mRNA Display Based Selections Using Synthetic Peptide and Natural Protein Libraries

Steve W. Cotten, **Jianwei Zou**, **Rong Wang**, **Bao-cheng Huang**, and **Rihe Liu**[§]

Eshelman School of Pharmacy and Carolina Center for Genome Sciences, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599-7568

Abstract

mRNA display is a powerful *in vitro* selection technique that can be applied towards the identification of peptides or proteins with desired properties. The physical conjugation between a protein and its own RNA presents unique challenges in manipulating the displayed proteins in an RNase free environment. This protocol outlines the generation of synthetic peptide and natural proteome libraries as well as the steps required for generation of mRNA-protein fusion libraries, *in vitro* selection, and regeneration of the selected sequences. The selection procedures for the identification of Ca^{2+} dependent calmodulin binding proteins from synthetic peptide and natural proteome libraries are presented.

Keywords

mRNA display; genotype-phenotype conjugation; *in vitro* selection; synthetic combinatorial peptide library; natural proteome library; conditional protein-protein interaction

1. Introduction

mRNA display is a genotype-phenotype conjugation method that allows the amplification-based, iterative rounds of *in vitro* selection to be applied to peptides and proteins (1-4). mRNA display can be used to display both long natural proteome and short synthetic peptide libraries with high diversity. Compared to prior peptide or protein selection methods, mRNA display has several major advantages. First, the genotype is covalently linked to and is always present with the phenotype. This stable linkage makes it possible to use any desired conditions during the selection process and titrate the stringency of the selection. Second, the complexity of the peptide or protein library can be close to that of the RNA or DNA pools. Peptide or protein libraries containing as many as $10^{12} \sim 10^{14}$ unique sequences can be readily generated and selected, a few orders of magnitude higher than that can be achieved using phage display and other selection platforms. Therefore, both the likelihood of isolating rare sequences and the diversity of the sequences isolated in a given selection are significantly increased. In principle, mRNA display can be used for any *in vitro* selection that aims at identifying peptide or protein sequences with desired properties. The various applications of mRNA display can be classified based on the type library

[§] To whom correspondence should be addressed. Tel: (919) 843-3635 rihe_liu@unc.edu.

constructed for the selection. The first type of libraries that can be used for mRNA display-based selection are synthetic combinatorial peptide libraries. These libraries can be designed from structured protein scaffolds containing totally or partially randomized amino acids on surface loops or unstructured peptides consisting of randomized residues (5-12). The second type of libraries are natural proteome libraries derived from the mRNAs of any organism, tissue, or treatment (*i.e.* drug/environmental insult) of interest (13-21). To date, mRNA display has been successfully applied in the identification of drug-binding targets, mapping of the protein-protein interactions and DNA-protein interaction networks, elucidation of the enzyme-substrate interactions, and improvement of the binding affinities of existing affinity molecules (13-21).

Much insight into the function of a protein can be gained by studying its interaction with other proteins. Often, such protein-protein interactions only occur under specific conditions. One effective strategy to get a thorough understanding of a protein of interest is to map its conditional protein-protein interactions on a proteome-wide scale. mRNA display can provide a global picture of conditional protein-protein interactions and allows for simultaneous search of the sequence space to understand the nature and specificity of the target protein with its natural or synthetic interacting partners.

Functional selections using mRNA display approach can be challenging due to the necessity of manipulating from nanomolar to low micromolar amounts of radiolabeled proteins in an RNase-free environment. The success of a selection from a highly diversified synthetic peptide or natural proteome library displayed on its own mRNA relies on a selection scheme that allows specific enrichment of sequences with desired properties while minimizing isolation of nonspecific sequences. In general, immobilization of a target of interest followed by competitive elution of bound molecules using an excess of unmodified target is an effective approach for specific enrichment. If the protein interaction is conditional, binding and elution steps for the selection can be designed such that the presence or absence of small molecules or conditions such as Ca^{2+}, cofactors, light, temperature, or ionic strength dictate when molecules are released from the target. These "binary" binding events used for mRNA display selection are very effective at rapidly enriching target sequences. Selections that use binding events where conditions that achieve specificity are not yet known often take many more cycles of selection and can be challenging to enrich the desired functional sequences from non-specific ones (See Note 1).

