PROTEIN-ASSISTED TARGETING OF GENES IN YEAST AND HUMAN CELLS

A Dissertation
Presented To
The Academic Faculty

By

Patrick Thomas Ruff

In Partial Fulfillment
Of the Requirements for the Degree
Doctor of Philosophy in Biology

Georgia Institute of Technology

August 2013

PROTEIN-ASSISTED TARGETING OF GENES IN YEAST AND HUMAN CELLS

Approved by:

Dr. Francesca Storici, Advisor
School of Biology
Georgia Institute of Technology

Dr. Raquel Lieberman
School of Chemistry/Biochemistry
Georgia Institute of Technology

Dr. Yury Chernoff
School of Biology
Georgia Institute of Technology

Dr. Yuhong Fan
School of Biology
Georgia Institute of Technology

Dr. Kirill Lobachev
School of Biology
Georgia Institute of Technology

Date approved: June 4th, 2013

For my family and friends

ACKNOWLEDGMENTS

I would like to acknowledge my parents, my siblings, and all of my family members for their moral support and for asking “So when are you going to graduate?” many, many times. I would also like to thank Dr. Francesca Storici for giving me the opportunity to pursue a career in Molecular Biology. I often think that Francesca saved me from a life of staring at a computer all day and gave me a chance to be a true “wet lab” scientist. She is one of the most generous people when it comes to new students, so many of the people in our lab switched from Master’s to PhD because of her encouragement. When I first started I had the choice of joining Francesca in the Fall or to wait and join King Jordan’s lab (a Bioinformatics lab) in the Spring. I would like to thank King for being a great person and a constant friend throughout my grad school years. It’s interesting to note that my good friend Daudi Jjingo is graduating at the same time I am but he came a year after I did (he started in King’s lab ☺). I do not regret my decision but I am sometimes curious as to what might have been. I would like to acknowledge Dr. Yuhong Fan for her continued support and insightful questions at each of my meetings. I would like to acknowledge Dr. Raquel Lieberman for her support and for asking many tough but good questions at my oral exam. I would also like to thank Dr. Yury Chernoff and Dr. Kirill Lobachev for being outstanding scientists in the lab and decent soccer players outside the lab. Yury has an amazing ability to recall in great detail both scientific and historical data. Kirill and his wife, Natasha Degtyareva, have been inspirational in their passion for science and just generally wonderful people both inside and outside of the lab.

I would like to thank all of the people behind-the-scenes that made this research possible. Marc Pline particularly for being incredibly patient with me and the lab budgets! Angie Lessard for helping all of the labs I've ever TAed to run smoothly (along with Marc they make the labs possible). I'd like to thank Frank Canella for always fixing everything not computer-related (with the help of Kendall Carey) and Troy Hilley for fixing all things computer-related (and for loaning me a laptop and always joking and for many other things). I would like to thank Nadia Boguslavsky for effectively running the IBB Core Facilities and allowing me access to using the capillary electrophoresis (CE) machine as much as I wanted to. I'd also like to thank Kevin Roman for being one of the best people to interact with whenever I had a graduate school question. Basically, Kevin has a mountain of paperwork shoved at him on a daily basis and yet he's able to remain calm and work through it. He's never let me down whenever I've come to him for anything. There were several graduate coordinators during my stay here, but he's the best by far.

There is no way I could thank everyone related to the research, but I am fairly confident I would not have made it this far without the help of Dr. Samantha Stuckey. Ying Shen, Samantha, and I were the first Storici lab members and Samantha not only graduated recently (Spring 2013) but she also helped every single person in our lab in some minor or major way. She was incredibly thoughtful and always went out of her way to help others, even if it meant that she sacrificed her own time. On that note, I'd like to thank her husband, Lee Katz, who was one of the first people I met at GA Tech and who helped to inspire me. I would like to thank all of my lab members including Kyung-Duk (K.D.) Koh, Havva Keskin, Lin Zhiqiang, Sathya (Sat) Balachander, Taehwan Yang, who have

all helped me in some way or another. I would also like to thank Dr. Rekha Pai who, along with Francesca, was a great mentor to me and taught me many basic lab techniques including dialyzation and Western blotting. Rekha was absolutely essential in the BSA aptamer work and I am very grateful for her. I would also like to thank Dr. Kuntal Mukherjee for being a great friend and mentor to me during my time here. I'd like to thank everyone in my class (who started in 2007) and most of the people the year after I started and the year before I started. Some names I can recall right now are Zeng Jia, Xu Ke, Khairat Elbaradie, Vinay Mittal, Pan Chenyi, Wang Jianrong, Andy Conley, Ho Po-Yi, Zhang Yu, Natalie Saini, Sunyoung Goo, Brian Ondov, Steve Heitner, Seng, Gabriel Mitchell, Lava, Daudi Jjingo, Eishita Tyagi, Pan Minmin, Taylor Updegrove, Vidhya Narayanan, Gaurav Arora, Anju Ondov, Bee, and many many more. There are lots of people that stand out but I won't name them all (you know who you are!).

Additionally, I would like to thank my girlfriend Wang Qing, or in English Jenny Wang. She did not break up with me even though I decided to extend my graduate schooling for an unknown number of years by switching from the Master's Bioinformatics program to the Biology PhD program. Also she has been a constant driving force encouraging me to work hard and persevere. Without her sympathetic ear I probably would have went insane with frustration after years of failed experiments. I also would like to thank God, who must exist or I must just be very lucky, because one of the two has to be true for my project, which had so many chances for failure (and for a long time it was a failure) to eventually work out. Lastly, I would like to thank the reader of this thesis, for whatever

your reason may be for reading it, you have added value (or “impact”) to this work by spending your time, which is valuable, reading it. Thank you.

TABLE OF CONTENTS

LIST OF TABLES

Page

LIST OF FIGURES

LIST OF SYMBOLS AND ABBREVIATIONS

5-FOA	5-fluoroorotic acid
AAV	adeno-associated virus
ADA	adenosine deaminase
BIR	break-induced replication
bp	base pair
°C	degrees Celsius
Cas	CRISPR associated
CBA	chicken β actin
cDNA	complementary DNA
CE	capillary electrophoresis
cm	centimeter
CMV	cytomegalovirus
CO_2	carbon dioxide
CORE	COunterselectable REporter
CRISPR	clustered regularly interspaced short palindromic repeat
DMEM	Dulbecco's modified Eagle's medium
DNA	deoxyribonucleic acid
dNTP	deoxynucleotide triphosphate
DSB	double-strand break
DSBR	double-strand break repair

EDTA	ethylenediaminetetraacetic acid
EtOH	ethanol
FACS	fluorescence-activated cell sorting
FBS	fetal bovine serum
g	gram
g	gravity
gRNA	guide RNA
GCR	gross chromosomal repeat
GFP	green fluorescent protein
h	hour
H_3BO_3	boric acid
HEK-293	human embryonic kidney
HR	homologous recombination
in/dels	insertions/deletions
kb	kilobase
l	liter
LiOAc	lithium acetate
M	molar
mg	milligram
Mg	magnesium
Mg^{2+}	magnesium ions
min	minute
ml	milliliter

mM	millimolar
MMR	mismatch repair
Mn	manganese
MRX	*MRE11*/*RAD50*/*XRS2*
mtDNA	mitochondrial DNA
NaOAc	sodium acetate
$(NH_4)_2SO_4$	ammonium sulfate
NHEJ	non-homologous end-joining
nmole	nanomole
nt	nucleotide
oligo	oligonucleotide
O/N	overnight
PARP	poly(ADP-ribose) polymerase
PCR	polymerase chain reaction
PEG	polyethylene glycol
PEI	polyethyleneimine
pmol	picomole
*RAD51*p	*RAD51* protein
*RAD52*p	*RAD52* protein
RFP	red fluorescent protein
RNA	ribonucleic acid
ROS	reactive oxygen species
rpm	revolutions per minute

s	second
SCID	severe combined immunodeficiency
SCID-X	x-linked SCID
SC-Leu	synthetic complete medium lacking leucine
SC-Trp	synthetic complete medium lacking tryptophan
SC-Ura	synthetic complete medium lacking uracil
SD-complete	synthetic dextrose complete
SDSA	synthesis-dependent strand annealing
SELEX	systematic evolution of ligands by exponential enrichment
ss	single-stranded
SSA	single-strand annealing
SSB	single-strand break
SSD	salmon sperm DNA
ssDNA	single-stranded DNA
TALEN	transcription activator-like effector nuclease
TBE	tris/borate/EDTA
TE	tris/EDTA
U	unit
UV	ultraviolet
V(D)J	variable, diverse, joining
wt	wild-type
YPD	yeast extract/peptone/dextrose
YPG	yeast extract/peptone/glycerol

YPGal	yeast extract/peptone/galactose
YPLac	yeast extract/peptone/lactic acid
ZFN	zinc finger nuclease
α	alpha
β	beta
Φ	phi
µl	microliter
µM	micromolar

SUMMARY

This work was designed as a proof-of-principle concept or prototype to show the effect of protein-assisted targeting of DNA to specific genomic loci. Two strategies were employed to deliver the DNA with the aim that once inside the cell the DNA would be delivered to the target sequence by the assistance of a protein. In our case, the chosen protein was the site-specific meganuclease I-*Sce*I. The first strategy described herein was to bind the targeting DNA to I-*Sce*I by the use of a fusion protein between I-*Sce*I and a known DNA-binding domain, the *GAL4*-DBD. The second strategy involved using a DNA aptamer to I-*Sce*I to link the targeting DNA and I-*Sce*I. Testing *in vivo* revealed that in our human cells (HEK-293) single-stranded DNA was more efficient at gene targeting than double-stranded DNA. In order for the first strategy to work, we needed to have some region of double-stranded DNA. We found that in human cells, it was better for gene targeting to have that double-stranded DNA on the 5' side of our targeting DNA. We also used gel shift assays to confirm binding by our candidate DNA-binding domain, the *GAL4*-DBD. We were unable to detect expression of the fusion protein of I-*Sce*I and the *GAL4*-DBD.

For the second strategy we were able to construct an aptamer to I-*Sce*I using a variant of the systematic evolution of ligands by exponential enrichment (SELEX). The I-*Sce*I aptamer was synthesized as part of a longer DNA molecule containing homology to a target locus. Using this chimeric oligonucleotide (part aptamer, part DNA repair region) testing was done in both yeast and human cells. Aside from instances where the

aptamer's secondary structure may have been compromised, the aptamer containing oligonucleotide stimulated repair at a rate 2 to 15-fold higher than the non-selected control sequence. These experimental results show that by delivering targeting DNA within close proximity to the site of modification, gene targeting frequencies can be increased.

CHAPTER 1

INTRODUCTION

1.1 DOUBLE-STRAND BREAK (DSB) REPAIR AND GENE TARGETING

A DNA double-strand break (DSB) is generally regarded as the most severe DNA damage. DSBs can be caused by a variety of exogenous sources including ionizing radiation (IR), radiomimetic chemicals, and ultraviolet light, as well endogenous sources like collapsed replication forks or programmed endonucleolytic cleavage by programmed cellular endonucleases during meiosis or V(D)J recombination (Chapman, Taylor et al. 2012). A DSB if left unrepaired can be a lethal cellular event. There are two main repair pathways used by the cell to correct DSBs, specifically homologous recombination (HR) and non-homologous end-joining (NHEJ). HR involves resection of the two 5' DNA ends generated by the DSB, a search for homologous DNA by the 3' overhanging DNA tails, annealing to homologous DNA, and repair of the DSB (Aylon and Kupiec 2004). Also there is a variant of HR known as single-strand annealing (SSA). SSA is a special case of DSB repair in which a DSB occurs between two repeated DNA sequences oriented in the same direction (Ivanov, Sugawara et al. 1996). The two 5' DNA ends generated after the DSB are resected in a 5' to 3' manner and the remaining two single-stranded 3' tails find homology with each other at the repeated sequence. The ssDNA sequences anneal, the unannealed tails are removed, and the resulting repair causes a deletion since the two repeats become one and the sequence between them is lost. The other pathways of HR

include break-induced replication (BIR), synthesis-dependent strand annealing (SDSA), or homologous recombination involving the formation of two Holliday junctions known as double Holliday junction (dHJ) (Heyer, Ehmsen et al. 2010). Each of these repair pathways is different but they are all similar in the fact that after the DSB is generated in order for these repair pathways to function, the DNA ends must be resected in a 5' to 3' manner. Likewise, in each of these pathways the 3' ssDNA tails then invade a homologous template, displacing its DNA and forming a D loop (the name for the structure after displacement), and then the homologous template is used to synthesize through the gap that was generated by the DSB.

The other main pathway for repairing DSBs involves NHEJ, which repairs the DSB by religating the free DNA ends generated from the break. The difference between NHEJ and HR is that NHEJ relies on very little (in the case of microhomology-mediated end joining) or no homology while HR requires a homologous template. Although simple religation can lead to accurate repair by NHEJ, the majority of DSBs in cells arise from DNA damage in the form of ionizing radiation, reactive oxygen species (ROS), and chemical agents that typically do not leave behind easily ligatable products, but rather "dirty ends," which must undergo processing before NHEJ can occur (Woodbine, Brunton et al. 2011). In this capacity, NHEJ often leads to misalignment of the DNA ends, deletion of genetic information, or the insertion of DNA fragments at the break site.

Gene targeting is the *in situ* modification of a specific genomic locus by HR through the use of exogenously introduced DNA as a template for repair. By using DNA with

homology to the targeted DNA sequence at or near the DSB, the DSB can be repaired. Additionally, by modifying the exogenous DNA to have non-native sequence information flanked by regions of homology to the DSB, HR repair can actually introduce new DNA sequences or conversely remove pre-existing DNA sequences. Thus, gene targeting can be used to insert or delete sequences of DNA at a specific genomic locus.

The alternative to gene targeting is known as random integration or illegitimate recombination, whereby the exogenously introduced DNA integrates at a random genomic locus. Random integration occurs due to NHEJ between the exogenously introduced DNA and a DSB somewhere in the endogenous genomic DNA (Iiizumi, Kurosawa et al. 2008). The exogenous DNA is ligated in between the ends of the DSB, leading to integration of the exogenous DNA while at the same time disrupting the native DNA sequence.

1.2 THE ADVANTAGES OF GENE TARGETING OVER RANDOM INTEGRATION

As stated, gene targeting and random integration are two outcomes when exogenous DNA is introduced into the cell. Several examples of "gene augmentation" exist which employ random integration to introduce an exogenous gene into the human genome for a therapeutic outcome. Severe combined immunodeficiency (SCID) was the first disease to be treated with gene therapy, specifically gene augmentation, in the early 90s (Blaese, Culver et al. 1995, Bordignon, Notarangelo et al. 1995, Kohn, Weinberg et al. 1995). In adenosine deaminase deficient SCID (ADA- SCID), patients lacking the ADA gene were

given a good copy of the ADA gene. Since then, gene augmentation has been used to treat not only ADA- SCID (Aiuti, Slavin et al. 2002), but X-linked SCID (Cavazzana-Calvo, Hacein-Bey et al. 2000), epidermolysis bullosa (EB) (Mavilio, Pellegrini et al. 2006), β thalassaemia (Sadelain 2006), and many other genetic disorders (Ginn, Alexander et al. 2013).

Although not as widely used as gene augmentation strategies, gene targeting has several advantages over random integration. As can be derived from its name, random integration can lead to unforeseen consequences including *cis* or *trans*-activation of a gene close to where the transgene was inserted, disruption of regulatory elements, or the creation of aberrant fusion proteins all of which could lead to diseases such as cancer. These kinds of insertional mutagenesis, although rare, present a real risk for gene therapy patients. In 2002 two patients, followed by one patient in 2006 and one patient in 2007, in X-linked SCID trials in France and London were shown to have developed a leukemia-like disorder after the insertional mutagenesis of the therapeutic vector which activated an oncogene (Hacein-Bey-Abina, Von Kalle et al. 2003). Aside from the risk associated with gene augmentation, from a therapeutic aspect there are many genetic disorders caused by dominant negative mutations (Huntington's, myotonic dystrophy, and brittle bone disease to name a few) where the mutated gene product has an antagonistic effect on the wild-type protein. In these disorders, even successful integration of a transgene will do nothing to alter the disease. Likewise, random integration of a transgene cannot generate a gene deletion or single base substitutions which are important for studying the

function of a particular gene. Gene targeting is logically the best choice for gene therapy and functional studies because it proceeds in a controlled manner, modifying only the DNA of interest, and not causing unnatural expression in surrounding genes.

1.3 STIMULATING GENE TARGETING IN MAMMALIAN CELLS

As stated, gene targeting has many advantages over random integration and in human cells would be a powerful tool for researchers interested in functional analysis of genes as well as patients suffering from genetic disorders. The primary limitation of gene targeting is the low frequency with which it spontaneously occurs in mammalian cells, happening in roughly 1 cell for every 10^5 to 10^7 treated cells (Vasquez, Marburger et al. 2001). The low frequency of gene targeting, which relies on homologous recombination (HR), is due in part to the much higher frequency of non-homologous end joining (NHEJ), which occurs in roughly 1 cell for every 10^2 to 10^4 treated cells (Vasquez, Marburger et al. 2001).

Currently, there are several strategies for increasing gene targeting in mammalian cells. Stimulation of gene targeting by generating a DSB at the target site increased the frequency of gene targeting several orders of magnitude in bacteria (Nussbaum, Shalit et al. 1992), yeast (Storici, Durham et al. 2003), plants (Puchta, Dujon et al. 1993), fruit flies (Banga and Boyd 1992), mice (Rouet, Smih et al. 1994), human embryonic stem cells (Smih, Rouet et al. 1995), and many other cell types. Another strategy to increase gene targeting in mammalian cells has been achieved through the over-expression of key recombination proteins from HR proficient organisms. Over-expression of bacterial RecA

led to a 10-fold increase in gene targeting in mouse cells (Shcherbakova, Lanzov et al. 2000), likewise over-expression of yeast Rad52 led to a 37-fold increase in gene targeting in human cells (Di Primio, Galli et al. 2005). Conversely, another strategy for increasing gene targeting in human cells involves decreasing the amount of the DSB repair through the pathway of NHEJ. In mouse embryonic stem cells an increase in gene targeting was seen in Ku70 (6-fold), XRCC4 (2-fold), and DNAPK-cs (2-fold) deficient cell lines (Pierce, Hu et al. 2001) and a 3-fold increase in Chinese hamster ovary cells lacking DNAPK-cs (Allen, Kurimasa et al. 2002). Similarly, knockdown of KU70 and XRCC4 in human colon cancer cells lead to a 30-fold increase in gene targeting (Bertolini, Bertolini et al. 2009).

Another strategy for increasing gene targeting not focused on increasing HR or decreasing NHEJ was developed whereby knocking down human SMC1, important for a certain type of HR, gene targeting increased (Potts, Porteus et al. 2006). The sister chromatid is the normal donor DNA for HR repair, but in the case of gene targeting an exogenous DNA acts as the donor for repair. HR with the sister chromatid actually hinders gene targeting by exogenously introduced DNA. By knocking down hSMC1 which is required for sister chromatid HR, gene targeting increased four-fold. The proteins hSMC1 and hSMC3 form the cohesin complex which is responsible for keeping sister chromatids in close proximity to each other during a DSB. Without close proximity to the DSB site the sister chromatid was used less frequently as a homologous donor,

shifting repair of the DSB more toward HR with the exogenous sequence (gene targeting).

1.4 SITE-SPECIFIC ENDONUCLEASES

As previously discussed, a targeted DSB stimulates gene targeting several orders of magnitude. There are two main methods to generate a targeted DSB. The first characterized method for generating a site-specific DSB was to make use of naturally occurring "homing" endonucleases or meganucleases. Homing endonucleases are restriction enzymes which generate DNA double-strand breaks (DSBs) at defined genetic loci, and they have high specificity due to a long recognition sequence (12-40bp) (Belfort and Roberts 1997). Although homing endonucleases could stimulate gene targeting 1,000-fold or more there was an inherent limitation to their use for gene therapy or functional analysis. In order to increase gene targeting at a specific locus, that locus would need to contain the long recognition sequence of the homing endonuclease close to the desired area of modification. Obviously, a more modular approach was needed and zinc finger nucleases (ZFNs) fulfilled that role. ZFNs are chimeric proteins composed of a DNA-binding domain, a series of zinc finger motifs, fused to the non-specific nuclease domain of the FokI endonuclease. When two ZFNs dimerize, the FokI domains are able to create a DSB at the targeted site (Kim, Cha et al. 1996). ZFNs seemed to be a modular solution to the problem of site-specific DSBs, although binding by ZFNs was complicated. ZFN binding was theoretically modular because each "finger" domain recognized a specific three nucleotide base pairs. The problem was that this binding was context specific meaning that, for example, while a single finger might recognize the

sequence AAT alone, when fused with another finger it might recognize AAG. Due to the inability to predict ZFN context specific binding properties, empirical studies needed to be done for each locus to find the best ZFN pair. ZFNs could be engineered and selected for almost any sequence, but not in a straightforward or cost-effective manner.

Fortunately, an alternative to zinc fingers known as transcription activator-like effectors (TALEs) have been developed. Working in much the same way as a ZFN, a TALEN is composed of a single TALE protein fused to a FokI nuclease domain. The key difference between TALEs and ZFs is that instead of combining several ZFs that are context specific a single TALE can be constructed with an easily designed sequence specificity. The TALE consists of a repeat variable domain (RVD) which contains several (2 to 40) repeats of typically 34 amino acids, with the 12^{th} and 13^{th} amino acids being hypervariable. It was discovered that these 12^{th} and 13^{th} amino acids are responsible for recognizing a single base pair (Moscou and Bogdanove 2009). Following this TALE "code" one can efficiently design a site-specific TALE *in silico*, construct the TALEN, and then use it *in vivo* without the need for laborious empirical testing (Moscou and Bogdanove 2009).

Although ZFNs may be replaced by TALENs for the reasons stated, recently there have been reports of a new class of modular site-specific endonucleases called clustered regularly interspaced short palindromic repeats (CRISPRs) and their CRISPR associated (Cas) proteins (Cong, Ran et al. 2013). These CRISPR/Cas systems differ from ZFNs and TALENs in that they do not make use of the FokI nuclease and targeted recognition relies

on guide RNA (gRNA) (Cong, Ran et al. 2013). Specific RNA generated from the CRISPRs acts to guide a nuclease which creates a DNA DSB at the target bound by the gRNA (Mali, Yang et al. 2013).

1.5 THE HOMING ENDONUCLEASE I-SCEI

As stated previously, homing endonucleases are restriction enzymes which generate DNA double-strand breaks (DSBs) at defined genetic loci, and they have high specificity due to a long recognition sequence (12-40bp) (Belfort and Roberts 1997). The term homing comes from self-splicing group I or group II introns or inteins that "home" in on a specific sequence on the host gene, and splice themselves into the unoccupied allelic site by generating a break at the site which then homologously recombines with the intron-containing allele. Homing endonucleases have been studied since the late 1970s, and one of the first homing endonucleases studied was called "Omega" which later became known as I-*Sce*I, Intron (where it was found) – *Saccharomyces cerevisiae* (the genus and species) I (the first discovered) (Stoddard 2011). The I-*Sce*I endonuclease's natural function is to recognize a nonsymmetrical 18bp sequence of 5' TAG GGA TAA CAG GGT AAT 3' (Colleaux, D'Auriol et al. 1988) on the intron-less allele and generate a DNA double-strand break (DSB) at that location, propagating the intron containing allele and overwriting the previously intron-less allele through homologous recombination and gene conversion. I-*Sce*I binds to its recognition sequence very specifically as a monomer, but without the presence of Mn or Mg it will only bind, and not cleave the target (Beylot and Spassky 2001). I-*Sce*I is naturally found in the yeast mitochondria, but can generate DSBs in the nucleus if the recognition sequence is

present. I-*Sce*I enters the nucleus, but has no known NLS and is thought to enter the nucleus by passive diffusion (Plessis, Perrin et al. 1992) (Thermes, Grabher et al. 2002). Although it does not have a known NLS, I-*Sce*I is small (27 kDa) and basic in nature (pI = 10.08), and therefore should be able to freely diffuse through the nuclear pore (size limit of ~50kDa) (Lungwitz, Breunig et al. 2005). For this work I-*Sce*I was chosen as our protein of interest due to its ability to stimulate gene targeting and its ability to target a specific DNA sequence.

1.6 THE *GAL4* DNA-BINDING DOMAIN (DBD)

The 881 amino acid *GAL4* protein from *Saccharomyces cerevisiae* is one of the most well-characterized and studied transcription factors, serving as an activator of the galactose metabolizing enzyme genes *GAL*1, *GAL*7, and *GAL*10 (Oshima 1982) (Laughon and Gesteland 1982). The *GAL4* protein consists of an N-terminal DNA binding domain and dimerization domain (amino acids 1-74 or 1-147, with 1-147 showing better DNA binding) that binds DNA through two cadmium ions in a Cd2Cys6 helix-turn bimetal thiolate cluster (Baleja, Thanabal et al. 1997) as well as two transcription activating domains (amino acids 148-196 and 768-881) (Ma and Ptashne 1987). Since 1989, the DNA binding domain (DBD) and the activating domain (AD) have been used separately to study protein-protein interactions by creation of two fusion proteins, one of *GAL4*'s DBD (1-147) with a protein of interest, and another fusion protein made up of *GAL4*'s AD (768-881) the idea being that if two proteins of interest interacted, they could thus activate transcription (Fields and Song 1989). This technique came to be known as the 2 hybrid assay/system or the yeast 2 hybrid assay/system and

the components are commercially available.

In addition to its use in the yeast 2 hybrid assay, the *GAL4*-DBD has started to be used in non-viral gene delivery strategies. The *GAL4*-DBD binds to a 17bp consensus sequence of 5' CGGAGGACAGTCCTCCG 3' (with a general binding pattern of 5' cggrnnrcynyncnccg 3') (Baleja, Thanabal et al. 1997). The delivery strategies usually involve using the *GAL4*-DBD as a fusion protein (not surprising given the availability of fusion protein vectors) where the *GAL4*-DBD acts as the DNA binding portion and the other parts of the chimeric protein are involved in DNA transport into the cell and/or other functions such as endosome escape. As examples, the *GAL4*-DBD has been fused to a single chain antibody for delivery (Wels, Groner et al. 1996), invasin (Paul, Weisser et al. 1997), anthrax toxin (Gaur, Gupta et al. 2002), and more recently to short cell penetrating peptides (CPPs) like the Tat peptide from HIV (Xu, Chi et al. 2010). It was shown that by using the *GAL4*-DBD and a plasmid containing two recognition sequences there was an enhancement of even calcium phosphate transfection, however in the same paper they showed that the nuclear localization sequence (NLS) of the *GAL4*-DBD is mutually exclusive of its DNA binding (Chan, Hubner et al. 1998). The authors reasoned that the *GAL4*-DBD enhanced the transfection by either DNA compaction or by protecting DNA from degradation once inside the cell, but not by its NLS. Additionally, in the same year, there was developed a fusion protein between the *GAL4*-DBD and the endonuclease *Fok*I, generating a site-specific endonuclease in the process (Kim, Smith et al. 1998). The *GAL4*-DBD was chosen for our research because it is small (17 kDa),

highly specific (low nM Kd), and well-characterized (Taylor, Workman et al. 1991). In order to achieve our strategy of protein-assisted targeting we explored using a *GAL4*-DBD I-*Sce*I fusion protein.

1.7 APTAMERS AND APTAMER SELECTION BY SYSTEMATIC EVOLUTION OF LIGANDS BY EXPONENTIAL ENRICHMENT (SELEX)

Nucleic acid aptamers are short single-stranded DNA or RNA oligonucleotides that are capable of binding a ligand (protein, small molecule, or even living cells) with high affinity. They are also known as artificial antibodies because in addition to binding with high affinity, they also bind with high specificity. As an example, aptamers were selected for the small molecule theophylline and tested to see if they reacted with caffeine (another small molecule that differs from theophylline by only a single methyl group) and not only did the aptamer have strong binding with a 100nM Kd, but the aptamer bound theophylline 10,000 times stronger than it did caffeine (**Figure 1.1**) (Jenison, Gill et al. 1994). Aptamers have several advantages over antibodies, including ease and low cost of production which does not involve animals. Aptamers are less immunogenic than antibodies and are already being used as a therapeutic for human (Singerman, Masonson et al. 2008). A recent market report predicts that the aptamer market may be worth $1.8 billion by 2014 (Jackson 2010). Aptamers are "evolved" from random sequences of DNA/RNA by a process known as systematic evolution of ligands by exponential enrichment (SELEX). The SELEX procedure involves the use of the random library of DNA/RNA sequences being incubated with the target, followed by a partitioning step to remove unbound sequences, then followed by an elution step to recover the binding

sequences, and then followed by an amplification step to generate a library of sequences enriched for binding (**Figure 1.2**). Using capillary electrophoresis (CE) allows for SELEX to be performed in a much shorter amount of time due to much more efficient partitioning and the obfuscation of aptamers binding to the ligand support (the ligand flows freely in buffer). In as little as one round of selection (Berezovski, Drabovich et al. 2005), and almost always less than five, strong binding highly specific aptamers may be selected, as opposed to traditional SELEX which typically takes 10 or more rounds of selection. CE-SELEX generated aptamers have nM and sometimes pM level disassociation constants (Mosing, Mendonsa et al. 2005). CE-SELEX is a new technology first developed for use in 2004 and has yet to become commonly used (Mendonsa and Bowser 2004). In addition to strong binding aptamers in few rounds of selection, CE-SELEX typically generates a greater variety of unique aptamers compared to traditional SELEX. Aptamers are a versatile and interesting newly developing technology that we planned to use for our protein-assisted targeting strategy.

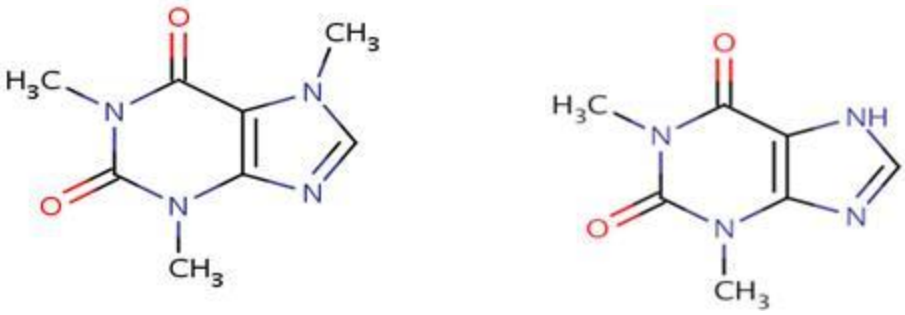


Figure 1.1 Comparison of theophylline and caffeine structures. Shown above are the chemical structures for caffeine and theophylline, which differ only in a single methyl group. An aptamer was selected to bind to theophylline which showed an approximately 320 nm K_d for theophylline but had a 3.5 mM K_d for caffeine (approximately 10,000 fold difference in binding affinity).

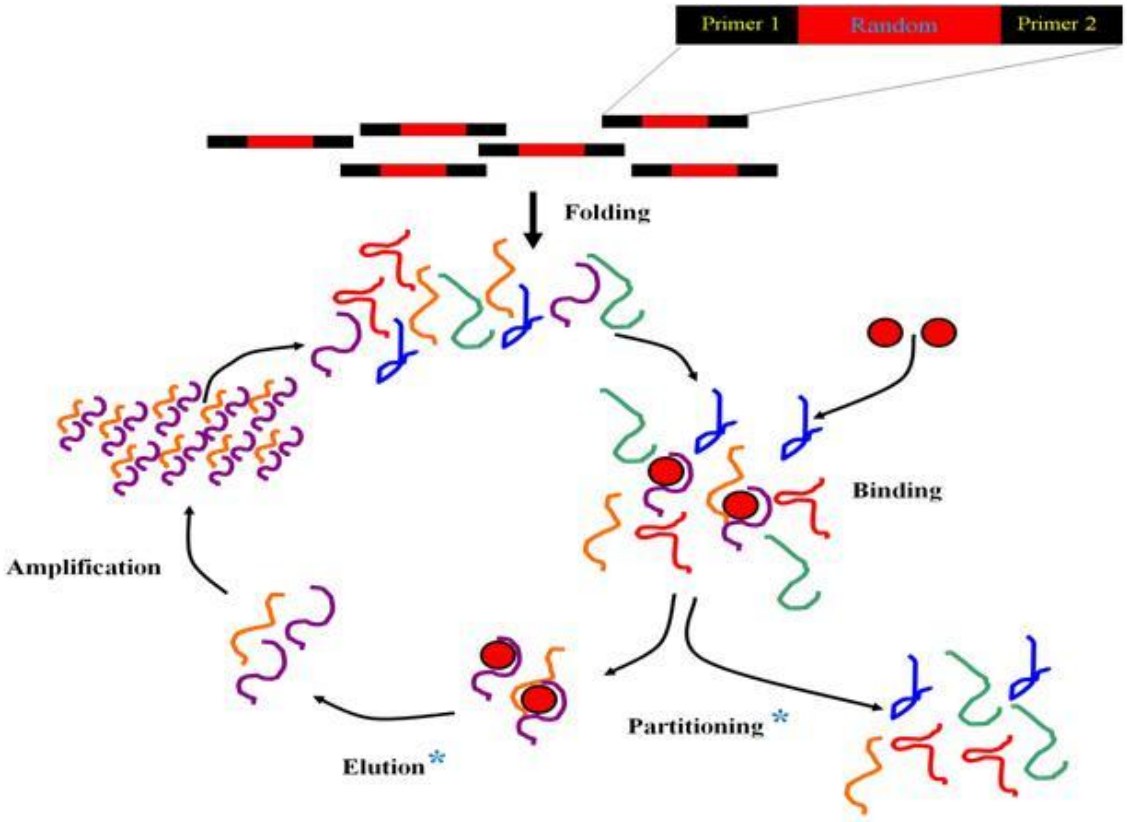


Figure 1.2 The Selective Evolution of Ligands by EXponential enrichment (SELEX) procedure. The ligand of interest is incubated with a random library of DNA. Sequences that form a suitable secondary structure bind to the ligand. These sequences are partitioned from the weaker binders and are eluted. After amplification of these binders, the process is repeated.

