Supplementary information for:

ARCaDia: Single-round selection of a DNA-type targeted covalent binder possessing a latent warhead

Masumi Taki,* Masayasu Kuwahara, Chaohui Li, Naoko Tomoda, Naoyuki Miyashita, Tetsuo Kan, and Jay Yang*

**Department of Engineering Science, Bioscience and Technology Program, The Graduate School of Informatics and Engineering, The University of Electro-Communications (UEC), 1-5-1 Chofugaoka, Chofu, Tokyo 182-8585, Japan.*

Tel: +81-42-443-5980; E-mail: taki@pc.uec.ac.jp (M.T.)
E-mail: jyang75@wisc.edu (J.Y.)

Contents

1. General

All the reagents and solvents were purchased commercially and used without further purification. Human α-thrombin (Haematologic Tech Inc, # HCT-0020, USA) was purchased commercially. TBA with introduction of a relatively long spacer and a terminal alkyne (*i.e.,* 5-octadinyl-dU; U^*) replacing a thymine residue (TBA_x), and carboxyfluorescein (FAM)-tagged complementary strand against TBA (Fl-CS) were synthesized by Integrated DNA Technologies (IDT) Inc. (Table S1).

Table S1. Nucleic acids purchased from IDT.

sample		sequence
alkyne (**5**)-containing TBA (TBA_3)		5'-GGU*TGGTGTGGTTGG-3'
alkyne (**5**)-containing TBA (TBA_4)		5'-GGTU*GGTGTGGTTGG-3'
alkyne (**5**)-containing TBA (TBA_7)		5'-GGTTGGU*GTGGTTGG-3'
alkyne (**5**)-containing TBA (TBA_9)		5'-GGTTGGTGU*GGTTGG-3'
alkyne (**5**)-containing TBA (TBA_{12})		5'-GGTTGGTGTGGU*TGG-3'
alkyne (**5**)-containing TBA (TBA_{13})		5'-GGTTGGTGTGGTU*GG-3'
wild-type TBA		5'-GGTTGGTGTGGTTGG-3'
FAM-tagged CS	(Fl-CS)	FAM-5'-CCAACCACACCAACC-3'

NMR experiments were performed at 25 °C using a 500 MHz spectrometer (JNM-ECA500, Jeol Resonance, Japan). Liquid chromatography (LC) analysis was performed on an Agilent 1100 HPLC system (Agilent Technologies, USA) using a 0–100% gradient of acetonitrile containing 0.1% formic acid at a flow rate of 300 μL per minutes, equipped with a C18 reverse-phase column (Hypersil GOLD, 2.1 × 100 mm, Thermo Fisher Scientific, USA) connected to a photodiode array (PDA) and/or a LCQ-Fleet ion trap mass spectrometer. A small-scale quantitative analysis of aptamers was carried out by using a reversed-phase semi-micro HPLC system (PU-2085 with C18 column, JASCO, Japan) connected to a fluorescence detector followed by a PDA.

All images of stained gel and in-gel fluorescence were captured by ChemDoc XRS+ (BioRad Laboratories Inc., USA) and band intensities were quantified using Image Lab software (BioRad Laboratories Inc., USA).

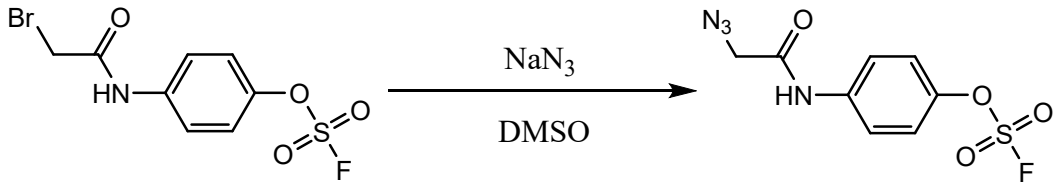
In addition to Hoon's pioneering work of the non-SELEX single-round selection using high-throughput sequencing (HTS) which is cited as Reference #24 in the main

text,[3] a review by Qian *et al.* [5] lists at least 14 references on single-round selection of non-covalent DNA aptamers published between 1990 to 2020.

2. Synthesis of covalent-binding DNA library

Benzenesulfonyl fluoride (BSF; Ar-SO_2F) is previously utilized by us as a successful warhead that can be conjugated to TBA_x, identified by the traditional SELEX, to create a covalently binding aptamer against thrombin.[6] Aryl-fluorosulfate (AFS; aryl-OSO_2F), also known as fosylate, is an analogous warhead which demonstrates minimal reactivity in bulk condition but with an enhanced reactivity when present in a correct microenvironment (*i.e.,* a latent warhead), and this warhead was successfully used in a combinatorial phage library for a covalently binding peptide screening.[7] Both warheads form a covalent bond with a target electrophile through the sulfur fluoride exchange (SuFEx) click chemistry.[8]

2.1. Synthesis of warhead 2 (N_3-AFS)

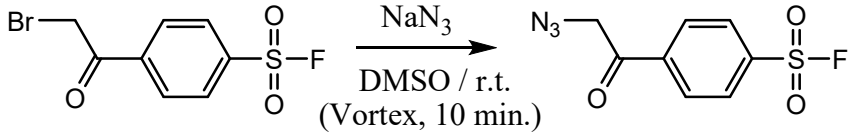


Warhead **2** (N_3-AFS) was synthesized on a preparative scale according to the following procedure. 4-bromoacetamide-benzenfluorosulfate (Br-AFS[7]; 22 mg, 70 μmol) and sodium azide (80 μmol, WAKO #195-11092, Japan) were mixed in 0.32 mL of dimethyl sulfoxide (DMSO) in a microtube. The reaction mixture was vortexed for 10 minutes at room temperature, then mixed with cold water (0.5 mL) and extracted with ethyl acetate (1 mL). The collected organic phase was washed with saturated $NaHCO_3$ (0.5 mL × 2) and brain (0.5 mL × 2), dried over Na_2SO_4, and evaporated to yield a yellow-brownish solid pure product (14 mg, yield 75%; TLC R_f = 0.19, hexane:AcOEt = 3:1). Identification of the warhead **2** was performed by ^1H NMR (Fig. S1).

As we could not observe the expected *m/z* value of the warhead because of its low ionization efficiency, both in positive and negative ionization modes of LC-MS analysis, the warhead **2** (N_3-AFS) was further derivatized by trimethylsilylacetylene (TMSA, WAKO #325-48221, Japan) via the CuAAC reaction as follows. TMSA (5.9 mg, 60 μmol), **2** (N_3-AFS; 14 mg, 51 μmol), $CuSO_4 \cdot 5H_2O$ (0.5 mg, 2 μmol), and sodium ascorbate (0.16 mg, 0.81 μmol, WAKO #195-01252, Japan) were mixed in 0.12

mL of DMSO and 0.24 mL of water using a microtube. The reaction mixture was vortexed overnight at room temperature, extracted with ethyl acetate, and the organic phase analyzed by LC-MS which revealed the expected *m/z* value at 373.12 [$(M+H)^+$; calculated for $C_{13}H_{18}FN_4O_4SSi$: 373.08].

