The Dissertation Committee for Letha Jane Sooter Certifies that this is the approved version of the following dissertation:

Automation of *in vitro* selections

Committee:

Andrew D. Ellington, Supervisor

George Georgiou

Barrie Kitto

Dean Neikirk

C. Grant Willson

Automation of *in vitro* selections

by

Letha Jane Sooter, B.S.

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

December, 2004

To my family.

Acknowledgements

Thank you to Andy Ellington and past and present lab members.

Automation of *in vitro* selections

Publication No._____

Letha Jane Sooter, Ph.D.

The University of Texas at Austin, 2004

Supervisor: Andrew D. Ellington

Automation is a powerful tool, which may be used to increase the throughput of many otherwise laborious manual manipulations. Aptamer and deoxyribozyme selections are prime examples of processes, which require substantial amounts of time at the bench, but which are amenable to automation. Double-stranded DNA binding sites that bound with high affinity to the nuclear factor kappa B (NFκB) p50 homodimer were selected using a Tecan Genesis workstation. This was followed by selections against whole cell lysates. The resultant sequences represented an array of transcription factor binding sites within the *E. coli* genome. Finally, a Biomek2000 was used to perform a deoxyribozyme ligase selection, which formed an unnatural phosphorothioate linkage.

Table of Contents

Chapter 1: Introduction

IN VITRO SELECTIONS

Aptamers are nucleic acid species, which bind a target molecule with both high affinity and specificity. The nucleic acids can be either DNA or RNA, double stranded or single stranded. The target for binding can be anything. Aptamers that bind specifically have been selected against everything from ions to small organics, peptides, proteins, viruses and tissues. (Famulok and Mayer, 1999; Brody and Gold, 2000; Hesselberth et al., 2001).

Deoxyribonucleic acid (DNA) and ribonucleic acid (RNA) are composed of sugar phosphate backbones and a series of nucleobases. The difference between the two is that DNA bears a hydrogen at the 2' carbon on the sugar moiety while RNA bears a hydroxyl in this position, and RNA utilizes uracil instead of thymine, which contains a methyl group on the five carbon of the base. Adenine and Guanine are purines, which have a bicyclic structure. Pyrimidines, thymine, cytosine, and uracil, have a a single six member ring as their basic structure. In a double stranded form, the highly charged phosphate backbone is primarily what is exposed on the surface of the molecule. However, in the single stranded form an array of polar and nonpolar surfaces are available to participate in the formation of secondary and tertiary structures. This provides a host of possibilities for binding surfaces. Single stranded DNA and RNA also fold into structures that place specific functional groups at precise locations to allow optimal target binding properties. The 2' hydroxyl, which is present in RNA gives these molecules additional binding functionality, however is also makes them more susceptible to degradation.

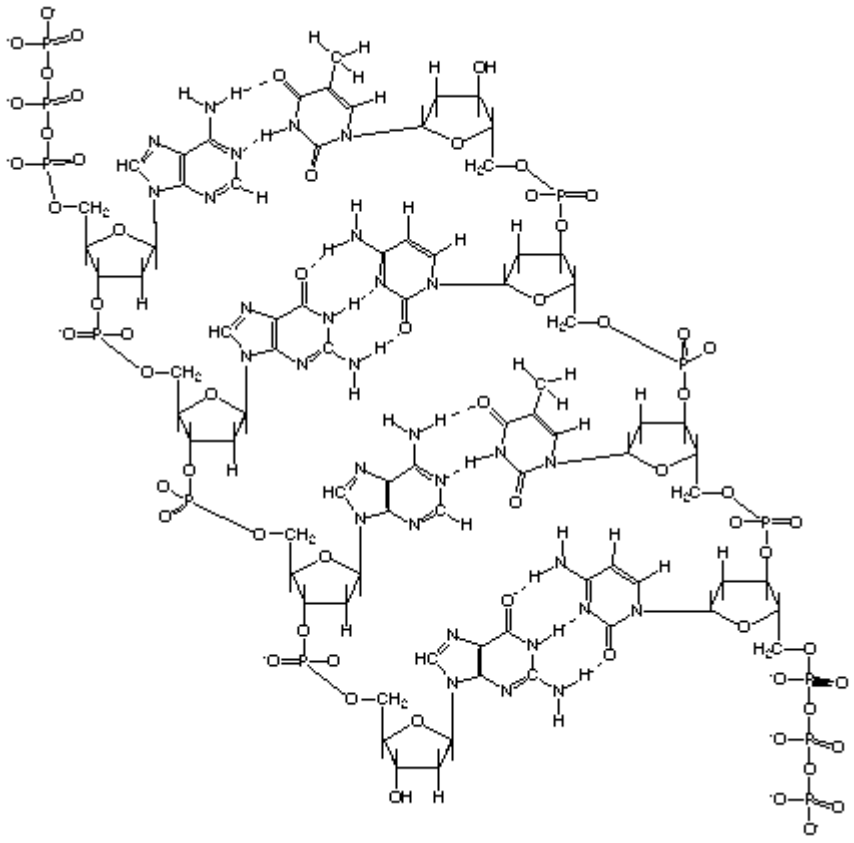


Figure 1: Double stranded DNA.

(http://web.mit.edu/esgbio/www/lm/nucleicacids/dna.html**)**

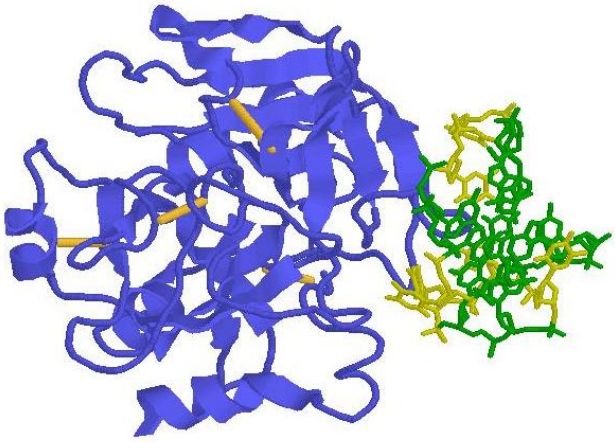


Figure 2: Structure of ssDNA aptamer bound to thrombin. (www.pdb.org) Thrombin is shown as a ribbon structure in blue. The ssDNA aptamer is shown as a stick structure in yellow and green.

Aptamers are selected through a cyclic process of binding species isolation followed by amplification. They can be selected from random sequence populations. Oligonucleotides are synthesized that contain a random sequence region of 30 to 200 residues, flanked by constant regions required for amplification via the polymerase chain reaction (PCR). Chemically synthesized single-stranded DNA pools can be converted into double-stranded DNA pools through PCR, while RNA pools and modified RNA pools can be generated by inclusion of a T7 RNA polymerase promoter in one of the constant regions and *in vitro* transcription. Nucleic acid pools typically contain from 10^{13} to 10^{15} different sequences. Functional species are separated from non-functional species by immobilization. For example, aptamers are typically captured by affinity chromatography or filtration partition, while catalytic nucleic acids (ribozymes, deoxyribozymes) can be captured following the addition of an activated substrate (e.g., a

biotinylated oligonucleotide) to themselves. Following isolation, functional species can be amplified by a combination of reverse transcription, PCR, and *in vitro* transcription, as is necessary to regenerate the pool. Multiple cycles of selection and amplification result in the preferential enrichment of those binding or catalytic species with the highest affinities or activities.

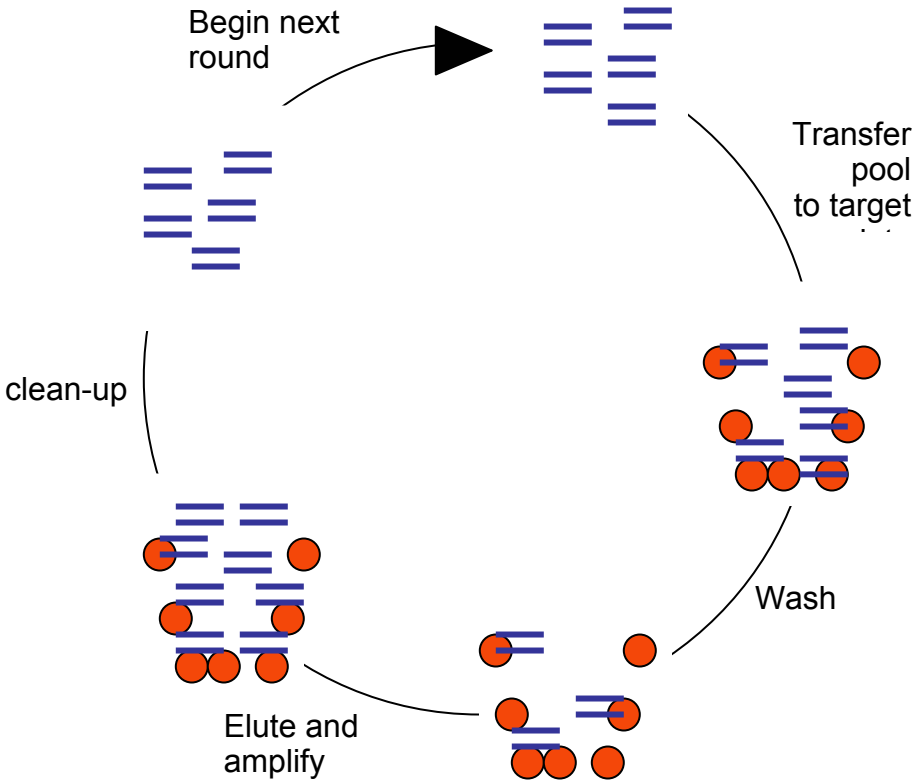


Figure 3: General Selection Scheme. Target molecules are shown as orange circles and nucleic acids are shown as double blue lines.

Aptamers have been selected against a wide variety of proteins, including both nucleic acid binding proteins, such as T4 DNA polymerase and HIV-1 Rev, and a surprising number of non-nucleic acid binding proteins. In general, anti-protein aptamers

recognize basic patches on protein surfaces. For example, the arginine-rich motifs (ARMs) of many viral proteins are recognized by aptamers (Ellington et al., 1996), the phosphate-binding pockets of both kinases (Conrad et al., 1994) and phosphatases (Bell et al., 1998). Aptamers also have an affinity for pockets on protein surfaces, such as the combining sites of antibodies (Tsai et al., 1992) or the active sites of enzymes (Tuerk et al., 1992). Most importantly, aptamers recognize their targets with high specificity, and can typically discriminate between protein targets that are highly homologous (Conrad et al., 1994; Hirao et al., 1998). There are numerous examples of aptamers that discriminate strongly between proteins that are greater than 95% similar in sequence, and some that differ by only a few amino acids (Conrad et al, 1994; Hicke et al., 2001).

Aptamers can not only be selected against purified targets or antigens, but also against complex targets, such as whole cells. In one of the most intriguing examples to date, Homann and Goringer (1999) were able to select aptamers against whole trypanosomes; these were eventually found to bind to the variant surface glycoprotein. (Lorger et al., 2003). Another example is a selection that was carried out against red blood cell (RBC) membranes, and a series of aptamers specific to different targets on the RBC membrane surface were identified (Morris et al., 1998). Minimized versions of some of the aptamers were as small as 22 nucleotides, yet retained high affinity (K_d = 1.6nM) and specificity for their target ligand. Similarly, Pan et al. (1995) generated both RNA and modified nuclease-resistant RNA aptamers to Rous sarcoma virus (RSV).

Aptamers can be applied in target identification as substitutes for antibodies or other staining reagents. For example, aptamers have been adapted to sandwich assays ('ELONAs,' for enzyme-linked oligonucleotide assay) and were able to sensitively quantitate target proteins (Drolet et al., 1996). Later adaptations of the sandwich assay

even allow the detection and quantitation of physiological concentrations of proteins in extracts (Rye and Nustad, 2001). Fluorescent aptamers have even been used to label cells in flow cytometry (Davis et al., 1996).

Aptamers have proven to be valuable both as therapeutic and diagnostic agents. Aptamers against potential therapeutic targets including IgE, thrombin, PTPase, and others have shown efficacy in tissue culture experiments and some have even been successful in animal models. An anti-thrombin aptamer was used in place of heparin for anti-coagulation during heart bypass surgery in canines (DeAnda et al., 1994). In another animal model, aptamers against inflammation factor human neutrophil elastase (hNE) were shown to significantly reduce lung inflammation in rats and had better specificity for their target than an anti-elastase IgG control (Bless et al., 1997). An aptamer that inhibits vascular endothelial growth factor (VEGF) was introduced into humans for the treatment of macular degeneration, making it the first aptamer to reach clinical trials (Tucker et al., 1999).

Despite the many useful attributes of aptamers, there are some drawbacks. The two most notable are the large amount of time required to select them and the importance of specific selection conditions. To obviate these difficulties, robots can be used to perform the selections. Instead of several weeks to complete a selection, the robots can complete them several hours (Cox et al., 2002). In addition, parallel selections may be performed under different buffer conditions. The robots can therefore more quickly identify aptamers and the optimal conditions under which they should be selected.

LIQUID HANDLING ROBOTICS

Beckman Biomek 2000

The Biomek 2000 (Figure 4) was introduced by Beckman Coulter. It is run using Bioworks software which has a very object oriented interface. The machine has detachable tools which hold one or eight pipetting channels. Each of the pipetting heads has an optimum volume range and uses a particular set of tips. The Biomek also has a detachable gripper tool which can be used for moving microplates and tip boxes around the worksurface. Third party devices such as MJ Research thermal cyclers, Dynal auto-96 magnetic bead separators, and Beckman stacker carousels may also be integrated using special drivers. Custom devices were also constructed by Tim Riedel including an "enzyme cooler" which will keep contents of a microplate at -20C and a "refrigerator plate" which will keep the contents of a microplate at 4C.

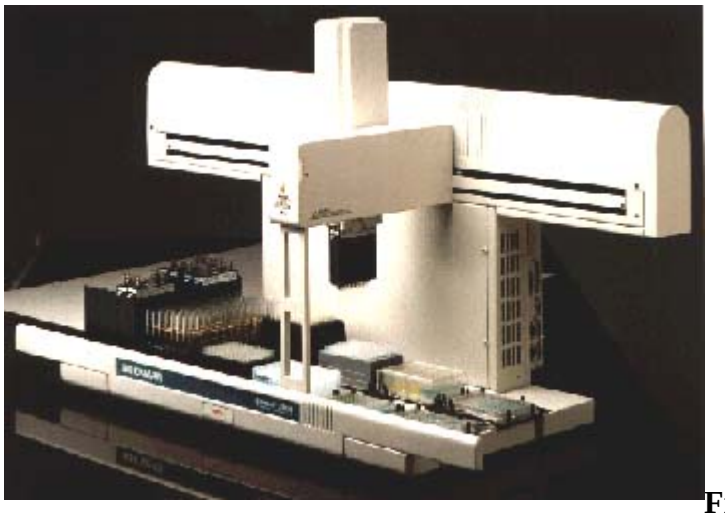


Figure 4: Biomek2000 **(**http://www.beckman.com/products/instrument/automatedsolutions/biomek/biomek2000_inst_dcr.asp**)**

Beckman Biomek FX

The Biomek FX in our lab is a one pod system. This pod contains a fixed 96 channel pipetting head and with a gripper tool. The gripper tool may be used to move microplates and tip boxes around the worksurface, but it cannot reach off the worksurface. The Biomek FX has an integrated stacker carousel and an integrated solid phase extraction device, thermal plate, and shaker plate. We have been unable to integrate third party devices such as a Dynal auto-96 or an MJ Research thermal cycler with this unit. The software is very user friendly. For the average user, the learning time is minimal compared to other robots. The interface is very object oriented and it does much of the "thinking" for the programmer. This robot is optimal for serial dilutions and aliquotting solutions from reservoirs into plates.

Figure 5: BiomekFX

(http://www.beckman.com/products/instrument/automatedsolutions/biomek/biomekfx_inst_dcr.asp**)**

Tecan Genesis Workstation 200

The most advanced automated workstation available to us is the Tecan Genesis Workstation 200 (Figure 6). The Tecan has two pods, a liquid handling pod and a robotic manipulator arm. The liquid handling pod is composed of eight, independently controlled pipetting tips that have liquid sensing capabilities and can accurately pipette between 0.5 and 1000μL into plates with 1 – 384 wells. The robotic manipulator arm can move items on and off the worksurface, allowing access to additional equipment to the side of the workstation. The worksurface also contains a number of different carriers and devices; for example, temperature control units are available for buffer and enzyme storage, incubators are present for temperature-controlled microplate storage, and a vacuum filtration unit and a centrifuge can be interfaced with various purification tasks or kits, such as those from Qiagen, Promega, or Millipore. Successful integration between centrifugation and liquid-handling robots is rare, and we are one of the few academic labs to have accomplished this. An integrated Tecan Columbus plate washer and MJ Research thermal cycler are essential for automated panning selections. The custom integration of an Invitrogen 96-well agarose gel system has facilitated high throughput visualization of PCR products, and a Safire microplate reader has been integrated for rapid sample screening.

Figure 6: Tecan Genesis **(**http://www.tecan-us.com/us-index/com-pr-in/com-pr-in-ro_li_entry-3/com-pr-in-ro_li-genesis_rsp.htm**)**

REFERENCES

Bell SD, Denu JM, Dixon JE, Ellington AD (1998). RNA molecules that bind to and inhibit the active site of a tyrosine phosphatase. J Biol Chem 273:14309-14.

Bless NM, Smith D, Charlton J, Czermak BJ, Schmal H, Friedl HP, Ward PA (1997). Protective effects of an aptamer inhibitor of neutrophil elastase in lung inflammatory injury. Curr Biol 7:877-80.

Conrad R, Keranen LM, Ellington AD, Newton AC (1994). Isozyme-specific inhibition of protein kinase C by RNA aptamers. J Biol Chem 269:32051-4.

Cox JC, Hayhurst A, Hesselberth J, Bayer TS, Georgiou G, Ellington AD (2002). Automated selection of aptamers against protein targets translated *in vitro*: from gene to aptamer. Nucleic Acids Res 30:e108.

Davis KA, Abrams B, Lin Y, Jayasena SD (1996). Use of a high affinity DNA ligand in flow cytometry. Nucleic Acids Res 24:702-6.

DeAnda A Jr, Coutre SE, Moon MR, Vial CM, Griffin LC, Law VS, Komeda M, Leung LL, Miller DC (1994). Pilot study of the efficacy of a thrombin inhibitor for use during cardiopulmonary bypass. Ann Thorac Surg 58:344-50.

Drolet DW, Moon-McDermott L, Romig TS (1996). An enzyme-linked oligonucleotide assay. Nat Biotechnol 14:1021-5.

Ellington AD, Leclerc F, Cedergren R (1996). An RNA groove. Nat Struct Biol 3:981-4.

Famulok M, Mayer G (1999). Aptamers as tools in molecular biology and immunology. Curr Top Microbiol Immunol 243:123-36.

Hesselberth J, Robertson MP, Jhaveri S, Ellington AD (2000). *In vitro* selection of nucleic acids for diagnostic applications. J Biotechnol 74:15-25.

Hicke BJ, Marion C, Chang YF, Gould T, Lynott CK, Parma D, Schmidt PG, Warren S (2001). Tenascin-C aptamers are generated using tumor cells and purified protein. J Biol Chem 276:48644-54.

Hirao I, Spingola M, Peabody D, Ellington AD (1998). The limits of specificity: an experimental analysis with RNA aptamers to MS2 coat protein variants. Mol Divers 4:75-89.

Homann M, Goringer HU (1999). Combinatorial selection of high affinity RNA ligands to live African trypanosomes. Nucleic Acids Res 27:2006-14.

Lorger M, Engstler M, Homann M, Goringer HU (2003). Targeting the variable surface of African trypanosomes with variant surface glycoprotein-specific, serum-stable RNA aptamers. Eukaryot Cell 2:84-94.

Morris KN, Jensen KB, Julin CM, Weil M, Gold L (1998). High affinity ligands from *in vitro* selection: complex targets. Proc Natl Acad Sci U S A 95:2902-7.

Pan W, Craven RC, Qiu Q, Wilson CB, Wills JW, Golovine S, Wang JF (1995). Isolation of virus-neutralizing RNAs from a large pool of random sequences. Proc Natl Acad Sci U S A 92:11509-13.

Rye PD, Nustad K (2001). Immunomagnetic DNA aptamer assay. Biotechniques 30:290-2, 294-5.

Tsai DE, Kenan DJ, Keene JD (1992). *In vitro* selection of an RNA epitope immunologically cross-reactive with a peptide. Proc Natl Acad Sci U S A 89:8864-8.

Tucker CE, Chen LS, Judkins MB, Farmer JA, Gill SC, Drolet DW (1999). Detection and plasma pharmacokinetics of an anti-vascular endothelial growth factor oligonucleotide-aptamer (NX1838) in rhesus monkeys. J Chromatogr B Biomed Sci Appl 732:203-12.

Tuerk C, MacDougal S, Gold L (1992). RNA pseudoknots that inhibit human immunodeficiency virus type 1 reverse transcriptase. Proc Natl Acad Sci U S A 89:6988-92.

http://web.mit.edu/esgbio/www/lm/nucleicacids/dna.html

http://www.beckman.com/products/instrument/automatedsolutions/biomek/biomekfx_inst_dcr.asp

http://www.beckman.com/products/instrument/automatedsolutions/biomek/biomek2000_inst_dcr.asp

http://www.pdb.org

http://www.tecan-us.com/us-index/com-pr-in/com-pr-in-ro_li_entry-3/com-pr-in-ro_li-genesis_rsp.htm

Chapter 2: Nuclear Factor kappa B selection

INTRODUCTION

Double-Stranded DNA Aptamers

Double-stranded DNA (dsDNA) aptamer selections have proven to yield accurate *in vivo* binding sequences for dsDNA binding proteins. For example, manual selections against the NFκB p50 homodimer have produced the *in vivo* binding site (Kunsch et al, 1992). Benbrook and Jones performed selections against CREB1, CREB2, and the CREB2/cJun heterodimer (1994). All of the selected sequences varied from the known *in vivo* binding sequence by three or less nucleotides out of the 10bp binding site. Selections have also been used to successfully identify unknown binding sites, for example, the HspR binding site in *H. pylori* (Delany et al. 2002). Selection experiments yielded two binding sites which were later confirmed by DNA footprinting.

