SELECTION OF

CELL-INTERNALIZING CIRCULAR DNA APTAMERS

SELECTION OF

CELL-INTERNALIZING CIRCULAR DNA APTAMERS

By JIMMY GU, B.Sc.

A Thesis Submitted to the School of Graduate Studies in Partial Fulfillment of the Requirements for the Degree Master of Science

MASTER OF SCIENCE (2011) **McMaster University**
(Biochemistry & Biomedical Sciences) **Hamilton, Ontario**

TITLE:	**Selection of Cell-Internalizing Circular DNA Aptamers**
AUTHOR:	**Jimmy Gu, B.Sc. (McMaster University)**
SUPERVISOR:	**Professor Yingfu Li (Simon Fraser University)**
NUMBER OF PAGES:	**x, 51**

ABSTRACT

Adaptation of nucleic acid *in vitro* selection for whole cell targets has been demonstrated to be an effective means of isolating useful sequences with applications in biomarker detection and therapeutics. The problem of efficient delivery of materials across cell membranes is common to a variety of research and medical fields. Existing aptamers isolated in surfacing binding selections have been successfully adapted for cell targeted therapies through complex modifications. However, better aptamers may be derived from a selection optimized to isolate internalized sequences directly. A cell selection experiment with the goal of identifying circular random-sequence DNA aptamers with the ability to facilitate their own internalization into MCF7 cells was conducted. Several classes of sequences isolated from this selection were shown to target cell nuclei at a rate significantly greater than control sequences as determined by qPCR relative recovery assays supported by *in situ* RCA fluorescence microscopy data. The localization of functional DNA sequences at the subcellular and intercellular levels suggests a receptor mediated mechanism. Techniques for the selection, purification and fluorescent detection of small circular DNAs were also developed for this study. Further work to characterize and identify targets should be pursued to better understand the mechanism of internalization and judge the suitability of G18d sequences as a delivery platform.

ACKNOWLEDGEMENTS

My sincere thanks to Dr. Yingfu Li for the learning opportunities he has offered me over the past five years. I greatly value his enthusiastic approach to research and his helpful guidance. I consider myself very fortunate to have had his support in pursuing a diverse range of research interests during my time in the Li Lab.

I am also appreciative to my supervisory committee members, Dr. David Andrews and Dr. Zhou Xing for their expert advice and insight in this research project.

Finally, I am indebted to the members of the Li Lab and the Andrews Lab for their invaluable technical expertise and advice with this project.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

BSA	Bovine serum albumin
Ct	Cycle threshold
DNA	Deoxyribonucleic acid
dNTP	Deoxyribonucleotide triphosphate
EDTA	Ethylenediaminetetraacetic acid
FBS	Fetal bovine serum
HBSS	Hank's buffered salt solution
LB	Lysogeny broth
MEM	Minimal essential medium
mRNA	Messenger ribonucleic acid
NLS	Nuclear localization signal
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PSMA	Prostate specific membrane antigen
qPCR	Quantitative polymerase chain reaction
RCA	Rolling circle amplification
RNA	Ribonucleic acid
SDS	Sodium dodecyl sulfate
SELEX	Systematic evolution of ligands by exponential enrichment

siRNA	Small interfering ribonucleic acid
tRNA	Transfer ribonucleic acid

Chapter 1: Introduction

1.1 *Aptamers and in vitro Selection*

A common problem in the fields of molecular medicine, diagnostic imaging and genetic manipulation has been one of cellular targeting. Many small molecule drugs are designed to target specific enzymes or receptors in pathways of a specific cell type, broad distribution of a given drug increases the risk of side effects caused by nonspecific activity. The problem of off-target effects facing siRNA technologies could benefit from targeting at the cellular level (1). Additionally, there is a need for molecular probes capable of identifying specific cell types for use in disease diagnosis. Currently, applications requiring high specificity probes rely on small molecules or monoclonal antibodies. However, with the introduction of the *in vitro* selection technique in 1990, also referred to as systematic evolution of ligands by exponential enrichment (SELEX), nucleic acids with specific binding abilities called aptamers have been shown to rival antibodies in specificity and versatility while surpassing antibody technologies in cost, ease of production and stability (2-6). The *in vitro* selection method assumes that functional sequences exist within a large random sequence population of oligonucleotides due to the variety of three dimensional structures possible through intramolecular hydrogen bonding and metal ion interactions; analogous to protein folding. The key features of *in vitro* selection mirror the processes of natural selection and involve the use of a random sequence library of oligonucleotides and a cyclic progression of selecting increasingly functional molecules from the random pool until the final population is enriched for sequences with the desired properties and functionality. Accelerated by the decreasing cost of DNA synthesis, sequencing and PCR technologies, the use of *in vitro*

selection methods to generate functional nucleic acids has increased rapidly. An evolution of the *in vitro* selection method initially used to select for cell surface markers termed "cell SELEX" first appeared in the Pubmed database in 2003 with 32 publications to date. Cell selection differs from traditional *in vitro* selection by its use of whole cells as targets versus single or multi-molecular complexes.

1.2 *Developments in Targeting*

Early small molecule drugs had limited ability to target specific cell types. Properties of the small molecules themselves were manipulated chemically to achieve a balance between efficacy, distribution and toxicity. The problem of targeting a specific cell type requires that a cell possess a unique surface feature that can be reliably and specifically recognized. In the case of cancer cells, these cell surface markers can often have high similarity to wild-type cells. Significant work has been done developing antibody based targeting systems capable of binding unique surface markers with high specificity (7). Antibodies can serve a dual purpose in the sense that binding a specific surface receptor of a target cell not only requires specificity but can also interfere with cellular signalling within a pathway of the target cell that results in a therapeutic effect (7). This dual functionality would also be expected to apply to DNA and RNA based aptamer therapeutics due to the parallels between the nucleic acid target recognition and antibody target recognition and their comparable nanomolar scale binding affinities (4,5). There has also been rapid development in the use of nanoparticle and lipid based drug delivery systems in recent years to deliver small molecule drugs that would otherwise not have suitable distribution profiles *in vivo*. Nanomaterials have been an active area of research for creating drug carriers with the ability for controlled release within a specific

environmental context; a broader form of targeting as compared to the capabilities of antibodies or aptamers. Properties such as pH, light or temperature can be used to trigger the release or activation of drug carrying nanomaterials (8). Studies have shown liposomes with surfaces modified with anti-tumour antibodies for the targeted delivery of siRNA to cancer cells (9). A growing area of research in cell targeting and drug delivery investigates the use of aptamers as tools alongside antibodies.

1.3 *Comparing Aptamers to Antibodies*

Many antibody based therapeutics and several aptamer based therapeutics are currently in clinical trials. The repertoire of cell targeting aptamers is currently limited to several cell surface markers, with the most well characterized aptamers targeting the prostate specific membrane antigen (PSMA) commonly used as a marker for prostate cancer(6,10). As an alternative to antibodies, aptamers have several advantages that make them attractive for diagnostics and therapeutics. Firstly, oligonucleotides can be easily and cheaply produced through chemical means which contribute to fewer complications, consistent yields and simplified regulatory approval (6). The increased chemical stability of aptamers, particularly DNA based, makes them attractive as well. Limited immune responses have been reported for DNA based therapeutics making them a safer choice for use *in vivo* (11). To be an effective tool for drug delivery and diagnostics, a wider range of aptamers must be developed against a variety of target cell types important in disease. Selecting a target among a diverse cell surface landscape presents unique difficulties for cell selection experiments and adding the requirement of efficient cellular internalization further limits the field of targets (12,13). Beyond the challenges of conducting selection with whole

cells, the stability and distribution of nucleic acid therapeutics *in vivo* also must be considered for an effective delivery platform.

1.4 *Cell Selection*

Cell selection experiments are generally conducted using whole live cells as the target. Advantages of whole cell targets include proteins that are more likely to be in their native conformations. Extracellular faces of proteins are preserved and prior information about the ultimate target of a cell selection experiment is not required. A major feature of the cell selection method is that it simply looks for a specific end result with no regard to the mechanisms or the pathways required in achieving it. This simplicity allows for all possible pathways to be considered and increases the likelihood of achieving the selected functionality. However, cell selection conditions must be considered thoughtfully as the quality of selected aptamers depends greatly on the conditions of selection. While the principle behind cell selection is simple, there are significant challenges to carrying out cell selection experiments. Cleavage of surface proteins with trypsin can be employed to remove surface bound sequences in experiments searching for internalized aptamers, in addition blocking components such as tRNA and BSA can be used to reduce non-specific binding (14,15). The majority of the cell internalizing aptamers discovered thus far have been found through cell selection experiments searching for surface binding aptamers with the goal of applications in biomarker detection, such as the sgc8 RNA aptamer that was found to bind to tyrosine kinase 7 transmembrane protein expressed on T-cell acute lymphoblastic leukemia cells (16). The sgc8 aptamer was later discovered to also be internalized into the cell (17). Subsequent studies have used sgc8 to study the ability of aptamers to function as carriers for drugs and fluorescent dyes or as targeting sequences

to direct liposomes (18,19). To find truly effective aptamers that are specialized at cellular internalization, the selection experiments should select for aptamers that internalize into cells while excluding aptamers that simply bind the surface of cells. As in the case with sgc8, it would be expected that aptamers capable of entering the cell would likely also be good surface binders. It would also be expected that many surface binding aptamers are eventually internalized into the cell through plasma membrane recycling pathways. However, the identification of cell internalizing aptamers is also dependent upon the stability of the sequence post internalization, as rapidly degraded sequences may not survive long enough to be detected. Nuclease resistance would be an important feature of internalization aptamers for delivery of nucleic acid based therapeutics such as mRNA cleaving deoxyribozymes, antisense oligonucleotides or siRNA for gene knockdown. A nuclear targeted aptamer that exploits a pathway that minimizes nuclease exposure would be ideally suited for gene delivery. A further advantage of selecting for internalization is the potential for isolating aptamer species capable of localizing to specific subcellular compartments. A panel of aptamers capable of localizing to the endoplasmic reticulum, the Golgi apparatus, mitochondria or the nucleus would have applications in probe development for use in biosensors and fluorescence microscopy probes as well as subcellular targeting of therapeutics. Once a library of cell internalizing aptamers has been isolated, a reselection experiment using fluorescently tagged aptamers could be done to determine the subcellular fate of specific aptamer sequences. A feature of the cyclic nature of the cell selection method is the ability to branch off selection experiments at various rounds of selection. For example, a cell selection experiment generating a pool of cell internalizing aptamers could be diverged after several rounds to select for rapidly internalizing species by decreasing the target presentation time for one

group while increasing the time for another group to select for exceptionally stable species. This versatility in the cell selection method allows for optimization of aptamers for specific applications.

1.5 *Stability of Nucleic Acid Therapeutics in vivo*

The problem of nucleic acid stability *in vivo* has generally been ignored during the development of the selection experiment to avoid adding complexity to an already complex and labour intensive experiment. It is well known that nucleic acids, especially RNA, are susceptible to nucleases found in the cellular environment. Therefore, the successful application of aptamers as cell penetrating molecules must include a consideration of *in vivo* stability. A variety of chemical modifications are available to increase the stability of nucleic acids *in vivo*, much of the research thus far has been done on stabilizing RNA molecules using a variety of modifications at the 2` position such as 2`-F, 2`-$NH_2$ and 2`-O-CH_3 modifications. These modifications have been shown to significantly increase the stability of RNA in serum containing media (20). Similarly, locked nucleic acids have been found to enhance base pairing interactions as well as provide nuclease resistance (21). An important concern when applying stabilizing modifications to nucleic acids after selection is the potential for disturbing the secondary structure that gives the aptamer its functionality. Incorporating chemical modifications into the selection experiment itself would avoid this problem as any functional features dependent upon the modifications would be incorporated into the functional aptamer. An alternative method to generate stable nuclease resistant aptamers for use in the cell involves selecting for naturally nuclease resistant sequences. A recent study comparing the stability of aptamers derived from G-rich libraries found that the formation of G-

quadruplex structures enhanced the stability of aptamers in serum as well as their uptake into cells (22,23). G-quadruplexes are well characterized secondary structures of DNA that are highly stable (24). The structures were also found to target and kill cancer cells preferentially to non-cancer cells (23). It is hypothesised that the G-quadruplex structure has a natural ability to bind the highly expressed nucleolin protein found on the surface of cancer cells which facilitates the aptamer's internalization. AS1411 is a DNA aptamer currently in clinical trials for use as a cancer therapeutic (25). It appears to be effective against several different types of cancer and features a G-rich sequence and G-quadruplex structures. Circular nucleic acids may also contribute to *in vivo* aptamer stability as there are no 5` or 3` ends available for exonuclease digestion. A combination of these methods can be used to design cell selection experiments where the stability of aptamers *in vivo* is an initial concern factored into the cell selection experiment itself potentially resulting in more robust cell internalizing aptamers.

