# Comparison of mRNA-Display-Based Selections Using Synthetic Peptide and Natural Protein Libraries†

Bao-cheng Huang and Rihe Liu*

*School of Pharmacy and Carolina Center for Genome Sciences, The University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599*

*Received February 1, 2007; Revised Manuscript Received May 22, 2007*

ABSTRACT: 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. Compared to prior protein selection techniques, mRNA display can be used to select functional sequences from both long natural protein and short combinatorial peptide libraries with much higher complexities. To investigate the basic features and problems of using mRNA display in studying conditional protein−protein interactions, we compared the target-binding selections against calmodulin (CaM) using both a natural protein library and a combinatorial peptide library. The selections were efficient in both cases and required only two rounds to isolate numerous $Ca^{2+}$/CaM-binding natural proteins and synthetic peptides with a wide range of affinities. Many known and novel CaM-binding proteins were identified from the natural human protein library. More than 2000 CaM-binding peptides were selected from the combinatorial peptide library. Unlike sequences from prior CaM-binding selections that correlated poorly with naturally occurring proteins, synthetic peptides homologous to the $Ca^{2+}$/CaM-binding motifs in natural proteins were isolated. Interestingly, a large number of synthetic peptides that lack the conventional CaM-binding secondary structures bound to CaM tightly and specifically, suggesting the presence of other interaction modes between CaM and its downstream binding targets. Our results indicate that mRNA display is an ideal approach to the identification of $Ca^{2+}$-dependent protein−protein interactions, which are important in the regulation of numerous signaling pathways.

One of the most effective strategies to get a thorough understanding of a protein of interest is through mapping its conditional interactions with other proteins on a proteome-wide scale, while simultaneously searching the peptide sequence space for the binding specificity. The ideal method should allow the identification of target-interacting sequences from both natural protein and combinatorial peptide libraries under the conditions that are specific for such interactions. A number of approaches have been developed to address this problem. The yeast two-hybrid method has been widely used to isolate interacting sequences of a target protein from natural cDNA libraries (*1*, *2*). Phage display is another approach to the selection of sequences with desired functions, often from a short combinatorial peptide library (*3*). Unfortunately, most peptide or protein selection methods have limitations in addressing conditional protein−protein interactions.

mRNA display is a relatively new in vitro selection technique that allows the identification of conditional protein−protein interactions from both a natural protein library and a combinatorial peptide library (*4*−*6*). The central feature of this method is that the polypeptide chain is covalently linked to the 3′ end of its own mRNA. This is accomplished by synthesis and in vitro translation of an mRNA template with puromycin attached to its 3′ end via a short DNA linker. During in vitro translation, when the ribosome reaches the RNA−DNA junction and translation pauses, puromycin, an antibiotic that mimics the aminoacyl moiety of tRNA, enters the ribosome "A" site and accepts the nascent polypeptide by forming a peptide bond. This results in tethering of the nascent peptide to its own mRNA. Since the genotype coding sequence and the phenotype polypeptide sequence are united in a single molecule, mRNA display provides a powerful means for reading and amplifying a peptide or protein sequence after it has been selected on the basis of its function. Multiple rounds of selection and amplification can be performed, enabling enrichment of rare sequences with desired properties. 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 arbitrary and stringent conditions in the functional 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 $10^{12}$−$10^{13}$ unique sequences can be readily generated and selected, a few orders of magnitude higher than can be achieved using phage display. Therefore, both the likelihood of isolating rare sequences and the diversity of the sequences isolated in a given selection are significantly increased. It is of great interest to investigate the basic

† This work was supported by startup funds from the Carolina Center for Genome Sciences and School of Pharmacy at the University of North Carolina at Chapel Hill (to R.L.) and by a grant from the American Cancer Society (RSG-06-073-01-TBE) (to R.L.).

* To whom correspondence should be addressed. Tel: (919) 843-3635. Fax: (919) 968-0081. E-mail: rihe_liu@unc.edu.

10.1021/bi700220x CCC: $37.00 
Published on Web 08/09/2007

features and potential problems of using mRNA display in studying conditional protein−protein interactions. Ideally, this question should be addressed by comparing two selections against the same target using both a natural protein and combinatorial peptide libraries, so that the interacting partners in both the proteome of interest and the sequence space can be examined. The model target we chose to compare here is CaM,[1] which is considered as the primary transducer of $Ca^{2+}$ signals in eukaryotes. Although numerous CaM-binding proteins have been revealed, novel signaling pathways mediated by CaM are continuously identified (*6*, *7*). Like many other $Ca^{2+}$-sensing proteins, CaM contains EF-hand motifs as $Ca^{2+}$ binding sites and undergoes distinct conformational changes upon binding with $Ca^{2+}$. CaM is involved in a wide variety of biological processes through its interaction with many different protein targets, which come in various shapes, sizes, and sequences (*8*−*14*). Many $Ca^{2+}$-dependent CaM-binding proteins have one or more CaM-binding motifs that are characterized by a basic amphipathic helix with approximately 20 amino acids in length (*15*−*17*). On the basis of the positions of the conserved hydrophobic residues (Φ), three related $Ca^{2+}$/CaM-binding motifs have been grouped, namely 1-10, 1-14, and 1-16 motifs, which refer to groups of sequences whose key bulky hydrophobic residues are spaced 8, 12, and 14 residues apart, respectively. Another well-known CaM-binding motif is the IQ or IQ-like motif, which has a consensus sequence of (FILV)$QX_3$(RK)$GX_3$(RK)$X_2\Phi$ or (FILV)$QX_3$(RK), respectively (*18*, *19*). Due to the presence of numerous CaM-binding motifs, a robust selection strategy should allow the identification of peptides or proteins with all of the subclasses of CaM-binding motifs.

Phage display had been used to identify CaM-binding peptides from combinatorial peptide libraries. Dedman and co-workers reported the first CaM-binding screening using a phage library displaying random peptides 15 amino acids in length (*20*). Twenty-eight independent CaM-binding sequences were isolated in this screen, and all contained a tryptophan residue within the 15 amino acid insert. Kay and co-workers used a phage library displaying random peptides 26 amino acids in length for CaM-binding sequences (*21*). Twenty CaM-binding peptides were identified in this work, and 17 contained one of the three consensus sequence motifs. Interestingly, the sequences of the selected peptides from these two selections had little in common, which was attributed to different types of combinatorial peptide libraries used in the selection. It is now clear that phage display could preferentially select high-affinity interacting peptides at the expense of low-affinity binders, presumably due to biases against certain sequences at different steps (*22*). In the work reported by Nairn and co-workers, for example, two CaM-binding sequences (AWDTVRISFG and AWPSLQAIRG) comprised 80% of the pool after two rounds of affinity selection using a phage-displayed combinatorial peptide library (*23*). Another problem in phage-display-based selection is that the selected peptide sequences usually correlate poorly with naturally occurring proteins in GenBank. Indeed, none of the selected CaM-binding peptides was found to be homologous to natural CaM-binding proteins. We hypothesized that the observed poor correlation could be due to low complexity of the initial libraries. Accordingly, screening a synthetic peptide library with much higher complexity could result in better correlation. However, most in vitro selection techniques only allow screening of peptide libraries with less than $10^9$ sequences. Therefore, a novel strategy should be used to address the problem.

In this work, we describe the $Ca^{2+}$/CaM-binding selection using an mRNA-displayed combinatorial peptide library with about $10^{12}$ unique sequences. More than 2000 $Ca^{2+}$/CaM-binding peptides were identified, including several that were highly homologous to the CaM-binding motifs in naturally occurring proteins. We also compared the CaM-binding selection from a human protein domain library with that from the combinatorial peptide library. The selection was very efficient in both cases, and CaM-binding natural proteins and synthetic peptides with a wide range of affinities were enriched. Some unique features and potential problems of mRNA-display-based selection were revealed. Our results indicate that mRNA display is an ideal approach to the identification of $Ca^{2+}$-dependent protein−protein interactions, which are important in the regulation of numerous signaling pathways.