2.2. Synthesis of warhead 1 (N_3-BSF)



The synthesis and characterization of warhead **1** (N_3-BSF) is fully described in our previous report.[4]

2.3. Synthesis of warhead-functionalized DNA library via CuAAC reaction

A DNA library, whose sequence is 5' – GGGAGGACGATGCGG – ACGT – N_{12} – [OctdU] – N_{12} – TGCA – CAGACGACGAGCGGGA – 3', was synthesized by Nihon Gene Research Laboratories (NGRL) Inc. Equal molar (2.5 nmol) of the DNA library and CS for both 5' and 3' ends were mixed in D-PBS (50 µL scale), and the mixture was heated for 2 min at 95 °C, and then brought to room temperature over 1 h. During this step, the terminal double-stranded structures were formed (underlined). Meanwhile, tris (3-hydroxypropyltriazolylmethyl) amine (THPTA in water, 0.12 µmol, Aldrich # 762342, USA), copper (II) sulfate ($CuSO_4$ in water, 0.03 µmol, Aldrich # 451657, USA), and freshly-prepared ascorbic acid (neutralized by NaOH in water, 0.75 µmol, Aldrich # A92902, USA) were successively mixed in water (30 µL scale) using a separated microtube, and the resulting catalytic solution was incubated at room temperature for 10 min. Then, the alkyne-containing library solution (2.5 nmol, 50 µL), azidated warhead (in DMSO, 0.05 µmol, 2 µL), and the pre-incubated catalytic solution (7.5 µL) were added, and the mixture was reacted for 1 hour at 4 °C. The crude reaction product was purified by ethanol precipitation. To the crude product, sodium acetate (in water, 18 µmol) and -20 °C cooled ethanol was added. After centrifugation (15000 rpm, 40 minutes, 4 °C), the supernatant was removed, and the pellet washed with 70% ethanol. The residue was dissolved in PBS supplemented with KCl (15 mM), to obtain the warhead-functionalized DNA library (25 µM). As a negative control, benzyl azide (Wako #357-40761) was used instead of the azidated warhead to obtain the corresponding 'non-reactive' library (R = **4** in Fig. 2A).

3. Screening of the library against the thrombin target

3.1. Immobilization of thrombin (*i.e.*, targeted protein) or BSA (*i.e.*, target unrelated protein) on FG-Nanobeads

Proteins were covalently immobilized on nano-sized magnetic beads (NHS Beads, Tamagawa Seiki #TAS8848N1141, Japan) using Protein Immobilization Buffer Kit (#TAB1200N0319) according to the manufacturer's protocol. Briefly, methanol-washed NHS Beads (1 mg, functional groups: ca. 0.2 μmol) was suspended in 50 μL of HEPES-NaOH reaction buffer (25 mM, pH 7.0), and briefly sonicated. Then, 50 μg of protein (*e.g.*, thrombin, 1.4 nmol) in the same buffer (50 μL) was added to the beads suspension, and the mixture was incubated overnight at 4°C using a rotator, followed by a removal of the supernatant. After the immobilized protein was indirectly quantified from the supernatant by the Bradford method via Quick Start Bradford Reagent, 1X kit (BioRad), the unreacted NHS functional groups on beads were blocked using 0.25 mL of aminoethanol (1.0 M, pH 8.0) containing 0.1% (w/v) NP-40. The suspension was incubated overnight at 4°C using a rotator, followed by a removal of the supernatant. The beads were washed 3 times using Wash/Storage Buffer containing (10 mM HEPES-NaOH (pH 7.9), 50 mM KCl, 1 mM EDTA, 10% (v/v) glycerol) at 4°C. The protein immobilized beads were stored in the same buffer in a refrigerator. The protein on the beads were quantified with the same Bradford kit, and ca. 20 and 45 μg (0.54 and 0.70 nmol) of thrombin and BSA was immobilized on 1 mg beads suspended in the buffer (0.20 mL). Just before the biopanning, 10 μL (*i.e.*, 50 μg beads for a single batch) of the suspension was pre-washed with D-PBS (0.20 mL), and the supernatant was removed.

3.2. Selection against Nanobeads-thrombin by using the warhead-functionalized DNA library

Before the targeted selection, a negative counter-selection against BSA-immobilized beads was performed using the warhead-functionalized DNA library to remove non-specific binders (*i.e.,* beads-, plastic-, and BSA-binders). The library (19 μL; 0.48 nmol, 2.9 x 10^{14} molecules / sample) was pre-incubated with the BSA-conjugated beads (50 μg beads; 35 pmol BSA / sample) for overnight at 37 °C with shaking at 800 rpm using a maximizer (MBR-022UP, Taitec, Saitama, Japan). For the targeted selection, the supernatant was further incubated with the thrombin-immobilized beads (50 μg; 27 pmol thrombin / sample) for overnight at 37 °C. Then, the beads were washed with 0.20 mL of a stringent wash buffer (50 mM Tris-HCl (pH 8.0), 1.0% (w/v)

SDS, 4.0 M urea, 0.15 M NaCl, 5 mM DTT) under constant sonication (30 minutes), and this washing process was repeated three times. Then, the washing was performed once again at 50 °C (30 minutes) under occasional sonication with shaking at 800 rpm by using the maximizer. The beads were further washed with D-PBS (0.20 mL) twice.

Three negative control libraries were tested to confirm that non-specific covalent conjugation with the target did not form. We used those possessing natural T (**3**), nonreactive benzyl mock warhead (**4**), or left unmodified as an alkyne (**5**), instead of the reactive warheads (**1** or **2**). In theory, the aforementioned control libraries should not covalently react with thrombin although a recent report suggests a possible enhanced non-covalent interaction between a terminal alkyne and a target protein.[9] A non-SELEX single round selection against thrombin (Fig. 2A in the main text) with the control libraries showed negligible PCR product band of the expected size (*i.e.,* 64 base pairs; 64bp) with a post-selection amplification cycle of 12, suggesting that essentially almost all of the non-covalent aptamers including both thrombin-specific and other non-specific binders were washed away during the harsh wash using the denaturing buffer (Fig. 2B, lanes **3**–**5**). When the post-selection amplification cycle was increased to 18, the 64bp band from the negative controls appeared with intensities like the AFS-library. This band was completely absent even with an 18-cycle PCR amplification after a mock selection without using any library DNA (data not shown). We tentatively consider that trace amounts of non-covalent binders in the system were present even after the harsh wash process but did not pursue this point further since the PCR band intensities were greatly reduced for the control libraries. We have confirmed the reproducibility of the order of the PCR band intensity as consistently AFS (**2**) >> BSF (**1**) > negative controls (**3**–**5**) over many experiments.

