GRAPs).

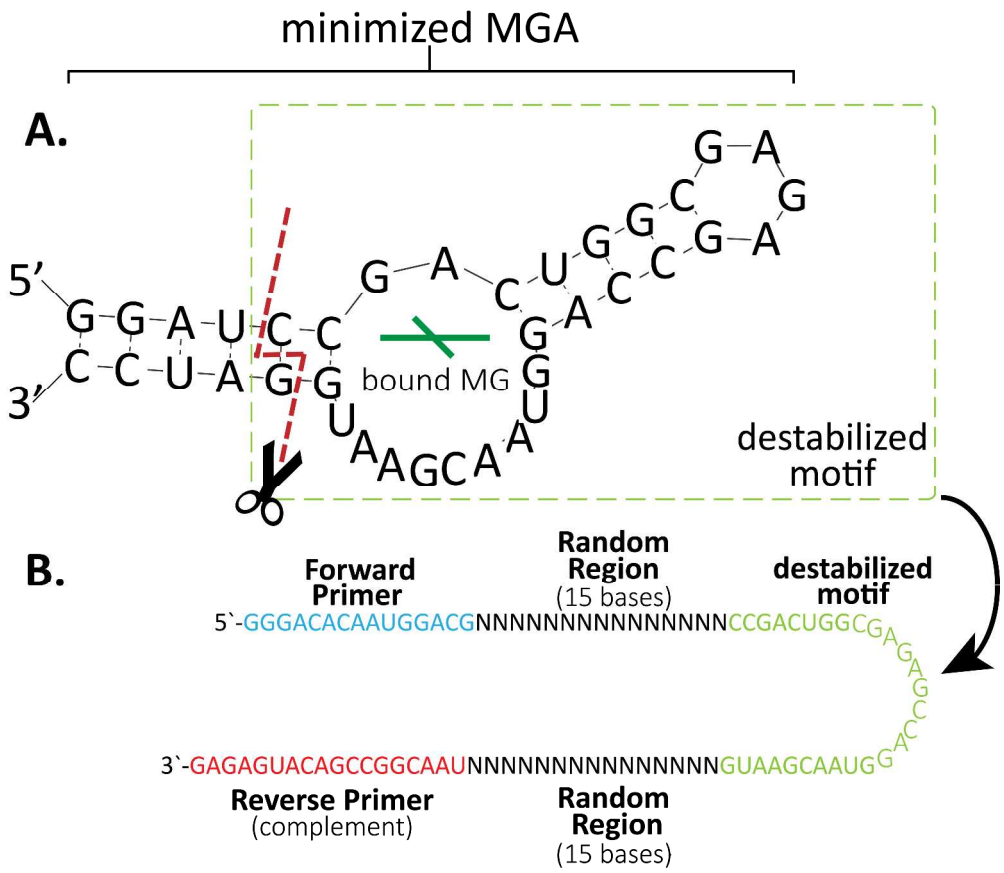
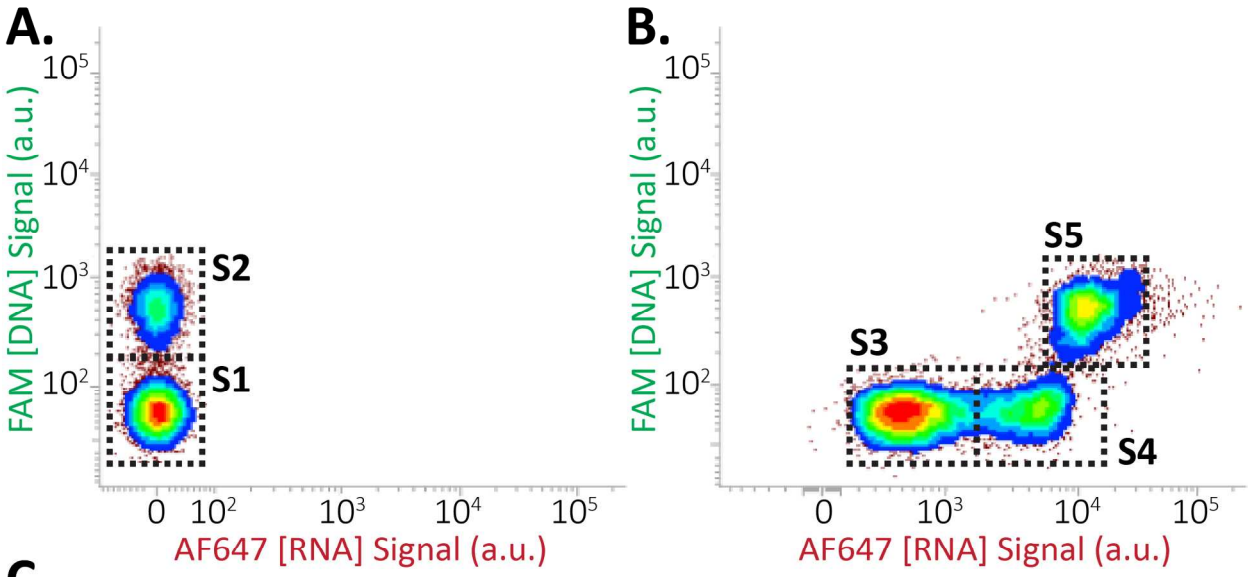


