In Vitro Selection of Proteins with Desired Characteristics Using mRNA-display

C. Alexander Valencia[a,b,*], **Jianwei Zou**[c,d], and **Rihe Liu**[c,d,*]

[a]Division of Human Genetics, Cincinnati Children's Hospital Medical Center, Cincinnati, OH 45229, U.S.A.

[b]Department of Pediatrics, University of Cincinnati School of Medicine, Cincinnati, OH 45229, U.S.A.

[c]Division of Chemical Biology and Medicinal Chemistry, UNC Eshelman School of Pharmacy, University of North Carolina, Chapel Hill, North Carolina 27599-7568, U.S.A.

[d]Carolina Center for Genome Sciences, University of North Carolina, Chapel Hill, North Carolina 27599-7568, U.S.A.

Abstract

mRNA-display is an amplification-based, iterative rounds of *in vitro* protein selection technique that circumvents a number of difficulties associated with yeast two-hybrid and phage display. Because of the covalent linkage between the genotype and the phenotype, mRNA-display provides a powerful means for reading and amplifying a peptide or protein sequence after it has been selected from a library with very high diversity. The purpose of this article is to provide a summary of the field and practical framework of mRNA-display-based selections. We summarize the advantages and limitations of selections using mRNA-display as well as the recent applications, namely, the identification of novel affinity reagents, target-binding partners, and enzyme substrates from synthetic peptide or natural proteome libraries. Practically, we provide a detailed procedure for performing mRNA-display-based selections with the aim of identifying protease substrates and binding partners of a target protein. Furthermore, we describe how to confirm the function of the selected protein sequences by biochemical assays and bioinformatic tools.

Keywords

1. Introduction

1.1. Principle of mRNA-display selection

mRNA-display permits the efficient identification of peptide sequences, via *in vitro* selection, with desired characteristics from a natural protein or a combinatorial peptide library [1–3]. The key feature of the technique is the peptide covalent linkage to its own mRNA (at 3' end). 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

---

*To whom correspondence should be addressed. rihe_liu@unc.edu, alexander.valencia@cchmc.org.

"A" site and accepts the nascent polypeptide by forming a peptide bond. This results in tethering the nascent peptide to its own mRNA (Fig. 1). Since the mRNA (genotype) and peptide (phenotype) sequences are covalently linked in a single molecule, mRNA-display permits effective amplification of a peptide or protein sequence following selection based on its function. Iterative rounds of selection and amplification enable enrichment of rare sequences with desired characteristics. In theory, any peptide or protein sequence with desired characteristics may be selected by mRNA-display. Thus far, the applications of mRNA-display notably include, but are not limited to, mapping of the protein-protein and nucleic acid-protein interaction networks, identification of drug-binding targets, elucidation of the enzyme-substrate interactions, the development of novel and improvement of existing affinity reagents [4–11]. Fig. 1 shows the general procedure for the selection of the downstream substrates of a protease of interest (option A) and for the identification of the binding partners of a target of interest (option B).

Here, we briefly summarize the major developments, advantages, limitations and applications of mRNA-display, provide a practical step-by-step procedure of the selection process from mRNA-displayed protein libraries, focusing on the selection for binding partners and protease substrates, and conclude by pointing out the future direction of the field.

1.2. Advantage of mRNA-display selection

Compared to related protein selection methods, mRNA-display has notable advantages. First, the covalent linkage of the genotype to the phenotype allows rapid identification of isolated proteins with wide variety of desired characteristics following a functional selection. Second, the library complexity, with as many as 10^{12}–10^{14} unique sequences, close to that of the RNA or DNA pools, can be readily generated and selected, several orders of magnitude higher than phage display can reach. In this case, the main advantages are that the likelihood of selecting rare sequences and that the diversity of the selected sequences isolated is significantly increased. Third, the covalent linkage between the protein and its mRNA permits the utilization of any arbitrary condition in a functional selection with tunable stringency without compromising the recovery of the selected sequences. The success of a selection is highly dependent on the development of a selection scheme that allows for specific enrichment of sequences with desired characteristics, while minimizing nonspecific sequences. Many of the current technologies, including ribosome-display, are limited in terms of the binding conditions that may be used in a given screen. For example, ribosome-display with the fragile noncovalent conjugation between genotype and phenotype requires mild selection conditions. In contrast, the phenotype–genotype linkage in mRNA-display is covalent and highly stable, making it possible to perform many types of selections under very stringent conditions. mRNA-display is an entirely *in vitro* selection technique that exploits cell-free translation systems for the efficient generation of peptide sequences. Since the expression and selection steps are both performed *in vitro* and transformation is not necessary, the library complexity size is determined by the genetic material, which can be scaled up in the laboratory to increase library diversity. The amount of DNA oligonucleotides that can be synthesized, the volume of PCR that can be performed to generate the cDNA library and the volume of *in vitro* translation reaction to express the protein library can also be scaled up and tailored to increase the diversity [12]. As such, peptide or protein libraries with 10^{12}–10^{14} unique sequences can be obtained. In contrast, *in vivo* protein selection methods are limited to small, low complexity libraries due to inefficient transformations or transfections of the starting cDNA library into the organism of choice. The phage display is typically around 10^9, whereas the library size for cell-based selections, such as the yeast two-hybrid system, bacteria and yeast surface display, is

approximately 10^6 [13,14]. In the IVC (*in vitro* compartmentalization) system using water-in-oil micro-droplets, the diversity of the library is typically limited to less than 10^9 [15].

For totally or partially randomized peptide or protein libraries, the much higher diversity provides the opportunity for exploration of the sequence space to a greater depth for the search of target-binding partners. Practically, this larger library complexity raises the probability of finding target-binding sequences with improved affinity and specificity. According to Lancet and coworkers' model of olfactory receptors and antibodies, the affinity of the best binding partners increased approximately 300-fold when the library size increased 10,000-times [12,16]. Compared with totally synthetic protein libraries, natural proteome libraries have much lower diversities. In theory, all of the full-length proteins in the proteome can be displayed [17,18]. However, this task requires the amplification of all genes of interest using gene-specific primers, which significantly reduces the throughput and increases the costs of the selection. To circumvent this issue, one approach is to construct a cDNA library from isolated mRNAs using random reverse-transcription primers to cover all the open reading frame regions in the genome. Therefore, a much larger cDNA library, a feature of mRNA-display, is necessary to completely cover all protein sequences with desired characteristics. Additionally, numerous copies of each gene and its individual domains can be found in an mRNA-displayed proteome library. Therefore, this added complexity increases the possibility of finding rare and diversity sequences in a given selection.

In selection methods, the protein expression can be significantly influenced by the environment where the library generated. Being a totally *in vitro* method, mRNA-display permits the generation of fusion proteins from any candidate libraries, with the random regions present in isolation or within the context of the fusion. In an mRNA-display-based selection, the proteins are typically expressed in cell-free eukaryotic translation systems, such as the rabbit reticulocyte lysate, allowing the synthesis of proteins with reasonable post-translational modifications and correct folding [19]. In contrast, protein expression of phage display relies on using the translational machinery in bacteria, whereas, yeast two-hybrid is limited to yeast cells, thus, some sequences are poorly expressed [20,21]. In ribosome display, the proteins are displayed on very large ribosomes; hence there would be unpredictable interactions between the huge ribosome complexes with the target. The mRNA-display cell-free expression system offers the generation of protein molecules with presumably reduced bias with a large complexity.

One of the major challenges of most protein selection platforms is that some abundant sequences, either specific or nonspecific binders, may dominate the pool as selection proceeds from one round to the next. Experimentally, in selections using unnormalized natural cDNA libraries, the most abundant mRNA species could be more than thousands of times higher than the least abundant ones [22]. In mRNA-display-based selection, abundant sequences that obscure the discovery of rare, potentially interesting sequences, can be efficiently subtracted at the mRNA level because the genotype of each molecule in the selected pool is covalently linked with its own protein or peptide sequence [8,9,23]. This unique feature significantly increases the chance of discovering rare sequences.

Another advantage of mRNA-display over other methods is the use of long proteins for selection. In an mRNA-displayed protein library, the protein length is typically between 50–300 amino acids. It is possible to generate larger proteins by DNA fractionation and purification of the cDNA library using agarose gels. Inarguably, peptides or proteins of this size (5.5–33 kDa) are much more likely than short ones to adopt native conformations, allowing the identification of residues or motifs that play a role in mediating protein–protein interactions or enzyme-substrate recognitions [9,24].

Random mutagenesis, by PCR-based methods such as error-prone PCR, DNA shuffling, random insertion and deletion, and random insertional-deletional strand exchange, is widely utilized to introduce mutations in directed evolution for the isolation of sequences with desirable characteristics [25]. Due to the necessity to perform ligation and transformation of PCR products into cells, the use of mutagenesis approaches is limited in i*n vivo* screening methods. mRNA-display uses PCR to amplify the cDNA following each round of selection and as such is readily compatible with PCR-based mutagenesis techniques.

1.3. Limitations of mRNA-display selection

In the previous section mRNA-display's powerful features were described. However, the method has some limitations. Even though large proteins may be displayed at lower efficiency, mRNA-display works better when the displayed proteins size is less than 300 residues [26]. Moreover, protein interactions that depend on the formation of complexes with their binding partners are difficult to address via mRNA-display. Due to the limited expression of membrane-bound proteins by *in vitro* translation systems, mRNA-display cannot be utilized to address membrane protein related questions [27]. mRNA-display–based selection is relatively difficult to perform because of the manipulation of protein, DNA and RNA in an RNAse-free environment and requires extensive expertise [26]. Another concern is whether the covalently attached mRNA portion interferes with the function of the displayed protein or with the interaction between the target molecule and the displayed protein. The interference of interaction is possible when the mRNA is present as the flexible single-stranded form, particularly for proteins that nonspecifically bind to nucleic acids. In addition, the mRNA portion of the fusion molecules is highly negatively charged, which might interfere in a selection where the target is highly positively charged [28]. Furthermore, some protein sequences may not be efficiently enriched because the biases of mRNA-display depend on several factors, namely, the abundance in the initial mRNA library, the overall efficiency of amplification, the efficiency of protein expression, formation of mRNA-protein fusion and its interaction with the target.

To continually improve this technology several attempts have been made to address the limitations of mRNA-display. To improve the stability of the fusion molecules, the single-stranded mRNA of the fusion can be converted to an mRNA/DNA hybrid after the fusion formation. A variant of mRNA-display, called cDNA-display, has been developed [29–31]. Selections using mRNA-display can also be performed by using microfluidic systems that provide more efficient separation and purification [32]. However, the cost for higher enrichment efficiency is the loss of the ability to obtain binding partners with low affinity, which could be very useful for many applications.

