Chemistry Dissertations   Department of Chemistry

Spring 2-26-2020

Synthesis of Selenium and Tellurium Modified Nucleic Acids For DNA Crystallization, Structure and Function Studies

cen Chen

Follow this and additional works at: https://scholarworks.gsu.edu/chemistry_diss

SYNTHESIS OF SELENIUM AND TELLURIUM MODIFIED NUCLEIC ACIDS FOR DNA

CRYSTALLIZATION, STRUCTURE AND FUNCTION STUDIES

by

CEN CHEN

Under the direction of Dr. Zhen Huang and Dr. Jun Yin

ABSTRACT

Nucleic acid X-ray crystallography faces crystallization as well as phasing challenge. The Phasing problem could be rationally solved by incorporation of selenium into the DNA or RNA oligonucleotides, while the crystallization of nucleic acid is still challenging. To address these challenges without structural perturbation, we decided to explore the atom-specific incorporation and place a selenium atom on the 2'-beta position to control B-form DNA formation during crystallization process. Herein we report the first synthesis of the β-2'-MeSe-thymidine ($^{2'Se}$T) phosphoramidite and Se-DNAs. We found that the Se-DNAs and Se-DNA-protein complexes formed crystals most of time with higher quality and diffraction resolution than the non-modified ones. Surprisingly, the Se-DNA can form crystals up to 600 microns in size, which were frequently hundreds of times larger in volume than the corresponding native. Moreover, we discovered that the high-quality Se-DNA crystals offered the diffraction resolution up to 1.15 Å and the Se-

derivatized structure was virtually identical to the native one. In the meanwhile, we synthesized the β-2'-MeSe-cytidine ($^{2'Se}$C) phosphoramidite and its DNA oligos. To our surprise, the selenium modification greatly facilitate crystallization when the modified cytidine was placed at the terminals of the oligo. The crystal of the Se-DNA formed within a few hours, where the corresponding native crystalized over 1 week.

Furthermore, we describe the first synthesis of 2'-MeSe-arabinouridine ($^{2'Se}$U) phosphoramidite and its DNAs to investigate the structure of DNAs containing the uracil. The Se-derivatized DNA X-ray crystal structure (1.25 Å) was virtually identical to the native one. Interestingly, during the Se incorporation step, we also found that the MeSe group attack the C-4 of the uracil generating a oxazolinyl-selenolester when the N-3 was protected with tert-butyloxycarbonyl (Boc) group which may activate the C-4 position through a n-π* interaction.

Besides the structure studies, we also synthesized the 5-phenylselenium and 5-phenyltelenium modified deoxyuridine and incorporated them into DNA oligos to investigate their charge transport properties. The conductance and current-voltage (I-V) characteristics measurement indicates that the Te modification more effectively manipulate the electronic structure of the DNA compared to the Se modification and the corresponding native.

INDEX WORDS: Selenium, Tellurium, Nucleic acids, DNA, X-ray crystallography, crystallization, Uracil in DNA, charge transport, selenol esters

SYNTHESIS OF SELENIUM AND TELLURIUM MODIFIED NUCLEIC ACIDS FOR DNA

CRYSTALLIZATION, STRUCTURE AND FUNCTION STUDIES

by

CEN CHEN

A Dissertation Submitted in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

In the College of Arts and Sciences

Georgia State University

2020

SYNTHESIS OF SELENIUM AND TELLURIUM MODIFIED NUCLEIC ACIDS FOR DNA

CRYSTALLIZATION, STRUCTURE AND FUNCTION STUDIES

by

CEN CHEN

Committee Chair: Jun Yin

Committee: Ming Luo

Markus Germann

Electronic Version Approved:

Office of Graduate Studies

College of Arts and Sciences

Georgia State University

May 2020

DEDICATION

This dissertation is dedicated to my loving wife **Ping Li** who always support and encourage me through this journey with her love and understanding. Thank you for flying thousands of miles from China and stand by me after all these years. Without your company, it would be impossible to overcome the arduous PhD study. And to our lovely son **Jason Chen** who is my biggest motivation.

I further dedicate this dissertation to my parents, **Zuxuan Chen** and **Lin Wan**, who had instilled in me the strength of character that navigates me through each challenge. Thank you for your education which endued me with the ability to accomplish this work, as well as your unconditional love!

This dissertation is also dedicated to my uncle **Lin Pu** who give me the most boost in my life. Thank you for your help and wise advices. Finally, I want to dedicate this dissertation to my Grandma, **Mingyu Wan**, as well as my Grandpa, **Songbo Pu** who have left us forever. Thank you for your caring and unconditional love.

ACKNOWLEDGEMENTS

At the end of my PhD study, I would like to express my sincere thanks to many people who encourage me and help me overcome all difficulties.

I would like to express my deepest gratitude to my advisor Dr. Zhen Huang for giving me the opportunity to do the research and providing invaluable support and guidance through each stage of the research. Without his instruction, I will not be able to accomplish my dissertation. His vision, motivation and enthusiasm greatly inspired me and would become precious assets in my future life. It was a great honor to work and study under his guidance. I am extremely grateful for what he has offered me.

I am extremely grateful to Dr. Jun Yin who kindly accepted me to join his research family and be my dissertation committee chair. Thank you for your kindness, encouragement, understanding, concern and all the help. And I also would like to thank my dissertation committee members, Dr. Ming Luo and Dr. Markus Germann, for their understanding, insightful comments, and constructive advices. My sincere thanks also go to Dr. Donald Hambelberg and Dr. Suri Iyer for their help and advices through the hard time.

I would also like to thank my friends and lab mates for their support and instructive discussion. I am very grateful to Dr. Sibo Jiang and Dr. Huiyan Sun for the discussion in DNA synthesis, to Dr. Wen Zhang and Dr. Ziyuan Fang for their help in solving structures, to Dr. Jozef Salon for his advice in nucleic acid chemistry and oligonucleotide synthesis, to Shuo Liu and Mohammad Sazid Hassan for the help in organic synthesis, to Dr. Manindar Kaur, Dr. Abdur Rob, Yifei Wang, Cuilun Kong, Razieh Esmaeili, Edwin Ogbonna, Xinhua Chen, Fukang Yang, Qin Lin, Liuyi Shi, Linrui Zhen, Babatunde Bello, James Compbell, Travon Haynes, for their support and discussion.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SCHEMES

1   GENERAL INTRODUCTION

Nucleic acids (including DNA and RNA) is essential and ubiquitous genetic materials in living cells. Although targeting proteins has always been the mainstream of therapeutic development and drug discovery, there is a trend in recent years to choose nucleic acids as the targets for clinical drugs, such as anticancer drugs. Therefore, more 3D structural information of nucleic acids is required for designing this kind of novel and effective drug. In the past decade, our group has been focusing on the selenium atom-specific modification (SAM) on nucleic acids, which has already been proved to have the potential for crystallization facilitation and structure determination of nucleic acid and protein-nucleic acid complexes. The brief background and current progress of SAM for structural and functional studies of nucleic acid as well as drug discovery are summarized in this chapter.

1.1   Background

DNAs and RNAs are essential biomolecules in cells that is responsible for genetic information storage, replication and transcription, and also have been considered as potential drug targets, especially for anticancer drugs.[1-8] Moreover, they have also been widely applied to biotechnology and nanotechnology[9-16]. DNAs and RNAs can fold into many different shapes and structures in addition to classical duplexes[17-19], such as triplexes[20], quadruplexes[21, 22], Hollidays Junctions[23] and i-motifs[24], let alone the structures formed between nucleic acids and proteins. The 3D-structure determination of nucleic acids and nucleic acid-protein complexes can significantly enhance nucleic acid-targeted anticancer drug discovery, as well as selective gene silencing research.

In the past two decades, the development of X-ray crystallographic technologies has largely facilitated the structure determination of biological macromolecules, of course including nucleic

acids. More and more novel structure of nucleic acid and nucleic acid-protein complexes have emerged. At the same time, there are many promising advancements in potential nucleic acid-based therapeutics, including antisense oligonucleotides, siRNAs, microRNAs and aptamers. These investigations make structural and functional studies of nucleic acid-protein complexes an increasingly active research field, thereby demanding new methodologies, especially novel strategies for nucleic acid X-ray crystallography.[25]

1.2   Challenges in nucleic acid crystallography

1.2.1   Crystallization

However, compared with protein crystallography, there are unique challenges in nucleic acid X-ray crystallography. One of the long-existing challenge among these is the poor ability of nucleic acids to form crystals. While a protein molecule has a variety of structural and chemical groups on its surface to enable crystal contact, the surface of a nucleic acid molecule is full of negatively charged repetitive phosphate groups, these intrinsic charges as well as its structural dynamics make the crystal packing of DNA and RNA molecules much more difficult.[25] Moreover, during the crystallization process, the high salt buffer condition would favor A-form DNA duplex which makes B-DNA crystallization more challenging. The crystallization buffer screening process is usually labor intensive, and the formation of high-quality crystals is usually rare and time-consuming. These issues have significantly slowed down the structure determination of nucleic acids.

Based on previous research, it is possible to tailor the local structure and molecular dynamics of DNA and accelerate the crystallization process of DNA molecules by the Se atom-specific modification (SAM) strategy[26-28]. During the past decade, selenium atom in the same

elemental family of oxygen has been introduced to various locations of nucleotides, especially nucleobases, to study its unique structural and functional effect on nucleic acids.[29-33]

1.2.2   Phase problem

As mentioned previously, crystal growth is the first challenge exist in nucleic acid X-ray crystallography, due to the repulsion of negatively charged backbone and multiple conformations of oligonucleotides. Besides this, phase problem is another challenge need to be overcome. During the diffraction data analysis, the application of a Fourier transformation is required to construct the electron density map of the macromolecule. In the Fourier transformation, both of the structure factor amplitude F and phase of the reflection α are needed for calculation of the electron density. The former can be measured during the X-ray diffraction experiment. However, the phase information is missing which is critical to produce the correct crystal structure.

There are several different strategies were applied to address this issue, such as isomorphism replacement (SIR; MIR), molecular replacement (MR) and anomalous dispersion techniques (SAD; MAD). (Table 1.1) With an available similar model, MR should be the most convenient method which don't need to obtain a heavy-atom derivative. However, for novel structures, especially nucleic acid which containing highly repetitive and stacked units[34], molecular replacement will be very difficult to solve the phase problem due to the lack of a homolog model. Thus heavy-atom method was necessary in order to solve the problem. Isomorphous replacement was successfully used in protein crystallography by soaking and co-crystallization with heavy-metal cations which often cause randomly hydrolyzation of the phosphate backbone. Moreover, it also required to prepare many isomorphous native crystals and heavy-atom derivatives to get the structure determined. Compared with isomorphous replacement, MAD technique only require one single crystal to obtain all the diffraction needed for the structure

determination, which provides a good solution to the phase problem of nucleic acid structure determination.

Table 1.1 Methods to solve the phase problem of nucleic acid crystallography

Methods	Requirement	Techniques
Direct method	Short oligonucleotides and high-resolution data	
Molecular replacement	Good homolog model	
Isomorphous replacement	Heavy-atom derivatives	Heavy metal cation soaking or co-crystallization
MAD/SAD phasing	Heavy-atom derivatization	Bromine derivative
		Indirect selenium derivatization (Se-Met-U1A method)
		Direct selenium derivatizations

1.3   Advantages of SAM on NAs

All heavy-atom used in MIR plus many that are too light for MIR can be used as anomalous scattering center for MAD, such as Se, Br, Hg, Zn, et al.[35] Selenium is the most popular anomalous scattering atom used for proteins because of the ideal K absorption edge (0.9795 Å) for X-ray diffraction experiment and the minimal structure perturbation. Selenium derivatives of nucleic acids was also developed and achieved great success in facilitating crystal growth and solving phase problem which are the two major problems in nucleic acid crystallography. During data collection, traditional 5-halogen derivatives are significantly affected by radiation damage. In contrast, radiolysis had little effect on selenium modified nucleic acids.[36] Selenium could also provide similar phasing power as bromine[37] and have more modification site to choose than bromine since it is in the same group as oxygen in the periodic table which can be replaced with selenium in principle (Figure 1.1). Previous studies on selenium modified nucleic acid X-ray crystallography show that the selenium modification facilitate the crystal growth generally[28, 38-44] and without significant structure perturbations[37, 45]. It also facilitates the structure determination of

protein-nucleic acid complex, which helped to study the function and mechanism of enzyme such as RNase-H[46, 47].
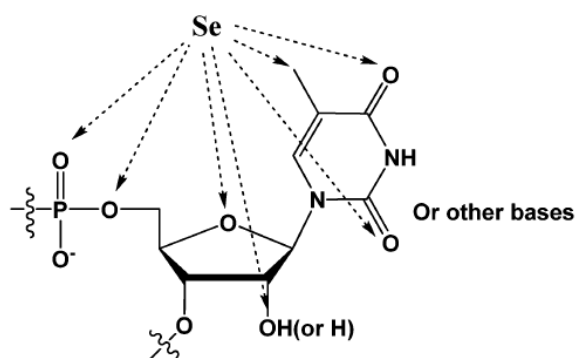
Figure 1.1 Se-modifications on nucleosides

A theoretical study[48] reveals that the 2'-methylseleno modification facilitate the crystallization of A-form DNA by destabilize the B-form helix. The large methylseleno-group cannot tolerate a B-helix geometry due to the steric hindrance with neighboring residues, but it is fitting well in the minor groove of the A-form helix, which serve as the origin of B- to A-conversion and further facilitate the crystallization. (Figure 1.2)
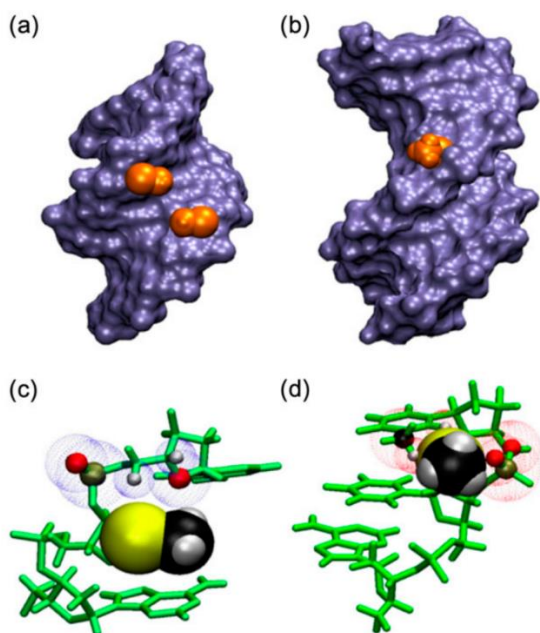
Figure 1.2 2'-SeCH3 groups implanted in A and B helical structures[48]

1.4 Recent advance in Se-derivatized nucleic acids (SeNA)

The study of Se-derivatized nucleic acids (SeNA) was started since 2001 and different kind of Se-modifications were incorporated into DNA and RNA for structural and functional studies. Those work has been well reviewed[25]. Recent years, more novel selenium modified nucleosides and nucleotides have been synthesized. Herein, we will discuss the synthesis of those novel modifications and their applications.

1.4.1 New modifications and structures

1.4.1.1 Se-modifications on sugar and phosphate

By using a protection-free one-pot triphosphates synthesis strategy [36], Lin et al. successfully synthesized nucleoside 5'-(α-P-seleno)-triphosphates (NTPαSe) and incorporate it into RNAs by T7 RNA polymerase [49]. During the triphosphates synthesis, the phosphitylating reagent was generated in situ without any purification and the reaction was highly regioselective at the 5'-hydroxyl group of nucleosides containing no protecting group on the sugar nor nucleobases. The selenium was introduced after being treated with the phosphitylating reagent by adding 3H-1,2-benzothaselenol-3-one (BTSe) at room temperature followed by hydrolysis. (Scheme 1.1) The NTPαSe analogs were purified by boronate affinity method which using the strong interaction between the cis-diols at the 2' and 3' positions of NTPs and boronate to separate NTPαSe analogs with other impurities. Compared with conventional HPLC method, the boronate affinity column give larger capacity for large-scale synthesis with less purification time and steps. Another new modification containing selenium at the 5'-position of thymidine was summarized in chapter 2.
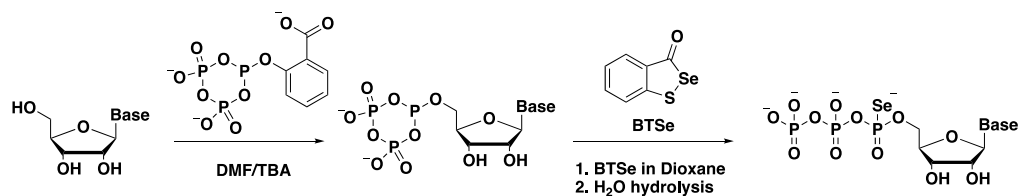
Scheme 1.1 Synthesis of the NTPαSe analogs

1.4.1.2 Se-modifications on nucleobase

2-selenium modified thymidine and uridine synthesis and modified DNA/RNA structure studies were reported which showing unique ability to enhance the base pairing specificity at the atomic level. Both studies were reviewed in chapter 2 and chapter 4 respectively. Besides the 2-Se-derivatized uridine, the selenium was also incorporated into the 4-positon of uridine to probe the U/U and U/A base pair which is also summarized in chapter 4.

6-Se-guanosine phosphoramidite, which can also be used as probe to study the RNA secondary structures, and RNAs was synthesized by Salon et al. in 2013[46]. The [Se]G-RNA also shows yellow color which is useful for protein-RNA co-crystallization visualization. It was observed that the Se-modification fit better in the bulge and wobble structure than a duplex structure. 6-Se-G-modified-RNA/DNA/RNase H complex crystallization was conducted by Salon et al.[46] The study shows that Se-modification was remove by hydrolysis naturally during the crystallization. However, surprisingly, high quality crystals were generated, and the structure of RNA/DNA/RNase H complex was determined by the molecular replacement with high resolution (1.60 Å, PDB: 3ULD) contrasted to the corresponding native RNA complex with DNA and RNase H (2.70 Å).

1.4.2 Functional study of SAM

Selenium-derivatized nucleic acids were also applied to functional studies such as RNase cleavage. The duplexes of 6-Se-G-modified-DNA/RNA were complexed with RNase H and the structure was reported with 1.80 Å resolution (PDB: 3TWH). A mechanism study was conducted

based on this high-resolution structure and compared with the corresponding native RNA/DNA/RNase H complex structure (PDB: 2G8U)[47]. In the native structure, the RNase H binds to the junction of two RNA/DNA duplexes. However, with modified Se-DNA/RNA duplex, the RNase H recognizes the internal position, which has the same sequence as in the native complex structure. The structure also shows that the selenium modification introduces a subtle duplex unwinding by 0.5-0.7 Å, and the scissile phosphate was shifted closer to the transition state of the catalytic hydrolysis which significantly accelerating the RNA cleavage by 6.2-fold faster. In the meanwhile, the scissile phosphate also positions the water nucleophile in the structure by forming a hydrogen bond (3.10 Å). These studies help us have better understanding of the guide-dependent RNA cleavage.

1.5   Potential in clinical therapeutic and drug development

Compared to protein, nucleic acid attracted much less attention as drug targets in the structure-based drug design (SBDD), partially because the lack of structural information of nucleic acids with drug candidates. By facilitating the crystallization and improving crystal quality (ref), SAM strategy can greatly promote the SBDD of nucleic acids. Those detailed 3D structure can help researcher understand the structure of drug target or even the interactions between the targets and small molecule ligands.

Moreover, oligonucleotides itself has been studied for over 30 years as potential therapeutics. The major oligonucleotides therapeutics containing antisense oligonucleotides (ASOs), aptamers, ribozymes and siRNAs, were studied extensively and lots of progress has been achieved. However, until 2017, there are only six drugs approved by FDA.[50] Many obstacles slow down the development of DNA/RNA-based therapies including the vulnerability to nucleases, off-target effects, poor delivery and low affinity. Lots of chemical modifications of nucleic acids were

developed to conquer those challenges such as phosphorothioate (PS), 2'-O-methyl (2'-OMe), locked nucleic acids (LNAs), phosphorodiamidate morpholino oligomer (PMO) and so on[51], which can increase affinity, nuclease resistant, delivery and reduce off-target effects by altering the structure and charge of the oligonucleotides.

Previous study on Se-modified nucleic acid shows that replacement of the non-bridging oxygen of DNA and RNA backbone phosphates protect the oligonucleotide from degradation by nucleases[52, 53]. Thus, the Se-modified siRNA may have improved RNase resistance, suggesting a potential therapeutic application on RNAi. Moreover, certain modifications, such as 2'-MeSe, reduce multiple conformation of nucleic acid by destabilizing other unfavorable structure[48] which can potentially increase the affinity and specificity of DNA/RNA duplex by reducing entropic penalty. The consequent longer half-life allowing effect delivery to tissues of interest. Furthermore, anticancer activity of selenium-modified nucleotides was also reported.[54]

Overall, Se-derivatized nucleic acids (SeNA) possess many unique properties and show great potential in clinical therapeutic as well as drug development. However, the studies on SeNA are still limited and the synthesis is relatively costly, which may prevent broader applications. We expect that further progress and applications of SeNA will be achieved in future.

1.6 Conclusions and Perspectives

In conclusion, selenium modification has been introduced into nucleosides and nucleotides at different position. The phasing problem could be solved rationally with the introduced selenium atom via anomalous diffraction. The modified DNAs/RNAs has been applied to a series of structure and function studies[46, 47, 55-58] and exhibit various functions and applications such as speed up A-DNA crystallization or enhance base pair specificity.

Although the B-DNA crystallization still a challenge, it is reasonable and possible to facilitate crystallization, tailoring dynamics of DNA molecules, especially local dynamics by the Se-atom-specific modification (SAM) strategy, without causing significant structure and function perturbations. By replacing oxygen atom with selenium for the phase determination in crystallography, selenium atom in the same elemental family of oxygen has been introduced to various locations of nucleic acids, especially nucleobases, without causing significant structure and function perturbations. In addition, the selenium-atom-specific modification (SAM) provides the mechanistic insights into the DNA crystallization and inter- and intra-molecular interactions, thereby facilitating discovery of the small-molecule drugs that target nucleic acids as well as nucleic acid therapeutics such as siRNA, microRNA, aptamer and antisense oligonucleotides.

2   2'-MeSe-ARABINOTHYMIDINE MODIFICATION FACILITATES LARGE B-DNA CRYSTALLIZATION FOR NEUTRON AND X-RAY STRUCTURE STUDY

2.1   Introduction

Structural biology, where X-ray crystallography is one of the most powerful methods, can greatly promote the biomolecular mechanistic research and drug discovery through examining the 3D structures of macromolecules, particularly proteins and nucleic acids. Currently, more than 140,000 protein structures and more than 3,000 nucleic acid structures have been determined, and over 85% of them have been solved through X-ray crystallography.[59, 60] However, there are two major challenges in crystallography: crystallization and phase determination (phasing), which largely slow down structure determination of these bio-macromolecules. Selenium derivatization strategy for proteins has been developed to rationally solve the phasing problem of protein crystallography, via multi-wavelength or single-wavelength anomalous diffraction (MAD or SAD), using Se-modified proteins. So far, over two thirds of novel protein structures are determined in this fashion.[35, 61-63]

2.1.1   Neutron diffraction for nucleic acid

Inspired by this protein strategy, the selenium-atom-derivatized nucleic acid (SeNA) for X-ray crystallography has been pioneered and developed for nucleic acids and their protein complexes, which solves the phasing problem rationally.[25, 64, 65] After addressing this problem in nucleic acid X-ray crystallography, we are taking on the remaining challenge: crystallization of nucleic acids. It is notoriously difficult and more challenging than that of proteins, due to nucleic acid negative charges, multiple structures, and chemical heterogeneity caused by synthesis and purification.[66-69] Moreover, it is much more difficult to grow large crystals (approximately 0.5-0.6 mm in size), required for neutron diffraction crystallography.

Generally, complete information about the nature of macromolecular H-bonds are lacked in X-ray crystallography due to the low X-ray scattering power of hydrogen atoms. Neutron nucleic acid crystallography as a complementary of X-ray can determine the orientation of water molecule with a relatively low-resolution diffraction data (2.5-2.6 Å) because of the neutrons scatter off atomic nuclei instead of electron clouds.[70, 71] Thus, even the lightest atom, hydrogen, can be observed directly. In the meanwhile, the neutrons do not cause direct radiation damage compared to X-ray diffraction, which can produce damage-free biomacromolecular structures and the data can be collected equally well at room and cryo temperatures. A joint X-ray/neutron structure of a self-complementary A-DNA oligonucleotide at both room and cyro temperatures was reported recently which containing 2'-SeMe modification on Cyt5.[72] Interestingly, a subtle difference in the structure between different temperature was observed in which a proton is replaced with hydrated Mg^{2+} upon cooling the crystal to 100 K on a backbone phosphate oxygen indicating a favorable metal binding at low temperature (Figure 2.1). Moreover, because most of crystal structures were obtained under cyro temperature, the enzyme mechanism or biological function established base on those structures may leading to significant error. The neutron crystallography has the advantage over X-rays diffraction in several aspects. However, the requirement of large crystal volume, generally >0.1 mm^3 for nucleic acids and >1 mm^3 for protein, largely restrict the application and development of neutron crystallography.
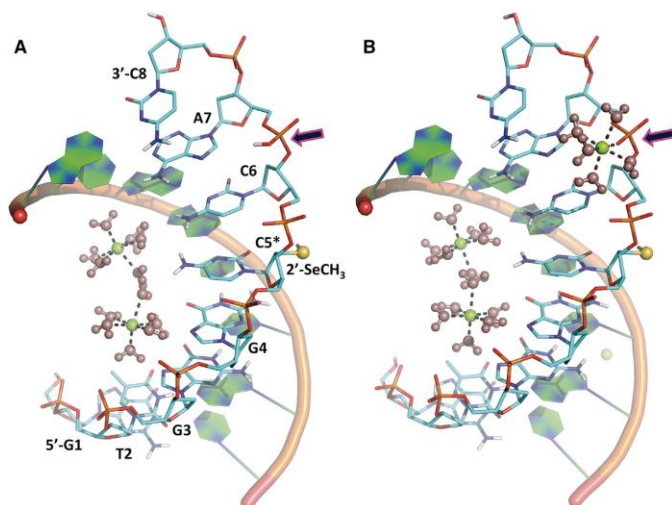
Figure 2.1 Joint X-ray/Neutron Structure of the A-DNA

2.1.2   *Selenium modifications on thymidine*

2.1.2.1 Selenium modified on nucleobase of thymidine

Selenium functionality was incorporated in the nucleobase of thymidine as different position including replacement of 2-[28] and 4-oxygen[42, 73] as well as 5-methyl group[38]. To synthesize the 4-Se-thymidine (Scheme 2.1), the 4-position needs to be activated via the formation of triazolide, followed by the Se introduction using di(2-cynoethyl) diselenide reagent[74]. Interestingly, When the selenium moiety was introduced to the 2-positon of the nucleobase, the corresponding nucleoside triphosphate as well as the DNA exhibits yellow color and a significant red shift was observed in the UV absorption (Figure 2.2).[73] Crystal structure study reveals that the replacement of O with Se on the nucleobase does not cause significant perturbation on the duplex structure as well as the formation of Se-mediated hydrogen bond with the T:A pair.[42]
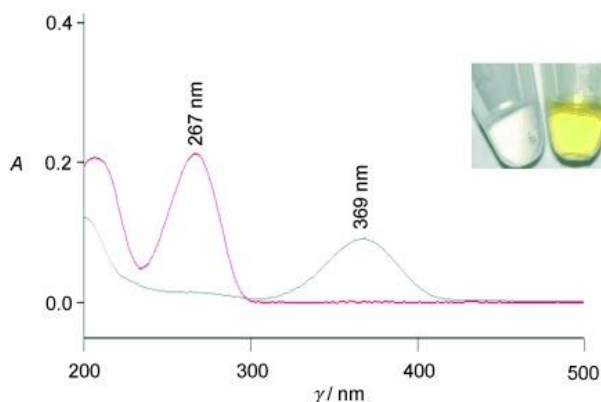
Figure 2.2 UV spectra of TTP (red) and ^{Se}TTP (gray). Inset: left: TPP (colorless); right: ^{Se}TPP (yellow).
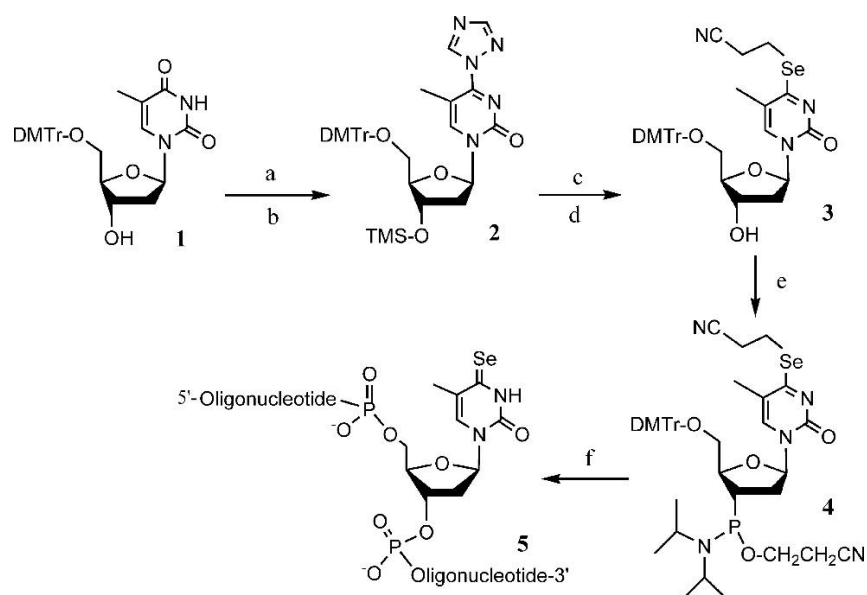
Scheme 2.1 Synthesis of 4-Se-T phosphoramidite and Se-DNA. (a) TMS-Im and CH₃CN; (b) triazole-POCl₃-Et₃N; (c) (NCCH₂CH₂Se)₂/NaBH4, EtOH; (d) 10% Et₃N in MeOH; (e) 2-cyanoethyl N,N-diisopropylchlorophosphoramidite and N,N-diisopropylethylamine in CH₂Cl₂; (f) the solid-phase synthesis.

Selenium has also been successfully incorporated into the 2-position of thymidine started from 2-thiothymidine derivative (Scheme 2.2). The biophysical and structure studies indicate that the bulky 2-Se atom discourage the T:G wobble and T:C base pairs due to the steric hindrance as well as the electronic effect, which enhance base-pairing fidelity without destabilize the DNA duplex.[28] In the 5-Se-thymidine modified DNA structure[38], a novel C-H···O hydrogen bond between the MeSe- and the phosphate group was observed which provide new insight into

atomic mechanisms of phosphorylation and dephosphorylation. Moreover, a 4, 5-diSe-modified thymidine and its DNA synthesis was also reported[75].
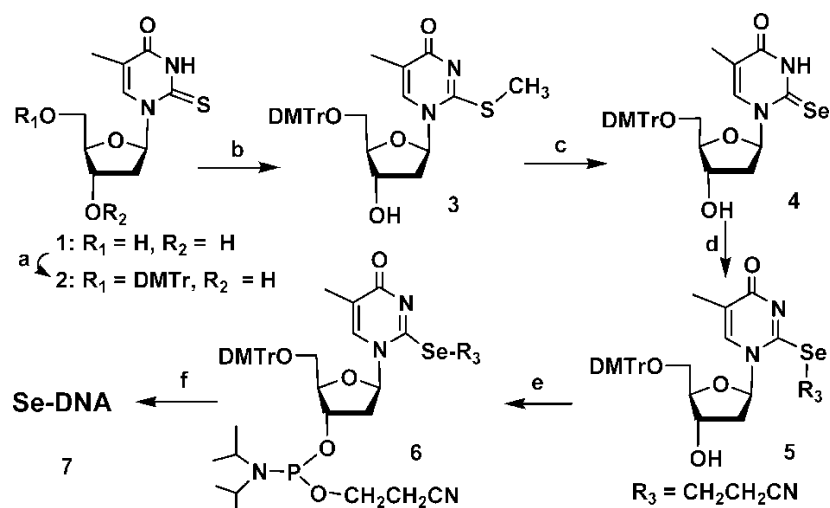
Scheme 2.2 Synthesis of 2-Se-T phosphoramidite and Se-DNA. (a) DMTr-Cl, Pyridine, DMAP, rt; (b) DBU, DMF, CH₃I; (c) Se, NaBH₄, EtOH; (d) I-CH₂CH₂CN, i-Pr₂NEt, CH₂Cl₂; (e) (i-Pr₂N)₂P(Cl)OCH₂CH₂CN, (i-Pr)₂NEt, CH₂Cl₂; (f) Solid-phase synthesis.

2.1.2.2 Selenium modified on sugar of thymidine

The synthesis of 5'-Se-derivatized thymidine and its phosphoramidite was reported by Zhang et al., which was then successfully incorporated into DNA oligo through solid-phase synthesis[76]. The protected Se-functionality was introduced within two steps with 71% yield in total. With the 5'-Se-moiety, the fluorescent group can be easily incorporated into DNA during the deprotection step under a mild condition. It was also observed that the labeled DNA is well recognized by DNA polymerase which could be used in bioreaction monitoring or disease diagnosis.

Another selenium modification which contains a methylseleno group on the 2'-position of thymidine was synthesized (Scheme 2.3) and found largely facilitated the crystallization of modified A-DNA compared to the native DNA. The thermal denaturing study reveals that the Se derivatization did not cause significant perturbation of the duplexes structure, which is consistent

with the crystal structure results[43]. Base on the structure, the 2'-Se-furanose displays the 3'-endo sugar pucker and the 2'-methylseleno group is placed in the minor groove of the duplex.
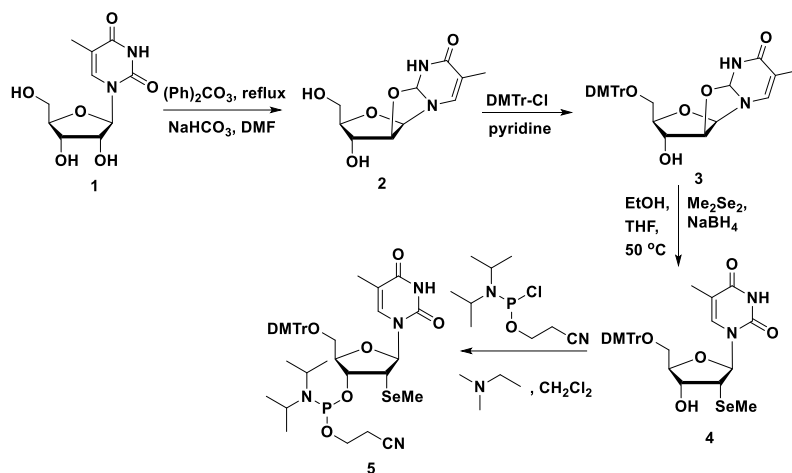
Scheme 2.3 Synthesis of 5'-DMTr-2'-MeSe-thymidine phosphoramidite

As mentioned above, crystallization is a bottle-neck step in crystallography. While, the crystallization of B-form DNA is even more challenging because of the high salt conditions during the crystallization process disfavoring the B-form duplex formation. Base on the previous study that the 2'-MeSe-thymidine modified oligonucleotides was found to largely facilitate the crystallization of A-form DNA while retaining high diffraction quality and did not cause significant structure perturbation.[77] We hypothesize that 2'-MeSe-arabinothymidine may facilitates the B-DNA crystal packing which helps the formation of large crystals.

We reported here the first synthesis of 2'-MeSe-arabinothymidine phosphoramidite and its incorporation into oligonucleotides. The Se-DNA give high quality crystal comparing to the corresponding native DNA and facilitates large crystal growth up to 0.6 mm in size. Moreover, our biophysical and structural studies reveal that the Se-DNAs retain a similar structure compared to the native DNAs without significant perturbation. The B-factor comparison between the Se-DNA and the native reflects the impact of the selenium derivation on the local structure flexibility without affects the global molecular dynamics.

2.2 Results and discussion

2.2.1 The synthesis of 2'MeSe-arabinothymidine and its DNA oligonucleotide

The synthesis of 2'-MeSe-arabinothymidine derivative (**8**) started from the protection of the 5'-, 3'-hydroxyl groups of thymidine (**1**) with tetraisopropyldisilylene (TIPDS). To avoid the generation of 2'-MeSe-ribothymidine,[43] protection of the 3-N is required before the MeSe functionality incorporation. The protective group needs to be stable enough to tolerate the Se introduction condition, and, in the meanwhile, deprotected easily afterwards without affect the incorporated MeSe group. Initially, we attempted different protective groups including benzyl, benzoyl and tert-butyloxycarbonyl (BOC). The benzyl group is stable enough for the Se incorporation, however, difficult to remove possibly due to the selenium functionally "poisoned" the catalyst during the hydrogenation deprotection. The problem was later overcome by using Benzyloxymethyl acetal (BOM) as protecting group, which can be later cleaved with strong acid. After the BOM protection, the 2'-OH was activated by mesylate. Then the MeSe incorporation was conducted at a high temperature (90 °C) in THF in a pressure vessel, which give 80% yield in 6 h. After the deprotection of BOM, **6** was treated with HF to remove the TIPDS group and the 5'-OH was selectively protected with the DMTr group. Finally, the 2'-Se-arabinothymidine (**8**) was converted to corresponding phosphoramidite (**9**) in 90% yield. (Scheme 2.4)
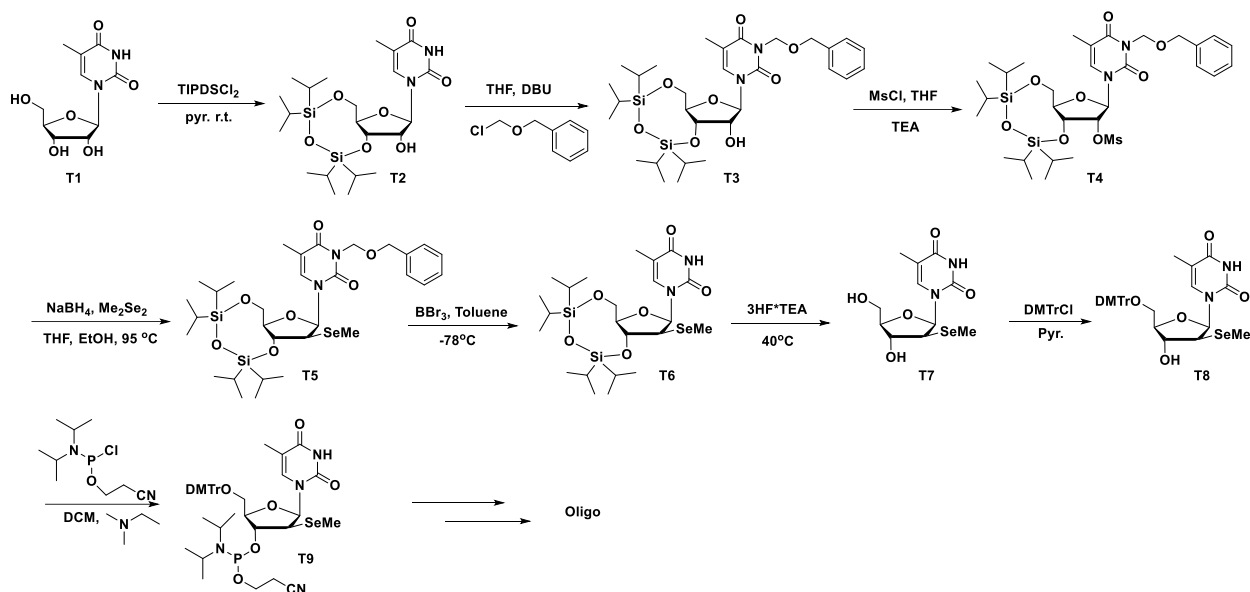
Scheme 2.4 Synthesis of 5'-DMTr-2'-MeSe-arabinothymidine phosphoramidite and Se-modified Oligonucleotides

The phosphoramidite **9** was incorporated into DNA oligos via standard solid-phase synthesis. The Se functionality was found stable with the treatment of I_2 (20 mM, 20 s) and trichloroacetic acid (TCA) during the oligo synthesis. The synthesized DNA oligo was cleaved off the solid support and fully deprotected by treating with concentrated ammonia at 55 ^{o}C for 5 h, followed by the reverse-phase HPLC purification. All the purified DNA oligo was further confirmed by MALDI-TOF MS analysis (Table 2.1). Figure 2.3 shows the typical selenium isotopic distribution of the Se-DNA.

Table 2.1 MALDI-TOF MS Data of Se-DNAs

Entry	Se-oligonucleotides	Measured (calcd) m/z
a	B-Ts1 5'-CGCGAA$^{2'Se}$TTCGCG-3' $C_{117}H_{149}N_{46}O_{70}P_{11}Se$ FW 3739.4	[M+H]$^+$: 3740.1 (3740.4)
b	B-Ts2 5'-CGCGAAT$^{2'Se}$TCGCG-3' $C_{117}H_{149}N_{46}O_{70}P_{11}Se$ FW 3739.4	[M+H]$^+$: 3740.1 (3740.4)
c	B-Td3 5'-CGCGAA$^{2'Se}$T$^{2'Se}$TCGCG-3' $C_{118}H_{151}N_{46}O_{70}P_{11}Se_2$ FW 3833.5	[M+H]$^+$: 3834.8 (3834.5)
d	Rnase-s1	[M+H]$^+$: 1901.5 (1901.3)

	5'-A$^{2'Se}$TGTCG-3' $C_{60}H_{77}N_{22}O_{35}P_5Se$ FW 1900.3	
e	Rnase-s2	[M+H]$^+$: 1901.4 (1901.3)
	5'-ATG$^{2'Se}$TCG-3' $C_{60}H_{77}N_{22}O_{35}P_5Se$ FW 1900.3	
f	Rnase-d3	[M+H]$^+$: 1995.4 (1995.2)
	5'-A$^{2'Se}$TG$^{2'Se}$TCG-3' $C_{61}H_{79}N_{22}O_{35}P_5Se_2$ FW 1994.2	
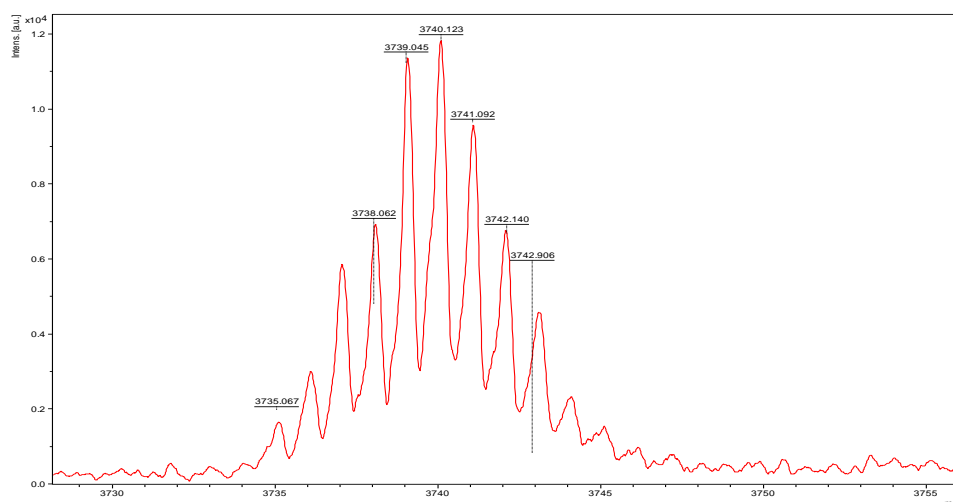
Figure 2.3 Typical selenium isotopic distribution of Se-DNA

2.2.2   Thermostability and Circular dichroism studies

To compare the thermostability of the Se-modified oligonucleotides with the corresponding native DNA, the UV denaturation experiment was conducted. The results, presented in Table 2.2, show that the melting temperatures are similar between the Se-modified DNA and the native ones, indicating that the Se functionality does not cause significant perturbation. Circular dichroism (CD) was also carried out to study their structure change in the solution. Positive peaks at 280 nm, 220 nm and negative peak at 245 nm were observed in all oligos, and no shift was founded comparing to the native DNA, which suggested that the native and Se-DNAs has similar helixes in solution. (Figure 2.5)

Table 2.2 Melting temperature of Native and Se-DNA duplexes

Entry	Sequence	Melting temperature (°C)	ΔT_m(°C)

a	5'-CGCGAATTCGCG-3'	64.0	-
b	5'-CGCGAA^{Se}TTCGCG-3'	61.5	2.5
c	5'-CGCGAAT^{Se}TCGCG-3'	61.1	2.9
d	5'-CGCGAA^{Se}T^{Se}TCGCG-3'	55.8	8.2
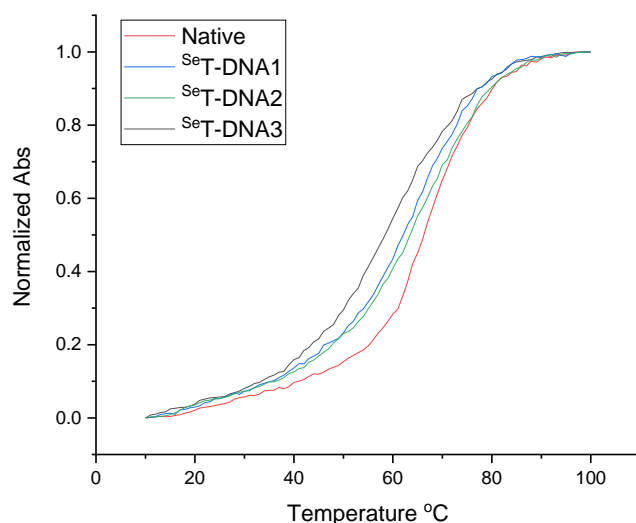
Figure 2.4 Normalized UV melting curves of DNAs.
The native DNA duplex: (5'-CGCGAATTCGCG-3')$_2$. The ^{Se}T-DNA1 duplex: (5'-CGCGAA^{Se}TTCGCG-3')$_2$. The ^{Se}T-DNA2 duplex: (5'-CGCGAAT^{Se}TCGCG-3')$_2$. The ^{Se}T-DNA3 duplex: (5'-CGCGAA^{Se}T^{Se}TCGCG-3')$_2$.
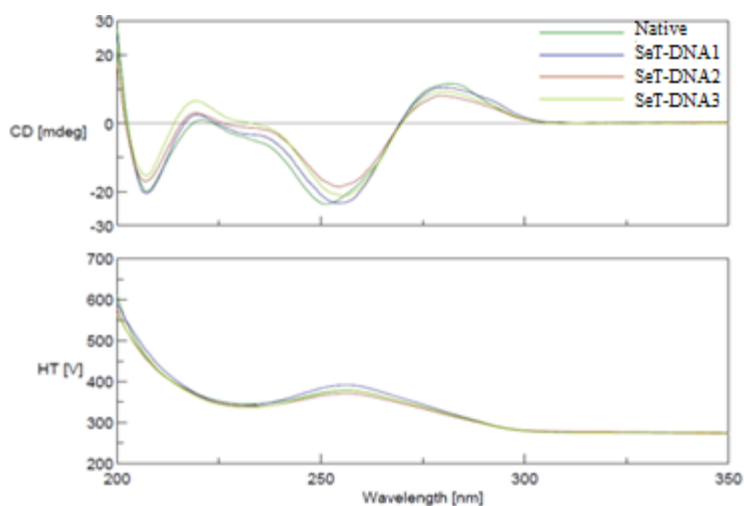
Figure 2.5 circular dichroism spectroscopy of native and ^{Se}T-DNAs.
The native DNA duplex: (5'-CGCGAATTCGCG-3')$_2$. The SeT-DNA1 duplex: (5'-CGCGAA^{Se}TTCGCG-3')$_2$. The SeT-DNA2 duplex: (5'-CGCGAAT^{Se}TCGCG-3')$_2$. The SeT-DNA3 duplex: (5'-CGCGAA^{Se}T^{Se}TCGCG-3')$_2$.

2.2.3 Large crystal growth facilitated by the modification

In the meanwhile, we also monitored the crystal growth of the native and modified DNA oligos under different buffer conditions (Appendix B). Interestingly, the SeT-modification slow down the crystal growth dramatically (Figure 2.6, left). And, in each droplet, there is one crystal formed compare to the native which generally form 2-10 small crystals. Attributing to the slow growth, the SeT-DNA form crystals that generally 40 – 60 times larger in volume than the native under same buffer condition, and the largest crystal is up to 600 microns (Table 2.3). 1 mM single strand concentration is required for the crystal growth. The improved crystal volume with Se-modified DNA may help break the bottleneck of crystal volume for neutron crystallography. We are currently in the process of neutron diffraction experiment.
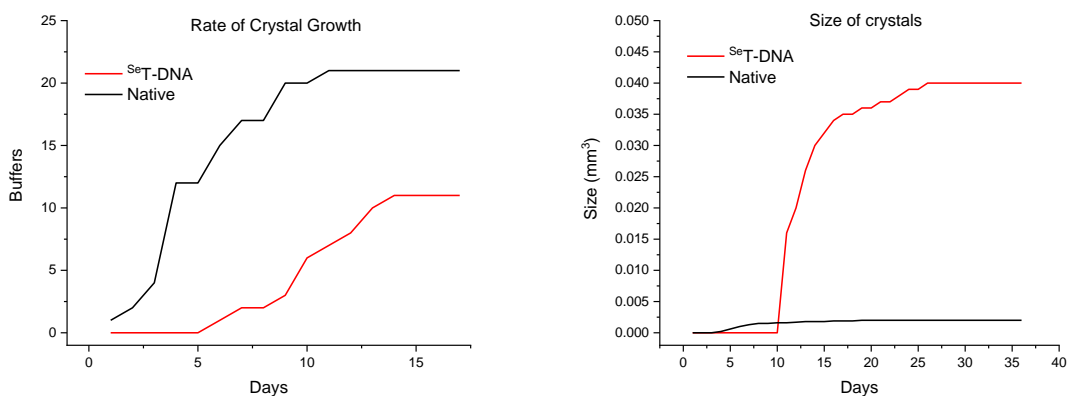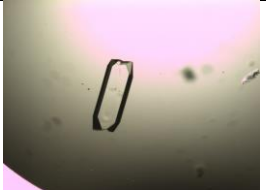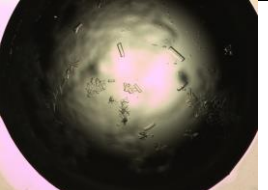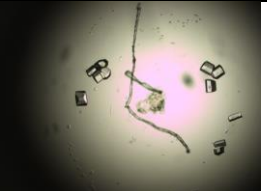
Figure 2.6 Comparison of crystal growth between native and ^{Se}T-DNA2.
Left: the comparison of the rate of crystal growth between SeT-DNA2 and the corresponding native; Right: the comparison of the size of crystals between SeT-DNA2 and the corresponding native.

Table 2.3 crystal comparison of SeT-DNA2 and the native

Se-DNA2 buffer #9	Se-DNA2 buffer #10	Native-DNA buffer #9	Native-DNA buffer #10

2.2.4   Structure determination of modified DNA and protein-DNA complex

The determined crystal structure (1.15 Å resolution) of Se-DNA (Figure 2.7) is superimposable to the corresponding native structure (1.26 Å resolution) with same space group, and the structures were virtually identical indicating that the selenium modification at 2'-position does not cause significant structure perturbation. The structure reveals that the 2'-Se-arabino-furanose displays the 2'-endo sugar pucker and the 2'-methylseleno group is in the major groove of the B-DNA duplex. The Se-modified and the native structures are also virtually identical with a RMSD of 0.176Å for all 454 atoms which support that the incorporation of 2'-MeSe-arabinothymidine does not alter the global structure of the complex (Figure 2.7). The data collection and structure refinement statistics are summarized in Table 2.4.

