

Epitope mapping using mRNA display and a unidirectional nested deletion library

William W. Ja¹, Brett N. Olsen^{1,2} and Richard W. Roberts^{1,3}

¹Division of Chemistry and Chemical Engineering, California Institute of Technology, M/C 147-75, Pasadena, CA 91125, USA

²Present address: Program in Cellular and Molecular Biology, University of Wisconsin, Madison, WI 53706, USA

³To whom correspondence should be addressed.
E-mail: rroberts@caltech.edu

***In vitro* selection targeting an anti-polyhistidine monoclonal antibody was performed using mRNA display with a random, unconstrained 27-mer peptide library. After six rounds of selection, epitope-like peptides were identified that contain two to five consecutive, internal histidines and are biased for arginine residues, without any other identifiable consensus. The epitope was further refined by constructing a high-complexity, unidirectional fragment library from the final selection pool. Selection by mRNA display minimized the dominant peptide from the original selection to a 15-residue functional sequence (peptide Cmin: RHDAGDHHHHGVRQ; $K_D = 38$ nM). Other peptides recovered from the fragment library selection revealed a separate consensus motif (ARRXA) C-terminal to the histidine track. Kinetics measurements made by surface plasmon resonance, using purified Fab (antigen-binding fragment) to prevent avidity effects, demonstrate that the selected peptides bind with 10- to 75-fold higher affinities than a hexahistidine peptide. The highest affinity peptides ($K_D \approx 10$ nM) encode both a short histidine track and the ARRXA motif, suggesting that the motif and other flanking residues make important contacts adjacent to the core polyhistidine-binding site and can contribute >2.5 kcal/mol of binding free energy. The fragment library construction methodology described here is applicable to the development of high-complexity protein or cDNA expression libraries for the identification of protein–protein interaction domains.**

Keywords: anti-polyhistidine/epitope mapping/HIS-1/*in vitro* selection/mRNA display/nested deletion/peptide libraries

Introduction

Epitope mapping, the identification of regions of an antigen recognized by an antibody, is an important subset of protein–protein interaction analysis that is relevant in a wide range of disciplines where antibodies are used as molecular reagents. Conventional methods for epitope mapping involve the synthesis or expression of numerous overlapping polypeptides followed by probing for antibody reactivity (Geysen *et al.*, 1984; Lenstra *et al.*, 1990; Frank, 1992; Frank and Overwin, 1996; Kramer *et al.*, 1999; Reineke *et al.*, 1999). Although these methods can achieve very fine mapping (single amino acid resolution), they involve tedious, time-consuming and

often cost-intensive steps. These techniques also typically require *a priori* knowledge of one of the interacting partners (i.e. the antigen sequence).

Display technologies such as phage (Scott and Smith, 1990) and cell surface display on *Escherichia coli* or yeast (Boder and Wittrup, 1997; Georgiou *et al.*, 1997) permit the assay of millions of polypeptides simultaneously for the identification of functional properties. In these systems, each display vehicle expresses multiple copies of a single polypeptide sequence on its surface. Active peptides are recovered by affinity selection (e.g. by biopanning or fluorescence-activated cell sorting) and identified by DNA sequencing of the library inserts. Random peptide libraries (Miceli *et al.*, 1994; Parhami-Seren *et al.*, 1997; Murthy *et al.*, 1998), antigen- or gene-fragment libraries (Kuwabara *et al.*, 1999; Christmann *et al.*, 2001; Mullaney *et al.*, 2001) or a combination of both (Stephen *et al.*, 1995; Fack *et al.*, 1997; Coley *et al.*, 2001) have previously been used for the epitope mapping of a wide variety of monoclonal antibodies (mAbs) (reviewed by Irving *et al.*, 2001). Generally, these libraries suffer from low starting complexities and do not always achieve fine mapping of antibodies unless the epitope is short (about five residues) and well defined.

More recently, entirely *in vitro* techniques for protein selection such as ribosome (Mattheakis *et al.*, 1994; Hanes and Plückthun, 1997; He and Taussig, 1997) and mRNA display (Roberts and Szostak, 1997) have emerged. In mRNA display, peptides are covalently attached to the 3'-end of their encoding mRNA via a tethered puromycin moiety. Pools of RNA–peptide fusions are selected for binding via their attached peptides and recovered fusions are RT-PCR amplified for the next round of selection and/or cloned for DNA sequencing (Figure 1). The mRNA display system generates libraries that are robust (functional in a wide variety of conditions), encode high complexities (>10¹³ unique sequences, compared with ~10⁸–10⁹ for techniques requiring an *in vivo* transformation step) and lack avidity effects as only one peptide is displayed per mRNA sequence. By accessing larger libraries, extremely rare sequences (such as long, discontinuous epitopes or peptides with better functional properties) can be selected and amplified (Takahashi *et al.*, 2003). Epitope-like consensus motifs that define the core determinants of binding for the anti-c-Myc antibody, 9E10, have previously been identified using mRNA display with a random peptide library (Baggio *et al.*, 2002).

Published methods for generating gene or fragment libraries from DNA typically involve degenerate oligonucleotide priming (Whitcomb *et al.*, 1993; Hampson *et al.*, 1996; Santi *et al.*, 2000), random fragmentation of DNA (Gupta *et al.*, 1999) or the iterative removal of bases from either end of a gene (Henikoff, 1984; Milavetz, 1992; Poes *et al.*, 1997), followed by amplification of the library. These techniques have been employed for a variety of purposes, including epitope mapping and the determination of protein interaction domains (Fack

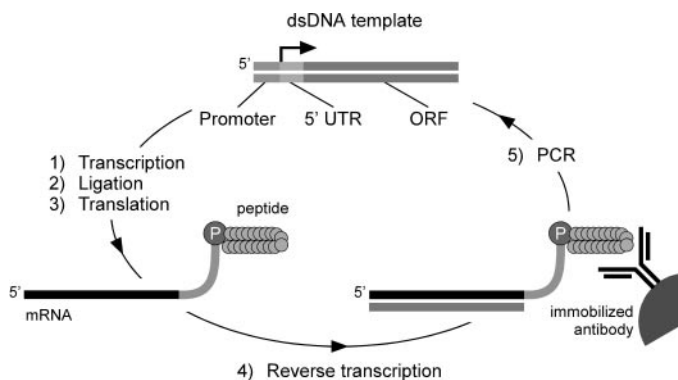


Fig. 1. *In vitro* selection scheme using mRNA display. The starting dsDNA pool (top, center) which encodes the peptide library is transcribed *in vitro*. Purified mRNA is enzymatically ligated to a puromycin–DNA oligo prior to RNA–peptide fusion formation via *in vitro* translation. Purified RNA–peptide fusions are reverse transcribed and affinity selected on to the immobilized antibody target. Eluted cDNA is used as the template for PCR for the next cycle of selection.

et al., 1997; Kuwabara *et al.*, 1999; Christmann *et al.*, 2001; Coley *et al.*, 2001; Mullaney *et al.*, 2001; McPherson *et al.*, 2002). Because of the random nature of library construction, the majority of sequences in these libraries are non-viable owing to frameshifts and ligations in the anti-sense orientation. Techniques have been described to maintain gene orientation using a pair of degenerate primers with constant 5' sequences used sequentially in the amplification of cDNA (Hampson *et al.*, 1996) or mRNA (Hammond *et al.*, 2001; McPherson *et al.*, 2002). However, these methods are technically challenging and may be prone to poor library coverage owing to biased hybridization to target sequences (Telenius *et al.*, 1992; Zhang and Byrne, 1999).

A further advancement of mRNA display technology is described here, where we describe a robust method for generating unidirectional, nested deletion libraries. As mRNA display facilitates selection from pool sizes larger than previously possible, improvements are needed for generating libraries with broad coverage while maintaining high sequence complexity. Here, parent DNA sequences are partially digested with DNase I. The resulting fragments are then directionally amplified, maintaining the sense orientation and used to generate an mRNA display library. We used this method to identify a 15-mer peptide that binds with high affinity to an anti-polyhistidine mAb from an initially enriched population of 35-mer, epitope-like sequences. The fragment library selection also revealed a new motif important for high affinity, demonstrating how sequence length may be an important factor in delineating a specific binding requirements. The methods described here should be highly applicable towards the isolation of minimal protein interaction domains from cDNA or protein expression libraries using mRNA display.

