Chapter 1

mRNA display – a novel combinatorial peptide library

Abstract: mRNA display provides a strategy to make up to 10^{13} unique peptides that fuse with their own genes via a covalent linkage. *In vitro* selection using mRNA display libraries has been applied to ligand discovery and analysis. Compared to other *in vitro* selection methodologies, such as phage- and surface-display libraries, mRNA display libraries have several unique features, including the incorporation of non-native residues, the chemical derivatization of libraries, and the feasibility of mutagenic PCR and DNA recombination. These libraries should facilitate the discovery of novel ligands with functionalities beyond those provided by the 20 naturally occurring amino acids.

1.1 Introduction

1. Biological and chemical combinatorial peptide libraries

Combinatorial peptide libraries have been applied to select many functional peptides, such as receptor agonists[1] or antagonists[2], antibody epitopes[3,4], and enzyme inhibitors[5]. During the last decade, many powerful techniques to construct these libraries have been developed and provided various means for the isolation of peptides with desired chemical or biochemical properties. In general, combinatorial peptide libraries can be made from biological systems or chemical synthesis. Phage-display libraries[6,7], as well as ribosome-display[8], cell surface-display[9] and other biological libraries, share the common feature that each peptide is somehow associated with its encoding gene, allowing iterative cycles of enrichment and amplification. The identity of functional peptides from these libraries can be obtained directly by sequencing their tethered genes. Furthermore, some techniques, such as mRNA-display[10,11], can generate libraries containing up to 10^{13} unique peptides without any technical hurdles. Synthetic peptide libraries, by contrast, are unable to be amplified for repeating cycles of selection. The identification of active molecules also requires tedious de-convolution procedures[12] or sophisticated encoding-decoding schemes[13]. The complexity of synthetic libraries, for example, the one-bead-one-compound library, is usually only about 10^6. Nevertheless, the synthetic library can contain any unnatural amino acid residues, including fluorescent moieties, affinity tags and D-enantiomers[14,15]. The biological libraries, on the other hand, are limited to 20 naturally occurring amino acids. Thus, it would be a great advantage if unnatural amino acid residues could be introduced into a biologically generated library. The majority of this thesis covers several strategies to construct these libraries containing

unnatural residues. The application of these libraries to the isolation of high-affinity and high-specificity ligands is also discussed.

2. Common peptide libraries made from biological systems

Phage display, the most commonly used method for creating peptide biological libraries, is based on the expression of coat protein fusion on the surface of bacteriophage[6,16]. One of the coat proteins of the phage M13, termed gp13, is fused with N-terminal recombinant peptides encoded by synthetic random oligonucletides inserted at the 5'-end of the gene (gIII). The phages displaying various peptides on their surface can then be subjected to affinity selection. Those phages remaining bound to the immobilized targets after extensive washing are recovered and used to infect the bacteria for the next cycle of selection. Cell surface display libraries are very similar to phage display libraries in that peptides are expressed as a fusion to a membrane protein on the surface of bacteria or yeast.

In peptide-on-plasmid libraries, peptides are fused with a DNA-binding protein, which recognizes a specific region of plasmid containing the library sequences to form a stable non-covalent complex. During selection, the enriched library is amplified by liberating the plasmid for polymerase chain reaction.

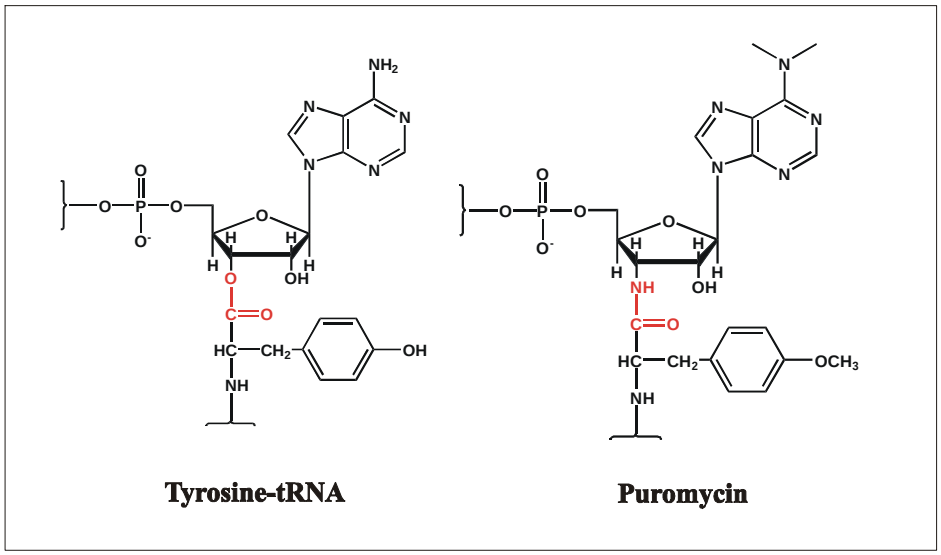
Ribosome display is another distinct technique for the construction of peptide libraries. The peptides are linked to their own messenger RNA via the formation of a tertiary mRNA-ribosome-peptide complex at low temperature and high magnesium concentrations.

