Title

Biophysical and Biochemical Applications of Fluorescent Tricyclic Cytidine for Nucleic Acid Labeling and DNA Structural Dynamics

Author

Turner, Marc Benjamin

Publication Date

2020

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA SAN DIEGO
SAN DIEGO STATE UNIVERSITY

Biophysical and Biochemical Applications of Fluorescent Tricyclic Cytidine for Nucleic Acid Labeling and DNA Structural Dynamics

A dissertation submitted in partial satisfaction of the
requirements for the degree Doctor of Philosophy

in

Chemistry

by

Marc Benjamin Turner

Committee in charge:

University of California San Diego

Professor Ulrich Muller
Professor Charles Perrin

San Diego State University

Professor Byron Purse, Chair
Professor Ralph Feuer
Professor Jeffrey Gustafson

2020

The Dissertation of Marc Benjamin Turner is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

_____

_____
Chair

University of California San Diego

San Diego State University

2020

TABLE OF CONTENTS

Signature Page iii

Table of Contents iv

List of Abbreviations viii

List of Symbols vviiii

List of Figures ix

List of Schemes xiiii

List of Tables xiiiiii

Acknowledgements xiv

Vita xviiiii

Abstract of the Dissertation xxx

Chapter 1: Fluorescent Nucleoside Analogues to Study Nucleic Acids while Preserving Native Biological Context 1

1.1 Introduction 1

1.2 Structural biophysics of nucleic acids 3

1.3 Examining structure using fluorescence spectroscopy 7

1.4 Applications of fluorescent nucleoside analogues 11

1.5 Incorporating fluorescent nucleosides in oligonucleotides 18

1.6 Concluding remarks and the work of this dissertation 22

1.7 References 23

Chapter 2: Spectroscopic Analysis of the Fluorescent Turn-on Nucleoside 8-diethlamino-tC in DNA-RNA Hybrids to Identify Mechanism of Emissive Response 28

2.1 Introduction 28

2.2 Greater fluorescence turn-on in DNA-RNA duplexes 30

2.3 Lifetime measurements and structural investigation 33

2.4 Evaluating extent of helical shielding from external quenching mechanisms 35

2.5 Acid-induced fluorescence of ^{DEA}tC nucleoside 39

2.6 Discussion 41

2.7 Conclusion 45

2.8 Supporting Data and Figures 46

2.9 Experimental Methods 50

2.10 Acknowledgements 55

2.11 References 55

Chapter 3: Fluorescent Tricyclic Cytidine Analogues as Substrates for Retroviral Reverse Transcriptases 58

3.1 Introduction 58

3.2 Overview of methods and results 62

3.3 RNA-templated single nucleotide insertion and read-through 64

3.4 DNA-templated single nucleotide insertion kinetics and read-through 69

3.5 Modeling in vitro reverse transcription with fluorescent tricyclic cytidine analogues 76

3.6 Discussion 78

3.7 Conclusion 80

3.8 Supporting Figures 86

3.9 Experimental Methods 89

3.10 Acknowledgements 93

3.11 References 93

Chapter 4: Live-cell Studies using tC nucleotides with Prokaryotic and Eukaryotic Cells 97

4.1 Introduction 97

4.2 Metabolic labeling of transformed prokaryotes 100

4.3 Delivering fluorescent nucleotides to eukaryotic cells 104

4.3 Future metabolic labeling studies 109

4.4 Experimental Methods 109

4.5 References 123

Chapter 5: Conclusion 126

5.1 References 97

LIST OF ABBREVIATIONS

298T	surface-adherent line of human T-cells
AMV	avian myeloblastosis virus
BL-21	*Escherichia coli* strain optimized for heterologous gene expression
DNA	deoxynucleic acid; single-stranded (ssDNA), double-stranded (dsDNA)
CD	circular dichroism
FISH	Fluorescence *in situ* hybridization
HIV-1	human immunodeficiency virus 1
Huh7	human hepatocyte derived from a carcinoma cell line
KIE	kinetic isotope effect
M-MLV	Moloney murine leukemia virus
ODN	oligonucleotide
PAGE	polyacrylamide gel electrophoresis
*Pt*NTT2	*Phaeodactylum tricornutum* nucleoside triphosphate transporter
PBS	phosphate buffered saline
SUPT1	human T-cells derived from a lymphoma cell line
RNA	ribonucleic acid; single-stranded (ssRNA), double-stranded (dsRNA), messenger (mRNA), ribosomal (rRNA), transfer (tRNA)
RT	reverse transcriptase
YT	yeast extract tryptone growth media

LIST OF SYMBOLS

E_R	reduction potential
eV	electron volts
$E_{\pi\text{-}\pi}$	base-stacking energy in DNA/RNA
k	Log growth rate of bacteria
k_f	radiative rate
K_M	Michaelis-Menten constant observed in enzyme catalysis
k_{nr}	non-radiative decay rate
k_q	bimolecular quenching constant
K_{SV}	Stern-Volmer quenching coefficient
T_m	melting temperature
V_{MAX}	maximum reaction rate observed in Michaelis-Menten enzyme catalysis
α_F	fraction folded/annealed
ε	molar absorptivity
η	index of refraction
λ	wavelength
τ	excited-state lifetime
Φ	quantum yield of emission
ω	twist angle between adjacent pairs of DNA/RNA bases in a double-stranded helix
$\tilde{\nu}$	wavenumber

LIST OF FIGURES

Figure 1.1. Base-pairing modes of canonical DNA bases 2

Figure 1.2. Biophysical parameters of double-stranded nucleic acids. 4

Figure 1.3. Natural non-canonical DNA structures. 6

Figure 1.4. Jablonski diagram illustrating sequence of steps resulting in fluorescence. 8

Figure 1.5. Time-resolved fluorescence spectrum of tC deoxyribose nucleoside. 10

Figure 1.6. Select examples of intrinsically fluorescent nucleoside analogues. 14

Figure 1.7. Structure and fluorescence of ^{DEA}tC nucleoside in DNA. 17

Figure 1.8. Common incorporation strategies of nucleoside analogues into nucleic acids. 19

Figure 2.1. Structure and fluorescence of ^{DEA}tC nucleoside in DNA. 29

Figure 2.2. Comparing quantum yields of ^{DEA}tC oligonucleotides 32

Figure 2.4. Stern-Volmer analysis of GXC duplexes. 37

Figure 2.5. Solvent isotope effects on time-resolved ^{DEA}tC oligonucleotides fluorescence 38

Figure 2.6. Induced fluorescence of ^{DEA}tC under acidic conditions. 40

Figure 2.S1. Time-resolved emission decays of ^{DEA}tC DNA-RNA duplexes. 47

Figure 2.S2. Minor lifetime $<\tau 2>$ of probe plotted against reduction potential. 47

Figure 2.S3. Circular dichroism spectra and melting curves of ^{DEA}tC DNA-RNA sequences. 48

Figure 2.S4. Melting curves of ^{DEA}tC DNA-RNA sequences. 49

Figure 2.S5. Stern-Volmer analysis of sequences TXA and CXA. 50

Figure 3.1. A simplified sequence of retroviral nucleic acid replication. 60

Figure 3.2. Structure of tC and ^{DEA}tC nucleotide analogue fluorophores. 62

Figure 3.3. Primer–template sequences for polymerization experiments. 64

Figure 3.4. Michaelis-Menten plots of single insertion using an RNA_G template. 66

Figure 3.5. Read-through polymerization assay of RNA_G primer-template complex. 68

Figure 3.6. Read-through assay of RNA_G primer-template complex without dTTP. 69

Figure 3.7. Michaelis-Menten plots of matched single insertion using DNA_G template..... 71

Figure 3.8. Read-through polymerization assay of DNA_G primer-template complex. 75

Figure 3.9. Read-through assay of reverse transcriptases across G and A quartets. 76

Figure 3.10. Modeling reverse transcription with nucleotide analogues. 78

Figure S3.1. Michaelis-Menten plots for mismatch insertion with DNA_A. 86

Figure S3.2. Read-through polymerization assay of DNA_A primer-template complex..... 87

Figure S3.3 ^{1}H NMR spectrum of 8-Diethylamino-tC 2'-deoxy-β-$_D$-ribonucleoside-5'-*O*-triphosphate..... 88

Figure S3.4 ^{31}P NMR spectrum of 8-Diethylamino-tC 2'-deoxy-β-$_D$-ribonucleoside-5'-*O*-triphosphate..... 89

Figure 4.1. Strategy to deliver fluorescent nucleotides to bacterial cells. 98

Figure 4.2. Metabolic conversion of nucleoside phosphoramidate to triphosphate..... 105

Figure 4.3. Select single-cell fluorescence microscopy of Huh7 cells. 107

Figure 4.4. Flow cytometry histograms from human cells. 109

Figure 4.5. ^{1}H NMR spectrum of 2',3',5'-tri-*O*-acetyl-8-diethylamino-tC-ribonucleoside 116

Figure 4.6. ^{1}H NMR spectrum of 8-diethylamino-tC-ribonucleoside 117

Figure 4.7. ^{13}C NMR spectrum of 8-diethylamino-tC-ribonucleoside 118

Figure 4.8. ^{1}H NMR spectrum of 8-diethylamino-tC-ribonucleoside 5'-*O*-triphosphate..... 119

Figure 4.9. ^{31}P NMR spectrum of 8-diethylamino-tC-ribonucleoside 5'-*O*-triphosphate 120

Figure 4.10. ^{1}H NMR spectrum of (S)-2-(S)-phenoxyphosphorylamino tC 2'-deoxy-β-$_D$-2'-deoxyribonucleoside propionic acid isopropyl ester..... 121

Figure 4.11. ^{13}C NMR spectrum of (S)-2-(S)-phenoxyphosphorylamino tC 2'-deoxy-β-$_D$-2'-deoxyribonucleoside propionic acid isopropyl ester..... 122

Figure 4.12. ^{31}P NMR spectrum of (S)-2-(S)-phenoxyphosphorylamino tC 2'-deoxy-β-$_D$-2'-deoxyribonucleoside propionic acid isopropyl ester..... 123

LIST OF SCHEMES

Scheme 1.1. Optimized synthesis of ^{DEA}tC phosphoramidite 20

Scheme 3.1 Synthesis of 8-Diethylamino-tC 2'-deoxy-β-$_D$-ribonucleoside-5'-triphosphate..... 90

Scheme 4.1. Synthesis of tC 2'-deoxyribose nucleoside phosphoramidate. 99

Scheme 4.2. Synthesis of ^{DEA}tC ribose nucleoside triphosphate..... 101

LIST OF TABLES

Table 1.1. Select physical parameters distinguishing A-form and B-form DNA. 5

Table 1.2. Fluorescent parameters for select nucleoside analogues. 13

Table 1.3. Spectroscopic properties of tC-derived nucleoside analogs 16

Table 2.1. Steady-State fluorescence measurements from ^{DEA}tC DNA-RNA duplexes. 31

Table 2.2. Time-resolved fluorescence measurements and relaxation rate 344

Table 2.3. Quenching efficiency measured by Stern-Volmer analysis 36

Table 2.4. Solvent isotope effects on ^{DEA}tC oligonucleotide fluorescence 39

Table 2.S1. Thermostability of ^{DEA}tC DNA-RNA oligonucleotides 46

Table 2.S2. Quantitative analysis of ^{DEA}tC oligonucleotides. 51

Table 3.1. Kinetics measurements for RNA_G-templated single dNTP insertion. 66

Table 3.2 Kinetics measurements for DNA_G-templated single dNTP insertion. 72

Table 3.3 Kinetics Measurements from DNA_A-templated mismatch insertion. 74

Table 3.4 Efficacy summary of RT with d($^{(DEA)}$tC)TP, and RNA or DNA templates 80

Table 4.1. Fluorescence latency with tC 2'-deoxyribose nucleotides. 102

Table 4.2. Fluorescence latency with tC ribose nucleotides. 102

Table 4.3. Fluorescence latency with ^{DEA}tC nucleotides after 24 hours. 103

ACKNOWLEDGEMENTS

The past five years of graduate education has undoubtedly been the toughest challenge I have endeavored in my life, though as I am still quite young, perhaps I will have equally challenging and meaningful experiences ahead. Managing through the immense rigor and difficulty of my graduate experience would not have been feasible without the people whom I typically enjoyed knowing along the way. Without the encouragement and inspiration from my family, friends, and lab mates, I very likely would have forfeited higher education.

Firstly, Byron, as my PI you have been the most constant individual these past five years and I still greatly appreciate you indulging the young biologist's aspiration of becoming an organic chemist, even with me having practically no chemical laboratory experience. You were always very kind and polite, even when providing constructive criticism, and while sometimes the harsher critiques were only subtly veiled, I still appreciated your tactful delivery. I myself learned subtle tactics to communicate my feelings without being completely outright, such as responding with "yes, professor" when your requests became mildly excessive. So in addition to critical thinking, project planning, and physical organic chemistry points, I value most what you taught me about communication. Whether engaging professional colleagues, superiors, subordinates, or even the average person on the street, being able to clearly, concisely, and tactfully communicate our intentions or feelings is an essential life skill. Now that is not at all to devalue the chemistry-based lessons you taught me, but this is my dissertation, so I can write what I want.

Secondly, my lab mates have been absolutely critical in my graduate years, whether that was through emotional support, professional assistance, social comfort, or inadvertently demonstrating how I ought not to be. Kristine Teppang, Dillon Burns, Breanna Canter, and Ray

Lee were the first individuals I met in the Purse lab and they will likely remain some of the most memorable people in my life. Though I thought Kristine detested me at first, I later learned she was just shy and we became really good friends and continue to do so. Quite in fact, becoming friends with Kristine was one of my favorite experiences I would have otherwise missed had I not selected graduate education. Breanna was always so kind and helpful to me, and I greatly appreciated someone patient like her while learning basic lab techniques. Dillon was a fun person to learn synthetic chemistry from and I feel omitting some of the more memorable instances is prudent for maintaining a professional dissertation. Ray was my unofficial mentor and I thoroughly missed his quick wit and critical thinking after he graduated. He taught me HPLC in one day and useful phrases like "it's not bad data, it's just not good news, and maybe that's the truth" when experiment results do not meet our expectations. Acknowledgements to Sadie, an exceptionally cheerful person to be around, to Sara, a great example for all aspiring women in science, and to Michael, a very nice, albeit at times difficult, person to work with.

Through the five years I met amazing undergraduates like Katrina Ngo, Cesar Garcia, Karina, Julian, Mackenzie Wyllie, and Grace Kim. Mackenzie was socially one of the most fascinating people I have ever met—a day with her in lab is not boring day at all. Grace is a great researcher and despite her shy demeanor she has the boldness and tenacity that research work requires. I hope she does well wherever she goes. I have enjoyed meeting other graduate students like George and Ani, who are both very quiet and reserved but are excellent workers and good people to be around. Jesus (our lab mate, not our Lord) Ceja, has been one of the most important friends I have met in life. In addition to creating an enjoyable social environment to get through the last two years of graduate work, you have constantly provided wise life lessons, such looking at personal struggle not at the time of experiencing it, but appreciating in retrospect later in life

how we develop from our strife, or how the romanticized allure of honorable service in uniform is not necessarily worth the unnecessary anguish that often accompanies military life. Though I did enjoy hearing stories about what people do when imprisoned on a ship for several months at a time. I am genuinely grateful you were in lab for the final homestretch of my graduate studies.

I would like to thank Diego Avalleneda and Lucas Luna from Professor Christal Sohl's biochemistry research lab for helping with some of the bacterial culturing and Dr. Sohl for allowing me to use their lab's materials, necessary equipment, and reagents. For the experiments with the human cells we greatly appreciate Danielle Slemons in the biology department for managing the cells and Cameron for operating the flow cytometers. I personally would like to thank Dr. Roland Wolkwowicz for not only being an excellent collaborator but also for being an outstanding professor and one of the best people whom I met in my 9 years at SDSU. Final acknowledgments to my cohort classmates like Dan, Sean, and Andrew—we helped each other out through our coursework and research challenges.

The contents of Chapter 2, in part, is currently being prepared for submission for publication of the following material: Turner, M. B., Ceja, J., Kim, G. Y., and Purse, B. W. Spectroscopic analysis of the fluorescent turn-on nucleoside 8-diethlamino-tC in DNA-RNA hybrids to identify mechanism of emissive response. *J. Phys. Chem. B.* 2020. The dissertation author was the lead author of this paper.

The contents of Chapter 3 contains material that is similar to the material published in the following manuscript: Turner, M. B., and Purse B. W. Fluorescent tricyclic cytidine analogues as substrates for retroviral reverse transcriptases. *ChemPlusChem*. 2020, 85, 855-865. The dissertation author was the lead author of this paper.

VITA

2015	Bachelor of Science, San Diego State University
2015-2017	Teaching Assistant, San Diego State University
2015-2020	Graduate Researcher, San Diego State University
2020	Doctor of Philosophy, University of California San Diego & San Diego State University

PUBLICATIONS

1. Turner, M. B., Ceja, J., Kim, G. Y., Pearce, H., Cooksy, A., and Purse, B. W., "Increasing the fluorescence turn-on of the nucleoside probe ^{DEA}tC in DNA-RNA helices" J. Phys. Chem. C. 2020. [In Preparation].
2. Turner, M. B., and Purse B. W. Fluorescent tricyclic cytidine analogues as substrates for retroviral reverse transcriptases" ChemPlusChem. 2020, 85, 855-865.
3. Turner, M. B., Anderson, B. A., Samaan, G. N., Coste, M., Burns, D. D., & Purse, B. W. "Synthesis of the fluorescence turn-on DNA hybridization probe using the ^{DEA}tC 2'-deoxycytidine analog" Curr. Protoc. Nucleic Acid Chem. 2018, 75, e59.
4. Teppang, K. ., Lee, R. ., Burns, D. ., Turner, M., Lokensgard, M. ., Cooksy, A., and Purse, B. Electronic modifications of fluorescent cytidine analogues control photophysics and fluorescent responses to base stacking and pairing. Chem. Eur. J. 2019, 25, 1249-1259.

PRESENTATIONS

Oral

1. “Synthesis and Spectroscopic Properties of Fluorescent ^{DEA}tC Oligonucleotides” American Chemical Society General Meeting, San Diego International Convention Center, San Diego CA. August 29, 2019.

Poster

1. “Increasing the Fluorescence Turn-On of ^{DEA}tC in DNA-RNA Heteroduplexes” Methods and Applications of Fluorescence, University of California San Diego, La Jolla CA. August 21, 2019.
2. “Synthesis and Photophysical Properties of Modified Nucleic Acid Oligonucleotides with Fluorescent Tricyclic Cytidine” CSU Program for Education & Research in Biotechnology, Hyatt Regency, Anaheim, CA. January 3-5, 2019.
3. “Fluorescent nucleoside analogues towards monitoring retroviral replication in live cells” International Round Table on Nucleosides, Nucleotides, and Nucleic Acids, University of California San Diego, La Jolla CA. August 27-30, 2018.
4. “DNA Polymerization Kinetics of d(tC)TP by Reverse Transcriptase” Student Research Symposium, San Diego State University, San Diego CA. March 2, 2018

ABSTRACT OF THE DISSERTATION

Biophysical and Biochemical Applications of Fluorescent Tricyclic Cytidines in Nucleic Acids

by

Marc Benjamin Turner

Doctor of Philosophy in Chemistry

University of California San Diego, 2020
San Diego State University, 2020

Professor Byron Purse, Chair

Intrinsically fluorescent nucleoside analogues with minimally perturbing architecture offer a non-disruptive method for optically studying nucleic acid structure, dynamics, and metabolism. These nucleoside analogues have inherently fluorescent nucleobases, thus replacing the natural canonical bases of DNA, but retain Watson-Crick base-pairing. Minimally perturbing structures are those that induce few or mild adverse deviations from native, unmodified biomolecular conformations. Using fluorescent nucleobases eliminates fluorophore conjugation chemistry often employed in conventional labeling strategies and their minimally perturbing

structure better preserves the natural biomolecular topography and function of DNA. Since the discovery of 2-aminopurine 51 years ago, a multitude of nucleoside analogue scaffolds have been synthesized, including the tricyclic cytidine (tC) structure that the Purse Lab uses as a platform to develop novel derivatives with varied photophysical properties. Applying these novel derivatives towards biological applications requires characterization of their fluorescent, biophysical properties in DNA helices, and their compatibility with polymerase enzymes for metabolic labeling. This dissertation describes the biophysical characterization of a fluorescence turn-on probe in DNA-RNA hybrids, the polymerization kinetics exhibited by viral reverse transcriptase when incorporating the fluorescent nucleotides into nascent DNA strands, and preliminary live cell studies with fluorescent nucleotides. To better understand environmental factors modulating the turn-on response in the derivative 8-diethylamino tC, spectroscopic experiments with the probe in DNA-RNA hybrids were performed. Shielding from bulk water attenuated excited-state proton transfer and was identified to be a major factor increasing the brightness by up to 37-fold compared to the nucleoside in water. Michaelis-Menten kinetics for viral reverse transcriptase inserting the nucleotide analogues during DNA polymerization show efficiencies from 0.09 – 5 times that of natural dCTP across from G, with continued strand elongation. In a simplified reverse transcription cycle model, HIV-1 RT effectively recognized the tC analogues as the incoming nucleotide and as the templating base during complementary DNA synthesis. Live cell work with bacterial and human cells has suggested that tC nucleotides are retained in cells following delivery and that cytotoxicity is not a major concern. The findings of these studies suggest these tC-derivatives are appropriate biophysical tools for applications including sequence detection, DNA/RNA structural changes, and metabolic labeling.

Chapter 1

Fluorescent Nucleoside Analogues to Study Nucleic Acids while Preserving Native Biological Context

1.1 Introduction

Nucleic acids are one of the four principle macromolecules in biology, in addition to carbohydrates, peptides, and lipids. Each of these macromolecules can form a biopolymer such as proteins, fats, or polysaccharides, and in the case of nucleic acids, nucleotides are the monomeric building blocks that consist of a nitrogenous nucleobase, a five-carbon ribose, and a phosphate group often appended to *C5'* or *C3'*. Nucleotides lacking a hydroxyl on *C2'* are termed 2'-deoxyribonucleotides to form DNA, whereas nucleotides retaining the *C2'* hydroxyl form RNA. The structure and function of DNA and RNA macromolecules can vary distinctly and understanding what functions these molecules may perform as a result of their secondary, tertiary, or quaternary structure aside from their primary sequence remains an actively investigated topic.

When considering the biophysical structure of double-stranded DNA (dsDNA), we may immediately envision an ideal, right-handed, anti-parallel double-stranded helix exemplified by the structure published by Watson and Crick in 1954.[1] While double-stranded DNA (dsDNA) predominantly adopts B-form conformation, A-form conformation has been observed for double-stranded RNA or DNA-RNA hybrid helices as an alternate right-handed conformation.[2] Despite not sharing in the 1962 Nobel Prize, Rosalind Franklin generated essential X-ray diffraction patterns of dsDNA, which were unethically delivered to Watson and Crick and facilitated the structural determination of DNA, indicating that water content modulated the conformational structure of DNA helices.[3] The hydrogen-bonding mode predominantly observed in double-

stranded nucleic acids is termed Watson-Crick H-bonding, in which there are three intermolecular H-bonds between G:C pairs and two for A:T/U pairs (**Figure 1.1**). Conventional intermolecular H-bonding within DNA has alternative H-bonding modes between nucleobases. One example of alternate H-bonding includes Hoogsteen base-pairing in which H-bonds are formed with heterocyclic atoms on the nucleobase not involved in canonical Watson-Crick base-pairing.[4] These alternative modes endow DNA molecules with additional biological functions besides encoding genes for transcription and translation.

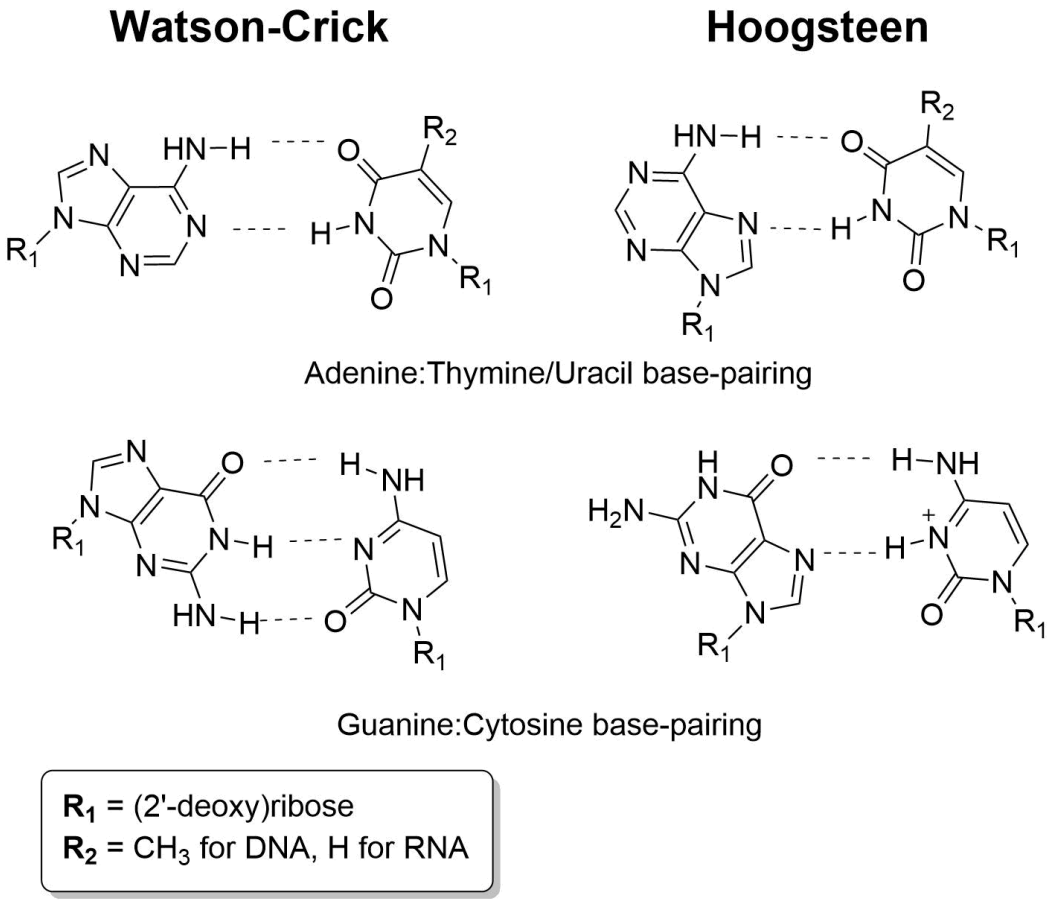


Figure 1.1. Base-pairing modes of canonical DNA bases.

In this review, fluorescent nucleoside analogues are discussed and how their spectroscopic properties make them highly informative tools for studying nucleic acids. Examples of nucleic acid physical diversity are described, followed by the basic mechanics of

fluorescence, and then canonical nucleoside analogues with intrinsic fluorescence. Chemists have developed a wide array of tools for studying the biological role of DNA/RNA molecules such as G-quadruplexes, i-motifs, or long non-coding RNA.[5–7] Fluorescent nucleoside analogues as bioanalytical tools have been essential on reporting on dynamic, subtle environmental changes or structural conformations that might not be apparent from classic techniques. Minimally perturbing, fluorescent nucleoside structures are expected to preserve as much as possible the most native structure and activity of nucleic acids. Lastly, a brief overview for introducing these unnatural analogues into DNA is discussed followed by concluding perspectives.

1.2 Structural biophysics of nucleic acids

DNA/RNA biopolymers in solution adopt a variety of physical conformations, with increasing diversity at greater sizes and lengths. Studying the biophysical differences between A- and B-form DNA helices may be experimentally performed using with representative oligonucleotides 10-20 bases long or computationally using shorter sequences with optimized geometries.[8] A- and B-form DNA helices are not mutually exclusive, binary conformations but rather they are two representative sets of average structural physical parameters including twist angle, inclination, and base-pair step spacing (**Figure 1.2**). Actual DNA structures may adopt conformations matching idealized parameters for either form or may even adopt an intermediate conformation between A-form and B-form. Determining which helical structure a nucleic acid adopts includes preferred ribose ring conformation and intramolecular interactions between ribose hydroxyls and neighboring nucleobases.[9,10] Average physical parameters for the two helix types are shown in **Table 1.1**. While the helical diameter is approximately 20 Å for A- and B-form DNA, the preferred C-3’ endo ribose conformation in A-form causes the base pairs to be displaced by 4-5 Å, while the B-form C-2’ endo conformation leads to near zero displacement

from the central helical axis to the periphery of the helix.[11] This translates to 3D models displaying a prominent "hole" in A-form helices when viewed down the helical axis, while the aperture in B-form is near non-existent. Examining the global macromolecular structure has historically been performed using crystallography,[2] however favoring dynamic analytical techniques such as 2D NMR and circular dichroism spectroscopy operating on samples in solution permit more natural conformations as opposed to rigidified crystal structures.[12,13] Determining the dynamic 3D structures of A-form and B-form has enabled measuring average biophysical parameters distinguishing the two types.

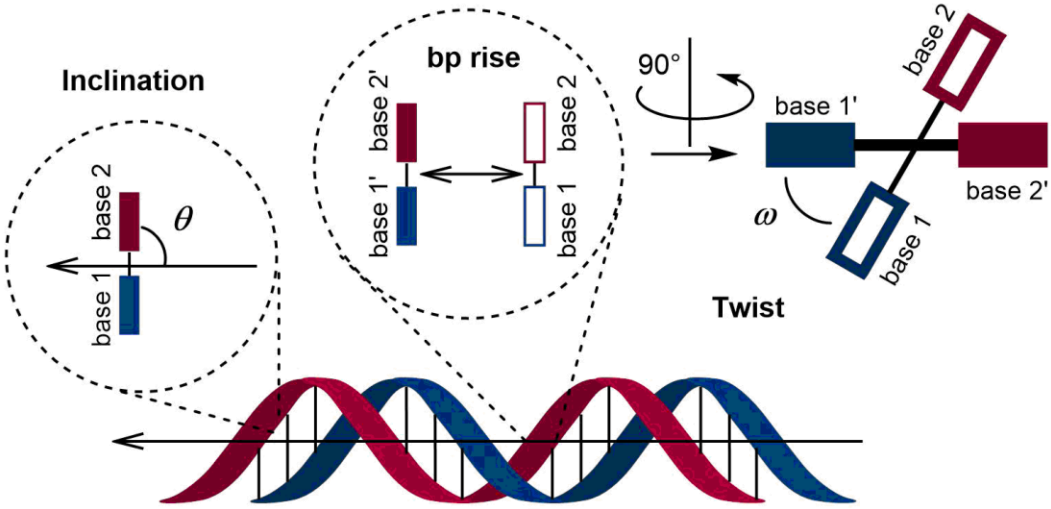


Figure 1.2. Biophysical parameters of double-stranded nucleic acids.

Transitions between A-form and B-form DNA can be forcibly induced by modulating the environment such as solvation with ethanol instead of water, concentration of salt ions, or DNA-protein binding events.[14–16] For example, the dsDNA genome of *Sulfolobus islandicus* rudivirus-2 (SIRV2) retains A-form structure in the viral capsid due to DNA-binding proteins interacting with the entire genome.[17] Longer DNA polymers may have a global average of A or B-form, which can be easily determined using CD spectroscopy,[13] however longer strands may also have

local regions of alternate forms modulated by specific sequences. Establishing a consistent, concrete correlation between certain sequences and their preferred A/B form conformation is statistically challenging, however localized stretches of DNA exhibit different propensities for their 3D topology depending on sequence.[18] Structures affected by sequence may occur through classic Watson-Crick base-pairs interacting with the DNA backbone or through alternated hydrogen-bonding, base-pairing modes.

Table 1.1. Select physical parameters distinguishing A-form and B-form DNA.

Parameter	**A-Form**	**B-Form**
Rise	260 pm	340 pm
Twist	32°	36°
Bases per turn	11	10
Inclination	70°	90°
Diameter	20 Å	20 Å
Sugar puckering	C-3’ endo	C-2’ endo

One example of alternate base-pairing is Hoogsteen base-pairing, which occurs naturally in G-quadruplexes (G4s). These non-canonical structures entail four-stranded complexes that form under physiological conditions and have been known to occur in G-rich sequences especially in eukaryotic telomere regions of eukaryotic genomes.[19] The tertiary structure originates from Hoogsteen base-pairing between four guanines while coordinating a central metal cation (**Figure 1.3A**), such as potassium, between the four *O6* atoms. G4s arising from intramolecular base-pairing may adopt varying arrangements, for instance anti-parallel as illustrated in **Figure 1.3B**,[20] and their presence in telomere and promoter sequences has been implicated in regulation of gene expression and epigenetic alterations.[5,21,22] Another non-canonical four-stranded structure is intercalated motifs (i-motifs) that may occur from cytosine-rich sequences, in which $C:C^+$ base-pairs are arranged in an intercalated fashion as opposed to the typical anti-parallel arrangement.[23] Similar to G4s, i-motifs occur naturally in human cell

nuclei and are implicated in gene regulation,[24,25] however they are cell-cycle and pH dependent.[7,26,27] To form a stable base-pair one of the cytosine bases must be cationic with *N3* protonated, whose pK_a is approximately 4.5 (**Figure 1.3C**).[28] The resulting base-pairs are stacked with an intercalated arrangement, meaning each base-pair step inserts between two other base-pairs from two other strands (**Figure 1.3D**).

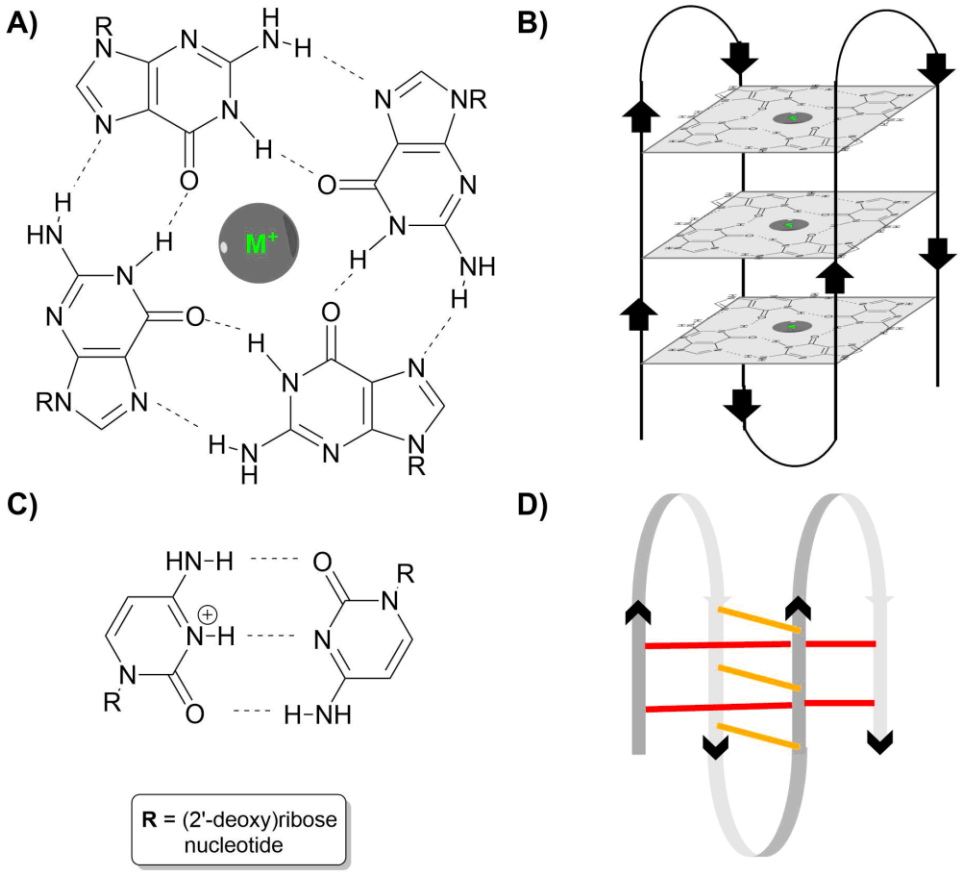


Figure 1.3. Natural non-canonical DNA structures. **A**) Four guanosine bases form tetramer from Hoogsteen base-pairing, coordinating a central metal cation. **B**) Stacking of G-tetramers gives rise to G-quadruplex, a four-stranded complex originating from a G-rich single strand. **C**) Non-canonical hemiprotonated cytosine base-pair at physiological conditions. **D**) Four-stranded complex composed of intercalating C^+:C base-pairs. Base-pair lengths are exaggerated for clarity.

By adopting discrete tertiary structures, G4s and I-motifs can serve as recognition sites for cellular proteins involved in genetic regulation. Examples include telomere DNA-binding

proteins at the ends of chromosomes or the regulatory promoter regions upstream of gene sequences.[28–32] Protein-DNA interactions in this case are determined not directly by sequence, rather from three-dimensional shape. Different G-quadruplex configurations exist in addition to the configuration shown in **Figure 1.3B**.[5] Both G4s and I-motifs as non-canonical structures discovered nearly 30 years following the published structure of classic DNA underscore the importance in studying how DNA impacts biological function as a result of physical structure. G4s, i-motifs, and Hoogsteen base-pairing were initially discovered by physical organic chemists and biochemists studying nucleic acids *in vitro*, however their occurrence in cells was later realized.

1.3 Examining structure using fluorescence spectroscopy

Spectroscopic methods for studying the structure and dynamics of DNA/RNA are highly informative techniques that do not require extensive sample manipulation or destruction as with X-ray crystallography, electron microscopy, and mass spectrometry. Furthermore, fluorescence spectroscopy using synthetically designed biomimetic analogues offers the advantage of easily detectable signals in the broad span of visible and near-infrared light, and can focus on specific structures or cellular locations.[33] Natural canonical nucleobases exhibit very short excited-state lifetimes and low quantum yields,[34,35] making them ill-suited as probes for fluorescence spectroscopy. Characterizing a fluorophore's fluorescence properties is essential when developing novel molecules, and especially for fluorescent nucleoside analogues where molecular size constraints, excited-state reactivity, and excitation/emission profiles have certain limitations so as to optimally preserve nucleic acid context and cellular function. For nucleoside analogues as an example, retaining natural biological function is more likely when employing structures that have molecular structures that minimally perturb DNA helices and do not induce

other adverse side-effects like photon-induced radical reactions. A large factor in preserving as much native DNA structure requires nucleobase analogues with conserved purine and pyrimidine architecture, for which many successful analogues have accomplished.

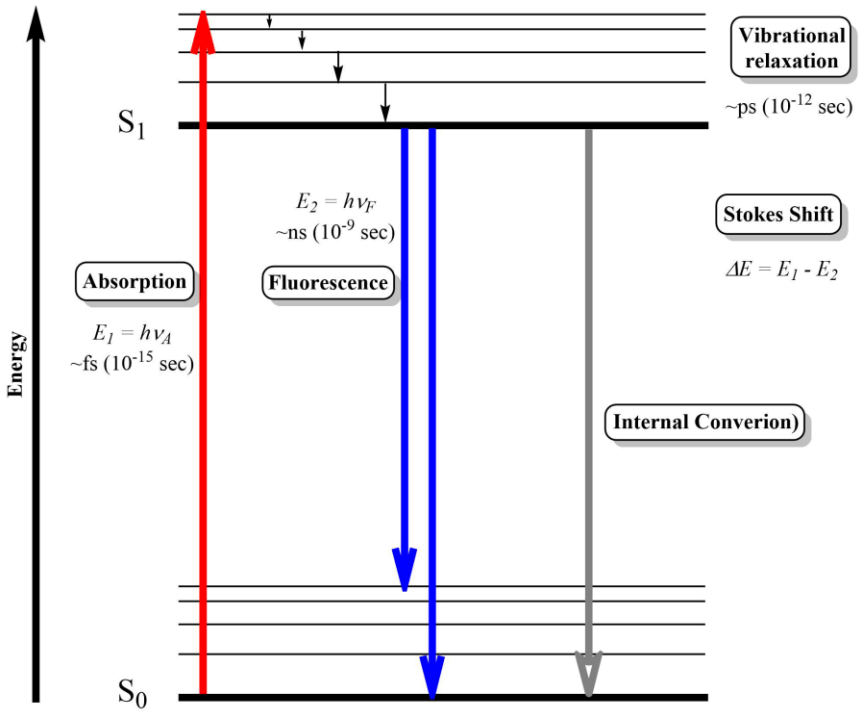


Figure 1.4. Jablonski diagram illustrating sequence of steps resulting in fluorescence (blue). Absorption (red) and excitation are used interchangeably. Energy resulting from internal conversion (gray) transforms into thermal energy which dissipates into the surroundings.

Fluorescence describes the spin-state allowed emission of a photon following absorption of a higher energy photon, and is one possible relaxation pathway a molecule may follow descending from a high electronic energy state to a lower one in photoluminescence. In simple models the low energy electronic state is termed the ground state and is denoted as S_0 while higher electronic states are termed excited states, S_n, where n is an integer ≥1. Photon-induced fluorescence occurs following excitation from S_0 to S_1 (for simplicity purposes n=1, but transitions to S_2 or higher can occur) that is induced by the molecular absorption of a photon. A chromophore's metric for light absorption is represented by its molar extinction coefficient (ε)

and is how strongly light of a certain wavelength is absorbed per chromophore quantity. The wavelength, frequency, or energy of that photon can be of a broad range of levels with different probabilities of raising an electron from the ground-state HOMO to the LUMO. This manifests as broad absorption spectra relative to other spectroscopy techniques, like NMR or mass spectrometry, which often provide sharp, narrow signals.

Within each electronic state there are multiple vibrational states and while in **Figure 1.4** excitation leads to the highest S_1 vibrational state, transitions to other discrete vibrational modes within a given S_n are possible, which causes signal broadening. Non-radiative decay within S_1, called vibrational relaxation, includes conformational changes and energy released as thermal radiation. Absorption is exceptionally fast, occurring on the femtosecond timescale, and leads to a S_1 vibrational state whose geometry initially resembles the original S_0 vibrational state. This phenomenon is known as the Franck-Condon principle. The time-scale for vibrational relaxation is 10^3 slower than that of absorption. Fluorescence refers specifically to the spontaneous radiative relaxation from S_1 to S_0 with an emitted photon lower in energy than the absorbed photon, with the energy difference known as the Stokes shift. Like absorption, emission spectra are broad due to the multiple vibrational modes. The S_1 to S_0 transition may also occur through non-radiative internal conversion. A fluorophore's quantum yield of emission (Φ_{em}) is the ratio of photons emitted per photons absorbed. Measuring Φ_{em} is typically done using steady-state fluorescence spectroscopy, which is measuring the total emission over all wavelengths from constant irradiation at a particular excitation wavelength, in comparison to a reference standard with a known Φ_{em} value.[36] Multiplying a fluorophore's Φ_{em} by its ε provides the overall brightness, which is an essential metric of fluorophore properties for biological and other applications.

During the typical nanosecond-timescale lifetime of a molecule's excited state, there is a certain probability that relaxation will proceed through spontaneous photon emission or non-radiative decay. Time-resolved fluorescence spectroscopy identifies the average excited-state lifetimes (τ), which can inform upon structural dynamics, molecular environment, and even temperature.[37] For the fluorophore tricyclic cytidine (shown as a 2'-deoxyribonucleoside), the time-resolved fluorescence spectrum is shown in **Figure 1.5**. An instrument timer coincides with excitation pulses of light and the instrument measures the time delay between excitation and detection of fluoresced photons from the sample. The baseline delay region inherent to instruments results from delays in the instrument components and radio frequency interference, but the real sample data is indicated by a signal spike often normalized to 1.0.[38] For accurate time-correlated single photon counting (TCSPC) data the detection rate is typically 1 fluoresced photon per 100 excitation pulses, with counted photons stored in histograms depending on the measured time delay. The heights of these histograms are represented as data points at discrete time intervals.