Table 2.4 Data collection and refinement statistics of ^{Se}T-DNA and the corresponding native

Data Collection	SeT-DNA2	Native
Wavelength (Å)	1	1
Resolution range (Å)	50 - 1.15	50 - 1.26
	(1.17 – 1.15)	(1.28 - 1.26)
Space group	P 21 21 21	P 21 21 21
Unit-cell a, b, c (Å)	25.4, 40.0, 65.7	25.5, 40.2, 65.7
α, **β**, **γ** (°)	90, 90, 90	90, 90, 90
Number of reflections	24146 (1078)	18494 (904)
Completeness (%)	98.6 (93.5)	98.0 (97.0)
I/σ	45 (10)	26 (4.8)
R merge (%)	3.8 (24.9)	5.7 (50.3)
Refinement		
Resolution range (Å)	34.18 - 1.15	34.31 - 1.26
Number of reflections	22851	17319
Completeness (%)	98.4	96.5
R value (%)	20.7	20.6
R free (%)	22.5	23.2
Average B value (Å^2)	17.9	16.4
Rms Bond Length (Å)	0.019	0.014
Rms Bond Angle (°)	2.08	1.91
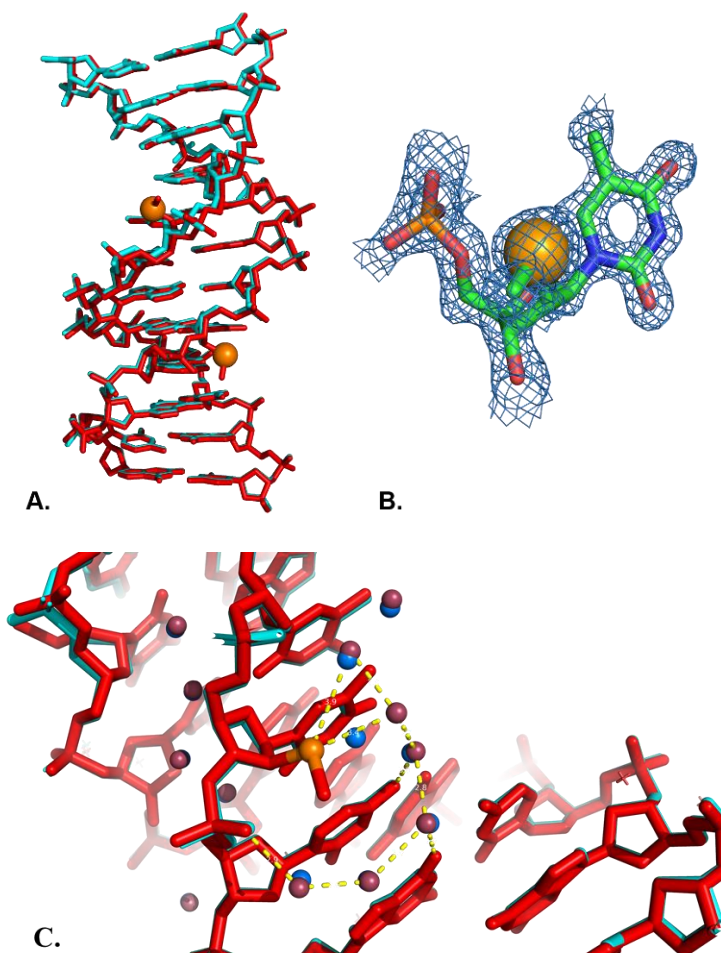
Figure 2.7 A. Global and B. local structure of the ^{Se}T-DNA2 [5'-CGCGAAT-^{Se}T-CGCG-3'] with a resolution of 1.15 Å. (Se-DNA, red; Native DNA, cyan); C. Hydration pattern comparison.

The impact of 2'-arabino-SeMe-dT modification on duplex hydration was also studied. The comparison between the modified DNA and the corresponding native show similar hydration pattern in both minor groove as well as major groove (Figure 2.7, C). The introduction of MeSe group repulses one of the water molecule away which may cause minor hydration alteration in the major groove of duplex.

When examine the crystal contact between DNA molecules, we found that the selenium functionality at the second thymidine (5'-CGCGAAT**T**CGCG-3') would not affect their interaction (Figure 2.8). However, the methylseleno-group located on the first thymidine (5'-

CGCGAA**T**TCGCG-3') may have direct interaction with the nearby DNA molecule which could further influence the crystal packing and growth.
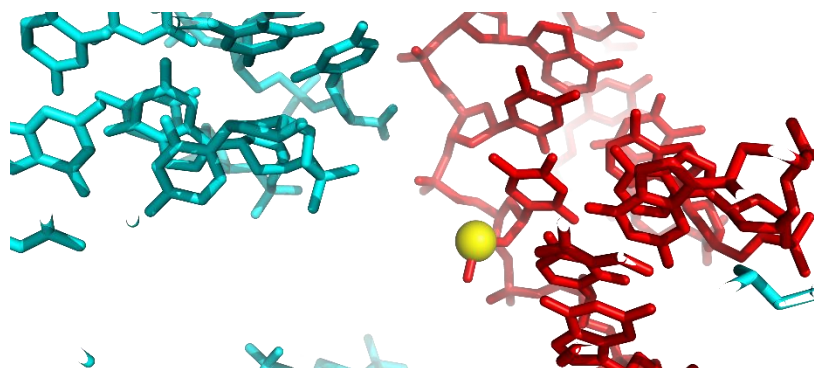
Figure 2.8 Packing of 2'SeT-modified DNA (5'-CGCGAATTCGCG-3')

Bst DNA polymerase is a member of the DNA polymerase I family responsible for maintaining the genomic integrity which has been extensively studied as a model system for DNA synthesis. Here, the crystal of the fragment DNA polymerase with DNA oligo containing 2'-β-MeSe located on two different thymidine (5'-CGCGAA-^{Se}T-TCGCG-3' and 5'-CGCGAAT-^{Se}T-CGCG-3') were grown (Figure 2.9) and the structures were determined at 2.64 Å and 2.90 Å respectively. The data collection and structure refinement statistics are summarized in Table 2.5. The modified BST-DNA structures were superimposed with the corresponding native structure and show no significant perturbation in its global geometry (Figure 2.10, left). However, by comparing the sugar conformation in the duplex, we found that the location of the modification will directly change the sugar puckering. When the 2'-MeSe group was located at the first T of the sequence from the 5'-end, the sugar conformation of the modified thymidine was changed to 3'-endo, which normally exist in the A-form DNA, on only one of the chain in the duplex with another chain retain the same sugar puckering. Moreover, the MeSe group was pointed into the minor groove of the DNA duplex (Figure 2.10, right). While, when the selenium moiety was

located on the second T, the sugar pucker of the thymidine exhibits same conformation as the native oligonucleotides with the MeSe group pointed to the major groove (Table 2.6).
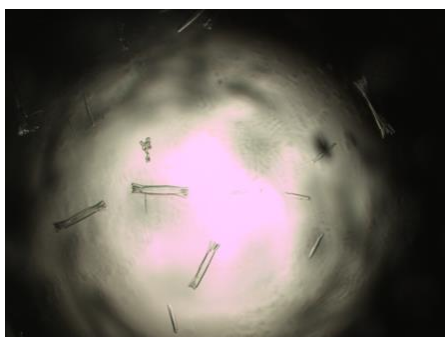
Figure 2.9 BST-DNA complex crystal

Table 2.5 Crystallographic data collection and refinement statistics of ^{Se}T-DNA/Bst complex

DNA #	1	2	3
Wavelength (Å)	1	1	1
Resolution range (Å)	50.0-1.85 (1.92-1.85)	50.0-2.90 (2.95-2.90)	50.0-2.64 (2.69-2.64)
Space group	P2$_1$2$_1$2$_1$	P2$_1$2$_1$2$_1$	P2$_1$2$_1$2$_1$
Unit-cell a, b, c (Å)	86.7, 94.1, 105.0	86.9, 94.5, 105.8	89.4, 93.7, 105.3
α, β, γ (°)	90, 90, 90	90, 90, 90	90, 90, 90
Unique reflections	73185(6683)	20086 (980)	26773 (1297)
Completeness (%)	99.1(91.5)	100.0 (100.0)	99.9 (98.8)
Rmerge (%)	7.4 (43.1)	17.4 (55.9)	10.9 (45.9)
I/σ(I)	32.7 (3.6)	17.3 (5.2)	16.5 (3.2)
R value (%)	18.7	16.8	19.0
Rfree (%)	22. 4	23.4	25.2
Average B value (Å^2)	22.6	44.6	39.6
R.m.s.d. bond length (Å)	0.019	0.009	0.009
R.m.s.d. bond angle (°)	1.94	1.64	1.59
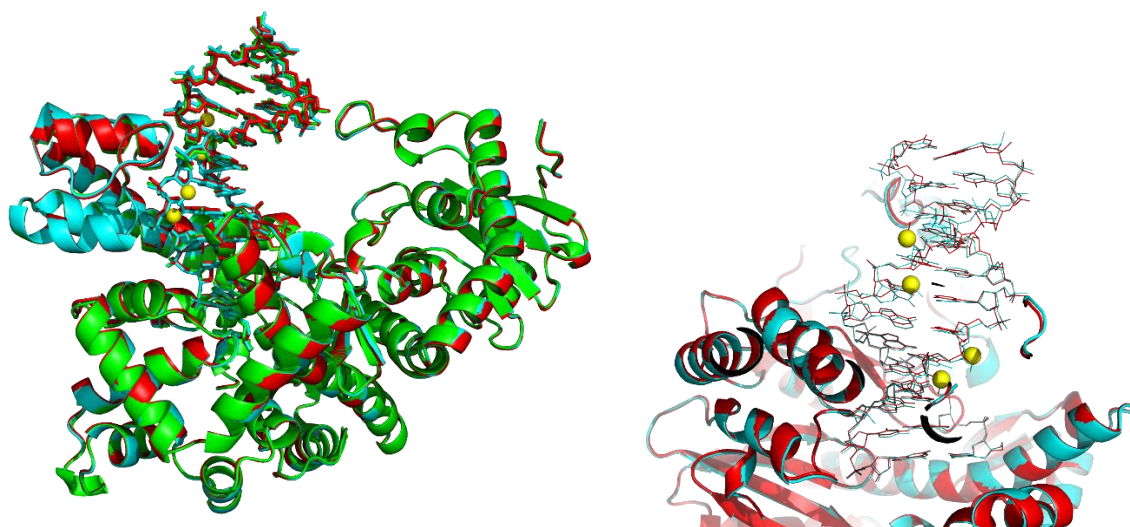
Figure 2.10 The superimposed structure of the fragment of Bst polymerase with DNAs. (Green, native; cyan, Se-DNA, 5'-CGCGAATTCGCG-3'; Red, Se-DNA, 5'-CGCGAATTCGCG-3')

Table 2.6 Sugar Pucker Confirmation of Bst-DNA complexes

#	Base	Native		Ts1		Ts2	
		Chain C	Chain B	Chain C	Chain B	Chain C	Chain B
1	C	C3'-endo	O4'-exo	C3'-endo	O4'-exo	O4'-exo	O4'-exo
2	G	C3'-endo	C3'-exo	C3'-endo	C3'-exo	C3'-endo	C3'-exo
3	C	C3'-endo	C1'-exo	C3'-endo	C2'-endo	C3'-endo	C2'-endo
4	G	C2'-endo	C2'-endo	C3'-exo	C2'-endo	C2'-endo	C2'-endo
5	A	C2'-endo	C2'-endo	C3'-exo	C3'-exo	C2'-endo	C2'-endo
6	A	C1'-exo	C1'-exo	C4'-exo	C2'-endo	C1'-exo	C2'-endo
7	T	O4'-endo	C1'-exo	C3'-endo	O4'-endo	O4'-endo	C2'-endo
8	T	C1'-exo	C1'-exo	C2'-endo	C1'-exo	C1'-exo	C1'-exo
9	C	C2'-endo	C2'-endo	C2'-endo	C2'-endo	C2'-endo	C1'-exo
10	G	C2'-endo	C3'-endo	C2'-endo	C3'-endo	C2'-endo	C3'-endo
11	C	C2'-endo	C3'-endo	C2'-endo	C3'-endo	C2'-endo	C3'-endo
12	G	C1'-exo	O4'-endo	C2'-endo	C3'-endo	C1'-exo	C3'-endo

Rnase H is a sequence-nonspecific endonuclease that digest the RNA of a DNA/RNA duplex which is involved in many important biological processes as well as silencing gene expression through antisense mechanism. Here, we also grown the Rnase H-DNA/RNA complex crystals (Figure 2.11) with a selenium modified thymidine at two different location (5'-A-[Se]T-GTCG-3' and 5-ATG-[Se]T-CG-3'). The X-ray crystal structure of RNase H/RNA/Se-DNA

complex was determined at a resolution of 1.47 Å. Interestingly, when the 2'-β-MeSe group is located on the first thymidine (5'-A-^{Se}T-GTCG-3'), the complex structure is virtually identical to the corresponding native (Figure 2.12). However, when the 2'-β-MeSe group is located on the second thymidine (ATG-^{Se}T-CG-3'), a disordered structure was obtained based on the electron density map that the DNA/RNA duplex exhibits two possible binding positions with the RNase H protein (Figure 2.13), which indicates that the selenium moiety may influence the binding site.
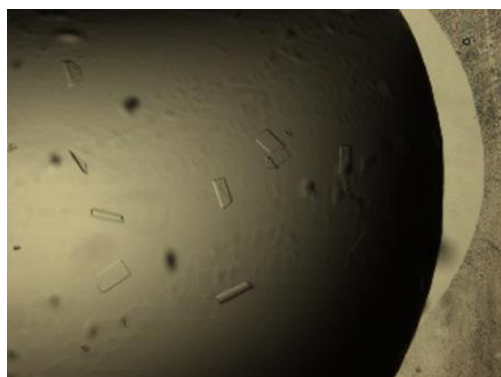
Figure 2.11 RNaseH-(^{Se}T)DNA/RNA complex crystals

Table 2.7 Crystallographic data collection and refinement statistics of RNaseH-(^{Se}T)DNA/RNA complex

DNA #	1	2	3*
Wavelength (Å)	1	1	1
Resolution range (Å)	50.0-1.70 (1.76-1.70)	50.0-1.48 (1.51-1.48)	50.0-1.45 (1.48-1.45)
Space group	C2	C2	C2
Unit-cell a, b, c (Å)	80.2, 37.6, 62.1	81.2, 37.7, 62.1	80.9, 37.7, 62.1
α, β, γ (°)	90, 96.0, 90	90, 96.3, 90	90, 96.8, 90
Unique reflections	18866 (1081)	30461 (1283)	33258 (1608)
Completeness (%)	92.1 (53.7)	97.4 (83.7)	100.0 (100.0)
Rmerge (%)	7.7 (23.5)	6.8 (34.2)	3.3 (46.2)
I/σ(I)	21 (3.5)	26 (4.0)	54 (6.7)
R value (%)	18.9	19.8	-
Rfree (%)	23.4	22.3	-
Average B value (Å²)	15.6	15.3	-
R.m.s.d. bond length (Å)	0.019	0.014	-
R.m.s.d. bond angle (°)	1.85	1.91	-
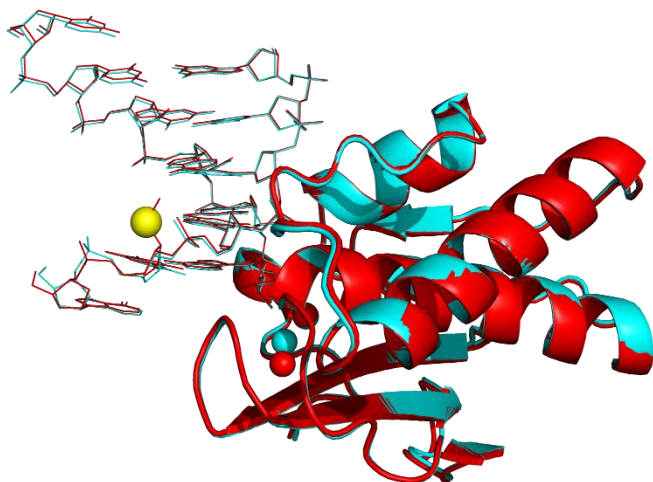
Figure 2.12 Superimposed structure of RNaseH-^{Se}T-DNA complex (5'-A-^{Se}T-GTCG-3') with the native
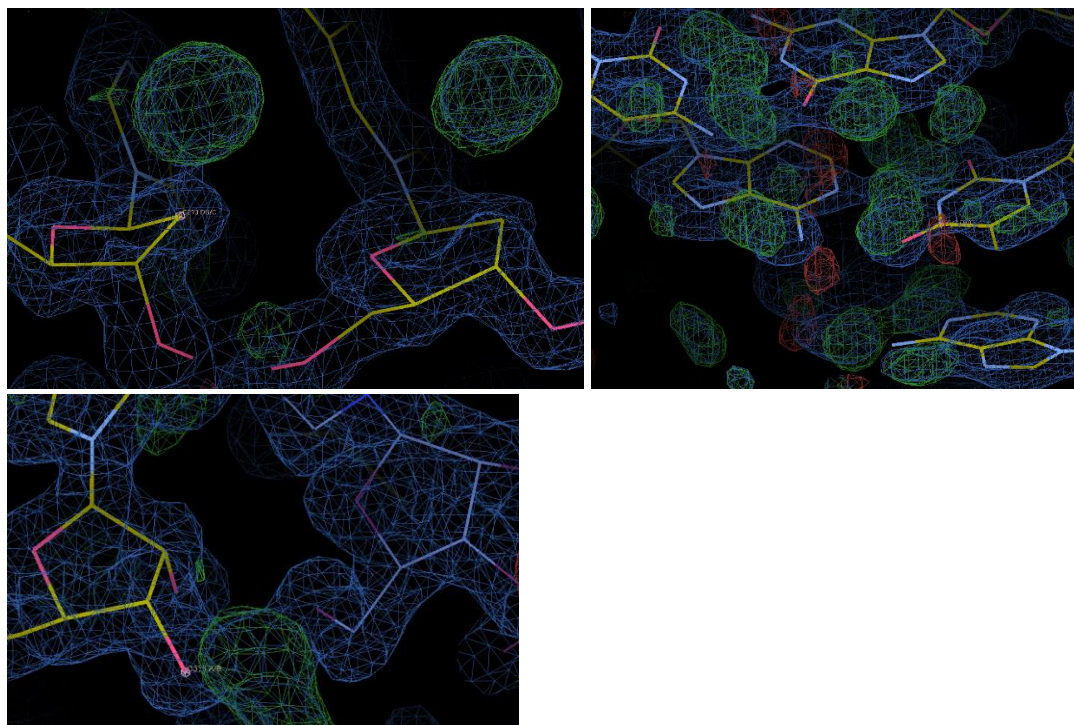
Figure 2.13 Electron density map of RNaseH-^{Se}T-DNA complex (ATG-^{Se}T-CG-3')

2.2.5   B-factor comparison

Besides providing the average static picture, X-ray diffraction can further explore the molecular dynamics through a well-defined parameter so-called B-factor. Crystallographic B-factor describes the attenuation of X-ray diffraction result from the thermal motion, which reflects

the rigidity, flexibility and internal motion of the structure.[78] However, their interpretation should be treated cautiously due to the B-factor also correlates with the resolution of the structure. The B-factor can be up to 100-200 with a resolution as low as 3-5 Å which is ineligible to draw specific conclusions. Thus, high resolution data is required when interpret the B-factor regarding the molecular dynamics. The high-resolution data we obtained allows the comparison of the B-factors between the Se-DNA and the native one, which give insight into the impact of the Se functionality to the local molecular dynamics. B factors were extracted from the analyzed structures and normalized using unity-based scaling for comparison.[79] The scaled B-factor $B_{scaled(x)}$ for atom x was calculated according to the formula

$$B_{scaled(x)} = 99[B_x - B_{min}]/[B_{max} - B_{min}] + 1 \quad (1)$$

Interestingly, large fluctuations of phosphates were observed for both Se-modified and native DNAs, while the nucleobases are much less flexible (Figure X.). Surprisingly, significant decrease of the B-factor was observed on one of the DNA strands at the phosphate group of the Se-T as well as the adjacent cytidine residue compared to the native one which suggesting that the selenium derivation restricts its movement and enhances rigidity. However, the similar average B-value of the Se-DNA (17.9 Å^2) and the corresponding native one (16.4 Å^2) indicating that the Se functionality does not change the global molecular dynamics of the DNA duplex significantly. Consistently, the B-factor comparison between the RNase H-RNA/Se-DNA complex and the corresponding native shows no significant changes (Figure X).
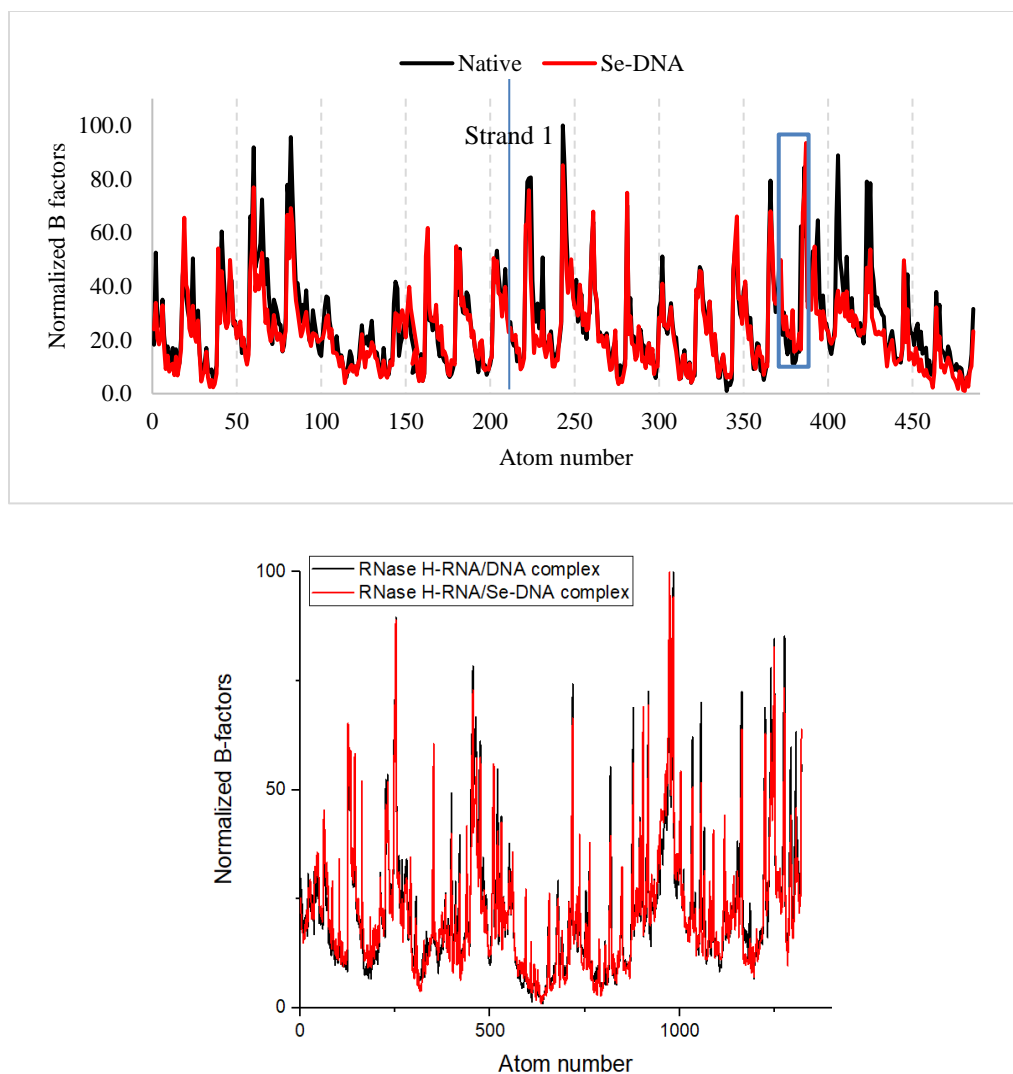
Figure 2.14 Comparison of normalized B-factor profiles from Se-DNA (red line) with the corresponding native (black line)
(The MeSe-group was not included in the profile. Every nucleoside and phosphate group in the DNA sequence has been labelled and the Se-modified thymidine are highlighted with red font.)

2.3   Conclusion

In conclusion, we developed a route to synthesize the 2'-MeSe-arabinothymide analogues and oligonucleotides containing selenium modification. High resolution structures were obtained through X-ray crystallography for both DNA duplex and RNase H/RNA/DNA complex (1.15 Å and 1.47 Å respectively). The structure shows that the 2'-β-position selenium derivatization retains the native 2'-endo conformation in the B-form DNA duplex and the modification does not cause

any significant structure perturbation which is consistent with the UV melting and CD studies results. The structures of Bst-DNA complex show that the location of the modified nucleoside will affect the sugar pucker conformation. B-factor comparison shows that the Se-modification restrict the local molecular flexibility instead of the global molecular dynamics. Furthermore, we observed that the 2'-Se-araT modified oligo formed larger crystal than the native oligo, which it has potential to aid the structure study for neutron crystallography.

2.4 Experimental section

2.4.1 General

Most solvents and reagents were purchased from Sigma, Fluka, or Aldrich (p.a.) and used without purification unless mentioned otherwise. Triethylamine (TEA) was dried over KOH (s) and distilled under argon. When necessary, solid reagents were dried under high vacuum. Reactions with compounds sensitive to air or moisture were performed under argon. Solvent mixtures are indicated as volume/volume ratios. Thin layer chromatography (TLC) was run on Merck 60 F254 plates (0.25 mm thick; Rf values in the text are for the title products), and visualized under UV-light or by a Ce-Mo staining solution (phosphomolybdate, 25 g; $Ce(SO_4)_2 \cdot 4H_2O$, 10 g; H_2SO_4, 60 mL, conc.; H_2O, 940 mL) with heating. Flash chromatography was performed using Fluka silica gel 60 (mesh size 0.040-0.063 mm) using a silica gel:crude compound weight ratio of ca. 30:1. ^1H, ^{13}C and ^{31}P-NMR spectra were recorded using Bruker-300 or 400 (300 or 400 MHz). All chemical shifts (δ) are in ppm relative to tetramethylsilane and all coupling constants (J) are in Hz. High resolution (HR) MS were either obtained with electrospray ionization (ESI) on a Q-TOFTM Waters Micromass at Georgia State University.

2.4.2 Synthesis of phosphoramidite and characterization
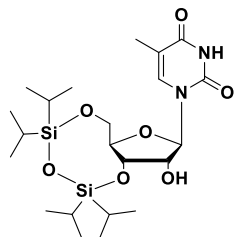
3′,5′-O-(1,1,3,3-tetraisopropyldisiloxane-1,3-diyl)- 5-methyl-uridine (**2.1**) 5-methyl-uridine (10 g, 38.7 mmol) was suspended in 200 mL dry pyridine with argon and the mixture was placed in an ice-bath. To the suspension, 1,3-dichloro-1,1,3,3-tetraisopropyl-disiloxane (13.4g, 42.6 mmol) was added dropwise. Then the ice-bath was removed, and the reaction was stirred at room temperature for 5 h. After the reaction completed, 5 mL of Methanol was added. The solvent was evaporated under reduced pressure and dissolved in methylene chloride (300 mL). The organic solution was washed with HCl (3 M, 75 mL), saturated sodium bicarbonate (75 mL), water (100 mL), brine (100 mL), dried over anhydrous $MgSO_4$, and concentrated in vacuum to given **2.1** as white solid. The product was used in the next step without further purification. ^{1}H-NMR (400 MHz, $CDCl_3$) δ (ppm): 1.03-1.09 (m, 28H, 4×iPrSi), 1.91 (s, 3H, CH_3), 3.31 (br, 1H, OH), 4.01 (dd, J_1=2.2 Hz, J_2=13.1 Hz, 1H, H-5'), 4.09 (d, J=8.4 Hz, 1H, H-4'), 4.19 (m, 2H, H-2', H-5'), 4.40 (dd, J_1=5.1 Hz, J_2=8.4 Hz, 1H, H-3'), 5.71 (s, 1H, H-1'), 7.41 (s, 1H, H-6), 8.91 (br, 1H, NH). ^{1}H-NMR spectrum is identical to the literature[80].
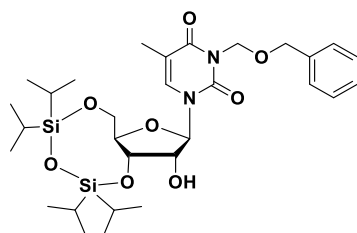
3′,5′-O-(1,1,3,3-tetraisopropyldisiloxane-1,3-diyl)-N3-(benzyloxymethylacetal)-5-methyl-uridine (2.2) Compound **2.1** (5 g, 10 mmol) was dissolved in 50 mL THF followed by the addition of 1,8-Diazabicyclo(5.4.0)undec-7-ene (DBU, 4.5 mL, 30 mmol). The mixture was

treated with benzyl chloromethyl ether (2.7 mL, 20 mmol) and stirred at room temperature for 2 h. The solvent was evaporated under reduced pressure and washed with water (30 mL), brine (50 mL), dried over $MgSO_4$ and concentrated. The residue was purified by silica gel chromatography (25 % ethyl acetate in hexane) to give desired product **2.2** (5.6 g, 90 %). ^{1}H-NMR (400 MHz, $CDCl_3$) δ (ppm): 1.03-1.10 (m, 28H, 4×iPrSi), 1.91 (s, 3H, CH_3), 2.89 (d, 1H, OH), 4.01 (dd, J_1=2.9 Hz, J_2=13.1 Hz, 1H, H-5'), 4.06 (td, J_1=3.7 Hz, J_2=8.5 Hz, 1H, H-4'), 4.14 (d, J=5.1 Hz, 1H, H-2'), 4.19 (dd, J_1=2.0 Hz, J_2=13.1 Hz, 1H, H-5'), 4.39 (dd, J_1=5.2 Hz, J_2=8.5 Hz, 1H, H-3'), 4.72 (s, 2H, $CH_2OC\underline{H}_2Ph$), 5.49 (d, J=3.1 Hz, 2H, $C\underline{H}_2OCH_2Ph$), 5.72 (s, 1H, H-1'), 7.38-7.23 (m, 6H, aromatic, H-6). ^{13}C-NMR (100 MHz, $CDCl_3$) δ (ppm): 12.7, 12.6 (i-Pr), 13.0 (CH_3), 13.5, 13.3, 17.4, 17.4, 17.3, 17.2, 17.08, 17.0, 17.0, 16.9 (i-Pr), 60.4 (C-5'), 69.2 (C-3'), 70.5 (CH_2), 72.3 (CH_2), 75.1 (C-2'), 81.9 (C-4'), 91.4 (C-1'), 109.9 (C-5), 127.6, 127.7, 128.3 (Ar-C), 134.3 (C-6), 138.0 (Ar-C), 150.7 (C-2), 163.4 (C-4). HRMS (ESI): $C_{30}H_{48}N_2O_8Si_2$; $[M+H]^+$: 621.1884 (calc. 621.3027). ^{1}H-^{13}C HSQC, ^{1}H-^{13}C HMBC and ^{1}H-^{1}H COSY (Appendix D.1).
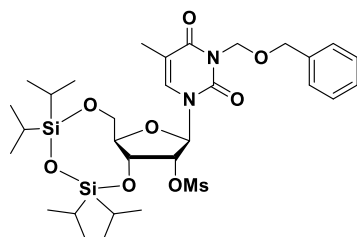
2'-O-mesyl-3′,5′-O-(1,1,3,3-tetraisopropyldisiloxane-1,3-diyl)-N3-(benzyloxymethyl acetal)-5-methyl-uridine (2.3) Compound **2.2** (3.5 g, 5.6 mmol) was dissolved in 35 mL THF followed by the addition of triethylamine (3.1 mL, 22.4 mmol). To the solution, methanesulfonyl chloride (0.88 mL, 11.2 mmol) was added dropwise and the reaction was stirred at room temperature for 1 h. The solvent was evaporated under reduced pressure and dissolved in methylene chloride (100 mL). The organic solution was washed with water (30 mL) and brine (40 mL), dried over $MgSO_4$ and concentrated. The residue was purified by silica gel chromatography

(15 % ethyl acetate in hexane) to give desired product **2.3** (3.5 g, 90 %). ^{1}H-NMR (400 MHz, CDCl$_3$) δ (ppm): 1.04-1.10 (m, 28H, 4×iPrSi), 1.93 (s, 3H, CH$_3$), 3.19 (s, 3H, O$_3$SCH$_3$), 4.07-4.16 (m, 2H, H-5'), 4.43-4.44 (m, 1H, H-4'), 4.68 (s, 2H, CH$_2$OC$\underline{H}_2$Ph), 5.23 (t, *J*=5.3 Hz, 1H, H-2'), 5.41 (t, *J*=4.7 Hz, 1H, H-3'), 5.50 (s, 2H, C$\underline{H}_2$OCH$_2$Ph), 6.14 (d, *J*=5.2 Hz, 1H, H-1'), 7.25-7.37 (m, 5H, aromatic), 7.40 (s, 1H, H-6). ^{13}C-NMR (100 MHz, CDCl$_3$) δ (ppm): 12.6 (i-Pr), 12.9 (CH$_3$), 13.2, 13.2, 13.4, 17.2, 17.2, 17.3, 17.3, 17.3 (i-Pr), 38.7 (O$_3$SCH$_3$), 60.9 (C-5'), 70.6 (CH$_2$), 72.2 (CH$_2$), 74.7 (C-3'), 77.2 (C-2'), 83.0 (C-4'), 87.2 (C-1'), 111.2 (C-5), 127.7, 127.8, 128.3 (Ar-C), 133.3 (C-6), 137.8 (Ar-C), 151.1 (C-2), 163.1 (C-4). HRMS (ESI): C$_{31}$H$_{50}$N$_2$O$_{10}$SSi$_2$; [M+Na]$^+$: 721.2646 (calc. 721.2622). ^{1}H-^{13}C HSQC, ^{1}H-^{13}C HMBC and ^{1}H-^{1}H COSY (Appendix D.1).
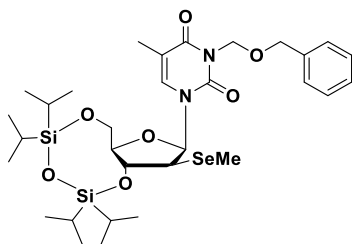
3',5'-O-(1,1,3,3-tetraisopropyldisiloxane-1,3-diyl)-N3-(benzyloxymethylacetal)-5-methyl-2'-methylseleno-arabinouridine (2.4) NaBH$_4$ (106 mg, 2.8 mmol) was placed in a pressure vessel and suspended in 5 mL anhydrous THF under argon. To the solution, Me$_2$Se$_2$ (0.26 mL, 2.8 mmol) was injected, followed by the drop wise addition of 1 mL anhydrous ethanol. The yellowish color suspension was stirred at room temperature for 30 min. Then, the starting material **2.3** (1 g, 1.4 mmol) was dissolved in 5 mL anhydrous THF and added into the solution. The reaction mixture was heated up to 95 °C and stirred for 16 h. The solution was concentrated under reduced pressure and re-dissolved in ethyl acetate (60 mL). The organic solution was washed with water (30 mL), brine (40 mL), dried over MgSO$_4$, and evaporated to dryness to afford crude product. The residue was then purified by silica gel chromatography (10 % ethyl acetate in hexane)

to give desired product **2.4** (0.8 g, 80 %) as white solid. ^{1}H-NMR (400 MHz, CDCl$_3$) δ (ppm): 1.04-1.13 (m, 28H, 4×iPrSi), 1.95 (d, J=1.0 Hz, 3H, CH$_3$), 2.06 (s, 3H, SeCH$_3$), 3.63 (dd, J_1=7.4 Hz, J_2=9.8 Hz, 1H, H-2'), 3.73 (td, J_1=2.4 Hz, J_2=8.1 Hz, 1H, H-4'), 4.05 (dd, J_1=2.9 Hz, J_2=13.2 Hz, 1H, H-5'), 4.12 (dd, J_1=2.0 Hz, J_2=13.2 Hz, 1H, H-5'), 4.23 (dd, J_1=8.5 Hz, J_2=9.5 Hz, 1H, H-3'), 4.68 (d, J=1.7 Hz, 2H, CH$_2$OC<u>H$_2$</u>Ph), 5.53 (s, 2H, C<u>H$_2$</u>OCH$_2$Ph), 6.41 (d, J=7.3 Hz, 1H, H-1'), 7.38-7.26 (m, 6H, aromatic, H-6). ^{13}C-NMR (100 MHz, CDCl$_3$) δ (ppm): 6.6 (SeCH$_3$), 12.5, 12.7, 13.0 (i-Pr), 13.3 (CH$_3$), 13.9, 17.0, 17.1, 17.3, 17.4 (i-Pr), 50.2 (C-2'), 60.0 (C-5'), 70.4 (CH2), 71.9 (CH2), 73.6 (C-3'), 83.5 (C-4'), 85.1 (C-1'), 109.9 (C-5), 127.6, 127.7, 128.3 (Ar-C), 134.3 (C-6), 138.0 (Ar-C), 151.3 (C-2), 163.4 (C-4). HRMS (ESI): C$_{31}$H$_{50}$N$_2$O$_7$SeSi$_2$; [M+Na]$^+$: 721.2209 (calc. 721.2219). ^{1}H-^{13}C HSQC, ^{1}H-^{13}C HMBC and ^{1}H-^{1}H COSY (Appendix D.1).
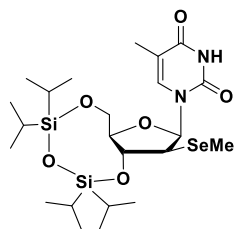
3′,5′-O-(1,1,3,3-tetraisopropyldisiloxane-1,3-diyl)-5-methyl-2'-methylseleno-arabinouridine (2.5) Compound **2.4** (1.8 g, 2.6 mmol) was placed in a round bottom flask, purge with argon and dissolved in 15 mL anhydrous toluene. The solution was cooled down to -78 ^{o}C by placing in a dry ice-acetone bath. To the solution, BBr$_3$ solution (0.31 mL, 3.1 mmol, 1 M in hexane) was injected and the reaction was stirring at -78 ^{o}C for 1 h. Then, the reaction was quenched by adding a mixture of triethylamine and anhydrous isopropanol (2 mL, 1:1, v/v), followed by the remove of the dry ice-acetone bath. The solution was stirred at room temperature for another 1 h. The reaction was concentrated to around 2 mL and diluted with 50 mL of ethyl acetate. The organic solution was washed with water (15 mL). The water layer was then extracted with ethyl acetate (3 x 10 mL), and the combined organic layer was washed with brine (40 mL),

dried over MgSO$_4$ (s), followed by filtration. The solvent was evaporated under reduced pressure to afford the crude product. The residue was then purified by silica gel chromatography (15 % ethyl acetate in hexane) to give desired product **2.5** (1.1 g, 71 %). ^{1}H-NMR (400 MHz, CDCl$_3$) δ (ppm): 1.04-1.12 (m, 28H, 4×iPrSi), 1.93 (d, 3H, CH$_3$), 2.09 (s, 3H, SeCH$_3$), 3.62 (m, 1H, H-2'), 3.72 (d, *J*=7.4 Hz, 1H, H-4'), 4.05 (d, *J*=13.2 Hz, 1H, H-5'), 4.13 (d, *J*=14.2 Hz, 1H, H-5'), 4.24 (t, *J*=8.9 Hz, 1H, H-3'), 6.36 (d, *J*=7.4 Hz, 1H, H-1'), 7.25 (s, 1H, H-6), 8.34 (br, 1H, NH). ^{13}C-NMR (100 MHz, CDCl$_3$) δ (ppm): 6.5 (SeCH$_3$), 12.5 (i-Pr), 12.6 (CH$_3$), 12.7, 13.0, 13.9, 17.0, 17.1, 17.3, 17.4 (i-Pr), 49.9 (C-2'), 60.0 (C-5'), 73.7 (C-3'), 83.5 (C-4'), 84.5 (C-1'), 110.5 (C-5), 135.6 (C-6), 150.3 (C-2), 163.4 (C-4). HRMS (ESI): C$_{23}$H$_{42}$N$_2$O$_6$SeSi$_2$; [M+Na]$^+$: 601.1631 (calc. 601.1644).
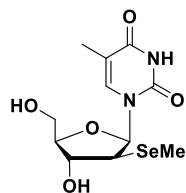
5-methyl-2'-methylseleno-arabinouridine (2.6) Compound **2.5** (1 g, 1.8 mmol) was dissolved in 10 mL anhydrous THF, and treated with 3HF·Et$_3$N (0.29 mL, 1.8 mmol) at 40 °C for 2 h. The mixture was evaporated to dryness, and the residue was purified by flash column chromatography on silica gel (5% MeOH in methylene chloride) to afford pure product **2.6** (597 mg, 99%) as white solid. ^{1}H-NMR (400 MHz, DMSO) δ (ppm): 1.77 (d, *J*=1.0 Hz, 3H, CH$_3$), 1.99 (s, 3H, SeCH$_3$), 3.61-3.65 (m, 3H, H-2', H-4', H-5'), 3.74 (dd, *J$_1$*=5.0 Hz, *J$_2$*=9.5 Hz, 1H, H-5'), 4.11 (dd, *J$_1$*=6.6 Hz, *J$_2$*=14.8 Hz, 1H, H-3'), 5.15 (t, *J*=5.0 Hz, 1H, 5'-OH), 5.67 (d, *J*=5.9 Hz, 1H, 3'-OH), 6.23 (d, *J*=7.2 Hz, 1H, H-1'), 7.67 (d, *J*=0.7 Hz, 1H, H-6), 11.30 (s, 1H, NH). ^{13}C-NMR (100 MHz, DMSO) δ (ppm): 5.2 (SeCH$_3$), 12.7 (CH$_3$), 49.9 (C-2'), 59.7 (C-5'), 73.9 (C-3'), 84.8 (C-1'), 85.1 (C-4'), 108.9 (C-5), 137.1 (C-6), 150.9 (C-2), 164.1 (C-4). HRMS (ESI):

$C_{11}H_{16}N_2O_5Se$; [M+Na]$^+$: 359.0018 (calc. 359.0122). ^{1}H-^{13}C HSQC, ^{1}H-^{13}C HMBC and ^{1}H-^{1}H COSY (Appendix D.1).
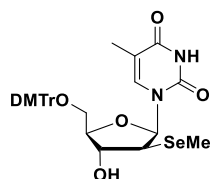
5'-O-(4,4'-dimethoxytrityl)-5-methyl-2'-methylseleno-arabinouridine **(2.7)**

Compound **2.6** (500 mg, 1.5 mmol) was dried over high vacuum and co-evaporated with anhydrous pyridine (2 x 10 mL). Then, the starting material was dissolved in 15 mL anhydrous pyridine and cooled by ice bath, followed by the treatment of dimethoxytrityl chloride (DMTrCl) (542 mg, 1.6 mmol) under argon. The reaction was warmed up to room temperature and stirred for 2 h. Then the solvent was removed under vacuum, the residue was dissolved in 30 mL methylene chloride. The organic solution was washed with water (20 mL), brine (30 mL), dried over MgSO$_4$ (s), followed by filtration and evaporation. The crude product was purified by silica gel chromatography (3% MeOH in methylene chloride with 1% Et$_3$N) to afford the desired product **2.7** (851 mg, 89%) as white solid. ^{1}H-NMR (400 MHz, CDCl$_3$) δ (ppm): 1.61 (s, 3H, 5-CH$_3$), 2.09 (s, 3H, SeCH$_3$), 2.66 (d, J=4.1 Hz, 1H, 3'-OH), 3.48 (dd, J_1=3.8 Hz, J_2=10.7 Hz, 1H, H-5'), 3.57 (dd, J_1=3.3 Hz, J_2=10.8 Hz, 1H, H-5'), 3.61 (dd, J_1=7.3 Hz, J_2=8.3 Hz, 1H, H-2'), 3.81 (s, 6H, OCH$_3$), 3.88 (dt, J_1=3.6 Hz, J_2=6.1 Hz, 1H, H-4'), 4.30 (td, J_1=4.0 Hz, J_2=7.9 Hz, 1H, H-3'), 6.41 (d, J=7.2 Hz, 1H, H-1'), 6.85-6.87 (m, 4H, aromatic), 7.25-7.46 (m, 10H, aromatic, H-6), 8.75 (br, 1H, NH). ^{13}C-NMR (100 MHz, CDCl$_3$) δ (ppm): 6.3 (SeCH$_3$), 12.1 (5-CH$_3$), 50.4 (C-2'), 55.3 (OCH$_3$), 61.8 (C-5'), 75.4 (C-3'), 82.5 (C-4'), 85.2 (Ar-C), 86.8 (C-1'), 110.4 (C-5), 113.3, 127.1, 128.0, 128.2, 130.1, 135.4, 135.5 (Ar-C), 135.8 (C-6), 144.4 (Ar-C), 150.4 (C-2), 158.7 (Ar-C), 163.7 (C-4). HRMS (ESI): $C_{32}H_{34}N_2O_7Se$; [M+Na]$^+$: 661.1431 (calc. 661.1429).
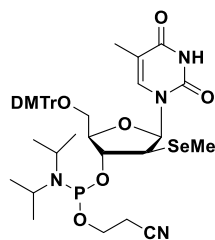
3'-O-(2-cyanoethyl-N,N-diisopropylphosphoramidite)-5'-O-(4,4'-dimethoxytrityl)-5-methyl-2'-methylseleno-arabinouridine (2.8) The starting material **2.7** (500 mg, 0.8 mmol) was dried under high vacuum and dissolved in 5 mL anhydrous CH_2Cl_2. To the solution, dimethylethylamine (0.5 mL, 4.8 mmol) was injected, followed by the addition of the 2-cyanoehtyl *N,N*-diisopropylchlorophosphoramidite (0.2 mL, 0.8 mmol). The reaction mixture was stirred at room temperature under dry argon for 1 h. Then, the solvent was evaporated under reduced pressure. The residue was re-dissolved in 2 mL CH_2Cl_2 and precipitated in 400 mL hexane under vigorous stirring. The hexane solution was then decanted carefully, and the crude product was further purified by silica gel chromatography (30% ethyl acetate in CH_2Cl_2 containing 1% dimethylethylamine). The product was precipitated in hexane (400 mL) again, and the precipitate was re-dissolved in CH_2Cl_2 and evaporated to afford the desired pure product **2.8** (542 mg, 81%) as a white foam. HRMS (ESI): $C_{41}H_{51}N_4O_8PSe$; $[M+Na]^+$: 861.2525 (calc. 861.2507).

2.4.3   2'-Se-functionalized oligonucleotides synthesis and HPLC purification

All DNA oligonulceotides were synthesized by solid-phase synthesis (Figure 2.15) using an ABI3400 DNA/RNA synthesizer on a 1 μmol scale. 2'-MeSe-5-methyl-arabinouridine phosphoramidite was dried under high vacuum for 3 h and prepared as 0.1 M concentration solution in anhydrous acetonitrile prior of use. Normal phosphoramidites purchasing from Glenresearch were used for the synthesis of natural nucleosides. The coupling reaction was carried out using a tetrazole solution (0.45 M) in acetonitrile with a coupling time of 300 seconds. All the oligonucleotides were prepared in DMTr-On mode, followed by the cleavage from CPG solid

support (beads) and the deprotection with concentrated ammonia solution under 55 °C overnight. The volatile ammonias were evaporated on a speed-vac concentrator, and the remaining beads were extracted with water (3 x 0.5 mL). The DNA oligonucleotides were analyzed and purified by reversed-phase high performance liquid chromatography (RP-HPLC), flow rate 6 ml/min [buffer A: 20mM triethylammonium acetate (TEAAc, pH 7.1) in water; buffer B: 20mM TEAAc (pH 7.1) in 50% acetonitrile]. The HPLC analysis was performed with a linear gradient from buffer A to 100% buffer B in 20 min. The detritylation was performed by treatment of the oligonucleotide with an acetic acid solution (pH 4.5) under 40 °C for 1 h. The DMTr-off oligonucleotides were desalted by a Water C-18 column and purified again by RP-HPLC. Native RNAs were purchased from Integrated DNA Technologies. The native, as well as the Se modified DNAs were characterized by matrix assisted laser desorption/ionization-time of fight mass spectrometry (MALDI-TOF MS).

Figure 2.15 Steps of DNA solid-phase synthesis

2.4.4   UV-melting and CD experiments

The thermal denaturing experiments were performed on Cary 300 UV-Vis spectrometer equipped with a six-sample cell changer and a temperature controller. the sample annealed by heated to 85 ^{o}C for 2 min and slowly cooling down to 4 ^{o}C before data acquisition. Both denaturing and annealing curves were acquired at 260 nm with 1 cm path length at heating or cooling rates of 0.5 ^{o}C/min and data interval of 0.5 ^{o}C. The samples (2 µM DNA duplexes) were dissolved in the buffer of 350 mM NaCl, 10 mM NaH_2PO_4-Na_2HPO_4 (pH 6.8), 0.1 mM EDTA.

The circular dichroism (CD) experiments were performed on Jasco J-1500 spectrometer. The samples (3 µM DNA duplexes) were dissolved in the buffer of 1 mM NaCl, 1 mM NaH_2PO_4-Na_2HPO_4 (pH=7.0), 0.1 mM EDTA.

2.4.5   DNA and DNA/protein complex Crystallization

For the DNA crystallization by itself, the purified DNA oligonucleotides were heated to 80 ^{o}C for 2 min, then cooled down slowly to room temperature. The Nucleic Acid Mini Screen Kit (Hampton Research) was applied to screen the crystallization conditions room temperature using the hanging-drop method by vapor diffusion. The concentrations of the native, Se modified DNAs were adjusted to 1.0 mM in water for crystallization. During the crystallization screening process, the DNA solution was mixed with crystallization buffer at 1:1 ratio (0.5 µL+ 0.5 µL) and equilibrate against 500 µL of the 35% MPD (2-methyl-2,4-pentanediol) solution within a hanging-drop plate.

For the DNA/Bst DNA polymerase complex crystallization, the purified DNA oligonucleotides (1 mM) were annealed by first heating to 80°C for 2 min, and then slowly cool down to room temperature. The resulting DNA solution was mixed with the protein (final concentration: 24 mg/mL) at 1:1 molar ratio. Co-crystallization of Se-DNA/Bst DNA polymerase

was achieved by screening with the QIAGEN Classics Suite Kit (www.qiagen.com) using the sitting-drop vapor diffusion method at 25°C, and the mixture #67 was found to be the best buffer condition for the crystal growth. [Salt: 0.2 M Calcium acetate; Buffer: 0.1 M sodium cacodylate, pH 6.5; precipitant: 18% (w/v), PEG 8000].

2.4.6   *Diffraction data collection and structure refinement*

Diffraction data were collected on beamline 8.2.1 of the ALS (Advanced Light Source) at the Lawrence Berkeley National Laboratory. X-ray data were collected under a liquid-nitrogen stream at 99 K. Each crystal was exposed for 0.5 second per image with 1° rotation and a total of 360 images were obtained. The data were integrated and scaled with the programs HKL2000 and DENZO/SCALEPACK.[81] The structures were solved by molecular replacement method using Phaser[82] within CCP4i[83, 84]. The resulting models were refined using REFMAC5.5[85] within CCP4i[83, 84]. The DNA, protein, ions and water molecules were added and modeled using Coot[86].

3   2'-MeSe-ARABINOCYTIDINE MODIFICATION FACILITATES B-DNA CRYSTAL GROWTH AND X-RAY STRUCTURE STUDY

3.1   Introduction

3.1.1   The difficulties of B-form DNA crystallography

It is important to study the structure of nucleic acids which reveals the mechanisms of diseases, viral infections as well as cancers, and provides information for treatment and drug design. X-ray crystallography is one of the most powerful and successful method to determine the 3D-structure. To get high-resolution data for X-ray diffraction and gain insight into the structure details, large, well-ordered, single crystal is required. As mentioned in the general introduction section, it is more difficult to form crystal for nucleic acids than for proteins due to the negatively charged repetitive phosphate backbone repelling the crystal packing as well as the conformational heterogeneity. It is even more challenging to get a B-DNA crystal due to the high salt condition during the crystallization process which disfavoring the B-DNA duplex formation.

Besides the challenges in the crystallization, the phasing problem will need to be solved before determining the 3D structure through X-ray crystallography. The selenium-atom-derivatized nucleic acid (SeNA) method has been pioneered and developed to solve this problem[25]. The selenium atom can be used as an anomalous scattering center for phase determination via either multiple or single wavelength anomalous diffraction (MAD or SAD), which become a powerful method to solve the phasing problem rationally.

3.1.2   Selenium modifications on cytidine

To explore the selenium modification on nucleic acids for structural and functional studies, different modified cytidine and its DNA/RNA oligonucleotides have been synthesized. 5-methylselenyl-cytidine and its corresponding DNA, mimicking the 5-methylcytosine (m^5C), were

synthesized and determined.[87] Virtually identical structure were observed for the Se-DNA and the corresponding native, which is consistent with the thermostability studies, indicating the well accommodation of the selenium functionality without perturbing the DNA duplex.

