

Efficient Identification of Murine M2 Macrophage Peptide Targeting Ligands by Phage Display and Next-Generation Sequencing

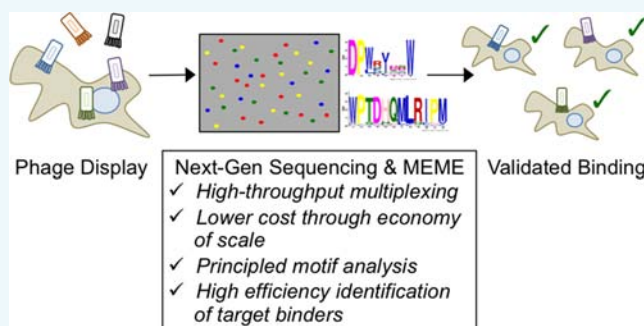
Gary W. Liu,^{†,‡} Brynn R. Livesay,^{†,‡} Nataly A. Kacherovsky,[†] Maryelise Cieslewicz,[†] Emi Lutz,[†] Adam Waalkes,[‡] Michael C. Jensen,^{†,§} Stephen J. Salipante,^{*,‡} and Suzie H. Pun^{*,†}

[†]Departments of Bioengineering and Molecular Engineering and Sciences Institute and [‡]Laboratory Medicine, University of Washington, Seattle, Washington 98195, United States

[§]Ben Towne Center for Childhood Cancer Research, Seattle Children's Research Institute, Seattle, Washington 98145, United States

Supporting Information

ABSTRACT: Peptide ligands are used to increase the specificity of drug carriers to their target cells and to facilitate intracellular delivery. One method to identify such peptide ligands, phage display, enables high-throughput screening of peptide libraries for ligands binding to therapeutic targets of interest. However, conventional methods for identifying target binders in a library by Sanger sequencing are low-throughput, labor-intensive, and provide a limited perspective (<0.01%) of the complete sequence space. Moreover, the small sample space can be dominated by nonspecific, preferentially amplifying “parasitic sequences” and plastic-binding sequences, which may lead to the identification of false positives or exclude the identification of target-binding sequences. To overcome these challenges, we employed next-generation Illumina sequencing to couple high-throughput screening and high-throughput sequencing, enabling more comprehensive access to the phage display library sequence space. In this work, we define the hallmarks of binding sequences in next-generation sequencing data, and develop a method that identifies several target-binding phage clones for murine, alternatively activated M2 macrophages with a high (100%) success rate: sequences and binding motifs were reproducibly present across biological replicates; binding motifs were identified across multiple unique sequences; and an unselected, amplified library accurately filtered out parasitic sequences. In addition, we validate the Multiple Em for Motif Elicitation tool as an efficient and principled means of discovering binding sequences.



■ INTRODUCTION

Targeting ligands facilitate cell recognition and are key technologies in targeted drug delivery and molecular imaging. In particular, peptides are an attractive class of targeting ligands: compared to proteins and antibodies, peptides are smaller and can therefore better penetrate tissues to reach target cells; peptides are less prone to clearance by the reticuloendothelial system; and finally, peptides can be easily functionalized for well-defined conjugation to cargo. Indeed, several peptides have been used clinically for various targets, such as cancer to thrombi.^{1–3}

A core methodology for ligand identification, phage display, has led to the discovery of new functional ligands through high-throughput screening of peptide and antibody fragment libraries for binding interactions with targets of interest.^{4–6} In peptide phage display, bacteriophage are engineered to display random peptide sequences off of coat proteins to create a peptide library. One popular commercial platform, the Phage Display (Ph.D.) system, uses M13 bacteriophage modified for pentavalent peptide display, where the randomized peptide is fused to each of the five pIII bacteriophage coat proteins.⁴ The input phage library is enriched for target binders through

iterative rounds of negative (removal of nontarget binders) and positive (enrichment of target binders) selection and amplification. Since displayed peptides are genetically encoded by the phage genome, the identity of target-binding peptides expressed by adherent phage may be retrieved by DNA sequencing. Using standard Sanger sequencing workflows, 20–100 randomly selected phage clones are typically isolated for sequencing. The strategy used to select potential target-binding peptides relies on identifying consensus sequences or motifs, which appear multiple times within a given subset of the phage population, that are presumptive of a target-binding sequence enriched during selection.

Given the vast diversity encoded by peptide phage display libraries (~10⁹ unique sequences), the current sequence analysis scheme is low-throughput (~8 h to isolate 20–40 clones) and provides a limited perspective (<0.01%) of the complete sequence space. This small sample size is particularly sensitive to bias from nonspecific, preferentially amplifying

Received: June 17, 2015

Revised: July 9, 2015

Published: July 10, 2015

“parasitic sequences”⁷ and plastic-binding sequences,⁸ leading to identification of false positives and exclusion of target-binding sequences (Figure 1). These limitations impose a major bottleneck in ligand identification, as functionally evaluating potential binding sequences can be costly in both time and resources.