A number of different selections have been performed against protein targets in lysates (Pollock and Treisman, 1990; Benbrook and Jones, 1994) Blackwell and Weintraub performed selections against MoyD and E2A homo- and heterodimers (1990). They found that the different dimers recognized a consensus CANNTG, but varied on the two internal residues according to which dimer was being selected against. These proteins were purified for the selection experiments and were later tested for their *in vivo* activation abilities (Huang et al, 1996). Interestingly, optimal *in vitro* binding sites were inactive *in vivo* while other selected sites exhibited binding activity similar to naturally occurring promoters. Also, manual selections against both purified myogenin and myogenin in nuclear extracts produced the known *in vivo* binding sequence (Wright et al

1991; Funk and Wright, 1992). The significance of this is that the selection in the lysate allowed the myogenin target to be a natural complex of proteins instead of a single purified protein.

In general, the sequence data from *in vitro* selection experiments corresponds to identified, natural binding sites and hence can be matched to promoter regions within genomes and used to identify genes that may be regulated by a transcription factor. There are a number of programs and databases that are available to assist with transcription factor binding site identification. For example, TRANSFAC, The Transcription Factor Database, is a compilation of known transcription factor binding sites, and EPD, the Eukaryotic Promoter Database, is a collection of eukaryotic promoters for which the transcriptional start sites have been experimentally determined. There is also software, such as SiteSeer (Boardman et al., 2003), that can interface directly with the TRANSFAC database and can aid in binding site identification.

Double-stranded DNA selections have been used to not only define DNA-binding sites for proteins, but also to determine the relative contributions of individual residues within a site to interactions with the target protein (Roulet et al., 2002). For example, after four rounds of *in vitro* selection against the CTF/NFI transcription factor, Bucher and co-workers sequenced over 10,000 possible binding sites and constructed a binding model based on the sequence distribution. The binding model was used to predict natural binding sites, which were subsequently experimentally verified.

With the availability of structural and sequence data for transcription factors, selections are now not the only means of identifying binding sites. Statistical weighting algorithms (Sinha and Tompa, 2002; 2003) can assist in the identification of new sites or subtle alterations in specificity. He and coworkers performed a selection against MetJ

and successfully identified the *in vivo* holo-repressor and apo-repressor binding sites. Liu and coworkers used known binding and selection data from MetJ to devise a model for predicting transcription factor binding sites. This model was used to identify additional, previously unknown, MetJ binding sites within the genome. (Liu et al, 2001).

In one of few examples, Church and co-workers have constructed a phage display library that expressed a variety of mouse Zif268 zinc finger domains and analyzed the specificities of individual fingers using a double-stranded DNA microarray (Bulyk et al., 2001). Newman and Keating (2003) examined the partnering specificities of all possible dimeric interactions between human leucine zipper transcription factors, using a protein (as opposed to double-stranded DNA) array. In addition, some transcription factor microarrays are commercially available (www.clontech.com; www.panomics.com).

We have established an automated selection protocol that targets double-stranded DNA binding proteins. The Tecan Genesis liquid handling robot was used to automate double-stranded DNA aptamer selections (Figure 6). The Tecan robot was chosen for this work, as opposed to the Beckman platforms, because its plate washer and versatile robotic arm enabled the implementation of high-stringency panning selections.

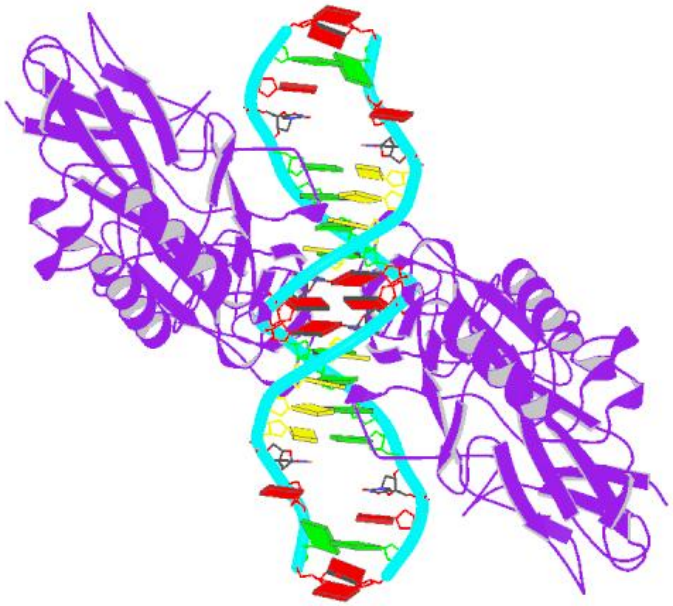


Figure 7: NFKB bound to DNA (www.pdb.org)

The *in vitro* selection process for dsDNA aptamers is known by a variety of names, inlcuding: CASTing (cyclic amplification and selection of targets)(Wright et al., 1991), SELEX (systematic evolution of ligands by exponential enrichment)(Tuerk and Gold, 1990), SAAB (selected and amplified binding site)(Blackwell and Weintraub, 1990), and TDA (target detection assay)(Thiesen and Bach, 1990). Regardless of the name, the selection process is essentially the same. As diagramed in Figure 8, a "pool" of oligonucleotides is synthesized with a randomized core region which ranges in size from 20 to 100 bases (Wright and Funk, 1993). Constant regions at the 5' and 3' ends of the DNA serve as the necessary priming areas for the polymerase chain reaction (PCR). The single-stranded pool is converted into double-stranded DNA via PCR, and the resultant product is purified. The initial dsDNA pool containing 10^{13} to 10^{15} different sequences is then incubated with a target protein or protein complex. Non-binding

species are partitioned away and the binding species are amplified via the polymerase chain reaction (PCR). The product is purified and subjected to additional rounds of selection for a total of four to eight rounds. The resultant pool is composed of the species which bind the target with the highest affinity. Following selection, the individual DNA species are sequenced. One or several families of sequences are typically present.

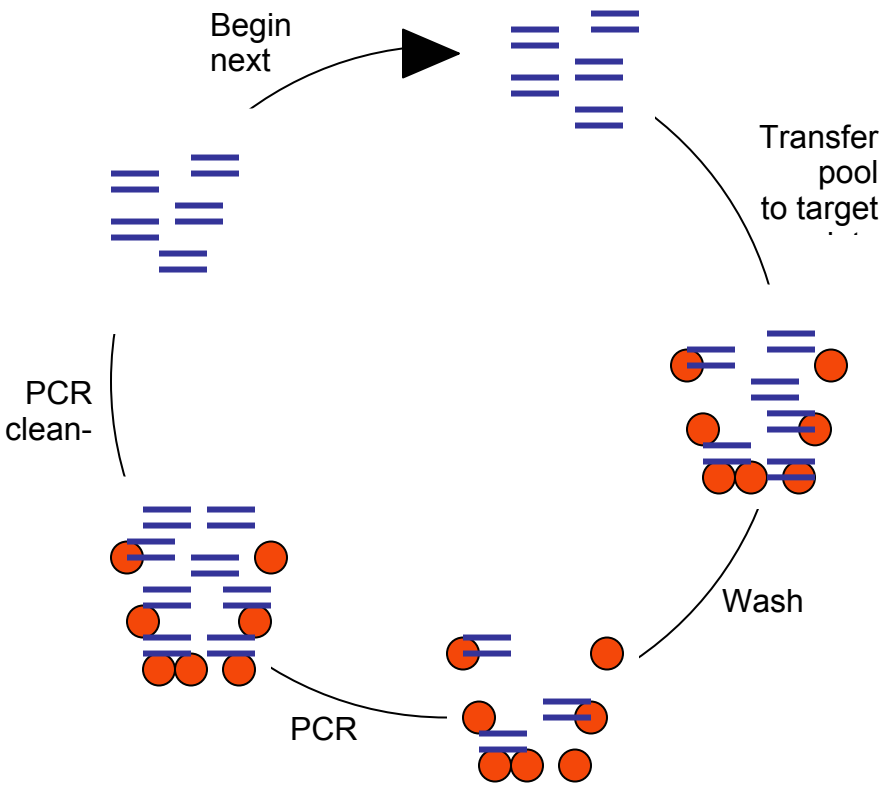


Figure 8: General *in vitro* dsDNA selection scheme. Nucleic acids are shown as parallel blue lines and target molecules are shown as orange circles.

Double-stranded DNA aptamer selections are typically used to identify the sequence of a binding site for a protein that binds dsDNA. This application is unlike typical aptamer applications. Single-stranded DNA and RNA aptamers are capable of folding into complex secondary and tertiary structures. The surfaces of these species are composed of both hydrophobic and hydrophilic regions. In contrast, dsDNA aptamers are linear. The negatively charged phosphodiester backbone is exposed to the environment while the bases reside within the double helix. As a result, dsDNA aptamers are not well suited for binding to a wide variety of targets (Figure 1). The proteins containing sites specific for binding dsDNA are, on the other hand, excellent selection targets.

Cellular functions are composed of a complex network of signals. In order to better understand how cells operate, it is important to understand which proteins control the expression of which genes. To this end, a variety of methods have been employed. DNA footprinting is one example. A protein is incubated with genomic DNA and then DNAse is added. All exposed DNA is digested, while the DNA that is protected by the protein remains intact. An alternative method is dsDNA aptamer selection. Families of binding sequences will be isolated. This is advantageous because there may be slight variations in promoter sequences *in vivo*, however sequences which bind but are not present in promoters may also be identified. Current literature shows that dsDNA aptamer selections typically yield species that are highly homologous to known promoter regions. When coupled with automation, a large number of transcription factor/promoter pairs can be identified in a short amount of time. Genome sequences can then be searched for the selected promoter sequences, and regulatory networks can be mapped. The identification of multiple transcription factor binding sites by a combination of

selection, sequence analysis, and database mining can potentially lead to the construction of a full description of the regulatory pathways in a cell. Unfortunately, the *in vitro* selection process can be extremely time-consuming.

A number of selections have been performed in extracts to identify double-stranded DNA sequences that bind to either specific proteins or protein complexes, including estrogen receptors (Medici et al., 1999); CRE (Benbrook and Jones, 1994); serum response factor and Fos (Pollock and Treisman, 1990); p53 (Funk et al., 1992), myogenin (Wright et al., 1992; Funk and Wright, 1992), and retinoblastoma-containing complexes (Oulette et al., 1992).

Manual selection methods have previously yielded binding sequences for a variety of transcription factors, including NFκB (Kunsch, 1992), estrogen receptor (Medici et al., 1999); p53 (Funk et al., 1992), myogenin (Wright et al., 1991; Funk and Wright, 1992), and CTF/NFI (Roulet et al., 2002). Binding sites have been identified not only for purified proteins, but also for protein complexes (Wright et al, 1991; Funk and Wright, 1992). For example, selections against purified myogenin and myogenin in nuclear extracts each yielded the same sequence families. The selected protein binding sites corresponded to natural sequences known to bind to the myogenin homodimer.

The first double-stranded DNA selection was performed *in vivo* in 1986, prior to the widespread use of PCR (Horwitz and Loeb, 1986). A nineteen base pair random region was inserted at the -35 promoter region of the tetracycline resistance gene (tet^r) on a pBR322 plasmid. *Escherichia coli* was transformed with the plasmid to produce a library of about 1000 bacteria. The bacteria were then plated on Luria-Bertani (LB) media containing tetracycline. If RNA polymerase could recognize the sequence in the randomized region, the gene for tetracycline resistance would be transcribed and the

bacterium would survive. Conversly, if RNA polymerase could not recognize the randomized region as a promoter, tet^r would not be transcribed and the bacterium would die. E. coli which survived and formed colonies had the randomized region of their plasmid sequenced. Those sequences shared some homology with the -35 promoter region they had replaced.

To facilitate high-throughput binding site identification, we have attempted to automate the selection of transcription factor binding sites. The NFκB p50 homodimer was chosen as an initial target for the development of automated selection methods. This transcription factor is well-known to bind double-stranded DNA (Muller et al., 1995; Ghosh et al., 1998; Ghosh and Karin, 2002), and has previously been a target for manual selection experiments (Kunsch et al., 1992). In addition, a Tecan Genesis workstation was chosen for this project; the flexibility of the Tecan allowed the implementation of high-stringency panning and the separation of high-affinity binding sites from non-specific binding sequences.

MATERIALS AND METHODS

Liquid Handling Robot

A Tecan Genesis workstation 200 was used as the platform for the automation of double-stranded DNA binding site selections. This robot has two pods, a liquid handling (LiHa) pod and a robotic manipulator (RoMa) arm. The LiHa is composed of eight, independently-controlled pipetting tips that have liquid sensing capabilities and can accurately pipette between 0.5 and 1000uL. The RoMa arm can reach off of the worksurface, which was essential for integrating the auxiliary equipment necessary for automated selection.

The Tecan Genesis worksurface holds a number of items (all from Tecan, unless otherwise indicated), including a twelve position microplate carrier (MP-12), a solid-phase extraction unit (SPE) with an adapter for Qiagen (Valencia, CA) kits, a two position orbital shaker, a 4°C cooled microplate carrier with a recirculating temperature bath (Julabo, Allentown, PA), a -20°C cooled microplate carrier (Mecour, Groveland, MA) with a recirculating temperature bath (Neslab, Waltham, MA), disposable tips (DiTi's), and reservoirs for buffers, reagents, and other solutions. Items off the worksurface but accessible by the RoMa arm included a thermal cycler (MJ Research, Waltham, MA), Tecan microplate hotels, and a Tecan 16-channel Columbus plate washer. The plate washer was essential to the success of the selection, and its operation is described here in some detail. The Columbus washes two columns of eight wells in parallel, 16 total wells. Two needles are inserted into each well: an aspiration needle and a dispense needle. A defined volume of liquid flows out of one of the four solution reservoirs and into the microplate well through the dispense needle. The liquid remains in the well for a defined time and is then removed by the aspiration needle. All liquid is finally deposited in a solution waste reservoir.

Tecan null modem cable for integration of MJ thermal cycler.

In order to integrate the MJ Research thermal cycler to the Tecan workstation, a null modem cable was required. The cable pinouts are diagramed in Figure 9.

A) Standard	B) Null Modem	C) Tecan
DB9 ---- DB9	DB9 ---- DB9	DB9 ---- DB9
1 -------- 1		1 -------- 7
2 -------- 2	2 -------- 3	2 -------- 3
3 -------- 3	3 -------- 2	3 -------- 2
4 -------- 4	4 -------- 1,6	4 ------- 6,8
5 -------- 5	5 -------- 5	5 -------- 5
6 -------- 6	1,6 ------ 4	6,8 ------ 4
7 -------- 7	7 -------- 8	7 -------- 1
8 -------- 8	8 -------- 7	9 -------- 9

D)

1) Carrier Detect
2) Receive Data
3) Transmit Data
4) Data Terminal Ready
5) System Ground
6) Data Set Ready
7) Request to Send
8) Clear to Send

Figure 9: Pinout configuration for 9-pin RS-232 cables. A) Standard cable, B) Standard null modem cable, C) Null modem cable used to connect the Tecan

Genesis workstation with the MJ Research thermal cycler, D) Pin designations.

Oligonucleotides

The N30 pool contains 30 random nucleotides between a 5' constant region (5' GATAATACGACTCACTATAGGGAATGGATCCACATCTACGAATTC) and a 3' constant region (5' TTCACTGCAGACTTGACGAAGCTT; Bell et al., 1998). Following amplification, the double-stranded N30 pool (10^{13} molecules) was used in the first round of selection. A positive control for the double-stranded DNA selection was constructed by inserting a NFκB p50 homodimer-binding sequence (5' TGACTGATTGGGGGATTCCCGAAGCTTATC; Kunsch et al, 1992) between the two constant regions.

Target Plate Preparation

Target plates were prepared by hydrophobic immobilization of NFκB p50 homodimer protein (0.3μg per well; Sigma, St. Louis, MO) in wells in TopYield microtitre plates (Nunc, Rochester, NY). The NFκB was dissolved in 100uL of 1X selection buffer (20mM HEPES, pH 7.9, 100mM KCl, 0.2mM EDTA, 5mM DTT). The solution was added to wells, the wells were sealed, and the plates were incubated without agitation at 4°C for approximately 18 hours. Following incubation, the solution was removed and the wells were washed with a casein blocking solution (Pierce, Rockford, IL). Remaining hydrophobic sites were blocked by incubation with casein solution at 4°C for greater than three hours. Plates used for negative selections were prepared in an identical manner except that NFκB was not added to the selection buffer.

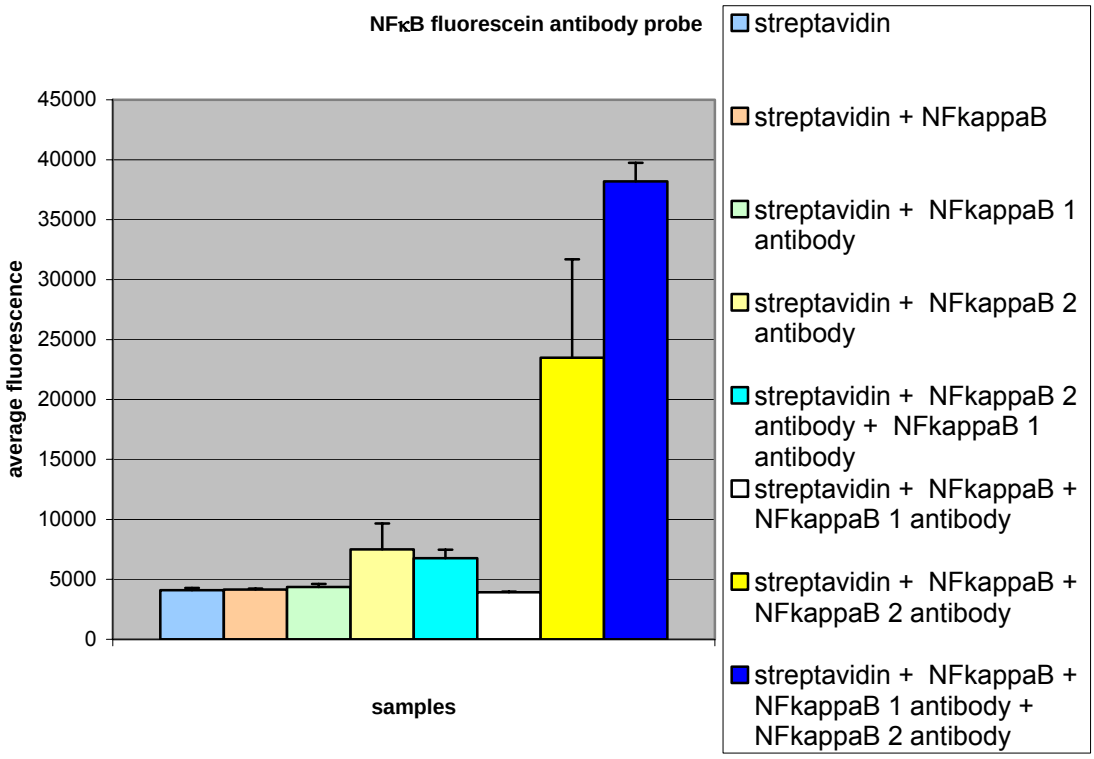


Figure 10: NFκB bound to TopYield plates and probed with antibodies.

Automated Selection

The selection process is diagramed in Figure 11 and the details of the selection cycles are provided in Table 1. The negative selection plate and the target plate were placed on the MP-12 microplate carrier on the Tecan work surface. The casein blocking solution was removed from the negative plate and the plate was rinsed with 175µL selection buffer. The Round 0 double-stranded DNA pool (100µL; 1.5µg; 10^{13} molecules was spiked with 10^{9} molecules NFκB p50 homodimer-binding sequence positive control. The spiked pool was transferred from the 4°C cooled microplate carrier to the negative selection plate. The negative selection plate was transferred to the orbital shaker where it

underwent four cycles of alternating incubations (three minutes at 500 rpm and then five minutes stationary), and was then moved back to the MP-12. The casein blocking solution was removed from the target plate which was transferred to the Columbus plate washer and sequentially washed with 1.5mL selection buffer and 300µL dH_2O. The spiked pool in the negative selection plate was transferred to the target plate, which was in turn transferred to the orbital shaker. After one to four cycles of alternating incubations (as described above), the target plate was transferred to the Columbus plate washer. The microtitre plate wells were washed with seven or eleven wash cycles (10.5mL or 16.5mL total) of selection buffer, then 300µL dH_2O. The target plate was moved back to the MP-12, and PCR master mix (100µL; 10mM Tris, pH 8.4; 50mM KCl; 2.5mM $MgCl_2$; 0.2mM dNTPs; 0.4µM each of the 41.30 5' primer and the 24.30 3' primer) and 5U Taq polymerase were added. The target plate was transferred to the thermal cycler and 15 or 20 cycles of PCR amplification (denaturation for 10 min. at 90°C, then cycled for 90 sec. at 90°C, 30 sec. at 60°C, and 90 sec. at 72°C; final extension for 3 min at 72°C) were carried out. During the thermal cycling procedure, the Columbus probes were cleaned with 6mL of a 7M urea solution followed by 6mL of dH_2O. Following DNA amplification, the plate was returned to the MP-12 and 15µL 3M sodium acetate (pH 5.2) was added to the well to lower the pH of the solution to pH 6-7. The PCR solution was then added to 345µL Qiagen Buffer PM in a 2mL deepwell plate on the MP-12 worksurface. The contents were mixed and transferred to the Qiagen filter plate on the SPE. A 500 mbar vacuum was applied for 5 minutes to pull the solution through the filter. Then 900µL of Qiagen Buffer PE were added, followed again by application of a vacuum. The final wash was an addition of 900µL of Buffer PE. Following filtration the filter was dried (as required by the protocol), and 120µL of

selection buffer was added to the well. For the collection of the DNA eluate, the RoMa arm transferred the SPE block to the second position on the manifold, and a vacuum of 500mbar was applied for 5 minutes. The purified PCR product was ultimately eluted into a Qiagen deepwell plate, and then the SPE block was transferred back to the first position on the manifold by the RoMa arm. The final 100μL of the DNA was then transferred from the SPE to a negative selection microtitre plate to begin the next round of selection and amplification. Following the negative selection, the pool was transferred to a new well coated with the target.

Table 1: Selection conditions and stringency. In order to modulate the stringency of the selection through successive rounds, four different conditions were varied: length of incubation time for negative and positive selections, wash volumes, and the number of PCR cycles. Each of these variables is described in greater detail in Materials and Methods.