The synthesis of 2'-α-MeSe-C modified DNA and RNA as well as the structure study have also been reported[88-90]. To introduce the selenium functionality to the 2'-α-position, an indirect method via conversion of the corresponding uridine derivative **4** was used due to the low yield giving from the direct incorporation (Figure 3.1). The structure of modified DNAs and RNAs were determined through X-ray diffraction with MAD phasing. The modified nucleotides exhibit same 3'-endo sugar pucker and had no effect on the stability of the duplexes in the UV melting study which demonstrating that the 2'-selenium functionality is suitable for RNA and A-DNA derivatization for X-ray crystallography. Further study using the T4 RNA ligase successfully ligate the selenium-containing RNA shows the high flexibility of the selenium approach (Figure 3.2).[89]
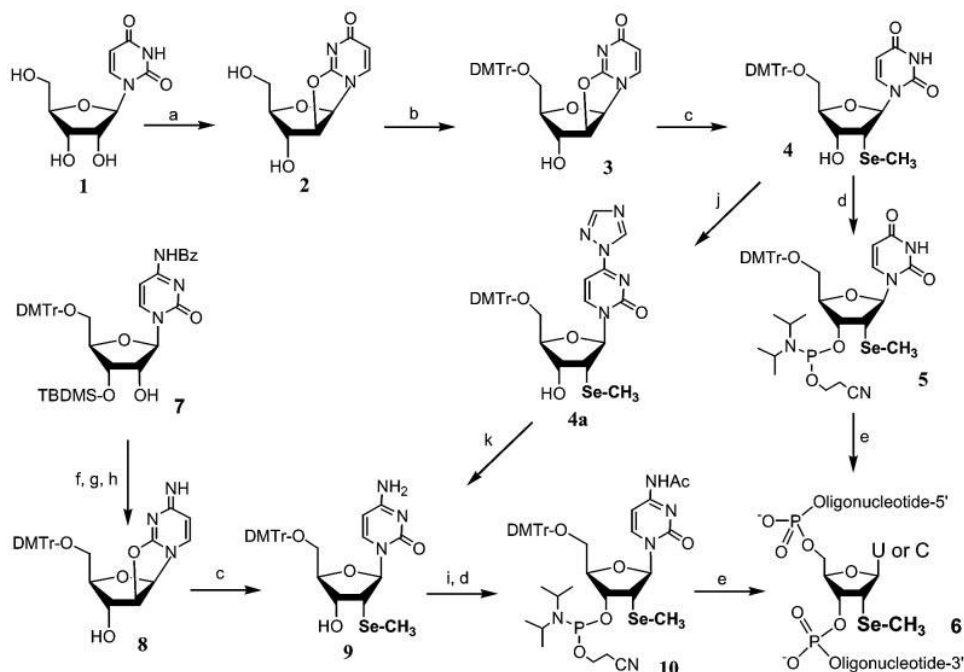

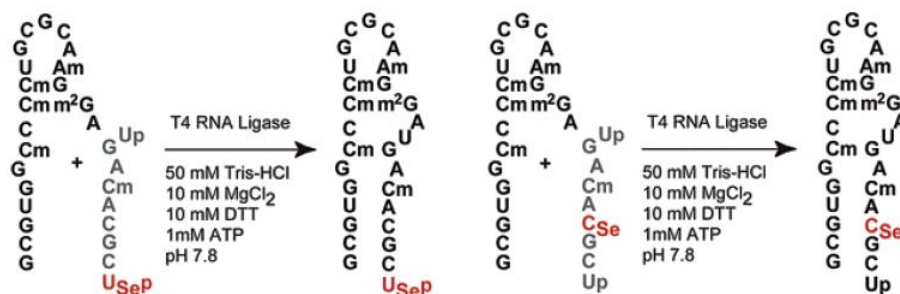*Figure 3.1 Synthesis of 2'-α-MeSe-C/U modified DNA and RNA*

Figure 3.2 Enzymatic ligation of 2'-Se-methyluridine and 2'-Se-methylcytidine derivatized oligoribonucleotides with T4 RNA ligase

3.2   Results and discussion

3.2.1   *Synthesis of 2'βSe-C phosphoramidite and oligonucleotides*

In the very beginning, we try to directly synthesize the 2'-β-MeSe-modified cytidine through activating the 2'-OH group and substituted with MeSe via a S_N2 reaction. However, the synthesis was unsuccessful with different protective and leaving groups (Scheme 3.1). A computational study reveals that the selenium incorporation step has a much higher energy barrier (3.2 kcal/mol) than the corresponding uridine derivative with same leaving group (Appendix A). Therefore, to introduce the selenium functionality will need either extremely high temperature which requires a highly stable base protective group giving deprotection problem later, or an extremely active leaving group which may result in various side-reactions.

R_1=Bz, R_2=H, LG=Ms, Ts, R_3=R_4=TIPDS
R_1=Boc, R_2=H, LG=Ms, Ts, Tf, R_3=R_4=TIPDS
R_1=Boc, R_2=Boc, LG=Ms, Ts, Ns, Tf, R_3=R_4=TIPDS
R_1=Ac, R_2=H, LG=Ms, Ts, Ns, R_3=R_4=TIPDS
R_1=DMTr, R_2=H, LG=Ms, R_3=R_4=TIPDS
R_1=H, R_2=H, LG=Ms, Tf, R_3=R_4=TIPDS
R_1=Ac, R_2=H, LG=Ms, R_3=R_4=H
R_1=DMTr, R_2=H, LG=Ms, R_3=R_4=H
R_1=Ac, R_2=H, LG=Ms, R_3=DMTr, R_4=H
R_1=H, R_2=H, LG=Ms, R_3=DMTr, R_4=H
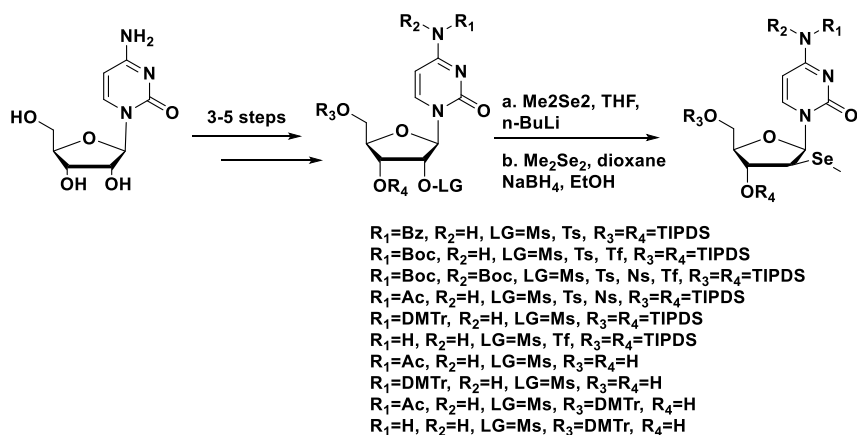
Scheme 3.1 Direct incorporation of selenium functionality into cytidine

Based on the computational results, we designed an indirect method via the synthesis of a 2'β-MeSe-modified uridine derivative and conversion of the corresponding uridine to cytidine (Scheme 3.2). The 5'.3'-OH groups were protected with tetraisopropyldisilylene (TIPDS), followed by the protection of 3-N with benzyloxymethyl acetal (BOM). The 2'-OH group was activated with mesylate, then treated with sodium methylselenolate generating from the reduction of dimethyldiselenide by sodium borohydride in THF. After the selenium incorporation, the BOM group was deprotected through the treatment of BBr₃ in hexane, followed by quenching with a mixture of anhydrous isopropanol and triethylamine. To convert the 2'βMeSe-uridine to cytidine, the 4-O was activated with phosphoryl chloride and triazole. The activated intermediate was then treated with ammonia given the 2'βMeSe-cytidine derivative **C7**. After acetylation of the base, the TIPDS group was remove by treated with triethylamine trihydrofloride at elevated temperature. The 5'-OH group was then protected with trityl and converted to 2'-Se-cytidine phosphoramidite by reacting with 2-cyanoethyl N,N-diisopropylchlorophosphoramidite in the presence of dimethylamine in dry CH₂Cl₂.
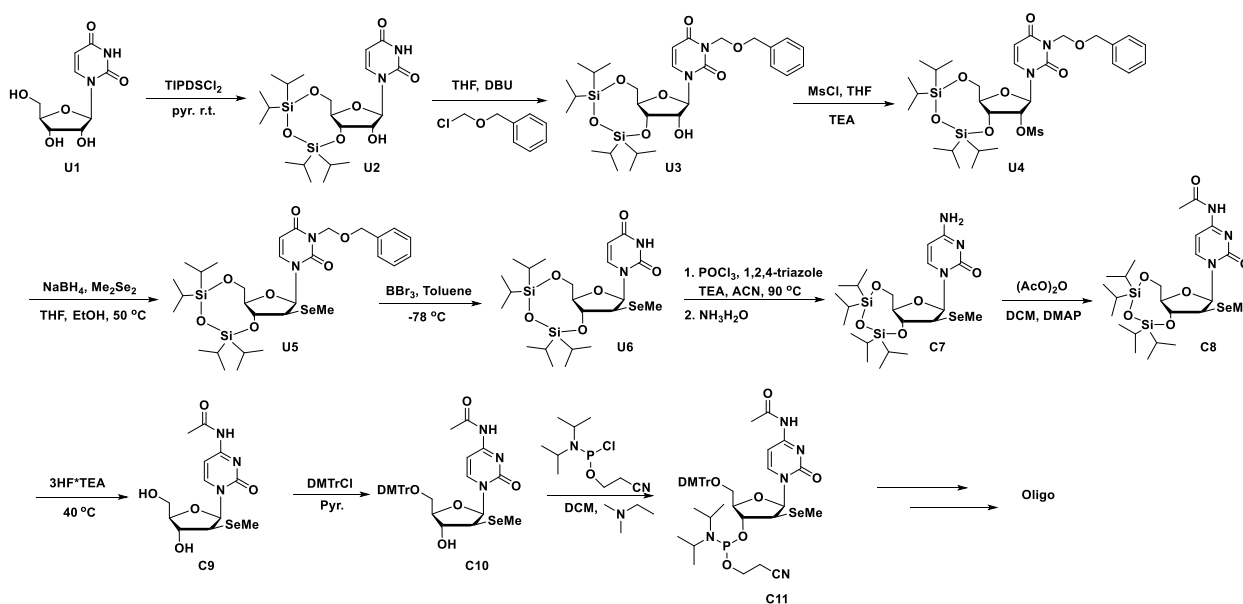
Scheme 3.2 Synthesis of 2'-β-MeSe-cytidine phosphoramidite and DNA oligonucleotides.

Several DNA oligonucleotides were synthesized via the solid-phase synthesis showing the compatibility of 2'βSe-C phosphoramidite. The 2'-MeSe group was stable under mild I_2 treatment for the phosphite oxidation and trichloroacetic acid treatment for detritylation. Moreover, coupling yields of the DNA solid-phase synthesis using this novel 2'βSe-C phosphoramidite were higher than 95%. The result DNA oligonucleotides were analyzed and purified with reversed-phase HPLC and characterized by MALDI-TOF (Table 3.1).

Table 3.1 MALDI-TOF analytical data of native and [Se]C-DNA oligonucleotides

Entry	Se-oligonucleotides	Measured (calcd) m/z
1	C-s1 5'-[Se]CGCGAATTCGCG-3' $C_{117}H_{149}N_{46}O_{70}P_{11}Se$ FW 3739.4	[M]+: 3739.7 (3739.4)
2	C-s2 5'-CG[Se]CGAATTCGCG-3' $C_{117}H_{149}N_{46}O_{70}P_{11}Se$ FW 3739.4	[M]+: 3739.6(3739.4)
3	C-s3 5'-CGCGAATT[Se]CGCG-3' $C_{117}H_{149}N_{46}O_{70}P_{11}Se$ FW 3739.4	[M]+: 3739.6 (3739.4)
4	C-s4 5'-CGCGAATTCG[Se]CG-3' $C_{117}H_{149}N_{46}O_{70}P_{11}Se$ FW 3739.4	[M]+: 3739.7 (3739.4)
5	Native 5'-CGCGAATTCGCG-3' $C_{115}H_{145}N_{46}O_{70}P_{11}$ FW 3646.4	[M]+: 3645.9 (3646.4)

3.2.2 Thermostability and circular dichroism comparison of DNAs containing 2'βSe-Cytidine

The UV melting temperatures of the [Se]C-DNAs were measured to examine the impact of the modification on the thermostability of the duplexes. The melting curves are shown in Figure 3.3. The melting temperatures indicate that the position of [Se]C is directly relate to the stability of the DNA duplex. With the shift of the modification from terminal into the middle of the sequence, the destabilization effect is getting stronger from 4.4 ºC to 37 ºC compared to the corresponding native DNA (Table 3.2), which reveals that the [Se]C modification may interrupt the base pairing or

packing of the DNA. Modification at the terminal cytidine give the smallest effect to the stability. CD measurements were also conducted at room temperature, and the results suggest that the modified DNAs still exhibit similar B-form conformation in the buffer solution as the native (Figure 3.4). However, a significant increase in intensity at 280 nm indicates there is some difference in the conformation comparing with the native DNA.
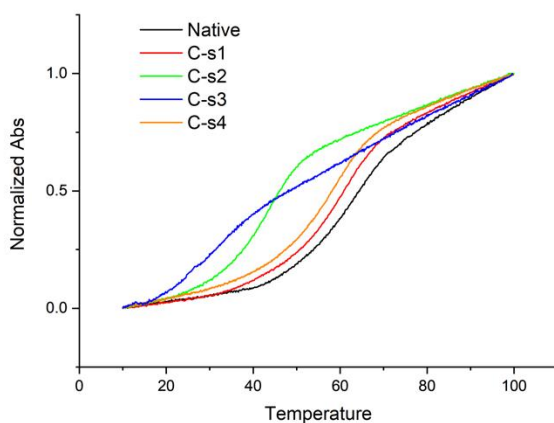

Figure 3.3 Normalized UV-melting curves of the native and ^{Se}C-DNAs

Table 3.2 UV-melting temperature of the native and ^{Se}C-DNAs

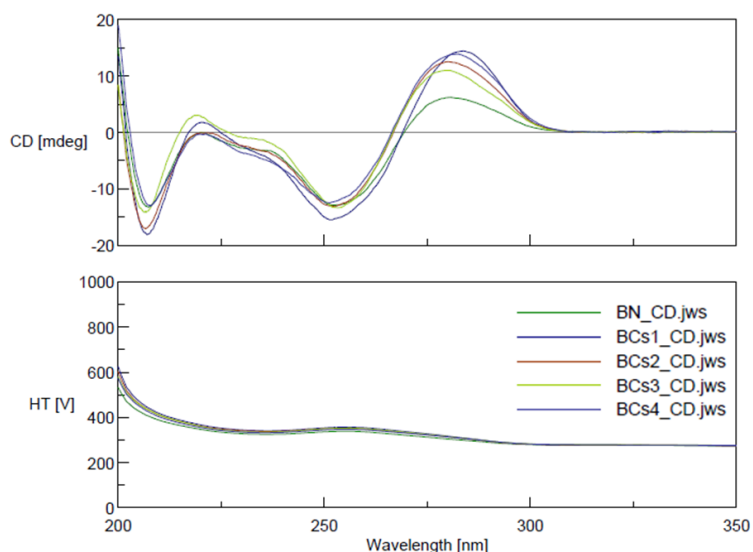| Entry	Sequence	Melting temperature (°C)	ΔT_m(°C)
native	5'-CGCGAATTCGCG-3'	64.4	-
C-s1	5'-^{Se}CGCGAATTCGCG-3'	60.0	4.4
C-s2	5'-CG^{Se}CGAATTCGCG-3'	43.4	21
C-s3	5'-CGCGAATT^{Se}CGCG-3'	27.4	37
C-s4	5'-CGCGAATTCG^{Se}CG-3'	58.9	5.5

Figure 3.4 CD spectra comparison of ^{Se}C-DNAs with the corresponding native

3.2.3 B-DNA and DNA/BST complex crystallization studies

The crystallization of ^{Se}C-DNAs was examined to demonstrate the ability of Se-modification strategy to facilitate crystal growth (Figure 3.5). For the crystallization of DNA by itself, only the sequence with the ^{Se}C at the terminal of the sequence (C-s1) form crystals, which is consistent with our UV-melting study since it has the smallest destabilization effect. However, to our surprise, this modification greatly speeds up the crystal growth (Figure X.). The first crystal formed within 3 hours. It only took 20 hours to form crystals in 20 different buffer conditions, comparing to the native DNA which needs at least two weeks.

Moreover, the ^{Se}C-DNAs also co-crystallized with a fragment of DNA polymerase I from Bacillus stearotherophilus. However, for the ^{Se}C-DNA/BST complex crystallization, opposite results were observed. Only the oligonucleotides with the modification at the terminal cytidine (C-s1) do not form crystal with BST polymerase. The C-s2 forms the best crystal with BST protein compared with other modified oligonucleotides (Figure 3.6).
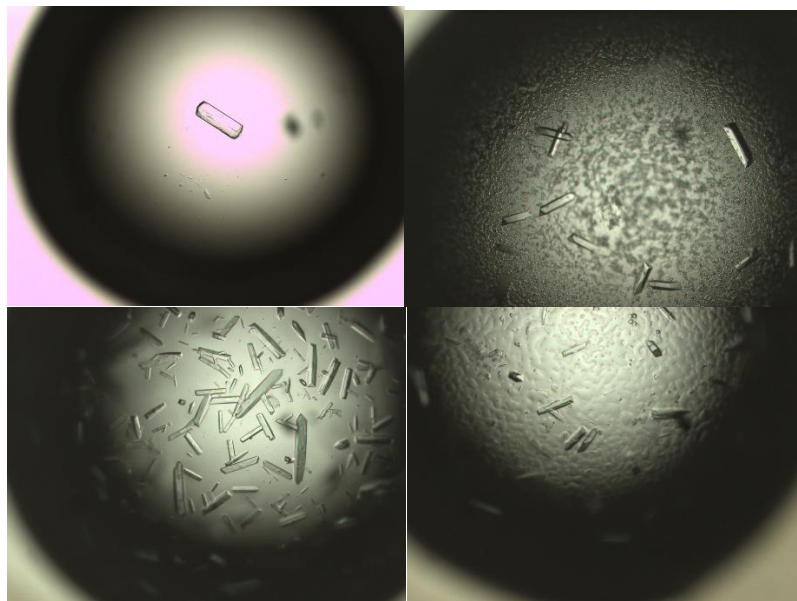
Figure 3.5 Pictures of crystals with ^{Se}C modification
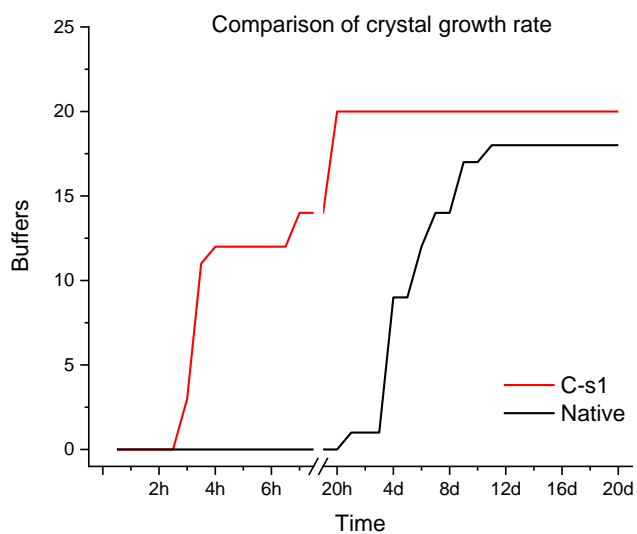

Figure 3.6 The rate comparison of crystal growth

3.2.4 Structure of ^{Se}C-DNA/Bacillus fragment complex

The X-ray diffraction was also performed for crystals of ^{Se}C-DNAs. However, poor resolution was obtained for all ^{Se}C-DNAs. The structure determine with ^{Se}T modification (5'-CGCGAATTCGCG-3') may provide an explanation (Fig. 3.7). The terminal cytidine was

surrounded by another two DNA molecules which could be affected by the 2'MeSe-modification
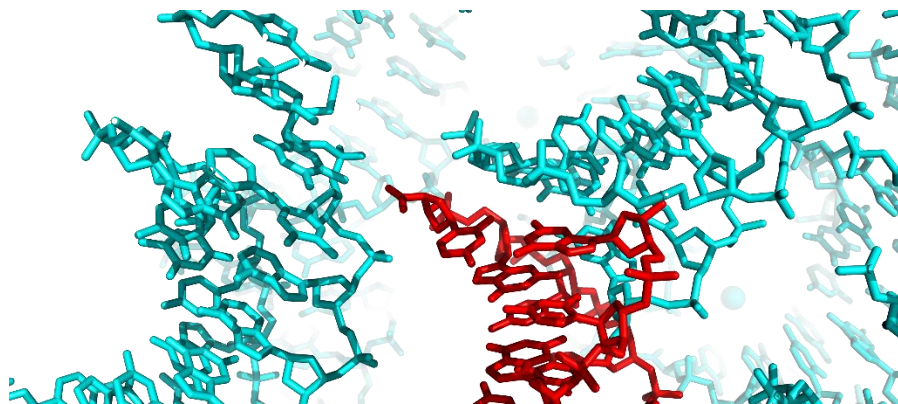
on the cytidine and given an irregular packing.

Figure 3.7 Packing of 2'SeT-modified DNAs (5'-CGCGAATTCGCG-3')

The crystal structure of fragment DNA polymerase I from Bacillus stearotherophilus with

DNA modified with methylselenium group at the 2'-arabino position of cytidine was determined

at 2.00 Å resolution. The modified DNA accommodated very well in the DNA/protein complex

structure which is virtually identical to the corresponding native complex with a RMSD of 0.176Å

for all 4387 atoms (Figure 3.8). The 2'-MeSe functionalities can be undoubtedly recognized from

the electron density map. The data collection and structure refinement statistics are summarized in

Table 3.3.

Figure 3.8 Superimposed structure of the ^{Se}C-DNA/BST complex with the native complex

Table 3.3 Crystallographic data collection and refinement statistics of ^{Se}C-DNA/BST complex

| DNA #	Cs2/BST
Wavelength (Å)	1
Resolution range (Å)	50.0-2.00 (2.03-2.00)
Space group	$P2_12_12_1$
Unit-cell a, b, c (Å)	86.3, 94.2, 105.9
α, β, γ (°)	90, 90, 90
Unique reflections	58840(2886)
Completeness (%)	100.0(100.0)
Rmerge (%)	6.7 (46.6)
I/σ(I)	24.1 (4.2)
R value (%)	18.5
Rfree (%)	21. 2
Average B value (Å^2)	27.6
R.m.s.d. bond length (Å)	0.012
R.m.s.d. bond angle (°)	1.72

3.3   Conclusion

In conclusion, we successfully synthesized the 2'-β-MeSe-Cytidine phosphoramidite and its DNAs through the conversion of the modified uridine to cytidine. We also demonstrate the solid-phase synthesis, deprotection, and purification of 2'-β-MeSe-Cytidine modified oligonucleotides. In the thermostability study, we found that the position of the ^{Se}C modification is crucial to the stability of the DNA duplex, which also affect the ability of the DNAs to form crystal. Interestingly, the terminal modification greatly speeds up the B-DNA crystallization with broader crystallization conditions compared to the native. The modified DNAs were also co-crystallized with a fragment of DNA polymerase I from Bacillus stearotherophilus. The modification site was well accommodated in the DNA/Bst protein complex structure without significantly perturb the global structure, which suggests its excellent compatibility for protein/DNA complex structure study. This modification could be further exploited in structure determination of B-DNA, quadruplex, i-motif as well as DNA/drug complexes.

3.4   Experimental section

3.4.1   General

Most solvents and reagents were purchased from Sigma, Fluka, or Aldrich (p.a.) and used without purification unless mentioned otherwise. Triethylamine (TEA) was dried over KOH (s) and distilled under argon. When necessary, solid reagents were dried under high vacuum. Reactions with compounds sensitive to air or moisture were performed under argon. Solvent mixtures are indicated as volume/volume ratios. Thin layer chromatography (TLC) was run on Merck 60 F254 plates (0.25 mm thick; Rf values in the text are for the title products), and visualized under UV-light or by a Ce-Mo staining solution (phosphomolybdate, 25 g; $Ce(SO_4)_2 \cdot 4H_2O$, 10 g; H_2SO_4, 60 mL, conc.; H_2O, 940 mL) with heating. Flash chromatography was performed using Fluka silica gel 60 (mesh size 0.040-0.063 mm) using a silica gel:crude compound weight ratio of ca. 30:1. ^1H, ^{13}C and 2D-NMR spectra were recorded using Bruker 400 (400 MHz). All chemical shifts (δ) are in ppm relative to tetramethylsilane and all coupling constants (J) are in Hz. High resolution (HR) MS were either obtained with electrospray ionization (ESI) on a Q-TOFTM Waters Micromass at Georgia State University.

3.4.2   Synthesis protocol and characterization
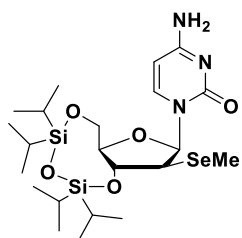
3′,5′-O-(1,1,3,3-tetraisopropyldisiloxane-1,3-diyl)-2'-methylseleno-arabinocytidine

(3.1) Phosphorus oxychloride (1.2 mL, 13.2 mmol) was added to a solution of 1,2,4-triazole (3.6 g, 52.8 mmol) in dry acetonitrile (30 mL) under argon. After stirring for 1 hr. at room temperature, dry triethylamine (15 mL, 105.6 mmol) was injected. The reaction was stirred another hour, then

the suspension was filtered directly into a pressure vessel containing compound **2.5** (2.5 g, 4.4 mmol) in dry acetonitrile (10 mL) under argon. Then the reaction was heated up to 90 °C and stirred at that temperature overnight. After the reaction was complete, the solution was cooled to 50 °C and treated with concentrated $NH_3 \cdot H_2O$(10 mL, 15 M). After stirring at 50 °C for 2 hr, the reaction mixture was evaporated to approximate 15 mL before extraction with ethyl acetate (3 x 100 mL). The combined organic layer was washed with brine (100 mL) and dried over anhydrous $MgSO_4$ (s) before evaporation. The crude product was purified by silica gel chromatography (2 % methanol in dichloromethane) to give desired product **3.1** (1.9 g, 75 %). ^{1}H-NMR (400 MHz, CDCl₃) δ (ppm): 1.01-1.11 (m, 28H, 4×iPrSi), 2.03 (s, 3H, SeCH₃), 3.70 (dd, J_1=6.8 Hz, J_2=8.1 Hz, 1H, H-2'), 3.76 (dt, J_1=7.2 Hz, J_2=3.3 Hz, 1H, H-4'), 4.05 (m, 2H, H-5'), 4.25 (t, J=7.7 Hz, 1H, H-3'), 5.74 (d, J=7.4 Hz, 1H, H-5), 6.44, (d, J=6.6 Hz, 1H, H-1'), 7.58 (d, J=7.4, 1H, H-6). ^{13}C-NMR (100 MHz, CDCl₃) δ (ppm): 6.2 (SeCH₃), 12.5, 12.9, 13.1, 13.8, 17.0, 17.1, 17.3, 17.4, 17.5 (i-Pr), 49.9 (C-2'), 61.2 (C-5'), 75.2 (C-3'), 83.6 (C-4'), 85.1 (C-1'), 94.1 (C-5), 141.5 (C-6), 156.0 (C-2), 165.6 (C-4). HRMS (ESI): $C_{22}H_{41}N_3O_5SeSi_2$; [M+Na][+]: 586.2769 (calc. 586.1648). ^{1}H-^{13}C HSQC, ^{1}H-^{13}C HMBC and ^{1}H-^{1}H COSY (Appendix D.2).
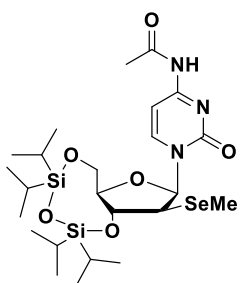
N⁴-acetyl-3′,5′-O-(1,1,3,3-tetraisopropyldisiloxane-1,3-diyl)-2'-methylseleno-arabinocytidine (3.2) Compound **3.1** (1.3 g, 2.3 mmol) was dissolved in 20 mL THF followed by the addition of triethylamine (2.6 mL, 18.4 mmol) in a pressure vessel. The mixture was then treated with N,N-dimethyl-aminopyridine (DMAP, 281 mg, 2.3 mmol) and acetic anhydride (1.3 mL, 13.8 mmol) under argon. The reaction was heated up to 85 °C and stirred at that temperature

overnight. After the reaction complete, MeOH (2 mL) was added to the mixture. The reaction was stirred for another 20 min to consume the excess acetic anhydride. The solvents were evaporated under reduced pressure and the residue was re-dissolved in ethyl acetate (100 mL). The organic solution was washed with water (30 mL), brine (30 mL) and dried over anhydrous $MgSO_4$ (s) before evaporation. The crude product was purified by silica gel chromatography (1 % methanol in dichloromethane) to give desired product **3.2** (1.3 g, 95 %). ^{1}H-NMR (400 MHz, $CDCl_3$) δ (ppm): 1.05-1.13 (m, 28H, 4×iPrSi), 2.07 (s, 3H, $SeCH_3$), 2.27 (s, 3H, CH_3), 3.73 (dd, J_1=6.7 Hz, J_2=8.7 Hz, 1H, H-2'), 3.76 (dt, J_1=7.5 Hz, J_2=3.0 Hz, 1H, H-4'), 4.09 (d, J=3.0, 2H, H-5'), 4.24 (t, J=8.1 Hz, 1H, H-3'), 6.47, (d, J=6.6 Hz, 1H, H-1'), 7.45 (d, J=7.5 Hz, 1H, H-5), 7.97 (d, J=7.5, 1H, H-6), 10.27 (br, 1H, NH). ^{13}C-NMR (100 MHz, $CDCl_3$) δ (ppm): 6.7 ($SeCH_3$), 12.5, 12.9, 13.1, 13.8, 17.0, 17.1, 17.3, 17.4 (i-Pr), 24.8 (CO$\underline{C}$H$_3$), 49.8 (C-2'), 60.8 (C-5'), 74.1 (C-3'), 83.9 (C-4'), 85.8 (C-1'), 96.2 (C-5), 145.0 (C-6), 154.9 (C-2), 162.7 (C-4), 171.0 ($\underline{C}$OCH$_3$). HRMS (ESI): $C_{24}H_{43}N_3O_6SeSi_2$; [M+H]$^+$: 606.1866 (calc. 605.1934). ^{1}H-^{13}C HSQC (Appendix D.2).
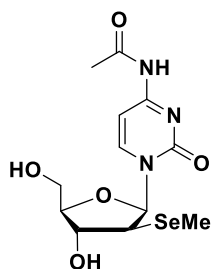
N^4-acetyl-3′,5′-O-(1,1,3,3-tetraisopropyldisiloxane-1,3-diyl)-2'-methylseleno-arabinocytidine (3.3) Compound **3.2** (1 g, 1.7 mmol) was dissolved in 10 mL anhydrous THF, and treated with 3HF·Et$_3$N (0.26 mL, 1.7 mmol) at 40 °C for 2 h. The mixture was evaporated to dryness, and the residue was purified by flash column chromatography on silica gel (5% MeOH in methylene chloride) to afford pure product **3.3** (609 mg, 99%) as white solid. ^{1}H-NMR (400 MHz, DMSO-d$_6$) δ (ppm): 1.88 (s, 3H, $SeCH_3$), 2.11 (s, 3H, CH_3), 3.59-3.75 (m, 4H, H-2', H-4', H-5'), 4.09 (m, 1H, H-3'), 5.07 (t, J=5.3, 1H, OH-5'), 5.74 (d, J=5.4, 1H, OH-3'), 6.29, (d, J=6.4

Hz, 1H, H-1'), 7.20 (d, *J*=7.4 Hz, 1H, H-5), 8.19 (d, *J*=7.5, 1H, H-6), 10.86 (s, 1H, NH). ^{13}C-NMR

(100 MHz DMSO-d$_6$) δ (ppm): 5.4 (SeCH$_3$), 24.8 (CO$\underline{C}$H$_3$), 50.1 (C-2'), 60.6 (C-5'), 75.0 (C-3'),

85.7 (C-4'), 86.4 (C-1'), 95.2 (C-5), 146.1 (C-6), 154.9 (C-2), 162.7 (C-4), 171.5 ($\underline{C}$OCH$_3$). HRMS

(ESI): C$_{12}$H$_{17}$N$_3$O$_5$Se; [M+Na]$^+$: 386.0223 (calc. 386.0231). ^{1}H-^{13}C HSQC, ^{1}H-^{13}C HMBC and

^{1}H-^{1}H COSY (Appendix D.2).
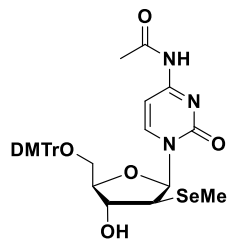
N^4-acetyl-5'-O-(4,4'-dimethoxytrityl)-2'-methylseleno-arabinocytidine (3.4)

Compound **3.3** (420 mg, 1.2 mmol) was dried over high vacuum and co-evaporated with anhydrous

pyridine (2 x 10 mL). Then, the starting material was dissolved in 15 mL anhydrous pyridine and

cooled by ice bath, followed by the treatment of dimethoxytrityl chloride (DMTrCl) (432 mg, 1.3

mmol) under argon. The reaction was warmed up to room temperature and stirred for 2 h. Then

the solvent was removed under vacuum, the residue was dissolved in 30 mL methylene chloride.

The organic solution was washed with water (20 mL), brine (30 mL), dried over MgSO$_4$ (s),

followed by filtration and evaporation. The crude product was purified by silica gel

chromatography (3% MeOH in methylene chloride with 1% Et$_3$N) to afford the desired product

3.4 (718 mg, 90%) as white solid. ^{1}H-NMR (400 MHz, CDCl$_3$) δ (ppm): 1.97 (s, 3H, SeCH$_3$), 2.25

(s, 3H, CH$_3$), 3.50 (dd, *J$_1$*=4.4 Hz, *J$_2$*=10.7 Hz, 1H, Ha-5'), 3.57 (dd, *J$_1$*=3.4 Hz, *J$_2$*=10.7 Hz, 1H,

Hb-5'), 3.83 (s, 6H, OCH$_3$), 3.94 (t, *J*=6.4 Hz, 1H, H-2'), 4.02-4.05 (m, 1H, H-4'), 4.28 (t, *J*=6.2,

1H, H-3'), 6.54, (d, *J*=6.2 Hz, 1H, H-1'), 6.88 (d, *J*=8.8, 4H, Ph), 7.27-7.36 (m, 8H, Ph, H-5), 7.45

(d, *J*=7.4 Hz, 2H, Ph), 8.16 (d, *J*=7.5, 1H, H-6), 9.32 (br, 1H, NH). ^{13}C-NMR (100 MHz, CDCl$_3$)

δ (ppm): 6.4 (Se$\underline{C}$H$_3$), 24.9 (CO$\underline{C}$H$_3$), 50.5 (C-2'), 53.4 (DMTr), 55.3 (O$\underline{C}$H$_3$), 62.1 (C-5'), 75.6

(C-3'), 83.6 (C-4'), 86.9 (C-1'), 96.2 (C-5), 113.3 (DMTr), 127.1, 128.0, 128.2, 130.1, 135.5, 144.4 (DMTr), 145.3 (C-6), 155.6 (C-2), 158.7 (DMTr), 162.4 (C-4), 170.4 ($\underline{C}$OCH$_3$). HRMS (ESI): C$_{33}$H$_{35}$N$_3$O$_7$Se; [M+Na]$^+$: 688.1511 (calc. 688.1538).
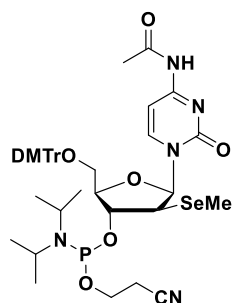
N^4-acetyl-3'-O-(2-cyanoethyl-N,N-diisopropylphosphoramidite)-5'-O-(4,4'-dimethoxytrityl)-2'-methylseleno-arabinocytidine (3.5) The starting material **3.4** (240 mg, 0.4 mmol) was dried under high vacuum and dissolved in 2 mL anhydrous CH$_2$Cl$_2$. To the solution, dimethylethylamine (0.2 mL, 2.4 mmol) was injected, followed by the addition of the 2-cyanoehtyl *N,N*-diisopropylchlorophosphoramidite (80 μL, 0.4 mmol). The reaction mixture was stirred at room temperature under dry argon for 1 h. Then, the solvent was evaporated under reduced pressure. The residue was re-dissolved in 2 mL CH$_2$Cl$_2$ and precipitated in 400 mL hexane under vigorous stirring. The hexane solution was then decanted carefully, and the crude product was further purified by silica gel chromatography (20-30% ethyl acetate in CH$_2$Cl$_2$ containing 1% dimethylethylamine). The product was precipitated in hexane (400 mL) again, and the precipitate was re-dissolved in CH$_2$Cl$_2$ and evaporated to afford the desired pure product **3.5** (260 mg, 75%) as a white foam. HRMS (ESI): C$_{42}$H$_{52}$N$_5$O$_8$PSe; [M+Na]$^+$: 887.8305 (calc. 887.8285).

3.4.3   *Solid-phase synthesis of SeC-DNA oligonucleotides*

All DNA oligonulceotides were synthesized by solid-phase synthesis using an ABI3400 DNA/RNA synthesizer on a 1 μmol scale. 2'-MeSe-arabinocytidine phosphoramidite was dried under high vacuum for 3 h and prepared as 0.1 M concentration solution in anhydrous acetonitrile

prior of use. Normal phosphoramidites purchasing from Glenresearch were used for the synthesis of natural nucleosides. The coupling reaction was carried out using a tetrazole solution (0.45 M) in acetonitrile with a coupling time of 300 seconds. All the oligonucleotides were prepared in DMTr-On mode, followed by the cleavage from CPG solid support (beads) and the deprotection with concentrated ammonia solution under 55 $^{\circ}$C overnight. The volatile ammonias were evaporated on a speed-vac concentrator, and the remaining beads were extracted with water (3 x 0.5 mL). The DNA oligonucleotides were analyzed and purified by reversed-phase high performance liquid chromatography (RP-HPLC), flow rate 6 ml/min [buffer A: 20mM triethylammonium acetate (TEAAc, pH 7.1) in water; buffer B: 20mM TEAAc (pH 7.1) in 50% acetonitrile]. The HPLC analysis was performed with a linear gradient from buffer A to 100% buffer B in 20 min. The detritylation was performed by treatment of the oligonucleotide with an acetic acid solution (pH 4.5) under 40 $^{\circ}$C for 1 h. The DMTr-off oligonucleotides were desalted by a Water C-18 column and purified again by RP-HPLC. Native RNAs were purchased from Integrated DNA Technologies. The native, as well as the Se modified DNAs were characterized by matrix assisted laser desorption/ionization-time of fight mass spectrometry (MALDI-TOF MS).

3.4.4 UV-melting and CD measurements

The thermal denaturing experiments were performed on Cary 300 UV-Vis spectrometer equipped with a six-sample cell changer and a temperature controller. the sample annealed by heated to 85 $^{\circ}$C for 2 min and slowly cooling down to 4 $^{\circ}$C before data acquisition. Both denaturing and annealing curves were acquired at 260 nm with 1 cm path length at heating or cooling rates of 0.5 $^{\circ}$C/min and data interval of 0.5 $^{\circ}$C. The samples (2 μM DNA duplexes) were dissolved in the buffer of 350 mM NaCl, 10 mM NaH_2PO_4-Na_2HPO_4 (pH 6.8), 0.1 mM EDTA.

The circular dichroism (CD) experiments were performed on Jasco J-1500 spectrometer. The samples (3 μM DNA duplexes) were dissolved in the buffer of 1 mM NaCl, 1 mM NaH$_2$PO$_4$-Na$_2$HPO$_4$ (pH=7.0), 0.1 mM EDTA.

3.4.5   DNA and DNA/protein complex crystallization

The purified DNA oligonucleotide was heated to 80 °C for 2 min, then cooled down slowly to room temperature. The Nucleic Acid Mini Screen Kit (Hampton Research) was applied to screen the crystallization conditions room temperature using the hanging-drop method by vapor diffusion. The concentrations of the native, Se modified DNAs were adjusted to 1.0 mM in water for crystallization. During the crystallization screening process, the DNA solution was mixed with crystallization buffer at 1:1 ratio (0.5 μL+ 0.5 μL) and equilibrate against 500 μL of the 35% MPD (2-methyl-2,4-pentanediol) solution within a hanging-drop plate.

For the DNA/Bst DNA polymerase complex crystallization, the purified DNA oligonucleotides (1 mM) were annealed by first heating to 80°C for 2 min, and then slowly cool down to room temperature. The resulting DNA solution was mixed with the protein (final concentration: 24 mg/mL) at 1:1 molar ratio. Co-crystallization of Se-DNA/Bst DNA polymerase was achieved by screening with the QIAGEN Classics Suite Kit (www.qiagen.com) using the sitting-drop vapor diffusion method at 25°C, and the mixture #67 was found to be the best buffer condition for the crystal growth. [Salt: 0.2 M Calcium acetate; Buffer: 0.1 M sodium cacodylate, pH 6.5; precipitant: 18% (w/v), PEG 8000].

3.4.6   Diffraction data collection and structure determination

Diffraction data were collected on beamline 8.2.2 of the ALS (Advanced Light Source) at the Lawrence Berkeley National Laboratory. X-ray data were collected under a liquid-nitrogen stream at 99 K. Each crystal was exposed for 0.5 second per image with 1° rotation and a total of

180 images were obtained. The data were integrated and scaled with the programs HKL2000 and DENZO/SCALEPACK.[81] The structures were solved by molecular replacement method using Phaser[82] within CCP4i[83, 84]. The resulting models were refined using REFMAC5.5[85] within CCP4i[83, 84]. The DNA, protein, ions and water molecules were added and modeled using Coot[86].

4 SYNTHESIS OF 2'-MeSe-ARABINOURIDINE FOR THE STRUCTURE STUDIES OF DNA CONTAINING URACIL

4.1 Introduction

4.1.1 *The biological significance of uracil in DNA*

Uracil, naturally occurring in RNA, was also present in DNA in small quantities as a result of either the spontaneous cytosine deamination or misincorporation of dUMP during DNA replication. The biochemical, and structural studies reveal its biological and clinical significance.[91] Spontaneous cytosine deamination is the hydrolysis reaction of cytosine to uracil which occurs readily under physiological circumstances primarily in ss DNA regions, such as transcription bubbles or replication forks, producing 100-500 uracil residues per day, per cell.[92] By replacing the cytosine with uracil, the deamination will lead to U:G mispairing which is highly mutagenic. Mammalian cells contain four main uracil-DNA glycosylases (UNGs), which is the machinery for repair of C deamination, recognizing and excising the uracil from the genome and generating an abasic (AP) site.[93, 94] Compared to cytosine deamination, the misincorporation of dUMP in the place of dTMP in replication is the predominant source of uracil in DNA resulting in U:A pairing which is not directly mutagenic, but may be cytotoxic. The uracil can still be recognized and removed by UNGs and formed an abasic site which is mutagenic and will increase the level of DNA strand breaks leading to cell death.[95, 96] Study shows that the aberrant uracil incorporation leads to cytotoxicity by inhibiting the thymidylate biosynthesis via inhibition of thymidylate synthase (TS) which is an important target of several anticancer drugs exhibiting significance in clinic studies.[97] Uracil could also be introduced into DNA by enzymatic deamination of cytosine in the Ig locus of B-cells which is important for diversification of antibodies after antigen exposure.[98]

The study of uracil in DNA could also reveals the evolution of DNA by answering the question that why the nature chooses thymine instead of uracil. DNA replaces RNA because of its superior information-storage capacity took place in two steps as suggested by modern biochemical pathways.[99, 100] The first step is the replacement of ribose with deoxyribose, followed by the replacement of uracil with thymine. The tinkering nature of the evolution provide an explanation that the replacement of uracil driven by the occasional removal of U opposite A by the proto-UNG/MUG solved the cytosine deamination problem which results in U:G mispairs.[101] The greater photochemical mutation resistance of thymine could be another reason which making the genetic message more stable.[102] This explanation was further supported by the facts that the only different base between DNA and RNA is the photosensitive one and the thymine is only in DNA which has a requirement for long-term stability. However, the final answer to this question still requires more fundamental studies.

4.1.2   *Structural biology studies of DNA containing Uracil*

X-ray crystallography is one of the most powerful methods in structural biology which provide insight into the structure of DNA containing uracil and could further facilitate the biomolecular mechanistic research, such as the human uracil-DNA glycosylase, and drug discovery. The DNA of the Bacillus subtilis transducing bacteriophage PBS2 containing uracil in the place of thymine and the structure was determined by X-ray diffraction which indicates that the substitution of uracil for thymine has no major effect on the overall structure.[103] The structure studies of human UNG bound to DNA containing uracil reveals the extrahelical base recognition and a catalytic mechanism for hydrolytic base excision.[104-106] Further study indicates that the base-pair dynamics plays a critical role in the genome-wide search for uracil by UNG which is initiated by thermally induced opening of T:A and U:A base pairs.[107] Another structure study of archaeal

DNA polymerases suggests its ability to recognize the uracil in DNA and stall replication on encountering the pro-mutagenic base.[108] However, the uracil recognition capability is limited to the archaea, not occurring with bacteria and eukarya.[109]

4.1.3 Selenium modifications on uridine

4.1.3.1 2'-MeSe-U for A-form DNA crystallization

As mentioned previously, selenium functionality has been successfully introduced into different position of uridine and incorporated into oligonucleotides to solve the phase problem for nucleic acid crystallography through multiwavelength anomalous dispersion (MAD) or single-wavelength anomalous dispersion (SAD). The synthesis of 2'-selenium uridine analogues and the oligonucleotides including DNA as well as RNA were reported by Du and Teplova et al. in 2002 (Scheme 4.1).[37, 77] The 2'-α-position selenium modification retains the 3'-endo sugar puckering of A-form DNA and RNA molecules. The crystal structure shows that the methylselenium moiety is point to the minor groove of the duplex and the selenium-derivatized DNA exhibit virtually identical conformation to the corresponding native without alter the A-DNA duplex geometry or the minor groove hydration significantly (Figure 4.1).
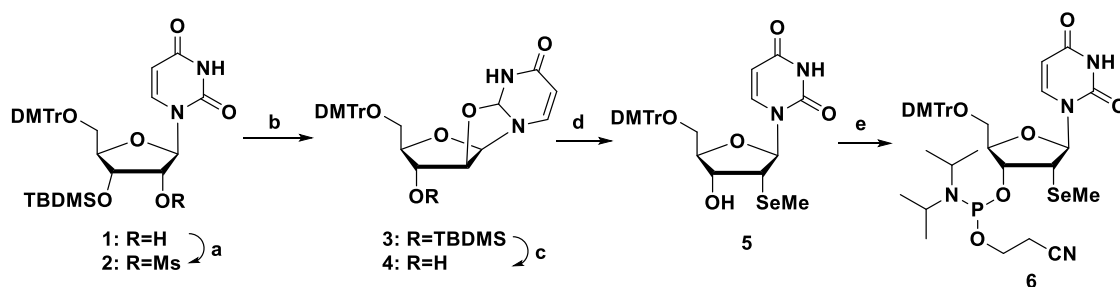
Scheme 4.1 The synthesis of 2'-MeSe-uridine phosphoramidite
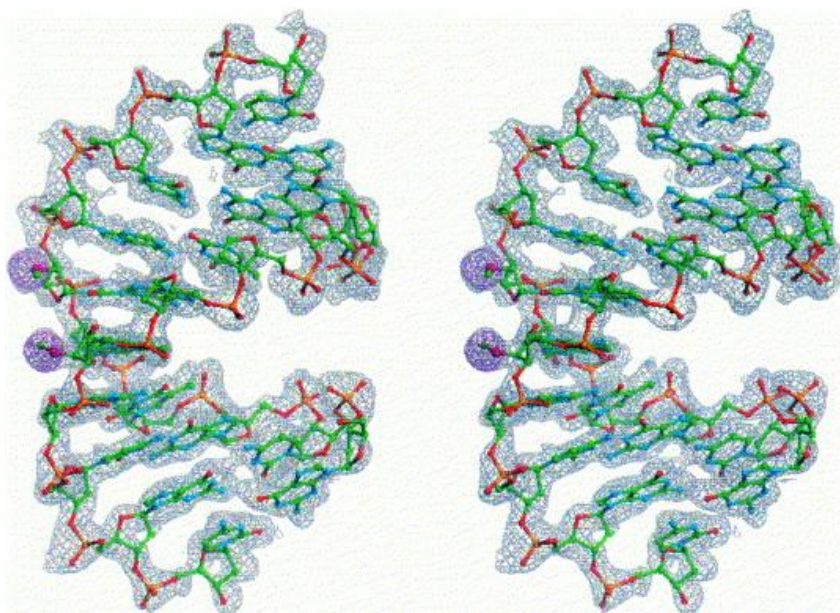
Figure 4.1 global structure and electron density map of the decamer DNA duplex (1.3 Å, PDB ID: 1MA8) [d(GCGTA)^{Se}Ud(ACGC)]$_2$

4.1.3.2 2-Se modified U/T and their oligo crystal structure

To study the high fidelity of base pairing in both DNA and RNA, 2-selenium modified thymidine and uridine was synthesized and incorporated into oligonucleotides (Scheme 4.2).[27, 28] Both the X-ray crystal structure of Se-DNA (5'-G-$^{2'MeSe}$dU-G-^{2Se}T-ACAC-3')$_2$ and Se-RNA r(5'-GUAUA-^{2Se}U-AC-3')$_2$ were solved by molecular replacement at high resolution (1.58 Å, 2.2 Å respectively) (Figure 4.2). The results indicate that 2-Se-modified DNA/RNA has very similar structure to the corresponding native. And the 2-Se-substitution enhanced the base pair fidelity by further discouraging the T/C and T/G or U/G wobble pair in DNA and RNA respectively, but without alter the native base pairing and overall duplex structure significantly. The steric hindrance introduced by the bulky Se atom and the electronic effect, inability of forming a stable hydrogen bond, are main factors responsible for the discrimination against wobble pair, hence increasing the base pair recognition. Theoretical studies including quantum mechanics calculation, molecular dynamics simulation and free-energy perturbation simulation further confirmed the experimental conclusions.[110] The free-energy calculation indicates that both the disfavored hydrogen bonding

interaction and the unfavorable solvent entropic contribution lead to the high specificity of the 2-Se-T/A pair.
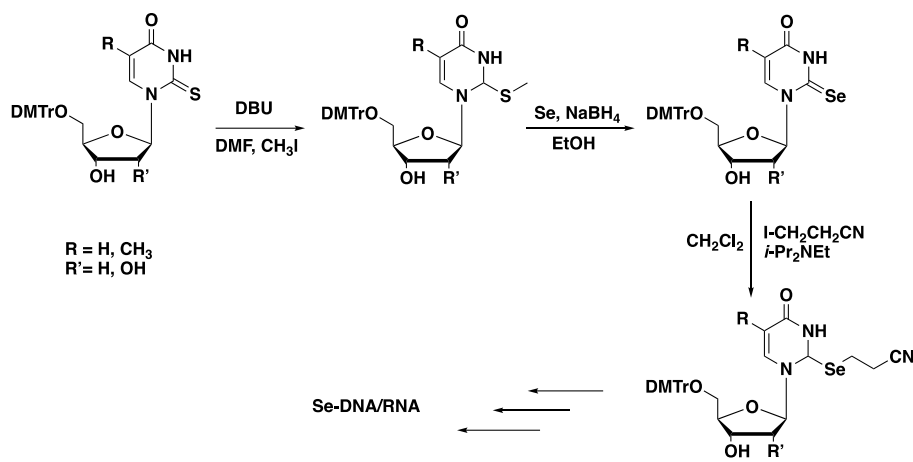
Scheme 4.2 Synthesis of 2-Se-U/T modified DNA and RNA
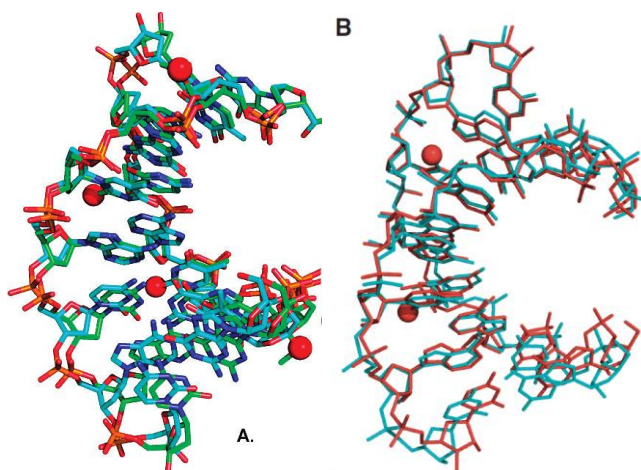
Figure 4.2 A. SeT-DNA; B. the superimpose comparison of one ^{Se}U-RNA duplex (red; PDB ID: 3S49) with its native counterpart (cyan; PDB ID: 246D).