## EXPERIMENTAL PROCEDURES

*Construction and Generation of mRNA-Displayed Protein and Peptide Libraries*. The mRNA-displayed natural protein domain library was generated with a mixture of poly(A)$^+$ mRNAs from human brain, heart, spleen, thymus, and muscle as reported recently (*6*). For the combinatorial peptide library, a preselected, high-quality DNA library that codes for peptides with 20 consecutive random amino acids was used as templates for in vitro peptide synthesis (*24*). Each sequence in the DNA library contains a T7 RNA polymerase promoter, a TMV translation enhancer sequence, an N-terminal FLAG tag coding sequence, a random cassette encoding 20 consecutive random codons, and a $His_6$ tag coding sequence. The library was preselected using mRNA display to remove sequences with stop codons and frame shifts. Unlike prior CaM-binding selections in which all of the random residues were coded by an NNK codon (N = G, A, T, or C; K = G or T) (*20*, *21*, *23*), we started from a library that encoded short peptides with more balanced amino acid compositions (Supporting Information, Table 1). This was realized by 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 (*24*, *25*). The DNA library was in vitro transcribed using T7 RNA polymerase (Ambion, Austin, TX) to get mRNAs. The mRNA templates with puromycin at the 3′ ends were generated by cross-linking with an oligonucleotide containing a psoralen moiety and a puromycin residue at its 5′ and 3′ ends, respectively (*26*). In vitro translation was performed using rabbit reticulocyte lysate (Novagen, Madison, MI), and mRNA−protein fusion formation was accomplished under optimized conditions (*5*).

[1] Abbreviations: CaM, calmodulin; CAMK2D, calcium/calmodulin-dependent protein kinase II δ isoform; CBSPP, CaM-binding site prediction program; CDC5-L, *Schizosaccharomyces pombe* cell division cycle 5 homologue-like; GLUT4, glucose transporter member 4; $K_d$, binding affinity; MLCK, myosin light chain kinase; MYBL1, v-myb myeloblastosis viral oncogene homologue (avian)-like 1; Ni-NTA, $Ni^{2+}$-nitrilotriacetic acid; PPP3C, catalytic subunit of protein phosphatase 3 α isoform; snRNA, small nuclear RNA; TBC1D4, TBC (Tre-2, BUB2, CDC16) domain-containing family member 4; TMV, tobacco mosaic virus; TNT, coupled in vitro transcription and translation.

mRNA templates and mRNA−protein fusions were then isolated from the lysate using an oligo(dT) column, taking advantage of oligo(dA) residues at the puromycin-containing DNA linker [psoralen-$(ATAGCCGGTG)_2$-OMe-$dA_{15}$-CC-puromycin]. To remove secondary RNA structures that might interfere with the selection step, the fusion molecules were converted to a DNA/RNA hybrid by reverse transcription. The resulting mRNA-displayed peptide library was successively purified on the basis of each of the affinity tags at the N- and C-termini. More than 85% of the purified sequences had contiguous ORF without stop codons or frame shifts (*24*). On the basis of the large number of clones we have picked up from the initial library for sequencing analysis, none was found to be identical to others.

*Selection of $Ca^{2+}$-Dependent CaM-Binding Sequences from mRNA-Displayed Peptide or Protein Libraries*. The purified combinatorial peptide or natural protein library displayed on their own mRNAs (∼1.5−3 pmol each round) was first diluted in a selection buffer (50 mM Tris-HCl, pH 7.5, 150 mM KCl, 0.05% Tween-20, 1 mg/mL BSA, 5 mM 2-mercaptoethanol, and 0.5 mM $Ca^{2+}$) and passed through a precolumn with 100 $\mu$L of streptavidin beads (Pierce, Rockford, IL) to minimize the enrichment of matrix-binding sequences. The flow-through was incubated with 25 $\mu$g of biotinylated CaM (*6*) for 1 h at 4 °C. After binding, the mixture was incubated with 100 $\mu$L of streptavidin beads for another 0.5 h followed by loading to an empty column. Unbound and nonspecifically bound molecules were washed off the column using 18 and 40 column volumes of selection buffer for the natural protein library and combinatorial peptide library, respectively. Molecules that bound to CaM in a $Ca^{2+}$-dependent manner were eluted using the same buffer containing 2 mM EGTA. The selected molecules were PCR amplified for the next round of selection or cloned into a TOPO vector (Invitrogen, La Jolla, CA) for sequencing and analysis.

*Synthesis and Purification of Selected Peptides or Proteins*. To examine the CaM-binding properties of a large number of selected peptide sequences, a parallel CaM-binding assay was used. First, individually cloned sequences selected from round 1 or round 2 were PCR amplified and used as templates for a coupled in vitro transcription and translation (TNT) reaction in the presence of 10 $\mu$Ci of [$^{35}$S]methionine (Perkin-Elmer, Wellesley, MA) in a total volume of 25 $\mu$L for 90 min at 30 °C. Expressed peptides or proteins were purified using Ni-NTA agarose (Qiagen, Valencia, CA), either individually or in a 96-well microplate format. To purify peptides or proteins using 96-well plates, 25 $\mu$L of the TNT reaction mixture was diluted into 100 $\mu$L of Ni-NTA binding buffer (50 mM HEPES, pH 7.4, 300 mM NaCl, 0.05% Tween 20) and incubated with 25 $\mu$L of Ni-NTA agarose at room temperature for 1 h with gentle shaking. The mixtures were then transferred to a 96-well filter plate. The supernatant was removed by applying vacuum in a manifold for 3 min. The Ni-NTA resin in each well was washed with 800 $\mu$L of binding buffer for four times. Bound peptides or proteins were then eluted twice each with 50 $\mu$L of elution buffer (50 mM HEPES, pH 7.4, 300 mM NaCl, 250 mM imidazole, 0.05% Tween 20). Peptides or proteins thus purified were passed through a Zeba desalt spin column (Pierce) to further remove residual free [$^{35}$S]methionine before being used in the binding assay.

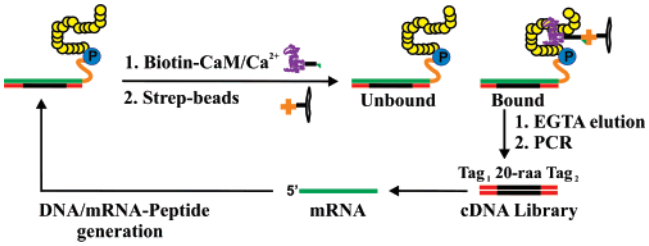


FIGURE 1: Selection of $Ca^{2+}$-dependent CaM-binding sequences using an mRNA-displayed combinatorial peptide library: mRNA, green; DNA, red; protein, yellow circles; puromycin, blue circle labeled with P; streptavidin beads, orange cross. The 20 random amino acids (raa) are shown as a black bar at the DNA level, flanked with two affinity tags at both sides.

*In Vitro CaM-Binding Assays*. An aliquot of purified peptide or protein was incubated with an appropriate amount of biotinylated CaM in a well of 96-well microplate in a CaM-binding buffer (25 mM Tris-HCl, pH 8.0, 150 mM NaCl, 1 mg/mL BSA, 5 mM 2-mercaptoethanol) at room temperature for 60 min. The binding was performed in the presence of 1 mM $Ca^{2+}$ or 2 mM EGTA. The CaM−peptide or CaM−protein complexes were then captured by adding 20 $\mu$L of a streptavidin−agarose slurry on the same plate, and the mixture was incubated at room temperature for 30 min with gentle mixing. After removal of the supernatant, the beads were washed four times each with 100 $\mu$L of corresponding binding buffer to wash away unbound molecules. Peptides or proteins captured by CaM were then gently released from the beads by chelating $Ca^{2+}$ using 100 $\mu$L of elution buffer (CaM-binding buffer supplemented with 2 mM EGTA). Each fraction, including flow-through, washes, eluates, and agarose beads, was recovered. The binding percentage was determined from the scintillation counts of each fraction.

*Binding Affinity of Selected CaM-Binding Peptides or Proteins*. To determine the binding affinity, biotinylated CaM with different final concentrations (from 5 nM to 1.0 $\mu$M) was used to interact with an appropriate amount of purified, [$^{35}$S]methionine-labeled peptides or proteins in a polypropylene 96-well microplate. The binding assay was performed in the presence of 1 mM $Ca^{2+}$. The radioactive counts of the flow-through, washes, eluates, and agarose beads were counted. The binding assay was performed twice, and the averaged data were fitted to a binding curve using GraphPad Prism software.

