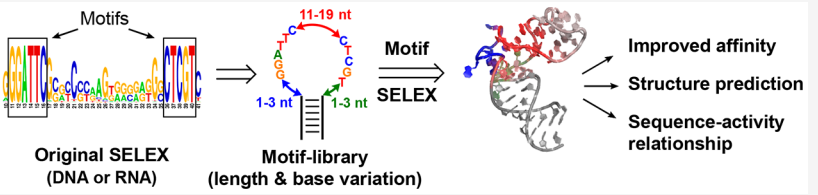
JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY
pubs.acs.org/JACS
Article

Improving Aptamer Affinity and Determining Sequence–Activity Relationships via Motif-SELEX

Haixiang Yu, Juan Canoura, Caleb Byrd, Obtin Alkhamis, Adara Bacon, Amy Yan, Bruce A Sullenger,* and Yi Xiao*

Cite This: J. Am. Chem. Soc. 2025, 147, 9472–9486
Read Online
ACCESS | Metrics & More | Article Recommendations | Supporting Information



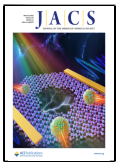
ABSTRACT: The affinity of nucleic acid aptamers isolated *in vitro* via Systematic Evolution of Ligands by Exponential Enrichment (SELEX) is often limited because the entire potential sequence space cannot be screened. In this study, we introduce Motif-SELEX, a novel method that enables the optimization of existing underperforming aptamers by generating libraries that broadly represent both the sequence and length variations of the parent sequence. This approach enables the isolation of sequences with improved affinity without the biases and limitations of traditional mutagenesis methods like doped SELEX and error-prone PCR. As a demonstration, we applied Motif-SELEX to a DNA-based morphine aptamer and a 2′ fluoro- and methoxy-RNA-based apixaban aptamer, discovering new, better-performing sequences with differing random domain lengths from their parents and up to 10-fold improvements in affinity. These new sequences would be inaccessible to traditional post-SELEX methods. Critically, our analysis of Motif-SELEX pools also enabled us to identify sequence and structural elements crucial for target binding and to predict secondary and tertiary structures for a given aptamer family—even when those structures involve noncanonical nucleotide interactions. We believe that Motif-SELEX offers an effective and generalizable solution for optimizing the structure and binding properties of functional nucleic acid molecules for diverse applications.

INTRODUCTION

Aptamers are short, single-stranded DNA or RNA molecules that exhibit high affinity for specific targets.[1] They can be isolated *in vitro* from random oligonucleotide libraries through a directed evolution process called Systematic Evolution of Ligands by Exponential Enrichment (SELEX).[2,3] To date, thousands of aptamers have been isolated against various small molecule, protein, and cell targets. Aptamers offer many advantages as affinity reagents, such as high affinity and specificity, excellent chemical stability, and ease of sequence and chemical engineering, making them useful in numerous applications including biosensing, diagnostics, and therapeutic development.[1,4,5] Despite their appeal, a longstanding challenge associated with aptamers is that SELEX often fails to generate aptamers with binding properties that are sufficient for real-world applications. This issue primarily results because of the inverse relationship between the length of the random domain of a library and the ability to represent all possible sequences in the starting library. Most SELEX libraries contain a random region of ≥30 nucleotides to confer sufficient structural complexity for aptamers to recognize targets with high affinity and specificity.[6] The sequence space for such libraries theoretically comprises $\geq 4^{30}$ (1.2×10^{18}) different sequences, but the initial screening round with SELEX is practically limited to an input of just 0.1–1 nmol of library, which comprises no more than 6×10^{13}–6×10^{14} different sequences. This large difference means that at most, < 0.05% of all possible sequences are being examined, making it highly likely that optimal sequences are being excluded from the beginning and that the isolated aptamers will have suboptimal properties. As Li and co-workers recently described it aptly, the ability to cover the sequence space of initial oligonucleotides libraries is mathematically equivalent to taking a single glass of water from nearly 100 Olympic-sized swimming pools.[7]

To address this problem, evolution-based strategies have been developed in which existing aptamers are subjected to new SELEX screens based on partially randomized libraries comprising variants that are intended to explore the sequence space around that aptamer. These libraries, typically generated

Received: November 29, 2024
Revised: February 22, 2025
Accepted: February 24, 2025
Published: March 7, 2025



https://doi.org/10.1021/jacs.4c17041
J. Am. Chem. Soc. 2025, 147, 9472–9486

by error-prone PCR[8,9] or "doped" library synthesis,[10−14] introduce mutations at each nucleotide position, and when subjected to *in vitro* selection, can yield aptamers with binding affinities that are improved by several fold relative to the original aptamer. The first report of doped-SELEX was by the Szostak group, which used a doped library to identify the binding motifs in a protein-binding RNA molecule.[12] The same group later implemented doped-SELEX to perform affinity maturation of an RNA cyanocobalamin aptamer, discovering variants with 3−4-fold improved affinity relative to the original aptamer.[14] Additionally, doped-SELEX has been employed to alter the specificity of aptamers, for instance, converting an aptamer for citrulline into arginine,[15] and improving signaling activity in the case of an artificial riboswitch.[16]

Despite the utility of previous aptamer affinity maturation approaches, they are unable to include nucleotide insertions and deletions in the library, which may prove critical for optimizing binding performance. Recently, the Stojanovic group devised a new approach to improve aptamer binding performance termed insertion-reselection,[17] wherein selections are performed with a library containing a known aptamer sequence flanked by additional randomized nucleotides that serve to enhance affinity or specificity. As an example, Yang et al. isolated a cocaine-binding aptamer with 3-fold improved target affinity using a library based on a previously discovered 36-nt aptamer with a 22-nt random domain insertion.[18] In this approach, however, the length of the inserted random domain is fixed, and the position of this insertion is arbitrary. This library design could prove disadvantageous if improved binding requires shorter or longer binding domains. New methods for incorporating both length and sequence variation into SELEX libraries have been reported—for instance, via nonhomologous random recombination[19] or skewed oligo synthesis[20]—but these are often complex and cannot equally present sequence length variation.

In this work, we demonstrate a new library generation method based on the assembly of multiple templates, where insertions, deletions, and point mutations can be simultaneously introduced to explore a large sequence space surrounding a given aptamer in an unbiased fashion. Our library creation strategy is based on the fact that even if the original SELEX screen is unlikely to find the optimal binding sequence, it can reveal shared consensus sequence motifs among the selected aptamers. These consensus motifs are usually directly involved in target binding and cannot be altered without impairing target affinity, while flanking sequence domains may play a supporting role in target recognition. We hypothesized that these supporting domains can be further optimized to improve binding performance. To do so, we devised a method to construct a "motif" library that uses the consensus motifs as scaffolds, which are then interspersed with variable domains consisting of randomized sequences with different length combinations. By fully randomizing the variable domain, we can eliminate the bias toward the original sequence, creating equitable opportunities for selective enrichment among all sequences. In parallel, by fixing the consensus domains, the extent of the randomized sequence can be reduced to such a scale that we can screen every possible individual sequence within the motif library.

By subjecting motif libraries to *in vitro* selection in this process, termed motif-SELEX, we enable the highly efficient searching of the sequence space around known aptamer domains to discover variants with greatly improved binding performance; the results in turn provide information about sequence-activity relationships. To evaluate this new evolution strategy, we have performed motif-SELEX with previously isolated DNA and chemically modified RNA aptamers that recognize for morphine and apixaban, respectively. Motif-SELEX successfully identified new sequences based on both aptamers that exhibit a 3−10-fold improvement in binding affinity. These superior sequences, which have variable domains that differ in length from the original aptamer, could not have been isolated from either the original library or discovered via traditional mutagenesis-based methods or insertion-reselection. Moreover, by analyzing the evolution of sequences enriched by motif-SELEX, we were able to identify sequence and structural elements that are beneficial to target binding, and which ultimately allowed us to predict the folded structures of the aptamers. Thus, motif-SELEX offers the important advantage of providing valuable information to guide aptamer sequence optimization, structure prediction, and post isolation engineering.

MATERIALS AND METHODS

Reagents and Materials. Molecular biology grade water and 10× phosphate buffered saline were purchased from Corning. Ultrapure water with a resistivity of 18.2 MΩ·cm was obtained from a Milli-Q eq 7000 water purification system. Morphine sulfate hydrate, epinephrine HCl, (−)-norepinephrine bitartrate hydrate, and fentanyl HCl were purchased from Cayman Chemicals. Diphenhydramine HCl, lidocaine HCl, quinine hemisulfate monohydrate, L-phenylalanine, glycine, GABA, naloxone HCl dihydrate, naltrexone HCl, dopamine HCl, serotonin HCl, sodium dodecyl sulfate, and Triton X-100 were purchased from Sigma-Aldrich. Normorphine, morphine-3-glucuronide, and morphine-6-glucuronide were purchased from Cerilliant. GoTaq Hot Start Master Mix was purchased from Promega. QIAquick PCR purification kit was purchased from Qiagen. SYBR Gold, streptavidin-coated agarose resin (capacity: 1−3 mg biotinylated BSA/ml resin), 0.5 M EDTA solution (pH 8.0), formamide, and Greiner 384-well flat-bottom transparent plates were purchased from Thermo Fisher Scientific. X-732−91B dye was retrieved from the Max Weaver Dye Library at North Carolina State University. All other chemicals were purchased from Sigma-Aldrich unless otherwise specified. All DNA aptamers were purchased from Integrated DNA Technologies. RNA aptamers (with 2'OMe A, C, and U, and 2'F G) were synthesized using a Mermade 12 synthesizer and purified by HPLC. A list of sequences used in this work is provided in Supporting Information (SI), Table S1.

Library Generation. Libraries were generated as described in SI, Figures S1, S2, and S3. For the morphine motif DNA library (SI, Figure S1), the templates were mixed at a ratio such that each sequence has the same frequency in the final mixture (i.e., F1:F2:F3 = 1:4:16, M1:M2:M3:M4:M5 = 1:16:256:4,096:65,536, and R1:R2:R3 = 1:4:16). PCR was used to assemble the morphine library with a input of 1 nmol F1/F2/F3 mixture, 10 pmol M1/M2/M3/M4/M5 mixture and 1 nmol R1/R2/R3 mixture, with the following conditions: 2 min at 95 °C; 15 cycles of 95 °C for 15 s, 50 °C for 30 s, and 72 °C for 45 s; and 5 min at 72 °C. Single-stranded DNA generation was performed using streptavidin agarose resin and NaOH treatment as described previously.[21]

For generation of the chemically modified RNA library for the first APX1 phase (SI, Figure S2), the templates were mixed at a ratio so that each sequence has the same frequency in the final mixture (i.e., APX-F1−1:APX-F1−2:APX-F1−3 = 1:4:16, APX-M1−1:APX-M1−2:APX-M1−3 = 1:16:256, and APX-R1−1:APX-R1−2:APX-R1−3 = 1:4:16). One nmol APX-F1−1/APX-F1−2/APX-F1−3 mixture and 1 nmol APX-R1−1/APX-R1−2/APX-R1−3 mixture were assembled with 30 U Klenow Fragment (3′ → 5′ exonuclease activity, New England Biolabs) in 500 μL of NEBuffer 2 at 37 °C for 90 min.

https://doi.org/10.1021/jacs.4c17041
J. Am. Chem. Soc. 2025, 147, 9472−9486

Single-stranded DNA templates containing two random regions were generated using streptavidin agarose resin and NaOH treatment as described above, and then assembled with 1 nmol APX-F1−1/APX-F1−2/APX-F1−3 mixture using the same Klenow reaction to generate dsDNA templates containing all three variable domains. The single-stranded RNA library was then transcribed from this double-stranded DNA template library with LAL T7 transcriptase in 20 μL buffer (40 mM Tris-HCl (pH 8.3), 40 mM DTT, 1 mM spermidine, 0.01% Triton X-100, 5% PEG-8000, 25 mM $MgCl_2$, 10 mM $MnCl_2$, 5 mM NTP (2'OMe A, C, and U, and 2'F G)) at 37 °C overnight. The transcribed product was treated with 10 U DNase I (New England Biolabs) at 37 °C for 1 h to remove DNA templates, followed by PAGE gel purification. The motif library for the second phase was generated using the same method but with different templates (APX-F2−1, APX-F2−2, APX-F2−3, APX-M2−1, APX-M2−2, APX-M2−3, APX-R2−1, APX-R2−2, and APX-R2−3) (SI, Figure S3), and the ratio of APX-M2−1:APX-M2−2:APX-M2−3 was changed to 1:4:16.

SELEX Procedure. All library-immobilized SELEX experiments were performed as described previously.[21] For morphine, selection conditions are detailed in SI, Table S2. At the beginning of each round, a 250 μL solution of 1 μM library was prepared with 5 μM biotinylated 15-nt cDNA in selection buffer (1× PBS, pH 7.4, 1 mM $MgCl_2$), heated to 90 °C for 10 min, and slowly cooled to room temperature over 20 min in a water bath. The library-cDNA complex was then passed through a gravity column (0.8 mL) containing 250 μL streptavidin-coated agarose beads preconditioned with selection buffer. The library-immobilized beads were washed several times with a selection buffer to remove nonspecifically eluting library strands. Starting from Round 2, counter-SELEX was performed to remove aptamers that bind to interferents. Afterward, several washing steps with selection buffer were performed to remove excess interferents, and positive selection with the target was performed by adding the target to the column and collecting all eluted library molecules. The eluted strands were purified using a 10 kDa molecular weight cutoff filter and PCR amplified using the following conditions: 2 min at 95 °C; 11 cycles of 95 °C for 15 s, 58 °C for 30 s, and 72 °C for 45 s; and 5 min at 72 °C. Successful PCR amplification of the library was confirmed via PAGE. Single-stranded DNA generation was then performed as described above. The pool was subjected to additional rounds of SELEX until the elution of pool by target exceeded was ∼5−10× pool elution by buffer alone. After each round, the pool target affinity was characterized using a previously reported gel elution assay.[22]

For apixaban selection, the above procedure was modified for RNA libraries. Specifically, selection was done in selection buffer (20 mM HEPES pH 7.4, 150 mM NaCl, 2 mM $CaCl_2$, 2% DMSO) before the PCR step, and the enriched sequences were reverse transcribed with 20 U AMV reverse transcriptase (Roche) with 0.2 nmol 3′ primer and 25 nmol dNTP in 50 μL 1x AMV reverse transcriptase buffer at 37 °C for 1 h. After PCR, the double-stranded template was purified via QIAquick column (Qiagen) and ∼250 pmol of the template was transcribed to regenerate the RNA library as described above. Details about selection conditions are in SI, Tables S3 and S4.

Isothermal Titration Calorimetry (ITC). ITC experiments were performed in selection buffer at 23 °C using a Malvern MicroCal iTC200 instrument. 360 μL of aptamer in selection buffer without $MgCl_2$/$CaCl_2$ was first heated to 95 °C for 10 min, and then cooled immediately on ice for 1 min; divalent salts were subsequently added after equilibration at room temperature. 300 μL of the solution was loaded into the cell, and the syringe was loaded with target in selection buffer. An initial purge injection of 0.4 μL was performed, followed by 19 successive injections of 2 μL with a spacing of 180 s between each injection. Additional titrations were performed if sufficient saturation was not achieved. The data was then fitted with a one-site binding model using the MicroCal analysis kit integrated into Origin 7 software. Specific conditions and experimentally determined binding affinity and thermodynamic parameters are presented in SI, Table S5.

HTS Analysis. Double-stranded DNA pools generated after different rounds of motif-SELEX were prepared for HTS by Azenta Life Sciences. Prior to submission, adapters were added via PCR, and the product was purified using the QIAquick PCR purification kit. The HTS raw result in .fastq was first processed with Aptasuite 0.98[23] to filter sequences without primers, trim primer sequences, and calculate the frequency of each sequence in different selection rounds. The sequences were further processed in Microsoft Excel. The length of each variable domain of each sequence was obtained using the FIND() function. The sequence of each variable domain was then obtained using the MID() function. The frequency summation of sequences possessing specific domain lengths, sequences, or combinations were calculated using the SUMIFS() function. For sequence logo generation, sequences were first aligned using the MAFFT algorithm by DNASTAR. The sequence logos were then plotted using the weblogo server.[24] The Circos plot was made by the Circos online table viewer.[25] Mutation and comutation maps were produced using Microsoft Excel; to simplify analysis, we chose all sequences with the most popular combinations of variable domain lengths so that all variable domains were aligned, and each base was assigned to a fixed position. The base identity of each position in every sequence was then obtained using the MID() function, and the base distribution of every position was calculated using the SUMIFS() function. The base with the highest frequency at a given position was used as the reference sequence; other bases at a given site were considered as point mutations, and the mutation rate of each position was calculated accordingly. The comutation rate (P_{CM}) between every two positions was calculated as

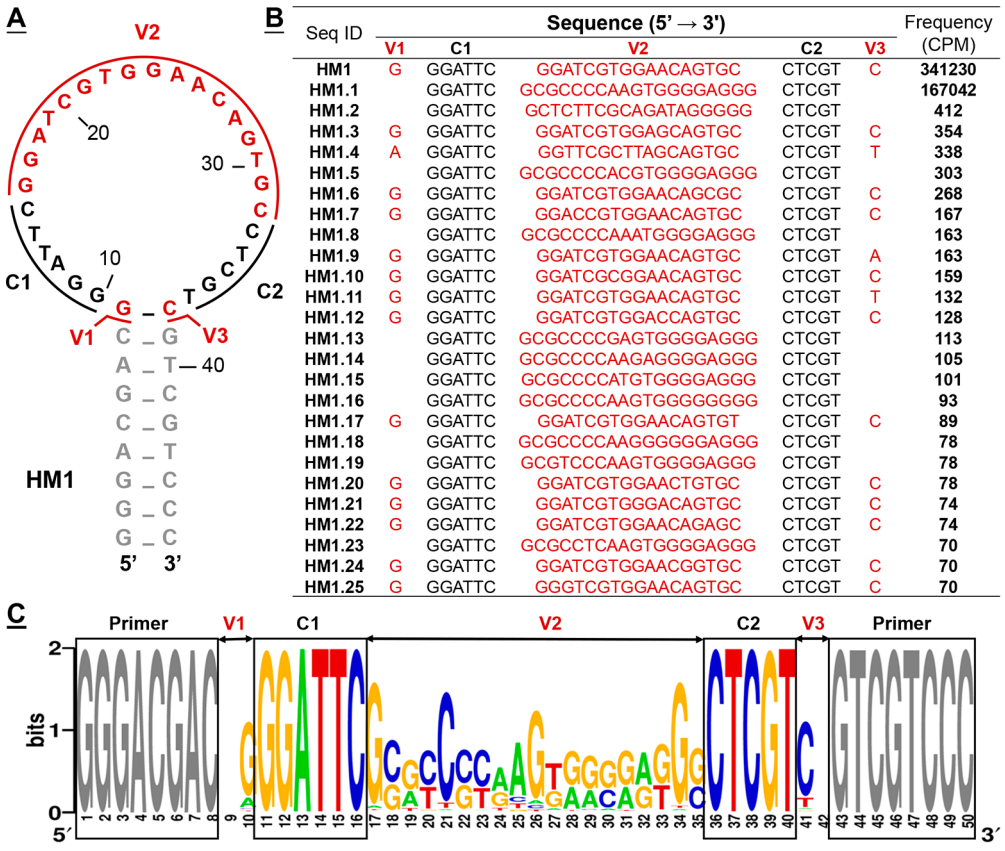
$$P_{CM} = P_{AB}/(P_A + P_B - P_{AB})$$

where P_A and P_B are the frequency summation of all sequences carrying mutations at the first and second positions, respectively, and P_{AB} is the frequency summation of all sequences carrying mutations at both positions. A P_{CM} of 0 means two positions never mutate together, while a P_{CM} of 1 means two positions always mutate together. The calculation was performed using the SUMIFS() function. The base-pairing rate (P_P) between every two positions was also calculated by summing the frequencies of all sequences that form Watson−Crick base pairs (A-T, T-A, C-G, and G-C) between two given positions using the SUMIFS() function.

Structure Modeling. Aptamer secondary structures were predicted based on the reference sequence from the mutation maps. For the HM1 family, the structure was predicted using mfold with forced base-pairing between positions 1−29, 10−27, and 14−22. Since mfold does not predict pseudoknot structure, the secondary structure of APX1 family was predicted manually based on forced base-pairing between positions 1−29, 4−14, and 10−22. The tertiary structure modeling was done using the RNAComposer web server.[26] For HMR10.2, the structure was predicted by replacing T with U since the server only accepts RNA sequences. For APXd2.2, this modified RNA aptamer was treated and modeled as canonical RNA.

