Identification of epitope-like consensus motifs using mRNA display

The mRNA display approach to in vitro protein selection is based upon the puromycin-mediated formation of a covalent bond between an mRNA and its gene product. This technique can be used to identify peptide sequences involved in macromolecular recognition, including those identical or homologous to natural ligand epitopes. To demonstrate this approach, we determined the peptide sequences recognized by the trypsin active site, and by the anti-c-Myc antibody, 9E10. Here we describe the use of two peptide libraries of different diversities, one a constrained library based on the trypsin inhibitor EETI-II, where only the six residues in the first loop were randomized (6.4 \times 10^7 possible sequences, 6.0 \times 10^{11} sequences in the library), the other a linear-peptide library with 27 randomized amino acids (1.3 \times 10^{35} possible sequences, 2 \times 10^{13} sequences in the library). The constrained library was screened against the natural target of wild-type EETI, bovine trypsin, and the linear library was screened against the anti-c-myc antibody, 9E10. The analysis of selected sequences revealed minimal consensus sequences of PR(I,L,V)L for the first loop of EETI-II and LISE for the 9E10 epitope. The wild-type sequences, PRILMR for the first loop of EETI, bovine trypsin, and the anti-c-Myc antibody, 9E10, were selected with the highest frequency, and in each case the complete wild-type epitope was selected from the library. Copyright © 2002 John Wiley & Sons, Ltd.

**Keywords:** molecular recognition; mRNA; protein; peptide, display; selection; 9E10; EETI-II; Myc; trypsin

INTRODUCTION

Specific interactions between proteins, such as between an enzyme and its inhibitor or between an antibody and its antigen, are mediated through molecular recognition of subsites on both binding partners. To determine such binding subsites, one strategy is to use randomized peptide or protein libraries and a screening method to identify the minimal binding domain for an interaction with a given target. Armed with this information, it should in principle be possible to search databases of protein sequences to identify previously unknown protein:target binding pairs. A method for determining the sequences of natural protein-binding peptides would be a valuable tool in proteomic research. Such a method would allow, for example, the identification of the unknown binding partners of signaling proteins, and the identification of the targets of antibodies with defined biological properties but unknown antigens.

One example of a well-characterized complex between a protein and a peptide is the complex between trypsin and its inhibitor, the Ecballium elaterium trypsin inhibitor two (EETI-II; Favel et al., 1989a,b). Trypsin is a 24 kDa pancreatic serine protease and EETI-II is a 28-residue trypsin inhibitor of the knottin family; like most knottins, it contains a triple-stranded beta sheet and three disulfide bonds (Heitz et al., 1989). Only the first loop of EETI-II (residues 3–8) is involved in the interaction with trypsin (Le-Nguyen et al., 1989, 1990).

Another example of a well-characterized complex between a protein and a peptide is that of the c-Myc antigen and the anti-c-Myc antibody 9E10, which recognizes the c-Myc epitope between residues 408 and 439 (Evan et al., 1985; Schiweck et al., 1997). Human c-Myc is a transcription factor of 65 kDa which, as a heterodimer with the Max protein, binds to genes whose promoters contain the E box motif (CACGTG, Amati et al., 2001).

A limitation of previous methods (Deroos and Muller, 2001) used to achieve the goal of finding epitope-like consensus motifs is the overall size of the randomized library that allows for adequate interaction space to be sampled. In affinity-based peptide library selections (Matheakis et al., 1994; Schatz et al., 1996), selection pressure is applied on a random population, resulting in the enrichment of specific binders. The association of phenotype with genotype is essential both to enrich binders at each round of selection and to identify the selected molecules. Among the technologies that take advantage of this link are...
in vitro technologies such as phage display (Hoogenboom et al., 1998), E. coli display (Christmann et al., 1999; Wentzel et al., 1999), yeast display (Kieke et al., 1997; Boder and Wittrup, 2000) and in vitro technologies such as ribosome display (He and Taussig, 1997), display technologies using DNA-binding proteins (Fitzgerald, 2000; Speight et al., 2001), and mRNA display (Roberts and Szostak, 1997; Nemoto et al., 1997). mRNA display uses puromycin to covalently attach an mRNA to its translated product (Fig. 1). The advantages of this method are that the libraries are generated in vitro, which allows large libraries (≥10^12 different sequences) to be made, and that the covalent link between the genotype and the phenotype ensures that the mRNA–peptide fusion remains intact under a range of conditions. mRNA display has been used to select high-affinity streptavidin-binding peptides (Wilson et al., 2001) and functional proteins (Keefe and Szostak, 1997) from a randomized peptide library.

