

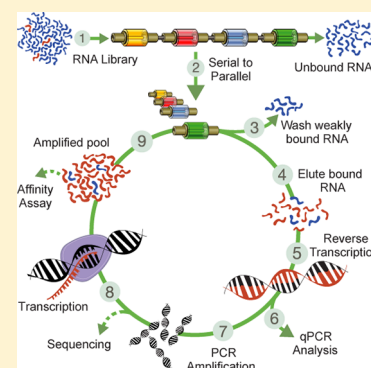
Multiplexed Microcolumn-Based Process for Efficient Selection of RNA Aptamers

David R. Latulippe,^{†,§,||} Kylan Szeto,^{†,§} Abdullah Ozer,[‡] Fabiana M. Duarte,[‡] Christopher V. Kelly,^{†,⊥} John M. Pagano,[‡] Brian S. White,^{‡,▽} David Shalloway,[‡] John T. Lis,[‡] and Harold G. Craighead^{*,†}

[†]School of Applied and Engineering Physics and [‡]Department of Molecular Biology and Genetics, Cornell University, Ithaca, New York 14853, United States

Supporting Information

ABSTRACT: We describe a reusable microcolumn and process for the efficient discovery of nucleic acid aptamers for multiple target molecules. The design of our device requires only microliter volumes of affinity chromatography resin—a condition that maximizes the enrichment of target-binding sequences over non-target-binding (i.e., background) sequences. Furthermore, the modular design of the device accommodates a multiplex aptamer selection protocol. We optimized the selection process performance using microcolumns filled with green fluorescent protein (GFP)-immobilized resin and monitoring, over a wide range of experimental conditions, the enrichment of a known GFP-binding RNA aptamer (GFPapt) against a random RNA aptamer library. We validated the multiplex approach by monitoring the enrichment of GFPapt in de novo selection experiments with GFP and other protein preparations. After only three rounds of selection, the cumulative GFPapt enrichment on the GFP-loaded resin was greater than 10^8 with no enrichment for the other nonspecific targets. We used this optimized protocol to perform a multiplex selection to two human heat shock factor (hHSF) proteins, hHSF1 and hHSF2. High-throughput sequencing was used to identify aptamers for each protein that were preferentially enriched in just three selection rounds, which were confirmed and isolated after five rounds. Gel-shift and fluorescence polarization assays showed that each aptamer binds with high-affinity ($K_D < 20$ nM) to the respective targets. The combination of our microcolumns with a multiplex approach and high-throughput sequencing enables the selection of aptamers to multiple targets in a high-throughput and efficient manner.



Nucleic acid aptamers are short (~100 nucleotide, nt) structured oligonucleotides that have been selected from large sequence-diverse libraries and shown to display high-affinity and specificity for a wide range of targets ranging from simple metal ions¹ to complex surface proteins on living cells.² This combination of properties has led to growing interest in applications of aptamers in fields including therapeutics, chemical analysis, biotechnology, chemical separations, and environmental diagnostics.³ Aptamers are identified from large libraries of random nucleic acid sequences via an iterative in vitro process called SELEX (systematic evolution of ligands by exponential enrichment).^{4–6} A typical SELEX round includes the following three steps: (i) binding, incubation of the library with the target; (ii) partitioning, separation of target-bound library sequences from unbound ones; and (iii) amplification, generation of a new pool of nucleic acids by making multiple copies of the sequences that bound to the target. These steps are then repeated in an iterative fashion to obtain an enriched pool and the target binding aptamers are identified via cloning and sequencing processes.

Different selection strategies have been developed to separate or partition the free and target-bound sequences, a critical step to ensure the successful identification of only the strongest binding aptamers. Affinity chromatography is one such

partitioning strategy that uses specific binding onto resin-immobilized targets to purify macromolecular solutes from dilute solutions.⁷ It is a well-documented aptamer selection technique, given that it can achieve a level of purification greater than 95% in a single step and that numerous types of resin media are available to bind a wide variety of targets. However, there is limited understanding of the relationship governing the process parameters (target loading, resin volume, etc.) and the selection quality, and thus, many selection rounds (typically 12–15) are required to identify aptamers with the desired specificity and affinity for the target. This approach is particularly challenging when working with RNA libraries because it takes approximately 2 days just to complete the amplification step. Some work has been done to parallelize the use of libraries and the selection process to multiple targets in order to save time and reagents.^{8,9}

Affinity chromatography-based selections are typically done in two different modes of operation. In batch mode, a small amount of target-immobilized resin (~20–200 μ L) is incubated with the nucleic acid library, or alternatively, target-

Received: January 11, 2013

Accepted: February 11, 2013

Published: February 11, 2013

free resin is incubated with a mixed suspension of target and nucleic acid library.^{10–12} Any unbound sequences in the supernatant are removed, and the target-bound sequences remaining on the resin are then exposed to other solutions for the subsequent processing steps. For this approach, the entire selection process is quite laborious, because each step must be done manually and repeated several times. This is especially true when multiple targets are considered for selection. The second mode of operation, flow mode, uses small columns (~0.5–3 mL) packed with resin.^{1,13,14} The primary advantage of this approach is that the resin is physically confined within the column, allowing all of the selection steps to be automated by use of pumps and/or centrifuges and thus completed more efficiently than the batch strategy. This approach was used in one of the landmark papers on aptamers; Ellington and Szostak⁵ used a 3.5 mL column filled with dye-immobilized resin. However, there are a few limitations to this approach. The standard columns that have been used previously are not practical for the simultaneous selection of aptamers for multiple targets. These columns require more resin than the batch mode, as well as more of the immobilized target (e.g., protein), which can be both limiting and expensive. Thus, with the current affinity chromatography-based strategies, there is a noticeable lack of means to rapidly screen for aptamers to multiple targets in a high-throughput and efficient manner.