Dye Displacement Assay. The dye-displacement assays were performed as reported previously.[27] First, to optimize the concentration ratio of aptamer and dye, the UV−vis absorbance spectrum of a fixed concentration of X-732−91B dye (final concentration: 4 μM) dissolved in 1× PBS with 1 mM $MgCl_2$, 0.01% (w/v) SDS, 0.001% Triton X-100 (v/v), and 1% (v/v) DMSO was determined in the presence of various concentrations of either parent aptamer HM1 (final concentration: 0−20 μM) or matured aptamer HMR10.6 (final concentration: 0−10 μM). To do so, the aptamer was first dissolved in 1× PBS, heated at 90 °C for 5 min, and then immediately cooled to room temperature. Afterward, $MgCl_2$, Triton X-100, and SDS were added to the solution. To create aptamer solutions of different concentrations, the aptamer solution was diluted with the same buffer. Then, 99 μL of aptamer solution was mixed with 1 μL of 100× X-732−91B dissolved in DMSO and mixed rapidly by inverting the tube several times. To record absorbance spectra, 75 μL of the sample was loaded in a 384-well flat bottom transparent microplate and scanned with a Tecan Spark plate reader from 400−800 nm with 5 nm steps. Using the optimal ratio of aptamer and dye,

https://doi.org/10.1021/jacs.4c17041
J. Am. Chem. Soc. 2025, 147, 9472−9486



Seq ID	Sequence (5' → 3') V1	C1	V2	C2	V3	Frequency (CPM)
HM1	G	GGATTC	GGATCGTGGAACAGTGC	CTCGT	C	341230
HM1.1		GGATTC	GCGCCCCAAGTGGGGAGGG	CTCGT		167042
HM1.2		GGATTC	GCTCTTCGCAGATAGGGGG	CTCGT		412
HM1.3	G	GGATTC	GGATCGTGGAGCAGTGC	CTCGT	C	354
HM1.4	A	GGATTC	GGTTCGCTTAGCAGTGC	CTCGT	T	338
HM1.5		GGATTC	GCGCCCCACGTGGGGAGGG	CTCGT		303
HM1.6	G	GGATTC	GGATCGTGGAACAGCGC	CTCGT	C	268
HM1.7	G	GGATTC	GGACCGTGGAACAGTGC	CTCGT	C	167
HM1.8		GGATTC	GCGCCCCAAATGGGGAGGG	CTCGT		163
HM1.9	G	GGATTC	GGATCGTGGAACAGTGC	CTCGT	A	163
HM1.10	G	GGATTC	GGATCGCGGAACAGTGC	CTCGT	C	159
HM1.11	G	GGATTC	GGATCGTGGAACAGTGC	CTCGT	T	132
HM1.12	G	GGATTC	GGATCGTGGACCAGTGC	CTCGT	C	128
HM1.13		GGATTC	GCGCCCCGAGTGGGGAGGG	CTCGT		113
HM1.14		GGATTC	GCGCCCCAAGAGGGGAGGG	CTCGT		105
HM1.15		GGATTC	GCGCCCCATGTGGGGAGGG	CTCGT		101
HM1.16		GGATTC	GCGCCCCAAGTGGGGGGGG	CTCGT		93
HM1.17	G	GGATTC	GGATCGTGGAACAGTGT	CTCGT	C	89
HM1.18		GGATTC	GCGCCCCAAGGGGGGAGGG	CTCGT		78
HM1.19		GGATTC	GCGTCCCAAGTGGGGAGGG	CTCGT		78
HM1.20	G	GGATTC	GGATCGTGGAACTGTGC	CTCGT	C	78
HM1.21	G	GGATTC	GGATCGTGGGACAGTGC	CTCGT	C	74
HM1.22	G	GGATTC	GGATCGTGGAACAGAGC	CTCGT	C	74
HM1.23		GGATTC	GCGCCTCAAGTGGGGAGGG	CTCGT		70
HM1.24	G	GGATTC	GGATCGTGGAACGGTGC	CTCGT	C	70
HM1.25	G	GGATTC	GGGTCGTGGAACAGTGC	CTCGT	C	70

Figure 1. Characterization of the HM1 family identified from the Round 10 pool of the original SELEX trial.[28] (**A**) Secondary structure of HM1 with identified constant and variable domains. The simple stem-loop secondary structure is only used as a framework to show the aptamer's sequence and motifs of the HM family rather than its precise fold. (**B**) Selected top sequences of the HM1 family and their frequency in the original SELEX pool. (**C**) Sequence logo of the HM1 family. A larger font size in the sequence logo represents higher sequence conservation.

a target calibration curve was then created by challenging a fixed concentration of aptamer-dye complex (final concentrations of HM1 and HMR10.6 were 15 μM and 5 μM, respectively) with various concentrations of morphine (0−150 μM). Specifically, the aptamer was prepared in buffer as explained above and mixed with X-732−91B dissolved in DMSO. Then, 72 μL of this solution was mixed with 8 μL of morphine in buffer (or buffer as blank) at a 10× concentration. Samples were immediately loaded in the wells of a 384-well transparent plate and the absorbance spectrum of these solutions from 400−800 nm was determined using the Tecan Spark plate reader. For data analysis, the areas under the curve (AUC) for the H-aggregate (410−495 nm) and monomer/dimer (500−650 nm) were used to calculate signal gain according to the following equations: $R = AUC_{410-495\ nm}/AUC_{500-650\ nm}$ and signal gain = $(R/R_0) - 1$, where R and R_0 represent the ratio of the areas under the curve of aggregate and monomer absorbance in the presence and absence of target, respectively. All experiments were performed independently three times, and the limit of detection is defined as the lowest target concentration that yielded a signal (R) which is greater that the signal of the blank (R_0) plus three times the standard deviation of the blank.

RESULTS

Design, Generation, and Characterization of Motif Libraries Based on an Existing DNA Aptamer. To start to explore this approach, we utilized motif-SELEX to attempt to improve the affinity of HM1, a DNA aptamer we previously isolated for morphine after 10 rounds of conventional library-immobilized SELEX.[28] HM1 binds to morphine with a dissociation constant (K_D) of 12.6 ± 0.2 μM in 1× PBS with 1 mM $MgCl_2$ (SI, Figure S4).

We hypothesized that further refinement of the sequence through motif-SELEX could identify a derivative of this aptamer with higher affinity. We first determined the essential binding nucleotides and motifs in HM1. Notably, mfold[29] analysis of this aptamer predicted three possible secondary structures with similar folding energy (from −10.18 to −9.49 kcal/mol) (SI, Figure S5). We present a simplified version of HM1 in Figure 1A. We then examined the final pool from that SELEX screen to identify clustered sequences related to HM1. In total, we found 94 sequences in this "HM1 family" (Figure 1B and SI, Table S6), which comprised 51.5% of the pool. All members of this family contain two conserved motifs, C1 (GGATTC) and C2 (CTCGT), interspersed with three variable domains (V1, V2, and V3) that showed considerable diversity of sequence and length (Figure 1B,C). V1 and V3 varied between 0−2 nt in length while V2 was 15, 17, or 19 nt. Both the diversity of the variable domains and the ambiguity of the secondary structure confound efforts at sequence optimization via rational truncation and mutagenesis methods.

Based on this information, we designed a new motif library that contained conserved elements C1 and C2 interspersed by randomized domains in regions V1−V3. In this library, V1 and V3 could be 0−2 nt in length, while V2 could be 11, 13, 15, 17, or 19 nt (Figure 2A). For V2, we opted to include shorter domains than what was observed in the HM1 family in order to explore further opportunities to achieve enhanced binding. In theory, this library contains 1.29×10^{14} possible unique sequences: $(4^0 + 4^1 + 4^2) \times (4^{11} + 4^{13} + 4^{15} + 4^{17} + 4^{19}) \times (4^0 + 4^1 + 4^2)$ (see Supporting Note 1 for detailed calculations).

https://doi.org/10.1021/jacs.4c17041
J. Am. Chem. Soc. 2025, 147, 9472−9486

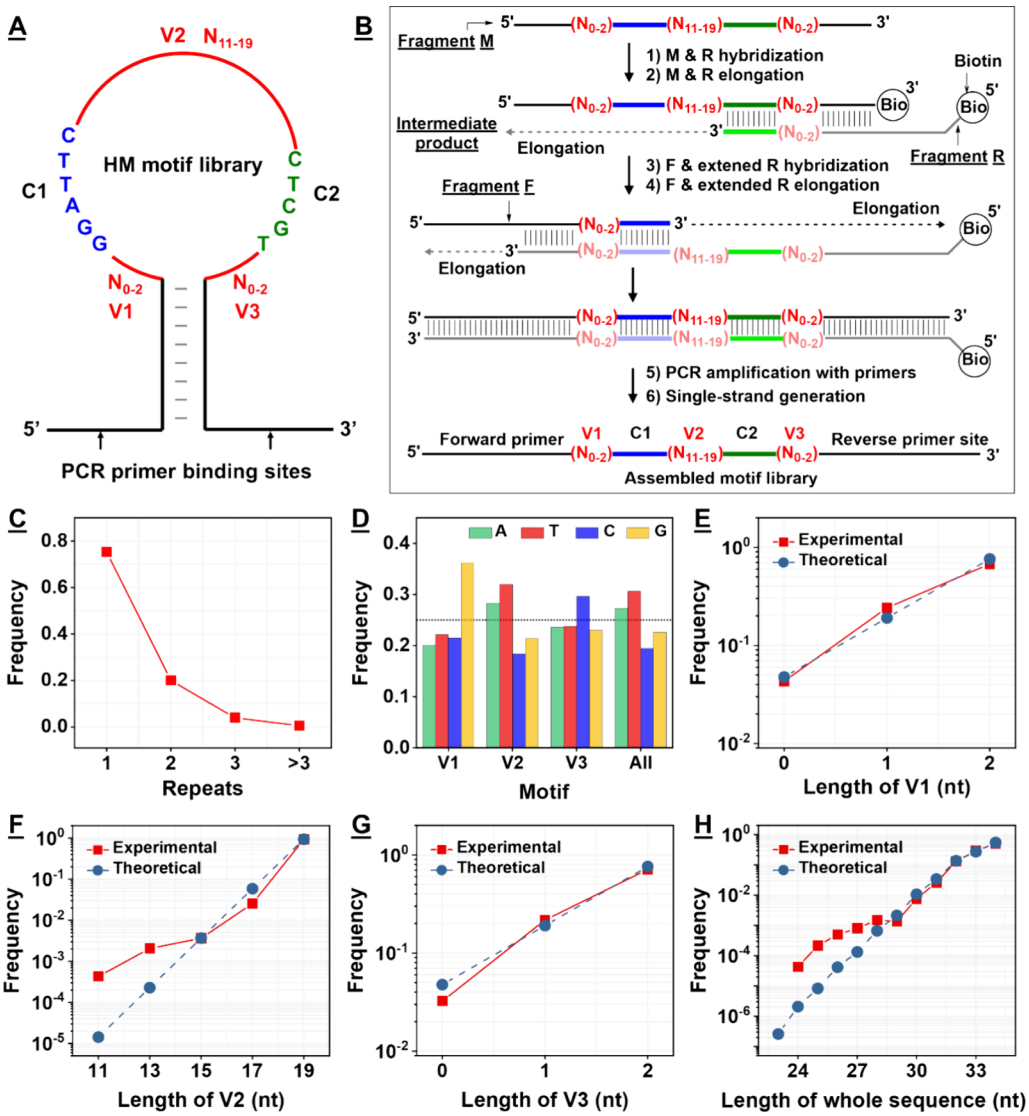


Figure 2. Motif library generation for aptamer family HM1. (**A**) HM1 motif library design. (**B**) Schematic of the PCR-based library generation process. First, the templates (fragment M) hybridize to the reverse primer (fragment R), and both strands are extended by the polymerase, generating an intermediate product containing all variable and constant domains without the forward primer site. In the next step, the forward primer (fragment F) hybridizes with the intermediate product and both strands are extended, yielding the full intact motif library. The library is then amplified by PCR, and then single-stranded library is generated by immobilization of the amplicons via the biotin tag on the reverse strand with streptavidin-coated resin followed by treatment with base. We analyzed the resulting library in terms of (**C**) pool diversity, (**D**) base distribution in V1, V2, and V3, and variable domains overall, and the length distribution of (**E**) V1, (**F**) V2, (**G**) V3, and (**H**) the whole sequence. Solid lines (experimental data) offer comparison to theoretical values (dashed lines) in panels **E**–**H**.

To represent every unique sequence once in this starting library, a minimum input of 215 pmol is required, which is well within the typical library input employed for *in vitro* selections. Due to the inclusion of all possible point mutations, insertions, and deletions, this motif library in theory offers more unique and informative diversity in terms of sequence, length, and structure relative to a library that would be generated by doped-SELEX or segment randomization.

To generate the HM1 motif library, we implemented a library assembly strategy facilitated by PCR. We synthesized a set of three forward primers (F1–3) that respectively contain a 0-, 1-, or 2-nt randomized V1 domain; five biotinylated templates (M1–5) that respectively contain an 11-, 13-, 15-, 17-, or 19-nt randomized V2 domain; and three biotinylated reverse primers (R1–3) that respectively contain a 0-, 1-, or 2-nt randomized V3 domain (Figure 2B and SI, Figure S1). All sets of primers and templates were mixed at a ratio such that each possible sequence has an equal chance of being assembled and amplified by PCR. The ratio of F1:F2:F3 was 4^0:4^1:4^2 (1:4:16); M1:M2:M3:M4:M5 was 4^{11}:4^{13}:4^{15}:4^{17}:4^{19} (1:16:256:4,096:65,536); and R1:R2:R3 was 4^0:4^1:4^2 (1:4:16). In the first PCR cycle, the reverse (R) primers are hybridized with the template (M) strands and elongated to produce double-stranded intermediate sequences containing V1, V2, and V3 domains that lack the forward primer region. In the second PCR cycle, the forward (F) primers hybridize with the intermediates, which are then elongated to yield the double-stranded motif library product containing all three variable domains, constant domains, and PCR primer-binding sites. In subsequent PCR cycles, the double-stranded motif library is further amplified by the F and R primers. After PCR, the single-stranded aptamer-motif library is purified using

https://doi.org/10.1021/jacs.4c17041
J. Am. Chem. Soc. 2025, 147, 9472–9486

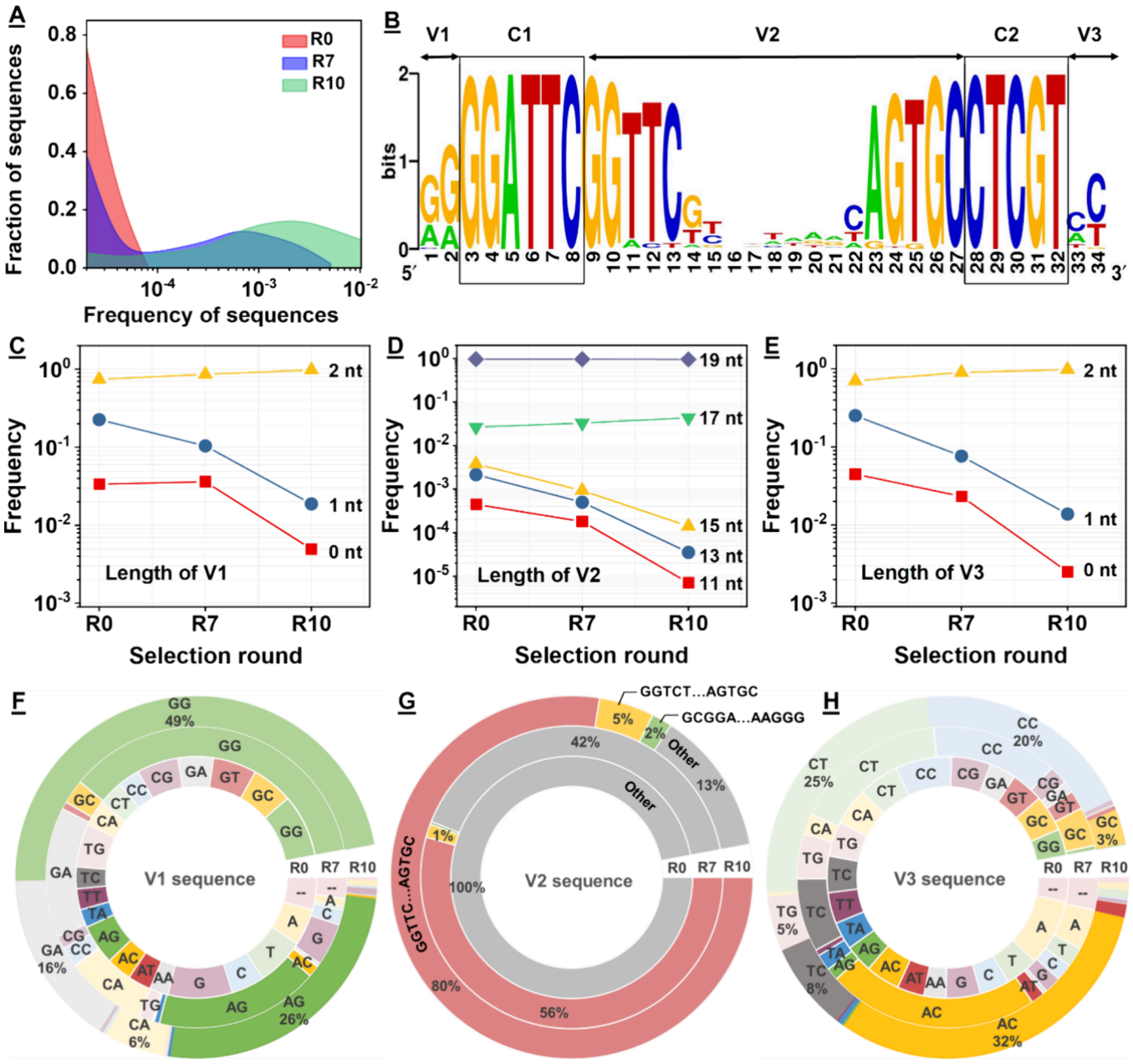


Figure 3. Evolution of aptamer sequences during motif-SELEX. (**A**) Frequency distribution in the initial (R0) library, Round 7 (R7), and Round 10 (R10) pools. (**B**) Sequence logo generated from the top 500 sequences in the Round 10 pool. Length distribution of (**C**) V1, (**D**) V2, and (**E**) V3 in the Round 0, 7, and 10 pools. Sequence distribution of (**F**) V1, (**G**) V2, and (**H**) V3 in the Round 0, 7, and 10 pools.

streptavidin-coated agarose resin, which remove all biotinylated strands from both the products and the intermediates (Figure 2B). Thus, we were able to generate a library containing sequences with 45 (3 × 5 × 3) length combinations of the three variable domains using only 11 (3 + 5 + 3) different lengths of chemically synthesized single-stranded templates.

We performed HTS analysis to verify that the motif library was correctly constructed. Out of a total of 93,284 reads, 96.1% contained intact constant domains with all variable domains within the expected length ranges, demonstrating that the library assembly method worked as designed. We also observed that the library retained high diversity, with more than 75% of the 77,999 unique sequences sampled in the library comprising a single copy (Figure 2C). All randomized domains had near-equal base distribution, with an A:T:C:G composition of 26:29:20:23, but the nucleotide composition of certain variable domains was somewhat skewed. Domain V1 had a bias for G (35%), domain V3 was biased toward C (27%), and domain V2 was somewhat biased toward A (27%) and T (30%) relative to C (19%) and G (20%) (Figure 2D). It is possible that the base distribution was unequal due to biases during the chemical synthesis of primers and templates, PCR amplification, or HTS. Nonetheless, we concluded that the library should provide sufficient variability for *in vitro* selection. We next examined the length distribution of each variable motif as well as the library sequences as a whole (Figure 2E–H). Since we used enough primer and template during the library construction process to theoretically represent each possible sequence with similar abundance, we expected that sequences

https://doi.org/10.1021/jacs.4c17041
J. Am. Chem. Soc. 2025, 147, 9472−9486

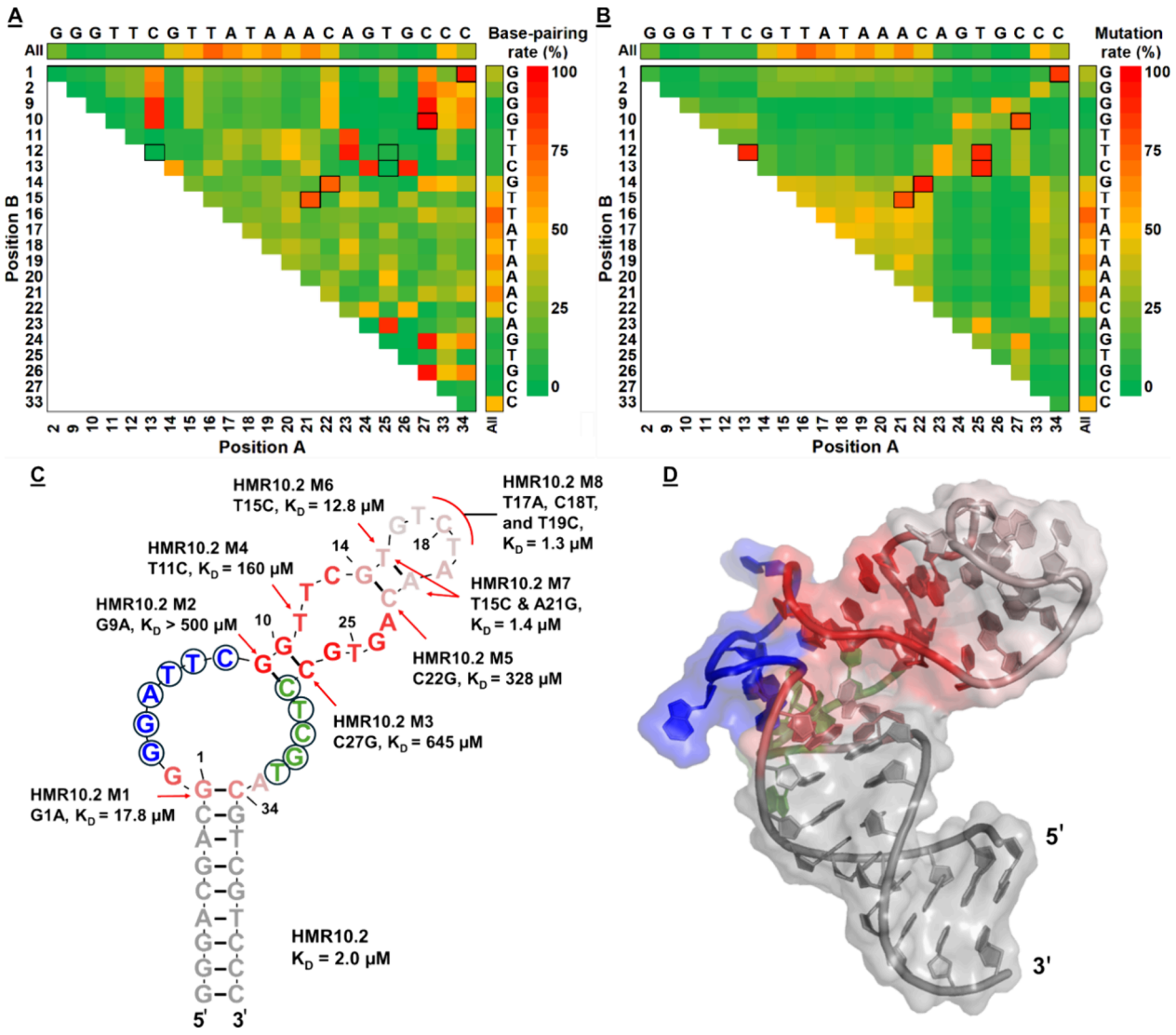


Figure 4. Predicting the structure of HM1 aptamers based on sequences in the Round 10 motif-SELEX pool. (**A**) Base-pairing and (**B**) comutation maps of sequences in the R10 pool in which V1, V2, and V3 are respectively 2, 19, and 2 nt in length. The bars along the top and right sides of the plot show the most popular base at each position as a reference, with its overall base-pairing/mutation rate depicted using the color scale. Position pairs with a comutation rate >75% are labeled with bold frames. Predicted (**C**) secondary structure of HMR10.2 using motif-SELEX data to constrain the prediction and (**D**) the corresponding tertiary structure. In both structures, the primer-binding regions are colored gray; C1 and C2 are colored blue and green, respectively; and V1−3 are colored red (bases with higher mutation rates are light red). **C** also depicts point mutants of HMR10.2 evaluated in this work. Aptamer tertiary structure was obtained with the RNA Composer software[26] based on the HTS-predicted secondary structure.

with longer random domains should be more popular than those with shorter lengths. For example, for V1 or V3, we predicted the relative abundance of sequences in the library spanning 0, 1, or 2 nt to be 4.8%, 19.0%, and 76.2%, respectively. We experimentally confirmed this, with a proportion of sequences with 0, 1, or 2 nt in domain V1 of 4.5%, 25.1%, and 70.3%, respectively. The only departure from predicted values was for V2 when the length was 11 or 13 nt, where we observed experimental values (0.04% and 0.21%, respectively) that were greater than the theoretical values (0.001% and 0.02%, respectively) (Figure 2F). Since the error in abundance measurements becomes relatively high for sequences with less than ∼0.1% abundance—at least for the number of reads that we obtained—it is possible that these values are affected by error due to the limited precision of HTS.[30] We observed similar results in the length distribution of the reassembled aptamer library, where the experimental results aligned with predicted values when the length was >28 nt but showed greater deviation with shorter sequences (Figure 2H). These results therefore established that the motif library was adequately constructed.

