Supplementary material for:

Laboratory Evolution of Artificially Expanded DNA gives Redesignable Aptamers that Target the Toxic Form of Anthrax Protective Antigen

Elisa Biondi[1,*], Joshua D. Lane[1], Debasis Das[2], Saurja Dasgupta[3], Joseph A. Piccirilli[3,4], Shuichi Hoshika[1], Kevin M. Bradley[1], Bryan A. Krantz[2] and Steven A. Benner[1,5,*]

[1] Foundation for Applied Molecular Evolution, Alachua, FL, 32615, USA
[2] School of Dentistry, The University of Maryland, Baltimore, MD, 21201, USA
[3] Department of Chemistry, University of Chicago, Chicago, IL, 60637, USA
[4] Department of Biochemistry and Molecular Biology, University of Chicago, Chicago, IL, 60637, USA
[5] Firebird Biomolecular Sciences LLC, Alachua, FL, 32615, USA

EXPANDED METHODS

Oligonucleotide synthesis

Oligonucleotides and the library containing AEGIS nucleotides were prepared as previously reported (9,10) by solid phase phosphoramidite synthesis using standard nucleoside phosphoroamidites (Bz-dA, Ac-dC, dmf-dG, and dT, from Glen Research, Sterling, VA) and AEGIS phosphoramidites (Firebird Biomolecular Sciences LLC, Alachua, FL). CPG having standard nucleosides was from Glen Research. Solid phase synthesis was performed using an ABI 394 DNA Synthesizer following standard phosphoramidite chemistry procedures. The randomized sites in the library were prepared by coupling with a 1:1:1:1:1:1 mixture of the six (GACT**ZP**) nucleotide phosphoramidites. After synthesis was complete, the solid phases carrying the protected synthetic oligonucleotides were treated with concentrated NH_4OH (55 °C, 16 h). After evaporation of the NH_4OH, the oligonucleotides were purified by ion-exchange HPLC, and then desalted using Sep-Pac® Plus C18 cartridges (Waters).

In vitro selection scheme

A synthetic library of GACT**ZP** oligonucleotides (59 nucleotides in length) containing 25 randomized positions flanked by two primer binding sites (17 nt each) was used to start the AEGIS laboratory *in vitro* evolution (AEGIS-LIVE) experiment. The library was 59 nucleotides in length, with sequence: 5'-AGAGAGCGTCGTGTGGA-N25-TGAGGAGGTGCGCAAGT-3'.

The initial step was a negative selection on the magnetic beads lacking PA. Protective antigen (PA63) was presented immobilized on magnetic beads (Dynabeads M-270 Carboxylic Acid, Invitrogen), binding oligonucleotides were recovered magnetically, and AEGIS-PCR (9,10) with a single biotinylated primer was performed directly on survivors bound to the bead-coupled PA63. Following amplification, single-stranded DNA was recovered with streptavidin immobilized on magnetic beads (Dynabeads M-270 Streptavidin, Invitrogen), and used in the next round of selection.

In detail, PCR amplifications were performed at 5/10/15/20 and 25 cycles in the following conditions: 0.6 µM DNA, 5 µM Fwd Primer (PA-F-17, 5'-AGAGAGCGTCGTGTGGA-3'), 5 µM Rev Primer (PA-R-17Bio, 5'-Biotin-ACTTGCGCACCTCCTCA-3'), 0.1 mM dZ/dA/dG/dTTP, 0.2 mM dCTP, 0.6 mM dPTP, 5 U/µL Takara HF-Taq Pol, 20 mM Tris-HCl pH 8.0, 10 mM $(NH_4)SO_4$, 10 mM KCl, 2 mM $MgSO_4$, 0.1% Triton-X-100. PCR cycles included 60" @ 94°C, [20" @ 94°C, 30" @ 55°C, 5' @ 72°C] X 5/10/15/20 or 25, 10' @ 72°C, ∞ @ 4°C.

