# Chapter 2

# mRNA Display: Ligand Discovery, Interaction Analysis and Beyond

This work has been adapted from the following publication:

Takahashi, T.T., Austin, R.J. and Roberts, R.W. mRNA display: ligand discovery, interaction analysis and beyond. (2003) *Trends Biochem Sci* **28**, 159-165.

# Abstract

*In vitro* peptide and protein selection using mRNA display* enables the discovery and directed evolution of new molecules from combinatorial libraries. These selected molecules can serve as tools to control and understand biological processes, enhance our understanding of molecular interactions, and potentially treat disease in therapeutic applications. In mRNA display, mRNA molecules are covalently attached to the peptide or protein they encode. These mRNA-protein fusions enable *in vitro* selection of peptide and protein libraries of more than $10^{13}$ different sequences. mRNA display has been used to discover novel peptide and protein ligands for RNA, small molecules, and proteins, as well as to define cellular interaction partners of proteins and drugs. In addition, several unique applications are possible with mRNA display, including self-assembling protein chips and library construction with unnatural amino acids, and chemically modified peptides.

*mRNA display has been referred to as mRNA-protein fusions (1), *in vitro* virus and *in vitro* virus virion (2), and PROfusion™ technology (3).

# Introduction

Functional approaches, such as *in vitro* selection, currently provide the best means available for isolating peptides and proteins with desired chemical or biochemical properties. Over the last decade, display technologies have been essential tools in the discovery of peptide and protein ligands and in delineating *in vivo* interaction partners. The phage (4) and ribosome display systems (5) have been principally used for discovery, while the yeast two-hybrid method (6) has been used for *in vivo* interaction analysis.

Despite their power, technologies that require *in vivo* step, such as phage display and the yeast two-hybrid system face certain limitations. In phage display, libraries must be transformed into bacteria, limiting the number of possible independent sequences to $10^9$–$10^{10}$. The total number of sequences represented can be further decreased by other issues including: degradation of unfolded molecules, poor expression in the bacterial host, failure in processing to the phage surface, failure to fold in the oxidizing periplasmic space of *Escherichia Coli*, or toxicity of the gene product. Similarly, the two-hybrid system requires that the interaction partners be cloned into yeast, limiting the number of constructs examined to $10^6$-$10^7$. Additionally, in the two-hybrid approach, interactions must occur in the nucleus, limiting control over the binding stringency, appropriate binding partners, and biochemical conditions.

Totally *in vitro* techniques, such as ribosome and mRNA display, overcome many limitations of phage display and the two-hybrid system. These approaches reduce biases due to expression and routinely generate libraries of more than $10^{12}$ independent

molecules since no transformation step is required. In addition, more control can be exercised over the binding conditions as well as the stringency of selection.

**mRNA Display**

The mRNA display peptide and protein selection system provides an alternative method that can be applied to both ligand discovery and interaction analysis problems (2, 7). In this approach, encoded peptide and protein libraries are covalently fused to their own mRNA (Figure 2.1). Fusion synthesis is possible because the message can act as

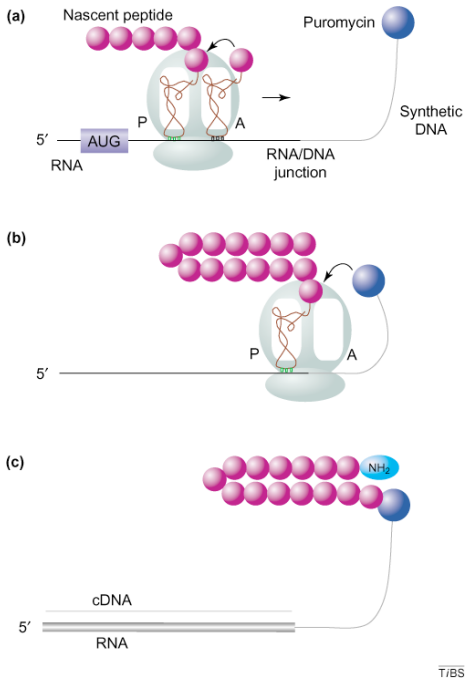


**Figure 2.1.** Formation of an mRNA-protein fusion. **(a)** mRNA (black) is ligated photochemically (11) or enzymatically (12) to a synthetic oligonucleotide (grey) containing puromycin (blue) at its 3' end. The ribosome (pale green) initiates synthesis of the template and reads in a 5'→3' direction. tRNAs (brown) and amino acids (pink) are shown in the P- and A-site of the ribosome. **(b)** Puromycin enters the ribosome attaching the template to the C-terminus of the nascent peptide. This entry occurs almost exclusively at the last or next to last codon (T. Snyder, A. Balakin, and R. W. Roberts, unpublished observation). **(c)** Reverse transcription generates cDNA (grey) that can be amplified by PCR.