Phage-, surface-, and plasmid-display libraries all require an *in vivo* step where the transformation efficiency limits the complexity of these libraries to approximately 10^9 members. Degradation of unfolded molecules, poor expression in the bacterial host, failure of processing to the phage surface, and failure to fold in the oxidizing periplasm, may further decrease the effective sequence representation. This *in vivo* step also makes it hard to manipulate the library sequences during selection through mutagenic PCR and DNA recombination. Ribosome display can achieve 10^{13} complexity easily. However, the mRNA-ribosome-peptide complex can only be maintained at very specific conditions, confining the system to applications that can be performed within a narrow range of thermal and chemical parameters.

The mRNA-display peptide library, originally described in 1997, has provided a novel approach for the selection of functional peptides and proteins with desired properties[10]. In this technique, messenger RNA templates are modified with a 3'-puromycin, a common translational inhibitor. Puromycin is a structural mimic of a 3'-tyrosinylated nucleotide, differing only in the replacement of unstable ester linkage between tyrosine and 3'-hydroxyl group by a stable amide bond (Figure 1.1). During translation, tyrosine-charged tRNA enters the ribosome, adds the tyrosine into the nascent peptide, and is released from the ribosome in the uncharged form. Puromycin, instead, stops the extension of the peptide, once it adds itself to the terminus of elongating peptide chain. Thus, translation of mRNA templates containing 3'-puromycin in rabbit retic lysate yields up to 10^{13} unique peptides fused with their encoding genes, serving as both templates and functional moieties[17] (Figure 1.2). The sequence information of the

peptides is stored in the covalently attached mRNA, and can be amplified for the recovery of the enriched library during affinity selection (Figure 1.3).

Figure 1.1 Chemical structure comparison between tyrosine-tRNA and puromycin.



The mRNA display library has certain advantages over other techniques in that 1) it is a fully *in vitro* strategy, and 2) the peptide is covalently linked to its encoding mRNA. Several strategies that are incompatible with other techniques, such as nonsense suppression, and chemical derivatization, have been demonstrated in mRNA-peptide fusion and considerably extend the application of mRNA display libraries. The optimization of fusion formation also makes it routine to produce more than $2x10^{13}$ fusions per milliliter of translation reaction, representing 10^4-fold increase in complexity over phage display, 10^6-fold over surface display, and 10^9 over colony screening approaches.

Figure 1.2 Process to prepare a mRNA display library. The mRNA template is modified with a 3'-puromycin molecule. During translation, puromycin can enter the ribosome and add to the nascent peptide chain to form a covalent link between mRNA template and its encoding peptide.