	Negative Selection	Positive Selection	Selection Buffer Wash	# PCR cycles
Round 1	4 shaking/stationary incubation cycles	4 shaking/stationary incubation cycles	10.5mL	20
Round 2	4 shaking/stationary incubation cycles	4 shaking/stationary incubation cycles	10.5mL	20
Round 3	4 shaking/stationary incubation cycles + 1 hour stationary incubation	3 shaking/stationary incubation cycles	10.5mL	20
Round 4	4 shaking/stationary incubation cycles + 1 hour stationary incubation	2 shaking/stationary incubation cycles	10.5mL	20
Round 5	4 shaking/stationary incubation cycles + 24 hours stationary incubation	1 shaking/stationary incubation cycles	10.5mL	15
Round 6	4 shaking/stationary incubation cycles + 24 hours stationary incubation	1 shaking/stationary incubation cycles	16.5mL	15

Sequencing

The double-stranded DNA pools from Rounds 0, 3, and 6 were cloned into TOPO TA vectors (Invitrogen, Carlsbad, CA) and transformed into Top 10 (Invitrogen) competent cells. Following transformation, cells were plated on Luria-Bertoni media (LB) plates supplemented with 50 μg/mL kanamycin and 1600 μg X-gal per plate. The

plates were incubated at 37°C until small colonies were visible. White colonies were picked and used to inoculate 1.5mL cultures of LB containing 50 μg/mL ampicillin in a 2mL 96-well deepwell plate (Corning, Acton, MA). The antibiotics used for growth were changed between plates and media in order to ensure that the transformants were derived from the original TOPO TA vector. Cell cultures were grown overnight at 37°C with shaking, and 2μL of cells were used directly as templates for PCR reactions. The 2μL of cells were boiled at 100°C in 78μL dH_2O for 10 minutes, then 19μL of PCR master mix (final concentrations 10mM Tris, pH 8.4; 50mM KCl; 2.5mM $MgCl_2$; 0.2mM dNTPs; 0.4μM each of the M13(-40)F and M13R primers) and 1μL (5U) of Taq polymerase were added. Following fifteen thermal cycles (denaturation for 3 min. at 95°C, then cycled 45 sec. at 95°C, 30 sec. at 45°C, and 90 sec. at 72°C; final extension for 3 min. at 72°C), PCR products were purified with a Millipore (Billerica, MA) PCR clean-up kit and sequenced with Big Dye v3.0 mix (ABI, Foster City, CA) (Harkey, 2003). Sequencing reactions were analyzed on an ABI 3700 automated sequencer.

RESULTS AND DISCUSSION

The process of *in vitro* selection was automated by converting molecular biology steps that were normally carried out at the bench to steps that could be carried out by an automated workstation (Figure 1). In order to carry out selection experiments a PCR machine, orbital shaker, solid phase extraction device (SPE), plate washer, and microplate carriers that maintained reagents at 4°C and -20°C had to be introduced on or adjacent to the worksurface.

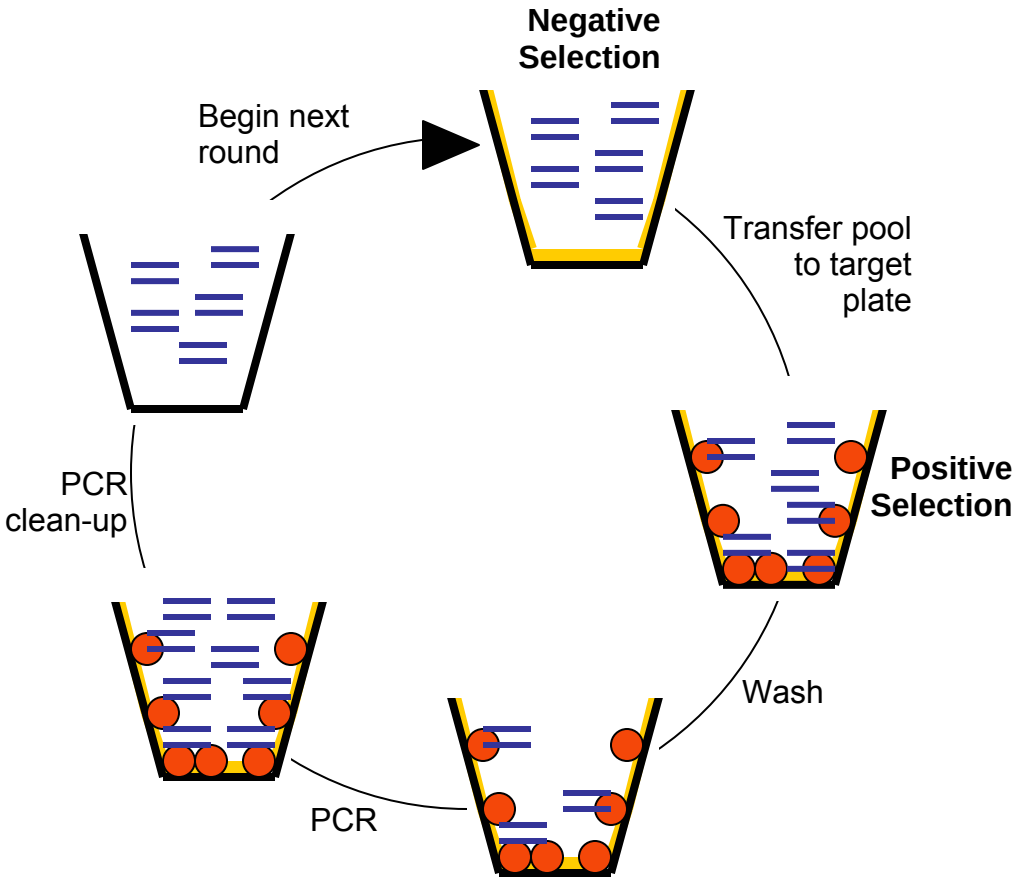


Figure 11: Automated panning protocol for *in vitro*, double-stranded DNA selections.

This figure is a simple schematic of the protocol described in Materials and Methods. In short, first a negative selection of a double-stranded DNA library (dual lines) is carried out against a microtitre plate well containing only a casein block (yellow). Those DNA molecules that do not stick to the block or to the plate are then transferred to a different microtitre plate well containing target protein (orange circles). Loosely or non-specifically bound DNA species are washed away. PCR mix is added directly to the well and any remaining DNA molecules are amplified. The amplified products are purified and then added to a new negative selection well to begin the next cycle of selection and amplification.

As a test of the automated system, a double-stranded DNA selection was initiated against the transcription factor NFκB p50 homodimer. Binding sites for this transcription factor had previously been identified by both the examination of promoter sequences and by manual selection experiments (Kunsch et al, 1992). In order to determine whether the automated selection could potentially yield a NFκB p50 homodimer binding site a previously selected high-affinity site (5' TGACTGATTGGGGGATTCCCGAAGCTTATC) was doped into a double-stranded DNA pool that contained a similar sized random sequence region (N30). The high-affinity site was included at a molar proportion of 1 to 10,000.

As with most molecular biology protocols, the primary steps in the automated selection protocol involved liquid handling. Initially, the DNA pool was moved from the 4°C microplate carrier to the MP-12 microplate carrier. Following incubation in the microtitre plates coated with NFκB p50 homodimer, unbound aptamers were removed via a panning protocol. The advantage of using panning relative to other selection methods, such as filtration, is that the majority of the selection process (binding, washing, and PCR) could occur in the same well, reducing the number of liquid handling steps and manipulations, and decreasing the possibility that rare binding sequences might be lost. Additionally, the amount of protein used to coat the microplate well in each round (0.3μg for the NFκB p50 homodimer) is less than the amount that would be used in a round of a typical manual selection protocol (4.5μg). One potential disadvantage of using a panning protocol is that nucleic acid sequences might be selected that would bind to the hydrophobic surface of the plates, rather than to the immobilized target; for example, nucleic acids that bind to hydrophobic nitrocellulose filters frequently arise during filtration selections (Tuerk et al., 1992). To reduce the hydrophobic surface area that

nucleic acids would be exposed to, the wells in the Top Yield plates were blocked following the immobilization of the NFκB target. A wide variety of blocking agents were tested for their ability to reduce background binding, and Pierce Casein block was ultimately selected. To help prevent the selection of matrix-binding sequences, a negative selection was first carried out using blocked microtitre plates that did not contain NFκB. Any nucleic acids that bound via non-specific hydrophobic interactions should have been lost from the selection at an early round. Such non-specific interactions would have been much more likely in the course of a single-stranded DNA selection, as the hydrophobicity of single-stranded DNA is much greater than that of double-stranded DNA.

Following the negative selection, the DNA solution was transferred to wells containing NFκB and thoroughly mixed via an orbital shaker. Stringency was varied by increasing the time allowed for plate-binding during the negative selection and decreasing the time allowed for target-binding during the positive selection (Table 1). Non-binding DNA species in solution were removed from the binding species immobilized on the surface of the plate. In manual selections this is one of the most critical but also one of the most tedious steps. In our panning protocol, a plate washer was used to rapidly wash the wells with 10.5mL to 16.5mL of buffer. In contrast, wash steps for bead- or filter-based selections typically rely upon repeatedly washing immobilized complexes with buffer aliquots of around 300μL. This slows the overall procedure and frequently results in researchers carrying out far less stringent selections than would otherwise be possible. While panning with the plate washer proved to be extremely efficient, one problem that was initially encountered was that pool DNA could stick to the aspiration needles on the plate washer, leading to cross-contamination

between wells washed by the same needles. This problem was eliminated by washing the needles with 7M urea after each use.

In order to amplify bound sequences, PCR reagents were added directly to the microtitre wells after the wash step. The PCR master mix was stored at 4°C and the Taq polymerase was stored at -20°C on the surface of the robot in cooled carriers connected to recirculating temperature baths. This allowed the automated selection to run essentially autonomously without the need for the addition of reagents at each step. Following reagent addition, the RoMa arm transfers the microtitre plate to the integrated thermal cycler, where the lid closes and a pre-set amplification program runs. The template DNA was conveniently eluted off of the target protein during the initial denaturation step of the PCR program. The number of thermal cycles required for amplification was determined by separating PCR products on agarose gels. After an initial optimization, it was determined that 15 to 20 thermal cycles would generally yield PCR products that could be carried forward into the next round. Following cycling, the plate was held at 4°C for 30 minutes to reduce aerosol formation, and then the lid of the PCR machine opens and the plate is transferred by the RoMa arm to the worksurface for the PCR product purification step.

While automation of selection procedures helps to ensure reproducibility and increases throughput, consistency and attention to fine detail are essential for successful method development. As an example, one problem that initially arose was the cross-contamination of PCR products between different wells or cycles during or following amplification. In order to successfully integrate the thermal cycler and eliminate this problem, four separate optimizations were required. First, an MJ Research Microseal 'P' adhesive-backed sealing pad was manually placed on the inside of the thermal cycler's

motorized Power Bonnet lid prior to the selection, and the height of the lid was adjusted so that seals were formed around the top of each well. If the height was not adjusted properly, the PCR product was found to evaporate out of the well. Secondly, a heated lid was used to keep the PCR product from condensing on the lid. Third, an MJ Research Microseal 96 Plate Lifter was modified and placed in the thermal cycler so as to slightly lift the plate out of the unit when the lid of the thermal cycler opened. Without the Plate Lifter the lid pressed down on the plate so hard that it proved impossible for the RoMa arm to transfer the plate back to the Tecan work surface. The lid and Plate Lifter were adjusted until the plate was consistently available to the RoMa arm. Fourth, a slight vacuum was sometimes created in the wells during the heating and cooling process of the PCR program, causing the plate to stick to the lid of the thermal cycler when it opened. A final incubation at 4°C for 30 minutes not only minimized aerosol formation but also helped eliminate vacuum formation. All four of these optimizations were intertwined; for example, setting the Plate Lifter too high led to a greater probability of the plate sticking to the lid of the thermal cycler. These various improvements had to be iteratively implemented and tested in order to ensure that the final program would operate smoothly. As a standard of performance, if liquid was found to accumulate anywhere except in the microplate wells during a selection, the selection was terminated and the program was further modified.

Once amplification was completed and the microtitre plate transferred to the Tecan worksurface, the PCR products were purified with a Qiagen kit. In order to automate this process, the two position SPE unit was equipped with an adapter that fit the Qiagen filter plate. The pure product was eluted into selection buffer, and was ready to initiate the next round of selection. Automation of the selection process provides a

significant increase in throughput. While roughly the same amount of time is spent on reagent preparation and sequencing for manual and automated selections, there is a large difference in the time required for the selection process itself. Manual selections employ time-consuming techniques such as ethanol precipitation and purification via gel electrophoresis. While these precautions can help to avoid the accumulation of amplification artefacts, they also typically extend a manual round of selection to several days, as opposed to the few hours required for an automated round. The total time savings over 6 to 18 cycles of selection (the number of rounds typically required for the purification of binding species) is therefore considerable, the difference between days and weeks.

Six rounds of automated selection were carried out, with the only human intervention being addition of PCR reagents to the target plate (although it should also be possible to automate this step, as well). The double-stranded DNA pool from Rounds 0, 3, and 6 were cloned and sequenced. The number of recognizable NFκB binding sites progressively increased (Figure 2). The results from the automated selection experiment were similar to those obtained from manual selection experiments. Rosen and co-workers identified a consensus, 10 base-pair (bp) NFκB p50 homodimer binding sequence (5' GGGGATYCCC; Kunsch et al., 1992). While the selected Rounds 0 and 3 did not contain the consensus NFκB p50 binding sequence, it was present in 10% of the sequences by Round 6 (Figure 2(A)). Gorenstein and coworkers have suggested a more general consensus binding sequence (Figure 2(B)) (5' GGGRNNYYCC; King et al., 2002). By Round 6 recognizable NFκB binding sites were present in 93% of the sequenced clones (Figure 2(F)).

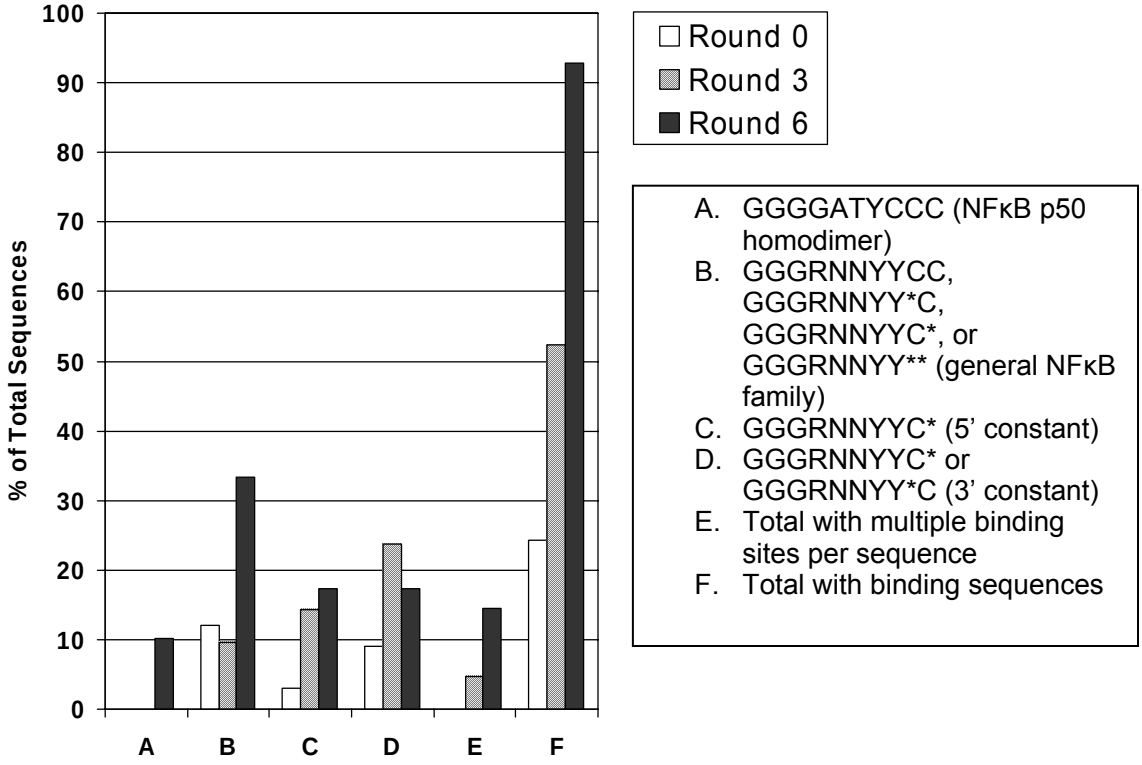


Figure 12: Frequencies of selected binding sequences.

A variety of different NFκB binding sequences were known prior to the start of the selection, and were recovered by the automated selection procedure. (A) represents the frequency of a core NFκB p50 homodimer binding site. (B) represents the frequency of a more general representation of the NFκB family binding site. (C-D) represent the frequency of the general NFκB family binding sites that utilize either the 5' or 3' (respectively) constant regions from the original N30 pool. (E) represents the frequency of species which contain dimeric binding sites. The values in (E) are not included in the other tallies (A-D). (F) represents the total frequency of species with selected NFκB p50 homodimer binding sequences, both core and more general family. In each instance, the white bar represents the 33 sequences derived from Round 0, the stippled bar represents

the 21 sequences derived from Round 3, and the solid bar represents the 69 sequences derived from Round 6. Values in (A-E) sum to the value in (F).

In addition to identifying NFκB p50 homodimer binding sequences that corresponded to consensus binding sites, more subtle sequence contributions to protein recognition could also be discerned. For example, Kunsch et al. (1992) had observed that the guanosine triplet at the 5' end of the 10 bp consensus NFκB p50 homodimer binding sequence was essential for binding, while variations at the 3' cytidine doublet were tolerated. Similarly, we have found that additional sequence variations can occur in the 3' portion of this motif (5' GGGRNNYY*C, GGGRNNYYC*, and GGGRNNYY**, Figure 2(B)). Indeed, by Round 6, 62% of the binding sequences contained mutations in one or both of the 3' terminal cytidines. Kunsch et al. (1992) also observed that there was frequently an additional guanine present at the 5' end of the consensus binding sites and an additional cytosine present at the 3' end (5' gGGGGATYCCCc). Similarly, all of the selected binding sites from Round 6 that contained the 10 bp consensus NFκB p50 homodimer binding sequence also contained one or both of these additional 5' or 3' bases. Overall, 68% of the selected binding sites from Round 6 contained one or both of the additional bases; that is the core decamer binding site had apparently expanded to either a undecamer or dodecamer binding site.

The expansion of the previously determined core decamer binding site that is predicted by our selection experiments has recently been confirmed by other studies. The undecamer (5' GGGGATTCCCc) is palindromic about the central thymidine residue and is identical to the high affinity human major histocombatibility complex H-2 binding site for NFκB (Angelov et al., 2003). Crystal structures and DNA-protein crosslinking studies have shown that there are in fact specific base contacts between the NFκB p50

homodimer and all four guanosine residues at the 5' end of the decamer core binding site sequence (Angelov et al., 2003). The palindromic undecamer therefore of necessity contains a guanosine quadruplet in each strand that each monomer of the homodimer can bind to.

Interestingly, selected DNAs that contained two NFκB binding sites comprised 14% of the population by Round 6 (Figure 12(E)). The spacing between the two sites varied from zero (the core 10 bp consensus sites touched one another) to 22 nucleotides. Since selected DNAs that contained dimeric sites predominated only in the later rounds of selection, it seems likely that the presence of two sites resulted in a competitive advantage, by allowing multiple opportunities for interactions with the protein target.

The control sequence added to the pool did not take over the selection, as planned. Instead, the constant sequences in the N30 pool contributed to numerous NFκB p50 homodimer binding sites (Figure 13). Seven of the ten bases in the NFκB family general consensus sequence were present in the 5' constant region, therefore only three bases were needed to complete the sequence. The probability of this was $1 / 4^3$ or 1 in every 64 molecules. Likewise, only five to six bases were needed to complete the consensus sequence utilizing the 3' constant region. The selection of N30 binding sites was also assisted by the previously mentioned sequence flexibility allowed at the 3' end of the consensus site, and again should have resulted in 1 in every 64 molecules containing a NFκB binding site. Interestingly, by Round 6, utilization of the 5' and 3' constant regions to form NFκB binding sites was equal (20 instances each), as predicted (Figure 12(C-D)). In contrast, the control sequences that were spiked (1 in 10,000 molecules) into the pool was much less populous than the binding sequences that occurred by chance in the N30 pool, and this helps to explain why they were not found by the conclusion of

the selection. This tendency for selections, automated or manual, to utilize the most common, functional motifs has previously been observed and is known as the 'tyranny of small motifs' (Ellington, 1994). The fact that constant regions sometimes participate in binding sequences suggests that it may sometimes be desirable to compare the results of selection experiments with different pools, in order to more fully examine the range of binding sites that are possible.

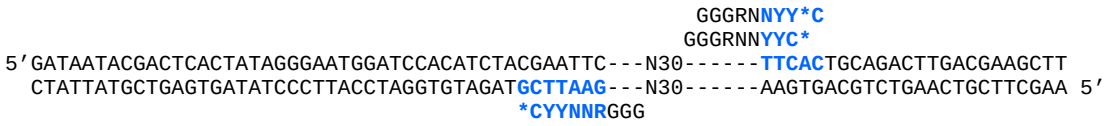


Figure 13: Utilization of constant regions in selected binding sequences.

The original N30 pool is shown as a double-stranded DNA molecule. Superimposed on the pool are the locations of potential NFκB binding sequences that utilize the 5' or 3' constant regions. These sequences correspond to the frequencies shown in Figure 12(C-D).

Nonetheless, it should be noted that the same consensus NFκB p50 homodimer binding site that was present on the control (5' GGGGATTCCC) was also recovered from the completely random region alone (Figure 12(A)). This site would have been present in the population at roughly the same frequency as the positive control (taking into account multiple possible registers), but by the conclusion of the selection it was present in 10% of the population. However, as we discussed above, many of the NFκB binding sites that were recovered from the selection likely formed even more contacts with the protein than did the previously identified core decamer binding site, and thus would have enjoyed a selective advantage relative to the control. Therefore, our results actually highlight the extraordinary potential of robotic selection experiments to overcome even the tyranny of small motifs (in this instance, both degenerate but

populous binding motifs and the positive control). That is, with additional rounds of robotic selection, the best sequences can be culled from a population, even though they may be only incrementally better than a majority sequence.

CONCLUSIONS

Double-stranded DNA aptamer selections against the NFκB p50 homodimer were successfully automated using a Tecan Genesis workstation. The consensus DNA binding site for NFκB was isolated from a pool of 10^{13} double-stranded, random sequence oligonucleotides. Although the consensus binding sequence that was originally spiked into the pool did not rise to the fore in the automated selection, the fact that an unanticipated but more likely set of binding sequences was ultimately chosen was also proof that the automated method worked well. Moreover, the fact that variations observed between individual selected sequences could be largely explained based on the binding propensities of known NFκB p50 homodimer binding sites indicated that the automated selection method should be capable of fully describing binding sites for other transcription factors. A significant increase in throughput was achieved, from several days for a round of manual selection to four hours for a round of automated selection.