1.4. Types of biological and biochemical questions that can be addressed by mRNA-display

1.4.1. Applications using synthetic libraries—mRNA-display can be employed to select peptides or protein sequences with desired features from synthetic combinatorial libraries that contain varying length of totally or partially randomized amino acids, to address a number of biological problems, including the discovery of binding partners for both proteins and small molecules, and the discovery of novel enzymatic activities from the peptide sequence space. Specifically, synthetic libraries have been used to identify peptides that bind to streptavidin, the small molecule ATP, GPCR (G protein-coupled receptors) Methuselah (Mth), calmodulin (CaM) and thrombin [33–37]. For example, in an effort to search for ATP binding peptides, Keefe and coworkers showed that an *in vitro* selection of an mRNA-displayed peptide library, composed of peptides containing 80 contiguous random amino acids, yielded four new ATP binding proteins that appear to be unrelated to each other or to anything found in the current databases of biological proteins [34]. Ja and

coworkers demonstrated that mRNA-display can be used to identify peptides that bind to Methuselah (Mth), a class B GPCR, with high affinity (K_d = 15–30 nM) from a 39-residue synthetic library with 27 totally randomized amino acids. Significantly, the optimized peptides could function as antagonists, and *Drosophila melanogaster* expressing these Mth antagonistic peptides could extend their life span presumably by binding to the N-terminal ectodomain of Mth and inhibiting the Mth signaling pathway [35]. mRNA-display has been utilized to synthesize and select peptides with many unnatural amino acids [38]. In particular, using a reconstituted *Escherichia coli* translation system, Szostak and coworkers demonstrated that the reprogrammed genetic code could synthesize peptides with up to 13 different unnatural amino acids [39–41]. Unnatural peptide libraries of 10^{14} unique members may be synthesized via mRNA-display using this translation system.

In addition to synthetic peptide libraries containing totally randomized cassettes, various protein domain libraries with randomized binding loops have also been used for mRNA-display-based selections in an effort to generate target-binding affinity molecules. These protein domains include the 10th fibronectin type III domain (FN3), single chain antibody fragment (scFv), and zinc finger-containing retinoid-X-receptor (RXR) scaffold. Xu and co-workers created an FN3-based domain library by randomizing the three loops, including BC, DE, and FG, that could be involved in target interaction and used it for the selection of TNF-alpha-binding FN3 domains that bind TNF-alpha with low pM affinities [42]. By coupling with error-prone DNA shuffling, Fukuda and co-workers demonstrated the selection and affinity improvement of fluorescein-binding sequences from an mRNA-displayed single-chain antibody fragment (scFv) library [11]. Similarly, Cho and Szostak identified protein domains that specifically bound to ATP with moderate to high affinities, but not GTP, from a protein domain library based on the DNA binding domain of human retinoid-X-receptor (RXR) in which the two closely juxtaposed recognition loops were extensive randomized [43]. To inquire the amino acid sequence space, Seelig and Szostak used an mRNA-displayed RXR scaffold library with very high diversity to evolve synthetic proteins with genuinely novel RNA ligase activities [44]. It was shown that new enzymatic activities can be created *de novo* without the need for prior mechanistic information by selection from such naive protein domain library. Potentially, novel enzymes that catalyze other biochemical reactions may be discovered in this manner [44].

1.4.2. Applications using natural proteome libraries—In addition to synthetic libraries generated from randomization of amino acids, natural libraries derived from the mRNAs of tissues have also been used for mRNA- display-based selection to understand biological problems such as drug-protein, protein-protein and nucleic acid-protein interactions. The large library size achievable through mRNA-display allows for a comprehensive coverage of potential targets in a proteome of interest and therefore creation of network maps important for understanding the relevant signaling pathways. In the application of drug-protein interactions, McPherson and coworkers demonstrated that FKBP12 (FK506 binding protein 12 kDa), the physiological binding partner of immunosuppressive drug FK506, could be readily identified using FK506-biotin as the bait, from an mRNA-displayed human proteome library constructed from human liver, kidney, and bone marrow transcripts [6]. Hammond and coworkers used mRNA-display to isolate the binding partners of the anti-apoptotic protein Bcl-XL from a mixture of four libraries derived from human kidney, liver, bone marrow, and brain tissue, each barcoded with an 8-nucleotide tag within the 5'-UTR that can be deconvoluted with library-specific primers [4]. The selection resulted in the identification of three members of the Bcl-2 family known to bind to Bcl-XL and several novel Bcl-XL binding partners. Using mRNA-display-based selections, the binding partners of the basic leucine zipper domain (amino acids 167–319) of the transcription factor Jun have been described [7]. Sixteen novel Jun-associated protein candidates, in addition to four known interactors, were identified. Shen and coworkers

scanned the human and *C. elegans* proteomes for Ca^{2+}/CaM-binding proteins using an mRNA-display proteome library constructed from mRNAs of the brain, heart, spleen, thymus, and muscle tissues of human and of the adult worm, respectively [9,24]. A large number of both known and previously uncharacterized proteins that interact with CaM in a Ca^{2+}-dependent manner were identified. From an mRNA-displayed mouse brain proteome library, Tateyama and co-workers demonstrated the binding partners of the TPA-responsive element (TRE) can be efficiently identified [10]. These studies demonstrate that mRNA-display is a robust approach for the comprehensive mapping of the protein-protein and nucleic acid-protein interaction networks at a proteome-wide scale.

mRNA-display has also been used to study the challenging enzyme-substrate interactions. Cujec and coworkers investigated the substrate specificity of the kinase v-ABL using mRNA-displayed random peptide and human bone marrow libraries [5]. Ju and coworkers developed a high-throughput screen platform based on mRNA-display for the identification of caspase family member specific substrates on a proteome-wide scale [8]. The method can be widely applied for efficient and systematic identification of the family member specific natural substrate repertoire of any caspase in an organism of interest, in addition to that of numerous other proteases with high specificity.

2. Reagent and equipment list

Oligo name	Sequence (5′ to 3′)	Purpose
Random primer	TTNNNNNN	First-strand cDNA library synthesis
EcoRI/Hind III linker	CTTGAATTCAAGC	Cassette ligation (5′ phosphorylation)
Left cassette	TTCTTATACGACTCACTATAGGGACAATTACT ATTTACAATTACAATGTCCGGCCTGAACGACA TCTTTGAGGCTCAGAAAATCGAATGGCACGA AGGTTCTGGTGGGAATTCATATTG	Cassette ligation
Right cassette	TCCGTCGACAAGCTTGGGCGGAAGCGACTAC AAGGACGATGACGACAAGGGCGGCAGCCATC ACCACCATCACCATATGGGAATGTCTGGATCT GGCACCGGCTAT	Cassette ligation
Forward library primer	TTCTTATACGACTCACTATAGGGACAATTACT ATTTACAATTACAATG	Library amplification
Reverse library primer	ATAGCCGGTGCCAGATCCAGACATTCCCATAT G	Library amplification
Reverse transcription primer	TTTTTTTTTTTNNCCAGATCCAGACATTCCCAT	Reverse transcription

Tris-HCl (VWR, cat. no. EM-9210)

HEPES (Gibco, cat. no. 15630-106)

Glycine (VWR, cat. no. EM-4810)

First-strand buffer (5×). First-strand buffer is prepared with 250 mM Tris-HCl, pH 8.3, 15 mM $MgCl_2$ and 375 mM KCl. The buffer is stable at $-20°C$ for at least 3 months.

Second-strand buffer (5×). Second-strand buffer is prepared with 200 mM Tris-HCl, pH 7.5, 22 mM $MgCl_2$ and 425 mM KCl. The buffer is stable at $-20°C$ for at least 3 months.

T4 DNA polymerase buffer (10×). To prepare T4 DNA polymerase buffer, mix 100 mM Tris-HCl, pH 7.9, 500 mM NaCl, 100 mM $MgCl_2$ and 10 mM DTT. The buffer is

stable at − 20°C for at least 1 month. Avoid excessive freeze-thaw cycles to prevent oxidation of DTT.

T4 DNA ligation buffer (10×). To prepare T4 DNA ligation buffer, mix 500 mM Tris-HCl, pH 7.6, 100 mM $MgCl_2$ and 100 mM DTT. The buffer is stable at −20°C for at least 1 month. Avoid excessive freeze-thaw cycles to prevent oxidation of DTT.

RNA polymerase buffer (10×). RNA polymerase buffer consists of 400 mM Tris-HCl, pH 7.9, 60 mM $MgCl_2$ and 20 mM spermidine. DTT is freshly added prior to the reaction to a final concentration of 10 mM. The buffer is stable at − 20 °C for at least 1 month.

Oligo(dT) binding buffer (1×). Oligo(dT) binding buffer consists of 100 mM Tris-HCl, pH 8.0, 1 M NaCl, 10 mM EDTA and 0.2% (vol/vol) Triton X-100. The buffer is stable at room temperature (∼25°C) for at least 3 months.

Oligo(dT) wash buffer. Oligo(dT) wash buffer consists of 20 mM Tris-HCl, pH 8.0, 300 mM KCl and 0.1% (vol/vol) Tween-20 (may be omitted if desired). The buffer is stable at room temperature for at least 3 months.

TBST buffer (1×). To prepare TBST buffer, mix 20 mM Tris-HCl, pH 8.0, 150 mM NaCl and 0.2% (vol/vol) Tween-20. The buffer is stable at room temperature for at least 3 months.

TE buffer (1×). To prepare TE buffer, mix 10 mM Tris-HCl, pH 8.0, and 1 mM EDTA. The buffer is stable at room temperature for at least 3 months.

DNase buffer (10×). DNase buffer is prepared by combining 400 mM Tris-HCl, pH 8.0, 100 mM MgSO4 and 10 mM $CaCl_2$. The buffer is stable at − 20°C for a least 1 month.

Glycine buffer, pH 3.5 (100 mM). After autoclaving, the buffer is stable at room temperature for at least 1 month.

Preparation of left and right cassettes for ligation. The left and right cassettes are prepared by PCR amplification, digested with either EcoRI or HindIII before use, and then purified by agarose gel electrophoresis. The digested DNA fragments should be prepared no more than a few days in advance and stored at 4 °C until use.

Acidic phenol/chloroform/isoamyl alcohol (Ambion, cat. no. AM9730)

LiCl RNA precipitation solution (7.5 M; Ambion, cat. no. AM9480)

Chroma Spin-1000 column (Clontech, cat. no. 636093)

QIAquick PCR purification kit (Qiagen, cat. no. 28104)

QIAquick gel extraction kit (Qiagen, cat. no. 28704)

Ethanol (EMD Chemicals, cat. no. EX0278-1)

Glycogen (Ambion, cat. no. AM9510)

Chloroform (EMD Millipore, cat. no. CX1054-1)

T4 DNA polymerase (NEB, cat. no. M0203)

NEB No. 2 buffer (NEB, cat. no. B7002S)

T4 DNA ligase (NEB, cat. no. M0202)

Synthetic deoxyoligonucleotides (Integrated DNA Technologies)

5-Me dCTP (Fermentas, cat. no. R0431)

Methylation dNTP mix (10×; 10 mM of dATP, dTTP, dGTP, 5-Me dCTP; ZymoResearch, cat. no. D1030)

RNase H (NEB, cat. no. M0297)

DNA polymerase I (NEB, cat. no. M0209)

T7 RNA polymerase (NEB, cat. no. M0251)

T7 RNA polymerase buffer, 10X

$MgCl_2$ (VWR, cat. no. EM-5980)

rNTPs mix (25 mM) (Promega, cat. no. E6000)

Puromycin oligo linker:

5′-Psoralen-(TAGCCGGTG)$_m$-(AAAAAAAAAAAAAAA)$_d$C$_9$C$_9$(ACC)$_d$-Puromycin-3′ where, ( )$_m$ = 2'-O methyl RNA, ( )$_d$ = standard DNA, C$_9$ = triethylene glycol phosphoramidite (Spacer 9) (custom-synthesized at the W.M. Keck Oligo Facility at Yale University)