3.3. Aptamer unleashing from the beads and PCR amplification

Before the aptamer unleashing from the beads by proteinase-K (ProK) digestion, thrombin on the beads was reduced with 10 mM dithiothreitol in 30 mM NH_4HCO_3 aqueous solution at 60 °C for 10 min, and then alkylated with 0.10 M iodoacetamide in the aqueous solution at room temperature for 60 min. The beads were successively washed with 15 mM NH_4HCO_3 aqueous solution containing 50% (v/v) acetonitrile, then with 100% acetonitrile, and dried briefly in vacuo. Digestion of thrombin on the beads was carried out with 2.5 μL of 1 μg/μL proteinase K (Sigma, St. Louis, MO, USA) in a digestion buffer (50 mM Tris-HCl (pH 8.0), 10 mM $CaCl_2$) for 1 hour at 50 °C with shaking at 800 rpm by using the maximizer, and 3.5 μL of pure water was added. The

beads were further incubated overnight at 50 °C without shaking in the dark, then 0.2 μL of 0.10 M phenylmethanesulfonyl fluoride (PMSF) in ethanol was added to deactivate the enzyme. The mixture was incubated at room temperature for 30 minutes, and 6.0 μL of 2xPCR buffer for PWO polymerase was added. The unleashed DNA aptamer was separated in the liquid phase while boiled in the buffer at 98 °C (10 minutes) under occasional sonication with shaking, and 1.0 μL out of the liquid (12 μL) was used as a template for a hot-start polymerase chain reaction (PCR). The PCR was carried out with 12 cycles repeating three steps of denaturation at 95 °C for 30 seconds, annealing at 52 °C for 30 seconds, and extension at 72 °C for 1 minute) in a 10 μL scale using a forward primer (GGGAGGACGATGCGG) and a reverse primer (TCCCGCTCGTCGTCTG). Each PCR product was separated by polyacrylamide gel electrophoresis (PAGE), and the gel was stained by ethidium bromide.

Under the condition, ProK-treated eluent of the AFS-DNA library from the negatively counter-selected BSA beads never gave the robust PCR product of 64bp. This indicates that the covalent conjugation between the AFS-library and BSA did not occur due to an insufficient activation of the latent warhead at any microenvironment on the protein.

3.4. High throughput sequencing

The amplified selected aptamer band (64bp) was excised from the PAGE gel, dissolved in water (20 μL) under occasional sonication and boiling, and 1.0 μL of it was used as a template for PCR (11 cycles repeating three steps of denaturation at 95 °C for 30 sec, annealing at 58 °C for 30 sec, and extension at 72 °C, 1 min) to further amplify the library-coding region using a forward primer (ACACTCTTTCCCTACACGACGCTCTTCCGATCTGGGAGGACGATGCGGAC), a reverse primer (GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTTCCCGCTCGTCGTCTGTG), and GoTaq DNA polymerase (Promega). The 1st PCR product (131bp) for high throughput sequencing (HTS) was purified by 20% PAGE followed by ethanol precipitation and supplied to a vendor (Bioengineering Lab. Co., Ltd.) for the HTS analysis on Illumina MiSeq (ca. 10K reads) and MGI DNBSEQ-G400 (ca. 100K reads). In the latter case, the PCR amplification for making the DNBSEQ library was performed by the vendor.

4. Analysis of the HTS data

4.1. Preprocessing cleanup of the raw HTS reads

The forward and reverse raw reads in fastq format provided by the vendor were read into AptaSuite (v0.9.8)[10] with the designated forward and reverse primer sequences. Reads with 5' or 3' errors, invalid alphabet, or inconsistency between the two reads were eliminated, leaving the cleaned HTS reads with a 25-base randomized insert. Typically, greater than 90% of the raw reads were accepted. The HTS reads after cleanup was output as an HTS.fasta file using the Export option for further analysis.

4.2. K-mer distribution analysis

Tallymer[11] running under Ubuntu Linux (v22.04.3) implemented in Oracle VM Virtual Box on an MSI laptop computer with Windows 10 operating system was used for the k-mer analysis. Tallymer can be installed with the line command '$ sudo apt install genometools' from a desired directory in Ubuntu. The cleaned fasta file (HTS.fasta) is read by '$ gt suffixerator -dna -pl -tis -suf -lcp -v -db HTS.fasta -indexname HTSreads' and the k-mer read distribution in the desired range (minmersize 10 and maxmersize 20 in this example) obtained by '$ gt tallymer occratio -scan -output unique nonunique -minmersize 10 -maxmersize 20 -esa HTSreads'. The output, plotted as Figure 3A, lists the overall k-mer distribution:

distribution of non-unique k-mers (counting each non-unique k-mer only once)
10 952841
11 1987414
12 1863995
13 967982
14 359318
15 118611
16 43138
17 20636
18 13126
19 9863
20 7860

For a given k-mer (12 in the example), the read count is given by the command 'gt tallymer mkindex -scan -mersize 12 -minocc 110 -esa HTSreads' with the output (truncated between 6 and 79 reads):
1 3747856

2 1143688
3 394478
4 157973
⋮
80 1
81 1
98 1
110 1
tggtgtggttgg
124 1
tggtgtgggtgg

The output states that of all k-mers of length 12, there were 3747856 k-mers with one read each, 1143688 k-mers with 2 reads, up to one k-mer (tggtgtggttgg) with 110 reads, and one k-mer (tggtgtgggtgg) with 124 reads. The k-mer with the most reads is designated as the 'top seed' and the second most read as the '2nd seed'. The same analyses repeated for the different k-mer are plotted as Fig. S4B and the respective top and 2nd seeds are noted on the right. Full information on the optimal site of warhead conjugation required a 100K read but the k-mer alignment for identification of a potential consensus sequence could be obtained from a 10K read (Fig. S7). Negative control before the screening with any target protein gave no obvious consensus sequence (Fig. S8).

4.3. Alignment of k-mer seeds or HTS reads containing the seeds

Sequences and their ID tag identified in AptaSuite using the Search function were output as a tab delimited file, converted to a fasta file (Tab to Fasta Converter, choose file and convert it now (bugaco.com)) and aligned using CLUSTALW Result (genome.jp). The aligned *.aln file was read into SnapGene Viewer (SnapGene | Software for everyday molecular biology) and the consensus sequence with a threshold of >50% highlighted in yellow and noted on the top.

5. Evaluation of TBA_x conjugation to thrombin on gel electrophoresis

For the evaluation, AFS-conjugated TBA isomers (AFS-TBA_x) were synthesized. THPTA in water (40 nmol), $CuSO_4$ in water (10 nmol), and ascorbic acid (neutralized by NaOH in water and stored in -20 °C, 0.25 μmol) were successively mixed in water (10 μL scale) using a microtube, and the resulting catalytic solution was incubated at room temperature for 10 min. Then, the selected and chemically-

synthesized aptamer solution (TBA_x in water, Table S1, 5.0 nmol, 5.0 μL), the azidated warhead (in DMSO, 25 nmol, 1.0 μL) were added, and the mixture (25 μL) was reacted for 2 hour at room temperature using a rotator. The crude reaction product was purified by ethanol precipitation; sodium acetate (in water, 7.5 μmol) and -20 °C cooled ethanol were added to the crude product. After centrifugation, the supernatant was removed, and the pellet was washed with 70% ethanol. The pellet was briefly dried, dissolved in water (20 μL). The molar concentration and the reaction conversion of the functionalized TBA_x were estimated by UV absorbance using NanoPhotometer (Implen, German) and using a reversed-phase semi-micro HPLC system (PU-2085 with C18 column, JASCO, Japan) connected to a PDA, respectively.