Performing Motif-SELEX to Identify New High-Quality Morphine Aptamers. Before initiating selection with the motif library, we assessed the affinity of the library for morphine using a gel elution assay. We observed no detectable library elution by morphine at concentrations up to 100 μM, indicating that the naïve library does not bind to the target well. We then performed 10 rounds of library-immobilized SELEX in 1× PBS with 1 mM $MgCl_2$ (see SI, Table S2 for detailed selection conditions). We began the selection using 250 pmol of the motif library (∼1.5 × 10^{14} molecules), which is theoretically sufficient to fully represent the potential

https://doi.org/10.1021/jacs.4c17041
J. Am. Chem. Soc. 2025, 147, 9472−9486

sequence diversity of this assembled library. For the first seven rounds, we used a constant morphine concentration of 100 μM to elute binders. From the second round on, we further enriched for highly specific aptamers by performing counter-SELEX against 17 counter-targets, including ligands known to bind three-way junction structured oligonucleotides, endogenous compounds, metabolites of morphine, and opioid receptor antagonists.[28] Over the first six rounds, library elution by morphine was relatively low (0.2−1.5%). In Round 7, pool elution increased to 6.2%, suggesting successful enrichment of morphine binders. To further enrich high-affinity aptamers, we gradually decreased the morphine concentration by 20-fold over the following three rounds. In Round 8, we used 25 μM morphine and pool elution rose to 7.7%. We then decreased morphine concentration to 10 μM in Round 9 and 5 μM in Round 10, and observed 4.7 and 3.4% pool elution, respectively (SI, Figure S6A). We characterized the binding affinity of the enriched pools for morphine using a gel elution assay and obtained K_Ds of 16 and 11 μM for the Round 7 and 10 pools, respectively, indicating that the pools had been sufficiently enriched with morphine binders (SI, Figure S6B,C).

High-Throughput Sequencing (HTS) Analysis of Morphine Selected Pools. To identify aptamers and examine the evolution of sequences derived through motif-SELEX, we subjected the DNA pools from Rounds 7 and 10 to HTS. In comparison to the starting motif library, which was mostly comprised of single-copy sequences, we saw considerable sequence enrichment in Round 7 and 10 pools (Figure 3A). For instance, in the starting library, 99% of the pool comprised sequences with abundance <0.0022%, while 50% of the Round 7 and 10 pools comprised sequences with abundance >0.0089% and >0.082%, respectively. The proportion of unique sequences also decreased sharply, from 87% in the initial library to 17% in Round 7 and 5.5% in Round 10. However, unlike traditional SELEX starting from a fully random library, where a few sequences typically dominate the final pool in terms of abundance (Figure 1B), no particular sequence dominated the Round 10 pool. The most abundant sequence, HMR10.1, represented just 1.1% of the pool. We hypothesized that since the library was constructed with conserved elements of the HM1 binding domain, many sequences in the library might bind morphine with similar affinity, such that no single sequence dominates the selection process. To assess this hypothesis, we selected 12 aptamer candidates with relatively high abundance (generally, abundance >0.02%) in Round 10 and tested their affinity with isothermal calorimetry (ITC). All aptamers bound morphine with a K_D of 1−2 μM, which is 5−10-fold stronger than HM1 (SI, Figure S7A−L). This confirms our hypothesis that several superior sequences were enriched in parallel.

We next assessed which sequence elements were shared among the high-affinity aptamers discovered via motif-SELEX. We created a sequence logo for the 500 most abundant sequences in the Round 10 pool to graphically present conserved nucleotides among these sequences. We observed new conserved regions at both the 5′ and 3′ ends of the V2 domain, namely GGTTC and AGTGC, respectively, (Figure 3B) that were identified in only three of the 94 sequences in the original HM1 family with abundance >0.05%. We then investigated the evolution of both the length and sequence of each variable domain (Figure 3C−H). The V1 and V3 domains were nearly all 2 nt in length, while V2 was generally 19 nt in length. In contrast, shorter sequences were gradually depleted during SELEX (Figure 3C−E). We also observed a strong sequence preference in the motif-SELEX data (Figure 3F−H), and although all 21 possible sequences for V1 were generally equally represented in the initial library, several sequences became undetectable in Round 7, and only four dominated the Round 10 pool (Figure 3F). Similar results were observed for V3, where only six sequences dominated by Round 10 (Figure 3H). For V2, we observed that three different 10-nt motifs (areas highlighted in pink, yellow and green), which were apparently undetectable in the starting library, dominated the Round 10 pool with a total abundance of 87% (Figure 3G). These results highlight the importance of these variable regions for aptamer-target recognition and indicate that particular combinations of lengths and sequences are preferred for optimal target binding.

Using Data from Motif-SELEX to Guide Aptamer Structure Prediction and Structure−Activity Relationships. We then investigated potential interactions between the different variable domains as a means for understanding aptamer structure and how structural features contribute to target binding. We first analyzed the potential base-pairing frequency for the 10,478 sequences in the Round 10 pool (92% of the pool) that had lengths of 2 nt, 19 nt, and 2 nt for V1, V2, and V3, respectively (Figure 4A) (see *Methods* for detail). We identified several different nucleotide pairs that had a high base-pairing probability (>75%). A few of these were conformationally prohibited, such as the base-pairing of adjacent or near-adjacent nucleotides (e.g., nucleotide 23 with 25 or nucleotide 26 with 27). In contrast, many other nucleotides had high probabilities of base-pairing to multiple different positions, making it difficult to determine the most probable secondary structure from these data alone. We hypothesized that true base-pairing interactions could be identified by examining the comutation rate of nucleotide pairs: if a particular base-pair is critical for aptamer functionality, a mutation in one nucleotide should be accompanied by a mutation in the other to maintain base pairing. For example, we consistently observed a strong correlation between sequences in the V1 and V3 domains. When the sequence of V1 was GG, the sequence of V3 was almost always AC or CC, and when the sequence of V3 was CT, the sequence of V1 was almost always AG (SI, Figure S8). This covariation demonstrates some measure of coevolution between variable domains.

To continue our coevolution analysis, we first generated a reference sequence based on the most frequently observed bases at each position and considered all other bases as point mutations (Figure 4B). We examined the comutation rate between each pair of positions across V1, V2, and V3. Out of all previously identified base-pairing candidates, only nucleotides 1−34, 10−27, 14−22, and 15−21 had high comutation rates (>75%), which indicated a high likelihood of base-pairing. Using this information, we predicted the shared secondary structure of the HM1 family (Figure 4C). Since C1 and C2 have little complementarity, they most likely form a bulge and could serve as the target binding site. V1 and V3 are close to this putative binding site and may also be involved in binding. The bulge is closed by a short two-base-pair stem (nucleotides 9−28 and 10−27). Base-pair 9−28 was excluded from the comutation map because nucleotide 28 was not randomized during selection. Nevertheless, the low mutation rate of nucleotide 9 suggests a high possibility of base pairing

https://doi.org/10.1021/jacs.4c17041
J. Am. Chem. Soc. 2025, 147, 9472−9486

with nucleotide 28. A second, smaller bulge forms after the stem within the V2 domain, which possibly facilitates target binding. Adjacent to this bulge is a two-base-pair stem (nucleotides 14−22 and 15−21) within V2 that is likely to maintain an optimal conformation for binding, followed by a 5-nt loop that is distal to the putative binding site. Since the loop has a high degree of sequence variability, we hypothesized that it probably does not contribute to binding. Using our predicted secondary structure, we modeled the tertiary structure of the aptamer and observed that the two bulges within the V2 domain fold to form a cavity that possibly serves as the target-binding pocket (Figure 4D).

We next evaluated the accuracy of our predicted aptamer structure and examined the structure−activity relationship of the HM1 family by determining the binding affinity of several point mutants of top sequences from the Round 10 pool with ITC (Figure 4C and SI, Figure S7 and S9). We began our mutation studies with the aptamer HMR10.2 (K_D = 2.0 μM for morphine) (SI, Figure S7B), which was both one of the most abundant aptamers in the pool (abundance >1%) and one of the most representative members of the HM1 family (≤5 point-mutations from the reference sequence). First, we examined the importance of the G1-C34 base pair in HMR10.2 by mutating G1 to A (HMR10.2 M1) and determined the affinity of this sequence for morphine using ITC. We observed a K_D of 17.8 μM (SI, Figure S9A), 8.9-fold weaker than the parent aptamer, which indicates that the G1-C34 base-pair is important for target binding. We also examined other pairs of aptamers that differed only at base-pair 1−34 (HMR10.1/HMR10.14, HMR10.7/HMR10.15, and HMR10.11/HMR10.16) (SI, Table S7), and found that the disruption of this base-pair generally reduced binding affinity by 7−9-fold (SI, Figure S7A,G,K and S9B−D). When we mutated G9 in the G9-C28 base pair to A (HMR10.2 M2), binding to morphine was completely lost (SI, Figure S9E), indicating that this base-pair is critical for target binding. Likewise, mutating C27 in the G10-C27 base pair to G (HMR10.2 M3) dramatically reduced the K_D to 645 μM (SI, Figure S9F), highlighting the importance of this dinucleotide stem. Interestingly, when we extended this stem by one more base-pair by mutating T11 to C to create a C11-G26 base pair (HMR10.2 M4), the affinity of the aptamer was reduced by 80-fold (K_D = 160 μM) (SI, Figure S9G). This observation is consistent with our structural prediction that nucleotides 11 and 26 are most likely not base-paired, and therefore might be involved directly in ligand binding.

We next examined the significance of the distal stem. We disrupted the predicted G14-C22 base pair with a C22 to G mutation (HMR10.2 M5), which reduced affinity by 164-fold (K_D = 328 μM) (SI, Figure S9H). By mutating the T15-A21 base pair with T15 to C (HMR10.2 M6), we observed a 6.4-fold decrease in affinity (SI, Figure S9I). This reduction in affinity for HMR10.2M6 could be rescued by further mutating A21 to G, forming a C-G base-pair (HMR10.2 M7, K_D = 1.4 μM) (SI, Figure S9J). These results indicate that base-pairing in the distal stem, rather than the identity of the nucleotides, is important to facilitate morphine binding. Finally, we assessed the role of the middle region of the V2 domain (nucleotides 16−20), which we predict to form a loop that is not directly involved in target binding. We mutated the GTCTA loop in HMR10.2 to a GATCA loop (HMR10.2 M8), and as predicted, found that this mutant retained a K_D of 1.3 μM (SI, Figure S9K). We found similarly high morphine affinity among several pairs of aptamers that differ only in this 16−20 loop region (HMR10.3 vs HMR10.2M7 (SI, Figure S7C), HMR10.4 vs HMR10.5 (SI, Figure S7D,E), and HMR10.8 vs HMR10.10 (SI, Figure S7H,J)). These mutagenesis data thus closely corroborate our structural predictions, confirming that base-pairs 9−28 and 10−27 are critical for target binding, base pair 1−34 is moderately important, base pairs 14−22 and 15−21 are important for binding, and the nucleotide 16−20 loop does not meaningfully impact affinity.

The above aptamer variants have a wide range of K_D (1.0−645 μM) and Round 10 pool abundances (0.0004−1.1%). Interestingly, we observed among this data set that as pool abundance increased, aptamer affinity increased, and vice versa. Specifically, when we plotted the logarithm of aptamer abundance in the Round 10 pool of motif-SELEX against the logarithm of aptamer target-binding K_D, a linear fit of the plot produced an R^2 of 0.85 (SI, Figure S10 and Table S7). This trend implies that highly enriched sequences have better binding affinities. However, we find that for aptamers with abundance >0.2%, there is no correlation between sequence abundance and affinity, with all aptamer K_Ds falling within a narrow range of 1−2 μM (SI, Figure S10). It is possible that the sensitivity of abundance as a surrogate of affinity is constrained by the limited precision of *in vitro* selection experiments, PCR, and sequencing. For example, the aptamer HMR10.2 M6 was a major outlier in our data set. This is unsurprising, however, considering that the pool abundance of this sequence is only 0.0007%, and the error in abundance readings is relatively high when sequence abundance is <0.1%.[30]

Aptamers Matured by Motif-SELEX Have Improved Sensing Performance. We next determined whether the matured morphine aptamers discovered through motif-SELEX had improved sensing performance relative to the parent aptamer HM1. Specifically, we compared the performance of the matured aptamer HMR10.6 (K_D for morphine = 1.2 μM) and HM1 (K_D = 12.6 μM) in our previously developed colorimetric dye-displacement assay with the cyanine dye X-732−91B.[27] Here, the aptamer is initially noncovalently bound to the dye as a monomer, which absorbs maximally at 570 nm, and the binding of target to the aptamer displaces the dye into solution, causing the dye to form H-aggregates with maximum absorbance at 450 nm and the solution to change color from pink to yellow. After optimizing the concentration of aptamer and dye in the assay (SI, Figure S11A−D), we challenged aptamer-dye complexes with various concentrations of morphine up to 150 μM. The addition of morphine caused an increase in H-aggregate absorbance and a decrease in monomer absorbance (SI, Figure S11E,F). We observed that the assay based on HMR10.6 was more sensitive than the HM1-based assay with a limit of detection (LOD) of 97 nM, which is a ∼ 16-fold improvement over that of the parent aptamer (LOD = 1.6 μM) (SI, Figure S11G,H). Importantly, the matured aptamer can be potentially used to detect clinically relevant concentrations of morphine in oral fluid.[31] A control experiment performed by challenging X-732−91B alone with various concentrations of morphine demonstrated that the target does not meaningfully perturb the aggregation state of the dye (SI, Figure S11I). These results therefore demonstrate that motif-SELEX represents a viable route to greatly improving the performance of underperforming aptamers.

https://doi.org/10.1021/jacs.4c17041
J. Am. Chem. Soc. 2025, 147, 9472−9486

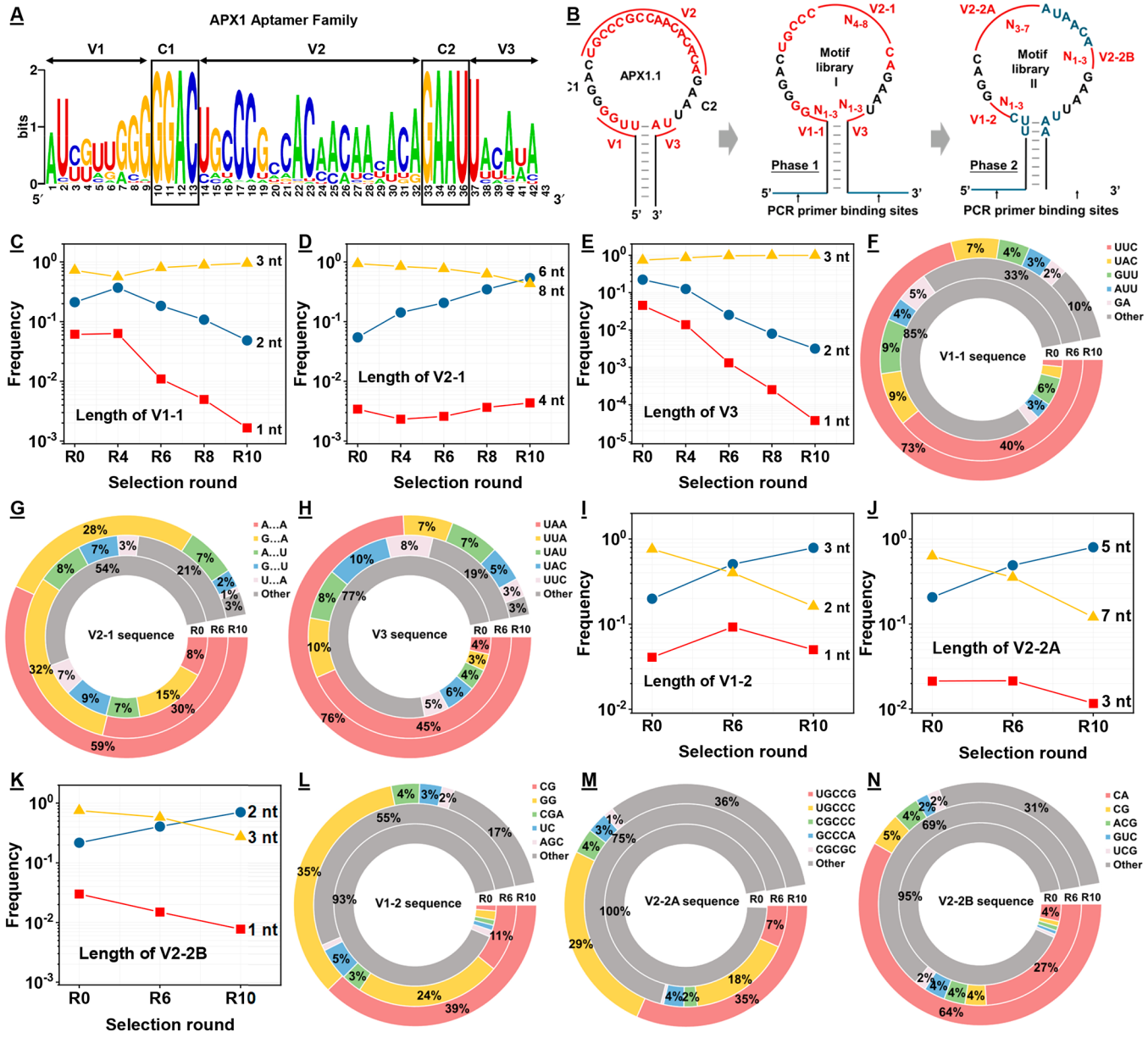


Figure 5. Evolution of aptamer sequences during a two-phase motif-SELEX screen for apixaban. (**A**) Sequence logo generated from 549 sequences in the APX1 family identified in the original SELEX screen. Consensus sequences are highlighted in boxes. (**B**) Library design for the first and second phases of motif-SELEX. Variable domains are colored red, and optimized sequences from the first phase are colored blue. Evolution of the lengths of (**C**) V1−1, (**D**) V2−1 and (**E**) V3, and (**F**−**H**) evolution of sequences of each variable domain during the first phase of motif-SELEX. We also assessed evolution of the lengths of (**I**) V1−2, (**J**) V2−2A and (**K**) V2−2B, as well as the (**L**−**N**) sequence composition for these variable domains for the second phase of motif-SELEX.