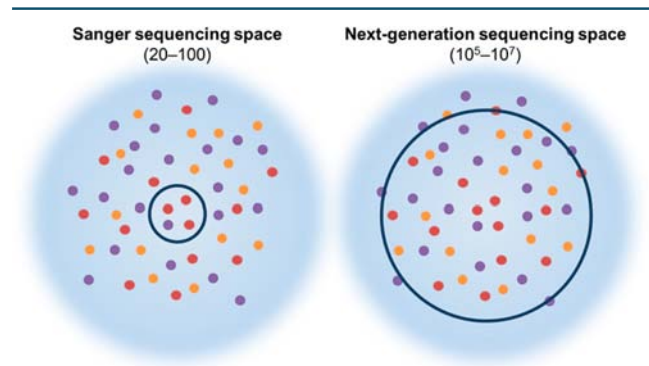


Figure 1. Relative perspectives offered by Sanger (left) and next-generation (right) sequencing of the sequence space. Represented in purple and yellow are targeting-binding sequences; represented in red are nonspecific sequences. Sanger sequencing may be biased by nonspecific, highly proliferating sequences (red) that result in false positives; next-generation sequencing provides a much broader perspective.

To address these limitations, several groups have recently employed next-generation sequencing technologies to analyze multiple phage libraries on a high-throughput scale ($\sim 10^5$ – 10^7 reads/sample), using a variety of strategies to define and identify target-binding peptides. 't Hoen and colleagues used Illumina sequencing to analyze libraries selected against KS483 osteoblasts.⁸ To eliminate potential parasitic sequences, they discarded sequences that appeared at least twice in an unselected, amplified library, and subsequently identified and

confirmed KS483-binding peptides by confocal microscopy. Ngubane and colleagues similarly used Illumina sequencing, and compared sequence enrichment of the selected libraries to the input library with a Pearson χ -squared test to identify an *M. tuberculosis*-binding peptide.⁹ Rebollo and colleagues utilized Ion Torrent sequencing and developed MATLAB scripts to identify potential binding motifs; however, these binding motifs were not experimentally validated.¹⁰

These reports and others have mostly been prospective, in which the identity of target-binding sequences was unknown during data processing and analysis. However, applying data filters in prospective analyses may potentially eliminate legitimate target-binding sequences; for example, discarding sequences that appear more than twice in an unselected, amplified library⁸ may inadvertently remove target-binding sequences displayed on phage that also amplify well in bacteria. As such, development of an empiric, experimentally validated method for identifying target-binding peptide sequences by next-generation sequencing could substantially improve the yield and diversity of the biologically relevant sequences identified.

In this study, we define the hallmarks of target-binding sequences analyzed by next-generation sequencing data, and utilize these hallmarks to develop a method that identifies target-binding sequences with a high success rate. We have previously used peptide phage display coupled with Sanger sequencing to identify a murine M2 “anti-inflammatory” macrophage-binding peptide that mediates M2 macrophage-specific delivery of pro-apoptotic peptides, and have employed this peptide to selectively eliminate M2-like tumor-associated macrophages and improve survival in murine tumor models.¹¹ As we have established a successful phage display strategy that has yielded an experimentally validated peptide, here we characterize how target-binding sequences are represented in high-throughput sequencing readouts of phage display libraries performed in biological duplicate. Through this work, a recommended method for efficient identification of peptide

Rank	Illumina sequencing							
	Replicate 1				Replicate 2			
			UA1	UA4			UA1	UA4
1	WPTDHQMLRIPM	40.15%	342	0	WPTDHQMLRIPM	32.14%	342	0
2	QSDPWLVSRLF	8.55%	66	0	TYPSTQWFFAKF	9.51%	60	0
3	DWSSWVYRDPQT	7.82%	143	3030	SFPDVQWWRNRY	5.16%	50	0
4	TYPSTQWFFAKF	1.90%	60	0	WPWDPLRISDWL	5.08%	36	0
5	ELPSVEQLWDF	1.87%	37	0	TLPPEAWWRTMY	4.42%	52	0
6	YPSSEQLLAWWG	1.74%	40	0	QSDPWLVSRLF	3.59%	66	0
7	WPSDNQLLIHV	1.64%	2	0	YPSSEQLLAWWG	3.52%	40	0
8	RLPTSMELLAAF	1.64%	24	0	RLPTSMELLAAF	3.21%	24	0
9	TWVSDLDMLGA	1.53%	28	0	GLPSSAELERLW	2.47%	14	0
10	WDPWQIDRWLL	1.47%	8	0	ELPSVEQLWDF	2.44%	37	0
11	NAPSIYDWLTL	1.20%	5	0	TWVSDLDMLGA	1.89%	28	0
12	KLPSYDLYLFL	1.15%	9	0	QDPWKVKWNINQ	1.78%	15	0
13	TLPPEAWWRTMY	0.72%	52	0	FFPSEQLLIAAL	1.64%	12	0
14	WPWDPLRISDWL	0.62%	36	0	SDSWVRGLNYWY	1.27%	16	0
15	QDPWKVKWNINQ	0.56%	15	0	FDPLNIRHWIAG	1.20%	7	0
16	FFPSEQLLIAAL	0.43%	12	0	WDPWQIDRWLL	1.19%	8	0
17	SWDPLLVSAWIK	0.43%	1	0	LPWPSDQIILMW	1.17%	14	0
18	WPLWSFDWPQNA	0.38%	4	2	DPWFINEWIRIP	0.88%	4	0
19	WLTEDMSAFMSY	0.34%	4	0	LPSSAELLWALR	0.81%	9	0
20	DPWFINEWIRIP	0.32%	4	0	SYWVPDVIWAGL	0.74%	15	0

