DNA APTAMERS HAVE AN ANALOG IN GENOMIC DNA

By

MEGHAN BRADLEY O'DONOGHUE

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

2011

To Mom, daD, Carrie, Peter
and the two great loves of my life:
Josh and DNA

ACKNOWLEDGMENTS

I would like to thank the scientists who prodded me over the years to think harder, dig deeper, and then reveled with me in the ecstasy of discovery, especially when the journey was excruciating. In particular, I want to thank my dear “lab-sisters,” Yan Chen, Lin Wang, Zeyu Xiao, Youngmi Kim, Prabodhika Mallikarachy, Maria-Carmen Estevez, Karen Milians, Jennifer Martin, Zhi Zhu, Yanrong Wu, Ling Meng, Hui Wang, Suwussa Bamrungsap, Jian Wang, and Dalia Lopez Colon; you taught me with style and grace—how to work hard, cut loose, and earn a Ph.D. I am also grateful to my mentors in the scientific life, who have shaped my professional growth, and guided me down my path, including, David Ostrov, Richard Condit, Nancy Denslow, Bradley Fletcher, Joshua Socolar, Georgia Dadoly, Gregory Schultz, Chen Liu, Katherine Williams, and Weihong Tan. Much gratitude also goes to Kwame Sefah, Joseph Phillips, Dihua Shangguan, Haipeng Liu, Parag Parekh, Ibrahim Shukoor, and Yunpeng Cai who shared with me their expertise and scientific rigor. They got me asking the right questions and provided me with the tools to get the answers I needed. Thanks to Xiaohong Fang, and her lab in Beijing, especially Xiaoli Shi and Zilong Zhao, for showing me immense hospitality, and helping me succeed during the Olympic summer of 2008. Thanks to Stephen Larner for starting me in research and Josie Cooke for keeping me focused at the end of it. Finally, I would like to thank my dear friends who will remain, for now, in Florida, especially Weijun Chen, Xiangling Xiong, Guizhi Zhu, Tao Chen, Dimitri van Simaeys, Elizabeth Jimenez, Yunmin Chang, Sena Cansiz, Diane Turek, Jen Maier, Catherine Buchanan-McGrath, and Jeff Boissoneault: may all of your pipettes be calibrated, and may all of your experiments work the first time.

Please keep in touch, Josh and my’s home will be open to you all.

TABLE OF CONTENTS

page

ACKNOWLEDGMENTS 4

LIST OF TABLES 8

LIST OF FIGURES 9

LIST OF ABBREVIATIONS 11

ABSTRACT 17

CHAPTER

1 INTRODUCTION 21

Aptamer Selection 23
The PTK7 Aptamer Sgc8c 25
Biology of PTK7 28
Wnt Signaling Pathway 28
B-Catenin/Wnt Pathway 30
PCP/Wnt Pathway 31
PTK7 Structure 34
Pseudo-Tyrosine Kinase Domain of PTK7 35
Transmembrane Domain of PTK7 36
Extracellular PTK7 and its MT-MMP1 Cleavage Site 37
PTK7 Gene Structure and Splicing 39
sgc8c Aptamer 40
DIXDC1 Biology 40
DIXDC1 Increases Wnt/β-catenin Signaling 41
DIXDC1 Decreases Wnt/PCP Signaling 42
DIXDC1 Gene Structure 43

2 PTK7 APTAMERS HAVE AN ANALOG IN GENOMIC DNA 46

Introduction 46
Results and Discussion 47
38nt sgc8c Binds as well as 41nt sgc8c 47
Bioinformatic Analysis of cell-SELEX Aptamers 48
Statistical Analysis of Consensus Region 53
Consensus Region is Important to Aptamer Binding 53
Mutational analysis 54
Consensus sequence blocking 55
Sgc8c pyrene mutants suggest sgc8c interacts intimately with PTK7 57
BLAST of the Consensus Sequence against the Human Genome 60

DIXDC1b has the Consensus Sequence 60
KC2D4 has Identity with DIXDC1b Negative Strand 63
Conclusions 66
Materials and Methods 67
DNA Sequences 67
Cell Culture 68
Bioinformatics 70
Significance Simulations 70
Competition and Binding Experiments 71
Off-Rate Experiments 72
Kd Measurements 72
Azobenzene Mutation Measurements 72
cDNA Blocking 73
Pyrene-sgc8c Binding 73

3 POSSIBLE FUNCTIONAL ROLE FOR PTK7-DIXDC1B INTERACTION 74

Introduction 74
Results and Discussion 75
PTK7 Aptamers have Sequence Identity with DIXDC1b 5' UTR 75
The Consensus Region is Conserved among Species 76
DNA is Double Stranded, but our Aptamers are Single Stranded 77
PTK7 Structural Analysis 79
PTK7 Has a Weak Predicted Nuclear Localization Sequence 82
Confocal Microscopy 83
Summary of PTK7 Known and Predicted Features 84
Cleaved PTK7 Fragment is Higher in Cellular Nuclear Fraction 84
Sgc8c Prevents Wound Healing in HeLa Scratch Assay 85
Conclusions 86
Materials and Methods 90
DNA Sequences 90
Cell Culture 91
dsDNA Binding Experiments 91
Western Blot 92
Scratch Assay 92

4 ADDITIONAL SEQUENCE IDENTITIES FOUND BETWEEN APTAMERS BY CELL-SELEX BIOINFORMATIC ANALYSIS 93

Introduction 93
Vaccinia Aptamers 93
BLAST of the Consensus Sequence against the Human Genome 95
Conclusions 96

5 SGC8C-APTAMER AND PTK7-ANTIBODY RUPTURE FORCES ARE COMPARABLE ON LIVE HELA CELLS 98

Introduction 98
Results and Discussion 99
Blocking Controls 100
Fitting the Collected Data 101
Conclusions 101
Materials and Methods 103
Preparation of AFM Tips and Substrate 103
Cell Culture 104
AFM Measurements 104

6 MODIFYING CELLULAR PROPERTIES USING ARTIFICIAL LIPID APTAMER-RECEPTORS 106

Introduction 106
Results and Discussion 108
Streptavidin Aptamer-Receptor Anchors on the Cell Membrane 108
Streptavidin on Modified Cells Remains Functional 112
SA-ARs Can Be Used to Capture Cells for Analysis 114
Thrombin Aptamer-Receptor Captures Thrombin 114
Thrombin Modified Cells Cause Clotting 116
Conclusion 117
Materials and Methods 117
DNA Synthesis 117
Cell Culture 118
Flow Cytometry 119
Fluorescence Microscopy 119
MTS Assay 120
Cell Capture with Streptavidin DynaBeads 120
Clotting Assay 121

7 FUTURE DIRECTIONS AND CONCLUSIONS 123

Future Directions 123
Conclusions 124

APPENDIX

A COMPLEX TARGET SELEX ssDNA APTAMER DATABASE 128

B LIST OF SIMILARITY BETWEEN APTAMERS IN DATABASE 135

LIST OF REFERENCES 138

BIOGRAPHICAL SKETCH 150

LIST OF TABLES

Table page

1-1 The many uses of sgc8c 27

1-2 PTK7 orthologs in various species 35

2-1 31nt sgc8c binds as well as 41nt sgc8c 47

2-2 Aptamer sequences showing sequence similarity 51

2-3 Azobenzene mutated sgc8c sequences 55

2-4 BLAST hits 14/15nt identity for GCTGCGCCGCCGGGA in genome 62

2-5 Aptamers share sequence similarity with DIXDC1b DNA sequence 63

2-6 KC2D4 has identity to the negative DIXDC1b strand 65

2-7 All DNA sequences used in this chapter 68

3-1 Consensus region is conserved among species 77

3-2 Sequences used in this chapter 91

4-1 Aligned aptamer sequences from three Vaccinia related selection 94

4-2 BLAST results of VV genome revealed 96

6-1 Sequences used for aptamer-receptors 118

LIST OF FIGURES

Figure page

1-1 Sgc8 structure and optimization 26
1-2 Schematic Wnt signaling pathway 29
1-3 Examples of planar cell polarity 32
2-1 Hyperbolic fitting of data collected at different aptamer concentrations 48
2-2 Competition studies 52
2-3 Azobenzene-sgc8c mutants 54
2-4 Blocking consensus region with complementary DNA, blocks binding 56
2-5 Pyrene sgc8c-mutants 59
2-6 Pyrene sgc8c-mutant data 61
3-1 The consensus sequence is found in DIXDC1b 5' UTR 75
3-2 NuPACK structures for DIXDC1b DNA 78
3-3 dsDIX DNA does not bind CEM cells 78
3-4 Structural features of Ig-like folds 80
3-5 Model of PTK7 extracellular domain 81
3-6 Annotated PTK7 primary amino acid sequence 84
3-7 Western blot of various cellular compartments of HEK293 cells 85
3-8 Scratch assay 87
3-9 Scheme of possible PTK7-aptamer interactions 88
3-10 Model for PTK7-DIXDC1b interaction 89
4-1 Predicted secondary structures of consensus sequence 95
4-2 Sequence identity between two aptamers selected in different labs 97
5-1 Scheme of AFM measurements on live cells 99
5-2 Binding probabilities of tip with cells 100

5-3 Histograms of binding forces between tips and HeLa cells 101
5-4 Rupture forces from literature of different lengths of dsDNA and sgc8c..... 102
5-5 Representative force-distance curves for sgc8c AFM tip and HeLa..... 105
6-1 Characteristics of streptavidin-artificial receptors 109
6-2 Aptamer insertion begins after 5min and saturates after an hour..... 111
6-3 SA-AR does not inhibit cell growth 112
6-4 Streptavidin-modified cells bind biotin to make cell assemblies 113
6-5 SA-AR modified cells are collected with magnetic beads..... 114
6-6 Cells modified with thrombin via a thrombin aptamer artificial receptor..... 115
6-7 Thrombin-Modified Cells Cause Clotting 116

LIST OF ABBREVIATIONS

α-PTK7	PTK7 antibody
5’-UTR	5’-untranslated region
aa	Amino acids
AFM	Atomic force microscopy
AFU	Arbitrary Flourescence Units
AKT	Protein Kinase B
AMA	Ammonium hydroxide: methylamine 1:1
AML	Acute myeloid leukemia
ANOVA	Analysis of variance
APC	Adenomatous polyposis coli
ATCC	American Type Culture Collection
ATP	Adenosine triphosphate
BB	Binding Buffer (WB, 1g/L BSA, and 100mg/L tRNA)
bp	Base pair
BLAST	Basic local alignment search tool
BLASTn	Nucleotide-BLAST
BTG4	BTG/TOb-family-protein
BSA	Bovine serum albumin
Ccd1	Coiled-coil domain 1 protein
cDNA	Complementary DNA
CEM	Human T-Cell Acute Lymphoblastic Leukemia cell line
CH	Calponin homology domain
chz	Chuzhoi mutant
CLUSTAL	Multiple sequence alignment computer program

CPG	Controlled pore glass bead
DAAM1/2	Disheveled-associated activator of morphogenesis 1/2
DAPI	4',6-diamidino-2-phenylindole
DIXDC1	DIX domain containing 1 protein
DIXDC1a	Long DIXDC1 isoform
DIXDC1b	Short DIXDC1 isoform
DLAT	Dihydrolipoamide S-acetyltransferase
DMEM	Dulbecco's Modified Eagle Media
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DOX	Doxorubicin
ds	Double-stranded
dsDNA	Double-stranded DNA
Dtrk	PTK7 ortholog in zebrafish
Dvl	Disheveled
ECM	Extracellular matrix
EDTA	Ethylenediaminetetraacetic acid
EGFR	Epidermal growth factor receptor
ESPRIT	Bioinformatics algorithm for sequence alignment
FASTA	Bioinformatics algorithm for sequence alignment
FBS	Fetal bovine serum
FGFR	Fibroblast growth factor receptor
Flt-1	Vascular endothelial growth factor receptor 1
FRET	Fluorescence resonance energy transfer
Fz	Frizzled receptor

GAPDH	Glyceraldehyde 3-phosphate dehydrogenase
GPCR	G-protein coupled receptor
GSK3β	Glycogen synthase kinase 3
H23	Lung adenocarcinoma cell line
HA	Hemagglutinin
HEK293	Human embryonic kidney cell line
HeLa	Henrietta Lacks's cervical cancer cell line
Her3	Human epidermal growth factor receptor 3
HF	Hydrofluoric acid
HPLC	High pressure liquid chromatography
HRP	Horseradish peroxidase
Ig	Immunoglobulin-like domains
JNK	c-Jun N-terminal kinases
K_d	Disassociation constant
kDa	KiloDalton
KLG	Chicken ortholog of PTK
LRP5/6	Low-density lipoprotein receptor-related protein 5/6
MBTPS1	Membrane-bound transcription factor peptidase
MCOLN1	Mucolipin 1
MDCK	Canine kidney cell line
MEKK1	Mitogen-activated protein kinase kinase kinase 1/4
MINA	Myc induced nuclear antigen
miRNA	MicroRNA
MT1-MMP	Membrane type 1 metalloprotease; MMP-14
MPTMS	(3-mercaptopropyl)-trimethoxysilane

MRI	Magnetic resonance imaging
mRNA	Messenger ribonucleic acid
MTH	Central myosin tail homology domain
NCBI	National center for biotechnology information
NFκB	Nuclear factor kappa-light-chain-enhancer of activated B cells
NFAT/AP1	Nuclear factor of activated T-cells/Fos/Jun
NHS-PEG-MAL	ω-N-hydroxysuccinimide ester-poly(ethylene glycol)-α-maleimide
NLS	Nuclear localization signal
NLStradamus	Nuclear localization prediction alogorithm
NP	Nanoparticle
NPC	Nuclear pore complex
NPH-II	Vaccinia viral helicase nucleoside triphosphate phosphohydrolase-II
NR	Nanorod
nt	Nucleotide
NuPACK	DNA and RNA secondary structure prediction program
PAGE	Poly-acrylimide gel electrophoresis
PBS	Phosphate buffered saline
PCP	Planar cell polarity
PCR	Polymerase chain reaction
PE	Phycoerythrin
PEG	Polyethylene glycol
Pk	Prickle
PP2A	Protein phosphatase 2
PredictNLS	Nuclear localization prediction algorithm
pTK	Tyrosine pseudokinase domain

PTK7	Protein tyrosine kinase 7
qRT-PCR	Quantitative reverse transcriptase PCR
RACK1	Receptor of activated protein kinase C 1
Rho/ROCK	Rho-associated protein kinase
RhoA GTPase	Ras homolog gene family, member A
RNA	Ribonucleic acid
Ror	Receptor tyrosine kinase-like orphan receptor
RPMI-1640	Commonly used cell media
Runx1	Runt-related transcription factor 1
RYK	Related to receptor tyrosine kinase
SA	Streptavidin
SELEX	Systematic Evolution of Ligands by EXponential enrichment
SET	Surface energy transfer
SHANK1	SH3 and multiple ankyrin repeat domains protein 1
SIK2	Salt-inducible serine/threonine kinase 2
siRNA	Small interfering RNA
sPTK7	Soluble PTK7; the cleavage fragment of PTK7 produced by MT1-MMP
ssDNA	Single stranded deoxyribonucleic acid
STAT	Signal Transducers and Activators of Transcription protein
SV40	Simian virus 40
TBE	Tris/Borate/EDTA buffer
TCF/LEF	T-cell factor/lymphoid enhancer factor
TEAA	Triethylammonium acetate buffer
TGF-β	Transforming growth factor beta
TK	Tyrosine kinase

Transfac	Transcription factor prediction program
Tris-HCl	Tris(hydroxymethyl)aminomethane-HCl
TrkB	TrkB tyrosine kinase
tRNA	Transfer RNA
UBE2D2	Ubiquitin-conjugating Enzyme E2D2
UF	University of Florida
UV	Ultra-violet
UV-Vis	UV-Visible
Vangl1/2	van Gogh protein
VEGF-A	Vascular endothelial factor A
VEGFR1	Vascular endothelial growth factor receptor 1
WB	Washing buffer (PBS, 4.5g/L glucose, 1M MgCl2)
Wnt	Wingless-Int
VV	Vaccinia virus

Abstract of Dissertation Presented to the Graduate School
of the University of Florida in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy

DNA APTAMERS USED TO DETECT, TREAT, AND PROBE
THE BASIC BIOLOGY OF CANCER

By
Meghan Bradley O'Donoghue

December 2011

Chair: Weihong Tan
Major: Medical Sciences with a Concentration in Physiology and Pharmacology

Cell-SELEX generates artificial DNA and RNA molecules (aptamers) that bind to biological targets of interest with applications for research and therapy. Each whole cell-SELEX uses a large library of random DNA sequences (~10^{15}), amplified by unique primers, against cells with more than 10,000 unique protein targets. Using this tool, our lab has selected a number of DNA aptamers that bind to leukemia, breast, and colon cancer cell lines. Bioinformatic and competition analysis revealed that aptamers binding 1 protein epitope were independently selected 4 separate times in selections against different cell lines. Searching for a biological explanation for this result, we discovered a genomic DNA analog to a common DNA sequence in these aptamers.

The first of these aptamers to be selected, sgc8c, was found to bind the extracellular portion of the receptor tyrosine kinase PTK7, which is important for non-canonical Wnt signaling, and is misregulated in numerous cancers. Bioinformatic analysis of 148 aptamer sequences selected by whole cell-SELEX revealed significant sequence identity between sgc8c and 3 other aptamers within a 15nt "consensus region" (GCTGCGCCGCCGGGA). As predicted, all 4 aptamers competed with each other and bound to the cells with similar affinity, implying that they bound to the same

place on the PTK7 protein. Furthermore, mutational analysis of sgc8c indicated that the consensus region, but not the surrounding nucleotides, is important to aptamer binding.

We were curious whether the consensus region's repeat selection, as well as its importance in aptamer-protein binding, reflected a functional role for the consensus sequence. A BLAST search found that the consensus region, minus the first G base, appears 5 times in the human genome. One of these sites is in the 5'-untranslated region(5'-UTR) of human DIXDC1b DNA, which, along with PTK7, is a regulator of non-canonical Wnt signaling. Surprisingly, aligning the aptamers with DIXDC1b DNA, we found 2 of the aptamers, KMF9b and H01, had additional nucleotides in common with DIXDC1b DNA outside the consensus region. This 22nt region, shared by the PTK7 aptamers and DIXDC1b DNA, is unique in the human genome. We were also surprised that another aptamer, KC2D4, which also competes with the PTK7 aptamers, but which shares no sequence similarity with them, has a region of significant sequence identity to the opposite strand of DIXDC1b DNA.

In total, 5 aptamers compete for binding the same site on PTK7. Four of these aptamers share sequence identity to the positive strand of DIXDC1b DNA's 5'-UTR. The fifth aptamer shares sequence identity 5 bases from this region, but on the opposite strand of DIXDC1b DNA. The aptamers' sequence identity to both the positive and negative strands of DIXDC1b DNA could be consistent with the protein PTK7 melting the genomic DNA and interacting with the resulting ssDNA hairpins, which share sequence identity to the aptamers formed by each melted strand. If this is indeed the case, it could affect DIXDC1b transcription. These sequences identities are conserved

in the putative DIXDC1b regulatory regions of *Pan troglodytes*, *Rattus norvegicus*, and *Mus musculus*.

PTK7 has a weak predicted nuclear localization sequence. By cell fractionation, followed by Western blot, we found an abundance of cleaved PTK7 in the nuclei of several different cell lines. This suggests that whole cell-SELEX has not just found an aptamer for an extracellular protein, but may also have identified a genomic DNA sequence, which the protein binds to naturally—after it is cleaved, internalized, and transported to the nucleus.

While we have only begun to address this receptor tyrosine kinase and this regulatory pathway, these findings suggest that whole cell-SELEX might be used more generally to identify novel extracellular transcription factors with highly specific binding motifs. Our bioinformatic analysis of other aptamer sequences selected against whole cells has yielded other examples of disparate selections yielding similar sequences, hinting that there may be other proteins that interact with DNA on the plasma membrane yet to be discovered.

Our results are important for three main reasons: (1) The proteins PTK7 and DIXDC1b are murkily understood, yet key players in Wnt signaling, which is crucial to embryonic development and cancer progression. Understanding their relationship could be important for understanding Wnt signaling. For instance, there are two isoforms of DIXDC1, a and b; only DIXDC1b has the consensus sequence. PTK7 binding this region could act as a transcriptional switch for this isoform's production. (2) This family of PTK7-binding aptamers has sequence identity to both strands of DIXDC1b DNA, implying that, if melted, the DNA might be forming ssDNA hairpins that bind to PTK7 in

a manner similar to the aptamers. This would be a new type of transcription factor of interest to molecular biologists. (3) Finally, this finding is probably not a one-time occurrence. Our bioinformatic analysis of 148 DNA sequences selected by whole-cell SELEX identified other aptamers from disparate selections, like those for *Vaccinia* infected cells and pure virus, which share significant sequence identity. Further comparison of existing DNA and RNA aptamers may yield other examples of SELEX identifying natural DNA or RNA sequences with functional roles, not initially envisioned. Future selections should also not be considered complete until the newly selected aptamers are compared with all other existing aptamers for sequence identity.

CHAPTER 1
INTRODUCTION

Genetic alterations in cells cause cancer, which result in different cell behavior due to changes to the cell at the molecular level. There has been a big push in recent years to develop biomarkers for these changes to improve cancer diagnosis and treatment. To identify unique molecular features of target cancer cells, our lab has previously developed a cell-SELEX (cell-systematic evolution of ligands by exponential enrichment) method for the selection of a panel of aptamers that specifically recognize leukemia, liver, and lung cancer cells among others [1,2,3,4,5,6,7,8,9]. These aptamers are single-stranded DNA molecules that bind, through their unique secondary structures, to molecular targets on the cancer cell surface. A counter selection strategy is used to collect those DNA sequences that interact with the target cells but not with the control cells. Consequently, this process enriches aptamer candidates that exclusively bind to the target cells.

One such aptamer is sgc8c, which binds to PTK7 [6], a protein tyrosine pseudo-kinase misregulated in numerous cancers, including various colon cancer [10], leukemia [11], and melanoma [12]. PTK7 has been found to play a role in development by regulating planar cell polarity in vertebrates through the Wnt pathway [13], and it is also believed to be important for invasiveness and metastasis in cancer [14,15,16].

Aptamers including sgc8c are easily functionalized to many different surfaces. Throughout the first part of my doctoral work I collaborated with several members of my lab to attach sgc8c to various nanomaterials including dye-doped silica nanoparticles (NPs) [17], liposomes [18], and gold NPs [19] to target PTK7 expressing cells. These

sgc8c targeted materials could improve current diagnostic and therapeutic approaches to treat cancers that express PTK7.

After several intense years of study about PTK7 biology and the selection process for sgc8c conducted throughout my work on aptamer-functionalized nanomaterials, I was perplexed by the observation that several different aptamers selected on different cell lines, by different people, years apart, competed with each other for binding to PTK7. So I created a database of 148 aptamers selected by whole cell-SELEX against various cancer types and performed bioinformatic analysis on them. Curiously, I found that three other aptamers, out of 148, shared sequence identity with sgc8c.

I further found that these four aptamers competed with each other, and that the 14nt sequence that the four different aptamers shared was not only important for the aptamers' binding, but also was the same as a DNA sequence in the 5'-untranslated region (5'-UTR) of a protein also involved in Wnt signaling, DIXDC1b. A fifth aptamer, which also competed with the four other PTK7 aptamers, but which did not share sequence identity with them, shared identity instead with the opposite DNA strand of the DIXDC1b 5'-UTR, adjacent to the region sharing sequence identity with the other aptamers.

This suggested that whole cell-SELEX has not just found an aptamer for an extracellular protein, but may also have identified a genomic DNA sequence the protein binds to naturally. Throughout this dissertation, I will try to develop a new concept, that whole cell-SELEX might be used more generally to identify novel extracellular

transcription factors with highly specific binding motifs. Furthermore, it may shed light on possible therapeutic pathways of current and future ssDNA aptamers.

This Introduction chapter will review, in depth, aptamer selection, and what is currently known about PTK7 and another protein that is important to our story, DIXDC1b. These discussions consist of a comprehensive review of all literature relating to these two proteins. For the casual reader, not interested in all of these details, the review of PTK7 and DIXDC1b in the Introduction can be skipped over, and mainly used as a reference when reading the later chapters.

Chapter 2 will detail the discovery of the DIXDC1b DNA analog to the PTK7 aptamers in the genome. Chapter 3 will try to determine how PTK7 might be interacting with DIXDC1b DNA. Chapter 4 will explore other potential repeatedly selected aptamers identified through bioinformatics. Chapter 5 will show how we used atomic force microscopy (AFM) to determine that sgc8c aptamer-PTK7 protein interactions on live cells have similar forces to α-PTK7 antibody-protein interactions. Chapter 6 will detail the modification of cell surfaces with aptamer that can act as artificial receptors. Finally, Chapter 7 will present some conclusions and detail further experiments that will help clarify and deepen our understanding of the relationship between PTK7 and DIXDC1b.

Aptamer Selection

Molecular aptamers are single-stranded DNAs (ssDNAs) or RNAs, 15-80nt in length, which can recognize target proteins, peptides and other small molecules. The dissociation constants of aptamers to targets can range from 10^{-12} M-10^{-8} M. Aptamers recognize their targets with high specificity, and can typically discriminate between protein targets that are highly homologous or differ by only a few amino acids [20,21]. The secondary structures formed by the ssDNAs are the basis for target protein

recognition [22,23]. These aptamers are selected by a process called SELEX (Systematic Evolution of Ligands by Exponential enrichment), where the aptamers are selected from libraries of random sequences of synthetic DNA or RNA by repetitive binding of these oligonucleotides to the target molecules [24,25,26].

Through this iterative *in vitro* selection process, aptamers with high specificity and affinity to their targets can be obtained. Most of the aptamers reported so far have been selected using pure molecules, such as purified proteins as the targets. Aptamer selection against complex targets (such as red blood cells and single protein on live trypanosomes) was also demonstrated [24,27,28,29,30,31]. Aptamers have several key advantages over antibodies in molecular recognition and imaging, such as low molecular weight, easy and reproducible synthesis, simple modification, fast tissue penetration, low toxicity or immunogenicity, easy storage, high binding affinity and specificity that are very comparable with antibodies [24,29]. Aptamers have shown great promise in molecular recognition, diagnosis and therapy.

To produce probes for molecular analysis of tumor cells, our lab has developed a novel method, the cell-based aptamer selection process (cell-SELEX) for aptamer selection [2,5,6,7,8]. Instead of using a single type of molecule as a target, the cell-SELEX process uses whole cells as targets to select ssDNA aptamers that can distinguish target cells from control cells. In addition to the aptamer advantages mentioned above, a big advantage of the cell-SELEX technology is that there is no need of prior knowledge about potential biomarkers for cancer on these cells. A group of cell-specific aptamers can be selected using a subtraction strategy in a relatively

short period without knowing which target molecules are present on the cell surface. Moreover, the selections may identify important biomarkers.

Compared to 2-D gel electrophoresis and mass spectrometry used for proteomic studies aimed at identifying proteins, cell-SELEX produces molecular probes first, which can then be used to identify the probe's target proteins. Thus, not only can the selected aptamers be used as molecular probes for molecular analysis of cancer, but also they can be used as tools for identifying new biomarkers expressed by tumor cells or other cells, indicating disease status.

The PTK7 Aptamer Sgc8c

Sgc8, was among the first aptamers selected by whole cell-SELEX in the Weihong Tan lab. It was discovered by Dihua Shangguan through selection for the T-cell leukemia cell line, CEM-CCRF but not Ramos, a B-cell leukemia [7]. The full sequence of sgc8 is: ataccagcttattcaatt AGT CAC ACT TAG AGT TCT AAC TGC TGC GCC GCC GGG AAA ATA CTG TAC GGT TA gatagtaagtgcaatct. The lower case, underlined regions, are the primers, and the upper case region is the variable part of the aptamer. This aptamer was interesting because it had a very low Kd of 0.8nM, and bound strongly to the target CEM cells, but not at all to the negative Ramos cells. Furthermore the aptamer showed a clear stem-loop hairpin structure [Figure 1-1], which allowed Shangguan, by progressive optimization, to shorten the 88nt aptamer to a 41nt version ATC TAA CTG CTG CGC CGC CGG GAA AAT ACT GTA CGG TTaga without altering the binding properties. Over time more was discovered about the aptamer: it binds at a physiological relevant temperature, 37°C; it was swiftly internalized into cells [32]; and most importantly it selectively binds a membrane protein tyrosine kinase called PTK7 [6].

While aptamers hold a lot of promise through their size and ease of functionalization, compared to antibodies they are relatively new and untested. Therefore, to show the full potential of aptamers for the diagnosis and treatment of cancer, a model aptamer can be used. Sgc8c is a great model aptamer: it is short; it binds at physiological temperature, with a low K_d; it has a known target, PTK7, which is upregulated in a lot of cancer types; it internalizes which is a major problem for drug delivery; and there are control cell lines, like Ramos, which do not express PTK7. As a result of these properties, sgc8c has been conjugated with different imaging and therapy modalities to specifically target a particular tumor type. Table 1-1 has an extensive list of different dyes and drugs sgc8c has been conjugated with and to what purpose.

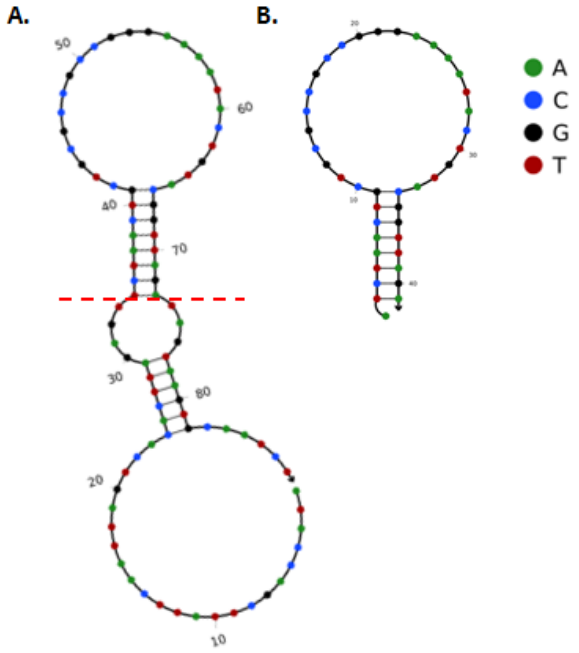


Figure 1-1. A) Full length sgc8 structure. B) Shortened sgc8c structure. Both sequences were modeled by NuPACK [33]. A clear stem-loop hairpin is visible in both. The red dashed line shows where the full length sgc8 was cut to make sgc8c. A: adenine; C: cystosine; G: guanine; T: thymine.

Each of these strategies has taken a non-selective dye or drug and, by attaching the sgc8c aptamer to its surface, made it able to selectively target those cells that express PTK7 on their surface. Throughout this dissertation we will return again and again to why and how this aptamer binds PTK7. To begin, the next section will be an in-depth look at our current understanding of PTK7; its structure, functions, and role in development and cancer.

Table 1-1. The many uses of sgc8c.

Purpose	sgc8c Conjugated Material	Modality
Targeted Detection	Dye [34]	Profiling Patient Blood
	Fluorophore-doped and magnetic silica NPs [35]	Extraction then Fluorescent
	Au NPs [36]	Colorimetric
	Ultrasound microbubbles [37]	Ultrasound
	Different sized AU NPs	Molecular ruler, SET
	Microfluidic device [38,39,40]	Enrichment and capture
	Multi-dye loaded silica NPs [41]	Fluorescence
	Dye doped silica NPs [42]	Flow Cytometer
	Square capillary channels [43]	Capture, microscopy
	Dye-with a dynamic sgc8 structure [44]	*In vivo* fluorescence
Targeted Treatment	Doxorubicin [45]	Chemotherapy
	Doxorubicin loaded liposomes or micelles [18,46]	
	Danunorubicin [47]	
	Acrylamide polymer [48]	
	Au-Ag NR [49]	Photo-thermal therapy
	DNA G-Quadruplex loaded with TMPyP4 [50]	Photodynamic therapy
	Virus particle [51]	Gene delivery
Multi-modal	Hollow magnetic silica NPs loaded with DOX [52]	*In vivo* MRI imaging and chemotherapy
	Au-Ag NRs coated in DNA-nanogel with DOX [53]	Photo-thermal and Chemotherapy

sgc8: anti-PTK7 aptamer; **NP:** nanoparticle; MRI: magnetic resonance imager; **DOX:** doxorubicin; **Au**: gold; **Ag**: silver; **TMPyP4**: 5,10,15,20 tetrakis-(1-methyl-4-pyridyl)-21*H*,23*H*-porphine (TMPyP4)

Biology of PTK7

Protein tyrosine kinase 7 (PTK7) is an ancient protein found as far back in the evolutionary tree of life as the fresh water animal Hydra. While conserved and clearly important—PTK7 mutants lacking the first 114 amino acids (aa) or the cytoplasmic domain are embryonic lethal in mice [13]—its exact purpose and signaling details are still unclear. Since its characterization in 1995 [54,55] various knockout models and mutation experiments have identified a central role for PTK7 in a process called planar cell polarity (PCP) [56,57,58]. PCP causes changes in the underlying architecture of cells by rearranging the cells' actin filaments. Such rearrangements are crucial to early, dynamic developmental events such as gastrulation, neural tube closure, and convergent extension. Animal experiments show PTK7 mutation produces stunted embryos with open spinal cords and other defects [13]. As with many proteins important for development, dysregulation of PTK7 expression is also a factor in many different cancers especially colon cancer [10], leukemia [11], and melanoma [12].

Wnt Signaling Pathway

To understand PTK7's role in PCP requires a basic understanding of the Wnt signaling pathway. Wnt signaling is crucial during embryonic development, and is one of eight major pathways mutated in cancer—the others are AKT, Hedgehog, Death receptor NFκB, TGF-β, Ras, Notch, and GPCRs. The Wnt pathway is actually not just one pathway, but three that share common proteins, but which have vastly different signaling outcomes.

In Wnt signaling, one of nineteen different extracellular Wnt ligands binds to a Frizzled receptor (Fz) on the cell membrane which in turn interacts with one of several different types of co-receptor membrane proteins, such as low-density lipoprotein-

related receptor proteins 5/6 (LRB5/6) or PTK7, to bind a cytoplasmic protein called Disheveled (Dvl). Depending on which of the nineteen Wnts binds Fz, and what co-receptor Fz is associated with, one of three different signaling cascades is activated: the canonical β-catenin/Wnt pathway, or one of two non-canonical Wnt pathways, either the PCP/Wnt pathway, or the Ca^{2+}/Wnt pathway. Please refer to Figure 1-2 for a pictorial description of the components of Wnt signaling important for this discussion.

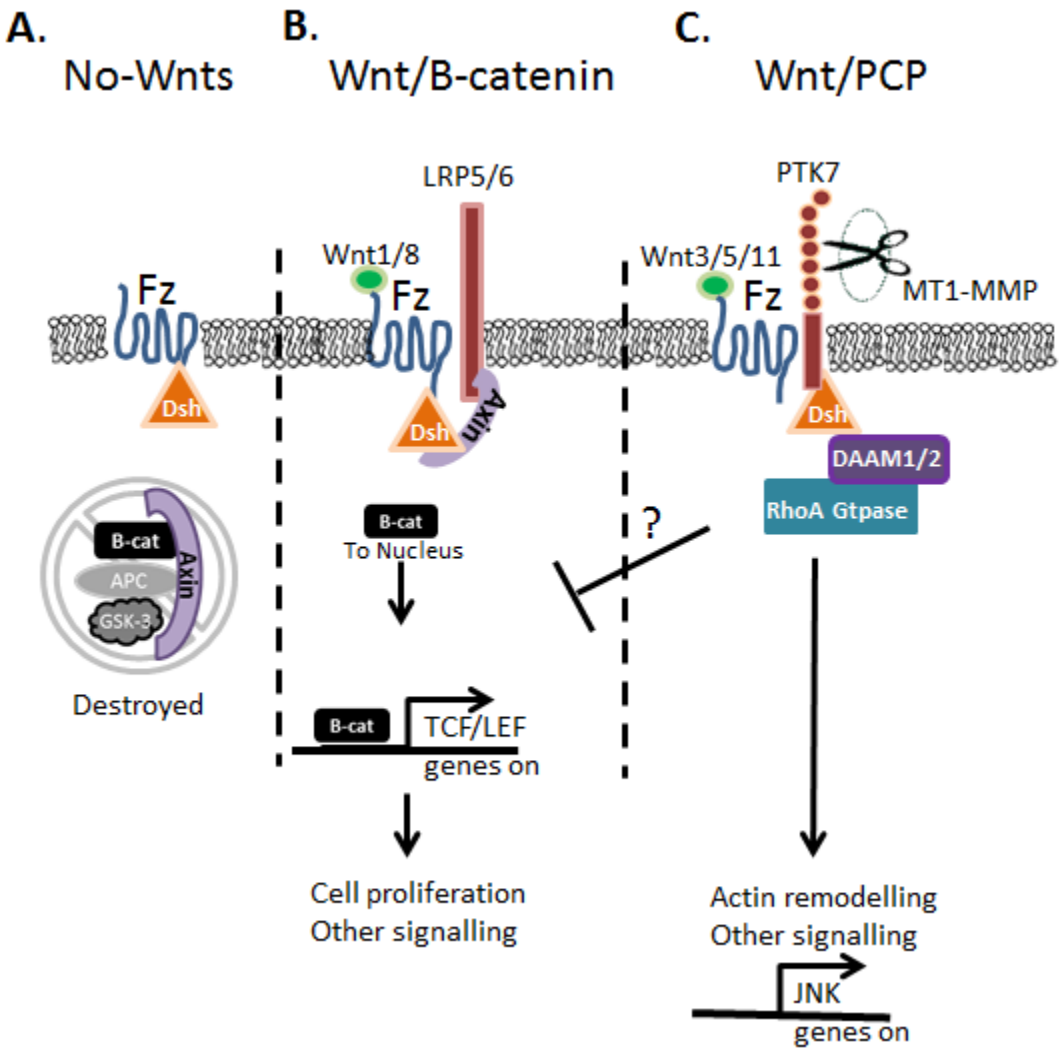


Figure 1-2 Schematic of the canonical and non-canonical Wnt signaling pathway. Please see accompanying text for figure details.