1.8 RESEARCH GOALS

In order to assess the efficacy of our protein-assisted targeting system, three separate aims were devised.

1.8.1 Construction and testing of a fusion protein between the I-SceI endonuclease and the GAL4-DNA binding domain (DBD)

The essential part of protein-assisted targeting is the linking of the targeting DNA used to modify the endogenous DNA with our protein of interest, I-*Sce*I. The first strategy for linking our targeting DNA to I-*Sce*I involved a fusion protein between I-*Sce*I and the *GAL4*-DBD. The *GAL4*-DBD is a specific DNA binding domain that recognizes a 17 bp sequence. Construction of fusion proteins along with *in vivo* testing of these constructs and different oligonucleotides designed to bind to the *GAL4*-DBD was done.

1.8.2 Selection of a DNA aptamer using CE-SELEX

One of the two strategies for protein-assisted targeting involved using a DNA aptamer to I-*Sce*I. A modification on SELEX using capillary electrophoresis (CE) was performed. The goal was to use a different protein, bovine serum albumin (BSA), and select aptamers against it. BSA was used as a proof-of-principle for aptamer selection using CE-SELEX.

1.8.3 Selection of an aptamer for I-*Sce*I *and in vivo testing*

As stated, an I-*Sce*I induced DSB stimulates gene targeting several orders of magnitude. Our goal was to generate an I-*Sce*I aptamer so that we could create a bifunctional DNA

targeting molecule, with one part being an aptamer with the ability to bind to I-*Sce*I and the other part being homologous to a target locus for gene correction. I-*Sce*I would then target this DNA to the genomic locus to be modified. We planned to first select for I-*Sce*I aptamers and then to assess their *in vivo* capacity to stimulate gene targeting.

CHAPTER 2

CONSTRUCTION AND TESTING OF A FUSION PROTEIN BETWEEN THE I-*SCE*I ENDONUCLEASE AND THE *GAL4*-DNA BINDING DOMAIN (DBD)

2.1 SUMMARY

The goal of this work was to increase the frequency of gene targeting in human cells by attaching the targeting DNA to a protein involved in the DSB or its repair. The concept was that by bringing the targeting DNA into close proximity of the DSB that it would be preferentially used to repair the break, thus increasing the frequency of gene targeting. In order to test our hypothesis, two parallel strategies were attempted. The first strategy involved the construction of a fusion protein between I-*Sce*I and the *GAL4* DNA-binding domain (*GAL4*-DBD). From transient transfections in HEK-293 cells, repair was shown using oligonucleotides that contained the *GAL4*-DBD site. These targeting oligonucleotides were also shown to bind specifically and strongly to the *GAL4*-DBD protein in EMSA gels. Several vectors and also several yeast strains were engineered to contain the *GAL4*-DBD I-*Sce*I fusion protein with various linkers. Whole cell extracts from both yeast and HEK-293 cells did not show any expression of the fusion protein.

2.2 INTRODUCTION

2.2.1 Targeting DNA choice

In order to carry out gene targeting, an exogenous DNA molecule must be transformed or transfected into the cell. There are several different targeting DNA types that could be used including plasmids, chromosomally integrated truncated genes, PCR products, and oligonucleotides. A comparison of several DNA targeting molecules (short DNA oligonucleotides, RNA/DNA chimeric oligonucleotides, and PCR products ~500 bp to

~2kbp) showed DNA oligonucleotides to have the highest frequency of chromosomal gene correction in HEK-293 cells (Nickerson and Colledge 2003). Additionally, our lab has sufficient experience with oligonucleotide-based gene correction in yeast and human cells (Storici, Durham et al. 2003). Also, oligonucleotides are relatively cheap and do not require time-consuming steps such as generating PCR products or isolating plasmids. For these reasons, we chose to use DNA oligonucleotides as our repairing DNA molecule of choice.

2.2.2 Gene targeting assay in human cells

We had in our possession a monoclonal cell line known as 658-D which is a derivative of human embryonic kidney (HEK-293) cells. The cell line contains an integrated GFP under the strong cytomegalovirus/chicken β actin (CMV/CBA) promoter but cannot express functional GFP because it is truncated by the presence of a STOP codon and the 18 basepair I-*Sce*I recognition site. Using this cell line, along with the corresponding plasmid-based assay in HEK-293 cells, we were able to detect the frequency of gene targeting by flow cytometry.

2.3 RESULTS

2.3.1 HEK-293 oligonucleotide preference

As stated, oligonucleotides were chosen as our targeting DNA molecule. After making this decision the question remained whether to use complementary oligonucleotides to form double-stranded DNA (dsDNA) or to use single-stranded DNA oligonucleotides (ssDNA). In yeast, it was shown that complementary DNA oligonucleotides were more efficient at gene targeting than either single oligonucleotide alone (Storici, Durham et al. 2003). However, in mammalian cells there is not a clear preference for single-stranded oligonucleotides or double-stranded oligonucleotides but rather the preference varies between different mammalian systems (Lu, Lin et al. 2003, Nickerson and Colledge 2003, Radecke, Radecke et al. 2004). In order to determine which oligonucleotide structure we should have chosen several transfections were done with single-stranded oligonucleotides, gapped duplex oligonucleotides, and fully complementary oligonucleotides. Gapped duplex oligonucleotides were shown to condense more than complementarity oligonucleotides or a 3kbp plasmid in the presence of poly(ethylenimine) (PEI), which is the polycationic transfection reagent that used in our mammalian cell transfections (Sarkar, Conwell et al. 2005). **Figure 2.1** and **Figure 2.2** show the results from our studies in both a plasmid-based and chromosomal assay. It is clear that in both our plasmid as well as our chromosomal assay there is a strong preference for ssDNA over dsDNA. Interestingly, there was a strand bias in favor of repair with the coding strand for the chromosomal locus but no such bias existed for the plasmid assay. This strand bias is likely due to the chromosomal architecture at the integration site of GFP in 658-D. The strand bias is not likely due to cell line differences since the HEK-293 and 658-D cell lines are identical except for the integrated disrupted GFP in the 658-D cell line. Even in the gapped duplex-forming oligonucleotide (the F

oligonucleotide with the 12H5 linker oligonucleotide), the majority of the molecule being single-stranded, the level of repair decreases significantly. This is consistent with previous findings in mammalian cell lines that found ssDNA oligonucleotides were a more favorable substrate for gene targeting in episomal or chromosomal loci compared to double-stranded oligonucleotides (Nickerson and Colledge 2003, Radecke, Radecke et al. 2004). The reason for this preference is unclear but it was proposed that double-stranded oligonucleotides might have to unfold prior to hybridization with a target locus (Radecke, Radecke et al. 2004). Also, it is possible that a blunt-ended dsDNA sequence, such as that used in our studies, is recognized as DNA damage, triggering a checkpoint response and degradation.

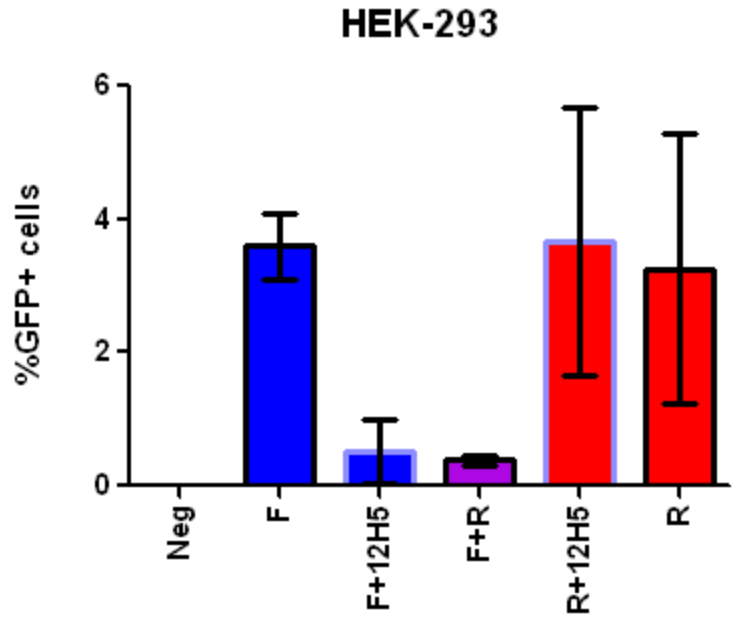


Figure 2.1 Comparison between complementary DNA and ssDNA in HEK-293 cells. Level of GFP repair following transfection with oligonucleotides to repair a nonfunctional GFP gene on a plasmid in HEK-293 cells. The plasmid was digested *in vitro* by I-*Sce*I forming a DSB in GFP that could be repaired resulting in a GFP+ phenotype. F = oligonucleotide homologous to the coding strand of GFP (40 bases of homology to either side of the DSB site), R=oligonucleotide homologous to the template strand of GFP, and 12H5 is an oligonucleotide with 12 bases of complementarity to the 3' end of the F oligo and 12 bases of homology to the 5' end of the F oligo with a 5 base gap (non-complementary stretch of bases) in between.

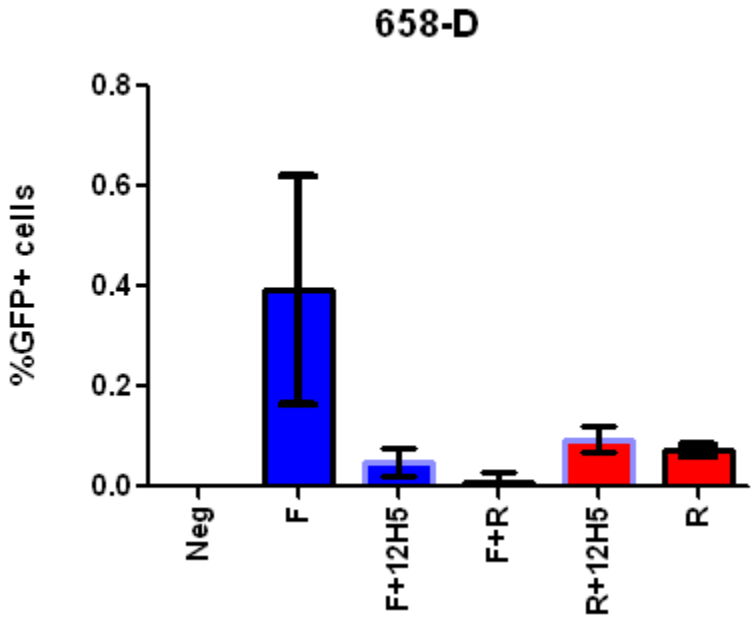


Figure 2.2 Comparison between complementary DNA and ssDNA in 658-D cells. Level of GFP repair following transfection with oligonucleotides to repair a nonfunctional GFP gene integrated somewhere in the genome along with an I-*Sce*I expression vector. F = oligonucleotide homologous to the coding strand of GFP (40 bases of homology to either side of the DSB site), R=oligonucleotide homologous to the template strand of GFP, and 12H5 is an oligonucleotide with 12 bases of complementarity to the 3' end of the F oligonucleotide and 12 bases of homology to the 5' end of the F oligonucleotide with a 5 base gap (non-complementary stretch of bases) in between. As can be seen, there is a strand bias favoring the F oligo over the R oligo.

2.3.2 Plasmid design and construction

The *GAL4*-DBD I-*Sce*I fusion protein was constructed using the Clontech Matchmaker Yeast 2-Hybrid Assay System (Clontech, Mountainview, CA). The I-*Sce*I coding sequence was PCR amplified with primers that contained BglII and SalI restriction enzyme sites, respectively. PCR product was digested along with the vector and the I-*Sce*I sequence was cloned into the multiple cloning site (MCS), downstream of the *GAL4*-DBD (**Figure 2.3**). After construction of the yeast vector, the *GAL4*-DBD and the *GAL4*-DBD I-*Sce*I fusion protein sequences were PCR amplified and cloned into the pFLAG CMV 6C vector (Sigma-Aldrich) such that the fusion protein would be under the strong cytomegalovirus (CMV) promoter for use in mammalian cells. Additionally, several plasmid constructs were generated to express the *GAL4*-DBD or the *GAL4*-DBD fusion protein with I-*Sce*I in the bacterial vector pMAL c4X. Bacterial plasmids contained the *GAL4*-DBD or the *GAL4*-DBD I-*Sce*I fusion protein and the maltose binding protein (MBP) as a fusion protein. The MBP provided solubility and good expression of the *GAL4*-DBD which was subsequently purified by the Lieberman lab.

The *GAL4*-DBD *LexA*-DBD fusion protein was constructed by cloning the *GAL4*-DBD sequence into the pFLAG CMV 6C vector to generate the pFLAG GAL plasmid. After sequencing and verification of the pFLAG GAL plasmid, the *LexA*-DBD sequence was cloned downstream of the *GAL4*-DBD. Two different constructs were made, one where

the *LexA*-DBD sequence alone was cloned downstream of the *GAL4*-DBD and another construct where a flexible glycine-serine linker $(GGGGS)_2$ was cloned in between the *GAL4*-DBD and the *LexA*-DBD.

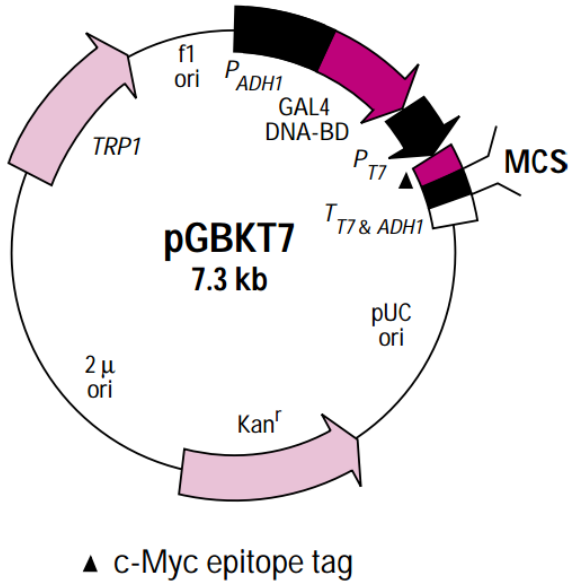


Figure 2.3 The pGBKT7 plasmid. This plasmid was used as the starting plasmid for the *GAL4*-DBD I-*Sce*I fusion protein. The fusion protein was expressed under the strong constitutive yeast promoter for alcohol dehydrogenase 1a (ADH1). Also, the region between the *GAL4*-DBD and the multiple cloning site (MCS) contains the T7 promoter, a c-Myc epitope tag, and restriction enzyme sites. Despite this being acceptable for the 2-hybrid assay, this region was not designed specifically as a linker.

2.3.3 Oligonucleotide structure design

After determining that ssDNA oligonucleotides should be used for repair in our human cells instead of dsDNA oligonucleotides, each of the linkers needed oligonucleotide designs. For the *GAL4*-DBD linker strategy at least part of the oligonucleotide needed to be dsDNA. The *GAL4*-DBD binds to a 17 bp dsDNA sequence and so we needed to determine where on our oligonucleotide the 17 bp double-stranded DNA region should be. Several different DNA oligonucleotides containing the *GAL4*-DBD recognition sequence were tested for their ability to repair GFP in HEK-293 cells (and the modified HEK-293 cell line 658-D). The different oligonucleotides are shown in **Figure 2.4**. As stated, in our HEK-293 and 658-D cell systems, ssDNA repairs more efficiently than dsDNA. Therefore, the oligonucleotides generated contained only a portion of dsDNA, enough to contain the *GAL4*-DBD recognition sequence and a 5 bp spacer. It was discovered that oligonucleotides containing the *GAL4*-DBD recognition sequence at the 5' end of the oligo (Aa and Cc) were more efficient at repairing the DSB (**Figure 2.5**). Also it should be noted that the 658-D cell line showed a strand bias in favor of the repairing molecule having homology to the coding strand of GFP (hence, Aa repaired more efficiently than Cc), and this is consistent with previous data (see **Figure 2.2**).

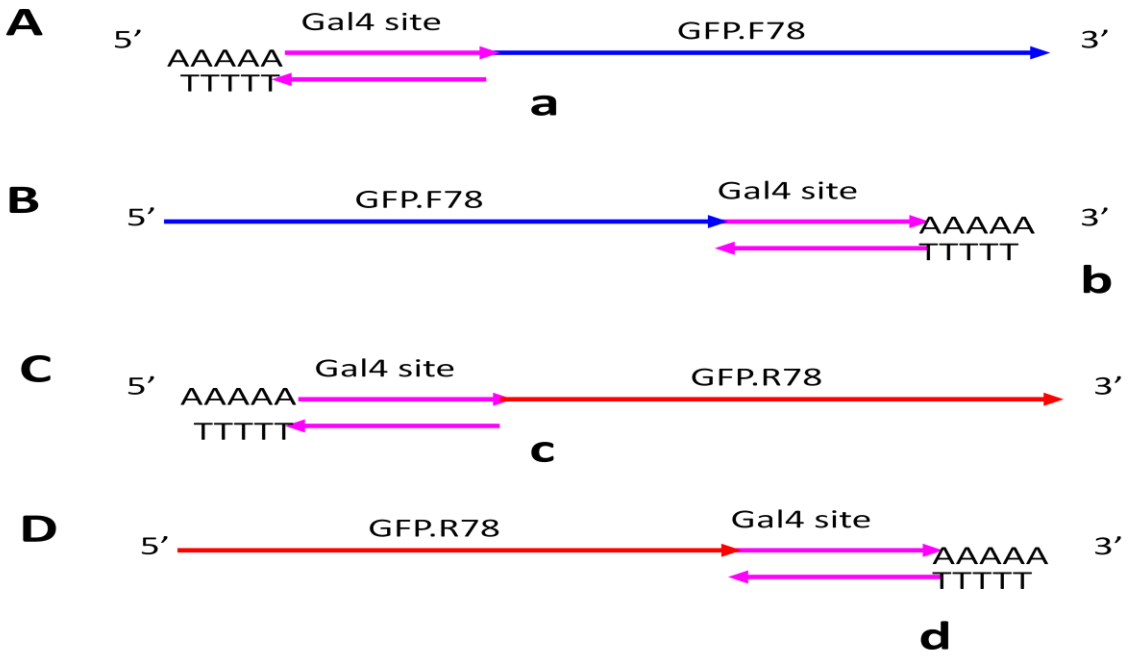


Figure 2.4 The different oligonucleotides tested in the HEK-293 and 658-D human cell lines. Each oligo pair Aa, Bb, Cc, or Dd contains the 17bp *GAL4*-DBD recognition sequence. The differences in the oligos are the positioning of the *GAL4*-DBD recognition sequence (either at the 5' or 3' end of the repairing oligo) and also the repairing sequence. For oligos Aa and Bb, the repairing sequence (GFP.F78) has homology to 39 bases of the coding strand on either side of the DSB, whereas oligos Cc and Dd's repairing sequence (GFP.R78) have homology to the template strand.

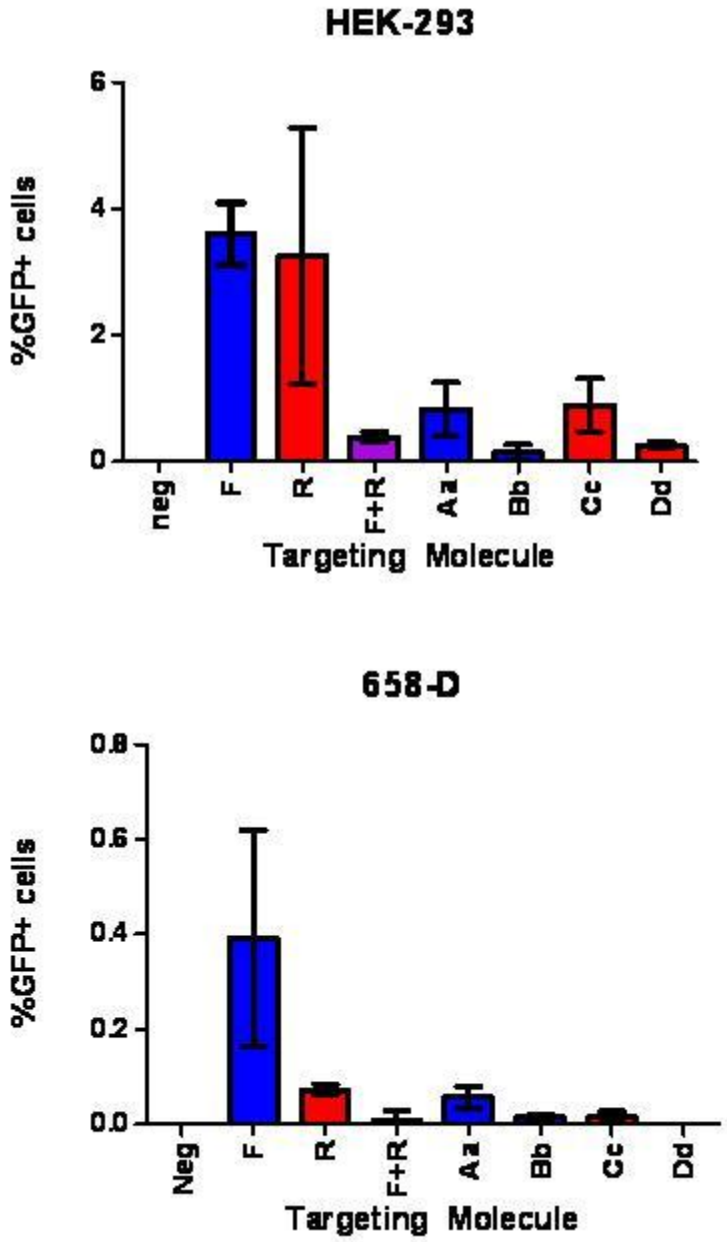


Figure 2.5 The efficiency of different oligonucleotides to accurately repair GFP by the DSB generated by I-*Sce*I. The two different cell lines used are HEK-293 and 658-D representing both plasmid and chromosomal gene repair respectively. The F oligo is a single-stranded oligo that has 80 bases of homology (40b on either side of the break) to the GFP coding strand and the R oligo is a single-stranded oligo that has 80 bases of homology (40b on either side of the break) to the template strand of GFP. The other oligonucleotides are described in **Figure 2.4**.

After determining that the optimal oligonucleotide structure for the *GAL4*-DBD in our human cells would be to contain the *GAL4*-DBD recognition site at the 5' end of the molecule and the targeting DNA will be the coding sequence we needed to ascertain whether or not this molecule can still be bound by the *GAL4*-DBD. In order to verify the DNA binding capability of the *GAL4*-DBD to the optimal oligonucleotide structure, we used electophoretic mobility shift assay (EMSA) gels. EMSA works by using electrophoresis to separate different samples in a polyacrylamide gel. Larger, more positively charged samples migrate slower than smaller, more negatively charged samples. Initial studies were done with purified *GAL4*-DBD protein (confirmed by Western blot), and several different DNA oligonucleotides, each containing a 5' *GAL4*-DBD site, as shown in **Figure 2.6**. The *GAL4*-DBD/I-*Sce*I fusion protein was not purified in the same manner so all tests for binding were done with the *GAL4*-DBD alone. We radiolabeled the different oligonucleotide designs containing the double-stranded *GAL4*-DBD region and binding to the *GAL4*-DBD was shown with P32-labeled oligonucleotides containing the 17bp consensus recognition sequence. In addition to the *GAL4*-DBD recognition sequence, these oligonucleotides contained homology to GFP such that they could be used to restore function of the marker. Of the several initial oligonucleotides tested, two specific structures, the stem-loop forming single oligonucleotide as well as the Aa oligonucleotide mentioned previously, were chosen for further testing. Both these structures contained the *GAL4*-DBD site at the 5' end of the

oligonucleotide, making them ideal for use in human cells. Each of these constructs showed that the binding by the *GAL4*-DBD was specific, as excess unlabeled competitor abolished the shift seen by EMSA (**Figure 2.7**).

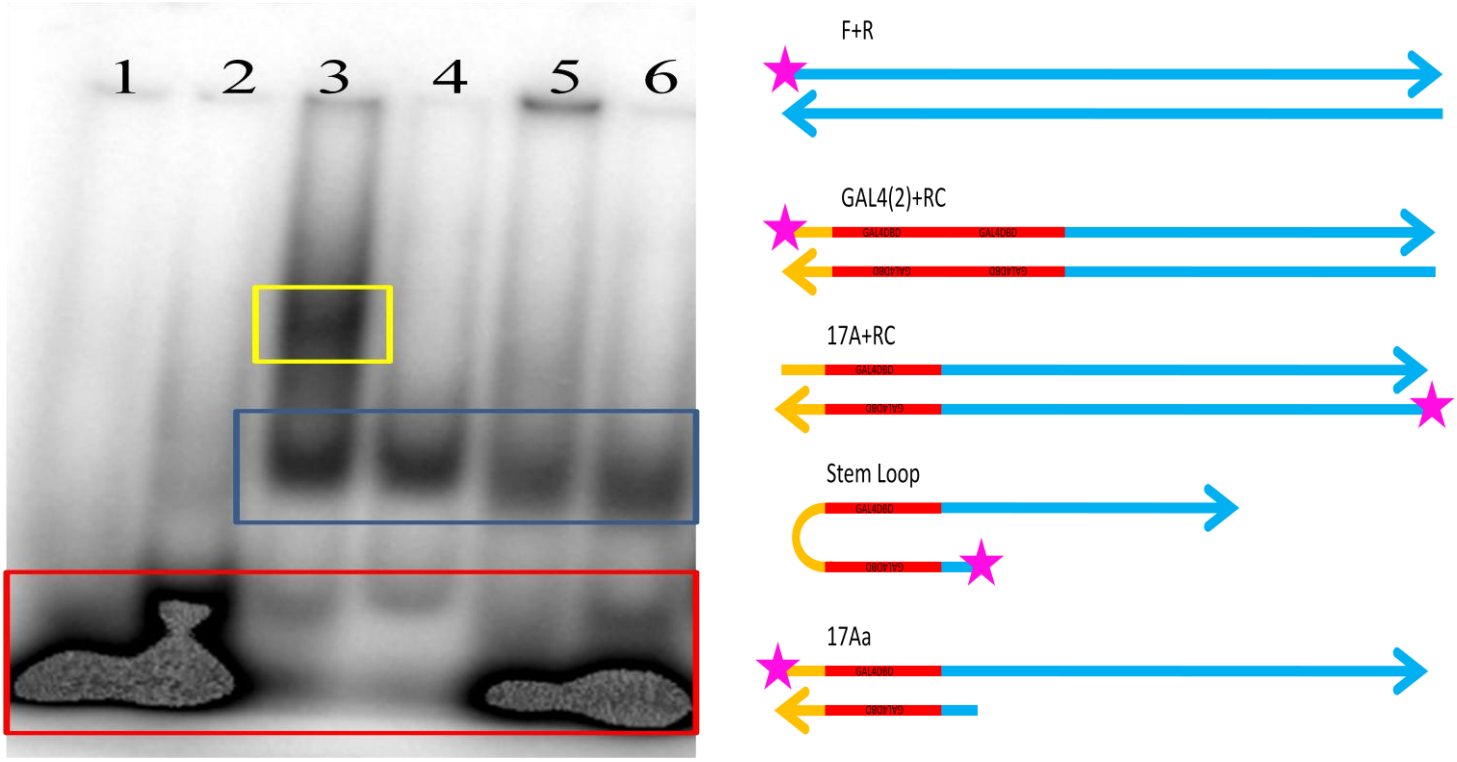


Figure 2.6 The different P32-labeled oligonucleotides used in the *GAL4*-DBD EMSAs in the presence of the *GAL4*-DBD. The red box contains the free probe, the blue box contains the protein-DNA complex and the yellow box contains a super complex where the DNA is bound by 2 *GAL4*-DBD dimers. Lane 1 represents a negative control, a single stranded oligo that does not contain the *GAL4*-DBD recognition sequence or the spacer (the F oligo), but otherwise is homologous to the other oligos. Lane 2 represents another negative control, a double stranded oligonucleotide (F+R) that does not contain the *GAL4*-DBD recognition sequence. Lane 3 represents an oligo with 2 closely spaced *GAL4*-DBD recognition sequences (*GAL4*-DBD(2)). Lane 4 represents the 17A nucleotide along with its reverse complement (17A+RC). Lane 5 represents the stem loop forming oligo that contains the *GAL4*-DBD recognition sequence (Stem Loop). Lane 6 represents a long single stranded oligo with a portion that is double stranded which contains the *GAL4*-DBD recognition sequence (the 17Aa oligo). Again, a schematic of the oligos used is on the right, the red area representing the *GAL4*-DBD recognition sequence, the orange representing a spacer and the blue region representing the GFP repairing sequence. The pink star represents the P32 label.

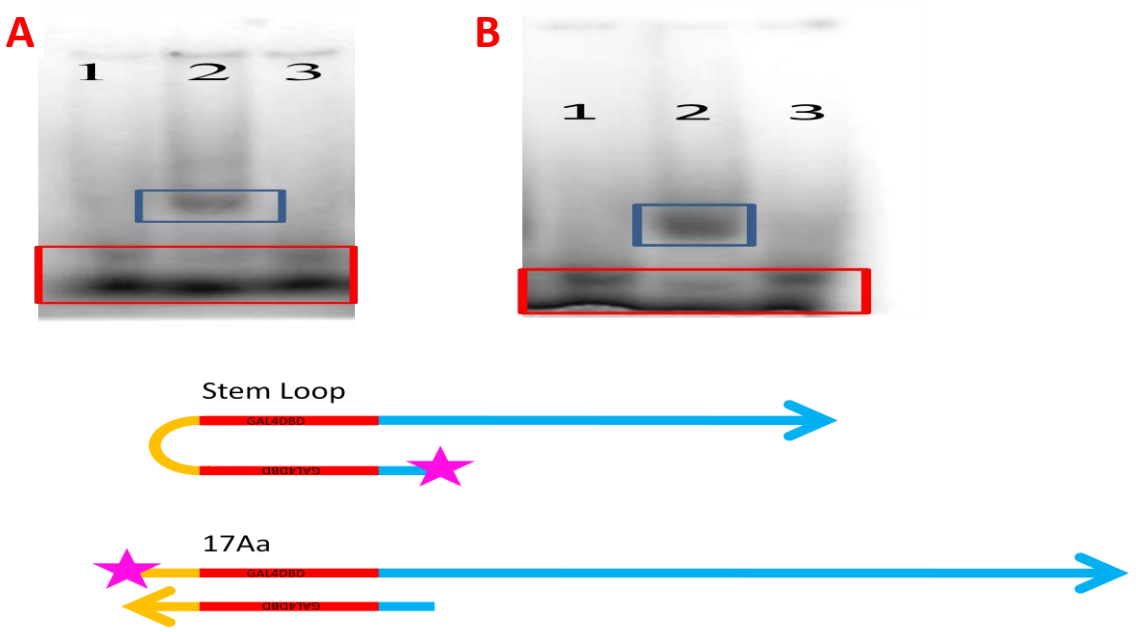


Figure 2.7 The Stem Loop and 17Aa oligos bind *GAL4*-DBD. (**A**) the single stranded oligonucleotide "Stem Loop" that folds to form a small loop and a long stem which contains the *GAL4*-DBD recognition sequence and (**B**) the 17Aa oligo which contains a long single strand and short double stranded region containing the *GAL4*-DBD recognition sequence. The red boxes contain the free probes and the blue boxes contain the protein-DNA complexes. Lane 1 represents the condition with *GAL4*-DBD, the P32-labeled oligonucleotide, and the same oligonucleotide unlabeled and in excess (a cold competitor to show the binding is specific). Lane 2 represents the *GAL4*-DBD and the P32-labeled oligo. Lane 3 represents the P32-labeled oligo without *GAL4*-DBD present. A schematic of the oligos used is on the right, the red area representing the *GAL4*-DBD recognition sequence, the orange representing a spacer and the blue region representing the GFP repairing sequence. The pink star represents the P32 label.

