Applications of Cooperative DNA Biosensors

by

Courtney Michelle DuBois

A Thesis Presented in Partial Fulfillment
of the Requirements for the Degree
Master of Science

Approved April 2018 by the
Graduate Supervisory Committee:

Michael Caplan, Chair
Brent Vernon
David Carpentieri

ARIZONA STATE UNIVERSITY

May 2018

ABSTRACT

Cooperativity can be used to manipulate binding affinities of DNA biosensors – improving specificity without sacrificing sensitivity; examples include tentacle probes (TPs) and cooperative primers (CPs). This thesis body of work: (1) used TPs to develop a rapid, low-cost diagnostic for detecting the point mutation leading to Navajo Neurohepatopathy (NNH) and (2) used CPs to amplify a symmetric bowtie-barcoded origami with captured t-cell receptor (TCR) α and β mRNA of a single cell.

NNH (affecting 1-in-1600 Navajo babies) is a fatal genetic disorder often caused by 149G>A mutation and is characterized by brain damage and liver disease/failure. Phoenix Children's Hospital currently uses gene sequencing to identify the 149G>A mutation. While this process is conclusive, there are limitations, as it requires both time (3-4 weeks) and money (≥$700). Ultimately, these factors create barriers that can directly impact a patient's quality of life. Assessment of the developed TP diagnostic, using genomic DNA derived from FFPE patient liver samples, suggests nearly 100% specificity and sensitivity while reducing cost to ~$250 (including cost of labor) and providing a diagnosis within 48 hours.

TCR specificity is dependent on V(D)J recombination as well as pairing of the αβ chains. Drs. Schoettle and Blattman have developed a solution in which a bowtie-barcoded origami strand nanostructure is transfected into individual cells of a heterogeneous cell population to capture and protect αβ mRNA. When PCR of the origami template is performed with Vα, X, Vβ, and Y primers, the α and β gene segments cannot be tied back to a barcode – and paired. Assessment of the developed CPs for PCR

suggests correct individual amplification using (1) Va + Xcp and (2) Vβ + Ycp primers, whereas combination of all the primers (Va, Xcp, Vb, and Ycp) suggests hybridization of the Vα + Xcp and Vβ + Ycp products due to the origami target symmetry.

ACKNOWLEDGMENTS

I would like to take this opportunity to express my deepest gratitude for my thesis advisor and mentor, Dr. Michael Caplan. We've encountered just about every hurdle possible with the NNH project, and I feel so fortunate to have had you by my side guiding us to the end. I have learned so much from you, and I cannot thank you enough for always supporting me and believing in me. Even with hundreds of emails, you always seem to find mine and respond. Michael, I can truly say I wouldn't be where I am today without you. Thank you for giving me this opportunity and helping shape me into who I am today.

Both of these projects have been possible because of the amazing support and collaboration of so many individuals. I would like to thank PCH collaborators Dr. David Carpentieri, Dr. Mitchell Shub, and William Marsh. Thank you for your diligence in ensuring the NNH project has always progressed, despite our numerous hurdles. You have provided an immense amount of support, and I truly appreciate your feedback and insights. Thank you also to Dr. Karmella Haynes for letting us use her LightCycler, and a special thank you to Stefan Tekel. Stefan, thank you for always putting a smile on my face and enabling me with the tools and skills to be the researcher I am today. I feel so honored to have had such a great Ph.D. mentor, and I hope I can someday be even half the mentor you are. Thank you for never giving up on me. I would also like to thank Dr. Louis Schoettle and Dr. Joseph Blattman for giving me an opportunity to continue expanding my knowledge on DNA biosensor applications. Thank you Louie for always

putting up with my questions and giving me insight to the field of immunology. I'd also like to thank Dr. Brent Satterfield, Allison Marley, Meilin Ossanna, Emily Thompson, Scott Johnson, Logan Tayson, and Dr. Brent Vernon, who have all provided contributions to these projects. I cannot thank you enough for your efforts and time.

I wish to also thank my family members and loved ones. Thank you to my BME friends and my roommates, Sam and Justin, for always believing in me and listening to me vent. You've kept me sane throughout this entire process, and I couldn't be more grateful for your friendships. Thank you also to my boyfriend, Kieran, who has always encouraged me to dream big. Thank you for never giving up on me. Finally, to my parents Dan, Michelle, and Bob, thank you for always attending every event and being there for me. Thank you for trusting me and supporting me when I decided to quit my SRP internship three years ago to pursue my research and teaching endeavors. You have all made so many sacrifices so I can pursue my dreams. I love you all so much, and I hope you someday understand how thankful I truly am.

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

DNA	Deoxyribonucleic acid
TPs	Tentacle probes
CPs	Cooperative primers
PEG	Polyethylene glycol
C_t	Concentration of target DNA
C_p	Concentration of TP/CP with both unbound capture and functional regions
$C_{p,cap}$	Concentration of TP/CP with only capture region bound to target DNA
$C_{p,func}$	Concentration of TP/CP with only functional region bound to target DNA
$C_{p,both}$	Concentration of TP/CP with both capture and functional regions bound to target DNA
Z_{AB}	Number of collisions for molecules A and B
d	Sum of molecules A and B radii
V	Average velocity of the molecules A and B
P_r	Reaction probability
K_f	Forward rate constant
$K_{f,cap}$	Forward rate constant of the TP/CP capture region binding to the target DNA

$K_{f,func}$	Forward rate constant of the TP/CP capture functional binding to the target DNA
l	x-distance of the linker length
N_A	Avogadro's number
$P_{r,coop}$	Cooperative reaction probability
$K_{r,cap}$	Reverse rate constant of the TP/CP capture region unbinding from the target DNA
$K_{r,func}$	Reverse rate constant of the TP/CP functional region unbinding from the target DNA
$\frac{dC_{p,cap}}{dt}$	Rate of consumption of $C_{p,cap}$
$\frac{dC_{p,func}}{dt}$	Rate of consumption of $C_{p,func}$
$\frac{dC_{p,both}}{dt}$	Rate of consumption of $C_{p,both}$
$\Delta G°$	Gibbs free energy per mole of reaction
R	Universal gas constant
T	Absolute temperature of reaction
K	Equilibrium constant
$\Delta H°$	Change in enthalpy per mole of reaction
$\Delta S°$	Change in entropy per mole of reaction
NNH	Navajo Neurohepatopathy
PCH	Phoenix Children's Hospital
ARMS	Amplification refractory mutation system

PCR	Polymerase chain reaction
RFLP	Restriction fragment length polymorphism
MBs	Molecular beacons
SG	SYBR Green
gDNA	Genomic DNA
bp	Base pairs
qPCR	Real-time PCR
NNH-N TP	NNH-negative TP
NNH-P TP	NNH-positive TP
FFPE	Formalin-fixed, paraffin-embedded
FP	Forward primer
RP	Reverse primer
SYN DNA	Synthetic single-strand DNA
P SYN DNA	Positive SYN DNA
N SYN DNA	Negative SYN DNA
N gDNA01	Patient 01 negative gDNA
N gDNA02	Patient 02 negative gDNA
N gDNA03	Patient 03 negative gDNA
N gDNA04	Patient 04 negative gDNA
P gDNA01	Patient 01 positive gDNA
TCR	T cell receptor
mRNA	Messenger ribonucleic acid

BCs	Barcodes
RT	Reverse transcription
cDNA	Complementary DNA
T_m	Melting temperature
E_{ff}	Effective primer binding efficiency
OT	Origami template
PT	Positive template
TILs	Tumor infiltrating lymphocytes

CHAPTER 1

INTRODUCTION

Cooperativity

Cooperativity is where the binding of one ligand to a receptor changes the binding of the second ligand to the same receptor (Mercer, 2013). If the first ligand-receptor binding event increases the affinity of the receptor to the second ligand, it is positive cooperativity; conversely, if the first ligand-receptor binding event decreases the affinity of the receptor to the second ligand, it is negative cooperativity (Mercer, 2013). There are two forms for cooperativity: allosteric and configurational. Allosteric cooperativity uses a conformational change of the receptor after the first binding event to then encourage the second binding event. In contrast to this, configuration cooperativity does not use a conformational change of the receptor. Instead, the first binding event keeps the second ligand in close proximity to the receptor to encourage the second binding event **(Figure 1.1)** (Whitty, 2008). Herein, the focus will be configurational cooperativity.

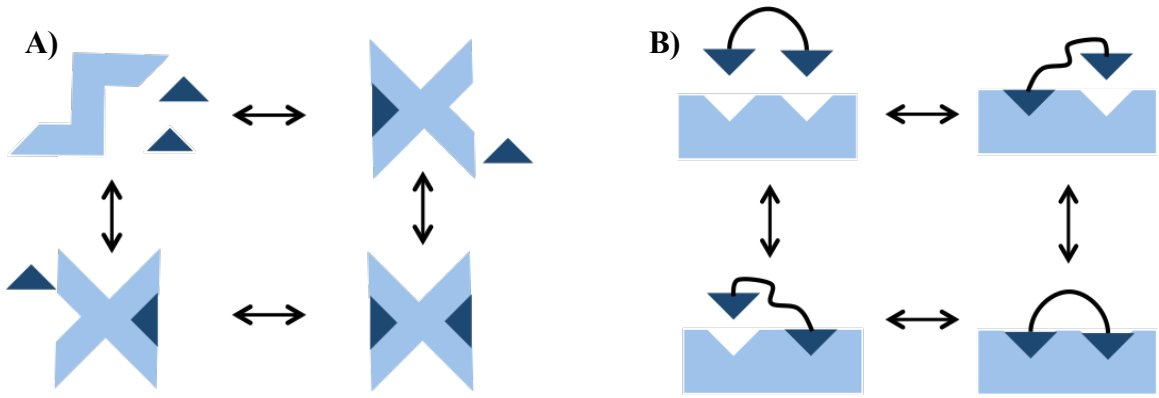


Figure 1.1. Allosteric vs. Configurational Cooperativity
[A] In allosteric cooperativity the receptor undergoes a conformational change after the binding of the first ligand to then encourage binding of the second ligand. **[B]** In configurational cooperativity the receptor does not undergo a conformational change. Instead, the binding of the first ligand keeps the second ligand in close proximity to the receptor to encourage the second binding event. Figure adapted from Whitty 2008.

Positive cooperativity is a tool that can be exploited to manipulate the binding affinities of reagents to improve overall accuracy. Examples of deoxyribonucleic acid (DNA) reagents that use this principle are tentacle probes (TPs) and cooperative primers (CPs), (Satterfield, 2014; Satterfield, West, & Caplan, 2007). The TP and CP reagents are composed of two DNA regions, labeled as capture and functional, which are connected by a polyethylene glycol (PEG) linker **(Figure 1.2)**. Both the capture and functional regions of the TPs/CPs have respective binding locations to target DNA to form DNA duplexes.

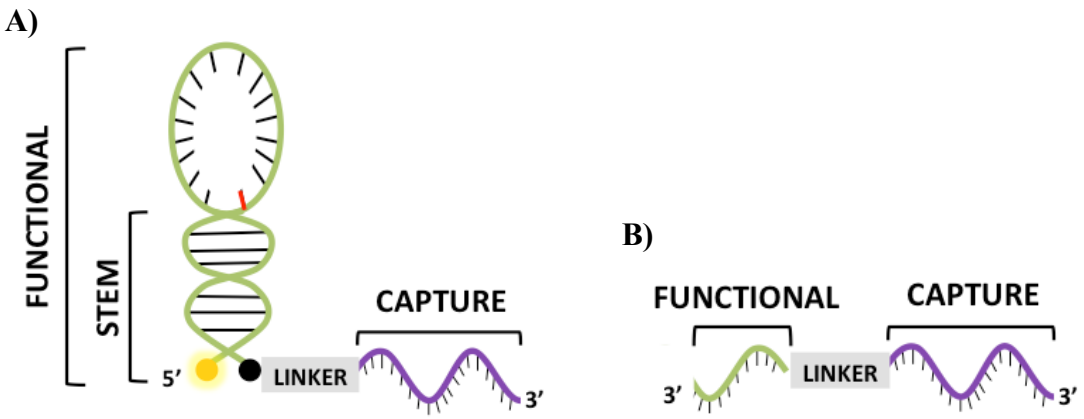


Figure 1.2. Schematic of TPs and CPs
[A] The Tentacle Probe has a hairpin conformation for the functional region (green), which keeps the quencher (black circle) near the FAM fluorophore (yellow circle) to suppress fluorescence produced by the fluorophore. This functional region is connected to a capture region (purple) via a PEG linker (grey box). **[B]** The Cooperative Primer has a short primer around ~12 base pairs for the functional region (green). This region is connected to a capture region (purple) via a PEG linker (grey box).

When discussing binding, the concentration of target DNA will be identified as C_t; the concentration of TP/CP reagent with both unbound capture and functional regions: C_p; with only capture region bound to target DNA: $C_{p,cap}$; with only functional region bound to target DNA: $C_{p,func}$; with both capture and functional region bound to DNA: $C_{p,both}$.

Regardless of the differing conformation for the functional region of the TP and CP, both experience similar cooperative binding reactions, as modeled in **(Figure 1.3)**. Initially, the target DNA strand and TP/CP are free and unbound (**Figure 1.3.A**). Then the first binding event of the TP/CP to the target DNA strand occurs, creating the intermediate product **(Figure 1.3.B.i-B.ii)**. This intermediate can be created one of two ways, where the functional region binds first **(Figure 1.3.B.i)** or the capture region binds

first **(Figure 1.3.B.ii)**. Through cooperativity, the intermediate from B.i increases the binding affinity for the capture region and the intermediate from B.ii increases the binding affinity for the functional region. This second binding event then creates the final product **(Figure 1.3.C)**, assuming no cross-reactions (Cera, 1998).

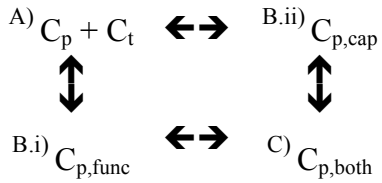


Figure 1.3. Cooperative binding reactions for TPs and CPs
[A] The TP/CP is unbound by both the functional and capture regions. The TP/CP reagent is free to move throughout the solution with target DNA strands. **[Bi]** The TP/CP functional region can bind to the target DNA strand to form one type of intermediate product. **[Bii]** Conversely, the TP/CP capture region can bind to the target DNA strand to form another type of intermediate product. **[C]** In the Bi intermediate, the functional region encourages positive cooperativity for the binding of capture region to create the final product. In the Bii intermediate, the capture region encourages positive cooperativity for the binding of the functional region to create the final product. The final product is both the functional region and capture region of the TP/CP bound to a target DNA strand.

The impact of cooperativity on binding affinity can best be assessed through mathematical modeling and understanding of collision theory and hybridization thermodynamics.

Collision Theory

Max Trautz, in 1916, and William Cudmore McCullagh Lewis, in 1918, developed the kinetic theory of collision. This theory is based on the idea that if every collision of molecule A with molecule B resulted in an A–B product, then the collision

rate would be equal to the rate of reaction, **Eq. (1.1)** (Laidler, 1987):

$$rate = Z_{AB} = (\pi d^2 v)[A][B] \quad (1.1)$$

where Z_{AB} is the number of collisions for molecules A and B, d is the sum of molecules A and B radii, and v is the average velocity of the molecules A and B. However, over the years, there have been additions to the simple collision theory. The original theory assumes all collisions are effective, which is not always the case, as molecules must not only collide to react, but also have mutual orientation and energy (Laidler, 1987). For these reasons, the term reaction probability, P_r, was introduced to the collision rate equation to account for the likelihood of the products to form, **Eq. (1.2)** (Laidler, 1987).

$$rate = P_r Z_{AB} = (P_r \pi d^2 v)[A][B] \quad (1.2)$$

Using rate law, the rate of reaction can then be expressed in terms of forward rate constant, k_f, and concentration of reactants, **Eq. (1.3)** (Laidler, 1987):

$$k_f[A][B] = (P_r \pi d^2 v)[A][B] \quad (1.3)$$

The rate of the first binding event for TPs/CPs can then be modeled based on Eq. 1.3. There are two resulting equations to account for the two intermediate products, $C_{p,cap}$ or $C_{p,func}$, as shown in **Eq. (1.4, 1.5)** respectively.

$$k_{f,cap}[C_P][C_T] = (P_r \pi d^2 v)[C_P][C_T] \quad (1.4)$$

$$k_{f,func}[C_P][C_T] = (P_r \pi d^2 v)[C_P][C_T] \quad (1.5)$$

Where $k_{f,cap}$ is the forward rate constant of the capture region binding to the target DNA, and $k_{f,func}$ is the forward rate constant of the functional region binding to the target DNA. As discussed by Satterfield et al., 2007, the second binding event of TPs/CPs in creating the final product $C_{p,both}$, follows the same collision theory model as the first

binding event; however, it is adjusted to account for the effects of cooperativity. Once again, there are two resulting equations to account for the varying forward rate constants, dependent on the intermediate reactant, $C_{p,cap}$ or $C_{p,func}$, used to form the $C_{p,both}$ product, as shown in **Eq. (1.6, 1.7)**, respectively.

$$k_{f,func}\left[\frac{1\ molecule}{\left(\frac{4}{3}\pi l^3\right)N_A}\right][C_{p,cap}] = \left(P_{r,coop}\pi d^2 v\right)\left[\frac{1\ molecule}{\left(\frac{4}{3}\pi l^3\right)N_A}\right][C_{p,cap}] \tag{1.6}$$

$$k_{f,cap}\left[\frac{1\ molecule}{\left(\frac{4}{3}\pi l^3\right)N_A}\right][C_{p,func}] = \left(P_{r,coop}\pi d^2 v\right)\left[\frac{1\ molecule}{\left(\frac{4}{3}\pi l^3\right)N_A}\right][C_{p,func}] \tag{1.7}$$

Where l is the x-distance of the linker length, N_A is Avogadro's number, and $P_{r,coop}$ is the cooperative reaction probability. Through the first binding event, the local concentration of the second binding event is increased, as the single unbound segment of the TP/CP is restricted to travel within the spherical space created by the length of the PEG linker. This is represented by the $\frac{1\ molecule}{\left(\frac{4}{3}\pi l^3\right)N_A}$ term.

Though collision theory was developed for sphere models, it can be applied to DNA binding, as the equations for collision theory are primarily to demonstrate that forward rate constants maintain the cooperative ratio differences between the first binding and second binding event.

Hybridization Thermodynamics

There are four reaction equations generated when considering the binding of a TP/CP to target DNA. The first two reaction equations, **Eq. (1.8, 1.9)**, correspond to the

first binding event, while the last two reaction equations, **Eq. (1.10, 1.11)**, correspond to the second binding event.

$$C_t + C_p \underset{k_{r,cap}}{\overset{k_{f,cap}}{\rightleftarrows}} C_{p,cap} \tag{1.8}$$

$$C_t + C_p \underset{k_{r,func}}{\overset{k_{f,func}}{\rightleftarrows}} C_{p,func} \tag{1.9}$$

$$\left[\frac{1\ molecule}{\left(\frac{4}{3}\pi l^3\right)N_A}\right] + C_{p,cap} \underset{k_{r,func}}{\overset{k_{f,func}}{\rightleftarrows}} C_{p,both} \tag{1.10}$$

$$\left[\frac{1\ molecule}{\left(\frac{4}{3}\pi l^3\right)N_A}\right] + C_{p,func} \underset{k_{r,cap}}{\overset{k_{f,cap}}{\rightleftarrows}} C_{p,both} \tag{1.11}$$

Where $k_{r,cap}$ is the reverse rate constant of the bound capture region unbinding from the target DNA and $k_{r,func}$ is the reverse rate constant of the bound functional region unbinding from the target DNA. From the **Eq. (1.8-1.11)** above, it can be seen there is both formation and reverse constants for all reaction equations. This indicates that the state of equilibrium is not reached based primarily on completion. Using this information of opposing reactions, the net rate change of $C_{p,cap}$, $C_{p,func}$, and $C_{p,both}$ concentrations can then be calculated, **Eq. (1.12-1.14)**:

$$\frac{dC_{p,cap}}{dt} = k_{f,cap}[C_t][C_p] - k_{r,cap}[C_{p,cap}] \tag{1.12}$$

$$\frac{dC_{p,func}}{dt} = k_{f,func}[C_t][C_p] - k_{r,func}[C_{p,func}] \tag{1.13}$$

$$\frac{dC_{p,both}}{dt} = k_{f,func}\left[\frac{1\ molecule}{\left(\frac{4}{3}\pi l^3\right)N_A}\right][C_{p,cap}] + k_{f,cap}\left[\frac{1\ molecule}{\left(\frac{4}{3}\pi l^3\right)N_A}\right][C_{p,func}] - (k_{r,func} + k_{r,cap})[C_{p,both}] \tag{1.14}$$

Where $\frac{dC_{p,cap}}{dt}$ is the rate of consumption of $C_{p,cap}$, $\frac{dC_{p,func}}{dt}$ is the rate of

consumption of $C_{p,func}$, and $\frac{dC_{p,both}}{dt}$ is the rate of consumption of $C_{p,both}$. These reactions are at equilibrium when the net rate is zero. Using this information, the equilibrium constants for the reactions can then be calculated, **Eq. (1.15-1.17)**. The $C_{p,both}$ equilibrium constant is calculated under the assumptions that (1) the functional region of the TP/CP is designed to have a weak binding affinity and (2) the capture region of the TP/CP is designed to have a strong binding affinity (Satterfield et al., 2007). Through these assumptions, the first binding event is favored to be the capture region binding to the target DNA. Thus, through the first assumption, the $k_{f,cap}\left[\frac{1\ molecule}{\left(\frac{4}{3}\pi l^3\right)N_A}\right]\left[C_{p,func}\right]$ term can be set to zero, and through the second assumption, the $k_{r,cap}$ term can also be set to zero.

$$k_{cap} = \frac{k_{f,cap}}{k_{r,cap}} = \frac{[C_{p,cap}]}{[C_t][C_p]} \tag{1.15}$$

$$k_{func} = \frac{k_{f,cap}}{k_{r,cap}} = \frac{[C_{p,func}]}{[C_t][C_p]} \tag{1.16}$$

$$k_{both} = \left[\frac{1\ molecule}{\left(\frac{4}{3}\pi l^3\right)N_A}\right] k_{func}k_{cap} = \frac{[C_{both}]}{[C_t][C_p]} \tag{1.17}$$

Where k_{cap} is the equilibrium constant of $C_{p,cap}$ formation, k_{func} is the equilibrium constant of $C_{p,func}$ formation, and k_{both} is the equilibrium constant of $C_{p,both}$.

Using the standard free energy change for a chemical process at equilibrium and Gibbs free energy, **Eq. (1.18, 1.19)**, respectively, the equilibrium constant can then be written in terms of thermodynamic properties, **Eq. (1.20)**.

$$\Delta G° = -RTln(k) \tag{1.18}$$

$$\Delta G° = \Delta H° - T\Delta S° \tag{1.19}$$

$$k = e^{-(\Delta H° - T\Delta S°)/RT} \quad (1.20)$$

Where $\Delta G°$ is Gibbs free energy change per mole of reaction, R is the universal gas constant, T is absolute temperature of reaction, k is equilibrium constant, $\Delta H°$ is change in enthalpy, or energy, of a system per mole of reaction, and $\Delta S°$ is change in entropy, or randomness, of a system per mole of reaction.

The change in enthalpy and change in entropy can be approximated using the nearest-neighbor model for Watson-Crick base pairs (SantaLucia & Hicks, 2004). Through this, ten possible pairing parameters have been defined **(Table 1.1)**. Thermodynamic parameters for Watston-Crick nearest neighbor base pairs parameters have been further investigated for internal mismatches **(Table 1.2)**. As can be seen in both tables, the most energetically favorable pairings are Watson-Crick pairings rather than internal mismatches. However, even internal mismatches have the potential to be energetically favorable due to a GC or CG nearest neighbor. For example, there is a ΔG° difference between the GG/CG mismatch pairing and the GG/CC Watson-Crick pairing of 0.73. The GC or CG nearest neighbors are more favorable as they are more stable due to the formation of three hydrogen bonds. The Watson-Crick and internal mismatch parameters are the foundation for the DNA software database Mfold (http://unafold.rna.albany.edu/?q=mfold) (Zuker, 2003) to predict the total ΔH° and ΔS° of two sequences. The two-state folding (http://unafold.rna.albany.edu/?q=DINAMelt/Two-state-folding) from Mfold is used to calculate the hybridization of the stem structure of the TP while the two-state melting (http://unafold.rna.albany.edu/?q=DINAMelt/Two-state-melting) is used for all

hybridization properties. The information produced from Mfold can then be plugged into **Eq. (1.20)** with a defined reaction temperature, to determine the equilibrium constant. For TPs/CPs, this can be used to calculate the equilibrium constants in the presence of both a target and non-target strand to determine the best TP/CP sequence to use.

Table 1.1. Thermodynamic parameters for Watson-Crick nearest neighbor base pairs
Table from SantaLucia & Hicks 2004 and calculated using 1 M NaCl.