In Fig. 4A, each AFS-conjugated TBA_x isomers (11 μM) were mixed with thrombin (11 μM) in Dulbecco's phosphate-buffered saline (D-PBS), and incubated for 3 or 24 hours at 37 °C. The samples were mixed with 4x SDS-PAGE Sample Buffer (SB; Wako) supplemented with 2-mercaptoethanol, briefly (*i.e.*, 1 minute) boiled on 95 °C, and separated by 13% sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE). Whole proteins were visualized by Oriole™ Fluorescent Gel Stain, and the mobility shift of thrombin by TBA_x was evaluated.

In Fig. 4B, specificity of covalent-binding between AFS-TBA_4 and thrombin was examined in the presence/absence of 40% (v/v) human serum (Aldrich #H4522, USA) in D-PBS. Thrombin (25 μM) with AFS-TBA_4 or wild-type TBA (0.10 mM) in the aqueous solution were incubated for 24 hours at 37 °C. Then, the mixture was supplemented with or without FAM-tagged CS (Fl-CS; 0.40 mM) for 30 minutes at 37 °C, mixed with SB without 2-mercaptoethanol, separated by 13% SDS-PAGE without boiling,[4] followed by an in-gel fluorescence imaging. Whole proteins were visualized by Coomassie Brilliant Blue (CBB) staining.

6. Statistical calculations

Values are presented as mean ± S.E.M. statistical significance of differences between groups was estimated by a paired t-test. A P-value < 0.05 was considered statistically significant. Analyses were performed using GraphPad Prism 6.0 software (GraphPad Software Inc., USA).

7. Supplementary figures

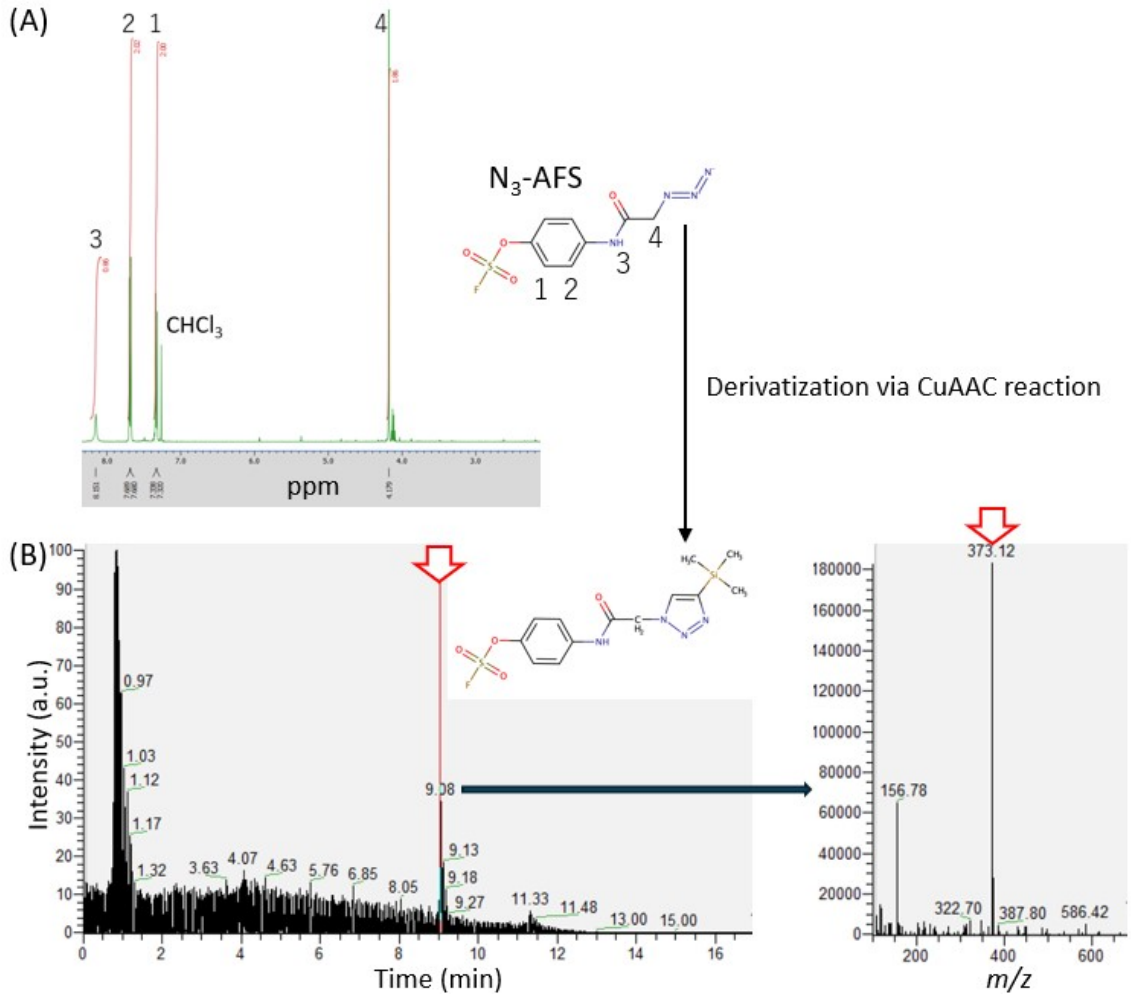


Fig. S1:[1]H NMR spectrum of N_3-AFS and LC-MS total ion profile of the derivatized warhead. (A) [1]H NMR spectrum ($CDCl_3$, 500 MHz) of the warhead **1** (N_3-AFS). (B) LC-MS total ion profile of the derivatized warhead by TMSA. After the derivatization via the CuAAC reaction, we could observe the expected *m/z* value of the *reacted* warhead derivatized at the N_3 moiety (right panel).