We describe here the procedures of using mRNA display to perform two selections against the same target from both a natural proteome and a synthetic combinatorial peptide library, so the interacting partners from both natural proteome and synthetic sequence space can be examined. Specifically, we use the identification of Ca^{2+}-dependent binding synthetic peptides or natural proteins against calmodulin, the major transducer of Ca^{2+} signals in eukaryotes, as examples. Since the desired binding requires Ca^{2+} ion, EGTA can be used to

---

[1]Typically, two to five rounds of selection are necessary to isolate peptides or proteins with desired properties for selections using a competitive or conditional elution. If the selection is simply based on nonspecific elution to disrupt the interaction between a target and mRNA displayed library, the eluted pool should be carefully monitored after each round of selection to prevent enrichment of non-specific sequences. This can be done through the monitoring of both radioactive counts and sequence analysis of the selected molecules from the eluted pool.

effectively remove Ca^{2+} from solution, resulting in a conformation change in calmodulin and specific release of the bound mRNA-protein fusion molecules under very mild conditions. The selection scheme for this conditional protein-protein interaction is illustrated in Figure 1. For other protein-protein interactions, gentle and competitive elution should be applied whenever possible.

2. Materials

2.1. Reagents

1. Expand long template PCR system (Roche)

2. T7 RNA polymerase (NEB)

3. RNase-free DNase (Promega)

4. Retic lysate IVT™ kit (Ambion)

5. [^{35}S]-L-methionine (Perkin Elmer)

6. Oligo(dT) cellulose (Ambion)

7. RNase-free 10 mL poly-prep chromatography column (Biorad)

8. SuperScript II RNase H$^-$ reverse transcriptase (Invitrogen)

9. anti-FLAG M2 affinity gel (Sigma)

10. FLAG peptide (Sigma)

11. Biotinylated calmodulin (CalBioChem)

12. Binding buffer: 50 mM Tris-HCl, pH 7.5, 150 mM NaCl, 0.05% Tween-20, 1 mg/ml BSA, 5 mM 2-mercaptoethanol, and 0.5 mM $CaCl_2$

13. Elution buffer: 50 mM Tris-HCl, pH 7.5, 150 mM NaCl, 0.05% Tween-20, 1 mg/ml BSA, 5 mM 2-mercaptoethanol, and 2 mM EGTA

2.2. Equipment

1. Thermal Cycler

2. Nanodrop Spectrophotometer

3. UV lamp (Black Ray Lamp 365 nm, 0.16 Amps)

4. Barnstead Labquake Shaker/Rotator

5. Scintillation Counter

3. Methods

3.1. Construction of Synthetic cDNA Library Coding Synthetic Combinatorial Peptide Library

For the synthetic peptide library, a high quality cDNA library is synthesized and assembled according to the published protocol (5, 10). Each sequence in the cDNA library contains a T7 RNA polymerase promoter, a TMV translation enhancer sequence, an N-terminal FLAG

tag coding sequence, a random cassette encoding twenty consecutive random codons, and a His× 6 tag coding sequence. The amino acid compositions in the random region can be designed to be close to the natural proteins, or contain any residues of interest at desired positions and levels, while the possibility of coding stop codons is minimized (5). The random region can be readily synthesized by oligo synthesizer using three phosphoramidite mixtures with appropriate proportions of each of the four phosphoramidites in the DNA synthesis, corresponding to each of the three positions in the random codons (5). Consensus sequences are flanked on both sides of the random region to facilitate annealing with other oligos and PCR amplification for assembly of the final library.