Figure 2. Illumina sequencing of the two biological duplicates. Common sequences from the duplicates are highlighted. Percentages indicate absolute abundance within the library; numbers indicate abundance in the unselected, 1 \times amplified (UA1) and unselected, 4 \times amplified (UA4) libraries.

targeting sequences is presented, as well as the discovery of six new peptide sequences that exhibit preferential binding to murine M2-like macrophages versus M1-like macrophages.

RESULTS AND DISCUSSION

Next-Generation Sequencing Provides a Broader Perspective. Subtractive phage display using murine IL-4-polarized M2 (“anti-inflammatory”) macrophage for positive selection and IFN- γ - and LPS-polarized M1 (“pro-inflammatory”) macrophage for negative selection was completed in biological duplicates with one initial enrichment selection followed by three rounds of paired negative/positive selection. From the last round of each duplicate phage display experiment, 20 phage plaques were randomly selected for Sanger sequencing (Figure S1). While the exact M2pep sequence (YEQDPWGVKWWY) was not recovered from either replicate, sequences containing the M2pep motif DPWXXXW (highlighted in orange), where X indicates any amino acid, were identified in two independent experiments. The absence of M2pep is likely due to batch variations between the phage libraries used: as New England BioLabs has produced and sold at least 10 independent lots of Ph.D.-12 libraries,⁷ the original M2pep sequence may not have been present in the initial phage display peptide library used in this study. While no consensus sequence was observed in the first replicate, multiple consensus sequences were observed in the second replicate (colored). Barcoded phage library amplicons were sequenced by Illumina, and the top 20 most abundant sequences from the last rounds are shown (6.3×10^5 and 8.7×10^5 reads, respectively) in Figure 2. The two replicates contained the same 13 sequences (highlighted in blue) in the top 20 most abundant sequences, indicating a high degree of reproducibility (65% match) between the two biological replicates.

Sanger sequencing provided only a limited representation of the phage population as revealed by next-generation sequencing: the most abundant sequence identified from next-generation sequencing data, WPTDHQMLRIPM (“WPT”, >40% and >32% of the selected phage library in replicates 1 and 2, respectively), appeared only 1 and 3 times, respectively, out of the 20 Sanger-sequenced clones in the first and second replicates. Based on the Sanger sequencing results in the first replicate, we would have been unable to identify a valid consensus sequence. In contrast, the second replicate revealed 5 distinct consensus sequences that did not reflect the proportions reported by next-generation sequencing. These observations underscore the high degree of variability inherent in a Sanger sequencing approach owing to sample size constraints, which are overcome by high-throughput next-generation sequencing.

We employed Clustal Omega alignment analysis of the top 20 sequences identified by next-generation sequencing in each biological replicate to identify potential motifs (defined as a minimum of three consecutive amino acid residues) for binding studies.¹² The most abundant sequence, “WPT”, and motifs that were present in multiple sequences were considered strong candidates for binding studies. Multiple sequence alignment analysis identified four hypothesized motifs (reported along with their parent peptide sequence, in parentheses): “DPW” (QSDPWLVSRLF), “YPS” (TYPSTQWFFAKF), “SEQ” (FFPSEQVLIAAL), and “LPS” (GLPSSAELERLW) (colored, Figure S2). In addition, a fifth sequence YPSSEQLLAWWG containing both the YPS and SEQ motifs in tandem was also selected (“YPSSEQ”). To identify potential nonspecific,

preferentially amplifying “parasitic” sequences,⁸ the absolute abundance of the top 20 sequences in unselected, 1 \times amplified (UA1), and 4 \times amplified (UA4) libraries were compared. Notably, many of the putative binding sequences were well represented in UA1 (>20). The most abundant sequence, “WPT”, was present in UA1 342 times. In contrast, only two of the top 20 sequences appeared in UA4, and these two sequences were present in only one of the biological replicates. Moreover, the common sequences of the two replicates (highlighted in blue) did not appear in UA4.

In general, an increase in phage titer throughout display rounds suggests an enrichment and amplification of target-binding sequences and concurrent collapse in library diversity. Indeed, these trends were confirmed by next-generation sequencing data. We observed that increases in phage titer corresponded with decreases in the number of unique peptide sequences (Figure S3), and there was a 5- and 20-fold decrease in library diversity from rounds 1 to 4 in the two biological replicates, respectively. Collectively, these data suggest that the two phage display experiments enriched for M2 macrophage-binding phage with successive panning rounds.