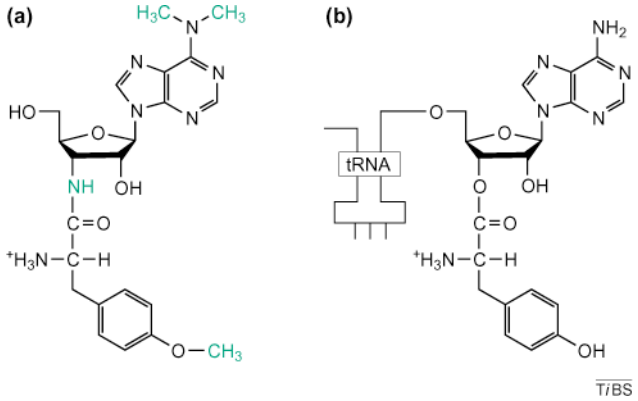


**Figure 2.2.** Puromycin **(a)** is a small molecule analog of Tyrosyl tRNA **(b)**. Differences between the two molecules are highlighted in green text.

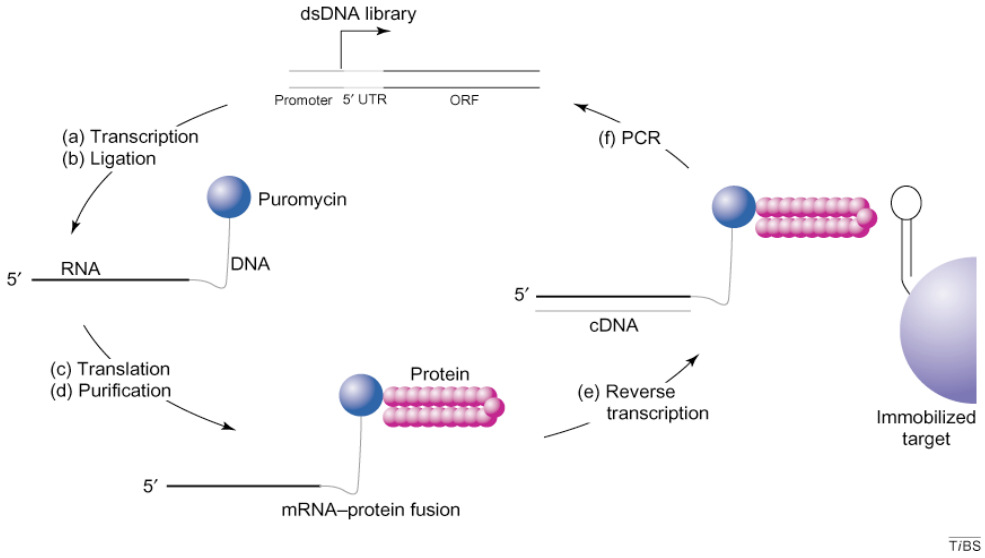


**Figure 2.3.** A typical mRNA Display selection cycle. **(a)** A library of double-stranded DNA sequences is transcribed to generate mRNA. **(b)** The mRNA is ligated to a puromycin oligonucleotide (blue) and used to program an *in vitro* translation reaction **(c)**. cDNA synthesis is performed **(e)** and the cDNA/mRNA-protein fusion is sieved using the target of interest. PCR is used to regenerate the full-length DNA construct **(f)**. For targets containing RNase or RNase H activity the cDNA can be crosslinked to the puromycin oligonucleotide to generate a cDNA-protein fusion (10).

both template and peptide acceptor if it contains a 3'-puromycin molecule. Puromycin serves as a chemically stable, small molecule mimic of aminoacyl tRNA (Figure 2.2). The selection cycle for a typical mRNA display experiment is shown in (Figure 2.3).

Detailed descriptions of experiments and protocols have been published elsewhere (8-12). Briefly, a synthetic oligonucleotide containing a 3' puromycin is ligated to the 3' end of an mRNA and the product is translated in rabbit reticulocyte lysate. The sequence present in the peptide is therefore encoded in the covalently attached mRNA, allowing the sequence information in the protein to be read and recovered after selection via reverse transcriptase (RT)-PCR. Thus, exceedingly small amounts of material can thus be amplified.

Since the original description of the mRNA display system in 1997 (1, 2), optimization has resulted in the ability to perform selection experiments on libraries containing more than $10^{13}$ molecules (8, 12, 13). Routinely 10–40% of the input mRNA template can be converted to fusion product, resulting in more than 5 x $10^{13}$ mRNA-protein fusions per milliliter of translation reaction. Overall, the mRNA display system allows libraries with sequence complexity approximately 10,000-fold that of phage display (4), $10^6$-fold over yeast display or yeast two- and three- hybrid systems (6, 14-16), and approximately $10^9$-fold over colony screening approaches (17).

In the majority of mRNA display experiments, polypeptides with relatively short chain lengths (10–110 amino acids) have been used. Larger proteins have been studied as well (e.g. λ protein phosphatase a 24 kDa enzyme (12) and β-lactamase, a 31 kDa enzyme; S. Li & R. W. Roberts unpublished observation), but these typically form fusion products with somewhat reduced efficiency. Even for such proteins, libraries can still be

readily constructed that are orders of magnitude larger than a typical phage display library. Another feature of the mRNA display constructs is that the mRNA appears to improve the solubility of the attached protein, enabling functional selection of sequences that can aggregate or are only partially soluble when expressed by themselves (see below).

# Ligand Discovery with mRNA Display

Selections to discover new peptides and proteins with desired features have now been completed. Sequences have been isolated that bind RNA, small molecules, and proteins. These results illustrate three important principles: (i) larger library size does, in fact, result in higher affinity molecules; (ii) larger libraries result in a greater diversity of sequences with similar function; and (iii) the vast number of sequences recovered after selection can be analyzed using informational techniques, such as sequence covariation analysis (see below).

