The use of mRNA display to select high-affinity protein-binding peptides

David S. Wilson*, Anthony D. Keefe, and Jack W. Szostak†

Howard Hughes Medical Institute and Department of Molecular Biology, Massachusetts General Hospital, Boston, MA 02114

We report the use of "mRNA display," an *in vitro* selection technique, to identify peptide aptamers to a protein target. mRNA display allows for the preparation of polypeptide libraries with far greater complexity than is possible with phage display. Starting with a library of $\approx 10^{13}$ random peptides, 20 different aptamers to streptavidin were obtained, with dissociation constants as low as 5 nM. These aptamers function without the aid of disulfide bridges or engineered scaffolds, yet possess affinities comparable to those for monoclonal antibody–antigen complexes. The aptamers bind streptavidin with three to four orders of magnitude higher affinity than those isolated previously by phage display from lower complexity libraries of shorter random peptides. Like previously isolated peptides, they contain an HPQ consensus motif. This study shows that, given sufficient length and diversity, high-affinity aptamers can be obtained even from random nonconstrained peptide libraries. By engineering structural constraints into these ultrahigh complexity peptide libraries, it may be possible to produce binding agents with subnanomolar binding constants.

affinity tag | peptide library | peptide aptamer | Strep-tag | streptavidin

Antibodies and other specific protein-binding reagents are indispensable tools for monitoring protein abundance and activity. With the number of known proteins increasing dramatically as a result of whole-genome-sequencing projects, it has become crucial to find alternatives to traditional time-consuming monoclonal antibody production for generating high-affinity binding reagents. Numerous techniques have been developed to select polypeptides that bind with high affinity and specificity to target proteins (reviewed in ref. 1). The most commonly used such technique is phage display, in which each bacteriophage "displays" a unique peptide or protein on its surface, allowing for the *in vitro* selection of ligands with high affinities (2). As would be expected, the affinity of ligands derived from this process increases with increasing complexity (i.e., diversity) of the starting library (3, 4). The starting complexity of phage display libraries is generally limited to $<10^9$ because of the bacterial transformation requirement. To overcome this limitation, two completely *in vitro* procedures for linking polypeptides to their encoding mRNAs have been developed. Ribosome display achieves this link by stalling the translating ribosome in an *in vitro* translation reaction, thus maintaining the mRNA-ribosome-nascent peptide complexes (5, 6). mRNA display takes advantage of the translation-terminating antibiotic puromycin, which functions by entering the A site of ribosomes and forming a covalent bond with the nascent peptide. By covalently attaching puromycin to the 3′ end of an mRNA, a covalent link between a polypeptide and its encoding message can be achieved *in situ* during *in vitro* translation (7–9). These mRNA-peptide fusions can then by purified and subjected to *in vitro* selection. Here we report the use of this procedure to isolate peptide ligands.

Several different types of library have been used for *in vitro* selections, including random linear or disulfide-constrained peptide libraries and scaffold-based libraries. Nonconstrained linear peptide libraries generally do not yield high-affinity ligands to proteins, except in cases where the protein normally functions in peptide recognition (10, 11). By constraining the peptides, either through disulfide formation or by presenting them on the surface of a known protein scaffold (such as that of the antibody), higher affinities have been achieved (12–21). We show here, however, that even random nonconstrained peptides with high affinity (5–10 nM) to protein targets can be identified if the starting library is of sufficiently high complexity and length, in this case almost 10^{13} peptides, each 88 amino acids long.

Materials and Methods

For a detailed experimental protocol, please refer to the supplemental data (which is published on the PNAS web site, www.pnas.org).

Library Synthesis. The library was constructed, transcribed, ligated to the puromycin-containing linker, translated, fused to the mRNA, purified, and reverse transcribed as described (22). By comparing the ^{35}S counts of the purified, reverse transcribed mRNA-peptide fusions to the [^{35}S]methionine stock, and taking into consideration the total methionine concentration in the translation reaction (10 μM), we estimated the number of displayed peptides in the sample to be 6.7×10^{12}. This number also represents the complexity of the library, because it contains virtually no redundancy (22).

Selection for Streptavidin (SA)-Binding Peptides. The displayed peptides were incubated with immobilized SA (Ultralink Immobilized Streptavidin Plus, about 4 mg/ml; Pierce) in SA-binding buffer [40 mM Tris(hydroxymethyl)aminomethane/300 mM KCl/2 mM EDTA/0.1% Triton X-100/5 mM 2-mercaptoethanol/100 μg/ml BSA/1 μg/ml tRNA, pH 7.4]. The amount of gel used was 0.5 ml in a total volume of 5.5 ml. After incubating for 20 min at room temperature, the contents were loaded onto a chromatography column, drained, washed with 14 column volumes (CV) of SA-binding buffer, and then eluted with 5 successive aliquots (at 10-min intervals) of SA-binding buffer plus 2 mM D-biotin (Sigma). Elution fractions were combined and then PCR-amplified as described (22). This concluded the first round of selection, and the remaining rounds followed the same protocol except that the translation was scaled down 10-fold, and the washing of the SA column was increased (32 CV for round 2; 40 for rounds 3, 4, and 6; 25 for rounds 5 and 7). The SA-binding selection for rounds 5 and 7 were performed directly on the SA-column eluate from the preceding selection rounds, without intervening amplification (the biotin was removed by gel filtration). PCR products amplified after the seventh selection round were cloned and sequenced.
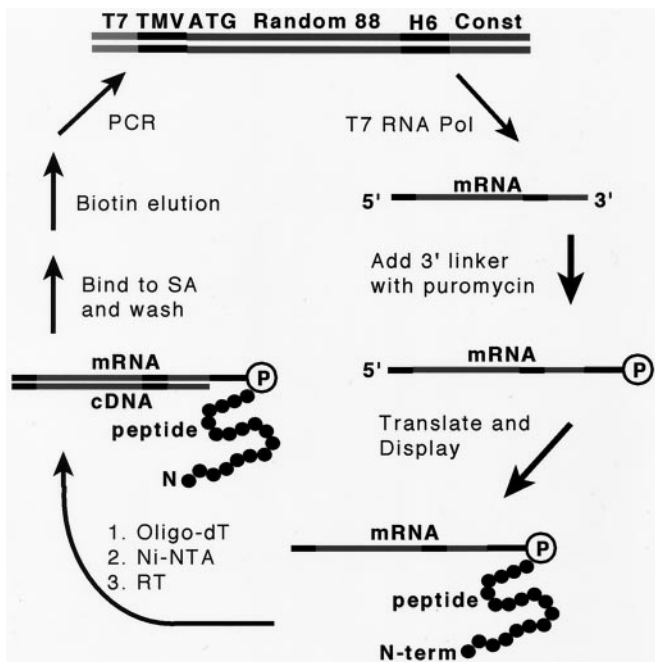
Analysis of Individual Peptide Sequences. To rapidly characterize selected peptides, a method for generating mRNA-peptide

