

Supporting Information

X-Aptamers: A bead-based selection method for random incorporation of drug-like moieties onto next-generation aptamers for enhanced binding

Weiguo He^{1,2}, Miguel-Angel Elizondo-Riojas^{1,3}, Xin Li^{1,2}, Ganesh Lakshmana Rao Lokesh^{1,2}, Anoma Somasunderam^{1,2}, Varatharasa Thiviyanathan^{1,2}, David E. Volk^{1,2}, Ross H. Durland⁴, Johnnie Englehardt⁴, Claudio N. Cavasotto⁵, David G. Gorenstein^{1,2,*}

¹Center for Proteomics and Systems Biology, The Brown Foundation Institute for Molecular Medicine for the Prevention of Human Diseases, and ²Department of Nanomedicine and Biomedical Engineering, The University of Texas Health Science Center – Houston, 1825 Pressler Street, Houston, TX 77030.

³Centro Universitario Contra el Cáncer, Hospital Universitario “Dr. José Eleuterio González”, Universidad Autónoma de Nuevo León, Monterrey, N.L. México.

⁴AM Biotechnologies, 12521 Gulf Freeway, Houston, TX 77034.

⁵Instituto de Investigación en Biomedicina de Buenos Aires (IBioBA)-CONICET-Partner Institute of the Max Planck Society, Godoy Cruz 2390 3rd. Floor, C1425FQA, Buenos Aires, Argentina.

**Corresponding Author: David G. Gorenstein, David.G.Gorenstein@uth.tmc.edu, Tel: 713-500-2233 Fax: 713-500-0319*

MATERIALS AND METHODS

Materials. The dA, dG, dC and dT cyanoethylphosphoramidites, and the Beaucage reagent (3H-1,2-benzodithiol-3-one 1,1-dioxide) were purchased from Glen Research (Sterling, VA). The *Taq* polymerase kits were from Applied Biosystems. The TOPO TA Cloning kit was from Invitrogen. Polystyrene beads (60–70 μm) with non-cleavable hexaethyleneglycol linkers with a loading of 36 $\mu\text{mol/g}$ were from ChemGenes Corp. (Ashland, MA). The oligodeoxynucleotides (ODNs) and monothioated S-ODNs used in the study were synthesized on a 1 μmol scale in an Expedite 8909 System (Applied Biosystems) DNA synthesizer. NHS-PEG₁₂-biotin was purchased from Pierce. *N*-acetyl-2,3-dehydro-2-deoxyneuraminic acid (ADDA) was purchased from Sigma-Aldrich.

Synthesis of S-ODN library. Standard phosphoramidite chemistry was used for the synthesis of the S-ODN library. The library was prepared using an automated four-column, split-pool synthesizer⁶ on a 1 $\mu\text{mol/column}$ scale on polystyrene beads. After first synthesizing the 3'-primers, the pseudo-random sequences were programmed, one section at a time, on the four columns of the synthesizer to create the combinatorial S-ODN library as shown in **Figure 1**. Subsequent addition of the 5'-primer completed the 73mer ODNs. A 'split and pool' occurred at each position indicated by a dash in order to synthesize the combinatorial region for the S-ODN. The step-wise coupling yields were approximately 99% as determined by the dimethoxytrityl cation assay. Sulfurization chemistry utilized the Beaucage reagent. The S-ODN combinatorial libraries on non-cleavable linker beads were deprotected with concentrated ammonium hydroxide at 37°C for 21 h. and then washed with doubly distilled water.

Expression and Purification of CD44-HABD. DNA encoding CD44-HABD (20-178 amino acid residues) was synthesized and cloned into expression vector pET19b between the *NdeI* and *BamHI* sites. Protein expression, refolding and purification followed published procedures¹. The purity of CD44-HABD was analyzed by gel electrophoresis (SDS-PAGE and native PAGE). The concentration of CD44-HABD was determined by UV-Vis Spectrometry (276 nm, $\epsilon = 12.95 \text{ mM}^{-1}\text{cm}^{-1}$) and BCA assay (Pierce).

¹H, ¹⁵N-HSQC NMR Data for CD44-HABD. ¹H, ¹⁵N-HSQC data were collected on an 800 MHz Varian UnityPlus spectrometer (Rice University) equipped with a cryogenic probe using 32 transients per fid, sweep widths of 3.2 kHz and 11.2 kHz for f1 (¹⁵N) and f2 (¹H), respectively, and 128 and 1912 complex data points for f1 and f2, respectively. Data were processed using VNMRJ (Varian, Inc.) or Felix (Felix, Inc.) software. The proton dimension was referenced to TSP and nitrogen shifts were referenced indirectly. CD44-HABD was prepared in the NMR buffer (20 mM Tris, 50 mM NaCl, 10% (v/v) D₂O, pH 6.7). ADDA was dissolved in DMSO (1.0 M) as stock solution. CD44-HABD (70 μM) with ADDA (7 mM) was prepared in the same buffer (20 mM Tris, 50 mM NaCl, 1.25 mM ADDA, 0.7% (v/v) DMSO, 5% (v/v) D₂O, pH 6.7).

ADDA binding to CD44-HABD. CD44-HABD was prepared in the NMR buffer (20 mM Tris, 50 mM NaCl, 5% (v/v) D₂O, pH 6.7). ADDA was dissolved in DMSO (1.4 M) as stock solution. The ADDA sample for NMR was prepared in the same buffer (20 mM Tris, 50 mM NaCl, 1.25 mM ADDA, 0.1% (v/v) DMSO, 5% (v/v) D₂O, pH 6.7). CD44-HABD (27 μM) with various

concentration of ADDA (14 μ M, 25 μ M, 50 μ M, 250 μ M and 1250 μ M) was also prepared in the same buffer. The $^1\text{H-NMR}$ spectra of CD44-HABD only, ADDA only and CD44-HABD in presence of ADDA were obtained.