Phage Clones Exhibit Selective M2 Macrophage Binding. All of the strong candidate binding sequences identified from next-generation sequencing data, as well as M2pep and scrambled M2pep (scM2pep), were cloned into phage (Table S2) and experimentally tested for binding to M1 and M2 macrophages by flow cytometry. All of the tested phage clones, except for the scrambled control, exhibited significantly greater binding to M2 compared to M1 macrophages (Figure 3A). These findings were somewhat unexpected, as the

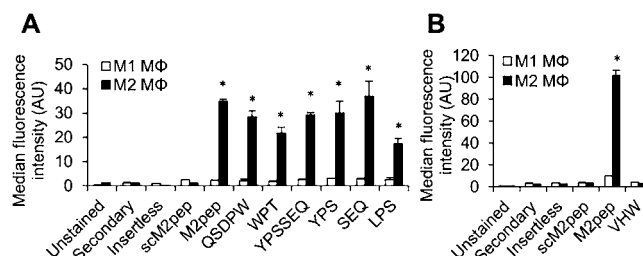


Figure 3. Phage clone binding of (A) selected phage clones and (B) a preferential amplifier to M1 and M2 macrophages by flow cytometry. Phage binding is expressed as median fluorescence intensity (MFI). Data reported are the mean MFI \pm standard deviation ($n = 3$). AU, arbitrary units. Statistical analysis performed with a two-tailed Student's t test compared to phage-treated M1 macrophage. * p -value <0.05.

sequence “WPT”, given its high abundance in UA1, was considered a putative parasitic sequence. ‘t Hoen et al. identified binding peptides by filtering out presumptive false positive sequences (defined as sequences that appeared more than once in an unselected, 1 \times amplified library or in online peptide databases).⁸ Here, all of the tested sequences exhibited abundance >2 in UA1. In a prospective approach following the strategy recommended by ‘t Hoen, these selective binders would have been discarded.

An improved strategy to identify selective binding sequences may be to assess sequence abundance in a UA4 library, which better represents the amplification cycles of the selected libraries. With this approach, only the two sequences DWSSWVYRDPQT (“DWSS”) and WPLWSFDWPQNA (“WPLW”) would be identified as false positive sequences.

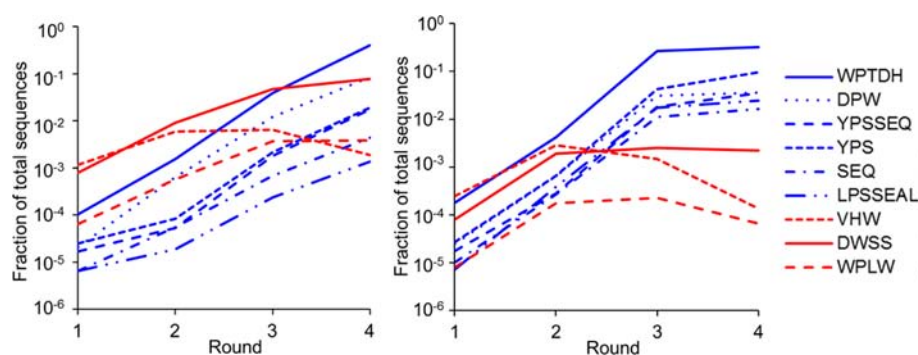


Figure 4. Fraction of total sequences in Illumina next-generation data for confirmed binding sequences (blue) and nonspecific preferential amplifying sequences (red) throughout four rounds of phage display panning in both the first (left) and second (right) biological replicates.

Both “DWSS” and “WPLW” contain the plastic binding motif WXXW.¹³ We have also observed “DWSS” in the top 20 of next-generation sequencing results from phage display experiments against other targets (data not shown). To confirm nonspecificity of parasitic sequences, we tested binding of VHWDFRQWWQPS (“VHW”), which was present in both the UA1 (abundance: 199, proportion of library: $\sim 0.0069\%$) and UA4 (abundance: 41, proportion of library: $\sim 0.0043\%$) libraries and has been identified as a consensus sequence in display strategies against other targets pursued by our group. As tested by flow cytometry, this sequence did not exhibit significant macrophage binding compared to controls (Figure 3B).

Binding Sequences Enrich Throughout Subtractive Phage Display Rounds. We next utilized multiple analytical tools to observe how binding sequences emerge in next-generation sequencing data and to find a method that would successfully identify binding sequences while eliminating false positives. In this context, the six validated binding sequences served as positive controls and the known parasitic sequences as negative controls.

To characterize enrichment dynamics, the fractional abundance of the six binding sequences and three parasitic sequences were tracked throughout the four rounds of phage display for each biological replicate (Figure 4). Binding sequences exhibited lower abundance in the first and second rounds of phage display, but consistently increased in proportion after each round. Parasitic sequences generally exhibited higher fractional abundance in early rounds of the panning experiments, but began to plateau or decrease in fractional abundance in the last two rounds as binding sequences began to dominate the sequence space. Similar trends were observed when comparing the rank in abundance of binding and parasitic sequences throughout each round of phage display (Figure S4). Therefore, performing multiple rounds of phage display enhances discrimination between binding and parasitic sequences through changes in rank and proportion throughout the display rounds. While binding peptides may theoretically be identified without multiple subtractive rounds,⁸ in this report we find that additional rounds of selection enhanced the representation of desired target-binding phage versus undesired nonspecific phage.