1.6 *Aptamer Complexes*

For aptamers obtained from a cell selection experiment to be truly useful, it will likely have to be complexed with other molecules. Some aptamers may have an intrinsic ability to perform a therapeutically useful function, in the same way monoclonal antibody based therapies function, by binding a receptor and altering a signalling pathway. Other aptamers may be adept at cellular internalization and nuclease resistance and will need to be functionalized by the addition of therapeutic molecules to become useful. A variety of methods exist for the chemical modification of nucleic acids, chemistry can be done at the 5`, 3`, sugar or base. Similar to nucleic acid stabilizing modifications, changes to an aptamer after selection can be difficult to predict. Published research has already

demonstrated the use of a PSMA RNA aptamer targeting prostate cancer cells covalently linked to a siRNA targeting eukaryotic elongation factor 2 mRNA, creating an aptamer-siRNA hybrid that can effectively induce apoptosis in the target cell type (26). Other non-covalent methods of attaching molecules to an aptamer have also been developed; these would simplify the creation of aptamer-drug complexes by avoiding additional chemistry. This approach has been demonstrated recently by the physical conjugation of the chemotherapeutic doxorubicin to the A10 RNA aptamer targeting prostate cancer cells (27). Instead of covalently linking the drug to the aptamer, the planar structure of doxorubicin allows it to be intercalated into the stem structure present in the A10 aptamer. This physical drug-aptamer conjugate was shown to bind prostate cancer cells effectively and deliver the chemotherapeutic agent. Delivery of nucleic acid therapeutics is also a growing area of interest with the mainstream popularity of siRNA technologies and the continued development of DNA based therapeutics for mRNA cleavage and signalling. Watson-Crick base pairing could allow the delivery of a therapeutic nucleic acid on one strand by base pairing to a targeting aptamer strand. Using circular aptamer and therapeutic nucleic acid molecules in combination allows for the creation of topologically linked ring complexes (28). Another method used to enhance the effectiveness of aptamer targeting involves complexing several different aptamers for the same target, termed multivalency (29). A similar approach has previously been used in the siRNA field to address off target effects (6). By combining several aptamers with a common target but unique off-target profiles, a more effective targeting complex can be created. Cell targeting aptamers may also be adapted to enhance existing drug delivery platforms. Liposomes are another commonly used method to deliver molecules into cells. A group has recently published a communication describing the use of a polyethylene

glycol-sgc8 aptamer embedded in the membrane of a fluorescent dye loaded liposome to target T-cell acute lymphoblastic leukemia cells (18). They reported effective delivery of the fluorescent dye to the target cells with minimal leakage, their design could be adapted to target other cell types by simply replacing the aptamer displayed on the surface. Similarly, another group has reported the use of nanoparticles of the chemotherapeutic docetaxel in complex with polyethylene glycol and a surface modified with A10 aptamer targeting prostate cancer cells. An interesting result of their study was the observation that not only did the aptamer functionalized nanoparticles localize with high specificity to prostate cancer cells, but also showed decreased absorption into non-target cell types (30). This observation has previously been observed with small molecule-aptamer conjugates as well (6). The positive reports of aptamer targeted drug delivery complexes published thus far suggest that the versatility of the aptamer has yet to be fully explored. Future cell selection experiments looking specifically for internalization can yield better aptamers, minimizing the need for additional modifications.

1.7 *Research Objective*

Given the need for an aptamer capable of facilitating its own cellular internalization with the potential to deliver a variety of payloads into cells, a cell selection experiment will be conducted. With respect to chemical stability *in vivo*, a circular DNA structure will be used which affords a degree of nuclease resistance and opportunities for both covalent and non-covalent modification. The selection experiment will be targeted against the MCF7 breast adenocarcinoma cell line. Aptamers isolated in this selection could potentially have specificities ranging from broad multi-cell type targeting sequences to highly specific sequences targeting motifs unique to MCF7. Low specificity aptamers

would be ideal for use as a general delivery tool in research, where as cell type specificity generated by counter selection steps could be useful in targeted drug delivery or as a probe. Characterization of aptamers isolated from selection may include concentration dependency, time courses, sequence minimization and targeting characteristics. Ultimately, aptamer sequences would be tested for their ability to carry a drug or nucleic acid into cells.

Chapter 2: Materials and Methods

2.1 *Library Preparation*

The CDL2 single stranded DNA library used for selection consisted of a 40 nt randomized region flanked by 18 nt and 19 nt fixed primer binding domains for a total sequence length of 77 nt. All unmodified DNA oligonucleotides were synthesized using phosphoramidite chemistry by Integrated DNA Technologies and provided desalted and lyophilized (Coralville, Iowa). CDL2 library was further purified using poly acrylamide gel electrophoresis (PAGE) on a 10% denaturing gel (8 M urea) and bands were visualized by UV shadow. Bands of desired length were excised and eluted in elution buffer overnight at room temperature. Ethanol precipitation was done by adding 2.5 volumes of chilled 100% ethanol to 1 volume of supernatant followed by centrifugation at 20,000 G for 15 minutes at 4°C. Supernatant was then decanted and the pellet was dried on the bench at room temperature. A working stock of CDL2 library was obtained by resuspending the purified pellet in Milli-Q ddH_2O and quantification by UV spectroscopy (Genesys 10UV, Thermo Scientific). Concentration was calculated using absorbance values and the base composition of the CDL2 sequence as determined by the Oligo Calc tool (31).

The CDL2 library was prepared for circularization by phosphorylation of 1660 pmol of linear CDL2 stock using T4 polynucleotide kinase (PNK) by manufacturer recommended protocol and extending incubation time to 1 hour at 37°C (EK00312, Fermentas, Burlington, Ontario). The phosphorylation reaction was then directly used for ligation by T4 DNA ligase and addition of 1660 pmol of CDL2 linker sequence to facilitate

circularization (EL0011, Fermentas, Burlington, Ontario). The reaction was diluted to yield a CDL2 concentration of 500 nM to favour intramolecular ligations and heated at 90°C for 1 minute and cooled at room temperature for 10 minutes prior to addition of T4 DNA ligase and incubation at 16°C overnight. Ligation reactions were ethanol precipitated and PAGE purified by the previously mentioned protocol. Circular bands were identified by UV shadow as the closest band with a decreased mobility versus a linear marker. Circular stock was purified and quantitated by the previously mentioned protocol. Circularity of the library was confirmed using EcoRI digestion (New England Biolabs, Ipswich, MA) and ExoI nuclease resistance assays (Fermentas, Burlington, Ontario). Refer to Appendix I Table A1 for sequences used in selection.

2.2 *Cell Lines*

MCF7 cells (human breast adenocarcinoma cell line) were used for cell selection as well as for qPCR and *in situ* RCA assays. The cell line was provided by the David Andrews lab, McMaster University. For all assays, cells were cultured to 80% confluence using α-MEM (MDCL prepared media, McMaster University) supplemented with 10% FBS (Gibco Invitrogen, Burlington, Ontario) on tissue culture treated dishes. Cells were cultured at 37°C and 5% CO^2 in a humidified incubator.

2.3 *Cell Selection*

To begin the initial round of selection, 100 pmol of circularized CDL2 library added to 1 mL Hank's buffered salt solution (HBSS) (14025-092, Gibco Invitrogen, Burlington, Ontario) was heated to 90°C and cooled to room temperature. This selection mixture was then supplemented with 1 mg/mL BSA and 0.1 mg/mL tRNA. MCF7 cells grown to 80%

confluence were prepared for selection by aspirating growth media and washing once with HBSS. An average of $3x10^5$ cells were used for each round of selection. Selection mixture was added to cells and incubated at 37°C and 5% CO^2 in a humidified incubator for 2 hours. After incubation the selection mixture was removed and cultures were washed twice with HBSS and once with 1x PBS (0.02 M inorganic phosphate) (MDCL prepared media, McMaster University). Cells were then incubated with 2X trypsin-EDTA (Gibco Invitrogen, Burlington, Ontario) for 20 minutes before being resuspended in HBSS. Cell suspensions were then pelleted by centrifugation for 5 minutes at 4000 G and the supernatant removed. Cell pellets were gently resuspended in 1 mL PBS and repelleted twice to remove unbound sequences. Final pellet was resuspended in 100 µL PBS. Refer to Figure 1 for selection scheme.

An incubation time of 2 hours was used for rounds 1 to 8 of the CDL2 G18d line of selection and was reduced to 1 hour for rounds 9 to 18. Rounds subsequent to the initial round used 20 pmol of regenerated library for selection. From round 13 to 18, trypsinized cells resuspended in HBSS were treated with Proteoblock (Fermentas, Burlington, Ontario) to deactivate trypsin before addition of 250 units of micrococcal nuclease (EN0181, Fermentas, Burlington, Ontario) and incubation at 37°C for 1 hour to digest unbound sequences. An additional treatment with PNGase F (New England Biolabs, Ipswich, MA) was added to the selection method from rounds 15 to 18, done concurrently with nuclease treatment.