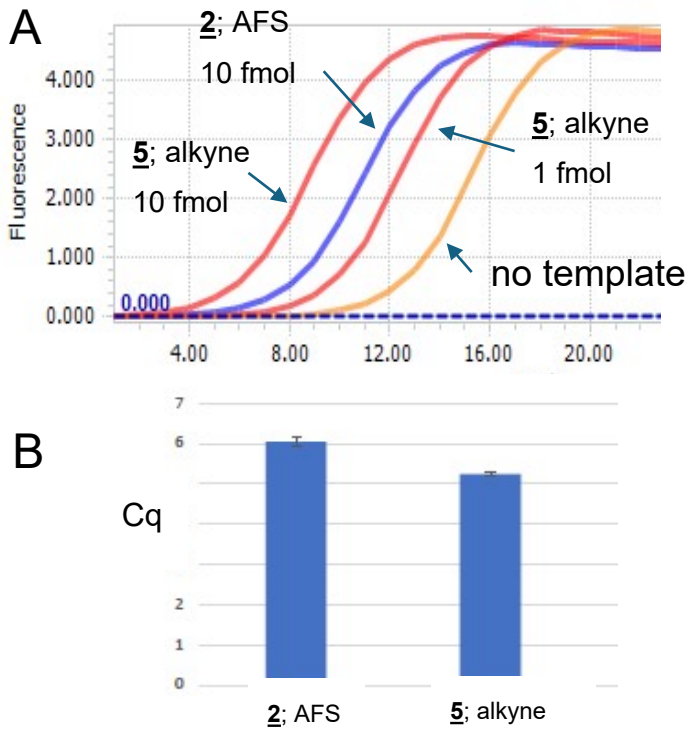


Fig. S2: Quantitative PCR (qPCR) amplification of aptamer library using PWO polymerase, followed by SYBR green detection. (A) Amplification curve of an aptamer library in Fig. 2A (*i.e.*, Octadinyl-dU at nucleotide 13) with (**2**; blue) or without (**5**; orange) the AFS-warhead conjugation. Equimolar template (10 fmol) shifted the Cq from 4.5 to 7.2 upon AFS conjugation. A Cq shift of 3.4 was observed when the control alkyne template amount was reduced from 10 (orange) to 1 (red) fmol. An amplification curve shifted to the right was observed for no template due to primer dimer formation (yellow). (B) Cq (mean ± S.E.M., n=3) of 10 fmol template with or without AFS. The conjugation of the AFS warhead resulted in an approximate 1.8-fold reduction in the PCR amplification efficiency.

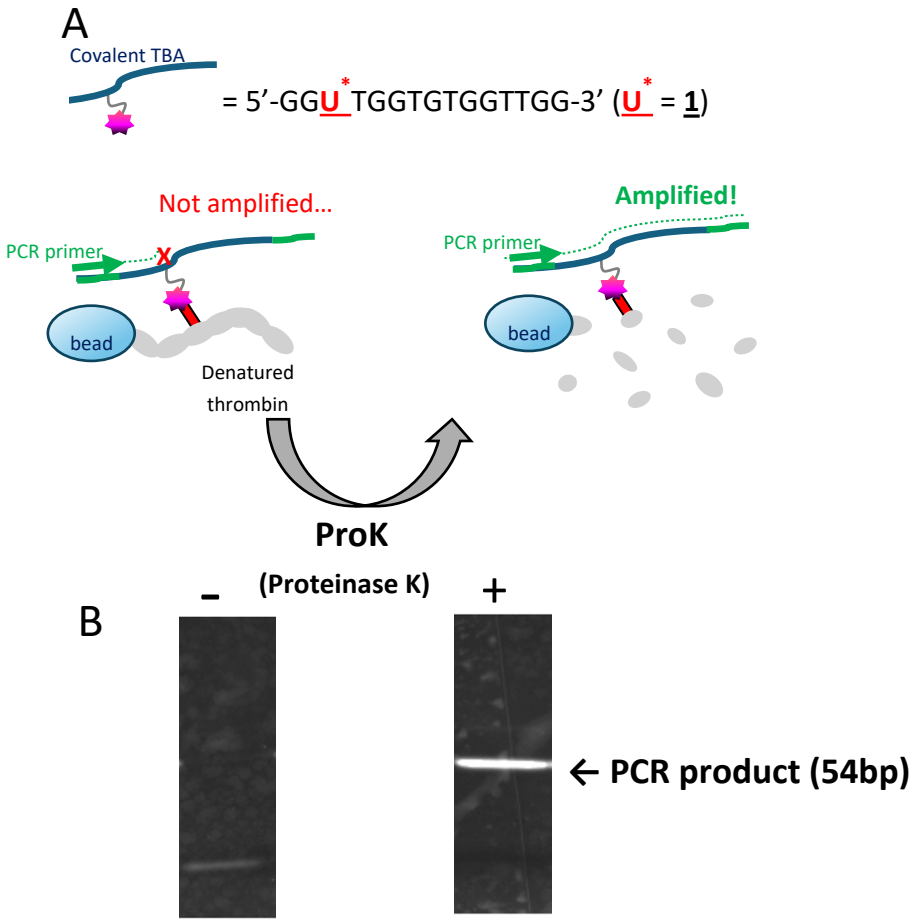


Fig. S3: Proteinase K (ProK) digestion enabled PCR amplification of covalently bound aptamer.[1] (A) TBA_3 aptamer (blue) with a single T replaced by an 5-octadinyl-dU (OctdU) residue with BSF (**1** in Fig. 2A)[4] denoted by a pink star at the 3rd residue (top). Whole sequence of the aptamer is: GGGAGGACGATGCGG – ACGT – GG – OctdU – TGGTGTGGTTGG – TGCA – CAGACGACGAGCGGGA (54 mer). (B) The aptamer covalently bound to the target protein immobilized on beads was not amplified by PCR but ProK treatment to digest thrombin leaving one (or a few) amino acids on the aptamer enabled PCR amplification giving the expected 54bp band seen on PAGE. The primers are GGGAGGACGATGCGG and TCCCGCTCGTCGTCTG.