### RNA-Binding Peptides

RNA binding proteins participate in regulation of transcription, splicing, and translation, and have been implicated in several diseases (18, 19). Selections for RNA-binding peptides also present a stringent functional test of mRNA display. Numerous mRNA display selections have isolated more than 100 chemically distinct RNA binding peptides (20-22). These selections demonstrate that even highly basic and unstructured molecules retain function and do not interact with the attached mRNA/cDNA hybrid. The majority of experiments have been conducted using the RNA binding domain from phage λ N protein as a model system due to its small size (22 amino acids), high affinity

(low nanomolar), and thorough characterization. Selections have resulted in numerous peptides with nanomolar affinity for their cognate target (21). The highest complexity RNA selection performed to date contained 10 random residues ($X_{10}$, where X is any of the twenty amino acids) and more than 9 x $10^{12}$ different sequences in the initial library (Figure 2.4A) (21). The selected peptides all bound the *boxB* RNA hairpin with very high affinity ($K_d$ = 0.5 to 5 nM) and most demonstrated equal or better specificity than the wild-type sequence. However, the selected peptides showed striking chemical diversity

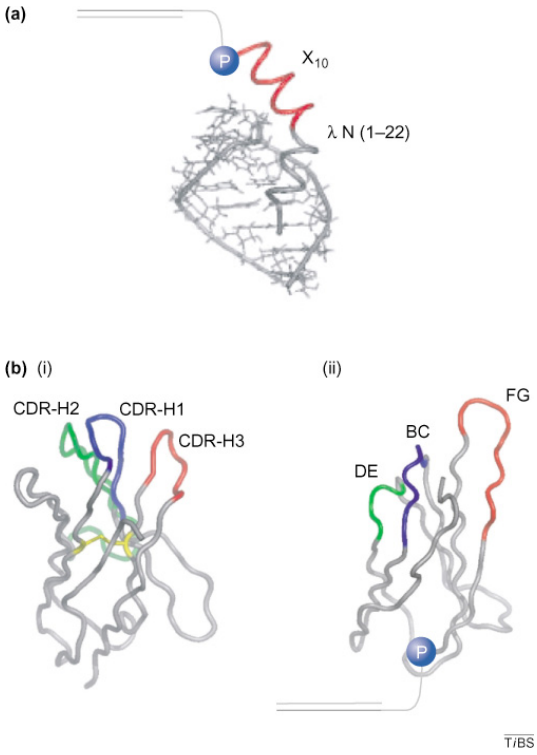


**Figure 2.4.** Examples of mRNA display libraries. **(a)** A library containing ten consecutive random residues (red) was constructed from the λ N RNA-binding protein and selected for binding RNA hairpins (21). **(b)** Structural comparison between $V_{HH}$ (i) and the 10th fibronectin type III domain ($^{10}$Fn3) (ii). Three libraries derived from $^{10}$Fn3 were constructed as antibody mimics of $V_{HH}$ and selected for binding to TNF-α and leptin (35). Residues of $^{10}$Fn3 that were randomized are shown in color. Abbreviations: CDR, complementarity-determining region; P, puromycin.

and bore little resemblance to wild-type. Only a single Arg at position 15 (glutamine in the wild-type) showed any significant conservation. Despite the lack of homology, sequence covariation analysis indicated that the molecules fold into helices, showing correlations between adjacent residues (i to i + 1) and residues located one turn away (i to i + 3 and i to i + 4) (20).

**ATP Aptamers**

Primordial proteins presumably evolved from random sequences and it is probable that one of these first proteins bound ATP. Keefe and Szostak (23) attempted to isolate a modern relative of these prebiotic proteins by selecting for ATP binding using a 109-mer protein containing an 80-amino acid random region. Libraries containing such large numbers of random positions present special problems since the probability of encountering a frame shift or stop codon can become substantial (12, 24). To solve these problems, the Szostak group utilized mRNA display to preselect library fragments for readability, selecting for the presence of N- and C-terminal epitope tags (13). The readable fragments were then digested and assembled into full-length libraries that contained greatly improved open reading frames (ORFs). After 8 rounds of selection for ATP binding, 4 distinct protein sequence families could be discerned (23). Further rounds of selection, combined with mutagenesis, resulted in a clone (18-19) that bound ATP with high affinity ($K_d$ = 100 nM) and could discriminate ATP from other nucleotide triphosphates with up to 2000-fold specificity. These protein aptamers contain a conserved Cys-Xaa-Xaa-Cys (CXXC) motif and function in a metal-dependent manner. The fact that the aptamer uses metals might indicate that chelation provides a simple way to create stable, functional protein structures, which is consistent with the large energy

seen for protein-metal interactions (25). One feature of these aptamers is that only a fraction of each clone appears folded and functional; the proteins themselves tend to aggregate when expressed as free proteins. Thus, selection of these proteins was probably facilitated by the improved solubility imparted by the mRNA/cDNA tail, and argues that such sequences would not be found in a typical phage display selection. The fact that the functional clones are not well behaved likely reflects the relative paucity of proteins that are both folded *and* functional in the vastness of sequence space.