Materials and methods

General

L-[³⁵S]Methionine (³⁵S-Met) was purchased from PerkinElmer Life Sciences. Other reagents and solvents were obtained from Sigma-Aldrich or VWR International, unless noted otherwise. All buffer components for RNA and RNA–peptide fusions were made with diethyl pyrocarbonate-treated doubly distilled

water. DNA oligos were synthesized at the Caltech Biopolymer Synthesis and Analysis Facility and were desalted by OPC purification, with the exception of DNA template 130.2, which was obtained from the W. M. Keck Foundation Biotechnology Resource Laboratory (<http://keck.med.yale.edu>) and purified by urea–PAGE (Ellington and Pollard, 2001). Oligo and peptide concentrations were determined by UV spectrophotometry using a calculated extinction coefficient (<http://paris.chem.yale.edu/extinct.html>). Protein concentrations were determined by measuring the UV absorbance at 205 nm (Scopes, 1974). The values obtained with this method were within 5% of those obtained using a calculated extinction coefficient at 280 nm for the protein (<http://paris.chem.yale.edu/extinct.html>).

mRNA display library construction

Construction of a random peptide library for mRNA display selections has been described in detail previously (Liu *et al.*, 2000; Keefe, 2001; Baggio *et al.*, 2002). Briefly, the anti-sense DNA oligo 130.2 [5'-AGC GCA AGA GTT ACG CAG CTG (SNN)₂₇ CAT TGT AAT TGT AAA TAG TAA TTG TCC C; S = C or G, N = A, C, G or T] was PCR amplified with primers 47T7FP (5'-GGA TTC TAA TAC GAC TCA CTA TAG GGA CAA TTA CTA TTT ACA ATT AC) and mycRP (5'-AGC GCA AGA GTT ACG CAG CTG) to produce the initial template containing a T7 promoter, a 5'-untranslated region (UTR), an ATG methionine start codon, 27 random amino acids and a constant 3'-end that encoded the peptide, QLRNSCA. *In vitro* transcription, purification of the mRNA and splint-mediated ligation of the puromycin linker oligo (pF30P: 5'-A₂₁[S9]₃ACC-P, S9 = spacer phosphoramidite 9, P = puromycin, 5'-phosphorylated with phosphorylation reagent II, Glen Research; splint: 5'-TTT TTT TTT TTN AGC GCA AGA GT) were performed as described (Ja and Roberts, 2004) to produce puromycin-conjugated templates (mRNA–F30P).

RNA–peptide fusion preparation and selection

Purified mRNA–F30P templates were translated in rabbit reticulocyte lysate (Red Nova lysate, Novagen) according to the manufacturer's instructions with optimized conditions (100 mM KOAc, 0.5 mM MgOAc and 0.5 μM mRNA–F30P) and additional L-Met (0.5 mM final, 1 ml total reaction volume) or ³⁵S-Met labeling (150 μl reaction). Following the incubation step at 30°C, KOAc and MgCl₂ were added to 585 and 50 mM (final), respectively, and the reactions were incubated on ice for 15 min to facilitate RNA–peptide fusion formation. Radioactively labeled and non-labeled RNA–peptide fusions were pooled and subsequently purified with oligo dT-cellulose (New England Biolabs) as described (Ja and Roberts, 2004). Purified fusions were concentrated (Microcon YM-30, Millipore) and reverse transcribed according to the manufacturer's instructions (Superscript II, Invitrogen) with excess mycRP primer in a 100 μl reaction.

The matrix preparation and all selection steps were performed at 4°C. The reverse-transcribed fusions, in 1 ml of selection buffer (50 mM HEPES–KOH, pH 7.5, 100 mM NaCl, 10 mM MgCl₂, 10 mM NaF, 30 μM AlCl₃, 0.05% Tween 20, 1 mM β-mercaptoethanol (β-ME) and 5 μM GDP), were pre-cleared by rotating with 20 μl of protein G–Sepharose (4B Fast Flow, Sigma) for >1 h. The supernatant was transferred to the target matrix [80 μg of His₆–TEV–G_{iα1} (Lee *et al.*,

1994) immobilized by 40 µg of anti-polyhistidine mAb (clone HIS-1, catalog No. H1029, Sigma) on 20 µl of protein G–Sepharose] and rotated for 1 h. The matrix was washed with 3 × 1 ml of selection buffer and the bound RNA–peptide fusions were eluted with 2 × 200 µl of 4% acetic acid through a 0.45 µm spin filter (SpinX, Costar). Washes and an aliquot of the elution were scintillation counted (LS 6500, Beckman Coulter) to determine the amount of bound fusions.

The eluted fusions were either desalted and concentrated by ultrafiltration (Microcon YM-30) or frozen and dried by vacuum centrifugation. After resuspension in doubly distilled water or 10 mM Tris–HCl, pH 8, samples were PCR amplified for the next cycle of selection and/or for DNA sequencing (TOPO TA cloning, Invitrogen). Subsequent selection rounds were performed similarly, except that smaller translation reactions were used (300 µl non-labeled, 100 µl ³⁵S-Met-labeled). Unblocked mAb (without the His₆-tagged protein) was used as the target in the sixth round of selection, when it was realized that the peptides were specific for the mAb.

RNA–peptide fusion binding assay

Aliquots of purified ³⁵S-Met-labeled RNA–peptide fusions were treated with RNase (DNase-free, Roche) and added to ~15 µl of protein G–Sepharose matrix (with or without ~10 µg of anti-polyhistidine mAb) in 1 ml of selection buffer. Mixtures were rotated at 4°C for 1 h and washed with 3 × 1 ml of selection buffer. The percentage binding was determined by scintillation counting of the washes and the matrix.

Fragment library construction and selection

First-strand cDNA was produced from the mRNA of a selected library by reverse transcription with dUTP instead of dTTP nucleotides (Superscript II). The mRNA was subsequently removed with RNase H (Roche) and the single-stranded cDNA was purified by spin-column (QIAquick, Qiagen). To generate cDNA fragments, a partial digest was performed on 30 pmol of cDNA (~1.2 µM final concentration) with 0.25 U of DNase I (Invitrogen) in 1× DNase I buffer (10 mM Tris–HCl, pH 7.4, 2.5 mM MgCl₂ and 0.1 mM CaCl₂) at 15°C for 10 min. DNase I was removed using DNase Removal Reagent (Ambion), according to the manufacturer's instructions.

Second-strand cDNA was generated by random hexamer priming and fill-in with a processive polymerase (Sequenase v2.0, Amersham Biosciences). cDNA fragments were mixed with 125 pmol of myc6-N6-FP (5'-ATC TCT GAA GAG GAC CTG NNN NNN) in T7 reaction buffer (Amersham Biosciences). After heating the sample to 60°C and cooling on ice to anneal the primers, dNTP (200 µM each nucleotide, final), DTT (10 mM, final) and Sequenase v2.0 (13 U) were added and the reaction was incubated at 37°C for 20 min. The enzyme was heat inactivated at 90°C for 5 min. First-strand cDNA was removed by adding uracil–DNA glycosylase (UDG, 30 U, New England Biolabs) and incubating at 37°C for 20 min. After heat inactivating the UDG, ssDNA longer than ~50 bases was gel purified with QiaEX II (Qiagen) from a 4% agarose gel (Frohlich and Parker, 2001). A second fill-in reaction was performed with 3myc-N6-RP (5'-AAA TGC ACA AGA GTT GCC CTC GNN NNN N) as before. The dsDNA was subsequently purified on a 2% agarose gel (QIAquick).

PCR using primers T7mycFP (5'-GGA TTC TAA TAC GAC TCA CTA TAG GGA CAA TTA CTA TTT ACA

ATT ACA ATG GAA CAG AAA CTG ATC TCT GAA GAG GAC CTG) and psn3mycRP (5'-AAA TGC ACA AGA GTT GCC CTC G) produces the initial library (containing a T7 promoter, a 5'-UTR, an ATG methionine start codon, the fragment domain and a constant 3'-end) suitable for mRNA display selection. The PCR resulted in a smear of products ranging from 100 to 200 bp and DNA corresponding to 150–200 bp was gel purified (QIAquick) and used as the starting template. RNA–peptide fusions were prepared as described above, except that the puromycin moiety was coupled to the mRNA by UV photo-crosslinking with oligo psn-mycF15P (5'-[Ps]-UGC ACA AGA GUU G-dA₁₅-[S9]₂-dCdC-P, where unlabeled bases are 2'-OMe RNA, Ps = psoralen C6, S9 = spacer phosphoramidite 9, P = puromycin, Glen Research) as published previously (Kurz *et al.*, 2000). A modified selection buffer [1× PBS, 1 mM β-ME, 1 mM EDTA, 0.05% Tween 20, 0.2% (w/v) BSA and 1 µg/ml yeast tRNA (Roche)] was used in the fragment library selection. In the second and third rounds of selection, the matrix was more stringently washed by incubating the mAb-bound RNA–peptide fusions in buffer containing poly-L-His (0.15 mg/ml, P2534, Sigma) and His₆ peptide (60 µM, Covance Research Products) for ~40 min at 4°C (Boder and Wittrup, 1998).