To address these limitations, we developed a process utilizing reconfigurable microcolumns of varying capacity for selecting RNA aptamers. The microcolumns require only microliter volumes of affinity chromatography resin (~2–50 μL), they can be easily assembled in various configurations to accommodate multiple targets, and they can be easily integrated with common laboratory equipment. In addition, these microcolumns are not restricted to RNA and other nucleic acid aptamer selections but are also suitable for other affinity chromatography needs where small column volumes are desired. The assembly of microcolumns and aptamer selection process are shown in Figure 1 with the experimental details provided below.

We first evaluated the design space of our process using computer simulations of the binding kinetics of a model library over a wide range of experimental conditions. Next, we evaluated the performance of our microcolumns at those same conditions by monitoring the behavior of a known RNA aptamer (GFPapt) that binds tightly to green fluorescent protein (GFP) and its derivatives.¹⁴ Our results show that predictions based on simple kinetics fail to reproduce the behavior of low-affinity binders under flow, suggesting that typical SELEX processes based on theoretical kinetics are likely to be far from optimal. Furthermore, we observed the best performances at protein concentrations 100 times less than the capacity of the resin. We also empirically validated a multiplex approach by monitoring the enrichment of GFPapt in de novo selection experiments with GFP, two nonrelated protein preparations, and an empty microcolumn. To examine the utility of our device, multiplex and in-line negative selections were performed on two human heat shock factor (hHSF) proteins, hHSF1 and hHSF2. High-throughput sequencing was used to identify enriched candidate aptamer sequences for hHSF1 and hHSF2. Fluorescence electrophoretic mobility shift assays (F-EMSA) as well as fluorescence polarization (FP) confirmed the selection of novel high-affinity aptamers for hHSF1 and hHSF2.

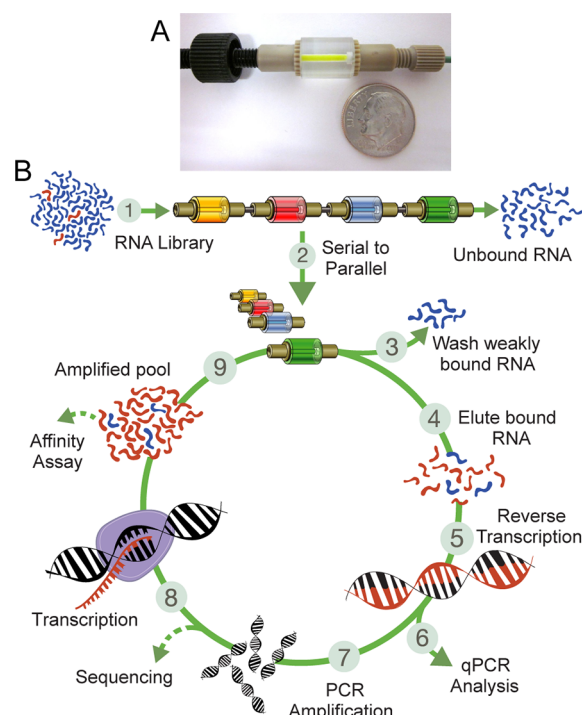


Figure 1. Aptamer selection workflow for multiple targets by use of microcolumns. (A) Microcolumn filled with 10 μL of GFP-immobilized chromatography resin. (B) Multiplexed selection of RNA aptamers: (1) The starting RNA library is dynamically loaded onto multiple microcolumns that are connected in a serial configuration. (2) The devices are rearranged into a parallel configuration and the subsequent cycles in the process are done independently but simultaneously. (3) Unbound and weakly bound RNAs are washed away. (4) The remaining bound RNAs are eluted from each column separately. (5) The RNA molecules are reverse-transcribed into cDNA and (6) a small fraction is analyzed via qPCR. (7) The remaining cDNA is PCR-amplified and then (8) transcribed back into RNA to make a new amplified pool for (9) the next selection round. The steps shown with dashed arrows are optional and are not necessarily done in each round.

MATERIALS AND METHODS

Selection Simulations. Computational simulations were performed with a custom-made Matlab routine to test a wide variety of experimental parameters. The differential equations of association and dissociation kinetics (first-order with respect to each species) were numerically integrated with respect to time, distance along the microfluidic channel, and each aptamer type within the modeled library. Additional details are available in Supporting Information

Preparation of Protein-Immobilized Resins. For each selection round, a fresh batch of protein-bound resin was prepared. Nickel–nitrilotriacetic acid (Ni-NTA) Superflow or glutathione–agarose resins were extensively washed with binding buffer [Ni-NTA binding buffer contained 25 mM tris(hydroxymethyl)aminomethane (Tris), pH 8.0, 100 mM NaCl, 25 mM KCl, and 5 mM MgCl_2 ; glutathione binding buffer contained 10 mM *N*-2-hydroxyethylpiperazine-*N'*-ethanesulfonic acid (HEPES)–KOH pH 7.6, 125 mM NaCl, 25 mM KCl, 1 mM MgCl_2 , and 0.02% Tween-20] to remove any residual storage components. Hexahistidine- or GST-tagged proteins were prepared as described in the Supporting Information and immobilized onto the washed resin at 4 $^{\circ}\text{C}$ with constant mixing. The protein-bound resin was then

degassed in a vacuum desiccator for approximately 20 min and carefully pipetted into the device (Figure 1A).

Nucleic Acid Library and GFP-Binding Aptamer. A random library containing $\sim 5 \times 10^{15}$ sequences of 120-nucleotide (nt) DNA templates, hereafter referred to as N70 library, was chemically synthesized by GenScript. It consists of a 70-nt random region flanked by two constant regions as described elsewhere.¹⁵ The GFP-binding RNA aptamer used in this work was previously identified as AP3-1 and characterized elsewhere.¹⁴ Details on the library preparation and oligos used in this work are available in Supporting Information.

Microcolumn Selection and Amplification Protocol.