Evaluating the Generality of Motif-SELEX with an RNA Aptamer. Finally, we tested the generality and versatility of motif-SELEX by applying it to an RNA aptamer with 2′ chemical modifications: APX1.1 is an RNA aptamer that binds to the small molecule anticoagulant drug apixaban with a K_D of 2.26 ± 0.09 μM according to ITC (SI, Figure S12). We isolated APX1.1 using library-immobilized SELEX with an N30 stem-loop structured chemically modified (2′ methoxyl-A, C, U, and 2′ fluoro-G) library, and identified the APX1 aptamer family (Figure 5A) via HTS analysis of the Round 12 pool from that selection experiment. APX1.1 is a major constituent (70.2%) of that family (SI, Table S8), which contains 549 sequences with the consensus motif GGAC···GAAU and comprises 75.5% of the overall Round 12 pool. These two constant domains (C1 and C2) were interspersed with three domains, V1, V2, and V3, that were highly variable in sequence and length. The most prevalent lengths for V1, V2, and V3 were 4 nt (51%), 16 nt (48%), and 2 nt (55%), respectively. We initially considered a motif library in which V1 was 2−6 nt, V2 was 8−18 nt, and V3 was 1−3 nt. However, such a library contains 4.2×10^{16} unique sequences ($\sum_{L_1=2}^{6} 4^{L_1} \times \sum_{L_2=8}^{18} 4^{L_2} \times \sum_{L_3=1}^{3} 4^{L_3}$), making it impractical to represent all possible sequences in a single selection experiment (see Supporting Note 2 for detailed calculations). To overcome this limitation, we implemented a two-phase motif-SELEX approach. In the first phase, we generated a motif library in which we randomized the 5′ region of V1 (termed V1−1) the internal region of V2 (termed V2−1), and the entire V3 domain

https://doi.org/10.1021/jacs.4c17041
J. Am. Chem. Soc. 2025, 147, 9472−9486

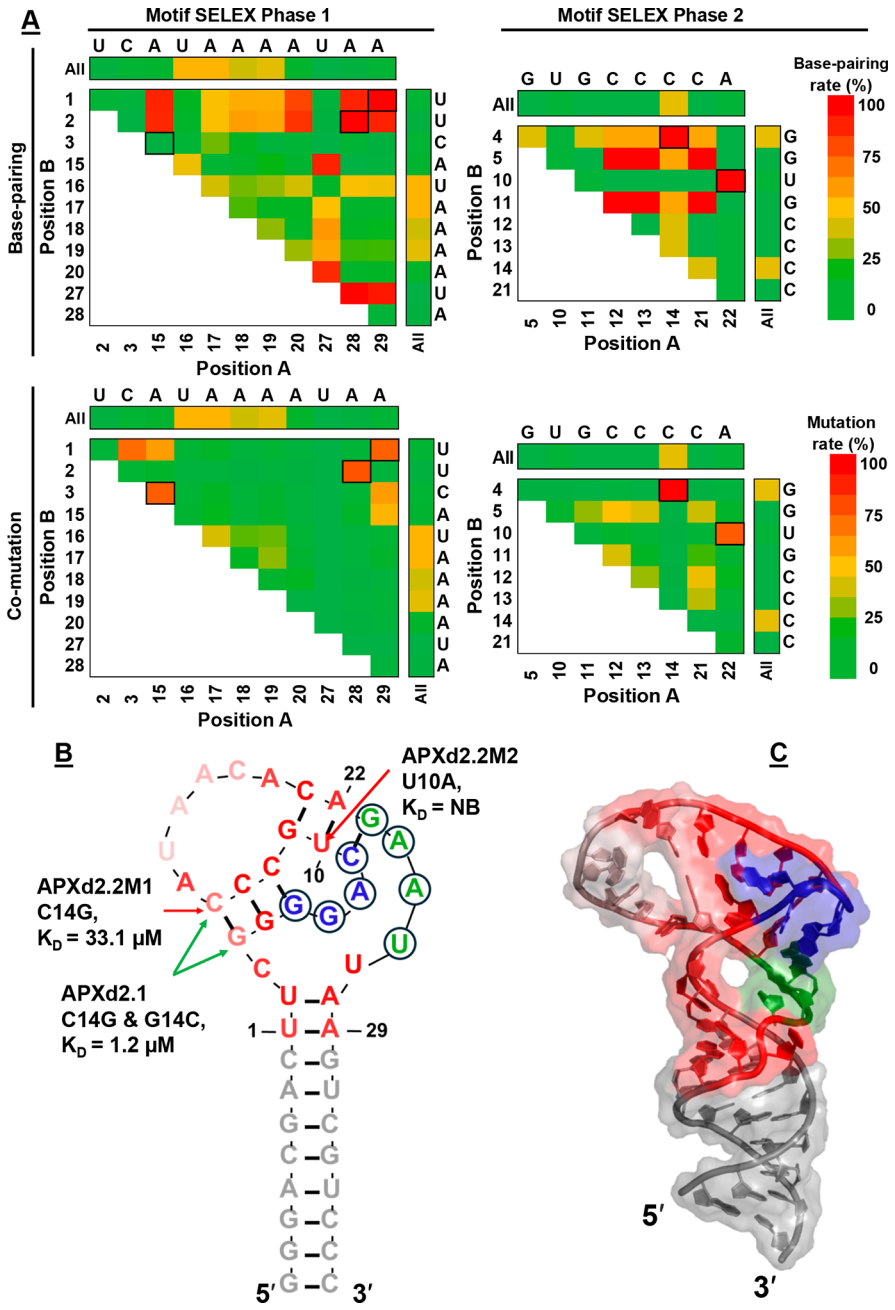


Figure 6. Optimized structure of the APX1 aptamers identified by motif-SELEX. (**A**) Base-pairing (top) and covariation (bottom) maps of the first (left) and second (right) motif-SELEX phases. The maps of phase 1 were constructed using sequences from the phase 1 Round 10 pool with V1−1, V2−1, and V3 lengths of 3, 6, and 3 nt, respectively. The maps for phase 2 were constructed using sequences from the phase 2 Round 10 pool with V1−2, V2−2A, and V2−2B lengths of 2, 5, and 2 nt, respectively. Position pairs with a covariation rate >75% are labeled with bold frames. Bars at top and right show the most common base at each position as the reference, with its overall mutation rate shown via the color scale. Predicted (**B**) secondary and (**C**) tertiary structures of the motif-SELEX optimized APXd2.2 aptamer. In both structures, the primer region is gray; C1 and C2 are blue and green, respectively; and V1−3 is red (bases with higher mutation rates are lighter). Aptamer tertiary structure was obtained with the RNA Composer software[26] treating the modified RNA aptamer as canonical RNA based on the HTS-predicted secondary structure.

(Figure 5B, motif library I); this library contains 5×10^8 unique sequences $((4^1 + 4^2 + 4^3) \times (4^4 + 4^6 + 4^8) \times (4^1 + 4^2 + 4^3)$, see Supporting Note 2 for detailed calculations). We performed library-immobilized SELEX against apixaban with this motif library to determine the most conserved sequence and length for each variable domain. We then designed a second motif library that incorporated constant sequences identified from this first experiment (V1−2, V2−2A, and V2−2B), and in which we randomized the remaining segments of domains V1 and V2 that had been kept constant in the first stage (Figure 5B, motif library II). This second library contains 1.2×10^8 unique sequences $((4^1 + 4^2 + 4^3) \times (4^3 + 4^5 + 4^7) \times (4^1 + 4^2 + 4^3)$, see Supporting Note 2 for detailed calculations). In this way, assuming that both sets of random domains are independent of each other, we were able to explore all of the sequence space in an efficient and practical manner. However, assuming complete codependency among

https://doi.org/10.1021/jacs.4c17041
J. Am. Chem. Soc. 2025, 147, 9472−9486

the bases in the randomized domains, this approach will only search 10^{-6} % of the sequence space.

We first assembled the library for the first motif-SELEX phase. Since the generation of an RNA library requires *in vitro* transcription by T7 polymerase, we assembled a double-stranded DNA template library by performing two Klenow reactions (SI, Figure S2). As with the HM1 motif library, HTS analysis of the APX1.1 motif library confirmed that the base composition and length distribution of the variable domains matched our randomized design (SI, Figure S13), demonstrating that motif libraries can be generated for RNA aptamers as well. We then used this motif library to perform 10 rounds of SELEX against apixaban (selection conditions shown in SI, Table S3) and used HTS to monitor the evolution process. The final apixaban pool was well enriched, with 14,929 unique sequences out of a total of 387,041 reads (3.9% unique), and the most abundant sequence only constituted 4.3% of the pool. Several combinations of lengths and sequences were highly enriched (Figure 5C−E), and domains V1−1, V2−1, and V3 had preferred lengths of 3, 6, and 3 nt, respectively. In Round 10, 73% of the sequences had UUC as domain V1−1 and 76% had UAA as domain V3. Domain V2−1 was more diverse, but 59% of the pool had a sequence with A at both ends (Figure 5F−H). Notably, we did not identify these preferred sequence elements in the original APX1 family. The most abundant sequence in the Round 10 pool, APXd2.2, is one nucleotide shorter than APX1.1 and has ∼3-fold higher target affinity as measured by ITC (K_D = 0.76 ± 0.03 μM) (SI, Figure S14A). These results offer further evidence that both the length and sequence of each variable domain are potentially important for target recognition, and that these parameters can be optimized using motif-SELEX.

For the second phase of apixaban motif-SELEX, we retained key features from aptamer APXd2.2 while also randomizing sequence elements that were not optimized in the first phase. Specifically, we randomized the 3′ region of V1 (V1−2) and the 5′ and 3′ ends of the V2 domain (V2−2A, and V2−2B, respectively) of APXd2.2 to generate the second motif library (Figure 5B, motif library II, and SI, Figure S3), and conducted 10 rounds of SELEX (conditions shown in SI, Table S4). Since this library only has 1.2×10^8 unique sequences, and thus full representation of all sequences is feasible, we skewed the ratio between different templates to include more copies of sequences with shorter V2−2A domains in the library to balance the quantity of shorter and longer sequences given the inherently higher abundance of the latter. The phase 2 Round 10 pool contained 15,409 unique sequences out of a total of 111,388 reads (13.8% unique), where sequences in V1−2, V2−2A, and V2−2B typically had lengths of 2, 5, and 2 nt, respectively (Figure 5I−K and SI, Figure S15), similar to APXd2.2. This result indicates that longer variable domains are not always preferred. The most abundant sequences for domains V1−2, V2−2A, and V2−2B were CG/GG, UGCGG/UGCGC, and CA, respectively (Figure 5L−N) again, very similar to APXd2.2. In fact, APXd2.2 was the second most abundant sequence in the Round 10 pool. The top sequence in the phase 2 Round 10 pool, APXd2.1, differs from APXd2.2 by two point-mutations but has a similar target affinity as characterized by ITC (K_D = 1.2 ± 0.1 μM) (SI, Figure S14B). These results indicate that the lengths and sequences of all variable domains in APXd2.2 are challenging to further improve.

Structure Prediction of an RNA Aptamer Guided by Motif-SELEX Data. We next used base-pairing and covariation maps to investigate the secondary structure of aptamers in the APX1 family by integrating data from both phases of motif-SELEX. For the first phase, we examined sequences with lengths of 3, 6, and 3 nt for V1−1, V2−1, and V3, which comprised 53.3% of the final round pool, and found that positions 1−29 and 2−28 are hot spots for both covariation and base-pairing (Figure 6A). This observation suggests that the constant stem formed by the primer region is preferentially extended by 2 bp. For the second phase, we examined sequences with lengths of 2, 5, and 2 nt for V1−2, V2−2A, and V2−2B, which comprised 69.3% of the phase 2 Round 10 pool, and found that positions 4−14 and 10−22 are hot spots for both covariation and base-pairing. Such a base-pairing pattern suggests a pseudoknot structure that is challenging to predict by common RNA secondary structure folding algorithms such as mfold (Figure 6B). The fact that few base positions demonstrated high comutation rates (and hence interdependency) indicates that the two-phase approach employed for the APX library indeed searched more than 10^{-6} % of the entire sequence space. Constrained by these preferred sequence patterns and the predicted secondary structure, we built a tertiary structure model of the optimized APX1 family sequence (Figure 6C). As with the HM1 family, consensus domains C1 and C2 are predicted to fold in adjacent to each other, forming a putative target binding site surrounded by variable domains that are more conserved (i.e., V1, V2−2A, V2−2B, and V3), while the more diverse V2−1 domain is distal to the consensus domains.

To assess the predicted pseudoknot structure of the aptamer, we generated two point mutants that disrupt either base-pairs 4−14 (APXd2.2M1) or 10−22 (APXd2.2M2) and found that the target affinity of these mutants is diminished or ablated (Figure 6B and SI, Figure S14C,D). In contrast, APXd2.1, a double mutant of APXd2.2 at positions 4 and 14 that restores base-pairing (G4-C14 to C4-G14) does not have impaired target-binding affinity. Our data also offers a potential explanation for why APX1.1 can bind apixaban even though it has different lengths and sequences for all variable domains compared to APXd2.1. This flexibility is because APX1.1 possesses the minimum sequence and structural requirements for binding: in particular, the U-A base pair between V1 and V3, three base-pairs between nucleotides 4−6 and 12−14, and three base-pairs between nucleotides 9−11 and 21−23. These results confirm that motif-SELEX can also be used to infer structural and sequence-activity relationship information for RNA aptamers.

DISCUSSION

Traditional SELEX methods based on fully randomized libraries often do not yield aptamer sequences with optimal target binding affinity in part due to the inability to include all possible sequences in starting pools. One common solution is to perform additional selection experiments using libraries comprising point mutants of an aptamer of interest, either via error-prone PCR or by doping random bases into the library during the chemical synthesis process. However, these methods are often ineffective due to their limited sequence coverage and biased sequence composition. In this work, we developed a new approach, motif-SELEX, to efficiently explore the sequence space around aptamers and discover variants with improved binding properties while also refining aptamer

https://doi.org/10.1021/jacs.4c17041
J. Am. Chem. Soc. 2025, 147, 9472−9486

structure prediction and elucidating sequence-activity relationships. Motif-SELEX uses RNA or DNA libraries that contain consensus sequence elements identified during initial aptamer selection, which are then separated by random domains that vary in both sequence and length. These more complex libraries can be constructed via the assembly of multiple single-stranded templates. Our motif libraries are highly diverse in terms of sequence and length. For instance, the HM1 motif library consisted of 1.29×10^{14} unique DNA sequences with minimal bias in sequence distribution and random region lengths ranging from 22−34 nt. This level of sequence randomness and diversity in length has not been reported previously with error-prone PCR or doped-SELEX and has allowed us to identify optimized aptamer sequences and make structural predictions based on information from those sequences.

We note that the utility and composition of motif libraries are different from previously reported structured SELEX libraries.[7] In terms of utility, structured libraries have been used to reduce the sequence space that needs to be explored in initial SELEX campaigns[32] and to identify aptamers that fold properly and function in biological environments.[20,33] In addition, the constant regions used in previously reported structured libraries are limited to a handful of simple, generic secondary structures, including double-stranded stems, three-way junctions, or guanine tracts that seed for G-quadruplexes.[7] In contrast, the constant regions used in motif libraries, which are discovered from previous selections for specific targets, include *in vitro* evolved, often noncanonical structures that directly contribute to aptamer-target binding. We note previous reports of screening motif-containing oligonucleotide libraries using microarray technology.[34,35] However, while microarrays offer the benefits of rapidity, simplicity, and screening without PCR bias, compared to motif-SELEX, they are limited in the ability to examine insertions due to length restrictions with microarray technology. Critically, microarrays also have a significantly lower capability to explore sequence space due to limits in their capacity (typically 10^6 sequences).

Motif-SELEX has many advantages compared to other reported affinity maturation approaches (SI, Table S9). One important advantage of motif libraries is that they introduce a new dimension of variation—the length of the variable domain—relative to traditional point-mutation libraries. Incorporating length as a variable in oligonucleotide libraries has been challenging for both error-prone PCR and chemical synthesis. Our new library generation method, which utilizes a handful of template strands, makes it possible to assemble libraries with multiple variable domains that each span a variety of lengths. Selection with these libraries allows one to discover improved aptamer sequences that could not otherwise be identified from the original SELEX library, point-mutation-based libraries, or structured libraries. Interestingly, the optimized aptamers that we discovered via motif-SELEX for morphine and apixaban had variable domain lengths of 34 and 29 nt, respectively, both of which differ from the original length of 30 nt. Motif-SELEX also allowed us to identify additional sequence and structure signatures important for target recognition. For example, in the HM1 family, we found that the GGTTC···AGTGC motif in the V2 domain was conserved regardless of the length of the domain. We also determined that complementarity between V1 and V3 is strongly preferred—independent of variation in the length and sequence of those domains—revealing a structural benefit from having additional base-pairs in the consensus stem.

Another advantage of motif libraries is that they can represent a greater number of point mutants compared to doped libraries. In traditional mutagenesis methods such as doped-SELEX, to preserve essential sequence elements of the original aptamer, a low mutation rate (<30%) is necessary. This restriction, however, results in the parent aptamer being highly abundant and greatly reduces the representation of unique, more heavily mutated sequences with potentially improved binding properties (see SI, Supporting Note 3 for details about sequence space calculation). In fact, one report demonstrated that a doped library can only confidently present all sequences with ≤4 point mutations relative to the original aptamer.[36] In contrast, motif libraries fully randomize the peripheral domains of conserved sequence regions in the aptamer, allowing for the inclusion of sequences with more than 20 point mutations relative to the original sequence using ∼300 pmol of starting library. Additionally, with the high diversity of the resulting motif libraries, we could identify conserved sequence and structural elements that are important for ligand binding.

Importantly, we found that the randomization of variable domains in our motif libraries was largely unbiased with respect to nucleotide identity, such that all sequences are represented with roughly equal frequency. This implies that the abundance of sequences in the late-stage motif-SELEX pools may roughly correlate to their target-binding affinity, at least at a logarithmic scale, as we observed in our experimental data (SI, Figure S10). The unbiased nature of motif-SELEX allows for easy identification of the optimal sequence, but also greatly improves the rigor of the resulting sequence-activity relationships. Based on simple HTS analysis for both morphine and apixaban aptamer families, we were able to identify several key sequences and structural features involved in target recognition. In addition, we could model the probable secondary and tertiary structure of the entire aptamer, confirming our predictions with mutagenesis studies and ITC data.

One limitation of motif-SELEX is that, unlike doped-SELEX or SELEX with error prone PCR, it requires prior knowledge of conserved motifs (SI, Table S9). Since HTS has become a widely accessible and affordable technology, we do not see this as a major impediment to the broad adoption of motif-SELEX. However, if in case motif information is lacking, one can still employ motif-SELEX by arbitrarily assigning variable regions and performing multiple phases of motif-SELEX until all regions of the sequence are surveyed. If the sequence being optimized happens to already be the optimal sequence, it will likely reemerge via motif-SELEX, as was the case with domain V2−2B of the apixaban aptamer. However, one important assumption that must be true for this "multiphasic" motif-SELEX approach to work is that the motifs identified from different phases are independent of each other. If multiple different motifs are dependent on each other, this approach is much less effective. For example, for motif-SELEX with the APX aptamers, if the motifs searched in the two phases of motif-SELEX were dependent on each other, then the sequence space could range anywhere between $\sim 10^8$ to $\sim 10^{14}$ sequences (see SI, Supporting Note 2 for more details).

Our results demonstrate that motif-SELEX is versatile and generalizable, and can be readily adapted to both DNA and RNA aptamers carrying different consensus and variable domains. In addition, our library assembly method allows for

https://doi.org/10.1021/jacs.4c17041
J. Am. Chem. Soc. 2025, 147, 9472−9486

considerable customizability in terms of the number and length of the consensus and variable domains. For instance, we were able to produce libraries for HM1 that incorporate three variable domains with lengths ranging from 0−19 nt interspersed by two consensus domains as short as 5 nt with our single-step library generation approach. For the even shorter consensus domains in the APX1 family, we used a two-phase motif-SELEX approach and successfully studied a family with only two 4-nt consensus domains. Although we have only demonstrated the optimization of aptamer affinity in this work, we anticipate that the same method can be applied for evaluating and optimizing other functional nucleic acid elements such as noncoding RNAs,[37] ribozymes/DNAzymes,[38,39] riboswitches,[40] and CRISPR-Cas guide RNAs.[10,41] And, because the design of motif libraries is independent of target identity, motif-SELEX should also be applicable to other targets like proteins.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/jacs.4c17041.

Sequence space calculations; sequences used in this work; motif-SELEX conditions employed for morphine and apixaban; ITC experimental conditions and data; HM1 family sequences identified from round 10 of the original morphine SELEX; sequences from the Round 10 motif-SELEX pool for morphine and point mutants of HMR10.2; top sequences in the APX1 aptamer family identified from the original apixaban SELEX; benefits and limitations of motif-SELEX and other aptamer affinity maturation approaches; scheme of the motif library generation processes; secondary structures predicted by mfold for HM1; selection progress of motif-SELEX for morphine; raw ITC data of aptamer binding affinity; circos plot of V1 and V3 domain sequences in the Round 10 HM1 motif-SELEX pool; correlation between frequency of sequences in the morphine motif-SELEX pool and their affinity; dye displacement assay; and characterization of the library for the first and second phase of motif-SELEX for apixaban (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

Bruce A Sullenger − *Department of Surgery, Duke University Medical Center, Durham, North Carolina 27710, United States*; Email: bruce.sullenger@duke.edu

Yi Xiao − *Department of Chemistry, North Carolina State University, Raleigh, North Carolina 27695, United States*; orcid.org/0000-0001-7278-9811; Email: yxiao34@ncsu.edu

Authors

Haixiang Yu − *Department of Surgery, Duke University Medical Center, Durham, North Carolina 27710, United States*; orcid.org/0000-0002-1100-9882

Juan Canoura − *Department of Chemistry, North Carolina State University, Raleigh, North Carolina 27695, United States*

Caleb Byrd − *Department of Chemistry, North Carolina State University, Raleigh, North Carolina 27695, United States*

Obtin Alkhamis − *Department of Chemistry, North Carolina State University, Raleigh, North Carolina 27695, United States*

Adara Bacon − *Department of Chemistry, North Carolina State University, Raleigh, North Carolina 27695, United States*

Amy Yan − *Department of Surgery, Duke University Medical Center, Durham, North Carolina 27710, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/jacs.4c17041

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported by the National Institute of Justice, Grant Office of Justice Programs, U.S. Department of Justice [2022-GG-04440-RESS], National Institutes of Health [R01DA51100 and P01 HL139420-06], and American Heart Association [23POST1018721].

■ REFERENCES

(1) Dunn, M. R.; Jimenez, R. M.; Chaput, J. C. Analysis of Aptamer Discovery and Technology. *Nat. Rev. Chem.* **2017**, *1*, No. 0076.

(2) Ellington, A. D.; Szostak, J. W. In Vitro Selection of RNA Molecules That Bind Specific Ligands. *Nature* **1990**, *346*, 818−822.

(3) Tuerk, C.; Gold, L. Systematic Evolution of Ligands by Exponential Enrichment: RNA Ligands to Bacteriophage T4 DNA Polymerase. *Science* **1990**, *249*, 505−510.

(4) Zhou, J.; Rossi, J. Aptamers as Targeted Therapeutics: Current Potential and Challenges. *Nat. Rev. Drug Discovery* **2017**, *16*, 181−202.

(5) Nimjee, S. M.; White, R. R.; Becker, R. C.; Sullenger, B. A. Aptamers as Therapeutics. *Annu. Rev. Pharmacol. Toxicol.* **2017**, *57*, 61−79.

(6) Yu, H.; Alkhamis, O.; Canoura, J.; Liu, Y.; Xiao, Y. Advances and Challenges in Small-Molecule DNA Aptamer Isolation, Characterization, and Sensor Development. *Angew. Chem., Int. Ed.* **2021**, *60*, 16800−16823.

(7) Brown, A.; Brill, J.; Amini, R.; Nurmi, C.; Li, Y. Development of Better Aptamers: Structured Library Approaches, Selection Methods, and Chemical Modifications. *Angew. Chem., Int. Ed.* **2024**, *63*, No. e202318665.

(8) Cadwell, R. C.; Joyce, G. F. Mutagenic PCR. *Genome Res.* **1994**, *3*, S136−S140.

(9) Autour, A.; Westhof, E.; Ryckelynck, M. iSpinach: A Fluorogenic RNA Aptamer Optimized for *in Vitro* Applications. *Nucleic Acids Res.* **2016**, *44*, 2491−2500.

(10) Bush, K.; Corsi, G. I.; Yan, A. C.; Haynes, K.; Layzer, J. M.; Zhou, J. H.; Llanga, T.; Gorodkin, J.; Sullenger, B. A. Utilizing Directed Evolution to Interrogate and Optimize CRISPR/Cas Guide RNA Scaffolds. *Cell Chem. Biol.* **2023**, *30*, 879−892.

(11) Mannironi, C.; Scerch, C.; Fruscoloni, P.; Tocchini-Valentini, G. P. Molecular Recognition of Amino Acids by RNA Aptamers: The Evolution into an L-Tyrosine Binder of a Dopamine-Binding RNA Motif. *RNA* **2000**, *6*, 520−527.

