

Syracuse University

SURFACE at Syracuse University

Dissertations - ALL

SURFACE at Syracuse University

5-14-2023

Development of a Novel Bistable DNA Sensor for Anti-HIV Drug Discovery and Re-engineering of Recombinant Cyclin T1-Tat Protein with SUMO Fusion in Escherichia Coli

Nan Thuzar Myint
Syracuse University

Follow this and additional works at: <https://surface.syr.edu/etd>

Recommended Citation

Myint, Nan Thuzar, "Development of a Novel Bistable DNA Sensor for Anti-HIV Drug Discovery and Re-engineering of Recombinant Cyclin T1-Tat Protein with SUMO Fusion in Escherichia Coli" (2023).
Dissertations - ALL. 1714.
<https://surface.syr.edu/etd/1714>

This Dissertation is brought to you for free and open access by the SURFACE at Syracuse University at SURFACE at Syracuse University. It has been accepted for inclusion in Dissertations - ALL by an authorized administrator of SURFACE at Syracuse University. For more information, please contact surface@syr.edu.

Abstract

This dissertation focuses on the interaction of proteins and nucleic acids and their applications. It includes two projects: I: Development of a Novel Bistable DNA Sensor for Anti-HIV Drug Discovery and II. Re-engineering of Recombinant Cyclin T1-Tat Protein with SUMO Fusion in Escherichia Coli.

I: Development of a Novel Bistable DNA Sensor for Anti-HIV Drug Discovery

Screening drug compounds targeting HIV-1 NCp7 provide attractive candidates for new anti-retroviral therapeutics because of the highly conserved nature of the zinc fingers in NCp7¹ in selecting and packaging RNA in the HIV-1 life cycle. The unique 3-segment, reversible switch for high throughput screening (HTS) drug targets will be for the HIV-1 nucleocapsid (NC) protein. The Probe is a natural binding element for the NCp7 protein target or, in our case, an aptamer hairpin with loop sequence, TGTGGT, having a nano-molar affinity ($K_d=16\text{nM}$).² Toggle is a damaged probe where the target-binding sequences are replaced with other bases. Cover is a mostly complementary strand for the probe and the toggle. A hairpin loop forms around the 5Me-dC-brancher in both the ON and OFF forms. Using Visual OMPTM simulation, the following two potential switch molecules (NM-1 and NM-2) were designed and successfully synthesized using a one-step ligation method with at least 90% purity as judged by mass spectrometry. Then fluorescence measurements using NM-1 and NM-2 with NCp7 protein were analyzed to demonstrate proof-of-principle for 3-segment nucleic acid switches. Increasing [NC] causes a dramatic decrease in CY3 fluorescence. The ON/OFF contrast ratio of NM-1 and NM-2 are 2.6 and 3.2 showing the feasibility of 3-segment switches being used to further modify the design of HTS switches for the HIV-1 NC.

II. Re-engineering of Recombinant Cyclin T1-Tat Protein with SUMO Fusion in Escherichia Coli.

Due to increasing drug resistance for current antiretroviral therapeutic (ART) treatments, it is important to add drugs targeting other aspects of HIV-1 infection to manage AIDS. Tat (trans-activator of transcription) protein up-regulates the transcription of viral-specific proteins by a factor of 1,000³, which makes Tat a very attractive drug target. High-level expression and purification of recombinant GST-CycT1(249-280)-linker(25aa)-Tat(1-101) fusion protein in *Escherichia coli* (*E. coli*) was challenging because CycT1-Tat forms inclusion bodies making it difficult to purify and obtain a high concentration of active CycT1-Tat. A commercially available pET-SUMO cloning vector was introduced for the high-level expression of four CycT1-Tat chimeras, F1-F4. The gene associated with the protein sequence was purchased from IDT and SUMO-CycT1-Tat (F1) fusion plasmid was prepared and transformed in *E. coli* BL21 (DE3). By inducing with IPTG, His₆-SUMO-CycT1-Tat fusion protein was able to promote soluble expression making it more efficient to purify CycT1-Tat (F1) compared to GST fusion tag protein. The molecular weight of purified CycT1-Tat was confirmed with mass spectrometry (MALDI-TOF). For initial characterization of the TAR-CycT1-Tat complex, an Electrophoretic Mobility Shift Assay (EMSA) was carried out with partially purified CycT1-Tat protein using TAR-31 and truncated versions that altered the hairpin loop (TAR-H24) and deleted the bulge loop (TAR-B25). A TAR-CycT1-Tat complex was formed and a band shift was observed in the binding of CycT1-Tat to TAR-31 and a lower affinity complex with TAR-B25, but TAR-H24 did not bind.

Development of a Novel Bistable DNA Sensor for Anti-HIV Drug Discovery

and

Re-engineering of Recombinant Cyclin T1-Tat Protein with SUMO Fusion in

Escherichia Coli

by

Nan Thuzar Myint

B.S., Utica University, 2012

Dissertation

Submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Chemistry

Syracuse University

May 2023

Copyright © Nan Thuzar Myint 2023
All Rights Reserved

Acknowledgments

I'd like to express my deepest gratitude to my advisor Dr. Philip N. Borer for his knowledge, invaluable patience, and feedback. I'm so grateful to be his last graduate student in the Borer lab, allowing me to grow as a scientist with ample support and encouragement.

Thank you, Dr. Ramesh Raina, for serving as the chair of my dissertation committee and to Dr. Yang-Yeung Luk as my co-advisor, Dr. John Franck, Dr. Joseph Chaiken, and Dr. Olga V. Makhlynets for serving on my dissertation committee. I'd like to extend my sincere thanks to Dr. Mark McPike and Dr. Deborah Kerwood for their insightful advice and support throughout my graduate study. Many thanks to my awesome lab mates, Dr. Caitlin Miller, and Dr. Raghuvanan Iyer for their expertise in deep knowledge of nucleic acid, oligonucleotide synthesis, molecular biology, and supportive conversations. I'm also thankful for all the friends I made in the United States for their moral support and great memories.

I'd like to acknowledge my parents and siblings who provided unconditional love and emotional support. Words can't describe how thankful I am for all the sacrifices that my parents made for me to get a better education in the United States. Their belief in me motivated and helped me through the dark times. Last but not least, I'd also like to thank my nephew and niece, Hendrix and Irene, for their laughs and cheerfulness to keep my spirits high during this process.

Table of Contents

Abstract.....	i
Acknowledgments	v
List of Figures.....	viii
List of Tables	xii
Chapter 1: HIV and AIDS	1
1.1. HIV-1 genome and virion structure	2
1.2. HIV-1 Replication Cycle.....	5
1.3. New Targets: HIV-1 Nucleocapsid protein (NCp7) and Trans-activator for Transcription (Tat) for anti-retroviral drug therapies	8
1.4. Aims of the Projects	17
Part I: Development of a Novel Bistable DNA Sensor for Anti-HIV Drug Discovery	19
Chapter 2: A Bistable aptamer-based DNA switch sensor for Drug Discovery	20
Chapter Summary.....	20
2.1. Aptamer	21
2.2. 3-segment DNA switch sensor for drug discovery	25
Chapter 3: Direct Synthesis of Switch and Optimization of Switch Synthesis	31
Chapter Summary.....	31
3.1. Direct Synthesis of switches using 3'-DABCYL CPG.....	33
3.2. Optimization of switch synthesis	41
3.3. Optimization of the Purification of Switch Molecules.....	72
3.4. Alternative synthesis of switch molecule using 3'-PT-amino-C6 Modifier	90
3.1. Direct Synthesis of a switch using 3'-DABCYL CPG	103
3.2. Optimization of switch synthesis	105
3.3. Optimization of the Purification of Switch Molecules.....	107
3.4. Alternative synthesis of switch molecule using 3'-PT-amino-C6 Modifier	108

Chapter 4: Synthesis, ligation, and characterization of oligonucleotide switches	109
Chapter Summary.....	109
4.1. Design criteria for switches.....	112
4.2. Design of a switch molecule for enzymatic ligation	117
4.3. Synthesis and Analysis of Phosphorylated Test Sequence	122
4.4. Synthesis and Purification of Switch NM-1 using ligation.....	129
4.5. Design and synthesis of Switch NM-2.....	141
4.6. Locked-ON and Locked-OFF molecules	145
4.7. Analysis of NCp7 binding to Switch NM-1 and 2.....	155
Supplemental Information	177
4.8. Ligation and Purification of NM-1 using two-step ligation.....	177
4.9. Determination of Extinction Coefficient.....	183
Part II: Re-engineering of Recombinant human Cyclin T1-Tat Chimera with SUMO Fusion in Escherichia Coli	194
Chapter 5: Re-engineering a human Cyclin T1 (hCycT1)- Tat Chimera.....	195
Chapter Summary.....	195
5.1. Cyclin T1 (CycT1)-Tat Chimera.....	197
5.2. Cloning, Overexpression, and Purification of His ₆ -SUMO tagged hCycT1-Tat proteins	205
5.3. Analysis of CycT1-Tat fusion protein.....	220
References	242
Bibliographic Information	254

List of Figures

Figure 1.1: (a) HIV-1 genome (adapted from ^{12, 13}) and (b) HIV-1 mature virion. ¹²	4
Figure 1.2: The Replication Cycle of the HIV-1 Virus. ¹²	7
Figure 1.3: Structure of NCp7 and the region of the unspliced viral RNA (Ψ site).	11
Figure 1.4: HIV-1 Tat-101 protein.....	13
Figure 1.5: TAR-CyclT1-Tat Interaction.....	15
Figure 2.1: An illustration of the AlloSwitch. ⁵⁵	24
Figure 2.2: Design of a branched switch molecule.....	26
Figure 3.1: Solid-phase synthesis of an oligonucleotide. ⁶⁶	34
Figure 3.2: Synthesis Scheme of 3-segment DNA Switch.	36
Figure 3.3: Switch NM-0 sequences for direct synthesis.	38
Figure 3.4: Analytical IE- HPLC of Switch NM-0.....	39
Figure 3.5: ESI-MS spectrum of NM-0.	40
Figure 3.6: Sequences for Optimization of Coupling Efficiency.	44
Figure 3.7: A truncated switch model for Optimization of Coupling Efficiency.	45
Figure 3.8: Test sequences composed of Thymidine phosphoramidite.....	48
Figure 3.9: Branched loops models (ACn).	49
Figure 3.10: Model to mimic Beta Branched Synthesis.	50
Figure 3.11: dT-CE phosphoramidite. ⁶⁹	52
Figure 3.12: Analytical IE-HPLC chromatogram of T7 control with original synthesis procedure.	53

Figure 3.13: Analytical IE-HPLC chromatogram of T7 control with modified synthesis procedure.....	54
Figure 3.14: IE-HPLC chromatogram of T7 branched molecule using the modified procedure.	57
Figure 3.15: IE-HPLC chromatogram of T7 branched molecule using the original procedure. ..	58
Figure 3.16: IE-HPLC chromatogram of the 37mer truncated switch molecule.....	60
Figure 3.17: ESI-MS spectrum of peak 12_1 from Figure. 3.16.	61
Figure 3.18: Analytical IE-HPLC of T6[10] molecule.....	66
Figure 3.19: Analytical IE-HPLC of T6[10] molecule without optimized procedure.....	67
Figure 3.20: Analytical IE- HPLC chromatogram of T _{7_2_7} beta-branched molecule.	70
Figure 3.21: Structure of non-branched DNA hairpin.	75
Figure 3.22: Unlabeled Switch Molecules.....	77
Figure 3.23: Controls for Unlabeled Switch Molecules.	78
Figure 3.24: Analytical IE-HPLC chromatogram of AC ₇ using GP purification.	81
Figure 3.25: Analytical IE-HPLC chromatogram of AC ₇ (control) without prior GP purification.	82
Figure 3.26: Analytical IE-HPLC analysis of full-length unlabeled Switch 1.	84
Figure 3.27: IE-HPLC analysis of the control for unlabeled Switch 1.....	85
Figure 3.28: Analytical IE-HPLC of full-length unlabeled Switch 1.	86
Figure 3.29: IE-HPLC analysis of CY3 labeled linear strand using GP purification.	88
Figure 3.30: Analytical IE-HPLC analysis of CY3 labeled 20mer.	89
Figure 3.31: Conjugation of DABCYL. ⁷³	91
Figure 3.32: Sequences for Conjugation of DABCYL succinimidyl ester.....	93

Figure 3.33: Short Branched Sequences for Conjugation of DABCYL Succinimidyl Ester Analysis.....	95
Figure 3.34: RP-HPLC purification chromatogram of DABCYL conjugated 20mer.....	97
Figure 3.35: RP-HPLC purification chromatogram of DABCYL conjugated 51mer.....	98
Figure 3.36: RP-HPLC purification chromatogram of DABCYL conjugated 51mer.....	101
Figure 3.37: ESI-MS chromatogram of DABCYL conjugated oligonucleotide.....	102
Figure 4.1: A 3-segment 70mer DNA switch, Switch NM-1.....	113
Figure 4.2: Output from a V-OMP simulation of Switch NM-1.....	114
Figure 4.3: Covalent connections in 3-base α loops (a) and β loops (b).....	116
Figure 4.4: Schematic of One-Step Ligation.....	118
Figure 4.5: Switch NM-1 sequences for one-step (a) and two-step (b) ligation.....	119
Figure 4.6: Schematic for one-step ligation.....	121
Figure 4.7: The mechanism of removal of CPR II. ⁷⁵	123
Figure 4.8: Analytical RP-HPLC chromatogram of a DMT-pT15 sample.....	126
Figure 4.9: Analytical RP-HPLC chromatogram of pT15.....	127
Figure 4.10: Ligation trials for the first step in two-step ligation (20% denaturing PAGE gel.)	133
Figure 4.11: Analytical 20% denaturing PAGE gel following one-step ligation.....	136
Figure 4.12: Purification of 70mer switch NM-1 after one-step ligation.....	138
Figure 4.13: ESI-MS analysis of ligated switch NM-1.....	140
Figure 4.14: 71mer switch NM-2.....	142
Figure 4.15: DNase I hydrolysis of switch NM-2.....	144
Figure 4.16: Sequences of Locked-ON and Locked-OFF states.....	146

Figure 4.17: Analytical 20% denaturing PAGE gel.....	149
Figure 4.18:Analytical 20% denaturing PAGE gel of ligation where the Core was treated with T4 PNK.....	150
Figure 4.19 Fluorescence emission of Locked-ON and Locked-OFF molecules.....	154
Figure 4.20: Theoretical curves for switches show the effect of K1 on $K_d(\text{app})$	158
Figure 4.21: Fluorescence measurements for switches NM-1 and NM-2.	160
Figure 4.22: Interpretation of the binding equilibrium for Switch NM-1.....	163
Figure 4.23: Interpretation of data for NM-2.....	164
Figure 5.1: CycT1-Tat Chimera.....	198
Figure 5.2: Optimized Chimera of HIV-1 CycT1-Tat.....	201
Figure 5.3: Four CycT1-Tat fusion proteins were prepared in this work. The N-terminus is on the left and the C-terminus is on the right.	203
Figure 5.4: Constructions of His ₆ -SUMO-CycT1-Tat fusion protein.	207
Figure 5.5 Construction of Champion TM pET SUMO vector system. ⁹³	209
Figure 5.6: 15% SDS-PAGE gel image of optimized HIV-1 CycT1-Tat fusion proteins.....	213
Figure 5.7: 15% SDS-PAGE gel of purification of His ₆ -SUMO-CycT1-Tat protein.	215
Figure 5.8: 15% SDS-PAGE of SUMO protease cleavage applied to His ₆ -SUMO-F1.	218
Figure 5.9: 15% SDS-PAGE analysis of purification of CycT1-Tat (F1) protein.....	219
Figure 5.10: MALDI-TOF mass spectrum of CycT1-Tat protein.	221
Figure 5.11: Various secondary structures of TAR-RNA.	224
Figure 5.12: Binding of recombinant CycT1-Tat (F1) protein to various HIV-1 TAR RNAs...	226

List of Tables

Table 3.1: Trityl yield of T7 branched molecule.	55
Table 3.2: ESI-MS data of a truncated switch molecule.	62
Table 3.3: Retention time of the prominent peaks on analytical IE-HPLC.	65
Table 3. 4: Comparison of % of full-length constructs built around © based on IE HPLC.	71
Table 3.5: ESI-MS analysis of DABCYL conjugated sample.	99
Table 4.1: Analysis of Phosphorylated Test Sequences.	128
Table 4.2: Data for Switch NM-1	165
Table 4.3: Data for Switch NM-2	166

Chapter 1: HIV and AIDS

Acquired immune deficiency syndrome (AIDS) is a blood-borne infectious and sexually transmitted disease (STD) affecting the human immune system. In the early part of the epidemic, it was also spread by blood transfusion, but blood banks now prevent this by careful screening of donors and the use of nucleic acid amplification testing for the HIV-1 genome. Human Immunodeficiency Virus (HIV) is a pathogenic RNA virus belonging to the family *Retroviridae*, the genus *Lentivirus*, that emerged by zoonotic transmission. HIV consists of types 1 and 2, and HIV-1 is the causative agent of AIDS worldwide.⁴ HIV-1 virus destroys and impairs the immune cells' function, weakening people's defense systems against infections, and rare types of cancers. AIDS is still considered a global health threat that has already caused the death of approximately 40.1 million people. About 38.4 million people have acquired HIV since the beginning of the epidemic, and 1.5 million people are newly infected, according to the Joint United Nations Program on HIV/AIDS (UNAIDS) in 2021.⁴

With progress in the development of highly active antiretroviral therapy (HAART), which combines several 25 antiretroviral drugs targeting the key HIV-1 enzymes, i.e., reverse transcriptase (RT), protease (PR), or integrase (IN), the mortality and morbidity have been reduced among HIV-infected patients.⁵ However, HIV has not been eradicated yet due to the emergence of drug resistance (DR) and toxicity from current antiretroviral regimes.⁶

HIV-1 virus has high mutation rates⁶⁻⁹ and can insert its genomic material into the host cell using viral and host enzymes and proteins, and then produce and package newly made genomic RNA and proteins to form new virus particles that go on to infect more cells. Despite the advances in HAART, the transcriptionally silenced virus persists in a small number of cells and leads to

resistance to current antiretroviral drugs. Thus, there is a need to discover other drug candidates targeting highly conserved viral proteins that prevent the rapid growth of resistant strains¹⁰ and augment long-term treatment strategies for people living with the virus.

1.1. HIV-1 genome and virion structure

HIV-1 mature virion is a round, enveloped retrovirus (figure 1.1b) with a 9.8 kb RNA genome (gRNA) (figure 1.1a) which is a dimer of single-stranded positively sensed gRNAs, encapsulating in the capsid. Three polyproteins (Gag, Pol, and Env) and six accessory proteins (Tat, Rev, Nef, Vpr, Vif, and Vpu) coded in gRNA are essential for replication and viral assembly in the infected host cells.¹¹ The genome is flanked by 5' and 3' - long terminal repeats (LTRs) including the viral promoter for gene expression, integration, and reverse transcription.

There are nine different open reading frames (ORFs) for genetic information to synthesize 15 proteins, and some viral genes have overlapped regions in the gRNA. The *gag* gene encodes the structural component of viral core proteins, matrix (MA), capsid (CA), and nucleocapsid (NC). The consecutive *pol* gene encodes viral enzymes protease (PR), reverse transcriptase (RT), and integrase (IN). The two envelop glycoproteins, gp120-SU (surface unit) and the gp41-TM (transmembrane unit), are encoded in the *env* gene.¹²

The coding regions of two regulatory genes, *rev* and *tat*, and four accessory genes *vif*, *vpr*, *vpu*, and *nef* are located after the *pol* gene. Tat and Rev are vital for viral replication in the host cell nucleus which binds to the Rev-responsive element (RRE) and Trans-activation response (TAR). The rate of virion production is controlled by *vif*, *vpr*, and *vpu*, and the last accessory gene, *nef*, regulates the HIV infectivity by downregulating several host cell proteins.¹²

In the structure of the HIV-1 mature virion (figure 1.1b), gRNA of the HIV-1 virus is protected mainly by three main layers where the outermost layer is a lipid bilayer membrane followed by matrix (MA, p17) and capsid (CA, p24). The structure of single-stranded gRNAs is supported by the nucleocapsid proteins (NC, p7). The *env* glycoproteins, gp120-SU, and gp41-TM encoded in the *env* gene are formed from the cleavage of the Env polyprotein (gp160) with the cellular furin-like protease. They are located on the lipid bilayer membrane to interact with the receptors on the host cell. The viral enzymes, protease (PR, p10), integrase (IN, p32), reverse transcriptase (RT, p66, and p51), and the accessory protein Vpr play an essential role in the virion maturation process.

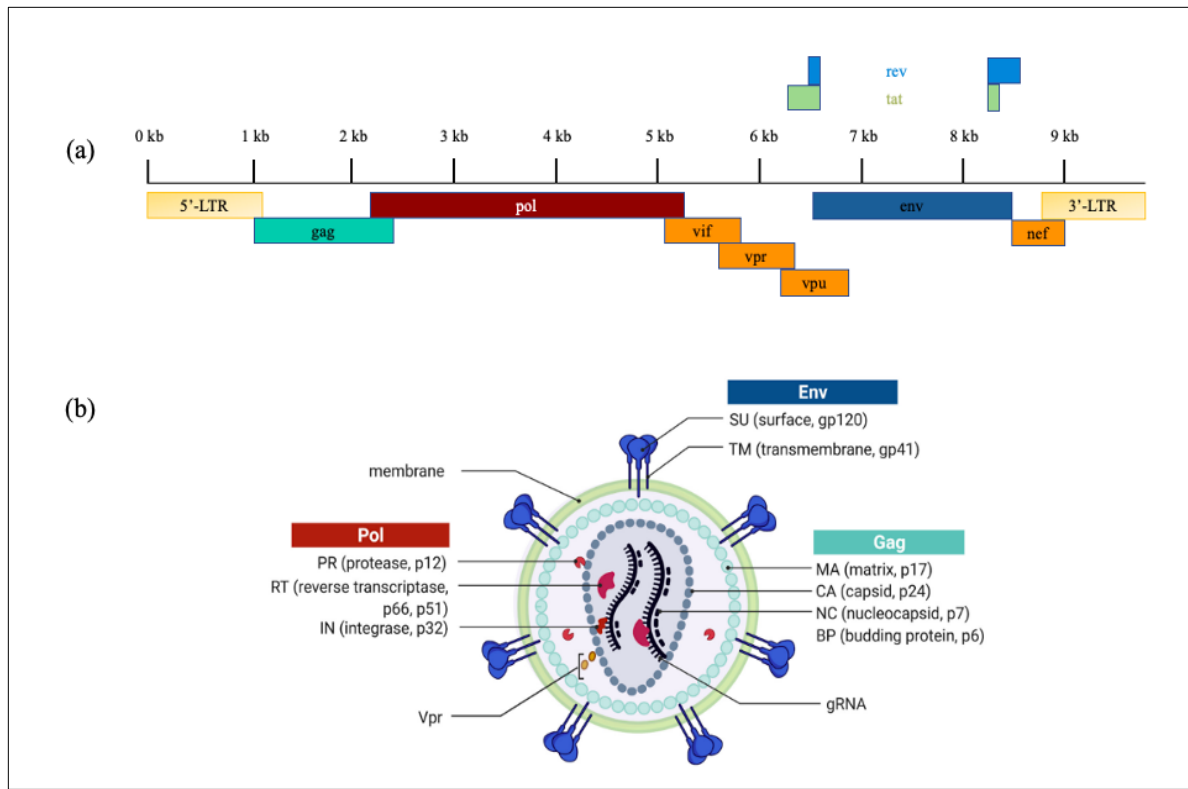


Figure 1.1: (a) HIV-1 genome (adapted from^{12, 13}) and (b) HIV-1 mature virion.¹²

(a) RNA genome (gRNA) of the HIV-1 virus is ~9 kb encoding 15 proteins required for replication and assembly to infect the host cells.

(b) The mature virion contains two viral gRNAs protected mainly by three main layers where the outermost layer is a lipid bilayer membrane followed by a matrix (MA, p17) and capsid (CA, p24). Envelope (Env) proteins are on the host-derived membrane surface and pol proteins play vital roles in reverse transcription and the virion maturation process.

Reprinted with permission from Heuvel et.al (2022). Copyright 2022 by the authors. This figure is from an open-access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license. The copyright page is attached at the end of the thesis.

1.2. HIV-1 Replication Cycle

HIV-1 virus is receptor-dependent and requires a host cell to copy its genetic information to make new virus particles. The replication cycle of HIV-1 (figure 1.2) consists of two stages; early and late. The binding of the virion to the host cell and integration of the viral genome is part of the early stage, and the post-integration and viral maturation occur in the late stage.¹²

In the early stage of the replication cycle, HIV-1 binds to CD4 and CCR5 or CXCR4 chemokine co-receptors of the host cell via the glycoprotein complex of the virion (Env). The binding triggers the conformational changes at the receptor binding site by forming a stable complex allowing the viral core to enter into the cytoplasm of the host cell. The capsid (CA) core transports along the microtubules of the cell to the nuclear capsid complex (NPC) for entering the nucleus. Uncoating and reverse-transcription of viral genome initiate within the core and complete after importing into the nucleus. As the capsid core disassembles, the HIV-1 genome and proteins essential for infection are released as a part of the pre-integration complex (PIC). The reverse transcription of the viral gRNA to proviral dsDNA is a vital step in the replication cycle because the error-prone RT and template-switching activities are responsible for the genetic variability of HIV.¹²

Integration of viral transcripts into the transcriptional active sites of the host DNA is a permanent process making the viral genes to be part of the infected host cell. After integration, the proviral dsDNA either remains transcriptional silencing to enter latency or initiates transcription to produce a new virion. During the transcription of the provirus, the binding of Tat and the activated positive transcription elongation factor b kinase (P-TEFb) with TAR RNA elevates the gRNA's production. The splicing of full-length gRNA into mRNAs enables protein expressions

of all viral proteins. The viral transcripts and mRNAs are then exported from the nucleus for translation in the cytoplasm. After translation of regulatory proteins Rev and Tat, and accessory protein Nef they are imported back into the nucleus to aid transcription phases of the provirus. The translated Gag-Pol, Gag, and Env proteins, and other accessory proteins such as Vif, Vpu, and Vpr are remained in the cytoplasm for the production of new virions.¹²

In the late stage of the replication cycle, the assembly of two virial gRNAs, structural Gag precursor protein, and other regulatory proteins mediates the packaging toward the plasma membrane for virion budding. The immature virion is then released into the outer membrane. Viral maturation is the final step of the viral life cycle where viral protease cleaves the Gag polyprotein into functional proteins to form the viral capsid to prepare for infection of the next host cell and initiate the viral replication cycle again.¹³

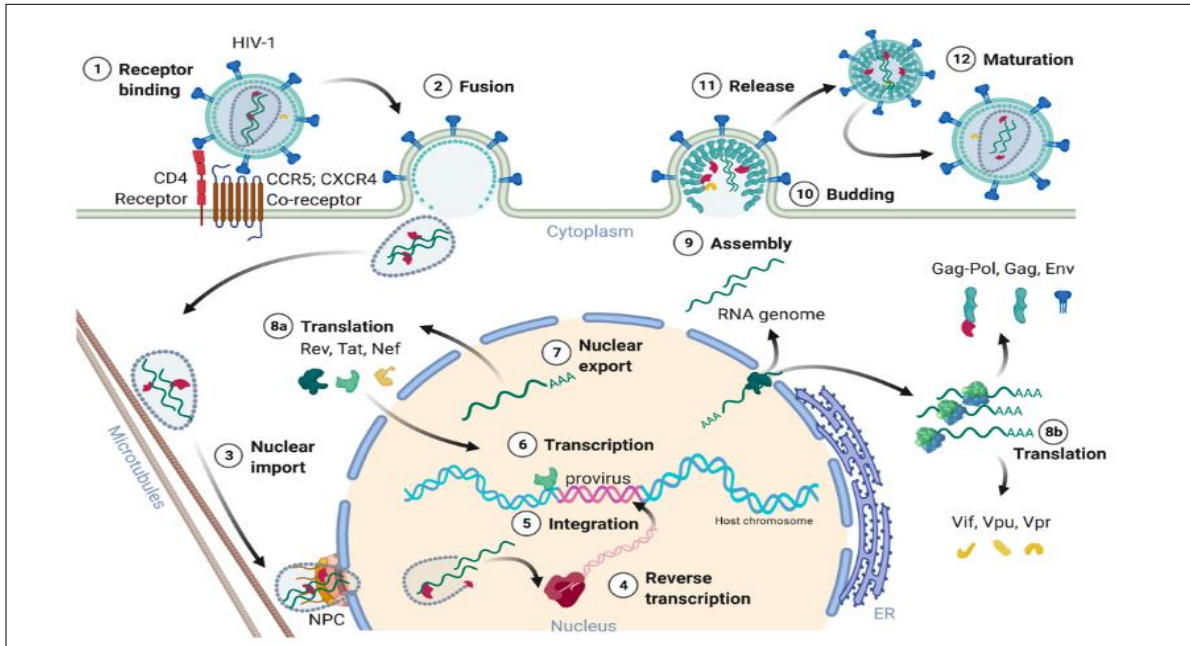


Figure 1.2: The Replication Cycle of the HIV-1 Virus.¹²

The viral replication cycle begins with the receptor binding of the HIV-1 virion and the host cell. Then the capsid core of the virion is released into the cytoplasm to transport to the nuclear pore complex (NPC) utilizing the microtubules. Uncoating and reverse transcription are initiated outside of the nucleus and completed after the core passes through the pore into the nucleus. The integration of provirus into the host dsDNA is facilitated by the integrase and other co-factors. The transcription of proviral DNA is mediated by the upregulation of Tat. mRNAs for different proteins are then exported into the cytoplasm for translation. Two viral gRNAs, Gag, Pol, Env, and Vpr are assembled at the host cell membrane for budding and releasing into the outer membrane. The maturation of the virion is the last step in the replication cycle where protease cleaves gag polyproteins into functional proteins. Reprinted with permission from Heuvel et.al (2022). Copyright 2022 by the authors. This figure is from an open-access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license. The copyright page is attached at the end of the thesis.

1.3. New Targets: HIV-1 Nucleocapsid protein (NCp7) and Trans-activator for Transcription (Tat) for anti-retroviral drug therapies

HIV-1 is a retrovirus whose genetic information is carried by RNA instead of DNA that attacks CD4 T cells (immune cells) and leads to AIDS. All retroviruses consist of nine overlapping open reading frames (ORFs)¹⁴ (Figure 1.1a) where three of these encode group-specific antigen (*Gag*), *polymerase (Pol)*, and envelope (*Env*) in conserved 5' to 3' direction. These polyproteins are later processed to generate viral structural proteins and enzymes. A *Gag* polyprotein is produced during the host cell's last stages of the infectious cycle and plays a vital role in encapsulating two copies of the unspliced viral genome.¹⁵ During the maturation of the virus, the viral protease cleaves the *gag* polyprotein into the matrix (MA), capsid (CA), and nucleocapsid (NC) proteins, and the virus rearranges to develop infectious particles.¹⁶ The *Pol* polyprotein is spliced into three enzymes, i.e., RT, which is responsible for transcribing viral RNA to DNA, IN, which is responsible for facilitating the entry of viral DNA into the host genome and PR, which is responsible for proteolytic cleavage of the viral gene products into functional proteins. Parts of the *Env* polyprotein act on the outer viral surface to bind to host cell receptors and trigger entry into the host cell.¹⁷

Due to the highly conserved nature of the zinc fingers in NCp7¹⁸, the requirement for NC domains in *Gag* to select and package genomic RNA, and NC's involvement in other parts of the HIV-1 life cycle, drugs that target HIV-1 NCp7 are attractive for new anti-retroviral therapeutics.

1.3.1. NCp7- SL3 Interaction

A mature NCp7 protein, shown in Figure 1.3a, is a very small basic protein (a 55 amino acid sequence) containing two retroviral zinc finger domains (Cys-X₂-Cys-X₄-His-X₄-Cys where X is a variable amino acid).¹⁹ NCp7 is equivalent to the NC protein domain; hence, it is a beneficial model for anti-retroviral studies.²⁰⁻²³ CCHC-type zinc finger motifs from the conserved region of most retroviral NC proteins are essential for viral replication, encapsulation, and genome recognition.¹⁹ NCp7 is also utilized as a nucleic acid chaperone to stabilize the dimeric linkage between the genomic RNA molecules in the virus particle.²⁴⁻²⁶

The interactions between NCp7 and the region of the unspliced viral RNA (Ψ site), which has four stem-loop structures (SL1- SL4) (Figure 1.3b), provide the recognition of the HIV-1 genome for packaging full-length genomic RNA into budding virus particles.²⁷⁻³³ SL3, a 20 nucleotide RNA sequence located just upstream of the *gag* initiation codon (AUG), is highly conserved in various strains of HIV-1.²⁸ It binds tightly to NCp7 and the *gag* polyprotein, and functions to direct the recognition and packaging of unspliced genomic RNA into viral particles.²⁹

Our work on 3-segment nucleic acid switches is described in Chapters 2-4. These switches offer a novel approach to high throughput screening (HTS) of drug candidates to block Ψ -RNA: NC interactions and inhibit the packaging of genomic RNA into new virus particles. General design principles for 3-segment bistable switches are described, so these switches could be used for HTS against other targets, and in other potential applications, including nanostructures and nanomachines. The switches described here are designed to be highly fluorescent, ON, in the absence of the target, and then to dim and switch OFF as the target binds the OFF-state upon increasing the concentration of the target. Switches specific for binding NCp7 were designed, built, and tested, establishing proof-of-principle for the operation of ON-OFF 3-segment switches.

This work involved a close collaboration with Dr. Raghu Iyer in the Borer lab, who assessed thermodynamic contributions to the equilibrium between the two states of the switch. Dr. Iyer's dissertation describes the optimization of chemical synthesis. He also designed, built, and tested switches that were OFF in the absence of NCp7 for the diagnostic assay, showing proof of principle for the operation of OFF-ON three-segment switches.³⁴ The current thesis work developed ligation as a method to greatly simplify purification.

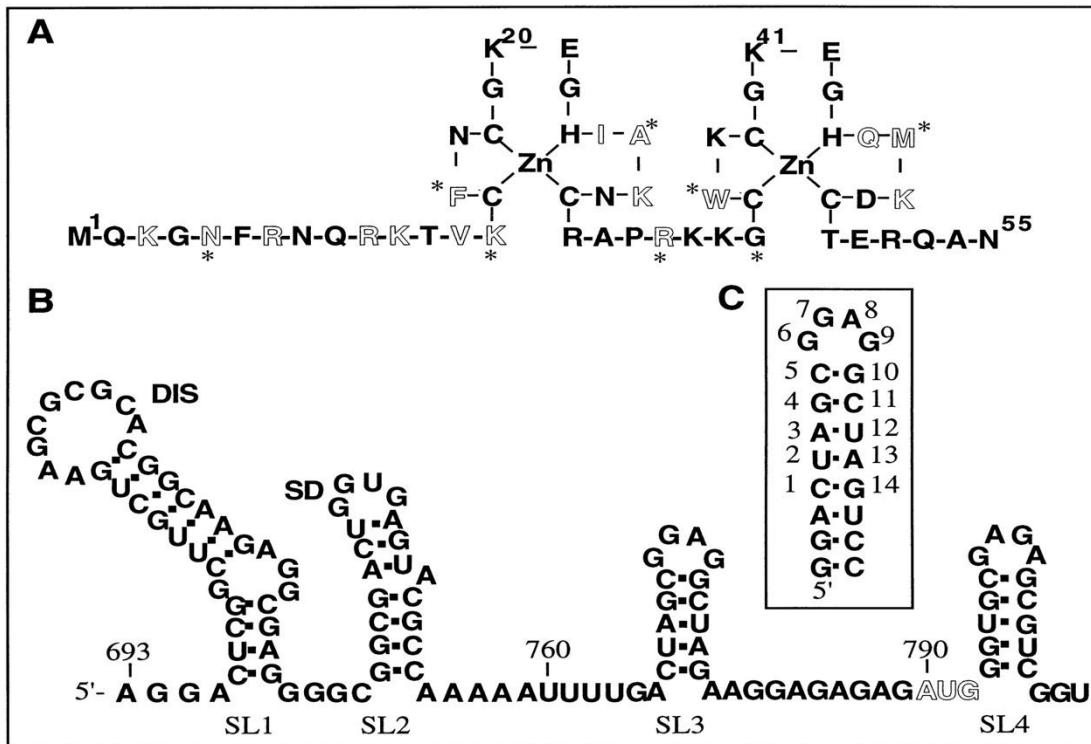


Figure 1.3: Structure of NCp7 and the region of the unspliced viral RNA (Ψ site).

(A) The structure of mature NCp7 protein consisting of two zinc finger (ZF) motifs. ZF domains chelate zinc ions with high affinity to enable proper folding for NCp7 functions.

(B) Nucleotide HIV-1 sequence and secondary structure of HIV-1 Ψ -sequences, which has four stem-loop structures (SL1- SL4).

(C) SL3 loop serves as the primary recognition site and is highly conserved across different HIV-1 virus subtypes.

De Guzman et al., *SCIENCE*, 16 Jan 1998, Vol 279, Issue 5349, pp. 384-388 DOI: 10.1126/science.279.5349.384. Reprinted with permission from AAAS.

1.3.2. TAR-CycT1-Tat Interaction

In the early stages of the HIV replication cycle, the proviral genome is reverse transcribed, transported into the nucleus, and then integrated into the host cell genome.³⁵ In the process of virus assembly, the transcription from the viral long terminal repeat (LTR) of viral RNA requires an HIV regulatory protein known as transcriptional activator protein (Tat) to promote high levels of transcription elongation from the integrated provirus.³⁶ The presence of Tat activates the RNA polymerase II complex found downstream of the promoter for expression and replication.³⁷

Tat protein (Figure 1.4a) has two zinc finger domains where Cysteine 261 (C261) from CycT1 is necessary to complete the second zinc finger in Tat.³⁸ Tat interacts with this host protein, the cyclin T1 (CycT1) subunit of the positive transcription elongation factor complex (P-TEFb). The interaction occurs through Tat's cysteine-rich activation domain and binds to the bulge region of Trans-activation response (TAR) RNA via its arginine-rich motif (ARM). Tat appears to interact with the residues in the carboxyl-terminal of the CycT1 domain especially the basic residues, Arginine 251 (R251) and Arginine 254 (R254), to stabilize the CycT1-Tat-TAR RNA complex.³

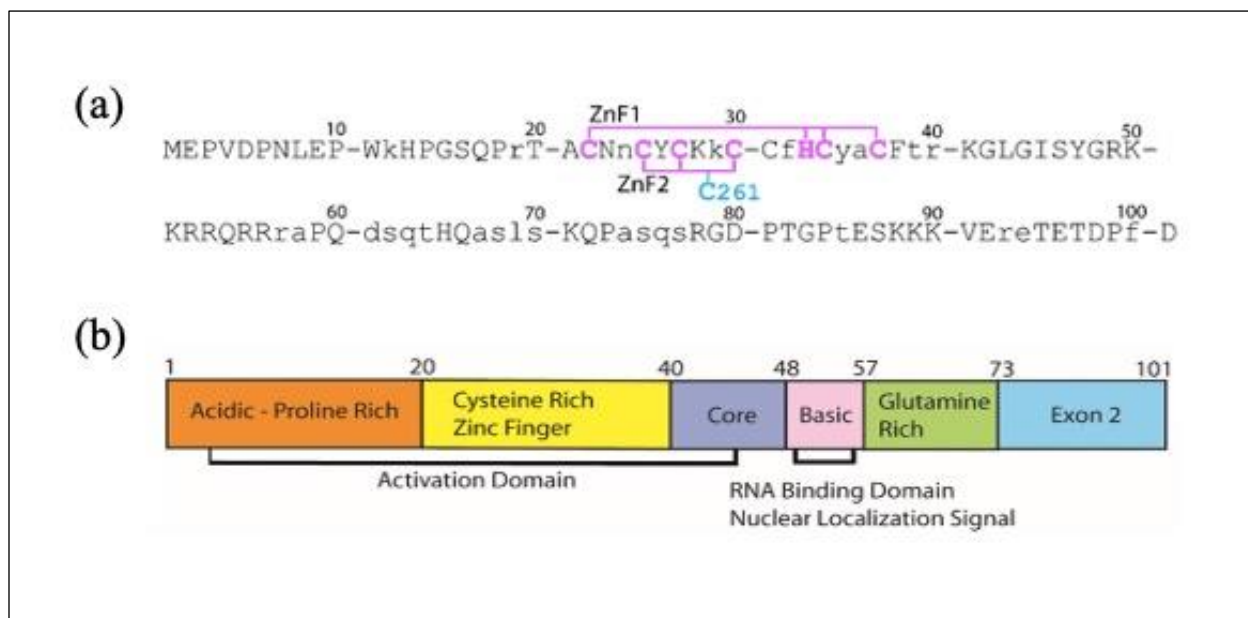


Figure 1.4: HIV-1 Tat-101 protein.

(a) Tat-101 protein (Prepared by Dr. Borer) and (b) Visual organization of functional domain.³⁹

(a) Tat-101 protein has two zinc finger domains where Cysteine 261 (C261) from CycT1 is necessary to complete the second zinc finger in Tat.^{38,40} The residues denoted in lower-case letters differ by $\geq 25\%$ in common HIV sub-types.⁴¹

(b) The functional domains of HIV-1 Tat(1-101) protein, where Exon 2 contains a splicing silencer that is not required for transcription but is required for in vivo activity.

Reprinted with permission from Asamitsu et.al. (2018). Copyright 2018 by the authors. This figure is from an open-access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license. The copyright page is attached at the end of the thesis.

Tat protein binds to a 31-base stem-loop structure (Figure 1.5) that is a model for the 57mer consensus TAR RNA, minor variations of which are located at the 5' end of all newly started HIV-1 transcripts. In deletion studies of TAR RNA, the region from nucleotide 19-42 containing a six-nucleotide loop and a three-nucleotide pyrimidine bulge was sufficient for tight binding with the CycT1-Tat protein.^{38, 42, 43} In the absence of Tat, neither CycT1 nor P-TEFb binds TAR RNA. Tat binds to TAR at a lower affinity in the absence of CycT1. Therefore, the formation of the TAR-CycT1-Tat complex is vital to stimulate the transcription elongation from the HIV-1 provirus.⁴⁴

Drug discovery targeting the TAR-Tat interaction has had limited success largely because the role of Cyclin T1 has not been given appropriate consideration. Also, early studies of Tat's structure were compromised due to solubility and folding issues that make this protein difficult to work with. Figure 1.4(a) provides an easy explanation of why these problems exist – the two overlapping zinc fingers involve seven residues from Tat and one from CycT1. In the absence of CycT1, the cysteines of Tat may crosslink in undesirable forms, including Tat dimers and higher aggregates.

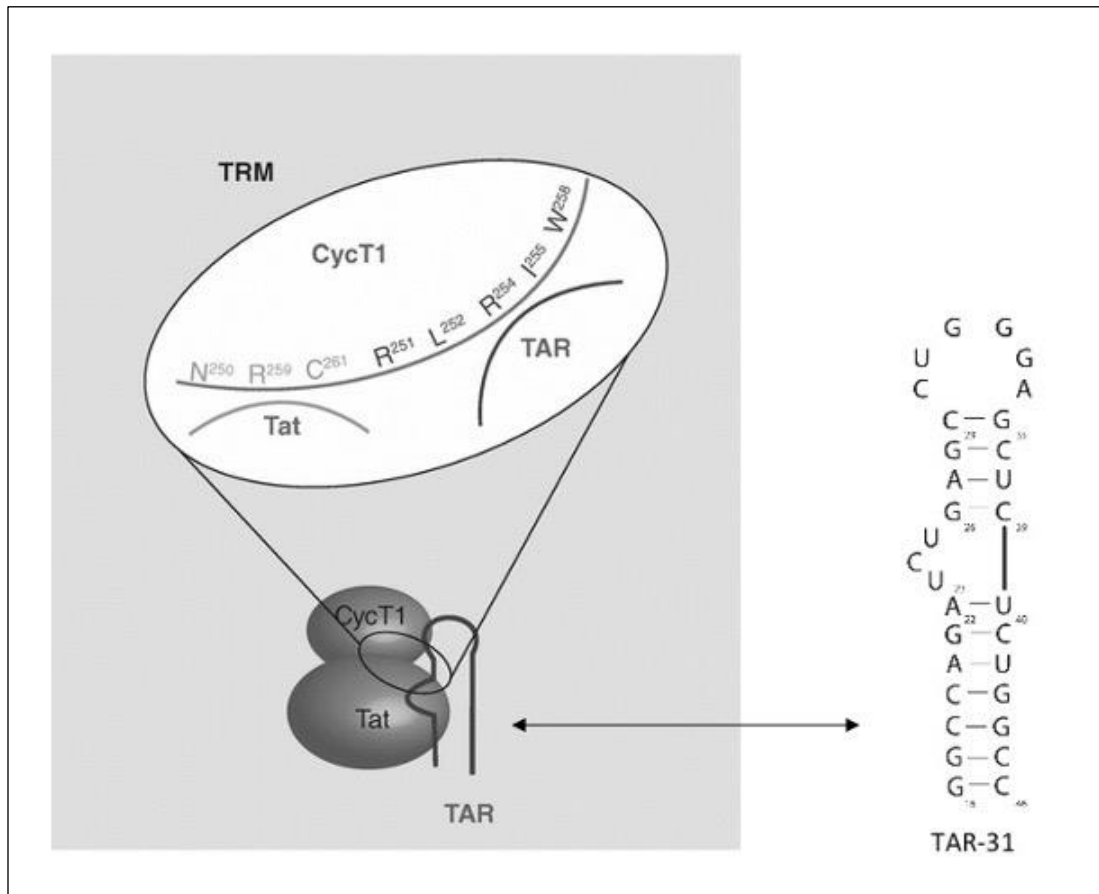


Figure 1.5: TAR-CycT1-Tat Interaction.

Wild-type HIV 1 TAR-31nt stem-loop structure (Right) and CycT1-Tat-TAR complex (Left). (adapted from⁴⁵) Tat interacts with a host protein, the cyclin T1 (CycT1) subunit of the positive transcription elongation factor complex (P-TEFb). The interaction occurs through Tat's cysteine-rich activation domain and binds to the bulge region of Trans-activation response (TAR) RNA via its arginine-rich motif (ARM).

Reprinted with permission from Future Science for CycT1-Tat-TAR complex (Left). The copyright page is attached at the end of the thesis.

Chapter 5 describes our work to design, overexpress, and test a CycT1-Tat fusion protein (F1) that binds TAR 31. Figure 5.1a shows a CycT1-Tat chimera that includes the key C261 from Tat.³ A flexible linker allows this cysteine to complete a proper set of two interlocking zinc fingers as in Figure 1.4(a). Our work builds on this previous chimera that was shown to up-regulate transcription of HIV-1 transcripts in human cells. It has a 32aa fragment of CycT1 (including C261) connected through a 25aa spacer to Tat. This chimera was not suitable for cloning and overexpression in *E. coli*, so it was redesigned by Dr. Collin Fischer in the Borer lab.⁴⁶ However, this revised chimera lacked R251 and R254, which are included in our new design discussed in Chapter 5. The design also simplifies cloning and was successful in the purification of a highly soluble CycT1-Tat fusion protein (which we refer to as, F1).

Results presented in Chapter 5 show that our new CycT1-Tat chimera binds to both the hairpin and bulge loops of TAR31 with high affinity. The TAR31:CycT1-Tat complex should be a good model for the RNA-P-TEFb interaction, and can easily be adapted for high throughput screening of drug candidates to disrupt the formation of the complex. Our 144aa chimera with TAR31 should also be small enough for 3D structure determination by NMR or x-ray diffraction. A 3D structure would be helpful in drug design, and in understanding the mechanism of elongation of HIV-1 RNA transcripts by P-TEFb.

The native TAR-Tat interaction in P-TEFb results in a roughly 1000-fold upregulation of transcription for viral mRNA in HIV-1 infected cells.³ Drugs that block TAR from binding to Tat could gain wide usefulness in AIDS therapy. One important unmet need is to deal with HIV-1 infections that emerge from otherwise normal cells that have the provirus integrated into their DNA. Rapid production of viral mRNA and genomic RNA from the provirus can cause a new

burst of infectious virus in an AIDS patient. Anti-Tat drugs could play a key role in preventing and managing these outbreaks.

1.4. Aims of the Projects

Part I: Development of a Novel Bistable DNA Sensor for Anti-HIV Drug Discovery

The goals of this work are to design and create a simple, low-cost, mix-and-measure NCp7-responsive DNA switch for High Throughput Screens (HTS) of drug candidates. This work establishes the principles for the design of a switch sensor for any target of interest; these sensors can be used in an assay with no immobilization and washing steps as in many traditional assays. A 3-segment switch molecule contains probe (P), cover (C), and toggle (T) segments, and each segment can form individual secondary structures. In the examples described here, P is an aptamer that selectively and specifically binds the HIV-1 nucleocapsid protein (NCp7), and each segment can form a hairpin stem loop. The specific aims of this work are:

- 1) Design branched switch molecules consisting of three unique segments (70-71mer total) using V-OMP software.
- 2) Synthesize and optimize the synthesis of DNA oligonucleotides with a fluorescent dye (CY3) and quencher (DABCYL).
- 3) Analyze and purify the full-length switches using HPLC (IE and RP) and denaturing polyacrylamide gel electrophoresis.
- 4) Evaluate switch candidates by measuring the ON/OFF contrast ratio in the presence and absence of NCp7 and determine thermodynamic equilibrium constants: $K_d(\text{app})$ for binding NCp7, and K_1 for the ON to OFF equilibrium in the absence of protein.

Part II: Re-engineering of Recombinant human Cyclin T1-Tat Chimera with SUMO Fusion in *E. coli*

The goals of this project are to optimize the length of CycT1 and Tat segments from the previously expressed fusion protein (His6-MBP-CycT1 (257-280)-Tat-101), which was studied by Dr. Fischer in Borer lab for High Throughput Screening (HTS) and anti-Tat drug discovery. The specific aims of this work are:

- 1) Prepare variations of hCycT1-Tat Chimera plasmids by using ChampionTM pET SUMO Protein Expression System
- 2) Perform cloning into the pET SUMO vector, transforming in the Mach 1TM-T1^R competent cells, isolating plasmid DNA using Qiagen plasmid purification kit, and analyze positive clones using Sanger sequencing
- 3) Perform pilot expressions using BL21 (DE3) *E. coli*
- 4) Overexpress the recombinant proteins in *E. coli* and purify His₆-SUMO tagged hCycT1-Tat proteins
- 5) Analyze hCycT1-Tat proteins using SDS-PAGE and MALDI-TOF.
- 6) Perform Electrophoretic mobility shift assay (EMSA) to semi-quantify the TAR-CycT1-Tat complex.

Part I: Development of a Novel Bistable DNA Sensor for Anti-HIV
Drug Discovery

Chapter 2: A Bistable aptamer-based DNA switch sensor for Drug Discovery

Chapter Summary

Aptamers are small nucleic acids that bind to their targets with high affinity and specificity; they may be considered chemical antibodies as they can be readily made on standard DNA/RNA synthesizers. There are advantages of aptamers over antibodies in a wide range of applications such as biomarker discovery, *in vitro* diagnosis, *in vivo* imaging, and targeted therapy. In this work, the design and development of an aptamer biosensor for drug discovery are described to establish proof of principle for an unprecedented class of 3-segment switch sensors. The 3-segment design can be applied to any target of interest.

A 3-segment switch (Figure 2a) is composed of **Probe**, **Cover**, and **Toggle** segments that should have a low development cost for a wide variety of interesting targets. This unimolecular reversible switch can be used in simple mix-and-measure assays to screen thousands to millions of drug candidates that bind to the chosen target, which in our case is NCp7. In switch design, we incorporate a fluorophore (CY3) on the 5'-end of the **Toggle** segment and a quencher (DABCYL) on the 3'-end of the **Cover** segment; this allows fluorescence to be used to quantify the populations of the two forms of the switch.

This Chapter describes the advantages of aptamers over antibodies to use in a simple mix and measure assay without additional wash steps (section 2.1). The design of the 3-segment switch, the calculation for thermodynamic parameters, and extinction coefficient correction for a branched molecule are described in section 2.2.

2.1. Aptamer

Aptamers are single-stranded RNA, DNA, or modified DNA oligonucleotides that have high affinity and specificity for their targets, similar to antibodies. The term, aptamer, is derived from the Latin *aptus* (to fit), and its unique property is that it folds into a specific three-dimensional (3D) structure to interact with targets selectively, having dissociation constants in the low pico- to the nano-molar range.⁴⁷ Unlike antibodies, the size, ease of manufacturing with low variability between batch to batch, non-immunogenicity, thermal stability, and modifiability make aptamers attractive for a variety of targets such as small ions, drugs, toxins, peptides, proteins, viruses, bacteria, cells, and tissues.⁴⁸

Aptamers are small in size with typical molecular weights of (8-25kDa) compared to antibodies (~150kDa), simplifying their application in the clinic and large batch synthesis. They are the equivalent of monoclonal antibodies in that they are a unique molecular species, distinct from a mixture, as in polyclonal antibodies raised against a target in a rabbit or other animal. Moreover, aptamers can be synthesized with a high yield and purified to >95% purity on a standard solid-phase DNA/RNA synthesizer. Due to their oligonucleotide properties, aptamers are usually non-immunogenic while antibodies can elicit a strong immune response *in vivo*.⁴⁹

Unlike antibodies which lose their activity permanently at a high temperature (95°C), aptamers can refold into their stable conformation after cooling to room temperature. They have a longer shelf life and can be shipped as lyophilized powders because of their thermal stability. A major advantage of aptamers over antibodies is that chemically modified aptamers can further extend their stability (higher affinity and specificity).⁵⁰ The first modified RNA aptamer, Macugen, was approved by the US Food and Drug Administration (FDA) to target vascular

endothelial growth factor (VEGF).⁵¹ Also, the widespread use of modified mRNA fragments in vaccines against Covid-19 has shown them to cause an undesirable immune reaction in only a tiny fraction of humans that received injections.

The Discovery of Aptamers for proteins of interest is traditionally carried out through *in vitro* selection and amplification in a procedure known as Systematic Evolution of Ligands by Exponential enrichment (SELEX).^{52, 53} In this approach, a randomized/ naïve library pool is exposed to a protein of interest over multiple rounds of selection so that aptamer candidates with high affinity and specificity will dominate in the final pool. However, this process is high-cost and labor-intensive. Advances in aptamer technology to make the discovery of aptamers more efficient and less time-consuming have been explored in recent years showing aptamers as a high interest in the research field.