Fig. 1. *In vitro* selection process. Schematic showing the structure of the library and the selection scheme. The DNA library has, from 5′ to 3′, a T7 RNA polymerase promoter (T7), a tobacco mosaic virus translation enhancer (TMV; ref. 36), a start codon (ATG), 88 random amino acids, a hexahistidine tag (H6), and a 3′ constant region (Const). This library is transcribed by using T7 RNA polymerase, after which the puromycin-containing linker is ligated onto the 3′ end of the mRNA. When this template is translated *in vitro*, the nascent peptide forms a covalent bond to the puromycin moiety. The resulting mRNA-peptide covalent fusion molecules are then purified on oligo-dT-cellulose (which anneals to the oligo-dA sequence in the puromycin-containing linker) and Ni-NTA agarose. The mRNA portion of this display construct is then reverse transcribed. The double-stranded DNA/RNA-peptide species are then incubated with the immobilized target protein (SA) and unbound library members are washed off. SA-bound peptides are then displaced with biotin. The eluted molecules are then amplified by PCR, thus completing the first round of selection and amplification.

covalent fusions was used (23). Plasmids containing single inserts were PCR-amplified with the same 5′ PCR primer as described for the library construction (22), and a new 3′ primer that altered the 3′ RNA sequence to ACUGGUCUGGCGGCUG-CAAGCUUGGCACCGGCUAU. This sequence was designed to anneal to the photo-crosslinking linker, which has the sequence 5′-psoralen-<u>TAGCCGGTG</u>-A17-CC-puromycin-3′, in which the underlined positions are 3′-methoxy nucleotides and the remaining ones are deoxy. This new primer changed the encoded constant C-terminal peptide sequence from WSG-GCHHHHHHSSA to WSGGCKLGTGY, of which the last three amino acids may not be translated because the RNA is annealed to the linker. Each DNA template was transcribed and gel purified as described (22), incubated with the psoralen linker, and irradiated at 366 nm. The translation/display reactions and oligo-dT-purification were carried out as above. Finally, RNase A was added to degrade the mRNA (complete degradation was confirmed by SDS/PAGE analysis).

The resulting purified DNA-tagged proteins (DTPs) were analyzed in the SA column-binding assay, in which ≈500 pM ^{35}S-labeled DTPs were mixed with 50 μl of the SA matrix in SA-binding buffer, in a total volume of 300 μl, for 10 min at room temperature (with agitation), after which the contents were loaded onto a chromatography column. The column was drained and then washed with 80 column volumes (CV) of SA-binding buffer, and then eluted with 3 consecutive aliquots (3 CV each)
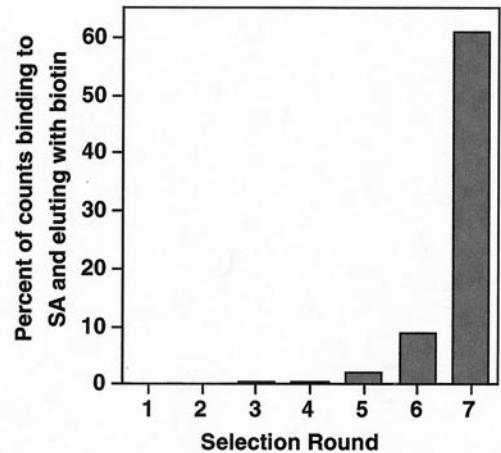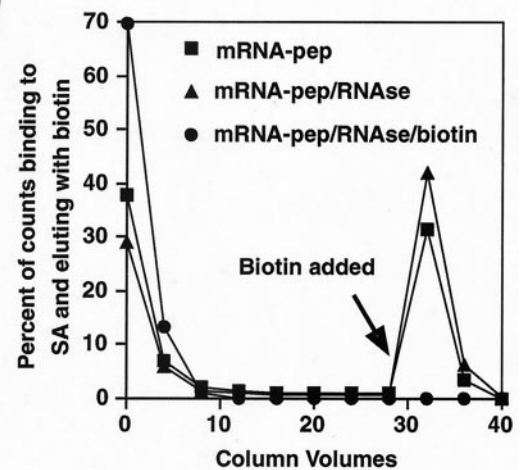
Fig. 2. Progress and result of the selection. (*A*) Fraction of ^{35}S counts from the displayed peptides that bound to SA and eluted with biotin, at each round of selection. (*B*) Elution profile for the peptide library generated from the output of the seventh round of selection. The first fraction represents the flow-through. Biotin was added at the point indicated. The plot compares the binding of the intact reverse-transcribed displayed peptides (mRNA-pep), the same sample treated with RNase A, and the RNase-treated sample applied to an SA column presaturated with biotin (excess biotin was washed away before exposing the library to the matrix).