4.1.3.3 4-Se-uridine and its RNA crystal structure

Subsequently, Sheng et al. reported the synthesis of 4-Se-uridine and Se-RNA to probe the U·U and U/A base pair.[111] The Se-modified RNA duplexes r(5'-U-^{4Se}U-CGCG-3')$_2$ and r(5'-GUG-^{4Se}U-AAC-3')$_2$ formed yellow crystals, and the structures were determined via SAD and molecular replacement (Figure 4.3). The Se-derivatized and native structure are virtually identical based on the crystal structure and thermostability studies which implies the RNA flexibility to accommodate the large Se atom. The comparison of the Se-modified and native structure also

suggested a stronger stacking interaction between Se-nucleobase to 3'-cytosine which can strengthen the RNA duplexes. The decrease of H-bonding in U/A pair by 4-Se modification of uridine was compensated by the base-stacking increase which may also increase the duplex junction thermostability. By introducing the probe to the RNA duplex junctions, the study reveals that each Se atom contributed additional stabilization about 0.4-0.8 kcal/mol to the RNA duplex due to the uracil stacking. Interestingly, color RNA was generated by a single Se-modification which shifts the uridine UV spectrum over 100 nm.
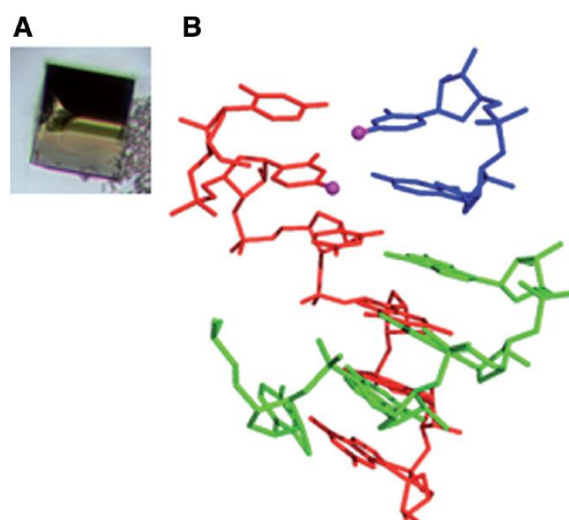
Figure 4.3 crystal and structure of the 4-Se-U RNA hexamer, (5'-U-^{Se}U-CGCG-3')$_2$

So far, several Se-modified uridines as well as their oligonucleotides have been synthesized. Their structures were also determined by X-ray diffraction. However, all modifications were limited to either A-form DNA or RNA where both of them containing the northern sugar conformation. In order to investigate the structure of DNA containing uracil, the modified uridine analogue needs to accommodate the southern sugar pucker conformation as well as the B-form DNA geometry. To solve this problem, herein, we describe the first synthesis of 2'-MeSe-arabinouridine phosphoramidite and its incorporation into DNA by solid-phase synthesis. The 2'-β-position selenium modification retain the 2'endo sugar conformation and fit to the B-

form DNA geometry. The X-ray crystal structure of the Se-DNA was determined at a resolution of 1.25 Å and the Se-derivatized structure was virtually identical to the native one.

4.2 Results and discussion

4.2.1 The synthesis of 2'MeSe-arabinouridine phosphoramidite and DNA oligos

The synthesis of 2'-MeSe-arabinouridine phosphoramidite (Scheme 4.3) is similar to the synthesis of corresponding selenium modified thymidine (Chapter 2, Section 2.4.2). The 5',3'-hydroxyl groups of the uridine were protected with tetraisopropyldisilylene (TIPDS), followed by the protection of 3-N with benzyloxymethyl acetal (BOM). To introduce the methylseleno group, the 2'-OH group was activated with mesylate. Then, dimethyldiselenide was dissolved in anhydrous THF and treated with n-butyllithium at -78 $^{\circ}$C to generate the MeSe nucleophile. The activated nucleoside **U4** was dissolved in anhydrous THF and injected into the reaction at 60 $^{\circ}$C which is lower than the selenium incorporation of corresponding thymidine derivative indicating a lower energy barrier of the reaction. This result is consistent with the computational study of the transition state and activation energy of the methylseleno group incorporation for uridine, thymidine, cytidine as well as 5-F-uridine (Appendix A.). After the selenium incorporation, the BOM group was deprotected by treated with BBr$_3$ under -78 $^{\circ}$C in anhydrous toluene followed by carefully quenching the reaction with a mixture of triethylamine (TEA) and *i*-PrOH (v/v, 1:1). Only treated with TEA will reproduce the starting material. In the meanwhile, it is important to use anhydrous *i*-PrOH due to the nucleophilic protic solvent, such as MeOH or water, can serve as a nucleophile to substitute the MeSe group which was also activated by BBr$_3$ (Scheme 4.4). The compound **U6** was treated with 3HF·TEA to remove the TIPDS group and the 5-OH was selectively protected with the DMTr group. Finally, the 2'-Se-arabinouridine **U8** was converted to corresponding phosphoramidite **U9** in 90% yield.
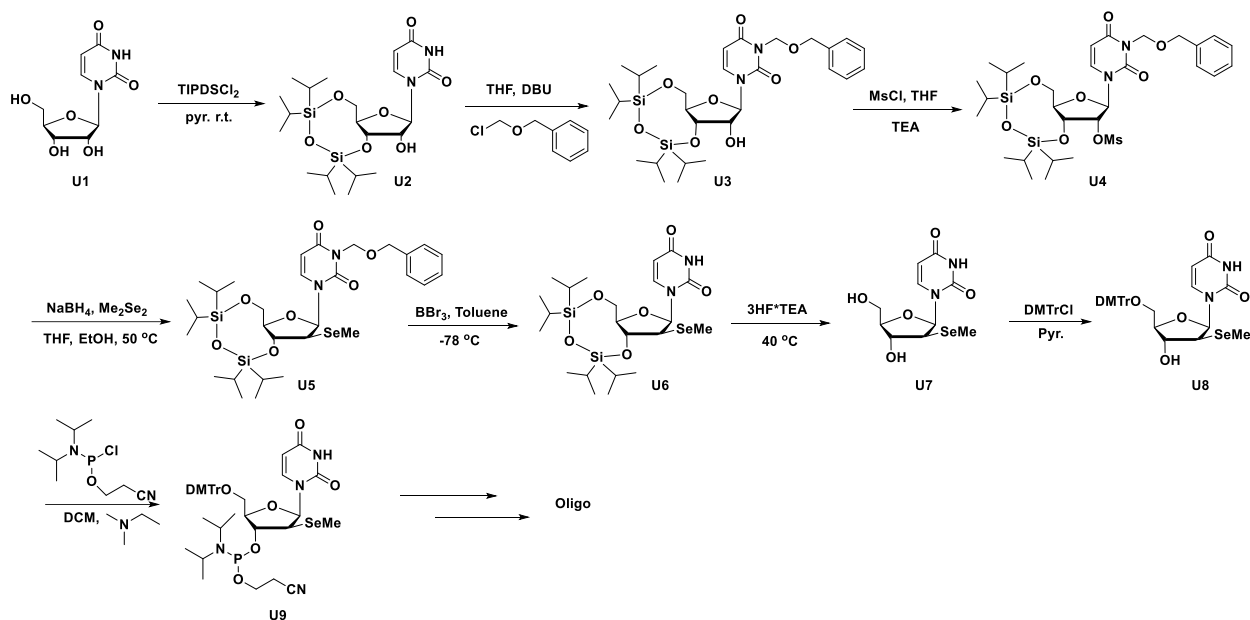
Scheme 4.3 Synthesis of 2'-MeSe-arabinouridine phosphoramidite
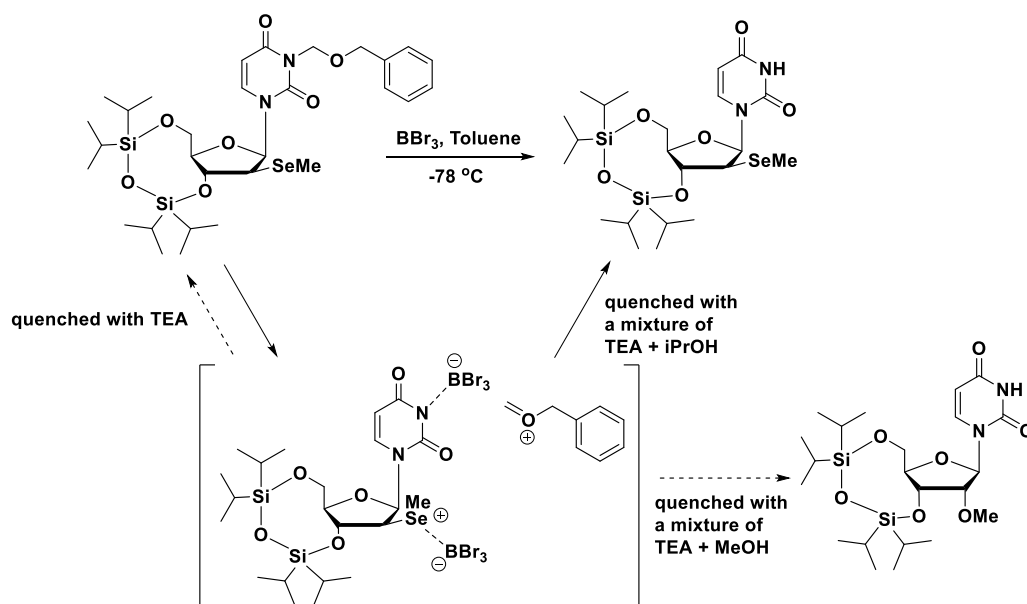
Scheme 4.4 Deprotect BOM group and quench the reaction with a mixture of TEA and iPrOH

The phosphoramidite **U9** was incorporated into DNA oligos through solid-phase synthesis with phosphoramidites containing standard base protection. The synthesized DNA oligo was cleaved off the solid support and fully deprotected by treating with concentrated ammonia at 55 °C for 5 h, followed by the reverse-phase HPLC purification. Table 4.1 show the MALDI-TOF MS analysis of all the purified DNA oligo.

Table 4.1 MALDI-TOF MS Data of ^{Se}U-DNAs

Entry	Se-oligonucleotides	Measured (calcd) m/z
1	dU-s1 5'-CGCGAA$^{2'Se}$UUCGCG-3' $C_{115}H_{145}N_{46}O_{70}P_{11}Se$ FW 3711.3	$[M]^+$: 3711.2 (3711.3)
2	dU-s2 5'-CGCGAAU$^{2'Se}$UCGCG-3' $C_{115}H_{145}N_{46}O_{70}P_{11}Se$ FW 3711.3	$[M]^+$: 3710.8 (3711.3)
3	U-s1 5'-CGCGAA$^{2'Se}$UTCGCG-3' $C_{116}H_{147}N_{46}O_{70}P_{11}Se$ FW 3725.4	$[M]^+$: 3725.1 (3725.4)
4	U-s2 5'-CGCGAAT$^{2'Se}$UCGCG-3' $C_{116}H_{147}N_{46}O_{70}P_{11}Se$ FW 3725.4	$[M]^+$: 3725.4 (3725.4)
5	Native 1 5'-CGCGAAUUCGCG-3' $C_{114}H_{143}N_{46}O_{70}P_{11}$ FW 3616.6	$[M+H]^+$: 3617.6 (3617.6)
6	Native 2 5'-CGCGAAUTCGCG-3' $C_{115}H_{145}N_{46}O_{70}P_{11}$ FW 3632.4	$[M+H]^+$: 3633.3 (3633.4)
7	Native 3 5'-CGCGAATUCGCG-3' $C_{115}H_{145}N_{46}O_{70}P_{11}$ FW 3632.4	$[M+H]^+$: 3633.0 (3633.4)

4.2.2 Thermostability studies

Two single modified Dickerson-Drew dodecamer were synthesized for the UV melting experiment to study the impact of 2'-MeSe-arabino-dU on duplex thermal stability. With the modification, about 8 ^{o}C drop in melting temperature were observed compared with the corresponding native DNA suggesting that the Se-modified uridine destabilized the DNA duplex (Table 4.2). This effect may come from the interaction between the methyl group on selenium with nucleobases based on our structure study (Figure 4.4). Changing the position of the modified uridine shows similar destabilization.

Table 4.2 Melting temperature of Native and ^{Se}U-DNA duplexes

Entry	Sequence	Melting temperature (°C)	ΔT_m(°C)
U-n1	5'-CGCGAAUTCGCG-3'	64.5	-
U-s1	5'-CGCGAA^{Se}UTCGCG-3'	56.2	8.3
U-n2	5'-CGCGAATUCGCG-3'	64.4	-
U-s2	5'-CGCGAAT^{Se}UCGCG-3'	56.4	8.0
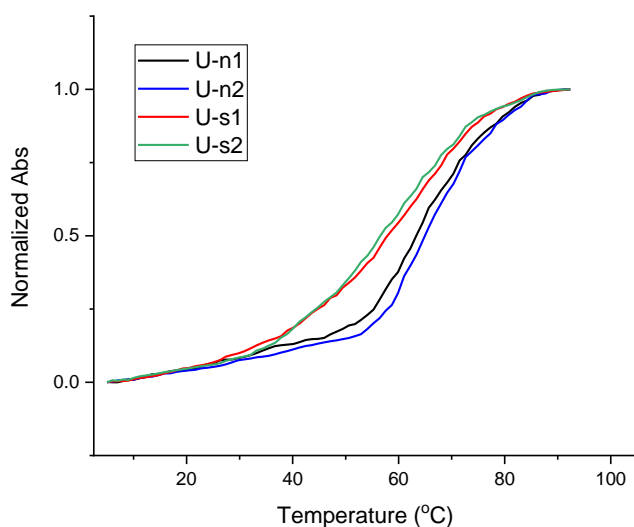
Figure 4.4 Normalized UV melting curve

4.2.3 Structure determination

The determined crystal structure (1.25 Å resolution) of Se-DNA is superimposable to the corresponding native structure (1.25 Å resolution) with same space group ($P2_12_12_1$). The structure reveals that the 2'-Se-arabino-furanose displays the 2'-endo sugar pucker and the 2'-methylseleno group is in the major groove of the B-DNA duplex. The Se-modified and the native structures are also virtually identical with a RMSD of 0.155Å for all 430 atoms which support that the incorporation of 2'-MeSe-arabinothymidine does not alter the global structure of the complex (Figure 4.5, left). 2'-SeMe-arabino-dU/dA base pair is also superimposable over native U/A base pair (Figure 4.5, right). The data collection and structure refinement statistics are summarized in Table 4.3.
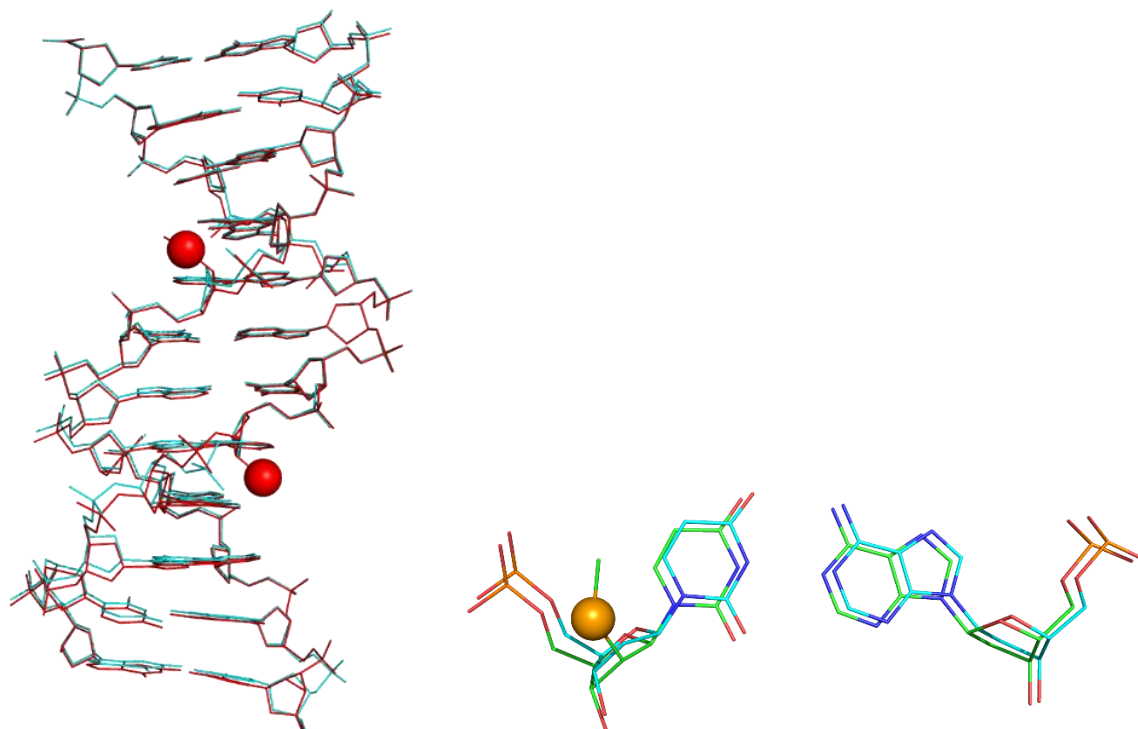
Figure 4.5 Superimposed global and local structure of ^{Se}U-modified DNA as well as the corresponding native

Table 4.3 Data collection and refinement statistics of ^{Se}U-DNA and the corresponding native

Data Collection	U-s2	Native 3
Wavelength (Å)	1	1
Resolution range (Å)	50 - 1.25 (1.27 – 1.25)	50 - 1.25 (1.27 - 1.25)
Space group	P 21 21 21	P 21 21 21
Unit-cell a, b, c (Å)	25.2, 39.8, 65.2	25.4, 39.7, 65.4
α, β, γ (°)	90, 90, 90	90, 90, 90
Number of reflections	18820 (905)	17963 (854)
Completeness (%)	99.6 (98.5)	94.2 (91.0)
I/σ	32 (4.7)	73 (5.3)
R merge (%)	6.0 (46.1)	6.3 (49.9)
Refinement		
Resolution range (Å)	34.00 - 1.25	39.94 - 1.25
Number of reflections	17603	17062
Completeness (%)	98.2	94.2
R value (%)	21.4	20.6
R free (%)	23.1	22.8
Average B value (Å^2)	15.4	20.0
Rms Bond Length (Å)	0.020	0.015

Rms Bond Angle (º)	3.15	1.79

Interestingly, by comparing the DNA structure containing Se-modified uridine and thymidine respectively, we found that the direction of the methyl group is different. For the 2'βSe-U modified DNA, the methyl group is located between two nucleobase and is quite close to the proximate cytosine (Figure 4.6, top). However, when comes to the 2'βSe-T, the methyl group is pointed away from the nucleobase and has a long distance between either the uracil or the cytosine (Figure 4.6, bottom). This effect may be explained by the electron deficient uracil attracting the electron-rich methyl group which may contribute to the destabilization observed in the UV melting studies for [Se]U modified instead of [Se]T modified DNA (Chapter 2, Section 2.2.2).
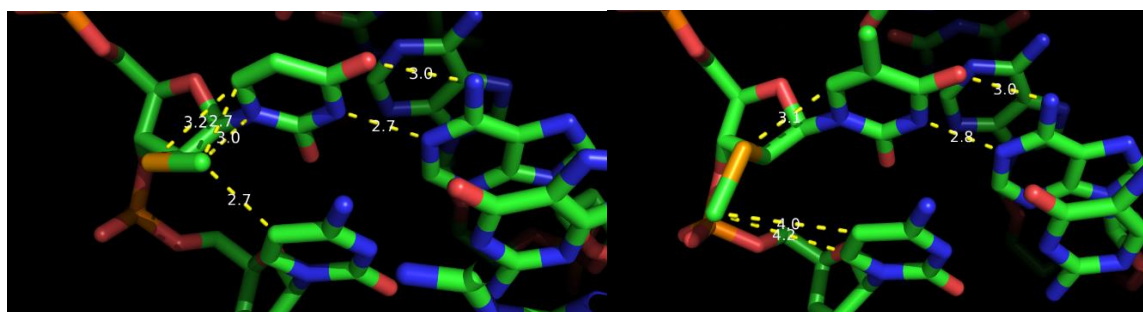
Figure 4.6 Local structures of the SeU-containing DNA [5'-CGCGAAT([Se]U)CGCG]$_2$ and SeT-containing DNA [5'-CGCGAAT([Se]T)CGCG]$_2$.
Left, the distance between the methyl group and uracil as well as cytosine; Right, the distance between the methyl group and thymine as well as cytosine.

In the meanwhile, we also compared the B-factors of the [Se]U-DNA and the native normalized (Figure 4.7) by the same method using for [Se]T-DNA (Chapter 2, Section 2.2.5). In contrast to [Se]T-DNA, instead of decrease, significant increase of the B-factor was observed on one of the DNA strands at the phosphate group of the Se-U as well as the adjacent cytidine residue compared to the native one which may also cause by the interaction between the methyl group and the adjunct nucleobases. However, the similar average B-value of the SeU-DNA (15.4 Å^2) and the corresponding native one (20.0 Å^2) indicating that the Se functionality does not change the global molecular dynamics of the DNA duplex significantly.
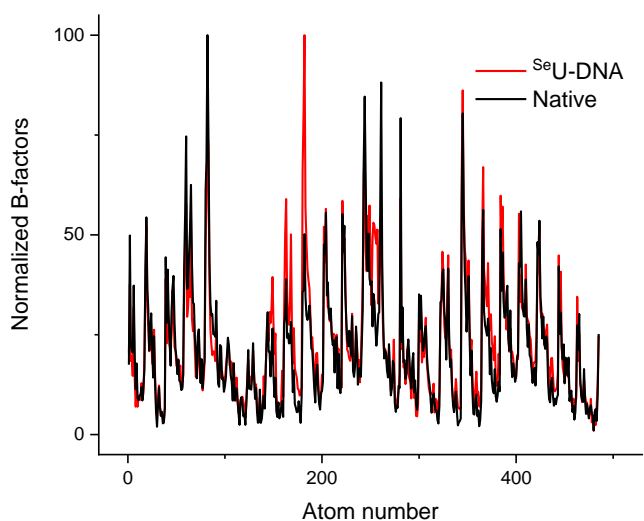
Figure 4.7 B-factor comparison between ^{Se}U-DNA and the native

4.3   Conclusion

In conclusion, we developed a route to synthesize the 2'-MeSe-arabinouridine analogues and oligonucleotides containing selenium modification. High resolution structures were obtained through X-ray crystallography for DNA duplex (1.25 Å). The structure shows that the 2'-β-position selenium derivatization retains the native 2'-endo conformation in the B-form DNA duplex and the modification does not cause any significant structure perturbation. Compared to the ^{Se}T-DNA, the electron deficient uracil attracts the methyl group on the selenium leading to interaction with the adjunct nucleobases which may result in the drop of melting temperature as well as the increase in local B-factors around the modified nucleoside. The incorporation of the selenium functionality solving the phasing problem could help the structure study of DNAs containing uracil.

4.4   Experimental section

4.4.1   General

Most solvents and reagents were purchased from Sigma, Fluka, or Aldrich (p.a.) and used without purification unless mentioned otherwise. Triethylamine (TEA) was dried over KOH (s) and distilled under argon. When necessary, solid reagents were dried under high vacuum. Reactions with compounds sensitive to air or moisture were performed under argon. Solvent mixtures are indicated as volume/volume ratios. Thin layer chromatography (TLC) was run on Merck 60 F254 plates (0.25 mm thick; Rf values in the text are for the title products), and visualized under UV-light or by a Ce-Mo staining solution (phosphomolybdate, 25 g; Ce$(SO_4)_2 \cdot 4H_2O$, 10 g; H_2SO_4, 60 mL, conc.; H_2O, 940 mL) with heating. Flash chromatography was performed using Fluka silica gel 60 (mesh size 0.040-0.063 mm) using a silica gel:crude compound weight ratio of ca. 30:1. ^{1}H, ^{13}C and ^{31}P-NMR spectra were recorded using Bruker-300 or 400 (300 or 400 MHz). All chemical shifts (δ) are in ppm relative to tetramethylsilane and all coupling constants (J) are in Hz. High resolution (HR) MS were either obtained with electrospray ionization (ESI) on a Q-TOFTM Waters Micromass at Georgia State University.

4.4.2   Synthesis protocol and characterization
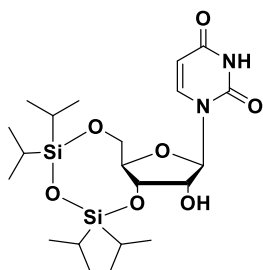
3′,5′-O-(1,1,3,3-tetraisopropyldisiloxane-1,3-diyl)-uridine (**4.1**) Uridine (11 g, 49.1 mmol) was suspended in 250 mL dry pyridine with argon and the mixture was placed in an ice-bath. To the suspension, 1,3-dichloro-1,1,3,3-tetraisopropyl-disiloxane (12.8 mL, 58.9 mmol) was added dropwise. Then the ice-bath was removed, and the reaction was stirred at room temperature

for 5 h. After the reaction completed, 5 mL of Methanol was added. The solvent was evaporated under reduced pressure and dissolved in methylene chloride (300 mL). The organic solution was washed with HCl (3 M, 75 mL), saturated sodium bicarbonate (75 mL), water (100 mL), brine (100 mL), dried over anhydrous MgSO$_4$, and concentrated in vacuum to given **4.1** as white solid. The product was used in the next step without further purification. ^{1}H-NMR (400 MHz, CDCl$_3$) δ (ppm): 1.03-1.09 (m, 28H, 4×iPrSi), 3.28 (br, s, 1H, OH), 4.01 (dd, J_1=2.2 Hz, J_2=13.1 Hz, 1H, H-5'), 4.09 (d, J=8.4 Hz, 1H, H-4'), 4.18 (m, 2H, H-2', H-5'), 4.38 (dd, J_1=4.7 Hz, J_2=8.6 Hz, 1H, H-3'), 5.67 (d,) 5.71 (d, J=8.1 Hz, 1H, H-1'), 7.66 (d, J=8.1 Hz 1H, H-6), 8.52 (br, s, 1H, NH). ^{1}H-NMR spectrum is identical to the literature[112].
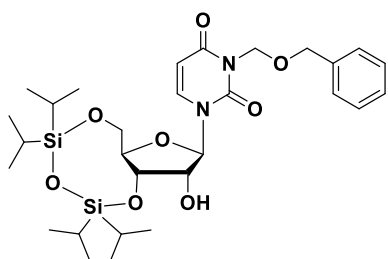
3′,5′-O-(1,1,3,3-tetraisopropyldisiloxane-1,3-diyl)-N3-(benzyloxymethylacetal)-uridine (4.2) Compound **4.1** (6.5 g, 13.4 mmol) was dissolved in 50 mL THF followed by the addition of 1,8-Diazabicyclo(5.4.0)undec-7-ene (DBU, 4 mL, 26.8 mmol). The mixture was treated with benzyl chloromethyl ether (2.8 mL, 20.1 mmol) and stirred at room temperature for 2 h. The solvent was evaporated under reduced pressure and washed with water (50 mL), brine (70 mL), dried over MgSO$_4$ and concentrated. The residue was purified by silica gel chromatography (25 % ethyl acetate in hexane) to give desired product **4.2** (6.6 g, 81 %). ^{1}H-NMR (400 MHz, CDCl$_3$) δ (ppm): 1.02-1.10 (m, 28H, 4×iPrSi), 2.90 (br, s, 1H, OH), 4.00 (dd, J_1=2.8 Hz, J_2=13.2 Hz, 1H, H-5'), 4.10 (dt, J_1=3.7 Hz, J_2=8.5 Hz, 1H, H-4'), 4.13 (d, J=4.9 Hz, 1H, H-2'), 4.21 (dd, J_1=1.4 Hz, J_2=13.3 Hz, 1H, H-5'), 4.34 (dd, J_1=4.9 Hz, J_2=8.8 Hz, 1H, H-3'), 4.72 (s, 2H, CH$_2$OC<u>H</u>$_2$Ph), 5.47 (d, J=2.4 Hz, 2H, C<u>H</u>$_2$OCH$_2$Ph), 5.71 (d, J=8.2 Hz, 1H, H-5), 5.73 (s, 1H, H-

1'), 7.38-7.26 (m, 6H, aromatic, H-6), 7.64 (d, J=8.2 Hz, 1H, H-6). ^{13}C-NMR (100 MHz, CDCl$_3$) δ (ppm): 12.5, 12.9, 13.4, 16.8, 16.9, 17.0, 17.1, 17.3, 17.4 (i-Pr), 60.2 (C-5'), 68.9 (C-3'), 70.3 (CH$_2$), 72.4 (CH$_2$), 75.3 (C-2'), 81.9 (C-4'), 91.1 (C-1'), 101.5 (C-5), 127.7, 128.3 (Ar-C), 137.9 (Ar-C), 138.4 (C-6), 150.7 (C-2), 162.7 (C-4). HRMS (ESI): C$_{29}$H$_{46}$N$_2$O$_8$Si$_2$; [M+Na]$^+$: 629.1387 (calc. 629.2690). ^{1}H-^{13}C HSQC, ^{1}H-^{13}C HMBC and ^{1}H-^{1}H COSY (Appendix D.3).
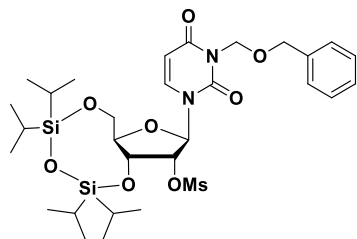
2'-O-mesyl-3′,5′-O-(1,1,3,3-tetraisopropyldisiloxane-1,3-diyl)-N3-(benzyloxymethyl acetal)-uridine (4.3) Compound **4.2** (12.5 g, 20.6 mmol) was dissolved in 150 mL THF followed by the addition of triethylamine (11.5 mL, 82.4 mmol). To the solution, methanesulfonyl chloride (3.2 mL, 41.2 mmol) was added dropwise and the reaction was stirred at room temperature for 3 h. The solvent was evaporated under reduced pressure and dissolved in methylene chloride (150 mL). The organic solution was washed with water (50 mL) and brine (50 mL), dried over MgSO$_4$ and concentrated. The residue was purified by silica gel chromatography (15 % ethyl acetate in hexane) to give desired product **4.3** (12.7 g, 90 %). ^{1}H-NMR (400 MHz, CDCl$_3$) δ (ppm): 1.04-1.11 (m, 28H, 4×iPrSi), 3.28 (s, 3H, O$_3$SCH$_3$), 3.99 (dd, J_1=2.3 Hz, J_2=13.7 Hz, 1H, H-5'), 4.10 (dd, J_1=1.9 Hz, J_2=9.7 Hz, 1H, H-4'), 4.25-4.30 (m, 2H, H-3', H-5'), 4.71 (s, 2H, CH$_2$OC$\underline{H}_2$Ph), 4.97 (d, J=4.4 Hz, 1H, H-2'), 5.48 (m, 2H, C$\underline{H}_2$OCH$_2$Ph), 5.72 (d, J=8.2 Hz, 1H, H-5), 5.79 (s, 1H, H-1'), 7.26-7.37 (m, 5H, aromatic), 7.74 (d, J=8.2 Hz, 1H, H-6). ^{13}C-NMR (100 MHz, CDCl$_3$) δ (ppm): 12.6, 12.8, 12.9, 13.6, 16.8, 16.9, 17.2, 17.4 (i-Pr), 39.3 (O$_3$SCH$_3$), 59.0 (C-5'), 66.6 (C-3'), 70.3 (CH$_2$), 72.4 (CH$_2$), 82.0 (C-4'), 82.8 (C-2'), 88.9 (C-1'), 101.9 (C-5), 127.6, 127.8, 128.3

(Ar-C), 137.3 (C-6), 137.8 (Ar-C), 150.7 (C-2), 162.4 (C-4). HRMS (ESI): $C_{30}H_{48}N_2O_{10}SSi_2$; $[M+H]^+$: 685.1212 (calc. 685.2646). ^1H-^{13}C HSQC and ^1H-^1H COSY (Appendix D.3).
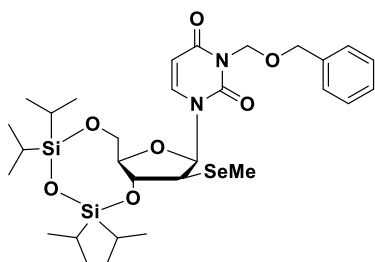
3′,5′-O-(1,1,3,3-tetraisopropyldisiloxane-1,3-diyl)-N3-(benzyloxymethylacetal)-2'-methylseleno-arabinouridine (4.4) Me_2Se_2 (2.4 mL, 25.3 mmol) was dissolved in anhydrous THF (25 mL) under argon. The solution was cooled to -78 °C in an acetone/dry ice bath followed by the injection of n-BuLi (5 mL, 2.5 M in hexane, 12.6 mmol). The reaction was allowd to warm up to room temperature slowly. Then, the starting material **4.3** (4.3 g, 6.3 mmol) was dissolved in 25 mL anhydrous THF and added into the solution. The reaction mixture was heated up to 65 °C and stirred for 16 h. The solution was concentrated under reduced pressure and re-dissolved in ethyl acetate (100 mL). The organic solution was washed with water (30 mL), brine (40 mL), dried over $MgSO_4$, and evaporated to dryness to afford crude product. The residue was then purified by silica gel chromatography (10 % ethyl acetate in hexane) to give desired product **4.4** (3.0 g, 69.8 %) as white solid. ^1H-NMR (400 MHz, $CDCl_3$) δ (ppm): 1.03-1.11 (m, 28H, 4×iPrSi), 2.08 (s, 3H, $SeCH_3$), 3.64 (dd, J_1=7.0 Hz, J_2=10.1 Hz, 1H, H-2'), 3.76 (d, J=8.1 Hz, 1H, H-4'), 4.04 (dd, J_1=2.7 Hz, J_2=13.2 Hz, 1H, Ha-5'), 4.13 (dd, J_1=1.7 Hz, J_2=13.2 Hz, 1H, Hb-5'), 4.18 (dd, J_1=8.6 Hz, J_2=9.7 Hz, 1H, H-3'), 4.68 (d, J=3.1 Hz, 2H, $CH_2OC\underline{H}_2Ph$), 5.51 (s, 2H, $C\underline{H}_2OCH_2Ph$), 5.75 (d, J=8.2 Hz, 1H, H-5), 6.40 (d, J=6.9 Hz, 1H, H-1'), 7.28-7.37 (m, 5H, Ph), 7,58 (d, J=8.2 Hz, 1H, H-6). ^{13}C-NMR (100 MHz, $CDCl_3$) δ (ppm): 6.7 ($SeCH_3$), 12.5, 12.9, 13.0, 13.8, 17.0, 17.1, 17.2, 17.3, 17.5 (i-Pr), 50.1 (C-2'), 60.1 (C-5'), 70.2 (CH_2), 71.8 (CH_2), 72.9 (C-3'), 83.8 (C-4'), 85.5 (C-1'), 101.7 (C-5), 127.0, 127.7, 127.8, 128.3, 128.6 (Ar-C), 137.9 (Ar-C), 138.5 (C-6), 151.3

(C-2), 162.7 (C-4). HRMS (ESI): $C_{30}H_{48}N_2O_7SeSi_2$; [M+Na]$^+$: 707.2105 (calc. 707.2063). ^{1}H-^{13}C HSQC, ^{1}H-^{13}C HMBC and ^{1}H-^{1}H COSY (Appendix D.3).
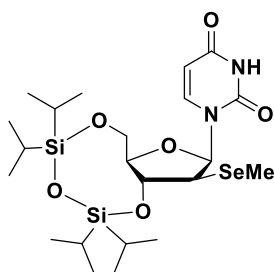
3′,5′-O-(1,1,3,3-tetraisopropyldisiloxane-1,3-diyl)-2'-methylseleno-arabinouridine

(4.5) Compound **4.4** (1.7 g, 2.5 mmol) was placed in a round bottom flask, purge with argon and dissolved in 15 mL anhydrous toluene. The solution was cooled down to -78 ^{o}C by placing in a dry ice-acetone bath. To the solution, BBr$_3$ solution (3.8 mL, 1 M in hexane, 3.8 mmol) was injected and the reaction was stirring at -78 ^{o}C for 1 h. Then, the reaction was quenched by adding a mixture of triethylamine and anhydrous isopropanol (2 mL, 1:1, v/v), followed by the remove of the dry ice-acetone bath. The solution was stirred at room temperature for another 1 h. The reaction was concentrated to around 2 mL and diluted with 50 mL of ethyl acetate. The organic solution was washed with water (30 mL). The water layer was then extracted with ethyl acetate (3 x 10 mL), and the combined organic layer was washed with brine (40 mL), dried over MgSO$_4$ (s), followed by filtration. The solvent was evaporated under reduced pressure to afford the crude product. The residue was then purified by silica gel chromatography (15 % ethyl acetate in hexane) to give desired product **4.5** (1.0 g, 70 %). ^{1}H-NMR (400 MHz, CDCl$_3$) δ (ppm): 1.05-1.12 (m, 28H, 4×iPrSi), 2.13 (s, 3H, SeCH$_3$), 3.64 (dd, J_1=7.0, J_2=10.1 Hz, 1H, H-2'), 3.77 (dt, J_1=2.4, J_2=8.1 Hz, 1H, H-4'), 4.05 (dd, J_1=2.8, J_2=13.2 Hz, 1H, Ha-5'), 4.13 (dd, J_1=1.9, J_2=13.2 Hz, 1H, Hb-5'), 4.21 (dd, J_1=8.3, J_2=10.0 Hz, 1H, H-3'), 5.72 (dd, J_1=2.0, J_2=8.1 Hz, 1H, H-5), 6.39 (d, J=7.0, 1H, H-1'), 7.59 (d, J=8.2 1H, H-6), 9.10 (br, 1H, NH). ^{13}C-NMR (100 MHz, CDCl$_3$) δ (ppm): 6.6 (SeCH$_3$), 12.5, 12.9, 13.0, 13.8, 17.0, 17.1, 17.2, 17.3, 17.4 (i-Pr), 49.9 (C-2'), 60.1 (C-

5'), 73.1 (C-3'), 83.8 (C-4'), 84.8 (C-1'), 102.1 (C-5), 139.9 (C-6), 150.5 (C-2), 163.1 (C-4).

HRMS (ESI): $C_{22}H_{40}N_2O_6SeSi_2$; [M-H]⁻: 563.1525 (calc. 563.1512). ^{1}H-^{13}C HSQC, ^{1}H-^{13}C

HMBC and ^{1}H-^{1}H COSY (Appendix D.3).
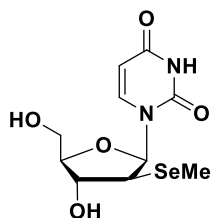
5-methyl-2'-methylseleno-arabinouridine (4.6) Compound **4.5** (1 g, 1.8 mmol) was

dissolved in 10 mL anhydrous THF, and treated with 3HF·Et$_3$N (0.29 mL, 1.8 mmol) at 40 °C for

2 h. The mixture was evaporated to dryness, and the residue was quickly purified by a short flash

column chromatography on silica gel (5% MeOH in methylene chloride) to remove salt and used

for next step without further purification. HRMS (ESI): $C_{10}H_{14}N_2O_5Se$; [M-H]⁻: 320.9982 (calc.

320.9990).
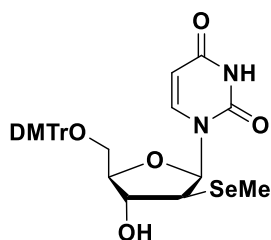
5'-O-(4,4'-dimethoxytrityl)-5-methyl-2'-methylseleno-arabinouridine (**4.7**)

Compound **4.6** (300 mg, 0.9 mmol) was dried over high vacuum and co-evaporated with anhydrous

pyridine (2 x 10 mL). Then, the starting material was dissolved in 15 mL anhydrous pyridine and

cooled by ice bath, followed by the treatment of dimethoxytrityl chloride (DMTrCl) (339 mg, 1.0

mmol) under argon. The reaction was warmed up to room temperature and stirred for 2 h. Then

the solvent was removed under vacuum, the residue was dissolved in 30 mL methylene chloride.

The organic solution was washed with water (20 mL), brine (30 mL), dried over MgSO$_4$ (s),

followed by filtration and evaporation. The crude product was purified by silica gel

chromatography (3% MeOH in methylene chloride with 1% Et$_3$N) to afford the desired product **4.7** (505 mg, 90%) as white solid. ^{1}H-NMR (400 MHz, CDCl$_3$) δ (ppm): 2.13 (s, 3H, SeCH$_3$), 2.66 (d, *J*=4.1 Hz, 1H, 3'-OH), 3.48 (dd, *J$_1$*=3.8 Hz, *J$_2$*=10.7 Hz, 1H, H-5'), 3.57 (dd, *J$_1$*=3.3 Hz, *J$_2$*=10.8 Hz, 1H, H-5'), 3.61 (dd, *J$_1$*=7.3 Hz, *J$_2$*=8.3 Hz, 1H, H-2'), 3.81 (s, 6H, OCH$_3$), 3.88 (dt, *J$_1$*=3.6 Hz, *J$_2$*=6.1 Hz, 1H, H-4'), 4.30 (td, *J$_1$*=4.0 Hz, *J$_2$*=7.9 Hz, 1H, H-3'), 5.72 (dd, *J$_1$*=2.0, *J$_2$*=8.1 Hz, 1H, H-5), 6.41 (d, *J*=7.2 Hz, 1H, H-1'), 6.85-6.87 (m, 4H, Ph), 7.25-7.46 (m, 9H, Ph), 7.59 (d, *J*=8.2 1H, H-6), 8.95 (br, 1H, NH). ^{13}C-NMR (100 MHz, CDCl$_3$) δ (ppm): 6.6 (SeCH$_3$), 50.4 (C-2'), 55.3 (OCH$_3$), 61.8 (C-5'), 75.4 (C-3'), 82.5 (C-4'), 85.2 (Ar-C), 86.8 (C-1'), 102.5 (C-5), 113.3, 127.1, 128.0, 128.2, 130.1, 135.4, 135.5 (Ar-C), 139.8 (C-6), 144.4 (Ar-C), 150.4 (C-2), 158.7 (Ar-C), 163.7 (C-4). HRMS (ESI): C$_{31}$H$_{31}$N$_2$O$_7$Se; [M-H]$^-$: 623.1298 (calc. 623.1296).
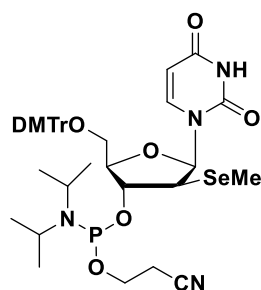
3'-O-(2-cyanoethyl-N,N-diisopropylphosphoramidite)-5'-O-(4,4'-dimethoxytrityl)-5-methyl-2'-methylseleno-arabinouridine (4.8) The starting material **4.7** (680 mg, 1.1 mmol) was dried under high vacuum and dissolved in 5 mL anhydrous CH$_2$Cl$_2$. To the solution, dimethylethylamine (0.7 mL, 6.6 mmol) was injected, followed by the addition of the 2-cyanoehtyl *N,N*-diisopropylchlorophosphoramidite (0.2 mL, 1.1 mmol). The reaction mixture was stirred at room temperature under dry argon for 1 h. Then, the solvent was evaporated under reduced pressure. The residue was re-dissolved in 2 mL CH$_2$Cl$_2$ and precipitated in 400 mL hexane under vigorous stirring. The hexane solution was then decanted carefully, and the crude product was further purified by silica gel chromatography (30% ethyl acetate in CH$_2$Cl$_2$ containing 1% dimethylethylamine). The product was precipitated in hexane (400 mL) again, and the precipitate

was re-dissolved in CH_2Cl_2 and evaporated to afford the desired pure product **4.8** (659 mg, 80%) as a white foam. HRMS (ESI): $C_{40}H_{49}N_4O_8PSe$; $[M+Na]^+$: 847.2376 (calc. 847.2351).

4.4.3   DNA oligo synthesis and purification

Solid-phase synthesis was used for All DNA oligonulceotides using an ABI3400 DNA/RNA synthesizer on a 1 µmol scale. 2'-MeSe-arabinouridine phosphoramidite was dried under high vacuum for 3 h and prepared as 0.1 M concentration solution in anhydrous acetonitrile prior of use. Normal phosphoramidites purchasing from Glenresearch were used for the synthesis of natural nucleosides. The coupling reaction was carried out using a tetrazole solution (0.45 M) in acetonitrile with a coupling time of 300 seconds. All the oligonucleotides were prepared in DMTr-On mode, followed by the cleavage from CPG solid support (beads) and the deprotection with concentrated ammonia solution under 55 °C overnight. The volatile ammonias were evaporated on a speed-vac concentrator, and the remaining beads were extracted with water (3 x 0.5 mL). The DNA oligonucleotides were analyzed and purified by reversed-phase high performance liquid chromatography (RP-HPLC), flow rate 6 ml/min [buffer A: 20mM triethylammonium acetate (TEAAc, pH 7.1) in water; buffer B: 20mM TEAAc (pH 7.1) in 50% acetonitrile]. The HPLC analysis was performed with a linear gradient from buffer A to 100% buffer B in 20 min. The detritylation was performed by treatment of the oligonucleotide with an acetic acid solution (pH 4.5) under 40 °C for 1 h. The DMTr-off oligonucleotides were desalted by a Water C-18 column and purified again by RP-HPLC. Native RNAs were purchased from Integrated DNA Technologies. The native, as well as the Se modified DNAs were characterized by matrix assisted laser desorption/ionization-time of fight mass spectrometry (MALDI-TOF MS).

4.4.4   UV-metling studies

The thermal denaturing experiments were performed on Cary 300 UV-Vis spectrometer equipped with a six-sample cell changer and a temperature controller. the sample annealed by heated to 85 $^{\circ}$C for 2 min and slowly cooling down to 4 $^{\circ}$C before data acquisition. Both denaturing and annealing curves were acquired at 260 nm with 1 cm path length at heating or cooling rates of 0.5 $^{\circ}$C/min and data interval of 0.5 $^{\circ}$C. The samples (2 μM DNA duplexes) were dissolved in the buffer of 350 mM NaCl, 10 mM NaH_2PO_4-Na_2HPO_4 (pH 6.8), 0.1 mM EDTA.

4.4.5   DNA Crystallization

The purified DNA oligonucleotide was heated to 80 $^{\circ}$C for 2 min, then cooled down slowly to room temperature. The Nucleic Acid Mini Screen Kit (Hampton Research) was applied to screen the crystallization conditions room temperature using the hanging-drop method by vapor diffusion. The concentrations of the native, Se modified DNAs were adjusted to 1.0 mM in water for crystallization. During the crystallization screening process, the DNA solution was mixed with crystallization buffer at 1:1 ratio (0.5 μL+ 0.5 μL) and equilibrate against 500 μL of the 35% MPD (2-methyl-2,4-pentanediol) solution within a hanging-drop plate. Crystals were formed within 2 weeks (Figure 4.8).
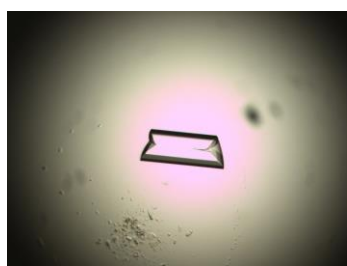
Figure 4.8 Crystal of ^{Se}U-DNA

4.4.6   Diffraction data collection and structure refinement

Diffraction data were collected on beamline 8.2.2 of the ALS (Advanced Light Source) at the Lawrence Berkeley National Laboratory. X-ray data were collected under a liquid-nitrogen

stream at 99 K. Each crystal was exposed for 0.5 second per image with 1° rotation and a total of 360 images were obtained. The data were integrated and scaled with the programs HKL2000 and DENZO/SCALEPACK.[81] The structures were solved by molecular replacement method using Phaser[82] within CCP4i[83, 84]. The resulting models were refined using REFMAC5.5[85] within CCP4i[83, 84]. The DNA, protein, ions and water molecules were added and modeled using Coot[86].

5 TELLURIUM MODIFIED OLIGONUCLEOTIDES SYNTHESIS FOR CHARGE TRANSPORT STUDY

5.1 Introduction

Beyond all doubt, nucleic acids are the most important biomolecules for all known forms of life, which storing all genetic information and passing it from one generation to the next. Extensive studies about nucleic acid reveals its structure, functions and properties[113-115] and stimulated the investigations of chemical modifications of nucleoside and nucleic acids which have been widely applied to lots of areas such as pharmaceutics[116, 117], crystallography[118, 119], asymmetric catalysis[120, 121] and material science[122]. With the remarkable structural stability, flexibility and its self-assembly properties, DNA has great potential as novel material in molecular electronics, biosensing and signaling devices. With the continuous development of single molecule break-junction technique, various attempts[123-125] have been made to manipulate, control, and modify a natural DNA for the request of both high conductivity and functional I-V characteristics.

As same group element of tellurium, selenium modifications was widely applied for protein structure determination, by replacing the sulfur in methionine, in which the selenium can be used as an ideal scattering centers for multiwavelength anomalous dispersion (MAD)[126-128]. The selenium-derivatized nucleic acids (SeNA) was also achieve great success in 3D crystal structure determination and selenium has been introduced to different positions of the ribose, the phosphate backbone as well as the nucleobases (Scheme 5.1)[118, 119, 129, 130]. It is worth to mention that the incorporation of 2'-selenium modified nucleoside into DNA oligo not only solved the phase problem but also greatly facilitated the crystallization, especially because, compared with protein crystallization, there are more challenge in nucleic acid crystallization due to the negatively

charged repetitive phosphate groups. However, compared with selenium, the application of tellurium in nucleic acid was still quite limited.
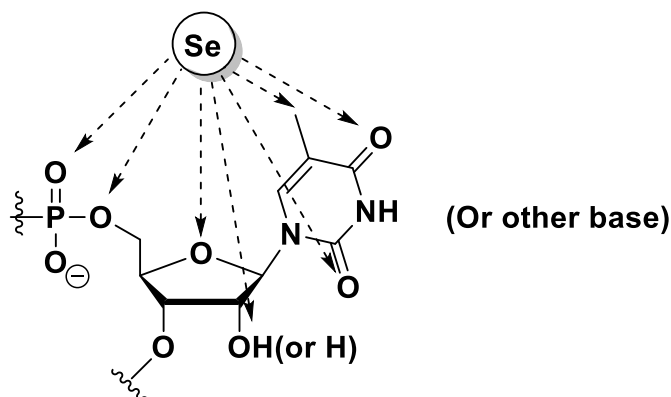
Scheme 5.1 Atom-specific selenium substitution of oxygen atoms in nucleic acids

5.1.1 Properties and applications of tellurium

Tellurium is a metalloid located in the group VI, also known as chalcogens, of the periodic table following sulphur and selenium. Its properties and functions have been well studied and shows multiple potential applications in different areas[131]. Tellurium has a larger radius (1.35 Å) and a weaker electronegativity (2.0) in the comparison to sulphur (1.04 Å, 2.58) and selenium (1.17Å, 2.55). The larger electrovalent and co-ordination sphere radius provide tellurium strong metallic properties and result in weak covalent bonds with carbon and hydrogen[132]. Naturally occurring tellurium containing a series of isotopes, including ^{120}Te (natural abundance 0.09%), ^{122}Te (2.55%), ^{123}Te (0.89%), ^{124}Te (4.74%), ^{125}Te (7.07%), ^{126}Te (18.84%), ^{128}Te (31.74%) and ^{130}Te (34.08%)[133], which result in a unique isotope pattern in mass spectrometry for Te-containing compounds. Meanwhile, the diamagnetic nucleus ^{125}Te (spin 1/2) enables Te-NMR studies and has wide chemical shifts ranging from -1400 ppm to 3400 ppm. Moreover, it also has excellent sensitivity due to the high natural abundance (7.07%) compared with ^{13}C (1.1%)[134].