Direct binding assay of *in vitro* translated peptides in lysate

Individual clones (in pCR4-TOPO vector, Invitrogen) were PCR amplified with primers 47T7FP and mycRP, *in vitro* transcribed, urea–PAGE purified and *in vitro* translated (Red Nova Lysate) with ³⁵S-Met labeling according to the manufacturer's instructions. An aliquot of the translation reaction (4 µl) was added directly to an assay tube (600 µl of fragment selection buffer, 10 µl of protein G–Sepharose, 5 µg of anti-polyhistidine mAb). After rotating at 4°C for 1 h, the Sepharose was washed with 6 × 600 µl of fragment selection buffer in a 0.45 µm spin filter (SpinX) and bound peptides were eluted with 2 × 20 µl of 0.05% SDS. Half of the sample was analyzed via tricine SDS–PAGE along with 2 µl of the original translation reaction for comparison. After electrophoresis, gels were destained (40% methanol and 10% acetic acid) for 20 min, dried under vacuum and imaged via autoradiography (Storm Phosphorimager, Amersham Biosciences). Peptide band intensities were analyzed with ImageQuant software (Amersham Biosciences).

Peptide synthesis/protein purification

Peptides were synthesized on an ABI 432A Synergy peptide synthesizer (Applied Biosystems) using Fmoc chemistry. Peptides included the sequence GGYK–NH₂ at their C-terminus, where **K** is biotinyllysine (biocytin, BAchem) and –NH₂ represents C-terminal amidation. The tyrosine residue, used for quantitation by UV absorbance, was omitted from the synthesis for peptides that already contained a tryptophan and/or tyrosine. Crude peptides were deprotected in TFA–thioanisole–1,2-ethanediol (450, 25, 25 µl, 2 h at room temperature), precipitated with methyl *tert*-butyl ether, purified to >95% purity by reversed-phase HPLC on a semi-preparative C₁₈ column (250 × 10 mm i.d., Vydac) and confirmed by MALDI-TOF mass spectrometry.

Several peptide sequences were expressed in *E. coli* as *in vivo* biotinylated maltose-binding protein (MBP) fusions using a vector derived from pDW363 (Tsao *et al.*, 1996). The MBP gene from pDW363 was amplified by successive PCR (primers 35.3 5'-GGA CTA GTA AAA TCG AAG AAG GTA AAC

TGG TAA TC and 35.4 5'-CCA TTG GAT CCT TAA TTA GTC TGC GCG TCT TTC AG, then primers 84.1 5'-GAG CAC TCG AGC GGT GCG AAT TCA AAC AAC ATC GAG GGG CGC GCC GGT GGC ACT AGT AAA ATC GAA GAA GGT AAA CTG GTA ATC and 29.3 5'-CCA TTG GAT CCT TAA TTA GTC TGC GCG TC). The PCR-amplified fragment and pDW363 were digested with *XhoI/BamHI*, purified and ligated to produce the pDW363B vector.

DNA templates encoding peptides B and C were amplified by PCR using the universal forward primer 29.4 (5'-TGA AGT CTG GAG TAT TTA CAA TTA CAA TG) and a template-specific reverse primer that added an *SpeI* site. *BpmI/SpeI*-digested dsDNA was co-ligated into *XhoI/SpeI* digested pDW363B with DNA linkers (*XhoI* linker 5'-TCG AGC TCT GGA GGC ATC GAG GGT CGC AT and *BpmI* linker 5'-GCG ACC CTC GAT GCC TCC AGA GC) to produce the expression vector. Inserts contained an N-terminal bio-tag (MAGGLNDIFEAQKIEWHEDTGGSS), peptide B or C and a C-terminal MBP fusion. The vectors produce a dicistronic mRNA which encodes the bio-tag-peptide-MBP fusion and biotin holoenzyme synthetase (*birA*), an enzyme that attaches biotin to the single lysine in the bio-tag *in vivo*. Protein expression with 30 ml cultures of *E.coli* BL21 cells was performed as described (Tsao *et al.*, 1996). Cells were lysed with B-PER (Pierce) and MBP fusions were purified on monomeric avidin-agarose (Pierce) according to the manufacturer's instructions. Purified proteins were concentrated and desalted into 1× PBS by ultrafiltration (Centriprep YM-10, Millipore).

The His₆-TEV-G_{1x1} was expressed and purified as described previously (Lee *et al.*, 1994). Briefly, *E.coli* expression cultures were lysed by French press and successively purified by FPLC on metal chelate affinity, anion-exchange and size-exclusion columns. Protein purity was assayed by SDS-PAGE. MALDI-TOF analyses on TEV protease-treated His₆-TEV-G_{1x1} were consistent with the epitope tag being removed. For the surface plasmon resonance (SPR) experiments, His₆-TEV-G_{1x1} was biotinylated using EZ-Link Sulfo-NHS-LC-biotin (5-fold excess, Pierce) according to the manufacturer's instructions. Excess reagent was quenched with ethanolamine and the biotinylated protein was desalted on an NAP-10 column (Amersham Biosciences).

Anti-polyhistidine mAb in ascites fluid was affinity purified on protein G-Sepharose in 1× PBS/0.1% Triton X-100, eluted with 0.1 M citric acid buffer, pH 3, and immediately neutralized with buffer. After concentration and buffer exchange (Centriprep YM-50) into papain buffer (20 mM phosphate, pH 7, 10 mM EDTA), the antigen-binding fragment (Fab) was generated and purified using an ImmunoPure Fab Preparation Kit (Pierce), according to the manufacturer's instructions.

Surface plasmon resonance

SPR measurements were made at 25°C on a Biacore 2000 (Biacore) equipped with either SA (streptavidin) sensor chips or research-grade CM5 sensor chips (Biacore) with amine-coupled streptavidin (ImmunoPure, Pierce). The CM5-streptavidin chips were prepared in-house by standard NHS/EDC amine coupling (Biacore) and achieved >1100 RU of immobilized streptavidin per flow cell. HBS-EP [20 mM HEPES, pH 7.4, 150 mM NaCl, 3 mM EDTA and 0.005% (v/v) surfactant P20 (Tween 20)] was used as the running buffer for all experiments. Biotinylated ligands were diluted in HBS-EP to 1 nM and immobilized to individual flow cells

(~10 RU for peptides and ~100 RU for proteins). Flow cell 1 was left as a streptavidin negative control in all sensor chips. To collect kinetic data, a concentration series of Fab in HBS-EP was injected for 2 min at 35 µl/min over all flow cells and dissociation was observed for 3 min. The Fab samples were injected in random order, interspersed with a number of buffer blank injections for double referencing (Myszka, 2000). Flow cells were regenerated between Fab injections with a 0.5 min wash of 2.5 M NaCl at 100 µl/min. Raw data were processed with Scrubber and analyzed with CLAMP using a 1:1 bimolecular interaction model (Myszka and Morton, 1998). K_D values were calculated (k_d/k_a) from the on and off rates determined by CLAMP. Standard free energies of binding were calculated from the K_D values [$\Delta G^\circ = -RT \ln(C/K_D)$, $R = 1.987 \times 10^{-3}$ kcal/mol.K, $T = 298.15$ K and $C = 1$ mol/l].

Results

Peptide selection against an anti-polyhistidine mAb

The peptide selection experiment, originally designed to target a His₆-tagged protein immobilized by an anti-polyhistidine mAb, utilized a random, unconstrained 27-mer peptide library. During PCR and transcription the complexity of the library was maintained by having at least 7×10^{13} sequences at the start of each reaction. The initial mRNA display pool contained at least 10^{12} unique peptide sequences, estimated from the initial mRNA and methionine concentrations in the translation reaction, out of a maximum complexity of 20^{27} peptides ($\sim 1.3 \times 10^{35}$).

Five rounds of selection were performed on the immobilized anti-polyhistidine mAb, pre-saturated with an N-terminal His₆-tagged protein (Figure 2A). Bound RNA-peptide fusions were eluted with acetic acid, which generally recovered >80% of the remaining ³⁵S counts. To determine the progress of the selection, a separate ³⁵S-Met-labeled RNA-peptide fusion pool from the fifth round was purified, RNase-treated and assayed for binding (Figure 2B). This assay revealed specific binding of the peptide pool (now modified only at the C-terminus with puromycin and a short DNA linker) to the antibody rather than to the immobilization matrix (protein G-Sepharose) or to the His₆-tagged protein. The reduced binding observed when a His₆ peptide competitor was added further evinced that the selected sequences specifically targeted the antigen-binding region of the mAb. A sixth round of selection, performed with unblocked mAb as the target, demonstrated that the enrichment for active peptides against the mAb was essentially complete (Figure 2A).