of SA-binding buffer plus 2 mM biotin (over a 15-min period). All fractions were analyzed by scintillation counting to determine the fraction of DTPs that bound SA and eluted with biotin. The *Strep*-tag II (24) template encoded the peptide sequence MSNWSHPQFEKNWSGGCGTGY. The nonselected clone in which two HPQ motifs (separated by 19 amino acids) were introduced encoded the sequence M<u>DEA</u>**HPQ**<u>AGP</u>VDQA-DARLVQQGA<u>L</u>**QHHPQ**<u>G</u>DRMMSGGCKLGTGY (the underlined portions are identical to the HPQ regions of clone SB2).

Quantitative analysis of binding was performed by an electrophoretic mobility shift assay (EMSA). DTPs were incubated with varying amounts of SA (Pierce Immunopure Streptavidin) in SA-binding buffer plus 5% glycerol (to increase the density of the solution). After incubating at room temperature for 20 min, the reactions were incubated at 4°C before being loaded onto a 10% polyacrylamide gel containing 2× 90 mM Tris/64.6 mM boric

Table 1. Sequences of the 20 clones from the seventh round of selection

Name	No.	Sequence
SB1	3	MDEKTHCTISMNGAVPLVPHH**HPQ**GDPLRLLHRPQPALLVR**HPQ**GDLVALVEHHEGVDRGLVALPELHAEELGEPVGDLVQGPVEQVQGVVDALVWRLPPS
SB2	2	MDEKTHCFHPGDHLVRLVEELQALAEGLQRQGGRQPHRLPRRRPHHLQLLLDEA**HPQ**AGPLRERAHQVDGRLLLQH**HPQ**GDRLLQQPQDHPLELVWRLPPS
SB3	4	MTRRPTASSSSCVRHLLLRQGEHGHQALEDRDKARHVRLVEGDVEVLGGLDRLARARHEAL**HPQ**AGLVHLPLHGGDLGGHLRLVLEA**HPQ**GDRLGLAVHHH
SB4	1	MDEKTHWGISTWRGEPLLH**HPQ**AGRLPLDRRRARHRRILGAEPGGVDHGLRLELLDDHRPLVPDH**HPQ**RGPLQRGDLPQVVPLVRLRHAHVLGLGLAAATIT
SB5	3	MDEKTHWVNVY**HPQ**GDLLVRGHGHDVEALHDQGLHQLDLLVGPPPEVVRALRGEVLGGLRRLVPLD**HPQ**GEALDQARQRPQHLLELHHRALPPALVWRLPPS
SB6	1	MDEKTHWLNNFEELLARLDGLREGEDHPLVLRH**HPQ**GDGLLDQPLGRHRALDGEVREGDRPLDQGGEEDLGALVDDDGEVLDGLVHVGVHVHDPLVCGCHHH
SB7	1	MDEKTHWFGTLNSFPTHWMSAVGNGKIDCSFNMNLSLNHWLSSGHPDGALDDQL**HPQ**GDALVGRDDGVVQALRLEGQHQHRRLAQRRADRHRQLVWRLPPS
SB8	1	MDEKTHCTIELNFSFTHWKLHH**HPQ**GDALLDDGVRPHHPLADEGGGLDQGLGHRRGVVAERLARRDPEVLEGLVERHRGLVPRLRHGGERHAEPLVWRLPPS
SB9	1	MDEKTHCNTGLYDGAADCFNELNKDVAPLVEGRHDLVEGLLLER**HPQ**GDPLVAHRQLVHHPLLGRGERHRRALVPQQEHQPHRLQPVVDLGRRRLVWRLPPS
SB10	1	MDEKTHWHERAQELVGGLLLHD**HPQ**RLLLEPRGPRPLRGLVHERGHQPQPLAGRVEEADGGLLRDGGGELEPLVREGEDHLEPLDDELDAGPRGLVWRLPHHH
SB11	1	MDEKTHWHERVHHLADGLEQ**HPQ**GQRRPLVERHRQVPRGLVRELQHEGLPLEHPAGVHVIRLHQGDDRDVDGLVDGHGRDVRGLEREVGDGPHRLVWRLPPS
SB12	4	MDKDPLLEELEELRERLVH**HPQ**GGLLPLRGQVGHDAERLGAEVDDLRGGLLDEPQRAVAGLHHVPHRVGQRLVHEVRELDEGLLDQRDDLRQRLVWRLPPS
SB13	2	MEREDPLDEQLRELREALVD**HPQ**GGAQALHRHDGGEHVPLRRVQHRLQPGLQHHLEPQPLGLLGELQARLQPLAGEHEGDGAGLQRVPGHQGRRLVWRLPPS
SB14	1	MDEKTHRTLSVSLSFNDWLGQTKACWRLVEGLHG**HPQ**GLVREHEVDVLPLAEEVQQVVGGLADGVEQPGGGLLHRAQRVDHPLPDHAGQVLGRLVWRLPPS
SB15	1	MDEKTHWLEDLKGVLKDCLKDLMDFTKDCRSPRVQPQPLLHHDRGEPVPLLREAGRDLGGLGPRAPRQARPLHHGRHDLHEPLVLQD**HPQ**GGPLVCGCHHH
SB16	1	MDEKTHWVLQL**HPQ**GDRLGPRHGGDDVRLVGQGEGVLEGLDGRPRRRRHRLPREDEHRVRALVDQVRDLAERLVEEVDGGVEALRHLGLPQDEPRSGGCHHH
SB17	2	MDEKTHWVGDLQEPLGPLHGGVGEVPGGLVLRH**HPQ**RDRLVDGVGPHGRALARRPHRVVEGLHHLLQRGGERLPPDGPRQLGLLGGELDRADPALVWRLPPS
SB18	1	MDEKTHCAVNVNVGLTHWCHRVAHLQPLDP**HPQ**GDHLRLEPLGHALVDPLVQGVEEVVRPLQLDVGVQRVALVEQVAEVGEGLDHEAGQAHGALVWRLPPS
SB19	1	MDEKTTGWRGGHVVEGLAGELEQLRARLEH**HPQ**GQREPLVQEVEDVDEGLVQDLHGVVAGLLDPVEKLLTDWFKKFKNVSKDCKMTFYLEMYDWSGGCHHH
SB20	1	MNEKTHCKLNFKVNIADWLAEFHGGGQGLLGRRDGVVQRLVDGVQERVERLDRDPGLGDLRLELHHRDHRLRLGGEHLLRDHPLEPDDHLVVGGLVWRLPPS