The equilibrium dissociation constant of ADDA to CD44-HABD was determined by $^1\text{H-NMR}$. **Figure S4B** shows a set of spectra measured on changing the concentration of the ADDA ligand, where $I_{\text{lig}}/I_{\text{lig-free}}$ represent the ratio between the intensity of the signal of the protein CD44-HABD with and without the ligand. The $^1\text{H-NMR}$ spectrum of the ligand alone showed no signal in the region of interest of **Figure S4B** which could have interfered with the signal intensity ratios. Several peaks were selected (7.7925, 7.9025, 7.9225, 7.9375, 7.9575 and 7.9925 ppm) in order to obtain a statistical value of the dissociation constant of the ligand ADDA, $K_D = 2.22 \pm 0.82$ mM, as shown in **Figure S4C** and summarized in the following table:

peak (ppm)	7.7925	7.9025	7.9225	7.9375	7.9575	7.9925	Average K_D
K_D [mM]	3.0823	1.4438	1.7213	1.9167	1.7083	3.4259	2.22 ± 0.82

Labeling CD44-HABD protein with biotin. To 130 μ L of CD44-HABD (40.5 μ M) in PBS was added 1.2 μ L of NHS-PEG₁₂-biotin (125 mM in DMF). The reaction was incubated on ice for two hours. Biotin labeled protein was purified from non reacted biotin reagent using Zeba desalt spin columns (Thermo Scientific). The labeled protein was stored at 4°C.

Biotin-labeled CD44-HABD binding to beads and selection of beads. The screen began with the ADDA X-aptamer library incubated in PBS (pH 7.4) containing Tween 20 (0.1%, v/v) and bovine serum albumin (BSA, 0.1%) for 1 h, with shaking, to block non-specific protein binding. The library was then washed with PBS (pH 7.4). A negative selection was carried out first by incubating the washed X-aptamer beads with streptavidin-coated Dyna beads (Invitrogen) for 2 h at room temperature. The slurry of the beads mixture was allowed to slowly pass in solution near a magnet to remove the Dyna beads and any X-aptamer beads which have the Dyna beads bound nonspecifically on their surface. The rest of the X-aptamer beads were then suspended in a dilute solution of biotinylated CD44-HABD (0.01 nM in PBS, pH 7.4) at room temperature overnight. After washing with PBS containing 0.1% Tween-20 and BSA, PBS containing 0.1% Tween-20, and PBS, the library was incubated with streptavidin-coated Dyna beads for 2 h at room temperature. The beads were washed thoroughly with PBS containing 0.1% Tween-20 and BSA, PBS containing 0.1% Tween-20, and PBS. The positive X-aptamer beads which had CD44-HABD bound tightly were isolated from the incubation by passing the beads in solution near a magnet. The positive beads were transferred onto a glass microscope slide and selected manually by pipette. To remove bound proteins, each positive bead was incubated in 8 M guanidine hydrochloride for 1 h then rinsed ten times with water.

One-bead one-PCR amplification and sequencing of PCR products. A selected single bead was mixed with the following PCR components: 6 μ L of 25 mM MgCl₂, 0.5 μ L of *Taq* polymerase (5 U/ μ L), 1 μ L of 8 mM dNTP, 10 μ L of PCR buffer, and 1 μ L of 40 mM primers. The PCR was run on a GeneAmp PCR system 2400 (Perkin Elmer). The PCR mixtures were thermal cycled using the following scheme for amplification: 94°C for 5 min (1 cycle); 94°C for 1 min, 40°C for 2 min and 72°C for 1 min 10 sec (35 cycles); 72°C for 10 min (1 cycle). The PCR product was inserted into pCR2.1 TOPO TA vector (Invitrogen) and sequenced. PCR amplification converts the original X modifications to Ts. X positions were determined by reference to the original

library design (**Figure 1**), using the adjacent bases as “barcoding”. **Table S1** lists the XA sequences obtained.

Conjugation of ADDA to XAs. To 1 μL of a 1 M ADDA solution in DMSO, 5 μL of a freshly prepared ethyldimethylaminopropyl carbodiimide (EDAC) solution (1 M in H_2O) and 5 μL of an *N*-hydroxysuccinimide (NHS) solution (1 M in DMSO) were added. The reaction was kept at room temperature for 1 h. To couple ADDA to the original X-aptamer bead library, 10% of the original XA library was added, and the resulting mixture was gently shaken overnight. Beads were washed to remove unused reagents, and the conjugation was confirmed by a ninhydrin test. To couple ADDA to XAs, XAs (0.05 μmol) were dissolved in 100 μL of sodium bicarbonate buffer (0.1 M, pH 8.5). ADDA-NHS ester solution was added to the solution of XAs. The mixture was first vortexed and then shaken at room temperature for 4 h. The ADDA-modified XAs were purified by reverse phase HPLC on a Hamilton PRP-1 column. Buffer A: 100 mM triethylammonium acetate in H_2O (pH 7.5); buffer B: acetonitrile. Gradient: 0-40% B, 0- 60 min; 40-100% B, 60-70 min; 100-0% B, 70-75 min. Flow rate is 2 mL/min. The fractions of ADDA-modified XAs were combined, lyophilized and analyzed by 15% PAGE.