Figure S2: Design of structured RNA library. A) The predicted mfold structure of the minimized **MGA** motif showing the active malachite green binding pocket. The complementary pairing at the 5' and 3' ends can be removed so as to destabilize the binding pocket. B) In the structured library, this destabilized, minimized motif is surrounded on either side by 15 random bases, which are in turn flanked by constant primer regions.



C.

Gate	% Population	DNA Signal (mean±SD)	RNA Signal (mean±SD)
S1	72.6	44±25	5±21
S2	27.4	510±228	5±21
S3	50.8	42±24	608±274
S4	22.0	49±28	4,596±2,017
S5	25.8	488±210	16,129±10,203

Figure S3: Characterization of GRAPs. After emulsion PCR (ePCR) in the presence of FAM-labeled reverse primer (FAM-RP-poly-T, Table S1), beads were subjected to emulsion transcription to yield a GRAP library under conditions where only 15-30% of beads contain both DNA and RNA to ensure the beads are monoclonal. **A)** FAM vs. Alexa Fluor 647 signals of beads immediately following emulsion transcription. Gate S1 contains forward primer beads that remain free of covalently-linked DNA after emulsion PCR. Gate S2 contains beads coated with multiple identical copies of FAM-labeled double-stranded DNA. **B)** FAM vs. Alexa Fluor 647 signals of beads following emulsion transcription and incubation with an Alexa Fluor 647-labeled FP-complementary oligo (647-FP-complement, Table S1). Because FP beads are coated with ~10-fold more poly-T than FP, all beads should fluoresce with the addition of 647-FP-complement. However, RNA-coated beads will capture greater amounts of 647-FP-complement due to the transcribed region of the FP, which remains single-stranded upon RNA capture. Gate S3 contains beads that display no DNA nor RNA. Gate S4 contains beads that display no DNA but have an RNA signal, possibly due to non-specifically-bound, carryover DNA becoming transcribed. Gate S5 contains beads displaying both RNA and parent DNA. During FACS sorting, only beads displaying a high DNA signal are considered to ensure that DNA-free, RNA-coated beads (*e.g.*, Gate S4) are not collected. **C)** Mean fluorescent signals (arbitrary units) of beads in each gate. DNA and RNA signals are not comparable to one another due to the different fluorophores, emission channels, and laser settings being used.

Supplementary Tables

Table S1: Oligos used during GRAP display. The following sequences were ordered from IDT with the modifications listed (using standard IDT abbreviations). Standard desalting was used except for oligos with modifications, for which HPLC purification was performed. All sequences were ordered as single-stranded DNA except for MGA, which was ordered as double-stranded DNA for use in transcription. /56-FAM/ represents 5' 6-FAM, /5AmMC6/ represents a 5' amino modification with a six-carbon linker, /iSp18/ represents an 18-atom hexa-ethyleneglycol spacer, /3InvdT/ represents a 3' inverted deoxythymidine, and /5Alex647N/ represents 5' Alexa Fluor 647.