All of the solutions were degassed prior to use and introduced into the microcolumns via a standard syringe pump (Harvard Apparatus). First, yeast tRNA (Invitrogen) in binding buffer was introduced to block any possible nonspecific RNA binding sites. For each loading step, the RNA library was diluted in 1 mL of RNase-free binding buffer, heat-denatured at 60 °C for 5 min, renatured by cooling down to room temperature while degassing, and then spiked with 200 units of RNase inhibitor (Invitrogen). A 10 μ L aliquot was collected and used as a standard for the quantitative polymerase chain reaction (qPCR) analysis. Each device was then washed with 3 mL of binding buffer (supplemented with 10 mM imidazole for selections with Ni-NTA resin) to remove unbound RNA. Finally, the RNA–protein complexes were eluted from individual microcolumns by flowing elution buffer [binding buffer + 50 mM ethylenediaminetetraacetic acid (EDTA)] at a rate of 50 μ L/min for 6 min. Eluted RNA and the input samples were phenol/chloroform-extracted and ethanol-precipitated together with 1 μ L of GlycoBlue (Ambion) and 40 μ g of yeast tRNA (Invitrogen), and the resulting pellet was resuspended in RNase-free water.

Both the resuspended pools and standards were reverse-transcribed with Moloney murine leukemia virus reverse transcriptase (MMLV-RT). For the optimization experiments, $\sim 10\%$ of the selected pool and the input sample were reverse-transcribed in a separate reaction for GFPapt quantitation. Residual RNA was eliminated by treating the samples with RNase H (Ambion). A small amount (less than 5%) of the cDNA product was analyzed on a LightCycler 480 qPCR instrument (Roche) to determine the amount of RNA library and GFPapt that was retained on each device. Two different sets of oligos (see Supporting Information for details) were used to independently evaluate the amount of the RNA library/pool and GFPapt, respectively. The cDNA samples from each round were PCR-amplified and then subjected to phenol/chloroform and chloroform extractions and a final purification step on DNA Clean & Concentrator (Zymo Research) spin columns. A fraction of the purified PCR product was used to make the RNA pool for the next round of SELEX. A typical 72 μ L transcription reaction consisted of 500 ng of DNA, 546 pmol of each ribonucleoside triphosphate (rNTP, Sigma), T7 RNA polymerase, 72 units of RNase inhibitor (Invitrogen), and 0.12 unit of yeast inorganic pyrophosphatase (New England Biolabs). The reactions were incubated at 37 °C for various times depending on the desired amount of RNA for the next selection round. The resulting RNA pool was treated with DNase I (Invitrogen) to remove the template DNA, verified by denaturing polyacrylamide gel electrophoresis (PAGE) for length and purity, and finally quantified by Qubit BR RNA assay (Invitrogen).

High-Throughput Sequencing. A small amount of the purified PCR product from each target pool for various selection rounds (e.g., hHSF1 round 5) was PCR-amplified by use of primers that contain a unique 6 nt barcode and the necessary adapters for the HiSeq 2000 (Illumina) sequencing platform. Sequences of the primers and the barcodes are available upon request. Additional details on the sequencing data filtering and clustering analyses are given in the Supporting Information.

Fluorescence Electrophoretic Mobility Shift Assay.

Candidate aptamer sequences were amplified from the final round 5 pool and prepared from sequence-verified plasmid constructs (see Supporting Information). The candidate aptamers were 3'-end-labeled with fluorescein 5-thiosemicarbazide (Invitrogen) as described previously.¹⁶ Binding reactions (50 μ L) were prepared with 2 nM fluorescently labeled RNA and decreasing amounts of protein (2000 to 0 nM) in binding buffer containing 0.01% IGEPAL CA-630, 10 μ g/mL yeast tRNA, and 3 units of RNase inhibitor (Invitrogen). Reactions were incubated at room temperature for 2 h, spiked with 6 \times loading dye, and then loaded into the wells of a refrigerated 1.5% agarose gel prepared with 0.5 \times Tris–borate–EDTA (TBE) buffer with 1 mM MgCl₂. The gel was run for 80 min at 100 V in refrigerated 0.5 \times TBE. Images were acquired at the fluorescein scan settings on a Typhoon 9400 imager (GE Healthcare Life Sciences). The resulting bands were quantified with ImageQuant software and the data were fit to the Hill equation by use of Igor (Wavemetrics) to estimate the equilibrium dissociation constant (K_D).

RESULTS

Fabrication of Microcolumns for Selection of Aptamers.

Our microcolumns were assembled from both custom-fabricated and commercially available parts. The column consists of a transparent biocompatible plastic rod fitted with a porous frit that retains the resin in the device (Figure 1A). By varying both the length and internal diameter of the column, we were able to fabricate columns with a range of volume capacities from 2 to 50 μ L. NanoPorts (IDEX Health and Science) that accommodate standard tubing connectors were bonded to either end of the column. Overall, this design has a number of important features: simple union connectors can be used to arrange multiple microcolumns into various configurations, the dead volume between the devices is minimized and is generally less than 1 μ L, and they can be connected to common laboratory equipment to automate the selection steps. To perform the multiple target aptamer selection, we developed a workflow process wherein a set of microcolumns, each prefilled with a different target-immobilized resin, are arranged into a serial configuration (Figure 1B). With our pump system, a typical arrangement could contain up to 10 parallelized assemblies of devices, but the general approach can easily be scaled up for a larger number of parallel processes. A single aliquot of the starting random library is then flowed through these devices, allowing the target binding aptamers to be captured on the resin within each individual column. The library molecules that do not bind to any target are discarded and then the individual microcolumns are disconnected and reorganized into a parallel configuration (Figure 1B). This arrangement allows for specific elutions from each target and thus separate processing of only the bound sequences to create target-specific amplified pools for the next selection round. Note that, following reverse transcription of

the RNA aptamer into cDNA, we used quantitative PCR to determine the amount of nucleic acid recovered from each device. This information was used to determine the optimal number of PCR cycles and thus minimize the chance of amplification artifacts.¹⁷ Although Figure 1B shows the microcolumns arranged exclusively in a parallel configuration after the first selection round, it is also possible to use a serial configuration in later rounds. This arrangement would allow for negative or counter selections to be done simultaneously with the selection step to enhance specificity for the target.