Filter binding assay. The equilibrium binding constants of selected XAs for CD44-HABD were determined by a filter binding assay. The biotinylated XAs (1 nM) were incubated with varying concentrations of CD44-HABD in 50 μL of 20 mM Tris buffer (150 mM NaCl, pH 8.0) for 40 min at room temperature and then transferred to a 96-well dot-blot apparatus and filtered under vacuum onto nitrocellulose membranes, which retain the CD44-HABD along with any bound XAs. The amount of biotinylated XA retained at each spot was determined by chemiluminescent detection using the Chemiluminexcent Nucleic Acid Detection Module (Thermo Scientific) following the manufacturer’s instructions. The chemiluminescent signals were collected on a Chemimager (Alpha Innotech). Image analysis and quantification of spot intensities were performed using ImageJ (version 1.42q)². Binding analysis was based on the spot intensities on the nitrocellulose membranes with subtraction of background spot intensity due to the buffer effect from all the data points. Saturation binding curves were generated by using GraphPad Prism with curve fits assuming a single binding site. The equilibrium dissociation constants, K_D , were derived from these curves and are listed in **Table S2**. The binding curves of ADDA XA3 and ADDA XA7 are shown in **Figure S5** and **Figure S6**, respectively. MFold-predicted secondary structures indicated that all selected XA sequences can form hairpin loop structures with the random region forming the loop and the primer regions making up the stem regions (**Figure S1**). We proposed some binding motifs based on the predicted secondary structures and made several smaller constructs of various stem-loop regions of these selected X-aptamers (**Table S1**). The equilibrium binding constants of the small constructs were determined by the filter binding assay. The binding curves of motif 3, motif 4 and motif 5 are shown in **Figures S7-S10**. We found that one site specific binding with Hill Slope always provides superior data fits (data not shown). One possible explanation is that the actual CD44-HABD concentrations could be lower than the apparent concentration calculated from the dilution factors due to binding with the plastic tubes. A mock binding experiment was carried out in which CD44-HABD of eleven varying concentrations were diluted and incubated without XAs for 1 h. The actual CD44-HABD before and after incubation were determined by microBCA assay (Pierce). The results showed a decrease in CD44-HABD concentration over the incubation time and confirmed our explanation. To correct for the plastic binding effects and find the actual concentration for all the binding data,

the eleven actual concentrations determined by microBCA were plotted against their apparent concentration of CD44-HABD. Linear regression was carried out with a fixed X-intercept (**Figure S2**). The value of the fixed X-intercept was chosen by averaging all the X-intercepts from the binding curves which were determined by having linear regression on the first 3 or 4 data points of each curve. The apparent concentrations of CD44-HABD in the binding data were converted to the actual concentration by using the regression line before performing the saturation binding curve fitting.

Molecular Dynamics Simulation protocol. The initial structure (human CD44-HABD, 1POZ.pdb) was neutralized by adding Na^+ counter ions using an algorithm of *xLeap* (AMBER9 suite of programs)³. The latter structure was then surrounded by a 10 Å layer of TIP3P model water molecules in an orthorhombic box of approximately 60 x 72 x 90 Å containing a total of 28386 atoms. The system was minimized by 1000 steps steepest descent method then 4000 steps of conjugate gradient. The solvated system was then equilibrated as follows i) 40 ps MD simulation to gradually heat the system from 0 to 300K keeping all molecules restrained by 25 kcal mol⁻¹ Å⁻² under NVT conditions, ii) 20 ps MD simulation at 300K with all molecules restrained by 25 kcal mol⁻¹ Å⁻² under NVT conditions, iii) the system was subsequently equilibrated in 7 MD simulation rounds over 1180 ps where the positional restraints were gradually relaxed under NPT conditions at 300K and finally a structure production run of 10 ns MD simulation at 300K under NPT conditions was performed. The long-range electrostatics were accounted for using the particle-mesh Ewald summation method, as implemented in the PMEMD module of AMBER9, and the force field *ff99SB* was applied. The SHAKE algorithm was used to constrain covalent bonds to hydrogen atoms allowing a time step of 2 fs. A cutoff of 9 Å was chosen for the non-bonded van der Waals interactions. During the heating protocol at NVT conditions, the Berendsen temperature coupling algorithm was used with a coupling constant of 2.0 ps. During the equilibration and production of the simulation, the Langevin dynamics were used with a collision frequency of 1.0 ps⁻¹. This simulation allowed us to find 4 representative structures using the *cluster* analysis algorithm on *ptraj* software (AMBER9 suite of programs)³. A second 10 ns MD simulation was performed including the hyaluronic acid ligand bound into the HABD this time. The same MD protocol was used as described before but applying the *ff99SB + GLYCAM06* force field because of the nature of the hyaluronic acid ligand. Therefore, a second set of 4 representative structures were found using the *cluster* analysis algorithm on *ptraj* software (AMBER9 suite of programs)³. These 8 representative structures plus the original 1POZ.pdb structure were used as possible binding pockets for a virtual screening of small molecules.

Virtual Screening protocol. A group of 2553 possible ligands with modifiable groups allowing attachment of the ligand to the X-aptamer ODN scaffold were selected from ChemBridge database and 245 ligands from Sigma-Aldrich. Each ligand followed a 0.5 ns MD simulation on implicit water using the Generalized Born method at 300K in order to find and select an ensemble of possible rotamers. The rotamers were chosen using the following protocol: i) 20 ps to gradually heat the system from 0 to 300K, ii) 460 ps of production run at 300K and iii) 20 ps to gradually cool down the system from 300 to 0K. A total of seven rotamers per ligand were initially selected: the first rotamer was the initial structure of stage i) which is the original structure downloaded from the database, five more rotamers were found using the *cluster* analysis algorithm on *ptraj* software (AMBER9 suite of programs)³ from stage ii) and the final

rotamer was selected at the end of stage iii). All rotamers were minimized with the Generalized Born method. Only rotamers > 2.0 rmsd from the original structure (first rotamer) were considered. After rotamer selection a semi-empirical quantum calculation (AM1-BCC method, AMBER9 suite of programs)³ was performed for each rotamer, and an average charge was calculated and used for all rotamers of the same ligand. The initial rotamer structures and their charges are critical components during the virtual screening process in order to get good lead compounds. Using DOCK6.4 software⁴ on the 9 representative protein structures with the 2798 small molecules, the ADDA ligand (*N*-acetyl-2,3-dehydro-2-deoxyneuraminic acid, PubChem: CID 65309, **Figure S3B**) was selected as lead compound.