For selection negative cycle I and selection cycle I, 5'-labeled and PAGE-purified ssDNA library (1 nmole) was used. Carboxylated magnetic beads (Dynabeads M-270 carboxylic acid, Invitrogen) were coated with PA63 (List Biologicals) following the manufacturer protocol, and non-reacted sites were blocked with ethanolamine. A mock coating process without PA63 was also performed on the beads to support negative selection cycles. ssDNA was denatured/renatured in

1X PBS at 80°C for 5 min followed by slow cooling to RT (0.1°C/second). Folded ssDNA was incubated with uncoated beads for 30 min at RT (negative cycle I). Beads were then collected on a magnet and the supernatant transferred to PA-coated beads. Uncoated beads were washed once with 1X PBS and this supernatant added to the binding reaction. Binding reactions were carried over for 30 min at RT. Subsequently, beads were collected and washed extensively with 1X PBS, until no radioactive signal could be detected in the washes. All fractions were then read with Cherenkov counting at a scintillation counter (Perkin Elmer TriCarb 2900TR). Negative cycle I and cycle I of the selection were performed using 1 nmole library and 3 nmoles PA63; all other cycles were performed with 150 pmoles ssDNA and 800 pmoles PA63. *In vitro* selection included fourteen cycles, at which point the library was prepared for conversion and deep sequencing.

AEGIS-DNA to standard DNA transliteration for high-throughput sequencing

Survivors of Cycle XIV were divided in two aliquots and PCR amplified under conditions that transliterated the **Z:P** pairs to C:G or T:A pairs in approximately equal amount (9,12). Tags and barcodes were then added by 12 cycles of PCR with tagged primers. The amplified products were purified by native PAGE, recovered by gel-extraction, and the collection of amplicons was submitted for Ion Torrent deep sequencing, according to the sequencing center guidelines.

Translation from **Z:P** to C:G used the following conditions: 0.4 µM each primer (AnthraxBcode2: 5'-CCATCTCTATCCCTGCGTGTCTCCGACTCAGGATGATTGCCAGAGAGCGTCGTGTGGA-3', Anthrax_trP_Rev: 5'-CCTCTCTATGGGCAGTCGGTGATACTTGCGCACCTCCTCA-3'), 0.01 mM dZTP, 0.4 mM each dC/dGTP, 0.04 mM each dA/dTTP, 20mM Tris-HCl pH 8.8, 10mM $(NH_4)_2SO_4$, 10 mM KCl, 2 mM $MgSO_4$, 0.1% Triton® X-100, 0.1 U/µL JumpStart Taq Polymerase (Sigma-Aldrich).

Conversion from **Z:P** to T:A had the following conditions: 0.4 µM each primer (AnthraxBcode3: 5'-CCATCTCTATCCCTGCGTGTCTCCGACTCAGAAGAGGATTCAGAGAGCGTCGTGTGGA-3', Anthrax_trP_Rev), 0.4 mM dPTP, 0.04 mM each dC/dGTP, 0.4 mM each dA/dTTP, 10 mM Tris-HCl, pH 8.3, 50 mM KCl, 1.5 mM $MgCl_2$, 0.001% (w/v) gelatin, 0.1 U/µL JumpStart Taq Polymerase.

Twelve cycles of PCR amplification were then carried out for each sequencing conditions.

High-throughput sequencing

Following conversion, the barcoded samples were combined and submitted for Ion Torrent deep sequencing. Ion torrent analysis was performed as previously reported (9,10). Reads missing exactly matched barcodes and/or forward and reverse priming sequences were discarded. To minimize miscalling, sequences present at 20 or fewer copies in the library were removed from the subsequent analysis. The remaining reads (~2 x 10^6 in total) were clustered using a custom algorithm. This algorithm ignored differing barcodes, concentrating only on the

variable 25 nucleotide region. Due to the short sequence size and unknown number of converted sites, sequence clusters were created in a step-wise process by grouping those with a single base change between sequence reads, starting with the most common sequence read and proceeding towards the least common. This was an iterative algorithm, meaning that sequences with a single-base difference from any other sequence in the cluster were continued to be added to the cluster until no further sequences could be added. Clustered sequences were then separated by barcode, and variable sites were compared between each barcode. The clustered sequences obtained under the first conversion conditions (**Z** to C and **P** to G) serve as reference for the clustered sequences obtained under the second.