These observations emphasize the importance of experimental design for successful enrichment of targeting-binding sequences. We have observed a much higher prevalence of parasitic sequences in phage display experiments following less stringent selection protocols, such as using immobilized

proteins for positive selection and the immobilizing substrate for negative selection (data not shown). The selection strategy and amplification in bacteria between rounds both contribute to the “evolutionary” pressure placed on bacteriophage, selecting for phage that both amplify (leading to parasitic sequences) and bind to nontarget materials (plastic binders, nonspecific sequences) used in the display throughout an experiment. Here, the UA4 library enabled discrimination between target-binding phage and parasitic phage.

Algorithmic Identification of Binding Motifs. As we observed the greatest phage titers and the least library diversity in the fourth round of each replicate (Figure S3), these libraries provided data to explore the use of automated motif analysis algorithms to provide a clear and principled representation of target binding peptide motifs with minimal representation of parasitic peptide motifs.

The MEME analysis (Multiple Expectation Maximization for Motif Elicitation) algorithm developed by Bailey utilizes an expectation maximization to fit a two-component, finite mixture model to identify multiple motifs of varying lengths within a DNA or protein sequence set.^{14,15} By using a “greedy” heuristic computation, MEME is able to sample a large number of sequences (up to 1000) in reasonable computing time while still identifying statistically significant motifs. Peptide sequences that appeared at least 10 times in the fourth round of both biological replicates were used to optimize analysis settings; we omitted sequences with abundance less than 10, as this method equally weights all sequences and low abundant sequences could introduce noise. Scanning for motifs between 6 and 12 amino acids in length resulted in the most robust representation of confirmed binding sequences and hypothesized motifs. MEME analysis without any comparison of the phage display libraries with the UA4 library identified binding sequence motifs as well as a DWSS parasitic sequence motif in round 1.4 (Figure S5). Removal of sequences that were present in the UA4 library before MEME analysis resulted in a robust representation of motifs and their binding sequences and good agreement between the two biological replicates (Figure 5). This approach effectively captured all of the hypothesized motifs present in the validated M2-binding sequences “DPW”, “YPSSEQ”, “YPS”, “SEQ”, and “LPS”.

In addition, we observed that MEME was most effective after a substantial collapse of the library to diversity below 2×10^5 unique sequences. MEME analysis was able to identify many of the hypothesized binding motifs using sequences from the third round of biological replicate 2 and the fourth round of both replicates, but was only able to identify one significant motif

Rank	Round 1.4 motifs	Number of sequences	Round 2.4 motifs	Number of sequences
1		60		59
2		19		105
3		34		57
4		15		82
5		10		294

Figure 5. Logos and number of sequences in motifs identified by multiple em for motif elicitation (MEME). Analysis includes sequences that were present 10 or more times in the fourth round of each biological replicate after removal of preferentially amplified sequences present in the UA4 library. All experimentally confirmed binding sequences are represented in these motifs. Only statistically significant motifs (E -value <0.05) are shown.

from the third round of biological replicate 1 (Figures S5 and S6). This further motivates the use of multiple rounds of subtractive phage display, which enhance the effectiveness of motif-finding algorithms.

Highly Efficient Identification of Binding Sequences by Next-Generation Sequencing. Prior to this study, our group utilized Sanger sequencing to identify prospective target-binding peptide sequences. This conventional approach was low-throughput (20–40 sequences) and time-intensive (~ 8 h), often generating artifacts including sequencing errors due to poor DNA quality or sequencing of insertless phage that further diminished the power of an already small sample size. Moreover, sequencing costs accumulated as displays were repeated, costing \$120–240 per experiment. In contrast, next-generation sequencing technology has overcome these limitations due to its high throughput and economy of scale.¹⁶ The use of barcoded primers enables multiple library samples to be sequenced on the same sequencing run, bringing the absolute per-sample cost of next-generation sequencing to a level comparable with Sanger sequencing and increasing the sequencing capacity by 4 orders of magnitude. Indeed, in this study, we were able to identify and validate 6/6 (100%) of the candidate sequences in ~ 1 month, using data generated from a single next-generation sequencing run.

CONCLUSIONS

Performing phage display in biological replicates and validating binding of selected phage clones to their target enabled us to define characteristics of target-binding sequences in the context of next-generation sequencing data. Specifically, we noted that (i) sequences and binding motifs (3–7 amino acid residues) were reproducibly present across biological replicates; (ii) an unselected library amplified the same number of times as selected libraries enabled more accurate discrimination of parasitic sequences; and (iii) binding motifs could be identified by their presence in multiple unique sequences. Moreover, we validated MEME as a powerful tool for identifying targeting-binding sequences. In this particular study, six new sequences

for targeting murine M2 macrophage were discovered by the combination of phage display selection, next generation sequencing, and data analysis.