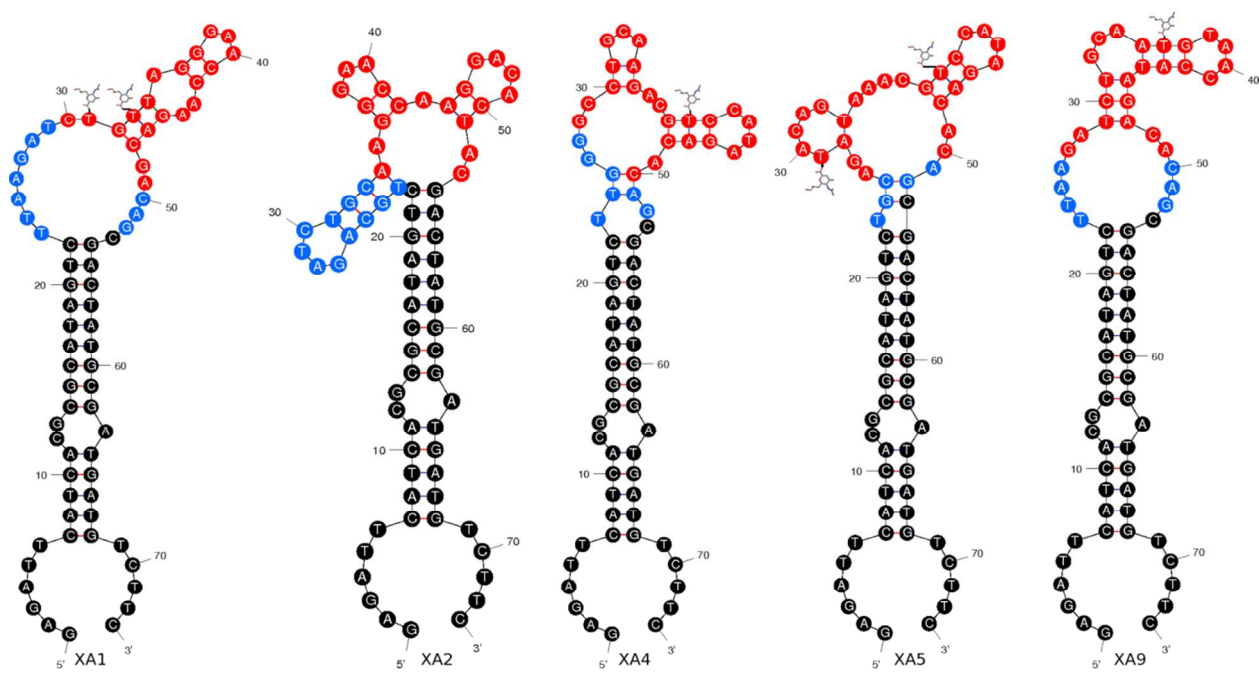


Figure S1. Secondary structures of some selected X-aptamers. Secondary structures were predicted with MFold. Primer regions are in black while random regions are in blue and red. Proposed binding motifs are shown in red. The positions where ADDA was coupled are also shown.

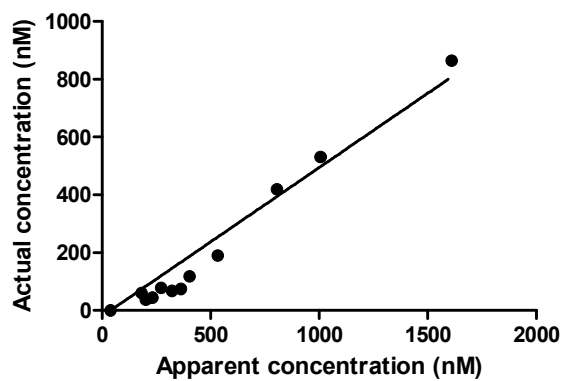


Figure S2. Correlation of the actual concentrations and apparent concentrations of CD44-HABD. Apparent concentrations were calculated based on known dilution factors. Actual concentrations were determined by MicroBCA after 1 h incubations under mock binding conditions. Actual concentrations were consistently lower than apparent concentrations, presumably due to adsorption to plastic tube walls.