Determination of the affinity constant with bead-based assay

After being coupled to PA63 (~1.6 nmoles), the beads (200 µL, as during selection) were divided into aliquots (20 µL each) and each fraction was incubated with increasing concentrations of pre-folded cold PA1, spiked with 1 µL (> 13 pM) of $5^{,32}$P-PA1. Concentrations of aptamer [A] were: >13 pM, 10 nM, 20 nM, 50 nM, 100 nM, 200 nM, 0.5 µM, 1 µM, 2 µM. Binding mixtures were tumbled (RT, 45 min), and the beads were recovered with a magnet. The supernatant was collected for cpm reading. Beads were washed 2 times with 100 µL 1X binding buffer (1X PBS), and all the fractions were separately read with liquid scintillation. Radioactivity readings of the samples aptamers+PA-beads were analyzed as follows:

First, the following assumptions were made:

- at 2 µM aptamer concentration, each PA molecule on the beads has 1 aptamer molecule bound

- adding the labeled aptamer to the measured amounts of non-labeled aptamers does not materially change the concentration of the aptamers.

- The beads do not materially quench the scintillation counting signal.

- labeled and unlabeled aptamer bind the same way

Under these assumption, taken the cpm at 2 µM as a starting point "cpm2", cpm at 1 µM should be cpm2 X 2, cpm at 0.5 µM should be cpm1 X 2, cpm at 200 nM should be cpm 0.5 X 2.5, and so on.

Dividing the actual cpm read for each sample by the expected cpm for that sample, we obtain the Fraction of PA molecules on the beads that actually have one aptamer bound (C/P_0, or $C/[P+C]$). The reciprocal of this number is P_0/C. Plotting P_0/C versus the reciprocal of aptamer concentration ($1/[A]$) will give a line the slope of which represents the K_{diss} value (K in this derivatization). The described stepwise calculation is as follows:

[A][P]/[C] = K,

P_0 = P + C

P = P_0 - C

[A][P_0 - C]/[C] = K

[A][P$_0$]/[C] – [A] [C]/[C] = K

[A][P$_0$]/[C] = K + [A] [C]/[C]

divide through by [A]

[P$_0$]/[C] = K/[A] + 1

Bead-based binding assay on CMG2-PA complex

PA63 (1 nmole) was incubated with biotinylated CMG2 (1 nmole) in buffer (20 mM Tris pH 7.5, 1 mM MgCl$_2$) at RT to give complex (1 nmole). The complex was adsorbed to magnetic beads bearing streptavidin (300 µL, Invitrogen). This corresponds to 600 pmoles of bound complex (the manufacturer's described binding capacity is ~ 2 pmoles/µL). After adsorption and washing according to manufacturer instructions, aliquots of the beads were placed in 10 tubes.

Unlabeled aptamer (pre-folded in 1X PBS) spiked with 0.1 nM 5'-labeled aptamer (also pre-folded) was added at increasing concentrations to the beads, and the mixtures were incubated at RT (45 min). The supernatants were then collected, the beads were washed twice (1X PBS), and all fractions were dissolved in scintillation cocktail for cpm reading. For each aptamer concentration, the percent bound of the labeled species was calculated by counting cpm of bound molecules (times 100) and dividing by total cpm for that sample (bound + supernatant + washes). Percent binding was then transformed in bound aptamer concentration assuming 0.1 nM (the concentration of labeled species) = 100%. Assuming that 5'-labeled species bind exactly as cold aptamer molecules, and can thus be regarded as representative of all populations, the concentration of bound labeled molecules was multiplied times its dilution to obtain the real nM concentration of bound aptamers. For each aptamer, data were plotted in a non-linear regression fitted to an equation of the type $a * \dfrac{x}{b+x}$, where a is maximum binding and **b** is the K$_{diss}$ of the aptamer.