In this study, we applied the mRNA display technology to test its ability to identify the binding domains for two target:ligand interactions, bovine trypsin inhibitor and the anti-myc 9E10 monoclonal antibody. We selected the peptides that bound to these targets by iteratively selecting and enriching for binders from randomized peptide libraries. In addition to selecting epitope-like motifs from randomized peptide libraries, our study sought to dissect the peptide–macromolecular interactions and to determine the minimal sequence required for the recognition of a peptide by a macromolecule.

**EXPERIMENTAL**

**Library construction**

**EETI-II.** Three gel-purified oligonucleotides (Oligos Etc. Inc.) were used to construct EETI-II with a randomized first loop (residues 3–8). In a final volume of 100 μl, 1 nmol of the template, 5′ GAT AAA TGT TAA TGT TAC CGG (NNS)₆ ACG TTC TGT TCG CTG ACG GAG CGG CGC ACG CAG ACG CGG GGG TTG CCG AAG ACG CGG 3′, N = A, G, C, T; S = C, G), was amplified for three rounds of PCR (94°C, 1 min; 65°C, 1 min; 72°C, 2 min) using 1 μM of the 5′ primer, 5′ TAA TAC GAC TCA CTA TAG GGA CAA TTA CTA TTT ACA ATT ACA ATG GGC TGC 3′ (Oligos Etc. Inc.), and the 3′ primer, 5′ CGG CGT CTT CGG CAA CCC CGG CTG 3′ (Oligos Etc. Inc.) in the supplied buffer (Promega) with MgCl₂ (2.5 mM), NTP (250 μM) and 5 μl of Taq polymerase (Promega, 5 units/μl).

One nanomole of DNA was transcribed into RNA in a 0.5 ml reaction using the Megascript in vitro Transcription kit (Ambion). After incubation for 1 h at 37°C, 20 units of DNase I (Ambion) were added and the incubation was continued for an additional 15 min. The reaction was twice purified with phenol/CHCl₃/isoamyl alcohol and excess NTPs were removed by NAP-25 gel filtration (Pharmacia). The RNA was precipitated by addition of 1/10 vol NaOAc pH 5.2 and 1 vol isopropanol and dissolved in 200 μl H₂O.

The puromycin-DNA oligonucleotide linker, 5′-A₂₇CC-Puromycin, was synthesized as described previously and conjugated to the 3′-end of the library RNA by template-directed enzymatic ligation (Roberts and Szostak, 1997). In a 100 μl reaction volume, the purified RNA (1 nmol) was ligated to 5′-A₂₇CC-Puromycin linker (2 nmol) in the presence of biotinylated DNA splint (2 nmol), 5′ biotin-G AAA CGA CCA ACT TTT TTT TTN GCC GCA GAA G 3′ (Oligos Etc. Inc.), with 10 μl of T4 DNA ligase (20 units/μl, Promega) for 4 h at room temperature. Ligated mRNA products that annealed to the biotinylated splint were captured on 250 μl of neutravidin-agarose beads (Pierce) suspended in 1 ml of PBS for 1 h at 35°C. The beads were washed three times with 1 ml of PBS to remove unbound species, and mRNA-A₂₇CC-Puromycin linker conjugates were eluted by incubating the beads for one hour at 35°C with an excess (10 nmol) of antislip, 5′ CTT CTG CGG CAA AAA AAA AAG TGG TTC TTG GCC GTA AAG G 3′ (Oligos Etc. Inc.). The mRNA-A₂₇CC-Puromycin linker conjugates were concentrated by isopropanol precipitation. The purified ligated mRNA was then used to generate mRNA–peptide fusion.