DNA sequencing of the final sixth round pool revealed a variety of peptides containing two to five consecutive His residues with no other apparent consensus except a bias for Arg residues C-terminal to the His-track (Table I). The His-track was seen in various positions in the random region of the library, suggesting that specific locations in the random domain were not favored for mAb recognition. One sequence, peptide C, emerged as the dominant member of the selected library (Table I). Further rounds of selection using His₆ peptide and/or poly-L-His as competitors in the selection buffer generally resulted in changes in the percentage of peptide C in the pool rather than the emergence of new, beneficial mutations or peptides defining a single consensus (data not shown). Peptide C remained the most prevalent sequence in all subsequent

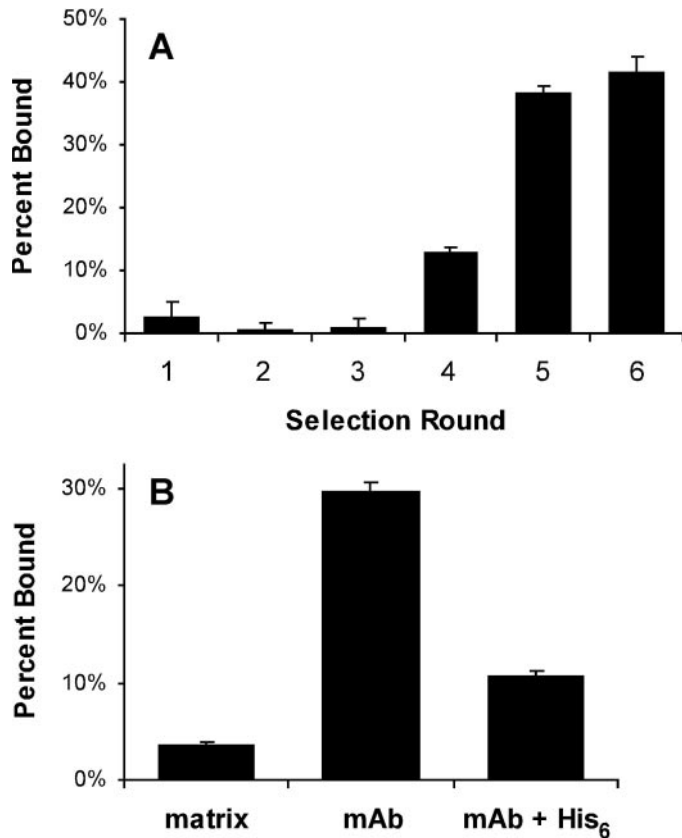


Fig. 2. Selection of peptides against the anti-polyhistidine mAb. (A) The percentage binding from each round of selection was determined by scintillation counting of an aliquot of the ³⁵S-Met-labeled RNA-peptide fusions before and after affinity selection on the immobilized antibody. (B) Binding assay of the fifth round mRNA display library. Purified, ³⁵S-labeled fusions from the fifth round pool were RNase treated (resulting in peptides covalently coupled to the puromycin-DNA linker) and assayed on the protein G-Sepharose matrix with and without immobilized anti-polyhistidine mAb. The addition of 10 mM hexahistidine peptide competitor resulted in reduced binding to the mAb, suggesting that the selected peptides interact specifically with the antigen-binding site.

selection rounds, with a collective frequency of 20 out of 53 sequences (Table I).

Selection with an mRNA display fragment library

To identify the minimal peptide sequence necessary for high-affinity binding, a nested deletion library was constructed from the peptide C-dominated library. This library is composed of fragments of DNA that encode shorter stretches of the parent peptides. Initial attempts to generate nested deletions using random priming on cDNA resulted in nearly full-length sequences (unpublished observations), possibly due to the strand-displacement abilities of the tested polymerases (I.N.Hampson, personal communication; Hamilton *et al.*, 2001). This attribute was exploited in the final fragmentation scheme (Figure 3A). DNase I was used to generate random fragments from the cDNA of a functional library. Various dilutions of DNase I were used to find the optimal conditions for producing a range of ssDNA products from ~50 to 130 bases (data not shown). Successive random priming and fill-in reactions with a modified T7 polymerase and primers containing 3'-random hexamers produced the initial DNA pool. PCR-amplified dsDNA was purified to retain fragments between

Table I. Peptide sequences from anti-polyhistidine mAb selection using a random 27-mer library^a

	YRTNHHYDVGRFAARGRRD	
	NGRSSMNWRSQEITRYTSEHHYRMAFL	
	PEQYDHHHLEARRRASSTRQVRRARR	
	RAYTPHHHAEGRLVRLPHAPYKNT	
	YVKNRLHHHRLARLVAAEH HH RLRVQ	
	NKRNLSPWSHHHQVARRTHMRAQHTM	
	RPTKNFEAEVVRSTGPMHHHTAKQRY	
	DFLTYNKSMGCRPTNFRHHHSSVVQSQ	
	DEPEVVGRVLGERPAGALADHHHMMKW	
	EVLHGHVVVARASCTGPTRRASCA	(6/53)
	HVYEKANNRLGKHHHLLAARRSKSWN	
	SNKGF SW RKKGMAVTPN RHL HHHMAH	
	TNHHHHGVLEERRQDILTGSLIEHKH	
	ILKRLREQHHHAAAHHVRRRRGRH	
	NYTT RR AEWN RQDA HHHHQEARRGAL	A (3/53)
	SKKDNAVGLQELRLREGHRRHHVDMLT	
	KKVRC HH HHHHQVALLDAAERGPGRMS	
	GIHHHHAMAVLAE LGMN PMGFALPDMW	
	AGVHHHHDAARGG TRRR STPRSATRR	*
	TMNWHHHHENGLRARMYDAGRR	
	KVRRDVMR WH HHHRMARRKANR	B (4/53)
	RVQDRLGHRVQ PVL HHHHQAARRRV	
	AALHHHHHDAGRASAMRRPGTPATSWR	
	DGH PERH DAGDHHHHHGVRQWRLISTG	C (20/53)

^aOnly the random domain is shown. Sequences contained between two and five consecutive histidines and were aligned at the C-terminal end of the His-track. A consensus was not observed except for a strong bias for Arg several residues C-terminal to the His-track. His and Arg residues are shown in bold. The frequency (out of 53) is shown for peptides that appeared more than once from DNA sequencing of individual clones. For these sequences, amino acids that differed between clones are in italics, with the most common residue at that position shown. Several sequences contained multiple deletions that shortened the random domain but left the C-terminal constant region intact and in-frame. The sequence marked with an asterisk contained a 2 bp insertion which resulted in a frameshift of the C-terminal constant region (not shown). Peptides A, B and C are named.

150 and 200 bp, corresponding to peptides ~10–30 amino acids long.

Because stop codons hinder RNA-peptide fusion formation, the 3'-constant sequence of the fragment library was chosen such that TAA, TAG and TGA codons did not exist in any frame. The 5'-constant region added a c-Myc epitope tag and provided a primer site for subsequent PCR amplification (for additional attachment of the T7 promoter and UTR sequence). This method resulted in a unidirectional fragmented pool; all transcribed RNA maintained the sense orientation. DNA sequencing of the initial pool demonstrated reasonable representation of the dominant sequence (peptide C) and confirmed that one-third of the sequences were in-frame, as expected (Figure 3B). DNA alignments with peptide C derivatives typically contained several mismatches at the beginning and end of the fragment region, most likely due to imperfect annealing of the random hexamer primers.

The nested deletion library was used for selection against the anti-polyhistidine mAb (Figure 3C). Poly-L-His and His₆ peptide were used as competitors in the second and third rounds. Although the binding of the second and third round pools was similar, more RNA-peptide fusions were retained after the stringent, competitive wash in the third round, suggesting that the washes were indeed enriching the pool for the highest affinity peptides. DNA sequencing of the final pool revealed three distinct classes of peptides (Table II). Class 1 sequences were fragments corresponding to N- and C-terminal deletions