B-Catenin/Wnt Pathway

A mutation considered the initiating event in 85% of colorectal cancers occurs in a protein that participates in β-catenin/Wnt signaling [59,60]. This is not a surprise because the end-target of this signaling cascade occurs when β-catenin arrives in the nucleus and activates a powerful set of genes called the TCF/LEF genes that drive, among other things, cell replication. When no Wnt ligands are present, a complex made up of β-catenin and three other proteins–GSK-3, APC, Axin–are found bound tightly together throughout the cell. If these protein-complexes remain intact, they are continually destroyed, causing no signaling [Figure 1-2a]. However, when a Wnt ligand binds to a Fz receptor that is interacting with LRP5/6 co-receptor, several changes take place and Dvl, which is bound to Fz, and LRP5/6, are then able to bind Axin. The β-catenin complex which was held together by Axin becomes unstable, and β-catenin is released, free to travel to the nucleus and activate the powerful TCF/LEF genes [Figure 1-2b].

Until recently, PTK7 was not considered a player in this β-catenin/Wnt canonical pathway; however, over the last year, two papers have challenged this assumption. The first paper found that PTK7 interacts directly with β-catenin through its intracellular domain, potentially stabilizes it, and allows it to signal [61]. Knocking down PTK7 expression prevents TCF/LEF gene transcription when cells are stimulated with Wnt3a, a ligand that activates β-catenin/Wnt signaling. The second paper asserts that the extracellular domain of PTK7 can interact directly with Wnts to inhibit β-catenin/Wnt signaling [62]. In both studies PTK7, when bound to Fz, was shown to interact with Wnt3a and Wnt8, two ligands known to increase β-catenin signaling. The papers suggest that because Wnt/PTK7/Fz complexes bind Dvl but not Axin, like the

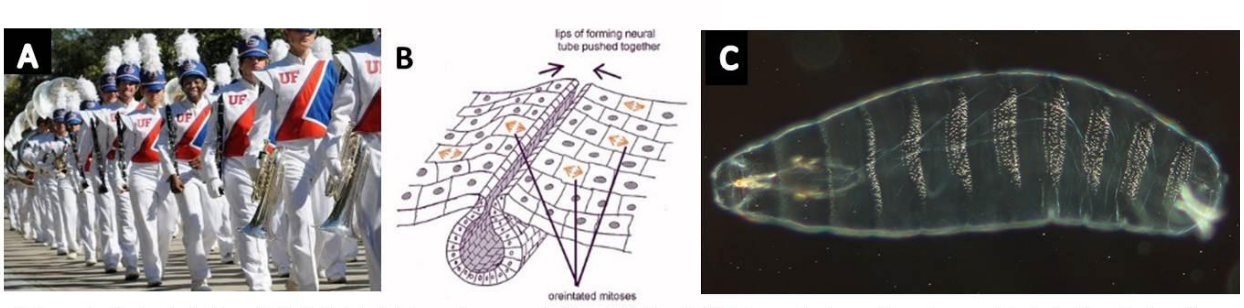
Wnt/LRP5/6/Fz complexes, Axin is left in the cytoplasm to maintain the β-catenin-destruction complex. This leaves less β-catenin free to enter the nucleus and activate TCF/LEF genes. The Wnt/PTK7/Fz/Dvl complexes are free, however, to signal through non-canonical PCP/Wnt signaling. It is this PCP/Wnt signaling that PTK7 has been most closely associated with. The papers conclude that PTK7 can shift signaling from the β-catenin/Wnt pathway to the PCP/Wnt pathway.

PCP/Wnt Pathway

In addition to their roles in β-catenin/Wnt signaling, Dvl and Fz proteins are also major players in PCP/Wnt signaling [Figure 1-2c]. The signaling cascade is very different with respect to PCP/Wnt signaling. For proper development and healthy functioning, cells align in relation to other cells on an xy-plane. For instance, if you look down at the hairs on your arm right now, you will see they are, for the most part, tidily aligned and pointing in the same direction. The same is true for mesenchymal cells growing along the neural crest during embryo development; these cells align their mitotic spindles in the same direction [Figure 1-3b]. This causes dividing cells to pull toward the newly formed neural tube, helping close it. Or more easily seen by researchers, the small hairs in animals' inner ears or the spiny projections on *Drosophila* larva, called denticles [Figure 1-3c], are carefully aligned along their respective xy-planes. PCP signaling is a major driver of this exquisite order, and its main method of action is remodeling actin.

Actin is a major component of a cell’s internal architecture, and changes to it alter cellular shape. PCP acts in a left versus right (or distal versus proximal) direction, perpendicular to the more commonly described apical/basal polarity seen in a typical epithelial cell of the gut, where the top or apical side of the cell faces out into the lumen

and catches nutrients using cilia, and the bottom or basolateral side of the cell is smooth and faces into the cell.



A) Image by Karina Isabel from Karinaisabelphoto.blogspot.com, used with permission. B) "Cellular mechanisms of morphogenesis" adapted from Davies JA (2005) Mechanisms of Morphogenesis (Elsevier) used under creative commons. C) "Drosophila Kutikula" from Wikipedia, used under creative commons.

Figure 1-3. Examples of planar cell polarity. A) UF marching band members aligned along an xy-plane. B) Neural tube closure aided by aligned mitotic spindles. C) *Drosophila* denticles are the white projections on each segment of a 22h old larva, mutations of proteins in the PCP pathway causes perturbations in these denticle patterns.

A number of proteins are involved in an complex, intricate dance to determine which way a cell is oriented in the xy plane, or which "side" of the cell is left and which is right relative to where that cell is positioned compared to others. Five proteins are especially important to development of the protein gradients important to forming this PCP: Flamingo, Fz (which we saw in the β-catenin/Wnt signaling mentioned above), Dvl (same), van Gogh (Vangl1/2), and Prickle (Pk). Flamingo is found on both sides of the cell, while the other proteins migrate in pairs comprised of one transmembrane protein and one cytosolic protein, with Fz and Dvl on one side, and Vangl1/2and Pk on the other. Before an xy-axis is patterned in the cell, these proteins are found mixed throughout the cytoplasm. Over time, in response to input from receptors on other cells, Fz and Dvl localize to one side of the cell and Vangl1/2 and Pk localize to the other, creating a gradient in the cell. The absolute level of Fz on a cell surface is not what determines polarity. Instead, it is differences in the level of Fz between cells that

determines polarity. In an elegant, early experiment on PCP, researchers poured hot wax on *Drosophila* pupae expressing an Fz transgene. The Fz transgene was driven by a temperature-sensitive heat shock promoter, which only made protein in areas where the hot wax fell. Regardless of where they would normally grow, wing hairs invariably grew away from where the wax had turned on Fz activation Please read A. Jennifer or N.A. Paul's reviews to help visualize this aspect of PCP development [63,64].

In addition to these 5 main proteins, there are a number of accessory proteins that appear to modulate PCP signaling. These include Ror, RYK and PTK7. When particular Wnt ligands, especially Wnt5a, Wnt5b, or Wnt11, bind to Fz with a PTK7 co-receptor, PTK7 recruits Dvl, which activates disheveled associated activator of morphogenesis 1 and 2 (DAAM1/2), which in turn activates RhoA GTPase, a key regulator of actin filaments through Rho/ROCK, among numerous other proteins (Figure 1.2c). DAAM1/2 and RhoA GTPase also activate longer-term signaling, including c-Jun N-terminal kinases (JNK) [56] and/or protein kinase C [65], each of which cause changes in cell polarity, migration, and adherence.

In addition to Fz, Dvl, β-catenin, and Wnt3a, PTK7 has been reported to interact with other proteins, including, Plexin-A1 [66], Vangl2 [13], Flt-1 (VEGFR1; vascular endothelial growth factor receptor 1) [67], and a receptor of activated protein kinase C 1 (RACK1) [65]. Recently, Flt-1/PTK7 interactions have been implicated in angiogenesis. In a mouse corneal model, PTK7 overexpression produced more angiogenesis, while siRNA knockdown of PTK7 produced less angiogenesis. The addition of VEGF-A (vascular endothelial factor A) serves to strengthen the interaction between Flt-1 and

PTK7 [67]. RACK1 also interacts with PTK7, and this interaction appears to be important for neural convergent extension in *Xenopus* [65].

PTK7 expression is also linked to apoptosis. When PTK7 is knocked down in HCT116, a colon cancer cell line, apoptosis is increased through the caspase 10 pathway [68]. Likewise increasing PTK7 appears to rescue patient derived acute myeloid leukemia (AML) cells from apoptosis, while decreasing PTK7 led to apoptosis through the effector caspases, 3 and 7 [69].

PTK7 plays important roles in many different cell types from epithelial to endothelial [70] and T-cells [71], and it appears that in different cell types, depending of what sort of proteins PTK7 interacts with, different signaling occurs. There is still a long way to go before the varied web of PTK7's protein interactions and behavior in various cell types is untangled and fully understood. PTK7 generally acts as a switch that turns the Wnt signaling machinery Fz and Dvl away from canonical Wnt/β-catenin signaling toward Wnt/PCP signaling.

PTK7 Structure

Structurally, PTK7 has 7 immunoglobulin-like domains (Ig) that hang outside of the cell, a single transmembrane domain, and a tyrosine pseudokinase domain (pTK) inside the cell. This protein is structurally unique among tyrosine kinases (TKs) for a number of reasons: its extracellular domain is composed of only Ig domains, which in humans and zebrafish contain a matrix metalloprotease (MMP) cleavage site; its transmembrane domain is more conserved than those of any other receptor TK; and its catalytically dead pTK domain, is also inert in all known orthologs [72]. These characteristics indicate a protein with conserved biological function involving all three

protein domains, which arose long ago, in the metazoan radiation, during the Cambrian explosion, 530 million years ago.

Pseudo-Tyrosine Kinase Domain of PTK7

The pTK domain has structural similarities with other tyrosine kinases, but two crucial areas important for phosphorylating tyrosines have been mutated: The GXGXXG motif, which acts as a clamp to help anchor transferable phosphates of ATP, is mutated to ^{803}GXSXXG; and the DFG motif, which chelates a Mg^{2+} ion bridge between the β- and γ-phosphates of ATP, orienting the γ-phosphate for transfer, is mutated to ^{948}ALG [54]. These PTK7 kinase mutations are highly conserved and, while apparently not active, the pTK domain does have other biologically conserved functions [Table 1-2]. Of the 58 TK proteins, 5—PTK7; RYK, another co-receptor for Fz mentioned above; Her3, an EGF receptor family member; and two ephrin receptor members, EphA10 and EphB6—are not functional tyrosine kinases [73].

Table 1-2. PTK7 orthologs in various species. The active kinase motif, DFG, is mutated in all of them.

Organism	PTK7 Ortholog	Mutated DFG Motif
Hydra	Lemon	FLD
Drosophila	Dtrk	YPA
Chicken	KLG	ALS
Human, mouse	PTK7/CCK4	ALG

A major binding partner of PTK7 pTK is Dvl, the cytoplasmic phosphoprotein mentioned previously, which is important in all three Wnt signaling pathways. The first Dvl was originally isolated from a *Drosophila* mutant with unruly wing hair (a symptom of poor PCP), hence the name “disheveled” [74]. Dvl has three domains DIX, PDZ, and DEP. Dvl’s DIX domain is not important for PCP signaling, but is for β-catenin/Wnt signaling as this is where Axin binds [75]. We will revisit Dvl’s DIX domain later in this

work. The DEP domain is important for all Wnt signaling. The PDZ domain is important mainly for PCP, as PTK7 and Dvl interact directly through the PDZ when in a complex with Fz [56].

In order for PTK7 to cause signaling, Dvl must first move from the cytoplasmic aggregates where it is stored, to the membrane where it interacts directly through Dvl's PDZ domain. In a *Xenopus* model transfected with mutant Dvl and wild-type PTK7, removal of either the DIX or DEP domains did not affect Dvl's movement to the cell membrane, but removal of PDZ did [56]. Similarly when the pTK domain of PTK7 was removed, Dvl movement to the membrane was prevented. PTK7 and Dvl cannot, however, interact on their own; they need another protein, like Fz, to aid their interaction [56].

Transmembrane Domain of PTK7

Many transmembrane proteins thread through the membrane several times, and have conserved transmembrane domains. By contrast, TKs thread through the membrane only once, and rarely have a highly-conserved transmembrane domain. PTK7 has the most-conserved transmembrane domain among TKs. It has a predicted 22aa helical structure, which is identical to chicken KLG, and it shares 55% homology with Hydra *Lemon* and Drosophila Dtrk, other PTK7 orthologs. The PTK7 transmembrane domain includes a GxxxG motif, which in some proteins promotes interactions between transmembrane helices [72]. These types of interactions, especially in TKs, promote signaling. However, studies where the transmembrane domain was made into a bacterial fusion protein found no increase in helix-helix interactions, indicating some other, yet unknown, method of action allows PTK7 to form homodimers and heterodimers with Fz [76].

Extracellular PTK7 and its MT-MMP1 Cleavage Site

The extracellular portion of PTK7 is composed of seven Ig-like domains, held together by an Ig fold. An Ig fold is where two β-sheets interact by a single disulfide cysteine bond. The Ig fold is most commonly associated with antibodies, but is also found in many proteins that encourage cell-cell interactions, as well a number of critical transcription factors such as p53, STAT, and NFκB [77]. Statistical analysis of Ig folds across receptor TKs, found that PTK7 evolved very early in receptor TK development [78]. To date, no ligand has been found that binds to PTK7, but the extracellular portion of the protein is clearly important for signaling.

The ST Lee group from the University of Korea, one of the groups who first cloned PTK7 mRNA, found that addition of a soluble fragment of PTK7 made of just the 6th and 7th Ig domains (~200aa) had profound effects on cancer-related signaling as seen by reduced VEGF-induced tube formation, wound healing, migration, invasion, and angiogenesis in HeLa [70]. This showed the fragment was competing with normal PTK7 for signaling. This result was repeated in blood cells from patient's with AML using the full length extracellular PTK7 which also caused an increase in the apoptosis [69].

Several years after this finding, the AY Strongin group from Stanford University might have discovered an explanation. They found an MT1-MMP cleavage site in the 7th Ig fold, above the transmembrane domain [79]. MT1-MMP, which is also known as MMP-14, is a membrane-bound endopeptidase that cleaves extracellular matrix (ECM) proteins; other soluble MMPs, activating them; and, other membrane-bound signaling receptors, including, PTK7. It is known to be overexpressed in cancer, with a higher level of expression yielding more invasive and malignant tumors [14,15,16]. Previous

studies had linked Wnt/PCP signaling and PTK7 in particular with MT1-MMP expression [79,80]. They confirmed that MT1-MMP cleaves PTK7 PKP^{621}↓LI into two pieces: a 50kDa C-Terminal membrane-bound pTK fragment and a 70kDa soluble PTK7 (sPTK7) fragment made up of the first 6 Ig-like domains. They found these sPTK7 fragments in the cell media and bound to full-length PTK7 on the surface of the cells. Furthermore, their mutant, L622D, which is resistant to MT1-MMP cleavage, could not cause downstream actin remodeling events. Finally, overexpression of sPTK7 in HT1080 primate fibrosarcoma cells did not affect actin remodeling, but did cause a large spike in RhoA GTPase expression, which is the upstream signaling molecule of actin remodeling that is activated by PTK7. Furthermore, although not explicitly stated in their conclusions, they found that MT1-MMP treatment of MDCK canine kidney cells caused PTK7 to move from the cell membrane and accumulate around the cell nucleus.

The AY Strongin group followed up on this work by characterizing a mutant PTK7 protein with not one, but two MT1-MMP cleavage sites. This mutant PTK7 protein was made in a mouse strain, *chuzhoi* (*chz*), created by exposing the mice to the mutagen N-ethyl-N-nitrosourea [57]. The mutant mice showed signs of classic PCP signaling problems, including neural tube defects and disrupted hairs in the inner ear. In addition to the normal MT1-MMP cleavage site in the 7th Ig-like domain, these mutant's PTK7 contained a second cleavage site in the linker between the 5th and 6th Ig-like domain where $QVLEKLK^{503}$ has been mutated to QVL**ANP**EK↓LK^{503} by the addition of **ANP**.

In experiments comparing the mutant to wild-type PTK7 in HT1080 cells, when MT1-MMP1 was added to wild-type cells, PTK7 migrated from the cell surface to surrounding the nucleus, and when sPTK7 was ectopically added to the cells, RhoA

GTPase expression rose. By contrast, when MT1-MMP1 was added to the *chz* mutant PTK7 was not able to localize to the membrane and the addition of sPTK7 caused no change in RhoA GTPase expression. However, it did cause an increase in the cells' invasion ability.

PTK7 Gene Structure and Splicing

PTK7 is located on chromosome 6 at 6p21.1, and has an 883 base pair (bp) promoter on the 5’ end of the gene that drives its expression. This promoter lacks both a TATA box and CAAT box, and, instead, has a CpG island with several GC box/SP1 binding motifs [81]. This promoter resembles those of housekeeping genes, which are constitutively expressed in most cells, and which are essential for general cell maintenance. It contains TCF/LEF-, SP1- and 5 bHLH-binding sites [10]. The TCF/LEF genes are turned on by the Wnt/β-catenin pathway activation, dysregulated in colon cancer.

There is one main PTK7 variant, which has been the focus of our discussion; it has 1070aa stitched together from 20 exons. There are, however, four other PTK7 splice variants expressed at lower levels shown to have different expression levels in cancer cell lines. These variants lack various portions of the protein. PTK7-2 is missing the N-terminal half of the 6th Ig (Exon 10). PTK7-3 is missing the 5th and N-terminal half of the 6th Ig (Exons 8-10) and has the highest mRNA expression levels next to the major PTK7. PTK7-4 is missing the 7th Ig (portion of Exon 12, and Exon 13)—this variant retains the MT1-MMP cleavage site, but removes a 55aa piece 5aa toward the C-terminal of that site. PTK7-4 is truncated due to a frame-shift mutation, which keeps the transmembrane domain, but which removes most of the pTK domain, and is only highly expressed in the testis. These splice variants may be interesting because in other

receptor TKs, splice variants exhibit different ligand binding affinities, as with FGFR2(IIIb) and FGFR2(IIIc), or form dominant negative heterodimers between major and minor splice variants, as with TrkB.

sgc8c Aptamer

In their paper [82], the A.D. Ellington group at the University of Texas was surprised that the anti-PTK7 aptamer, sgc8c, bound to so many different types of cancer cell lines, including colon, leukemia, cervical, and breast cell lines. They conclude that this might be because PTK7 is a marker for cell adherence to culture dishes and not relevant. However, based on our review of the literature and PTK7's role in Wnt signaling, we contest the A.D. Ellington group's conclusion that sgc8c binds to many different types of cancer cells lines because PTK7 is a marker for cell adherence. Instead, we argue that sgc8c binds to many cancer cell lines because PTK7 is overexpressed in many different cancer types.

DIXDC1 Biology

Although PTK7 acts as a switch that shunts Wnt signaling toward PCP pathways, there is another protein that has the opposite effect and will become important to our discussion: DIX Domain Containing 1 protein (DIXDC1) is also known as coiled-coil domain 1 protein (Ccd1) in zebrafish and mice, or KIAA1735. There are three known proteins with a DIX domain: Dvl and Axin, which we have previously mentioned; and DIXDC1, which we will detail here. First identified in humans in 2000 [83], DIXDC1 is the least studied of the three DIX domain-containing proteins. It is becoming clear, however, that it might act as a switch that enhances Wnt/β-catenin activity while inhibiting Wnt/PCP activity. For instance, DIXDC1 has been implicated in cancers with dysregulated Wnt signaling. Overexpression of DIXDC1 leads to survival of human

colorectal and lung adenocarcinomas through Wnt/β-catenin crosstalk with the PI3K/AKT pathway. By contrast, downregulation of DIXDC1, as in squamous cell carcinoma of the lung, leads to aberrant upregulation of Wnt/PCP signaling [84].

DIXDC1 Increases Wnt/β-catenin Signaling

DIXDC1 is spatially coexpressed with other Wnt signaling proteins in developing zebrafish and mouse embryos [85,86]. In embryos, overexpression of Ccd1, the homolog of DIXDC1 in zebrafish leads to a reduction in eyes and forebrain; a similar phenotype to Wnt8 overexpression, a ligand that causes Wnt/β-catenin signaling. Addition of a dominant negative Ccd1, has the opposite phenotype, more similar to Axin overexpression, and tamps down the Wnt/β-catenin signaling. Furthermore, Wnt8 overexpression rescues dominant negative DIXDC1 treatment. These *in vivo* results imply that DIXDC1 acts upstream of Axin but downstream of Wnt ligands [87].

At the molecular level also, DIXDC1 is a positive regulator of canonical and a downregulator of non-canonical Wnt signaling. The major downstream effect of canonical Wnt/β-catenin is expression of the TCF/LEF genes. siRNA knockdown of DIXDC1 suppresses Wnt3a TCF/LEF signaling, while upregulating DIXDC1 increases Wnt activation by the same amount as Dvl upregulation alone. Coexpression of both Dvl and DIXDC1 act synergistically to greatly increase TCF/LEF [88]. Furthermore, DIXDC1 levels were increased after Wnt-3a stimulation, but it was less phosphorylated. While no change in DIXDC1 mRNA was found, activation of canonical Wnt signaling decreased ubiquitin-dependent degradation of ectopic and endogenous DIXDC1. Wnt/β-catenin signaling might upregulate DIXDC1 through a post-translation mechanism by inhibiting degradation of the DIXDC1 protein [89].

One major way DIXDC1 affects Wnt signaling is by controlling the amount of bioavailable Dvl protein found in cells. On its own, endogenous Dvl forms large homomeric assemblies, where Dvl's DIX domains interact with each other and are found in discrete aggregates, called puncta, throughout the cell. In these large aggregates, Dvl is unavailable for signaling, because they need to be broken up into smaller assemblies that can travel to the cell membrane and bind with receptors such as Fz or PTK7.

By contrast, Dvl-DIXDC1 hetero-assemblies also occur through DIX domain interactions, but the hetero-assemblies are much smaller than homo-assemblies of Dvl-Dvl. Instead of having high molecular weight aggregates, Dvl-DIXDC1 hetero-assemblies typically have 3 Dvl and 1 DIXDC1 protein. This difference in oligomer size is due to the replacement of an important hydrophobic region of the Dvl's DIX domain by a hydrophilic histidine residue in DIXDC1 [88]. In fact, expression of DIXDC1's DIX domain alone is able to break up Dvl puncta in cells. Upon Wnt activation, increases of DIXDC1 occur in the in cytosol where they are recruited to Dvl aggregates, depolymerizing them into the active trimer. Thus, one way DIXDC1 controls Wnt activity is by making Dvl available for signaling [88].

DIXDC1 Decreases Wnt/PCP Signaling

DIXDC1 increases canonical Wnt signaling, but decreases non-canonical Wnt signaling by modulating JNK signaling in several different ways [90]. As mentioned in the above discussion of PTK7, JNK activation through PTK7-Dvl interactions is important for changes in PCP, including cell polarity, migration, and adherence. DIXDC1 inhibits JNK activation through independent interactions with both Axin and Dvl.

Axin activates JNK through the Mammalian Mitogen-activated protein Kinase Kinase Kinase 1 and 4 (MEKK1 and MEKK4), which are serine/threonine kinases that phosphorylate many targets in response to growth factors. DIXDC1's DIX domain binds to a region of the Axin protein, called PP2A, located 30aa from the N-terminal of Axin's DIX domain. This is the same region through which Dvl and Axin interact [91]. When DIXDC1 binds to Axin, the DIXDC1-Axin complex cannot bind to MEKK1, and thus doesn't cause JNK activation. DIXDC1 can also independently bind to MEKK4. When this occurs, MEKK4 is physically sequestered by DIXDC1 and cannot bind to Axin, preventing JNK signaling [90].

While Axin signals through MEKK1/4, as mentioned above, Dvl non-canonical Wnt signaling occurs when Fz interacts with its co-receptor PTK7, which in turn recruits Dvl, causing activation of DAAM1/2 and signaling through RhoA GTPase. While the details are not clear, when Dvl and DIXDC1 are bound together they can inhibit the downstream activation of JNK. This inhibition requires the coiled-coil domain, which is called the central myosin tail homology domain (MTH) of DIXDC1. DIXDC1 mutants missing MTH domains do not inhibit JNK [90]. Coiled-coil domains common in centrosomal proteins and DIXDC1 have been shown to co-localize with gamma tubulin during interphase and mitotic phase in HEK293, suggesting a further role for DIXDC1 in chromosome segregation [92].

DIXDC1 Gene Structure

Human DIXDC1 is found at 11q23.1, located between microsatellite markers D11S927 and D11S1347, in a region commonly deleted in sporadic breast cancer. Other genes in this region include KIAA1391, BTG/TOb-family-protein BTG4, and SIK2 (salt-inducible serine/threonine kinase 2) [93].

Human DIXDC1 has two different transcription initiation sites yielding two isoforms, a long form and short form: DIXDC1a and DIXDC1b, respectively. DIXDC1b is predominantly expressed during development and in the brain [83]. During adulthood, DIXDC1a is the more prevalent isoform [94]. Both isoforms have an MTH and a C-terminal DIX domain. In humans the MTH region contains 2 tyrosine phosphorylation sites and a leucine zipper motif. The longer form DIXDC1a is 683aa coded by a 6.3kb mRNA. This isoform has an N-terminal extension containing a calponin homology (CH) domain. The long form, DIXDC1a, but not the short form, DIXDC1b, directly binds to actin, with one DIXDC1a binding every 200 actin monomers, localizing it to focal adhesions at the tips of the cell. DIXDC1a binding to actin requires amino acids 127 to 300 of DIXDC1a; however, it does not require the DIX or MTH domains [94].

The shorter DIXDC1b is 472aa coded by a 5.1kb mRNA, and has a separate regulatory domain from DIXDC1a. Unlike DIXDC1a, DIXDC1b does not bind to actin, and is found diffusely throughout the cytoplasm. In humans, the DIXDC1b gene consists of 16 exons and is about 45kb. The gene is linked to dihydrolipoamide S-acetyltransferase (DLAT) in a tail to head manner, with less than 4kb between them [95]. DIXDC1b is highly conserved in vertebrates. Zebrafish only have the shorter DIXDC1b ortholog, called Ccd1 which is 54% conserved with humans [87]. The mouse gene, by contrast, has 3 main isoforms of DIXDC1, and the isoform most similar to DIXDC1b (Ccd1B) is 90.9% similar to the human gene [96]. Mouse DIXDC1 consists of 25 exons over 77kb in chromosome 8. All three mouse isoforms of DIXDC1—Ccd1A, Ccd1B, and Ccd1C—combine with Dvl, and all act along the same pathway to cause

TCF/LEF gene transcription. However, only independent transfection of Ccd1B, and not of Ccd1A or of Ccd1C, caused activation of TCF/LEF genes.

In the following chapters, I will endeavor to uncover more about PTK7 and DIXDC1b through the study of aptamers.

CHAPTER 2
PTK7 APTAMERS HAVE AN ANALOG IN GENOMIC DNA

Introduction

Cell-SELEX generates artificial DNA and RNA molecules (aptamers) that bind to biological targets of interest with applications for research and therapy. Each whole cell-SELEX uses a large library of random DNA sequences (~10^{15}), amplified by unique primers, against cells with more than 10,000 unique protein targets [97]. Using this tool, our lab has selected a number of DNA aptamers that bind to leukemia [7], breast [98], and colon [4] cancer cell lines .

In the process of cataloging colon cancer aptamers he selected, Kwame Sefah noticed several aptamers had the same binding profile as sgc8c. These three aptamers—KC2D8, KC2D4, and KDED19—all bound the same cell types with similar affinities, and competed with each other for binding [4]. I was intrigued by Kwame Sefah's finding, and looked to see if any of these aptamers was similar in sequence to sgc8c. To my surprise CLUSTALx alignment showed sgc8c and KC2D8 share 38 contiguous nucleotides (nt).

As previously discussed, sgc8c was selected by whole-cell SELEX for a target T-Cell leukemia cell line, CEM-CCRF, and against a B-Cell leukemia cell line, Ramos [7]. The second aptamer, KC2D8, was selected by a Kwame Sefah several years later using a colon cancer cell line, DLD1, as the target, and no negative cell line [4]. These two selections used two different primer pairs to PCR-amplify the portion of the library that binds to cells after washing [Table 2-1]. The sgc8 aptamer cannot prime the KC2D8 library, making the likelihood of contamination by sgc8c during KC2D8

selection remote. Starting with this initial observation, I was curious how and investigated why these two selections identified a nearly identical aptamer.

Results and Discussion

38nt sgc8c Binds as well as 41nt sgc8c

Sgc8c is the most commonly used aptamer in the Weihong Tan lab. It was one of the first aptamers selected by the lab. It started as an 88nt aptamer and was optimized by Dihua Shangguan to a shorter 41nt aptamer. As KC2D8 and sgc8c shared only 38nt of 41nt in common, I predicted these extra bases would be superfluous, and that an aptamer consisting of only shared bases would have the same binding ability as the 41nt aptamer.

I synthesized different aptamers that incorporated different mixtures of the 3 bases not found in common between the two sequences, and found the different sequences derived from sgc8c and varying in length from 38-41nt all had comparable K_ds [Table 2-1, Figure 2-1], indicating the three bases not shared between KC2D8 and sgc8c were not necessary for binding. In the future, all labs should save resources, and use the 38nt sgc8c. Shortening the aptamer to 38nt would use 7% less resources to synthesize, but would not sacrifice aptamer quality.

Table 2-1. 31nt sgc8c binds as well as 41nt sgc8c. Values averaged from two experiments with three replicates each. Bold, underlined bases are variable, depending on the sequence.

Name	Sequence	K_d (nM)
sgc8-41	**AT**CTAACTGCTGCGCCGCCGGGAAAATACTGTACGGTTAG**A**	0.8 ± 0.1
sgc8-38	CTAACTGCTGCGCCGCCGGGAAAATACTGTACGGTTAG	0.9 ± 0.9
sgc8x	CTAACTGCTGCGCCGCCGGGAAAATACTGTACGGTTAG**A**	1.4 ± 0.4
xsgc8x	**T**CTAACTGCTGCGCCGCCGGGAAAATACTGTACGGTTAG**A**	0.5 ± 0.1
xxsgc8	**AT**CTAACTGCTGCGCCGCCGGGAAAATACTGTACGGTTAG	0.5 ± 0.1

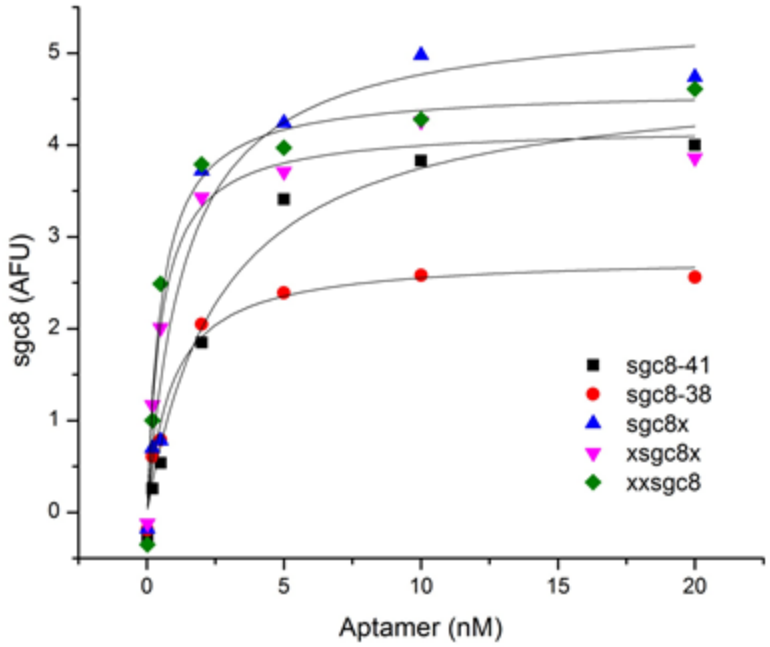


Figure 2-1. Hyperbolic fitting of data collected at different aptamer concentrations. Data is based on triplicate measurements of CEM cells. For each fitting $R^2 > 0.96$.

Bioinformatic Analysis of cell-SELEX Aptamers

Looking beyond saving resources by shortening the sgc8 sequence, the finding that two aptamers, selected by different researchers, on different cell lines, years apart, shared 38nt puzzled us. The selections that produced sgc8c and KC2D8 used initial libraries with different primers. Also, each selection started from a large pool of potential targets because, neglecting lipids and peripheral proteins and considering integral proteins alone, 20-30% of human proteins have a transmembrane domain [99], and a significant portion of these proteins are exported to the cell surface. This means upwards of 10,000 different proteins are expressed on each cell's surface. Moreover, each of these proteins has more than one unique domain serving as a potential aptamer target, much as there are typically several antibodies or aptamers that recognize different regions called epitopes on the same protein. These proteins can be further modified by carbohydrates, creating many more sites for possible aptamer interaction

and subsequent selection. An ovarian cancer selection that produced aptamers insensitive to proteases has been postulated to bind these glycoproteins [9].

Furthermore, aptamers are routinely selected against a wide array of single targets from small ions, to peptides, and purified proteins. A search of the literature did not uncover a single target that evaded selection. Building on this, we assumed that each potential target on the cell surface had an equal probability of binding one of the sequences in the library, such that a selection would agnostically find targets [100]. If this was the case, whole-cell SELEX would select for a panel of aptamers that profile the entire range of proteins, carbohydrates, and lipids on a cell surface.

Thus, the odds for repeat selection are remote because: (1) a large aptamer library is used for each selection, calculated to include 10^{15} unique sequences; (2) there are a large pool of potential targets on each cell membrane surface; (3) and we assumed each potential target on the cell surface has an equal probability of binding one of the sequences in the library.

The similarity between sgc8 and KC2D8 indicated one of these reasons did not hold. The first two were not in question; leading us to postulate the last assumption could be faulty. Instead of agnostically profiling the cell surface, perhaps cell-SELEX could be preferentially selecting for aptamers with a biological function; for instance, those with an underlying affinity for DNA binding.

Next I wondered if any other complex target selections had also found the same target and shared this common sgc8c-KC2D8 motif. I compiled a dataset of all ssDNA aptamers selected by complex selection or against membrane-bound targets (Appendix A). This yielded 148 unique aptamer sequences from 33 different selections. The

aptamers' primer regions were removed to avoid interference during alignment, and the identity between them was determined by alignment with ESPRIT [101].

Using this model, we made a set of identities with corresponding p-values. From the results, two further aptamers, H01 and KMF9b, which had sequence identity to each other and to sgc8, were identified (Appendix B).

When these sequences were aligned with sgc8c, their similarities were clustered around a core 15nt GC rich region: GCTGCGCCGCCGGGA [Table 2-2]. As predicted, all four aptamers--KC2D8, HO1, KMF9b, and sgc8c--competed with each other but not with a control, scrambled sgc8c sequence (Scr-sgc8c) [Figure 2-2].

Table 2-2 Aptamer sequences showing sequence similarity. Black with white lettering shows consensus region. Note sequences have been shortened for clarity. Please refer to experimental section for full sequences.