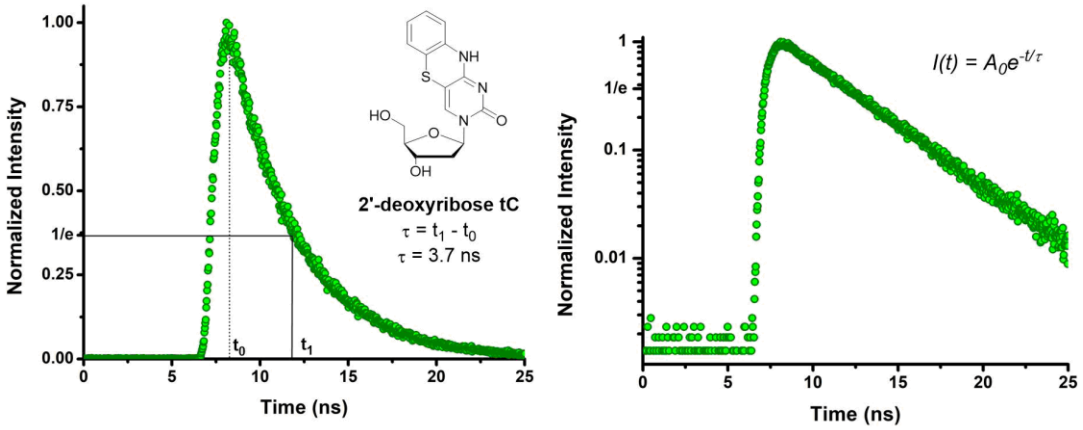


Figure 1.5. Time-resolved fluorescence spectrum of tC deoxyribose nucleoside. For exponential fitting the average scaling time (i.e. average lifetime) is the inverse base component, or 1/e = 0.368.

In time-resolved fluorescence spectra, each data point corresponds to a sum of spontaneous emission events collected at a certain time and therefore brighter fluorophores are often associated with longer lifetimes, because their comparatively longer-lived excited-state allows for higher probability of spontaneous emission. The radiative decay following excitation adheres to an exponential decay [$I(t)=A_0exp(-t/\tau)$], where the inverse of the base component, in this case Euler's number *e*, corresponds to the average excited-state lifetime, as indicated in **Figure 1.5**. For tricyclic cytidine, a single-component fitting is adequate,[39] however other fluorophores often have multiple excited-state lifetimes and require multi-component exponential fittings. Deciding how many components to include in the fitting can be visually approximated by representing intensity on a log scale, where single-component lifetimes have linear slopes whereas multi-component decays will have changes in the slope and may exhibit distinct regions of approximately linear slopes.

1.4 Applications of fluorescent nucleoside analogues

Commercial fluorophores conventionally employed in fluoresce-based analysis permit high sensitivity owing to their large quantum yields and molar extinction coefficients, however this results from large molecular scaffolds and extended conjugation and often flexible tethers linking the fluorophores to the biomolecules. Other conventional alternatives have used intercalating agents like ethidium bromide, SYBR green, or DAPI that offer high fluorescent signal at the expense of drastic structural disruption to nucleic acid helices.[40] Minimally perturbing nucleobase analogues must function like commercial fluorophores, but with molecular designs much smaller than commercial counterparts. Since the discovery of 2-aminopurine (2AP) in 1969, over 100 fluorescent nucleosides have been developed, differing in molecular designs, photophysical properties, and most appropriate biotechnological

applications.[41] A small set of minimally perturbing fluorescent nucleobase analogues are shown in **Figure 1.6**. Ideally, for most applications the molecular structures of fluorescent nucleosides should deviate minimally from their natural counterpart regarding size, shape, and base-pairing.

Fluorescent nucleobase analogue can report on nucleic acid biophysics by more than simple labeling and imaging, but through other physical properties including orientation and distance. For example, fluorescence anisotropy (i.e. unequal fluorescent properties depending on physical orientation) can report on molecular rigidity and sizes of nucleic acids.[42] In anisotropy experiments, a fluorophore is irradiated with polarized light and if the excited-state lifetime outpaces rotational motion, then emitted photons will retain equal polarization, allowing for detection through emission filters.[38] For fluorophores in rigid environments (e.g. nucleobases constrained in a DNA duplex) or with large hydrodynamic volumes, rotation is much slower and therefore more fluorescent signal is detected at the initial polarization.[43,44] Another fluorescence-based technique is Förster resonance energy transfer (FRET) where a donor fluorophore's excited-state transfers energy to a proximal acceptor fluorophore whose absorption energy overlaps with the donor's emission.[45,46] The transfer occurs through non-radiative dipole-dipole coupling and efficiency depends on the inverse exponential of distance between the donor and acceptor, making FRET a powerful tool for distance measurements, approximating accurate molecular distances 10 – 100 Å,[47] and conformational changes in DNA.[48] FRET also depends on fluorophore orientation, specifically aligning the transition dipole (not to be confused with the molecular dipole) parallel with the incident irradiation angle.[39] Therefore, anisotropy and FRET effectively function as molecular protractors and rulers for DNA/RNA. Spectroscopic techniques focusing not only on the location of fluorescently tagged nucleic acids, but also their biophysical

conformations based on orientation and distance of fluorophore reporters provides a deeper understanding of nucleic acid biophysics.

Table 1.2. Fluorescent parameters for select nucleoside analogues.

Compound	Solvent	λ_{ex} (nm)	λ_{em} (nm)	Molar Extinction (M^{-1} cm^{-1})	Quantum Yield	Brightness (M^{-1} cm^{-1})
2-AP	Water	303	370	6,800	0.68	4,600
Pyrrolo-dC	Buffer, pH 7.0	350	460	5,900	0.20	1,200
thG	Water	321	453	4,150	0.46	1,900
	1,4-Dioxane	333	424	4,530	0.50	2,300
tC	Buffer, pH 7.5	375	513	4,700	0.09	420
	1,4-Dioxane	369	472	4,300	0.34	1,500
tC^{O}	Buffer, pH 7.5	355	473	6,800	0.24	1,600
	1,4-Dioxane	357	445	7,400	0.46	3,400
qAN4	Water	356	445	7,300	0.30	2,200
^{DMA}T	Buffer, pH 7.0	357	522	2,900	0.03	87
	Buffer, pH 11.0	345	480	3,900	0.09	350

Parameters are from measurements performed on analogues as (2'-deoxy)ribose nucleosides.[33,49–51]

Molecular structures of nucleobase analogues that preserve natural arrangements as much as possible are known as isomorphic. For instance, thienoguanosine (thG) developed by the Tor research group,[52] which preserves a nearly identical shape to natural guanine but substitutes a thiophene for the imidazole 5-membered ring in purines.[53] Appending additional structural modifications to nucleobases may permit extended conjugation and enhanced fluorescent properties but risks compromising natural structures or functions. A reasonable settlement to this predicament entails designing ring-extended nucleobase analogues that mostly mimic isomorphic structures but with modifications inducing few to no adverse effects. Depicted in **Figure 1.6** and **Table 1.2** is a small subset of fluorescent purine and pyrimidine analogues with notable photophysical properties. Each molecule retains the necessary hydrogen-bond acceptor and donor atoms required for classic Watson-Crick base-pairing, enabling each analogue to function as a suitable base substitute in double-stranded DNA or RNA helices. The structural additions in

these ring-extended analogues project into the major groove of the DNA helix, which is more accommodating of structural perturbations than the minor groove.

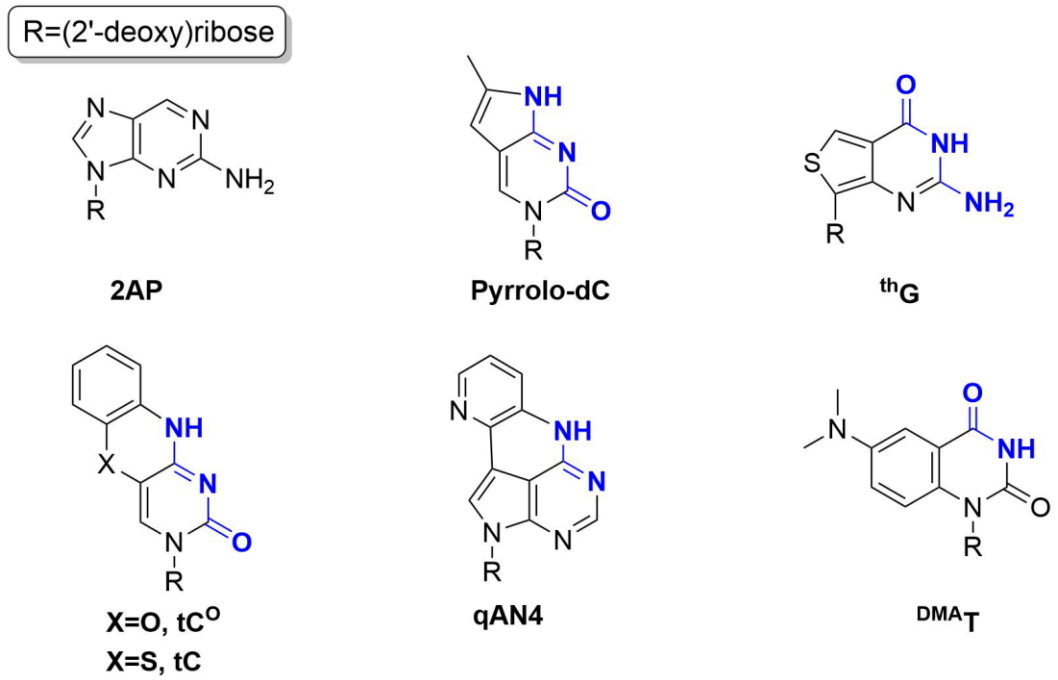


Figure 1.6. Select examples of intrinsically fluorescent nucleoside analogues. Molecular faces preserving Watson-Crick hydrogen-bonding are highlighted in blue.

One of the most informative capabilities that fluorescent nucleoside analogues have been focused on exploring, is reporting on physical changes that DNA/RNA molecules experience. Understanding how the structural and functional dynamics of DNA molecules are associated with subtle variations in the surrounding microenvironment (e.g. pH, local polarity, viscosity or rotational flexibility, binding events) can be achieved using fluorescent nucleoside analogues and sensitive spectroscopic methods.[41] In the cases of 2-AP and thG, base-stacking in the duplex strongly attenuates emission, in which case these analogues can be used to sense duplex denaturation.[53,54] As a double-stranded nucleic acid is denatured or degraded a large increase in fluorescence signal would be detected as base-stacking would no longer attenuate emission. Another highly emissive, recently reported purine analogue is the quadracyclic adenine

compounds (qAN) developed by the Wilhelmsson group.[50] Including a pyrimidine moiety (qAN4 shown in **Figure 1.6**) endowed greater brightness relative to the qAN parent nucleobase with a benzene moiety and more solvent sensitivity was observed.[55] While qAN4 likewise experiences a substantial drop in fluorescence when base-stacked in DNA, the analogue was shown to serve as a FRET donor with the non-emissive tC-derivative 8-nitro-tC^O to report on DNA structural dynamics.[55] A recent pyrimidine derivative developed by the Luedtke group, *N*,*N*-dimethylaniline-2'-deoxythymidine (^{DMA}T shown in **Figure 1.6**) exhibits pH-dependent fluorescence (**Table 1.2**) with greater brightness in basic environments.[51] ^{DMA}T was also used as a fluorescent reporter in measuring the binding kinetics between DNA and cytotoxic mercury cations (Hg^{2+}) to study the mechanism by which Hg^{2+} may inhibit DNA replication and enact cytotoxicity.[56] Another set of pyrimidine analogues, fluorescent tricyclic cytidines, is reviewed further.

The tricyclic cytidine (tC) family of compounds was originally developed by Gilead as a part of their antisense program because these compounds increase duplex stability when base-stacked, resulting from enhanced π-stacking in the secondary structure.[57] Altering the middle ring atom at position X (**Figure 1.6**) changes the photophysical properties, notably in the emission wavelengths or color, however substitutions at *C8* have yielded useful environmentally sensitive fluorophores (**Table 1.3**). Substituent modifications made by the Purse group at C8 were observed to modulate fluorescence more distinctly than modifications at C7, due to C8 substituents having direct resonance conjugation with the bridging oxygen or sulfur atom.[58] Installing electron-withdrawing substituents appears to both blue-shift the absorbance/emission maxima and also increases the overall brightness.[49] Conversely, electron-donating groups leads to red-shifted emission and comparatively less brightness. Oddly, emplacing a cyano group (CN)

did not produce the brightest derivative despite CN being the strongest electron-withdrawing substituent tested.

Table 1.3. Spectroscopic properties of tC-derived nucleoside analogues.[49,58,59]

Derivative	X	R group	Solvent	λ_{ex} (nm)	λ_{em} (nm)	Molar Extinction (M^{-1} cm^{-1})	Quantum Yield
$^{Cl}tC^{O}$	O	Cl	1X PBS, pH 7.4	361	458	5,600	0.35
			1,4-Dioxane	366	453	7,200	0.49
$^{CN}tC^{O}$	O	CN	1X PBS, pH 7.4	355[a]	429	5,100[a]	0.32
			1,4-Dioxane	n.d.	n.d.	n.d.	n.d.
$^{MeO}tC^{O}$	O	OMe	1X PBS, pH 7.4	361	499	4,000	0.05
			1,4-Dioxane	360	474	4,800	0.38
^{DEA}tC	S	NEt_2	1X PBS, pH 7.4	395	493	2,700	$6x10^{-3}$
			1,4-Dioxane	389	524	n.d.	n.d.
^{MeO}tC	S	Ome	1X PBS, pH 7.4	379	550	3,800	0.01
			1,4-Dioxane	372	500	3,800	0.28

a) Absorbance maximum and molar extinction coefficient for $^{CN}tC^{O}$ are displayed for excited-state transition matching energy of other tC^{O} compounds.

Substituting a chlorine at *C8* generates the brightest reported tC^{O} derivative in both aqueous and organic solvents. The derivative exhibits solvent sensitivity, observable in the molecule's fluorescence spectra As a free nucleoside in aprotic polar solvents such as 1,4-dioxane and dimethyl sulfoxide, vibrational features appear in the emission spectra. These fluctuations are separated from the peak emission wavelength by approximately 29 nm, corresponding to about 1400 cm^{-1} and vibrational transitions for aromatic C–C and C–N bonds. Recording the emission of $^{Cl}tC^{O}$ in protic solvents including ethanol or methanol dampens the appearance of these vibrational features, which are effectively ablated in aqueous solvent. As an environment-sensitive fluorophore to solvent polarity, $^{Cl}tC^{O}$ may be advantageous in reporting on polarity changes in the microenvironments of $^{Cl}tC^{O}$-labeled DNA/RNA in certain tertiary structures or bound to proteins. Another notable derivative, ^{DEA}tC, exhibits sensitivity to solvent and base-stacking, manifested as a fluorescence turn-on response when the analogue is base-

stacked in dsDNA helices (**Figure 1.7**). The obvious advantage that ^{DEA}tC provides is as a sequence-specific fluorescence turn-on probe for sensing duplex formation.

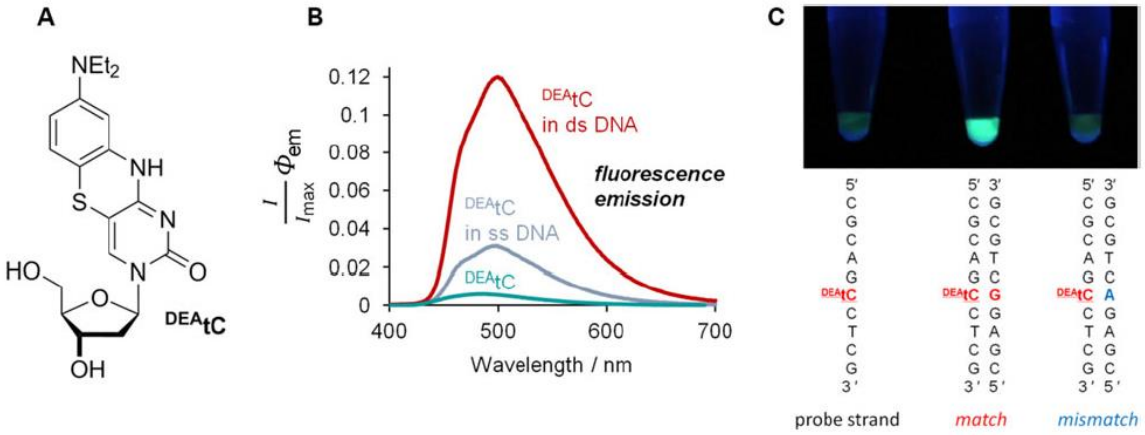


Figure 1.7. Structure and fluorescence of ^{DEA}tC nucleoside in DNA. **A**) Chemical structure of the cytidine analogue. **B**) Fluorescence turn-on responds to matched DNA duplex formation and base-stacking. **C**) The sequence specificity of turn-on response allows for visual sequence discrimination when samples are irradiated by a hand-held UV lamp.

Single-stranded oligonucleotides containing ^{DEA}tC behave largely like natural oligonucleotides with no defined secondary structure. Accordingly, the ^{DEA}tC fluorophore is minimally fluorescent but becomes substantially brighter when hybridized to a complementary sequence with a correct ^{DEA}tC:G base-pairing (**Figure 1.7C**). This fluorescence turn-on response does not appear when ^{DEA}tC is incorrectly base-paired with adenosine or an abasic lesion,[59] thereby making ^{DEA}tC oligonucleotides potential discriminating hybridization probes for fluorescence *in situ* hybridization (FISH) methods of sequence detection. False positives from FISH may occur because duplex thermodynamic stability (T_m) arises from preferential binding of two complementary sequences, but 1-2 base mismatches do not greatly affect T_m.[60] Therefore, ^{DEA}tC probes may be advantageous for detecting guanosine single nucleotide polymorphisms (SNP). The fluorescence turn-on intensity is sensitive to neighboring bases, an aspect that should be considered when planning applications. Another consideration is that ^{DEA}tC substitutions can

perturb the melting temperature of matched DNA duplexes. For the AXA sequence described, $\Delta T_m = -2.6$ °C, meaning that the ^{DEA}tC-containing duplex is less stable than the corresponding natural construct. For the GXC sequence, $\Delta T_m = -14.5$°C. The greater brightness of the GXC sequence comes at the expense of thermodynamic stability.

1.5 Incorporating fluorescent nucleosides into oligonucleotides

There are three frequently employed strategies for introducing fluorescent analogues into DNA or RNA oligonucleotides: solid-phase synthesis using fluorescent nucleoside phosphoramidites (**Figure 1.8A**), post-synthesis fluorophore conjugation to a nucleoside analogue, and enzymatic incorporation (**Figure 1.8B**). Each strategy entails certain benefits and limitations. Conjugating a fluorophore to an already existing DNA/RNA structure, derives from the first approach, but will not be described here.[61] The first method, solid-phase DNA synthesis, involves reagent-based DNA synthesis on a solid support called a resin, often consisting of controlled pore glass (CPG) or polystyrene, in repeating cycles to generate oligonucleotides synthesized in a 3' to 5' direction.[62–65] Nucleoside phosphoramidites with a 3' P(III) species and a 5' protected hydroxyl group enter the cycle and through a series of steps, which can be automated with appropriate instruments, unnatural substrates can be incorporated anywhere in the sequence.[66] As such, solid-phase DNA synthesis offers the advantages of automated production yielding oligonucleotides on large scales suitable for NMR or crystallography, and site-specific labeling.[67] Since synthesis utilizes reagents instead of polymerase enzymes, the procedure is amenable to nucleoside analogues with bulky or highly unnatural nucleobases that would otherwise pose challenges to polymerase active sites.[68] The major drawbacks to this strategy include the cost and resources synthesizing sufficient quantities of phosphoramidite for oligonucleotide production, and sequence fidelity is limited to fewer than ~100 bases in DNA

oligonucleotides and fewer than ~50 bases in RNA.[69] Improvements to solid-phase DNA synthesis methods will likely need to focus on efficiently synthesizing longer oligonucleotides.

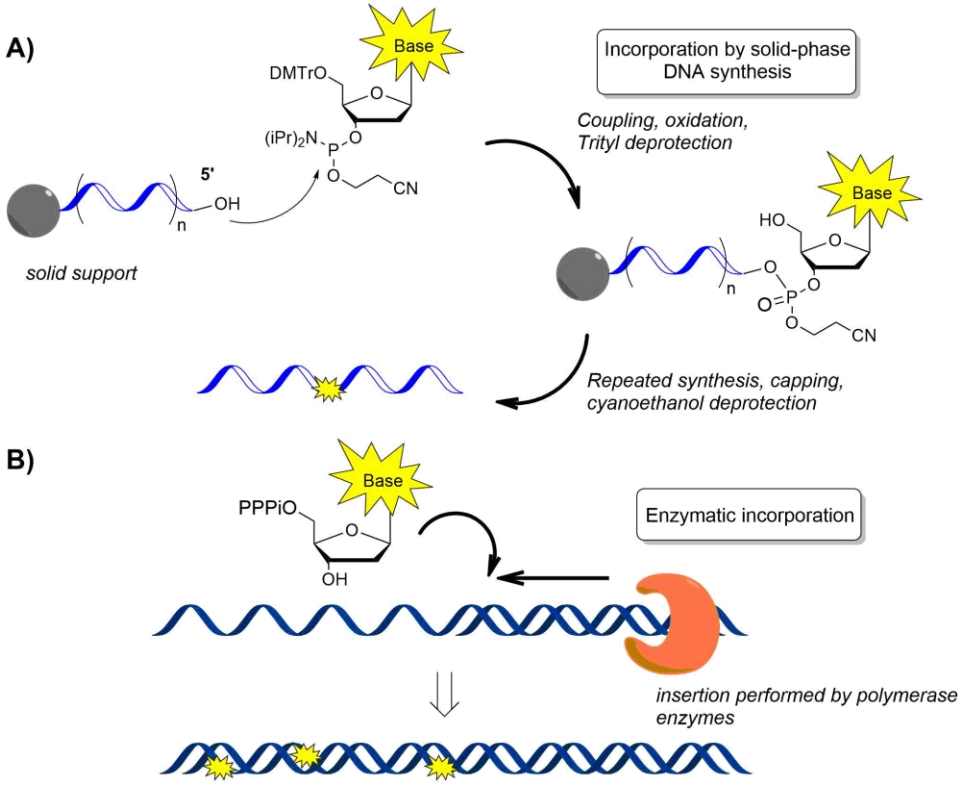
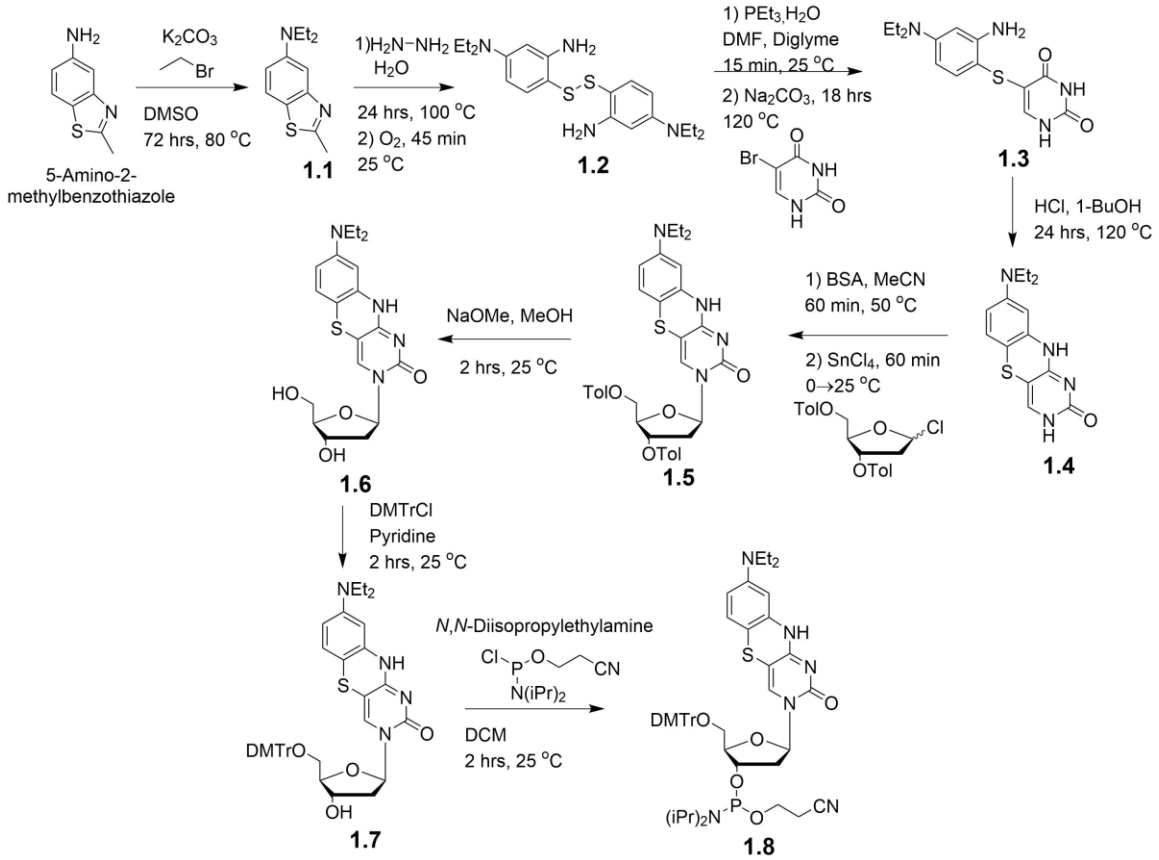


Figure 1.8. Common incorporation strategies of nucleoside analogues into nucleic acids. Solid-phase DNA synthesis utilizes nucleoside phosphoramidites while polymerase-mediated incorporation involves nucleoside triphosphates.

Our lab recently published an optimized synthetic procedure that was developed to efficiently generate ^{DEA}tC-labeled oligonucleotides with improved yields and purity, compared to the original 2017 procedure,[59] of intermediates leading to the ^{DEA}tC phosphoramidite (**Scheme1.1**).[70] The synthesis begins with the commercially available starting material 5-amino-2-methylbenzothiazole, which is first doubly ethylated at the amino group to provide the diethylamino derivative **1.1**. The thiazole ring is then opened using hydrazine as a nucleophile, and the resulting thiol is oxidized aerobically by exposure to the air, generating the disulfide **1.2**, which facilitates purification and handling. Formation of the disulfide occurs essentially as a dimerization reaction between two molecules of the preceding thiol. Reduction of the disulfide

for the following reaction is performed using triethylphosphine and subsequent conjugation to 5-bromouracil is performed in a one-pot procedure to give thioether **1.3**. Acid-catalyzed condensation completes the stable ^{DEA}tC nucleobase **1.4**. This mildly fluorescent nucleobase is glycosylated using silyl-Hilbert-Johnson conditions and commercially available Hoffer's chlorosugar to provide the protected nucleoside **1.5** as a mixture of α and β anomers,[71] which are easier to separate after toloyl (Tol) protecting group removal to afford **1.6**. Standard conditions for dimethoxytrityl-group (DMTr) installation and phosphoramidite synthesis[72,73] provide **1.7** and **1.8**, respectively in two separate reactions. Purifying the final product by liquid chromatography and precipitation in hexane efficiently removes impurities, including residual P(III) species that can inhibit solid-phase DNA synthesis.



Scheme 1.1. Optimized synthesis of ^{DEA}tC phosphoramidite. Product **8** is used as substrate in solid-phase DNA synthesis, generating sequence-specific oligonucleotides.[70]

An alternative to solid-phase DNA synthesis with phosphoramidites is enzymatic synthesis of DNA/RNA, which can generate much longer sequences and requires cheaper and fewer resources to perform. Published work from the Szostak group focuses on the template-based synthesis of oligonucleotides in the absence of enzymes by using imidazole-activated nucleotides,[74] which carry implications in the evolution of nucleic acid replication. The essential components for enzyme-mediated fluorescent oligonucleotide synthesis include a DNA or RNA polymerase, a template sequence usually hybridized to a primer sequence, and analogue substrates in the form of nucleoside triphosphates. Fluorophore incorporation can be performed in the absence of a template sequence if an appropriate polymerase, such as deoxynucleotidyl transferase, performs synthesis.[75] Polymerase-mediated incorporation proceeds in aqueous buffer, making the reaction much more forgiving than solid-phase reactions that require inert atmosphere conditions. Producing fluorescently-labeled oligonucleotides by polymerases presents two primary challenges: non-specific labeling and limitations imposed by the enzyme active site. Nucleobase analogues with ring sizes minimally perturbing the structural constraints of the polymerase active site, and retaining Watson-Crick base-pairing reliably with a natural base on the template strand are likely to provide somewhat predictable results, though still often with lower activity or fidelity than with natural nucleotides.[52] If the nucleoside analogue demonstrates promiscuous base-pairing, then sequence integrity can be compromised. To endow certain fluorescent properties, such as extended conjugation for red-shifted absorption and emission profiles or polar substituents for environmental sensitivity, the molecular size increases, which can decrease compatibility with enzyme binding pockets. Parent tC nucleoside triphosphates was recently demonstrated to efficiently substitute natural cytidine triphosphates as substrates in the RNA-, DNA-dependent modes of retroviral reverse transcription, as well as

template base for dGTP insertion, with reliable fidelity and retained polymerase function.[76] Published findings detailing this work are presented in a subsequent chapter. Future endeavors studying nucleic acid metabolism in more realistic contexts, for instance live cells or tissue, will require both efficient delivery and incorporation strategies.

1.6 Concluding remarks and the work of this dissertation

There have been major advances studying the complexities of nucleic acid structure and metabolism using fluorescent nucleoside analogues, however *in vitro* studies are limited to demonstrating what advances are possible and novel discoveries will rely on live-cell or *in vitro* endeavors. Within these more complicated biological environments, minimally perturbing analogues are expected to yield optimal results while preserving as much native function as possible. Two prominent challenge for chemists to solve include increasing the brightness of nucleobase fluorophores for greater sensitivity and developing analogues with more red-shifted absorption/emission maxima. In **Table 1.2** the absorbance/emission maxima for these isomorphic and minimally ring-extended analogues span 300-500 nm, as such analogues with red-shifted spectral profiles typically require extending their π-conjugation, though this risks compromising biological compatibility. A greater diversity of colors allows for multiplexing while longer wavelength, lower energy photons are less prone to photo bleaching and scattering, thereby providing greater sample viability and resolution.[77,78] Developing novel fluorescent nucleosides has mostly involved incorporating structural motifs of existing bright fluorophores without knowing the photophysical properties beforehand. A more rational-design approach would optimizes the development of new molecules to expedite their uses in relevant biological applications. Determining an analogues structure-function relationship and mechanisms underpinning their fluorescent properties in DNA is critical for synthetic chemists. In this

dissertation, the structure-function relationship of ^{DEA}tC is assessed to understand how the analogue's molecular structure affects the observed fluorescent properties and to identify possible biotechnological applications the analogue can be used for. Subsequent chapters describe the tC derivatives' biochemical compatibility with viral reverse transcriptase as well as delivering the fluorescent tC nucleosides into living cells as useful tools and methodology to study nucleic acid metabolism and replication. The current diverse array of fluorescent analogues provides an encouraging outlook into the various aspects of nucleic acid metabolism that can be explored.

1.7 References

(1) Crick, F. H. C.; Watson, J. D.; Bragg, W. L. *Proc. R. Soc. London. Ser. A. Math. Phys. Sci.* **1954**, *223* (1152), 80–96.

(2) Dickerson, R. E.; Drew, H. R.; Conner, B. N.; Wing, R. M.; Fratini, A. V; Kopka, M. L. *Science (80-.).* **1982**, *216* (4545), 475 LP – 485.

(3) Franklin, R. E.; Gosling, R. G. *Nature* **1953**, *172* (4369), 156–157.

(4) Hoogsteen, K. *Acta Crystallogr.* **1963**, *16*, 907–916.

(5) Spiegel, J.; Adhikari, S.; Balasubramanian, S. *Trends Chem.* **2020**, *2* (2), 123–136.

(6) Yao, R.-W.; Wang, Y.; Chen, L.-L. *Nat. Cell Biol.* **2019**, *21* (5), 542–551.

(7) Zeraati, M.; Langley, D. B.; Schofield, P.; Moye, A. L.; Rouet, R.; Hughes, W. E.; Bryan, T. M.; Dinger, M. E.; Christ, D. *Nat. Chem.* **2018**, *10* (6), 631–637.

(8) Arnott, S.; Hukins, D. W. L. *Biochem. Biophys. Res. Commun.* **1972**, *47* (6), 1504–1509.

(9) Zubatiuk, T.; Kukuev, M. A.; Korolyova, A. S.; Gorb, L.; Nyporko, A.; Hovorun, D.; Leszczynski, J. *J. Phys. Chem. B* **2015**, *119* (40), 12741–12749.

(10) Liebl, K.; Zacharias, M. *Nucleic Acids Res.* **2018**, *47* (3), 1132–1140.

(11) El Hassan, M. A.; Calladine, C. R. *Philos. Trans. Math. Phys. Eng. Sci.* **1997**, *355* (1722), 43–100.

(12) Imeddourene, A. Ben; Xu, X.; Zargarian, L.; Oguey, C.; Foloppe, N.; Mauffret, O.; Hartmann, B. *Nucleic Acids Res.* **2016**, *44* (7), 3432–3447.

(13) Kypr, J.; Kejnovska, I.; Renciuk, D.; Vorlıckova, M. *Nucleic Acids Res.* **2009**, *37* (6), 1713–1725.

(14) Malenkov, G.; Minchenkova, L.; Minyat, E.; Schyolkina, A.; Ivanov, V. *FEBS Lett.* **1975**, *51* (1–2), 38–42.

(15) Beard, W. A.; Wilson, S. H. *Biochemistry* **2014**, *53* (17), 2768–2780.

(16) Waters, J. T.; Lu, X.-J.; Galindo-Murillo, R.; Gumbart, J. C.; Kim, H. D.; Cheatham 3rd, T. E.; Harvey, S. C. *J. Phys. Chem. B* **2016**, *120* (33), 8449–8456.

(17) DiMaio, F.; Yu, X.; Rensen, E.; Krupovic, M.; Prangishvili, D.; Egelman, E. H. *Science (80-.).* **2015**, *348* (6237), 914 LP – 917.

(18) Svozil, D.; Kalina, J.; Omelka, M.; Schneider, B. *Nucleic Acids Res.* **2008**, *36* (11), 3690–3706.

(19) Henderson, E.; Hardin, C. C.; Walk, S. K.; Tinoco Jr., I.; Blackburn, E. H. *Cell* **1987**, *51* (6), 899–908.

(20) Burge, S.; Parkinson, G. N.; Hazel, P.; Todd, A. K.; Neidle, S. *Nucleic Acids Res.* **2006**, *34* (19), 5402–5415.

(21) Rhodes, D.; Lipps, H. J. *Nucleic Acids Res.* **2015**, *43* (18), 8627–8637.

(22) Hänsel-Hertsch, R.; Beraldi, D.; Lensing, S. V; Marsico, G.; Zyner, K.; Parry, A.; Di Antonio, M.; Pike, J.; Kimura, H.; Narita, M.; Tannahill, D.; Balasubramanian, S. *Nat. Genet.* **2016**, *48* (10), 1267–1272.

(23) Gehring, K.; Leroy, J.-L.; Guéron, M. *Nature* **1993**, *363* (6429), 561–565.

(24) Wright, E. P.; Huppert, J. L.; Waller, Z. A. E. *Nucleic Acids Res.* **2017**, *45* (6), 2951–2959.

(25) Brooks, T. A.; Kendrick, S.; Hurley, L. *FEBS J.* **2010**, *277* (17), 3459–3469.

(26) Garavís, M.; Escaja, N.; Gabelica, V.; Villasante, A.; González, C. *Chem. – A Eur. J.* **2015**, *21* (27), 9816–9824.

(27) Jin, K. S.; Shin, S. R.; Ahn, B.; Rho, Y.; Kim, S. J.; Ree, M. *J. Phys. Chem. B* **2009**, *113* (7), 1852–1856.

(28) Abou Assi, H.; Garavís, M.; González, C.; Damha, M. J. *Nucleic Acids Res.* **2018**, *46* (16), 8038–8056.

(29) Brázda, V.; Hároníková, L.; Liao, J. C. C.; Fojta, M. *Int. J. Mol. Sci.* **2014**, *15* (10), 17493–17517.

(30) Izumi, H.; Funa, K. *Cells* **2019**, *8* (5), 390.

(31) Soldatenkov, V. A.; Vetcher, A. A.; Duka, T.; Ladame, S. *ACS Chem. Biol.* **2008**, *3* (4), 214–219.

(32) Niu, K.; Zhang, X.; Deng, H.; Wu, F.; Ren, Y.; Xiang, H.; Zheng, S.; Liu, L.; Huang, L.; Zeng, B.; Li, S.; Xia, Q.; Song, Q.; Palli, S. R.; Feng, Q. *Nucleic Acids Res.* **2017**, *46* (4), 1710–1723.

(33) *Fluorescent Analogs of Biomolecular Building Blocks: Design and Applications*; Wilhelmsson, L. M., Tor, Y., Eds.; John Wiley & Sons, Inc.: Hoboken, NJ, 2016.

(34) Middleton, C. T.; de La Harpe, K.; Su, C.; Law, Y. K.; Crespo-Hernández, C. E.; Kohler, B. *Annu. Rev. Phys. Chem.* **2009**, *60* (1), 217–239.

(35) Michel, B. Y.; Dziuba, D.; Benhida, R.; Demchenko, A. P.; Burger, A. *Frontiers in Chemistry* . 2020, p 112.

(36) Williams, A. T. R.; Winfield, S. A.; Miller, J. N. *Analyst* **1983**, *108*, 1067–1071.

(37) Berezin, M. Y.; Achilefu, S. *Chem. Rev.* **2010**, *110* (5), 2641–2684.

(38) Lakowicz, J. R. *Principles of Fluorescence Spectroscopy*, 3rd ed.; Springer: New York, NY, 2006.

(39) Wilhelmsson, L. M.; Sandin, P.; Holmén, A.; Albinsson, B.; Lincoln, P.; Nordén, B. *J. Phys. Chem. B* **2003**, *107* (34), 9094–9101.

(40) Biebricher, A. S.; Heller, I.; Roijmans, R. F. H.; Hoekstra, T. P.; Peterman, E. J. G.; Wuite, G. J. L. *Nat. Commun.* **2015**, *6*, 7304.

(41) Sinkeldam, R. W.; Greco, N. J.; Tor, Y. *Chem. Rev.* **2010**, *110* (5), 2579–2619.

(42) Gradinaru, C. C.; Marushchak, D. O.; Samim, M.; Krull, U. J. *Analyst* **2010**, *135* (3), 452–459.

(43) Pawar, M. G.; Srivatsan, S. G. *J. Phys. Chem. B* **2013**, *117* (46), 14273–14282.

(44) Sandin, P.; Stengel, G.; Ljungdahl, T.; Borjesson, K.; Macao, B.; Wilhelmsson, L. M. *Nucleic Acids Res.* **2009**, *37* (12), 3924–3933.

(45) Stryer, L.; Haugland, R. P. *Proc. Natl. Acad. Sci. U. S. A.* **1967**, *58* (2), 719–726.

(46) Lilley, D. M. J.; Wilson, T. J. *Curr. Opin. Chem. Biol.* **2000**, *4* (5), 507–517.

(47) Sekar, R. B.; Periasamy, A. *J. Cell Biol.* **2003**, *160* (5), 629–633.

(48) Iqbal, A.; Arslan, S.; Okumus, B.; Wilson, T. J.; Giraud, G.; Norman, D. G.; Ha, T.; Lilley, D. M. J. *Proc. Natl. Acad. Sci.* **2008**, *105* (32), 11176 LP – 11181.

(49) Teppang, K. L.; Lee, R. W.; Burns, D. D.; Turner, M. B.; Lokensgard, M. E.; Cooksy, A. L.; Purse, B. W. *Chem. - A Eur. J.* **2019**, *25*, 1249–1259.

(50) Dierckx, A.; Miannay, F.-A.; Ben Gaied, N.; Preus, S.; Björck, M.; Brown, T.; Wilhelmsson, L. M. *Chem. – A Eur. J.* **2012**, *18* (19), 5987–5997.

(51) Mata, G.; Schmidt, O. P.; Luedtke, N. W. *Chem. Commun.* **2016**, *52* (25), 4718–4721.

(52) McCoy, L. S.; Shin, D.; Tor, Y. *J. Am. Chem. Soc.* **2014**, *136* (43), 15176–15184.

(53) Shin, D.; Sinkeldam, R. W.; Tor, Y. *J. Am. Chem. Soc.* **2011**, *133* (38), 14912–14915.

(54) Seo, Y. J.; Ryu, J. H.; Kim, B. H. *Org. Lett.* **2005**, *7* (22), 4931–4933.

(55) Füchtbauer, A. F.; Wranne, M. S.; Sarangamath, S.; Bood, M.; El-Sagheer, A. H.; Brown,

T.; Gradén, H.; Grøtli, M.; Wilhelmsson, L. M. *Chempluschem* **2020**, *85* (2), 319–326.

(56) Schmidt, O. P.; Mata, G.; Luedtke, N. W. *J. Am. Chem. Soc.* **2016**, *138* (44), 14733–14739.

(57) Lin, K.-Y. Y.; Jones, R. J.; Matteucci, M. *J. Am. Chem. Soc.* **1995**, *117* (13), 3873–3874.

(58) Rodgers, B. J.; Elsharif, N. A.; Vashisht, N.; Mingus, M. M.; Mulvahill, M. A.; Stengel, G.; Kuchta, R. D.; Purse, B. W. *Chem. - A Eur. J.* **2014**, *20* (7), 2010–2015.

(59) Burns, D. D.; Teppang, K. L.; Lee, R. W.; Lokensgard, M. E.; Purse, B. W. *J. Am. Chem. Soc.* **2017**, *139* (4), 1372–1375.

(60) Fontenete, S.; Guimarães, N.; Wengel, J.; Azevedo, N. F. *Crit. Rev. Biotechnol.* **2015**, *8551* (December), 1–12.

(61) Verma, S.; Eckstein, F. *Annu. Rev. Biochem.* **1998**, *67* (1), 99–134.

(62) Froehler, B. C.; Ng, P. G.; Matteucci, M. D. *Nucleic Acids Res.* **1986**, *14* (13), 5399–5407.

(63) Guzaev, A. P.; Manoharan, M. *J. Am. Chem. Soc.* **2003**, *125* (9), 2380–2381.

(64) Wada, T.; Sato, Y.; Honda, F.; Kawahara, S.; Sekine, M. *J. Am. Chem. Soc.* **1997**, *119* (52), 12710–12721.

(65) Garegg, P. J.; Lindh, I.; Regberg, T.; Stawinski, J.; Strömberg, R.; Henrichson, C. *Tetrahedron Lett.* **1986**, *27* (34), 4051–4054.

(66) Hughes, R. A.; Ellington, A. D. *Cold Spring Harb. Perspect. Biol.* **2017**, *9* (1).

(67) Kosuri, S.; Church, G. M. *Nat. Methods* **2014**, *11*, 499.

(68) Kool, E. T. *Annu. Rev. Biochem.* **2002**, *71* (1), 191–219.

(69) Xu, W.; Chan, K. M.; Kool, E. T. *Nat. Chem.* **2017**, *9* (11), 1043–1055.

(70) Turner, M. B.; Anderson, B. A.; Samaan, G. N.; Coste, M.; Burns, D. D.; Purse, B. W. *Curr. Protoc. Nucleic Acid Chem.* **2018**, e59.

(71) Niedballa, U.; Vorbrueggen, H. *J. Org. Chem.* **1974**, *39* (25), 3654–3660.

(72) Beaucage, S. L.; Caruthers, M. H. *Tetrahedron Lett.* **1981**, *22* (20), 1859–1862.

(73) Sinha, N. D.; Biernat, J.; Köster, H. *Tetrahedron Lett.* **1983**, *24* (52), 5843–5846.