5.1.2   Incorporation of tellurium into protein

In 1989, tellurium was successfully incorporated into protein in a tellurium-tolerant fungus which was achieved by growing the Te-resistant fungi on a sulfur-free medium and extraordinary high level of tellurium was detected[135]. Later, telluromethionine was reported to be selectively incorporated into dihydrofolate reductase[136]. Further studies optimized the bioincorporation technique of TeMet in to protein and provide a promising approach for X-ray structure study of protein[137]. The absorption edge of tellurium is about 0.3Å, which indicates that it is not as suitable as selenium (0.9795Å) as a scattering center in MAD experiment. But the ideal electronic and atomic properties of tellurium for generating clear isomorphous signals make it a suitable heavy-atom for isomorphous replacement without the need of synchrotron radiation[132].

5.1.3   Tellurium in nucleic acids

The incorporation of Te into nucleic acid was also achieved in the past ten years. The first 2'-Te modified nucleoside was reported by Huang et al. in 2008[138] and then successfully incorporated into DNA oligo through solid-phase synthesis[139]. Furthermore, the Te functionality was also incorporated to the 5-position of a pyrimidine and the crystal structure of modified Te-DNA was determined[140].
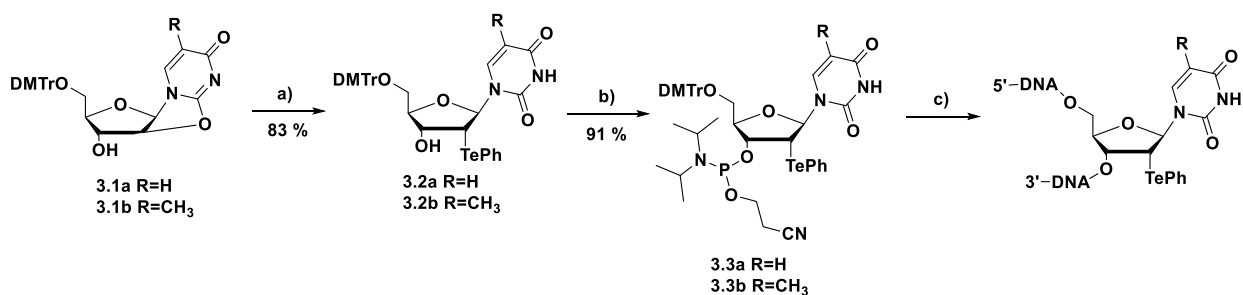
Scheme 5.2 Synthesis of 2'-phenyltelluro-nucleoside

5.1.3.1 2'-Te modification

The synthesis of 2'PhTe-modified was reported by Huang et al. (Scheme 5.3)[139]. TePh functionality was successfully introduced into 2'-position of both uridine and ribo-thymidine with

good yield. However, unlike the MeSe functionality[89, 141-145], the 1', 2'- and 2', 3'-eliminations was observed during Te functionalization when using sodium borohydride as reducing reagent at room temperature (Scheme 5.2).
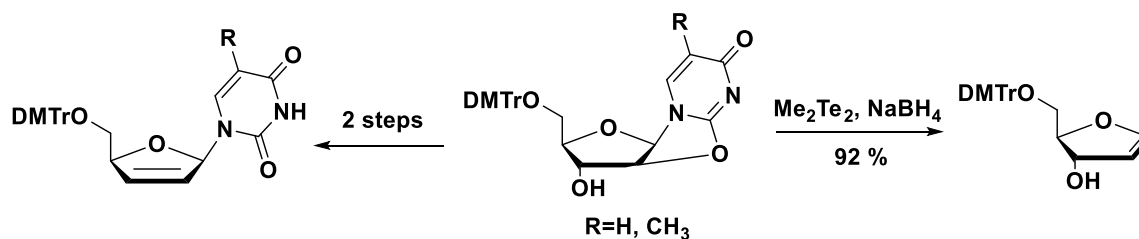
Scheme 5.3 Elimination reactions resulting from the 2'-TeMe functionality

To get the desired product X, a stronger reducing reagent and lower temperature (0 °C) was applied together with crown ether (12-*crown*-4) to chelate the lithium ions to enhance the MeTe reactivity. The desired product was obtained in 47 % yield with the 1', 2'-elimination products as the major byproduct. Both PhTe and MeTe modified nucleosides were incorporated into DNA oligos by solid-phase synthesis through standard protocol[88] and quantitative coupling yield was achieved. A few of Te-DNAs was oxidized to telluxides during the solid-phase synthesis which can be reduced by treating with diborane after the deprotection step (Scheme 5.3). It was found that both methyltelluride and phenyltelluride functionalities was stable with the treatment of mild acid and base during the deprotection and purification. Interestingly, under heating (50 °C) in the presence of B_2H_6 or I_2, 2'-TePh DNA undergoes 2', 3'-elimination at the modification site and generates the fragmented product. However, the 1', 2'-elimination was observed for 2'-TeMe DNA and creates the abasic product (Scheme 5.4).[139] The decrease of the melting temperature was observed during the UV melting study which probably caused by the perturbation introduced by the bulky Te functionality (Figure 5.1).
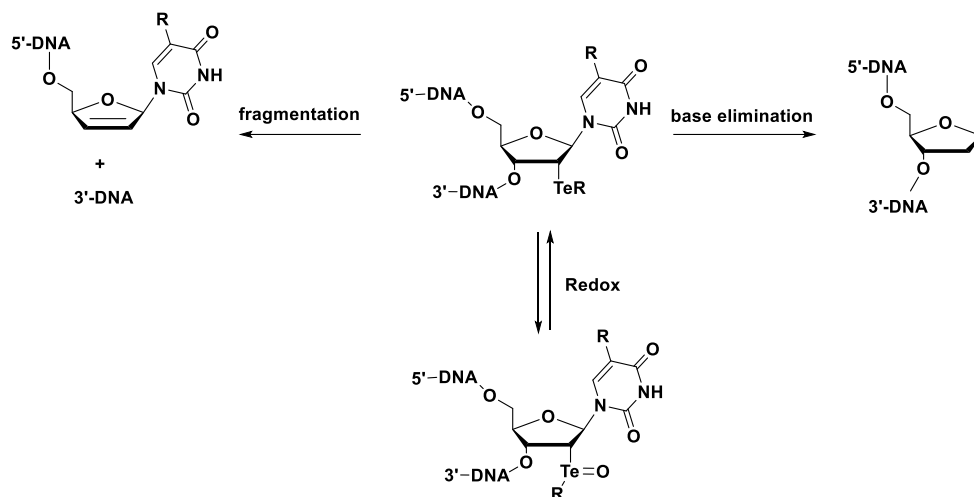
Scheme 5.4 Redox and fragmentation of 2'-Te nucleoside
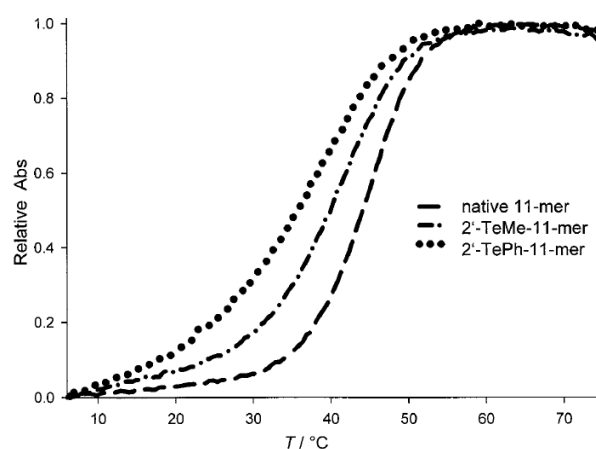
Figure 5.1 UV melting studies of the native and 2'Te-DNAs

5.1.3.2 5-Te modification

In 2011, 5-PhTe modified nucleoside was successfully synthesized by applying the lithium-halogen exchange reaction[146] on a protected 5-iodo-2'-deoxyuridine[140] and achieved medium yield (64 %) (Scheme 5.5). The key steps of the reaction are the deprotonation of the NH and the treatment with n-BuLi followed by the addition of Ph_2Te_2. Elevated concentration of the reactant (0.15 - 0.18 M) is necessary to avoid the generation of 6-PhTe isomer which is inseparable. The synthesis of the corresponding phosphoramidite was followed the standard protocol[140] and applied to solid-phase synthesis. The results show that the PhTe functionality is
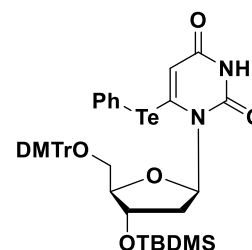
well compatible with the solid-phase synthesis condition, deprotection and purification. The UV-thermal denaturation studies show similar stability between the Te-derivatized duplex and the corresponding native which suggested that the bulky PhTe moiety is well accommodated and does not significantly change the duplex structure.
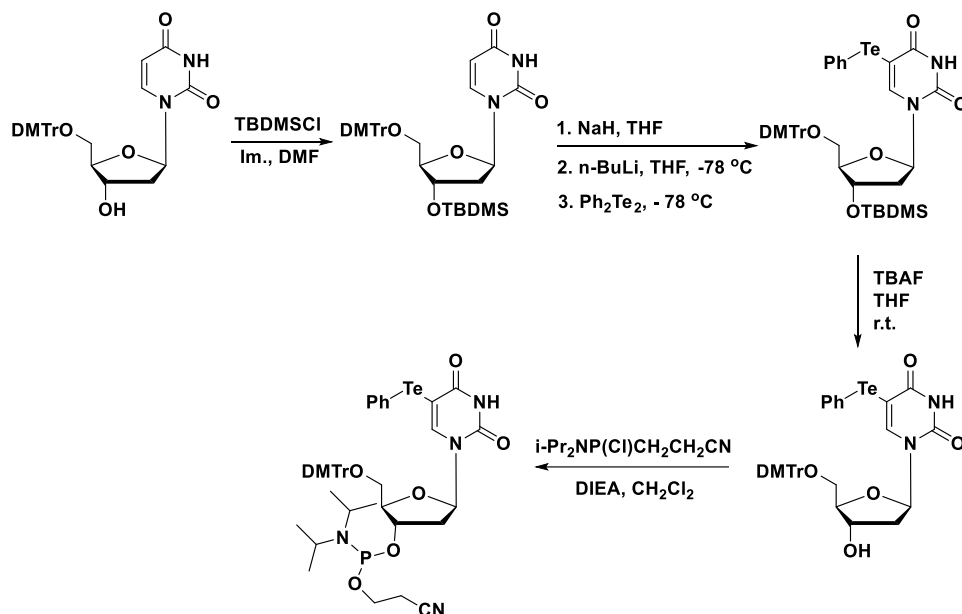
Scheme 5.5 Synthesis of 5-phenyltelluro-2'-deoxyuridine

The Te-DNA crystal structure was also obtained by the same author by using 2'-Se modification strategy[147-150]. The results revealing that Te-DNA has virtually identical global and local structures as the corresponding native DNA (Figure 5.3)[140]. This result further confirms that the Te-functionality does not cause significant perturbation. The Te-DNA was quite stable under high temperature (90 °C) but it was found to be sensitive to X-ray irradiation. Partial cleavage of Te-C bond was detected through MALDI-TOF-MS after X-ray irradiation. Due to the metallic property of the tellurium atom, STM imaging studies show stronger topographic and current peaks for Te-modified DNA duplex compared to the native one (Figure 5.4)[140].
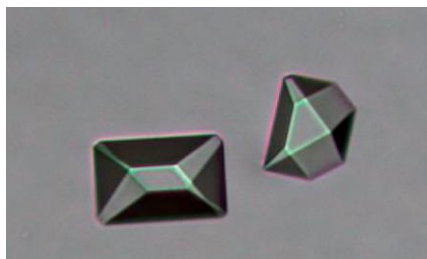
Figure 5.2 Crystal photo of the Te-DNA octomer [5'-G(2'-SeMe-dU)G(^{Te}T)ACAC-3]$_2$
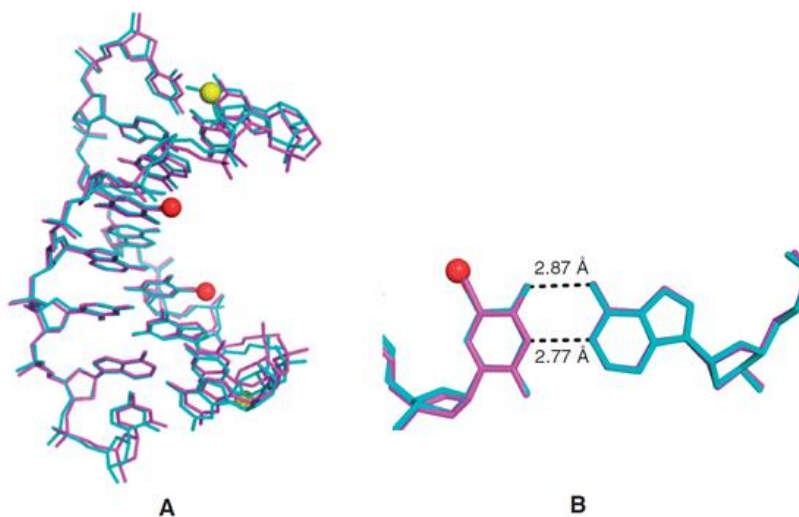
Figure 5.3 The global and local structures of the Te-DNA duplex
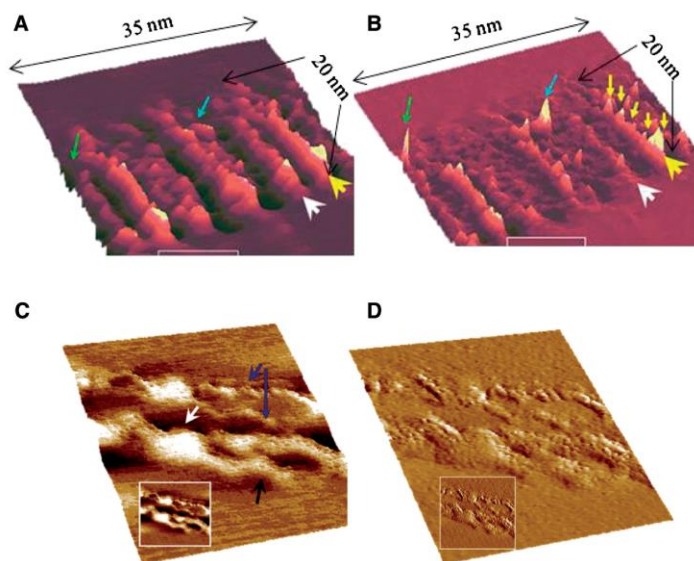
Figure 5.4 The STM images of the Te-modified DNA duplex

The Te-modified DNA and RNA are a promising strategy to investigate the structure and function of nucleic acid. However, studies about this area are still quite limited and only a few

papers published in last 10 years. 2'- and 5-position tellurium modified nucleoside has been successfully synthesized, and both are compatible with the solid-phase synthesis, deprotection and purification. The particular redox properties and selective elimination of the 2'-Te modified DNA oligo could be useful in studying the DNA fragmentation and nucleobase damage. The location of the Te functionality modification and the size of the protecting group directly affect the melting temperature of the duplex which could be used as a useful strategy for detecting DNA and RNA polymerization and catalysis. Furthermore, due to the metallic property of the tellurium atom, the Te-modified DNA duplex become visible under STM which suggesting a promising strategy to directly image DNA without structural perturbation. This will further help us accomplishing mechanism and function studies or even making novel nano-electronic materials.

Herein, we report the synthesis of 5-PhTe and 5-PhSe thymidine modified DNAs and the study of their charge transport properties. The conductance and current-voltage (I-V) characteristics of a series of Te-doped DNA molecules are measured with scanning tunneling microscope break junction technique (STM-BJ) in Au-DNA-Au junction systems. A comparison between Te-modified, Se-modified and the native DNAs show that the tellurium modification can effectively manipulate the electronic structure of DNA, and the location of the modification has a significant impact on the resulting I-V characteristics of DNA. This work provides new insights into charge transport through DNA molecules.

5.2   Results and discussion

5.2.1   5-PhTe/PhSe-thymidine phosphoramidite and oligo synthesis

N-DNA	5'-GTGTACAC-3'-SH HS-3'-CACATGTG-5'
^{Te}T-DNA 1	5'-GTGTACAC-3'-SH HS-3'-CACATGTG-5'
^{Te}T-DNA 2	5'-GTGTACAC-3'-SH HS-3'-CACATGTG-5'
^{Te}T-DNA 3	5'-GTGTACAC-3'-SH HS-3'-CACATGTG-5'
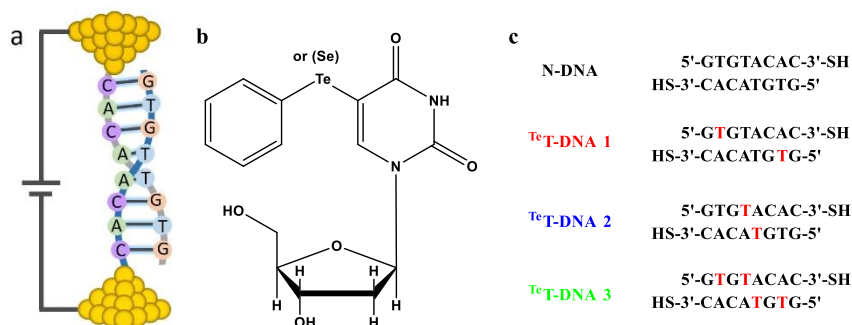
Figure 5.5 a. Au-DNA-Au molecular junction; b. chemical structure of Te and Se modified thymidine; c. Native as well as modified DNA sequences

In this work, a 5-PhTe-modified thymidine derivative (^{Te}T) were synthesized following the method reported[140] previously from our group with slight modification. Three corresponding Te-DNA duplexes (5'-GTGTACAC-3'-SH), ^{Te}T-DNA 1, ^{Te}T-DNA 2 and ^{Te}T-DNA 3, containing ^{Te}T bases at different locations were also synthesized (Figure 5.5 b and c). In the meanwhile, we also synthesized corresponding selenium nucleobase-derivatized nucleoside (Figure 5.6) as well as Se-DNA duplexes with same sequence for comparison. Note that one native DNA (N-DNA) of the same sequence was also synthesized for control experiment. To make it attachable to the Au(111) surface, all the DNAs containing a thiol tail at the 3'-end. By conducting the UV-thermal study, we found that the ^{Te}T-DNA duplex has the thermo-stability similar to the corresponding native duplex (Table 5.1). Circular dichroism was also used to study the structure change (Figure 5.11). Positive peaks at 280 nm, 220 nm and negative peak at 245 nm were observed in all oligos, and no shift was founded comparing to native DNA, which indicated that the native and Te-DNA has similar helixes. Furthermore, our previous X-ray crystal structure studies of the Te-derivatized DNAs indicate that the Te-derivatized and native structures are virtually identical, and that ^{Te}T and A interact as well as the native T–A pair.[140] These results suggest that the secondary structures, including the hydrogen bond, π-π stacking between neighboring base pairs and helicity of the

studied DNA duplexes are not perturbed by the tellurium modification. This allows isolating the influence of Te-modification on the charge transport properties of DNA for in-depth investigation.
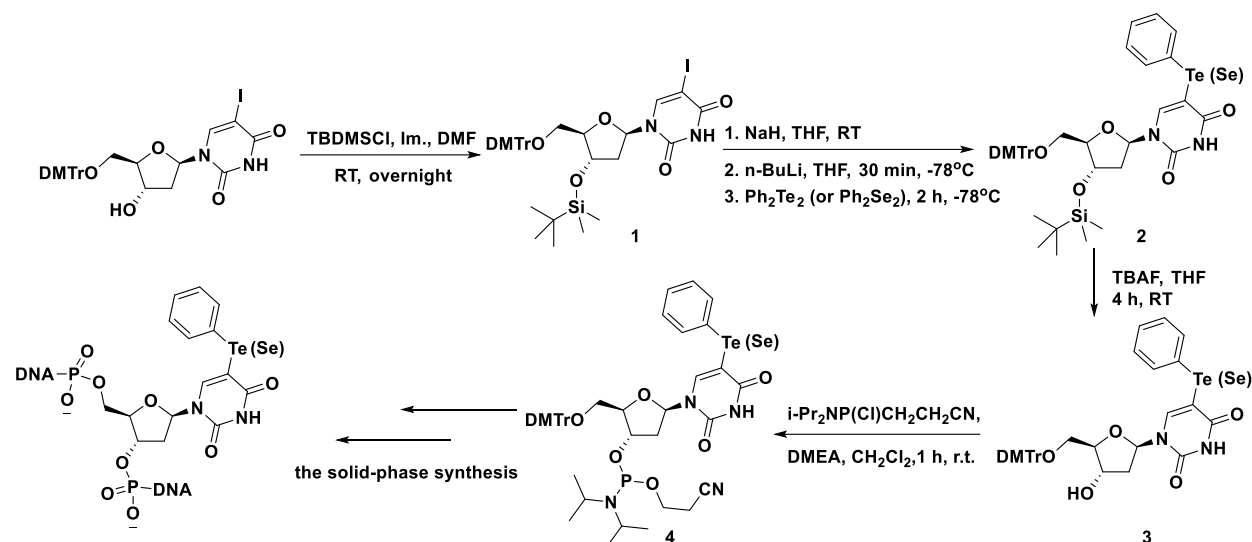
Scheme 5.6 The synthesis of 5-PhTe/Se-T phosphoramidite and DNAs

Table 5.1 UV melting study of self-complementary Te-DNA

Entry	DNA duplex	T_m (°C)
N-DNA	5'-GTGTACAC-3'	35.0 ± 0.2
[Te]T-DNA 1	5'-G([Te]T)GTACAC-3'	34.5 ± 0.2
[Te]T-DNA 2	5'-GTG([Te]T)ACAC-3'	34.2 ± 0.3
[Te]T-DNA 3	5'-G([Te]T)G([Te]T)ACAC-3'	33.1 ± 0.5

5.2.2 STM-BJ conductance and Current-voltage (I-V) characteristics measurements

Collaborating with Kun Wang and Binqian Xu at University of Georgia, we further studied the charge transportation properties of the DNAs containing [Se]T and [Te]T. By using scanning tunneling microscope break junction technique (STM-BJ), we measured the single-molecule conductance of these modified DNAs. An increase of single-molecule conductance was observed for both selenium and tellurium modified DNA comparing to the native (Figure 5.6). We also measured the single-molecule conductance for [Te]T-DNAs with the modification at different location under two different bias (0.3V and 0.75V). Under 0.3V, all [Te]T-DNAs show greater conductance than the native and follow the order of [Te]T-DNA 3 > [Te]T-DNA 2 > [Te]T-DNA 1 > N-

DNA (native). However, surprisingly, when the applied bias increased to 0.75V, the conductance difference follows the order of [Te]T-DNA 3 > [Te]T-DNA 1 > [Te]T-DNA 2 > N-DNA, caused by a dramatic conductance increase of [Te]T-DNA 1, triple its conductance, comparing to other [Te]T-DNAs which only have a slight increment.

We further performed the I-V characteristic measurements for both [Te]T-DNAs and [Se]T-DNA, as well as the corresponding native DNA. The I-V curves averaged from around 50 single-molecule I-V curves for each DNA molecule are shown in Fig. 5.7. Consistent with the static conductance measurement results, all [Te]T-DNAs yield higher current than the N-DNA. More importantly, the I-V curve of [Te]T-DNA 1 shows strong non-linear feature, which does not be observed for other [Te]T-DNAs (Figure 5.7). A sharp increase in current occurs for Te-DNA 1 when bias increases over 0.6V. A comparison between [Te]T-DNA 1 and [Se]T-DNA 1 reveals that the [Se]T-DNA 1 alse shows the non-linear feature, however, which is much weaker than the [Te]T-DNA 1 indicating the tellurium modification has a stronger effect on the charge transport properties of DNA than selenium.

5.3 Conclusion

We synthesized the 5-PhTe/Se-T phosphoramidite and DNAs containing these modifications. These DNAs containing a thiol tail which makes it capable to attach on the Au surface to form Au-DNA-Au junction system. Interestingly, with the tellurium modification on the first thymidine from the 5'-end, the DNA shows special charge transport properties in the conductance and I-V measurements. The position dependent, strong non-linear feature of the [Te]T-DNA give insight into charge transport properties of modified DNAs. Since the non-linear current-voltage characteristics are central to building functional electronic devices, this work could further spur future design of functional DNA-based molecular devices.

5.4 Experimental Section

5.4.1 General Section

Most solvents and reagents were purchased from Sigma, Fluka, or Aldrich and used without purification unless mentioned otherwise. When necessary, solid reagents were dried under high vacuum. Reactions with compounds sensitive to air or moisture were performed under argon. Solvent mixtures are indicated as volume/volume ratios. Thin layer chromatography (TLC) was run on Sorbtech Silica XHL TLC plates (0.25 mm thick; Rf values in the text are for the title products), and visualized under UV-light or by a Ce-Mo staining solution (phosphomolybdate, 25 g; $Ce(SO_4)_2 \cdot 4H_2O$, 10 g; H_2SO_4, 60 mL, conc.; H_2O, 940 mL) with heating. Flash chromatography was performed using Sorbtech Silica gel (mesh size 0.040-0.063 mm) using a silica gel:crude compound weight ratio of ca. 30:1. 1H-NMR and 13C-NMR spectra were recorded using a Bruker Avance 400 or 600. All chemical shifts (d) are in ppm relative to tetramethylsilane and all coupling constants (J) are in Hz. High resolution (HR) MS were either obtained with electrospray ionization (ESI) on a Q-TOF™ Waters Micromass at Georgia State University.

5.4.2 The synthesis of 5-PhTe-thymidine phosphoramidite
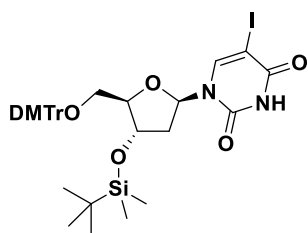
3'-O-tert-Butyldimethylsilyl-5'-O-(4,4'-dimethoxytriphenylmethyl)-2'-deoxy-5-iodouridine (5.1) was prepared with slight modification according to ref 1. 5'-O-(4,4'-dimethoxytrityl)-2'-deoxy-5-iodouridine (2 g, 3 mmol) and imidazole (0.61 g, 9 mmol) was placed in RBF and dried over high vacuum 20 min. Anhydrous DMF (20 ml) was injected under argon followed by the addition of tert-butyldimethylsilyl chloride (TBDMSCl) (0.69 g, 4.5 mmol). The

mixture was stirred at room temperature for 2 h, and then quenched with MeOH. The mixture was diluted with EtOAc (30 ml) and washed with H_2O (3×10 ml) and brine. The organic phase was dried over anhydrous $MgSO_4$. After filtration, the solvent was concentrated under reduced pressure and the residue was purified by a flash silica gel column (the silica gel was pre-equalized with 1% Et_3N in CH_2Cl_2, eluent: 20% EtOAc in hexanes) to give **5.1** (2.26 g, 98%). TLC condition: EA:hexanes = 1:2; rf = 0.5. ^{1}H-NMR (400 MHz, $CDCl_3$): δ 8.96 (1H, br, NH, exchangeable with D_2O), 8.23 (1H, s, H-6), 7.46–7.26 (9H, m, Ar), 6.89–6.87 (4H, m, Ar), 6.31 (1H, dd, H-1', J= 6, 7.2 Hz), 4.48 (1H, m, H-3'), 4.02 (1H, m, H-4'), 3.82 (6H, 2 s, CH_3O), 3.44 (1H, dd, H-5'a, J= 2.8, 10.8 Hz), 3.30 (1H, dd, H-5'b, J= 2.8, 10.8 Hz), 2.41 (1H, ddd, H-2'a, J= 2.8, 5.6, 13.2 Hz), 2.22 (1H, m, H-2'b), 0.86 (9 H, 3 s, $SiMe_3$), 0.05–0 (6H, 2s, $SiMe_2$); ^{13}C-NMR (100 MHz, $CDCl_3$) δ 160.01 (C4), 158.67 (Ar), 149.84 (C2), 144.38 (C-6), 144.30 (Ar), 135.48 (Ar), 135.37 (Ar), 130.11 (Ar), 130.05 (Ar), 128.12 (Ar), 128.05 (Ar), 127.09 (Ar), 113.39 (Ar), 113.36 (Ar), 87.38 (C4'), 86.99 (Ar), 85.79 (C-1'), 72.48 (C-3'), 68.42 (C-5), 63.01 (H5'), 55.30 (OMe), 42.05 (C2'), 25.74 (CCH3), 17.97 (CCH3), -4.67 (SiCH3), -4.86 (SiCH3); ^{13}C-HSQC, see Figure 1; UV (in MeOH), λ_{max}: 208 nm, 282 nm; HRMS (ESI-TOF): molecular formula: $C_{36}H_{42}N_2O_7Si$; [M-H]+: 769.1811 (calcd 769.1806). ^{1}H-^{13}C HSQC (Appendix D.4)
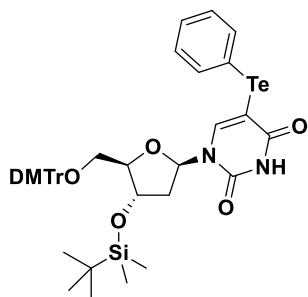
3'-O-tert-Butyldimethylsilyl-5'-O-(4,4-dimethoxytrityl)-2'-deoxy-5-phenyltellurouridine (5.2) was prepared with slight modification according to literature[140]. Two-neck bottle flask was used. The Ph_2Te_2 was packaged by a tissue bag and placed in the flask. The

bag was suspended above the solution during reaction. Under Argon, NaH 95 % (49 mg, 1.95 mmol) was placed in a two-neck round bottle flask. **5.1** (1 g, 1.3 mmol) was dissolve in dry THF (20 mL) and injected to the flask at room temperature. The Ph_2Te_2 was wrapped by a piece of tissue and suspended above the solution. The mixture was stirred for 30 min until no bubbling, then cool down to -78 °C (place in an Acetone-dye ice bath) and treated with n-BuLi (2.3 mL, 2.5 M solution in hexane; 5.9 mmol) dropwise over 10 min. The mixture was further stirred for 30 min, and then a needle was used to break the tissue and release all the Ph_2Te_2 in the bag. The mixture was further stirred for 2 h at -78 °C. Saturated NaCl solution (2 ml) was added to quench the reaction. The reaction flask was warmed up to room temperature. Water (10 ml) was added to the flask and EtOAc (3×20 ml) was used to extract the crude product. The organic phase was combined and dried over $MgSO_4$ (s). After filtration and evaporation, the residue was purified by flash silica gel chromatography (eluent: 40% EtOAc in hexanes containing 1% Et_3N) to give **5.2** (0.72 g, 65%) as a colorless foam. TLC condition: EA:hexanes = 1:2; rf = 0.4. ^{1}H-NMR (400 MHz, $CDCl_3$), δ: 8.54 (1H, br, NH, exchangeable with D_2O), 7.93 (1H, s, H-6), 7.67 (2H, d, Ar, J=6.8 Hz), 7.47 (2H, d, Ar, J=7.2 Hz), 7.39-7.23 (8H, m, Ar), 7.14 (2H, t, J=7.6 Hz), 6.88-6.85 (4H, m, Ar), 6.29 (1H, m, H-1'), 4.30 (1H, m, H-3'), 4.00 (1H, m, H-4'), 3.82 (6H, 2 s, CH_3O), 3.25 (2H, m, H-5'a,b), 2.37 (1H, ddd, H-2'a, J= 3.2, 9.2, 13.2 Hz), 2.09 (1H, m, H-2'b), 0.87 (9H, 3 s, $SiMe_3$), 0.04–0 (6H, 2s, $SiMe_2$); ^{13}C-NMR (100 MHz, $CDCl_3$) δ 162.75 (C4), 158.73 (Ar), 150.37 (C2), 147.15 (C6), 144.57 (Ar), 138.48 (Ar), 135.81 (Ar),135.61 (Ar), 130.24 (Ar), 130.20 (Ar), 128.57 (Ar), 128.36 (Ar), 128.23 (Ar), 128.11 (Ar), 127.13 (Ar), 113.37 (Ar), 113.22 (C5), 89.23 (Ar), 86.85 (C4'), 85.43 (C1'), 72.44 (C3'), 63.31 (C5'), 55.38 (OMe), 41.58 (C2'), 25.82 (CCH_3), 18.04 (CCH_3), -4.57 ($SiCH_3$), -4.77 ($SiCH_3$); ^{13}C-HSQC, see Figure 2; UV (in MeOH), λ_{max}: 204

nm, 266 nm; HRMS (ESI-TOF): molecular formula: $C_{42}H_{48}N_2O_7TeSi$; $[M+H^+]^+$: 851.2367 (calcd 851.2371). ^1H-^{13}C HSQC (Appendix D.4)
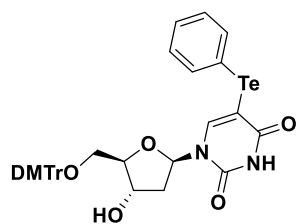
5'-O-(4,4-dimethoxytrityl-2'-deoxy-5-phenyltellurouridine (5.3) TBAF (1 ml, 1M solution in THF) was added to a solution of **5.2** (0.5 g, 0.7 mmol) in THF (10 ml) at 0 ^{o}C. The mixture was stirred for 3 h at room temperature. The solvent was evaporated under reduced pressure, and the residue was partitioned between EtOAc and H_2O. The organic phase was dried over anhydrous $MgSO_4$ (s) before evaporation. The residue was then purified by silica gel column chromatography (the silica gel was pre-equalized with 1% TEA in CH_2Cl_2; eluent: 4% MeOH in CH_2Cl_2 containing 1% TEA) to give (0.4 g, 83%) of **5.3** as pale-yellow foam. 1H-NMR (CDCl$_3$, d): 9.18 (1H, br, NH, exchangeable with D$_2$O), 7.83 (1H, s, H-6), 7.63–7.00 (15 H, m, Ar), 6.86–6.81 (4H, m, Ar), 6.26 (1H, dd, H-1', J=6.4, 6.8 Hz), 4.32 (1H, m, H-3'), 3.96 (1H, m, H-4'), 3.76 (6H, 2 s, CH$_3$O), 3.27 (1H, dd, H-5'a, J=4.4, 10.4 Hz), 3.18 (1H, dd, H-5'b, J=4.4, 10.8 Hz), 2.39 (1H, m, H-2'a), 2.15 (1H, m, H-2'b); ^1H-NMR spectrum is identical to the literature[140].
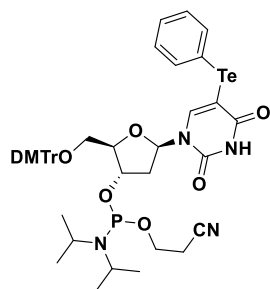
3'-O-(2-cyanoethyl-N,N-diisopropylamino)-5'-O-(4,4-dimethoxytrityl)-2'-deoxy-5-phenyltellurouridine (5.4). 5.3 (0.5 g, 0.68 mmol) was dried over high vacuum for 2 h and then dissolved in CH_2Cl_2 (5 mL) under argon. Dimethylethylamine (0.44 mL, 4.1 mmol) was injected, followed by the addition of N,N-diisopropylamino cyanoethylphosphamidic chloride (193 mg,

0.82 mmol). The mixture was stirred for 1 h at room temperature. The solvent was evaporated under reduced pressure and re-dissolved in minimal CH_2Cl_2. Slowly drop the solution into pentane (500 mL) under vigorous stirring. The white precipitate was filtered out and purified by a short silica gel column (eluent: 2 % MeOH and 1 % DMEA in DCM). The crude product was dissolved in minimal CH_2Cl_2 and re-precipitated in pentane. The precipitate was filtered out again and dried under high vacuum to give **5.4** (350 mg, 55%) as a mixture of two diastereomers. ^{31}P-NMR (CDCl3, d): 149.3, 149.8; HRMS (ESI-TOF): molecular formula: $C_{45}H_{51}N_4O_8PTe$; $[M-H^+]$: 935.252 (calcd 935.242)

5.4.3 *The synthesis of 5-PhSe-thymidine phosphoramidite*
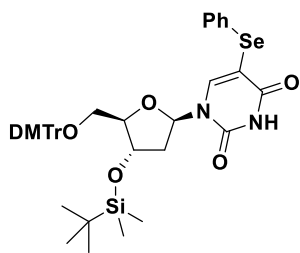
3'-O-tert-Butyldimethylsilyl-5'-O-(4,4-dimethoxytrityl)-2'-deoxy-5-phenylselenouridine (5.5). Two-neck bottle flask was used. The Ph_2Se_2 was packaged by a tissue bag and placed in the flask. The bag was suspended above the solution during reaction. Under Argon, NaH 95 % (49 mg, 1.95 mmol) was placed in a two-neck round bottle flask. **5.1** (1 g, 1.3 mmol) was dissolve in dry THF (20 mL) and injected to the flask at room temperature. The Ph_2Se_2 (1.4 g, 4.5 mmol) was wrapped by a piece of tissue and suspended above the solution. The mixture was stirred for 30 min until no bubbling, then cool down to -78 °C (place in an Acetone-dye ice bath) and treated with n-BuLi (2 mL, 2.5 M solution in hexane; 5.2 mmol) dropwise over 10 min. The mixture was further stirred for 30 min, and then a needle was used to break the tissue and release all the Ph_2Se_2 in the bag. The mixture was further stirred for 2 h at -78 °C. Saturated NaCl solution (2 ml) was added to quench the reaction. The reaction flask was warmed up to room

temperature. Water (10 ml) was added to the flask and EtOAc (3×20 ml) was used to extract the

crude product. The organic phase was combined and dried over MgSO$_4$ (s). After filtration and

evaporation, the crude product was moved to next step without further purification.
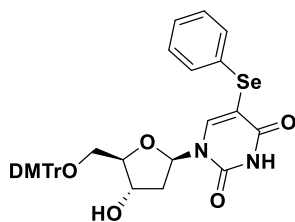
5'-O-(4,4-dimethoxytrityl-2'-deoxy-5-phenyltellurouridine (5.6) TBAF (1 ml, 1M

solution in THF) was added to a solution of **5.5** (0.6 g, 0.7 mmol) in THF (10 ml) at 0 ºC. The

mixture was stirred for 3 h at room temperature. The solvent was evaporated under reduced

pressure, and the residue was partitioned between EtOAc and H$_2$O. The organic phase was dried

over anhydrous MgSO$_4$ (s) before evaporation. The residue was then purified by silica gel column

chromatography (the silica gel was pre-equalized with 1% TEA in CH$_2$Cl$_2$; eluent: 4% MeOH in

CH$_2$Cl$_2$ containing 1% TEA) to give (0.4 g, 83%) of **5.6** as pale-yellow foam. 1H-NMR (400 MHz,

CDCl$_3$) δ (ppm): 8.07 (1H, s, H-6), 7.43–7.20 (14 H, m, Ar), 6.86 (1H, d, H-1', *J*=8.7 Hz), 6.81

(4H, dd, Ar, *J*=2.5, 8.9 Hz), 4.14 (1H, q, H-3', *J*=7.2), 4.05 (1H, m, H-4'), 3.78 (6H, s, CH$_3$O),

3.41 (1H, dd, H-5'a, *J*=3.2, 10.7 Hz), 3.33 (1H, dd, H-5'b, *J*=3.7, 10.1 Hz), 2.49 (1H, m, H-2'a),

2.29 (1H, m, H-2'b);
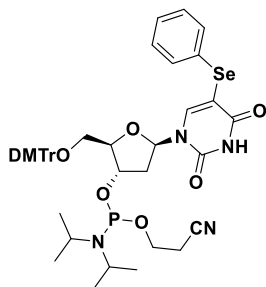
3'-O-(2-cyanoethyl-N,N-diisopropylamino)-5'-O-(4,4-dimethoxytrityl)-2'-deoxy-5-

phenyltellurouridine (5.7). 5.6 (0.5 g, 0.68 mmol) was dried over high vacuum for 2 h and then

dissolved in CH$_2$Cl$_2$ (5 mL) under argon. Dimethylethylamine (0.44 mL, 4.1 mmol) was injected,

followed by the addition of N,N-diisopropylamino cyanoethylphosphamidic chloride (193 mg, 0.82 mmol). The mixture was stirred for 1 h at room temperature. The solvent was evaporated under reduced pressure and re-dissolved in minimal CH_2Cl_2. Slowly drop the solution into pentane (500 mL) under vigorous stirring. The white precipitate was filtered out and purified by a short silica gel column (eluent: 2 % MeOH and 1 % DMEA in DCM). The crude product was dissolved in minimal CH_2Cl_2 and re-precipitated in hexane. The precipitate was filtered out again and dried under high vacuum to give **5.7** (350 mg, 55%) as a mixture of two diastereomers. HRMS (ESI-TOF): molecular formula: $C_{45}H_{51}N_4O_8PSe$; $[M+H^+]$: 887.2635 (calcd 887.2690)

5.4.4   5-PhTe/PhSe-Uridine containing DNA synthesis and purification

All the DNA oligonucleotides were chemically synthesized in a 1.0 mmol scale using an ABI3400 DNA/RNA Synthesizer. The ultra-mild nucleoside phosphoramidite reagents were used in this work (Glen Research). The concentration of the Te-deoxy-uridine phosphoramidite was identical to that of the conventional ones (0.1 M in acetonitrile). Coupling was carried out using a 5-ethyl thio-1H-tetrazole (ETT) solution (0.25 M) in acetonitrile. The coupling time was 25 s for both native and 300 s for modified samples. Trichloroacetic acid (3%) in methylene chloride was used for the 5'-detritylation. Synthesis were performed on control pore glass (3'-thiol-modifier C3 S-S CPG, see Figure 5.8) immobilized with the appropriate nucleoside through a succinate linker. The modified CPG containing a disulfide bond which can be cleaved at room temperature in 30 min with 100 mM DTT pH 8.3 – 8.5 in the buffer after purification. All the oligonucleotides were prepared with DMTr-on form. After synthesis, the DNA oligonucleotides were cleaved from the solid support and fully deprotected by the treatment of 0.05 M K_2CO_3 solution in methanol for 8 h at room temperature. The 5'-DMTr deprotection was performed in a 10% acetic acid solution for 1 h at 40 °C, followed by neutralization to pH 7.0 with a freshly made aqueous solution of

triethylamine (1.1 M) and desalt with a waters certified Sep-Pak® cartridges to remove DMTr-OH.

The typical MS results of the 5-Te-T-containing DNAs are presented in Figure 5.9 and Table 5.3.
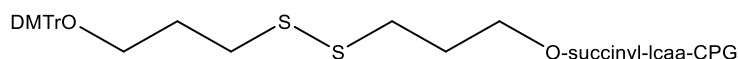
Figure 5.6 Structure of 3'-Thiol-Modifier C3 S-S CPG
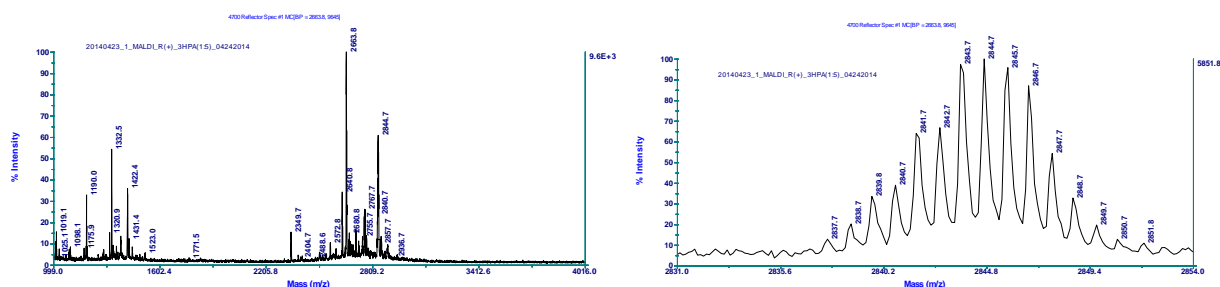
Figure 5.7 MALDI-TOF MS spectrum of 5'-G(^{Te}T)GTACAC-3'.
Molecular formula: $C_{89}H_{114}N_{30}O_{50}P_8S_2Te$; [M]$^-$: 2844.7 (calc. 2844.3)

Table 5.2 MS values of the 5-Te-T-containing DNAs

Entry	DNA sequences[a]	Measured (calc.) m/z
1	GTGTACAC $C_{84}H_{112}N_{30}O_{50}P_8S_2$	[M+H]$^+$: 2653.8 (2653.4)
2	G(^{Te}T)GTACAC $C_{89}H_{114}N_{30}O_{50}P_8S_2Te$	[M]$^-$: 2844.7 (2844.3)
3	GTG(^{Te}T)ACAC $C_{89}H_{114}N_{30}O_{50}P_8S_2Te$	[M+H]$^+$: 2845.6 (2845.3)
4	G(^{Te}T)G(^{Te}T)ACAC $C_{94}H_{116}N_{30}O_{50}P_8S_2Te_2$	[M+H]$^+$: 3037.5 (3037.2)

a. 3'-end of all DNA sequences: ;

5.4.5 *UV-melting temperature and Circular dichroism experiements*

The experiments were performed using the samples (2 µM DNA duplexes) dissolved in the

buffer of 150 mM NaCl, 10 mM NaH_2PO_4-Na_2HPO_4 (pH 6.5), 0.1 mM EDTA and 10 mM $MgCl_2$.

The samples were heated up to 60 ℃ and allowed to cool down to room temperature slowly. These

experiments were carried out by Cary 300 UV-Visible Spectrophotometer with a temperature

controller at a heating rate of 0.5 ℃/min.
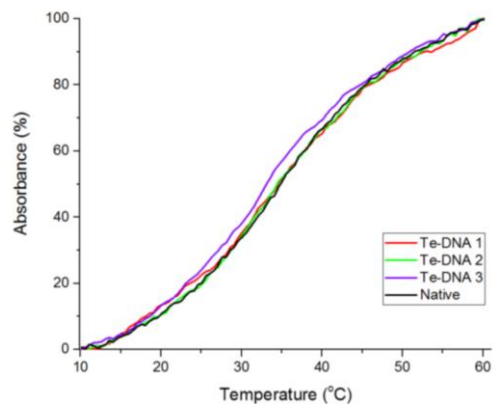
Figure 5.8 Normalized UV-melting curves of the native and corresponding Te-modified DNA.

The CD spectra were measured by using a Jasco J-810 Spectropolarimeter (JASCO, Japan) with a 0.1-cm path-length quartz cell at 25 ºC. The experiments were performed using samples (15 µM DNA duplexes) dissolved in the buffer of 25 mM Tris-HCl at pH 7.3 with 50 mM NaCl at 25 ºC. The CD spectrum was obtained by taking the average of three scans made at 0.1-nm intervals from 200 to 350 nm.
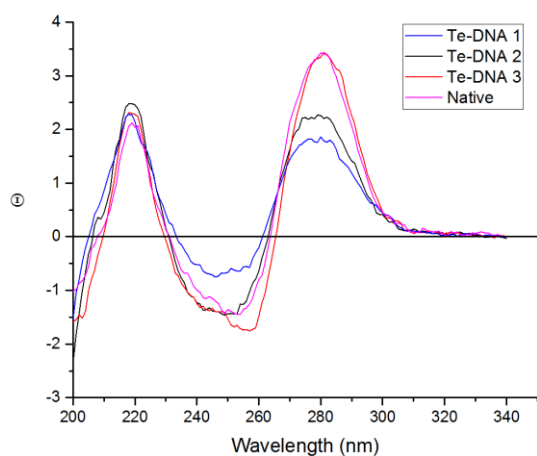
Figure 5.9 CD spectra of the native and Te modified DNA

REFERENCES

1.   Opalinska, J. B.; Gewirtz, A. M., Nucleic-acid therapeutics: basic principles and recent applications. *Nat. Rev. Drug Discov.* **2002,** *1* (7), 503-514.

2.   Palchaudhuri, R.; Hergenrother, P. J., DNA as a target for anticancer compounds: methods to determine the mode of binding and the mechanism of action. *Curr. Opin. Biotechnol.* **2007,** *18* (6), 497-503.

3.   Ding, S.; Qiao, X.; Kucera, G. L.; Bierbach, U., Using a build-and-click approach for producing structural and functional diversity in DNA-targeted hybrid anticancer agents. *J. Med. Chem.* **2012,** *55* (22), 10198-10203.

4.   Ueberschaar, N.; Dahse, H. M.; Bretschneider, T.; Hertweck, C., Rational design of an apoptosis-inducing photoreactive DNA intercalator. *Angew. Chem. Int. Ed. Engl.* **2013,** *52* (24), 6185-6189.

5.   Wang, S.; Kumar, A.; Aston, K.; Nguyen, B.; Bashkin, J. K.; Boykin, D. W.; Wilson, W. D., Different thermodynamic signatures for DNA minor groove binding with changes in salt concentration and temperature. *Chem. Commun. (Camb.)* **2013,** *49*, 8543-8545.

6.   Chenoweth, D. M.; Meier, J. L.; Dervan, P. B., Pyrrole-imidazole polyamides distinguish between double-helical DNA and RNA. *Angew. Chem. Int. Ed. Engl.* **2013,** *52*, 415-418.

7.   Ranjan, N.; Davis, E.; Xue, L.; Arya, D. P., Dual recognition of the human telomeric G-quadruplex by a neomycin-anthraquinone conjugate. *Chem. Commun. (Camb.)* **2013,** *49*, 5796-5798.

8.   Minarini, A.; Milelli, A.; Tumiatti, V.; Rosini, M.; Lenzi, M.; Ferruzzi, L.; Turrini, E.; Hrelia, P.; Sestili, P.; Calcabrini, C.; Fimognari, C., Exploiting RNA as a new biomolecular target for synthetic polyamines. *Gene* **2013,** *524*, 232-40.

9.   Sassolas, A.; Leca-Bouvier, B. D.; Blum, L. J., DNA biosensors and microarrays. *Chem. Rev.* **2008,** *108* (1), 109-139.

10.   Xue, X.; Wang, F.; Liu, X., One-step, room temperature, colorimetric detection of mercury (Hg2+) using DNA/nanoparticle conjugates. *J. Am. Chem. Soc.* **2008,** *130* (11), 3244-3245.

11.   Lan, T.; Furuya, K.; Lu, Y., A highly selective lead sensor based on a classic lead DNAzyme. *Chem. Commun. (Camb.)* **2010,** *46* (22), 3896-3898.

12.   Zhang, W.; Huang, Z., Synthesis of the 5′-se-thymidine phosphoramidite and convenient labeling of DNA oligonucleotide. *Org. Lett.* **2011,** *13* (8), 2000-2003.

13.   Yang, Z.; Chen, F.; Alvarado, J. B.; Benner, S. A., Amplification, mutation, and sequencing of a six-letter synthetic genetic system. *J. Am. Chem. Soc.* **2011,** *133* (38), 15105-15112.

14.   Cristovao, M.; Sisamakis, E.; Hingorani, M. M.; Marx, A. D.; Jung, C. P.; Rothwell, P. J.; Seidel, C. A.; Friedhoff, P., Single-molecule multiparameter fluorescence spectroscopy reveals directional MutS binding to mismatched bases in DNA. *Nucleic Acids Res.* **2012,** *40* (12), 5448-5464.

15.   Sha, R.; Birktoft, J. J.; Nguyen, N.; Chandrasekaran, A. R.; Zheng, J.; Zhao, X.; Mao, C.; Seeman, N. C., Self-assembled DNA crystals: the impact on resolution of 5'-phosphates and the DNA source. *Nano Lett.* **2013,** *13* (2), 793-797.

16.   Han, D.; Jiang, S.; Samanta, A.; Liu, Y.; Yan, H., Unidirectional scaffold-strand arrangement in DNA origami. *Angew. Chem. Int. Ed. Engl.* **2013,** *52*, 9031-9034.

17.   Zhang, J.; Ferre-D'Amare, A. R., Co-crystal structure of a T-box riboswitch stem I domain in complex with its cognate tRNA. *Nature* **2013,** *500* (7462), 363-6.

18.   Ke, A.; Zhou, K.; Ding, F.; Cate, J. H.; Doudna, J. A., A conformational switch controls hepatitis delta virus ribozyme catalysis. *Nature* **2004,** *429* (6988), 201-5.

19.   Doherty, E. A.; Doudna, J. A., Ribozyme structures and mechanisms. *Annu. Rev. Biophys. Biomol. Struct.* **2001,** *30*, 457-475.