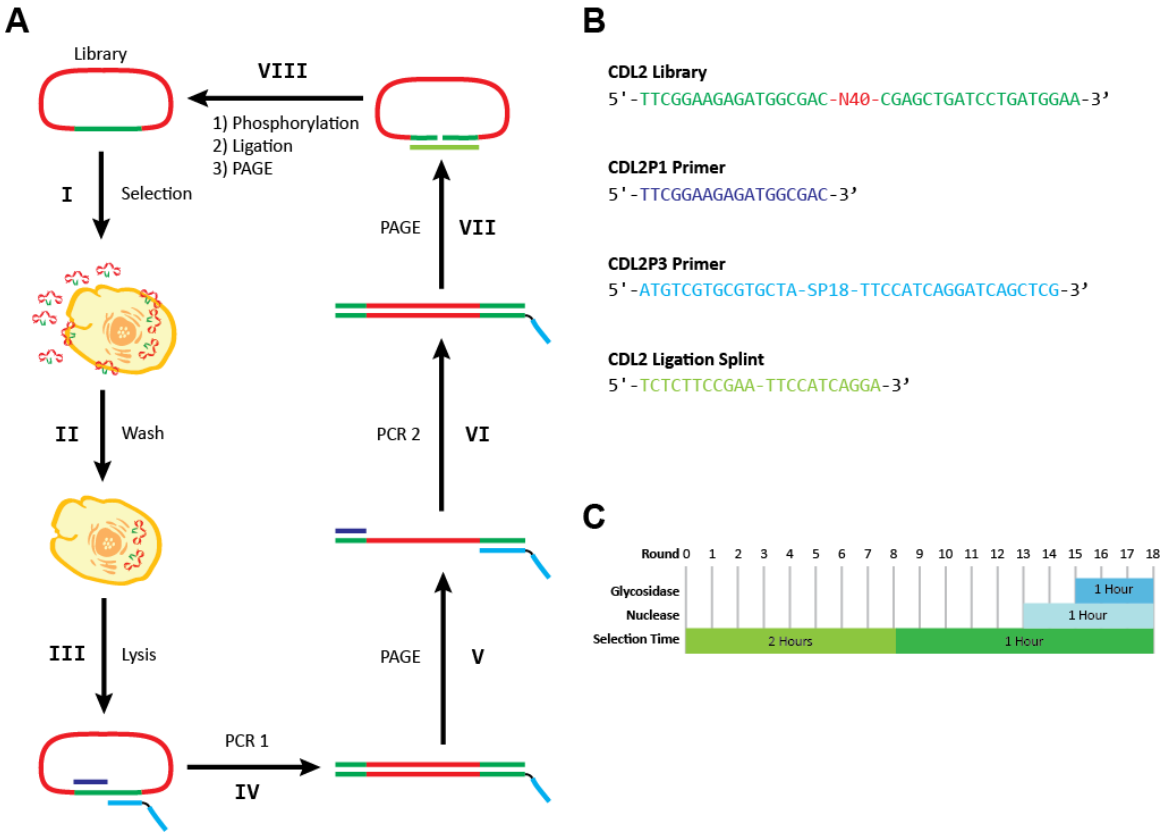

Figure 1 – *[A] CDL2 selection scheme. In the initial round of selection, (I) 100 pmol of circularized CDL2 library is mixed with 1 mg/mL BSA and 0.1 mg/mL tRNA and applied to 3x10^5 MCF7 cells at 80% confluence. In subsequent rounds, 20 pmol of library is used. (II) After incubation at 37°C and 5% CO_2 for 2 hour, cells are washed twice with HBSS and once with PBS. Cells are then trypsinized and resuspended in HBSS. From round 9 onward, incubation time is reduced to 1 hour. (III) Cell suspensions are lysed using TRIzol reagent and short DNAs purified. Micrococcal nuclease digestion is added from round 13 onwards and PNGase F digestion from round 15 onwards. (IV) Primary PCR incorporating a [α-^{32}P]dGTP radiolabel is done on cell lysates using CDL2P1 and CDL2P3 primers. (V) The polymerase blocked tail of CDL2P3 primer is exploited for sense strand separation on PAGE. (VI) Secondary PCR using purified primary PCR sense product is done using CDL2P1 and CDL2P3 primers, reactions are not radiolabeled. (VII) Sense strand is visualized by UV shadow and purified. (VIII) Secondary PCR sense strand is phosphorylated and ligated by annealing CDL2 Ligation Splint, followed by PAGE purification to regenerate the library.[B] CDL2 library sequence and sequences of PCR primers and ligation splints used in selection. [C] Overview of G18d selection experiment showing changes to the selection method introduced at various rounds. Nuclease refers to micrococcal nuclease and glycosidase refers to PNGase F.*

2.4 *Cell Lysate Preparation*

MCF7 cell pellets were lysed using TRIzol LS reagent (Invitrogen, Burlington, Ontario) according to manufacturer protocol for DNA isolation. Recovered DNA from the lysis procedure was resuspended in 50 µL ddH_2O. To reduce genomic DNA contamination the recovered sample was filtered through an Amicon Ultra-0.5 centrifugal filter unit with a 100 kDa membrane filter (Millipore, Billerica, MA).

2.5 *Cell Selection Two-Step PCR Amplification and Library Recircularization*

Primary PCR amplification of recovered CDL2 sequences from purified cell lysates were done using 38 µL of purified cell lysate and 1 µM each of CDL2P1 and CDL2P3 primers. Antisense primer CDL2P3 (Integrated DNA Technologies) contains an 18 atom hexa-ethyleneglycol spacer linking the complimentary primer region to a 15 nt random sequence to facilitate strand separation of the PCR products by PAGE. Primary PCR reactions also contained 5 µCi [α-^{32}P]dGTP (Perkin Elmer, Waltham, MA) to radiolabel PCR products. Other reaction components consisted of 200 µM dNTP, buffer and Biotools DNA polymerase (Tth DNA polymerase, Biotools, Madrid, Spain). The 20 cycle PCR program consisted of a 95°C 30 second denaturation step, a 50°C 45 second annealing step and a 72°C 10 second extension step run on a RoboCycler Gradient 96 (Stratagene). Reactions were purified by denaturing PAGE and visualized on storage phosphor screens (Molecular Dynamics) scanned on a Molecular Dynamics Typhoon 9200 imager. Band quantification was done using Molecular Dynamics Image Quant version 5.2 software. Bands corresponding to sense CDL2 sequences were excised and DNA eluted into 100 µL water for use in secondary PCR.

Secondary PCR was done using 2 µL of purified primary PCR product as template. Additional reaction components consisted of 200 µM dNTP, 1 µM each of CDL2P1 and CDL2P3 primers, buffer and VENT DNA polymerase (New England Biolabs, Ipswich, MA). The PCR program and purification methods used for secondary PCR were the same as used for primary PCR with the exception of 15 cycles of PCR amplification and band visualization by UV shadow in place of storage phosphor.

Pooled secondary PCR products were phosphorylated and circularized by the method previously described for generation of the initial library.

2.6 *Cloning and Sequencing*

The final population of the selection experiment was prepared for cloning by the secondary PCR protocol where CDL2P3 primer was replaced with an un-tailed variant, CDL2P3-notail, resulting in double stranded DNA suitable for use in a TA-cloning kit. Cloning was done using an InsTAclone PCR cloning kit (K1213, Fermentas, Burlington, Ontario) according to manufacturer instructions. Ligation products were transformed into *E.coli* DH5α by electroporation and plated on LB agar supplemented with 100 µg/mL ampicillin (Sigma-Aldrich, St.Louis, MO).

A small number of colonies were picked and inoculated into LB media supplemented with 50 µg/mL ampicillin and incubated overnight at 37°C with shaking. Cultures were then miniprepped using QIAprep Spin Miniprep Kit (Qiagen, Valencia, CA) according to manufacturer protocol. Purified plasmid samples were sent for small scale sequencing (MOBIX Lab, McMaster University) using M13 forward sequencing primer. Five G18d sequences and 5 initial library clones were sequenced.

An additional 96 clones were sent for sequencing by Functional Biosciences (Madison, WI).

2.7 *Ratiometric Recovery Assay and Quantitative PCR Analysis*

MCF7 cells at 80% confluence were simultaneously incubated with CDL2-G18d selected sequences and non-selected control sequences for 1 hour at 37°C, 5% CO^2 in a humidified incubator at a CDL2-G18d:Control sequence ratio of 0.25:1. Cells were then processed as described previously in the cell selection method.

Quantitative PCR (qPCR) assays were setup using 10 µL lysate per 25 µL PCR reaction in addition to 1 µM of sequence specific primer pairs, 200 µM dNTP, Eva Green fluorescent dye (Biotium, Hayward, CA) and Biotools DNA polymerase. Each lysate sample was tested using a CDL2-G18d sequence specific primer pair (Appendix I Table A2) as well as with a control sequence primer pair for comparison. PCR reactions were run on an Eppendorf Mastercycler EP Realplex real time PCR machine cycling 95°C for 30 seconds, 53°C for 45 seconds and 72°C for 10 seconds. Fluorescence data was analyzed using Realplex software to obtain threshold cycle (Ct) results. Standard curves and controls were created using cell lysate collected from untreated cells.

2.8 *In Situ Rolling Circle Amplification*

MCF7 cells grown to 80% confluence on glass cover slips were washed once with HBSS before addition of 10 nM to 200 nM CDL2-G18d selected sequence solutions in HBSS. Cells were incubated with DNA solutions at 37°C and 5% CO^2 in a humidified incubator for 1 hour. After incubation DNA solutions were removed and slides were washed twice with HBSS and once with PBS. Slides were then fixed with 4% formaldehyde (Bioshop,

Burlington, Ontario) for 10 minutes at room temperature, and then washed twice with PBS. Cells were permeabilized by treatment with 0.1% Triton X-100 for 15 minutes at room temperature followed by two washes with PBS. Rolling circle amplification (RCA) primer mixture consisting of 100 nM CDL2P3-untailed antisense primer, phi29 DNA polymerase buffer (EP0091, Fermentas, Burlington, Ontario) in ddH_2O was applied to the slide and heated at 90°C for 30 seconds and cooled at room temperature. Once cooled, RCA enzyme solution consisting of 1 mM dNTP, phi29 DNA polymerase buffer and phi29 DNA polymerase in ddH_2O was added to the cover slips. RCA reactions were incubated at 30°C for 1.5 hours after which cover slips were washed twice with PBS. Fluorescein labeled antisense probes RA-Fluor and LA-Fluor (Keck Oligonucleotide Synthesis facility, Yale University) were added in equimolar amounts at a combined concentration of 500 nM to the RCA treated slides (Appendix I Table A3). Samples were then heated at 90°C for 1 minute and cooled to room temperature to allow probe binding. Slides were then washed once with PBS to remove unbound probe and counter-stained for 5 minutes at room temperature with 40 µM DRAQ5 DNA dye (Biostatus Limited, Shepshed, Leicestershire, UK) followed by a final PBS wash. Slides were then mounted with ProLong Gold antifade reagent (Invitrogen, Burlington, Ontario).

2.9 *Fluorescence Microscopy*

Images of RCA treated cells were acquired on a Leica TCS SP5 fluorescence imaging microscope (Biophotonics Facility, McMaster University) using a 63x glycerol immersion objective. Samples were excited with 488 nm and 633 nm lasers for simultaneous fluorescein and DRAQ5 visualization. Fields of view were selected based on positive DRAQ5 staining and z-stacks were collected ranging from the glass surface to

the top of cells. Generally, 10 to 20 fields of view were collected, randomly spaced along a line bisecting the surface of the cover slip.

2.10 *Fluorescence Image Processing*

Z-stacks were processed using ImageJ (32). Fluorescein and DRAQ5 stacks for each sample were analyzed using custom written macros (Refer to Appendix II for macros) to identify fluorescent spots and nuclei respectively. To prepare the stacks for spot counting, a maximum projection image is first generated followed by rolling-ball background correction, Gaussian blur filter, manual thresholding, hole filling and watershed processing for DRAQ5 images. Particles and nuclei are then counted using the particle analysis function. Thresholding values are manually set for each sample to ensure proper segmentation and account for imaging variances between experiments.

Chapter 3: Results

3.1 *Cell Selection*

In recent years the method of *in vitro* selection for the isolation of aptamers has been expanded beyond the typical targets of small molecules and biomolecules to include whole cells. Studies have applied cell selection towards the isolation of cell surface binding RNA aptamers to serve as cell-type specific probes (16). From these studies it was found that some cell surface binding aptamers also have the ability to internalize within the cell (33, 34). A cell selection experiment designed with the goal of isolating cell internalizing aptamers to the exclusion of simple surface binding sequences would have the potential to isolate more effective aptamers than have previously been described. With respect to the goal of internalization, the intracellular stability of an aptamer may be improved by using DNA in place of less stable RNA and a covalently closed single stranded circular structure that is resistant to exonuclease digestion. A circular structure also presents unique opportunities for non-covalent attachment of payload molecules through linked ring structures as previously demonstrated by this lab (28).

Initial attempts at cell selection using the CDL2 library (G1a, G1b and G1c selection lines were aborted) were done using an SDS lysis protocol that resulted in poor downstream PCR. For selection lines G18d and G11f (not sequenced) an improved method using TRIzol reagent was used. Initial lines of selection required up to 25 cycles of primary PCR to generate a detectable signal. This contributed to significant background amplification in negative control reactions. The improved DNA recovery and lysate purity obtained by using the TRIzol purification method allowed primary PCR reactions

to be reduced to 20 cycles with minimal amplification in negative controls. In contrast to traditional *in vitro* selection where enrichment in functional sequences can be visualized by measure of a cleavage product or increase in recovery, no trends were observed in the amount of primary PCR product recovered during selection as measured by radiolabel incorporation in PCR.