(74) Walton, T.; Zhang, W.; Li, L.; Tam, C. P.; Szostak, J. W. *Angew. Chemie Int. Ed.* **2019**, *58* (32), 10812–10819.

(75) Hollenstein, M.; Wojciechowski, F.; Leumann, C. J. *Bioorg. Med. Chem. Lett.* **2012**, *22* (13), 4428–4430.

(76) Turner, M. B.; Purse, B. W. *Chempluschem* **2020**, *85* (5), 855–865.

(77) Dempsey, G. T.; Vaughan, J. C.; Chen, K. H.; Bates, M.; Zhuang, X. *Nat. Methods* **2011**, *8* (12), 1027–1036.

(78) Lukinavičius, G.; Umezawa, K.; Olivier, N.; Honigmann, A.; Yang, G.; Plass, T.; Mueller, V.; Reymond, L.; Corrêa Jr, I. R.; Luo, Z.-G.; Schultz, C.; Lemke, E. A.; Heppenstall, P.; Eggeling, C.; Manley, S.; Johnsson, K. *Nat. Chem.* **2013**, *5* (2), 132–139.

Chapter 2

Spectroscopic Analysis of the Fluorescent Turn-on Nucleoside 8-Diethylamino-tC in DNA-RNA Hybrids to Identify Mechanism of Emissive Response

2.1 Introduction

Fluorescent nucleoside analogues are an increasingly informative tool for studying nucleic acid metabolism, structural dynamics, and interactions with other biomolecules. Inherently fluorescent nucleobase replacements Watson-Crick base-pairing capability accomplish this while minimally perturbing natural biological structure and function.[1] Conversely, conventional fluorophores conjugated to DNA or RNA can interfere with recognition among other adverse effects, and do not report directly on local base-pair surroundings or dynamics.[2] Since the discovery of 2-Aminopurine (2AP) fluorescence in 1969,[3] more than 100 fluorescent nucleoside analogues have been reported.[4] Intrinsically fluorescent nucleobase analogues have been used to report on local structure, conformation, and dynamics of nucleic acids, sequence (mis)matches, enzyme-mediated nucleobase modifications, and changes in local polarity or pH.[5–9] By expanding on the intrinsically fluorescent tC molecular scaffold with a series of chemical modifications, we have developed novel derivatives with environmentally sensitive fluorescent properties for enhanced probative capabilities of DNA structure and metabolism.[5] For example, the recently published 8-diethylamino tC (^{DEA}tC) derivative is nearly non-emissive as a free nucleoside in aqueous solution but experiences up to a 20-fold fluorescence turn-on effect when base-stacked in double-stranded DNA (dsDNA) and base-paired with guanine (**Figure 2.1**),[10] thereby offering a potentially powerful tool for detecting hybridization events with G-dependent polymorphisms.

The mechanism underlying the fluorescence turn-on response or why fluorescence is quenched for the free nucleoside had not been firmly established in our prior investigations. Understanding the molecular quenching mechanism for ^{DEA}tC is an important aspect to designing novel probe structures with improved spectral properties and selecting the relevant applications that analogues are most appropriate for. In this present study, we measure the photophysical properties of 8-diethylamino tC, the brightest minimally perturbing nucleoside turn-on probe currently known, in 10-mer DNA-RNA hybrids to further understand how DNA helical properties modulate ^{DEA}tC fluorescence. DNA-RNA and dsRNA helices adopt global A-form DNA conformation which differs from B-form dsDNA helices in regards to average biophysical parameters such as twist, inclination, flexibility, groove topology, and base-pair spacing among others.[11–13] Steady-state and time-resolved fluorescence measurements inform upon the fluorescence and quenching kinetics of the analogue base-stacked in A-form DNA-RNA hybrids while circular dichroism (CD) spectroscopy is used to account for helical structure and thermodynamic stability.

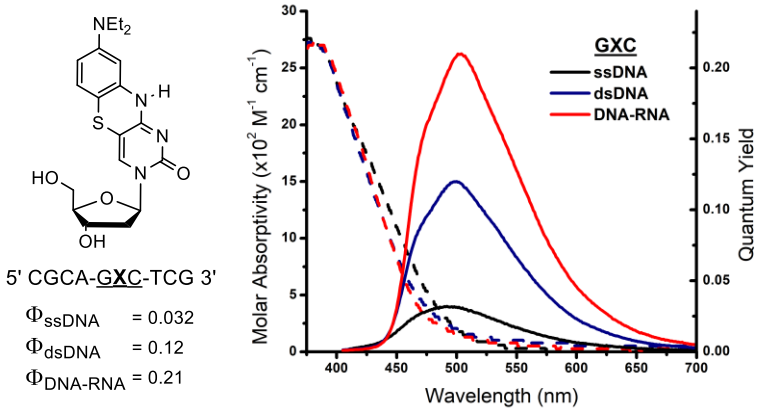


Figure 2.1. Structure and fluorescence of ^{DEA}tC nucleoside in DNA. Absorption and emission spectra are shown for a 10-mer DNA strand containing an internal 5' GXC 3' sequence, where X is ^{DEA}tC. Fluorescence turn-on is observed when analogue is stacked in duplex with a DNA or RNA complement.

During a fluorophore's excited-state lifetime there is an increased propensity to participate in undesirable acid-base reactions or redox chemistry.[14–16] Moreover, adding heteroatoms or functional groups that can function as proton donors or acceptors onto the molecular scaffolds of fluorescent nucleobase analogues often confers greater sensitivity to solvent polarity and pH.[17] For excited-state acid-base reactions, transfer of a proton during a fluorophores excited-state can lead to non-emissive transitions and is known as excited-state proton transfer (ESPT). The diethylamino group on ^{DEA}tC could serve as a possible proton acceptor site during the molecule's excited-state. When the bright fluorescence turn-on was first reported, one proposal for the observed behavior was that when ^{DEA}tC was base-stacked in dsDNA, insulation within the DNA duplex shielded the fluorophore's excited state from ESPT.[10] Quenched fluorescence resulting from ESPT has been investigated as an influential relaxation mechanism in other environmentally sensitive fluorophores.[18–20] Quenching experiments involving Stern-Volmer analysis with halide anions and solvent isotope effects are used in this study to further evaluate the extent to which A-form and B-form helices may insulate ^{DEA}tC from external quenchers.

2.2 Greater fluorescence turn-on in DNA-RNA duplexes

We first measured the steady-state fluorescence of ^{DEA}tC in DNA-RNA duplexes with varying adjacent bases using 10-mer oligonucleotides. The selection of sequences was based on previous measurements to enable direct comparisons with previous measurements on dsDNA duplexes.[10] Consistently greater ^{DEA}tC fluorescence intensity is observed in the DNA-RNA duplexes with GXC remaining as the highest performing sequence and CXT as the lowest performing (**Figure 2.2**). GXC exhibited the largest Φ_{em} of 0.22, a 6.9-fold increase with respect to the GXC 10-mer as ssDNA, while most other sequences had Φ_{em} values within 0.10 – 0.17

(**Table 2.1**). CXT exhibited the lowest of 0.048, but was still 2.4-fold brighter that the corresponding ssDNA probe strand. The AXA sequence experienced the largest fluorescence enhancement, near 14-fold brighter, when comparing ssDNA and DNA-RNA, while CXA showed the largest turn-on improvement (7.6-fold enhancement) when RNA substituted DNA as the complementary strand (**Figure 2.2**). Brightness [$B = \varepsilon\Phi$] is calculated using ε_{395} = 2,700 M^{-1} cm^{-1} though the molar absorptivity of DNA bases usually decreases 25 – 40% in oligonucleotides compared to free nucleoside monomers.[21] The Stokes shifts among the sequences are largely consistent, ranging from 83 – 110 nm, which corresponds to wavenumbers ($\tilde{\nu}$) of 4,100 – 5,600 cm^{-1}. The mean of the absorption and emission wavenumbers yields the excitation energy [$E = hc\tilde{\nu}$] averaging 2.78 eV, which is slightly lower than the average for parent tC, 2.9 eV, in DNA.[22] The excitation energy of ^{DEA}tC in dsDNA averages to 2.76 eV, and 2.83 eV as a free nucleoside. Similar excitation energies when stacked in a dsDNA or DNA-RNA duplex indicate that the energy gap between the ground (S_0) and excited states (S_1) remains consistent for ^{DEA}tC nucleoside, ssDNA, dsDNA, and DNA-RNA data.

Table 2.1. Steady-State fluorescence measurements from ^{DEA}tC DNA-RNA duplexes.

DNA Sequence (5'… 3')	$\lambda_{max, abs}$ (nm)	$\lambda_{max, em}$ (nm)	Stokes Shift (nm)	Φ_{em} (±)	ΔT_m (±)[a]	Brightness (M^{-1} cm^{-1})[c]
CGCA-GXC-TCG	404	504	100	0.22 (7.4×10^{-3})	-7.05 (0.77)	594
CGCA-GXA-TCG	394	501	107	0.17 (1.4×10^{-2})	-2.83 (1.8)	459
CGCA-TXA-TCG	409	503	94	0.14 (2.4×10^{-2})	+0.61 (0.60)	378
CGCA-CXA-TCG	410	497	87	0.13 (5.2×10^{-2})	-3.39 (0.85)	351
CGCA-CXC-TCG	411	494	83	0.12 (3.9×10^{-2})	-6.92 (0.64)	324
CGCA-AXA-TCG	392	502	110	0.11 (1.6×10^{-2})	-1.16 (1.0)	297
CGCA-GXG-TCG	402	503	101	0.10 (2.7×10^{-4})	-5.43 (0.54)	270
CGCA-CXT-TCG	410	499	89	0.048 (6.9×10^{-4})	+1.49 (1.1)	130

Analogue is substituted 6 DNA bases from 5' end, denoted as X. *a)* DNA-RNA thermodynamic stability is calculated by subtracting T_m (°C) of duplexes with natural cytidine from T_m (°C) of duplexes substituting ^{DEA}tC at position X. *c)* Brightness is calculated using ε = 2,700 M^{-1} cm^{-1} at 395 nm.

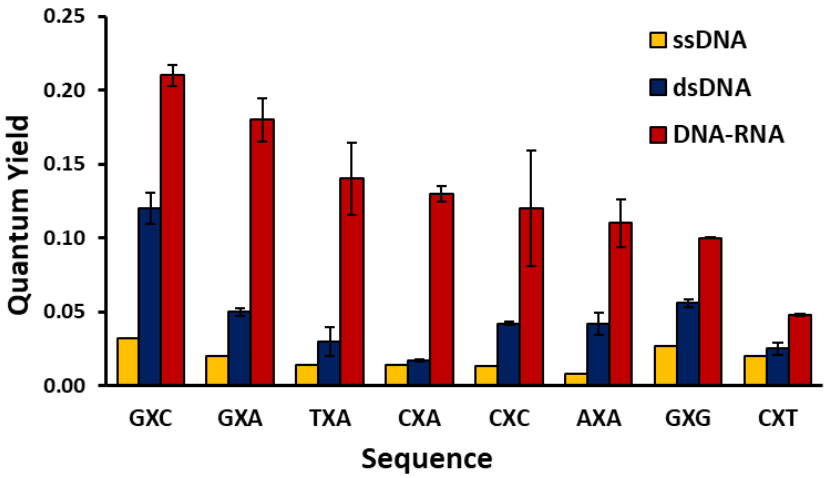


Figure 2.2. Comparing quantum yields of ^{DEA}tC oligonucleotides**.** 10-mer sequences with varying adjacent bases to ^{DEA}tC shown as 5'→3'. For CXA dsDNA, CXC dsDNA, GXG dsDNA, GXG DNA-RNA, and CXT DNA-RNA standard deviations were $<2.5x10^{-3}$.

Consistent with past work, ^{DEA}tC exhibits fluorescence turn-on in response to base-stacking and canonical Watson-Crick base-pairing with guanine in the complementary DNA or RNA strand. When the quantum yield was measured for a DNA-RNA ^{DEA}tC:A mismatch using the AXA sequence, the fluorescence was comparable to AXA ssDNA Φ_{em} revealing no significant turn-on response (**Figure 2.3A**). We performed a similar fidelity experiment with a DNA complement that included the non-canonical nucleoside inosine (**Figure 2.3B**). The natural nucleoside inosine contains a non-canonical purine nucleobase and arises from the enzymatic deamination of adenine as the most common post-transcriptional RNA modification in eukaryotes.[23–26] Inosine has a Watson-Crick hydrogen-bonding face resembling guanine and consequently may base-pair with cytosine.[27] I:C base-pairs differ from classic G:C base pairs in that hypoxanthine engages in two hydrogen bonds with cytosine, lacking an H-bond with the *O2* acceptor on cytosine (**Figure 2.3B**). Although inosine presents similar features to guanosine, we did not observe a significant fluorescence turn-on for a dsDNA ^{DEA}tC:I duplex. Engaging the three Watson-Crick H-bonding sites on ^{DEA}tC with a contraposing guanine is required for induced fluorescence turn-on.

2.3 Lifetime measurements and structural investigation

Time-resolved fluorescence spectroscopy was used and emission decays were plotted to measure the average excited-state lifetime $<\tau>$ following absorption (**Figure 2.S1** and **Table 2.2**). ^{DEA}tC exhibits very low fluorescence as a nucleoside in 1X PBS solution that results from a relatively short excited-state lifetime of 0.826 ns, which is still 10^2-10^3 longer than the lifetime of the canonical nucleobases.[28] Stacking ^{DEA}tC in DNA-RNA duplexes reduces solvent exposure and the average major lifetime $<\tau_1>$ extends nearly 10-fold longer. Unlike constitutively fluorescent parent tC, ^{DEA}tC has both a longer $<\tau>$ and an increased k_f when base-stacked.[22] Sequences with a 3' adenosine flanking ^{DEA}tC and CXT were observed to have a two-component decay ranging from 0.637-1.01 ns, as determined by exponential decay fittings requiring a second component. Since there is no correlation between $<\tau_2>$ and Φ_{em}, the shorter $<\tau_2>$ does not correspond to an emissive transition. Ranking the 5' neighbor's reduction potential (E_R) in the NXA sequences reveals that $<\tau_2>$ prolongs with decreasing E_R of the neighboring 5' nucleobase (**Figure 2.S2**). Guanine has the lowest reduction potential, -2.74 V, of the four canonical DNA bases (G<A<C<T),[29] and GXA experienced the longest $<\tau_2>$ while TXA accordingly exhibited the shortest. From the measured quantum yields and lifetimes, we calculated both the radiative decay rates (k_f) and the non-radiative rates (k_{nr}), which may occur via internal conversion or vibrational cooling.[30] The majority of tested sequences experienced a 10-fold difference in both rates when stacked in the DNA-RNA hybrid compared to the free nucleoside.

Table 2.2. Time-resolved fluorescence measurements and relaxation rates.

Sequence	Φ_{em}	$\langle\tau_1\rangle$ (ns)	$\langle\tau_2\rangle$ (ns)	k_f (x10^7 s^{-1})[a]	k_{nr} (x10^7 s^{-1})[b]
Nucleoside	0.006	0.826	--	0.73	120
GXC	0.22	10.0	--	2.20	7.80
GXA	0.17	9.12	1.01	1.86	9.10
TXA	0.14	8.97	0.655	1.56	9.59
CXA	0.13	9.80	0.854	1.33	8.88
CXC	0.12	9.38	--	1.28	9.38
AXA	0.11	9.30	0.906	1.18	9.57
GXG	0.10	9.47	--	1.06	9.50
CXT	0.048	9.12	0.637	0.56	11

Time-resolved measurements were performed using TCSPC and λ_{ex} = 371 nm. Sequences shown as internal triplet with flanking bases in 5' to 3' order. *a*) Radiative decay rate is calculated as Φ_{em} divided by fluorescence lifetime (τ_1). *b*) Non-radiative decay rates were calculated using the equation $k_{nr} = (k_f/\Phi_{em}) - k_f$.

Though ^{DEA}tC fluorescence responds to certain neighboring bases similarly in dsDNA and DNA-RNA duplexes, the brightness is considerably greater in the latter. We subsequently investigated possible factors that might rationalize why an RNA complement induces a greater fluorescence response than complementary DNA. DNA-RNA hybrids typically adopt A-form DNA conformation, similarly to dsRNA, which are structurally distinct from B-form dsDNA.[12] CD spectroscopy revealed that all of the ^{DEA}tC-containing DNA sequences adopted a global A-form structure when paired to their complementary 10-mer RNA (**Figure 2.S3**). These comparisons were performed by examining the CD spectral differences between sequences containing ^{DEA}tC to those containing natural cytosine and confirming the presence of a sharp transition at 210 nm.[31] Six of the eight sequences incurred a thermodynamic penalty when comparing the melting temperatures (T_m) of duplexes with analogue versus natural cytosine (**Table 2.1**, **Table 2.S1**). When the duplex contained ^{DEA}tC the loss of stability ranged from 1.16 – 7.05 °C, with GXC experiencing the greatest thermodynamic loss (**Figure 2.S4**). Sequences

AXA and CXT with ^{DEA}tC exhibited increased thermodynamic stability, although to a modest extent.

2.4 Evaluating extent of helical shielding from external quenching mechanisms

We initially hypothesized that the different duplex structure may enhance the turn-on effect by better insulating the analogue from external quenchers. Differences between A-form and B-form DNA include A-form having a greater tilt of the base-pair plane with respect to the central helical axis, a deeper major groove, and approximately 11 base pairs per turn instead of 10 in B-form. The increased overlap of the bases and narrower major groove were expected to endow greater shielding to the nucleobases from external quenchers in DNA-RNA. We began testing this hypothesis by performing Stern-Volmer quenching analysis using chloride and iodide, which are commonly used collisional quenchers.

Stern-Volmer quenching experiments were performed with three sequences and two halide ions in sodium phosphate buffer (for details, see the Supporting Information). Two halides, chloride and iodide, were selected and differ in their spherical ionic size as well as their quenching potency, which has been correlated to their ionization energies.[32] Chloride has ionic diameter of about 334 pm while iodide has an ionic diameter of approximately 412 pm and is a stronger quencher.[33–35] Oligonucleotide experiments were performed using DNA-RNA GXC, TXA, and CXA (**Figures 2.4** and **2.S5**) sequences. GXC was the brightest sequence observed while TXA and CXA showed the most pronounced fluorescence increase in A-form DNA-RNA duplexes relative to fluorescence in B-form dsDNA. Most samples exhibited a smaller K_{SV} in DNA-RNA duplexes, specifically 9-44% decrease with Cl^- or I^- (**Table 2.3**). Two exceptions, TXA using Cl^- and CXA using I^-, showed 5% and 10%, respectively, greater quenching in DNA-RNA relative to their respective dsDNA counterpart.

From the recorded fluorescence lifetimes and Stern-Volmer quenching coefficients, we calculated the bimolecular quenching constants (k_q), which describe the quencher's physical accessibility to the fluorophore and intrinsic quenching efficiency. In all measured cases except one, a 1.1–1.8-fold greater k_q was observed for the B-form dsDNA helices of the three sequences compared to their A-form DNA-RNA hybrid counterparts, suggesting that quenching of the excited-state ^{DEA}tC (denoted as ^{DEA}tC*) by halide anions was marginally mitigated by greater shielding of ^{DEA}tC in an A-form duplex. For TXA with sodium chloride, an 8% increased k_q was observed for dsDNA, suggesting that ^{DEA}tC* was better insulated in the B-form duplex. Nonetheless, the extent of attenuated quenching in A-form helices did not greatly differ from B-form for a given sequence, at least insufficiently to explain the doubling of brightness in DNA-RNA helices. Thus, we then surmised that increased insulation from external quenchers was an inadequate explanation for the higher ^{DEA}tC brightness in A-form DNA-RNA helices.

Table 2.3. Quenching efficiency measured by Stern-Volmer analysis

		NaCl		**NaI**	
Seq.	**Duplex**	**K_{SV} ($x10^{-3}$ M^{-1})**	**k_q ($x10^{8}$ M^{-1} s^{-1})[a]**	**K_{SV} ($x10^{-3}$ M^{-1})**	**k_q ($x10^{8}$ M^{-1} s^{-1})[a]**
GXC	dsDNA	3.22 ±0.13	3.99	3.22 ±0.23	3.99
	DNA-RNA	2.23 ±0.072	2.23	2.63 ±0.92	2.63
TXA	dsDNA	1.46 ±0.072	1.57	3.63 ±0.14	3.91
	DNA-RNA	1.54 ±0.076	1.72	3.28 ±0.38	3.66
CXA	dsDNA	2.02 ±0.090	2.74	2.93 ±0.18	3.98
	DNA-RNA	1.84 ±0.057	1.88	3.22 ±0.10	3.29

The Stern-Volmer quenching constant is graphically measured. *a*) The bimolecular quenching constant (k_q) is calculated from the relation $K_{SV} = k_q\tau$.

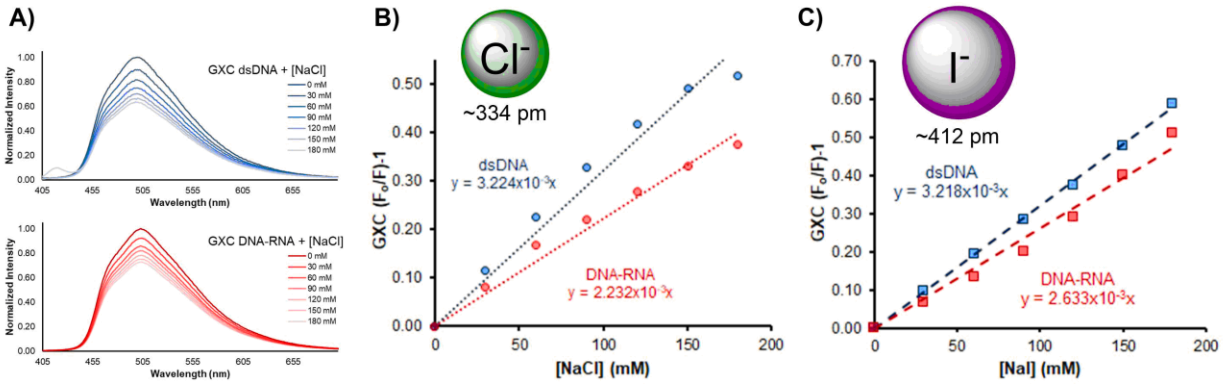


Figure 2.4. Stern-Volmer analysis of GXC duplexes. **A**) Fluorescence was measured as a function of [*Q*] and integration of emission spectra enabled graphical determination of Stern-Volmer quenching coefficients. **B)** Attenuated emission was recorded with varying [NaCl], and **C)** [NaI].

When ^{DEA}tC fluorescence was first studied in dsDNA,[10] the hypothesis explaining the fluorescence turn-on property was that the duplex attenuated excited-state proton transfer (ESPT) that ^{DEA}tC might otherwise experience as free monomer in aqueous solution. As a free nucleoside in buffered D_2O, ^{DEA}tC exhibited a Φ_{em} of 0.012, a 2-fold increase with respect to observations made in buffered H_2O solvent.[10] To evaluate if DNA-RNA duplexes attenuated ESPT better than B-form dsDNA helices, we measured time-resolved fluorescence spectra using time-correlated single photon counting (TCSPC) and compared Φ_{em} and τ of ^{DEA}tC-containing duplexes in protonated or deuterated buffer (experimental details are presented in the supporting information). As ESPT is a non-radiative relaxation pathway, non-radiative decay rates (k_{nr}) for ^{DEA}tC were calculated to quantify solvent isotope effects on fluorescence based on environment (i.e. as a free nucleoside, in B-form or A-form DNA).

Measuring the fluorescence of the ^{DEA}tC nucleoside in 1X PBS D_2O resulted in both a higher Φ_{em} and longer τ (time-resolved decays are shown in **Figure 2.5**). The nearly non-emissive ^{DEA}tC nucleoside monomer had a 2-fold higher Φ_{em} in deuterated buffer (**Table 2.4**), while the three sequences as either dsDNA or DNA-RNA had consistently longer lifetimes but inconsistent quantum yields. Though Φ_{em} remained higher for DNA-RNA duplexes, for CXA

and CXT the recorded Φ_{em} in D_2O was lower than Φ_{em} recorded in H_2O. For TXA the Φ_{em} values are near identical for dsDNA and DNA-RNA, while τ for TXA dsDNA is marginally higher than DNA-RNA in either solvent. Measuring the excited-state lifetimes revealed that the secondary lifetime observed in DNA-RNA and some dsDNA sequences with a 3' adenosine disappears in deuterated buffer.

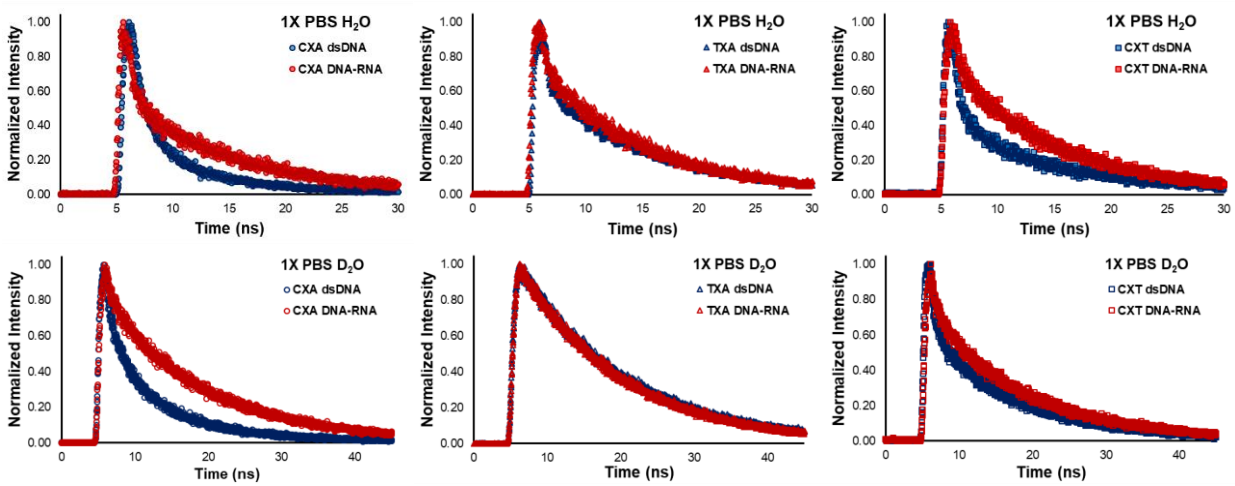


Figure 2.5. Solvent isotope effects on time-resolved ^{DEA}tC oligonucleotides fluorescence. Samples were measured in the indicated buffer with excitation at 371 nm using TCSPC.

From the non-radiative decay rates, we calculated the solvent kinetic isotope effect (KIE) of non-radiative decays. The free ^{DEA}tC nucleoside experiences a 6.3-fold higher k_{nr} in protonated buffer as compared with deuterated solvent, exhibiting a primary isotope effect with respect to proton transfer (**Table 2.4**). Therefore, ESPT is the predominant quenching mechanism for the nucleoside monomer. All three of the tested sequences exhibited normal KIE values, meaning non-radiative rates are lower when the probe is base-stacked in the double-stranded duplex solvated in D_2O as compared with H_2O. Sequences CXA and TXA exhibit <10% greater non-radiative decay in a dsDNA duplex compared to a DNA-RNA duplex, while CXT had 18% greater non-radiative decay in a DNA-RNA duplex. The isotopic effects on k_{nr}

between a dsDNA and DNA-RNA duplex of a given sequence have little to moderate variation, however all duplexes have a markedly decreased k_{nr} from the ^{DEA}tC nucleoside, indicating that the reduced rate of ESPT is the most important factor contributing to fluorescence turn-on.

Table 2.4. Solvent isotope effects on ^{DEA}tC oligonucleotide fluorescence

Sequence	Duplex	Φ_D	$\langle\tau_H\rangle$ (ns)	$\langle\tau_D\rangle$ (ns)	$k_{f\,H}$ ($x10^6$ s^{-1})[a]	$k_{f\,D}$ ($x10^6$ s^{-1})[a]	$k_{nr\,H}$ ($x10^7$ s^{-1})[b]	$k_{nr\,D}$ ($x10^7$ s^{-1})[b]	KIE
Nucleoside		0.012	0.826	5.09	7.26	2.36	120	19.4	6.2
CXA	dsDNA	0.017	7.37	10.8	2.31	1.57	13.3	9.07	1.5
	DNA-RNA	0.069	9.80	13.9	13.3	4.98	8.88	6.72	1.3
TXA	dsDNA	0.31	9.28	14.2	3.23	21.9	10.5	4.87	2.2
	DNA-RNA	0.32	8.97	13.5	15.6	23.7	9.59	5.03	1.9
CXT	dsDNA	0.017	11.2	12.5	2.23	1.37	8.71	7.90	1.1
	DNA-RNA	0.033	9.12	12.3	5.26	2.68	10.4	7.86	1.3

^{DEA}tC has Φ_H = 0.0060 as a nucleoside in 1X PBS. *a*) Radiative rates (k_f) are measured as Φ_{em}/τ for either protonated or deuterated PBS buffer. *b*) Non-radiative rates were calculated using $k_{nr} = (k_f/\Phi_{em}) - k_f$.

2.5 Acid-induced fluorescence of ^{DEA}tC nucleoside

As an environmentally sensitive fluorophore, ^{DEA}tC is responsive to pH changes, especially. The free ^{DEA}tC nucleoside exhibits a fluorescence turn-on at moderate acidic pH ranging from 2-4, in the absence of any base-stacking or base-pairing. For these experiments, a concentrated stock solution of ^{DEA}tC nucleoside in DMSO was added to solutions of varying H_2SO_4 concentrations providing a pH range of 1.50 to 6.00 in increments of 0.50 pH units. The fluorescence at each pH was measured with 120 μM nucleoside.

From pH 2-3.50 the absorbance both experienced a blue-shifted maximum to around 372 nm relative to the neutral pH $\lambda_{max\ abs}$ for ^{DEA}tC at 395 nm, and exhibited a greater molar absorptivity, or higher absorbance intensity (**Figure 2.6**). Within this pH range, the absorption spectra intersect at 410 nm, establishing an isosbestic point indicating a single preferential protonation site. The emission maxima for pH 2.00 – 7.00 (**Figure 2.6**, green) are focused at

493-496 nm, similar to λ_{em} recorded for ^{DEA}tC in dsDNA and DNA-RNA. A mild shouldering effect often observed near 475 nm disappears at the lower pH range. At pH 6-7.00 (**Figure 2.6**, dark red), the shouldering is visible but disappears as acidity increases. With a blue-shifted λ_{ex}, the excitation energy to transition from the ground to the excited-state increases from 2.72 eV at buffered neutral pH to 2.92 eV, indicating a greater energy gap between the HOMO and LUMO. At pH 3.50 the probe exhibited a Φ_{em} of 0.029, which resembles values for ssDNA emissions where ^{DEA}tC is partially solvent-exposed and situated between adjacent bases. Time-resolved measurements revealed a single-exponential decay with $\langle\tau\rangle$ = 4.50 ns, corresponding to a k_{nr} of 2.16×10^8 s^{-1}, intermediate between free nucleoside and DNA-RNA k_{nr} values (**Table 2.2**). Measuring the emission under basic conditions did not induce a turn-on effect and emission intensity was less than the intensity at neutral pH and ablated at pH >10.

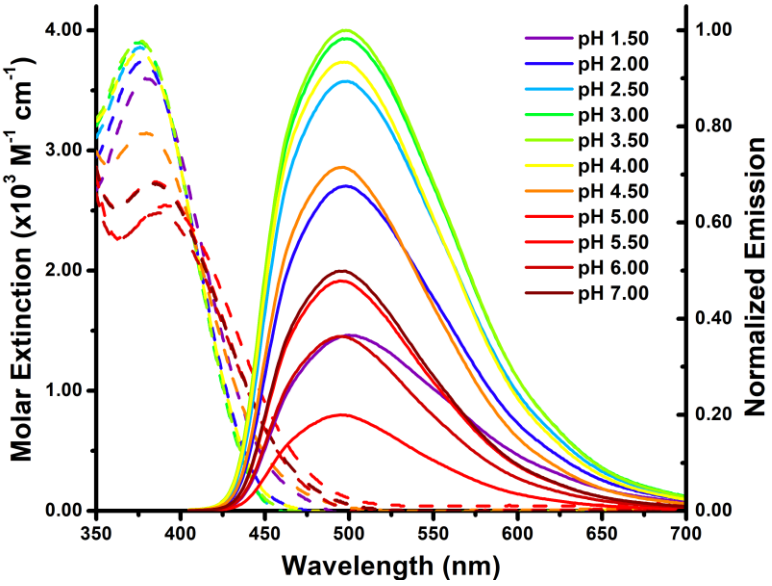


Figure 2.6. Induced fluorescence of ^{DEA}tC under acidic conditions. Recording the absorbance (dashed lines) and emission (solid lines) at λ_{ex} = 395 nm reveals fluorescence turn-on can be forcibly induced using an optimal pH 3.50. Measurements were performed with 120 µM ^{DEA}tC nucleoside and various concentrations of H_2SO_4.

2.6 Discussion

Initial steady-state fluorescence measurements of ^{DEA}tC contained in DNA-RNA heteroduplexes provided highly encouraging prospects as an RNA-targeting hybridization probe. While still exhibiting sensitivity to adjacent bases, sequence-based trends regarding Φ_{em} largely remained consistent. For instance, GXC was the highest performing sequence while CXT was the lowest performing (**Figure 2.2**), at least within the tested eight sequences of the possible 16 possibilities. TXA and CXA were two notable exceptions, meaning that the turn-on going from ssDNA to DNA-RNA was much greater than the corresponding ssDNA → dsDNA turn-on. Regardless, ^{DEA}tC clearly responds similarly to certain adjacent base pairs, whether stacked in B-form dsDNA or in A-form DNA-RNA helices. Coupled with remarkably improved brightness in DNA-RNA heteroduplexes, we hypothesized that helical structure (i.e. A-form vs. B-form) affected the magnitude of interactions between ^{DEA}tC and varying neighboring bases in the context of fluorescence turn-on.

As evidenced by similar λ_{ex} and excitation energies of ^{DEA}tC as free nucleoside, ssDNA, dsDNA, and DNA-RNA, the rationale underlying varying brightness among these structures stems from processes following excitation, namely, the extent of non-radiative decay. One relaxation mechanism observed in sequence-sensitive fluorescent nucleosides is photoinduced electron transfer (PET), which has been observed as a strong quenching factor for fluorescent nucleosides in oligonucleotides, especially when flanked by guanosine.[36–38] In double-stranded oligonucleotides, ^{DEA}tC fluorescence displays the opposite behavior, with flanking guanosines resulting in greater brightness. However, appearance of a second lifetime in NXA sequences with τ_2 durations inversely proportional to E_R of the 5' neighboring base (**Figure 2.S2**) may be suggestive of a PET as a possible non-radiative relaxation mechanism in these sequences. The

brightness however of the four NXA sequences displays no immediate trend relating to the reduction potential, as was noted in dsDNA helices, nor to τ_2 duration. CXT was the one non-NXA DNA-RNA sequence observed to have a second lifetime, though the fluorescence for the full set of possible sequences would need to be measured to verify if the trend persists. The ground-state of unsubstituted parent tC has an E_R of -1.1 V that raises to 1.8 V in the excited state, [22] whereas the $-NEt_2$ substituent on ^{DEA}tC confers higher electron density and likely lowers the ground-state reduction potential. Precise electrochemical measurements for the free nucleoside have not yet been performed.

One of the promising aspects of ^{DEA}tC is the probe's fidelity for increased fluorescence when base-paired with guanine in the double-stranded helix. The fluorescence turn-on effect is known not to occur when the complementary strand contraposes adenine or an abasic lesion across from ^{DEA}tC in dsDNA,[5,10] indicating that base-pairing with G is required for enhancing fluorescence. Without forming the typical three hydrogen-bonds, ^{DEA}tC may experience a more flexible base-pairing and base-stacking, enabling a higher opening rate for mismatched duplexes. Similar behavior was observed for the purine analogue quadracyclic adenosine (qAN) that normally experiences lower fluorescence when base-stacked but regained brightness in mismatched dsDNA.[39] In duplexes with ^{DEA}tC:A mismatches the analogue may adopt the imino tautomer of cytosine (**Figure 2.3A**), imitating the hydrogen bonding pattern of thymine, but with orbital configurations of the tautomeric ground and excited states expectedly distinct from the corresponding amino form and involving an apparent dark-state transition.

Forming the classic double-stranded structure of DNA arises from π–π dispersion energies ($E_{\pi\text{-}\pi}$) between stacked nucleobases as the major factor, and Hydrogen-bonding between base-pairs as an essential component.[40–42] In canonical B-form DNA the observed twist angle (ω)

is 36° offset from exact parallel alignment (ω = 0°), whereas in A-form DNA the twist between consecutive base-pairs is smaller at ω = 33°.[12] Increasing ω in B-form DNA was computationally shown to increase overall duplex thermodynamic stability, principally by reducing Pauli repulsion between neighboring bases.[43] In DNA-RNA A-form hybrids with greater base-pair overlap and reduced physical distance between base pairs, ^{DEA}tC should experience greater π–π interactions that may contribute to an increased fluorescence response. The average thermodynamic stability decreases from -2.2 °C in ^{DEA}tC dsDNA helices to -3.1 °C in ^{DEA}tC DNA-RNA hybrids, which is consistent with greater π–π interactions incurring a larger thermodynamic penalty as shown with computational modeling.[43] Parent tC is largely environmentally insensitive and confers increased duplex stability, however for ^{DEA}tC, an advantageous fluorescence turn-on response compromises double-stranded stability.

In addition to a more compressed helix, A-form DNA has deeper major grooves than that of classic B-form DNA.[12] While the major and minor grooves in B-form DNA have isometric depths, the narrower, deeper A-form DNA major grooves restrict access to solvent molecules and some ions. Unpublished NMR data of a GXC-containing 10-mer dsDNA complex demonstrate that the probe's diethylaniline portion projects into the major groove of the dsDNA helix, where the $–NEt_2$ substituent may freely rotate. The polarity in A-form major grooves is predicted to mimic the polarity of 20% H_2O in 1,4-dioxane, about half the estimated polarity of B-form.[44] Reduced interactions between the helix-insulated probe and bulk solvent were hypothesized to largely increase the magnitude of fluorescence, and indeed, we observed external quenching factors typically exerted decreased effects. [45] From Stern-Volmer quenching experiments, chloride and iodide were observed to access ^{DEA}tC in both dsDNA and DNA-RNA hybrids, although with some easier access to ^{DEA}tC in dsDNA. A drastic KIE of 6.3 observed for

^{DEA}tC nucleoside closely approaches the theoretical KIE limit of 6.5-7, and is compelling evidence that ESPT between bulk protonated solvent and ^{DEA}tC* is the predominant factor in generating fluorescence turn-on response. Oddly, the extent of attenuating ESPT were non-identical to Stern-Volmer experiments, however this discrepancy may be explained by the halide ions accessing the excited-state fluorophore in the DNA-RNA A-form minor groove, which is comparatively wider and shallower. Quenching by halide can occur through non-specific physical contacts with the molecule whereas solvent-mediated ESPT requires access to H-bond acceptors or donors on the molecule, and on ^{DEA}tC, those interactions are reduced by the DNA-RNA A-form topology.

pH-sensitive fluorescence has been observed for other reported nucleoside analogues that fluoresce more brightly or dimly in certain pH ranges.[46] The blue-shifted excitation and constant emission wavelength at lower pH indicates that acidity lowers the energy level of S_o but not S_1, due to the rapid process of excitation-relaxation relative to solvent rearrangement. Observing a lowered S_0 is reasonable under acidic conditions as protonation of the largely hydrophobic tricyclic ring structure may facilitate solubility in aqueous media. Smoothening of the emission spectra in acid pH is expected to occur from bright-state transitions of only one protonated excited-state species, as excited-states typically do not reach acid-base equilibrium except with highly acidic pH.[47] For example, a recent study on thG pH-based fluorescence identified no equilibrium between protonated and deprotonated forms of the excited-states.[48] The excited-state lifetime of ^{DEA}tC at pH 3.50 is shorter than the lifetimes measured in the thG 2020 study, further lowering the likelihood of ^{DEA}tC* acid-base equilibrium. The quantum yield of emission is comparable to measurements in single-stranded DNA, suggesting that the maximum emission that can be derived from the fluorophore in the absence of ordered base-stacking and base-

pairing is relatively low. Even with acid-induced fluorescent properties, factors contributing to k_{nr} are remain strongly quenching. As a pH-sensitive probe, ^{DEA}tC may be useful reporting on local acidic environments by recording the fluorescent excited-state lifetime or taking an excitation scan.

2.7 Conclusion

In this present work we sought to better understand how ^{DEA}tC's fluorescent turn-on response is modulated by its physical environment so as to better inform upon rationale-design approaches to developing novel fluorescent nucleoside analogues. An improved understanding of how physical environment and chemical structure relate to extent of turn-on also narrows biological contexts or applications in which the molecule's fluorescence are the most beneficial to studying nucleic acids metabolism and dynamics. Based on isotopic quenching kinetics, the causative factor leading to the large increase in brightness upon base stacking is attenuated ESPT, however variations in fluorescence depending on neighboring bases is likely more a product of electronic interactions between nucleobases in the duplex. Future computational modeling studies would be suitable to further examine the quantitative extent of these interactions. Nonetheless, ^{DEA}tC remains the brightest, minimally perturbing turn-on nucleoside probe and understanding how/why that turn-on occurs can inform upon on designing the synthesis of brighter turn-on probes. For majority of the DNA-RNA quantum yield values, ranging 0.10-0.22, the fluorescence is comparable to the constitutively emissive parent tC and tC^{O} molecules. While fluorescence *in situ* hybridization (FISH) probes conventionally involve appended fluorophores and quenchers to identify sequences based on preferential binding and not necessarily sequence complementarity,[49] alternative probes using ^{DEA}tC would be more sequence specific. Probe sequences using ^{DEA}tC fluorescence turn-on have multiple valuable

applications such as identifying single nucleotide polymorphisms (SNP), selectively detecting newly synthesized DNA in qPCR, and post-transcriptional RNA editing.[50–52]

As such, oligonucleotide probes with ^{DEA}tC offer encouraging alternatives for RNA-targeting hybridization probes considering single-stranded, ^{DEA}tC:A mismatches, and even ^{DEA}tC:H base-pairs have very low intrinsic fluorescence. Therefore, the turn-on is reliably restricted to ^{DEA}tC:G base-pair formations. Indeed, ^{DEA}tC oligonucleotide probes may function as informative biophysical tools for applications such as discriminating single nucleotide polymorphisms.

2.8 Supporting Data and Figures

Table 2.S1. Thermostability of ^{DEA}tC DNA-RNA oligonucleotides

Sequence	**Cytidine T_m (±)**	**^{DEA}tC T_m (±)**	**ΔT_m (±)**
GXC	58.36 (0.675)	51.31 (0.376)	-7.05 (0.77)
GXA	45.20 (1.74)	42.37 (0.220)	-2.83 (1.8)
AXA	40.53 (0.573)	41.14 (0.198)	+0.61 (0.60)
TXA	48.27 (0.740)	44.88 (0.417)	-3.39 (0.85)
CXA	56.72 (0.563)	49.80 (0.302)	-6.92 (0.64)
CXC	60.30 (0.693)	59.14 (0.731)	-1.16 (1.0)
GXG	52.03 (0.456)	46.60 (0.291)	-5.43 (0.54)
CXT	52.26 (0.927)	53.75 (0.672)	+1.49 (1.1)

Thermodynamic stability of duplexes was measured by recording CD signal at 270 nm as a function of temperature. Subtracting T_m (°C) of duplexes with natural cytidine from T_m (°C) of duplexes substituting ^{DEA}tC provided ΔT_m.

