Translation of Random Transcripts Generated by TdT: Potential Use in Polysome Peptide Libraries

---

A Thesis

Presented to

The Faculty of the College of Science and Technology

Morehead State University

In Partial Fulfillment

Of the Requirements for the Degree

Master of Science in Biology

---

by

Michael Lane Spencer

July, 1998

Accepted by the faculty of the College of Science and Technology, Morehead State University, in partial fulfillment of the requirements for the Master of Science degree.

Director of Thesis

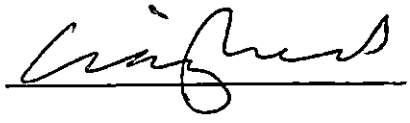
Master's Committee: _____, Chair

7/22/98

Date

Translation of Random Transcripts Generated by TdT: Potential Use in Polysome Peptide Libraries

Michael L. Spencer, M. S.
Morehead State University, 1998

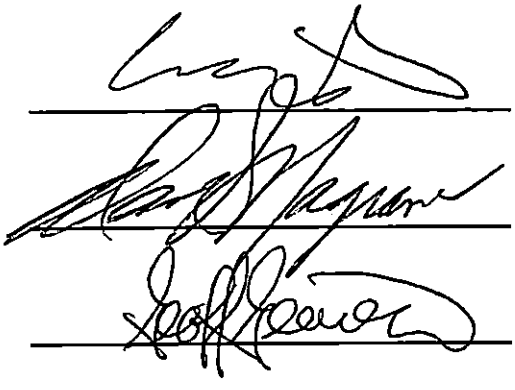
Director of Thesis: 

The use of peptide libraries combined with in vitro selection strategies has become a powerful protein engineering tool with all the advantages of evolution while lacking all the disadvantages of traditional methodologies (Clackson and Wells 1994).

Peptide libraries allow for the creation of a population of novel peptides, selection from the population, and amplification of selected peptides (Gates 1996). In recent years, powerful applications of peptide libraries have demonstrated the ability of peptide libraries to enhance our knowledge of proteins. However, current peptide libraries do not encode random proteins of a large molecular weight or three-dimensional, folded structure. We used terminal deoxynucleotidyl transferase (TdT), a template-independent DNA polymerase found in prelymphocytes, to generate large random protein coding sequences. We generated RNA transcripts of over 1000 nucleotides long using a cell-free methodology. Translation products from these transcripts ranged from 250-415 amino acids. This far exceeds the 6-38 random

amino acid sequences found in current peptide library systems (Kay et al. 1993). We suggest that this system can be used in polysome display libraries (Tuerk and Gold 1990; Mattheakis et al. 1994) to display large molecular weight proteins. Furthermore, this library creation strategy could be used in ribosome and ARM display methodologies to select proteins that have three dimensional, folded structure (Hanes and Pluckthun 1997; He and Taussig 1997).

Accepted by: _____, Chair



Acknowledgments

I wish to thank my graduate committee, Dr. Geoff Gearner, Dr. David Magrane, and Dr. Craig Tuerk. Each of the above people have played a valuable role in my professional and personal development. Their guidance has not only molded my scientific abilities but also shaped my personal values. I express my sincere gratitude and wish them many successful years to come. In addition, I express my thanks to the faculty of the department of biological and environmental sciences. It is my hope that the biology faculty continue the standard of excellence that was extended to me during my stay at Morehead State University.

I wish to thank the following core group of graduate students: Todd Jacobs, Chris Goshorn, Mike Fultz, John Cox, Mehrunissa Lambat, Les Meade, Marlin Lawson, Denise Patrick, Ryan McCall, Matt Burchett, and Nicole Griffith. I hope we all remember the good times.

Two people in particular played a huge role in the completion of this research. Dr. Craig Tuerk supplied the creative force that shaped the fundamental ideas of this work. I wish to think him for allowing my participation in this work. Mehrunissa Lambat was my faithful partner in crime during the duration of this work. I wish to think her for all her hard work during the two summers spent in the lab.

Finally, I wish to think my family. It is a strong family that pushes men and women to achieve. I appreciate their loving support during my undergraduate and graduate education. I would especially like to thank my brother, Matthew Spencer.

Table of Contents

List of Figures

Introduction

Nature has provided life with an effective array of useful proteins that have been gradually optimized over millions of years of natural selection (Sander 1994). Many of nature's proteins have proved useful in a medical, agricultural, or scientific setting (Blundell 1994). In recent years it has become a priority to be able to design proteins or peptides with specific properties. Traditionally, protein engineering has been an iterative process of mutant construction and subsequent assay of individual mutants for the desired property (Clackson and Wells 1994). This traditional approach is time consuming, expensive, and requires considerable knowledge of how protein structure affects function (Blundell 1994). The use of peptide libraries combined with in vitro selection strategies has become a powerful protein engineering tool with all the advantages of evolution while lacking all the disadvantages of traditional methodologies (Clackson and Wells 1994).

Peptide libraries allow for the creation of a population of novel peptides, selection from the population, and amplification of selected peptides (Gates 1996). In recent years, powerful applications of peptide libraries have demonstrated the ability of peptide libraries to enhance our knowledge of proteins. These libraries facilitated the discovery of receptor ligands (Martens 1995), antibody epitopes (Cull 1992; Scott 1990) , novel antimicrobial compounds (Blondelle and Houghten 1996) and enzyme substrates (Matthews 1993, Schatz 1993). Peptide libraries (sometimes called combinatorial peptide libraries) are generated by two differing methodologies. Biological peptide libraries express random peptides physically linked to the nucleic

acid sequence that encodes them and use biological machinery for amplification (Scott and Smith 1990; Cull et al. 1992; Mattheakis et al. 1994). Synthetic combinatorial peptide libraries use chemical synthesis procedures to generate peptides, usually on a fixed bead substrate (Dooley et al. 1994; Lam et al. 1991).

Peptide libraries

Biological peptide libraries use the techniques of molecular biology to engineer random peptides that are somehow linked to the genetic material that encoded them (Clackson and Wells 1994). Biological library strategies offer several advantages that make library production and maintenance practical.

M13 is a filamentous bacteriophage that has been utilized to express peptide sequences that are cloned into the phage gene III. This fusion phage displays the peptide sequences as part of the gene III protein, pIII (Scott and Smith 1990). M13 expresses five copies of pIII protein at one end of the phage particle where the proteins are involved in *E. coli* pilus attachment during infection (Fowlkes et. al. 1992). The foreign peptide sequences do not interfere with pilus attachment and are accessible for binding. Since each virion contains a copy of the M13 genome, there is a physical link between the foreign peptide sequence and the DNA that encodes it (Fowlkes et. al. 1992). This property of accessibility and genome linkage lends itself to the creation of a M13 fusion phage library that displays diverse peptide sequences that can be subjected to in vitro selection (Scott and Smith 1990).

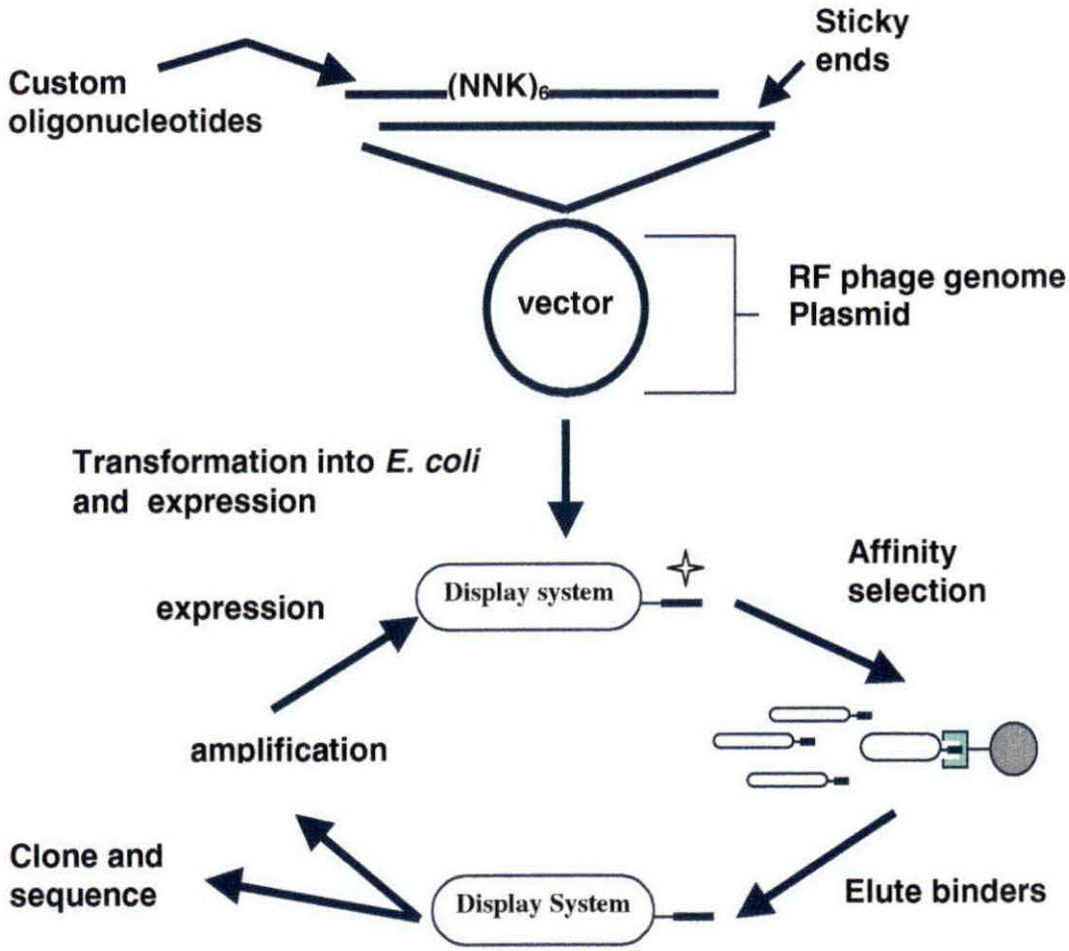


Figure 1: Peptide library creation strategy. Cloning oligonucleotides are synthesized with the random motif NNK. The size of the library, the number of different random sequences contained within the library, is dependant upon transformation efficiency of the vector, RF genome for phage or plasmid for peptides-on-plasmids. The random peptide (marked with a star) is displayed on one of a number of display systems (Figure 2) and is subjected to rounds of selection. Amplification requires that the selected genetic material be transformed or transfected into *E. coli*. Once selection is completed the selected protein is studied by cloning and sequencing.