20.   Frank-Kamenetskii, M. D.; Mirkin, S. M., Triplex DNA structures. *Annu. Rev. Biochem.* **1995,** *64* (1), 65-95.

21.   Davis, J. T., G‐Quartets 40 Years Later: From 5′‐GMP to Molecular Biology and Supramolecular Chemistry. *Angew. Chem. Int. Ed. Engl.* **2004,** *43* (6), 668-698.

22.   Lim, K. W.; Phan, A. T., Structural Basis of DNA Quadruplex-Duplex Junction Formation. *Angew. Chem. Int. Ed. Engl.* **2013,** *52*, 8566-8569.

23.   Ho, P. S.; Eichman, B. F., The crystal structures of DNA Holliday junctions. *Curr. Opin. Struct. Biol.* **2001,** *11* (3), 302-308.

24.   Phan, A. T.; Mergny, J. L., Human telomeric DNA: G‐quadruplex, i‐motif and Watson–Crick double helix. *Nucleic Acids Res.* **2002,** *30* (21), 4618-4625.

25.   Lin, L.; Sheng, J.; Huang, Z., Nucleic acid X-ray crystallography via direct selenium derivatization. *Chemical Society Reviews* **2011,** *40* (9), 4591-4602.

26.   Pallan, P. S.; Greene, E. M.; Jicman, P. A.; Pandey, R. K.; Manoharan, M.; Rozners, E.; Egli, M., Unexpected origins of the enhanced pairing affinity of 2′-fluoro-modified RNA. *Nucleic acids research* **2011,** *39* (8), 3482-3495.

27.   Sun, H.; Sheng, J.; Hassan, A. E.; Jiang, S.; Gan, J.; Huang, Z., Novel RNA base pair with higher specificity using single selenium atom. *Nucleic acids research* **2012,** *40* (11), 5171-5179.

28.   Hassan, A. E.; Sheng, J.; Zhang, W.; Huang, Z., High fidelity of base pairing by 2-selenothymidine in DNA. *Journal of the American Chemical Society* **2010,** *132* (7), 2120-2121.

29.   Lin, L.; Sheng, J.; Huang, Z., Nucleic acid X-ray crystallography via direct Selenium derivatization. *Chem. Soc. Rev.* **2011,** *40* (9), 4591-4602.

30.   Carrasco, N.; Ginsburg, D.; Du, Q.; Huang, Z., Synthesis of selenium-derivatized nucleosides and oligonucleotides for X-ray crystallography. *Nucleosides Nucleotides Nucleic Acids* **2001,** *20* (9), 1723-1734.

31.   Du, Q.; Carrasco, N.; Teplova, M.; Wilds, C. J.; Egli, M.; Huang, Z., Internal derivatization of oligonucleotides with selenium for X-ray crystallography using MAD. *J. Am. Chem. Soc.* **2002,** *124* (1), 24-25.

32.   Wilds, C. J.; Pattanayek, R.; Pan, C.; Wawrzak, Z.; Egli, M., Selenium-assisted nucleic acid crystallography: use of phosphoroselenoates for MAD phasing of a DNA structure. *J. Am. Chem. Soc.* **2002,** *124* (50), 14910-14916.

33.   Hobartner, C.; Micura, R., Chemical synthesis of selenium-modified oligoribonucleotides and their enzymatic ligation leading to an U6 SnRNA stem-loop segment. *J. Am. Chem. Soc.* **2004,** *126* (4), 1141-1149.

34.   Campbell, N. H.; Parkinson, G. N., Crystallographic studies of quadruplex nucleic acids. *Methods* **2007,** *43* (4), 252-263.

35.   Hendrickson, W. A., Determination of macromolecular structures from anomalous diffraction of synchrotron radiation. *Science* **1991,** *254* (5028), 51-58.

36.   Olieric, V.; Rieder, U.; Lang, K.; Serganov, A.; Schulze-Briese, C.; Micura, R.; Dumas, P.; Ennifar, E., A fast selenium derivatization strategy for crystallization and phasing of RNA structures. *Rna* **2009,** *15* (4), 707-715.

37.   Teplova, M.; Wilds, C. J.; Wawrzak, Z.; Tereshko, V.; Du, Q.; Carrasco, N.; Huang, Z.; Egli, M., Covalent incorporation of selenium into oligonucleotides for X-ray crystal structure determination via MAD: proof of principle. *Biochimie* **2002,** *84* (9), 849-858.

38.   Hassan, A. E.; Sheng, J.; Jiang, J.; Zhang, W.; Huang, Z., Synthesis and crystallographic analysis of 5-Se-thymidine DNAs. *Organic letters* **2009,** *11* (12), 2503-2506.

39.   Jiang, J.; Sheng, J.; Carrasco, N.; Huang, Z., Selenium derivatization of nucleic acids for crystallography. *Nucleic acids research* **2006,** *35* (2), 477-485.

40.   Moroder, H.; Kreutz, C.; Lang, K.; Serganov, A.; Micura, R., Synthesis, oxidation behavior, crystallization and structure of 2 '-methylseleno guanosine containing RNAs. *Journal of the American Chemical Society* **2006,** *128* (30), 9909-9918.

41.   Salon, J.; Sheng, J.; Gan, J.; Huang, Z., Synthesis and crystal structure of 2′-Se-modified guanosine containing DNA. *The Journal of organic chemistry* **2010,** *75* (3), 637-641.

42.   Salon, J.; Sheng, J.; Jiang, J.; Chen, G.; Caton-Williams, J.; Huang, Z., Oxygen replacement with selenium at the thymidine 4-position for the Se base pairing and crystal structure studies. *Journal of the American Chemical Society* **2007,** *129* (16), 4862-4863.

43.   Sheng, J.; Jiang, J.; Salon, J.; Huang, Z., Synthesis of a 2 '-Se-thymidine Phosphoramidite and Its Incorporation into Oligonucleotides for Crystal Structure Study. *Organic letters* **2007,** *9* (5), 749-752.

44.   Sheng, J.; Salon, J.; Gan, J.; Huang, Z., Synthesis and crystal structure study of 2′-Se-adenosine-derivatized DNA. *Science China Chemistry* **2010,** *53* (1), 78-85.

45.   Wilds, C. J.; Pattanayek, R.; Pan, C.; Wawrzak, Z.; Egli, M., Selenium-assisted nucleic acid crystallography: use of phosphoroselenoates for MAD phasing of a DNA structure. *Journal of the American Chemical Society* **2002,** *124* (50), 14910-14916.

46.   Salon, J.; Gan, J.; Abdur, R.; Liu, H.; Huang, Z., Synthesis of 6-Se-guanosine RNAs for structural study. *Organic letters* **2013,** *15* (15), 3934-3937.

47.   Abdur, R.; Gerlits, O. O.; Gan, J.; Jiang, J.; Salon, J.; Kovalevsky, A. Y.; Chumanevich, A. A.; Weber, I. T.; Huang, Z., Novel complex MAD phasing and RNase H structural insights using selenium oligonucleotides. *Acta Crystallographica Section D: Biological Crystallography* **2014,** *70* (2), 354-361.

48.   Thompson, R. A.; Spring, A. M.; Sheng, J.; Huang, Z.; Germann, M. W., The importance of fitting in: conformational preference of selenium 2′ modifications in nucleosides and helical structures. *Journal of Biomolecular Structure and Dynamics* **2015,** *33* (2), 289-297.

49.   Lin, L.; Caton-Williams, J.; Kaur, M.; Patino, A. M.; Sheng, J.; Punetha, J.; Huang, Z., Facile synthesis of nucleoside 5′-(α-P-seleno)-triphosphates and phosphoroselenoate RNA transcription. *Rna* **2011,** *17* (10), 1932-1938.

50.   Stein, C. A.; Castanotto, D., FDA-approved oligonucleotide therapies in 2017. *Molecular Therapy* **2017,** *25* (5), 1069-1075.

51.   Sharma, V. K.; Sharma, R. K.; Singh, S. K., Antisense oligonucleotides: modifications and clinical trials. *MedChemComm* **2014,** *5* (10), 1454-1471.

52.   Carrasco, N.; Huang, Z., Enzymatic Synthesis of Phosphoroselenoate DNA Using Thymidine 5 '-(α-P-seleno) triphosphate and DNA Polymerase for X-ray Crystallography via MAD. *Journal of the American Chemical Society* **2004,** *126* (2), 448-449.

53.   Brandt, G.; Carrasco, N.; Huang, Z., Efficient substrate cleavage catalyzed by hammerhead ribozymes derivatized with selenium for X-ray crystallography. *Biochemistry* **2006,** *45* (29), 8972-8977.

54.   Lin, L.; Sheng, J.; Momin, R. K.; Du, Q.; Huang, Z., Facile synthesis and anti-tumor cell activity of Se-containing nucleosides. *Nucleosides, Nucleotides and Nucleic Acids* **2009,** *28* (1), 56-66.

55.   Helm, M.; Petermeier, M.; Ge, B.; Fiammengo, R.; Jäschke, A., Allosterically Activated Diels−Alder Catalysis by a Ribozyme. *Journal of the American Chemical Society* **2005,** *127* (30), 10492-10493.

56.   Serganov, A.; Keiper, S.; Malinina, L.; Tereshko, V.; Skripkin, E.; Höbartner, C.; Polonskaia, A.; Phan, A. T.; Wombacher, R.; Micura, R., Structural basis for Diels-Alder ribozyme-catalyzed carbon-carbon bond formation. *Nature Structural and Molecular Biology* **2005,** *12* (3), 218.

57.   Egli, M.; Pallan, P. S.; Pattanayek, R.; Wilds, C. J.; Lubini, P.; Minasov, G.; Dobler, M.; Leumann, C. J.; Eschenmoser, A., Crystal structure of homo-DNA and nature's choice of pentose over hexose in the genetic system. *Journal of the American Chemical Society* **2006,** *128* (33), 10847-10856.

58.   Salon, J.; Jiang, J.; Sheng, J.; Gerlits, O. O.; Huang, Z., Derivatization of DNAs with selenium at 6-position of guanine for function and crystal structure studies. *Nucleic acids research* **2008,** *36* (22), 7009-7018.

59.   Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E., The Protein Data Bank. *Nucleic Acids Research* **2000,** *28* (1), 235-242.

60.   Berman, H. M.; Kleywegt, G. J.; Nakamura, H.; Markley, J. L., The future of the protein data bank. *Biopolymers* **2013,** *99* (3), 218-222.

61.   Elkayam, E.; Kuhn, C.-D.; Tocilj, A.; Haase, A. D.; Greene, E. M.; Hannon, G. J.; Joshua-Tor, L., The structure of human argonaute-2 in complex with miR-20a. *Cell* **2012,** *150* (1), 100-110.

62.   Hendrickson, W. A., Synchrotron crystallography. *Trends in biochemical sciences* **2000,** *25* (12), 637-643.

63.   Nguyen, P. T.; Lai, J. Y.; Lee, A. T.; Kaiser, J. T.; Rees, D. C., Noncanonical role for the binding protein in substrate uptake by the MetNI methionine ATP Binding Cassette (ABC) transporter. *Proceedings of the National Academy of Sciences* **2018,** *115* (45), E10596-E10604.

64.   Carrasco, N.; Ginsburg, D.; Du, Q.; Huang, Z., Synthesis of selenium-derivatized nucleosides and oligonucleotides for X-ray crystallography. *Nucleosides, Nucleotides and Nucleic Acids* **2001,** *20* (9), 1723-1734.

65.   Zhang, W.; Szostak, J. W.; Huang, Z., Nucleic acid crystallization and X-ray crystallography facilitated by single selenium atom. *Frontiers of Chemical Science and Engineering* **2016,** *10* (2), 196-202.

66.   Choi, J.; Majima, T., Conformational changes of non-B DNA. *Chemical Society reviews* **2011,** *40* (12), 5893-5909.

67.   Dock-Bregeon, A.; Moras, D.; Giegé, R., Nucleic acids and their complexes. *Crystallization of nucleic acids and proteins. A practical approach* **1999**, 209-243.

68.   Ke, A.; Doudna, J. A., Crystallization of RNA and RNA–protein complexes. *Methods* **2004,** *34* (3), 408-414.

69.   Mooers, B. H., Crystallographic studies of DNA and RNA. *Methods* **2009,** *47* (3), 168-176.

70.   Blakeley, M. P.; Langan, P.; Niimura, N.; Podjarny, A., Neutron crystallography: opportunities, challenges, and limitations. *Current opinion in structural biology* **2008,** *18* (5), 593-600.

71.   Chatake, T., Neutron Nucleic Acid Crystallography. In *Nucleic Acid Crystallography*, Springer: 2016; pp 283-300.

72.   Vandavasi, V. G.; Blakeley, M. P.; Keen, D. A.; Hu, L. R.; Huang, Z.; Kovalevsky, A., Temperature-Induced Replacement of Phosphate Proton with Metal Ion Captured in Neutron Structures of A-DNA. *Structure* **2018,** *26* (12), 1645-1650. e3.

73.   Caton‐Williams, J.; Huang, Z., Synthesis and DNA‐polymerase incorporation of colored 4‐selenothymidine triphosphate for polymerase recognition and DNA visualization. *Angewandte Chemie International Edition* **2008,** *47* (9), 1723-1725.

74.   Logan, G.; Igunbor, C.; Chen, G.-X.; Davis, H.; Simon, A.; Salon, J.; Huang, Z., A simple strategy for incorporation, protection, and deprotection of selenium functionality. *Synlett* **2006,** *2006* (10), 1554-1558.

75.   Zhang, W.; Hassan, E. A.; Huang, Z., Synthesis of novel di-Se-containing thymidine and Se-DNAs for structure and function studies. *Science China Chemistry* **2013,** *56* (3), 273-278.

76.   Zhang, W.; Huang, Z., Synthesis of the 5′-se-thymidine phosphoramidite and convenient labeling of DNA oligonucleotide. *Organic letters* **2011,** *13* (8), 2000-2003.

77.   Du, Q.; Carrasco, N.; Teplova, M.; Wilds, C. J.; Egli, M.; Huang, Z., Internal derivatization of oligonucleotides with selenium for X-ray crystallography using MAD. *Journal of the American Chemical Society* **2002,** *124* (1), 24-25.

78.   Sun, Z.; Liu, Q.; Qu, G.; Feng, Y.; Reetz, M. T., Utility of B-factors in protein science: interpreting rigidity, flexibility, and internal motion and engineering thermostability. *Chemical reviews* **2019,** *119* (3), 1626-1665.

79.   Schneider, B.; Gelly, J.-C.; de Brevern, A. G.; Černý, J., Local dynamics of proteins and DNA evaluated from crystallographic B factors. *Acta Crystallographica Section D: Biological Crystallography* **2014,** *70* (9), 2413-2419.

80.   Turkman, N.; Gelovani, J. G.; Alauddin, M. M., A novel method for stereospecific fluorination at the 2′‐arabino‐position of pyrimidine nucleoside: synthesis of [18F]‐FMAU. *Journal of Labelled Compounds and Radiopharmaceuticals* **2010,** *53* (13), 782-786.

81.   Otwinowski, Z.; Minor, W., [20] Processing of X-ray diffraction data collected in oscillation mode. In *Methods in enzymology*, Elsevier: 1997; Vol. 276, pp 307-326.

82.   McCoy, A. J.; Grosse-Kunstleve, R. W.; Adams, P. D.; Winn, M. D.; Storoni, L. C.; Read, R. J., Phaser crystallographic software. *Journal of applied crystallography* **2007,** *40* (4), 658-674.

83.   Potterton, E.; Briggs, P.; Turkenburg, M.; Dodson, E., A graphical user interface to the CCP4 program suite. *Acta Crystallographica Section D: Biological Crystallography* **2003,** *59* (7), 1131-1137.

84.   Winn, M. D.; Ballard, C. C.; Cowtan, K. D.; Dodson, E. J.; Emsley, P.; Evans, P. R.; Keegan, R. M.; Krissinel, E. B.; Leslie, A. G.; McCoy, A., Overview of the CCP4 suite and current developments. *Acta Crystallographica Section D: Biological Crystallography* **2011,** *67* (4), 235-242.

85.   Murshudov, G. N.; Skubák, P.; Lebedev, A. A.; Pannu, N. S.; Steiner, R. A.; Nicholls, R. A.; Winn, M. D.; Long, F.; Vagin, A. A., REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallographica Section D: Biological Crystallography* **2011,** *67* (4), 355-367.

86.   Emsley, P.; Lohkamp, B.; Scott, W. G.; Cowtan, K., Features and development of Coot. *Acta Crystallographica Section D: Biological Crystallography* **2010,** *66* (4), 486-501.

87.   Zhang, W.; Sheng, J.; Hassan, A. E.; Huang, Z., Synthesis of 2′‐Deoxy‐5‐(methylselenyl) cytidine and Se‐DNAs for Structural and Functional Studies. *Chemistry – An Asian Journal* **2012,** *7* (3), 476-479.

88.   Carrasco, N.; Buzin, Y.; Tyson, E.; Halpert, E.; Huang, Z., Selenium derivatization and crystallization of DNA and RNA oligonucleotides for X-ray crystallography using multiple anomalous dispersion. *Nucleic acids Res.* **2004,** *32*, 1638-1646.

89.   Höbartner, C.; Micura, R., Chemical Synthesis of Selenium-Modified Oligoribonucleotides and Their Enzymatic Ligation Leading to an U6 SnRNA Stem-Loop Segment. *J. Am. Chem. Soc.* **2004,** *126*, 1141-1149.

90.   Buzin, Y.; Carrasco, N.; Huang, Z., Synthesis of selenium-derivatized cytidine and oligonucleotides for X-ray crystallography using MAD. *Organic letters* **2004,** *6* (7), 1099-1102.

91.   Olinski, R.; Jurgowiak, M.; Zaremba, T., Uracil in DNA—its biological significance. *Mutation Research/Reviews in Mutation Research* **2010,** *705* (3), 239-245.

92.   Lindahl, T., Instability and decay of the primary structure of DNA. *nature* **1993,** *362* (6422), 709-715.

93.   Hagen, L.; Peña-Diaz, J.; Kavli, B.; Otterlei, M.; Slupphaug, G.; Krokan, H. E., Genomic uracil and human disease. *Experimental cell research* **2006,** *312* (14), 2666-2672.

94.   Visnes, T.; Doseth, B.; Pettersen, H. S.; Hagen, L.; Sousa, M. M.; Akbari, M.; Otterlei, M.; Kavli, B.; Slupphaug, G.; Krokan, H. E., Uracil in DNA and its processing by different DNA glycosylases. *Philosophical Transactions of the Royal Society B: Biological Sciences* **2008,** *364* (1517), 563-568.

95.   el-Hajj, H. H.; Zhang, H.; Weiss, B., Lethality of a dut (deoxyuridine triphosphatase) mutation in Escherichia coli. *Journal of Bacteriology* **1988,** *170* (3), 1069-1075.

96.   Gadsden, M. H.; McIntosh, E.; Game, J. C.; Wilson, P. J.; Haynes, R., dUTP pyrophosphatase is an essential enzyme in Saccharomyces cerevisiae. *The EMBO journal* **1993,** *12* (11), 4425-4431.

97.   Tinkelenberg, B. A.; Hansbury, M. J.; Ladner, R. D., dUTPase and uracil-DNA glycosylase are central modulators of antifolate toxicity in Saccharomyces cerevisiae. *Cancer research* **2002,** *62* (17), 4909-4915.

98.   Rada, C.; Williams, G. T.; Nilsen, H.; Barnes, D. E.; Lindahl, T.; Neuberger, M. S., Immunoglobulin isotype switching is inhibited and somatic hypermutation perturbed in UNG-deficient mice. *Current Biology* **2002,** *12* (20), 1748-1755.

99.   Gilbert, W., Origin of life: The RNA world. *nature* **1986,** *319* (6055), 618.

100.   Poole, A.; Penny, D.; Sjöberg, B.-M., Methyl-RNA: an evolutionary bridge between RNA and DNA? *Chemistry & biology* **2000,** *7* (12), R207-R216.

101.   Poole, A.; Penny, D.; Sjöberg, B.-M., Confounded cytosine! Tinkering and the evolution of DNA. *Nature Reviews Molecular Cell Biology* **2001,** *2* (2), 147.

102.   Lesk, A., Why does DNA contain thymine and RNA uracil? *Journal of theoretical biology* **1969,** *22* (3), 537-540.

103.   Langridge, R.; Marmur, J., X-ray diffraction study of a DNA which contains uracil. *Science* **1964,** *143* (3613), 1450-1451.

104.   Slupphaug, G.; Mol, C. D.; Kavli, B.; Arvai, A. S.; Krokan, H. E.; Tainer, J. A., A nucleotide-flipping mechanism from the structure of human uracil–DNA glycosylase bound to DNA. *Nature* **1996,** *384* (6604), 87.

105.   Savva, R.; McAuley-Hecht, K.; Brown, T.; Pearl, L., The structural basis of specific base-excision repair by uracil–DNA glycosylase. *Nature* **1995,** *373* (6514), 487.

106.   Mol, C. D.; Arvai, A. S.; Slupphaug, G.; Kavli, B.; Alseth, I.; Krokan, H. E.; Tainer, J. A., Crystal structure and mutational analysis of human uracil-DNA glycosylase: structural basis for specificity and catalysis. *Cell* **1995,** *80* (6), 869-878.

107.   Parker, J. B.; Bianchet, M. A.; Krosky, D. J.; Friedman, J. I.; Amzel, L. M.; Stivers, J. T., Enzymatic capture of an extrahelical thymine in the search for uracil in DNA. *Nature* **2007,** *449* (7161), 433.

108.   Firbank, S. J.; Wardle, J.; Heslop, P.; Lewis, R. J.; Connolly, B. A., Uracil recognition in archaeal DNA polymerases captured by X-ray crystallography. *Journal of molecular biology* **2008,** *381* (3), 529-539.

109.   Wardle, J.; Burgers, P. M.; Cann, I. K.; Darley, K.; Heslop, P.; Johansson, E.; Lin, L.-J.; McGlynn, P.; Sanvoisin, J.; Stith, C. M., Uracil recognition by replicative DNA polymerases is limited to the archaea, not occurring with bacteria and eukarya. *Nucleic acids research* **2007,** *36* (3), 705-711.

110.   Christofferson, A.; Zhao, L.; Sun, H.; Huang, Z.; Huang, N., Theoretical studies of the base pair fidelity of selenium-modified DNA. *The Journal of Physical Chemistry B* **2011,** *115* (33), 10041-10048.

111.   Sheng, J.; Gan, J.; Soares, A. S.; Salon, J.; Huang, Z., Structural insights of non-canonical U• U pair and Hoogsteen interaction probed with Se atom. *Nucleic acids research* **2013,** *41* (22), 10476-10487.

112.   Bobkov, G. V.; Mikhailov, S. N.; Van Aerschot, A.; Herdewijn, P., Phosphoramidite building blocks for efficient incorporation of 2′-O-aminoethoxy (and propoxy) methyl nucleosides into oligonucleotides. *Tetrahedron* **2008,** *64* (27), 6238-6251.

113.   Watson, J. D.; Crick, F. H., Molecular structure of nucleic acids. *Nature* **1953,** *171* (4356), 737-738.

114.   Blount, K. F.; Uhlenbeck, O. C., The structure-function dilemma of the hammerhead ribozyme. *Annu. Rev. Biophys. Biomol. Struct.* **2005,** *34*, 415-440.

115.   Eddy, S. R., Non–coding RNA genes and the modern RNA world. *Nature Reviews Genetics* **2001,** *2* (12), 919-929.

116.   Jordheim, L. P.; Durantel, D.; Zoulim, F.; Dumontet, C., Advances in the development of nucleoside and nucleotide analogues for cancer and viral diseases. *Nat. Rev. Drug Discov.* **2013,** *12* (6), 447-464.

117.   Burnett, J. C.; Rossi, J. J., RNA-based therapeutics: current progress and future prospects. *Chemistry & biology* **2012,** *19* (1), 60-71.

118.   Lin, L.; Sheng, J.; Huang, Z., Nucleic acid X-ray crystallography via direct selenium derivatization. *Chem. Soc. Rev.* **2011,** *40*, 4591-4602.

119.   Zhang, W.; Szostak, J. W.; Huang, Z., Nucleic acid crystallization and X-ray crystallography facilitated by single selenium atom. *Front. Chem. Sci. Eng.* **2016,** *10*, 196-202.

120.   Park, S.; Okamura, I.; Sakashita, S.; Yum, J. H.; Acharya, C.; Gao, L.; Sugiyama, H., Development of DNA Metalloenzymes Using a Rational Design Approach and Application in the Asymmetric Diels–Alder Reaction. *ACS Catal.* **2015,** *5* (8), 4708-4712.

121.   Roelfes, G.; Feringa, B. L., DNA‐Based Asymmetric Catalysis. *Angew. Chem. Int. Ed.* **2005,** *44* (21), 3230-3232.

122.   Niemeyer, C. M., Nanoparticles, proteins, and nucleic acids: biotechnology meets materials science. *Angew. Chem. Int. Ed.* **2001,** *40* (22), 4128-4158.

123.   Liu, S.; Clever, G. H.; Takezawa, Y.; Kaneko, M.; Tanaka, K.; Guo, X.; Shionoya, M., Direct conductance measurement of individual metallo‐DNA duplexes within single‐molecule break junctions. *Angewandte Chemie International Edition* **2011,** *50* (38), 8886-8890.

124.   Mallajosyula, S. S.; Pati, S. K., Toward DNA conductivity: a theoretical perspective. *The Journal of Physical Chemistry Letters* **2010,** *1* (12), 1881-1894.

125.   Tsutsui, M.; Matsubara, K.; Ohshiro, T.; Furuhashi, M.; Taniguchi, M.; Kawai, T., Electrical detection of single methylcytosines in a DNA oligomer. *Journal of the American Chemical Society* **2011,** *133* (23), 9124-9128.

126.   Hendrickson, W. A., Determination of Macromolecular Structures from Anomalous Diffraction of synchrontron radiation. *Science* **1991,** *254*, 5028.

127.   Hendrickson, W. A., Synchrotron crystallography. *Trends biochem. sci.* **2000,** *25* (12), 637-643.

128.   Hendrickson, W. A.; Horton, J. R.; LeMaster, D. M., Selenomethionyl proteins produced for analysis by multiwavelength anomalous diffraction (MAD): a vehicle for direct determination of three-dimensional structure. *The EMBO j.* **1990,** *9* (5), 1665.

129.   Sheng, J.; Huang, Z., Selenium Derivatization of Nucleic Acids for Phase and Structure Determination in Nucleic Acid X-ray Crystallography. *Int. J. Mol. Sci.* **2008,** *9*, 258-271.

130.   Sheng, J.; Huang, Z., Selenium Derivatization of Nucleic Acids for X-Ray Crystal-Structure and Function Studies. *Chemistry & Biodiversity* **2010,** *7*, 753-785.

131.   Lalla Aicha Ba; Mandy Döring; Vincent Jamier; Jacob, C., Tellurium: an element with great biological potency and potential. *Org. Biomol. Chem.* **2010,** *8*, 4203-4216.

132.   Moroder, L., Isosteric replacement of sulfur with other chalcogens in peptides and proteins. *J. Peptide Sci.* **2005,** *11*, 187-214.

133.   K. J. R. Rosman; Taylor, P. D. P., ISOTOPIC COMPOSITIONS OF THE ELEMENTS 1997. *Pure Appl. Chem.* **1998,** *70*, 217-235.

134.   Shinichi Saito; Jian Zhang; Kyoko Tanida; Shigemasa Takahashi; Koizumi, T., A Systematic [125]Te NMR Study of Organotellurium Compounds: The Effect of Oxidation States and Substituents. *Tetrahedron* **1999,** *55*, 2545-2552.

135.   Ramadan, S. E.; Razak, A.; Ragab, A. M.; El-Meleigy, M., Incorporation of tellurium into amino acids and proteins in a tellurium-tolerant fungi. *Biological trace element research* **1989,** *20* (3), 225-232.

136.   Boles, J. O.; Lewinski, K.; Kunkle, M.; Odom, J. D.; Dunlap, R. B.; Lebioda, L.; Hatada, M., Bio-incorporation of telluromethionine into buried residues of dihydrofolate reductase. *Nature Structural Biology* **1994,** *1*, 283-284.

137.   Budisa, N.; Karnbrock, W.; Steinbacher, S.; Humm, A.; Prade, L.; Neuefeind, T.; Moroder, L.; Huber, R., Bioincorporation of Telluromethionine into Proteins: A Promising New Approach for X-ray Structure Analysis of Proteins. *J. Mol. Biol.* **1997,** *270*, 616-623.

138.   Sheng, J.; Hassan, A. E. A.; Huang, Z., New Telluride-Mediated Elimination for Novel Synthesis of 2',3'-Didehydro-2'.3'-dideoxynucleosides. *J. Org. Chem.* **2008,** *73*, 3725-3729.

139.   Sheng, J.; Hassan, A. E. A.; Huang, Z., Synthesis of the First Tellurium-Derivatized Oligonucleotides for Structural and Functional Studies. *Chem. Eur. J.* **2009,** *15*, 10210-10216.

140.   Sheng, J.; Hassan, A. E. A.; Zhang, W.; Zhou, J.; Xu, B.; Soares, A. S.; Huang, Z., Synthesis, structure and imaging of oligodeoxyribonucleotides with tellurium-nucleobase derivatization. *Nucleic Acids Res.* **2011,** *39*, 3962-3971.

141.   Du, Q.; Carrasco, N.; Teplova, M.; Wilds, C. J.; Egli, M.; Huang, Z., Internal Derivatization of Oligonucleotides with Selenium for X-ray Crystallography Using MAD. *J. Am. Chem. Soc.* **2002,** *124*, 24-25.

142.   Höbartner, C.; Rieder, R.; Kreutz, C.; Puffer, B.; Lang, K.; Polonskaia, A.; Serganov, A.; Micura, R., Syntheses of RNAs with up to 100 Nucleotides Containing Site-Specific 2'-Methylseleno Labels for Use in X-ray Crystallography. *J. Am. Chem. Soc.* **2005,** *127*, 12035-12045.

143.   Teplova, M.; Wilds, C. J.; Wawrzak, Z.; Tereshko, V.; Du, Q.; Carrasco, N.; Huang, Z.; Egli, M., Covalent incorporation of selenium into oligonucleotides for X-ray crystal structure determination via MAD: proof of principle. *Biochimie* **2002,** *84*, 849-858.

144.   Salon, J.; Chen, G.; Portilla, Y.; Germann, M. W.; Huang, Z., Synthesis of a 2'-Se-uridine Phosphoramidite and Its Incorporation into Oligonucleotides for Structural Study. *Org. Lett.* **2005,** *7*, 5645-5648.

145.   Sheng, J.; Jiang, J.; Salon, J.; Huang, Z., Synthesis of a 2'-Se-thymidine Phosphoramidite and Its Incorporation into Oligonucleotides for Crystal Structure Study. *Org. Lett.* **2007,** *9*, 749-752.

146.   Karino, N.; Ueno, Y.; Matsuda, A., Synthesis and properties of oligonucleotides containing 5-formyl-2′-deoxycytidine: in vitro DNA polymerase reactions on DNA templates containing 5-formyl-2′-deoxycytidine. *Nucleic acids Res.* **2001,** *29* (12), 2456-2463.

147.   Jain, S.; Zon, G.; Sundaralingam, M., Base only binding of spermine in the deep groove of the A-DNA octamer d (GTGTACAC). *Biochemistry* **1989,** *28* (6), 2360-2364.

148.   Jiang, J.; Sheng, J.; Carrasco, N.; Huang, Z., Selenium derivatization of nucleic acids for crystallography. *Nucleic acids Res.* **2006,** *35* (2), 477-485.

149.   Salon, J.; Sheng, J.; Gan, J.; Huang, Z., Synthesis and crystal structure of 2′-Se-modified guanosine containing DNA. *J. Org. Chem.* **2010,** *75* (3), 637-641.

150.   Sheng, J.; Salon, J.; Gan, J.; Huang, Z., Synthesis and crystal structure study of 2′-Se-adenosine-derivatized DNA. *Sci. China: Chem.* **2010,** *53* (1), 78-85.

151.   Frisch, M. J.; Trucks, G.; Schlegel, H.; Scuseria, G.; Robb, M.; Cheeseman, J.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G., Gaussian 09, Revision D. 01, Gaussian. *Inc.: Wallingford, CT* **2009**.

152.   Zhao, Y.; Truhlar, D. G., The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: two new functionals and systematic testing of four M06-class functionals and 12 other functionals. *Theoretical Chemistry Accounts* **2008,** *120* (1-3), 215-241.

153.   Krishnan, R.; Binkley, J. S.; Seeger, R.; Pople, J. A., Self‐consistent molecular orbital methods. XX. A basis set for correlated wave functions. *The Journal of Chemical Physics* **1980,** *72* (1), 650-654.

154.   Dunning Jr, T. H., Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen. *The Journal of chemical physics* **1989,** *90* (2), 1007-1023.

155.   Dennington, R.; Keith, T.; Millam, J.; Eppinnett, K.; Hovell, W.; Gilliland, R., GaussView v. 5.0. 9 Visualizer and Builder. *Gaussian Inc, Wallingford, CT* **2009**.

156.   Becke, A. D., Density-functional exchange-energy approximation with correct asymptotic behavior. *Physical review A* **1988,** *38* (6), 3098.

157.   Lee, C.; Yang, W.; Parr, R. G., Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Physical review B* **1988,** *37* (2), 785.

158.   Becke, A. D., Density‐functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* **1993,** *98*, 5648-5652.

159.   Roothaan, C. C. J., New developments in molecular orbital theory. *Reviews of modern physics* **1951,** *23* (2), 69.

160.   Miehlich, B.; Savin, A.; Stoll, H.; Preuss, H., Results obtained with the correlation energy density functionals of Becke and Lee, Yang and Parr. *Chemical Physics Letters* **1989,** *157* (3), 200-206.

161.   Stephens, P. J.; Devlin, F.; Chabalowski, C.; Frisch, M. J., Ab initio calculation of vibrational absorption and circular dichroism spectra using density functional force fields. *The Journal of physical chemistry* **1994,** *98* (45), 11623-11627.

162.   Schäfer, A.; Horn, H.; Ahlrichs, R., Fully optimized contracted Gaussian basis sets for atoms Li to Kr. *The Journal of Chemical Physics* **1992,** *97* (4), 2571-2577.

163.   Glendening, E.; Reed, A.; Carpenter, J.; Weinhold, F., NBO Version 3.1. **1998**.

APPENDICES

Appendix A. Computational study of the MeSe incorporation at 2'-position of uridine, cytidine, thymidine and 5-F-uridine.

Appendix A.1 Calculation of the energy barrier for the MeSe incorporation
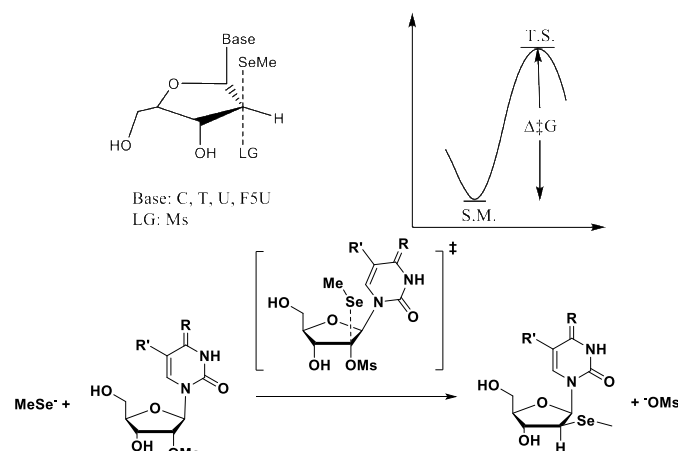
Figure A.1 Energy barrier calculation for MeSe incorporation.

The calculation of the energy barrier of MeSe incorporation for cytidine, thymidine, uridine and 5-F-uridine derivatives were conducted and compared. We found that the $\Delta^{\ddagger}G$ of the reaction is directly related to electron richness of the nucleobase following the order of cytosine > uracil > thymine > 5-F-uracil. The 5-F-U give the lowest activation energy and the cytidine give the highest one (Table A.1). Comparing with uridine, the activation energy is thymidine and cytidine give 0.3 and 3.2 kcal/mol higher respectively, which is consistent with our experiment that the temperature of MeSe incorporation to thymidine require 95 $^{\circ}$C instead of 50 $^{\circ}$C for uridine. The higher energy barrier may cause by the electron repelling between the electron-rich nucleobase with the selenium atom.

Table A.1 Energy barriers of MeSe incorporation

Starting material	Energy of S.M. (a.u.)	Energy of T.S. (a.u.)	$\Delta^{\ddagger}G$ (kcal/mol)	$\Delta\Delta^{\ddagger}G$ (kcal/mol)
2'-Ms-U + MeSe$^{\ominus}$	-1911.6947	-1911.6684	16.5437	-

2'-Ms-T + MeSe°	-1950.9676	-1950.9407	16.9026	0.3589
2'-Ms-C + MeSe°	-1891.8093	-1891.7778	19.7534	3.2097
2'-Ms-F5U + MeSe°	-2010.9082	-2010.8836	15.4505	-1.0932

Appendix A.2 Computational structures and coordinates

Methods: Calculations were executed at Georgia State University using the Gaussian 09[151] program with the M062X[152] method. The 6-311G+(d)[153] basis set for C, H, O, N and aug-cc-pVTZ-PP[154] basis set for Se and S were used. The calculations did converge to transition state as determined by the existence of one negative frequency, and further confirmed by Intrinsic reaction coordinate (IRC) calculation. Molecular structures are rendered in GaussView5.0.9.[155]
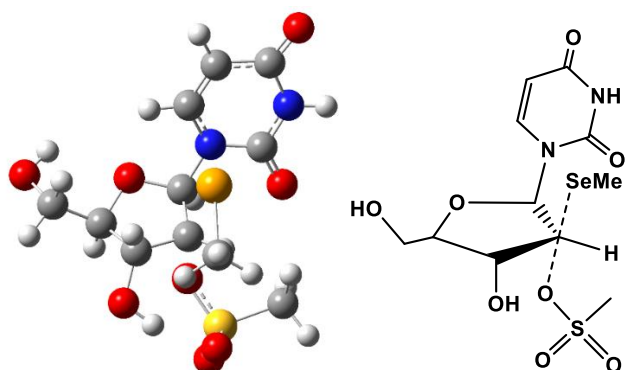
Figure A.2 Transition state complex of the 2'-mesylate-uridine MeSe incorporation.

Table A.2 Coordinates of the transition state of MeSe incorporation of 2'-Ms-uridine

Atomic Number	Coordinates (Angstroms)		
	X	Y	Z
8	-0.587796	4.433391	-1.661506
6	-0.857744	3.739902	-0.4226
1	-1.721221	4.217783	0.031803
1	-0.002313	3.810077	0.255743
6	-1.152338	2.286944	-0.68327
1	-2.013054	2.156962	-1.337593
8	0.04759	1.748414	-1.351783
6	0.115453	0.327177	-1.132987
1	-0.31649	-0.210879	-1.970246
7	1.51935	-0.064471	-1.052132
6	2.523508	0.841459	-0.778096
6	3.807315	0.456821	-0.627512
1	4.587564	1.163413	-0.408685

6	4.163879	-0.94039	-0.718587
8	5.306243	-1.41748	-0.617096
7	3.071828	-1.788827	-0.957079
6	1.749836	-1.426803	-1.076536
8	0.830235	-2.256627	-1.214742
6	-1.327566	1.425857	0.570164
6	-0.756761	0.08356	0.108624
8	-2.68764	1.465076	1.009338
1	-0.691585	1.807116	1.365092
1	-3.056882	0.570172	1.157471
1	2.189224	1.861099	-0.697564
1	0.049028	3.902407	-2.170031
34	1.295527	0.162592	2.011721
6	-0.209188	-0.18956	3.232151
1	0.109502	-0.006679	4.256638
1	-1.054495	0.466515	3.009598
1	-0.538516	-1.22458	3.142654
16	-3.124667	-1.514671	-0.402403
6	-2.190368	-3.013507	-0.57325
1	-2.779738	-3.813045	-0.130218
1	-2.026788	-3.184468	-1.634079
1	-1.229373	-2.903748	-0.075909
8	-4.31142	-1.625901	-1.229347
8	-3.32768	-1.287367	1.035738
1	-0.720875	-0.807799	0.704214
8	-2.198561	-0.421136	-0.948731
1	3.260587	-2.780379	-0.958042
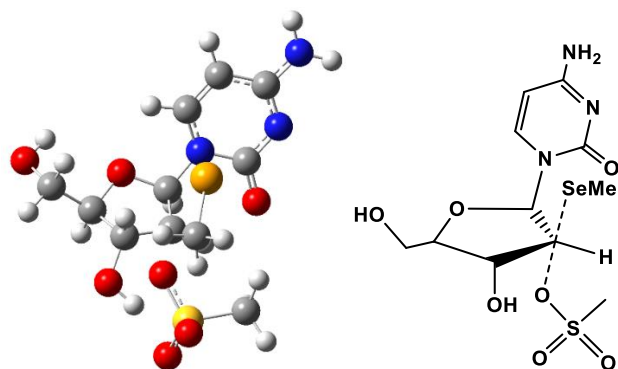
Figure A.3 Transition state complex of the 2'-mesylate-cytidine MeSe incorporation

Table A.3 Coordinates of the transition state of MeSe incorporation of 2'-Ms-cytidine

Atomic Number	Coordinates (Angstroms)		
	X	Y	Z
8	-0.685755	4.367188	-1.850324
6	-0.952199	3.702451	-0.595537

1	-1.848927	4.155174	-0.181333
1	-0.121148	3.839993	0.101045
6	-1.171203	2.22845	-0.812058
1	-1.990004	2.041394	-1.506436
8	0.079607	1.707006	-1.386872
6	0.153438	0.289337	-1.138106
1	-0.264323	-0.269486	-1.969071
7	1.551221	-0.099673	-1.013402
6	2.553516	0.813387	-0.853338
6	3.835196	0.401105	-0.690892
1	4.637587	1.107471	-0.563495
6	4.073074	-1.011255	-0.685948
7	3.084122	-1.902024	-0.825578
6	1.793604	-1.484815	-0.961054
8	0.807666	-2.262948	-1.035118
6	-1.381862	1.411364	0.466594
6	-0.73611	0.072228	0.094149
8	-2.767136	1.407263	0.832409
1	-0.810559	1.855201	1.275747
1	-3.064536	0.516637	1.11167
1	2.249779	1.846089	-0.866632
1	0.062039	3.924973	-2.287295
34	1.223127	0.317782	2.023866
6	-0.297253	0.28643	3.279093
1	-0.010621	-0.231178	4.191688
1	-0.611695	1.298437	3.53002
1	-1.141247	-0.244043	2.830344
16	-3.072972	-1.620789	-0.359478
6	-2.188763	-3.143692	-0.522508
1	-2.14523	-3.388921	-1.580924
1	-1.179136	-3.006146	-0.139652
1	-2.7335	-3.903999	0.031952
8	-4.310246	-1.7133	-1.122723
8	-3.230045	-1.344588	1.081438
1	-0.695919	-0.802732	0.712859
8	-2.161355	-0.55712	-0.970704
7	5.324639	-1.491769	-0.543056
1	6.115468	-0.887149	-0.414545
1	5.462808	-2.487357	-0.528701
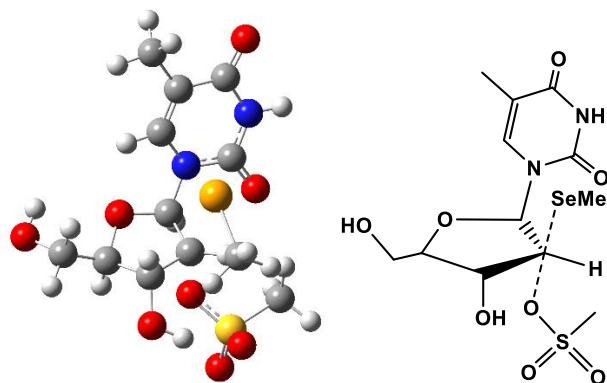
Figure A.4 Transition state complex of the 2'-mesylate-thymidine MeSe incorporation

Table A.4 Coordinates of the transition state of MeSe incorporation of 2'-Ms-thymidine

Atomic	Coordinates (Angstroms)		
Number	X	Y	Z
8	-0.449133	4.506427	-1.494338
6	-0.746372	3.770599	-0.287432
1	-1.574589	4.279513	0.198084
1	0.11627	3.759592	0.383908
6	-1.134725	2.352321	-0.612308
1	-1.987234	2.313798	-1.289138
8	0.035186	1.747472	-1.273356
6	-0.015252	0.318651	-1.11962
1	-0.481145	-0.145935	-1.98247
7	1.349189	-0.193631	-1.034807
6	2.434225	0.627246	-0.780005
6	3.682948	0.14366	-0.606122
6	3.888615	-1.291708	-0.681735
8	4.989589	-1.862318	-0.550107
7	2.738381	-2.044446	-0.924577
6	1.457833	-1.565483	-1.062521
8	0.462966	-2.310368	-1.197929
6	-1.395853	1.459434	0.604774
6	-0.916069	0.09558	0.105802
8	-2.759324	1.569743	1.028944
1	-0.750876	1.769538	1.422895
1	-3.167057	0.692354	1.183101
1	2.197706	1.676957	-0.730039
1	0.237088	4.031223	-1.993304
34	1.082413	-0.068075	2.031725
6	-0.450975	-0.338717	3.239244
1	-0.126146	-0.240735	4.272826
1	-1.223348	0.409246	3.046627
1	-0.876617	-1.330738	3.093484
16	-3.422864	-1.281628	-0.478165

6	-2.69013	-2.863263	-0.780877
1	-3.324964	-3.616602	-0.320737
1	-2.63645	-3.003952	-1.857637
1	-1.686917	-2.873598	-0.359521
8	-4.654446	-1.181293	-1.247437
8	-3.55897	-1.121681	0.980816
1	-0.979754	-0.825741	0.651299
8	-2.403193	-0.262733	-0.991594
1	2.843662	-3.049961	-0.943686
6	4.872673	1.006672	-0.319169
1	5.648818	0.861529	-1.071191
1	4.592573	2.05795	-0.305
1	5.306868	0.750756	0.647999
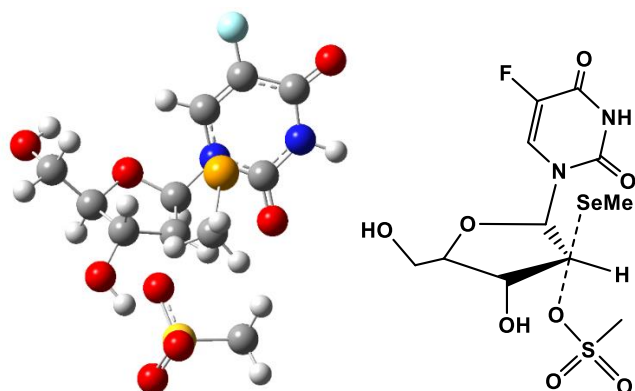
Figure A.5 Transition state complex of the 2'-mesylate-5-fluoro-uridine MeSe incorporation

Table A.5 Coordinates of the transition state of MeSe incorporation of 2'-Ms-5-F-uridine

Atomic Number	Coordinates (Angstroms)		
	X	Y	Z
8	-0.484896	4.529249	-1.434548
6	-0.756562	3.768745	-0.237418
1	-1.578887	4.263883	0.27184
1	0.117222	3.749436	0.419191
6	-1.145546	2.355034	-0.580651
1	-2.004138	2.325317	-1.24991
8	0.018584	1.760974	-1.264096
6	-0.022772	0.332507	-1.125083
1	-0.485523	-0.130576	-1.99016
7	1.349982	-0.163624	-1.051241
6	2.412778	0.675441	-0.791473
6	3.639837	0.169892	-0.613281
6	3.916494	-1.245631	-0.664043
8	5.0255	-1.777631	-0.507174
7	2.772272	-2.000887	-0.923721

6	1.482353	-1.53774	-1.073062
8	0.50239	-2.294549	-1.21405
6	-1.394889	1.443694	0.625452
6	-0.908431	0.088706	0.107317
8	-2.758141	1.54037	1.052781
1	-0.750421	1.747473	1.446306
1	-3.157337	0.659215	1.207975
1	2.190246	1.727732	-0.743251
1	0.206961	4.078446	-1.947903
34	1.112496	-0.076308	2.003566
6	-0.404918	-0.391185	3.218391
1	-0.071235	-0.305817	4.250098
1	-1.189761	0.348675	3.045883
1	-0.816835	-1.386837	3.059189
16	-3.410027	-1.299079	-0.474258
6	-2.671711	-2.87369	-0.799676
1	-3.305208	-3.635901	-0.35249
1	-2.61509	-2.998186	-1.878248
1	-1.670382	-2.888331	-0.374175
8	-4.647288	-1.196372	-1.233322
8	-3.534399	-1.155481	0.987251
1	-0.96213	-0.83926	0.642574
8	-2.398427	-0.270068	-0.985099
1	2.887572	-3.005711	-0.940303
9	4.713149	1.001687	-0.354198

Appendix B. The mechanism study of the ring-opening of uracil through the nucleophilic attack of methylselenium

During the synthesis of 2'-β-MeSe-uridine, a new reaction was observed when we use the tert-butyloxycarbonyl (BOC) as the protective for the N^3 (Scheme B.1). With the BOC protection, the MeSe nucleophile attack the C-4 directly instead of the leaving group on the 2'-position. Further computational study indicates that it is the intramolecular interaction between lone pair on the 4-O and the π^* of the carbonyl group of the BOC activating the nucleobase for the MeSe nucleophilic attack. The cleavage of the mesylate group is also require which probably compensate the energy rising due to the break of the aromatic ring. The effect of counterion was also examined and reveals that stronger Lewis acid would facilitate the cleavage of BOC group instead of ring-opening of the nucleobase.