To increase stringency the selection incubation time was decreased to one hour at round 8, however no changes in amplification were observed in primary PCR results. To further increase stringency and reduce unbound and surface bound sequences, treatments with micrococcal nuclease and PNGase F were introduced in rounds 13 and 15 respectively (Figure 1 Panel C). When tested against the initial library, a 32% decrease in primary PCR band intensity was observed versus untreated cells (Figure 2).

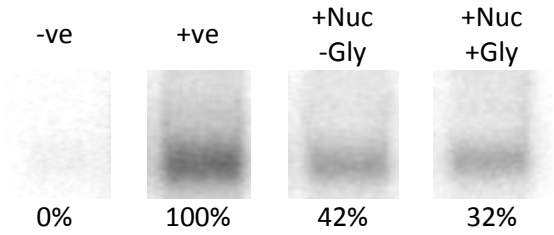

Figure 2 – *Primary PCR reactions of MCF7 cell lysates incubated with CDL2 library with and without micrococcal nuclease and PNGase F treatments. MCF7 cells at 80% confluence were incubated with 20 pmol initial CDL2 library for 1 hour at 37°C and washed and processed as described in the selection method. The negative PCR control (-ve) contained no template and the positive control (+ve) was untreated. Incubation of cell suspensions with glycosidase and to a greater extent nuclease resulted in a significant decrease in band intensity.*

3.2 *G18d Classes*

A total of 101 G18d selected clones were sent for sequencing resulting in 81 complete sequences used in class analysis. An additional 5 clones containing unselected DNAs from the initial CDL2 library were also sequenced for use as controls. CDL2 Control 1 (con1) and Control 2 (con2) sequences were derived from this pool of clones. The 81 sequences were sorted into 12 classes with sizes ranging from 2 members to 16 members (Table 1). Fifteen of 81 G18d sequences were not classified as only a single member was identified. Class 3, represented by sequence G18d-70, and class 4, represented by G18d-5, composed the largest classes representing 20% and 16% of all sequences respectively. Class 6, represented by G18d-14, was the only class to contain a deletion. None of the classes showed any stable secondary structures as determined by mfold (35). Classes 3, 4, 6 and 1 were selected for further study covering the most populous classes, a moderately populated class and the deletion class. While the goal of selection was to generate aptamers, the G18d sequences described herein have yet to be confirmed to function via a binding mechanism consistent with aptamers, therefore should be regarded as functional DNA molecules.

Class	Length (nt) 10 20 30 40 50 60 70 5`-....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|..-3`	Members
-	Library TTCGGAAGAGATGGCGAC----------N40----------CGAGCTGATCCTGATGGAA	
1	G18d-22 .....ACACATGCCTATTGATCTCTGTAACCCCGAATCTCCTGCC.....	5
2	G18d-2 .....ACGCTAAGGGGAGCAACGATTGTGCAACGGCACCGCCCTC.....	3
3	G18d-70 .....CACTCCCTCTGCGTGCGAATTGAGCCTATGACGCATTTTC.....	16
4	G18d-5 .....CCGTTGGTGGTGACGTGAACATTGTTCCTATGACGTTCCT.....	13
5	G18d-9 .....CCCCACCGGAATTCGACGCCTATTGATCTCTAGAACCTCC.....	12
6	G18d-14 .....ACATTCCGCTATGAATGACATCCGACCCCCCGGTCCTGA-.....	3
7	G18d-26 .....CTATGAAACACGTCGTACCGCCTATTGATCTCTAGTCCGA.....	2
8	G18d-46 .....CCGTTGGTGGTGACGTGAACACTGTTCCTATGACGTTCCT.....	2
9	G18d-52 .....TATGCTACCCCAAGTCAACCGCCTAACGATCTCTTTAGCC.....	2
10	G18d-55 .....CCCCGGGTAACTTGGTGCAACTCTGTGCTATTGACCAATA.....	2
11	G18d-69 .....CACTGGGGGAGTGGCCCTGCCCGGTCGATGTCAATACACT.....	3
12	G18d-72 .....CCCCACCGGAATTCGACGCCTATTGATCTCTAGAACCCCC.....	3
-	Con1 .....TAGTCGCGCGTTTTACTTTTCTGTGCGAGAGCCGTACTGT.....	-
-	Con2 .....GAATAGAATGCGGGTGGATGGTCGAAGTTGGGGGCAGCGC.....	-

Table 1 – *Summary of classes identified in CDL2 Library G18d sequencing data and control sequences. Sequences were grouped into classes when 2 or more identical sequences were present. Of the 81 G18d sequences, 66 were grouped into classes CDL2 library sequence at top indicates location of fixed primer binding domains relative to randomized domain. Fixed primer domains are omitted for clarity in class sequences. Colours represent unique bases A, T, C and G. A deletion mutation is observed in Class 6. Two of the largest classes (Class 3 and Class 4) and two smaller classes (Class 1 and Class 6) were selected for characterization. Control sequences con1 and con2 were selected from a pool of 5 sequences of initial library.*

3.3 *qPCR Analysis*

A recovery experiment was done to determine whether a mixture of a G18d sequence and a control sequence could be recovered from cells at a ratio greater than present in the selection medium. Cells incubated with a 0.25:1 ratio of G18d:control circular DNAs were lysed and the relative recovered quantities of each species were determined by qPCR. Of the four sequences studied, all sequences showed recovery ratios suggestive of selectivity (Figure 3). Standard deviations were generally large, however effects were still

well above control levels. The non-selective control2/control1 combination resulted in a ratio of 0.24 (SD=0.12) as would be expected of non-selective sequences. Given the initial bias of 0.25:1 G18d:control in the selection medium, an overall 52-times recovery enhancement is observed with G18d-22 versus control 1. The lowest recovery enhancement was observed with G18d-5 at 12.7-times versus control 1.

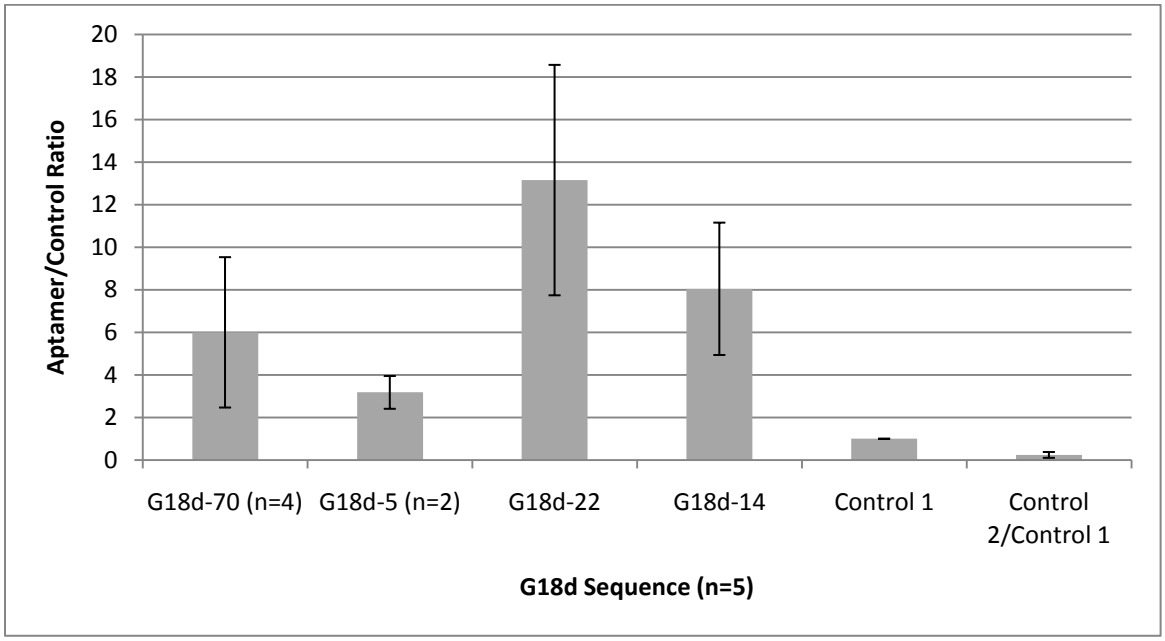

Figure 3 – *Summarized qPCR G18d recovery assay. MCF7 cells incubated with a 0.25:1 molar ratio of a G18d selected sequence to a control sequence for 1 hour at 37°C in HBSS. Cells were washed twice with HBSS and once with PBS before trypsinization and TRIzol lysis. Cell lysates were used for qPCR to determine the ratio of G18d selected sequences versus spiked controls. G18d-70, 5, 22 and 14 were derived from selection and show selectivity against a control sequence. A control 2 sequence tested against control 1, as with G18d selected sequences, showed a ratio of 0.24 (SD=0.14) consistent with expectations of non-selectivity.*

3.4 *In situ* RCA Imaging

An alternative method of measuring the ability of the selected sequences for internalization takes advantage of the circular nature of G18d sequences and rolling circular amplification. By incubating MCF7 cells with selected G18d sequences it was possible to detect internalization by binding an antisense primer to internalized sequences and using the RCA method to generate long antisense single stranded DNAs. These RCA products were then hybridized to sense fragments modified with fluorescein. This technique allows for signal amplification and *in situ* visualization of G18d sequences by fluorescence microscopy without having to modify the sequences themselves. Direct labeling of the G18d sequences was attempted however fluorescence intensity was not sufficient for imaging. Optimization of this method for short circular DNAs showed that fixation in 4% formaldehyde is superior to methanol fixation in terms of sensitivity and background reduction, and that staining with 500 nM fluorescent probe generates a strong signal with minimal background.

In situ RCA experiments were done on G18d-70, 5, 22 and 14 along with control 1 and control 2 sequences. However, the major focus for optimization and study was the G18d-14 sequence. When averaging particle and nuclei counts from available experimental replicates, the G18d-14 sequence showed 11-times greater uptake at 100 nM versus control 1 at 100 nM (Figure 4). A second control sequence was also tested and showed comparable results to control 1. The uptake of G18d-14 also increased with concentration up to 200 nM.