## RESULTS

*CaM-Binding Sequences Were Selected Very Efficiently from both Natural Protein and Combinatorial Peptide Libraries*. The length of the protein portion from the ORFs of the natural protein library was in the range of 80−200 amino acids. Typically, 1.5 pmol (∼$10^{12}$ molecules) of mRNA-displayed protein sequences was generated and used in the binding selection. Therefore, each protein and its fragments should have numerous copies. The combinatorial peptide library has a complexity of approximately $5 \times 10^{12}$, which is more than 3 orders of magnitude higher than those used in most phage-display-based selections (*24*). We then performed $Ca^{2+}$-dependent CaM-binding selection using either an mRNA-displayed natural protein domain library or a combinatorial peptide library. Figure 1 illustrates the

Table 1: Sequences and CaM-Binding Properties of Several CaM-Binding Protein Fragments Isolated from the mRNA-Displayed Human Protein Domain Library

| S no. | isolated protein fragment and its mapped CaM-binding motif[a] | potential CaM-binding motif predicted by CBSPP[b] | % binding with $Ca^{2+}$ (elution/on beads) | % binding with EGTA (elution/on beads) | $K_d$ (nM) |
|---|---|---|---|---|---|
| P1 | CaMK2D; D248−R390 from BAD92525; [327]RRKLKGAILTTMLATR | contains a database basic 1-5-10 motif | 27/13 | 0.7/2.6 | 32 ± 6 |
| P2 | PPP3CA; E381−R478 from NP_000935; [393]KEVIRNKIRAIGKMARVFSV | contains a database 1-5-8-14 motif | 54/9.5 | 0.1/0.1 | 19 ± 3 |
| P3 | MLCK; M1736−G1823 from NP_444253; [1742]RRKWQKTGNAVRAIGRLSSM | contains a database basic 1-8-14 motif | 42/12.7 | 1.8/1.4 | 26 ± 4 |
| P4 | TBC1D4; D826−G925 from NP_055647; [838]ELRSLWRKAIHQQILLLR | | 9.8/32.8 | 2.9/15.5 | 16 ± 3 |
| P5 | CDC5-L; A22−E115 from NP_001244; [31]WSRIASLLHRKSAKQCKAR | | 31.7/14.6 | 0.9/2.2 | 37 ± 6 |
| P6 | MYBL1; K30−E191 from XP_034274; [151]YEAHKRLGNRWAEIAKLLP | | 25.8/26.4 | 3.4/13.5 | 32 ± 5 |

[a] This column lists the name of the protein, the region of the fragment isolated from the selection, and the CaM-binding motif included in this region predicted using a CaM-binding site prediction program (CBSPP). The underlined residues are the key bulky hydrophobic residues as reported in the literature. [b] This column lists the type of the CaM-binding motifs predicted by CBSPP. Each of these motifs refers to a group of sequences whose key bulky hydrophobic residues are spaced 8, 12, and 14 residues apart, respectively, based on the classification detailed in http://calcium.uhnres.utoronto.ca/ctdb/ctdb/3cln-big.html.

selection scheme. In both selections, the signal was close to background in round 1 but rapidly increased to approximately 8% in round 2 (Supporting Information, Figure 1). Among the 269 annotated or hypothetic proteins isolated in round 1 from the natural protein library, 28 known and 99 novel CaM-binding proteins were confirmed using the CaM-binding assay (*6*). Interestingly, almost all of the sequences from the round 2 pool bound to CaM in a $Ca^{2+}$-dependent manner. Out of 120 sequences we thoroughly analyzed from round 2, 72 were from 12 classes of well-known CaM-binding proteins, including different isoforms (α, *β*, *γ*, and *δ*) of $Ca^{2+}$/CaM-dependent protein kinases, different isoforms (α, *β*, and *γ*) of $Ca^{2+}$/CaM-dependent protein phosphatases, $Ca^{2+}$ transporting ATPase, α II spectrin, phosphoinositide 3-kinase, myosin heavy polypeptides, voltage-dependent sodium channel II, DEAD-box protein, cyclic nucleotide phosphodiesterase, myosin light chain kinase, dystrophin, and nitric oxide synthase. The remaining 48 sequences were from potential novel CaM-binding proteins, including TBC1 family members. Other well-known CaM-binding proteins were found when more clones were picked up for analysis (data not shown).

We picked up several selected sequences to detail their CaM-binding affinity and specificity. These sequences include selected protein fragments originating from well-known CaM-binding proteins (CAMK2D, PPP3C, and MLCK), together with a few originating from potential novel CaM-binding proteins (TBC1D4, CDC5-L, and MYBL1). As shown in Table 1 and Figure 2A (also Supporting Information, Figure 2A), these protein fragments bound to CaM tightly and specifically. The binding was highly dependent on $Ca^{2+}$, as it was significantly reduced when $Ca^{2+}$ was chelated. The binding affinities were in the range of 16−37 nM, and those of known CaM-binding proteins were close to that reported in the literature (*27*, *28*).

We attributed the efficient enrichment of CaM-binding proteins from the natural protein library to its relatively lower complexity ($<10^7$ unique protein sequences) and the presence of a large number of CaM-binding proteins in the human proteome. Since the combinatorial peptide library is totally unnatural and contains more than $5 \times 10^{12}$ unique sequences, we postulated that it should take many more rounds of selection before the pool is dominated by $Ca^{2+}$-dependent CaM-binding peptides. Surprisingly, a similar selection profile was obtained (Supporting Information, Figure 1). Unlike in the selection using the natural protein library in

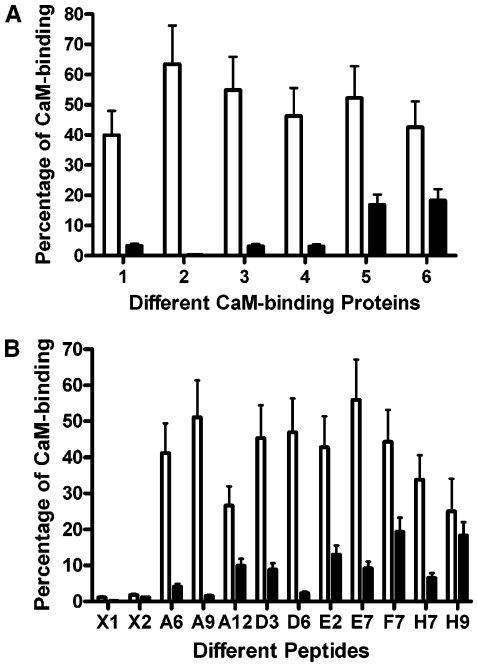


FIGURE 2: In vitro binding analysis of several selected natural protein fragments (A) or peptides (B) with biotinylated CaM. Radiolabeled protein fragments or peptides were generated by TNT and purified using Ni-NTA. The purified TNT product was incubated with biotinylated CaM in the presence of 1 mM $CaCl_2$ or 2 mM EGTA. The complex was captured with streptavidin−agarose beads. After extensive wash, the bound molecules were eluted using a buffer containing 2 mM EGTA. The *Y* axis is the percentage (%) of a protein or peptide present in the EGTA elution and beads calculated from its radioactive counts. Panel A: 1, CAMK2D; 2, PPP3C; 3, MLCK; 4, CDC5-L; 5, TBC1D4; 6, MYBL1. In panel B, X1 and X2 are negative controls, and others are different round 2 peptides. Only the combined percentage from elution and those that remained on beads are shown. White bars: binding in the presence of $Ca^{2+}$. Black bars: binding in the presence of EGTA. The percentages of all of the fractions are given in Supporting Information, Figure 2.

Table 2: Some Selected CaM-Binding Fragments Originated from Known CaM-Binding Proteins Whose CaM-Binding Motifs Have Been Reported

| protein | protein fragments isolated from our selection and its CaM-binding motif reported in the literature[a] |
|---|---|
| **1-10 Subclasses** | |
| basic 1-5-10 motif | |
| CaM kinase II α subunit (NP_741960) | H273−K328; [296]RRKLKGAILTTMLATR |
| CaM kinase II β (NP_001211) | E244−A364; [297]RRKLKGAILTTMLATR |
| CaM kinase II δ (BAD92525) | E316−R390; [327]RRKLKGAILTTMLATR |
| 1-5-10 motif | |
| CaM kinase I (NP_003647) | I230−S324; [303]WKQAFNAAVVRHMRKLQL |
| CaM kinase kinase α (NP_757343) | V371−E470; [441]IPSWTTVILVKSMLRKRSFGNPF |
| heat-shock 90 kDa α (NP_001017963) | Q610−K668; [626]NSAFVERLRKHGLEVI |
| heat-shock 90 kDa β (NP_001004082) | V381−P516; [496]NSAFVERVRKRGFEVV |
| heat-shock 70 kDa protein 8 (NP_006588) | G203−R301; [257]KRAVRRLRTACERAKRTLSSS |
| utrophin (NP_009055) | E687−L764; [690]KKFDAISAELLNWILKWKTA |
| **1-14 Subclasses** | |
| 1-8-14 motif | |
| nonerythroid α-spectrin (NP_003118) | M1178−L1256; [1189]ASPWKSARLMVHAVATFNSI |
| basic 1-8-14 motif | |
| myosin light chain kinase (NP_444253) | K1733−S1824; [1742]RRKWQKTGNAVRAIGRLSSM |
| 1-5-8-14 motif | |
| calcineurin A α (NP_000935) | G386−E472; [393]KEVIRNKIRAIGKMARVFSV |
| calcineurin A β (NP_066955) | E399−S455; [402]KEIIRNKIRAIGKMARVFSV |
| calcineurin A γ (NP_005596) | V364−G469; [387]KEIIRNKIRAIGKMARVFSI |
| titin (NP_596869) | S31513−G31622; [31527]HTLIKKDLNMVVSAARISCG |
| plasma membrane $Ca^{2+}$-ATPase 1b (NP_001673) | G1075−E1168; [1104]QILWFRGLNRIQTQIRVVNA |
| TBC1 domain member 4 (NP_055647) | K831−A884; [838]ELRSLWRKAIHQQILLLR |
| **1-16 Subclass** | |
| DEAH box polypeptide 57 (NP_945314) | V422−Q542; [497]VNLKKKISKRYDWQAKSV |
| **IQ or IQ-like Subclasses** | |
| myosin heavy chain 1 (NP_005954) | Q1015−E1125; [1104]LGMQLQKKIKELQARIEELE |
| dynein heavy polypeptide 1 (NP_001367) | H403−D534; [426]EYEKLQVLLRDIVKRKREEN |
| mitochondrial $H^+$-ATP synthase, F1-1 (NP_005165) | P17−K79; [35]RRLKSIKNIQKITKSMKMVAAA |