Aptamer	Sequence	Cell Target	Primers Forward (F), Reverse (R)
sgc8c [7]	GGCGCAGC**CTGCGCCGCCGGGAG**CCTCCCTCCCAGTGGGAGATG	CEM (+) Ramos (-)	F-ATACCAGCTTATTCAATT R-GATAGTAAGTGCAATCT
KC2D8 [4]	ATCTAACTG**CTGCGCCGCCGGGA**AAATACTGTACGGTTAGA	DLD1 (+)	F-ATCGTCCGCCACCACTACTC R-GTGAGACTGCCTGCCGATGT
KMF9b [102]	AGCGCAGCAG**CTG**T**GCC**A**CCGGGAG**AATTTACGTACGGCTGAGCGA	MCF7-10-AT-1 (+)	F-AGGCGGCAGTGTCAGAGT R-CTGAGCGACGAAGACCCC
H01 [103]	AAGCAGCAG**CTG**T**GCC**AT**CGGG**TTCGGATTTTCTTCCTACGACT	CEM (+)	F-ATCGTCTGCTCCGTCCAATAT R-TTTGGTGTGAGGTCGTGC

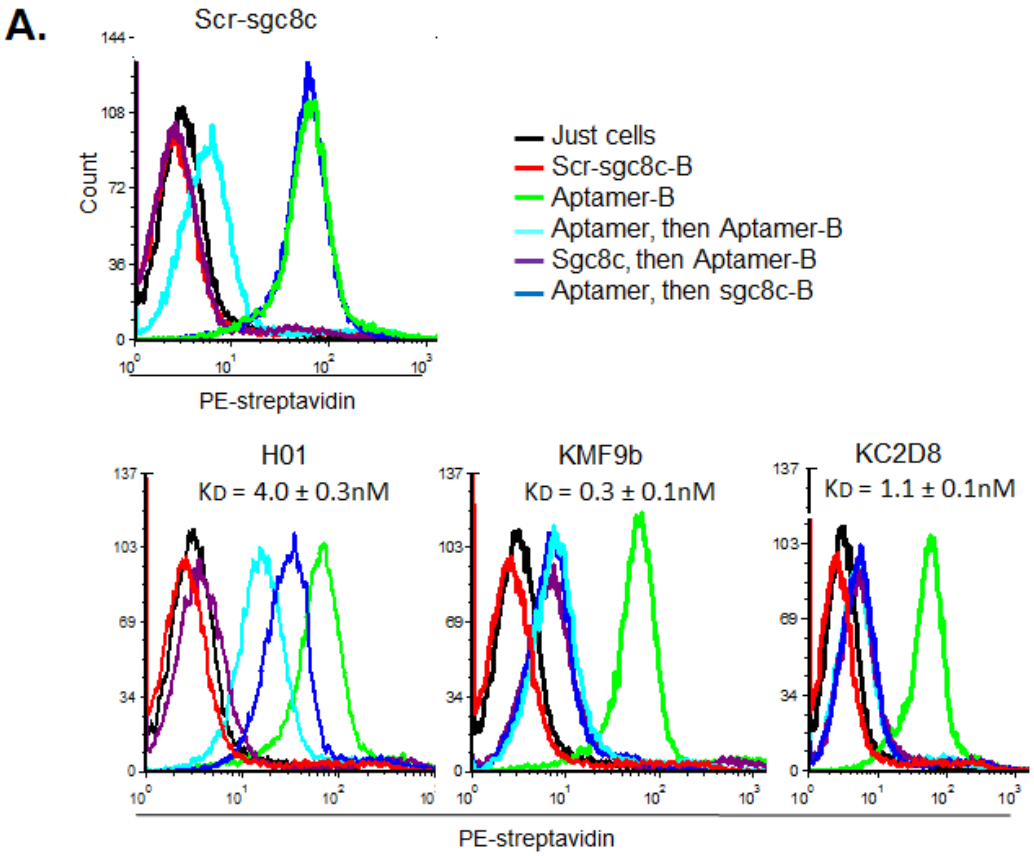


Figure 2-2. A) Competition studies between sgc8c and sgc8c, H01, KMF9b, and KC2D8. Scr-sgc8c: Scrambled sgc8c, a negative control. Apparent K_d for each aptamer on CEM cells is given under the aptamer's name.

From the competition experiments [Figure 2-2a], it was clear that H01 was competed off of the cell surface faster than the other aptamers; because addition of sgc8c first, then H01-B (purple line) reduced the binding to background levels, but addition of H01, then sgc8-B (dark blue line) or H01 then H01-B (light blue line) reduced binding, but at a much lower level. This means, if unlabeled sgc8c binds PTK7 first, H01 is not able to bind the PTK7. By contrast, H01 is readily replaced by both biotinylated sgc8c and 10x unlabeled H01. To investigate the off rate, a competition experiment where biotinylated aptamer was bound, washed off, and replaced by 10x unlabeled aptamer for varying periods of time was performed. After 120min the signal for H01

decreased by 75% compared to less than 15% for each of the other aptamers, showing H01 has a faster off-rate than the other PTK7 aptamers.

Statistical Analysis of Consensus Region

To determine whether the sequence identity between the aptamers could occur by chance, Yunpeng Cai, a bioinformatician, and I developed a bioinformatic simulation to determine the probability that four aptamer sequences, selected at random, would have the same level of identity. We performed 1 million random simulations to determine the probability that a random pair of sequences would have the same probability value, *p,* as the alignment between sgc8c and KC2D8. The overall probability of coincidence is P=$n(n-1)/2*p$, where n is 1 million. With these simulations, we confirmed that the pair sgc8c-KC2D8 is a confidently non-coincident match, with significance level $P_{sgc8\text{-}KC2D8}$ <0.01. Next, we did a random simulation with the third sequence, KMF9b. The overall probability between these three sequences is $P=2*(1-P_{sgc8\text{-}KC2D8})*(n-2)*p_{KMF9b}$. We found KMF9b is matched to the pair sgc8c-KC2D8, with significance level P_{KMF9b} <0.03. Similarly, sequence H01 is matched to the above triplet at significance level P_{H01} <0.05. Thus, we concluded that *each* of the 4 sequences matching with the others cannot be explained by coincidence.

Consensus Region is Important to Aptamer Binding

Although we found a statistically significant, conserved region between these four aptamers, we did not know if this region was relevant to how the aptamer bound PTK7. To determine if the consensus sequence had a larger influence on binding than sequences outside this region, we performed mutational, blocking, and structural probing of one of these PTK7 aptamers, sgc8c.

Mutational analysis

Azobenzene-phosphoramidites are DNA nucleoside analogs that have an azobenzene in place of the standard A, T, G, or C base [Figure 2-3]. Mutating aptamers with these bases perturbs the aptamers' secondary structure immediately surrounding the mutation, potentially having an effect on binding. I hypothesized that aptamer mutants with affected binding, determined by K_d, would have mutations in key aptamer-protein interacting nucleotides.

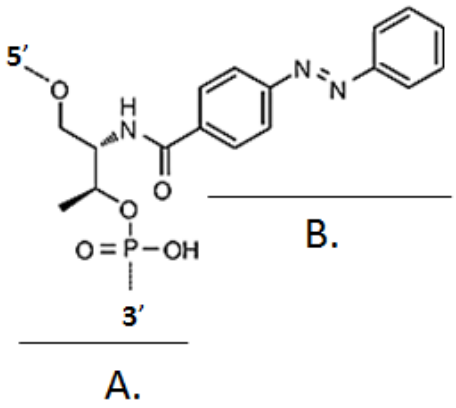


Figure 2-3. Azobenzene nucleotide base-analog. A) Phosphodiester backbone of the nucleoside. B) Azobenzene conjugated in place of a standard base.

Under the aegis of Joseph Phillips and Haipeng Liu, we synthesized a number of sgc8c sequences with azobenzene mutations at different locations, and determined their K_d with CEM cells. Sgc8c mutants with azobenzene bases inserted within the consensus region, the region shared by the aptamers, but not those with azobenzene mutations inserted outside that region, had significantly decreased aptamer binding ability, as measured by their K_d [104] [Table 2-3]. This indicated that the consensus region is more important than other aptamer regions, including the stem and loop of sgc8c, for binding to CEM cells.

Table 2-3 Azobenzene mutated sgc8c sequences. Black with white lettering shows consensus region. Z indicates an azobenzene insertion. One-way ANOVA on K_d between azobenzene "outside" group and "within consensus sequence" group gave a p-value < 0.05.

sgc8c Mutations	Sequence with Azobenzene Mutations (Z)	K_d (nM)
No Azobenzene	ATCTAACTGCTGCGCCGCCGGGAAAATACTGTACGGTTAGA	2
Azobenzene(Z)outside	AZTCTAACTGCTGCGCCGCCGGGAAAATACTGTACGGTTAGA	4
consensus sequence	ATCTAZACTGCTGCGCCGCCGGGAAAATACTGTACGGTTAGA	5
	ATCTAACTGCTGCGCCGCCGGGAAZAATACTGTACGGTTAGA	2
	ATCTAACTGCTGCGCCGCCGGGAAAATACTGTACZGGTTAGA	132
Azobenzene (Z) within consensus sequence	ATCTAACTGCTGZCGCCGCCGGGAAAATACTGTACGGTTAGA	1496
	ATCTAACTGCTGCGZCCGCCGGGAAAATACTGTACGGTTAGA	2429
	ATCTAACTGCTGCGCCGCCZGGGAAAATACTGTACGGTTAGA	523

Consensus sequence blocking

To determine the region of sgc8c most important for binding in another way, a visiting scholar, Jian Wang, and I, blocked the sgc8c consensus sequence with different lengths of complementary DNA (cDNA). Aptamers are ssDNA; addition of a complementary sequence to the aptamer sequence causes the aptamer and its complementary sequence to hybridize. This hybridized DNA does not have the same secondary structure as the single stranded aptamer. In order to bind a target, the aptamer binding confirmation must be more energetically favorable than the hybridized confirmation. The aptamer must be able bind in spite of having a piece of it hybridized, altering its secondary structure.

If an important region for aptamer binding is blocked by cDNA, then it cannot form the proper secondary structure and bind its target. If, however, the cDNA blocks a region unimportant for binding, then the remaining unblocked aptamer can form enough proper secondary structure to act as a toe-hold and displace the blocking cDNA from the aptamer [105]. Therefore, we made different lengths of cDNA, complementary to the

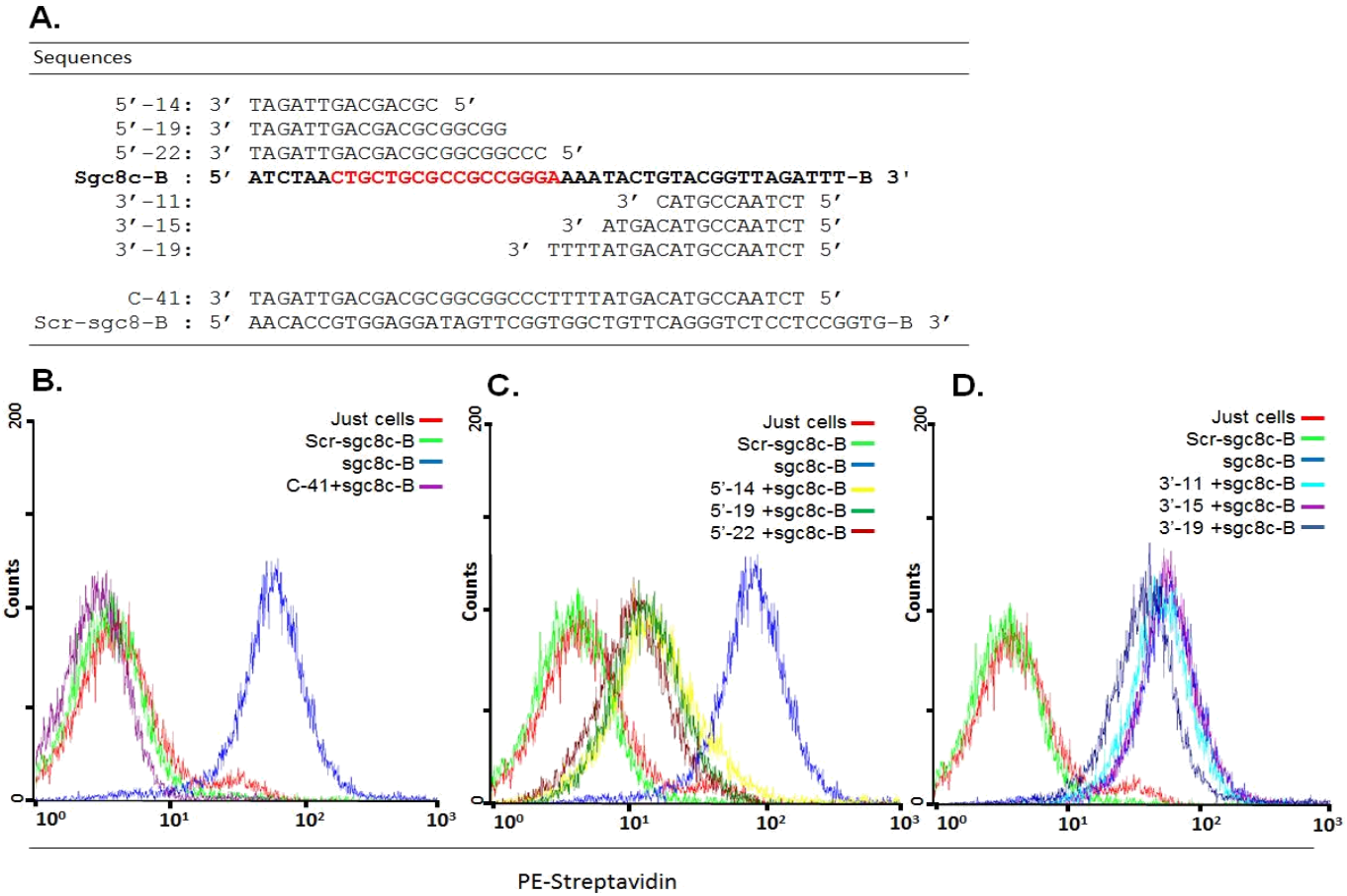


Figure 2-4. Blocking consensus region with complementary DNA, blocks binding. A) Sequences used in these experiments. All complementary sequences are aligned with the bold-faced sgc8 sequence they hybridize to. They are named 5' or 3' depending on which end of sgc8c they bind. The number after the name of each sequence is its length. The putative consensus sequence in sgc8c is in red. B: biotin. C-41: complementary to the entire sgc8c sequence, a positive control. Scr-sgc8c: scrambled sgc8c, a negative control. B) C-41, shifts binding all the way back. C) Complementary sequences that block the consensus sequence shift sgc8c binding back (5'-14, 19, and 22). D) Complementary sequences that do not block the consensus sequence do not shift sgc8c binding back (3'-11, 15, and 19). A slight shift backward is seen in 3'-19 which blocks one nucleotide of the consensus sequence.

3’ or the 5’ end of sgc8c, and measured the blocked aptamers' binding ability by flow cytometry [Figure 2-4].

The positive control was a strand entirely complementary to sgc8c, C-41. The sgc8 that was hybridized with C-41 shifted aptamer binding all the way back to the level of the negative control, scr-sgc8c [Figure 2-4b]. Blocking only the 5’ end of sgc8c with 14nt, 19nt, or 22nt cDNA resulted in a 10 fold decrease in aptamer binding [Figure 2-4c]. These 5’ cDNAs block the consensus sequence. However, blocking the 3’ end of sgc8c with 11nt, 15nt, or 19nt did not cause a shift. Of the 3’ blocking-cDNAs, only 3’-19 showed a small shift backward [Figure 2-4d]. This 3’-19 cDNA blocks 1nt of the consensus sequence. This indicated that the 5’ end, containing the consensus sequence, was the most important region for binding.

Sgc8c pyrene mutants suggest sgc8c interacts intimately with PTK7

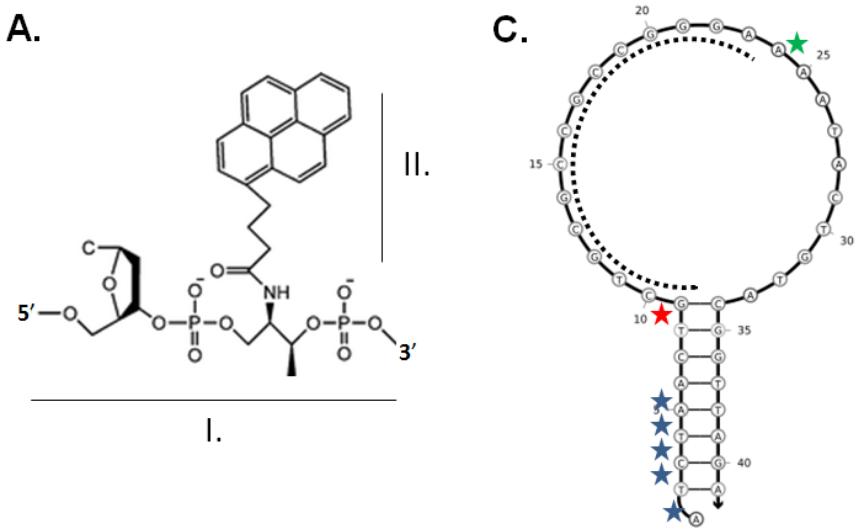
To more specifically probe the interactions between the protein and the aptamer, 3 pyrene phosophoramidites sgc8c mutants were synthesized [Figure 2-5]. When UV light is shone on these sequences, the pyrene dye is excited, and electron transfer occurs. This electron transfer has enough energy to break disulfide bridges and amide bonds that are within a 2nm radius of the pyrene [106]. Mingxu You and I hypothesized that pyrene added to aptamers within the consensus sequence would be in close enough physical proximity to amino acids of the PTK7 target that UV irradiation could break important bonds in the protein. If this occurred the PTK7-aptamer binding pocket would be disrupted, leading to decreased binding of those pyrene-sgc8c mutants to PTK7.

Our previous mutational analysis of sgc8c with azobenzene, discussed above, showed that insertion of a non-standard base directly into the consensus region caused decreased binding. To avoid this, we inserted pyrene mutations either at the cusp of the consensus sequence (red star in Figure 2-5c) or 1 base away from the consensus region (green star, same figure). Thinking perhaps multiple pyrene mutations in the stem could yield a more intense effect, even though they are not in the consensus region, we added 5–pyrene mutations to the 5' stem (blue stars).

As expected, the pyrene mutations alone did not affect the binding of the probes, which all bound as well as unmodified sgc8c to CEM cells [Figure 2-6a]. Furthermore, only the mutant with pyrene inserted closest to the consensus sequence showed a significant decrease in aptamer binding to the target cells [Figure 2-6 A.IV and B]. The other two mutants, with pyrene inserted in the loop or stem, had no significant change [Figure 2-6 A.III, A.V, and B.]. To ensure that UV alone was not damaging the cells or the DNA probes, all cells and aptamers received the same 15min dose of UV, regardless of their treatment group. Also, to ensure that internalization of the probes, which would decrease their abundance of the probes on the surface of the cells, and that heating from the UV lamp, did not interfere with our measurements, the entire protocol was performed at 4°C. The cells appeared healthy under the microscope with trypan blue staining, and when Scr-sgc8 was used as a negative control, no non-specific binding was seen [Figure 2-6 A.I].

When pyrene is excited by the UV lamp, electron transfer occurs in a localized environment, within ~2nm. Amide and disulfide bonds within this radius might possibly be disrupted. Disruption of the protein-aptamer binding site could weaken protein-

aptamer interactions, and decrease the signal seen by the flow cytometer. According to our results, there was a small but significant change in the binding ability of the pyrene-sgc8c mutant adjacent to the consensus region. The most likely reason for this change is that, in this region, the consensus sequence is in close physical proximity to the protein it binds.



B.

Key	Aptamer	Sequence
	sgc8c	5′ A-T-C-T-A-ACT-**GCTGCGCCGCCGGGA**A-AATACTGTACGGTTAGA-B 3′
★	sgc8c-stem	5′ APTPCPTPAPACT-**GCTGCGCCGCCGGGA**A-AATACTGTACGGTTAGA-B 3′
★	sgc8c-consensus	5′ A-T-C-T-A-ACTP**GCTGCGCCGCCGGGA**A-AATACTGTACGGTTAGA-B 3′
★	sgc8c-loop	5′ A-T-C-T-A-ACT-**GCTGCGCCGCCGGGA**APAATACTGTACGGTTAGA-B 3′

Figure 2-5. Pyrene sgc8c-mutants. A) Structure of pyrene-nucleoside DNA-base analog. I. Phosphodiester backbone. II. Pyrene base conjugated in place of standard DNA base. B) Sequences used in this experiment. Key shows “star” color in sgc8c structure, above right, Figure 2-5c. Pyrene bases are shown in red. Consensus sequence is bold-underlined. Dashes are inserted for clarity, they do not represent bases. C) NuPACK predicted structure of sgc8c with colored stars representing pyrene-insertion locations. Dashed line inside loop indicates consensus region.

Bioinformatic and competition analysis of aptamer sequences identified 4 aptamer sequences that shared a consensus region: GCTGCGCCGCCGGGA. Mutational analysis of sgc8c with azobenzene, blocking with sgc8c cDNA, and UV-mediated pyrene degradation of the protein-aptamer interaction, all showed that the region of sequence identity shared between sgc8c, KC2D8, KMF9b, and H01 aptamers was also the region most important for binding. Thus, I concluded the 15nt consensus region of these PTK7 aptamers is crucial for binding.

BLAST of the Consensus Sequence against the Human Genome

I hypothesized that these 4 independently selected, yet competing, aptamers might be mimics for a naturally occurring interaction between the PTK7 protein and natural DNA. If this was the case, the consensus sequence should be found in a target sequence in the DNA. To investigate this possibility, I blasted the consensus sequence against the human genome using the NCBI nucleotide BLASTn algorithm, adjusted for short sequences. This returned 8 hits with 14/15nt identity, detailed in Table 2-4. Six of these hits were within a gene: 4 in introns, 1 in the coding region, and 1 in the 5' untranslated region (5'UTR).

DIXDC1b has the Consensus Sequence

Looking over this list, one of these results, DIXDC1b, immediately peaked my interest because, like its name suggests, it has a DIX domain. DIX domains are very rare. Aside from DIXDC1, only two other proteins have them: Axin and Dvl. As is extensively discussed in the Introduction, these proteins are central in Wnt signaling, as is the aptamer's target protein, PTK7.

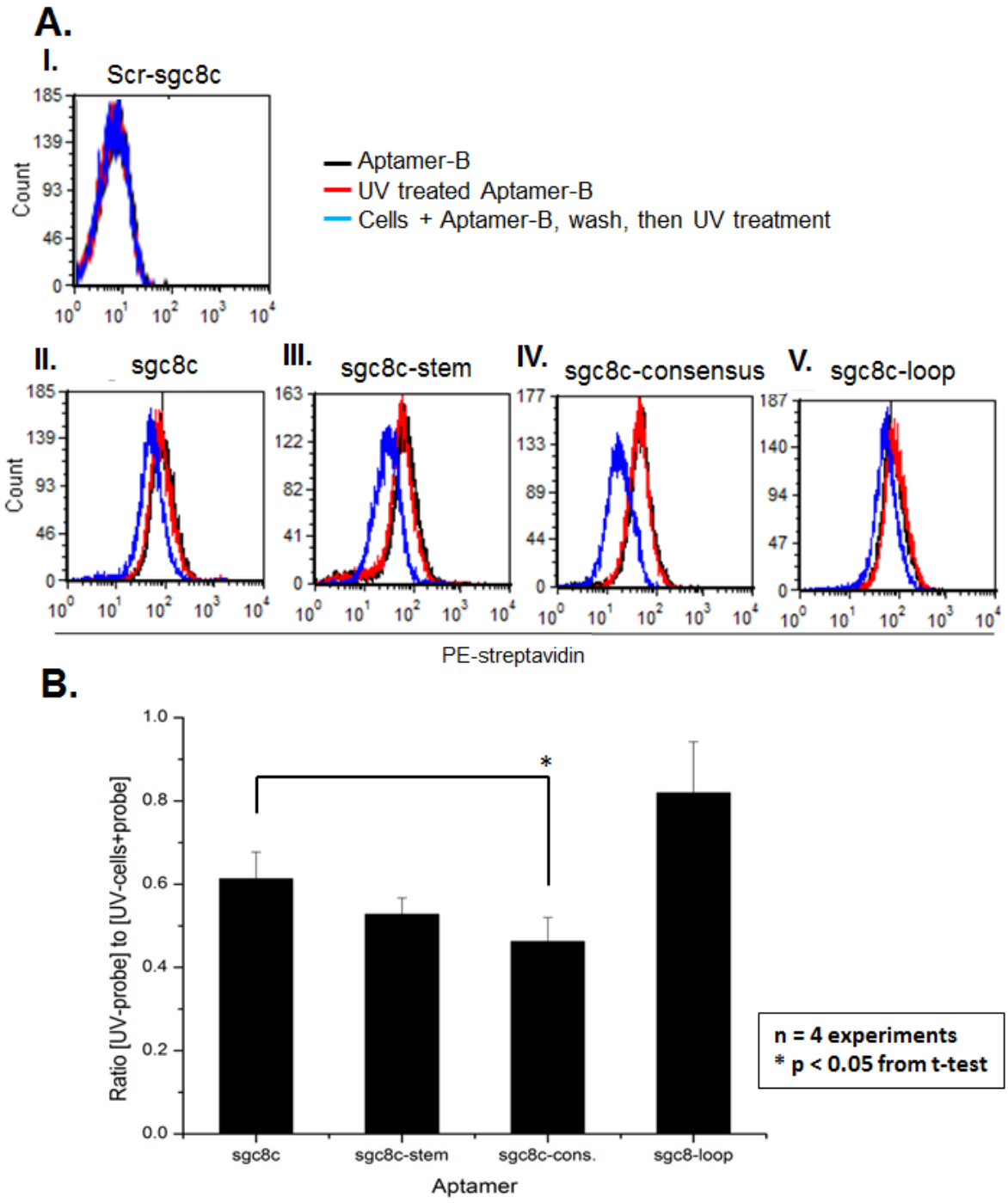


Figure 2-6. A) Representative data from one flow cytometry experiment. Scr-sgc8: scrambled sgc8c, a negative control. B: biotin. Sgc8c: another negative control. All probes and cells received 15min of UV irradiation at 4°C. B) Bar graph showing the ratio of fluorescence obtained by flow cytometry of UV-irradiated probe only versus UV-irradiated probe and cells together. p-value determined from Student's t-test. Data is averaged from 4 separate experiments with 3 replicates each.

Table 2-4. BLAST hits 14/15nt identity for GCTGCGCCGCCGGGA in human genome.

Protein Name, Abbreviation	Locus	Location in Gene
DIX Domain Containing 1, isoform b, DIXDC1b	NM_033425.3	5' UTR
Mucolipin 1, MCOLN1	NM_020533.2	Coding region
Membrane-bound transcription factor peptidase, MBTPS1	NM_003791.2	Intron
Ubiquitin-conjugating Enzyme E2D2, UBE2D2	NM_003339.2	Intron
Myc Induced Nuclear Antigen, MINA	NM_032778.4	Intron
SH3 and multiple ankyrin repeat domains protein 1, SHANK1	NT_011109.16	Intron
Nearest protein KIAA1875	NR_024207.1	Unknown
Nearest miRNA MIR302F	NW_001838467.2	Unknown

Recall from the Introduction that Dvl interacts with PTK7, the target of sgc8c, causing a switch in Wnt activation from the Wnt/β-catenin pathway toward the Wnt/PCP pathway. PTK7 interacts with Dvl through its PDZ domain. DIXDC1 and Dvl interact through their respective DIX domains, and DIXDC1 causes Dvl to depolymerize from large aggregates into smaller, 4 protein heteromers, which are biologically active. This depolymerization makes Dvl available for signaling. Furthermore, Dvl-DIXDC1 heteromers serve to inhibit JNK signaling, genes that are turned on by the Wnt/PCP pathway. Accordingly, DIXDC1 has the opposite effect of PTK7 on cell signaling. DIXDC1's interactions with Dvl make Dvl available to move to the membrane and signal, driving the expression of TCF/LEF genes, the prime target of the Wnt/β-catenin pathway, while simultaneously inhibiting Dvl expression of the pro-Wnt/PCP JNK genes.

Therefore, PTK7 and DIXDC1 have opposite effects—PTK7 switches Wnt signaling from being β-catenin dominant to being PCP dominant, whereas DIXDC1 switches Wnt signaling from being PCP dominant to being β-catenin dominant. These two proteins are intimately involved in the same signaling pathway; therefore, as with

many other signaling biofeedback loops within Wnt and other pathways, it would not be unexpected for them to modulate each other's expression [107].

Aligning the PTK7-aptamers with DIXDC1b DNA revealed a 1nt mismatch with sgc8 and KC2D8, 3nt mismatches with KMF9b, and 5nt mismatches with H01. Although H01 and KMF9b had more mismatches within the consensus region, they surprisingly both had additional bases in common with the DIXDC1b DNA adjacent to the consensus sequence: 5 additional nts for H01, and 8 additional nts for KMF9b [purple in Table 2-5]. Also of interest, H01 had a 10 fold larger K_d than KMF9b, indicating decreased binding affinity, and H01 also has 3 fewer bases in common with DIXDC1b DNA than KMF9b. The combined 22nt region in common between the PTK7 aptamers and DIXDC1b DNA is unique in the human genome.

KC2D4 has Identity with DIXDC1b Negative Strand

In addition to the 4 sgc8-type aptamers previously mentioned, there are 2 other aptamers, KDED19 and KC2D4, that had been shown to compete with sgc8 [4]. Neither of these two sequences shared significant identity with each other, the other

Table 2-5 Aptamers share sequence similarity with DIXDC1b DNA sequence. The sequence over the red line is the consensus sequence used for BLAST. Pink with white lettering denotes aptamer sequence similarity to DIXDC1b DNA within the consensus region. Purple with white lettering denotes sequence identity outside of the consensus region.

Aptamer	Sequence	K_d (nM)
DIXDC1b	GGCGCAGC–CTGCGCCGCCGGGAGCCTCCCTCCCAGTGGGAGATG	n/a
sgc8c	ATCTAACTGCTGCGCCGCCGGGAAAATACTGTACGGTTAGA	0.8 ± 0.1
KC2D8	TACTAACTGCTGCGCCGCCGGGAAAATACTGTACGGTTAGTT	1.1 ± 0.1
KMF9b	AGCGCAGCAGCTGTGCCACCGGGAGAATTTACGTACGGCTGAGCGA	0.3 ± 0.1
H01	AAGCAGCAGCTGTGCCATCGGGTTCGGATTTTCTTCCTACGACTGC	4.0 ± 0.3

PTK7 aptamers, the 146 other sequences in our SELEX database, or with the positive DIXDC1b DNA. Interestingly, however, the KC2D4's variable region showed sequence

identity with the negative strand of DIXDC1b DNA, five bases downstream from the sgc8 consensus sequence [Table 2-6].

Yunpeng Cai and I performed a simulation to determine the probability that this level of sequence identity—12/13nt between DIXDC1b DNA and KC2D4—was the result of chance. For this simulation, we took 1 million 39nt-long random sequences (the same length as KC2D4), and aligned them to a 52bp segment in the 5'-UTR of DIXDC1b DNA. Our simulation showed that 2,554 sequences out of 1 million sequences, or their reverse complement, have at least 1 13bp window that shares 12nt identity with the extracted DNA segment, which is the same similarity as the KC2D4 aptamer. Hence, the chance that a random sequence has the same similarity to the target region is p=0.0025.

Table 2-6. KC2D4 has identity to the negative DIXDC1b strand. Blue with white lettering is KC2D4 identity with the negative strand of DIXDC1b DNA. As in Table 2-3, Pink with white lettering denotes aptamer sequence similarity to DIXDC1b DNA within the consensus region. Purple with white lettering denotes sequence identity outside of the consensus region. p-value for KC2D4 alignment with DIXDC1b DNA is p = 0.0025.

Aptamer	Sequence	K_d (nM)
DIXDC1b DNA	+ GGCGCAGCCTGCGCCGCCGGGAGCCTCCCTCCCAGTGGGAGATGGGTTGAGA	n/a
	− CCGCGTCG—GACGCGGCGGCCCTCGGAGGGAGGGTCACCCTCTACCCAACTCT	
KC2D4	GAGGGTGACCATCGGTAAGGCGGGAATTGGCCCGGTAGC	54.3 ± 7.9

Conclusions

The 14nt identity among the 15nt consensus sequence is quite unusual, with a probability of 1.5×10^{-8}, or one occurrence in 33.6M bps. This estimate predicts we should find 94 other instances of our 14/15nt sequence in the human 3.1647×10^{9} bp genome. We only found this level of sequence identity for our consensus region 8 times. This discrepancy could be explained by the high number of repeated sequences throughout the human genome or because the BLASTn search only looks for contiguous bases. Unfortunately, the size of the human genome is so large that we are not able to enumerate all of the possible 14nt permutations. Hence, the coincidence cannot be judged by statistical inference, unless only a small part of the human genome is involved in whole cell-SELEX procedure, which we cannot assume.

Nevertheless, our simulations discourage the idea that these 5 aptamers share sequence similarity with DIXDC1b DNA through coincidence alone, because, if the 22nt sequence in common between the PTK7 aptamers and DIXDC1b DNA is a coincidence, then the KC2D4-DIXDC1b DNA match should also be a coincidence, as they would have happened independently. It is statistically improbable that both of these two coincidences would occur at the same time. Hence, if the PTK7-DIXDC1b DNA match is not a coincidence, then the KC2D4-DIXDC1b DNA match is also unlikely a coincidence.

These simulations do not take into account that all 5 of these aptamers compete with each other for binding to the same site on the cell, or that the consensus region is the most important aptamer region for binding. If we take these experimental results into account, the likelihood that this confluence of data is just a random coincidence becomes very, very remote. Therefore, we concluded that if our observations were not a product of chance, there must be some underlying biologically relevant reason for a

DNA aptamer having an analog in genomic DNA. Using this as the basis for our reasoning, we began to look for possible scenarios where PTK7 protein could interact directly with DIXDC1b DNA. These explorations will be detailed in the next chapter.

Materials and Methods

DNA Sequences

All aptamer sequences, shown in Table 2-7, were synthesized in-house at 1 μmole-scale synthesis using an ABI 3400 DNA/RNA synthesizer (Applied Biosystems). Special bases such as the azobenzene and pyrene were dissolved in dry acetonitrile before coupling for twice the time of standard bases on the synthesizer. Sequences were deprotected from CPG beads with AMA (ammonium hydroxide: methylamine 1:1) at 50°C for either 20min or 12h for the azobenzene modified sequences. All sequences longer than 30nt that were either unlabeled, or had 3' biotin-CPG (Glen Research) modifications, were purified using Glen-Pak columns (Glen Research). Aptamers that included special bases such as azobenzene or pyrene, or that were shorter than 30nt in length were ethanol precipitated, resuspended in 1M TEAA and purified on a reverse phase Prostar HPLC (Varian) using a C18 column (Econosil, 5U 250 x 4.6 mm from Alltech Associates) with a linear elution. These sequences were vacuum dried or ethanol precipitated, detritylated with 20% acetic acid, and stored at minus 20°C for future use.

The absorbance at 260nm for the purified aptamers was quantified using a UV-Vis spectrophotometer (Bio-Rad). Aptamer concentration was determined using the Lambert-Beer equation: Abs 260nm = εbc, where ε is the extinction coefficient, b is the cuvette pass length, and c is the DNA concentration. All sequences were tested for quality control by running 0.5μL of purified aptamer on a 3% agarose gel in TBE buffer

(89mM Tris-HCl, 89mM boric acid, 2mM EDTA, pH 8.0) for 40min at 120V and imaged under UV with Ethidium Bromide (EtBr).

Cell Culture

CCRF-CEM (CCL-119, a T-cell line, Acute Lymphoplastic Leukemia) and Ramos (CRL-1596, a B-cell line Burkitt's lymphoma) cell lines were obtained from American

Table 2-7. All DNA sequences used in this chapter. B: Biotin; UL: unlabeled; Z:Azobenzene; -$(T)_{10}$-B: TTTTTTTTTT P: Pyrene.