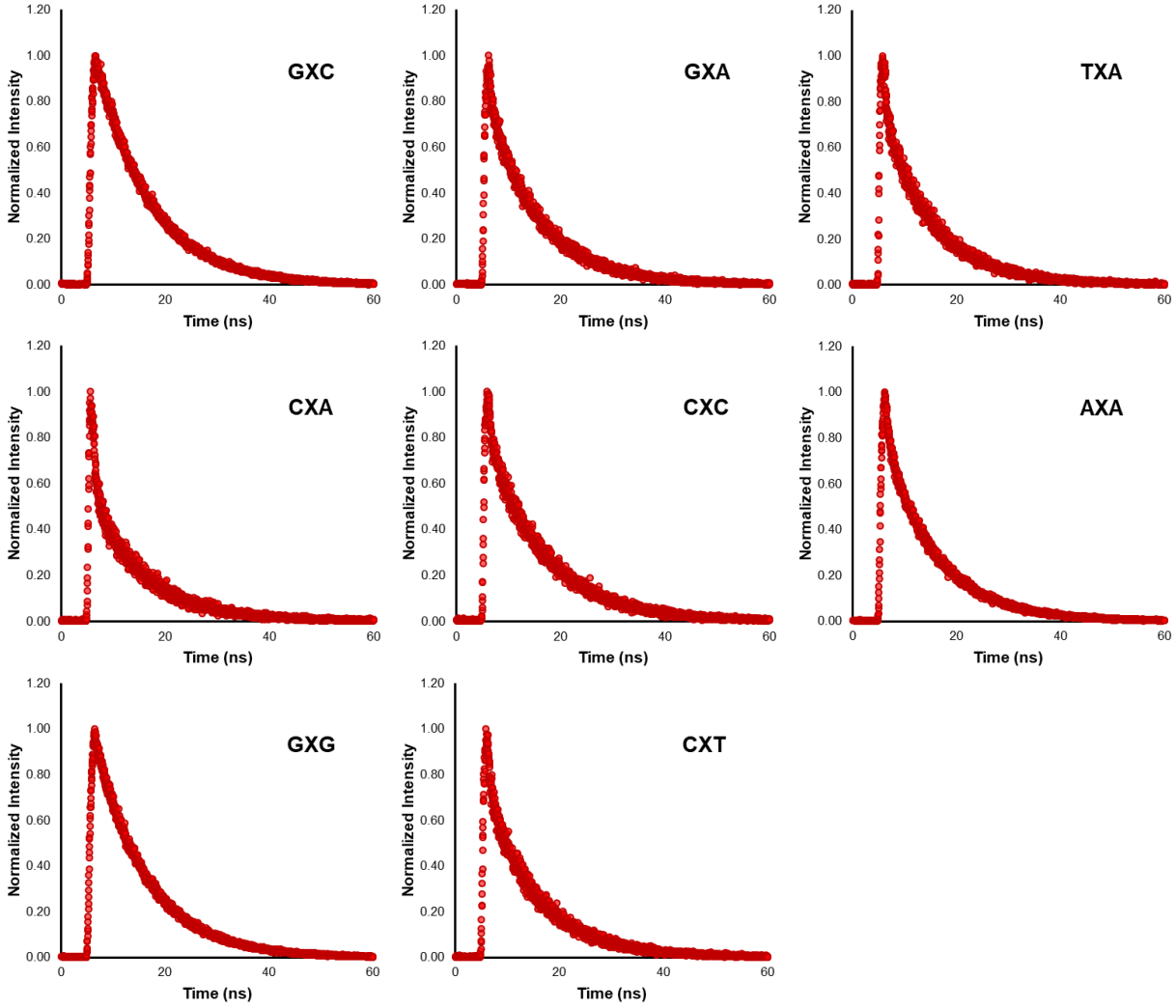


Figure 2.S1. Time-resolved emission decays of ^{DEA}tC DNA-RNA duplexes. Emission intensity within 250 – 650 nm is measured by TCSPC using an excitation of 371 nm.

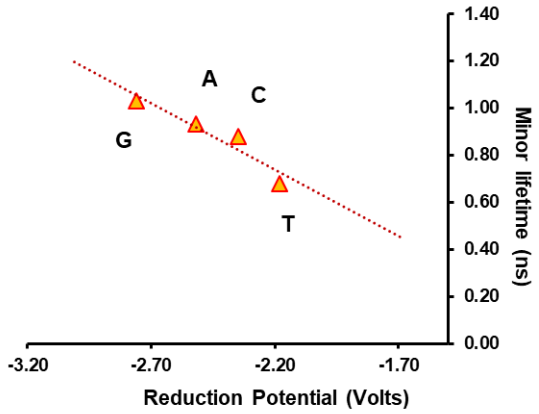


Figure 2.S2. Minor lifetime <τ2> of probe plotted against reduction potential. Average second-component <τ_2> of ^{DEA}tC in sequences with adjacent 3' adenosine correlate with reduction potential (E_R) of 5' neighbor. E_R are taken from published experimental recordings.[29]

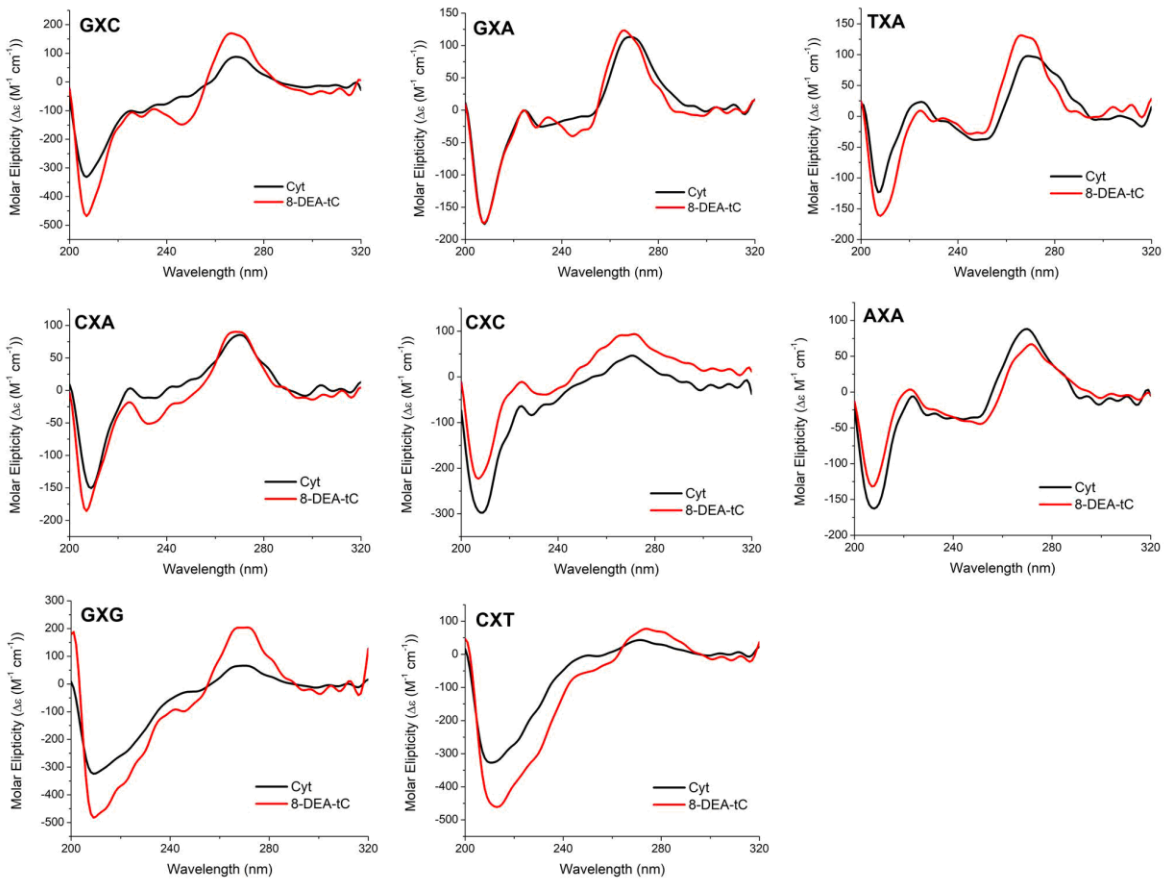


Figure 2.S3. Circular dichroism spectra and melting curves of ^{DEA}tC DNA-RNA sequences. Spectra were recorded at 25 °C with 2.5 μM sample duplex.

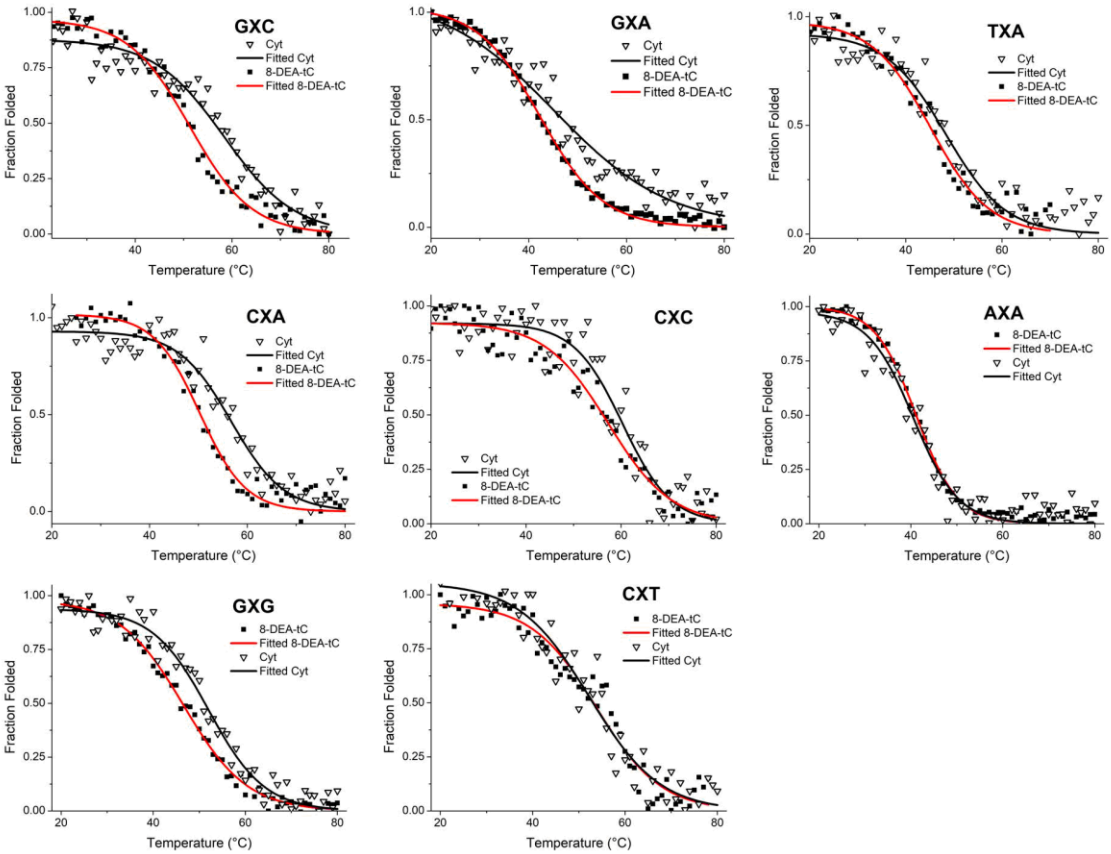


Figure 2.S4. Melting curves of ^{DEA}tC DNA-RNA sequences. Signal at 270 nm was recorded and data was fit to two-state model.

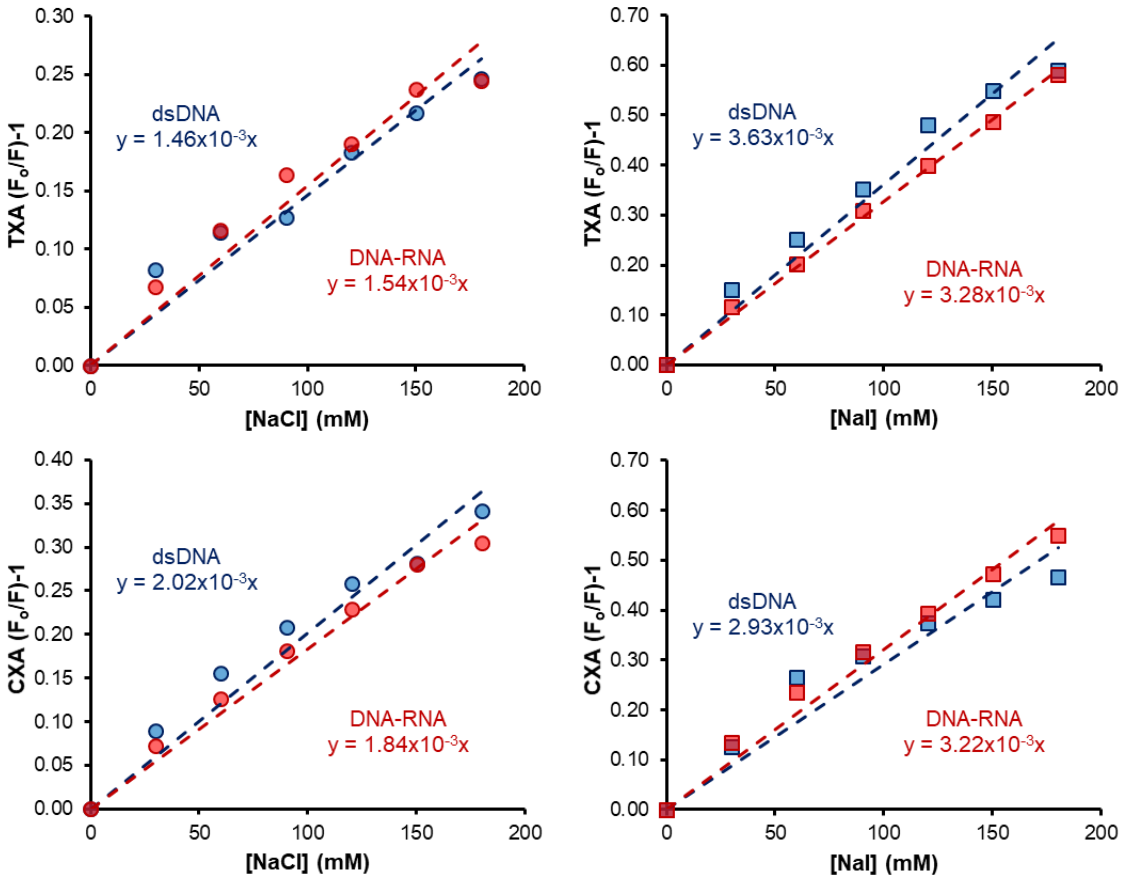


Figure 2.S5. Stern-Volmer analysis of sequences TXA and CXA. Experiments were performed using sodium chloride or sodium iodide.

2.9 Experimental Methods

Synthesis of oligonucleotide probes

Probe sequences were generated by synthesizing ^{DEA}tC phosphoramidites using published methods,[53] and subsequently incorporated into DNA strands using solid-phase DNA synthesis performed by TriLink Biotechnologies, Inc. (San Diego, CA). The HPLC-purified oligonucleotides were characterized by MALDI-TOF mass spectrometry and found to be consistent with calculated masses as shown in **Table S2**. Complementary RNA and unmodified DNA sequences were purchased from Integrated DNA Technologies, Inc. (San Diego, CA).

Table 2.S2. Quantitative analysis of ^{DEA}tC oligonucleotides.

Sequence Name	Sequence (5'…3')	Calculated Mass (g mol^{-1})	Observed Mass (amu)
GXC	CGCA-GXC-TCG	3166.7	3166.5
GXA	CGCA-GXA-TCG	3190.7	3190.1
AXA	CGCA-AXA-TCG	3174.8	3174.2
TXA	CGCA-TXA-TCG	3165.8	3165.2
CXA	CGCA-CXA-TCG	3150.8	3150
CXC	CGCA-CXC-TCG	3126.8	3126
GXG	CGCA-GXG-TCG	3206.7	3206.2
CXT	CGCA-CXT-TCG	3141.8	3140.9

Steady-state fluorescence spectroscopy to measure quantum yield of emission

Quantum yield values were measured using the comparative method with quinine sulfate (QS, Φ_{em} = 0.54) in 0.1 M H_2SO_4 as a standard.[54] Oligonucleotide stock solutions were diluted in 1X PBS pH 7.4 prepared with nuclease-free water and measurements were taken in quartz sub-micro cuvettes with a 1.0 cm path length (Starnacell Inc.) at 25 °C. Emission spectra were recorded under steady-state conditions using a PTI QuantaMaster QM-400 fluorometer (Horiba Scientific) and absorbance spectra were measured using a Shimadzu UV-1700 Pharmaspec spectrophotometer. ^{DEA}tC oligonucleotides were annealed to 1.2 molar equivalents of RNA complement and heated to 72 °C for 5 min, then cooled to room temperature before measuring absorbance and emission. Duplex samples were measured with an initial absorbance near 0.10, approximating 37 μM of fluorophore, at 390 nm. The same wavelength was used as excitation for collecting total visible emission, 405 – 700 nm. Absorbance and emission spectra recorded with further diluted samples for at least six data points. The slopes of integrated emission versus absorbance for QS standard and sample sequences were used in **Equation 2.1** to calculate quantum yield.

Equation 2.1

$$\Phi_F = \Phi_{ref}\left(\frac{slope_F}{slope_{ref}}\right)\left(\frac{\eta_F}{\eta_{ref}}\right)^2$$

Where *F* denotes values for the fluorophore ^{DEA}tC, quinine sulfate is the reference, and η is the refractive index of 0.10 M H_2SO_4 or 1X PBS. Sequences were measured at least in duplicate and averaged.

Time-dependent fluorescence spectroscopy and excited-state lifetime measurements

Lifetimes were recorded using a Delta Pro™ Fluorescence Lifetime System (Horiba Scientific) with a LED Delta Diode 371 nm excitation laser operating at 25 MHz pulse frequency. Detected photons were collected in the range of 250 – 650 nm. Time-dependent decays were recorded in a quartz sub-micro cuvette with a 1.0 cm path length (Starnacell Inc.) at 25 °C. Emission plots from time-correlated single photon counting (TCSPC) were fitted to an exponential decay using EzTime Processing Software (version 3.2.9.11 Horiba Scientific) with one or two components to calculate the excited-state lifetimes (τ). The instrument response function (IRF) was recorded using dilute colloidal silica (0.30 % solution, LUDOX™) and deconvoluted from sample spectra. Pre-exponential emission intensities (I_0) were normalized to 1.00. A two component model, as shown in **Equation 2.2**, was selected if a single-component fitting resulted in X^2>>1. Fitted linear regressions approximated X^2=1.

Equation 2.2

$$I(t) = I_0 e^{-t/\tau 1} + I_0 e^{-t/\tau 2}$$

From the measured quantum yields, the radiative decay rates (k_f) and non-radiative rate (k_{nr}) were calculated using **Equations 2.3** and **2.4**.

Equation 2.3

$$k_f = \frac{\Phi_F}{\tau}$$

Equation 2.4

$$\Phi_F = \frac{k_f}{k_f + k_{nr}}$$

Duplex structure and thermodynamic stability by circular dichroism spectroscopy

To measure the structure of ^{DEA}tC DNA-RNA heteroduplexes, CD spectra were scanned using 2.5 μM samples at 25 °C in a 0.20 cm quartz cuvette and an Aviv model 420 CD spectrophotometer. CD spectra were averaged from two scans ranging 320 nm to 200 nm, in 1 nm increments. Background 1X PBS spectra were subtracted and raw signal (θ_i) in mdeg was converted to molar elipticity ($\Delta\varepsilon$, $M^{-1}\,cm^{-1}$). Duplex melting temperatures were determined by measuring the absorbance at 270 nm from 20 °C to 80 °C in 1 °C increments. The data were normalized to a two-state model using **Equation 2.5**.

Equation 2.5

$$\alpha_F = \frac{\theta_i - \theta_U}{\theta_F - \theta_U}$$

Where fraction annealed (α_F) equals the difference between raw signal minus the signal of fully denatured duplex (θ_U), divided by the difference between fully annealed duplex signal (θ_F) and denatured duplex signal. The normalized values were fit to a logistic model using OriginLab and **Equation 2.6** to calculate the melting temperature.

Equation 2.6

$$\alpha_F = \frac{1}{1 + e^{-k(T - T_m)}}$$

Where T is temperature, T_m is the duplex melting temperature, and k is the melting rate (not reported).

Fluorescence quenching with halide anions

Stern-Volmer experiments were performed using ^{DEA}tC dsDNA and DNA-RNA samples suspended in 10 mM Na_2HPO_4 pH 7.4 buffer using degassed Ultrapure™ nuclease-free water. Samples were prepared to 37 μM in 150 μL buffer and fluorescence with 395 nm excitation was measured at 25 °C under steady-state conditions. A 4.5 M solution of sodium iodide or sodium

chloride (in degassed H_2O) was added in 1 μL increments and the emission was recorded for a total of six data points. The final sample volume amounted to 156 μL corresponding to a final fluorophore concentration of 35.6 μM and a 3.8 % molar reduction. The Stern-Volmer quenching efficiency (K_{SV}) was determined from **Equation 2.7** and a linear plot of the data. F_o/F is the ratio of emission in the absence of quencher (Q) to the presence of (Q). Calculating the bimolecular quenching rate (k_q) is achieved by substituting the unquenched excited-state lifetime in the equation.

Equation 2.7

$$\frac{F_0}{F} = 1 + K_{SV}[Q] = 1 + k_q\tau[Q]$$

Kinetic isotope effect on fluorescence efficiency and excited state lifetime

Duplex oligonucleotide samples for dsDNA and DNA-RNA were prepared in protonated buffer before dehydrating the samples by lyophilization. The samples were suspended in D_2O and the solvent evaporated by lyophilization for a total of two times. Quantum yields and fluorescence lifetimes were recorded in deuterated buffers to measure Φ_D and τ_D. Experimental parameters, instruments, and cuvettes were identical to previously detailed methods. KIE values for dsDNA and DNA-RNA were calculated comparing the non-radiative decay rates (k_{nr}) in protonated or deuterated solvents.

Equation 2.8

$$KIE = \frac{\mathrm{k}_{nr\ H}}{\mathrm{k}_{nr\ D}}$$

pH-dependent absorbance and emission measurements

A standard solution of 25 mM ^{DEA}tC nucleoside was prepared in 100% DMSO and diluted to individual 0.12 mM samples in varying concentrations of H_2SO_4. Absorbance and

emission spectra (with λ_{ex}=395 nm) were recorded for each individual sample starting at pH 1.50 and increasing pH increments of 0.50 pH units at 25 °C.

2.10 Acknowledgements

The contents of Chapter 2, in part, is currently being prepared for submission for publication of the following material: Turner, M. B., Ceja, J., Kim, G. Y., and Purse, B. W. Spectroscopic analysis of the fluorescent turn-on nucleoside 8-diethlamino-tC in DNA-RNA hybrids to identify mechanism of emissive response. *J. Phys. Chem. B.* 2020. The dissertation author was the lead author of this paper.

2.11 References

(1) *Fluorescent Analogs of Biomolecular Building Blocks: Design and Applications*; Wilhelmsson, L. M., Tor, Y., Eds.; John Wiley & Sons, Inc.: Hoboken, NJ, 2016.

(2) Neef, a. B.; Luedtke, N. W. *Proc. Natl. Acad. Sci.* **2011**, *108* (51), 20404–20409.

(3) Ward, D. C.; Reich, E.; Stryer, L. *J. Biol. Chem.* **1969**, *244* (5), 1228–1237.

(4) Xu, W.; Chan, K. M.; Kool, E. T. *Nat. Chem.* **2017**, *9* (11), 1043–1055.

(5) Teppang, K. L.; Lee, R. W.; Burns, D. D.; Turner, M. B.; Lokensgard, M. E.; Cooksy, A. L.; Purse, B. W. *Chem. - A Eur. J.* **2019**, *25*, 1249–1259.

(6) Schmidt, O. P.; Mata, G.; Luedtke, N. W. *J. Am. Chem. Soc.* **2016**, *138* (44), 14733–14739.

(7) Saito, Y.; Suzuki, A.; Okada, Y.; Yamasaka, Y.; Nemoto, N.; Saito, I. *Chem. Commun.* **2013**, *49* (50), 5684–5686.

(8) Wypijewska del Nogal, A.; Füchtbauer, A. F.; Bood, M.; Nilsson, J. R.; Wranne, M. S.; Sarangamath, S.; Pfeiffer, P.; Rajan, V. S.; El-Sagheer, A. H.; Dahlén, A.; Brown, T.; Grøtli, M.; Wilhelmsson, L. M. *Nucleic Acids Res.* **2020**.

(9) Gardarsson, H.; Kale, A. S.; Sigurdsson, S. T. *ChemBioChem* **2011**, *12* (4), 567–575.

(10) Burns, D. D.; Teppang, K. L.; Lee, R. W.; Lokensgard, M. E.; Purse, B. W. *J. Am. Chem. Soc.* **2017**, *139* (4), 1372–1375.

(11) Franklin, R. E.; Gosling, R. G. *Nature* **1953**, *172* (4369), 156–157.

(12) Dickerson, R. E.; Drew, H. R.; Conner, B. N.; Wing, R. M.; Fratini, A. V; Kopka, M. L. *Science (80-.).* **1982**, *216* (4545), 475 LP – 485.

(13) Liebl, K.; Zacharias, M. *Nucleic Acids Res.* **2018**, *47* (3), 1132–1140.

(14) Wan, P.; Shukla, D. *Chem. Rev.* **1993**, *93* (1), 571–584.

(15) Middleton, C. T.; de La Harpe, K.; Su, C.; Law, Y. K.; Crespo-Hernández, C. E.; Kohler, B. *Annu. Rev. Phys. Chem.* **2009**, *60* (1), 217–239.

(16) Markovitsi, D. *Photochem. Photobiol.* **2016**, *92* (1), 45–51.

(17) Michel, B. Y.; Dziuba, D.; Benhida, R.; Demchenko, A. P.; Burger, A. *Frontiers in Chemistry* . 2020, p 112.

(18) Cohen, B.; Huppert, D. *J. Phys. Chem. A* **2001**, *105* (30), 7157–7164.

(19) Zhou, P.; Han, K. *Acc. Chem. Res.* **2018**, *51* (7), 1681–1690.

(20) Dumas, A.; Luedtke, N. W. *Chem. – A Eur. J.* **2012**, *18* (1), 245–254.

(21) Bloomfield, V. . .; Crothers, D. .; Tinoco, I. *Nucleic Acids; Structures, Properties, and Functions*, 1st ed.; University Science Books: Sausolito, CA, 1999.

(22) Sandin, P.; Wilhelmsson, L. M.; Lincoln, P.; Powers, V. E. C.; Brown, T.; Albinsson, B. *Nucleic Acids Res.* **2005**, *33* (16), 5019–5025.

(23) Slotkin, W.; Nishikura, K. *Genome Med.* **2013**, *5* (11), 105.

(24) Gerber, A. P.; Keller, W. *Trends Biochem. Sci.* **2001**, *26* (6), 376–384.

(25) Pullirsch, D.; Jantsch, M. F. *RNA Biol.* **2010**, *7* (2), 205–212.

(26) Roth, S. H.; Levanon, E. Y.; Eisenberg, E. *Nat. Methods* **2019**, *16* (11), 1131–1138.

(27) Wright, D. J.; Force, C. R.; Znosko, B. M. *Nucleic Acids Res.* **2018**, *46* (22), 12099–12108.

(28) Kang, H.; Lee, K. T.; Jung, B.; Ko, Y. J.; Kim, S. K. *J. Am. Chem. Soc.* **2002**, *124* (44), 12958–12959.

(29) Seidel, C. A. M.; Schulz, A.; Sauer, M. H. M. *J. Phys. Chem.* **1996**, *100* (13), 5541–5553.

(30) Pecourt, J.-M. L.; Peon, J.; Kohler, B. *J. Am. Chem. Soc.* **2001**, *123* (42), 10370–10378.

(31) Kypr, J.; Kejnovska, I.; Renciuk, D.; Vorlıckova, M. *Nucleic Acids Res.* **2009**, *37* (6), 1713–1725.

(32) Treinin, A.; Hayon, E. *J. Am. Chem. Soc.* **1976**, *98* (13), 3884–3891.

(33) Shannon, R. D. *Acta Crystallogr. Sect. A* **1976**, *32* (5), 751–767.

(34) Giri, R. *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* **2004**, *60* (4), 757–763.

(35) Guo, Y.; Cao, F.; Qiu, P.; Wang, Z. *Luminescence* **2019**, *34* (4), 450–455.

(36) Somsen, O. J. G.; Hoek, van A.; Amerongen, van H. *Chem. Phys. Lett.* **2005**, *402* (1), 61–65.

(37) Sandin, P.; Börjesson, K.; Li, H.; Mårtensson, J.; Brown, T.; Wilhelmsson, L. M.;

Albinsson, B. *Nucleic Acids Res.* **2008**, *36* (1), 157–167.

(38) Wilson, J. N.; Cho, Y.; Tan, S.; Cuppoletti, A.; Kool, E. T. *ChemBioChem* **2008**, *9* (2), 279–285.

(39) Dierckx, A.; Miannay, F.-A.; Ben Gaied, N.; Preus, S.; Björck, M.; Brown, T.; Wilhelmsson, L. M. *Chem. – A Eur. J.* **2012**, *18* (19), 5987–5997.

(40) Černý, J.; Kabeláč, M.; Hobza, P. *J. Am. Chem. Soc.* **2008**, *130* (47), 16055–16059.

(41) Yakovchuk, P.; Protozanova, E.; Frank-Kamenetskii, M. D. *Nucleic Acids Res.* **2006**, *34* (2), 564–574.

(42) Riley, K. E.; Hobza, P. *Acc. Chem. Res.* **2013**, *46* (4), 927–936.

(43) Poater, J.; Swart, M.; Bickelhaupt, F. M.; Fonseca Guerra, C. *Org. Biomol. Chem.* **2014**, *12* (26), 4691–4700.

(44) Sinkeldam, R. W.; Greco, N. J.; Tor, Y. *ChemBioChem* **2008**, *9* (5), 706–709.

(45) Harteis, S.; Schneider, S. *Int. J. Mol. Sci.* **2014**, *15* (7), 12335–12363.

(46) Mata, G.; Schmidt, O. P.; Luedtke, N. W. *Chem. Commun.* **2016**, *52* (25), 4718–4721.

(47) Schulman, S. G.; Capomacchia, A. C. *J. Phys. Chem.* **1975**, *79* (14), 1337–1343.

(48) Didier, P.; Kuchlyan, J.; Martinez-Fernandez, L.; Gosset, P.; Léonard, J.; Tor, Y.; Improta, R.; Mély, Y. *Phys. Chem. Chem. Phys.* **2020**, *22* (14), 7381–7391.

(49) Fontenete, S.; Guimarães, N.; Wengel, J.; Azevedo, N. F. *Crit. Rev. Biotechnol.* **2015**, *8551* (December), 1–12.

(50) Lefever, S.; Pattyn, F.; Hellemans, J.; Vandesompele, J. *Clin. Chem.* **2013**, *59* (10), 1470–1480.

(51) Cui, C.; Shu, W.; Li, P. *Front. cell Dev. Biol.* **2016**, *4*, 89.

(52) Mizrahi, R. A.; Shin, D.; Sinkeldam, R. W.; Phelps, K. J.; Fin, A.; Tantillo, D. J.; Tor, Y.; Beal, P. A. *Angew. Chemie* **2015**, *54* (30), 8713–8716.

(53) Turner, M. B.; Anderson, B. A.; Samaan, G. N.; Coste, M.; Burns, D. D.; Purse, B. W. *Curr. Protoc. Nucleic Acid Chem.* **2018**, e59.

(54) Williams, A. T. R.; Winfield, S. A.; Miller, J. N. *Analyst* **1983**, *108*, 1067–1071.

Chapter 3

Fluorescent Tricyclic Cytidine Analogues as Substrates for Retroviral Reverse Transcriptases

3.1 Introduction

Fluorescent labeling of nucleic acids using chemically modified nucleotides is an increasingly important biophysical tool in studies on the replication, function, and expression of the nucleic acids. For example, Kraus *et al.* have recently shown that nucleic acids in living eukaryotic cells can be metabolically labeled using fluorescent nucleotides and a supramolecular nucleotide transporter that enables membrane permeation.[1] Damha *et al.* have used fluorescent nucleobase analogues to create intrinsically fluorescent siRNA that can be delivered and imaged in living cells.[2] The advantages of using intrinsically fluorescent base analogues retaining canonical Watson–Crick hydrogen bonding include precise probe placement with direct detection of biomolecular events, for instance DNA/RNA metabolism.[3–5] More than 100 fluorescent nucleoside analogues have been developed to date, motivated by the need for analogues of each natural nucleobase and a range of fluorescent properties and responsiveness.[6,7]

Site-specific labeling of nucleic acids with fluorescent nucleotide analogues is often, but not exclusively, performed using solid-phase synthesis,[8] whereas DNA/RNA polymerase-mediated insertion of fluorescent nucleotides is more suitable for multiple fluorophore incorporation in longer nucleic acid polymers and metabolic labeling.[7] When this polymerases-mediated approach is performed, the substitutions of a fluorescent nucleotide for the natural counterpart is determined by the discrimination of the acting polymerase and the ratio of fluorescent to natural nucleotides used in the reaction mixture. Fluorophores are distributed throughout the DNA/RNA product rather than at specific sites.[9–17] Most of these past studies

have used polymerases lacking 3′→5′ proofreading exonuclease activity, as proofreading often limits analogue substitution efficiency.[18]

Reverse transcriptases (RT) are virally encoded enzymes that catalyze the synthesis of complementary DNA from first an RNA or a DNA template during viral replication (**Figure 3.1**).[19,20] During RNA-templated DNA synthesis, the viral RNA is degraded by the RT's RNase H domain, which recognizes the DNA–RNA heteroduplex. These RTs are intriguing targets for metabolic labeling because the polymerase activity has a high error rate relative to other DNA polymerases and they lack 3′→5′ exonuclease proofreading activity, [21] suggesting that they will tolerate nucleoside analogues. Past studies have shown that propargylated 2′-deoxyribonucleosides are inserted preferentially by HIV-1-RT over cellular polymerases, however these non-emissive analogues require fluorophore conjugation for optical detection.[22] The emissive thieno[3,4-*d*]pyrimidine analog thA was demonstrated to be recognized by AMV-RT as a templating base in RNA-dependent DNA synthesis, but nucleotide analogues with intrinsically fluorescent, modified nucleobases have not been studied as substrates for insertion by reverse transcriptases to yield natively fluorescent complementary DNA.[3] In the present study, we sought to examine the incorporation of fluorescent cytidine analogues during both RNA-dependent and DNA-dependent DNA synthesis by the reverse transcriptases of three well characterized retroviruses: avian myeloblastosis virus (AMV), Moloney murine leukemia virus (M-MLV), and human immunodeficiency virus (HIV-1). We assess the kinetics and fidelity of fluorescent cytidine analogue incorporation under steady state conditions that fit a Michaelis–Menten kinetics model, we measure read-through following fluorescent analogue insertion, and we measure the ability of the analogues to function as templates in DNA-dependent DNA synthesis.

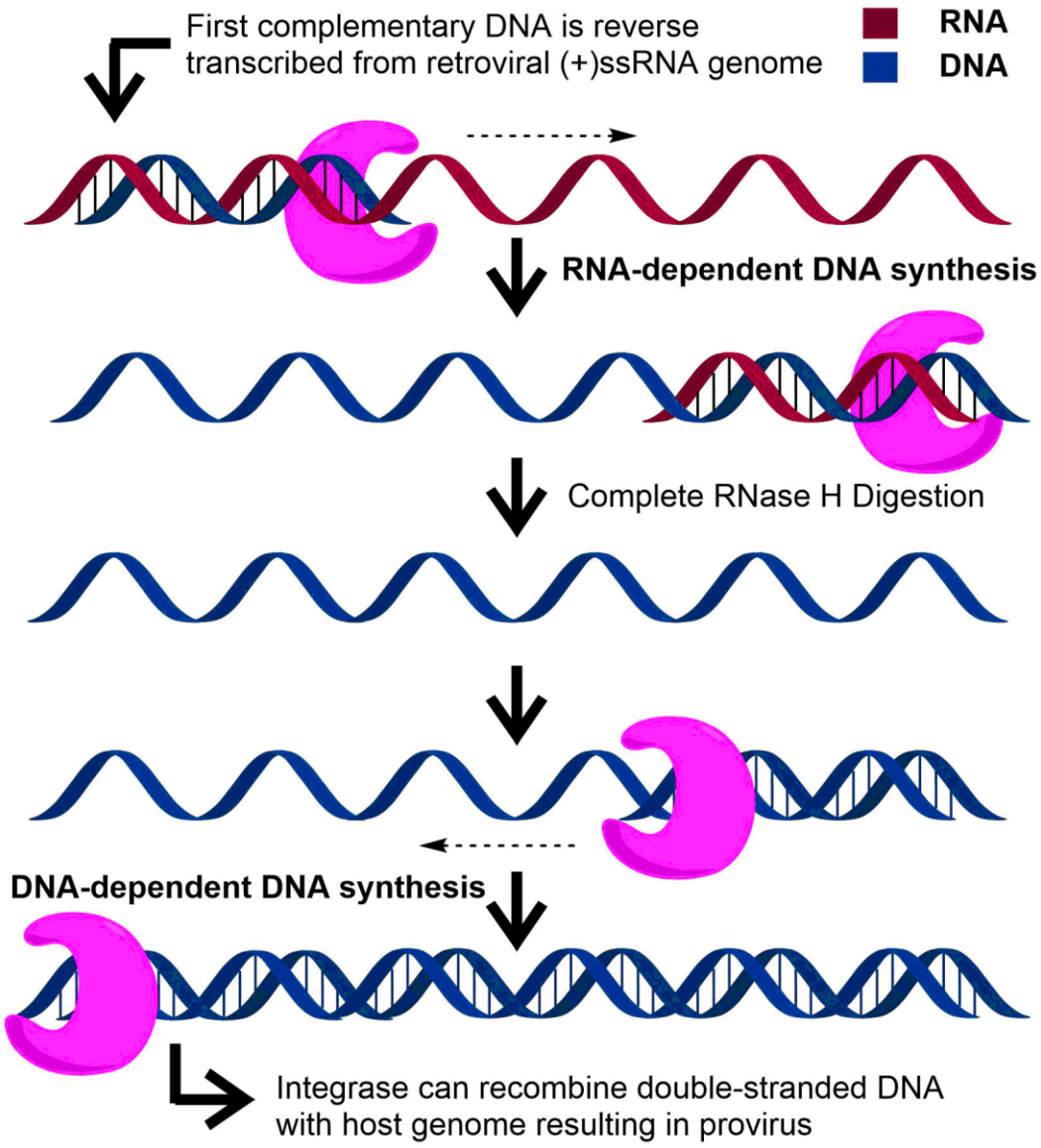


Figure 3.1. A simplified sequence of retroviral nucleic acid replication. (A) reverse transcription (RNA-dependent DNA synthesis) of the viral (+)ssRNA genome with concomitant RNA degradation by RNase H, and (B) DNA-dependent DNA synthesis, to generate dsDNA that can be integrated into the host genome.

For this study, we selected two members the tricyclic cytidine family of nucleotide analogue fluorophores because of their well-characterized fluorescent properties and the availability of data from past studies on their insertion by A and B family DNA polymerases and T7 RNA polymerase.[4,10,12] tC compounds are fluorescent nucleotide derivatives of

phenothiazine.[23] They retain the capacity for Watson–Crick hydrogen bonding with guanine (**Figure 3.2A**) and their fluorescent properties have been tuned by systematic substitution with electron donating and withdrawing groups.[24] Parent tC has fluorescence that is relatively insensitive to the molecule's local environment and is among the first nucleobase analogues known to retain fluorescence (i.e. exhibit minimal quenching) in the base stack.[25,26] In this context, its average Φ_{em} = 0.2, varying from 0.16 to 0.21 depending on the identity of neighboring bases. We recently reported that 8-diethylamino-tC (^{DEA}tC) has low intrinsic fluorescence as a nucleoside monomer (Φ_{em} = 0.006) or in single-stranded DNA (Φ_{em} = 0.012) but exhibits up to a 20-fold fluorescence turn-on upon incorporation into duplex DNA (**Figure 3.2B**), with some dependence on the identity of neighboring bases.[24,27] Past collaborative work from our lab and the Kuchta group showed that polymerases lacking 3′→5′ exonuclease activity, including the Klenow fragment, human pol α, and T7 RNA polymerase, incorporate tC (2'-deoxy)ribonucleoside triphosphates into DNA or RNA with kinetics comparable to natural nucleotides, although the DNA polymerases insert tC across from adenine with approximately 10% the efficiency of thymine, likely a result of tautomerism.[10,12] Here, we show that d(tC)TP and d(^{DEA}tC)TP are generally good substrates for single nucleotide insertion by AMV, M-MLV, and HIV-1 RTs in both RNA-dependent DNA synthesis and DNA-dependent DNA synthesis modes. Elongation after insertion works well when using RNA templates and less well with DNA templates. The compatibility of d(tC)TP and d(^{DEA}tC)TP with HIV-1-RT is demonstrated in a complete RT assay, starting with a ssRNA template, and resulting in a fluorescently labeled dsDNA product.

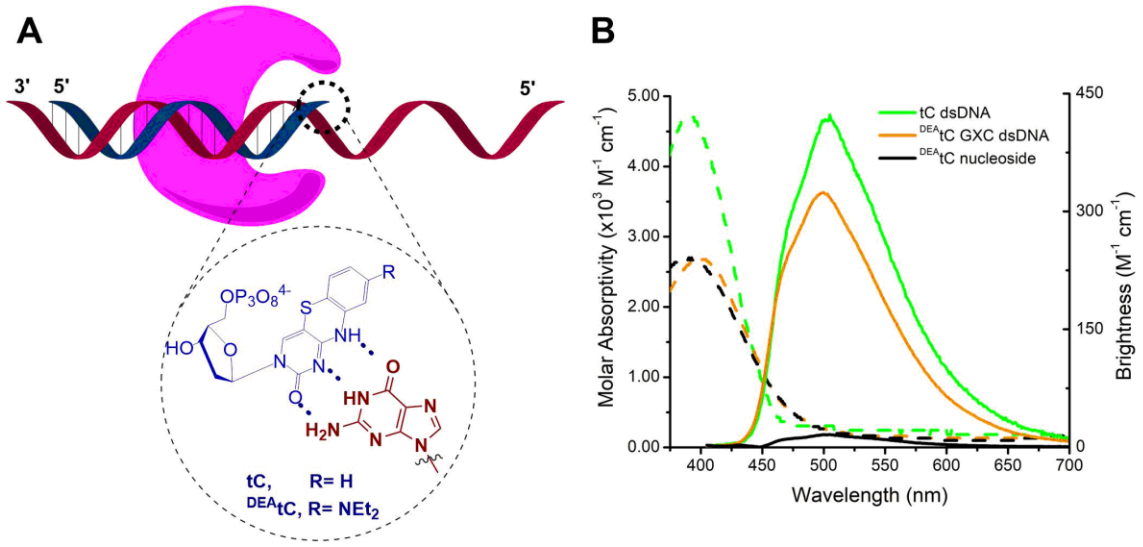


Figure 3.2. Structure of tC and ^{DEA}tC nucleotide analogue fluorophores. (A) Structures illustrated in a cartoon drawing of the active site of reverse transcriptase. (B) Emission spectra (solid) for parent tC (R=H) and ^{DEA}tC (R = NEt_2) are scaled to relative fluorescence brightness. The dsDNA sequence used is 5′-CGCA-GXC-TCG-3′ and its complement, where X is the position of the analogue. Measurements are recorded in 1× PBS buffer, pH 7.4.