(12) Bartel, D. P.; Zapp, M. L.; Green, M. R.; Szostak, J. W. HIV-1 Rev Regulation Involves Recognition of Non-Watson-Crick Base Pairs in Viral RNA. *Cell* **1991**, *67*, 529−536.

(13) Lotz, T. S.; Halbritter, T.; Kaiser, C.; Rudolph, M. M.; Kraus, L.; Groher, F.; Steinwand, S.; Wachtveitl, J.; Heckel, A.; Suess, B. A Light-Responsive RNA Aptamer for an Azobenzene Derivative. *Nucleic Acids Res.* **2019**, *47* (4), 2029−2040.

(14) Lorsch, J. R.; Szostak, J. W. In Vitro Selection of RNA Aptamers Specific for Cyanocobalamin. *Biochemistry* **1994**, *33*, 973−982.

(15) Famulok, M. Molecular Recognition of Amino Acids by RNA-Aptamers: An L-Citrulline Binding RNA Motif and Its Evolution into an L-Arginine Binder. *J. Am. Chem. Soc.* **1994**, *116*, 1698−1706.
(16) Boussebayle, A.; Torka, D.; Ollivaud, S.; Braun, J.; Bofill-Bosch, C.; Dombrowski, M.; Groher, F.; Hamacher, K.; Suess, B. Next-Level Riboswitch Development-Implementation of Capture-SELEX Facilitates Identification of a New Synthetic Riboswitch. *Nucleic Acids Res.* **2019**, *47*, 4883−4895.
(17) Yang, K.; Mitchell, N. M.; Banerjee, S.; Cheng, Z.; Taylor, S.; Kostic, A. M.; Wong, I.; Sajjath, S.; Zhang, Y.; Stevens, J.; Mohan, S.; Landry, D. W.; Worgall, T. S.; Andrews, A. M.; Stojanovic, M. N. A Functional Group-Guided Approach to Aptamers for Small Molecules. *Science* **2023**, *380*, 942−948.
(18) Yang, K.; Alkhamis, O.; Canoura, J.; Bryant, A.; Gong, E. M.; Barbu, M.; Taylor, S.; Nikic, D.; Banerjee, S.; Xiao, Y.; Stojanovic, M. N.; Landry, D. W. Exploring the Landscape of Aptamers: From Cross-Reactive to Selective to Specific, High-Affinity Receptors for Cocaine. *JACS Au* **2024**, *4*, 760−770.
(19) Bittker, J. A.; Le, B. V.; Liu, D. R. Nucleic Acid Evolution and Minimization by Nonhomologous Random Recombination. *Nat. Biotechnol.* **2002**, *20*, 1024−1029.
(20) Dey, S. K.; Filonov, G. S.; Olarerin-George, A. O.; Jackson, B. T.; Finley, L. W. S.; Jaffrey, S. R. Repurposing an Adenine Riboswitch into a Fluorogenic Imaging and Sensing Tag. *Nat. Chem. Biol.* **2022**, *18*, 180−190.
(21) Yang, K. A.; Pei, R.; Stojanovic, M. N. In Vitro Selection and Amplification Protocols for Isolation of Aptameric Sensors for Small Molecules. *Methods* **2016**, *106*, 58−65.
(22) Yang, W.; Yu, H.; Alkhamis, O.; Liu, Y.; Canoura, J.; Fu, F.; Xiao, Y. In Vitro Isolation of Class-Specific Oligonucleotide-Based Small-Molecule Receptors. *Nucleic Acids Res.* **2019**, *47*, No. e71.
(23) Hoinka, J.; Backofen, R.; Przytycka, T. M. AptaSUITE: A Full-Featured Bioinformatics Framework for the Comprehensive Analysis of Aptamers from HT-SELEX Experiments. *Mol. Ther.–Nucleic Acids* **2018**, *11*, 515−517.
(24) Crooks, G. E.; Hon, G.; Chandonia, J. M.; Brenner, S. E. WebLogo: A Sequence Logo Generator. *Genome Res.* **2004**, *14*, 1188−1190.
(25) Krzywinski, M.; Schein, J.; Birol, I.; Connors, J.; Gascoyne, R.; Horsman, D.; Jones, S. J.; Marra, M. A. Circos: An Information Aesthetic for Comparative Genomics. *Genome Res.* **2009**, *19*, 1639−1645.
(26) Sarzynska, J.; Popenda, M.; Antczak, M.; Szachniuk, M. RNA Tertiary Structure Prediction Using RNAComposer in CASP15. *Proteins* **2023**, *91*, 1790−1799.
(27) Wang, L.; Canoura, J.; Byrd, C.; Nguyen, T.; Alkhamis, O.; Ly, P.; Xiao, Y. Examining the Relationship Between Aptamer Complexity and Molecular Discrimination of Low-Epitope Targets. *ACS Cent. Sci.* **2024**, *10*, 2213−2228.
(28) Canoura, J.; Alkhamis, O.; Venzke, M.; Ly, P. T.; Xiao, Y. Developing Aptamer-Based Colorimetric Opioid Tests. *JACS Au* **2024**, *4*, 1059−1072.
(29) Zuker, M. Mfold Web Server for Nucleic Acid Folding and Hybridization Prediction. *Nucleic Acids Res.* **2003**, *31*, 3406−3415.
(30) Canoura, J.; Alkhamis, O.; Byrd, C.; Wang, L.; Bryant, A.; Xiao, Y. Determining the Precision of High-Throughput Sequencing and Its Influence on Aptamer Selection. *Anal. Chem.* **2024**, *96*, 17720−17729.
(31) Cone, E. J.; Clarke, J.; Tsanaclis, L. Prevalence and Disposition of Drugs of Abuse and Opioid Treatment Drugs in Oral Fluid. *J. Anal. Toxicol.* **2007**, *31*, 424−433.
(32) Li, J.; Zhang, Z.; Gu, J.; Stacey, H. D.; Ang, J. C.; Capretta, A.; Filipe, C. D. M.; Mossman, K. L.; Balion, C.; Salena, B. J.; Yamamura, D.; Soleymani, L.; Miller, M. S.; Brennan, J. D.; Li, Y. Diverse High-Affinity DNA Aptamers for Wild-Type and B.1.1.7 SARS-CoV-2 Spike Proteins from a Pre-Structured DNA Library. *Nucleic Acids Res.* **2021**, *49*, 7267−7279.
(33) Porter, E. B.; Polaski, J. T.; Morck, M. M.; Batey, R. T. Recurrent RNA Motifs as Scaffolds for Genetically Encodable Small-Molecule Biosensors. *Nat. Chem. Biol.* **2017**, *13*, 295−301.
(34) Kinghorn, A. B.; Dirkzwager, R. M.; Liang, S.; Cheung, Y. W.; Fraser, L. A.; Shiu, S. C. C.; Tang, M. S. L.; Tanner, J. A. Aptamer Affinity Maturation by Resampling and Microarray Selection. *Anal. Chem.* **2016**, *88* (14), 6981−6985.
(35) Martin, J. A.; Mirau, P. A.; Chushak, Y.; Chávez, J. L.; Naik, R. R.; Hagen, J. A.; Kelley-Loughnane, N. Single-Round Patterned DNA Library Microarray Aptamer Lead Identification. *J. Anal. Methods Chem.* **2015**, *2015*, No. 137489.
(36) Svehlova, K.; Lukšan, O.; Jakubec, M.; Curtis, E. A. Supernova: A Deoxyribozyme That Catalyzes a Chemiluminescent Reaction. *Angew. Chem., Int. Ed.* **2022**, *61*, No. e202109347.
(37) Cech, T. R.; Steitz, J. A. The Noncoding RNA Revolution - Trashing Old Rules to Forge New Ones. *Cell* **2014**, *157*, 77−94.
(38) Long, M. B.; Jones, J. P.; Sullenger, B. A.; Byun, J. Ribozyme-Mediated Revision of RNA and DNA. *J. Clin. Invest.* **2003**, *112*, 312−318.
(39) Wang, Y.; Nguyen, K.; Spitale, R. C.; Chaput, J. C. A Biologically Stable DNAzyme That Efficiently Silences Gene Expression in Cells. *Nat. Chem.* **2021**, *13*, 319−326.
(40) Kavita, K.; Breaker, R. R. Discovering Riboswitches: The Past and the Future. *Trends Biochem. Sci.* **2023**, *48*, 119−141.
(41) Jinek, M.; Chylinski, K.; Fonfara, I.; Hauer, M.; Doudna, J. A.; Charpentier, E. A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science* **2012**, *337*, 816−821.

https://doi.org/10.1021/jacs.4c17041
J. Am. Chem. Soc. 2025, 147, 9472−9486

Supporting information for:

Improving Aptamer Affinity and Determining Sequence-Activity Relationships via Motif-SELEX

Haixiang Yu[1], Juan Canoura[2], Caleb Byrd[2], Obtin Alkhamis[2], Adara Bacon[2], Amy Yan[1], Bruce A Sullenger[1*] and Yi Xiao[2*]

[1]Department of Surgery, Duke University Medical Center, 2 Genome Ct., Durham, North Carolina, 27710, USA; [2]Department of Chemistry, North Carolina State University, 2620 Yarbrough Dr., Raleigh, North Carolina, 27695, USA.

*Corresponding author: bruce.sullenger@duke.edu and yxiao34@ncsu.edu

Supplementary notes. Sequence space calculation for motif-libraries and doped libraries.

Note 1. HM1 motif-library: The HM1 motif-library contains three variable domains (V1, V2, and V3). V1 and V3 can be 0–2 nt in length, while V2 can be 11, 13, 15, 17, or 19 nt. Consequently, the total sequence space of the library is given by:

$\sum_{L=0}^{2} 4^L \times \sum_{\substack{L\ odd,\\ 11\le L\le 19}} 4^L \times \sum_{L=0}^{2} 4^L = 129{,}302{,}444{,}113{,}920,$

which corresponds to 215 pmoles of individual library molecules.

Note 2. APX motif-libraries: The APX motif library also contains three variable domains (V1, V2, and V3). V1 can be 2–6 nt in length, V2 can be 8–18 nt in length, and V3 can be 1–3 nt in length. Thus, the total sequence space is:

$\sum_{L=2}^{6} 4^L \times \sum_{L=8}^{18} 4^L \times \sum_{L=1}^{3} 4^L = 41{,}951{,}539{,}740{,}278{,}784,$

equivalent to 69.7 nmoles of individual library molecules, which cannot be effectively represented by a single SELEX library. Consequently, we divided this sequence space into two dimensions and screened them consecutively using a two-phase SELEX procedure with two motif libraries.

In the first motif library, we randomized the 5' region of V1 (termed V1-1, length varied between 1–3 nt), the internal region of V2 (termed V2-1, length varied between 4, 6, and 8 nt), and the entire V3 domain (length varied between 1–3 nt). This library provided a sequence dimension of:

$\sum_{L=1}^{3} 4^L \times \sum_{\substack{L\ even,\\ 4\le L\le 8}} 4^L \times \sum_{L=1}^{3} 4^L = 458{,}535{,}168,$

equivalent to 0.76 fmoles of individual library molecules, thoroughly surveyed in the phase 1 motif SELEX.

In the second phase motif library, we fixed the optimal sequence of V1-1, V2-1, and V3 to survey the second sequence dimension, including the 3′ region of V1 (V1-2) and the 5′ and 3′ ends of the V2 domain (V2-2A and V2-2B, respectively). V1-2 and V2-2B varied between 1–3 nt, and V2-2B varied between 3, 5, and 7 nt, creating a sequence dimension of:

$\sum_{L=1}^{3} 4^L \times \sum_{\substack{L\ odd,\\ 3\le L\le 7}} 4^L \times \sum_{L=1}^{3} 4^L = 114{,}633{,}792,$

equivalent to 0.19 fmoles of individual library molecules, thoroughly surveyed in the phase 2 motif SELEX.

Ideally, the two dimensions surveyed by phase-1 and phase-2 selection are completely independent; the optimal sequence in domains V1-1, V2-1, and V3 remains constant regardless of changes in domains V1-2, V2-2A, and V2-2B, and vice versa. In this ideal case, the actual sequence space searched via the two motif-SELEX phases is:

$(\sum_{L=1}^{3} 4^L \times \sum_{\substack{L\ even,\\ 4\le L\le 8}} 4^L \times \sum_{L=1}^{3} 4^L) \times (\sum_{L=1}^{3} 4^L \times \sum_{L\ odd,\ 3\le L\le 7} 4^L \times \sum_{L=1}^{3} 4^L) =$

52,563,625,073,197,056, covering the whole possible sequence space generated by V1, V2, and V3.

However, the two sequence dimensions are not fully independent due to possible overlap between domains V1-1 and V1-2, and domains V2-1, V2-2A, and V2-2B. Additionally, cooperativity between different sequence domains would reduce the independence between the two sequence dimensions.

In the unlikely worst-case scenario where the two sequence dimensions are completely dependent, each sequence in the first dimension would have a unique optimal sequence in the second dimension. In this extreme case, the sequence space searched via the two motif-SELEX phases is reduced to:

$$\left(\sum_{L=1}^{3} 4^L \times \sum_{\substack{L\ even,\\ 4\le L\le 8}} 4^L \times \sum_{L=1}^{3} 4^L\right) + \left(\sum_{L=1}^{3} 4^L \times \sum_{L\ odd,\ 3\le L\le 7} 4^L \times \sum_{L=1}^{3} 4^L\right) - 1 = 616{,}412{,}159$$

Therefore, the sequence space searched by the two-phase APX motif-SELEX was between 616,412,159 and 41,951,539,740,278,784, depending on the independency between the two sequence dimensions.

We note that based on data in Figure 4B and Figure 6A in the manuscript, interdependencies between different randomized domains (which are reflected by a high co-mutation rate) was observed for a very limited number of nucleotides, implying that a majority of the base positions are independent of each other. Thus, between the possibility of screening 0.000001% versus 100% of the sequence space, we infer that we screened somewhere in between these two figures.

Note 3. Doped libraries: For doped-SELEX libraries, the theoretical number of unique sequences in the sequence space is 4^n, where n is the number of randomized nucleotides. A sequence with k mutations in a doped library where the randomization rate is p occurs with a frequency of $(1-p)^{n-k} \cdot \left(\frac{p}{3}\right)^k$, such that there will be $C\ (n,k) \cdot\ 3^k$ sequences with k mutations. For a doped library with 30 nucleotides where each nucleotide is randomized at a rate of 30% (*i.e.*, 70% chance of retaining the original nucleotide), there are 4^{30} or ~10^{18} possible unique sequences. Doped libraries, however, face three problems in terms of traversing this sequence space.

First, the frequency with which any particular sequence with >12 mutations will appear is lower than 1 in 10^{15}. Thus, in 1 nmol of a doped library, it is highly unlikely that each of these mutants will be present.

Second, the parent aptamer is theoretically present at a copy number of 1.4×10^{10}, or 0.0023% of the library. For any given 5-nt mutant, the original sequence outnumbers the mutant by a factor of 1.7×10^4. For a 10-nt mutant, the original sequence outnumbers it by a staggering 2.8×10^8-fold. Because of this disparity in abundance, the parent aptamer and closely-related sequences are much more likely to become enriched and dominate the selection process.

Third, doped libraries only encompass a narrow search window in terms of the number of mutations explored. For the hypothetical doped library discussed above, 93% of the pool would consist of sequences with 5–13 mutations relative to the parent sequence.

Table S1. All sequences used in this work.

Sequence ID	Sequence (5'→3')
HM1	GGGACGACGGGATTCGGATCGTGGAACAGTGCCTCGTCGTCGTCCC
HMR10.2	GGGACGACGGGGATTCGGTTCGTGTCTAACAGTGCCTCGTACGTCGTCCC
HMR10.2 M1	GGGACGACAGGGATTCGGTTCGTGTCTAACAGTGCCTCGTACGTCGTCCC
HMR10.2 M2	GGGACGACGGGGATTCAGTTCGTGTCTAACAGTGCCTCGTACGTCGTCCC
HMR10.2 M3	GGGACGACGGGGATTCGGTTCGTGTCTAACAGTGGCTCGTACGTCGTCCC
HMR10.2 M4	GGGACGACGGGGATTCGGCTCGTGTCTAACAGTGCCTCGTACGTCGTCCC
HMR10.2 M5	GGGACGACGGGGATTCGGTTCGTGTCTAAGAGTGCCTCGTACGTCGTCCC
HMR10.2 M6	GGGACGACGGGGATTCGGTTCGCGTCTAACAGTGCCTCGTACGTCGTCCC
HMR10.2 M7	GGGACGACGGGGATTCGGTTCGCGTCTAGCAGTGCCTCGTACGTCGTCCC
HMR10.2 M8	GGGACGACGGGGATTCGGTTCGTGATCAACAGTGCCTCGTACGTCGTCCC
HMR10.1	GGGACGACGGGGATTCGGTTCTCGATAAGTAGTGCCTCGTACGTCGTCCC
HMR10.3	GGGACGACGGGGATTCGGTTCGCGATGAGCAGTGCCTCGTACGTCGTCCC
HMR10.4	GGGACGACGGGGATTCGGTTCGTGTTCAGCAGTGCCTCGTCCGTCGTCCC
HMR10.5	GGGACGACGGGGATTCGGTTCGTGATTAGCAGTGCCTCGTCCGTCGTCCC
HMR10.6	GGGACGACAGGGATTCGGTTCGTGGTTAACAGTGCCTCGTCTGTCGTCCC
HMR10.7	GGGACGACGGGGATTCGCGGATTCGGAGAAAAGGGCTCGTACGTCGTCCC
HMR10.8	GGGACGACAGGGATTCGGTCTTGGTACATAGGGGCCTCGTCTGTCGTCCC
HMR10.9	GGGACGACGGGGATTCGGTCTTGGTACATAGGGGCCTCGTACGTCGTCCC
HMR10.10	GGGACGACAGGGATTCGGTCTTGGAACAAAGGGGCCTCGTCTGTCGTCCC
HMR10.11	GGGACGACGGGGATTCGCGGATTCATAGAAAAGGGCTCGTCCGTCGTCCC
HMR10.12	GGGACGACCAGGATTCGGTTCTTTGTAAATAGTGCCTCGTTGGTCGTCCC
HMR10.14	GGGACGACAGGGATTCGGTTCTCGATAAGTAGTGCCTCGTACGTCGTCCC
HMR10.15	GGGACGACAGGGATTCGCGGATTCGGAGAAAAGGGCTCGTACGTCGTCCC
HMR10.16	GGGACGACGGGGATTCGCGGATTCATGGAAAAGGGCTCGTCTGTCGTCCC
APX1.1	GGGACGACUUGGGGACUGCCCGCCACAACACAGAAUUAGUCGUCCC
APXd2.1	GGGACGACUUCCGGGACUGCCGAUAACACAGAAUUAAGUCGUCCC
APXd2.2	GGGACGACUUCGGGGACUGCCCAUAACACAGAAUUAAGUCGUCCC
APXd2.2M1	GGGACGACUUCGGGGACUGCCGAUAACACAGAAUUAAGUCGUCCC
APXd2.2M2	GGGACGACUUCGGGGACAGCCCAUAACACAGAAUUAAGUCGUCCC
APXd2.2M3	GGGACGACUUUGGGGACUGCCCAUAACACAGAAUCAAGUCGUCCC
F1	GGAGGCTCTCGGGACGACGGATTC
F2	GGAGGCTCTCGGGACGACNGGATTC
F3	GGAGGCTCTCGGGACGACNNGGATTC
M1	GGGACGACNGGATTCNNNNNNNNNNNCTCGTNGTCGTCCCGCCTTT/3BioTEG/
M2	GGGACGACNGGATTCNNNNNNNNNNNNNCTCGTNGTCGTCCCGCCTTT/3BioTEG/
M3	GGGACGACNGGATTCNNNNNNNNNNNNNNNCTCGTNGTCGTCCCGCCTTT/3BioTEG/
M4	GGGACGACNGGATTCNNNNNNNNNNNNNNNNNCTCGTNGTCGTCCCGCCTTT/3BioTEG/
M5	GGGACGACNGGATTCNNNNNNNNNNNNNNNNNNNCTCGTNGTCGTCCCGCCTTT/3BioTEG/
R1	GAGCACAGCAGGGCGGAAATCCTAAATGTC/3BioTEG/
R2	GAGCANCAGCAGGGCGGAAATCCTAAATGTC/3BioTEG/
R3	GAGCANNCAGCAGGGCGGAAATCCTAAATGTC/3BioTEG/
HM1-cDNA	GTCGTCCCGAGAGCCATA/3BioTEG/
HM1-FP	GGAGGCTCTCGGGACGAC
HM1-RP	/5BiotinTEG/CTGTAAATCCTAAAGGCGGGACGAC
APX-F1-1	GGGGGAATTCTAATACGACTCACTATAGGAGGCTCTCGGGACGACNGGGGACTGCCC
APX-F1-2	GGGGGAATTCTAATACGACTCACTATAGGAGGCTCTCGGGACGACNNGGGGACTGCCC

APX-F1-3	GGGGGAATTCTAATACGACTCACTATAGGAGGCTCTCGGGACGACNNNGGGGACTGCCC
APX-M1-1	/5BiotinTEG/GGGGACTGCCCNNNNCAGAATNNGTCGTCCCGCCTTTAGGATTTACAG
APX-M1-2	/5BiotinTEG/GGGGACTGCCCNNNNNNCAGAATNNGTCGTCCCGCCTTTAGGATTTACAG
APX-M1-3	/5BiotinTEG/GGGGACTGCCCNNNNNNNNCAGAATNNGTCGTCCCGCCTTTAGGATTTACAG
APX-R1-1	CTGTAAATCCTAAAGGCGGGACGACNATTCTG
APX-R1-2	CTGTAAATCCTAAAGGCGGGACGACNNATTCTG
APX-R1-3	CTGTAAATCCTAAAGGCGGGACGACNNNATTCTG
APX-F2-1	GGGGGAATTCTAATACGACTCACTATAGGAGGCTCTCGGGACGACTTCNGGACNNN
APX-F2-2	GGGGGAATTCTAATACGACTCACTATAGGAGGCTCTCGGGACGACTTCNNGGACNNN
APX-F2-3	GGGGGAATTCTAATACGACTCACTATAGGAGGCTCTCGGGACGACTTCNNNGGACNNN
APX-M2-1	/5BiotinTEG/NTGTTATNNNGTCCNNGAAGTCGTCCCGAGAGCCTCC
APX-M2-2	/5BiotinTEG/NTGTTATNNNNNGTCCNNGAAGTCGTCCCGAGAGCCTCC
APX-M2-3	/5BiotinTEG/NTGTTATNNNNNNNGTCCNNGAAGTCGTCCCGAGAGCCTCC
APX-R2-1	CTGTAAATCCTAAAGGCGGGACGACTTAATTCNTGTTAT
APX-R2-2	CTGTAAATCCTAAAGGCGGGACGACTTAATTCNNTGTTAT
APX-R2-3	CTGTAAATCCTAAAGGCGGGACGACTTAATTCNNNTGTTAT
APX-cDNA	GTCGTCCCGAGAGCCTATA/3BioTEG/
APX-FP	GGGGGAATTCTAATACGACTCACTATAGGAGGCTCTCGGGACGAC
APX-RP	CCCTGTAAATCCTAAAGGCGGGACGAC

/3BioTEG/ and /5BiotinTEG/ are biotinylation modifications with TEG linkers on 3' and 5' termini, respectively. For APX1.1, APXd2.1, APXd2.2, APXd2.2M1, APXd2.2M2, and APXd2.2M3, all A, C, and U are 2' O-Methyl modified, and G are 2' F modified.