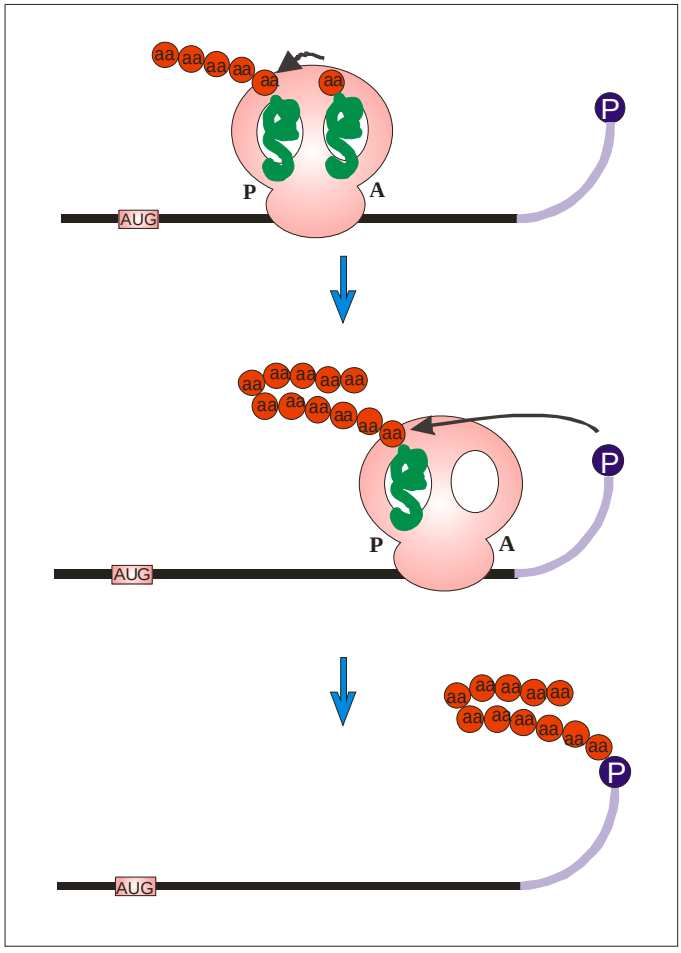
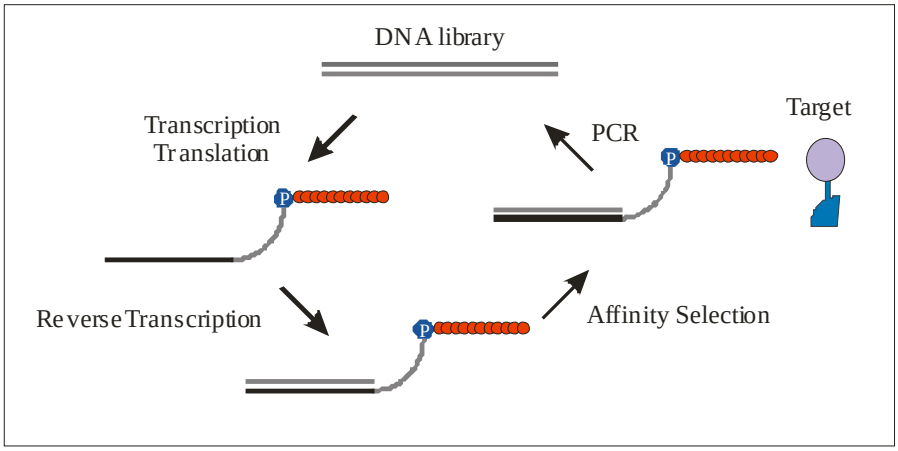


Figure 1.3 Affinity selection scheme with mRNA display library. The synthetic DNA library is transcribed to make mRNA, which is modified with a 3'-puromycin. The engineered mRNA template is translated to prepare mRNA-peptide fusion. After reverse transcription that generates cDNA·mRNA duplex, the fusion library is subjected to affinity selection against targets of interest. The enriched library is amplified by PCR for the next cycle of selection.



3. Applications of mRNA display libraries

mRNA display libraries have been used to identify novel ligands for many targets of interest, including ATP-binding aptamers[18], antibody epitopes[19], and RNA-binding peptides[20]. These results have demonstrated that diverse ligands can be selected from high-complexity peptide libraries.

a. ATP aptamers

An 80-residue peptide library was prepared from synthetic oligonucleotides by Keefe and Szostak to select ATP binding peptides[18]. The construction of libraries containing large numbers of random positions imposes several problems, including: 1)

the longer oligonucleotides are more prone to deletion and insertion during synthesis, and 2) the theoretical 3% rate of TAG stop codon incorporation at each random position (NNS) can dramatically reduce the fusion formation efficiency. To overcome these issues, Keefe adopted a prescreening procedure by selecting short fragments for the presence of N- and C-terminal epitope tags. This prescreen step considerably reduced the oligos containing stop codons and/or frame shifts. The full length library assembled from these readable fragments contains a higher percentage of members with the correct open reading frame.