Phage libraries are created (Figure 1) by inserting randomized oligonucleotides into the phage vector, the RF form of the genome. Oligonucleotide synthesis is used to construct degenerate coding sequences (NNK) within a fragment

that has sticky ends compatible with the cloning position on the phage vector (Scott and Smith 1990). The NNK motif (N=equal mixture of deoxynulceotides, K=equal mixture of G and T) codes for all 20 amino acids including the amber stop codon (Scott and Smith 1990). The NNK motif avoids the UAA and UGA stop codons increasing the number of complete peptides. Phage display libraries typically contain fragments that code for 6 to 38 random amino acids (Kay, et al. 1993). The diversity of peptides is on the order of 10^9 with each individual phage carrying a vector with one random coding sequence and displaying the random peptide on its five pIII proteins (Scott and Smith 1990).

Another library creation strategy is the "peptide on plasmids" or lac repressor display methodology (Figure 2). Random oligonucleotides, as used in phage display, are used to create an insert that is cloned into the carboxy terminus of the lac repressor (Cull, et al. 1992). The vector for this method is a plasmid that is propagated in bacteria. The genomic linkage between the peptide and its genetic material is through the lac repressor (Lac I) binding to two lac operator sequences forming a "peptide-Lac I-plasmid" complex. This complex can be transformed into *E. coli* after selection for amplification. The selection method is like the phage display selection methods and has been used to isolate ligands for antibodies and cell surface receptors (Gates, et al. 1996). These isolated proteins have similar binding affinities with phage displayed libraries for their targets, 10 to 100 μM (Gates, et al. 1996).

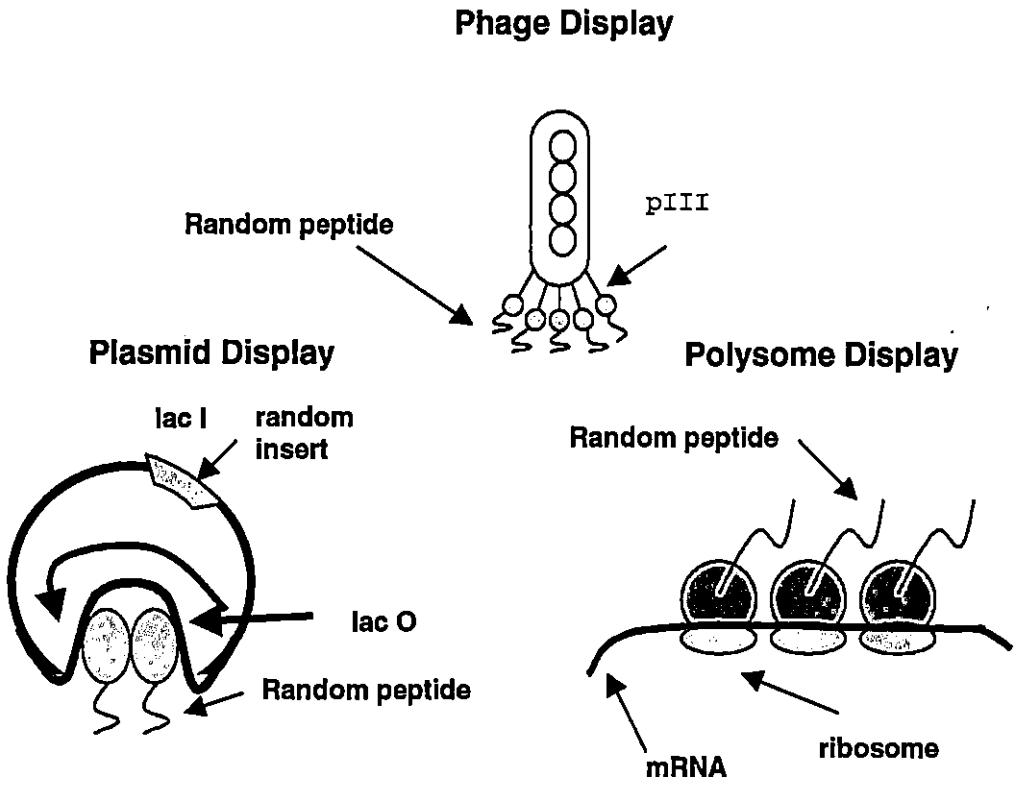


Figure 2: Display Systems for Biological Libraries. Biological display systems provide a physical link between the random peptide that they display and the nucleic acid sequence that encodes the random peptide. Selection for the random peptide also selects the nucleic acid sequence. This property of biological libraries allow for easy cloning, peptide characterization, and library amplification.

Tuerk and Gold (1990) suggested that their RNA combinatorial library, SELEX, could be modified to generate randomized transcripts capable of being translated and displayed on polysomes. In this scenario a drug such as cyclohexamide is used to freeze ribosomes to the RNA transcripts to create a stable

polysome structure. The ribosome would provide a physical link between the random peptide and the genetic material, RNA in this case, that encodes the peptide (Tuerk and Gold 1990). This idea was used to create a polysome library that yielded proteins with binding affinities that ranged from 7-140 n*M* (Mattheakis, et al. 1994). Phage and "peptides-on-plasmid" libraries typically yield proteins with affinities ranging from 10 to 100 μ*M* (Gates, et al. 1996). Polysome libraries have the potential to display 10^{12} random peptides and outperform phage libraries in obtaining high affinity binders (Mattheakis, et al. 1994).

Once a library is constructed, a selection protocol (Figure 1) is used to find ligands to the target molecule. This procedure, sometimes called biopanning (Cull, et al. 1992), involves rounds of incubation of the library with a target immobilized to a microtiter plate or other substrates such as beads (Scott and Smith 1990). Incubation of a peptide library with the target molecule allows peptides that have affinity for the target to bind the target. Peptides that do not bind, termed non-binders, are washed away, while binders, peptides attached to the target, are eluted and propagated in their bacterial host to amplify the library (Scott and Smith 1990). This process is repeated in subsequent rounds of selection under conditions that select for peptides with the highest affinities. The peptides from the last rounds of selection and amplification are cloned and sequenced (Scott and Smith 1990).

An alternative to the above biological peptide libraries are synthetic peptide libraries. Synthetic libraries, either peptide or small organic molecules, are created by

iterative synthesis methods (Clackson and Wells 1994). These libraries fall into two categories, libraries synthesized on a fixed array, the "one bead-one peptide strategy" (Lam, et al. 1991) or synthetic peptide combinatorial libraries (Houghten, et al. 1991; Dooley, et al. 1994; Blondelle and Houghten 1996).

Synthetic peptide libraries such as the "one bead, one peptide" approach are created by the iterative chemical coupling of amino acids (Lam, et al. 1991). Synthesis begins by distributing resin beads which have an attached amino acid to 19 columns. Each column contains an individual species of amino acid for coupling to the library members. Once the coupling reaction has gone to completion the beads are pooled and redistributed for the coupling of the next amino acid. The process of repeated cycles of pooling and random distribution to generate peptides is termed "split synthesis" (Lam, et al. 1991). A library of this type is screened by incubating a fluorescein or alkaline phosphatase labeled target molecule with the mixture of random peptides. Selected peptides are identified by washing away non-binding target molecules. Labeled beads are dissected from the mixture of beads that did not display peptides with affinity to the target molecule and microsequenced (Lam, et al. 1991).

Synthetic peptide combinatorial libraries (SPCL; Houghten et al. 1991) are generated by a methodology which is nearly identical to "split synthesis", but is termed DCR (divide, couple and recombine). SPCL peptides are cleaved from their resin beads to yield a soluble library (Houghten, et al. 1991). SPCL libraries scan each amino acid position within the peptide for optimal activity (Houghten, et al.

1991). The selection of peptides from SPCL peptide libraries mimics the selection and library amplification steps of biological libraries by a novel selection strategy.

SPCL peptides are synthesized with the first two amino acid positions specifically defined, while the remaining amino acids are randomly defined (Houghten, et al. 1991). This strategy gives rise to 400 different pools of peptides with 160,000 individual members (assuming the peptide is a hexapeptide). Selection identifies the pool with the desired activity. The knowledge from the first round of selection is used to generate a second library with the third position defined. This yields 20 different pools with 8000 different members (Houghten, et al. 1991). Subsequent rounds of selection reduce the number of peptides within the pools. The peptide is selected when all amino acid positions have been defined (Houghten, et al. 1991).

Although synthetic peptide libraries are not as readily reproduced or amplified when compared to biological peptide libraries, they do offer distinct advantages. During synthesis of chemical peptide libraries, it is possible to incorporate D-amino acids, chemically modified amino acids or other carboxylic acids. This flexibility allows the creation of diverse peptides that cannot be cleaved by proteases (Dooley, et al. 1994 and Blondelle and Houghten 1996). Despite the relatively high cost of synthetic peptide libraries, their versatility has led to increased use by the pharmaceutical industry to facilitate drug discovery (Blondelle and Houghten 1996; Udaka 1996).