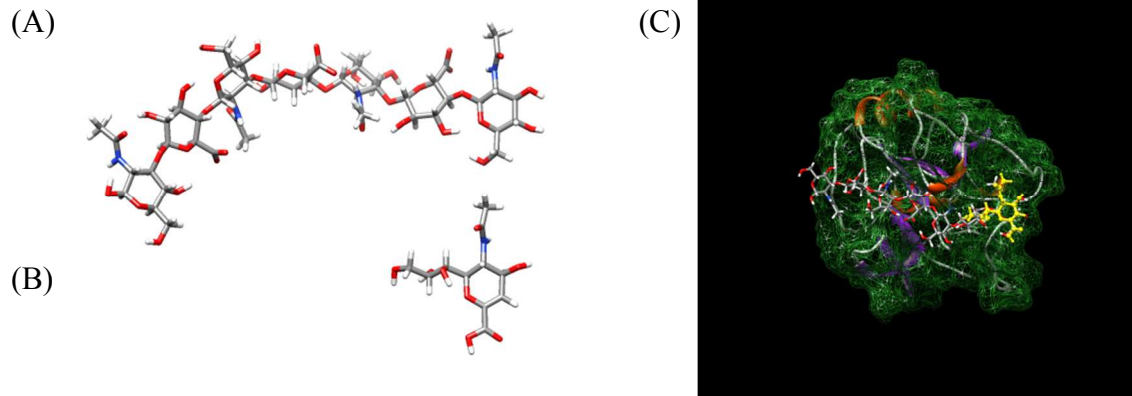
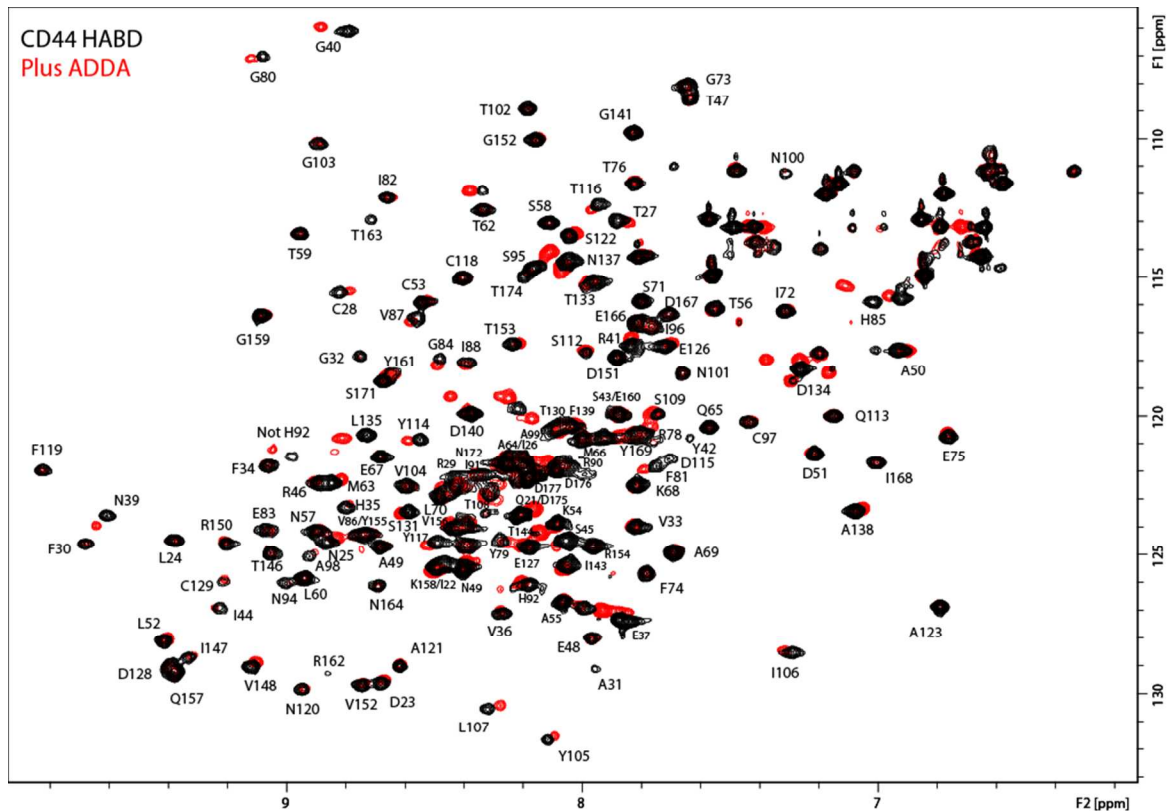
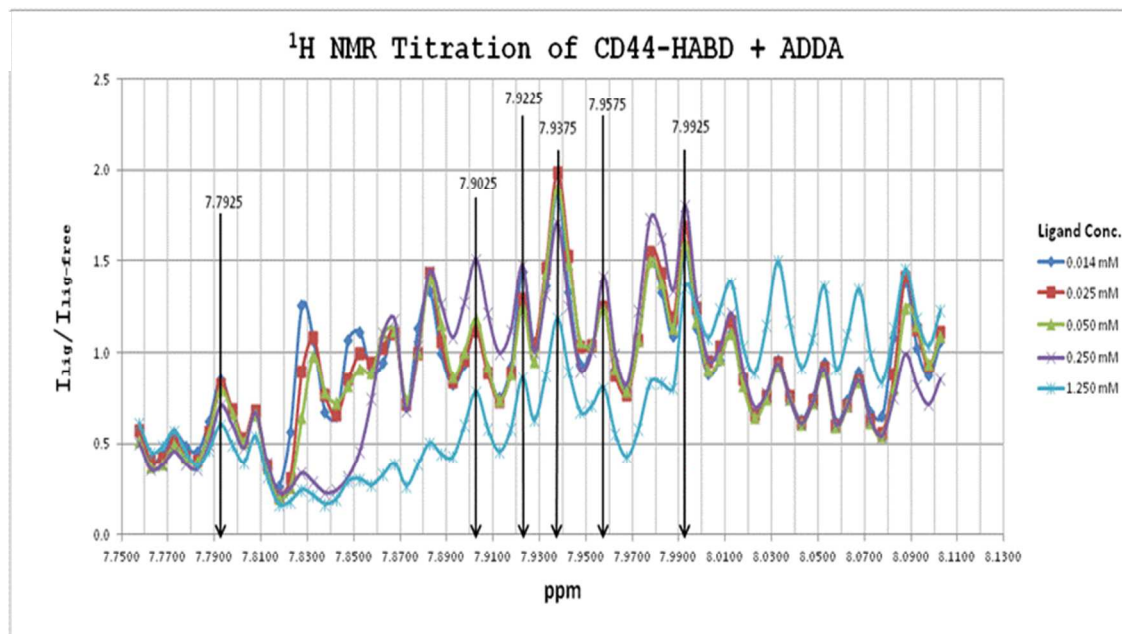