2.3.4 *in vivo* testing

The *GAL4*-DBD I-*Sce*I fusion protein was tested *in vivo* for gene targeting stimulation. For the *in vivo* testing of the fusion protein transfections in mammalian cells as well as in yeast cells were performed. In the HEK-293and 658-D cell lines, transfections of the expression vector for the *GAL4*-DBD I-*Sce*I fusion along with either of the two *GAL4*-DBD site containing oligonucleotides previously tested (see **Figure 2.7**) did not yield any GFP+ (gene corrected) cells. In yeast, a strain was made in which the *GAL4*-DBD I-*Sce*I fusion protein under the inducible *GAL1-10* promoter was integrated into the *TRP5* gene. This strain was used to compare gene correction using oligonucleotides containing the *GAL4*-DBD site or a scrambled *GAL4*-DBD sequence that should not be bound by the *GAL4*-DBD. In our yeast strains, double-stranded DNA oligonucleotides are more efficient substrates for homologous recombination than single-stranded DNA, so we were not limited to have only one *GAL4*-DBD site as was the case for studies in human cells. Oligonucleotides containing five tandemly repeated *GAL4*-DBD sites or scrambled *GAL4*-DBD sites were used as well as oligonucleotides containing only a single *GAL4*-DBD or scrambled *GAL4*-DBD site. There was no significant difference between the scrambled oligonucleotides and *GAL4*-DBD site containing oligonucleotides for any of the yeast strains tested, including the original strain that did not contain the *GAL4*-DBD.

In addition to the *in vivo* testing of the *GAL4*-DBD I-*Sce*I fusion protein, *in vitro* testing for protein expression was performed. A plasmid expressing the *GAL4*-DBD under the yeast constitutive alcohol dehydrogenase 1 (ADH1) promoter was transformed into yeast.

The vector, pGBKT7 was previously described (see **Figure 2.3**). Whole cell extracts from the yeast strain transformed with the pGBKT7 vector were obtained using Yeast Protein Extraction Reagent (Y-PER) (Pierce, Rockford, IL) and Western blots were done using a monoclonal antibody to the *GAL4*-DBD (Clontech, Mountainview, CA). As can be seen in **Figure 2.8**, both the purified *GAL4*-DBD as well as the *GAL4*-DBD from the yeast cell extracts were able to be detected by chemiluminescence. Similarly, a plasmid containing the fusion protein between the *GAL4*-DBD and I-*Sce*I was transformed into yeast. This fusion protein was generated by inserting the I-*Sce*I coding sequence into the pGBKT7 vector downstream of the T7 promoter (see **Figure 2.3**). The whole cell extracts from yeast containing this plasmid do not show a detectable fusion protein of the *GAL4*-DBD and I-*Sce*I by Western blot (see **Figure 2.9**). Interestingly, Western blots with this plasmid do show the individual *GAL4*-DBD, suggesting cleavage of the fusion protein or aberrant expression. Suspecting a problem with the linker region between the *GAL4*-DBD and I-*Sce*I, several different plasmids and yeast strains with different linkers were created. As stated, the pGBKT7 plasmid does not contain a true linker sequence, but rather it contains a c-Myc epitope tag, followed by several restriction enzymes and the T7 promoter. Use of a rigid proline linker (Gustavsson, Lehtio et al. 2001), a flexible glycine-serine linker (Zhao, Yao et al. 2008), a PRAMA linker suggested by the Doetsch lab Emory, as well as an alpha helix-forming linker (Zhao, Yao et al. 2008)were all created (the amino acid sequences are shown in **Table 2.1**). In addition to the *GAL4*-DBD western blots, western blots were done using a polyclonal antibody to I-*Sce*I (Santa Cruz Biotechnology, Dallas, Texas) but again there was no expression seen of the fusion protein (**Figure 2.10**). Of all the *GAL4*-DBD and I-*Sce*I fusion protein constructs none

showed expression by Western blot. Additionally, changing the orientation of the fusion protein (such that the *GAL4*-DBD was at the carboxyl terminal instead of the amino terminal) did not improve expression. No expression was seen in yeast using a polyclonal antibody to I-*Sce*I or a monoclonal antibody to the *GAL4*-DBD. Western blots were also done using extracts from HEK-293 cells transiently transfected with the *GAL4*-DBD and the I-*Sce*I *GAL4*-DBD fusion protein. While there was expression detected for the *GAL4*-DBD alone there was no expression detected of the fusion protein between the *GAL4*-DBD and I-*Sce*I (**Figure 2.11**).

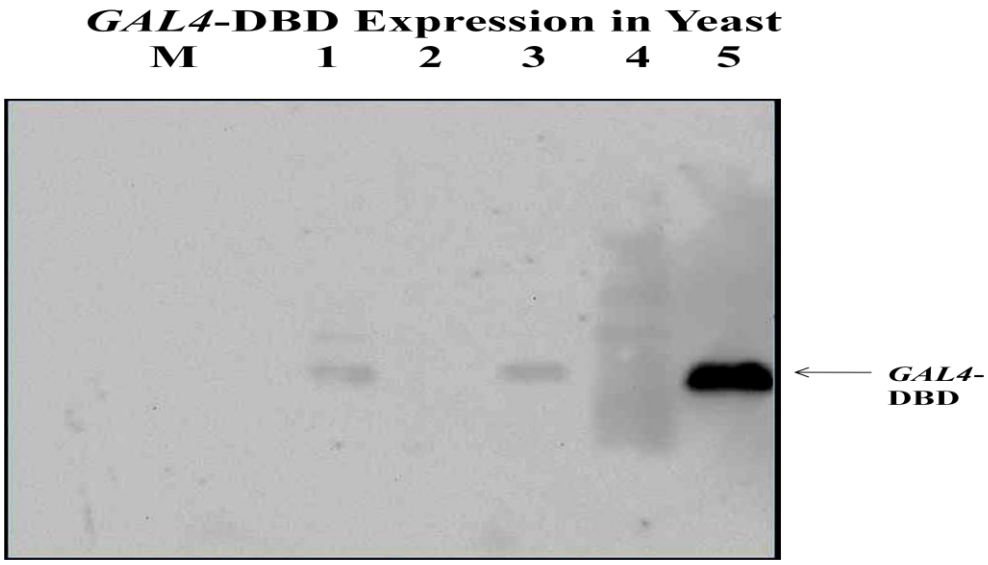


Figure 2.8 A Western blot of purified *GAL4*-DBD and *GAL4*-DBD from a yeast extract. The cell extract is of a strain expressing the *GAL4*-DBD stably replicated under a 2 µm plasmid origin of replication and under the ADH1 promoter. M is the colorimetric molecular weight marker which does not appear on this chemiluminescent blot. Lanes 2 and 4 are both from a yeast strain not expressing the *GAL4*-DBD under the ADH1 promoter. Lanes 1 and 3 are from the yeast strain KM85 where the *GAL4*-DBD is expressed under the ADH1 promoter. Lane 5 is the positive control, purified *GAL4*-DBD from the Lieberman lab. This indicates that cell extracts expressing the *GAL4*-DBD can be used to quantify the presence or absence of our fusion protein.

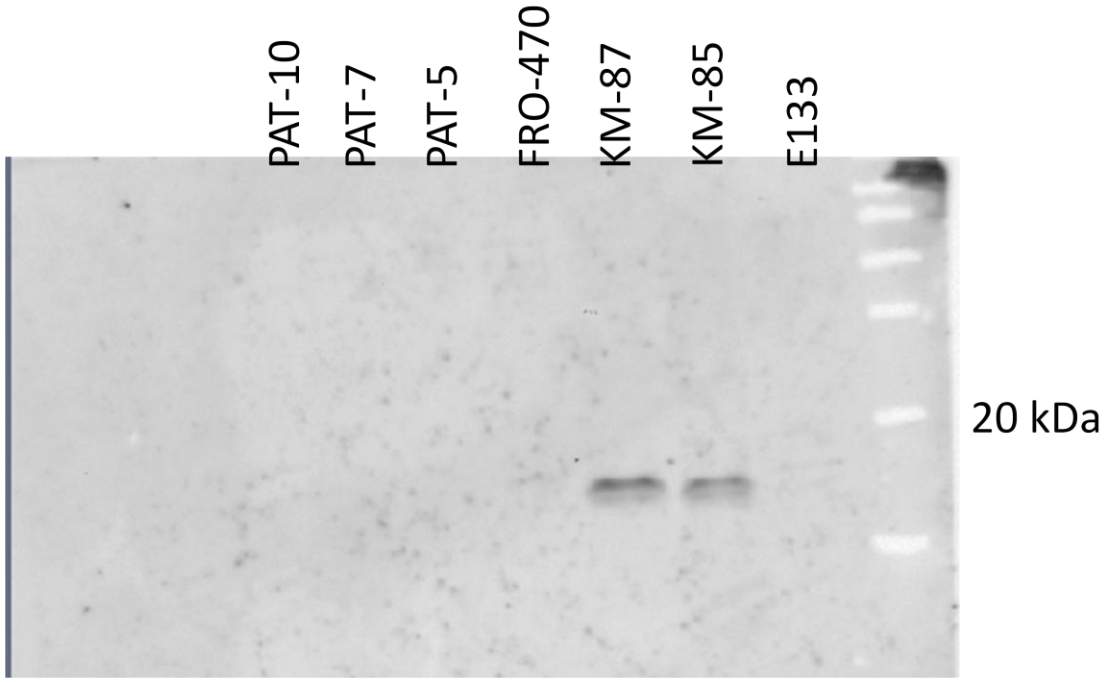


Figure 2.9 A Western blot of several whole cell extracts from different strains using a monoclonal antibody for the *GAL4*-DBD. E133 is the background strain for KM-85 and KM-87 and does not contain the *GAL4*-DBD. The strains KM-85 and KM-87 are different isolates of the same strain, where the *GAL4*-DBD I-*Sce*I fusion protein was expressed on a plasmid under the ADH1 promoter. FRO-470 is the background strain for PAT-5, PAT-7, and PAT-10 and should not have the *GAL4*-DBD. PAT-5 is FRO-470 except the *GAL4* gene has been knocked out. The PAT-7 strain contains the *GAL4*-DBD under the *GAL1-10* promoter but interestingly did not show expression. PAT-10 is a strain where the *GAL4*-DBD I-*Sce*I fusion protein (joined by a flexible glycine/serine $(GGGGS)_2$ linker) is under the *GAL1-10* promoter. As can be seen, none of these strains shows expression of the fusion protein.

Linker	Sequence	Description
pGBKT7 "linker"	PEFVIRLTIGRAAIMEEQKLISEEDLHMAMEAEFPGIR	The "linker" provided with the pGBKT7 vector
Proline	GTPTPTPTPTGEF	Rigid linker
Alpha-helix forming	AEAAAKEAAAKA	2ndary structure forming rigid linker
Glycine-serine	GGGGSGGGGS	Flexible linker
PRAMA	PRAMA	Linker design for a fusion protein from the Doetsch lab at Emory

Table 2.1 Sequences of different linkers for the *GAL4*-DBD and I-*Sce*I fusion protein The table contains the sequences of the different linkers used to construct the *GAL4*-DBD and I-*Sce*I fusion protein along with a description. The pGBKT7 vector contains a region between the *GAL4*-DBD and I-*Sce*I but is not a true linker, hence it is described as a "linker."

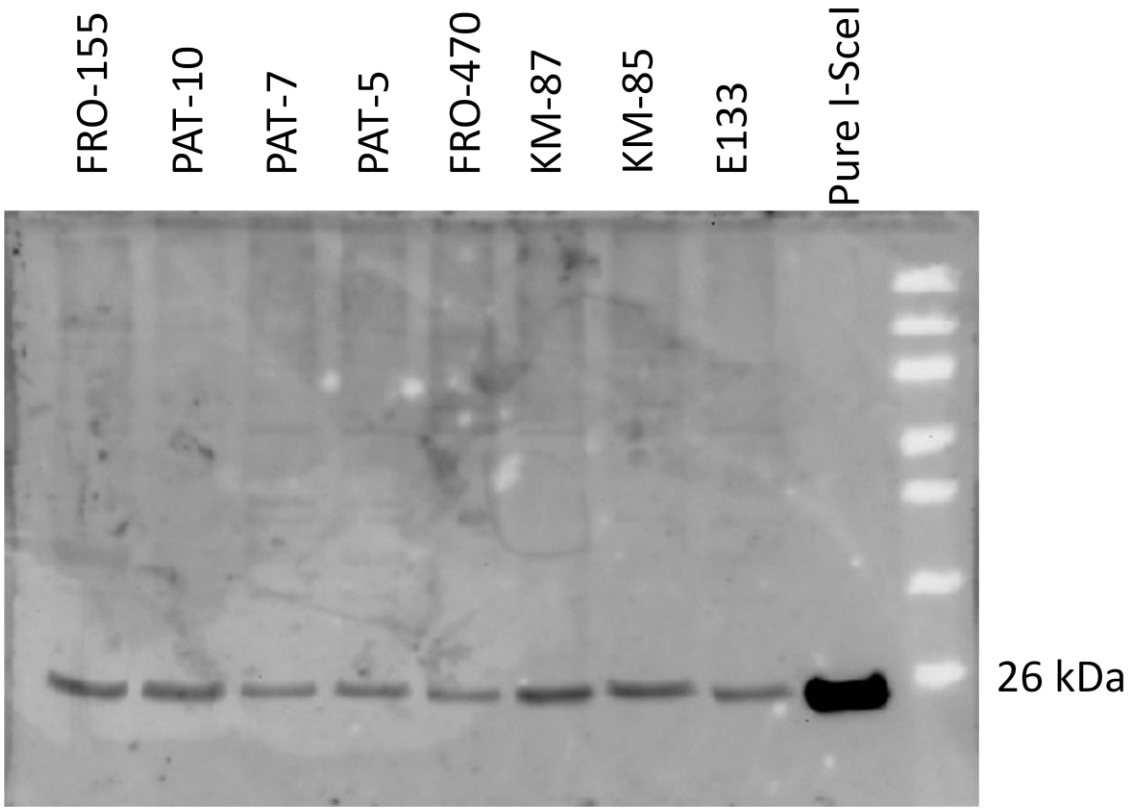


Figure 2.10 A Western blot of several whole cell extracts from different strains using a polyclonal antibody for I-*Sce*I. E133 is the background strain for KM-85 and KM-87 and does not contain the *GAL4*-DBD. FRO-155 is a strain that has the I-*Sce*I gene under the *GAL1-10* promoter but does not have the fusion protein. The strains KM-85 and KM-87 are different isolates of the same strain, where the *GAL4*-DBD I-*Sce*I fusion protein was expressed on a plasmid under the ADH1 promoter. FRO-470 is the background strain for PAT-5, PAT-7, and PAT-10 and should not have the *GAL4*-DBD. PAT-5 is FRO-470 except the *GAL4* gene has been knocked out. The PAT-7 strain contains the *GAL4*-DBD under the *GAL1-10* promoter but interestingly did not show expression. PAT-10 is a strain where the *GAL4*-DBD I-*Sce*I fusion protein (joined by a flexible glycine/serine GGGGS linker) is under the *GAL1-10* promoter. As can be seen, none of these strains shows expression of the fusion protein.

In addition to the *GAL4*-DBD I-*Sce*I fusion protein, we also attempted to construct a fusion protein between the *GAL4*-DBD and the *LexA*-DBD. The *LexA*-DBD is another well characterized, small (16 kDa) DNA binding domain responsible for binding to a 16 bp palindromic sequence 5' CTGTNNNNNNNNACAG 3' (Knegtel, Fogh et al. 1995). For this particular set of constructs a new system was designed. Whereas in the *GAL4*-DBD I-*Sce*I fusion protein we relied on the *GAL4*-DBD binding to a double-stranded region on a small DNA oligonucleotide, for the *GAL4*-DBD *LexA*-DBD experiments we constructed a plasmid that contained the gene of interest flanked by *LexA*-DBD sites. In this way, our aim was to conjugate the *LexA*-DBD site containing plasmid targeted for repair with a separate plasmid that contained 200 bases of homology to the targeted gene. The repairing vector either had these 200 bases of homology flanked with *GAL4*-DBD recognition sites or scrambled *GAL4*-DBD sequences. In this way we sought to directly compare repair between the targeted DNA plasmid and a separate targeting DNA plasmid. Fusion protein constructs were made for use in our human cell lines. In this assay we saw no significant difference between a targeting plasmid containing the *GAL4*-DBD sites and a targeting plasmid containing scrambled *GAL4*-DBD sites flanking the region of homology with either of the *GAL4*-DBD *LexA*-DBD fusion protein constructs. Western blots were done on mammalian cell extracts to verify expression and the *GAL4*-DBD alone was detected, as was the *GAL4*-DBD *LexA*-DBD fusion protein without a linker. As stated previously, the *GAL4*-DBD I-*Sce*I fusion protein was not detected (**Figure 2.11**).

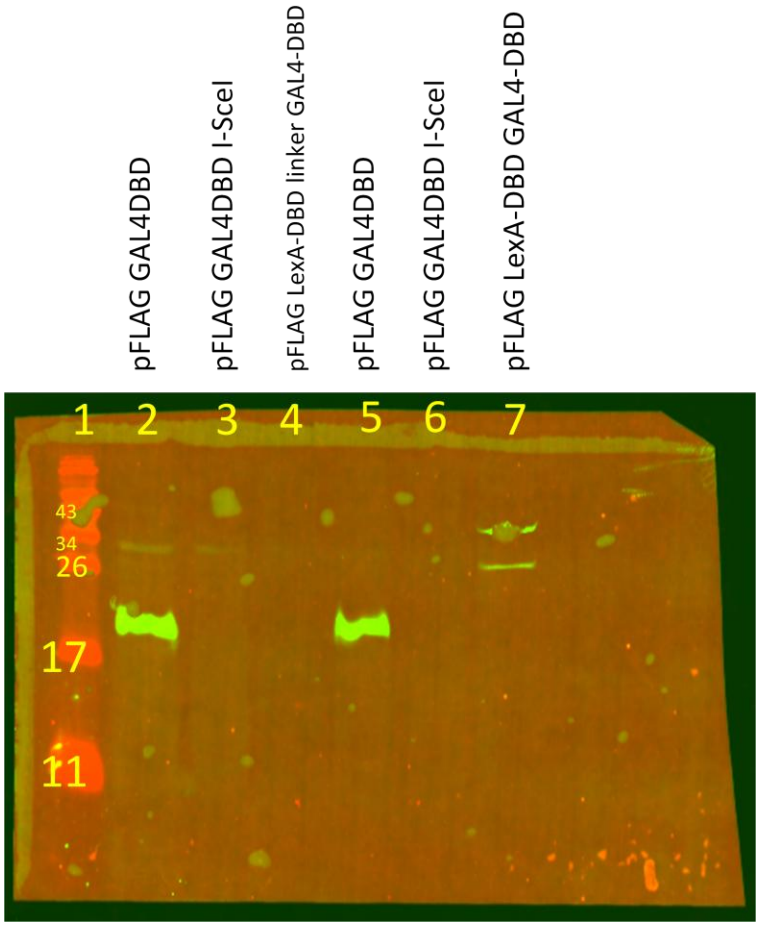


Figure 2.11 Western blot of several whole cell extracts from HEK-293. Several whole cell extracts were taken after transient transfections of various plasmids. Lane 1 is a molecular weight marker, lane 2 and 5 are two different extracts after transient transfection of plasmids expressing the *GAL4*-DBD alone (~19 kDa), lane 3 and 6 are extracts after transfection with the *GAL4*-DBD I-*Sce*I fusion protein, lane 4 is an extract after transfection with the *LexA*-DBD *GAL4*-DBD fusion protein with a flexible $(GGGGS)_2$ linker between the two proteins, and lane 7 is the same plasmid except without a linker between the *LexA*-DBD and the *GAL4*-DBD (~16 kDa + ~19 kDa = ~35 kDa).

2.4 CONCLUSIONS

EMSA gels showed specific and strong binding of the *GAL4*-DBD by several oligonucleotides. The oligonucelotides shown to be capable of being bound by the *GAL4*-DBD were also shown to be capable of repairing GFP *in vivo*. Western blots done using a monoclonal antibody to the *GAL4*-DBD, a polyclonal antibody to I-*Sce*I, as well as a monoclonal antibody to the FLAG tag peptide sequence (dykddddk) did not show any expression of the *GAL4*-DBD I-*Sce*I fusion protein. There are many potential reasons why the fusion protein was not expressed. As can be seen in **Figure 2.9**, the individual *GAL4*-DBD protein can be seen in both the strain containing the vector with the *GAL4*-DBD only (KM-85) as well as strain containing the vector with the *GAL4*-DBD I-*Sce*I fusion protein (KM-87). The presence of the *GAL4*-DBD alone in this strain could be the result of proteolytic cleavage of the linker between the two proteins. Interestingly, this cleavage was not seen in the Western blot using the FLAG antibody. Alternatively, the fusion protein may have been unstable, unable to fold properly, and/or targeted for degradation.

Due to the inability of the *GAL4*-DBD I-*Sce*I fusion protein and the *GAL4*-DBD *LexA*-DBD to stimulate gene targeting, we opted to further develop the I-*Sce*I DNA aptamer strategy. Although these particular fusion proteins, the *GAL4*-DBD and I-*Sce*I along with the fusion protein between the *GAL4*-DBD and the *LexA*-DBD, did not stimulate gene targeting, without knowing the binding characteristics of the *GAL4*-DBD I-*Sce*I and *GAL4*-DBD *LexA*-DBD fusion proteins and without proper expression of the *GAL4*-DBD

I-*Sce*I fusion protein we cannot conclude that the concept itself was a failure. Further testing may yet reveal our hypothesis to be true. It also may be possible that only one of the two domains was able to bind DNA at any given time. Many proteins upon binding to their substrate undergo conformational changes and it could be that a conformational change in either of the domains inhibited the binding ability of the other domain. There are many possibilities to explain why the fusion protein strategy did not work but the work done here provided us with a better understanding of the protein-assisted targeting system.

CHAPTER 3

Real-Time PCR-Coupled CE-SELEX for DNA Aptamer Selection

The work presented in Chapter 3 was published as a research article under the same name in ISRN Molecular Biology:

Patrick Ruff, Rekha B. Pai, and Francesca Storici, "Real-Time PCR-Coupled CE-SELEX for DNA Aptamer Selection," ISRN Molecular Biology, vol. 2012, Article ID 939083, 9 pages, 2012. doi:10.5402/2012/939083. (2012). School of Biology, Georgia Institute of Technology, Atlanta, GA.

3.1 ABSTRACT

Aptamers are short nucleic-acid or peptide sequences capable of binding to a target molecule with high specificity and affinity. Also known as “artificial antibodies,” aptamers provide many advantages over antibodies. One of the major hurdles to aptamer isolation is the initial time and effort needed for selection. The systematic evolution of ligands by exponential enrichment (SELEX) is the traditional procedure for generating aptamers, but this process is lengthy and requires a large quantity of target and starting aptamer library. A relatively new procedure for generating aptamers using capillary electrophoresis (CE), known as CE-SELEX, is faster and more efficient than SELEX, but requires laser-induced fluorescence (LIF) to detect the aptamer-target complexes. Here we implemented an alternative system without LIF using Real Time (RT)-PCR to indirectly measure aptamer-target complexes. In three rounds of selection, as opposed to ten or more rounds common in SELEX protocols, a specific aptamer for bovine serum albumin (BSA) was obtained. The specificity of the aptamer to BSA was confirmed by electrophoretic mobility shift assays (EMSAs), an unlabeled competitor assay, and by a supershift assay. The system used here provides a cost effective, highly efficient means of generating aptamers.

3.2 INTRODUCTION

Aptamers are short single-stranded oligomers made up of DNA, RNA, or peptides that are capable of binding a target ligand (proteins, small molecules, or even living cells) with high affinity. They are also known as artificial antibodies because in addition to binding with high affinity, they also bind with high specificity. Aptamers have several advantages over antibodies, including ease and low cost of production which does not involve animals. Aptamers are less immunogenic than antibodies and are already being used as therapeutic agents in humans (Singerman, Masonson et al. 2008). Nucleic acid aptamers also are able to act in ways that antibodies cannot. Nucleic acid aptamers, unlike antibodies, can be selected for and used under non-physiological conditions, such as high salt conditions and varying pH (Xu and Lu 2010). Also, nucleic acid aptamers are able to undergo specific conformational changes that antibodies cannot. For example, nucleic acid aptamer binding can be "turned off" by the addition of the complementary strand (Cao, Tong et al. 2009). Additionally, nucleic acid aptamers can undergo a conformational change when binding to their target, and can be used as molecular beacons, fluorescently "off" when unbound, and "on" when bound (Morse 2007). The field of aptamers is rapidly growing as is the number of applications for their use.

Nucleic acid aptamers are "evolved" from random sequences of DNA/RNA by a process known as systematic evolution of ligands by exponential enrichment (SELEX) (Tuerk and Gold 1990)[5]. The SELEX procedure involves the use of the random library of DNA/RNA sequences being incubated with the target, followed by a partitioning step to remove unbound sequences, an elution step to recover the binding sequences, and then an

amplification step to generate a library of sequences enriched for binding. The SELEX procedure generally takes months to complete, with a typical selection requiring 10 or more rounds before completion (Jayasena 1999). Also, traditional SELEX requires a support for the target (magnetic beads, membranes, etc.) to bind with. The supports themselves can be targets for selection, and often rounds of negative selection must be done to avoid aptamers for the support.

Use of capillary electrophoresis (CE) allows for SELEX to be performed in a much shorter amount of time due to much more efficient partitioning and without the aptamers binding to the ligand support (the ligand flows freely in buffer, there is no support). In as little as one round of selection (Berezovski, Drabovich et al. 2005), and generally less than five rounds of selection, strong binding highly specific aptamers may be obtained. CE-SELEX is a new technology first developed for use in 2004 and has yet to be commonly used (Mendonsa and Bowser 2004). One of the main advantages to CE-SELEX over traditional SELEX is that the aptamer-target complex can be detected in the first round of selection. This early detection contrasts traditional SELEX, where several rounds must be done before being able to detect any DNA (Sefah, Shangguan et al. 2010). Most CE-SELEX is done with laser induced fluorescence (LIF) to increase the detection sensitivity to the analysed samples. Using CE with LIF a laser excites fluorescently-labeled samples passing through the glass capillary tube which then emit light that is captured by an on-board detector attached to the CE machine itself.

We have developed a technique for selection of DNA aptamers using CE but without the

need for an on-board laser/detector system. The system takes advantage of real-time polymerase chain reaction (RT-PCR). RT-PCR is able to sensitively detect DNA-target complexes early on in the selection procedure with an efficacy greater than that of traditional SELEX and equal to that of CE-SELEX with LIF. We believe that this system could be beneficial for researchers that have access to CE, but do not have access to CE with LIF, and are seeking to perform CE-SELEX. The protein that we chose to use as a target for aptamer selection was bovine serum albumin (BSA).

BSA is one of the most commonly used proteins in biochemical studies. BSA is widely used for stabilization of enzymes, preventing enzymes of interest from adhering to tubes or pipettes, or as a protein comparison standard. There are several advantages to using BSA in that it is relatively stable, abundantly available from cow's blood, and is of low cost. Due to the low cost and high abundance of BSA, we chose it to serve as the protein target for our RT-PCR coupled selection system. Here we report the isolation of a DNA aptamer with specificity for BSA using the RT-PCR coupled capillary electrophoresis selection system.

3.3 MATERIALS AND METHODS

3.3.1 Aptamer Selection

The target protein of interest, bovine serum albumin (BSA), was purchased as a

lyophilized powder through Sigma-Aldrich and was greater than 98% pure. The BSA was dissolved in RB1 buffer (50 mM Tris-HCl, pH 8.2) and stored at 4 °C at a stock concentration of 1 mM. The RB1 buffer was also used as the run buffer for capillary electrophoresis. The DNA library was purchased from Alpha DNA and contained a sequence of 5'_CTTCTGCCCGCCTCCTTCC-(N)36-GACGAGATAGGCGGACACT_3' (36 random nucleotides flanked by two fixed 19 base regions used later as primers for PCR amplification).

The protocol for SELEX using capillary electrophoresis (CE) was essentially as described earlier (Berezovski, Musheev et al. 2006) but with a few modifications. The initial bulk affinity assay was performed with 50 µM BSA and 25 µM DNA in order to view any DNA-protein complexes and determine the collection window. Capillary electrophoresis was done using a Beckman Coulter Proteomelab PA 800 with a photon diode array (PDA) capable of reading wavelengths in the UV range (10 nm - 400 nm), separating with a voltage at 10 kV. After determination of the collection window based on the bulk affinity analysis, the first round of selection began. The initial *in vitro* selection procedure involved 500 nM BSA and 3.3 µM DNA. The DNA library (0.3 µl at 100 µM) was mixed with 0.3 µl of SB3 (100 mM Tris-HCl at pH 8.2, 200 mM NaCl, and 10 mM $MgCl_2$) for a final concentration of 50 µM DNA library, 50 mM Tris-HCl at pH 8.2, 100 mM NaCl, and 5 mM $MgCl_2$. This mixture was heated in the BioRad iCycler to 94 °C for 1 minute, and then cooled to 20 °C at a rate of 0.5 °C/second. After the folding of the DNA library, 5 µl of 1 µM BSA dissolved in RB1 buffer was added, and additional run buffer was added to make the final volume 10 µl. This brought the final concentrations to 3.3 µM DNA library, 500 nM BSA, 6 mM NaCl, 300 µM MgCl2, and

50 mM Tris-HCl (pH 8.2). The collection window was partitioned into seven different fractions.

3.3.2 RT-PCR

After each round of selection, fractions were analyzed through real-time PCR (RT-PCR) using the ABI StepOne Plus. RT-PCR was done with two primers, the forward aptamer-amplifying primer P1 (5'_CTTCTGCCCGCCTCCTTCC_3') and the reverse primer P2 (5'_AGTGTCCGCCTATCTCGTC_3') respectively. The primers were designed using OligoAnalyzer (http://www.idtdna.com/analyzer/Applications/OligoAnalyzer/) to limit complementarity to each other, in order to decrease non-specific amplification of self-dimerizing primers. For amplification, 20 µl of PCR mix was prepared consisting of 10 µl of 2X Quanta SYBR Green PCR Master Mix (Roche), 0.6 µl of 10 µM P1, 0.6 µl of 10 µM P2, 1 µl of collected fraction as template, and 7.8 µl H2O. The PCR setup is shown in **Table 3.1**.

Table 3.1 Real-Time PCR cycling conditions

Cycle #	Denaturation	Annealing	Extension
1	94 °C for 30 seconds		
2-50	94 °C for 10 seconds	55 °C for 10 seconds	72 °C for 10 seconds
51			72 °C for 1 minute
Hold at 4 °C			

3.3.3 Amplification and isolation of aptamer strand

Following RT-PCR, the fraction containing the complex was amplified using standard PCR. PCR was done in a 100 µl volume consisting of 1 µl of 5U/µl X-Taq polymerase from Takara, 3 µl of 10 µM forward primer P1, 3 µl of 10 µM reverse primer P2, 10 µl of 10X Mg^{2+} buffer (Takara Ex Taq), 8 µl of 2.5 mM each dNTP, and 5 µl of the collected fraction from capillary electrophoresis. PCR was done using primer P1 and P2 as noted previously, except that the number of cycles for PCR amplification was based on 50% of the maximum yield as determined by RT-PCR. Additionally, primer P2 was biotinylated at its 5' end. The biotin-labeled primer was used subsequent to PCR in order to separate the strand of interest and the non-aptamer strand after PCR. Magnetic beads with streptavidin coating from Bangs Laboratories (Biomag Streptavidin Nuclease-Free) were used to bind the biotin-labeled DNA. Strands were separated with 10 mM NaOH after three washes with wash buffer (10 mM Tris-HCl at pH 8 with 500 mM NaCl and 1 mM EDTA). The single-stranded aptamer pool was used in subsequent rounds of selection.