Nearest Neighbor Sequence (5'→3' / 3'→5')	**ΔH°** (kcal mol^{-1})	**ΔS°** (cal K mol^{-1})	**$\Delta G°_{37}$** (kcal mol^{-1})
AA/TT	-7.6	-21.3	-1.00
AT/TA	-7.2	-20.4	-0.88
TA/AT	-7.2	-21.3	-0.58
CA/GT	-8.5	-22.7	-1.45
GT/CA	-8.4	-22.4	-1.44
CT/GA	-7.8	-21.0	-1.28
GA/CT	-8.2	-22.2	-1.30
CG/GC	-10.6	-27.2	-2.17
GC/CG	-9.8	-24.4	-2.24
GG/CC	-8.0	-19.9	-1.84

Table 1.2. Gibbs free energy parameters for internal mismatch of Watson-Crick nearest neighbor base pairs

Table from SantaLucia & Hicks 2004 and calculated using 1 M NaCl.

		ΔG° (kcal mol^{-1})			
		Y			
Nearest Neighbor Sequence (5'→3' / 3'→5')	**X**	**A**	**C**	**G**	**T**
GX/CY	A	0.17	0.81	-0.25	-
	C	0.47	0.79	-	0.62
	G	-0.52	-	-1.11	0.08
	T	-	0.98	-0.59	0.45
CX/GY	A	0.43	0.75	0.03	-
	C	0.79	0.70	-	0.62
	G	0.11	-	-0.11	-0.47
	T	-	0.40	-0.32	-0.12
AX/TY	A	0.61	0.88	0.14	-
	C	0.77	1.22	-	0.64
	G	0.02	-	-0.13	0.71
	T	-	0.73	0.07	0.69
TX/AY	A	0.69	0.92	0.42	-
	C	1.33	1.05	-	0.97
	G	0.74	-	0.44	0.43
	T	-	0.75	0.34	0.68

CHAPTER 2

DEVELOPING A RAPID DIAGNOSTIC TEST FOR DETECTING NAVAJO NEUROHEPATOPATHY USING TENTACLE PROBES

Introduction

Navajo Neurohepatopathy. Navajo neuropathy or neurohepatopathy (NNH) is a fatal genetic disorder caused by a R50Q (149G>A) mutation in the *MPV17* gene. It is a point mutation that changes an arginine to glutamine in the DNA sequence, thus impacting the resulting amino acid (Wong et al., 2007). The disease is autosomal recessive, as both the mother and father have to be carriers of the mutation in order to pass it on to the child (Vu et al., 2001). It affects 1-in-1600 Navajo babies in the southwestern region of the U.S., and on average, most patients are diagnosed around thirteen months. These patients often don't live past the age of ten (Vu et al., 2001). The main side effects of NNH are characterized by brain damage and liver disease/failure (Karadimas et al., 2006). As of now, the only treatment options for this disease are dietary control (to prevent hypoglycemia) or liver transplant (Karadimas et al., 2006). Despite these treatment options, there is still a significant concern for this disease. Evidence shows patients develop neurological problems (i.e. development delay, muscle weakness, and reduced sensation) regardless of implementation of one of the treatment options (Wong et al., 2007).

Patients diagnosed with NNH can be grouped into three different categories: (1) infantile NNH, (2) childhood NNH, and (3) classic NNH (Vu et al., 2001). These categories are based on the age of diagnosis as well as severity of the disease. Patients

diagnosed before six months fall within the infantile NNH category. While patients diagnosed between one to five-years-old fall within the childhood NNH category. Patients of both of these phenotypes experience severe liver dysfunction (hepatopathy), and thus, typically have an early death due to liver failure. Though patients do fall within these phenotype categories, they are not as common as classic NNH. For classic NNH, patients are diagnosed across a spectrum of ages and experience moderate liver dysfunction and progressive neurological deterioration (Karadimas et al., 2006).

Upon exhibiting symptoms, children in the American Southwest region (specifically Arizona) seek help from Phoenix Children's Hospital (PCH). At PCH, doctors currently use differential diagnosis in order to detect the root cause of these symptoms. Through this process, doctors will run multiple tests in order to distinguish a particular disease from other diseases with similar symptoms. In order to diagnose a patient with NNH, it was determined that patients must exhibit four of the six symptoms: "(1) sensory neuropathy, (2) motor neuropathy, (3) corneal anesthesia, ulcers or scarring, (4) liver disease, (5) documented metabolic or immunologic derangement, and (6) central nervous system demyelination" (Vu et al., 2001). Following the identification of these symptoms, gene sequencing is utilized, as it is a definitive method to diagnosing NNH. While this process is conclusive for diagnosing NNH, there are limitations, as it requires both time and money. It can take up to several weeks to receive the results of a gene sequence. Furthermore, it costs approximately $700 to run a gene sequence. These factors create barriers that can directly impact a patient's quality of life. Thus, there is a need to develop a rapid diagnostic tool for PCH that can accurately detect the 149G>A mutation.

Specificity and sensitivity. When developing this diagnostic tool, two of the requirements determined by PCH were to create a diagnostic that would be at least 95% specific and 95% sensitive. Specificity is the ability to identify the proportion of negatives that are correctly identified as negatives; the percentage of samples without the mutation that are correctly identified as not having the mutation (Balboni et al., 2014). While sensitivity is the ability to identify the proportion of positives that are correctly identified as positives; the percentage of samples with the mutation that are correctly identified as having the mutation (Balboni et al., 2014). A specific test will not give a positive for anything that is not the target, and a sensitive test will not overlook a positive if the target is there. Thus, a highly specific test has few false positives, and a highly sensitive test has few false negatives.

Types of molecular method detections. There are several molecular methods that can be used for point mutation detection. Some of these methods include amplification refractory mutation system (ARMS) polymerase chain reaction (PCR), restriction fragment length polymorphism (RFLP), molecular beacons (MBs) and TPs. Each method was assessed to determine which detection technique would be best for fulfilling PCH's requirements.

Amplification refractory mutation system. Through ARMS, primers are used for the detection of a specific DNA sequence (Whitcombe, Newton, & Little, 1998). For a robust ARMS test, two primers are designed. The first primer binds to the region with the point mutation while the second primer binds to the same region but without the point mutation. Products of ARMS can be detected via gel electrophoresis or in real-time with

SYBR Green (SG). One of the benefits of this method is that is has the potential to detect a mutation in genomic DNA (gDNA) (Whitcombe et al., 1998). However, there are several disadvantages to this method. Using SG provides a quicker result, however, users assume the fluorescence output of SG did correctly detect the target and that no primer-dimers or other non-target sequences were formed and thus detected by SG – ultimately, suggesting a positive diagnosis when it is negative. Furthermore, it can be difficult to ensure the design of a primer in which it will thermodynamically favor one sequence with the point mutation and not the sequence without the point mutation (see Watson-Crick and internal mismatch free energy parameters from Chapter 1). For these reasons, a high specificity and sensitivity are not guaranteed when using ARMS as point mutation diagnostic.

Restriction fragment length polymorphism. RFLPs recognize restriction sites and create various fragments of DNA after digestion that can then be identified by electrophoresis (Mahdieh & Rabbani, 2013). Point mutations have the ability to create or destroy restriction sites (Mahdieh & Rabbani, 2013). For these reasons, a change in the number/length of fragments can be used to determine a diagnosis. However, a limitation with this application is that it requires a mutation to be in a restriction site. In the case of NNH, the point mutation does not alter any restriction sites. Even if RFLPs could be used to detect NNH, the test requires large amounts of DNA, limits automation possibilities, and in some cases, requires the use of the highly toxic ethidium bromide during electrophoresis to more easily detect DNA fragments (Mahdieh & Rabbani, 2013).

Molecular beacons. MBs are a hairpin structure with a nucleic acid stem loop, fluorophore, and quencher. The loop is approximately 15-25 base pairs (bps) long and acts as the probe to detect the target sequence (Tyagi & Kramer, 2012). The stem loop is approximately 5-7 bps long and serves the function of keeping the fluorophore and quencher in close proximity to limit background fluorescence when the probe is unbound (Tyagi & Kramer, 2012). When the MB binds to its target, it undergoes a conformational change, where the fluorophore and quencher become separated to produce fluorescence.

There is typically a tradeoff seen with MBs in that an increase in sensitivity causes a decrease in specificity (Satterfield, Caplan, & West, 2008). To increase specificity, the probe must be designed to have a more positive ΔG for the neighbor nucleotides of the point mutation. This more positive ΔG would thermodynamically favor the unbound state over the bound state. However, this particular bias can also create a more positive ΔG for the probe to the point mutation, since the thermodynamic difference between point mutation and no point mutation can be such a small change. This would again favor the unbound state over the bound state even in the presence of the point mutation and thus increase false negatives (Satterfield et al., 2008). Conversely, to increase sensitivity, the probe must be designed to have a more negative ΔG for the point mutation, but this can also create a more negative ΔG of the probe to the neighbor nucleotides of the point mutation. These more negative ΔG's would thermodynamically favor the bound state over the unbound state even when no point mutation is present and thus increase false positives (Satterfield et al., 2008). In the case of a diagnosis, typically, false positives are preferred over false negatives. However, false positives can still have

dire consequences, as they can lead to negative psychosocial impacts (Brodersen & Siersma, 2013; Satterfield et al., 2007).

In addition to manipulation of the binding affinity to the target, the binding affinity of the stem can also be altered. The stem hybridization primarily impacts the background fluorescence, regardless of the presence of a target or non-target sequence (Satterfield et al., 2007). A longer stem will create a more negative ΔG – increasing the affinity of the stem hybridization. This will decrease background fluorescence as most MB conformations will be in the stem loop structure and have the fluorescence of the fluorophore suppressed by the quencher. However, having a strong stem can also negatively impact the binding kinetics of the probe as the thermodynamics can favor the stem structure over the probe bound to the target (Satterfield et al., 2007). Conversely, a shorter stem will create a more positive ΔG – decreasing the affinity of the stem hybridization. This will increase the background fluorescence, as most MB conformations will be not be in the hairpin structure, causing the fluorescence to emit due to separation of the fluorophore from the quencher. While a weaker stem hybridization helps with binding kinetics of the probe to the target, it will also limit the fluorescence increase of the system when the probe binds to the target sequence (Satterfield et al., 2007). Ultimately, all of these design considerations create limitations and can prevent an overall high specificity and sensitivity from being obtained.

Tentacle probes. TPs are similar to MBs, as there is still a hairpin structure with a nucleic acid stem loop, fluorophore, and quencher; however the main difference between TPs and MBs is the addition of the capture region to the TPs **(Figure 2.1)**

(Satterfield et al., 2007). When referring to the earlier cooperativity models mentioned in the introduction, the functional region of the TP is known as the detection region. The capture region binds upward of the target point of interest. Ultimately, this helps with the binding kinetics as the probe can be designed to have a more positive ΔG for the neighbor nucleotides of the point mutation. Even though this also create a more positive ΔG of the probe to the point mutation, this is overcome by the TP due to the capture region holding the detection region in closer proximity to the target. When the stem loop unfolds and binds to the target region, the fluorophore and quencher are separated, thus causing fluorescence (Satterfield et al., 2008). Previous research has demonstrated that TPs (when compared to MBs) have the fastest assay time, most sensitive detection limit, and highest specificity (Satterfield et al., 2008).

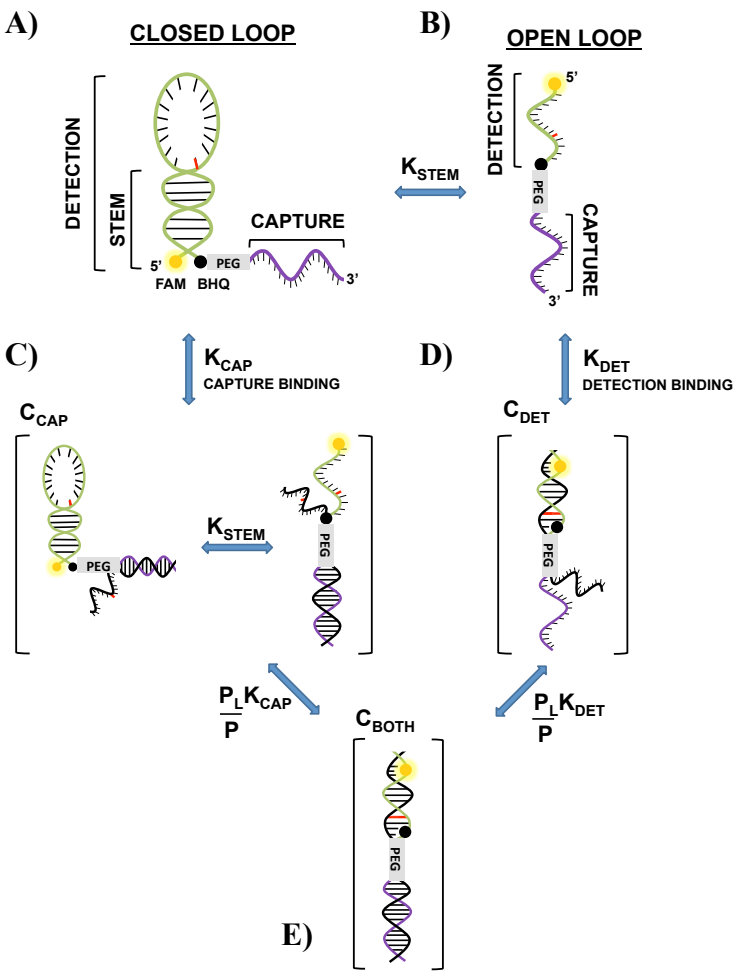


Figure 2.1. Schematic of TP Binding
[A] The TP can begin in a closed loop conformation, where the quencher and fluorophore are near, thus suppressing fluorescence. **[B]** The TP can also begin in on open loop conformation, where the quencher and fluorophore are separated, emitting fluorescence. The ability of the TP to be in either the closed loop or open loop conformation is based on the equilibrium constant of the stem loop hybridization, K_{stem}. There are two pathways to create the intermediate product of the TP and target DNA binding. **[C]** The first binding event can be the capture region to the target DNA, K_{cap}, to form the C_{cap} intermediate. When bound to the target DNA, the stem loop structure can still switch between closed and open conformation. **[D]** The first binding event can also be the detection region to the target DNA, K_{det}, to form the C_{det} intermediate. **[E]** Both of these intermediates lead to positive cooperative binding to create the C_{both} product.

TPs can be used in conjunction with real-time polymerase chain reaction (qPCR) in order to obtain both low-cost and quick results while maintaining efficacy. With qPCR, DNA is amplified through three steps: (1) denaturation, (2) annealing, and (3) extension. These three steps make up one cycle, and after each cycle, the DNA is doubled. With the TP incorporated in the qPCR, fluorescence can be measured in order to generate a graph detailing fluorescence versus cycle number. By including a positive and negative control in the experiment the output of the qPCR can be used to determine if a sample is positive or negative for a particular disease. Furthermore, including two TPs: NNH-negative TP (NNH-N TP) to detect the wild type and NNH-positive TP (NNH-P TP) to detect the 149G>A mutation, creates a more robust test. No fluorescence from either probe would indicate an inconclusive test and something wrong in the diagnostic, fluorescence from only the NNH-P TP would indicate a positive diagnosis, fluorescence from only the NNH-N TP would indicate a negative diagnosis, and fluorescence from both the NNH-P and NNH-N TPs would indicate a carrier **(Figure 2.2)**. Herein this chapter is the verification and assessment of the TP diagnostic.

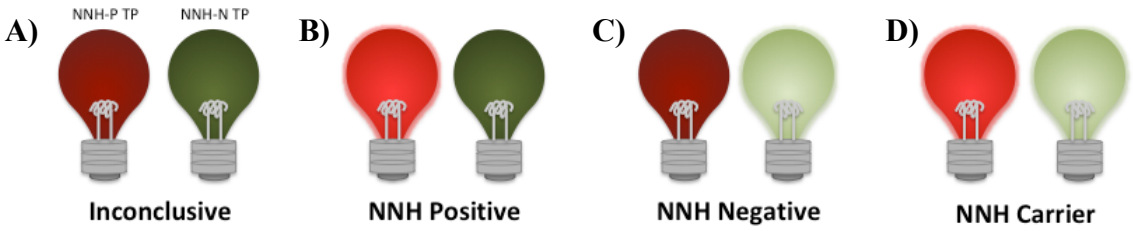


Figure 2.2. Purpose of Dual TPs
The red light bulb represents the NNH-P TP, while the green light bulb represents the NNH-N TP. A darker light bulb represents little to no fluorescence being emitted by the TP, while a brighter light bulb represents fluorescence being emitted by the TP. **[A]** If neither TP fluoresces, then the diagnostic is inconclusive and it can be concluded that

some aspect of the test has failed. **[B]** If only the NNH-P TP fluoresces, then the diagnosis can be confirmed as NNH positive. **[C]** If only the NNH-N TP fluoresces, then the diagnosis can be confirmed as NNH negative. **[D]** If both the NNH-P and NNH-N TPs fluoresce, then the diagnosis can be confirmed as positive for NNH carrier.

Materials and Methods

Oligonucleotide synthesis. Human gDNA was extracted from H1975, a non-small cell lung cancer, as well as PCH-stored NNH-N and NNH-P formalin-fixed, paraffin-embedded (FFPE) liver samples. All human gDNA samples were stored in nuclease-free water. Forward (FP) and reverse primer (RP) sequences were designed close to the mutation site using OligoAnalyzer 3.1 to give an amplicon of 107 bases. Synthetic single-strand DNA (SYN DNA) of 107 bases was designed as positive synthetic DNA (P SYN DNA) containing the 149G>A mutation and negative synthetic DNA (N SYN DNA) without the 149G>A mutation in wild-type form. Integrated DNA Technologies (Skokie, Illinois) synthesized all SYN DNA, which were stored in nuclease-free water. The TP sequences were designed to be within the boundary of the FP and RP. The NNH-P TP was designed to identify the 149G>A mutation while the NNH-N TP was designed to identify the wild-type. TP folding, hybridization, and thermodynamic parameters were modeled using Mfold software (Zuker, 2003). These thermodynamic parameters were then used to model fluorescence, and TPs were chosen based on >50% and <1% fluorescence in target and non-target sequences, respectively. BioSearch (Petaluma, CA) synthesized all TPs, which were suspended in 10 mM Tris, 1 mM EDTA, pH 8.0 (see **Table 2.1** for primer, synthetic DNA, and TP sequences).

Table 2.1. Primer, synthetic DNA, and TP sequences for NNH diagnostic

Name	Sequence 5' → 3'
FP	AGAGGCGGGGTCTGCAG
RP	TGAGTCCACTGAAGCCCTGTTG
N SYN DNA	AGAGGCGGGGTCTGCAGGAACACCAGAGAGGCC**G**GACTCTG ACCATGGTGTCCCTGGGCTGTGGCTTTGTGGTAAGTTCTCCCT CAACAGGGCTTCAGTGGACTCA
P SYN DNA	AGAGGCGGGGTCTGCAGGAACACCAGAGAGGCC**A**GACTCTG ACCATGGTGTCCCTGGGCTGTGGCTTTGTGGTAAGTTCTCCCT CAACAGGGCTTCAGTGGACTCA
NNH-N TP	[FAM]CGGAGTCCC**G**GACTCCG[BHQ-1][SPACER 18]GTGTCCCTGGGCT
NNH-P TP	[FAM]CCGGAGTCTGAGGCC**A**GACTCCGG[BHQ-1][SPACER 18]GTGTCCCTGGGCTGTGGCTTTG

Primer design and optimization. To determine optimal primer annealing conditions, 8 samples of 50 µL were diluted with water to have 10 ng/µL of H1975 DNA, 1 µM FP, 1 µM RP, and 1X GoTaq Green Master Mix (Promega). Samples were loaded in a C100 Touch Thermal Cycler. Cycling conditions were an initial denaturation of 95°C for 30 s, annealing temperature gradient of 50°C-70°C for 20 s, and extension of 72°C for 30 s for a total of 25 cycles. Gel electrophoresis using 1X TAE buffer was then

performed on the samples using a 1% agarose gel for 40 min at 110 V with a 1 Kb plus DNA ladder (ThermoFisher) and 100 bp DNA Ladder (ThermoFisher) for reference. Optimal annealing temperature was determined to be 66°C.

NNH-N and NN-P TP design and optimization. NNH-N and NNH-P TPs were first designed in silico using calculations of the enthalpy and entropy hybridization of both the detection and capture regions to target and non-target strands. Using a set reaction temperature of 55°C, equilibrium constants were then calculated. These equilibrium constants were then used to calculate the fluorescence of the TP, **Eq. (2.1)**, derived by Satterfield:

$$F = \propto \frac{K_{func} + \left[\frac{1\,molecule}{\left(\frac{4}{3}\pi l^3\right)N_A}\right] K_{func} K_{cap}}{K_{both}} \left(\frac{T_0 K_{both}}{1 + T_0 K_{both}}\right) \tag{2.1}$$

Where F is the fluorescence and T_0 is the initial target concentration and P_0. In Satterfield *et al.* 2007, the equation also includes a β and γ term. The β term represents the fluorescence when the TP is unbound and closed in stem loop structure, and the γ term represents the fluorescence when the TP is unbound and open from stem loop structure. These terms are not included as both β and γ fluorescence are used to calculate the background fluorescence, which should be the same in the presence of both a target and non-target sequence. Furthermore, the equilibrium constant of the stem loop structure can be used to ensure there is little to no fluorescence occurring for the γ term. The best TPs were determined to be >50% fluorescence in presence of target sequence and <0.1% fluorescence in presence of non-target sequence.

To determine optimal TP annealing conditions, 4 samples (P SYN DNA + NNH-P TP, N SYN DNA + NNH-P TP, P SYN DNA + NNH-N TP, and N SYN DNA + NNH-N TP) of 20 µL were diluted with water to have 10 ng/µL of specified pre-amplified synthetic DNA, 1X ThermoPol Reaction Buffer (NEB), and 0.1 µM specified TP. Samples were loaded in a Roche LightCycler 480. A melting curve was performed starting at 95°C with a 2 min hold and 1°C decrease until 37°C. Optimal annealing temperatures of the NNH-P TP and NNH-N TP was determined to be 55°C and 44°C, respectively.

NNH-N and NNH-P TP synthetic DNA verification. To assess NNH-N TP efficacy with synthetic DNA, 2 samples (N SYN DNA + NNH-N TP and P SYN DNA + NNH-N TP) were prepared to be 400 µL each diluted with water to have 1 pg/µL of specified SYN DNA, 1X GoTaq Colorless Master Mix (Promega), 1 µM FP, 1 µM RP, and 0.1 µM NNH-N TP. Two controls (N SYN DNA + SG and P SYN DNA + SG) were prepared to be 60 µL each diluted with water to have 1 pg/µL of specified synthetic DNA, 1 µM FP, 1 µM RP, and 1X SYBR Green I (Roche). Twenty-µL samples sizes were pipetted from each master mix sample into a 96-well plate; there were 20 replicates for each TP sample and 3 replicates for each SG sample. The 96-well plate was loaded in a Roche LightCycler 480. Cycling conditions were an initial denaturation of 95°C for 30 s, primer annealing of 66°C for 20 s, extension of 72°C for 30 s, denaturation of 95°C for 30 s, and TP annealing of 44°C for 2 min for a total of 25 repeats. A fluorescence measurement was made after the extension phase and after the TP annealing phase. Data was exported to excel and split into fluorescence values specific to SG and TP. The first

derivative of fluorescence for the TP was then plotted versus cycle number.

To assess NNH-P TP efficacy with synthetic DNA, 2 samples (N SYN DNA + NNH-P TP and P SYN DNA + NNH-P TP) were prepared to be 400 µL each diluted with water to have 1 pg/µL of specified synthetic DNA, 1X GoTaq Colorless Master Mix, 1 µM FP, 1 µM RP, and 0.1 µM NNH-P TP. Two controls (N SYN DNA + SG and P SYN DNA + SG) were prepared to be 40 µL each diluted with water to have 1 pg/µL of specified synthetic DNA, 1 µM FP, 1 µM RP, and 1X SG. Twenty-µL samples sizes were pipetted from each master mix sample into a 96-well plate; there were 20 replicates for each TP sample and 2 replicates for each SG sample. The 96-well plate was loaded in a Roche LightCycler 480. Cycling conditions were an initial denaturation of 95°C for 30 s, primer annealing of 66°C for 20 s, extension of 72°C for 30 s, denaturation of 95°C for 30 s, and TP annealing of 55°C for 2 min for a total of 25 repeats. A fluorescence measurement was made after the extension phase and after the TP annealing phase. Data was exported to excel and split into fluorescence values specific to SG and TP. The first derivative of fluorescence for the TP was then plotted versus cycle number.

NNH-N and NNH-P TP genomic DNA verification. Five patient gDNA samples (patient 01 negative gDNA (N gDNA01), N gDNA02, N gDNA03, N gDNA04, and patient 01 positive gDNA (P gDNA01)) were prepared to be 100 µL diluted with water to have 10 ng/µL of specified DNA, 1 µM FP, 1 µM RP, and 1X GoTaq Green Master Mix. Fifty-µL samples sizes were pipetted from each master mix sample into a 0.2 mL PCR tube; there were 2 replicates per sample. Samples were loaded in a C100 Touch Thermal Cycler. Cycling conditions were an initial denaturation of 95°C for 30 s,

annealing of 66°C for 20 s, and extension of 72°C for 30 s for a total of 25 repeats. Replicates were then combined in the same PCR tube to perform PCR purification kit (QIAquick). DNA was eluted in nuclease-free water.