A peptide with high affinity (K_d=100 nM) was identified after eighteen rounds of selection. In order to cover more sequence space, mutations were introduced into enriched library by error-prone PCR after eight cycles of selection. This peptide contains a conserved CXXC motif and binds an ATP molecule in a metal dependent manner, indicating a metal ion may help to maintain the peptide's overall structure.

b. Streptavidin aptamers

A binary patterned library was also designed in the Szostak laboratory. This library contained an 88-residue random region assembled from two 11-residue fragments with distinct hydrophobic and polar amino acid patterns[19]. Based on the patterning, the library was expected to form either amphipathic α-helices or β-strands. Surprisingly, after selection against immobilized streptavidin agarose, the enriched peptides showing high affinity with streptavidin were found to be frame-shifted, thus altering the designed patterning. The shifted frame was preferred in the selection due to the high percentage of the tripeptide (HPQ) in this frame. This HPQ tripeptide was first identified in a phage-display library selection as the minimum unit needed to bind with streptavidin. Several

peptides containing one or more HPQ motifs or the analogs, including HPA and LPQ motifs, were able to bind to streptavidin at sub-nanomolar level, representing more than 1000-fold improvement over strept-tag II peptide obtained from previous phage display experiments.

c. RNA binding peptides

The boxB RNA hairpin, the binding target of λ N protein, was used as a model system for RNA binding peptides selection[20,21]. In this selection, the initial library containing 10 random residues and representing about 10^{13} unique sequences has generated more than 100 distincts peptides that show strong affinity (K_d ~ 5 nM) against the boxB RNA hairpin. The lack of homology in these selected high-affinity RNA binding peptides indicate that the distinct peptides distributed throughout sequence space of the library possess the same binding characteristics. The research on how various peptides converge to the single wild-type RNA binding domain can shed light on further understanding of molecular evolution.

d. Epitope recovery

The mRNA display library was also applied to determine the sequences critical for antibody recognition. Baggio and coworkers prepared a 27 random residue library and used it for affinity selection against anti c-MYC monoclonal antibody 9E10[22]. The selection revealed a consensus core X-Q/E-X-L-I-S-E-X-X-L/M was required for effective recognition by antibody 9E10. This study demonstrated additional potential recognition sites for antibody 9E10 whose wild-type epitope is EQKLISEEDLN.

e. Self-assembling protein microarrays

Weng and coworkers demonstrated a novel application of mRNA-display library[23]. A DNA probe chip, whose oligo compositions are spatially addressable, can be converted to a protein chip. Each unique chimera molecule in mRNA display library can be directed to the proper coordinates on the DNA chip by a simple hybridization step. The resulting protein microarray preserves the functionality of the displayed proteins, keeps them in a uniform orientation, and has sub-attomole detection limits.

4. mRNA display library containing unnatural amino acid residues

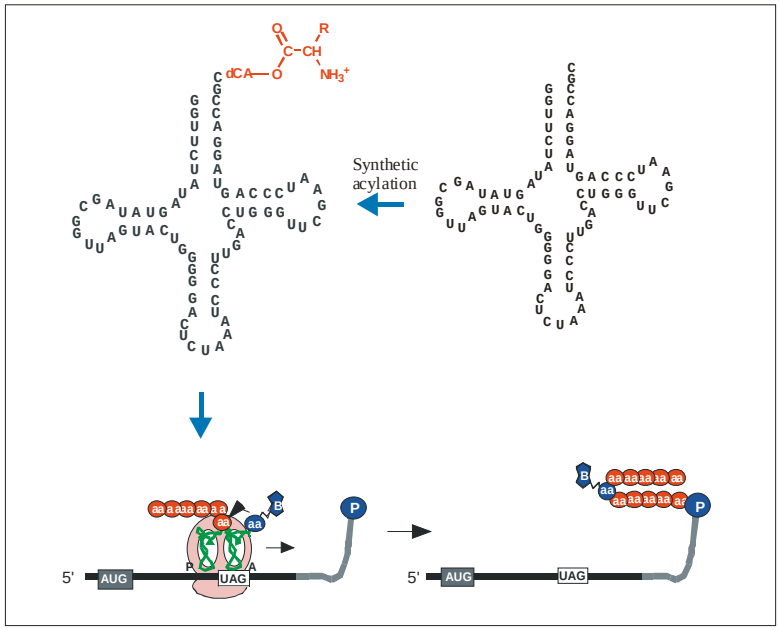
a. Incorporation of unnatural amino acid with UAG stop codon suppression