3.2 Overview of methods and results

We synthesized the tC and ^{DEA}tC nucleosides using published methods and converted them to the corresponding triphosphates d(tC)TP and d(^{DEA}tC)TP by the application of standard Ludwig phosphorylation chemistry (see the Supporting Information).[28–30] The AMV-RT, M-MLV-RT, and HIV-1-RT enzymes were obtained from commercial sources and the kinetics of primer extension were measured using fluorescently labeled primer–templates and standard *in vitro* protocols (for details, see the Supporting Information). We were interested in determining the relative preference for native nucleotides versus analogues, and an excess of primer/template substrate relative to enzyme (i.e. Michaelis–Menten conditions) allows some approximation of values for V_{max} and K_M if the polymerase is processive.[31] While there are limitations in using steady-state kinetics to measure single nucleotide incorporation rates in that the overall rate limiting step is typically product release, these studies can be instructive in approximating V_{max}

and K_M,[32] and for this reason we are reporting our kinetic parameters as $V_{max,obs}$ and $K_{M,obs}$. We measured the kinetics and fidelity of d(tC)TP and d(^{DEA}tC)TP insertion during both RNA- and DNA-dependent DNA synthesis by all three RTs. To check for retained processivity after fluorescent analogue incorporation or when the analogues were the templating DNA base, we performed a series of read-through assays. Last, we performed a complete reverse transcription model assay using HIV-1-RT, showing that DNA-dependent DNA synthesis is compatible with tC or ^{DEA}tC in the templating strand.

To measure the catalytic efficiency of d(tC)TP and d(^{DEA}tC)TP insertion by the reverse transcriptases, we designed primers and templates with sequences based on previous work on the polymerization kinetics of tC nucleoside triphosphates (**Figure 2.3**).[10] d(^{DEA}tC)TP is a novel nucleoside triphosphate (synthetic details and full characterization are presented in the Supporting Information). The primers were 5′-end labeled with FAM or Cy5 for quantification of extension reactions using sequencing gels and a Typhoon FLA 9500 imager. Nucleotide insertion reactions were performed under steady-state conditions and made up to 10–20 μL with a range of concentrations of the nucleotide substrates. After adding one volume of pre-warmed enzyme solution, the mixture was incubated for 2 to 45 minutes depending on the activity of the enzyme and type of assay. The initial rate of primer extension was measured using a minimum of six different nucleotide concentrations, with experiments run at least in duplicate (experimental details and all data plots not included in the main text are provided in the Supporting Information). Gel images were quantified using ImageQuant software (GE Healthcare Life Sciences). Kinetic parameters for $K_{M,obs}$ and $V_{max,obs}$ were for insertion of d(tC)TP and d(^{DEA}tC)TP both across from guanine and across from adenine (the most likely mismatch), and were calculated using Origin (OriginLab, Northampton, MA).

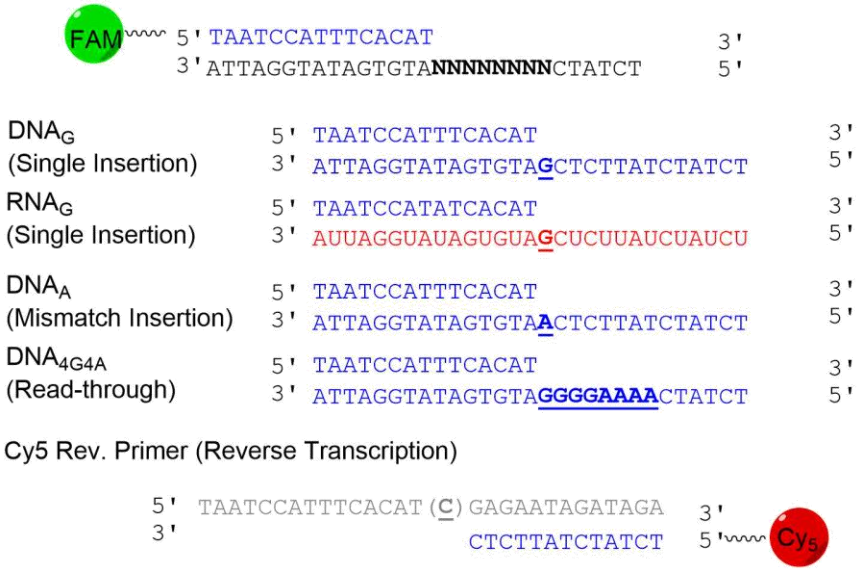


Figure 3.3. Primer–template sequences for polymerization experiments. Sequences in blue are DNA, red are RNA, and gray is the template sequence for reverse transcription generated enzymatically. In this sequence (**C**) refers to C, tC, or ^{DEA}tC, depending on which nucleotide was used in reverse transcription.

3.3 RNA-templated single nucleotide insertion and read-through

We first examined RNA-dependent DNA synthesis by the RTs using an RNA template and a DNA primer. These experiments assess the ability of d(tC)TP and d(^{DEA}tC)TP to substitute for dCTP during the first stage of viral replication (**Figure 2.1**). Plots of the initial primer extension rate against substrate concentration reveal that AMV-RT, M-MLV-RT, and HIV-1-RT are all appropriately modeled by standard Michaelis–Menten kinetics when inserting each of d(tC)TP and d(^{DEA}tC)TP. Representative plots for each of d(tC)TP, d(^{DEA}tC)TP, and natural dCTP are shown (**Figure 3.4**). Nonlinear regression was used to fit the data to the Michaelis–Menten equation, thereby providing the kinetic parameters $V_{max,obs}$ and $K_{M,obs}$. Catalytic efficiency is measured by the ratio $V_{max,obs}/K_{M,obs}$ and fidelity is determined by comparing the catalytic efficiency of the formation of the tC:G and ^{DEA}tC:G base pairs with a natural C:G base pair (**Table 3.1**).

Excepting one outlier, AMV-RT, M-MLV-RT, and HIV-1-RT insert d(tC)TP and d(^{DEA}tC)TP with catalytic efficiencies in the same order of magnitude as natural dCTP (**Table 3.1**). Several differences in catalytic efficiency for insertion of the fluorescent analogues are within experimental error. The outlier is d(^{DEA}tC)TP, which is inserted by HIV-1-RT with 9% the catalytic efficiency of natural dCTP insertion, the lowest performance observed in this RNA-templated insertion study. The greatest efficiency observed for analogue insertion is d(tC)TP by M-MLV-RT, approximately five-fold more efficient than natural dCTP. $K_{M,obs}$, which provides insight into substrate binding affinities, for d(tC)TP and dCTP do not vary greatly among the three RTs, all 2.8–7.8 μM, while there is greater variation in the d(^{DEA}tC)TP $K_{M,obs}$ values. The data suggest that the diethylamino group of ^{DEA}tC slightly diminishes M-MLV-RT and HIV-1-RT compatibility, as $K_{M,obs}$ is raised for M-MLV-RT and $V_{max,obs}$ is lowered for HIV-1-RT and with respect to the values for tC. But, within the limits of precision of our measurements, AMV-RT does not discriminate either analogue from natural dCTP.

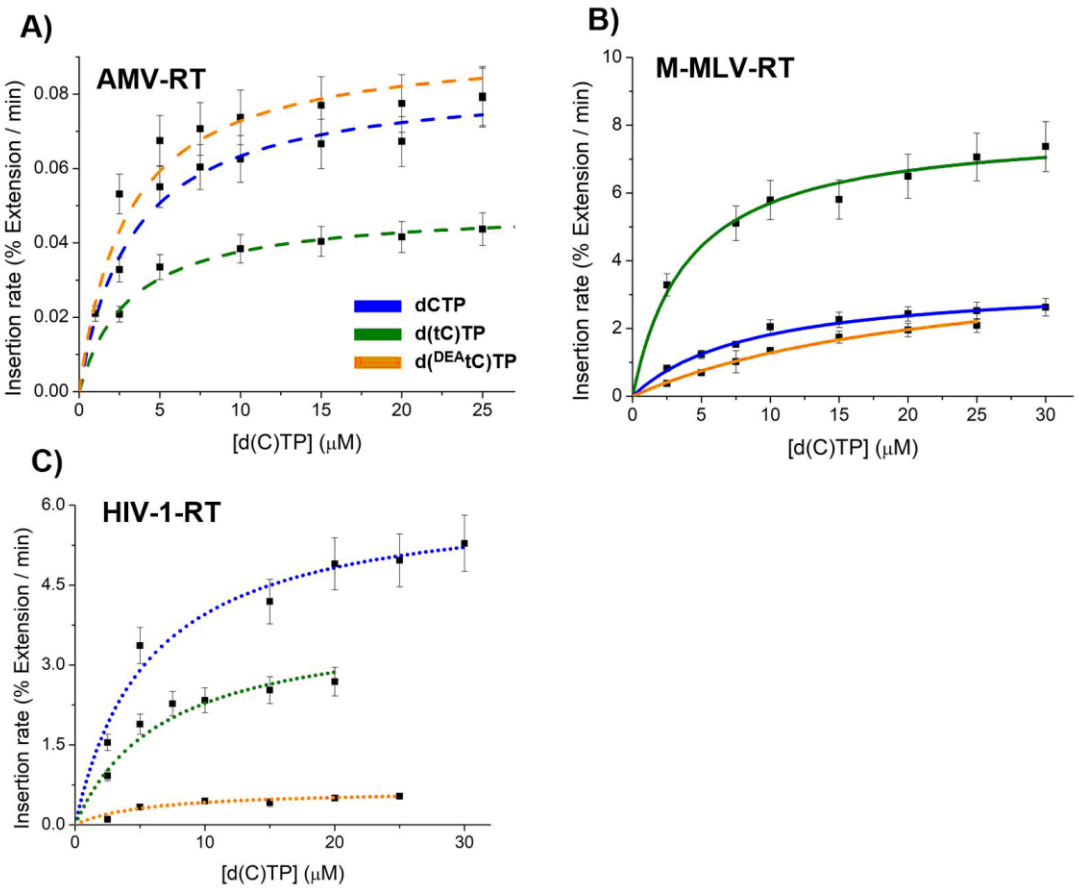


Figure 3.4. Michaelis-Menten plots of single insertion using an RNA_G template. Plots are shown for (A) AMV-RT, dashed, (B) MLV-RT, solid, and (C) HIV-1-RT, dotted.

Read-through assays were designed to test the ability of the RTs to elongate past a newly incorporated analogue. These assays were performed using the same conditions as for single nucleotide insertion assays, but with the addition of dATP, dGTP, and dTTP, to allow for complete reverse transcription of the RNA template. M-MLV-RT and HIV-1-RT generate full length 29-mer product and exhibit similar or less stalling with the analogues when compared to insertion and extension with natural dCTP, demonstrating effective read-through (**Figure 3.5**). The additional steric bulk of tC and ^{DEA}tC slows electrophoretic migration as compared with natural C, resulting in band misalignments between lanes. These band shifts show that tC and ^{DEA}tC are inserted across from the initial templating G in these experiments; it is not the formation natural mispairs that allows read-through without analogue incorporation. While

d(tC)TP is an excellent substrate for insertion by M-MLV-RT and a good substrate for insertion by HIV-1-RT, its insertion does induce some stalling and inhibit read-through (**Figure 3.5; lanes 5 and 8**). HIV-1-RT exhibited, in general, less stalling than M-MLV-RT when using both d(^{DEA}tC)TP and d(tC)TP as substrates, as indicated by the decreased presence of incomplete primer elongation products bands (**Figure 3.5**, **lanes 5-10**). Under these conditions, AMV-RT showed lower catalytic efficiency for nucleotide insertion across from an RNA template (**Table 3.1**). With AMV-RT, the d(tC)TP functions well as a substitute for dCTP, resulting in full length product and stalling comparable to when natural dCTP is used, but no complete elongation product was observed when using d(^{DEA}tC)TP (**Figure 3.5**, **lanes 2-4**).

Table 3.1. Kinetics measurements for RNA_G-templated single dNTP insertion.

	dNTP	$V_{max,obs}$ / % Ext min^{-1}	$K_{M,obs}$ / µM	$V_{max,obs}/K_{M,obs}$ / % Ext min^{-1} $µM^{-1}$	Fidelity[a]
AMV-RT	d(tC)TP	0.049 ± 0.001	2.8 ± 0.4	0.017 ± 0.002	0.65 ± .09
	d(^{DEA}tC)TP	0.088 ± 0.004	2.0 ± 0.4	0.043 ± 0.008	1.6 ± 0.3
	dCTP	0.084 ± 0.005	3.3 ± 0.7	0.026 ±0.006	1
M-MLV-RT	d(tC)TP	8.0 ± 0.3	4.0 ± 0.7	2.0 ± 0.4	4.7 ± 0.9
	d(^{DEA}tC)TP	3.9 ± 0.4	21 ± 3	0.19 ± 0.04	0.44 ± 0.08
	dCTP	3.3 ± 0.2	8 ± 1	0.4 ± 0.06	1
HIV-1-RT	d(tC)TP	3.4 ± 0.3	5 ± 1	0.7 ± 0.2	0.7 ± 0.2
	d(^{DEA}tC)TP	0.7 ± 0.1	7 ± 3	0.10 ± 0.04	0.09 ± 0.03
	dCTP	6.2 ± 0.4	5 ± 1	1.1 ± 0.3	1

[a] Fidelity is measured as the ratio of catalytic efficiency ($V_{max,obs}/K_{M,obs}$) for analogue insertion divided by catalytic efficiency for dCTP insertion. Errors are standard error derived from non-linear global fit of dataset.

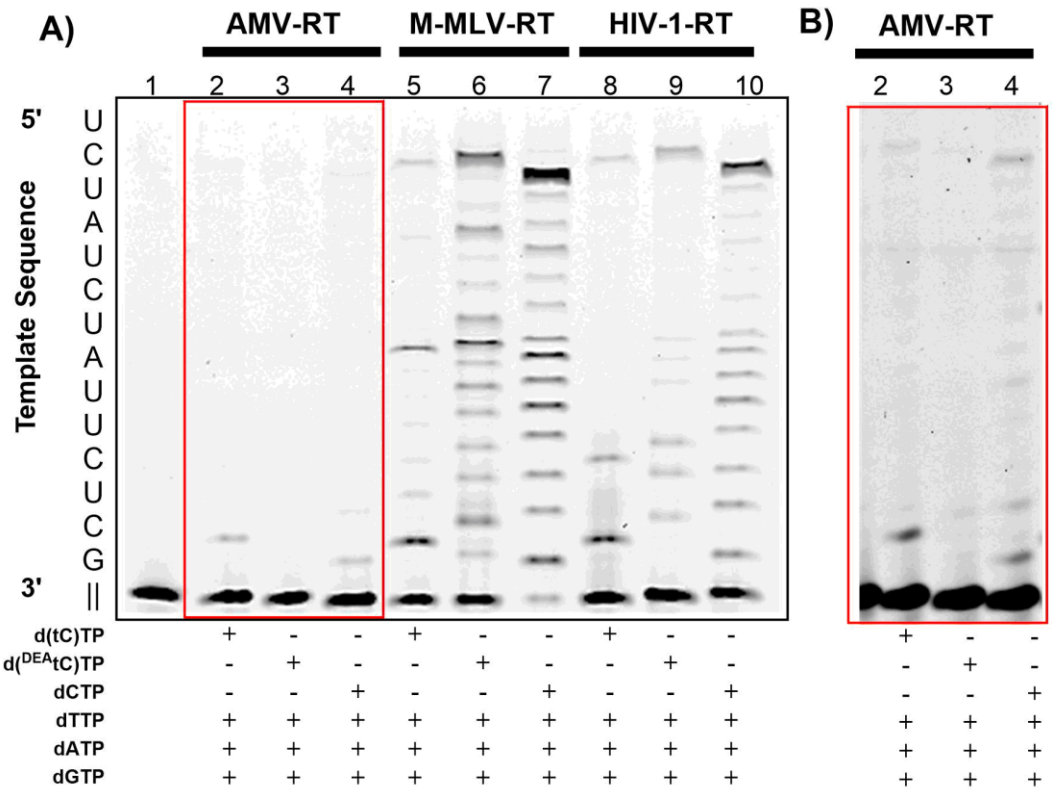


Figure 3.5. Read-through polymerization assay of RNA_G primer-template complex. (A) Reactions contained indicated cytidine nucleoside triphosphate and additional dNTP's at 15 μM to complete extension. Reactions with M-MLV-RT and HIV-1-RT were incubated for 20 min while AMV-RT reactions were incubated for 30 min. Lane 1 contains 15-mer FAM labelled primer for reference. (B) Contrast enhancement of lanes 2–4 to reveal fully extended 29-mer product using d(tC)TP and dCTP.

To assess the RTs' fidelities for pairing d(tC)TP and d(^{DEA}tC)TP only with G, we performed a similar read-through assay, omitting dTTP. Seven templating bases down from the primer's 3′ terminus is a templating A. Under these conditions, each RT is expected to cease polymerization or, in the case of poor fidelity, "misinsert" tC or ^{DEA}tC across from A, which is consistent with purine:pyrimidine mismatches being preferred.[10,33] The data show that, for M-MLV-RT, read-through works nearly as well in the absence of dTTP when d(^{DEA}tC)TP is present (**Figure 3.6, lanes 5-7**). Complete primer extension is observed, with lower electrophoretic mobility and the differences in observed stalling supporting that it is tC or ^{DEA}tC inserted across from A. When using d(tC)TP, polymerization past the first templating A is efficient (**lane 5**), but

stalling occurs at the second templating A. These data show that d(^{DEA}tC)TP substitutes very efficiently for dTTP and d(tC)TP less so. Under these conditions, in the absence of dTTP, AMV-RT and HIV-1-RT both show stalling prior to the first templating A (**Figure 3.6**, **lanes 2**, **3** and **8**, **9**, respectively), so the insertion of d(tC)TP or d(^{DEA}tC)TP across from A could not be observed.

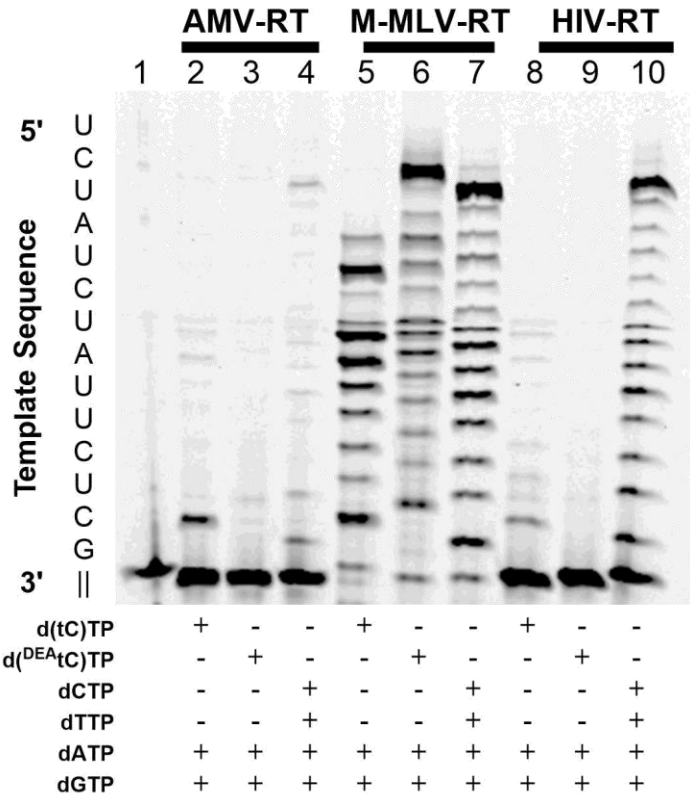


Figure 3.6. Read-through polymerization assay of RNA_G primer-template complex without dTTP. Reactions with d(tC)TP and d(^{DEA}tC)TP in the absence of dTTP to assess the enzymes' ability to polymerize past template adenosine seven bases down from the primer's 3′ end. Reactions contained 30 μM dNTPs and were incubated for 40 min at 42 °C for AMV-RT and M-MLV-RT, and 37 °C for HIV-1-RT.

3.4 DNA-templated single nucleotide insertion kinetics and read-through

Similar kinetic measurements of single-nucleotide insertion were taken using a DNA primer and a DNA template, mimicking the second mode of nucleic acid polymerization by RT (**Figure 3.7**, **Table 2**). As in the case of an RNA template, almost all measurements of catalytic

efficiency were within an order of magnitude of those measured for natural dCTP insertion. As was observed when using an RNA template, AMV-RT shows the least discrimination of the natural nucleotides with a DNA template. The catalytic efficiency varies only by up to a factor of three. M-MLV-RT has approximately a 10-fold reduced ability to insert d(tC)TP as compared with dCTP, and HIV-1-RT has approximately a 10-fold reduced ability to insert d(^{DEA}tC)TP as compared with the natural nucleotide. The remaining pairings—M-MLV-RT with d(^{DEA}tC)TP and HIV-1-RT with d(tC)TP—show less than a three-fold decrease in catalytic efficiency as compared with the natural substrate.

Mismatch insertion assays were performed similarly, but using DNA_A with adenine in the templating position (**Figure 3.3**). As has been observed in past studies of d(tC)TP insertion across from a DNA template by non-proofreading polymerases,[10] this tricyclic cytidine is much more prone than natural dCTP to insertion across from adenine. The likelihood for incorrectly inserting tC across from adenine was approximately 5% in all cases, as measured by the catalytic efficiencies of tC vs. T insertion across from A (**Table 3.3**; the Michaelis–Menten plots are shown in **Figure S3.1**). Most of the basis for this discrimination is the attenuated d(tC)TP binding affinity to form a ternary complex, as indicated by an increased $K_{M,obs}$, but $V_{max,obs}$ is also reduced.

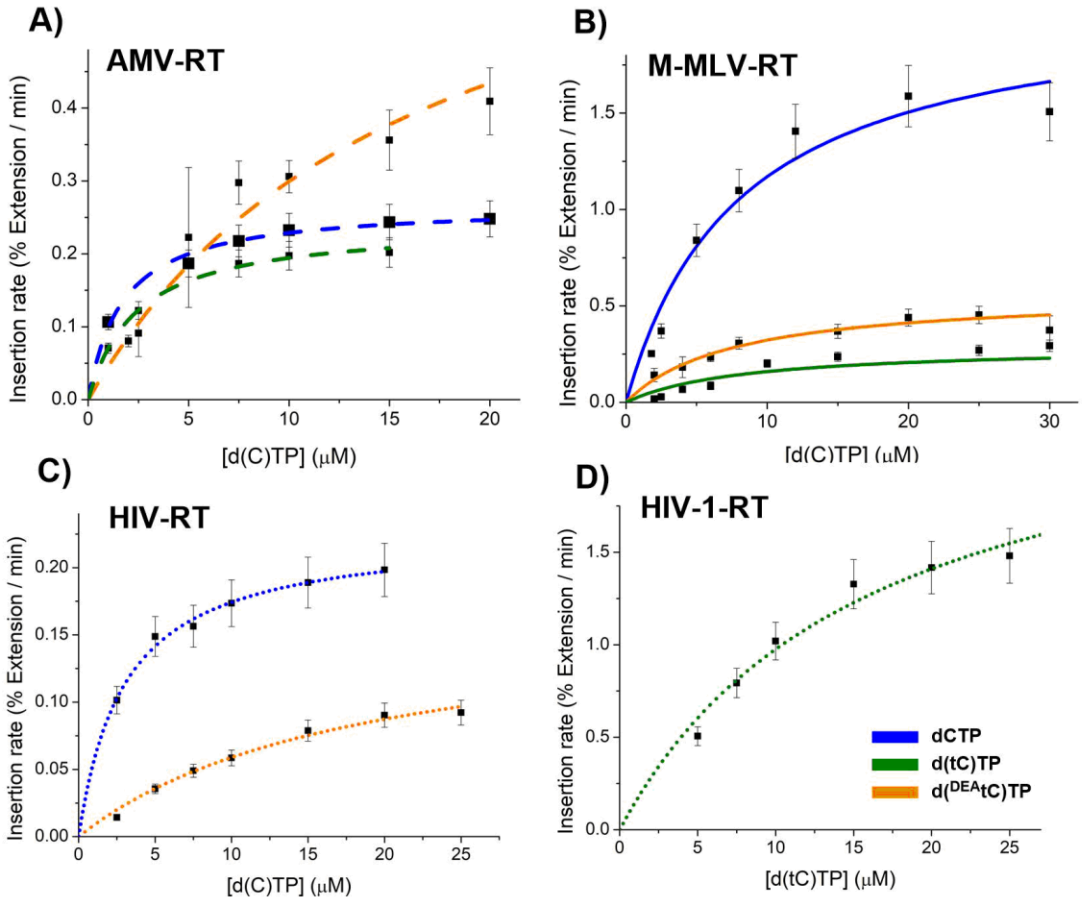


Figure 3.7. Michaelis-Menten plots of matched single insertion using DNA_G template. Plots are shown for (A) AMV-RT, dashed, (B) M-MLV-RT, solid, and (C) HIV-1-RT with dCTP and d(^{DEA}tC)TP, dotted. (D) HIV-1-RT with d(tC)TP necessitated a separate plot due to a large difference in *y*-axis scaling.

The probability of incorrectly inserting ^{DEA}tC across from A is three to eight times more likely than for a tC:A mismatch insertion among all three RT variants. MLV-RT and HIV-1-RT exhibited the largest infidelity with 39% the efficiency of d(^{DEA}tC)TP insertion across from A as compared with natural dTTP. While the $V_{max,obs}$ for these two RTs are <1/3 of that with dTTP, the $K_{M,obs}$ for d(^{DEA}tC)TP is lower, possibly a result of ^{DEA}tC's hydrophobicity. For HIV-1-RT, the catalytic efficiency for incorrectly inserting ^{DEA}tC across from adenine is the same as for correct insertion across guanine. Kinetic measurements for natural dCTP insertion with a templating adenine generated no observable product when using any RT and 500 μM dCTP, and incubating with the primer–template for 1 hr, indicating a natural discrimination factor of $< 10^{-5}$.

Experiments were performed using both DNA_G and DNA_A to measure extension activity following analogue insertion across from G and A, respectively (**Figures 3.8, S3.2**). Assays using DNA_G showed complete primer extension following some stalling after d(tC)TP was inserted by HIV-1-RT or AMV-RT (**Figure 3.8**, **lanes 2** and **8**), similar to when natural dCTP was used (**lanes 4** and **10**). In contrast, little complete 29-mer product is observed for M-MLV-RT, demonstrating less tolerance for the analogue. Assays with ^{DEA}tC did not yield observable fully-extended primer with any of the RTs. M-MLV-RT exhibited some continued extension following ^{DEA}tC insertion without full-length product formation (**lane 6**), while AMV-RT and HIV-1-RT ceased polymerization after the initial ^{DEA}tC insertion (**lanes 3** and **9**).

Table 3.2 Kinetics measurements for DNA_G-templated single dNTP insertion.

	dNTP	$V_{max,obs}$ / % Ext min^{-1}	$K_{M,obs}$ / µM	$V_{max,obs}/K_{M,obs}$ / % Ext min^{-1} $µM^{-1}$	Fidelity[a]
AMV-RT	d(tC)TP	0.240 ± 0.006	2.3 ± 0.2	0.10 ± 0.01	0.56 ± 0.05
	d(^{DEA}tC)TP	0.62 ± 0.08	10 ± 3	0.06 ± 0.02	0.3 ± 0.1
	dCTP	0.267 ± 0.002	1.5 ± 0.1	0.18 ± 0.01	1
M-MLV-	d(tC)TP	0.5 ± 0.1	20 ± 10	0.02 ± 0.01	0.09 ± 0.04
	d(^{DEA}tC)TP	0.53 ± 0.05	6 ± 2	0.08 ± 0.02	0.31 ± 0.09
	dCTP	2.1 ± 0.2	8 ± 2	0.36 ± 0.08	1
HIV-1-RT	d(tC)TP	2.5 ± 0.3	16 ± 4	0.16 ± 0.05	2.1 ± 0.6
	d(^{DEA}tC)TP	0.17 ± 0.02	19 ± 4	0.009 ± 0.002	0.12 ± 0.03
	dCTP	0.226 ± 0.006	3.0 ± 0.3	0.08 ± 0.01	1

[a] Fidelity is measured as the ratio of catalytic efficiency ($V_{max,obs}/K_{M,obs}$) for analogue insertion divided by catalytic efficiency for dCTP insertion. Errors are standard error derived from non-linear global fit of dataset.

For DNA_A, the assays were performed with either 120 µM d(tC)TP or 75 µM d(^{DEA}tC)TP and a longer incubation time was used (45-60 min) given the lower efficiencies observed in the

mismatch insertion assays (**Table 3.3**). For M-MLV-RT, insertion of d(tC)TP and d(^{DEA}tC)TP across from A was effective and read-through could continue, but no full-length product was detected with tC and only a modest amount was generated with ^{DEA}tC, indicating stalling (**Figure S3.2, lanes 5, 6**). dTTP was not provided in these mismatch extension assays and a second template A is found after six bases, where RT must insert a second $^{(DEA)}$tC:A mismatch or cease extension, yielding a 22-mer primer. M-MLV-RT yielded some product for sequences longer than 22 bases when using ^{DEA}tC (**Figure S3.2 lane 6**). HIV-1-RT generated a comparable amount of the 29-mer product when using ^{DEA}tC to mispair with A, but full-length 29-mer product was not detected with tC read-through (**lanes 8-10**). AMV-RT was very inefficient with at inserting either analogue across from T, and no full-length 29-mer product was detected (**lanes 2-4**).

An additional template sequence used, DNA_{4G4A}, included two consecutive base quartets, first four G bases followed by four A bases with the remaining sequence identical to DNA_G (**Figure 3.9**). The sequence was intended to pose a challenge to the RTs and to assess the effects of cumulative structural perturbation when multiple $^{(DEA)}$tC bases are inserted across from tracts of four templating guanines. The results from read-through experiments were similar among all three RT variants in that all three RTs were able to insert exactly four tC analogues, without continuing to mis-insert the analogues across from adenine (**lanes 2**, **8**, and **14**), however no more than two ^{DEA}tC analogues were observed to be inserted consecutively by any RT (**lanes 3**, **9**, and **15**). Although AMV-RT incorporated the first two ^{DEA}tC nucleotides with 5–6 times more product yield than the other RTs, the enzyme was unable to polymerize further. When these experiments were performed with the added dATP, dGTP, and dTTP, synthesis of the full-length 29-mer product was observed with d(tC)TP when using HIV-1-RT and AMV-RT (**Figure 3.9**

lanes 5 and **11**); there is a faint band for fully extended 29mer when using M-MLV-RT (**lane 17**), accounting for 0.82% of the total intensity in that lane.

Table 3.3 Kinetics Measurements from DNA_A-templated mismatch insertion.

	dNTP	$V_{max,obs}$ / % Ext min^{-1}	$K_{M,obs}$ / μM	$V_{max,obs}/K_{M,obs}$ / % Ext min^{-1} μM^{-1}	Fidelity[a]
AMV-RT	d(tC)TP	0.12 ± 0.03	120 ± 60	0.0010 ± 0.0006	0.04 ± 0.02
	d(^{DEA}tC)TP	0.24 ± 0.02	70 ± 10	0.0035 ± 0.0007	0.15 ± 0.03
	dTTP	0.43 ± 0.05	18 ± 4	0.024 ± 0.006	1
M-MLV-RT	d(tC)TP	0.27 ± 0.04	100 ± 30	0.0026 ± 0.0008	0.05 ± 0.02
	d(^{DEA}tC)TP	0.23 ± 0.02	10 ± 3	0.022 ± 0.006	0.4 ± 0.10
	dTTP	0.7 ± 0.1	13 ± 4	0.06 ± 0.02	1
HIV-1-RT	d(tC)TP	0.052 ± 0.003	30 ± 6	0.0017 ± 0.0004	0.07 ± 0.02
	d(^{DEA}tC)TP	0.041 ± 0.001	4.5 ± 0.9	0.009 ± 0.002	0.39 ± 0.08
	dTTP	0.16 ± 0.01	7 ± 2	0.023 ± 0.005	1

[a] Fidelity is measured as the ratio of catalytic efficiency ($V_{max,obs}/K_{M,obs}$) for analogue insertion divided by catalytic efficiency for dTTP insertion. Incubation with any of the three tested RTs and 500 μM dCTP for 1 hour generated < 0.01 % extension, resulting in < 3.3×10^{-7} % extension min^{-1} μM^{-1} efficiency. Errors are standard error derived from non-linear global fit of the data sets.

Quantifying the electrophoresis results shows that AMV-RT and HIV-1-RT have similar capacities to insert four nucleotides across from the four-G tract when using either d(tC)TP or natural dCTP, whereas M-MLV-RT performs approximately half as well when using d(tC)TP as compared with dCTP. In the HIV-1-RT extension, little to no polymerase stalling occurs when polymerizing the natural dNTPs, but stalling does occur with d(tC)TP. Moreover, the appearance of multiple bands shows that insertion of four consecutive tC residues by HIV-1-RT attenuates the processivity during the remaining nucleotide insertions to complete primer extension (**Figure 3.9**, **lanes 17-19**). M-MLV-RT and AMV-RT both show stalling and a loss of processivity immediately after the G-tract when either d(tC)TP or dCTP is used (**lanes 5**, **7** and **11**, **13**). As

would be expected from the results described above, complete primer extension was not achieved by any of the RTs when using d(^{DEA}tC)TP.

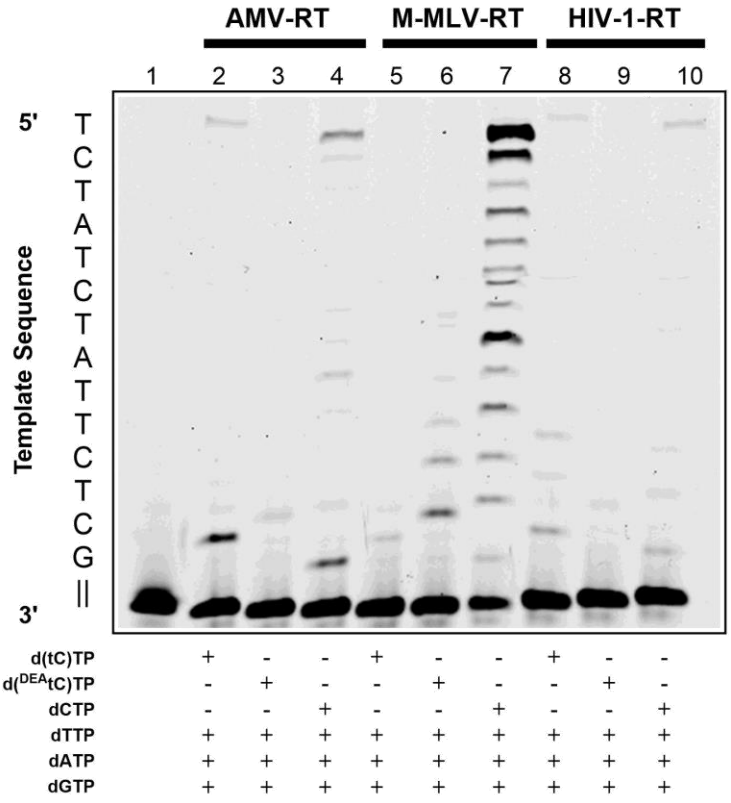


Figure 3.8. Read-through polymerization assay of DNA_G primer-template complex. Reactions contained indicated cytidine nucleoside triphosphate, additional dNTP's at 15 μM to complete extension, and were incubated for 20 min. Lane 1 contains 15-mer FAM labelled primer for reference.

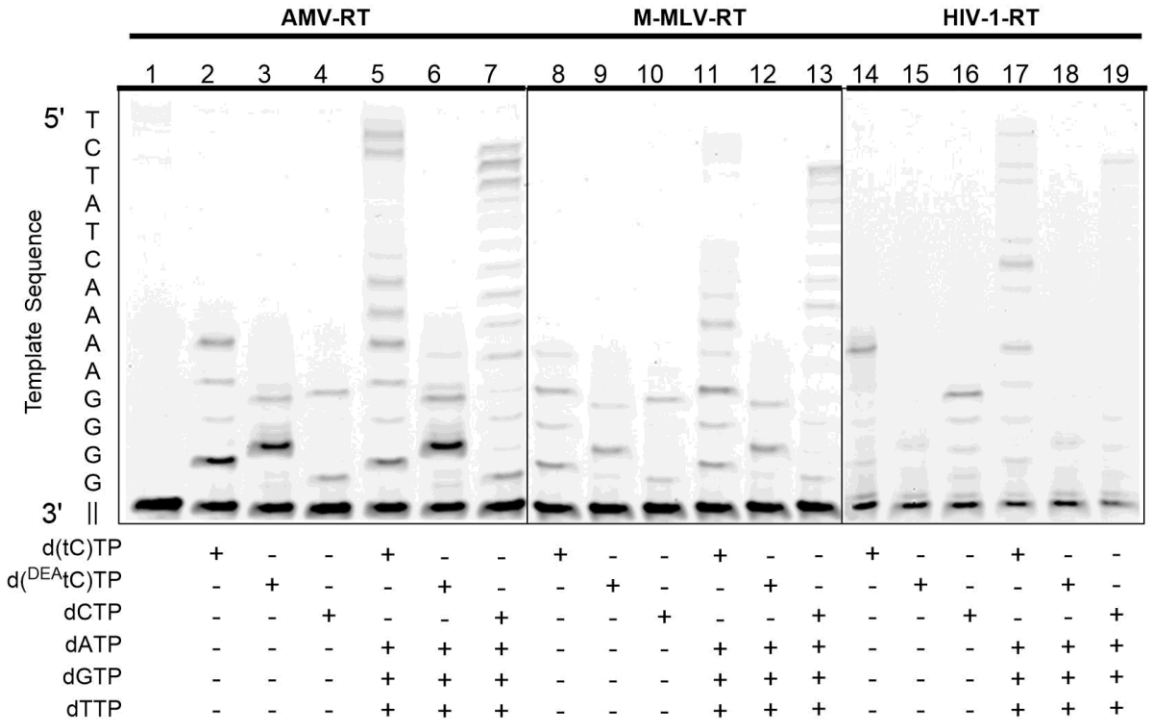


Figure 3.9. Read-through assay of reverse transcriptases across G and A quartets. Experimental conditions were 0.5 μM DNA_{4G4A} primer-template, 15 μM of indicated dNTP(s), and either 10 nM AMV-RT, 150 nM MLV-RT, or 9.75 nM HIV-1-RT. The additional steric bulk of tC and ^{DEA}tC slows electrophoretic migration as compared with natural C, a property that is manifest in band misalignment. Samples were incubated at 42 °C for AMV-RT and M-MLV-RT, 37 °C for HIV-1-RT.

3.5 Modeling in vitro reverse transcription with fluorescent tricyclic cytidine analogues

The purpose of the reverse transcription assay was to model the performance of d(tC)TP and d(^{DEA}tC)TP as substrate analogues during the complete cycle of reverse transcription of single-stranded RNA to produce a of double-stranded DNA product. We selected HIV-1-RT for this assay, given the virus's biomedical significance and good performance in the prior experiments. The complete process of reverse transcription is complex and has been studied extensively,[137,138] however we used a simplified model to validate the enzyme's competence in using tC and ^{DEA}tC as incoming and templating nucleotides (**Figure 3.10A**).

In the first part of the experiment, a FAM-labeled DNA primer was added to an RNA template to give product (I) and the primer was extended enzymatically to completion to give (II) using HIV-1-RT, dCTP, d(tC)TP or d(^{DEA}tC)TP, and the remaining natural nucleotides dGTP, dATP, and dTTP. Degradation of the RNA template occurs too, catalyzed by the RT's RNase H activity. Denaturing gel electrophoresis of an aliquot from the initial enzymatic reaction mixture showed that completely extended primer (III) was formed in this initial extension using dCTP, d(tC)TP and d(^{DEA}tC)TP, with moderate stalling using tC and more apparent stalling with ^{DEA}tC (**Figure 3.10B, lanes 3, 5,** and **7**). Next, a Cy5-labelled reverse primer was added and a brief heating–cooling cycle was used for denaturation–annealing to ensure primer hybridization to the newly synthesized 29-mer complementary DNA (cDNA) to give (IV). To account for a lower primer–template melting temperature, the second extension was performed at 32 °C for 50 minutes using only natural dNTPs, resulting in the synthesis of double-stranded DNA oligonucleotides (V). Denaturing gel electrophoresis imaged in the FAM channel showed that the first complementary DNA strands were intact (**Figure 3.10B, lanes 4, 6, and 8**). Imaging the same gel in the Cy5 channel showed the synthesis of the second complementary DNA strand, consisting only of natural nucleotides, but with templating by C, tC, or ^{DEA}tC (**Figure 3.10C**, lanes 4, 6, 8). The two bands indicative of polymerase stalling during this second DNA strand synthesis likely correspond to a dGTP and dATP insertion, respectively, with dGTP being inserted across $^{(DEA)}$tC and dATP insertion in the next position. Following these two insertions, HIV-1-RT has robust processivity for the remaining polymerization.

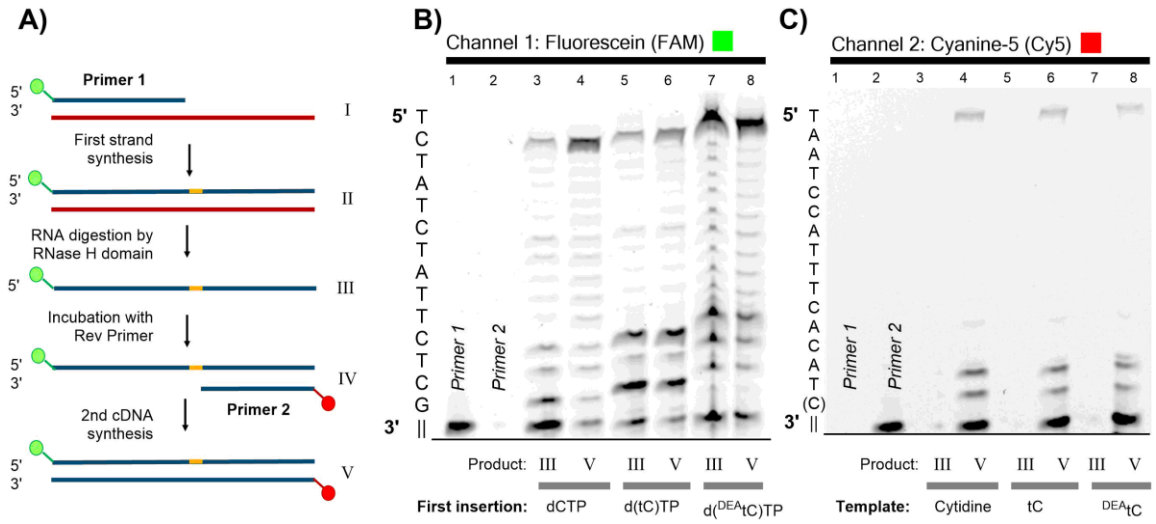


Figure 3.10. Modeling reverse transcription with nucleotide analogues. (A) The experiment involves DNA primer extension using an RNA template and a FAM-labeled 15-mer DNA primer 1 (I) to give heteroduplex (II), RNA digestion by HIV-1-RT's RNase H domain to give complementary DNA (III), and synthesis of a second DNA strand using an added Cy5-labeled primer 2 (IV) to give double-stranded DNA product (V). The yellow bar shows the site of tC, ^{DEA}tC, or natural C insertion. Denaturing gel electrophoresis of products (III) and (V) imaged in the FAM channel (B) shows the full-length, fluorescein labeled DNA in both. Imaging the same gel in the Cy5 channel (C) shows the second complementary strand of DNA only in products (V). Separate experiments were performed using dCTP, d(tC)TP, and d(DEAtC)TP, demonstrating comparable cDNA synthesis.

3.6 Discussion

The reverse transcription assays performed in the present study indicate that both d(tC)TP and d(^{DEA}tC)TP are viable substrates for insertion by AMV-RT, M-MLV-RT, and HIV-1-RT (**Table 3.4**). They are inserted in many cases with catalytic efficiency close to that observed when using natural dCTP, and not less than 9% of this natural efficiency, using both RNA and DNA templates. They are prone to mismatch incorporation across from adenine, and they induce some polymerase stalling when present in the primer 3′ end, especially with a DNA template. Here, we will examine the lessons that can be drawn from the polymerase studies described above, from both the perspectives of practical applications and chemical biology of reverse transcriptase enzymes.