Oligo Name	DNA Sequence (5′ → 3′)
Structured Library	TAATACGACT CACTATAGGG ACACAATGGA CGNNNNNNNN NNNNNNCCG ACTGGCGAGA GCCAGGTAAC GAATGNNNNN NNNNNNNNNN TAACGGCCGA CATGAGAG
Naïve Library	TAATACGACT CACTATAGGG ACACAATGGA CGNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNTAACGGCC GACATGAGAG
FP	TAATACGACT CACTATAGGG ACACAATGGA CG
RP	CTCTCATGTC GGCCGTTA
RP-poly-T	TTTTTTTTTT TTTTTTTTTT TTTTTCTCTC ATGTCGGCCG TTA
FAM-RP-poly-T	/56-FAM/- TTTTTTTTTT TTTTTTTTTT TTTTTCTCTC ATGTCGGCCG TTA
AmM-FP	/5AmMC6//iSp18//iSp18//iSp18//iSp18/- TAATACGACT CACTATAGGG ACACAATGGA CG
AmM-pT	/5AmMC6//iSp18//iSp18//iSp18//iSp18/- TTTTTTTTTT TTTTTTTTTT TTTT/3InvdT/
647-FP-complement	/5Alex647N/- CGTCCATTGT GTCCCTATAG TGAGTCGTAT TA
FAM-FP-complement	/56-FAM/- CGTCCATTGT GTCCCTATAG TGAGTCGTAT TA
FAM-poly-A	/56-FAM/- AAAAAAAAAA AAAAAAAAAA AAAAA
MGA	TAATACGACT CACTATAGGG AACACTATCC GACTGGCACC CCCCTGCCAG GTAACGAATG AAGTGCTTTT CTCGATCTCG TGACCCGCGC ACTAGTCGCG AAGGTGTATG TCCTTGGTCA TTAGGATCC
minimized MGA	TAATACGACT CACTATAGGG ATCCGACTGG CGAGAGCCAG GTAACGAATG GATCC

Table S2: Sequences of aptamers selected from the structured library. Sequences **MG-D1** and **MG-D2** are identical except for the underlined base. Sequences **MG-D7**, **MG-D8**, and **MG-D9** are identical except for the underlined base. **MG-D6** has two deletions from the starting library, which are represented by blanks to maintain alignment. The conserved, destabilized minimal **MGA** region is bolded, and random regions are separated from conserved regions by hyphens.

Name (MG-)	DNA Sequence (5` → 3`)
D1	GGGACACAATGGACG-ACACGCCATTGTCCT-**CCGACTGGCGAGAGCCAGGTAACGAATG**-AGGTTTCTAAGTGTA-TAACGGCCGACATGAGAG
D2	GGGACACAATGGACG-ACACGCCATTGTCCT-**CCGACTGGTGAGAGCCAGGTAACGAATG**-AGGTTTCTAAGTGTA-TAACGGCCGACATGAGAG
D3	GGGACACAATGGACG-CAGATAATCGGATTC-**CCGACTGGCGAGAGCCAGGTAACGAATG**-GGCCGGTTTCTGTGT-TAACGGCCGACATGAGAG
D4	GGGACACAATGGACG-CCTTTAAGGGGCTCG-**CCGACTGGCGAGAGCCAGGTAACGAATG**-CGAGGCTCGGAAAAG-TAACGGCCGACATGAGAG
D5	GGGACACAATGGACG-CGCCCAGGATCTCTG-**CCGACTGGCGAGAGCCAGGTAACGAATG**-CAAGACGTCCCCTGT-TAACGGCCGACATGAGAG
D6	GGGACACAATGGACG- CGCCTCCAATTTGG-**CCGACTGGCGA AGCCAGGTAACGAATG**-CCAATAGCCCCTCCG-TAACGGCCGACATGAGAG
D7	GGGACACAATGGACG-CGTACCATAAGGCCT-**CCGACTGGCGAGAGCCAGGTAACGAATG**-AGGGGATTAACACTC-TAACGGCCGACATGAGAG
D8	GGGACACAATGGACG-CGTACCATAAGGCCT-**CCGACTGGCGAGCGCCAGGTAACGAATG**-AGGGGATTAACACTC-TAACGGCCGACATGAGAG
D9	GGGACACAATGGACG-CGTACCATAAGGCCT-**CCGACTGGCGAGGGCCAGGTAACGAATG**-AGGGGATTAACACTC-TAACGGCCGACATGAGAG
D10	GGGACACAATGGACG-GGTCTTAATGGTAAG-**CCGACTGGCGAGAGCCAGGTAACGAATG**-GCACGCGACAAACCA-TAACGGCCGACATGAGAG