3.3.4 Cloning and sequencing

Post-selection DNA cloning of the aptamer pool was done with the TOPO Zero Blunt Cloning Kit (Invitrogen). Standard PCR with unlabeled primers P1 and P2 was used to generate double-stranded DNA containing the aptamer sequence, which was then blunt-end ligated into the pCR-Blunt II-TOPO vector that contains the kanamycin resistance

gene. Colonies were selected for growth on kanamycin-containing media (kanamycin final concentration was 40µg/mL) and plasmid DNA was isolated using the GeneJET Plasmid Miniprep Kit (Fermentas). Asymmetric PCR with unlabeled primers P1 and P2 was used on the plasmid DNA to predominately generate the strand of interest which was then analyzed using CE and RT-PCR. 12 colonies were pooled into a group and each group was assayed for BSA binding using CE. Individual plasmids from each pooled group that showed binding were sequenced by Eurofins MWG Operon. Based on sequencing results, several candidate aptamers were chosen and ordered as salt-free oligonucleotides. Consensus sequence was analyzed using ClustalW2 (http://www.ebi.ac.uk/Tools/msa/clustalw2/). Additionally, mFold (http://mfold.rna.albany.edu/?q=mfold/DNA-Folding-Form) was used on each candidate aptamer to identify secondary structure.

3.3.5 Electrophoretic Mobility Shift Assay (EMSA)

Potential aptamer oligonucleotides and a negative control oligonucleotide were 5' labeled with P32 γ-ATP using T4 Polynucleotide Kinase (New England Biolabs). The negative control consisted of an oligonucleotide of the same length as the random DNA library oligonucleotides (74 bases), contained the same flanking primer regions, and had a fixed sequence for its internal region 5'-

CTTCTGCCCGCCTCCTTCCGGTCGGGCACACCTGTCATACCCAATCTCGAGG

CCAGACGAGATAGGCGGACACT-3'. The internal region was chosen using a random

DNA sequence generator with a specified GC content of 50% (http://www.faculty.ucr.edu/~mmaduro/random.htm). EMSAs were performed under different binding conditions. The binding conditions for the initial EMSA were done by adding equal amounts of SB3 buffer to the labeled oligonucleotides (20,000 cpm equivalent) and incubating at 94 ^{o}C for 1 minute in a PCR machine and a quick chilling to 20 ^{o}C at a rate of 0.5 ^{o}C per second (total time taken is ~4 minutes). BSA stock of 1 mM was made in RB1 buffer (the run buffer used in the CE-SELEX protocol). Different dilutions of BSA as required for assessing aptamer binding were also made in RB1 buffer. With fixed concentration of the γP32-labeled candidate aptamer, the concentration of the ligand (BSA) was varied from 50 to 800 µM.

The buffer used for binding was a 6X buffer consisting of 600 mM ammonium chloride, 300 mM potassium chloride, 30 mM sodium chloride, 120 mM Tris-HCl pH 7.5, 30% glycerol, and bromophenol blue (BPB) 0.25%. The entire binding conditions were made up to 10 µl with RB1 buffer. The reaction was incubated at room temperature for 30 minutes, after which 1.5 µl of 10X loading buffer (200 mM Tris-HCl pH8.2, 50% glycerol, and BPB 0.25%) was added and the reaction was left on ice for 5-15 minutes before loading into the gel. The reaction complex was run on an 8% polyacrylamide gel under non-denaturing conditions. Mini-gels were made with stock solutions of 40% acrylamide/bis-acrylamide (29:1), 1X Tris-borate EDTA (TBE), 10% ammonium persulfate (APS), and tetramethylethylenediamine (TEMED). Gels were run using the Mini-PROTEAN Tetra Cell apparatus from BioRad. Pre-run was done in 1X TBE buffer for 1 hour prior to loading of the samples. The samples were run at 150 V until the bromophenol blue dye reached the bottom of the gel. The radioactivity in the gel was

analyzed by Phosphor Imager (Molecular Dynamics – Typhoon Trio Imager). After the initial EMSA, binding conditions were slightly changed. Apart from the above mentioned 6X buffer, an additional 5X binding buffer containing 100 mM Tris-HCl pH 8.5, 250 mM NaCl, 10 mM $MgCl_2$, 10 mM $ZnCl_2$, and 10% glycerol was added in the incubation mixture. Furthermore, the radio-labeled oligonucleotides were added directly to the incubation mixture and not heated to 94 ^{o}C as done previously. Lastly, the incubation was carried out at room temperature for 1 hour and 30 minutes, after which the samples were left on ice for 5-15 minutes on the 4% non-denaturing gels which were pre-run for 1 hour in 1X TBE as described above.

3.3.6 Competition assays

Increasing concentrations of the specific unlabeled oligonucleotide was added during incubation of the ligand (BSA)-aptamer complex (conditions described above). Three different amounts (1.25, 2.5, and 5 pmol) were used. Complex formation was allowed to go on for 30 minutes at room temperature before the radioactively-labeled oligonucleotide was added and the incubation was continued for another hour and 30 minutes. In the control sample all the conditions were similar except for the absence of the unlabeled competing oligonucleotide.

3.3.7 Supershift assays to determine specificity of binding of BSA to the potential aptamer

EMSAs were set up using 4 µg of the anti-BSA polyclonal antibody (obtained from Invitrogen). Essentially, the antibody (4 µg) was mixed with BSA (500 µM) along with the 5X binding buffer and left on ice for 1 hour. The radioactively-labeled I1-5 aptamer and the 6X buffer were added next and the incubation continued for 1 hour and 30 minutes at room temperature before stopping the incubation by leaving the samples on ice for 5-15 minutes. Next, the EMSA proceeded as previously described.

3.4 RESULTS AND DISCUSSION

3.4.1 CE machine calibration and initial runs

The general scheme of RT-PCR coupled CE-SELEX is shown in **Figure 3.1**. To test the concept that aptamer selection with CE-SELEX and RT-PCR could be feasible, we first began by calibrating the capillary electrophoresis (CE) machine. The CE machine used, the Proteomelab PA 800, is incapable of running solutions that contain high concentrations of salts. Due to this limitation, our buffers had to be tested to find the proper salt balance to ensure that the machine did not fail but at the same time have enough salt to stabilize our aptamer structures. Therefore the buffer condition used for the initial step of folding the DNA library was done in a high salt buffer (50 mM Tris-HCl at pH 8.2, 100 mM NaCl and 5 mM $MgCl_2$), followed by incubation and running the folded DNA library in low salt concentrations (50 mM Tris-HCl at pH 8.2, 6 mM NaCl, and 300 nM $MgCl_2$) to ensure both proper folding and running of the DNA library. Although the low salt problem may have excluded many potential binders to BSA, low salt concentrations actually mirror the physiological conditions more accurately since the cellular level of magnesium is only ~1-2 mM (Carothers, Goler et al. 2010). After optimizing for the salt concentrations we used CE and ran a large amount (100 µM) of BSA alone in order to visualize the free protein peak run timing. Next we performed the same analysis using a large (3 µM) amount of DNA alone.

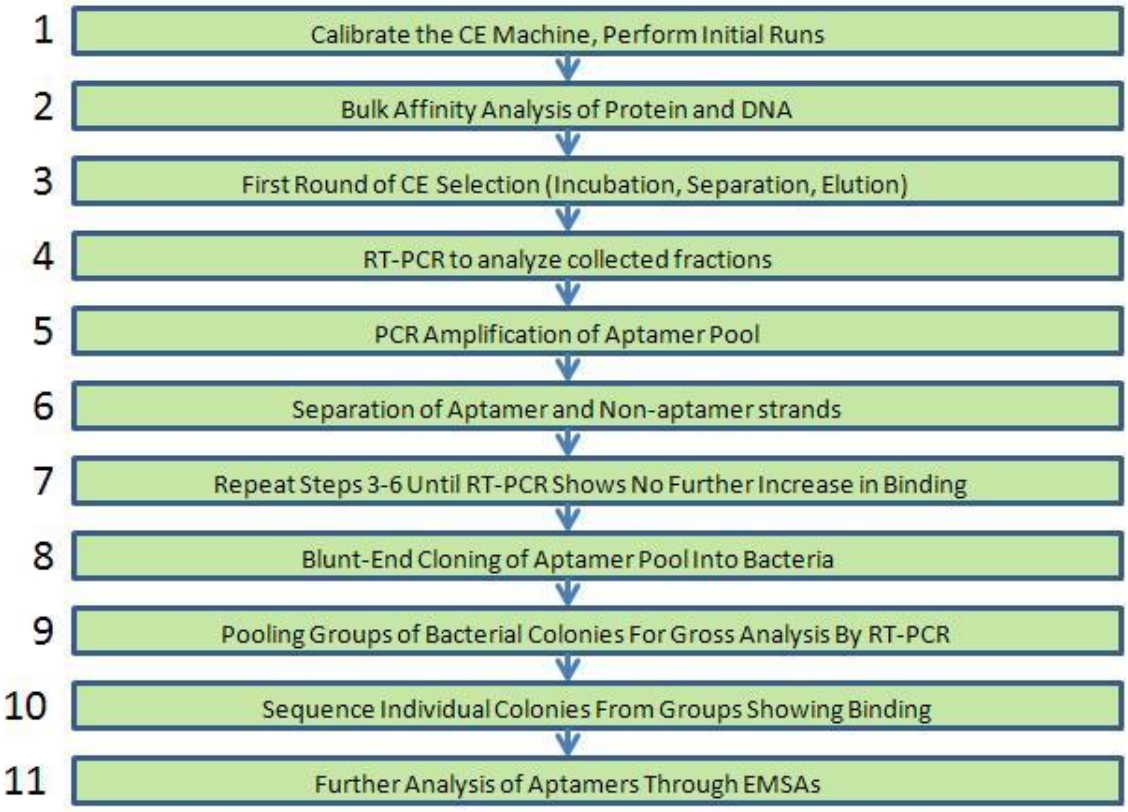


Figure 3.1 Summary of RT-coupled CE-SELEX The overall scheme of the protocol for selection is laid out sequentially. The scheme omits the steps after selection of the aptamers. Further analysis may be carried out to confirm the specificity and affinity of the aptamers, but this is specific to the user and separate from the selection

3.4.2 Bulk affinity analysis and selection

Once the individual free protein and free DNA run times were established, we proceeded to combine BSA and DNA for a bulk affinity analysis **Figure 3.2**. The bulk affinity analysis allowed us to visualize peaks for the free BSA and the free DNA in combination, and this information was used to determine the collection window. The CE machine used was the Proteomelab PA 800 from Beckman Coulter, which did not have laser-induced fluorescence (LIF). To sensitively detect protein-DNA complexes between the free DNA and free protein, we relied on RT-PCR. The collection window started from the end of the free protein peak to the start of the free DNA peak **Figure 3.3**. This collection window was further subdivided into seven separate two-minute fractions. The seven fractions, instead of just one fraction, allowed us to analyze with greater precision the region of the DNA-protein complex. In the first round of selection the concentration of BSA was reduced to 500 nM in order to increase the stringency of selection. In the first round of selection the initial collected fraction of DNA detected by RT-PCR needed36 cycles of amplification to reach 50% of the maximum yield, the midpoint cycle, of PCR **Figure 3.3**. The second fraction analyzed had a higher amount of DNA compared to the first fraction collected, taking only 34 cycles of amplification to reach the midpoint cycle. The amount of DNA collected for the third and fourth fractions was lower than that of the second fraction collected. The fifth fraction collected showed a marked increase in DNA indicative of the beginning of the free DNA peak and so this fraction was not used. The pattern detected by RT-PCR implied a DNA-protein complex present in the second

fraction collected since this fraction was collected by CE after the free protein peak, but well before the free DNA peak.

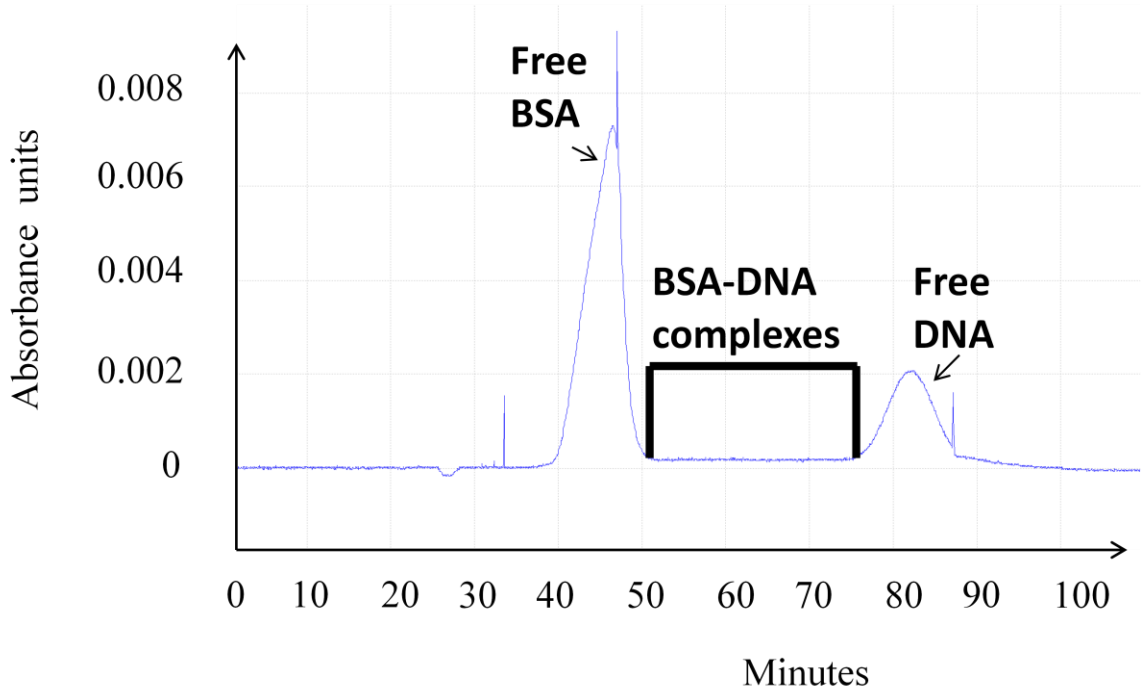


Figure 3.2 Bulk affinity assay of BSA and DNA This electropherogram is the bulk affinity assay in which a high concentration of BSA (100 μM) and DNA (3 μM) was run together. The *y*-axis represents the absorbance at 280 nm, and the *x*-axis is the run time to the absorbance detector. The tall peak starting around 40 minutes is the free BSA, and the smaller peak starting around 77 minutes is the free DNA. The collection window for the BSA-DNA complexes is boxed in black.

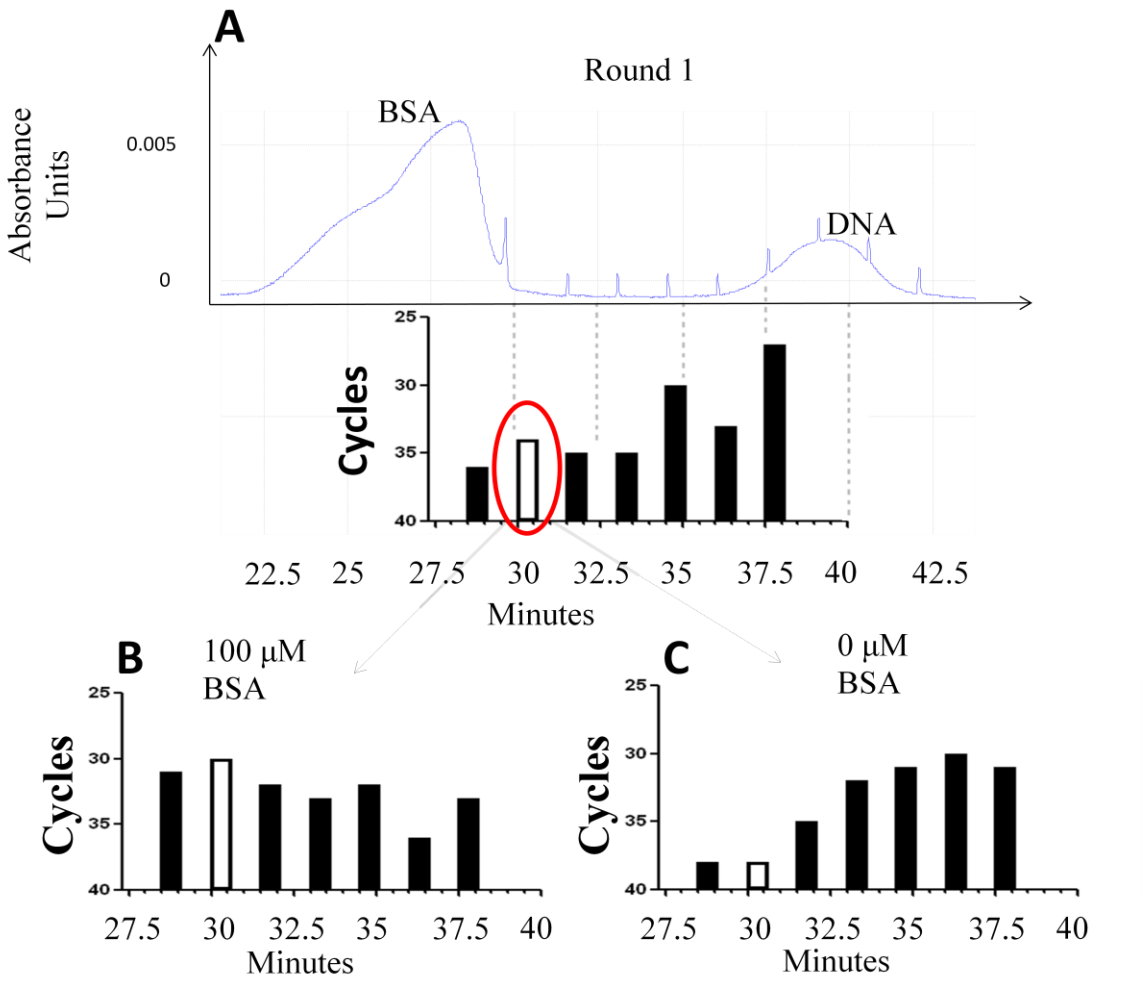


Figure 3.3 CE fractions collected and analyzed (a) The time points at which the seven fractions were collected during the first round of CE selection are represented by the bars in the graph. The *x*-axis for the bar graph and the electropherogram is the run time to the absorbance detector in minutes. The *y*-axis of the electropherogram represents the absorbance at 280 nm. The *y*-axes for the bar graphs are the number of RT-PCR cycles it took to reach 50% of the maximum or the midpoint cycle. The earlier the fraction reached the midpoint, the more DNA the fraction contained. The electropherogram above the bar graph is shown as a guide since the actual amount of BSA used for round 1 of selection is very low and it cannot be detected by CE. A supposed complex was detected in the second fraction collected during round 1 of selection (white bar). The supposed complex in the 2nd fraction of round 1 was analyzed in two test rounds using 100 μM BSA (b) or no BSA (c). Seven fractions were collected at the same time points as the seven fractions shown in (a) and the corresponding RT-PCR bar graphs of the midpoints are shown, with the 2nd fraction collected shown in white.

The collected DNA containing the supposed DNA-protein complex was RT-PCR amplified using primers flanking the internal random sequence. The forward primer, P1, amplified the aptamer-containing strand of interest, while the reverse primer, P2, amplified the non-aptamer complementary strand. In a previous protocol to select DNA aptamers, it was shown that over-amplification of the random library leads to formation of non-specific products (Berezovski and Krylov 2005), therefore the random library was only amplified to ~50% of the maximum yield as measured by RT-PCR. After RT-PCR analysis revealed the optimum number of cycles to amplify the collected DNA, a regular PCR using more of the collected DNA was performed. In this regular PCR, the P2 primer was labelled at its 5' end with biotin while P1 had no modifications. After the PCR, the DNA products were attached to streptavidin-coated beads due to the strong interaction between the streptavidin on the beads and the biotin-labelled primer P2. After immobilization on the beads and several washing steps, the forward aptamer-containing strand was released from the complementary strand by incubation with 10 mM NaOH. After this step, the binding of the new aptamer pool was analyzed using CE and RT-PCR.

To assess the binding of the newly generated aptamer pool to the target, two separate CE runs were done, one with a 200-fold molar excess of BSA used for the first round of selection (100 µM) (**Figure 3.3**) and another without any BSA (**Figure 3.3**). Based upon these CE test runs it was clear that the second fraction collected in round one of CE-SELEX did show binding to BSA. The aptamer pool (5 µl of the collected fraction) was then subjected to another round of selection. Selection continued until the fourth round, at which point RT-PCR did not show any change in the amount of DNA at the complex and we did not continue with any further rounds of selection.

3.4.3 Post-selection characterization of potential aptamers

Potential aptamers collected during the third round of selection were again amplified and then used for cloning. Instead of amplifying with a biotin-labelled P2 primer, the potential aptamers were amplified with a non-labelled P2 primer since the biotin label at the 5' end of the DNA might interfere with ligation during cloning. TOPO blunt-end cloning was performed with the aptamer pool and colonies were selected by growth on LB plus kanamycin plates. Two hundred colonies were selected and analyzed for BSA binding using CE and RT-PCR. Groups of 12 colonies were combined in order to facilitate faster analysis. In order to amplify the aptamer strand from the vector with the aptamer insert, asymmetric PCR was performed using a 10-fold excess of the aptamer strand primer. Four groups, I1, R2, G2, and O3 showed stronger binding than the others by RT-PCR and then the individual clones from these groups were sequenced.

After sequencing, in order to confirm that the potential aptamer sequences were binding to BSA, salt-free oligonucleotides containing the potential aptamer sequences were used in EMSA gels. EMSAs work on the principle that DNA bound in a DNA-target complex will run slower than free DNA. The free DNA will appear at the bottom of the gel, while the DNA bound in the DNA-target complex will be shifted higher up in the gel. The potential aptamers, as well as a negative control (described in Materials and Methods), were initially screened with a set of binding conditions similar to that of the CE experiments. Only those sequences that were 76 or 75 bases in length, and identical or almost identical in length to the original library length of 76 bases, were ordered and

subsequently screened by EMSA. The candidate aptamer sequences were analyzed *in silico* by using ClustalW2 and mFold, but no consensus sequence or specific secondary structure between the sequences was found.

Of the initial group of oligonucleotides screened by EMSA, the I1 group, one member of this group of potential aptamers, I1-5, did bind to BSA and a distinct bound complex could be visualized at 500 and 800 μM BSA in the initial EMSA (**Figure 3.4**). The bound complex showed a modest, but distinct shift. Also in the initial EMSA, at higher concentrations of BSA (>500 μM), a faint band could be seen. This band was considered a nonspecific complex because both the negative control and the I1-5 aptamer had this band. In this EMSA, and in all subsequent EMSAs, the appearance of radio-labeled DNA at the top of each lane, in the well, can be seen. The DNA at the top of the lanes is caused due to the single-stranded nature of the DNA. The single-stranded DNA is capable of forming aggregates with itself and with the loading dye. The size of these aggregates prevents the DNA from entering the gel and thus it remains in the well. After the initial EMSA modeled after the CE experiments, binding conditions were changed to include higher salt conditions that were not possible to run with the Proteomelab PA 800. These higher salt conditions were used to stabilize the aptamer secondary structure. In addition to buffer conditions, we also did not subject the aptamers to the heating at 94 ºC prior to incubation. Again the potential aptamer I1-5 showed a bound complex using these modified binding conditions (**Figure 3.4**). After the initial screen for potential aptamers with the I1 group, a subsequent screen using oligos from the G2 group and the modified EMSA conditions revealed several additional aptamers for BSA (**Figure 3.5**). The list of the potential aptamers' sequences ordered is shown in **Table 3.2**, among these aptamers,

I1-5 was chosen for further characterization.

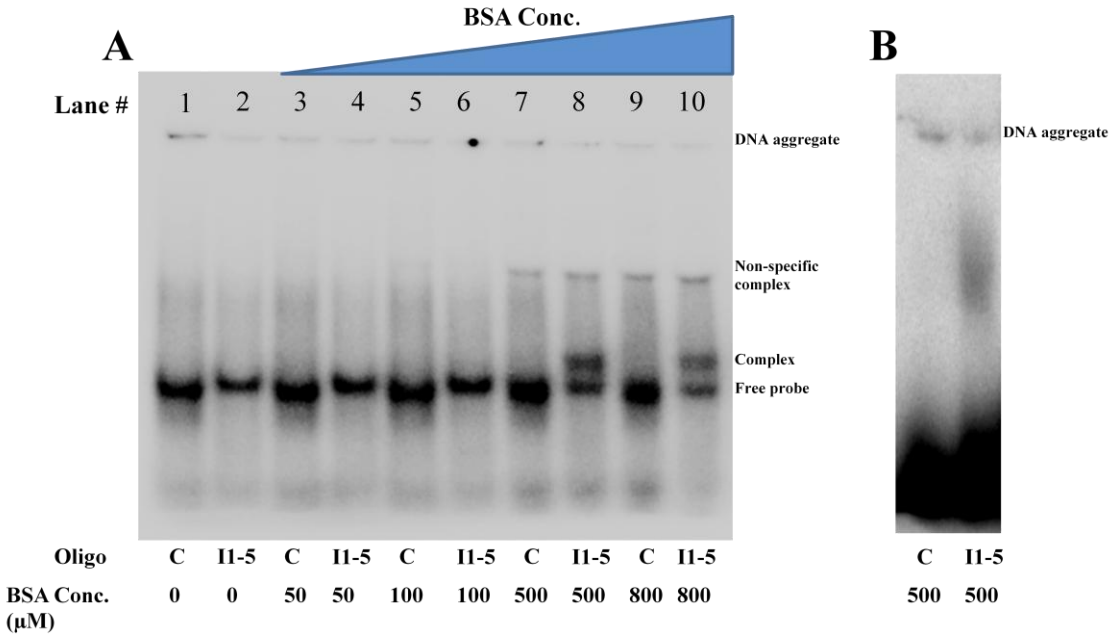


Figure 3.4 Confirmation of CE-SELEX aptamer selection by EMSA (a) Shown is the initial EMSA done with the BSA aptamer, I1-5, selected through RT-PCR coupled CE-SELEX, along with a negative control of the same length ("C" in the figure) described in Section 2. At 500 μM BSA a specific complex, formed between the BSA and DNA. Also, there is a higher band corresponding to nonspecific interaction at high concentrations of BSA in both the negative control and the aptamer I1-5 sequences. (b) EMSA of I1-5 and negative control with the modified binding conditions. The binding pattern was the same with these two different binding conditions and for clarity only the BSA at 500 μM is shown.

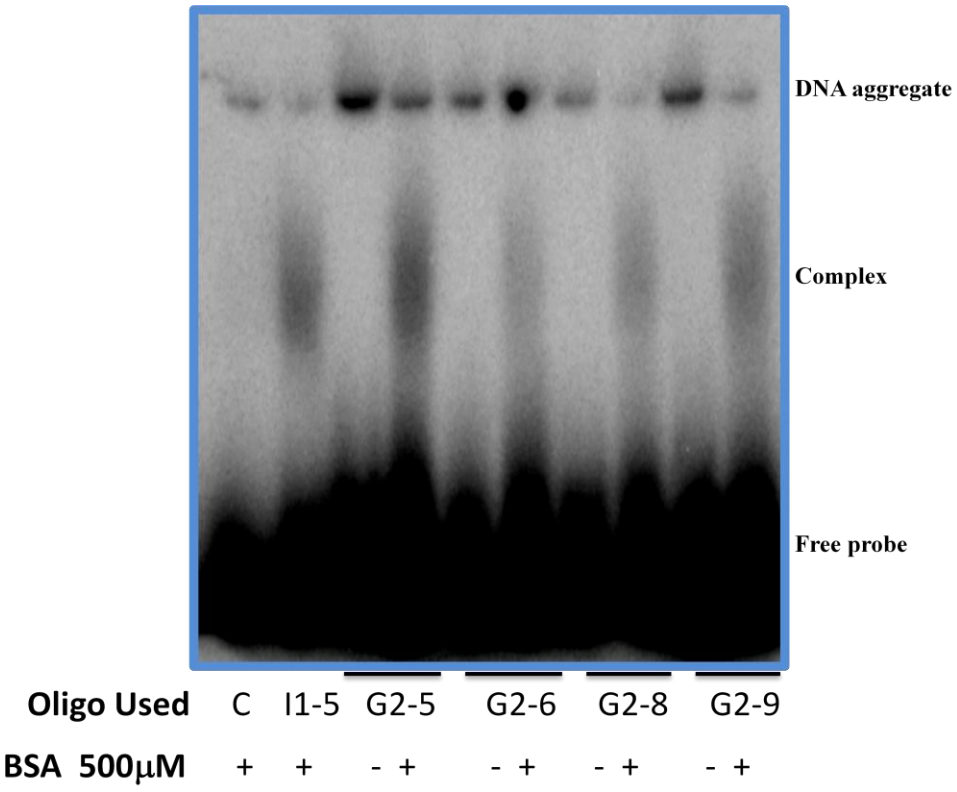


Figure 3.5 EMSA gel confirming binding of group G2 aptamers Shown is a screen done with potential aptamers from the second group tested, G2. Several from this group were capable of binding to BSA. I1-5 was used as a positive control for binding, and the negative control ("C") was also used.

Potential Aptamer	Internal Sequence
O3-9	CAGTCACCTCAGCGTCCACAATTATGGCAGCGCGCG
G2-1	TTTATTGGGGGAATATAACCGATCATGTGGCGGTTA
G2-2	ATACAATGATTTGCACACCGGGTGGCGTGAATTGT
G2-3	GGATTGTGGTAAGCTCTATCTACCTTGCTACGCCC
G2-4	GGCTGATCGACAATAAGCGTGCCGGCCCTCAACTCC
G2-5	**TCCGATGCCACACTTGACGTGCTGGGTTGGCTGCGT**
G2-6	**AGCGAGAGGCGACTTCATTATTGGGAGGTGGCTTCG**
G2-8	**AGTAGGTGTTATGAGCACGACCGGTTTACTTCACG**
G2-9	**GGAGTGACTGGAGGATCTGGTCCCCGGAGTTCCTTG**
G2-11	GCATGTGGTGCTTCTTGACCGCCTGCGCCTGCGGGT
R2-6	AACGACTGGACTCCGTCATACAGTTTGGGGAAAGG
R2-8	TCCCTACCCGGTGTTTGAAACGGCCGTAAGGAACTT
R2-10	GATAGTAAGACCTGATTGGGGTCAAGCTAACGTCGA
R2-11	GCTCGCGTTGTAAGCATTTGTTGGNCGAGATAGGCG
I1-5	**GCCCGCCGTGGCTGGGTCTTCCTTGGTCGGTCTAC**
I1-11	GTATCTACAGGGTGGCCAGGGCGCAACCGGGCGAGT
I1-12	TGGCTAGCGATGTATTCCGTTTTGGGGATAAATCCA

Table 3.2 Sequences of potential aptamers for BSA The table contains the sequences of the potential aptamers from each of the pooled groups of clones that showed binding to BSA by RT-PCR analysis. These sequences were ordered as salt-free oligonucleotides and analyzed using electrophoretic mobility shift assays (EMSAs), and sequences that showed binding are bolded. I1-5, boxed in black, was chosen for further analysis.

3.4.4 Further characterization of I1-5 aptamer

Only the I1-5 oligonucleotide was further pursued as it was the first aptamer discovered by the screen that consistently showed binding ability to BSA in the CE-SELEX as well as under the two different EMSA conditions. To further confirm the specificity of this aptamer to BSA, a supershift assay using an antibody to BSA was performed. In the presence of 4 µg of the anti-BSA polyclonal antibody, the binding capacity was reduced as seen by an appreciable reduction in the intensity of the radioactive oligo banding pattern (**Figure 3.6**). This supports the idea that the anti-BSA polyclonal antibody and the BSA aptamer I1-5 recognize the same epitope. A similar pattern of banding was reported for the human neurofilament mRNA binding to superoxide dismutase1 (SOD1) in a supershift assay (Ge, Wen et al. 2005). Also, competition assays were performed by incubating BSA with an excess of unlabeled I1-5 aptamer of three different amounts (1.25, 2.5, and 5 pmoles) before the addition of the labeled I1-5, the BSA-aptamer bound complex was abolished in the presence of higher amounts of the unlabeled competitor (**Figure 3.7**). Although by EMSA the I1-5 aptamer showed low binding affinity to BSA, the cold-competitor assay, as well as the supershift assay indicated a specific interaction was occurring between BSA and the I1-5 aptamer, thus supporting the effectiveness of RT-PCR coupled CE-SELEX.