Figure S3. (A) Structure of a 7-mer hyaluronic acid (HA) built by *xLeap* (AMBER9 suite of programs); (B) Structure of the ADDA ligand. (C) Docking overlay of HA (bond representation) with ADDA ligand (CPK representation in yellow) on the CD44-HABD. Hydrogen atoms are indicated by white. Nitrogen atoms are indicated by blue. Oxygen atoms are indicated by red. Carbon atoms are indicated by gray.

(A)



(B)



(C)

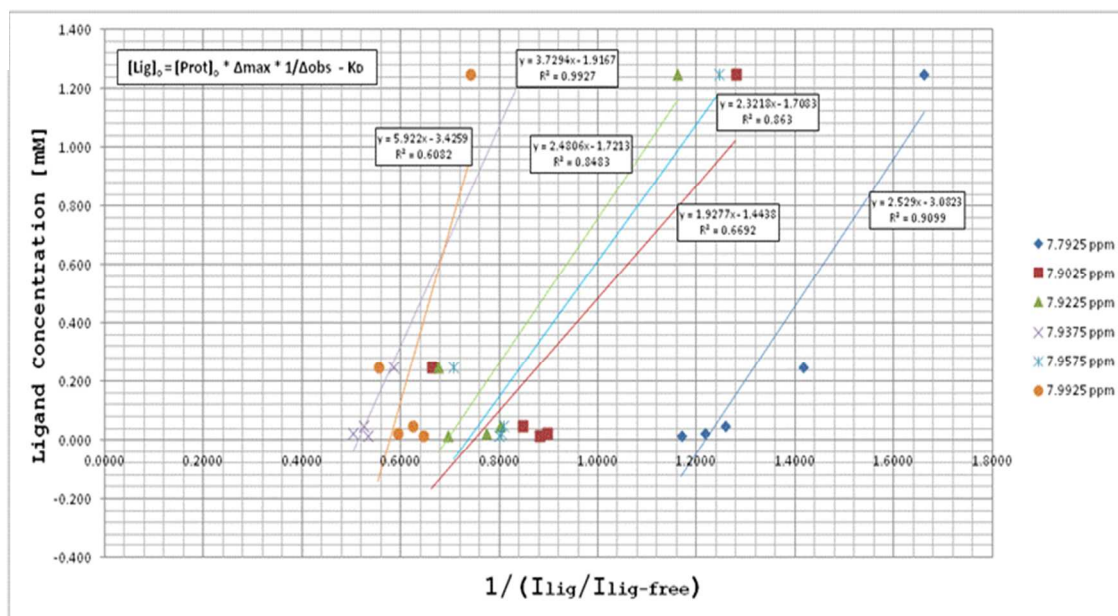


Figure S4. (A) ^1H , ^{15}N -HSQC spectrum of CD44-HABD +/- the ADDA ligand (the red peaks represent + ADDA). Assignments based on Takeda, M. et al.⁵ (B) 800 MHz ^1H -NMR titration spectra of ADDA ligand and CD44-HABD protein. $I_{\text{lig}}/I_{\text{lig-free}}$ represent the ratio between the intensity of the signal of the protein CD44-HABD with and without the ligand. Several peaks were selected (7.7925, 7.9025, 7.9225, 7.9375, 7.9575 and 7.9925 ppm) in order to obtain a statistical value of the dissociation constant of the ligand. (C) The intensities values of the peaks shown on Figure S4B were used for a linear regression fitting of the equation, $[\text{Lig}]_0 = [\text{Prot}]_0 * \Delta_{\text{max}} * 1/(I_{\text{lig}}/I_{\text{lig-free}}) - K_D$, yielding the dissociation constants of the ligand, maximum bound parameters and the correlation coefficients. Average $K_D = 2.22 \pm 0.82$ mM

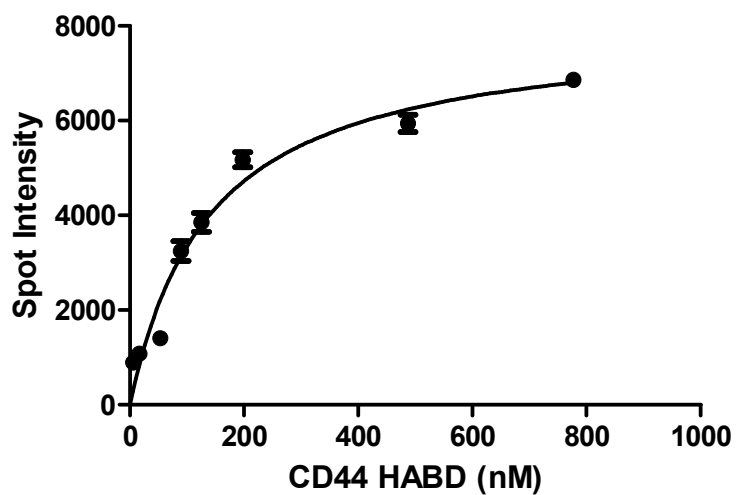


Figure S5. Saturation binding curves for binding of ADDA XA3 to CD44-HABD. Biotinylated XAs were used in the assays.