Limitations of Peptide Libraries

Peptide libraries, as well as other combinatorial libraries, have far surpassed expectations for their ability to generate efficient binders. However, peptide libraries only encode short oligomeric peptides. The length of random amino acid sequence is dependant upon the maximum limit of oligonucleotide synthesis for biological libraries and upon chemical synthesis yield for synthetic libraries. Biological peptide libraries have been created that code for a random peptide ranging from 6-38 amino acids (Kay, et al. 1993). This size restriction limits current peptide libraries ability to find highly specific binders due to lack of shape constraint and does not allow for the display of completely random proteins that exhibit folded motifs.

Currently peptide libraries display random linear peptides. Highly specific binding interactions can be missed due to a lack of constraint within the peptide backbone (Clackson and Wells 1994 and Ladner 1995). Phage and "peptide-on-plasmid" libraries display linear, unconstrained peptides that are free to rotate about their alpha carbon bonds. The most important example of constraint affecting highly specific interactions are the cone snail toxins. Cone snails of the genus *Conus* are a group of highly efficient marine predators (Olivera, et al. 1995). Cone snails attack by shooting a dart tipped with a toxin that contains numerous pharmacologically active peptides into the prey (Olivera, et al. 1990).

Cone snail peptide toxins bind to ion channels and receptors in the preys neuromuscular system. The toxin mixture of *Conus geographus* contain peptides

that target voltage-sensitive calcium channels, sodium channels, N-methyl-D-aspartate (NMDA) receptors, acetylcholine receptors, and vasopressin receptors (Olivera, et al. 1990). Cone snail toxin peptides, which are cleaved from large precursor proteins that guide folding and disulfide bond formation, use a stiff framework made possible by multiple disulfide bonds to display diverse arrays of amino acids within looped structures (Olivera, et al. 1990). The mature peptide is constrained by disulfide bridges and ranges from 10-30 amino acids in length (Olivera, et al. 1990).

The constrained arrangement allows the peptide to bind to receptor subtypes to knock out the prey's neuromuscular system at many levels (Olivera, et al. 1990). Peptide library designers try to mimic the highly specific interactions of cone snail toxins by designing libraries with constrained loop domains similar to that of cone snail toxins or by designing cyclic peptides (Clackson and Wells 1994; Eichler, et al. 1995; Ladner 1995)

Natural proteins exhibit elaborate three-dimensional, folded structure. This folded structure aids in the functionality of the protein. At the present peptide libraries do not encode random, large molecular weight, folded proteins. Proteins such as human growth hormone (hGH) have been expressed as pIII fusion proteins in phage libraries (Lowman and Wells 1993). However, libraries of this type only randomize short sequences, such as the hGH binding site (Lowman and Wells 1993) or zinc finger DNA binding sites (Greisman and Pabo 1997). Peptide libraries can not display proteins with large random sequences due to the size restriction placed on them by oligonucleotide synthesis and synthetic peptide chemistry. It is not possible

sample the entire sequence space of a protein during selection. Therefore the selection of random proteins in varied structural contexts from peptide libraries is not feasible.

Research Objectives

Terminal deoxynucleotidyl transferase (TdT) is a template-independant DNA polymerase that catalyzes the addition of nucleotides to a 3' terminal hydroxyl (Chang and Bollum 1970). TdT was used by Damiani, et al. (1982) to create DNA duplexes that contained random sequences. It is proposed that long random protein coding sequences can be created by using the enzyme TdT (Figure 3a and 3b). These sequences can be used to create long random transcripts that could be used in polysome peptide display libraries (Figure 3c and 3d). Such a system would allow the display and selection of large random proteins.

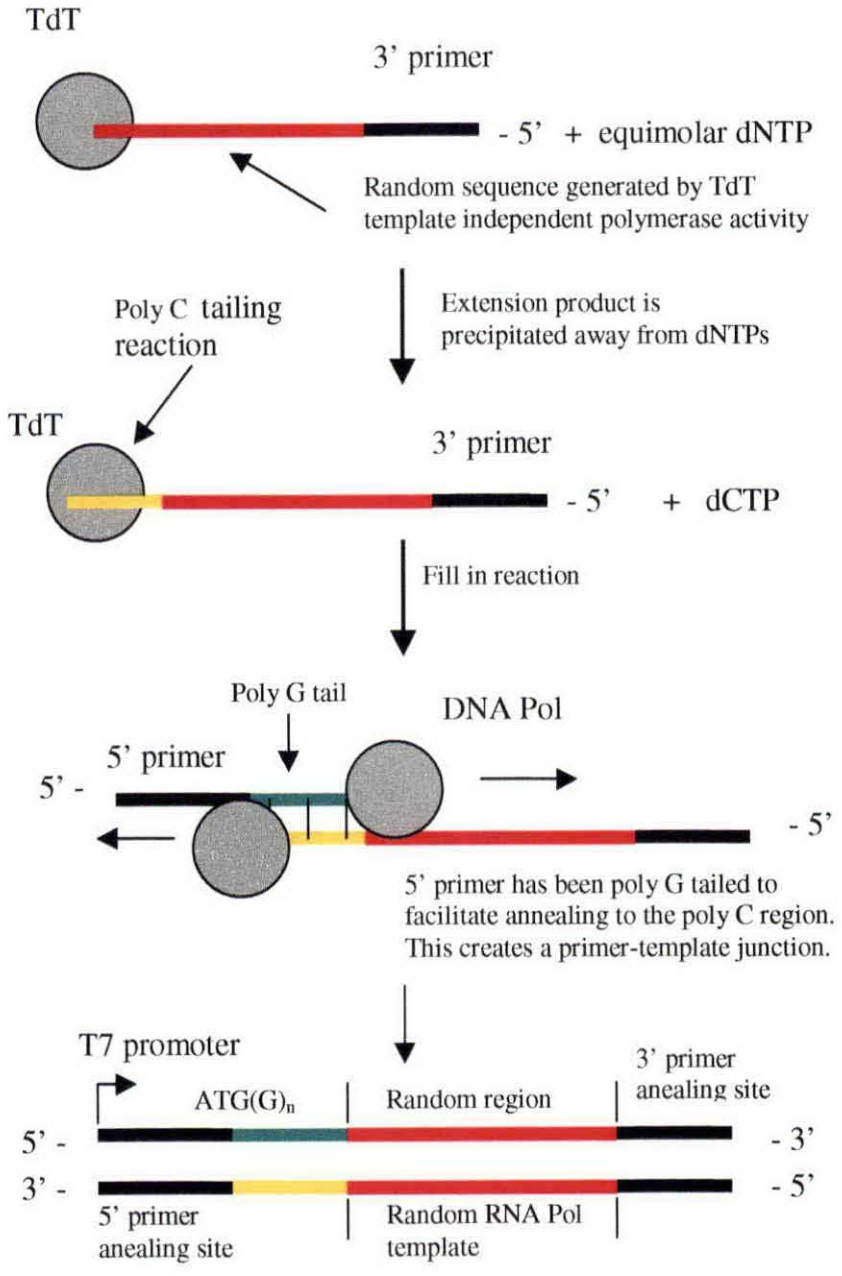


Figure 3a: Generation of random DNA duplexes. TdT is used in an extension reaction to generate a ssDNA that contains a random nucleotide sequence. The ssDNA product is converted to a DNA duplex that contains a T7 transcription promoter in front of the random sequence.

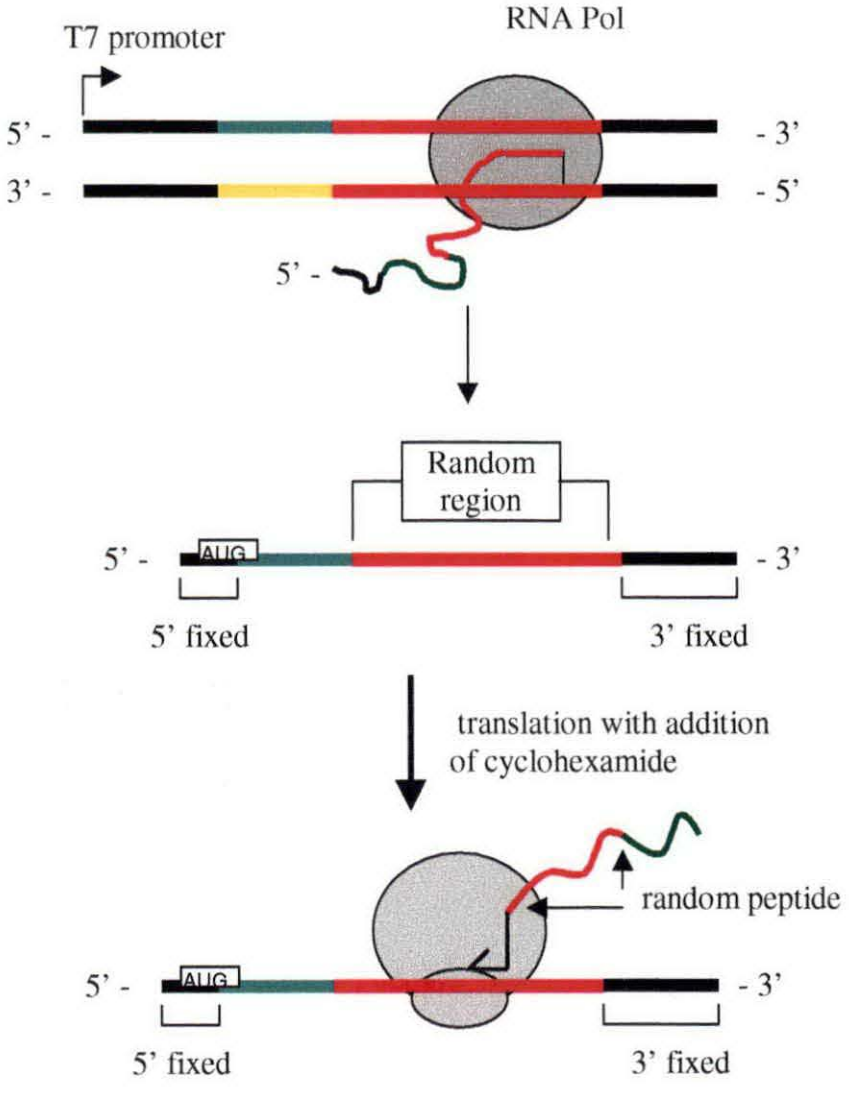


Figure 3b: Creation of polysomes displaying random proteins. The ramdom duplex (bottom 3a) is transcribed to generate a pool of random RNA. The pool of random RNA are translated in vitro with cyclohexamide being added to stop translation and freeze the ribosomes to the RNA. For simplicity, only one ribosomes is shown on the RNA strand. In reality ribosomes are stacked along the length of the RNA.