The ''No.'' indicates the number of times each sequence was observed. The HPQ sequence is in bold type. Defined sequences at the termini are underlined. The six C-terminal residues are not shown.

acid/2.5 mM EDTA (TBE), 0.1% Triton X-100, and 5% glycerol. The gel (prerun for 30 min) and running buffer were precooled to 4°C and run in the cold room (the temperature of the gel was about 20°C) for 45–120 min, depending on the mobility of the particular DTP. The gel was then fixed in 10% acetic acid, 10% methanol for 15 min, transferred to electrophoresis paper (Ahlstrom, Mt. Holly Springs, PA), dried, and analyzed with a PhosphorImager (Molecular Dynamics). The concentration of DTPs was <1 nM in each titration, and thus the K_D can be approximated by the concentration of SA that results in half of the DTPs being mobility-shifted. To determine the K_D, several different measurements were taken in the range of 25–75% of DTPs bound (values outside of this range were unreliable because of background and close proximity of the bound and unbound bands in the gel). The K_D was determined by using the equation K_D = [SA]*R, where R is the ratio of unbound to bound DTPs (ratio of unshifted to shifted band). Independent measurements on gels prepared at different times were used for each clone (the number of different measurements, n, is shown in Table 2). SA concentrations were measured by UV_{282}, by using the molar extinction coefficient of 57,000 per monomer (25). Deletion constructs of clone SB19 were generated by amplification with nested PCR primers.

For the SA-coated plate-based binding assay, the following [^{35}S]methionine labeled protein was generated by *in vitro* translation: MDYKDDDDKMDEKTTGWRGGHVVEGLAGELEQLRARLEHHPQGQREPGSHHHHHHMGM. After translation, it was purified under native conditions on FLAG M2 agarose and Ni-NTA agarose as described (22), and then dialyzed into SA-binding buffer. Radiolabeled peptide (100 pM) was then incubated with a range of SA concentrations spanning 10 pM to 10 μM, or with no SA. After incubation at 4°C for 1 hour, these samples were transferred to a Reacti-Bind Streptavidin High Binding Capacity Coated 96-well Plate (Pierce) and incubated at 4°C for 5 min. The supernatant was then counted in a scintillation counter. During this brief incubation, 22–24% of the counts bound to the plate (in the absence of competing SA). This experiment was repeated in duplicate and gave a K_D of 2.4 nM each time. All binding was inhibited by either 10 μM SA or 1 mM biotin.

Results and Discussion

Generation of the Peptide Library. A DNA library encoding polypeptides of length 108 was synthesized as described (22). The library consists of short cassettes (each encoding 11 amino acids) concatamerized together. Forty-four percent of the incorporated cassettes encode a random peptide with a pattern of polar versus nonpolar amino acid side chains compatible with forming an amphipathic α-helix; 45% were similarly patterned to form β-strands; 11% were unpatterned random sequence (22). After concatamerization, the random region is 88 amino acids long, followed by a C-terminal invariant region containing a hexahistidine tag (Fig. 1).

The library was transcribed by T7 RNA polymerase (Fig. 1), after which a "linker" oligonucleotide was added to the 3′ end by using T4 DNA ligase (9, 22). The linker consists of an oligo-dA stretch, followed by 3 triethylene-glycol-phosphate units, followed by the sequence dA-dC-dC-puromycin (9). This puromycin-terminated mRNA was translated *in vitro*, yielding 1.2×10^{14} polypeptides linked via the puromycin moiety to their encoding mRNAs. These mRNA-displayed peptides were then purified on oligo-dT cellulose and Ni-NTA agarose, and the mRNA portion was then reverse transcribed, resulting in a library consisting of 6.7×10^{12} different displayed polypeptides.

Selection of SA-Binding Peptides. We chose SA as a protein target for selection because it does not normally bind to peptides, and because it has been the target of several previous phage display selections (14, 24, 26–31). The mRNA display library was incubated with immobilized SA, washed, and then eluted with biotin (Fig. 1). These molecules were then amplified by PCR to complete the first round of selection. Six subsequent rounds were carried out, and the fraction of the library that bound and eluted from the SA column increased in each round, reaching 61% at round seven (Fig. 2A).