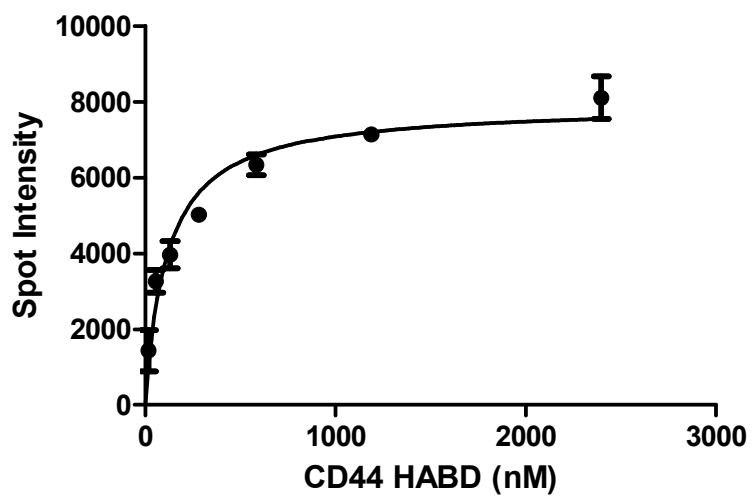


Figure S6. Saturation binding curves for binding of ADDA XA7 to CD44-HABD. Biotinylated XAs were used in the assays.

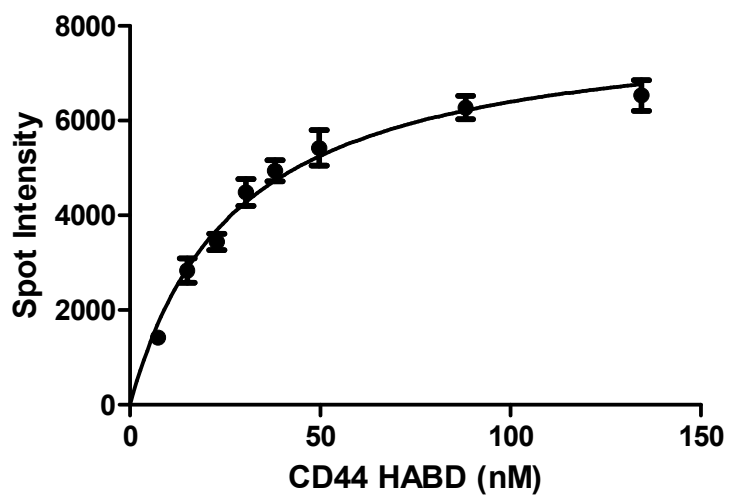


Figure S7. Saturation binding curves for binding of phospho aminodU motif 3 to CD44-HABD. Biotinylated XAs were used in the assays.

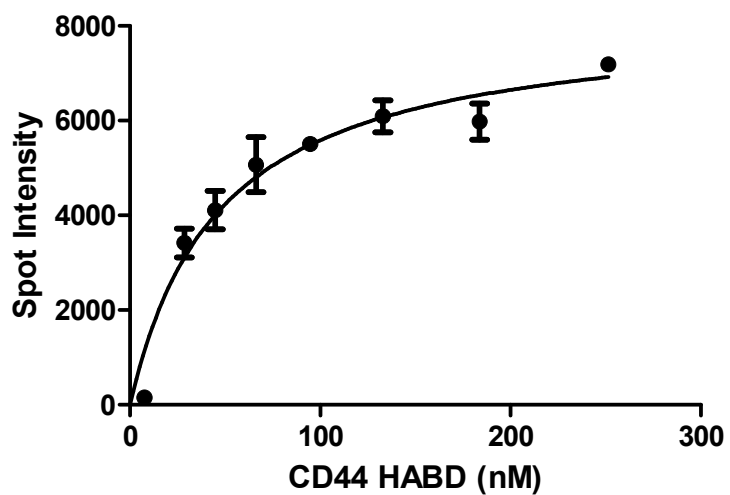


Figure S8. Saturation binding curves for binding of thiophospho ADDAdU motif 3 to CD44-HABD. Biotinylated XAs were used in the assays.

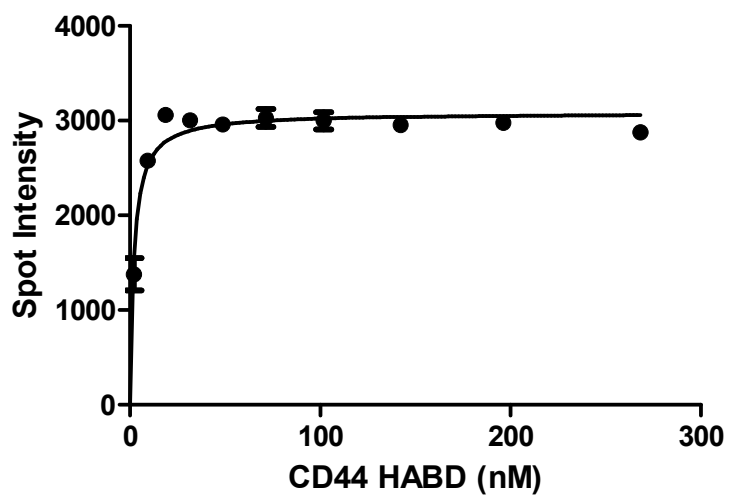


Figure S9. Saturation binding curves for binding of phospho ADDAdU motif 4 to CD44-HABD. Biotinylated XAs were used in the assays.