3.2 Construction of Natural cDNA Library Coding Natural Proteome Library

Natural cDNA libraries suitable for mRNA display-based selection can be generated from any organisms or tissues where approximately 1.0 μg of high quality mRNA can be isolated. Specifically, high-quality poly (A$^+$) mRNA sequences are prepared by removing genomic DNAs, ribosomal RNAs and tRNAs through three rounds of stringent oligo(dT) purification. If the mRNAs of interest do not contain poly (A$^+$) tail (*i.e.* bacterial mRNAs), appropriate approaches such as use of a combination of microExpress bacterial mRNA enrichment kit and RiboPure-Bacteria kit (Ambion) should be employed for at least three cycles to rigorously remove ribosomal RNAs and tRNAs. The resulting mRNAs are reverse transcribed to generate the corresponding cDNAs with a nucleotide mix containing 5-Me dCTP to protect ORF regions from the subsequent restriction digestion of the directional linker. After the generation of blunt ends with T4 DNA polymerase, a directional linker (*i.e. EcoR* I/*Hind* III linker (GCTTGAATTCAAGC)) is ligated to both ends that allows for directional ligation of different left and right consensus sequences upstream and downstream of the coding cDNA region. The left consensus sequence contains a T7 promoter and a deletion mutant of TMV 5'-UTR for efficient *in vitro* transcription and translation, respectively. The right consensus cassette contains a short sequence for hybridizing with a puromycin-containing oligo linker. Sequences that code for various affinity tags can be included on both ends to facilitate the purification of mRNA-displayed peptide or protein library. Depending on the design of the selection scheme, some special sequences can be incorporated. For instance, an N-terminal BirA site allowing the site-specific introduction of a biotin molecule at each of the mRNA-protein fusion molecules could greatly facilitate the immobilization of mRNA-displayed proteome library onto a solid support matrix. A highly specific protease recognition site (*e.g.* for TEV protease) could facilitate the gentle release of the fusion molecules at any desired selection step. A 5-amino-acid PKA recognition site (RRASV) should allow for the radiolabeling of each peptide or protein sequence in the library with the same efficiency. The ligated dsDNA is PCR amplified using two primers complementary to the consensus T7 promoter and 3' linker-hybridization region at the 5' and 3' ends, respectively. Finally, the PCR product is carefully fractionated using a spin column or gel electrophoresis to generate a cDNA library with desired length distribution (*i.e.* 0.5 - 2 kb). A critical requirement of mRNA display is the removal of constructs that contain stop codons, which prematurely terminate translation and prevent the conjugation between mRNA and its coding protein. This is achieved by generating and purifying the mRNA-protein fusion molecules directly followed by regeneration back of the cDNA.

Constructs containing frame-shifts or stop codons will not be purified using both nucleic acid and protein affinity tags, effectively removing them from the library (17, 20).

3.3. Generation of mRNA-displayed Synthetic Peptide or Natural Proteome library

After the construction of a cDNA library and prior to the selection, one round of pre-selection should be performed in order to remove out of frame sequences and sequences containing stop codons (5, 17, 20). One round of mRNA display consists of the following steps: *in vitro* transcription, DNase digestion, conjugation with the puromycin oligo linker, *in vitro* translation/fusion formation, oligo(dT) mRNA purification, reverse transcription, protein affinity purification, pre-selection or functional selection, and regeneration of the selected sequences. The details for most of these procedures are described in the previous chapter. Briefly, the cDNA library is *in vitro* transcribed using T7 RNA polymerase; the mRNA templates with puromycin at the 3' ends are generated by crosslinking with an oligonucleotide containing a psoralen residue and a puromycin residue at its 5' and 3' ends, respectively; and *in vitro* translation is performed using rabbit reticulocyte lysate and mRNA-protein fusion formation accomplished under optimized conditions (3). After the fusion formation, free mRNA templates and mRNA-protein fusions are first isolated from the translation reaction mixture using an oligo(dT) column, followed by converting into DNA/RNA hybrids through reverse transcription to remove secondary mRNA structures that might interfere with the subsequent selection. The resulting mRNA-displayed synthetic peptide or natural proteome library is purified using an affinity column (*i.e.* anti-FLAG) and used for selection.

3.4. Functional Selection of Synthetic Peptides or Natural Proteins that Bind to Calmodulin in a Ca^{2+}-dependent Manner

One critical issue in using mRNA display for functional selection is to design the selection scheme that minimizes the capture of nonspecific sequences while maximizes efficient enrichment of sequences with desired properties. All recombinant protein targets, small molecules, and other components used in a selection should be RNase- and DNase-free to maintain intact mRNA-protein fusion molecules. Typically, binding buffer should be supplemented with RNase-free tRNA and BSA to reduce non-specific protein and RNA binding. It is important to optimize the wash steps that will be used prior to elution. A stringent wash will effectively remove molecules that may bind to the target non-specifically but may also result in the loss of desired weakly bound sequences. An appropriate wash volume and stringency should be determined prior to the selection by using both positive and negative control sequences.