Techniques such as phage-display and cell surface display libraries demand an *in vivo* process, restricting the composition of these libraries to 20 naturally occurring amino acids. Recently, there are some advances in incorporating nonnative amino acids into proteins site specifically *in vivo*[24]. The unnatural amino acid is simply supplied as a nutrient to feed *E.Coli* and used by a modified tRNA synthetase to charge the UAG stop codon suppressor tRNA. When an mRNA template containing an in-frame UAG stop codon is translated, this nonnative residue is inserted into the elongating polypeptide at the position corresponding to the UAG stop. Even though several unnatural amino acids have been successfully incorporated by this method, the tRNA synthetase mutants can only take analogs of native residues. In addition, the incorporation of each unnatural amino acid requires tedious steps for selecting tRNA synthetase mutants that effectively charge tRNA with this amino acid. Thus, this technique is still hard to apply to combinatorial peptide libraries that require an *in vivo* step.

The mRNA display approach is a totally *in vitro* process, making it feasible to manipulate the library during the selection cycles. Techniques such as mutagenic PCR[25] and DNA shuffling[26] can be combined with mRNA display to enable the survey of more sequence space. More importantly, the *in vitro* UAG stop codon suppression enables us to insert unnatural amino acids into mRNA display library[27]. Incorporation of unnatural residues in protein via stop codon suppression was first demonstrated in the early 1990's and has been widely applied to incorporate fluorescent molecules, affinity tags, and spectroscopic probes into proteins at desired positions for purification and detection[28,29]. Unlike the *in vivo* UAG stop codon suppression developed recently, the problems associated with the charge of tRNA suppressor *in vivo* are avoided by making unnatural amino acid charged tRNA in a semi-synthetic manner. This semi-synthetic strategy starts with the protection of α-amine of amino acid with NVOC (4,5-dimethoxyl-2-nitrobenzyl) group and activation of carboxylic acid with chloroacetonitrile. The activated amino acid then reacts with a dinucleotide (pdCpA), and is subsequently coupled to the 3'-end of a transcribed 73-nucleotide tRNA by T4 RNA ligase to form intact charged suppressor tRNA. Prior to the translation, the NVOC protective group is cleaved from tRNA by UV irradiation. The addition of this semi-synthetic tRNA to the lysate in addition to a messenger RNA template containing an in-frame UAG stop codon can produce a peptide bearing the desired unnatural residue. It was demonstrated that the unnatural amino acids that are compatible with translational apparatus include not only various side chains, but also main chain structures. For example, 2 hydroxy acids can be inserted into proteins to form an acid-labile ester bond instead of normal backbone amide bond[30], and the insertion of N-methyl amino acids can produce a protease-resistant site[30]. Therefore, in

mRNA display, an engineered mRNA template containing both an in-frame stop UAG stop codon and 3'-puromycin should be able to form the mRNA-peptide fusion containing the unnatural amino acid, provided that the unnatural amino acid charged UAG suppressor is supplied (Figure 1.4). To demonstrate this strategy, biocytin-charged tRNA, a biotin derivative of lysine, was synthesized to insert the biocytin into mRNA display libraries. The incorporation of biocytin in the library was confirmed in two cycles of selection against streptavidin for enrichment of UAG stop codons in a library containing NNS random positions.

Figure 1.4 The incorporation of unnatural amino acids into mRNA-display library by amber stop codon suppression. The semi-synthetic charged tRNA can insert an unnatural amino acid in response to the UAG stop codon on the mRNA template.