Appendix B.1. Synthesis of selenolester and characterization
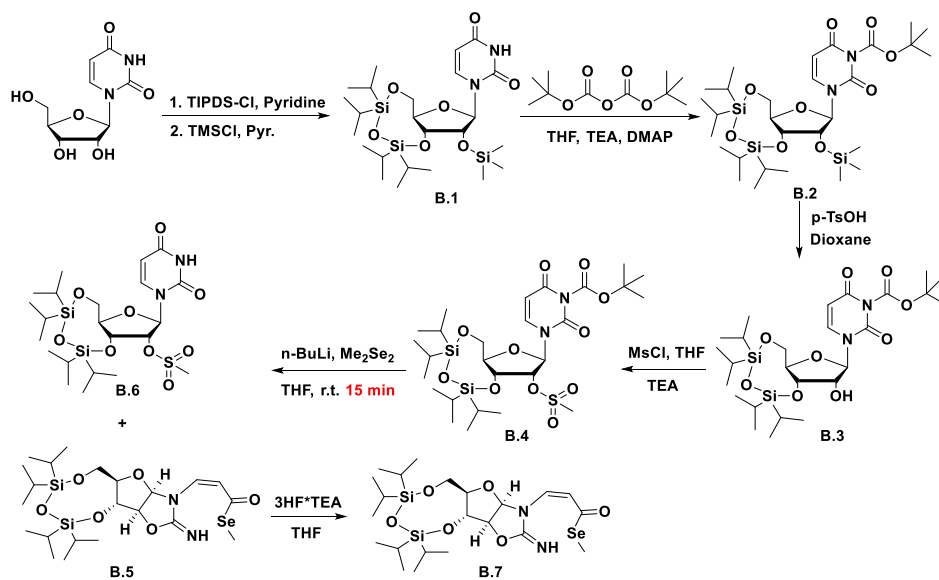
Scheme B.1 Conversion of uridine to its selenoester

The 3',5'-OH groups of the uridine were protected with tetraisopropyldisilylene (TIPDS), followed by the in-situ protection of 2'-OH group with trimethylsilyl (TMS) give **B.1**. With the hydroxyl group fully protected, the BOC group can be introduced to N^3 give **B.2**, followed by the

deprotection of 2'-OTMS by treating with p-toluenesulfonic acid (TsOH). Then the 2'-OH group of **B.3** was activated with mesylate. The MeSeLi nucleophile was generated by treating the dimethyldiselenide with n-BuLi under -78°C followed by the injection of **B.4**. The reaction only took 15 min to complete under room temperature generating a new compound **B.5** with 50% yield and a byproduct **B.6**. B.5 was further deprotected with 3HF·TEA to remove the TIPDS generating **B.7** as final product.
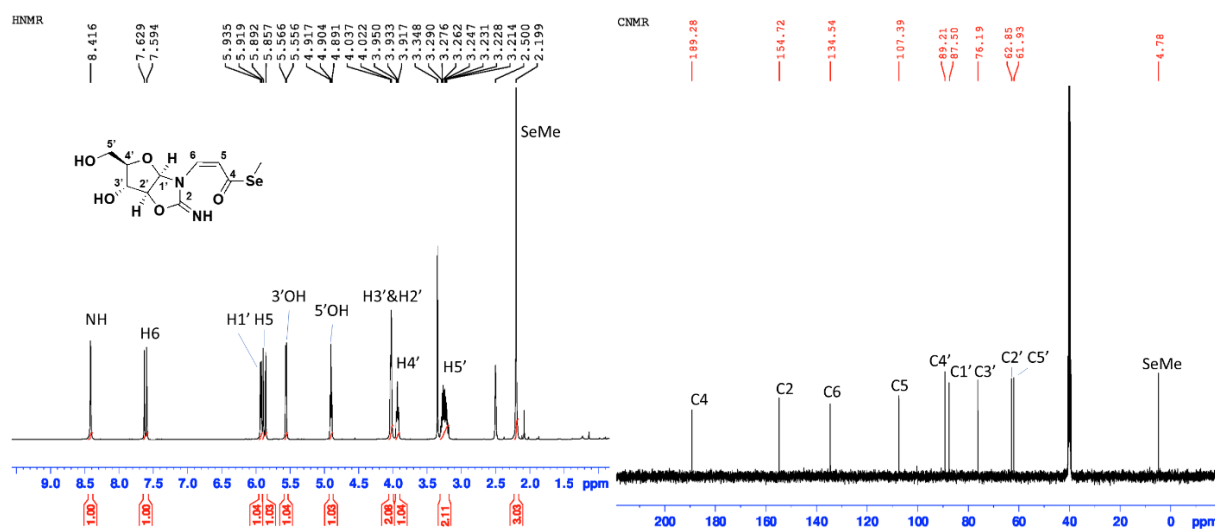
*Figure B.1 The ^{1}HNMR and ^{13}CNMR of **B.7***

To confirm the structure of the new compound **B.7**, lots of spectroscopy studies were conducted including NMR (Fig. B.1, B.2), IR (Fig. B.3), MS (Fig. B.4) and UV (Fig. B.5). MeSe functionality were observed in both ^{1}HNMR, ^{13}CNMR, and further confirmed with MS. Because of electron shield of selenium, MeSe group shift to high field in ^{13}CNMR at 4.8 ppm. The connection between the proton and carbon of MeSe group was confirmed with 2D HSQC (Fig. B.2). Based on the ^{1}HNMR, both the BOC group and 2'-mesylate were cleaved. Both H-6 and H-5 were observed with an increased the J-coupling of 14 Hz instead of 8 Hz for uridine, which indicates the break of the aromatic system. The doublet of the 1'-H exhibiting in **B.5** and **B.7** suggests a conformation change of the sugar pucker caused by nucleophilic attack of the O-2 at

the 2'-mesylate. Due to the direct connection with selenium, the ^{13}C chemical shift of C-4 largely

shift from 160 ppm in **B.4** to 190 ppm (Fig. B.1).

The 2D HMBC and HSQC reveals the connections between C-C and C-H (Fig. B.2). The

N-H shows cross peak between C-2, C-1' and C-2' instead of C-4 or C-5 confirming the break of

the nucleobase ring. The cross peaks between Se**C**H₃ – 5-**H** and Se**C**H₃ – **C**-4 were observed in

HMBC suggesting that the MeSe functionality is connected to C-4 rather than C-2'.
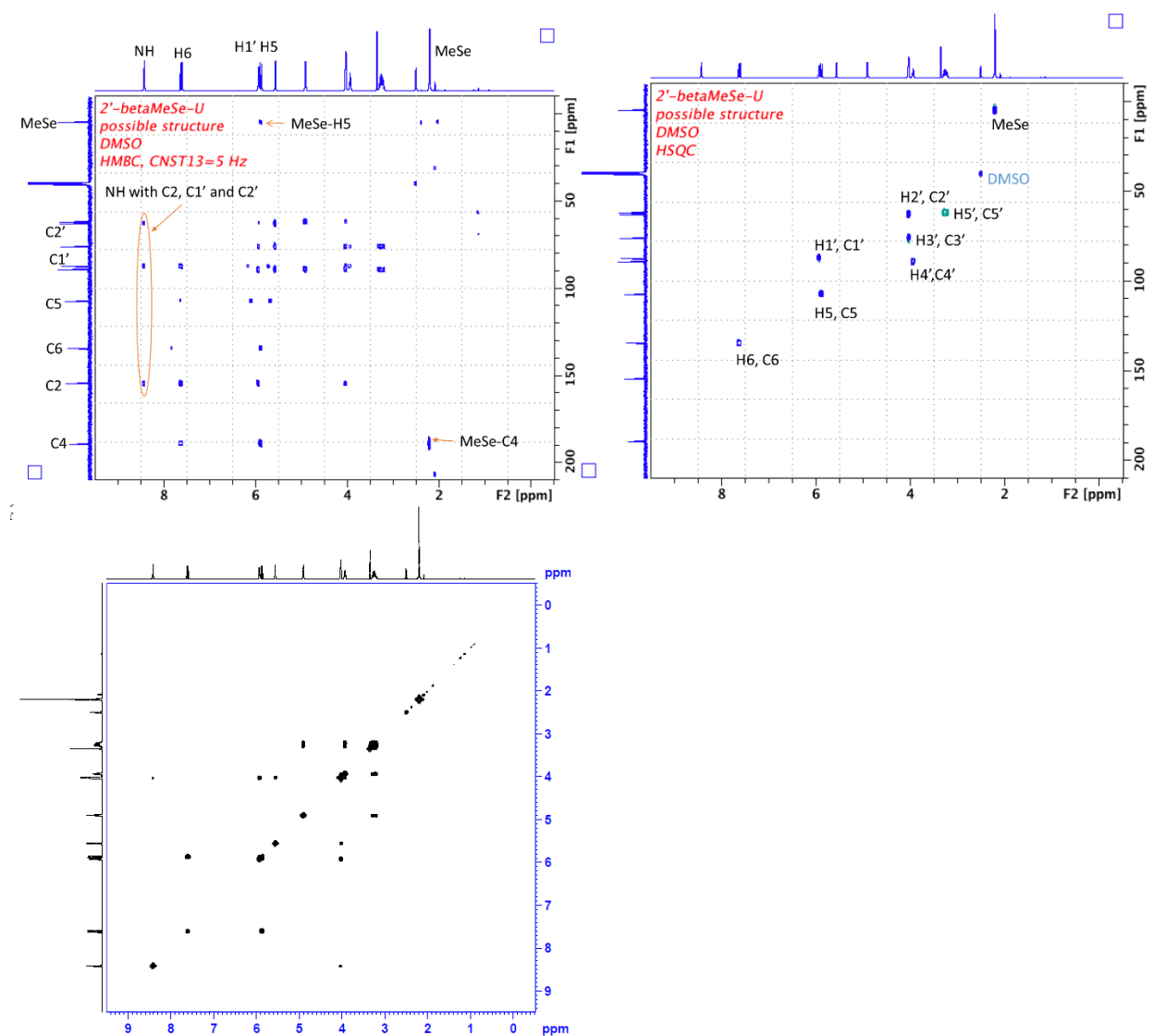
*Figure B.2 ^1H-^{13}C HMBC, HSQC and ^1H-^1H COSY of **B.7***

In the infrared (IR) spectrum, three different double bond, C=O, C=C, C=N, were observed at 1720 cm⁻¹, 1669 cm⁻¹ and 1594 cm⁻¹ respectively (Fig. B.3, a). Moreover, a UV comparison between B.7 and uridine was also conducted. A new peak at 300 nm was observed indicating the change of the nucleobase which is consistent with our NMR results (Fig. B.3, b). The molecule weight was measured by HRMS using electrospray ionization (ESI) as ion source. The molecular ion peak was observed at 323.0125 with the typical selenium isotopes distribution (Fig. B.4).
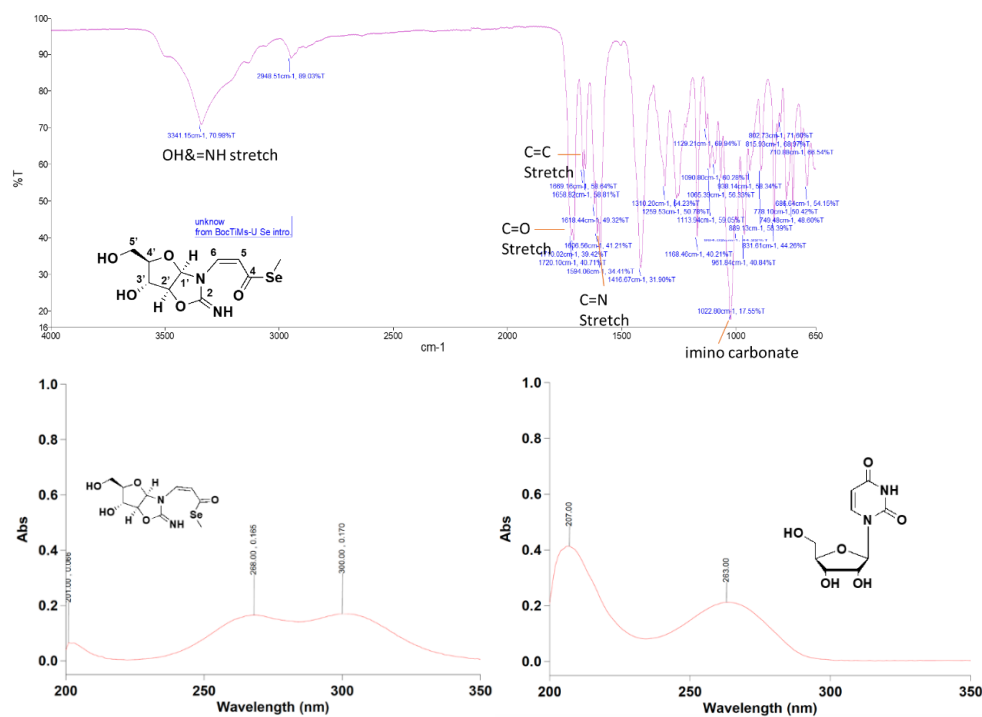
*Figure B.3 a. IR of **B.7**; b. UV spectra of **B.7** and uridine.*
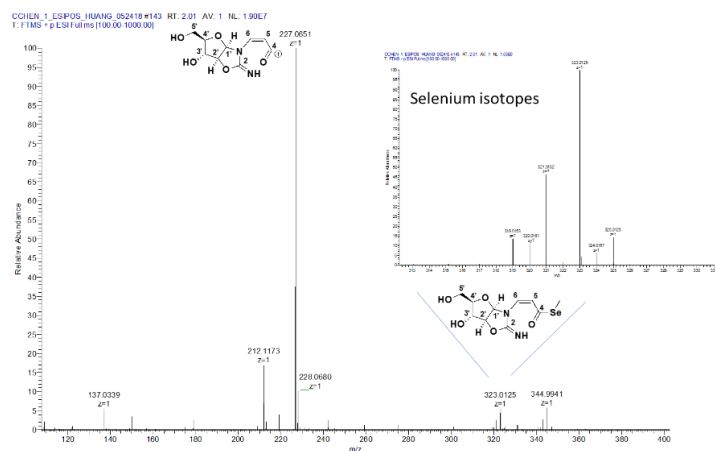
*Figure B.4 HRMS(ESI) of **B.7***

Appendix B.2 The effect of BOC group for ring-opening

The reaction generated B.6 as the major byproduct which only lose the BOC group with the mesylate retained. This imply that the BOC group is necessary for the reaction. To understand the effect of BOC group activating the nucleobase, we conducted the same reaction with different protective groups on N3 (Scheme B.2, b). However, without BOC group, no reaction happened at room temperature overnight. When rise the temperature to 60 °C, the MeSe functionality substitute the leaving group at 2'-position generating the 2'-β-MeSe-modified uridine derivatives. Without any protective groups on the N3 (Scheme B.2, c), under room temperature and 50 °C, only 2'-β-MeSe-incorporated products were observed. These results demonstrate that the BOC group is required for the reaction.
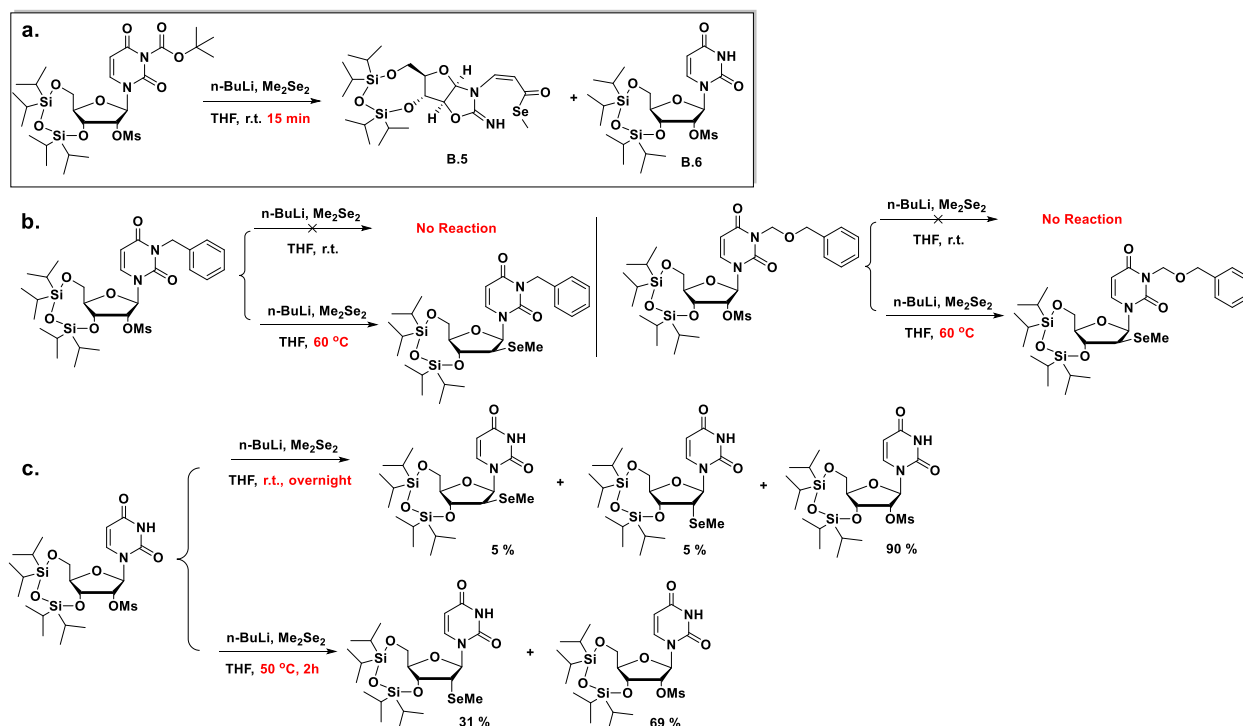
Scheme B.2 Examine the effect of the BOC group

The role of BOC playing in the reaction is now clean. However, how BOC group activated the nucleobase is still not understood. A computational study of the second order orbital perturbation using the natural bond orbital (NBO) calculation in Gaussian09W shows that there is

a significant interaction between the long pair of O-4 and the π^* orbital of the BOC (Fig. B.5) which is about 2.0 kcal/mol (Table B.1). Thus, the BOC group may serve as an intramolecular Lewis acid, which lowering the energy of π^* orbital of the carbonyl group at 4-position making it more vulnerable to the attack of nucleophiles.
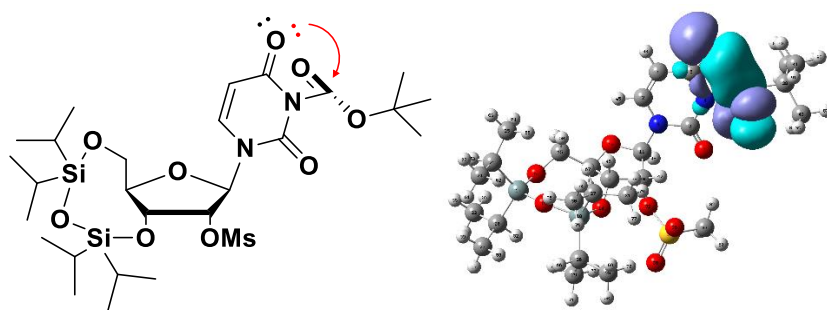
Figure B.5 Orbital interaction of n-π^*

Table B.1 Calculated energy of the n-π^* interaction

Method	Second order perturbation (kcal/mol)
B3lyp/def-TZVP	1.64
B3lyp/6-31g+(d)	2.00
HF/6-31g+(d)	2.04

Appendix B.2 The effect of Leaving group and counterion

Even the activation of BOC group makes the reaction kinetically possible by lowering the energy barrier. However, the reaction might be still thermodynamically unfavored since the break of the aromatic system would cause significant energy penalty. Thus, we expect that the cleavage of the leaving group is also necessary to compensate the energy rising. Several reactions with different group at 2'-position were performed (Scheme B.3). Different leaving group give the same product. However, with a poor leaving group TMS on the 2'-O, only the BOC cleaved by-product was observed. These results are in accordance with our hypothesis that a good leaving group at 2'-position is required.

As we proved previously, the BOC group acts as an intramolecular Lewis acid to activate the nucleobase. In the reaction condition, there is Li$^+$ ion which could also serves as Lewis acid

coordinating with carbonyl group. To examine the effect of counterions of MeSe, we generated MeSe nucleophiles under different conditions with different ion including Na[+], Li[+] and Mg[2+] (Scheme B.4). When using Na+ as the counterion which is a poor Lewis acid, the MeSe nucleophile still attack the 4-position give the ring-opening product (Scheme B.4, b). Interestingly, unlike treated with MeSeLi, the BOC group still retain on the N[3]. When treated with (MeSe)$_2$Mg, the stronger Lewis acid Mg[2+] directly cleave the BOC group without giving the desired product (Scheme B.4, c). The results reveal that the counterion only activate the cleavage of BOC group instead of activating the nucleobase.
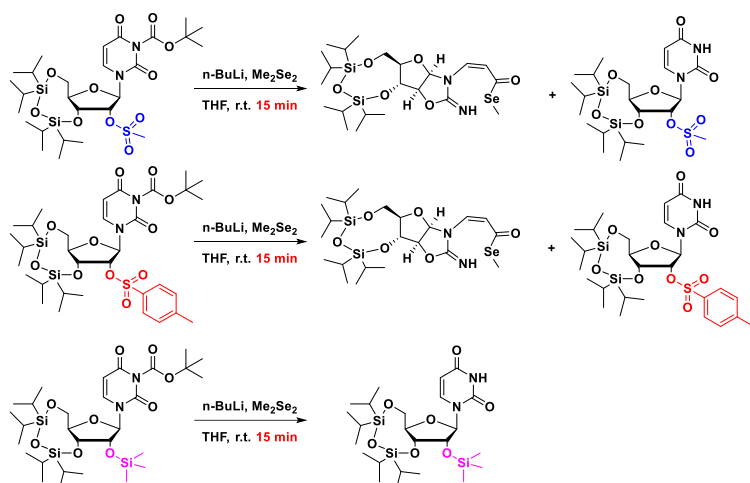
Scheme B.3 Examine the effect of leaving group at 2'-position
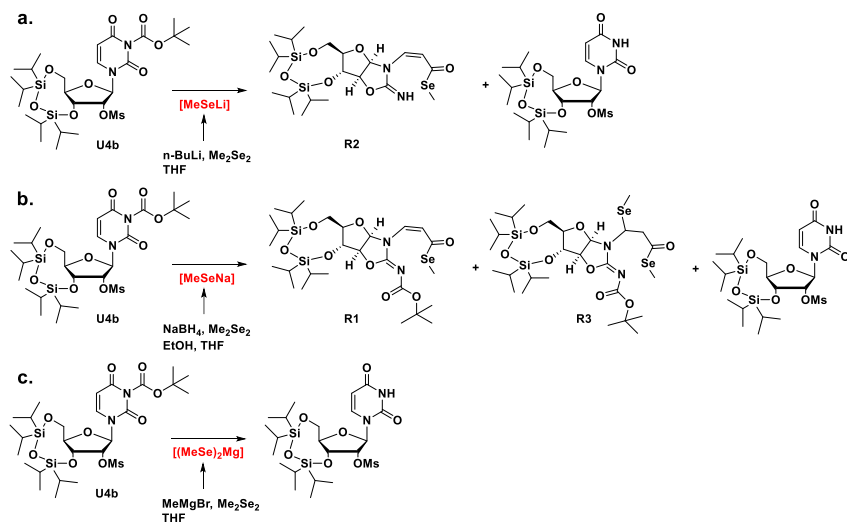
Scheme B.4 Examine the effect of different counterion of the MeSe nucleophile.

Appendix B.3 Proposed mechanism

Here, we proposed a possible mechanism for the ring-opening reaction (Scheme B.5). With the activation of BOC group, the MeSe⁻ acts as nucleophile attack the C-4. Then, the 2-O serves as another nucleophile substitute the 2'-OMs which compensate the energy increasing due to the break of the aromatic ring. The Li⁺ ion acts as Lewis acid facilitate the in-situ deprotection of the BOC group giving the selenolester as the final product.
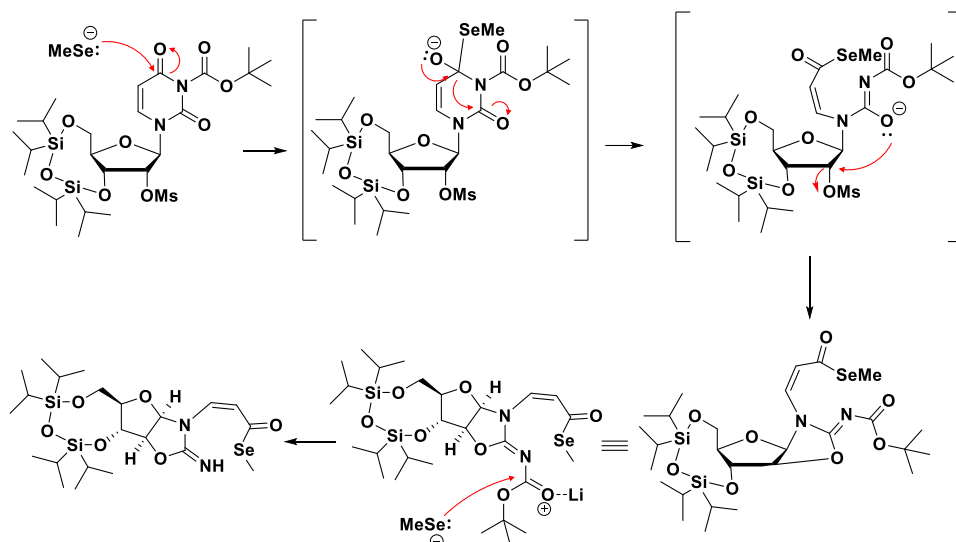
Scheme B.5 Proposed mechanism.

Appendix B.4 Convert selenolester to its derivatives

Both selenolester B.5 and B.7 are relatively reactive to nucleophilic attack and several functional groups can be easily introduced (Scheme B.6).
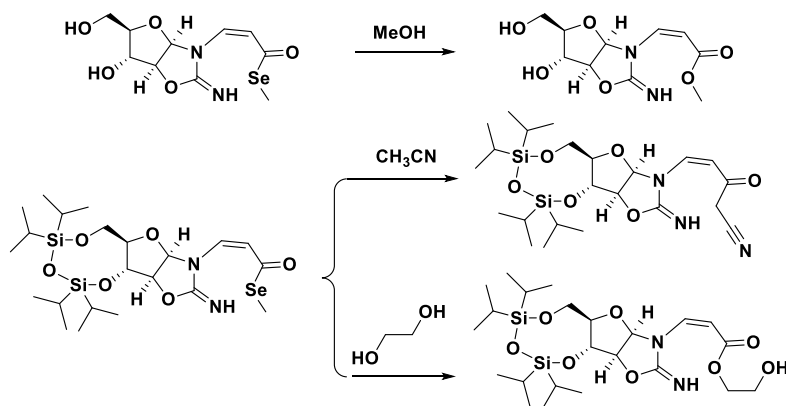
Scheme B.6 Convert selenolester to its derivatives

Appendix B.5 Experimental section
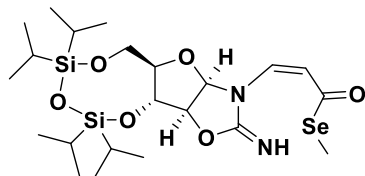
Se-methyl-(Z)-3-((3aR,5R,6R,6aS)-5,6-O-(1,1,3,3-tetraisopropyldisiloxane-1,3-diyl)-

5-methylene-2-iminotetrahydrofuro[2,3-d]oxazol-3(2H)-yl)prop-2-eneselenoate **(B.5)**

Dimethyldiselenide (1.7 mL, 18 mmol) was dissolved in 10 mL anhydrous THF in a flask and

cooled down to -78 °C in a acetone-dry ice bath. n-BuLi (4.8 mL, 2.5 M in hexane, 12 mmol) was

slowly injected into the solution and stirred for 10 min and then warm up to 0 °C. B.4 (2 g, 3.0

mmol) was dissolved in 20 mL anhydrous THF and injected into the reaction. The reaction was

warmed up to room temperature and stirred for 15 min then quenched with water (1 mL). The

mixture was diluted with EtOAc (30 ml) and washed with H_2O (3×10 ml) and brine. The organic

phase was dried over anhydrous $MgSO_4$. After filtration, the solvent was concentrated under

reduced pressure and the residue was purified by a flash silica gel column (eluent: 10 to 25 % ethyl

acetate in hexane) given **B.5** as white solid (0.8 g, 45 %). ^{1}H-NMR (400 MHz, $CDCl_3$) δ (ppm):

1.04-1.10 (m, 28H, 4×iPrSi), 2.31 (s, 3H, $SeCH_3$), 3.86 (dd, J_1=3.2 Hz, J_2= 6.7 Hz, 1H, H-4'), 3.90

(dd, J_1=5.9 Hz, J_2= 12.1 Hz, 1H, Ha-5'), 4.06 (dd, J_1=3.2 Hz, J_2= 12.1 Hz, 1H, Hb-5'), 4.21-4.24

(m, 1H, H-2'), 4.27 (dd, J_1=4.2 Hz, J_2= 6.9 Hz, 1H, H-3'), 5.69 (d, J=7.2 Hz, 1H, H-1'), 5.93 (d,

J=14.1 Hz, 1H, H-5), 7.86 (d, J=14.1 Hz, 1H, H-6). ^{13}C-NMR (100 MHz, DMSO-d_6) δ (ppm): 4.8

($SeCH_3$), 12.5, 12.7, 13.2, 13.4, 16.9, 17.2, 17.3, 17.4 (i-Pr), 61.9 (C-2'), 62.3 (C-5'), 80.0 (C-3'),

83.3 (C-4'), 85.6 (C-1'), 109.9 (C-5), 133.3 (C-6), 154.9 (C-2), 190.1 (C-4). HRMS (ESI):

$C_{22}H_{40}N_2O_6SeSi_2$; $[M+H]^+$: 564.1675 (calc. 564.1670).
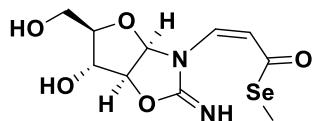
Se-methyl-(Z)-3-((3aR,5R,6R,6aS)-6-hydroxy-5-(hydroxymethyl)-2-imino-tetrahydrofuro[2,3-d]oxazol-3(2H)-yl)prop-2-eneselenoate (B.7) Compound **B.5** (0.56 g, 1.0 mmol) was dissolved in 10 mL anhydrous THF, and treated with $3HF \cdot Et_3N$ (0.16 mL, 1.0 mmol) at 40 ^{o}C for 2 h. The mixture was evaporated to dryness, and the residue was purified by flash column chromatography on silica gel (5% MeOH in methylene chloride) to afford pure product **B.7** (318 mg, 99%) as white solid. ^{1}H-NMR (400 MHz, DMSO-d_6) δ (ppm): 2.20 (s, 3H, SeCH$_3$), 3.19-3.30 (m, 2H, H-5'), 3.93 (t, J=6.4 Hz, 1H, H-4'), 4.02-4.03 (m, 2H, H-2', H-3'), 4.90 (t, J=5.3 Hz, 1H, OH-5'), 5.56 (d, J=4.2 Hz, 1H, OH-3'), 5.87 (d, J=14.0 Hz, 1H, H-5), 5.92 (d, J=6.5 Hz, 1H, H-1'), 7.61 (d, J=14.0 Hz, 1H, H-6), 8.42 (s, 1H, NH). ^{13}C-NMR (100 MHz, DMSO-d_6) δ (ppm): 4.8 (SeCH$_3$), 61.9 (C-5'), 62.8 (C-2'), 76.2 (C-3'), 87.5 (C-1'), 89.2 (C-4'), 107.4 (C-5), 134.5 (C-6), 154.7 (C-2), 189.3 (C-4). HRMS (ESI): $C_{10}H_{14}N_2O_5Se$; $[M+H]^+$: 323.0125 (calc. 323.0146). ^{1}H-^{13}C HSQC, ^{1}H-^{13}C HMBC and ^{1}H-^{1}H COSY (Figure B.2).
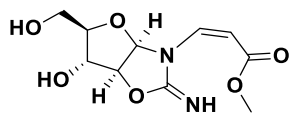
Methyl-(Z)-3-((3aR,5R,6R,6aS)-6-hydroxy-5-(hydroxymethyl)-2-imino-tetrahydrofuro[2,3-d]oxazol-3(2H)-yl)-acrylate (B.8) Compound B.7 was treated with 5 mL K_2CO_3 MeOH solution

(50 mM). The reaction stirred at room temperature for 3 hr. The mixture was evaporated to dryness, and the residue was purified by flash column chromatography on silica gel (5% MeOH in methylene chloride) to afford pure product **B.8** (25 mg, 95%). ^{1}H-NMR (400 MHz, DMSO-d_6) δ (ppm): 3.20-3.30 (m, 2H, H-5'), 3.64 (s, 3H, OCH$_3$), 3.93 (t, J=6.5 Hz, 1H, H-4'), 4.03 (m, 2H, H-2', H-3'), 4.89 (t, J=4.9 Hz, 1H, OH-5'), 5.34 (d, J=14.5 Hz, 1H, H-5), 5.55 (d, J=3.7 Hz, 1H, OH-3'), 5.91 (d, J=6.4 Hz, 1H, H-1'), 7.63 (d, J=14.1 Hz, 1H, H-6), 8.29 (s, 1H, NH).
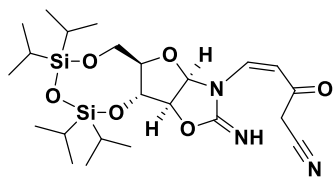
(Z)-5-((3aR,5R,6R,6aS)-5,6-O-(1,1,3,3-tetraisopropyldisiloxane-1,3-diyl)-5-methylene-2-imino-tetrahydrofuro[2,3-d]oxazol-3(2H)-yl)-3-oxopent-4-enenitrile (B.9) 0.01 mL acetonitrile was added to 2 mL THF and was cooled to -78°C in a acetone-dry ice bath, followed with the injection of n-BuLi (0.2 mL, 2.5 M, 1.6 mmol). After stirring for 20 min, the compound **B.5** (300 mg, 0.5 mmol) was dissolved in 3 mL THF and injected under room temperature. The reaction was stirred at room temperature for another 15 min and quenched with water. The mixture was evaporated to dryness, and the residue was purified by flash column chromatography on silica gel (40% ethyl acetate in hexane) to afford pure product **B.9** (140 mg, 55%). ^{1}H-NMR (400 MHz, CDCl$_3$) δ (ppm): 1.01-1.06 (m, 28H, 4×iPrSi), 3.56 (s, 2H, CH$_2$CN), 3.81-3.84 (m, 1H, H-4'), 3.88 (dd, J_1=5.6 Hz, J_2=12.2 Hz, 1H, Ha-5'), 4.02 (dd, J_1=3.1 Hz, J_2=12.2 Hz, 1H, Ha-5'), 4.21-4.26 (m, 2H, H-2', H-3'), 5.70 (d, J=6.8 Hz, 1H, H-1'), 5.93 (d, J=14.2, 1H, H-5), 6.88 (s, 1H, NH), 7.90 (d, J=14.2 Hz, 1H, H-6). ^{13}C-NMR (100 MHz, CDCl$_3$) δ (ppm): 12.5, 12.7, 13.1, 13.4, 16.9, 17.2, 17.3, 17.4 (i-Pr), 30.7 (CH$_2$CN), 61.9 (C-2', C-5'), 79.5 (C-3'), 83.4 (C-4'), 85.2 (C-1'), 104.3 (C-5), 114.4 (CN), 138.9 (C-6), 154.7 (C-2), 185.3 (C-4).
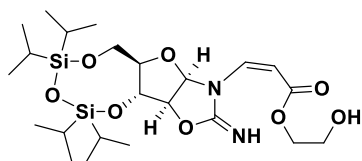
2-hydroxyethyl-(Z)-3-((3aR,5R,6R,6aS)-6-hydroxy-5-(hydroxymethyl)-2-imino-tetrahydrofuro[2,3-d]oxazol-3(2H)-yl)-acrylate (B.10) The compound B.5 (300 mg, 0.5 mmol) was dissolved in a mixture of THF and ethylene glycol (5 mL, v/v 1:1), followed with the addition of triethylamine (0.14 mL, 1 mmol). The reaction was heated up to 60 °C overnight. The mixture was evaporated to dryness, and the residue was purified by flash column chromatography on silica

gel (4 % MeOH in methylene chloride) to afford pure product **B.10** (252 mg, 95%). ^{1}H-NMR (400 MHz, CDCl$_3$) δ (ppm): 1.02-1.07 (m, 28H, 4×iPrSi), 2.56 (br, 1H, OH), 3.83-3.89 (m, 4H, Ha-5', H-4', CH$_2$C$\underline{H}_2$OH), 4.04 (dd, J_1=2.6 Hz, J_2=11.7 Hz, 1H, Hb-5'), 4.20-4.22 (m, 1H, H-2'), 4.24-4.27 (m, 3H, H-3', C$\underline{H}_2$CH$_2$OH), 5.51 (d, J=14.3 Hz, 1H, H-5), 5.68 (d, J=7.1 Hz, 1H, H-1'), 7.90 (d, J=14.3 Hz, 1H, H-6). ^{13}C-NMR (100 MHz, CDCl$_3$) δ (ppm): 12.5, 12.7, 13.2, 13.4, 16.9, 17.2, 17.3, 17.4 (i-Pr), 61.4 (CH$_2$C$\underline{H}_2$OH), 62.1 (C-2'), 62.4 (C-5'), 65.9 (C$\underline{H}_2$CH$_2$OH), 80.1 (C-3'), 83.3 (C-4'), 85.7 (C-1'), 98.8 (C-5), 137.5 (C-6), 155.1 (C-2), 167.5 (C-4).

Computational study

Calculations were executed at Georgia State University using the Gaussian09W[151] program with the B3LYP[156-158] and HF[159] method. The basis set 6-31G+(d)[160, 161], def-TZVP[162] were used. The second order perturbation was calculated by natural bond orbital (NBO) version 3.1[163] in Gaussian09W. Molecular structures are rendered in GaussView5.0.9.[155]

Appendix C. Nucleic Acid Mini Screen Reagent Formulation

Tube#	Precipitant	Buffer	Polyamine	Monovalent Ion	Divalent Ion
1.	10% v/v (+/-)-2-Methyl-2,4-pentanediol	0.040 M Sodium cacodylate trihydrate pH 5.5	0.020 M Hexamine cobalt(III) chloride	None	0.020 M Magnesium chloride hexahydrate
2.	10% v/v (+/-)-2-Methyl-2,4-pentanediol	0.040 M Sodium cacodylate trihydrate pH 5.5	0.020 M Hexamine cobalt(III) chloride	0.080 M Sodium chloride	0.020 M Magnesium chloride hexahydrate
3.	10% v/v (+/-)-2-Methyl-2,4-pentanediol	0.040 M Sodium cacodylate trihydrate pH 5.5	0.020 M Hexamine cobalt(III) chloride	0.012 M Sodium chloride, 0.080 M Potassium chloride	None
4.	10% v/v (+/-)-2-Methyl-2,4-pentanediol	0.040 M Sodium cacodylate trihydrate pH 5.5	0.020 M Hexamine cobalt(III) chloride	0.040 M Lithium chloride	0.020 M Magnesium chloride hexahydrate
5.	10% v/v (+/-)-2-Methyl-2,4-pentanediol	0.040 M Sodium cacodylate trihydrate pH 6.0	0.012 M Spermine tetrahydrochloride	0.080 M Potassium chloride	0.020 M Magnesium chloride hexahydrate
6.	10% v/v (+/-)-2-Methyl-2,4-pentanediol	0.040 M Sodium cacodylate trihydrate pH 6.0	0.012 M Spermine tetrahydrochloride	0.080 M Potassium chloride	None
7.	10% v/v (+/-)-2-Methyl-2,4-pentanediol	0.040 M Sodium cacodylate trihydrate pH 6.0	0.012 M Spermine tetrahydrochloride	0.080 M Sodium chloride	0.020 M Magnesium chloride hexahydrate
8.	10% v/v (+/-)-2-Methyl-2,4-pentanediol	0.040 M Sodium cacodylate trihydrate pH 6.0	0.012 M Spermine tetrahydrochloride	0.080 M Sodium chloride	None
9.	10% v/v (+/-)-2-Methyl-2,4-pentanediol	0.040 M Sodium cacodylate trihydrate pH 6.0	0.012 M Spermine tetrahydrochloride	0.080 M Sodium chloride, 0.012 M Potassium chloride	0.020 M Magnesium chloride hexahydrate
10.	10% v/v (+/-)-2-Methyl-2,4-pentanediol	0.040 M Sodium cacodylate trihydrate pH 6.0	0.012 M Spermine tetrahydrochloride	0.012 M Sodium chloride, 0.080 M Potassium chloride	None
11.	10% v/v (+/-)-2-Methyl-2,4-pentanediol	0.040 M Sodium cacodylate trihydrate pH 6.0	0.012 M Spermine tetrahydrochloride	0.080 M Sodium chloride	0.020 M Barium chloride
12.	10% v/v (+/-)-2-Methyl-2,4-pentanediol	0.040 M Sodium cacodylate trihydrate pH 6.0	0.012 M Spermine tetrahydrochloride	0.080 M Potassium chloride	0.020 M Barium chloride
13.	10% v/v (+/-)-2-Methyl-2,4-pentanediol	0.040 M Sodium cacodylate trihydrate pH 6.0	0.012 M Spermine tetrahydrochloride	None	0.080 M Strontium chloride hexahydrate
14.	10% v/v (+/-)-2-Methyl-2,4-pentanediol	0.040 M Sodium cacodylate trihydrate pH 7.0	0.012 M Spermine tetrahydrochloride	0.080 M Potassium chloride	0.020 M Magnesium chloride hexahydrate
15.	10% v/v (+/-)-2-Methyl-2,4-pentanediol	0.040 M Sodium cacodylate trihydrate pH 7.0	0.012 M Spermine tetrahydrochloride	0.080 M Potassium chloride	None
16.	10% v/v (+/-)-2-Methyl-2,4-pentanediol	0.040 M Sodium cacodylate trihydrate pH 7.0	0.012 M Spermine tetrahydrochloride	0.080 M Sodium chloride	0.020 M Magnesium chloride hexahydrate
17.	10% v/v (+/-)-2-Methyl-2,4-pentanediol	0.040 M Sodium cacodylate trihydrate pH 7.0	0.012 M Spermine tetrahydrochloride	0.080 M Sodium chloride	None
18.	10% v/v (+/-)-2-Methyl-2,4-pentanediol	0.040 M Sodium cacodylate trihydrate pH 7.0	0.012 M Spermine tetrahydrochloride	0.080 M Sodium chloride, 0.012 M Potassium chloride	0.020 M Magnesium chloride hexahydrate
19.	10% v/v (+/-)-2-Methyl-2,4-pentanediol	0.040 M Sodium cacodylate trihydrate pH 7.0	0.012 M Spermine tetrahydrochloride	0.012 M Sodium chloride, 0.080 M Potassium chloride	None
20.	10% v/v (+/-)-2-Methyl-2,4-pentanediol	0.040 M Sodium cacodylate trihydrate pH 7.0	0.012 M Spermine tetrahydrochloride	0.080 M Sodium chloride	0.020 M Barium chloride
21.	10% v/v (+/-)-2-Methyl-2,4-pentanediol	0.040 M Sodium cacodylate trihydrate pH 7.0	0.012 M Spermine tetrahydrochloride	0.080 M Potassium chloride	0.020 M Barium chloride
22.	10% v/v (+/-)-2-Methyl-2,4-pentanediol	0.040 M Sodium cacodylate trihydrate pH 7.0	0.012 M Spermine tetrahydrochloride	0.040 M Lithium chloride	0.080 M Strontium chloride hexahydrate,
23.	10% v/v (+/-)-2-Methyl-2,4-pentanediol	0.040 M Sodium cacodylate trihydrate pH 7.0	0.012 M Spermine tetrahydrochloride	0.040 M Lithium chloride	0.020 M Magnesium chloride hexahydrate
24.	10% v/v (+/-)-2-Methyl-2,4-pentanediol	0.040 M Sodium cacodylate trihydrate pH 7.0	0.012 M Spermine tetrahydrochloride	None	0.080 M Strontium chloride hexahydrate

Appendix D. Characterization of key compounds

Appendix D.1 NMR and MS Spectra for 2'-MeSe-arabinothymidine modification

BOMTi-MeU
CDCl3
CNMR

163.52
150.85
138.04
134.34
128.26
127.65
127.62
109.86
91.35
81.88
75.10
72.34
70.54
69.23
67.09
60.38
17.42
17.36
17.27
17.24
17.09
17.01
16.96
16.88
13.46
13.26
12.97

BOMTi-MeU
CDCl3
COSY

BOMTi-MeU
CDCl3
HSQC

BOMTi-MeU
CDCl3
HMBC

N-BOM-3',5'-TIPDS-2'-Ms-MeU
CDCl3
HNMR

7.405
7.403
7.371
7.353
7.341
7.323
7.304
7.267
6.151
6.138
5.496
5.421
5.409
5.398
5.241
5.228
5.215
4.682
4.441
4.437
4.432
4.154
4.147
4.129
4.111
4.102
4.097
4.094
3.190
3.126
3.033
2.044
1.926
1.276
1.259
1.241
1.105
1.097
1.080
1.070
1.065
1.051
1.045
-0.000
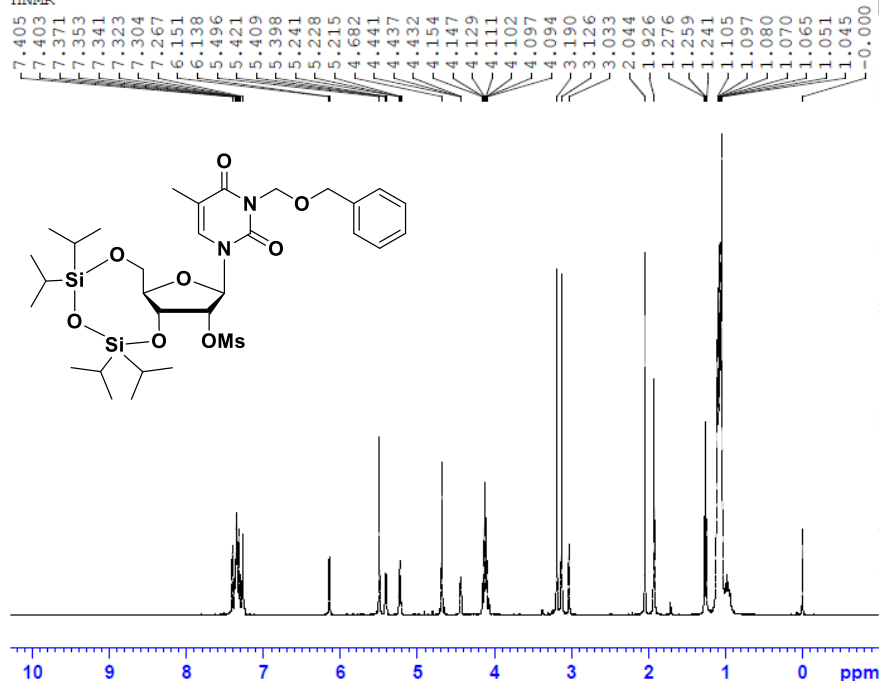
Current Data Parameters
NAME   20190120_BOMTiMs-T
EXPNO   1
PROCNO   1

F2 - Acquisition Parameters
Date_   20190121
Time   10.23
INSTRUM   spect
PROBHD   5 mm PABBO BB-
PULPROG   zg30
TD   65536
SOLVENT   CDCl3
NS   16
DS   2
SWH   8278.146 Hz
FIDRES   0.126314 Hz
AQ   3.9583745 sec
RG   57
DW   60.400 usec
DE   6.50 usec
TE   296.3 K
D1   1.00000000 sec
TD0   1

======== CHANNEL f1 ========
NUC1   1H
P1   11.00 usec
PL1   0 dB
SFO1   400.1324710 MHz

F2 - Processing parameters
SI   32768
SF   400.1300066 MHz
WDW   EM
SSB   0
LB   0.30 Hz
GB   0
PC   1.00

10 9 8 7 6 5 4 3 2 1 0 ppm

6.85 1.00 1.84 1.01 1.00 1.02 3.99 3.08 3.15 2.80 3.16 28.50

N-BOM-3',5'-TIPDS-2'-Ms-MeU
CDCl3
CNMR

171.17
163.11
151.11
137.82
133.30
128.33
127.77
127.72
111.23
87.19
83.01
77.23
74.75
72.20
70.59
60.87
60.40
38.70
38.53
21.06
17.33
17.31
17.26
17.21
17.17
14.20
13.40
13.22
13.18
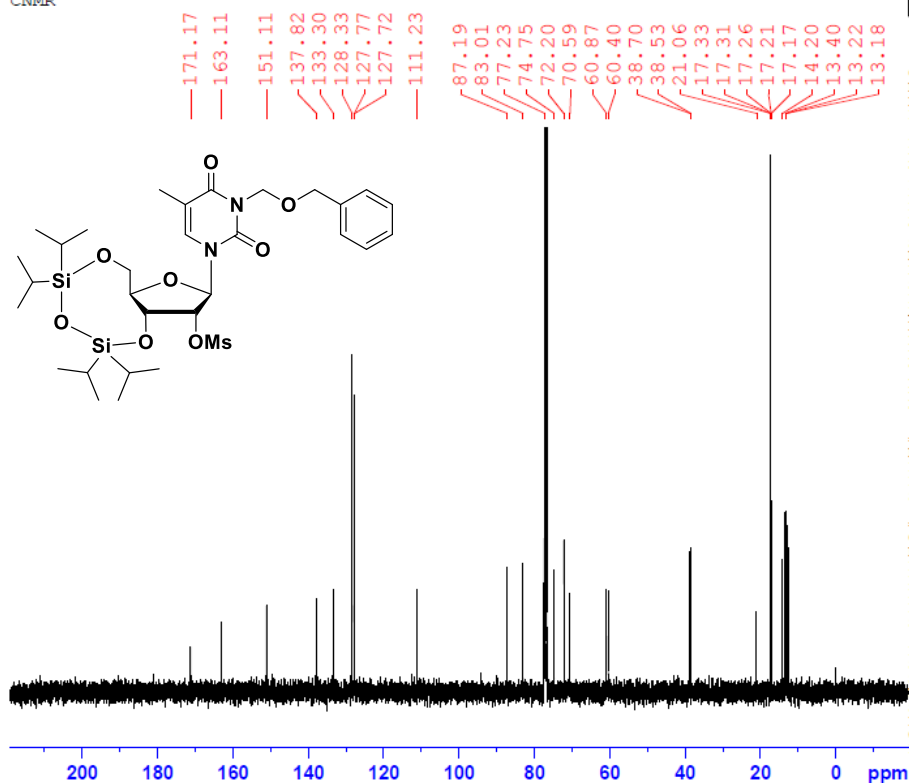
Current Data Parameters
NAME   20190120_BOMTiMs-T
EXPNO   2
PROCNO   1

F2 - Acquisition Parameters
Date_   20190121
Time   10.35
INSTRUM   spect
PROBHD   5 mm PABBO BB-
PULPROG   zgpg30
TD   65536
SOLVENT   CDCl3
NS   202
DS   4
SWH   23980.814 Hz
FIDRES   0.365918 Hz
AQ   1.3664256 sec
RG   1149.4
DW   20.850 usec
DE   6.50 usec
TE   296.8 K
D1   2.00000000 sec
D11   0.03000000 sec
TD0   1
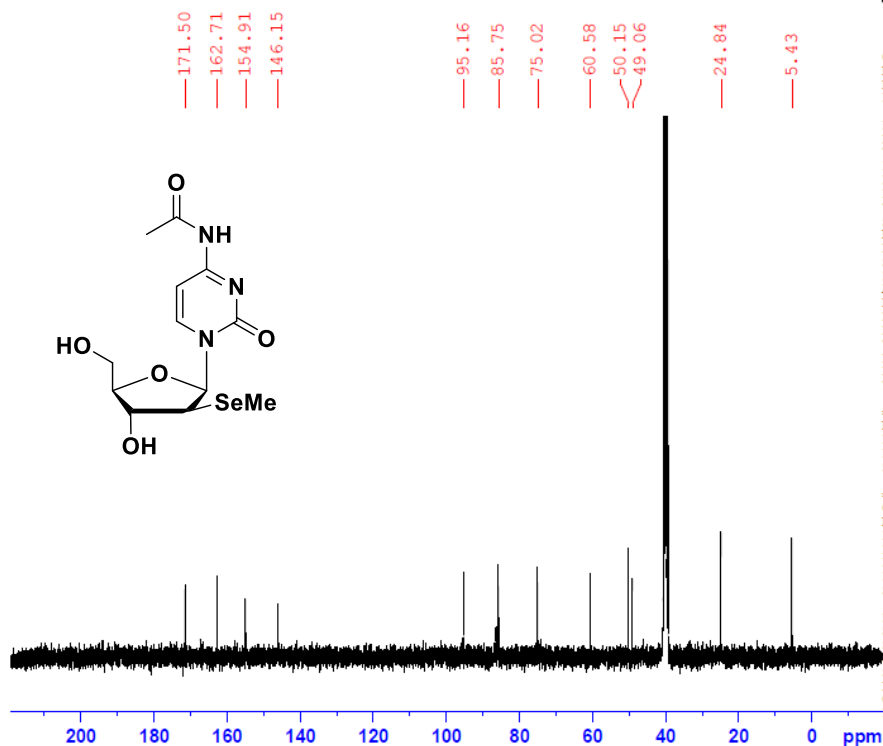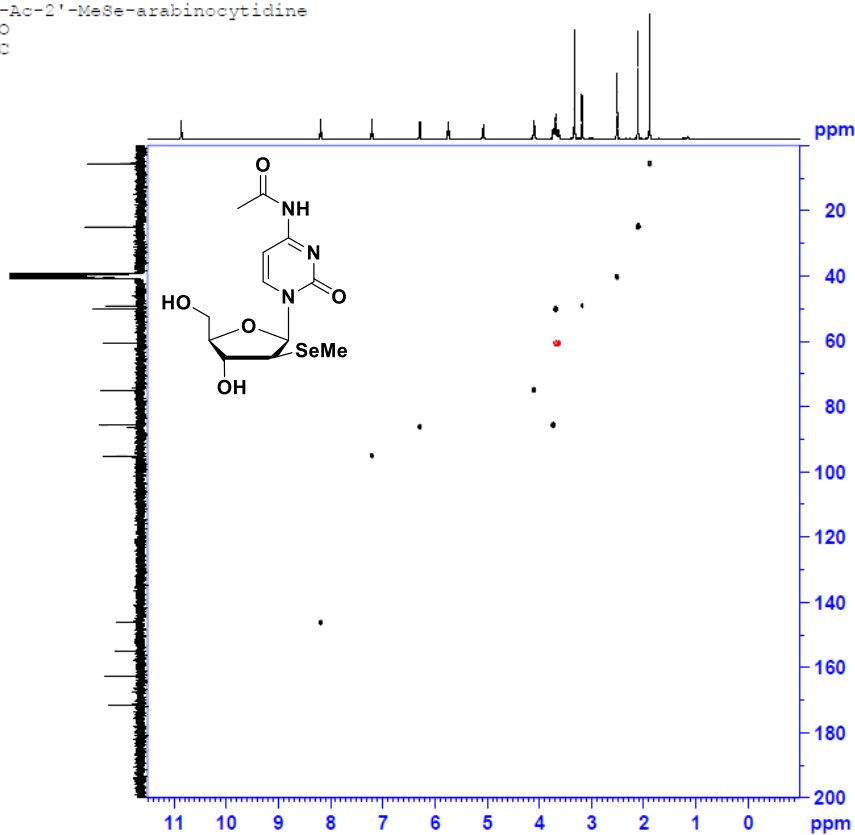
======== CHANNEL f1 ========
NUC1   13C
P1   9.00 usec
PL1   3.00 dB
SFO1   100.6228298 MHz