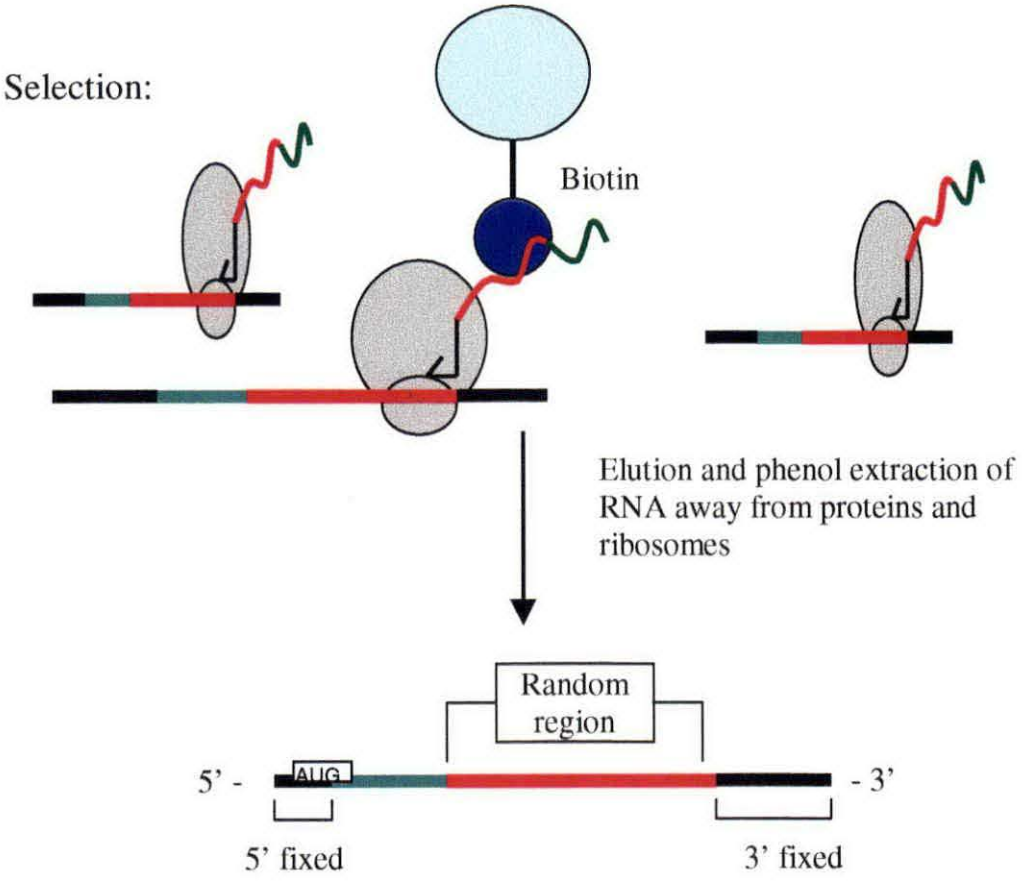
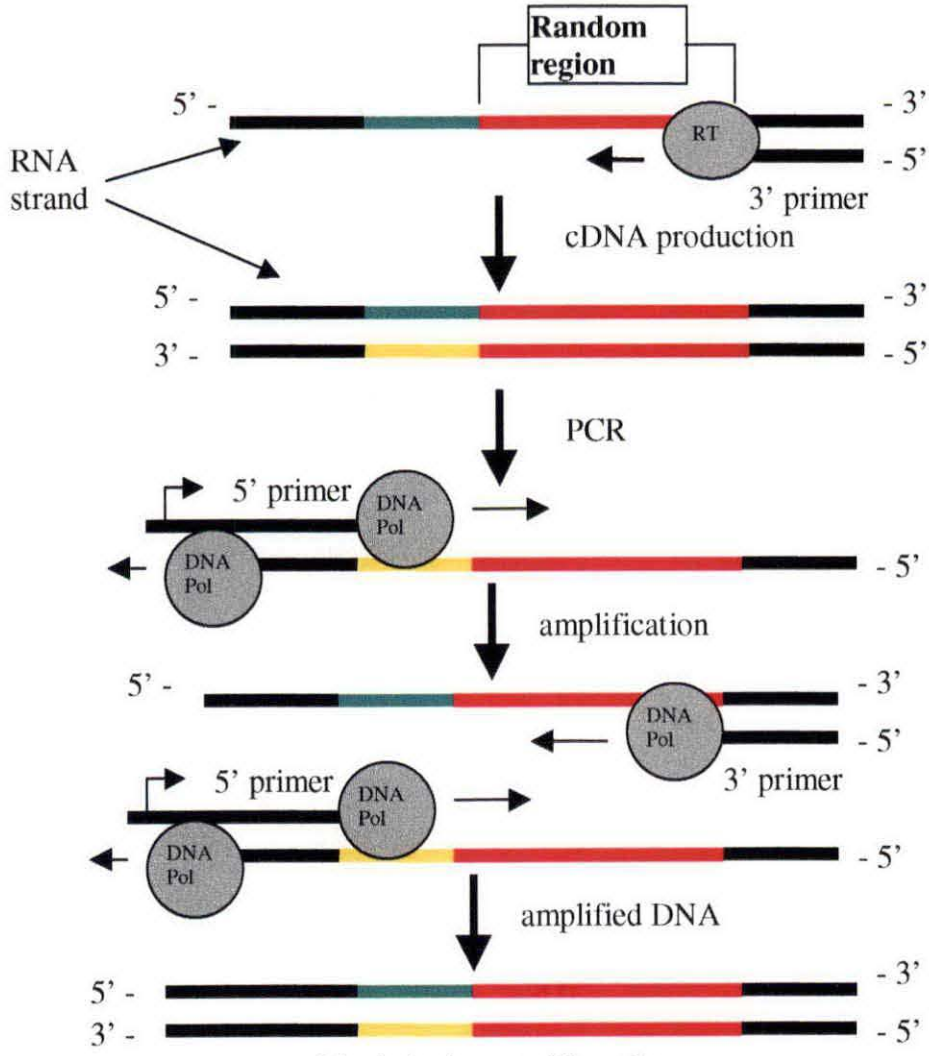


Figure 3c: Selection of random proteins. Polysomes displaying random peptides are selected for binding specific targets by the process of affinity chromatography. In this case, it is the binding to biotin, attached to an agarose bead within an affinity column that selects the polysomes with affinity for biotin. After an incubation time, peptides that do not bind (non-binders) are washed away leaving behind attached polysomes, which are composed of a random protein linked to its encoding RNA by the ribosomes that translated the protein. Therefore, selection of proteins with desired properties also selects for the genetic material the encode the proteins.



Identical to bottom of figure 3a

Figure 3d: Amplification of selected mRNA. The selection of the RNA at the bottom of figure 3c is the due to the properties of the protein that the polysome is displaying. This RNA is reverse transcribed to generate a cDNA and then amplified by PCR to generate an amplified library. The amplified library, which contains many protein coding sequences that can bind to the target, is put through numerous rounds of expression, selection and amplification. This process evolves protein coding sequences that encode the best binders. When selection is completed the final amplified library can be cloned in expression vectors that allow for DNA sequencing or protein purification.

Materials And Methods

Enzymatic reactions were carried out under manufacturer's recommendations unless otherwise noted. Polyacrylamide and agarose gel electrophoresis and other molecular biology techniques were performed as in Sambrook, et al. (1989). Other methods not described below are noted in individual results for ease in description. General lab supplies were purchased from Sigma.

Extension Reactions

The oligonucleotide TDT3P (Figure 4) was extended with TdT (lot # 74209, Promega, Madison, Wisconsin) under the following optimized conditions: 20 picomoles TDT3P, 2 m*M* Co^{2+} (1m*M* of this value is present in 1x reaction buffer), 3 m*M* BSA, 0.5 m*M* dNTP, 0.8 U of TdT, and 1x reaction buffer (100 m*M* cacodylate buffer at pH 6.8 and 1 m*M* DTT). The optimal concentration of Co^{2+}, BSA, dNTP, and TdT was determined experimentally (data not shown). Reactions were incubated at 37°C overnight. Additional TdT (0.8 U) and dNTP (0.5 m*M*) were added at the 5 and 10 hour timepoints. Reactions were examined on 8% denaturing polyacrylamide gel media to assay reaction completion (Sambrook, et al. 1989). The length of extension products were estimated from lambda *Hin*dIII size standards on 1% agarose gels.