The eluate from the seventh round of selection was amplified by PCR. The resulting PCR DNA was used to synthesize a library of displayed peptides. Treatment of the library with RNase A did not reduce the extent of binding/elution from the matrix (Fig. 2B), confirming that the peptides, rather than the RNA portion of the library constructs, were responsible for the interaction

Table 2. Binding characteristics of selected peptides

Peptide	Structure	% binding and eluting	K_d, nM	Standard deviation (n)
Strep-tag II	SNWS**HPQ**FEK	0.04	13,000–72,000	
Nonselected	**HPQ** 19 **HPQ**	0.16		
Two **HPQ** motifs				
SB1	**HPQ** 19 **HPQ**	86	50	5.7 (4)
SB2	**HPQ** 19 **HPQ**	48	4.8	0.91 (8)
SB3	**HPQ** 23 **HPQ**	20		
SB4	**HPQ** 43 **HPQ**	49		
SB5	**HPQ** 52 **HPQ**	72	110	22 (6)
One **HPQ** and one similar tripeptide motif				
SB6	HPL 4 **HPQ**	49		
SB7	HPD 7 **HPQ**	28		
SB8	**HPQ** 12 HPL	27		
SB9	**HPQ** 12 HPL	64		
SB10	**HPQ** 21 QPQ	15		
SB11	**HPQ** 28 HPA	68		
SB12	**HPQ** 30 EPQ	73		
SB13	**HPQ** 32 EPQ	64		
SB14	**HPQ** 43 HPL	11		
SB15	QPQ 50 **HPQ**	44	92	16 (4)
SB16	**HPQ** 74 LPQ	50		
One **HPQ** motif				
SB17		8.3		
SB18		58		
SB19		85	10	1.8 (10)
SB19—C4		88	4.9	0.88 (10)
SB19—C4-FLAG			2.4	0.1 (2)
No **HPQ** motif				
SB20	HPL	34		

First column: clone name (SB1–SB20). For comparison, the *Strep*-tag II peptide (24) and a nonselected sequence with 2 HPQ motifs spaced by 19 residues are also shown. SB-19–C4 is a deletion mutant of peptide SB19 (see text). The clones are grouped according to the number of HPQ and similar tripeptide motifs they possess. Second column: the tripeptide motifs in each clone and the number of amino acid residues separating them. Third column: percentage of peptide binding and specifically eluting from the SA column. Fourth column: K_D, when known, for the interaction between SA and the peptides, as measured in the EMSA (except for the *Strep*-tag II sequence; values from refs. 24 and 33); the K_D for SB10–C4-FLAG was measured by the competition assay; Fig. 4). The standard deviation in the K_D is shown in the fifth column, based on the number of independent measurements (n, shown in parentheses).

with SA. Biotin-saturated SA showed no binding to the peptide library (Fig. 2B), demonstrating that the interaction of the selected peptides with the SA matrix is specific for the unliganded protein, rather than for any other component of the matrix.

Sequence Analysis of Selected Peptides. Thirty-three randomly chosen clones from the PCR DNA from round seven were chosen for sequencing. Twenty different sequences were observed (Table 1). Surprisingly, all 20 sequences were frame-shifted from the intended frame (frame 1) to frame 3 (achieved by deleting two or adding one nucleotide). The designed pattern of polar and nonpolar residues was therefore discarded, leaving an essentially random sequence. Before the selection, about half of the library members were in frame 1 throughout their entire open reading frames (22). Frame 3 appears to have been enriched over frame 1 because of the increased frequency of the sequence HPQ, which is known to bind SA with micromolar affinity (24, 32). Frame 1 has a low incidence (1:45,000 library members) of the sequence HPQ because of the designed polar/nonpolar pattern. By contrast, frame 3 has a much higher expected incidence of the HPQ sequence (1:64), similar in frequency to that of a library of the same length and with equal mixtures of all four nucleotides at each position (1:193). Frame 2 has a high incidence of stop codons.

Nineteen of the 20 clones have at least one HPQ motif, and five clones contain two such motifs (Table 1). The number of amino acids between the two motifs, when present, ranged from 4 to 74 (Table 2). In addition to the HPQ sequence, some of the flanking positions show a nonrandom composition, but this may simply reflect the patterned structure of the library, which limits the subset of available residues at each position relative to the HPQ core.

Binding Affinities of Peptides. The highest reported affinity for a peptide with an internal HPQ sequence under reducing conditions is in the micromolar range (27, 31–33). We compared our selected clones to that of the decapeptide SNWSHPQFEK ("*Strep*-tag II"), which binds SA with a K_D of 13–72 μM (24, 33). To rapidly screen each of the 20 clones, a method for preparing, tagging, and purifying the peptides was used. First, the mRNA for individual clones was photocrosslinked to a puromycin-containing DNA oligonucleotide (23); these templates were translated *in vitro* with added [^{35}S]methionine and then purified on oligo-dT resin, as above. RNase A was then added to degrade the mRNA portion, leaving peptides fused to a short DNA oligonucleotide. These DTPs were then assayed for binding to immobilized SA and compared with the *Strep*-tag II sequence prepared in the same way.

The results of this analysis are shown in Table 2. All 20 of the selected peptides bound to a much higher degree (200- to 2,200-fold) than the *Strep*-tag II peptide. For another comparison, we introduced 2 HPQ motifs, separated by 19 residues, into an unselected member of the library. It bound only 4-fold better than the single *Strep*-tag II sequence, showing that the presence of 2 HPQ motifs is not sufficient for high-affinity binding.