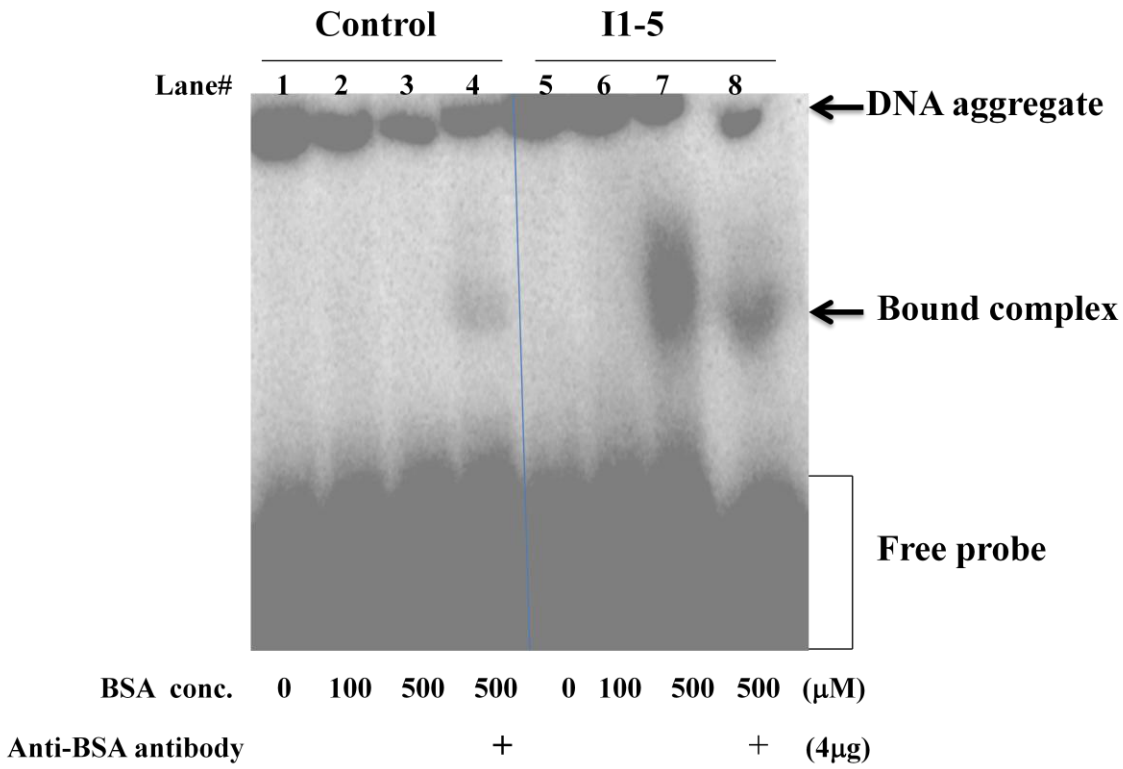


Figure 3.6 Supershift assay with inclusion of antibody to BSA binding in gel retardation assays A polyclonal antibody to BSA was tested for binding to BSA in the reaction complex and analyzed by EMSA. The negative control previously shown not to bind to BSA was used as well as I1-5, the BSA aptamer. For the negative control, no binding to BSA was seen at 100 μM or 500μM BSA (lane 2 and 3). With addition of 4 μg of the antibody, there was some nonspecific binding, as seen by the faint band, by the control sequence (lane 4). However, with I1-5 there was binding to 500 μM BSA (lane 7), but this binding was disrupted when the antibody was incubated with BSA prior to the EMSA (lane 8). As stated previously, some of the radiolabeled DNA formed aggregates incapable of entering the gel, giving rise to the bands at the top of the lanes. A slight leak occurred during gelling such that the top of lane 8 appears disjointed due to the shape of the gel.

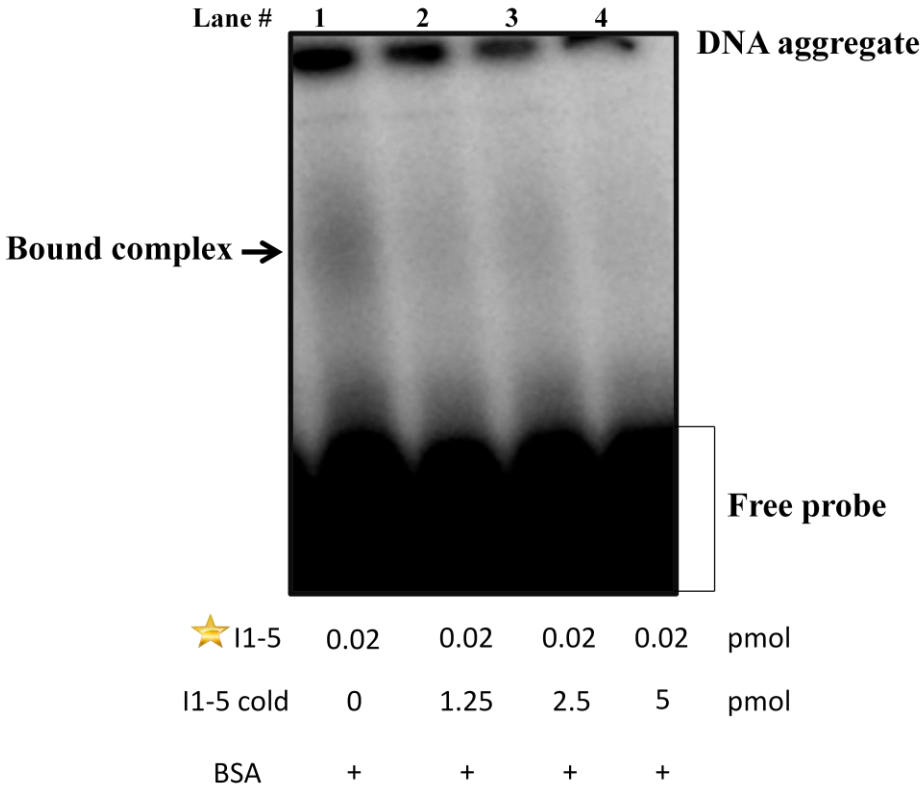


Figure 3.7 Competition assay with unlabeled I1-5 aptamer Unlabeled "cold" I1-5 aptamer at 0, 1.25, 2.5, and 5 pmol as indicated were incubated with 500 μM BSA (indicated by the "+") before addition of the P32-labeled I1-5 (represented by a star) aptamer. Again, the formation of DNA aggregates is present at the top of the lanes. As unlabeled DNA competitor increases, the amount of radiolabeled DNA trapped in the aggregates decreases as expected.

3.5 CONCLUSIONS

The conclusion drawn from our studies is that we have been able to generate an aptamer for BSA with appreciable specificity in a few rounds of selection from a random library of DNA oligonucleotides.

The protocol described here is very efficient compared to the traditional SELEX which takes much more time and reagents. As noted previously, traditional SELEX makes use of a solid support to which the target must be bound. The coupling of the target to the support is not completely efficient and often much of the target remains unbound and washed away. Using CE-SELEX there is no need for a solid support and because of the small size of the capillary, a much smaller amount of target is required. Even though CE-SELEX is much more efficient than traditional SELEX, CE-SELEX still requires additional equipment (LIF) beyond the standard capillary electrophoresis machine. CE is not restricted to the isolation of aptamers, and while many labs may have a CE machine, much fewer will have one with LIF. Our system can be used with any CE machine, with or without LIF.

Another advantage of the system is the high level of sensitivity. RT-PCR and LIF both work under the principle of fluorescence. While LIF can very sensitively detect DNA-target complexes directly, RT-PCR has the potential to be more sensitive because unlike LIF which directly detects the fluorescent signal, RT-PCR can amplify the signal. By amplifying a signal undetectable by LIF, DNA-target complexes that would have been missed by LIF could be detected and collected using our system. Taken together, the efficiency, time, and sensitivity of the protocol for DNA aptamer selection by coupling

CE-SELEX with RT-PCR described here can benefit other researchers that are also interested in selecting for aptamers and would like to use CE-SELEX, but are unable to do so for lack of LIF.

BSA was chosen to test our RT-PCR coupled selection system, but the BSA aptamer we selected may be beneficial itself. Due to the high amount of BSA used in many biochemical assays, sometimes the amount of BSA present in a mixture of proteins obscures accurate readings of proteins of interest. It would be useful to have a way to deplete this unwanted BSA from a reaction. Also, BSA can be allergenic (Natale, Bisson et al. 2004) and in certain situations it might be beneficial to detect small amounts of BSA contaminant present in cow's milk. Currently antibodies to BSA provide an answer to these issues, but as previously mentioned there are several advantages to using aptamers over antibodies.

3.6 ACKNOWLEDGEMENTS

We would like to acknowledge David Scarborough and Grant Wise from Beckman Coulter for their help in troubleshooting the CE machine. This work was supported in part by the Georgia Cancer Coalition grant R9028 and the NIH grant R21EB9228.

CHAPTER 4

PROTEIN-ASSISTED TARGETING OF GENES BY A DNA APTAMER TO I-*SCE*I

The work presented in Chapter 4 is part of a research article under the same name in preparation for submission to the journal Nature Biotechnology:

Ruff[1], P., Pai[1], R.B., Pohl[2], J., and Storici[1], F. (2013) [1]School of Biology, Georgia Institute of Technology, Atlanta, GA. [2]Biotechnology Core Facilities, Center for Disease Control and Prevention, Atlanta, GA. (in preparation)

4.1 ABSTRACT

DNA aptamers are sequences of DNA selected for their ability to bind a specific target. We selected a DNA aptamer to the I-*Sce*I endonuclease, which binds an 18-bp DNA sequence and generates a DNA double-strand break (DSB). Bifunctional oligonucleotides containing the I-*Sce*I aptamer sequence or a non-binding control of equal length were generated as part of a longer DNA molecule that contained a sequence tract with homology to repair the I-*Sce*I generated DSB in the chosen genomic locus. The aptamer portion binds I-*Sce*I which effectively targets the entire DNA molecule to the site of the break. In yeast the I-*Sce*I aptamer stimulated gene targeting two to fifteen-fold over the non-binding control. In addition to the work in yeast, correcting DNA oligonucleotides containing the I-*Sce*I aptamer sequence were used for gene targeting in human cells and showed stimulation of gene targeting two to sixteen-fold over the control oligonucleotides. Taken together, this work provides a novel strategy to increase gene targeting efficiency and lays the groundwork for future studies using aptamers for gene targeting.

4.2 INTRODUCTION

Targeted gene modification in human cells would be a powerful tool for researchers interested in functional analysis of genes as well as patients suffering from genetic disorders. The primary limitation of gene targeting is the low frequency with which it spontaneously occurs in mammalian cells, happening in roughly 1 cell for every 10^5 to 10^7 treated cells(Vasquez, Marburger et al. 2001). The low frequency of gene targeting, which relies on homologous recombination (HR), is due in part to the much higher frequency of non-homologous end joining (NHEJ), which occurs in roughly 1 cell for every 10^2 to 10^4 treated cells(Vasquez, Marburger et al. 2001). HR and NHEJ are the two major pathways known to repair a DNA double-strand break (DSB), which if left unrepaired is lethal to the cell. NHEJ joins the two ends of the DSB in a sequence independent manner and can even insert foreign DNA at the site of a DSB in a process known as random integration. Random integration by its nature results in unpredictable outcomes, which can include *cis* or *trans*-activation of a gene downstream of where the transgene was inserted, disruption of regulatory elements, or creation of aberrant fusion proteins all of which could lead to diseases such as cancer(Kohn, Sadelain et al. 2003). Gene targeting, in which a sequence homologous to the DNA at the DSB is used to repair the DSB, offers a safer and more precise means to integrate a gene of interest or modify an existing gene.

As stated, the level of spontaneous gene targeting is low, but there are several strategies used to increase the frequency of gene targeting. It was shown that a DSB at the target site increases the frequency of gene targeting several orders of magnitude in

bacteria(Nussbaum, Shalit et al. 1992), yeast(Storici, Durham et al. 2003), plants(Puchta, Dujon et al. 1993), fruit flies(Banga and Boyd 1992), mice(Rouet, Smih et al. 1994), human embryonic stem cells(Smih, Rouet et al. 1995), and many other cell types. Another strategy to increase gene targeting in mammalian cells has been achieved through the overexpression of key recombination proteins from HR proficient organisms. Overexpression of bacterial RecA led to a 10-fold increase in gene targeting in mouse cells(Shcherbakova, Lanzov et al. 2000), likewise overexpression of yeast Rad52 led to a 37-fold increase in gene targeting in human cells(Di Primio, Galli et al. 2005). Conversely, another strategy for increasing gene targeting in human cells involves decreasing the amount of the DSB repair through the pathway of NHEJ. In mouse embryonic stem cells an increase in gene targeting was seen in Ku70 (6-fold), XRCC4 (2-fold), and DNAPK-cs (2-fold) deficient cell lines(Pierce, Hu et al. 2001) and a 3-fold increase in Chinese hamster ovary cells lacking DNAPK-cs(Allen, Kurimasa et al. 2002). Similarly, knockdown of KU70 and XRCC4 in human colon cancer cells lead to a 30-fold increase in gene targeting(Bertolini, Bertolini et al. 2009).

Different from the strategies mentioned above focused on increasing HR or decreasing NHEJ, it was shown that by knocking down human SMC1, important for HR, gene targeting increased(Potts, Porteus et al. 2006). The sister chromatid is the normal donor DNA for HR repair, but in the case of gene targeting an exogenous DNA acts as the donor for repair. HR with the sister chromatid actually hinders gene targeting by exogenously introduced DNA. By knocking down hSMC1 which is required for sister chromatid HR, gene targeting increased four-fold. hSMC1 and hSMC3 form the cohesin complex which is responsible for keeping sister chromatids in close proximity to each

other during a DSB. Without close proximity to the DSB site the sister chromatid was used less frequently as a donor, shifting repair of the DSB more toward HR with the exogenous sequence (gene targeting).

We have developed a novel delivery system, which we call protein-assisted targeting, to increase gene targeting by guiding the exogenous DNA to the site of modification. The system is designed to increase gene targeting by increasing the amount of exogenous DNA available at the target site, while simultaneously reducing random integration of exogenous DNA outside of the target. By tethering the exogenous DNA to the site-specific homing endonuclease I-*Sce*I, we achieve targeted delivery of exogenous DNA to the site of the DSB (**Figure 4.1**). DNA aptamers were selected for I-*Sce*I using a variant of capillary electrophoresis systematic evolution of ligands by exponential enrichment (CE-SELEX) called "Non-SELEX" (Berezovski, Musheev et al. 2006). DNA aptamers are sequences of DNA that because of their unique secondary structure are able to bind to a specific target with high affinity. By synthesizing a DNA oligonucleotide that contained the aptamer sequence as well as homology to restore the DSB, we were able to increase gene targeting frequencies two to fifteen-fold over a non-binding control in yeast and two to sixteen-fold over a non-binding control in human cells. Our strategy offers a novel way to increase gene targeting and represents the first study to use aptamers in the context of gene repair.

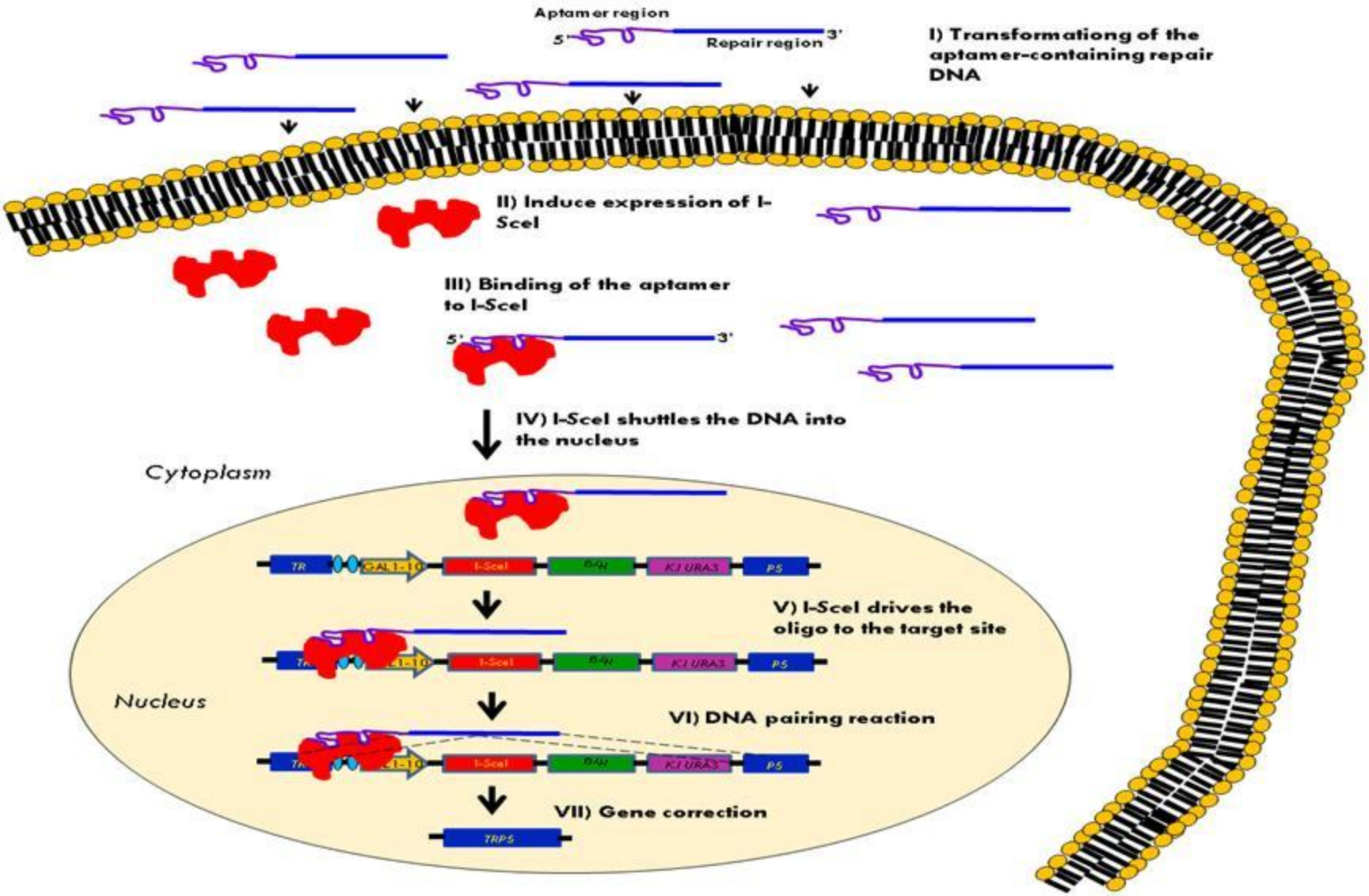

Aptamer region
5'
Repair region
3'
I) Transformationg of the aptamer-containing repair DNA
II) Induce expression of I-SceI
III) Binding of the aptamer to I-SceI
5'
3'
IV) I-SceI shuttles the DNA into the nucleus
Cytoplasm
TR
GAL1-10
I-SceI
BAR
KI.URA3
P5
V) I-SceI drives the oligo to the target site
Nucleus
VI) DNA pairing reaction
VII) Gene correction
TRP5

Figure 4.1 Protein-assisted targeting model (for yeast *TRP5*) Above is the model for protein-assisted targeting in yeast. Briefly, oligonucleotides containing the ISB7 aptamer along with a region of homology to restore the *TRP5* gene are transformed into the cell. Simultaneously to the oligonucleotide transformation, the I-*Sce*I gene under the *GAL1-10* promoter is expressed. The ISB7 aptamer then binds to the I-*Sce*I protein, which shuttles the oligonucleotide into the cytoplasm. I-*Sce*I drives the oligonucleotide to the *TRP5* locus and causes a DSB. Afterwards the oligonucleotide restores the function of the *TRP5* gene.

4.3 MATERIALS AND METHODS

4.3.1 Aptamer Selection

The target protein of interest, I-*Sce*I, was kindly provided by Dr. Frederick Gimble in storage buffer (10mM KPOi pH 7.4, 100nM EDTA, 1mM DTT, 100mM NaCl, and 50% glycerol). Prior to selection, in order to remove storage buffer components, I-*Sce*I was dialyzed. I-*Sce*I was dialyzed in Run Buffer 1 (RB1), 50 mM Tris-HCl at pH 8.2, yielding a concentration of 3 µM I-*Sce*I in RB1 post dialysis. RB1 was the run buffer used for the capillary electrophoresis. The DNA library was purchased from Alpha DNA (Montreal, Quebec) and contained a sequence of

5'_CTTCTGCCCGCCTCCTTCC-(N)36-GACGAGATAGGCGGACACT_3' (36 random nucleotides flanked by two fixed 19 base regions used later as primers for PCR amplification).

The protocol for SELEX using capillary electrophoresis (CE) was essentially as described earlier [10] but with a few modifications. Initial calibrations were done with a serial dilution of the aptamer library in RB1. The initial bulk affinity assay was performed with 1.5 µM I-*Sce*I and 100 nM DNA in order to view any DNA-protein complexes and determine the collection window. Capillary electrophoresis was done using a Beckman Coulter P/ACE MDQ with laser-induced fluorescence (LIF) detection. The LIF was composed of a 488 nm air-cooled argon ion laser along with an on-board detector. Separation was carried out with a voltage of 10 kV. After determination of the collection window based on the bulk affinity analysis, the first round of selection began. The initial round of *in vitro* selection procedure involved 100 nM I-*Sce*I and 50 µM

DNA. The DNA library (5 µl at 100 µM) was mixed with 5µl of SB3 (100 mM Tris-HCl at pH 8.2, 200 mM NaCl, and 10 mM $MgCl_2$) for a final concentration of 50 µM DNA library, 50 mM Tris-HCl at pH 8.2, 100 mM NaCl, and 5 mM $MgCl_2$. This mixture was heated in the BioRad iCycler to 94 °C for 1 minute, and then cooled to 20 °C at a rate of 0.5 °C/second. After the folding of the DNA library, 5 µl of 200 nM I-*Sce*I dissolved in SB1 buffer (50 mM Tris-HCl at pH 8.2, 100 mM NaCl, and 5 mM $MgCl_2$) was added to make the final volume 10 µl. This brought the final concentrations to 50 µM DNA library, 100 nM I-*Sce*I, 100 mM NaCl, 5 mM $MgCl_2$, and 50 mM Tris-HCl (pH 8.2). The collection window was from the beginning of the first complex peak to the end of the last complex peak, well before the free DNA peak. The fraction collected was typically 0.3 µl to 0.5 µl which was collected into a tube containing 10 µl of the above mixture except without any additional DNA. After 15 minutes of incubation at room temperature, this new mixture was used in subsequent rounds of selection. Two additional rounds of selection proceeded in this manner.

4.3.2 qRT-PCR

After the aptamer selection, the collected fractions containing the aptamer pools were analyzed through quantitative real-time PCR (qRT-PCR) using the ABI StepOne Plus. qRT-PCR was done with two primers, the forward aptamer-amplifying primer P1 (5'_CTTCTGCCCGCCTCCTTCC_3') and the reverse primer P2 (5'_AGTGTCCGCCTATCTCGTC_3') respectively. The primers were designed using OligoAnalyzer (http://www.idtdna.com/analyzer/Applications/OligoAnalyzer/) to limit complementarity to each other, in order to decrease non-specific amplification of self-

dimerizing primers. For amplification, 20 µl of PCR mix was prepared consisting of 10 µl of 2X Quanta SYBR Green PCR Master Mix (Roche), 0.6 µl of 10 µM P1, 0.6 µl of 10 µM P2, 1 µl of collected fraction as template, and 7.8 µl H2O. The PCR setup is described in **Table 4.1**.

Table 4.1 Real-Time PCR cycling conditions

Cycle #	Denaturation	Annealing	Extension
1	94 °C for 30 seconds		
2-50	94 °C for 10 seconds	55 °C for 10 seconds	72 °C for 10 seconds
51			72 °C for 1 minute
Hold at 4 °C			

4.3.3 Amplification and isolation of aptamer strand

Following qRT-PCR, the fraction containing the potential aptamers was amplified using standard PCR. PCR was done in a 100 µl volume consisting of 1 µl of 5U/µl X-Taq polymerase from Takara, 3 µl of 10 µM forward primer P1, 3 µl of 10 µM reverse primer P2, 10 µl of 10X Mg^{2+} buffer (Takara Ex Taq), 8 µl of 2.5 mM each dNTP, 70 µl H2O and 5 µl of the collected fraction from capillary electrophoresis. PCR was done using primer P1 and P2 as noted previously, except that the number of cycles for PCR amplification was based on 50% of the maximum yield as determined by qRT-PCR. Additionally, primer P2 was biotinylated at its 5' end. The biotin-labeled primer was used subsequent to PCR in order to separate the strand of interest and the non-aptamer strand after PCR. Magnetic beads with streptavidin coating from Bangs Laboratories (Biomag Streptavidin Nuclease-Free) were used to bind the biotin-labeled DNA. Strands were separated with 10 mM NaOH after three washes with wash buffer (10 mM Tris-HCl at pH 8 with 500 mM NaCl and 1 mM EDTA). The single-stranded aptamer pool was used in subsequent cloning.

4.3.4 Cloning and sequencing

Post-selection DNA cloning of the aptamer pool was done with the TOPO Zero Blunt Cloning Kit (Invitrogen). Standard PCR with unlabeled primers P1 and P2 was used to generate double-stranded DNA containing the aptamer sequence, which was then blunt-end ligated into the pCR-Blunt II-TOPO vector that contains the kanamycin resistance

gene. Colonies were selected for growth on kanamycin-containing media (kanamycin final concentration was 40µg/mL) and plasmid DNA was isolated using the GeneJET Plasmid Miniprep Kit (Fermentas). Plasmid DNA was crudely extracted by placing selected colonies into 50 µL of RNAse/DNAse-free H_2O and incubated in a boiling water bath for 5 minutes. Debris was pelleted by centrifugation for 10 minutes at 10,000 X g and supernatant was used for an asymmetric PCR. Asymmetric PCR with 5 µL unlabeled primers P1 and P2 were used on the plasmid DNA to predominately generate the strand of interest which was then analyzed using CE with LIF. PCR product was used with 1.5 µM I-*Sce*I in the same manner described previously. Individual plasmids that showed strong binding through asymmetric PCR product were sequenced by Eurofins MWG Operon. Based on sequencing results, several candidate aptamers were chosen and ordered as salt-free oligonucleotides. Consensus sequence was analyzed using ClustalW2 (http://www.ebi.ac.uk/Tools/msa/clustalw2/). Additionally, mFold (http://mfold.rna.albany.edu/?q=mfold/DNA-Folding-Form) was used on each candidate aptamer to identify secondary structure.

4.3.5 Electrophoretic Mobility Shift Assay (EMSA)

Potential aptamer oligonucleotides and a negative control oligonucleotide were 5' labeled with P32 γ-ATP using T4 Polynucleotide Kinase (New England Biolabs). The negative control consisted of an oligonucleotide of the same length as the random DNA library oligonucleotides (74 bases), contained the same flanking primer regions, and had a fixed sequence for its internal region 5'-

CTTCTGCCCGCCTCCTTCCGGTCGGGCACACCTGTCATACCCAATCTCGAGG

CCAGACGAGATAGGCGGACACT-3'. The internal region was chosen using a random DNA sequence generator with a specified GC content of 50% (http://www.faculty.ucr.edu/~mmaduro/random.htm). I-*Sce*I was dialyzed before running the EMSA gels in Run Buffer 1 (RB1), 50mM Tris-HCl at pH 8.2, as previously described. Bovine serum albumin (BSA) was purchased as a lyophilized powder through Sigma-Aldrich and was greater than 98% pure. BSA stock of 10 mg/mL was made in RB1 buffer.

Equal amounts of SB3 (100 mM Tris-HCl at pH 8.2, 200 mM NaCl, and 10 mM $MgCl_2$) and γP32-labeled DNA were added together for a final concentration of 50 mM Tris-HCl at pH 8.2, 100 mM NaCl, and 5 mM $MgCl_2$. Labeled oligonucleotides and SB3 were incubated at 94 ^oC for 1 minute in a PCR machine and a quick chilling to 20 ^oC at a rate of 0.5 ^oC per second (total time taken is ~4 minutes).

The buffer conditions used for binding had several components. Each reaction consisted of 2 µL 5X EMSA buffer 1 (100 mM Tris-HCl at pH 8.5, 250 mM NaCl, 10 µM $ZnCl_2$, 10 mM $MgCl_2$, 10% glycerol), 1 µL BSA (10 mg/ml), 1 µL freshly prepared 20 mM DTT, and 1 µL 100 mM $MgCl_2$ for a final buffer concentration of 20 mM Tris-HCl, 50 mM NaCl, 2 µM $ZnCl_2$, 22 mM $MgCl_2$, 1mg/mL BSA, 4 mM DTT, and 2% glycerol. After mixing these components together, 2 µL of dialyzed I-*Sce*I (3 µM) for each reaction was added, bringing the volume to 7 µL. Reactions were aliquoted and 0.5 µl to 1 µl (20,000 cpm equivalent) of γP32-labeled oligonucleotides were added. The reaction mixture of DNA and I-*Sce*I was incubated for 30 minutes at room temperature. After incubation 2 µl of EMSA buffer 2 (120 mM Tris-HCl at pH 8, 600 mM NH_4Cl, 300 mM

NaCl, 300 mM KCl, 30% glycerol, 0.25% bromophenol blue) was added bringing the final buffer concentration to 44 mM Tris-HCl, 1mg/mL BSA, 4 µM $ZnCl_2$, 160 mM NaCl, 120 mM NH_4Cl, 60 mM KCl, 14 mM $MgCl_2$, 2 mM DTT, 10% glycerol, 0.05% bromophenol blue, and 600 nM dialyzed I-*Sce*I. After addition of EMSA buffer 2 the samples were put on ice until being loaded into the gel. The reactions were run on 4% polyacrylamide gels under non-denaturing conditions. Mini-gels were made with stock solutions of 40% acrylamide/bis-acrylamide (29:1), 1X Tris-borate EDTA (TBE), 10% ammonium persulfate (APS), and tetramethylethylenediamine (TEMED). Gels were run using the Mini-PROTEAN Tetra Cell apparatus from BioRad. Pre-run was done in 1X TBE buffer for 1 hour prior to loading of the samples. The samples were run at 150 V until the bromophenol blue dye reached the bottom of the gel. The radioactivity in the gel was analyzed by Phosphor Imager (Molecular Dynamics – Typhoon Trio Imager).

4.3.6 Human cell lines, plasmids, and procedures

Human embryonic kidney (HEK-293) cells were grown in Dulbecco's modified Eagle's medium, DMEM (Mediatech, Inc. Manassas, VA), supplemented with 10% heat-inactivated fetal bovine serum (Gemini, Bio-Products, West Sacramento, CA) and 1X Penicillin/Streptomycin (Lonza, Walkersville, MD). Cells were grown at 37˚C in a water-jacketed 5% CO_2 humidified incubator (NuAire, Plymouth, MN). Cell line 658-D (kindly provided by Matthew Porteus, Standford University) is a HEK-293 derivative where plasmid pA658 was randomly integrated which contains a non-functional GFP gene under the CMV/CBA promoter(Porteus and Baltimore 2003). Plasmid pLDSL contains the DsRed2 gene disrupted by a 45-bp region containing the 18-bp site for the I-*Sce*I

endonuclease preceded by 2 STOP codons. Plasmid p67 contains the I-*Sce*I endonuclease gene expressed under the CMV/CBA promoter (Porteus and Baltimore 2003). Cells were transfected using polyethylenimine (PEI, Polysciences, Warrington, PA) transfection reagent in 24-well plates seeded at a density of ~150,000 cells per well (Grieger, Choi et al. 2006, Hirsch, Storici et al. 2009) 24 hours prior to transfection. In all transfection experiments in HEK-293 cells, the plasmid DNA was used in the amount of 0.5 μg for the expression vector as well as 0.5 μg for the targeted vector and the repairing DNA oligonucleotide used was 1 μg, unless otherwise indicated. For experiments in 658-D cells, 0.5 μg for the expression vector p67 and 1.5 μg of the repairing DNA oligonucleotide were used. In these experiments, the oligos and the plasmid were diluted in DMEM without supplements, vortexed in the presence of PEI, and then added to the wells 10-15 minutes later. Red fluorescent cells were visualized by fluorescent microscopy using a Zeiss Observer A1 microscope and an AxioCam MRm camera (Zeiss, Thornwood, NY). Frequencies of RFP and GFP positive cells were obtained by flow cytometric analysis using the BD FACS Aria II Cell Sorter (BD Biosciences, Sparks, MD) for RFP detection or the BD LSR II Flow Cytometer (BD BioSciences, Sparks, MD) for GFP detection only 5-8 days following transfection. Sequences of oligonucleotides used to repair the GFP or DsRed2 genes are listed in **Table 4.2**.