REFERENCES

Angelov D.; Charra M.; Muller C.W.; Cadet J.; Dimitrov S. (2003). Solution study of the NF-kappaB p50-DNA complex by UV laser protein-DNA cross-linking. Photochem Photobiol. 77:592-6.

Bell, S.D.; Denu, J.M.; Dixon, J.E.; Ellington, A.D. (1998). RNA molecules that bind to and inhibit the active site of a tyrosine phosphatase. J Biol Chem 273:14309-14314.

Benbrook D.M.; Jones N.C. (1994). Different binding specificities and transactivation of variant CRE's by CREB complexes. Nucleic Acids Res 22:1463-9.

Blackwell T.K.; Weintraub H. (1990). Differences and similarities in DNA-binding preferences of MyoD and E2A protein complexes revealed by binding site selection. Science 250:1104-10.

Boardman P.E.; Oliver S.G.; Hubbard S.J. (2003) SiteSeer: Visualisation and analysis of transcription factor binding sites in nucleotide sequences. Nucleic Acids Res 31: 3572-5.

Bulyk M.L.; Huang X.; Choo Y.; Church G.M. (2001). Exploring the DNA-binding specificities of zinc fingers with DNA microarrays. Proc Natl Acad Sci U S A 98:7158-63.

Delany I.; Spohn G.; Rappuoli R.; Scarlato V. (2002). *In vitro* selection of high affinity HspR-binding sites within the genome of Helicobacter pylori. Gene 283:63-9.

Ellington A.D. (1994). Empirical explorations of sequence space: host-guest chemistry in the RNA world. Ber Bunsenges Phys Chem 98:1115–21.

Funk W.D.; Pak D.T.; Karas R.H.; Wright W.E.; Shay J.W. (1992). A Transcriptionally Active DNA binding Site for Human p53 Protein Complexes. Mol Cell Biol 12:2866-2871.

Funk W.D.; Wright W.E. Cyclic amplification and selection of targets for multicomponent complexes: Myogenin interacts with factors recognizing binding sites for basic helix-turn-helix, nuclear factor 1, myocyte-specific enhancer-binding factor 2, and COMP1 factor. (1992). Proc Natl Acad Sci USA 89:9484-9488.

Ghosh S.; May M.J.; Kopp E.B. (1998). Evolutionarily Conserved Mediators of Immune Responses. Annu Rev Immunol 16:225-60.

Ghosh S.; Karin M. (2002). Missing Pieces in the NF-kB Puzzle. Cell 109:S81-S96.

Harkey C. Sample prep for automated sequencing. http://www.icmb.utexas.edu/core/DNAFacility/Sequencing%20Handout.pdf (accessed Feb 2003).

Horwitz M.S.; Loeb L.A. (1986). Promoters selected from random DNA sequences. Proc Natl Acad Sci U S A 83:7405-9.

Huang J.; Blackwell T.K.; Kedes L.; Weintraub H. (1996). Differences between MyoD DNA binding and activation site requirements revealed by functional random sequence selection. Mol Cell Biol 16:3893-900.

King D.J.; Bassett S.E.; Li X.; Fennewald S.A.; Herzog N.K.; Luxon B.A.; Shope R.; Gorenstein D.G. (2002). Combinatorial Selection and Binding of Phosphorothioate Aptamers Targeting Human NF-κB RelA(p65) and p50. Biochemistry 41:9696-9706.

Kunsch C.; Ruben S.M.; Rosen C.A. (1992). Selection of Optimal kB/Rel DNA binding Motifs: Interaction of Both Subunits of NF-kB with DNA Is Required for Transcriptional Activation. Mol Cell Biol 12:4412-4421.

Liu R.; Blackwell T.W.; States D.J. (2001). Conformational model for binding site recognition by the E.coli MetJ transcription factor. Bioinformatics 17:622-33.

Medici N.; Abbondanza C.; Nigro V.; Rossi V.; Piluso G.; Belsito A.; Gallo L.; Roscigno A.; Bontempo P.; Puca A.A.; Molinari A.M.; Moncharmont B.; Puca G.A. (1999). Identification of a DNA Binding Protein Cooperating with Estrogen Receptor as RIZ (Retinoblastoma Interacting Zinc Finger Protein). Biochem Biophys Res Commun 264:983-989.

Muller C.W.; Rey F.A.; Sodeoka M.; Verdine G.L.; Harrison S.C. (1995). Structure of the NF-kB p50 homodimer bound to DNA. Nature 373:311-317.

Newman J.R.; Keating A.E. (2003). Comprehensive identification of human bZIP interactions with coiled-coil arrays. Science 300:2097-101.

Pollock R.; Treisman R. (1990). A sensitive method for the determination of protein-DNA binding specificities. Nucleic Acids Res. 18:6197-204.

Roulet E.; Busso S.; Camargo A.A.; Simpson A.J.; Mermod N.; Bucher P. (2002). High-throughput SELEX SAGE method for quantitative modeling of transcription-factor binding sites. Nat Biotechnol 20:831-5.

Sinha S.; Tompa M. (2002). Discovery of novel transcription factor binding sites by statistical overrepresentation. Nucleic Acids Res 30:5549-60.

Sinha S.; Tompa M. (2003). YMF: A program for discovery of novel transcription factor binding sites by statistical overrepresentation. Nucleic Acids Res 31:3586-8.

Sooter L.J.; Ellington A.D. (2004). Automated selection of transcription factor binding sites. JALA 9:277-284.

Thiesen H.J.; Bach C. (1990). Target Detection Assay (TDA): a versatile procedure to determine DNA binding sites as demonstrated on SP1 protein. Nucleic Acids Res 18:3203-9.

Tuerk C.; Gold L. (1990). Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. Science 249:505-10.

Tuerk C.; MacDougal S.; Gold L. (1992). RNA pseudoknots that inhibit human immunodeficiency virus type 1 reverse transcriptase. Proc Natl Acad Sci USA 89:6988-92.

Wright W.E.; Binder M.; Funk W. (1991). Cyclic Amplification and Selection of Targets (CASTing) for the Myogenin Concensus Binding Site. Mol Cell Biol 11:4104-4110.

Wright W.E.; Funk W.D. (1993). CASTing for multicomponent DNA-binding complexes. Trends Biochem Sci 18:77-80.

Chapter 3: Cell Lysate Selections

INTRODUCTION

As new genomes are identified, methods are necessary for the high-throughput identification of transcription factor binding sites and other regulatory sequences. Few lysate selections are present in the literature. Although cellular lysates are a popular blocking agent, the lysate itself has not been a popular target. One example of a lysate selection was a phage display antibody selection against the lysate of hair cells found in the sensory receptors of a bull frog's inner ear (Cyr and Hudspeth, 2000). The resultant antibodies recognized a single protein that was found to be expressed in the inner ear. The general lack of enthusiasm for lysate selections is understandable considering the wide possibility for binding targets within the cell extract. Furthermore it would be difficult to determine what had been selected for at the conclusion of the experiment. By performing double-stranded DNA aptamer selections against cellular lysates, the selection is biased toward dsDNA binding proteins and other positively charged species within the lysate. With the availability of modern computing power, it is possible to do sequence searches for known binding species. These two factors allowed us to perform selections against two different cellular lysates and then search the selection products for known cellular protein DNA binding sites.

For double stranded DNA aptamer background, refer to the introduction of chapter 2.

MATERIALS AND METHODS

Liquid Handling Robot

A Tecan Genesis workstation 200 was used to automated the selection process against cellular lysates. This robot has both a liquid handling pod and a pod containing a robotic manipulator arm. The liquid handling pod contains eight independently controlled pipette tips which are capable of pipetting between 0.5 and 1000uL. The tips are also capable of fine liquid sensing. The RoMa arm can reach off of the worksurface and can apply 7.5 newtons of force to carry an object. These capabilities were essential for integrating additional equipment to the Tecan workstation. By reaching off the worksurface, a tremendous amount of space became available for integration of additional equipment.

The Tecan Genesis worksurface holds a variety of equipment (all from Tecan, unless otherwise indicated), including a twelve position microplate carrier, a solid-phase extraction unit with an adapter for Qiagen (Valencia, CA) kits, a two position orbital shaker, a 4°C cooled microplate carrier with a recirculating temperature bath which holds 1.5mL tubes (Julabo, Allentown, PA), a -20°C cooled microplate carrier (Mecour, Groveland, MA) with a recirculating temperature bath (Neslab, Waltham, MA), disposable tips, and reservoirs for other reagents. Items off the worksurface but accessible by the RoMa arm include a thermal cycler (MJ Research, Waltham, MA) and a Tecan 16-channel Columbus plate washer.

Oligonucleotides

The LS.N65 pool contains 56 random nucleotides between a 5' constant region (5' GATAATACGACTCACTATAGCTTA) and a 3' constant region (5' ACGTCTCGTCAAGTCTGCAATGTA). Following amplification, the double-stranded N56 pool was purified using a Qiagen PCR clean up kit (Qiagen, Valencia, CA) and 10^{13} molecules were used in the first round of selection.

Target Plate Preparation

Cell lysate was prepared from two different cell lines. The first was BL21(DE3) cells from Novagen (San Diego, CA) containing pACYC(CAM), a plasmid which overexpressed LacI (Novagen, San Diego, CA). The second was BL21(DE3) cells from Novagen containing pASK-IBA3, a plasmid which overespressed TetR (IBA, St.Louis, MO). One liter cultures containing chloramphenicol were grown for each cell line. The cultures were grown to saturation and spun down at 4000g for 15 minutes. The pellets were washed with 5mL of selection buffer and spun down again. Then the pellets were transferred to 50mL conicals where 5mL of selection buffer and 800uL of Roche Complete protease inhibitor cocktail (Roche, Indianapolis, IN) were added. Each pellet was resuspended and sonicated on ice for ten minutes with alternating 10 second intervals of sonication at 13% and resting. Following sonication, the tubes were spun at 10,000g. The supernatant was removed and used to prepare the target plates.

Target plates were prepared by hydrophobic immobilization of cell lysate in wells of TopYield microtitre plates (Nunc, Rochester, NY). 100uL of lysate and 200uL of selection buffer were added to each well. The wells were sealed with a microplate seal, and the plates were incubated without agitation at 4°C for approximately 18 hours.

Automated Selection

The selection process is diagramed in **Figure 14** and the details of the selection cycles are provided in **Table 2**. The target plate was placed on the twelve position microplate carrier on the Tecan work surface. The lysate solution was removed from the plate and it was rinsed with 175µL selection buffer. The Round 0 double-stranded DNA pool (100µL; 1.5µg; 10^{13}) was transferred from the 4°C cooled microplate carrier to the selection plate. The selection plate was transferred to the orbital shaker where it underwent varied cycles of alternating incubations (three minutes at 500 rpm and then five minutes stationary), and was then moved to the Columbus plate washer. The microtitre plate wells were washed with varied amounts of selection buffer then dH_2O. The target plate was moved back to the twelve position microplate carrier, and PCR master mix (100µL; 10mM Tris, pH 8.4; 50mM KCl; 2.5mM $MgCl_2$; 0.2mM dNTPs; 0.4µM each of the 5.24.N56 5' primer and the 3.24.N56 3' primer) and 5U Taq polymerase were added. The target plate was transferred to the thermal cycler and a varied number of cycles of PCR amplification (denaturation for 10 min. at 90°C, then cycled for 90 sec. at 90°C, 30 sec. at 49°C, and 90 sec. at 72°C; final extension for 3 min at 72°C) were carried out. During the thermal cycling procedure, the Columbus probes were cleaned with 6mL of a 7M urea solution followed by 6mL of dH_2O. Following DNA amplification, the program was paused and a 4% agarose gel was run to make sure that an appropriate number of PCR cycles had been performed. Additional reagents for the next round were also added at this time. Following the agarose gel, the plate was returned to the twelve microplate carrier on the worksurface and 15µL 3M sodium acetate (pH 5.2) was added to the well to lower the pH of the solution to pH 6-7. The

PCR solution was then added to 345μL Qiagen Buffer PM in a 2mL deepwell plate on the MP-12 worksurface. The two were mixed and transferred to the Qiagen filter plate on the solid phase extraction device. A 500 mbar vacuum was applied for 5 minutes to pull the solution through the filter on the Qiagen plate. Then 900μL of Qiagen Buffer PE were added, followed again by application of a vacuum. The final wash was an addition of 900μL of Buffer PE from Qiagen. Following filtration the filter was dried (as required by the protocol), and 120μL of selection buffer was added to the well. For the collection of the DNA eluate, the robotic manipulator arm transferred the block of the solid phase extraction device to the second position on the manifold, and a vacuum of 500mbar was applied for 5 minutes. The purified PCR product was finally eluted into a Qiagen deepwell plate, and then the block was transferred back to the first position on the manifold by the robotic manipulator arm. The final 100μL of the DNA was then transferred from the solid phase extraction position to a freshly washed, lysate coated microtitre plate to begin the next round of selection and amplification.

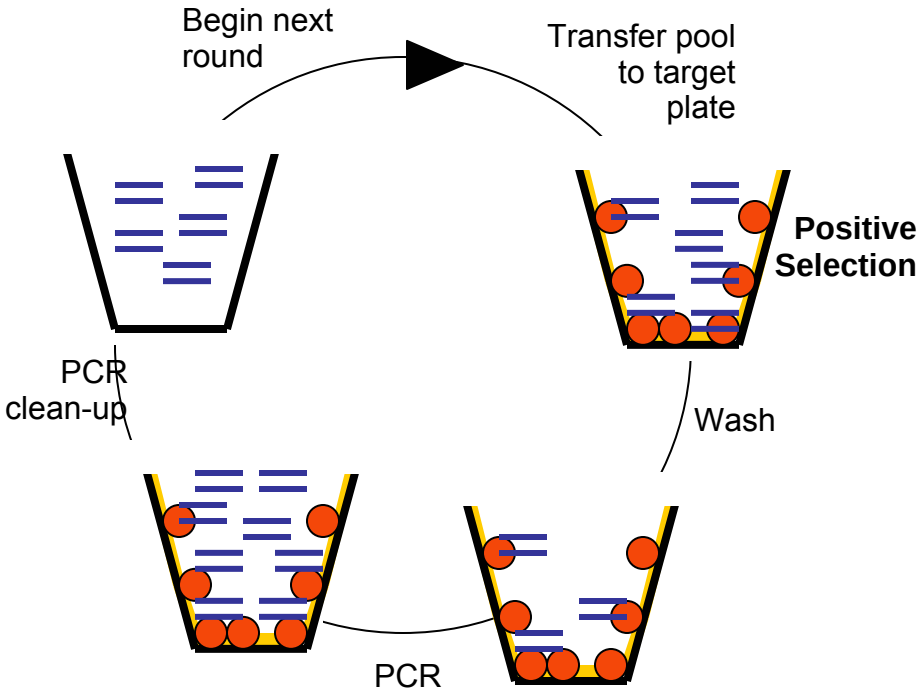


FIGURE 14: Automated panning protocol for *in vitro*, double-stranded DNA selections. This figure is a simple schematic of the protocol described in Materials and Methods. In short, a double-stranded DNA library (dual lines) is transferred to a microtitre plate well containing target lysate (grey circles). Loosely or non-specifically bound DNA species are washed away. PCR mix is added directly to the well and any remaining DNA molecules are amplified. The amplified products are purified and then added to a new selection well to begin the next cycle of selection and amplification.

TABLE 2: Selection conditions and stringency. To modulate the stringency of the selection through successive rounds, three different conditions were varied: length of incubation time for the selections, wash volumes, and the number of PCR cycles. Each of these variables is described in greater detail in Materials and Methods.

	Binding incubation	Washes	PCR cycles
Round 1	4 shake/stand	2 SXN 2 H2O	10
Round 2	4 shake/stand	2 SXN 2 H2O	11
Round 3	4 shake/stand	2 SXN 2 H2O	12
Round 4	2 shake/stand	2 SXN 2 H2O	10*
Round 5	1 shake/stand	4 SXN 2 H2O	10
Round 6	1 shake/stand cycle with no shaking	4 SXN 2 H2O	10
Round 7	1 shake	6 SXN 2 H2O	13
Round 8	1 shake	6 SXN 2 H2O	11

*Tet- sample cycled an additional 4 times

Sequencing

A standard automated sequencing protocol was used to sequence the pools from these selections. The double-stranded DNA pools from Rounds 0 and Rounds 5 and 8 of both LacI overexpressing lysates and TetR overexpressing lysates were cloned into TOPO TA vectors (Invitrogen, Carlsbad, CA) and transformed into Top 10 (Invitrogen) competent cells. Following transformation, cells were plated on Luria-Bertoni media (LB) plates supplemented with 50 μg/mL kanamycin and 1600 μg X-gal per plate. The plates were incubated at 37°C until small colonies were visible. White colonies were picked and used to inoculate 1mL cultures of LB containing 50 μg/mL ampicillin in a 2mL 96-well deepwell plate (Corning, Acton, MA). Cell cultures were grown overnight at 37°C with shaking, and 2μL of cells were used directly as templates for 100uL PCR reactions. The PCR reaction was constructed as follows: the 2μL of cells were first boiled at 100°C in 78μL dH_2O for 10 minutes, then 19μL of PCR master mix (final concentrations 10mM Tris, pH 8.4; 50mM KCl; 2.5mM $MgCl_2$; 0.2mM dNTPs; 0.4μM each of the M13(-40)F and M13R primers) and 1μL (5U) of Taq polymerase were added. Following fifteen thermal cycles (denaturation for 3 min. at 95°C, then cycled 45 sec. at 95°C, 30 sec. at 45°C, and 90 sec. at 72°C; final extension for 3 min. at 72°C), PCR products were purified with a Millipore (Billerica, MA) PCR clean-up kit and sequenced with Big Dye v3.0 mix (ABI, Foster City, CA) (Harkey, 2003). Sequencing reactions were analyzed on an ABI 3700 automated sequencer.

Sequence Analysis

The selected sequences from each of the five pools were analyzed using a series of Perl scripts (Phil Shannon). The programs read in the sequences from each pool along

with the set of known *E. coli* transcription factor binding sequences, obtained online from Regulon DB (http://www.cifn.unam.mx/Computational_Genomics/regulondb/) (Salgado et al., 2004), or with a randomized set of sequences. The randomized sets of sequences contained the same number of sites of each length-class as the known set, only the base contents were varied. One randomized set matched the base content of the Round 0 pool (A 30.66%, C 26.06%, G 19.48%, T 23.80%), while a second randomized set contained equimolar amounts of the four bases. Each binding site was then broken into hexamer registers for both the forward and reverse strands. The script looked for each register within all of the aptamers in each pool. When a match was found, the script extracted a sequence from the aptamer which corresponded to the alignment between the transcription factor binding site and the aptamer, starting at the identified hexamer. Once the transcription factor-like sequences were extracted from the aptamers, the scripts were then able to tabulate a variety of statistics: the greatest similarity match for each binding site (or randomized site) from each pool; the number of times each pool showed better similarity to a known sequence over a randomized sequence; and a 'quality score' that was defined as the greatest percent similarity multiplied by the greatest number of matching bases in each length-class of binding site.

Another series of Perl scripts was developed to analyze the sequences using an approach that had previously been employed to identify Medline abstracts that discussed protein-protein interactions (Marcotte et al, 2001)(Phil Shannon). In adapting this technique, the set of known binding sites, again from RegulonDB, was used as the "training set" and the entire *E. coli* K-12 genome from Genbank acted as the background "dictionary." The frequency of individual hexamer "words" in the binding sites, genome, and aptamer pools was determined. Based on whether the frequency n_i for a given

hexamer i in the aptamer pool with total hexamers N more closely resembled that hexamer's frequency in the binding sites $f_{I,i}$ or the genome $f_{I,i}$, we can generate a score indicating whether the pool is more closely related to the genome in general or the binding sites in particular. By summing these scores for all the hexamers in the training set, we calculate a log likelihood score S for the whole pool:

$$S = \sum_i \left(n_i \ln \frac{f_{I,i}}{f_{N,i}} - N\left(f_{I,i} - f_{N,i}\right) \right)$$

The value of S is positive for pools enriched for the binding sites and negative for pools closely tied to the genome.

Results and Discussion

Selection of DNA sequences that bind to cellular lysates.

Initially, we chose to select double-stranded DNA molecules from a N30 pool that could bind to lysates from two different derivatives of the same parental *E. coli* strain. One strain (BL21(DE3) (Novagen, San Diego, CA)) contained the pACYC(CAM) plasmid (Novagen, San Diego, CA), which overexpressed the LacI protein, and the other strain (BL21(DE3) (Novagen, San Diego, CA)) contained the plasmid pASK-IBA3 (IBA, St. Louis, MO), which overexpressed the TetR protein.

The selection procedure itself was relatively straightforward. *E. coli* lysates were incubated in hydrophobic microplates. Varied species within the lysate bound to the microplate wall through hydrophobic interactions. The wells were then rinsed with selection buffer to remove any non-bound species. A double-stranded DNA library (N30) with 10^{13} species was incubated in the wells in 100uL of selection buffer. Species that bound to the target lysate remained behind when the plate was washed with 1.8mL to

5.6mL of selection buffer and 600uL of water. The captured double-stranded DNA binding species were then eluted in 100uL of selection buffer and amplified via the polymerase chain reaction (PCR). At this point, the robot was paused while a 4% agarose gel was run to check for the presence of product. Following confirmation that a band of the correct size had been obtained, the PCR product was purified with a Qiagen PCR clean-up kit and used for additional rounds of selection and amplification.

Analysis of selected sequences.

It seemed unlikely that selected binding sites would be perfect replicas of genomic DNA binding sites for several reasons. First, previous selection experiments had shown that double-stranded DNA molecules selected to bind individual DNA-binding proteins did not perfectly mimic genomic binding sites **(**Sooter and Ellington, 2004; Kunsch et al, 1992; Shultzaberger and Schneider, 1999). For example, Shultzaberger and Schneider found that SELEX experiments identified the binding site for a probable dimeric or trimeric form of Lrp, while *in vivo* the Lrp monomer bound an alternate site. Kunsch and co-workers identified many binding sites for p50 homodimers. These sites could generally be put into families such as GGGRNNYYCC, but the vast majority did not exactly match the strong *in vivo* binding sequence of GGGGATTCCC. Second, the selection conditions were not particularly stringent, compared to selections that targeted individual DNA-binding proteins. The use of non-stringent selection conditions meant that sequences that targeted proteins that were relatively rare or that had relatively low binding affinities would be retained during the selection. However, this also meant that many, non-canonical variants of higher affinity binding sites would also be retained in the population. Finally, given that each selected binding sequence was

exposed to a variety of targets during each round of selection, those sequences that could bind to multiple, different targets might ultimately predominate. Such chimeric binding sequences would obviously not necessarily resemble a single genomic binding site.