Retic lysate IVT kit (Ambion, cat. no. AM1200)

[^{35}S]-L-Met (PerkinElmer, cat. no. NEG009A05MC)

BirA enzyme (Biotin Ligase; Avidity, cat no. BirA500)

Biotin (Sigma, cat. no. B4639)

HABA (4′-hydroxyazobenzene-2-carboxylic acid) dye (Pierce, cat. no. 28010)

Nuclease-free water (Ambion, cat. no. AM9922)

NaCl (VWR, cat. no. EM-7710)

$MgCl_2$ (VWR, cat. no. EM-5980)

KCl (VWR, cat. no. BDH4532-500GP)

$CaCl_2$ (VWR, cat. no. EM-3000)

DTT (VWR, cat. no. EM-3860)

Spermidine (Sigma, cat. no. S0266)

Polyoxyethylene (20) sorbitan monooleate (VWR, cat. no. EM-9490)

EDTA disodium salt hydrate (VWR, cat. no. BDH4528-500GP)

Oligo(dT) cellulose (Ambion, cat. no. AM10020)

RNase-free 10-ml poly-prep chromatography column (Bio-Rad, cat. no. 731–1550)

Anti-FLAG M2 affinity gel (Sigma, cat. no. A2220)

FLAG peptide (Sigma, cat. no. F3290)

NAP-5 column (GE Healthcare, cat. no. 17–0853-01)

NAP-10 column (GE Healthcare, cat. no. 17–0854-01)

Streptavidin plus ultralink agarose resin (Pierce, cat. no. 53117)

BSA (Ambion, cat. no. AM2616)

RNase-free DNase (Promega, cat. no. M6101)

SuperScript II RNase H – reverse transcriptase (Invitrogen, cat. no. 18064)

Diethylpyrocarbonate (DEPC) (Sigma, cat. no. D5758-100ML)

Sodium acetate (Ambion, cat. no. AM9740)

tRNA (Ambion, cat. no. AM7119)

dNTPs (Fermentas, cat. no. R0182)

dATP (Fermentas, cat. no. R0141)

Expand long template PCR system (Roche, cat. no. 11681842001)

Synthetic deoxyoligonucleotides (Integrated DNA Technologies)

Thermal cycler

NanoDrop spectrophotometer (Thermo Fisher Scientific)

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

Labquake shaker/rotator (Barnstead)

Scintillation counter

Agarose and PAGE gel electrophoresis equipment

Gel documentation system

Microcentrifuge

Microcentrifuge tubes, nuclease- and RNase-free (e.g., Eppendorf tubes)

RNase-free PCR strip tubes

Vacuum concentrator (e.g., Speedvac)

Equipment for cloning and sequencing

Aluminum foil

3. Detailed Description of Methods

3.1. Overview of the procedure

One round of mRNA-display-based selection typically requires 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. In more detail, the newly constructed library serves as template for *in vitro* transcription to synthesize mRNAs that will be conjugated with the puromycin-containing linker. Conjugation is achieved by hybridization of the purified mRNA with the psoralen- and puromycin-containing linker, followed by UV cross-linking to form a stable 3' terminal hairpin. Other methods have been used to introduce the puromycin moiety needed for fusion formation including enzymatic splint- and Y-ligations [3,45–47]. *In vitro* translation, using puromycin-containing mRNAs as templates, is performed using rabbit reticulocyte lysate to express the proteins. Fusion formation is initiated, after translation, by the addition of high concentrations of Mg^{2+} and K^+ to favor the reaction between puromycin at the 3' end of the mRNA strand with the C-terminus of the nascent proteins. The poly(dA) tract on the oligonucleotide linker is used to purify the fusion molecules from the rabbit reticulocyte lysate via oligo(dT) resin. A second fusion purification is accomplished by using protein affinity tags on the N or C termini. Prior to the start of the selection, one round of pre-selection is typically performed to remove out-of-frame sequences and those with stop

codons following the cDNA library construction. The sequences included in the pre-selected library can be characterized by deep sequencing or by using cDNA microarrays, when available, for the organism or tissue of interest [48].

3.2. Generation of the random library

The random cassette can be synthesized on a DNA synthesizer using standard phosphoramidite chemistry [49]. For the combinatorial peptide library, we used a pre-selected, high-quality DNA library that codes for peptides with 20 consecutive random amino acids as templates for *in vitro* transcription and translation [49]. 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 C terminal His×6 tag coding sequence. Any cDNA libraries that encode short peptides with either balanced or appropriately biased amino acid compositions can be readily generated. This can be realized by using three phosphoramidite mixtures with appropriate proportions of each of the four phosphoramidites during DNA synthesis, corresponding to each of the three positions in the random codons.

3.3. Generation of the natural library

Useful natural libraries for mRNA-display can be created from any organisms, including *Caenorhabditis elegans, Drosophila melanogaster*, mouse and human, at different developmental stages, ages, tissues and treatments [8,9,24]. In general terms, the generation of an mRNA-displayed proteome library starts by making a cDNA library that contains all possible open reading frames. The cDNA library is made from high-quality poly (A$^+$) mRNA by stringent oligo (dT) purification to remove genomic DNA, rRNA and tRNA. To permit nonbiased coverage of all genes, first-strand cDNA synthesis is performed through reverse transcription by using a degenerate primer TTNNNNNN. Second-strand synthesis is performed by DNA polymerase I in the presence of RNase H. 5-Me dCTP is used during the synthesis of both cDNA strands as a way to make the ORF regions resistant to restriction digestion when the linker and cassette are ligated. T4 DNA polymerase mediated blunt end formation is followed by the ligation of a directional linker (EcoRI/HindIII linker) to permit the ligation of the left cassette (containing a T7 promoter and a deletion mutant of the tobacco virus 5' untranslated region for efficient *in vitro* transcription and translation) and right cassette (contains a short sequence for hybridizing with the puromycin-containing oligo linker) with different consensus sequences. Affinity tags such as FLAG epitope, E epitope, CaM-binding, AviTag, Myc, or His×6, may be incorporated at both ends for purification or immobilization of mRNA-displayed proteome library during selection. Amplification of the library is accomplished by using two common primers complementary to the consensus T7 promoter and 3' linker-hybridization region at the 5' and 3' end, respectively. The cDNA library with desired length is then purified and recovered.

3.3.1. Step by step procedure for the generation of a natural library

1. **First-strand synthesis**. Prepare premix A in a nuclease-free microcentrifuge tube on ice, as described in the table below:

Component	Amount per reaction (µg)	Final concentration
mRNA from species/tissue of choice	0.5–5	0.01–0.1 µg/µl
Random primer	0.125–1.0	—
Nuclease-free water	To a final volume of 30 µl	—

2. Heat the mixture to 70 °C for 10 min and quickly cool on ice.

3. Centrifuge briefly to collect the contents.

4. Prepare premix B in a nuclease-free microcentrifuge tube on ice, as described in the table below:

Component	Volume per reaction (µl)	Final concentration
First-strand buffer, 5×	10	1×
DTT, 100 mM	5	10 mM
Methylation dNTP mix, 10×	2.5	0.5 mM each
Superscript II (200 U/µl)	2.5	10 U/µl

5. Combine premix A and B with gentle pipetting (final volume 50 µl).

6. Incubate the mixture at 25 °C for 2 min and then at 37 °C for 1 h.

7. Inactivate the enzyme by heating at 70 °C for 15 min and cool on ice.

8. **Second-strand synthesis**. Assemble the following in a nuclease-free microcentrifuge tube on ice:

Component	Volume per reaction (µl)	Final concentration
First-strand product from Step 7	50	—
Second-strand buffer, 5×	30	1×
DTT, 100 mM	3.6	2.4 mM
Methylation dNTP mix, 10×	1.25	0.25 mM
DNA polymerase I (10 U/µl)	3	0.2 U/µl
RNase H (5 U/µl)	0.2	1 U per 150 µl
Nuclease-free water	To a final volume of 150 µl	—

9. Mix contents gently and incubate at 15 °C for 90 min.

10. Purify double-stranded cDNAs with the QIAquick PCR purification kit, according to the manufacturer's instructions. Dissolve the resulting cDNAs in 30 µl of TE buffer and proceed immediately with blunt ending (Step 11).

11. **Blunt ending the cDNA**. Assemble the following in a nuclease-free microcentrifuge tube on ice:

Component	Volume per reaction (µl)	Final concentration
Double-stranded cDNAs in TE buffer from Step 10	30	—
T4 DNA polymerase buffer, 10×	5	1×
DTT, 100 mM	2.5	5 mM
dNTP mix (2.5 mM each)	2.0	0.1 mM each
T4 DNA polymerase (3 U/µl)	0.83	0.05 U/µl

Component	Volume per reaction (μl)	Final concentration
Nuclease-free water	To a final volume of 50 μl	—

12. Mix gently and incubate at 11 °C for 20 min.

13. Purify the product with QIAquick PCR purification kit according to the manufacturer's instructions, and then dissolve the resulting blunt-ended cDNAs in 25 μl of TE buffer. Proceed immediately to Step 14.

14. Ligation with directional linker. Assemble the following 50 μl reaction in a nuclease-free microcentrifuge tube on ice:

Component	Volume per reaction (μl)	Final concentration
Blunt-ended cDNA from Step 13	25	—
T4 DNA ligase buffer, 10×	5	1×
DTT, 100 mM	5	10 mM
ATP, 2.5 mM	2	0.1 mM
PEG4000, 50%	8	8%
Phosphorylated EcoRI/Hind III linker	3	3 μM
T4 DNA ligase (6 Weiss Units)	8	0.12 U/μl
Nuclease-free water	To a final volume of 50 μl	

15. Mix gently and incubate the mixture at 16 °C for 20 h.

16. Add 150 μl nuclease-free DEPC-water to the ligation product and extract with 200 μl phenol:chloroform:isoamyl alcohol (PCI = 25:24:1 vol/vol/vol).

17. Add 1 μl glycogen (10 mg/ml), 200 μl of 4 M ammonium acetate and 1 ml of 100% ethanol; maintain at − 80 °C for 2 h to precipitate the ligated cDNAs.

18. Remove the supernatant after centrifugation at 12,000g for 10 min at 4 °C, and rinse the pellet with 300 μl of 70% (vol/vol) ethanol twice.

19. Allow the pellet to dry and resuspend the product in 20 μl nuclease-free DEPC-water.

20. Generation of sticky ends for ligation with cassettes. Assemble the following 50 μl reaction mixture:

Component	Volume per reaction (μl)	Final concentration
dsDNA from Step 19	20	—
NEB no. 2 buffer, 10×	5	1×
HindIII (100 U/μl)	1	2 U/μl
Nuclease-free water	To a final volume of 49 μl	

21. Mix gently and incubate the mixture at 37 °C for 2 h.

22. Add 1 μl of EcoRI-HF (100 U/μl).

23. Mix well and incubate at 37 °C for 4 h.

24. Purify the digested dsDNA with the QIAquick PCR purification kit, according to the manufacturer's instructions, and elute the product in 50 μl of 10 mM Tris-HCl, pH 8.0.

25. Pass the sample through a Chroma Spin-1000 column according to the manufacturer's instructions, and proceed immediately to ligation with left and right cassettes (Step 26).