Purified ligated RNA (~1 nmol) was translated using 1 ml of lysate from the Rabbit Reticulocyte IVT kit.
A27CC-puromycin, and splint (5 strands used) Superscript II Reverse Transcriptase (200 fusion molecules containing the corresponding cDNA strands were used for the reverse transcription reaction, necessary to generate the low diversity, constrained peptide library construction. MgCl₂ were added to a final concentration of 530 mM and 150 mM, respectively. mRNA–protein fusion molecules were isolated by incubation with ~250 mg oligo dT₂₅ cellulose (Pharmacia) in 10 ml of incubation buffer (100 mM Tris–HCl pH 8.0, 10 mM EDTA, 1 M NaCl and 0.25% Triton X-100) for 1 h at 4 °C. Bound mRNA–protein fusion molecules were isolated by filtration, washed with incubation buffer and eluted with ddH₂O. The concentration of mRNA–protein fusion molecules was determined by scintillation counting. To generate fusion molecules containing the corresponding cDNA strands, isolated mRNA–protein fusion molecules were reverse transcribed with 10 μl of Superscript II Reverse Transcriptase (200 units/μl, Gibco BRL) in the manufacturer’s buffer at 42 °C using a 5-fold excess of ligation splint as primer. In order to promote disulfide pairing of the EETI-II fusion library, DTT was omitted for the reverse transcription reaction.

c-Myc. Three gel-purified oligonucleotides (Oligos Etc.) were used to construct the linear library. One naromole of the template [5’-AGC TTT TGG TGC TGG ATC (SNN)₂₇ CTC CTC GCC CTT GCT CAT-3’], N = A, G, C, T; S = C, G] was amplified by three rounds of PCR (94 °C, 1 min; 65 °C, 1 min; 72 °C, 2 min) using 1 μM of the S’ primer (5’-AGC TTT TGG TGC TGG ATC-3’) and the 3’ primer (5’-TAA TAC GAC TCA TTA TAG GGA CAA TTA CTA TTT ACA ATT ACA ATG GTG AGC-3’) in the supplied buffer (Promega) with MgCl₂ (2.5 mM), NTP (250 μM) and Taq polymerase (Promega, 5 units) in a final volume of 5 ml. The reaction mixture was extracted twice with phenol/CHCl₃/isoamylalcohol, precipitated with ethanol, and dissolved in 100 μl TE. The final construct contains a randomized region where each codon has the sequence NNS, where ‘N’ represents equal probability of G and C, an upstream T7 transcription promoter and a variant of the TMV translation initiation sequence. Downstream from the randomized region is a sequence designed to accommodate primers for PCR, reverse transcription and ligation. Ligation followed the protocols described in the EETI-II section.

For ligation, equal molar amounts (~2 mmol) of RNA, 5’-A₂₂₃C-puromycin, and splint (5’-TTT TTT TTT TTA GNT TTT GGT GCT TG) were mixed in a 1.5 ml reaction with T4 DNA ligase (1200 units, Promega) in buffer supplied by the manufacturer. Following a 4 h incubation at room temperature, the reaction was extracted twice with phenol/ CHCl₃ isoaoylalcohol and the nucleic acids were precipitated with isopropanol as described above.

Ligated RNA was separated from unligated RNA on a 6% TBE/urea denaturing polyacrylamide gel, eluted from the gel with 200 μl H₂O. Ligated RNA (1.25 mmol) was translated in a total volume of 7.5 ml using the Rabbit Reticulocyte IVT kit (Ambion) in the presence of 150 μCi ³⁵S-methionine (1000 Ci/mmoul). Subsequent steps in preparing the library were similar to those described for the low diversity, constrained peptide library construction. The reverse transcription reaction, necessary to generate fusion molecules containing the corresponding cDNA strands used Superscript II Reverse Transcriptase (200 units/μl, Gibco BRL) and followed the manufacturer’s recommended protocol.