2.1.1. Aptamer-based Biosensors

To date, sensing and molecular detection has heavily relied on monoclonal antibodies as a gold standard. However, developing antibody biosensors is difficult because of structural complexity, large size, high immunogenicity, and high-cost production. In addition, they often require low-temperature storage and often vary in reproducibility from batch to batch. Therefore, aptamers with high sensitivity and selectivity are attractive biosensors for a variety of targets in many fields of application such as drug discovery, medicine, and environmental or food analytics. Aptamer-based biosensors developed by several research groups have been published for a broad range of targets.^{54,55,56}

Among them, fluorescent biosensors offer great advantages in structure-switching by use of the fluorescence of a dye or the fluorescence resonance energy transfer (FRET) between two

dyes. As a result of a conformational change in a fluorescent biosensor upon target binding, an increase or decrease in fluorescence intensity or polarization can be detected and the binding affinity can be calculated. Aptamer-based fluorescent biosensors are established by using existing aptamers that have been previously validated for affinity and specificity. The challenges in engineering reversible switch sensors are sequence design, the fluorophore and quencher selection, the optimal thermodynamic equilibrium between bound and unbound states, and experimental validation.

An RNA and DNA chimeric switch sensor,⁵⁷ a FRET sensor (Figure 2.1), was designed previously in Borer Lab using a modified SL3-RNA that interacts with the mature HIV-1 NCp7 protein with high affinity and specificity. This switch sensor called an AlloSwitch⁵⁸ consists of two thermodynamically stable conformations, where one conformation has 170 times higher affinity to NCp7 protein than the other conformation.⁵⁷ The FRET pair of fluorescein and Dabsyl are used to detect the conformational change upon target binding. The AlloSwitch is predominantly in the ON state (H-form), and the equilibrium shift to the OFF state (O-form) happens in the presence of NC. AlloSwitch reverts to the ON state if high-affinity competitors (drug candidates) disrupt the interaction between AlloSwitch in O-form and NC.⁵⁷

The reversible switch design of AlloSwitch has great potential for HTS assay of anti-NC drug discovery. However, it has several disadvantages including a low contrast ratio (~2) between the ON and OFF forms; background fluorescence from many drug candidates interfering with the sensor readout due to labeling with the switch FAM (green fluorescence) at 5' end, and that probe strands might encounter kinetic traps in switching between the two stable states of the sensor. Hence, a novel 3-segment DNA switch sensor was proposed to overcome these problems.

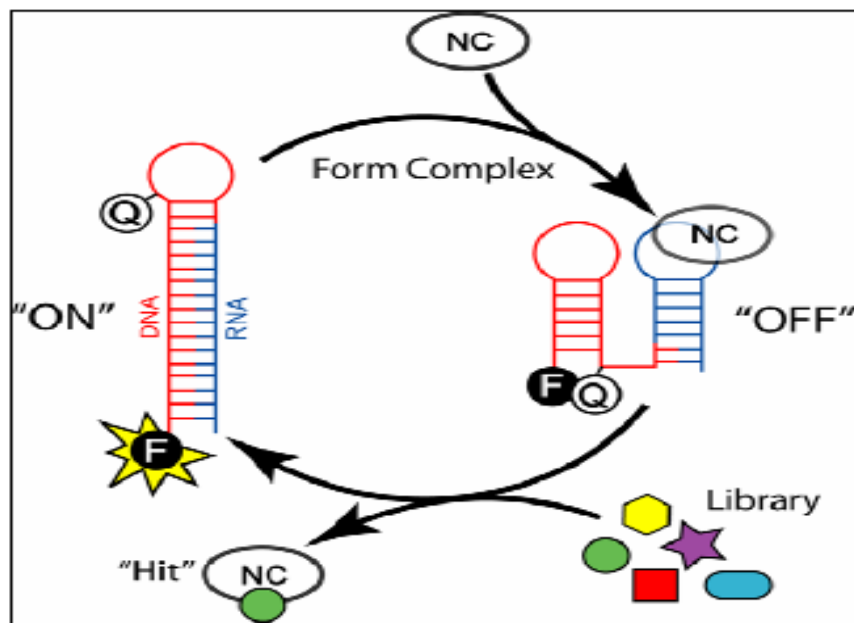


Figure 2.1: An illustration of the AlloSwitch.⁵⁸

In the absence of NC, the AlloSwitch is in the ON state (left). As [NC] increases, the AlloSwitch forms a complex with NC, which shifts the equilibrium to the OFF state (right). If there are high-affinity competitors in the library of drug candidates, they will compete with NC to bind to the AlloSwitch and revert it to the ON state.

2.2. 3-segment DNA switch sensor for drug discovery

In the design of the HTS switch sensor, three unique DNA segments can form individual hairpins, the **Probe (P)**, the **Cover (C)**, and the **Toggle (T)**. The DNA **Probe** segment is a DNA aptamer that similarly binds to NCp7 as does SL3 RNA; the aptamer was discovered by Dr. M. McPike in our lab. The TGTGGT bases in the DNA **Probe** segment have a high resemblance to the GGAG bases in the conserved region of SL3 RNA. This aptamer⁵⁹ was selected using a microarray-based screening approach of NCp7 against a library of DNA hairpin loops (3-6 nucleotides of all possible recognition elements); the DNA aptamer has a low nanomolar affinity for NCp7. Hence, design considerations for cover and toggle will avoid unpaired regions with two or more unpaired G or T- bases near each other, and will instead have an abundance of C and A bases in the unpaired regions. The segments of this novel bistable, unimolecular switch are joined by a 5-methyl-dC brancher (abbreviated as ©), which has a standard phosphoramidite on O3', dimethoxytrityl on O5', and the third branch connects via cytosine-N4, where a -C₆-O- linker is protected by a levulinyl group (Figure 2.2; deprotection with hydrazine hydrate releases N4-C₆-OH to synthesize the third segment).

This project designed sequences of the segments to make an NCp7-responsive switch and measured fluorescence emission to characterize the state of the switch. This is done by adjusting the sequences of the segments such that the ON form is favored over OFF in the absence of the target (the equilibrium constant $K_1 < 1$). The fluorescence intensity decreases as increasing [NCp7] forces the switch into the target-bound, OFF form (Figure 2.2(b)). In the presence of NCp7, the fluorescence will increase if drug candidates bind to the aptamer site on the protein; this releases the switch and it reverts to the ON form.

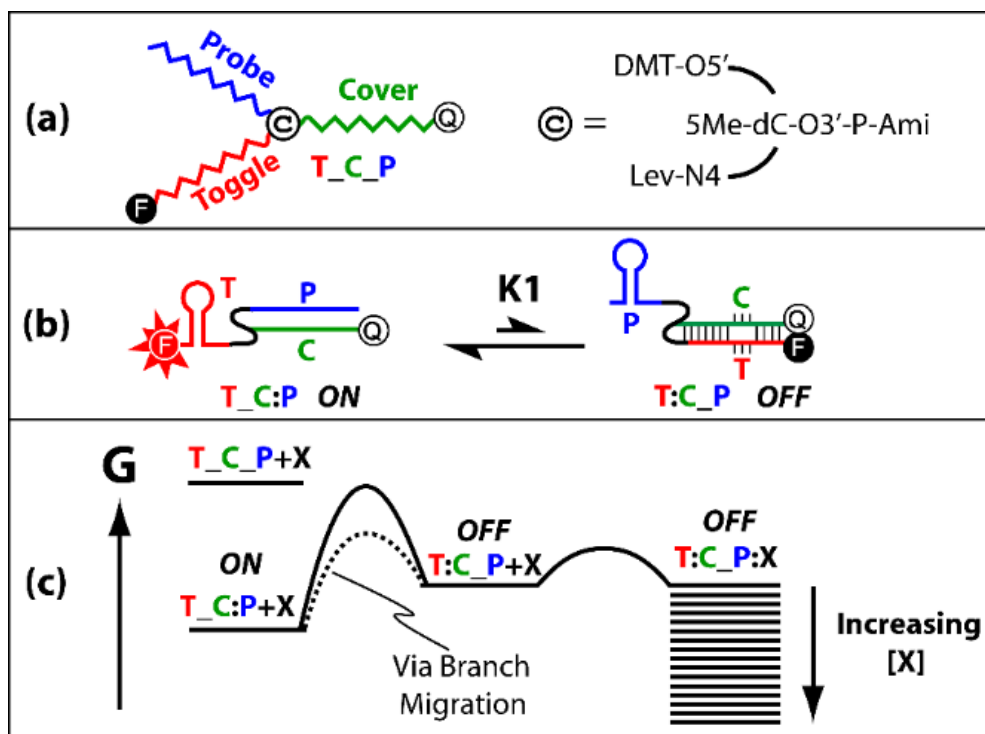


Figure 2.2: Design of a branched switch molecule.

(a) 3-segment DNA switch illustrating the **Probe**, **Toggle**, and **Cover** segments as unstructured single strands joined at a 5Me-dC brancher (©). Figure prepared by Dr. Borer.

(b) ON and OFF forms of the switch. The ON state has a fluorophore and quencher separated in space, while OFF has them adjacent at the **Cover** and **Toggle** termini. Parallel lines denote a stem with extensive base pairing. Sequences are adjusted to make the ON conformation more stable than OFF as characterized by the equilibrium constant K_1 , which can be estimated from the calculated ΔG° value of each state. Figure prepared by Dr. Borer.

(c) A free energy diagram illustrating conversion of ON to OFF form with and without branch migration. X is the target that binds to the aptamer in **Probe**. Figure prepared by Dr. Borer.

For a reversible switch sensor, the toggle strand (containing a 5'-fluorophore) competes with the probe strand to bind to the cover strand (containing a 3'-quencher). For the desired application of the switch, K_1 should be in the range of 0.001-0.1 so that the novel switch will have a **T** hairpin and **P:C** duplex (ON) in a population of 90-99.99% and a **P** hairpin and **T:C** duplex (OFF) around 0.01-10% of the total population. When the target, NCp7, is present at a low nanomolar concentration, it will bind to the open probe and shift the equilibrium in favor of the OFF form. If a drug candidate, which binds with moderate to high affinity to the protein, is present at $\sim 10\mu\text{M}$, it will compete with the probe for the NCp7 binding site and the switch equilibrium will shift back to the ON form.

In the free energy diagram shown in Figure 2.2(c), **T_C:P** indicates that the **Toggle** is free, and the **C:P** duplex is in the ON form on the left side of the diagram. The OFF form in the middle of the diagram indicates that the target is bound to the switch. Branch migration, where only a few base pairs are broken at any time, can hasten equilibration. **T_C:P** is designed to have more negative free energy than that **T:C_P**, where the **Probe** is free and available to bind the target. Therefore, the equilibrium will favor the OFF form as the concentration of the target is increased as shown on the right side of the diagram. The reference state for free energy is shown in the upper left corner. This state has each segment unstructured (random coils) and connected by the brancher; the target, X, is in its native, but unbound state.

Proper functioning of the switch requires that at 25°C the free energy difference between the two conformations ($\Delta\Delta G^\circ$) should be between 1.36-2.73 kcal/mol ($2.303RT = 1.364$ kcal/mol). The free energy of formation for the ON and OFF states can be calculated using tabulated free energies for base-pair stacking in stems and the various kinds of unpaired loops.⁶⁰ In early work in our lab, an additional -2 kcal/mol was included for the OFF state to account for the (favorable)

free energy of bringing the fluorophore and quencher into close proximity [R. Iyer dissertation]. The Visual-Oligonucleotide Modeling Platform (V-OMP) software from the company, DNA Software, now includes free energy contributions for bringing various dyes together, along with the usual free energies for forming base-paired stems and loops. As detailed in Chapter 4, this version of V-OMP was used to evaluate sequence alterations in potential switch molecules to obtain $\Delta\Delta G^\circ$ between -1.4 to -2.8 kcal/mol (in favor of the ON form).

2.2.1. Calculations for the theoretical equilibrium constant

The sequences of Cover and Toggle can be adjusted to obtain an equilibrium constant, $K_1 = 0.1$ to 0.001 using free energy rules for DNA.⁶⁰ The Visual-Oligonucleotide modeling platform (V-OMP) developed by DNA Software, Inc. is a very useful package that has thermodynamic data to predict free energies to predict possible DNA secondary structures under various salt conditions and using the current DNA thermodynamic database. The program also predicts and ranks the populations of alternative structures. In the prediction of potential switches using V-OMP, changes in the sequence of **T:C** or **P:C** duplexes and **T-** or **P-**hairpins give different free energy contributions relative to the unstacked single strands. Adding mismatches, hairpins, bulges, internal loops, etc. will lower the stability of the **P: C** duplex in the ON state. In contrast, adding base pairs will increase the stability of the **T:C** duplex to increase the stability of the OFF state.

Watson-Crick NN Parameters

By following the strict pairing rules, Watson-Crick base-pairing interactions allow both DNA and RNA to have stable secondary structures. With the number of possible secondary structures, 1.8^N , where N is the sequence length⁶¹, dynamics programming algorithms⁶² can predict accurate DNA secondary structure. The dataset of the Watson-Crick NN parameters, the unified NN, in 1M NaCl, was collected from various labs by solving for 12 unknowns (10 NN propagation parameters, 1 initiation parameter, and 1 correction for terminal AT pairs) of 108 sequences.⁶³ This provides the foundation to measure the thermodynamics of a motif in a larger sequence to create the thermodynamic database by subtracting the Watson-Crick contribution.⁶³ The following equation shows how to calculate the total ΔG°_{37} using the unified NN parameters:

$$\Delta G^\circ_{37} (\text{total}) = \sum_i n_i \Delta G^\circ_{37 \text{ init}} + \Delta G^\circ_{\text{init w/AT}} + \Delta G^\circ_{\text{init w/GC}} + \Delta G^\circ_{37 \text{ symmetry}}$$

where $\Delta G^\circ_{37 \text{ init}}$ (kcal/mol) is the standard free-energy changes for all possible Watson-Crick NNs, n_i is the number of the same nearest neighbor, and $\Delta G^\circ_{37 \text{ symmetry}}$ is +0.43 kcal/mol to accounts for the self-complementary duplex.

Sodium Dependence

Thermal stability and folding kinetics of DNA and RNA depend on the presence of metal ions. Different salt conditions must be corrected to predict accurate thermodynamic parameters. From the thermodynamic data for 26 oligonucleotide duplexes dissolved in 0.01M to 0.3 M NaCl, the salt effect is assumed to depend on the length but not on the sequence.⁶³ The following free energy equation is derived from accounting for a variety of salt conditions:

$$\Delta G^\circ_{37} [\text{Na}^+] = \Delta G^\circ_{37} [1\text{M NaCl}] - 0.114 \times N/2 \times \ln [\text{Na}^+]$$

where $[\text{Na}^+]$ is the total concentration of Na^+ ion for the desired solution, and N is the total number of phosphates in the duplex.

V-OMP calculations require the specification of the salt and strand concentrations. The salt environment was set to replicate our usual buffer for measurement of these NCp7-RNA or DNA interactions: (0.205 M NaCl, $1\mu\text{M Zn}^{2+}$; pH: 7). The stability of NCp7-SL3 RNA interactions are strongly salt-dependent⁶⁴, as NCp7 has a formal +9 charge and the SL3 RNA 20mer has a formal charge of -19. Any DNA or RNA interacting with NCp7 will be influenced strongly by salt. Working at ionic strengths of around 0.2 M duplicates near-physiological conditions in the blood (ionic strength ~ 0.18 M disregarding charged macromolecules). As the calculations are computer simulations, we set the concentrations of all strands to the 1M standard states for ΔG° , which also ensures that duplex formation competes effectively with unimolecular hairpin formation. Details of the setup for the calculation depend on details that are left for the specific sequences discussed in Chapter 4.

Chapter 3: Direct Synthesis of Switch and Optimization of Switch Synthesis

Chapter Summary

A unimolecular bistable oligonucleotide, a branched switch that consists of three unique segments, was designed based on an aptamer that selectively and specifically binds the HIV-1 nucleocapsid protein (NCp7). This novel DNA sensor contains cover, toggle, and probe sequences that can form individual hairpins. Without base pairing, the molecule resembles a Y-shaped structure with the three strands joined at a fork (5-Me dC brancher). The probe segment contains an aptamer sequence that binds NCp7 with a high affinity and specificity ($K_d = 16$ nM).

This chapter describes the direct synthesis of a switch molecule using a 3'-DABCYL controlled pore glass (CPG) beads column in section 3.1. A simple branched oligonucleotide is synthesized on CPG beads from a 3' to 5' direction. In the Part I synthesis, cover and probe sequences are synthesized, then the levulinyl protecting group (Lev) from the 5-Me dC brancher is deprotected. The synthesis of the branching sequence, toggle, can then be proceeded.

Due to the low recovery and purity of a directly synthesized switch, the optimization of switch synthesis and purification are described in sections 3.2 and 3.3. The failures in coupling efficiency are greatly reduced by using fresh phosphoramides and increasing coupling time for capping and oxidation steps. Attempts were made for the purification of full-length oligonucleotides utilizing comparing IE and RP HPLC (described in this chapter), but the results were disappointing. We then developed highly successful purification by denaturing polyacrylamide gels (Chapter 4).

The preliminary experiments in Chapter 3 are lengthy and included here as they could be useful in developing commercial-scale synthesis and purification of 3-segment switches in the future. However, many readers might read this chapter through the discussion of Figure 3.2, and then skip to chapter 4, which presents definitive experiments establishing proof-of-principle for ON-OFF switches using the free energy of the NCp7-aptamer interaction to drive structure switching. Returning to Chapter 3 could help to clarify important topics embodied in Figure 3.9 (Models for branched loops), section 3.2.3 (on effects of steric hindrance in solid-phase synthesis of branched Structures), and section 3.2 (on the importance of optimizing synthesizer performance to reduce unwanted side reactions).

3.1. Direct Synthesis of switches using 3'-DABCYL CPG

An automated synthesizer was used to chemically synthesize oligonucleotides sequentially on the solid support (Figure 3.1) with steps to deblock, couple, cap the uncoupled bases, oxidize, and repeat these steps until the last base is added to the strand.⁶⁵ A standard synthesis of DNA oligonucleotide is initiated from 3' to 5' with the first base already incorporated in the CPG support. The following base is then attached as a phosphoramidite derivative where a hydroxyl group with dimethoxytrityl (DMT) is used to protect the 5' end. As each base is added to the support, trichloroacetic acid (TCA) is delivered to remove the DMT, which gives out a bright orange color. The coupling efficiency of each cycle is determined by monitoring the conductivity of DMT since only bases successfully incorporated into the growing strand are present during the removal of DMT. In the next step, activation of the 5' hydroxyl with a weak acid, tetrazole, and delivery of the next amidite ensure coupling of the incoming base. Then a mixture of methylimidazole and acetic anhydride, an acetylating agent, is delivered to cap the end of failure chains that did not incorporate a base in this cycle. Iodine, an oxidizer, is added to the newly attached base to form a phosphotriester, a more stable product, before adding the next base.⁶⁶

After the completion of the synthesis, treatment with 10% diethylamine (DEA) is carried out to remove the protecting group of the phosphate backbone, the beta-cyanoethyl group, and cleave the synthesized oligo from the support with ammonium hydroxide.

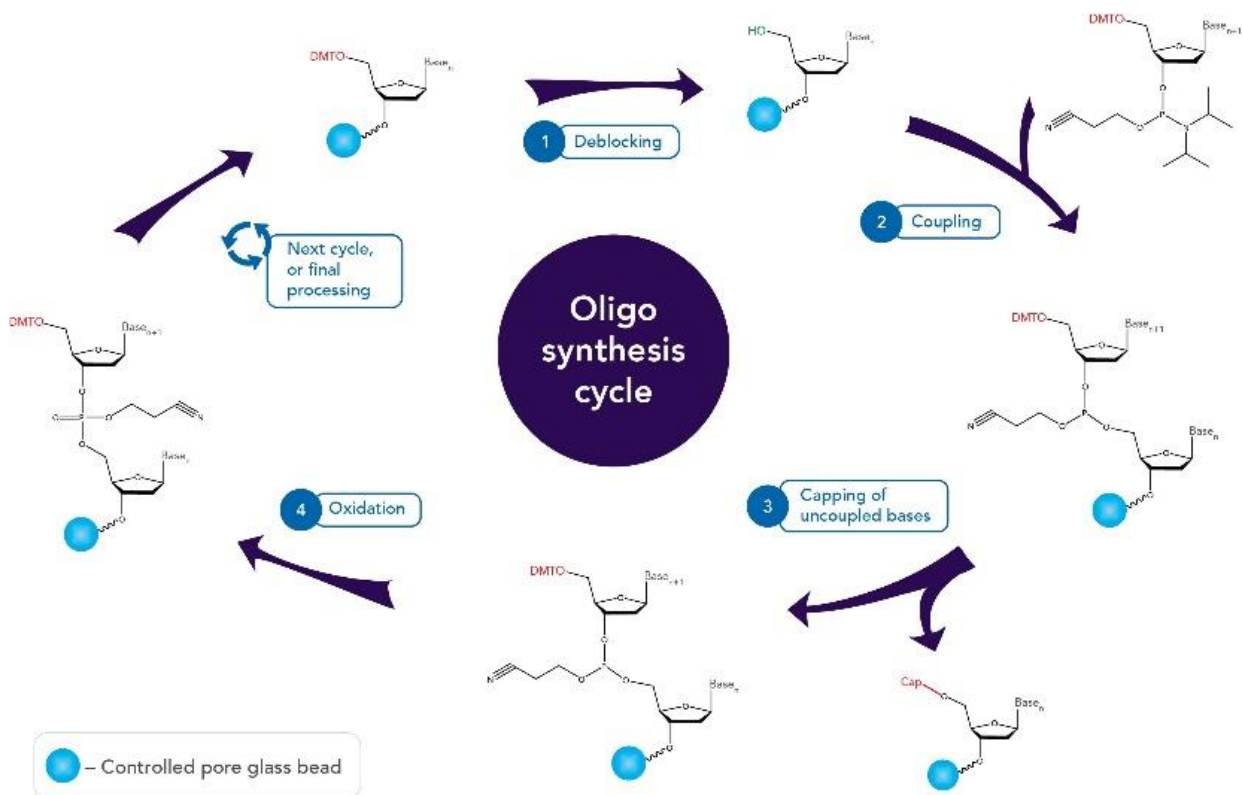


Figure 3.1: Solid-phase synthesis of an oligonucleotide. ⁶⁷

The synthesis of an oligonucleotide is carried out on an insoluble control pore glass (CPG) support. The steps include deblocking the DMT group, coupling with the incoming base, capping uncoupled bases, and oxidation to form a phosphotriester. The next cycle is continued until the last base is added to the growing strand.

Reprinted from Oligo Synthesis: Why IDT leads the oligo industry Published on Dec 18, 2015 (Revised/updated Aug 31, 2020) by Nan Pazdernik, Ph.D., Science Writer, IDT and Nolan Speicher, former IDT associate.

The 3-segment switch molecule (Figure 3.2) was synthesized using an ABI 394 DNA/RNA synthesizer (Applied Biosystems) on the support with controlled pore glass (CPG) (Glen Research). The generation of a molecular switch consists of a two-part process. In Part I of the synthesis, a 3'-DABCYL containing sequence (cover) was made, a 5-Me-dC brancher was added, and an aptamer containing sequence (probe) was synthesized. The 5' end of this chain must be capped so that no other bases will be added during Part II of the synthesis. Subsequently, the levulinyl group on the brancher was deprotected off the synthesizer with hydrazine hydrate to form cytosine-N4-C₆-OH on the brancher, allowing the second part of the synthesis to begin (C₆-OH represents n-hexanol).

In Part II of the synthesis, the third sequence (toggle) was incorporated at the N4- C₆-OH site, and CY3 phosphoramidite was introduced in the last cycle. The CPG column was removed to eliminate the β -cyanoethyl groups and cleavage from the solid support. Then, the sample was dried in a Speed-Vac DNA concentrator 120 (Thermo Scientific) and resuspended in 150 μ L Milli-Q purified water (Millipore Corp.). We gratefully acknowledge prior work done by Dr. R. Iyer and Dr. M. McPike and their collaborators in developing the above procedure.

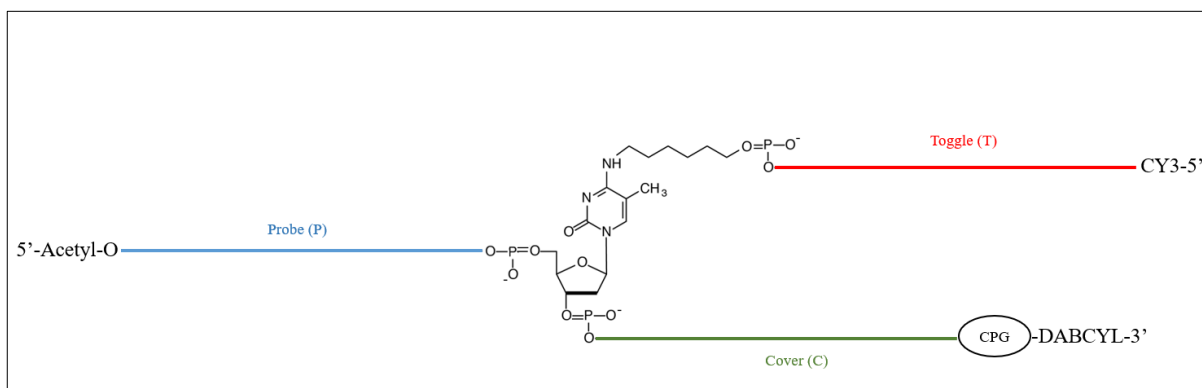


Figure 3.2: Synthesis Scheme of 3-segment DNA Switch.

The synthesis scheme of the 3-segment DNA Switch centers around a 5-Me dC brancher, making phosphate esters from its three hydroxyl (OH) groups. Standard solid-phase DNA synthesis assembles cyanoethyl-phosphoramidite-2'-deoxynucleosides (Figure 3.1) in the 3'to 5' direction. Part I of the synthesis begins from 3'DABCYL on controlled pore-glass (CPG) beads, until the cover (C) sequence is completed, Then a 5-Me dC brancher phosphoramidite is coupled to the last nucleotide residue (nt) in C, and the brancher's DMT group is removed to open the brancher's 5'-OH, followed by standard synthesis of probe (P). Finally, the DMT is removed from the 5'-OH of the last base in P, which is then capped (acetylated). The synthesis column is removed from the machine and the levulinyl group protecting the brancher's N4-C₆-OH is removed. The column is dried and restored to the machine to proceed with the synthesis of the toggle (T) sequence that concludes upon coupling a CY3-amidite at the 5'-OH of the last nt in T.

Results and Discussion

After the direct synthesis and clean-up of Switch NM-0, 68mer, MW= 21960.77g/mol (Figure 3.3), IE-HPLC analysis was monitored at the absorption maxima of 260nm for DNA, 550nm for CY3, and 475nm for DABCYL. Figure 3.4 shows an IE-HPLC chromatogram for Switch NM-0, two prominent UV peaks are observed at similar retention times. Peaks on the left side at around 8 min retention time were failures from part II of the synthesis, which has only UV (A_{260}) and DABCYL (A_{475}) absorption. The right-side peaks at around 9 min were tentatively assigned as containing our switches, with strong contributions from all three absorbance maxima. Fractions were collected as marked by vertical lines in Figure 3.4, and then analytical HPLC was run to estimate the purity of the fractions for further analysis. 1nM of the peak with strong absorbance at all three wavelengths was prepared for characterization at Novatia using electrospray ionization liquid chromatography-mass spectrometry (ESI/LC/MS). From the mass spec spectrum shown in figure 3.5, putative switch NM-0 (MW= 21963.3 g/mol) was in the acceptable error range of $\pm 0.03\%$ and $\sim 40\%$ pure.

In a short sequence trial run used for optimizing synthesis conditions, failure sequences were seen clearly as N-1, N-2, and N-3 peaks, where N is the chain length of the desired switch peak (data not shown). Because of the longer chains for switch molecules (68 bases for NM-0), the resolution of HPLC typically cannot display all the failure peaks, so there were overlapping peaks where only UV and DABCYL absorbance could be observed. Direct synthesis was optimized to simplify the recovery of authentic switch molecules.

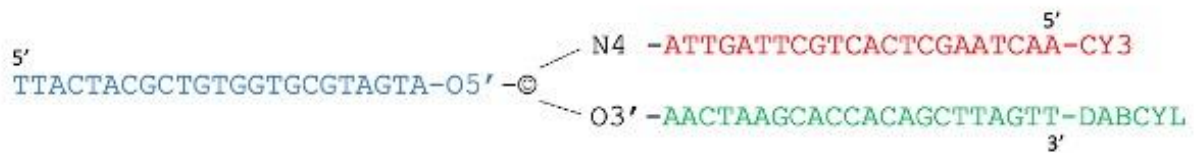


Figure 3.3: Switch NM-0 sequences for direct synthesis.

The full 68mer NM-0 switch ($K_1=0.02$) is designed using V-OMP and synthesized using a 3'-DABCYL CPG column.

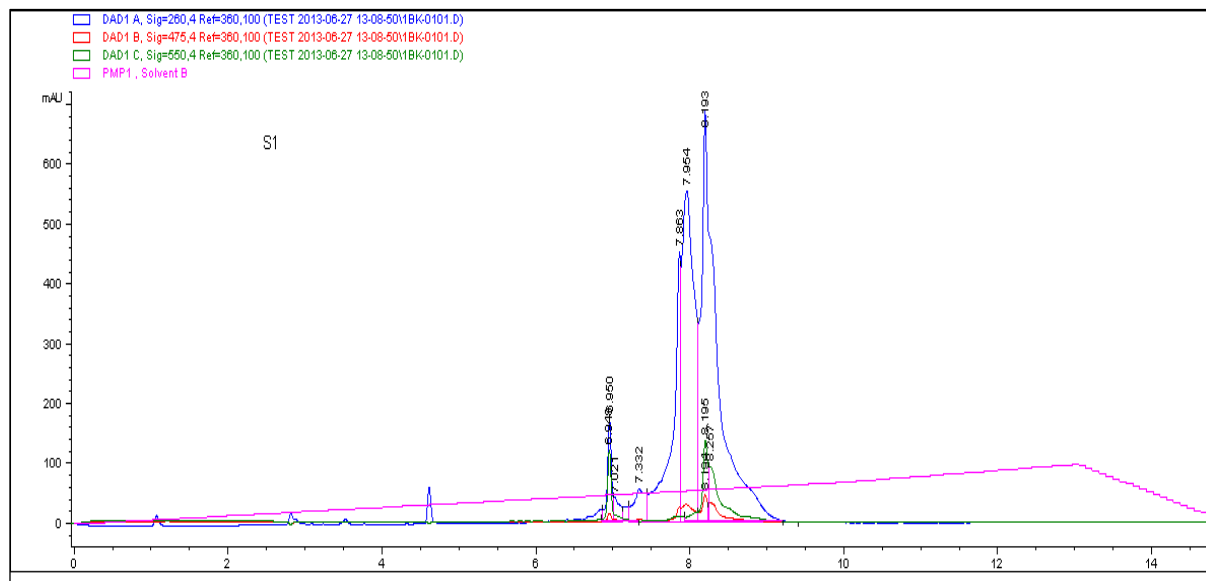


Figure 3.4: Analytical IE- HPLC of Switch NM-0.

Two prominent peaks are observed, which could be two conformations of the switch molecule. The analysis was run using UV absorbance at 260nm (blue line), DABCYL absorbance at 475nm (red), and CY3 absorbance at 550nm (green).

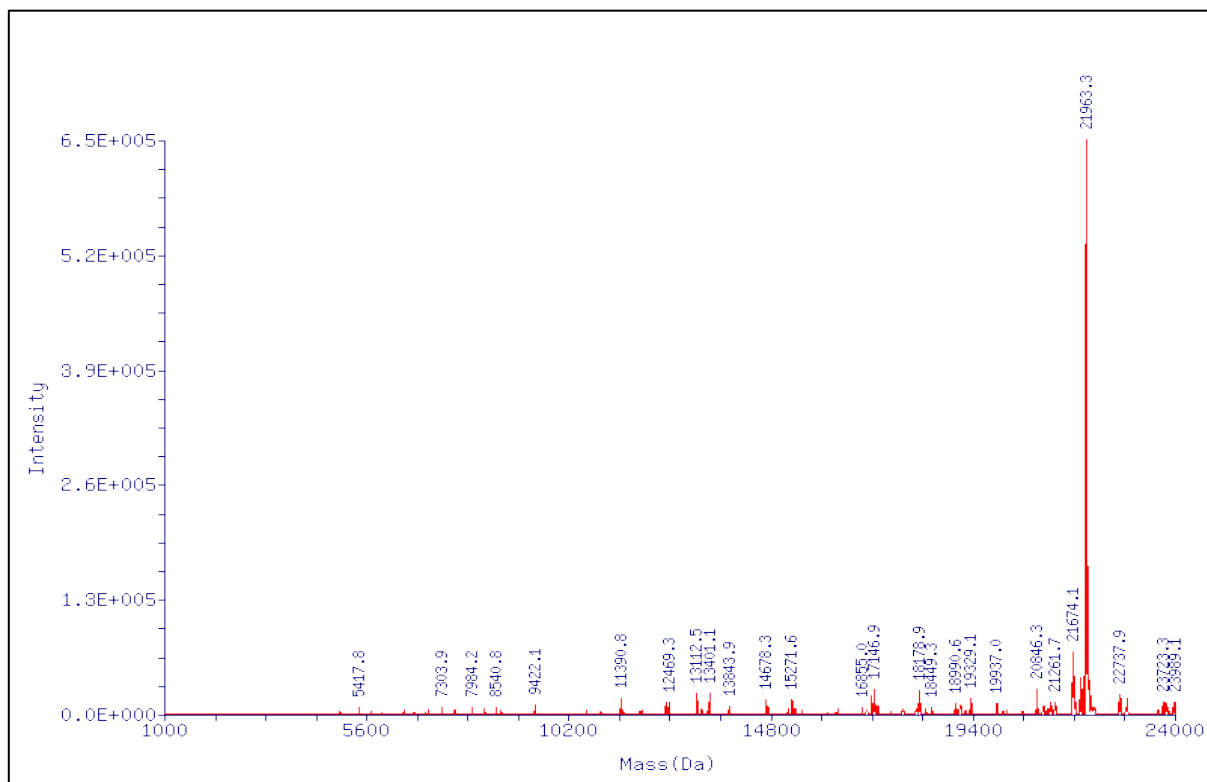


Figure 3.5: ESI-MS spectrum of NM-0.

The largest peak in the spectrum corresponds to the actual mass of switch NM-0 which is ~40 % pure.

3.2. Optimization of switch synthesis

A primary concern in synthesizing switches is failures in the addition of bases after levulinyl group deprotection from N4-C₆-OH of the 5-Me-dC brancher and resuming the synthesis for Part II. One possible explanation is that the stepwise yield of phosphoramidite coupling decreases as the length of oligonucleotides increases during the synthesis. To achieve a successful synthesis of switch oligonucleotides (>50 bases), all the DNA reagents in all the steps in the synthesis cycle: coupling, capping, oxidation, and detritylation, are required to be nearly optimal.

Maintaining the coupling efficiency of bases as high as possible is critical for a good synthesis. Theoretically, a 20mer with an average coupling efficiency of 98.0% will yield 68% full-length product, while a 100mer will only yield 13% full-length product (Glen Research). One of the main concerns for achieving high coupling efficiency is the presence of moisture either in the air or in the solution. Water reacts with the activated tetrazolide of the incoming nucleotide, which reduces the coupling efficiency.

Moreover, water catalyzes the phosphoramidite in the synthesizer by converting the phosphoramidite to phosphonate. Therefore, reducing the water content in all reagents, such as the acetonitrile (ACN), the activator solution, and argon used on the synthesizer, is critical. A transparent acrylic (Plexiglass) glovebox was used in our lab to enclose the synthesizer. Argon gas flowing through the box kept humidity very low, according to a monitor placed inside.

For the capping step, maintaining a high capping efficiency is vital for downstream purification. If the capping efficiency is low, shortmers with deletions (n-1, n-2, etc.) will accumulate, and deletions of 1 or 2 nt may be difficult to separate from the desired full-length product.⁶⁸ If detritylation is not complete at the end of a synthesis cycle, base skipping can occur.

Deprotection of the levulinyl group involves taking the synthesis column off the synthesizer and exposing it to a hydrazine hydrate in water for several minutes. In drying the column, clumps and voids in the solid support may occur. When the column is put back on the synthesizer, it is dried extensively with dry ACN, but the solvent could pass through voids and around clumps. If sufficient water remains after the ACN wash, amidites and coupling reagents can be degraded causing reduced coupling efficiency. Also, inefficient mass transfer into the bead pores can occur when passing reagents through the column for coupling, oxidation, capping, and detritylation. Insufficient removal of water will reduce stepwise yields, and cause failure chains, and side reactions.

In the synthesis of switches, steric hindrance can occur in the pores of the CPG beads. For instance, the 5'-end of the primary strand may interfere with the addition of bases off N4-C₆-OH. An analysis of steric hindrance on synthesis was carried out using alpha and beta-branched loops, and different sizes of beta-branched loops.

In collaboration with Dr. R. Iyer, the synthesis was optimized by changing parts of the procedure. Several modifications in the DNA synthesizer were made to:

- i. Decreasing the moisture content of phosphoramidites
- ii. Increasing capping and oxidation times
- iii. Increasing the concentration of phosphoramidites for part II synthesis
- iv. Allowing an additional two-minute coupling time after levulinyl group deprotection to drive efficient coupling of phosphoramidites

3.2.1. Optimization of coupling efficiency on the synthesizer

To investigate synthesis off the brancher, test sequences were designed and synthesized. Firstly, a synthesis of a simple sequence, 5'-TTTTTTT-3' (T7 control), and the sequence with a brancher (T3 control) (Figure 3.6) were performed using with and without the modified procedure. For the modified procedure, the water content of all reagents in the synthesis cycle, including the coupling, capping, oxidation, and detritylation, was dried with molecular sieves. Fresh phosphoramidites were prepared by diluting them with high-quality anhydrous ACN.

In the synthesis of a sequence with a brancher, capping off the 5'-hydroxyl of the probe strand, washing with anhydrous acetonitrile (ACN), and drying columns with argon were performed on the synthesizer after hydrazine hydrate treatment; columns were removed from the synthesizer in a fume hood for the levulinyl group deprotection. Other molecules containing the brancher were synthesized using all steps in the modified procedure (Figure 3.6).

Furthermore, a truncated version of a switch (Figure 3.7) with 12 nucleotides on each strand was synthesized using the modified procedure. This is also a great way to check the coupling efficiency on the synthesizer with a sequence that mimics a short version of switch molecules. Possible steric hindrance or secondary structure may also interfere with Part II of the switch synthesis for the longer authentic switches, so the truncated switch may also help evaluate this possibility.

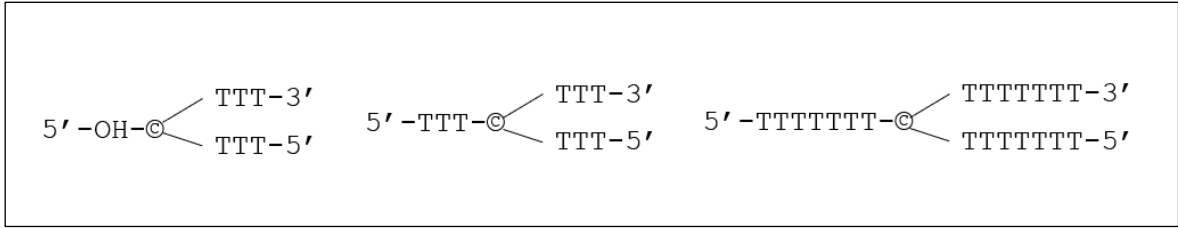


Figure 3.6: Sequences for Optimization of Coupling Efficiency.

(Left to Right): The sequence with a brancher (T3 control) with open HO-C6-N4-C, T3 branched molecule, and T7 branched molecule.

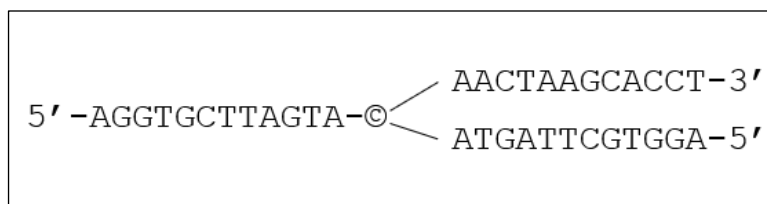


Figure 3.7: A truncated switch model for Optimization of Coupling Efficiency.

A truncated switch version with 12 nucleotides (nt) in each strand without fluorophore and quencher labeling. Including the brancher, there are 37 nt. This is not a functioning switch. The analogs of Toggle (left) and Probe (bottom right) are identical in the 5' to 3' direction. The Cover analog (top right) is fully complementary to each of the other two segments.

3.2.2. Optimization of capping and oxidation times on the DNA synthesizer

The capping and oxidation steps during oligonucleotide synthesis are as crucial as the coupling step in optimizing synthesis. As new phosphoramides are introduced to the growing strand, the failure sequences or unincorporated bases are capped with acetic anhydride to acetylate the 5'-OH of truncated sequences so that no additional bases will be added to these by-products. Since low capping efficiency can affect the purity and the coupling yield of the switch, optimization in capping and oxidation is essential in the synthesis. The following modifications were made:

- (i) Increasing the time for running capping reagents through the column by 25% (from 10 seconds to 13 seconds),
- (ii) Increasing the waiting time with cap reagents on the column from 8 seconds to 12 seconds, and
- (iii) Increasing the concentration of capping reagents and additional waiting time aimed to reduce failures in the base coupling that result in shortmers or shortmers with deletion.

More than thirty years ago, when the ABI 394 synthesizer was first introduced, a 1 M iodine mixture was used for oxidation. However, a side reaction of iodine and aqueous ammonia can form nitrogen triiodide, a highly oxidative solid, which can be unsafe and explosive.⁶⁹ Therefore, the concentration of iodine was reduced to 0.02 M. All synthesis steps are automated for safety concerns to minimize contact with harmful chemicals. However, there was an adverse effect: lowering the concentration of iodine the formation of the desired stable phosphate triester [P(V)] from its unstable intermediate form, and phosphite triesters [P(III)] are decreased due to rapid hydrolysis in the presence of either acid or base.⁶⁹ According to Glen Research's technical

support staff, the ABI 394 synthesizer did not consider the lower concentration of iodine for the oxidation step, so the time for the oxidation step was increased from 8 to 12 seconds to compensate for the reduced concentration of iodine.

Three test sequences were synthesized for capping efficiency optimization using the above-modified conditions (Figure 3.8). In a T6 [10] sequence, acetonitrile was delivered in place of the detritylation step needed to add the next thymidine phosphoramidite. Therefore, coupling the following thymine base should not be carried out in the next cycle after the capping step, adding no additional bases after the 6th base. T₁₀ and T₆ sequences are also synthesized along with the T6 [10]. The completed sequences, with or without the optimization, were analyzed with IE-HPLC.

3.2.3. Analysis of Steric Hindrance in Oligonucleotide Synthesis

Besides sequences to optimize coupling efficiency (Figure 3.6), sequences mimicking the base pairing in a switch molecule were checked for effects due to steric hindrance during the synthesis. In the experiments for determining the free energies of branched molecules, two conformations of branched molecules, alpha and beta loops (Figure 3.9) were synthesized for thermal denaturation profiles (done by Dr. R. Iyer).³⁴ For the alpha loop, the synthesis of 21 nucleotides was initiated from 3'-CPG through the brancher's O3' and O5', ending Part I of the synthesis at dG-C5'-OH. After levulinyl group deprotection, two additional bases were added off the brancher's N4-C₆-O as part II of the synthesis. In the beta-loop, part I of the synthesis was carried out by incorporating 13 nucleotides, through the brancher's O3' and O5', ending at dC-C5'-OH and then ten more bases were added to the brancher's N4 as part II of the synthesis. Along with alpha and beta-branched loops, effects on the length of sequences at the 5'-end of the primary strand in the synthesis were examined using Thymidine bases that varied from 2-7 bases (Figure 3.10)

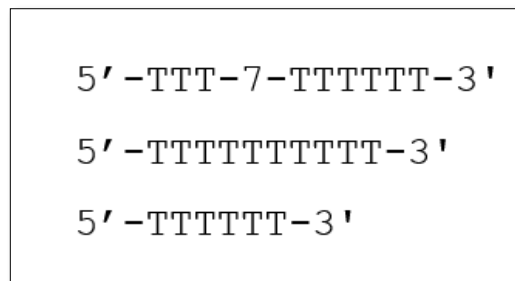


Figure 3.8: Test sequences composed of Thymidine phosphoramidite.

(Top to Bottom): T6[10] sequence (top) has the thymidine 7th base from the 3' end replaced by acetonitrile to end the synthesis, and so should be identical to capped T₆. The authentic T₁₀ sequence (middle) and T₆ sequence (bottom) were also synthesized to validate the capping efficiency.

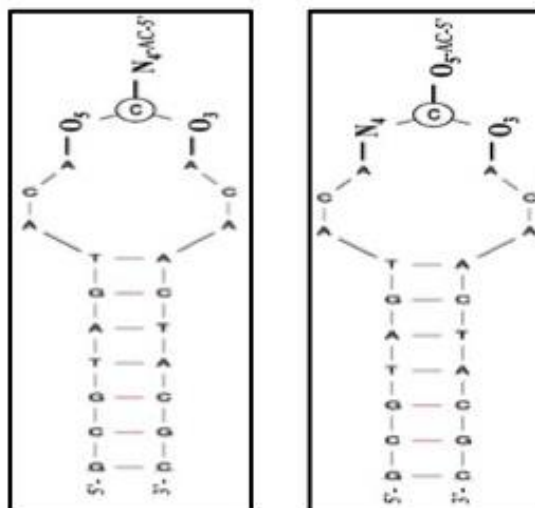


Figure 3.9: Branched loops models (ACn).

(a) A 7-base alpha loop, “AC7,” is prepared by a primary synthesis from 3’-CPG through O5’ of the brancher, ending at the final G at the bottom left. After levulinyl group deprotection, the secondary synthesis adds two additional bases (A-C) off ©-N4-C6-O- as shown at the top.

(b) Primary synthesis of beta AC7 (right), goes through the brancher as in (a) but adds only the two bases (A-C). This strand has only standard 3’-5’ phosphodiester linkages. After levulinyl group deprotection, the secondary synthesis adds ten additional bases (A-C-A-T-G-A-T-G-C-G) off ©-N4-C6-O- as shown on the left side. (This Figure was prepared by Dr. Iyer.)

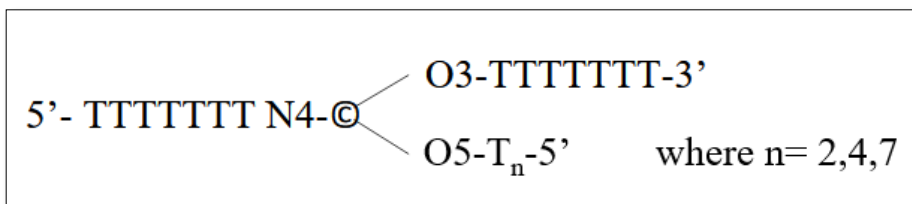


Figure 3.10: Model to mimic Beta Branched Synthesis.

Different lengths, n , of T nucleotides, were incorporated at the 5'-end of the primary synthesis (bottom right) to determine whether increasing “ n ” affects the efficiency of synthesis off ©-N4-C₆-O- (at left).

Results and Discussion

3.2.1. Analysis of coupling efficiency on the synthesizer

Thymidine phosphoramidite (Figure 3.11) differs from the protected nucleoside phosphoramidites dG^{iBu}, dC^{Bz}, and dA^{Bz}; as T has no base-protecting groups. Removal of DMT, and base protecting groups, benzoyl or isobutyryl, require incubation in ammonia at room temperature for 17- 24 h. By contrast, a synthesized sequence with only T phosphoramidites can be deprotected for 2 h at 65°C to remove cyanoethyl groups and cleave the oligo-T from the CPG. Using T phosphoramidites for optimization reduces the analysis time significantly.

The coupling efficiency of switch synthesis was evaluated using HPLC analysis and calculated trityl percentage on the synthesizer. In HPLC analysis, T7 controls with and without the modified procedure were compared (Figures 3.12 and 3.13). The intensity of failure peaks (shorter sequences at n-1, n-2, etc., where n= the full-length chain) decreases by five times, making the modified procedure a very promising putative switch.

The coupling efficiency of each base indicates how successfully subsequent bases are added to the growing strand. The ABI 394 synthesizer is equipped with conductivity flow cells for each column. The DMT efficiency of each base is calculated by monitoring the dimethoxytrityl (DMT) cation or signal value as it is removed from the base and the background or noise value after the elimination of DMT from the conductivity flow cell. The measurement of coupling efficiency of incoming DNA base is the difference between signal value and noise value in percentage output.⁶⁵

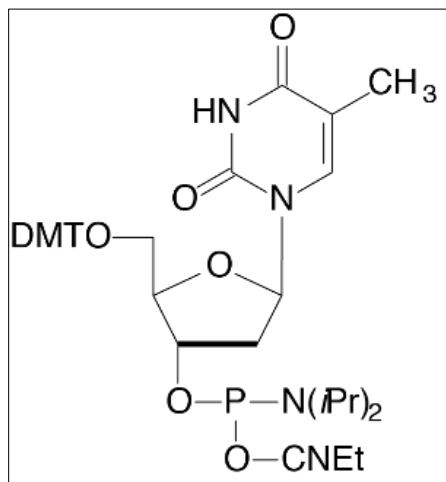


Figure 3.11: dT-CE phosphoramidite.⁷⁰

5'-Dimethoxytrityl-2'-deoxyThymidine,3'-[(2-cyanoethyl)-(N,N-diisopropyl)]-phosphoramidite does not have a base-protecting group which requires minimal incubation time to cleave oligo-T from the CPG.

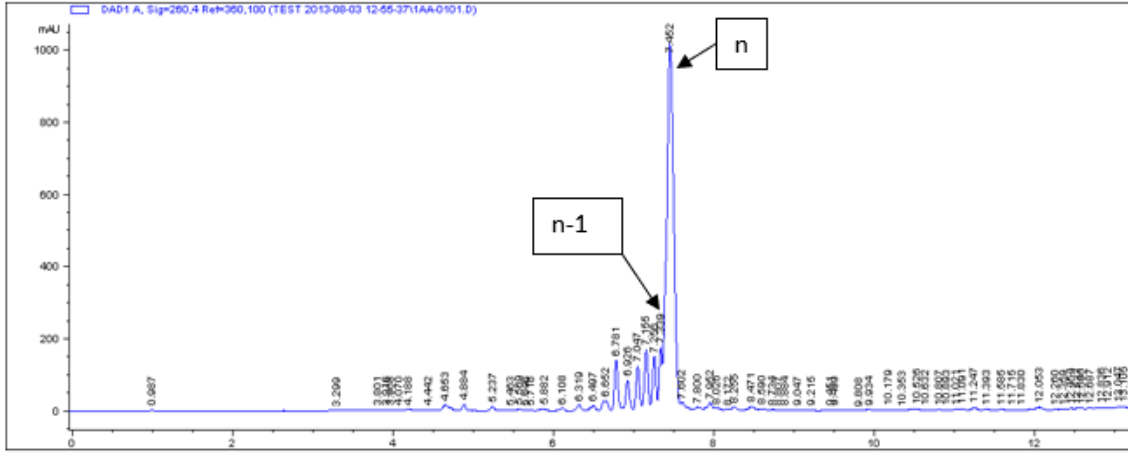


Figure 3.12: Analytical IE-HPLC chromatogram of T7 control with original synthesis procedure.

Assuming n = full-length oligonucleotide, the subsequent peaks on the left side of the main mountain are the failure sequences that can denote $n-1$.

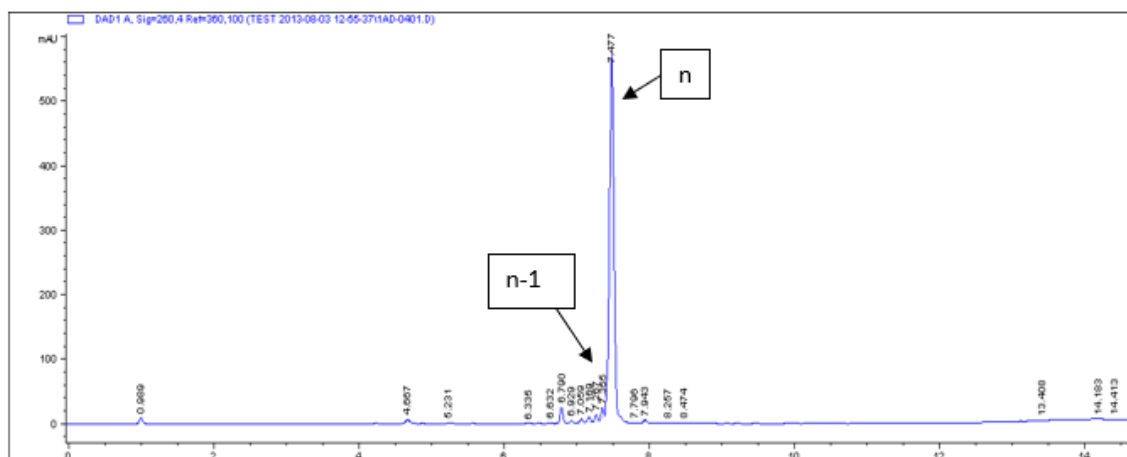


Figure 3.13: Analytical IE-HPLC chromatogram of T7 control with modified synthesis procedure.

Failure peaks with lower intensities were observed using a modified procedure by comparing them with the original procedure (Figure 3.12).

Base	Trityl (%)	
	T7_Modified	T7
T	100	100
T	100	97.1
T	98.2	100
T	98.6	95.9
T	98.9	97.3
T	98.6	97.9
T	98.8	97.9

Table 3.1: Trityl yield of T7 branched molecule.

Dimethoxytrityl (DMT) measurement for T7 branched molecule with and without modified procedure. An increase in the coupling efficiency of each addition base was observed by using the T7_modified procedure.

A general increase in coupling yield (trityl %) was observed with the T7 branched molecule (Figure 3.6) synthesized using the modified procedure. A similar result was obtained using the modified procedure with the T3 branched molecule (data not shown). For HPLC analysis of branched molecules, the T3 control (Figure 3.6 left) served as a benchmark for the retention time of part I of the synthesis. By comparing the retention time on HPLC chromatograms of the T3 control and T3 branched molecules, we confirmed that no visible peaks were observed before part II of the synthesis (data not shown).

In chromatograms of the T7 branched molecule (Figure 3.15), there are seven failure peaks before the main peak. We observed that the intensities of failure peaks are visibly decreased in the sample using the modified procedure (Figure 3.14). From the analysis, we are confident that the coupling efficiency is greatly improved by implementing new parameters in the oligonucleotide synthesis.