The structure of one of these aptamers has been recently determined by X-ray crystallography. By screening a few different constructs, Sollazzo and coworkers were able to obtain single crystals of one aptamer they termed artificial nucleotide-binding protein (ANBP) (26). The protein folds into a novel three-stranded β-sheet flanked by two α-helices and binds ATP in a manner similar to natural ATP binding proteins. It also exhibits biophysical characteristics reminiscent of natural proteins such as good NMR chemical shift dispersion and thermal unfolding.

Keefe and Szostak estimate that one in $10^{12}$ molecules in their initial library have the ability to bind ATP – approximately the same fraction seen for ATP-binding RNA aptamers (27). This result is somewhat surprising given the greater chemical diversity of proteins (twenty sidechains) relative to nucleic acids (four sidechains). While functionally impoverished, nucleic acid aptamers may benefit from the ease of forming higher-order structures through simple base-pairing interactions, in contrast to proteins, which require a hydrophobic core for folding. It remains an open question if catalytic proteins can also be found with similar frequencies to their nucleic acid counterparts.

**Streptavidin Aptamers**

Szostak and coworkers also created long open reading frames for a binary patterned library (13). This library contained a random region of 87-88 amino acids with an initial complexity of ~$10^{13}$ sequences, and was assembled from two distinct 11 amino acid segments containing hydrophobic and polar amino acid patterning that result in either amphipathic α-helices or β-strands (28). mRNA display selections against streptavidin resulted in a number of sequences that bound streptavidin with nanomolar affinity ($K_d$ ~5 nM ) (29), and bind 200- to 2,200-fold better than the *Strep*-tag II peptide obtained previously by phage display (30, 31).

Although the library had been patterned to form helices and sheets in reading frame one, all of the selected molecules were shifted into reading frame three, effectively eliminating the patterning. The shifted frame seems to have been greatly preferred due to the presence of His-Pro-Gln (HPQ) tripeptide sequences. The HPQ peptide represents the minimal core of the *Strep*-tag II peptide and has been shown to bind streptavidin (30, 31). Frame one of the patterned library contained very few HPQ sequences (1/45,000 clones), owing to the library design, whereas in the third frame, HPQ peptides were present in 1/64 sequences.

The majority of the sequences contained at least one HPQ motif, one similar tripeptide motif (e.g., HPQ, His-Pro-Ala (HPA), and Leu-Pro-Gln (LPQ)) and do not appear to contain any disulfide bonds. A 38 amino acid sequence, termed the "SBP-tag," has been used for one-step affinity purification on streptavidin agarose and western blot detection using streptavidin-horseradish peroxidase for visualization (32). Despite frameshifting, the patterned library still contained ~10,000-fold greater sequence

complexity than a standard phage display selection, likely leading to the high affinity of the resulting aptamers. Finally, this experiment demonstrates the difficulty in designing random libraries with imposed structural features *a priori*.

**TNF-α Aptamers using the $^{10}$Fn3 Domain**

Monoclonal antibodies are useful both as a biochemical tool and as potential therapeutics. mRNA display has been used to isolate novel antibody mimetics based on a fibronectin domain. The tenth type III domain of human fibronectin ($^{10}$Fn3) displays an Arg-Gly-Asp (RGD) sequence involved in cell-surface recognition by integrins (33). The $^{10}$Fn3 domain has a similar β-sheet architecture to antibody $V_H$ domains, with three structurally analogous loops (Figure 2.4B). The antibody-like structure, exposure to the immune system, small size (94 amino acids), lack of disulfide bonds, high bacterial expression levels, and high stability ($T_m$ = 90°C) all make the make the $^{10}$Fn3 domain an excellent potential scaffold. However, previous attempts to isolate $^{10}$Fn3 derivatives using phage display resulted in molecules with only modest affinity ($IC_{50,\ ubiquitin}$ = 5 μM) and relatively non-specific binding (34).

Xu et al. constructed three libraries based on $^{10}$Fn3, randomizing either one loop (libraries 1 and 2) or all three loops simultaneously (library 3) (35). An mRNA display selection was then performed against tumor necrosis factor-α (TNF-α). After 9-10 rounds of selection, diverse, high affinity ($K_d$ = 1 – 24 nM), and high specificity ligands were isolated, primarily originating from library 3. Further selection for a total of 14 rounds resulted in clones with sub-nanomolar affinity ($K_d$ = 90-110 pM). Returning to round 8, mutagenic PCR was added to the selection cycle, duplicating the affinity

maturation process of antibodies. Further rounds of selection resulted in a clone with very high affinity ($K_d$ = 20 pM). While less stable than wild-type $^{10}$Fn3, the best clone (12.21) nonetheless was monomeric and showed good expression and protease resistance at 30°C. Immobilized versions of a round 9 clone (9.12) could be used to capture TNF-α from a solution of 10% fetal bovine serum, demonstrating the high specificity of these reagents, even when immobilized on a solid support.