Table S3: Characterization of aptamers from the structured library. For each sequence, we determined and ranked the reads-per-million (RPM) rank and fitness in the Round 2 and Round 3 pools, where fitness is defined as the ratio of the current-round RPM over the previous-round RPM. If fitness is not determined (n.d.), this is because we detected <20 copies of the sequence in the raw sequencing data for the previous round. K_d is reported with standard deviation measured from three separate samples. Fluorescence enhancement (FE) represents the fluorescent signal fold-increase above background. Data for **MGA** included as a comparator.

Name (MG-)	Round 3 RPM	Round 3 RPM Rank	Round 2 Fitness	Round 3 Fitness	Round 3 Fitness Rank	K_d (nM)	FE
D1	26,679	3	156.6	5.4	90	63 ± 6	1638
D2	1,035	72	n.d.	53.9	2	59 ± 5	1445
D3	15,568	4	347.3	1.7	278	44 ± 19	3660
D4	817	86	n.d.	42.5	3	60 ± 17	1338
D5	27,590	2	409.7	1.6	282	42 ± 9	1262
D6	518	180	n.d.	37.2	4	55 ± 16	1224
D7	420,106	1	212.4	5.5	86	20 ± 7	2815
D8	1,335	51	n.d.	59.3	1	21 ± 7	2392
D9	10,981	5	118.5	18.2	16	24 ± 6	1955
D10	529	168	n.d.	33.3	5	n.d.	n.d.
MGA	-	-	-	-	-	103 ± 31	2360

Table S4: Sequences of aptamers selected from the naïve library. 22 sequences were identified through copy number and fitness analysis. For each sequence, we compared brightness and affinity with and without the 3' poly-A tail that coupled each RNA aptamer to the bead surface. For sequences unaffected (o) or negatively affected (-) by its presence, we removed the tail sequence before conducting any further functional characterization. R3-7 contained a deletion in the forward primer region which is represented as a blank to maintain alignment. The starting library's random region is separated from the constant regions by hyphens.