Based on our experience, we suggest the following workflow and methodological considerations to increase the productivity of phage display experiments and to improve the success rate of identifying target-binding sequences:

1. Experimental Design. A 100-fold representation of the naïve library should be amplified the same number of rounds as the experimental library (UAX, where X denotes the number of amplifications), with no selection step between amplifications to aid in identifying parasitic sequences. Performing experiments in biological replicates enables comparison analyses that can be helpful in confirming conserved binding motifs.

2. Next-Generation Sequencing. Sample-specific barcodes can be used to multiplex multiple experimental samples onto a single sequencing run in order to minimize cost and time. In these studies, allocating 1×10^6 sequencing reads per sample gave an adequate representation of the sequence space. Sequencing the last round of phage display performed and the UAX library is sufficient for identification of false positives and can be used to identify conserved motifs. Fidelity of sequencing data is also a consideration, and efforts should be taken to utilize high-fidelity next-generation sequencing platforms^{17–19} to limit the number of artifactual sequences recovered.

3. Data Formatting and Analysis. Translate quality-filtered sequencing data into counts of unique peptide sequences. Eliminate sequences from the selected library that are present in the UAX library and retain filtered sequences with abundance >10 sequence reads. We have developed an analysis pipeline, “PepRS”, to facilitate these steps.

4. MEME Analysis. Perform MEME motif analysis over a range of 6–12 amino acids in motif length (this may have to be adjusted depending on the size of the displayed peptide) and search for up to 10 motifs. Only consider motifs with E -values <0.05 .

5. Candidate Selection for Binding Validation. In general, strong candidates for binding exhibit consensus across the biological replicates and a high number of sequences within the displayed motif in MEME analysis. Select for the most abundant sequence within a strong candidate motif. Discard sequences that are present in the online databases SAROTUP (immunet.cn/sarotup)^{20,21} and PepBank (pepbank.mgh.harvard.edu).²²

We have developed a method that allows for highly efficient identification of target-binding peptide sequences within a phage display library, as well as providing an automated software package that generates and pools peptide sequences from next-generation sequencing reads. This method offers improvements in laboratory workflow efficiency, increases data quality and volume, and expedites target-binding sequence identification.

MATERIALS AND METHODS

Peptide Phage Display. Peptide phage display was conducted against murine M2 macrophage as previously reported except using lot number 0131402 of the Ph.D.-12 linear dodecapeptide library (New England BioLabs).¹¹ Briefly, negative and positive selection was performed using bone marrow-derived murine IFN- γ - and LPS-polarized M1 and IL-4-polarized M2 macrophage, respectively. An initial enrichment round (positive selection only) was performed followed by three subtractive rounds (negative and positive selection). For

the first round, 2×10^{11} PFU of the Ph.D. library was used as input; for each subsequent round, 2×10^{11} PFU of the amplified eluate from the previous round was used as the input. Stringency was applied throughout the rounds by increasing the Tween-20 content (0.1%, 0.3%, 0.5%) in wash steps. Phage titering, amplification, and DNA sequencing was performed according to the New England BioLabs manual. The experiment was performed in biological duplicate. From each experiment, 20 plaques from the third subtractive round were randomly selected for DNA Sanger sequencing (GENEWIZ). In addition, 2×10^{11} PFU of the Ph.D. library was amplified once or four times with no selection and used as the unselected library controls, UA1 and UA4, respectively.

Phage DNA Isolation and Purification for Next-Generation Sequencing. DNA amplicons were prepared as previously described with some modifications.¹⁰ Single-stranded DNA (ssDNA) was isolated from 1×10^{11} PFU of the amplified phage eluate from each round of peptide phage display using QIAprep Spin M13 Kit (QIAGEN), and 50 ng of ssDNA was amplified by PCR using primers containing Illumina-compatible sequencing adaptors and sample-specific barcodes (Table S1). For each library sample, the PCR reaction (50 μ L) contained 500 nM of both the forward primer and reverse barcode primer unique to that sample, 250 μ M dNTP, 1 U Phusion High Fidelity DNA Polymerase (New England BioLabs), and 1 \times Phusion High Fidelity buffer. Thirteen PCR cycles (10 s at 98 $^{\circ}$ C, 30 s at 59 $^{\circ}$ C, 30 s at 72 $^{\circ}$ C) were performed and PCR products were purified from a 2% agarose gel (AquaPor LE GTAC Agarose, National Diagnostics) using QIAquick Gel Extraction Kit (QIAGEN). The concentration of DNA was determined by a quantitation assay using a Qubit 3.0 Fluorometer (Life Technologies). Purified PCR products prepared from the unselected and selected phage libraries were combined in a 5:1 molar ratio, with respect to each individual selected sample, prior to sequencing.

Illumina Next-Generation Sequencing. Sequencing and data processing was performed as previously described with some modifications.⁸ Sequencing was performed on an Illumina MiSeq system in the presence of 7% PhiX control, using 150 bp single-ended sequencing chemistries and custom sequencing primers (Read_1_primer, 5'-GCTCGACCTGTTCTTT-AGTGGTACCTTTCTATTCTCACTCT-3'; Index_read_primer, 5'-AGCAAAATCCCATACAGAAATTCATTAC-CGCAGGTCGCTCC-3') according to manufacturer instructions. Raw sequence reads for all experiments are available from the Sequence Read Archive (SRA; <http://www.ncbi.nlm.nih.gov/sra>) under Study Accession Number SRP059048.