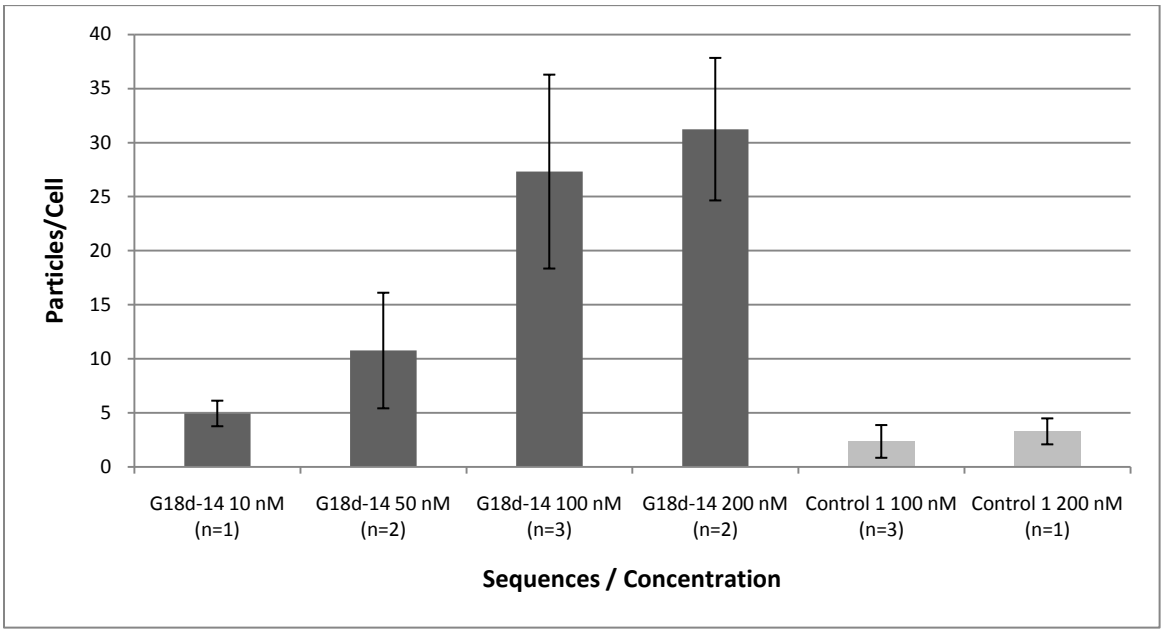

Figure 4 – *Summary of in situ RCA data for G18d-14 and control. MCF7 cells at 80% confluence were incubated with varying concentrations of G18d-14 circular DNA for 1 hour at 37°C, followed by washing and in situ RCA. Cells are labeled with fluorescein modified DNA complementary to the RCA product and counter stained with DRAQ5 to facilitate nuclei counting. Each in situ RCA experiment imaged 250 cells on average and slices at and below the glass surface were discarded. Particle and nuclei counts were pooled from all available experimental replicates for calculation of particle/cell value and standard deviation. 100 nM and 200 nM concentrations of G18d-14 resulted in significantly greater particle count versus control.*

In a qualitative inspection of the fluorescence microscopy data the increased RCA particles generated in the G18d-14 sample versus control DNA or no DNA samples can be easily observed (Figure 5). G18d-14 RCA products appear densest in cell nuclei with few particles observed in cytoplasm. Particles are generally compact and roughly spherical, however resolution is limited by the 488 nm excitation wavelength. Z-stacks also show that particles are distributed throughout the volume of the nuclei. The particles observed in the control 1 experiment show no selectivity for cell nuclei with equal numbers of particles inside and outside. For comparison, a no DNA experiment was performed which showed no fluorescent particles generated without the presence of

circular DNA. Some diffuse staining of the cytoplasm by the fluorescent DNA probe is also observed, however this is effectively removed by rolling-ball background subtraction during image processing.

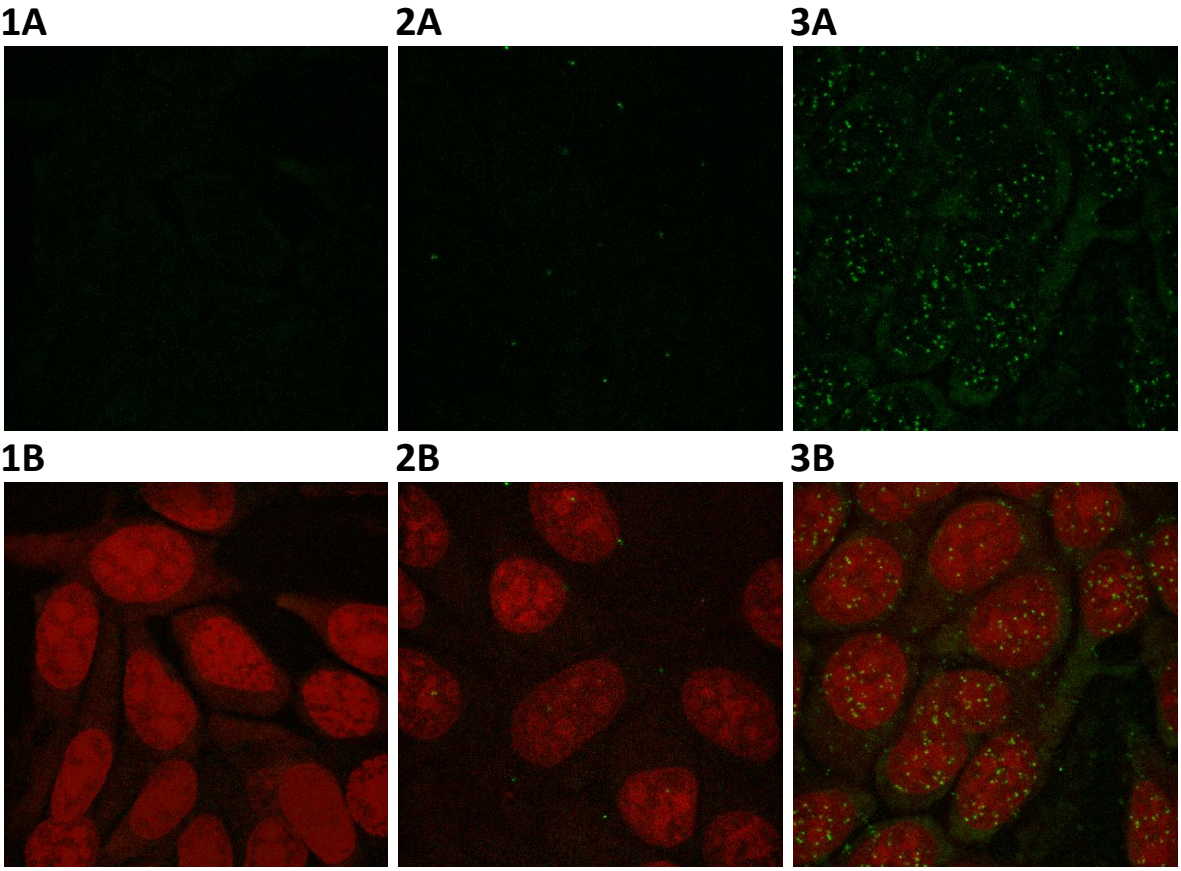

Figure 5 – *Comparison of control in situ RCA to G18d-14. In situ RCA reactions on MCF7 cells using 100 nM DNA concentration. Slices at and below glass surface were removed. Maximum projection images were generated from z-stacks and background subtracted. Panels sets 2 and 3 show control 1 and G18d-14 DNA in situ RCA experiments respectively. Panel set 1 shows a control with no DNA added during incubation. Panels 1A, 2A and 3A are fluorescein emission images indicating the location of RCA products, while panels 1B, 2B and 3B are corresponding composite images of fluorescein and DRAQ5 channels. G18d-14 shows a significant number of RCA particles versus control 1.*

Classes 1, 3 and 4 represented by sequences G18d-22, G18d-70 and G18d-5 were also studied qualitatively. As observed with G18d-14, all sequences appear to primarily target to the nuclei of cells (Figure 6). G18d-70 and G18d-5 sequences generate a greater

number of fluorescent particles as compared to both G18d-5 and G18d-14 at the same concentration. G18d-5 also shows more particles in cytoplasm as compared to other G18d classes. While in most experiments the particles appear dense and spherical, it has been observed in some replicates that particles are diffuse, making quantification difficult, G18d-70 in particular.

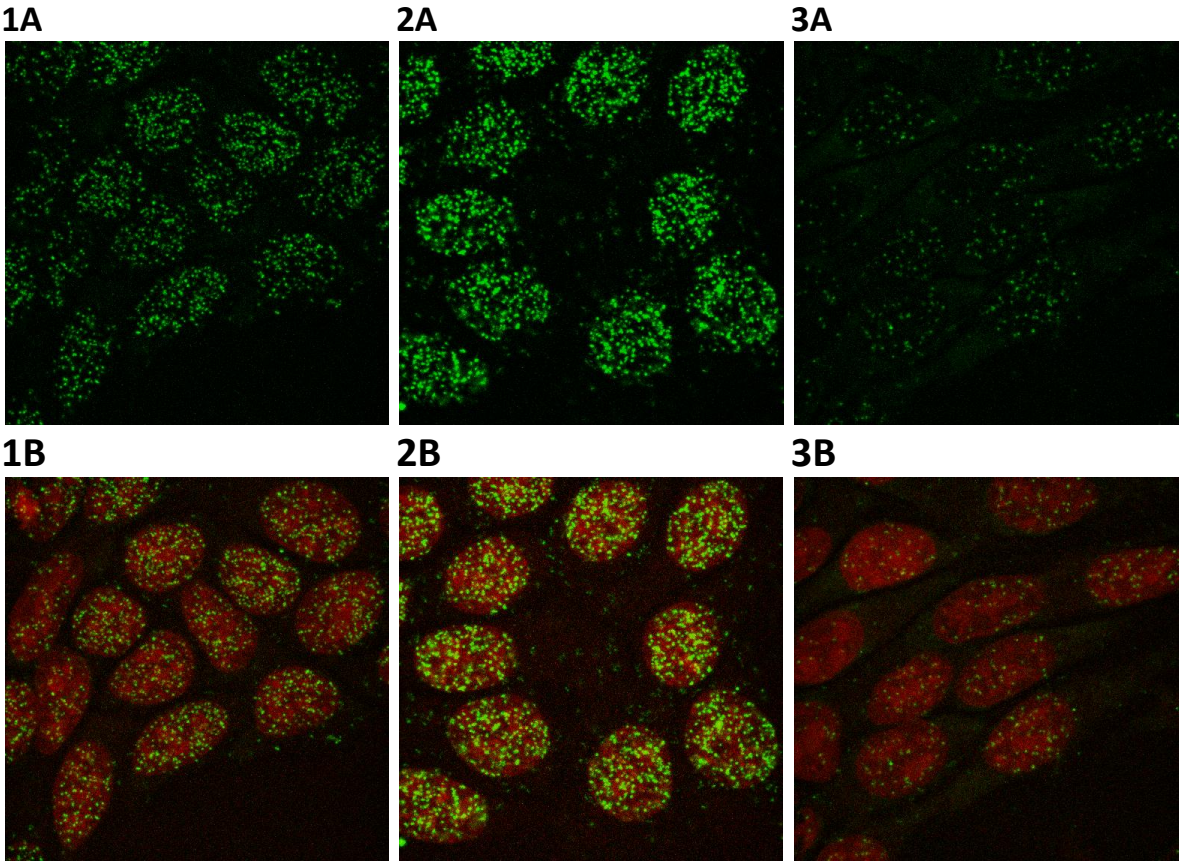

Figure 6 – *Examples of in situ RCA fluorescence images for G18d-70, 5 and 22. In situ RCA reactions on MCF7 cells using 100 nM DNA concentration. Slices at and below glass surface were removed. Maximum projection images were generated from z-stacks and background subtracted. Panels in the top row show fluorescein emission and panels on the bottom row show a composite of fluorescein and DRAQ5 emission channels. Panel set 1 shows G18d-70, panel set 2 shows G18d-5 and panel set 3 show G18d-22 treated samples. All G18d sequences appear to generate RCA particles in the nuclei with a small number of faint cytoplasmic particles as seen in G18d-5. G18d-22 shows fewer RCA particles versus other G18d sequences.*

In all experiments it was observed that a significant number of particles were bound at the level of the glass surface. These particles appear both in areas with and without cell growth, however they are not observed in control reactions incubated without DNA. In general, glass surface binding was observed to contribute more to control reactions which typically have lower overall particle counts as illustrated in Figure 7, panels 1A and 1B. Glass surface binding also has a significant effect on G18d samples, but contributes less to particle count as a proportion of total count than is seen in control DNA experiments. The number of glass surface level particles also appears to relate to the overall number of particles observed in a field of view. The greater the overall number of particles in the image, the greater the number of particles at glass surface level. Particles at glass level have also been observed to have a more spherical, dense morphology in areas without cell growth versus a diffuse fiber-like morphology in areas overlapped by cells. Thus, glass surface level slices have been excluded from particle count analysis to improve the accuracy of scripted particle counting and simplify interpretation of the data.