Experiment	Name	Sequence 5' to 3'
sgc8c Optimization	sgc8c-41	ATCTAACTGCTGCGCCGCCGGGAAAATACTGTACGGTTAGA-B
	sgc8c-38	CTAACTGCTGCGCCGCCGGGAAAATACTGTACGGTTAG-B
	sgc8cx	CTAACTGCTGCGCCGCCGGGAAAATACTGTACGGTTAGA-B
	xsgc8cx	TCTAACTGCTGCGCCGCCGGGAAAATACTGTACGGTTAGA-B
	xxsgc8	ATCTAACTGCTGCGCCGCCGGGAAAATACTGTACGGTTAG-B
Competition and K_d Experiments	sgc8c-B	ATCTAACTGCTGCGCCGCCGGGAAAATACTGTACGGTTAGA-B
	sgc8-no	ATCTAACTGCTGCGCCGCCGGGAAAATACTGTACGGTTAGA
	KC2D8-B	ATCGTCCGCCACCACTACTAACTGCTGCGCCGCCGGGAAAATACTGTACGGTT AGTTTGAGACTGCCTGCCGATGT-B
	KC2D8-UL	ATCGTCCGCCACCACTACTAACTGCTGCGCCGCCGGGAAAATACTGTACGGTT AGTTTGAGACTGCCTGCCGATGT
	KMF9b-B	AGCGCAGCAGCTGTGCCACCGGGAGAATTTACGTACGGCTGAGCGA-B
	KMF9b-no	AGCGCAGCAGCTGTGCCACCGGGAGAATTTACGTACGGCTGAGCGA
	H01-B	AAGCAGCAGCTGTGCCATCGGGTTCGGATTTTCTTCCTACGACTGC-B
	H01-no	AAGCAGCAGCTGTGCCATCGGGTTCGGATTTTCTTCCTACGACTGC
	Scr-sgc8-no	ACATGGAGCGTCCAAATCGAGTCGGATATCACGTGCTCGAT-B
	Scr-sgc8-no	ACATGGAGCGTCCAAATCGAGTCGGATATCACGTGCTCGAT
	TD05-B	AACACCGTGGAGGATAGTTCGGTGGCTGTTCAGGGTCTCCTCCGGTG-B
	TD05-no	AACACCGTGGAGGATAGTTCGGTGGCTGTTCAGGGTCTCCTCCGGTG
sgc8c Azobenzene Mutational Analysis	Stem1A-10T-B	AZTCTAACTGCTGCGCCGCCGGGAAAATACTGTACGGTTAGA-$(T)_{10}$-B
	Stem1B-10T-B	ATCTAZACTGCTGCGCCGCCGGGAAAATACTGTACGGTTAGA-$(T)_{10}$-B
	Loop1C-10T-B	ATCTAACTGCTGCGCCGCCGGGAAZAATACTGTACGGTTAGA-$(T)_{10}$-B
	Stem1D-10T-B	ATCTAACTGCTGCGCCGCCGGGAAAATACTGTACZGGTTAGA-$(T)_{10}$-B
	Stem1C-10T-B	ATCTAACTGCTGZCGCCGCCGGGAAAATACTGTACGGTTAGA-$(T)_{10}$-B
	Loop1A-10T-B	ATCTAACTGCTGCGZCCGCCGGGAAAATACTGTACGGTTAGA-$(T)_{10}$-B
	TD05-10T-B	AACACCGTGGAGGATAGTTCGGTGGCTGTTCAGGGTCTCCTCCGGTG-$(T)_{10}$-B

Table 2-7. Continued

Experiment	Name	Sequence 5' to 3'
cDNA Blocking Experiments	sgc8c-B	ATCTAACTGCTGCGCCGCCGGGAAAATACTGTACGGTTAGATTT-B
	5'-14	TAGATTGACGACGC
	5'-19	TAGATTGACGACGCGGCGG
	5'-22	TAGATTGACGACGCGGCGGCCC
	3'-11	CATGCCAATCT
	3'-15	ATGACATGCCAATCT
	3'-19	TTTTATGACATGCCAATCT
	C-41	TAGATTGACGACGCGGCGGCCCTTTTATGACATGCCAATCT
	Scr-sgc8-B	AACACCGTGGAGGATAGTTCGGTGGCTGTTCAGGGTCTCCTCCGGTG-B
Pyrene-UV disruption	sgc8c-B	ATCTAACTGCTGCGCCGCCGGGAAAATACTGTACGGTTAGA-B
	sgc8c-stem-B	APTPCPTPAPACTGCTGCGCCGCCGGGAAAATACTGTACGGTTAGA-B
	sgc8c-cons-B	ATCTAACTPGCTGCGCCGCCGGGAAAATACTGTACGGTTAGA-B
	sgc8c-loop-B	ATCTAACTGCTGCGCCGCCGGGAAPAATACTGTACGGTTAGA-B

Type Culture Collection (ATCC). Both cell lines were cultured in RPMI-1640 medium (GIBCO) with 10% FBS (GIBCO) at 37°C under 5% CO_2. 100 units/mL penicillin-streptomycin (Cellgro) were added to the cultures, occasionally, if bacterial contamination was suspected in the lab. Cultures were routinely monitored for *Mycoplasma* infection.

Immediately before experiments, cells were collected by centrifugation at 1,260 RPM for 3min and washed with 2mL ice cold Washing Buffer (WB: PBS, 4.5g/L glucose, 1M $MgCl_2$). The cells were resuspended in Binding Buffer (BB: WB with 1g/L bovine serum albumin, and 100mg/L tRNA) at a concentration of 20 x 10^6 cells/mL and placed on ice. For each experiment 1 x 10^6 cells were used. These conditions remained constant for all experiments unless otherwise noted.

Bioinformatics

We compiled a dataset of 148 unique aptamer sequences from 33 different selections, removed their primers, and aligned them with ESPRIT (Appendix A). Using the convergent property of SELEX [108], the p-value *p* obtained in our simulation indicates that, given a target sequence *S1*, a reference sequence *R,* which has a known similarity value *t* to *S1*, and a random sequence *S2,* with the same length as *R*, the probability that *S2* is at least as similar to *S1* as *R* is *p*. Under this non-deterministic model, we supposed we have *n* (here *n* ~10,000) aptamers in the system. The probability that a random aptamer can be more similar than *R* is $1-(1-p)^n$. If we examine *k* targets (here *k*=148), the probability of one random hit will be $1-(1-p)^{nk}$, which is approximately *nk*p* when *p* is small. Hence, if *nk*p* << *1*, we claim that *R* is not likely a random match. This means that *R* should be identical, or at least correlated to *S1*.

In this case, we had *k**(*k*-1)/2 random pairs in total where *k*=148. For a given pair, the chance of one random hit is $1-(1-p)^{k(k-1)/2}$. If *k(k-1)p*/2 << 1, we claim that the pair is not a random match, but either identical or correlated. From the results, sgc8c and KC2D8 had the greatest alignment. Two other aptamers, H01 and KMF9b, which had sequence identity to each other and to sgc8, were identified. Based on this result, a GenBank BLASTn [109] search was performed on the following consensus region: GCTGCGCCGCCGGGA.

Significance Simulations

We performed random simulations to find the probability that a random pair of sequences *a* and *b* resembles the similarity is *P*. If we have *n* sequences, then we have *n*(*n*-1)/2 pairs. Hence, the overall probability of coincidence is *n*(*n*-1)/2**P*. With this

criterion, we confirmed the pair, sgc8c-KC2D8, is a confidently non-coincident match, with significance level P_0<0.01. Suppose we have found a matched pair *a*-*b*, with probability 1-P_0 that the match is not by chance, and we have found a third sequence *c* which shares some degree of similarity with *a* or *b*, while random simulation showed the probability of resembling this similarity is P_1. Then, the overall probability of coincidence is 2*(1-P_0)*(*n*-2)* P_1. With this criterion, we find that sequence KMF9b is matched to the pair sgc8c-KC2D8, with significance level P_1<0.03. Similarly, sequence H01 is matched to the above triplet at significance level P_2<0.05. We can conclude that *each* of the 4 sequences matching with the others cannot be explained by coincidence.

Furthermore, the 22nt common region between the 4 PTK7 aptamers and DIXDC1b DNA is unique in the human genome. We examined the probability that another random sequence can match the same DNA in adjacent regions. We carried out a simulation by generating 1 million random sequences 39nt long, the same as KC2D4, and aligned them to a 52bp segment in the 5'-UTR of DIXDC1b DNA. Because local alignment is required to measure the similarity between the sequences and the DNA, we used FASTA with the following command:

fasta35 -a -z -1 -n -q -H -b 1 -d 1 -O <outputfile> <querysequences> <DIXDC1bDNA.fas>

Competition and Binding Experiments

One million cells were washed with 2mL washing buffer (WB: PBS, 4.5g/L glucose, 1M $MgCl_2$), probed with the first aptamer in 100μL binding buffer (BB: WB with 1g/L bovine serum albumin, and 100mg/L tRNA) for 20min at 4°C then probed with the 10x unlabeled second aptamer for 20min at 4°C, washed, stained with 1:400 streptavidin-PE, washed again, and measured using a FACScan flow cytometer (Becton Dickenson), counting 20k cells. The data was analyzed with FCS Express. The different

aptamer treatments were as follows: no aptamer; 80nM scrambled sgc8-biotin; 80nM aptamer-biotin; 800nM unlabeled-aptamer then 80nM aptamer-biotin; 800nM unlabeled-sgc8c then 80nM aptamer-biotin; and 800nM unlabeled-aptamer then 80nM sgc8c-biotin. For each experiment Ramos cells, which do not bind PTK7 aptamers, were used as a negative control.

Off-Rate Experiments

CEM cells were incubated with biotinylated aptamers (125nM) for 20min. The cells were washed and then incubated with 10x of the same, unlabeled aptamer (1.25μM) for different amounts of time: 120m, 60m, 30m, 20m, 10m, 6m, 1m. Cells were then washed and probed with 1:400 PE-streptavidin and measured by flow cytometry. The experiment was carried out at 4°C to prevent aptamer internalization.

Kd Measurements

The apparent equilibrium dissociation constant (K_d) for the aptamer-cell interactions was determined by adding varying concentrations of biotin-aptamer to $1x10^6$ CEM-CCRF cells in BB at 4°C. Cells were washed and incubated with 1:400 streptavidin-PE (Invitrogen), washed again, and analyzed by flow cytometry, counting 20k cells. The mean fluorescence intensity of background binding with scr-sgc8c aptamer or TDO5 aptamer (for the Azobenzene measurements) was subtracted for each aptamer concentration. Data from 2 or 3 separate experiments with 3 replicates was averaged. Intensity and aptamer concentration were fit to the hyperbolic equation $Y= B\ max\ X/(K_d +X)$, using Origin8.5 to measure K_d.

Azobenzene Mutation Measurements

Azobenzene phosphoramidite was synthesized according the published protocol [104]. K_d for each sgc8c-azobenzene mutant was determined as described above

without modification. For each aptamer tested, Ramos cells, which do not bind sgc8c, were used as a negative control.

cDNA Blocking

Sgc8c-B aptamers were hybridized by mixing equal molar equivalents of the various cDNA in H_20, which was heated to 95°C for 5min and allowed to slowly cool to room temperature. Hybridization was confirmed by running a small aliquot on a 3% agarose gel in TBE for 40min at 120V and noting a shift in molecular weight. The sequences were then tested for binding with CEM cells similar to the competition experiments described above, using 250nM concentrations of probe. For each experiment Ramos cells, which do not bind sgc8, was used as a negative control.

Pyrene-sgc8c Binding

For the photo-regulation of the pyrene-sgc8c, cells (1×10^6) and aptamer (250nM) were irradiated in BB at 350 nm with a UV−B lamp centered at 302 nm (SANKYO DENKI, Japan) with a 352 nm optical filter (3 nm half bandwidth; Oriel Instruments, Stratford, CT, Newport).

Aptamers and cells were separated into 3 groups: aptamer with UV-irradiated cells; UV-irradiated aptamer with UV-irradiated cells; and UV-irradiated cells-pre-bound to aptamer. The aptamers and cells (except for the aptamer that received no UV), either separately or together, all received a total of 15min UV-irradiation. UV-irradiation and aptamer binding was all conducted at 4°C. Cells were stained afterward with trypan blue and checked under the microscope to ensure the UV-irradiation did not cause cell death during the experiment. For each experiment Ramos cells, which do not bind sgc8, was used as a negative control.

CHAPTER 3
POSSIBLE FUNCTIONAL ROLE FOR PTK7-DIXDC1B INTERACTION

Introduction

Our lab focuses on whole cell-SELEX of DNA libraries against cancer cell lines, identifying aptamers to target specific tumor types. After many of these selections, we noticed a curious phenomenon: aptamers from different selections competed against each other for the same target, a membrane protein tyrosine kinase, PTK7. These selections involved different cell lines, DNA pools, and primers. They were performed years apart by different researchers. The chances of selecting aptamers against the same target multiple times were remote. Why was this happening?

We hypothesized there might be a natural, physiological role for these sequences. We whittled down the smallest region of nucleotide consensus between the aptamers, and BLASTed it against the human genome. Surprisingly, we found the consensus sequence appears in the regulatory region for the protein, DIXDC1b, which is involved in the same Wnt signaling pathway as our aptamer's target protein, PTK7. Moreover, aligning the aptamer sequences with this regulatory region, we found more bases in common than just the predicted consensus sequence. This family of PTK7-binding aptamers has significant sequence identity to both the positive and negative strands of the DIXDC1b DNA, implying PTK7 may interact naturally with this regulatory region, and explaining why we repeatedly selected aptamers for this protein.

Results and Discussion

PTK7 Aptamers have Sequence Identity with DIXDC1b 5' UTR

So if what we have found is not a coincidence, what is happening? Why have our DNA aptamers been found in the genome? To help answer these questions we needed to determine where in the DIXDC1b gene the aptamer sequence identity was found.

DIXDC1 protein in humans is coded by two main isoforms with separate regulatory regions. There is a deeper discussion of DIXDC1 in the Introduction. One isoform, DIXDC1a, is longer. It has an extra 200aa N-terminal region with a CH domain and another domain that lets it bind to actin. DIXDC1a is found in many adult tissues, and it localizes with actin to focal adhesions at the edge of the cell. The second isoform, DIXDC1b, lacks the CH- and actin-binding domains, does not bind actin, and is found diffusely throughout the cytoplasm of cells, especially in neurons, during development. The region of PTK7-aptamer identity is only found in the 5'UTR of DIXDC1b [Figure 3-1]. The 5' UTR is a well-known site for translational regulation, but it can also be important for transcriptional regulation. Therefore, it can act as a control region for mRNA expression. The PTK7 aptamers' sequence identity is found in the 5' UTR of DIXDC1b.

```
  1 ATCCGGAAGG TGGCACGGAG TGGGATCGCC GCTGGGGACT CGAGGCGCAG CCTGCGCCGC
 61 CGGGAGCCTC CCTCCCAGTG GGAGATGGGT TGAGATGCCC CCGCCAGGGG GGATGCCCGG
121 CACCGTGCGT CCGCGGAGGC CAAGATGCAG CGGCCAGGGG CCGGCAGCCT GCGAGGGGAG
181 GCAGCTTCCG CCGGGGCCGG GCTGCTGCAC AGTCTGAGCG GCCGGGACTG CGCGCTTCAG
241 AGCCTGGAGC ATCCCAGTCG CTGGGGCCGA GACGCCGCCG CCGCCGCCGT TCCCGCTTTC
301 TCCCGCGAGC CGGGCCAGTA GCTTTGCTAG CTGGCCTTCC CGTGGAGGCG TTTTCCAGCC
361 CCAGCGCGGG GAGACATGCC TGAATTTGGG AGCGATGTGA CTCTCAGCCT CCCACTTCAC
421 CCGGGGACGC AGGCTTGCTG AAGCCCGAGA CAGGAGGGGG ACCATGGGAG GGACGCAAGT
```

Figure 3-1. The consensus sequence is found in DIXDC1b 5' UTR. DIXDC1b mRNA locus: NM_033425. Black with white lettering is sgc8c-KMF9b-type aptamer sequence similarity. Grey with black lettering is cKC2D4 similarity. DIXDC1b's ATG methionine start site is in red. A rather strong Kozak sequence surrounds the start site: GGGACCATGG [110].

The Consensus Region is Conserved among Species

As both the DIXDC1 and PTK7 proteins are highly conserved across vertebrates, if there was an interaction between the protein PTK7 and DIXDC1 we might hypothesize that it would be conserved. By BLASTn searches of various published genomes, this is indeed what we find. In the DIXDC1b putative regulatory regions (as determined by Transfac analysis) of human (*Homo sapiens*), chimpanzee (*Pan troglodytes*), mouse (*Mus musculus*), and rat (*Rattus norvegicus*), we find sequence identity with the aptamers [Table 3-1].

Zebrafish (*Danio* rerio) also has PTK7 and DIXDC1 orthologs, but, unlike the mammals mentioned above, *D. rerio* only has one isoform of DIXDC1, which corresponds to DIXDC1b and is called Ccd1. There is no sequence identity between the putative regulatory region of the *D. rerio* gene Ccd1 and the aptamers. However, a 23nt sequence within a gene for a hypothetical protein called loc797129 was found in chromosome 12. Ccd1 is on chromosome 21. Nothing is known about this hypothetical protein, according the PubMed citation, it was found through automated computational analysis using GNOMON. *D. rerio* has ~331M base pairs in its non-repetitive genome [111]; the chance that a random sequence has 23bp in common is $1.4 \times 10^{-14.}$ This means that, randomly, only 1 genome of ~1 million different genomes the same size of *D. rerio* would have this exact sequence.

This result in *D. rerio* is intriguing, but it remains highly speculative that the hypothetical protein in *D. rerio* has anything to do with PTK7, as nothing is known about the potential protein. Of more importance to this discussion are the mammalian sequences that have retained some of the bases found in the regulatory regions, but not necessarily the 5'UTR, of DIXDC1b orthologs. This indicates these sequences are

conserved, and lends credence to the supposition that the sequences play a functional role.

Table 3-1. Consensus region is conserved among species. Blue with white lettering is KC2D4 identity with the negative strand of DIXDC1b. As in Table 2-3, Pink with white lettering denotes aptamer sequence similarity to DIXDC1b DNA within the consensus region. Purple with white lettering denotes sequence identity outside of the consensus region. P. troglodytes, M. musculus, and R. norvegicus sequences are all in the regulatory regions for DIXDC1b. D. rerio sequence is not.

Species	Sequence
H. sapiens	CGCAGCCTGCGCCGCCGGGAGCCTCC----CTCCCAGTGGGAGATGGG
P. troglodytes	CGCAGCCTGCGCCGCCGGGAGCCTCC----CTCCCAGTGGGAGATGGG
M. musculus	CGGTGTCTCAGCTACCGGGACAGTACGCTTCACCCTAAGGGAGCATCC
R. norvegicus	TGAATGAGGCGCCGCCGGGAGCCTGGCCTACCTTGCACTGGTCGAGTG
D. rerio	CGCAGCCTGCGCCGCCGGGAGCCCAGCGGGGCCTGCAGCGGGACGGGC

DNA is Double Stranded, but our Aptamers are Single Stranded

These aptamers are ssDNA, and their specific recognition is contingent on their secondary structure. If sgc8c- and KC2D4-type aptamers are indeed a mimic for natural interactions between DIXDC1b DNA and PTK7, they need to interact with dsDNA. We modeled a 70nt region surrounding the aptamer consensus region with NuPACK DNA secondary structure prediction software. As expected, the region had apparent secondary structure with two prominent stem-loop hairpins [Figure 3-2].

However, various attempts to see if double-stranded regions of the DIXDC1b DNA could bind with cells or compete with sgc8c showed no binding or competition [Figure 3-3]. We know there is a large amount of basal transcription machinery in place on the DNA during transcription that can melt small regions of dsDNA, allowing for a secondary structure to form. It is possible that some other factors are needed to allow the protein to bind to the DNA *in situ*. Or, perhaps we have not identified the correct dsDNA region.

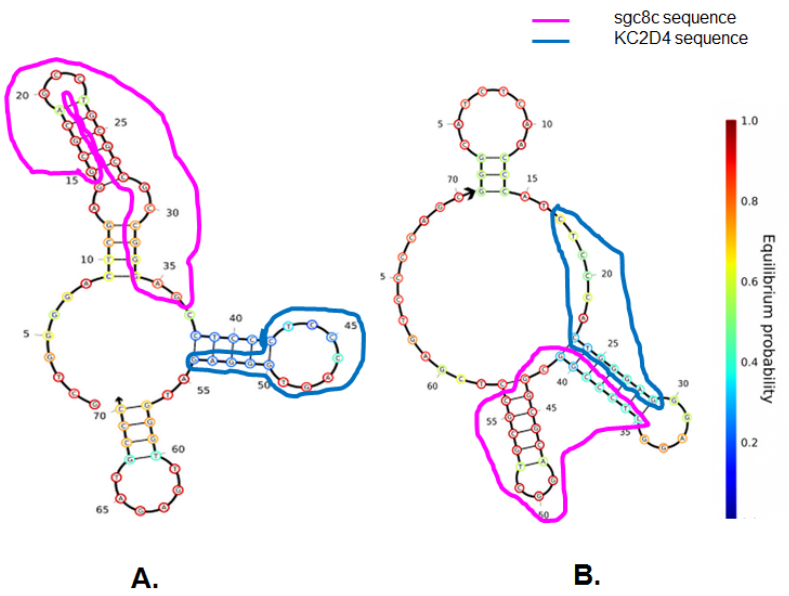


Figure 3-2. NuPACK structures for 70nt surrounding the aptamer consensus region in DIXDC1b DNA. A) DIXDC1b forward, positive strand. B) DIXDC1b reverse, negative strand. Pink line denotes 22nt region of aptamer identity to sgc8c-KMF9b regions. Blue line denotes 15nt region of KC2D4 sequence identity. Bar on right shows strength of predicted structure. Red being most strong.

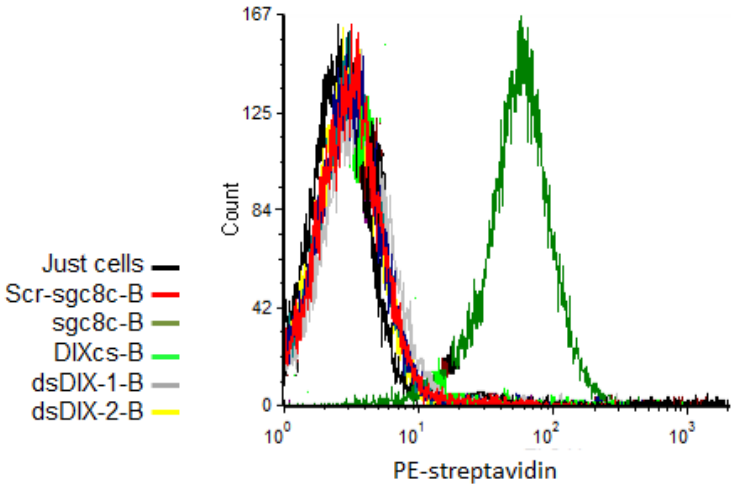


Figure 3-3. dsDIX DNA does not bind CEM cells. Scr-sgc8c: Scrambled sgc8c; B: biotin; DIXcs: DIX consensus sequence, 22nt sequence in common between the KC2D8 and sgc8c aptamers and DIXDC1b DNA. dsDIX-1: Double stranded oligo of 22nt bases. dsDIX-2: Double stranded oligo of 40nt surrounding the consensus sequence.

PTK7 Structural Analysis

The extracellular portion of PTK7 interacts with the PTK7 aptamers. Thus, the region of PTK7 that interacts with the aptamers must have a protein structure that would be conducive to such protein-DNA interactions. To determine where on the PTK7 extracellular portion the aptamers would be most likely to bind, we used Swiss Modeler to model the extracellular region of PTK7, and analyzed that model for areas of likely DNA-protein interaction.

As detailed in the Introduction, the extracellular portion of PTK7 is made up of 7 Ig-like domains. There are many important human transcription factor families that use an Ig-like fold to bind dsDNA, including Runx, NFAT/AP1, NKκB, p53, STAT, and Brachyhury T-box proteins [112].

Ig folds are made when two β-sheets are held together by a disulfide cysteine bridge. Each Ig-like fold has 8 strands, labeled A-G in Figure 3-4a. While the individual structures of the 6 families of transcription factors are unique to the families, there are several aspects that they all share. DNA is positively charged; consequently it has an attraction for the basic amino acids: lysine, arginine, and, under the right conditions, histidine. Each member of the Ig-like fold families of transcription factors interacts with DNA through basic residues on the A-B loop, the E-F loop, and the C-Tail. An example of such an interaction is shown in Figure 3-4b. Here, the crystal structure of Runx1 (PDB: 1h9d [113]), in complex with its dsDNA target, is shown with the important basic residues (shown as red stars) that determine its interaction with DNA.

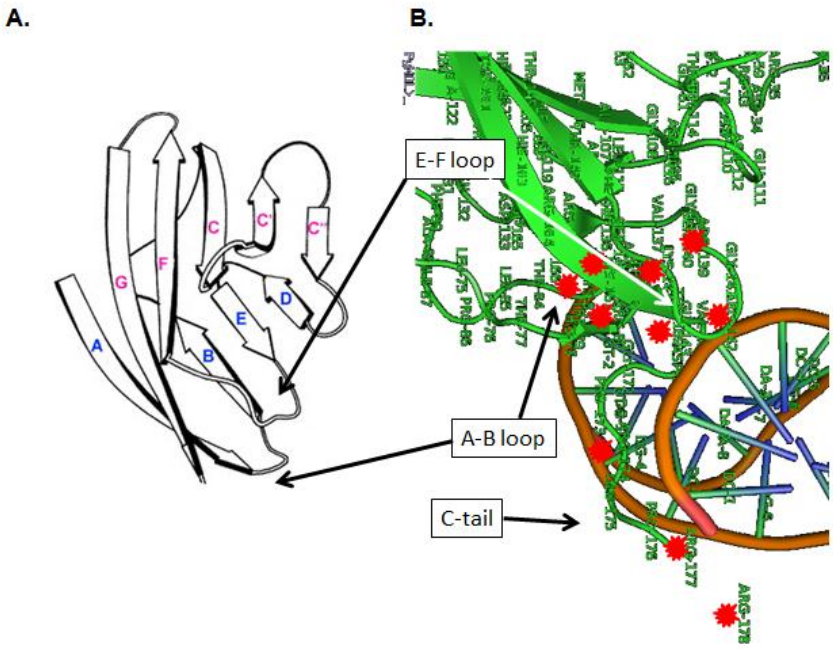


Figure 3-4. Structural features of Ig-like folds. A) Scheme showing an Ig-like fold with each strand labeled. The pink letters (G, F, C) correspond to one β-sheet and the blue letters (A, B, E, D) correspond to the other. The A-B loop and E-F loop are indicated by black arrows. A disulfide bridge is typically formed between the B and F strands. B) Crystal structure of Runx1 transcription factor, green, bound to DNA, brown backbone with blue-green DNA bases). The A-B loop, E-F loop, and C-tail are indicated by arrows. The red stars indicate basic residues important for binding.

We modeled the 3-dimensional structure of the 7 extracellular Ig-like domains of the PTK7 protein using titin (PDB 2nzi for Ig 6-7 and 3B43 for Ig 1-5) as a template [Figure 3-5a]. Then, using the classic Ig-like fold architecture, and the common binding pattern of transcription factors with Ig-like folds as a starting point, we tried to predict where PTK7 would most likely bind the aptamers. Of all the Ig-like domains only 2 had a preponderance of basic residues: Ig7 and Ig3 [Figure 3-5b, c]. The Ig7 domain of PTK7 contains the MT1-MMP cleavage site discussed in the Introduction. The PTK7 receptor is cleaved at this site, releasing a 70kDa protein fragment that affects PTK7 signaling. As proteolytic cleavage at the Ig7 domain would leave the basic residue area

behind, still attached to the membrane, we do not predict that the PTK7-aptamers would bind here.

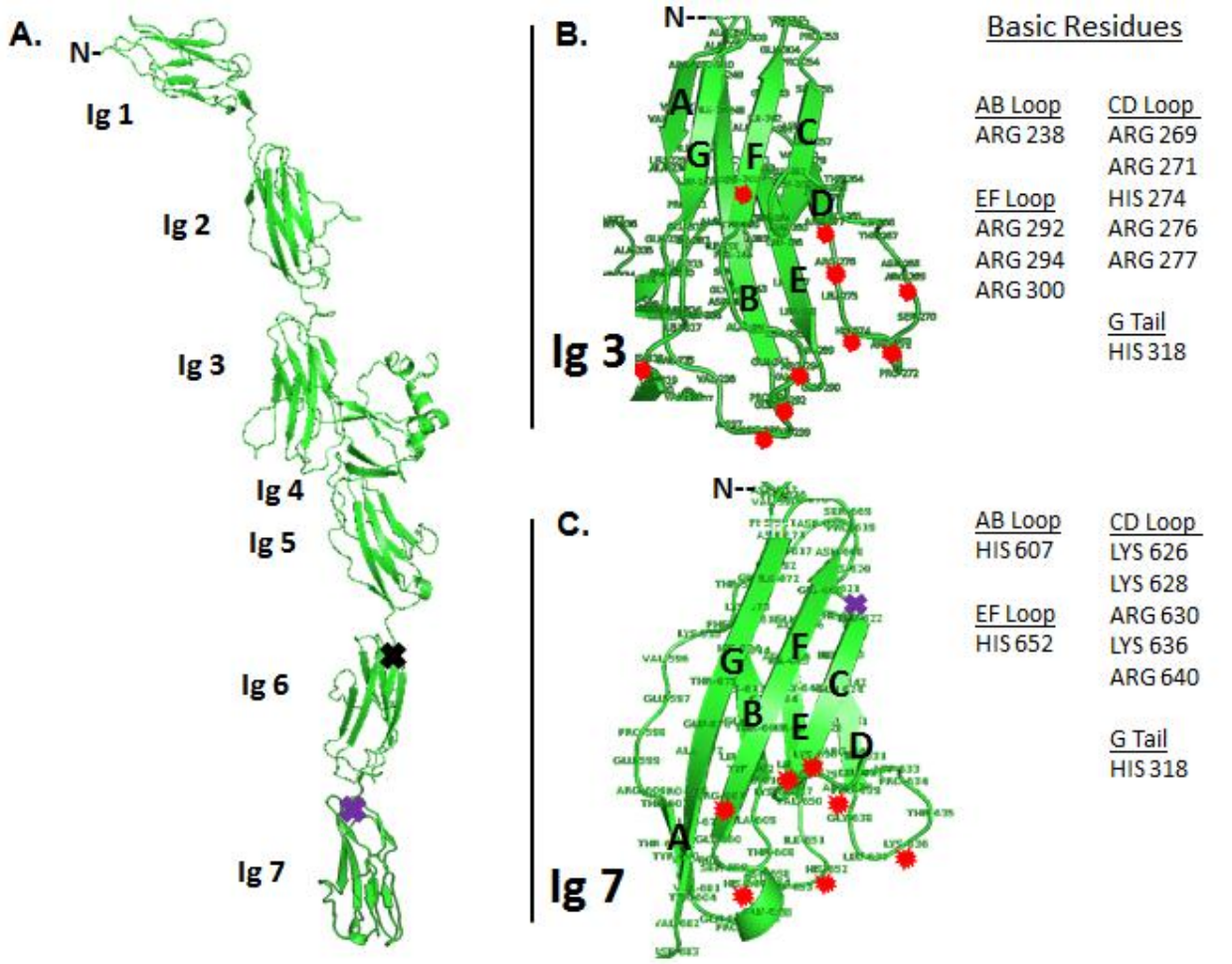


Figure 3-5. Model of PTK7 extracellular domain. A) Model of PTK7's 7 Ig-like extracellular domains. The transmembrane domain and cytosolic pTK domain are not shown, but would be below the Ig7 domain. This protein fragment is ~70kDa. The purple X denotes the natural MT1-MMP cleavage site; the black X is the additional *chuzhoi* mutant MT1-MMP cleavage site. B) Detail of the Ig3 domain. C) Detail of the Ig7 domain. Basic residues are shown by red stars. Each strand of the Ig β-sheet is labeled. A list of basic residues and their location on the protein are shown to the right side of each model.

The Ig3 domain also has many basic resides, including 4 arginine residues in the A-B and E-F loops that are important for other Ig-like transcription factor-DNA interactions mentioned above. Furthermore, when PTK7 is cleaved by MT1-MMP, the

Ig3 domain would remain with the cleaved PTK7 fragment. Therefore, we predict the Ig3 domain would be the most likely domain to bind the PTK7 aptamers. This prediction is tenuous, but may serve as a guide for future PTK7 mutation experiments that can pin-point the actual site of PTK7-aptamer binding.

PTK7 Has a Weak Predicted Nuclear Localization Sequence

If PTK7 does indeed interact with the regulatory region of DIXDC1b DNA in the nucleus, it needs to first get into the nucleus. Although smaller proteins can diffuse through the nuclear pore complex (NPC) relatively quickly, transport of proteins larger than 60kDa through the NPC by diffusion alone is very slow. Larger proteins therefore require a nuclear localization signal (NLS) to facilitate their nuclear transport. NLS can be specific amino acid sequences, typically a short string of basic residues, such as PPKKKRKV for the SV40 virus T-antigen, or basic patches on the protein's secondary structure. These NLSs interact with proteins found in the NPC and/or with specialized nuclear import receptors, which shuttle proteins from the cytoplasm into the nucleus. There are several algorithms developed in recent years that attempt to predict from a protein's primary amino acid structure the likelihood that a protein contains an NLS.

Using two of these tools, PredictNLS [114] and NLStradamus [115], I searched the PTK7 protein for a predicted NLS. When searching for these sites, I used another membrane tyrosine kinase, the protein epidermal growth factor receptor (EGFR), as a positive control. EGFR is transported to the nucleus using the NLS: RRRHIVRKRTLRR [116]. If the algorithm found EGFR's NLS, then it might be more likely to find one in PTK7. The PredictNLS did not find an NLS in PTK7, but it also did not identify the validated NLS in EGFR. NLStradamus did a better job identifying the EGFR NLS, and made a weak prediction for a NLS in the PTK7 protein (RPPHLRR). This predicted

amino acid residue is in the Ig3 domain of PTK7's extracellular domain, the same region I predicted for aptamer binding in the preceding section. Strangely, while this RPPHLRR motif is conserved in mouse, chimpanzee, rat, and cow, it is deleted in its entirety from chicken PTK7, called KLG. KLG and PTK7 share 70% homology over the extracellular region. This is the only region deleted between PTK7 and KLG extracellular domains, and it is the longest contiguous stretch of difference between the two proteins.

From this prediction analysis, PTK7 does not contain a validated NLS. However, not all NLSs have been characterized, and the absence of an NLS does not necessarily eliminate the possibility that PTK7 can be transported to the nucleus. More experiments need to be performed to determine what PTK7's role is in the nucleus.

Confocal Microscopy

The most direct way to see if PTK7 is found in the nucleus is through confocal microscopy. By staining the nucleus with DAPI and performing slice-by-slice confocal analysis of cells stained with either an aptamer or antibodies against PTK7, colocalization of PTK7 in the nucleus can be determined. I tried numerous times to co-stain cells for PTK7 and nuclei using many different parameters. Unfortunately, while internalized sgc8c-labeled aptamers are clearly seen by flow cytometry, under the microscope, the staining was too weak to determine where in the cells PTK7 was located. Due to the low quality of these images, I have not included them here. More optimization of the immunocytochemistry protocol for PTK7 needs to be done to produce pictures that can conclusively determine if PTK7 is localized to the nucleus. However, the AY Strongin group has succeeded in fluorescent microscopy of PTK7 in cells. They found addition of the PTK7 protease MT1-MMP to PTK7-expressing cell

lines caused PTK7 to be removed from the cell membrane and accumulate around the nucleus [117].

Summary of PTK7 Known and Predicted Features

In our discussion of PTK7 protein we have mentioned many structural and predicted features. Figure 3-6 summarizes and marks the features of PTK7 primary sequence mentioned in the text for clarity.

MGAARGSPARPRRLPLLSVLLLPLLGGTQTAIVFIKQPSSQDALQGRRALLR[1.CEVEAPGPVHVYWLLD
GAPVQDTERRFAQGSSLSFAAVDRPQDSGTFQCVARDDVTGEEARSANASFNIKWIEAGPVVLKH]PASEA
EIQPQTQVTLR[2.CHIDGHPRPTYQWFRDGTPLSDGQSNHTVSSKERNLTLRPAGPEHSGLYSC]CAHSA
FGQACSSQNFTLSIADESFARVVLAPQDVVVARYEEAMF[3.HCQFSAQPPPSLQWLFEDETPITNRSRPP
HLRRATVFANGSLLLTQVRPRNAGIYRC]IGQGQRGPPIILEATLHLAEIEDMPLFEPRVFTAGSEERVT
[4.CLPPKGLPEPSVWWEHAGVRLPTHGRVYQKGHELVLANIAESDAGVYTC]HAANLAGQRRQDVNITVA
TVPSWLKKPQDSQLEEGKPGYLD[5.CLTQATPKPTVVWYRNQMLISEDSRFEVFKNGTLRINSVEVYDGT
WYRC]MSSTPAGSIEAQARVQVLEK↓LKFTPPPQPQQCMEFDKEATVP[6.CSATGREKPTIKWERADGSS
LPEWVTDNAGTLHFARVTRDDAGNYTC]IASNGPQGQIRAHVQLTVAVFITFKVEPERTTVYQGHTA[7.L
LQCEAQGDPKP↓LIQW4KGKDRILDPTKLGPRMHIFQNGSLVIHDVAPEDSGRYTC]IAGNSCNIKHTEAP
LYVVDK4PVPEESEGPGSPPPYKMIQTIGLSVGAAVAYIIAVLGLMFYCKKRCKAKRLQKQPEGEEPEMEC
LNGGPLQNGQPSAEIQEEVALTSLGSGPAATNKRHSTSDKMHFPRSSLQPITTLGKSEFGEVFLAKAQGLE
EGVAETLVLVKSLQSKDE5QQQLDFRRELEMFGKLNHANVVRLLGLCREAEPHYMVLEYVDLGDLKQFLRI
SKSKDEKLKSQPLSTKQKVALCTQVALGMEHLSNNRFVHKDLAARNCLVSAQRQVKVSALGLSKDVYNSEY
YHFRQAWVPLRWMSPEAILEGDFSTKSDVWAFGVLMWEVFTHGEMPHGGQADDEVLADLQAGKARLPQPEG
CPSKLYRLMQRCWALSPKDRPSFSEIASALGDSTVDSKP

Figure 3-6. PTK7 primary amino acid sequence with identified features marked [GenBank AAH71557.1]. Protein is 1,070aa with 118kDa full form and 68kDa cleaved fragment. Single underlined numbers and brackets: denotes each of 7 Ig-like fold domains; Grey highlighted text: area covered by the α-PTK7 M02 antibody; Pink highlight and white lettering: region I predicted could bind PTK7; MGAARGSPARPRRLPLLSVLLLPLLGGTQT: signal peptide; RPPHLRR predicted NLS; K↓LK: *chuzhoi* mutant cleavage site; G: Site of possible GGG→GGA glycine SNP; PKP↓LI: MT1-MMP cleavage site; GKSEFG and ALG: mutated regions of tyrosine kinase domain that inactivated the kinase, active kinase have GXGXXG and DFG respectively instead; TIGLSVGAAVAYIIAVLGLMFY: transmembrane domain; 4: Region between these two 4s is deleted in PTK7-4 isoform. 5: Truncation site for PTK7-5 isoform.