Table S2. Motif-SELEX conditions employed for morphine with the HM1 family.

Selection round	Pool size (pmole)	Buffer wash 1	Counter-target wash	Buffer wash 2	[Target] (μM)
1	250	10	NA	NA	100
2	250	20	100 μM QUI×3, 100 μM DPH×3, 100 μM LID×3, 100 μM PRO×3	20	100
3	250	30	200 μM QUI×3, 200 μM DPH×3, 200 μM LID×3, 200 μM PRO×3, 250 μM TYR×3, 250 μM PHE×3, 250 μM GLY×3, 100 μM GABA×3,100 μM EPI×3, 100 μM NEP×3, 100 μM SER×3, 100 μM DOP×3	30	100
4	250	40	200 μM QUI×3, 200 μM DPH×3, 200 μM LID×3, 200 μM PRO×3, 250 μM TYR×3, 250 μM PHE×3, 250 μM GLY×3, 200 μM GABA×3, 200 μM EPI×3, 200 μM NEP×3, 200 μM SER×3, 200 μM DOP×3, 100 μM NOL×3, 100 μM NAL×3, 100 μM NMOR×3, 100 μM M3G×3	40	100
5	250	40	200 μM QUI×3, 200 μM DPH×3, 200 μM LID×3, 200 μM PRO×3, 250 μM TYR×3, 250 μM PHE×3, 250 μM GLY×3, 200 μM GABA×3, 200 μM EPI×3, 200 μM NEP×3, 200 μM SER×3, 200 μM DOP×3, 100 μM NOL×3, 100 μM NAL×3, 100 μM NMOR×3, 100 μM M3G×3, 50 μM M6G×3	40	100
6	250	40	200 μM QUI×3, 200 μM DPH×3, 200 μM LID×3, 200 μM PRO×3, 200 μM TYR×3, 250 μM PHE×3, 250 μM GLY×3, 100 μM NOL×3, 100 μM NAL×3, 100 μM NMOR×3, 100 μM M3G×3,100 μM M6G×3, 200 μM GABA×3, 200 μM EPI×3, 200 μM NEP×3, 200 μM DOP×3, 200 μM SER×3	40	100
7	250	40	Same as round 6	40	100
8	250	40	Same as round 6	40	25
9	250	40	Same as round 6	40	10
10	250	40	Same as round 6	40	5

Counter-targets used include quinine (QUI), diphenhydramine (DPH), lidocaine (LID), procaine (PRO), tyrosine (TYR), phenylalanine (PHE), glycine (GLY), naloxone (NOL), naltrexone (NAL), normorphine (NMOR), morphine-3-glucuronide (M3G), morphine-6-glucuronide (M6G), gamma-aminobutyric acid (GABA), epinephrine (EPI), norepinephrine (NEP), dopamine (DOP), and serotonin (SER). The volume of each buffer and counter-target wash was 250 μL. The volume of target wash was 250 μL ×3.

Table S3. Phase 1 motif-SELEX conditions for apixaban with APX1 family.

Selection round	Pool size (pmole)	Buffer wash	Target wash volume (µL)	[Target] (µM)
1	250	20	500	100
2	250	20	500	100
3	250	20	500	25
4	250	20	375	10
5	250	20	250	10
6	250	20	250	5
7	250	20	375	2.5
8	250	20	250	2
9	250	40	500	2
10	250	40	500	1

The volume of each buffer wash was 250 µL.

Table S4. Phase 2 motif-SELEX conditions for apixaban with APX1 family.

Selection round	Pool size (pmole)	Buffer wash	Target wash volume (µL)	[Target] (µM)
1	250	20	750	100
2	250	20	750	100
3	250	20	500	50
4	250	20	500	50
5	250	30	500	25
6	250	30	750	100
7	250	30	500	50
8	250	30	500	50
9	250	30	500	25
10	250	30	750	25

The volume of each buffer wash was 250 µL.

Table S5. ITC experimental conditions and data for sequences evaluated in this work.

Aptamer ID	[Ligand] (μM)	[Aptamer] (μM)	K_D (μM)	ΔH (kcal mol^{-1})	ΔS (cal mol^{-1} K^{-1})
HM1	300	30	12.6 ± 0.20	-29.0 ± 0.3	-75.5
HMR10.2	400	30	2.02 ± 0.05	-26.0 ± 0.1	-61.8
HMR10.2 M1	1,000	30	17.8 ± 0.40	-30.6 ± 0.8	-81.7
HMR10.2 M2	4,000	50	>500	NA	NA
HMR10.2 M3	4,000	50	645 ± 31	-9.5 ± 2.6	-17.4
HMR10.2 M4	4,000	50	160 ± 2.4	-42.3 ± 3.7	-125.6
HMR10.2 M5	4,000	50	328 ± 16	-29.5 ± 12.2	-83.8
HMR10.2 M6	1,000	30	12.8 ± 0.20	-37.7 ± 0.4	-104.7
HMR10.2 M7	400	30	1.42 ± 0.02	-26.0 ± 0.1	-61.1
HMR10.2 M8	300	30	1.30 ± 0.04	-29.4 ± 0.1	-72.2
HMR10.1	400	30	1.75 ± 0.04	-26.1 ± 0.1	-61.8
HMR10.3	400	30	1.59 ± 0.03	-24.9 ± 0.1	-57.4
HMR10.4	400	30	1.20 ± 0.03	-25.4 ± 0.1	-58.8
HMR10.5	400	30	1.01 ± 0.03	-24.4 ± 0.1	-54.7
HMR10.6	400	30	1.15 ± 0.03	-26.9 ± 0.1	-63.5
HMR10.7	300	30	2.30 ± 0.06	-28.3 ± 0.1	-69.8
HMR10.8	300	30	1.10 ± 0.03	-30.2 ± 0.1	-74.4
HMR10.9	300	30	1.40 ± 0.04	-28.0 ± 0.1	-67.6
HMR10.10	300	30	1.10 ± 0.02	-29.4 ± 0.1	-71.9
HMR10.11	300	30	2.50 ± 0.06	-29.3 ± 0.2	-73.4
HMR10.12	300	30	1.80 ± 0.04	-35.0 ± 0.2	-91.8
HMR10.14	800	20	13.3 ± 0.28	-28.1 ± 0.7	-72.6
HMR10.15	800	20	22.2 ± 0.69	-26.9 ± 1.3	-69.6
HMR10.16	800	20	20.6 ± 0.40	-26.5 ± 0.8	-68.2
APX1.1	400	30	2.26 ± 0.09	-21.4 ± 0.2	-46.1
APXd2.1	400	20	1.22 ± 0.07	-17.9 ± 0.2	-33.4
APXd2.2	400	30	0.76 ± 0.03	-17.7 ± 0.1	-31.9
APXd2.2M1	400	30	33.1 ± 6.20	-3.6 ± 0.3	8.1
APXd2.2M2	400	30	No binding	NA	NA

Table S6. HM1 family sequences identified from round 10 of the original morphine SELEX (***JACS Au*** 2024, 4, 1059–1072).

Sequence ID	Sequence (5'→3')							Frequency (CPM)
	Primer	V1	C1	V2	C2	V3	Primer	
HM1.26	GGGACGAC	G	GGATTC	GGATCATGGAACAGTGC	CTCGT	C	GTCGTCCC	**70**
HM1.27	GGGACGAC	G	GGATTC	GGATCGAGGAACAGTGC	CTCGT	C	GTCGTCCC	**70**
HM1.28	GGGACGAC	T	GGATTC	GGATCGTGGAACAGTGC	CTCGT	C	GTCGTCCC	**70**
HM1.29	GGGACGAC	G	GGATTC	GGCTCGTGGAACAGTGC	CTCGT	C	GTCGTCCC	**62**
HM1.30	GGGACGAC	G	GGATTC	GGATCGTGGAATAGTGC	CTCGT	C	GTCGTCCC	**62**
HM1.31	GGGACGAC	G	GGATTC	GGATCGTTGAACAGTGC	CTCGT	C	GTCGTCCC	**62**
HM1.32	GGGACGAC	G	GGATTC	GGATCGTGTAACAGTGC	CTCGT	C	GTCGTCCC	**62**
HM1.33	GGGACGAC	G	GGATTC	GGTTCGTGGAACAGTGC	CTCGT	C	GTCGTCCC	**62**
HM1.34	GGGACGAC		GGATTC	GCGCCCCAAGTGGGGTGGG	CTCGT		GTCGTCCC	**58**
HM1.35	GGGACGAC		GGATTC	GCGCCCCAAGCGGGGAGGG	CTCGT		GTCGTCCC	**54**
HM1.36	GGGACGAC	A	GGATTC	GGATCGTGGAACAGTGC	CTCGT	C	GTCGTCCC	**54**
HM1.37	GGGACGAC	G	GGATTC	GGATCGTAGAACAGTGC	CTCGT	C	GTCGTCCC	**54**
HM1.38	GGGACGAC		GGATTC	GCGCCCCAGGTGGGGAGGG	CTCGT		GTCGTCCC	**51**
HM1.39	GGGACGAC		GGATTC	GCGCCCCTAGTGGGGAGGG	CTCGT		GTCGTCCC	**51**
HM1.40	GGGACGAC	G	GGATTC	GGATCGTGGTACAGTGC	CTCGT	C	GTCGTCCC	**51**
HM1.41	GGGACGAC	G	GGATTC	TGATCGTGGAACAGTGC	CTCGT	C	GTCGTCCC	**51**
HM1.42	GGGACGAC	G	GGATTC	GGATCGTGGAACAGTAC	CTCGT	C	GTCGTCCC	**47**
HM1.43	GGGACGAC	G	GGATTC	GGATCGTGGAACAGTGA	CTCGT	C	GTCGTCCC	**47**
HM1.44	GGGACGAC		GGATTC	GCGCCCCCAGTGGGGAGGG	CTCGT		GTCGTCCC	**43**
HM1.45	GGGACGAC	G	GGATTC	GGATCGTGAAACAGTGC	CTCGT	C	GTCGTCCC	**43**
HM1.46	GGGACGAC	G	GGATTC	GGATTGTGGAACAGTGC	CTCGT	C	GTCGTCCC	**43**
HM1.47	GGGACGAC	G	GGATTC	GGTTCGCTTAGCAGTGC	CTCGT	T	GTCGTCCC	**39**
HM1.48	GGGACGAC	AG	GGATTC	GCGCTGTTAAGAGGG	CTCGT	CT	GTCGTCCC	**39**
HM1.49	GGGACGAC	G	GGATTC	GGATAGTGGAACAGTGC	CTCGT	C	GTCGTCCC	**35**
HM1.50	GGGACGAC		GGATTC	GCGCCCAAAGTGGGGAGGG	CTCGT		GTCGTCCC	**31**
HM1.51	GGGACGAC		GGATTC	GCGCCCCAAGTGTGGAGGG	CTCGT		GTCGTCCC	**31**
HM1.52	GGGACGAC		GGATTC	GCGCCCCAATTGGGGAGGG	CTCGT		GTCGTCCC	**31**
HM1.53	GGGACGAC		GGATTC	GCGCCACAAGTGGGGAGGG	CTCGT		GTCGTCCC	**27**
HM1.54	GGGACGAC		GGATTC	GCGCCCCAAGTGGGAAGGG	CTCGT		GTCGTCCC	**27**
HM1.55	GGGACGAC	G	GGATTC	GGAGCGTGGAACAGTGC	CTCGT	C	GTCGTCCC	**27**
HM1.56	GGGACGAC		GGATTC	GTGCCCCAAGTGGGGAGGG	CTCGT		GTCGTCCC	**23**
HM1.57	GGGACGAC		GGATTC	ACGCCCCAAGTGGGGAGGG	CTCGT		GTCGTCCC	**23**
HM1.58	GGGACGAC		GGATTC	GCGCTCCAAGTGGGGAGGG	CTCGT		GTCGTCCC	**23**
HM1.59	GGGACGAC		GGATTC	GCGACCCAAGTGGGGAGGG	CTCGT		GTCGTCCC	**23**
HM1.60	GGGACGAC	G	GGATTC	AGATCGTGGAACAGTGC	CTCGT	C	GTCGTCCC	**23**

HM1.61	GGGACGAC		GGATTC	GCGCCCCAAGTGGGGAGTG	CTCGT		GTCGTCCC	**19**
HM1.62	GGGACGAC	G	GGATTC	GGATCGTGGAAAAGTGC	CTCGT	C	GTCGTCCC	**19**
HM1.63	GGGACGAC		GGATTC	GCGCCCTAAGTGGGGAGGG	CTCGT		GTCGTCCC	**16**
HM1.64	GGGACGAC		GGATTC	GCGCCCCAAGTGGGTAGGG	CTCGT		GTCGTCCC	**16**
HM1.65	GGGACGAC		GGATTC	GCGCCCCAACTGGGGAGGG	CTCGT		GTCGTCCC	**16**
HM1.66	GGGACGAC		GGATTC	GCGCCCCAAGTGGGGAAGG	CTCGT		GTCGTCCC	**16**
HM1.67	GGGACGAC		GGATTC	GCGCACCAAGTGGGGAGGG	CTCGT		GTCGTCCC	**16**
HM1.68	GGGACGAC		GGATTC	GCACCCCAAGTGGGGAGGG	CTCGT		GTCGTCCC	**16**
HM1.69	GGGACGAC		GGATTC	GCGCCCCATGAGGGGAGGG	CTCGT		GTCGTCCC	**16**
HM1.70	GGGACGAC		GGATTC	GCGCCCCAAGTGGTGAGGG	CTCGT		GTCGTCCC	**12**
HM1.71	GGGACGAC		GGATTC	TCGCCCCAAGTGGGGAGGG	CTCGT		GTCGTCCC	**12**
HM1.72	GGGACGAC		GGATTC	GCGGCCCAAGTGGGGAGGG	CTCGT		GTCGTCCC	**12**
HM1.73	GGGACGAC		GGATTC	GCGCCCCATGTGGGGTGGG	CTCGT		GTCGTCCC	**12**
HM1.74	GGGACGAC		GGATTC	GCGCCCCAAGTGGAGAGGG	CTCGT		GTCGTCCC	**8**
HM1.75	GGGACGAC		GGATTC	GCGCCCCAAGTTGGGAGGG	CTCGT		GTCGTCCC	**8**
HM1.76	GGGACGAC		GGATTC	GCGCCCCAAGTAGGGAGGG	CTCGT		GTCGTCCC	**8**
HM1.77	GGGACGAC		GGATTC	GCGCCCCAAGTGGGGAGGA	CTCGT		GTCGTCCC	**8**
HM1.78	GGGACGAC		GGATTC	GCGCCCCAAGTGGGGAGGT	CTCGT		GTCGTCCC	**8**
HM1.79	GGGACGAC		GGATTC	GCGCCCCAAGTGGGGATGG	CTCGT		GTCGTCCC	**8**
HM1.80	GGGACGAC		GGATTC	GCGCCCCAAGTGAGGAGGG	CTCGT		GTCGTCCC	**8**
HM1.81	GGGACGAC		GGATTC	CCGCCCCAAGTGGGGAGGG	CTCGT		GTCGTCCC	**8**
HM1.82	GGGACGAC		GGATTC	GCGCCCCAAGTGGCGAGGG	CTCGT		GTCGTCCC	**8**
HM1.83	GGGACGAC		GGATTC	GCGCCCCACGTGGGGTGGG	CTCGT		GTCGTCCC	**8**
HM1.84	GGGACGAC		GGATTC	GCGCCGCAAGTGGGGAGGG	CTCGT		GTCGTCCC	**8**
HM1.85	GGGACGAC	A	GGATTC	GGATCGCTTAGCAGTGC	CTCGT	C	GTCGTCCC	**8**
HM1.86	GGGACGAC		GGATTC	GCGCCCCAGGTTGGGAGGG	CTCGT		GTCGTCCC	**4**
HM1.87	GGGACGAC		GGATTC	GCTCCCCAAGTGGGGAGTG	CTCGT		GTCGTCCC	**4**
HM1.88	GGGACGAC		GGATTC	GCGCCCCAAGTGTGGCGGG	CTCGT		GTCGTCCC	**4**
HM1.89	GGGACGAC		GGATTC	GCGCCCCTTGTGGGGAGGG	CTCGT		GTCGTCCC	**4**
HM1.90	GGGACGAC		GGATTC	GCGCGCCAAGTGGGGAGGG	CTCGT		GTCGTCCC	**4**
HM1.91	GGGACGAC		GGATTC	GCGCCCCAAGTGGGGCGGG	CTCGT		GTCGTCCC	**4**
HM1.92	GGGACGAC		GGATTC	GCGCCCCAAGTGCGTAGTG	CTCGT		GTCGTCCC	**4**
HM1.93	GGGACGAC		GGATTC	GCGCCCCAAGTGGGGAGAG	CTCGT		GTCGTCCC	**4**

Table S7. Sequences from the HM1 Round 10 motif-SELEX pool and point mutants of HMR10.2. Mutation sites in HMR10.2 M1–8 relative to HMR10.2 are underlined.

Sequence ID	Sequence (5'→3')							Frequency (CPM)
	Primer	V1	C1	V2	C2	V3	Primer	
HM1	GGGACGAC	G_	GGATTC	GGATCGTGG__AACAGTGC	CTCGT	_C	GTCGTCCC	**0**
HMR10.2	GGGACGAC	GG	GGATTC	GGTTCGTGTCTAACAGTGC	CTCGT	AC	GTCGTCCC	**10,860**
HMR10.2 M1	GGGACGAC	**A**G	GGATTC	GGTTCGTGTCTAACAGTGC	CTCGT	AC	GTCGTCCC	**190**
HMR10.2 M2	GGGACGAC	GG	GGATTC	**A**GTTCGTGTCTAACAGTGC	CTCGT	AC	GTCGTCCC	**14**
HMR10.2 M3	GGGACGAC	GG	GGATTC	GGTTCGTGTCTAACAGTG**G**	CTCGT	AC	GTCGTCCC	**4**
HMR10.2 M4	GGGACGAC	GG	GGATTC	GG**C**TCGTGTCTAACAGTGC	CTCGT	AC	GTCGTCCC	**4**
HMR10.2 M5	GGGACGAC	GG	GGATTC	GGTTCGTGTCTAA**G**AGTGC	CTCGT	AC	GTCGTCCC	**0**
HMR10.2 M6	GGGACGAC	GG	GGATTC	GGTTCG**C**GTCTAACAGTGC	CTCGT	AC	GTCGTCCC	**7**
HMR10.2 M7	GGGACGAC	GG	GGATTC	GGTTCG**C**GTCTAGCAGTGC	CTCGT	AC	GTCGTCCC	**0**
HMR10.2 M8	GGGACGAC	GG	GGATTC	GGTTCGTG**ATC**AACAGTGC	CTCGT	AC	GTCGTCCC	**3,648**
HMR10.1	GGGACGAC	GG	GGATTC	GGTTCTCGATAAGTAGTGC	CTCGT	AC	GTCGTCCC	**11,423**
HMR10.3	GGGACGAC	GG	GGATTC	GGTTCGCGATGAGCAGTGC	CTCGT	AC	GTCGTCCC	**10,213**
HMR10.4	GGGACGAC	GG	GGATTC	GGTTCGTGTTCAGCAGTGC	CTCGT	CC	GTCGTCCC	**10,051**
HMR10.5	GGGACGAC	GG	GGATTC	GGTTCGTGATTAGCAGTGC	CTCGT	CC	GTCGTCCC	**9,891**
HMR10.6	GGGACGAC	AG	GGATTC	GGTTCGTGGTTAACAGTGC	CTCGT	CT	GTCGTCCC	**9,502**
HMR10.7	GGGACGAC	GG	GGATTC	GCGGATTCGGAGAAAAGGG	CTCGT	AC	GTCGTCCC	**8,598**
HMR10.8	GGGACGAC	AG	GGATTC	GGTCTTGGTACATAGGGGC	CTCGT	CT	GTCGTCCC	**7,190**
HMR10.9	GGGACGAC	GG	GGATTC	GGTCTTGGTACATAGGGGC	CTCGT	AC	GTCGTCCC	**6,627**
HMR10.10	GGGACGAC	AG	GGATTC	GGTCTTGGAACAAAGGGGC	CTCGT	CT	GTCGTCCC	**4,205**
HMR10.11	GGGACGAC	GG	GGATTC	GCGGATTCATAGAAAAGGG	CTCGT	CC	GTCGTCCC	**3,851**
HMR10.12	GGGACGAC	CA	GGATTC	GGTTCTTTGTAAATAGTGC	CTCGT	TG	GTCGTCCC	**2,691**
HMR10.13	GGGACGAC	CA	GGATTC	GGTTCGGAAAGTCCAGTGC	CTCGT	TG	GTCGTCCC	**1,779**
HMR10.14	GGGACGAC	AG	GGATTC	GGTTCTCGATAAGTAGTGC	CTCGT	AC	GTCGTCCC	**233**
HMR10.15	GGGACGAC	AG	GGATTC	GCGGATTCGGAGAAAAGGG	CTCGT	AC	GTCGTCCC	**127**
HMR10.16	GGGACGAC	GG	GGATTC	GCGGATTCATGGAAAAGGG	CTCGT	CT	GTCGTCCC	**92**

Table S8. Top sequences (>2,000 CPM) in the APX1 aptamer family identified from the original SELEX.