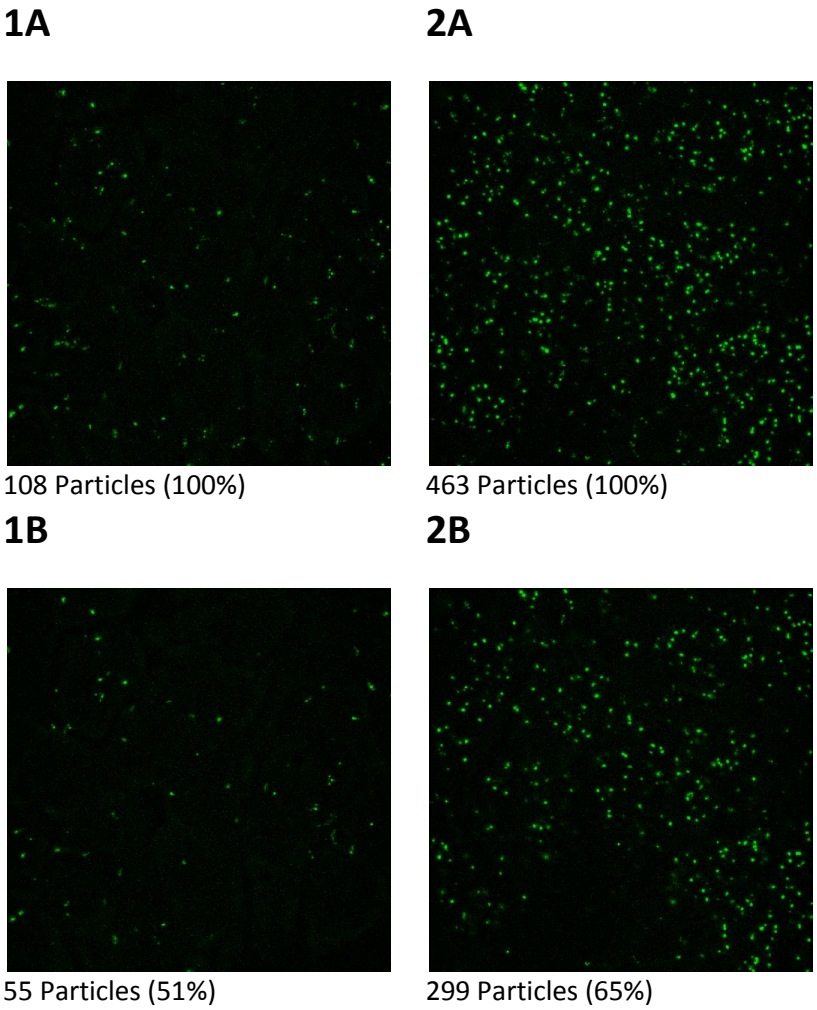

Figure 7 – *Comparison of the effect of slice removal at and below glass surface level on particle counts. In situ RCA reactions on MCF7 cells using 100 nM DNA concentration. Maximum projection images were generated from z-stacks and background subtracted. Particle counts are generated following the method previously described in this paper. The same thresholding values were used for all images in this figure. Panel set 1 shows control 1 treated cells before (1A) and after (1B) slice removal resulting in a 49% decrease in particle count. Similarly, panel set 2 shows G18d-14 treated cells before (2A) and after (2B) slice removal resulting in a 35% decrease.*

While the distribution of RCA particles within the cell appears to concentrate in the nucleus, at an intercellular level RCA density is dependent on local cell density in G18d-70 treated cell cultures (Figure 8). Cells with few close neighbours showed higher amounts of RCA particles versus cells in the centre of clusters. This cell density dependency has been observed in other G18d treated samples as well however to a lesser extent.

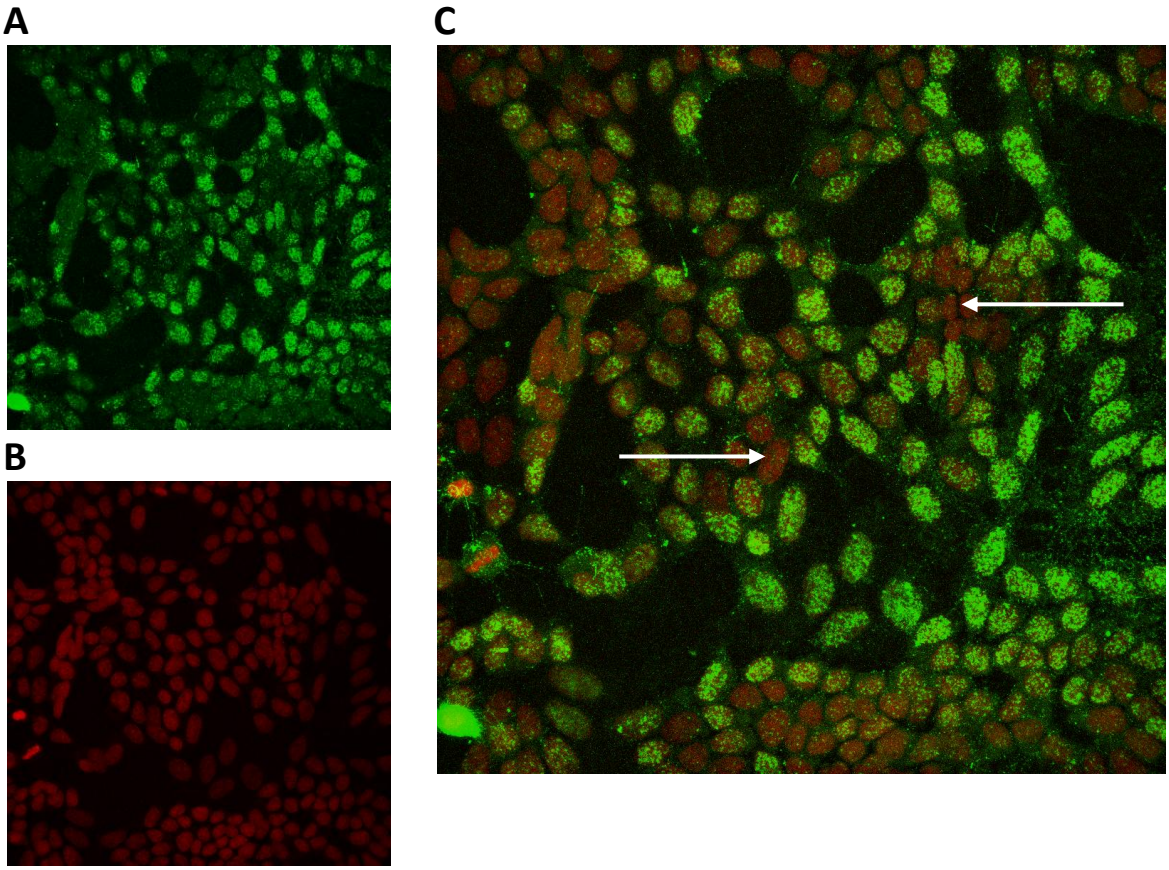

Figure 8 – *Spatial distribution of fluorescent RCA particles in G18d-70 treated MCF7 cell culture. In situ RCA reactions on MCF7 cells using 100 nM DNA concentration. Maximum projection images were generated from z-stacks and background subtracted. Cells were imaged using a 20x objective. Slices at and below glass level have been removed. Panels A and B show the fluorescein emission and DRAQ5 emission channels respectively. Panel C is a composite of both channels to illustrate fluorescent particle density relative to local cell density. Arrows indicate centres of cell clusters where fluorescent particle density is low versus cells at the perimeter of the cluster.*

Fluorescent particles observed at glass level in cells treated with G18d-70 were also observed to have a different appearance as compared to other G18d sequences. In areas without cell growth, typical spherical small fluorescent particles are observed, however particles at the perimeters of cell clusters show an elongated fiber-like appearance (Figure 9 Panel A). The fibers run perpendicular to the line of cell growth beneath the level of the nucleus. Several fibers are also observed at mid-cell level extending out from cells however no particles are observed penetrating nuclei at any depth (Figure 9 Panel B). Fibers identified at mid-cell level can be tracked extending to the tops of cells where the longest fibers are observed to interconnect. Nodes of brighter fluorescence are also seen along the length of fibers at the tops of cells (Figure 9 Panel C).

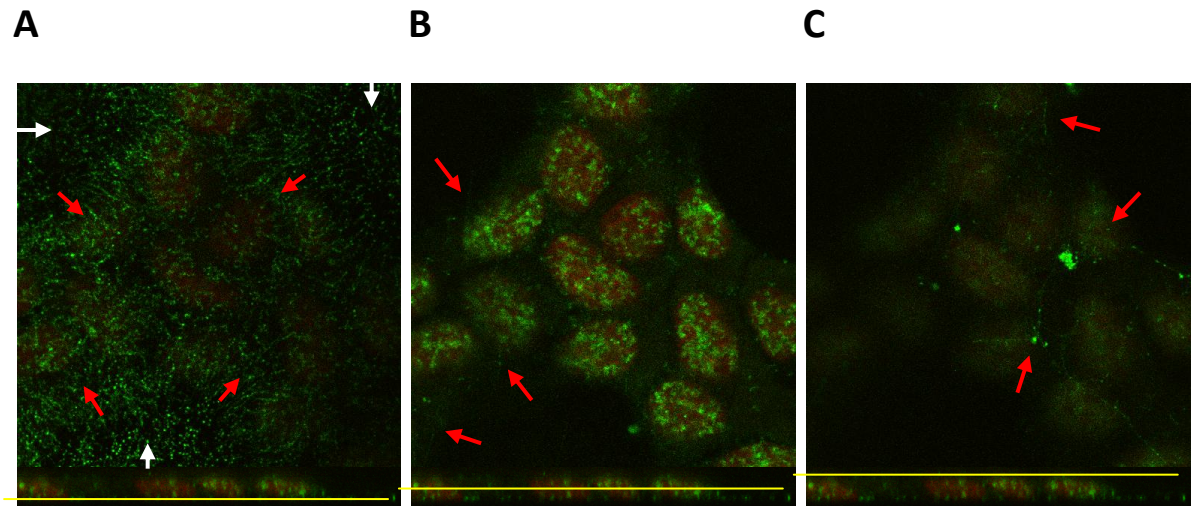

Figure 9 – *Examples of fluorescent fibers observed in G18d-70 treated cells. In situ RCA reactions on MCF7 cells treated with 100 nM G18d-70 DNA. Panel A shows a glass level image, spherical RCA particles are indicated by white arrows found in areas without cell growth. Fluorescent fibers perpendicular to the line of cell growth are observed at the perimeter of the central cell cluster as indicated by red arrows. Panel B shows a mid-cell slice with fibers extending from cells indicated by red arrows. Panel C show the tops of cells where large fibers indicated by red arrows are observed. The fibers interconnect and brightly fluorescent nodes are also observed along their length.*

Chapter 4: Discussion

4.1 *Significance of Cell Internalizing Aptamers*

In this study, a cell selection experiment was conducted to isolate functional DNA sequences that could efficiently internalize into MCF7 cells. A library of self-internalizing aptamers provides an attractive alternative to existing transfection methods for delivery of drugs and nucleic acids into cells. The importance of this work is reflected in its potential applications in basic research, diagnostics and therapeutics. Previous research has shown that aptamers can be raised against cell surface markers in complex cell selection experiments (35-39). Furthermore, surface binding sequences have also been found to internalize into cells (19, 27). The novelty of the research presented here is primarily the change in focus of cell selection from a search for high binding affinity and specificity to a search for sequences that internalize efficiently. While these goals may seem similar, the sequence and targeting requirements of the optimal aptamers for these different types of selection may be vastly different. The importance of aptamer stability and flexibility in downstream applications were also considered in this selection, resulting in the use of a circular DNA based structure.