Cleaved PTK7 Fragment is Higher in Cellular Nuclear Fraction

Another way to determine if PTK7 is in the nucleus is to fractionate cells and probe them by western blot. Using this technique, we saw full-length PTK7 in the

nuclear and pellet fractions of human embryonic kidney cells (HEK293). There was a larger level of full-length PTK7 in the pellet fraction. Full-length PTK7 was expected in the pellet fraction, which would contain the membrane; however, the presence of full-length PTK7 in the nuclear fraction is unexplained. We saw cleaved PTK7 in all three fractions, and at much higher level in the nuclear fraction. We would expect to find cleaved PTK7 in the pellet fraction, as cleaved PTK7 has been found in complex with full-length PTK7 on the membrane [79]. The high level of cleaved PTK7 in the nuclear fraction may indicate that cleaved PTK7 is indeed present in the nucleus; however, this needs to be validated by confocal microscopy. These results were similar for the lung adenocarcinoma cell line H23, and in the cervical cancer cell line HeLa (Data not shown).

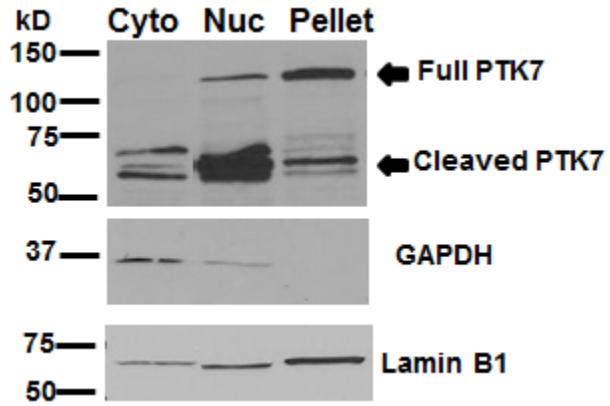


Figure 3-7. Western blot of various cellular compartments of HEK293 cells. Full-length PTK7 is ~118kDa, cleaved ~68kDa. GAPDH is a cytoplasmic marker. Lamin B1 is a nuclear membrane marker. This blot is representative of three experiments.

Sgc8c Prevents Wound Healing in HeLa Scratch Assay

As discussed in the Introduction, PTK7 has been implicated in tumorigensis, invasiveness, and metastasis. The ST Lee group found that *in vitro* addition of a soluble

fragment of PTK7 made of just the 6th and 7th Ig domains (~200aa) had profound effects on cancer-related signaling as seen by reduced wound healing in HeLa [70].

Following their work, we hypothesized that, if sgc8c binds to a region of PTK7 important for downstream signaling, then addition of PTK7 aptamers could potentially block the region of the protein and affect binding. One way to assay this would be to repeat ST Lee's wound-healing assay in the presence of sgc8c. In this assay, a monolayer of cells is serum-starved and then the surface of them is scratched. Cells migrate into the wound and heal the scratch. The speed and efficacy by which the cells heal the wound can be considered an indicator of their invasiveness. If sgc8c blocked a region important to PTK7 function, then the ability of the cells to migrate would be diminished. This is what we saw. Treatment of cells with sgc8c for 24h significantly reduced their ability to heal a wound [Figure 3-8]. This indicates that the addition of sgc8c might block a region of PTK7 important for downstream signaling events. In the future, more experiments, using transwell plates or other assays of invasiveness, could help to strengthen this conclusion.

Conclusions

Bioinformatic, competition, and mutational analysis of PTK7 aptamers revealed a consensus sequence important for aptamer binding. An analog to the PTK7 aptamers sequence was found within the 5'-untranslated region of DIXDC1b. This region of DIXDC1b was also conserved in the regulatory regions of chimpanzee, mouse, and rat DIXDC1b homologs. These results indicated there might be a natural interaction between PTK7 protein and DIXDC1b DNA. Structural modeling of PTK7 identified a region of potential aptamer binding at PTK7 Ig3. A possible scheme for this interaction is shown in Figure 3-9.

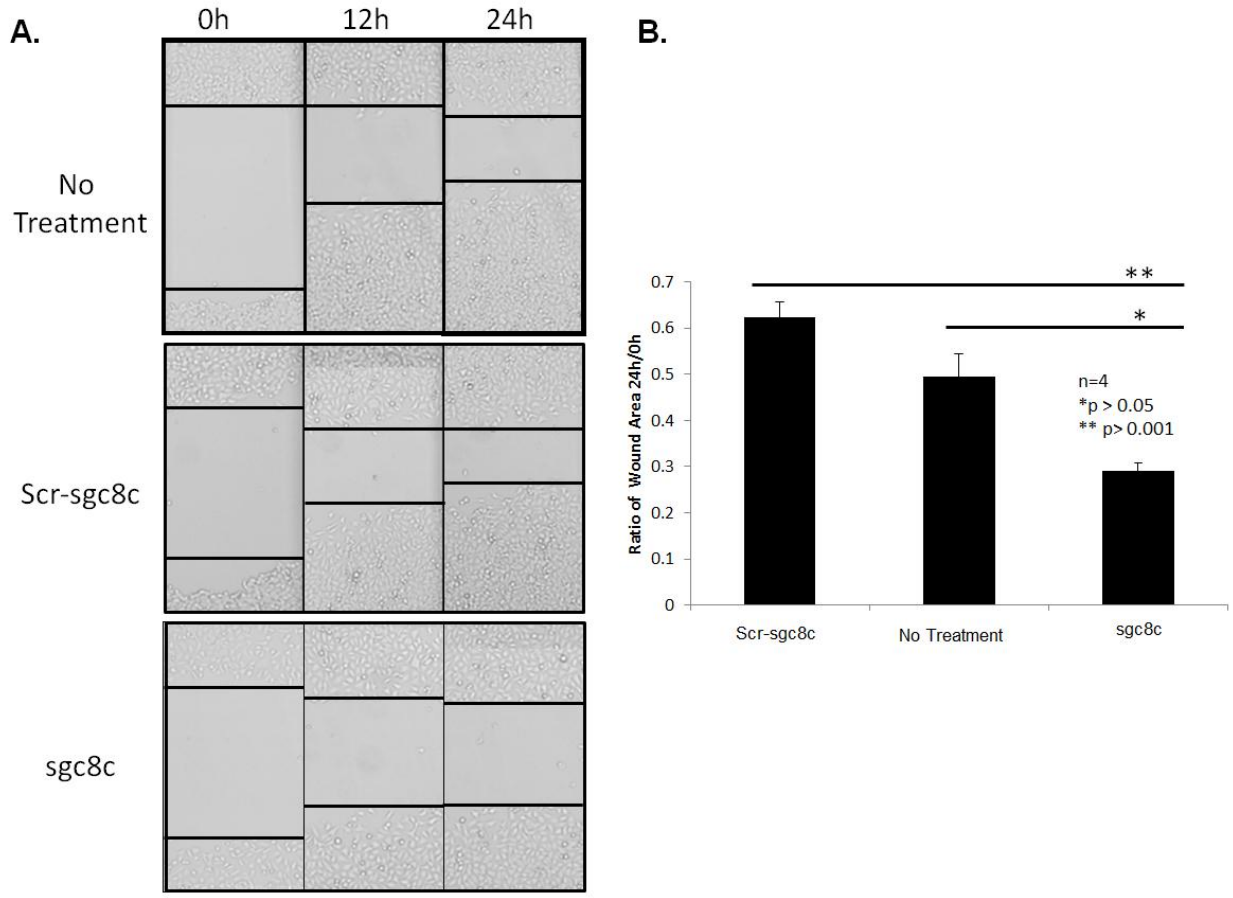


Figure 3-8. Scratch assay. A) Representative images of scratch in monolayer of HeLa cells at 0h, 12h, 24h: 10x resolution. Scr-sgc8: scrambled sgc8. B). Quantitation of scratch assay data. The ratio of wound area at 24h and 0h were determined by ImageJ software. Results are from 1 experiment with 4 replicates. Several other independent experiments had similar results. P-values determined by ANOVA.

In order for this interaction to occur, cleaved PTK7 must be transported to the nucleus. PTK7 has a weak predicted nuclear localization sequence. By cell fractionation, followed by Western blot, we found an abundance of cleaved PTK7 in the nuclei of several different cell lines. Furthermore, addition of sgc8c to HeLa cells reduced their ability to migrate in a scratch assay, indicating sgc8c may block a region important for PTK7 signaling.

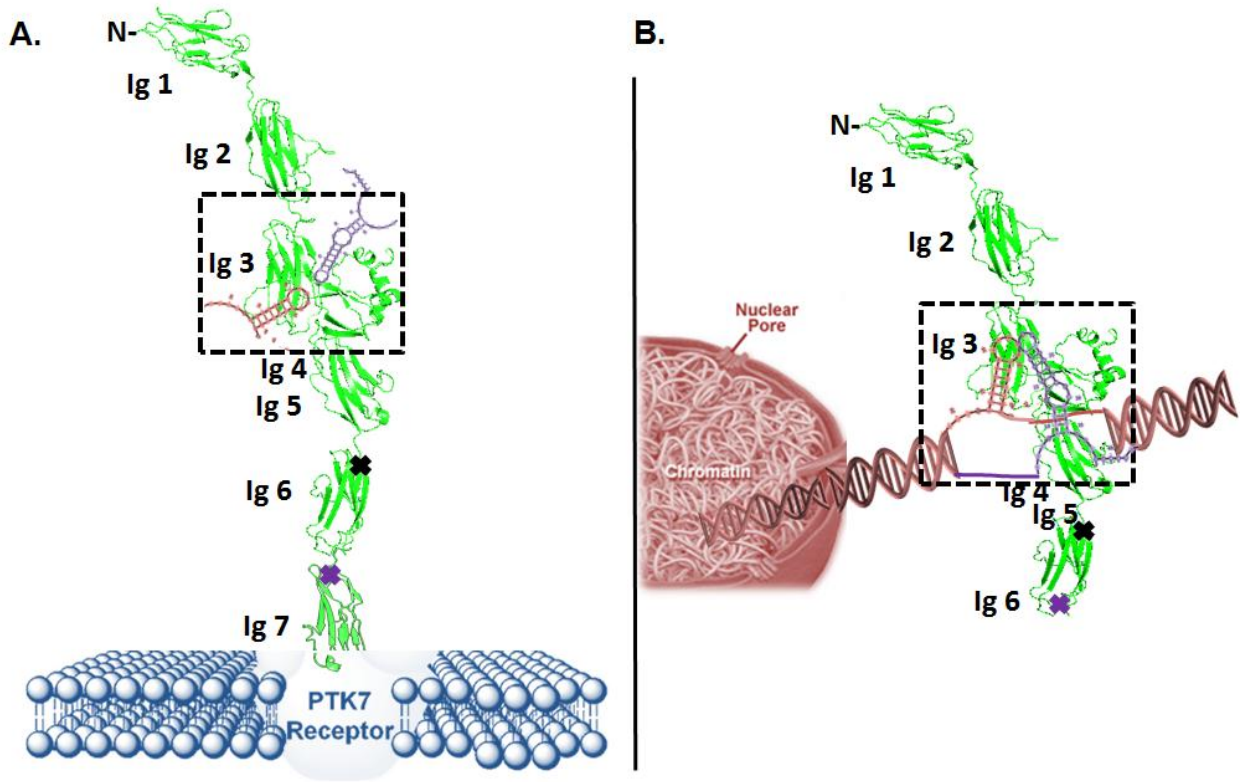


Figure 3-9. A) Scheme of possible PTK7 aptamer interactions on the surface of the cell. B) Scheme of possible PTK7 protein interaction with aptamer similar regions in DIXDC1b DNA. Pink stem-loop structure in dashed box: KC2D4 consensus sequence; Blue stem-loop structure in dashed box: sgc8c consensus sequence. Green protein structure: model of PTK7 extracellular region. The purple X denotes the natural MT1-MMP cleavage site; the black X is the additional *chuzhoi* mutant MT1-MMP cleavage site.

Using the sum of this information, we developed a model for the possible functional role for PTK7-DIXDC1b DNA interaction [Figure 3-10]. In this model, PTK7's extracellular portion is cleaved by MT1-MMP outside of the cell [1]. The soluble PTK7 fragment (sPTK7) is transported to the nucleus [2]. In the nucleus, sPTK7 binds the 5'UTR region of DIXDC1b at the sites of aptamer sequence identity [3.], and causes the modulation of DIXDC1b transcription [4.]. The altered DIXDC1b affects the amount of active Dvl and the level of JNK gene expression [5.]. These changes affect Wnt

signaling [6.]. Much work still needs to be done to interrogate each one of the steps in the model below; it is presented only as a starting point for discovery.

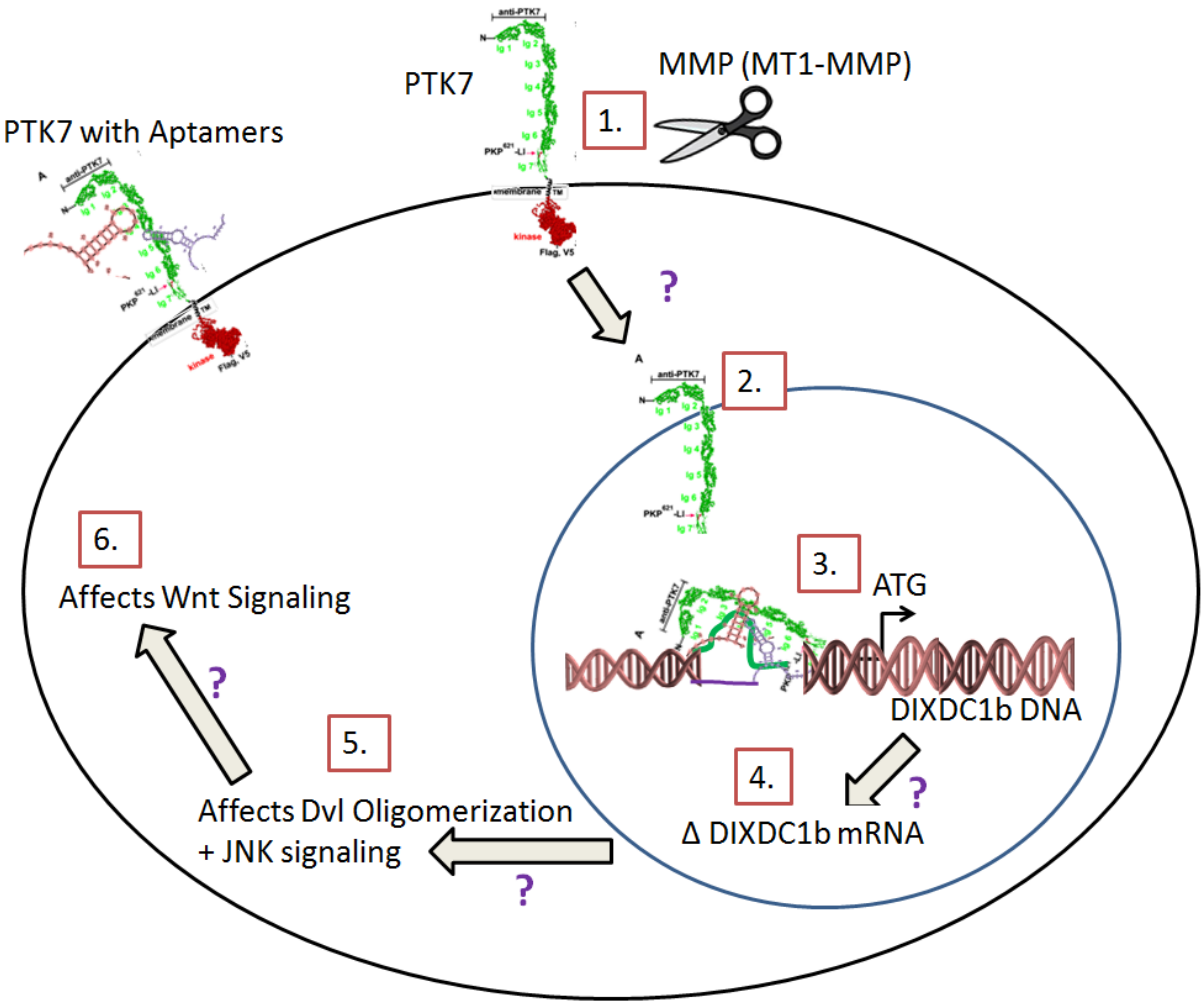


Figure 3-10. Model for PTK7-DIXDC1b interaction and its effect on Wnt signaling. Please see text for description.

Another possibility is that PTK7 interacts with the mRNA of DIXDC1b instead of the DNA and could modulate DIXDC1b by translational regulation instead of transcriptional regulation. This is not expected for two main reasons. Although DNA and RNA share similar structures, RNA has a hydroxyl group attached to every ribose. This modification has several effects; namely, it makes RNA more structurally unstable,

alters the secondary structure, and prevents the robust double-stranded conformation seen with DNA. As aptamer binding is intimately connected to ssDNA secondary structure, the perturbations caused by replacement of RNA would affect binding. In the last chapter, we saw how introduction of a single azobenzene within the consensus region decreased the aptamer binding affinity 1,000 fold. Our review of the literature has not revealed one example of an ssDNA aptamer binding a target when it is translated into RNA, or vice versa, however, this does not mean it is not possible.

The second main reason why it is unlikely PTK7 would interact with DIXDC1b mRNA instead of DIXDC1b DNA is that, when RNA is transcribed from DNA, mRNA is only made from 1 strand of the DNA. Our PTK7 aptamers have sequence identity to both strands of DIXDC1b DNA, and would presumably interact with both strands of the DNA. Another option, if PTK7 does indeed bind mRNA, could be that an miRNA interacts with DIXDC1b mRNA at the region on the negative strand of DIXDC1b RNA that has sequence identity to the KC2D4 aptamer. The PTK7 protein would then bind the DIXDC1b mRNA and its miRNA. The question of whether PTK7 binds DIXDC1b mRNA or DNA can only be conclusively answered by synthesizing the sgc8c aptamer in RNA and seeing if it binds.

Materials and Methods

DNA Sequences

DNA aptamers used in this chapter [Table 3-2] were synthesized, purified, and validated as described in Chapter 2 Materials and Methods.

Table 3-2 Sequences used in this chapter. B: Biotin; DIXcs: aptamer consensus region with DIXDC1b DNA; dsDIX-40-80: nucleotides 40-80 of the DIXDC1b mRNA sequence, positive direction; c-dsDIX-40-80: sequence complementary to dsDIX-40-80.; Scr-sgc8c: scrambled sgc8c.

Experiment	Name	Sequence 5' to 3'
	DIXcs	GCGCAGCCTGCGCCGCCGGGA-B
	dsDIX-40-80	CACTGGGAGGGAGGCTCCCGGCGGCGCAGGCTGCGCCTCG-B
	c-dsDIX-40-80	GCTGGGGACTCGAGGCGCAGCCTGCGCCGCCGGGAGCCTC-B
dsDNA Binding + Scratch Assay	sgc8c	TCTAACTGCTGCGCCGCCGGGAAAATACTGTACGGTTAGA-B
	Scr-sgc8c	ATCTAACTGCTGCGCCGCCGGGAAAATACTGTACGGTTAG-B

Cell Culture

CEM-CCRF, HeLa, and HEK cells lines were obtained from American Type Culture Collection (ATCC). CEM cells were cultured in RPMI-1640; HeLa and HEK cell lines were cultured in DMEM-1640 medium (GIBCO). All media was supplemented with 10% FBS (GIBCO) and cell were cultured at 37°C under 5% CO_2. 100 units/mL penicillin-streptomycin (Cellgro) were added to the cultures, occasionally, if bacterial contamination was suspected in the lab. Cultures were routinely monitored for *Mycoplasma* infection.

dsDNA Binding Experiments

One million CEM cells were washed with 2mL washing buffer (WB: PBS, 4.5g/L glucose, 1M $MgCl_2$), probed with 1μM of sgc8c, Scr-sgc8c, DIXcs-B, hybridized dsDIX and c-dsDIX-40-80; or hybridized dsDIX-40-80 and c-dsDIX-40-80. in 100μL binding buffer (BB: WB with 1g/L bovine serum albumin, and 100mg/L tRNA) for 20min at 4°, washed, stained with 1:400 streptavidin-PE, washed again, and measured using a FACScan flow cytometer (Becton Dickenson), counting 20k cells. The data was analyzed with FCS Express.

Aptamers were hybridized by mixing equal molar equivalents of the various DNA in H_20, which was heated to 95°C for 5min and allowed to slowly cool to room temperature. Hybridization was confirmed by running a small aliquot on a 3% agarose gel in TBE for 40min at 120V and noting a shift in molecular weight.

Western Blot

HEK293 cells were cultured in DMEM with 10% FBS under 5% CO_2. One million cells were collected and fractionated using a NE-PER kit (Pierce) following all manufacturer's instructions. They were run on a 4-20% Tris-Glycine gel (Invitrogen) and probed with α-PTK7 M02 (Abnova) 1:1000 followed by α-mouse horseradish peroxidase (HRP) 1:5000 (Pierce). The membranes was stripped and re-probed for α-GAPDH (ABcam), a cytosolic marker, and Lamin B1 1:1000 (ABcam), a nuclear marker, followed by α-Rabbit HRP 1:5000 (Pierce) and imaged on Kodak film.

Scratch Assay

The night before the experiment, a 24 well plate, fully confluent with HeLa cells (100k cells per well) was plated in the presence of no treatment, 5μM Scr-sgc8c, or 5μM-sgc8c. The next day each well was scratched with a 200uL pipet tip. Cell were washed 2x in PBS, and 100μL DMEM with 2% FBS with no treatment, 5μM Scr-sgc8c, or 5μM-sgc8c was added to each well. Images were taken at 10x magnification under bright field using a Zeiss microscope at 0h, 12h, and 24h. To ensure the same region was photographed each time the plates were pre-marked with 2 parallel lines black on top, blue on bottom. PowerPoint was used to align the images from different time points using the black and blue line guides. ImageJ was used to measure the area of the wound. Every well had the same-size region analyzed for wound area.

CHAPTER 4
ADDITIONAL SEQUENCE IDENTITIES FOUND BETWEEN APTAMERS BY CELL-SELEX BIOINFORMATIC ANALYSIS

Introduction

Thus far our discussion has centered around the PTK7 aptamers, which were found to share a common consensus region that had an analog in genomic DNA. The bioinformatic dataset that uncovered the commonality between these PTK7 aptamers included 145 other sequences from 28 independent selections (Appendix A). Among these other aptamers, there were also pairs from different selections that shared common sequence regions greater than would be expected from chance. This chapter will highlight another set of aptamers from one of these selections that share sequence identity—sequences selected against Vaccina virus and Vaccina virus infected cells. After a discussion of the Vaccina aptamers, we will briefly explore general trends surrounding the GC/AT content of the aptamers, and conclude with a brief catalog of some other aptamer pairs we have found that share suspiciously high sequence identity.

Vaccinia Aptamers

Vaccinia Virus (VV) is responsible for saving many millions of lives. It is the active constituent of the smallpox vaccine which eradicated smallpox in 1980. The virus is a roughly 350x250nm football shaped particle enclosed in a lipid membrane stolen from the host as it exits the cell. This purloined cell membrane is studded with many viral proteins, including, hemagglutinin (HA). HA is a nonessential gene for VV replication, and is known to interact with other VV encoded proteins such as VP37K and SPI-3 late in VV infection. HA interactions with these proteins prevent cell fusion. HA is

also expressed early in VV infection, but its function at early infection time points is unclear.

Due to the need for rapid detection methods for viruses in the pox family, two independent aptamer selections were conducted in the Weihong Tan lab using VV infected cells as the target [118,119]. These selections used different virus infected cell lines, different primer pairs and libraries, and were carried out several years apart by different people. Parag Parekh determined HA was the aptamer target for his selection of PP5 aptamer. In addition to these selections several years previously, the Andreas Kage group in Germany performed a ssDNA selection directly against VV particles themselves, in a one-step selection using a VV particle affinity column, in a procedure they termed MonoLEX [120].

Surprisingly, bioinformatic analysis of our ssDNA aptamer dataset revealed a common region of sequence identity in aptamers selected by these three independent aptamer selections [Table 4-1]. Structural prediction of aptamer secondary structure by NuPACK software showed all the sequences formed a strong stem-loop structure with the consensus region found in the loop of each structure [Figure 4-1]. Based on these results, I predict all four sequences will bind to the HA, and these aptamers will compete against each other.

Table 4-1. Aligned aptamer sequences from three Vaccinia related selection. Black highlighted sequence is shared between all the aptamers. SELEX target either VV infected cells: HeLa or A549, or target VV particles. Note: TOV2, PP5, and PP3 only show the aptamer-variable region for clarity. Please refer to Materials and Methods for full sequences.

Aptamer	Sequence	SELEX Target
TOV2	CACTTGCATATACACTTTGCATTATAGGG	HeLa
PP5	CCTGCATATACACTTTGCATGTGG	A549
PP3	CGAGCCAGACATCTCACACCTGTTGCATATACATTTTGCAT	A549
A38	TACGACTCACTATAGGGATCCTGTATATATTTTGCAACTAATTGAATTCCCTTTAGTGAG	VV

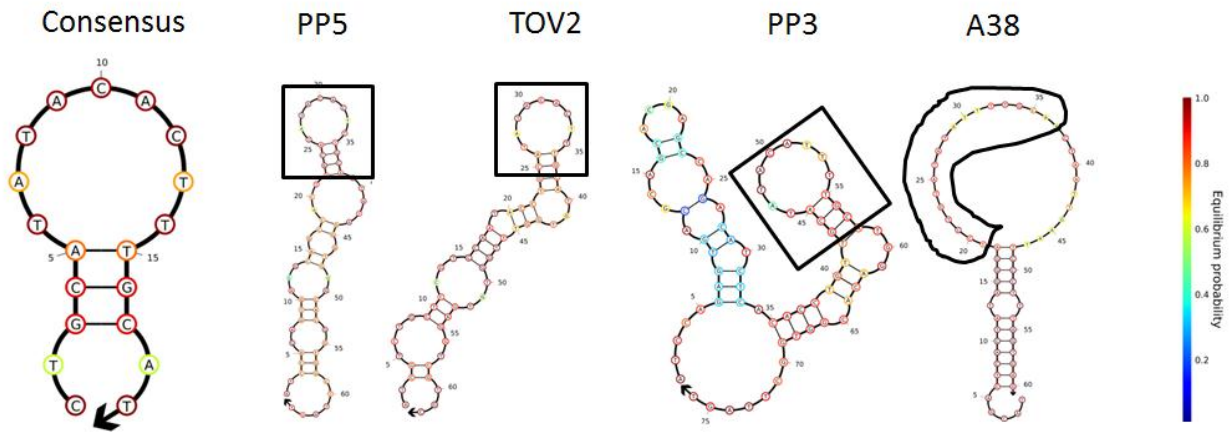


Figure 4-1. Predicted secondary structures of consensus sequence and each of the full aptamer sequences. The bar on the right indicates Bar on right shows strength of predicted structure. Red being most strong.

The selections that produced these aptamers for VV involved two different infected cell lines and a third selection on the virus particle alone. They were the product of three entirely different SELEX procedures carried out in two labs, on two different continents years apart. As with selection of the PTK7 aptamer, the chances of selecting aptamers against the same target multiple times were remote; even more so because HA is only a minor protein in the VV particle. It is not by any means the most highly expressed protein. Why was this happening?

BLAST of the Consensus Sequence against the Human Genome

VV is a dsDNA virus. Furthermore its DNA replication and particle maturation takes place entirely in the cytosol of the host cell. As a result, VV proteins, including HA which is expressed early, intermediately, and late in VV maturation, could have an opportunities to come in contact with the VV genome. Therefore we hypothesized there might be a natural, physiological role for these sequences and possibly HA interacts directly with the viral genome. To determine if this was indeed the case, we performed a

BLAST search for the 19nt consensus region of the aptamers: TGC ATA TAC ACT TTG CAT. The loop part of the consensus sequence was found twice in the VV genome in the coding regions these two proteins: D3R, GenBank and NPH-II, GenBank: AAA48064. Neither of these proteins has any known direct link to HA.

Table 4-2. BLAST results of VV genome revealed two proteins D3R and I8R share DNA sequences in common with the VV aptamers. Black highlighted sequences are shared between all the aptamers and VV DNA. Purple highlighted sequence is additional sequence identity found between D3R and I8R.

Aptamer	Sequence
TOV2	CACTTGCATATACACTTTGCATTATAGGG
PP5	CCTGCATATACACTTTGCATGTGG
PP3	CGAGCCAGACATCTCACACCTGTTGCATATACATTTTGCAT
A38	TACGACTCACTATAGGGATCCTGTATATATTTTGCAACTAATTGAATTCCCTTTAGTGAG
VV DNA	
D3R	TTGTCATTTATGATA----TATACACTTTTGACGCTTTCAAGAA
I8R	GATGATGATACGTATATACACTTTTGTAAAATATTATTCG

Conclusions

While this story about VV aptamers is incomplete, there is much intriguing evidence that more analysis might uncover a functional role for the aptamer sequence identity. Perhaps, HA is a naturally DNA binding protein. PTK7 and VV aptamers is not the end of sequence identities produced from SELEX bioinformatic analysis. Below are the aligned sequence from two selections performed in different labs against a neuronal cancer cell line (PC12 [121]) and a mesenchymal stem cell line (20MSC [122]). These two aptamers share an 18/20nt contiguous sequence. Unfortunately the protein targets for these aptamers have not yet been discovered, but when they are, I would bet these two aptamers bind the same target; perhaps for a functional reason. There are 3 other pairs of aptamers from our dataset that share similar levels of sequence identity.

```
10              ------GCTGGGGTGTTGGGTGTGGGGGTGA------------------
20MSC           GAGTAATGTAGGGTGAAGGGTGTGGGGGCTATGGGGATAGTGGCACGGCC
                       *.*****::***********  *
```

Figure 4-2. Sequence identity between two aptamers with unknown targets selected in different labs on different cell lines.

CHAPTER 5
SGC8C-APTAMER AND PTK7-ANTIBODY RUPTURE FORCES ARE COMPARABLE ON LIVE HELA CELLS

Introduction

Aptamers and antibodies rely on a similar set of non-covalent interactions to bind their targets—hydrogen bonding, electrostatic interactions, and van der Waals forces, except antibodies can use hyrophopic interactions as well. Even though aptamers are much smaller that antibodies, i.e., 8-15kDa versus 150kDa antibodies, and even though they are made of only 4 different nucleic acid bases compared to the 20 amino acids of antibodies, surprisingly, both antibodies and aptamers can have very strong non-covalent interactions, producing binding affinities in the low nanomolar K_d range.

Due to their similar ability to target a wide range of proteins, yet their very different structural natures, there has been some concern that aptamers may not bind as robustly as antibodies. To address this issue, we used single-molecule atomic force microscopy (AFM) to measure the rupture force between a protein and its respective aptamer and antibody.

As we have mentioned before, the pseudokinase protein tyrosine kinase 7 (PTK7), is important in development [13,32,123] and cancer [67,68,70]. Both a monoclonal antibody, α-PTK7, and a DNA aptamer, sgc8c, have been identified that exhibit strong binding to receptor PTK7 with a K_d in the subnanomolar range [6]. The specific interactions of protein receptor PTK7 with its two ligands have been confirmed by siRNA silencing and PTK7 plasmid insertion [68].

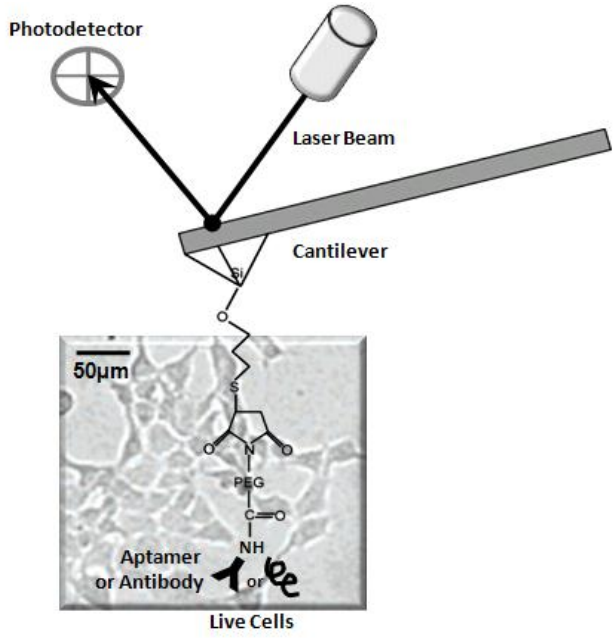


Figure 5-1. Scheme of AFM measurements on live cells.

With the goal of using sgc8c for targeted delivery of chemotherapy to tumors, several questions surrounding sgc8c-PTK7 interactions needed to be answered: What is the force of interaction between the protein and the aptamer? And how does that force compare with the force for the antibody? A straightforward and novel way to answer these questions is via examination of aptamer-target adducts using microscopes capable of single-molecule tracking on live cells. In this study we used the single-molecule AFM technique developed by XH Fang et al. [124] to compare the rupture forces between PTK7 with its aptamer, sgc8c, and its antibody α-PTK7. The experimental scheme is shown in Figure 5-1.

Results and Discussion

The rupture force was measured using AFM tips functionalized with either sgc8c aptamer, α-PTK7 antibody, or thrombin antibody as a control (there is no thrombin on

the HeLa cell surface) to scan cell membrane surfaces for changes in the force curve that indicate a binding event. Three typical force curves are shown in Figure 5-5.

Blocking Controls

To ensure that actual binding was observed, several binding probability controls were performed [Figure 5-2] Blocking by addition of excess sgc8c aptamer significantly diminished the number of rupture peaks observed with HeLa cells from 13.2±1.3% of the total to only 4.7±0.9%. Likewise, when excess α-PTK7 antibody was used to block the anti-PTK7 tip, the binding probability decreased from 11.5±2.0% to 4.2±0.7%. The probability after blocking was only slightly higher than the background controls when a thrombin modified tip was used (2.3±0.6%). This result indicates that there is specific binding between the sgc8c and α-PTK7 modified AFM tips and HeLa cells.

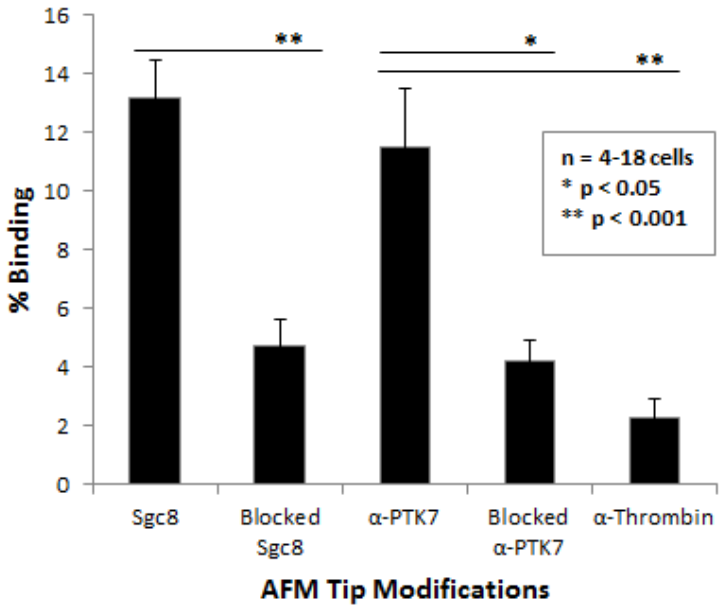


Figure 5-2. Binding probabilities of tip with cells. Obtained using different blocking, cell, and tip modification conditions. 1,000 data points were recorded per cell.

Fitting the Collected Data

At least 100 force curves were recorded to construct force histograms, which were then fitted using Gaussian peak functions [Figure 5-3]. The single-molecule rupture force between sgc8c and HeLa PTK7 was found to be 46±26pN, while the rupture force between anti-PTK7 and HeLa PTK7 was 68±33pN with the error representing 1σ.