Oligonucleotide	Size (mer)	Sequence (aptamer in red)
ISB7_GFP_F54	90	5'_GCGGGCGCTGTTGACAGCGGTCAGGTGGATGGGATGTTCAAGGACGACGGCAACTACAAGACGCGCGCCGAGGTGAAGTTCGAGGGCGAC_3'
ISB7_GFP_F40	76	5'_GCGGGCGCTGTTGACAGCGGTCAGGTGGATGGGATGACGACGGCAACTACAAGACGCGCGCCGAGGTGAAGTTCGA_3'
ISB7_DsRed2_F54	90	5'_GCGGGCGCTGTTGACAGCGGTCAGGTGGATGGGATG GCGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCTGCTTCATCTACAAGGTG_3'
ISB7_DsRed2_F40	76	5'_GCGGGCGCTGTTGACAGCGGTCAGGTGGATGGGATG TGACCCAGGACTCCTCCCTGCAGGACGGCTGCTTCATCTA_3'
ISB7_DsRed2_F30	66	5'_GCGGGCGCTGTTGACAGCGGTCAGGTGGATGGGATG CAGGACTCCTCCCTGCAGGACGGCTGCTTC_3'
P1P2_GFP_F54	90	5'_TTCTGCCCGCCTCCTTCCGACGAGATAGGCGGACACTTCAAGGACGACGGCAACTACAAGACGCGCGCCGAGGTGAAGTTCGAGGGCGAC_3'

P1P2_GFP_F40	76	5'_TTCTGCCCGCCTCCTTCCGACGAGATAGGCGGACACACGACGGCAACTACAAGACGCGCGCCGAGGTGAAGTTCGA_3'
P1P2_DsRed2_F54	90	5'_TTCTGCCCGCCTCCTTCCGACGAGATAGGCGGACACGCGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCTGCTTCATCTACAAGGTG_3'
P1P2_DsRed2_F40	76	5'_TTCTGCCCGCCTCCTTCCGACGAGATAGGCGGACACTGACCCAGGACTCCTCCCTGCAGGACGGCTGCTTCATCTA_3'
P1P2_DsRed2_F30	66	5'_TTCTGCCCGCCTCCTTCCGACGAGATAGGCGGACACCAGGACTCCTCCCTGCAGGACGGCTGCTTC_3'

Table 4.2 Oligonucleotides used for mammalian cells The table above shows the oligonucleotides used in HEK-293 and 658-D cell lines. The aptamer sequence is shown in red.

4.3.7 Yeast Strains

Two different strain backgrounds were used for these studies, BY4742 (MATα his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0) and 55R5-3C (MATa ura1). For the BY4742 background, *TRP5*, *ADE2*, and *LEU2* loci were tested. For the 55R5-3C background, *TRP1*, *ADE2*, and *LEU2* loci were tested.

For the *TRP5* locus (BY4742 background), yeast strains FRO-155 and FRO-526 were used. Yeast haploid strain FRO-155 (*MATα his3Δ1 leu2Δ0 lys2Δ0 trp5*::GSHU *lys2*::Alu IR) contains the GSHU CORE cassette (including the I-*Sce*I gene under the inducible *GAL1* promoter, the hygromycin resistance gene *hyg*, and the counterselectable URA3 gene from *Kluyveromyces lactis* (KlURA3) marker gene) and the I-*Sce*I site (HOT site) in *TRP5*(Storici, Durham et al. 2003). FRO-526 is identical to FRO-155 except that instead of the CORE cassette GSHU, the *TRP5* gene is disrupted with the CORE cassette UK (the counterselectable URA3 gene from Kluyveromyces **lactis** (KlURA3) marker gene along with the *KanMX4* gene conferring G418 resistance). These strains, along with separate isolates FRO-156 (identical to FRO-155), and FRO-527 (identical to FRO-526) were used to characterize the *TRP5* locus.

All other strains were generated by integrating the CORE cassette GSH (including the I-*Sce*I gene under the inducible *GAL1* promoter and the hygromycin resistance gene *hyg*) into each respective locus and strain background. For the strains to contain the I-*Sce*I recognition site, the CORE cassette was PCR-amplified from plasmid pGSHU(Storici,

Durham et al. 2003) using primers with homology tails to the respective integration site along with the I-*Sce*I site upstream of the GAL1 promoter. The strains lacking the I-*Sce*I site were generated in the same way except with primers lacking the I-*Sce*I site.

4.3.8 Yeast Transformations

Transformations were done as previously described with minor variations(Stuckey, Mukherjee et al. 2011). 50 ml of YPLac liquid culture was inoculated approximately 24 hours prior to transformation and incubated with vigorous shaking at 30ºC. Transformations were done with 1 nmol of total oligo DNA. Sequences of oligonucleotides used for repair can be found in **Table 4.3**. Cells from each transformation were diluted appropriately and plated to synthetic complete medium lacking the respective amino acid and containing 2% galactose for I-*Sce*I induction. Viability was assessed by diluting and plating control cells that did not contain oligonucleotides for repair on synthetic complete medium.

4.3.9 Data presentation and statistics

Graphs were made using GraphPad Prism 5 (Graphpad Software, Inc.). Data are plotted as mean values with 95% confidence intervals shown. Statistical significance was determined using two-tailed *t*-tests (Mann-Whitney U).

Oligonucleotide	Size (mer)	Sequence (aptamer in red)
ISB7_TRP5_F54	90	5' GCGGGCGCTGTTGACAGCGGTCAGGTGGATGGGATG GGAAAAGGGTTTTGATGAAGCTGTCGCGGATCCCACATTCTG GGAAGACTTCAA 3'
ISB7_TRP5_F40	76	5' GCGGGCGCTGTTGACAGCGGTCAGGTGGATGGGATG GGTTTTGATGAAGCTGTCGCGGATCCCACATTCTGGGAAG 3'
ISB7_TRP1_F54	90	5’ GCGGGCGCTGTTGACAGCGGTCAGGTGGATGGGATG GTGGCAAGAATACCAAGAGTTCCTCGGTTTGCCAGTTATTAA AAGACTCGTATT 3’
ISB7_ADE2_F54	90	5’ GCGGGCGCTGTTGACAGCGGTCAGGTGGATGGGATG GGACATTATACCATTGATGCTTGCGTCACTTCTCAATTTGAAG CTCATTTGAGA 3’
ISB7_ADE2_F40	76	5’ GCGGGCGCTGTTGACAGCGGTCAGGTGGATGGGATG ATACCATTGATGCTTGCGTCACTTCTCAATTTGAAGCTCA 3’
ISB7_LEU2_F54	90	5’ GCGGGCGCTGTTGACAGCGGTCAGGTGGATGGGATG CGCTTTCATGGCCCTACAACATGAGCCACCATTGCCTATTTGG TCCTTGGATAA 3’
ISB7_LEU2_F40	76	5’ GCGGGCGCTGTTGACAGCGGTCAGGTGGATGGGATG ATGGCCCTACAACATGAGCCACCATTGCCTATTTGGTCCT 3’
P1P2_TRP5_F54	90	5' TTCTGCCCGCCTCCTTCCGACGAGATAGGCGGACAC GGAAAAGGGTTTTGATGAAGCTGTCGCGGATCCCACATTCTGG GAAGACTTCAA 3'

P1P2_TRP5_F40	76	5' TTCTGCCCGCCTCCTTCCGACGAGATAGGCGGACAC GGTTTTGATGAAGCTGTCGCGGATCCCACATTCTGGGAAG 3'
P1P2_TRP1_F54	90	5′ TTCTGCCCGCCTCCTTCCGACGAGATAGGCGGACAC GTGGCAAGAATACCAAGAGTTCCTCGGTTTGCCAGTTATTAAAAGACTCGTATT 3′
P1P2_ADE2_F54	90	5’ TTCTGCCCGCCTCCTTCCGACGAGATAGGCGGACAC CGCTTTCATGGCCCTACAACATGAGCCACCATTGCCTATTTGGTCCTTGGATAA 3’
P1P2_ADE2_F40	76	5’ TTCTGCCCGCCTCCTTCCGACGAGATAGGCGGACAC ATGGCCCTACAACATGAGCCACCATTGCCTATTTGGTCCT 3’
P1P2_LEU2_F54	90	5′ TTCTGCCCGCCTCCTTCCGACGAGATAGGCGGACAC CGCTTTCATGGCCCTACAACATGAGCCACCATTGCCTATTTGGTCCTTGGATAA 3′
P1P2_LEU2_F40	76	5′ TTCTGCCCGCCTCCTTCCGACGAGATAGGCGGACAC ATGGCCCTACAACATGAGCCACCATTGCCTATTTGGTCCT 3′

Table 4.3 Oligonucleotides used for yeast The table above shows the oligonucleotides used in yeast. The aptamer sequence is shown in red.

4.4 RESULTS

4.4.1 Bulk affinity analysis and selection

Serial dilutions of the aptamer library were made and the individual free DNA run time was determined (**Figure 4.1**). Next, we proceeded to combine I-*Sce*I and DNA for a bulk affinity analysis (**Figure 4.2**). The bulk affinity analysis allowed us to visualize peaks corresponding to I-*Sce*I DNA complexes. Knowing the migration times of these complexes along with the free DNA alone, a collection window could be determined stretching from the beginning of the first complex peak to the end of the last complex peak. In order to increase the stringency of the selection, the concentration of I-*Sce*I was greatly reduced compared to the bulk affinity for the first round of selection. The amount of DNA used in the first round of selection was increased compared to the bulk affinity assay in order to increase the number of binding sequences. Despite the reduction in I-*Sce*I concentration, complexes were still observed for the first round of selection (**Figure 4.3**). The fraction containing the aptamer-I-*Sce*I complexes was collected and without amplification was used in a subsequent round of selection. In the second round of selection, the ratio of DNA forming a complex compared to free DNA was much higher than in the first round. Selection proceeded to a third round, but no complexes were observed probably due to the very low amount of total DNA.

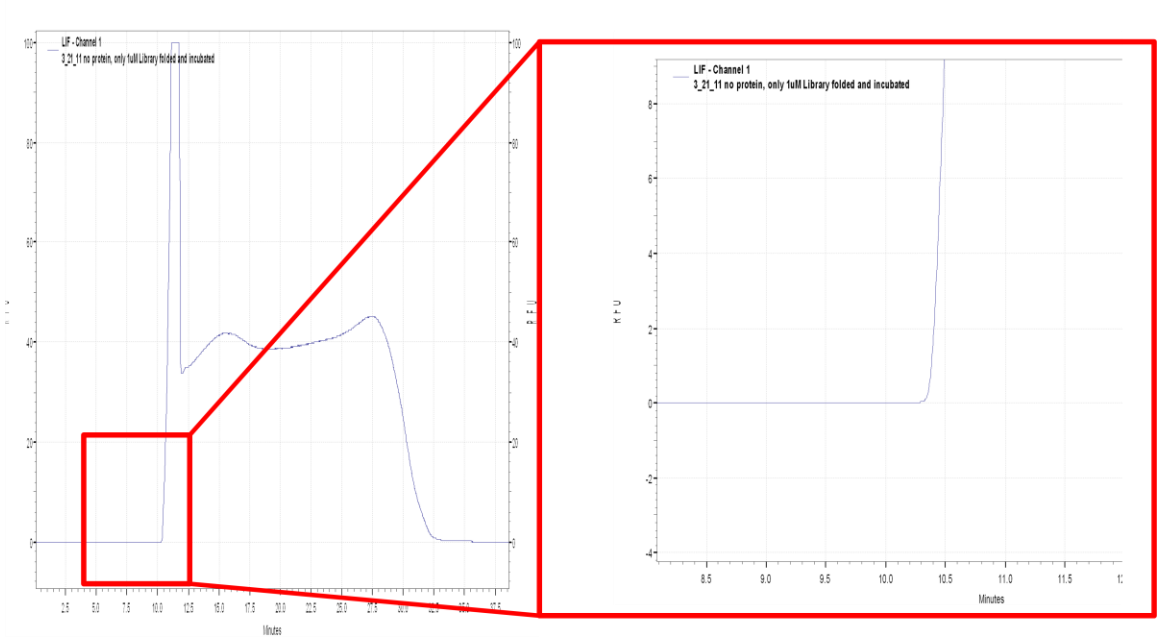


Figure 4.2 DNA library run without protein. The random ssDNA library (1 µM) was run alone, in the absence of I-*Sce*I. As can be seen the free DNA begins to appear at approximately 10 minutes. There are no complexes prior to 10 minutes.

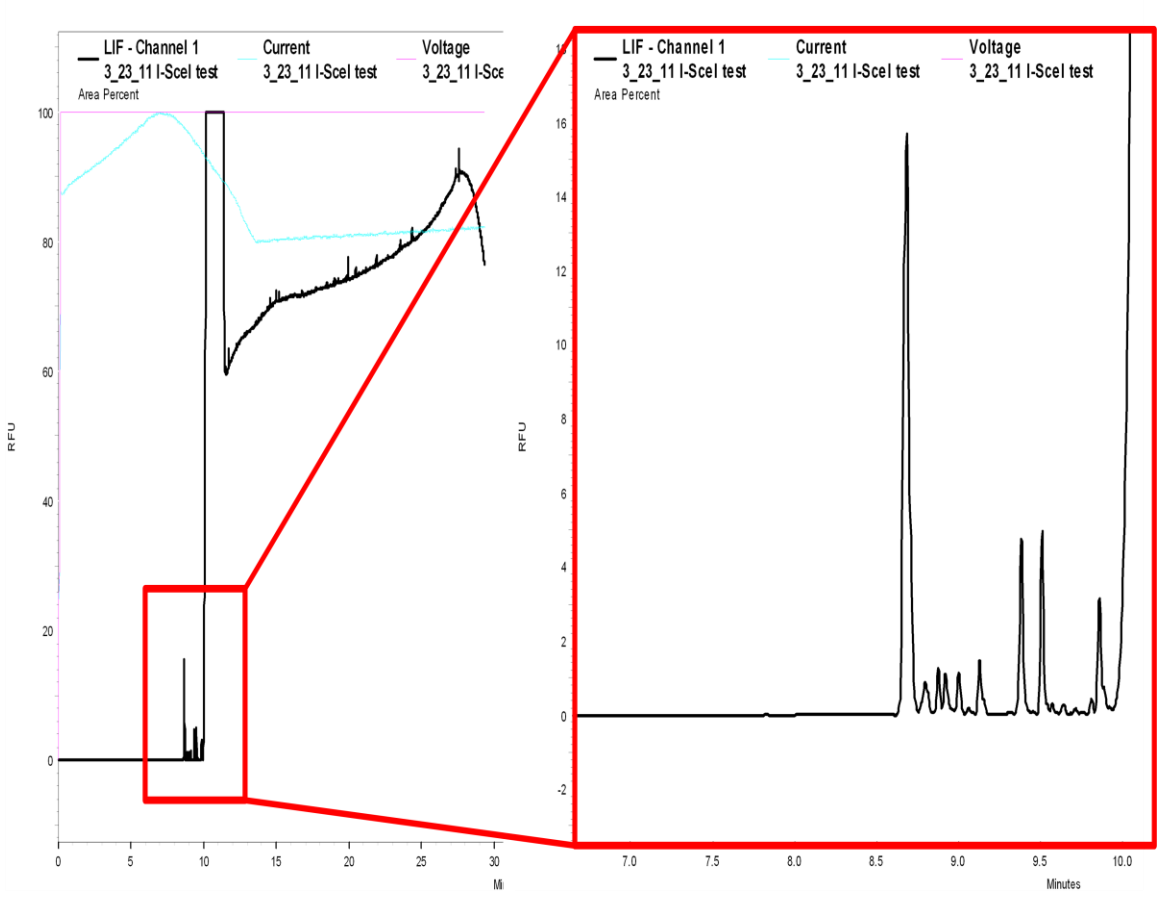
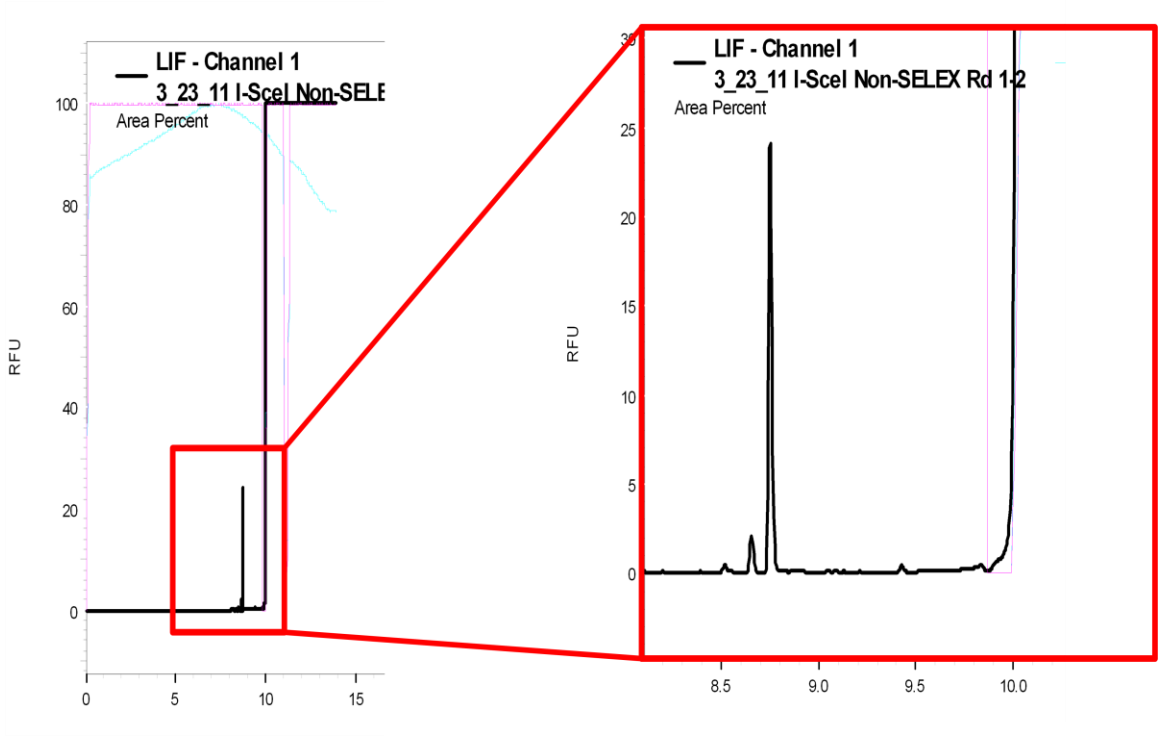
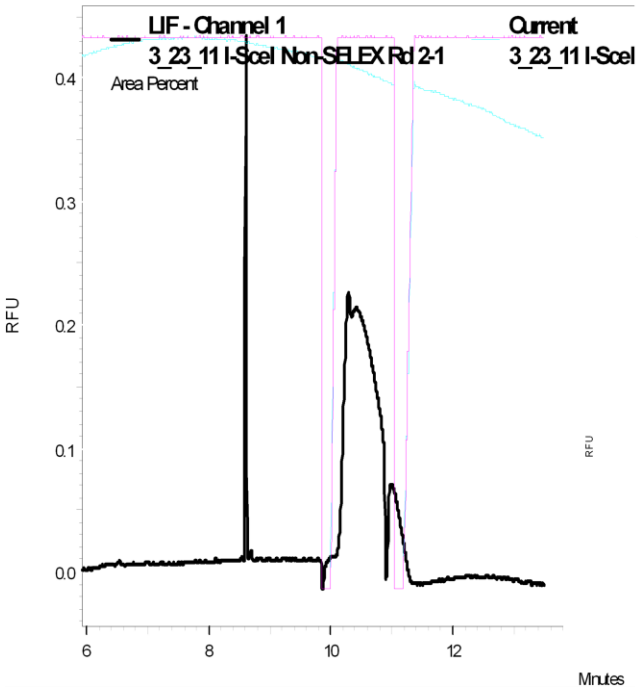


Figure 4.3 Bulk affinity assay for I-*Sce*I aptamer selection. A bulk affinity assay was done with 1.5 µM I-*Sce*I and 100 nM ssDNA aptamer library. The free DNA not bound to I-*Sce*I has a run time of approximately 10 minutes. Prior to the free DNA peak there are several I-*Sce*I DNA complexes that form.

Round 1



Round 2



Round 3

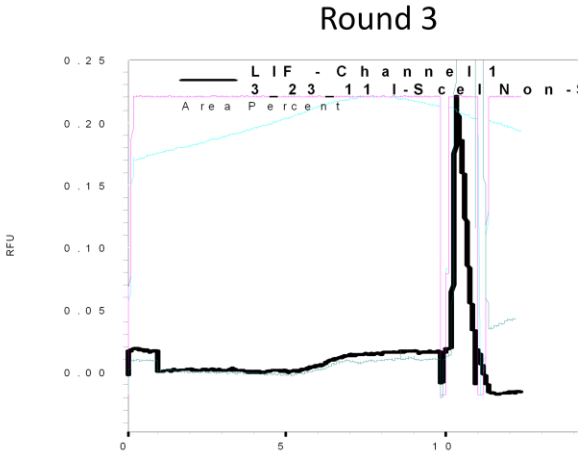


Figure 4.4 Non-SELEX Round 1, 2, and 3. The first round of Non-SELEX ran with a low concentration of I-*Sce*I (100 nM) and a large concentration (50 µM) of ssDNA library. Even in the first round of selection, I-*Sce*I DNA complex peaks could be detected prior to the free DNA. The second round of selection was done using the complexes collected in the first round. As can be seen for round 2, the amount of total DNA drastically decreased but there was still a complex peak that formed. In the third round of selection the amount of total DNA was so low that no complex was detected.

4.4.2 Post-selection processing of aptamers

The collected DNA from the second round of selection containing the DNA I-*Sce*I complexes underwent qRT-PCR amplification using primers flanking the internal random sequence. The forward primer, P1, amplified the aptamer-containing strand of interest, while the reverse primer, P2, amplified the non-aptamer complementary strand. In a previous protocol to select DNA aptamers, it was shown that over-amplification of the random library leads to formation of non-specific products [12], therefore the random library was only amplified to ~50% of the maximum yield as measured by qRT-PCR. After qRT-PCR analysis revealed the optimum number of cycles to amplify the collected DNA, a regular PCR using more of the collected DNA was performed with unlabeled P1 and P2 primers. TOPO blunt-end cloning was performed with the aptamer pool and colonies were selected by growth on LB-kanamycin plates. Forty-eight colonies were chosen and plasmid DNA was crudely extracted. An asymmetric PCR was used on the extracted DNA to predominantly amplify the potential aptamer strand. The P1 primer that amplified the potential aptamer strand was FAM-labeled so the potential aptamer could be analyzed using CE with LIF. Potential aptamers were run with I-*Sce*I and the strongest binding aptamers were selected for sequencing.

After sequencing, several candidate aptamers to I-*Sce*I, called "I-*Sce*I strong binders" were ordered and P32-labeled for further characterization using electrophoretic mobility shift assay (EMSA) gels. The binding conditions were modified from the CE

experiments. Of the candidate aptamers, three sequences were chosen for their ability to bind to I-*Sce*I, namely I-*Sce*I strong binder 4 (ISB4), I-*Sce*I strong binder 7 (ISB7), I-*Sce*I strong binder 10 (ISB10). Of these sequences, ISB7 showed the strongest binding through EMSA gels. These three sequences were then ordered as PAGE purified FAM-labeled oligonucleotides and again underwent testing by CE with LIF. The binding affinities by CE were calculated to be ~3.16 µM for ISB7, ~52.49 µM for ISB4, and ~5.83 µM for ISB10 by a method described previously(Berezovski and Krylov 2002).

4.4.3 Characterization of I-SceI aptamers in yeast

4.4.3.1*TRP5* locus

Oligonucleotides were ordered from Eurofins MWG Operon (Huntsville, AL) that contained the aptamer sequences from ISB4, ISB7, and ISB10 attached to the 5' end of a DNA sequence containing 54 bases of homology to restore the disrupted *TRP5* gene in yeast strains FRO-155 and FRO-526 described previously. The primer regions from the aptamer library were removed as they were not seen to influence binding (data not shown). In addition to the aptamer-containing oligonucleotides used to correct the *TRP5* gene a negative control not selected to bind to I-*Sce*I was used. Due to the inability of the library primers P1 and P2 to bind I-*Sce*I, these were used in place of the aptamer sequence in a new oligonucleotide, called P1P2, as the non-binding negative control.

Using each of the different oligonucleotides to repair the *TRP5* gene it was shown that the ISB7 aptamer-containing oligonucleotide significantly increased the level of gene correction compared to the negative control P1P2 and the other aptamer-containing oligonucleotides ISB4 and ISB10 in the FRO-155 strain where the I-*Sce*I gene was expressed and the I-*Sce*I site was present at the target site. FRO-526, the strain that did not have the I-*Sce*I site and also did not express the I-*Sce*I gene, showed no significant difference between the ISB7 aptamer-containing oligonucleotide and the negative control P1P2 or the other aptamer-containing oligonucleotides (ISB4 and ISB10). Also, because there was no DNA double-stranded break (DSB) in FRO-526 the level of gene correction was reduced several orders of magnitude. As an additional control, the strain containing the I-*Sce*I gene and site was grown and plated to glucose media. Without galactose for the expression of I-*Sce*I, there was no significant difference between the ISB7 aptamer-containing oligonucleotide and the other oligonucleotides.

4.4.3.2 Aptamer testing at various loci in yeast

After the results from the *TRP5* locus, testing was done at several other loci to verify this was not a locus specific event. In these new loci the ISB4 and ISB10 aptamers were not used since they did not show an increase in gene targeting at the *TRP5* locus. At each of these loci the GSH cassette containing the I-*Sce*I gene under the inducible *GAL1-10* promoter along with the hygromycin resistance gene *hyg* were integrated generating auxotrophs for the respective amino acid. For the BY4742 strain background, the *ADE2* and *LEU2* loci were chosen, and for the 55R5-3C strain background, the *TRP1*, *ADE2*, and *LEU2* loci were chosen. For each locus two strains were made in which one had the

integrated GSH cassette with the 18-bp I-*Sce*I recognition site and one strain that had the cassette but did not have the I-*Sce*I site. At every loci tested there was a significant increase in gene targeting with the ISB7 aptamer-containing oligonucleotide compared to the negative control P1P2 oligonucleotide when I-*Sce*I was induced by galactose and the I-*Sce*I site was present (**Figure 4.4**). There was an approximately 3-fold increase for the *ADE2* locus, approximately 2-fold increase for the *LEU2* locus, and approximately 2.5-fold for the TRP1 locus. However, in the strains lacking the I-*Sce*I site there was no significant difference between the oligonucleotides (**Figure 4.5**). Likewise, when the strains containing the I-*Sce*I site were grown and plated to glucose-containing media there was no significant difference between the oligonucleotides (**Figure 4.5**).

Seeing the increase in gene targeting using the ISB7 aptamer-containing oligonucleotide, it was postulated that by shortening the region of homology we might see an even greater effect from the aptamer. While shorter homology was expected to reduce the overall repair, it was expected that the fold-difference between the aptamer-containing oligonucleotide would increase relative to the non-binding oligonucleotide control. Due to the lower recombination level in the 55R5-3C strain background, the shorter oligonucleotides were only used in the BY4742 derived strains. By shortening the homology region of the oligonucleotide from fifty bases to forty bases, the level of repair decreased dramatically as expected (**Figure 4.6**). As we expected, the level of repair at the *TRP5* locus using the shorter oligonucleotides showed a greater fold difference between the ISB7 aptamer-containing oligonucleotide and the P1P2 non-selected negative control (from six-fold to fifteen-fold). Unexpectedly, this was not the case for the *ADE2* or the *LEU2* loci where the shorter oligonucleotides were not significantly

different (**Figure 4.6**).

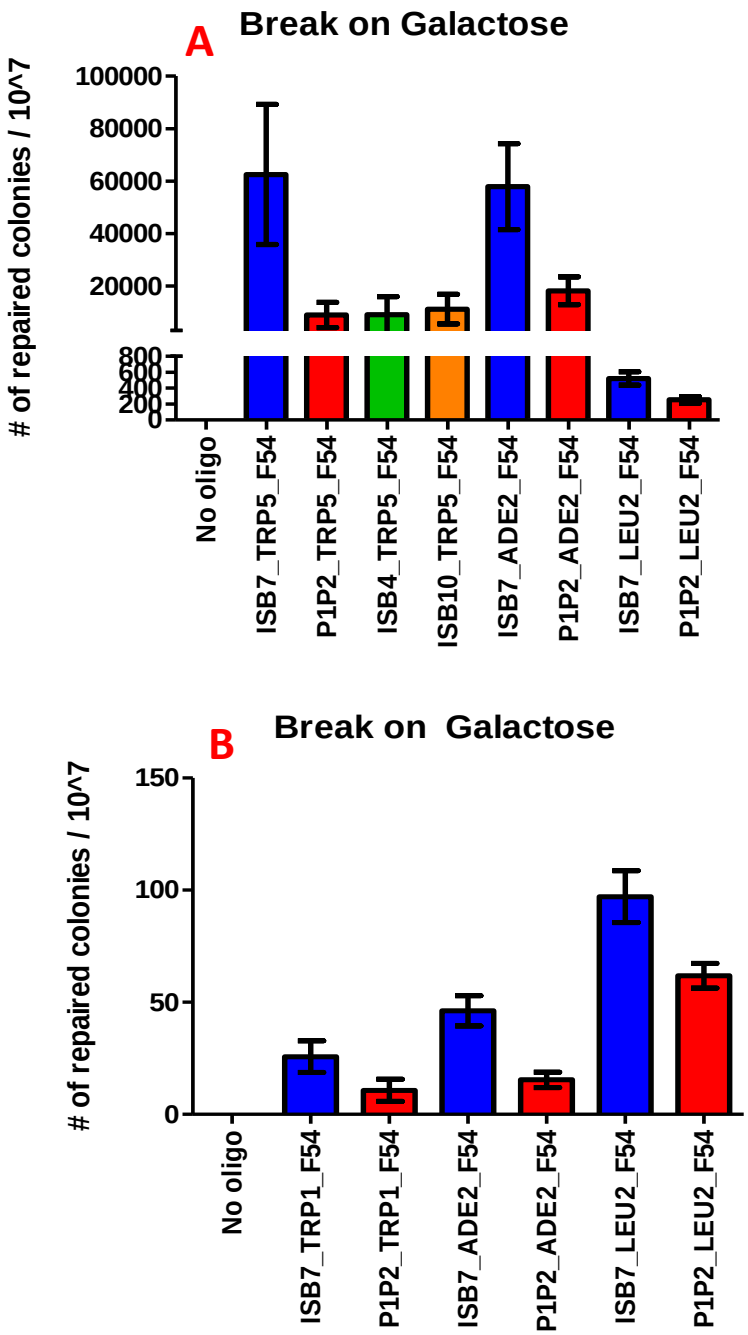


Figure 4.5 Yeast transformation results. The strains containing the I-*Sce*I break site as well as the I-*Sce*I gene under the *GAL1-10* promoter grown on galactose media. As can be seen, for every locus tested there was a significant increase in gene targeting using the aptamer-containing oligonucleotides compared to using the non-selected control. A) The strains from the BY4742 background. B) The strains from the 55R5-3C background.