Selection

EETI-II. Potential column binders were removed from the reverse transcribed EETI-II fusion library (1 pmol in 700 μl) by incubating the library with 100 μl CL-sepharose beads (Sigma) for 15 min at room temperature. The suspension was spun at 1000 rpm for 30 s and the supernatant applied to 100 μl of trypsin-agarose bead slurry (Sigma) in selection buffer (PBS pH 7.4, 1% Tween-20), and, in the instances where high stringency conditions were used, 200 μM benzamidine (Sigma). The mixture was incubated for 30 s at room temperature. The beads were centrifuged at 1000 rpm and washed 4–8 times with 1 ml selection buffer. Bound material was eluted with 4 × 100 µl of 0.33 N NaOH. During the course of the selection, the enrichment of the selected binders was determined by measuring the amount of radio labeled protein fusion bound to a 10% fraction of the trypsin–agarose bead slurry. All radiolabel measurements were performed on a scintillation counter (Wallac 1409). The eluates were neutralized and the DNA recovered by isopropanol precipitation.

c-Myc. The starting library (33 pmol, 2 × 10¹³ molecules) was incubated with a 12-fold excess of the c-myc binding antibody (Santa Cruz). An excess of protein A–sepharose was used to precipitate the peptide fusion–antibody complexes. After washing the sepharose with 5 vols of selection buffer (1 × PBS pH 7.4, 1 mg/ml BSA, 0.05% Tween 20) to remove non-specific binders, bound species were eluted by the addition of 15 mM acetic acid. The cDNA portion of the eluted fusion molecules was amplified by PCR and the resulting DNA was used to generate an enriched population of fusion products for ensuing rounds of selection.

Binding assay

EETI-II. Representative clones were subjected to further analysis. Selected clones of the EETI-II selection were prepared as mRNA-peptide fusions, the RNA in the fusion was digested with RNase One (Promega), and the target binding of the remaining peptide was determined in the following way. mRNA-peptide fusions (2–7 pmol) were prepared as described in the EETI-II experimental section. Following oligo dT₂₅ cellulose purification, 1.5 μl of RNase One (Promega) were added to, and incubated with, each mRNA–peptide fusion for 1 h at room temperature. Constant amounts of the mase-treated mRNA–peptide fusions were then separately incubated in 500 μl selection buffer (PBS pH 7.4, 1% Tween-20) at room temperature for 20 min with various concentrations (20–200 nM) of biotinylated bovine trypsin (Sigma). Streptavidin magnetic beads (200 μl; Dynal) were added to each incubate. The resulting suspensions were rotated for 10 min. The beads were washed twice with 500 μl selection buffer, and the total radioactive counts for the beads (bound radioligand value), unbound material and washes (free radioligand peptide bound).
value) was determined on a scintillation counter (Wallac). Dissociation constants for the different EETI-II variants were determined by plotting the data to the equation:

\[
\text{Fractional radioligand bound to trypsin} = \frac{[\text{trypsin concentration}]}{([\text{trypsin concentration}] + K_d)}
\]

The equation was derived from the law of mass action model for a ligand receptor interaction:

\[
\text{Radioligand} + \text{trypsin} \rightleftharpoons \text{Radioligand} - \text{trypsin}
\]

c-Myc. 0.2 pmol aliquot 35S-labeled monoclonal RNA-peptide fusion (R6-51, R6-52, R6-53, R6-55, R6-56, R6-58, R6-60, R6-61, R6-63, R6-66, R6-67, R6-68) or wild-type fusion construct (WT) were incubated with 100 pmol anti-myc monoclonal antibody 9E10, with 100 pmol anti-myc monoclonal antibody 9E10 in the presence of 2 nmol synthetic myc peptide, or without monoclonal antibody, respectively. The peptide fusion–antibody complexes were precipitated by addition of protein A–sepharose. The values represent the average percentage of fusion molecules that bound to the antibody and could be eluted with 15 mM acetic acid determined in duplicate binding reactions.

Sequence determination and analysis

DNA recovered from the fifth round of selection was PCR amplified, gel-purified (Qiagen) and cloned into the pCR® 2.1-TOPO® cloning system (Invitrogen). The resulting plasmids were purified (Qiagen). Cycle sequencing was performed on chosen samples and resolved on an ABI 310 sequencer. Sequences were analyzed with the MacVector analysis package (Oxford Molecular) and aligned according to amino acid sequence.