1. Dilute the purified mRNA-displayed library in 0.5-2 mL of binding buffer.

2. Apply the mRNA-displayed library to a pre-column of 100-300 μl UltraLink Plus Streptavidin Agarose beads to remove sequences that may bind matrix non-specifically. Collect the flow-through for subsequent selection steps.

3. Incubate the flow-through with 5-25 μg of biotinylated calmodulin for 30-90 mins at 4° C.

4. Add 100-300 μl of streptavidin-agarose beads pre-equilibrated in binding buffer to the mixture and incubate for 30 minutes at room temperature with gentle mixing.

5. Load the slurry mixture to an empty nuclease-free 10 mL poly-prep chromatography column. Retain flow-through for further analysis.

6. Wash unbound molecules from the column with 9-24 column volumes (3×3-8×3 CV) of binding buffer. Retain each wash fraction for further analysis.

7. Elute the fusion molecules that bind to the target protein under desired conditions (see Note 2). Elute molecules that bind to calmodulin in a Ca^{2+}-dependent manner with 1 column volume of elution buffer (binding buffer minus $CaCl_2$ plus 2 mM EGTA) 4 times. Retain each elution fraction for further analysis.

8. Count 1/100 of each fraction, including flow-through, wash, elution, and 1/10 of the beads using liquid scintillation (see Note 3).

9. PCR-amplify the eluted mRNA-protein fusion sequences under the conditions that have been previously titrated on a small scale (See Note 4).

10. Perform the next round of selection using the regenerated cDNA library (see Note 5). Enrichment data measuring radioactive counts from the flow-through, wash and elution steps can be used to monitor the enrichment of the desired sequences as demonstrated in Figure 2.

3.5. Functional Confirmation

After selection, several hundred clones are typically sequenced to determine the identity of the selected proteins. When the selection is from an mRNA-displayed natural protein library, cDNA microarrays can also be used to reveal the identity of the genes that are present in the pool.

It is of great importance to develop a high throughput assay that allows for biochemical characterization of the selected clones to confirm which of the sequences indeed possess the desired properties (See Note 6). The nucleic acid portion of the fusion molecule contains all the necessary sequences for efficient *in vitro* transcription and translation, and therefore can be directly used after PCR amplification as template to generate radiolabeled proteins. Autoradiography can be used in binding assays against the immobilized target protein,

---

[2]For other protein-protein interactions, 10-20 folds excess of unmodified target protein can be used to compete with an immobilized target protein to release the target-binding partners. The milder and more selective the elution conditions, the more efficiently each round of selection will enrich for the desired functional sequences.

[3]Radioactive counts should return to baseline levels prior to elution of target binding molecules. The amount of radioactivity present in elution fractions will generally be low in the first several rounds of selection. However, after several rounds of enrichment, the percentage of radioactivity present in the elution fractions from the selection will dramatically increase.

[4]The number of PCR cycles required to effectively regenerate the cDNA library without over amplification is critical for the next round of selection. Too many cycles of PCR have the potential to generate artificial sequences that will overwhelm the selected pools. A small-scale diagnostic PCR should be performed at the end of the each selection cycle to determine the appropriate number of cycles necessary to regenerate the library. For a synthetic cDNA library, the length of the PCR products is fixed and therefore the quality of the library is easy to determine. For a natural cDNA library, the distribution of the original library should be used as a benchmark to determine the number of PCR cycles required for regeneration.

[5]For the iterative rounds of selection, the general selection procedure remains the same for the generation and purification of the mRNA-protein fusion molecules, but the selection pressure can be gradually increased for subsequent selection cycles. This can be achieved by reducing the amount of target protein used during binding, shortening the incubation time or increasing the incubation temperature, adding binding competitors, increasing the wash stringency (times and volumes) or elution specificity.

allowing for sensitive analysis of binding between the two molecules. As illustrated in Figure 3, Ca^{2+}-dependent calmodulin-binding analysis of individual natural proteins (Figure 3A) or synthetic peptides (Figure 3B) isolated from the selection showed that the selected protein or peptide sequences bound to calmodulin in a Ca^{2+}-dependent manner (10, 17).