The size of our microcolumns was chosen for primarily two reasons. First, they are small enough that they require only small amounts of material for each selection round, yet their internal dimensions are sufficiently large enough that they could be easily filled with a variety of different resins. Second, we simulated the binding kinetics of a model library binding to a model target molecule within our device and discovered that aptamers with strong binding affinities for the target (i.e., equilibrium dissociation constant $K_D = 0.5$ nM) were preferentially retained at the input end (i.e., in the first few microliters) of the microfluidic column (Figure 2). Aptamers

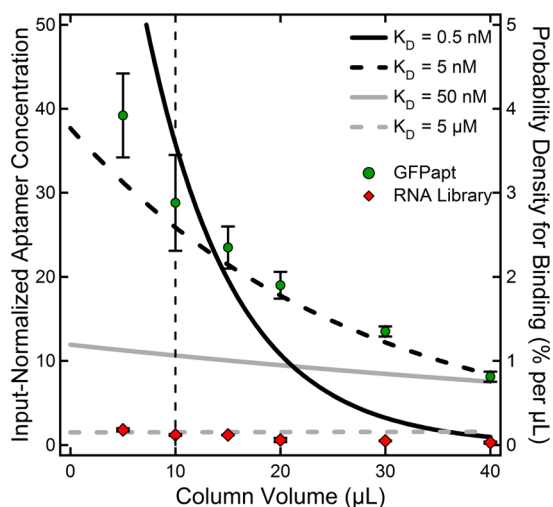


Figure 2. Dependence of aptamer recovery on microcolumn volume. Solid and dashed lines are the simulation results for the input-normalized concentrations of aptamers with various binding affinities (K_D ranging from 0.5 nM to 5 μ M) as a function of microcolumn volume. Data points are the experimentally measured binding densities, given as percent of the total, of high-affinity GFPapt (green solid circles) and low-affinity, nonspecific-binding N70 random library (red solid diamonds) as a function of microcolumn volume. Input-normalized aptamer concentrations are defined as the bound aptamer concentration at a point along the column divided by the initial aptamer concentration. Probability densities for binding are defined as the probability per unit volume for molecules to bind in the vicinity at a point along the column. The two y-axes are related by the total loaded sample volume. The dashed vertical line at 10 μ L marks the volume of the devices used for the optimization experiments.

with weaker binding affinities for the target ($K_D \geq 50$ nM) were distributed almost uniformly throughout the microfluidic column with concentrations nearly identical to the input concentration. Therefore, smaller columns increase the mean density of strong binders, a condition that would require fewer selection rounds to identify aptamer candidates. A previous study used an affinity capillary column that was physically cut

into smaller pieces to isolate the highest affinity aptamers in the earliest column segments.¹⁸

We evaluated experimentally the binding distribution of an RNA aptamer, hereafter referred to as GFPapt, on microcolumns that were filled with various amounts of GFP-immobilized resin. It had been previously shown that this aptamer has a strong binding affinity ($K_D \sim 5$ nM) for GFP¹⁴ and thus serves as a model molecule for the high-affinity target-binding aptamers that are presumed to be present in random libraries. In order to determine the amount of nonspecific, low-affinity (i.e., “background”) binding within the microcolumn, a small amount of the random RNA library (~ 5 pmol) was included in addition to the GFPapt (0.064 pmol). We used 0.6 μ g of GFP/ μ L of resin, binding and washing flow rates of 100 μ L/min, and qPCR assays to quantify the amount of both the GFPapt and the random library captured on the microcolumn. The experimental results are shown in Figure 2 as a percentage density of the amount of both the GFPapt and the N70 random library loaded onto the device. The experimental results are well fit by the simulations for both high-affinity ($K_D = 5$ nM for GFPapt) and low-affinity ($K_D \geq 10$ μ M for the N70 random library) binding.

Optimization of Aptamer Selection Conditions. To optimize the performance of our devices and to further test our simulation predictions, we used microcolumns filled with GFP-immobilized Ni-NTA resin to evaluate the binding behavior of GFPapt and the random N70 library over a wide range of experimental conditions. The results are shown as a percentage of the amount loaded onto the device (Figures 3A and 3C). For each condition, the GFPapt enrichment was calculated as the ratio of the percent amounts of GFPapt to random library (Figures 3B and 3D).

Affinity chromatography resins were developed primarily for protein purification applications and thus are capable of binding relatively large amounts of target molecules; the reported binding capacity of the resin used in our experiments is 20–50 μ g of protein/ μ L of resin. Despite the widespread use of these resins for aptamer selections, there is no information available on how this parameter affects the selection performance, since none of the previous studies optimized the resin binding conditions.⁷ In order to determine the optimum amount of bound target, we prepared five different batches of resin with varying amounts of immobilized GFP, from 0.024 to 15 μ g of protein/ μ L of resin in 5-fold increments. The amount of GFPapt captured within the device was strongly dependent on the amount of GFP target on the resin (Figure 3A). Interestingly, we found that the highest recovery ($\sim 40\%$) was obtained at the intermediate value of 0.6 μ g of protein/ μ L of resin (~ 15 μ M). Library recovery was essentially independent of the amount of target, and so the GFPapt enrichment was also maximized at the same intermediate value (Figure 3B). We had initially hypothesized that saturating the resin with GFP would maximize the number of target binding aptamer molecules while also minimizing nonspecific binding sites on the resin surface, yielding the highest GFPapt enrichment. Kinetic simulations done by a similar approach to that described for determining the optimal device size, except with a higher fraction of strong-binding aptamers ($\sim 1\%$) to match the experimental conditions, correctly predicted that there was no effect of target amount on recovery of the random library (Figure S-1B, Supporting Information). However, the simulations predicted that the highest recovery of the strong-binding aptamer would occur at the conditions with the highest amount