To assess NNH-N and NNH-P TP efficacy with gDNA, 14 samples (N gDNA01 + NNH-N TP, N gDNA02 + NNH-N TP, N gDNA03 + NNH-N TP, N gDNA04 + NNH-N TP, positive genomic DNA 01 P gDNA01 + NNH-N TP, N SYN + NNH-N TP (+ control), P SYN + NNH-N TP (- control), N gDNA01 + NNH-P TP, N gDNA02 + NNH-P TP, N gDNA03 + NNH-P TP, N gDNA04 + NNH-P TP, P gDNA01 + NNH-P TP, N SYN + NNH-P TP (- control), and P SYN + NNH-P TP (+ control)), were prepared to be 100 μL each diluted with water to have 1 pg/μL of specified pre-amplified gDNA or SYN DNA, 1X GoTaq Colorless Master Mix, 1 μM FP, 1 μM RP, and 0.1 μM specified TP. Seven controls (N gDNA01 + SG, N gDNA02 + SG, N gDNA03 + SG, N gDNA04 + SG, P gDNA + SG, N SYN + SG, and P SYN + SG) were prepared to be 40 μL each diluted with water to have 1 pg/μL of specified SYN DNA, 1 μM FP, 1 μM RP, and 1X SG. Twenty-μL samples sizes were pipetted from each master mix sample into a 96-well plate; there were 5 replicates for each type of TP sample and 2 replicates for each SG sample. The 96-well plate was loaded in a Roche LightCycler 480. Cycling conditions were an initial denaturation of 95°C for 30 s, primer annealing of 66°C for 20 s, extension of 72°C for 30 s, denaturation of 95°C for 30 s, TP annealing of 44°C for 2 min, denaturation of 95°C for 30 s, TP annealing of 55°C for 2 min for a total of 25 repeats. A fluorescence measurement was made after the extension phase and after each TP annealing phase. Data was exported to excel and split into fluorescence values

specific to SG, NNH-N TP, and NNH-P TP. The first derivative of fluorescence was averaged for both the NNH-N TP and NNH-P TP and plotted versus cycle number.

Simulated synthetic DNA carrier. To assess potential to identify an NNH carrier, 6 samples (N SYN DNA + NNH-N TP, P SYN DNA + NNH-N TP, N SYN DNA:P SYN DNA + NNH-N TP, N SYN DNA + NNH-P TP, P SYN DNA + NNH-P TP, and N SYN DNA:P SYN DNA + NNH-P TP) were prepared to be 60 µL each diluted with water to have 1 pg/µL of specified SYN DNA (N SYN DNA:P SYN DNA was prepared in a 1:1 ratio), 1X GoTaq Colorless Master Mix, 1 µM FP, 1 µM RP, and 0.1 µM specified TP. Three controls (N SYN DNA + SG, P SYN DNA + SG, and N SYN DNA:P SYN DNA + SG) were prepared to be 40 µL each diluted with water to have 1 pg/µL of specified SYN DNA, 1 µM FP, 1 µM RP, and 1X SG. Twenty-µL samples sizes were pipetted from each master mix sample into a 96-well plate; there were 3 replicates for each TP sample and 2 replicates for each SG sample. The 96-well plate was loaded in a Roche LightCycler 480. Cycling conditions were an initial denaturation of 95°C for 30 s, primer annealing of 66°C for 20 s, extension of 72°C for 30 s, denaturation of 95°C for 30 s, TP annealing of 44°C for 2 min, denaturation of 95°C for 30 s, TP annealing of 55°C for 2 min for a total of 25 repeats. A fluorescence measurement was made after the extension phase and after each TP annealing phase. Data was exported to excel and split into fluorescence values specific to SG, NNH-N TP, and NNH-P TP. The first derivative of fluorescence was averaged for both the NNH-N TP and NNH-P TP and plotted versus cycle number.

Results and Discussion

Primer verification. Results from the temperature annealing gradient **(Figure 2.3)** indicate that for every temperature there is a primary band just barely above 100 bps, which is close to the expected 107 bp product. At temperatures 66°C and above (**Figure 2.3**, lanes 2-4), there are no additional by-products created, while there are larger by-products created at temperatures 62.5°C and below (**Figure 2.3**, lanes 5-9). Ultimately, the bands for 66°C – 70°C show the primers are specific to the target amplicon of interest even when exposed to complex gDNA. The annealing temperature 66°C was chosen as the optimal temperature, as it produces the most product, indicated by the thicker and darker band (**Figure 2.3**, lane 4).

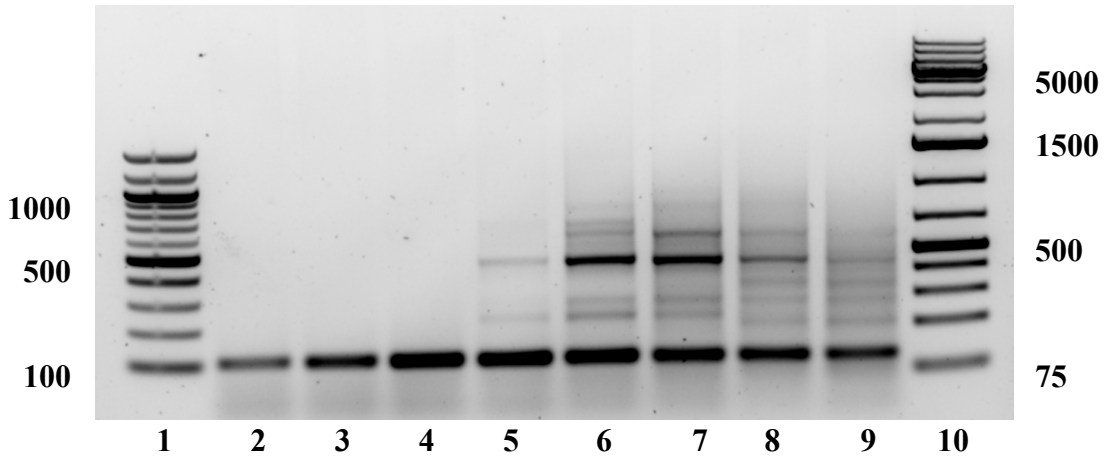


Figure 2.3. Annealing temperature gradient for NNH primers of gDNA
Lanes 1 and 10 are DNA ladders. Lane 1 is 100 bp DNA ladder and lane 10 is a 1 Kb plus DNA ladder. Lanes 2-9 are temperatures used for the annealing temperature gradient. Lane 2 is 70°C3, lane 3: 68.5°C, lane 4: 66°C, lane 5: 62.5°C, lane 6: 57.9°C, lane 7: 54.5°C, lane 8: 51.8°C, and lane 9: 50°C. Results suggest 66°C as the best primer annealing temperature in the presence of genomic DNA.

NNH-N and NNH-P TP synthetic DNA verification. SG showed all samples amplified efficiently (results not shown). **Figure 2.4** shows results for NNH-N TP while **Figure 2.5** shows results for NNH-P TP. From both graphs, it can be seen that there is a significant difference in fluorescence between positive and negative controls. Thus verifying both the NNH-N and NNH-P TPs can correctly differentiate between samples with and without the 149G>A mutation. Furthermore, the results indicate both the NNH-N and NNH-P TPs have 100% sensitivity and 100% specificity in the presence of synthetic DNA.

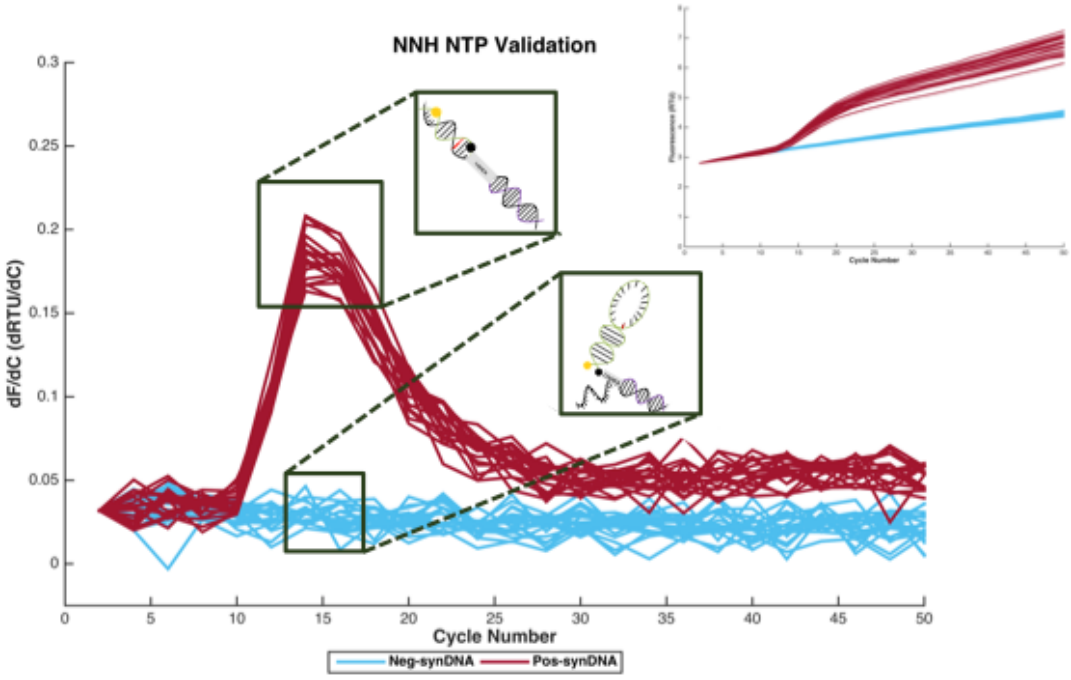


Figure 2.4. NNH-N TP synthetic DNA verification
Main graph shows derivative of NNH-N TP fluorescence versus cycle number. Top right graph shows raw fluorescence values. Raw fluorescence across all replicates was adjusted to begin at the most minimum fluorescence. The dark red is N SYN DNA and the cyan is P SYN DNA. Results of 20 replicates per sample shows 100% specificity and 100% sensitivity. NNH-N TP capture and detection region is bound for the N SYN DNA while only the capture region is bound for the P SYN DNA.

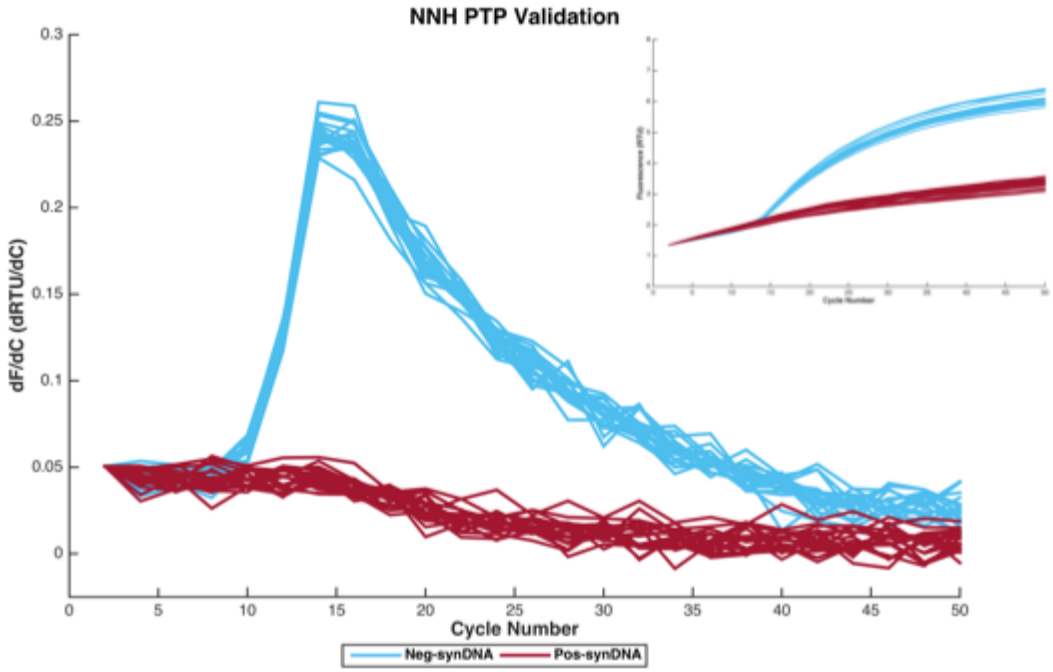


Figure 2.5. NNH-P TP synthetic DNA verification
Main graph shows derivative of NNH-P TP fluorescence versus cycle number. Top right graph shows raw fluorescence values. Raw fluorescence across all replicates was adjusted to begin at the most minimum fluorescence. The dark red is N SYN DNA and the cyan is P SYN DNA. Results of 20 replicates per sample shows 100% specificity and 100% sensitivity. NNH-P TP capture and detection region is bound for the P SYN DNA while only the capture region is bound for the N SYN DNA.

NNH-N and NNH-P TP genomic DNA verificaiton. SG showed all samples amplified efficiently (results not shown). **Figure 2.6** shows results for NNH-N TP while **Figure 2.7** shows results for NNH-P TP. Genomic DNA samples required pre-amplification prior to the introduction of a TP. Complex gDNA seemed to impact the specificity of the TP (results not shown). This similar type of result was obtained by Satterfield when evaluating CPs, which uses similar binding kinetics (Satterfield, 2014). It would have been preferred to not use pre-amplified gDNA, as this would have further

reduced run-time for the diagnostic and limited potential contaminations; however, pre-amplification is a necessary step to create more specificty in the diagnostic. From both graphs, it can be seen that there is a significant difference in fluorescence between positive and negative controls as well as positive and negative patient samples. Thus further verifying both the NNH-N and NNH-P TPs can correctly differentiate between samples with and without the 149G>A mutation. Furthermore, the results indicate both the NNH-N and NNH-P TPs nearly have 100% sensitivity and 100% specificity in the presence of patient gDNA. This is based on a limited number of available positive patient samples though so these numbers may decrease as more samples are tested.

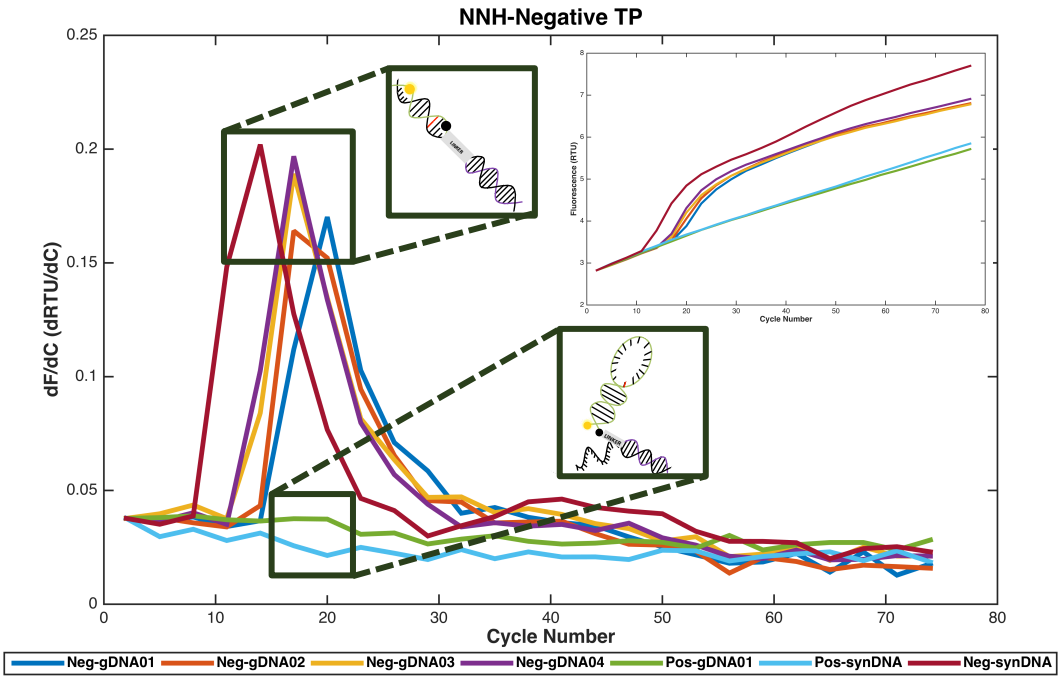


Figure 2.6. NNH-N TP genomic DNA verification
Main graph shows derivative of NNH-N TP fluorescence versus cycle number. Top right graph shows raw fluorescence values. Raw fluorescence across all replicates was adjusted to begin at the most minimum fluorescence. The dark blue, orange, yellow, purple, green, cyan, and dark blue lines represent NNH-N FFPE patient 1 tissue sample, NNH-N FFPE

patient 2 tissue sample, NNH-N FFPE patient 3 tissue sample, NNH-N FFPE patient 4 tissue sample, NNH-P FFPE patient 1 tissue sample, P SYN DNA, and N SYN DNA, respectively. Samples were run with 5 replicates and average of those replicates are displayed in the first derivate and raw fluorescence graphs. Results of 4 negative patient replicates and 1 positive patient replicate suggests nearly 100% specificity and 100% sensitivity. NNH-N TP capture and detection region is bound for the NNH-N patient samples and N SYN DNA while only the capture region is bound for the NNH-P patient sample and P SYN DNA.

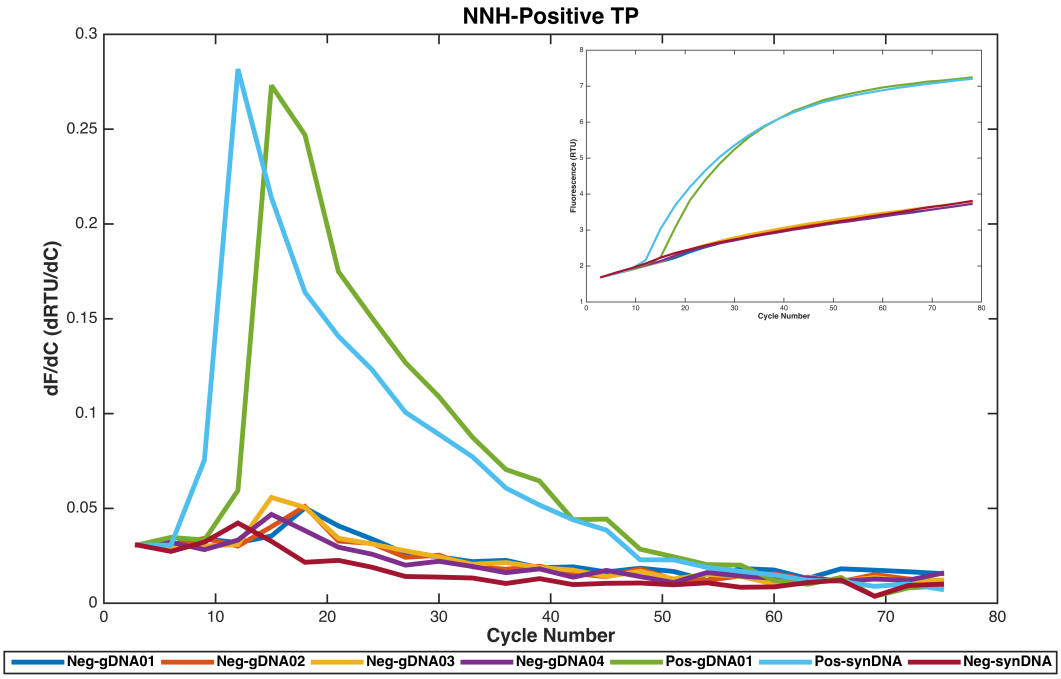


Figure 2.7. NNH-P TP genomic DNA verification
Main graph shows derivative of NNH-P TP fluorescence versus cycle number. Top right graph shows raw fluorescence values. Raw fluorescence across all replicates was adjusted to begin at the most minimum fluorescence. The dark blue, orange, yellow, purple, green, cyan, and dark blue lines represent NNH-N FFPE patient 1 tissue sample, NNH-N FFPE patient 2 tissue sample, NNH-N FFPE patient 3 tissue sample, NNH-N FFPE patient 4 tissue sample, NNH-P FFPE patient 1 tissue sample, P SYN DNA, and N SYN DNA, respectively. Samples were run with 5 replicates and average of those replicates are displayed in the first derivate and raw fluorescence graphs. Results of 4 negative patient replicates and 1 positive patient replicate suggests nearly 100% specificity and 100% sensitivity. NNH-P TP capture and detection region is bound for the NNH-P patient sample and P SYN DNA while only the capture region is bound for the NNH-N patient samples and N SYN DNA.

Assessment of the raw SG results from the genomic DNA data (data not shown), indicated there was amplification of DNA occurring at the 72°C extension and potentially 44°C annealing, 55°C annealing, or both. Though Taq polymerase works optimally at 72°C, it can still work at lower temperatures, just less efficiently. Using excel, a mathematical model (results not shown) was created to try and match the raw SG results and determine at which temperatures the DNA was amplifying during qPCR. Results indicated amplification of DNA occurred at 72°C, 44°C, and 55°C. Prior to this assessment, fluorescence results were reported as being for cycles 1-25, however, with this new information, fluorescence was reported for cycles 1-75. This impacted the derivative as the change in fluorescence was divided by a difference of 3 cycles instead of 1 cycle. This primarily impacted the maximum fluorescence obtained but not the difference in fluorescence between the positive and negative samples as both samples were decreased by the same magnitude. Though overall fluorescence is minimal (<0.3), it is relative and can change with every LightCycler machine.

To verify the output products of qPCR, 2 samples (N SYN DNA + SG and P SYN DNA + SG) were prepared to be 20 µL each diluted with water to have 1 pg/µL of specified synthetic DNA, 1 µM FP, 1 µM RP, and 1X SG. Twenty-µL of each sample was pipetted into a 96-well plate, which was loaded in a Roche LightCycler 480. Cycling conditions were an initial denaturation of 95°C for 30 s, primer annealing of 66°C for 20 s, extension of 72°C for 30 s, denaturation of 95°C for 30 s, TP annealing of 44°C for 2 min, denaturation of 95°C for 30 s, TP annealing of 55°C for 2 min for a total of 25 repeats. A fluorescence measurement was made after the extension phase and after each

TP annealing phase. Following qPCR, 12 µL master samples were created with 10 µL of each sample from the 96-well plate combined with 6X Gel Loading Dye diluted to 1X. The 12 µL samples were then run with a 1 Kb plus DNA ladder and 100 bp DNA ladder at 110V for 40 min on a 1% agarose gel prepared with 1X TAE buffer **(Figure 2.8)**. Results show a large smear for both samples (**Figure 2.8**, lanes 2 and 3), which is most likely due to the large amount of sample loaded. Furthermore, results primarily show the amplified band of interest with no additional by-products despite indicated amplification from qPCR SG results.

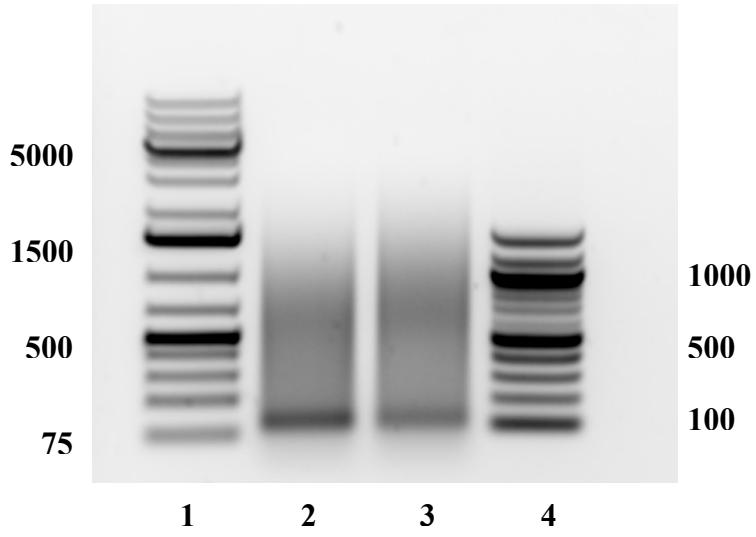


Figure 2.8. Amplified products following qPCR with three annealing steps
Lanes 1 and 4 are DNA ladders. Lane 1 is 1 Kb plus DNA ladder and lane 4 is a 100 bp DNA ladder. Lane 2 is the results of the amplified products from P SYN DNA. Lane 3 is the results of the amplified products from N SYN DNA.

Simulated synthetic NNH carrier. SG showed all samples amplified efficiently (results not shown). **Figure 2.9** shows results for both NNH-N TP and NNH-P TP. Both

TPs will equally bind to both target and non-target sequences due to their capture regions. However, only the TPs bound to the target sequence by the capture region will fluoresce. Since this is only half of the TPs, only half the fluorescence will be produced. From the graph, it can be seen that this theory does align with the results. The NNH-N and NNH-P TPs have the ability to not only identify homozygous individuals but also heterozygous. Thus providing the potential to develop the diagnostic into a screening tool.