SUPPLEMENTARY TABLES

Supp Table 1. sequences of the six most represented aptamers and their frequency obtained by high-throughput sequencing of survivors of cycle XIV

Aptamer #	Standard Sequence	AEGIS Sequence	# of Sequences
Aptamer #1	CCATGGAGTGAGTGAGTAGGTGTCC	CCATGGPGTGAGTGPGTAGGTGTCC	595149
Aptamer #2	CAGGTGTGAGGTCGGAGGTGACGCG	CAGGTGTGAGGTCGGAGGTGACGCG	9083
Aptamer #3	CCGGGGGGAGTGGGGAGTTGACGGG	CCPGGGGGPGTGGGGPGTTGACGGG	9490
Aptamer #4	CCATGGGAGTAGTGAGTAAGTGTCC	CCATGGGPGTAGTGPGTAPGTGTCC	636
Aptamer #5	CCATGGGTGAGTGAGTAGGTGTCCG	CCATGGGTGAGTGPGTAGGTGTCCG	151
Aptamer #6	CCATGGGTGAGTGAGTGAAGTGTCC	CCATGGGTGAGTGPGTPAAGTGTCC	221

Supp Table 2. Sequence analysis of G-quadruplex probability for aptamer PA1T4 performed with the program QGRS Mapper (**Q**uadruplex forming **G**-**R**ich **S**equences). P's were substituted with G'sfor the analysis. G-score indicates the probability for that specific pattern. In bold is the most likely pattern, with highest G-score. The yellow star indicates the only putative G-quadruplex series that does not involve P's. This is also the only GQ option for the sequence obtained when P's are substituted with A's.

Position	Length	QGRS	G-Score
6	18	GGACCATGGPGTGAGTGP	16
6	19	GGACCATGGPGTGAGTGPG	15
6	23	GGACCATGGPGTGAGTGPGTAGG	11
6	23	GGACCATGGPGTGAGTGPGTAGG	17
6	23	GGACCATGGPGTGAGTGPGTAGG	15
6	23	**GGACCATGGPGTGAGTGPGTAGG**	18
6	23	GGACCATGGPGTGAGTGPGTAGG	16
6	23	GGACCATGGPGTGAGTGPGTAGG	17
6	23	GGACCATGGPGTGAGTGPGTAGG	16
13	16	GGPGTGAGTGPGTAGG	16
13	16	GGPGTGAGTGPGTAGG	15
13	26	GGPGTGAGTGPGTAGGTGTCCTGAGG	8
13	26	GGPGTGAGTGPGTAGGTGTCCTGAGG	9
13	26	GGPGTGAGTGPGTAGGTGTCCTGAGG	11
13	26	GGPGTGAGTGPGTAGGTGTCCTGAGG	16
13	26	GGPGTGAGTGPGTAGGTGTCCTGAGG	15
13	29	GGPGTGAGTGPGTAGGTGTCCTGAGGAGG	5
13	29	GGPGTGAGTGPGTAGGTGTCCTGAGGAGG	6
13	29	GGPGTGAGTGPGTAGGTGTCCTGAGGAGG	10
13	29	GGPGTGAGTGPGTAGGTGTCCTGAGGAGG	1
13	29	GGPGTGAGTGPGTAGGTGTCCTGAGGAGG	13
13	29	GGPGTGAGTGPGTAGGTGTCCTGAGGAGG	9
13	29	GGPGTGAGTGPGTAGGTGTCCTGAGGAGG	12
13	29	GGPGTGAGTGPGTAGGTGTCCTGAGGAGG	10
13	29	GGPGTGAGTGPGTAGGTGTCCTGAGGAGG	10 ⭐
14	25	GPGTGAGTGPGTAGGTGTCCTGAGG	16
14	25	GPGTGAGTGPGTAGGTGTCCTGAGG	15
14	28	GPGTGAGTGPGTAGGTGTCCTGAGGAGG	13
14	28	GPGTGAGTGPGTAGGTGTCCTGAGGAGG	9
14	28	GPGTGAGTGPGTAGGTGTCCTGAGGAGG	12
14	28	GPGTGAGTGPGTAGGTGTCCTGAGGAGG	10