[a] This column lists the region of the fragment isolated from the selection and the included CaM-binding motif reported in the literature. The underlined residues are the key bulky hydrophobic residues as reported in the literature. Each of these motifs refers to a group of sequences whose key bulky hydrophobic residues are spaced 8, 12, and 14 residues apart, respectively, based on the classification detailed in http://calcium.uhnres.utoronto.ca/ctdb/ctdb/3cln-big.html.

which the round 2 pool was dominated by a limited number of known CaM-binding proteins, the complexity of the round 2 pool from the combinatorial peptide library was still very diverse. By detailed analysis of 300 clones, none of them was found to be identical to others. Using a high throughput CaM-binding assay, we were able to examine the conditional interaction of a large number of selected peptide sequences with CaM. Almost all of the round 2 peptide sequences we picked up for analysis bound to CaM, suggesting that the enrichment is as efficient as that from the natural protein library.

*Selected CaM-Binding Proteins from the Natural Protein Library Cover Almost All Known Subclasses of CaM-Binding Motifs.* The most characteristic feature of the selected natural protein fragments is that all those from known CaM-binding proteins contain the corresponding CaM-binding motifs reported in the literature. Table 2 lists some selected CaM-binding fragments originating from known CaM-binding proteins whose CaM-binding motifs could be found in the literature (*17*). Significantly, all known subclasses of CaM-binding motifs can be found from these proteins, including 1-8-14, 1-5-8-14, basic 1-8-14, 1-5-10, basic 1-5-10, 1-16, and IQ or IQ-like motifs.

The selected CaM-binding sequences from round 1 were diverse with approximately 125 unique proteins, whereas those from round 2 were more focused. These CaM-binding proteins include proteins in many different signaling pathways. In addition to known CaM-binding proteins that are involved in protein phosphorylation, dephosphorylation, second-message generation, and cytoskeletal elements regulation, novel CaM-binding proteins that belong to other protein families were also identified, including numerous DEAD/H box proteins, ribosomal proteins, splicing factors, and proteasome subunits (*6*).

*Analysis of CaM-Binding Motifs on Selected Natural CaM-Binding Proteins.* One big advantage of the mRNA-display approach is that the CaM-binding motif of a selected protein can be readily mapped by localizing the shortest overlapping region of multiple selected fragments originating from the same parental protein. Using this method, CaM-binding motifs from a number of well-characterized CaM-binding proteins were mapped and found to be the same as reported in the literature (Table 2). This approach is particularly useful in determining the binding motifs on novel CaM-binding proteins with long cDNAs. In the round 2 pool from the human protein domain library, for example, we found that 9 out of 120 sequences were originated from human TBC1 domain family member 4, a large protein with 1298 amino acids in length. TBC1D4 contains a functional Rab GTPase-activating protein domain and was found to be a substrate of AKT kinase involved in the regulation of GLUT4 translocation (*29*). The 9 selected TBC1D4 sequences have different lengths, but all share a common region with 50 amino acids, corresponding to K835−A884 in TBC1D4

(Supporting Information, Figure 3). Recently, the GTPase activating protein domain region of TBC1D4 was found to interact with $Ca^{2+}$/CaM through the yeast two-hybrid method (*30*), and the CaM-binding motif was mapped to a small domain just amino terminal to the GAP domain using the conventional protein truncation method. The CaM-binding motif thus mapped is just 27 amino acids shorter than the shortest overlapping region we identified by aligning the selected sequences, indicating that our method is particularly useful for mapping the potential CaM-binding motifs on very large proteins.

*Most Selected CaM-Binding Peptides from the Combinatorial Peptide Library Bound to CaM in a $Ca^{2+}$-Dependent Manner.* The selection was also efficient when an mRNA-displayed combinatorial peptide library was used. Almost all of the peptides we picked up for analysis from the round 2 pool bound to CaM in a $Ca^{2+}$-dependent manner. The complexity of the round 2 pool was still very high. We estimated by large-scale sequencing that more than 2000 unique sequences could be present in this pool (data not shown). The percentage of binding ranged from 8% to 50%, indicating that peptides with different binding affinities could be enriched. Since the number of CaM-binding peptides isolated was far more than what we were able to characterize, we picked up 41 peptide sequences for detail analysis, together with a number of positive and negative controls. For negative controls, we hypothesized that most of the peptide sequences before selection should not bind to CaM. Indeed, none of the negative control sequences (X1 to X6) we randomly picked up from the initial peptide library bound to CaM (data not shown). The CaM binding of these 41 peptide sequences was detailed in the presence or absence of $Ca^{2+}$. Significantly, all 41 peptides bound to CaM in a $Ca^{2+}$-dependent manner (Table 3), consistent with the specific elution condition applied in the selection. We then picked up ten sequences that bound to CaM tightly for quantitative analysis (Table 3, column 6, Figure 2B, and Supporting Information, Figure 2B). The $Ca^{2+}$/CaM-binding affinities of these peptides vary from 16 to 64 nM, comparable to those of CaM-binding motifs from well-characterized CaM-binding proteins. In addition to these tight binders, we found that peptide sequences with binding affinities in the range of the micromolar scale were also isolated in round 2 (data not shown). These results indicate that mRNA display can be used to efficiently isolate both high-affinity and low-affinity ligands from a highly complex peptide library.

*Most Selected CaM-Binding Peptides from the Combinatorial Peptide Library Have Nonconventional Secondary Structures.* The selected CaM-binding peptides are very diverse in sequence. However, the first two residues in the random region are limited to only several amino acids. This is because these two residues were coded by a restriction enzyme cleavage site (NACTGN) for the purpose of library construction (*24*). Other than that, no obvious homology was detected among these 41 peptides using various bioinformatic tools. To investigate whether the selected peptides contain the conventional positively charged amphiphilic helical or IQ-motif secondary structures that are often adopted by natural CaM-binding motifs, we systematically analyzed these sequences using the CBSPP program (http://calcium.uhnres.utoronto.ca/ctdb/ctdb/3cln-big.html). Fifteen out of 41 peptide sequences contain the conventional CaM-binding secondary structures. Interestingly, these sequences cover most subclasses of CaM-binding motifs, including basic 1-5-10 motif (B7 and C6), 1-5-10 motif (H8, D6, and F7), 1-14 motif (C8, D5, and G7), 1-8-14 motif (D3 and E2), 1-16 motif (D9 and F6), and IQ motif (A6 and H7). Significantly, 26 out of 41 sequences do not contain the conventional CaM-binding secondary structures. A1, A3, A9, C10, C12, D8, D10, E7, F3, G11, and H3, for example, are not predicted to have the conventional CaM-binding secondary structures. Nevertheless, they still bound to CaM tightly in a $Ca^{2+}$-dependent manner with a binding percentage over 35%. These results indicate that many sequences other than the conventional CaM-binding motifs are able to interact with CaM in a $Ca^{2+}$-dependent manner. Since a large number of such sequences were isolated, we were not able to characterize their secondary structures using experimental approaches. However, all of these peptides are relatively short with 20 amino acids in the random region. Therefore, their secondary structures predicted by the Network Protein Sequence Analysis (NPS@) and CBSPP programs (Table 3) are still valuable.