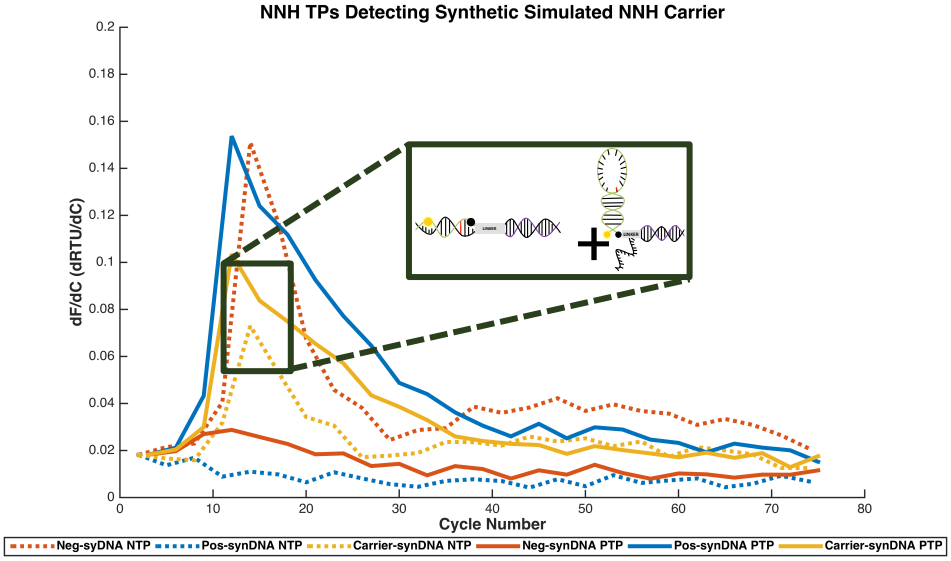


Figure 2.9. NNH-N and NNH-P TPs synthetic simulated DNA carrier verification First derivative of NNH-N and NNH-P TP fluorescence versus cycle number. Raw fluorescence across all replicates was adjusted to begin at the most minimum fluorescence with respect to TP assessed. The orange, dark blue, and yellow lines represent N SYN DNA, P SYN DNA, and 1:1 ration of N SYN DNA:PSYN DNA, respectively. The dashed lines represent the NNH-N TP, and the solid lines represent the NNH-P TP. Samples were run with 3 replicates and average of those replicates are displayed in the first derivate. Results suggest an ability to detect NNH carriers. For carriers, the NNH TPs capture regions can equally bind to either N SYN DNA or P SYN DNA, however, the detection region will only bind for the corresponding target DNA, producing half the fluorescence.

Conclusion

TPs have been developed as a diagnostic tool for NNH. The NNH-N TP was designed to detect a wild-type sequence, without the 149G>A mutation, while the NNH-P TP was designed to detect the 149G>A mutation. Verification of the TPs was performed with P SYN DNA and N SYN DNA as well as NNH-N patient FFPE liver gDNA and NNH-P patient FFPE liver gDNA. Results suggest 100% specificity and 100% sensitivity for SYN DNA as well as patient gDNA. The specificity and sensitivity may decrease, as the number of positive patient samples available increases.

As more NNH-P patient samples become available, they will be tested through this diagnostic to build the sample size and continue verification of the sensitivity and specificity of the TP diagnostic test. Future work will also look at beginning clinical trials for detecting carriers of NNH, as this type of sample is more readily available than the NNH-P sample type.

In the meantime, melting curves will be re-created for the NNH-N and NNH-P TP optimization tests. The original melting curves show very little difference in fluorescence between the P SYN and N SYN data. After determining the greatest change in fluorescence of the probes occurs around cycle 6, a calculation was performed to determine the number of copies of DNA at that cycle number based on 1 pg/μL initial DNA. This copy number was then converted to a concentration for the melting curves. Previously, the melting curves were performed with 10 ng/μL of SYN DNA, however at this concentration, the fluorescent difference between the NNH-N and NNH-P TPs was minimal. The melting curves will be re-run with an initial concentration of 65 pg/μL to

obtain the greatest difference in fluorescence.

A contamination experiment will also be conducted to determine the specificity of the probes. P SYN DNA will range form 100% to 0.01% contamination in N SYN DNA and run with NNH-P TP. Similarly, N SYN DNA will range form 100% to 0.01% contamination in P SYN DNA but be run with NNH-N TP. This will help to validate the specificity of the probes and help develop the robustness of the test.

This thesis body of work has focused on the role of amplification and detection of the NNH diagnostic. Assuming no automation and over exaggerating of preparation times, it will take approximately 7 hours of technician time to prepare the diagnostic (from pre-amplification to running the samples with the TP) and approximately 12 hours for a total run-time. For these reasons, it estimated that the diagnostic could have an approximate turn-around time of 48 hours from the time the patient enters PCH, compared to the current 3-4 week timeframe. Furthermore, most of the costs are for the amplification and detection role of the NNH diagnostic. This cost breakdown is shown in **Table 2.2**. This cost is created for one run and is inclusive of a positive control, negative control, and one patient sample. Each of these samples is run under the assumption of 3 replicates each for the NNH-N TP and NNH-P TP as well as 3 replicates each for SG. The most expensive cost is the clinical technician labor, which was estimated at $24.48/hr, as based on the 2016 median pay outlined by Bureau of Labor Statistics (Bureau of Labor Statistics, 2018). Because the cost of the item purchased does not include shipping or tax and because the cost breakdown does not include sample collection and extraction, the overall cost of the diagnostic was estimated to be

approximately $250 for the in-house diagnostic, compared to the current $700 for outsourcing the diagnostic.

Future work could continue to decrease the time and cost for this diagnostic by introducing automated processes, finding quicker methods for pre-amplification of gDNA, and assessing if a 2 min annealing time is needed per TP. Even if a 2 min annealing step is required, new TPs could be designed to have the same annealing temperature to reduce on the number of annealing steps required and thus overall time.

This diagnostic has the potential to be life-saving in cases of NNH. Thus far, it has shown to have nearly 100% specificity and 100% sensitivity in the presence of positive and negative patient FFPE liver gDNA samples. When evaluating the future sample type: blood, the use of a QIAamp DNA Blood Mini Kit produces, at minimum, a 400 ng yield of DNA from a 20 µL whole blood sample (equivalent to a drop of blood) (Qiagen, 2018). This yield is more than enough to perform the pre-amplification of gDNA since this requires anywhere from 50 – 250 ng of DNA for a 50 µL reaction (Barrick Lab, 2017). Ultimately, the low volume of blood, low cost, quick turnaround time, and high specificity and sensitivity makes the NNH TP diagnostic a favorable molecular method for use by PCH.

This diagnostic creates a translatable protocol for any point mutation detection. Further investigation of the diagnostic has demonstrated its potential as a screening tool, as the probes are capable of detecting a simulated SYN DNA carrier. Ultimately, early diagnosis and screening can help with genetic counseling and ensuring families have the information needed to understand the genetic disorder and make necessary decisions.

Table 2.2. Cost breakdown for amplification and detection role of NNH diagnostic
Cost breakdown is inclusive for one positive control, one negative control, and one patient sample with 3 replicates each of the NNH-N TP and NNH-P TP as well as 3 replicates of SG. Cost of items does not include shipping or tax.

Item	Quantity Purchased	Cost	Quantity Used	Cost per Test
FP (1 μM)	6951 μL	$17.00	2.3 μL	$0.008
RP (1 μM)	6968 μL	$22.00	2.3 μL	$0.01
P SYN DNA (4 nmol)	132525.6 μL	$85.60	0.0018 μL	$1.16E-06
N SYN DNA (4 nmol)	132589.6 μL	$85.60	0.0018 μL	$1.16E-06
GoTaq Green 2X Master Mix (1,000 rxns)	25000 μL	$379.00	75.0 μL	$1.14
GoTaq Colorless Master Mix (1,000 rxns)	25000 μL	$379.00	90.0 μL	$1.36
NNH-N TP (200 nmol)	2000 μL	$490.00	0.18 μL	$0.04
NNH-P TP (200 nmol)	2000 μL	$490.00	0.18 μL	$0.04
SYBR Green I (500 rxns)	5000 μL	$391.00	90.0 μL	$7.04
PCR Purification Kit	250 rxns	$550.00	3 rxns	$6.60
LightCycler480 Multiwell Plate 96	50 plates	$467.74	30 wells	$2.92
Consumables	-	-	-	$10
Technician Labor	-	$24.48/hr	7 hr	$153.64
				Total: $182.81

CHAPTER 3

AMPLFICIATION OF BOWTIE-BARCODED ORIGAMI WITH CAPTURED TCRA AND TCRB USING COOPERATIVE PRIMERS

Introduction

The adaptive immune system has a broad antigen receptor repertoire as well as its specificity (Attaf, Huseby, & Sewell, 2015). It is comprised of B and T cells, where each cell has its own receptor for recognizing an antigen. The receptor on the T cell is known as the T cell receptor (TCR). It is heterodimer, as it is composed of two chains; there are two sets of chains used to form the TCR: (1) α and β chains or (2) δ and γ chains (Attaf et al., 2015). The αβ TCRs are more common than δγ, comprising around 90% of circulating T cells, therefore they are the primary focus of this research (Mirzaei, Mirzaei, Lee, Hadjati, & Till, 2016).

It has been previously suggested that the large, complex repertoire for αβ TCRs is dependent on genetic factors; one impact stemming from these genetic factors is V(D)J recombination (Attaf et al., 2015). Each chain has a V, J, and C region, and the β chain has an additional D region (Davis & Bjorkman, 1988). There are multiple V(D)J gene segments, and these gene segments are randomly assembled next to a specific C region to create various combinations for the α and β chains. Further pairing of the α chain with the β chain to form the TCR develops an even greater number of combinations **(Figure 3.1)**. Knowing the assignment of the particular α and β chain pairing can help to identify TCR specificity, thus leading to development of potential diagnostic biomarkers as well as immunotherapy options (Mirzaei et al., 2016; Redmond, Poran, & Elemento, 2016).

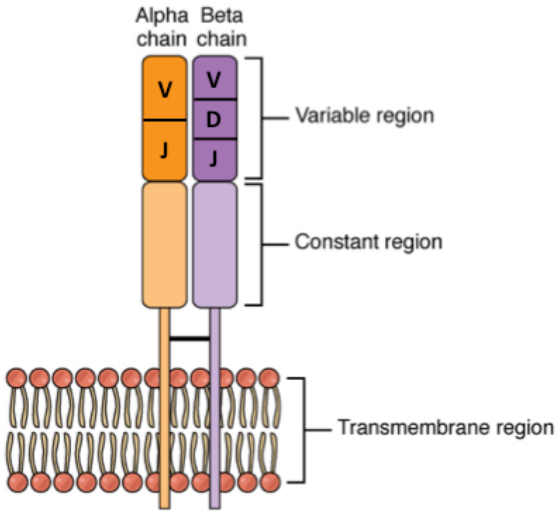


Figure 3.1. TCR V(D)J Recombination with α and β Chains
The alpha chain (shown in orange) can be made up through a variety of V and J combinations while the beta chain (shown in purple) can be made up through a variety of V, D, and J combinations. The combination of the alpha and beta chain creates large repertoire for the TCR. Figure adapted from OpenStax College "Illustration from Anatomy & Physiology Connexions Web site".

Specificity of TCR is dependent on the pairing of the α and β chains; therefore, there is a need to develop a system that identifies both the α and β chains gene sequences of a single-cell. To create a high-throughput system, this further needs to identify α and β chains of a single-cell but within a heterogeneous cell population. Drs. Louis Schoettle and Joseph Blattman have developed a solution in which a nanostructure is transfected into individual cells of a heterogeneous cell population to capture and protect α and β messenger ribonucleic acid (mRNA). In the nanostructure is a bowtie-barcoded origami strand **(Figure 3.2)**. The 12 bp barcodes (BCs) on the left and right side of a single origami strand are complimentary and unique to a single origami strand. This BC linkage provides a system for pairing α and β mRNA derived from a single-cell **(Figure 3.3)**.

Figure 3.2. Bowtie-barcoded origami strand pre-mRNA capture
Diagram from Dr. Louis Schoettle. The BC on the left of the strand matches the BC on the right. Each origami strand has its own unique BC.

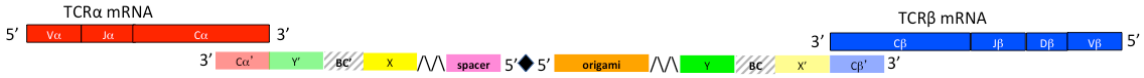


Figure 3.3. Bowtie-barcoded origami strand with captured mRNA
Diagram from Dr. Louis Schoettle. Both the α and β mRNA strands are captured from within the same cell.

Following intracellular capture, the cell population is lysed to obtain the bowtie-barcoded origami strand with the captured mRNA. Reverse transcription (RT) is then performed on the strand, to elongate the bowtie-barcoded origami strand with complementary DNA (cDNA) **(Figure 3.4)**.

Figure 3.4. RT of origami captured α and β mRNA
Diagram from Dr. Louis Schoettle. The bowtie-barcoded origami strand is further elongated to include the α and β mRNA strands. These mRNA strands are tied to the same BC so it is known what pairing of α and β mRNA makes up the TCR of a single-cell.

Using PCR, this cDNA is then amplified for sequencing. The Vα primer binds to the left side of the bowtie-barcoded origami strand while the Vβ primer binds to the right side **(Figure 3.5)**. The products generated from the Vα and Vβ primers **(Figure 3.6)** are then used as the template for the X and Y primers, respectively **(Figure 3.7)**.

Figure 3.5. Vα and Vβ primer anneal and extension
Diagram from Dr. Louis Schoettle. The Vα and Vβ primers bind to the ends of the bowtie-barcoded origami to create a copy of the α and β gene segments with the BCs. The elongation of the primers ends just before the zigzags, as this is the location of a linker.

Figure 3.6. Products generated following PCR of Vα and Vβ primers
Diagram from Dr. Louis Schoettle. These are the products in the solution following Vα and Vβ elongation. The Vα and Vβ products must first be generated prior to elongation with the X and Y primers.

Figure 3.7. X and Y primers theoretical anneal and extension
Diagram from Dr. Louis Schoettle. The X and Y primers bind to the ends of the Vα and Vβ products to create and additional copy of the α and β gene segments with the BCs.

When the PCR of the bowtie-barcoded origami template is performed using only the left-side primers (Vα + X), the expected left-product is produced (**Figure 3.8**, lane 2). Similarly, when using only the right-side primers (Vβ + Y), the expected right-product is produced (**Figure 3.8**, lane 3). However, when PCR of the bowtie-barcoded origami template is performed with the left-side primers and right-side primers together (Vα + X + Vβ + Y), the two bands of the expected products are not produced (**Figure 3.8**, lane 4). This is due to the symmetry within the bowtie-barcoded origami template. The X primer can bind to either the Vα product or the right-side of the bowtie-barcoded origami

template, and the Y primer can bind to either the Vβ product or the left-side of the bowtie-barcoded origami template **(Figure 3.9)**. As PCR cycles continue, conditions favor the smaller product **(Figure 3.10)**. Because of this amplification, the α and the β V(D)J segments cannot be associated with a BC and thus paired. The left- and right-side PCR of the bowtie-barcoded origami template cannot be performed separately, as each origami will have its own BC and separation of the primers will provide data for only the α or β chain.

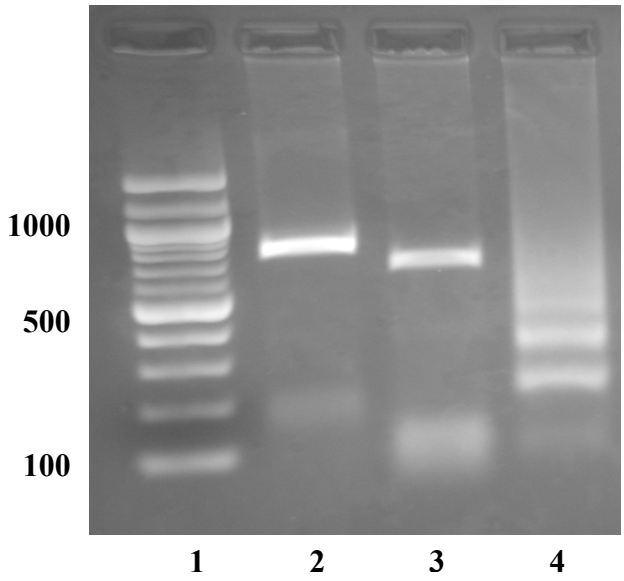


Figure 3.8. Amplified products using Vα, Vβ, X, and Y primers
RT, PCR, and gel electrophoresis produced by Dr. Louis Schoettle. Lane 1 is 100 bp DNA ladder. Lane 2 is the result of the amplified products from Vα + X primers. Lane 3 is the result of the amplified products from Vβ + Y primers. Lane 4 is the result of the amplified products from Vα + X + Vβ + Y primers.

Figure 3.9. Two binding locations for X and Y primers
Diagram adapted from Dr. Louis Schoettle. The X and Y primers have two potential binding locations. The X primer can either bind to the Vα product or the right side of the origami product. The Y primer can either bind to the Vβ product or the left side of the origami product.

Figure 3.10. Product favored by X and Y primers
Diagram adapted from Dr. Louis Schoettle. Both X and Y primers favor the origami product since it produces a smaller product. These products created by the X and Y primers include the necessary BCs but it is unassociated with a α or β gene segment.

Schoettle and Blattman requested a collaboration to resolve this problem with the PCR amplification. CP technology was utilized to overcome the formation of the short-products. CPs consist of a short primer and capture region. When referring to the earlier cooperativity models mentioned in the introduction, the functional region of the CP is known as the short primer. The short primer consists of a sequence that has a low melting temperature (T_m) and thus would not normally amplify a template at higher reaction temperature. This short primer is connected to a capture region via a PEG linker. The capture region keeps the short primer in close proximity to the template to then encourage binding and extension. Connecting the 5' end of both primers via a PEG linker encourages displacement of the capture region following the binding of the short primer

due to the strain on the inverted linkage (Satterfield, 2014). This ensures Taq polymerase extends from the short primer until the end of the DNA template. Furthermore, the capture region has a blocking carbon chain on the 3' end to ensure Taq polymerase does not begin extension from the capture region but rather the short primer (Satterfield, 2014). Herein this chapter is the design of the two CPs for the bowtie-barcoded origami template as well as initial testing **(Figure 3.11)**.

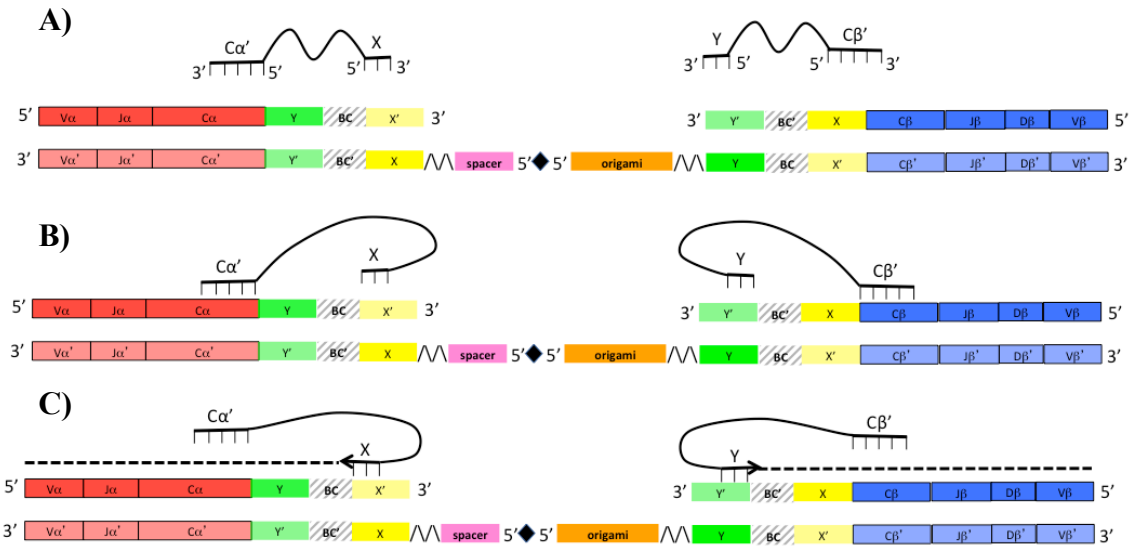


Figure 3.11. CP design to bind to Vα and Vβ PCR products
Diagram adapted from Dr. Louis Schoettle. **[A]** The left CP is composed of two regions: (1) a short primer, X, which binds to X' and (2) a capture region, Cα', which binds to Cα. The right CP is composed of two regions: (1) a short primer, Y, which binds to Y' and (2) a capture region, Cβ', which binds to Cβ. A PEG linker connects the short and capture regions for both CPs. **[B]** The capture region first binds to the target sequence, bringing the short primer in closer proximity to its target. **[C]** The inverted PEG linker encourages displacement of the capture region as the short primer binds and elongates.

Materials and Methods

Brent Satterfield's mathematical model for CPs was adapted from his earlier model derived for cooperative TPs (Satterfield, 2014). Satterfield defined the effective primer binding efficiency (E_{ff}) as **Eq. (3.1)**:

$$E_{ff} = \frac{C_{p,func}+C_{both}}{T_0} = \frac{(K_{func}+\left[\frac{1\ molecule}{\left(\frac{4}{3}\pi l^3\right)N_A}\right]K_{func}K_{cap})P_0}{1+K_{both}P_0} \quad (3.1)$$

Where T_0 is the initial target concentration and P_0 is the initial primer concentration. With this model, it is assumed the equilibrium constant of the functional region, in this case the short primer, is greater in the presence of a target than non-target sequence. Furthermore, the equilibrium constant of the capture region, in this case the capture region, is the same in the presence of a target and non-target sequence. This is because the target and non-target strand have the same capture regions but differing functional regions. However, in the case of the CPs for the bowtie-barcoded origami, the equilibrium constant of the short primer in the presence of a target and non-target sequence is the same; however, the equilibrium constant of the capture region is greater in the presence of a target than non-target sequence. This is because the target and non-target strand have the same functional regions but differing caption regions. For these reasons, the equation to calculate E_{ff} was modified to be **Eq. (3.2)**:

$$E_{ff} = \frac{C_{p,cap}+C_{both}}{T_0} = \frac{(K_{cap}+\left[\frac{1\ molecule}{\left(\frac{4}{3}\pi l^3\right)N_A}\right]K_{func}K_{cap})P_0}{1+K_{both}P_0} \quad (3.2)$$

Target sequences of the products created by Vα and Vβ primers were used to generate the left CP (compatible with the Vα product) and the right CP (compatible with

the Vβ product). The short primer and capture region of the left CP targets X' and Cα, respectively; while the short primer and capture region of the right CP targets Y' and Cβ, respectively **(Figure 3.11)**.

Creating Short Primer, Capture Region, Target, and Non-Target Sequences. In silico prototyping was utilized to perform calculations of possible short primer and capture region combinations in the presence of both a target and non-target sequence. For the first iteration, short primers for the left CP were generated by taking 15 bps of the complement X' sequence either starting at the 5' end or ending at the 3' end of the 20 bp sequence. To generate a variety of short primers starting at the 5' end, one bp was removed from the 3' end to gradually decrease the bp length until it was only 4 bps long. This same process was repeated with the short primers ending at the 3' end; however, the one bp was gradually removed from the 5' end. This process generated 24 varying short primers for the left CP. Capture regions for the left CP were generated by taking 22 bps of the complement Cα sequence ending at the 3' end. One bp was then removed from the 5' end to gradually decrease the bp length until it was only 10 bps long. This process generated 13 varying capture regions for the left CP. This same process was repeated for the right CP using the complement Y' sequence for the short primer and complement Cβ sequence for the capture region (see **Appendix A** for all first iteration sequences).

Of the left cDNA sequence, 5'-Vα-Jα-Cα-Y-BC-X'-3', the target sequence for the both the short and capture region of the left CP was 5'-Cα-Y-X'-3'. Similarly, of the right cDNA sequence, 5'-Vβ-Jβ-Cβ-X-BC-Y'-3', the target sequence for the short and capture region of the right CP was 5'-Cβ-X-Y'-3'. Only Cα/Cβ are retained, as this sequence

remains constant with each mRNA captured, while V, (D), and J are variable due to V(D)J recombination. Furthermore, the barcode is excluded, as it will be a different sequence with each bowtie-barcoded origami generated.

Of the bowtie-barcoded origami sequence, 3'-Vα'-Jα'-Cα'-Y'-BC'-X-5'•5'-Y-BC-X'-Cβ'-Jβ'-Vβ'-3', the non-target sequence for both the short and capture region of the left CP was 5'-Y-X'-Cβ'-3'. Whereas the non-target sequence for both the short and capture region of the right CP was 5'-X-Y'-Cα'-3'. The first iteration produced results with a primer efficiency that has high in both target and non-target sequences and thus required further iterations.

For the second iteration, short primers for the left and right CP were user-generated 12 bp sequences with varying %GC content: 66.7, 58.3, and 50. Two alternatives were generated per %GC content by altering an A for a T or vice versa to create subtle changes in T_m. This process generated six short primers each for the left CP and right CP. Capture regions for the left and right CP were the same as the first iteration (see **Appendix B** for second iteration sequences).