Based on the relative rates of single nucleotide analogue insertion by the three RTs using the RNA_G template, AMV-RT has the least discrimination of analogues from natural dCTP and the catalytic efficiencies for analogue insertion are within an order of magnitude of dCTP when using each of M-MLV-RT and HIV-1-RT. Very similar results are obtained when using DNA_G: AMV-RT is unable to distinguish d(tC)TP and d(^{DEA}tC)TP from natural dCTP, and M-MLV-RT and HIV-1-RT have up to around 10-fold slower incorporation of these analogues as compared with the natural substrate. The additional steric size and the increased electron richness of ^{DEA}tC does not result in any consistent difference from tC in the kinetics of these matched insertions across from G. Previous work has demonstrated that the HIV-1 reverse transcriptase active site has more flexibility with an RNA template and is consequently more accommodating to enlarged nucleobase analogues during RNA-dependent DNA synthesis, as indicated by lower $K_{M,obs}$ values.[36] Our results are consistent with this past finding, but the trend does not extend to AMV-RT and M-MLV-RT. Ultimately, pre-steady-state kinetics will provide a more complete picture of discrimination, as the rates described here can be complicated by the slow kinetics associated with product dissociation. Nonetheless, the data provide qualitatively reliable evidence that the family of reverse transcriptase enzymes can incorporate these analogues.

Table 3.4 Efficacy summary of RT with d(tC)TP and d(^{DEA}tC)TP, and RNA or DNA templates.

	tC				^{DEA}tC			
RT	Match ins.	Match ext.[b]	Mismatch ins.	Mismatch ext.[b]	Match ins.	Match ext.[b]	Mismatch ins.	Mismatch ext.[b]
RNA								
AMV	+++	+	+[a]	+	+++	–	–[a]	–
M-MLV	+++	++	+++[a]	++	+++	+++	+++[a]	+++
HIV-1	+++	++	+[a]	+	++	+++	–[a]	–
DNA								
AMV	+++	++	++	+	+++	–	++	–
M-MLV	++	+	++	–	+++	+	+++	++
HIV-1	+++	+	++	–	+++	–	+++	+

Single insertion efficacies were assigned based on efficiency ≥10% comparable to natural dCTP (+++), <10% (++), ≤1% (+), or no apparent product on the gel (−). [a] For RNA mismatch insertion, intensity of bands corresponding to $^{(DEA)}$tC:A mismatch were compared to T:A band intensity. [b] For extension assessments, gel band intensity of 29mer product when using d($^{(DEA)}$tC)TP were compared to 29mer band intensity to assays supplied with dTTP.

As has been observed using parent tC and other A and B family polymerases, these tricyclic cytidines are prone to mismatch insertion across adenine, likely due to tautomerism to a Thymidine-like hydrogen bonding pattern.[10] Measurements of this infidelity using RNA templates were performed in read-through assays omitting dTTP. Mismatch insertion across from adenine could not be measured using AMV-RT and HIV-1-RT because of stalling, but M-MLV-RT is comparably efficient at forming tC:A and ^{DEA}tC:A base pairs as natural T:A base pairs. With templating DNA, we measured insertion efficiency across from A using the Michaelis–Menten model. Natural dCTP is inserted by all three RTs across from adenine at less than 0.01% the catalytic efficiency of dTTP insertion, but d(tC)TP is inserted with approximately 5% the catalytic efficiency and d(^{DEA}tC)TP up to 40%. With both analogues and all three RTs, $V_{max,obs}$ is at least 25% that of dTTP insertion across from A. Compared with d(tC)TP, d(^{DEA}tC)TP is more inserted incorrectly across from adenine because of a lower $K_{M,obs}$.

d(^{DEA}tC)TP is more hydrophobic than d(tC)TP, which may increase the affinity of the polymerase active site for this analogue. Interestingly, however, gas-phase DFT calculations (not shown) predict that ^{DEA}tC is less prone than tC to tautomerism, which would enable T-like base-pairing with adenine. Nonetheless, ^{DEA}tC is the analogue more prone to this form of infidelity.

The three RTs used in this study have nuanced structural differences, which may influence the efficiency with which they insert the tC analogues. There is a published crystal structure for HIV-1-RT and M-MLV-RT,[37,38] however no reported crystal structure for AMV-RT to date. Both AMV-RT and HIV-1-RT are heterodimeric proteins[37,39] while MLV-RT is monomeric,[38] and all have both RNA- and DNA-dependent DNA synthesis, and RNase H activity. Past research on the fidelity of RTs, particularly HIV-1-RT, led to some contrasting findings regarding the fidelity of this RT when polymerizing with a DNA or RNA template. Johnson, Anderson, and Loeb have found that HIV-1-RT maintains higher fidelity during polymerization using an RNA template,[40–42] while Goodman, Krauss, and Kool have reported higher fidelity with a DNA template.[36,43,44] The published ternary structures of HIV-1-RT show that the primer-template proximal to the polymerase active site adopts the A-form conformation, even if a dsDNA primer-template is used.[37,45] While there is no ternary crystal structure known for M-MLV-RT or AMV-RT, the conserved amino acid residues involved in coordination and catalysis are expected to be the same, irrespective of whether the template is RNA or DNA. HIV-1-RT does not require any minor groove hydrogen bonding between polymerase and substrate, while AMV-RT requires a single interaction with either the incoming nucleotide or the templating base, and M-MLV-RT requires minor groove interactions with the templating base and incoming nucleotide.[46] The added rings of the tC and ^{DEA}tC project into the major groove, but do not perturb minor groove hydrogen bonding. M-MLV-RT has been shown to have poor

compatibility with nonpolar isosteric nucleoside analogues.[47] In the present work, M-MLV-RT has kinetics of nucleotide insertion with tC analogues comparable to natural dCTP. While the added ring structure confers more hydrophobicity than natural cytosine, the analogues retain the canonical cytosine structure in the first ring that would allow for hydrogen bonding interactions between the pyrimidine O^2 and the terminal amide of residue Q190 in the M-MLV-RT active site, or Q151 for HIV-1-RT.[48] Hydrogen bonding between polymerase, O^2 of pyrimidines and *N*-3 of purines has been shown to be an important factor in catalytic efficiency, while structural changes to other parts of the nucleobase are more tolerated.[49] There is no reported crystal structure for AMV-RT, but the amino acid sequence has the conserved LPQG at the active site sequence, with Q155 available to form the essential hydrogen bond with the pyrimidine O^2.

Like other polymerases, as reverse transcriptases incorporate nucleotides during primer extension, the growing primer–template must be translocated within the polymerase active site to allow for the insertion of subsequent nucleotides. The structural perturbations of nucleotide analogues may inhibit translocation or induce conformational change that misaligns the 3′-OH for nucleophilic attack on the α-phosphate of the incoming nucleotide.[50] For this reason, even catalytically efficient single nucleotide insertion is no guarantee that further primer extension can continue. Because RTs perform both RNA- and DNA-templated DNA synthesis, read through efficiency may differ in each of these modes.

HIV-1-RT inserts d(tC)TP and d(^{DEA}tC)TP across from the RNA_G template and retains processivity (**Figure 3.5, lanes 8** and **9**), supporting the idea that the HIV-1-RT active site is more accommodating to structural perturbations with an RNA template.[36] M-MLV-RT generated the same full-length DNA product, but with considerable stalling when using d(^{DEA}tC)TP, as was also observed with natural dCTP and this enzyme. AMV-RT generated full-length product

following d(tC)TP insertion, but not following d(^{DEA}tC)TP insertion. Read through is considerably less efficient with DNA templates when using these RTs. After insertion of ^{DEA}tC, considerable stalling is observed with DNA_G and DNA_A and little to no final 29-mer product is observed with any of the RT variants used. d(tC)TP performed better, resulting in a small amount of complete primer extension with all three RTs.

In a more challenging variant on this experiment, we repeated the measurements using a template with four consecutive templating G residues beginning immediately at the 3′ end of the primer. Here, the four-G tract induced some polymerase stalling even when using natural dCTP, and d(tC)TP performed comparably to dCTP with each of the three RTs. In this assay, however, additional perturbation was clearly manifest when using d(^{DEA}tC)TP and all three polymerases. No more than two ^{DEA}tC residues were inserted consecutively, and primer extension ceased.

The complete reverse transcription process presents the greatest challenge when using nucleotide analogues because these analogues must function as enzyme-compatible incoming nucleotides and as templates, and they must also be compatible with RNase H activity. As would be expected from the robust insertion kinetics of d(tC)TP and d(^{DEA}tC)TP with HIV-1-RT, the enzyme that we chose for these experiments, synthesis of the first cDNA strand proceeded efficiently to generate full-length product, following some polymerase stalling after the insertion of the fluorescent analogue. RNase H activity was not assayed directly, but was implicitly indicated by the success of RT-mediated synthesis of the second DNA strand that replaces the original RNA template. This synthesis proceeded to yield full-length Cy5-labelled 29mer product, following some polymerase stalling that we observed when either tC or ^{DEA}tC served as the templating base. The extent of stalling using either analogue appears comparable to that observed with a natural cytidine template (**Figure 3.10C**), suggesting that polymerase activity in

this specific context may be the result of the limited duplex stability of the reverse primer-template complex rather than adverse interactions between HIV-1-RT and the templating analogue. Past studies have indicated that hydrophobic interactions largely influence primer-template binding to HIV-1-RT,[49] and residue P157 contacts the nucleobase and sugar of the third template base from the hanging end.[45,51] After tC or ^{DEA}tC has been translocated to this third position, stalling largely stops. This effect could result from the increased thermostability of the primer-template following two base incorporations or the greater hydrophobicity of the tricyclic ring structure, which may strengthen interactions with P157.

3.7 Conclusion

The fluorescent cytidine analogues d(tC)TP and d(^{DEA}tC)TP are substrates for insertion across from both DNA and RNA templates by the three reverse transcriptases HIV-1-RT, AMV-RT, and M-MLV-RT. In many cases, according to a Michaelis–Menten model, they are not significantly distinguished from natural dCTP during nucleotide insertion, and in no case is an analogue's rate of insertion across from G less than 9% the rate of natural dCTP. They are prone to mismatch insertion across from A with a tC:A base pair 4–7% as likely to form as a T:A base pair, and a ^{DEA}tC:A base pair 15–39% as likely, depending on polymerase and with a DNA template. This mispairing likely derives from the analogues' hydrophobicity and increased potential to adopt a thymine-like tautomer as compared with dCTP. Read-through was efficient when the analogues were inserted across from G in a RNA template with M-MLV-RT and HIV-1-RT, and was satisfactory with tC and AMV-RT. For M-MLV-RT, elongation after insertion of a tC: or ^{DEA}tC:A base pair was nearly as efficient as after insertion of a T:A pair. With a DNA template, read-through after analogue insertion across from G is less efficient, resulting in full-length product only when using d(tC)TP and HIV-1-RT or AMV-RT. Efficient elongation after

d(^{DEA}tC)TP insertion or analogue mismatch insertion across from adenine on a DNA template was not observed.

Our modeling study of a full reverse transcription cycle—RNA-templated DNA synthesis, RNA template degradation, and finally DNA-templated DNA synthesis—shows that all three steps succeed and full-length, fluorescently labeled dsDNA can be produced using either d(tC)TP and d(^{DEA}tC)TP and HIV-1-RT. Some polymerase stalling is observed when dGTP is inserted across from tC or ^{DEA}tC, but is not severe enough to prevent the formation of full-length product.

Collectively, these results point to future applications of fluorescent nucleotides in monitoring retroviral replication, and they invite further chemical biology studies of fluorescent nucleotide analogues. Overall, as a substrate analogue, tC performs better than ^{DEA}tC, due to the greater fidelity of its insertion and more efficient read-through following its insertion across from G on a DNA template. d(tC)TP is compatible with the entire process of dsDNA formation from an RNA template using HIV-1-RT. For this process to succeed with d(^{DEA}tC)TP, the analogue must not be inserted during DNA-templated polymerization, as elongation would cease. ^{DEA}tC is more electron-rich than tC, protrudes further into the direction of the major groove, and induces some local duplex destabilization. Future studies to determine how the electronic and structural differences of fluorescent nucleotides influence fidelity and read-through, leading to analogues that excel at both, will be greatly enabling as these nascent metabolic technologies emerge for biophysical studies in living cells.

3.8 Supporting Figures

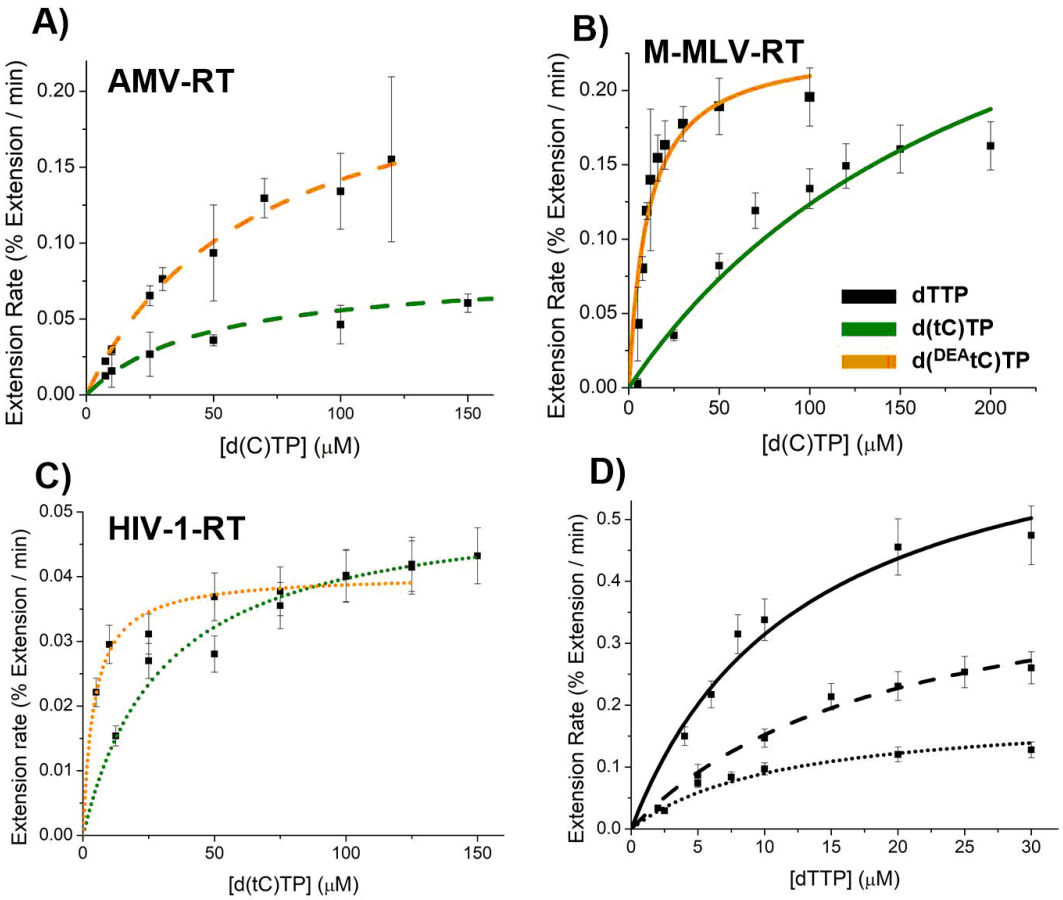


Figure S3.1. Michaelis-Menten plots for mismatch insertion with DNA_A. Plots are shown for **A**) AMV-RT, dashed, **B**) M-MLV-RT, solid, and **C**) HIV-1-RT, dotted. **D**) Plots using dTTP as control shown with black curves.

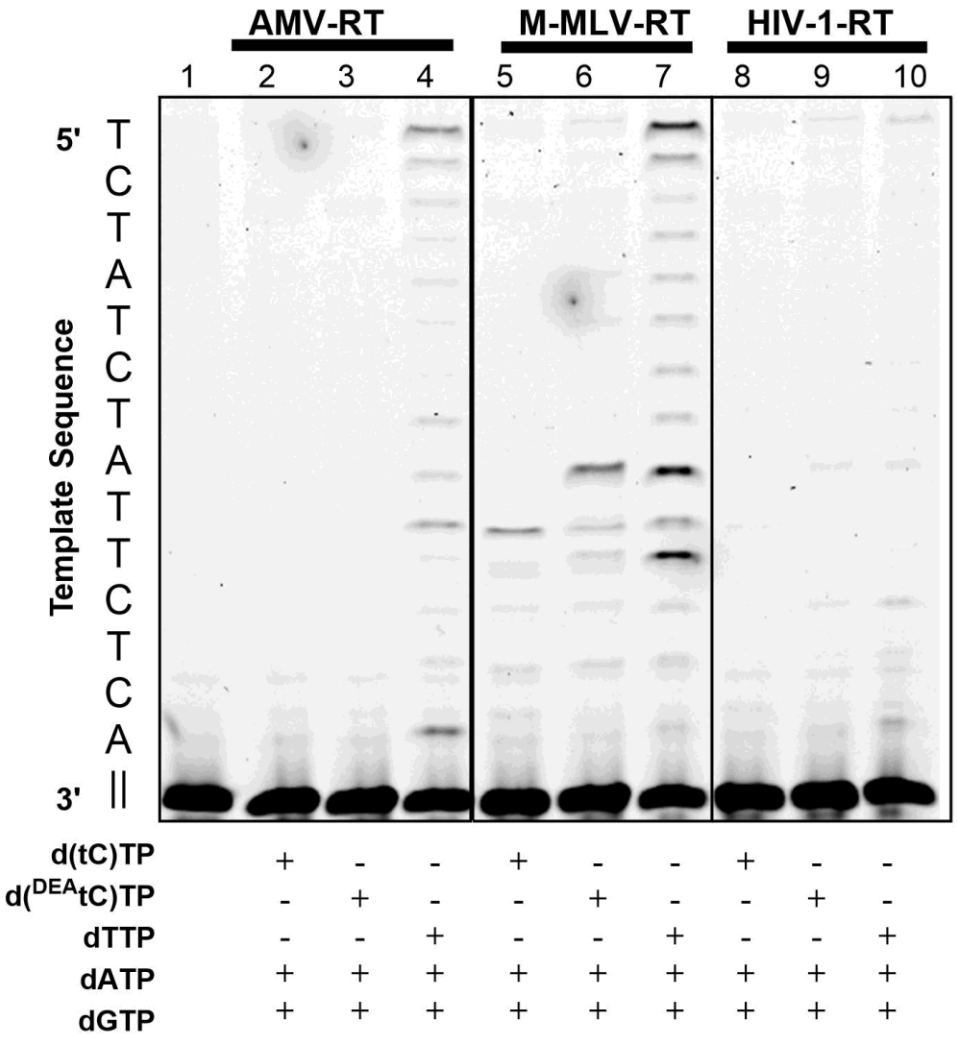


Figure S3.2. Read-through polymerization assay of DNA_A primer-template complex. Reactions contained indicated cytidine nucleotide and additional dNTPs to complete extension, at least after seven insertions, at which point the polymerase encounters another templating adenosine. Reactions were incubated with 120 μM d(tC)TP or 75 μM d(^{DEA}tC)TP for 40-60 min at 42 °C for AMV-RT and M-MLV-RT, 37 °C for HIV-1-RT.

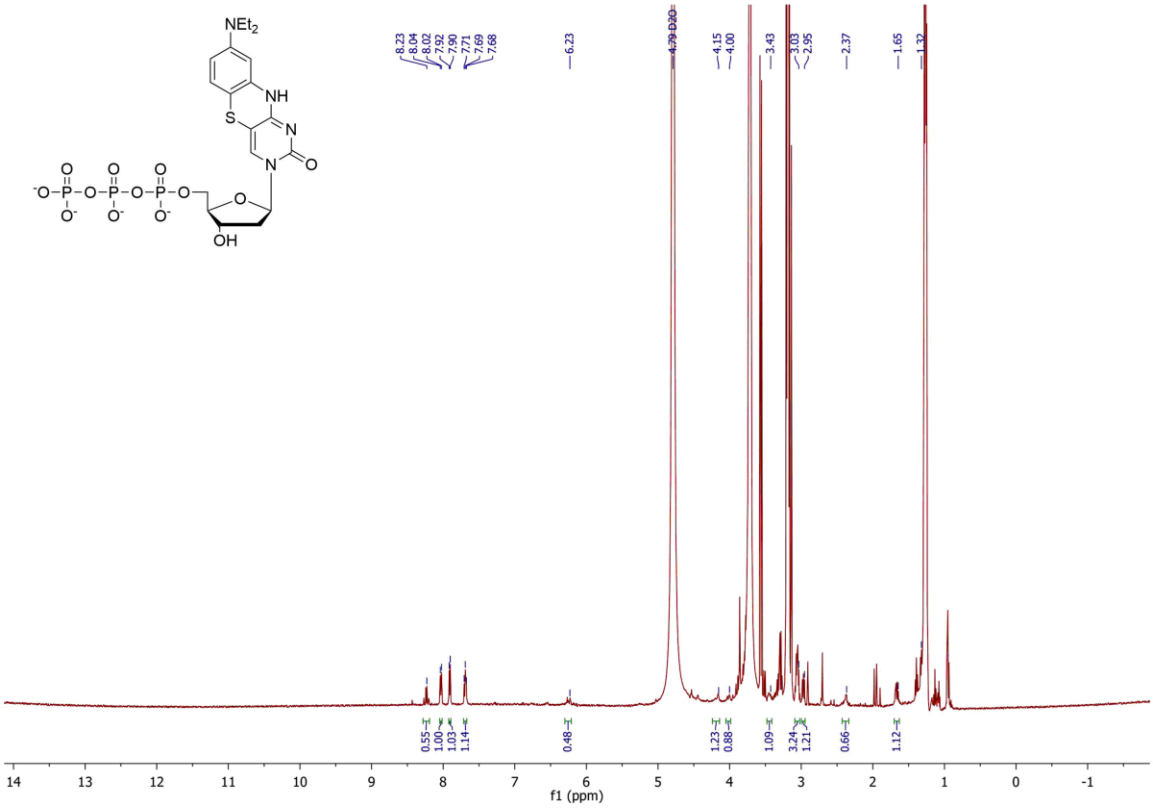


Figure S3.3 ^{1}H NMR spectrum of 8-Diethylamino-tC 2'-deoxy-β-$_{D}$-ribonucleoside-5'-*O*-triphosphate. Spectrum acquired in D_2O at 298 K in a 500 MHz NMR spectrometer.

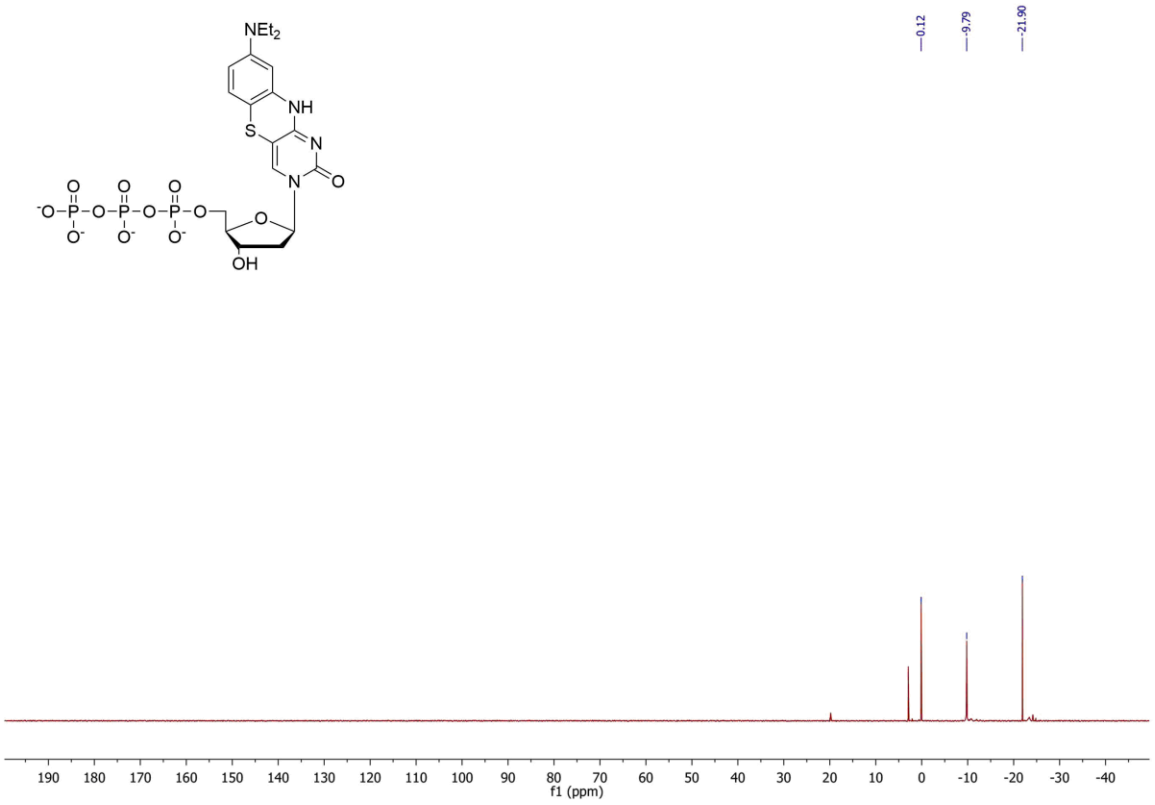
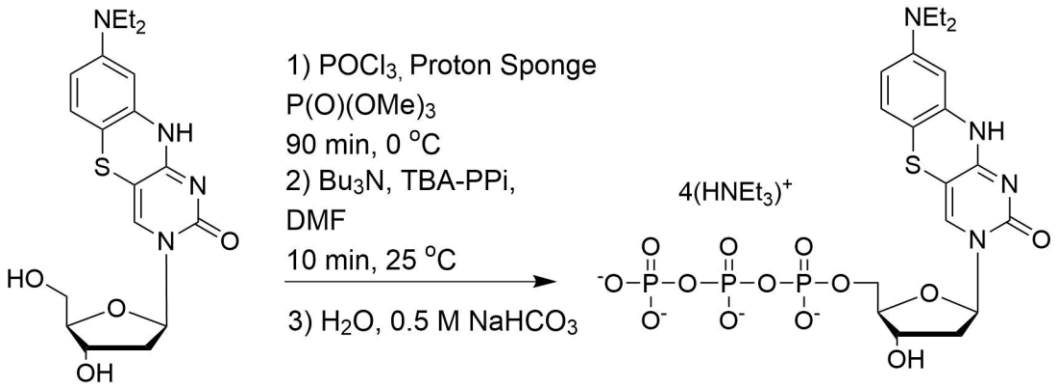


Figure S3.4 ^{31}P NMR spectrum of 8-Diethylamino-tC 2'-deoxy-β-$_D$-ribonucleoside-5'-*O*-triphosphate. Spectrum acquired in D_2O at 298 K in a 500 MHz NMR spectrometer.

3.9 Experimental Methods

Synthesis of tC-derived nucleoside triphosphates

Nucleoside analogues were synthesized and then converted to the corresponding triphosphates according to published procedures.[24,29,30] The crude products were purified by C18 RP-HPLC using a gradient of acetonitrile in 0.10 M TEAB buffer. Purified products were confirmed by ^{1}H and ^{31}P NMR before being lyophilized and stored at -20 °C.



Scheme 3.1 Synthesis of 8-Diethylamino-tC 2'-deoxy-β-$_D$-ribonucleoside-5'-triphosphate.

8-Diethylamino-tC 2'-deoxy-β-$_D$-ribonucleoside (24 mg, 0.060 mmol) and proton sponge (64 mg, 0.30 mmol) were dried in a 25-mL Schlenk tube under vacuum for at least two hours. The solids were suspended in 350 μL of trimethyl phosphate and stirred at 0 °C for 10 min before adding phosphorous(V) oxychloride (17 μL, 0.18 mmol) in two doses—8.5 μL per 20 min. The reaction was stirred for an additional 40 min at 0 °C and monitored by TLC using 10 % MeOH in DCM. Product formation was indicated by lack of migration up the plate and retained fluorescence at the spot site. The reaction vessel was then allowed to equilibrate to room temperature before adding tributylamine (84 μL, 0.36 mmol) and a solution of tributylammonium pyrophosphate (163 mg, 0.30 mmol) dissolved in 650 μL of anhydrous DMF. The reaction proceeded for 10 min before quenching the reaction with 5 mL of deionized water and 1 mL of 0.5 M $NaHCO_3$ pH 7.5 buffer. The solvent was removed by rotary evaporation and dried overnight using a vacuum pump. The dry crude was suspended in 0.1 M TEAB pH 7.5 buffer and purified using reversed-phase (C18) HPLC and a gradient of 20-25 % acetonitrile in 0.1 M TEAB pH 7.5 buffer. Eluted fractions containing product were pooled and the solvent was removed by lyophilization to yield product as a yellow solid as a triethylammonium salt. **^{1}H NMR (500 MHz, D_2O pH 7.5)** δ 8.23 (m, 1H), 8.02 (d, J=8.1 Hz, 1H), 7.90 (d, J=7.8 Hz, 1H), 7.68 (t, J=7.7 Hz, 1H), 6.23 (m, 1H), 4.15 (m, 1H), 4.00 (m, 1H), 3.43 (m, 1H), 3.05 (q, J=7.1

Hz, 4H), 2.95 (m, 1H), 2.37 (m, 1H), 1.65 (m, 1H), 1.32 (t, J=6.9 Hz, 6H). **^{31}P NMR (162 MHz, D_2O pH 7.5)** δ 0.12, -9.79, -21.90.

Preparation of stock samples

Lyophilized d(tC)TP and d(^{DEA}tC)TP were each suspended in 10 mM Tris buffer pH 7.9 and concentrations calculated by UV-vis absorbance before diluting to working concentrations. Molar absorptivity for analogues are $\varepsilon = 4700\ M^{-1}\ cm^{-1}$ at $\lambda = 377$ nm for parent tC, and $\varepsilon = 2700\ M^{-1}\ cm^{-1}$ at $\lambda = 395$ nm for ^{DEA}tC. AMV-RT and M-MLV-RT were purchased from New England Biolabs (NEB) and HIV-1-RT was purchased from Worthington Biochemical Corporation. Varying concentrations of enzyme were initially tested with 10-20 μM substrate to gauge extension over 15-30 min and determine working concentrations of reverse transcriptase by not exceeding 20 % extension. Enzymes were diluted with a solution of 1X buffer (50 mM Tris-HCl, 75 mM KCl, 3 mM $MgCl_2$, 10 mM DTT for MLV-RT; 50 mM Tris-OAc, 75 mM KOAc, 8 mM $Mg(OAc)_2$, 10 mM DTT for AMV-RT and HIV-1-RT), 50 ng mL^{-1} bovine serum albumin, and Ultrapure nuclease-free water. Buffer contents are based on what manufacturers provided. DNA primers (15-mer) and templates (29-mer) were purchased from Integrated DNA Technologies (IDT) with a fluorescein (FAM) label conjugated to the 5′ end of the primer sequence and suspended in nuclease-free water. Primer and template sequences were hybridized by heating at 90 °C for 5 min, 72 °C for 15 min, and 37 °C for 10 min. Hybridized primer-template complexes were confirmed on native PAGE.

Single insertion and mismatch assays

Conditions for single insertion and mismatch insertion assays included 0.5 μM of appropriate primer-template complex, 1X buffer specific to the enzyme, varying concentrations of dNTP and either 10 nM AMV-RT, 150 nM MLV-RT, or 9.25 nM HIV-1-RT to a final volume

of 20-30 μL. Solutions with primer-template, buffer, and substrate were prepared separately from a solution containing reverse transcriptase. Both solutions were pre-warmed at 42 °C for AMV-RT and MLV-RT or 37 °C for HIV-1-RT for 5 min before adding an equal volume of enzyme to primer-template and substrate mixture. A slightly higher incubation temperature for MLV-RT and AMV-RT was used as per the manufacturer's recommendation and higher reported activity at 42-45 °C.[155] Single insertion reactions proceeded for 2-30 min, depending on activity of the enzyme, at 42 °C or 37 °C and were terminated by adding two volumes of stop solution consisting of 90 % formamide, 50 mM EDTA, bromophenol blue, and xylene cyanol. Oligonucleotides were resolved on an 8 M urea denaturing gel with 1X TBE buffer and 15 % polyacrylamide. Gels were scanned on a GE Typhoon FLA 7000 and band intensities were measured using ImageQuant software (GE Healthcare Life Sciences). Resulting kinetics data were processed in Origin (OriginLab, Northampton, MA) using Hill1 non-linear regression to calculate kinetic parameters V_{MAX} and K_M. Individual data points were measured in 2-3 replicates and output graphs show standard deviation for individual points.

Read-through and reverse transcription assays

Read-through experiments were performed with similar parameters as single-insertion experiments and analyzed using the same instrumentation and data processing programs. Reactions were run either at 42 °C for AMV-RT and MLV-RT or 37 °C for HIV-1-RT for 30 minutes with a final concentration of 15 μM dNTPs or higher in mismatch read-through assays. Two volumes of stop solution were added and samples were resolved on an 8 M urea denaturing gel with 1X TBE buffer and 15 % polyacrylamide. To simulate the multi-step biochemical synthesis of cDNA from RNA, HIV-1-RT was incubated with 15/29-mer DNA-RNA heteroduplex and dNTPs to generate a 29-mer DNA template *in situ*. Higher concentrations of

HIV-1-RT were used than in our polymerase assays, to accelerate both the production of the first cDNA 29-mer and the degradation of the template RNA by the enzyme's RNase H activity. Final concentrations of first-strand cDNA synthesis were 1.0 μM RNA_G, 1X RT buffer, 30 μM dNTPs, and 75 nM HIV-1-RT. The reaction proceeded at 37 °C for 50 minutes. Half of the reaction solution was removed and added to two volumes of stop solution. The remaining sample half was then incubated with 1.1 equivalents of Cy5-labeled reverse primer (13-mer) and was incubated at 72 °C for 5 minutes, 56 °C for 5 minutes, and then 32 °C for 5 minutes to anneal the reverse primer to the enzymatically-synthesized template. Second-strand cDNA synthesis was performed by adding additional materials to a final concentration of 25 μM dNTPs, 1X buffer, and 55 nM HIV-1-RT, accounting for the quantity of moles of dNTP remaining from first-strand synthesis. This was incubated for 50 minutes at 32 °C before adding stop solution. This step needed to be run at a lower temperature due to the denaturation of the primer-template complex at 37 °C. Samples were run on denaturing PAGE.

3.10 Acknowledgments

The contents of Chapter 3 contains material that is similar to the material published in the following manuscript: Turner, M. B., and Purse B. W. Fluorescent tricyclic cytidine analogues as substrates for retroviral reverse transcriptases. *ChemPlusChem*. 2020, 85, 855-865. The dissertation author was the lead author of this paper. Helpful discussions with Christal Sohl and Gudrun Stengel are gratefully acknowledged.

3.11 References

(1) Zawada, Z.; Tatar, A.; Mocilac, P.; Buděšínský, M.; Kraus, T. *Angew. Chemie Int. Ed.* **2018**, *57* (31), 9891–9895.

(2) Wahba, A. S.; Azizi, F.; Deleavey, G. F.; Brown, C.; Robert, F.; Carrier, M.; Kalota, A.; Gewirtz, A. M.; Pelletier, J.; Hudson, R. H. E.; Damha, M. J. *ACS Chem. Biol.* **2011**, *6*

(9), 912–919.

(3) Mizrahi, R. A.; Shin, D.; Sinkeldam, R. W.; Phelps, K. J.; Fin, A.; Tantillo, D. J.; Tor, Y.; Beal, P. A. *Angew. Chemie* **2015**, *54* (30), 8713–8716.

(4) Stengel, G.; Gill, J. P.; Sandin, P.; Wilhelmsson, L. M.; Albinsson, B.; Nordén, B.; Millar, D. *Biochemistry* **2007**, *46* (43), 12289–12297.

(5) Frener, M.; Micura, R. *J. Am. Chem. Soc.* **2016**, *138* (11), 3627–3630.

(6) Sinkeldam, R. W.; Greco, N. J.; Tor, Y. *Chem. Rev.* **2010**, *110* (5), 2579–2619.

(7) Xu, W.; Chan, K. M.; Kool, E. T. *Nat. Chem.* **2017**, *9* (11), 1043–1055.

(8) Kosuri, S.; Church, G. M. *Nat. Methods* **2014**, *11*, 499.

(9) Sandin, P.; Stengel, G.; Ljungdahl, T.; Borjesson, K.; Macao, B.; Wilhelmsson, L. M. *Nucleic Acids Res.* **2009**, *37* (12), 3924–3933.

(10) Stengel, G.; Purse, B. W.; Wilhelmsson, L. M.; Urban, M.; Kuchta, R. D. *Biochemistry* **2009**, *48* (31), 7547–7555.

(11) Stengel, G.; Urban, M.; Purse, B. W.; Kuchta, R. D. *Anal. Chem.* **2009**, *81* (21), 9079–9085.

(12) Stengel, G.; Urban, M.; Purse, B. W.; Kuchta, R. D. *Anal. Chem.* **2010**, *82* (3), 1082–1089.

(13) Shin, D.; Sinkeldam, R. W.; Tor, Y. *J. Am. Chem. Soc.* **2011**, *133* (38), 14912–14915.

(14) McCoy, L. S.; Shin, D.; Tor, Y. *J. Am. Chem. Soc.* **2014**, *136* (43), 15176–15184.

(15) Li, Y.; Fin, A.; McCoy, L.; Tor, Y. *Angew. Chemie Int. Ed.* **2016**, 2–7.

(16) Passow, K. T.; Harki, D. A. *Org. Lett.* **2018**, *20* (14), 4310–4313.

(17) Hocek, M. *Acc. Chem. Res.* **2019**, *52* (6), 1730–1737.

(18) Anderson, J. P.; Angerer, B.; Loeb, L. A. *Biotechniques* **2005**, *38* (2), 257–264.

(19) Das, K.; Martinez, S. E.; DeStefano, J. J.; Arnold, E. *Proc. Natl. Acad. Sci.* **2019**, *116* (15), 7308 LP – 7313.

(20) Hu, W.; Hughes, S. H. *Cold Spring Harb. Perspect. Med.* **2012**, *2* (10), a006882.

(21) Roberts, J. D.; Bebenek, K.; Kunkel, T. A. *Science (80-.).* **1988**, *242* (4882), 1171 LP – 1173.

(22) De Wit, F.; Pillalamarri, S. R.; Sebastián-Martín, A.; Venkatesham, A.; Van Aerschot, A.; Debyser, Z. *J. Biol. Chem.* **2019**, *294* (4), 11863–11875.

(23) Lin, K.-Y. Y.; Jones, R. J.; Matteucci, M. *J. Am. Chem. Soc.* **1995**, *117* (13), 3873–3874.

(24) Teppang, K. L.; Lee, R. W.; Burns, D. D.; Turner, M. B.; Lokensgard, M. E.; Cooksy, A. L.; Purse, B. W. *Chem. - A Eur. J.* **2019**, *25*, 1249–1259.

(25) Wilhelmsson, L. M.; Sandin, P.; Holmén, A.; Albinsson, B.; Lincoln, P.; Nordén, B. *J. Phys. Chem. B* **2003**, *107* (34), 9094–9101.

(26) Sandin, P.; Wilhelmsson, L. M.; Lincoln, P.; Powers, V. E. C.; Brown, T.; Albinsson, B. *Nucleic Acids Res.* **2005**, *33* (16), 5019–5025.

(27) Burns, D. D.; Teppang, K. L.; Lee, R. W.; Lokensgard, M. E.; Purse, B. W. *J. Am. Chem. Soc.* **2017**, *139* (4), 1372–1375.

(28) Sandin, P.; Lincoln, P.; Brown, T.; Wilhelmsson, L. M. *Nat. Protoc.* **2007**, *2* (3), 615–623.

(29) Turner, M. B.; Anderson, B. A.; Samaan, G. N.; Coste, M.; Burns, D. D.; Purse, B. W. *Curr. Protoc. Nucleic Acid Chem.* **2018**, e59.

(30) Burgess, K.; Cook, D. *Chem. Rev.* **2000**, *100* (6), 2047–2059.

(31) Reardon, J. E.; Miller, W. H. *J. Biol. Chem.* **1990**, *265* (33), 20302–20307.

(32) Lim, S. E.; Ponamarev, M. V; Longley, M. J.; Copeland, W. C. *J. Mol. Biol.* **2003**, *329* (1), 45–57.

(33) Menéndez-Arias, L. *Viruses* **2009**, *1* (3), 1137–1165.

(34) Sarafianos, S. G.; Marchand, B.; Das, K.; Himmel, D. M.; Parniak, M. A.; Hughes, S. H.; Arnold, E. *J. Mol. Biol.* **2009**, *385* (3), 693–713.

(35) Acheson, N. H. *Fundamentals of Molecular Virology, 2nd Edition*, 2nd ed.; John Wiley & Sons, Inc., 2011.

(36) Silverman, A. P.; Garforth, S. J.; Prasad, V. R.; Kool, E. T. *Biochemistry* **2008**, *47* (16), 4800–4807.

(37) Jacobo-Molina, a; Ding, J.; Nanni, R. G.; Clark, a D.; Lu, X.; Tantillo, C.; Williams, R. L.; Kamer, G.; Ferris, a L.; Clark, P. *Proc. Natl. Acad. Sci. U. S. A.* **1993**, *90* (13), 6320–6324.

(38) Das, D.; Georgiadis, M. M. *Structure* **2004**, *12* (5), 819–829.

(39) Grandgenetts, D.; Quinn, T.; Hippenmeyer, P. J.; Oroszlanll, S. *J. Biol. Chem.* **1985**, *260* (14), 8243–8249.

(40) Kerr, S. G.; Anderson, K. S. *Biochemistry* **1997**, *36* (46), 14056–14063.

(41) Ji, J.; Loeb, L. A. *Biochemistry* **1992**, *31* (4), 954–958.

(42) Kati, W. M.; Johnson, K. A.; Jerva, L. F.; Anderson, K. S.; Science, T. A. *J. Biol. Chem.* **1992**, *267* (36), 25988–25997.

(43) Yu, H.; Goodman, M. *J. Biol. Chem.* **1992**, *267* (15), 10888–10896.

(44) Hubner, A.; Kruhoffer, M.; Grosse, F.; Krauss, G. *Journal of molecular biology.* London : 1992, pp 595–600.

(45) Huang, H.; Chopra, R.; Verdine, G. L.; Harrison, S. C. *Science (80-.).* **1998**, *282* (5394), 1669 LP – 1675.

(46) Morales, J. C.; Kool, E. T. *Biochemistry* **2000**, *39* (42), 12979–12988.

(47) Morales, J. C.; Kool, E. T. *J. Am. Chem. Soc.* **2000**, *122* (6), 1001–1007.

(48) Georgiadis, M. M.; Jessen, S. M.; Ogata, C. M.; Telesnitsky, A.; Goff, S. P.; Hendrickson, W. A. *Structure* **1995**, *3* (9), 879–892.

(49) Klarmann, G. J.; Eisenhauer, B. M.; Zhang, Y.; Gotte, M.; Pata, J. D.; Chatterjee, D. K.; Hecht, S. M.; Le Grice, S. F. J. *Biochemistry* **2007**, *46* (8), 2118–2126.

(50) Sarafianos, S. G.; Clark, A. D.; Tuske, S.; Squire, C. J.; Das, K.; Sheng, D.; Ilankumaran, P.; Ramesha, A. R.; Kroth, H.; Sayer, J. M.; Jerina, D. M.; Boyer, P. L.; Hughes, S. H.; Arnold, E. *J. Biol. Chem.* **2003**, *278* (18), 16280–16288.