Name	DNA Sequence (5` → 3`)	Poly-A Tail Effect
R3-1	GGGACACAATGGACG-TGCACGGAGGCGCCTCTGGTCCTTAGTTTAATCCATAACG-TAACGGCCGACATGAGAG	+
R3-2	GGGACACAATGGACG-TACAATCTCAGGTGCACCTTTAGACCACTGTGTCTCTACA-TAACGGCCGACATGAGAG	o
R3-3	GGGACACAATGGACG-CTAGCTTCGAGTAACGAAACGCGTGAGGCTGCCCCGCTCC-TAACGGCCGACATGAGAG	o
R3-4	GGGACACAATGGACG-AGTCTAGAGCGGGTACCTGTGCTGCAAATCTCCCGTCACT-TAACGGCCGACATGAGAG	+
R3-5	GGGACACAATGGACG-GTCCCAAGGAGGTGGGTGGGTGGTGTCGACTCCTTTTCTT-TAACGGCCGACATGAGAG	o
R3-6	GGGACACAATGGACG-GTGCCACCAGACTCACGCACTAGATCCGGCCGAGTAACGA-TAACGGCCGACATGAGAG	–
R3-7	GGGACACA TGGACG-GTCCCAAGGAGGTGGGTGGGTGGTGTCGACTCCTTTTCTT-TAACGGCCGACATGAGAG	o
R3-8	GGGACACAATGGACG-AGAAATTGTGTAAGACCTTATTGAATGAGGCGCAACTCGC-TAACGGCCGACATGAGAG	–
R3-9	GGGACACAATGGACG-TACAATCTTAGGTGCACCTTTAGACCACTGTGTCTCTACA-TAACGGCCGACATGAGAG	–
R3-10	GGGACACAATGGACG-CTAGCTCCGAGTAACGAAACGCGTGAGGCTGCCCCGCTCC-TAACGGCCGACATGAGAG	–
R3-B-1	GGGACACAATGGACG-TTACAATACAAGTCTTCGAAGACTGAGTTTTCCATCTCCA-TAACGGCCGACATGAGAG	+
R3-B-2	GGGACACAATGGACG-TTACAATACAAGTCTTCGAAGACTGAGTTTCCCATCTCCA-TAACGGCCGACATGAGAG	+
R3-B-3	GGGACACAATGGACG-TTCATCAACATAGGAGTGGGAGGGGTAGCTAGGGCTCGTG-TAACGGCCGACATGAGAG	–
R3-B-4	GGGACACAATGGACG-AGTCTAGAGCGGGTACCTGTGCTGCAAATCTCCCGTCATT-TAACGGCCGACATGAGAG	–
R3-R-1	GGGACACAATGGACG-ATCGATTCAATGCCGCCAAGGGCTAGCCCGATCCATAACG-TAACGGCCGACATGAGAG	+
R3-R-2	GGGACACAATGGACG-CGCGTGCAACAGGGCCCTGATCCAGCACTCACTTATAACG-TAACGGCCGACATGAGAG	+
R3-R-3	GGGACACAATGGACG-TACAAGAGATTCAAGCGCACGGTCTTCCACACCCAACTCG-TAACGGCCGACATGAGAG	+
R3-R-4	GGGACACAATGGACG-TGCACGGAGGGGCCTCTGGTCCTTAGTTTAATCCATAACG-TAACGGCCGACATGAGAG	–
R3-R-5	GGGACACAATGGACG-ATCGATTCAATGCCGCCAAGGGCTGGCCCGATCCATAACG-TAACGGCCGACATGAGAG	o
R3-R-6	GGGACACAATGGACG-CGCGCGGGTCTCGAGACCCTGAATCATTTCTCGAAAAACG-TAACGGCCGACATGAGAG	+
R3-R-7	GGGACACAATGGACG-CGCGTCACCAGGGACACTCGGATCCACTCACATCATAACG-TAACGGCCGACATGAGAG	o
R3-R-8	GGGACACAATGGACG-CGCGTGCAACAGGGCCTTGATCCAGCACTCACTTATAACG-TAACGGCCGACATGAGAG	+

Table S5: Characterization of aptamers from the naïve library. RPM, fitness and K_d data are presented for our 22 selected sequences as described in Table S3. Sequences whose K_d exceeded the limit of detection (~2 μM) were not determined (n.d.). Fluorescence enhancement (FE) was determined only for sequences with a measured K_d and represents the fluorescent signal fold-increase above background. Because these enhancements were measured under MG titration, FE values represent a minimum value, except for **MGA**. Data for **MGA** included as a comparator.