For these reasons, we initially decided to determine whether the selection had been successful by analyzing whether short sequences (hexamers) within the selected molecules were found in greater prevalence in the known set of *E. coli* protein-binding DNA sites**,** relative to hexamers in unselected molecules (Figure **15**). For each selected pool, there is clearly a statistically significant increase in the number of hexamers corresponding to genomic, protein-binding DNA sites, relative to the starting pool.

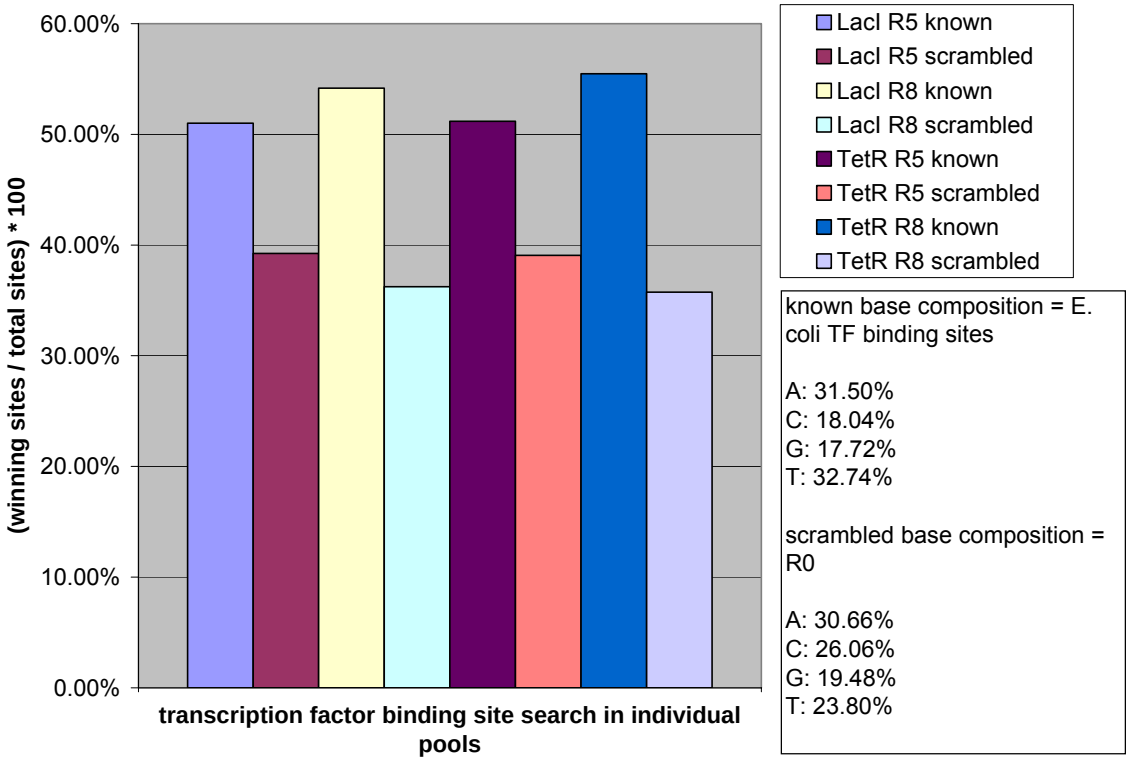


Figure 15: Transcription factor binding sites found in individual pools.

In order to determine to what extent this increase in information content in the selected pool faithfully represented the complement of genomic binding sites, we further analyzed the data relative to the sizes of genomic protein-binding DNA sites (Figure **16**) and also determined whether the selected hexamers more completely and accurately represented genomic protein-binding 'space' than did the random sequence pool (Figure **16**). As can be seen, in general selected sequences contained more protein-binding DNA site hexamers than did the original random sequence pool, over virtually every size class. While there were some differences between the selected sequences from Round 5 and Round 8, and between sequences selected from lysates that contained either LacI or TetR, these differences were much less significant than the overall differences between selected pools and random pools.

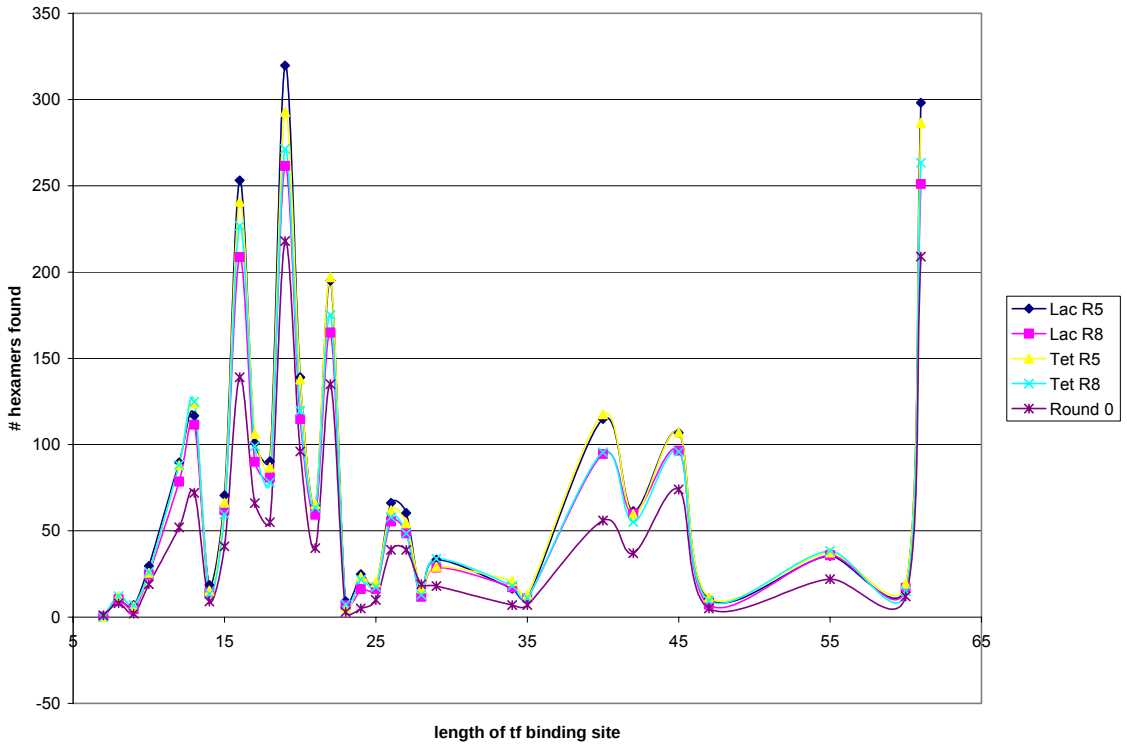


Figure 16: Number of hexamers found in transcription factor binding sites.

Finally, we needed to investigate how relevant the selected sequences were to the *E. coli* genome. Selected sequences were found to match transcription factor binding sites in the *E. coli* genome more frequently than non-selected sequences.

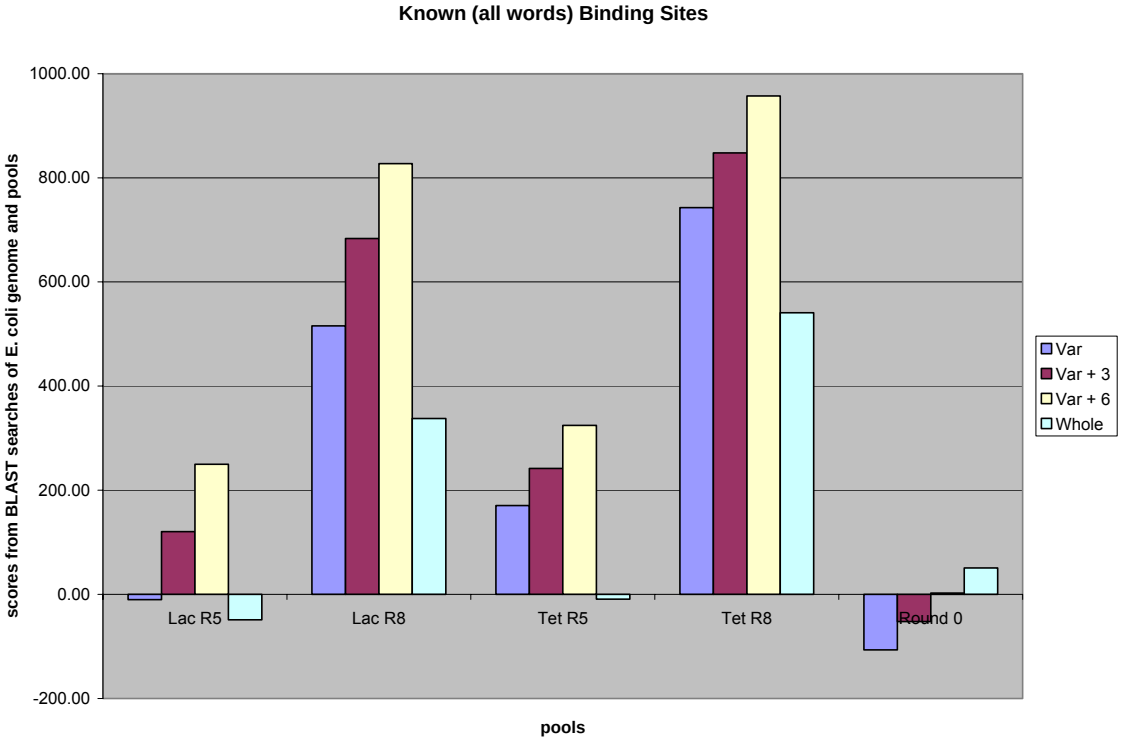


Figure 17: Score of seaches of *E. coli* genome and transcription factor binding sites.

References

Cyr J.L.; Hudspeth A.J. (2000). A library of bacteriophage-displayed antibody fragments directed against proteins of the inner ear. Proc Natl Acad Sci U S A 97:2276-81.

Salgado H.; Gama-Castro S.; Martinez-Antonio A.; Diaz-Peredo E.; Sanchez-Solano F.; Peralta-Gil M.; Garcia-Alonso D.; Jimenez-Jacinto V.; Santos-Zavaleta A.; Bonavides-Martinez C.; Collado-Vides J. (2004). RegulonDB (version 4.0): transcriptional regulation, operon organization and growth conditions in Escherichia coli K-12. Nucleic Acids Res 32:D303-6.

Marcotte E.M.; Xenarios I.; Eisenberg D. (2001). Mining literature for protein-protein interactions. Bioinformatics 17:359-63.

Phil Shannon is a graduate student in the Ellington Lab who performed the analysis of the sequence data from the lysate selections.

Chapter 4: Deoxyribozyme Ligase

INTRODUCTION

Functional nucleic acid species can be selected from random sequence libraries, and in many instances have been shown to have attributes that rival those of proteins. For example, selected nucleic acid binding species (aptamers) frequently interact with their targets with affinities and specificities that rival those of monoclonal antibodies (Osborne and Ellington, 1997; Famulok and Jenne, 1998; Famulok and Mayer, 1999), while selected allosteric nucleic acid catalysts (aptazymes) have been shown to have activation parameters far in excess of those normally observed in allosteric protein enzymes (Soukup and Breaker, 2000).

Methods for automation of nucleic acid selections have been successfully developed. The selection of aptamers has been successfully automated using a Biomek 2000 workstation. Several binding species with nanomolar affinities were isolated from diverse populations. Automation of a deoxyribozyme ligase selection has also been completed. The development of automated selections of nucleic acid catalysts has proven to be very challenging. In order to make catalyst selections as amenable to automation as possible, many changes were necessary. The process requires eleven times more robotic manipulations than an aptamer selection. The random sequence pool contained a 5' iodine residue and the ligation substrate contained a 3' phosphorothioate. Several difficulties were encountered during the automation of DNA catalyst selection, including effectively washing bead-bound DNA, pipetting 50% glycerol solutions, purifying single strand DNA, and monitoring the progress of the selection as it is performed. Nonetheless,

automated selection experiments for deoxyribozyme ligases were carried out starting from a naive pool.

The process of *in vitro* selection was automated by converting molecular biology steps that were normally carried out at the bench to steps that could be carried out by an automated workstation. As with most molecular biology protocols, the primary steps involved liquid handling, and therefore we initially chose a robust automated workstation for liquid handling, the Beckman Biomek 2000. The work surface of this robot could also be readily manipulated, and we were therefore able to integrate a PCR machine with the liquid handling system. We also adapted an enzyme cooler and other devices to the surface, ultimately converting the Biomek 2000 to a selection robot.

The most difficult problem that had to be resolved was nucleic acid purification. The Biomek (and, indeed, almost any robotic system) is incapable of such common molecular biology procedures as gel electrophoresis, centrifugation, and so forth. Therefore, we attempted to design selection experiments that would require little or no purification of nucleic acids. We limited the number of amplification cycles that were carried out by the robot so that we would in turn limit the accumulation of DNA templates that had folded back on themselves and amplified via a single primer (so called 'one-primer artefacts,' see also Green et al., 1991). In the end, these precautions allowed us to readily carry out selections that produced amplicons of discrete sizes throughout the course of a selection experiment.

The final automated selection system can carry out 4 rounds of selection in a 24 hour period without the need of human intervention. An entire selection typically takes from 8 to 12 cycles (from 2 to 3 days). The progress of the selection is monitored by carrying out activity assays during the selection.

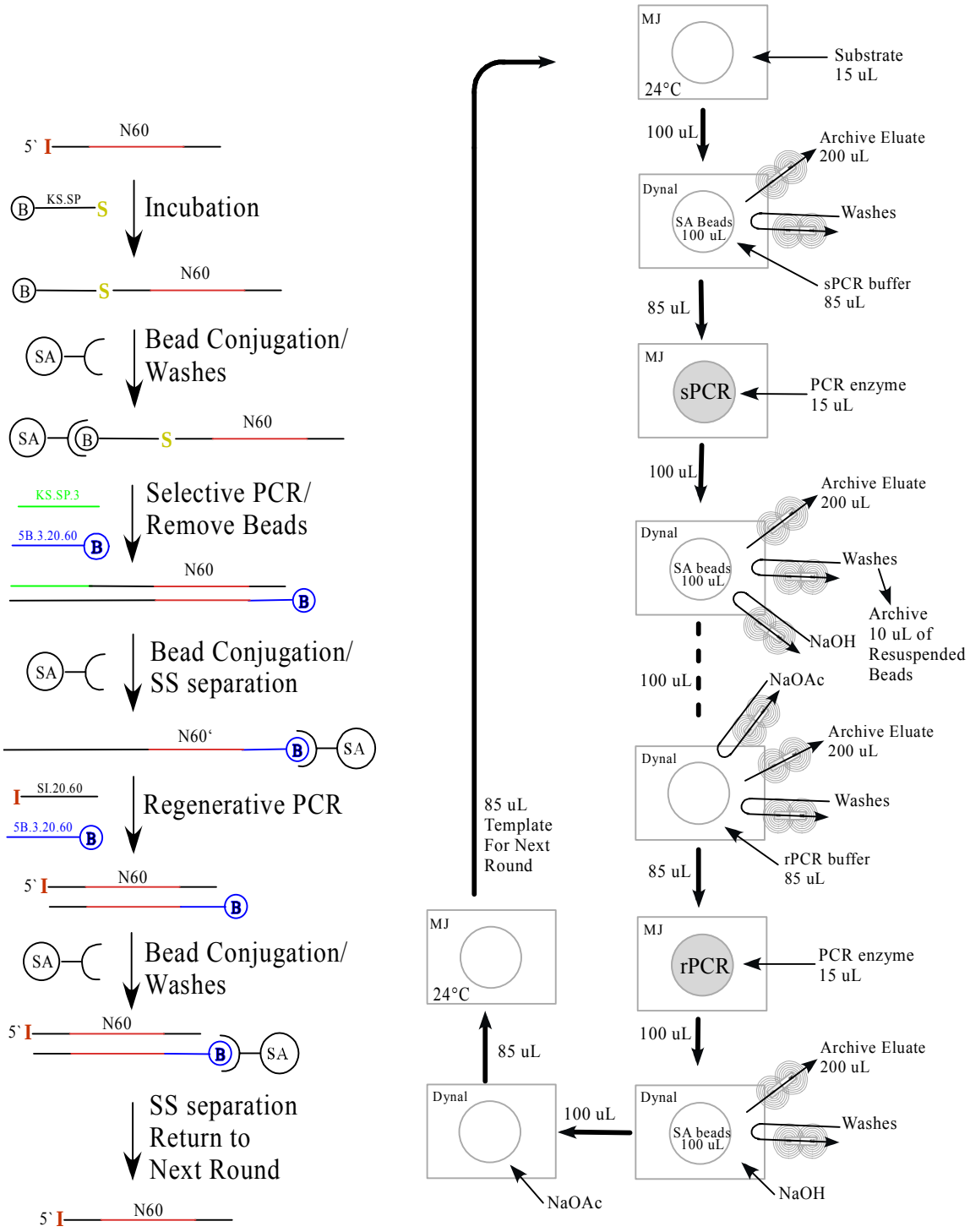
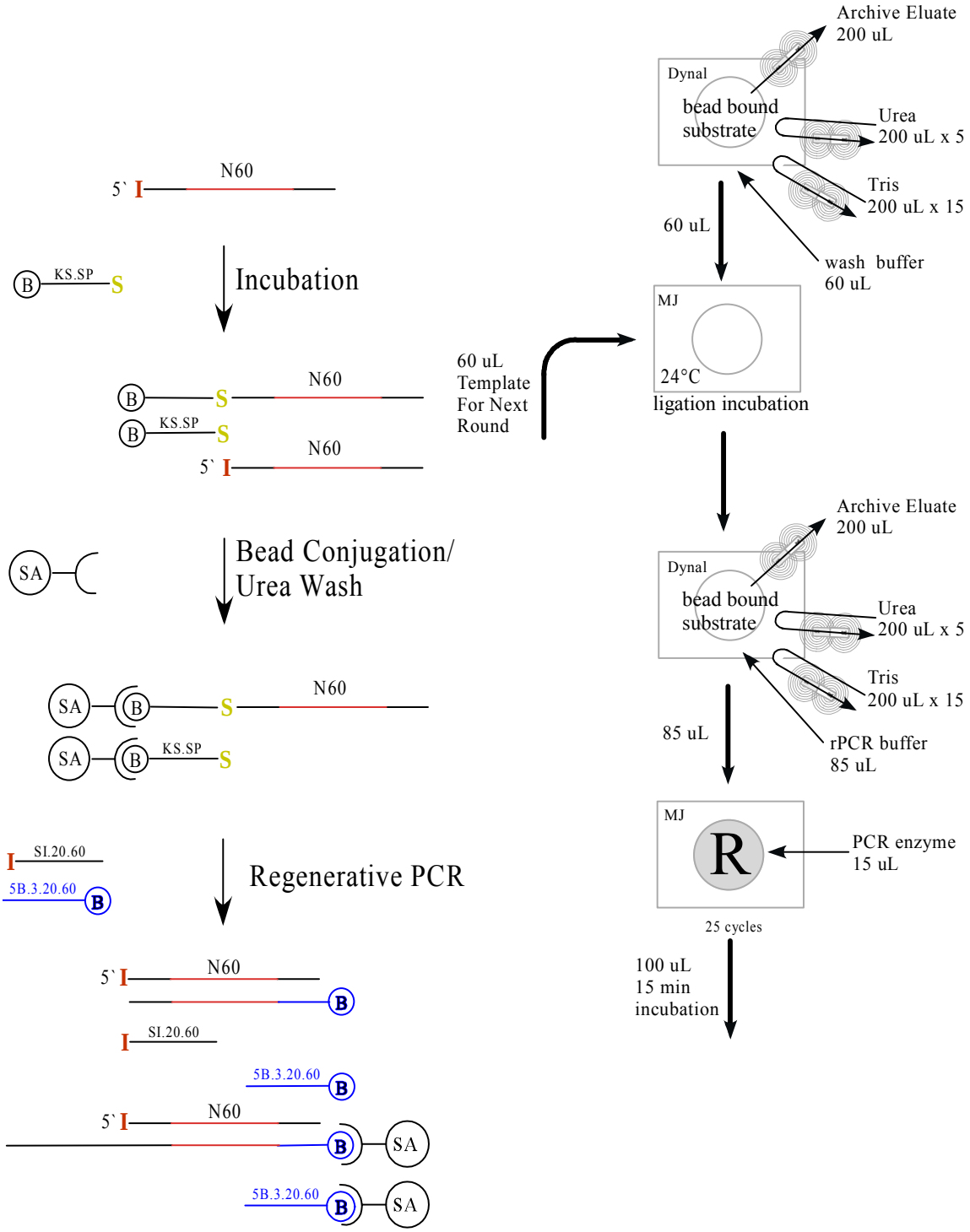


FIGURE 18: Automated method with "selective" and "regenerative" PCR steps.