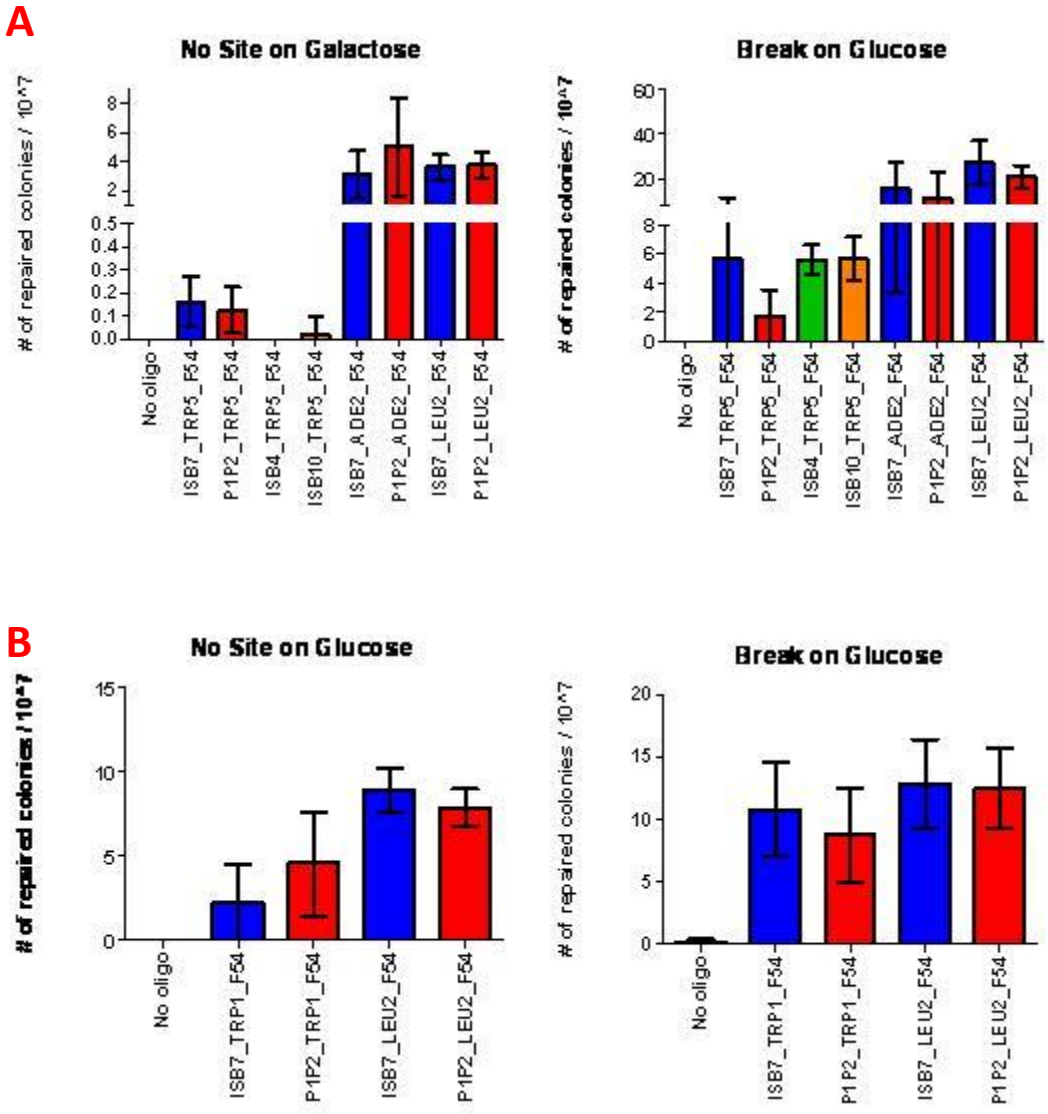


Figure 4.6 Yeast transformation results. The strains with the I-*Sce*I gene under the *GAL1-10* promoter but lacking the I-*Sce*I site (No Site). In addition, the strains with the I-*Sce*I break site were grown on glucose media, but since the I-*Sce*I gene is under the galactose-inducible promoter it should not be expressed under this condition. As can be seen, for every locus tested there was no significant difference in gene targeting using the aptamer-containing oligonucleotides compared to using the non-selected control. A) The strains from the BY4742 background. B) The strains from the 55R5-3C background (no ADE+ colonies were observed for the strain with or without the break site on glucose)

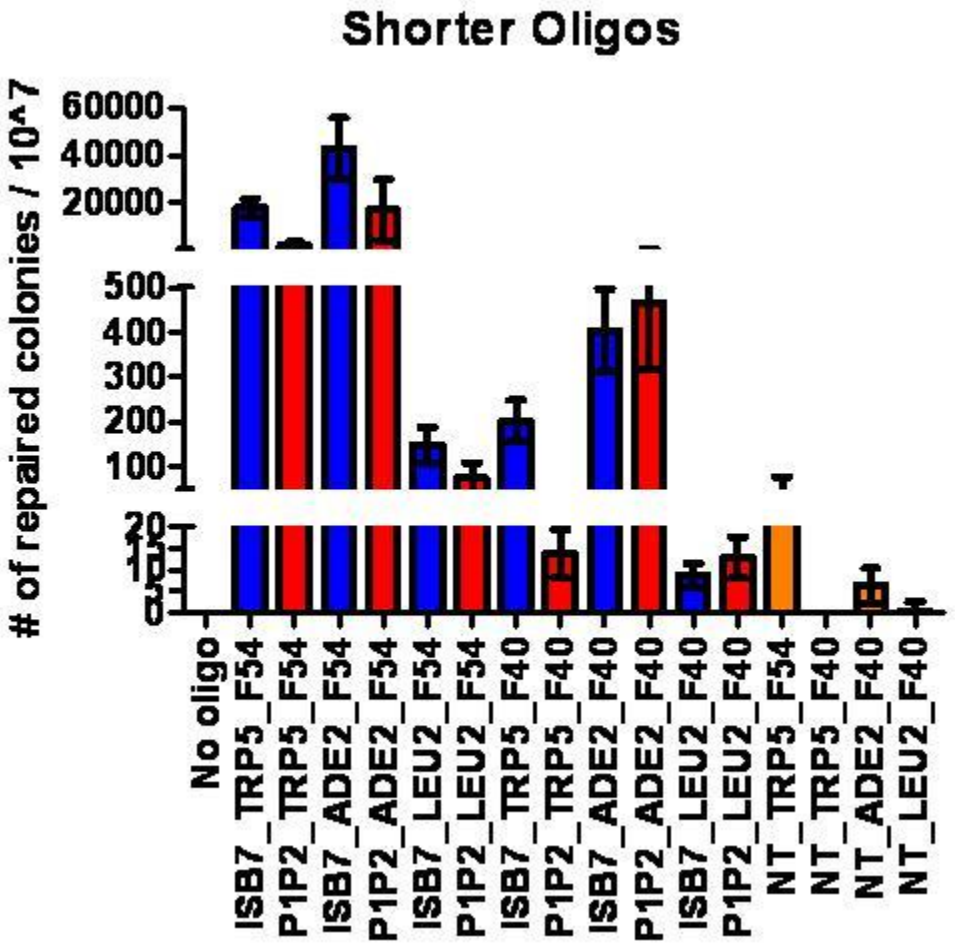


Figure 4.7 Yeast transformation with shorter oligonucleotides. The strains with the I-*Sce*I gene under the *GAL1-10* promoter and the I-*Sce*I site (all BY4742 strain background) were grown on galactose, inducing expression of I-*Sce*I. In addition to using aptamer or non-selected control containing oligonucleotides with 54 bases of homology, shown before to stimulate gene targeting, shorter oligonucleotides with only 40 bases of homology were used. As expected, the level of overall repair decreased with the shorter oligonucleotides because they had less homology. The shorter *TRP5* oligonucleotides showed a greater fold difference between the aptamer containing oligonucleotide and the non-selected control oligonucleotide. Unexpectedly, the shorter *ADE2* and *LEU2* repairing oligonucleotides both showed no significant difference between the aptamer and the non-selected control containing oligonucleotides.

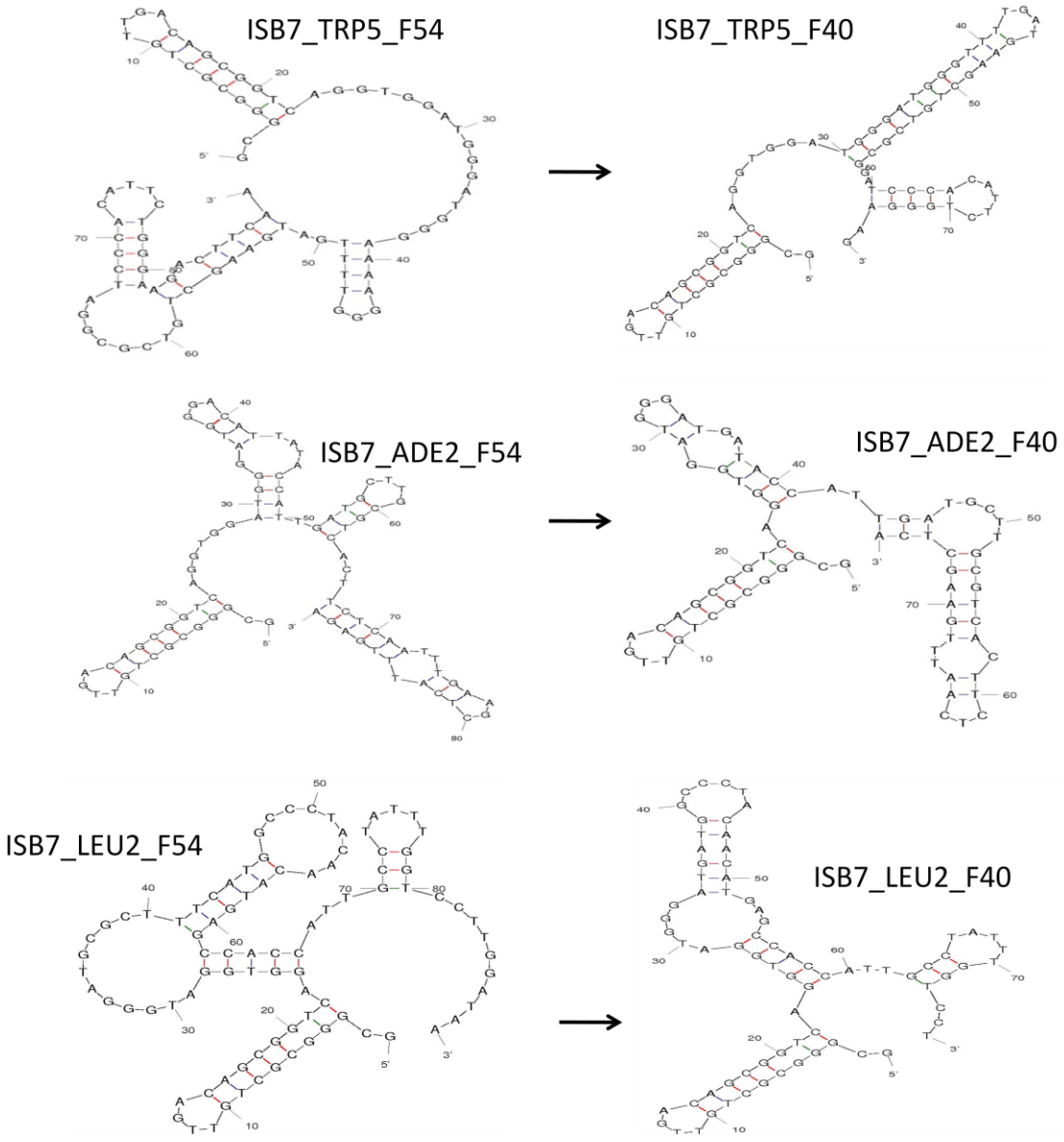


Figure 4.8 Yeast transformation with shorter oligonucleotides structural explanation. The shorter *TRP5* oligonucleotides showed a greater fold difference between the aptamer containing oligonucleotide and the non-selected control oligonucleotide as expected. Unexpectedly, the shorter *ADE2* and *LEU2* repairing oligonucleotides both showed no significant difference between the aptamer and the non-selected control containing oligonucleotides. Looking at the predicted secondary structures of the shorter aptamer containing oligonucleotides in the 5' aptamer region the shorter *TRP5* repairing oligonucleotide maintains the aptamer hairpin structure followed

by several bases not predicted to form any secondary structure (similar to the longer ISB7_TRP_F54 oligonucleotide). However, for the shorter *ADE2* oligonucleotide the proximity of the aptamer hairpin to a subsequent hairpin increases compared to the longer oligonucleotide which contained several single-stranded bases between the aptamer hairpin and the next hairpin. Lastly, while the *LEU2* oligonucleotide overall structure does not seem to change much, the size of the internal loops of the hairpin following the aptamer hairpin decreases significantly compared to the longer oligonucleotide. The proximity of the subsequent hairpin following the aptamer hairpin for the *LEU2* oligonucleotide may explain why the aptamer at this locus, while significantly different from the non-selected control, did not show a very high fold difference.

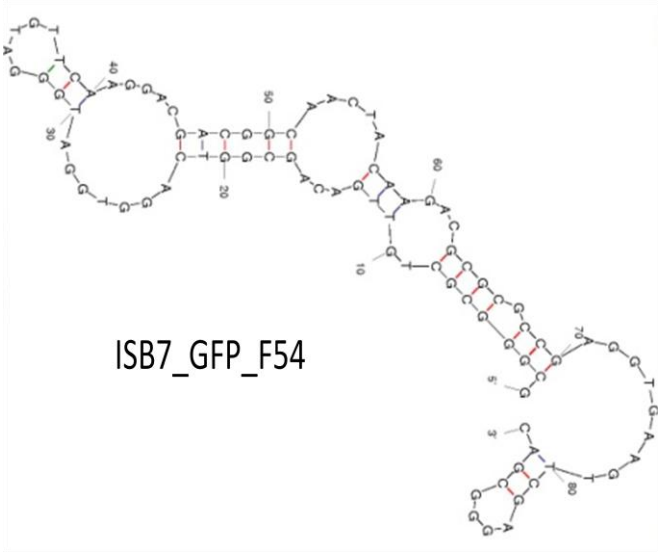
A possible explanation for the unexpected results at the *ADE2* and *LEU2* loci with shorter oligonucleotides could be that by shortening the oligonucleotides the secondary structure of the oligonucleotide was changed and the formation of the aptamer was no longer energetically favorable. To investigate this possibility, secondary structure prediction software mFold was used on the oligonucleotides (http://mfold.rna.albany.edu/?q=mfold/dna-folding-form). The lowest free-energy (most stable) structures predicted for the ISB7 aptamer-containing oligonucleotides with 54-base homology regions all formed the suspected aptamer hairpin near the 5' end of the oligonucleotide (**Figure 4.8**). It was interesting to note, however, that while the *TRP5* and *ADE2* oligonucleotides had several bases without secondary structure following the aptamer hairpin the oligonucleotide to repair *LEU2* contained only a single base between the aptamer hairpin and another stem-loop structure. This might explain why the *LEU2* oligonucleotide with the 54-base homology region, while still capable of increasing gene targeting, showed the least fold-difference in repair and was not as significantly different from the nonbinding control as either the *ADE2* or *TRP5* oligonucleotides. When analyzing the secondary structures of the ISB7 aptamer-containing oligonucleotides with 40-base homology regions, there was a noticeable change in secondary structure for each of the oligonucleotides (**Figure 4.8**). However, for the shorter *TRP5* oligonucleotide, the aptamer hairpin was still followed by several bases not predicted to form any secondary structure and hence not likely to interfere with the aptamer binding. This is consistent with the results seen *in vivo*. The *ADE2* oligonucleotide with 54 bases of homology compared to the *ADE2* oligonucleotide with 40 bases changed secondary structure such that the number of bases after the aptamer hairpin not forming a secondary structure went

from six to one, bringing another stem-loop structure closer to the aptamer hairpin, possibly interfering with binding. The *LEU2* oligonucleotide with 40 bases of homology had a similar structure to the *LEU2* oligonucleotide with 54 bases of homology, except that the size of the internal loops for the stem-loop structure close to the aptamer hairpin were reduced. The reduction in the size of the loops may have altered the binding capabilities of the aptamer. The potentially altered binding of the *LEU2* and *ADE2* 40-base homology aptamer-containing oligonucleotides would explain the reduction in gene targeting with these oligonucleotides and highlight a disadvantage to our system. Although the aptamer stimulates gene targeting, the secondary structure of the aptamer must be taken into account when designing oligonucleotides for repair.

4.4.4 Characterization of I-SceI aptamers in human cells

After testing the ISB7 aptamer in yeast, we also tested the aptamer in human embryonic kidney (HEK-293) cells. We have a chromosomal assay for detecting repair of GFP in a monoclonal cell line derived from HEK-293 known as 658-D, described previously (Porteus and Baltimore 2003). The 54 bp of homology aptamer containing oligonucleotide generated to repair this locus was not predicted to have secondary structure in which the aptamer hairpin would form (**Figure 4.9**). Without the aptamer formation, it was not expected that there would be an increase in gene targeting. In the 658-D cell line, as expected we did not see an increase in gene correction when transfecting an ISB7 aptamer-containing oligonucleotide and a plasmid for I-*Sce*I expression (**Figure 4.10**). Additionally, the assay was performed with a plasmid

containing the same disrupted GFP sequence used to generate the 658-D cell line in HEK-293 cells along with the expression vector and the ISB7 containing oligonucleotide or the non-binding control. Again, there was no increase in gene correction by the ISB7 containing oligonucleotide over the non-binding control and again there was a significant decrease in repair. Interestingly there was a significant increase in repair with the P1P2 containing oligonucleotide over the ISB7 containing oligonucleotide. An explanation for this increase could be that the secondary structure predicted for the ISB7 containing oligonucelotide is largely dsDNA (**Figure 4.9**), whereas the P1P2 containing oligonucleotide is mostly ssDNA, which was shown to be preferential for DSB repair in the HEK-293 and 658-D cell lines.



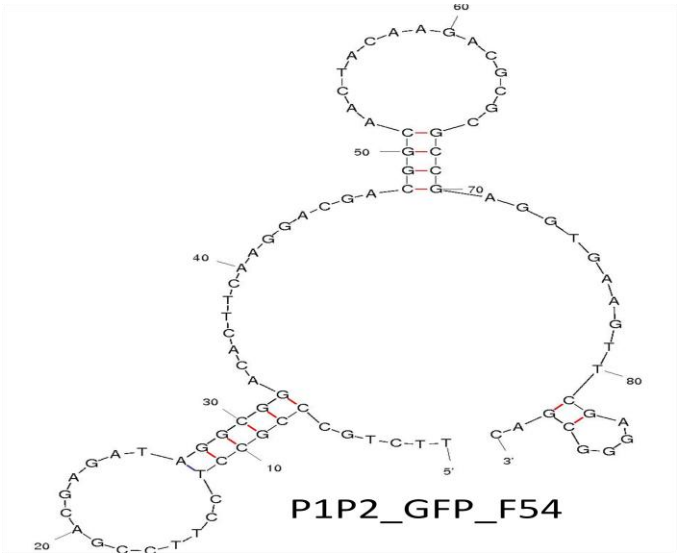


Figure 4.9 mFold structures for the GFP locus In both the HEK-293 and the 658-D cell line there is a significant decrease in repair by the ISB7 containing oligonucleotide compared to the nonselected control P1P2 containing oligonucleotide. Based upon the mFold predicted secondary structure it can be seen that the aptamer hairpin does not form for the ISB7 containing oligonucleotide. Also, the P1P2 containing oligonucleotide is mostly single-stranded DNA, which was shown previously to be a better substrate for repair in these cell lines.

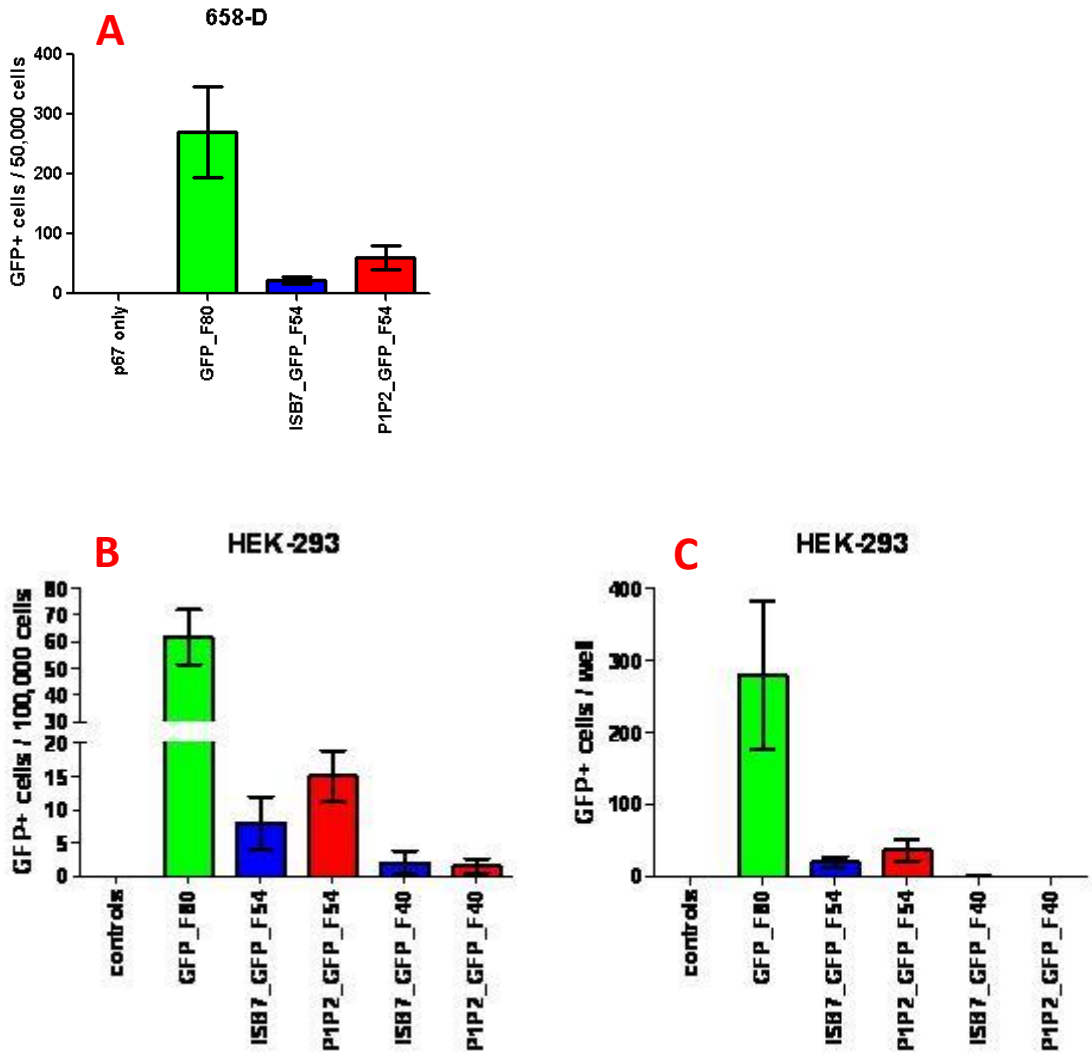


Figure 4.9 Testing of the ISB7 aptamer at the GFP locus in HEK-293 and 658-D cells. The GFP_F80 oligonucleotide does not contain the aptamer sequence or the non-selected control but is used as a positive control because it has 80 bases of homology (40 to either side of the DSB) to GFP. A) In the 658-D cell line there is a significant decrease in repair by the ISB7 containing oligonucleotide compared to the nonselected control P1P2 containing oligonucleotide which is also true for B) the plasmid based assay. C) Represents data from hand counting the number of GFP+ cells in each well and this data corroborates the flow cytometry data. The level of repair with the shorter oligonucleotides was not significantly different from the background.

In addition to testing at the GFP locus, the ISB7 aptamer was tested using a plasmid based assay to repair the red fluorescent protein, DsRed2. The assay consisted of transfecting a plasmid containing the DsRed2 gene disrupted with the I-*Sce*I recognition site and two STOP codons along with the I-*Sce*I expression vector and the oligonucleotides for repair of DsRed2. Unlike for the GFP locus, the secondary structure predicted for each of the ISB7 containing oligonucleotides used to repair DsRed2 contained the aptamer hairpin. Using oligonucleotides containing the ISB7 aptamer and 54 bases of homology to DsRed2, there was a significant increase in repair over the non-binding control. As in yeast, oligonucleotides with shorter homology regions were designed and tested, this time using oligonucleotides with 40 bases of homology and 30 bases of homology. Both the ISB7 aptamer containing oligonucleotides with 40 and 30 bases of homology were predicted to form the aptamer hairpin. Similar to our results at the *TRP5* locus in yeast, the shorter oligonucleotides increased repair relative to the corresponding shorter non-binding control sequence as measured by flow cytometry (**Figure 4.10**) beyond the difference seen with the longer oligonucleotide. In the case of the shortest oligonucleotides, those with only 30 bases of homology to DsRed2, the flow cytometer was unable to accurately detect the level of repair since it was too close to background levels. Hand counts of the RFP+ cells in each well were compared to the readings by flow cytometry. For the oligonucleotides with 54 and 40 bases of homology, the hand counts and the readings by flow cytometry were in agreement, but for the shorter oligonucleotides with only 30 bases of homology, the hand counts did not agree with the flow cytometry readings which seemed to underreport the level of repair. Using

the hand counts for the ISB7 aptamer-containing oligonucleotide with 30 bases of homology to DsRed2, a fifteen-fold increase in repair relative to the nonbinding control was observed.

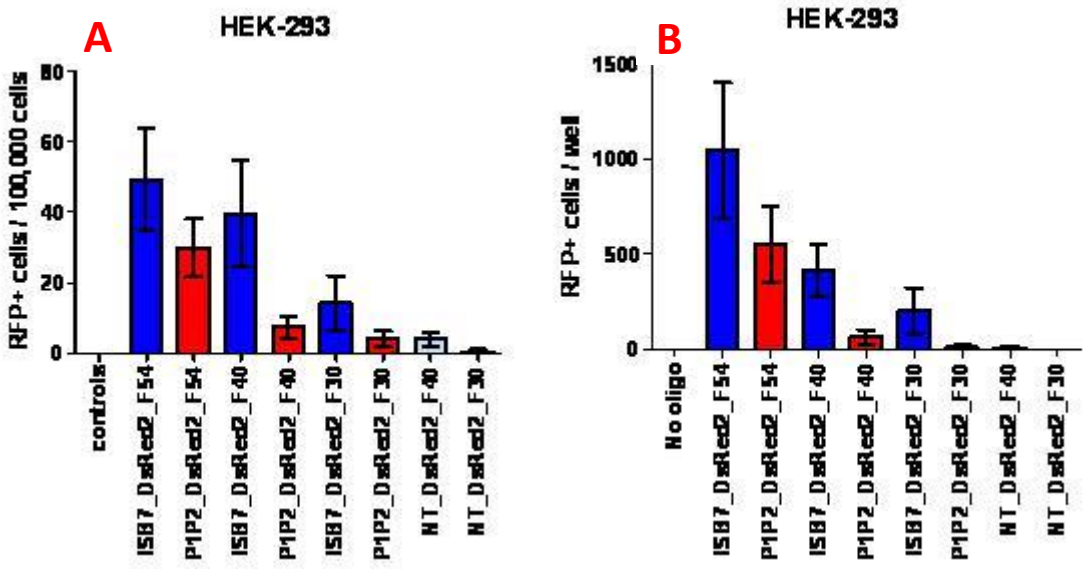


Figure 4.11 Testing of the ISB7 aptamer at the DsRed2 locus in HEK-293 cells. A) Flow cytometry analysis reveals a significant difference between the ISB7 aptamer containing oligonucleotides and the non-selected P1P2 control oligonucleotides. When the homology length is shortened from 54 bases to 40 there is a greater fold difference between the ISB7 and P1P2 containing oligonucleotides. For the oligonucleotides with 30 bases of homology, while the fold difference between the ISB7 aptamer containing oligonucleotide and the non-selected P1P2 control containing oligonucleotide is greater than the oligonucleotides with 54 bases of homology, it is less than the oligonucleotides with 40 bases of homology. The flow cytometer seems to have reached its lower detection limit and may be over reporting the number of RFP+ cells in the controls, the P1P2_DsRed2_F30 oligonucleotide, and the NT_DsRed2_F40 oligonucleotide; hence it is underreporting the fold difference. B) Represents data from hand counting the number of RFP+ cells in each well and this data corroborates the flow cytometry data except for those oligonucleotides previously mentioned as being over reported. Based upon the hand counts, the fold difference is actually greatest with the shortest oligonucleotides as expected.

4.5 CONCLUSIONS

The system described here takes advantage of the fact that a single DNA molecule can have multiple functions. By constructing a single DNA molecule to contain both an aptamer region as well as a region to repair a genomic locus we were able to tether homologous DNA to the site-specific endonuclease I-*Sce*I. In every genomic locus tested we stimulated gene targeting when the I-*Sce*I site was present along with expression of the I-*Sce*I endonuclease.

Although I-*Sce*I was chosen, protein-assisted targeting in principle could be applied to many other proteins. Other site-specific endonucleases, including modular endonucleases such as ZFNs, TALENs, or Cas proteins should theoretically also provide stimulation of gene targeting similarly to I-*Sce*I. I-*Sce*I generates a DSB in a site-specific manner, but alternatively to a site-specific protein, protein-assisted targeting could be used for other components of DSB repair that are not site-specific. For example, an aptamer to a general HR repair protein, such as Rad51, might provide an increase to gene targeting as well. Conversely, the mild but statistically significant stimulation of gene targeting even in the absence of Rad52 (data not shown) might mean that protein-assisted targeting could be used in an HR-independent fashion. The work described here represents a novel proof-of-principle study where we show that aptamers can be used as tools for gene targeting.

CHAPTER 5

GENERAL CONCLUSIONS

5.1 MAJOR FINDINGS

- **Selection of BSA aptamers using CE-SELEX without LIF**
- **Selection of I-*Sce*I aptamers using Non-SELEX with LIF**
- ***in vivo* testing of I-*Sce*I aptamer ISB7 reveals stimulation of gene targeting in both yeast and human cells**
- **Generation of a novel system to increase gene targeting**

5.2 CONCLUSIONS

Gene therapy has recently been gaining attention again after safety concerns in the early 2000s slowed, and in some cases halted, its progress due to the death of a clinical trial subject and the development of a leukemia-like disorder in several patients (Thomas, Ehrhardt et al. 2003). One of the major problems of gene augmentation therapy, the most commonly taken approach to gene therapy, is the possibility of insertional mutagenesis of the transgene. Random integration of foreign genetic material is the normal outcome for exogenously introduced DNA, but it presents a risk for any gene therapy patient.

In addition to the safety problem random integration poses, gene augmentation is also limited from a researcher's perspective. In order to study the function of a gene, addition of another copy of that gene will provide little or no information. Gene targeting is essential for research because it allows for the deletion and/or modification of endogenous genes, which cannot be achieved with random integration.

As described in the introduction, there are currently many different strategies for increasing gene targeting, but all of these strategies are harmful in some way. Increase of gene targeting by a DSB can introduce mutations in cells that do not undergo homologous recombination. Additionally, DSBs, generated by homing endonucleases (Petek, Russell et al. 2010), ZFNs(Pattanayak, Ramirez et al. 2011) (Gabriel, Lombardo et al. 2011), or TALENs (Hockemeyer, Wang et al. 2011) all have off-target cleavage which presents a safety concern for mutagenesis beyond the targeted site. Increase of gene targeting by over-expression of recombination proteins from HR proficient organisms could also lead to unwanted recombination elsewhere in the genome (Di Primio, Galli et al. 2005). In the same way, the increase of gene targeting by knocking down or knocking out expression of key NHEJ proteins could lead to increased recombination at unwanted sequences. Also, knocking down or knocking out key NHEJ proteins, while increasing gene targeting, limits the cell's ability to repair DSBs which in most cases leads to a hypersensitivity to DNA damage (Adachi, Suzuki et al. 2003). A third reason that knocking down or knocking out classical NHEJ proteins could be harmful is that this will then cause an increase in alternative end-joining, also known as

mutagenic end-joining or micro-homology mediated end-joining (MMEJ), which would increase mutations on a genome-wide scale (Bindra, Goglia et al. 2013).

In order to address the need for a safer, more effective means to increase gene targeting we began this project of protein-assisted targeting. There were two parallel strategies employed, the use of two proteins capable of binding to DNA, specifically the *GAL4*-DBD and the I-*Sce*I endonuclease, as well as the strategy using a DNA aptamer to I-*Sce*I. The *GAL4*-DBD I-*Sce*I fusion protein was unable to be expressed which limits our understanding of the efficacy of protein-assisted targeting using this strategy. The DNA aptamer to I-*Sce*I strategy relied on the successful selection of a DNA sequence capable of binding to I-*Sce*I. After several attempts with traditional SELEX (see **APPENDIX A**) we were successful in generating a high affinity aptamer to I-*Sce*I using a modification of capillary electrophoresis (CE)-SELEX. Use of this aptamer as part of a longer DNA molecule that contained homology to a target sequence stimulated gene targeting in every locus tested when the aptamer secondary structure was able to form. In yeast cells, the aptamer stimulated repair up to 15-fold over the non-binding control and in mammalian cells, the aptamer stimulated gene targeting up to 16-fold over the control. The system we have developed represents an effective way to increase gene targeting.

Protein-assisted targeting provides a strategy that increases gene targeting, in a way that is not expected to increase mutations and produce damage for the cell. The binding of our

targeting DNA to a protein does not cause any damage to the endogenous DNA targeted and should not increase recombination outside of the targeted area. For the purposes of gene targeting, it benefits us to reduce NHEJ (Iiizumi, Kurosawa et al. 2008) (Bertolini, Bertolini et al. 2009), but it is clear that doing so on a genome-wide scale is harmful to the cell. Our strategy deals with the problems associated with NHEJ while at the same time not affecting NHEJ outside of the targeted gene. Using our system random integration should decrease because the exogenously introduced DNA, when bound to I-*Sce*I, is now specifically targeted to the I-*Sce*I site while at the same time the DNA attached to the protein may be sterically hindered to randomly integrate. Also, by bringing the targeting DNA into close proximity of the DSB, NHEJ relative to HR may decrease.

In addition to the efficacy of our system at increasing gene targeting while simultaneously decreasing random integration, the uniqueness of the concept provides benefits of its own. Protein-assisted targeting uses DNA aptamers, which themselves are a relatively new discovery with aptamer selection only being done since the early 1990s (Tuerk and Gold 1990) (Ellington and Szostak 1990). Aptamers have been used as biosensors (Cheng, Ge et al. 2007) (Stojanovic and Kolpashchikov 2004) and as therapeutics (Barbas, Mi et al. 2010) but their use can be simplified to binding and fluorescing (sensor) or binding and inhibiting (therapy) or binding and being endocytosed (therapy). Our aptamer for I-*Sce*I binds and is targeted to a specific DNA site. This represents not only a novel gene targeting strategy but also a novel use of an aptamer.