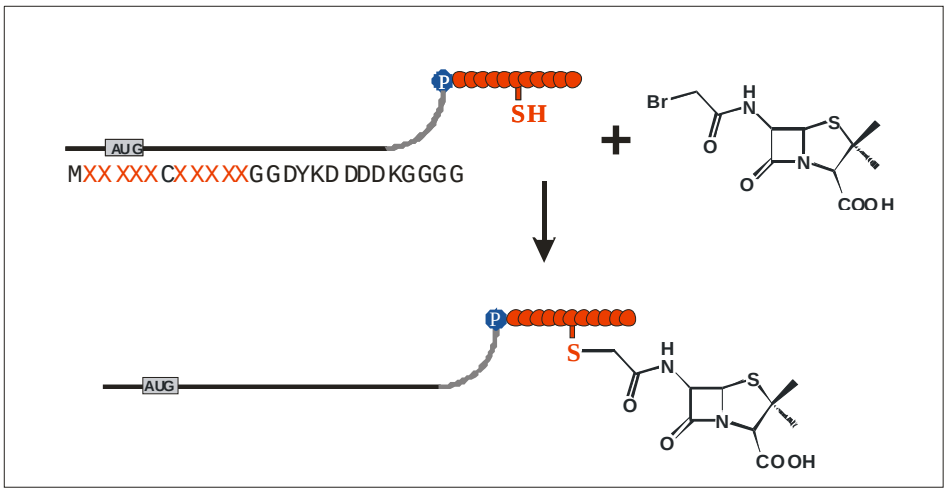
b. Incorporation of unnatural amino acid with chemical derivatization

One limitation of *in vitro* UAG stop codon suppression is that the unnatural residues incorporated must be compatible with the protein synthesis apparatus. Amino acids that 1) inhibit translation, 2) are reactive in retic lysate, and 3) are incompatible with ribosome can't be introduced into proteins. However, in mRNA display, peptides fuse with their own messenger RNA template through a covalent bond, making possible posttranslational chemical derivatization to incorporate an unnatural side chain.

Several specific chemical reactions have been extensively used for protein modification, including the coupling between the thio group of cysteine with maleimide or bromoacetyl functionality, and NHS with primary amine. The majority of protein labeling reagents available contain either cysteine reactive groups or lysine reactive groups due to their high-specificity and efficiency. Therefore, 6-bromoacetyl-penicillin was synthesized and incorporated into an mRNA display library containing a fixed cysteine[31] (Figure 1.5). This conjugation step is chemically orthogonal in that the bromoacetyl groups can not react with mRNA template and other 19 naturally occurring amino acids. The mRNA display library containing a pendant penicillin core was enriched for more potent inhibitors of penicillin binding protein 2a (PBP2a), a low-affinity penicillin binding protein that functions as a D,D-transpeptidase during bacterial cell wall synthesis of methicillin resistant *Staphylococcus aureus* (MRSA).

This chemical derivatization occurs after the translation so that any side chain, regardless its size, charge or chirality, can be introduced into fusion, considerably extending the chemical diversity of these libraries.

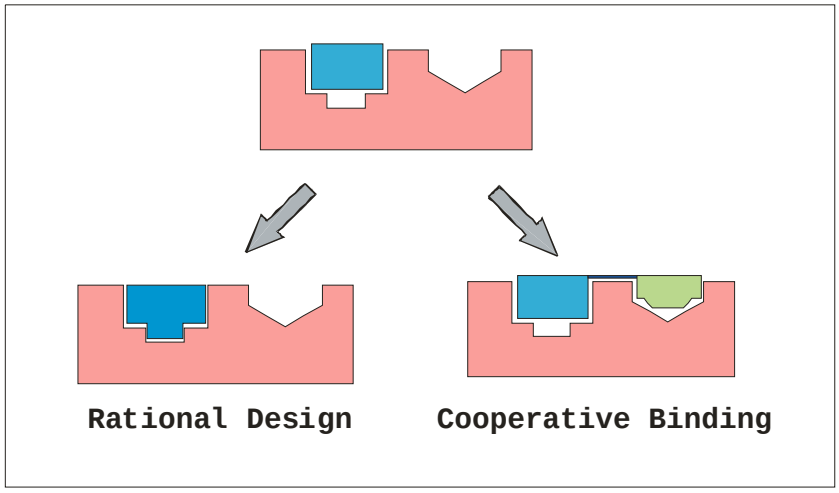
Figure 1.5 Chemical derivatization of the fixed cysteine residue with bromoacetyl-penicillin in an mRNA display library.



5. Applications of mRNA display library containing unnatural amino acids

Ligands with high affinity and high specificity are good drug candidates and excellent probes to explore the biological functions of target proteins. In pharmaceutical industries, low- or medium-affinity ligands can usually be obtained by screening a combinatorial library. To optimize the affinity of the lead compounds identified in the screen, a common approach is based on rational drug design, where the structure of lead compounds is modified to fit the active site of target proteins for better binding. Although extremely powerful, this strategy is dependent on the high-resolution structure of the target proteins, which is not always available. One interesting alternative is to utilize the cooperative binding of two molecules. The coupling of two modest-affinity non-competitive ligands can result in affinity that approaches the product of the affinities of two individual molecules, given the appropriate linker is used (Figure 1.6).