“Paste” DNA Enzyme Selection Scheme Rounds 0-3



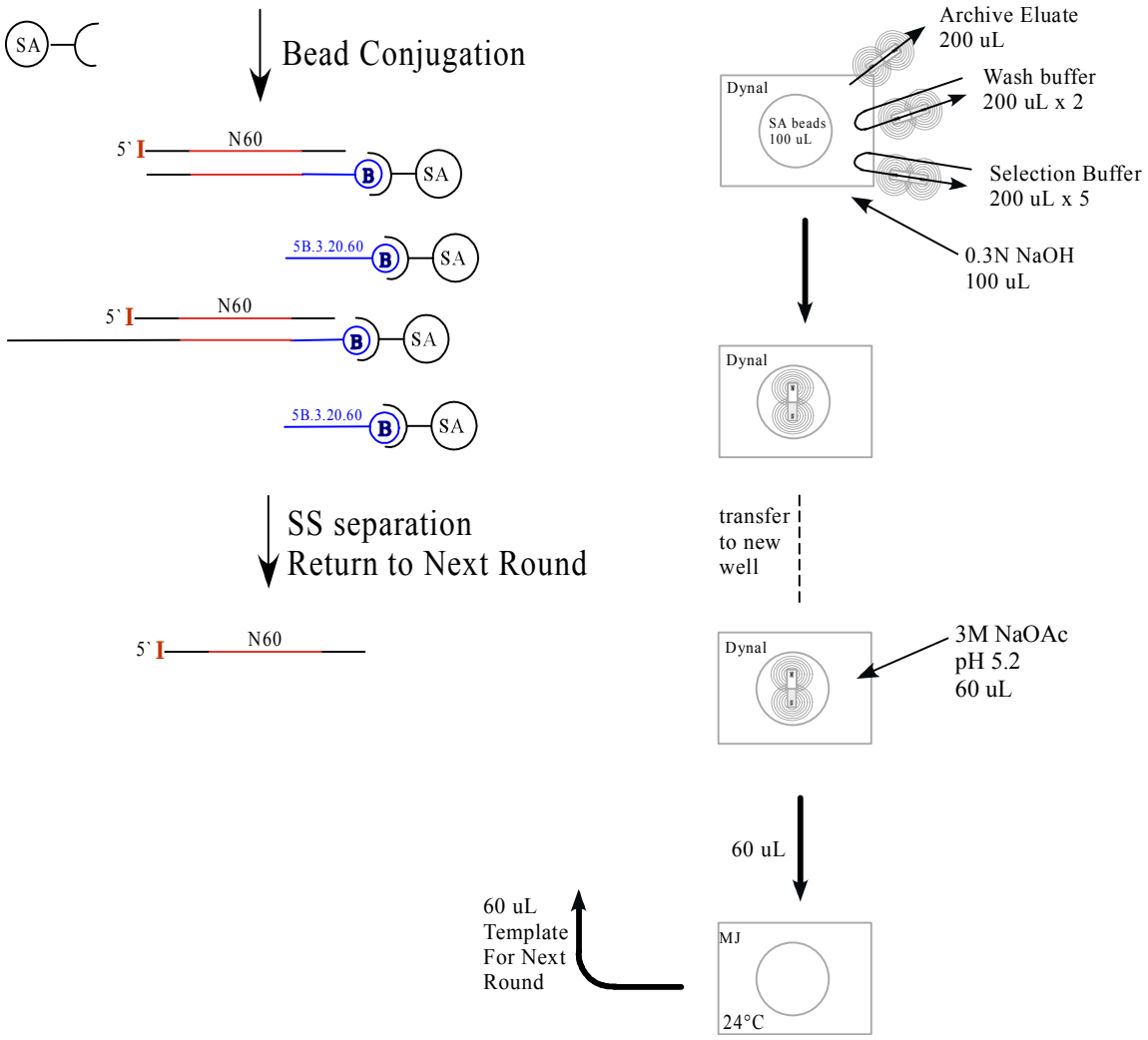


Figure 19: Automated selection with only regenerative PCR.

The various manual manipulations that are typically performed during *in vitro* selection experiments are straightforward, but their adaptation to a robotic workstation nonetheless proved difficult (reported more completely in Cox et al., 1998, and Cox and Ellington, 2001). In particular, the workstation is currently incapable of gel isolation and ethanol precipitation, and this necessitated the development of methods that either required alternative purification or no purification. This facet of experimental design is especially important given that the workstation cannot monitor the success of each cycle of selection and amplification, and thus the input into standard PCR reactions will vary as the selection progresses. The partitioning of selected species is also different on a robotic workstation than it is for a human experimentalist.

Integration of the MJ thermal cycler required the use of a null modem cable with the following pinout configuration.

A) Standard	B) Null Modem	C) Biomek2000
DB9 ---- DB9	DB9 ---- DB9	DB9 ---- DB9
1 -------- 1		
2 -------- 2	2 -------- 3	2 -------- 3
3 -------- 3	3 -------- 2	3 -------- 2
4 -------- 4	4 -------- 1,6	4 -------- 6
5 -------- 5	5 -------- 5	5 -------- 5
6 -------- 6	1,6 ------ 4	6 -------- 4

7 -------- 7	7 -------- 8	7 -------- 8
8 -------- 8	8 -------- 7	8 -------- 7

D)

1) Carrier Detect
2) Receive Data
3) Transmit Data
4) Data Terminal Ready
5) System Ground
6) Data Set Ready
7) Request to Send
8) Clear to Send

Figure 20: Pinout configuration for 9-pin RS-232 cables. A) Standard cable, B) Standard null modem cable, C) Null modem cable used to connect the Biomek2000 workstation with the MJ Research thermal cycler, D) Pin designations.

Automated Selection of Nucleic Acid Catalysts

Given that it proved possible to select deoxyribozyme ligases from random sequence pools, we immediately attempted to adapt the selection procedure to the automated workstation. The automated schema is shown in Figure 18, and differs from the previously described manual selection procedure in several important ways. First, the substrate for the reaction was biotinylated. Following reaction with the DNA pool, ligated

species were captured on magnetic beads bearing streptavidin. This allowed the facile separation of reacted and unreacted DNA species via the integrated magnetic bead separator (MPC-auto96, Dynal, Oslo, Norway). Second, in the manual selection procedure unreacted DNA molecules were removed using 80 ml of wash buffer. To decrease the volume of washes required while maintaining the stringency of separation, the magnetic beads were washed with 7 M urea five times (200 µl) followed by fifteen washes with 10 mM Tris (pH 7.8, 200 µl). Retained, reacted DNA molecules were amplified directly on beads (a 'selective' amplification step). The Taq polymerase solution (50% glycerol) proved difficult for the Biomek to pipette, and a special transfer step involving a slow aspiration was introduced. While some of the amplified DNA products likely remained bound to the streptavidin beads, the remaining amplified DNA was captured on a second set of magnetic beads and the non-biotinylated strand was removed by alkaline denaturation and neutralization. Several methods of single-strand DNA purification were initially explored: heat, urea, sodium hydroxide, heat and urea, and heat and sodium hydroxide. Double-stranded, biotinylated PCR product radiolabeled with [32P]-dATP was incubated with streptavidin coated magnetic beads. The beads were washed and then incubated for varying amounts of time with one of the five denaturing conditions. Afterwards, the solutions were separated from the beads and run on an acrylamide gel. Phosphoimage analysis revealed that sodium hydroxide, with or without heat, was the most effective denaturant. In additional experiments, incubation times and temperatures were systematically varied. Following separation, single-stranded products were incubated with substrate (KS.S.3) and a splint oligonucleotide (KA.T.1.18) for 24 hours at 24 °C.

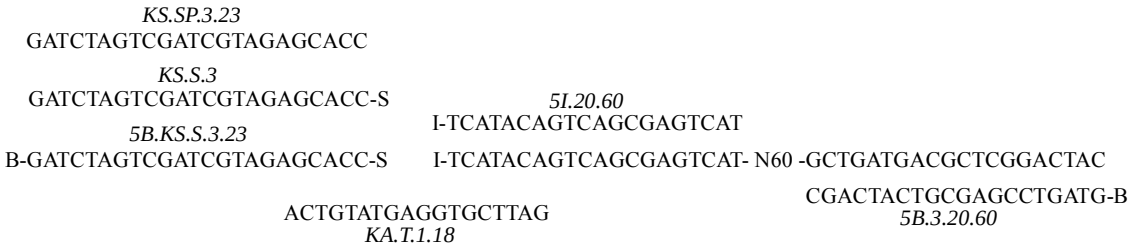


Figure 21: N60 pool and all oligos used.

The splint brings the pool and substrate into close proximity and should allow efficient ligation. Of course, if no 5' iodine moiety is present, ligation will not occur. The optimal combination of single strand separation and 5' iodine moiety retention was achieved with a sodium hydroxide wash at room temperature. The remaining, biotinylated strand was then amplified using a second set of primers that restored the original form of the deoxyribozyme, including the 5' iodine moiety (a 'regenerative' amplification step). Single-stranded DNA was purified via alkaline denaturation and neutralization. The selected, single-stranded pool was then subjected to further rounds of selection and amplification. Overall, a single round of automated deoxyribozyme selection requires 11 000 individual, programmed movements of the robot. To gauge the complexity of the procedure, it should be realized that the automated aptamer selections which have only recently been successfully implemented require only 1000 movements per round.

The automated selection procedure was attempted with a new pool, N60. After the first cycle of selection and amplification, deoxyribozyme ligase activity appeared to have accumulated, but was lost in subsequent cycles. In order to better gauge whether the loss of ligase activity was due to the automated procedure or was somehow a function of the

new pool, a second automated selection experiment was set up involving the deoxyribozyme pool from round 8 of a manual selection. Again, after one cycle of selection and amplification, deoxyribozyme ligase activity improved, but was lost in subsequent rounds (Figure 22).

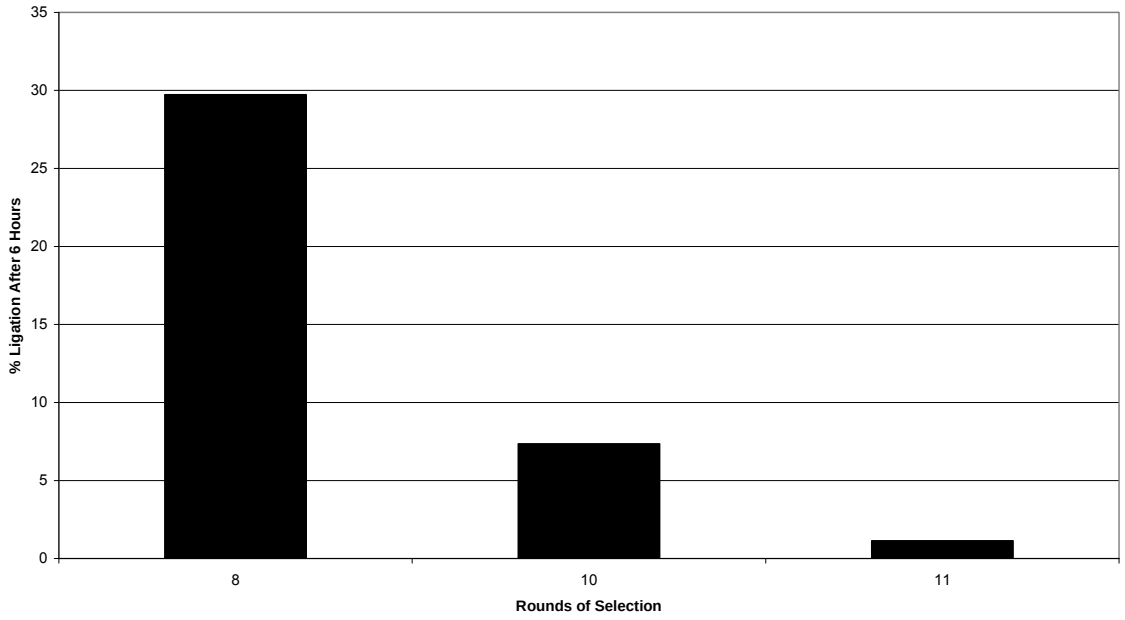


Figure 22: Deoxyribozyme ligase selection with both "selective" and "regenerative" PCR

A close examination of the PCR products that accumulated during both the 'selective' and 'regenerative' steps revealed a series of bands, indicating that mis-priming was a serious problem during amplification reactions. (Figure 23)

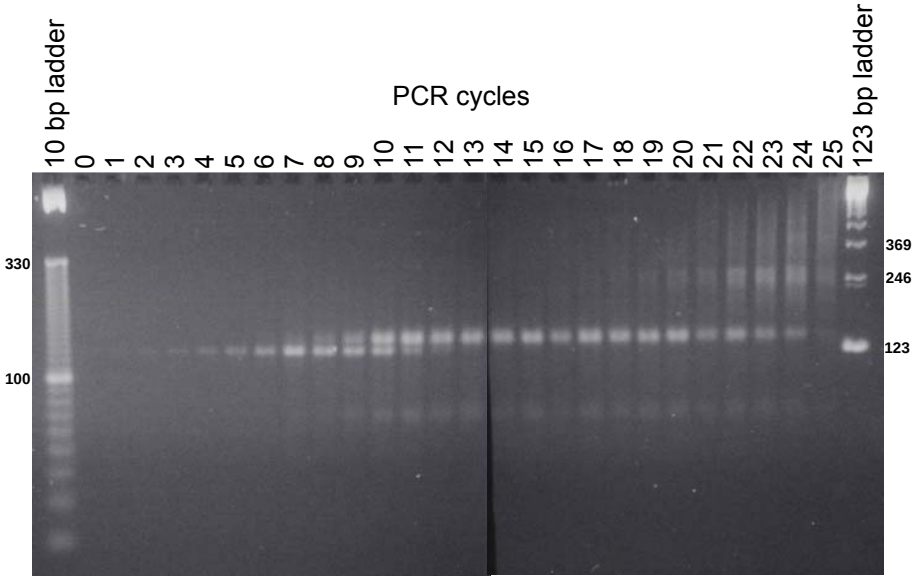


Figure 23: PCR cycle course using N60 pool.

Cycle-course reactions that mimicked the automated 'selective' and 'regenerative' amplifications also showed the accumulation of higher molecular weight bands after nine PCR amplification cycles (Figure 23). It seems likely that the pools that were used for deoxyribozyme selection may be prone to the accumulation of amplification artifacts. There are several possible solutions to this problem. First and most obviously, the N60 pool can be re-designed with alternative constant regions that are potentially less prone to the accumulation of amplification artifacts. Second, fewer rounds of amplification can be used during both the PCR amplification steps. However, this latter alternative points up the single most important problem with robotic selection: in its current form it tends to be blind. Rounds of selection and amplification can be carried out, but until the machine has completed its run there is no way to ascertain how well the selection is progressing.

Ultimately, the "selective" PCR step was removed from the process and only the "regenerative" PCR step was used. (See Figure 18 vs. Figure 19)

Results and Discussion

Selection and Amplification Procedures

The selection begins with the incubation of a single-stranded DNA pool and biotinylated substrate in a thermal cycler (PTC-200, MJ Research, Waltham, USA) held at 24 °C. Following the incubation, the solution is transferred to a well in the magnetic bead separator (Dynal) containing streptavidin coated magnetic beads in 2× binding buffer (10 mM Tris, 2 M NaCl, pH 7.5). A bead binding incubation occurs with mixing, and then the beads undergo a stringent wash with five 200 µl washes of 7 M urea and fifteen 200 µl washes of 10 mM Tris, pH 7.5. This co-immobilizes the species in the pool that have successfully ligated to the substrate. The beads are resuspended in a 'selective' PCR (sPCR) mixture and transferred to the thermal cycler where the immobilized DNA molecules serve as amplification templates. Fifteen µl of a 50% glycerol solution containing five units of Taq polymerase (Display Systems Biotech) are transferred from a homemade enzyme cooling unit and the thermal cycler automatically performs 10 PCR amplification cycles. The amplified DNA molecules are then transferred to the Dynal magnetic bead separator and the biotinylated double strands are captured on a new set of streptavidin beads. Following capture, the beads undergo a stringent wash with 7 M urea and 10 mM Tris as before. The beads are then incubated with 0.3 N NaOH to denature the double-stranded DNA, releasing the non-biotinylated single strands. The solution containing single-stranded DNA is removed and discarded, and the beads are returned to a neutral pH with 3 M sodium acetate (NaOAc), pH 5.2. The beads bearing biotinylated,

single-stranded DNA are washed with wash buffer (5 mM Tris-HCl, 1 M NaCl, pH 7.5) and used as the template for a 'regenerative' PCR (rPCR) reaction; 10 PCR amplification cycles are performed. This solution is transferred to the magnetic bead separator where the beads are captured, and the supernatant is transferred to a fresh set of streptavidin beads. Bead binding and stringent washing are followed by a single strand separation using 100 microliters of 0.3 N NaOH. The alkaline solution is transferred to a new well on the magnetic bead separator and neutralized with 60 µl of 3 Msodium acetate, pH 5.2. This solution contains the regenerated, iodinated, single stranded DNA pool that will be used in the next round of selection.

In order to simplify the procedure, the "selective" PCR step was removed. With this modification, fewer PCR steps were required and therefore the pool had fewer opportunities to form artifacts. This change allowed the successful selection of deoxyribozyme ligases.

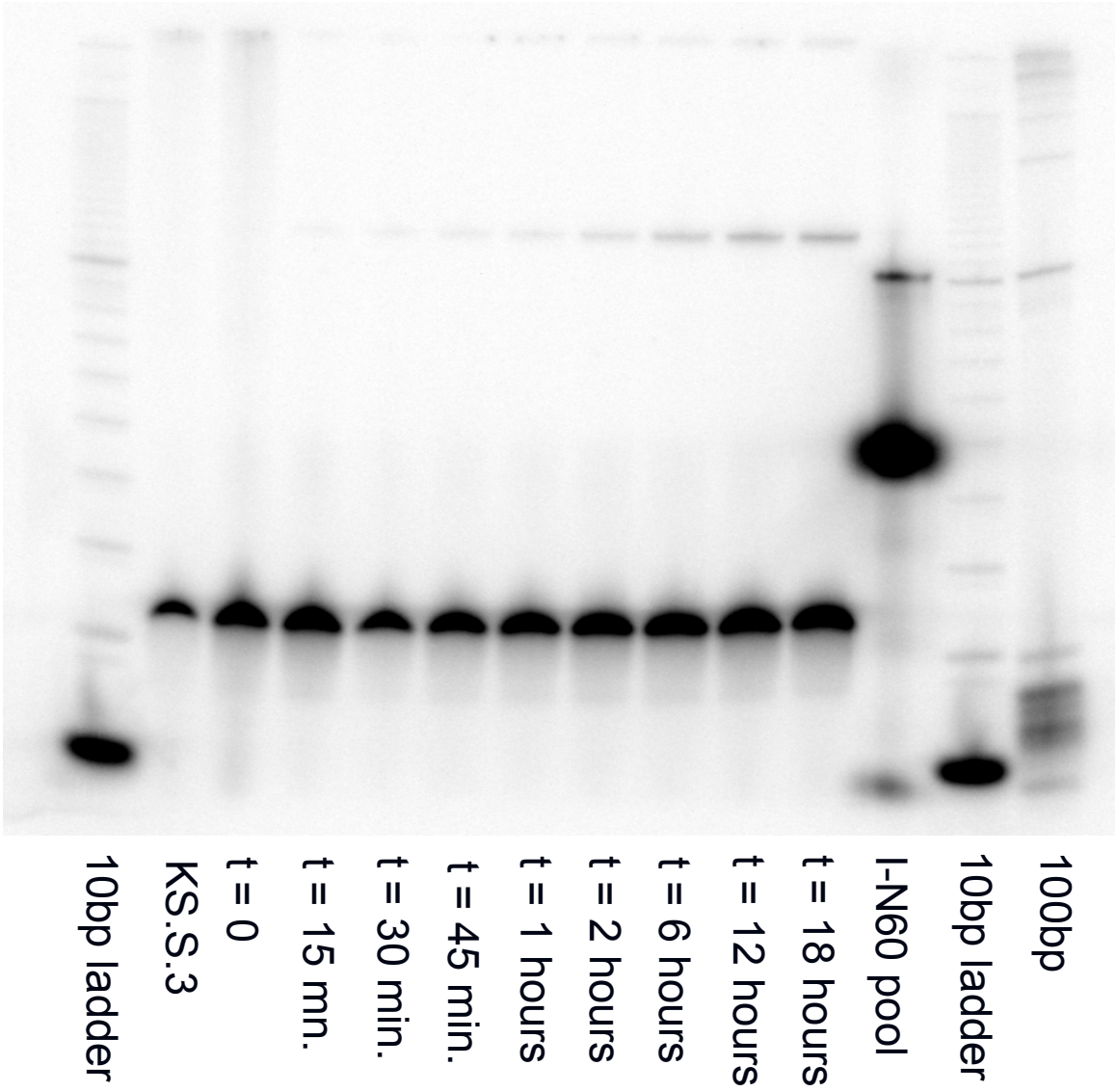


Figure 24: Ligation of pool to substrate over a time course.

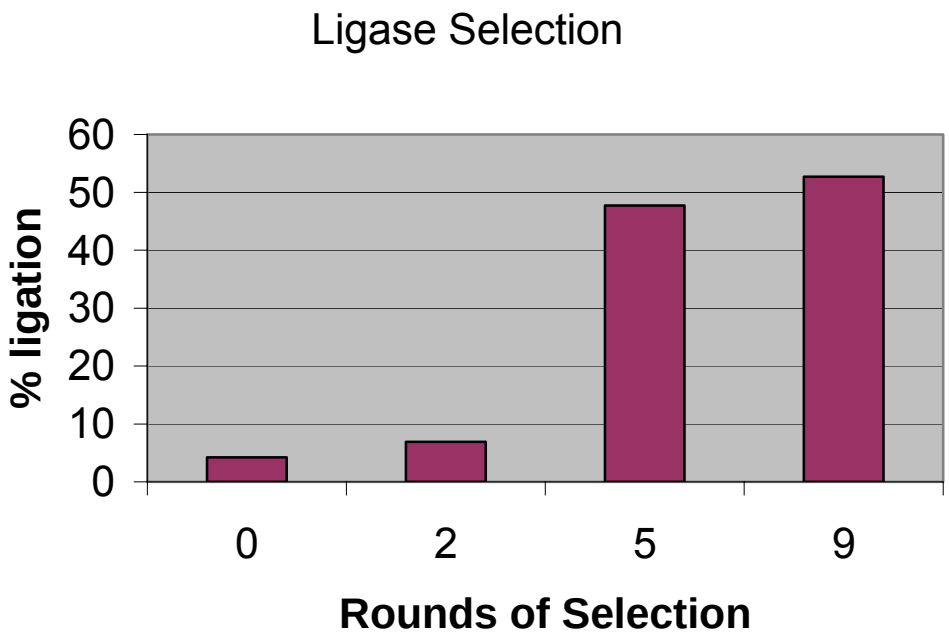


Figure 25: Ligation of pool to substrate over 24 hour period.

REFERENCES

Bell, S.D., Denu, J.M., Dixon, J.E., and Ellington, A.D. (1998). RNA molecules that bind to and inhibit the active site of a tyrosine phosphatase. J. Biol. Chem. 273, 14309 – 14314.

Collins, R.A., and Lambowitz, A.M. (1985). RNA splicing in Neurospora mitochondria. Defective splicing of mitochondrial mRNA precursors in the nuclear mutant cyt18-1. J. Mol. Biol. 184, 413 – 428.

Cox, J.C., and Ellington, A.D. (2001). Automated selection of anti-protein aptamers. Bioorg. Med. Chem., in press.

Cox, J.C., Rudolph, P., and Ellington, A.D. (1998). Automated RNA selection. Biotechnol. Prog. 14, 845 – 850.

Crameri, A., and Stemmer, W.P. (1993). 10(20)-fold aptamer library amplification without gel purification. Nucleic Acids Res. 21, 4410.

Famulok, M., and Jenne, A. (1998). Oligonucleotide libraries – variatio delectat. Curr. Opin. Chem. Biol. 2, 320 – 327.

Famulok, M., and Mayer, G. (1999). Aptamers as tools in molecular biology and immunology. Curr. Top. Microbiol. Immunol. 243, 123 – 36.

Guo, Q., and Lambowitz, A.M. (1992). A tyrosyl-tRNA synthetase binds specifically to the group I intron catalytic core. Genes Dev. 6, 1357 – 1372.