Illumina Next-Generation Data Processing. Sequence reads are an exact match of the first 10 nucleotides of the flanking sequence immediately downstream of the library insert variable region (5'-GGTGGAGGTT-3') were retained for further analysis. Sub-sequences corresponding to the phage display variable region (excluding the terminal 3 glycines of the library insert) were extracted and translated into amino acids, substituting the amber stop codon (UAG) with glutamine to reflect the nonsense suppression activity of *E. coli* K12 ER2738. For each library, the count of sequence reads encoding each unique peptide sequence was tabulated. The relative abundance of specific peptides was calculated as the fraction of total reads in the library that encoded the peptide. Source code and executables for our analysis pipeline, "PepRS" (Peptide Retrieval Software), are freely available to academic users at <https://bitbucket.org/stevesal/peprs>.

Secondary Analysis. Clustal Omega Motif Analysis. Clustal Omega (<http://www.ebi.ac.uk/Tools/msa/clustalo>)¹² was used for initial motif identification using the 20 most abundant peptide sequences of the fourth round from each biological replicate. Afterward, aligned sequences were visually inspected for binding motifs (defined as a minimum of three consecutive amino acids) that appeared multiple times across different peptide sequences. The most abundant peptide sequence that contained each motif was selected for cloning and subsequent in vitro binding studies.

Multiple EM for Motif Elicitation (MEME) Analysis. Unique sequences that were present over 10 times in the third or fourth rounds of phage display for each biological replicate were analyzed for conserved motifs using Multiple EM for Motif Elicitation (MEME, <http://meme-suite.org/tools/meme>).^{14,15} The normal motif discovery mode, binary site distribution, and zero-order model of sequences were used to scan for up to 10 motifs with a width of 6–12 amino acids. Motifs with an *E*-value <0.05 were considered statistically significant. This analysis was repeated for the fourth round sequences after eliminating all sequences that also appeared in the UA4 library.

Construction of Phage Clones. Oligonucleotides encoding the peptides of interest were designed with 5' overhangs and an overlap region with $T_m \sim 55$ –60 $^{\circ}$ C and synthesized by IDT (Table S2). Oligonucleotide pairs were annealed, filled in with Klenow (New England BioLabs), digested with Acc65I/EagI (New England BioLabs), and cloned into M13KE dsDNA vector according to the New England BioLabs Ph.D. manual. The resulting plaques were used for amplification and production of phage. Peptide sequences were confirmed by DNA sequencing.

Phage Binding Study. The selected phage clones (Table S2) were tested for differential binding to murine M1 and M2 macrophages. Macrophages (1×10^5 cells/well) were incubated with 1×10^9 PFU of phage diluted in PBS+1% BSA (PBBSA)+Fc Block (BD Biosciences) for 30 min on ice. The cells were then washed twice with 1% PBBSA, fixed with 4% paraformaldehyde for 15 min on ice, washed, stained with 1:1000 rabbit anti-M13 antibody (Sigma) for 20 min, washed, and stained with 1:500 goat anti-rabbit antibody (Sigma) for 20 min. After washing, cells were analyzed by a MACSQuant flow cytometer (Miltenyi Biotec). Gating was set with unstained and positively stained cells.

■ ASSOCIATED CONTENT

● Supporting Information

Sequencing primers and peptide oligonucleotide sequences, Sanger sequencing results, Clustal Omega alignment, phage titer and rank in abundance of specific clones throughout rounds, and additional MEME analysis. The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.bioconjchem.5b00344.

■ AUTHOR INFORMATION

Corresponding Authors

*E-mail: stevesal@uw.edu. Tel: +1 206 616 1831. Fax: +1 206 598 6189.

*E-mail: spun@uw.edu. Tel: +1 206 685 3488. Fax: +1 206 543 6124.

Author Contributions

#Gary W. Liu and Brynn R. Livesay contributed equally to the work.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported by the National Institutes of Health [grant numbers 1R01CA177272 and R21NS086500]; and by National Science Foundation Graduate Research Fellowships [grant number DGE-1256082] to GWL and BRL. We would also like to thank Charles Grant for insightful discussion about MEME.