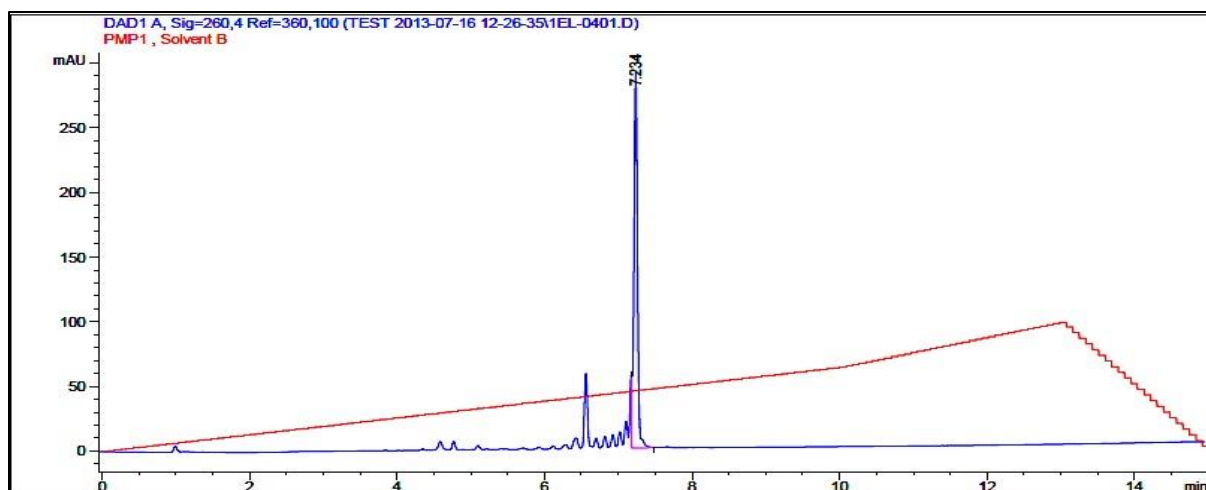


Figure 3.14: IE-HPLC chromatogram of T7 branched molecule using the modified procedure.

The analytical IE-HPLC was run using a gradient of 0-60% Buffer B at 55°C for 10 min. The decrease in intensities of truncated peaks was observed using the modified parameters.

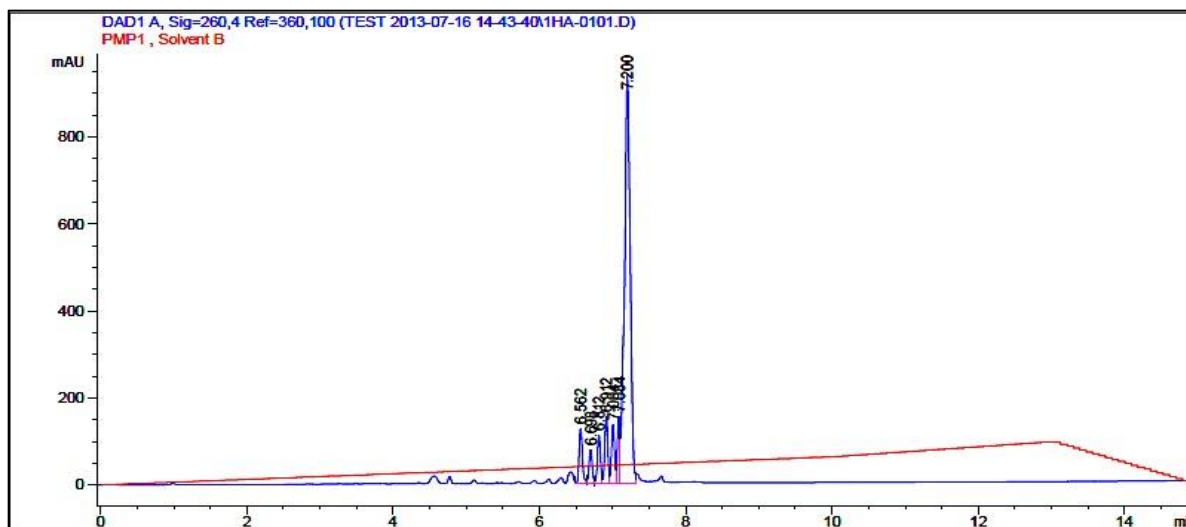


Figure 3.15: IE-HPLC chromatogram of T7 branched molecule using the original procedure.

IE-HPLC utilized a gradient of 0-60% Buffer B at 55°C for 10 min. Multiple truncated peaks were observed on the left side of the main peak.

The IE-HPLC analysis of the truncated 37mer molecule (Figure 3.7 and Figure 3.16), shows that failure chains eluting from 4 to 7 min are quite small. However, there are shoulders on both sides of the main peaks (12_1 and 12_2). This suggests that HPLC may not be useful to prepare 60-80 nt active switches in high purity.

The two prominent peaks (12_1 and 12_2) at the retention time of 7.058 min and 8.053 min were observed. Interestingly, IE-HPLC analysis of a T7 branched molecule synthesized with or without the modified procedure shows a single prominent peak. In contrast, two high-intensity peaks are detected in the analytical IE-HPLC of a truncated switch molecule. We suggest that it could be due to (1) incomplete deprotection or (2) a slow equilibrium on the column between two distinguishable conformations that give the split peaks. In the case of incomplete deprotection, a beta-cyanoethyl group remains attached to the 3'-oxygen on the phosphate backbone, and/or protecting groups on dA, dC, and dG may not be deprotected completely. These protecting groups may interact with the IE HPLC column, resulting in different peaks. We collected the peaks separately to test these hypotheses and obtained their mass spectra.

From the ESI-MS data (Novatia, LLC), both peaks match the calculated molecular weights (Table 3.2) within the acceptable error range. Since both HPLC peaks possess very similar molecular weights, this supports the hypothesis that the two peaks in IE-HPLC are due to conformational differences in the switch.

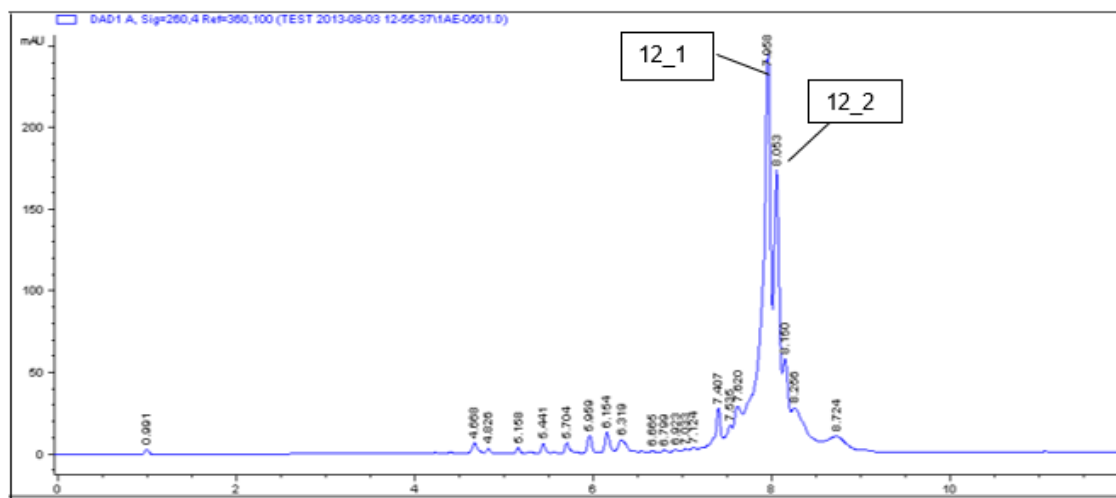


Figure 3.16: IE-HPLC chromatogram of the 37mer truncated switch molecule.

This sequence (Figure 3.7) was used to test the coupling efficiency of the optimized synthesis procedure. Two prominent peaks were observed which could be due to a slow equilibrium between two conformations that give the split peaks.

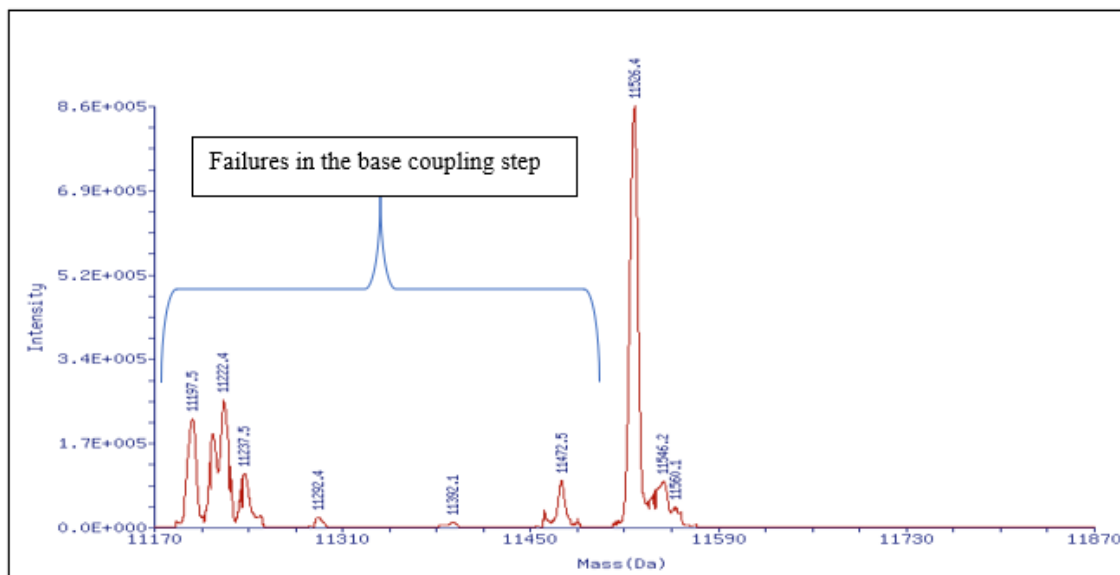


Figure 3.17: ESI-MS spectrum of peak 12_1 from Figure. 3.16.

The highest intensity was the actual molecular weight of a truncated switch. The peaks on the left side could be due to failures in base coupling resulting in shortmers or shortmers with deletion, while the peaks observed on the right side are undetermined.

Peak	Calculated MW (g/mol)	Actual MW (g/mol)	Purity (%)
12_1	11525.7	11526.4	39.59
12_2	11525.7	11526.1	34.75

Table 3.2: ESI-MS data of a truncated switch molecule.

Two prominent peaks in a truncated switch molecule sample confirmed to be the two conformations with a similar equilibrium giving split peaks in figure 3.16.

It is also interesting that the mass spectrum (Figure 3.17), has masses at the left of the spectrum that differs from the main peak by 288 to 328 amu, in the range of the four deoxynucleotide monophosphates (307-347). The acetyl group used for capping would contribute 43 amu. These peaks occur within the envelope of peak 12_1 from the IE chromatogram (Figure 3.16) and contribute about 30% of the total peak intensity in the mass spectrum. These peaks probably include failures due to sequences due to coupling or capping failures. Thus, IE-HPLC could not remove these failure sequences, reducing the purity of the final product. Contaminants of the desired peak will be even more challenging to remove in the switches compared to this truncated version. Therefore, other possibilities for purification are explored.

The mass spectrum reveals small peaks to the right of peak 12_2. The largest is in the position of the monosodium adduct, which is common in the ESI-MS of DNA and RNA. There is little evidence for remnants of base-protecting groups or remnant cyanoethyl groups.

Regarding the identical molecular weights for HPLC peaks, 12_1 and 12_2, it is often assumed that secondary structures of DNA and RNA are denatured in acetonitrile. This is because water is important in maintaining helical structures, where the hydrophobic bases lie in the center of the helix, and the hydrophilic sugars and phosphates on the outside. In single strands, water forms an icelike structure surrounding the aromatic bases in single strands; this water is released to the bulk upon forming a helical stem or duplex, contributing to the stability of the helix. However, we are unaware of any measurements demonstrating the absence of a secondary structure for RNA and DNA in dry acetonitrile.

If some remnant helical structure occurs in ACN, it is still hard to understand how two separate peaks can occur in the ion exchange chromatogram of Figure 3.14, because both Probe and Toggle segments are fully complementary to Cover. Both helical segments are identical, and

both single-stranded segments are identical. However, in a structure built around the 5Me-dC brancher, the Cover:Probe stem is connected through an alpha-loop, and the Cover:Toggle stem is connected through a beta-loop. This might be sufficient to generate distinct interactions with the ion-exchange column that lead to separate peaks.

3.2.2. Analysis of capping and oxidation time on the DNA synthesizer

As the peaks observed on the left side of the ESI-MS spectrum (Figure 3.17) could be due to the capping failures during the synthesis. Capping limits failure chains so that by-products due to base skipping can be minimal. Three sequences (Figure 3.8) were synthesized and then analyzed by HPLC to optimize the additional times for capping and oxidation. The T6 and T6[10] molecules should be identical with proper capping since only acetonitrile was introduced at the 7th base of the T6[10] synthesis. By comparing the retention times of analytical IE-HPLC for the three sequences (Table 3.3), the prominent peaks of the T6 and T6[10] are eluted at a very similar retention time. This validates that the increase in capping and oxidation time reduces the chance of failure sequences being carried over to the incorporation of the next base. T6[10] synthesized without optimized capping and oxidation times was used as a control. By comparing analytical IE-HPLC chromatograms (Figure 3.18 and 3.19), there was a peak on the right side of the main peak (Figure 3.19), which could be due to the by-products (T8 or T9) due to base skipping.

Sample	Retention Time
T6	4.829
T6[10]I	4.798
T10	6.410

Table 3.3: Retention time of the prominent peaks on analytical IE-HPLC.

The prominent peaks of the T6 and T6[10] are eluted at a very similar retention time. This validates that the increase in capping and oxidation time reduces the chance of failure sequences being carried over to the incorporation of the next base. T6[10] synthesized without optimized capping and oxidation times was used as a control.

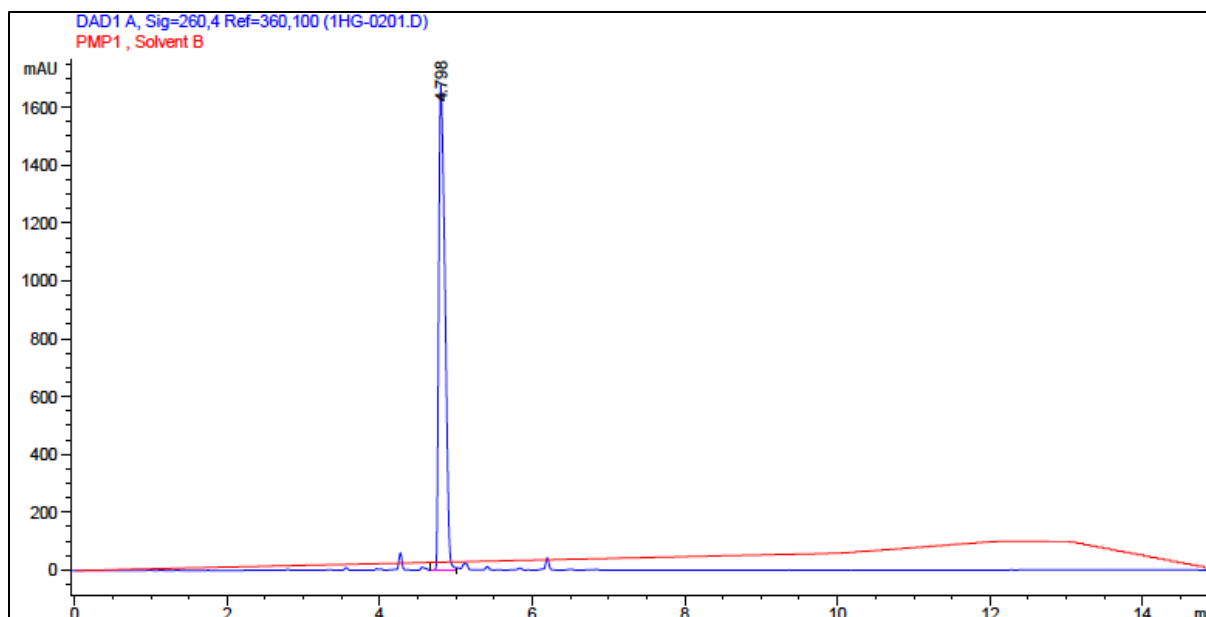


Figure 3.18: Analytical IE-HPLC of T6[10] molecule.

The T6[10] sequence was synthesized using the optimized oxidation and capping times and analyzed using 0-60% Buffer B at 55°C for 10 min. Minimal by-products are obtained.

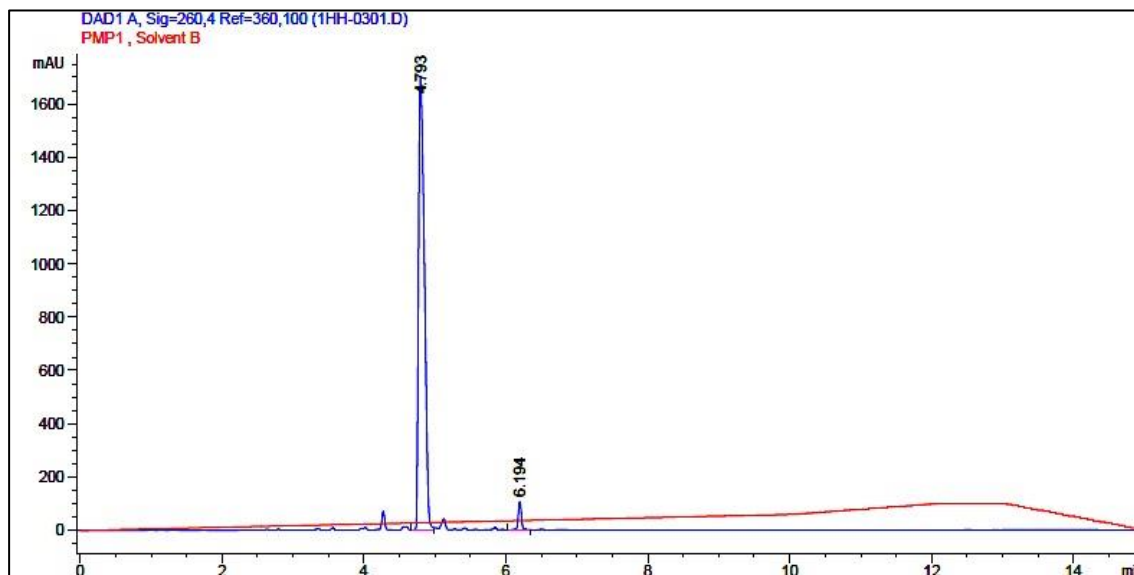


Figure 3.19: Analytical IE-HPLC of T6[10] molecule without optimized procedure.

The T6[10] sequence was synthesized without the optimized oxidation and capping times and analyzed using 0-60% Buffer B at 55°C for 10 min. The peak eluted at 6.19 min could be due to insufficient capping.

3.2.3. Analysis of Potential Steric Hindrance in a Branched Oligonucleotide Synthesis

It is possible that having DNA chains off both the HO5'-©-3'-OH of the brancher interferes with access to ©-N4-C₆-OH in the secondary synthesis after the removal of the levulinyl-protecting group. This would decrease coupling off ©-N4-C₆-O- and increase the number of failure chains. The steric crowding effect could become more severe as the chain length (n) increases off the brancher's O5' in the primary synthesis. This hypothesis would also explain Dr. Iyer's repeated observation of low coupling efficiency and larger amounts of failures in the synthesis of alpha-linked hairpins (Figure 3.9a) compared to hairpins with beta-linkage (Figure 3.9b).³⁴ In his primary syntheses of beta-loops, only A-C (n=2) was added off HO5'-©, while his syntheses added n=7 residues in alpha-loops. An unstructured random coil with n=7 would be more likely to interfere with access to ©-N4-C₆-O- than n=2.

To further test this hypothesis, constructs were prepared with primary syntheses of T₇-O3'-©-5'O-T_n (n=2,4,7), then the levulinyl was removed, and the secondary synthesis completed with ©-N4-C₆-O-T₇ as depicted in Figure 3.10. For convenience, these constructs are named T_{7-n-7}. The secondary synthesis used the modified procedure (sections 3.2.1 and 3.2.2). In addition, a 6 min pause was included to increase the coupling time for each base in the secondary synthesis (as suggested by the technical support at Glen Research).

The T₇₋₂₋₇ construct showed a very low intensity for failure sequences in the IE-HPLC (Figure 3.20); T₇₋₂₋₇ is nearly 100% of this full-length construct. The % Full length was calculated by using the area under the curve (AUC) of the desired peak divided by the HPLC chromatogram's total area. It is apparent from Table 3.4 that failures increase and that the amount of full-length construct decreases slightly for T₇₋₄₋₇ (96%), and T₇₋₇₋₇ (81%). This supports the hypothesis of steric interference on polymerization from ©-N4-C₆-O- by increasing the length of the segment

polymerized off ©5'-OH in the last part of the primary synthesis. However, the % Full length for T_{7_7_7} using the revised procedure is still 30% more than for the identical molecule, T_{7_branched} (Figure 3.6) prepared earlier, suggesting that a longer coupling time after levulinyl group deprotection is essential for the synthesis of a branched oligonucleotide.

Active 3-segment switches, such as those prepared and evaluated in this work, have segment lengths greater than 20 nt each. Direct solid-state synthesis, as opposed to ligation-based methods discussed in Chapter 4, cannot avoid steric interference from the long segment polymerized off ©5'-OH in the last part of the primary synthesis on the secondary synthesis off ©-N₄-C₆-O-. If the steric interference hypothesis is correct, as suggested in the previous paragraphs, yields of full-length active switches will be low, with substantial failure chains resulting from each step in the synthesis of the third segment off ©-N₄-C₆-O-.

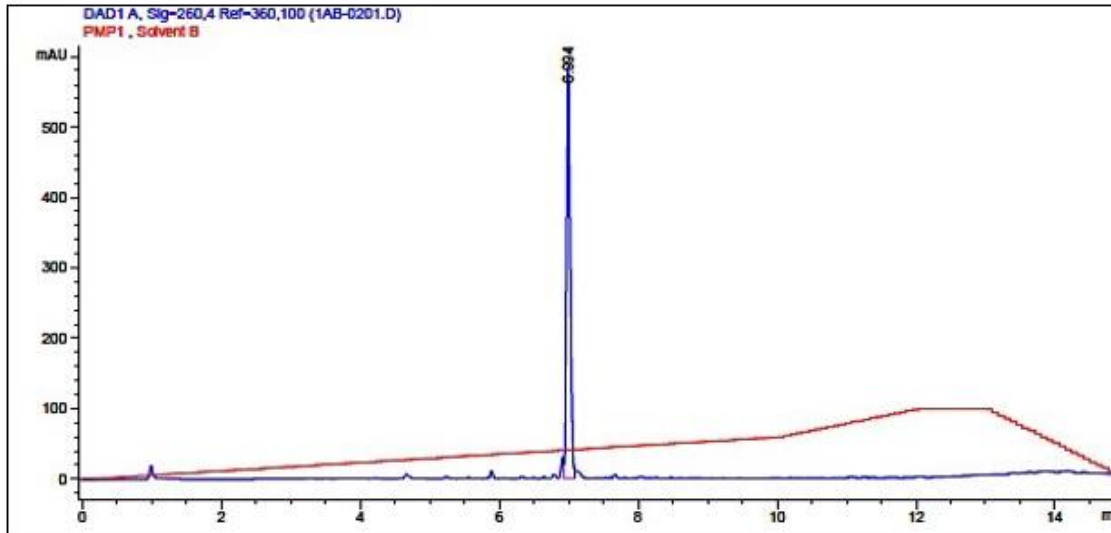


Figure 3.20: Analytical IE- HPLC chromatogram of T_{7_2_7} beta-branched molecule.

IE-HPLC was conducted using a gradient of 0-60% Buffer B at 55°C for 10 min. No significant failure peaks were observed.

Construct	% Full length
T_{7_2_7}	100
T_{7_4_7}	96.28
T_{7_7_7}	80.55
T_{7_branched}	50.21

Table 3.4: Comparison of % of full-length constructs built around © based on IE HPLC.

See Figure 3.10 for details. Column two gives the area of the full-length constructs divided by the sum of failures and full-length expressed as a percent. T_{7_7_7} is the same molecule as T_{7_branched} (Figure 3.6) but the latter was prepared without a 6 min pause to increase the coupling time for each base in the synthesis off ©-N₄-C₆-O-.

3.3. Optimization of the Purification of Switch Molecules

The purity of the switch molecule is as vital as the coupling efficiency of nucleotides during the synthesis for downstream assays. Its purity was still very low from the ESI-MS analysis of the truncated switch using the modified synthesis procedure. Even with 98% coupling efficiency of each base in a 20mer oligonucleotide synthesis, only 67% will be a full-length product. The end-product of a 70mer switch molecule synthesis would be only 24% of the total synthesis scale. Therefore, any failure sequences must be removed by purification. Purification of switch molecules was investigated by comparing IE and RP HPLC (described in this chapter) before settling on purification by denaturing polyacrylamide gels (Chapter 4).

IE-HPLC applies the principle of charged molecule separation, while RP-HPLC utilizes the difference in hydrophobicity of the molecules. In IE-HPLC, the ionic interaction between the oppositely charged groups in the sample molecule and the ligand in the column support determines the absorption of the molecules into the column support. By increasing the salt concentration in a linear gradient, the oligonucleotides with the weakest ionic interaction (shorter length) will elute first from the column, and the longer-length oligonucleotides will elute later.

In RP-HPLC, the adsorption of hydrophobic molecules onto the hydrophobic column support in a polar mobile phase is applied. By increasing the amount of organic solvent in the eluent, de-sorption results depend upon differences in hydrophobic interaction between sample molecules and column support. Therefore, the more hydrophobic a molecule is, the longer the time it will bind to the column support in a gradient that increases the organic content of the eluent.

There are several criteria to consider when purifying a 3-segment switch molecule since it consists of about 70 bases total where each segment can form an individual hairpin. There is a quencher on the 3' end of the cover and a fluorophore on the 5' end of the toggle strand. Oligonucleotides with dye molecule(s) are suitable to perform RP-HPLC due to the intrinsic hydrophobicity of the dyes. This allows a significant separation between the desired product and sequences without dyes. However, the resolution of RP-HPLC decreases as the length of oligonucleotide increases significantly beyond 50 bases. Also, a stable secondary structure like a hairpin loop can give rise to a broad peak or multiple peaks, limiting RP-HPLC for purification. An advantage of purification with IE-HPLC is that it can tolerate alkaline pH that can disrupt DNA secondary structures. At high pH, N (3) of thymidine and N (1) of guanine are deprotonated, and inter-base hydrogen bonds are destabilized.⁷¹

Since the purity of the truncated switch was still low using modified synthesis procedures, we thought that the application of combined RP and IE techniques could result in a final product with high purity. Instead of performing dual HPLC purification on the HPLC system, a Glen-Pak™ DNA/RNA purification cartridge was selected for RP-HPLC because it will be less time-consuming and waste fewer reagents.

A Glen-Pak™ (GP) DNA/RNA purification affinity column contains a proprietary resin that binds to 5'-DMT in full-length sequences to remove the failure sequences efficiently.⁷² Glen Research claims that oligos up to 150 bases in length can be purified using a DNA cartridge on the scale of 40 nmol up to 1 μmol. It will also be useful for purifying the putative switch because HPLC analysis has low sensitivity in resolving >20 nt as the failure peaks merge into the desired peak. Incorporating GP columns in our current direct synthesis purification was straightforward. The switch molecule was synthesized with 5'-DMT retained.

After DNA synthesis, the DMT-ON oligonucleotide was deprotected from the column using ammonium hydroxide then it was applied to the GP cartridge by mixing with an equal volume of 100mg/mL NaCl. The DMT-ON oligo will bind to the packing material while failure sequences without DMT are washed away. Then, DMT was removed by rinsing with 2% trifluoroacetic acid (TFA, a faint orange band appeared as DMT was removed from the full-length oligo). Lastly, The desired product was eluted from the column using 50% acetonitrile in water containing 0.5% ammonium hydroxide.⁷² The purification efficiency of GP columns was tested using two non-branched hairpin loops, AC7 and AC9 (Figure 3.20), and a full-length switch molecule.

3.3.1. Synthesis of a non-branched hairpin loop

Initially, the efficiency of the GP cartridge was tested using two non-branched DNA hairpin loops, AC7 and AC9 (Figure 3.21), with DMT of the last base retained on, the last base during the synthesis; that is the detritylation of the previous base added was not carried out during the synthesis. These controls (DMT-ON oligonucleotides) are selected because all segments of our switch molecules can form a stable secondary structure (hairpin loop). 50 μ L of DMT-ON oligo (control) was removed from the sample to compare the purification with and without the GP column. DMT was manually released using 80% acetic acid in water⁷³, while the rest of the sample was purified using a GP column.

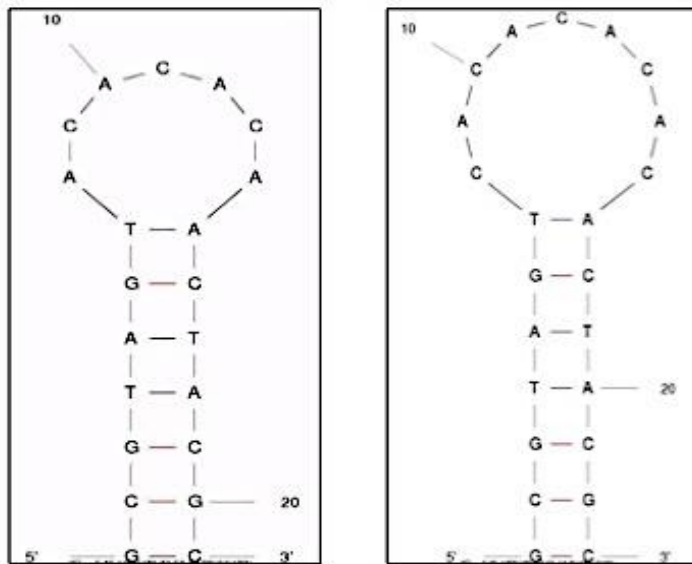


Figure 3.21: Structure of non-branched DNA hairpin.

AC7 (left) and AC9 (right). These control oligos, which have stable secondary structures, were synthesized to test the purification efficiency of the GP column. (This Figure was prepared by Dr. Iyer)

3.3.2. *Synthesis of unlabeled switch molecules*

In the optimization experiments, samples with high purity are obtained for DNA hairpin loops which have ~25 bases and are non-branched. Since most of the failures were observed in the second strand after levuliny1 group deprotection in the synthesis of switches, the formation of a secondary structure may interfere with the purification of the switches. It was necessary to determine the purification efficiency of GP for a three-way branched switch molecule and examine the effect of steric hindrance by forming a secondary structure if any. Unlabeled Switch-1 and Switch-2 (Figure 3.22) were designed in a way that the cover strand and the probe strand are base-paired in Unlabeled Switch 1, while there was minimal base pairing between the toggle strand and the probe strand in Unlabeled Switch 2. To use GP in purifying the unlabeled switches, DMT was retained on the 5' end of the desired product.

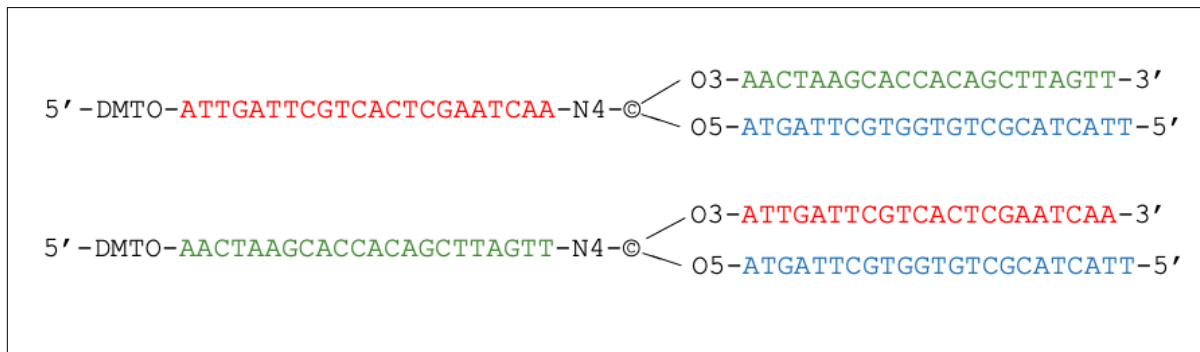


Figure 3.22: Unlabeled Switch Molecules.

Unlabeled switch 1 (top), where the cover strand and probe strand were synthesized in Part I, and the toggle strand was synthesized in Part II. Unlabeled switch 2 (bottom) was synthesized to determine the effect of a steric hindrance.

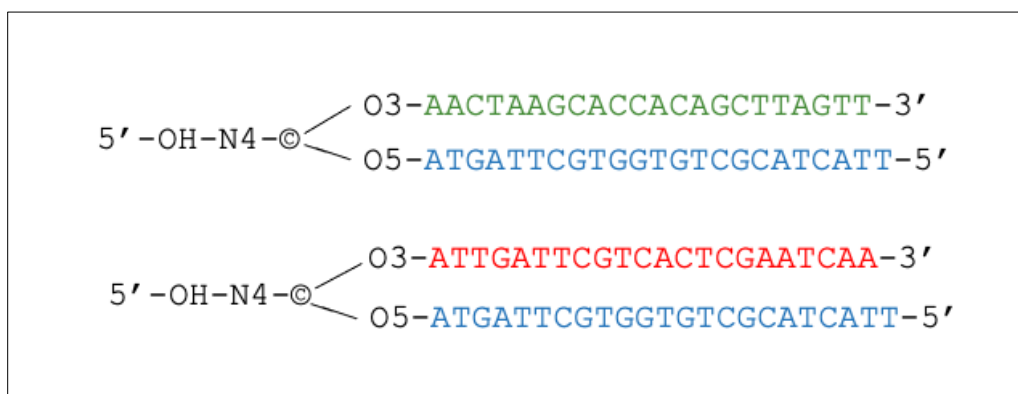


Figure 3.23: Controls for Unlabeled Switch Molecules.

(Top to Bottom): The two controls for Switch 1 (top) and Switch 2 (bottom) were synthesized to monitor failure sequences after levulinyl group deprotection.

The unlabeled switch molecules were synthesized using the modified procedure to increase the coupling efficiency of bases during the synthesis. In unlabeled Switch 1 (Figure 3.22 top), the synthesis was initiated from 3' of the cover strand (green) to the 5' end of the probe (red) strand as part I of the synthesis. The third strand, toggle (blue), was synthesized in part II, where the last base was not detritylated. A different synthesis order was tested in unlabeled Switch 2 (Figure 3.22 bottom). Unlike unlabeled Switch 1, the toggle replaces the cover in Part I synthesis. The control sequences (Figure 3.23) without a third segment were also synthesized to identify the switch fragment retention time from part I synthesis for IE-HPLC analysis.

3.3.3. Synthesis of CY3 labeled oligonucleotides

Since GP purification utilized the hydrophobicity of the DMT moiety in a synthesized oligo and gave very promising results in purity for switches 1 and 2, CY3 labeled 20mer oligonucleotide (5'-CY3-CATTTAAACATATAATACGG-3') was synthesized to perform GP purification and analyzed using IE-HPLC. The fluorescent dye, CY3, is hydrophobic like DMT and has a high affinity to the Glen-Pak cartridge. Since CY3 is added last in the switch synthesis, the failure sequences arising from the synthesis after levulinyl group deprotection should be eliminated first using a GP cartridge and then the purified product eluted using 80% ACN in water with 0.5% ammonium hydroxide.

Results and Discussion

3.3.1. Analysis of non-branched hairpin loops

A GP DNA purification cartridge can utilize the hydrophobicity of 5'-DMT (dimethoxytrityl) on the last base of the oligonucleotide to bind to the support. Therefore, failure sequences without 5'-DMT on the previous base will elute first from the GP. Then DMT is removed with 2% TFA, eluting the purified product using 50% acetonitrile in water. For comparison with or without GP, a 50 μ l 5'-DMT-on sample (control) from the synthesis was aliquoted, and DMT was removed with 20% acetic acid in water. The oligonucleotides were then recovered using either sodium acetate/ethanol precipitation or ethyl acetate phase extraction. In the analytical IE-HPLC of the control sample recovered using the two methods (Figure 3.25), the failure peaks are observed on the left side of the main peak (eluted at 7.243 min). The GP purification effectively reduces the number of failure sequences by comparing IE-HPLC of oligos purified using GP (Figure 3.24) and control (Figure 3.25).

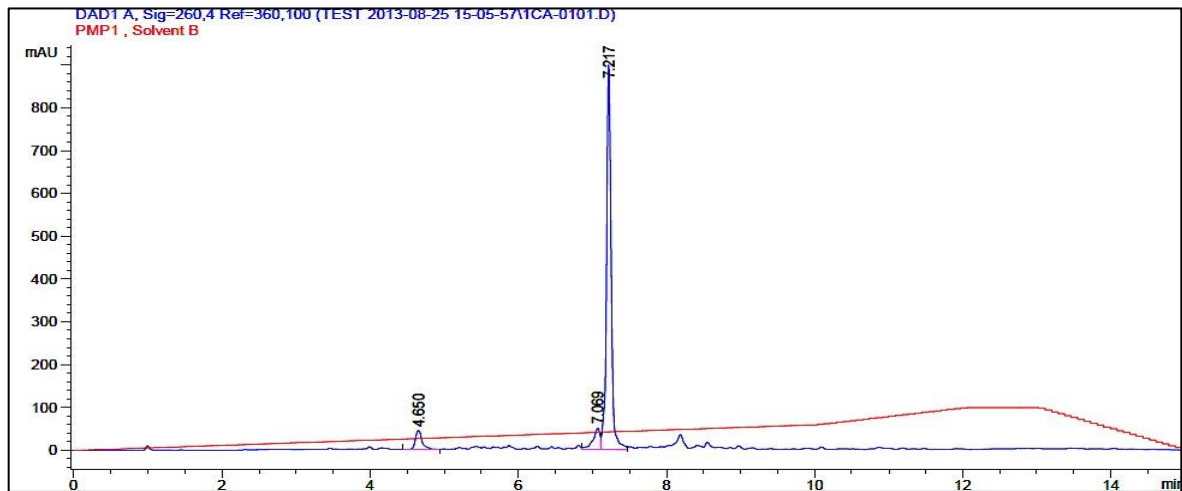


Figure 3.24: Analytical IE-HPLC chromatogram of AC₇ using GP purification.

A gradient of 0-60% buffer B at 55°C in 10 min was used to analyze after GP purification. The GP purification effectively reduces the number of failure sequences by comparing IE-HPLC of oligos purified using GP (Figure 3.24) and control (Figure 3.25).

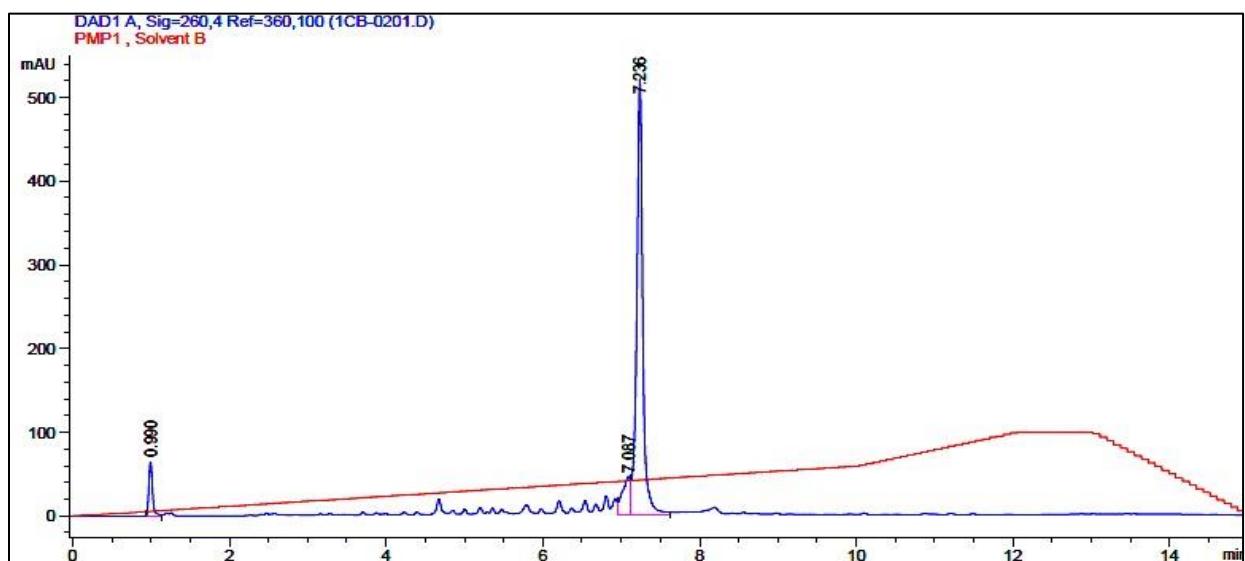
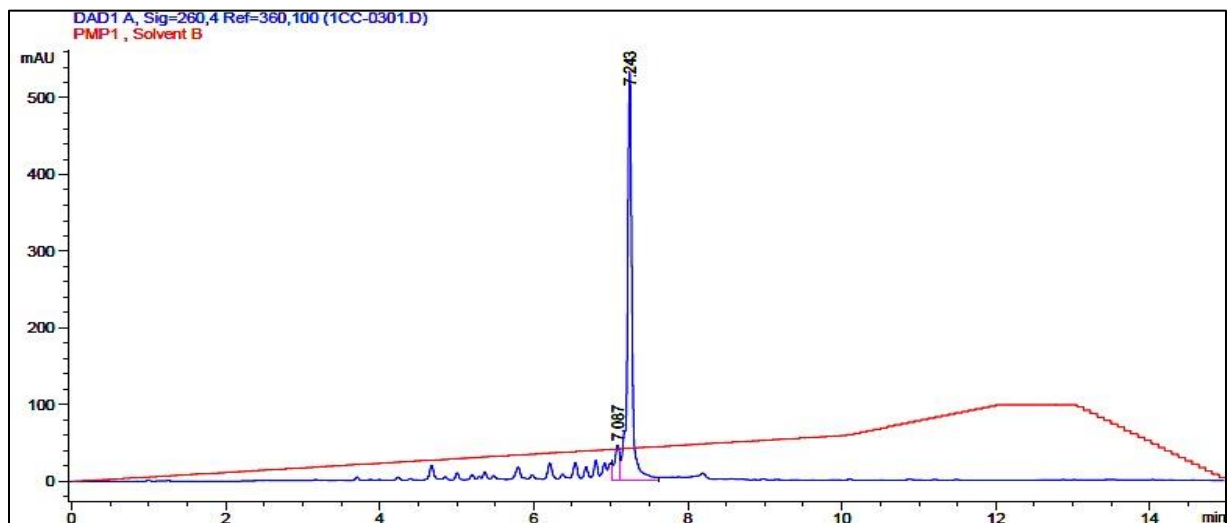


Figure 3.25: Analytical IE-HPLC chromatogram of AC₇ (control) without prior GP purification.

5'-DMT of the control was eliminated using acetic acid and then purified using sodium acetate/ ethanol precipitation (top) and ethyl acetate phase extraction (bottom). IE-HPLC was run using a gradient of 0-60% Buffer B at 55°C for 10 min.

3.3.2. Analysis of unlabeled switch molecules

The unlabeled switches (Figure 3.22) and their controls (Figure 3.23) were synthesized with the DMT group retained on the last base of the third strand. They were removed from the solid support column and deprotected using ammonium hydroxide for 17 h. Before being purified with GP, a portion of each sample was treated with 20% acetic acid in water to deprotect DMT for comparison with GP purified sample. These two controls were used to determine the retention time for the synthesis's first part and compare it with the full-length switches. In the HPLC chromatogram of the control sample (Figure 3.27), minimal failures are observed, which supports our hypothesis that most failures were from part II of the synthesis.

A full-length switch with many failure sequences is observed in the analytical IE-HPLC of a full-length unlabeled switch without GP purification (Figure 3.26). By utilizing prior GP purification, nearly all truncated or deletion sequences are effectively removed. This can be seen in the comparison of the control for Switch 1 (Figure 3.27) and the result from the complete two-part synthesis (Figure 3.28). A minimal number of failures were observed in the first part of the synthesis.

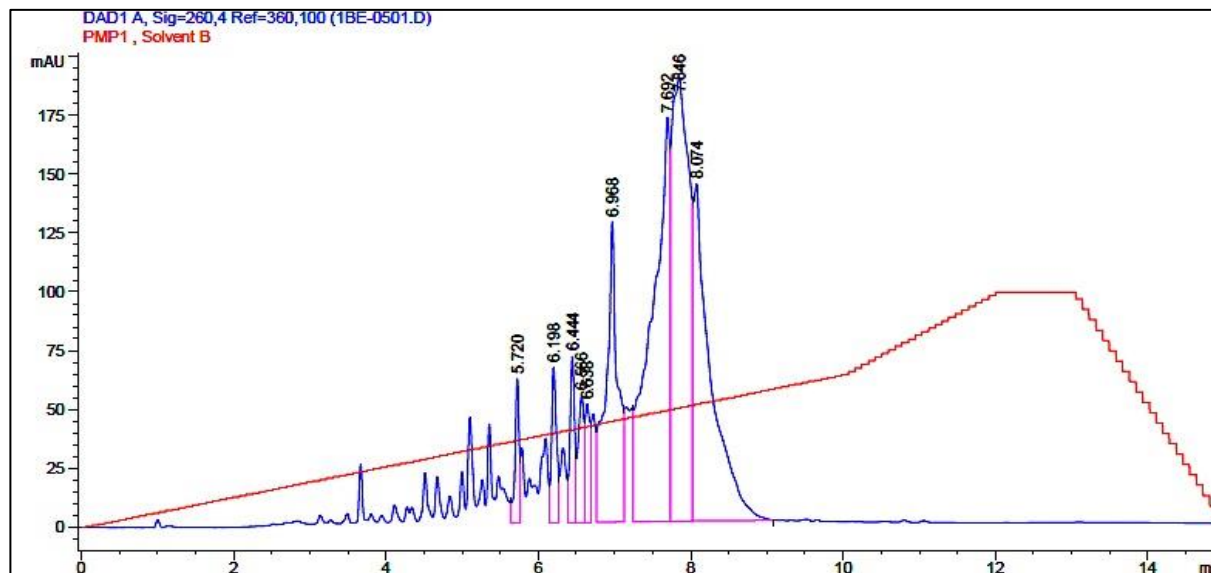


Figure 3.26: Analytical IE-HPLC analysis of full-length unlabeled Switch 1.

A full-length unlabeled Switch 1 without GP and many truncated sequences peaks were observed where DMT was removed with acetic acid and ethanol/sodium acetate precipitation.

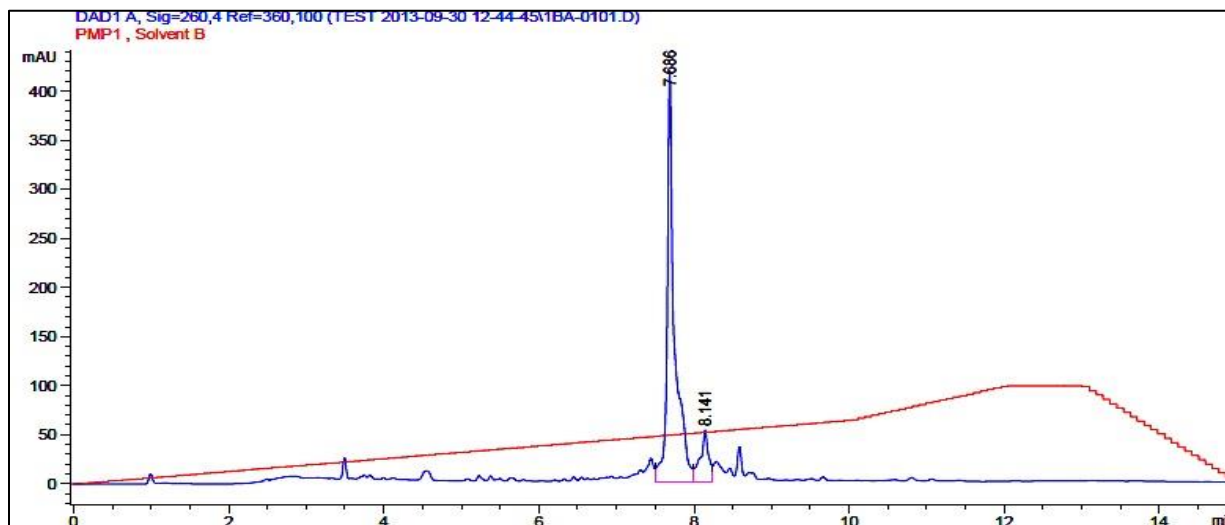


Figure 3.27: Analytical IE-HPLC of the control for unlabeled Switch 1.

A gradient of 0-60% buffer B at 55°C in 10 min was used. The retention time for the synthesis's first part (control) is 7.686 min.

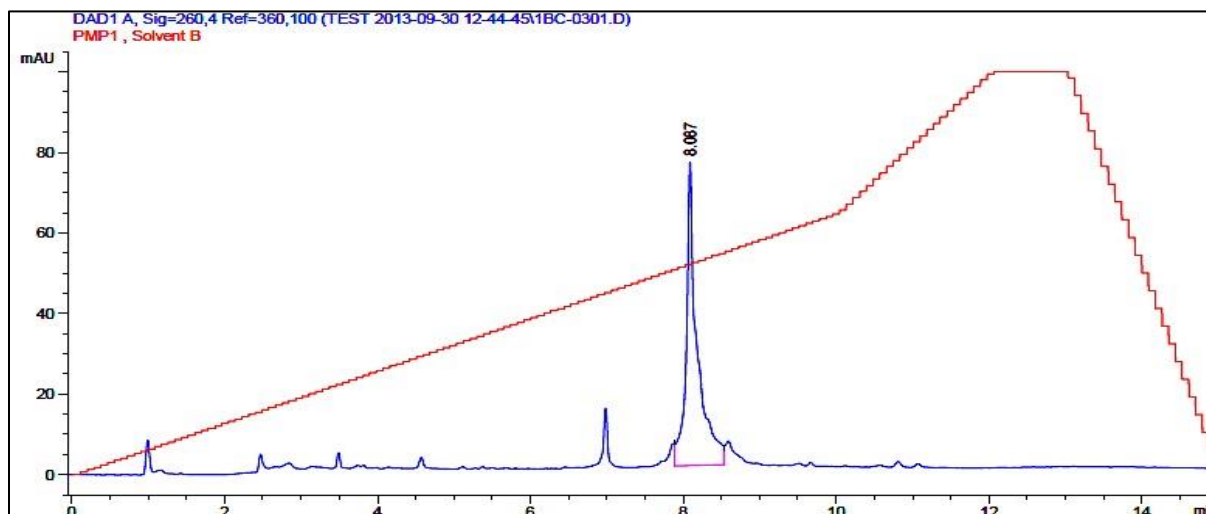


Figure 3.28: Analytical IE-HPLC of full-length unlabeled Switch 1.

A gradient of 0-60% buffer B at 55°C in 10 min was used to analyze after GP purification. The retention time of the full-length switch (switch 1) is 8.087 min.

3.3.3. Analysis of CY3 labeled oligonucleotide

Efficient purification by GP on the oligonucleotide labeled with indodicarboxyleyanine (CY3) dye is necessary because the last base at the 5' end of the third segment to be synthesized will have CY3 phosphoramidite coupled. After synthesizing a 5'-CY3 labeled 20mer, the oligo was removed from the solid support and deprotected in ammonium hydroxide for 17 h. For purification by GP, an additional salt wash was performed in place of the removal of the DMT step because it is the hydrophobic CY3 that was used to attach to GP instead of DMT. A second salt wash may help to remove any remaining failure sequences or excess dyes which are non-specifically bound to the GP column. CY3 labeled oligo was then eluted using 80% acetonitrile in water with 0.5% ammonia. By comparing the analytical IE-HPLC chromatogram (Figure 3.29 and 3.20), only a clean peak with oligo signal and CY3 signal was observed, meaning the GP cartridge was very efficient in removing failure sequences.

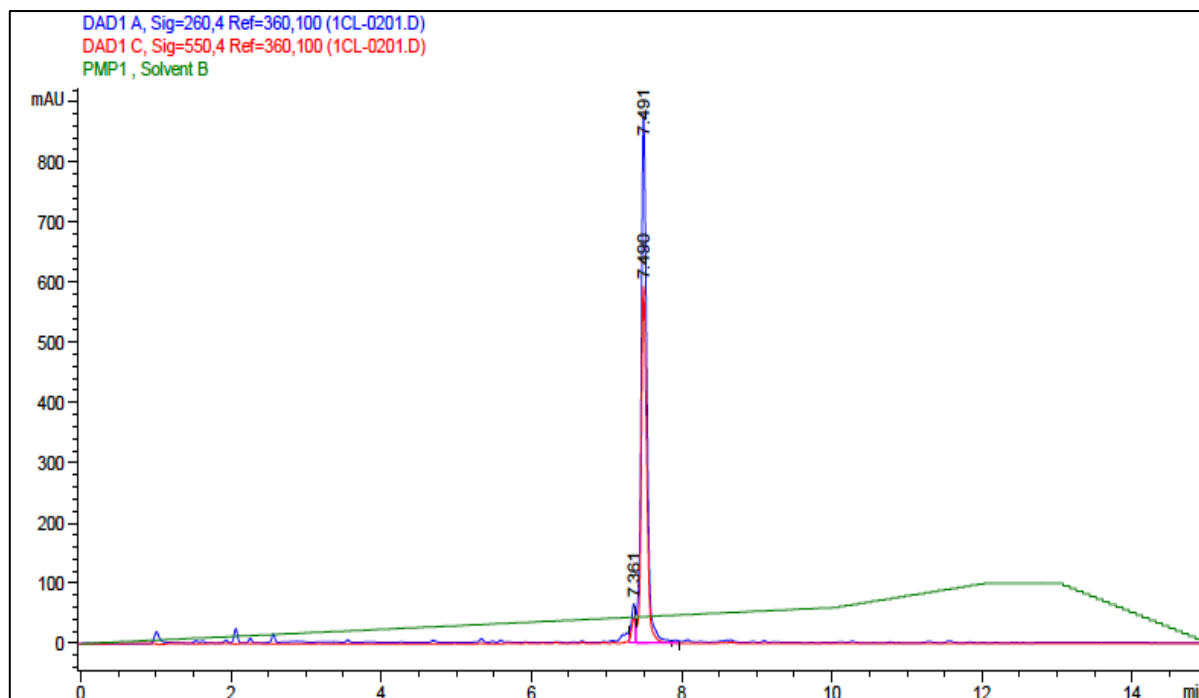


Figure 3.29: IE-HPLC analysis of CY3 labeled linear strand using GP purification.

Minimal to no failure sequence peaks were observed. The run used 0-60% Buffer B in 15 min at 55°C. The UV spectrum for oligo at 260nm is shown in blue, while CY3 at 550nm is shown in red.