We measured the dissociation constants of the selected peptides for SA by using an EMSA. The short DNA oligonucleotide tag on the DTPs allows them to migrate in a native gel, and the addition
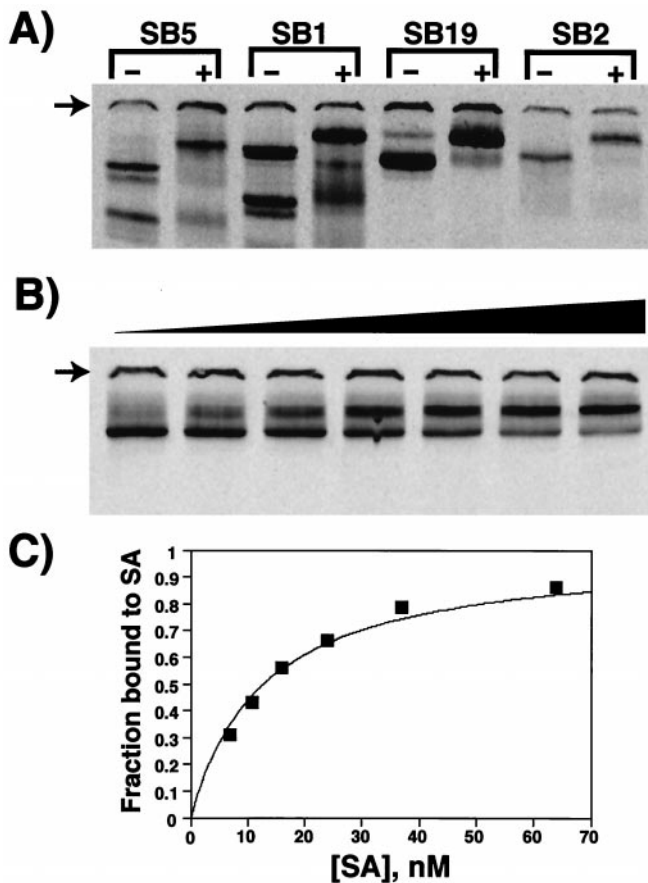
Fig. 3. EMSA analysis of peptide-SA interactions. (*A*) Qualitative demonstration of the binding of four different DNA-tagged peptides to SA. The migration of each clone is shown in the absence (−) and presence (+) of 1 μM SA. Some of the clones show multiple bands, presumably representing different conformations. The arrows show the position of the gel well, which often contains a fraction of the counts. (*B*) Titration of clone SB19 (full-length) with SA. The SA concentration in each lane, from left to right, is: 3.8, 6.6, 10, 15, 23, 35, and 61 nM. (*C*) Curve fitted to the data shown in *B* (the fraction of peptide bound could not be accurately determined for the point with the lowest concentration of SA). Assuming that the peptide is homogeneous and 100% active, the data from this experiment give a K_D of 11 nM ($K_D = 10 +/- 1.8$ nM from multiple measurements; see Table 2).

of unlabeled SA causes a mobility shift for several of the clones. Examples of these shifts are shown in Fig. 3*A*. Other clones showed either no shift or poorly defined bands, suggesting that the lifetime of these complexes in the gel was too short for detection by this method. We then chose five clones and quantitatively examined their mobility shifts in response to a range of SA concentrations. An example of an SA titration for clone SB19 is shown in Fig. 3 *B* and *C*. The dissociation constants for the clones range from over 100 nM to as low as 5 nM (Table 2). These affinities are comparable to those for monoclonal antibody–antigen interactions.

Dissection of Clone SB19. We further characterized clone SB19, which has a single HPQ motif. A series of deletion constructs were assayed in the SA column-binding assay (Table 3). Construct C4, which retains only the first 38 residues from the selected construct (plus the C-terminal invariant sequence SG-GCKLG), retains full binding activity in this assay. Mutating the HPQ motif to HGA reduced the extent of binding by three orders of magnitude (compare construct C4 to M1). Comparison of the N-terminal deletion constructs (N1-N3) suggests that binding determinants are spread throughout the N-terminal 38 residues of the peptide. Construct C4 is therefore the minimal peptide retaining full activity in this assay. EMSA of construct C4 (not shown) confirmed high-affinity SA binding. A fraction (13%) of the peptide was inactive even at SA concentrations >1 μM, but the majority (87%) has an apparent K_D of 4.9 nM.

To demonstrate that this high-affinity binding did not require the DNA tag, we generated the free C4 peptide with a FLAG tag and purified it on a FLAG-affinity matrix. This peptide bound to SA immobilized on a microtiter plate. We measured the binding affinity by competition with varying concentrations of free SA (Fig. 4; $K_D = 2.4 +/- 0.1$ nM).

Conclusions

SA has been used as a model target protein for numerous phage display selections from libraries of linear peptides ranging in length from 5 to 38 amino acids (24, 26–29, 31). The primary isolates from such selections have dissociation constants of >100 μM, and the HPQ consensus peptide that emerged from these selections was subsequently optimized to the 10-aa-long "*Strep*-tag II" sequence (24). The K_D of this peptide for SA has been measured at 72 μM (24) or 13 μM (33), depending on the method. By contrast, using the ultrahigh complexity libraries accessible by mRNA display, we obtained primary isolates with binding affinities as low as 5 nM, demonstrating that even random nonconstrained peptide libraries can be a source of avid ligands to proteins that do not normally function in peptide binding. The library used in this study differed from previous phage display libraries in its sequence complexity (>10^4-fold more sequences sampled), the length of the random displayed peptides (88 vs. 5–38 amino acids), and an increased incidence of HPQ sequences (about 3-fold per unit length) because of amino acid composition bias. Library diversity is the most dramatic of these differences and is probably the dominant factor