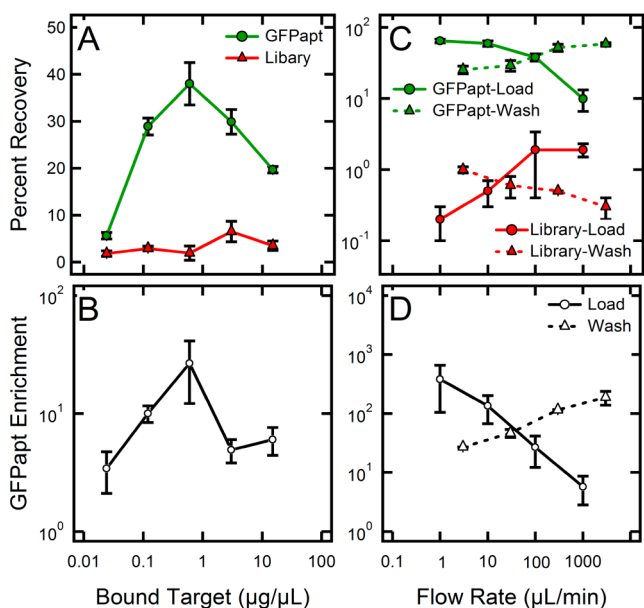


Figure 3. Optimization of microcolumn-based selection. (A) Percent recovery of GFPapt (green) and N70 random library (red) and (B) GFPapt enrichment for different amounts of GFP immobilized onto Ni-NTA resin in a 10 μL microcolumn; loading and washing flow rates were 100 $\mu\text{L}/\text{min}$. (C) Percent recovery of GFPapt (green) and N70 random library (red) and (D) GFPapt enrichment for different flow rates; the amount of GFP immobilized onto Ni-NTA resin was 0.6 $\mu\text{g}/\mu\text{L}$. Solid lines show the effect of loading flow rate at a fixed washing flow rate of 100 $\mu\text{L}/\text{min}$; dashed lines show the effect of washing flow rate at a fixed loading flow rate of 100 $\mu\text{L}/\text{min}$. Error bars represent the standard deviation of triplicate experiments and measurements at each condition.

of bound target (Figure S-1A, Supporting Information). We believe that this discrepancy between our experimental observations and simulation results is primarily due to steric hindrance, where macromolecular crowding effects on the resin surface decrease the binding affinity of the specific aptamer sequence; a similar phenomenon has been reported for the binding of soluble proteins to surface lipid vesicles.¹⁹ Formation of the aptamer–target complex could also hinder the binding of other aptamer molecules. Our results support the use of concentrations below a critical packing density that is determined by the larger of the two biomolecules. All of the subsequent optimization experiments were conducted with resin prepared at the optimum value of 0.6 μg of protein/ μL of resin.

It is well known that the best affinity chromatography performance is realized when the loading step is operated at the lowest possible flow rates to approach equilibrium conditions. However, there is a practical limitation to the experiment time. In order to determine the optimum condition while also keeping within a practical experimental timeline, we varied the loading flow rate from 1 to 1000 $\mu\text{L}/\text{min}$. The recoveries of GFPapt and library were both strongly dependent on the flow rate but with considerably different trends (Figure 3C). At the lowest flow rate, we observed the highest GFPapt recovery, the lowest library recovery, and thus the best GFPapt enrichment. By operating the process at higher flow rates, we observed a gradual decrease in performance as measured by a decrease in GFPapt enrichment (Figure 3D). Our kinetic simulations correctly predicted that the highest recovery of the strong-binding GFPapt molecule would be obtained at the lowest flow

rate (Figure S-2A, Supporting Information). However, the simulations also predicted the same trend of decreasing recovery with increasing flow rate for the random library (Figure S-2B, Supporting Information). This disagreement between simulation and experimental results for the library is discussed below.

After completion of the loading step, a fixed-volume washing step was employed to improve the separation performance by removing any unbound or weakly bound sequences. We varied the washing flow rate from 3 $\mu\text{L}/\text{min}$ to 3 mL/min . We observed similar results to those seen before in that the recoveries of both GFPapt and library had different trends with increasing flow rate. Whereas the GFPapt recovery increased at higher flow rates, the library recovery decreased (Figure 3C). Thus, the best GFPapt enrichment was obtained at the highest washing flow rate (Figure 3D). Together, all of the optimization experiments enabled us to choose the optimal experimental conditions that maximize the enrichment of strong-binding aptamers for each selection round, while keeping within practical experimental and time constraints. These conditions are particularly important for the earliest selection rounds when there are only a few copies of each aptamer sequence. Our results also revealed the importance of empirical validation and characterization of different selection conditions, since the kinetic simulations were unable to properly predict all the experimental trends. For example, simulation results for the washing step predicted a gradual increase in the recovery of random library (Figure S-3, Supporting Information)—the exact opposite trend to that seen experimentally. In our simulations, the behavior of each species was determined by the on- and off-rate kinetic constants. They do not include other phenomena such as shear or pressure-related flow effects that could preferentially affect the behavior of weak binders and the ultimate separation performance of our microcolumns.

After fully characterizing the operating conditions to maximize the enrichment of strongly binding aptamers in our microcolumns, we then proceeded to validate our selection strategy empirically by monitoring the enrichment of GFPapt molecules in de novo selections with multiple protein targets, including GFP. For each round, three microcolumns were filled with Ni-NTA resin that had been preimmobilized with either GFP or two unrelated proteins, CHK2 and UBLCP1. The latter two were chosen because they have similar size and charge properties as the specific target GFP and were readily available with hexahistidine affinity tags (Table S-1, Supporting Information). An empty microcolumn was also included as a control to enable us to discriminate target-binding from device bias.