Acknowledgments

References

1. Roberts RW, Szostak JW. RNA-peptide fusions for the in vitro selection of peptides and proteins. Proc Natl Acad Sci U S A. 1997; 94:12297–12302. [PubMed: 9356443]

2. Nemoto N, Miyamoto-Sato E, Husimi Y, Yanagawa H. In vitro virus: bonding of mRNA bearing puromycin at the 3′-terminal end to the C-terminal end of its encoded protein on the ribosome in vitro. FEBS Lett. 1997; 414:405–408. [PubMed: 9315729]

3. Liu R, Barrick JE, Szostak JW, Roberts RW. Optimized synthesis of RNA-protein fusions for in vitro protein selection. Methods Enzymol. 2000; 318:268–293. [PubMed: 10889994]

4. Szostak, J.; Roberts, R.; Liu, R. 1998. 2000. 2001. 2001. 2001. 2001. 2001. WO/1998/031700 WO/2000/047775 U.S. Patent 6,207,446 U.S. Patent 6,214,553 U.S. Patent 6,258,558 U.S. Patent 6,261,804 U.S. Patent 6,281,344

5. Cho G, Keefe AD, Liu R, Wilson DS, Szostak JW. Constructing high complexity synthetic libraries of long ORFs using in vitro selection. J Mol Biol. 2000; 297:309–319. [PubMed: 10715203]

6. Wilson DS, Keefe AD, Szostak JW. The use of mRNA display to select high-affinity protein-binding peptides. Proc Natl Acad Sci U S A. 2001; 98:3750–3755. [PubMed: 11274392]

7. Baggio R, et al. Identification of epitope-like consensus motifs using mRNA display. J Mol Recognit. 2002; 15:126–134. [PubMed: 12203838]

8. Xu L, et al. Directed evolution of high-affinity antibody mimics using mRNA display. Chem Biol. 2002; 9:933–942. [PubMed: 12204693]

9. Getmanova, Ev, et al. Antagonists to human and mouse vascular endothelial growth factor receptor 2 generated by directed protein evolution in vitro. Chem Biol. 2006; 13:549–556. [PubMed: 16720276]

10. Huang BC, Liu R. Comparison of mRNA-display-based selections using synthetic peptide and natural protein libraries. Biochemistry. 2007; 46:10102–10112. Epub 12007 Aug 10109. [PubMed: 17685586]

11. Cho GS, Szostak JW. Directed evolution of ATP binding proteins from a zinc finger domain by using mRNA display. Chem Biol. 2006; 13:139–147. [PubMed: 16492562]

12. Olson, Ca; Liao, Hi; Sun, R.; Roberts, RW. mRNA display selection of a high-affinity, modification-specific phospho-IkappaBalpha-binding fibronectin. ACS Chem Biol. 2008; 3:480–485. Epub 2008 Jul 2001. [PubMed: 18590330]

---

[6]One major challenge in the functional selection is how to remove abundant sequences that could dominate the pool as selection goes on. Some peptide or protein sequences might be preferentially enriched, which could interfere with the isolation of other sequences with the same or even better properties. These abundant sequences can be effectively removed at the mRNA level. Specifically, after determining the identity of the abundant molecules by sequencing a couple of hundred clones, biotinylated antisense oligos are designed against the variable region mapped by aligning the abundant sequences (17, 20). Following RNA hydrolysis and neutralization, hybridization of the biotinylated oligos with the complementary cDNA, and passage through a streptavidin column, these abundant sequences can be effectively removed. This unique feature significantly increases the chance of discovering non-abundant sequences.

13. Hammond PW, Alpin J, Rise CE, Wright M, Kreider BL. In vitro selection and characterization of Bcl-X(L)-binding proteins from a mix of tissue-specific mRNA display libraries. J Biol Chem. 2001; 276:20898–20906. [PubMed: 11283018]

14. Cujec TP, Medeiros PF, Hammond P, Rise C, Kreider BL. Selection of v-abl tyrosine kinase substrate sequences from randomized peptide and cellular proteomic libraries using mRNA display. Chem Biol. 2002; 9:253–264. [PubMed: 11880040]

15. McPherson M, Yang Y, Hammond PW, Kreider BL. Drug receptor identification from multiple tissues using cellular-derived mRNA display libraries. Chem Biol. 2002; 9:691–698. [PubMed: 12079780]