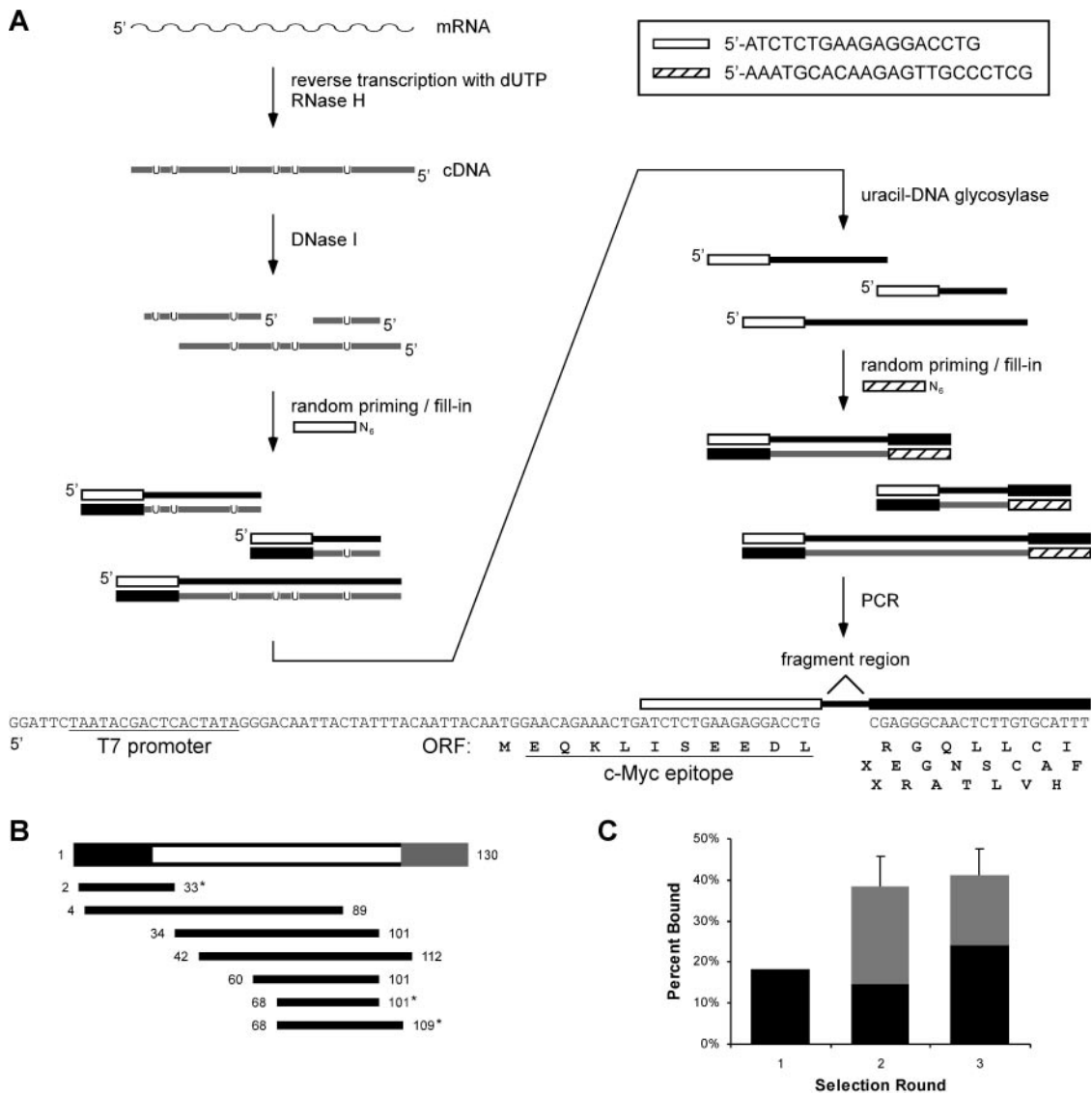


Fig. 3. Construction of a unidirectional nested deletion library. (A) Single-stranded cDNA is obtained by reverse transcription of mRNA templates with dUTP. After fragments are produced by a partial DNase I digest, a randomly-primed fill-in reaction is performed with degenerate DNA hexamers containing a constant 5' sequence, resulting in complete second-strand cDNA for each fragment. Uracil-DNA glycosylase (UDG) is used to remove first-strand cDNA and the anti-sense strand is filled-in again by random priming. The constant region of the second primer encodes a suitable peptide sequence in all three frames (lacking stop codons) and serves as the reverse primer site for subsequent PCR. PCR of the resulting dsDNA produces the initial library suitable for *in vitro* selection. (B) Representation of the peptide C parent DNA sequence in the initial fragment library. The 5'-UTR, peptide coding region and 3'-constant region are in black, white and gray, respectively. The bases spanned by each library member are shown. Sequences marked with an asterisk are in-frame with the 5'-constant region added during the generation of the library. The sequence spanning bases 4–89 is also viable assuming translation occurs at the first Met codon. (C) Selection of the peptide fragment library on anti-polyhistidine mAb. The percentage of recovered fusions (black) was determined as in Figure 2. In the second and third rounds, the competitive washes (gray) removed a portion of the initially bound counts.

of peptide C. A sequence alignment of the fragments identified RHDAGDHHHHHGVRQ (peptide Cmin) as a minimal functional sequence for peptide C.

The majority of fragments recovered after the selection came from parent sequences other than peptide C (Table II, Class 2). An alignment of peptides D and E (which collectively represented 40% of the final, third round selection pool) revealed the consensus motif ARRHA. This exact motif was not seen in the original selection, although three peptide sequences contained ARRXA [X = R, G (peptide A) or K (peptide B)] two residues C-terminal to the His-track (Table I), as in peptide D. Additional N- and C-terminal deletions for peptides D and E were not observed. Hence these sequences may already

represent minimal high-affinity binding epitopes. Alternatively, there may have been an insufficient number of clones sequenced to find other corresponding fragments. Other recovered sequences in this peptide class retained at least part of the ARRXA, suggesting that the first few residues of the consensus motif are more critical for mAb interaction.

Several additional peptides were discovered that encoded a weak consensus sequence unrelated to the mAb-binding peptides (Table II, Class 3). Binding assays with two of these peptides revealed significantly weaker affinity for the mAb than a His₆-containing peptide control (data not shown). These peptides may bind to an alternate interaction site and were consequently enriched when high stringency, competitive

Table II. Peptide sequences from fragment library selection^a

Class 1	* <u>MDGHPERHDAGDHHHHHGVRQ</u> <u>ERHDAGDHHHHHGVRQWRLIS</u> <u>RHDAGDHHHHHGVRQWRLIS</u>	
Class 2	<i>I</i> <u>TNSPGRF</u> RHHH <u>VLARR</u> H ALYR M <u>T</u> SAGW <u>T</u> AM <u>H</u> YI <u>S</u> ARR H <u>A</u> M <u>R</u> SMK <u>F</u> AQ NY <u>T</u> TQRAEW <u>N</u> RQDA <u>H</u> RHHH <u>Q</u> ARR GQ * MKVRRD <u>V</u> MR <u>W</u> HHHH <u>R</u> MARR <u>K</u> ANR <u>D</u> HHHH <u>G</u> AAR <u>P</u> V <u>F</u> RRGLYQ <u>K</u> RG <u>D</u> RRHH <u>G</u> VAR <u>V</u> REQ <u>M</u> ARYV	D (6/20) E (2/20) A1 B F
Class 3	VT <u>M</u> FD <u>V</u> DA <u>Y</u> F GLAVWSSGD <u>L</u> RAFQ VT <u>M</u> FD <u>V</u> DA <u>Y</u> F GLAVW * <u>M</u> FD <u>Y</u> DA <u>F</u> Y <u>G</u> YNGSA <u>V</u> GS <u>P</u> TL <u>Q</u> H <u>V</u> RL <u>Q</u> P * <u>M</u> N <u>F</u> D E <u>Y</u> LR <u>L</u> LLR	(2/20)

^aOnly the fragment domain of the peptides is shown. Class 1 peptides are derived from peptide C (Table I) and the putative minimal active sequence is underlined. Class 2 sequences contain portions of the ARRXA motif. Conserved residues are in bold. Sequences derived from parent peptides A and B, as well as new peptides D, E and F, are labeled. The C-terminal RGQ in the sequence derived from peptide A is encoded by part of the 3'-constant region. Class 3 peptide sequences were aligned using CLUSTALW (<http://npsa-pbil.ibcp.fr>) with key residues (bold) determined automatically. Clone frequency (out of 20) is shown and differing residues are italicized as described in Table I. Peptide sequences translated from alternate start codons are marked with an asterisk.

washes were introduced for the last rounds of selection. Site-specific, competitive washes (e.g. with poly-L-histidine) would result in the enrichment of peptides with higher affinity for the antigen-binding region, as well as of peptides with any affinity for other sites.

Immunoprecipitation of selected peptides

Selected clones were qualitatively assessed for binding by immunoprecipitation with the anti-polyhistidine mAb. ³⁵S-Met-labeled peptides were assayed directly from the *in vitro* translation reactions. The selected peptides demonstrated significantly increased binding compared with a C-terminal His₆-tagged peptide control (Figure 4). Non-specific binding was shown to be minimal with a c-Myc epitope control peptide. The fragment-selected peptides and the Myc control were immunoprecipitated with the 9E10 anti-c-Myc mAb to confirm that they were correctly translated (data not shown).

Kinetics by surface plasmon resonance

Various peptides from the fragment selection were synthesized or expressed for kinetic analysis by surface plasmon resonance (SPR). In an SPR experiment, one binding partner (the ligand) is immobilized on a sensor chip while the other reactant (the analyte) is in solution. Binding of the analyte to the ligand is observed as a refractive index change on the sensor chip surface and is measured in real time in resonance units (RU). Biotinylated ligands were immobilized on streptavidin sensor chips for the SPR analyses.

Absolute and relative binding constants obtained from mAb interactions with immobilized antigens are unreliable owing to the effects of rebinding and bivalency (Nieba *et al.*, 1996). To avoid these problems, Fab was prepared from anti-polyhistidine mAb and used as the analyte. Using the peptides as the immobilized ligands and Fab as the analyte ensured fair comparisons between the kinetics measurements, avoiding bias in protein quantitation since all Fab concentrations were prepared from a single stock solution. Kinetic parameters were

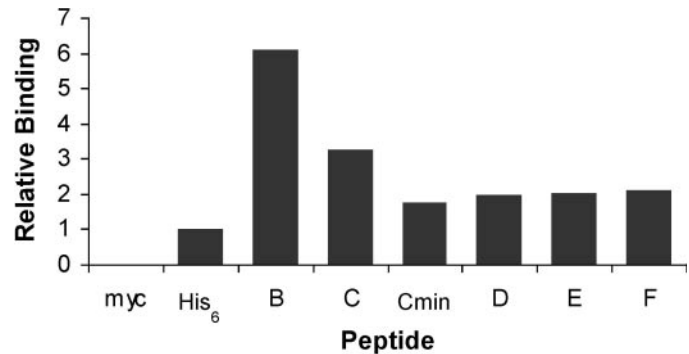


Fig. 4. Binding of *in vitro* translated peptides to anti-polyhistidine mAb. Free, ³⁵S-labeled peptides were assayed for binding directly from the translation reactions. Relative binding is shown as a fold-change versus the His₆ sequence. Peptide sequences are given in Tables I and II. Myc is a peptide encoded by the constant regions of the fragment library primers with only an arginine residue in between. The His₆ control sequence encoded a 31-mer peptide with a C-terminal His₆ tag.

determined using a simple 1:1 bimolecular interaction model (Table III).