Of the left cDNA sequence, 5'-Vα-Jα-Cα-Y_n-BC-X'_n-3', where X'_n is the complement of the particular X' short primer being evaluated and Y'_n is the complement of the particular Y' short primer being evaluated, the target sequence for the short primer of the left CP was 5'-Cα-X'_n-3', and the target sequence for the capture region of the left CP was 5'-Cα-3'. Y_n was further eliminated from the short primer target, as there would need to be 6 target sequences created per left CP short primer to accommodate for all possible Y_n variations. X'_n and Y_n were eliminated from the capture region target, as

there would need to be 36 target sequences created per left CP capture region to accommodate for all possible variations in both X'_n and Y_n. Ultimately, these components were removed from the target sequences to limit the number of outputted combinations for E_{ff}. Similarly, of the right cDNA sequence, 5'-Vβ-Jβ-Cβ-X_n-BC-Y'_n-3', the target sequence for the short primer of the right CP was 5'-Cβ-Y'_n-3', and the target sequence for the capture region of the right CP was 5'-Cβ-3'.

Of the bowtie-barcoded origami sequence, 3'-Vα'-Jα'-Cα'-Y'_n-BC'-X_n-5'•5'-Y_n-BC-X'_n-Cβ'-Jβ'-Vβ'-3', the non-target sequence for the short primer of the left CP was 5'-X'_n-Cβ'-3', and the non-target sequence for the capture region of the left CP was 5'-Cβ'-3'.Whereas the non-target sequence for the short primer of the right CP was 5'-Y'_n-Cα'-3', and the non-target sequence for the capture region of the right CP was 5'-Cα'-3'

For the third iteration, the short primers and capture regions with the best E_{ff} combination from the second iteration were used, however, the complement of these 12 bp short primer sequences were incorporated into new, longer 20 bp X' and Y' target sequences (see **Appendix C** for third iteration sequences).

Of the left cDNA sequence, 5'-Vα-Jα-Cα-Y-BC-X'-3', the target sequence for the both the short and capture region of the left CP was 5'-Cα-Y-X'-3'. Similarly, of the right cDNA sequence, 5'-Vβ-Jβ-Cβ-X-BC-Y'-3', the target sequence for the short and capture region of the right CP was 5'-Cβ-X-Y'-3'.

Of the bowtie-barcoded origami sequence, 3'-Vα'-Jα'-Cα'-Y'-BC'-X-5'•5'-Y-BC-X'-Cβ'-Jβ'-Vβ'-3', the non-target sequence for both the short and capture region of the left CP was 5'-Y-X'-Cβ'-3'. Whereas the non-target sequence for both the short and

capture region of the right CP was 5'-X-Y'-Cα'-3'.

Creating Excel File Mathematical Model. After determining the short primer, capture region, target, and non-target sequences, an excel file was created to calculate the E_{ff} of the CPs. The file had eight excel sheets: SEQUENCES, CP, LCP THERMO, LCP KEFF, LCP EFF, RCP THERMO, RCP KEFF, and RCP EFF **(Figure 3.12)**. The SEQUENCES sheet included the sequences making up the bowtie-barcoded origami. Using bioinformatics (https://www.bioinformatics.org/sms/rev_comp.html) ("Reverse Complement," n.d.), complements were derived from the sequences and presented in a 5' → 3' format **(Figure 3.12)**.

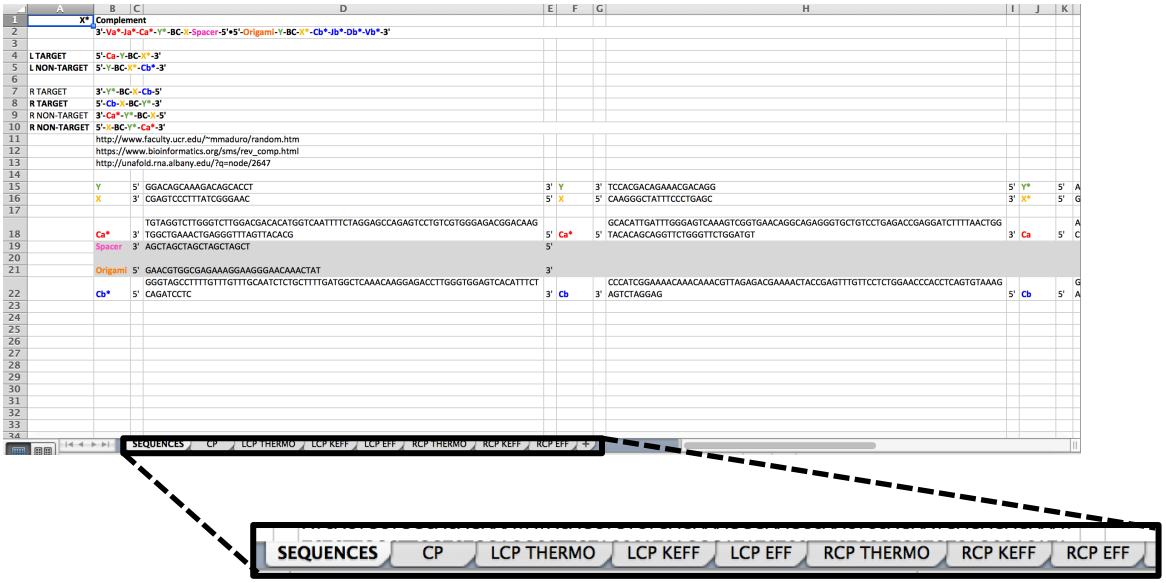


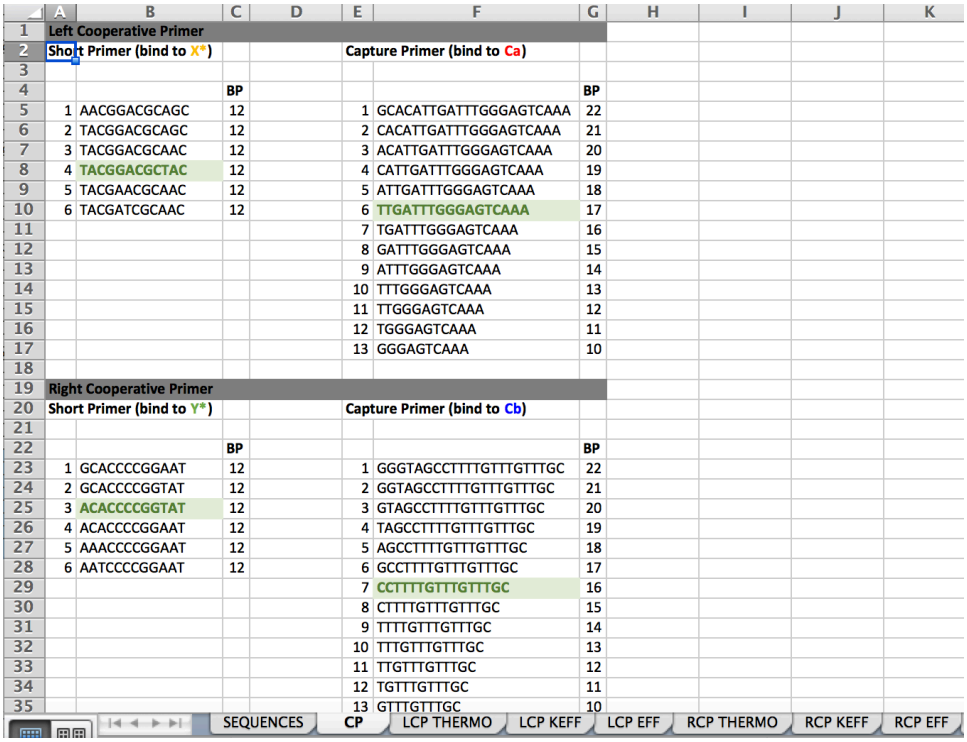
Figure 3.12. Excel file for CP Design
There are eight excel sheets used for the CP Design. The SEQUENCES sheet includes all target and non-target sequences for the CPs. The CP sheet includes all possible combinations of short primers and capture regions for both the left and right CPs. The LCP THERMO, LCP KEFF, and LCP EFF all refer to thermodynamic equations for the left CP. The RCP THERMO, RCP KEFF, and RCP EFF all refer to thermodynamic equations for the right CP.

The CP sheet generates the short primer and capture region for both the left CP and right CP. Because of the short primer creation differences in the first and second iterations, the CP sheet uses different formulas to create the short primers for these iterations **(Figure 3.13 and 3.14)**. The third iteration has CP sequences already defined based on the output determined by the second iteration, so there is no need to generate new short primers and capture regions.

	A	B	C	D	E	F	G	H	I	J
1	**Left Cooperative Primer**									
2	**Short Primer (bind to X*)**							**Capture Primer (bind to Ca)**		
3	CAAGGGCTATTTCCCTGAGC									
4	**5' end 3' reduced**		**BP**	**3' end 5' reduced**		**BP**				**BP**
5	1	CAAGGGCTATTTCCC	15	13	GCTATTTCCCTGAGC	15		1	GCACATTGATTTGGGAGTCAAA	22
6	2	CAAGGGCTATTTCC	14	14	CTATTTCCCTGAGC	14		2	CACATTGATTTGGGAGTCAAA	21
7	3	CAAGGGCTATTTC	13	15	TATTTCCCTGAGC	13		3	ACATTGATTTGGGAGTCAAA	20
8	4	CAAGGGCTATTT	12	16	ATTTCCCTGAGC	12		4	CATTGATTTGGGAGTCAAA	19
9	5	CAAGGGCTATT	11	17	TTTCCCTGAGC	11		5	ATTGATTTGGGAGTCAAA	18
10	6	CAAGGGCTAT	10	18	TTCCCTGAGC	10		6	TTGATTTGGGAGTCAAA	17
11	7	CAAGGGCTA	9	19	TCCCTGAGC	9		7	TGATTTGGGAGTCAAA	16
12	8	CAAGGGCT	8	20	CCCTGAGC	8		8	GATTTGGGAGTCAAA	15
13	9	CAAGGGC	7	21	CCTGAGC	7		9	ATTTGGGAGTCAAA	14
14	10	CAAGGG	6	22	CTGAGC	6		10	TTTGGGAGTCAAA	13
15	11	CAAGG	5	23	TGAGC	5		11	TTGGGAGTCAAA	12
16	12	CAAG	4	24	GAGC	4		12	TGGGAGTCAAA	11
17								13	GGGAGTCAAA	10
18										
19	**Right Cooperative Primer**									
20	**Short Primer (bind to Y*)**							**Capture Primer (bind to Cb)**		
21	GGACAGCAAAGACAGCACCT									
22	**5' end 3' reduced**		**BP**	**3' end 5' reduced**		**BP**				**BP**
23	1	GGACAGCAAAGACAG	15	13	GCAAAGACAGCACCT	15		1	GGGTAGCCTTTTGTTTGTTTGC	22
24	2	GGACAGCAAAGACA	14	14	CAAAGACAGCACCT	14		2	GGTAGCCTTTTGTTTGTTTGC	21
25	3	GGACAGCAAAGAC	13	15	AAAGACAGCACCT	13		3	GTAGCCTTTTGTTTGTTTGC	20
26	4	GGACAGCAAAGA	12	16	AAGACAGCACCT	12		4	TAGCCTTTTGTTTGTTTGC	19
27	5	GGACAGCAAAG	11	17	AGACAGCACCT	11		5	AGCCTTTTGTTTGTTTGC	18
28	6	GGACAGCAAA	10	18	GACAGCACCT	10		6	GCCTTTTGTTTGTTTGC	17
29	7	GGACAGCAA	9	19	ACAGCACCT	9		7	CCTTTTGTTTGTTTGC	16
30	8	GGACAGCA	8	20	CAGCACCT	8		8	CTTTTGTTTGTTTGC	15
31	9	GGACAGC	7	21	AGCACCT	7		9	TTTTGTTTGTTTGC	14
32	10	GGACAG	6	22	GCACCT	6		10	TTTGTTTGTTTGC	13
33	11	GGACA	5	23	CACCT	5		11	TTGTTTGTTTGC	12
34	12	GGAC	4	24	ACCT	4		12	TGTTTGTTTGC	11
35								13	GTTTGTTTGC	10

SEQUENCES | CP | LCP THERMO | LCP KEFF | LCP EFF | RCP THERMO | RCP KEFF | RCP EFF | +

Figure 3.13. First iteration short primer and capture region possibilities for both left and right CPs
The top half is possible short primers and capture regions for the left CP. The bottom half is possible short primers and capture regions for the right CP. Column B and E are short primer possibilities and column I is capture region possibilities.



	A	B	C	D	E	F	G	H	I	J	K
1	Left Cooperative Primer										
2	Short Primer (bind to X*)				Capture Primer (bind to Ca)						
3											
4			BP				BP				
5	1	AACGGACGCAGC	12		1	GCACATTGATTTGGGAGTCAAA	22				
6	2	TACGGACGCAGC	12		2	CACATTGATTTGGGAGTCAAA	21				
7	3	TACGGACGCAAC	12		3	ACATTGATTTGGGAGTCAAA	20				
8	4	TACGGACGCTAC	12		4	CATTGATTTGGGAGTCAAA	19				
9	5	TACGAACGCAAC	12		5	ATTGATTTGGGAGTCAAA	18				
10	6	TACGATCGCAAC	12		6	TTGATTTGGGAGTCAAA	17				
11					7	TGATTTGGGAGTCAAA	16				
12					8	GATTTGGGAGTCAAA	15				
13					9	ATTTGGGAGTCAAA	14				
14					10	TTTGGGAGTCAAA	13				
15					11	TTGGGAGTCAAA	12				
16					12	TGGGAGTCAAA	11				
17					13	GGGAGTCAAA	10				
18											
19	Right Cooperative Primer										
20	Short Primer (bind to Y*)				Capture Primer (bind to Cb)						
21											
22			BP				BP				
23	1	GCACCCCGGAAT	12		1	GGGTAGCCTTTTGTTTGTTTGC	22				
24	2	GCACCCCGGTAT	12		2	GGTAGCCTTTTGTTTGTTTGC	21				
25	3	ACACCCCGGTAT	12		3	GTAGCCTTTTGTTTGTTTGC	20				
26	4	ACACCCCGGAAT	12		4	TAGCCTTTTGTTTGTTTGC	19				
27	5	AAACCCCGGAAT	12		5	AGCCTTTTGTTTGTTTGC	18				
28	6	AATCCCCGGAAT	12		6	GCCTTTTGTTTGTTTGC	17				
29					7	CCTTTTGTTTGTTTGC	16				
30					8	CTTTTGTTTGTTTGC	15				
31					9	TTTTGTTTGTTTGC	14				
32					10	TTTGTTTGTTTGC	13				
33					11	TTGTTTGTTTGC	12				
34					12	TGTTTGTTTGC	11				
35					13	GTTTGTTTGC	10				

SEQUENCES | CP | LCP THERMO | LCP KEFF | LCP EFF | RCP THERMO | RCP KEFF | RCP EFF

Figure 3.14. Second iteration short primer and capture region possibilities for both left and right CPs
The top half is possible short primers and capture regions for the left CP. The bottom half is possible short primers and capture regions for the right CP. Column B is short primer possibilities and column F is capture region possibilities.

In the first iteration CP sheet **(Figure 3.13)**, for the left CP, cell A3 is "=SEQUENCES!H16", which pulls the 5' → 3' complement of X' from the SEQUENCES sheet. Cell B5 is "=LEFT(A3,LEN(A3)-5)", which removes the last 5 bps of the X' complement sequence defined in cell A3. Cell B6 is "=LEFT(B5,LEN(B5)-1)", which removes the last bp from the sequence defined in the cell above. This formula was dragged down to cell B16 to create the first set of 12 short primers. Cell E5 is "=RIGHT(A3,LEN(A3)-5)", which removes the first 5 bps of the X' complement sequenced defined in cell A3. Cell E6 is "=RIGHT(E5,LEN(E5)-1)", which removes the

first bp from sequence defined in the cell above. This formula was dragged down to cell E16 to create the last set of 12 short primers. Cell I5 is "=LEFT(SEQUENCES!H18,LEN(SEQUENCES!H18)-80)", which pulls the first 22 bps of the 5' → 3' complement of Cα from the SEQUENCES sheet. Cell I6 is "=RIGHT(I5,LEN(I5)-1)", which removes the first bp from sequence defined in the cell above. This formula was dragged down to cell E17 to create the capture regions.

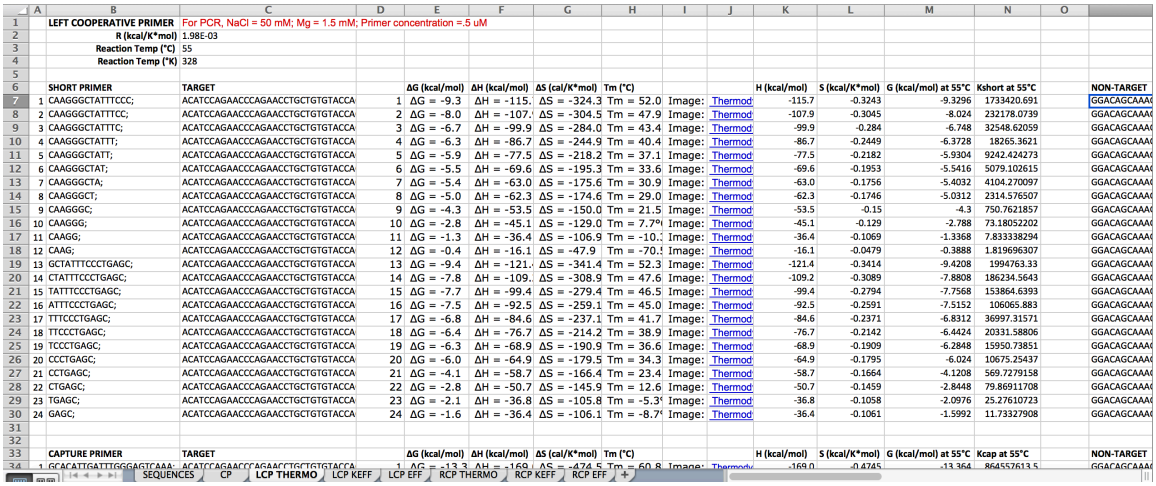
For the right CP, cell A21 is "=SEQUENCES!D15", which pulls the 5' → 3' complement of Y' from the SEQUENCES sheet. Cell B23 is "=LEFT(A21,LEN(A21)-5)", which removes the last 5 bps of the Y' complement sequence defined in cell A21. Cell B24 is "=LEFT(B23,LEN(B23)-1)", which removes the last bp from the sequence defined in the cell above. This formula was dragged down to cell B34 to create the first set of 12 short primers. Cell E23 is "=RIGHT(A21,LEN(A21)-5)", which removes the first 5 bps of the Y' complement sequenced defined in cell A21. Cell E24 is "=RIGHT(E23,LEN(E23)-1)", which removes the first bp from sequence defined in the cell above. This formula was dragged down to cell E34 to create the last set of 12 short primers. Cell I23 is "=LEFT(SEQUENCES!D22,LEN(SEQUENCES!D22)-80)", which pulls the first 22 bps of the 5' → 3' complement of Cβ from the SEQUENCES sheet. Cell I24 is "=RIGHT(I23,LEN(I23)-1)", which removes the first bp from sequence defined in the cell above. This formula was dragged down to cell E35 to create the capture regions.

In the second iteration CP sheet **(Figure 3.14)**, for the left CP cell B5 is "=SEQUENCES!H15", which pulls the first 5' → 3' complement of X' from the

SEQUENCES sheet. This formula was dragged down to cell B10 to capture all short primer variations. Cell F5 is "=LEFT(SEQUENCES!H29,LEN(SEQUENCES!H29)-80)", which pulls the first 22 bps of the 5' → 3' complement of Cα from the SEQUENCES sheet. Cell F6 is "=RIGHT(F5,LEN(F5)-1)", which removes the first bp from sequence defined in the cell above. This formula was dragged down to cell F17 to create the capture regions.

For the right CP, cell B23 is "=SEQUENCES!D22", which pulls the first 5' → 3' complement of Y' from the SEQUENCES sheet. This formula was dragged down to cell B28 to capture all short primer variations. Cell F23 is "=LEFT(SEQUENCES!D33,LEN(SEQUENCES!D33)-63)", which pulls the first 22 bps of the 5' → 3' complement of Cβ from the SEQUENCES sheet. Cell F24 is "=RIGHT(F23,LEN(F23)-1)", which removes the first bp from sequence defined in the cell above. This formula was dragged down to cell F35 to create the capture regions.

The LCP THERMO and RCP THERMO sheets hold the thermodynamic values for the short primer as well as the capture region in presence of target and non-target sequences for the left CP and right CP, respectively. These sheets reformat the ΔH and ΔS values to be in similar units and calculate the ΔG in terms of a dependent reaction temperature. Finally, using these values, the equilibrium constant is calculated. The first, second, and third iterations use the same calculations **(Figure 3.15)**.



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	
1		LEFT COOPERATIVE PRIMER	For PCR, NaCl = 50 mM; Mg = 1.5 mM; Primer concentration =.5 uM													
2		R (kcal/K*mol)	1.98E-03													
3		Reaction Temp (°C)	55													
4		Reaction Temp (°K)	328													
5																
6		SHORT PRIMER	TARGET		ΔG (kcal/mol)	ΔH (kcal/mol)	ΔS (cal/K*mol)	Tm (°C)			H (kcal/mol)	S (kcal/K*mol)	G (kcal/mol) at 55°C	Kshort at 55°C		NON-TARGET
7	1	CAAGGGCTATTTCCC;	ACATCCAGAACCCAGAACCTGCTGTGTACCA	1	ΔG = -9.3	ΔH = -115.	ΔS = -324.3	Tm = 52.0	Image:	Thermod	-115.7	-0.3243	-9.3296	1733420.691		GGACAGCAAA
8	2	CAAGGGCTATTTCC;	ACATCCAGAACCCAGAACCTGCTGTGTACCA	2	ΔG = -8.0	ΔH = -107.	ΔS = -304.5	Tm = 47.9	Image:	Thermod	-107.9	-0.3045	-8.024	232178.0739		GGACAGCAAA
9	3	CAAGGGCTATTTC;	ACATCCAGAACCCAGAACCTGCTGTGTACCA	3	ΔG = -6.7	ΔH = -99.9	ΔS = -284.0	Tm = 43.4	Image:	Thermod	-99.9	-0.284	-6.748	32548.62059		GGACAGCAAA
10	4	CAAGGGCTATTT;	ACATCCAGAACCCAGAACCTGCTGTGTACCA	4	ΔG = -6.3	ΔH = -86.7	ΔS = -244.9	Tm = 40.4	Image:	Thermod	-86.7	-0.2449	-6.3728	18265.3621		GGACAGCAAA
11	5	CAAGGGCTATT;	ACATCCAGAACCCAGAACCTGCTGTGTACCA	5	ΔG = -5.9	ΔH = -77.5	ΔS = -218.2	Tm = 37.1	Image:	Thermod	-77.5	-0.2182	-5.9304	9242.424273		GGACAGCAAA
12	6	CAAGGGCTAT;	ACATCCAGAACCCAGAACCTGCTGTGTACCA	6	ΔG = -5.5	ΔH = -69.6	ΔS = -195.3	Tm = 33.6	Image:	Thermod	-69.6	-0.1953	-5.5416	5079.102615		GGACAGCAAA
13	7	CAAGGGCTA;	ACATCCAGAACCCAGAACCTGCTGTGTACCA	7	ΔG = -5.4	ΔH = -63.0	ΔS = -175.6	Tm = 30.9	Image:	Thermod	-63.0	-0.1756	-5.4032	4104.270097		GGACAGCAAA
14	8	CAAGGGCT;	ACATCCAGAACCCAGAACCTGCTGTGTACCA	8	ΔG = -5.0	ΔH = -62.3	ΔS = -174.6	Tm = 29.0	Image:	Thermod	-62.3	-0.1746	-5.0312	2314.576507		GGACAGCAAA
15	9	CAAGGGC;	ACATCCAGAACCCAGAACCTGCTGTGTACCA	9	ΔG = -4.3	ΔH = -53.5	ΔS = -150.0	Tm = 21.5	Image:	Thermod	-53.5	-0.15	-4.3	750.7621857		GGACAGCAAA
16	10	CAAGGG;	ACATCCAGAACCCAGAACCTGCTGTGTACCA	10	ΔG = -2.8	ΔH = -45.1	ΔS = -129.0	Tm = 7.7°(	Image:	Thermod	-45.1	-0.129	-2.788	73.18052202		GGACAGCAAA
17	11	CAAGG;	ACATCCAGAACCCAGAACCTGCTGTGTACCA	11	ΔG = -1.3	ΔH = -36.4	ΔS = -106.9	Tm = -10.:	Image:	Thermod	-36.4	-0.1069	-1.3368	7.833338294		GGACAGCAAA
18	12	CAAG;	ACATCCAGAACCCAGAACCTGCTGTGTACCA	12	ΔG = -0.4	ΔH = -16.1	ΔS = -47.9	Tm = -70.!	Image:	Thermod	-16.1	-0.0479	-0.3888	1.819696307		GGACAGCAAA
19	13	GCTATTTCCCTGAGC;	ACATCCAGAACCCAGAACCTGCTGTGTACCA	13	ΔG = -9.4	ΔH = -121.	ΔS = -341.4	Tm = 52.3	Image:	Thermod	-121.4	-0.3414	-9.4208	1994763.33		GGACAGCAAA
20	14	CTATTTCCCTGAGC;	ACATCCAGAACCCAGAACCTGCTGTGTACCA	14	ΔG = -7.8	ΔH = -109.	ΔS = -308.9	Tm = 47.6	Image:	Thermod	-109.2	-0.3089	-7.8808	186234.5643		GGACAGCAAA
21	15	TATTTCCCTGAGC;	ACATCCAGAACCCAGAACCTGCTGTGTACCA	15	ΔG = -7.7	ΔH = -99.4	ΔS = -279.4	Tm = 46.5	Image:	Thermod	-99.4	-0.2794	-7.7568	153864.6393		GGACAGCAAA
22	16	ATTTCCCTGAGC;	ACATCCAGAACCCAGAACCTGCTGTGTACCA	16	ΔG = -7.5	ΔH = -92.5	ΔS = -259.1	Tm = 45.0	Image:	Thermod	-92.5	-0.2591	-7.5152	106065.883		GGACAGCAAA
23	17	TTTCCCTGAGC;	ACATCCAGAACCCAGAACCTGCTGTGTACCA	17	ΔG = -6.8	ΔH = -84.6	ΔS = -237.1	Tm = 41.7	Image:	Thermod	-84.6	-0.2371	-6.8312	36997.31571		GGACAGCAAA
24	18	TTCCCTGAGC;	ACATCCAGAACCCAGAACCTGCTGTGTACCA	18	ΔG = -6.4	ΔH = -76.7	ΔS = -214.2	Tm = 38.9	Image:	Thermod	-76.7	-0.2142	-6.4424	20331.58806		GGACAGCAAA
25	19	TCCCTGAGC;	ACATCCAGAACCCAGAACCTGCTGTGTACCA	19	ΔG = -6.3	ΔH = -68.9	ΔS = -190.9	Tm = 36.6	Image:	Thermod	-68.9	-0.1909	-6.2848	15950.73851		GGACAGCAAA
26	20	CCCTGAGC;	ACATCCAGAACCCAGAACCTGCTGTGTACCA	20	ΔG = -6.0	ΔH = -64.9	ΔS = -179.5	Tm = 34.3	Image:	Thermod	-64.9	-0.1795	-6.024	10675.25437		GGACAGCAAA
27	21	CCTGAGC;	ACATCCAGAACCCAGAACCTGCTGTGTACCA	21	ΔG = -4.1	ΔH = -58.7	ΔS = -166.4	Tm = 23.4	Image:	Thermod	-58.7	-0.1664	-4.1208	569.7279158		GGACAGCAAA
28	22	CTGAGC;	ACATCCAGAACCCAGAACCTGCTGTGTACCA	22	ΔG = -2.8	ΔH = -50.7	ΔS = -145.9	Tm = 12.6	Image:	Thermod	-50.7	-0.1459	-2.8448	79.86911708		GGACAGCAAA
29	23	TGAGC;	ACATCCAGAACCCAGAACCTGCTGTGTACCA	23	ΔG = -2.1	ΔH = -36.8	ΔS = -105.8	Tm = -5.3(	Image:	Thermod	-36.8	-0.1058	-2.0976	25.27610723		GGACAGCAAA
30	24	GAGC;	ACATCCAGAACCCAGAACCTGCTGTGTACCA	24	ΔG = -1.6	ΔH = -36.4	ΔS = -106.1	Tm = -8.7(	Image:	Thermod	-36.4	-0.1061	-1.5992	11.73327908		GGACAGCAAA
31																
32																
33		CAPTURE PRIMER	TARGET		ΔG (kcal/mol)	ΔH (kcal/mol)	ΔS (cal/K*mol)	Tm (°C)			H (kcal/mol)	S (kcal/K*mol)	G (kcal/mol) at 55°C	Kcap at 55°C		NON-TARGET
34	1	GCACATTGATTTGGGGAGTCAAA;	ACATCCAGAACCCAGAACCTGCTGTGTACCA	1	ΔG = -13.3	ΔH = -169.	ΔS = -474.5	Tm = 60.8	Image:	Thermody	-169.0	-0.4745	-13.364	864557613.5		GGACAGCAAA