16. Horisawa K, et al. In vitro selection of Jun-associated proteins using mRNA display. Nucleic Acids Res. 2004; 32:e169. [PubMed: 15576676]

17. Shen X, Valencia CA, Szostak JW, Dong B, Liu R. Scanning the human proteome for calmodulin-binding proteins. Proc Natl Acad Sci U S A. 2005; 102:5969–5974. Epub 2005 Apr 5919. [PubMed: 15840729]

18. Shen X, et al. Ca(2+)/Calmodulin-binding proteins from the C. elegans proteome. Cell Calcium. 2008; 43:444–456. Epub 2007 Sep 2012. [PubMed: 17854888]

19. Tateyama S, et al. Affinity selection of DNA-binding protein complexes using mRNA display. Nucleic Acids Res. 2006; 34:e27. [PubMed: 16478713]

20. Ju W, et al. Proteome-wide identification of family member-specific natural substrate repertoire of caspases. Proc Natl Acad Sci U S A. 2007; 104:14294–14299. Epub 12007 Aug 14229. [PubMed: 17728405]

21. Fukuda I, et al. In vitro evolution of single-chain antibodies using mRNA display. Nucleic Acids Res. 2006; 34:e127. Epub 2006 Sep 2029. [PubMed: 17012279]
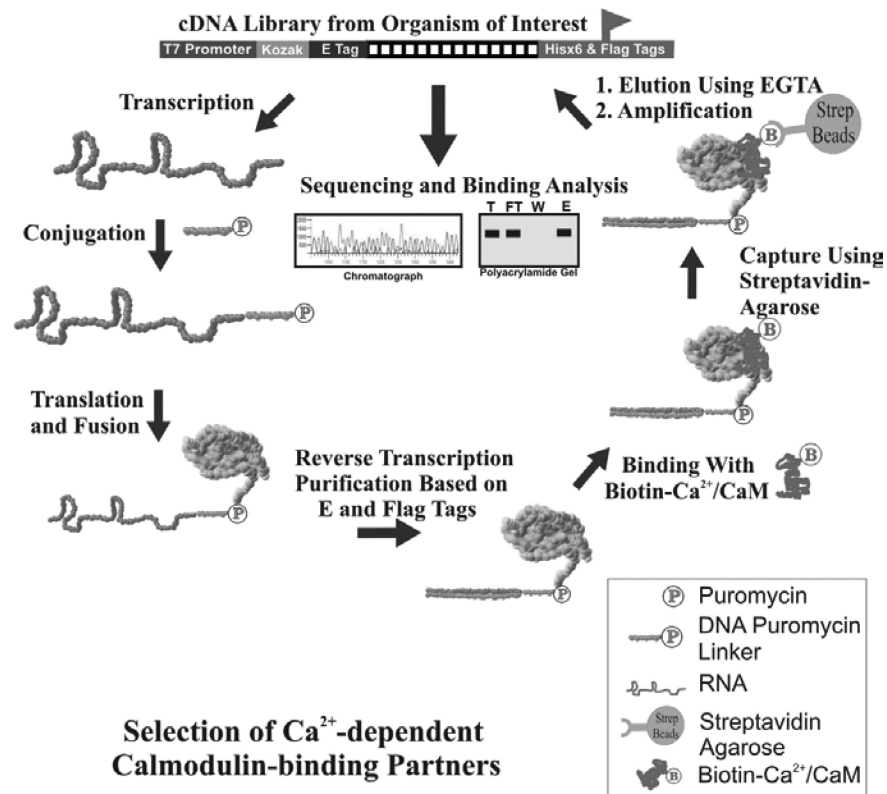
Figure 1.
Scheme for the selection of Ca^{2+}-dependent calmodulin-binding partners from an mRNA-displayed synthetic peptide or natural proteome library.
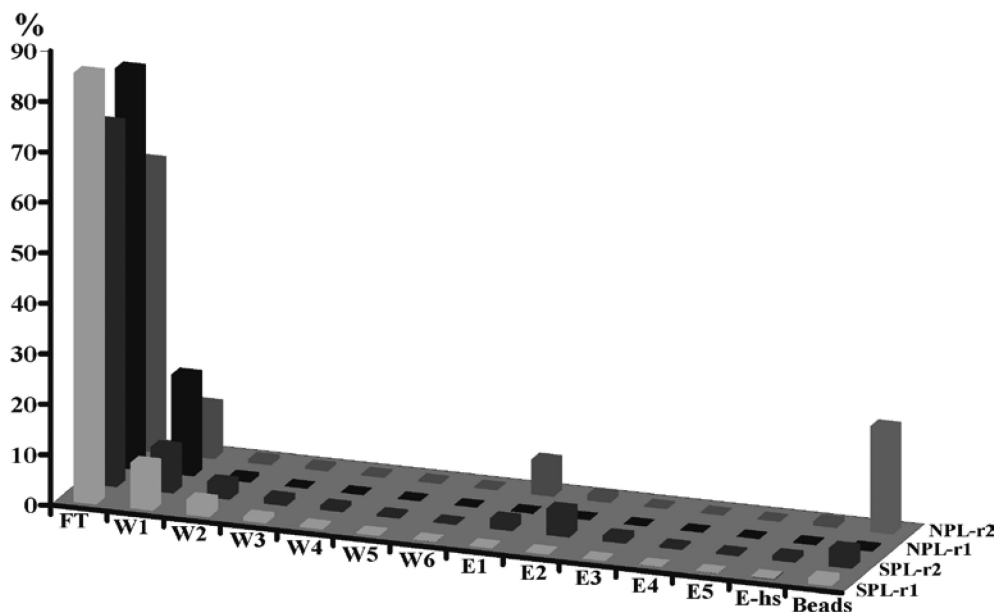
Figure 2.