RESULTS

Library complexity

The two libraries constructed were a constrained, low-diversity library based on the EETI-II scaffold, which consists of a randomly mutated first loop, and a high-diversity linear peptide library, which consists of 27 randomized residues within a linear peptide. EETI-II has been shown to be tolerant of mutagenesis while retaining its knottin fold (Favel et al., 1989a,b; Christmann et al., 1999; Wentzel et al., 1999). The choice of these two libraries was based upon the basic structural requirements for high affinity binders to trypsin and the 9E10 anti-myc antibody. Whereas a constrained peptide would be a more likely trypsin inhibitor candidate than a linear peptide, the 9E10 anti-myc antibody recognizes a linear epitope. Therefore, the library based upon the EETI-II scaffold was used against trypsin, and the high diversity linear library was used against the 9E10 anti-myc antibody.

The theoretical complexity of the EETI-II library (a randomized 6-mer) is 206 or 6.40 × 106 different sequences; that of the linear library (a randomized 27-mer), 2027 or 1.34 × 1035 different sequences. For the EETI-II library, the starting library consisted of 1011 molecules (1 pmol), which exceeds the potential complexity of the EETI library. In contrast to the linear library, its starting library consisted of 1013 molecules (33 pmol), which only samples 10−23 of the potential complexity of the linear library.

EETI-II selection

In the first three rounds of selection the amount of radiolabeled protein fusion bound to immobilized trypsin increased from 1.8 to 22%, the latter being equivalent to the binding of wild-type EETI. Since these measurements of percent binding represent the ability of in vitro produced radiolabeled fusion to bind to trypsin, a dramatic extension of the c-terminus of the EETI scaffold by way of puromycin-poly-A linker and mRNA/DNA stills yields a construct capable of substantial binding to trypsin. In an attempt to further enrich the population, 200 μM benzamidine, a competitive inhibitor of trypsin (Kd = 18.4 μM; Mares-Guia and Shaw, 1964), was introduced into the incubation and wash buffers in the fourth round. Additionally, the binding time was reduced to 30 s. This increased stringency decreased the amount of binding to 0.5% in the fourth round. By the fifth round under the same set of increased stringency conditions, the binding percentage of radiolabeled protein fusions to immobilized trypsin increased to 2%, indicating a possible enrichment for higher affinity binders.

Sequence analysis. The cDNA pool from the fifth round of selection was cloned and 108 clones were sequenced. The sequence analysis revealed that wild-type EETI-II represented 20% of the clones (Table 1). The proline at the P2 position (nomenclature from Schechter and Berger, 1967) is conserved in all the analyzed sequences, and the arginine at the P4 position is conserved except for one instance where it is substituted by lysine. The hydrophobic residues isoleucine, leucine, or valine occupy the P1′ and P2′ positions. The P3′ and P4′ positions were highly variable (Table 1).

Characterization. Representative peptides were produced in vitro as mRNA-protein fusions, treated with RNase and

<table>
<thead>
<tr>
<th>Table 1. Representative clones from the EETI-II selection. Conserved residues are bold; residues that have similar properties are italic. The motifs PRIL or PRL are present in most of the clones</th>
</tr>
</thead>
<tbody>
<tr>
<td>wt EETI-II</td>
</tr>
<tr>
<td>20%</td>
</tr>
<tr>
<td>7.5%</td>
</tr>
<tr>
<td>3.75%</td>
</tr>
<tr>
<td>3.75%</td>
</tr>
<tr>
<td>3.75%</td>
</tr>
<tr>
<td>2.5%</td>
</tr>
<tr>
<td>2.5%</td>
</tr>
<tr>
<td>2.5%</td>
</tr>
<tr>
<td>2.5%</td>
</tr>
<tr>
<td>1.25%</td>
</tr>
<tr>
<td>Consensus</td>
</tr>
</tbody>
</table>

Copyright © 2002 John Wiley & Sons, Ltd.