SEQUENCES | CP | LCP THERMO | LCP KEFF | LCP EFF | RCP THERMO | RCP KEFF | RCP EFF

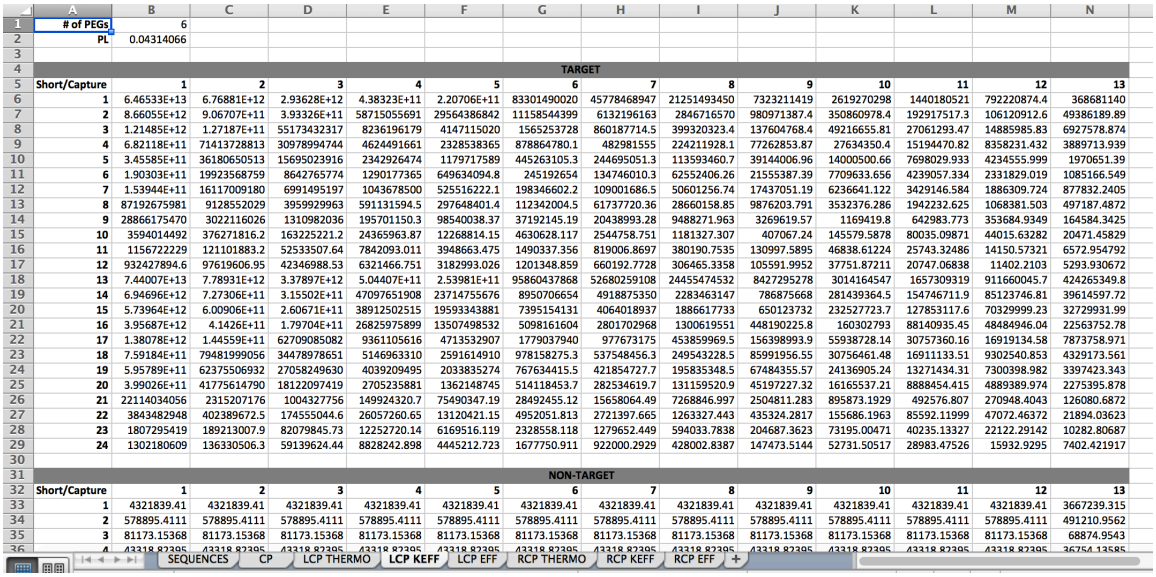
Figure 3.15. Thermodynamic parameters for CP regions in presence of target and non-target sequences

Diagram only shows left CP, but the same sheet is repeated for the right CP but with the corresponding sequences. Top half is the thermodynamic parameters for the short primers and bottom half is the parameters for the capture regions. Left half corresponds to the target sequence and the right half corresponds to the non-target sequence.

The thermodynamic calculations are repeated in four sections of the sheet: (1) short primer + target, (2) capture region + target, (3) short primer + non-target, and (4) capture region + non-target. Only the first section will be used to demonstrate the formulas. Columns D – J are copied and pasted output values from Mfold (Zuker, 2003) after entering either the short or capture region sequences as well as either target or non-target sequence into the program. Cell K7 is "=RIGHT(F7,6)*1", which removes the 6 characters in front of the numerical ΔH value from the F7 cell. Cell L7 is "=RIGHT(G7,7)*1/1000", which removes the 6 characters in from of the numerical ΔS value from the G7 cell and converts the units from cal K^{-1} mol^{-1} to kcal K^{-1} mol^{-1} to match the units outputted for the ΔH value. The cell M7 is "=K7-($C$4*L7)", which calculates the ΔG value by using the ΔH value generated in cell K7, ΔS value generated in L7, and

reaction temp value assigned in cell C4. Finally, cell N7 is "=EXP((-M7)/($C$2*$C$4))", which calculates the equilibrium constant of a binding event by using the ΔG value generated in cell M7, reaction temp value assigned in cell C4, and universal gas constant R value 0.00198 kcal K^{-1} mol^{-1} assigned in cell C2. These formulas are dragged down for all primer sequences available.

The LCP KEFF and RCP KEFF sheets calculate the effective equilibrium constant (K_{eff}) of the various short primer and capture region combinations in the presence of target and non-target sequences for the left CP and right CP, respectively. The first, second, and third iterations use the same calculations **(Figure 3.16)**.



	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	# of PEGs	6												
2	PL	0.04314066												
3														
4							TARGET							
5	Short/Capture	1	2	3	4	5	6	7	8	9	10	11	12	13
6	1	6.46533E+13	6.76881E+12	2.93628E+12	4.38323E+11	2.20706E+11	83301490020	45778468947	21251493450	7323211419	2619270298	1440180521	792220874.4	368681140
7	2	8.66055E+12	9.06707E+11	3.93326E+11	58715055691	29564386842	11158544399	6132196163	2846716570	980971387.4	350860978.4	192917517.3	106120912.6	49386189.89
8	3	1.21485E+12	1.27187E+11	55173432317	8236196179	4147115020	1565253728	860187714.5	399320323.4	137604768.4	49216655.81	27061293.47	14885985.83	6927578.874
9	4	6.82118E+11	71413728813	30978994744	4624491661	2328538365	878864780.1	482981555	224211928.1	77262853.87	27634350.4	15194470.82	8358231.432	3889713.939
10	5	3.45585E+11	36180650513	15695023916	2342926474	1179717589	445263105.3	244695051.3	113593460.7	39144006.96	14000500.66	7698029.933	4234555.999	1970651.39
11	6	1.90303E+11	19923568759	8642765774	1290177365	649634094.8	245192654	134746010.3	62552406.26	21555387.39	7709633.656	4239057.334	2331829.019	1085166.549
12	7	1.53944E+11	16117009180	6991495197	1043678500	525516222.1	198346602.2	109001686.5	50601256.74	17437051.19	6236641.122	3429146.584	1886309.724	877832.2405
13	8	87192675981	9128552029	3959929963	591131594.5	297648401.4	112342004.5	61737720.36	28660158.85	9876203.791	3532376.286	1942232.625	1068381.503	497187.4872
14	9	28866175470	3022116026	1310982036	195701150.3	98540038.37	37192145.19	20438993.28	9488271.963	3269619.57	1169419.8	642983.773	353684.9349	164584.3425
15	10	3594014492	376271816.2	163225221.2	24365963.87	12268814.15	4630628.117	2544758.751	1181327.307	407067.24	145579.5878	80035.09871	44015.63282	20471.45829
16	11	1156722229	121101883.2	52533507.64	7842093.011	3948663.475	1490337.356	819006.8697	380190.7535	130997.5895	46838.61224	25743.32486	14150.57321	6572.954792
17	12	932427894.6	97619606.95	42346988.53	6321466.751	3182993.026	1201348.859	660192.7728	306465.3358	105591.9952	37751.87211	20747.06838	11402.2103	5293.930672
18	13	7.44007E+13	7.78931E+12	3.37897E+12	5.04407E+11	2.53981E+11	95860437868	52680259108	24455474532	8427295278	3014164547	1657309319	911660045.7	424265349.8
19	14	6.94696E+12	7.27306E+11	3.15502E+11	47097651908	23714755676	8950706654	4918875350	2283463147	786875668	281439364.5	154746711.9	85123746.81	39614597.72
20	15	5.73964E+12	6.00906E+11	2.60671E+11	38912502515	19593343881	7395154131	4064018937	1886617733	650123732	232527723.7	127853117.6	70329999.23	32729931.99
21	16	3.95687E+12	4.1426E+11	1.79704E+11	26825975899	13507498532	5098161604	2801702968	1300619551	448190225.8	160302793	88140935.45	48484946.04	22563752.78
22	17	1.38078E+12	1.44559E+11	62709085082	9361105616	4713532907	1779037940	977673175	453859969.5	156398993.9	55938728.14	30757360.16	16919134.58	7873758.971
23	18	7.59184E+11	79481999056	34478978651	5146963310	2591614910	978158275.3	537548456.3	249543228.5	85991956.55	30756461.48	16911133.51	9302540.853	4329173.561
24	19	5.95789E+11	62375506932	27058249630	4039209495	2033835274	767634415.5	421854727.7	195835348.5	67484355.57	24136905.24	13271434.31	7300398.982	3397423.343
25	20	3.99026E+11	41775614790	18122097419	2705235881	1362148745	514118453.7	282534619.7	131159520.9	45197227.32	16165537.21	8888454.415	4889389.974	2275395.878
26	21	22114034056	2315207176	1004327756	149924320.7	75490347.19	28492455.12	15658064.49	7268846.997	2504811.283	895873.1929	492576.807	270948.4043	126080.6872
27	22	3843482948	402389672.5	174555044.6	26057260.65	13120421.15	4952051.813	2721397.665	1263327.443	435324.2817	155686.1963	85592.11999	47072.46372	21894.03623
28	23	1807295419	189213007.9	82079845.73	12252720.14	6169516.119	2328558.118	1279652.449	594033.7838	204687.3623	73195.00471	40235.13327	22122.29142	10282.80687
29	24	1302180609	136330506.3	59139624.44	8828242.898	4445212.723	1677750.911	922000.2929	428002.8387	147473.5144	52731.50517	28983.47526	15932.9295	7402.421917
30														
31							NON-TARGET							
32	Short/Capture	1	2	3	4	5	6	7	8	9	10	11	12	13
33	1	4321839.41	4321839.41	4321839.41	4321839.41	4321839.41	4321839.41	4321839.41	4321839.41	4321839.41	4321839.41	4321839.41	4321839.41	3667239.315
34	2	578895.4111	578895.4111	578895.4111	578895.4111	578895.4111	578895.4111	578895.4111	578895.4111	578895.4111	578895.4111	578895.4111	578895.4111	491210.9562
35	3	81173.15368	81173.15368	81173.15368	81173.15368	81173.15368	81173.15368	81173.15368	81173.15368	81173.15368	81173.15368	81173.15368	81173.15368	68874.9543
36	4	43318.82395	43318.82395	43318.82395	43318.82395	43318.82395	43318.82395	43318.82395	43318.82395	43318.82395	43318.82395	43318.82395	43318.82395	36754.13585

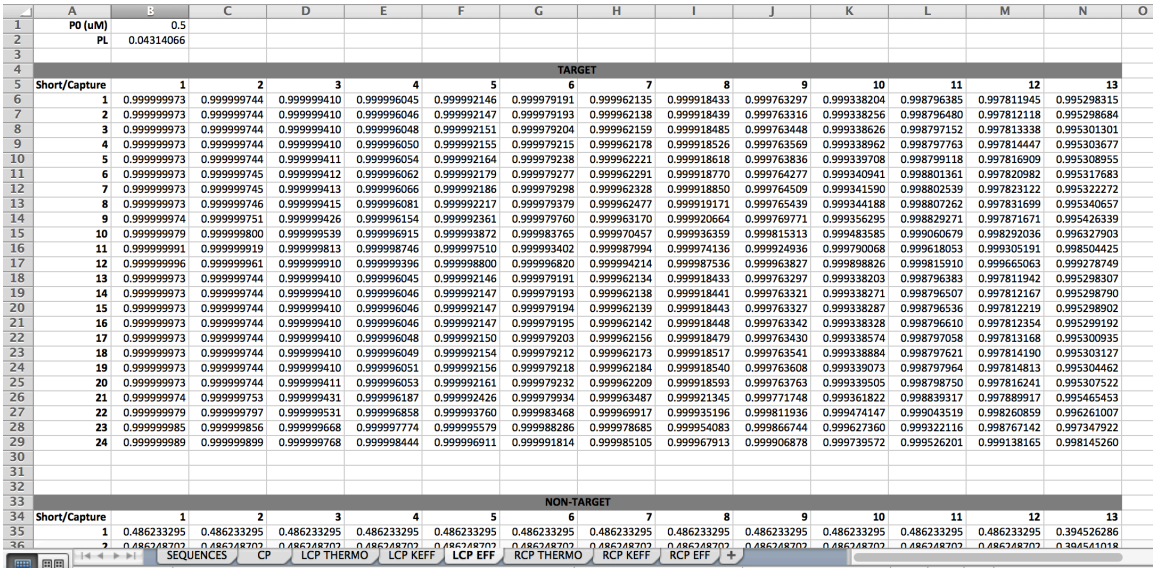
SEQUENCES | CP | LCP THERMO | LCP KEFF | LCP EFF | RCP THERMO | RCP KEFF | RCP EFF

Figure 3.16. Effective equilibrium constant for all possible combinations of short primers and capture regions in presence of target and non-target sequences
Diagram only shows left CP, but the same sheet is repeated for the right CP but with the corresponding sequences. Top half is the effective equilibrium constant for the target sequence and the bottom half is for the non-target sequence.

The K_{eff} calculations are repeated in two sections of the sheet: (1) short primer and capture region combination + target and (2) short primer and capture region combination + non-target. Only the first section will be used to demonstrate the formulas. Cell B2 is "=1/(4/3*PI()*(((1.54*SIN(52.5*(PI()/180))*(B1)+1.43*COS(37.5*(PI()/180))*(2*B1))*0.1)^3*1E-24*6.023E+23))", which calculates the increased concentration created by the first binding event. The equation uses trigonometry of the C-C and C-O Å bond length of an ethylene glycol to calculate the x-distance of the total polyethylene glycol linker. The total number of ethylene glycols in the polyethylene glycol is defined in cell B1. This number of ethylene glycols is dependent on the number of bps between the end of the short primer and beginning of the capture region; one bp is approximately one ethylene glycol. This x-distance is then converted from Å to nm and it is cubed, as it is the radius of the volumetric space. This value is then multiplied by a unit conversion to convert nm to L and then multiplied by Avogadro's number to convert molecules to mol so the final value has M units. Cell B6 is "='LCP THERMO'!$N7+'LCP THERMO'!$N$34+('LCP THERMO'!$N7*'LCP THERMO'!$N$34*'LCP KEFF'!$B$2)", which calculates the K_{eff} by adding the equilibrium constant of the short primer binding (from the LCP/RCP THERMO sheet) with the equilibrium constant of the capture region binding (from the LCP/RCP THERMO sheet) with the equilibrium constant of both primers binding (modeled by multiplying the equilibrium constant of short primer binding (LCP/RCP THERMO sheet) with the equilibrium constant of capture region binding (LCP/RCP THERMO sheet) with the increased local concentration created by the first binding event

(cell B2)). This formula is manipulated for all capture sequences available and then dragged down for all short primer sequences available.

The LCP EFF and RCP EFF sheets calculate the E_{ff} of the various short primer and capture region combinations in the presence of target and non-target sequences for the left CP and right CP, respectively. The first, second, and third iterations use the same calculations **(Figure 3.17)**.



	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	P0 (uM)	0.5												
2	PL	0.04314066												
3														
4							TARGET							
5	Short/Capture	1	2	3	4	5	6	7	8	9	10	11	12	13
6	1	0.999999973	0.999999744	0.999999410	0.999996045	0.999992146	0.999979191	0.999962135	0.999918433	0.999763297	0.999338204	0.998796385	0.997811945	0.995298315
7	2	0.999999973	0.999999744	0.999999410	0.999996046	0.999992147	0.999979193	0.999962138	0.999918439	0.999763316	0.999338256	0.998796480	0.997812118	0.995298684
8	3	0.999999973	0.999999744	0.999999410	0.999996048	0.999992151	0.999979204	0.999962159	0.999918485	0.999763448	0.999338626	0.998797152	0.997813338	0.995301301
9	4	0.999999973	0.999999744	0.999999410	0.999996050	0.999992155	0.999979215	0.999962178	0.999918526	0.999763569	0.999338962	0.998797763	0.997814447	0.995303677
10	5	0.999999973	0.999999744	0.999999411	0.999996054	0.999992164	0.999979238	0.999962221	0.999918618	0.999763836	0.999339708	0.998799118	0.997816909	0.995308955
11	6	0.999999973	0.999999745	0.999999412	0.999996062	0.999992179	0.999979277	0.999962291	0.999918770	0.999764277	0.999340941	0.998801361	0.997820982	0.995317683
12	7	0.999999973	0.999999745	0.999999413	0.999996066	0.999992186	0.999979298	0.999962328	0.999918850	0.999764509	0.999341590	0.998802539	0.997823122	0.995322272
13	8	0.999999973	0.999999746	0.999999415	0.999996081	0.999992217	0.999979379	0.999962477	0.999919171	0.999765439	0.999344188	0.998807262	0.997831699	0.995340657
14	9	0.999999974	0.999999751	0.999999426	0.999996154	0.999992361	0.999979760	0.999963170	0.999920664	0.999769771	0.999356295	0.998829271	0.997871671	0.995426339
15	10	0.999999979	0.999999800	0.999999539	0.999996915	0.999993872	0.999983765	0.999970457	0.999936359	0.999815313	0.999483585	0.999060679	0.998292036	0.996327903
16	11	0.999999991	0.999999919	0.999999813	0.999998746	0.999997510	0.999993402	0.999987994	0.999974136	0.999924936	0.999790068	0.999618053	0.999305191	0.998504425
17	12	0.999999996	0.999999961	0.999999910	0.999999396	0.999998800	0.999996820	0.999994214	0.999987536	0.999963827	0.999898826	0.999815910	0.999665063	0.999278749
18	13	0.999999973	0.999999744	0.999999410	0.999996045	0.999992146	0.999979191	0.999962134	0.999918433	0.999763297	0.999338203	0.998796383	0.997811942	0.995298307
19	14	0.999999973	0.999999744	0.999999410	0.999996046	0.999992147	0.999979193	0.999962138	0.999918441	0.999763321	0.999338271	0.998796507	0.997812167	0.995298790
20	15	0.999999973	0.999999744	0.999999410	0.999996046	0.999992147	0.999979194	0.999962139	0.999918443	0.999763327	0.999338287	0.998796536	0.997812219	0.995298902
21	16	0.999999973	0.999999744	0.999999410	0.999996046	0.999992147	0.999979195	0.999962142	0.999918448	0.999763342	0.999338328	0.998796610	0.997812354	0.995299192
22	17	0.999999973	0.999999744	0.999999410	0.999996048	0.999992150	0.999979203	0.999962156	0.999918479	0.999763430	0.999338574	0.998797058	0.997813168	0.995300935
23	18	0.999999973	0.999999744	0.999999410	0.999996049	0.999992154	0.999979212	0.999962173	0.999918517	0.999763541	0.999338884	0.998797621	0.997814190	0.995303127
24	19	0.999999973	0.999999744	0.999999410	0.999996051	0.999992156	0.999979218	0.999962184	0.999918540	0.999763608	0.999339073	0.998797964	0.997814813	0.995304462
25	20	0.999999973	0.999999744	0.999999411	0.999996053	0.999992161	0.999979232	0.999962209	0.999918593	0.999763763	0.999339505	0.998798750	0.997816241	0.995307522
26	21	0.999999974	0.999999753	0.999999431	0.999996187	0.999992426	0.999979934	0.999963487	0.999921345	0.999771748	0.999361822	0.998839317	0.997889917	0.995465453
27	22	0.999999979	0.999999797	0.999999531	0.999996858	0.999993760	0.999983468	0.999969917	0.999935196	0.999811936	0.999474147	0.999043519	0.998260859	0.996261007
28	23	0.999999985	0.999999856	0.999999668	0.999997774	0.999995579	0.999988286	0.999978685	0.999954083	0.999866744	0.999627360	0.999322116	0.998767142	0.997347922
29	24	0.999999989	0.999999899	0.999999768	0.999998444	0.999996911	0.999991814	0.999985105	0.999967913	0.999906878	0.999739572	0.999526201	0.999138165	0.998145260
30														
31														
32														
33							NON-TARGET							
34	Short/Capture	1	2	3	4	5	6	7	8	9	10	11	12	13
35	1	0.486233295	0.486233295	0.486233295	0.486233295	0.486233295	0.486233295	0.486233295	0.486233295	0.486233295	0.486233295	0.486233295	0.486233295	0.394526286
36	2	0.486248702	0.486248702	0.486248702	0.486248702	0.486248702	0.486248702	0.486248702	0.486248702	0.486248702	0.486248702	0.486248702	0.486248702	0.394541018

SEQUENCES | CP | LCP THERMO | LCP KEFF | LCP EFF | RCP THERMO | RCP KEFF | RCP EFF

Figure 3.17. Primer efficiency for all possible combinations of short primers and capture regions in presence of target and non-target sequences
Diagram only shows left CP, but the same sheet is repeated for the right CP but with the corresponding sequences. Top half is the effective equilibrium constant for the target sequence and the bottom half is for the non-target sequence.

The E_{ff} calculations are repeated in two sections of the sheet: (1) short primer and capture region combination + target and (2) short primer and capture region combination + non-target. Only the first section will be used to demonstrate the formulas. Cell B2

references the value of the increased local concentration calculated from the LCP/RCP KEFF sheet. Cell B6 is "=(('LCP THERMO'!$N$34+('LCP THERMO'!$N7*'LCP THERMO'!$N$34*'LCP EFF'!$B$2))*'LCP EFF'!$B$1)/(1+('LCP KEFF'!B6*'LCP EFF'!$B$1))", which calculates the E_{ff} using equation Y. The equilibrium constant of the capture region binding (from the LCP/RCP THERMO sheet) is added with the equilibrium constant of both primers binding (modeled by multiplying the equilibrium constant of short primer binding (LCP/RCP THERMO sheet) with the equilibrium constant of capture region binding (LCP/RCP THERMO sheet) with the increased local concentration created by the first binding event (cell B2)). This value is then divided by the corresponding K_{eff} (LCR/RCP KEFF sheet), which is multiplied by the initial primer concentration defined in cell B1. This formula is dragged down for all short primer sequences available and across for all capture region sequences available.