Table 3. Deletion analysis of clone SB19

Type	Sequence	% binding
FL	MDEKTTGWRGGHVVEGLAGELEQLRARLEH**HPQ**GQREPLVQEVEDVDEGLVQDLHGVVAGLLDPVEKLLTDWFKKFKNVSKDCKMTFYLEMYDWSGGCKLG	85
C1	MDEKTTGWRGGHVVEGLAGELEQLRARLEH**HPQ**GQREPLVQEVEDVDEGLVQDLHGVVAGLLDPVEKLLTDWFKKFKNVSMMSGGCKLG	87
C2	MDEKTTGWRGGHVVEGLAGELEQLRARLEH**HPQ**GQREPLVQEVEDVDEGLVQDLHGVVAGLLDPVEMMSGGCKLG	88
C3	MDEKTTGWRGGHVVEGLAGELEQLRARLEH**HPQ**GQREPLVQEVEDVDEGLVQMMSGGCKLG	89
C4	MDEKTTGWRGGHVVEGLAGELEQLRARLEH**HPQ**GQREPMMSGGCKLG	88
M1	MDEKTTGWRGGHVVEGLAGELEQLRARLEH**HGA**GQREPMMSGGCKLG	0.065
N1	MDGHVVEGLAGELEQLRARLEH**HPQ**GQREPMMSGGCKLG	69
N2	MDEGLAGELEQLRARLEH**HPQ**GQREPLVQEVEDVDEGLVQDLHGVVAGLLDPVEKLLTDWFKKFKNVSKDCKMTFYLEMYDWSGGCKLG	30
N3	MELEQLRARLEH**HPQ**GQREPMMSGGCKLG	0.058

The full-length (FL), C-terminal deleted (C1–C4), N-terminal deleted (N1–N3), and point mutated (M1) peptide sequences are shown. The "% binding" refers to the performance of these peptides in the SA column-binding assay.
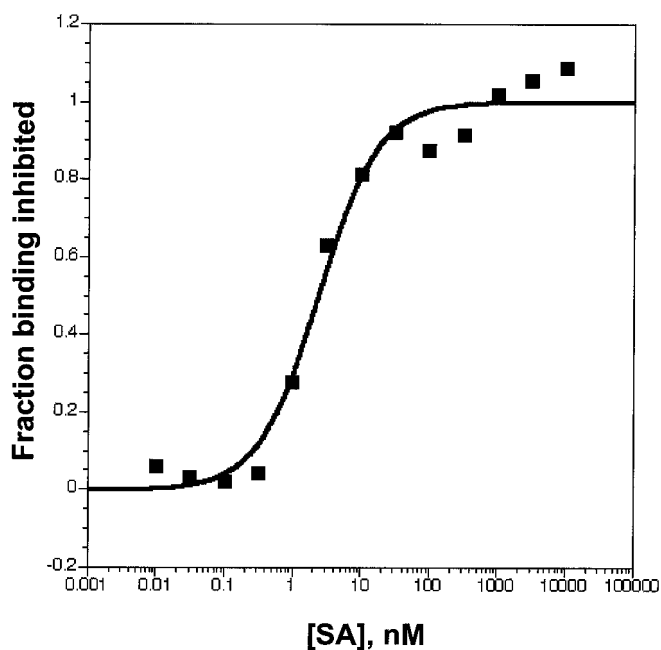
Fig. 4. Binding of peptide SB19-C4-FLAG to immobilized streptavidin. [^{35}S]methionine labeled SB-19 was incubated with a range of concentrations of SA for 1 hour before being transferred to an SA-coated plate and incubated for 5 min. The y axis shows the fraction of surface-bound peptide that is competed by the free SA. This experiment was repeated in duplicate and gave a K_D of 2.4 nM each time.

contributing to the isolation of higher-affinity peptides. However, the use of longer peptides may also have been critical, because our 38-aa peptide could only have been isolated from the 38-aa phage display library and not from the other shorter libraries. Our use of longer peptides may also have effectively increased library diversity by another 50-fold, because a given 38-aa peptide could occur in any of 50 registers in the 88-aa random region.

The affinities of the SA-binding peptides could presumably be further improved by evolutionary optimization. The affinity of simple peptide ligands can also be dramatically increased by introducing disulfide constraints (14, 30) or by constraining them

as loops within known structural scaffolds (12). By combining these library design features with the high-complexity libraries that can be generated by mRNA display, it may be possible to select subnanomolar binding peptides to proteins.

Structures of HPQ-containing peptides bound to SA have been determined by x-ray crystallography (24, 31, 32, 34). The HPQ motif inserts itself into the biotin-binding cleft and forms hydrogen bonds and hydrophobic interactions with SA (24, 34). The high affinity of the HPQ-containing clones isolated in the present study could result from either of two different factors. First, the flanking amino acids could stabilize the active conformation of the HPQ tripeptide motif; this type of mechanism appears to be responsible for the high-affinity binding of a disulfide-constrained HPQ-containing peptide to SA ($K_D = 230$ nM; refs. 30, 34). Second, several distinct peptide elements could be interacting with different surfaces on SA, providing additional binding energy from either specific or non-specific interactions.

These SA-binding peptides may provide useful affinity tags for protein purification. The *Strep*-tag II sequence (SNWSH-PQFEK), which binds to SA with a K_D of ≈13 μM, has been used for this purpose. Effective column retention requires the use of an SA mutant with a 13-fold higher affinity for the *Strep*-tag II sequence (33, 35). The 38-aa SB19-C4 peptide identified in the present study allows for a one-step purification of proteins from *Escherichia coli* extract by using normal SA and results in a high yield of very pure fusion protein (data not shown).

Complete expression and activity analysis of the proteome will require the use of hundreds of thousands of protein-binding and inhibiting reagents. Traditional methods for generating monoclonal antibodies are expensive and time-consuming; *in vitro* selection of binding peptides from polypeptide libraries may help to meet this need. This study shows that, given sufficient library complexity, even random peptide libraries can be used to generate ligands with strong binding constants ($K_D = 5$ nM). It should be possible to increase the affinities of peptides for their targets by constraining them as loops in known protein scaffolds (12). Such constrained polypeptides may also have superior properties with respect to solubility and stability against proteolysis.