(51) Klarmann, G. J.; Smith, R. A.; Schinazi, R. F.; North, T. W.; Preston, B. D. *J. Biol. Chem.* **2000**, *275* (1), 359–366.

(52) Yasukawa, K.; Mizuno, M.; Inouye, K. *J. Biochem.* **2009**, *145* (3), 315–324.

Chapter 4

Live-cell Studies using tC nucleotides with Prokaryotic and Eukaryotic Cells

4.1 Introduction

Fluorescent labeling of intracellular structures has yielded pivotal results in understanding basic cellular phenomena. Studying nucleic acid metabolism and dynamics in living cells provides essential information on gene expression, regulation, and replication.[1] Conventional fluorophores with high quantum yields offer great sensitivity but their molecular structures may induce adverse effects to natural cellular function.[2–4] Intrinsically fluorescent nucleoside analogues may function as powerful biophysical tools for studying these processes in living cells, monitoring changes in real time, with high resolution and molecular structure minimally perturbing native biology.[5] A 2017 review stated that *in vivo* and live cell applications with fluorescent nucleobase analogues are currently in the stages of early development.[6] Recent examples include fluorescent *N*,*N*-dimethylaniline-2'-deoxythymidine analogue that reports on Hg-induced DNA metabolism,[7] fluorescent etheno-cytidine and uridine-furan nucleosides measuring the expression of nucleoside transporter membrane proteins,[8] and 1-methyladenine base-stacked over emissive pyrene to monitor DNA repair in human cells.[9] Characterizing the live cell applications of fluorescent nucleoside analogues beyond their *in vitro* biochemical capabilities enables researchers to better understand the diverse functions of nucleic acids.

Prokaryotes like bacteria are excellent model organisms to use as a starting platform for studies with fluorescent nucleoside. In bacteria, during each round of binary fission (reproductive division), a new set of DNA is synthesized from available cytosolic 2'-deoxyribose nucleoside triphosphates. Bacterial genomes are typically circular chromosomes averaging ~3.87×10^6 base-pairs but may contain comparatively smaller chromosomes called plasmids, averaging ~78.9×10^3

base-pairs.[10] Studying nucleic acid metabolism in bacterial cells, or phage-infected bacteria, with nucleoside analogues can be challenging as many bacteria synthesize their DNA/RNA building blocks *de novo* and as such do not necessarily rely on extracellular availability of nucleotides for metabolism or have lost altogether nucleoside salvage pathways through eveolution.[11] Work by the Romesberg research group introduced an efficient strategy for transporting nucleoside triphosphates to cells by transforming competent strains to express the *Phaeodactylum tricornutum* nucleoside triphosphate transporter (*Pt*NTT2) membrane protein.[12] The membrane protein originally isolated from a planktonic diatom species actively imports extracellular (2'-deoxy)ribonucleoside triphosphates into the cell.[13] This technology is adapted here to facilitate metabolic labeling of *E. coli* nucleic acids using tC-derived nucleotides, as shown in **Figure 4.1**.

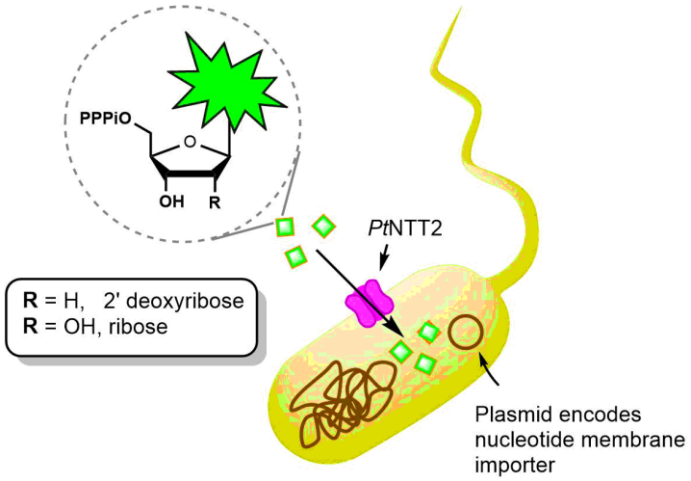
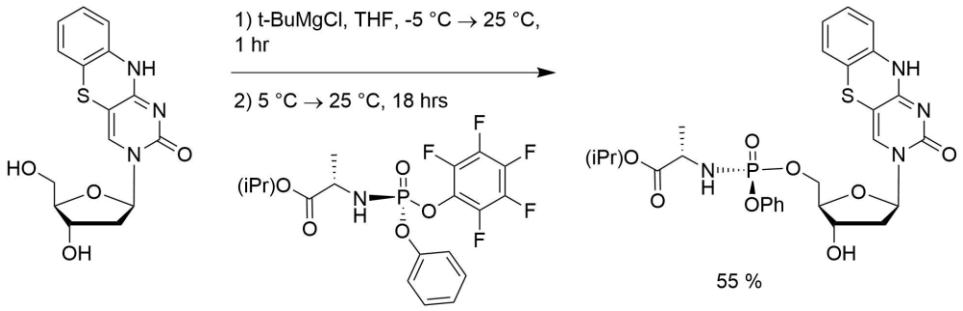


Figure 4.1. Strategy to deliver fluorescent nucleotides to bacterial cells. *E. coli* cells are transformed with pSC plasmid encoding *P. tricornutum* nucleoside triphosphate transporter II and chloramphenicol resistance.

Nucleoside analogues applications with eukaryotic cells have largely included developing effective inhibitors for viral pathogens such as HIV-1, hepatitis B, and Ebola, by targeting the virally encoded polymerase.[14] Even without inhibitory purposes, the first obstacle using nucleoside analogues in eukaryotic cells is delivering the unnatural substrates into cytoplasm. Negative charges on the phosphate moieties of nucleotides restrict passive diffusion across

phospholipid membranes, while administering free nucleosides may result in some passive diffusion. However the obstacle then becomes recognition by endogenous kinases to phosphorylate the chemically modified nucleosides into the corresponding triphosphate.[15] Nucleoside phosphoramidates, originally developed by Christopher McGuigan *et al*. in the 1990s,[16] are inactive nucleoside monophosphates with lipophilic groups masking the otherwise anionic charges, thereby facilitating passive diffusion across cell membranes.[17–19] **Scheme 4.1** provides an example converting the tC 2'-deoxyribose nucleoside into the corresponding inactive nucleotide. The masking groups shown in **Scheme in 4.1** are identical to the groups used on β-$_D$-2'-deoxy-2'-fluoro-2'-methylcytidine, an FDA-approved inhibitor to treat hepatitis C.[20–22] After entering cells, phosphoramidates undergo a series of enzymatic and intramolecular reactions, generating the corresponding triphosphate and substrate for intracellular polymerases.



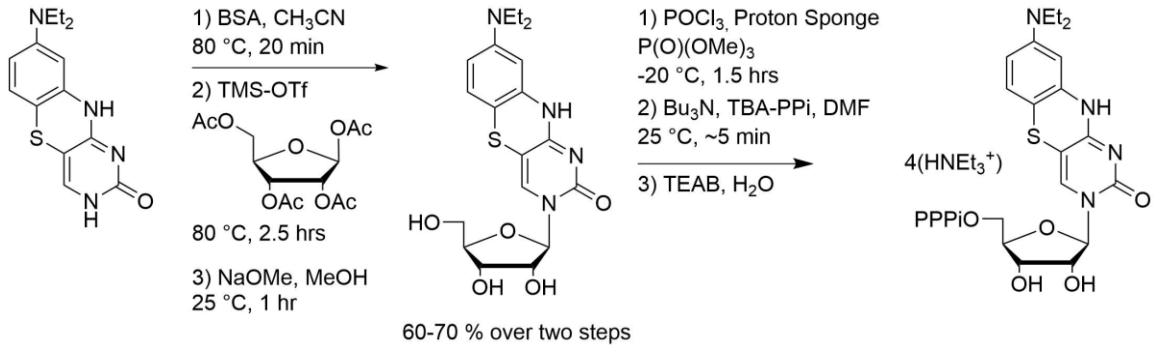
Scheme 4.1. Synthesis of tC 2'-deoxyribose nucleoside phosphoramidate.

In this present investigation, fluorescent tC molecules are delivered to *E. coli* as representative prokaryotes and three types of human cells, as an initial proof of concept demonstrating that the fluorescent nucleotides can be delivered to live cells as valuable tools for studying DNA/RNA dynamics and metabolism. Three human cell lines were used: Huh7 carcinoma hepatocytes (liver cells), surface-adherent 293T lymphoma helper T-cells, and non-adherent SUPT1 lymphoma T-cells. Constitutively fluorescent parent tC and the turn-on probe

^{DEA}tC are used. Preliminary optical measurements recorded either using a 96-well plate reader or flow cytometer indicate the fluorescent metabolites remain in the cell cytosol for about 24 hours and that cytotoxicity is not a major concern for nucleotide concentrations <150 μM. Subsequent investigations will likely focus on confirming and quantifying the kinetics of DNA/RNA metabolism using the unnatural substrates ideally to generate fluorescently labeled transcripts.

4.2 Metabolic labeling of transformed prokaryotes

For this study, ribonucleotides (rNTPs) were used in addition to 2'-deoxy ribonucleotides (dNTPs) to fluorescently label intracellular RNA transcripts, such as messenger, ribosomal, and transfer RNA. The syntheses for novel ^{DEA}tC ribonucleoside and r(^{DEA}tC)TP are shown in **Scheme 4.2**. Ribose nucleoside synthesis is comparatively simpler than synthesis of 2'-deoxyribose nucleosides, particularly because ribose analogues do not require separation of α/β diastereomers, as with 2'-deoyribose counterparts, by virtue of the β-$_D$-ribofuranose tetraacetate structure and anchimeric assistance ribosylation. [23] Like with standard silyl-Hilbert-Johnson conditions, bis(trimethylsilyl)acetamide appends a TMS to the cytosine nucleobase *O2* and abstracts the amide proton from *N1*, generating a strong nucleophile.[24] In the same reaction vessel, trimethylsilyl trifluoromethanesulfonate (TMS-OTf) attaches to the *O*-linked acetyl on *C-1*' which consequently generates a highly labile leaving group. Contributions from the ribose oxygen forming an oxonium cation and the 2'-*O*-acetyl group serving as a transient nucleophile, result in the nucleobase coupling step to generate only the β-anomer. Both nucleoside types were converted to the corresponding nucleoside triphosphate using Ludwig phosphorylation chemistry.[25] Purified substrates were then added to bacterial cultures for metabolic labeling.



Scheme 4.2. Synthesis of ^{DEA}tC ribose nucleoside triphosphate.

Transformed *E. coli* (BL-21 Gold) cultures with the *PtNTT2* plasmid were grown to an OD_{600} of 0.40 – 0.90, corresponding to log phase of bacterial growth, at 37 °C before being transferred to 96-well plates and incubated with varying concentrations of nucleoside triphosphate overnight. Bacterial growth follows an exponential model during the log phase of their population expansion and the time required to double an initial quantity of cells is the generational time. Media supplied with chloramphenicol was used to select for *Pt*NTT2^{+} cells. To assess persistent lability of fluorescent nucleotides within bacteria, after 24 – 48 hours, the liquid media from each individual well was exchanged for fresh growth media, while retaining the same cell populations. Recording the OD_{600} at 24-hour time intervals provided an approximation for cell density, with the assumption that OD_{600} 1.0 = $8x10^8$ cells.[26] After pelleting and washing the cells after 24 and 48 hours, the optical density and emission intensity were recorded. Procedural details are further described in the experimental methods section. Quantified results including logarithmic growth rates (*k*) and percent reductions in fluorescence intensity from latency experiments are shown in **Table 4.1** for incubation with tC 2'-deoxribose nucleoside triphosphates and **Table 4.2** for fluorescent ribonucleotides. During the logarithmic phase of population growth [$N_t=N_0 2^{kt}$], *k* modulates the specific growth rate.

Table 4.1. Fluorescence latency with tC 2'-deoxyribose nucleotides.

[d(tC)TP] (μM)	At 24 Hours		At 48 Hours	
	Log Growth Rate ($x10^{-3}$ hr^{-1})[a]	**Intensity Reduction (%)[b]**	**Log Growth Rate ($x10^{-3}$ hr^{-1})[a]**	**Intensity Reduction (%)[c]**
300	0.365	44.7	52.2	96.1
250	0.228	50.1	49.4	95.1
200	3.06	60.3	58.4	94.9
150	14.3	76.7	54.5	94.3
100	24.6	88.1	41.4	91.5
50	24.4	85.6	41.8	91.6
0	26.5	--	35.7	--

Averaged values are from duplicate experiments (n=2). *a*) Growth rate constants are calculated from log of cell population, estimated from OD_{600} recordings. *b*) Fluorescence reductions are calculated as the ratio of intensity after 24 hours to initial intensity for a given nucleotide concentration, or *c*) after 48 hours.

From the quantified results, using dtCTP at 100 μM concentration or lower does not impose high cytotoxic concerns, comparing the logarithmic growth constants of treated cells with non-treated bacterial cells. At 150 μM there is a 54.0 % reduction in growth rate but the value remains within the same order of magnitude as the negative control cells, while at higher nucleotide concentrations cytotoxicity becomes more concerning. The doubling time varies among bacterial species and the concentration of added substrates also affects growth kinetics.[27] For *E. coli* strains the doubling time is about 20 min (0.33 hours),[28] and thus over the course of 24 hours, the *Pt*NTT2$^+$ populations would be expected to have undergone approximately 72 generations. After 48 hours, the average growth rate among treated cells are actually higher than untreated cells, however fluorescence has largely cleared from cell populations. The correlation between emission intensity and growth rate is an expected result as higher nucleotide concentrations may inhibit bacterial growth but after 48 hours cells have cleared available dtCTP, through metabolism or other removal mechanism. Control experiments with *Pt*NTT2$^+$ *E. coli* demonstrated no internal fluorescence when cells were incubated with nucleoside

monomers, indicating that transport of tC molecules was accomplished using the triphosphate species.

Table 4.2. Fluorescence latency with tC ribose nucleotides.

[r(tC)TP] (μM)	At 24 Hours		At 48 Hours	
	Log Growth Rate ($x10^{-3}$ hr^{-1})[a]	**Intensity Reduction (%)**	**Log Growth Rate ($x10^{-3}$ hr^{-1})[a]**	**Intensity Reduction (%)**
300	27.2	93.5	24.55	99.6
250	32.4	93.9	37.98	99.2
200	33.9	94.7	35.09	99.6
150	33.6	94.3	35.61	99.7
100	31.6	96.1	39.12	99.8
50	28.3	85.1	36.72	99.8
0	23.9	--	43.96	--

Averaged values are from duplicate experiments (n=2). *a*) Growth rate constants are calculated from log of cell populations, estimated from OD_{600} recordings.

As parent tC is constitutively fluorescent, similar incubation experiments were performed with the nucleoside probe ^{DEA}tC, whose fluorescence would be more suggestive of metabolic incorporation. The quantitative data reporting log growth rates and ratios of emission after 24 hours to initial fluorescence are presented in **Table 4.3**. Though similar concentrations of tC and ^{DEA}tC nucleotide substrate were used for these experiments, ^{DEA}tC conferred less cytotoxicity on bacterial growth than the parent analogue, even at 300 μM. This was the case when using either 2'-deoxyribonucleotides or ribonucleotides. The average increase in brightness ranged from 60 – 70% from 2'-deoxyribonucleotides among the different concentrations and marginally greater fluorescence was unexpectedly observed with lower concentrations. When fluorescent ribonucleotides were used, there was a larger span of fluorescence ratios, indicating more turn-on with the higher r(^{DEA}tC)TP concentrations, ranging 10 – 79% increased brightness. Initially, we had suspected that fluorescence turn-on resulting from RNA metabolism would yield greater increases in brightness compared to DNA metabolic labeling, however from the

current experiments the results between the two nucleotides are similar. While the intracellular copy number of RNA transcripts (e.g. messenger, ribosomal, transfer) largely exceeds their DNA counterparts,[29] RNA turnover rates will be higher during a single cell's lifetime,[30] while DNA is essentially replicated once per cell division. Therefore, we might observe similar fluorescence values between the 2'-deoxyribose and ribonucleotides because the latter are simply metabolized and degraded faster. Another possible factor to elevated fluorescence intensity with either ^{DEA}tC nucleotide involves non-specific hydrophobic effects, such as aggregation or associations with phospholipid membranes. Similar to parent tC metabolic labeling, quantitative mass spectrometry data from extracted and digested DNA/RNA would effectively confirm the incorporation of the fluorescent nucleotides into bacterial nucleic acids.

Table 4.3. Fluorescence latency with ^{DEA}tC nucleotides after 24 hours.

	2'-deoxyribonucleotide		**ribonucleotide**	
[(^{DEA}tC)TP] (μM)	**Log Growth Rate (hr^{-1})*a***	**Emission Ratio*b***	**Log Growth Rate (hr^{-1})*a***	**Emission Ratio*b***
300	0.0420	1.60±0.034	0.0394	1.79±0.030
250	0.0373	1.61±0.024	0.0322	1.67±0.0023
200	0.0365	1.66±0.0070	0.0358	1.56±0.013
150	0.0355	1.71±0.014	0.0350	1.46±0.022
100	0.0350	1.71±0.026	0.0363	1.30±0.0068
50	0.0363	1.76±0.034	0.0350	1.10±0.014
0	0.0375	--	0.0371	--

Averaged values are from duplicate experiments (n=2). *a*) Growth rate constants are calculated from log of cell populations, estimated from OD_{600} recordings. b) Emission ratios are calculated by comparing fluorescence at 510 nm (with λ_{ex} = 405 nm) of cell populations to fluorescence before 24-hour incubation interval.

4.3 Delivering fluorescent nucleotides to eukaryotic cells

The synthesis of the inactive nucleotide is reported to generate a single diastereomer from the pure diastereomerically pure phosphoramidate reagent,[22] however both diastereomers are

biologically active albeit with different metabolic kinetics.[31] The metabolic pathway of conversion is shown in **Figure 4.2**. After diffusing across the lipophilic cell membrane, the initial hydrolysis of the isopropyl ester by cathepsin A and/or carboxylesterase 1 is essential for mitigating passive diffusion and initiating the subsequent transformations. The anionic charge on the carboxylate restricts the nucleotide from passively diffusing across the outer lipid membrane and back out into the extracellular environment, effectively containing the metabolite within the cell. The following step is an enzyme-independent intramolecular addition-elimination reaction involving 5-member ring formation and release of phenolate anion. The subsequent enzymatic reaction involves hydrolysis performed by Hint1, a histidine triad protein acting on the α-phosphate of nucleotides,[32] liberating D-alanine as a side product. The resulting nucleoside monophosphate subsequently undergoes two kinase-mediated phosphorylation steps to yield the complete nucleoside triphosphate molecule.[33–35]

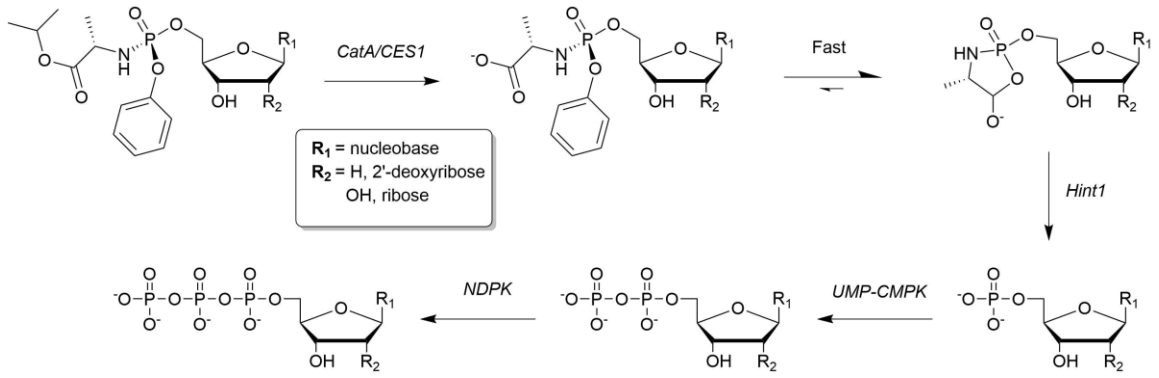


Figure 4.2. Metabolic conversion of nucleoside phosphoramidate to triphosphate. Sequence of steps occurs in cytosol after passive diffusion across cell membrane.

Microscopy images, taken using an Amnis ImageStream MkII flow cytometer (Luminex Corporation, Austin TX), of carcinoma human liver cells, Huh7, incubated with tC phosphoramidate demonstrated intracellular fluorescence distinct from background cellular autofluorescence. **Figure 4.3** presents representative images taken simultaneous with flow

cytometry quantification. Some cells displayed near homogenous distribution of fluorescence, suggesting that the nucleotide was distributed evenly throughout the cell (first row), however others displayed punctate or localized regions of fluorescence (second row). The fluorescent image in the third row indicates distributed fluorescence but with a large localized intensity. To confirm if the fluorescence in this case is indicative of nucleotide incorporation into genomic DNA, a nuclear stain control with spectral properties distinct from tC would need to be added. An alternative approach includes DNA extraction, purification, and digestion for LC-MS analysis,[36] however this method does not distinguish nuclear DNA from mitochondrial DNA. Both DNA polymerase ε, the main polymerase during S-phase replication, and pol γ, replicating mitochondrial DNA, possess proofreading capabilities for DNA repair.[37,38] While polymerases lacking exonuclease proofreading exhibit comparable insertion kinetics for the tC analogues compared to natural cytidine,[39–41] similar investigation with proofreading polymerases has not yet been conducted. Punctate fluorescence intensity as observed in **Figure 4.X** row 2, may demonstrate metabolic labeling of mitochondrial DNA with some incorporation in a larger structure corresponding to nuclear genomic DNA. Other possibilities may include cytosolic aggregations or sequestering to endosomal compartments of tC nucleotide intermediate metabolites. Nonetheless, the microscopy images indicate effective delivery of the fluorescent nucleotides into cells, thereby overcoming the first major obstacle of applications for studying nucleic acid metabolism involving nucleoside analogues.

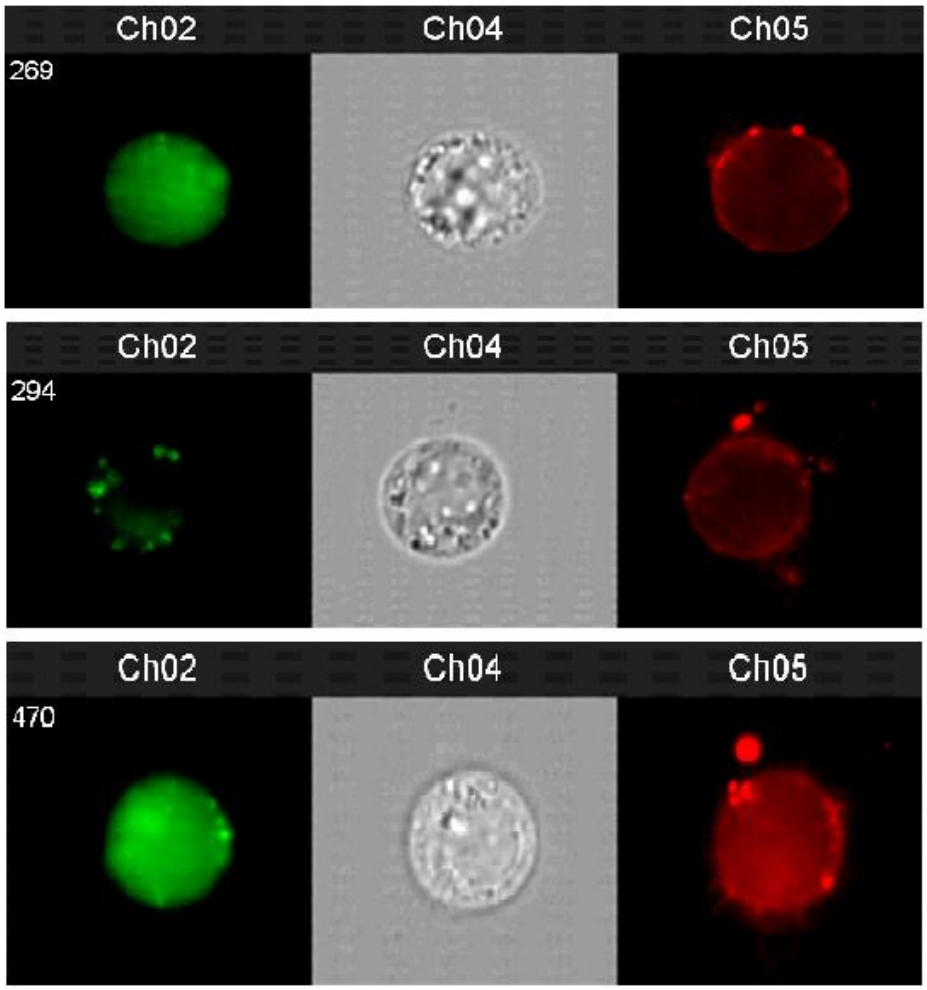


Figure 4.3. Select single-cell fluorescence microscopy of Huh7 cells. Channels from left to right are fluorescence of tC fluorophore (λ_{ex} = 375 nm), brightfield, and cellular autofluorescence (λ_{ex} = 642 nm).

From fluorescence-activated cell sorting (FACS) performed by a FACS Aria III flow cytometer (BD Scientific, San Diego CA), we were able to group populations of Huh7 and human T-cells based on their fluorescence intensity. Histogram plots of the data are presented in **Figure 4.4A** for human hepatocyte carcinoma cells and in **Figure 4.4B** for two types of $CD4^+$ lymphoma T-cells. In both experiments, treated cells were run in biological duplicate samples and the population quantities based fluorescence intensity were averaged. The pro-nucleotide strategy was expected to be highly successful with the Huh7 cells since the same phosphate

protecting groups were used to target liver cells. Indeed, populations of Huh7 cells experienced nearly 100-fold increase in the median fluorescence intensity of cells incubated with phosphoramidate compared to non-treated control populations after 24 hours (**Figure 4.4A**). Furthermore, the lowest intensity of phosphoramidate-treated cells were about 3.3-fold brighter than the highest intensity of cellular fluorescence. $CD4^+$ T-cells like SUPT1 and 293T likewise express the necessary enzymes to enzymatically convert phosphoramidates into nucleoside triphosphates.[42,43] In both lymphoma cell lines, we observed approximately 30-fold increase in median fluorescence of treated populations compared to control cells (**Figure 4.4B**, upper panel), with no significant variation between the surface-adherent and freely suspended lymphocytes. Similar to prokaryotic cell experiments, fluorescence intensity returns to baseline at 48 hours. Controls with tC nucleoside monomer demonstrated no change in fluorescence intensity compared to baseline, and therefore provide evidence affirming the importance of pro-nucleotide molecular structure for successful delivery.

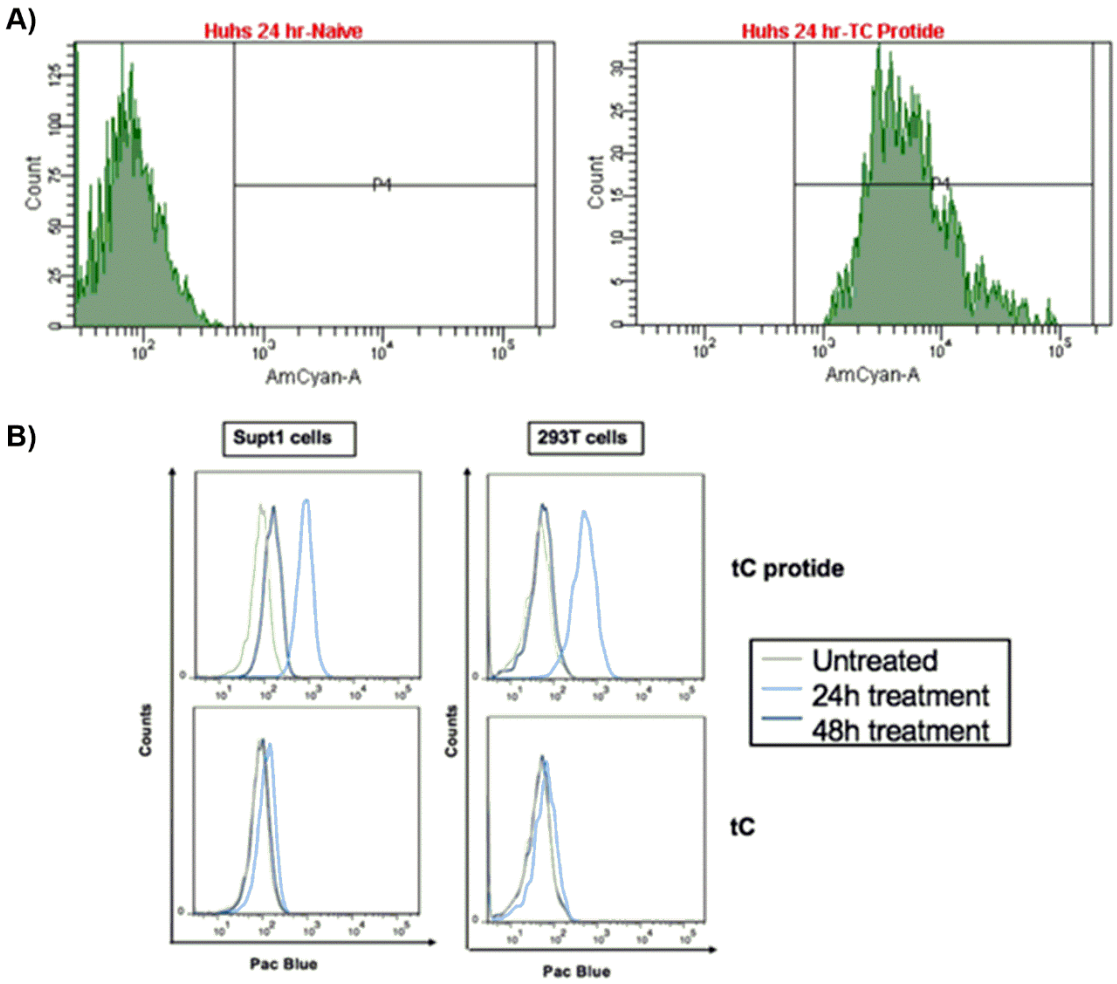


Figure 4.4. Flow cytometry histograms from human cells. Cell populations were quantified by FACS. Excitation laser set to 405 nm (AmCyan or Pacific Blue fluorophore settings on the instruments) and emission collection at 485 nm.

4.3 Future metabolic labeling studies

The results from these preliminary live-cell studies provide evidence that tC molecules can be integrated into the natural biology of living cells. Future endeavors will likely focus on quantifying the efficiency by which cellular polymerases metabolize the unnatural nucleotide analogues into newly synthesized DNA/RNA transcripts. For example, in the metabolic conversion of phosphoramidates to triphosphates, the first phosphorylation step of tC nucleoside monophosphate may be challenging for UMP-CMP kinase, as the tC nucleobase architecture is

larger than natural pyrimidines. As such, metabolomics would be suitable in confirming the presence of intermediate metabolites such as the tC nucleoside diphosphate and triphosphate. Confirming incorporation of tC nucleotides in prokaryotic DNA can potentially be accomplished using fourth-generation sequencing technology that threads single-stranded DNA through a membrane channel and can identify non-canonical nucleobases based on changes in electrical current across the membrane.[44] Furthermore, while the series of tC-derived compounds are considerably less bright than conventionally employed fluorophores, we have shown that common instrumentation for recording fluorescence in biological samples can detect adequate fluorescent signal at tC concentrations that avoid cytotoxicity. The set of current tC derivatives with varying photophysical properties can function as informative biophysical tools reporting on intracellular localization, physical conformation, and levels of nucleic acid metabolism in real time concurrent with normal cellular function.

4.4 Experimental Methods

Nucleoside triphosphate analogues were synthesized and purified according to published procedures.[25,39,45,46] The ^{DEA}tC ribose nucleoside, corresponding r(^{DEA}tC)TP, and tC phosphoramidate were novel compounds and their synthetic details and NMR data are described here. NMR spectra are shown in **Figures 4.5** – **4.12**.

Synthesis of 2',3',5'-tri-*O*-acetyl-8-diethylamino-tC-ribonucleoside. 8-diethylamino-3H-pyrimido[5,4-b][1,4]benzothiazin-2(10H)-one (150 mg, 0.52 mmol) was dissolved in 1 mL of anhydrous acetonitrile in a dry 25-mL Schlenk tube with magnetic stirring. After adding *N*,*O*-bis(trimethylsilyl)acetamide (148 μL, 0.73 mmol), the reaction was heated at 50 °C under N_2 with magnetic stirring. After 20 min the reaction vessel was removed from heating and cooled to room temperature, before adding 1,2,3,5-Tetra-*O*-acetyl-β-$_D$-ribofuranose (197 mg, 0.62 mmol)

and trimethylsilyl trifluoromethanesulfonate (112 μL, 0.62 mmol). The reaction ran for 2.5 hours at 50 °C under nitrogen gas. Reaction progress was monitored by TLC with 10 % methanol in methylene chloride. Once the reaction was finished, the flask was cooled to room temperature before transferring to a separation funnel with 50 mL of DCM and 50 mL of 5 % $NaHCO_3$ solution. After dispensing the first bottom organic layer, a second 50 mL of DCM was added to extract remaining product from the aqueous phase. The organic phases were combined, dried over anhydrous Na_2SO_4, and solvent removed by rotary evaporation. The crude mixture was purified by flash chromatography using 2% methanol in DCM to yield the product as a yellow film (225 mg, ~80% yield). ^{1}H NMR (400 MHz, CD_3OD) δ 7.52 (s, 1H), 6.79 (d, *J* = 8.6 Hz, 1H), 6.36 (dd, *J* = 2.7, 8.6 Hz, 1H), 6.34 (d, *J* = 2.5 Hz, 1H), 5.95 (d, *J* = 4.25 Hz, 1H), 5.47 (t, *J* = 4.3, 10.1 Hz, 1H), 5.39 (m, *J* = 5.6, 11.2 Hz, 1H), 4.38 (m, 3H), 3.33 (q, *J* = 7.1 Hz, 4H), 2.18 (s, 3H), 2.12 (s, 3H), 2.11 (s, 3H), 1.14 (t, *J* = 7.1 Hz, 6H).

Synthesis of 8-diethylamino-tC ribonucleoside. 2',3',5'-tri-*O*-acetyl-8-diethylamino-tC-ribonucleoside (225 mg, 0.412 mmol) was transferred to a dry 100-mL round bottomed flask and dissolved in 20 mL methyl alcohol with magnetic stirring. 30% sodium methoxide (405 μL, 2.25 mmol) in methanol was added to the reaction flask and the mixture stirred at 25 °C for 2.5 hours. Reaction progress was monitored by TLC with 10 % methanol in methylene chloride. When the reaction progressed to completion acetic acid (120 μL, 2.10 mmol) was added to the reaction mixture and briefly stirred. The solvent was removed by rotary evaporation and the crude product was purified by liquid chromatography using a gradient of 10-20% methanol in DCM to yield the product as a yellow solid (170 mg, 98%). ^{1}H NMR (500 MHz, CD_3OD) δ 7.96 (s, 1H), 6.76 (d, 8.6 Hz, 1H), 6.34 (dd, 8.6 Hz, 2.5 Hz, 1H), 6.32 (d, 2.5 Hz, 1H), 5.84 (d, 3.1 Hz, 1H), 4.13 (m, 1H), 4.03 (m, 1H), 3.91 (m 1H), 3.75 (12 Hz, 2.5 Hz, 1H), 3.32 (m, 4H), 1.13 (t,

7.1 Hz, 6H). ^{13}C NMR (100 MHz, CD_3OD) 162.05, 157.77, 149.02, 138.07, 135.81, 127.77, 109.75, 102.04, 101.55, 100.08, 92.33, 85.83, 76.31, 70.42, 61.59, 45.42, 12.83.

Synthesis of 8-diethylamino-tC ribonucleoside 5'-*O*-triphosphate. 8-diethylamino-tC ribonucleoside (120 mg, 0.286 mmol) was transferred to a clean 25-mL Schlenk tube with proton sponge (306 mg, 1.43 mmol) and dried 18 hours using a vacuum pump. The solids were suspended in 2 mL of trimethylphosphate with magnetic stirring and the reaction vessel was placed in an acetone/water-ice cooling bath. $POCl_3$ (80 μL, 0.858 mmol) was added in two doses over the course of 90 min while stirring at -15 °C. In a separate vial, tributylammonium pyrophosphate (784 mg, 1.43 mmol) was dried using a vacuum pump for 1-2 hours and dissolved in 2.86 mL of anhydrous DMF, making a 0.5 M solution. Phosphorylation of the nucleoside substrate was monitored by TLC with 10 % methanol in methylene chloride and checking for absence of migration. After initial phosphorylation of the starting nucleoside, the reaction vessel was removed from the cooling bath to equilibrate to room temperature. Diisopropylethylamine (249 μL, 1.43 mmol) was added to the reaction mixture before injecting the 0.5 M TBA-PPi solution and stirred at room temperature for 10 minutes. After which, the reaction was quenched by adding 5 mL of 0.5 M TEAB, pH 7.4. The solvents were removed by rotary evaporation and further dried with a vacuum pump overnight. The crude product was purified by reverse-phase liquid chromatography using 10-40% acetonitrile in 0.10 M TEAA buffer. Eluted fractions containing the purified product were pooled and solvent removed using lyophilization to yield the product as a yellow salt. ^{1}H NMR (400 MHz, D_2O) 8.40 (s, 1H), 7.71 (m, 1H), 7.04 – 6.98 (m, 2H), 5.89 (m, 1H), 4.25 (m, 1H), 3.81 (q, 7.3 Hz, 4H), 3.55 (m, 1H), 3.34, (m, 1H), 2.98 (m, 1H), 0.92 (t, 7.3 Hz, 6H). ^{31}P NMR (162 MHz, D_2O) -11.19, -12.03, -23.63.

Synthesis of (S)-2-(S)-phenoxyphosphorylamino tC 2'-deoxy-β-$_D$-ribonucleoside propionic acid isopropyl ester. tC nucleoside (10 mg, 0.030 mmol) was transferred to a dry 25-mL Schlenk tube and suspended in 250 μL of anhydrous THF using magnetic stirring. The reaction vessel was then submerged in an ice bath (0 °C) and 1.7 M tert-butylmagnesium chloride (37 μL, 0.063 mmol) in THF was added dropwise. The reaction was stirred for 30 min at 0 °C, then removed from the ice bath and allowed to equilibrate to room temperature (25 °C) before an additional 30 min of stirring. The reaction vessel was submerged in an ice-water bath (~5 °C) before adding (S)-2-[(S)-(2,3,4,5,6-pentafluorophenoxy)-phenoxyphosphorylamino] propionic acid isopropyl ester (16 mg, 0.036 mmol). The reaction was stirred in the thawing ice-water bath for 18 hrs. Disappearance of starting material was monitored by TLC using 10 % methanol in DCM. The solvent was removed by rotary evaporation and the product was purified by flash chromatography using 10 % methanol in DCM with 1 % triethylamine. The product was purified as a pale yellow solid (10 mg, 55 % yield). ^{1}H NMR (400 MHz, $CDCl_3$) δ 8.56 (bs, 1H), 7.62 (s, 1H), 7.33 (m, 2H), 7.16 (m, 1H), 7.07 (m, 1H), 6.94 (m, 2H), 6.87 (m, 1H), 6.14 (t, 6.5 Hz, 1H), 4.96 (m, 1H), 4.38 (m, 1H), 4.28 (m, 1H), 4.14 (m, 1H), 3.93 (m, 1H), 3.22 (m, 1H), 2.37 (m, 1H), 1.94 (m, 1H), 1.33 (d, 7.1 Hz, 1H), 1.20 (m, 6H). ^{13}C NMR (125 MHz, CD_3OD) δ 213.51, 206.61, 200.33, 152.80, 152.80, 147.68, 144.95, 130.78, 128.56, 127.13, 126.26, 125.69 121.58, 121.53, 118.06, 88.26, 72.04, 70.27, 51.69, 24.47, 22.00, 21.89, 20.66, 20.59, 19.74. ^{31}P NMR (162 MHz, $CDCl_3$) δ 3.05.

For labeling prokaryotic nucleic acids, BL21-Gold competent *E. coli* were transformed to be *PtNTT2*$^{+}$ by heat-shock transformation. Plasmid DNA encoding *P. tricornutum* nucleoside triphosphate transporter and chloramphenicol resistance genes were a generous gift from Dr. Floyd Romesberg from the Scripps Research Institute. Cells and plasmid DNA (10 ng) were

added in 2X YT (yeast extract and tryptone) liquid media and subjected to 42 °C for 22 seconds prior to immediate transfer to 0 °C for 2 min. Cultures were then plated on 2X YT agar plates supplemented with chloramphenicol (5 μg mL^{-1}) and incubated at 37 °C to isolate individual colonies of successfully transformed cells. Cells from genetically identical colonies were grown in liquid 2X YT supplemented with chloramphenicol (5 μg mL^{-1}) and 50 mM K_2HPO_4 at 37 °C. Glycerol was added to aliquots of culture fluid for final 40% glycerol content, flash-frozen in liquid nitrogen, and stored at -80 °C.

Fluorescence latency experiments were performed by growing cultures to OD_{600} 0.50 – 1.0 in 2X YT at 37 °C before being transferred to individual wells in a 96-well plate. Varying volumes of fluorescent nucleotide were added to culture aliquots and were brought to equal volumes of 200 μL with 50 mM K_2HPO_4 solution. The initial OD_{600} and fluorescence intensity (λ_{ex} = 380 nm, λ_{em} = 520 nm) were recorded using an Infinite M200 plate reader (Tecan Life Sciences) before incubating at 37 °C. After 24 hours the OD_{600} and fluorescence were recorded again and incubated cells were transferred to 1.5 mL microcentrifuge tubes. Cells were spun down at 13,000 g for 15 min and the supernatant was discarded, then washed with 50 mM K_2HPO_4 and gentle mixing. After pelleting the cells, the supernatant was discarded and replaced with 200 μL of fresh 2X YT with 5 μg mL^{-1} of chloramphenicol. This process was repeated after an additional 24 hours. Calculated growth rate constants (k) was performed using **Equation 4.1**. Where N is the number of cells at time t.

Equation 4.1

$$k = \frac{Log(N_t) - Log(N_0)}{Log(2)}$$

Reductions in fluorescence were calculated by taking ratio of emission intensity at time *t* over initial intensity and subtracting from 1.0.

Incubation experiments using human-derived cells were performed using tC phosphoramidates dissolved in dimethyl sulfoxide. Stock solutions of nucleotide substrates were diluted in nuclease-free water and added to cell cultures such that the final concentration was 50 μM and ≤1% DMSO. Cell cultures were incubated for 24-48 hours prior to analysis by flow cytometry. Cell cultures for 48 hour analysis were washed by first spinning down suspended cells or gently detaching adherent cells from surfaces and then spinning down, then subsequently exchanging the growth media for fresh media. Experiments with Huh7 hepatocytes were quantified and imaged on an Amnis ImageStream MkII (Luminex Corporation, Austin TX), while 293T and SUPT1 T-cells were sorted on FACS Aria III (BD Scientific, San Diego CA), both run in duplicate trials.