Optimal primers were determined by having an E_{ff} >95% in presence of the target sequence and <1% in presence of the non-target sequence. After finding optimal primer combinations within the desired E_{ff} ranges, the primers bp length and predicted T_m's were evaluated. Requirements for predicted T_m conditions of the primers and bp length were based on Satterfield's findings (Satterfield, 2014). An optimal combination was determined to be a short primer with 10 to 14 bps and a T_m slightly below the T_m of the capture region, which is 3°C to 4°C below the reaction temperature. After determining the optimal left CP and right CP from the second iteration and validating the combination in the third iteration, a custom order was fulfilled with BioSearch (Petaluma, CA). The order used a medium synthesis scale and Dual HPLC purification; nomenclature for

specific sequences ordered is shown in results **(Table 3.1)**. The "id[C/G/T/A]" nomenclature is used to represent an inverted based, so instead of being 5' → 3' it is 3' → 5'. To order the specific CPs, a Custom Oligos order form has to be obtained from BioSearch. The CP sequences, given in **Table 3.1**, can then be copied and pasted into the Oligo Info sheet under the sequence column. The synthesis scale and purification can then be selected for each sequence.

Evaluating Cooperative Primers. To determine the capture of the α and β mRNA, two sample types: (1) origami template (OT) and (2) positive template (PT) were prepared for RT. Using a sample size of 20 μL, a 0.2 mL PCR tube was prepared with the Omniscript Reverse-Transcription kit (Qiagen) and RNA (from TCRα and TCRβ cells extracted from a transgenic mouse spleen). For the OT sample, the origami strand (acting as the RT primers) was added and diluted with water to make a final concentration of 0.1 μM. For the PT sample, Cα and Cβ primers were added and diluted with water to make a final concentration of 0.1 μM. These samples were incubated at 37°C for 1 hr to perform RT. To determine the optimal annealing conditions of the cooperative primers, a temperature annealing gradient using PCR was then conducted on the products generated from RT.

The PCR samples included (1) Cα + Vα – PT, (2) Cβ + Vβ – PT, (3) left CP + Vα – OT, (4) right CP + Vβ – OT, (5) left CP + Vα + right CP + Vβ – OT, (6) X + Vα – OT, and (7) Y + Vβ – OT. The X and Y primers are the full 20 bp complements of the X' and Y' sequences of the origami strand, respectively. Samples 1 and 2 are positive controls to ensure there are no problems with either the RNA used for the RT or the RT/PCR

components. Samples 6 and 7 are positive controls to ensure there are no problems with the RT of the origami strand. To prepare the 50 μL PCR samples, 3 μL of the respective RT sample was added to a 0.2 mL PCR tube. Primers were added with 2X hot start DreamTaq PCR Master Mix (ThermoFisher) and diluted with water make a final concentration of 0.2 uM and 1X, respectively.

An annealing temperature gradient was performed for four temperatures: 45.2°, 48.5°, 51.4°, and 53.5°C. For the PCR protocol, it was necessary to create two annealing temperature steps, as the CPs are designed to bind to the X' and Y' products created by the Vα and Vβ primers – not the origami strand. Thus, if the products generated by the Vα and Vβ primers were never generated, then there would be no target template for the CPs. Previous data for Vα and Vβ primers (not shown) demonstrated these primers were optimized at 53°C. The PCR protocol was set to have an initial denaturation at 95°C for 1-3 min, followed by a repeatable process of denaturation at 95°C for 30 s, annealing at 53°C for 30 s (for the Vα and Vβ primers), extension at 72°C for 30 s, denaturation at 95°C for 30 s, annealing temperature gradient (45.2°, 48.5°, 51.4°, and 53.5°C) for 2 min (for the CPs), and extension at 72°C for 30 s.

Following PCR, 10μL of each sample was prepared with 2 μL of 6X Gel Loading Dye. The 12 μL samples were then run with a 1 Kb plus DNA ladder at 110V for 1 hr on a 2% agarose gel prepared with 1X TAE buffer.

Results and Discussion

For the first iteration, Schoettle and Blattman provided the X 'and Y' sequences from the bowtie-barcoded origami strand, which was based on their pre-constructed strand. Because these sequences were pre-defined, the only way to generate a variety of short primer sequences was to change the bp length and location of binding. Results produced 312 possible combinations of short primers and capture regions. The E_{ff} of the left CP indicated that while generally greater than 95% in the presence of the target, the minimum E_{ff} in the presence of non-target was 39.45%. For the right CP, the E_{ff} was generally greater than 95% in the presence of the target; however, the minimum E_{ff} in the presence of non-target was 17.1%.

For these results, the primers were evaluated assuming a PEG linker distance of 6 ethylene glycols, equivalent to about 6 bps. This linker length was used as each short primer and capture region had a varying bp length in its design, and when used in the combination table, this would have produced a different linker length distance for each combination. Thus, to simplify this concern, the linker length distance typically used for TPs, 6 ethylene glycols, was utilized for the primer efficiency analysis. As indicated by the results, none of the E_{ff}'s in the presence of non-target were below 1%, and since there were additional design constraints as based on Satterfield's findings (Satterfield, 2014), it was concluded to develop new primers based on backwards design. The first iteration of CPs was never ordered; all assessment was conducted in silico.

For the second iteration, six primers were designed to be 12 bps long, with varying %GC content, and with T_m's ranging from ~40-52°C. The same capture regions

were used from the first iteration. For these results, the primers were evaluated assuming a PEG linker distance of 24 ethylene glycols, equivalent to about 24 bps. This linker length was utilized, as this was the shortest distance that would occur between the short primer and capture region. The distance included the 12 bps of the random BC as well as the 12 bps of the Y/X segments of the Cα/Cβ products. After determining the optimal left and right short primers, the corresponding targets were implemented in the bowtie-barcoded origami strand. This backwards design didn't create as many limitations as the first iteration and was only possible as the bowtie-barcoded origami strand could be re-constructed with any determined X' and Y' segments.

Using six variations of short primers and 13 variations of capture regions generated 78 combinations. Though there were fewer combinations, better primer efficiencies were obtained. The E_{ff} of the left CP ranged from ~76.8% to 99.9% in the presence of the target and ~0.4% to 0.6% in the presence of non-target. For the right CP, the E_{ff} ranged from ~65.0% to 99.9% in the presence of the target and ~0.3% - 7% in the presence of non-target. For the left CP, the best E_{ff}'s were found to be for the 4 short primer – 6 capture region sequence, which had a 98.8% binding in the presence of the target, and a 0.05% binding in the presence of the non-target. For the right CP, the best E_{ff}'s were found to be for the 3 short primer – 7 capture region sequence, which had a 98.4% binding in the presence of the target, and a 0.08% binding in the presence of the non-target.

These combinations for the left and right CPs were determined to be the best, as they fit the necessary E_{ff} conditions and had capture region sequences with a T_m ~51°C

and short primer sequences with a T_m ~48°C for the target sequences. Having a T_m below the reaction temperature ensure specifics amplification, however, a trade-off with this is the possibility of no primer binding occurring. In the case of these capture regions, the T_m was below the selected reaction temperature but still indicated formation of product. Furthermore, these short primers were chosen, as they had a lower T_m than the capture regions. This was necessary as binding of the capture region holds the short primer in close proximity to the DNA template, thus increasing the local concentration. This effect shifts the T_m of the short primer above the reaction temperature – allowing binding to occur (Satterfield, 2014).

In the third iteration, the optimal short primer and capture region sequences for the left and right CPs of the second iteration were utilized. However, Schoettle and Blattman required the X' and Y' target sequences to actually be 20 bp in length, rather than the 12 bps determined from the second iteration. The extension of the X' and Y' target sequences was deemed as a necessary requirement to create a stronger binding affinity between X' and X and Y' and Y segments so the bowtie-barcoded component of the origami structure could be formed **(Figure 3.18)**.

Figure 3.18. Assembly of bowtie-barcoded origami strand
To assemble the bowtie-barcoded origami strand, there needs to be a strong binding affinity between X' and X and Y' and Y. This binding followed by PCR creates the arms of the origami strand to capture the α and β mRNA strands within the cell.

In the calculations, an additional 8 random bps were added to both the X' and Y' target sequences. Furthermore, now knowing the particular capture region, the actual length of the linker could be utilized in the E_{ff} calculations. The linker was updated from 24 ethylene glycols to 51, equivalent to about 51 bps. This linker length was utilized, as it incorporated the 12 bp distance of the random BC, 20 bp distance of the Y/X segments of the Cα/Cβ products, 5/6 bp distance of the 5' → 3' removal for the Cα/Cβ capture sequences, and 12 bp distance of the X'/Y' segments from the CP due to the need for an inverted linkage. This totals to a maximum of 50 bp distance and since each Spacer 9 for ordering incorporates 3 ethylene glycols, the distance was rounded to 51 for both the left and right CPs. The left CP had a 98.8% binding in the presence of the target, and a 0.01% binding in the presence of the non-target. The right CP had a 98.3% binding in the presence of the target, and a 0.08% binding in the presence of the non-target. These sequences were then ordered from BioSearch (Petaluma, CA) **(Table 3.1)**.

Table 3.1. CP sequences ordered from BioSearch
The [iX] signifies an inverted base pair.

Left CP
3'-short primer-5'-LINKER-5'-capture region-3'
[idC][idA][iT][idC][idG][idC][idA][idG][idG][idC][idA][iT][Spacer 9]TTGATTTGGGAGTCAAA[Spacer C3]
Right CP
3'-short primer-5'-LINKER-5'-capture region-3'
[iT][idA][iT][idG][idG][idC][idC][idC][idC][idA][idC][idA][Spacer 9]CCTTTTGTTTGTTTGC[Spacer C3]

Results of the annealing temperature gradient **(Figure 3.19)** indicate the left and right CPs are working efficiently at 51.4°C and 53.5°C annealing temperatures (**Figure 3.19.C and 3.19.D**, lanes 4 and 5, respectively), as these dark bands closely correspond to the X and Y controls (**Figure 3.19.C and 3.19.D**, lanes 7 and 8, respectively). Even though there is a larger by-product also produced by the right CP, both Schoettle and Blattman have indicated this is not a concern for identifying the β gene segment. When observing the product of the combined left and right CPs, instead of seeing two separate bands, around 500 and 400 bps, respectively aligning to the left and right CP controls,

there is one large band around 1000 bps (**Figure 3.19.C and 3.19.D**, lane 6). Because this band size is essentially the length of the two separate bands, it was believed this product is formed due to hybridization of the respective PCR products **(Figure 3.20)**.

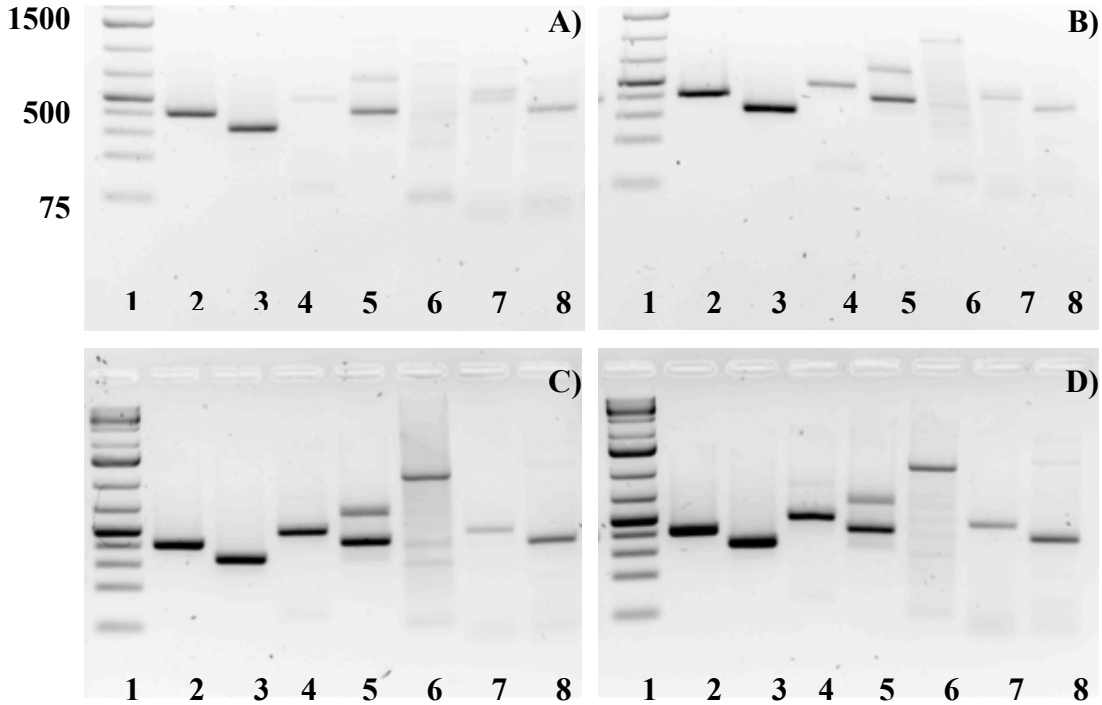


Figure 3.19. PCR annealing temperature gradient using Vα, left CP, Vβ, and right CP primers
Lane 1 is 100 bp DNA ladder. Lane 2 is the result of the amplified products from Cα + Vα primers. Lane 3 is the result of the amplified products from Cβ + Vβ primers. Lane 4 is the result of the amplified products from left CP + Vα primers. Lane 5 is the result of the amplified products from right CP + Vβ primers. Lane 6 is the result of the amplified products from left CP + Vα + right CP + Vβ primers. Lane 7 is result of the amplified products from X + Vα primers. Lane 8 is result of the amplified products from Y + Vβ primers. **[A]** Uses a 45.2°C annealing temperature for the CPs. **[B]** Uses a 48.5°C annealing temperature for the CPs. **[C]** Uses a 51.4°C annealing temperature for the CPs. **[D]** Uses a 53.5°C annealing temperature for the CPs.

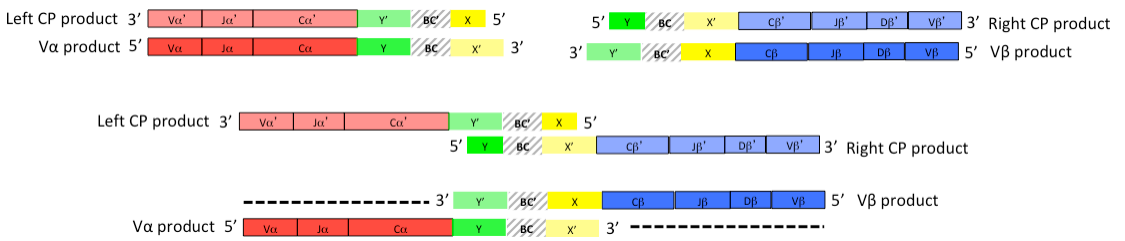


Figure 3.20. PCR product hybridization using Vα, left CP, Vβ, and right CP primers
The left CP product has the ability to bind to the right CP product, however, these strands do not elongate, as Taq polymerase can only synthesize DNA in the 5' → 3' direction. The Vα product has the ability to bind to the Vβ product, and these strands have the potential to elongate, leaving one BC associated with both an α and β mRNA. Because the BCs can be complete mismatches and still hybridize it cannot be guaranteed the one BC is for both the α and β gene segments tied to it. So it will be unknown which gene segment originally identified with the BC and preventing pairing of α and β mRNAs for TCR specificity.

Conclusion

From the results, it could be seen that the Vα + left CP primers specifically amplify the left product of interest, which includes the α gene segment paired with a BC. It could also be seen that the Vβ + right CP primers specifically amplify the right product of interest, which includes the β gene segment paired with a BC. However, when Vα + left CP + Vβ + right CP primers are all put together with the same cDNA, the resulting product is one that appears to be caused hybridization and elongation of the Vα and Vβ products. The hybridization of the Vα to the Vβ product then causes the left CP product and right CP product to be single strands as their complements are used in the hybridization. For the gel electrophoresis, SYBR safe was used as the DNA stain, so it is possible that the single strand products created by the left CP and right CP products are there but not visible, as SYBR safe works best for double stranded products due to its

intercalating properties. Using a different stain, such as SYBR gold, which is specific to single stranded products could further validate this hypothesis.

In reviewing the design of the CPs, it was noticed that to create the varying capture regions, a bp was removed from 5' → 3', which actually creates a longer linker. It would have been better to remove the bp from 3' → 5'. However, when comparing the linker length of 51 bp used (from the 5' → 3' removal) versus the potential 44 bp (from the 3' → 5' removal), it wouldn't make a huge difference in the calculation of the volumetric space created by the linker: 7.02E-05 compared to 1.09E-04. It would however make a difference in the cost of the custom order. A Spacer 9 modification costs $30.00/each, so instead of 17 Spacer 9 modifications, equating to $510, it would be 15 Spacer 9 modifications, equating to $450.

Both the products created by the left CP and right CP were entered in Mfold (Zuker, 2003). The ΔG produced by the hybridization was -19.7 kcal mol^{-1} with a T_m of 71.9°C. Even with an additional 12 random bps added between the Y' and X of the left CP product and Y and X' of the right CP product, Mfold still produced a ΔG of -12.7 kcal mol^{-1} with a T_m of 59.0°C. Because of the range of possibilities for the number of BCs, it's possible not all the 12 bps between the two BCs will mismatch, and if some of them do match, it will further decrease the ΔG and encourage hybridization. For these reasons, the hybridization of the two products can become favored over the hybridization of the Vα or Vβ primers. These products have the potential to hybridize but will not elongate since Taq only reads in a 5' → 3' directionality. So it is possible when the strands are denatured for sequencing, there will still be only one BC with one gene segment.

However, this sample is most likely mixed in with the Vα and Vβ hybridized products when displayed through gel electrophoresis.

Furthermore, the products created by the Vα and Vβ products were also entered in Mfold (Zuker, 2003). The ΔG produced by the hybridization was -36.2 kcal mol^{-1} with a T_m of 82.0°C. This ΔG is even more negative than the left CP and right CP products as there are 40 matching bps compared to 24. Even with an additional 12 random bps added between the Y and X' of the Vα product and Y' and X of the Vβ product, Mfold still produced a ΔG of -29.6 kcal mol^{-1} with a T_m of 74.1°C – despite the 12 bp mismatch between the two product strands. Again, this ΔG can become even more negative if there are some matches between the BCs. When comparing the lowest ΔG of hybridization to the ΔG of the CP capture regions (-9.0 kcal mol^{-1} for left CP and -8.8 kcal mol^{-1} for right CP), it can be seen that the product hybridization would be favored over the primers.

Even if the X' and Y' segments were altered for the Cα and Cβ products to include a more AT-rich region, thus reducing the %GC content and decreasing the T_m to prevent binding at the higher reaction temperature, the hybridized product would still most likely occur. This is because if the X' and Y' regions are altered to be more AT-rich, then the corresponding short primers of the CPs would also have to be altered to account for the new AT-rich region, thus reducing the T_m of the short primers. As of now, there has not been enough research on CPs to know the number of degrees the short primer T_m can be below the T_m of the capture region and still function. However, Satterfield did demonstrate that as the linker length increased between the short primer and capture region, the closer the T_m of the short primer would need to be to the T_m of the

capture region (Satterfield, 2014). This is because a longer linker length creates a larger spherical space the short primer can travel. The T_m of the short primer is associated with the probability of the short primer colliding with the target sequence. As the number of collisions increases, T_m of the short primer shifts above the reaction temperature and binds. With a long linker length, the frequency of collisions is less so the T_m of the short primer needs to be closer to the T_m of the capture region.

It's also possible that the Vα or Vβ primers could anneal to one of the hybridized products and elongate. This has the potential of associating the wrong BC with the wrong gene segment. If the 3' end of the template could be blocked after elongation of the Vα and Vβ primer, then this would prevent elongation of the product-product hybridization and ensure the BC is associated with the correct α or β gene segment. As described earlier, the hybridized result seems to be inevitable with every design choice. However, the initial strand used to make the origami could potentially be altered to introduce two different BCs in the final hybridized product; one BC next to the α gene segment and one next to the β **(Figure 3.21)**. This relocation of the BC would even further encourage product-product hybridization, as it would prevent the possible 12 bp mismatch and thus produce the lowest possible ΔG's. However, since the Y', Y, X', and X sequences are known and since the location of the BC is known to be next to the Cα, Cα', Cβ, or Cβ' sequences, then during sequencing, each gene segment can be associated with its correct BC despite product hybridization. Furthermore, relocation of the BC would not require new CPs to be designed, as it does not impact the short primer, capture region, or linker length. This idea was proposed to Schoettle, however, this would be unfeasible, as IDT

DNA cannot guarantee that the first randomly generated 12 bp sequences for the BC are exactly replicated later on in the sequence.

Future research suggests proceeding with the current bowtie-barcoded origami strand and current primers but finding a way to verify if the left CP and right CP products are hybridized but able to form the desired products upon denaturation. Future work should run the samples using a denaturing gradient gel electrophoresis to separate the hybridized strands. Theoretically, this should produce three bands with the topmost band being the Vα and Vβ hybridized products, which can be disregarded.

If the hybridization of the system were to be resolved or if the BCs of the hybridized product are identified to its α or β gene segment, then this system has the potential to develop novel immunotherapies, such as cancer therapies. Tumor infiltrating lymphocytes (TILs) are tumor-antigen-specific T cells that can be isolated following removal of the tumor tissue (Sharpe & Mount, 2015). Tumor-specific T cells can then be engineered as based on the identification of the α and β chains of the TIL receptors. Previous clinical research on TCR therapy has demonstrated overall tumor regression and showed no signs of significant toxicity, thus providing the possibility of a promising cancer therapy (Sharpe & Mount, 2015). However, mispairing of the α and β chains of the receptor can impact both efficacy and safety so it is necessary to ensure the bowtie-barcoded origami system is specific and accurate and relays the correct α and β chain to the correct BC.

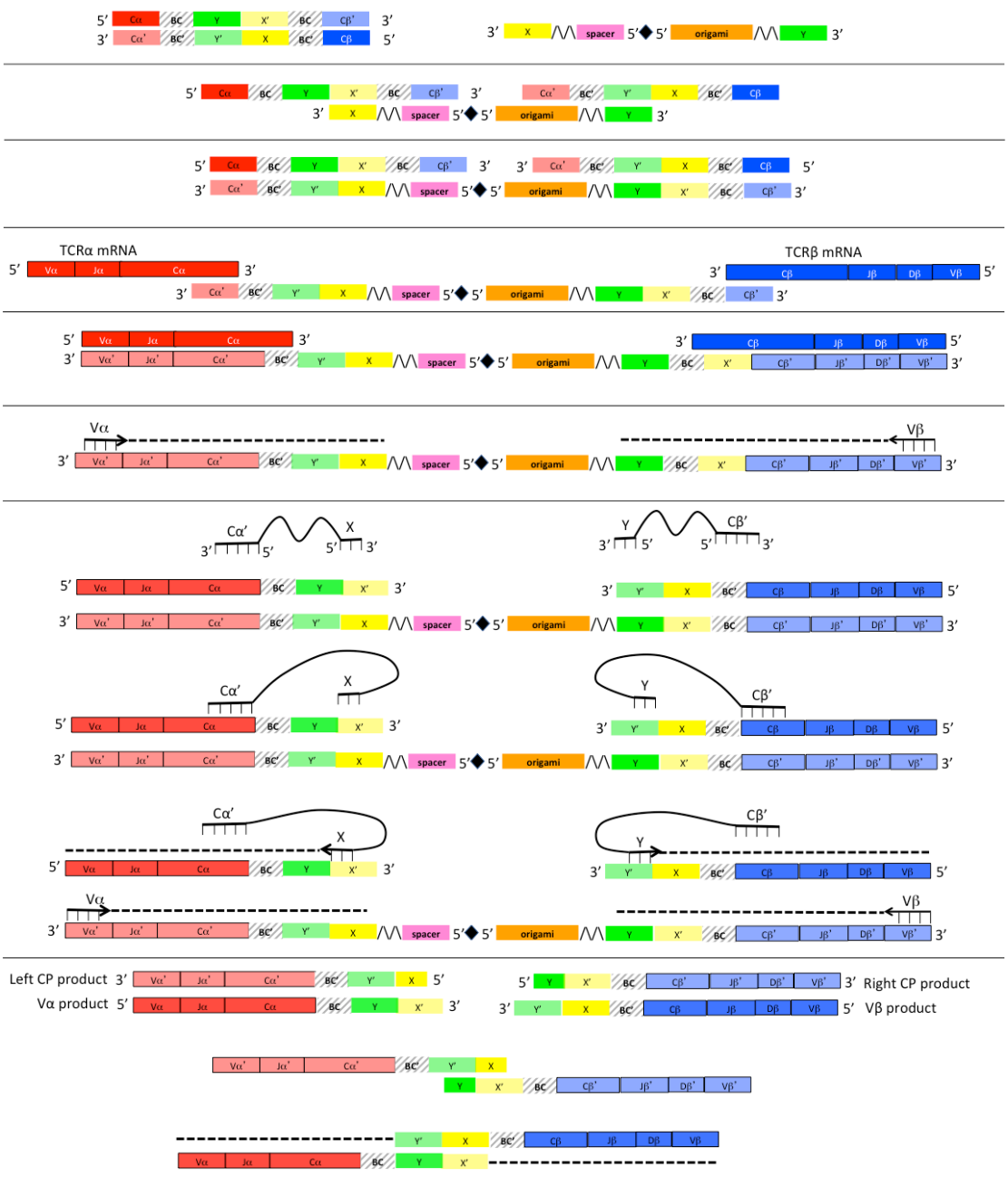


Figure 3.21. Proposed two barcode origami strand
This system would create two BCs in the hybridized product and ensure each α and β gene segment was associated to its own BC, however, IDT DNA cannot generate 12 random bp for the first barcode and ensure the same 12 bps are repeated later on in the sequence for the second barcode.