SUPPLEMENTARY FIGURES

Supp. Figure 1: Protective Antigen AEGIS-LIVE progress. Red asterisks indicate negative selection cycles.
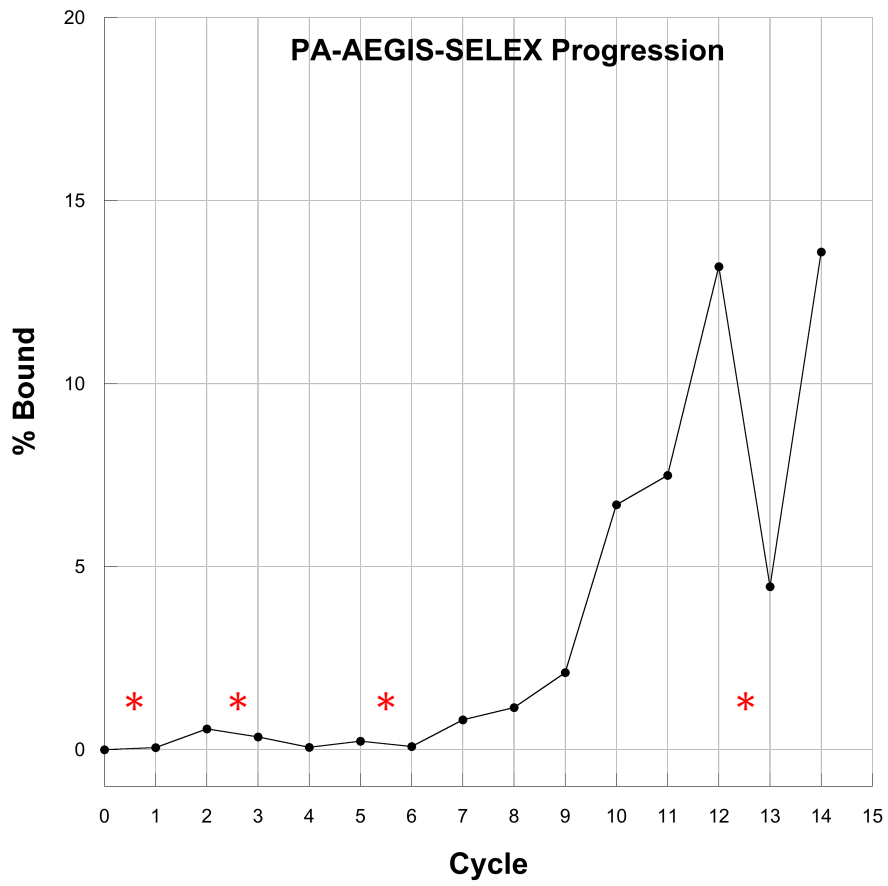
Supp. Figure 2: Color-coded representation of PA-Apt1 truncations on KineFold 2D prediction. 3' end of each molecule is represented by a dot. PA1T4, in yellow, is the truncation that retained best binding, and was studied in more details.