For the first selection round, four devices were arranged in a serial configuration in the following order: empty, UBLCP1, CHK2, GFP. The GFP target was put at the end so that all of the GFPapt molecules that were combined with the random RNA library would have an opportunity to bind to the other targets; the loading solution for the first selection round contained 40 nmol of the random RNA library (~ 5 copies of each sequence in the 5×10^{15} library) and 6.4 fmol of GFPapt in 1 mL of binding buffer. Thus, the initial molar ratio of library to GFPapt molecules was greater than 6 million to one. Enrichment of GFPapt was determined for each microcolumn from qPCR analyses of the eluted samples (Figure 4). Only the microcolumn filled with GFP-loaded resin showed significant GFPapt enrichment. In the first round, the recoveries of the

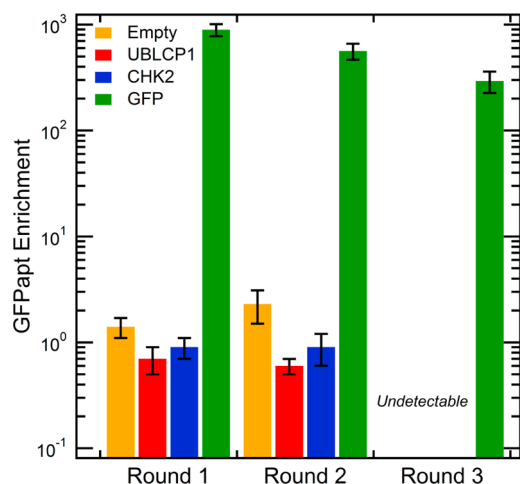


Figure 4. Validation of specific aptamer-target enrichment for multiplex SELEX. UBLCP1, CHK2, and GFP were preimmobilized on Ni-NTA resin at a density of $0.6 \mu\text{g}$ of protein/ μL of resin and an empty microcolumn was also included as a control. The loading flow rate for all the rounds was $1 \mu\text{L}/\text{min}$. The enrichment of GFPapt was monitored on all four microcolumns for three consecutive selection rounds. Results are presented in the same order that the four devices were arranged in the serial configuration used in round 1. GFPapt enrichment was calculated as the ratio of the percent amounts of GFPapt to random library. The error bars represent the standard deviation of triplicate measurements for each microcolumn.

GFPapt and the random library were 20% and 0.022% of the corresponding inputs, respectively, yielding a GFPapt enrichment of approximately 900-fold. However, for the other three devices, the GFPapt enrichments were all near unity, indicating that there was no affinity for GFPapt over the random library. To decrease the material used in the selections, for the next two rounds the amount of the amplified RNA pool loaded onto each device was decreased by 20-fold from the amount used in the previous round. An appropriate amount of GFPapt was “spiked-in” with the amplified RNA pool to maintain the same molar ratio that was recovered in the previous round. After round 3, the amounts of GFPapt that were recovered on the three non-GFP devices were well below the qPCR detection limit, and therefore, no GFPapt enrichment results could be obtained. However, for the GFP device the GFPapt enrichments were much greater than 100-fold per selection round, giving rise to a cumulative enrichment of over 10^8 -fold. At the end of round 3, approximately 95% of the GFP selected pool was composed of GFPapt.

Microcolumn SELEX for Human Heat Shock Factor Proteins. To conclusively demonstrate the multiplex utility of the microcolumns, we performed multiplex SELEX for a set of GST-tagged proteins: hHSF1, hHSF2, and four other related proteins. hHSF1 and hHSF2 are transcription factors that regulate stress response, including heat shock in human cells, and hHSF1 plays a critical role during tumor formation and maintenance.²⁰ However, important questions remain unanswered regarding their molecular interactions and the specific mechanisms that are used to execute their functions. An RNA aptamer against the *Drosophila melanogaster* HSF (dHSF) has been selected previously and used to inhibit the binding and recruitment of dHSF to the promoters of heat shock genes.^{21,22} However, only moderate cross-reactivity with the hHSF proteins limits the use of this aptamer in functional studies in human cells.²³ Given the biological importance of HSF, we

decided to select for aptamers directly to hHSF1 and hHSF2, which can be used as inhibitory tools to dissect the mechanisms of actions of these proteins.

Each protein was preimmobilized onto glutathione-agarose resin at approximately $1 \mu\text{g}$ of protein/ μL of resin. The GST tag was also included as a control and thus seven $20 \mu\text{L}$ microcolumns were arranged in a serial configuration for round 1. Following the procedure outlined above, another aliquot of the starting RNA library (~ 5 copies of each sequence in the 5×10^{15} library) was loaded onto the seven-microcolumn assembly and then separated into a parallel arrangement for all the subsequent rounds.

A total of five selection rounds were completed. For rounds 2 through 5, in-line negative selections were done by connecting a $10 \mu\text{L}$ microcolumn filled with GST-immobilized resin to the inlet of each of the six microcolumns for the GST-tagged proteins. These precolumns were removed after the loading step for the subsequent wash and elution of the target protein-bound aptamers from each microcolumn.

We analyzed the sequences from the RNA pools from rounds 3 and 5 for both hHSF1 and hHSF2 by high-throughput sequencing.^{24,25} The total number of sequencing reads per pool ranged from 6 million to 9 million. For both proteins, there was a noticeable shift toward higher multiplicity values from round 3 to round 5. In round 3, the top 20 highest multiplicity sequences for each protein represented only $\sim 0.04\%$ of the total pool. However, in round 5, the top 20 sequences represented 85.0% and 76.5% of the hHSF1 and hHSF2 selected pools, respectively (Table S-2, Supporting Information). In addition, of the top 20 highest multiplicity sequences in round 3, between a quarter and a half of them were also among the top 20 highest multiplicity sequences in round 5 for hHSF2 and hHSF1. The detection of enriched candidate aptamer sequences in earlier selection rounds was possible because of the high-throughput sequencing, which allowed us to select candidate aptamers for subsequent analysis.

We decided to investigate the potential binding of one of the top candidate sequences for hHSF1 and hHSF2. We chose the highest ranked sequences from the round 5 pools that were also highly ranked in round 3. For hHSF1, this was the first-ranked sequence in round 5, hereafter referred to as hHSF1-R5-1, which was also the 11th-ranked sequence in round 3 (Table S-2, Supporting Information). For hHSF2, this was the second-ranked sequence in round 5, hereafter referred to as hHSF2-R5-2, and also the sixth-ranked sequence in Round 3 (Table S-2, Supporting Information). The full-length sequences and predicted structures of these two aptamer candidates are available in the Supporting Information (Figure S-4). These two candidates represented 13.4% and 6.8% of the corresponding round 5 pools. The full-length aptamer candidates (including the constant regions) were PCR-amplified from their respective pools by use of a candidate-specific forward oligo and the reverse constant region oligo, and then cloned into plasmid vectors to obtain a pure template.