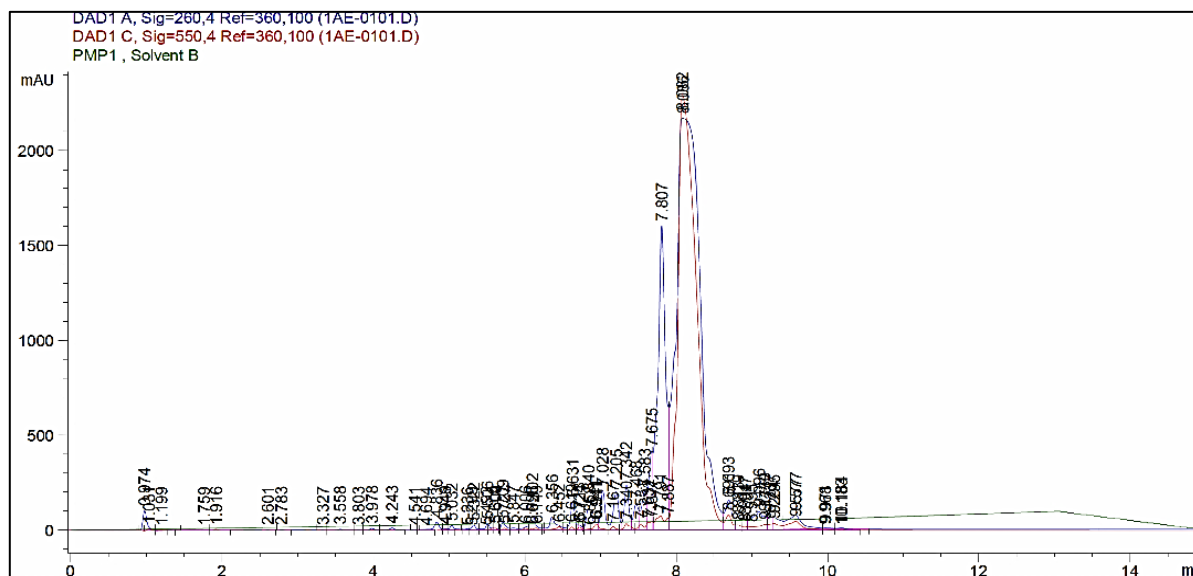


Figure 3.30: Analytical IE-HPLC analysis of CY3 labeled 20mer.

An aliquot of synthesized CY3 labeled sequence was used to compare with the same oligo purified by GP (Figure 3.29). The run used 0-60% Buffer B for 15 min at 55°C. The UV spectrum for oligo at 260nm is shown in blue while CY3 at 550nm is shown in red. Several failure sequences are observed, and the main band appears to have shoulders in the UV that do not appear at 550 nm.

3.4. Alternative synthesis of switch molecule using 3'-PT-amino-C6 Modifier

As both CY3 and DABCYL dyes are hydrophobic, subjecting a switch containing both fluorophore and quencher to a GP column will be problematic. Because a 3'-DABCYL column began the synthesis, all the failure sequences will have DABCYL that can adhere to the column support of the GP cartridge. This would defeat the purpose of using GP as the final purified product will contain both failures and desired outcomes.

Therefore, a 3'-PT-amino-C6 Modifier column (Figure 3.31 Top) was used in switch synthesis for the post-synthetic addition of DABCYL, and CY3 was added as the last step for the 5' end of the third segment on the synthesizer. The 3'-PT-amine modifier C6 was selected over 3'-amine or Fluorenylmethyloxycarbonyl chloride (FMOC) because the 3'-amino group is protected by a phthalimide (PT) ring attached to the support from an amide group attached to the aromatic ring.⁷⁴ This protecting group can withstand harsh conditions during oligonucleotide synthesis and can be removed using an extended ammonium hydroxide treatment at 55°C for 17 h. In addition, as this linkage prevents many side reactions from occurring during DNA synthesis, it should yield a higher purity product to conjugate with DABCYL (Figure 3.31 Bottom).⁷⁴

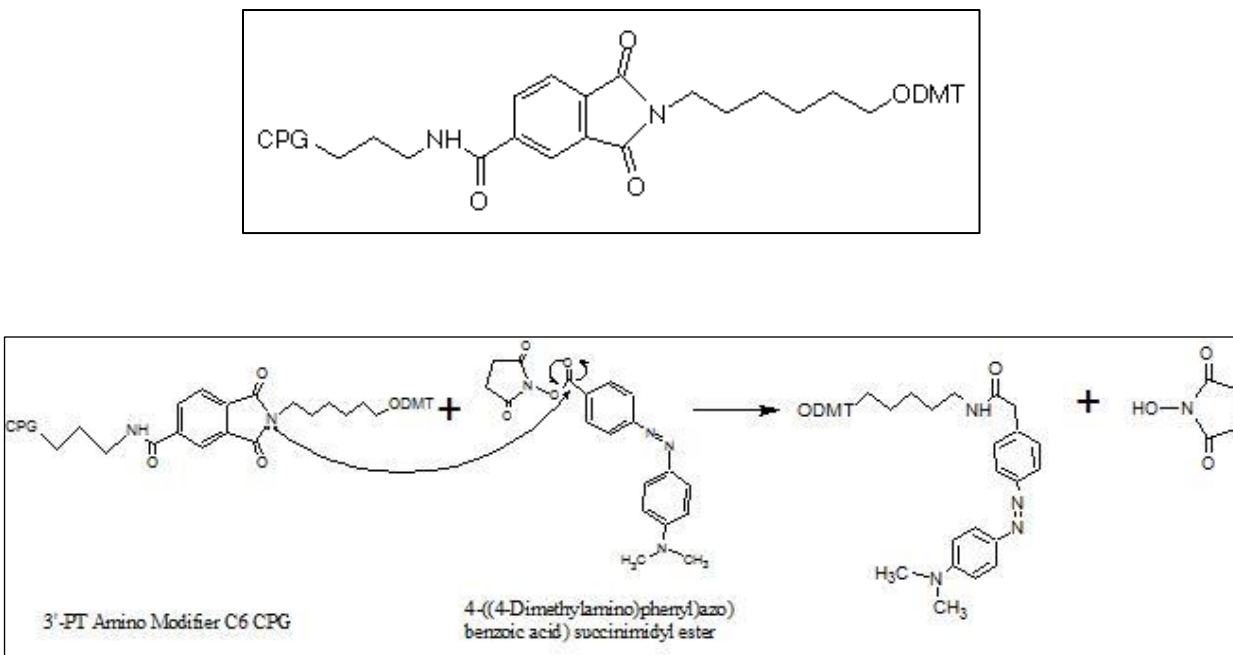


Figure 3.31: Conjugation of DABCYL.⁷⁴

(Top): Structure of 3'-PT- amine modifier C6 CPG column (Glen Research)

The 3'-amino group is protected by a phthalimide (PT) ring attached to the support from an amide group attached to the aromatic ring.⁷⁴ This protecting group can withstand harsh conditions during oligonucleotide synthesis and can be removed using an extended ammonium hydroxide treatment at 55°C for 17 h.

(Bottom): The chemistry of DABCYL succinimidyl ester and 3'-PT-amine. (Glen Research)

The conjugation of the 3'-PT amine of the oligo to the DABCYL succinimidyl ester.

Reprinted the structure of 3'-PT-amine modifier C6 CPG column and the chemistry of DABCYL succinimidyl ester and 3'-PT amine from Glen Research.

3.4.1. Conjugation of DABCYL succinimidyl ester with unlabeled 3'amine oligonucleotide

To verify the feasibility of the conjugation, a single strand (20mer) oligonucleotide (Figure 3.32) from Marras. et.al was synthesized using a 3'-PT- amine modifier C6 CPG column to perform DABCYL succinimidyl ester conjugation. To test the efficiency of DABCYL conjugation, a hairpin molecule (51mer) was synthesized using a 3'PT-amine column (Figure 3.32 bottom). A double-stranded DNA base pairing switch may pose a steric hindrance and interfere with conjugation. The oligo synthesis was carried out the same way as previously described, and DABCYL was conjugated post-synthetically by utilizing the 3'-PT-amine terminus of the oligo and DABCYL succinimidyl ester.

After the synthesis, the oligo with 5'-CY3 labeled will be purified using GP, and then DABCYL succinimidyl ester will be conjugated at the 3'-PT-amine. Conjugation of DABCYL succinimidyl ester with 3'-amine oligonucleotides was performed according to the procedures from Tabatadze et.al.²³ An excess of DABCYL succinimidyl ester in DMF was added to ensure a high yield of the conjugated product; DNA was purified using Lithium Perchlorate in acetone precipitation of conjugated oligonucleotide.

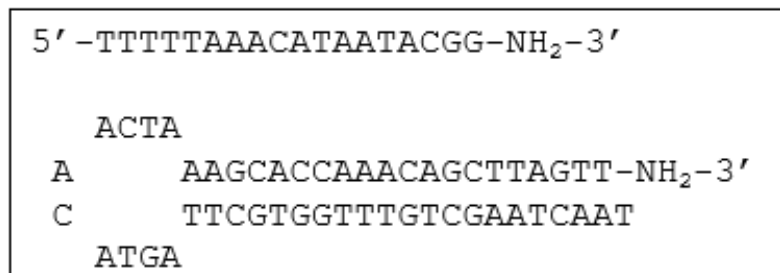


Figure 3.32: Sequences for Conjugation of DABCYL succinimidyl ester.

(**Top to bottom**): A 3'PT amine modifier C6 CPG synthesis column was used to make a linear 20mer oligonucleotide and a 51mer hairpin molecule.

3.4.2. Synthesis of the truncated switch to compare the efficiency of the DABCYL column and 3'-PT amine columns

To compare the purity of the putative switch, truncated switches utilizing a 3'-DABCYL and 3'-amine column (Figure 3.33) were synthesized. The synthesis started with 3'-DABCYL or 3'-amine columns in a 0.2 μ mol synthesis cycle.

After the first part of the synthesis, the 5'-OH of the primary strand was capped by acetate using the capping program in the synthesizer. Then the levulinyl group was deprotected using hydrazine hydrate treatment. Part II of the synthesis was carried out using the modified procedure, including increased coupling time and double base addition, with the last base at the 5' end labeled with CY3. After deprotection using 10% DEA and ammonium hydroxide, only a 3' amine labeled switch was used to perform GP purification. The purified sample was then conjugated with DABCYL succinimidyl ester, as described in the previous section. For both DABCYL conjugated and the oligo directly connected to 3' DABCYL, analysis of RP-HPLC was performed, and the peak containing absorption at 260nm (UV in blue), 550nm (CY3 in red), and 475nm (DABCYL in green) was purified.



Figure 3.33: Short Branched Sequences for Conjugation of DABCYL Succinimidyl Ester Analysis.

(Top to Bottom): A truncated switch sequence built from a 3'-PT-amine-C6 modifier CPG column, and conjugation with DABCYL and synthesis starting from a 3'-DABCYL CPG column.

Results and Discussion

3.4.1. Analysis of Conjugation of DABCYL succinimidyl ester with unlabeled 3'amine oligonucleotide

A single strand (20mer) oligonucleotide (Figure 3.32 top) was synthesized using 3'-PT-amine modifier C6 CPG column to perform DABCYL succinimidyl ester conjugation. Since the amine is deprotected by ammonium hydroxide at the end of the synthesis, it should not affect the coupling efficiency of the synthesis. Moreover, the impact of base deprotection in saturated aqueous ammonia by either heating at 55°C for 17 h or incubating at room temperature for 24 h was examined to optimize the yield of the product.

Analytical RP HPLC was performed on samples with different incubation times: 24 h of incubation at room temperature showed higher intensity and sharp peaks, which can be explained by the benefit of extra deprotection time for the phthalimide-protecting group. From the analytical RP-HPLC of the DABCYL conjugated sample (Figure 3.34), about 70% of the sample was successfully conjugated; this product was purified to analyze with ESI-MS. Analytical RP-HPLC of a hairpin molecule, 51mer (Figure 3.32 bottom), was run and proved that base pairing in double-stranded DNA did not significantly affect DABCYL conjugation (Figure 3.35). Both purified 20mer and 51mer were sent to Novatia, LLC, for ESI-MS analysis (Table 3.5).

From the ESI-MS analysis, the molecular weights of a linear 20mer and a branched 51mer were in an acceptable mass error range of $\pm 0.03\%$ with more than 80% purity, showing that DABCYL succinimidyl ester conjugation is a promising approach.

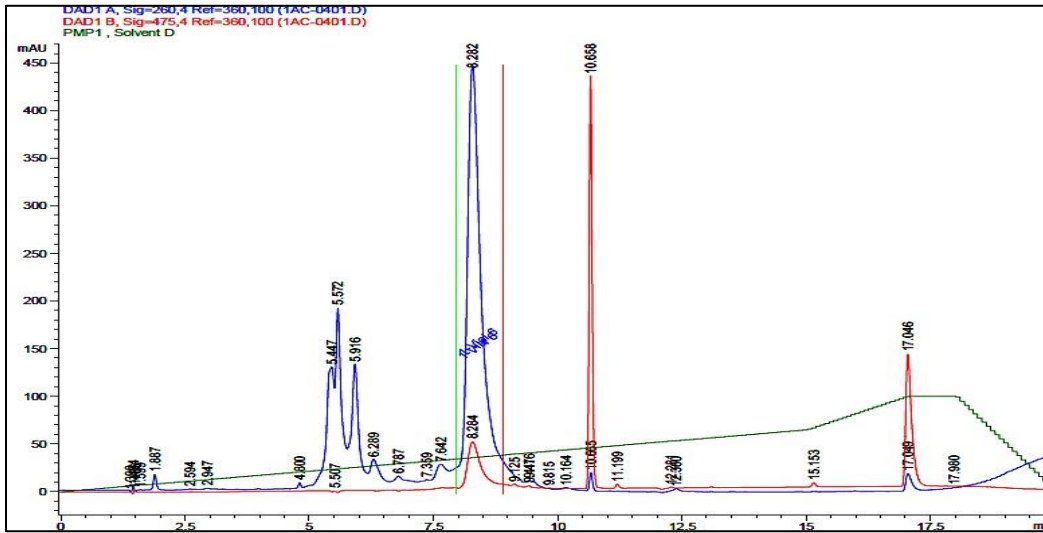


Figure 3.34: RP-HPLC purification chromatogram of DABCYL conjugated 20mer.

DABCYL conjugated 20mer was purified for MS analysis using a gradient of 0-60% acetonitrile at 55°C for 15 min. UV absorption at 260nm is shown in blue and DABCYL at 475nm is shown in red.

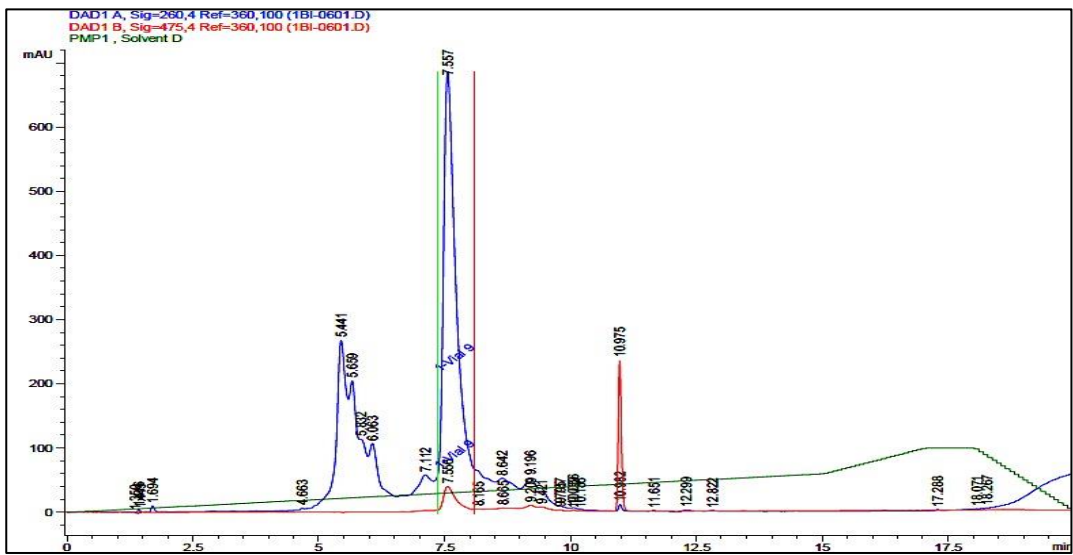


Figure 3.35: RP-HPLC purification chromatogram of DABCYL conjugated 51mer.

DABCYL conjugated 20mer was purified for MS analysis using a gradient of 0-60% acetonitrile at 55°C for 15 min. UV absorption at 260nm is shown in blue and DABCYL at 475nm is shown in red.

Sample	Targeted Mass	Observed mass	% Purity
20mer-Dab	6544.5	6543.4	87.40
51mer-Dab	16109.6	16108.5	81.08
NH-Dab switch	11776.3	11777.1	58.38
Dab switch	12390.6	12391.3	44.08

Table 3.5: ESI-MS analysis of DABCYL conjugated sample.

ESI-MS analysis of truncated switches with 3'-amine column (top) and 3'-DABCYL columns (bottom)

3.4.2. Analysis of a truncated switch to compare the efficiency of the DABCYL column and 3'-PT amine column

The efficiency of the 3'-DABCYL column and the 3'-PT amine column were compared by synthesizing truncated switch molecules (Figure 3.33). After the synthesis, 5'-CY3 labeled oligos were purified using GP columns. The purified samples are then conjugated with DABCYL succinimidyl ester in the same way described in the previous section. For both DABCYL conjugated and 3' DABCYL oligonucleotide, analytical RP-HPLC was performed (data not shown). For mass-spec analysis, the peak containing absorption at 260nm (UV in blue), 550nm (CY3 in red), and 475nm (DABCYL in green) were purified (Figure 3.36).

In the ESI-MS analysis of both samples (Table 3.5), the molecular weights were in an acceptable error range of $\pm 0.03\%$. From the ESI-MS chromatogram of DABCYL conjugated oligonucleotide (Figure 3.36), the impurities on the left and right sides of the main peak on the spectra contributed about 35-40% of the samples, which could be due to insufficient capping (left side of the main peak at 11,777) and incomplete deprotection (right side). After examining the overall and stepwise yield from the synthesizer, coupling efficiencies for G and C bases, just the C and G closest to © decreased by about 10% in making both the conjugated and the 3' DABCYL switches. In the switch synthesis, the unreacted detritylated nucleosides are acetylated to prevent them from participating in the coupling reaction of the next cycle. If the capping efficiency were low, the failure sequences could still couple to the next coming base, resulting in base skipping.

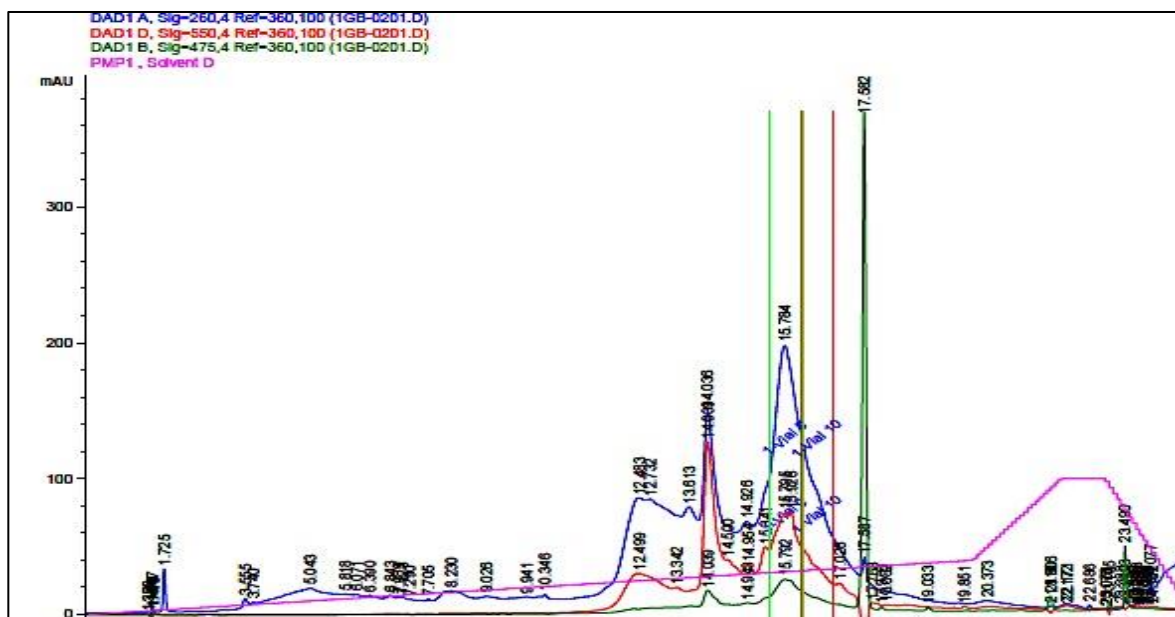


Figure 3.36: RP-HPLC purification chromatogram of DABCYL conjugated 51mer.

DABCYL conjugated 51mer was purified for MS analysis using a gradient of 0-60% acetonitrile at 55°C in 15 min, where the peak containing all three absorbances was collected for ESI-MS analysis. Table 3.5: ESI-MS analysis of truncated switches with 3'-amine column (top) and 3'-DABCYL columns (bottom)

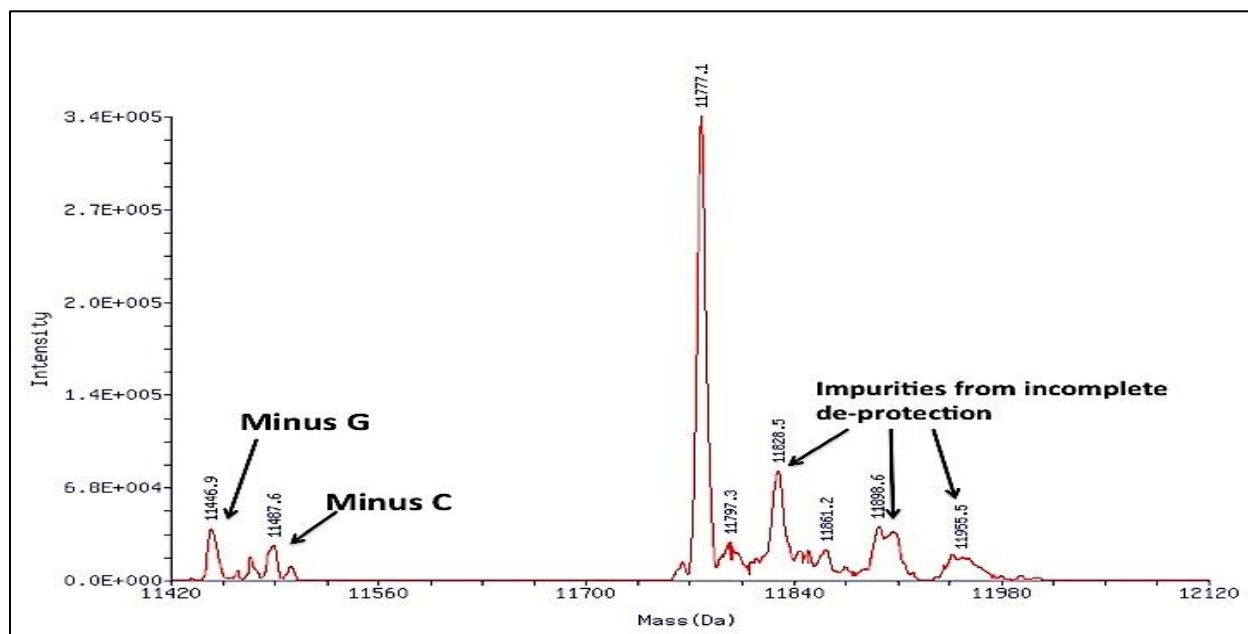


Figure 3.37: ESI-MS chromatogram of DABCYL conjugated oligonucleotide.

Impurities are identified on the main peak's left and right sides. contributed about 35-40% of the samples, which could be due to insufficient capping (left side of the main peak at 11,777) and incomplete deprotection (right side).

Materials and Methods

3.1. Direct Synthesis of a switch using 3'-DABCYL CPG

Synthesis of a branched switch molecule

Phosphoramidites, DABCYL-controlled pore glass (CPG) columns, CY3, 5 Me-dC-brancher, and other required solvents for ABI 392/394 synthesizer were purchased from Glen Research. The synthesizer was cleaned extensively using NH₄OH and ACN before the switch synthesis. The required phosphoramidites, CY3, and brancher are diluted to 0.1 M using diluent ACN. The synthesis was performed as a two-part synthesis (0.2 μm scale) in the dark because of the light sensitivity of DABCYL and CY3.

Direct synthesis of a branched switch was carried out in two-part.

3'-DABCYL-TTGATTCGACACCACGAATCAA©ATGATGCGTGGTGTCCGCATCATT-5' was synthesized in part 1 of the synthesis, and the 5' end nucleotide (probe) was capped using 1 mL of Cap A and B mixture for 30 min, and then the CPG columns were dried under argon. The levulinylyl group on the brancher was then deprotected using 0.5 M hydrazine hydrate in 1:1 pyridine: acetonitrile for 15 min. The CPG columns were washed with 1.5mL ACN 3 times and dried under argon before putting back into the system. This step is required to form a free hydroxyl group that allows the second part of the synthesis to begin. Toggle sequence, ATTGATTCGTCACTCGAATCAA-CY3-5' was synthesized, and the CPG column was removed from the synthesizer for extra deprotection workup. To eliminate the β-cyanoethyl protecting group, 10% DEA in acetonitrile is used for 10 min and then washed using 1mL ACN (HPLC grade) three times and dried under argon. Deprotection from the solid support was done by using 1 mL 30% NH₄OH for 2 h, followed by 0.5 mL NH₄OH wash, and the solution was stored at room

temperature for 17 h. The samples were dried in a SpeedVac DNA concentrator 120 (Thermo Scientific) and resuspended in 150 μ L Milli-Q purified water (Millipore Corp.). We gratefully acknowledge prior work by R. Iyer and M. McPike and their collaboration in accomplishing the above procedures.

HPLC purification

The samples were desalted using MicroSpin G-25 Columns (GE Healthcare), and a nanodrop UV absorption spectrometer was used to calculate the concentration. The samples were heated at 90°C for 3 min and snap-cooled to eliminate the formation of the duplex structure before HPLC purification on an Agilent 1200 Analytical HPLC equipped with a Dionex DNAPac PA200 4x250 mm analytical IE column. Buffer A (0.01M Tris-HClO₄, 20% ACN, pH:8) and buffer B (0.01M Tris-HClO₄, 0.33M NaClO₄, 20% ACN, pH:8) were made and filtered for Ion exchange HPLC. Ion exchange purification was carried out using multiple 5 μ L injections at 55°C, 1.2 mL/minute flow rate, 0-60% gradient of HPLC buffer B for 10 min and 2 min post gradient conditions for an overall time of 17 min. The peaks were collected using a time-based method, and the samples were dried in the Speed-Vac and then desalted using Econo-Pac® 10DG columns (Bio-Rad). The sample was dried by the Speed-Vac and resuspended with 150 μ L DNase-free water. 100 μ L stock solutions of the switch were prepared for purified peaks, and analytical HPLC of purified peaks was run for validation.

Mass Spectrometry

1 nM of each sample of the peak from switch synthesis with strong absorbance at all three wavelengths (DNA – 260nm, CY3 – 550 nm, and DABCYL – 475 nm) was prepared for characterization at Novatia using electrospray ionization liquid chromatography-mass

spectrometry (ESI/LC/MS). The molecular weight of the potential switch was calculated using the Oligo analyzer tool from the Integrated DNA technology (IDT) website, and 1nmol of the sample was sent to perform mass spec characterization (Novatia, LLC)

3.2. Optimization of switch synthesis

Test sequences using dT phosphoramidites for optimization of switch synthesis

Sequences (Figure 3.6, 3.8 and 3.10) were synthesized on standard 0.2 μm scale dT CPG columns and 5 Me-dC brancher phosphoramidites. The synthesis was carried out in either one part (Figure 3.8) or two parts (Figures 3.6 and 3.10). After the synthesis, the β -cyanoethyl protecting group is removed using 3 mL of 10% DEA in acetonitrile is used for 3 min and then washed using 1mL ACN (HPLC grade) three times and dried under argon. The synthesized sequence with only dT phosphoramidites can be deprotected from the solid support by using 1 mL 30% NH_4OH for 2 h at RT, then followed by 0.5 mL NH_4OH wash and heated at 65°C. The samples were dried in a SpeedVac DNA concentrator 120 (Thermo Scientific) and resuspended in 150 μl Milli-Q purified water (Millipore Corp.) and a desalting procedure using MicroSpin G-25 Columns (GE Healthcare) was carried out.

Synthesis of non-branched DNA hairpin loops

DNA Phosphoramidites, DABCYL-controlled pore glass (CPG) columns, CY3, 5 Me-dC-brancher phosphoramidites, and other required solvents for the synthesizer were purchased from Glen Research. Before the switch synthesis, the synthesizer was cleaned extensively using ammonium hydroxide and acetonitrile (ACN). The phosphoramidites, CY3, and brancher were diluted to 0.1M using diluent ACN. The synthesis was performed using a 0.2 μm scale, and 5'-DMT of the last base was kept, which means that the final detritylation step was not included. In

deprotection of the β -cyanoethyl group, a by-product can be formed that reacts with DNA. Thus, CPG was treated with 10% DEA in acetonitrile for 10 min, then washed using 1 mL ACN (HPLC grade) three times and dried under argon. Deprotection from the solid support was done by using 1 mL 30% of NH_4OH for 2 h, followed by 0.5 mL NH_4OH wash, and the solution was incubated at room temperature for 17 h.

Synthesis of truncated switch

The synthesis was performed in the dark in two parts (0.2 μm scales). The first part of the synthesis was carried out traditionally. However, a capping step was done on the system using the end-cap program to reduce the contamination of CPGs. Fresh hydrazine hydrate was prepared, and pyridine was redistilled a day before and dried on molecular sieves. The levulinyl group on the brancher was then deprotected using 0.5 M hydrazine hydrate in 1:1 pyridine: acetic acid for 15 min. The CPG columns were washed twice with 1.5 mL of ACN and dried under argon before putting back onto the synthesizer. This step is required to introduce a free hydroxyl group that allows the final part of the synthesis to begin. After the levulinyl group was deprotected, an extra washing step on the synthesizer was performed prior to the last part of the synthesis. Double additions of phosphoramidites were included in the synthesizer program to ensure a high coupling yield. For deprotection, 10% DEA in ACN (diluent) was used, followed by NH_4OH deprotection as described in the prior procedure. The samples were heated at 65°C for 2 h to accelerate the base-deprotection process and evaporated to dryness using a SpeedVac and HPLC analysis.

3.3. Optimization of the Purification of Switch Molecules

Glen-Pak (GP) purification

The ammonia-containing oligonucleotide samples were mixed with equal 100 mg/mL NaCl volumes. GP cartridges were prepped using 0.5 mL ACN and 1 mL 2% TEAA (2.0 M pH 7.0) with a flow rate of 2 drops per second to equilibrate the column using a 5 mL syringe. The ammonia salt solution was then loaded to the column, and only DMT-containing oligos adhered to the support, while failure sequences were rinsed off the column using 2mL of 100mg/mL NaCl. Then DMT was removed using 2 mL of 2% TFA and another wash step with 2mL of deionized water. The purified sample was eluted from the column using 80% acetonitrile in water with 0.5% ammonium hydroxide. The samples were dried in a SpeedVac DNA concentrator 120 (Thermo Scientific) and resuspended in 150 μ L Milli-Q purified water (Millipore Corp.).

3.4. Alternative synthesis of switch molecule using 3'-PT-amino-C6 Modifier

Synthesis of 3'-amine-labeled oligonucleotides

Phosphoramidites, 3'-PT-amino-C6 Modifier controlled pore glass (CPG) columns (Glen research) were used, and the synthesis was performed the same way described above. However, the phthalimide-protecting group required a longer ammonia deprotection time (24 h at room temperature). The samples were dried in a SpeedVac DNA concentrator 120 (Thermo Scientific) and resuspended in 150 μ L Milli-Q purified water (Millipore Corp.). 15 μ L of 3 M sodium acetate (pH 5.2) and 450 μ L of cold ethanol were added to the samples and vortexed well. The samples were incubated at -20°C overnight and centrifuged at max speed at 4°C for 30 min. The supernatant was then decanted, and the pellets were washed with 70% ethanol and air-dried for 15 min. Analysis of IE-HPLC was performed using a gradient of 0-60% buffer B (0.01 M tris-HClO₄, 0.33 M NaClO₄, 20% acetonitrile; pH 8.0) at 55°C over 10 min.

Conjugation of the 3'-amine oligonucleotide with DABCYL N-hydroxyl succinimidyl (NHS) ester

The conjugation was performed using the protocol by Tabatadze. et al.²³ 15 nmol of the 3'-amine sample was resuspended with 40 μ L of 0.1 M NaHCO₃ buffer (pH 9.0). Then 10 μ L of saturated DABCYL NHS ester (Invitrogen) in DMF (1 mg/60 μ L) was added over a 20 min interval with shaking. The sample was incubated at room temperature overnight with agitation so that DABCYL NHS ester did not precipitate. The excess DABCYL was removed by precipitating the DNA with 100 μ L of 2 M LiClO₄ and 0.7 mL of cold acetone.

Chapter 4: Synthesis, ligation, and characterization of oligonucleotide switches

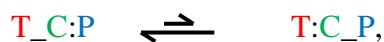
Chapter Summary

Synthetic oligonucleotides have been of great interest in many biological applications including DNA sequencing, genotyping, and quantitative polymerase chain reaction (PCR). Aptamer switches for molecular diagnostic and drug discovery are of considerable interest in recent years and can be used to detect proteins, small organic molecules, or potential drug candidates.⁷⁵ Numerous studies are done on the possibility of DNA aptamer-based molecular switches which undergo conformational changes to target/ ligand binding enabling quantitative in situ imaging, programmed tumor-targeting⁷⁵ and rapid clinical diagnostics.^{50,75} Direct synthesis and HPLC purification of a 68mer full-length switch molecule (NM-0) described in Chapter 3 did not achieve a high-purity product.

In this chapter, a 3-segment switch was prepared and gave proof-of-principle as a fluorescent molecular switch that turns OFF upon adding the HIV-1 nucleocapsid protein, NC. Although not tested in this work, the switch would turn ON when a drug candidate binds to the aptamer binding pocket of NC and displaces the switch (see the discussion of figures 2.1 and 2.2). Drug candidates screened in this fashion could be tested for the development of HIV inhibitors, and the platform can be adapted for high throughput screens of other drug targets.

As described in Chapter 2, the three active segments of the switch are Cover, Probe, and Toggle segments connected through a commercially available 5-methyl-deoxycytidine (5Me-dC) brancher (see Figure 3.2). The brancher joins Cover and Probe via standard 3'-5'-phosphodiester linkages; a linker (-C₆-O-) off cytosine-N4 is esterified to the 5'-phosphate of the next amidite

added on the synthesizer (the first nucleotide (nt) of Toggle). Probe contains a high-affinity aptamer for the HIV-1 nucleocapsid protein (NC), and Cover sequesters the critical aptamer binding sequence of the Probe in base pairs. Toggle competes with Probe for binding to Cover in rapid unimolecular equilibrium, with an equilibrium constant, K1.



$$K1 = \frac{[T:C_P]}{[T_C:P]}$$

Here C:P and T:C are stemmed with extensive base pairing to Cover, and T_ and _P are the isolated Toggle and Probe stem-loop hairpins. Using estimates of the ΔG° of folding for the two species (described in chapter 2), we manipulated the sequences of the three segments to create a switch with $K1 = 0.01$ to 0.001 , in the absence of NC. This favors the concentration $[T_C:P]$, about 100 to 1,000-fold over $[T:C_P]$. Adding increasing concentrations of NC binds _P, which drags the K1 equilibrium to the right. As the concentration of $[T:C_P\text{-NC}]$ increases, a DABCYL modifier at the 3'-end of Cover quenches the fluorescence of CY3 at the 5'-end of Toggle in the T:C stem.

However, it is quite challenging to design and execute the synthesis of switch molecules due to their chain length and labeling with both fluorophore and quencher. Isolation of oligonucleotides with more than 50 nucleotides in high purity (>90%) is difficult, even after optimization of switch synthesis and purification by IE- or RP-HPLC.

This chapter also describes the efficient use of T4 DNA ligase to join two fragments, a 55mer Core and a 15mer CY3-Toggle fragment prepared by solid-phase DNA synthesis. The design and successful synthesis are described in sections 4.1-4.4. Section 4.4 describes the ligation

of the Core and CY3-Toggle fragments using a DNA splint strand, followed by the isolation of high-purity switches by denaturing polyacrylamide gel electrophoresis (PAGE). Our ligation approach allows easy purification of full-length switches away from abundant failure chains that accumulate in solid-phase syntheses of long DNA molecules. Full-length Core and ligated switch molecules were analyzed by ESI mass spectrometry to assess purity. Truncated switches with dual labels were also prepared that are locked in the ON or OFF forms (see section 4.5).

Fluorescence measurements that are reported in section 4.7 demonstrate proof-of-principle for 3-segment nucleic acid switches. Increasing [NC] causes a dramatic decrease in CY3 fluorescence. Assuming that the decrease in fluorescence is caused by an increasing fraction of switch bound to NC affords an experimental estimate of the K_1 equilibrium constant, which is within the experimental error of the free energy estimates used to design these unprecedented DNA secondary structures.

4.1. Design criteria for switches

The sequence in Figure 4.1, called Switch NM-1 ($K_1=0.03$), is the result of trial-and-error changes that alter the individual sequences of Cover, Probe, and (mostly) Toggle segments, adjusting nearest-neighbor base pairs, loop sizes, introducing terminal mismatches, and other changes. The goal was to obtain a difference in free energy $\Delta\Delta G^\circ = \Delta G^\circ(\text{ON}) - \Delta G^\circ(\text{OFF})$ in the approximate range of -2 to -4 kcal/mol ($\Delta\Delta G^\circ = 1.36$ kcal/mol corresponds to a 10-fold change in K_1 at 25°C). Free energy estimates for standard secondary structure elements were made using visual-OMP (V-OMP) software. This version of the software includes the free energy contribution of dyes at the termini of stems (previously determined as -2 kcal/mol for CY3 opposite DABCYL³⁴).

The OMP database does not include three DNA segments joined at a vertex. Therefore, we simulated our unimolecular structures using duplexes with each strand at 1M (see simulation conditions at the left of Figure 4.2). The ON form in panel (a) of Figures 4.1 and 4.2 was approximated as an isolated Probe strand (P), and an isolated Cover-⊙-Toggle strand (CT), that forms the P:CT duplex as shown. The OFF form was simulated by a T:CP duplex, panel (b) using an isolated Toggle strand (T) and an isolated Cover-⊙-Probe strand (CP). This is an approximation, but it has served to generate working 3-segment switches with ON favored over OFF (“turn-off switch, this work), and OFF over ON (“turn-on” switch³⁴). Each of these switches exhibited the expected response upon increasing concentrations of NC protein – a large decrease in fluorescence for turn-off, and a large increase for turn-on switches.

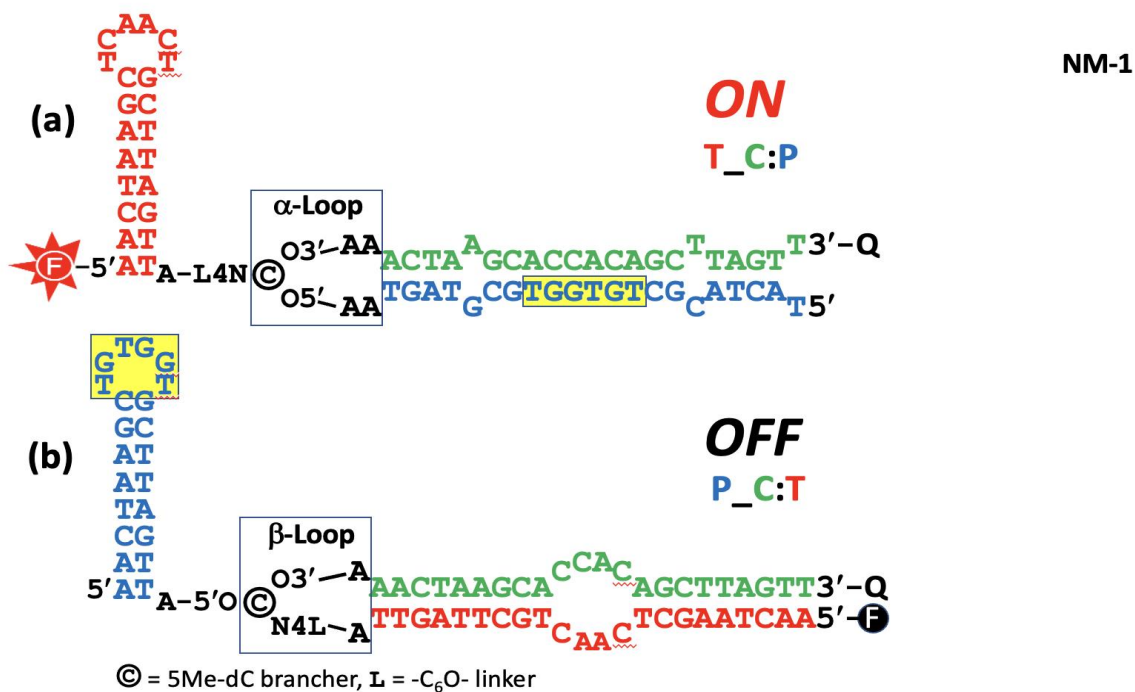


Figure 4.1: A 3-segment 70mer DNA switch, Switch NM-1.

Switch NM-1 is turned OFF upon the addition of the HIV-1 nucleocapsid protein. The segments are **Toggle** (red), **Cover** (green), and **Probe** (blue), joined at a 5Me-dC brancher (denoted by ©); important residues for binding NC are outlined in yellow. There are two different hairpin loops surrounding the brancher, an α -loop with normal 3'-5' linkages (panel a), and a β -loop with a 3'-dC-N4-L-O5' linkage (panel b), where L is -C₆O-. (Figure prepared by Dr. Borer.)

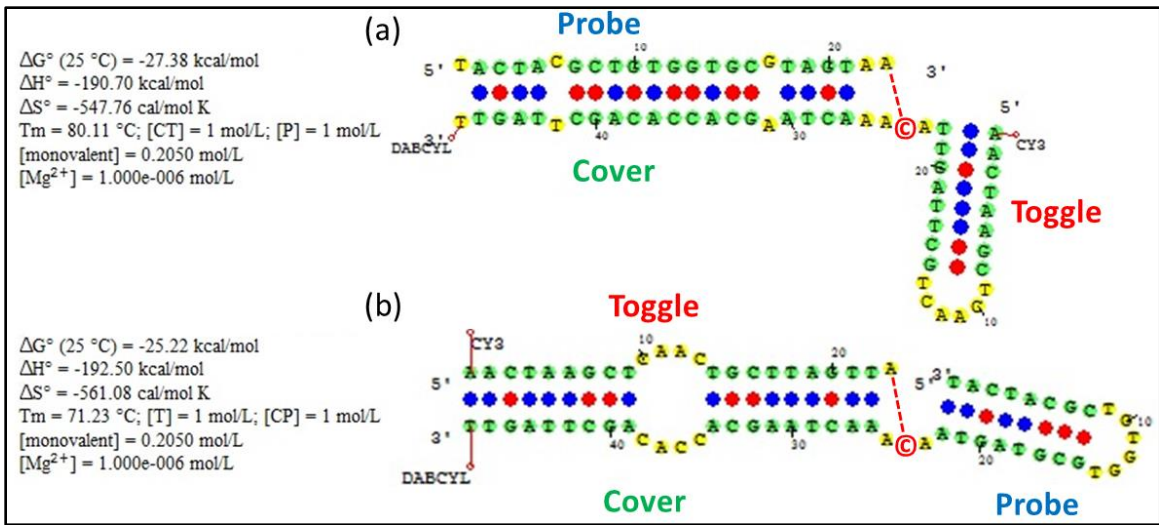


Figure 4.2: Output from a V-OMP simulation of Switch NM-1.

Switch NM-1 is in the ON form (a) and OFF form (b). The estimated $\Delta\Delta G^\circ = -2.16 \text{ kcal/mol}$ corresponds to $K_1 = 0.03$. The brancher $\textcircled{\text{C}}$ is simulated by a standard dC residue; a red dashed line shows where a covalent connection to the brancher occurs in the actual 3-segment switch.

Another important approximation that is unique to the 3-segment switch is the nature of the two distinct hairpin loops that form around the brancher: α and β (see Figure 4.3 and the boxed region at the center of Figure 4.1). Dr. Iyer, in our lab, synthesized stem-loop models of α and β loops and determined their thermodynamic properties.³⁴ He found that both 3-base, A-C-A α and β loops in his switches had nearly identical contributions to the free energy of folding the loop and stem, even though there are many more rotatable covalent bonds in the β -loop than in α . Therefore, it wasn't necessary to account for a special free energy contribution for forming these loops; the two individual contributions will simply cancel in the subtraction to calculate $\Delta\Delta G^\circ$.

There is a further complication for the central hairpin loops in Figure 4.1 for Switch NM-1. While the β -loop is still a 3-base loop, the α -loop is a 5-base loop. ΔG° of formation is positive for hairpin loops and the destabilization generally increases with loop size (see Table 4 of reference⁶⁰) The 5-base α -loop has standard phosphodiester linkages throughout, as in SantaLucia's table, where the free energy difference between 3-base and 5-base loops is only about $\Delta\Delta G^\circ = 0.2$ kcal/mol. For DNA structures with about 20 nt the errors are about ± 1 kcal/mol in calculating the ΔG° of formation from the unstructured strands. The error bounds increase with the increasing size of the molecules. Neglecting a difference due to the loop size of $\Delta\Delta G^\circ = 0.2$ kcal/mol is unlikely to be significant for the free energy balance between the folded states of the 70mers in Figure 4.1. Further, the errors inherent in approximating $\Delta\Delta G^\circ$ from the constructs in Figure 4.2 lead to estimates of K1 that can easily differ by factors of 10 to 100 from the values determined from the fluorescence of real switches.

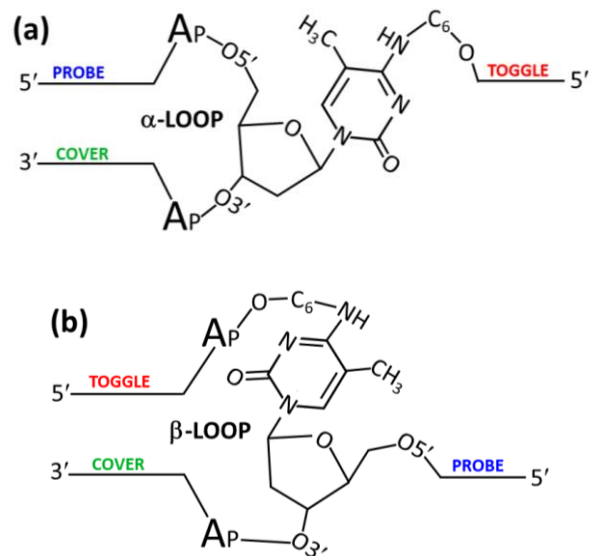


Figure 4.3: Covalent connections in 3-base α loops (a) and β loops (b).

Compare with

Figure 4.1. α -Loops have six rotatable bonds between the two Ap phosphates surrounding the brancher, while β -loops have fourteen.

4.2. Design of a switch molecule for enzymatic ligation

The overall design to assemble a 3-segment switch with T4 DNA ligase is easily understood for one-step ligation (shown) in Figure 4.4. Direct synthesis of (1) a DNA Core molecule begins with a 3'-DABCYL quencher amidite, followed by the full-length Cover, the brancher, and Probe. After removing the levulinyl protecting group on the brancher's cytosine N4-C₆-OH, an 8mer Toggle fragment is synthesized bearing a 5'-phosphate. Using the nomenclature for ligation, the Toggle donor is the 8mer fragment as it donates the 5'-phosphate to the CY3-Toggle acceptor bearing a free 3'-OH.

An option for two-step ligation was also explored and is described below and in the Supplementary Material to this chapter. While this option might find useful applications in the future, one-step ligation was used to prepare functional switches in this study. Some early studies of ligation and characterization of ligated products are also described in the Supplement. This includes preliminary phosphorylation and ligation attempts and attempts that failed to yield high-purity Cores.

V-OMP was also used to design ligation sites in the switch, along with appropriate splint strands. Figure 4.5 (a) shows our setup for one-step ligation. The 55mer Core surrounds the brancher, and the CY3-labeled Toggle acceptor (red 15mer). Panel (b) shows our setup for two-step ligation, which now includes the DABCYL-labeled Cover donor (blue 15mer).

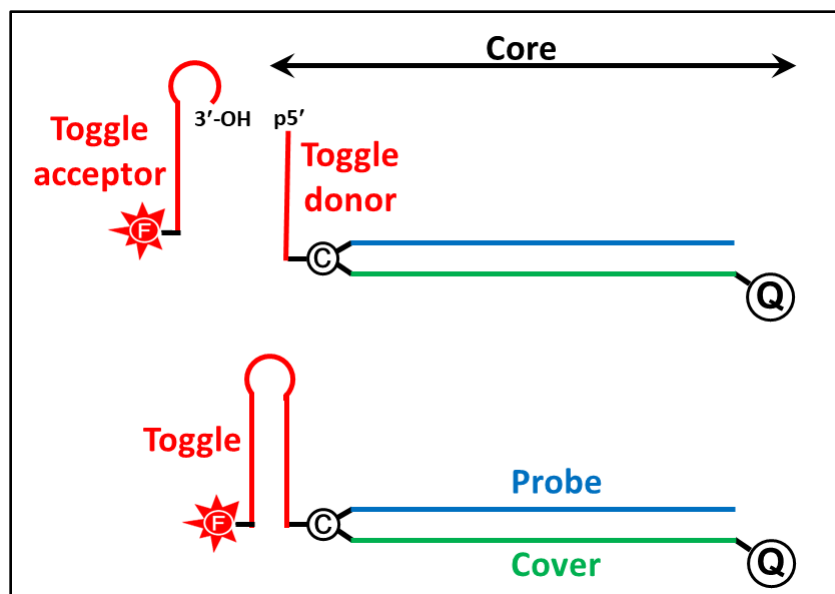


Figure 4.4: Schematic of One-Step Ligation.

Top: A 14mer Toggle acceptor is ligated to a 5' phosphorylated Core (55mer) containing an 8mer Toggle donor. Bottom: The product is a 70mer 3-segment switch. The quencher, Q, is DABCYL, and the fluorophore, F, is CY3. Sequences for each segment are shown in Figure 4.5. (Figure prepared by Dr. Borer.)

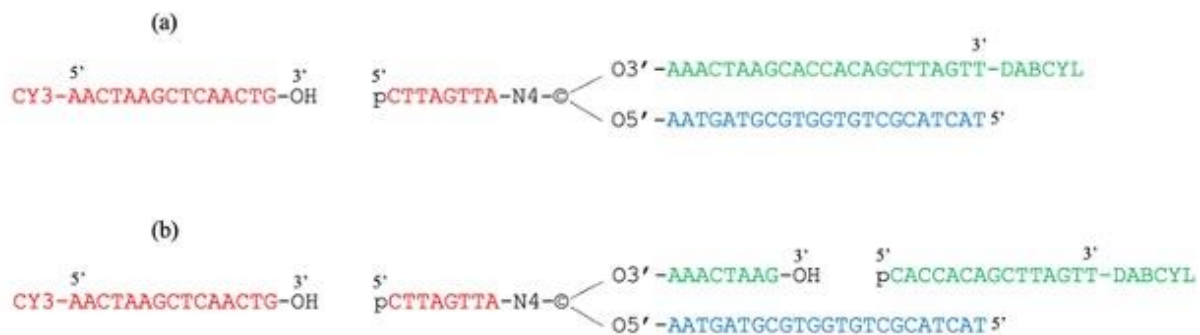


Figure 4.5: Switch NM-1 sequences for one-step (a) and two-step (b) ligation.

(a) The full 70mer NM-1 sequence is shown, including a gap for the ligation site used to join the 8mer Toggle donor to a 15mer Toggle acceptor. The 55mer Core is shown at the right, surrounding the brancher including the 23mer Cover and 23mer Probe segments. (b) In two-step ligation, the Toggle fragments are joined as in panel (a) and a second ligation joins a DABCYL-labeled Cover donor (15mer) to the 8mer Cover acceptor that is the first sequence synthesized in a 40mer Core.

4.2.1. Preparation of molecules for one-step ligation

Splint-assisted ligation works well for sites in unpaired regions, like the large internal loop toward the bottom right of panel (b) in Figure 4.1. To ligate Toggle fragments, the internal loop sequence, CAAC, is appropriate. Likewise, the CCAC sequence in the same loop is favorable for Cover ligation.

Figure 4.6 shows a schematic for the Cover and Probe duplex with the Toggle donor strand designed. The 8mer Toggle donor in this Core is too short to compete effectively with Probe for Cover. The splint binds the Toggle donor and a few nucleotides across the brancher and into the Cover; it also holds the CY3-labeled Toggle acceptor in an orientation suitable for ligation. After designing the Core, candidate splints were analyzed using V-OMP software. The length of the splint plays an important role in annealing to Toggle and Cover and avoid hairpin formation. The 23mer GTTTGTA ACTAAGCAGTTGAGCT sequence, (Splint-1) was selected (purchased from IDT). A terminal 5' phosphate group is required on the last nucleotide of the Core so that ligation with T4 DNA ligase can be carried out. In the next section, phosphorylation was evaluated using a test sequence, T₁₅.

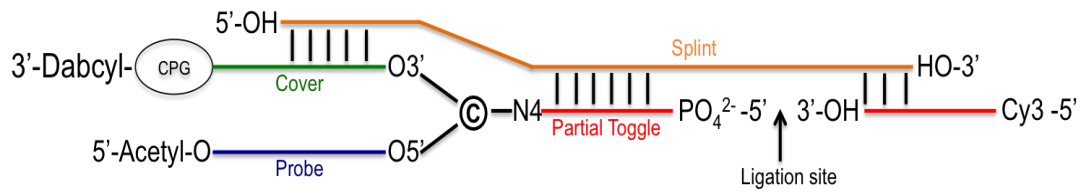


Figure 4.6: Schematic for one-step ligation.

The 8mer **Toggle** donor in this Core is exposed and its 5'-end is phosphorylated. For efficient ligation, the complementary splint binds the 3'-side of **Cover**, A-©-A, the **Toggle** donor, and the 15mer **Toggle** acceptor.

4.3. Synthesis and Analysis of Phosphorylated Test Sequence

Adenosine 5'-triphosphate (ATP) is commonly used with a T4 polynucleotide kinase to catalyze the 5'-phosphorylation of oligonucleotides. However, this would phosphorylate all sequences in the Core having a free 5'-OH, which could lead to ligation side products. Chemical phosphorylating reagent II (CPR II, Glen Research) can be used for 5'-phosphorylation on the synthesizer. The 5'-DMT of CPR II can be retained for reverse phase (Glen Pak™) purification. As shown below, purification on polyacrylamide gels is sufficient to obtain a high yield of the switch, so we eventually discarded the use of GlenPak™ purification.

The mechanism for the production of a terminal 5'-phosphate, and the elimination of the protecting group on CPR II (Figure 4.7) is as follows:

- I. A proton from the hydroxyl group is removed upon the addition of base and forms formaldehyde as a leaving group.
- II. The β -elimination reaction is carried out to release 5'-phosphate from the CPR II.