26. **Ligation with left and right cassettes**. Assemble the following 80 μl reaction mixture:

Component	Volume per reaction (μl)	Final concentration
Digested cDNA from Step 25	50	—
T4 ligation buffer, 10× (NEB)	8	1×
Digested/purified EcoRI-HF left cassette (see Oligos Table)	X	—
Digested/purified HindIII right cassette (see Oligos Table)	X	—
T4 DNA ligase (6 Weiss U/μl)	8	0.6 U/μl
Nuclease-free water	To a final concentration of 80 μl	

27. Mix gently and incubate the ligation mixture at 16 °C for 20 h.

28. **Size separation and PCR amplification of the cDNA library**. Purify and fractionate the ligation products with desired length distribution (i.e., 500 bp to 3 kb) using 1% (wt/vol) agarose gel electrophoresis.

29. Cut the gel containing PCR products greater than 500 bp and recover the cDNA library with QIAquick gel extraction kit and dissolve the product in 100 μl TE buffer.

30. Proceed to selection section 4.1.

3.4. Preparation of RNA and conjugation

mRNA is generated by *in vitro* transcription using the newly constructed cDNA library and subsequently conjugated with the puromycin/psoralen-containing oligo linker that will be used to covalently attach the protein portion in the next subsequent step. Conjugation is achieved by hybridization of purified mRNA with psoralen/puromycin-containing oligo linker followed by UV cross-linking, to form a stable 3' terminal hairpin structure [45,46,50].

3.5. Generation and purification of mRNA-protein fusion

Creation of the mRNA-protein fusion is performed by using puromycin-containing mRNA as templates for protein synthesis. Specifically, *in vitro* translation is performed using a cell-free system (i.e., rabbit reticulocyte lysate). Moreover, following translation a high concentration of Mg^{2+} and K^+ is added to promote efficient mRNA-protein fusion formation [3,45]. To recover mRNA-containing molecules from the lysate, an oligo(dT) purification step is performed by taking advantage of oligo (dA) residues in the puromycin-containing DNA linker. To remove the secondary structures of mRNA that may interfere with the selection step, the fusion molecules are often converted to DNA/RNA hybrids by reverse transcription. The resulting mRNA-displayed proteome library is then purified on the basis

of each of the affinity tags at the N and C termini and used for subsequent selection. The genes included in the mRNA-displayed protein library can be characterized by using gene chips, if such microarrays are available for the organism of interest or by conventional Sanger sequencing a desired numbers of clones from transformed *E. coli* colonies.

3.6. Characterization of the selected library

The selected sequences displayed on mRNAs are characterized after several rounds of selection to check whether these selected proteins possess the desired properties. There are two factors that determine which round of selection is to be characterized or when the selection needs to be stopped, namely, moderately high sequence diversity and percentage of proteins in the pool with desired properties. For example, an important question in the field of apoptosis or programmed cell death, carried out by a family of proteases known as caspases, is what protein targets are specifically cleaved by a caspase of interest. To address this question, we performed an mRNA-display selection to determine the substrates of one of the caspase family members, caspase-3 (Fig. 1). In this example, sequences from rounds 3, 4 and 5 of the caspase-3 selection were PCR-amplified and cloned into a TOPO vector, sequenced and analyzed by a high-throughput digestion assay [8]. At this time, it is important to see how diverse each selected pool is. Higher diversity increases the probability of finding novel target binding proteins or protease substrates. In the case of the caspase selection, it was important to determine the number of unique genes from clones that were sequenced. Alternatively, genes were identified by hybridizing labeled, transcribed RNA from the selected pool to an expression microarray. By analyzing the intensity of each sequence, it was readily distinguishable whether a sequence was present or absent. The second step in determining whether to continue the selection or characterize a particular round is to evaluate the percentage of the selected sequences that possess the desired properties. In our case, we must determined what percent of proteins can be cleaved by a protease of interest or what percent of selected sequences bind to a target of interest. For the proteolytic selection, the cleavage of each selected protein sequence by a protease of interest may be tested by an *in vitro* proteolytic assay, using ^{35}S-radiolabeled protein generated by coupled *in vitro* transcription and translation (TNT), in the presence or absence of a protease inhibitor. mRNA display can also be performed without using radio-labeling [7], but the interpretation of the enrichment signal should be cautious. Each reaction mixture is loaded to a high resolution Tricine-SDS/PAGE for analysis. Numerous positive and negative controls are tested to confirm that this *in vitro* proteolytic assay is robust, provided that the protein of interest is translated well in rabbit reticulocyte lysate. In our caspase-3 selection, we decided to characterize round 3 pool because among unique proteins from a total of 580 clones we picked up for analysis, 173 were translated well by the high-throughput *in vitro* transcription and translation (TNT), among which 115 proteins (66%) were found to be specifically cleaved by caspase-3 [8]. In this case, the diversity of round 3 was high, meaning, there were many novel substrates that were likely to be found as well as a high percent of all of substrates that were specifically cleaved by the caspase of interest.

3.7. Different selection scheme design

3.7.1. Salt, buffer, and pH—Successful selections that investigate protein-protein, protein-drug and protein-enzyme interactions depend on the selection design that optimally allows for specific enrichment of sequences with desired characteristics, while minimizing nonspecific sequences that might be co-selected due to various biases [51]. To this end, the highly stable covalent linkage between mRNA and protein permits the employment of many types of selection with diverse and stringent conditions. The flexibility of being able to use various selection conditions is important to distinguish proteins with desired characteristics from nonspecific ones. The selection stringency can be carefully controlled and tuned by using detergents, chelating agents, unusual pH, temperature and ionic strength, in addition to

nonspecific or specific competitors, cofactors and metal ions, as long as they do not interfere with the functions of the peptides or proteins [51]. In addition, the expression of the displayed proteins in rabbit reticulocyte lysate allows for some post-translational modifications, thereby enhancing their functionality during the selection. Furthermore, *in vitro* protein expression in mammalian cell lysate also minimizes biases from misfolding or subcellular transport *in vivo*.

3.7.2. Length of library and folding of proteins—Another tunable feature of mRNA-display is the ability of an investigator to manipulate the length of the cDNA and consequently protein size. Typically, we have used an mRNA-displayed proteome library whose length of peptides ranges from 50 to 300 amino acids. The protein size distribution can be changed as desired by fractionation of the initial cDNA library. Functionally, longer proteins are much more likely than short synthetic peptides to adopt native conformations involved in mediating protein–protein, protein-drug and enzyme-substrate interactions. Also, the use of longer proteins during selection allows for the identification of motifs that play a range of major or minor roles in these interactions.

4. Selection Results and Discussion

Successful mRNA-display based selections are highly dependent on a robust and effective selection scheme [26]. In one hand, an appropriate selection scheme will result in the isolation of proteins with desired characteristics after a few rounds. In contrast, poorly designed selections will enrich irrelevant sequences that quickly overwhelm the regenerated library pool. To aid in the selection of specific sequences, competitive elution, using the free target following target immobilization, is an effective strategy that should be employed. Furthermore, selection approaches based on 'binary' events, such as protease cleavage or conditional binding in the presence or absence of a third partner or cofactor, can also be effective in identifying proteins with desired characteristics after several rounds [8,9].

Key points on how to perform successful selections and procedures of two selection examples, protein-protein and enzyme-substrate interactions, respectively, are described in the following sections. The aim of the first selection was to identify Ca^{2+}-dependent CaM binding proteins and of the second selection was to identify the downstream substrates of a protease of interest. In the design of the Ca^{2+}-dependent CaM-binding protein selection, gentle and competitive elution was applied by specifically eluting with EGTA to release the bound mRNA-protein fusion molecules (Fig. 1). For the selection of the downstream substrates of a protease, the proteome library was immobilized on streptavidin-agarose beads through a single biotin residue introduced during *in vitro* translation. The addition of a protease of interest results in cleavage of endogenous substrates present in the library and release of mRNA-protein fusions from the agarose beads (Fig. 1). One major challenge in understanding caspase-mediated signaling pathways is the identification of the caspase family member-specific substrates. In our substrate identification platform, selective identification of member-specific substrates can be readily achieved by first treating the immobilized library with related caspases prior to the selection using the caspase of interest. To identify the family member-specific substrates of a caspase or caspase-like enzyme (granzyme B), for example, the immobilized proteome library is first incubated with other caspase members to remove the overlapping, redundant substrates. The pre-cleared library is then used for selection to enrich the natural substrates that are only cleaved by the protease of interest, such as granzyme B, but not by other caspase members. Using this method, we successfully identified many family member-specific substrates for different human caspase family members and granzyme B [8] (Fig. 2).

To minimize nonspecific sequences, there are several steps that should be employed to all selections. The binding buffer should have RNase-free tRNA and BSA to reduce nonspecific binding of fusion molecules to solid surfaces or immobilized targets. Only RNase- and DNase-free reagents should be utilized in the selections. The wash stringency should be tested prior to elution of each cycle by using both positive and negative control sequences to determine the optimal stringency. In a target-binding selection, a stringent wash may effectively remove nonspecific sequences that bind to the target, but may also result in loss of the desired sequences if their binding to target is weak.

Another important parameter in any given selection is the number of selection cycles. Often, the enrichment efficiency during each selection cycle is unknown; therefore, one approach is to perform 3 to 4 rounds of selection followed by sequencing one 96-well plate of constructs from the eluted pools obtained after each round. Analysis of the frequency of unique genes present in each pool provides insight into the diversity and enrichment of each cycle. Pools that are dominated by only several sequences may be unsuitable for further characterization. Conversely, pools that show moderate enrichment of several sequences (<20%) may provide greater diversity for validation screening.

4.1. Binding-based selection

1. Following the library construction, a large-scale amplification is performed by setting up 100 separate 50 μl PCR amplifications of the cDNA library. As opposed to small PCR amplifications that are sequentially performed by using the previous PCR product for the amplification each time, this large-scale amplification prevents decreasing the library diversity. Most of the resulting amplified cDNA library should be stored – 80 °C and should serve as the starting point of many selections.

2. To start the binding selection, usually 100 separate 50-μl PCR amplifications of the stock library should be setup as follows:

Component	Volume per reaction (μl)	Final concentration
Gel-purified cDNA library	1	—
Expand long template buffer no. 1, 10 ×	5	1×
dNTP mix (2 mM each)	8.75	0.35 mM
Forward library primer (10 μM)	2.5	0.5 μM
Reverse library primer (10 μM)	2.5	0.5 μM
Expand long template Taq (5 U/μl)	0.7	3.5 U per 50 μl
Nuclease-free water	To a final volume of 50 μl	

3. Run the following PCR program:

Cycle number	Denature	Anneal	Extend
1	95 °C for 2 min		
2–17 (or up to 27)	95 °C for 30 sec	65 °C for 30 sec	68 °C for 6 min
Final			68 °C for 7 min

4. Analyze the PCR products by agarose gel (1%, wt/vol) electrophoresis to check the length distribution of the cDNA library.

5. The number of PCR cycles needed to amplify the library should be titrated on a small scale first to avoid overamplification of nonspecific products.