Extension products were ethanol precipitated away from the dNTP mix and homopoly C tailed (Sambrook, et al. 1989). Homopoly C tailing reactions contained 0.5 m*M* dCTP, 1x reaction buffer (see above), 2 m*M* Co^{2+}, 3 m*M* BSA and 0.8 U of

T7 Promoter

5'—**TAATACGACTCACTATAGG**GAGAATAAACGCTCAACTTTGACC**AT**—3'

GS5, 5' Primer (45 nucleotides). contains T7 Promoter, Beta Globin-Like Start. This primer was homopoly G tailed. GS5 contains a AT sequence at its 3' end that allowed for creation of a start codon once poly G tailed (Operon Technologies, Alameda, California).

5'—GGCC***GAATTC***TTTATTACGAGGA —3'

*Eco*RI

TDT3P, 3' Primer (23 nucleotides). 3' Primer for TdT extension reactions. This primer was randomized using a equimolar dNTP mix and homopoly C tailed (Operon Technologies).

5'—CCC***GGATCC***TAATACGACTCACTATAGG —3'

*Bam*HI

CLO5GLO (28 nucleotides). 5' Cloning Primer for cloning selected random sequences. CloGlo5 contains BamHI for directional coloning (Operon Technologies).

Figure 4: Oligonuceleotide primers. The above primers were used in the creation of random transcripts. They allow for transcription and translation to occur in a cell free system.

TdT. Reactions were incubated at 37°C for 1 to 5 hours. A poly G tail was added to the oligonucleotide GS5 (Figure 4) under the conditions described previously except with dGTP as the nucleotide.

Poly C tailing of band purified extension products were under optimal conditions with 5 µl of band purified extension product input as template. Band purification was by freeze fracture. Briefly, an extension reaction was run on a 8% polyacrylamide gel, and a section of the lane containing the extension products was removed. The gel fragment was placed in a 1.5 ml microcentrifuge tube and the volume was adjusted to 500 µl of distilled water. The gel slice was then snap frozen in dry ice. After adding 200 µl of distilled water the frozen gel slice was crushed to a slurry with a 1 ml syringe plunger. The gel particles were removed by centrifugation in a microcentrifuge at high speed for 5 minutes. The supernatant was transferred to a new tube (supernatant 1) while 100 µl of 3M sodium acetate (NaAc) and 200 µl of water were added to the gel pellet. The gel pellet was resuspended, snap frozen and crushed as before. The gel material was removed by centrifugation with the supernatant being combined with supernatant 1. The supernatant was phenol/ether extracted and then ethanol precipitated. The band purified pellet was resuspended in 25 µl of distilled water.

Fill-in Reactions

Poly G tailed GS5 (20 picomoles) was annealed to poly C tailed extension products (95°C for 15 min. 37°C for 5 min., and ice bath for 5 min.) to create a primer template junction. Reactions were incubated 37°C overnight under the following conditions: 1x RT buffer (New England Biolabs, Beverly, Massachusetts), 0.5 m*M* dNTP, and 100 U M-MuLV reverse transcriptase (New England Biolabs).

Transcription Reactions

Fill-in products or amplified library DNA was transcribed for 2 h at 37°C under the following conditons: 1 x RNA Pol buffer (New England Biolabs), 0.5 m*M* NTP, +/- 25 m*M* methyl G cap (lot # 6114635011, Pharmacia Biotech), and 200 U of T7 RNA Pol (New England Biolabs). The methyl G cap was used in all reactions in which the RNA products were translated. Aliquots of transcription reactions were electrophoresed on 1% agarose gels +/- DNase I (cat. # 27-0514, lot # 7020514011, Pharmacia Biotech) to examine the efficiency of the transcription reaction. Treatment with 20 units of DNase I for 10 minutes at 37°C was enough to remove all template DNA.

RT-PCR

RNA from the above transcription reactions was phenol/ether extracted, ethanol precipitated, and resuspended in 50 µl of distilled water. Phenol/ether extraction involved two extractions with equal volumes of equilibrated phenol followed by two equal volume ether extractions. The following components were added: 1x RT buffer

(New England Biolabs), 20 pmoles TDT3P, 0.5 m*M* dNTP, and 100 U M-MuLV reverse transcriptase. The volume was brought to 75 µl and incubated at 37°C for 2 hours.

Reverse transciption (RT) products were amplified under the following conditions : 1 µl RT product, 1x PCR buffer (Promega), 0.5 m*M* dNTP, 10 pmoles GS5, 10 pmoles TDT3P, and 4 U of Taq (Promega). A PCR cycle consisted of the following steps: 95°C for 1 min., 50 or 55°C for 10 sec., and 72°C for 1 min. Pilot experiments were performed to optimize the number of cycles and annealing temperature. The remaining RT product was amplified by replicating the optimal pilot experiment many times.

Translation

Transcription products, after DNAse I (Pharmacia) treatment were translated for 2 hours at room temperature under the following conditions: 25 ml wheat germ extract (lot # 6141301, Promega), 60 m*M* KOAc, 2 ml RNA, 0.08 m*M* amino acid mix (- met), and 0.5 mCi/ml of [^{35}S] methionine. TCA precipitation for amino acid incorporation assay was performed as in Promega Protocols and Applications Guide (1991).

Results

Extension Reactions

Conditions that resulted in the optimal TdT extension lengths were shown to be 2 m*M* Co^{2+}, 3 m*M* BSA, 0.5 m*M* dNTP, 0.8 U of TdT, and 100 m*M* cacodylate buffer at pH 6.8 and 1 m*M* DTT for 20 picomoles of input primer, TDT3P (data not shown). Although there was no appreciable increase in extension product length after 5 hours, there was an increase in the number of extension products as the reaction continues past 5 hours (Figure 5).

It was determined that when TdT was incubated at 37°C in only its storage buffer, it lost its ability to extend after 2 hours (Figure 6). To further complicate matters, increasing the concentration of TdT in 5 hour reactions did not increase the length of extension products, but did increase the amount of primer that was extended (data not shown). Figure 7 shows that the longest extensions and most complete reactions, based on the amount of primer extended, were obtained by supplementing the reaction after 5 hours with an additional amount of TdT (0.8 U) and dNTP (0.5 m*M*).

Maximum extension lengths were obtained under optimal conditions with TdT and dNTP supplementation (Figure 8). The length of extension products were estimated on a 1% non-denaturing agarose gel using *Hin*dII lambda DNA size standards. The use of a non-denaturing agarose does not allow for precise size determination of the ssDNA extension products because of the opportunity to base pair or fold. It does allow for an estimation of sizes that exist in the population of

products from an extension reaction. We were able to perform extension reactions under optimal conditions where the major portion of extension products ranged from 500-2000 bp (Figure 8, lane 2). It was possible to increase the number of products in this size range by running the reaction an additional 24 hours with TdT (0.8 U) and dNTP (0.5 m*M*) supplementation at the 5 and 10 hour time points (Figure 8, lane 3).

After limited random extension of precipitated extension products, poly C tails were added to band purified extension products to more accurately gauge the poly C tailing reaction. The poly C tailing reaction was performed under the optimal conditions with dCTP being used in place of the equimolar dNTP mix. Extension products were band purified by freeze fracture. Figure 9 shows that the addition of poly C is visible in a 5 hour reaction, note the shift upward of staining intensity of the "extension + poly C" lane. The "band purified + poly C" lane further substantiates that extension in the presence of dCTP occurred.

Conversion of Extension Products to Random Protein Coding Sequences

TdT extension products were converted to dsDNA by filling in the primer template junction created by annealing a poly G tailed GS5 oligonucleotide to the poly C region of extension products. The dsDNA fill-in product was transcribed (minus the methyl G cap) to produce a pool of random RNAs that contained, from 5' to 3', a GS5 sequence, a poly G sequence, a random sequence generated by TdT, and a sequence complementary to TDT3P. This initial library of random RNA was amplified by RT-PCR.

Figure 10 shows a pilot RT-PCR for the optimization of PCR rounds. As expected amplified product increased with the number of PCR rounds. Optimal amplification for this initial library occurd at 30 rounds, since this yielded a library with the most diversity in length. There was a smooth transition from small PCR products to the largest transition products. PCR at 25 rounds did not produce large products whereas PCR at 35 rounds was biased for shorter products. Since one of the goals was to increase the size of random peptides that can be displayed, the initial library was amplified for 30 rounds. At this point the library is ready to be transcribed and translated.

Translation of Random Transcripts

The random protein coding sequences from above were transcribed in the presence of +/- methyl G cap analog and DNase treated. This yielded a pool of random RNA transcripts that were translated in vitro with wheat germ extract. RNA's within the pool contained part of the GS5 primer sequence (the most 5' portion of the GS5 sequence, the T7 promoter, is lost during transcription) that had been poly G tailed. The poly G tail combined with the AT at the 3' end of GS5 created a start codon for translation.

Translation efficiency of random library RNA was determined by TCA precipitation and scintillation counting of ^{35}S-methionine. Translation of capped library mRNA reactions had incorporation rates 80% of BMV control reactions at optimal potassium acetate concentrations (Figure 11). Translation of capped RNA yielded a smear of protein with similar intensity to that of the positive BMV control.

The molecular weight range for this protein smear was estimated from a standard BMV curve. The upper and lower size limits of the protein smear were 50 kilodaltons and 30 kilodaltons respectively (Figure 12). Although wheat germ extract is supposed to translate uncapped RNA near the efficiency of capped RNA, little translation product can be observed in the uncapped lane.