*Several Selected CaM-Binding Peptides from the Combinatorial Peptide Library Correlate with the CaM-Binding Motifs in Naturally Occurring Proteins.* In prior selections of CaM-binding peptides from phage-displayed combinatorial peptide libraries, none of the selected CaM-binding sequences was found to match any CaM-binding proteins in the protein database (*20*, *21*, *23*). Since our library has much higher complexity, it is likely that such correlation exists. To address this question, we systematically blasted the characterized 41 CaM-binding motifs in the protein database. Although most matches were to nonrelevant proteins, at least 3 out of 41 peptides, namely, A12, A6, and D6, were homologous to the CaM-binding motifs in some natural CaM-binding proteins.

The first such peptide is A12, which has a $Ca^{2+}$/CaM-binding affinity of approximately 41 nM. This peptide contains a CaM-binding site (RRASSKLTKK) that is homologous to the potential CaM-binding motifs in 12 different natural proteins (Table 4). The matched CaM-binding motifs on these proteins are very different among themselves, although they match to the selected A12 sequence at different residues. When analyzed using the CBSPP program, it appears that all of the 12 proteins could interact with CaM using the motifs that are homologous to the selected A12 peptide. We were able to generate the full-length zebrafish G protein-coupled receptor 19 (gpr19) by TNT and found that it indeed bound to CaM in a $Ca^{2+}$-dependent manner, although the binding was relatively weak (data not shown).

A6 is an IQ-motif-containing peptide that bound to $Ca^{2+}$/CaM with an affinity of approximately 19 nM. Four natural proteins were found to contain a potential CaM-binding IQ motif homologous to that found in A6 (Table 5), including human transglutaminase 5 (TGM5), human IQ motif containing GTPase activating protein 2 (IQGAP2), human voltage-gated sodium channel type III-α (SCN3A), and zebrafish similar to myosin XV (LOC559486). IQGAP2 and SCN3A are well-characterized CaM-binding proteins. The

Table 3: Sequences and CaM-Binding Properties of Some CaM-Binding Peptides Isolated from the mRNA-Displayed Combinatorial Peptide Library

| S# | Isolated sequence and its secondary structure predicted by NPS@[a] | Binding score predicted by CBSPP (highest 9; lowest 0) | % binding with $Ca^{2+}$ (elution/ on beads) | % binding with EGTA (elution/ on beads) | $K_{d:}$ (nM) |
|---|---|---|---|---|---|
| | **Negative controls randomly picked up before selection** | | | | |
| **X1** | **DWNTVMVPLIMVPSLHNSDSY**<br>C C C EE E E EEE E EC CCC CCC CC | 0 | 0.7/0.4 | 0.1/0.6 | |
| **X2** | **HWQLYLYDEGPQGMKFLMLNY**<br>C E E E EE EC C CC C C C E EE E E C C | 0 | 0.8/1.0 | 0.2/0.9 | |
| **X3** | **YWGNQRYRLNSNDEDMNKTAN**<br>C C C C CC E E EC CCC CC C C C CC C | 0 | 0.3/0.3 | 0.2/0.8 | |
| **X4** | **HCLLTLTETNPAQEWDLWWAN**<br>C E EE EE EC CCC CC C C H E H HC C | 0 | 0.8/0.8 | 0.4/0.9 | |
| **X5** | **DWQMTAPLSMGPAKGPERMPH**<br>C C C C CC CCC C CC CC CC CC C C C | 0 | 0.6/0.3 | 0.2/0.5 | |
| **X6** | **YCRNPTSLQVRHDVPVIAPKN**<br>C CC CCCC E E EC CC CC ECCC CC | 0 | 0.6/0.6 | 0.5/0.5 | |
| | **41 peptide sequences picked up for binding analysis from round 2 pool** | | | | |
| **A1** | **NWENKSDVLNKMLHLRLITN**<br>C C CC H HHHHH H HH HHHHHCC | 0 | 33.5/3.4 | 0.4/1.5 | |
| **A3** | **HWEKIRDNVIHKSDRSYYQKH**<br>C C CC CCC CE EE CCC CCC C CC C | 0 | 41.8/6.7 | 0.2/2.3 | |
| **A6** | **NWKSIIQRNLRWNKFKRFYQD**<br>C C HHHHHHHHH H HH HHHH CCC | 0.5; contains a potential IQ motif | 22.8/18.4 | 0.9/3.9 | 19 ± 5 |
| **A9** | **NWNILRQEVMKMGPAKDTVRN**<br>C C CHHH HHH H C C CC CC C CC CC | 0 | 48.3/2.8 | 0.2/1.3 | 24 ± 5 |
| **A12** | **NWNTVWSAYRRASSKLTKKTY**<br>C C C HH H HHHH HHHHHH CC CCC | 9; contains a database CBM | 21.4/5.2 | 0.6/9.2 | 41 ± 7 |
| **B1** | **NWNRWRWIYLAQKNHKECLKN**<br>C C C C H H HHHH HHH HH HH HHC C | 0 | 16.7/4.1 | 1.5/3.1 | |
| **B7** | **YWKRYRLMALKANFGSRYSRN**<br>C C H H HHH H HH HHHC CCC CCC C | 1; basic 1-5-10 motif | 14.3/11.8 | 1.2/10.1 | |
| **C6** | **NCPSTRKRWIVYLYRGRLINN**<br>CCCC CC C E E EEE E E E CC ECCC | 2; basic 1-5-10 motif | 7.9/4.7 | 0.8/4.5 | |
| **C8** | **NWNRVRRYCKCKYTLVAKNNY**<br>C C CC C E EC E E E E E EE E EC C C C | 1; 1-14 motif | 33.5/5.1 | 0.5/6.4 | |
| **C10** | **NWYTVKQMMIKDRRRSQAMYN**<br>C C H HHH H H H HHHH HHHH HH CC | 0 | 35.7/6.4 | 0.7/8.6 | |
| **C12** | **NWSDIRMHGLMRRREKNVLNH**<br>C C CCHH H HHH H H HHHH H HCC C | 0 | 35.0/3.6 | 0.2/5.8 | |
| **D3** | **NWVKLRQRVTLAKRVAVNLNY**<br>C C C HH HHH HHH HH HHH HHCC C | 1; 1-8-14 motif | 40.0/5.4 | 0.7/8.2 | 18 ± 4 |
| **D5** | **HWNKVASERINKMRRKAILNY**<br>C C C H HHHHHHHH H HH HHHHCC | 2; 1-14 motif | 29.8/5.1 | 0.6/8.6 | |
| **D6** | **HWNAIKNVVHSQRMTKKIAMN**<br>C C H HHHHH HH HHHH H HHHH CC | 9; contains a database 1-5-10 CBM | 38.2/8.8 | 0.5/1.9 | 25 ± 4 |
| **D8** | **NWEILRSKYNSHMTKSNVFTY**<br>C C EEE EEC C CCC C CC CC E ECC | 0 | 38.7/3.2 | 0.3/2.2 | |
| **D9** | **YWSELRRKKWWKVWKLSSCPH**<br>C H HHHHH HH H E E E E E CC CCCC | 1; 1-14 or 1-16 motif | 10.8/6.7 | 0.6/5.2 | |
| **D10** | **YWNHELRVRLINRIKMTN**<br>C C C HHH HH HHHHHEE C CC | 0 | 34.8/4.1 | 0.6/3.4 | |