tested for binding to trypsin (Fig. 2). Each peptide was therefore modified at the C-terminus due to the presence of the puromycin-DNA linker. Selected clones were further characterized by solution-based radioligand affinity analysis (Table 2). The binding efficiency of each clone was determined and, unexpectedly, a lower percentage of wild-type bound than any of the other screened clones; despite the fact that the wild-type clone bound the tightest, followed closely by the P1-lysine variant. The dissociation constant obtained for the wild-type clone (16 ± 3.2 nM) is close to that obtained by Wentzel et al. (1999; 6.09 ± 1.97 nM) and Le-Nguyen et al. (1990; 1.8 nM) for wild-type EETI-II constructs with an extended C-terminus. Under the conditions chosen for production, purification and binding, the discrepancy between percentage binding and dissociation constant for the wild-type clone may be a result of differences in the percentage of active (i.e. folded) wild-type clone relative to similar percentages for the other selected anti-trypsin clones.

c-Myc selection

Five rounds of selection were performed using the c-Myc monoclonal antibody, 9E10, as the target. After four rounds of selection, the amount of radiolabeled mRNA–peptide fusion recovered increased (Fig. 3). After the fifth round, 34% of the linear peptide library bound to the c-Myc antibody; this is comparable to the binding of wild-type c-Myc in control experiments. No further increase in binding was observed after round five. Eluants from the fifth and the sixth round were used to characterize and sequence of the resulting clones.

Sequence analysis. The cDNA pools from the fifth and sixth rounds were cloned and 116 colonies were sequenced. As presented in Table 3, two selected clones contained the sequence of the wild-type c-Myc epitope (EQKLISEEDL). A third clone differed from the wild-type by two point mutations in the nucleotide sequence, only one of which altered an amino acid (Ile to Val). The consensus sequence selected [X(Q,E)XLISEXX(L,M)] included the full length 10 amino acid wild-type sequence; identical or homologous sequences to the four of the core residues, LISE, were conserved in 86% of the 116 clones examined.

Characterization. A series of experiments was performed to determine if the pool of selected peptides bound to the anti-c-Myc antibody 9E10. Fusion products from the sixth round of selection were evaluated under different immunoprecipitation conditions. In order to confirm that selected fusion molecules interact with the antigen-binding site of the anti-c-Myc antibody 9E10, synthetic c-Myc peptide was introduced as a competitor for binding to the anti-c-Myc antibody. The percentage of fusion bound to the antibody

Table 2. Affinities of the selected EETI-II variants measured in three different experiments. The clones used for affinity characterization have the puromycin–DNA (30mer poly-dA) linker at their C-terminus

<table>
<thead>
<tr>
<th>Amino acid sequence (residues 3–8)</th>
<th>$K_d$ (nM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PRILMR</td>
<td>16 ± 3.2</td>
</tr>
<tr>
<td>PRVLAL</td>
<td>52 ± 21.0</td>
</tr>
<tr>
<td>PRLVQR</td>
<td>53 ± 9.0</td>
</tr>
<tr>
<td>PRLAAP</td>
<td>81 ± 21.0</td>
</tr>
<tr>
<td>PKLLAP</td>
<td>21 ± 2.4</td>
</tr>
</tbody>
</table>
decreased with increasing amounts of c-Myc peptide, which implies that the selected population binds to the c-Myc antigen binding site on the 9E10 antibody. No binding was detected in experiments without an antibody, which indicates that the selected peptides do not bind to protein A-agarose. The selected peptides were shown not to bind an irrelevant antibody target, an anti-integrin monoclonal antibody. The combined data suggests that the selected population interacts specifically with the target molecule against which it was raised.

Similar immunoprecipitation assays were performed on 12 selected sequences from the sixth round pool. These experiments confirmed the specific binding of the RNA–peptide fusions to the antigen-binding site of the anti-c-Myc monoclonal antibody (Fig. 4). The selected sequences bound the c-Myc antibody with an affinity similar to that of the wild-type Myc fusion.

DISCUSSION

mRNA display was applied to selections against two targets, the serine protease trypsin and the anti-c-Myc antibody 9E10, in order to determine the minimal binding sequence of the peptide to its protein partner. Following multiple rounds, each selection converged to a consensus binding sequence. Remarkably, in both randomized peptide selections, the highest affinity selected sequences were identical to the wild-type binding peptide, which suggests that the selection of these peptides is driven almost entirely by optimal binding interactions and not by other steps in the selection protocol.