# Specificity and Interaction Analysis

### Epitope Recovery

In addition to enriching sequences containing a known epitope (7), mRNA display can also be used to determine which sequences are critical for recognition. Baggio et al. used two random libraries to investigate the specificity of peptides binding the anti-*c-myc* antibody 9E10 (36) and bovine trypsin (37). Selection against the 9E10 antibody with a 27 random residue ($X_{27}$) library revealed a consensus sequence x[Q/E]xLISExx[L/M] (the *c-myc* tag is EQKLISEEDLN), demonstrating that the Leu-Ile-Ser-Glu (LISE) sequence was the core element recognized by the antibody. In the same work, a six random residue library (positions 3-8) was created using the *Ecballium elaterium* trypsin inhibitor two protein (EETI-II) as a scaffold (37). EETI-II, a 28-residue protein with three disulfide bonds, is a member of the knottin family and inhibits bovine trypsin via interaction at positions 3-8 (38). Selection against trypsin yielded a sequence consensus of Pro-Arg-Xaa-Leu-Xaa-Xaa (PRxLxx), with 20% of the selected clones matching the wild-type sequence of Pro-Arg-Ile-Leu-Met-Arg (PRILMR).

mRNA display has also been applied to define a recognition epitope for the oncogenic v-abl tyrosine kinase, which is a target of great biological and therapeutic interest (39). Initial experiments demonstrated that mRNA-peptide fusions containing a v-abl phosphorylation site (the tyrosine residue in EAI**Y**AAPFAKKK) could be phosphorylated by the v-abl kinase and immunoprecipitated with α4G10, an anti-phosphotyrosine monoclonal antibody. Libraries of the form $GCGGX_5\mathbf{Y}X_5GCG$ were subjected to phosphorylation with v-abl and precipitation with α4G10. The majority of clones contained an [Ile/Leu/Val]-Tyr-$Xaa_{1-5}$-[Pro/Phe] ([I/L/V]$Yx_{1-5}$[P/F]) consensus. Interestingly, despite the sequence variations, the kinase effectively phosphorylated all 12 of the consensus clones, indicating a broader specificity than previously thought.

**Cellular Interaction Partners**

mRNA display libraries constructed from cDNA offer the potential of isolating biologically relevant interaction partners. Hammond and coworkers used a random priming approach to create mRNA display libraries from several different human tissues (40). This approach yields libraries of various lengths and in three reading frames, and it also allows the experimenter construct libraries with tissue-specific primer tags. After selection, PCR using these primers can be used to deconvolute the library and obtain binders from specific tissues. Sieving cellular libraries against the anti-apoptotic protein Bcl-$X_L$ resulted in isolation of over 20 different proteins including the known interaction partners Bim, Bax and BCL2L12. The diversity in the cellular mRNA display libraries means that hundreds to millions of fragments of various lengths will be present from each gene. In that vein, the Bcl-$X_L$ selection demonstrated that alignment of multiple positive

clones is equivalent to typical deletion analysis, providing a clear indication of the sequence boundaries necessary for recognition.

Cellular libraries may also be used to characterize and discover cellular proteins or receptors that interact with a drug of interest. McPherson et al. used the immunosuppressive drug FK506 as a target for cellular libraries (41). This work resulted in isolation of the known target, FK506 binding protein (FKBP) and defined a region within FKBP that was necessary and sufficient for interaction with the drug (41).

# Unique Applications of mRNA Display

### Self-Assembling Protein Microarrays

The mRNA-protein fusions used in mRNA display can also be used for high-throughput screening applications. Protein chips offer the promise of quick analysis of the expressed protein content in a sample and performing *in vitro* interaction analysis. Weng et al. demonstrated that a standard DNA chip could be converted to a protein chip by hybridization of mRNA-protein fusions (Figure 2.5) (3). mRNA-protein fusions coding for the MYC, FLAG, or HA11 epitopes were synthesized and incubated with a DNA chip. The chip was imprinted with DNA complementary to a unique (MYC or FLAG or HA11) or common (5' or 3') nucleic acid portion of the fusions. Hybridization of the fusions to complementary DNA directs the self-directed assembly of the protein chip. The experiments demonstrate that at least for antibody-epitope interactions, these protein arrays preserve the functionality of the displayed proteins, present them in a uniform orientation, and have sub-attomole detection limits.