Table 3 (Continued)

| | | | | | |
|---|---|---|---|---|---|
| **D12** | **HWGIIRQKIVVAHDIFQCKDY**<br>C C EEEEE EE HHH HHHHHHC C C | 0 | 25.8/3.4 | 0.3/0.9 | |
| **E2** | **NWINNVRLRIHTKRWLLKSNH**<br>C CHH HHHH HHHHHH E E H HCC C | 1;<br>1-8-14 motif | 31.1/11.7 | 1.7/11.3 | 20 ± 5 |
| **E7** | **NWLRLVPRIKALNKVQVKNHN**<br>C H HHH HHHHHHHHH E E E C CC C | 0 | 50.1/5.8 | 0.3/8.9 | 16 ± 3 |
| **E8** | **NCKMQPQNWYHVYRMSRLVKN**<br>C CC C C CC C H H E E HH H HHHCC C | 0 | 21.5/3.6 | 0.4/4.4 | |
| **E11** | **DCEYRDKVTLFNLVRLVMTKN**<br>C CCH HH HHHH HHHHH HE E CCC | 0 | 21.4/3.4 | 0.5/3.2 | |
| **F1** | **YWSNRVTQSIKARYVIDSWQD**<br>C C CH CC HHHHEE E EEHHH C C C | 0 | 22.4/2.7 | 0.8/1.1 | |
| **F3** | **NWHKVFIRRQSKKLVYNTIKN**<br>C C C E E EEEC C CC E E EE E ECCC | 0 | 33.6/9.7 | 0.5/15.0 | |
| **F4** | **YCHKYTVANESHWSKIRLKMY**<br>C CC C E EE C CCCH H HHEE E CC C | 0 | 20.0/4.0 | 0.5/2.3 | |
| **F6** | **YWVFLKKKMKLENKCVRVVKN**<br>C H H HHHH H H HCC CC E E E E E CC | 2;<br>1-14 or 1-16 motif | 21.8/5.8 | 1.1/6.5 | |
| **F7** | **DWNMLKVKLYALRVRRRRMAN**<br>C C C H HH H HHH HHHH HHH H H CC | 1;<br>1-5-10 motif | 24.8/19.5 | 1.5/17.9 | 30 ± 5 |
| **G2** | **HWMRTSQRVRVNNAFHKYMGY**<br>C C C C CCH HE HHH HH HHH HC CC | 0 | 30.5/6.5 | 1.0/11.6 | |
| **G5** | **NCEVLKTQRWRKVLQQHIIRH**<br>C CC CHC HH H H HH HH HHHHECC | 0 | 22.5/5.5 | 0.7/6.5 | |
| **G7** | **HWSKTQGGNKRWRMIGAVVAH**<br>C C CC C C C CC C E E E EE E EE E C C | 2;<br>1-14 motif | 35.9/5.7 | 0.5/6.7 | |
| **G9** | **NWDKVRSTFKKCHSIVIFKRN**<br>C C HH HHHHHC CC E EEEEEE C C | 0 | 31.0/6.5 | 0.2/4.9 | |
| **G10** | **YWSLIVSKLRRRKVMNDPSTY**<br>C HHHEH HHHCH CC E C CC CCCC | 0 | 23.4/7.6 | 0.7/7.5 | |
| **G11** | **HWDRVRLTMLRSRLKDDKKKH**<br>C C H HHH HH HH HHHHH HH C CC C | 0 | 42.8/9.2 | 0.5/13.7 | |
| **G12** | **NCLKQRLLRTPYLMMSRAVTH**<br>C CH HHHH HHC HHHH H HHH CC C | 0 | 25.0/2.5 | 0.3/2.3 | |
| **H3** | **NWRKAINLVRKWRNNDDPNKD**<br>C C H HHHHH HHH H C CC CC CC C C | 0 | 39.3/8.5 | 0.5/3.9 | |
| **H4** | **HWAMTRWHILANNVMNRRTCD**<br>C C C HH H HHHH HHH HH C CC C CC | 0 | 26.1/3.6 | 0.3/3.2 | |
| **H7** | **YWQRLQSMLKRVDPRPYRVRD**<br>C H H HH HHHH H HCC CCC CC C CC | 2;<br>contains a potential IQ motif | 28.3/5.6 | 0.5/6.0 | 64 ± 11 |
| **H8** | **HWNKLSYTVRIGEIKRYVWRN**<br>C C C C E EEE E EECCHH H EE E C C | 2;<br>1-5-10 motif | 23.4/3.4 | 0.4/5.0 | |
| **H9** | **YWRQKAKDDLMVRRLRRAVKH**<br>C C C C C CC HH HH H HHH HHH HC C | 0 | 29.0/16.0 | 1.3/17.0 | 28 ± 6 |
| **H11** | **NWRNLRLKWRRLNSVHRH**<br>C C HHHH HH H HH HHHHH C C | 0 | 23.4/24.6 | 1.2/20.8 | |
| **H12** | **NCLKQRLLRTPYLMMSRAVTH**<br>C CH HHH HHH CHHH H H HHH CC C | 0 | 16.8/16.5 | 1.2/10.8 | |

[a] The secondary structures were predicted using the Network Protein Sequence Analysis (NPS@) program (http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa_nn.html). C, E, and H stand for random coil, extended strand, and α helix, respectively. Scoring is based on the evaluation criteria of the program, including hydropathy, α-helical propensity, residue charge, hydrophobic residue content, and helical class, for determining putative CaM-binding sites as detailed in http://calcium.uhnres.utoronto.ca/ctdb/ctdb/3cln-big.html. The higher the score, the more likely it contains a conventional CaM-binding motif.

Table 4: Natural Proteins That Contain a Motif Homologous to the CaM-Binding A12 Peptide Selected from the Combinatorial Peptide Library

| protein name | accession number | motif homologous to A12: RRASSKLTKKTY[a] | CaM-binding motif predicted by CBSPP program |
|---|---|---|---|
| human erythrocyte membrane protein band 4.1-like 2 | NP_001422 (EPB41L2) | $^{207}$ASQKVTKK$^{214}$ | $^{207}$ASQKVTKK$^{214}$ |
| human DNA (cytosine-5-)-methyltransferase 3-b | NP_008823 (DNMT3B) | $^{49}$RRSSSRLSK$^{57}$ | $^{51}$SSSRLSKR$^{58}$ |
| human oxysterol binding protein-like 8 | NP_065892 (OSBPL8) | $^{112}$SSKLTTK$^{118}$ | $^{111}$TSSKLTKK$^{118}$<br>$^{572}$YAHCKGILYGTMTL$^{585}$ |
| human inner centromere protein antigens 135/155 kDa | NP_001035784 (INCENP) | $^{333}$RRASRRLAKKT$^{343}$ | $^{70}$RISYVQDENRDPIRRRLSRR$^{89}$<br>$^{321}$KYSLVAKQES VVRRASRRLA KKT$^{343}$<br>$^{578}$LRK VLQARERVEQ MKEEKKK$^{697}$ |
| human establishment of cohesion 1 homologue | NP_443143 (ESCO1) | $^{12}$SSKVTKK$^{18}$ | $^{11}$NSSKVTKK$^{18}$ |
| mouse similar to spindlin-like protein 2 | XP_621036 (SPIN-2) | $^{137}$SSKLTKK$^{143}$ | $^{136}$ESSKLTKK$^{143}$ |
| mouse schlafen 1 | NP_035537 (Slfn1) | $^{181}$SKLTKK$^{186}$ | $^{179}$KQSKLTKK$^{186}$ |
| mouse calmodulin binding protein 1 | NP_033921 | $^{151}$ASSKHTKRT$^{159}$ | $^{1469}$QAQKLYKR$^{1476}$ |
| mouse α-kinase 1 | XP_929293 (Alpk1) | $^{271}$ASSKMSKK$^{278}$ | $^{271}$ASSKMSKK$^{278}$ |
| zebrafish G protein-coupled receptor 19 | NP_957288 (gpr19) | $^{360}$ASSRMAKKNY$^{369}$ | $^{132}$PAAACRILR F$^{141}$<br>$^{360}$ASSRMAKK$^{367}$ |
| *Arabidopsis* unknown protein | NP_001032066 | $^{153}$RRVSSRLNKK$^{162}$ | $^{155}$VSSRLNKK$^{162}$ |
| *Arabidopsis* ATP binding/ ATP-dependent helicase | NP_201200 (AT5G63950) | $^{625}$ATSKLSKK$^{632}$ | $^{625}$ATSKLSKK$^{632}$ |

[a] The residues that match the corresponding positions on A12 are underlined.

Table 5: Natural Proteins That Contain an IQ Motif Homologous to the CaM-Binding A6 Peptide Selected from the Combinatorial Peptide Library

| protein name | accession number | motif homologous to A6: IQRNLRWNKFK | CaM-binding IQ motif predicted by CBSPP Program |
|---|---|---|---|
| human transglutaminase 5 | NP_963925 (TGM5) | $^{696}$IQANMRSNKFK$^{706}$ | $^{696}$IQANMRSNKFK$^{706}$ |
| human IQ motif containing GTPase activating protein 2 | NP_006624 (IQGAP2) | $^{1374}$IQRNLR$^{1379}$ | $^{697}$IQAFWKGYKQRKEY$^{710}$<br>$^{727}$IQSWFRMATARKSY$^{740}$<br>$^{757}$IQSLLRANKARDDY$^{770}$<br>$^{1374}$IQRNLR$^{1379}$ |
| human sodium channel, voltage-gated, type III, α | NP_008853 (SCN3A) | $^{1906}$IIQRNFR$^{1912}$ | $^{1906}$IIQRNFRCYLLKQRL$^{1920}$ |
| zebrafish similar to myosin XV | XP_687921 (LOC559486) | $^{1325}$IQRNIR$^{1330}$ | $^{187}$LQGQKKMLKSKILQK$^{201}$<br>$^{1325}$IQRNIRGFICRRNF$^{1328}$<br>$^{1348}$IQSHIRGHQARKYF$^{1361}$ |

IQ motif in A6 has 72% identity to a motif (I696−K706) in human transglutaminase 5, which is predicted to be a CaM-binding protein by the CBSPP program. These results strongly suggest that the selected CaM-binding peptides from the combinatorial peptide library could be related to CaM-binding motifs in natural proteins, and novel CaM-binding proteins might be identified using this approach.