The novelty and uniqueness of our system lays the foundation for other systems to come. By showing increased gene targeting with our I-*Sce*I aptamer, others could select aptamers for other site-specific endonucleases including recently popular modular endonucleases such as ZFNs, TALENs, and CRSPR systems. The potential aptamer targets are not limited to endonucleases but theoretically any protein involved in recombination. Finally, perhaps the most exciting aspect of the novelty of our system is that because it is stimulating gene targeting by a different means than the other systems described, our protein-assisted targeting strategy can be used in conjunction with these systems. We have shown that our system increases gene targeting in human cells with a DSB, but it is possible that gene targeting can be stimulated even further with a DSB and transient knock-down of NHEJ or overexpression of HR proteins from HR proficient organisms.

A PubMed search of "DNA aptamer" or "RNA aptamer" in the title reveals only 374 papers. Assuming that not all of these aptamers are for unique targets, the number of aptamers is very low. There could be several reasons for the lack of aptamers discovered, but probably the largest hurdle would be the selection process. The difficulty of selection poses a deterrent to many researchers attempting to select for an aptamer. Another likely reason why the number of aptamers is low is because there are several well-characterized aptamers (to thrombin, PMSA, theophylline, etc.) that are used in a variety of applications and it is "safer" to study something well-characterized and use it instead of investing much time and energy to risk not even selecting strong-binding aptamers. Our selection of a DNA aptamer to I-*Sce*I itself is important because it represents a new class

of aptamer targets, namely DNA binding proteins. Not only did we select a new kind of aptamer, we also used this aptamer to affect a functional change *in vivo* using our protein-assisted targeting system. The protein-assisted targeting system we have developed has introduced a new area of aptamer-based research. Our system can be used as a model to other researchers interested in using aptamers to stimulate gene targeting.

5.3 FUTURE DIRECTIONS

The system developed here demonstrates a proof-of-principle for the concept of protein-assisted targeting, and based upon this initial work further studies to elucidate, validate, and improve the system will logically follow. For example, there is the limitation we found with the aptamer system whereby different single-stranded homology regions may disrupt the aptamer region's secondary structure to inhibit binding. To address the question of aptamer modularity in terms of different homology regions, there are several potential directions. Perhaps the most straightforward solution would be to anneal the homology region to its complementary sequence, such that the aptamer region would exist as a 5' single-stranded tail to a mostly double-stranded oligonucleotide. This would eliminate annealing between the aptamer region and the homology region to form a non-aptameric secondary structure. This strategy would be effective in systems where dsDNA is favorable to ssDNA, but unfavorable for systems where ssDNA is favored as a repair substrate. A second possibility is that a linker region could be designed between the aptamer and the homology regions. Several linker regions could be envisioned, the obvious nucleic acid linker, with sequence unlikely to be bound by either the aptamer or

the homology region, or other linkers such as a sulfide linker (Chu, Twu et al. 2006) or a carbon linker (Zhou, Neff et al. 2013)between the two nucleic acids.

There are several potential ways to increase the efficiency of our system from the standpoint of the DNA targeting molecule. Oligonucleotide stability, which could be a limitation to gene correction, could be enhanced by changing the chemistry of the oligonucleotides. Oligonucleotide modification (such as the use of phosphothioate linkages, addition of a 5' cap, 3' thiophosphate, etc.) could improve the stability of the oligonucleotides which could improve gene targeting frequencies. Another way to bolster the gene targeting efficiency of the system would be to use purified oligonucleotides instead of the non-purified oligonucleotides used in our studies. Non-purified 100-mer oligonucleotides synthesized at a coupling efficiency of 99.5% will contain only 60% full-length product, with the other 40% being truncated oligonucleotides (Stafford and Brun 2007) (truncated at the 5' end of the oligonucleotide, which in our system would be the aptamer region). Preliminary data suggests that use of PAGE purified oligonucelotides does increase the level of repair of the aptamer oligonucleotide as compared to the control (data not shown). Additionally, the aptamer itself could be improved for enhanced binding to I-*Sce*I. Although the aptamer was selected from a large pool of oligonucleotides, only ~10^{14} molecules were used in the selection process, out of $4^{36} \sim 5*10^{20}$ potential sequences, meaning that less than 1 sequence for every 5 million possible sequences was tested. This raises the possibility that the strongest binder to I-*Sce*I may not yet be selected. Assuming the aptamer binding affinity could be improved;

modifications to the ISB7 aptamer sequence might lead to a further increase in gene targeting.

In addition to improvements that could be made to the oligonucleotide sequence of the aptamer, verification of the binding could be done by altering the protein at the site of aptamer binding. Electrostatic mapping of the I-*Sce*I protein, along with analysis of the I-*Sce*I structure, could potentially reveal epitopes of the DNA aptamer ISB7. Mutagenesis of these epitopes would lead to a change in the binding affinity of the aptamer and thus confirm the binding site of our ISB7 aptamer. Although there are several different improvements that could be made to the current aptamer for I-*Sce*I, as stated previously a more modular system could be achieved using a different aptamer, to a designer endonuclease such as a ZFN or TALEN. Aptamer technology is constantly evolving and hopefully will continue to be a factor in not only biochemistry but also in the field of molecular biology. The system we developed here will act as the foundation for future studies to come.

APPENDIX A

SUPPLEMENTARY MATERIALS FOR CHAPTER 5

A.1 SELEX procedure using magnetic beads

Preliminary Work

1. Design the oligo (our oligo has 20 bp primers on either end and a 50 bp variable NNNNNN region) and order it PAGE purified

2. Order the magnetic beads that will bind to the protein (for our purpose we are using carboxyl terminated magnetic beads from Bangs Laboratory BM570 that come in a kit BP611)

3. Follow the procedure for protein coupling from Bangs Laboratory

4. Make the binding buffer by adding NaPO4 0.1M, MgCl2 0.01M, and NaCl 0.01M into Tris 0.01M at pH 8 bringing it to a final pH of 7.4

Actual Procedure

5. Starting with 10nmol of purified oligo, place in PCR machine and denature DNA at 95C for 5 minutes to eliminate any double stranded DNA

6. Put the DNA on ice for 10 minutes

7. Wash the protein and beads (which should have coupled) 3 or more times with the tris binding buffer to eliminate any leftover storage buffer. Aspirate (remove with pipette) the supernatant (liquid) each time after washing and vortexing.

8. Continue until there is no suspension after vortexing (clear buffer only)

9. Mix the denatured DNA from step 6 with the beads at 37C and shake for 1 hour

10. Aspirate unbound DNA into a labeled Eppendorf tube after using the magnet to hold the beads in place

11. Wash the bound oligo and beads (coupled with the protein) with 100 microliters of binding buffer

12. Remove unbound oligo and repeat 11 two more times

13. Add 100 microliters of binding buffer and then heat inside the PCR machine at 95C for 10 minutes

14. Transfer this supernatant into a tube labeled “bound oligo”

15. Precipitate DNA with ethanol (100%) at -80C

16. Leave at -80C for 30 minutes

17. Centrifuge at MAX for 10 minutes

18. Aspirate the ethanol, leaving the pellet

19. Use 95% ethanol to wash (use pipette to wash)

20. Vortex the pellet with the 95% ethanol

21. Mash the pellet

22. Vortex again

23. Centrifuge at MAX for 5 minutes

24. Vortex and mash until you have a good pellet

25. Aspirate the ethanol

26. Dry in SpeedVac for 30 minutes

27. Resuspend the pellet in 50 microliters of sterile water

28. Vortex H20 and the pellet

29. Mash the pellet

30. Vortex

31. Perform PCR using 10 microliters of the pellet/H2O

1st PCR

32. Add 1 microliter of Roche Taq polymerase

5 microliters of Mg+2 buffer (10X concentrated)

0.5 microliters of 50 mM DNTPs

0.5 microliters of 100 mM P1 (the primer starting at the 5' end of the oligo)

0.5 microliters of 100 mM P2

10 microliters of the pellet/H2O

32.5 microliters of sterile H2O

Total = 50 microliters

Cycle = 1. 94C for 2 minutes

2. 94C for 15 minutes
3. 44C for 15 minutes
4. 72C for 30 minutes

Repeat steps 2-4 a total of 29 times

2nd PCR

33. After the 1st PCR, then perform an asymmetric PCR. Instead of using equal amounts of each primer, we use 100 times more of the forward primer so that we end up with only the sequences we want and not their reverse complement.

Add 1 microliter of Roche Taq polymerase

5 microliters of Mg+2 buffer (10X concentrated)

0.5 microliters of 50 mM DNTPs

0.5 microliters of 100 mM P1 (1 uM final)

X microliters of 100 mM P2 (want 0.01 uM final)

25 microliters of oligos from PCR1

X microliters of sterile H2O (can't know till P2 is calculated)

Total = 50 microliters

Cycle = 1. 94C for 2 minutes

5. 94C for 15 minutes
6. 44C for 15 minutes
7. 72C for 30 minutes

Repeat steps 2-4 a total of 10 times

34. PCR product 2 is used (10 nmol) in the next round of SELEX

35. Every five rounds of SELEX (starting with the first), use 250 uCi P32 labeled dATPs (1 uL) and then measure

% = oligo bound / unbound + bound + beads

And at 80%, stop SELEX

REFERENCES

Adachi, N., H. Suzuki, S. Iiizumi and H. Koyama (2003). "Hypersensitivity of nonhomologous DNA end-joining mutants to VP-16 and ICRF-193: implications for the repair of topoisomerase II-mediated DNA damage." J Biol Chem **278**(38): 35897-35902.

Aiuti, A., S. Slavin, M. Aker, F. Ficara, S. Deola, A. Mortellaro, S. Morecki, G. Andolfi, A. Tabucchi, F. Carlucci, E. Marinello, F. Cattaneo, S. Vai, P. Servida, R. Miniero, M. G. Roncarolo and C. Bordignon (2002). "Correction of ADA-SCID by stem cell gene therapy combined with nonmyeloablative conditioning." Science **296**(5577): 2410-2413.

Allen, C., A. Kurimasa, M. A. Brenneman, D. J. Chen and J. A. Nickoloff (2002). "DNA-dependent protein kinase suppresses double-strand break-induced and spontaneous homologous recombination." Proc Natl Acad Sci U S A **99**(6): 3758-3763.

Aylon, Y. and M. Kupiec (2004). "DSB repair: the yeast paradigm." DNA Repair (Amst) **3**(8-9): 797-815.

Baleja, J. D., V. Thanabal and G. Wagner (1997). "Refined solution structure of the DNA-binding domain of GAL4 and use of 3J(113Cd,1H) in structure determination." J Biomol NMR **10**(4): 397-401.

Banga, S. S. and J. B. Boyd (1992). "Oligonucleotide-directed site-specific mutagenesis in Drosophila melanogaster." Proc Natl Acad Sci U S A **89**(5): 1735-1739.

Barbas, A. S., J. Mi, B. M. Clary and R. R. White (2010). "Aptamer applications for targeted cancer therapy." Future Oncol **6**(7): 1117-1126.

Belfort, M. and R. J. Roberts (1997). "Homing endonucleases: keeping the house in order." Nucleic Acids Res **25**(17): 3379-3388.

Berezovski, M., A. Drabovich, S. M. Krylova, M. Musheev, V. Okhonin, A. Petrov and S. N. Krylov (2005). "Nonequilibrium capillary electrophoresis of equilibrium mixtures: a universal tool for development of aptamers." J Am Chem Soc **127**(9): 3165-3171.

Berezovski, M. and S. N. Krylov (2002). "Nonequilibrium capillary electrophoresis of equilibrium mixtures--a single experiment reveals equilibrium and kinetic parameters of protein-DNA interactions." J Am Chem Soc **124**(46): 13674-13675.

Berezovski, M. and S. N. Krylov (2005). "Thermochemistry of protein-DNA interaction studied with temperature-controlled nonequilibrium capillary electrophoresis of equilibrium mixtures." Anal Chem **77**(5): 1526-1529.

Berezovski, M., M. Musheev, A. Drabovich and S. N. Krylov (2006). "Non-SELEX selection of aptamers." J Am Chem Soc **128**(5): 1410-1411.

Berezovski, M. V., M. U. Musheev, A. P. Drabovich, J. V. Jitkova and S. N. Krylov (2006). "Non-SELEX: selection of aptamers without intermediate amplification of candidate oligonucleotides." Nat Protoc **1**(3): 1359-1369.

Bertolini, L. R., M. Bertolini, E. A. Maga, K. R. Madden and J. D. Murray (2009). "Increased gene targeting in Ku70 and Xrcc4 transiently deficient human somatic cells." Mol Biotechnol **41**(2): 106-114.

Beylot, B. and A. Spassky (2001). "Chemical probing shows that the intron-encoded endonuclease I-SceI distorts DNA through binding in monomeric form to its homing site." J Biol Chem **276**(27): 25243-25253.

Bindra, R. S., A. G. Goglia, M. Jasin and S. N. Powell (2013). "Development of an assay to measure mutagenic non-homologous end-joining repair activity in mammalian cells." Nucleic Acids Res.

Blaese, R. M., K. W. Culver, A. D. Miller, C. S. Carter, T. Fleisher, M. Clerici, G. Shearer, L. Chang, Y. Chiang, P. Tolstoshev, J. J. Greenblatt, S. A. Rosenberg, H. Klein, M. Berger, C. A. Mullen, W. J. Ramsey, L. Muul, R. A. Morgan and W. F. Anderson (1995). "T lymphocyte-directed gene therapy for ADA- SCID: initial trial results after 4 years." Science **270**(5235): 475-480.

Bordignon, C., L. D. Notarangelo, N. Nobili, G. Ferrari, G. Casorati, P. Panina, E. Mazzolari, D. Maggioni, C. Rossi, P. Servida, A. G. Ugazio and F. Mavilio (1995). "Gene therapy in peripheral blood lymphocytes and bone marrow for ADA-immunodeficient patients." Science **270**(5235): 470-475.

Cao, Z., R. Tong, A. Mishra, W. Xu, G. C. Wong, J. Cheng and Y. Lu (2009). "Reversible cell-specific drug delivery with aptamer-functionalized liposomes." Angew Chem Int Ed Engl **48**(35): 6494-6498.

Carothers, J. M., J. A. Goler, Y. Kapoor, L. Lara and J. D. Keasling (2010). "Selecting RNA aptamers for synthetic biology: investigating magnesium dependence and predicting binding affinity." Nucleic Acids Res **38**(8): 2736-2747.

Cavazzana-Calvo, M., S. Hacein-Bey, G. de Saint Basile, F. Gross, E. Yvon, P. Nusbaum, F. Selz, C. Hue, S. Certain, J. L. Casanova, P. Bousso, F. L. Deist and A. Fischer (2000). "Gene therapy of human severe combined immunodeficiency (SCID)-X1 disease." Science **288**(5466): 669-672.

Chan, C. K., S. Hubner, W. Hu and D. A. Jans (1998). "Mutual exclusivity of DNA binding and nuclear localization signal recognition by the yeast transcription factor GAL4: implications for nonviral DNA delivery." Gene Ther **5**(9): 1204-1212.

Chapman, J. R., M. R. Taylor and S. J. Boulton (2012). "Playing the end game: DNA double-strand break repair pathway choice." Mol Cell **47**(4): 497-510.

Cheng, A. K., B. Ge and H. Z. Yu (2007). "Aptamer-based biosensors for label-free voltammetric detection of lysozyme." Anal Chem **79**(14): 5158-5164.

Chu, T. C., K. Y. Twu, A. D. Ellington and M. Levy (2006). "Aptamer mediated siRNA delivery." Nucleic Acids Res **34**(10): e73.

Colleaux, L., L. D'Auriol, F. Galibert and B. Dujon (1988). "Recognition and cleavage site of the intron-encoded omega transposase." Proc Natl Acad Sci U S A **85**(16): 6022-6026.

Cong, L., F. A. Ran, D. Cox, S. Lin, R. Barretto, N. Habib, P. D. Hsu, X. Wu, W. Jiang, L. A. Marraffini and F. Zhang (2013). "Multiplex genome engineering using CRISPR/Cas systems." Science **339**(6121): 819-823.

Di Primio, C., A. Galli, T. Cervelli, M. Zoppe and G. Rainaldi (2005). "Potentiation of gene targeting in human cells by expression of Saccharomyces cerevisiae Rad52." Nucleic Acids Res **33**(14): 4639-4648.

Ellington, A. D. and J. W. Szostak (1990). "In vitro selection of RNA molecules that bind specific ligands." Nature **346**(6287): 818-822.

Fields, S. and O. Song (1989). "A novel genetic system to detect protein-protein interactions." Nature **340**(6230): 245-246.

Gabriel, R., A. Lombardo, A. Arens, J. C. Miller, P. Genovese, C. Kaeppel, A. Nowrouzi, C. C. Bartholomae, J. Wang, G. Friedman, M. C. Holmes, P. D. Gregory, H. Glimm, M. Schmidt, L. Naldini and C. von Kalle (2011). "An unbiased genome-wide analysis of zinc-finger nuclease specificity." Nat Biotechnol **29**(9): 816-823.

Gaur, R., P. K. Gupta, A. Goyal, W. Wels and Y. Singh (2002). "Delivery of nucleic acid into mammalian cells by anthrax toxin." Biochem Biophys Res Commun **297**(5): 1121-1127.

Ge, W. W., W. Wen, W. Strong, C. Leystra-Lantz and M. J. Strong (2005). "Mutant copper-zinc superoxide dismutase binds to and destabilizes human low molecular weight neurofilament mRNA." J Biol Chem **280**(1): 118-124.

Ginn, S. L., I. E. Alexander, M. L. Edelstein, M. R. Abedi and J. Wixon (2013). "Gene therapy clinical trials worldwide to 2012 - an update." J Gene Med **15**(2): 65-77.

Grieger, J. C., V. W. Choi and R. J. Samulski (2006). "Production and characterization of adeno-associated viral vectors." Nat Protoc **1**(3): 1412-1428.

Gustavsson, M., J. Lehtio, S. Denman, T. T. Teeri, K. Hult and M. Martinelle (2001). "Stable linker peptides for a cellulose-binding domain-lipase fusion protein expressed in Pichia pastoris." Protein Eng **14**(9): 711-715.

Hacein-Bey-Abina, S., C. Von Kalle, M. Schmidt, M. P. McCormack, N. Wulffraat, P. Leboulch, A. Lim, C. S. Osborne, R. Pawliuk, E. Morillon, R. Sorensen, A. Forster, P. Fraser, J. I. Cohen, G. de Saint Basile, I. Alexander, U. Wintergerst, T. Frebourg, A. Aurias, D. Stoppa-Lyonnet, S. Romana, I. Radford-Weiss, F. Gross, F. Valensi, E. Delabesse, E. Macintyre, F. Sigaux, J. Soulier, L. E. Leiva, M. Wissler, C. Prinz, T. H. Rabbitts, F. Le Deist, A. Fischer and M. Cavazzana-Calvo (2003). "LMO2-associated clonal T cell proliferation in two patients after gene therapy for SCID-X1." Science **302**(5644): 415-419.

Heyer, W. D., K. T. Ehmsen and J. Liu (2010). "Regulation of homologous recombination in eukaryotes." Annu Rev Genet **44**: 113-139.

Hirsch, M. L., F. Storici, C. W. Li, V. W. Choi and J. Samulski (2009). "AAV Recombineering with Single Strand Oligonucleotides." Plos One **4**(11): -.

Hockemeyer, D., H. Wang, S. Kiani, C. S. Lai, Q. Gao, J. P. Cassady, G. J. Cost, L. Zhang, Y. Santiago, J. C. Miller, B. Zeitler, J. M. Cherone, X. Meng, S. J. Hinkley, E. J. Rebar, P. D. Gregory, F. D. Urnov and R. Jaenisch (2011). "Genetic engineering of human pluripotent cells using TALE nucleases." Nat Biotechnol **29**(8): 731-734.

Iiizumi, S., A. Kurosawa, S. So, Y. Ishii, Y. Chikaraishi, A. Ishii, H. Koyama and N. Adachi (2008). "Impact of non-homologous end-joining deficiency on random and targeted DNA integration: implications for gene targeting." Nucleic Acids Res **36**(19): 6333-6342.

Ivanov, E. L., N. Sugawara, J. Fishman-Lobell and J. E. Haber (1996). "Genetic requirements for the single-strand annealing pathway of double-strand break repair in Saccharomyces cerevisiae." Genetics **142**(3): 693-704.

Jackson, G. W., Strych, U. (2010). Nucleic acid aptamers for diagnostics and therapeutics: global markets.

Jayasena, S. D. (1999). "Aptamers: an emerging class of molecules that rival antibodies in diagnostics." Clin Chem **45**(9): 1628-1650.

Jenison, R. D., S. C. Gill, A. Pardi and B. Polisky (1994). "High-resolution molecular discrimination by RNA." Science **263**(5152): 1425-1429.

Kim, Y. G., J. Cha and S. Chandrasegaran (1996). "Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain." Proc Natl Acad Sci U S A **93**(3): 1156-1160.

Kim, Y. G., J. Smith, M. Durgesha and S. Chandrasegaran (1998). "Chimeric restriction enzyme: Gal4 fusion to FokI cleavage domain." Biol Chem **379**(4-5): 489-495.

Knegtel, R. M., R. H. Fogh, G. Ottleben, H. Ruterjans, P. Dumoulin, M. Schnarr, R. Boelens and R. Kaptein (1995). "A model for the LexA repressor DNA complex." Proteins **21**(3): 226-236.

Kohn, D. B., M. Sadelain and J. C. Glorioso (2003). "Occurrence of leukaemia following gene therapy of X-linked SCID." Nat Rev Cancer **3**(7): 477-488.

Kohn, D. B., K. I. Weinberg, J. A. Nolta, L. N. Heiss, C. Lenarsky, G. M. Crooks, M. E. Hanley, G. Annett, J. S. Brooks, A. el-Khoureiy and et al. (1995). "Engraftment of gene-modified umbilical cord blood cells in neonates with adenosine deaminase deficiency." Nat Med **1**(10): 1017-1023.

Laughon, A. and R. F. Gesteland (1982). "Isolation and preliminary characterization of the GAL4 gene, a positive regulator of transcription in yeast." Proc Natl Acad Sci U S A **79**(22): 6827-6831.

Lu, I. L., C. Y. Lin, S. B. Lin, S. T. Chen, L. Y. Yeh, F. Y. Yang and L. C. Au (2003). "Correction/mutation of acid alpha-D-glucosidase gene by modified single-stranded oligonucleotides: in vitro and in vivo studies." Gene Ther **10**(22): 1910-1916.

Lungwitz, U., M. Breunig, T. Blunk and A. Gopferich (2005). "Polyethylenimine-based non-viral gene delivery systems." Eur J Pharm Biopharm **60**(2): 247-266.

Ma, J. and M. Ptashne (1987). "Deletion analysis of GAL4 defines two transcriptional activating segments." Cell **48**(5): 847-853.

Mali, P., L. Yang, K. M. Esvelt, J. Aach, M. Guell, J. E. DiCarlo, J. E. Norville and G. M. Church (2013). "RNA-guided human genome engineering via Cas9." Science **339**(6121): 823-826.

Mavilio, F., G. Pellegrini, S. Ferrari, F. Di Nunzio, E. Di Iorio, A. Recchia, G. Maruggi, G. Ferrari, E. Provasi, C. Bonini, S. Capurro, A. Conti, C. Magnoni, A. Giannetti and M. De Luca (2006). "Correction of junctional epidermolysis bullosa by transplantation of genetically modified epidermal stem cells." Nat Med **12**(12): 1397-1402.

Mendonsa, S. D. and M. T. Bowser (2004). "In vitro evolution of functional DNA using capillary electrophoresis." J Am Chem Soc **126**(1): 20-21.

Morse, D. P. (2007). "Direct selection of RNA beacon aptamers." Biochem Biophys Res Commun **359**(1): 94-101.

Moscou, M. J. and A. J. Bogdanove (2009). "A simple cipher governs DNA recognition by TAL effectors." Science **326**(5959): 1501.

Mosing, R. K., S. D. Mendonsa and M. T. Bowser (2005). "Capillary electrophoresis-SELEX selection of aptamers with affinity for HIV-1 reverse transcriptase." Anal Chem **77**(19): 6107-6112.

Natale, M., C. Bisson, G. Monti, A. Peltran, L. P. Garoffo, S. Valentini, C. Fabris, E. Bertino, A. Coscia and A. Conti (2004). "Cow's milk allergens identification by two-dimensional immunoblotting and mass spectrometry." Mol Nutr Food Res **48**(5): 363-369.

Nickerson, H. D. and W. H. Colledge (2003). "A comparison of gene repair strategies in cell culture using a lacZ reporter system." Gene Ther **10**(18): 1584-1591.

Nussbaum, A., M. Shalit and A. Cohen (1992). "Restriction-stimulated homologous recombination of plasmids by the RecE pathway of Escherichia coli." Genetics **130**(1): 37-49.

Oshima, Y. (1982). Regulatory circuits for gene expression: the metabolism of galactose and phosphate. Molecular Biology of the Yeast Saccharomyces Cerevisiae: Metabolism and Gene Expression. J. N. Strathern. Cold Spring Harbor, Cold Spring Harbor Laboratory**:** 159-180.

Pattanayak, V., C. L. Ramirez, J. K. Joung and D. R. Liu (2011). "Revealing off-target cleavage specificities of zinc-finger nucleases by in vitro selection." Nat Methods **8**(9): 765-770.

Paul, R. W., K. E. Weisser, A. Loomis, D. L. Sloane, D. LaFoe, E. M. Atkinson and R. W. Overell (1997). "Gene transfer using a novel fusion protein, GAL4/invasin." Hum Gene Ther **8**(10): 1253-1262.

Petek, L. M., D. W. Russell and D. G. Miller (2010). "Frequent endonuclease cleavage at off-target locations in vivo." Mol Ther **18**(5): 983-986.

Pierce, A. J., P. Hu, M. Han, N. Ellis and M. Jasin (2001). "Ku DNA end-binding protein modulates homologous repair of double-strand breaks in mammalian cells." Genes Dev **15**(24): 3237-3242.

Plessis, A., A. Perrin, J. E. Haber and B. Dujon (1992). "Site-specific recombination determined by I-SceI, a mitochondrial group I intron-encoded endonuclease expressed in the yeast nucleus." Genetics **130**(3): 451-460.

Porteus, M. H. and D. Baltimore (2003). "Chimeric nucleases stimulate gene targeting in human cells." Science **300**(5620): 763.

Potts, P. R., M. H. Porteus and H. Yu (2006). "Human SMC5/6 complex promotes sister chromatid homologous recombination by recruiting the SMC1/3 cohesin complex to double-strand breaks." EMBO J **25**(14): 3377-3388.

Puchta, H., B. Dujon and B. Hohn (1993). "Homologous recombination in plant cells is enhanced by in vivo induction of double strand breaks into DNA by a site-specific endonuclease." Nucleic Acids Res **21**(22): 5034-5040.

Radecke, F., S. Radecke and K. Schwarz (2004). "Unmodified oligodeoxynucleotides require single-strandedness to induce targeted repair of a chromosomal EGFP gene." J Gene Med **6**(11): 1257-1271.

Rouet, P., F. Smih and M. Jasin (1994). "Expression of a site-specific endonuclease stimulates homologous recombination in mammalian cells." Proc Natl Acad Sci U S A **91**(13): 6064-6068.

Sadelain, M. (2006). "Recent advances in globin gene transfer for the treatment of beta-thalassemia and sickle cell anemia." Curr Opin Hematol **13**(3): 142-148.

Sarkar, T., C. C. Conwell, L. C. Harvey, C. T. Santai and N. V. Hud (2005). "Condensation of oligonucleotides assembled into nicked and gapped duplexes: potential structures for oligonucleotide delivery." Nucleic Acids Res **33**(1): 143-151.

Sefah, K., D. Shangguan, X. Xiong, M. B. O'Donoghue and W. Tan (2010). "Development of DNA aptamers using Cell-SELEX." Nat Protoc **5**(6): 1169-1185.

Shcherbakova, O. G., V. A. Lanzov, H. Ogawa and M. V. Filatov (2000). "Overexpression of bacterial RecA protein stimulates homologous recombination in somatic mammalian cells." Mutat Res **459**(1): 65-71.

Singerman, L. J., H. Masonson, M. Patel, A. P. Adamis, R. Buggage, E. Cunningham, M. Goldbaum, B. Katz and D. Guyer (2008). "Pegaptanib sodium for neovascular age-related macular degeneration: third-year safety results of the VEGF Inhibition Study in Ocular Neovascularisation (VISION) trial." Br J Ophthalmol **92**(12): 1606-1611.

Smih, F., P. Rouet, P. J. Romanienko and M. Jasin (1995). "Double-strand breaks at the target locus stimulate gene targeting in embryonic stem cells." Nucleic Acids Res **23**(24): 5012-5019.

Stafford, P. and M. Brun (2007). "Three methods for optimization of cross-laboratory and cross-platform microarray expression data." Nucleic Acids Res **35**(10): e72.

Stoddard, B. L. (2011). "Homing endonucleases: from microbial genetic invaders to reagents for targeted DNA modification." Structure **19**(1): 7-15.

Stojanovic, M. N. and D. M. Kolpashchikov (2004). "Modular aptameric sensors." J Am Chem Soc **126**(30): 9266-9270.

Storici, F., C. L. Durham, D. A. Gordenin and M. A. Resnick (2003). "Chromosomal site-specific double-strand breaks are efficiently targeted for repair by oligonucleotides in yeast." Proc Natl Acad Sci U S A **100**(25): 14994-14999.

Stuckey, S., K. Mukherjee and F. Storici (2011). "In vivo site-specific mutagenesis and gene collage using the delitto perfetto system in yeast Saccharomyces cerevisiae." Methods Mol Biol **745**: 173-191.

Taylor, I. C., J. L. Workman, T. J. Schuetz and R. E. Kingston (1991). "Facilitated binding of GAL4 and heat shock factor to nucleosomal templates: differential function of DNA-binding domains." Genes Dev **5**(7): 1285-1298.

Thermes, V., C. Grabher, F. Ristoratore, F. Bourrat, A. Choulika, J. Wittbrodt and J. S. Joly (2002). "I-SceI meganuclease mediates highly efficient transgenesis in fish." Mech Dev **118**(1-2): 91-98.

Thomas, C. E., A. Ehrhardt and M. A. Kay (2003). "Progress and problems with the use of viral vectors for gene therapy." Nat Rev Genet **4**(5): 346-358.

Tuerk, C. and L. Gold (1990). "Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase." Science **249**(4968): 505-510.

Vasquez, K. M., K. Marburger, Z. Intody and J. H. Wilson (2001). "Manipulating the mammalian genome by homologous recombination." Proc Natl Acad Sci U S A **98**(15): 8403-8410.

Wels, W., B. Groner and N. E. Hynes (1996). "Intervention in receptor tyrosine kinase-mediated pathways: recombinant antibody fusion proteins targeted to ErbB2." Curr Top Microbiol Immunol **213 (Pt 3)**: 113-128.

Woodbine, L., H. Brunton, A. A. Goodarzi, A. Shibata and P. A. Jeggo (2011). "Endogenously induced DNA double strand breaks arise in heterochromatic DNA regions and require ataxia telangiectasia mutated and Artemis for their repair." Nucleic Acids Res **39**(16): 6986-6997.

Xu, W. and Y. Lu (2010). "Label-free fluorescent aptamer sensor based on regulation of malachite green fluorescence." Anal Chem **82**(2): 574-578.

Xu, X.-h., B.-r. Chi, X. Li, E.-c. Yang, P. Gao, Y. Liu, P. Jia, S.-f. Kan, Z.-m. Wen and N.-y. Jin (2010). "Expression and Activities Experiment of DNA Transduction Motif Based on GAL4 in Pichia Pastoris." Chemical Research Chinese Universities **26**(2): 221-224.

Zhao, H. L., X. Q. Yao, C. Xue, Y. Wang, X. H. Xiong and Z. M. Liu (2008). "Increasing the homogeneity, stability and activity of human serum albumin and interferon-alpha2b fusion protein by linker engineering." Protein Expr Purif **61**(1): 73-77.

Zhou, J., C. P. Neff, P. Swiderski, H. Li, D. D. Smith, T. Aboellail, L. Remling-Mulder, R. Akkina and J. J. Rossi (2013). "Functional in vivo delivery of multiplexed anti-HIV-1 siRNAs via a chemically synthesized aptamer with a sticky bridge." Mol Ther **21**(1): 192-200.