The putative RNA aptamers were fluorescently end-labeled and then tested for binding to their hHSF targets by F-EMSA and FP assays.¹⁶ An image of a typical F-EMSA result is shown in Figure 5A for hHSF1-R5-1 aptamer binding to hHSF1 protein. The fraction of bound aptamer was calculated as a function of protein concentration and then plotted as shown in Figures 5B and 5C for various aptamer-protein pairings; K_D values were determined by fitting each data set to the Hill equation. Overall, both aptamers showed high-affinity binding

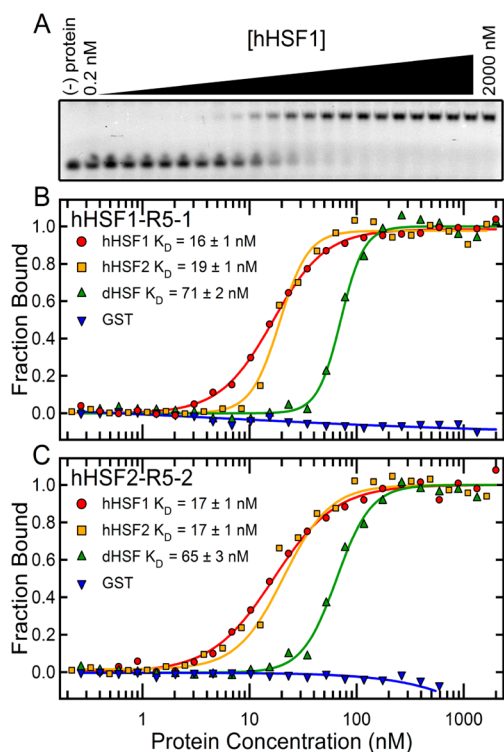


Figure 5. Evaluation of candidate aptamers binding to target proteins. (A) Typical F-EMSA results for hHSF1-R5-1 aptamer binding to a two-thirds dilution series (from 2000 nM to 0.2 nM) of hHSF1 protein. (B, C) Binding curves measured by F-EMSA for hHSF1-R5-1 and hHSF2-R5-2 aptamers to hHSF1, hHSF2, dHSF, and GST tag. The same dilution series in panel A was used in panels B and C. The solid lines are the best fits of the Hill equation to the experimental data for each aptamer–target pair with the appropriate K_D values given in the figure legends.

to hHSF1 and hHSF2 ($K_D < 20$ nM). Interestingly, both aptamers also bound to hexahistidine-tagged dHSF, although slightly more weakly ($K_D \sim 70$ nM), and no binding was observed to the GST-tag alone. The F-EMSA results were confirmed by the FP assays (Figures S-5 and S-6, Supporting Information). Thus, the observed binding is not due to the affinity tags on the protein targets but rather to specific domains of the targets themselves. Given that the highest degree of sequence similarity between hHSF1, hHSF2, and dHSF is in the DNA binding and trimerization domains (DBD-TD)²⁶ and that the previously selected dHSF aptamer was found to bind the DBD-TD of dHSF,²¹ we predict these novel hHSF aptamers are likely to bind the HSF proteins in a similar fashion. Contrary to their functional similarity, these two aptamers did not show any similarity in secondary structure as predicted by mFold²⁷ (Figure S-4, Supporting Information). Although beyond the scope of the present work, the detailed mechanism of these and other potential aptamers' binding to HSF proteins as well as the consequences of binding await further study. However, successful selection of two distinct high-affinity aptamers, hHSF1-R5-1 and hHSF2-R5-2, targeting two closely related proteins, hHSF1 and hHSF2 respectively, in a single selection demonstrates that our microcolumn-based SELEX technology is capable of yielding high-affinity aptamers ($K_D < 20$ nM) in as little as five rounds of selection, whereas most conventional SELEX methods require typically 12 rounds of selection.⁷

CONCLUSIONS

We have developed a microcolumn-based technology for the efficient selection of nucleic acid aptamers for multiple targets. Our microcolumns have a number of advantages over other chromatography-based processes. First, they can be readily assembled in various configurations to accommodate multiple targets during the selection step; as a proof of principle, we used a single aliquot of starting RNA aptamer library to perform selections for seven different immobilized target proteins in separate microcolumns. Second, as either assembled or disassembled units, they do not require any specialized equipment to perform the selection step; we used a multitrack syringe pump to control the solution flow rates during the selection step. Third, they require very small amounts of resin and target molecules. For this study, we focused on the selection of RNA aptamers; however, our approach would also work for DNA aptamers. The multiplex selection presented in this work is simpler and thus easier to perform than previous microfluidic techniques that use sol-gel chemistry.⁸ Also, compared to filter-plate-based methods,⁹ our modular columns allow “counter-selections” to be done simultaneously with the selection step to enhance target specificity.

The overall performance of our devices was evaluated in two different sets of experiments. In the first, we looked at a single selection step and optimized its performance by monitoring the enrichment of GFPapt, a known RNA aptamer for GFP. The performance was strongly dependent on the amount of target immobilized onto the resin and the flow rates for both the loading and washing steps. In the second, we used the optimal conditions as part of a selection and amplification strategy to verify specific GFPapt partitioning over three protein preparations (UBLCP1, CHK2, and GFP), followed by a complete multiplex selection against hHSF1 and hHSF2. High-throughput sequencing analysis of the selected pools from multiple rounds showed an enrichment of specific aptamer sequences. For hHSF1 and hHSF2, the sequences from round 5 with the highest multiplicity values also had high interround enrichments. Those sequences were target-specific and could be easily identified as being preferentially enriched after just three selection rounds. We tested a single candidate each to hHSF1 and hHSF2 and found both to be high-affinity aptamers to full-length hHSFs.