6. Purify the DNA following the detailed instructions below.

7. Pool 500 μl of amplified cDNA library into 1.5 ml microcentrifuge tube and add 3.2 μl of 0.5 M EDTA (pH, 8.0).

8. Mix and briefly centrifuge.

9. Add 0.5 ml of PCI.

10. Mix 10 times and hold at room temperature for 5 min.

11. Centrifuge at 14,000 rpm for 6 min.

12. Remove the supernatant and repeat the PCI purification.

13. Remove the supernatant and purify the DNA by performing a chloroform extraction

14. Transfer the supernatant and precipitate by the addition of 0.1 volumes of 3 M KOAc and 2.2 volumes of 100% ethanol and store at −20 °C overnight.

15. Centrifuge at 4 °C for 20 min at 13,400 rpm and wash the pellet with 500 μl of 70% (vol/vol) ethanol.

16. Dry the pellet at room temperature until dry and resuspend the DNA in 30 μl of DEPC-water and quantify the library using a nanodrop spectrophotometer.

17. **Generation of mRNA by *in vitro* transcription**. Assemble the following in a nuclease-free microcentrifuge tube on ice:

Component	Volume per reaction (μl)	Final concentration
cDNA library	X	200 nM
T7 RNA polymerase buffer, 10×	50	1×
rNTPs mix (25 mM each)	100	5 mM
MgCl$_2$ (1.0 M)	9.5	19 mM (or 25 mM including 6 mM in 1× T7 RNA polymerase buffer)
T7 RNA polymerase (50 U/μl)	30	3 U/μl
Nuclease-free water	To a final volume of 500 μl	

18. Incubate at 37 °C for 6–12 h.

19. Add 30 μl of 0.5 M EDTA (1.2 equivalent of [Mg^{2+}]) and incubate at room temperature for 5 min to dissolve the white precipitate (magnesium pyrophosphate).

20. Extract the reaction mixture with 530 μl of acidic phenol chloroform, followed by 500 μl chloroform.

21. Desalt the product using NAP-5 column and collect the product in 1 ml of nuclease-free DEPC-water.

22. Concentrate the sample to 400 μl in a vacuum concentrator.

23. **Removal of cDNA from mRNA by DNase digestion**. Assemble the following in a nuclease-free microcentrifuge tube:

Component	Volume per reaction (µl)	Final concentration
Desalted *in vitro* transcription product from Step 9	400	—
DNase buffer (10×)	50	1×
RNase-free DNase (1 U/µl)	25	0.05 U/µl
Nuclease-free water	To a final volume of 500 µl	

24. Incubate at 37 °C for 30 min.

25. Add 13.2 µl of 0.5 M EDTA to stop the reaction.

26. Extract the reaction mixture with 500 µl of acidic phenol chloroform, followed by 500 µl of chloroform.

27. Desalt using NAP-5 column and collect the product in 1 ml of nuclease-free DEPC-water.

28. Add 500 µl (0.5 volumes) of 7.5 M LiCl RNA precipitate solution and incubate at −80 °C for 2 h.

29. Spin at 13,200 g for 15 min at 4 °C and wash the pellet with 500 µl of 75% (vol/vol) ethanol.

30. Allow the pellet to air-dry until no ethanol is detectable. Resuspend the purified mRNA in 150 µl nuclease-free DEPC-water.

31. Measure optical density (OD) using the NanoDrop; calculate the molar concentration of mRNA on the basis of the average length of the cDNA library.

32. **Conjugation with puromycin-containing oligo linker**. Assemble the following reaction mixture:

Component	Volume per reaction (µl)	Final concentration
500 mM HEPES, pH 7.4	10	20 mM
KCl (500 mM)	50	100 mM
Purified mRNA from Step 17 (1.25 nmol)	X	5 µM
Puromycin-containing oligo linker	3.125 nmol	12.5 µM
Nuclease-free water	To a final volume of 250 µl	

33. Mix well and aliquot 50 µl of the sample into an RNase-free eight-strip PCR tube.

34. Anneal RNA and puromycin oligo linker on a thermal cycler under the following conditions:

Cycle number	Denature	Anneal	Extend
1	85 °C for 8 min		
2	80 °C for 60 sec		
3		− 1 °C every 20 sec until 25 °C is reached	
4			25 °C 25 min

35. Transfer strip tubes to an ice bath and irradiate by using UV light at 365 nm for 20 min at 4 °C. Exposure to UV light results in cross-linking between the puromycin linker and RNA through activation of the psoralen linker.

36. Pool the samples into an RNase-free 0.5-ml microcentrifuge tube.

37. Add 125 μl (0.5 volumes) of LiCl precipitate solution and incubate at − 20 °C for 4 h to precipitate RNA.

38. Wash the pellet with 250 μl of 75% (vol/vol) ethanol, air-dry the pellet and resuspend the RNA product in 100 μl DEPC-water.

39. Measure OD using the NanoDrop and calculate the molar concentration of RNA.

40. *In vitro* **translation and fusion formation**. Assemble the following reaction mixture as follows:

Component	Volume per reaction (μl)	Final concentration
Translation mix without Met, 12.5×	40	1×
KCl (2.0 M)	12.5–30	50–120 mM
Mg(OAc)$_2$ (25 mM)	6–20	0.3–1.0 mM
Puromycin-linked mRNA templates (from Step 25)	200 nM (assuming 40% cross-linking efficiency at Step 22)	—
[^{35}S]-L-Met (10 μCi/μl)	20	—
Retic lysate	200	40%
Nuclease-free water	To a final volume of 500 μl	

41. Incubate at 30 °C for 90 min. If the introduction of a biotin residue at the N terminus of the protein moiety (with an AviTag engineered) is desired for immobilization of the mRNA-displayed proteome library, add 10 μg of purified BirA enzyme, 50 μM free d-biotin and 10 mM Mg^{2+}/ATP to the translation reaction mixture and incubate at 30 °C for another hour.

42. Add 38 μl of 1 M MgCl$_2$ (final concentration 50 mM) and 220 μl of 2 M KCl (final concentration 580 mM), mix gently (total volume 758 μl) and incubate at room temperature for 30 min.

43. Incubate at − 20 °C overnight.

44. O**ligo(d**T**) purification to remove free proteins**. Weigh 60 mg of oligo(dT) cellulose into a 1.5-ml nuclease-free microcentrifuge tube.

45. Wash cellulose beads three times with 1 ml of nuclease-free DEPC-water, and two times with 1 ml of oligo(dT) binding buffer.

46. Resuspend oligo(dT) cellulose in 1 ml of binding buffer and transfer to a 15-ml tube containing 9 ml of oligo(dT) binding buffer and 1 mM DTT.

47. Add the post-translation mix (758 μl from Step 30) to the slurry, wrap the tube in aluminum foil and rotate for 2 h at 4 °C.

48. Load the slurry mixture into a 10 ml nuclease-free poly-prep chromatography column.

49. Collect flow-through and reload to the resin bed; retain the final flow-through.

50. Wash oligo(dT) cellulose on the column twice each with 1 ml of oligo(dT) binding buffer, followed by three times each with 1 ml of oligo(dT) wash buffer; retain each wash fraction.

51. Elute mRNA and mRNA-protein fusion molecules four times each with 600 μl of DEPC-water plus 1 mM DTT.

52. Take 1/100 of each fraction and quantify the radioactivity using a liquid scintillation counter. The quantity of radioactivity present in the final wash step should be at baseline to ensure that unincorporated Met is fully removed. Elution fractions typically show stepwise elevations in radioactivity with the highest counts in the second or third elution fractions. Typically, the second elution fraction is sufficient for subsequent steps if it shows the majority of the radioactivity.

53. **Reverse transcription**. Assemble the following reaction mixture and incubate at room temperature for 15 min:

Component	Volume per reaction (μl)	Final concentration
Oligo(dT) elution fraction (from Step 38)	600	
Reverse-transcription primer (50 μM)	50	2.5 μM
First-strand synthesis buffer, 10×	100	1×
dNTP mix (5 mM each)	100	0.5 mM each
DTT, 1M	9.4	10 mM
Nuclease-free water	To give a final volume of 975 μl	

54. Add 25 μl Superscript II (200 U/μl) to the mixture from Step 40 to yield a final concentration of 5 U/μl. Incubate the reaction mixture at 42 °C for 50 min.

55. Add EDTA to 3.6 mM ($1.2\times$ [Mg^{2+}]).

56. Exchange the buffer by applying to a NAP-10 column equilibrated with 1× TBST buffer.

57. Collect the reverse-transcribed product in 1.35 ml of 1× TBST buffer.

58. **Anti-FLAG purification to remove free mRNA**. Quickly wash 600 μl of anti-FLAG M2 affinity resin in a 10-ml poly-prep chromatography column five times, each with 1 ml of 100 mM glycine buffer at pH 3.5.

59. Wash the resin three times, each with 5 ml of 1× TBST buffer.

60. Cap the bottom of the column and load the product (1.35 ml) from Step 43.

61. Cap the top of the column with a lid, and rotate at 4 °C for 2 h at a 45 °C angle such that the resin is mixed well and stays at the bottom of the column. Prevent mRNA-displayed protein loss during this purification step by minimizing the attachment of solution beads to the top and sides of the chromatography column by maintaining a proper rotation angle.

62. Place the column upright and allow the resin to settle for 2 min. Remove top and bottom caps and collect the flow-through.

63. Wash the resin five times, each with 1 ml of 1× TBST, and collect each wash fraction.

64. Cap the bottom of the column and add 60 μl of 1× TBST plus 30 μl of 5 mg/ml FLAG peptide to the slurry and rotate at 4 °C for 30 min to elute the captured fusion molecules.

65. Collect the elution fraction and repeat Step 51 three times.

66. Take 1/100 of flow-through, wash and elution fractions and count the amount of fusion molecules recovered with a liquid scintillation counter.

67. **Functional selection using mRNA-protein fusion molecules**. This step should be designed and optimized for each different selection experiment. Here we provide examples of two different selection schemes that could be modified as required. Option A outlines a scheme for selecting protein-protein interactions. Option B outlines a scheme for selecting substrates of a specific protease.

   A) selection for conditional (Ca^{2+}-dependent) protein-protein interactions

   i. Dilute purified mRNA-displayed proteome library from Step 53 in 0.5–2.0 ml of an appropriate binding buffer. For selection of Ca^{2+}-dependent calmodulin-binding proteins, the binding buffer contains 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 $CaCl_2$.

   ii. Apply the mRNA-displayed proteome library to a precolumn of 100–300 μl streptavidin plus ultralink agarose beads. Collect the flow-through.

   iii. Incubate flow-through with 5–25 μg of biotinylated target protein (i.e., biotinylated calmodulin) for 30–90 min at 4 °C. Note: It is important to measure the average number of biotin per target protein by using a colorimetric biotin quantification assay based on the use of the HABA reagent. Ideally, one biotin per target molecule is desired.

   iv. Add the same amounts of prewashed streptavidin-agarose beads as in Step 54A(ii) (100–300 μl) to the mixture and incubate for 30 min at room temperature to capture the biotinylated target protein complexed with its binding partners.

   v. Load the slurry mixture into an empty nuclease-free 10 ml poly-prep chromatography column and wash unbound molecules from the column with 9–24 column volumes (3 × 3–8 × 3 column volumes) of binding buffer.

   vi. Elute the fusion molecules that bind to calmodulin in a Ca^{2+}-dependent manner using 1 column volume of elution buffer containing chelating agent (binding buffer minus $CaCl_2$ plus 2 mM EGTA) four times.

   vii. Collect each elution fraction, desalt with NAP-10 column in DEPC-water and proceed with regeneration (Step 54).