The assayed peptides could be categorized by their dissociation rates from the Fab (Figure 5). The cited epitope, His₆, bound weakest to the Fab; the His₆ peptide and the His₆-tagged protein used in the original selection exhibited K_D values of 0.6 and 3 μ M, respectively. Additional His residues (His₁₀ peptide) increased the association rate 6-fold without changing the dissociation rate significantly. Peptides from the selection demonstrated K_D values of <75 nM, ~10- to 75-fold better than the control His₆ sequence, with increased affinities as a result of faster association (up to 5-fold) and considerably slower (6- to 21-fold) dissociation rates (Table III). Class 2 peptides with the ARRXA motif demonstrated the highest affinities, with ~3-fold slower dissociation rates compared with sequences derived from peptide C (Figure 5C). While the flanking residues on peptide Cmin contribute at least 1.6 kcal/mol to the binding free energy compared with the His₆ peptide, sequences with the ARRXA motif demonstrate 2.6 (peptide B) and 2.2 (peptide D) kcal/mol improvements. The contributions from these flanking residues is likely even greater, as these calculations do not account for any loss of binding free energy from having shorter (<6) stretches of His residues in the core site.

Discussion

During an *in vitro* selection experiment against a His₆-tagged target protein immobilized using an anti-polyhistidine antibody, mAb-binding peptides were inadvertently enriched. The His₆-tagged fusion protein was of high quality and purity (see Materials and methods) and was previously used in functional assays (Ja and Roberts, 2004). Additionally, the presence of the target protein was confirmed in the elutions from each selection round (data not shown). Hence the enrichment of peptides that bind the mAb was most likely due to the existence of short sequences that confer significantly higher affinity than a hexahistidine tag. A preclearing step that included the mAb may not have been totally effective in preventing the selection of antibody-specific peptides, as even the final selection round resulted in an incomplete, ~40% pull-down of the RNA-peptide fusions.

Table III. Kinetic parameters for peptide interactions with Fab determined by surface plasmon resonance^a

	Peptide sequence	k_a ($M^{-1} s^{-1}$) ($\times 10^4$)	k_d (s^{-1}) ($\times 10^{-2}$)	K_D (nM)	χ^2	ΔG° (kcal/mol)
	HHHHHH-protein	7.3	23.82	3260	0.80	-7.5
	HHHHHH	9.9	5.78	580	0.72	-8.5
	HHHHHHHHHH	62.4	6.56	105	1.19	-9.5
	MDGHPERHDAGDHHHHHGVRQ	11.8	0.85	72	1.28	-9.7
Cmin	RHDAGDHHHHHGVRQ	21.5	0.82	38	1.19	-10.1
C	MDGHPERHDAGDHHHHHGVRQWRLISTG-MBP	52.5	0.97	18.5	1.48	-10.5
D	NSPGRFRHHHVLAARRHALYR	17.4	0.27	15.5	0.77	-10.7
B	MKVRDVMRWHHHHRMARRKANR-MBP	40.9	0.31	7.6	1.28	-11.1

^aSPR experiments monitored binding between immobilized peptides and purified Fab. On and off rates were determined by global fit analysis on CLAMP using a 1:1 bimolecular interaction model (Myszka and Morton, 1998). K_D values were calculated from k_d/k_a . Synthetic peptides include a short, C-terminal biotin-containing sequence (not shown). Full-length peptides B and C were assayed as MBP fusion proteins.

Although the cited mAb epitope is hexahistidine, the recovered peptides surprisingly each contained a shorter (≤ 5) stretch of consecutive His residues and a bias for Arg. Because His is clearly a key determinant in mAb recognition, we consider the selected peptides to be 'epitope-like' rather than mimotopes [linear peptides that mimic the binding mode of an epitope, a term previously reserved for distinct, alternative ligand structures (Stephen *et al.*, 1995)]. The original immunogen used to develop the mAb was an N-terminal His₆-tagged fusion protein of unknown sequence and identity (proprietary information, Sigma-Aldrich). Hence whether the regions flanking the His-tracks in the selected peptides have homology to the original immunogen cannot be determined. However, given the antigen sequence, the epitope (centering on the His₆ tag) clearly would have been recognized.

To characterize the mAb epitope better and demonstrate the feasibility of gene-fragment mRNA display, a nested deletion library was constructed from the final selection pool. A previously described protocol, directional random oligonucleotide primed (DROP) synthesis of cDNA (Hampson *et al.*, 1996), was modified to maintain as many viable library fragments as possible. Owing to the difficulty in obtaining a broad size distribution of sequences with degenerate oligos, DNase I was used for the random fragmentation of cDNA. DROP synthesis using a highly processive DNA polymerase, capable of potent strand displacement, yielded intact copies of the cDNA fragments while maintaining the sense strand (Figure 3A).

In vitro selection with the fragment library resulted in the identification of a 15-mer functional sequence (peptide Cmin: RHDAGDHHHHHGVRQ; $K_D = 38$ nM) derived from the full-length 35-mer, peptide C. Because the initial fragment library was produced from a pool dominated by peptide C, we expected to recover and identify numerous overlapping peptides that defined a minimal epitope for this sequence. Surprisingly, the majority of recovered sequences came from unknown parents. The enrichment of these peptides implies that these fragments were more highly favored after truncation. The flanking regions of the original peptides may have hindered access to the epitope by the mAb, suggesting that peptide length may be an important attribute in the fine-tuning of affinity and/or function. Alternatively, these particular sequences may have been negatively biased by the constant C-terminal peptide used in the original random peptide library. The three-frame constant sequence used in the fragment library construction increases the sensitivity of the selection when one of the

translation frames causes negative bias. Additionally, a random distribution between the three translation frames in the selected peptides would indicate that the 3'-constant region does not affect functional selectability or bias RNA-peptide fusion formation. The six independent clones of peptide D, for example, had all three frames represented in the 3'-constant region (Table II and data not shown).

Based on the selected peptide sequences, two major protein interaction motifs were identified: a core epitope consisting of at least three consecutive His residues and a second interaction site encoded by the consensus motif, ARRXA. SPR experiments demonstrated a significant increase in the association rate of His₁₀ compared with His₆, suggesting that additional His residues present a more accessible core interaction rather than slow dissociation by enhancing rebinding from multivalency effects. Only additional contacts, made by the addition of interacting residues such as the ARRXA motif, result in significantly slower dissociation rates. These flanking residues can contribute significantly to the binding free energy, at least 2.6 kcal/mol in the case of peptide B in comparison with His₆, which assumes that the loss of two out of six histidines in the core has no effect. The two interaction cassettes we have identified here may be juxtaposed sites from the original fusion protein antigen.

Our results also highlight the importance of flanking residues outside of the two consensus motifs and their contribution to binding affinity with antibodies. Residues adjoining core amino acids in an epitope can substantially influence antibody binding, the effects of which can only be assessed through quantitative affinity measurements (Choulier *et al.*, 2001; Coley *et al.*, 2001). This is demonstrated in our experiments, where the rank order of binding in the immunoprecipitation assay did not entirely correspond with quantitative kinetic measurements. Epitope tags are often appended to proteins and used as molecular handles for detection, isolation and analysis of protein-protein interactions. Their functionality in this context, however, is highly variable. Tandem repeats of tags (e.g. the popular c-Myc or FLAG epitopes) have been used to ensure robust affinity and recognition by antisera (Nakajima and Yaoita, 1997; Hernan *et al.*, 2000). By identifying longer functional peptides with appropriate flanking residues, high affinity can be maintained with less variability depending on the linker region and the protein to which the epitope is attached.

The fragment library selection resulted in a disproportionate number of peptides that did not contain an N-terminal deletion.