Although HSF proteins have been extensively studied and characterized, we are still limited by the available methods and lack approaches to perturb the activity of specific factors to tease apart molecular interactions. Aptamers can act as inhibitors that bind to a protein surface and disrupt specific interactions or functions. When expressed in vivo in a temporally and spatially controlled manner, these aptamers provide a way to rapidly disrupt targeted domains of proteins and efficiently assess their primary functions and mechanisms of actions. We previously demonstrated the utility of inhibitory RNA aptamers to study macromolecular interactions in vivo.^{21,22,28} However, there were some limitations on the methodology used to select those aptamers, and we believe that our new method will significantly improve the SELEX efficiency by: (1) allowing the selection of aptamers for many targets, including different domains of a single protein, at the same time and (2) reducing the number of SELEX rounds to achieve the selection of high-affinity aptamers. With this multiplexed technology, we believe it will be possible to efficiently select aptamers that bind to the distinct domains of HSF and other

proteins, which will be extremely valuable to study the interactions of these proteins.

Extensions of the microcolumn devices and approach developed in this work could be used for selection strategies with various combinatorial libraries, including but not limited to genomic sequences,²⁹ mRNA display,³⁰ and peptide nucleic acids.³¹ Also, our multiplex approach would easily facilitate the discovery of multivalent aptamers to distinct target binding sites³² by performing a final selection by use of serially arranged microcolumns with different target subunits in each. Finally, our microcolumn devices could be used to discriminate for aptamers based on their on- or off-rate binding characteristics by enforcing certain restrictions on the flow rates used for the loading step and washing step, respectively.³³

■ ASSOCIATED CONTENT

Supporting Information

Additional text, two tables, and six figures as described in the text. This material is available free of charge via the Internet at <http://pubs.acs.org>.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail hgc1@cornell.edu; telephone 607-255-8707; fax 607-255-7658.

Present Addresses

^{||}D.R.L.: Department of Chemical Engineering, McMaster University.

[†]C.V.K.: Department of Physics and Astronomy, Wayne State University.

[∇]B.S.W.: Department of Internal Medicine, Division of Oncology and The Genome Institute, Washington University School of Medicine.

Author Contributions

[§]D.R.L. and K.S. contributed equally to this work.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported by the National Institutes of Health R01 GM090320. We thank Dr. Peter Schweitzer at the Cornell University Genomics Facility for his help with the high-throughput sequencing, Colin Waters for sequence clustering and analysis, and Dr. Seung-min Park for early discussions. Also, we thank Heather Deal and Harvey Tian for preparing the workflow diagram in Figure 1B.

■ REFERENCES

- (1) Ciesiolka, J. M.; et al. *RNA* **1995**, *1*, 538–550.
- (2) Shangquan, D.; et al. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103* (32), 11838–11843.
- (3) Tombelli, S.; Mascini, M. *Curr. Opin. Mol. Ther.* **2009**, *11* (2), 179–188.
- (4) Joyce, G. F. *Gene* **1989**, *82* (1), 83–87.
- (5) Ellington, A. D.; Szostak, J. W. *Nature* **1990**, *346*, 818–822.
- (6) Tuerk, C.; Gold, L. *Science* **1990**, *249* (4968), 505–510.
- (7) Gopinath, S. C. B. *Anal. Bioanal. Chem.* **2007**, *387* (1), 171–182.
- (8) Park, S.; et al. *Lab Chip* **2009**, *9* (9), 1206–1212.
- (9) Jolma, A.; J.; et al. *Genome Res.* **2010**, *20* (6), 861–873.
- (10) Nieuwlandt, D.; et al. *Biochemistry* **1995**, *34*, 5651–5659.
- (11) Liu, J.; Stormo, G. D. *Nucleic Acids Res.* **2005**, *33* (17), e141.
- (12) Cheng, C.; et al. *Biochem. Biophys. Res. Commun.* **2008**, *366* (3), 670–674.

- (13) Holeman, L. A.; et al. *Folding Des.* **1998**, *3* (6), 423–431.
- (14) Shui, B.; et al. *Nucleic Acids Res.* **2012**, *40* (5), e39.
- (15) Cox, J. C.; et al. *Biotechnol. Prog.* **1998**, *14* (6), 845–850.
- (16) Pagano, J. M.; et al. *J. Biol. Chem.* **2007**, *282* (12), 8883–8894.
- (17) Eulberg, D.; et al. *Nucleic Acids Res.* **2005**, *33* (4), e45.
- (18) Nitsche, A.; et al. *BMC Biotechnol.* **2007**, *7*, 48.
- (19) Leventis, R.; Silviu, J. R. *Biophys. J.* **2010**, *99* (7), 2125–2133.
- (20) Dai, C.; et al. *Cell* **2007**, *130* (6), 1005–1018.
- (21) Zhao, X.; et al. *Nucleic Acids Res.* **2006**, *34* (13), 3755–3761.
- (22) Salamanca, H. H.; et al. *Nucleic Acids Res.* **2011**, *39* (15), 6729–6740.
- (23) Unpublished work, Cornell University, 2009.
- (24) Zimmermann, B.; et al. *PLoS One* **2010**, *5* (2), e9169.
- (25) Cho, M.; et al. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107* (35), 15373–15378.
- (26) Ankar, J.; Sistonen, L. *Adv. Exp. Med. Biol.* **2007**, *594*, 78–88.
- (27) Zuker, M. *Nucleic Acids Res.* **2003**, *31* (13), 3406–3415.
- (28) Shi, H.; et al. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96* (18), 10033–10038.
- (29) Singer, B. S.; et al. *Nucleic Acids Res.* **1997**, *25* (4), 781–786.
- (30) Wilson, D. S.; et al. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98* (7), 3750–3755.
- (31) Brudno, Y.; et al. *Nat. Chem. Biol.* **2010**, *6*, 148–155.
- (32) Gong, Q.; et al. *Anal. Chem.* **2012**, *84* (12), 5365–5371.
- (33) Gold, L.; et al. *PLoS One* **2010**, *5* (12), e15004.