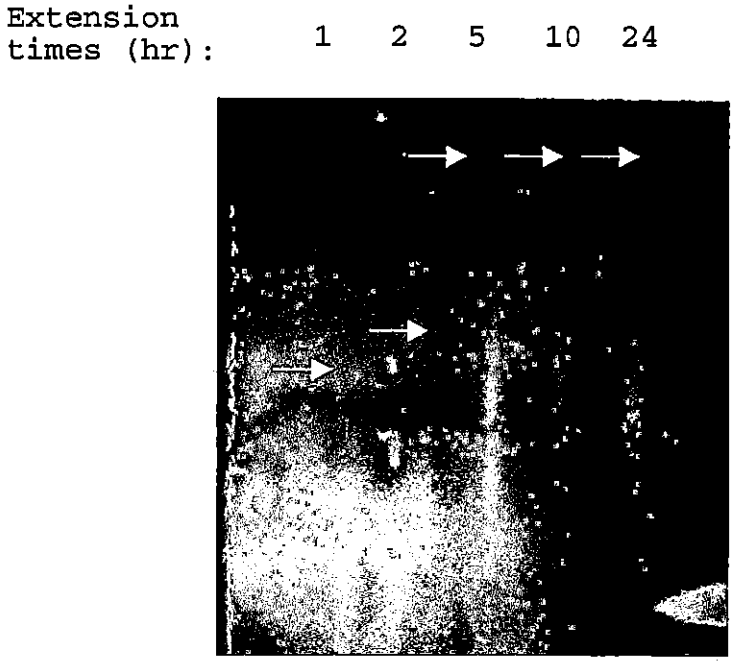


Figure 5: Timed extensions of the primer TDT3P. TdT was used to extend the oligonucleotide TDT3P under optimal extension conditions. The maximum extension length is achieved after 5 hours of reaction time (product smears for 5, 10, and 24 hours are of equal heights). The major product size, intensely staining area of the product smear, does increase with increased extensions. This is evident by a shift upward of the major product in the 10 and 24 hour extensions as compared to the 5 hour extension. Line arrows indicate maximum height of the product smear. Block arrow indicates size of TDT3P.

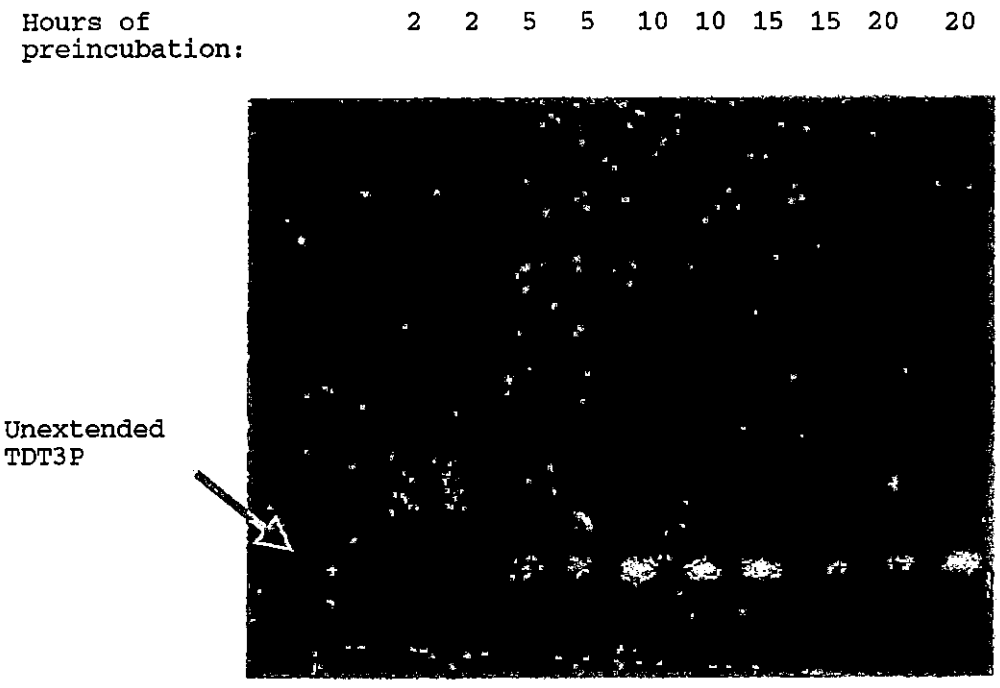


Figure 6: Effects of TdT preincubation. 5 hour TdT extension reactions were performed after preincubating TdT at 37°C for the stated amount of time. Preincubations of TdT were done in 2X reaction buffer and with a 2X amount of TdT. After incubation time, an aliquote was diluted to 1X with optimal concentrations of primer, dNTPs, and Co^{2+}. Extension reactions were carried out in duplicate.

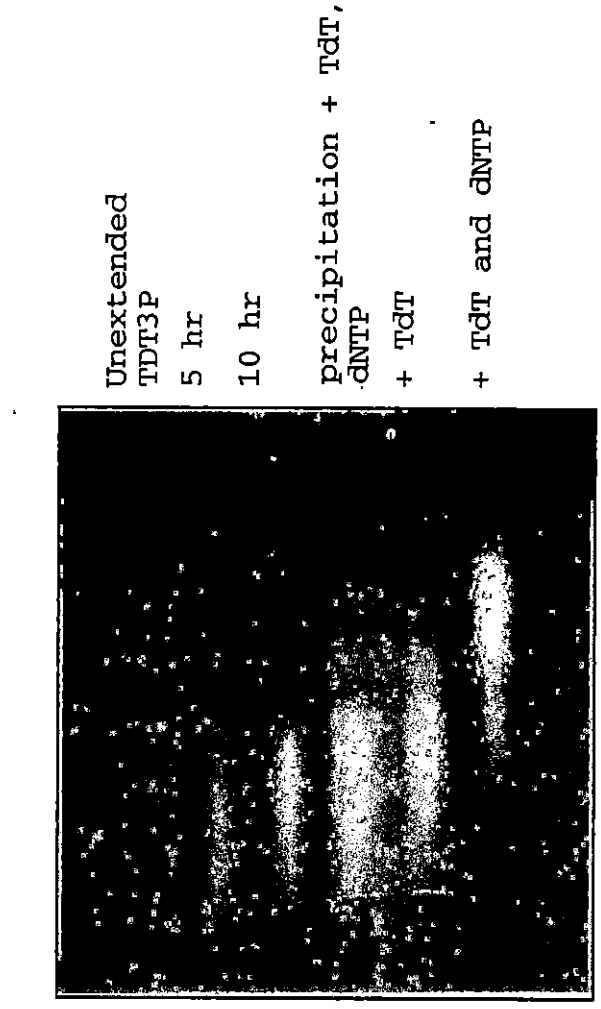


Figure 7: Effect of TdT supplementation on extension length. TDT3P was extended under optimal conditions for 5 hours and then received either 5 more hours of incubation or various treatments. Extension reactions were electrophoresed on a 8% denaturing polyacrylamide gel. Once again it is shown that TdT lacks the ability increase the length of TdT3P after 5 hours. Although precipitation and TdT supplementation increase the obtainable extension length these procedures do not allow complete extension of all 20 picomoles of TDT3P present in the reaction. Supplementing the extension reaction at 5 hours with TdT (0.8 U) and dNTP (0.5 mM) allowed for the longest extension products and extension of nearly all TDT3P present in the reaction.

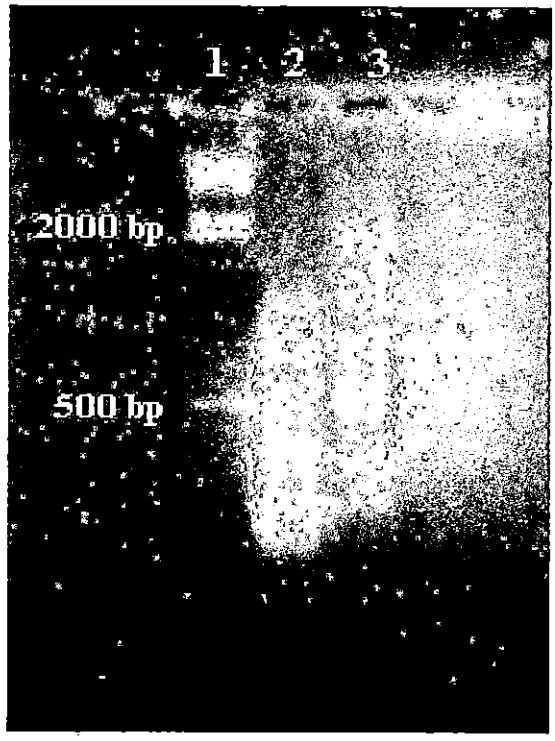


Figure 8: Maximum Extension Lengths. TDT3P primer extension reactions under optimal conditions are shown on a 1% agarose gel. Lane 1 is Lambda HindIII size standards with 2000 bp and 500 bp bands labeled. Lane 2 is a 24 hour extension with additional TdT (0.8 U) and dNTP (0.5 mM) added at the 5 and 10 hour time points. Lane 3 is the reaction from lane 2 with additional TdT (0.8 U) and dNTP (0.5 mM) added at the 5 and 10 hour time points. Lane 3 shows the bulk of the reaction is between 1000 and 2000 bp.