Elution profile of Ca^{2+}-dependent calmodulin-binding selection using an mRNA-displayed random peptide library (S-r1 and S-r2) or human proteome library (N-r1 and N-r2). The Y axis is the percentage (%) of the radiolabeled fusion molecules present in each fraction. The percentage was calculated based on the input of fusion molecules that were labeled with [^{35}S]-methionine. FT: flowthrough; W1-W6: washes in the presence of Ca^{2+}; E1-E5: elution with a buffer containing 2 mM EGTA and 150 mM NaCl; Ehs: elution with a buffer containing 2 mM EGTA and 1 M NaCl; Bds, on streptavidin-agarose beads. Round 1 and round 2 are labeled as r1 and r2, respectively.
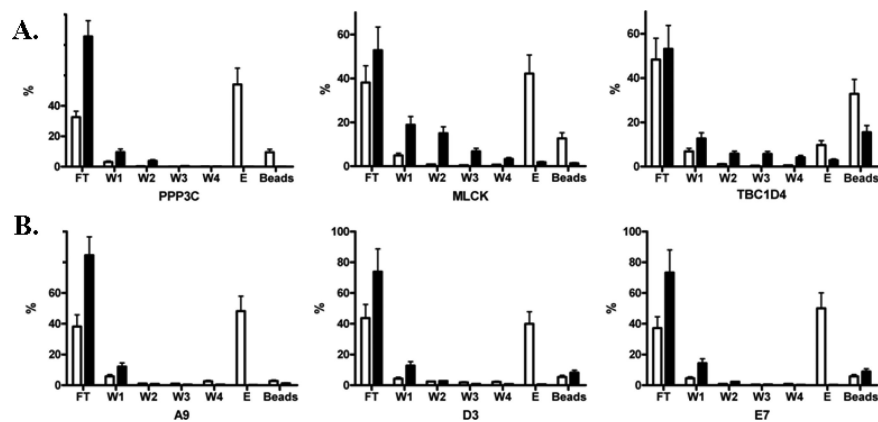
Figure 3.

In vitro binding analysis of three selected natural protein fragments from human proteome library (A) and three selected short peptides from combinatorial peptide library (B) with biotinylated calmodulin. Radiolabeled proteins or peptides were generated by coupled transcription/translation and purified using Ni-NTA beads. The purified protein product was incubated with biotinylated calmodulin in the presence of 1 mM $CaCl_2$ or 2 mM EGTA. The complex was captured with streptavidin-agarose beads. After extensive wash, the bound proteins or peptides were eluted using a buffer containing 2 mM EGTA. The Y axis is the percentage (%) of a protein or peptide present in each fraction calculated from its radioactive counts. FT: flowthrough; W1-W4: washes; E: elution with EGTA. White bars: binding in the presence of Ca^{2+}; black bars: binding in the presence of EGTA.