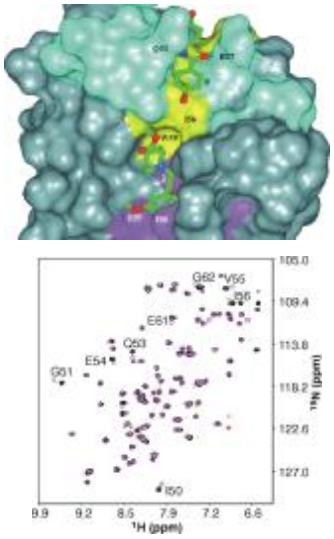
Figure 1.6 Two approaches for the optimization of ligands. One method to optimize a low-affinity ligand is rational design. Based on structural information, the ligand can be modified to fit the active site of target protein and improves its affinity. The other method is based on cooperative binding. The coupling of two modest-affinity non-competitive ligands can result in affinity that approaches the product of the affinities of two individual molecules.

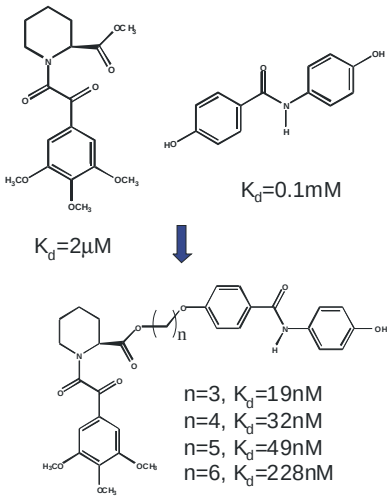


One of the examples that demonstrate this concept relied on structure-based linker design[32]. Shuker and coworkers used nuclear magnetic resonance (NMR) to optimize the linker between two weak binders of FKBP (FK506-binding protein), and were able to produce several chimeric molecules that bind to FKBP at least 100-fold better than either molecule alone (Figure 1.7). In most cases, however, only one weak ligand of the target protein is already known, thus making it necessary to search for both the second ligand and an appropriate linker. The mRNA display library containing the unnatural amino acid can be selected for these chimeric inhibitors where the unnatural amino acid serves

as the first binder. This affinity selection strategy searches for the second binder and the linker simultaneously, allowing the quick identification of high-affinity ligands.

Figure 1.7 High-affinity ligands of FKBP prepared from the cooperative binding of two weak non-competitive inhibitors of FKBP. NMR spectroscopy is used to optimize the connection between two inhibitors, whose K_d values are 2μm and 0.1mM, respectively. The resulting molecules have subnanomolar affinity.





Shuker S.B.,etal., *Science*, 274, 1531-1534 (1996)

1.2 Conclusion

mRNA display has provided a powerful tool for ligand discovery and interaction analysis. Many unique features are offered by this technique, including the option of mutagenic PCR and DNA recombination during selection, inclusion of nonnative residues, and chemical derivatization of libraries, create new opportunities in the field of molecule engineering. It also potentially enables the isolation of minimum artificial enzymes and the creation of wholly unnatural libraries with non-peptidic frameworks.

1.3 References

1. Livnah, O. et al. Functional mimicry of a protein hormone by a peptide agonist: the EPO receptor complex at 2.8 A. *Science* **273**, 464-471 (1996).

2. Norris, J. D. et al. Peptide antagonists of the human estrogen receptor. *Science* **285**, 744-746 (1999).

3. Parmley, S. F. and Smith, G. P. Filamentous fusion phage cloning vectors for the study of epitopes and design of vaccines. *Adv. Exp. Med. Biol.* **251**, 215-218 (1989).

4. Irving, M. B., Pan, O. and Scott, J. K. Random-peptide libraries and antigen-fragment libraries for epitope mapping and the development of vaccines and diagnostics. *Curr. Opin. Chem. Biol.* **5**, 314-324 (2001).

5. Huang, W., Zhang, Z. and Palzkill, T. Design of potent β-lactamase inhibitors by phage display of β-lactamase inhibitory protein. *J. Biol. Chem.* **275**, 14964-14968 (2000).