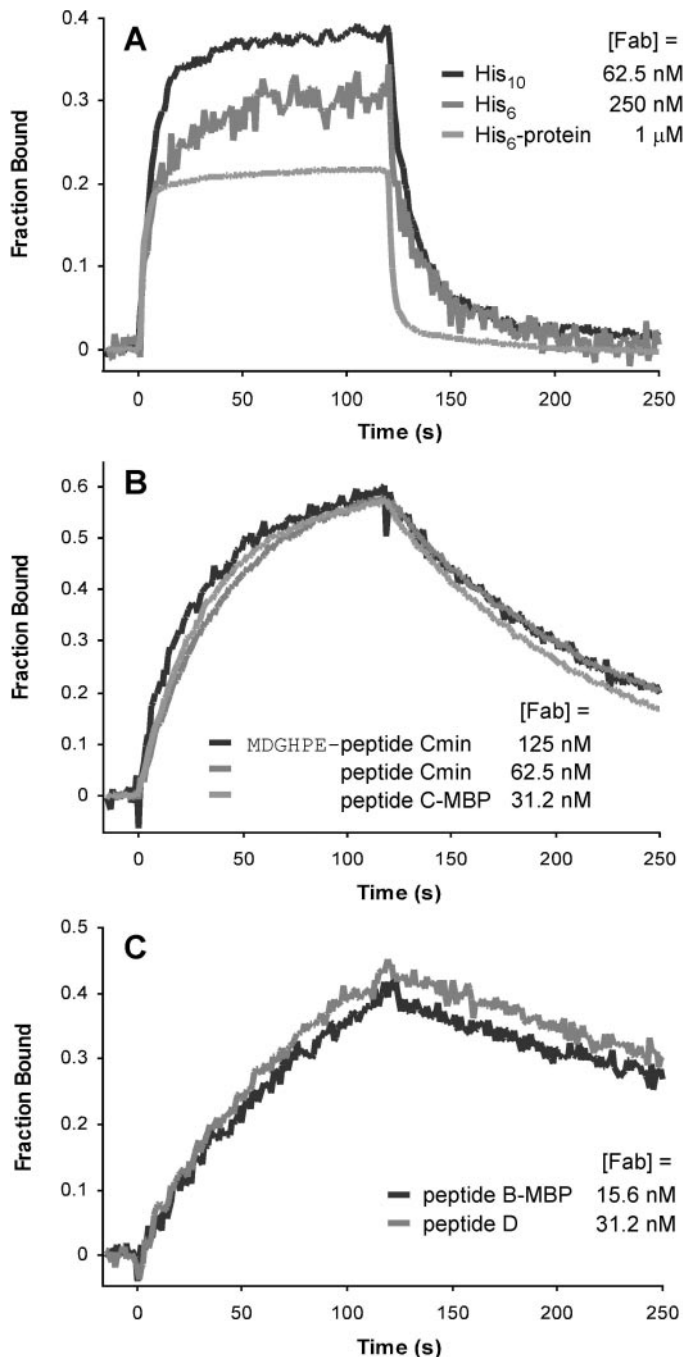


Fig. 5. Representative sensorgrams from SPR experiments. Purified anti-polyhistidine Fab at concentrations corresponding to $\sim 0.5 K_D$ was injected over immobilized peptides or peptide-MBP fusions. Peptides fell into three categories describing (A) weak (His_6 , His_{10} and His_6 -tagged blocking protein), (B) intermediate (peptide C-derived sequences) and (C) strong (sequences containing the ARRXA motif) binding for the Fab. For comparison, sensorgrams were divided by the computed maximum signal. The Fab concentration for each injection is indicated.

Because of the 5'-UTR on the mRNA used to make the fragment library, more fragments containing the first start codon (with varying lengths of UTR sequence) were probably present in the initial fragment pool. 5'-UTR and/or promoter sequences most likely do not hinder the fragment selection process, as ribosome scanning can initiate translation at the correct start codon, regardless of which frame was amplified. This was seen in several of the selected fragment sequences (Table II). This

property increases the number of viable (i.e. translatable) templates, but introduces some bias favoring intact N-terminal sequences.

Although not utilized in this experiment, the c-Myc tag introduced in the fragmentation library can be used to generate and purify a fragment library enriched with in-frame sequences. Although the tag is at the N-terminus of the library, in general RNA-peptide fusions will form only when the ribosome can translate most of the sequence and reach the end of the mRNA (unpublished results). Hence only sequences that lack stop codons (and therefore are most likely in-frame) will form fusions and be purified and amplified after a Myc-epitope pre-selection. Another improvement to the protocol includes using Exonuclease I to remove excess degenerate primers during DROP synthesis, preventing the amplification of sequences without 'inserts,' as DNA size fractionation by agarose gel is not completely effective in removing these smaller fragments (data not shown).

The ability to access high-complexity libraries is a great advantage for mRNA display over other selection systems. Library construction methods that involve PCR and DNA reassembly are better suited for the mRNA display format, thereby avoiding cloning steps that are required in techniques such as phage display. A comparative study on epitope mapping using random 6-mer and 15-mer peptide phage display libraries successfully identified consensus motifs for only two of the four mAbs examined (Fack *et al.*, 1997). For one of the mapped mAbs, the random peptide selection succeeded only with the 6-mer library, identifying a short consensus motif that was not discovered with the 15-mer library, which the authors attributed to a statistical lack of representation. In contrast, an mRNA display selection with a random 27-mer library identified epitope-like consensus motifs for the anti-c-Myc antibody, 9E10 (Baggio *et al.*, 2002). The complete, 10-mer wild-type epitope was also selected from the library. The selection revealed the core determinants and some of the allowed flanking residues for mAb interaction. By using high-complexity, long peptide libraries, mRNA display selections can identify extremely rare sequences such as discontinuous or conformational epitopes, as well as novel mimotopes. The full-length consensus peptide, $H_m-X_2-ARRXA$, found here, for example, may not have been identified with more traditional X_6 or X_{10} phage display libraries.

While selection with biological libraries is an inexpensive and technically straightforward approach for epitope mapping, synthetic combinatorial libraries (SCLs) offer an alternative method for generating peptide ligands (Houghten *et al.*, 1992; Pinilla *et al.*, 1994a). Screens of mixture-based SCLs have resulted in the delineation of linear and discontinuous epitopes *a priori* (Geysen *et al.*, 1986; Pinilla *et al.*, 1993; 1994b). In these experiments, peptide SCLs are composed of mixtures containing positions defined with a specific amino acid while the remaining positions contain mixtures of residues. Various libraries are screened to determine the most active amino acids at each position of a peptide sequence. The optimal residues can be determined in parallel ('positional scanning' approach; Pinilla *et al.*, 1994a) or iteratively, where each subsequent library is synthesized to expand on the most active libraries determined in the previous screen (Houghten *et al.*, 1992). The construction of SCLs can be very specifically controlled and library members can be screened free in solution, avoiding any bias from a fusion partner.

While selection methods employ an amplification step for the enrichment of functional library members, individual peptides in an SCL must exist at sufficient concentrations for producing a signal (e.g. through ELISA screens). As the library size increases, the concentration of individual peptides decreases significantly. Additionally, these larger libraries exponentially increase the deconvolution step (e.g. with positional scanning SCLs) after a screen because individual peptides made up of combinations of the optimal amino acids at each position in a library need to be synthesized and analyzed. The interdependence of individually determined optimal amino acids at each position may cause significant deviations from the optimum full-length peptide sequence. This contrasts greatly from peptide selections where complete peptide information is generated by DNA sequencing and each library member is presumably active for binding and already optimized in the context of other positions. These problems may not be significant in epitope mapping experiments, however, as most antibodies recognize linear epitopes composed of less than six residues. While SCLs seem appropriate for generating short peptide ligands, biological libraries can easily search hexamer libraries, with techniques such as mRNA and ribosome display permitting exhaustive searches of decapeptide libraries.

Owing to the higher efficiency of synthesizing the nested deletion library completely *in vitro*, the fragment library construction described here maintains a higher number of unique sequences, in contrast to DNA libraries produced by enzymatic ligation and cloning, which are limited by *in vivo* transformation efficiencies. Additionally, the library construction method we used is unidirectional for all amplified sequences so that the sense orientation is maintained and only the minimal two-thirds of the fragments are non-viable due to frameshifts. This protocol produces a well-distributed library and is technically less challenging as the random oligonucleotide priming is used only to 'copy' the cDNA fragments produced by DNase digestion and need not be optimized for generating a fragment distribution. mRNA display with fragment libraries combines the ease and versatility of working with cDNA *in vitro* with the benefits of expression cloning. The method permits the minimization of functional domains, as well as the isolation of optimal binding contexts through the removal of negative-acting flanking regions. Although the technique may not be sufficiently processive for the fine mapping of short peptide sequences, it should be highly applicable for constructing cDNA or tissue-specific expression libraries and the subsequent determination of minimal binding domains and novel protein-protein interactions.

Acknowledgements

We thank Dr David S.Waugh (National Cancer Institute at Frederick) for the pDW363 *in vivo* biotinylation vector, Professor Pamela J.Bjorkman and Anthony M.Giannetti for time and support on the Biacore 2000, William Hunter (Biacore, Piscataway, NJ) for technical advice on SPR, Cindy I.Chen and Christopher T.Balmaseda for preparative and technical assistance with library construction and protein purification and Professor David G.Myszka (University of Utah) for generously providing the kinetic analysis software, Scrubber and CLAMP. We greatly appreciate Dr Yuri Peterson (Duke University) for suggestions on the paper. We are indebted to Dr Ian N.Hampson (St Mary's Hospital, Manchester, UK) for valuable discussions and technical expertise on random priming and the synthesis of the fragment library. This work was supported by grants from the NIH (RO1 GM 60416) and the Beckman Foundation to R.W.R. W.W.J. was supported in part by a DOD National Defense and Engineering Graduate Fellowship.