4.2. Cleavage-based selection

   B) Selection for the downstream substrates of a protease of interest

i. Dilute the purified mRNA-displayed proteome library from Step 53 in 0.5–2.0 ml of an appropriate reaction buffer for the protease of interest. For selection of the downstream substrates of human caspases, for example, the reaction buffer contains 50 mM HEPES, pH 7.4, 50 mM NaCl, 0.1% (vol/vol) CHAPS, 5% (vol/vol) glycerol, 1 mg/ml BSA, 10 mM DTT and 10 mM EDTA.

ii. Apply the proteome library to a precolumn of 100 µl agarose beads.

iii. Collect the flow-through and apply to a new column of 100 µl streptavidin plus ultralink agarose beads. Incubate for 1 h at room temperature to immobilize the biotinylated mRNA-protein fusion molecules on the resin.

iv. Wash unbound molecules with 250 column volumes of protease reaction buffer. This helps to lower the number of false positive molecules passing through the column before protease-mediated digestion. Add an appropriate amount of a highly purified protease of interest (free of nucleases and other proteases) in the reaction buffer and incubate for 30–90 min at 37 °C with rotation.

v. Collect flow-through of the proteolytic reaction, desalt with NAP-10 column in DEPC-water and proceed with regeneration (Step 54).

54. **Regeneration of the selected sequences**. Assemble the following reaction mixture:

Component	Volume per reaction (µl)	Final concentration
Eluted fusion molecules in nuclease-free water	70	70%
Forward library primer (15 µM)	2	300 nM
Reverse library primer (15 µM)	2	300 nM
dNTP mix (5 mM each)	7	0.35 mM
Expanded long template PCR buffer, 10×	10	1×
Long template Taq enzyme (5 U/µl)	1	5 U per 100 µl
Nuclease-free water	To a final volume of 100 µl	

55. Mix well and aliquot 15-µl samples into six different PCR tubes.

56. Run the following PCR program:

Cycle number	Denature	Anneal	Extend
1	95 °C for 2 min		
2–16 (or up to 31)	95 °C for 30 sec	65 °C for 30 sec	68 °C for 6 min
Final			68 °C for 7 min

57. Remove a PCR tube at cycles 15, 18, 21, 24, 27 and 30, respectively. Transfer all tubes back to the thermocycler for the last extension step (68 °C for 6 min).

58. Run each sample on a 1% (wt/vol) agarose gel to determine how many cycles of PCR are necessary to regenerate the library without overamplification. Overamplification occurs when too many cycles of PCR are performed on a

library resulting the upward skewing of the library length. To prevent overamplification, a small volume of the library to be amplified is used as a template and numerous amplification cycles are tested. The PCR products of various cycles are separated and compared on an agarose gel and the lane with the highest amount of DNA, based on intensity, having the expected average size should be the library to be used for subsequent steps.

59. Repeat the PCR outlined in Steps 54–56 on a large scale using 30–80% of the selected molecules. Clone, sequence and bioinformatically analyze the selected molecules.

60. To begin the next round of selection, carry out *in vitro* transcription (Step 4) on the regenerated cDNA library and repeat the entire selection cycle (Steps 4–59)

4.3. Selection results analysis

4.3.1. Bioinformatics—After appropriate rounds of selection, hundreds of clones are typically sequenced and identified by BLASTing the selected sequences or by hybridizing the selected sequences onto cDNA microarrays (Fig. 1). Once identified, the percent of unique clones that bind to target or is cleaved by a protease of interest is determined. Once the identity of the clones has been determined, the binding or cleavage sites can be easily mapped (Fig. 3). Typically, the selected sequences are often isolated as multiple fragments of different lengths derived from the same parental ORF, and the shortest overlapping region can be readily mapped [8,9]. Based on the results from many known caspase substrates isolated from the selection, the shortest overlapping region often contains the cleavage site(s). Similarly, in the CaM binding protein selection, the shortest fragments indicated the motifs involved in binding to CaM, thus, known and novel Ca^{2+}/CaM-binding domains were readily discovered.

Using this method, the cleavage sites of the selected caspase substrates were determined following the caspase-substrate selection. Significantly, it was found that many of the mapped cleavage sites are located within or close to a structural domain, suggesting that the functions of these substrates could be altered or abolished by caspase-mediated proteolysis. In general, the mapped cleavage sites are consistent with the assigned classes reported in the literature [52]. In the case of caspase-3 substrates, P1 (amino acid that is C-terminally cleaved, additional sites are N-terminal from P1 labeled as P2, P3, etc.) is completely conserved with an aspartic acid, followed by P4, which is dominated by an aspartic residue. P3 and P1 are not highly conserved, but glutamic acid and glycine occur at the highest frequencies, respectively. At the P6, P5 and P2 positions, the variance is high, suggesting these positions are not critical for enzyme specificity.

The CaM binding motifs are also easily mapped, specifically, by locating the shortest overlapping region from different selected fragments of the parental protein. As an example, among 499 clones analyzed from round 1 of the CaM binding protein selection, 14 were from α-fodrin, a well characterized CaM-binding protein. These 14 protein fragments ranged from 80 to 189 residues and overlapped from M1158-N1192 (MMPRDETDSKTASPWKSARLMVHTVATFNSIKELN)[9]. This 34 amino acid fragment is just 14 amino acids longer than the previously mapped CaM-binding motif of α-fodrin (A1169-I1188). Similar results were obtained from many other well characterized CaM-binding proteins. The CaM-binding motifs in the previously uncharacterized CaM-binding proteins were generally predicted to contain a positively charged amphiphilic α-helix when the selected sequence fragments were analyzed by using a Web-based motif analysis program called Calmodulin Target Database (http://calcium.uhnres.utoronto.ca/ctdb/ctdb/sequence.html).

A challenge in the proteome-wide selections is the interference of sequences that dominate a selected pool. Fragments from several genes may be preferentially enriched, which could interfere with the isolation of lower abundance substrates. However, mRNA-display offers a solution by removing the abundant sequences from a given round of selection. Specifically, after obtaining the identity of the abundant genes by sequencing a couple of hundred clones, biotinylated antisense oligos are designed against the common region mapped by aligning the abundant protein sequences derived from the parent proteins [23]. Following RNA hydrolysis and neutralization, hybridization of the biotinylated oligos with the complementary cDNA, and passage through a streptavidin column, it was found that these abundant sequences were effectively removed. This unique feature significantly increases the chance of discovering non-abundant protein binding partners or protease substrates [23].

4.3.2. Biochemical assays—Although after each round of selection there is a percentage of proteins that have the desired characteristics, it is important to biochemically confirm the enriched clones by a high-throughput assay to address cleavage or binding specificity, inhibition, and kinetics. Several key features of mRNA-display are that the recovery of the coding sequences from a selection allows for rapid identification, regeneration, and functional characterization of the enriched proteins. Specifically, selected mRNA-display molecules contain all the necessary sequences for efficient *in vitro* transcription and translation, such as the T7 promoter, start codon, and Kozak sequences, and therefore can be directly used as templates to regenerate the proteins. Since the proteins can be radiolabeled with [^{35}S]-methionine during coupled *in vitro* transcription and translation, highly sensitive proteolytic or binding assays can be performed without using antibodies.

4.3.3. Protease-substrate orthologs in model organism—The identification of novel human protease substrates, along with the readily mapped cleavage sites, will greatly facilitate our understanding of such signaling pathways in other organisms. For example, although caspase-3 in mouse and human (HCASP3) has been thoroughly investigated enzymatically, its homolog in zebrafish has not yet been detailed. The amino acid sequence identities of zebrafish caspase-3 (ZCASP3) with human, mouse, *Drosophila*, and *C. elegans* caspase-3 are 62%, 61%, 42%, and 33%, respectively [53–55]. Functional analysis shows that overexpressing ZCASP3 had caspase activity and induced apoptosis in fathead minnow tailbud cells and zebrafish embryos [53]. Our knowledge about the signaling pathways mediated by ZCASP3 is limited. To elucidate the potential ZCAP3 substrates, we examined the conservation of the caspase-3 cleavage sites, found in the human caspase-3 selection, in zebrafish orthologs (Fig. 4). These results prompted us to investigate whether the zebrafish orthologs of these human caspase-3 substrates are indeed cleaved by ZCASP3. Among approximately 60 novel human caspase-3 substrates we analyzed, 27 of them have the putative DXXD cleavage sites conserved in their zebrafish orthologs. To examine whether these zebrafish proteins were zebrafish ZCASP3 substrates, they were expressed and subjected to a proteolytic assay (Fig. 4). Using this approach, we identified 14 natural ZCASP3 substrates that have not yet been previously studied [56]. In general, the database exploration approach can be used to define the function of homologous proteins in other biological systems, including model organisms.

5. Conclusions

mRNA-display can be used to identify polypeptide sequences with desired functions from both synthetic peptide and natural proteome libraries. mRNA-display has been used for isolation of novel and improvement of existing affinity reagents, evolution of enzymes with desired catalytic activities, identification of drug-binding targets, mapping and elucidation of protein-protein, nucleic acid-protein, and enzyme-substrate interaction networks.

Since we only rely on the presence of mRNA, mRNA-display can be widely used to address various biological questions in any organism of interest. Its unique features, together with the freedom of using arbitrary selection conditions, make mRNA-display-based selection an ideal approach for the comprehensive cataloging of the proteins that participate in a specific signaling pathway. In the biotechnology realm, mRNA-display selections have the great potential to be applied for the development of novel affinity proteins against new biomarkers for *in vivo* applications such as cancer drug delivery. Furthermore, mRNA-display may aid in the discovery of next generation biopharmaceuticals that highly specifically block the activation of signaling pathways.

mRNA-display is a totally *in vitro* peptide or protein selection method. Owing to its unique features, we expect that this technique will be widely applied to investigate many different types of biological questions that are difficult to address using other protein selection techniques. In particular, it is ideal for the investigation of conditional protein–protein interactions that rely on the present of a third partner. We expect that improvement will be made to make mRNA-display suitable for selections using bioactive receptors on the surface of living cells. We also expect that mRNA-displayed full-length proteins, such as all the proteins involved in a particular signaling pathway, will be used to understand their comprehensive protein–protein interaction networks and the regulatory mechanisms governing them.