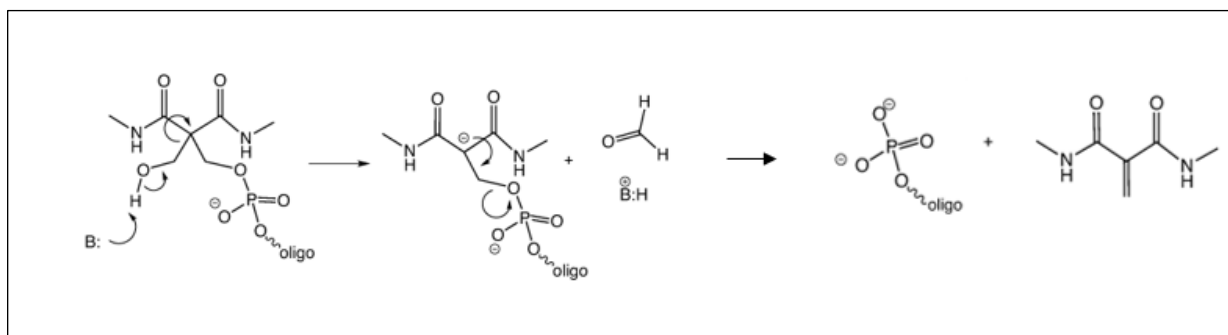


Figure 4.7: The mechanism of removal of CPR II.⁷⁶

A proton from the hydroxyl group is removed upon the addition of base and forms formaldehyde as a leaving group. The β -elimination reaction is carried out to release 5'-phosphate from the CPR II.

A test sequence, T₁₅ (non-phosphorylated), and DMT-pT₁₅ were synthesized using the modified procedure on the synthesizer. For comparison, the DMT group was deprotected from half the amount of DMT-pT₁₅ using 20% acetic acid in water. These samples were ethanol precipitated and analyzed by HPLC.

For the synthesis of the 3'-DABCYL-labeled Core (55mer), the primary sequence was synthesized using 3'-DABCYL CPG; the synthesis of the 8 nt sequence from N4 of the brancher ended at the 5'-terminal nucleotide to which the phosphate group will be added. The original CPR II offered by Glen, which we used for early trials, is a viscous oil that requires vigorous swirling in ACN for 15-30 min before placing it on the synthesizer. Even so, we got poor results from its use. All syntheses of the molecules for one-step ligation used Glen's "solid CPR II," which dissolves readily in dry ACN, and gave us much better yields of phosphorylated Core.

Results and Discussion

In the RP and IE-HPLC chromatograms of T₁₅, the retention time (RT) is reported in Table 4.1. In the RP-HPLC of DMT-pT₁₅ (figure 4.8), a trace of T₁₅ was observed at the retention time of 5.771 which corresponds to the T₁₅ control molecule (chromatogram not shown). Since DMT is strongly hydrophobic, it should elute later in RP-HPLC, thus the peak at RT = 10.619 was DMT-pT₁₅. The highest intensity peak at RT 6.57 is pT₁₅ and appears doubled, as sometimes happens in HPLC of oligonucleotides.

In figure 4.9, the putative DMT-pT₁₅ has nearly disappeared from the RP-chromatogram, showing that DMT was successfully removed to form pT₁₅ with a trace of DMT-pT₁₅ at RT= 10.627. Similar results were obtained by comparing the retention time of IE-HPLC analysis (chromatograms not shown).

Although the phosphorylated T₁₅ was synthesized with DMT ON, only one-eighth of the intensity of phosphorylated species were detected. Thus, deprotection of DMT either occurred in the synthesizer or upon drying in the speed-vac. Despite the advantage of DMT ON oligonucleotide for purification using a GP cartridge, unwanted loss of DMT would result in low recovery after GP purification. Also, yields from ligation will decrease if the phosphorylated oligonucleotide is still protected with DMT as shown by the detection of DMT-pT₁₅ in pT₁₅ RP-HPLC.

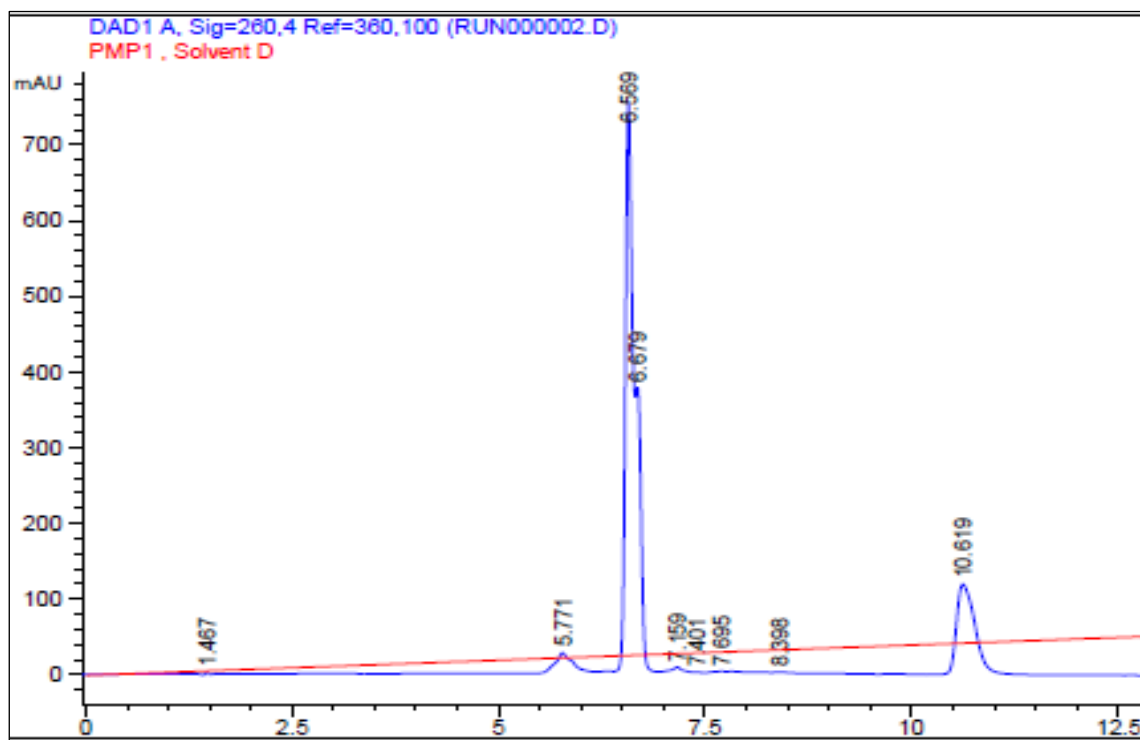


Figure 4.8: Analytical RP-HPLC chromatogram of a DMT-pT15 sample.

The analysis was carried out using a gradient of 0-60% acetonitrile at 55°C in 15 min. UV absorption at 260nm is shown in blue. Since DMT is strongly hydrophobic, it should elute later in RP-HPLC, thus the peak at RT = 10.619 was DMT-pT15.

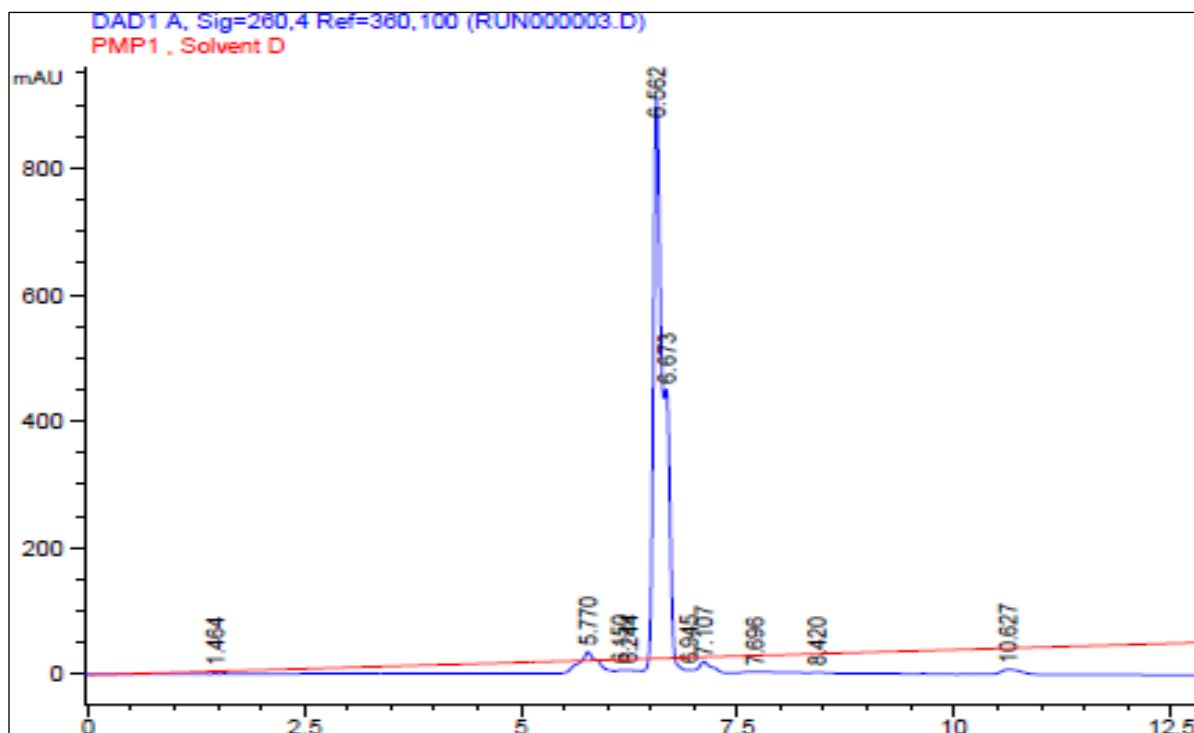


Figure 4.9: Analytical RP-HPLC chromatogram of pT15.

The analysis was carried out using a gradient of 0-60% acetonitrile at 55°C in 15 min. UV absorption at 260nm is shown in blue. Removal of DMT from DMT-pT15 was successful as the peak at 10.627 is minimal.

	T₁₅ (control)	DMT-pT15	pT15
RP-HPLC (retention time)	5.76	6.57, 10.619	6.56, 10.627
IE-HPLC (retention time)	11.14	11.626, 13.64	11.60

Table 4.1: Analysis of Phosphorylated Test Sequences.

Retention times (min) for T₁₅, DMT-pT15, and DMT OFF-pT15 for RP and IE HPLC analysis are shown. Retention times vary slightly from day to day. DMT-pT15 has nearly disappeared from the RP-chromatogram, showing that DMT was successfully removed to form pT15 with a trace of DMT-pT15 at RT= 10.627.

4.4. Synthesis and Purification of Switch NM-1 using ligation

4.4.1. Preparation of molecules for ligation

For one-step ligation, a Core molecule (55mer) was designed to be synthesized using a 3'-DABCYL column, with full-length Cover and Probe segments, and the 5' end of the Probe segment was capped (acetylated) on the system (this “primary” sequence has only standard 3'-5'-phosphodiester linkages, including the sequence through the 5Me-dC brancher). The cartridge was then removed from the synthesizer and the levulinyl group deprotected using hydrazine hydrate (in a fume hood) to free the brancher's N4-C₆OH for the synthesis of the “secondary sequence.”, the column was returned to the synthesizer, and water removed with a great excess of dry ACN, and the 8mer Toggle fragment was synthesized – see Figure 4.5, panel (a).

The Toggle fragment originates from N4 of the 5Me-dC brancher, and the 5'-end of this fragment was detritylated, and then 5'-phosphorylated with solid CPR II from Glen research, and DMT was removed. Off the synthesizer, ligation using T4-DNA ligase and complementary DNA splint-1 was used to join the 55mer Core to a 15mer “Toggle acceptor” strand to make the active switch (70mer). We synthesized the Toggle acceptor separately with a 5'-CY3 label and a free 3'-OH, using the modified synthesis described in Chapter 3 resulting in >98% average stepwise yield according to the synthesizer readout. After the synthesis of the phosphorylated Core and the partial CY3-labelled and the partial DABCYL-labelled strands, the products were removed from the synthesis cartridge, using 10% diethylamine (DEA) in acetonitrile (ACN) followed by 30% ammonium hydroxide incubated at room temperature for 17 h to remove the protecting groups on the bases.

4.4.2 Purification of switch NM-1 using ligation method

The advantage of denaturing PAGE gel is that the yield of individual bands can have very high purity in preparative gels. To recover the product from the gel, the first step is to excise the band of interest from the gel with a clean razor blade. Sybr gold stain is not needed for visualizing the band since CY3 fluoresces under UV which facilitates product recovery.

The traditional, crush and soak, a method to recover DNA from the gel can be used to slowly diffuse DNA from the polyacrylamide gel in the elution buffer. Although this method does not require any special equipment, it usually takes an 18-36 h incubation period and results in a low recovered yield (30%). Alternatively, electroelution can be used to force the DNA out of the gel into a buffer. DNA recovery is usually higher.

The D-tube dialyzer (Millipore) is a small electroelution device for proteins, RNA, DNA, and oligonucleotides in polyacrylamide or agarose gel. Using this device is a simple one-step procedure and only requires a support tray to use in standard agarose gel apparatus. Efficient extraction (>60%) for denaturing PAGE gel with a highly purified product can be achieved using this method as claimed by Millipore. The efficiency of the D-tube dialyzer was tested using a previously synthesized switch molecule (HPLC purified). The conditions provided in the protocol were modified to maximize the recovered yield from highly crosslinked PAGE gels used for preparative-scale separations of switch molecules.

As the percentage of crosslinking in a gel increases, it becomes more difficult to electroelute the product out of the gel. After some preliminary experiments, we settled on 12% denaturing gels, which gave the highest recovered yields using a D-tube dialyzer (3500 Da MWCO). Since CY3 can be visualized under white light, the sample elution was monitored very easily. After the

electroelution, samples were ethanol precipitated, and the percent recovery was calculated using the absorbance at 260nm.

Results and Discussion

4.4.1 Preliminary trials using two-step ligation and gel purification

Two-step ligation trials were conducted prior to the successful approach that used one-step ligation to produce a high-purity switch for fluorescence studies. A trial for two-step ligation is described next as it illustrates some important features of ligation and gel purification. Briefly, a 40mer Core (shown in Figure 4.5(b)) was prepared, except that the CPR II oil was used rather than the now-preferred “solid” version. This resulted in only a small fraction of the 45mer Core being phosphorylated.

An analytical 20% denaturing PAGE gel is shown in Figure 4.10 to evaluate the ligation of this 40mer Core to 15mer partial CY3. The positive ligation control in lane 3 is a phosphorylated 15mer purchased from IDT and had the same concentration as the Core and can be used to compare intensities to the other lanes. Lanes (4-9) describe the ligation results and shows bands that are described from the bottom of the gel (shortest fragments) to the top (longest).

The negative controls, minus-enzyme (lane 4), and minus splint-1 (lane 9) show the distribution of bands of samples prior to ligation. Lane (4) lacks T4 DNA ligase and shows the 15mer CY3-Toggle ligation acceptor (bottom heavy band, confirmed by CY3-fluorescence). Next up is the 23mer splint-1, and the unphosphorylated 40mer (HO-Core, (heaviest band) is near the top; only the faint band above this Core (called pCore) is the ligation donor bearing a 5'-phosphorylated 8mer Toggle fragment. Below HO-Core are 8 failure bands with substantial intensities, and then the prominent band for the 32mer “primary” sequence.” The primary sequence

was prepared on the synthesizer, and the levulinyl group was removed off the synthesizer to release ©-N4-C6-OH to synthesize the 8mer Toggle fragment. Stepwise synthesis off ©-N4-C6-OH was far less efficient than in the synthesis of the primary strand. Also, phosphorylation using the CPR oil reagent was very inefficient. This low coupling efficiency may be due to insufficient drying of the synthesis cartridge after levulinyl deprotection, steric hindrance of the growing chains on the CPG-50 solid support, or some other reason. Lane (9) lacks the splint, demonstrating that ligation is negligible in the absence of a splint, but is otherwise similar to lane (4).

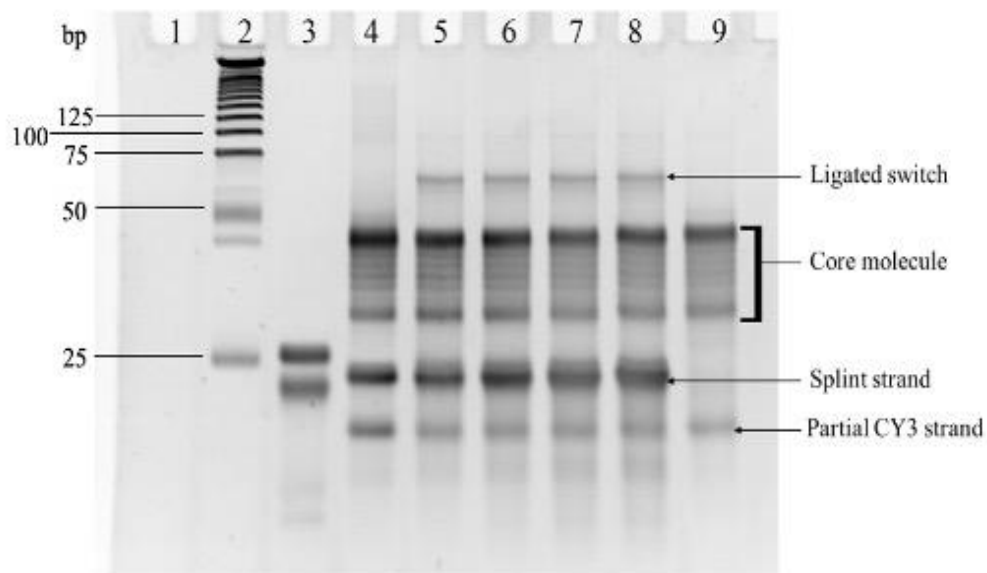


Figure 4.10: Ligation trials for the first step in two-step ligation (20% denaturing PAGE gel.)

Lane (2) shows 25bp DNA Step Ladder (Promega) of differing chain lengths. Lane (3) shows a positive control ligation of DNA fragments purchased from IDT (see text). Lane (4) is a negative control lacking DNA ligase. Lane (9) is a negative control lacking the 23mer splint strand. Lanes (5-8) are active ligations of 20 pmol of 40mer Core and 22 pmol of 15mer partial CY3 under different concentrations of splint-1. Lane (5): 40 pmol of 23mer splint-1, Lane (6): 60 pmol of 23mer splint-1, Lane (7): 80 pmol of 23mer splint-1, Lane (8) 100 pmol of 23mer splint-1. Bands were also examined for CY3 fluorescence (not shown). See text for analysis of the gel.

Lanes (5-8) are for active ligations. All show the desired 60mer product as the fluorescent, top band, ready for the 2nd ligation step to add a phosphorylated 15mer pCover-DABCYL fragment as the ligation donor to complete the 70mer switch. These bands all show the disappearance of the active ligation donor, pCore, and the faint band above the heavy HO Core band. Ligation works well under all conditions in the trial. The Core off the synthesizer was not purified prior to ligation, but standard 5'-acetylation of failure chains prevents phosphorylation and subsequent ligation. The product band (topmost) is well isolated from the other highly impure components, which bodes well for the ligation approach leading to the production of highly pure switch constructs using denaturing PAGE with single-nt resolution. It is also possible that the final products could be purified effectively by IE or RP-HPLC, although we did not attempt this.

The main advantages of the two-step scheme are that (i) it exposes the DABCYL tag is exposed to fewer rounds of harsh chemical treatment on the synthesizer that might reduce or destroy its ability to quench fluorescence, and (ii) short dye-labeled fragments can be prepared in large quantities and at very high purity using HPLC. However, if purification is required after ligation with splint-1 and again after ligation with splint-2 then workup losses could make this scheme unrealistic.

It is possible that both ligations could proceed simultaneously as splint-2 (GCTGTGGTGCTTAGTTTGT-AAC), used to guide ligation of the Cover acceptor, has a substantially different sequence from splint-1 (GTTTGTAAGTAAGCAGTTGAGCT). The desired 70mer product would run far above singly ligated by-products on denaturing PAGE (by-products would include the DABCYL fragment ligated to acetylated Core failures). For the proof-of-concept studies, here, we decided that one-step ligation should provide enough material for 3-segment switches. Later enhancements could use single-pot, simultaneous ligations to improve the

purity and recovered yield of 3-segment switches, and reduce costs for production at commercial scales.

A 20% analytical denaturing gel for one-step ligation is presented in Figure 4.11. In the minus-Ligase control (Lane 1) the bands are read from bottom to top as the CY3-Toggle acceptor, followed by the splint and the 55mer Core with failure sequences. We were able to reduce the amounts of failure bands off ©-C₆-O- by adding an extra 6 min pause time for every incoming base after levulinyl group deprotection (section 3.2.3). The failures are still more abundant than during synthesis of the 47mer primary sequence that begins with DABCYL on the CPG support, so commercial efforts should explore ways to reduce losses due to failures off ©-C₆-O-. The intensity of the 55mer pCore band was much higher than the pCore for two-step ligation which is likely due to using the solid CPR II reagent. Ligation was performed at 4°C for 48 h.

A possible explanation for visualizing two ligated bands at the top of Lane (2) is that this 3-segment branched molecule may not be completely denatured by formamide prior to loading in the well, and by 7M urea in the gel, run at ambient temperature. Of the two ligated bands, the upper one was more highly fluorescent, suggesting that this was the ON form and perhaps the bottom band was the OFF form. Further work could test this hypothesis.

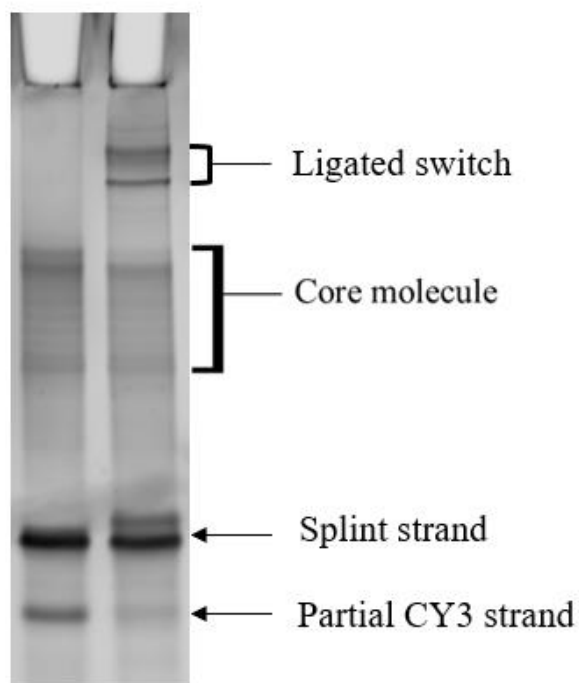


Figure 4.11: Analytical 20% denaturing PAGE gel following one-step ligation.

50 pmol of a 55mer pCore bearing DABCYL at the 3'-end of its Cover segment was ligated with 100 pmol of a 15mer partial CY3 and 100pmol of 23mer splint-1. A control without T4 DNA Ligase is shown (left) after ligation. The top band is pink under the UV transilluminator.

4.4.2 Purification of 70mer switch NM-1 after one-step ligation

To extract the ligated molecule for mass-spec analysis and fluorescence studies, a single-well gel was used to maximize the amount of sample that can be loaded. 100 μ L of the sample was prepared and the ligation mixture was incubated for 48 h to ensure complete ligation. Using the optimized conditions for running denaturing PAGE gel, the high intensity of the ligated band was observed (Figure 4.12). This band was then excised from the gel and cut into very small pieces (~2 mm). The sample was electroeluted and ethanol precipitated to recover the ligated product and remove excess salt. The recovery of this ligated molecule was about 60% which was lower than the reported extraction efficiency (90%) from Millipore. The ligated sample was then sent for ESI-mass spec analysis to verify its purity.

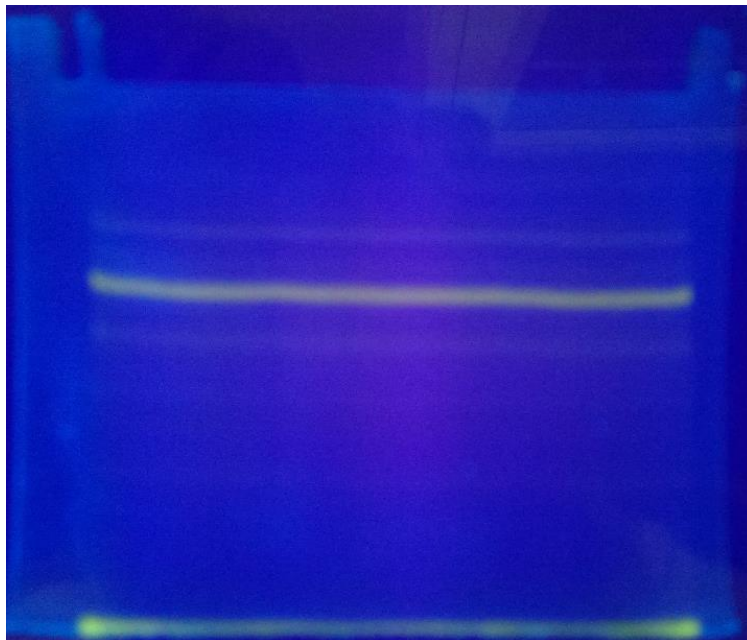


Figure 4.12: Purification of 70mer switch NM-1 after one-step ligation.

12% denaturing PAGE gel for extraction of 70mer switch NM-1 prepared by one-step ligation of a 15mer CY3-Toggle fragment to a 55mer pCore molecule. The 15mer fragment (bottom of the gel) and active switch (intense middle band of the set of bands near the top) are brightly fluorescent.

The ESI-MS spectrum shown in (Figure 4.13) is for the switch excised from the gel (sequence in Figure 4.1). The most prominent peak, MW= 22597.4Da corresponds to the ligated switch molecule; its molecular weight was in the acceptable error range of $\pm 0.03\%$. This peak has 68% abundance in the spectrum and 90% estimated purity, which is the highest purity compared to different purification methods that we had used previously. The sample purity is higher, as the peak at 22647.5 has the mass of the 2•Na adduct from ethanol precipitation, and most of the others to the right of the main peak have the mass of known adducts due to the matrix used for ionization. It is also possible that very small amounts of impurities arise from cyanoethyl groups that were not fully deprotected, modified bases, etc.³⁴ There are also very small impurities at the left of the spectrum. These could arise from “skipped” bases during the synthesis of Core, but they aren’t a major concern for this proof-of-principle study.

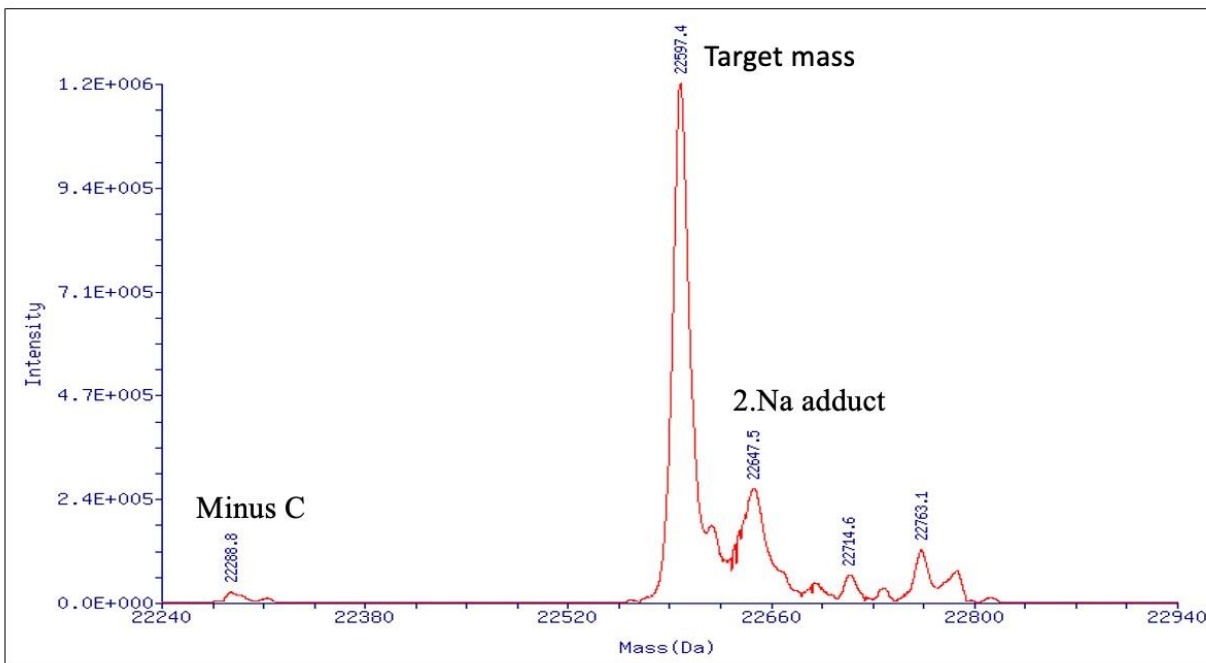


Figure 4.13: ESI-MS analysis of ligated switch NM-1.

Full-length NM-1 was isolated from the gel in figure 4.12 where minor impurities occur on the right side of the spectrum (see text).

4.5. Design and synthesis of Switch NM-2

The 71mer switch, NM-2, should have $K1 \sim 0.1$ (figure 4.14). It has a T hairpin and P:C duplex (ON form) in the population of $\sim 90\%$ and a P-hairpin and T: C duplex (OFF form) in around 10% of the total population, according to the ΔG° values estimated for the two states using V-OMP. The synthesis of NM-2 was carried out the same way as NM-1. The ligation site was determined such that the CY3-AACTATGCTACAAATG-OH ligation acceptor and the pCTAGTTAC-Core donor will have perfect base pairing to the splint to make the 23mer Toggle segment. The perfect alignment on the splint will prevent any pairing between bases from the CY3-acceptor fragment and other parts of the pCore donor fragment. The pCore and splint were designed using V-OMP. The ligation scheme of this switch molecule is very similar to that described in figure 4.11.

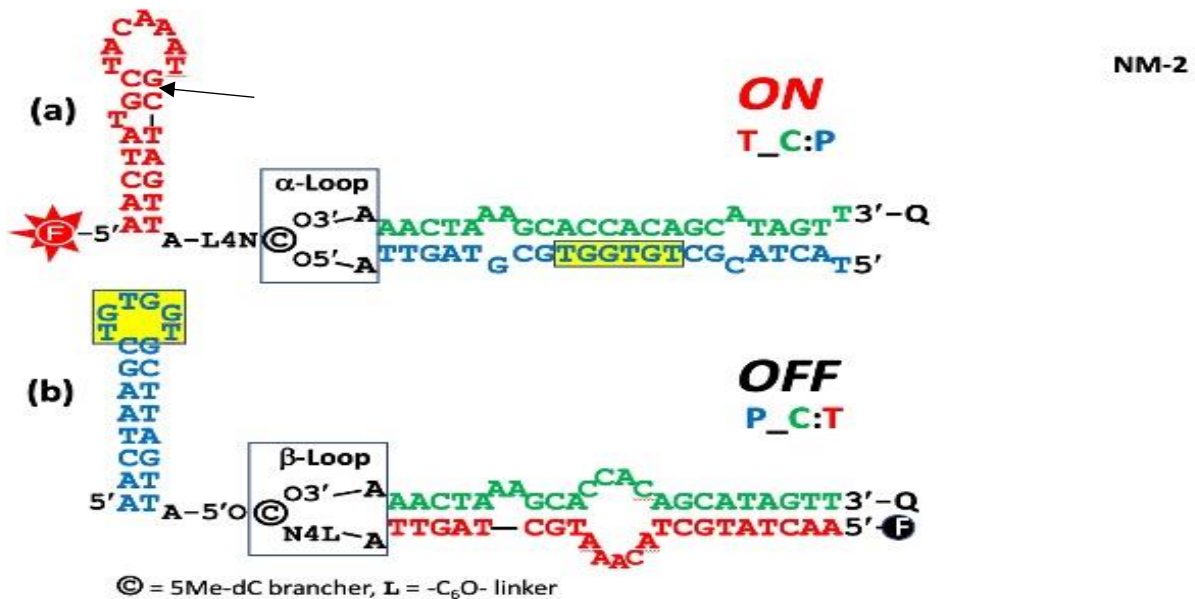


Figure 4.14: 71mer switch NM-2.

The segments are **Toggle**, **Cover**, and **Probe**, joined at a 5Me-dC brancher (denoted by ©); important residues for binding NC are outlined in yellow. 3-nt hairpin loops surround the brancher, an α -loop and a β -loop, where L is -C₆O-. (Figure prepared by Dr. Borer.)

The arrow shown in the ON conformation indicates the ligation site of the 16mer CY3 Toggle acceptor fragment with the pCore. The V-OMP estimated $\Delta\Delta G^\circ = -1.3$ kcal/mol, favoring the ON-state, with K1 ~0.1.

Results and Discussion

Using the optimized parameters for ligation, the 15mer CY3-AACTATGCTACAAATG-OH was joined to pCTAGTTAC -Core. The band corresponding to the ligated switch in 12% denaturing PAGE gel (similar to figure 4.12) was extracted and ethanol precipitated to recover the ligated molecules. The concentration of the switch was then calculated using the extinction coefficient as described in the supplementary material.

After the purification of ligated NM-2, hydrolysis of NM-2 using Dnase I was performed. Figure 4.15 shows the fluorescence spectrum for the 71mer switch (40 nM concentration) before and after treatment with DNase I. There is little, if any, significant difference between these spectra. There is also very little difference between this spectrum and the locked-ON version in figure 4.19, thus we might expect the K1 equilibrium in this active switch to favor the ON-form by a considerable margin over OFF.

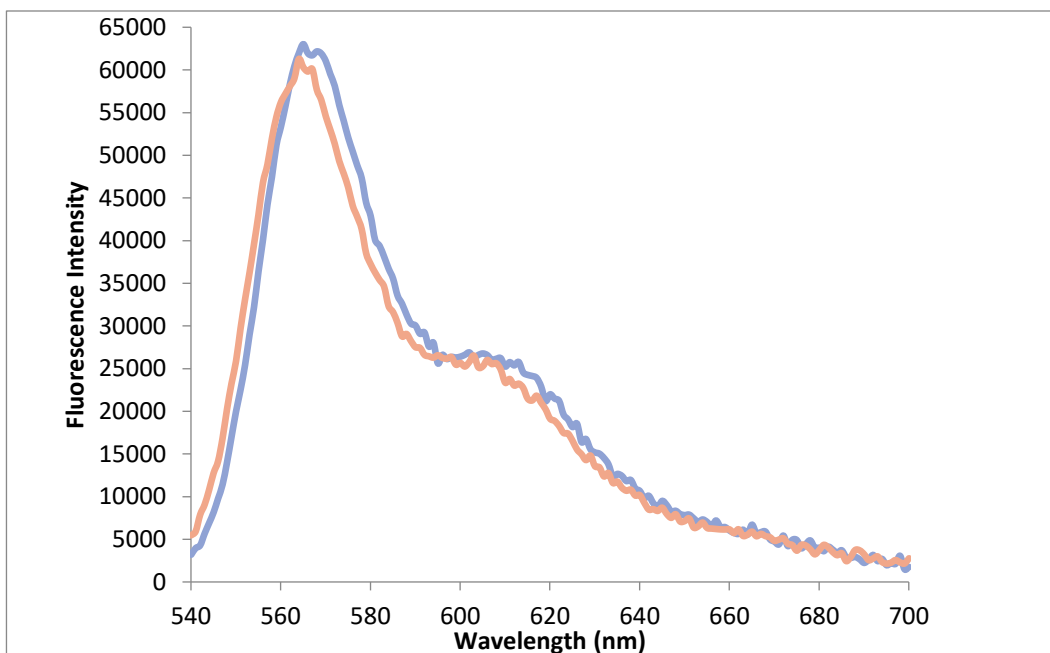


Figure 4.15: DNase I hydrolysis of switch NM-2.

The fluorescence intensity was measured before (blue) and after (red) the addition of DNase I. Fluorescence emission spectrum measured on a PTI photon counting fluorometer. No significant difference between the two spectra assuming NM-2 favors the ON form.

4.6. Locked-ON and Locked-OFF molecules

Locked ON and OFF molecules, which constrain the switch molecule in an ON or OFF conformation, are essential to verify the maximal and minimal fluorescence intensities to define the limits on contrast ratio, which is (ON fluorescence) / (OFF fluorescence). The Locked-ON molecule consists of a permanent Cover:Probe duplex and a CY3-AAACACCAAAA segment replacing Toggle shown in figure 4.16, (compare with figure 4.14). This separates the fluorophore and quencher and approximates the maximal fluorescence intensity for the switch. The Locked-OFF molecule has a permanent CY3-Toggle:Cover-DABCYL duplex where the fluorophore and quencher are in close contact thus resulting in minimal fluorescence intensity; The Probe segment is replaced by an A₁₀ segment. Locked-ON and locked-OFF molecules were synthesized and hydrolysis of these molecules with DNase I was carried out.

The sequences of Locked-ON and Locked-OFF molecules were designed using V-OMP as illustrated in Figure S.4.4. Unlike the ten adenosines to eliminate switching in the Locked-OFF form, three cytosines were placed within the CY3-AAACACCAAAA segment to provide a site for efficient ligation. Ligation sites are at the junctions of black and red fonts in figure 4.16. For the ligation reaction, the splints corresponding to the Locked-ON and Locked-OFF molecules were designed using V-OMP. The splints also prevent hairpin formation.

The ideal contrast ratio between Locked-ON and Locked-OFF molecules is important to verify the maximal and minimal fluorescence intensity of the binding of the CY3 and DABCYL hybrid. Deoxyribonuclease I, DNase I, is an endonuclease (breaks phosphodiester bonds within the oligonucleotide). Therefore, the fluorescence emission of Locked-ON and Locked-OFF molecules (figure 4.16) after purification was measured using a PTI Quantamaster fluorescence spectrophotometer.

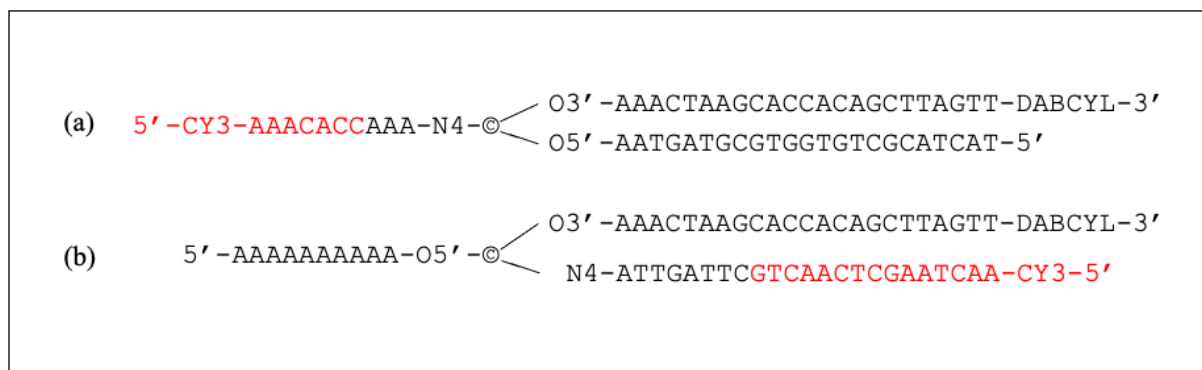


Figure 4.16: Sequences of Locked-ON and Locked-OFF states.

Locked-ON molecule (a), the Toggle sequence is replaced with seven adenosines and three cytosines covalently attached to CY3 to observe the maximal fluorescence intensity. Locked OFF molecule (b) exhibits minimal fluorescence due to the replacement of the Probe sequence with ten adenosines by locking the molecule in an OFF state. The sequences (red) in both Locked-ON and Locked-OFF structures indicate the ligation sites of the molecules.

Results and Discussion

4.6.1 Design and Synthesis of Locked-ON and Locked-OFF molecules using ligation

The Cores and partial CY3 strands of Locked-ON and Locked-OFF molecules were synthesized using the optimized parameters, which increase capping and oxidation time and double the addition of amidites after levulinyl group deprotection. After synthesis, the molecules were removed from the column by deprotecting with 10% diethyl amine (DEA) in acetonitrile (ACN) followed by 30% ammonium hydroxide to remove the protecting groups on the bases. The solutions were incubated at room temperature for 17 h and then sodium acetate/ethanol precipitation was carried out to recover the pellet. By utilizing the optimal condition for ligation reaction, the ligation mixture of the Core, partial CY3, and splint were prepared and analyzed with 20% denaturing gel.

Initially, we attempted to join the ligation donor, CY3-AAAAAAAA-OH, to the pAAACore of the Locked-ON molecule. However, no ligation was observed on a 20% denaturing gel. To increase the chance of successful ligation, we switched to a CY3-AAACACC-OH donor and adjusted the splint sequence. The ligated band was visualized on the denaturing gel. Unlike the Locked-ON molecule, preparing Locked-OFF by ligation of CY3-AACTAAGCTCAACTG-OH to pCTTAGTTA-Core succeeded on the first try due to having a mixture of various bases surrounding the ligation site.

We still did not achieve a high level of phosphorylation at the 5' terminus of Core synthesis; only ~35% of Cores were successfully phosphorylated as determined by measuring the intensity of the bands on the denaturing gel (Kodak Gel Logic 100 Digital Imaging System). Since the

5'-OH of Cores must be phosphorylated prior to ligation, we turned to T4 polynucleotide kinase (PNK), which can introduce the 5'-phosphate group post-synthetically.

T4 PNK catalyzes the transfer of the terminal phosphate from adenosine triphosphate (pppA) to the 5'-OH of polynucleotides. The conditions for phosphorylation using T4 PNK and ATP, and incubation time were established specifically for 1 μ M HO-Core using a mixture of pCore (from CPR-II on the synthesizer) and HO-Core. A sample of Locked-ON Core molecules was synthesized with partial phosphorylation by CPR II. A gel from this sample is shown in Figure 4.15 (left lane), while the same Core with additional end labeling via T4 PNK is shown in the right lane. The intensity of the pCore in band 1 increased by ~40% from the left lane as determined by the Gel Doc. The data in Figure 4.15 also shows that most, if not all, of the Core molecules in band 2 are capable of being phosphorylated. They have not capped failures or “wounded” Cores. Future work could be directed at increasing the potency of phosphorylation on the synthesizer, as for commercial production it is desirable to avoid the use of an enzyme.

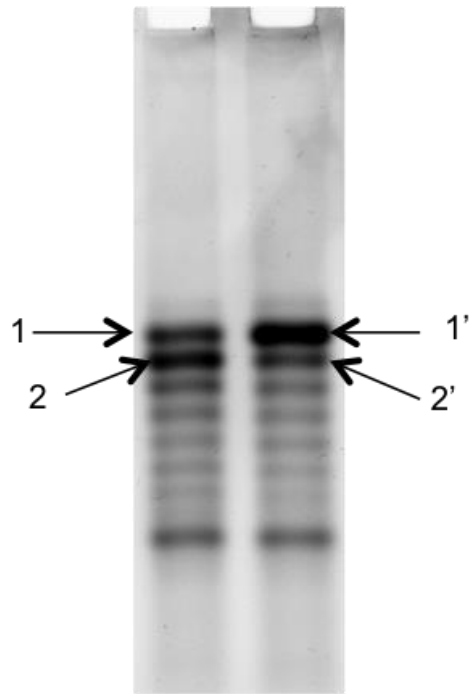


Figure 4. 17: Analytical 20% denaturing PAGE gel.

The band intensity of before (left lane) and after (right lane) phosphorylation with T4 PNK is shown. The intensity of bands 2 to 2' decreases as more phosphate groups were reacted to form band 1.

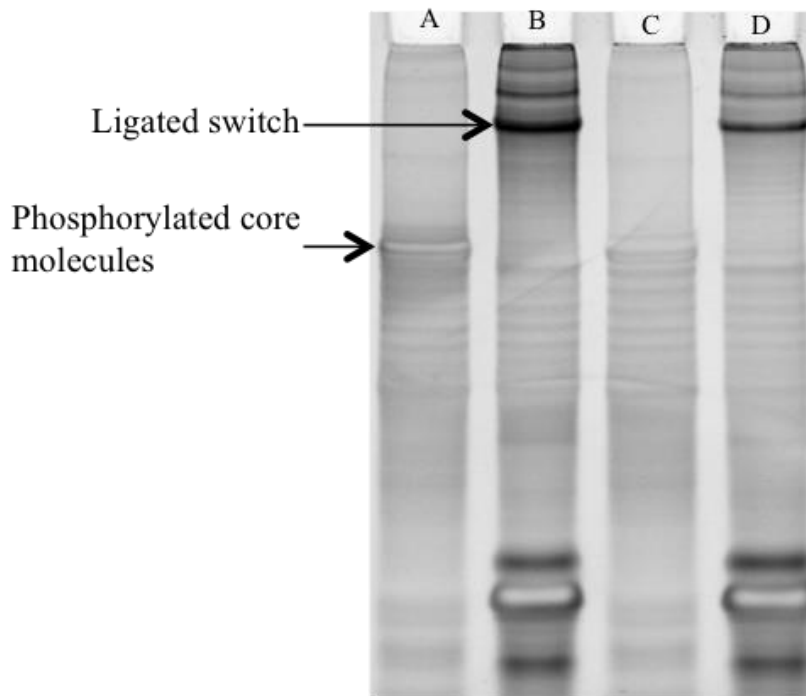


Figure 4.18: Analytical 20% denaturing PAGE gel of ligation where the Core was treated with T4 PNK.

Lane A shows is the Core treated with T4 PNK, and lane B shows the ligation of this pCore and 15mer CY3-Toggle-OH acceptor – the band for a ligated switch is quite intense. Lane C is the Core that was partially phosphorylated with CPR II, and lane D shows a weaker band for the Ligated switch that results from ligating CY3-Toggle-OH with the CPR II Core.

The analysis is carried a step further as shown in figure 4.18. We started with a Core sample that was partially phosphorylated by CPR II on the synthesizer (lane C). This Core sample was treated with T4 PNK with the result shown in lane A, where the phosphorylated band is more intense. Ligation of this sample with CY3-Toggle-OH acceptor produces an intensely fluorescent band marked near the top of lane B, considerably higher than the similar band in lane D where the partially phosphorylated Core molecules were not increased by PNK treatment. Utilizing a single-well gel, the extractions of ligated Locked-ON and Locked-OFF molecules were carried out by electro-elution with D-tube dialyzers. The samples were then recovered by ethanol precipitation.

4.6.2: Hydrolysis of Locked-ON and Locked-OFF molecules

The contrast ratio was ~11 between (Locked-ON) / Locked-OFF molecules calculated based on the fluorescence intensity at the peak maxima (near 560 nm wavelength) in figure 4.19. The fluorescence intensity of the Locked-ON molecule exhibited no change upon the addition of DNase I, suggesting that Locked-ON already had its maximal fluorescence.

An increase in fluorescence intensity of the Locked-OFF molecule was detected after DNase I treatment but not close to the intensity of Locked-ON. It is possible that the concentration of the Locked-OFF molecule was calculated incorrectly, or the purity of the ligated Locked-OFF molecule is low, but those explanations seem unlikely. DNase I is an endonuclease that nicks DNA strands, leaving oligonucleotides that may not fully dissociate from a duplex. So, the hydrolysis may not be complete. We note that the DABCYL and CY3-dyes lie at one end of a duplex and DNase I may require several nicks to release either DABCYL or CY3 from the duplex see figure 4.16. For Locked-ON, however, a nick in either strand could release one dye or the other. Also, a mixture of endo and exonucleases (DNase I and snake venom phosphodiesterase) has been used to efficiently degrade DNA to single nucleotides.⁷⁷ This should cleave all of the DABCYL and CY3 from the switch sequences.

Of more concern, however, is the relatively high fluorescence of the unhydrolyzed Locked-OFF molecule. R. Iyer in our lab built 20mer perfect duplexes with CY3 and DABCYL next to each other on the adjacent strands and found that their residual fluorescence was nearly zero.³⁴ Decreasing the fluorescence of the OFF state to lower values could produce excellent contrast ratios, preferably > 20:1.

Another possible explanation for the relatively high residual fluorescence in Locked-OFF is that some of the DABCYL dye was damaged during the many rounds of chemical synthesis and the hydrazine treatment required to produce our pCore molecules for the ligation of locked and active switches. This situation could be minimized by preparing locked and active switches using two-step ligation as described earlier. A definitive investigation into this issue would be to synthesize a single nucleotide linked to DABCYL-CPG, detritylating, capping, and subjecting it to 50 more rounds of synthesis chemistry adding coupling reagents but leaving out the amidite. Half of the sample could then be treated with hydrazine. Then the two nt-DABCYL samples could be deprotected, purified by RP-HPLC⁷⁸, and mass spectra obtained for the various peaks. If substantial modifications to DABCYL are detected, further work should adopt two-step ligation to obtain switches with a high contrast ratio. The same procedure could be applied to differing nucleotides to determine whether modifications occur on the nucleobases or sugars in substantial amounts due to synthesis conditions.

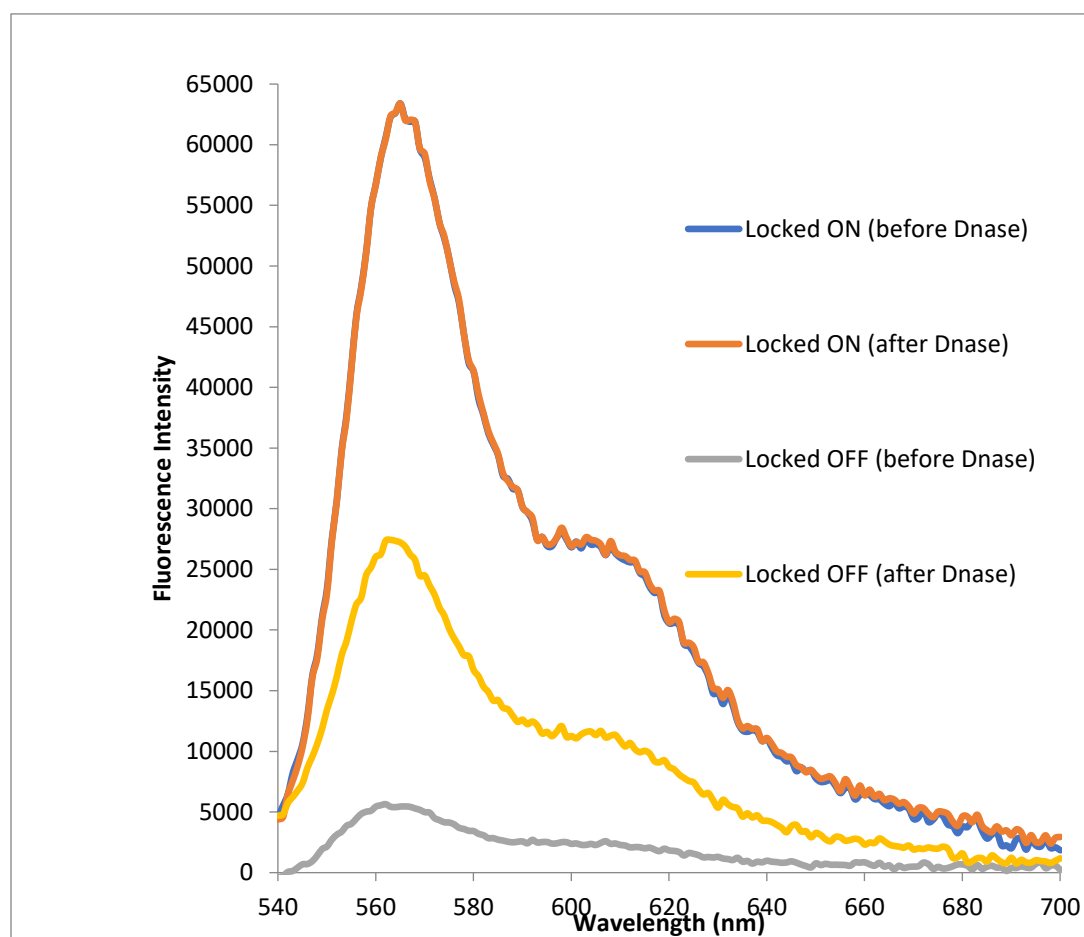
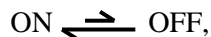


Figure 4.19 Fluorescence emission of Locked-ON and Locked-OFF molecules.

The spectrum shows before and after hydrolysis with DNase I of Locked-ON and Locked-OFF molecules. Concentrations prior to hydrolysis were the same, and dilution was negligible upon the addition of DNase I.

4.7. Analysis of NCp7 binding to Switch NM-1 and 2*

The equilibrium for switching in NM-1 and NM-2 (in the absence of NCp7) is



with an equilibrium constant, $K1 = [\text{OFF}]/[\text{ON}]$, which also defines the maximum range of fluorescent signal that can be expected from an $\text{ON} \rightarrow \text{OFF}$ switch.

We have assumed that the most useful range for $K1$ is 0.1 to 0.001 for applications to high throughput screening of drug candidates, as that will provide a large range in fluorescence. The mole fraction of bound, f , and free, $1-f$, switches can be estimated from $K1$ and the conservation equation for the total concentration of switch, $S_t = [\text{OFF}] + [\text{ON}]$. For $K1 = 0.03$, $[\text{ON}] = 0.0971 \times S_t$, so about 97% of the maximum fluorescence from Cy3 is present before any protein is added. For $K1 = 0.1$, the populations are $[\text{ON}] = 0.909 \times S_t$, so about 9% of the potential signal is lost in the OFF-state due to the $K1$ equilibrium. For $K1 = 1$, 50% of the switch population is already OFF before any protein is added. NM-1 was designed to have $K1 \cong 0.03$ and NM-2 to have $K1 \cong 0.1$.

When NCp7 binds to the DNA Probe hairpin alone, the dissociation equilibrium constant, $K_d(\text{probe}) = 16 \text{ nM}$.³⁴ However, most of the Probe in currently designed switch molecules is base-paired to Cover. Thus, in switches like NM-1, a free energy cost must be paid for NCp7 to bind a large fraction of Probe, due to switching to the higher energy conformation that has Probe available to bind protein. Another way to say this is that switching attenuates binding, and $K_d(\text{apparent}) > K_d$.

*This section and associated figures were drafted by Dr. Borer. NM designed the experiment and measured the fluorescence data in Figure 4.20.

The equation that accounts for the effect of switching on binding⁵⁷ is

$$\text{Eq. 4.1}^\dagger: \quad K_d(\text{app}) = K_d(\text{probe}) \times (1 + 1/K1),$$

where the factor $(1 + 1/K1)$ is always greater than 1. Measuring $K_d(\text{app})$ usually allows the calculation of $K1$ for the switch using the previously measured value of $K_d(\text{probe}) = 16 \text{ nM}$. An exception is when $K1 > 0.1$, and a substantial amount of an ON→OFF switch is already in the OFF form, as will be discussed below-regarding switch NM-2.