6. Smith, G. P. and Petrenko, V. A. Phage display. *Chem. Rev.* **97**, 391-410 (1997).

7. Zwick, M. B., Shen, J. and Scott, J. K. Phage-displayed peptide libraries. *Curr. Opin. Biotechnol.* **9**, 427-436 (1998).

8. Hanes, J. and Pluckthun, A. *In vitro* selection and evolution of functional proteins by using ribosome display. *Proc. Natl. Acad. Sci. USA* **94**, 4937-4942 (1997).

9. Boder, E. T. and Wittrup, K. D. Yeast surface display for directed evolution of protein expression, affinity, and stability. *Method Enzymol.* **328**, 430-444 (2000).

10. Roberts, R. W. and Szostak, J. W. RNA-peptide fusions for the *in vitro* selection of peptides and proteins. *Proc. Natl. Acad. Sci. USA* **94**, 12297-12302 (1997).

11. 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*. *FEBS Lett.* **414**, 405-408 (1997).

12. Geysen, H. M., Rodda, S. J., Mason, T. J., Tribbick, G. and Schoofs, P. G. Strategies for epitope analysis using peptide synthesis. *J. Immunol. Methods.* **102**, 259-274 (1987).

13. Brenner, S. and Lerner, R. A. Encoded combinatorial chemistry. *Proc. Natl. Acad. Sci. USA* **89**, 5381-5383 (1992).

14. Pinilla, C. et al. A review of the utility of soluble peptide combinatorial libraries. *Biopolymers* **37**, 221-240 (1995).

15. Gallop, M. A., Barrett, R. W., Dower, W. J., Fodor, S. P. and Gordon, E. M. Applications of combinatorial technologies to drug discovery. 1. Background and peptide combinatorial libraries. *J. Med. Chem.* **37**, 1233-1251 (1994).

16. Parmley, S. F. and Smith, G. P. Antibody-selectable filamentous f_d phage vectors: affinity purification of target genes. *Gene* **73**, 305-318 (1988).

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

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

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

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

21. Barrick, J. E. and Roberts, R. W. Sequence Analysis of an Artificial Family of RNA-Binding Peptides. *Protein Sci.* **11**, 2688-2696 (2002).

22. Baggio, R. et al. Identification of epitope-like consensus motifs using mRNA display. *J. Mol. Recognit.* **15**, 126-134 (2002).

23. Weng, S. et al. Generating addressable protein microarrays with PROfusion covalent mRNA-protein fusion technology. *Proteomics* **2**, 48-57 (2002).

24. Wang, L., Brock, A., Herberich, B. and Schultz, P. G. Expanding the genetic code of *Escherichia coli*. *Science* **292**, 498-500 (2001).

25. Cadwell, R. C. and Joyce, G. F. Randomization of genes by PCR mutagenesis. *PCR Methods. Appl.* **2**, 28-33 (1992).

26. Stemmer, W. P. C. Rapid evolution of a protein *in vitro* by DNA shuffling. *Nature* **370**, 389-391 (1994).

27. Li, S., Millward, S. and Roberts, R. *In vitro* selection of mRNA display libraries containing an unnatural amino acid. *J. Am. Chem. Soc.* **124**, 9972-9973 (2002).

28. Ellman, J., Mendel, D., Anthony-Cahill, S., Noren, C. J. and Schultz, P. G. Biosynthetic method for introducing unnatural amino acids site-specifically into proteins. *Methods Enzymol.* **202**, 301-336 (1991).

29. Bain, J. D., Glabe, C. G., Dix, T. A. and Chamberlin, A. R. Biosynthetic site-specific incorporation of a non-natural amino acid into a polypeptide. *J. Am. Chem. Soc.* **111**, 8013-8014 (1989).

30. Bain, J. D., Wacker, D. A., Kuo, E. E. and Chamberlin, A. R. Site-specific incorporation of non-natural residues into peptides: effect of residue structure on suppression and translation efficiencies. *Tetrahedron* **47**, 2389-2400 (1991)

31. Li, S. and Roberts, R. W. Novel PBP2a inhibitors isolated from a mRNA display library containing a penicillin side chain. *Chem. Biol.* in press.

32. Shuker, S. B., Hajduk, P. J., Meadows, R. P. and Fesik, S. W. Discovering high-affinity ligands for proteins: SAR by NMR. *Science* **274**, 1531-1534 (1996).