References

- Baggio,R. *et al.* (2002) *J. Mol. Recog.*, **15**, 126–134.
- Boder,E.T. and Wittrup,K.D. (1997) *Nat. Biotechnol.*, **15**, 553–557.
- Boder,E.T. and Wittrup,K.D. (1998) *Biotechnol. Prog.*, **14**, 55–62.
- Choulier,L., Laune,D., Orfanoudakis,G., Wlad,H., Janson,J.-C., Granier,C. and Altschuh,D. (2001) *J. Immunol. Methods*, **249**, 253–264.
- Christmann,A., Wentzel,A., Meyer,C., Meyers,G. and Kolmar,H. (2001) *J. Immunol. Methods*, **257**, 163–173.
- Coley,A.M., Campanale,N.V., Casey,J.L., Hodder,A.N., Crewther,P.E., Anders,R.F., Tilley,L.M. and Foley,M. (2001) *Protein Eng.*, **14**, 691–698.
- Ellington,A. and Pollard,J.D.,Jr (2001) In Ausubel,F.M., Brent,R., Kingston,R.E., Moore,D.D., Seidman,J.G., Smith,J.A. and Struhl,K. (eds) *Current Protocols in Molecular Biology*. Wiley, Hoboken, NJ, Unit 2.12.
- Fack,F., Hügle-Dörr,B., Song,D., Queitsch,I., Petersen,G. and Bautz,E.K.F. (1997) *J. Immunol. Methods*, **206**, 43–52.
- Frank,R. (1992) *Tetrahedron*, **48**, 9217–9232.
- Frank,R. and Overwin,H. (1996) *Methods Mol. Biol.*, **66**, 149–169.
- Frohlich,M.W. and Parker,D.S. (2001) *Biotechniques*, **30**, 264–266.
- Georgiou,G., Stathopoulos,C., Daugherty,P.S., Nayak,A.R., Iverson,B.L. and Curtiss,R.,III (1997) *Nat. Biotechnol.*, **15**, 29–34.
- Geysen,H.M., Meloen,R.H. and Barteling,S.J. (1984) *Proc. Natl Acad. Sci. USA*, **81**, 3998–4002.
- Geysen,H.M., Rodda,S.J. and Mason,T.J. (1986) *Mol. Immunol.*, **23**, 709–715.
- Gupta,S., Arora,K., Sampath,A., Khurana,S., Singh,S.S., Gupta,A. and Chaudhary,V.K. (1999) *Biotechniques*, **27**, 328–330, 332–334.
- Hamilton,S.C., Farchaus,J.W. and Davis,M.C. (2001) *Biotechniques*, **31**, 370–376, 378–380, 382–383.
- Hammond,P.W., Alpin,J., Rise,C.E., Wright,M. and Kreider,B.L. (2001) *J. Biol. Chem.*, **276**, 20898–20906.
- Hampson,I.N., Hampson,L. and Dexter,T.M. (1996) *Nucleic Acids Res.*, **24**, 4832–4835.
- Hanes,J. and Plückthun,A. (1997) *Proc. Natl Acad. Sci. USA*, **94**, 4937–4942.
- He,M. and Taussig,M.J. (1997) *Nucleic Acids Res.*, **25**, 5132–5134.
- Henikoff,S. (1984) *Gene*, **28**, 351–359.
- Herman,R., Heuermann,K. and Brizzard,B. (2000) *Biotechniques*, **28**, 789–793.
- Houghten,R.A., Appel,J.R., Blondelle,S.E., Cuervo,J.H., Dooley,C.T. and Pinilla,C. (1992) *Biotechniques*, **13**, 412–421.
- Irving,M.B., Pan,O. and Scott,J.K. (2001) *Curr. Opin. Chem. Biol.*, **5**, 314–324.
- Ja,W.W. and Roberts,R.W. (2004) *Biochemistry*, **43**, 9265–9275.
- Keefe,A.D. (2001) In Ausubel,F.M., Brent,R., Kingston,R.E., Moore,D.D., Seidman,J.G., Smith,J.A. and Struhl,K. (eds) *Current Protocols in Molecular Biology*. Wiley, Hoboken, NJ, Unit 24.5.
- Kramer,A., Reineke,U., Dong,L., Hoffmann,B., Hoffmüller,U., Winkler,D., Volkmer-Engert,R. and Schneider-Mergener,J. (1999) *J. Pept. Res.*, **54**, 319–327.
- Kurz,M., Gu,K. and Lohse,P.A. (2000) *Nucleic Acids Res.*, **28**, e83.
- Kuwabara,I., Maruyama,H., Kamisue,S., Shima,M., Yoshioka,A. and Maruyama,I.N. (1999) *J. Immunol. Methods*, **224**, 89–99.
- Lee,E., Linder,M.E. and Gilman,A.G. (1994) *Methods Enzymol.*, **237**, 146–164.
- Lenstra,J.A., Kusters,J.G. and van der Zeijst,B.A. (1990) *Arch. Virol.*, **110**, 1–24.
- Liu,R., Barrick,J.E., Szostak,J.W. and Roberts,R.W. (2000) *Methods Enzymol.*, **318**, 268–293.
- Mattheakis,L.C., Bhatt,R.R. and Dower,W.J. (1994) *Proc. Natl Acad. Sci. USA*, **91**, 9022–9026.
- McPherson,M., Yang,Y., Hammond,P.W. and Kreider,B.L. (2002) *Chem. Biol.*, **9**, 691–698.
- Miceli,R.M., DeGraaf,M.E. and Fischer,H.D. (1994) *J. Immunol. Methods*, **167**, 279–287.
- Milavetz,B. (1992) *Nucleic Acids Res.*, **20**, 3529–3530.
- Mullaney,B.P., Pallavicini,M.G. and Marks,J.D. (2001) *Infect. Immun.*, **69**, 6511–6514.
- Murthy,K.K., Shen,S.-H. and Banville,D. (1998) *Biochem. Biophys. Res. Commun.*, **248**, 69–74.
- Myszka,D.G. (2000) *Methods Enzymol.*, **323**, 325–340.
- Myszka,D.G. and Morton,T.A. (1998) *Trends Biochem. Sci.*, **23**, 149–150.
- Nakajima,K. and Yaoita,Y. (1997) *Nucleic Acids Res.*, **25**, 2231–2232.
- Nieba,L., Krebber,A. and Plückthun,A. (1996) *Anal. Biochem.*, **234**, 155–165.
- Parhami-Seren,B., Keel,T. and Reed,G.L. (1997) *J. Mol. Biol.*, **271**, 333–341.
- Pinilla,C., Appel,J.R. and Houghten,R.A. (1993) *Gene*, **128**, 71–76.
- Pinilla,C., Appel,J.R., Blondelle,S.E., Dooley,C.T., Eichler,J., Ostresh,J.M. and Houghten,R.A. (1994a) *Drug Dev. Res.*, **33**, 133–145.
- Pinilla,C., Appel,J.R. and Houghten,R.A. (1994b) *Biochem. J.*, **301**, 847–853.
- Pues,H., Holz,B. and Weinhold,E. (1997) *Nucleic Acids Res.*, **25**, 1303–1304.
- Reineke,U., Kramer,A. and Schneider-Mergener,J. (1999) *Curr. Top. Microbiol. Immunol.*, **243**, 23–36.

- Roberts,R.W. and Szostak,J.W. (1997) *Proc. Natl Acad. Sci. USA*, **94**, 12297–12302.
- Santi,E., Capone,S., Mennuni,C., Lahm,A., Tramontano,A., Luzzago,A. and Nicosia,A. (2000) *J. Mol. Biol.*, **296**, 497–508.
- Scopes,R.K. (1974) *Anal. Biochem.*, **59**, 277–282.
- Scott,J.K. and Smith,G.P. (1990) *Science*, **249**, 386–390.
- Stephen,C.W., Helminen,P. and Lane,D.P. (1995) *J. Mol. Biol.*, **248**, 58–78.
- Takahashi,T.T., Austin,R.J. and Roberts,R.W. (2003) *Trends Biochem. Sci.*, **28**, 159–165.
- Telenius,H., Carter,N.P., Bebb,C.E., Nordenskjöld,M., Ponder,B.A.J. and Tunnacliffe,A. (1992) *Genomics*, **13**, 718–725.
- Tsao,K.-L., DeBarbieri,B., Michel,H. and Waugh,D.S. (1996) *Gene*, **169**, 59–64.
- Whitcomb,J.M., Rashtchian,A. and Hughes,S.H. (1993) *Nucleic Acids Res.*, **21**, 4143–4146.
- Zhang,J. and Byrne,C.D. (1999) *Biochem. J.*, **337**, 231–241.

**Received November 23, 2004; revised April 22, 2005;
accepted May 18, 2005**

Edited by Dario Neri