The EETI-II selection mapped the first loop of six amino acids, which binds to the trypsin active-site. The trypsin-binding sequences from the EETI-II selection are highly similar to the wild-type EETI-II first loop, including three highly conserved residues and a fourth position with a preference for hydrophobic amino acids. These results

<table>
<thead>
<tr>
<th>c-Myc epitope</th>
<th>E</th>
<th>Q</th>
<th>K</th>
<th>L</th>
<th>I</th>
<th>S</th>
<th>E</th>
<th>E</th>
<th>D</th>
<th>L</th>
</tr>
</thead>
<tbody>
<tr>
<td>R6-51</td>
<td>C</td>
<td>A</td>
<td>S</td>
<td>V</td>
<td>I</td>
<td>S</td>
<td>E</td>
<td>R</td>
<td>E</td>
<td>C</td>
</tr>
<tr>
<td>R6-52</td>
<td>E</td>
<td>E</td>
<td>Y</td>
<td>L</td>
<td>V</td>
<td>S</td>
<td>E</td>
<td>Y</td>
<td>V</td>
<td>M</td>
</tr>
<tr>
<td>R6-53</td>
<td>R</td>
<td>Q</td>
<td>Y</td>
<td>L</td>
<td>I</td>
<td>S</td>
<td>E</td>
<td>Y</td>
<td>E</td>
<td>H</td>
</tr>
<tr>
<td>R6-55</td>
<td>L</td>
<td>Q</td>
<td>R</td>
<td>L</td>
<td>I</td>
<td>S</td>
<td>E</td>
<td>Q</td>
<td>M</td>
<td>F</td>
</tr>
<tr>
<td>R6-56</td>
<td>I</td>
<td>V</td>
<td>R</td>
<td>L</td>
<td>L</td>
<td>S</td>
<td>E</td>
<td>Y</td>
<td>H</td>
<td>M</td>
</tr>
<tr>
<td>R6-58</td>
<td>E</td>
<td>E</td>
<td>Y</td>
<td>L</td>
<td>L</td>
<td>S</td>
<td>E</td>
<td>Y</td>
<td>V</td>
<td>M</td>
</tr>
<tr>
<td>R6-60</td>
<td>M</td>
<td>Q</td>
<td>N</td>
<td>L</td>
<td>I</td>
<td>S</td>
<td>E</td>
<td>H</td>
<td>E</td>
<td>L</td>
</tr>
<tr>
<td>R6-61</td>
<td>T</td>
<td>M</td>
<td>D</td>
<td>L</td>
<td>I</td>
<td>P</td>
<td>E</td>
<td>H</td>
<td>Y</td>
<td>M</td>
</tr>
<tr>
<td>R6-63</td>
<td>E</td>
<td>Q</td>
<td>K</td>
<td>L</td>
<td>I</td>
<td>S</td>
<td>E</td>
<td>E</td>
<td>D</td>
<td>L</td>
</tr>
<tr>
<td>R6-66</td>
<td>D</td>
<td>M</td>
<td>M</td>
<td>L</td>
<td>I</td>
<td>S</td>
<td>E</td>
<td>K</td>
<td>E</td>
<td>L</td>
</tr>
<tr>
<td>R6-67</td>
<td>F</td>
<td>Q</td>
<td>A</td>
<td>L</td>
<td>I</td>
<td>A</td>
<td>E</td>
<td>E</td>
<td>E</td>
<td>L</td>
</tr>
<tr>
<td>R6-68</td>
<td>Q</td>
<td>R</td>
<td>V</td>
<td>L</td>
<td>I</td>
<td>S</td>
<td>E</td>
<td>F</td>
<td>W</td>
<td>L</td>
</tr>
<tr>
<td>consensus</td>
<td>X</td>
<td>Q</td>
<td>E</td>
<td>X</td>
<td>L</td>
<td>I</td>
<td>S</td>
<td>E</td>
<td>X</td>
<td>L/M</td>
</tr>
</tbody>
</table>
imply that the PRI(L, V)L is essential for trypsin binding. The observation that the P3' and P4' positions vary the most in the selected EETI-II variants is in accord with the co-crystal structure of CMTI-trypsin (Helland et al., 1999). In the crystal structure of the complex between CMTI and trypsin, the two residues at the P3' and P4' positions are more than 4 Å from the trypsin surface (Helland et al., 1999). Since these positions are not making any direct contact with the trypsin surface, we would expect that the selection pressure at the P3' and P4' positions would be relatively low.