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. Meth. Enzymol. 2000; 318:268–293. [PubMed: 10889994]

4. 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]

5. 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]

6. 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]

7. Horisawa K, Tateyama S, Ishizaka M, Matsumura N, Takashima H, Miyamoto-Sato E, et al. In vitro selection of Jun-associated proteins using mRNA display. Nucleic Acids Res. 2004; 32:e169. [PubMed: 15576676]

8. Ju W, Valencia CA, Pang H, Ke Y, Gao W, Dong B, 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. [PubMed: 17728405]

9. 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. [PubMed: 15840729]

10. Tateyama S, Horisawa K, Takashima H, Miyamoto-Sato E, Doi N, Yanagawa H. Affinity selection of DNA-binding protein complexes using mRNA display. Nucleic Acids Res. 2006; 34:e27. [PubMed: 16478713]

11. Fukuda I, Kojoh K, Tabata N, Doi N, Takashima H, Miyamoto-Sato E, et al. In vitro evolution of single-chain antibodies using mRNA display. Nucleic Acids Res. 2006; 34:e127. [PubMed: 17012279]

12. Roberts RW. Totally in vitro protein selection using mRNA-protein fusions and ribosome display. Curr Opin Chem Biol. 1999; 3:268–273. [PubMed: 10359713]

13. Conrad U, Scheller J. Considerations on antibody-phage display methodology. Comb. Chem. High Throughput Screen. 2005; 8:117–126. [PubMed: 15777175]

14. Huang H, Jedynak BM, Bader JS. Where have all the interactions gone? Estimating the coverage of two-hybrid protein interaction maps. PLoS Comput. Biol. 2007; 3:e214. [PubMed: 18039026]

15. Rothe A, Surjadi RN, Power BE. Novel proteins in emulsions using in vitro compartmentalization. Trends Biotechnol. 2006; 24:587–592. [PubMed: 17055094]

16. Lancet D, Sadovsky E, Seidemann E. Probability model for molecular recognition in biological receptor repertoires: significance to the olfactory system. Proc. Natl. Acad. Sci. U.S.A. 1993; 90:3715–3719. [PubMed: 8475121]

17. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, et al. Initial sequencing and analysis of the human genome. Nature. 2001; 409:860–921. [PubMed: 11237011]

18. International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. Nature. 2004; 431:931–945. [PubMed: 15496913]

19. Endo Y, Sawasaki T. Cell-free expression systems for eukaryotic protein production. Curr. Opin. Biotechnol. 2006; 17:373–380. [PubMed: 16828277]

20. Baneyx F, Mujacic M. Recombinant protein folding and misfolding in Escherichia coli. Nat. Biotechnol. 2004; 22:1399–1408. [PubMed: 15529165]

21. Porro D, Sauer M, Branduardi P, Mattanovich D. Recombinant protein production in yeasts. Mol. Biotechnol. 2005; 31:245–259. [PubMed: 16230775]

22. Bishop JO, Morton JG, Rosbash M, Richardson M. Three abundance classes in HeLa cell messenger RNA. Nature. 1974; 250:199–204. [PubMed: 4855195]

23. Valencia CA, Cotten SW, Dong B, Liu R. mRNA-display-based selections for proteins with desired functions: a protease-substrate case study. Biotechnol. Prog. 2008; 24:561–569. [PubMed: 18471027]

24. Shen X, Valencia CA, Gao W, Cotten SW, Dong B, Huang B-C, et al. Ca(2+)/Calmodulin-binding proteins from the C. elegans proteome. Cell Calcium. 2008; 43:444–456. [PubMed: 17854888]

25. Yuan L, Kurek I, English J, Keenan R. Laboratory-directed protein evolution. Microbiol. Mol. Biol. Rev. 2005; 69:373–392. [PubMed: 16148303]

26. Cotten SW, Zou J, Valencia CA, Liu R. Selection of proteins with desired properties from natural proteome libraries using mRNA display. Nat Protoc. 2011; 6:1163–1182. [PubMed: 21799486]

27. Lin H, Cornish VW. Screening and selection methods for large-scale analysis of protein function. Angew. Chem. Int. Ed. Engl. 2002; 41:4402–4425. [PubMed: 12458502]

28. Lamboy JA, Tam PY, Lee LS, Jackson PJ, Avrantinis SK, Lee HJ, et al. Chemical and genetic wrappers for improved phage and RNA display. Chembiochem. 2008; 9:2846–2852. [PubMed: 18973165]

29. Kurz M, Gu K, Al-Gawari A, Lohse PA. cDNA - protein fusions: covalent protein - gene conjugates for the in vitro selection of peptides and proteins. Chembiochem. 2001; 2:666–672. [PubMed: 11828503]

30. Tabuchi I, Soramoto S, Nemoto N, Husimi Y. An in vitro DNA virus for in vitro protein evolution. FEBS Lett. 2001; 508:309–312. [PubMed: 11728441]

31. Yamaguchi J, Naimuddin M, Biyani M, Sasaki T, Machida M, Kubo T, et al. cDNA display: a novel screening method for functional disulfide-rich peptides by solid-phase synthesis and stabilization of mRNA-protein fusions. Nucleic Acids Res. 2009; 37:e108. [PubMed: 19528071]

32. Tabata N, Sakuma Y, Honda Y, Doi N, Takashima H, Miyamoto-Sato E, et al. Rapid antibody selection by mRNA display on a microfluidic chip. Nucleic Acids Res. 2009; 37:e64. [PubMed: 19336414]

33. 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]

34. Keefe AD, Szostak JW. Functional proteins from a random-sequence library. Nature. 2001; 410:715–718. [PubMed: 11287961]

35. Ja WW, West AP Jr. Delker SL, Bjorkman PJ, Benzer S, Roberts RW. Extension of Drosophila melanogaster life span with a GPCR peptide inhibitor. Nat. Chem. Biol. 2007; 3:415–419. [PubMed: 17546039]

36. Huang B-C, Liu R. Comparison of mRNA-display-based selections using synthetic peptide and natural protein libraries. Biochemistry. 2007; 46:10102–10112. [PubMed: 17685586]

37. Raffler NA, Schneider-Mergener J, Famulok M. A novel class of small functional peptides that bind and inhibit human alpha-thrombin isolated by mRNA display. Chem. Biol. 2003; 10:69–79. [PubMed: 12573700]

38. Li S, Millward S, Roberts R. In vitro selection of mRNA display libraries containing an unnatural amino acid. J. Am. Chem. Soc. 2002; 124:9972–9973. [PubMed: 12188645]

39. Hartman MCT, Josephson K, Szostak JW. Enzymatic aminoacylation of tRNA with unnatural amino acids. Proc. Natl. Acad. Sci. U.S.A. 2006; 103:4356–4361. [PubMed: 16537388]

40. Hartman MCT, Josephson K, Lin C-W, Szostak JW. An expanded set of amino acid analogs for the ribosomal translation of unnatural peptides. PLoS ONE. 2007; 2:e972. [PubMed: 17912351]

41. Subtelny AO, Hartman MCT, Szostak JW. Ribosomal synthesis of N-methyl peptides. J. Am. Chem. Soc. 2008; 130:6131–6136. [PubMed: 18402453]

42. Xu L, Aha P, Gu K, Kuimelis RG, Kurz M, Lam T, et al. Directed evolution of high-affinity antibody mimics using mRNA display. Chem. Biol. 2002; 9:933–942. [PubMed: 12204693]

43. 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]

44. Seelig B, Szostak JW. Selection and evolution of enzymes from a partially randomized non-catalytic scaffold. Nature. 2007; 448:828–831. [PubMed: 17700701]

45. Kurz M, Gu K, Lohse PA. Psoralen photo-crosslinked mRNA-puromycin conjugates: a novel template for the rapid and facile preparation of mRNA-protein fusions. Nucleic Acids Res. 2000; 28:E83. [PubMed: 10982894]

46. Leemhuis H, Stein V, Griffiths AD, Hollfelder F. New genotype-phenotype linkages for directed evolution of functional proteins. Curr. Opin. Struct. Biol. 2005; 15:472–478. [PubMed: 16043338]

47. Tabuchi I, Soramoto S, Suzuki M, Nishigaki K, Nemoto N, Husimi Y. An Efficient Ligation Method in the Making of an in vitro Virus for in vitro Protein Evolution. Biol Proced Online. 2002; 4:49–54. [PubMed: 12734569]

48. Metzker ML. Sequencing technologies - the next generation. Nat. Rev. Genet. 2010; 11:31–46. [PubMed: 19997069]

49. 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]

50. Shimizu Y, Kuruma Y, Ying B-W, Umekage S, Ueda T. Cell-free translation systems for protein engineering. FEBS J. 2006; 273:4133–4140. [PubMed: 16930131]

51. Wang H, Liu R. Advantages of mRNA display selections over other selection techniques for investigation of protein-protein interactions. Expert Rev Proteomics. 2011; 8:335–346. [PubMed: 21679115]

52. Valencia CA, Ju W, Liu R. Matrin 3 is a Ca2+/calmodulin-binding protein cleaved by caspases. Biochem. Biophys. Res. Commun. 2007; 361:281–286. [PubMed: 17658460]

53. Yabu T, Kishi S, Okazaki T, Yamashita M. Characterization of zebrafish caspase-3 and induction of apoptosis through ceramide generation in fish fathead minnow tailbud cells and zebrafish embryo. Biochem. J. 2001; 360:39–47. [PubMed: 11695990]

54. Chakraborty C, Nandi SS, Sinha S, Gera VK. Zebrafish caspase-3: molecular cloning, characterization, crystallization and phylogenetic analysis. Protein Pept. Lett. 2006; 13:633–640. [PubMed: 16842121]

55. Chakraborty C, Saha G, Sarkar B, Pal S, Chatterjee TK, Sadhu AK. Caspase-3 induced apoptosis in transgenic zebrafish. Biotechnol. Lett. 2006; 28:189–196. [PubMed: 16489497]

56. Valencia CA, Bailey C, Liu R. Novel zebrafish caspase-3 substrates. Biochem. Biophys. Res. Commun. 2007; 361:311–316. [PubMed: 17643392]
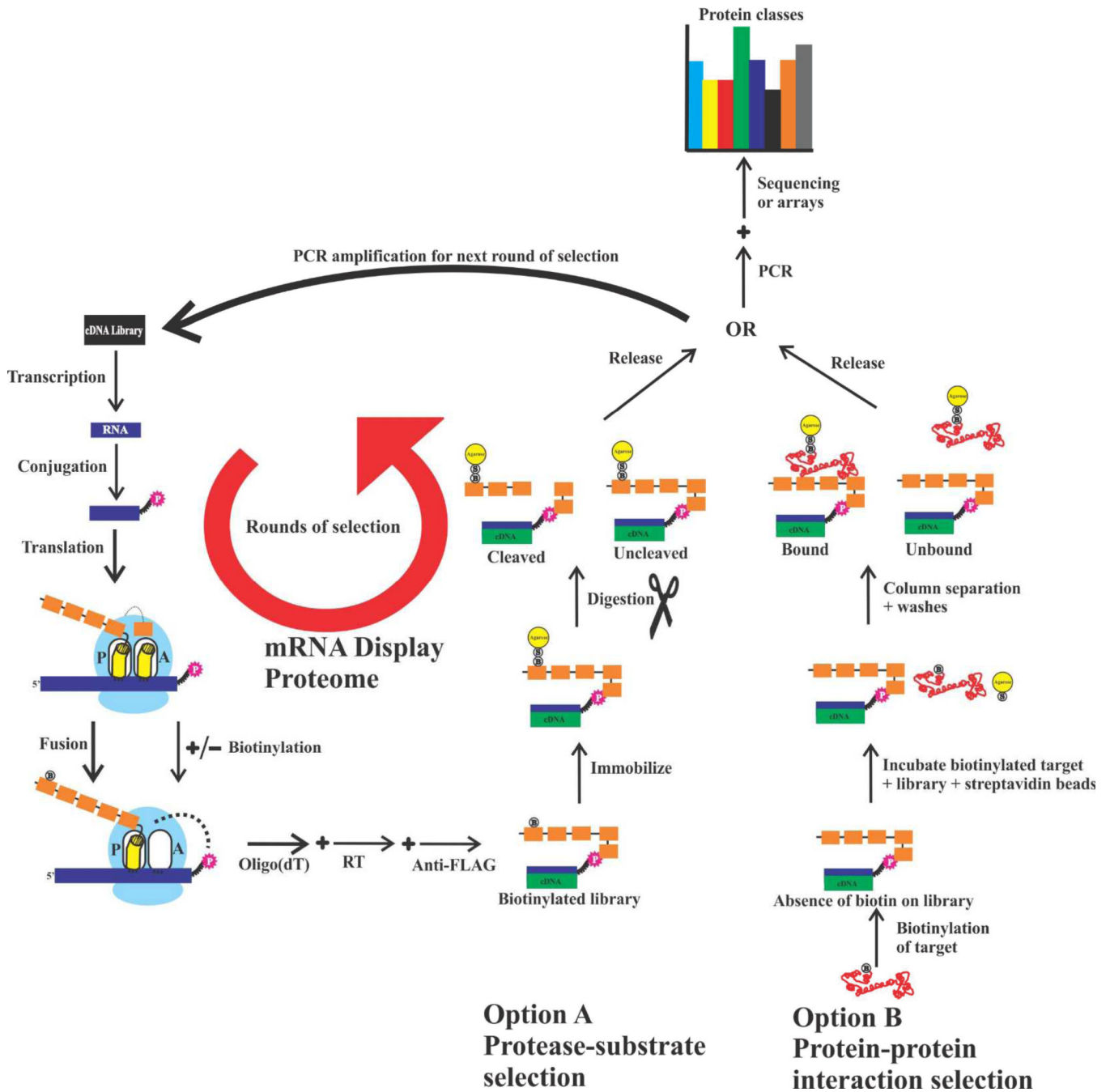
Fig. 1.
Sample selection types. Schematic representation of the selection procedure for the natural substrates of a protease of interest from an mRNA-displayed proteome library (Option A). Schematic representation of the selection procedure for the binding partners of a target of interest from an mRNA-displayed proteome library (Option B). mRNA, blue; DNA, black; protein, orange; puromycin, pink circle labeled with a P; biotin, black circle labeled with a B; streptavidin, black circle labeled with an S; agarose beads, yellow circle labeled as agarose; protease, black scissors; target protein, red.