## DISCUSSION

The conditions that modulate numerous protein−protein interactions are often difficult to manipulate in vivo. Recently, the yeast two-hybrid system has been used to construct comprehensive protein−protein interaction maps of several model organisms, including *Saccharomyces cerevisiae*, *Drosophila melanogaster*, and *Caenorhabditis elegans* (*31*−*33*). In each of these genome-wide protein−protein interaction maps, however, only very few $Ca^{2+}$/CaM-binding proteins were identified. One possible explanation is that the tightly regulated $Ca^{2+}$ level in the yeast nucleus does not allow monitoring of strict $Ca^{2+}$-dependent interactions (*34*, *35*). These results imply that the yeast two-hybrid method has limitations in finding cellular protein−protein interactions that are dependent on $Ca^{2+}$ signals. Such conditional protein−protein interactions are more likely to be identified using in vitro methods. The proteome chip is probably the most high throughput approach to addressing the problem. Indeed, the yeast proteome chip has been successfully used to identify CaM-binding proteins in yeast (*7*). However, the preparation of a proteome chip from any multicellular organism is a major undertaking. In addition to phage display, ribosome display is another powerful in vitro technique to study protein−protein interactions. However, it requires metal ions to maintain its tertiary structure and therefore cannot be applied to selections that rely on chelating agents. Unlike most of the prior methods, mRNA display can be used to display both long natural protein and short combinatorial peptide libraries with more than $10^{12}$ unique sequences. Although it still suffers from sparse sampling, the library complexity thus generated is a few magnitudes higher than most of the prior methods. Therefore, it is more likely to isolate peptide sequences that are homologous to naturally occurring proteins.

It is of great interest to ask why prior phage-display-based selections resulted in only limited numbers of CaM-binding

sequences. The most likely explanation is that the diversity of the sequences displayed on phage was at least 3 orders of magnitude lower than that in the mRNA display. The second possibility is that the selection pressures were raised against certain peptide sequences. A large number of CaM-binding peptides are positively charged amphipathic helices, which could affect protein expression, transport, membrane insertion, toxicity on host cells, phage viability, and infectivity. In mRNA-display-based selection, however, all of the steps are totally in vitro; therefore, most selection biases present in phage display are greatly reduced. Compared to the combinatorial peptide libraries encoded by the same NNK codon, the amino acid compositions we used in the peptide library were more balanced, particularly for those residues that are highly enriched in CaM-binding motifs. Whether this feature of the library affected the selection efficiency is not clear.

Our results demonstrate that the mRNA-displayed combinatorial peptide library can be used to isolate target-binding peptides that are not only useful for specificity studies but also correlate with natural proteins. Indeed, the results from the identification of c-myc binding peptides using an mRNA-displayed peptide library support this observation (*36*). Since few mRNA-display-based selections have been published, it is not yet known whether it is also true for other types of mRNA-display experiments with combinatorial peptide libraries. The successful enrichment of $Ca^{2+}$/CaM-binding sequences from both natural and synthetic libraries does not necessarily mean that other protein−protein interactions can be identified with similar efficiency. Several potential problems of mRNA-display-based selection should be made aware of. First, the selection conditions should be carefully designed to minimize the enrichment of nonrelevant sequences. It appeared that preselection procedures to remove matrix-binding and nonspecific sequences were helpful but not very effective. It was found that the most effective approach was through specific elution to enrich sequences with desired properties. In $Ca^{2+}$/CaM-binding selection, a low concentration of EGTA was used to chelate $Ca^{2+}$, which resulted in conformation change of CaM and therefore specific release of bound sequences. For other protein−protein interactions, gentle and competitive elution should be applied whenever possible.

The wash step before elution is also important. While a stringent wash effectively removes sequences that bind to the target nonspecifically, it might also result in loss of the desired sequences, if they bind to the target relatively weakly. It was found that while the 15 column volume wash allowed us to isolate CaM-binding sequences with binding affinities in the range from 10 nM to 2 $\mu$M, most selected sequences had affinities ranging from 5 to 300 nM when a 40 column volume wash was applied. These results indicate that a wash with 12−15 column volumes can remove most nonspecific sequences, while still retain sequences that weakly bind to the target. It appeared that CaM-binding peptides with a wide range of affinities were successfully isolated from an mRNA-displayed peptide library. We attributed this to a combination of less biased selection pressures, flexibility of in vitro procedures, and high complexity of the initial library. It is of great interest to ask whether there is a minimum dissociation constant necessary to recover a ligand through this technique. In the selection we reported here, CaM-binding peptides with less than 5 $\mu$M could still be isolated. However, the amplification-based selection is a competitive process between weak and tight binders and is also highly dependent on the interaction between the target and its binding ligands. Therefore, the minimum dissociation constant around 5 $\mu$M is for the current $Ca^{2+}$/CaM-binding selection, but not necessary for the selections using other targets.

The number of selection cycles is probably one of the most critical parameters. The percentage of selected sequences that bound to $Ca^{2+}$/CaM from the natural protein library was approximately 50% and 100% in round 1 and round 2 pools, respectively. When a combinatorial peptide library was used, few sequences from round 1 bound to CaM, presumably due to the much higher complexity of the initial library. However, almost all of the sequences we tested from round 2 bound to CaM tightly and specifically. We attributed this efficient enrichment to a combination of very stringent wash (40 column volumes) and gentle and specific elution by EDTA. Typically, fewer rounds of selection should be performed if the target of interest has many binding partners. On the other hand, selection should continue until the final pool is dominated by few tight binders, if the target protein has a limited number of binding partners or the aim is to find sequences that bind to the target with high affinities. We found that one of the most effective approaches was to perform two to four rounds of selection, followed by determining which selected pool should be used for further analysis. The pool that has limited diversity but has not yet been dominated by just a few sequences is then chosen for detail characterization.

The efficient selection of a large number of diverse $Ca^{2+}$/CaM-binding sequences from both natural protein and combinatorial peptide libraries indicates that mRNA display is a robust system for the investigation of conditional protein−protein interactions. Our results clearly demonstrate that the mRNA-displayed synthetic peptide library can be used to isolate target-binding peptides that are not only useful for specificity studies but also correlate with natural proteins.

## ACKNOWLEDGMENT

We thank Drs. Biao Dong and C. Alexander Valencia for assistance in experimental techniques.

## SUPPORTING INFORMATION AVAILABLE

One table and three figures as described in the text. This material is available free of charge via the Internet at http://pubs.acs.org.

## REFERENCES

1. Prichard, L., Deloulme, J. C., and Storm, D. R. (1999) Interactions between neurogranin and calmodulin in vivo, *J. Biol. Chem. 274*, 7689−7694.
2. Brill, S., Li, S., Lyman, C. W., Church, D. M., Wasmuth, J. J., Weissbach, L., Bernards, A., and Snijders, A. J. (1996) The Ras GTPase-activating-protein-related human protein IQGAP2 harbors a potential actin binding domain and interacts with calmodulin and Rho family GTPases, *Mol. Cell. Biol. 16*, 4869−4878.
3. Kay, B. K., Kasanov, J., and Yamabhai, M. (2001) Screening phage-displayed combinatorial peptide libraries, *Methods 24*, 240−246.
4. Roberts, R. W., and Szostak, J. W. (1997) RNA-peptide fusions for the in vitro selection of peptides and proteins, *Proc. Natl. Acad. Sci. U.S.A. 94*, 12297−12302.