PA1Short
PA1T1
PA1T2

PA1T3
PA1T4
PA1T5
PA1T6
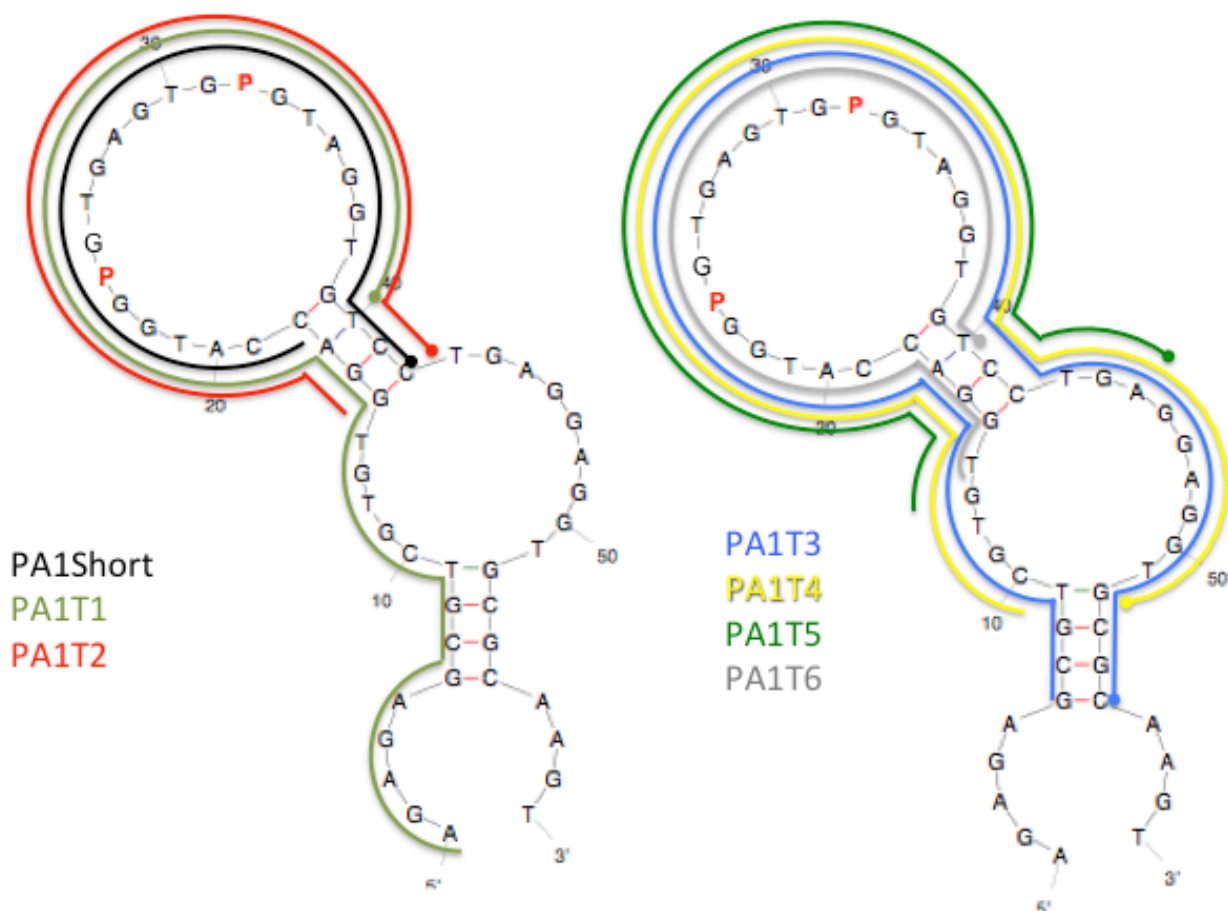
Supp. Figure 3: Results of bead-based binding assay for PA-Apt1. The reciprocal of the fraction of PA molecules on the beads that actually have one aptamer bound (P_0/C) is plotted against the inverse of aptamer concentration (1/[A]). The slope of the plot reflects aptamer's kd.

PA-Apt1 on PA63

y = 1.3225 + 308.22x   R= 0.99729

$[P_0]$ / [C]

1 / [A]