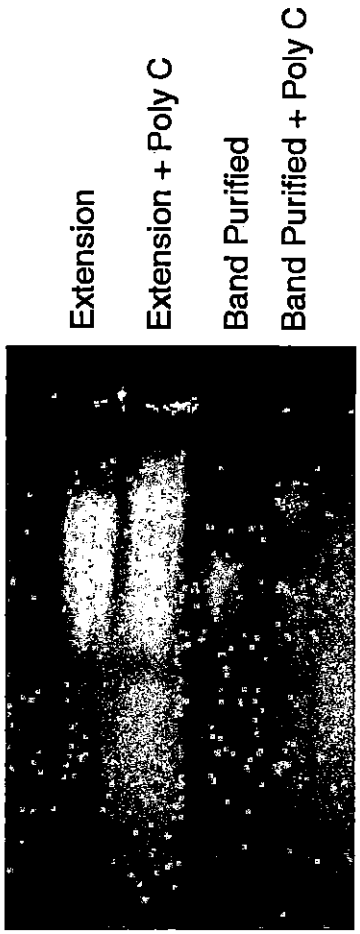


Figure 9: Poly C tailing reactions. A random extension reaction was ethanol precipitated away from the equimolar dNTP mix and extended using only dCTP. Band purification of the same extension reaction performed to examine the amount of extension during a 5 hour Poly C tailing reaction.

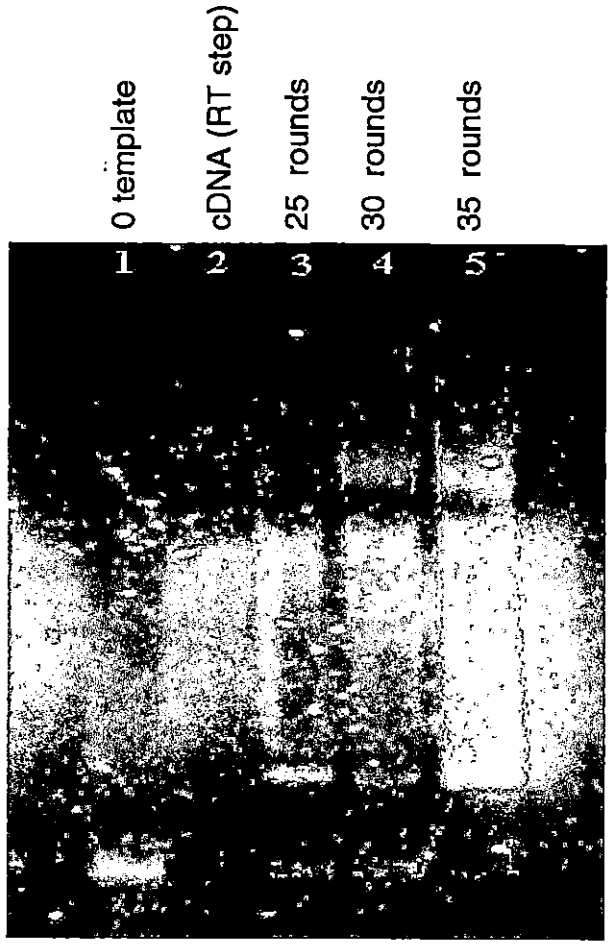


Figure 10: RT-PCR amplification. RNA transcribed from the fill-in product was amplified by RT-PCR. PCR amplification of 35 rounds produced a greater amount of product compared to 25 and 30 rounds. However, the smallest products began to dominate the product mixture. The optimal number of rounds for amplification was 30 because it had no product bias and contained larger product when compared to 25 rounds. This set of experiments was carried out with a 50°C annealing temperature. It was necessary to carry out pilot experiments to determine the best annealing temperature.

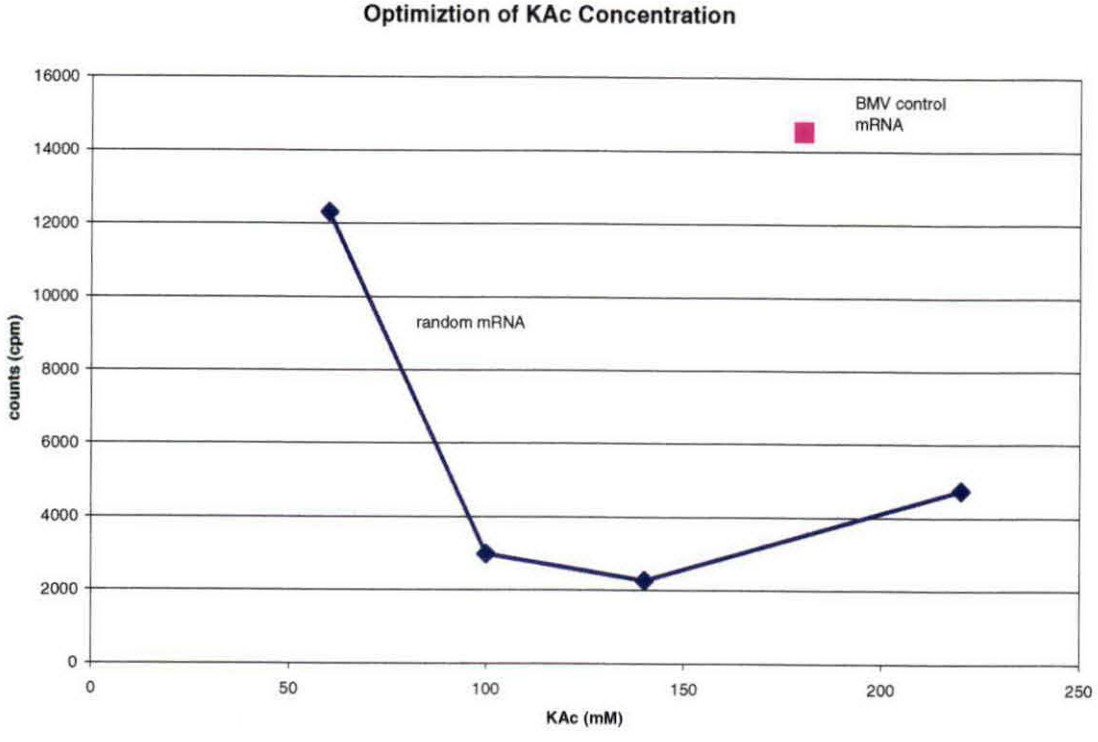


Figure 11: Optimization of KAc concentration. The optimal potassium acetate concentration for random mRNA was assayed by ^{35}S met incorporation. Translation reactions were TCA precipitated onto glass filter paper and placed in scintillation fluid. Scintillation counts for the translation reaction was subtracted from 0 mRNA reactions with a identical KAc concentration to yield total counts. The optimal KAc concentration for the BMV control was as suggested by the supplier.

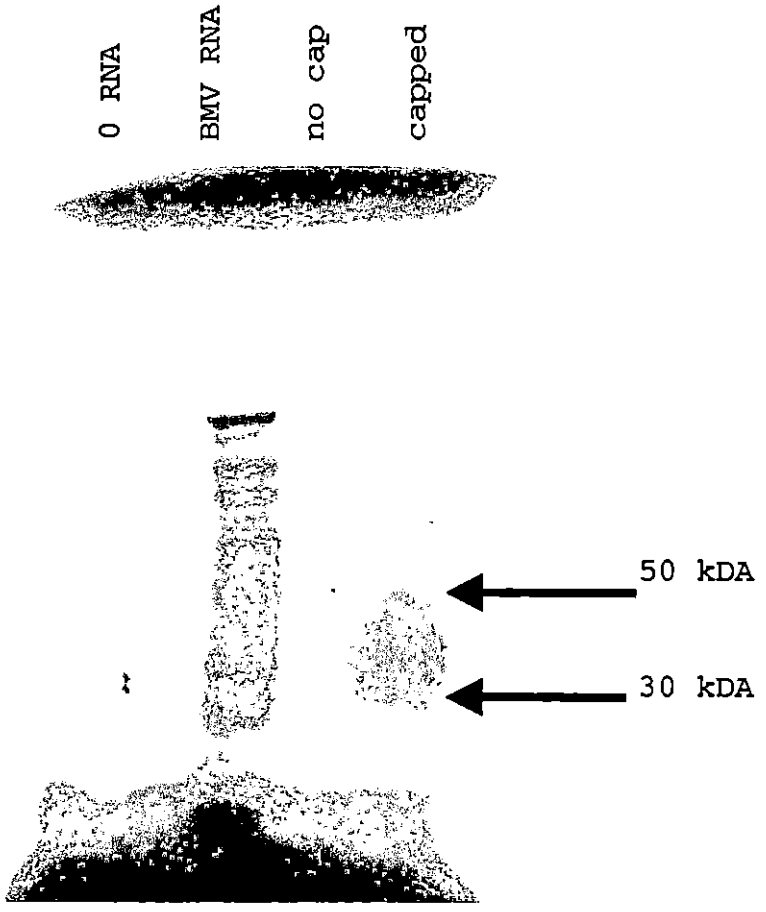


Figure 12: Translation of random transcripts. Transcription products from the initial amplified library were translated in the presence of ^{35}S-met. Translation of capped RNA produced a smear of proteins the ranged from 30 to 50 kDa. RNA without a cap did not translate as efficiently as capped RNA. The product sizes predict proteins with 250 to 415 amino acids.

Discussion

Terminal deoxynucleotidyl transferase (TdT) was used to generate long random protein coding sequences. TdT demonstrated the ability to add over 1500 random nucleotides to the oligonucleotide primer TDT3P during extension reactions. These ssDNA extension products were converted to dsDNA duplexes that contained T7 promoter sequences for transcription. Translation of RNA transcribed from these protein coding sequences yielded proteins that ranged from 30 kDa to 50 kDa.

Damiani, et al. (1982) used TdT to create random DNA duplexes that were cloned and sequenced. Sequence analysis revealed that TdT had a preference for purines which accounted for approximately 60% of the nucleotides incorporated during an extension. This problem was avoided by using Co^{2+} as the divalent cation in the extension reactions. The use of Co^{2+} negated any preference for nucleotides and allowed for longer extensions by increasing primer binding (Roychoudhury, et al. 1976). Damiani, et al. (1982) did detect open reading frames in their TdT generated DNA duplexes.