4.2 *Optimizing Selection of Cell-Internalizing DNA Sequences*

In traditional cell selection experiments for surface binding aptamers, bound sequences are eluted by treatment with trypsin or an elution buffer. In selection for cell internalizing aptamers, only the sequences within cells are desirable. This presents the challenge of recovering minute pieces of DNA at low concentration among other cellular components while ideally excluding genomic DNA. Purification using TRIzol reagent successfully

removed RNA, however the sample still required size filtration to reduce genomic DNA found to interfere with qPCR. It is likely most of this loss can be attributed to the multiple wash steps in the TRIzol method. Selection may be significantly improved by increasing the efficiency of lysate purification especially in early rounds of selection where some sequences may only exist in single copy. To reduce the number of steps where DNA may be lost, urea PAGE could be used to simultaneously denature, purify, concentrate and size separate library sequences.

Functional DNA sequences isolated in this selection were all found to localize to the nuclei of cells, however it should be reiterated that a mechanistic study is needed to classify the G18d sequences as true aptamers. Further characterization will be required to determine the targeting mechanism of the G18d sequences, it is possible that the selection method was biased towards nuclear localization. Wash steps and treatments meant to remove surface bound sequences that rupture the cell membrane could lead to loss of cytoplasmic sequences. Nuclear localized DNAs have additional protection within the nuclear membrane possibly resulting in their increased representation in cell lysates. Furthermore, differences in the subcellular environment between nuclei, cytoplasm or organelles may affect the efficiency of assays such as *in situ* RCA, this may skew the apparent relative localization of G18d sequences. Future cell selection experiments could investigate the effects of a variety of wash methods on the subcellular targeting of the resultant aptamer library. Additionally, cell fractionation could also be employed to specifically select sequences targeting an organelle of interest. Post-selection fractionation has the advantage of requiring aptamer sequences to first penetrate the cell membrane before organelle localization, this key requirement would not be met in a

selection using purified organelle targets and is a further example of an advantage of cell selection versus *in vitro* selection.

4.3 *G18d Sequences Successfully Internalize in MCF7 cells*

The investigated sequences G18d-14, 22, 5 and 70 represent a cross-section of sequences isolated from the CDL2 library selection. Quantitative PCR suggests that G18d sequences were recovered preferentially to control sequences however a large variability between replicates is also observed (Figure 3). This large error likely stems from difficulties in sample purification relating to low yields or lysate contaminants that interfere with PCR, as previously mentioned the lysate purification method should be further optimized for future selections.

A primary step with regard to cell internalizing DNAs is identifying the means of internalization. Intrinsic cellular uptake systems that may play a part include clathrin-mediated endocytosis, the caveolae pathway and macropinocytosis. The receptor mediated nature of the clathrin-mediated and caveolae pathways make them likely targets for cell internalizing DNAs. As with previously described cell surface binding aptamers, the unique three dimensional fold of the DNA determines specificity and binding affinity, therefore it is predicted that functional DNAs of the G18d selection have leveraged this mechanism to achieve cell membrane penetration. It is also possible that a specific primary nucleotide sequence is responsible for cell penetration and nuclear targeting by serving as a binding site for transcription factors possessing a nuclear localization signal (NLS) thereby co-opting an existing pathway into the cell. In the latter case, the G18d selection DNAs would represent one of the smallest transcription factor binding sites

capable of facilitating nuclear transport, as known sequences are generally greater than 75 nt in length (40). The qPCR recovery assay shows preferential internalization of G18d sequences as indicated by the increased recovery of G18d sequences versus controls (Figure 3). This is consistent with either clathrin-mediated or caveolae pathway endocytosis, however it is also possible G18d sequences have found alternative internalization targets such as binding extracellular structural components, cell membrane components or binding ligands that are then actively internalized through an existing mechanism. Strong binding of G18d sequences to membrane components such as lipid or carbohydrate would gradually internalize sequences via the membrane recycling pathways, however it would then be expected that RCA particles are found not only in nuclei but throughout other membrane containing organelles of the cell. A time course assay to determine how rapidly sequences are internalized may give further clues as to the mechanism of internalization. The potential also exists that selected G18d sequences do not bind cells at all, but rather bind compounds found in media that cells specifically uptake, this hypothesis could also explain the nuclear localization of RCA particles if a nuclear localizing molecule is the target of G18d DNAs. In complement to these hypotheses the recovery of two different non-selected control sequences showed no selectivity and minimal internalization, consistent with non-specific uptake via macropinocytosis or micropinocytosis. Given a 100 nM DNA concentration, and an average clathrin vesicle diameter of 100 nm, a non-selected sequence would be expected to be internalized once per 32 vesicle internalization events via the clathrin pathway alone (42).

Additional evidence of efficient internalization was provided by *in situ* RCA analysis of G18d sequences, G18d-14 in particular (Figure 4). The number of RCA particles corresponds to increasing concentrations of G18d-14 suggesting that the pathway for internalization is not yet saturated at 200 nM DNA concentrations. The *in situ* RCA method provides visual evidence for internalization and is highly specific due to the requirement of not only a complementary primer binding site but also for a circular template (Figure 5). The process of generating fluorescent *in situ* RCA particles however involves multiple steps and may underestimate the number of G18d sequences actually present and their subcellular localization as mentioned previously. For the RCA reaction to occur, DNA circles must have enough freedom to allow phi29 DNA polymerase to amplify around its circumference unencumbered. This would require that the sequences be dissociated from any potential targets. Therefore, sequences have the potential to be lost from the cell after the permeabilization step required for entry of RCA components. Similarly, products may be lost from cells after the RCA reaction, but this is less likely due to the large size of RCA products. The leakage of G18d sequences in cells during *in situ* RCA may explain the observation of increased background with G18d sequences versus control sequences (Figure 7). If background RCA products were due to non-specific binding, then equal numbers of background particles would be expected for a given concentration from all samples. The identification of G18d targets may provide further explanation for glass surface binding background.

4.4 *RCA Particle Localization*

The localization of all G18d sequences to the nucleus would suggest that a good place to start looking for targets would involve pathways with cell surface receptors that traffic to

the nucleus (Figure 6). The lack of cytosolic RCA particles may also suggests that DNAs are not being sequestered in endosomes as is observed with other methods of DNA transfection (40, 41). However as previously mentioned, varying subcellular environments may affect RCA efficiency such as the low pH environment found in endosomes therefore a second method for determining the localization of G18d sequences should be used to confirm this observation such as direct fluorescent or biotin labeling. In traditional transfection, efficient nuclear trafficking depends upon the breakdown of the nuclear membrane during mitosis, thereby allowing diffusion of transfected sequences and their entrapment when the nuclear membrane is regenerated. In the case of G18d sequences, RCA particles in the nucleus are observed in almost all cells within one hour of incubation with DNA. This provides further evidence for a trafficking mechanism that allows passage through nuclear pore complexes without breakdown of the nuclear membrane. Further studies on the G18d sequences should investigate the effects of time on internalization efficiency. The differing numbers and morphologies of RCA particles generated by the G18d sequences could suggest that the sequences target unique receptors, or target the same receptor with differing efficiencies.

The suggestion that G18d internalization is mediated by surface receptor binding would seem to be supported by the observation that cells with greater exposed surface area show higher numbers of RCA particles (Figure 8). This observation also explains some of the variability in the count of G18d-14 RCA particles because fields of view containing dense cell growth will uptake fewer particles per cell versus sparsely populated fields. Future experiments may attempt to address this by changing cell culturing methods to achieve a

more uniform distribution of cells. Alternatively, normalizing particle count to cell surface area instead of cell count may also improve consistency.

In visual inspections of G18d *in situ* RCA experiments, G18d-70 treated samples have unique features not observed in other sequences. Fluorescent strands are observed at glass surface level running perpendicular to perimeters of cell clusters and extending along the cell membrane to the tops of cells (Figure 9). The fact that these fibers are not seen in other G18d samples or controls suggests some sequence dependence. It is possible that this observation is not biologically significant, but an artifact of the RCA reaction on this particular sequence. However, the parallel alignment of the fibers and their relation to cells strongly suggests the involvement of some cell structure. Perhaps G18d-70 has the ability to bind extracellular matrix components, as no fibers are observed within cytoplasm and only diffuse particles are observed within nuclei.

4.5 *Further Characterization of G18d Sequences*

Given the promising qPCR and *in situ* RCA data presented thus far, several further lines of study should be considered to better define the properties of G18d sequences such as determining the extents of their concentration dependence, time dependence, sequence dependence and their cell type specificity. Systematic mutation of the G18d sequences to determine the minimal essential sequence will be valuable in further functionalization experiments where covalent modification will be required. The applicability of G18d sequences identified in this selection will also depend upon their cell type specificity, this may also give clues as to the targets and pathways exploited for internalization.

Ultimately the goal of this study is to develop a functional DNA that can carry a therapeutically relevant molecule into cells. To this extent, an attempt should be made to modify G18d sequences with compounds that can exert a therapeutically relevant effect. This may include demonstrating the efficient internalization and activity of a chemotherapeutic coupled to a G18d sequence, or possibly gene knockdown via an anti-sense DNA coupled to a G18d sequence. Chemotherapeutics such as doxorubicin have been used previously to show therapeutic applications of cell internalizing aptamers and therefore may be good starting point for comparing G18d sequences. Given the apparent nuclear targeting of G18d sequences, developing its use as a transfection platform is particularly interesting and relevant to challenges currently faced by existing gene delivery and knockdown systems. The true success of this selection experiment and the G18d sequences will be measured by their ability to deliver such payloads and show an enhanced effect compared to existing methods.

While not necessarily essential to the goal of cell internalization, the identification of the target or targets of G18d sequences is a major question. A known target may give useful insight into designing better cell internalizing functional DNAs as well as help to define the limits of that particular pathway and the technology in general. Preliminary attempts at identification may include surface protein ablation via protease digestion prior to G18d sequence presentation, a loss of internalization would suggest that a surface protein is a target. Similarly, digestion of surface carbohydrates prior to G18d incubation resulting in loss of internalization would suggest a target with a carbohydrate component. G18d internalization despite these treatments may suggest a mechanism binding some other membrane component such as lipid.

The novelty of this cell internalization selection leaves much room for optimization at the level of selection. Numerous combinations of selection conditions give the potential for a wide range of cell internalizing sequences to be identified. While not employed in this selection, counter-selection steps can also play an important role in generating specificity within the library in future experiments. As mentioned previously, the method could also be modified to select for subcellular localization to organelles. Overall, cell selection for cell internalizing DNA sequences has yet to be fully exploited.

Chapter 5: Conclusion

Previous research has shown that aptamers can be selected against cell surface markers in cell selection experiments (35-39). Sequences from these studies were then found to internalize into cells (19, 27) . A library of self-internalizing aptamers provides an attractive alternative to existing transfection methods for delivery of drugs and nucleic acids into cells. The significance of this work is reflected in its potential applications in basic research, diagnostics and therapeutics. With the goal of developing an aptamer that can efficiently internalize into cells, a cell selection experiment was carried out using a random-sequence circular DNA library selected against MCF7 cells. G18d sequences isolated from this selection have been shown to target cell nuclei at a rate significantly greater than control sequences as shown by qPCR relative recovery assays and *in situ* RCA fluorescence microscopy. Further characterization and target identification should be done to better understand the mechanism of internalization, to determine whether G18d sequences are true aptamers and to judge the suitability of G18d sequences as a delivery platform.