	Name	RPM	RPM Rank	Fitness	Fitness Rank	K_d (nM)	FE	λ_{exc}^{max} (nm)	λ_{em}^{max} (nm)
Round 3 Pool	**R3-1**	157,406	1	6.34	333	2017 ± 167	111	646	675
	R3-2	111,565	2	13.01	72	42 ± 12	19	641	669
	R3-3	110,502	3	6.04	357	n.d.	n.d.	634	652
	R3-4	96,745	4	8.43	216	n.d.	n.d.	631	649
	R3-5	95,075	5	10.04	141	92 ± 22	50	635	658
	R3-6	2,295	16	25.9	1	111 ± 11	259	633	651
	R3-7	165	242	22.08	2	125 ± 34	42	635	658
	R3-8	131	303	21.29	3	29 ± 8	18	628	651
	R3-9	195	199	20.92	4	38 ± 10	38	641	669
	R3-10	3,249	13	17.19	16	220 ± 39	317	634	652
Round 3-B Pool	**R3-4**	707,584	1	7.12	69	n.d.	n.d.	631	649
	R3-3	13,082	2	0.12	235	n.d.	n.d.	634	652
	R3-2	11,302	3	0.10	237	42 ± 12	19	641	669
	R3-1	8,281	4	0.05	247	2017 ± 167	111	646	675
	R3-5	7,517	5	0.08	243	92 ± 22	50	635	658
	R3-B-1	1,102	45	42.67	1	320 ± 26	620	631	650
	R3-B-2	422	95	17.20	3	144 ± 26	495	634	652
	R3-B-3	1,213	37	11.45	4	2094 ± 347	219	631	649
	R3-B-4	1,493	33	8.67	24	579 ± 75	4273	637	654
Round 3-R Pool	**R3-1**	571,285	1	3.57	165	2017 ± 167	111	646	675
	R3-2	95,829	2	0.84	350	42 ± 12	19	641	669
	R3-R-1	28,654	3	5.03	59	1301 ± 345	275	646	675
	R3-R-2	18,279	4	2.49	270	1549 ± 441	238	646	675
	R3-R-3	8,753	5	0.59	383	n.d.	n.d.	636	656
	R3-R-4	708	68	14.92	1	1045 ± 124	189	646	675
	R3-R-5	249	142	13.39	2	1488 ± 370	79	646	675
	R3-R-6	82	307	11.25	4	1827 ± 175	148	646	675
	R3-R-7	399	104	10.1	6	1561 ± 221	198	646	675
	R3-R-8	76	316	10.01	7	1590 ± 226	237	646	675
	MGA	-	-	-	-	103 ± 31	2360	633	652

Table S6: Motif minimization of red-shifted aptamers. A conserved motif was identified in sequences exhibiting the unique 675-nm emission maximum. Pair-wise truncation analysis revealed a short, 15-nt motif that is responsible for this fluorescence profile. The following RNA sequences were ordered from IDT and prepared in triplicate at concentrations of 100 nM RNA and 1 μM MG in PBSMT, pH 6.1. Fluorescence intensity was measured at 675±20 nm using a Tecan M1000 plate reader with excitation at 630±20 nm. Intensity was normalized to the signal from the 21-bp sequence.

Sequence ($5' \rightarrow 3'$)	Length (bp)	Normalized Intensity
CAUAACGUAACGGCCGACAUG	21	1.00
AUAACGUAACGGCCGACAU	19	0.70
UAACGUAACGGCCGACA	17	0.92
AACGUAACGGCCGAC	15	0.80
ACGUAACGGCCGA	13	0.05
CGUAACGGCCG	11	0.00

Table S7: Summary of FACS procedures. The number of particles sorted and collected from the respective emission channels during each sort with the structured library and naïve library. Each set of collected particles was combined prior to PCR amplification, such that only a single pool was prepared and sorted in subsequent rounds of GRAP display.

	Pool Sorted	Resulting Pool	MG Concentration	Particles Sorted	670±30nm Particles	730±45nm Particles	780±60nm Particles
Structured Library	Starting	Round 1	50 nM	$5.3 * 10^7$	8,428	-	-
	Round 1	Round 2	1 nM	$8 * 10^6$	1,198	-	-
	Round 2	Round 3	400 pM	$1.0 * 10^7$	1,473	-	-
Naïve Library	Starting	Round 1	10 μM	$6.0 * 10^7$	6,646	6,713	4,946
	Round 1	Round 2	1 μM	$1.7 * 10^7$	7,532	5,536	5,210
	Round 2	Round 3	100 nM	$8 * 10^6$	21,332	18,795	18,640
	Round 3	Round 3-B	1 μM	$3 * 10^5$	20,000	0	0
	Round 3	Round 3-R	1 μM	$3 * 10^5$	0	0	20,000