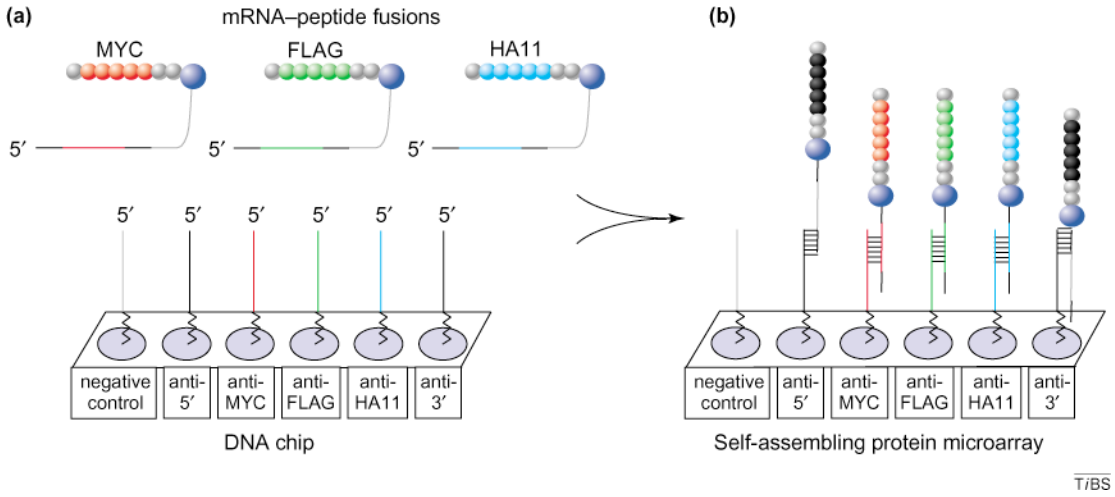


**Figure 2.5.** A self-assembling protein chip. **(a)** A mixture of mRNA-protein fusions containing either the MYC (red), FLAG (green), or HA11 epitope (light blue) (all three, black) epitope was incubated with a standard DNA chip. The nucleic acid component directs the fusions to regions on the chip containing complementary DNA. **(b)** A DNA complementary to the 5' or 3' sequence hybridizes to all three fusions, whereas the anti-MYC DNA will only isolate fusions containing the MYC epitope. This results in spatially addressable peptide micro arrays that can be recognized by monoclonal antibodies (35).

### Non-natural Libraries

Phage display and the yeast two-hybrid system contain an obligate *in vivo* step and thus are generally limited to display only the 20 natural amino acids. Expansion of the amino acid alphabet would increase the chemical diversity that can be displayed and facilitate the discovery of greatly improved ligands. Unnatural amino acids can be introduced either during translation or post-translationally.

Recently, Li et al. demonstrated that the suppressor tRNA strategy for incorporating unnatural amino acids (42, 43) could be used to create mRNA display libraries bearing biocytin, a biotinylated lysine derivative (44) (Figure 2.6). After selection, the library was enriched in sequences containing the amber stop codon (TAG), which was suppressed by biocytin. The combination of these two powerful technologies

increases the chemical diversity that can be displayed and should facilitate discovery of ligands with improved affinity, specificity, stability or reactivity.

Using the nonsense suppression strategy, a maximum of three unnatural amino acids could be incorporated – one for each stop codon. However, recent work has shown that sense suppression is also possible, allowing the incorporation of up to 64 unnatural amino acids and essentially rewriting the genetic code. Frankel and Roberts used a biocytin-charged tRNA to select for a GUA codon that would allow unnatural amino acid incorporation with efficiencies comparable to that of nonsense suppression (45). Using the GUA codon, unnatural polymers (termed "encodamers") of N-methyl phenylalanine (N-MePhe) were synthesized and exhibited marked protease resistance (46).

Other work to incorporate unnatural amino acids has focused on post-translationally derivatizing libraries. A library bearing a pendant penicillin sidechain was used to select for peptides that increased the inhibitory activity of the attached penicillin to penicillin binding protein 2a (PBP2a) by more than 100-fold (47). This strategy should be applicable to improve a variety of small molecule compounds, and also increase the chemical diversity of mRNA display libraries.

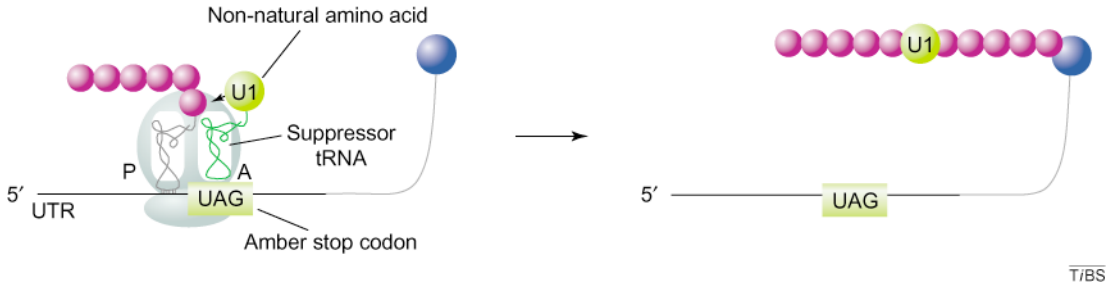


**Figure 2.6.** Inserting non-natural residues into mRNA display libraries. *In vitro* nonsense suppression using a chemically aminoacylated suppressor tRNA was used to insert biocytin into mRNA display libraries and select for the presence of the unnatural residue (44).

# Conclusions

Techniques for performing mRNA display are now well established and allow facile synthesis and selection of mRNA-protein fusion libraries (8, 12, 13). Completed selections demonstrate that mRNA display is a powerful tool for both ligand discovery and interaction analysis. Notable features of the resulting ligands are high affinity (nanomolar to picomolar) and sriking sequence diversity present (21, 29, 35). The *in vitro* nature of the system provides a unique opportunity for *in vitro* affinity maturation and evolution (23, 35), inclusion of non-natural residues (44), chemical derivatization of libraries, and the opportunity for *in vitro* recombination experiments (48). Future applications point toward the isolation of new catalysts and the creation of libraries composed entirely of unnatural sidechains or non-peptidic backbones.

**Acknowledgements**

This work supported by NIH Grant GM01416 (R.W.R), NSF Grant 9876246 (R.W.R), and NIH training grant T32 GM08501. We thank S.R. Starck and Dr. A. Frankel for helpful comments on the manuscript.