Plots of f vs. P_t/S_t were generated from eq. 4.1 at several values of $K1$ as shown in Figure 4.20 (P_t is the total concentration of added protein). The figure graphically depicts the attenuation of binding of the switch to protein at $K1 < 0.1$. $K_d(\text{app})$ approaches $K_d(\text{probe})$ in the limit as $K1 \rightarrow \infty$, as illustrated in Figure 4.20(b). The plots were generated using the solution to the quadratic equation for binding of the switch to protein,

$$f = \{1 + K_b(P_t + S_t) - \text{Sqrt}[(1 + K_b(P_t + S_t))^2 - 4 K_b P_t S_t]\} / 2S_t,$$

where $f = S_{\text{bound}}/S_t$ and $K_b = 1/K_d(\text{app})$. This is eq. 4 in section 6.3.5 of Iyer's dissertation.³⁴

In the case of the ON → OFF switches described here, the fraction of free switch, $1-f$ correlates with fluorescence, and the dashed line plots in Figure 4.20 are most useful for the experimental measurements reported in Figure 4.21 for NM-1 and NM-2. The measured values of fluorescence, I , as a function of P_t/S_t can be converted to $1-f$ by

$$1-f = 1 - (I_0 - I)/(I_0 - I_\infty)$$

where I is a particular fluorescence measurement, $I_0 = I$ at $P_t/S_t = 0$, and $I_\infty = I$ at a very large value of P_t . $K_d(\text{app})$ and I_∞ can be estimated by curve fitting, using the quadratic function and procedure described in section 6.3.5 of Iyer's dissertation.³⁴ Fluorescence measurements at

[†]Note, eq. 5 in Dr. Iyer's thesis (p. 222) was written incorrectly, where the \times sign (eq. 4.1, above) was replaced by $/$; however, his calculations were done with the correct equation.

constant S_t , can be used to calculate the relative fluorescence, $I_{rel} = I/I_0$, as plotted in Figure 4.21.

$K_d(app)$ can be used to determine K_1 from $K_d(app)$ according to eq. 4.1.

Translating the fluorescence data to a curve of $1-f$ vs. P_t/S_t affords a consistency check on the curve fits by comparing the fitted theoretical curve and experimental data at $f = 0.5$, as described next.

$K_d(app)$ can be expressed in terms of the concentrations of free switch, $[S]$, free protein, $[P]$, and bound switch, $[S:P]$, according to

$$K_d(app) = [S][P]/[S:P]$$

At $f = 0.5$ the following statements apply:

$$[S] = [S:P], \text{ so } K_d(app) = [P]$$

$$[P] = P_t - [S:P] = P_t - S_t/2 \text{ because the switch half-saturated at } f = 0.5, \text{ so}$$

$$K_d(app) = P_t - S_t/2, \text{ and}$$

Eq. 4.2:
$$P_t/S_t = 0.5 + K_d(app)/S_t$$

Thus, at $f = 0.5 = 1-f$, the value of P_t/S_t should match that calculated from Eq. 4.2. A sometimes-difficult issue in fitting experimental data to theoretical curves, as in Figures 4.22 and 4.23, is a realistic estimate of I_∞ . Eq. 4.2 provides an easy visual check on the consistency of the fit, as will be illustrated below.

The two experiments summarized in Figure 4.21 show the decrease in relative fluorescence, I_{rel} vs. P_t/S_t ; the total NCp7 concentration, P_t , increases along the x-axis and the total switch concentration, $S_t = 4.0 \times 10^{-8}$, is held constant for each data point ($I_{rel} = I/I_0$).

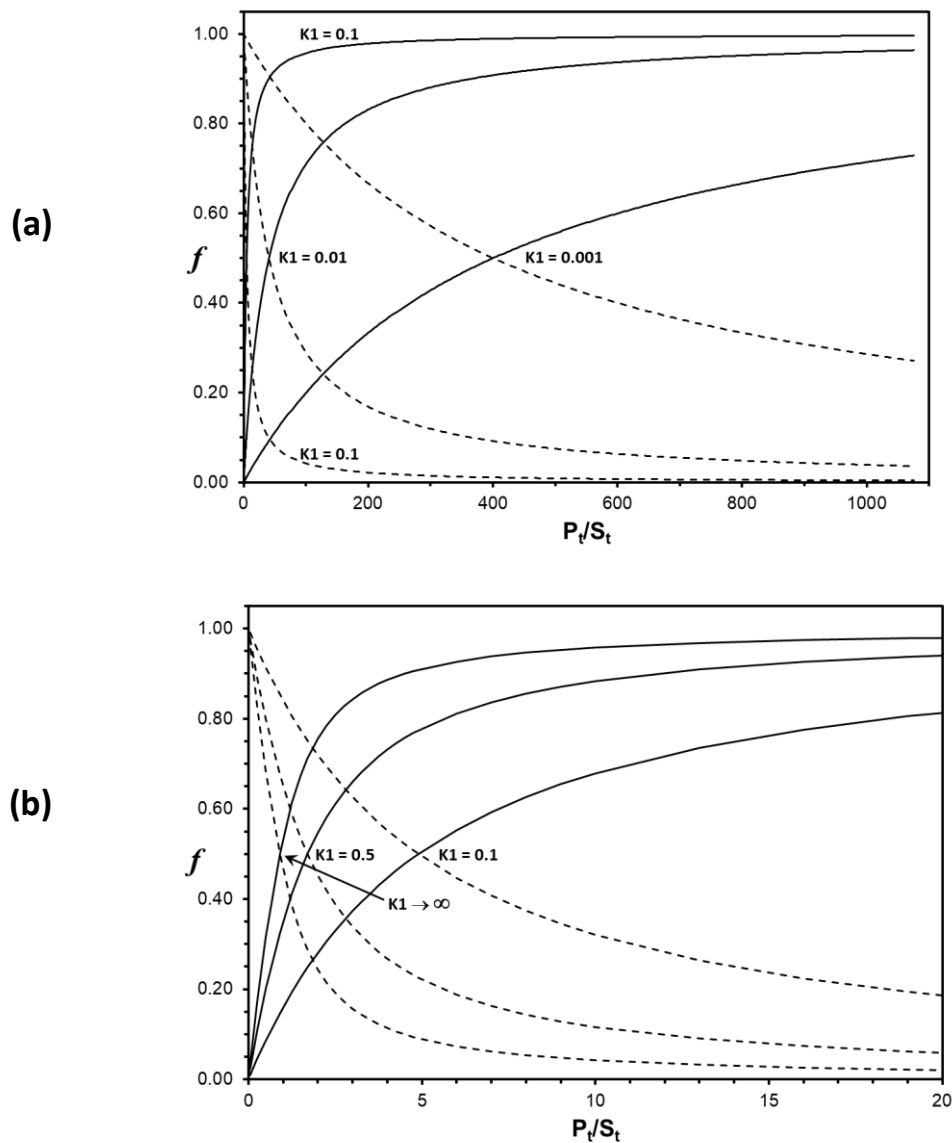


Figure 4.20: Theoretical curves for switches show the effect of $K1$ on $K_d(\text{app})$.

The fraction of the bound switch, f , (solid lines), and free switch ($1-f$, dashed lines) are plotted against P_t/S_t for switches at several values of the $K1$ equilibrium constant. The curves describe binding according to Eq. 4.1, with the dissociation constant of the Probe segment alone taken as $K_d(\text{probe}) = 16 \text{ nM}$. Panel (a) describes binding for values of $K1$ (0.001, 0.01, and 0.1). Panel (b) describes binding on an expanded x-axis for values of $K1$ (0.1, 0.5, and the limit when $K1 \gg 1$, and $K_d(\text{app}) \rightarrow K_d(\text{probe})$).

Fluorescence measurements were made on a plate reader; experimental details are given in section 4.2. $1-f$ can be calculated for each point if the limit fluorescence, I_∞ , is known at a very large concentration of protein. NM-1 was designed to have $K1 \sim 0.03$. However, visual inspection of the measured data in Figure 4.21 and Figure 4.20(a) suggests that a $K1$ value close to 0.001 will fit the data better than using a $K1$ close to 0.01.

Figure 4.22 and Table 4.2 show the results from curve fitting, with $K_d(\text{app}) = 5.3 \times 10^{-6}$ M. Using eq. 4.1, $K1 = 0.003$, less than the designed $K1$ by a factor of 10. This is not unexpected as predictions of ΔG° from the thermodynamic database for DNA folding typically have error bounds of ± 1 kcal/mol. The error bounds increase with number of DNA residues and there are additional uncertainties associated with the unusual nature of 3-segment structures. A 1.4 kcal/mol error in the predicted value for the difference in free energy between the ON and OFF forms of a switch, $\Delta\Delta G^\circ$, would result in a factor of 10 error in $K1$. Iyer's experience with designing OFF \rightarrow ON switches is similar.³⁴ At this point we can say that it is possible to use the thermodynamic databases to design switches that function, but tests require their synthesis, purification, and real-world experimental measurements. Combinatorial methods to evaluate switch constructs would allow the synthesis and testing of multiple switches to produce a switch in the desirable range of $0.05 > K1 > 0.01$. Such switches are relatively easy to trigger by small changes in P_t and can give a change in signal that is 95-99% of the maximum upon saturation with P.

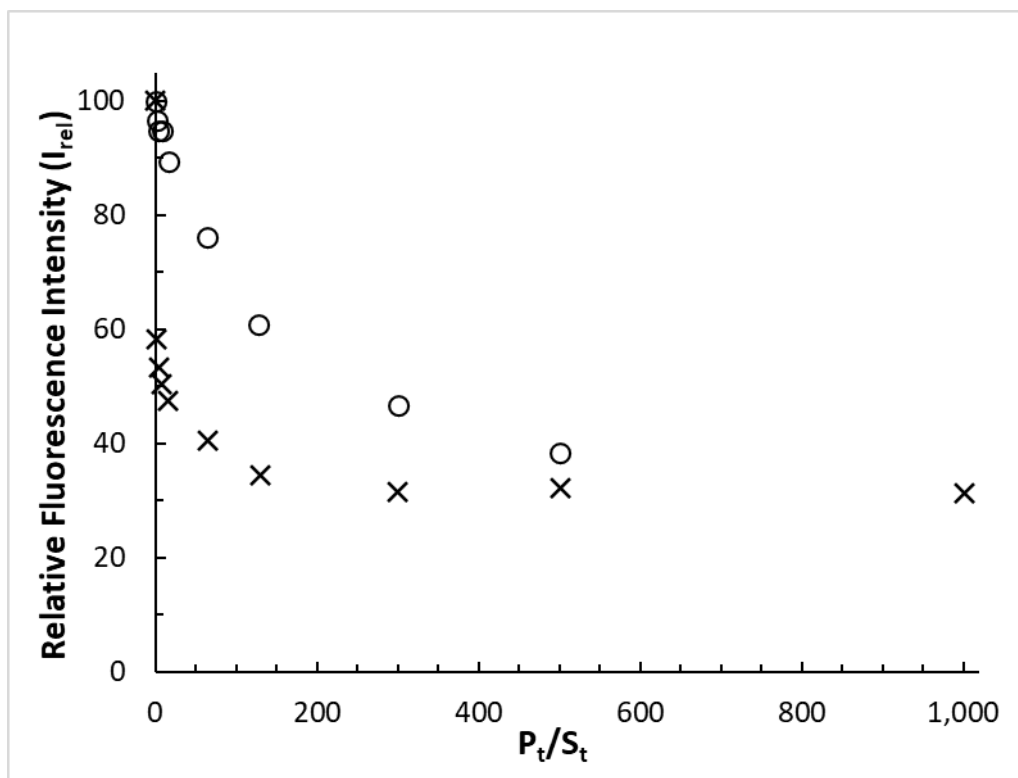


Figure 4.21: Fluorescence measurements for switches NM-1 and NM-2.

The relative fluorescence intensities for NM-1 (O) and NM-2 (x) are plotted against the ratio of total moles of NCp7 protein (P_t) added to the solution divided by the total moles of switch (S_t); $S_t = 40$ nM is held constant. Data were measured on a fluorescence plate reader with 100 μ L of solution in each well.

Figure 4.22 shows an excellent fit to the experimental data for NM-1, and the bottom of Table 4.2 shows the derived parameters for fitting with a single exponential. The best fit was obtained with the limiting fluorescence, $I_{\infty} = I_{rel} = 22$, which is 22% of I_0 . Trials at 20% and 24% required slight adjustments to $K_d(app)$ that suggest reasonable error bounds are $K_d(app) = 5.3 \pm 0.3 \times 10^{-6}$ M, as the trials at $I_{\infty} = 20$ and 24 resulted in slightly worse fits than at $I_{\infty} = 22$. The two items highlighted in yellow at the bottom of Table 4.2 are adjustable parameters that update the Excel spreadsheet and graphs used for visual evaluation of the quality of fit in Figures 4.22 and 4.23.

The consistency check using eq. 4.2 showed the expected result of $P_t/S_t = 133$ at $1-f = 0.5$ for the fit. Iyer used Kaleidograph software to fit his data on OFF \rightarrow ON switches, with results described in section 6.3.5 of his dissertation.³⁴ The two parameters used by the software for fitting are $I_0 - I_{\infty}$ and $K_d(app)/S_t$. The latter usually dominates the term at the right of eq. 4.2, typically in the range of P_t/S_t from 140 to 170 for his switch experiments; in such cases, the factor of 0.5 can be neglected from eq. 4.2. Similar curve fitting software will be used to obtain best fits to $K_d(app)$ and I_{∞} for NM-1 and 2 when the current measurements are submitted for publication, although it is unlikely they will change significantly from the visualization method used here.

There is a more complex situation for NM-2, which was designed with $K1 = 0.1$. The data in Figure 4.21 (\times) show an extremely abrupt decrease in fluorescence for the first five points, which then quickly levels off. It was impossible to fit the whole collection of data with single values of $K_d(app)$ and I_{∞} . Table 4.3 (top) and Figure 4.23(a) show a fit to the first five points on a greatly expanded scale for the x-axis. The derived value for $K_d(app)$ is 16 nM, indistinguishable from $K_d(probe)$; this might imply that $K1 > 100$. The fit to the last five points

(squares) is shown in panel (b) with data and parameters in Table 4.3 (bottom). The first point (at $P_t = 0$) is used to fit both curves. The four other early points (circles) are also plotted, but were not used in the second fit; Figure 4.23(b) illustrates that these early points do not fit well to the plotted quadratic (solid line) with $K_d(\text{app}) = 450$ nM. As lettered in the graphs of Figure 4.23 the consistency checks give reasonable results, with the factor of 0.5 being significant for a switch with $K1 \geq 0.1$.

Implicit in the previous analyses is the assumption that I_{rel} is directly proportional to $1-f$, or nearly so. However, NM-2 is probably a switch with a $K1 > 0.1$. This has interesting consequences. For example, if $K1 = 1$, then half of the switch is already in the OFF state at $P_t = 0$. Further, both ON- and OFF-states have the same ΔG° , so there is no free energy cost for switching. The Probe hairpin is exposed to molecules in the OFF state and ready to bind NCp7 protein, so $K_d(\text{app})$ should equal $K_d(\text{probe})$ for OFF molecules. As P_t increases, bound members of the original OFF-pool will then be replaced from the ON-pool by the rapid $K1$ equilibrium, causing fluorescence to decrease. I_{rel} for points at large values of P_t should level off rapidly. Rather than $K1 = 1$, it is more likely that $K1$ is in the range from 0.5 to 0.1 with a small switching free energy, $\Delta\Delta G^\circ$, from -0.4 to -1.4. Beyond that, we are reluctant to speculate.

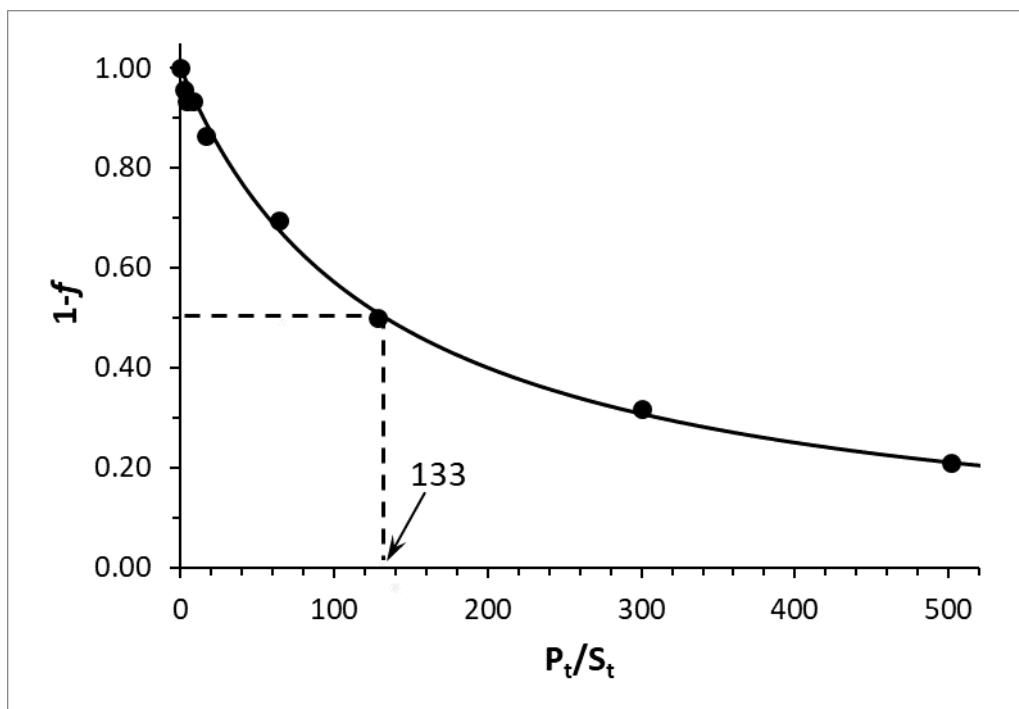


Figure 4.22: Interpretation of the binding equilibrium for Switch NM-1.

The mole fraction of the free switch ($1-f$) is plotted against P_t/S_t . The solid line shows a two-parameter fit to the data in Figure 4.21 generated with $K_d(\text{app}) = 5.3 \times 10^{-6}$ M, and limiting fluorescence, $I_\infty = 22$. A check on the quality of fit is shown at $1-f = 0.5$; a good agreement is shown on the graph where the predicted value of P_t/S_t of 133 should lie on the x-axis according to eq. 4.2. The experimental data and derived parameters are reported in Table 4.2

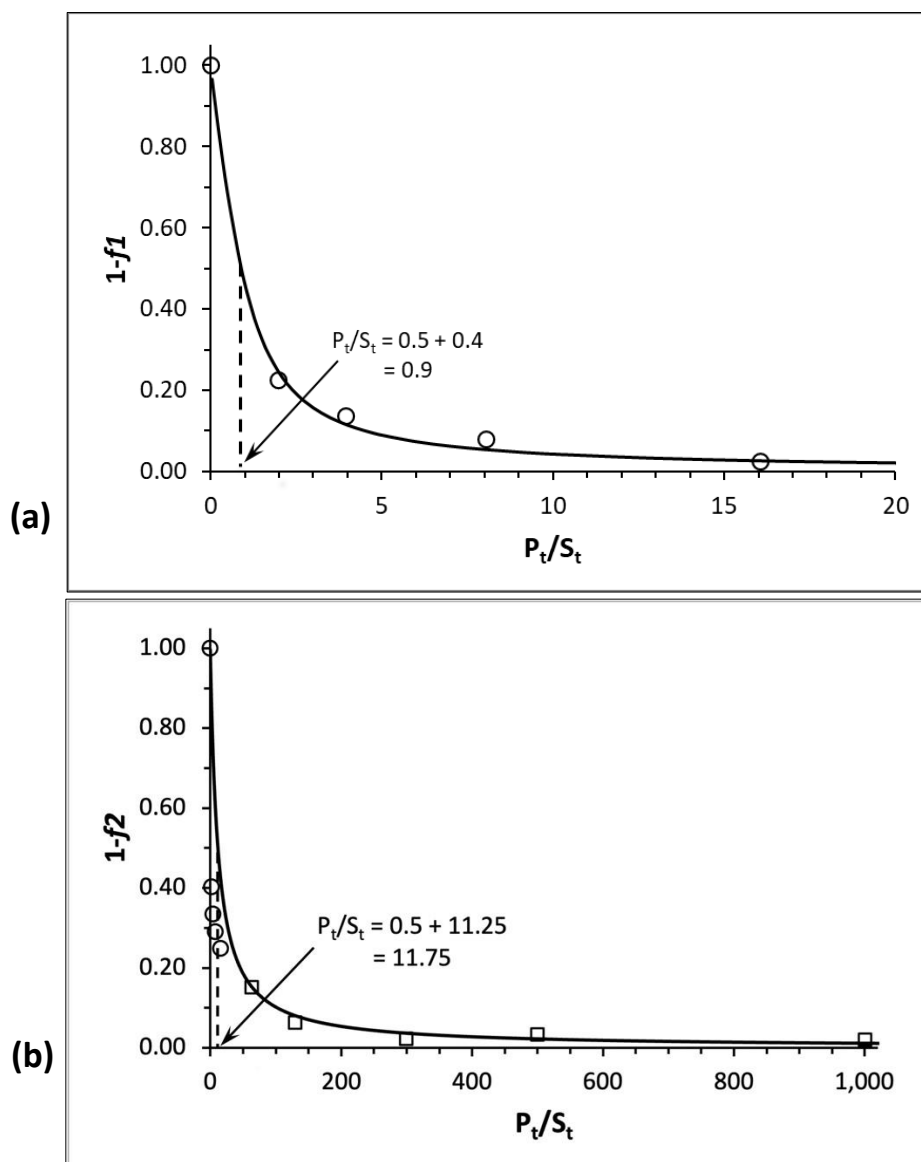


Figure 4.23: Interpretation of data for NM-2.

The data values of $1-f$ for NM-2 are plotted against P_t/S_t in two groupings. (a) The first five data points fit a curve with $K_d(\text{app}) = 16 \text{ nM}$, the same as $K_d(\text{probe})$. (b) The last five data points (squares) fit a curve with $K_d(\text{app}) = 450 \text{ nM}$. Checks on the quality of fit are shown at $1-f = 0.5$, where the corresponding values of P_t/S_t agree with calculations from eq. 4.2. The experimental data and derived parameters are reported in Table 4.3.

	P_t/S_t	I	I_{rel}	f	$1-f$
0.00E+00	0.0	43,555	100.0	1.000	0.000
8.00E-08	2.0	42,087	96.6	0.957	0.043
1.60E-07	4.0	41,340	94.9	0.935	0.065
3.20E-07	8.0	41,303	94.8	0.934	0.066
6.40E-07	16.0	38,963	89.5	0.865	0.135
2.56E-06	64.0	33,196	76.2	0.695	0.305
5.12E-06	128.0	26,550	61.0	0.499	0.501
1.20E-05	300.0	20,329	46.7	0.316	0.684
2.01E-05	501.3	16,702	38.3	0.210	0.790
	$I_\infty =$	9,582	22.0	0.000	1.000
	$S_t =$	4.00E-08 M			
		22%	of I_{max} is best fit for I_∞		
		5.30E-06	= Best fit $K_d(app)$		
		133	= $0.5 + K_d(app)/S_t$		

Table 4.2: Data for Switch NM-1

Fitting the first 5 Points							
Pt	Pt/St	I	Irel	f2	1-f2	f1	1-f1
0.00E+00	0.0	47,524	1.000	0.000	1.000	0.000	1.000
7.91E-08	2.0	27,645	0.582	0.598	0.402	0.775	0.225
1.58E-07	4.0	25,381	0.534	0.666	0.334	0.863	0.137
3.21E-07	8.0	23,907	0.503	0.710	0.290	0.920	0.080
6.42E-07	16.0	22,537	0.474	0.751	0.249	0.974	0.026
		21,861	= I_{∞}				
		4.00E-08	= S_t				
		46.0%	of I_{max} is best fit for I_{∞}				
		1.60E-08	= Best fit $K_d(app)$				
		0.90	= $0.5 + K_d(app)/S_t$				
Fitting the last 5 Points				f2	1-f2		
2.56E-06	64.1	19,259	0.405	0.850	0.150		
5.21E-06	130.2	16,344	0.344	0.937	0.063		
1.20E-05	300.3	14,988	0.315	0.978	0.022		
2.00E-05	500.6	15,341	0.323	0.967	0.033		
4.00E-05	1001.0	14,900	0.314	0.981	0.019		
		14,257	= I_{∞}				
		4.0E-08	= S_t				
		30%	of I_{max} is best fit for I_{∞}				
		4.50E-07	= Best fit $K_d(app)$				
		11.75	= $0.5 + K_d(app)/S_t$				

Table 4.3: Data for Switch NM-2

Conclusions and Prospects for Future Work

The nucleocapsid (NC) is highly conserved across HIV-1 subtypes and strains in the structural component of viral core proteins. Therefore, screening drug compounds that target HIV-1 NCp7 provide attractive targets for new anti-retroviral therapeutics because of the highly conserved nature of the zinc fingers in NCp7¹⁸ in the selection and packaging of RNA in the HIV-1 life cycle.

In this project, the engineering and synthesis of highly purified switch molecules are explored. Our work on 3-segment nucleic acid switches is described in Chapters 2-4. As described in Chapter 2, the unique 3-segment, reversible switch developed for high throughput screening (HTS) of drug targets was applied to the HIV-1 nucleocapsid (NC) protein. The Probe segment of the switch is a DNA hairpin that consists of an aptamer for the NCp7 protein. The theoretical equilibrium constant ($K_1 = 0.1$ to 0.001) of putative switches was calculated using V-OMP software. Attempts were made to directly synthesize the branched oligonucleotides using 3'-CPG columns and purify a 68mer full-length switch molecule as explained in Chapter 3. However, a high-purified product was not achieved. Optimizations of switch synthesis provided promising solutions to design and synthesize our switch molecules using the ligation method.

In Chapter 4, successful 3-segment switches of both types were constructed by a combination of (1) solid-support DNA synthesis on an in-house synthesizer adapted for our purposes, and (2) by enzymatic ligation to produce switch molecules with at least 90% purity as judged by mass spectrometry. Then fluorescence measurements using NM-1 and NM-2 with NCp7 were analyzed to demonstrate proof-of-principle for 3-segment nucleic acid switches. Increasing [NC] causes a dramatic decrease in CY3 fluorescence. The ON/OFF contrast ratio of NM-1 and NM-2 are 2.6 and 3.2, respectively. Assuming that the decrease in fluorescence is caused by an increasing fraction of switch bound to NC affords an experimental estimate of

the K_1 equilibrium constant, which is within the experimental error of the free energy estimates used to design these unprecedented DNA folded structures.

It should be possible to analyze switch equilibria similar to NM-2 by designing, building, and measuring the fluorescence of several switches with K_1 in the range of 0.05 to 10. Two-parameter fits to such data may not be sufficient to explain the variance in the data.

Practical switches with $K_1 > 0.1$ may be more useful than we have assumed. Although they will have a smaller dynamic range for fluorescence, they are very easily triggered. Such OFF \rightarrow ON switches may be useful in detecting low concentrations of a target, and ON \rightarrow OFF switches could detect drug candidates that bind their targets weakly. Weakly binding candidates sometimes have properties that make them useful as drugs (e.g., favorable absorption and biological half-life, lack of toxicity, etc.) and can be used as platforms to develop daughter molecules with greater efficacy, while retaining the useful drug-like properties of their parent.

As seen in Table 4.3 (bottom), the fluorescence of switch NM-2 levels off at 30% of I_0 , and other switches commonly level off at around 20%. Locked-OFF analogs can reach levels around 10% or less. An explanation is that a substantial fraction of the DABCYL quencher is damaged by the 50 or more cycles of harsh synthesis chemistry. If this is a major source of the reduced dynamic range, two-step ligation as described in section 4.8 should increase the ON/OFF contrast ratio.

Another possible explanation for the reduction from the expected dynamic range is that the 3D structures are distorted such that the ON-state does not have a large separation between fluorophore and quencher. Note that there are large flexible internal loops and bulges in the secondary structure (Figure 4.1. and Figure 4.14) for switches NM-1 and NM-2.

Still another contributor to the reduction in dynamic range might be the presence of substantial populations of intermediate states in structures that are designed for equilibration

by branch migration. The ~50 nsec decay time for fluorescence will be faster than the steps in branch migration but would average quencher-fluorophore distances in such intermediates. This effect would be strongest when the $\Delta\Delta G^\circ$ barrier for conformational exchange is zero at $K1 = 1$ and could contribute to the reduced contrast ratio for NM-2.

It is worth exploring other methods than fluorescence quenching in bulk solution to detect equilibrium populations of switches. Also, 3-segment switches could be incorporated into large RNA or DNA structures to monitor conformational changes, or used in single-molecule pulling by optical tweezers and similar methods to examine the thermodynamics and kinetics of DNA and RNA folding.^{78,79} Finally, a simplified and higher yield approach to solid-phase synthesis from N4 of the 5Me-dC brancher would be very useful, as would exploring other categories of branchers.

In addition to binding target species as we have illustrated, 3-segment molecules may prove useful in engineering nanostructures from DNA, RNA, or modified nucleic acids. As 3-segment switches are rapidly reversible they could perform critical functions in nanomolecular machines.

Completion of this analysis of ON \rightarrow OFF switches has revealed very useful aspects of the design, construction, purification, validation of purity, real-world testing, and interpretation of experimental results for 3-segment switches. Combined with Dr. Iyer's examination of OFF \rightarrow ON switches³⁴ there is now a strong basis for further research and development of these switches made from nucleic acids.

Materials and Methods

4.3. Synthesis and Analysis of Phosphorylated Test Sequence

dT Phosphoramidite, CPR II (oil), and other required solvents for ABI 392/394 synthesizer were purchased from Glen Research. The synthesizer was cleaned extensively using NH_4OH and ACN before the switch synthesis. dT phosphoramidite is diluted to 0.1 M using diluent ACN.

A test sequence, T_{15} (non-phosphorylated), and DMT-p T_{15} were synthesized on standard 0.2 μm scale dT CPG columns using the modified procedure on the synthesizer. After the synthesis, the β -cyanoethyl protecting group is removed using 3 mL of 10% DEA in acetonitrile used for 3 min and then washed using 1 mL ACN (HPLC grade) three times and dried under argon. The synthesized sequence with only dT phosphoramidites can be deprotected from the solid support by using 1 mL 30% NH_4OH for 2 h at RT, then followed by 0.5 mL NH_4OH wash and heated at 65°C. The samples were dried in a SpeedVac DNA concentrator 120 (Thermo Scientific) and resuspended in 150 μL Milli-Q purified water (Millipore Corp.) and a desalting procedure using MicroSpin G-25 Columns (GE Healthcare) was carried out. For comparison, the DMT group was deprotected from half the amount of DMT-p T_{15} using 20% acetic acid in water. These samples were ethanol precipitated and analyzed by HPLC.

4.4. Synthesis and Purification of Switch NM-1 using ligation

Synthesis of Core and a partial CY3 strand for one-step ligation

Phosphoramidites, DABCYL-controlled pore glass (CPG) columns, CY3, 5 Me-dC-brancher, and other required solvents for ABI 392/394 synthesizer were purchased from Glen Research. The synthesizer was cleaned extensively using NH₄OH and ACN before the switch synthesis. The required phosphoramidites, CY3, and brancher are diluted to 0.1 M using diluent ACN. The synthesis was performed as a two-part synthesis (0.2 μm scale) in the dark because of the light sensitivity of DABCYL and CY3.

The part I synthesis of the Core (55mer) began by using 0.2 μm 3'-DABCYL column where the Cover sequence was synthesized and the levulinyl protecting group was removed using hydrazine hydrate treatment. When the second part synthesis of Core with DMT-on was performed, a 0.2 μm dG column was used to synthesize a partial CY3 strand. To eliminate the β-cyanoethyl protecting group, 10% DEA in acetonitrile is used for 10 min and then washed using 1 mL ACN (HPLC grade) three times and dried under argon. Deprotection from the solid support was done by using 1 mL 30% NH₄OH for two hours, followed by 0.5 mL NH₄OH wash, and the solution was stored at room temperature for 17 h. The samples were dried in a SpeedVac DNA concentrator 120 (Thermo Scientific) and resuspended in 150 μL Milli-Q purified water (Millipore Corp.). Desalting was carried out using MicroSpin G-25 Columns (GE Healthcare) according to the manufacture protocol and a nanodrop UV absorption spectrometer was used to calculate the concentration.

One-step Ligation reaction (NM-1 and NM-2)

The concentration of the Cores and the partial Cy3-labeled strand were calculated using the reading from the absorbance at 260 nm on the nanodrop. The ligation mixtures were then prepared using the stock solutions of Core, partial CY3 strand, and corresponding splint. Before adding 10X ligase buffer and ligase to the mixtures, the sample containing snap top tubes was run using the program set as heat at 95°C for 5 min, then gradually cool to 4°C for 10 min by decreasing 1°C/ 30 sec in the thermocycler (Thermo Scientific™). The ligation mixtures were incubated at 4°C for 24 h, and the analytical denaturing gel was carried out the next day. To make the desired urea gel concentration, the following equations provided by national diagnostic were used;

$$V_c = (V_t) (X) / 25$$

$$V_b = 0.1 (V_t)$$

$$V_d = V_t - (V_c + V_b)$$

Where V_t = total volume; V_c = Volume of UreaGel Concentrate; V_b = volume of UreaGel Buffer; V_d = volume of UreaGel Diluent; X = % Gel Desired.

After the solution was prepared, 6 μ L TEMED and 120 μ L per 15 mL solution were added, and the gel was cast. The gel was allowed to polymerize for 2 h, and then a pre-run using 1X TBE buffer was carried out for 1 h. Since it is difficult to monitor the gel cassette's temperature, the buffer chamber temperature was measured to reach 40°C. Then the marker with tracking dye and samples were loaded, and the gel was typically run for between 1-2 h. The denaturing PAGE gel was then stained with Sybr gold and incubated by shaking for an hour. The denaturing PAGE gel was analyzed with a transilluminator and Kodak Gel Logic 100.

Ligated switch (NM-1 and NM-2) extraction from denaturing PAGE gel

D-tube Dialyzer Midi, MWCO 3.5 kDa, and supporting tray Midi were purchased from Millipore. This electroelution protocol is modified based on the reported protocol by Millipore. Before electroelution, the D-tube dialyzer was equilibrated with 1 mL of deionized water for 5 min. The D-tube dialyzer containing samples with ~400 μ L of 1X TBE buffer in a supporting tray was then placed in a horizontal agarose electrophoresis chamber, and different electric currents for elution (100-120 V) were applied. The minimum electroelution time for ligated switch molecules from various percent denaturing PAGE gel is about 2 h. Then the sample was released from the membrane by reversing the polarity of the electric current for 2 min. The elute was pipetted up and down ten times by avoiding the gel slice and transferred into a snap-top tube. An additional 400 μ L of 1X TBE buffer was added and electroeluted for 30 min, then reverse the polarity of the electric current was for a min. The sample containing snap-top tubes was centrifuged for 2 min at 14,000 \times g to pelletize the excess gel residues. The supernatant was transferred into new snap-top tubes and dried in Speed-Vac for 5 h. The pellet was resuspended in 200 μ L of Dnase-free water, and the ligated molecule was recovered by ethanol precipitation.

4.6. Locked-ON and Locked-OFF molecules

Synthesis of Locked-ON and Locked-OFF molecules

Phosphoramidites, DABCYL-controlled pore glass (CPG) columns, CY3, 5 Me-dC-brancher, and other required solvents for ABI 392/394 synthesizer were purchased from Glen Research. The synthesizer was cleaned extensively using NH₄OH and ACN before the switch synthesis. The required phosphoramidites, CY3, and brancher are diluted to 0.1 M using diluent ACN. The synthesis was performed as a two-part synthesis (0.2 μm scale) in the dark because of the light sensitivity of DABCYL and CY3.

The part I synthesis of the Locked-On and Locked-OFF was carried out by using 0.2 μm 3'-DABCYL column where the Cover, brancher, and Probe sequences (Locked-ON), and the Cover, brancher, and 5'-AAAAAAAAAAA-3' (Locked-OFF) were synthesized and the levulinyl protecting group was removed using hydrazine hydrate treatment. When part II synthesis, 5'-AAA-3' (Locked-ON) and 5'-CTTAGTTA-3' (Locked-OFF) were added, a 0.2μm dC column (Locked-ON molecule) and a 0.2μm dG column (Locked-OFF molecule) were used to synthesize partial CY3 strands. To eliminate the β-cyanoethyl protecting group, 10% DEA in acetonitrile is used for 10 min and then washed using 1mL ACN (HPLC grade) three times and dried under argon. Deprotection from the solid support was done by using 1 mL 30% NH₄OH for two hours, followed by 0.5 mL NH₄OH wash, and the solution was stored at room temperature for 17 h. The samples were dried in a SpeedVac DNA concentrator 120 (Thermo Scientific) and resuspended in 150 μL Milli-Q purified water (Millipore Corp.). Desalting was carried out using MicroSpin G-25 Columns (GE Healthcare) according to the manufacture protocol and a nanodrop UV absorption spectrometer was used to calculate the concentration

Phosphorylation of core molecules with T4 polynucleotide kinase (PNK)

T4 PNK was purchased from the New England Bio lab (NEB). Since ATP was not supplied in the kit, 1 μ L of 10X DNA ligase buffer containing 1 mM ATP was substituted with PNK buffer. Before phosphorylation, 1 μ M of the core molecule was heated at 95°C for 5 min, then gradually cooled to 4°C for 10 min by decreasing 1°C/ 30 seconds in the thermocycler (Thermo Scientific). Then ten units of T4 PNK and 1X DNase ligase buffer were added and incubated for 24 h at 4°C. Analytical 20% denaturing gel was carried out the next day.

Hydrolysis of switch NM-2, Locked-ON, and Locked-OFF molecules.

Prepare 5-10 μ L of 40nM sample with the appropriate amount of DNase I buffer (15 mM MgCl₂, 15 mM KCl, 10 mM Tris-HCl; pH 8.0) in the total volume of 2000 μ L was added to a quartz fluorescence cuvette (NSG precision cells) consisting of 10x 10 mm path length and 8x 1.5 mm stir bar. A Quantamaster QM- 4/2005SE fluorometer (Photon Technology International, Birmingham, NJ) was used to measure the fluorescence emission of CY3 at 564 nm with the excitation at 550 nm, a 2 nm excitation band-pass and 5 nm emission bandpass. After adding 20 units of DNase I (Promega), the fluorescence reading was acquired after 10 min using Felix 5.1 software.

4.7. Fluorescence Analysis of ligated switches

Switch binding assay

Individual wells of a 96-well, flat-bottom black plate (Greiner bio. one), were filled with corresponding NCp7 titration buffer, a final concentration of 40 nM NM-1, and a final concentration between 0 and 20 μM (500X excess) of NCp7 protein. The solution in each well was mixed by pipetting up and down. The excitation filter of 540/25 nm and the emission filter of 590/35 nm were used on a plate reader (BioTek SynergyTM 2, Multi-Mode Microplate Reader). The concentration of switches was calculated from the absorbance at 260nm obtained from nanodrop UV absorbance and an extinction coefficient of $575.52 \text{ cm}^{-1}\text{M}^{-1}$ using Beer law.

Supplemental Information

4.8. Ligation and Purification of NM-1 using two-step ligation

In two-step ligation, CY3- and DABCYL-labeled oligonucleotides will be ligated to the Core sequence using two different splint strands purchased from IDT. Initially, the CY3-labeled DNA strand will be ligated to the Core and the DABCYL-labeled DNA strand ligated later. Ligation of the CY3 strand will be guided by a splint that anneals to both the Toggle and the Cover strands by base-pairing across the brancher. Optimal conditions for ligation will be determined by varying the ratio of Core: CY3: Splint and then the relative amounts of ligated product will be estimated using denaturing PAGE.

The analytical and purification techniques described so far (IE, and RP HPLC). However, IE analysis of the Core for two-step ligation (figure S.4.1), and mass spectrometry analysis show that the phosphorylated Core coexists with failure, non-phosphorylated, and other sequences that make it challenging to purify by HPLC. Although GP purification (section 4.4.3) may be used for purification, recovery of a sufficient amount of the desired product is generally a concern. PAGE often provides single-nucleotide resolution and a high-grade product (>98%) but is challenging to scale up for oligonucleotide syntheses beyond 0.2-1 μ moles.

H-bonded stem-loop and duplex structures can be disrupted by denaturing agents including urea and formamide, effectively decreasing their melting temperatures. Denaturing urea polyacrylamide gel electrophoresis (PAGE) separates single-stranded DNA molecules based on molecular weight and can be used to analyze synthesized oligonucleotides and purify products from enzymatic reactions. For convenience and reproducible results, a pre-mixed urea gel system (National Diagnostics) was used to prepare various percentages of gel by altering the concentration of UreaGel concentrate and UreaGel diluent.

4.7.1 Ligation efficiency test and Optimization of Ligation conditions

Before performing the ligation reactions, the gel system and running buffer were considered, as well as ligation controls. For the running buffer in PAGE gels, Tris Borate EDTA (TBE) was selected over Tris Acetate EDTA (TAE) because TBE has a higher buffering capacity than TAE and can be used to run at high voltage for hours.⁸⁰ Polyacrylamide gels have several major advantages over agarose gels including greater resolving power for large quantities of DNA and higher purity in recovered DNA.⁸¹

The gel temperature is the most critical parameter for the UreaGel system where highly concentrated urea, 7.5M was used as a denaturant.⁸⁰ According to a technical bulletin from national diagnostics, the gel temperature must be above 40°C (usually in the range of 45-60°C) for denaturing PAGE gels. To achieve optimal separation, denaturing gels with various crosslinking percentages (12, 15, and 20%) were examined with different voltage and pre-run conditions. Analytical gels are small and thin to achieve high resolution with quick run times and have multiple wells to load different samples for comparison. Preparative gels can be much larger and thicker and may require overnight runs; they often have only a single well.⁸² It is also important to avoid overloading a gel and to load as much DNA as possible on preparative gels (to achieve good separation with as few gels as possible).

The amount of DNA per reaction is also important to maximize the production of a ligated switch. Concentrations of the Core in the range of 10-50 pmol per well were carried through the ligation reaction and analyzed using 20% denaturing gel. The efficiency of ligation decreases at high concentrations of Core, with the optimized condition determined using a 1 µM sample.

Results and Discussion

A control using IDT samples was analyzed using a 12% UREA PAGE gel. A modified ligation protocol developed by Dr. Caitlin Miller was used for this ligation reaction. After incubation of the ligation mixture overnight at 4°C, the samples were denatured by diluting them into 95% formamide and heating them to 95°C for three minutes before loading them on the gel. The addition of 6% loading dye containing glycerol was included to increase the density of the sample relative to the running buffer and sample tracking. After running the gel for 45 minutes, Sybr gold nucleic gel fluorescent stain was used to stain the gel for an hour.

Attempts using agarose gel electrophoresis purification were carried out because an easy and high-recovery gel extraction kit can be used. But the resolution of the ligated Core plus partial CY3 was not visible on a 4% gel after ethidium bromide staining (figure not shown). Agarose gels are routinely used to resolve DNA strands that are considerably longer than our switch components, so we chose not to pursue this option further.

In the analytical PAGE gel for ligation of an IDT control (figure not shown), swelling around gel bands was observed in the 1 μM sample due to the 6% loading dye containing glycerol. The disadvantage of TBE buffer is that it forms complexes with glycerol to generate a “salt wave” of altered conductivity which can distort gel patterns of a low range of DNA sizes.⁸⁰ Others have addressed this problem using either Tris-Taurine EDTA (TTE) buffer or sucrose in the loading dye as an alternative non-borate buffer or non-glycerol in the loading dye.⁸⁰ However, formamide in the sample buffer has a higher density than glycerol, and using TBE running buffer for denaturing PAGE gel, it is unnecessary to change the buffer or loading dye, however, loading of samples must be cautious due to the colorless solutions.

Since the samples were analyzed using 20% gel, no smeared pattern resulting from insufficient elongation of the second sequence can be observed. By comparing the intensity of the IDT control and the ligated product, improvements in the recovered yield of the Core will

be required to obtain sufficient amounts of a ligated switch for ESI-MS analysis and fluorescence studies.

All synthesized samples (the Core, CY3-Toggle donor, and Cover acceptor-DABCYL molecules) were analyzed with both IE HPLC and RP HPLC (figure not shown). In the IE-HPLC chromatogram of the 40mer Core for two-step ligation (Figure S.4.1), there were three peaks near each other. The peak with a retention time of 11.021 is assumed to be an accumulation of failure sequences after levulinyl group deprotection, the most prominent peak at 11.444 may be a non-phosphorylated Core, and a peak at 11.594 may correspond to the phosphorylated Core. Based on these assumptions, analysis of the IE-HPLC chromatogram, suggests that only 33% abundance of the phosphorylated product was detected.

To measure the efficiency of phosphorylation at the 5'-terminus, the Core was sent to Novatia, LLC for ESI-MS analysis. From the mass spec analysis of the phosphorylated Core (Figure S.4.2), the most prominent peak corresponds to the non-phosphorylated Core for two-step ligation (MW= 12448 Da), as predicted above from the HPLC chromatogram, and the second-highest peak is the phosphorylated Core for two-step ligation (MW =12527 Da). However, some impurities remained on the left and right side of these peaks corresponding to about 30-35% of the sample. The purity of the Core for two-step ligation was much lower than the 3'-DABCYL conjugated and directed synthesis of full-length oligonucleotide (NM-0) using 3'-DABCYL CPG (Chapter 3) which had ~40% purity after HPLC purification.

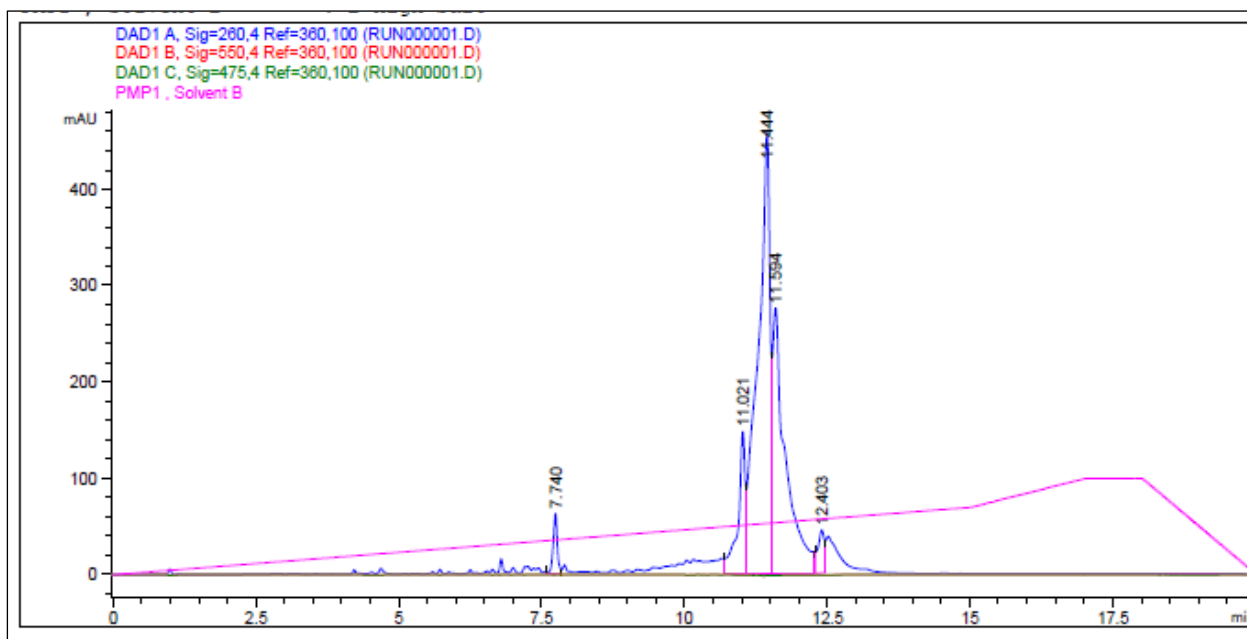


Figure S.4.1: Analytical IE-HPLC chromatogram of the Core for two-step ligation

The Core for two-step ligation was analyzed using a gradient of 0-60% buffer B at 55°C for 15 minutes. UV absorption at 260nm is shown in blue.

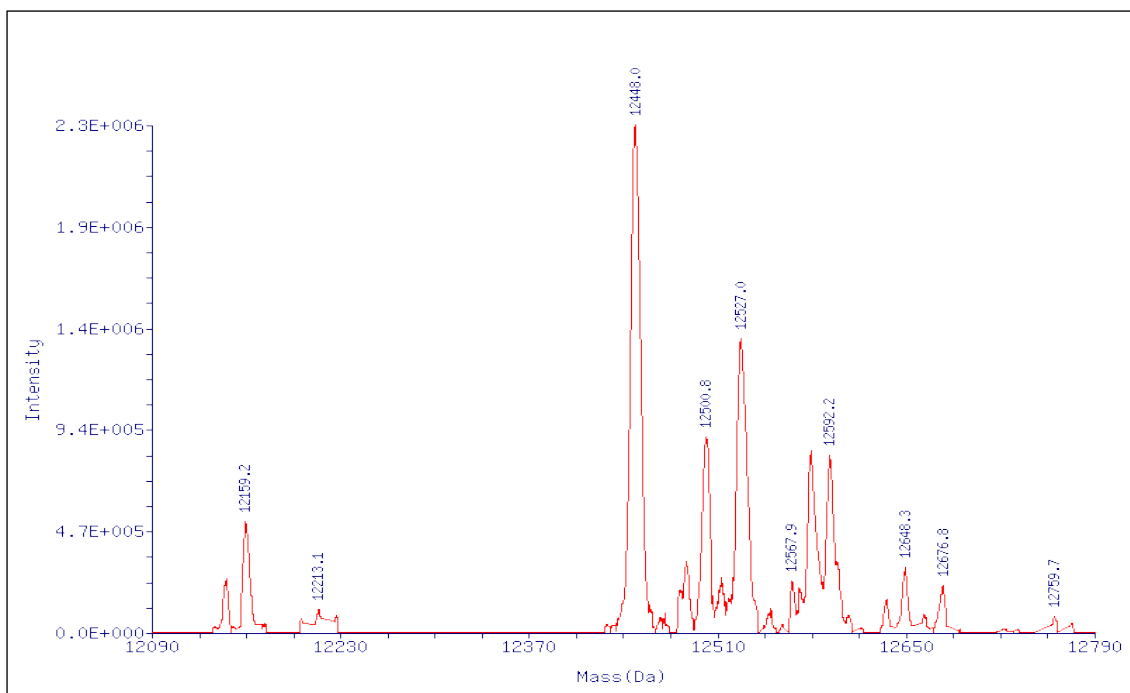


Figure S.4.2: ESI-MS analysis of a Core (40mer) for two-step ligation.

The most prominent peak corresponds to the non-phosphorylated Core for two-step ligation (MW= 12448 Da), as predicted above from the HPLC chromatogram, and the second-highest peak is the phosphorylated Core for two step ligation (MW =12527 Da). Impurities are identified on both left and right sides of the main peak.

4.9. Determination of Extinction Coefficient

Previously, we have relied on the calculated concentration of oligonucleotides from Nanodrop™ in various studies. The program assumes that 33ng/ μL and 50 ng/ μL for single and double-stranded oligonucleotide will have an absorbance of 1 to estimate the concentration using Beer's law. Since our switch molecules consist of three sequences capable of forming individual hairpins, the extinction coefficient of the molecules is required to determine experimentally.

The Beer-Lambert Law, absorbance $(A) = \epsilon cl$, where ϵ refers to the extinction coefficient, c is the concentration, and l is the pathlength, is used based on the extinction coefficient at 260nm to determine the concentration of DNA samples. Several methods calculate the extinction coefficient of oligodeoxynucleotide (ODN) at 260nm. The first method assumes a single-stranded oligodeoxynucleotide (SSO) and a double-stranded oligodeoxynucleotide (DSO) at 33 ng/ μL and 50 ng/μL will have an absorbance of 1 by neglecting the base composition and sequence. Therefore, the extinction coefficient of an SSO and DSO with length n can be estimated by $10 \text{ mM}^{-1}\text{cm}^{-1} \times n$ and $13.2 \text{ mM}^{-1}\text{cm}^{-1} \times n$. The second method calculates the extinction coefficient of a molecule by summing the individual nucleotide extinction coefficients ($\epsilon_{\text{sum}} = n_A\epsilon_A + n_C\epsilon_C + n_G\epsilon_G + n_T\epsilon_T$), ignoring the sequence. The third method is the nearest neighbor (NN) approximation, which requires the extinction coefficients of 4-mono and 16-dinucleotide possibilities by taking account of base composition and sequence. The extinction coefficient calculation using the NN method constrains a long chain sequence due to variations in base stacking caused by the secondary and tertiary structure of DNA. These methods can only approximately calculate the extinction coefficient within 20% of the values measured using the exhaustive hydrolysis method.⁷⁷

Enzyme degradation of oligodeoxynucleotide (ODN) can be used to measure ODN's extinction coefficient accurately. An increase in absorption intensity, termed the hyperchromic effect, is observed when total hydrolysis degrades oligonucleotides to mononucleotides.⁷⁷ For SSOs, the 3' to 5' phosphodiesterase activity of snake venom phosphodiesterase (SVP), known as an exonuclease, is used to accomplish complete hydrolysis. For SSOs containing double-stranded regions and DSOs, the combination of SVP with deoxyribonuclease I (DNase I), an endonuclease, is required to hydrolyze these molecules' mononucleotides rapidly. The extinction coefficient of the molecules can then be calculated using the following equation:

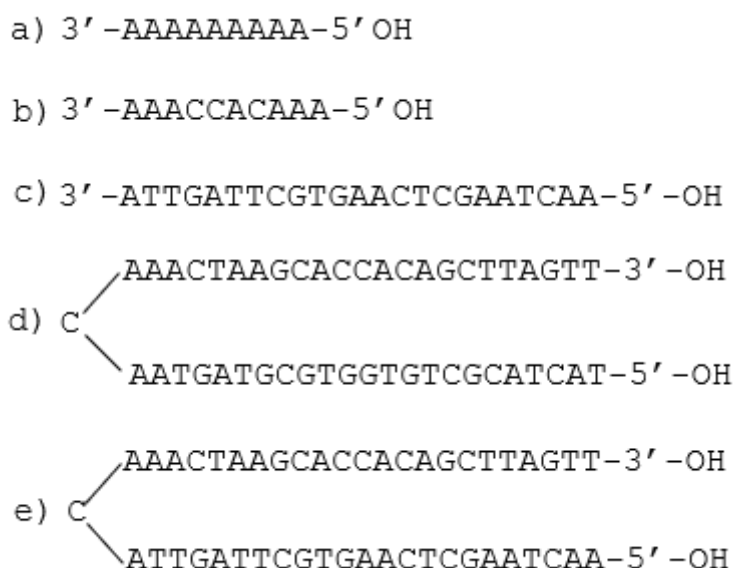
$$\epsilon_{\text{ODN}} = A_0 / A_{\infty} (\epsilon_{\text{sum}})$$

A_0 / A_{∞} is the ratio of pre and post-hydrolysis, and $\epsilon_{\text{sum}} = n_A \epsilon_A + n_C \epsilon_C + n_G \epsilon_G + n_T \epsilon_T$ where n_i and ϵ_i are the numbers and the extinction coefficient corresponding to the respective individual nucleotide in the ODN. The extinction coefficient at 260nm for pdA, pdC, pdG, and pdT are 15.06, 7.10, 12.18, and 8.56 $\text{mM}^{-1}\text{cm}^{-1}$ from work published by Michael J. Cavaluzzi.⁸³

Due to the presence of hairpin loops in the switch structure, the calculated concentration of the molecules in NanodropTM (method 1) remains uncertain. Therefore, the extinction coefficients of the molecules mimic the cover base pairing with the probe, and the toggle and the cover base pairing with the toggle were analyzed by exhaustively hydrolyzing these molecules with the combination of SVP with DNase I.