Levy, M., and Ellington, A.D. (2001). Selection of deoxyribozyme ligases that catalyze the formation of an unnatural internucleotide linkage. Bioorg. Med. Chem., in press.

Liao D., and Dennis, P.P. (1992). The organization and expression of essential transcription translation component genes in the extremely thermophilic eubacterium Thermotoga maritima. J. Biol. Chem. 267, 22787 – 22797.

Marshall, K.A., and Ellington, A.D. (1999). Molecular parasites that evolve longer genomes. J. Mol. Evol. 49, 656 – 663.

Marshall, K.A., and Ellington, A.D. (2000). *In vitro* selection of RNA aptamers. Methods Enzymol. 318, 193 – 214.

Opperman T., and Richardson J.P. (1994). Phylogenetic analysis of sequences from diverse bacteria with homology to the Escherichia coli rho gene. J. Bacteriol. 176, 5033 – 5043.

Osborne, S.E., and Ellington, A.D. (1997). Nucleic acid selection and the challenge of combinatorial chemistry. Chem. Rev. 97, 349 – 370.

Pollard, J., Bell, S.D., and Ellington, A.D. (2000). Design, synthesis, and amplification of DNA pools for *in vitro* selection. In: Current Protocols in Nucleic Acid Chemistry, S.L. Beaucage et al., eds. (New York, USA: John Wiley & Sons, Inc.), pp. 9.2.1 – 9.2.23.

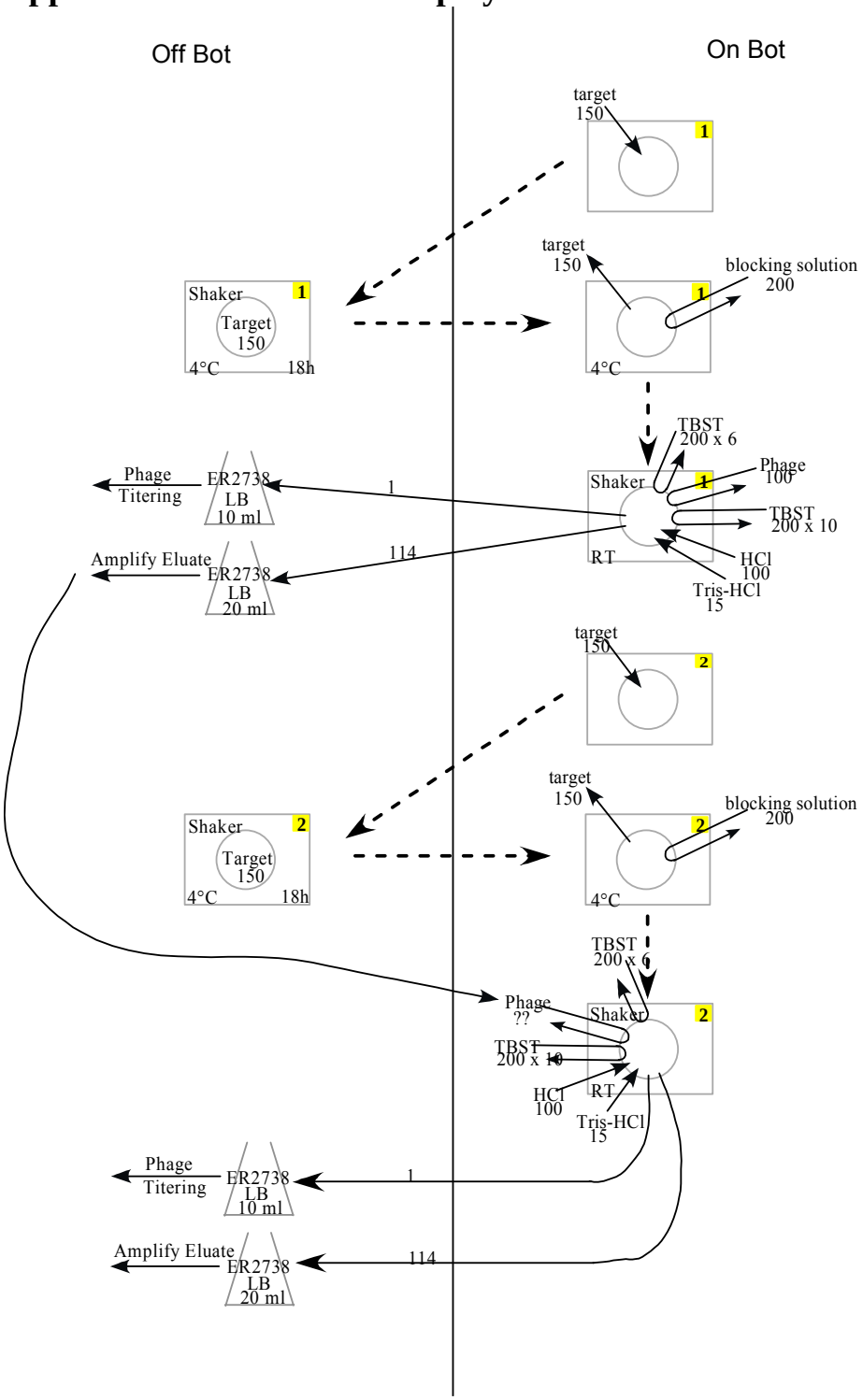
Roller R.J., Monk L.L., Stuart D., and Roizman B. (1996). Structure and function in the herpes simplex virus 1 RNA-binding protein U(s)11: mapping of the domain required for ribosomal and nucleolar association and RNA binding *in vitro*. J. Virol. 70, 2842 – 2851.

Schneider, D., Gold, L., and Platt, T. (1993). Selective enrichment of RNA species for tight binding to Escherichia coli rho factor. FASEB J. 7, 201 – 207.

Sooter L.J.; Riedel T.; Davidson E.A.; Levy M.; Cox J.C.; Ellington A.D. (2001). Toward automated nucleic acid enzyme selection. Biol Chem 382:1327-34.

Soukup, G.A., and Breaker, R.R. (2000). Allosteric nucleic acid catalysts. Curr. Opin. Struct. Biol. 10, 318 – 325.

Appendix 1: Cell Surface Display Selection



Automated Method:

The process will begin with target preparation. The target is immobilized on the surface of a 96-well microplate via hydrophobic interactions. First the Biomek FX workstation will transfer the protein targets into specific target 96well microplates. Target incubation will occur for off the robot for 18 hours at 4C. The wells will be washed with blocking solution by robotically pipetting the buffer in and out of the wells. Then blocking agent will be allowed to incubate in the wells for at least two hours at room temperature. Following the blocking wash, the wells will be washed with buffer to remove any non bound target or blocking agent. While the target microplates are being blocked, the cell library must be blocked as well. This step helps prevent cells which preferentially bind background from being carried through to subsequent rounds of selection. After the target plates have been prepared, the blocked cell library is added directly to the target plate. The microplate is placed on a room temperature shaker for one hour, and then the cell solution is removed. The plate is washed with buffer to remove any non-bound cells and then hydrochloric acid is added to elute the cells. Following neutralization of the solution, the cells are transferred to warm, rich media and allowed to grow overnight. Following the growth period, an aliquot of the cell culture is diluted in blocking solution and added to a new target well

Appendix 2: Quantitation and Normalization Software

Quantitation and normalization software was designed for the Tecan Genesis workstation using the Tecan Gemini software, Visual Basic Script, and Excel. The purpose of this software is to take a number of samples, all of different concentration, and dilute them so that they are all the same concentration.

First, the samples are placed in an optically clear microplate and the A260 or A280 value is read. The output data is placed in an Excel spreadsheet where the concentration of the sample and the necessary dilutions are calculated. A visual basic script then takes these numbers, assigns variable names to them, and inputs them into the Tecan Gemini software. The Tecan then pipettes according to the variable values and dilutes all of the samples to the same concentration.

Below is the online guide to using the software and setting up the Tecan for a run.

Quantitation and Normalization on the Tecan Genesis Workstation

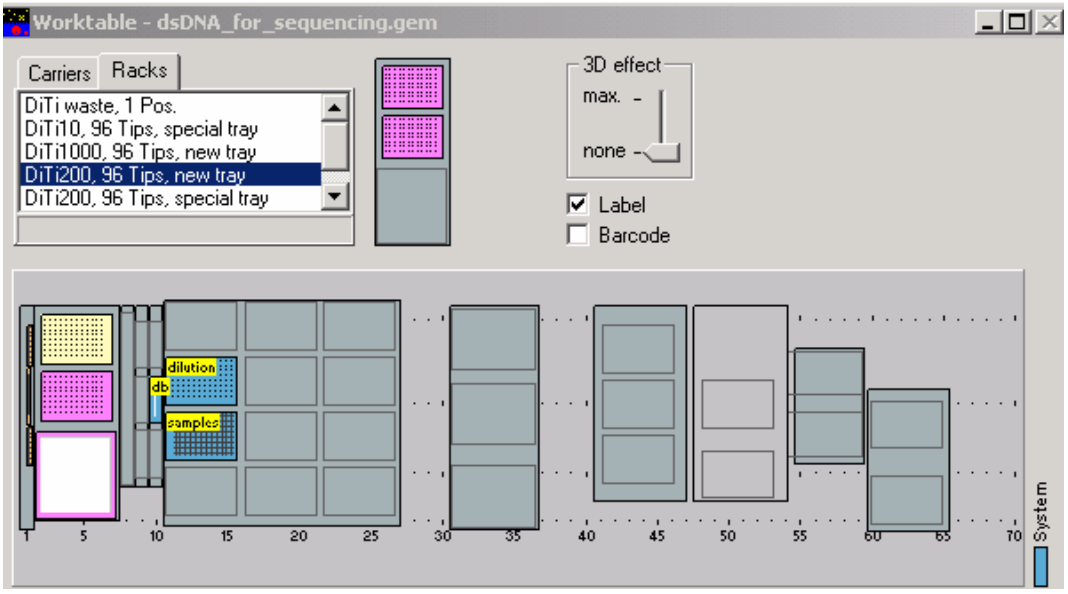
Carefully go through this list every time you run the Tecan. If you forget ANYTHING, even the tinest detail, the Tecan will not run properly.

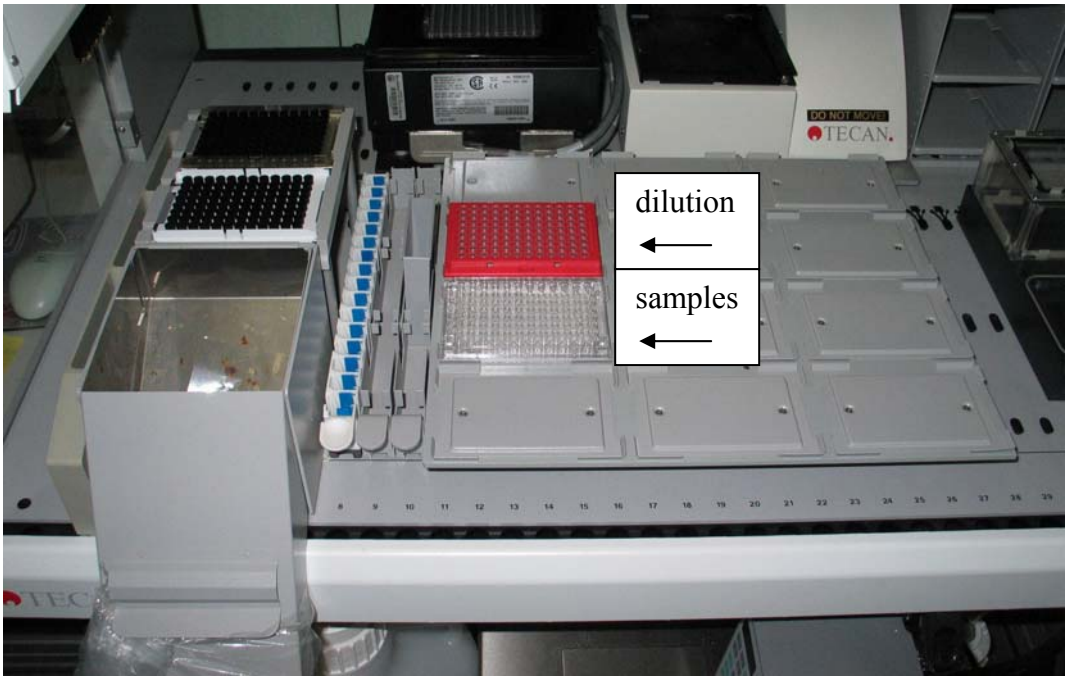
The Tecan is located in ESB 445. The doors are always locked, so plan accordingly.

The consumables are as follows:

dilution	MJ Hardshell plate (HSP-9601)	Empty
sample	Costar UV transparent plate (3635)	Contains samples to be diluted
db	Tecan 100mL reservoir	Contains dilution buffer
DiTi200	Tecan 200uL barrier tips	

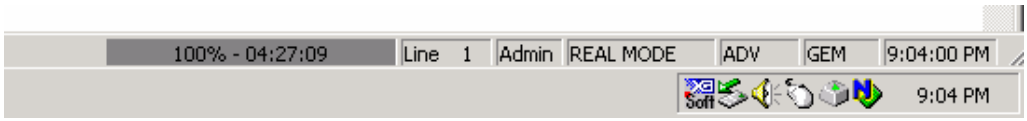
The worksurface layout is as follows:





Once you have all of your solutions aliquotted into your plates, work through the following checklist:

1. Make sure that the Tecan is on (The switch is on the far right side of the top front of the machine. A green light is on when the machine is on).
2. Open Excel and Gemini. There are two forms of Gemini: Gemini and Gemini Simulation. Make sure you open Gemini. The software will operate normally, but the robot will not run under Gemini Simulation. Another way to double check this is to look at the bottom right-hand area of the screen. If it says "REAL MODE" then the robot will run. If it says "SIMULATION" you must shut down the program and open Gemini.



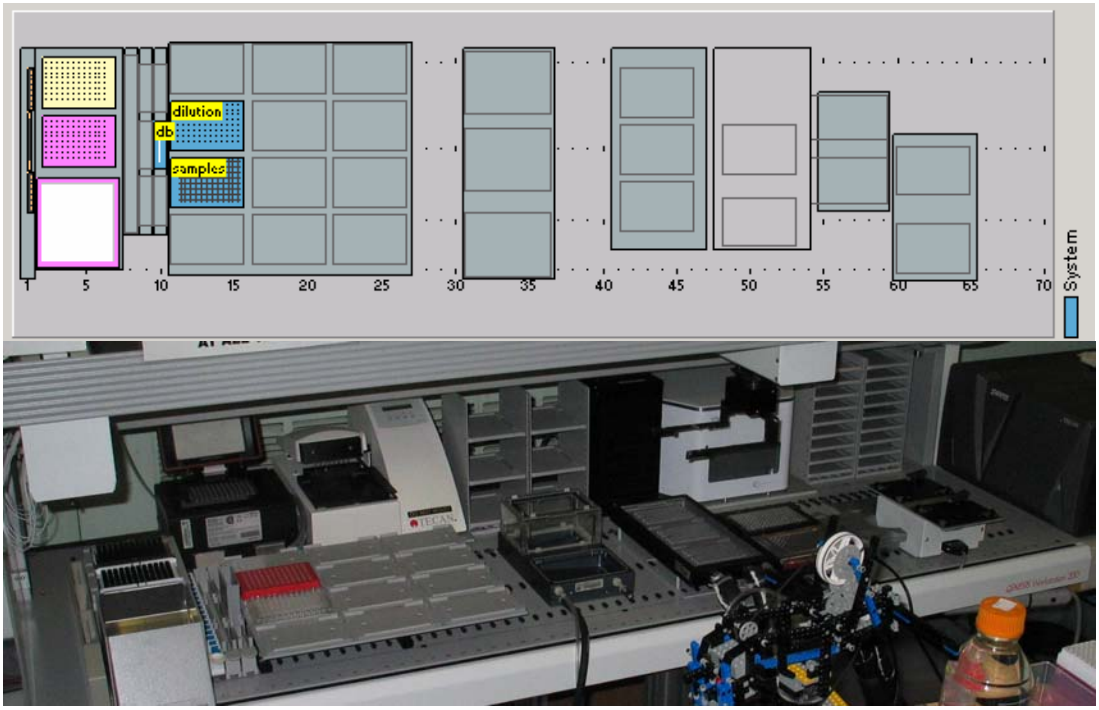
3. Open the Q&N program in Gemini (such as 'dsDNA_5ug_dilution'). Open the Excel file with the corresponding name (i.e. 'dsDNA_5ug_dilution'). Icons for both programs are on the desktop.

My Computer → C: → Q&N → dsDNA_for_sequencing.xls

My Computer → C: → Gemini → Data → Q&N → dsDNA_for_sequencing.gem

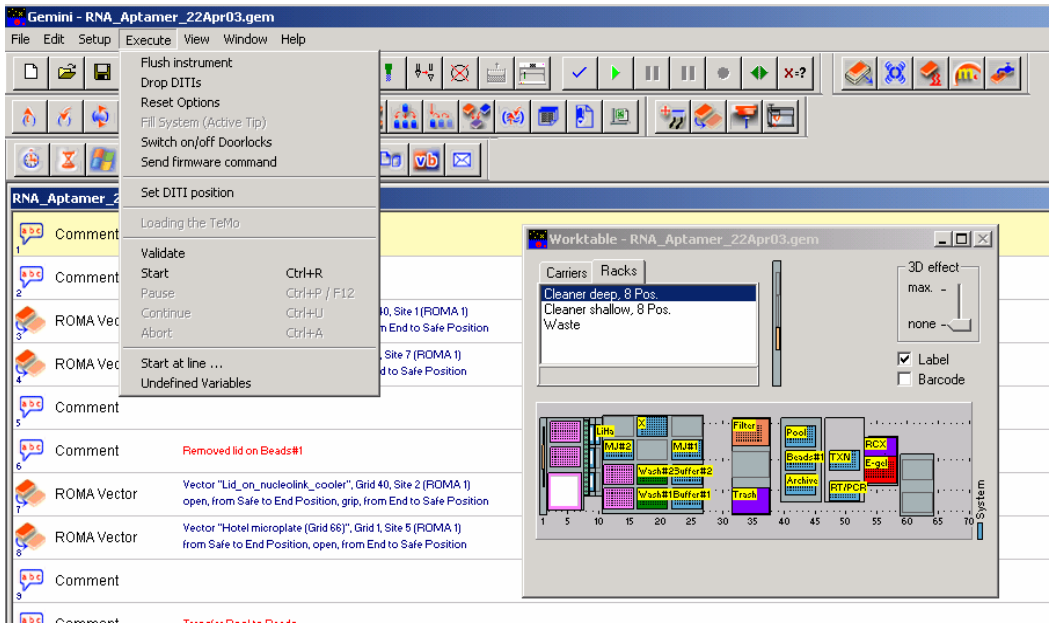
4. Make sure the worksurface is clean. If you see any dirt or splatters, clean the worksurface with water and alconox.

5. Check that the positions of the carriers on the worksurface match the position of the carriers in the program. Double check that the area immediately in front of the Safire is clear.

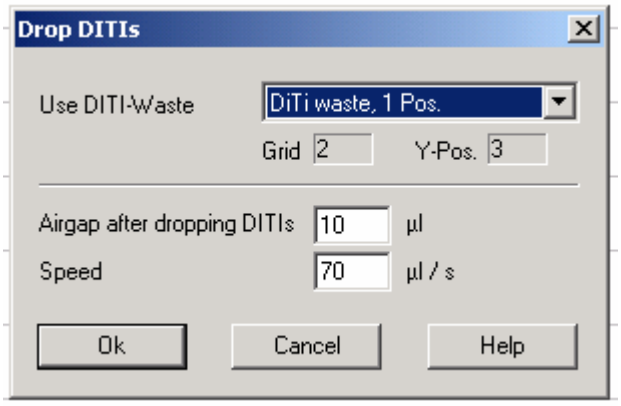

dilution
db
samples
1
5
10
15
20
25
30
35
40
45
50
55
60
65
70
System


SAFIRE
TECAN
GENESIS

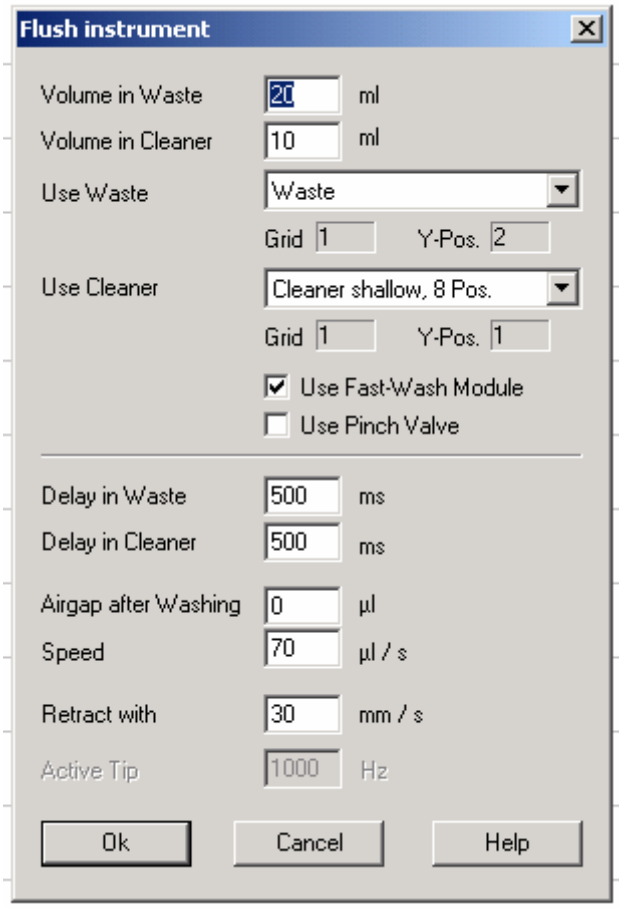
6. Place the racks on the worksurface (plates, reservoirs, and tips), and make sure that the position of the racks on the worksurface match the position of the racks in the program.
7. Adjust all racks so that they sit as far back and to the left as they can go (towards the A1 position).
8. Drop all DiTi's, even if there are no DiTi's mounted.
    a. Execute --> Drop DITIs



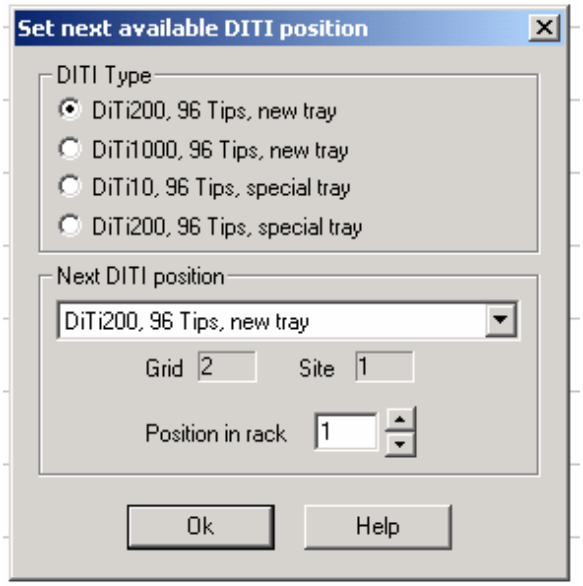
    b. The following window should appear. Click 'OK'.