# References


1. Roberts, R.W. and Ja, W.W. *In vitro* selection of nucleic acids and proteins: What are we learning? (1999) *Curr Opin Struct Biol* **9**, 521-529.
2. Nemoto, N., Miyamoto-Sato, E., Husimi, Y. and 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. (1997) *FEBS Lett*. **414**, 405-408.
3. Weng, S., Gu, K., Hammond, P.W., Lohse, P., Rise, C., Wagner, R.W., Wright, M.C. and Kuimelis, R.G. Generating addressable protein microarrays with PROfusion™ covalent mRNA-protein fusion technology. (2002) *Proteomics* **2**, 48-57.
4. Smith, G. and Petrenko, V. Phage display. (1997) *Chem. Rev*. **97**, 391-410.
5. Hanes, J. and Pluckthun, A. *In vitro* selection and evolution of functional proteins by using ribosome display. (1997) *Proc Natl Acad Sci U S A* **94**, 4937-4942.
6. Fields, S. and Song, O.K. A novel genetic system to detect protein-protein interactions. (1989) *Nature* **340**, 245-246.
7. Roberts, R.W. and Szostak, J.W. RNA-peptide fusions for the *in vitro* selection of peptides and proteins. (1997) *Proc. Natl. Acad. Sci. U.S.A.* **94**, 12297-12302.
8. Barrick, J.E., Takahashi, T.T., Balakin, A. and Roberts, R.W. Selection of RNA-binding peptides using mRNA-peptide fusions. (2001) *Methods* **23**, 287-293.
9. Kurz, M., G., K. and Lohse, P.A. An efficient synthetic strategy for the preparation of nucleic acid-encoded peptide and protein libraries for *in vitro* evolution protocols. (2000) *Molecules* **5**, 1259-1264.

10. Kurz, M., Gu, K., Al-Gawari, A. and Lohse, P.A. cDNA-protein fusions: covalent protein-gene conjugates for the *in vitro* selection of peptides and proteins. (2001) *Chembiochem* **2**, 666-672.

11. Kurz, M., Gu, K. and Lohse, P.A. Psoralen photo-crosslinked mRNA-puromycin conjugates: a novel template for the rapid and facile preparation of mRNA-protein fusions. (2000) *Nucleic Acids Res.* **28**, E83.

12. Liu, R., Barrick, J.E., Szostak, J.W. and Roberts, R.W. Optimized synthesis of RNA-protein fusions for *in vitro* protein selection. (2000) *Methods Enzymol.* **318**, 268-293.

13. Cho, G., Keefe, A.D., Liu, R., Wilson, D.S. and Szostak, J.W. Constructing high complexity synthetic libraries of long ORFs using *in vitro* selection. (2000) *J. Mol. Biol.* **297**, 309-319.

14. Boder, E.T. and Wittrup, K.D. Yeast surface display for screening combinatorial polypeptide libraries. (1997) *Nat. Biotechnol.* **15**, 553-557.

15. Gyuris, J., Golemis, E., Chertkov, H. and Brent, R. Cdi1, a human G1 and S phase protein phosphatase that associates with Cdk2. (1993) *Cell* **75**, 791-803.

16. Sengupta, D.J., Zhang, B.L., Kraemer, B., Pochart, P., Fields, S. and Wickens, M. A 3-hybrid system to detect RNA-protein interactions *in vivo*. (1996) *Proc. Natl. Acad. Sci. U.S.A.* **93**, 8496-8501.

17. Arnold, F.H. and Volkov, A.A. Directed evolution of biocatalysts. (1999) *Curr. Opin. Chem. Biol.* **3**, 54-59.

18. Burd, C.G. and Dreyfuss, G. Conserved structures and diversity of functions of RNA-binding proteins. (1994) *Science* **265**, 615-621.

19. Draper, D.E. Themes in RNA-protein recognition. (1999) *J Mol Biol* **293**, 255-270.

20. Barrick, J.E. and Roberts, R.W. Sequence analysis of an artificial family of RNA-binding peptides. (2002) *Protein Science* **11**, 2688-2696.

21. Barrick, J.E., Takahashi, T.T., Ren, J., Xia, T. and Roberts, R.W. Large libraries reveal diverse solutions to an RNA recognition problem. (2001) *Proc. Natl. Acad. Sci. U.S.A.* **98**, 12374-12378.

22. Xia, T., Frankel, A., Takahashi, T.T., Ren, J. and Roberts, R.W. Context and conformation dictate function of a transcription antitermination switch. (2003) *Nat Struct Biol* **10**, 812-819.

23. Keefe, A.D. and Szostak, J.W. Functional proteins from a random-sequence library. (2001) *Nature* **410**, 715-718.

24. Labean, T.H. and Kauffman, S.A. Design of synthetic gene libraries encoding random sequence proteins with desired ensemble characteristics. (1993) *Protein Sci.* **2**, 1249-1254.

25. Kuntz, I.D., Chen, K., Sharp, K.A. and Kollman, P.A. The maximal affinity of ligands. (1999) *Proc. Natl. Acad. Sci. U.S.A.* **96**, 9997-10002.