Sequence ID	Sequence (5'→3')							Frequency (CPM)
	Primer	V1	C1	V2	C2	V3	Primer	
APX1.1	GGGACGAC	UUGG	GGAC	UGCCCGCCACAACACA	GAAU	UA	GUCGUCCC	**519,700**
APX1.2	GGGACGAC	CGUUGG	GGAC	UGCCCGAAUACA	GAAU	UACA	GUCGUCCC	**66,600**
APX1.3	GGGACGAC	UCGCG	GGAC	CCACCGUUAAUGG	GAAU	UCUA	GUCGUCCC	**35,050**
APX1.4	GGGACGAC	UUGG	GGAC	UGCCCGCCACAAUACA	GAAU	UA	GUCGUCCC	**22,170**
APX1.5	GGGACGAC	UUGAGG	GGAC	UAUCCUACAAUA	GAAU	UUCA	GUCGUCCC	**20,340**
APX1.6	GGGACGAC	UUGAGG	GGAC	UAUCCUACUAUA	GAAU	UUCA	GUCGUCCC	**12,810**
APX1.7	GGGACGAC	UUGG	GGAC	UGCCCGGCACAACACA	GAAU	UA	GUCGUCCC	**10,720**
APX1.8	GGGACGAC	UUGG	GGAC	UGCCCGCCAUAACACA	GAAU	UA	GUCGUCCC	**4,180**
APX1.9	GGGACGAC	CGUUGG	GGAC	UGCCCAAAUACA	GAAU	UACA	GUCGUCCC	**3,844**
APX1.10	GGGACGAC	UUGG	GGAC	UGCCCGCUACAACACA	GAAU	UA	GUCGUCCC	**2,709**
APX1.11	GGGACGAC	UUGG	GGAC	UGCCCGAAUACA	GAAU	UACA	GUCGUCCC	**2,464**
APX1.12	GGGACGAC	UUGG	GGAC	UGCCCGUCACAACACA	GAAU	UA	GUCGUCCC	**2,170**
APX1.13	GGGACGAC	UUGG	GGAC	UGCCCGCAACAACACA	GAAU	UA	GUCGUCCC	**2,107**
APX1.14	GGGACGAC	GAGG	GGAC	CGCCCACGAUACG	GAAU	UUAAU	GUCGUCCC	**2,072**
APX1.15	GGGACGAC	UUGG	GGAC	UGCCCGCCACAACACA	GAAU	UA	GUCGUCCC	**2,002**

All A, C, and U are 2' O-Methyl modified, and G are 2' F modified.

Table S9. Benefits and limitations of motif-SELEX and other aptamer affinity maturation approaches.

Approach	Benefits	Limitations
Motif libraries and motif-SELEX (this work)	• Equal representation of library members • Variable-length random domains (*i.e.*, insertions and deletions) • Can identify binding elements and aid prediction of aptamer secondary & tertiary structure • Simple and cost-effective library construction (PCR & strand separation)	• Requires identification of binding site/aptamer motifs
Doped libraries and doped-SELEX (*Cell* 1991, 67, 529 – 536)	• No prior knowledge required of binding site/important motifs • Can identify binding elements and aid prediction of aptamer secondary & tertiary structure	• Unable to incorporate length variation in library • Unequal representation of library members • Less informative about aptamer structure • Cannot present highly mutated sequences • Doped library synthesis is more costly relative to completely randomized libraries • May require trial-and-error adjustment of nucleotide doping
SELEX with error-prone PCR (*Genome Res.* 1994, 3, S136–S140)	• Simple to execute (perform PCR with altered conditions) • No prior knowledge required of binding site/important motifs	• Unable to control types of mutations • Unable to control mutation rate • Difficult to implement length variation in random domain
Insertion-reselection strategy (*Science* 2023, 380, 942–948)	• Library synthesis is straightforward	• Principles of library design are not well established; sites of random nucleotide insertions are currently arbitrary • Aptamers are long (>60 nt) • Insertions are fixed in length, unlike in motif libraries
Oligo microarrays (*Anal. Chem.* 2016, 88, 6981–6985 & *J. Anal. Methods Chem.* 2015, 2015, 137489)	• Rapid and simultaneous screening of sequences without bias • Single round	• The sequence space searched is limited to 10^6 sequences • Requires special equipment

Figure S1. Scheme of the generation of the HM1 motif library via PCR amplification and single-strand generation.

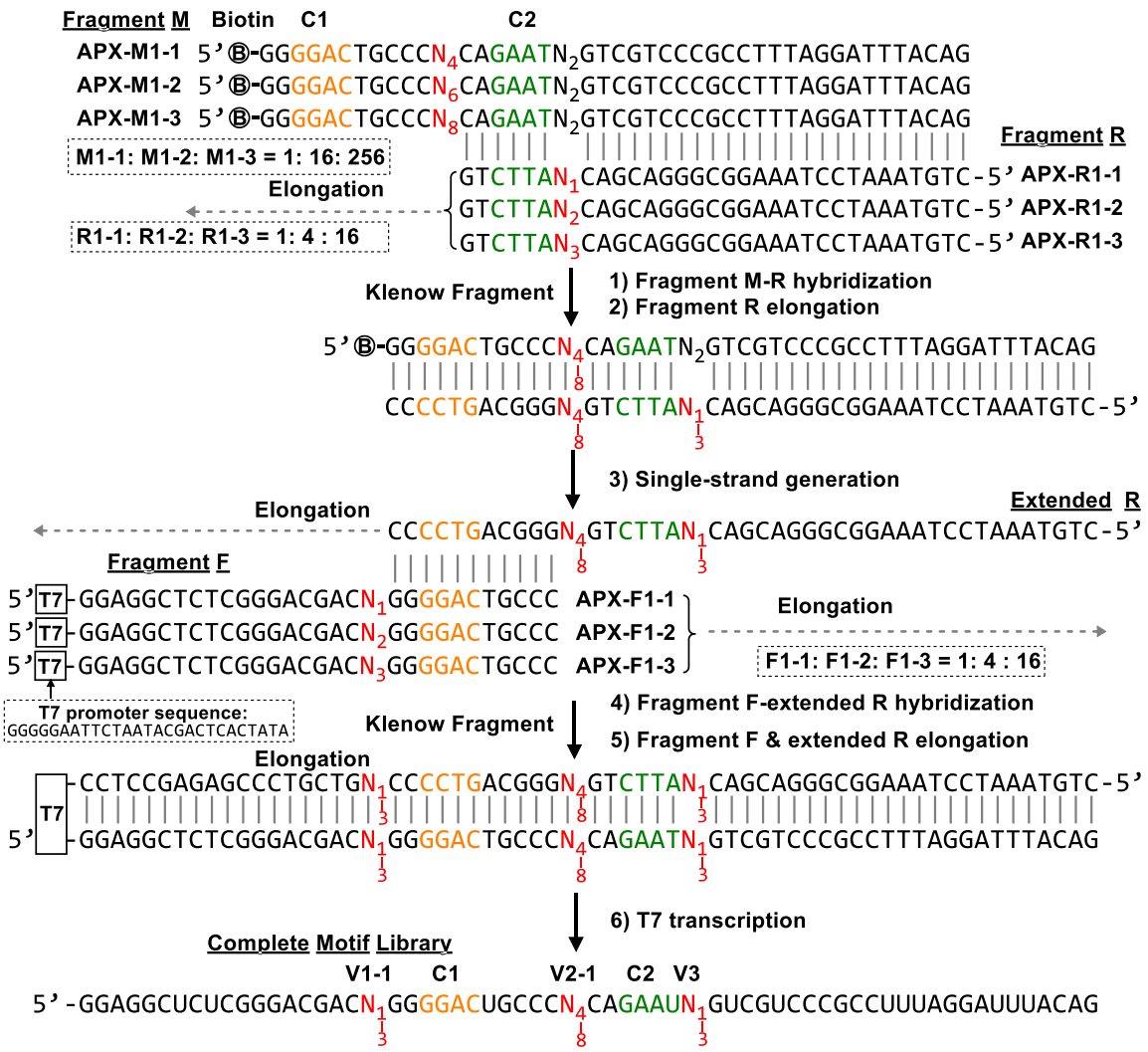


Figure S2. Scheme of the motif library generation process for phase 1 of motif-SELEX for apixaban.

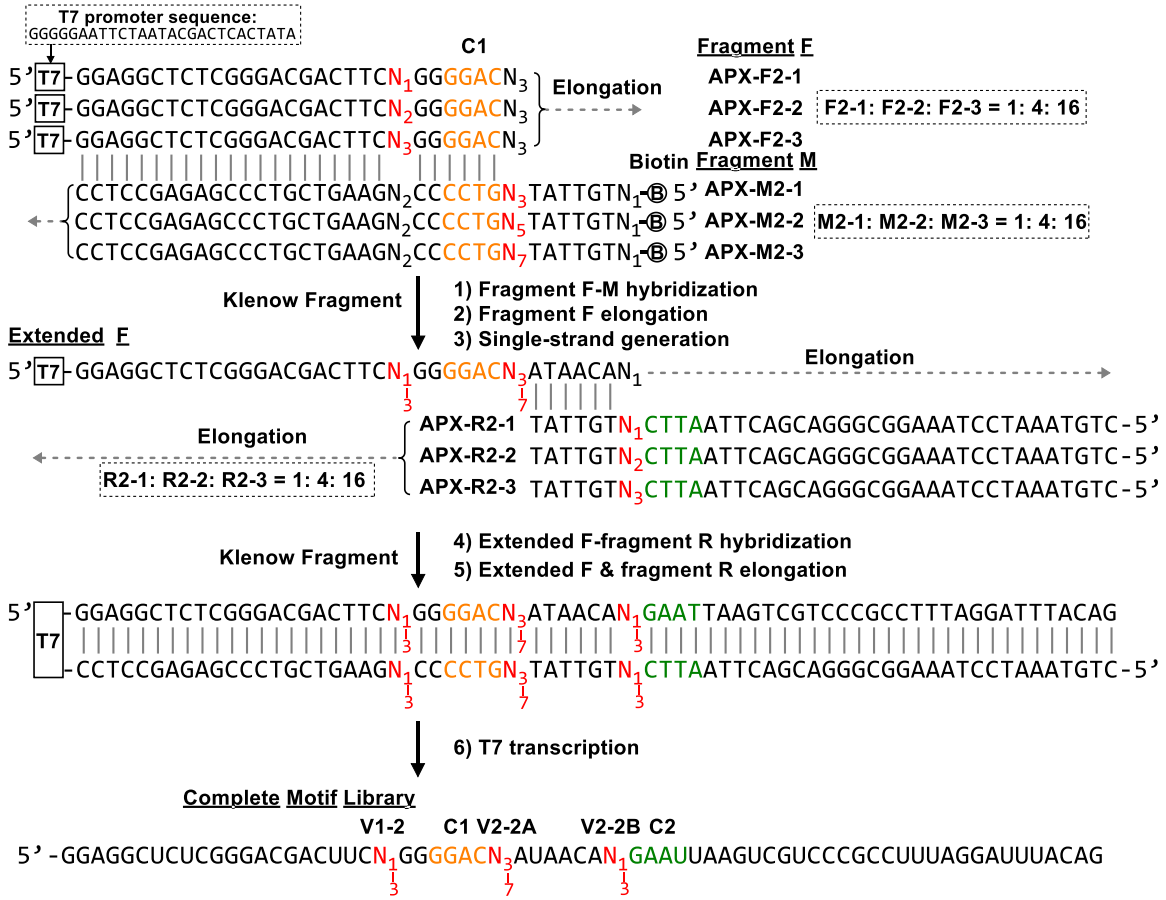


Figure S3. Scheme of the motif library generation process for phase 2 of motif-SELEX for apixaban.

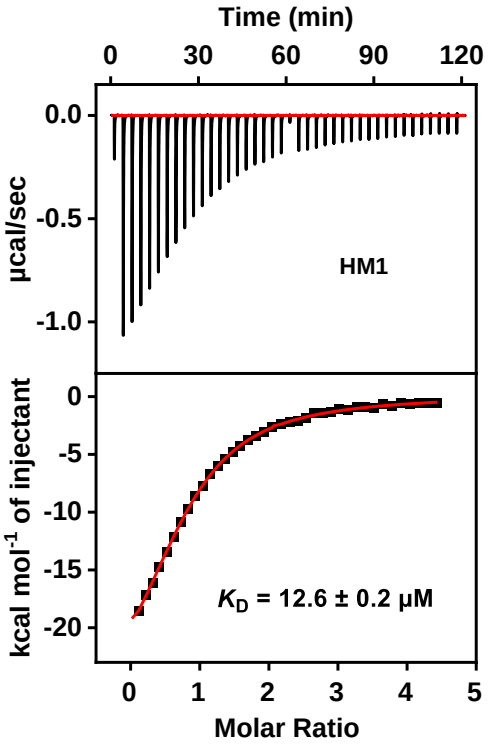


Figure S4. Isothermal titration calorimetry (ITC) measurements of binding affinity of HM1 for morphine in binding buffer (1× PBS with 1 mM $MgCl_2$). Top panel presents the raw data showing heat generated from each target titration into the aptamer, bottom panel depicts the integrated heat of each titration after correcting for the dilution heat of morphine. Data were fitted with a one-site binding model.

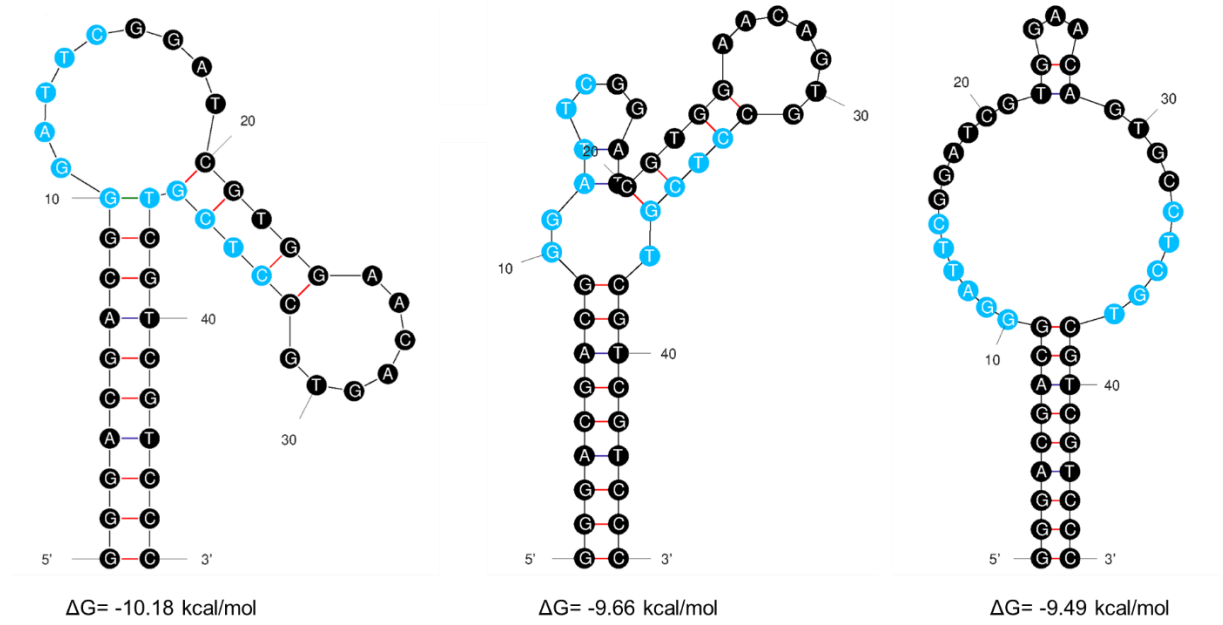


Figure S5. Different secondary structures predicted by mfold for HM1 in 150 mM Na^{+}, 1 mM Mg^{2+}, 25°C. The consensus domains are highlighted blue.

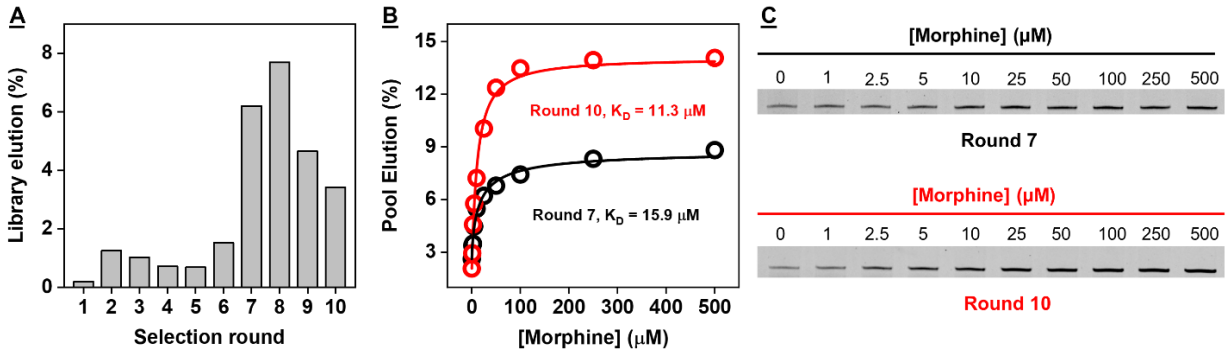


Figure S6. Monitoring selection progress of motif-SELEX for morphine. (**A**) Target elution in each round in motif-SELEX. (**B**) Pool affinity for Round 7 and Round 10 as characterized by (**C**) gel elution assay.

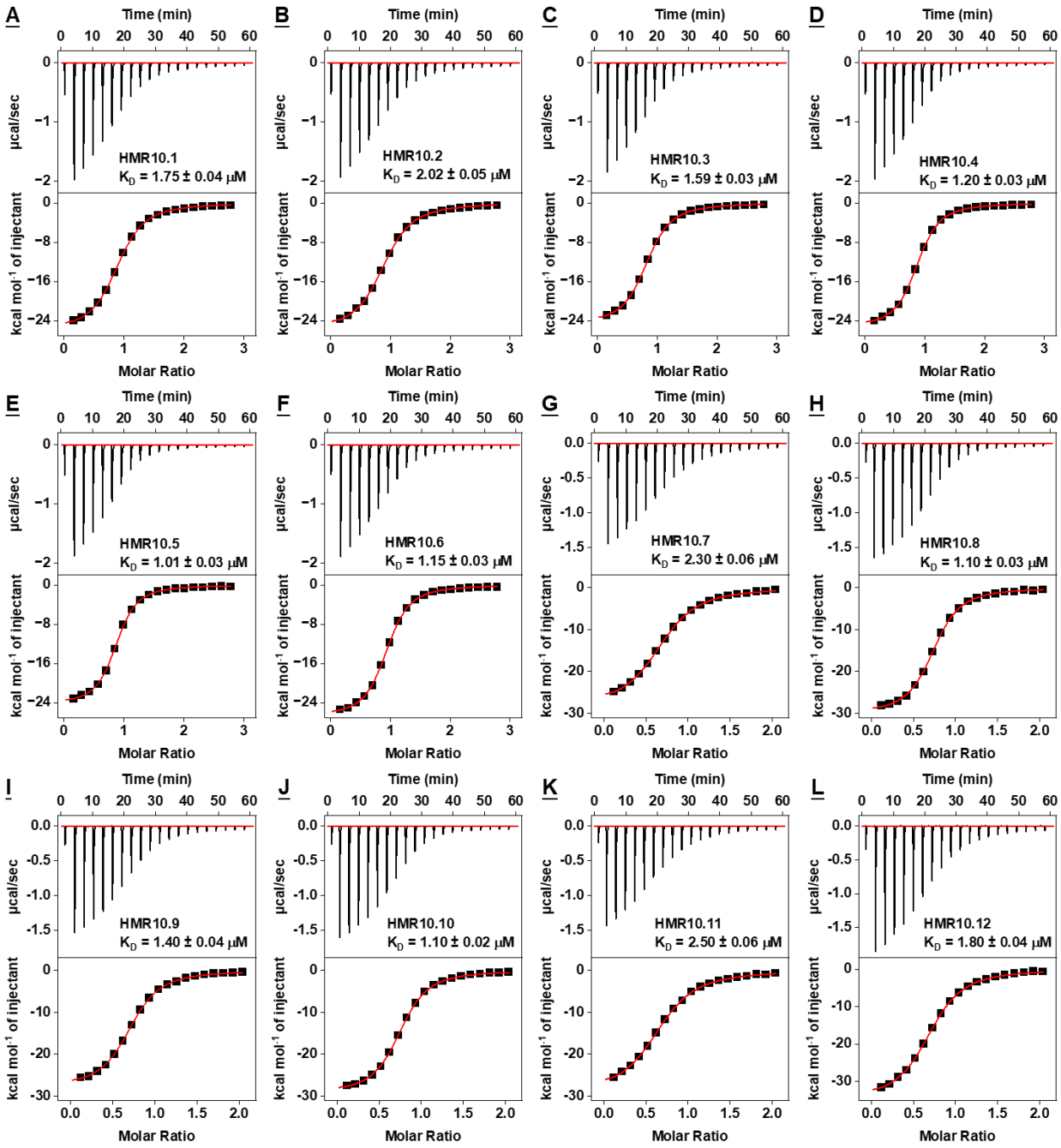


Figure S7. ITC characterization of morphine binding to aptamers isolated from motif-SELEX and their mutants. Top panels show the raw titration heat, bottom panels show the integrated heat from each titration and the calculated binding parameters for (**A**) HMR10.1, (**B**) HMR10.2, (**C**) HMR10.3, (**D**) HMR10.4, (**E**) HMR10.5, (**F**) HMR10.6, (**G**) HMR10.7, (**H**) HMR10.8, (**I**) HMR10.9, (**J**) HMR10.10, (**K**) HMR10.11, (**L**) HMR10.12. The titration parameters are shown in table S5.

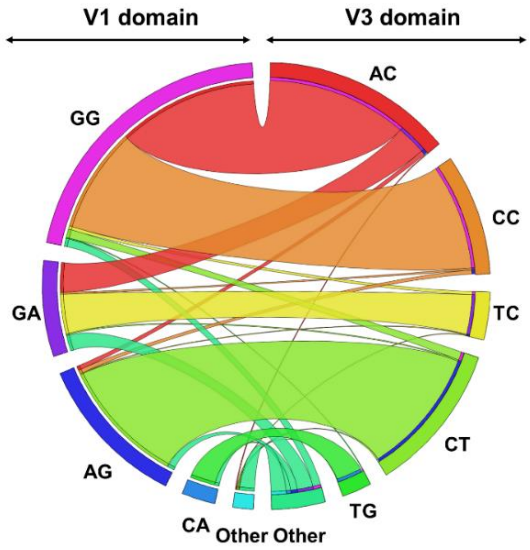


Figure S8. Circos plot showing linkage between V1 and V3 domain sequences in the Round 10 HM1 motif-SELEX pool. Segment sizes show sequence popularity in V1 and V3; ribbon sizes show the frequency of each V1 and V3 sequence combination.