REFERENCES

Attaf, M., Huseby, E., & Sewell, A. K. (2015). ab T cell receptors as predictors of health and disease. *Cellular & Molecular Immunology*, *12*, 391–399. https://doi.org/10.1038/cmi.2014.134

Balboni, G., Tassé, M. J., Schalock, R. L., Borthwick-Duffy, S. A., Spreat, S., Thissen, D., … Navas, P. (2014). The Diagnostic Adaptive Behavior Scale: Evaluating its diagnostic sensitivity and specificity. *Research in Developmental Disabilities*, *35*(11), 2884–2893. https://doi.org/10.1016/j.ridd.2014.07.032

Barrick Lab. (2017). Standard Polymerase Chain Reaction (PCR). Retrieved March 24, 2018, from http://barricklab.org/twiki/bin/view/Lab/ProtocolsStandardPCR

Brodersen, J., & Siersma, V. D. (2013). Long-Term Psychosocial Consequences of False-Positive Screening Mammography. *The Annals of Family Medicine*, *11*(2), 106–115. https://doi.org/10.1370/afm.1466

Bureau of Labor Statistics. (2018). Medical and Clinical Laboratory Technologists and Technicians : Occupational Outlook Handbook: : U.S. Bureau of Labor Statistics. Retrieved March 24, 2018, from https://www.bls.gov/ooh/healthcare/medical-and-clinical-laboratory-technologists-and-technicians.htm

Cera, E. Di. (1998). Site-Specific Thermodynamics: Understanding Cooperativity in Molecular Recognition. *Chemical Reviews*, *98*, 1563–1591. https://doi.org/10.1021/CR960135G

Davis, M. M., & Bjorkman, P. J. (1988). T-cell antigen receptor genes and T-cell recognition. *Nature*, *334*(6181), 395–402. https://doi.org/10.1038/334395a0

Karadimas, C. L., Vu, T. H., Holve, S. A., Chronopoulou, P., Quinzii, C., Johnsen, S. D., … Hirano, M. (2006). Navajo neurohepatopathy is caused by a mutation in the MPV17 gene. *American Journal of Human Genetics*, *79*(3), 544–548. https://doi.org/10.1086/506913

Laidler, K. J. (1987). Chemical Kinetics. In *Chemical Kinetics* (3rd ed., pp. 80–136). New York: Harper Collins.

Mahdieh, N., & Rabbani, B. (2013). An overview of mutation detection methods in genetic disorders. *Iranian Journal of Pediatrics*, *23*(4), 375–388. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/24427490

Mercer, J. (2013). Cooperativity. In *Brenner's Encyclopedia of Genetics* (2nd ed., pp. 183–187). Elsevier. https://doi.org/10.1016/B978-0-12-374984-0.00339-9

Mirzaei, H. R., Mirzaei, H., Lee, S. Y., Hadjati, J., & Till, B. G. (2016). Prospects for chimeric antigen receptor (CAR) γδ T cells: a potential game changer for adoptive T cell cancer immunotherapy. *Cancer Letters*, *380*(2), 413. https://doi.org/10.1016/J.CANLET.2016.07.001

Qiagen. (2018). QIAamp DNA Blood Mini Kit.

Redmond, D., Poran, A., & Elemento, O. (2016). Single-cell TCRseq: paired recovery of entire T-cell alpha and beta chain transcripts in T-cell receptors from single-cell RNAseq. *Genome Medicine*, *8*(1), 80. https://doi.org/10.1186/s13073-016-0335-7

Reverse Complement. (n.d.). Retrieved March 24, 2018, from https://www.bioinformatics.org/sms/rev_comp.html

SantaLucia, J., & Hicks, D. (2004). The Thermodynamics of DNA Structural Motifs. *Annual Review of Biophysics and Biomolecular Structure*, *33*(1), 415–440. https://doi.org/10.1146/annurev.biophys.32.110601.141800

Satterfield, B. C. (2014). Cooperative Primers. *The Journal of Molecular Diagnostics*, *16*(2), 163–173. https://doi.org/10.1016/j.jmoldx.2013.10.004

Satterfield, B. C., Caplan, M. R., & West, J. A. A. (2008). Tentacle probe sandwich assay in porous polymer monolith improves specificity, sensitivity and kinetics. *Nucleic Acids Research*, *36*(19), e129. https://doi.org/10.1093/nar/gkn564

Satterfield, B. C., West, J. A. A., & Caplan, M. R. (2007). Tentacle probes: eliminating false positives without sacrificing sensitivity. *Nucleic Acids Research*, *35*(10), e76. https://doi.org/10.1093/nar/gkm113

Sharpe, M., & Mount, N. (2015). Genetically modified T cells in cancer therapy: opportunities and challenges. *Disease Models & Mechanisms*, *8*(4), 337–350. https://doi.org/10.1242/dmm.018036

Tyagi, S., & Kramer, F. R. (2012). Molecular beacons in diagnostics. *F1000 Medicine Reports*, *4*(10), 1–6. https://doi.org/10.3410/M4-10

Vu, T., Tanji, K., Holve, S. A., Bonilla, E., Sokol, R. J., Snyder, R. D., … Vivo, D. De. (2001). Navajo neurohepatopathy: A mitochondrial DNA depletion syndrome? *Hepatology*, *34*(1), 116–120. https://doi.org/10.1053/jhep.2001.25921

Whitcombe, D., Newton, C. R., & Little, S. (1998). Advances in approaches to DNA-based diagnostics. *Current Opinion in Biotechnology*, *9*, 602–608. Retrieved from http://biomednet.com/elecref/0958166900900602

Whitty, A. (2008). Cooperativity and biological complexity. *Nature Chemical Biology*, *4*(8), 435–439. https://doi.org/10.1038/nchembio0808-435

Wong, L.-J. C., Brunetti-Pierri, N., Zhang, Q., Yazigi, N., Bove, K. E., Dahms, B. B., … Craigen, W. J. (2007). Mutations in theMPV17 gene are responsible for rapidly progressive liver failure in infancy. *Hepatology*, *46*(4), 1218–1227. https://doi.org/10.1002/hep.21799

Zuker, M. (2003). Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Research*, *31*(13), 3406–3415. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12824337

APPENDIX A

FIRST ITERATION SEQUENCES FOR COOPERATIVE PRIMERS

Table A.1. Target and non-target sequences for left CP

Target: 5'-Vα-Jα-**Cα**-**Y**-BC-**X'**-3'

Cα	5'-ACATCCAGAACCCAGAACCTGCTGTGTACCAGTTAAAAGATCC TCGGTCTCAGGACAGCACCCTCTGCCTGTTCACCGACTTTGACTC CCAAATCAATGTGC-3'
Y	5'-GGACAGCAAAGACAGCACCT-3'
X'	5'-GCTCAGGGAAATAGCCCTTG-3'
Cα-Y-X'	5'-ACATCCAGAACCCAGAACCTGCTGTGTACCAGTTAAAAGATCC TCGGTCTCAGGACAGCACCCTCTGCCTGTTCACCGACTTTGACTC CCAAATCAATGTGCGGACAGCAAAGACAGCACCTGCTCAGGGAA A TAGCCCTTG-3'

Non-Target: 5'-**Y**-BC-**X'-Cβ'**-Jβ'-Dβ'-Vβ'-3'

Y	5'-GGACAGCAAAGACAGCACCT-3'
X'	5'-GCTCAGGGAAATAGCCCTTG-3'
Cβ'	5'-GGGTAGCCTTTTGTTTGTTTGCAATCTCTGCTTTTGATGGCTCA AACAAGGAGACCTTGGGTGGAGTCACATTTCTCAGATCCTC -3'
Y-X'-Cβ'	5'- GGACAGCAAAGACAGCACCTGCTCAGGGAAATAGCCCTTGGG GTAGCCTTTTGTTTGTTTGCAATCTCTGCTTTTGATGGCTCAAACA AGGAGACCTTGGGTGGAGTCACATTTCTCAGATCCTC-3'

Table A.2. Target and non-target sequences for right CP

Target: 5’-Vβ-Dβ-Jβ-**Cβ-X-**BC-**Y’**-3’	
Cβ	5’- GAGGATCTGAGAAATGTGACTCCACCCAAGGTCTCCTTGTTTG AGCCATCAAAAGCAGAGATTGCAAACAAACAAAAGGCTACCC-3’
X	5’-CAAGGGCTATTTCCCTGAGC-3’
Y’	5’-AGGTGCTGTCTTTGCTGTCC-3’
Cβ-X-Y’	5’-GAGGATCTGAGAAATGTGACTCCACCCAAGGTCTCCTTGTTTG AGCCATCAAAAGCAGAGATTGCAAACAAACAAAAGGCTACCCCA AGGGCTATTTCCCTGAGCAGGTGCTGTCTTTGCTGTCC-3’
Non-Target: 5’-**X-Y’-Cα’**-Jα’-Vα’-3’	
X	5’-CAAGGGCTATTTCCCTGAGC-3’
Y’	5’-AGGTGCTGTCTTTGCTGTCC-3’
Cα’	5’-GCACATTGATTTGGGAGTCAAAGTCGGTGAACAGGCAGAGGG TGCTGTCCTGAGACCGAGGATCTTTTAACTGGTACACAGCAGGTT CTGGGTTCTGGATGT-3’
X-Y’- Cα’	5’-CAAGGGCTATTTCCCTGAGCAGGTGCTGTCTTTGCTGTCCGCA CATTGATTTGGGAGTCAAAGTCGGTGAACAGGCAGAGGGTGCTG TCCTGAGACCGAGGATCTTTTAACTGGTACACAGCAGGTTCTGGG TTCTGGATGT-3’

Table A.3. Short primer sequences for left CP

5’-GCTCAGGGAAATAGCCCTTG-3’

Short Primer #	Primer Sequence (5’ → 3’) 1bp removal from 5’ → 3’	bp
1	CAAGGGCTATTTCCC	15
2	CAAGGGCTATTTCC	14
3	CAAGGGCTATTTC	13
4	CAAGGGCTATTT	12
5	CAAGGGCTATT	11
6	CAAGGGCTAT	10
7	CAAGGGCTA	9
8	CAAGGGCT	8
9	CAAGGGC	7
10	CAAGGG	6
11	CAAGG	5
12	CAAG	4

5’-GCTCAGGGAAATAGCCCTTG-3’

Short Primer #	Primer Sequence (5’ → 3’) 1bp removal from 5’ → 3’	bp
13	GCTATTTCCCTGAGC	15
14	CTATTTCCCTGAGC	14
15	TATTTCCCTGAGC	13

16	ATTTCCCTGAGC	12
17	TTTCCCTGAGC	11
18	TTCCCTGAGC	10
19	TCCCTGAGC	9
20	CCCTGAGC	8
21	CCTGAGC	7
22	CTGAGC	6
23	TGAGC	5
24	GAGC	4

Table A.4. Capture region sequences for left CP

Capture Region #	Primer Sequence (5' → 3') 1bp removal from 5' → 3'	bp
1	GCACATTGATTTGGGAGTCAAA	22
2	CACATTGATTTGGGAGTCAAA	21
3	ACATTGATTTGGGAGTCAAA	20
4	CATTGATTTGGGAGTCAAA	19
5	ATTGATTTGGGAGTCAAA	18
6	TTGATTTGGGAGTCAAA	17
7	TGATTTGGGAGTCAAA	16
8	GATTTGGGAGTCAAA	15
9	ATTTGGGAGTCAAA	14
10	TTTGGGAGTCAAA	13
11	TTGGGAGTCAAA	12
12	TGGGAGTCAAA	11
13	GGGAGTCAAA	10

Table A.5. Short primer sequences for right CP

5’-AGGTGCTGTCTTTGCTGTCC-3’

Short Primer #	Primer Sequence (5’ → 3’) 1bp removal from 3’ → 5’	bp
1	GGACAGCAAAGACAG	15
2	GGACAGCAAAGACA	14
3	GGACAGCAAAGAC	13
4	GGACAGCAAAGA	12
5	GGACAGCAAAG	11
6	GGACAGCAAA	10
7	GGACAGCAA	9
8	GGACAGCA	8
9	GGACAGC	7
10	GGACAG	6
11	GGACA	5
12	GGAC	4

Closer to 5’-AGGTGCTGTCTTTGCTGTCC-3’

Short Primer #	Primer Sequence (5’ → 3’) 1bp removal from 5’ → 3’	bp
13	GCAAAGACAGCACCT	15
14	CAAAGACAGCACCT	14
15	AAAGACAGCACCT	13

16	AAGACAGCACCT	12
17	AGACAGCACCT	11
18	GACAGCACCT	10
19	ACAGCACCT	9
20	CAGCACCT	8
21	AGCACCT	7
22	GCACCT	6
23	CACCT	5
24	ACCT	4

Table A.6. Capture region sequences for right CP

Capture Region #	**Primer Sequence (5' → 3')** 1bp removed from 5'→3'	**bp**
1	GGGTAGCCTTTTGTTTGTTTGC	22
2	GGTAGCCTTTTGTTTGTTTGC	21
3	GTAGCCTTTTGTTTGTTTGC	20
4	TAGCCTTTTGTTTGTTTGC	19
5	AGCCTTTTGTTTGTTTGC	18
6	GCCTTTTGTTTGTTTGC	17
7	CCTTTTGTTTGTTTGC	16
8	CTTTTGTTTGTTTGC	15
9	TTTTGTTTGTTTGC	14
10	TTTGTTTGTTTGC	13
11	TTGTTTGTTTGC	12
12	TGTTTGTTTGC	11
13	GTTTGTTTGC	10

APPENDIX B

SECOND ITERATION SEQUENCES FOR COOPERATIVE PRIMERS

Table B.1. Short primer sequences for left CP

Short Primer #	Primer Sequence (5' → 3')	%GC Content	Target/Non-Target Sequences (5' → 3')
1	AACGGACGCAGC	66.7	**Target:** ACATCCAGAACCCAGAACCT GCTGTGTACCAGTTAAAGATCCTCG GTCTCAGGACAGCACCCTCTGCCTG TTCACCGACTTTGACTCCCAAATCAA TGTGCGCTGCGTCCGTT **Non-Target:** GCTGCGTCCGTTGGGTA GCCTTTTGTTTGTTTGCAATCTCTGC TTTTGATGGCTCAAACAAGGAGACC TTGGGTGGAGTCACATTTCTCAGATC CTC

2	TACGGACGCAGC	66.7	**Target:** ACATCCAGAACCCAGAACCT GCTGTGTACCAGTTAAAAGATCCTC GGTCTCAGGACAGCACCCTCTGCCT GTTCACCGACTTTGACTCCCAAATCA ATGTGCTGCGTCCGTA **Non-Target:** GCTGCGTCCGTAGGGTA GCCTTTTGTTTGTTTGCAATCTCTGC TTTTGATGGCTCAAACAAGGAGACC TTGGGTGGAGTCACATTTCTCAGATC CTC
3	TACGGACGCAAC	58.3	**Target:** ACATCCAGAACCCAGAACCT GCTGTGTACCAGTTAAAAGATCCTC GGTCTCAGGACAGCACCCTCTGCCT GTTCACCGACTTTGACTCCCAAATCA ATGTGTTGCGTCCGTA **Non-Target:** GTTGCGTCCGTAGGGTA GCCTTTTGTTTGTTTGCAATCTCTGC TTTTGATGGCTCAAACAAGGAGACC TTGGGTGGAGTCACATTTCTCAGATC CTC

4	TACGGACGCTAC	58.3	**Target:** ACATCCAGAACCCAGAACCT GCTGTGTACCAGTTAAAAGATCCTC GGTCTCAGGACAGCACCCTCTGCCT GTTCACCGACTTTGACTCCCAAATCA ATGTGTAGCGTCCGTA **Non-Target:** GTAGCGTCCGTAGGGTA GCCTTTTGTTTGTTTGCAATCTCTGC TTTTGATGGCTCAAACAAGGAGACC TTGGGTGGAGTCACATTTCTCAGATC CTC
5	TACGAACGCAAC	50.0	**Target:** ACATCCAGAACCCAGAACCT GCTGTGTACCAGTTAAAAGATCCTC GGTCTCAGGACAGCACCCTCTGCCT GTTCACCGACTTTGACTCCCAAATCA ATGTGTTGCGTTCGTA **Non-Target:** GTTGCGTTCGTAGGGTA GCCTTTTGTTTGTTTGCAATCTCTGC TTTTGATGGCTCAAACAAGGAGACC TTGGGTGGAGTCACATTTCTCAGATC CTC

6	TACGATCGCAAC	50.0	**Target:** ACATCCAGAACCCAGAACCT GCTGTGTACCAGTTAAAAGATCCTC GGTCTCAGGACAGCACCCTCTGCCT GTTCACCGACTTTGACTCCCAAATCA ATGTGTTGCGATCGTA **Non-Target:** GTTGCGATCGTAGGGTA GCCTTTTGTTTGTTTGCAATCTCTGC TTTTGATGGCTCAAACAAGGAGACC TTGGGTGGAGTCACATTTCTCAGATC CTC

Table B.2. Capture region sequences for left CP

Target Sequence: 5'-ACATCCAGAACCCAGAACCTGCTGTGTACCAGTTAAAAG ATCCTCGGTCTCAGGACAGCACCCTCTGCCTGTTCACCGACTTTGACTCCCAA ATCAATGTGC-3'

Non-Target Sequence: 5'-GGGTAGCCTTTTGTTTGTTTGCAATCTCTGCTTTTGA TGGCTCAAACAAGGAGACCTTGGGTGGAGTCACATTTCTCAGATCCTC-3'

Primer #	**Primer Sequence (5' → 3')** 1bp removal from 5' → 3'
1	GCACATTGATTTGGGAGTCAAA
2	CACATTGATTTGGGAGTCAAA
3	ACATTGATTTGGGAGTCAAA
4	CATTGATTTGGGAGTCAAA
5	ATTGATTTGGGAGTCAAA
6	TTGATTTGGGAGTCAAA
7	TGATTTGGGAGTCAAA
8	GATTTGGGAGTCAAA
9	ATTTGGGAGTCAAA
10	TTTGGGAGTCAAA
11	TTGGGAGTCAAA
12	TGGGAGTCAAA
13	GGGAGTCAAA

Table B.3. Short primer sequences for left CP

Short Primer #	Primer Sequence (5' → 3')	%GC Content	Target/Non-Target Sequences (5' → 3')
1	GCACCCCGGAAT	66.7	**Target:** GAGGATCTGAGAAATGTGACT CCACCCAAGGTCTCCTTGTTTGAGCCA TCAAAAGCAGAGATTGCAAACAAACA AAAGGCTACCCATTCCGGGGTGC **Non-Target:** ATTCCGGGGTGCGCACAT TGATTTGGGAGTCAAAGTCGGTGAAC AGGCAGAGGGTGCTGTCCTGAGACCG AGGATCTTTTAACTGGTACACAGCAG GTTCTGGGTTCTGGATGT
2	GCACCCCGGTAT	66.7	**Target:** GAGGATCTGAGAAATGTGACT CCACCCAAGGTCTCCTTGTTTGAGCCA TCAAAAGCAGAGATTGCAAACAAACA AAAGGCTACCCATACCGGGGTGC **Non-Target:** ATACCGGGGTGCGCACAT TGATTTGGGAGTCAAAGTCGGTGAAC AGGCAGAGGGTGCTGTCCTGAGACCG AGGATCTTTTAACTGGTACACAGCAG GTTCTGGGTTCTGGATGT

3	ACACCCCGGTAT	58.3	**Target:** GAGGATCTGAGAAATGTGACT CCACCCAAGGTCTCCTTGTTTGAGCCA TCAAAAGCAGAGATTGCAAACAAACA AAAGGCTACCCATACCGGGGTGT **Non-Target:** ATACCGGGGTGTGCACAT TGATTTGGGAGTCAAAGTCGGTGAAC AGGCAGAGGGTGCTGTCCTGAGACCG AGGATCTTTTAACTGGTACACAGCAG GTTCTGGGTTCTGGATGT
4	ACACCCCGGAAT	58.3	**Target:** GAGGATCTGAGAAATGTGACT CCACCCAAGGTCTCCTTGTTTGAGCCA TCAAAAGCAGAGATTGCAAACAAACA AAAGGCTACCCATTCCGGGGTGT **Non-Target:** ATTCCGGGGTGTGCACAT TGATTTGGGAGTCAAAGTCGGTGAAC AGGCAGAGGGTGCTGTCCTGAGACCG AGGATCTTTTAACTGGTACACAGCAG GTTCTGGGTTCTGGATGT

5	AAACCCCGGAAT	50.0	**Target:** GAGGATCTGAGAAATGTGACT CCACCCAAGGTCTCCTTGTTTGAGCCA TCAAAAGCAGAGATTGCAAACAAACA AAAGGCTACCCATTCCGGGGTTT **Non-Target:** ATTCCGGGGTTTGCACAT TGATTTGGGAGTCAAAGTCGGTGAAC AGGCAGAGGGTGCTGTCCTGAGACCG AGGATCTTTTAACTGGTACACAGCAG GTTCTGGGTTCTGGATGT
6	AATCCCCGGAAT	50.0	**Target:** GAGGATCTGAGAAATGTGACT CCACCCAAGGTCTCCTTGTTTGAGCCA TCAAAAGCAGAGATTGCAAACAAACA AAAGGCTACCCATTCCGGGGATT **Non-Target:** ATTCCGGGGATTGCACAT TGATTTGGGAGTCAAAGTCGGTGAAC AGGCAGAGGGTGCTGTCCTGAGACCG AGGATCTTTTAACTGGTACACAGCAG GTTCTGGGTTCTGGATGT

Table B.4. Capture region sequences for right CP

Target: 5'-GAGGATCTGAGAAATGTGACTCCACCCAAGGTCTCCTTGTTTGAGC CATCAAAAGCAGAGATTGCAAACAAACAAAAGGCTACCC-3'

Non-Target: 5'-GCACATTGATTTGGGAGTCAAAGTCGGTGAACAGGCAGAGGG TGCTGTCCTGAGACCGAGGATCTTTTAACTGGTACACAGCAGGTTCTGGGTTC TGGATGT-3'

Capture Region #	**Primer Sequence (5' → 3')** 1bp removed from 5'→3'	**bp**
1	GGGTAGCCTTTTGTTTGTTTGC	22
2	GGTAGCCTTTTGTTTGTTTGC	21
3	GTAGCCTTTTGTTTGTTTGC	20
4	TAGCCTTTTGTTTGTTTGC	19
5	AGCCTTTTGTTTGTTTGC	18
6	GCCTTTTGTTTGTTTGC	17
7	CCTTTTGTTTGTTTGC	16
8	CTTTTGTTTGTTTGC	15
9	TTTTGTTTGTTTGC	14
10	TTTGTTTGTTTGC	13
11	TTGTTTGTTTGC	12
12	TGTTTGTTTGC	11
13	GTTTGTTTGC	10

APPENDIX C

THIRD ITERATION SEQUENCES FOR COOPERATIVE PRIMERS

Table C.1. Primer sequences for left CP

Target Sequence: 5'-ACATCCAGAACCCAGAACCTGCTGTGTACCAGTTAAAAG ATCCTCGGTCTCAGGACAGCACCCTCTGCCTGTTCACCGACTTTGACTCCCAA ATCAATGTGCGCCGGATCACACCCCGGTATGTAGCGTCCGTAGCCCGTGG-3'

Non-Target Sequence: 5'-GCCGGATCACACCCCGGTATGTAGCGTCCGTAGCCC GTGGGGGTAGCCTTTTGTTTGTTTGCAATCTCTGCTTTTGATGGCTCAAACAA GGAGACCTTGGGTGGAGTCACATTTCTCAGATCCTC-3'

Primer Type	**Primer Sequence (5' → 3')**
Short	TACGGACGCTAC
Capture	TTGATTTGGGAGTCAAA

Table C.2. Primer sequences for right CP

Target: 5'-GAGGATCTGAGAAATGTGACTCCACCCAAGGTCTCCTTGTTTGAGC

CATCAAAAGCAGAGATTGCAAACAAACAAAAGGCTACCCCCACGGGCTACG

GACGCTACATACCGGGGTGTGATCCGGC-3'

Non-Target: 5'-CCACGGGCTACGGACGCTACATACCGGGGTGTGATCCGGCGC

ACATTGATTTGGGAGTCAAAGTCGGTGAACAGGCAGAGGGTGCTGTCCTGAG

ACCGAGGATCTTTTAACTGGTACACAGCAGGTTCTGGGTTCTGGATGT-3'

Primer Type	**Primer Sequence (5' → 3')**
Short	ACACCCCGGTAT
Capture	CCTTTTGTTTGTTTGC