Caspase/Granzyme 0 GB 3 1 2 4 5 6   7 8 9 10a 10b
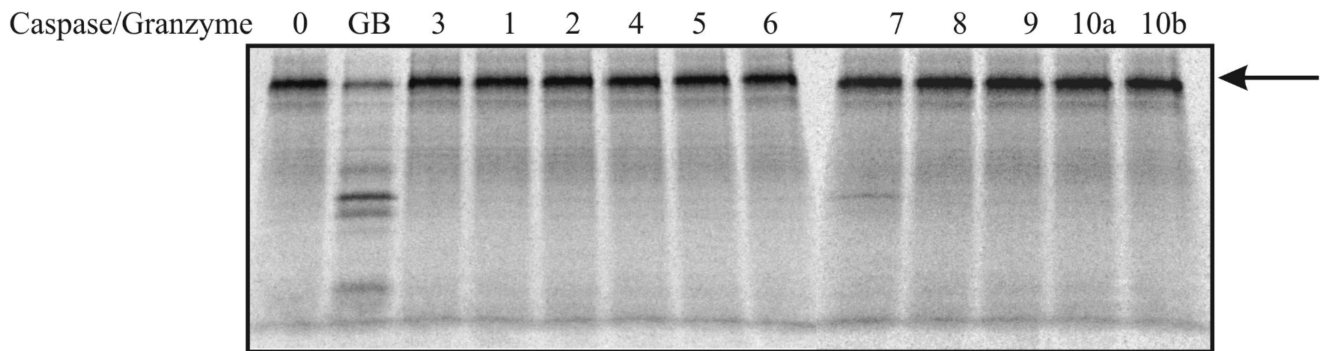
calpastatin isoform i [Homo sapiens]
NP_001035908.1

Caspase/Granzyme 0 GB 3 1 2 4 5 6   7 8 9 10a 10b
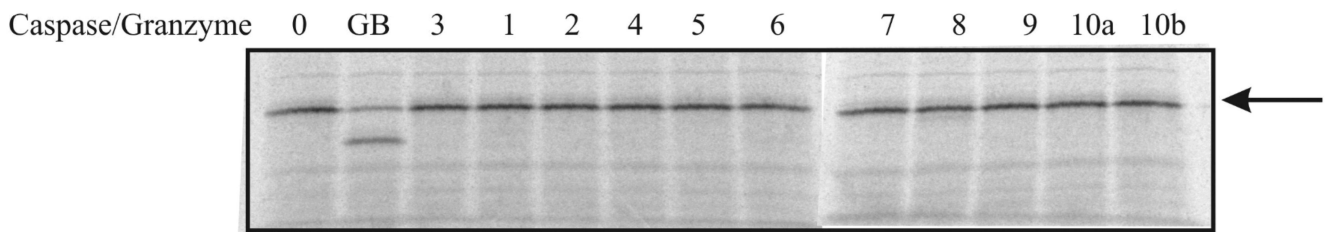
RWD domain containing 1 isoform a [Homo sapiens]
NP_057036.2

Fig. 2.
Identification of two granzyme B substrates by using a selection procedure that allows the identification of family member specific substrates. Prior to selection, the mRNA-displayed proteome library is immobilized on a streptavidin-agarose matrix and cleaved with other caspase members to remove the overlapping substrates. After stringent washing, the pre-cleared proteome library was used to enrich for the family member-specific substrates that are cleaved by the protease of interest, in this case, granzyme B (GB). The numbers 1 to 10 indicate the caspase used in the proteolytic assay. Caspase-10a and 10b refer the two isoforms of caspase-10. The ^{35}S-labeled full-length proteins are shown below (arrow).

Screening

⬇

Sequencing

⬇

Fragment Alignment and Mapping

EVNGKVRKKLMAPDISLTLDPSDGSVLSADLDESGEIDLAGLDTPSENEFEW
PEDDSIEADILAITGPEDQPGSLEVNGKVRKKLMAPDISLTLDPSDGSVLSAD
PEDDSIEADILAITGPEDQPGSLEVNGKVRKKLMAPDISLTLDPSDGSVLSADLDESGEIDLAGLDTPSENEFEW
LAITGPEDQPGSLEVNGKVRKKLMAPDISLTLDPSDGSVLSADLDESGEI
KVRKKLMAPDISLTLDPSDGSVLSADLDESGEIDLAGLDTPSE

⬇

Cleavage Site Analysis

KVRKKLMAPDISLTL**DPSD**GSVLSAD

⬇

Protein Domain Analysis

⬇

Functional Significance

Fig. 3.
Cleavage site(s) or binding domain mapping diagram. After each round of selection, the selected fragments are cloned into a plasmid and transformed bacteria are selected on an antibiotic containing agar plate. Plasmid DNAs are extracted from the bacterial colonies and sequenced. The sequence of each clone is bioinformatically translated to find the correct reading frame. Often fragments of varying lengths from the same parental sequence appear and the shortest sequence marks either the cleavage site(s), as in the diagram below, or binding domain. Thus, the functional significance of the newly mapped cleavage site(s) can be further tested by *in vivo* assays.

A

```
HCAST    9   YTGPEVSDPTSSTYIEEL-GKREVTIPPKYR-ELLAKKEGITGP------PADSSKPIGP   60
             YTGPE+ +P   +    L G+R+ T+PP YR E L KK G   P        P + KPI
ZstCAST  200 YTGPEIKEPKLKSEKGVLCGERDDTLPPGYRKEDLEKKHGAGVPAGVPEKPKEVPKPIST  259

HCAST    61  DDAIDALSSDFTCGSPT-AAGKKTE   84
             DDA+++LSS F   SP  AA KKT+
ZstCAST  260 DDALESLSSTFV--SPAQAAPKKTD   282
```


Z-stCAST FG$_{1-454}$

B

```
HCGNL1   1   LQLEVENQQNIKEERERMRANLEELRSQHNEKMEENSTLQQRLEESEGELQKNLEELFQV   60
             L+ E + + ++ E + +  A +++LR +  +    E S L+ +L ++E +L+  ++EL+QV
ZCGNL1   129 LENESKKRSDLSEMQMKAGAGVKDLRRELEQSQAECSRLRDKLSKTEADLRTTVDELYQV  188

HCGNL1   61  KMEREQRQTEIRDLQDQLSEMHDELDSAKRS--EDREKGALIEEllqakqdlqdllIAKE  118
             KME+EQ QTEIRDLQDQLSEMHDELD AK+S   + +K A++E+ +Q K DLQDLL+ +E
ZCGNL1   189 KMEKEQYQTEIRDLQDQLSEMHDELDRAKQSQTDSTDKEAMMEDFMQLKIDLQDLLLTRE  248

HCGNL1   119 ER   120
             E+
ZCGNL1   249 EQ   250
```


Z-CGNL1 FL$_{1-487}$

C

```
HSCG3    19  SSPLDNKLNVEDVDSTKNRKLIDDYDSTKSGLDHKFQDDPDGL   61
             S   + +L  +D DSTKNR+L DDYDSTK+G+D+K+QDDP+
ZSCG3    94  SDSANERLGADDADSTKNRRLADDYDSTKNGMDYKYQDDPESF   136
```


Z-SCG3 FL$_{1-478}$

D

```
HGolGB1  8    QQIDQLKERIAGLEEEKQKNKEFSQTLENEKNTLLSQISTKDGEL----KMLQEEVTKMN   63
              Q ++LKE    LE + K KE    LE+      QI  D EL    + LQE + + +
ZgolGB1  2002 QTEERLKELETSLETAENKIKELEIALEHHMEFRNEQIL--DAELTSIKQQLQESLERED  2059

HGolGB1  64   LLNQQIQEELSRVtklketaeeekddleeRLMNQLAELNGSIGNYCQDVTDAQIKNELLE  123
              +  ++I ++ +++ +L+    E E+DDLEERLMNQLA+LNGSI  Y Q+ +D++ +   ++
ZgolGB1  2060 IHKEEISKKENQLQELRMNLEAERDDLEERLMNQLAQLNGSIAGYQQEASDSRDRLTDMQ  2119

HGolGB1  124  SEMKNLKK   131
              E++ L++
ZgolGB1  2120 RELEKLER   2127
```


Z-stGOLGB FG$_{1860-2340/3080}$

Fig. 4.

In vitro proteolysis of full-length or fragment zebrafish proteins by ZCASP3, HCASP3, and DCP1 identified by homology searches. Novel human caspase-3 substrates protein sequences were BLASTed to check for mapped cleavage site homology with zebrafish proteins. Below are four zebrafish proteins that contain conserved cleavage sites (red letters).When digestion assays were performed on zebrafish proteins, the zebrafish caspase-3 (ZCASP3) efficiently digested zebrafish substrates. In contrast, when the zebrafish substrates were cleaved by the ortholog human caspase-3 (HCASP3), inefficient cleavage of all four substrates was observed, indicating a different enzyme specificity. The cleavage pattern of the zebrafish substrates following zebrafish and human caspase-3 digestions was not identical. Z in Z-stCAST is the zebrafish ortholog. FL indicates full-length protein and FG means large fragment, when the full-length protein could not be expressed.