5. Liu, R., Barrick, J. E., Szostak, J. W., and Roberts, R. W. (2000) Optimized synthesis of RNA-protein fusions for in vitro protein selection, *Methods Enzymol. 318*, 268−293.
6. Shen, X., Valencia, C. A., Szostak, J., Dong, B., and Liu, R. (2005) Scanning the human proteome for calmodulin-binding proteins, *Proc. Natl. Acad. Sci. U.S.A. 102*, 5969−5974.
7. Zhu, H., Bilgin, M., Bangham, R., Hall, D., Casamayor, A., Bertone, P., Lan, N., Jansen, R., Bidlingmaier, S., Houfek, T., Mitchell, T., Miller, P., Dean, R. A., Gerstein, M., and Snyder, M. (2001) Global analysis of protein activities using proteome chips, *Science 293*, 2101−2105.
8. Ghosh, A., and Greenberg, M. E. (1995) Calcium signaling in neurons: molecular mechanisms and cellular consequences, *Science 268*, 239−247.
9. Clapham, D. E. (1995) Calcium signaling, *Cell 80*, 259−268.
10. Bootman, M. D., and Berridge, M. J. (1995) The elemental principles of calcium signaling, *Cell 83*, 675−678.
11. Berridge, M. J., Bootman, M. D., and Roderick, H. L. (2003) Calcium signalling: dynamics, homeostasis and remodelling, *Nat. Rev. Mol. Cell Biol. 4*, 517−529.
12. Carafoli, E., Santella, L., Branca, D., and Brini, M. (2001) Generation, control, and processing of cellular calcium signals, *Crit. Rev. Biochem. Mol. Biol. 36*, 107−260.
13. Orrenius, S., Zhivotovsky, B., and Nicotera, P. (2003) Regulation of cell death: the calcium-apoptosis link, *Nat. Rev. Mol. Cell Biol. 4*, 552−565.
14. Ikura, M. (1996) Calcium binding and conformational response in EF-hand proteins, *Trends Biochem. Sci. 21*, 14−17.
15. Rhoads, A. R., and Friedberg, F. (1997) Sequence motifs for calmodulin recognition, *FASEB J. 11*, 331−340.
16. Crivici, A., and Ikura, M. (1995) Molecular and structural basis of target recognition by calmodulin, *Annu. Rev. Biophys. Biomol. Struct. 24*, 85−116.
17. Yap, K. L., Kim, J., Truong, K., Sherman, M., Yuan, T., and Ikura, M. (2000) Calmodulin target database, *J. Struct. Funct. Genomics 1*, 8−14.
18. Houdusse, A., Silver, M., and Cohen, C. (1996) A model of Ca(2+)-free calmodulin binding to unconventional myosins reveals how calmodulin acts as a regulatory switch, *Structure 4*, 1475−1490.
19. Munshi, H. G., Burks, D. J., Joyal, J. L., White, M. F., and Sacks, D. B. (1996) $Ca^{2+}$ regulates calmodulin binding to IQ motifs in IRS-1, *Biochemistry 35*, 15883−15889.
20. Dedman, J. R., Kaetzel, M. A., Chan, H. C., Nelson, D. J., and Jamieson, G. A., Jr. (1993) Selection of targeted biological modifiers from a bacteriophage library of random peptides. The identification of novel calmodulin regulatory peptides, *J. Biol. Chem. 268*, 23025−23030.
21. Adey, N. B., and Kay, B. K. (1996) Identification of calmodulin-binding peptide consensus sequences from a phage-displayed random peptide library, *Gene 169*, 133−134.
22. Zozulya, S., Lioubin, M., Hill, R. J., Abram, C., and Gishizky, M. L. (1999) Mapping signal transduction pathways by phage display, *Nat. Biotechnol. 17*, 1193−1198.
23. Nevalainen, L. T., Aoyama, T., Ikura, M., Crivici, A., Yan, H., Chua, N. H., and Nairn, A. C. (1997) Characterization of novel calmodulin-binding peptides with distinct inhibitory effects on calmodulin-dependent enzymes, *Biochem. J. 321* (Part 1), 107−115.
24. Cho, G., Keefe, A. D., Liu, R., Wilson, D. S., and Szostak, J. W. (2000) Constructing high complexity synthetic libraries of long ORFs using in vitro selection, *J. Mol. Biol. 297*, 309−319.
25. Wolf, E., and Kim, P. S. (1999) Combinatorial codons: a computer program to approximate amino acid probabilities with biased nucleotide usage, *Protein Sci. 8*, 680−688.
26. Kurz, M., Gu, K., and Lohse, P. A. (2000) Psoralen photo-crosslinked mRNA-puromycin conjugates: a novel template for the rapid and facile preparation of mRNA-protein fusions, *Nucleic Acids Res. 28*, E83.
27. Hubbard, M. J., and Klee, C. B. (1987) Calmodulin binding by calcineurin. Ligand-induced renaturation of protein immobilized on nitrocellulose, *J. Biol. Chem. 262*, 15062−15070.
28. Lukas, T. J., Burgess, W. H., Prendergast, F. G., Lau, W., and Watterson, D. M. (1986) Calmodulin binding domains: characterization of a phosphorylation and calmodulin binding site from myosin light chain kinase, *Biochemistry 25*, 1458−1464.
29. Kane, S., Sano, H., Liu, S. C., Asara, J. M., Lane, W. S., Garner, C. C., and Lienhard, G. E. (2002) A method to identify serine kinase substrates. Akt phosphorylates a novel adipocyte protein with a Rab GTPase-activating protein (GAP) domain, *J. Biol. Chem. 277*, 22115−22118 (Epub 22002 May 22116).
30. Kane, S., and Lienhard, G. E. (2005) Calmodulin binds to the Rab GTPase activating protein required for insulin-stimulated GLUT4 translocation, *Biochem. Biophys. Res. Commun. 335*, 175−180.
31. Uetz, P., Giot, L., Cagney, G., Mansfield, T. A., Judson, R. S., Knight, J. R., Lockshon, D., Narayan, V., Srinivasan, M., Pochart, P., Qureshi-Emili, A., Li, Y., Godwin, B., Conover, D., Kalbfleisch, T., Vijayadamodar, G., Yang, M., Johnston, M., Fields, S., and Rothberg, J. M. (2000) A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*, *Nature 403*, 623−627.
32. Giot, L., Bader, J. S., Brouwer, C., Chaudhuri, A., Kuang, B., Li, Y., Hao, Y. L., Ooi, C. E., Godwin, B., Vitols, E., Vijayadamodar, G., Pochart, P., Machineni, H., Welsh, M., Kong, Y., Zerhusen, B., Malcolm, R., Varrone, Z., Collis, A., Minto, M., Burgess, S., McDaniel, L., Stimpson, E., Spriggs, F., Williams, J., Neurath, K., Ioime, N., Agee, M., Voss, E., Furtak, K., Renzulli, R., Aanensen, N., Carrolla, S., Bickelhaupt, E., Lazovatsky, Y., DaSilva, A., Zhong, J., Stanyon, C. A., Finley, R. L., Jr., White, K. P., Braverman, M., Jarvie, T., Gold, S., Leach, M., Knight, J., Shimkets, R. A., McKenna, M. P., Chant, J., and Rothberg, J. M. (2003) A protein interaction map of *Drosophila melanogaster*, *Science 302*, 1727−1736.
33. Li, S., Armstrong, C. M., Bertin, N., Ge, H., Milstein, S., Boxem, M., Vidalain, P. O., Han, J. D., Chesneau, A., Hao, T., Goldberg, D. S., Li, N., Martinez, M., Rual, J. F., Lamesch, P., Xu, L., Tewari, M., Wong, S. L., Zhang, L. V., Berriz, G. F., Jacotot, L., Vaglio, P., Reboul, J., Hirozane-Kishikawa, T., Li, Q., Gabel, H. W., Elewa, A., Baumgartner, B., Rose, D. J., Yu, H., Bosak, S., Sequerra, R., Fraser, A., Mango, S. E., Saxton, W. M., Strome, S., Van Den, Heuvel, S., Piano, F., Vandenhaute, J., Sardet, C., Gerstein, M., Doucette-Stamm, L., Gunsalus, K. C., Harper, J. W., Cusick, M. E., Roth, F. P., Hill, D. E., and Vidal, M. (2004) A map of the interactome network of the metazoan *C. elegans*, *Science 303*, 540−543.
34. Yang, Q., O'Hanlon, D., Heizmann, C. W., and Marks, A. (1999) Demonstration of heterodimer formation between S100B and S100A6 in the yeast two-hybrid system and human melanoma, *Exp. Cell Res. 246*, 501−509.
35. Deloulme, J. C., Assard, N., Mbele, G. O., Mangin, C., Kuwano, R., and Baudier, J. (2000) S100A6 and S100A11 are specific targets of the calcium- and zinc-binding S100B protein in vivo, *J. Biol. Chem. 275*, 35302−35310.
36. Baggio, R., Burgstaller, P., Hale, S. P., Putney, A. R., Lane, M., Lipovsek, D., Wright, M. C., Roberts, R. W., Liu, R., Szostak, J. W., and Wagner, R. W. (2002) Identification of epitope-like consensus motifs using mRNA display, *J. Mol. Recognit. 15*, 126−134.

BI700220X