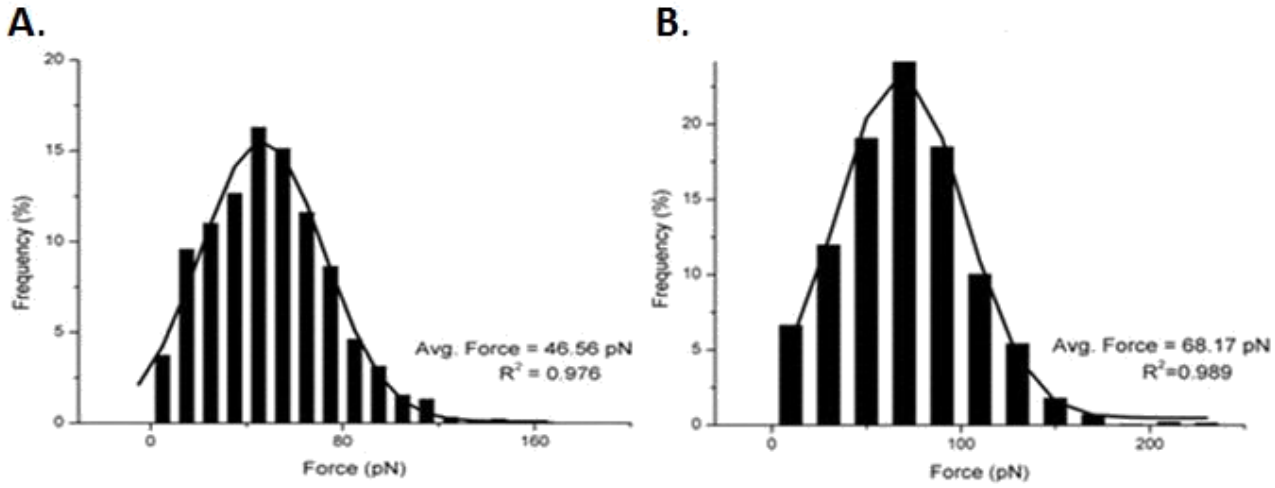


Figure 5-3. Histograms of binding forces between tips and HeLa cells. Bars: experimental data; solid line: Gaussian fit which reveals the most likely rupture force, and its respective theoretical Gaussian distribution curve. A) Binding force between sgc8c and HeLa cells. B) Binding force between α-PTK7 antibody and HeLa cells.

Conclusions

For the first time the single-molecule rupture force between a protein, PTK7, and both an aptamer and an antibody was measured using single-molecule AFM on live cells. The single-molecule rupture force between sgc8c and HeLa α-PTK7 was found to be 46±26pN, while the rupture force between α-PTK7 and HeLa PTK7 was 68±33pN. Blocking, by addition of excess sgc8c aptamer or α-PTK7 antibody,

significantly diminished the number of rupture peaks observed, indicating that specific binding between the sgc8c or α-PTK7 modified AFM tips and HeLa cells was being observed. The measured force values between the two are very similar, indicating that, despite the differences in size and component diversity, they have similar binding profiles, and that DNA aptamer interactions with proteins can be as robust as those of antibodies in terms of rupture force.

Interestingly, when we compare these results to previous AFM studies that have looked at rupture forces between hybridized dsDNA [Figure 5-4], we find the mean value for sgc8c-PTK7 interaction live cells equates to 20-30bp dsDNA. Thus, even though the consensus region of sgc8c, important for binding, is only 14bp's long, the aptamer rupture force could be commensurate with 20-30bp interactions or 45-60 hydrogen bonds within the error [125]. This indicates there might be unconventional, non-hydrogen binding interactions occurring between the PTK7 and sgc8c.

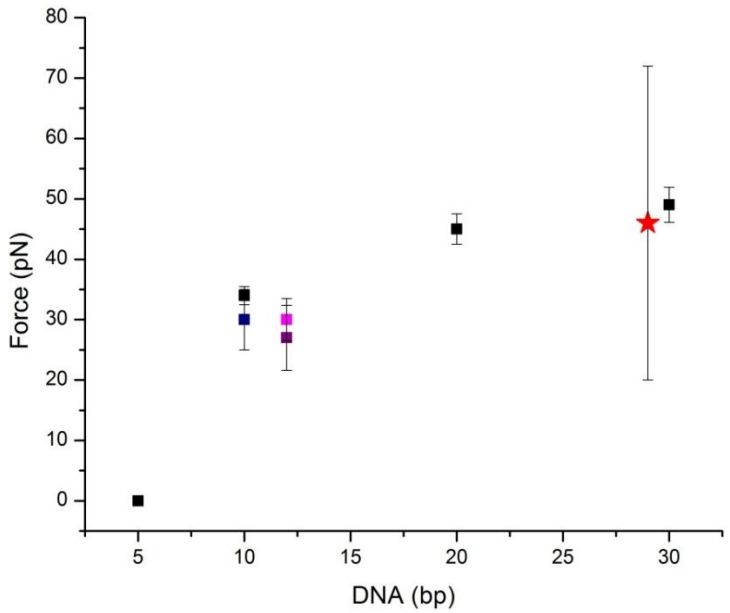


Figure 5-4. Rupture forces from literature of different lengths of dsDNA and sgc8c. Black squares [126]. Blue square [127]. Purple square [128]. Pink square [129]. Red star: Sgc8c rupture force.

Materials and Methods

Preparation of AFM Tips and Substrate

The natural oxide layer on the AFM silicon nitride (Si3N4) tips (type: NP, from Veeco, Santa Barbara, CA, USA) was removed by etching the substrates in HF for 30s at room temperature in a hood with proper protective equipment. The tips were then dipped in alkaline solutions of NH4OH:H2O2:H2O, 1:1:5 v/v for 30min, followed by oxidation in 90ºC piranha solution with 98% H2SO4:H2O2, 7:3 v/v for 30min. Between each of these steps the tips were washed well with water. To functionalize the tips, they were transferred into a solution of (3-mercaptopropyl)-trimethoxysilane (MPTMS) 1% v/v in toluene and incubated for 2h at room temperature, followed by thorough rinsing with toluene to remove any unbound silane. The tips were activated by incubating in 1mg/mL ω-N-hydroxysuccinimide ester-poly(ethylene glycol)-α-maleimide (NHS-PEG-MAL; Nektar Therapeutics) in DMSO for 3h at room temperature, followed by extensive rinsing with DMSO to remove unbound NHS-PEG-MAL. Finally, the tips were immersed for 30min in 2mg/mL of either α-Thrombin antibody as a control (Haematologic Technologies Inc, VT); α-PTK7 (Miltenyi Biotec, Germany); or in 100nM aptamer in Tris-HCl buffer. The sgc8c aptamer with 9 linker thymines (5’ NH2-TTT TTT TTT ATC TAA CTG CTG CGC CGC CGG GAA AAT ACT GTA CGG TTA GA 3’) was synthesized in-house on an IDT DNA synthesizer and purified using a reverse phase Prostar HPLC (Varian) on a C18 column (Econosil, 5U 250 x 4.6 mm from Alltech Associates) with a linear elution. These sequences were vacuum dried or ethanol precipitated, detritylated with 20% acetic acid, and stored at minus 20°C for future use.

The absorbance at 260nm for the purified aptamers was quantified using a UV-Vis spectrophotometer (Bio-Rad). Aptamer concentration was determined using the

Lambert-Beer equation: Abs 260nm = εbc, where ε is the extinction coefficient, b is the cuvette pass length, and c is the DNA concentration. The sequence was tested for quality control by running 0.5μL of purified aptamer on a 3% agarose gel in TBE buffer (89mM Tris-HCl, 89mM boric acid, 2mM EDTA, pH 8.0) for 40min at 120V and imaged under UV with Ethidium Bromide (EtBr).

The density of protein on the AFM tips was sufficiently low to ensure that only one protein or DNA molecule was measured at a given time. After rinsing in buffer, the tips were stored in buffer at 4ºC until use. The tips were calibrated using the thermal fluctuation method for the range of 0.040-0.075 N/m.

Cell Culture

All cells used in our experiments were cultured in incubators at 37ºC with 5% CO_2. The day before use, HeLa were subcultured into 35mm cell culture dishes at 70% confluency in 2mL Dulbecco's Modified Eagle Media (DMEM), supplemented with 1% penicillin/streptomyocin (Sigma) and 5% Fetal Bovine Serum (Gibco). Just before use, the cells were washed three times with 2mL phosphate buffered saline (PBS; Sigma), and then 2mL DMEM was added. For blocking experiments, 1μM sgc8c aptamer or αPTK7 antibody was added to the cells in serum-free DMEM, followed by incubation for 10min prior to imaging.

AFM Measurements

All AFM force measurements were performed with a Nano Scope III AFM (Veeco, Santa Barbara, CA) with live cells in plates filled with freshly prepared DMEM buffer. The force curves [Figure 4-5] were recorded and analyzed by Nanoscope 5.30b4 software (Veeco, Santa Barbara, CA), and statistics were calculated using Origin 8.5 software. Seven AFM tips were measured in total: 3 modified with Sgc8c, 3 modified

with PTK7 antibody, and 1 modified with thrombin protein as a control. Data from 4-6 cells were collected for each tip with and without blocking. For each cell, 1,000 data points were collected from at least 50 different locations on the cell membrane. Each measurement took no longer than 1h.

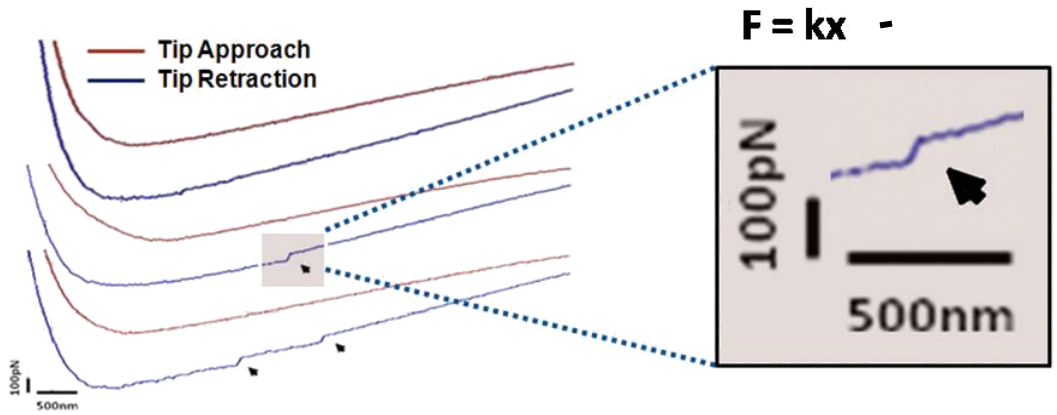


Figure 5-5. Representative force-distance curves for sgc8c AFM tip and HeLa cells.Red curve is tip approach; blue curve is tip retraction. Top curve pair (red and blue) shows no binding, while black arrows on the bottom two pairs point to the measured binding. Expanded inset shows how measurements were made. Hooke's Law is on top of the inset, which says the force, *F*, is directly proportional to spring constant, *k*, and spring displacement, *x*.

CHAPTER 6
MODIFYING CELLULAR PROPERTIES USING ARTIFICIAL LIPID APTAMER-RECEPTORS

Introduction

Cell membranes are a central hub for receiving signals from outside the cell and transmitting those signals into the cell itself. Phospholipids make up most of a cell's membrane surface area, keeping its integrity intact. These phospholipids are studded with numerous proteins acting as recognition moieties, transporters, receptors, or structural connectors. Membrane proteins are vital to many cell functions and can control cell binding, fate, and signaling. Thus, the ability to choose which proteins are present on a cell's surface could be an important tool for modulating cell behavior.

A number of methods have been used to alter which proteins are found on the cell surface. These include the use of recombinant proteins [130], the attachment of proteins to lipids which then insert themselves into the membrane, such as glycosylphosphatidylinositols anchored proteins [131,132,133], NHS functionalized poly (ethylene glycol) oleyl derivatives [134], and palmitated protein A complexes [135]. Covalently coupling proteins to the surface by chemically modifying them with azide [136] or another chemistry has also been reported. However, these approaches all have shortcomings. Recombinant strategies are time-consuming and require optimization to produce proteins that are reliably trafficked to the membrane. Insertion of lipid-functionalized proteins, or covalently linked proteins, cannot be modulated, nor can they be strategically placed on the cell membrane. Moreover, covalently bound proteins can potentially interfere with the protein's function.

In addition to proteins, groups have also linked cDNA to the surface of cells via lipid attachment [137] or broad chemical modification of the cell surface [138]. These

modifications were made in order to attach cells to surfaces or to build artificial cell architectures for tissue engineering. Building on these attempts, we sought to create artificial aptamer-aptamer receptors that capture proteins onto the cell surface in a rapid, reversible, and dose-controllable manner. Aptamers are short DNA or RNA sequences, typically 15 to 80 nt in length. Selected from large libraries of many different sequences, aptamers bind to a particular target through a process known as SELEX. While an aptamer's affinity for its target is often comparable to antibodies, with dissociation constants (K_d) in the low nanomolar to micromolar range, aptamers have many advantages over antibodies, including increased stability, small size, simple chemistry for easy functionalization, reversible binding after addition of an aptamer's cDNA, and flexible target recognition. Previously, our lab synthesized a diacyllipid phosphoramidite [46], a nucleoside building block that has two long saturated fatty acid chains held together with a glycerol. This lipid end can form micelles which, when added to cells, can insert into cell membranes. This diacyllipid phosphoramidite can easily be attached to the 5' end of any synthesized oligonucleotide. Thus, in theory, any aptamer can be easily functionalized with the lipid, and when added to cells, the lipid will anchor the aptamer in the membrane where it will protrude from the cell, thereby providing target -binding capability. Our lab has previously made TD05 aptamer-micelles, where the aptamer TD05 recognized the B-cell leukemia Ramos cell line[139]. A dye, which was loaded into the micelle interior, was specifically delivered to Ramos cells by these micelles.

Streptavidin (SA) and thrombin are two proteins having unique functions which, when attached to the cell surface, can dramatically modify cell behavior. SA is a

tetravalent protein that binds the small molecule biotin with a very high affinity $K_d > 10^{-14}$, making SA a useful tool in cell biology, especially when a rapid, specific interaction is needed. Many different fluorescently labeled SAs are available, as well as several aptamers that bind SA with high affinity.

Thrombin is an enzyme important in the clotting of blood. In blood bound to cell membranes, it exists in an inactive, prothrombin, state. When activated by Factor Xa, it cleaves soluble fibrinogen protein into insoluble fibrin fragments. These fibrin fragments agglomerate, are covalently linked through their lysine and glutamine side chains by Factor XIII, and promote clot stabilization. Thrombin is a complex protein with several different active sites; in addition to cleaving fibrinogen, it also activates protein C, platelets, cell thrombin receptors, and Factors V, VII, and VIII, which regulate thrombin production. Thrombin has two very well-studied aptamers: a 15nt aptamer (T-15), which binds thrombin's fibrinogen cleavage site, and a 27nt aptamer (T-27), which binds exosite 2, which is important for heparin binding. In this work, we will make aptamer-receptors for streptavidin and thrombin. We will show that cells modified with the receptors are able to capture proteins on their surface in a manner that is rapid and retains the protein's activity.

Results and Discussion

Streptavidin Aptamer-Receptor Anchors on the Cell Membrane

In the first set of experiments, cells were modified with streptavidin aptamer-receptors (SA-ARs), enabling them to capture fluorescently labeled streptavidin. The captured SA stained the cell membrane in a dose-controllable manner. Streptavidin (SA) is a tetravalent protein known for its high affinity toward the small molecule biotin. SA-ARs were made by attaching a lipid tail to a 29nt aptamer that binds streptavidin (40nM K_d)

[140]. To confirm that SA-ARs retained their binding ability to SA, FITC-labeled SA-aptamers were competed off of SA-coated magnetic beads by SA-AR (data not shown). All cell lines tested were able to capture Alexa-488-labeled SA (SA-488) on their cell membranes after insertion of SA-ARs [Figure 6-1a].

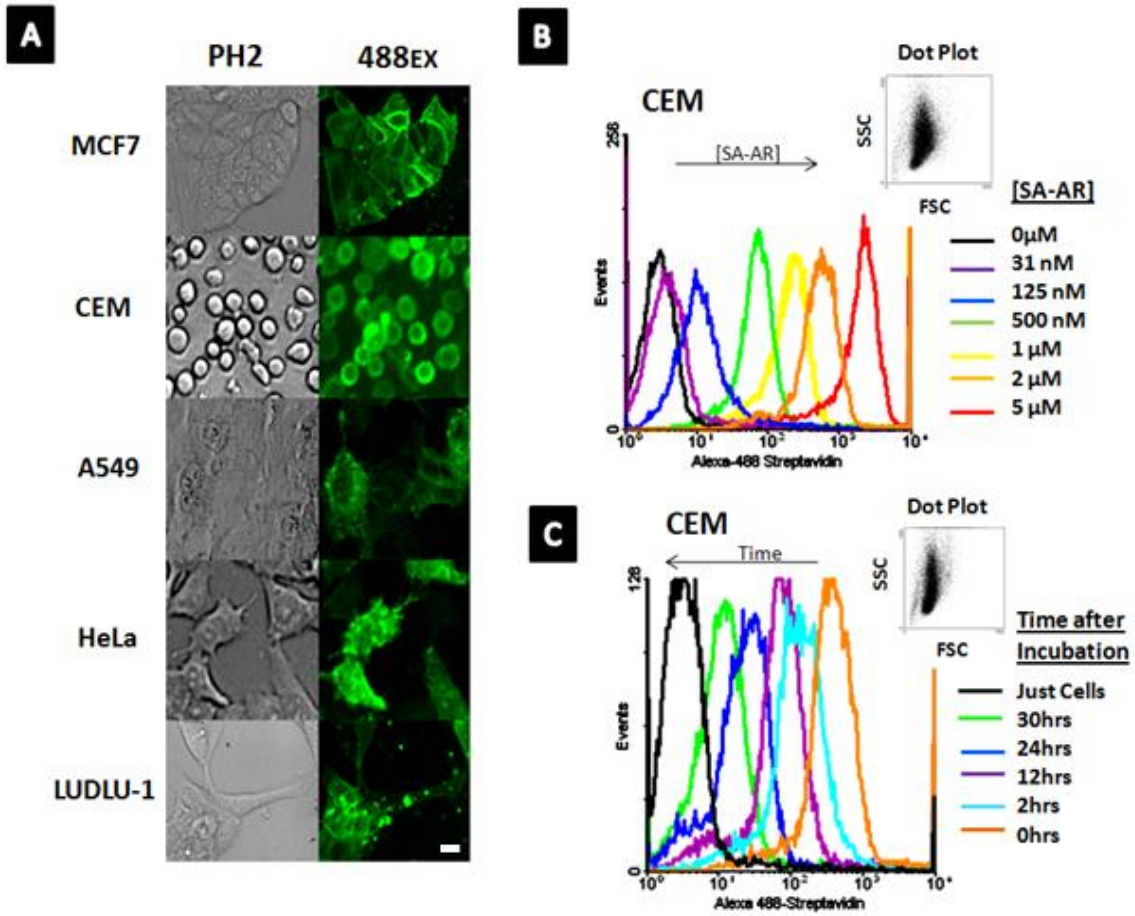


Figure 6-1. Characteristics of streptavidin-artificial receptors. (A) Artificial receptors can be inserted in many different cell types. Cells were incubated with 2µM SA-ARs for 2h at 37°C, followed by addition of 1:400 SA-488, and imaged using bright field and fluorescence (488ex) microscopy. (B) CEM cells were incubated with different concentrations from 31nM to 5µM SA-ARs for 2h at 37°C, followed by addition of 1:400 SA-488. SA-488 on the cell surface was analyzed by flow cytometry using channel 1 for SA-488. (C) CEM cells were incubated with 2µM SA-AR for 2h at 37ºC, followed by washes and incubation at various times from 0 to 30h before addition of 1:400 SA-488. Following treatment, the cells were analyzed by flow cytometry. All histograms are ungated. Scale bar is 10µM.

Each cell type had a unique staining profile relative to (1) the amount of SA-488 incorporated: CEM and HeLa were high, while A549 were lower; (2) the smoothness of staining: HeLa and A549 were punctate, while CEM, Ramos, and Ludlu-1 were smoother; and (3) the predominant location of staining: junctions were between cells as in MCF7 or were on the main body of the cell as in HeLa and A549. These differences probably resulted from differences in the fluidity and topology of each particular cell's plasma membrane and the ease of SA-AR insertion. While SA-488 was captured with ease, phycoerythrin (PE)-labeled SA could not be captured on the cell surfaces by the SA-ARs, possibly because PE is larger than Alexa-488, and PE sterically interferes with the aptamer's binding site on SA.

Incubation of CEM cells with different concentrations of SA-AR resulted in the dose-dependent capture of SA on the cell surface [Figure 6-1b]. Specifically, incubation with as little as 31nM of SA-AR, which is below the K_d for the SA aptamer, was enough to detect SA-488 on the cell surface with flow cytometry. Increasing the SA-AR concentration increased the amount of SA-488 signal in a linear fashion until it plateaued at around 5μM. SA-ARs persisted on the cell membrane for an extended time, but the amount of aptamer slowly decreased over two days after incubation [Figure 6-1c]. After two days, fluorescence became undetectable, indicating SA-AR modification is temporary, and cells return to normal after being cultured for two days. Furthermore the insertion of SA-ARs is rapid, after addition to cells, detectable levels of SA-AR insertion were apparent within 5min, and reached saturation levels within an hour [Figure 6-2].

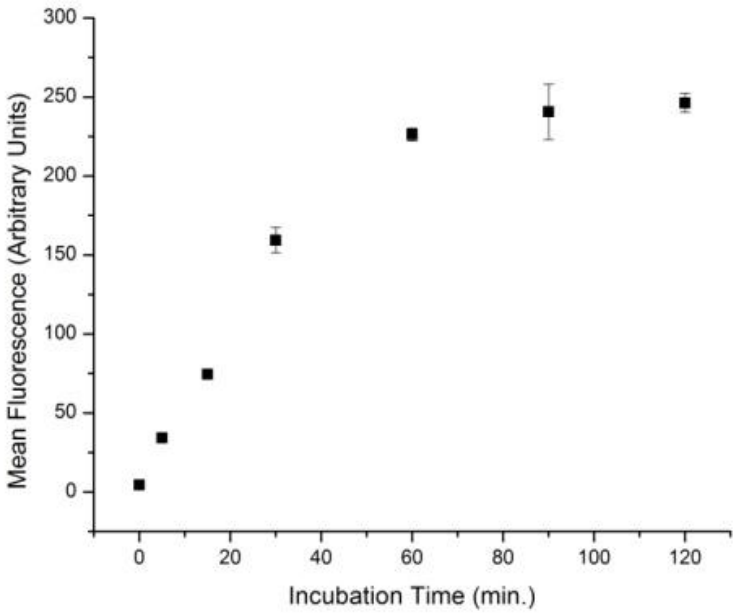


Figure 6-2. Aptamer insertion begins after 5min and reaches saturation after an hour**.** CEM cells were incubated with 1µM SA-AR for different times from 5-120min, after which cells were washed and stained with SA-488.

To demonstrate that the SA-aptamer was responsible for localizing SA on the membrane, we attached SA-488 to the cell surface at 4ºC. However, when the temperature was increased to 37ºC, the aptamer did not bind, and we lost all SA-488 signaling on the cells. The fluorescence measured by flow cytometry for treated cells went from 393 arbitrary fluorescent units (AFU) at 4ºC to 3 AFU after 30 minutes at 37ºC, which was the same as background, unmodified cells. These results demonstrate that the binding of SA-AR to its target SA-488 can be modulated by altering such environmental conditions as temperature. Furthermore, SA-ARs insertion had no effect on cell proliferation as measured by MTS [Figure 6-3]. These results indicate SA-AR treatments do not negatively affect cell growth, an important criteria for future applications.

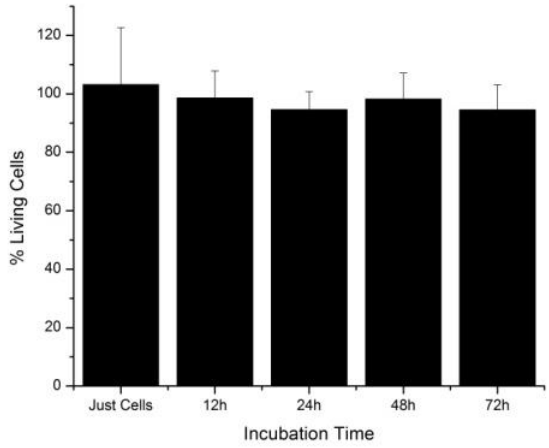


Figure 6-3. SA-AR does not inhibit cell growth**.** 20,000 CEM cells were incubated with 2µM SA-AR in 100µL media over 72h. After which proliferation was determined by MTS.

Streptavidin on Modified Cells Remains Functional

Proteins captured by aptamer-receptors need to remain functional to be useful. To confirm the active, i.e. biotin-binding, state of SA captured by SA-ARs, a sandwich assay was performed, as seen in Figure 6-4a. SA is tetravalent, with four distinct biotin binding sites. Therefore, even though one site is taken up by the aptamer, there are theoretically three other sites on each SA-488 capable of binding.

In this assay, two different suspension cell lines that do not naturally bind to each other, were induced to bind each other when a biotinylated aptamer that recognized the other cell type was added to SA-modified cells. This aggregation did not occur when the control aptamer, which only recognized the modified cell type, was added. Sgc8 is an aptamer that binds a protein called PTK7, which is found on CEM cells, but not on Ramos cells. TD05 is an aptamer that binds IgM, which is found on Ramos cells, but not on CEM cells. In the first case, CEM cells were modified with SA-AR and then coated with SA-488. Biotinylated TD05 aptamer was then added to the cells, and after washing, a 5x number of Ramos cells were mixed in. Following this, aggregates of cells “glued”

together by the SA-modified CEM cells were formed. However, for the control, when biotinylated Sgc8-aptamer, which only binds CEM cells, was added, instead of TD05, no aggregation was seen [Figure 6-4b, TOP]. In the second case with Ramos, the same phenomenon occurred, but in reverse [Figure 6-4b, BOTTOM]. These experimental findings indicate that the SA-488 modified on the surface of the cells remained functional, and able to form complex assemblies.

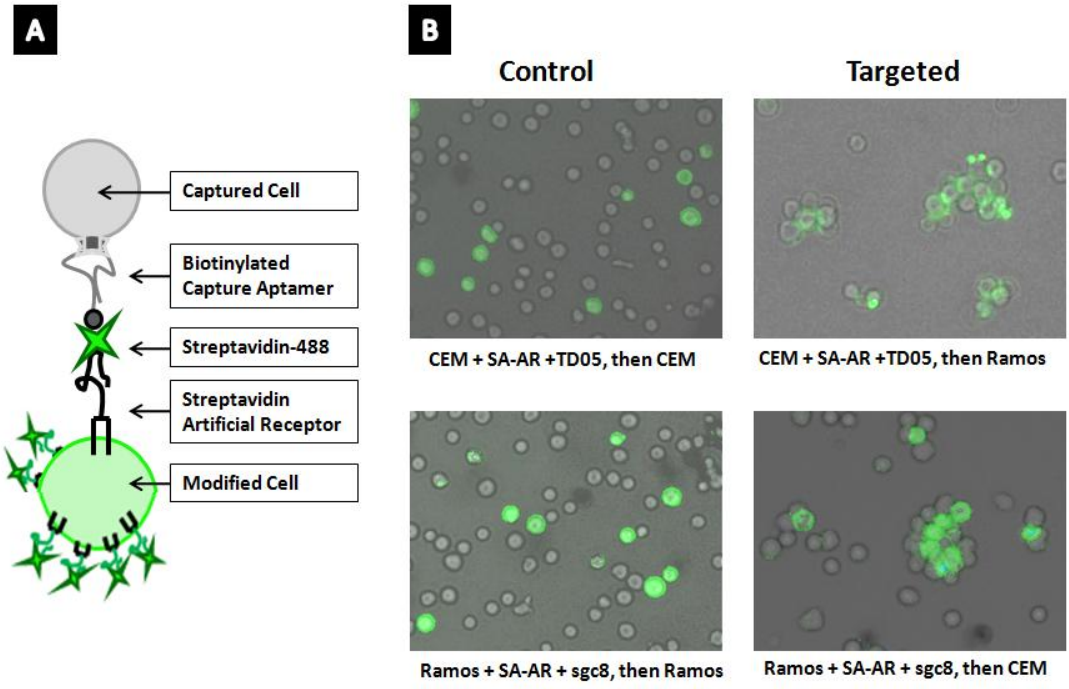


Figure 6-4. Streptavidin-modified cells bind biotin to make cell assemblies.
(A) Scheme showing the cellular assembly. (B) Top: CEM cells modified with streptavidin-488 via SA-AR incubated with TDO5 aptamer that binds Ramos cells and then either CEM (control, left) or Ramos (target, right). Bottom: Ramos cells modified with streptavidin-488 via SA-AR incubated with sgc8 aptamer that binds CEM and then either Ramos (control, left) or CEM (target, right).

SA-ARs Can Be Used to Capture Cells for Analysis

In order to determine how robust the SA-AR system was, and to show we could isolate SA-AR modified cells, we captured SA-AR functionalized cells with streptavidin coated DynaBeads [Figure 6-5]. In this experiment, CEM cells were modified with either the control, PDGF-AR, or the target, SA-AR, aptamer-receptors. When they were mixed with streptravidin-coated magnetic beads in a buffer, the cells were enriched on the beads. After lysing the cells and probing them via western blot against β-actin, the positive controls, either sgc8-biotin bound cells (well 3) or pure cell lysate (well 7) had β-actin. In addition SA-AR modified cells (well 5) were able to enrich the cells in buffer and consequently had β-actin. The wells with the untreated cells (well 1), TD05-biotin which does not bind CEM (well 2), or with PDGF-AR, which does not bind streptavidin, had no β-actin. Thus, SA-AR can be used to modify cells and then capture them again with SA coated beads.

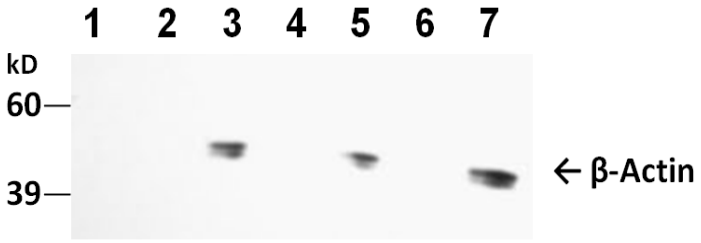


Figure 6-5. SA-AR modified cells are collected with streptavidin coated magnetic beads. CEM cells modified with SA-AR or a non-specific PDGF-AR were collected with streptavidin coated DynaBeads. Cells were lysed, and the lysates were probed by western blot for β-Actin. 1) Unmodified cells. 2) TD05-biotin 3) sgc8-biotin. 4) PDGF aptamer-lipid. 5) SA-AR. 6) Empty. 7) CEM lysate.

Thrombin Aptamer-Receptor Captures Thrombin

Thrombin is an enzyme important in the clotting of blood. Thrombin has two very well-studied aptamers: a 15nt (T-15), which binds the fibrinogen cleavage site on

thrombin with relatively low affinity (K_d = 450nM), and a 27nt aptamer (T-27), which binds exosite 2, important for heparin binding, but has a lower binding affinity (K_d = 0.7nM). We made our thrombin aptamer artificial receptor (TA-AR) by synthesizing T-27 with a lipid tail. Consequently, when thrombin binds the TA-AR, the protein will be tethered to the cell membrane through T-27's interaction with the heparin binding site. The fibrinogen cleaving active site is found on the other side of the thrombin protein, allowing it to remain free and active after interacting with the TA-AR. To show that we were able to localize thrombin to the cell surface via the TA-AR, we made an assembly on the cell surface where the TA-AR captured the thrombin protein, which, in turn, captured a biotinylated T-15. This was visualized by adding SA-488, which bound the biotin on T-15, and was analyzed using a flow cytometer [Figure 6-6].

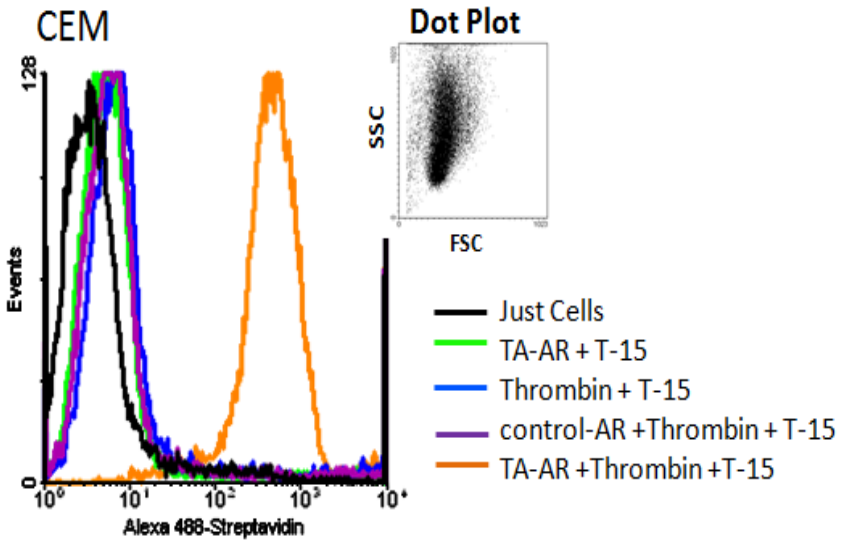


Figure 6-6. Cells are modified with thrombin via a thrombin aptamer artificial receptor. 500k healthy CEM cells were washed and incubated for two hours at 37°C in 50µL media (black and blue traces), 5uM TA-AR (green and orange), or 5µM control PDGF-AR (purple). After washing, all tubes, except black and green, were incubated with 500nM thrombin. After further washing, all tubes were incubated with 500nM biotinylated T-15 followed by streptavidin-488. Cells were examined in FL-1 by flow cytometry for streptavidin-488 signal.

Thrombin Modified Cells Cause Clotting

Since T-15 bound the fibrinogen cleavage site on thrombin, and, as a result, inhibited the ability to cleave fibrinogen, we asked whether the cells modified with thrombin would induce clotting through cleavage of fibrinogen to fibrin. To address this question, thrombin protein was modified on the surface of CEM cells using the TA-AR, followed by adding fibrinogen to a clotting buffer. Cells that were pre-modified with thrombin via the TA-AR, and exposed to fibrinogen caused a clot to form within 15 sec [Figure 6-7, Tube 3]. The clots had a gel-like consistency, which could be visualized by staining the mixtures with trypan blue. However when a T-15 aptamer, which inhibits thrombin's ability to cleave fibrinogen to fibrin, was added to the mixture before washing, there was no clotting [Tube 4]. If a scrambled T-15 was was added, clotting was recovered [Tube 5]. Likewise use of a control protein, BSA [Tube 6] or control aptamer-receptor, PDGF-AR, did not cause clotting.

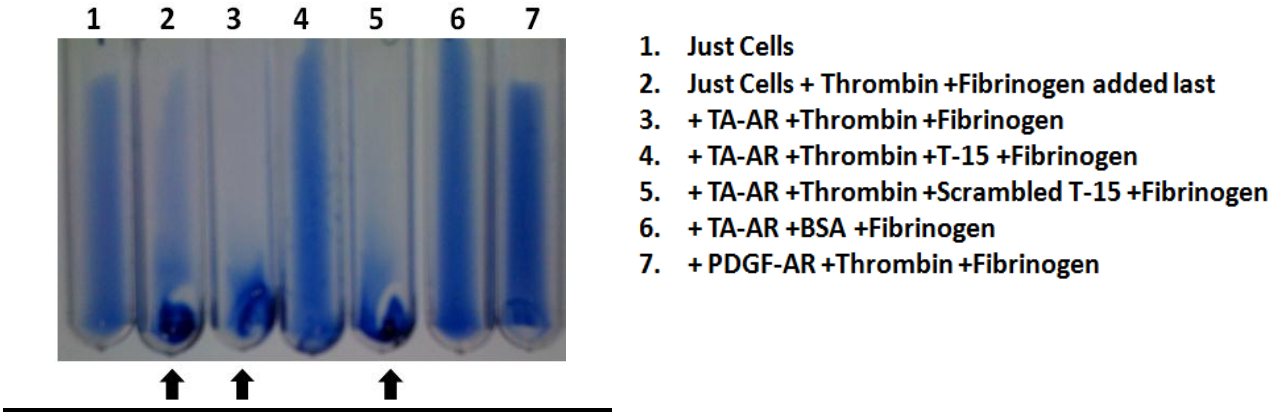


Figure 6-7. Thrombin-Modified Cells Cause Clotting. 500k healthy CEM cells were washed and incubated for two hours at 37°C in 50µL media (Tubes 1+2), 5uM TA-AR (Tubes 3-6), or PDGF-AR (Tube 7). After washing, cells were incubated with buffer (Tube 1), 500nM thrombin (Tube 1+7), 500nM thrombin pre-incubated with 1µM T-15 (Tube 4), 1µM scrambled T-15 (Tube 5), or 200nM BSA for 30min at RT. After five washes, fibrinogen was added to Tubes 2-7. Clots formed and were visualized by adding 50µL trypan blue stain.