Results and Discussion

Determining the accurate extinction coefficient of ODNs can be accomplished by enzyme degradation using a combination of SVP and DNase I enzymes. These powerful enzymes can effectively hydrolyze the molecules that are DSOs, SSOs, or SSOs containing base pairing regions. The following sequences were chosen to correspond to the SSO of Locked-OFF (a), SSO of Locked-ON (b), a toggle sequence (c), DSO of a cover, and probe (d), and DSO of a cover and toggle (e) to perform enzyme degradation.



When the absorbance vs. time for the hydrolysis experiment was plateaued for at least ten min, hydrolysis of oligonucleotides by a combination of SVP and Dnase I was completed. All molecules were able to hydrolyze within 15 min exhaustively. The hydrolysis of a double-stranded sequence containing a cover and a probe is shown in the figure. The sample was initially treated with 0.15-unit SVP for 5 mins, followed by eight units of Dnase I for 10 min. Hyperchromic shift as a function of time at 25°C was observed as SVP was added, and a sudden increase of intensity resulted upon the addition of Dnase I as expected. One trial with 30 min absorbance readings (data not shown) was carried out to ensure the oligonucleotides were exhaustively hydrolyzed. Since it is infeasible to distinguish between enzyme decay and

complete hydrolysis from the hyperchromic shift, thymidine *p*-nitrophenyl ester was added to observe the change in color from colorless to yellow, indicating the presence of *p*-nitrophenoxide. Thus, SVP is still active and is not limiting hydrolysis.

The hydrolysis of other molecules was achieved in the same way as shown in figure S.4.3. The extinction coefficients of these ODNs were then calculated using $\epsilon_{\text{ODN}} = A_0 / A_{\infty}$ (ϵ_{sum}) equation. The complete data set from enzyme degradation is shown in table S.4.1. The absorbance of a buffer containing 0.15 unit SVP or a combination of 0.15 unit SVP and eight units DNase I was measured to subtract from the initial and the final absorbance reading. The percentage absorbance increase of various samples is in the range of 27-39%. The most significant hyperchromic shifts were observed in SSO Locked-ON and Locked-OFF samples. It is expected because fewer bonds hold the monomers together compared to DSOs and SSOs containing base pairing regions. These experiments' coefficient of variance (CV) is in the range of 0.5-2%. The extinction coefficient of the switch, Locked-ON, and Locked-OFF sequences was then calculated using the average extinction coefficients obtained from enzyme degradation experiments.

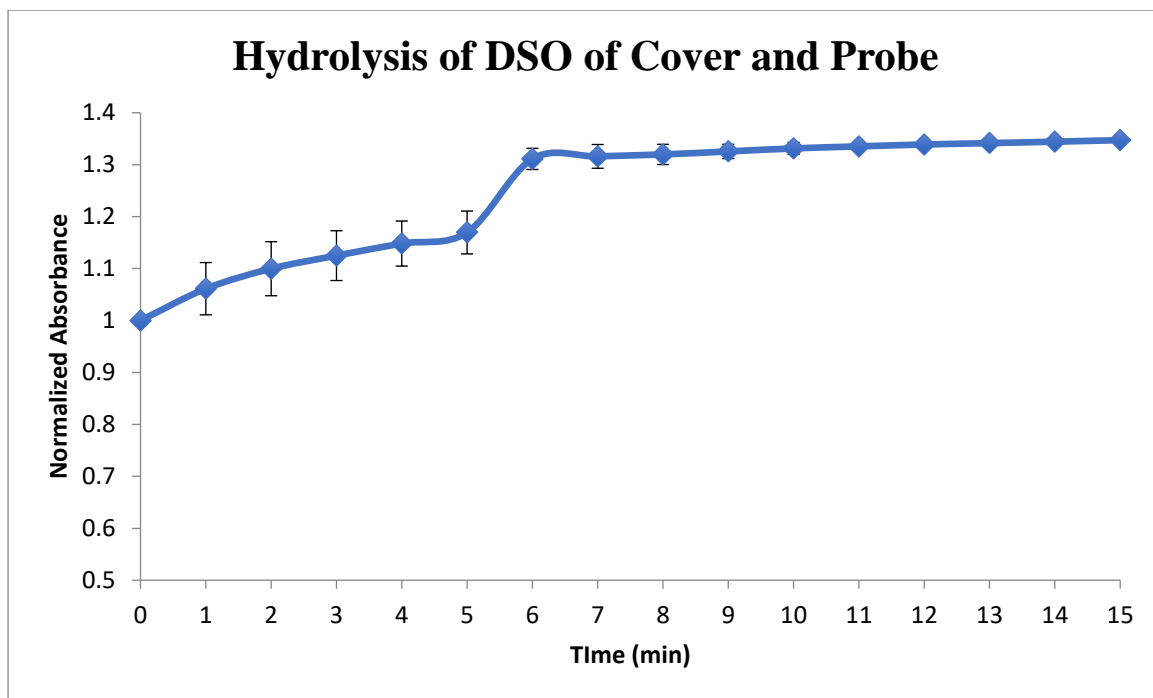


Figure S.4.3: Enzyme degradation of a double-stranded molecule consisting of a cover and a probe.

Three trials of hydrolysis of DSO of Cover and Probe were carried out with SVP and DNase

I.

Sample	Run	Initial Abs.	Final Abs.	Init/final	% Abs Increase	ϵ sum	ext. coeff.	Avg. ϵ
	1	0.1168	0.1608	0.7264	37.67	126.7	92.03	
sso Locked-ON	2	0.117	0.1628	0.7187	39.15	126.7	91.06	92.23
	3	0.1174	0.1589	0.7388	35.35	126.7	93.61	
	1	0.1451	0.1917	0.7569	32.12	150.6	113.99	
sso Locked-OFF	2	0.1371	0.1874	0.7316	36.69	150.6	110.18	113.64
	3	0.1476	0.1904	0.7752	29.00	150.6	116.75	
	1	0.1101	0.1399	0.7870	27.07	252.4	198.64	
Toggle	2	0.1297	0.1698	0.7638	30.92	252.4	192.79	194.23
	3	0.1095	0.1445	0.7578	31.96	252.4	191.27	
	1	0.1064	0.1434	0.7420	34.77	513.5	381.01	
Cover+ Probe	2	0.1299	0.1737	0.7478	33.72	513.5	384.02	381.59
	3	0.1042	0.1409	0.7395	35.22	513.5	379.75	
	1	0.1123	0.1474	0.7619	31.26	517	393.89	
Cover+ Toggle	2	0.1281	0.1696	0.7553	32.40	517	390.49	392.81
	3	0.125	0.164	0.7622	31.20	517	394.05	

Table S.4.1: Extinction coefficient data for the various molecules.

The correction factors are applied for initial and final absorbance. The unit of extinction coefficient is in $\text{mM}^{-1}\text{cm}^{-1}$.

The mean per base and standard deviation per base were calculated depending on the molecules' length and the data set's precision (see table S.4.2). Since there are DABCYL and CY3 covalently attached to the switch, Locked-ON and Locked-OFF, their extinction coefficients which are $7.3 \text{ mM}^{-1}\text{cm}^{-1}$ and $4.9 \text{ mM}^{-1}\text{cm}^{-1}$, are obtained from the technical bulletin of Glen Research. The total extinction coefficient is then calculated using the following equation:

$$\epsilon_{\text{total}} = \left[\left(\frac{n_1}{\text{total number of bases}} * \mu_1 \right) + \left(\frac{n_2}{\text{total number of bases}} * \mu_2 \right) + \frac{\epsilon_{\text{dabcyl}}}{n_1} + \frac{\epsilon_{\text{cy3}}}{n_2} \right] * n_3;$$

where n_1 and n_2 are the length of cover and probe, and toggle for switch molecule. n_3 is the total number of bases + 2. μ_1 and μ_2 are the mean per base corresponding to cover, probe, and toggle sequences.

Exception for Locked-OFF molecule, the denominator will have a total number of bases of 47 instead of n_1 and n_2 due to having both CY3 and DABCYL on the cover and toggle duplex.

For the switch molecule, the total extinction coefficient, ϵ_{total} , is $618.7 \text{ mM}^{-1}\text{cm}^{-1}$, the ϵ_{total} of Locked-ON is $528.8 \text{ mM}^{-1}\text{cm}^{-1}$, and Locked-OFF has ϵ_{total} of $540.4 \text{ mM}^{-1}\text{cm}^{-1}$. Since the base compositions of various switch molecules are similar, the calculated extinction coefficient of $618.7 \text{ mM}^{-1}\text{cm}^{-1}$ will be used to determine the concentration of switches.

Sample	Number of bases	Mean per base	Stdev per base
ss0 Locked-ON	10	9.28	0.15
ss0 Locked-OFF	10	11.4	0.26
Toggle	23	8.44	0.14
Cover+ Probe	47	8.12	0.04
Cover+ Toggle	47	8.36	0.09

Table S4.2: Data required to calculate the extinction coefficient of various molecules.

Materials and Methods

HPLC Purification of Core (40mer) for two-step ligation.

The samples were heated at 90 degrees for 3 min and snap-cooled to eliminate the formation of the duplex structure before HPLC purification on an Agilent 1200 Analytical HPLC equipped with a Dionex DNAPac PA200 4x250 mm analytical IE column. Buffer A (0.01M Tris-HClO₄, 20% ACN, pH8) and buffer B (0.01M Tris-HClO₄, 0.33M NaClO₄, 20% ACN, pH8) were made and filtered for Ion exchange HPLC. Ion exchange purification was carried out using multiple 5 μ L injections at 55°C, 1.2 mL/min flow rate using 0-60% gradient of HPLC buffer B for 30 min and 2 min post gradient conditions for an overall time of 32 min. The peaks were collected using a time-based method, and the samples were dried on the Speed-Vac and then desalted using MicroSpin G-25 columns (GE Healthcare). The desalted samples were dried by Speed-Vac and resuspended with 150 μ L of DNA-free water. 100 μ L stock solutions of switches were prepared for purified peaks, and analytical HPLC of purified peaks was run for validation.

Enzyme Degradation.

Snake venom phosphodiesterase (SVP) was purchased from Worthington Biochemical Corp., and thymidine 5'-monophosphate *p*-nitrophenyl ester was obtained from Sigma. DNase buffer (15mM MgCl₂, 15mM KCl, 10mM Tris-HCl; pH:8.0) was prepared. All ODN samples were heated at 94°C for 3 min and snap-cooled before experiments. The concentration of ODN samples in 200 µL DNase buffer gave the initial absorbance between 0.1 to 0.3 was determined. The enzyme degradation experiment then proceeded: The absorbance of buffer background, DNase buffer with 0.15 unit SVP, or the combination of 0.15 U SVP and 8 U DNase I was routinely taken before the experiment. 160-180 µL DNase buffer with 20-40 µL ODN samples was added in a clean cuvette. The initial absorbance (A_0) was determined by subtracting the absorbance of a buffer. 0.15 unit SVP was added, and the increase in absorbance was monitored for 5 min. Then eight units of DNase I were dissolved in the solution and monitored for 10 min. The final absorbance (A_∞) was recorded after the absorbance vs. time graph was plateaued for at least 10 min. To confirm that the hyperchromic shift is only because of enzyme degradation, 2 µL of a 100 nM thymidine 5'-monophosphate *p*-nitrophenyl ester was added to 100 µL of a sample. The change in color from colorless to yellow indicated that SVP was still active.



Locked ON

$\Delta G^\circ (25\text{ }^\circ\text{C}) = -20.18\text{ kcal/mol}$
 $\Delta H^\circ = -133.47\text{ kcal/mol}$
 $\Delta S^\circ = -379.95\text{ cal/mol K}$
 $T_m = 76.85\text{ }^\circ\text{C}$; [Locked_ON] = 1 mol/L; [CY3] = 1 mol/L
 [monovalent] = 0.2050 mol/L
 [Mg²⁺] = 1.000e-006 mol/L



Locked OFF

$\Delta G^\circ (25\text{ }^\circ\text{C}) = -17.80\text{ kcal/mol}$
 $\Delta H^\circ = -129.53\text{ kcal/mol}$
 $\Delta S^\circ = -374.73\text{ cal/mol K}$
 $T_m = 71.23\text{ }^\circ\text{C}$; [Locked_OFF] = 1 mol/L; [OLIGO4] = 1 mol/L
 [monovalent] = 0.2050 mol/L
 [Mg²⁺] = 1.000e-006 mol/L

Figure S.4.4 The designs of core molecules of Locked-ON (Top) and Locked-OFF (Bottom) using V-OMP.

**Part II: Re-engineering of Recombinant human Cyclin T1-Tat
Chimera with SUMO Fusion in Escherichia Coli**

Chapter 5: Re-engineering a human Cyclin T1 (hCycT1)- Tat Chimera

Chapter Summary

Due to increasing drug resistance for current antiretroviral therapeutic (ART) treatments, it is important to add drugs targeting other aspects of HIV-1 infection to manage AIDS. Tat (trans-activator of transcription) protein up-regulates the transcription of viral-specific proteins by a factor of 1,000³, which makes Tat a very attractive drug target. However, Tat is a difficult protein for structural studies and drug discovery. It often aggregates and is difficult to clone and overexpress. Tat also interacts with TAR, a long stem loop that begins HIV-1 RNA transcripts and a host protein, the cyclin T1 (CycT1) subunit of the positive transcription elongation factor complex (P-TEFb). Thus, the conformation of isolated Tat, or even Tat in a Tat-TAR complex, may not resemble the active structure.

A chimeric protein consisting of a short segment of CycT1, and Tat in a complex with TAR offers what could be a very effective target for anti-Tat drug discovery.³ However, high-level expression, and purification of this fusion protein in *Escherichia coli* (*E. coli*) were difficult because CycT1-Tat forms inclusion bodies, which makes it difficult to purify and obtain a high concentration of active CycT1-Tat. This project introduces a commercially available pET-SUMO cloning vector for the high-level expression of various CycT1-Tat chimeras. A fusion protein, His₆-SUMO-X, where X is the protein of interest can be purified using Ni²⁺ resin. Then, a SUMO protease, His₆-Ulp1₄₀₃₋₆₂₁-His₆ cleaves His₆-SUMO to produce the target protein.

Four variant sequences, differing at the N-terminus of CycT1, and the C-terminus of Tat were designed and tested for their ability to be expressed in *E. coli*. The genes associated with these protein sequences were purchased from IDT with codons that are used by *E. coli*

BL21. The SUMO-CycT1-Tat fusions were inserted into a plasmid and transformed in *E. coli* BL21 (DE3). After induction of mRNA production from the plasmid, SUMO-CycT1-Tat was expressed as a soluble protein from all four variants. We chose the most efficiently expressed variant for purification, which was much simpler compared to prior attempts with GST- and MBP fusions. After cleavage of the SUMO tag, the molecular weight of partially purified CycT1-Tat (F1) was confirmed with mass spectrometry (MALDI-TOF).

For initial characterization of the TAR-CycT1-Tat complex, an Electrophoretic Mobility Shift Assay (EMSA) was carried out with partially purified CycT1-Tat protein using TAR-31 and truncated versions that altered the hairpin loop (TAR-H24) and deleted the bulge loop (TAR-B25). A TAR-CycT1-Tat complex was formed and a band shift was observed in the binding of CycT1-Tat to TAR-31 and a lower affinity complex with TAR-B25, but TAR-H24 did not bind.

5.1. Cyclin T1 (CycT1)-Tat Chimera

The Tat-CycT1 chimeric proteins produced in this work present a novel target for drugs that could decrease the production of HIV-1 RNA transcripts from proviral DNA in infected cells by a factor of 100 or more. The original chimera of CycT1(249-280)-25aa linker-Tat(1-101) was developed by Fujinaga at UC San Francisco.³ We constructed variants of this sequence and modifications were made to simplify the cloning and overexpression for drug discovery and structural studies.

Previously in the Borer Lab, Dr. Collin Fischer expressed a chimera derived from a plasmid for a GST-CycT1(249-280)-(25aa linker)-Tat-101 fusion protein⁴⁶ that was generously provided by Dr. Fujinaga. During attempts at large-scale production of this chimera in *E. coli* for structural and thermodynamic studies, GST fusion protein did not express well and accumulated in an insoluble form, inclusion bodies. Dr. Fischer developed a His₆-MBP-CycT1 (257-280)-Tat-101 (figure 5.1) fusion. The Maltose binding protein (MBP) tag is used to enhance the protein's solubility; the tag was then removed using a tobacco etch virus (TEV) cleavage site, located prior to the CycT1 sequence for purification of the desired chimera.

(a)					
PN	RLKRIWNWRA	CEAAKKTAD	DRGTDEKTSE	QTMPEQKLIS	EEDLAMEFLE
	IDPVD	MEPVDPNLEP	WKHPGSQPRT	ACNNCYCKKC	CFHCYACFTR
	KGLGISYGRK	KRRQRRAPQ	DSQTHQASLS	KQPASQSRGD	PTGPTESKKK
	VERETETDPF	D			
(b)					
	NWRA	CEAAKKTAD	DRGTDEKTSE	QTMPEQKLIS	EEDLAMEFLE
	IDPVD	MEPVDPNLEP	WKHPGSQPRT	ACNNCYCKKC	CFHCYACFTR
	KGLGISYGRK	KRRQRRAPQ	DSQTHQASLS	KQPASQSRGD	PTGPTESKKK
	VERETETDPF	D			

Figure 5.1: CycT1-Tat Chimera.

(a) A chimeric protein³ consists of CycT1 (249-280), a 25 aa spacer with a Myc antibody tag, and Tat-101. A leading proline and trailing aspartate are inserted.

(b) The chimeric protein⁴⁶ consists of CycT1 (257-280), a 25mer spacer with a Myc antibody tag, and Tat-101 studied by Dr. Collin Fischer.

Dr. Fischer prepared a fusion gene (TEV Protease cleavage site-CycT1-Tat) and then the pDEST-HisMBP-CycT1(257-280)-25aa-Tat-101 expression vector was constructed, where she chose a gene sequence that removed the initial eight aa-residues (PNRLKRIW) from Fujinaga's CycT1 fragment. The recombinant protein chimera, His₆-MBP-CycT1-25aa-Tat protein, was purified on a HisTrap column (GE Bioscience) and then the His₆-MBP tag was cleaved using TEV to yield CycT1-25aa-Tat fusion protein.

Due to the multiple steps involved in cloning to produce the pDEST-HisMBP (257-280) hCycT1-Tat chimera expression plasmid, the expression of variant chimeras to analyze the binding affinity with wild-type TAR RNA may be challenging. Therefore, a commercially available pET-SUMO cloning vector is introduced for the high-level expression of variant CycT1-Tat chimeras. The goal of this project was to optimize the length of the CycT1 and Tat segments for High Throughput Screening (HTS) and anti-Tat drug discovery.

5.1.1. Optimization of CycT1 and Tat segments

To activate transcription elongation of the human immunodeficiency virus (HIV-1), Tat protein binds to Cyclin T1 (CycT1) of the positive transcription elongation factor (p-TEFb) and interacts with transactivation response (TAR) RNA in the long terminal repeat (LTR) region. Tat consists of two functional regions which are an arginine-rich motif (ARM) for TAR RNA binding, 49RKKRRQRRR57, and a cysteine-rich region (residues 21-40) where Tat forms a ZnF2 with Cys261 of CyT1.^{84, 38}

Similar to the full-length chimera (CycT1-280-Tat-101), the mutant CycT1 (249-280)-Tat-101 formed a complex with TAR RNA to promote transcription elongation of HIV mRNA in the murine cell.³ Other studies have also shown that CycT1 (249-280) is involved by providing a stable binding surface with Tat for TAR RNA binding. Therefore, it is beneficial to include CycT1 (249-256) in the chimera for structural and thermodynamic studies. Exon 2 of Tat contains a splicing silencer essential for *in vivo* activity but is not required for transcription^{41, 85}. So, a truncated carboxyl terminal of Tat protein should be sufficient for structural and thermodynamic studies. Therefore, modification to include more of the N-terminal region of CycT1(250-280) and to remove residues from Tat (87-101) was carried out.

All amino acids except an N-terminal proline of the target protein are effectively cleaved by SUMO protease. SUMO cleavage occurring at the active site of SUMO protease requires SUMO to pass through a constrictive hydrophobic tunnel. The structural change imposed by prolines near the cleavage site may inhibit the proteolytic cleavage.⁸⁶ Since the first amino acid residue of the CycT1 segment (figure 5.1) is a proline, which may reduce the cleavage efficiency of SUMO protease due to hydrophobicity, it was replaced with a serine residue.

```
SN-RLKRIWNWRA-CEAAKKTAD-DRGTDEKTSE-  
QTMPEQKLIS-EEDLAMEFLE-IDPVD-  
MEPVDPNLEP-WKHFGSQPRT-ACNNCYCKKC-CFHCYACFTR-KGLGISYGRK-  
KRRQRRRAPQ-DSQTHQASLS-KQPASQSRGD-PTGPTEKSKK-VERETETDPF-D
```

Figure 5.2: Optimized Chimera of HIV-1 CycT1-Tat.

CycT1 (249-280) top line (with the substitution, P249S), a **25mer spacer with a Myc antibody tag** second line, and **Tat (1-86)**, bottom two lines. Amino acid residues marked in black are not part of the CycT1, myc, or Tat sequences. P249S facilitates SUMO protease cleavage. This sequence is 143mer, Theoretical pI / Mw: 8.94 / 16416.47

After overexpressing the recombinant fusion protein, purification of His₆-SUMO tag protein can be carried out with a Nickel-NTA column. The native protein is then generated by using SUMO protease to cleave the N-terminal peptide containing the 6xHis tag and SUMO. Studies have shown that SUMO protease might not cleave the SUMO protein efficiently when the first amino acid residue following SUMO is a proline, lysine, valine, or leucine. An additional amino acid, preferably serine should be included in the N-terminus of the protein of interest.

The DNA for preparing CycT1 and Tat³ came directly from the human sequence for CycT1, and the Tat sequence came from infected humans. Therefore, the set of mRNA codons from the plasmid DNA was also adapted for protein biosynthesis in humans. Bacteria and humans have differing amounts of tRNA molecules for the redundant codons that specify one amino acid, and it is common for a humanized codon to specify a low abundance tRNA in bacteria. Inefficient expression is often the result of cloned human gene sequences. In this work, DNA sequences for genes were altered to contain codons optimized for expression in *E. coli*. This will maximize the rate and amount of protein biosynthesis for the cloned gene.

The sequence of Fujinaga's original fusion protein is shown in Figure 5.1(a), with the **CyclinT1** fragment shown in green, the 25 aa spacer in black and blue (a *myc* antibody binding sequence), and **Tat** in magenta. The flexible spacer can allow C251 of CycT1 to participate in zinc finger formation as noted in Figure 1.4(a). The *myc* antibody can be used to verify that the plasmid has not ejected the cloning insert.

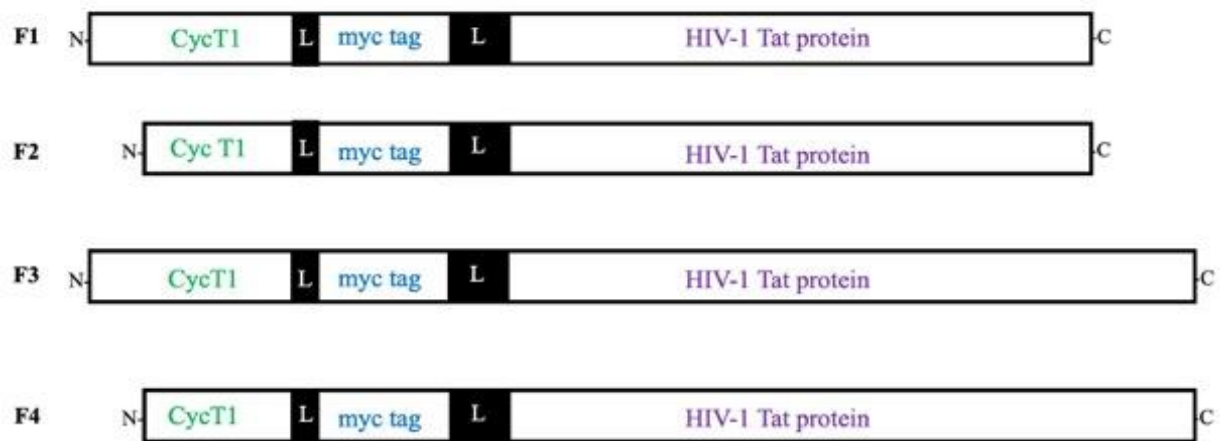


Figure 5.3: Four CycT1-Tat fusion proteins were prepared in this work. The N-terminus is on the left and the C-terminus is on the right.

F1 has the full CycT1(249-280) fragment used by Fujinaga³ but Tat(1-86) instead of Tat(1-100), and F1 also has the P249S substitution.

F2 (-N) is CycT1(257-280)-25aa-Tat (1-86) is truncated at the N-terminus compared to F1,

F3 (+C) is CycT1(249-280)-25aa-Tat (1-100) is longer at the C-terminus compared to F1,

F4 (-N,+C) is CycT1(257-280)-25aa-Tat(1-100) used by Fischer.⁴⁶

To optimize the length of CycT1 and Tat segments, four different fusion proteins (Figure 5.3) were cloned and expressed in Mach1-T1 competent *E. coli* cells. The F4 chimera, consisting of CycT1 (257-280)-25aa-Tat-101 is the sequence that Dr. Fischer studied for her work on the TAR-CycT1-Tat complex. As mentioned in the Introduction chapter, R251 and R254 in the CycT1 protein segment appear to be involved in stabilizing the TAR-CycT1-Tat complex (see discussion of Figure 1.5). Exon 2 in Tat (figure 1.4(b), residues 73-101) functions as a splicing silencer, so it probably is not required to up-regulate transcription. Also, for *in vivo* activity, a truncated C-terminal Tat (1-86) may be sufficient for thermodynamic studies and simplifying structural analysis by NMR or x-ray diffraction. Therefore, we chose F1, CycT1(249-256)-25aa-Tat(1-86), as the reference chimera to which the others would be compared.

5.2. Cloning, Overexpression, and Purification of His₆-SUMO tagged hCycT1-Tat proteins

Recombinant proteins are widely expressed in microbes for large-scale production due to their fast and inexpensive, high-yield production, well-characterized genetics, and the availability of various molecular tools.⁸⁷ 90% of structural and functional studies of proteins in the Protein Data Bank are based on the proteins produced in *E. coli*.⁸⁸ The affinity-based method, immobilized metal affinity chromatography (IMAC) with a polyhistidine tag (His₆), is widely used for protein purification. However, the limitations of *E. coli* recombinant protein production systems are the absence of downstream machinery to perform posttranslational modifications, which may cause proteins to misfold, resulting in aggregation into inclusion bodies.⁸⁸

Therefore, a fusion partner protein, a stable peptide, or a solubility-enhancing protein with affinity tags linked with the protein of interest can be used to improve the solubility and aid in purification.⁸⁹ In the solubility study of Maltose-binding protein (MBP), glutathione S-transferase (GST), and thioredoxin (TRX)⁹⁰, MBP not only is a better solubilizing agent but also may assist in the proper folding of the target protein into its active conformation. The protease cleavage site is designed between the expression tag and a target protein to yield highly purified recombinant protein. Most proteases such as factor Xa, tobacco etch virus protease (TEV) and enterokinase may have low cleavage activity because of steric hindrance caused by a fusion partner at or near the cleavage site for the target protein; protease recognition sites typically include 4-8 residues.^{89, 91, 92}

To overcome the solubility and cleavage efficiency for expression and purification of the recombinant protein, a small ubiquitin-related modifier (SUMO) serves as a solubility enhancer (chaperoning effect), and a protease cleavage site.^{92, 93} Because of its structure with a hydrophobic inner core and a hydrophilic outer surface, attachment of SUMO on the N-

terminal of the target protein mediates the solubility of the recombinant protein.⁹¹ Additionally, SUMO protease is a robust enzyme that recognizes the tertiary structure of SUMO, unlike other proteases. The digestion by SUMO protease occurs at the conserved Gly-Gly motif at the C terminal of SUMO to produce the desired protein. The protease cleavage reaction can be carried out under a wide range of conditions, a broad range of pH (5.5-10.5), temperature (4-37°C), and various salt concentrations.⁹¹

The advantages of SUMO fusion demonstrate that the expression and purification of His₆-SUMO-CycT1-Tat fusion protein (figure 5.4) can be successfully carried out using a commercially available pET-SUMO cloning vector. After cloning, the plasmid is then transformed into either Mach1-T1 competent cells for stable propagation and maintenance of the recombinant plasmid or BL21 (DE3) *E. coli* strain for expressing T7-regulated genes. Then, the overexpression of recombinant fusion protein, and purification of the His₆-SUMO tag protein can be carried out with the Ni-NTA column. The native protein is then generated by using SUMO protease which cleaves the N-terminal peptide containing the His₆-SUMO.

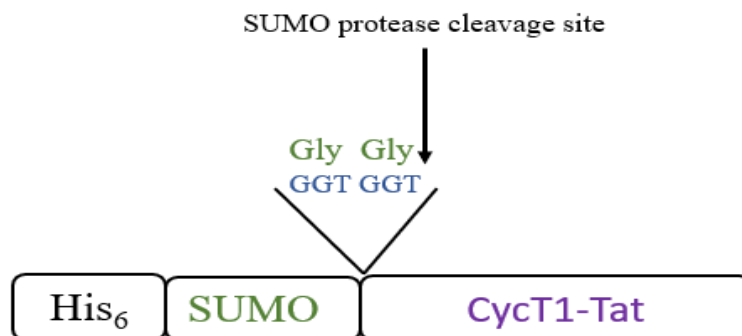


Figure 5.4: Constructions of His₆-SUMO-CycT1-Tat fusion protein.

The conserved Gly-Gly residues at the C-terminus of SUMO⁹³ can be cleaved by SUMO protease, resulting in the release of the CycT1-Tat protein. The last DNA base (3' T) of glycine serves as an overhang for a high-efficiency ligation of PCR product into the plasmids.

5.2.1 *Champion*TM pET SUMO Expression System

There are several advantages of the pET-SUMO cloning vector for high levels of expression of fusion proteins. SUMO protein is the *S. cerevisiae* (Smt3), a yeast protein that has a molecular weight of 11 kDa. SUMO is highly stable, a fast-folding protein that can be used as a fusion partner protein to increase the yield of overexpressed recombinant protein. In recent years, utilizing N-terminal SUMO fusion provides a useful biotechnological tool to enhance the solubility of expressed recombinant proteins.^{86, 94} A commercial vector, *Champion*TM pET-SUMO (figure 5.5), containing a His₆-SUMO tag becomes a promising expression system to efficiently express and enhance the solubility of the recombinant or native protein. His₆-tagged Smt3 fusion protein can be purified by Ni²⁺ resin, and His₆-Smt3 is removed by SUMO (His₆-Ulp1₄₀₃₋₆₂₁-His₆) protease on Ni²⁺ column to produce the desired product. Due to the dual His₆ tags on SUMO protease exhibiting high affinity for Ni²⁺ resin, purification and proteolytic cleavage reaction can be performed in a single Nickel-NTA column, if desired.

The construction of the *Champion*TM pET-SUMO vector shown in figure 5.5 is designed to facilitate the cloning of PCR products for expression in *E. coli*. It consists of a T7lac promoter, N-terminal His₆ tag, SUMO, TA® cloning site, and a kanamycin antibiotic resistance gene for selection in *E. coli*. A T7lac promoter is used for gene expression induced by IPTG. The His₆-SUMO tag is used for protein purification and enhancing solubility. *Taq*-amplified PCR products add 3'-A overhang, for efficient cloning into the pET-SUMO vector via ligation. The plasmid can be transformed into Mach1-T1 competent cells for propagating and maintaining the plasmid. For overexpression of T7-regulated genes B121 (DE3) *E. coli* cells are used.



Figure 5.5 Construction of Champion™ pET SUMO vector system.⁹⁵

Champion™ pET SUMO vector Thermo Fisher Scientific has His₆-SUMO fusion with 3' terminal thymidine on both ends where the gene of interest can be inserted for cloning.

Reprinted from Thermo Fisher Scientific.

5.2.2 Cloning of Optimized *CycT1-Tat* genes

Since the first amino acid residue of the *CycT1* fragment (Figure 5.1a) was a proline, which may reduce the cleavage efficiency of SUMO protease, it was replaced with a serine. The genes encoding F1 and F3 sequences (see supplemental data) were purchased from Integrated DNA Technologies (IDT). The forward (FP) and reverse (RP) primers associated with these genes were also purchased from IDT, which are

FP1: 5'-AGC AAC CGT TGA AGC GCA TCT-3'

FP2: 5'-AGC TGG CGT GCC TGC GAA-3'

RP1: 5'-TTA TTA CTC GGT AGG GCC TGT AGG-3'

RP2: 5'-CTA CTA GTC GAA AGG GTC AGT TTC-3'

For PCR amplification of the F1 gene, FP1 and RP1 were used. For PCR amplification of the F2 gene, FP2 and RP1 were used.

For PCR amplification of the F3 gene, FP1 and RP2 were used. For PCR amplification of the F4 gene, FP2 and RP2 were used.

The genes were first amplified by PCR, which was performed as follows: the reaction mixture was denatured at 94°C for 3 min. This was followed by 30 consecutive cycles of denaturation: 95°C for 30 seconds, annealing at 52°C for 1 minute and extension at 72°C for 1 minute, and then a final extension at 72°C for 7 min. The PCR product was purified with a PCR clean-up kit (Qiagen) before cloning into the pET-SUMO vector. Since *Taq* polymerase provides a 3'-dA overhang during the PCR reaction, the PCR product was directly cloned into the pET SUMO vector (Invitrogen) that has a single 3'-dT residue using ligation at the insertion site without the need for restriction digestion.

The plasmids were then transformed into Mach1-T1 competent cells, and the transformations were spread on prewarmed LB/kanamycin plates. The plates were incubated overnight at 37°C without shaking. About seven to ten colonies were picked and inoculated in 5mL fresh Luria-Bertani (LB) medium and incubated at 37°C with shaking at 200 rpm overnight. The plasmid DNA was isolated using the QIAprep Spin Miniprep kit (Qiagen) and then the concentrations of prepared plasmid DNA were obtained from the absorbance at 260nm on a Nanodrop spectrophotometer.

The sequences of recombinant plasmids were obtained by automated DNA sequencing (SUNY Upstate University), and the 4peaks software was used to analyze the purity and accuracy of the sequences. Due to the deletion of the amino acids (249-257) at the N-terminal of the CycT1 segment for the F2 and F4 chimera, two more rounds of cloning and isolation of plasmid DNA were done to get the correct sequences.

After the DNA sequences were confirmed, the plasmid DNA with the correct clone was transformed into BI21 (DE3) *E. coli* cells and the transformations were spread on prewarmed LB/kanamycin plates. The plates were incubated at 37°C overnight without shaking. A single colony was picked from each plate and inoculated with a fresh 5mL LB medium with 50 µg/mL kanamycin. The samples were incubated at 37°C with shaking at 200 RPM overnight. A fresh 5 mL of LB containing 50 µg/mL kanamycin was inoculated with 250 µL of the overnight culture and grown at 37°C with shaking until the culture reached the stationary phase. 850 µL of culture was then mixed with 150 µL of sterile glycerol and stored at -80°C.

5.2.3. Pilot Fusion Protein Expression

To confirm overexpression, a pilot experiment was performed by transferring 250 μL of the overnight culture into a new 5 mL of LB containing 50 $\mu\text{g}/\text{mL}$ Kanamycin. The samples were grown for 2 h at 37°C with shaking ($\text{OD}_{600} \sim 0.6$). 100 μL aliquots from each culture were removed to use as an uninduced sample, and a final concentration of 1mM IPTG was added to the samples and incubated at 37°C with shaking for another 4 h. 100 μL aliquot of induced samples was transferred into snap-top tubes. All 8 samples were mixed with 1X SDS- sample buffer in a 1:1 ratio and then incubated for 5 min at 94°C. The samples were centrifuged and 20 μL of each sample was loaded on a 15% SDS-PAGE gel and electrophoresed.

The results in Figure 5.6 showed that the target proteins are more than 30% of the total cellular protein. Expression levels for F1 (lane 2) and F2 (lane 4) appear larger than for F3 and F4. As the CycT1 fragment of F1 includes R251 and R254 and has the shorter Tat(1-86), and is robustly expressed, we focused our efforts on F1.

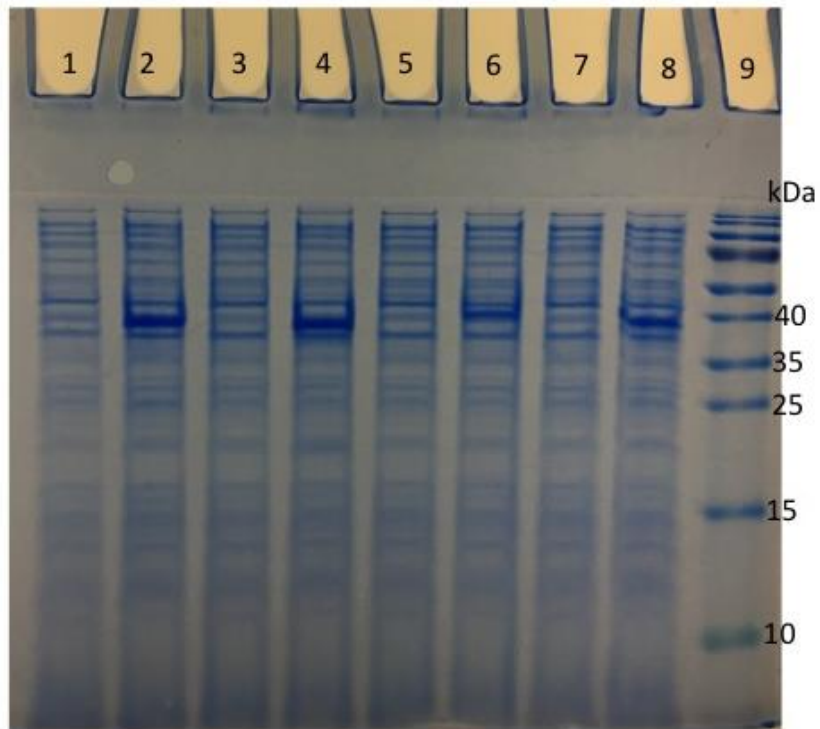


Figure 5.6: 15% SDS-PAGE of optimized HIV-1 CycT1-Tat fusion proteins.

Lanes 1, 3, 5 and 7: uninduced sample of F1, F2, F3 and F4, respectively; Lane 2, 4, 6, 8: induced sample of corresponding protein; Lane 9: Protein molecular weight ladder. Stacking gel: 5%. Separating gel: 15%. Run condition: 120V for 1.5 h.

5.2.4. Overexpression and purification of Optimized CycT1-Tat chimera

The protocol of CycT1-Tat (F1) protein overexpression and purification was established using His₆-MBP-CycT1-Tat protein expression⁴⁶, and instructions from the pET SUMO protein expression system protocols. Firstly, a 5mL seed culture was grown overnight at 37°C, shaking at 200 RPM. The recombinants were inoculated in fresh LB medium and incubated at 37°C until OD₆₀₀ was ~0.6 (mid-log). IPTG was added to a final concentration of 1mM to induce the expression of the cloned gene. The cells were harvested by centrifugation at 4,000 g at 4°C and lysed by sonication. The sample was centrifuged at 17,000 g at 4°C to obtain a supernatant containing soluble SUMO-F1 protein and pellet.

The purification of His₆-SUMO-F1 was carried out at a native condition with a Ni-NTA affinity column. Proteins without His₆ tag were first removed using (10- 25 mM) imidazole; Phosphate buffered saline (PBS) containing 250mM imidazole was then used to elute His₆-SUMO- F1. Since six imidazole rings are part of the structure of the His₆, the tag binds to the nickel ions immobilized by the NTA groups on the matrix. Imidazole can also bind to the nickel ions and compete with the histidine residues in the non-tagged proteins. Therefore, a low concentration of imidazole was included in the lysis and wash buffer to prevent the binding of background contaminants while His₆-SUMO- F1 can strongly bind to the Ni-NTA matrix.

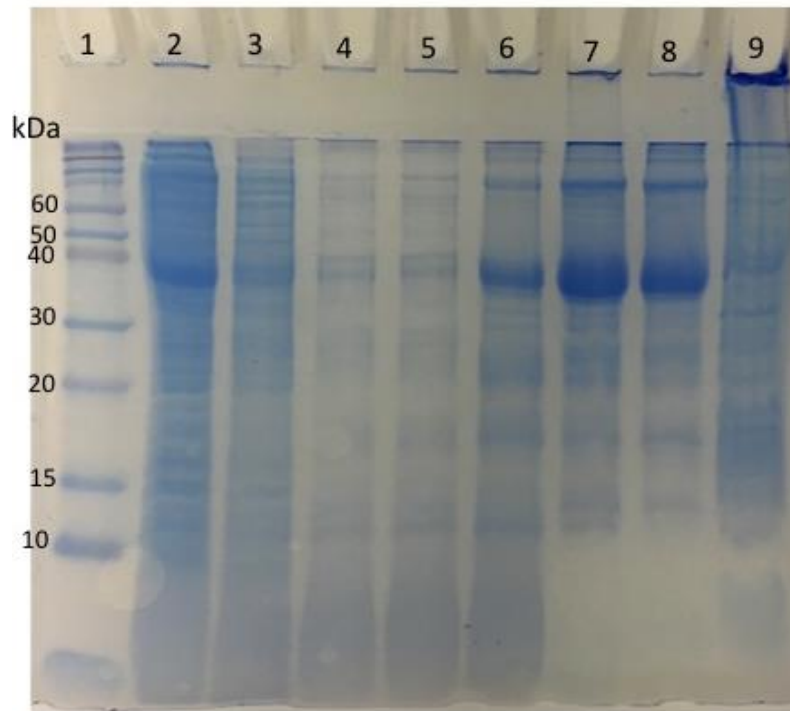


Figure 5.7: 15% SDS-PAGE gel of purification of His₆-SUMO-CycT1-Tat protein.

Lane 1: Protein molecular weight ladder; lane 2: flow-through after binding to Ni-NTA resin; lanes 3,4,5: fractions from wash buffer; lanes 6,7,8: purified SUMO-CycT1-Tat eluted from Ni-NTA column; lane 9: fraction from lysate pellet.

The SDS-PAGE analysis of purification of His₆-SUMO- F1 is shown in Figure 5.7. The results showed that the molecular weight is ~25kDa, which corresponds to the predicted size of the His₆-SUMO-CycT1-25aa-Tat (F1) protein. The supernatant containing soluble His₆-SUMO-F1 was mixed with an equal volume of equilibrium buffer and Ni-NTA resin on an end-over-end rotator for 30 min. The mixture was then loaded onto a column and the flow through was collected. Since there was a visible band associated with the protein of interest in the flow-through shown in lane 2, the mixing time should be increased or the flow-through should be reloaded onto the column in future purifications. The band intensity of the target His₆-SUMO-F1 protein is more than 70% in lanes 6,7 and 8. Since there was no appreciable intensity of this protein in lane 9, His₆-SUMO-F1 is not present in the pellet.

The SUMO expression tag is a reliable system to increase expression efficiency and enhance the solubility of recombinant proteins in *E. coli*. While one-step purification to isolate His₆-SUMO-F1, shows that it is about 70% of protein after the Ni-NTA treatment, a two-step purification should be helpful to improve the purity of His₆-SUMO- F1 in future work.

5.2.5 Cleavage of SUMO and purification of CycT1-Tat (F1) Protein

The conserved Gly-Gly residues at the C-terminus of SUMO⁹³ in the SUMO-CycT1-Tat protein, can be cleaved by SUMO protease (~27 kDa), releasing the desired CycT1(249-280)-25aa-Tat(1-86) protein F1. Since SUMO protease (Invitrogen) is not active above 300mM NaCl and 150mM imidazole, the sample was dialyzed into 50mM Tris, 150mM NaCl, 1mM DTT; pH 8.0 overnight at 4°C. A time course experiment with 10 units of SUMO protease cleavage reaction was performed by incubating the samples at 4°C for reaction time optimization. The SDS-PAGE analysis of the cleaved and uncleaved protein remaining after digestion is shown in Figure 5.8. After incubation overnight at 4°C shown in lane 4, most of F1 was released by SUMO protease. Overnight cleavage appears to be sufficient. The addition of 10 units of SUMO protease and incubation at 3 more hours at 4°C (lane 5) is not necessary. Since His₆-SUMO-CycT1-Tat protein and SUMO protease have similar molecular weights, the band between 35-40 kDa in Lanes 1-5 is comprised of these molecules.

After optimizing the reaction time for SUMO protease cleavage, the purified His₆-SUMO-F1 protein sample was treated with SUMO protease by incubating overnight at 4°C. Both His₆-SUMO and SUMO protease will bind to a Ni-NTA column, but the target F1 protein should elute. From the SDS-PAGE analysis of purification of CycT1-Tat (F1) protein (~16kDa) in Figure 5.9, there was also a band of ~50kDa in the fraction of purified protein in lanes 2,3 and 4, which could be due to an aggregation of F1. Some of the cleaved His₆-SUMO co-eluted as shown in Lanes 2, 3, and 4, which could be due to the short equilibration time with Ni-NTA resin. The results showed that CycT1-Tat (F1) was isolated as the major band in fractions 2, 3, and 4; these were pooled for characterization by mass spectrometry and studies of TAR binding to F1.

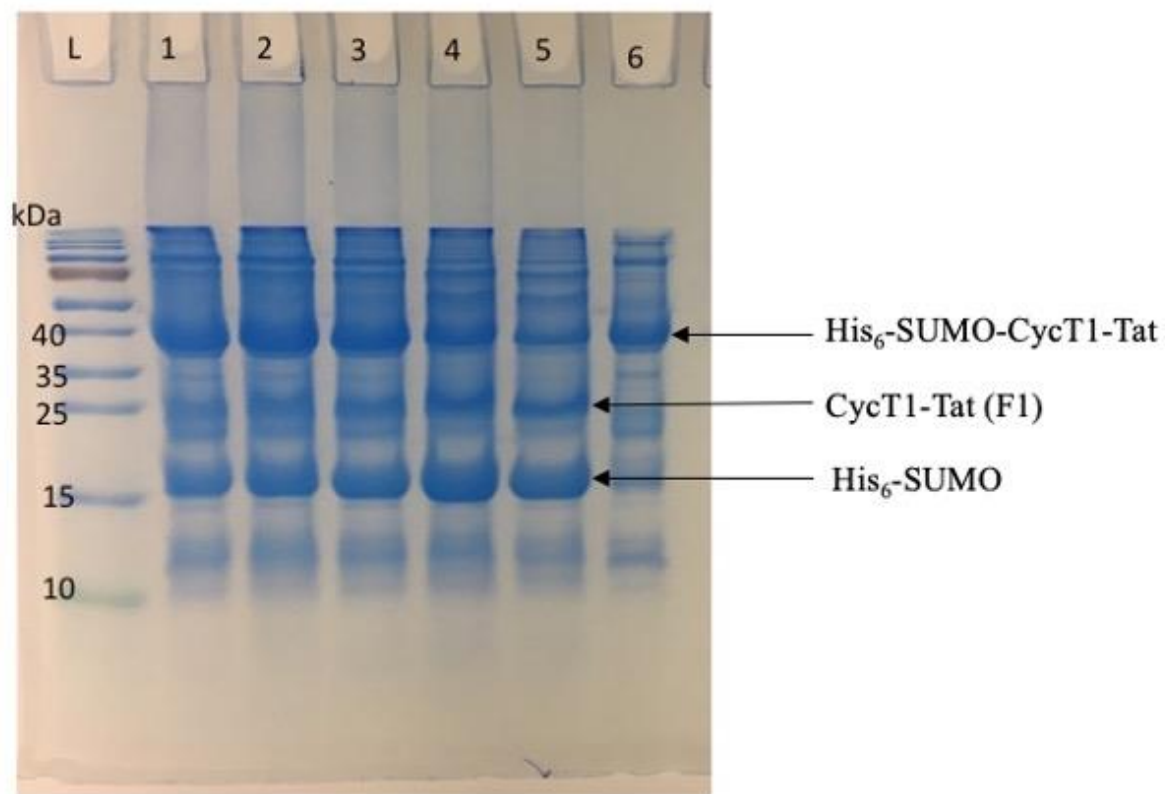


Figure 5.8: 15% SDS-PAGE of SUMO protease cleavage applied to His₆-SUMO-F1.

L=Protein molecular weight ladder. Lane 1: after 3 hours cleavage; Lane 2: after 5 hours cleavage; Lane 3: after 7 hours cleavage; Lane 4: after overnight cleavage; Lane 5: 3 hours after additional 10 units of SUMO protease used on the sample from overnight cleavage; Lane 6: Control (before SUMO protease treatment).

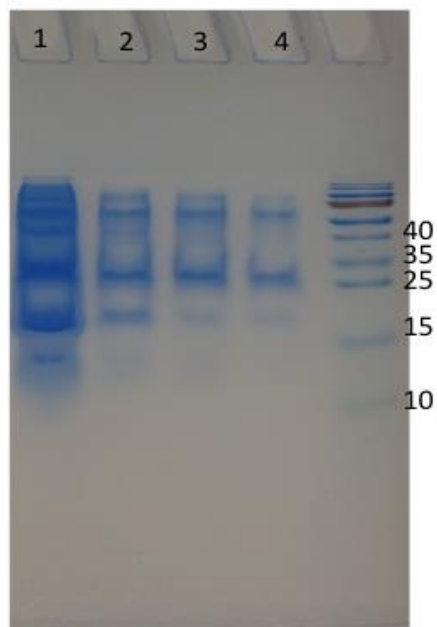


Figure 5.9: 15% SDS-PAGE analysis of purification of CycT1-Tat (F1) protein.

Lane 1: the fraction before loading to Ni-NTA column; Lanes 2, 3, and 4: purified F1 protein.

Molecular weights of expected components are His₆-SUMO- F1: 28 kDa, F1: 16.4 kDa, His₆-SUMO: 11.6 kDa. SUMO protease: 27 kDa may present only in a (very small) catalytic amount.

5.3. Analysis of CycT1-Tat fusion protein

Two methods, SDS-PAGE and mass spectrometry were used to analyze the overexpressed CycT1-Tat (F1) protein. 15% SDS-PAGE assay was used to check the overexpression and purity of the recombinant protein (Figures 5.6-5.9) while mass spectroscopy (MS) was used to confirm the molecular weight. Matrix-assisted laser desorption/ ionization in combination with a time of flight (MALDI-TOF) was used to analyze the F1 protein preparation. The protein matrix is exposed to ultraviolet light to generate charged ions of various sizes and the ratio of their mass to charge (m/z) is measured. According to the m/z value of various ions, the times of ion flight to reach the detector are different. The MALDI-TOF mass spectrum of the F1 protein preparation is shown in Figure 5.10. The spectrum of the target protein was acquired in reflection mode and recorded between 3000 and 25000 m/z range.

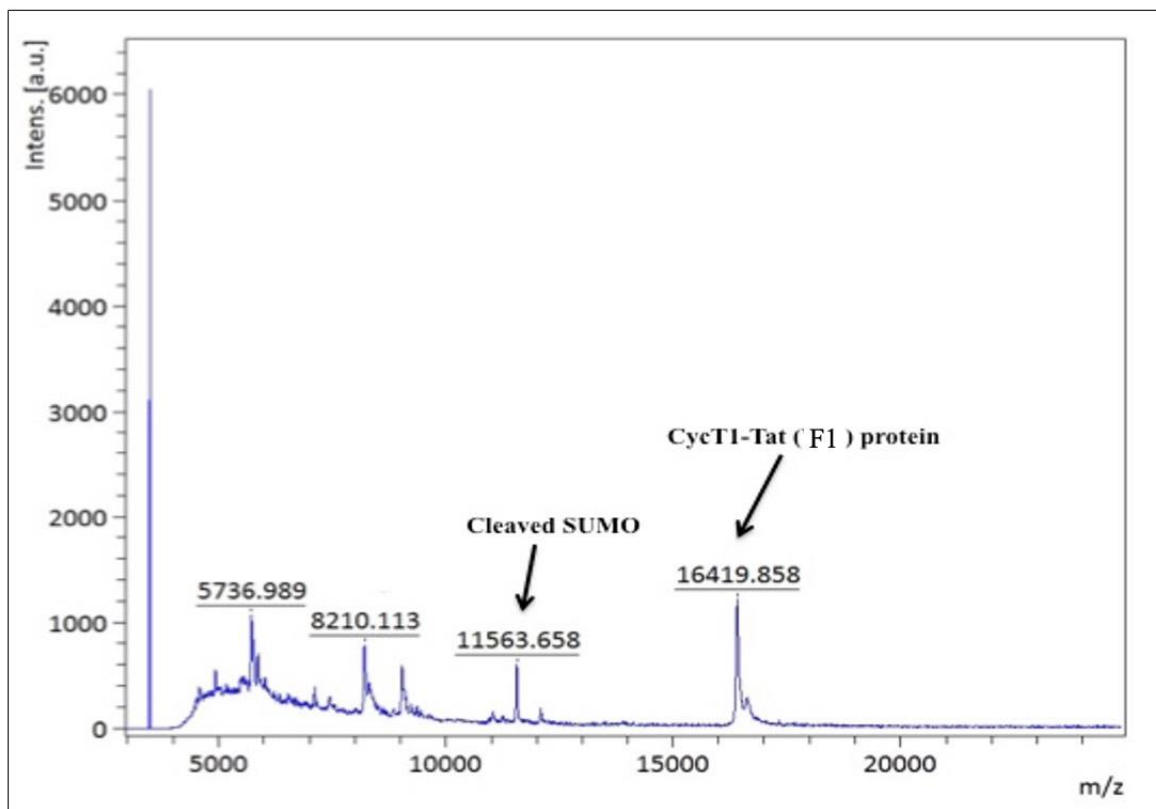


Figure 5.10: MALDI-TOF mass spectrum of CycT1-Tat protein.

The molecular weight of the target protein determined by three trials is approximately the same, with the theoretically calculated value of 16416.47 Da.

The MS results have a reasonable agreement with the SDS-PAGE analysis, with some cleaved SUMO remaining in the sample at $m/z = 11,564$ ($z=1$). Analysis of peak positions suggests that the peak at $m/z = 16,420$ ($z = 1$) is the parent molecular ion with the expected mass of the F1 protein, and the peak at $m/z = 8,210$ ($z = 2$) has the same mass. The expected mass for F1 at $z=3$ is $m/z = 5,473$; a peak at $m/z = 5,737$ misses the expectation by -264 mass units, so it could represent an impurity from partial degradation of the sample. The expected mass for His₆-SUMO at $z = 2$, is $m/z = 5,782$ in this same region with multiple peaks near the lower mass limit of the run. There is a likely impurity peak near $m/z = 9,100$, and other small peaks that could be partial degradation products of F1. Overall, possible degradation products of F1 are not present in large amounts, suggesting that F1 is not highly sensitive to decomposition in His₆-based separations, electrophoresis, dialysis, and common laboratory procedures in protein purification.