■ REFERENCES

- (1) Gaertner, F., Kessler, H., Wester, H.-J., Schwaiger, M., and Beer, A. (2012) Radiolabelled RGD peptides for imaging and therapy. *Eur. J. Nucl. Med. Mol. Imaging* 39, 126–138.
- (2) Putzer, D., Kroiss, A., Waitz, D., Gabriel, M., Traub-Weidinger, T., Uprimny, C., von Guggenberg, E., Decristoforo, C., Warwitz, B., and Widmann, G. (2013) Somatostatin receptor PET in neuroendocrine tumours: 68Ga-DOTA0, Tyr3-octreotide versus 68Ga-DOTA0-lanreotide. *Eur. J. Nucl. Med. Mol. Imaging* 40, 364–372.
- (3) Vymazal, J., Spuentrup, E., Cardenas-Molina, G., Wiethoff, A. J., Hartmann, M. G., Caravan, P., and Parsons, E. C., Jr (2009) Thrombus imaging with fibrin-specific gadolinium-based MR contrast agent EP-2104R: results of a phase II clinical study of feasibility. *Invest. Radiol.* 44, 697–704.
- (4) Smith, G. P. (1985) Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. *Science* 228, 1315–1317.
- (5) Hammers, C. M., and Stanley, J. R. (2014) Antibody phage display: technique and applications. *J. Invest. Dermatol.* 134, e17.
- (6) Dantas-Barbosa, C., de Macedo Brigido, M., and Maranhao, A. Q. (2012) Antibody phage display libraries: contributions to oncology. *Int. J. Mol. Sci.* 13, 5420–5440.
- (7) Matochko, W. L., Cory, L. S., Tang, S. K., and Derda, R. (2014) Prospective identification of parasitic sequences in phage display screens. *Nucleic Acids Res.* 42, 1784–1798.
- (8) 't Hoen, P. A., Jirka, S. M., Ten Broeke, B. R., Schultes, E. A., Aguilera, B., Pang, K. H., Heemskerk, H., Aartsma-Rus, A., van Ommen, G. J., and den Dunnen, J. T. (2012) Phage display screening without repetitious selection rounds. *Anal. Biochem.* 421, 622–631.
- (9) Ngubane, N. A., Gresh, L., Ioerger, T. R., Sacchettini, J. C., Zhang, Y. J., Rubin, E. J., Pym, A., and Khati, M. (2013) High-throughput sequencing enhanced phage display identifies peptides that bind mycobacteria. *PLoS One* 8, e77844.
- (10) Rentero Rebollo, I., Sabisz, M., Baeriswyl, V., and Heinis, C. (2014) Identification of target-binding peptide motifs by high-throughput sequencing of phage-selected peptides. *Nucleic Acids Res.* 42, e169.
- (11) Cieslewicz, M., Tang, J., Yu, J. L., Cao, H., Zavaljevski, M., Motoyama, K., Lieber, A., Raines, E. W., and Pun, S. H. (2013) Targeted delivery of proapoptotic peptides to tumor-associated macrophages improves survival. *Proc. Natl. Acad. Sci. U. S. A.* 110, 15919–15924.
- (12) Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Soding, J., et al. (2014) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* 7, 539.
- (13) Vodnik, M., Zager, U., Strukelj, B., and Lunder, M. (2011) Phage display: selecting straws instead of a needle from a haystack. *Molecules* 16, 790–817.
- (14) Bailey, T. L., and Elkan, C. (1994) Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc. Int. Conf. Intell. Syst. Mol. Biol.* 2, 28–36.
- (15) Bailey, T. L., Williams, N., Misleh, C., and Li, W. W. (2006) MEME: discovering and analyzing DNA and protein sequence motifs. *Nucleic Acids Res.* 34, W369–373.
- (16) Dias-Neto, E., Nunes, D. N., Giordano, R. J., Sun, J., Botz, G. H., Yang, K., Setubal, J. C., Pasqualini, R., and Arap, W. (2009) Next-generation phage display: integrating and comparing available molecular tools to enable cost-effective high-throughput analysis. *PLoS One* 4, e8338.
- (17) Lam, H. Y., Clark, M. J., Chen, R., Natsoulis, G., O'Huallachain, M., Dewey, F. E., Habegger, L., Ashley, E. A., Gerstein, M. B., Butte, A. J., et al. (2012) Performance comparison of whole-genome sequencing platforms. *Nat. Biotechnol.* 30, 78–82.
- (18) Loman, N. J., Misra, R. V., Dallman, T. J., Constantinidou, C., Gharbia, S. E., Wain, J., and Pallen, M. J. (2012) Performance comparison of benchtop high-throughput sequencing platforms. *Nat. Biotechnol.* 30, 434–439.
- (19) Salipante, S. J., Kawashima, T., Rosenthal, C., Hoogstraat, D. R., Cummings, L. A., Sengupta, D. J., Harkins, T. T., Cookson, B. T., and Hoffman, N. G. (2014) Performance comparison of Illumina and ion torrent next-generation sequencing platforms for 16S rRNA-based bacterial community profiling. *Appl. Environ. Microbiol.* 80, 7583–7591.
- (20) Huang, J., Ru, B., Zhu, P., Nie, F., Yang, J., Wang, X., Dai, P., Lin, H., Guo, F. B., and Rao, N. (2012) MimoDB 2.0: a mimotope database and beyond. *Nucleic Acids Res.* 40, D271–277.
- (21) Huang, J., Ru, B., Li, S., Lin, H., and Guo, F. B. (2010) SAROTUP: scanner and reporter of target-unrelated peptides. *J. Biomed. Biotechnol.* 2010, 1–7.
- (22) Duchrow, T., Shtatland, T., Guettler, D., Pivovarov, M., Kramer, S., and Weissleder, R. (2009) Enhancing navigation in biomedical databases by community voting and database-driven text classification. *BMC Bioinf.* 10, 317.