Conclusion

This work demonstrates a simple, biomimetic, non-toxic, reversible and dose-controllable strategy for modifying cell membranes with any protein for which there is a known aptamer. This is an improvement over current methods to modify the surface of cells with new proteins because this method is rapid, one-step, dosable, reversible and does not alter the captured protein. Using artificial aptamer-receptors for streptavidin and thrombin, we further showed that proteins captured on the cell membrane surface retained their activity, giving cells new properties, such as enzymatic ability, fluorescence, or a way to collect modified cells from a mixture. In the future, these aptamer-receptors could be used for various bioanalytical applications or to modify cells before their use in therapeutic interventions such as targeting moieties for stem-cell homing after autologous bone marrow transplants.

Materials and Methods

Unless otherwise noted, all chemicals and buffers came from Sigma-Aldrich and were not further purified. All DNA bases, except for the lipid-phosphoramidite, which was synthesized in house, were purchased from Glen Research. Oligonucleotides were synthesized in house on an automated ABI 3400 DNA synthesizer from Applied Biosystems. HPLC of the DNA sequences was done with a Varian Prostar Instrument. UV/Vis measurements for purity and concentration determination were carried out on a Varian Cary 100 spectrophotometer.

DNA Synthesis

All DNA sequences were synthesized with an ABI 3400 synthesizer on a 1.0 micromolar scale. Biotinylated CPG Lipid phosphoramidite was dissolved in 0.4mL

dichloromethane for coupling. For the lipid-DNA after synthesis, the DNA was cut from the CPG beads and deprotected in ammonia hydroxide at 55°C for 14 hours. Next, the DNA was dissolved in 100mM triethylamine-acetic acid buffer (TEAA, pH7.5) and purified by reverse-phase HPLC using a C4 column with an acetonitrile gradient (0-30 min, 10-100%) as an eluent. For other DNA sequences with no biotin or dye modifications, the sequences were deprotected for 20min in AMA (1:1 ammonia hydroxide: 40% methylamine) and purified using Gel-Pak Purification Columns (Glenn Research) followed by desalting on a Nap-5 column (GE Healthcare). The sequences used can be found in Table 4-1.

Table 6-1. Sequences used for Aptamer-Receptors

Name	Sequence
TA-AR	5’ Lipid TT TTT TTG TCC GTG GTA GGG CAG GTT GGG GTG AC-3’
SA-AR (St-2-1)	5′ Lipid TT TTT TTA TTG ACC GCT GTG TGA CGC AAC ACT CAA T-3’
SA complement	5’-ATT GAG TGT TGC GTC ACA CAG CGG TCA AT-3’
T-15 FITC	5’-GGT TGG TGT GGT TGG FITC-3’
T-15 complement	5’-CCA ACC ACA CCA ACC-3’
T-15 scrambled	5’-CAC CAC CAA CAC CAC-3’
T-27 complement	5’-GTC ACC CCA ACC TGC CCT ACC ACG GAC-3’
VEGF-complement	5’ CCC TGC ACT CTT GTC TGG AAG ACG GGA 3’
PDGF-AR	5’ Lipid TT TTT TTC AGG CTA CGG CAC GTA GAG CAT CAC CAT GAT CCT G-3’
VEGF-AR	5’ Lipid TT TTT TTC CCG TCT TCC AGA CAA GAG TGC AGG G-3’

SA: streptavidin; **AR**: aptamer receptor; **T-15**: 15nt thrombin aptamer; **T-27**: 27nt thrombin aptamer; **VEGF:** Vascular endothelial growth factor; **PDGF:** platlet derived growth factor.

Cell Culture

CCRF-CEM cells (T-cell, human acute lymphoblastic leukemia), Ramos (B-cell, human Burkitt's lymphoma), HeLa (human cervical adenocarcinoma), and A549 (human

lung adenocarcinoma) were obtained from ATCC (American Type Culture Association). Ludlu-1 cells were obtained from the European Collection of Cell Cultures (ECACC). CEM, Ludlu-1, and Ramos cells were grown in RPMI-1640 media (GIBCO). HeLa cells were grown in Dulbecco's modified Eagle's medium (DMEM, GIBCO). MCF7 cells were grown in Minimal Essential Media (MEM, GIBCO), and A549 cells were grown in F-12k media (GIBCO). All media were supplemented with 10% fetal bovine serum (Invitrogen), and cells were incubated at 37 °C in 5 % CO2.

Flow Cytometry

500k cells were washed with 2mL washing buffer (WB; 1x PBS, 5mM MgCl2, 4.5g/L glucose) and spun down at 1200g for 3min. CEM cells were incubated with different concentrations of SA-ARs, ranging from 31nM to 5μM, for 2h at 37°C and washed 3x. This was followed by addition of 1:400 streptavidin-488 (SA-488) (Invitrogen) or streptavidin-phycoerythrin (SA-PE) (Invitrogen) and 30min incubation at 4°C in WB. Cells were then washed once in WB and analyzed by flow cytometry on a FACScan (Becton Dickinson) using FL-1 for SA-488 and FL-2 for SA-PE. Data were analyzed using either Win-MDI or FCS Express 2.0. All data shown are ungated. To study aptamer-receptor permanence on the cell surface, CEM cells were incubated with 2μM of SA-AR for 2h at 37ºC, followed by washes and incubation at various times from 0 to 30h before adding 1:400 SA-488, washing, and performing flow analysis.

Fluorescence Microscopy

Cells were plated at low confluence on Lab-Tek four-chambered slides and allowed to grow for 24 hours before washing with WB and incubating with 2μM of SA-ARs for 2h at 37°C. Afterwards, the cells were washed again 3x and incubated with 1:400 SA-488

for 30min at 4°C. After washing, cells were imaged at 40x magnification using incubation and imaged using bright field and fluorescence (488ex) microscopy using a Leica DM6000B microscope.

MTS Assay

MTS works on the principle of a cell's ability to reduce the tetrazolium reagent (Owen's Reagent) via NADH or NADPH when it is alive. The reduced produce absorbs at 490nm and can be read at that wavelength. The higher the is the absorbance, the more viable are the cells. 250k CEM cells were washed and put in sterile flow tubes with 1mL of RPMI media. For each time point, each tube was treated in the same manner. 2mL of PBS was added, and cells were washed by spinning at 1300RPM for three minutes. The cell media was removed by pouring out the supernatant, and cells were incubated with either 50uL 500nM streptavidin-Lipid #3 for 1-2hours at 37°C. 2mL PBS was used to wash the cells. Then the cells in each tube were resuspended in 1mL RPMI. Finally cells were washed, resuspended in 100uL RPMI split into three 96 wells to which 20uL MTS reagent CellTiter 96® AQ_{ueous} Non-Radioactive Cell Proliferation (Promega)was added. Cells were measured at 490nm after 4hours. The wells were all bright purple.

Cell Capture with Streptavidin DynaBeads

12 $x10^6$ Healthy CEM-CCRF cells were washed in 10mL washing buffer. Cells were resuspended in 6 flow tubes with the following treatments: (1) Just Cells. Incubated 1hour RT, washed 2x with 2mL WB. (- control). (2) TD05-biotin 500nM in 200uL BB. Incubated 30min @ 4°C, washed 1x with WB. (- control). (3) Sgc8-Biotin 500nM in 200uL BB. Incubated 30min @ 4°C, washed 1x with WB. (+ control). (4) PDGF-lipid aptamer 2uM in RPMI media. Incubated 1hour , washed 2x with 2mL WB.

(- control). (5) Streptavidin-lipid aptamer 2uM in RPMI media. Incubated 1hour RT, washed 2x with 2mL WB. (6) Just Cells no DynaBead extraction (+ control). Tubes 1-5 were resuspended in 500uL BB. 400uL of each were added to a fresh tube with 50uL DynaBeads (Invitrogen). Cells were placed on a rotator at 4°C for 30min. 150µL RIPA buffer with protease inhibitor was added to each tube. Cells were shaken at 4°C for 30min and sonicated briefly. 150µL 2x SDS Lammeli Sample Buffer was added, and boiled for 5 min. 20µL sample was added to each lane of a 4-12% Bis-Tris Nupage NOVEX Gel (Invitrogen) and run at 200V in MOPs running buffer for 1hour. Blot was transferred at 30V for 1hour on a PDVF membrane. Blots were blocked with 5% milk for one hour then incubated with 1:1000 1° anti-rabbit β-actin (Cell Signaling #4967S) overnight at 4°C. The next day blots were washed and probed with 2° goat anti-rabbit HRP (Pierce #1858414, 1:2000). Blots were visualized with SuperSignal West Dura Extended Duration Substrate (Thermo Scientific) and imaged on Kodak film.

Clotting Assay

500k healthy CEM cells were washed and incubated for two hours at 37°C in 50µL RPMI media (Tubes 1+2), 5uM TA-AR (Tubes 3-6) or PDGF-AR (Tube 7). After washing 3x, cells were incubated with 100µL WB alone (Tube 1), 500nM thrombin (human α-thrombin; Haematologic Technologies, HCT-0020) (Tube 1+7), 500nM thrombin pre-incubated with 1µM T-15 (Tube 4), 1µM scrambled T-15 (Tube 5), or 200nM bovine serum albumin (BSA, Invitrogen) for 30min at RT. After five washes, cells were resuspended in 200µL clotting buffer (25mM Tris-HCL, 150mM NaCl, 5mM KCl, 1mM MgCl2, 1mM CaCl2, 10% glycerol, pH 7.5). Then, 4µL 20mg/mL fibrinogen (fraction I, type I from human plasma; Sigma F3879) was added to Tubes 2-7. Clots

formed and were visualized by adding 50µL trypan blue stain (GIBCO). Pictures were taken by placing the tubes on their sides and imaging against a white background with a digital camera.

CHAPTER 7
FUTURE DIRECTIONS AND CONCLUSIONS

Future Directions

The next step in this work is to further confirm, experimentally, PTK7's interaction with DIXDC1b DNA in the 5'-UTR. A trinity of experiments would confirm the conclusions suggested by our prior results: (1) Performing a gel shift assay where primers specific for DIXDC1b are used to PCR-amplify a dsDNA region surrounding the consensus region, and where the dsDNA is run on a gel. If the DIXDC1b region containing the consensus sequence does indeed interact with PTK7 protein, the addition of purified PTK7 extracellular domain to the amplified dsDNA DIXDC1b region, but not to a control region lacking the consensus sequence, would cause the DNA to run more slowly on a PAGE gel, causing the DNA to shift. (2) Measuring any changes in DIXDC1b mRNA levels compared to control after the following modulations: a. PTK7 knockdown by siRNA; b. PTK7 overexpression by addition of a PTK7 plasmid; c. addition of sgc8c and KC2D4 aptamers; and d. addition of MT1-MMP, which would cause increased cleavage of PTK7 into sPTK7. (3) Confocal immunocytochemistry of cells co-stained for PTK7 (either with an antibody or an aptamer), a nuclear marker, and a membrane marker.

Other useful experiments would determine exactly where the PTK7 aptamers bind the PTK7 protein. The specific binding region could be found by making a series of PTK7 mutants, each with a progressively shorter N-terminal extracellular domain. If mutants lacking a particular Ig-like domain cannot bind the aptamers, the location of the missing Ig-like domain would be in close proximity to the site of aptamer-protein

interaction. When this is found, a mutant that does not bind the aptamers can be made, and the phenotype for this PTK7 protein can be studied.

Lastly, there is one PTK7 aptamer, KDED19, which remains a puzzle. This aptamer competes with the other PTK7 aptamers, but does not share sequence identity with them or with the DIXDC1b sequence. Determining KDED19's relationship to PTK7 and/or the rest of the PTK7-binding aptamer family will prove interesting and informative.

Conclusions

Test tube evolution (SELEX) of RNA and DNA aptamers has become common since Szostak, Ellington, and Gold's pioneering work 15 years ago. Cell-SELEX generates artificial DNA and RNA molecules (aptamers) that bind to biological targets of interest with applications for research and therapy. Each whole cell-SELEX uses a large library of random DNA sequences (~10^{15}), amplified by unique primers, against cells with more than 10,000 unique protein targets. Using this tool, we have selected a number of DNA aptamers that bind to leukemia, breast, and colon cancer cell lines. Bioinformatic and competition analysis revealed that aptamers binding 1 protein epitope were independently selected 4 separate times in selections against different cell lines. Searching for a biological explanation for this result, we discovered a genomic DNA analog to a common DNA sequence in these aptamers.

The first of these aptamers to be selected, sgc8c, was found to bind the extracellular portion of the receptor tyrosine kinase PTK7, which is important for non-canonical Wnt signaling, and is misregulated in numerous cancers. Bioinformatic analysis of 148 aptamer sequences selected by whole cell-SELEX revealed significant sequence identity between sgc8c and 3 other aptamers within a 15nt "consensus

region” (GCTGCGCCGCCGGGA). As predicted, all 4 aptamers competed with each other and bound to the cells with similar affinity, implying that they bound to the same place on the PTK7 protein. Furthermore, mutational analysis of sgc8c indicated that the consensus region, but not the surrounding nucleotides, is important to aptamer binding.

We were curious whether the consensus region’s repeat selection, as well as its importance in aptamer-protein binding, reflected a functional role for the consensus sequence. A BLAST search found that the consensus region, minus the first G base, appears 5 times in the human genome. One of these sites is in the 5'-untranslated region(5’-UTR) of human DIXDC1b DNA, which, along with PTK7, is a regulator of non-canonical Wnt signaling. Surprisingly, aligning the aptamers with DIXDC1b DNA, we found 2 of the aptamers, KMF9b and H01, had additional nucleotides in common with DIXDC1b DNA outside the consensus region. This 22nt region, shared by the PTK7 aptamers and DIXDC1b DNA, is unique in the human genome. We were also surprised that another aptamer, KC2D4, which also competes with the PTK7 aptamers, but which shares no sequence similarity with them, has a region of significant sequence identity to the opposite strand of DIXDC1b DNA.

In total, 5 aptamers compete for binding the same site on PTK7. Four of these aptamers share sequence identity to the positive strand of DIXDC1b DNA’s 5’-UTR. The fifth aptamer shares sequence identity 5 bases from this region, but on the opposite strand of DIXDC1b DNA. The aptamers’ sequence identity to both the positive and negative strands of DIXDC1b DNA could be consistent with the protein PTK7 melting the genomic DNA and interacting with the resulting ssDNA hairpins, which share sequence identity to the aptamers formed by each melted strand. If this is indeed the

case, it could affect DIXDC1b transcription. These sequences identities are conserved in the putative DIXDC1b regulatory regions of *Pan troglodytes*, *Rattus norvegicus*, and *Mus musculus*.

PTK7 has a weak predicted nuclear localization sequence. By cell fractionation, followed by Western blot, we found an abundance of cleaved PTK7 in the nuclei of several different cell lines. This suggests that whole cell-SELEX has not just found an aptamer for an extracellular protein, but may also have identified a genomic DNA sequence, which the protein binds to naturally—after it is cleaved, internalized, and transported to the nucleus. Consistent with this hypothesis, a matrix metalloprotease cleavage site that frees the extracellular portion of PTK7 has been recently reported. Adding this protease to PTK7-expressing cells caused PTK7 to be removed from the cell membrane and accumulate around the nucleus [117].

While we have only begun to address this receptor tyrosine kinase and this regulatory pathway, these findings suggest that whole cell-SELEX might be used more generally to identify novel extracellular transcription factors with highly specific binding motifs. Our bioinformatic analysis of other aptamer sequences selected against whole cells has yielded other examples of disparate selections yielding similar sequences, hinting that there may be other proteins that interact with DNA on the plasma membrane yet to be discovered.

Our results are important for three main reasons: (1) The proteins PTK7 and DIXDC1b are murkily understood, yet key players in Wnt signaling, which is crucial to embryonic development and cancer progression. Understanding their relationship could be important for understanding Wnt signaling. For instance, there are two isoforms of

DIXDC1, a and b; only DIXDC1b has the consensus sequence. PTK7 binding this region could act as a transcriptional switch for this isoform's production. (2) This family of PTK7-binding aptamers has sequence identity to both strands of DIXDC1b DNA, implying that, if melted, the DNA might be forming ssDNA hairpins that bind to PTK7 in a manner similar to the aptamers. This would be a new type of transcription factor of interest to molecular biologists. (3) Finally, this finding is probably not a one-time occurrence. Our bioinformatic analysis of 148 DNA sequences selected by whole-cell SELEX identified other aptamers from disparate selections, like those for *Vaccinia* infected cells and pure virus, which share significant sequence identity. Further comparison of existing DNA and RNA aptamers may yield other examples of SELEX identifying natural DNA or RNA sequences with functional roles, not initially envisioned. Future selections should also not be considered complete until the newly selected aptamers are compared with all other existing aptamers for sequence identity.

APPENDIX A
COMPLEX TARGET SELEX ssDNA APTAMER DATABASE

SELEX	GQ?	Target	Name	Nt	Sequence
1	Yes	hnRNP-A1	BC-15	74	TGTGGCGAGGTAGGTGGGGTGTGTGTGTAT
2	No	IgE	IgE	21	TTTATCCGTTCCTCCTAGTGG
3	No	small cell lung cancer	16-1	25	GAATCCTTCTTTGTCCCGGGCCCGT
	No	small cell lung cancer	0-25	25	TACTCAATTACTCTCTTGTCCCTCT
4	Yes	Shp2 Phosphatse	HJ24	80	GGGGTTTTGGTGGGGGGGGCTGGGTTGTCTTGGGGGTGGG
5	No	Tenascin-C	GB-10	34	CCCAGAGGGAAGACTTTAGGTTCGGTTCACGTCC
6	Yes	Mucin	S1.3/S2.2	72	GCAGTTGATCCTTTGGATACCCTGG
7	No	RET Kinase	D24	50	GCGCGGGAATAGTATGGAAGGATACGTATACCGTGCAATCCAGGGCAACG
8	Yes	Nucleolin	AS1411	26	GGTGGTGGTGGTTGTGGTGGTGGTGG
9	Yes	IL-17RA	RA10-2	30	CTAAGGATCGGATCCACGGCCTACCAGGTC
	No	IL-17RA	RA10-6	30	CTTGGATCACCATAGTCGCTAGTCGAGGCT
	No	IL-17RA	RA10-7	30	ACGCGCTAGGATCAAAGCTGCACTGAAGTG
	No	IL-17RA	RA10-13	30	CCAGAAGAAGCCCACTAGCGTGCTTTTGTC
	No	IL-17RA	RA10-14	30	CCAGACGTGAGCACTAGATCAGTACGGAAG
10	Yes	NSCLC: A549 v HLAMP	S1	45	GGTTGCATGCCGTGGGGAGGGGGGTGGGTTTTATAGCGTACTCAG
	Yes	NSCLC: A549 v HLAMP	S6	45	GTGGCCAGTCACTCAATTGGGTGTAGGGGTGGGGATTGTGGGTTG
	Yes	NSCLC: A549 v HLAMP	S11a	45	AGAGTGGGGGGGTGGGTGGATTTGACAGGTGGCATGCTGGAGAGT
	Yes	NSCLC: A549 v HLAMP	S11b	45	TGGGGTTATTAATTTTGGGTGGGGGGGGAAGATGTAGCATCCGACG
	Yes	NSCLC: A549 v HLAMP	S11c	45	AGCTTGAGGGTGGGCGGGTGGACGCGGTAGTGGTATATAGGTCGG
	Yes	NSCLC: A549 v HLAMP	S11d	45	GATCGGTGGGTGGGGGGGGTTGGAGATCATCCTCAGGGATTACGTC
	Yes	NSCLC: A549 v HLAMP	S11e	45	ATGCGAACAGGTGGGTGGGTTGGGTGGATTGTTCGGCTTCTTGAT
	Yes	NSCLC: A549 v HLAMP	S11f	45	GGTCGCAGATGGATTAAGTATGTGGGTGGGGGGGTGGAAGTTAAT
	Yes	NSCLC: A549 v HLAMP	S15	45	GCTATCTTATGGAAATTTCGTGTAGGGTTTGGTGTGGCGGGGCTA
11	Yes	PigPen	III.1	96	AGGCGGTGCATTGTGGTTGGTAGTATACATGAGGTTTGGTTGAGACTAGTCGCA
12	No	RBC Ghost: CD71	C56t	26	AACTCAGTAATGCCAAGGTAACGGTT
	No	RBC Ghost	Motif 2a	33	CGAATCGCATTGCCCAACGTTGCCCAAGATTCG
13	Yes	Differentiated PC12	1	25	TGGTTGGGGATAGAGGTGGGTGTTT
	Yes	Differentiated PC12	2	25	TGAGGGTCTAGGGTGGTGGGGTGGA

	Yes	Differentiated PC12	3	25	TGATGGATGTGGGGATGCGGGGGCG
	Yes	Differentiated PC12	4	25	TATGGGGGTGGGTCAGGTTTCGGTA
	Yes	Differentiated PC12	5	24	GGGAGGTTGGGGTATCAGGGGGGG
	Yes	Differentiated PC12	7	25	GGGTGTGGGAGGTGATGGGGTAGGT
	Yes	Differentiated PC12	8	24	AGGGGGGTTCGGCGGAGGTATCAG
	Yes	Differentiated PC12	10	25	GCTGGGGTGTTGGGTGTGGGGGTGA
	No	Differentiated PC12	12	25	GTGCGACATAGCTAAACCGGTTCGT
	Yes	Differentiated PC12	13	25	GAGGAGGGAGAATAGGGGTGGGTGG
	No	Differentiated PC12	14	24	AGTCAGACAGGGGGGAGGATCCGT
	Yes	Differentiated PC12	15	25	TGGGTAGGTTCGAGGGGTGGGTGTG
	Yes	Differentiated PC12	16	25	AGAGTGGGGGGGATGTAGGTGGGTT
	Yes	Differentiated PC12	17	25	GGTTGGATGTAAGGTTGGAGGGGGG
	No	Differentiated PC12	18	25	GTGTCCGTGGACTAAACCGGCCTGT
	No	Differentiated PC12	20	24	GTGGAAGCCTCCTAAGCGGTGTGT
	No	Differentiated PC12	22	24	TGGGTGAGTTCAATGGGGGTATGT
	Yes	Differentiated PC12	23	25	GGGTGTGAGAGGTTGAGGGGGTTCG
14	No	Vaccinia Virus A549	TVO1	25	GTGCATTGAAACTTCTGCATCCTCG
	No	Vaccinia Virus A550	TVO2	24	CCTGCATATACACTTTGCATGTGG
	No	Vaccinia Virus A551	TVO4	33	AACCTGCATAATTTATAAGTCTAGACTGCTGCA
	No	Vaccinia Virus A552	TVO6	27	GGACCGATAGGAACCACGGACTGCATG
15	No	Vaccinia Virus Hela	PP2	38	ACACCGTTTGTATTCTGCATTGTTTTGCATTCTACATG
	(Primer)	Vaccinia Virus Hela	PP5	31	CACTTGCATATACACTTTGCATTATAGGGTG
16	No	Mucin	MUC15TR1	25	GAAGTGAAAATGACAGAACACAACA
	No	Mucin	MUC15TR2	25	GGCTATAGCACATGGGTAAAACGAC
	No	Mucin	MUC15TR3	25	CAAACAATCAAACAGCAGTGGGGTG
	No	Mucin	MUC15TR4	25	TACTGCATGCACACCACTTCAACTA
17	No	HL60/CEM	KH1C12	42	TGCCCTAGTTACTACTACTCTTTTTAGCAAACGCCCTCGCTT
	No	HL60/CEM	KHG11	45	TGCTCATCCACGATTCTGGCGAATTTAGTGCCTGTCTTTTTCTCT
	No	HL60	KH2B05	42	CACACAACCTGCTCATAAACTTTACTCTGCTCGAACCATCTC

	No	Ramos	KH1A02	44	GGCATAGATGTGCAGCTCCAAGGAGAAGAAGGAGTTCTGTGTAT
	No	Ramos	KK1B10	45	GATCAGTCTATCTTCTCCTGATGGGTTCCTATTTATAGGTGAAGC
	No	HL60, NB4, K562/CEM	KH1B08	45	TTCAAATCACACGACGCATTGAAACACTCTACAATATCACATTTA
	No	HL60, NB4, K562/CEM	KH3H03	45	CTGGCGCCTTCTACTTCAAGGCAATAAGCTCAATCAATATCATCG
18	No	CEM	H01	46	AAGCAGCAGCTGTGCCATCGGGTTCGGATTTTCTTCCTACGACTGC
	No	CEM	H04	45	TATCAAAGGCGAATTTTGTCAAGGTGTTAAACGATAGTCCCTACC
	No	CEM, Ramos, Toledo	H11	44	TCGCCTGTACATAGACTGTTGCGTTAGGGTCTGCCTTTATCTTG
	No	CEM, Ramos, Toledo	B07	44	CATAGAGACTTGGATGCAACTTAGCTACTAACGCTAGCTCTATG
19	Yes	Ramos	TD05	47	AACACCGTGGAGGATAGTTCGGTGGCTGTTCAGGGTCTCCTCCGGTG
	No	Ramos, CEM, Toledo	TD08	85	TACTCTAATTGCCGTATAAGGTCAGGGGGTTGGTTGGTTCCTAGTGCTT
	No	Ramos	TE02	44	GCAGTGGTTTGACGTCCGCATGTTGGGAATAGCCACGCCTCGGG
	No	Ramos, CEM	TE04	44	CACTCCTCGATGCACCAGTTCACCTTATTTGCTTCTTCTCTCTG
	No	Ramos, CEM, Toledo	TE13	42	GCCCCCAGGCTCGGTGGATGCAAACACATGACTATGGGCCCG
	No	Ramos	TE17	52	ACCTGCTTGACCGACCGATACAGCTACGCAATACAAAACTCCGAACACCTGC
20	No	CEM, Ramos, Toledo	TC01	33	CCAAACACAGATGCAACCTGACTTCTAACGTCA
	No	CEM	TC02	46	AGCATCAACAAGGTCATAAAAACACGTCAGCTCCTTCACATTTGCC
21	No	CEM, Jurkat	sgd3	53	AGGGGGAGCTTGCGCGCATCAAGGTGCTAAACGAAAGCCTCATGGCTTCTATA
	No	CEM, Jurkat, Ramos	sgc4a	36	CGAGTGCGGATGCAAACGCCAGACAGGGGGACAGGA
	No	CEM, Jurkat	sgc5	45	ACCGACGACGAACTATCTATCACTATCTTACACATCATACCTCGA
	No	CEM, Jurkat	sgc7	52	ACCGCAGCGACTATCTCGACTACATTACTAGCTTATACTCCGATCATCTCTA
	No	CEM, Jurkat	Sgd2	53	GAGTGAAGCAAGGATGCAACCTCGGCTCCAACCCGTGAGAGTCGCGAAACTCA
	Yes	CEM, Jurkat, Toledo	Sgd5a	66	ACTTATTCAATTATCGTGGGTCACAGCAGCGGTTGTGAGGAAGAAAGGCGGATAACAGATAATAAG
	No	CEM, Jurkat	sgc8c	41	ATCTAACTGCTGCGCCGCCGGGAAAATACTGTACGGTTAGA
	Yes	CEM, Jurkat	sgc3b	51	ACTTATTCAATTCCTGTGGGAAGGCTATAGAGGGGCCAGTCTATGAATAAG
22	Yes	Liver Cancer mouse: IMEA	TLS1c	55	ACAGGAGTGATGGTTGTTATCTGGCCTCAGAGGTTCTCGGGTGTGGTCACTCCTG
	No	Liver Cancer mouse: IMEA	TLS3	45	TGGGAATATTAGTACCGTTATTCGGACTCCGCCATGACAATCTGG
	yes	Liver Cancer mouse: IMEA	TLS4	45	ACGGTGGTCGTACACGGCCATTTTATTCCCGGAATATTTGTCAAC
	No	Liver Cancer mouse: IMEA	TLS7	45	TGCGCCCAAAGTTCCCATATTGCTTCCCTGTTGGTGAGTGCCGAT
	No	Liver Cancer mouse: IMEA	TLS11a	63	ACAGCATCCCCATGTGAACAATCGCATTGTGATTGTTACGGTTTCCGCCTCATGGACGTGCTG

	No	Small Cell Lung Cancer	HCH07	35	GCCGATGTCAACTTTTTCTAACTCACTGGTTTTGC
23	No	Small Cell Lung Cancer	HCA12	35	GTGGATTGTTGTGTTCTGTTGGTTTTTGTGTTGTC
	Yes	Small Cell Lung Cancer	HCC03	35	CCGGGGACCGGGGCACCGGGGGCCAGTGGCACGGA
	No	Small Cell Lung Cancer	HCH01	71	GTCAACCGAATGCGTCAGCTGGATCTTAAAGATTGCATGCGCTCACTATGGGACTGAGCATCGCACTGGTA
24	Yes	Colon Cancer	KMF2-1a	42	GAATAGGGGATGTACAGGTCTGCACCCACTCGAGGAGTGACT
	Yes	Colon Cancer	KMF3	42	AGGATAGCCATACCACCGGGGAGTTTATAACGGTACGGTCCT
	No	Colon Cancer	KMF9b	46	AGCGCAGCAGCTGTGCCACCGGGAGAATTTACGTACGGCTGAGCGA
25	No	Colon Cancer	KDED1	39	CTAAACAAAATACGAGCAGGGAGACTTCTATCCGATTGT
	Yes	Colon Cancer	KDED2	55	AACTGCTATTACGTGTGAGAGGAAAGATCACGCGGGTTCGTGGACACGGTGGCTT
	No	Colon Cancer	KDED3	39	GGGGTGGTTTTCAAAGAGTCTTGCCTGACTCCCCTGTGG
	No	Colon Cancer	KDED7	40	GCGGACGCACTTTTAGCAAGCAAGTCGACAATGGAGGTTT
	No	Colon Cancer	KDED9	40	GCAACTGAAGCTAGAACTGTGTGGGGTTTGGGGTATAATT
	Yes	Colon Cancer	KDED20	45	TAGGTTGGATAGGGATGGTAGAGCAGGCTAAGCACTTTTTTTTAT
26	Yes	Colon Cancer	KCHA10a	59	ACGCAGCAGGGGAGGCGAGAGCGCACAATAACGATGGTTGGGACCCAACTGTTTGGACA
	No	Colon Cancer	KCHB10	63	ATCCAGAGTGACGCAGCAGATCTGTGTAGGATCGCAGTGTAGTTGACATTTGATACGACTGGC
27	Yes	Colon Cancer	KC2D3	37	CGGGAAAGGAACAAACTGCTATTAGGTCGCAGGCCGG
	No	Colon Cancer	KC2D4	37	CCCACTGGTAGCCATTCCGCCCTTAACCGGGCCATCG
	No	Colon Cancer	KC2D8	38	AACTGCTGCGCCGCCGGGAAAATACTGTACGGTTAGTT
28	No	Ovarian TOV Cells	aptTOVl	45	GATCTGTGTAGGATCGCAGTGTAGTGGACATTTGATACGACTGGC
	No	Ovarian TOV Cells	aptTOV2a	42	CAATCTCTACAGGCGCATGTAATATAATGGAGCCTATCCACG
	No	Ovarian TOV Cells	aptTOV3	42	CTCACTCTGACCTTGGATCGTCACATTACATGGGATCATCAG
	No	Ovarian TOV Cells	aptTOV4	42	GGCACTCTTCACAACACGACATTTCACTACTCACAATCACTC
	No	Ovarian TOV Cells	aptTOV5	42	CAACATCCACTCATAACTTCAATACATATCTGTCACTCTTTC
	No	Ovarian TOV Cells	aptTOV6	42	CGGCACTCACTCTTTGTTAAGTGGTCTGCTTCTTAACCTTCA
	No	Ovarian TOV Cells	aptTOV7	42	CCAACTCGTACATCCTTCACTTAATCCGTCAATCTACCACTC
	No	Ovarian TOV Cells	aptTOV8	42	CCAGTCCATCCCAAAATCTGTCGTCACATACCCTGCTGCGCC
	No	Ovarian TOV Cells	aptTOV9	42	GCAACACAAACCCAACTTCTTATCTTTTCGTTCACTCTTCTC
29	Yes	Ovarian DOV Cells	DOV3	37	ATGCAGAGGCTAGGATCTATAGGTTCGGACGTCGATG
	Yes	Ovarian DOV Cells	DOV6	37	AATGTTGGGGTAGGTAGAAGGTGAAGGGGTTTCAGTT

30	No	Adenocarcinoma: H23	EJD1	42	CCCTCACCACCAAACAACAATATTAGAGACAATGAGTTCCCT
	No	Adenocarcinoma: H23	EJD2	41	AGTGGTCGAACTACACATCCTTGAACTGCGGAATTATCTAC
	No	Adenocarcinoma: H23	EJD4	41	GAAGACGAGCGGCGAGTGTTATTACGCTTGGAAACAACCCC
	No	Adenocarcinoma: H23	EJD5	41	TACGGGCTGGATCCACTGTTACGGCGTGTATCCGCTATCAA
	No	Adenocarcinoma: H23	EJD7	42	CAACTCTTAAGTAAATACCTTTTTCTGGCGTGTAAGAAAATG
	No	Adenocarcinoma: H23	ADE1	42	GGCAAAGCACGACGACATGGTATTACACGAACTACAATCCCT
	No	Adenocarcinoma: H23	ADE2	42	GAGCCCTATCTCACACCGCACCCGCAAACTATCATCCTACAT
31	Yes	Cancer Stem Cells: DU145	CSC01	40	AGGTGGTTTGCTGCGGTGGGCTCAAGAAGAAAGCGCAAAG
	No	Cancer Stem Cells: DU146	CSC08	43	GCTCTGAGCCTAGCTTGACCACTTTTCTTTATTCGCTCTGAGG
	No	Cancer Stem Cells: DU147	CSC13	43	GGGGTGTCGTATCTTTCGTGTCTTATTATTTTCTAGGTGGAGG
	No	Cancer Stem Cells: DU148	CSC17	42	CACCAGCTCCATAACGACACGACCCTCATTCCAACACACAGG
	No	Cancer Stem Cells: DU149	CSC22	43	GTGGGGCTGTGATACTTTACATCTTATTTCTCTAGTGACTAGG
32	Yes	Activated Protein C	HS02-52G	52	GCCTCCTAACTGAGCTGTACTCGACTTATCCCGGATGGGGCTCTTAGGAGGC
33	Yes	Mesenchymal Stem Cells	1MSC	40	CGACTTCGGTTATTACGTTGTTGGCCTCACAAGGACGCCC
	Yes	Mesenchymal Stem Cells	2MSC	39	CACGATCCAGATGTCATAGTTTAGGCTCTCTCTACTACT
	Yes	Mesenchymal Stem Cells	3MSC	40	GGCGGGAGGTCACGTTGAGAATTTACGAGGCAGGGGGCAC
	No	Mesenchymal Stem Cells	4MSC	39	GAGGGGCCGCCAAAGCTAGCTCAAGTGATATCCTGTACT
	No	Mesenchymal Stem Cells	5MSC	41	CACCCGTATGCCAAGTCAGATCCAGTGTAGATGCGCGCCCC
	No	Mesenchymal Stem Cells	6MSC	41	CGACACGCGCACGGTTCTCATCAATACTGCCTCGCCGGTAC
	No	Mesenchymal Stem Cells	7MSC	38	CAGCATGCAGAGGCGTCAAATAACGGGACCTCTCGGAC
	Yes	Mesenchymal Stem Cells	8MSC	53	GGGGAGTGGTGGAGAAAGGCTTACAGGGTAGATAAGGTTCAGGTGCTTCGTTC
	Yes	Mesenchymal Stem Cells	9MSC	50	GGGTCATTGCAGGGTAAGGTTGGATTTATTGATGCCTCGGAGTTGGGTGG
	No	Mesenchymal Stem Cells	10MSC	50	GTAGGCGTTGCCTTAGTTATTGTTTTGAGGTAGAGCAGAGTTTTACTCAG
	Yes	Mesenchymal Stem Cells	11MSC	50	CGAGGTGGATGACAGGGTATGTGGATTGGTAGTGTGTTTGGTGCTAACGC
	Yes	Mesenchymal Stem Cells	12MSC	50	GGAGGAAGGGTTACGGAGGAAGAGTTAGGATCGGTGGGGATGATGATGGG
	Yes	Mesenchymal Stem Cells	13MSC	50	GGTTTAATGTGTGGGTAGTTGGGCGTGACGGGGTAGTCCTGGGGGTTAGG
	Yes	Mesenchymal Stem Cells	14MSC	50	GTGGAGTGGCCGTAGTCTGGCCAGGTCCCGTTGGTGATGGGTAGAGTGGG
	No	Mesenchymal Stem Cells	15MSC	50	TTTGCGCTGGATGCGATAACGTGTTCGACATGAGGCCCGGATCCACTCCC
	No	Mesenchymal Stem Cells	16MSC	50	TGTGCTTATGCTCGAGATGGTGTTATCCGTGTTGCCACGATGGGGGGACC

Yes	Mesenchymal Stem Cells	17MSC	50	TGGATGGGTGGGCGTAGGTGAGGTGTTGTAAGAGCCTCTCCACAGGTGCG
Yes	Mesenchymal Stem Cells	18MSC	50	TGCTCCAAGGGACAGGGCAAGGGATCTATCCTGCCGCGGGGATGTAAGGC
Yes	Mesenchymal Stem Cells	19MSC	50	TGGGGGGAAGCGGACTGTTCGCACTTAGGGCGTATGATGGTAGTGGACCG
Yes	Mesenchymal Stem Cells	20MSC	50	GAGTAATGTAGGGTGAAGGGTGTGGGGGCTATGGGGATAGTGGCACGGCC