1. Colas, P. (2000) *Curr. Opin. Chem. Biol.* **4,** 54–59.
2. Smith, G. P. & Petrenko, V. A. (1997) *Chem. Rev.* **97,** 391–410.
3. Perelson, A. S. & Oster, G. F. (1979) *J. Theor. Biol.* **81,** 645–670.
4. Vaughan, T. J., Williams, A. J., Pritchard, K., Osbourn, J. K., Pope, A. R., Earnshaw, J. C., McCafferty, J., Hodits, R. A., Wilton, J. & Johnson, K. S. (1996) *Nat. Biotechnol.* **14,** 309–314.
5. Hanes, J. & Pluckthun, A. (1997) *Proc. Natl. Acad. Sci. USA* **94,** 4937–4942.
6. Hanes, J., Jermutus, L., Weber-Bornhauser, S., Bosshard, H. R. & Pluckthun, A. (1998) *Proc. Natl. Acad. Sci. USA* **95,** 14130–14135.
7. Roberts, R. W. & Szostak, J. W. (1997) *Proc. Natl. Acad. Sci. USA* **94,** 12297–12302.
8. Roberts, R. W. & Ja, W. W. (1999) *Curr. Opin. Struct. Biol.* **9,** 521–529.
9. Liu, R., Barrick, J. E., Szostak, J. W. & Roberts, R. W. (2000) *Methods Enzymol.* **318,** 268–293.
10. Clackton, T. & Wells, J. A. (1994) *Trends Biotechnol.* **12,** 173–184.
11. Katz, B. A. (1997) *Annu. Rev. Biophys. Biomol. Struct.* **26,** 27–45.
12. Nygren, P. A. & Uhlen, M. (1997) *Curr. Opin. Struct. Biol.* **7,** 463–469.
13. O'Neil, K. T., Hoess R.H., Jackson, S. A., Ramachandran, N. S., Mousa, S. A. & DeGrado, W. F. (1992) *Proteins* **14,** 509–515.
14. McLafferty, M. A., Kent, R. B., Ladner, R. C. & Markland, W. (1993) *Gene* **128,** 29–36.
15. DeLano, W. L., Ultsch, M. H., de Vos, A. M. & Wells, J. A. (2000) *Science* **287,** 1279–1283.
16. Clackson, T., Hoogenboom, H. R., Griffiths, A. D. & Winter, G. (1991) *Nature (London)* **352,** 624–628.
17. Persson, M. A., Caothien, R. H. & Burton, D. R. (1991) *Proc. Natl. Acad. Sci. USA* **88,** 2432–2436.
18. Marks, J. D., Hoogenboom, H. R., Griffiths, A. D. & Winter, G. (1992) *J. Biol. Chem.* **257,** 16007–16010.
19. Orum, H., Andersen, P. S., Oster, A., Johansen, L. K., Riise, E., Bjornvad, M., Svendsen, I. & Engberg J. (1993) *Nucleic Acids Res.* **21,** 4491–4498.
20. Sheets, M. D., Amersdorfer, P., Finnern, R., Sargent, P., Lindqvist, E., Schier, R., Hemingsen, G., Wong, C., Gerhart, J. C. & Marks, J. D. (1998) *Proc. Natl. Acad. Sci. USA* **95,** 6157–6162.
21. de Haard, H. J., van Neer, N., Reurs, A., Hufton, S. E., Roovers, R. C., Henderikx, P., de Bruine, A. P., Arends, J. W. & Hoogenboom, H. R. (1999) *J. Biol. Chem.* **274,** 18218–18230.
22. Cho, G., Keefe, A. D., Wilson, D. S., Liu, R. & Szostak, J. W. (2000) *J. Mol. Biol.* **297,** 309–319.
23. Kurz, M., Gu, K. & Lohse, P. A. (2000) *Nucleic Acids Res.* **28,** E83.
24. Schmidt, T. G. M., Koepke, J., Frank, R. & Skerra, A. (1996) *J. Mol. Biol.* **255,** 753–766.
25. Green, N. M. (1970) *Methods Enzymol.* **18,** 418–424.
26. Devlin, J. J., Panganiban, L. C. & Devlin, P. E. (1990) *Science* **249,** 404–406.
27. Lam, K. S., Salmon, S. E., Hersch, E. M., Hruby, V. J., Kazmierski, W. M. & Knapp, R. J. (1991) *Nature (London)* **354,** 82–84.
28. Kay, B. K., Adey, N. B., He, Y.-S., Manfredi, J. P., Mataragnon, A. H. & Fowlkes, D. M. (1993) *Gene* **128,** 59–65.
29. Schmidt, T. G. M. & Skerra, A. (1993) *Protein Eng.* **6,** 109–122.
30. Giebel, L. B., Cass, R. T., Milligan, D. L., Young, D. C., Arze, R. & Johnson, C. R. (1995) *Biochem. J.* **34,** 15430–15435.
31. Katz, B. A. (1999) *Biomol. Eng.* **16,** 57–65.
32. Weber, P. C., Pantoliano, M. W. & Thompson, L. D. (1992) *Biochem. J.* **31,** 9350–9354.
33. Voss, S. & Skerra, A. (1997) *Protein Eng.* **10,** 975–982.
34. Katz, B. A. (1995) *Biochem. J.* **34,** 15421–15429.
35. Maier, T., Drapal, N., Thanbichler, M. & Bock, A. (1998) *Anal. Biochem.* **259,** 68–73.
36. Sleat, D. E., Gallie, D. R., Jefferson, R. A., Bevan, M. W., Turner, P. C. & Wilson, T. M. (1987) *Gene* **60,** 217–225.

BIOCHEMISTRY