26. Lo Surdo, P., Walsh, M.A. and Sollazzo, M. A novel ADP- and zinc-binding fold from function-directed *in vitro* evolution. (2004) *Nat Struct Mol Biol* **11**, 382-383.

27. Sassanfar, M. and Szostak, J.W. An RNA motif that binds ATP. (1993) *Nature* **364**, 550-553.

28. Kamtekar, S., Schiffer, J.M., Xiong, H., Babik, J.M. and Hecht, M.H. Protein design by binary patterning of polar and nonpolar amino acids. (1993) *Science* **262**, 1680-1685.

29. Wilson, D.S., Keefe, A.D. and Szostak, J.W. The use of mRNA display to select high-affinity protein-binding peptides. (2001) *Proc. Natl. Acad. Sci. U.S.A.* **98**, 3750-3755.

30. Schmidt, T., Koepke, J., Frank, R. and Skerra, A. Molecular interaction between the Strep-tag affinity peptide and its cognate target, streptavidin. (1996) *J. Mol. Biol.* **255**, 753-766.

31. Voss, S. and Skerra, A. Mutagenesis of a flexible loop in streptavidin leads to higher affinity for the *Strep*-tag II peptide and improved performance in recombinant protein purification. (1997) *Protein Eng.* **10**, 975-982.

32. Keefe, A.D., Wilson, D.S., Seelig, B. and Szostak, J.W. One-step purification of recombinant proteins using a nanomolar-affinity streptavidin-binding peptide, the SBP-Tag. (2001) *Protein Expr. Purif.* **23**, 440-446.

33. Schwarzbauer, J.E. and Sechler, J.L. Fibronectin fibrillogenesis: a paradigm for extracellular matrix assembly. (1999) *Curr. Opin. Cell. Biol.* **11**, 622-627.

34. Koide, A., Bailey, C.W., Huang, X. and Koide, S. The fibronectin type III domain as a scaffold for novel binding proteins. (1998) *J. Mol. Biol.* **284**, 1141-1151.

35. Xu, L., Aha, P., Gu, K., Kuimelis, R.G., Kurz, M., Lam, T., Lim, A.C., Liu, H., Lohse, P.A., Sun, L., Weng, S., Wagner, R.W. and Lipovsek, D. Directed

evolution of high-affinity antibody mimics using mRNA display. (2002) *Chem. Biol.* **9**, 933-942.

36. Evan, G.I., Lewis, G.K., Ramsay, G. and Bishop, J.M. Isolation of monoclonal antibodies specific for human c-myc proto-oncogene product. (1985) *Mol. Cell. Biol.* **5**, 3610-3616.

37. 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. Identification of epitope-like consensus motifs using mRNA display. (2002) *J. Mol. Recognit.* **15**, 126-134.

38. Le-Nguyen, D., Heitz, A., Chiche, L., Castro, B., Boigegrain, R.A., Favel, A. and Coletti-Previero, M.A. Molecular recognition between serine proteases and new bioactive microproteins with a knotted structure. (1990) *Biochimie* **72**, 431-435.

39. Cujec, T.P., Medeiros, P.F., Hammond, P.W., Rise, C. and Kreider, B.L. Selection of v-Abl tyrosine kinase substrate sequences from randomized peptide and cellular proteomic libraries using mRNA display. (2002) *Chem. Biol.* **9**, 253-264.

40. Hammond, P.W., Alpin, J., Rise, C.E., Wright, M. and Kreider, B.L. *In vitro* selection and characterization of Bcl-X(L)-binding proteins from a mix of tissue-specific mRNA display libraries. (2001) *J. Biol. Chem.* **276**, 20898-20906.

41. Mcpherson, M., Yang, Y., Hammond, P.W. and Kreider, B.L. Drug receptor identification from multiple tissue using cellular-derived mRNA display libraries. (2002) *Chem. Biol.* **9**, 691-698.

42. Noren, C.J., Anthony-Cahill, S.J., Griffith, M.C. and Schultz, P.G. A general method for site-specific incorporation of unnatural amino acids into proteins. (1989) *Science* **244**, 182-188.

43. Saks, M.E., R., S.J., Nowak, M.W., Kearney, P.C., Du, F., Abelson, J.N., Lester, H.A. and Dougherty, D.A. An engineered *Tetrahymena* $tRNA^{Gln}$ for *in vivo* incorporation of unnatural amino acids into proteins by nonsense suppression. (1996) *J. Biol. Chem.* **271**, 23169-23175.

44. Li, S., Millward, S. and Roberts, R.W. Applying an Unnatural Strategy to mRNA Display Libraries. (2002) *J. Am. Chem. Soc.* **124**, 9972-9973.

45. Frankel, A. and Roberts, R.W. *In vitro* selection for sense codon suppression. (2003) *Rna* **9**, 780-786.

46. Frankel, A., Millward, S.W. and Roberts, R.W. Encodamers: unnatural peptide oligomers encoded in RNA. (2003) *Chem Biol* **10**, 1043-1050.

47. Li, S. and Roberts, R.W. A novel strategy for *in vitro* selection of peptide-drug conjugates. (2003) *Chem Biol* **10**, 233-239.

48. Stemmer, W.P.C. Rapid evolution of a protein *in vitro* by DNA shuffling. (1994) *Nature* **370**, 389-390.