APPENDIX B
LIST OF SIMILARITY BETWEEN APTAMERS IN DATABASE

	Aptamer 1	Aptamer 2	%Identity	p-value	95% CI
1	sgc8c	KC2D8	97.3	0	3.69E-06
2	KCHB10	aptTOVl	97.78	0	3.69E-06
3	S11c	S11f	73.33	2.70E-05	3.93E-05
4	S11a	16	83.33	4.00E-05	0.000102413
5	MUC1-5TR-3	TE02	81.82	6.00E-05	0.00013059
6	TVO2	PP5	84.62	8.00E-05	0.000157626
7	sgc8c	KMF9b-TMR	72.09	0.0001	0.000183896
8	H01	KMF9b-TMR	72.73	0.00011	0.000196812
9	KH1B08	aptTOV4	71.05	0.00015	0.00024739
10	sgc8c	18MSC	71.87	0.00015	0.00024739
11	7	23	80	0.00018	0.000284463
12	KH1A02	KCHB10	68.97	0.00019	0.000296693
13	B07	KDED9	71.43	0.0002	0.000308867
14	AS1411	10	73.91	0.00022	0.000333064
15	S6	S11a	68.75	0.00024	0.00035708
16	KDED9	20MSC	68.29	0.00024	0.00035708
17	RA10-2	20MSC	72.73	0.00027	0.000392812
18	S11b	13MSC	67.44	0.00033	0.000463412
19	sgd3	KDED1	70	0.00034	0.000475084
20	DOV3	12MSC	68.57	0.00035	0.000486732
21	TC01	HCH07	74.07	0.00037	0.000509961
22	BC-15	15	76	0.0004	0.00102384
23	HJ24	TD08	60.76	0.0004	0.00102384
24	1	16	77.27	0.0004	0.00102384
25	3	10	77.78	0.0004	0.00102384
26	17	23	75	0.0004	0.00102384
27	7	20MSC	76	0.0004	0.000544647
28	7	DOV6	75	0.00043	0.000579164
29	5	8	76.19	0.0005	0.00116644
30	S6	7	76	0.00051	0.000670502
31	S1.3/S2.2	KCHA10a	64.15	0.0006	0.00130549
32	AS1411	3	71.43	0.0006	0.00130549
33	15	17	76.19	0.0006	0.00130549
34	RA10-2	15MSC	71.43	0.0007	0.00144173
35	III.1	PP2	65.91	0.0007	0.00144173
36	7	13MSC	73.08	0.0007	0.00144173
37	TVO4	TC01	71.43	0.0007	0.00144173
38	sgc8c	6MSC	69.57	0.0007	0.00144173

39	AS1411	S11a	73.33	0.0008	0.00157571
40	S11a	MUC1-5TR-3	71.43	0.0008	0.00157571
41	S11c	2	72.73	0.0008	0.00157571
42	sgc5	sgc7	67.35	0.0008	0.00157571
43	D24	RA10-14	70	0.0009	0.00170779
44	S11c	13MSC	63.64	0.0009	0.00170779
45	S11e	11MSC	65.12	0.0009	0.00170779
46	2	7	73.91	0.0009	0.00170779
47	KMF3-TMR	KMF9b-TMR	66.67	0.0009	0.00170779
48	AS1411	S11e	73.08	0.001	0.00183826
49	S11a	13	72.73	0.001	0.00183826
50	S11f	17	73.33	0.001	0.00183826
51	2	10	75	0.001	0.00183826
52	CSC13	CSC22	66.67	0.001	0.00183826
53	GB-10	DOV3	70.59	0.0011	0.00196735
54	S1	15	73.91	0.0011	0.00196735
55	S1	S11d	65	0.0012	0.00209522
56	S11e	16	76	0.0012	0.00209522
57	10	20MSC	72	0.0012	0.00209522
58	23	20MSC	70.83	0.0012	0.00209522
59	TLS1c	HCA12	66.67	0.0012	0.00209522
60	S1	17	72	0.0013	0.00222201
61	1	12MSC	70	0.0013	0.00222201
62	KH2B05	aptTOV9	65.71	0.0013	0.00222201
63	DOV6	20MSC	65.79	0.0013	0.00222201
64	BC-15	S1	63.83	0.0014	0.00234785
65	S11b	S11f	64.86	0.0014	0.00234785
66	1	4	73.68	0.0014	0.00234785
67	2	8	75	0.0014	0.00234785
68	2	20MSC	71.43	0.0014	0.00234785
69	16	12MSC	70.37	0.0014	0.00234785
70	aptTOV5	aptTOV7	67.5	0.0014	0.00234785
71	aptTOV7	EJD2	67.57	0.0014	0.00234785
72	HJ24	S11f	61.11	0.0015	0.00247282
73	S11d	S11f	64.86	0.0015	0.00247282
74	20	HS02-52G	72.22	0.0015	0.00247282
75	TC02	sgd3	65.12	0.0015	0.00247282
76	KC2D4	aptTOV7	66.67	0.0015	0.00247282
77	RA10-6	B07	69.7	0.0016	0.002597
78	1	DOV6	72.41	0.0016	0.002597
79	S11d	S11e	65	0.0017	0.00272047
80	S1	KDED3	65.52	0.0018	0.00284329
81	2	15	72.73	0.0018	0.00284329

82	15	KDED20	72.73	0.0018	0.00284329
83	H04	sgd3	64.44	0.0018	0.00284329
84	TLS3	TLS11a	64.58	0.0018	0.00284329
85	HJ24	S6	60.87	0.002	0.00308716
86	2	19MSC	70.83	0.002	0.00308716
87	S1	S11a	63.16	0.0021	0.00320829
88	KH1B08	CSC17	64.71	0.0021	0.00320829
89	Sgd5a	8MSC	62.5	0.0021	0.00320829
90	TLS3	KC2D8	66.67	0.0021	0.00320829
91	KDED9	KCHB10	64.29	0.0021	0.00320829
92	AS1411	S11c	70.37	0.0022	0.00332894
93	S11d	17MSC	63.83	0.0022	0.00332894
94	S15	8	71.43	0.0022	0.00332894
95	CSC13	10MSC	64.1	0.0022	0.00332894
96	1	7	71.43	0.0023	0.00344914
97	5	17	72.73	0.0023	0.00344914
98	12	Sgd5a	70.83	0.0023	0.00344914
99	TVO2	H11	72	0.0023	0.00344914
100	TLS4	EJD2	65.12	0.0023	0.00344914

LIST OF REFERENCES

1. Bayrac AT, Sefah K, Parekh P, Bayrac C, Gulbakan B, et al. (2011) In vitro Selection of DNA Aptamers to Glioblastoma Multiforme. ACS Chem Neurosci 2: 175-181.

2. Chen HW, Medley CD, Sefah K, Shangguan D, Tang Z, et al. (2008) Molecular recognition of small-cell lung cancer cells using aptamers. ChemMedChem 3: 991-1001.

3. Mallikaratchy P, Tang Z, Kwame S, Meng L, Shangguan D, et al. (2007) Aptamer Directly Evolved from Live Cells Recognizes Membrane Bound Immunoglobin Heavy Mu Chain in Burkitt's Lymphoma Cells. Mol Cell Proteomics 6: 2230-2238.

4. Sefah K, Meng L, Lopez-Colon D, Jimenez E, Liu C, et al. (2010) DNA Aptamers as Molecular Probes for Colorectal Cancer Study. PLoS ONE 5: e14269.

5. Sefah K, Tang ZW, Shangguan DH, Chen H, Lopez-Colon D, et al. (2009) Molecular recognition of acute myeloid leukemia using aptamers. Leukemia 23: 235-244.

6. Shangguan D, Cao Z, Meng L, Mallikaratchy P, Sefah K, et al. (2008) Cell-specific aptamer probes for membrane protein elucidation in cancer cells. J Proteome Res 7: 2133-2139.

7. Shangguan D, Li Y, Tang Z, Cao ZC, Chen HW, et al. (2006) Aptamers evolved from live cells as effective molecular probes for cancer study. Proc Natl Acad Sci U S A 103: 11838-11843.

8. Tang Z, Shangguan D, Wang K, Shi H, Sefah K, et al. (2007) Selection of aptamers for molecular recognition and characterization of cancer cells. Anal Chem 79: 4900-4907.

9. Van Simaeys D, López-Colón D, Sefah K, Sutphen R, Jimenez E, et al. (2010) Study of the Molecular Recognition of Aptamers Selected through Ovarian Cancer Cell-SELEX. PLoS ONE 5: e13770.

10. Katoh M (2007) Comparative integromics on non-canonical WNT or planar cell polarity signaling molecules: transcriptional mechanism of PTK7 in colorectal cancer and that of SEMA6A in undifferentiated ES cells. Int J Mol Med 20: 405-409.

11. Prebet T, Lhoumeau AC, Arnoulet C, Aulas A, Marchetto S, et al. (2010) The cell polarity PTK7 receptor acts as a modulator of the chemotherapeutic response in acute myeloid leukemia and impairs clinical outcome. Blood 116: 2315-2323.

12. Easty DJ, Mitchell PJ, Patel K, Florenes VA, Spritz RA, et al. (1997) Loss of expression of receptor tyrosine kinase family genes PTK7 and SEK in metastatic melanoma. Int J Cancer 71: 1061-1065.

13. Lu X, Borchers AG, Jolicoeur C, Rayburn H, Baker JC, et al. (2004) PTK7/CCK-4 is a novel regulator of planar cell polarity in vertebrates. Nature 430: 93-98.

14. Uchibori M, Nishida Y, Nagasaka T, Yamada Y, Nakanishi K, et al. (2006) Increased expression of membrane-type matrix metalloproteinase-1 is correlated with poor prognosis in patients with osteosarcoma. Int J Oncol 28: 33-42.

15. Kanazawa A, Oshima T, Yoshihara K, Tamura S, Yamada T, et al. (2010) Relation of MT1-MMP gene expression to outcomes in colorectal cancer. Journal of Surgical Oncology 102: 571-575.

16. Cavalheiro BG, Junqueira CR, Brandão LG (2010) Expression of membrane type 1 matrix metalloproteinase in medullary thyroid carcinoma: Prognostic implications. Head & Neck 32: 58-67.

17. Estevez MC, Huang YF, Kang H, O'Donoghue MB, Bamrungsap S, et al. (2010) Nanoparticle-aptamer conjugates for cancer cell targeting and detection. Methods Mol Biol 624: 235-248.

18. Kang H, O'Donoghue MB, Liu H, Tan W (2010) A liposome-based nanostructure for aptamer directed delivery. Chem Commun (Camb) 46: 249-251.

19. Chen Y, O'Donoghue MB, Huang YF, Kang H, Phillips JA, et al. (2010) A Surface Energy Transfer Nanoruler for Measuring Binding Site Distances on Live Cell Surfaces. J Am Chem Soc 132: 16559-16570.

20. Conrad R, Keranen LM, Ellington AD, Newton AC (1994) Isozyme-specific inhibition of protein kinase C by RNA aptamers. J Biol Chem 269: 32051-32054.

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

22. Kelly JA, Feigon J, Yeates TO (1996) Reconciliation of the X-ray and NMR structures of the thrombin-binding aptamer d(GGTTGGTGTGGTTGG). J Mol Biol 256: 417-422.

23. Macaya RF, Schultze P, Smith FW, Roe JA, Feigon J (1993) Thrombin-binding DNA aptamer forms a unimolecular quadruplex structure in solution. Proc Natl Acad Sci U S A 90: 3745-3749.

24. Daniels DA, Chen H, Hicke BJ, Swiderek KM, Gold L (2003) A tenascin-C aptamer identified by tumor cell SELEX: systematic evolution of ligands by exponential enrichment. Proc Natl Acad Sci U S A 100: 15416-15421.

25. White R, Rusconi C, Scardino E, Wolberg A, Lawson J, et al. (2001) Generation of species cross-reactive aptamers using "toggle" SELEX. Mol Ther 4: 567-573.

26. Bock LC, Griffin LC, Latham JA, Vermaas EH, Toole JJ (1992) Selection of single-stranded DNA molecules that bind and inhibit human thrombin. Nature 355: 564-566.

27. Hicke BJ, Marion C, Chang YF, Gould T, Lynott CK, et al. (2001) Tenascin-C aptamers are generated using tumor cells and purified protein. J Biol Chem 276: 48644-48654.

28. Wang C, Zhang M, Yang G, Zhang D, Ding H, et al. (2003) Single-stranded DNA aptamers that bind differentiated but not parental cells: subtractive systematic evolution of ligands by exponential enrichment. J Biotechnol 102: 15-22.

29. Morris KN, Jensen KB, Julin CM, Weil M, Gold L (1998) High affinity ligands from in vitro selection: complex targets. Proc Natl Acad Sci U S A 95: 2902-2907.

30. Blank M, Weinschenk T, Priemer M, Schluesener H (2001) Systematic evolution of a DNA aptamer binding to rat brain tumor microvessels. selective targeting of endothelial regulatory protein pigpen. J Biol Chem 276: 16464-16468.

31. Homann M, Goringer HU (1999) Combinatorial selection of high affinity RNA ligands to live African trypanosomes. Nucleic Acids Res 27: 2006-2014.

32. Xiao Z, Shangguan D, Cao Z, Fang X, Tan W (2008) Cell-specific internalization study of an aptamer from whole cell selection. Chemistry 14: 1769-1775.

33. Zadeh JN, Steenberg CD, Bois JS, Wolfe BR, Pierce MB, et al. (2011) NUPACK: Analysis and design of nucleic acid systems. J Comput Chem 32: 170-173.

34. Shangguan D, Cao ZC, Li Y, Tan W (2007) Aptamers evolved from cultured cancer cells reveal molecular differences of cancer cells in patient samples. Clin Chem 53: 1153-1155.

35. Medley CD, Bamrungsap S, Tan W, Smith JE (2011) Aptamer-Conjugated Nanoparticles for Cancer Cell Detection. Analytical Chemistry 83: 727-734.

36. Liu X, Dai Q, Austin L, Coutts J, Knowles G, et al. (2008) A One-Step Homogeneous Immunoassay for Cancer Biomarker Detection Using Gold Nanoparticle Probes Coupled with Dynamic Light Scattering. Journal of the American Chemical Society 130: 2780-2782.

37. Chung-Hsin W, Chih-Kuang Y. One-step covalently conjugated aptamer microbubbles for ultrasound targeted imaging; 2010 11-14 Oct. 2010. pp. 1960-1963.

38. Xu Y, Phillips JA, Yan J, Li Q, Fan ZH, et al. (2009) Aptamer-Based Microfluidic Device for Enrichment, Sorting, and Detection of Multiple Cancer Cells. Analytical Chemistry 81: 7436-7442.

39. Phillips JA, Xu Y, Xia Z, Fan ZH, Tan W (2008) Enrichment of Cancer Cells Using Aptamers Immobilized on a Microfluidic Channel. Analytical Chemistry 81: 1033-1039.

40. Liu Y, Bae SW, Wang K, Hong J-I, Zhu Z, et al. (2010) The effects of flow type on aptamer capture in differential mobility cytometry cell separations. Analytica Chimica Acta 673: 95-100.

41. Chen X, Estévez MC, Zhu Z, Huang Y-F, Chen Y, et al. (2009) Using Aptamer-Conjugated Fluorescence Resonance Energy Transfer Nanoparticles for Multiplexed Cancer Cell Monitoring. Analytical Chemistry 81: 7009-7014.

42. Estévez M, O'Donoghue M, Chen X, Tan W (2009) Highly fluorescent dye-doped silica nanoparticles increase flow cytometry sensitivity for cancer cell monitoring. Nano Research 2: 448-461.

43. Martin JA, Phillips JA, Parekh P, Sefah K, Tan W (2011) Capturing cancer cells using aptamer-immobilized square capillary channels. Molecular BioSystems 7: 1720-1727.

44. Shi H, He X, Wang K, Wu X, Ye X, et al. (2011) Activatable aptamer probe for contrast-enhanced in vivo cancer imaging based on cell membrane protein-triggered conformation alteration. Proceedings of the National Academy of Sciences 108: 3900-3905.

45. Huang Y-F, Shangguan D, Liu H, Phillips JA, Zhang X, et al. (2009) Molecular Assembly of an Aptamer–Drug Conjugate for Targeted Drug Delivery to Tumor Cells. ChemBioChem 10: 862-868.

46. Liu H, Zhu Z, Kang H, Wu Y, Sefan K, et al. (2010) DNA-Based Micelles: Synthesis, Micellar Properties and Size-Dependent Cell Permeability. Chemistry – A European Journal 16: 3791-3797.

47. Taghdisi SM, Abnous K, Mosaffa F, Behravan J (2009) Targeted delivery of daunorubicin to T-cell acute lymphoblastic leukemia by aptamer. Journal of Drug Targeting 18: 277-281.

48. Yang L, Meng L, Zhang X, Chen Y, Zhu G, et al. (2011) Engineering Polymeric Aptamers for Selective Cytotoxicity. Journal of the American Chemical Society 133: 13380-13386.

49. Huang Y-F, Chang H-T, Tan W (2008) Cancer Cell Targeting Using Multiple Aptamers Conjugated on Nanorods. Analytical Chemistry 80: 567-572.

50. Wang K, You M, Chen Y, Han D, Zhu Z, et al. (2011) Self-Assembly of a Bifunctional DNA Carrier for Drug Delivery. Angewandte Chemie International Edition 50: 6098-6101.

51. Tong GJ, Hsiao SC, Carrico ZM, Francis MB (2009) Viral Capsid DNA Aptamer Conjugates as Multivalent Cell-Targeting Vehicles. Journal of the American Chemical Society 131: 11174-11178.

52. Chen T, Shukoor MI, Wang R, Zhao Z, Yuan Q, et al. (2011) Smart Multifunctional Nanostructure for Targeted Cancer Chemotherapy and Magnetic Resonance Imaging. ACS Nano.

53. Kang H, Trondoli AC, Zhu G, Chen Y, Chang Y-J, et al. (2011) Near-Infrared Light-Responsive Core–Shell Nanogels for Targeted Drug Delivery. ACS Nano 5: 5094-5099.

54. Mossie K, Jallal B, Alves F, Sures I, Plowman GD, et al. (1995) Colon carcinoma kinase-4 defines a new subclass of the receptor tyrosine kinase family. Oncogene 11: 2179-2184.

55. Park SK, Lee HS, Lee ST (1996) Characterization of the human full-length PTK7 cDNA encoding a receptor protein tyrosine kinase-like molecule closely related to chick KLG. J Biochem 119: 235-239.

56. Shnitsar I, Borchers A (2008) PTK7 recruits dsh to regulate neural crest migration. Development 135: 4015-4024.

57. Paudyal A, Damrau C, Patterson VL, Ermakov A, Formstone C, et al. (2010) The novel mouse mutant, chuzhoi, has disruption of Ptk7 protein and exhibits defects in neural tube, heart and lung development and abnormal planar cell polarity in the ear. BMC Dev Biol 10: 87.

58. Yen WW, Williams M, Periasamy A, Conaway M, Burdsal C, et al. (2009) PTK7 is essential for polarized cell motility and convergent extension during mouse gastrulation. Development 136: 2039-2048.

59. Korinek V, Barker N, Morin PJ, van Wichen D, de Weger R, et al. (1997) Constitutive transcriptional activation by a beta-catenin-Tcf complex in APC-/- colon carcinoma. Science 275: 1784-1787.

60. Morin PJ, Sparks AB, Korinek V, Barker N, Clevers H, et al. (1997) Activation of beta-catenin-Tcf signaling in colon cancer by mutations in beta-catenin or APC. Science 275: 1787-1790.

61. Puppo F, Thome V, Lhoumeau AC, Cibois M, Gangar A, et al. (2011) Protein tyrosine kinase 7 has a conserved role in Wnt/beta-catenin canonical signalling. EMBO Rep 12: 43-49.

62. Peradziryi H, Kaplan NA, Podleschny M, Liu X, Wehner P, et al. (2011) PTK7/Otk interacts with Wnts and inhibits canonical Wnt signalling. EMBO J 30: 3729-3740.

63. Jennifer A Z (2007) Planar Polarity and Tissue Morphogenesis. Cell 129: 1051-1063.

64. Paul N A (2002) Planar Signaling and Morphogenesis in Drosophila. Developmental Cell 2: 525-535.

65. Wehner P, Shnitsar I, Urlaub H, Borchers A (2011) RACK1 is a novel interaction partner of PTK7 that is required for neural tube closure. Development 138: 1321-1327.

66. Toyofuku T, Zhang H, Kumanogoh A, Takegahara N, Suto F, et al. (2004) Dual roles of Sema6D in cardiac morphogenesis through region-specific association of its receptor, Plexin-A1, with off-track and vascular endothelial growth factor receptor type 2. Genes Dev 18: 435-447.

67. Lee HK, Chauhan SK, Kay E, Dana R (2011) Flt-1 regulates vascular endothelial cell migration via a protein tyrosine kinase-7-dependent pathway. Blood 117: 5762-5771.

68. Meng L, Sefah K, O'Donoghue MB, Zhu G, Shangguan D, et al. (2010) Silencing of PTK7 in Colon Cancer Cells: Caspase-10-Dependent Apoptosis via Mitochondrial Pathway. PLoS ONE 5: e14018.

69. Prebet T, Lhoumeau A-C, Arnoulet C, Aulas A, Marchetto S, et al. (2010) The cell polarity PTK7 receptor acts as a modulator of the chemotherapeutic response in acute myeloid leukemia and impairs clinical outcome. Blood 116: 2315-2323.

70. Shin W-S, Maeng Y-S, Jung J-W, Min J-K, Kwon Y-G, et al. (2008) Soluble PTK7 inhibits tube formation, migration, and invasion of endothelial cells and angiogenesis. Biochemical and Biophysical Research Communications 371: 793-798.

71. Haines CJ, Giffon TD, Lu LS, Lu X, Tessier-Lavigne M, et al. (2009) Human CD4+ T cell recent thymic emigrants are identified by protein tyrosine kinase 7 and have reduced immune function. J Exp Med 206: 275-285.

72. Miller MA, Steele RE (2000) Lemon encodes an unusual receptor protein-tyrosine kinase expressed during gametogenesis in Hydra. Dev Biol 224: 286-298.

73. Robinson DR, Wu YM, Lin SF (2000) The protein tyrosine kinase family of the human genome. Oncogene 19: 5548-5557.

74. Wallingford JB, Habas R (2005) The developmental biology of Dishevelled: an enigmatic protein governing cell fate and cell polarity. Development 132: 4421-4436.

75. Tada M, Smith JC (2000) Xwnt11 is a target of Xenopus Brachyury: regulation of gastrulation movements via Dishevelled, but not through the canonical Wnt pathway. Development 127: 2227-2238.

76. Kobus FJ, Fleming KG (2005) The GxxxG-containing transmembrane domain of the CCK4 oncogene does not encode preferential self-interactions. Biochemistry 44: 1464-1470.

77. Rudolph MJ, Gergen JP (2001) DNA-binding by Ig-fold proteins. Nat Struct Biol 8: 384-386.

78. Grassot J, Gouy M, Perriere G, Mouchiroud G (2006) Origin and molecular evolution of receptor tyrosine kinases with immunoglobulin-like domains. Mol Biol Evol 23: 1232-1241.

79. Golubkov VS, Chekanov AV, Cieplak P, Aleshin AE, Chernov AV, et al. (2010) The Wnt/planar cell polarity protein-tyrosine kinase-7 (PTK7) is a highly efficient proteolytic target of membrane type-1 matrix metalloproteinase: implications in cancer and embryogenesis. J Biol Chem 285: 35740-35749.

80. Holmbeck K, Bianco P, Caterina J, Yamada S, Kromer M, et al. (1999) MT1-MMP-deficient mice develop dwarfism, osteopenia, arthritis, and connective tissue disease due to inadequate collagen turnover. Cell 99: 81-92.

81. Jung JW, Ji AR, Lee J, Kim UJ, Lee ST (2002) Organization of the human PTK7 gene encoding a receptor protein tyrosine kinase-like molecule and alternative splicing of its mRNA. Biochim Biophys Acta 1579: 153-163.

82. Li N, Ebright JN, Stovall GM, Chen X, Nguyen HH, et al. (2009) Technical and biological issues relevant to cell typing with aptamers. J Proteome Res 8: 2438-2448.

83. Nagase T, Kikuno R, Hattori A, Kondo Y, Okumura K, et al. (2000) Prediction of the Coding Sequences of Unidentified Human Genes. XIX. The complete Sequences of 100 New cDNA Clones from Brain Which Code for Large Proteins in vitro. DNA Research 7: 347-355.

84. Hu Y, Galkin AV, Wu C, Reddy V, Su AI (2011) CAFET Algorithm Reveals Wnt/PCP Signature in Lung Squamous Cell Carcinoma. PLoS ONE 6: e25807.

85. Ikeuchi Y, Stegmüller J, Netherton S, Huynh MA, Masu M, et al. (2009) A SnoN–Ccd1 Pathway Promotes Axonal Morphogenesis in the Mammalian Brain. The Journal of Neuroscience 29: 4312-4321.

86. Soma K, Shiomi K, Keino-Masu K, Masu M (2006) Expression of mouse Coiled-coil-DIX1 (Ccd1), a positive regulator of Wnt signaling, during embryonic development. Gene Expression Patterns 6: 325-330.

87. Shiomi K, Uchida H, Keino-Masu K, Masu M (2003) Ccd1, a Novel Protein with a DIX Domain, Is a Positive Regulator in the Wnt Signaling during Zebrafish Neural Patterning. Current Biology 13: 73-77.

88. Liu Y-T, Dan Q-J, Wang J, Feng Y, Chen L, et al. (2011) Molecular Basis of Wnt Activation via the DIX Domain Protein Ccd1. Journal of Biological Chemistry 286: 8597-8608.

89. Wang L, Li H, Chen Q, Zhu T, Zhu H, et al. (2010) Wnt signaling stabilizes the DIXDC1 protein through decreased ubiquitin-dependent degradation. Cancer Science 101: 700-706.

90. Wong CK, Luo W, Deng Y, Zou H, Ye Z, et al. (2004) The DIX Domain Protein Coiled-coil-DIX1 Inhibits c-Jun N-terminal Kinase Activation by Axin and Dishevelled through Distinct Mechanisms. Journal of Biological Chemistry 279: 39366-39373.

91. Luo W, Zou H, Jin L, Lin S, Li Q, et al. (2005) Axin Contains Three Separable Domains That Confer Intramolecular, Homodimeric, and Heterodimeric Interactions Involved in Distinct Functions. Journal of Biological Chemistry 280: 5054-5060.

92. Wu Y, Jing X, Ma X, Wu Y, Ding X, et al. (2009) DIXDC1 co-localizes and interacts with γ-tubulin in HEK293 cells. Cell Biology International 33: 697-701.

93. Katoh M (2003) Identification and characterization of human KIAA1391 and mouse Kiaa1391 genes encoding novel RhoGAP family proteins with RA domain and ANXL repeats. Int J Oncol 23: 1471-1476.

94. Wang X, Zheng L, Zeng Z, Zhou G, Chien J, et al. (2006) DIXDC1 isoform, l-DIXDC1, is a novel filamentous actin-binding protein. Biochemical and Biophysical Research Communications 347: 22-30.

95. Katoh M (2003) KIAA1735 gene on human chromosome 11q23.1 encodes a novel protein with myosine-tail homologous domain and C-terminal DIX domain. Int J Oncol 23: 145-150.

96. Shiomi K, Kanemoto M, Keino-Masu K, Yoshida S, Soma K, et al. (2005) Identification and differential expression of multiple isoforms of mouse Coiled-coil-DIX1 (Ccd1), a positive regulator of Wnt signaling. Molecular Brain Research 135: 169-180.

97. Wallin E, von Heijne G (1998) Genome-wide analysis of integral membrane proteins from eubacterial, archaean, and eukaryotic organisms. Protein science : a publication of the Protein Society 7: 1029-1038.

98. Zhang KJ, Tang LL, Sefah K, Zhao ZL, Zhu GZ, et al. (2011) Novel aptamer developed for breast cancer cell internalization. ChemMedChem In press.

99. Wallin E, Heijne GV (1998) Genome-wide analysis of integral membrane proteins from eubacterial, archaean, and eukaryotic organisms. Protein Science 7: 1029-1038.

100. Sefah K, Shangguan D, Xiong X, O'Donoghue MB, Tan W (2010) Development of DNA aptamers using Cell-SELEX. Nat Protocols 5: 1169-1185.

101. Cai Y, Sun Y (2011) ESPRIT-Tree: hierarchical clustering analysis of millions of 16S rRNA pyrosequences in quasilinear computational time. Nucleic Acids Res 39: e95.

102. Sefah K (2011) KMF9b Selection against MCF7-10-AT-1 cells. Unpublished.

103. Tang Z (2004) HO1 Selection against CEM cells. Unpublished.

104. Phillips Ja, Liu H, O'Donoghue MB, Xiong X, Wang R, et al. (2011) Using azobenzene incorporated DNA aptamers to probe molecular binding interactions. Bioconjugate chemistry 22: 282-288.

105. Xing Y, Yang Z, Liu D (2011) A Responsive Hidden Toehold To Enable Controllable DNA Strand Displacement Reactions. Angewandte Chemie International Edition: n/a-n/a.

106. You M, Zhu Z, Liu H, Gulbakan B, Han D, et al. (2010) Pyrene-assisted efficient photolysis of disulfide bonds in DNA-based molecular engineering. ACS Appl Mater Interfaces 2: 3601-3605.

107. Lustig B, Jerchow B, Sachs M, Weiler S, Pietsch T, et al. (2002) Negative Feedback Loop of Wnt Signaling through Upregulation of Conductin/Axin2 in Colorectal and Liver Tumors. Molecular and Cellular Biology 22: 1184-1193.

108. Seo Y-J, Chen S, Nilsen-Hamilton M, Levine Ha (2010) A mathematical analysis of multiple-target SELEX. Bulletin of mathematical biology 72: 1623-1665.

109. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, et al. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Research 25: 3389-3402.

110. Kozak M (1989) The scanning model for translation: an update. The Journal of Cell Biology 108: 229-241.

111. Moss SP, Joyce DA, Humphries S, Tindall KJ, Lunt DH (2011) Comparative Analysis of Teleost Genome Sequences Reveals an Ancient Intron Size Expansion in the Zebrafish Lineage. Genome Biology and Evolution 3: 1187-1196.

112. Rudolph MJ, Gergen JP (2001) DNA-binding by Ig-fold proteins. Nat Struct Mol Biol 8: 384-386.

113. Bravo J, Li Z, Speck NA, Warren AJ (2001) The leukemia-associated AML1 (Runx1)-CBF[beta] complex functions as a DNA-induced molecular clamp. Nat Struct Mol Biol 8: 371-378.

114. Rost B, Yachdav G, Liu J (2004) The PredictProtein server. Nucleic Acids Research 32: W321-W326.

115. Nguyen Ba A, Pogoutse A, Provart N, Moses A (2009) NLStradamus: a simple Hidden Markov Model for nuclear localization signal prediction. BMC Bioinformatics 10: 202.

116. Hsu S-C, Hung M-C (2007) Characterization of a Novel Tripartite Nuclear Localization Sequence in the EGFR Family. Journal of Biological Chemistry 282: 10432-10440.

117. Golubkov VS, Chekanov AV, Cieplak P, Aleshin AE, Chernov AV, et al. (2010) The Wnt/Planar Cell Polarity Protein-tyrosine Kinase-7 (PTK7) Is a Highly Efficient Proteolytic Target of Membrane Type-1 Matrix Metalloproteinase. Journal of Biological Chemistry 285: 35740-35749.

118. Parekh P, Tang Z, Turner PC, Moyer RW, Tan W (2010) Aptamers Recognizing Glycosylated Hemagglutinin Expressed on the Surface of Vaccinia Virus-Infected Cells. Analytical Chemistry 82: 8642-8649.

119. Tang Z, Parekh P, Turner P, Moyer RW, Tan W (2009) Generating Aptamers for Recognition of Virus-Infected Cells. Clin Chem 55: 813-822.

120. Nitsche A, Kurth A, Dunkhorst A, Panke O, Sielaff H, et al. (2007) One-step selection of Vaccinia virus-binding DNA aptamers by MonoLEX. BMC Biotechnol 7: 48.

121. Wang C, Zhang M, Yang G, Zhang D, Ding H, et al. (2003) Single-stranded DNA aptamers that bind differentiated but not parental cells: subtractive systematic evolution of ligands by exponential enrichment. Journal of Biotechnology 102: 15-22.

122. Schafer R, Wiskirchen J, Guo K, Neumann B, Kehlbach R, et al. (2007) Aptamer-based isolation and subsequent imaging of mesenchymal stem cells in ischemic myocard by magnetic resonance imaging. Rofo 179: 1009-1015.

123. Lhoumeau AC, Puppo F, Prebet T, Kodjabachian L, Borg JP (2011) PTK7: a cell polarity receptor with multiple facets. Cell Cycle 10: 1233-1236.

124. Jiang Y, Zhu C, Ling L, Wan L, Fang X, et al. (2003) Specific aptamer-protein interaction studied by atomic force microscopy. Anal Chem 75: 2112-2116.

125. Basnar B, Elnathan R, Willner I (2006) Following aptamer-thrombin binding by force measurements. Anal Chem 78: 3638-3642.

126. Strunz T, Oroszlan K, Schäfer R, Güntherodt H-J (1999) Dynamic force spectroscopy of single DNA molecules. Proceedings of the National Academy of Sciences 96: 11277-11282.

127. Pope LH, Davies MC, Laughton CA, Roberts CJ, Tendler SJ, et al. (2001) Force-induced melting of a short DNA double helix. Eur Biophys J 30: 53-62.

128. Nguyen T-H, Steinbock LJ, Butt H-Jr, Helm M, Berger Rd (2011) Measuring Single Small Molecule Binding via Rupture Forces of a Split Aptamer. Journal of the American Chemical Society 133: 2025-2027.

129. Sattin BD, Pelling AE, Goh MC (2004) DNA base pair resolution by single molecule force spectroscopy. Nucleic Acids Research 32: 4876-4883.

130. Grisshammer R (2006) Understanding recombinant expression of membrane proteins. Current Opinion in Biotechnology 17: 337-340.

131. Legler DF, Doucey MA, Schneider P, Chapatte L, Bender FC, et al. (2005) Differential insertion of GPI-anchored GFPs into lipid rafts of live cells. Faseb J 19: 73-75.

132. McELHINNY AS, Ginger E E, Carol M W (2000) Painting Qa-2 onto Preimplantation Embryos Increases the Rate of Cleavage. American Journal of Reproductive Immunology 44: 52-58.

133. Daniel RDP, Yoshihiro F, Daniel S, Elaine B, Terrone LR, et al. (2001) Properties of exogenously added GPI-anchored proteins following their incorporation into cells. Journal of Cellular Biochemistry 82: 234-245.

134. Koichi K, Chika I, Tohru Y, Teruyuki N (2004) Rapid Protein Anchoring into the Membranes of Mammalian Cells Using Oleyl Chain and Poly(ethylene glycol) Derivatives. Biotechnology Progress 20: 897-904.

135. Chen A, Zheng G, Tykocinski ML (2000) Hierarchical Costimulator Thresholds for Distinct Immune Responses: Application of a Novel Two-Step Fc Fusion Protein Transfer Method. J Immunol 164: 705-711.

136. Saxon E, Bertozzi CR (2000) Cell Surface Engineering by a Modified Staudinger Reaction. Science 287: 2007-2010.

137. Borisenko GG, Zaitseva MA, Chuvilin AN, Pozmogova GE (2009) DNA modification of live cell surface. Nucleic Acids Res 37: e28.

138. Hsiao SC, Shum BJ, Onoe H, Douglas ES, Gartner ZJ, et al. (2009) Direct cell surface modification with DNA for the capture of primary cells and the investigation of myotube formation on defined patterns. Langmuir 25: 6985-6991.

139. Wu Y, Sefah K, Liu H, Wang R, Tan W (2009) DNA aptamer-micelle as an efficient detection/delivery vehicle toward cancer cells. Proc Natl Acad Sci U S A 107: 5-10.

140. Bing T, Yang X, Mei H, Cao Z, Shangguan D (2010) Conservative secondary structure motif of streptavidin-binding aptamers generated by different laboratories. Bioorganic & Medicinal Chemistry 18: 1798-1805.

BIOGRAPHICAL SKETCH

Meghan Bradley O'Donoghue was born in New York to a librarian and a civil engineer. From the age of one, she grew up overseas with her family. For college she returned to the United States, studying at Duke University where she majored in botany and Chinese. After graduation, she sent herself down to do manual work in a hospital emergency room, a farm then a zoo. After three years, she returned to formal education, becoming fascinated by nanoparticles and aptamers while working toward her Ph.D. at the University of Florida. She is currently a post-doc at the NIH studying virus-cell interactions. After spending so much time trying to functionalize nanoparticles, her current advisor, Jon Yewdell, says she has a case of "virus envy". He might be right.