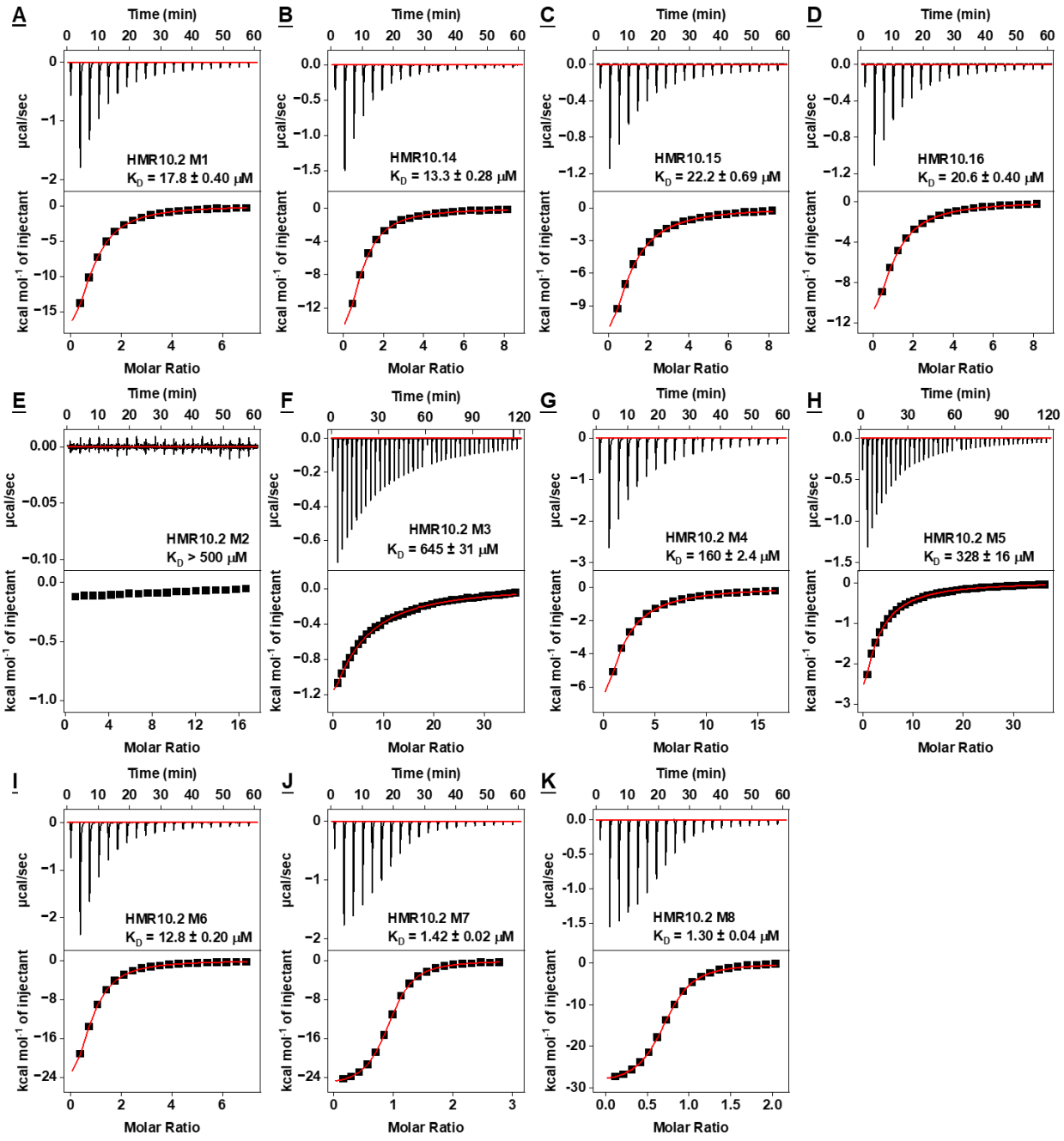


Figure S9. ITC characterization of morphine binding to new aptamers isolated from motif-SELEX and their mutants. The top panels show the raw titration heat, and the bottom panels show the integrated heat from each titration and the calculated binding parameters for (**A**) HMR10.2 M1, (**B**) HMR10.14, (**C**) HMR10.15, (**D**) HMR10.16, (**E**) HMR10.2 M2, (**F**) HMR10.2 M3, (**G**) HMR10.2 M4, (**H**) HMR10.2 M5, (**I**) HMR10.2 M6, (**J**) HMR10.2 M7, and (**K**) HMR10.2 M8. The titration parameters are shown in table S5.

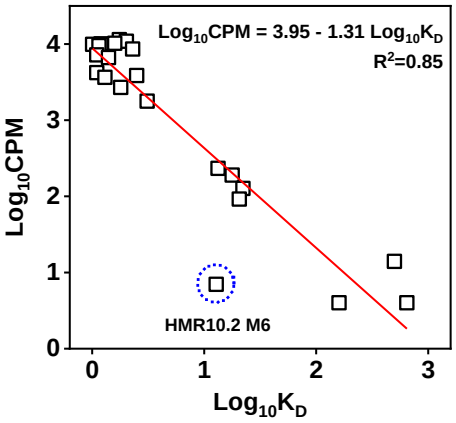


Figure S10. Correlation between frequency of sequences identified ($Log_{10}CPM$) in the morphine motif-SELEX pool and their affinity ($Log_{10}K_D$). The outlier (HMR10.2 M6) in the dataset is circled in blue.

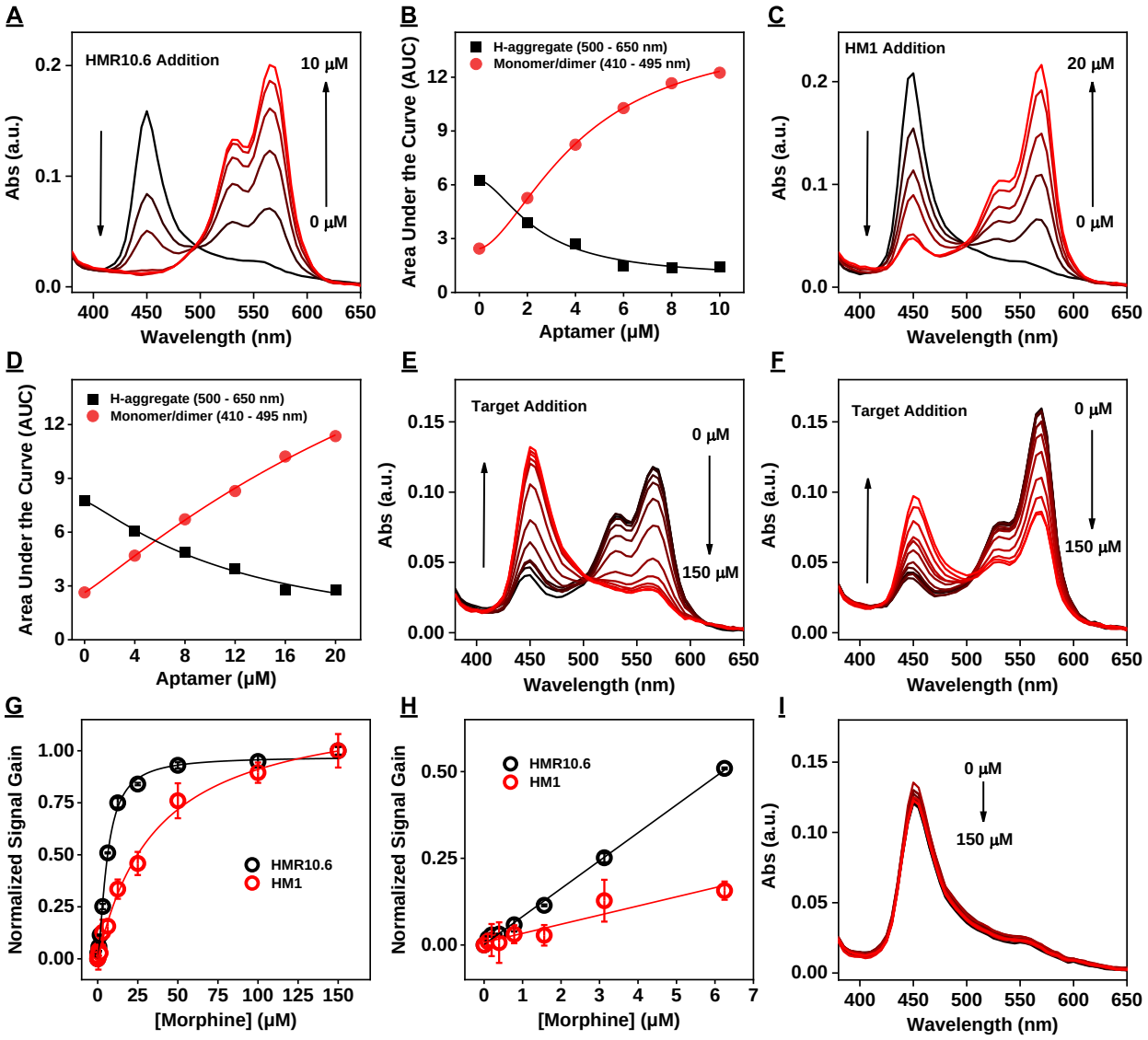


Figure S11. Comparing the analytical performance of an affinity matured aptamer for morphine (HMR10.6) to its parent aptamer (HM1) in a dye-displacement assay. (**A**) Absorbance spectrum of cyanine dye X-732-91B in the presence of various concentrations of HMR10.6 and (**B**) areas under the curves for the absorbance of dye monomer (400 – 495 nm) and H-aggregate (500 – 650 nm) plotted as a function of aptamer concentration. The black-to-red color gradient indicates increasing concentrations of aptamer. (**C & D**) Aptamer concentration optimization experiments were likewise performed for HM1. (**E**) Absorbance spectrum of the dye when HMR10.6-dye complex (final concentrations 5 µM and 4 µM for HMR10.6 and dye, respectively) or (**F**) HM1-dye complex (final concentrations 15 µM and 4 µM for HM1 and dye, respectively) is challenged with various concentrations of morphine (0 – 150 µM). The black-to-red color gradient indicates increasing concentrations of morphine. (**G**) Calibration curves for morphine detection based on assays employing HMR10.6 (black) and HM1 (red) with (**H**) signals in the linear range clearly shown. (**I**) Absorbance spectrum of X-732-91B alone in buffer challenged with various concentrations of morphine shows that morphine itself does not meaningfully interact with the dye.

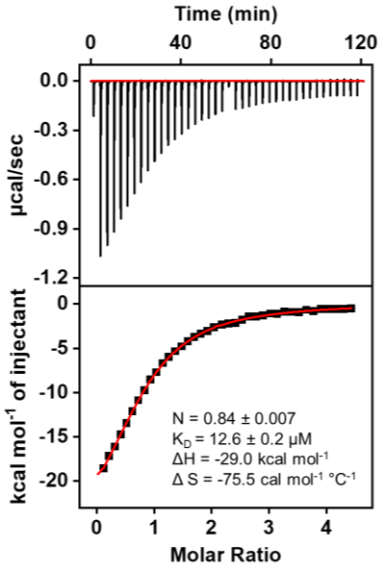


Figure S12. ITC characterization of apixaban binding to APX1.1 in binding buffer. Top panel shows the raw titration heat, bottom panel shows the integrated heat from each titration and the calculated binding parameters. The titration parameters are shown in table S5.

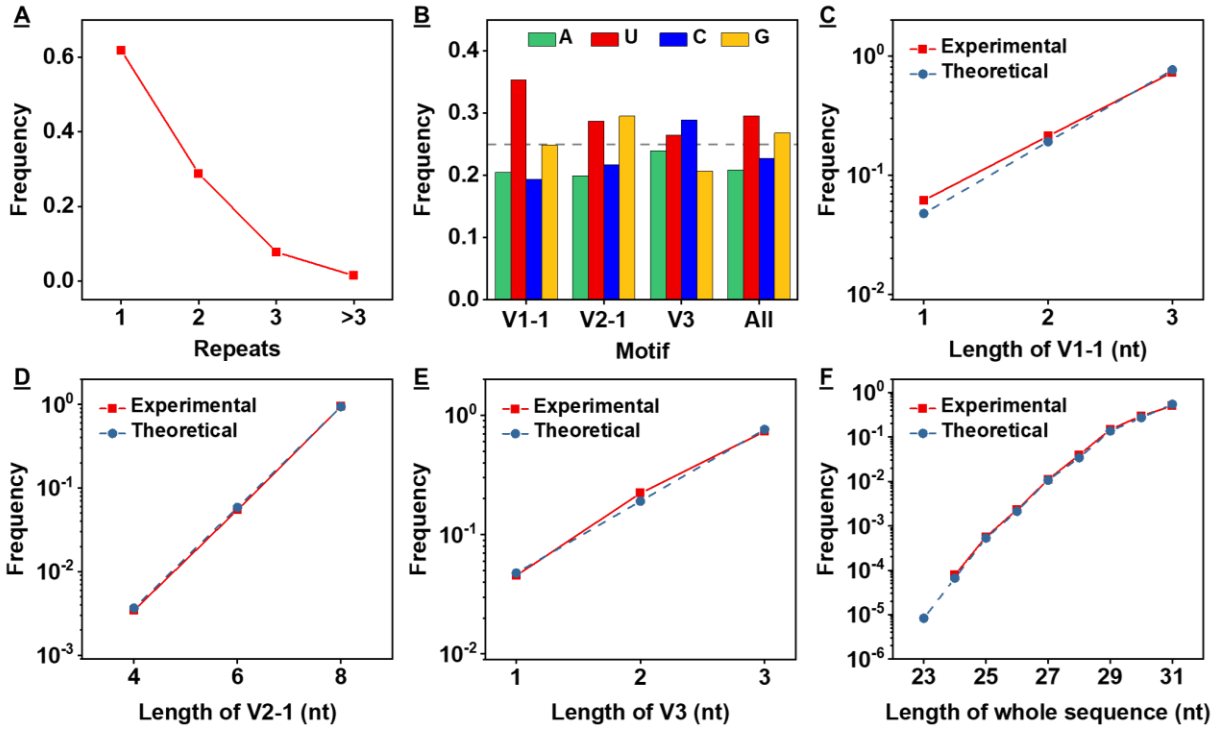


Figure S13. Characterization of the library for the first phase of motif-SELEX for apixaban. We subsequently analyzed (**A**) diversity of the library, (**B**) base distribution in V1-1, V2-1, V3, and all variable domains, and the length distribution of (**C**) V1-1, (**D**) V2-1, (**E**) V3, and (**F**) the whole sequence**.** Dashed lines in **C–F** are for comparison with theoretical values.

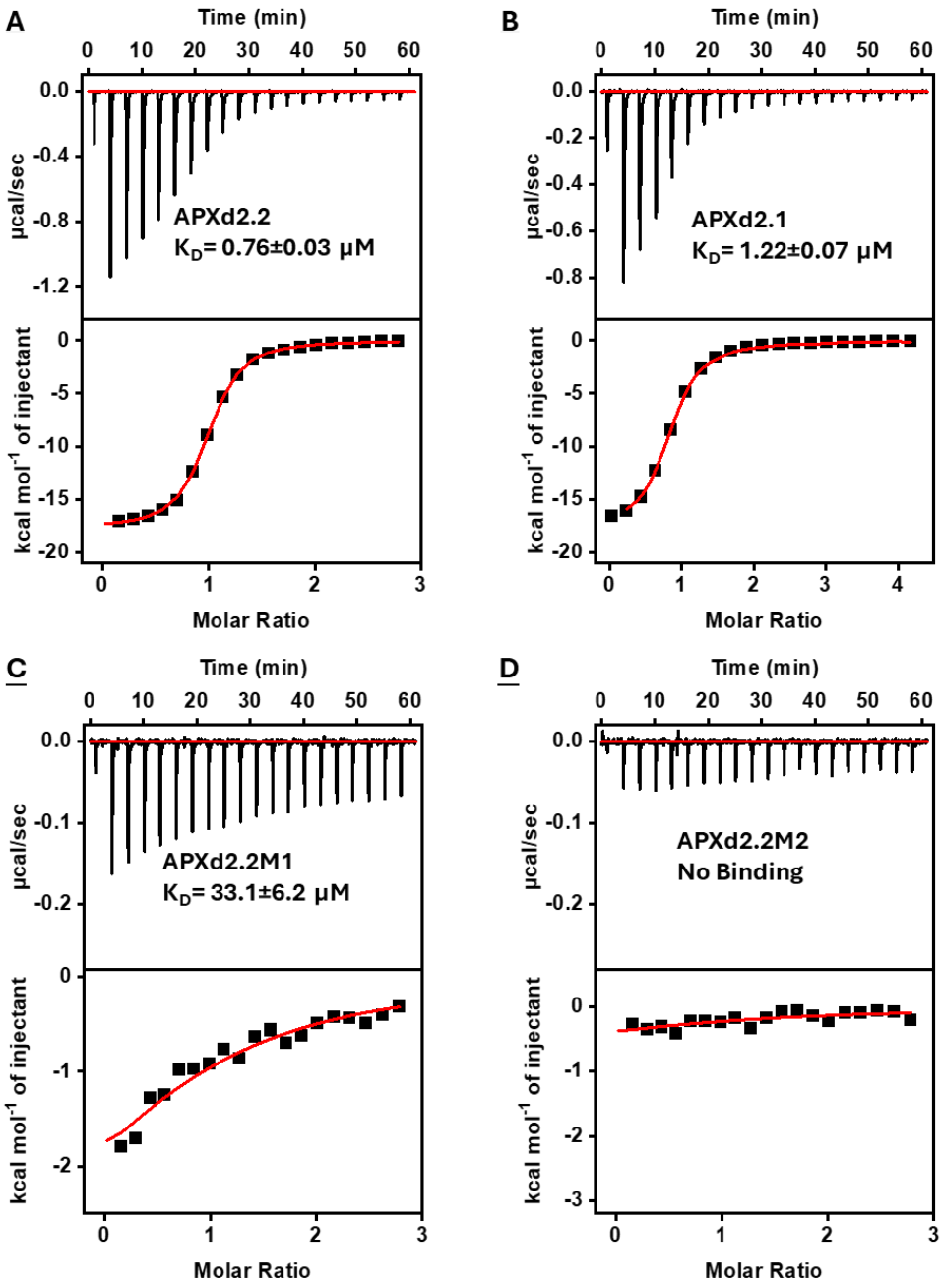


Figure S14. ITC characterization of apixaban binding to aptamers isolated from motif-SELEX and their mutants. Top panels display the heat generated from each titration of apixaban into (**A**) APXd2.2, (**B**) APXd2.1, (**C**) APXd2.2M1, and (**D**) APXd2.2M2. Bottom panels show the integrated heat of each titration after correcting for the heat of dilution of the titrant. The titration parameters are shown in table S5.

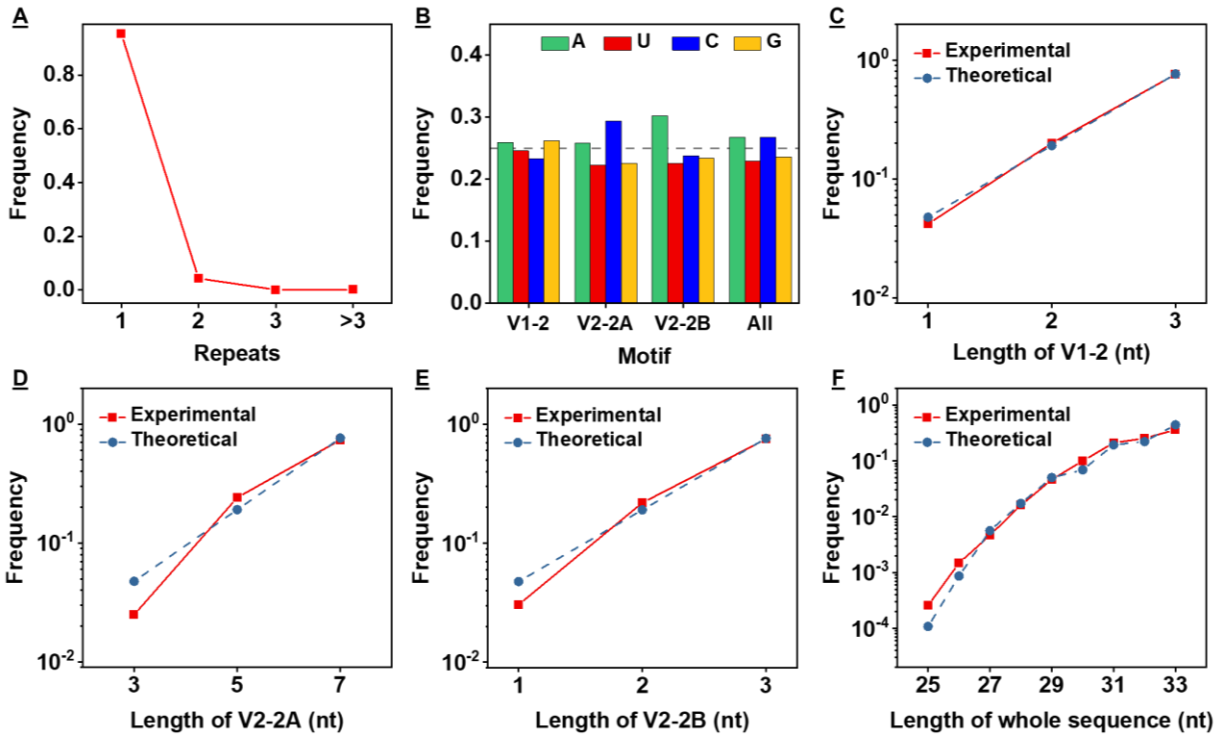


Figure S15. Characterization of the library for the second phase of motif-SELEX for apixaban. We subsequently analyzed (**A**) diversity of the library, (**B**) base distribution in V1-2, V2-2A, V2-2B, and all variable domains, and the length distribution of (**C**) V1-2, (**D**) V2-2A, (**E**) V2-2B, and (**F**) the whole sequence**.** Dashed lines in **C**–**F** are for comparison with theoretical values.