Analysis of peak intensities in Figure 5.10 can give estimates of the purity of the F1 protein. Assuming that only the $z=1$ and $z=2$ peaks for F1 are the full-size protein, about 40% of the mass in the sample was F1; subtracting the mass for His₆-SUMO ($z=1$) gives 45-50% mass purity for F1 in the remainder of other proteins and fragments after a single Ni-NTA treatment. If the possible $z=3$ peak for F1 is in fact F1, the purity increases to around 60%. Given the distribution of masses in the spectrum, it should be simple to purify to 90% or more using a gel filtration column. The range of masses acquired in the spectrum was 3,000 to 17,000 m/z , so a possible F1-dimer (top band in Figure 5.8) lies outside the spectral window.

Assuming that a dimer is not active in binding TAR-31, a conservative estimate for the mass of active F1 is 20-30% in the pooled sample from Figure 5.9. Due to time pressures associated with closing the lab after Prof. Borer's retirement, this sample was used without further purification for a binding study presented next.

5.3.1. Electrophoretic Mobility Shift Assay (EMSA)

Electrophoretic mobility shift assays use a native gel shift analysis (no denaturant) for semi-quantitative characterizations of complexes between nucleic acids and proteins. The negative pole is at the top of the gel, forcing oligonucleotides to migrate through the gel; usually, unbound protein stays in the well and does not enter the gel, especially for large proteins. Protein-RNA complexes will have a larger molecular weight than the unbound RNA, which results in band shifting. The nature and degree of shifting can be visualized by staining the RNA with an SYBR gold stain (Molecular Probes). The binding affinity of the complex can be characterized to an extent by comparing the relative intensities of the shifted and unshifted bands on the native PAGE gel.

For EMSA of the TAR-F1 complex, three different TAR RNA analogs were used (Figure 5.11). TAR-31 has all of the known binding features needed to interact with the CycT1:Tat heterodimer (see discussion of Figure 1.5. TAR-B25 mimics the absence of the hairpin loop sequence that is thought to interact primarily with CycT1. TAR-H24 mimics the absence of the U-Rich bulge region where the interaction F1 occurs with Tat.

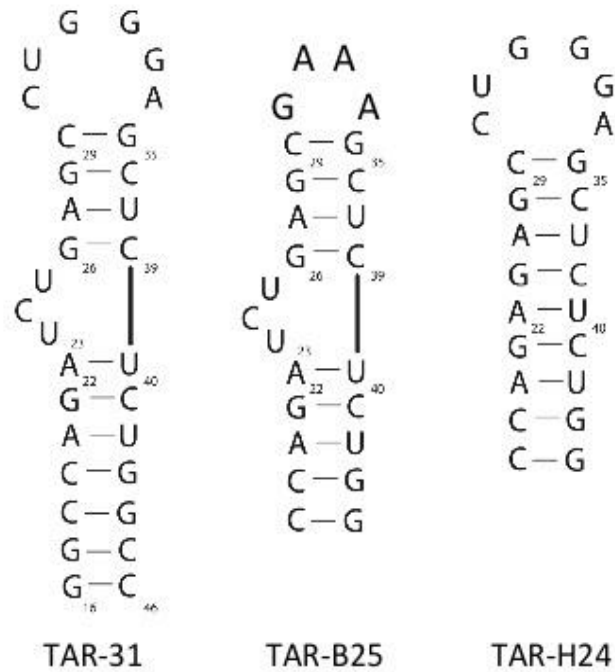


Figure 5.11: Various secondary structures of TAR-RNA.

TAR-31 contains the essential CycT1-Tat binding domains, TAR-B25 is a variant without the required bases in the hairpin loop, and TAR-H27 deletes the bulge loop.

The EMSA analysis of purified CycT1-Tat (F1) protein with TAR RNAs is shown in Figure 5.12. A TAR31-D1 complex forms, with the shifted band (lane 2) increasing in intensity at doubled [F1] concentration (lane 3). A similar band shift is observed with TAR-B25 at the doubled [F1] (lane 5), but the band intensity is weaker than in lane 3. Thus, TAR-B25 binds F1, but not as strongly as TAR31. This is backed up by the competition between these two RNAs for F1 in lane 6, which shows an increased intensity for the shifted band upon adding TAR31. By contrast, lanes 7 and 8 show that there is little affinity of Tar-H24 for F1. These results suggest that the TAR hairpin and bulge both contribute to the affinity of the complex with CycT1-Tat. There is some staining at the very top of the gel, especially the lanes that contain TAR-31, suggesting that an F1-dimer exists and that it has at least some affinity for TAR-31.

It is exciting that the EMSA of TAR-31 with F1 is consistent with our expectation that both the bulge and hairpin sequences of TAR-31 are required for high affinity. The key result from the work described in this chapter is that F1: CycT1(249-280)-25aa-Tat(1-86), contains the essential binding elements in a stable protein that is highly soluble in an aqueous solution. The 10 kDa TAR-31 complexed with our 16.4 kDa chimeric protein could be the long-sought model for TAR interacting with the P-TEFb initiation complex and used for 3D structure determination and drug discovery.

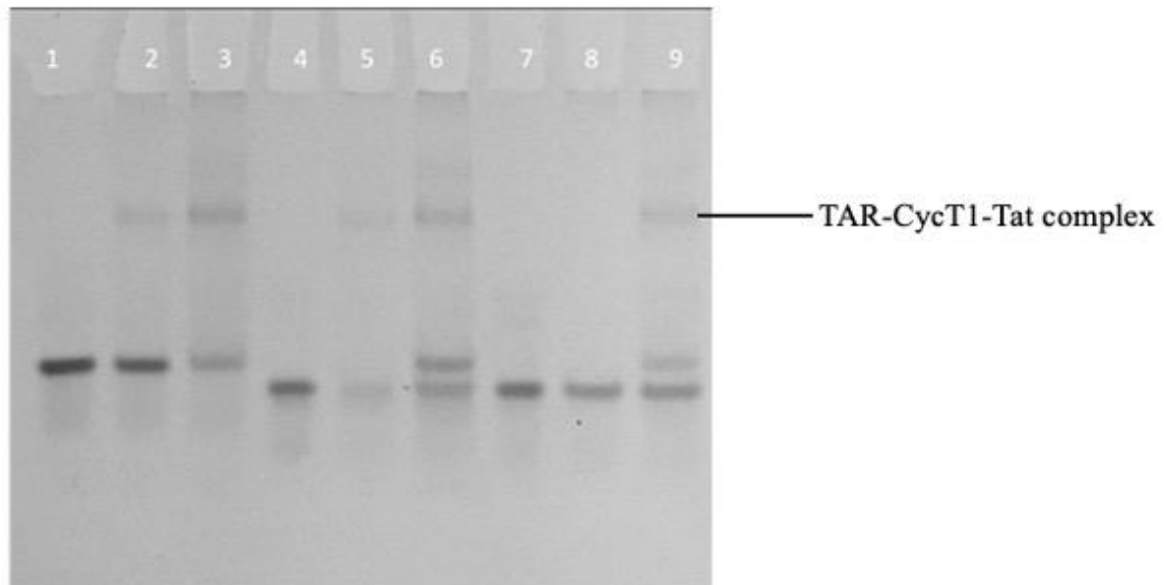


Figure 5.12: Binding of recombinant CycT1-Tat (F1) protein to various HIV-1 TAR RNAs.

Lanes 1, 4, and 7: 1 μ M TAR-31, TAR-B25, and TAR-H24, respectively, without protein.

Lanes 2, 3: TAR-31 at 10 μ M, 20 μ M with CycT1-Tat.

Lane 5: TAR-B25 with 20 μ M CycT1-Tat; lane 6: TAR-31 and TAR-B25 with 20 μ M CycT1-Tat.

Lane 8: TAR-H24 with 20 μ M CycT1-Tat; and lane 9: TAR-31 and TAR-H24 with 20 μ M CycT1-Tat.

Conclusions and Prospects for Future Work

This work (Chapter 5) conducted the design, cloning, and overexpression necessary to produce a working model system for HTS of drugs to disrupt the TAR:Tat complex, which is a key target in the HIV-1 infection cycle. The fusion protein, CycT1-Tat, was redesigned to include the key R251 and R254 residues for stabilization of the CycT1-Tat-TAR RNA complex and to simplify the overexpression in *E. coli*. Four variant sequences, F1-F4 (Figure 5.3) were expressed in *E. coli*. F1 and F2 were shown to be expressed more efficiently than F3 and F4, and only F1 includes the full N-terminal CycT1(249-280) sequence. Therefore, overexpression of F1 was performed for further work (section 5.2).

After cleavage of the SUMO tag, the molecular weight of partially purified CycT1-Tat (F1) was confirmed with mass spectrometry (MALDI-TOF). Then an initial characterization using EMSA to study the TAR-CycT1-Tat complex was performed. While EMSA can give a semi-quantitative picture of binding to these TAR analogs, even with partially purified proteins, a quantitative determination of K_d requires highly purified proteins using a technique like the Octet biolayer interferometry (BLI) assay.

To achieve CycT1-Tat fusion protein with high purity, a double purification column of Ni-NTA resin and/or a longer equilibration with Ni-NTA resin are proposed. The CycT1-Tat (F1) could have been degraded or aggregated before MALDI-TOF analysis. Therefore, overexpression, purification, and mass-spec, and binding analysis should be planned in rapid succession. F2 was shown to overexpress well in *E. coli* (figure 5.6). Since F2 lacks CycT1 (249-256), it would be interesting to compare the binding properties of F1 and F2.

This project is a proof-of-concept that a commercial SUMO protein expression system can be used to express a soluble CyT1-Tat fusion protein (F1) for further studies. Drug

candidates that bind to CycT1-Tat and disrupt the binding of TAR could be screened with a 3-segment switch containing TAR-31 (or an aptamer that mimics TAR) as the Probe segment. Such a switch could be substantially larger than the NCp7-responsive switches described earlier in this dissertation. Building larger switches on a commercial scale would also benefit from ligation strategies developed here to assemble the active segments. Finally, it is necessary to detect and quantify the release of the switch from the target-switch state caused by target-drug binding; this requires rapid switching between OFF and ON states of the switch. Large bistable nucleic acids are likely to fall into kinetic traps (stable intermediate folded states) that may slow conformational switching beyond usability. By design, 3-segment switches use branch-migration to move conformational change rapidly through kinetic traps.

Supplemental Information

F1 Gene (435 bp):

5'-

AGCAACCGTTTGAAGCGCATCTGGAATTGGCGTGCCTGCGAAGCAGCCAAGAAG
ACTAAAGCCGACGACCGTGGTACGGACGAGAAAACCTTCGGAACAGACAATGCCG
GAGCAAAAGCTGATTAGTGAAGAGGACCTTGCCATGGAATTCCTTGAAATTGAC
CCCGTCGACATGGAACCTGTCGATCCAAATTTAGAGCCGTGGAAGCATCCAGGC
AGCCAGCCGCGTACTGCTTGTAATAACTGTTATTGCAAGAAGTGTTGCTTCCATT
GTTACGCATGTTTTACTCGTAAAGGACTGGGAATCAGCTATGGTCGTAAAAAGCG
TCGCCAACGTCGCCGTGCCCCCAAGATTCCCAAACACACCAAGCGTCTCTTTCT
AAGCAGCCTGCTTCGCAGAGTCGTGGTGACCCTACAGGCCCTACCGAGTAATAA-

3'

F3 Gene (480 bp):

5'-

AGCAACCGTTTGAAGCGCATCTGGAATTGGCGTGCCTGCGAAGCAGCCAAGAAG
ACTAAAGCCGACGACCGTGGTACGGACGAGAAAACCTTCGGAACAGACAATGCCG
GAGCAAAAGCTGATTAGTGAAGAGGACCTTGCCATGGAATTCCTTGAAATTGAC
CCCGTCGACATGGAACCTGTCGATCCAAATTTAGAGCCGTGGAAGCATCCAGGC
AGCCAGCCGCGTACTGCTTGTAATAACTGTTATTGCAAGAAGTGTTGCTTCCATT
GTTACGCATGTTTTACTCGTAAAGGACTGGGAATCAGCTATGGTCGTAAAAAGCG
TCGCCAACGTCGCCGTGCCCCCAAGATTCCCAAACACACCAAGCGTCTCTTTCT
AAGCAGCCTGCTTCGCAGAGTCGTGGTGACCCTACAGGCCCTACCGAGAGTAAA
AAAAAAGTTGAACGTGAAACTGAAACTGACCCTTTCGACTAGTAG-3'

CycT1-Tat Fusion proteins:

This sequence has 143 aa (**F1**), Theoretical pI/Mw: 8.94 / 16416.47

CycT1 (249-280), a 25 aa spacer with a myc antibody tag, and Tat (1-86)

SN RLKRIWNWRA CEAAKKTKAD DRGTDEKTSE QTMPEQKLIS EEDLAMEFLE
IDPVD MEPVDPNLEP WKHPGSQPRT ACNNCYCKKC CFHCYACFTR
KGLGISYGRK KRRQRRRAPQ DSQTHQASLS KQPASQSRGD PTGPTE**

This sequence has 135 aa (**F2**), Theoretical pI/Mw: 8.42 / 15335.18

CycT1 (256-280), a 25 aa spacer with a myc antibody tag, and Tat (1-86)

SWRA CEAAKKTKAD DRGTDEKTSE QTMPEQKLIS EEDLAMEFLE IDPVD
MEPVDPNLEP WKHPGSQPRT ACNNCYCKKC CFHCYACFTR KGLGISYGRK
KRRQRRRAPQ DSQTHQASLS KQPASQSRGD PTGPTE**

This sequence has 158 aa (**F3**), Theoretical pI/Mw: 8.77 / 18207.42

CycT1 (249-280), a 25 aa spacer with a myc antibody tag, and Tat (1-100)

SN RLKRIWNWRA CEAARKTKAD DRGTDEKTSE QTMPEQKLIS EEDLAMEFLE
IDPVD MEPVDPNLEP WKHPGSQPRT ACNNCYCKKC CFHCYACFTR
KGLGISYGRK KRRQRRRAPQ DSQTHQASLS KQPASQSRGD PTGPTEKSKK
VERETETDPF D**

This sequence has 150 aa (**F4**), Theoretical pI/Mw: 8.09 / 17126.13

CycT1 (256-280), a 25 aa spacer with a myc antibody tag, and Tat (1-100)

SWRA CEAARKTKAD DRGTDEKTSE QTMPEQKLIS EEDLAMEFLE IDPVD
MEPVDPNLEP WKHPGSQPRT ACNNCYCKKC CFHCYACFTR KGLGISYGRK
KRRQRRRAPQ DSQTHQASLS KQPASQSRGD PTGPTEKSKK VERETETDPF D**

Note: * is a stop codon

Materials and Methods

5.2. Cloning, Overexpression, and Purification of His₆-SUMO tagged hCycT1-Tat proteins

The workflow for producing pET-SUMO-CycT1-Tat plasmid is to amplify the gene of interest, clone the gene sequence into an expression vector (pET-SUMO), transform it into *E. coli* cell (Mach1-T1 competent cells), grow on LB/kanamycin plate, select the colonies and inoculate in a fresh LB medium containing Kanamycin. The plasmid DNA was purified and analyzed with DNA sequencing. Cloning procedure of CycT1-Tat genes is performed according to the manufacturer instructions.

Genes of interest with associated forward and reverse primers are ordered from IDT (Integrated DNA Technologies). Polymerase chain reaction (PCR) of optimized CycT1-Tat gene was performed using PCR thermal cycler as follows: reaction mixtures containing a gene of interest, forward primer, reverse primer, dNTP, High fidelity Taq polymerase, 10X PCR buffer were incubated at 94°C for 3 min and denature at 94°C for 30 seconds. The annealing step was conducted at 52°C for 1 minute, and the extension was performed at 72°C for 1 minute. 30 cycles of denaturation, annealing, and extension were carried out, and then a final single extension at 72°C for 7 min to get amplified gene fragments.

Using the pET-SUMO expression vector system (Invitrogen), the cloning of the PCR product with the pET-SUMO vector was done by the ligation reaction. The recombinant plasmid was then transformed into a One Shot Mach1-T1 competent cell by incubating it on ice for 30 min. The cells were heat-shock for 30 seconds at a 42°C water bath without shaking and then immediately transferred the tubes to ice for 5 min. 250ul of room temperature S.O.C medium, a super optimal broth with Catabolite repression, was added to the tubes containing cells transformed with ligation reaction. S.O.C medium provides maximal transformation efficiency of *E. coli* cells. The tube was incubated for 1 hour at 37°C with shaking (200 rpm).

100 and 150 μ L from the transformation were plated onto prewarmed LB/Kanamycin plates to ensure well-spread colonies. The plates were incubated at 37°C overnight.

The transformations were analyzed by picking colonies and culturing them in fresh 5mL LB media containing 50 μ g/mL kanamycin at 37°C overnight, shaking at 200 rpm. The overnight culture was harvested by centrifugation in an Eppendorf centrifuge (Eppendorf, Hauppauge, NY) for 5 min at 3000 rpm. A QIAprep Spin Miniprep kit (Qiagen, Valencia, CA) was used to purify the plasmid DNA from the harvested cells by eluting them with 50ul Dnase-free water. The plasmids were stored at -20°C for long-term storage.

Sanger Sequencing

The prepared plasmids with different gene sequences were sent to the Molecular Core facility for DNA sequencing and Flow Cytometry at the State University of New York Upstate Medical University by using the universal T7 promoter primer as a sequencing primer.

Expression and Purification of SUMO-CycT1-Tat protein

Day 1: Seed culture

1. A 5mL seed culture is grown overnight @ 37°C, shaking @ 200~250rpm.

5mL seed culture

4.5mL LB media

500µl 10x M9 salts

5µl 50 mg/mL Kanamycin (final 50 µg/mL)

2.5µl 200 mM ZnCl₂ (final 100 µM)

5µl 1 M MgSO₄ (final 1 mM)

50µl 40% Glucose (final 0.4%)

Day 2: Large culture, induction, and cell harvest (~9hrs)

2. The following morning transfers the 5mL seed culture into 1L fresh media:

1L culture

900 mL LB media

100 mL 10x M9 salts

1 mL 50 mg/mL Kanamycin (final 50 µg/mL)

500 µl 200 mM ZnCl₂ (final 100 µM)

1 mL 1 M MgSO₄ (final 1 mM)

10 mL 40% Glucose (final 0.4%)

3. The culture is incubated @ 37°C, shaking @ 200~250 rpm till it reaches the desired cell density of OD₆₀₀ ~0.6. This step usually takes 3.5-4 h.

Note: You might need to check the OD every half hour to monitor the cell growth. If by any possibility the OD₆₀₀ gets above 1, I suggest you not to go on with the induction but start a new cell culture.

4. Induce the protein expression by adding 1mL 1M IPTG (final 1mM) into each 1000 mL culture at OD₆₀₀ ~0.6. Incubate @ 37°C, 200~250 rpm for another ~4 h.

After induction the cells are transferred to Thermo centrifuge bottles and pelleted on Sorvall Legend XTR centrifuge (Thermo Scientific), spinning at 4000 g for 15 min.

5. After spinning, the cleared LB media (supernatant) is disposed and the pellet is stored in the -20°C freezer overnight.

Day 3: Cell Lysis and Ni-NTA purification

Part 1. Cell Lysis:

6. The frozen pellets are placed on ice and resuspended in the following freshly made 15 mL Lysis Buffer containing

Lysis Buffer (pH: 7.8)

50 mM Tris-Base

400 mM NaCl

100 mM KCl

10% Glycerol

0.5% Triton X-100

10 mM Imidazole

100 μ L 10 mg/mL freshly made Lysozyme (final 30 μ g/mL)

7. After resuspending the cells, they are transferred to a 50mL falcon tube and left on ice for 20 min, 2 mL of a detergent mix (B-per (Pierce Biotechnology, www.piercenet.com)) is then added to enhance protein solubility. The lysate is then left on ice for an ADDITIONAL 10 min:
8. Using the Microson Ultrasonic Cell Disruptor, the lysate is pulsed 4x for 25 sec at an output power setting of 20~25 watts. Be sure to wait 1 minute on ice between each 25-sec pulse. Clean the sonicator tip before and after using the sonicator with ethanol.

The lysate is then transferred to 50 mL Oakridge tubes and is again pelleted on the Sorvall Legend XTR centrifuge (Thermo Scientific), spinning at 17000 g for 30 min.

9. After spinning, the cleared lysate is transferred to 50mL Falcon tubes and kept on ice before being loaded onto the Ni-NTA column.

Part 2: Ni-NTA resin Purification:

For native conditions, the following buffers are prepared:

Equilibrium Buffer: 20 mM sodium phosphate, 300 mM sodium chloride with 10 mM imidazole; pH 7.4

Wash Buffer: 20 mM sodium phosphate, 300 mM sodium chloride with 25 mM imidazole; pH 7.4

Elution Buffer: 20 mM sodium phosphate, 300 mM sodium chloride with 250 mM imidazole; pH 7.4

Preparation of Ni-NTA column

10. Add 8 mL of Ni-NTA resin to a tube and centrifuge for 2 min at 700 xg. Discard the supernatant.
 11. Add 2 resin-bed volumes of Equilibrium Buffer to resuspend the resin. Centrifuge 2 min at 700 x g. And discard the supernatant.
- Mix the protein extract with an equal volume of Equilibrium Buffer and place on an end-over-end rotator for 30 min.
 - Add the mixture onto a Gravity-flow Column and collect the flow-through in a 50 mL falcon tube. Re-apply the flow through to the column to maximize binding.
 - Wash the resin with two resin-bed volumes of Wash Buffer and collect the flow-through in a 15mL falcon tube.

-Elute His-tagged proteins from the resin with 20mL Elution buffer by collecting 1.0mL fraction.

Ni-NTA resin Regeneration

1. Wash resin with 10 resin-bed volumes of MES buffer.
2. Wash resin with 10 resin-bed volumes of dH₂O.
3. Store resin as 50% slurry in 20% ethanol.

Day 4: Gel analysis

A. SDS-PAGE analysis:

1. Pick samples for SDS-PAGE analysis.
2. Make a 15 % acrylamide/SDS Tris-Tricine gel and a 5% stacking gel on top of it. Take 10µl from each sample and combine it with 10µl of 2x SDS loading buffer. Samples are boiled and denatured for ~3 min before loading. Run the gel at 200 Volt until the smallest kDa of the protein ladder reaches the other end of the gel.
3. Stain the bottom gel with 50 mL Gel-Code Blue stain overnight
4. Destain the gel with 100 mL dH₂O, rinse the gel with water, and take a photo of the gel for the record.

Day 5: Buffer Exchange

-Combine the fractions containing the SUMO-CycT1-Tat protein band (~35 kDa) and place them in a 10000 MWCO dialyzer (Slide-A-Lyzer® Dialyzer Cassette) in a buffer (50 mM Tris, 150 mM NaCl, 1 mM DTT; pH 8.0) overnight at +4°C.

Day 6 & 7: Purification of Cyc-T1-Tat Protein

SUMO cleavage reaction

1. Prepare the cleavage reaction by mixing the fusion protein, 10x SUMO protease buffer, SUMO protease (Invitrogen), and dH₂O.
2. Incubate the mixture at 4°C overnight.

Removal of SUMO and SUMO protease

1. Apply the SUMO cleavage mixture to the Ni-NTA column.
2. Collect the flow-through (purified CycT1-Tat protein) in a 15mL falcon tube

Day 8: Buffer Exchange and Concentration

Buffer Exchange

- Combine the fractions containing the CycT1-Tat protein band (~17 kDa) and place in a 3000 MWCO dialyzer (Slide-A-Lyzer® Dialyzer Cassette) in a storage buffer (20 mM Tris, 200 mM NaCl, 5% glycerol, 10 µM ZnCl₂, 0.02% Sodium Azide, 5 mM TCEP; pH 7.4) overnight at +4°C

Concentration

-Place the purified protein, which is buffer-exchanged into the storage buffer in a VIVASPIN 15mL Concentrator (3000 MWCO). Put the concentrator in a 50mL falcon conical bottom tube. Balance the tubes and concentrate the sample to 1-2mL by centrifuging at 3000rpm, 30 minutes per run at room temperature. It would take several runs to concentrate the sample to the desired volume.


After concentrating, a UV measurement (280 nm) is taken, and sample concentration is calculated (using molar extinction coefficient $21970 \text{ M}^{-1}\text{cm}^{-1}$). Place the stock aliquots in the -80°C freezers for long-term storage.

Mass Spectroscopy (MALDI-TOF)

MS buffer consisting of 5mM sodium phosphate, 0.2 M NaCl, $1 \mu\text{M}$ ZnCl_2 , 1 mM TCEP.HCl, pH 7.0) was prepared to dialyze purified and concentrated protein samples overnight at $+4^{\circ}\text{C}$. The concentrations of the dialyzed samples were determined by UV absorbance at 280 nm using MS buffer as a blank. $20 \mu\text{M}$ protein samples were prepared in 0.1% aqueous trifluoroacetic acid (TFA) solution. A matrix solution containing 1:2 of sinapinic acid in acetonitrile: 0.1% aqueous TFA was prepared and mixed with the diluted protein sample in a 1:1 ratio. Multiple $1 \mu\text{L}$ protein samples were spotted on a MALDI sample plate, and it was air-dried for an hour. The molecular weight of CycT1-Tat chimera (F1) was confirmed using an Autoflex III Smartbeam MALDI-TOF system (Bruker Daltonics, Billerica, MA) at the State University of New York College of Environmental Science and Forestry (SUNY-ESF), Syracuse, NY.

Electrophoretic mobility shift (EMSA) assay

The diluted TAR RNA samples were prepared in Dnase-free water by heating at 94°C for 3 min and incubated on ice. TAR RNA samples ($1 \mu\text{M}$) were incubated with purified CycT1-Tat protein in gel shift buffer (20 mM Tris, 0.200 M NaCl, $10 \mu\text{M}$ ZnCl_2 , 0.02% Sodium Azide) at $+4^{\circ}\text{C}$ for 15 min. 12% acrylamide native gel was pre-run at 100 V for an hour in 1X TBE buffer. The binding reactions were loaded on the gel and electrophoresed at 100 V for 2 h at RT.

This page is available in the following languages: 



Creative Commons License Deed

Attribution 4.0 International (CC BY 4.0)



This is a human-readable summary of (and not a substitute for) the [license](#).

You are free to:

Share — copy and redistribute the material in any medium or format

Adapt — remix, transform, and build upon the material

for any purpose, even commercially.

The licensor cannot revoke these freedoms as long as you follow the license terms.

Under the following terms:



Attribution — You must give appropriate credit, provide a link to the license, and indicate if changes were made. You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use.

No additional restrictions — You may not apply legal terms or technological measures that legally restrict others from doing anything the license permits.

Notices:

You do not have to comply with the license for elements of the material in the public domain or where your use is permitted by an applicable exception or limitation.

No warranties are given. The license may not give you all of the permissions necessary for your intended use. For example, other rights such as publicity, privacy, or moral rights may limit how you use the material.

References

- (1) Ouyang, W.; Okaine, S.; McPike, M. P.; Lin, Y.; Borer, P. N. Probing the RNA binding surface of the HIV-1 nucleocapsid protein by site-directed mutagenesis. *Biochemistry* **2013**, *52* (19), 3358-3368. DOI: 10.1021/bi400125z.
- (2) Mohammad, M. M.; Iyer, R.; Howard, K. R.; McPike, M. P.; Borer, P. N.; Movileanu, L. Engineering a rigid protein tunnel for biomolecular detection. *J Am Chem Soc* **2012**, *134* (22), 9521-9531. DOI: 10.1021/ja3043646.
- (3) Fujinaga, K.; Irwin, D.; Taube, R.; Zhang, F.; Geyer, M.; Peterlin, B. M. A minimal chimera of human cyclin T1 and tat binds TAR and activates human immunodeficiency virus transcription in murine cells. *J Virol* **2002**, *76* (24), 12934-12939.
- (4) World Health Organization. *HIV*. 2021. <https://www.who.int/hiv/en/> (accessed 12/05/22).
- (5) Panos, G.; Samonis, G.; Alexiou, V. G.; Kavarnou, G. A.; Charatsis, G.; Falagas, M. E. Mortality and morbidity of HIV infected patients receiving HAART: a cohort study. *Curr HIV Res* **2008**, *6* (3), 257-260.
- (6) Johnson, V. A.; Brun-Vezinet, F.; Clotet, B.; Gunthard, H. F.; Kuritzkes, D. R.; Pillay, D.; Schapiro, J. M.; Richman, D. D. Update of the drug resistance mutations in HIV-1: December 2009. *Top HIV Med* **2009**, *17* (5), 138-145.
- (7) Duffy, S.; Shackelton, L. A.; Holmes, E. C. Rates of evolutionary change in viruses: patterns and determinants. *Nat Rev Genet* **2008**, *9* (4), 267-276. DOI: 10.1038/nrg2323.
- (8) Marcelin, A. G.; Ceccherini-Silberstein, F.; Perno, C. F.; Calvez, V. Resistance to novel drug classes. *Curr Opin HIV AIDS* **2009**, *4* (6), 531-537. DOI: 10.1097/COH.0b013e328331d4b1.
- (9) Das, A. T.; Berkhout, B. HIV-1 evolution: frustrating therapies, but disclosing molecular mechanisms. *Philos Trans R Soc Lond B Biol Sci* **2010**, *365* (1548), 1965-1973. DOI: 10.1098/rstb.2010.0072.

- (10) Johnson, V. A.; Brun-Vezinet, F.; Clotet, B.; Kuritzkes, D. R.; Pillay, D.; Schapiro, J. M.; Richman, D. D. Update of the drug resistance mutations in HIV-1: Fall 2006. *Top HIV Med* **2006**, *14* (3), 125-130.
- (11) Human Immunodeficiency Virus (HIV). *Transfus Med Hemother* **2016**, *43* (3), 203-222. DOI: 10.1159/000445852 From NLM.
- (12) van Heuvel, Y.; Schatz, S.; Rosengarten, J. F.; Stitz, J. Infectious RNA: Human Immunodeficiency Virus (HIV) Biology, Therapeutic Intervention, and the Quest for a Vaccine. **2022**, *14* (2), 138.
- (13) Rossi, E.; Meuser, M. E.; Cunanan, C. J.; Cocklin, S. Structure, Function, and Interactions of the HIV-1 Capsid Protein. **2021**, *11* (2), 100.
- (14) Freed, E. O. HIV-1 gag proteins: diverse functions in the virus life cycle. *Virology* **1998**, *251* (1), 1-15. DOI: 10.1006/viro.1998.9398.
- (15) De Guzman, R. N.; Wu, Z. R.; Stalling, C. C.; Pappalardo, L.; Borer, P. N.; Summers, M. F. Structure of the HIV-1 nucleocapsid protein bound to the SL3 psi-RNA recognition element. *Science* **1998**, *279* (5349), 384-388.
- (16) Mervis, R. J.; Ahmad, N.; Lillehoj, E. P.; Raum, M. G.; Salazar, F. H.; Chan, H. W.; Venkatesan, S. The gag gene products of human immunodeficiency virus type 1: alignment within the gag open reading frame, identification of posttranslational modifications, and evidence for alternative gag precursors. *J Virol* **1988**, *62* (11), 3993-4002.
- (17) Coffin, J. M.; Hughes, S. H.; Varmus, H. E. The Interactions of Retroviruses and their Hosts. In *Retroviruses*, Coffin, J. M., Hughes, S. H., Varmus, H. E. Eds.; 1997.
- (18) Lin, Y., Borer, P. N.,. Variation in 1700 NCp7 sequences. 2002.
- (19) South, T. L.; Summers, M. F. Zinc- and sequence-dependent binding to nucleic acids by the N-terminal zinc finger of the HIV-1 nucleocapsid protein: NMR structure of the complex

with the Psi-site analog, dACGCC. *Protein Sci* **1993**, 2 (1), 3-19. DOI:

10.1002/pro.5560020102.

(20) Cruceanu, M.; Urbaneja, M. A.; Hixson, C. V.; Johnson, D. G.; Datta, S. A.; Fivash, M. J.; Stephen, A. G.; Fisher, R. J.; Gorelick, R. J.; Casas-Finet, J. R.; et al. Nucleic acid binding and chaperone properties of HIV-1 Gag and nucleocapsid proteins. *Nucleic Acids Res* **2006**, 34 (2), 593-605. DOI: 10.1093/nar/gkj458.

(21) Darlix, J. L.; Cristofari, G.; Rau, M.; Pechoux, C.; Berthoux, L.; Roques, B.

Nucleocapsid protein of human immunodeficiency virus as a model protein with chaperoning functions and as a target for antiviral drugs. *Adv Pharmacol* **2000**, 48, 345-372.

(22) Musah, R. A. The HIV-1 nucleocapsid zinc finger protein as a target of antiretroviral therapy. *Curr Top Med Chem* **2004**, 4 (15), 1605-1622.

(23) Schafer, A.; Bogerd, H. P.; Cullen, B. R. Specific packaging of APOBEC3G into HIV-1 virions is mediated by the nucleocapsid domain of the gag polyprotein precursor. *Virology* **2004**, 328 (2), 163-168. DOI: 10.1016/j.virol.2004.08.006.

(24) Paoletti, A. C.; Shubsda, M. F.; Hudson, B. S.; Borer, P. N. Affinities of the nucleocapsid protein for variants of SL3 RNA in HIV-1. *Biochemistry* **2002**, 41 (51), 15423-15428.

(25) Shubsda, M. F.; Paoletti, A. C.; Hudson, B. S.; Borer, P. N. Affinities of packaging domain loops in HIV-1 RNA for the nucleocapsid protein. *Biochemistry* **2002**, 41 (16), 5276-5282.

(26) Yuan, Y.; Kerwood, D. J.; Paoletti, A. C.; Shubsda, M. F.; Borer, P. N. Stem of SL1 RNA in HIV-1: structure and nucleocapsid protein binding for a 1 x 3 internal loop. *Biochemistry* **2003**, 42 (18), 5259-5269. DOI: 10.1021/bi034084a.

- (27) Clever, J.; Sasseti, C.; Parslow, T. G. RNA secondary structure and binding sites for gag gene products in the 5' packaging signal of human immunodeficiency virus type 1. *J Virol* **1995**, *69* (4), 2101-2109.
- (28) Clever, J. L.; Parslow, T. G. Mutant human immunodeficiency virus type 1 genomes with defects in RNA dimerization or encapsidation. *J Virol* **1997**, *71* (5), 3407-3414.
- (29) Harrison, G. P.; Lever, A. M. The human immunodeficiency virus type 1 packaging signal and major splice donor region have a conserved stable secondary structure. *J Virol* **1992**, *66* (7), 4144-4153.
- (30) Hayashi, T.; Shioda, T.; Iwakura, Y.; Shibuta, H. RNA packaging signal of human immunodeficiency virus type 1. *Virology* **1992**, *188* (2), 590-599.
- (31) Hayashi, T.; Ueno, Y.; Okamoto, T. Elucidation of a conserved RNA stem-loop structure in the packaging signal of human immunodeficiency virus type 1. *FEBS Lett* **1993**, *327* (2), 213-218.
- (32) McBride, M. S.; Panganiban, A. T. The human immunodeficiency virus type 1 encapsidation site is a multipartite RNA element composed of functional hairpin structures. *J Virol* **1996**, *70* (5), 2963-2973.
- (33) Sakaguchi, K.; Zambrano, N.; Baldwin, E. T.; Shapiro, B. A.; Erickson, J. W.; Omichinski, J. G.; Clore, G. M.; Gronenborn, A. M.; Appella, E. Identification of a binding site for the human immunodeficiency virus type 1 nucleocapsid protein. *Proc Natl Acad Sci U S A* **1993**, *90* (11), 5219-5223.
- (34) Iyer, R. M. DESIGN AND SYNTHESIS OF A BISTABLE APTAMER BASED THREE-SEGMENT SWITCH AS A DIAGNOSTIC INDICATOR FOR HIV-1 NCp7 DETECTION. **2020**, Dissertations. (accessed 12/01/22).
- (35) Karn, J. Tackling Tat. *J Mol Biol* **1999**, *293* (2), 235-254. DOI: 10.1006/jmbi.1999.3060.

- (36) Sodroski, J.; Patarca, R.; Rosen, C.; Wong-Staal, F.; Haseltine, W. Location of the trans-activating region on the genome of human T-cell lymphotropic virus type III. *Science* **1985**, *229* (4708), 74-77.
- (37) Taube, R.; Peterlin, M. Lost in transcription: molecular mechanisms that control HIV latency. *Viruses* **2013**, *5* (3), 902-927. DOI: 10.3390/v5030902.
- (38) Garber, M. E.; Wei, P.; KewalRamani, V. N.; Mayall, T. P.; Herrmann, C. H.; Rice, A. P.; Littman, D. R.; Jones, K. A. The interaction between HIV-1 Tat and human cyclin T1 requires zinc and a critical cysteine residue that is not conserved in the murine CycT1 protein. *Genes Dev* **1998**, *12* (22), 3512-3527.
- (39) Asamitsu, K.; Fujinaga, K.; Okamoto, T. HIV Tat/P-TEFb Interaction: A Potential Target for Novel Anti-HIV Therapies. *Molecules* **2018**, *23* (4). DOI: 10.3390/molecules23040933 From NLM.
- (40) Tahirov, T. H.; Babayeva, N. D.; Varzavand, K.; Cooper, J. J.; Sedore, S. C.; Price, D. H. Crystal structure of HIV-1 Tat complexed with human P-TEFb. *Nature* **2010**, *465* (7299), 747-751. DOI: 10.1038/nature09131 From NLM.
- (41) Campbell, G. R.; Loret, E. P. What does the structure-function relationship of the HIV-1 Tat protein teach us about developing an AIDS vaccine? *Retrovirology* **2009**, *6*, 50. DOI: 10.1186/1742-4690-6-50.
- (42) Jakobovits, A.; Smith, D. H.; Jakobovits, E. B.; Capon, D. J. A discrete element 3' of human immunodeficiency virus 1 (HIV-1) and HIV-2 mRNA initiation sites mediates transcriptional activation by an HIV trans activator. *Mol Cell Biol* **1988**, *8* (6), 2555-2561.
- (43) Selby, M. J.; Bain, E. S.; Luciw, P. A.; Peterlin, B. M. Structure, sequence, and position of the stem-loop in tar determine transcriptional elongation by tat through the HIV-1 long terminal repeat. *Genes Dev* **1989**, *3* (4), 547-558.

- (44) Wei, P.; Garber, M. E.; Fang, S. M.; Fischer, W. H.; Jones, K. A. A novel CDK9-associated C-type cyclin interacts directly with HIV-1 Tat and mediates its high-affinity, loop-specific binding to TAR RNA. *Cell* **1998**, *92* (4), 451-462.
- (45) Tabarrini, O.; Desantis, J.; Massari, S. Recent advances in the identification of Tat-mediated transactivation inhibitors: progressing toward a functional cure of HIV. *Future Med Chem* **2016**, *8* (4), 421-442. DOI: 10.4155/fmc.16.3 From NLM.
- (46) Fischer, C. L. Recombinant Expression, Purification, and Characterization of an HIV-1 Tat-Human Cyclin T1 Chimera". **2013**, Dissertations. (accessed 12/01/22).
- (47) Nimjee, S. M.; Rusconi, C. P.; Sullenger, B. A. Aptamers: an emerging class of therapeutics. *Annu Rev Med* **2005**, *56*, 555-583. DOI: 10.1146/annurev.med.56.062904.144915.
- (48) Keefe, A. D.; Pai, S.; Ellington, A. Aptamers as therapeutics. *Nature Reviews Drug Discovery* **2010**, *9* (7), 537-550. DOI: 10.1038/nrd3141.
- (49) Eyetech Study, G. Preclinical and phase 1A clinical evaluation of an anti-VEGF pegylated aptamer (EYE001) for the treatment of exudative age-related macular degeneration. *Retina* **2002**, *22* (2), 143-152. DOI: 10.1097/00006982-200204000-00002.
- (50) Gold, L.; Ayers, D.; Bertino, J.; Bock, C.; Bock, A.; Brody, E. N.; Carter, J.; Dalby, A. B.; Eaton, B. E.; Fitzwater, T.; et al. Aptamer-based multiplexed proteomic technology for biomarker discovery. *PLoS One* **2010**, *5* (12), e15004. DOI: 10.1371/journal.pone.0015004.
- (51) Zuo, X.; Li, S.; Hall, J.; Mattern, M. R.; Tran, H.; Shoo, J.; Tan, R.; Weiss, S. R.; Butt, T. R. Enhanced expression and purification of membrane proteins by SUMO fusion in *Escherichia coli*. *J Struct Funct Genomics* **2005**, *6* (2-3), 103-111. DOI: 10.1007/s10969-005-2664-4.
- (52) Ellington, A. D.; Szostak, J. W. In vitro selection of RNA molecules that bind specific ligands. *Nature* **1990**, *346* (6287), 818-822. DOI: 10.1038/346818a0.

- (53) Tuerk, C.; Gold, L. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science* **1990**, *249* (4968), 505-510. DOI: 10.1126/science.2200121.
- (54) Wang, R. E.; Zhang, Y.; Cai, J.; Cai, W.; Gao, T. Aptamer-based fluorescent biosensors. *Curr Med Chem* **2011**, *18* (27), 4175-4184. DOI: 10.2174/092986711797189637.
- (55) Deng, B.; Lin, Y.; Wang, C.; Li, F.; Wang, Z.; Zhang, H.; Li, X. F.; Le, X. C. Aptamer binding assays for proteins: the thrombin example--a review. *Anal Chim Acta* **2014**, *837*, 1-15. DOI: 10.1016/j.aca.2014.04.055.
- (56) Musumeci, D.; Platella, C.; Riccardi, C.; Moccia, F.; Montesarchio, D. Fluorescence Sensing Using DNA Aptamers in Cancer Research and Clinical Diagnostics. *Cancers (Basel)* **2017**, *9* (12). DOI: 10.3390/cancers9120174 From NLM.
- (57) DeCiantis, C. L.; Jensen, D. K.; Hudson, B. S.; Borer, P. N. A nucleic acid switch triggered by the HIV-1 nucleocapsid protein. *Biochemistry* **2007**, *46* (32), 9164-9173. DOI: 10.1021/bi700031j.
- (58) DeCiantis, C. L. A FRET based bistable oligonucleotide switch, AlloSwitch, designed for specific recognition of HIV-1 NCp7 and use in High Throughput Screening. **2008**, Dissertation. (accessed 08/12/21).
- (59) McPike, M. P., Maqbool, S. B., Chen, L., Ouyang, W., Borer, P. N.,. Parallel analysis of protein binding to DNA libraries using multiplexed microarray chips.
- (60) SantaLucia, J., Jr.; Hicks, D. The thermodynamics of DNA structural motifs. *Annu Rev Biophys Biomol Struct* **2004**, *33*, 415-440. DOI: 10.1146/annurev.biophys.32.110601.141800.
- (61) Zuker, M.; Sankoff, D. RNA secondary structures and their prediction. *Bulletin of Mathematical Biology* **1984**, *46* (4), 591-621. DOI: [https://doi.org/10.1016/S0092-8240\(84\)80062-2](https://doi.org/10.1016/S0092-8240(84)80062-2).

- (62) D H Turner; N Sugimoto, a.; Freier, S. M. RNA Structure Prediction. **1988**, *17* (1), 167-192. DOI: 10.1146/annurev.bb.17.060188.001123.
- (63) SantaLucia, J., Jr. A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. *Proc Natl Acad Sci U S A* **1998**, *95* (4), 1460-1465. DOI: 10.1073/pnas.95.4.1460.
- (64) Athavale, S. S.; Ouyang, W.; McPike, M. P.; Hudson, B. S.; Borer, P. N. Effects of the nature and concentration of salt on the interaction of the HIV-1 nucleocapsid protein with SL3 RNA. *Biochemistry* **2010**, *49* (17), 3525-3533. DOI: 10.1021/bi901279e From NLM.
- (65) Bintzler, D. A.; Terrell, C. E. 10-nmol oligonucleotide synthesis for the ABI model 394 DNA synthesizer. *J Biomol Tech* **2000**, *11* (3), 122-134.
- (66) Surzhikov, S. A.; Timofeev, E. N.; Chernov, B. K.; Golova, J. B.; Mirzabekov, A. D. Advanced method for oligonucleotide deprotection. *Nucleic Acids Res* **2000**, *28* (8), E29. DOI: 10.1093/nar/28.8.e29.
- (67) Technologies, I. D. *Oligo synthesis: Why IDT leads the oligo industry*. 2015 (Updated 2018). <https://www.idtdna.com/pages/education/decoded/article/oligo-synthesis-why-idt-leads-the-oligo-industry> (accessed June 2, 2022).
- (68) Research, G. *Glen Report 21.211: Technical Brief – Synthesis of Long Oligonucleotides*. <https://www.glenresearch.com/reports/gr21-211> (accessed October 5, 2021).
- (69) Mullah, B.; Andrus, A.; Zhao, H.; Jones, R. A. Oxidative conversion of N-dimethylformamidinium nucleosides to N-cyano nucleosides. *Tetrahedron Letters* **1995**, *36* (25), 4373-4376. DOI: [https://doi.org/10.1016/0040-4039\(95\)00816-U](https://doi.org/10.1016/0040-4039(95)00816-U).
- (70) Research, G. *dT-CE Phosphoramidite*. https://www.glenresearch.com/structural-studies/10-1030.html?gclid=EAIaIQobChMI7fmI_ISb_AIVA4_ICh34JAe6EAAYAiAAEgJ69PD_BwE (accessed May 10,2016).

- (71) ATDBio. Nucleic acids book. (accessed October 15, 2021).
- (72) Research, G. User Guide to Glen-Pak™ Purification. (accessed 10 October, 2016).
- (73) Applied BioSystems, I. Evaluating and Isolating Synthetic Oligonucleotides- The Complete Guide. **1992**, A-7.
- (74) Research, G. Glen Report 14.12: A New Simplified 3'-Amino-Modifier CPG. (accessed February 3, 2015).
- (75) Sun, H.; Zhu, X.; Lu, P. Y.; Rosato, R. R.; Tan, W.; Zu, Y. Oligonucleotide aptamers: new tools for targeted cancer therapy. *Mol Ther Nucleic Acids* **2014**, 3 (8), e182. DOI: 10.1038/mtna.2014.32 From NLM.
- (76) Research, G. Glen Report 23.17: Technical Brief - Chemical Phosphorylation. Web Page. (accessed 06/10/2015).
- (77) Kallansrud, G.; Ward, B. A comparison of measured and calculated single- and double-stranded oligodeoxynucleotide extinction coefficients. *Anal Biochem* **1996**, 236 (1), 134-138. DOI: 10.1006/abio.1996.0141.
- (78) McFarland, G. D.; Borer, P. N. Separation of oligo-RNA by reverse-phase HPLC. *Nucleic Acids Res* **1979**, 7 (4), 1067-1080. DOI: 10.1093/nar/7.4.1067.
- (79) Tinoco, I., Jr.; Li, P. T.; Bustamante, C. Determination of thermodynamics and kinetics of RNA reactions by force. *Q Rev Biophys* **2006**, 39 (4), 325-360. DOI: 10.1017/s0033583506004446 From NLM.
- (80) Research, G. Technical Bulletin: Running conditions of Denaturing PAGE. Web Page. (accessed 05/25/2015).
- (81) Guillatt, A. M. Agarose and polyacrylamide gel electrophoresis. *Methods Mol Biol* **2002**, 187, 1-11. DOI: 10.1385/1-59259-273-2:001 From NLM.

- (82) Pappalardo, L.; Kerwood, D. J.; Pelczer, I.; Borer, P. N. Three-dimensional folding of an RNA hairpin required for packaging HIV-1. *J Mol Biol* **1998**, *282* (4), 801-818. DOI: 10.1006/jmbi.1998.2046 From NLM.
- (83) Cavaluzzi, M. J.; Borer, P. N. Revised UV extinction coefficients for nucleoside-5'-monophosphates and unpaired DNA and RNA. *Nucleic Acids Res* **2004**, *32* (1), e13. DOI: 10.1093/nar/gnh015.
- (84) Taube, R.; Fujinaga, K.; Wimmer, J.; Barboric, M.; Peterlin, B. M. Tat transactivation: a model for the regulation of eukaryotic transcriptional elongation. *Virology* **1999**, *264* (2), 245-253. DOI: 10.1006/viro.1999.9944 From NLM.
- (85) D'Orso, I.; Frankel, A. D. HIV-1 Tat: Its Dependence on Host Factors is Crystal Clear. *Viruses* **2010**, *2* (10), 2226-2234. DOI: 10.3390/v2102226 From NLM.
- (86) Butt, T. R.; Edavettal, S. C.; Hall, J. P.; Mattern, M. R. SUMO fusion technology for difficult-to-express proteins. *Protein Expr Purif* **2005**, *43* (1), 1-9. DOI: 10.1016/j.pep.2005.03.016.
- (87) Demain, A. L.; Vaishnav, P. Production of recombinant proteins by microbes and higher organisms. *Biotechnol Adv* **2009**, *27* (3), 297-306. DOI: 10.1016/j.biotechadv.2009.01.008 From NLM.
- (88) Costa, S.; Almeida, A.; Castro, A.; Domingues, L. Fusion tags for protein solubility, purification and immunogenicity in *Escherichia coli*: the novel Fh8 system. *Front Microbiol* **2014**, *5*, 63. DOI: 10.3389/fmicb.2014.00063 From NLM.
- (89) Lee, C. D.; Sun, H. C.; Hu, S. M.; Chiu, C. F.; Homhuan, A.; Liang, S. M.; Leng, C. H.; Wang, T. F. An improved SUMO fusion protein system for effective production of native proteins. *Protein Sci* **2008**, *17* (7), 1241-1248. DOI: 10.1110/ps.035188.108 From NLM.

- (90) Kapust, R. B.; Waugh, D. S. Escherichia coli maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused. *Protein Sci* **1999**, *8* (8), 1668-1674. DOI: 10.1110/ps.8.8.1668 From NLM.
- (91) Malakhov, M. P.; Mattern, M. R.; Malakhova, O. A.; Drinker, M.; Weeks, S. D.; Butt, T. R. SUMO fusions and SUMO-specific protease for efficient expression and purification of proteins. *J Struct Funct Genomics* **2004**, *5* (1-2), 75-86. DOI: 10.1023/b:Jsfg.0000029237.70316.52 From NLM.
- (92) Marblestone, J. G.; Edavettal, S. C.; Lim, Y.; Lim, P.; Zuo, X.; Butt, T. R. Comparison of SUMO fusion technology with traditional gene fusion systems: enhanced expression and solubility with SUMO. *Protein Sci* **2006**, *15* (1), 182-189. DOI: 10.1110/ps.051812706.
- (93) Mossessova, E.; Lima, C. D. Ulp1-SUMO crystal structure and genetic analysis reveal conserved interactions and a regulatory element essential for cell growth in yeast. *Mol Cell* **2000**, *5* (5), 865-876. DOI: 10.1016/s1097-2765(00)80326-3 From NLM.
- (94) Sun, Z.; Xia, Z.; Bi, F.; Liu, J. N. Expression and purification of human urodilatin by small ubiquitin-related modifier fusion in Escherichia coli. *Appl Microbiol Biotechnol* **2008**, *78* (3), 495-502. DOI: 10.1007/s00253-007-1330-0.
- (95) Scientific, T. F. *Invitrogen™ Champion™ pET SUMO Expression System*.
https://www.thermofisher.com/order/catalog/product/K30001?gclid=EAIaIQobChMIhr3fmv6a_AIVw9zICh20kwY5EAAYASAAEgKW3PD_BwE&ef_id=EAIaIQobChMIhr3fmv6a_AIVw9zICh20kwY5EAAYASAAEgKW3PD_BwE:G:s&s_kwcid=AL!3652!3!533624062684!!g!!!383441308!122660977177&cid=bid_elb_pex_r01_co_cp0000_pjt0000_bid00000_0se_gaw_dy_pur_con&s_kwcid=AL!3652!3!533624062684!!g!! (accessed November 5, 2017.)

Order Number: 1312899
Order Date: 18 Jan 2023

Payment Information

Nan Thuzar Myint
nmyint@syr.edu
Payment method: Invoice

Billing Address:
Nan Thuzar Myint
111 College Place
SYRACUSE, NY 13244
United States

3152785921
nmyint@syr.edu

Customer Location:
Nan Thuzar Myint
111 College Place
SYRACUSE, NY 13244
United States

Order Details

1. Future medicinal chemistry

Article: Recent advances in the identification of Tat-mediated transactivation inhibitors: progressing toward a functional cure of HIV.

Billing Status:
Open

Order License ID	1312899-1	Type of use	Republish in a thesis/dissertation
Order detail status	Completed	Publisher	Future Science Ltd
ISSN	1756-8927	Portion	Image/photo/illustration
			0.00 USD
			Republication Permission

LICENSED CONTENT

Publication Title	Future medicinal chemistry	Publication Type	e-Journal
Article Title	Recent advances in the identification of Tat-mediated transactivation inhibitors: progressing toward a functional cure of HIV.	Start Page	421
		End Page	442
		Issue	4
Date	01/01/2009	Volume	8
Language	English		
Country	United Kingdom of Great Britain and Northern Ireland		
Rightholder	Future Science Ltd.		

REQUEST DETAILS

Portion Type	Image/photo/illustration	Distribution	Worldwide
Number of Images / Photos / Illustrations	1	Translation	Original language of publication
Format (select all that apply)	Print, Electronic	Copies for the Disabled?	No
Who Will Republish the Content?	Academic institution	Minor Editing Privileges?	No
Duration of Use	Life of current edition	Incidental Promotional Use?	No
Lifetime Unit Quantity	Up to 499	Currency	USD
Rights Requested	Main product		

NEW WORK DETAILS

Title	Re-engineering of Recombinant Cyclin T1-Tat Protein with SUMO Fusion in Escherichia Coli	Institution Name	Syracuse University
Instructor Name	N/A	Expected Presentation Date	2023-01-18

ADDITIONAL DETAILS

The Requesting Person/Organization to Appear on the License Nan Thuzar Myint

REUSE CONTENT DETAILS

Title, Description or Numeric Reference of the Portion(s)	Figure 7: The Tat/trans-activating region recognition motif of CycT1	Title of the Article/Chapter the Portion Is From	Recent advances in the identification of Tat-mediated transactivation inhibitors: progressing toward a functional cure of HIV.
Editor of Portion(s)	Tabarrini, Oriana; Desantis, Jenny; Massari, Serena		

Bibliographic Information

NAME OF AUTHOR: Nan Thuzar Myint

PLACE OF BIRTH: Kyaing Tong, Shan State, Myanmar

DATE OF BIRTH: March 31, 1989

DEGREE AWARDED:

Bachelor of Science in Biochemistry, 2012, Utica University, Utica, NY, USA

Master of Philosophy in Chemistry, 2014, Syracuse University, Syracuse, NY, USA