REFERENCES

1. Jackson, A. L., and Linsley, P. S. (2010) *Nat. Rev. Drug Discov.* **9,** 57-67

2. Ellington, A. D., and Szostak, J. W. (1990) *Nature* **346,** 818-822

3. Tuerk, C., and Gold, L. (1990) *Science* **249,** 505-510

4. Conrad, R., Keranen, L. M., Ellington, A. D., and Newton, A. C. (1994) *J. Biol. Chem.* **269,** 32051-32054

5. Schneider, D. J., Feigon, J., Hostomsky, Z., and Gold, L. (1995) *Biochemistry* **34,** 9599-9610

6. Orava, E. W., Cicmil, N., and Gariepy, J. (2010) *Biochim. Biophys. Acta*

7. Santini, D., Fratto, M. E., Vincenzi, B., Napoli, N., Galluzzo, S., Tantardini, M., Abbruzzese, A., Caraglia, M., and Tonini, G. (2009) *Curr. Cancer. Drug Targets* **9,** 834-842

8. Stuart, M. A., Huck, W. T., Genzer, J., Muller, M., Ober, C., Stamm, M., Sukhorukov, G. B., Szleifer, I., Tsukruk, V. V., Urban, M., Winnik, F., Zauscher, S., Luzinov, I., and Minko, S. (2010) *Nat. Mater.* **9,** 101-113

9. Rothdiener, M., Muller, D., Castro, P. G., Scholz, A., Schwemmlein, M., Fey, G., Heidenreich, O., and Kontermann, R. E. (2010) *J. Control. Release*

10. Chu, T. C., Marks, J. W.,3rd, Lavery, L. A., Faulkner, S., Rosenblum, M. G., Ellington, A. D., and Levy, M. (2006) *Cancer Res.* **66,** 5989-5992

11. Wagner, H. (2008) *Curr. Opin. Immunol.* **20,** 396-400

12. Thiel, K. W., and Giangrande, P. H. (2009) *Oligonucleotides* **19,** 209-222

13. Cerchia, L., Giangrande, P. H., McNamara, J. O., and de Franciscis, V. (2009) *Methods Mol. Biol.* **535,** 59-78

14. Tang, Z., Shangguan, D., Wang, K., Shi, H., Sefah, K., Mallikratchy, P., Chen, H. W., Li, Y., and Tan, W. (2007) *Anal. Chem.* **79,** 4900-4907

15. Cerchia, L., Esposito, C. L., Jacobs, A. H., Tavitian, B., and de Franciscis, V. (2009) *PLoS One* **4,** e7971

16. Shangguan, D., Cao, Z. C., Li, Y., and Tan, W. (2007) *Clin. Chem.* **53,** 1153-1155

17. Xiao, Z., Shangguan, D., Cao, Z., Fang, X., and Tan, W. (2008) *Chemistry* **14,** 1769-1775

18. Kang, H., O'Donoghue, M. B., Liu, H., and Tan, W. (2010) *Chem. Commun. (Camb)* **46,** 249-251

19. Huang, Y. F., Shangguan, D., Liu, H., Phillips, J. A., Zhang, X., Chen, Y., and Tan, W. (2009) *Chembiochem* **10,** 862-868

20. Ulrich, H., Trujillo, C. A., Nery, A. A., Alves, J. M., Majumder, P., Resende, R. R., and Martins, A. H. (2006) *Comb. Chem. High Throughput Screen.* **9,** 619-632

21. Veedu, R. N., and Wengel, J. (2009) *RNA Biol.* **6,** 321-323

22. Dapic, V., Bates, P. J., Trent, J. O., Rodger, A., Thomas, S. D., and Miller, D. M. (2002) *Biochemistry* **41,** 3676-3685

23. Choi, E. W., Nayak, L. V., and Bates, P. J. (2009) *Nucleic Acids Res.*

24. Hardin, C. C., Perry, A. G., and White, K. (2000) *Biopolymers* **56,** 147-194

25. Teng, Y., Girvan, A. C., Casson, L. K., Pierce, W. M.,Jr, Qian, M., Thomas, S. D., and Bates, P. J. (2007) *Cancer Res.* **67,** 10491-10500

26. Wullner, U., Neef, I., Eller, A., Kleines, M., Tur, M. K., and Barth, S. (2008) *Curr. Cancer. Drug Targets* **8,** 554-565

27. Bagalkot, V., Farokhzad, O. C., Langer, R., and Jon, S. (2006) *Angew. Chem. Int. Ed Engl.* **45,** 8149-8152

28. Billen, L. P., and Li, Y. (2004) *Bioorg. Chem.* **32,** 582-598

29. McNamara, J. O., Kolonias, D., Pastor, F., Mittler, R. S., Chen, L., Giangrande, P. H., Sullenger, B., and Gilboa, E. (2008) *J. Clin. Invest.* **118,** 376-386

30. Farokhzad, O. C., Cheng, J., Teply, B. A., Sherifi, I., Jon, S., Kantoff, P. W., Richie, J. P., and Langer, R. (2006) *Proc. Natl. Acad. Sci. U. S. A.* **103,** 6315-6320

31. Kibbe, W. A. (2007) *Nucleic Acids Res.* **35,** W43-6

32. Abramoff, M. D., Magalhaes, P. J., and Ram, S. J. (2004) *Biophotonics International* **11,** 36

33. Xiao, Z., Shangguan, D., Cao, Z., Fang, X., and Tan, W. (2008) *Chemistry* **14,** 1769-1775

34. Huang, Y. F., Shangguan, D., Liu, H., Phillips, J. A., Zhang, X., Chen, Y., and Tan, W. (2009) *Chembiochem* **10,** 862-868

35. Zuker, M. (2003) *Nucleic Acids Res.* **31,** 3406-3415

36. Zhang, Y., Chen, Y., Han, D., Ocsoy, I., and Tan, W. (2010) *Bioanalysis* **2,** 907-918

37. Sefah, K., Tang, Z. W., Shangguan, D. H., Chen, H., Lopez-Colon, D., Li, Y., Parekh, P., Martin, J., Meng, L., Phillips, J. A., Kim, Y. M., and Tan, W. H. (2009) *Leukemia* **23,** 235-244

38. Chen, H. W., Medley, C. D., Sefah, K., Shangguan, D., Tang, Z., Meng, L., Smith, J. E., and Tan, W. (2008) *ChemMedChem* **3,** 991-1001

39. Tang, Z., Shangguan, D., Wang, K., Shi, H., Sefah, K., Mallikratchy, P., Chen, H. W., Li, Y., and Tan, W. (2007) *Anal. Chem.* **79,** 4900-4907

40. Lam, A. P., and Dean, D. A. (2010) *Gene Ther.* **17,** 439-447

41. Collins, E., Birchall, J. C., Williams, J. L., and Gumbleton, M. (2007) *J. Gene Med.* **9,** 265-274

42. Swanson, J. A., and Watts, C. (1995) *Trends Cell Biol.* **5,** 424-428

Appendix I: Synthetic Oligonucleotides

Oligonucleotide	Sequence (5`-3`)
CDL2 Library	TTCGGAAGAGATGGCGAC-N40-CGAGCTGATCCTGATGGAA
CDL2P1 Primer	TTCGGAAGAGATGGCGAC
CDL2P3 Primer	ATGTCGTGCGTGCTA-SP18-TTCCATCAGGATCAGCTCG
CDL2P3-untailed Primer	TTCCATCAGGATCAGCTCG
CDL2 Ligation Splint	TCTCTTCCGAATTCCATCAGGA

Table A1 – *Oligos used for G18d selection and universal CDL2 primers.*

Oligonucleotide Primer	Sequence (5`-3`)
G18d-70-F	GGCGACCACTCCCTC
G18d-70-R	AGCTCGGAAAATGCGTCATAG
G18d-5-F	GGCGACCCGTTGGT
G18d-5-R	AGCTCGAGGAACGTCATAG
G18d-9-F	GGCGACCCCCACC
G18d-9-R	AGCTCGGGAGGTTCTAGA
G18d-69-F	GGCGACCACTGGGG
G18d-69-R	AGCTCGAGTGTATTGACATCG
G18d-22-F	GGCGACACACATGCCT
G18d-22-R	AGCTCGGGCAGGAGATT
G18d-14-F	GGCGACACATTCCGCT
G18d-14-R	AGCTCGTCAGGACCG
CDL2 Con1-F	CGCGCGTTTTACTTTTCTG
CDL2 Con1-R	AGCTCGACAGTACGGC
CDL2 Con2-F	GGCGACGAATAGAATGCG
CDL2 Con2-R	CGCTGCCCCCAAC

Table A2 – *Sequence specific primers used in qPCR recovery assay to amplify selected G18d sequences from control sequences in lysate mixture.*

Oligonucleotide Probe	Sequence (5`-3`)
G18d-RA-Fluor	✶ ✶ ✶ TTCGGATGTGATGGCGAC
G18d-LA-Fluor	✶ ✶ CGAGCTGATCCTGATGGAA

```
                          ✶ ✶ ✶                    ✶ ✶
RA + LA Probes -    TTCGGATGTGATGGCGAC     CGAGCTGATCCTGATGGAA
                    |||||| | || ||||||     |||||||| || |||||||
CDL2 Antisense – ...AAGCCTTCTCTACCGCTG-N40-GCTCGACTAGGACTACCTT...

✶ Represents dT-Fluorescein modification
```

Table A3 – *Fluorescent probes for in situ RCA detection.*

Appendix II: ImageJ Macros

ImageJ Nuclei Counting Macro

```
macro "Nuclei Counting Batch"{

source_title=getTitle();
source_ID=getImageID();
source_directory=getDirectory("image");
root_directory=File.getParent(File.getParent(source_directory))+File.sep
arator;
//print(source_directory);
//print(root_directory);

selectImage(source_ID);
close();

if(File.exists(root_directory)) {
	root_list=getFileList(root_directory);
	//print(lengthOf(root_list));

	for (i=1; i<root_list.length; i++) {
		//print(i);

	instance_path=root_directory+"ROI"+i+File.separator+"Red"+File.se
parator;
		instance_list=getFileList(instance_path);
		instance_file=instance_path+instance_list[0];
		//print(instance_file);
		open(instance_file);
		instance_ID=getImageID();
		instance_Title=getTitle();
		nucleiCount(instance_Title);
		selectImage(instance_ID);
		close();
	}
}

function nucleiCount(ncTitle){

	selectWindow(ncTitle);
	run("Grouped ZProjector", "group="+nSlices+" projection=[Max
Intensity]");
	selectWindow("Projection of "+ncTitle);
	run("Subtract Background...", "rolling=100");
	run("8-bit");
	run("Gaussian Blur...", "sigma=3");
	selectWindow("Projection of "+ncTitle);
```

```
        setAutoThreshold("Default dark");
        //run("Threshold...");
        setThreshold(7, 255);
        run("Convert to Mask");
        run("Fill Holes");
        run("Watershed");
        run("Analyze Particles...", "size=1000-Infinity circularity=0.00-
1.00 show=Nothing summarize in_situ");
        close();

}

}
```

ImageJ Particle Counting Macro

```
source_title=getTitle();
source_ID=getImageID();
source_directory=getDirectory("image");
root_directory=File.getParent(File.getParent(source_directory))+File.sep
arator;
//print(source_directory);
//print(root_directory);

selectImage(source_ID);
close();

if(File.exists(root_directory)) {
	root_list=getFileList(root_directory);
	//print(lengthOf(root_list));

	for (i=1; i<root_list.length; i++) {
		//print(i);

	instance_path=root_directory+"ROI"+i+File.separator+"Green"+File.
separator;
		instance_list=getFileList(instance_path);
		instance_file=instance_path+instance_list[0];
		//print(instance_file);
		open(instance_file);
		instance_ID=getImageID();
		instance_Title=getTitle();
		particleCount(instance_Title);
		selectImage(instance_ID);
		close();
	}
}

function particleCount(ncTitle){

	selectWindow(ncTitle);
	run("Grouped ZProjector", "group="+nSlices+" projection=[Max
Intensity]");
	selectWindow("Projection of "+ncTitle);
	run("8-bit");
	run("Subtract Background...", "rolling=15");
	run("Gaussian Blur...", "sigma=0.90");
	selectWindow("Projection of "+ncTitle);
	//run("Threshold...");
	setAutoThreshold("Default dark");
	setThreshold(20, 255);
	run("Convert to Mask");
```

```
        run("Watershed");
        run("Analyze Particles...", "size=8-Infinity circularity=0.00-
1.00 show=Outlines summarize");
        close();
        close();

}
```