In other natural anti-trypsin knottins (Holak et al., 1989; Wilusz et al., 1983; Ling et al., 1993; Hatakeyama et al., 1991; Otlewski et al., 1987; Heitz et al., 1989; Wieczorek et al., 1985), the PRI(L)M sequence in the first loop is the most common; this agrees with our EETI-II selection results, where clones containing the PRI(L)MR sequence were dominant. The second most common sequence in this selection was PRI(L)AL. A similar sequence (PRI(L)MP), can be found in the trypsin inhibitor knottin from Luffa cylindrical (Ling et al., 1993).

Another interesting variant from the EETI-II selection was found to have as its first loop sequence PKLLAP. This variant is similar in affinity to the wild-type sequence, despite the substitution of Ile5 to Leu. NMR structural studies of the wild-type Ile5 to Leu EETI-II variant showed a first loop more disordered than the wild-type first loop, resulting in reduced trypsin binding (Nielsen et al., 1994). This decrease in binding relative to wild-type was not apparent in our PKLLAP variant, possibly because differences at other positions in the first loop compensated for the detrimental effect of the Ile5 to Leu substitution.

The c-Myc antibody, Myc1-9E10, was originally raised against the 32 amino acid C-terminus of the human c-myc oncoprotein (Chan et al., 1987). The longest contiguous selected peptide sequence with complete identity to c-myc was EQKLISEEDL. Despite the low initial abundance of peptides containing this 10-mer sequence in the 27-mer linear peptide library, the selection pressure favored its amplification and emergence. The sequence obtained is long enough that a simple PIR (Barker et al., 2001) search would have identified the c-Myc oncogene if the identity of the antigen for the 9E10 antibody had not been known. In fact, performing a PIR pattern search on the degenerative consensus sequence, Q/E-X-L-I-S-E-X-X-L/M, yielded...
predominantly matches for c-Myc. Thirty two percent of the matches (i.e. eight out of 25) from the PIR database were the c-Myc oncogene. All human matches were c-Myc. The report that the capacity of a monoclonal antibody binding site may be as short as 5–9 residues (Dunn et al., 1999) suggests that an antibody binding region can be shorter than the selected 10 amino acid identity of the c-Myc antigen. Indeed, the results of the selection against the 9E10 antibody show that enriched peptides converge on the consensus sequence LISE. Nearly 61% of the clones contain this sequence, and another 25% contain single homologous substitutions within the LISE consensus. The remaining 14% had no more than two nonhomologous substitutions within the LISE consensus. Two of the more frequent nonhomologous substitutions were a proline for the serine and a tryptophon for the glutamic acid of the LISE consensus. Although flanking amino acids may contribute to the affinity, the conservation of the LISE tetrapeptide sequence within the majority of analyzed clones suggest that LISE is the most important determinant of binding.

In summary, we used mRNA display with two peptide libraries of different diversities; a low diversity, constrained knottin library with six randomized residues in the first loop; and a high diversity 27-mer, linear peptide. Both selections converged to consensus sequences, and some clones contained the precise wild-type sequence of their respective ligands. In both cases, the sequence requirements for molecular recognition of a peptide by a macromolecule were mapped at a high resolution. The application of mRNA display to identify native ligands and precise primary sequence requirements of peptides binding to proteins may become a useful tool both in proteomic applications and peptidomimetic design.

Acknowledgements

The authors wish to thank Michael McPherson and Thomas Cujec for critical reading of the manuscript.

REFERENCES


Nemoto N, Miyamoto-Sato E, Husimi Y, Yanagawa H. 1997. In vitro virus: bonding of mRNA bearing puroMyctin at the 3′-terminal end to the C-terminal end of its encoded protein on