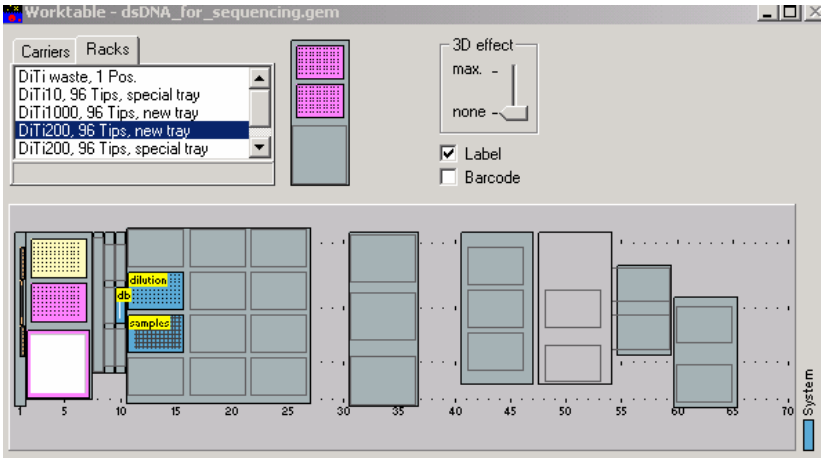


9. Flush Instrument. If you do not do this and remove the air bubbles from the lines, the pipetting will not be accurate. If one run of 'Flush Instrument' does not remove all of the air bubbles, do it again until none remain.
    a. Execute --> Flush Instrument
    b. The following window should appear. Click 'OK'.



10. Set DiTi Position
    a. Execute --> Set DITI position
    b. The following window should appear. Make sure everything matches this figure.
        i. The upper tip rack is highlighted in yellow on the worksurface layout. If it is not, just click on it and it should highlight it.
        ii. The "Next DITI position" should read "DiTi200, 96 Tips, new tray"
        iii. The "Position in rack" should read "1"





You should now be ready to run the program.

Click on the green ‘play’ arrow up in the rows of shortcut icons, stand back, and keep your fingers crossed.

Do not touch the robot while it is running.
Stay behind the caution tape.
The Tecan is a powerful piece of machinery. Even if you push as hard as you can, you cannot stop the pods from moving. It will poke, pinch, puncture, and cut and keep right on going.

After the program has finished. Remove both plates, the reservoir, and the plastic tip holders from the worksurface. Check for any spills or splatters and clean them up if you see any.

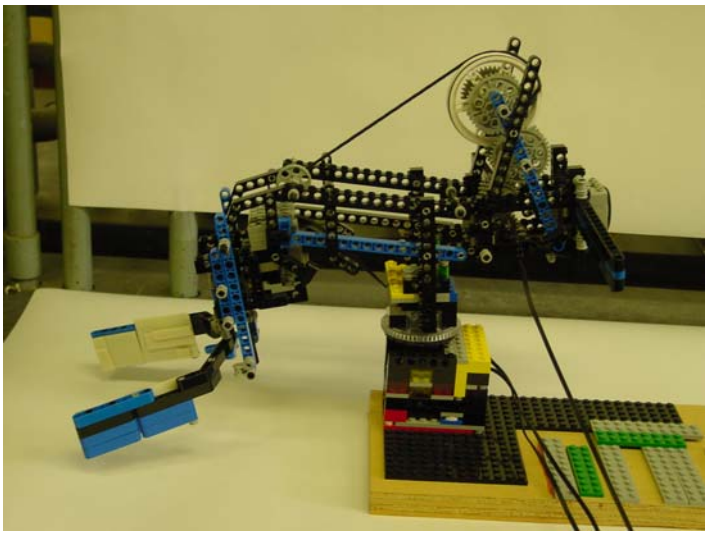
Close Gemini.

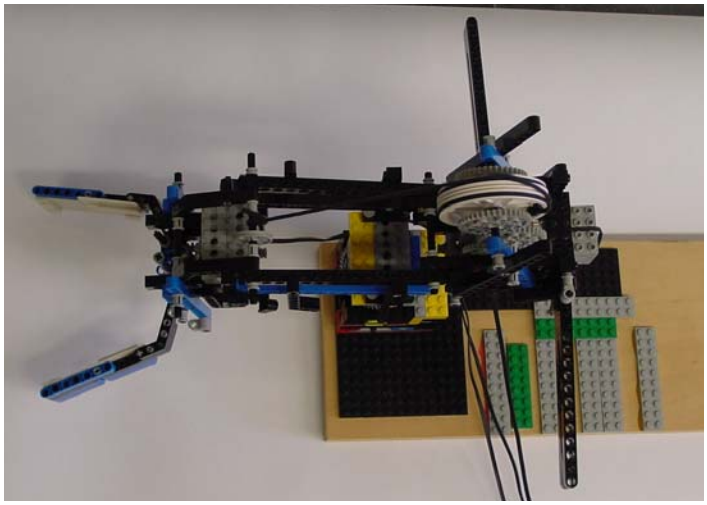
DO NOT SAVE ANY CHANGES!!!

Save the Excel file under a new name and close Excel.

DO NOT SAVE ANY CHANGES TO THE CURRENT EXCEL FILE!!!

Appendix 3: Lego Arm





A robotic arm was constructed out of a Lego Mindstorms kit in order to transfer microplates from the robotic worksurface into the trash can. Commercial robotic arms such as the Hudson plate crane and the Twister arm cost approximately $30,000. The Legos cost $300. The Mindstorms control brick was placed on the Tecan worksurface and the Tecan picked up a tip and used it to press the "on" and "go" buttons on the control brick. The result was that the Lego robotic arm picked up a microplate off of a custom designed platform, turned 90 degrees, and dropped the microplate into a trash can.

References

Angelov D.; Charra M.; Muller C.W.; Cadet J.; Dimitrov S. (2003). Solution study of the NF-kappaB p50-DNA complex by UV laser protein-DNA cross-linking. Photochem Photobiol. 77:592-6.

Bell S.D.; Denu J.M.; Dixon J.E.; Ellington A.D. (1998). RNA molecules that bind to and inhibit the active site of a tyrosine phosphatase. J Biol Chem 273:14309-14.

Benbrook D.M.; Jones N.C. (1994). Different binding specificities and transactivation of variant CRE's by CREB complexes. Nucleic Acids Res 22:1463-9.

Blackwell T.K.; Weintraub H. (1990). Differences and similarities in DNA-binding preferences of MyoD and E2A protein complexes revealed by binding site selection. Science 250:1104-10.

Bless N.M.; Smith D.; Charlton J.; Czermak B.J.; Schmal H.; Friedl H.P.; Ward P.A. (1997). Protective effects of an aptamer inhibitor of neutrophil elastase in lung inflammatory injury. Curr Biol 7:877-80.

Boardman P.E.; Oliver S.G.; Hubbard S.J. (2003). SiteSeer: Visualisation and analysis of transcription factor binding sites in nucleotide sequences. Nucleic Acids Res 31:3572-5.

Bulyk M.L.; Huang X.; Choo Y.; Church G.M. (2001). Exploring the DNA-binding specificities of zinc fingers with DNA microarrays. Proc Natl Acad Sci U S A 98:7158-63.

Collins R.A.; Lambowitz A.M. (1985). RNA splicing in Neurospora mitochondria. Defective splicing of mitochondrial mRNA precursors in the nuclear mutant cyt18-1. J Mol Biol 184:413 – 428.

Conrad R.; Keranen L.M.; Ellington A.D.; Newton A.C. (1994). Isozyme-specific inhibition of protein kinase C by RNA aptamers. J Biol Chem 269:32051-4.

Cox J.C.; Hayhurst A.; Hesselberth J.; Bayer T.S.; Georgiou G.; Ellington A.D. (2002). Automated selection of aptamers against protein targets translated *in vitro*: from gene to aptamer. Nucleic Acids Res 30:e108.

Cox J.C.; Ellington A.D. (2001). Automated selection of anti-protein aptamers. Bioorg. Med. Chem., in press.

Cox J.C.; Rudolph P.; Ellington A.D. (1998). Automated RNA selection. Biotechnol Prog 14:845 – 850.

Crameri A.; Stemmer W.P. (1993). 10(20)-fold aptamer library amplification without gel purification. Nucleic Acids Res 21:4410.

Cyr J.L.; Hudspeth A.J. (2000). A library of bacteriophage-displayed antibody fragments directed against proteins of the inner ear. Proc Natl Acad Sci U S A 97:2276-81.

Davis K.A.; Abrams B.; Lin Y.; Jayasena S.D. (1996). Use of a high affinity DNA ligand in flow cytometry. Nucleic Acids Res 24:702-6.

DeAnda A. Jr.; Coutre S.E.; Moon M.R.; Vial C.M.; Griffin L.C.; Law V.S.; Komeda M.; Leung L.L.; Miller D.C. (1994). Pilot study of the efficacy of a thrombin inhibitor for use during cardiopulmonary bypass. Ann Thorac Surg 58:344-50.

Delany I.; Spohn G.; Rappuoli R.; Scarlato V. (2002). *In vitro* selection of high affinity HspR-binding sites within the genome of Helicobacter pylori. Gene 283:63-9.

Drolet D.W.; Moon-McDermott L.; Romig T.S. (1996). An enzyme-linked oligonucleotide assay. Nat Biotechnol 14:1021-5.

Ellington A.D. (1994). Empirical explorations of sequence space: host-guest chemistry in the RNA world. Ber Bunsenges Phys Chem 98:1115–21.

Ellington A.D.; Leclerc F.; Cedergren R. (1996). An RNA groove. Nat Struct Biol 3:981-4.

Famulok M.; Jenne A. (1998). Oligonucleotide libraries – variatio delectat. Curr Opin Chem Biol 2:320 – 327.

Famulok M.; Mayer G. (1999). Aptamers as tools in molecular biology and immunology. Curr Top Microbiol Immunol 243:123 – 36.

Funk W.D.; Pak D.T.; Karas R.H.; Wright W.E.; Shay J.W. (1992). A Transcriptionally Active DNA binding Site for Human p53 Protein Complexes. Mol Cell Biol 12:2866-2871.

Funk W.D.; Wright W.E. (1992). Cyclic amplification and selection of targets for multicomponent complexes: myogenin interacts with factors recognizing binding sites for basic helix-loop-helix, nuclear factor 1, myocyte-specific enhancer-binding factor 2, and COMP1 factor. Proc Natl Acad Sci U S A 89:9484-8.

Ghosh S.; May M.J.; Kopp E.B. (1998). Evolutionarily Conserved Mediators of Immune Responses. Annu Rev Immunol 16:225-60.

Ghosh S.; Karin M. (2002). Missing Pieces in the NF-kB Puzzle. Cell 109:S81-S96.

Guo Q.; Lambowitz A.M. (1992). A tyrosyl-tRNA synthetase binds specifically to the group I intron catalytic core. Genes Dev 6:1357 – 1372.

Harkey, C. Sample prep for automated sequencing. http://www.icmb.utexas.edu/core/DNAFacility/Sequencing%20Handout.pdf (accessed Feb 2003).

Hesselberth J.; Robertson M.P.; Jhaveri S.; Ellington A.D. (2000). *In vitro* selection of nucleic acids for diagnostic applications. J Biotechnol 74:15-25.

Hicke B.J.; Marion C.; Chang Y.F.; Gould T.; Lynott C.K.; Parma D.; Schmidt P.G.; Warren S. (2001). Tenascin-C aptamers are generated using tumor cells and purified protein. J Biol Chem 276:48644-54.

Hirao I.; Spingola M.; Peabody D.; Ellington A.D. (1998). The limits of specificity: an experimental analysis with RNA aptamers to MS2 coat protein variants. Mol Divers 4:75-89.

Homann M.; Goringer H.U. (1999). Combinatorial selection of high affinity RNA ligands to live African trypanosomes. Nucleic Acids Res 27:2006-14.

Horwitz M.S.; Loeb L.A. (1986). Promoters selected from random DNA sequences. Proc Natl Acad Sci U S A 83:7405-9.

Huang J.; Blackwell T.K.; Kedes L.; Weintraub H. (1996). Differences between MyoD DNA binding and activation site requirements revealed by functional random sequence selection. Mol Cell Biol 16:3893-900.

King D.J.; Bassett S.E.; Li X.; Fennewald S.A.; Herzog N.K.; Luxon B.A.; Shope R.; Gorenstein D.G. (2002). Combinatorial Selection and Binding of Phosphorothioate Aptamers Targeting Human NF-κB RelA(p65) and p50. Biochemistry 41:9696-9706.

Kunsch C.; Ruben S.M.; Rosen C.A. (1992). Selection of optimal kappa B/Rel DNA-binding motifs: interaction of both subunits of NF-kappa B with DNA is required for transcriptional activation. Mol Cell Biol 12:4412-21.

Levy M.; Ellington A.D. (2001). Selection of deoxyribozyme ligases that catalyze the formation of an unnatural internucleotide linkage. Bioorg Med Chem 9:2581-7.

Liao D.; Dennis, P.P. (1992). The organization and expression of essential transcription translation component genes in the extremely thermophilic eubacterium Thermotoga maritima. J Biol Chem 267:22787 – 22797.

Liu R.; Blackwell T.W.; States D.J. (2001). Conformational model for binding site recognition by the E.coli MetJ transcription factor. Bioinformatics 17:622-33.

Lorger M.; Engstler M.; Homann M.; Goringer H.U. (2003). Targeting the variable surface of African trypanosomes with variant surface glycoprotein-specific, serum-stable RNA aptamers. Eukaryot Cell 2:84-94.

Marcotte E.M.; Xenarios I.; Eisenberg D. (2001). Mining literature for protein-protein interactions. Bioinformatics 17:359-63.

Marshall K.A.; Ellington A.D. (1999). Molecular parasites that evolve longer genomes. J Mol Evol 49:656 – 663.

Marshall K.A.; Ellington A.D. (2000). *In vitro* selection of RNA aptamers. Methods Enzymol 318:193 – 214.

Medici N.; Abbondanza C.; Nigro V.; Rossi V.; Piluso G.; Belsito A.; Gallo L.; Roscigno A.; Bontempo P.; Puca A.A.; Molinari A.M.; Moncharmont B.; Puca G.A. (1999). Identification of a DNA Binding Protein Cooperating with Estrogen Receptor as RIZ (Retinoblastoma Interacting Zinc Finger Protein). Biochem Biophys Res Commun 264:983-989.

Morris K.N.; Jensen K.B.; Julin C.M.; Weil M.; Gold L. (1998). High affinity ligands from *in vitro* selection: complex targets. Proc Natl Acad Sci U S A 95:2902-7.

Muller C.W.; Rey F.A.; Sodeoka M.; Verdine G.L.; Harrison S.C. (1995). Structure of the NF-kB p50 homodimer bound to DNA. Nature 373:311-317.

Newman J.R.; Keating A.E. (2003). Comprehensive identification of human bZIP interactions with coiled-coil arrays. Science 300:2097-101.

Opperman T.; Richardson J.P. (1994). Phylogenetic analysis of sequences from diverse bacteria with homology to the Escherichia coli rho gene. J Bacteriol 176:5033 – 5043.

Osborne S.E.; Ellington A.D. (1997). Nucleic acid selection and the challenge of combinatorial chemistry. Chem Rev 97:349 – 370.

Pan W.; Craven R.C.; Qiu Q.; Wilson C.B.; Wills J.W.; Golovine S.; Wang J.F. (1995). Isolation of virus-neutralizing RNAs from a large pool of random sequences. Proc Natl Acad Sci U S A 92:11509-13.

Pollard J.; Bell S.D.; Ellington A.D. (2000). Design, synthesis, and amplification of DNA pools for *in vitro* selection. In: Current Protocols in Nucleic Acid Chemistry, S.L. Beaucage et al., eds. (New York, USA: John Wiley & Sons, Inc.), pp. 9.2.1 – 9.2.23.

Pollock R.; Treisman R. (1990). A sensitive method for the determination of protein-DNA binding specificities. Nucleic Acids Res 18:6197-204.

Robison K.; McGuire A.M.; Church G.M. (1998). A comprehensive library of DNA-binding site matrices for 55 proteins applied to the complete Escherichia coli K12 genome. Journal of Molecular Biology 284:241-254.

Roller R.J.; Monk L.L.; Stuart D.; Roizman B. (1996). Structure and function in the herpes simplex virus 1 RNA-binding protein U(s)11: mapping of the domain required for ribosomal and nucleolar association and RNA binding *in vitro*. J Virol 70:2842 – 2851.

Roulet E.; Busso S.; Camargo A.A.; Simpson A.J.; Mermod N.; Bucher P. (2002). High-throughput SELEX SAGE method for quantitative modeling of transcription-factor binding sites. Nat Biotechnol 20:831-5.

Rye P.D.; Nustad K. (2001). Immunomagnetic DNA aptamer assay. Biotechniques 30:290-2, 294-5.

Salgado H.; Gama-Castro S.; Martinez-Antonio A.; Diaz-Peredo E.; Sanchez-Solano F.; Peralta-Gil M.; Garcia-Alonso D.; Jimenez-Jacinto V.; Santos-Zavaleta A.; Bonavides-Martinez C.; Collado-Vides J. (2004). RegulonDB (version 4.0): transcriptional regulation, operon organization and growth conditions in Escherichia coli K-12. Nucleic Acids Res. 32:D303-6.

Schneider D.; Gold L.; and Platt T. (1993). Selective enrichment of RNA species for tight binding to Escherichia coli rho factor. FASEB J 7:201 – 207.

Sinha S.; Tompa M. (2002). Discovery of novel transcription factor binding sites by statistical overrepresentation. Nucleic Acids Res 30:5549-60.

Sinha S.; Tompa M. (2003). YMF: A program for discovery of novel transcription factor binding sites by statistical overrepresentation. Nucleic Acids Res 31:3586-8.

Sooter L.J.; Ellington A.D. (2004). Automated selection of transcription factor binding sites. JALA 9:277-284.

Sooter L.J.; Riedel T.; Davidson E.A.; Levy M.; Cox J.C.; Ellington A.D. (2001). Toward automated nucleic acid enzyme selection. Biol Chem 382:1327-34.

Soukup G.A.; Breaker R.R. (2000). Allosteric nucleic acid catalysts. Curr Opin Struct Biol 10:318 – 325.

Thiesen H.J.; Bach C. (1990). Target Detection Assay (TDA): a versatile procedure to determine DNA binding sites as demonstrated on SP1 protein. Nucleic Acids Res 18:3203-9.

Tsai D.E.; Kenan D.J.; Keene J.D. (1992). *In vitro* selection of an RNA epitope immunologically cross-reactive with a peptide. Proc Natl Acad Sci U S A 89:8864-8.

Tucker C.E.; Chen L.S.; Judkins M.B.; Farmer J.A.; Gill S.C.; Drolet D.W. (1999). Detection and plasma pharmacokinetics of an anti-vascular endothelial growth factor oligonucleotide-aptamer (NX1838) in rhesus monkeys. J Chromatogr B Biomed Sci Appl 732:203-12.

Tuerk C.; Gold L. (1990). Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. Science 249:505-10.

Tuerk C.; MacDougal S.; Gold L. (1992). RNA pseudoknots that inhibit human immunodeficiency virus type 1 reverse transcriptase. Proc Natl Acad Sci U S A 89:6988-92.

Wright W.E.; Binder M.; and Funk W. (1991). Cyclic amplification and selection of targets (CASTing) for the myogenin consensus binding site. Mol Cell Biol 11:4104-10.

Wright W.E.; Funk W.D. (1993). CASTing for multicomponent DNA-binding complexes. Trends Biochem Sci 18:77-80.

http://web.mit.edu/esgbio/www/lm/nucleicacids/dna.html

http://www.beckman.com/products/instrument/automatedsolutions/biomek/biomekfx_inst_dcr.asp

http://www.beckman.com/products/instrument/automatedsolutions/biomek/biomek2000_inst_dcr.asp

http://www.pdb.org

http://www.tecan-us.com/us-index/com-pr-in/com-pr-in-ro_li_entry-3/com-pr-in-ro_li-genesis_rsp.htm

Vita

Letha Jane Sooter was born on July 29, 1977 in Ponca City, Oklahoma to Jean and Matt Sooter. She has one younger brother, Ben. Letha attended Elementary school in Ponca City, Oklahoma, Grades 5-7 at Memorial Parkway Junior High in Katy, Texas, Grades 8-10 at the American School in Japan in Chofu, Japan, and Grades 11-12 at the International School of Kuala Lumpur in Kuala Lumpur, Malaysia. She taught English as a second language in both Japan and Malaysia. She received her B.S. degree from Texas A&M University in College Station, Texas with a double major in biochemistry and genetics and a chemistry minor. There she performed research with Dr. Donald Pettigrew for two years and with Dr. Sherry Yennello for one year. She later worked in Dr. Andy Ellington's lab where she received her Ph.D. in biochemistry from the University of Texas at Austin in Austin, Texas.

Sooter, L.J.; Ellington, A.D. Automated Selection of Transcription Factor Binding Sites. Journal of the Association for Laboratory Automation. 2004.

Sooter, L.J.; Ellington, A.D. Reflections on a Novel Therapeutic Candidate. Chemistry & Biology. 2002, 9(8), 857-858.

Cox, J.C.; Rajendran, M.; Riedel, T.; Davidson, E.A.; Sooter, L.J.; Bayer, T.S.; Schmitz-Brown, M.; Ellington, A.D. Automated acquisition of aptamer sequences. Combinatorial Chemistry and High Throughput Screening. 2002, 5(4), 289-299.

Sooter, L.J.; Riedel, T.; Davidson, E.A.; Levy, M.; Cox, J.C.; Ellington, A.D. Toward automated nucleic acid enzyme selection. Biological Chemistry. 2001, 382(9), 1327-1334.

Ellington, A.D.; Hesselberth, J.; Marshall, K.; Robertson, M.; Sooter, L.; Davidson, E.; Cox, J.C.; Reidel, T.; Wilson, C.; Cload, S.T.; Keefe, A.D. Regulatable, catalytically active nucleic acids. PCT Int. Appl. 2002, 130 pp.

Permanent address: 91 Sidney St. #403, Cambridge, MA 02139

This dissertation was typed by the author.