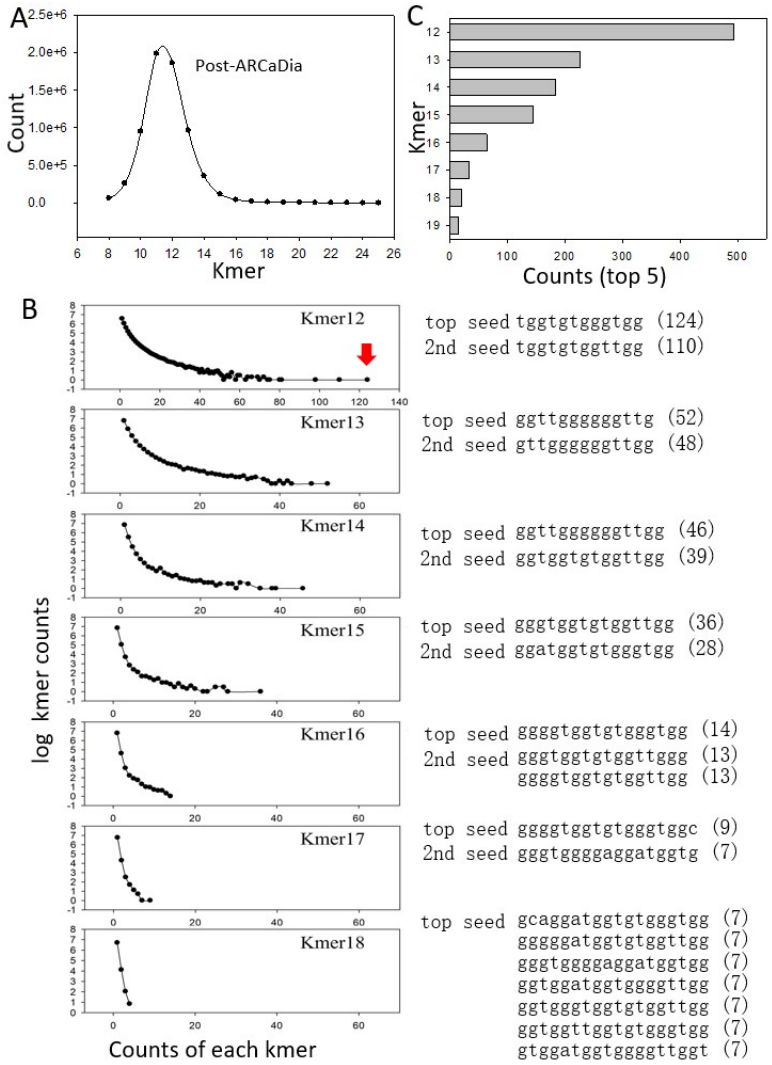


Fig. S4: A HTS read k-mer distribution of the AFS-library selected against thrombin. The gel isolated PCR product from lane 2 (Fig 2B) was subjected to a bi-directional HTS 100K read. The raw fastq files (forward and reverse reads) cleaned in AptaSuite were output as a single cleaned fasta file and k-mer analysis performed in Tallymer (see Methods for details). (A) Distribution of non-unique k-mers vs. k-mer length. The peak count was around 11mer with a decrease observed for shorter k-mers, probably from inadequate representation of the full library diversity under the experimental condition of 0.48 nmol input library. The distribution excludes all single reads (*i.e.,* unique reads). (B) Distribution of repetitive reads for each k-mer (log k-mer counts vs. counts of each k-mer) showing decreasing number of k-mers with multiple reads. For 12mer, the top read (red arrow) had 124 reads (denoted as top seed) and the second top read had 110 reads (denoted as 2nd seed). The top and 2nd seed sequences are denoted on the right for k-mers 12 to 18. (C) A bar graph of top 5 counts total for each k-mer. This analysis method is identical to that described by Hoon *et al.*[3]

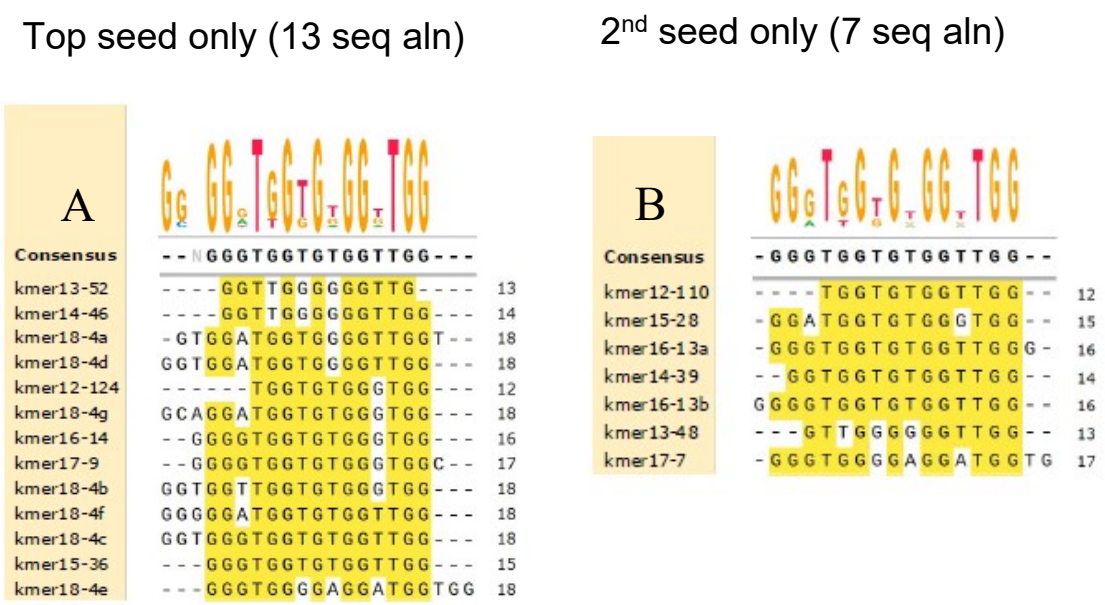


Fig. S5: Alignment of k-mer seeds from a single-round selection against thrombin. (A) Alignment of 13 top seeds for k-mers 12–18. (B) Alignment of seven 2nd seeds for k-mers 12–17. When equal count k-mers were present, all were included for the alignment. A k-mer seed alignment including both top and 2nd seeds is shown in Fig. 3A. Data from the same 100K HTS read. The Tallymer output of Thrombin_100K.fastq is appended as the Thrombin_100K tab on the K-mer Analysis Summary Excel file.