TdT did exhibit problems concerning its survivability during extension reactions with long incubation times. The enzyme exhibited marked decreases in the ability to extend a primer after 5 hours with extension totally disappearing after 10 hours (Figure 5). In fact, without reaction substrates TdT loses its ability to extend after 2 hours of incubation at 37°C (Figure 6). This problem has not been reported in the literature. Damiani, et al. (1982) performed their extensions for 13 hours and did not report any decrease in TdT's ability to extend their primers. It would be of

importance to determine which reaction components are responsible for the survivability of TdT in extension reactions. To get around this problem of marked decreases in TdT ability to perform extensions, the reaction was supplemented with fresh TdT and dNTP every 5 hours.

Translation products resulted in proteins that ranged from 30-50 kilodaltons. From these size estimations, it is predicted that proteins range from 250-415 amino acids. This number of random amino acids far exceed the current number of random amino acids, 6-38, used currently in peptide libraries (Scott and Smith 1990; Houghten, et al. 1991; Lam, et al. 1991; Cull, et al. 1992; Kay, et al. 1993; Lowman and Wells 1993). It is important to note that PCR amplification selects against long sequences (Figure 10), whereas translation selects for longer sequences (Figure 12). As to why shorter transcripts did not seem to translate, it is speculated that this is a peculiarity of the wheat germ translation system. Since a major goal was to optimize the system for the largest proteins, the lack of translation of short transcripts does not signify a bias within the system.

The diversity of this library far exceeds the number of allowable species. For example, the number of possible sequences for a 250 amino acid protein is 20^{250}. In our library creation strategy the input amount of primer to be extended was 20 picomoles or 10^{13} pieces of DNA. This means that all sequence possibilities can not be sampled by selection. However, this is no problem due to the ability of peptide libraries to efficiently find binding proteins (Clackson and Wells 1994). That is a protein will be selected for that wins out over the other proteins in the population.

As noted before our library protein coding sequences are completely random in regards to base composition. This is unlike current biological peptide libraries which use the NNK motif to code for proteins (Scott and Smith 1990; Cull, et al. 1992; Mattheakis, et al. 1994). One negative consequence of completely random sequence is the introduction of stop codons in the middle of the protein coding sequence. The NNK motif only codes for the amber stop codon or UAG (Scott and smith 1990). This library creation strategy allows for the creation of the UAG, UAA, and UGA stop codons. The increased probability of stop codon insertion would lead to more truncated proteins within the population. That is, the random sequence space following the stop codon could never be sampled during the selection process.

The major innovation of polysome display is the use of library creation and amplification enzymes entirely in vitro (Mattheakis, et al. 1994). The consequence of using a cell free system is that diversity will not be lost due to transformation efficiency (Clackson and Wells 1994; Mattheakis, et al. 1994; Hanes and Pluckthun 1997). Typical biological peptide libraries contain 10^8 members, enough to code for 4-5 random amino acids. Randomizing more residues than this would increase the diversity above the number of members allowable by transformation (Clackson and Wells 1994). Our library creation strategy is completely cell free and can be easily plugged into polysome display systems such as the Mattheakis, et al. (1994) system.

There is an increasing effort to extend peptide libraries beyond their current abilities to find novel ligands. The emphasis has been placed on discovering a technique to display and select large folded protein structures. Two systems that are

modifications of polysome display have been developed. Hanes and Pluckthun (1997) based their "ribosome display" method on the prokaryotic transcription system. Additional components such as protein disulfide isomerase and other protein chaperones were needed to assist in proper folding. He and Taussig (1997) used the eukaryotic, rabbit reticulocyte translation system which had previously been used to translate properly folded antibody fragments for their "ARM" display system. Both ribosome display and ARM display have displayed antibody fragments and used their epitopes as selection substrates to demonstrate their usefulness (Hanes and Pluckthun 1997; He and Taussig 1997).

Creators of both ribosome display and ARM display suggest that their systems can be adapted to display any protein in its native folded state and have the ability to be used as peptide libraries (Hanes and Pluckthun 1997; He and Taussig 1997). Libraries from the above methods still would not be able to display completely random proteins due to random size restrictions derived from the inability to synthesize long oligonucleotides. These systems, although major breakthroughs, would have to rely on randomization strategies such as the ones used to optimize human growth hormone by phage display (Lowman and Wells 1993).

Our library creation strategy using TdT as a random sequence generator would allow display of completely random proteins when plugged into either the ribosome display or ARM display systems. For the first time, the in vitro evolution of novel proteins would be possible.

Literature Cited

Blondelle, S. and Houghten, R. **1996.** Novel antimicrobial compounds identified using synthetic combinatorial library technology. *Trends in Biotechnology* **14**: 60-65.

Blundell, T. 1994. Problems and solutions in protein engineering – towards rational design. *Trends in Biotechnology* **12**: 145-148.

Chang, L. M. S. and Bollum, F. J. **1970.** *Proceedings of the National Academy of Sciences* **65**: 1041-1048.

Cull, M., Miller, J., and Schatz, P. **1992.** Screening for receptor ligands using large libraries of peptides linked to the C terminus of the *lac* repressor. *Proceedings of the National Academy of Sciences* **89**: 1865-1869.

Damiani, G., Scovassi, I., Romagnoli, S., Palla, E., Bertazzoni, U., and Sgaramella, V. **1982.** Sequence analysis of heteropolymeric DNA synthesized in vitro by the enzyme terminal deoxynucleotidyl transferase and cloned in *Escherichia coli*. *Nucleic Acid Research* **10**: 6401-6405.

Dooley, C., Chung, B., Wilkes, P., Schiller, J., Bidlack, G., Pasternak, G., and Houghten, R. **1994.** An all D-amino acid opioid peptide with central analgesic activity from a combinatorial library. *Science* **266**: 2019-2022.

Eichler, J., Appel, J., Blondelle, S., Dooley, C., Dorner, B., Ostresh, J., Perez-Paya, E., Pinilla, C., and Houghten, R. **1995.** Peptide, peptidomimetic, and organic synthetic combinatorial libraries. *Medicinal Research Reviews* **15**: 481-496.

Fowlkes, D., Adams, M., Fowler, A., and Kay, B. **1992.** Multipurpose vectors for peptide expression of the M13 viral Surface. *Biotechniques* **13**: 422-427.

Gates, C., Stemmer, W., Kaptein, R., and Schatz, P. **1996.** Affinity selective isolation of ligands from peptide libraries through display on a lac repressor "headpiece dimer". *Journal of Molecular Biology* **255**: 373-386.

Greisman, H. A. and Pabo, C. O. **1997.** A general strategy for selecting high-affinity zinc finger proteins for diverse DNA target sites. *Science* **275**: 657-661.

Hanes, J. and Pluckthun, A. **1997.** In vitro selection and evolution of functional proteins by using ribosome display. *Proceedings of the National Academy of Sciences* **94**: 4937-4942.

He, M. and Taussig, M. **1997.** Antibody-ribosome-mRNA (ARM) complexes as efficient selection particles for in vitro display and evolution of antibody combining sites. *Nucleic Acids Research* **25**: 5132-5134.

Houghten, R. A., Pinilla, c., Blondelle, S. E., Appel, J. R., Dooley, C. T. and Cuervo, J. H. **1991.** Generation and use of synthetic peptide combinatorial libraries for basic research and drug discovery. *Nature* **354**: 84-86.

Kay, B., Adey, N., He, Y., Manfredi, J., Mataragnon, A., and Fowlkes, D. **1993.** An M13 phage library displaying random 38-amino-acid peptides as a source of novel sequences with affinity to selected targets. *Gene* **128**: 59-65.

Ladner, R. **1995.** Constrained peptides as binding entities. *Trends in Biotechnology* **13**: 426-430.

Lam, K., Salmon, S., Hersh, E., Hruby, V., Kazmierski, W., and Knapp, R. **1991.** A new type of synthetic peptidde library for identifying ligand-binding activity. *Nature* **354**: 82-86.

Lowman, H. B. and Wells, J. A. **1993.** Affinity maturation of human growth hormone by monovalent phage display. *Journal of Molecular Biology* **234**: 564-578.

Martens, C., Cwirla, S., Lee, r., Whitehorn, E., Chen, Esther, Bakker, A., Martin, E., Wagstrom, C., Gopalan, P., smith, C., Tate, E., Koller, K., Schatz, P., Dower, W., and Barrett, R. **1995.** Peptides which bind to E-selectin and block neutrophil adhesion. *Journal of Biological Chemistry* **270**: 21129-21136.

Mattheakis, l., Bhatt, R., and Dower, W. **1994.** An *in vitro* polysome display system for identifying ligands from very large peptide libraries. *Proceedings of the National Academy of Sciences* **91**: 9022-9026.

Matthews, D., and Wells, J. **1993.** Substrate phage: selection of protease substrates by monovalent phage display. *Science* **260**: 1113-1117.

Roychoudhury, R., Jay, E., and Wu, R. **1976.** Terminal labeling and addition of homopolymer tracts to duplex DNA fragments by terminal deoxynucleotidyl transferase. *Nucleic Acid Research* **3**: 863-877.

Sambrook, J., Fritch, E., and Maniatis, T. **1989.** *Molecular cloning: a laboratory manual.* Cold Spring Harbor Press, Cold Spring Harbor, New York.

Sander, C. **1994.** Design of protein structures: helix bundles and beyond. *Trends in Biotechnology* **12**: 163-167.

Scott, J. and Smith, G. **1990.** Searching for peptide ligands with an epitope library. *Science* **249**: 386-389.

Ukaka, K. **1996.** Decrypting class I MHC-bound peptides with peptide libraries. *Trends in biochemical Sciences* **21**: 7-11.