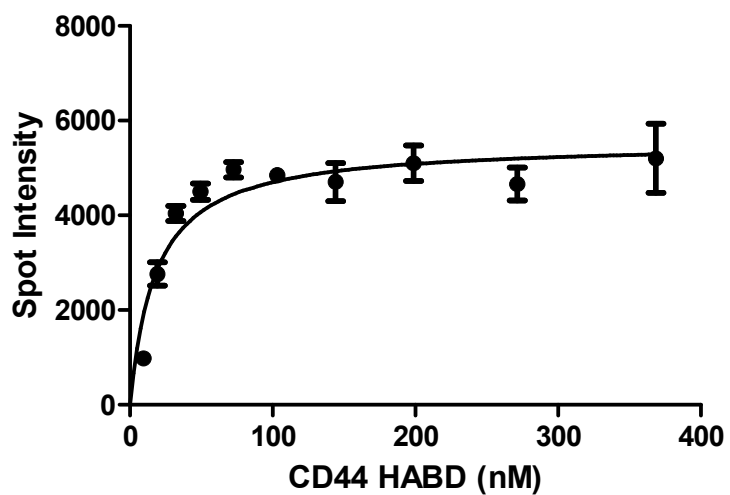


Figure S10. Saturation binding curves for binding of thiophospho aminodU motif 5 to CD44-HABD. Biotinylated XAs were used in the assays.

Table S1. XA sequences and motif sequences identified from the screen with CD44-HABD.

Name	Sequences of XAs (primers not shown)	Motif sequences
XA1/Motif 2	XXAA-GATC-XGX-TAG-GGA-ACC-AAG-ACG-AC-AG	C-XGX-TAG-GGA-ACC-AAG-ACG-A
XA2/Motif 1	XGCA-GATC-TGC-AAG-GGA-ACC-AAG-GAC-AC-TA	AAG-GGA-ACC-AAG-GAC-AC-TAC
XA3	CCAA-GGCC-TGC-AAG-GGA-ACC-AAG-TCC-AX-TA	
XA4/Motif 3	TTGG-GGCC-TGC-AAG-ACG-XCC-ATA-GAC-AC-AG	GCC-TGC-AAG-ACG-XCC-ATA-GAC-AC
XA5/Motif 5	XGCA-GAXA-CAG-TAA-ACG-XCC-ATA-GAC-AC-AG	A-GAXA-CAG-TAA-ACG-XCC-ATA-GAC-AC
XA6	TTGG-GGCC-TGC-AAG-ACG-ACC-XAA-ACG-AX-GA	
XA7	XXAA-GACG-CAX-TAA-XGA-ACC-AAG-GAC-GT-GA	
XA8	CCAA-GATC-XGX-AAG-GTA-XCC-GGG-GAC---XG	
XA9/Motif 4	XXAA-GATC-TGC-AAX-GTA-ACC-ATA-GAC-AC-AG	GATC-TGC-AAX-GTA-ACC-ATA-GAC-A
XA10	TTGG-GGCC-CAG-TAG-GTA-ACC-GGG-GAC----AG	
XA11	TTGG-GACG-----TAA-GTA-AXG-GGG-GAC-AX-GA	
XA12	XXAA-GGCC-XGX-TAA-----AXG-ATA-TCC-AC-TA	
XA13	XGCA-GATC-TGC-AAG-GGA-AAG-ATA-GAC-AC-AG	

Table S2. Equilibrium dissociation constants of selected permonothioated XAs toward CD44-HABD

Name	K_D (nM)	
	NH ₂ XA	ADDA XA
XA1	62.9 ± 10.3	108.7 ± 15.4
XA2	55.5 ± 13.4	78.3 ± 18.1
XA3	110.9 ± 18.5	139.6 ± 16.8
XA4	137.4 ± 37.4	102.7 ± 16.8
XA5	137.4 ± 37.4	124.8 ± 38.4
XA6	213.9 ± 31.6	150.0 ± 25.7
XA7	305.8 ± 69.3	116.3 ± 20.7
XA8	463.2 ± 82.4	274.0 ± 54.3
XA9	476.9 ± 89.5	555.5 ± 108.1
XA10	449.9 ± 180.3	N.A.
XA11	845.6 ± 151.4	1086.0 ± 390.5
XA12	962.6 ± 182.7	878.7 ± 134.9
XA13	1377.0 ± 237.0	312.5 ± 68.53

References

1. Banerji, S.; Day, A.J.; Kahmann, J.D. & Jackson, D.G. *Protein Expression and Purification* **1998**, *14*, 371-381
2. Abramoff, M.D.; Magelhaes, P.J. & Ram, S.J. *Biophotonics Intl.* **2004**, *11*, 36-42.
3. Case, D. A.; Darden, T.; III T. E. C.; Simmerling, C.; Wang, J.; Merz, K. M.; Wang, B.; Pearlman, D. A.; Duke, R. E.; Crowley, M.; Brozell, S.; Luo, R.; Tsui, V.; Gohlke, H.; Mongan, J.; Hornak, V.; Caldwell, J. W.; Ross, W. S.; Kollman, P. A. AMBER9 2006. University of California, San Francisco.
4. Lang, P.T.; Brozell, S.R.; Mukherjee, S.; Pettersen, E.T.; Meng, E.C.; Thomas, V.; Rizzo, R.C.; Case, D.A.; James, T.L.; Kuntz, I.D. *RNA* **2009**, *15*, 1219-1230.
5. Takeda, M.; Terasawa, H.; Sakakura, M.; Yamaguchi, Y.; Kajiwara, M.; Kawashima, H.; Miyasaka, M.; Shimada, I. *J. Biomol. NMR* **2004**, *29*(1), 97-98.
6. Engelhardt, J. P.; Gorenstein, D. G.; Luxon, B.; Herzog, N. U.S. Patent #7,576,037, 2009.