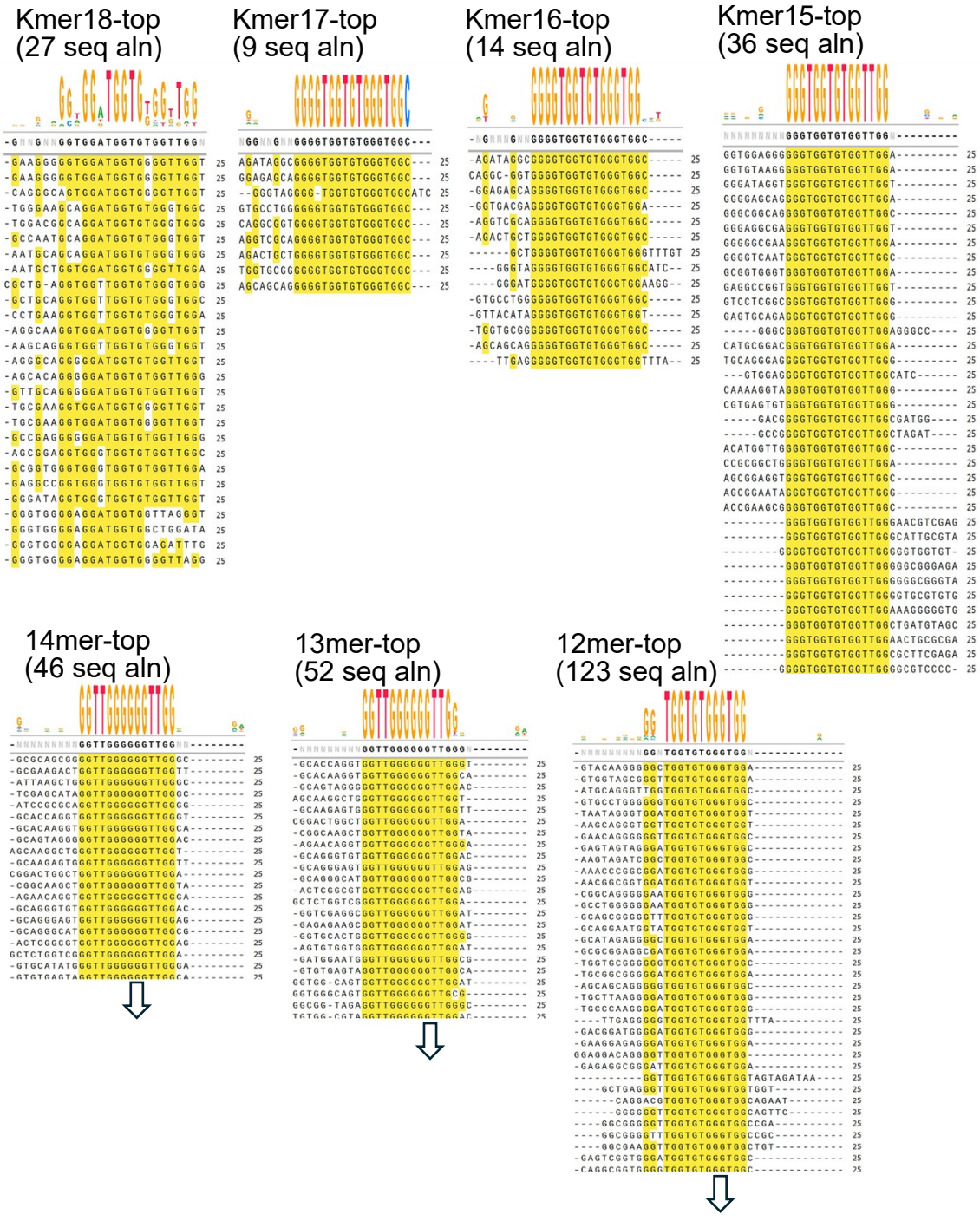


Fig. S6: Details of HTS reads identified by top seeds for each k-mer. The HTS reads were searched with the top read k-mer seeds for k-mer 18–12 and the full 25mer containing the seeds identified and aligned. The number of sequences identified is noted as # sequences aln (aligned) on the top. The full sequences for k-mer 12–14 abridged in the figure denoted by an open arrow can be found in the K-mer Analysis Summary Excel file appended in the Supplement.

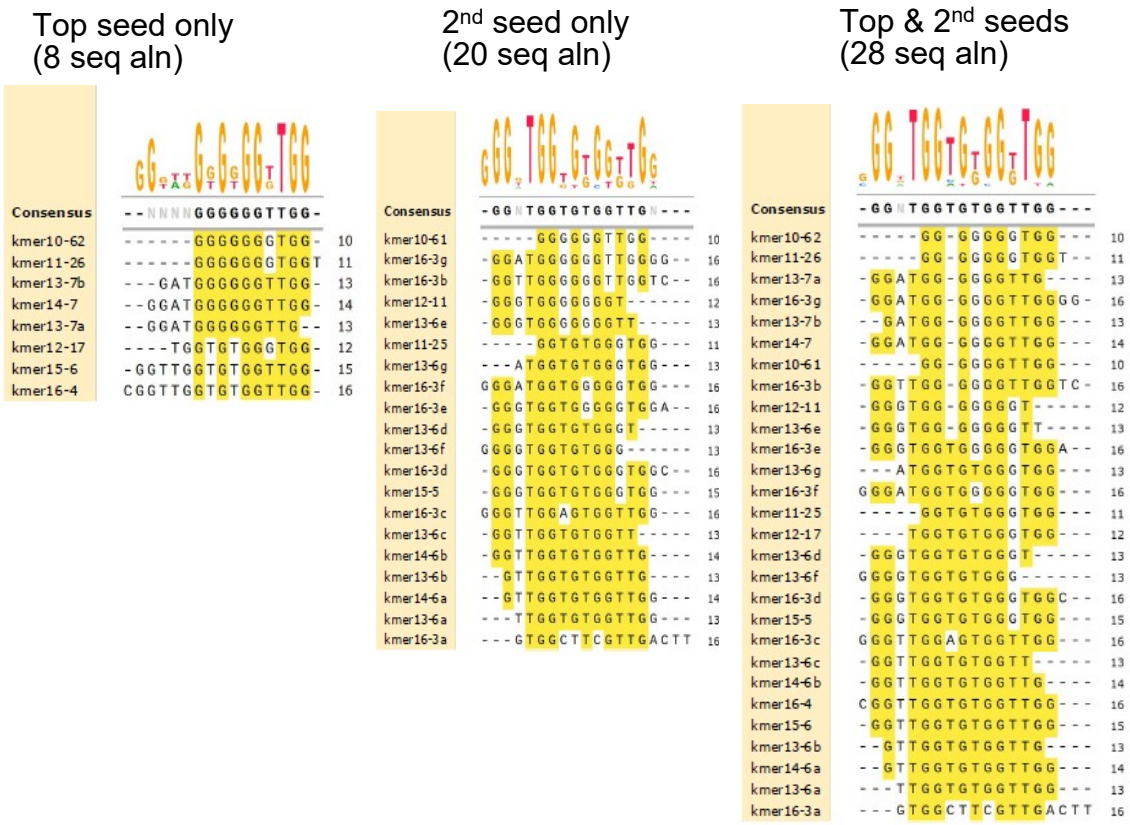


Fig. S7: Alignment of seed k-mers from a single-round selection and 10K HTS read. Identical k-mer analysis with data shown in Fig. S5, S6 but with the HTS read limited to 10K which is less costly and with a faster turnaround time from the vendor. The Tallymer output of Thrombin_10K.fastq is appended as the Thrombin_10K tab on the K-mer Analysis Summary Excel file.