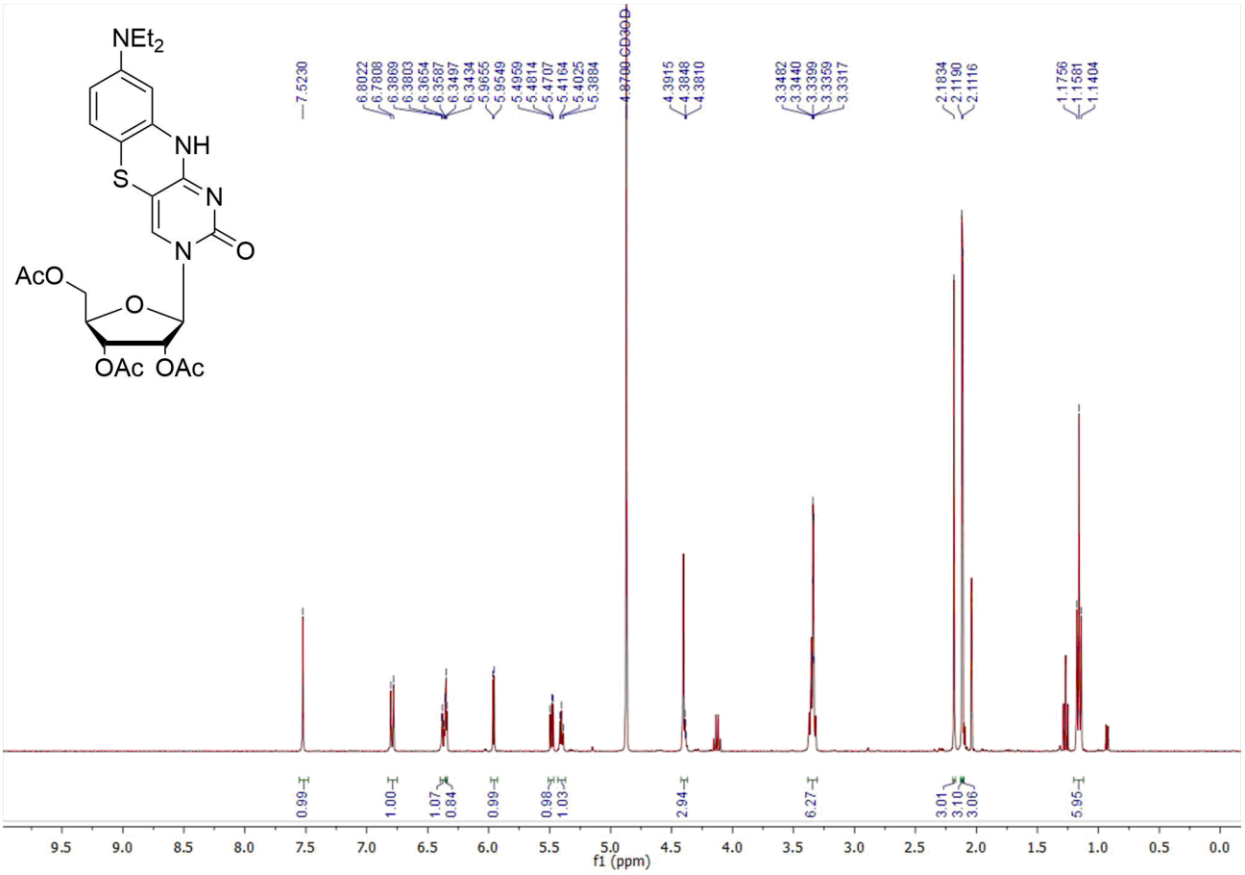


Figure 4.5. ^{1}H NMR spectrum of 2',3',5'-tri-*O*-acetyl-8-diethylamino-tC-ribonucleoside. Spectrum acquired in $CDCl_3$ at 298 K in a 500 MHz NMR spectrometer.

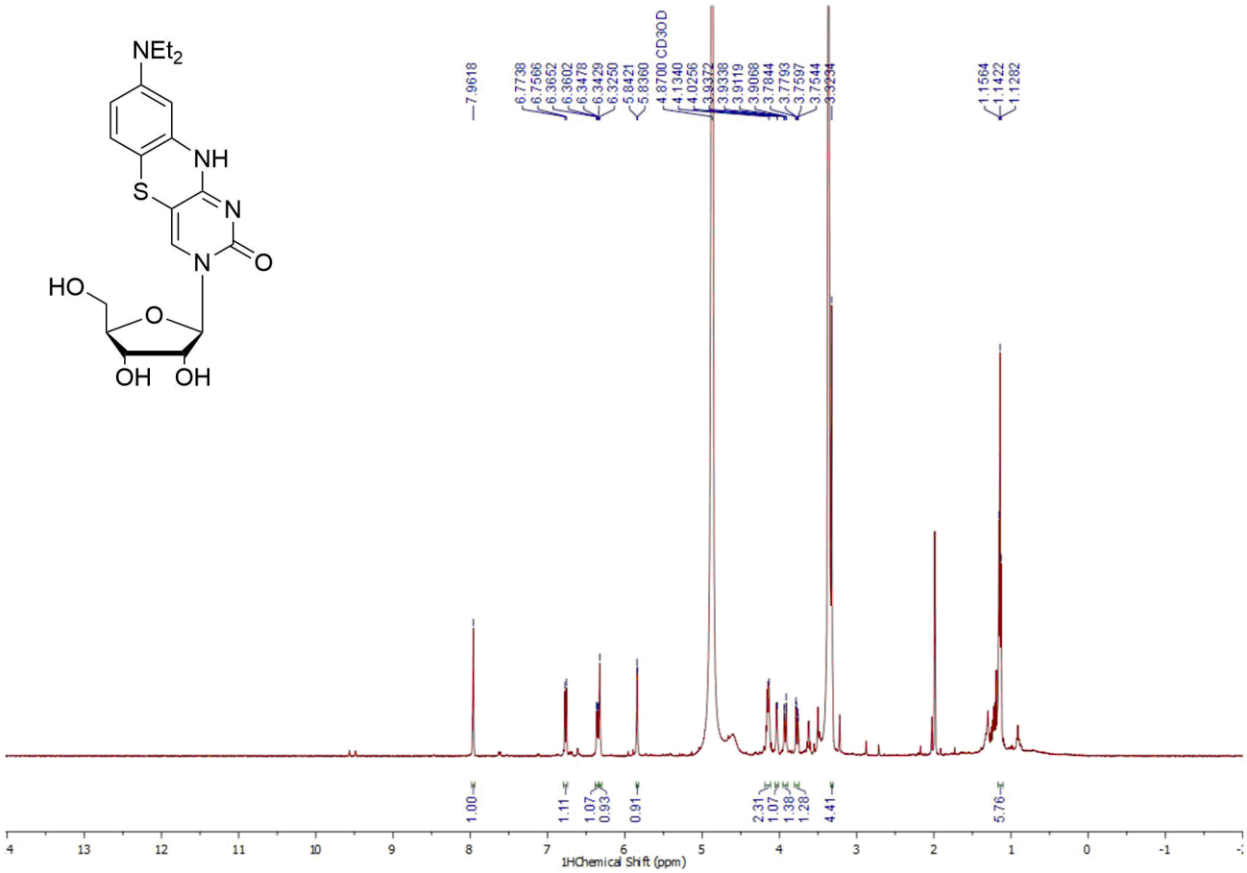


Figure 4.6. ^{1}H NMR spectrum of 8-diethylamino-tC-ribonucleoside. Spectrum acquired in CD_3OD at 298 K in a 500 MHz NMR spectrometer.

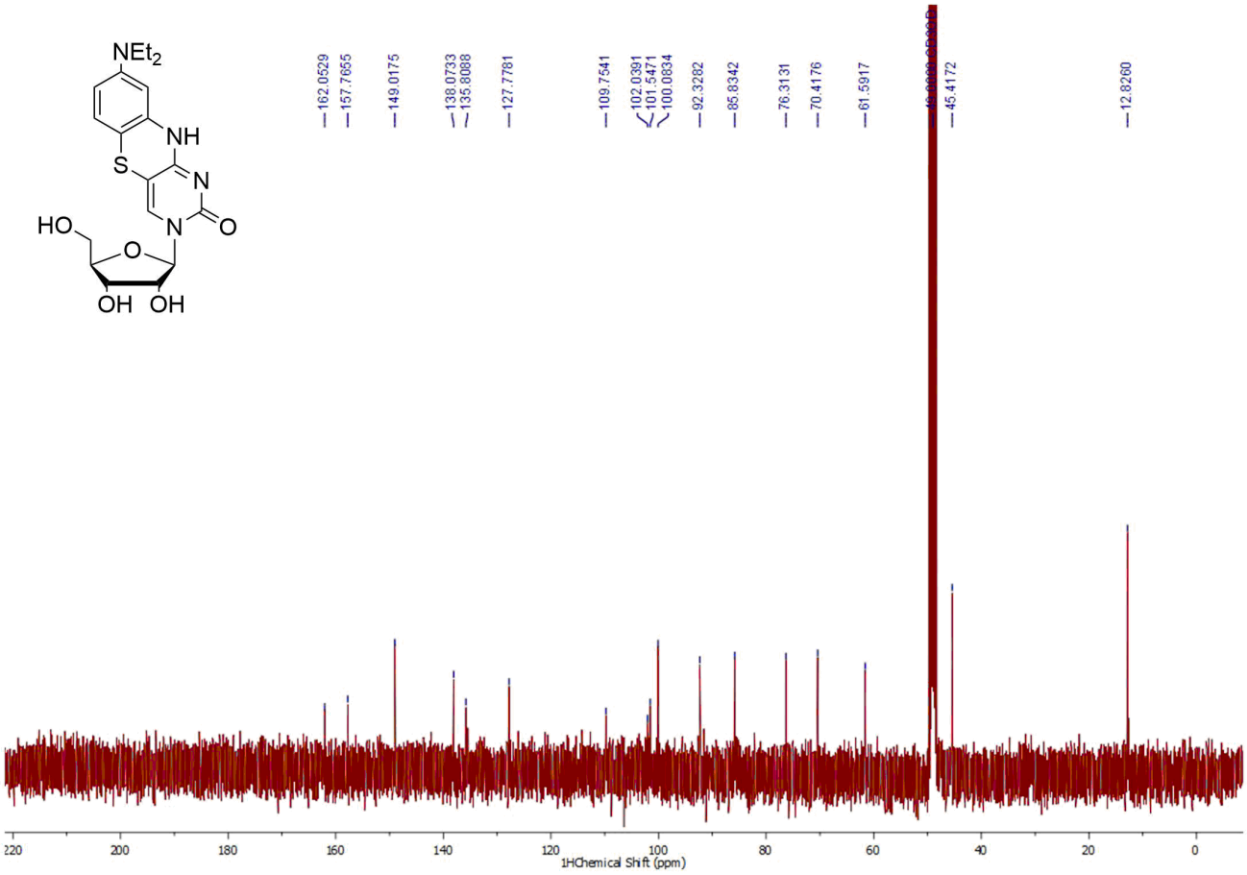


Figure 4.7. ^{13}C NMR spectrum of 8-diethylamino-tC-ribonucleoside. Spectrum acquired in CD_3OD at 298 K in a 100 MHz NMR spectrometer.

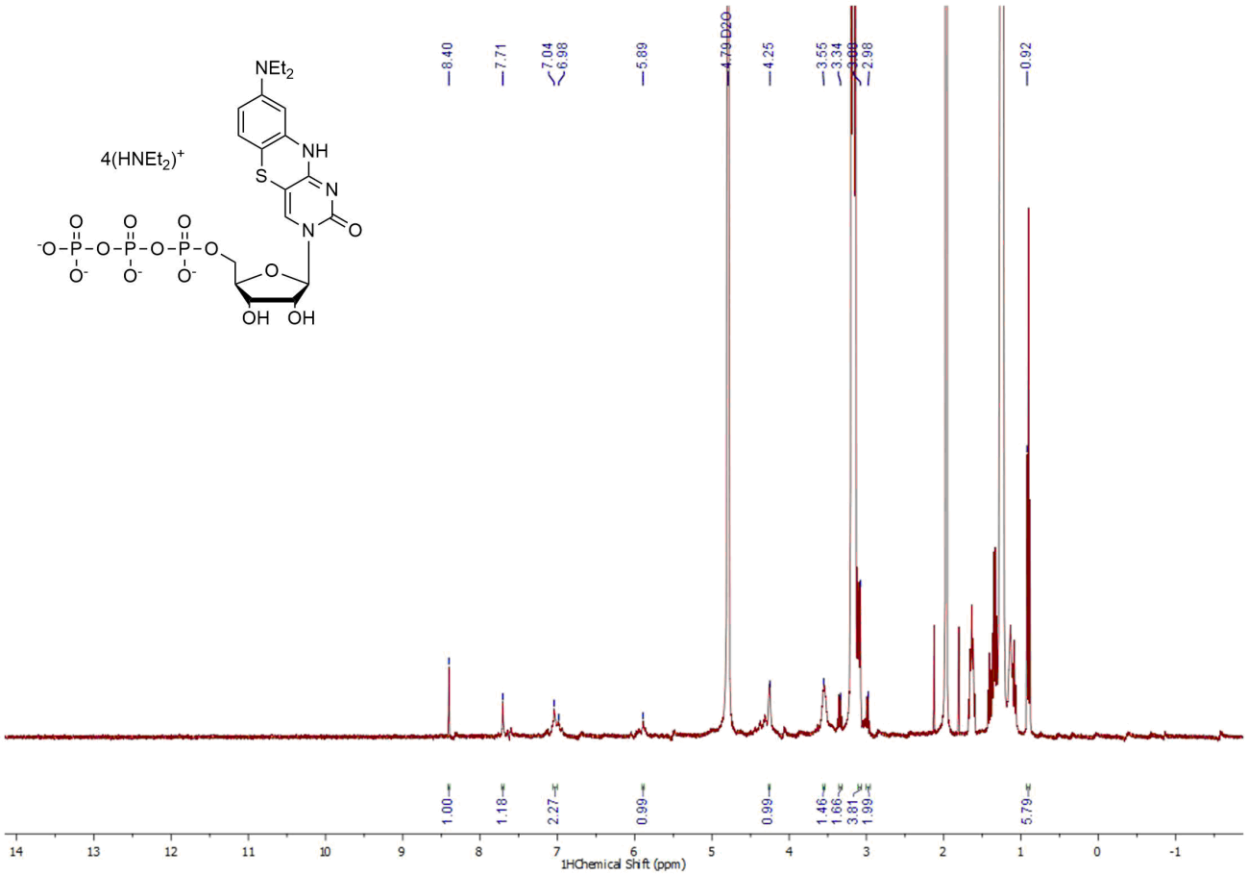


Figure 4.8. ^{1}H NMR spectrum of 8-diethylamino-tC-ribonucleoside 5'-*O*-triphosphate. Spectrum acquired in D_2O at 298 K in a 400 MHz NMR spectrometer.

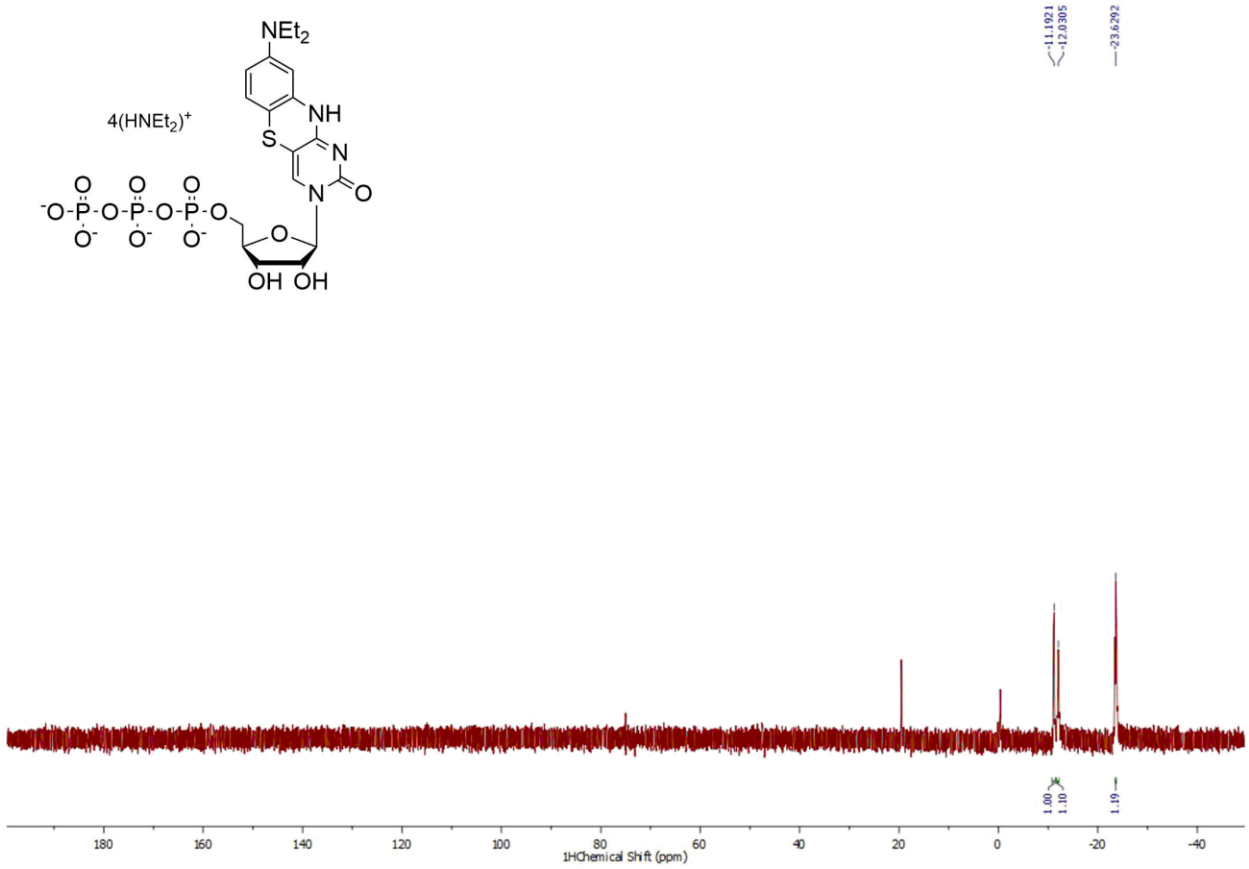


Figure 4.9. ^{31}P NMR spectrum of 8-diethylamino-tC-ribonucleoside 5'-*O*-triphosphate. Spectrum acquired in D_2O at 298 K in a 162 MHz NMR spectrometer. Note: residual signal around 20 ppm integrates to <5% of product.

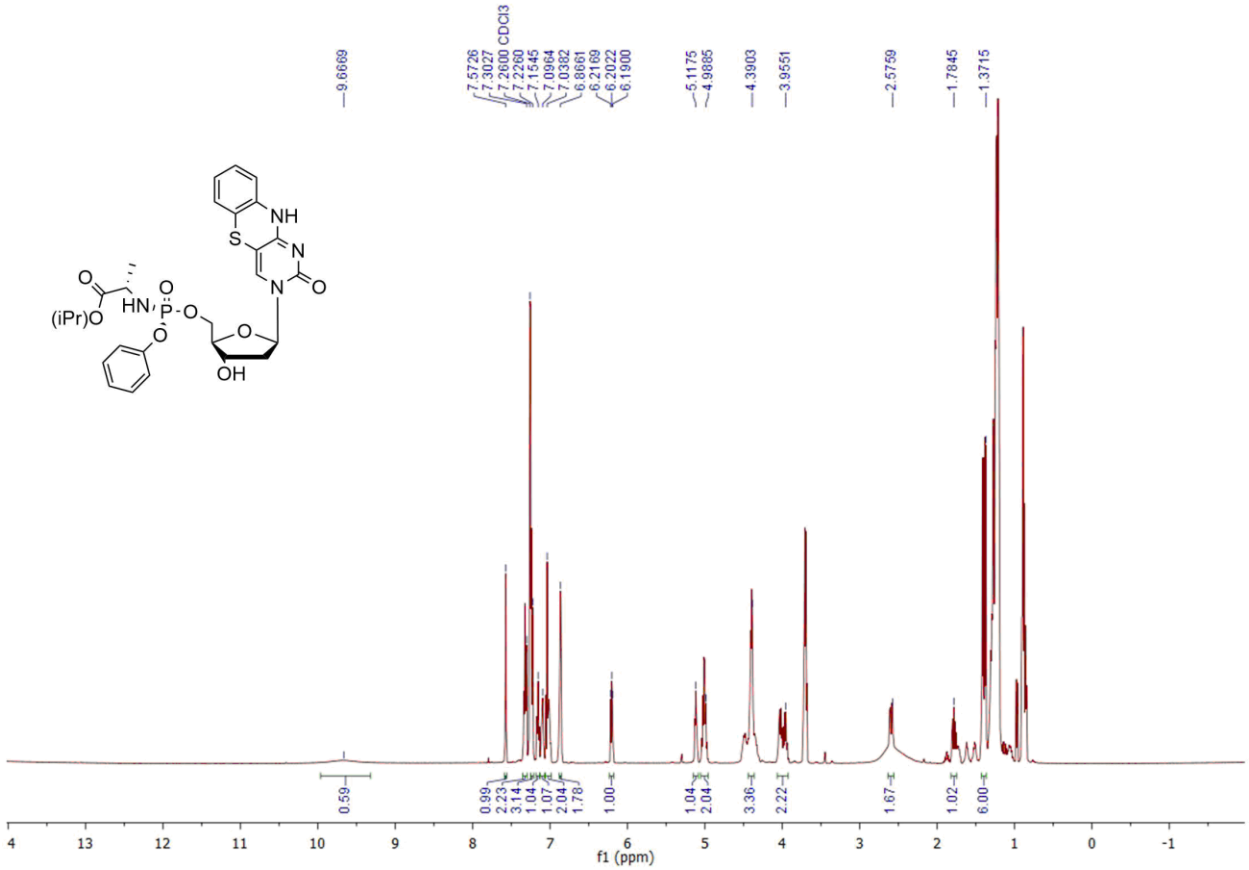


Figure 4.10. ^{1}H NMR spectrum of (S)-2-(S)-phenoxyphosphorylamino tC 2'-deoxy-β-D-2'-deoxyribonucleoside propionic acid isopropyl ester. Spectrum acquired in $CDCl_3$ at 298 K in a 400 MHz NMR spectrometer.

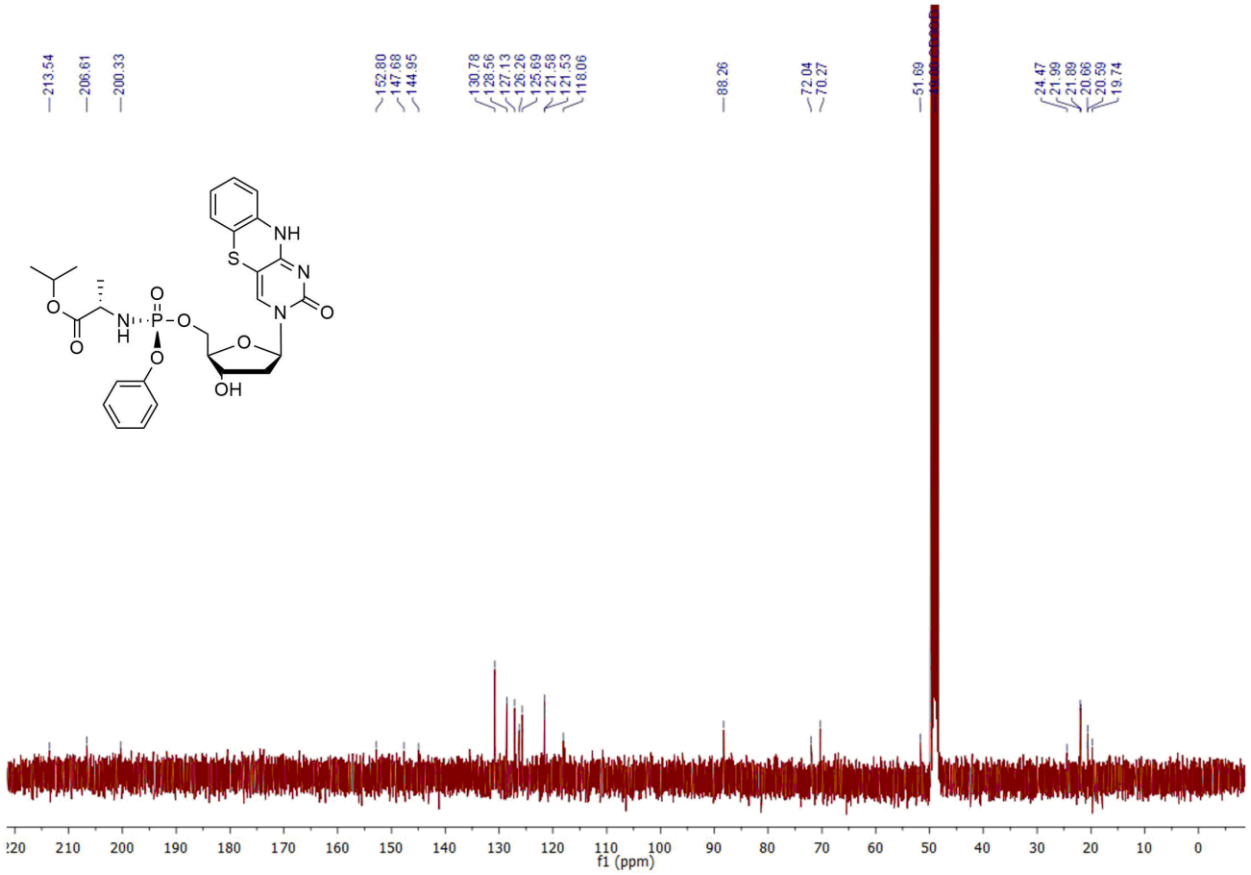


Figure 4.11. ^{13}C NMR spectrum of (S)-2-(S)-phenoxyphosphorylamino tC 2'-deoxy-β-$_D$-2'-deoxyribonucleoside propionic acid isopropyl ester. Spectrum acquired in CD_3OD at 298 K in a 125 MHz NMR spectrometer.

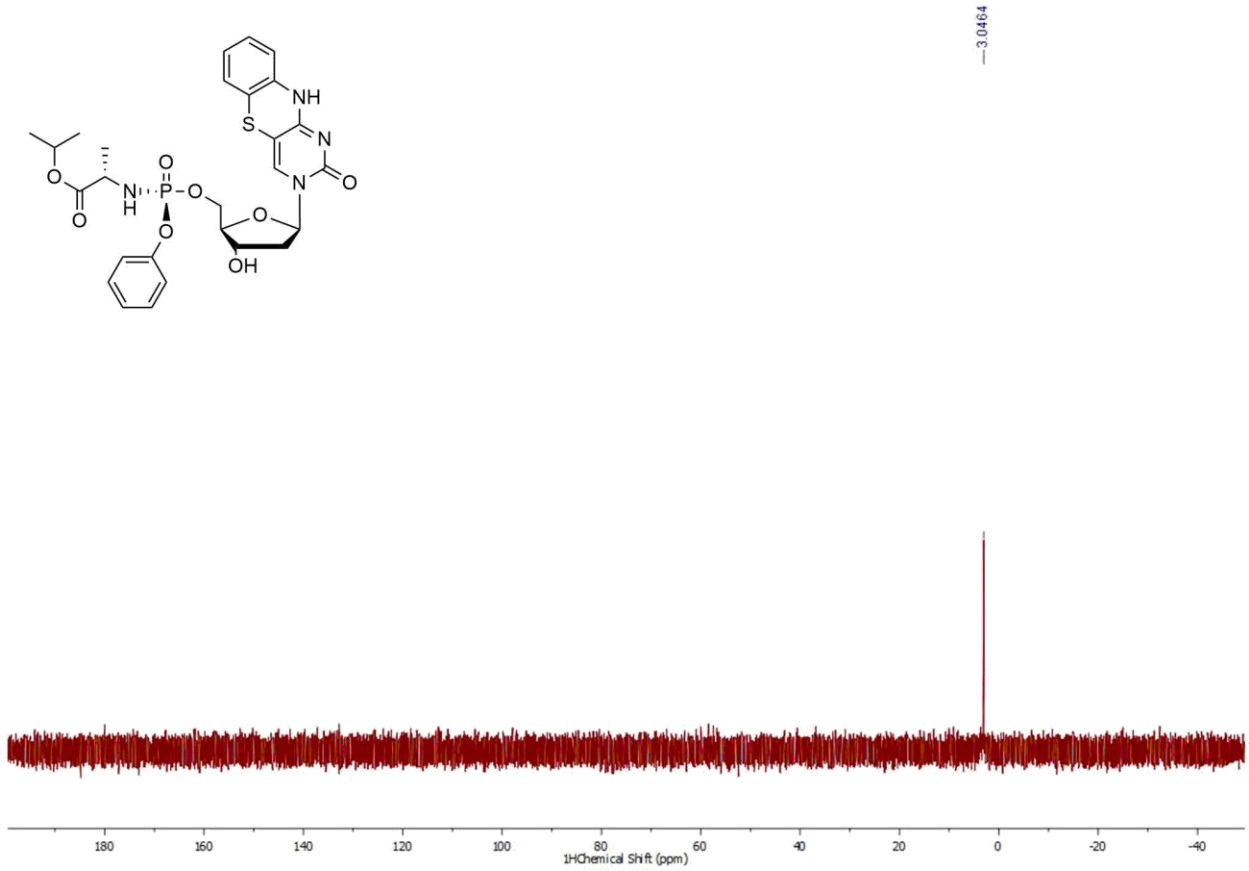


Figure 4.12. ^{31}P NMR spectrum of (S)-2-(S)-phenoxyphosphorylamino tC 2'-deoxy-β-$_D$-2'-deoxyribonucleoside propionic acid isopropyl ester. Spectrum acquired in $CDCl_3$ at 298 K in a 162 MHz NMR spectrometer.

4.5 References

(1) Sinkeldam, R. W.; Greco, N. J.; Tor, Y. *Chem. Rev.* **2010**, *110* (5), 2579–2619.

(2) Biebricher, A. S.; Heller, I.; Roijmans, R. F. H.; Hoekstra, T. P.; Peterman, E. J. G.; Wuite, G. J. L. *Nat. Commun.* **2015**, *6*, 7304.

(3) Güixens-Gallardo, P.; Zawada, Z.; Matyašovský, J.; Dziuba, D.; Pohl, R.; Kraus, T.; Hocek, M. *Bioconjug. Chem.* **2018**, *29* (11), 3906–3912.

(4) Neef, a. B.; Luedtke, N. W. *Proc. Natl. Acad. Sci.* **2011**, *108* (51), 20404–20409.

(5) *Fluorescent Analogs of Biomolecular Building Blocks: Design and Applications*; Wilhelmsson, L. M., Tor, Y., Eds.; John Wiley & Sons, Inc.: Hoboken, NJ, 2016.

(6) Xu, W.; Chan, K. M.; Kool, E. T. *Nat. Chem.* **2017**, *9* (11), 1043–1055.

(7) Schmidt, O. P.; Mata, G.; Luedtke, N. W. *J. Am. Chem. Soc.* **2016**, *138* (44), 14733–14739.

(8) Claudio-Montero, A.; Pinilla-Macua, I.; Fernández-Calotti, P.; Sancho-Mateo, C.; Lostao, M. P.; Colomer, D.; Grandas, A.; Pastor-Anglada, M. *Mol. Pharm.* **2015**, *12* (6), 2158–2166.

(9) Beharry, A. A.; Lacoste, S.; O'Connor, T. R.; Kool, E. T. *J. Am. Chem. Soc.* **2016**, *138* (11), 3647–3650.

(10) diCenzo, G. C.; Finan, T. M. *Microbiol. Mol. Biol. Rev.* **2017**, *81* (3), e00019-17.

(11) Konrad, A.; Yarunova, E.; Tinta, T.; Piškur, J.; Liberles, D. A. *Gene* **2012**, *492* (1), 117–120.

(12) Malyshev, D. A.; Dhami, K.; Lavergne, T.; Chen, T.; Dai, N.; Foster, J. M.; Corrêa, I. R.; Romesberg, F. E. *Nature* **2014**, *509* (7500), 385–388.

(13) Ast, M.; Gruber, A.; Schmitz-Esser, S.; Neuhaus, H. E.; Kroth, P. G.; Horn, M.; Haferkamp, I. *Proc. Natl. Acad. Sci.* **2009**, *106* (9), 3621 LP – 3626.

(14) Seley-Radtke, K. L.; Yates, M. K. *Antiviral Res.* **2018**, *154*, 66–86.

(15) Pertusat, F.; Serpi, M.; McGuigan, C. *Antivir. Chem. Chemother.* **2012**, *22* (5), 181–203.

(16) McGuigan, C.; Pathirana, R. N.; Mahmood, N.; Devine, K. G.; Hay, A. J. *Antiviral Res.* **1992**, *17* (4), 311–321.

(17) Slusarczyk, M.; Serpi, M.; Pertusati, F. *Antivir. Chem. Chemother.* **2018**, *26*, 2040206618775243–2040206618775243.

(18) McGuigan, C.; Harris, S. A.; Daluge, S. M.; Gudmundsson, K. S.; McLean, E. W.; Burnette, T. C.; Marr, H.; Hazen, R.; Condreay, L. D.; Johnson, L.; De Clercq, E.; Balzarini, J. *J. Med. Chem.* **2005**, *48* (10), 3504–3515.

(19) Mehellou, Y.; Rattan, H. S.; Balzarini, J. *J. Med. Chem.* **2018**, *61* (6), 2211–2226.

(20) Clark, J. L.; Hollecker, L.; Mason, J. C.; Stuyver, L. J.; Tharnish, P. M.; Lostia, S.; McBrayer, T. R.; Schinazi, R. F.; Watanabe, K. A.; Otto, M. J.; Furman, P. A.; Stec, W. J.; Patterson, S. E.; Pankiewicz, K. W. *J. Med. Chem.* **2005**, *48* (17), 5504–5508.

(21) Ma, H.; Jiang, W.; Robledo, N.; Leveque, V.; Ali, S.; Lara-Jaime, T.; Masjedizadeh, M.; Smith, D.; Cammack, N.; Klumpp, K.; Symons, J. *J. Biol. Chem.* **2007**, *282* (41), 29812–29820.

(22) Ross, B. S.; Ganapati Reddy, P.; Zhang, H. R.; Rachakonda, S.; Sofia, M. J. *J. Org. Chem.* **2011**, *76* (20), 8311–8319.

(23) Vorbrüggen, H.; Krolikiewicz, K.; Bennua, B. *Chem. Ber.* **1981**, *114* (4), 1234–1255.

(24) Niedballa, U.; Vorbrueggen, H. *J. Org. Chem.* **1974**, *39* (25), 3654–3660.

(25) Burgess, K.; Cook, D. *Chem. Rev.* **2000**, *100* (6), 2047–2059.

(26) Feldman, A. W.; Fischer, E. C.; Ledbetter, M. P.; Liao, J.-Y.; Chaput, J. C.; Romesberg, F. E. *J. Am. Chem. Soc.* **2018**, *140* (4), 1447–1454.

(27) Edwards, V. H. *Biotechnol. Bioeng.* **1970**, *12* (5), 679–712.

(28) Gibson, B.; Wilson, D. J.; Feil, E.; Eyre-Walker, A. *Proceedings. Biol. Sci.* **2018**, *285* (1880), 20180789.

(29) DeLong, E. F.; Wickham, G. S.; Pace, N. R. *Science (80-.).* **1989**, *243* (4896), 1360 LP – 1363.

(30) Anderson, K. L.; Dunman, P. M. *Int. J. Microbiol.* **2009**, *2009*, 525491.

(31) Murakami, E.; Tolstykh, T.; Bao, H.; Niu, C.; Micolochick Steuer, H. M.; Bao, D.; Chang, W.; Espiritu, C.; Bansal, S.; Lam, A. M.; Otto, M. J.; Sofia, M. J.; Furman, P. A. *J. Biol. Chem.* **2010**, *285* (45), 34337–34347.

(32) Brenner, C. *Biochemistry* **2002**, *41* (29), 9003–9014.

(33) Schaertl, S.; Konrad, M.; Geeves, M. A. *J. Biol. Chem.* **1998**, *273* (10), 5662–5669.

(34) Pasti, C.; Gallois-Montbrun, S.; Munier-Lehmann, H.; Veron, M.; Gilles, A.-M.; Deville-Bonne, D. *Eur. J. Biochem.* **2003**, *270* (8), 1784–1790.

(35) Varga, A.; Roy, C. L. and B. *Current Drug Metabolism*. **2016**, pp 237–252.

(36) Quinlivan, E. P.; Gregory 3rd, J. F. *Anal. Biochem.* **2008**, *373* (2), 383–385.

(37) Bailey, C. M.; Anderson, K. S. *Biochim. Biophys. Acta* **2010**, *1804* (5), 1213–1222.

(38) Bulock, C. R.; Xing, X.; Shcherbakova, P. V. *Proc. Natl. Acad. Sci.* **2020**, *117* (11), 6035 LP – 6041.

(39) Turner, M. B.; Purse, B. W. *Chempluschem* **2020**, *85* (5), 855–865.

(40) Stengel, G.; Urban, M.; Purse, B. W.; Kuchta, R. D. *Anal. Chem.* **2010**, *82* (3), 1082–1089.

(41) Stengel, G.; Purse, B. W.; Wilhelmsson, L. M.; Urban, M.; Kuchta, R. D. *Biochemistry* **2009**, *48* (31), 7547–7555.

(42) Galazka, G.; Domowicz, M.; Ewiak-Paszynska, A.; Jurewicz, A. *Cells* **2019**, *8* (8), 790.

(43) Perišić Nanut, M.; Sabotič, J.; Jewett, A.; Kos, J. *Front. Immunol.* **2014**, *5*, 616.

(44) Georgieva, D.; Liu, Q.; Wang, K.; Egli, D. *Nucleic Acids Res.* **2020**.

(45) Turner, M. B.; Anderson, B. A.; Samaan, G. N.; Coste, M.; Burns, D. D.; Purse, B. W. *Curr. Protoc. Nucleic Acid Chem.* **2018**, e59.

(46) Teppang, K. L.; Lee, R. W.; Burns, D. D.; Turner, M. B.; Lokensgard, M. E.; Cooksy, A. L.; Purse, B. W. *Chem. - A Eur. J.* **2019**, *25*, 1249–1259.

Chapter 5

Conclusion

The large diversity of fluorescent nucleoside analogues has been fundamental in analyzing nucleic acids structural dynamics, interactions with other biomolecules such as proteins, and cellular functions with increasing resolution. As detailed in Chapter 1, the biological function of nucleic acids extends far beyond the central dogma of biology (i.e. DNA transcribed to RNA, then translated to protein) and the probative capabilities endowed by minimally perturbing fluorescent nucleosides enable scientists to examine the complex functions of DNA/RNA. Chemists continue to synthesize novel fluorescent nucleoside analogues and characterize their biochemical capabilities *in vitro*, however the potential of those analogues' functions in biologically relevant contexts can be limited without *in vivo* studies. As was described by Eric Kool in a 2017 review on fluorescent nucleobase analogues, studying fluorescent nucleosides in living biological systems remains largely unexplored and therefore offers the greatest potential for research discovery.[1]

While the photophysical properties of parent tC and tC-derived analogues have been characterized, the development of more complex applications in living biological systems has not been thoroughly explored. Additionally, other tC-derived nucleosides show promising fluorescence properties for studying current major questions in nucleic acids research, such as local structure conformation, single nucleotide polymorphisms, and epigenetic alterations. This dissertation presents necessary work to advance purposes such as the aforementioned by examining the mechanism of fluorescence enhancement of ^{DEA}tC, incorporation kinetics with polymerase enzymes, and delivery of the unnatural nucleotide analogues to living prokaryotic and eukaryotic cells. Understanding the precise physical conditions or mechanisms of fluorescent

behavior exhibited by certain tC-derived analogues streamlines research efforts in identifying the best matches between analogue properties and applications.

Subsequent research involving the fluorescent tC molecules will likely include increasing the complexity of biological systems in which the tC derivatives are studied. For instance, ^{DEA}tC shows high potential as a SNP-discriminating probe based on the selective turn-on response reported in Chapter 2, however further investigations of the probe in hybridization assays with nucleic acids from cell lysates and in biological systems are needed to focus on which shortcomings of ^{DEA}tC to address to make the analogue a suitable alternative to current FISH probes. Another tC-derived analogue, ^{Cl}tC^{O}, exhibits emission spectra with vibrational fine structure that depends on solvation and base-stacking,[2] and future studies in the Purse Lab may focus on capitalizing on the analogue's sensitivity for live-cell studies, such as the condensation of genomes during cell division, or the physical conformations of viral genomes during and after encapsulation in virus particles. The contents of Chapter 3 demonstrate that reverse transcriptase from HIV-1 can incorporate the tC analogues during both RNA- and DNA-templated DNA synthesis with efficiencies comparable to natural cytidine. Accordingly, metabolic labeling studies similar to those in Chapter 4 may be performed in virally infected cells to generate fluorescently labeled viral particles using the tC analogues, such as with HIV-1 and human $CD4^{+}$ T-cells. Viruses have been a highly active area of biomedical research and studying their nucleic acid metabolism is pivotal to understanding their replication and development. The 2020 global pandemic caused by SARS-CoV-2 exemplifies the need for greater research into virology, and utilizing fluorescent nucleoside analogues with minimally perturbing effects to investigate viral DNA/RNA dynamics and metabolism may greatly advance efforts at mitigating infection.

Although the parent tC nucleoside analogue was initially developed in the 1990's for RNA antisense purposes,[3] the focus of tC research shifted towards examining the nucleoside's fluorescent properties and expanding the number of tC derivatives.[2,4–6] Multiple studies had been conducted characterizing the analogues' photophysical properties and even incorporating them into DNA or RNA via polymerase-mediated insertion.[7,8] Chemical modifications to the phenothiazine architecture were based on motifs that highly emissive commercial fluorophores tend to include. However, there was not much evidence rationalizing why certain modifications or substitutions endowed the observed fluorescent properties. The major lessons learned in this dissertation contributing to tC nucleoside research are those found from identifying excited-state proton transfer between bulk water and ^{DEA}tC is one of the major factors for the analogue's bright turn-on response, as well as initial studies using the analogues' in the next logical phase of development: using their fluorescent properties to study nucleic acids in living cells. Since the fluorescent turn-on response has significant potential for sensing DNA/RNA duplex formation, incorporating the diethylamino substituent into future tC-based nucleobase designs may result in turn-on probes with different fluorescent properties such as red-shifted emission or greater brightness. Performing biochemical analyses with tC nucleotides and polymerase enzymes *in vitro* demonstrates what biological applications (e.g. monitoring nucleic acid metabolism in real time, examining physical conformations) are possible, and live-cell studies work towards making those applications more biologically relevant. The value of those lessons on tC nucleoside research discussed in this dissertation come a long way from the nucleoside analogue's original function nearly 20 years ago.

5.1 References

(1) Xu, W.; Chan, K. M.; Kool, E. T. *Nat. Chem.* **2017**, *9* (11), 1043–1055.

(2) Teppang, K. L.; Lee, R. W.; Burns, D. D.; Turner, M. B.; Lokensgard, M. E.; Cooksy, A. L.; Purse, B. W. *Chem. - A Eur. J.* **2019**, *25*, 1249–1259.

(3) Lin, K.-Y. Y.; Jones, R. J.; Matteucci, M. *J. Am. Chem. Soc.* **1995**, *117* (13), 3873–3874.

(4) Rodgers, B. J.; Elsharif, N. A.; Vashisht, N.; Mingus, M. M.; Mulvahill, M. A.; Stengel, G.; Kuchta, R. D.; Purse, B. W. *Chem. - A Eur. J.* **2014**, *20* (7), 2010–2015.

(5) Wilhelmsson, L. M.; Sandin, P.; Holmén, A.; Albinsson, B.; Lincoln, P.; Nordén, B. *J. Phys. Chem. B* **2003**, *107* (34), 9094–9101.

(6) Sandin, P.; Wilhelmsson, L. M.; Lincoln, P.; Powers, V. E. C.; Brown, T.; Albinsson, B. *Nucleic Acids Res.* **2005**, *33* (16), 5019–5025.

(7) Stengel, G.; Urban, M.; Purse, B. W.; Kuchta, R. D. *Anal. Chem.* **2010**, *82* (3), 1082–1089.

(8) Sandin, P.; Stengel, G.; Ljungdahl, T.; Borjesson, K.; Macao, B.; Wilhelmsson, L. M. *Nucleic Acids Res.* **2009**, *37* (12), 3924–3933.