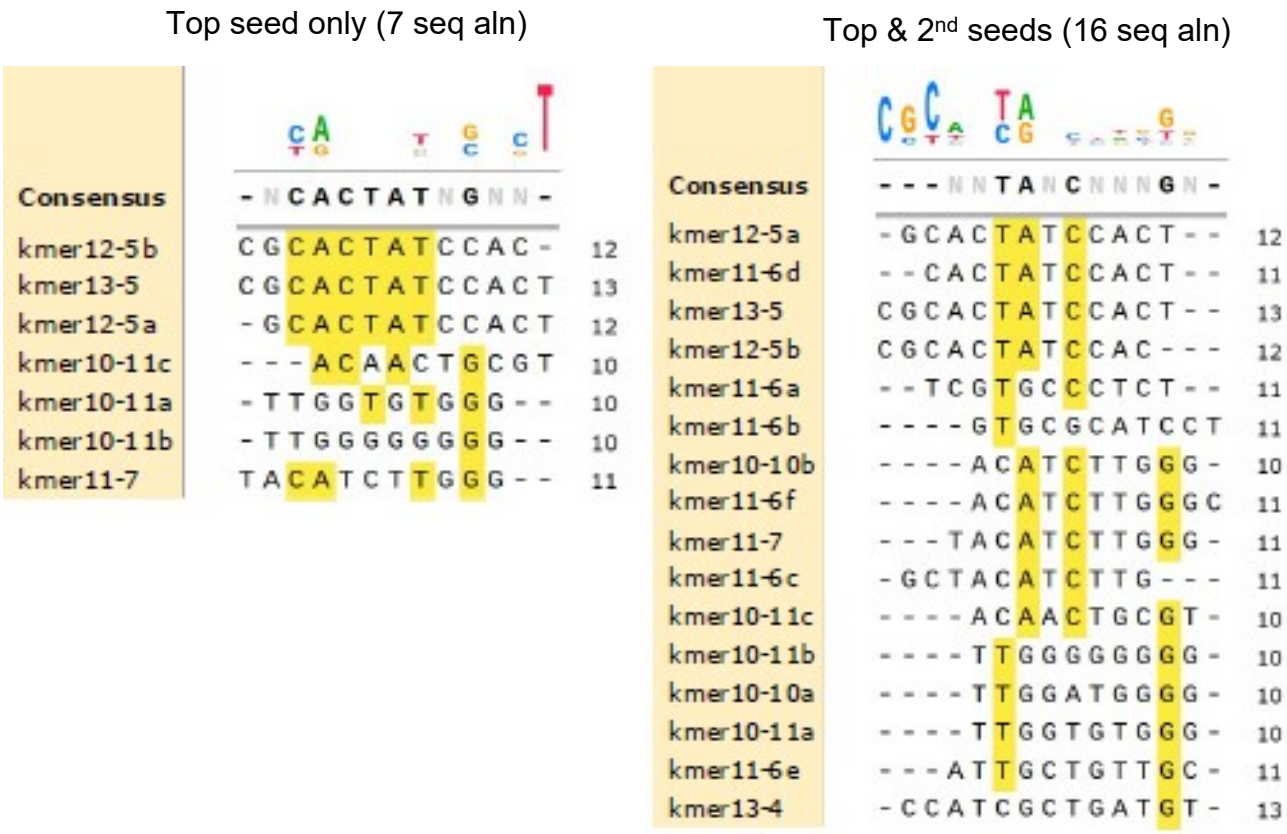


Fig. S8: Seed k-mers alignment *before* the selection against any protein. The alignments are for top seeds only (left) and both top and 2nd seeds combined (right). The Tallymer output of NoTarget_10K.fastq is appended as the NoTarg tab on the K-mer Analysis Summary Excel file.

8. Supplemental Files

- HTS read files (after clean up in AptaSuite):
  Thrombin_100K fastq
  Thrombin_10K fastq
  NoTarget_10K fastq
- K-mer Analysis Summary Excel file:

The sheet shows the overall non-unique and unique k-mer distributions, counts for each k-mer with the top (in yellow) and 2nd (in brown) highlighted. These are the "k-mer seeds" used to search for the actual HTS read 25mers containing the seeds. The tabs of the Excel file are for the Thrombin_100K, Thrombin_10K, and NoTarget_10K.

9. Appendix: History of covalent SELEX

In addition to the pioneering blended SELEX[12, 13] written in the main text, three major progresses have been made towards the generalization of direct screening of aptameric covalent binders by SELEX (Fig. 1A). In 1995, a photo-SELEX approach[14] was reported to conjugate with various responsive electron-rich amino acids (*i.e.*, W, Y, H, F, C, M, T) on target proteins.[1] A major limitation of the photo-SELEX is that it requires ultraviolet irradiation to initiate the cross-linking, which is not workable for *in vivo* use. Moreover, the conjugation efficiency is usually low because most of the warhead is degraded during the photo-crosslinking process. Later, in 2018, the direct screening of aptamer without irradiation, namely RNA-SELMA, was revisited by MacPherson. The success of this approach relies on a pi-conjugated planer structure of an uridine-containing amine-reactive warhead (*i.e.*, *N*-hydroxysuccinimidyl ester). The rigid structure containing the highly reactive warhead restricts the promiscuous covalent conjugation with unfavorable lysines, limiting non-specific binding.[15] In 2023, Qin *et al.* further broadened the generality of combinatorial screening by using SuFEx-type warheads, which potentially reacts with nucleophilic amino acids (*e.g.*, K, Y, H). They elegantly couple/decouple the warheads to the main chain of a phosphorothioate-modified DNA-aptamer library during the selection/PCR-amplification processes, respectively, and successfully selected covalent binders for the viral RBD binding protein target.[16]

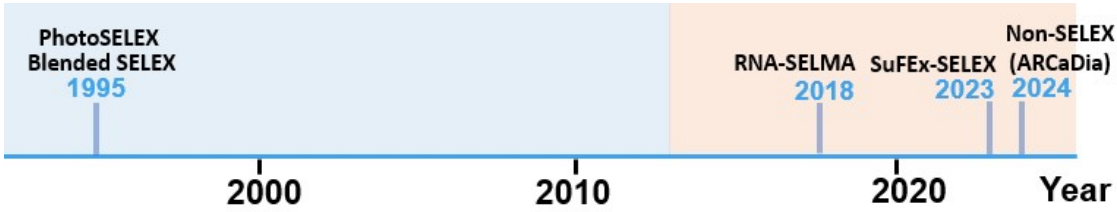


Fig. S9: A graphical summary of combinatorial screening methodology for obtaining nucleotidic (*i.e.*, aptameric) covalent binders. In our previous review article,[2] very important breakthroughs of the covalent SELEX (*i.e.*, PhotoSELEX and RNA-SELMA), which should have been cited, were lacking. We update the milestones in combinatorial covalent SELEX with references omitted in our recent review.

10. References

1. M. C. Golden, B. D. Collins, M. C. Willis and T. H. Koch, *J. Biotechnol.*, 2000, **81**, 167-178.
2. J. Yang, Y. Tabuchi, R. Katsuki and M. Taki, *Int. J. Mol. Sci.*, 2023, **24**, 3525.
3. S. Hoon, B. Zhou, K. D. Janda, S. Brenner and J. Scolnick, *Biotechniques*, 2011, **51**, 413-416.
4. Y. Tabuchi, J. Yang and M. Taki, *Chem. Commun.*, 2021, **57**, 2483-2486.
5. S. Qian, D. Chang, S. He and Y. Li, *Anal. Chim. Acta*, 2022, **1196**, 339511.
6. Y. Tabuchi, J. Yang and M. Taki, *Int. J. Mol. Sci.*, 2022, **23**, 7778.
7. Y. Tabuchi, T. Watanabe, R. Katsuki, Y. Ito and M. Taki, *Chem. Commun*, 2021, **57**, 5378-5381.
8. Q. Zheng, J. L. Woehl, S. Kitamura, D. Santos-Martins, C. J. Smedley, G. Li, S. Forli, J. E. Moses, D. W. Wolan and K. B. Sharpless, *Proc. Natl. Acad. Sci. U S A*, 2019, **116**, 18808-18814.
9. E. M. Kohn, K. Konovalov, C. A. Gomez, G. N. Hoover, A. K. Yik, X. Huang and J. D. Martell, *ACS Chem. Biol.*, 2023, **18**, 1976-1984.
10. J. Hoinka, R. Backofen and T. M. Przytycka, *Molecular Therapy-Nucleic Acids*, 2018, **11**, 515-517.
11. S. Kurtz, A. Narechania, J. C. Stein and D. Ware, *BMC Genomics*, 2008, **9**, 517.
12. D. Smith, G. P. Kirschenheuter, J. Charlton, D. M. Guidot and J. E. Repine, *Chem. Biol.*, 1995, **2**, 741-750.
13. J. Charlton, G. P. Kirschenheuter and D. Smith, *Biochemistry*, 1997, **36**, 3018-3026.
14. K. B. Jensen, B. L. Atkinson, M. C. Willis, T. H. Koch and L. Gold, *Proc. Natl. Acad. Sci. U S A*, 1995, **92**, 12220-12224.
15. I. S. MacPherson and T. Liedl, *Aptamers*, 2018, **2**, 74-81.
16. Z. Qin, K. Zhang, P. He, X. Zhang, M. Xie, Y. Fu, C. Gu, Y. Zhu, A. Tong, H. Wei, C. Zhang and Y. Xiang, *Nat. Chem.*, 2023, **15**, 1705-1714.