======== CHANNEL f2 ========
CPDPRG[2   waltz16
NUC2   1H
PCPD2   100.00 usec
PL2   0 dB
PL12   19.17 dB
PL13   19.00 dB
SFO2   400.1316005 MHz

F2 - Processing parameters
SI   32768
SF   100.6127690 MHz
WDW   EM
SSB   0
LB   1.00 Hz
GB   0
PC   1.40

200 180 160 140 120 100 80 60 40 20 0 ppm

N-BOM-3',5'-TIPDS-2'-Ms-MeU
CDCl3
HSQC

N-BOM-3',5'-TIPDS-2'-Ms-MeU
CDCl3
COSY

N-BOM-3',5'-TIPDS-2'-Ms-MeU
CDCl3
HMBC

N-BOM-3',5'-TIPDS-2'-beta-MeSe-MeU
CDCl3
HNMR

N-BOM-3',5'-TIPDS-2'-beta-MeSe-MeU
CDCl3
CNMR

163.45
151.27
137.96
134.26
128.28
127.76
127.62
109.89
83.54
73.65
71.86
70.44
60.03
50.17
17.44
17.42
17.36
17.27
17.12
17.03
13.91
13.32
13.01
12.73
12.49

N-BOM-3',5'-TIPDS-2'-beta-MeSe-MeU
CDCl3
HSQC

N-BOM-3',5'-TIPDS-2'-beta-MeSe-MeU
CDCl3
COSY

N-BOM-3',5'-TIPDS-2'-beta-MeSe-MeU
CDCl3
HMBC

3',5'-TIPDS-2'-beta-MeSe-MeU
CDCl3
HNMR

8.345
7.264 7.249
6.372 6.354
5.536 5.515
5.266 5.244
4.266 4.244 4.222
4.149 4.131 4.113 4.097
4.063 4.030
3.725 3.707
3.633 3.612 3.591
2.090
2.045 1.927
1.118 1.092 1.050 1.039
0.000

1.00 0.96 1.11 0.46 1.31 2.88 1.35 1.32 1.30 3.11 4.10 39.62

3',5'-TIPDS-2'-beta-MeSe-MeU
CDCl3
CNMR

163.39
150.31
135.59
110.50
84.52 83.53
77.22 73.70
60.39 60.00
49.88
21.05
17.42 17.36 17.26 17.11 17.01
14.20 13.90
13.00 12.73 12.56 12.49

2'-MeSe-arabinothymidine
DMSO
HNMR

7.676, 7.674, 6.243, 6.225, 5.678, 5.663, 5.162, 5.150, 5.137, 5.141, 4.124, 4.103, 4.087, 3.755, 3.742, 3.731, 3.719, 3.654, 3.642, 3.633, 3.615, 3.608, 3.332, 2.512, 2.508, 2.504, 1.990, 1.768

Integrals: 1.01, 1.00, 1.02, 1.00, 1.09, 1.08, 3.09, 3.07, 3.03

2'-MeSe-arabinothymidine
DMSO
CNMR

164.12, 150.87, 137.14, 108.87, 85.06, 84.80, 73.88, 59.66, 49.89, 12.75, 5.21

2'-MeSe-arabinothymidine
DMSO
HSQC

2'-MeSe-arabinothymidine
DMSO
HMBC

2'-MeSe-arabinothymidine
DMSO
COSY

ppm

-1

0

1

2

3

4

5

6

7

8

9

10

10 9 8 7 6 5 4 3 2 1 0 -1 ppm
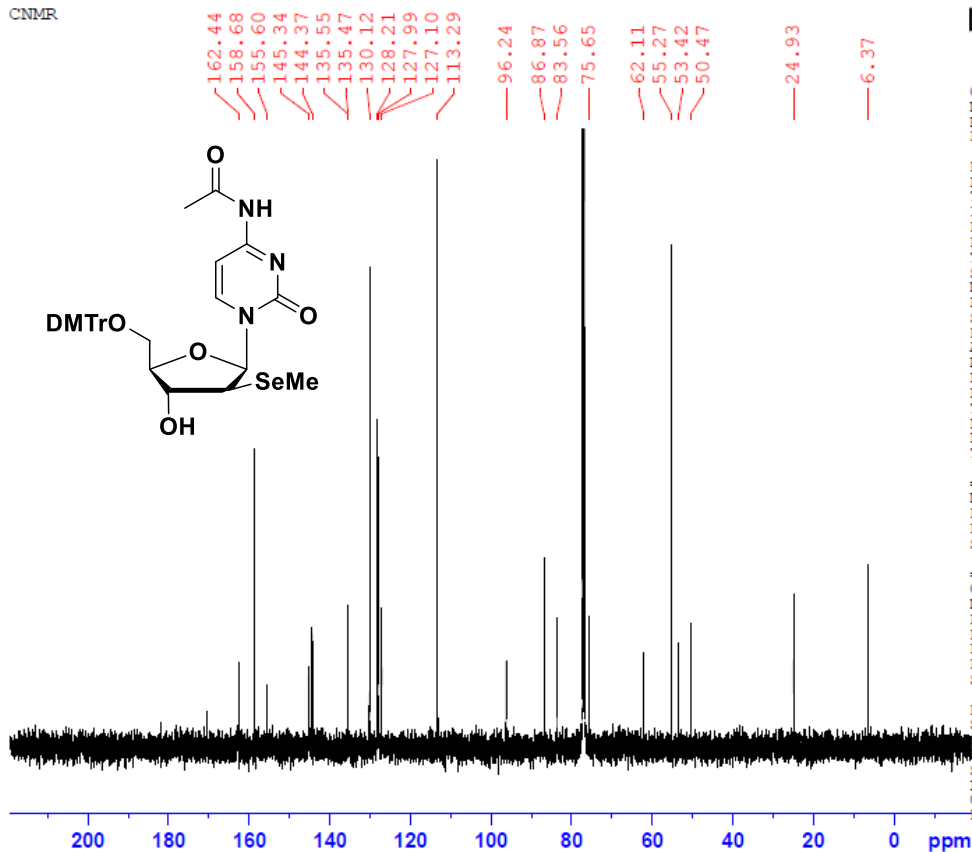
Current Data Parameters
NAME   20190802_2'bSe-T
EXPNO   6
PROCNO   1

F2 - Acquisition Parameters
Date_   20190808
Time   13.56
INSTRUM   spect
PROBHD   5 mm PABBO BB-
PULPROG   cosygpqf
TD   2048
NS   2
DS   8
SWH   4789.272 Hz
FIDRES   2.338512 Hz
AQ   0.2138112 sec
RG   574.7
DW   104.400 usec
DE   6.50 usec
TE   300.0 K
D0   0.00000300 sec
D1   1.45686196 sec
D13   0.00000400 sec
D16   0.00020000 sec
IN0   0.00020825 sec

======== CHANNEL f1 ========
NUC1   1H
P0   11.00 usec
P1   11.00 usec
PL1   0 dB
SFO1   400.1318006 MHz

====== GRADIENT CHANNEL =====
GPNAM[1]   SINE.100
GPZ1   10.00 %
P16   1000.00 usec

F1 - Acquisition parameters
TD   128
SFO1   400.1318 MHz
FIDRES   73.024912 Hz
SW   12.000 ppm
FnMODE   QF

F2 - Processing parameters
SI   1024
SF   400.1300000 MHz
WDW   SINE
SSB   0
LB   0 Hz
GB   0
PC   1.00

F1 - Processing parameters
SI   1024
MC2   QF
SF   400.1300000 MHz
WDW   SINE
SSB   0
LB   0 Hz
GB   0

2'bMeSe-DMTr-MeU
CDCl3
HNMR

8.751
7.458
7.455
7.437
7.392
7.390
7.351
7.346
7.340
7.329
7.323
7.302
7.284
7.274
7.257
6.872
6.850
6.419
6.401
4.151
4.133
3.892
3.883
3.873
3.809
3.617
3.615
3.596
3.563
3.497
2.664
2.654
2.101
2.088
2.074
2.065
1.727
1.612
1.611
1.297
1.280
1.262
0.022

10 9 8 7 6 5 4 3 2 1 0 ppm

1.07   2.92 9.68 4.49 1.00   1.23 1.64 2.19 5.90 2.38 1.05 1.08 5.39 1.46 2.98 2.22
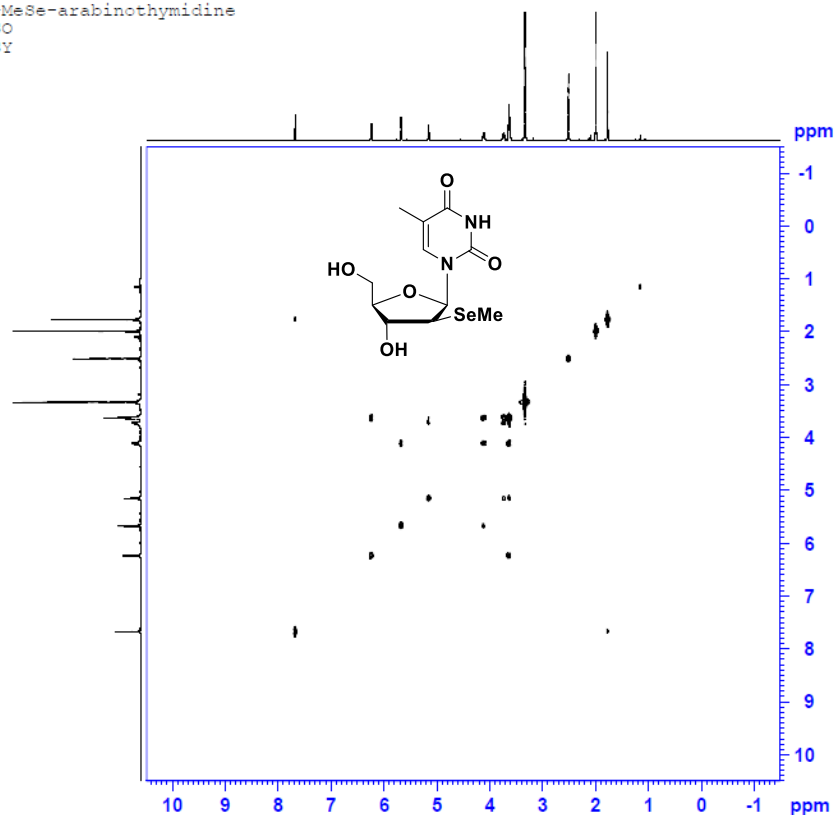
Current Data Parameters
NAME   20190521_2'bSeD-T
EXPNO   1
PROCNO   1

F2 - Acquisition Parameters
Date_   20190521
Time   14.21
INSTRUM   spect
PROBHD   5 mm PABBO BB-
PULPROG   zg30
TD   65536
SOLVENT   CDCl3
NS   16
DS   2
SWH   8278.146 Hz
FIDRES   0.126314 Hz
AQ   3.9583745 sec
RG   181
DW   60.400 usec
DE   6.50 usec
TE   299.1 K
D1   1.00000000 sec
TD0   1

======== CHANNEL f1 ========
NUC1   1H
P1   11.00 usec
PL1   0 dB
SFO1   400.1324710 MHz

F2 - Processing parameters
SI   32768
SF   400.1300000 MHz
WDW   EM
SSB   0
LB   0.30 Hz
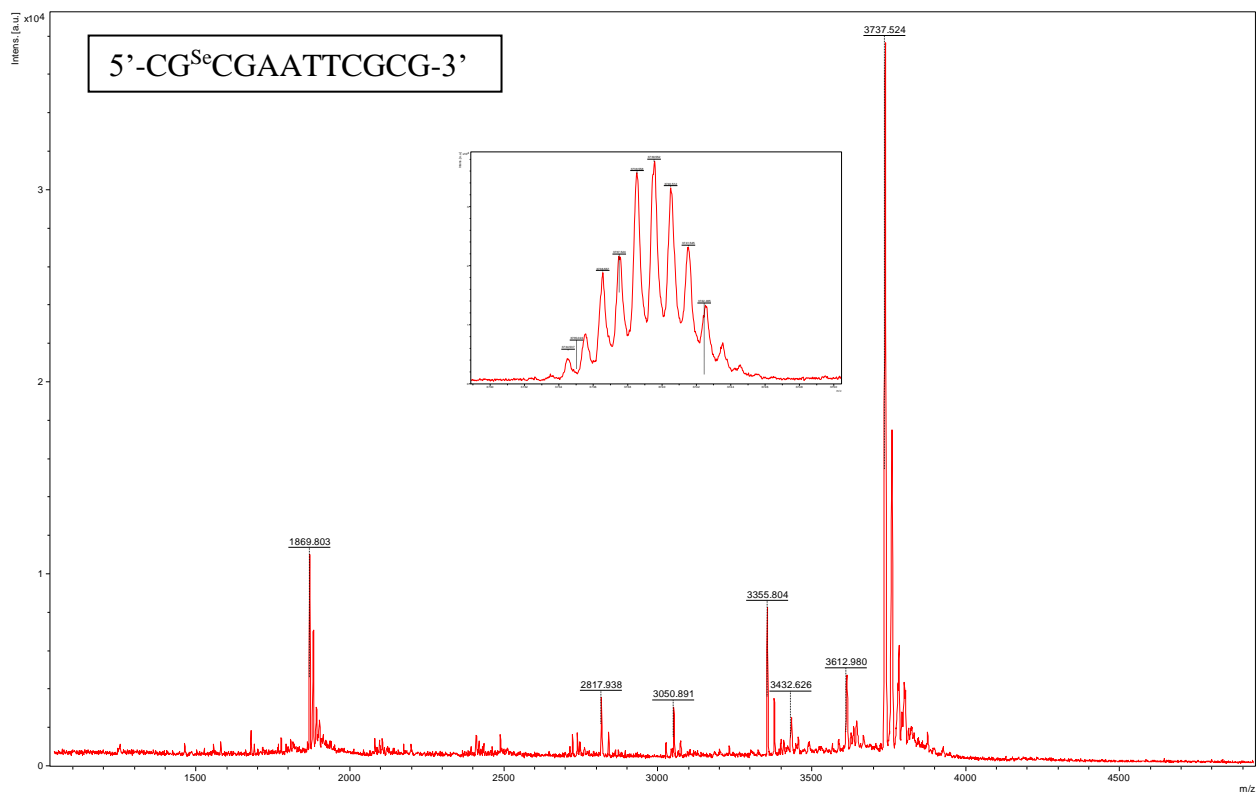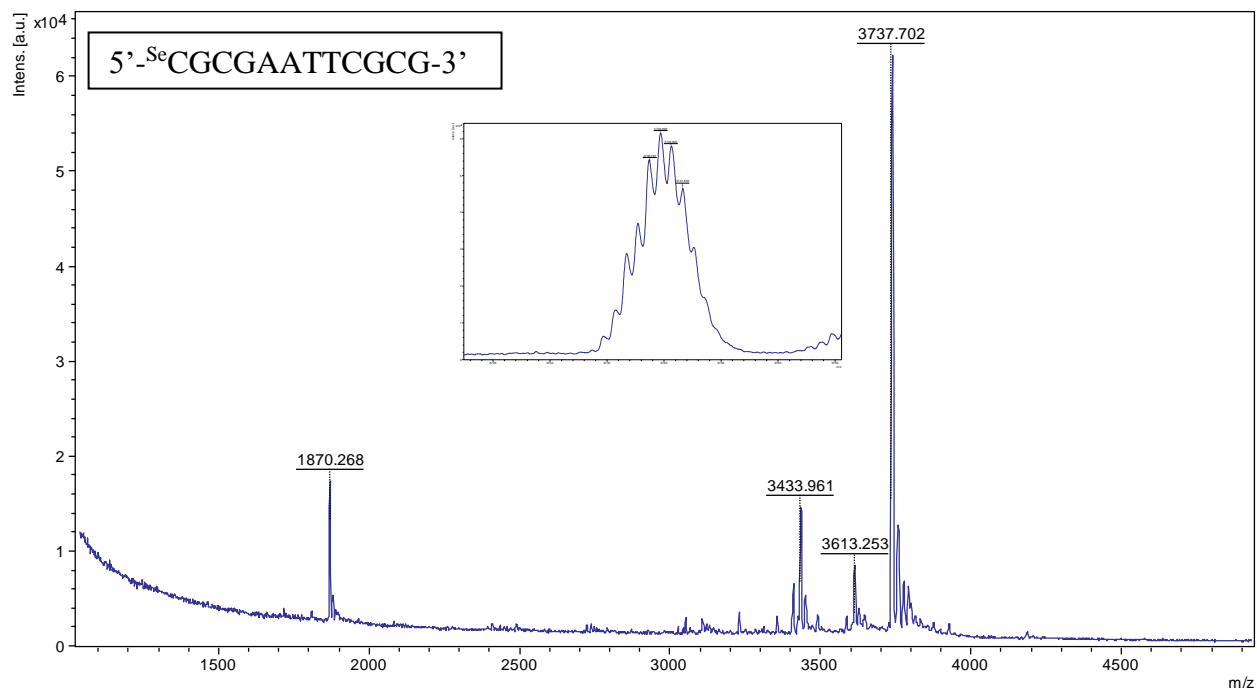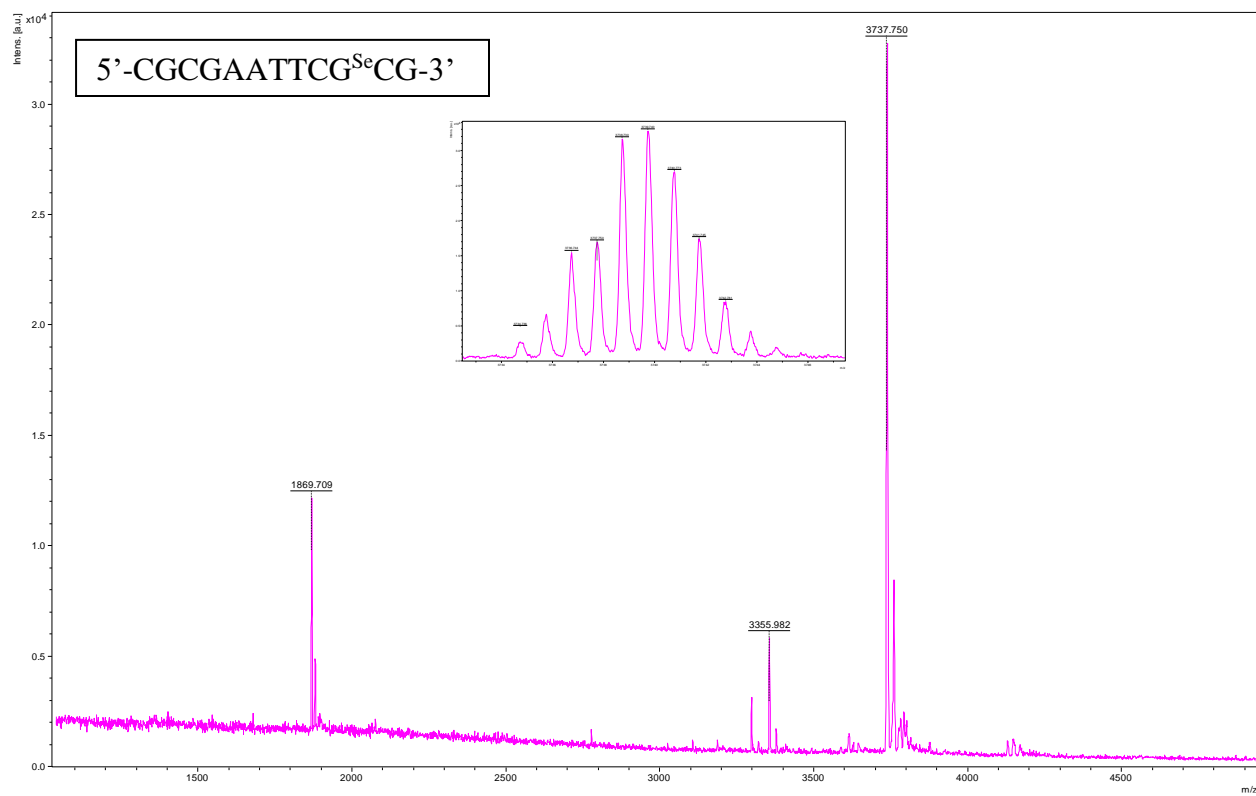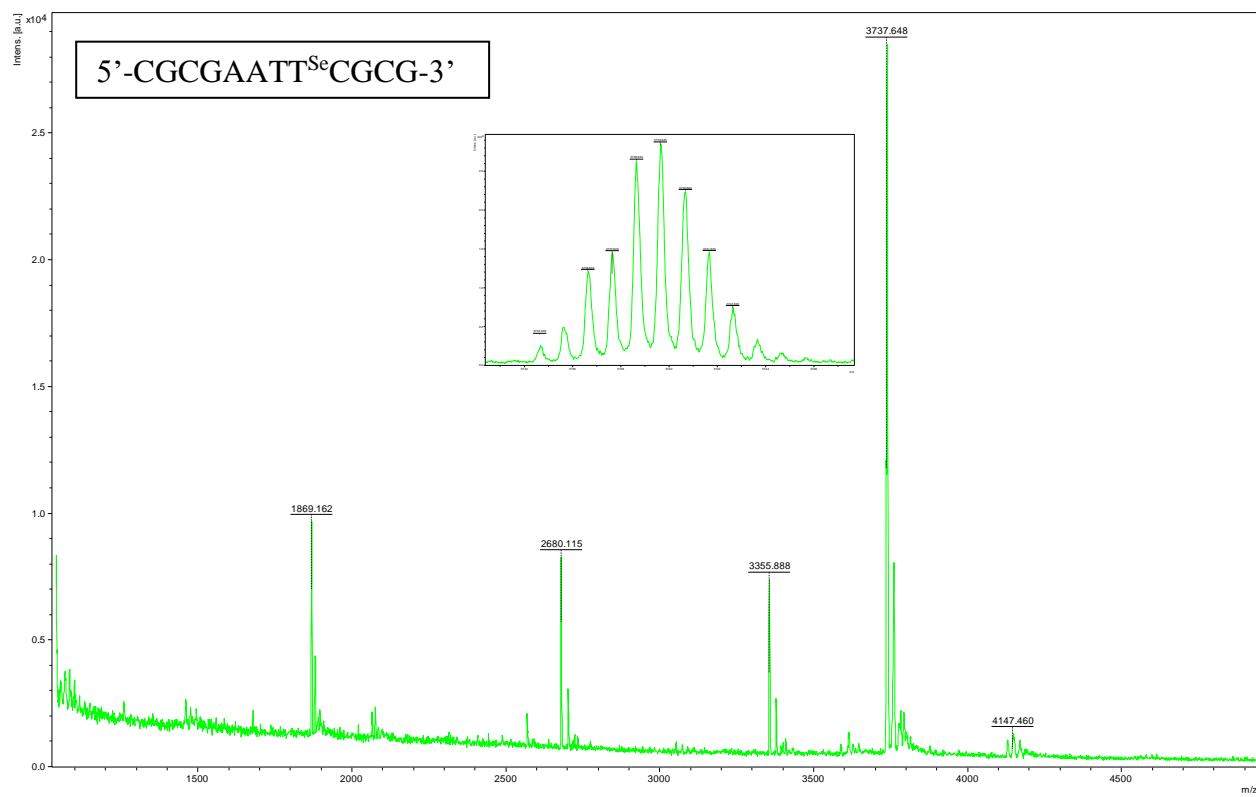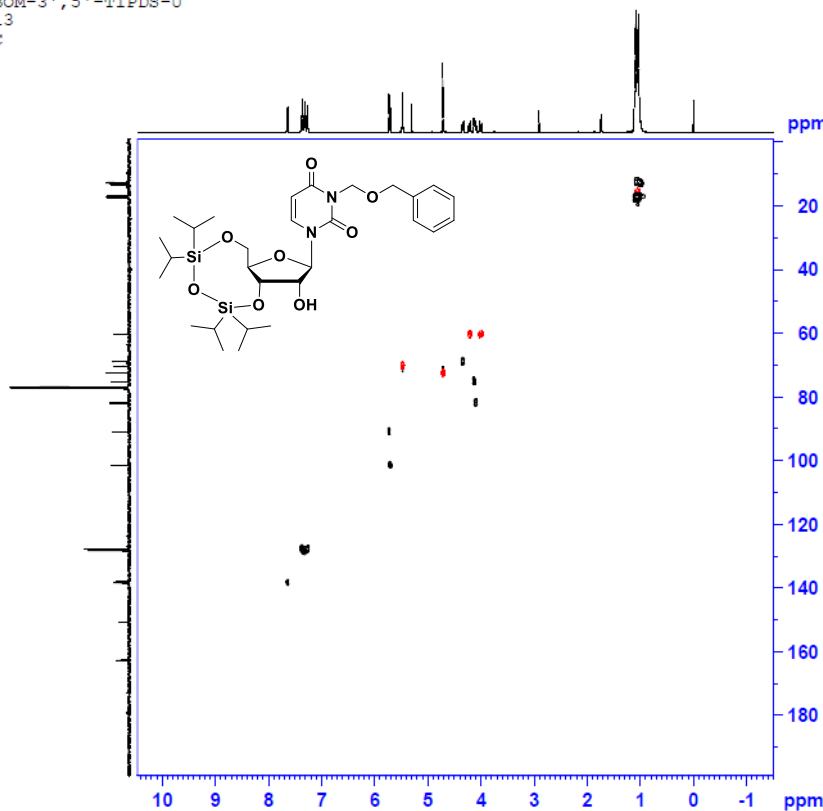GB   0
PC   1.00

2'bMeSe-DMTr-MeU
CDCl3
CNMR

163.65
158.72
150.38
144.36
135.84
135.50
135.43
130.14
130.10
128.19
128.01
127.14
113.29
110.43
86.79
85.21
82.54
75.37
61.81
60.42
55.26
50.39
21.06
14.20
12.12
6.31

CCHEN_T_3_ESIPOS_HUANG_080519_ 171 (3.204) Cm (171:178)

TOF MS ES+
2.05e4

241.1605

100

102.1204

359.1137

357.1272

242.1983

355.1259 361.1262

219.2014

275.1640

459.3604

m/z

CCHEN_T_3_ESIPOS_HUANG_080519_ 171 (3.204) Cm (171:178)

TOF MS ES+
8.12e3

359.1137

357.1272

335.1463

355.1259

361.1262

320.4108
321.3559
331.1668
333.1585
337.1697

381.1184

m/z

TOF MS ES+
3.42e4

241.1270

DMTrO

SeMe

N

P

O

CN

459.3143

102.1041
219.1935
242.1899
320.3530
331.2648
275.1640
441.3577
460.3647
553.3031
861.4891
859.5304 862.5325
103.1299

861.4891
859.5304
858.5204
857.5109
862.5325
863.5449
864.5579
903.8497

5'-CGCGAAT^Se TCGCG-3'

3738.063

674.261
1559.256
2087.397
3341.315

5'-CGCGAA^{Se}TTCGCG-3'

Appendix D.2 NMR and MS Spectra for 2'-MeSe-arabinocytidine modification

3',5'-TIPDS-2'bSe-C
CDCl3
CNMR

3',5'-TIPDS-2'bSe-C
CDCl3
HSQC

3',5'-TIPDS-2'bSe-C
CDCl3
HMBC

3',5'-TIPDS-2'bSe-C
CDCl3
COSY
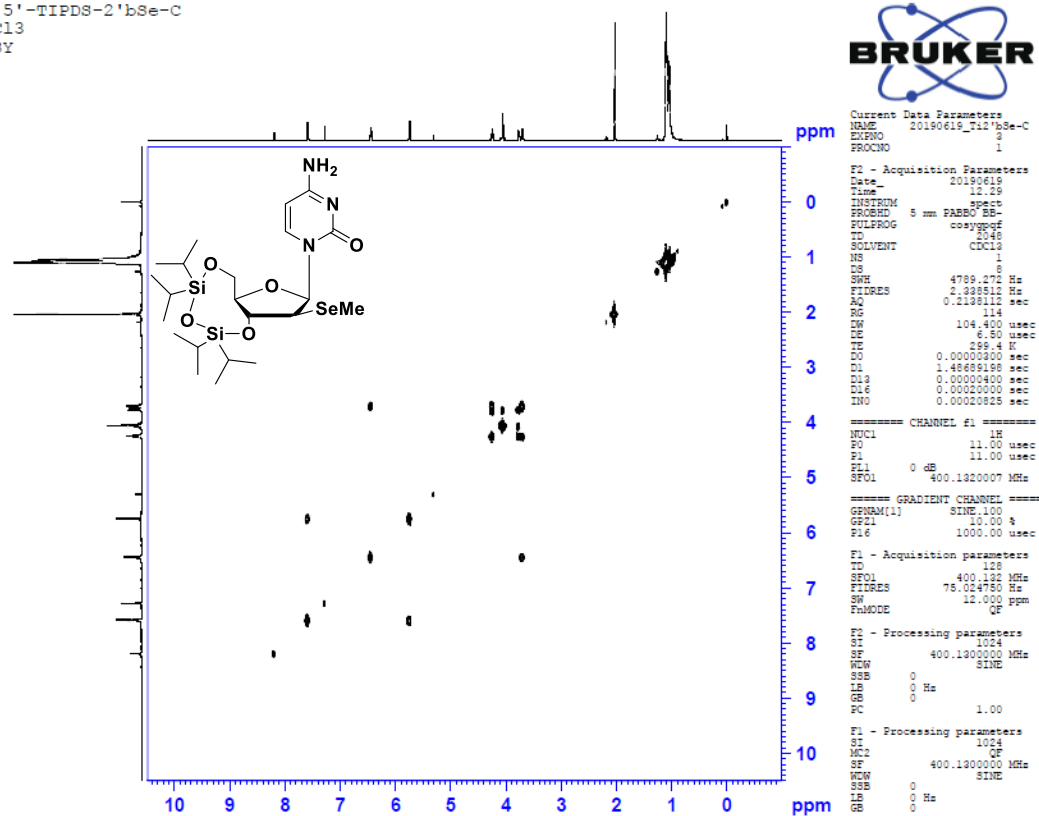
3',5'-TIPDS-4-N-Ac-2'-MeSe-arabinocytidine
CDCl3
HNMR

8.922 7.982 7.963 7.462 7.443 7.284 6.478 6.462 4.263 4.242 4.222 4.098 4.090 3.839 3.832 3.813 3.752 3.735 3.731 3.714 2.746 2.066 2.273 1.130 1.117 1.094 1.056 1.048

1.04 0.98 1.00 1.05 2.06 1.09 1.06 3.38 3.14 31.72

3',5'-TIPDS-4-N-Ac-2'-MeSe-arabinocytidine
CDCl3
CNMR

170.95 162.73 154.94 144.97 96.26 85.85 83.94 74.12 60.82 49.76 24.82 17.27 17.08 13.78 13.05 12.90 12.48 6.68

3',5'-TIPDS-4-N-Ac-2'-MeSe-arabinocytidine
CDC13
HSQC

ppm
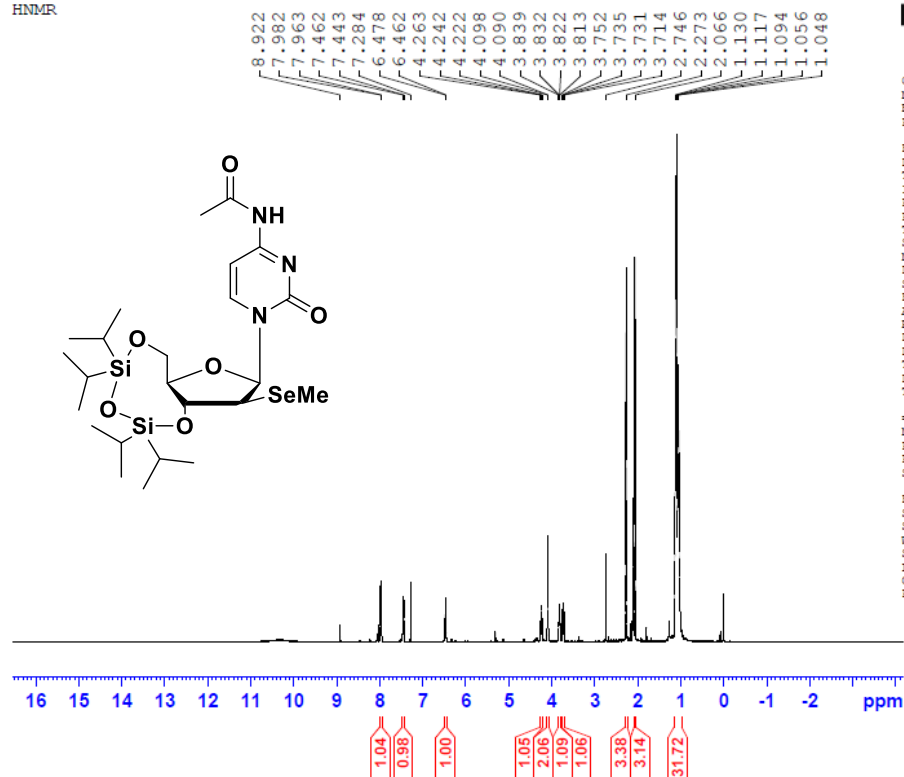
```
Current Data Parameters
NAME        20190813_AcTi2'bSe-C
EXPNO                3
PROCNO               1

F2 - Acquisition Parameters
Date_         20190813
Time              17.06
INSTRUM           spect
PROBHD   5 mm PABBO BB-
PULPROG      hsqcedetgp
TD                 1024
SOLVENT          CDC13
NS                    2
DS                   16
SWH          5592.841 Hz
FIDRES      5.461759 Hz
AQ          0.0915456 sec
RG             5792.6
DW            89.400 usec
DE             6.50 usec
TE            300.0 K
CNST2      145.0000000
D0         0.00000300 sec
D1         1.50000000 sec
D4         0.00172414 sec
D11        0.03000000 sec
D13        0.00000400 sec
D16        0.00020000 sec
D21        0.00345000 sec
ING        0.00002260 sec
ZGOPTNS

======== CHANNEL f1 ========
NUC1                 1H
P1             11.00 usec
P2             22.00 usec
P28            0 usec
PL1            0 dB
SFO1       400.1320007 MHz

======== CHANNEL f2 ========
CPDPRG[2           garp
NUC2                13C
P3              9.00 usec
P4             18.00 usec
PCPD2          70.00 usec
PL2             3.00 dB
PL12           20.82 dB
SFO2       100.6228303 MHz

====== GRADIENT CHANNEL ======
GPNAM[1]        SINE.100
GPNAM[2]        SINE.100
GPZ1           80.00 %
GPZ2           20.10 %
P16          1000.00 usec

F1 - Acquisition parameters
TD                  154
SFO1       100.6228 MHz
FIDRES      287.493805 Hz
SW          220.000 ppm
FnMODE    Echo-Antiecho

F2 - Processing parameters
SI                 1024
SF        400.1300000 MHz
WDW              QSINE
SSB                   2
LB             0 Hz
GB             0
PC             1.40

F1 - Processing parameters
SI                 1024
MC2        echo-antiecho
SF        100.6127690 MHz
WDW              QSINE
SSB                   2
LB             0 Hz
GB             0
```
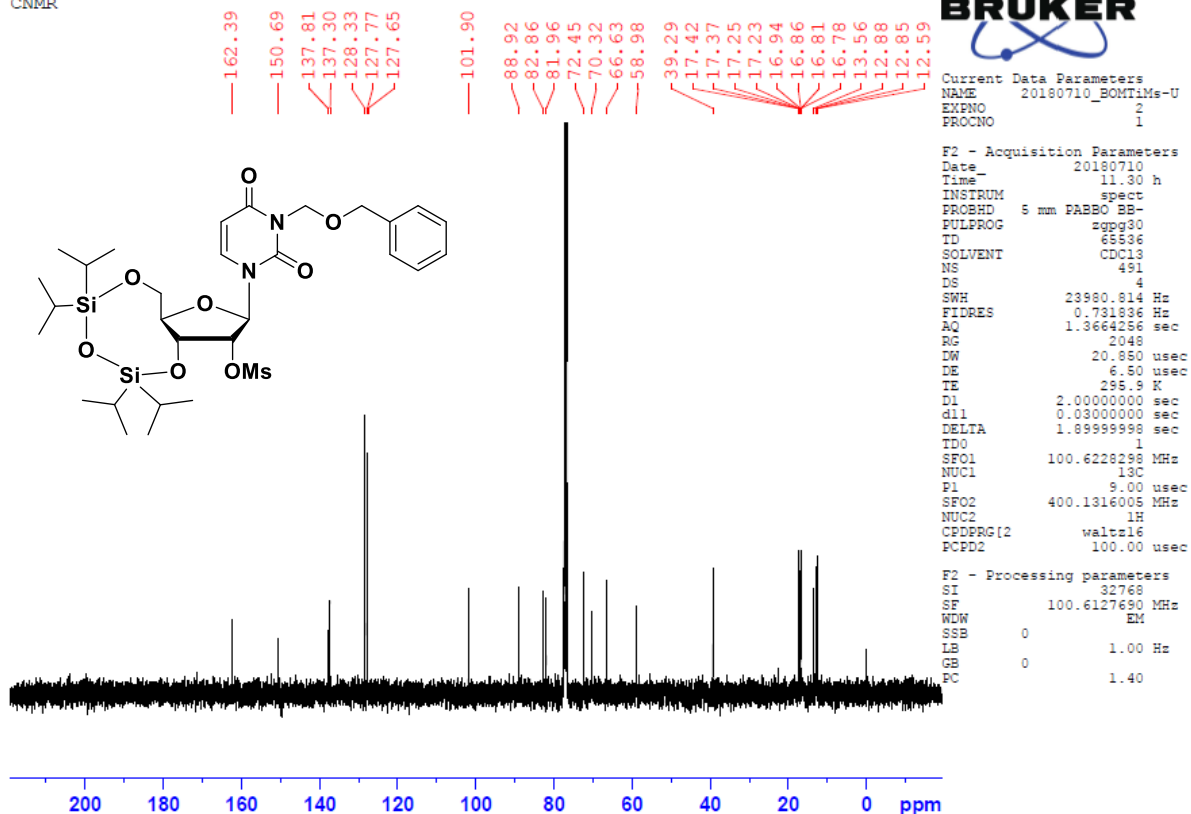
4-N-Ac-2'-MeSe-arabinocytidine
DMSO
HNMR

7.215
7.196
6.295
6.279
5.759
5.747
5.734
5.086
5.073
5.060
4.109
4.096
4.083
3.741
3.731
3.720
3.699
3.691
3.683
3.668
3.636
3.324
3.182
3.169
2.512
2.508
2.504
2.110
1.886
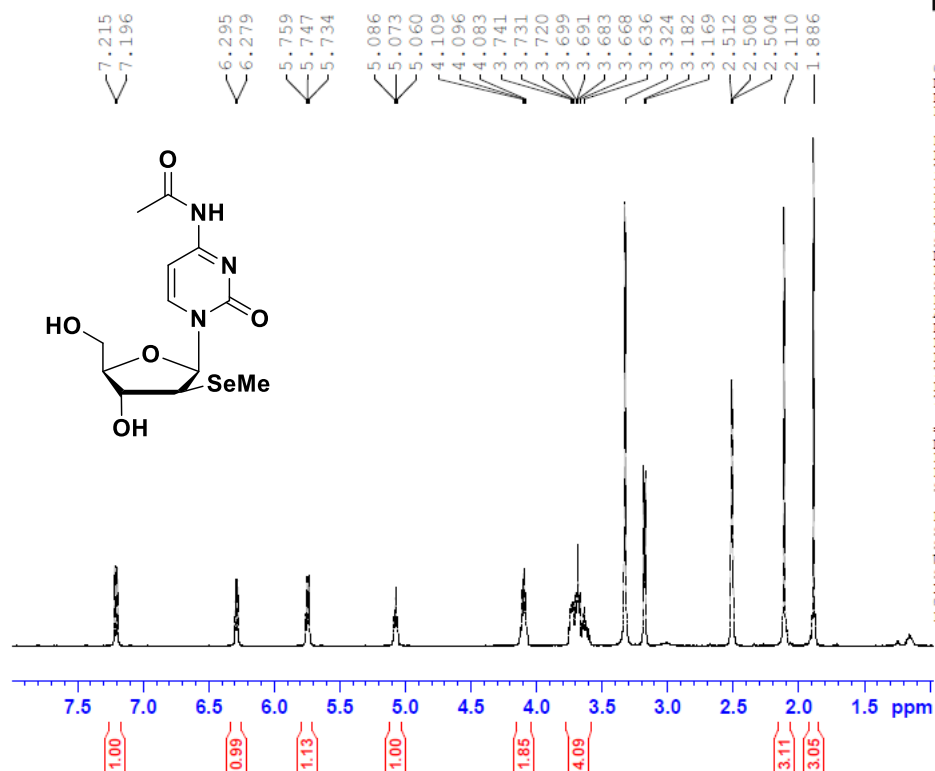

```
Current Data Parameters
NAME        20190819_Ac2'bSe-C
EXPNO                1
PROCNO               1

F2 - Acquisition Parameters
Date_         20190819
Time              16.54
INSTRUM           spect
PROBHD   5 mm PABBO BB-
PULPROG           zg30
TD                65536
SOLVENT           DMSO
NS                   16
DS                    2
SWH          8278.146 Hz
FIDRES      0.126314 Hz
AQ         3.9583745 sec
RG             322.5
DW            60.400 usec
DE             6.50 usec
TE            300.4 K
D1         1.00000000 sec
TD0                   1

======== CHANNEL f1 ========
NUC1                 1H
P1             11.00 usec
PL1            0 dB
SFO1       400.1324710 MHz

F2 - Processing parameters
SI                32768
SF        400.1300000 MHz
WDW                 EM
SSB                   0
LB             0.30 Hz
GB             0
PC             1.00
```
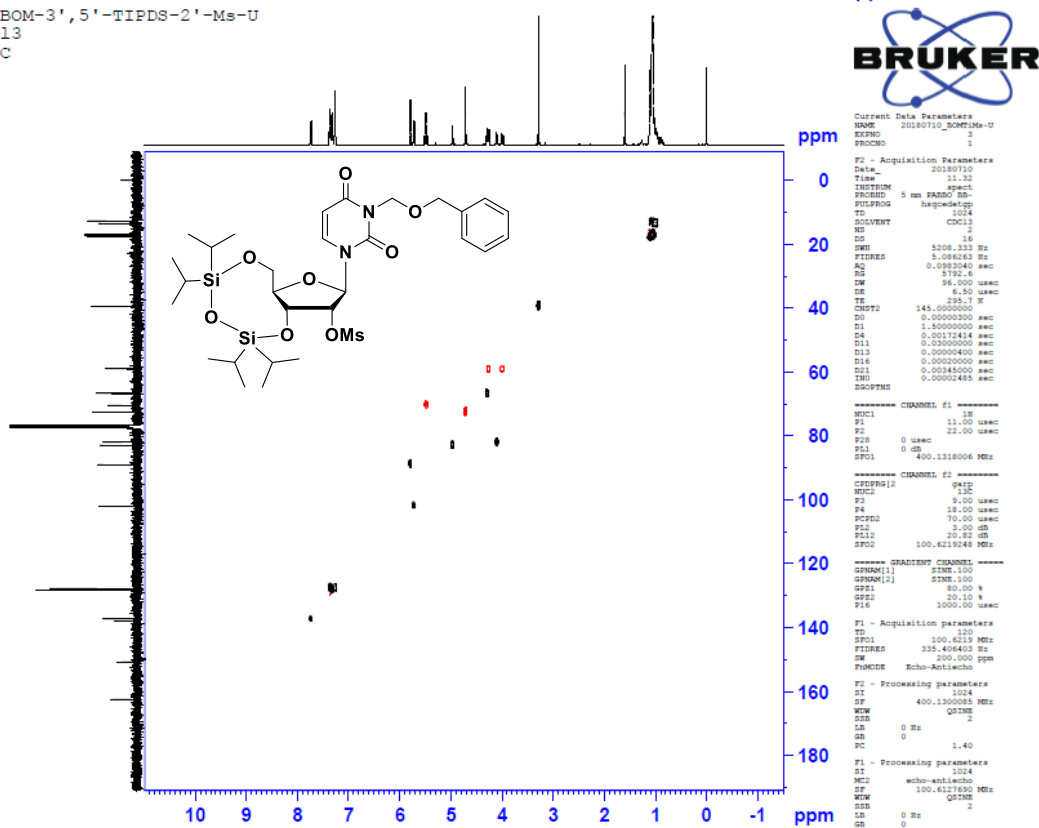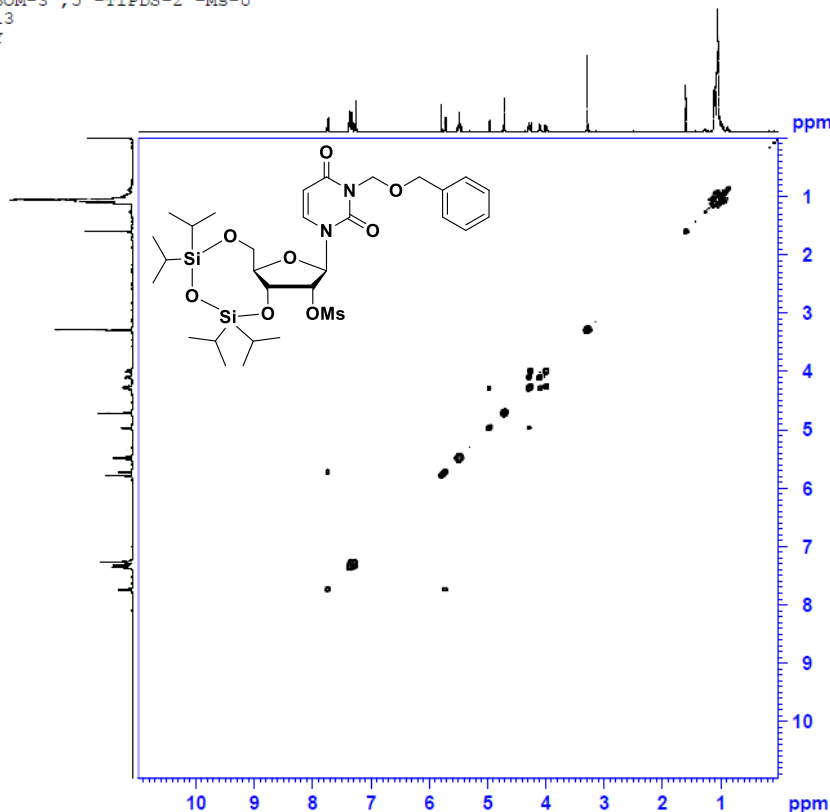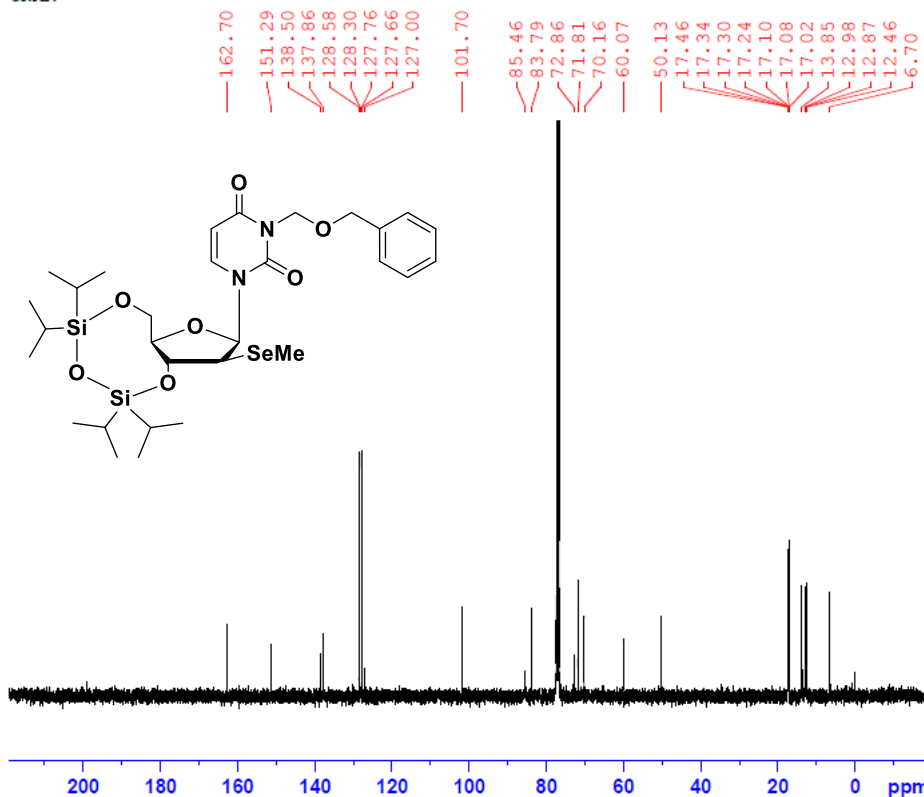
1.00 0.99 1.13 1.00 1.85 4.09 3.11 3.05

4-N-Ac-2'-MeSe-arabinocytidine
DMSO
CNMR

171.50 162.71 154.91 146.15   95.16 85.75 75.02 60.58 50.15 49.06   24.84   5.43

4-N-Ac-2'-MeSe-arabinocytidine
DMSO
HSQC

4-N-Ac-2'-MeSe-arabinocytidine
DMSO
HMBC

4-N-Ac-2'-MeSe-arabinocytidine
DMSO
COSY

5'-DMTr-4-N-Ac-2'-MeSe-arabinocytidine
CDCl3
HNMR

9.320
8.168
8.150
7.462
7.443
7.357
7.335
7.319
7.296
7.293
7.284
6.888
6.866
6.554
6.538
6.318
5.281
4.266
4.052
4.042
4.033
4.027
4.017
3.952
3.936
3.920
3.829
3.591
3.583
3.564
3.556
3.516
3.505
3.489
3.478
3.446
3.437
2.254

5'-DMTr-4-N-Ac-2'-MeSe-arabinocytidine
CDCl3
CNMR

162.44
158.68
155.60
145.34
144.37
135.55
135.47
130.12
128.21
127.99
127.10
113.29
96.24
86.87
83.56
75.65
62.11
55.27
53.42
50.47
24.93
6.37

5'-SeCGCGAATTCGCG-3'

5'-CGSeCGAATTCGCG-3'

5'-CGCGAATT^{Se}CGCG-3'

5'-CGCGAATTCG^{Se}CG-3'

Appendix D.3 NMR and MS spectra for 2'-MeSe-arabinouridine modification

3N-BOM-3',5'-TIPDS-U
CDC13
HSQC

Current Data Parameters
NAME   20180708_BOMTi-U
EXPNO   3
PROCNO   1

F2 - Acquisition Parameters
Date_   20180708
Time_   11.09
INSTRUM   spect
PROBHD 5 mm PABBO BB-
PULPROG   hsqcedetgp
TD   1024
NS   2
DS   16
SWH   4789.272 Hz
FIDRES   4.677023 Hz
AQ   0.1069056 sec
RG   5792.6
DW   104.400 usec
DE   6.50 usec
TE   295.4 K
CNST2   145.0000000
D0   0.00000300 sec
D1   1.50000000 sec
D4   0.00172414 sec
D11   0.03000000 sec
D13   0.00000400 sec
D16   0.00020000 sec
D21   0.00345000 sec
IN0   0.00002485 sec
ZGOPTNS

======= CHANNEL f1 ========
NUC1   1H
P1   11.00 usec
P2   22.00 usec
P28   0 usec
PL1   0 dB
SFO1   400.1318006 MHz

======= CHANNEL f2 ========
CPDPRG[2   garp
NUC2   13C
P3   9.00 usec
P4   18.00 usec
PCPD2   70.00 usec
PL2   3.00 dB
PL12   20.82 dB
SFO2   100.6227297 MHz
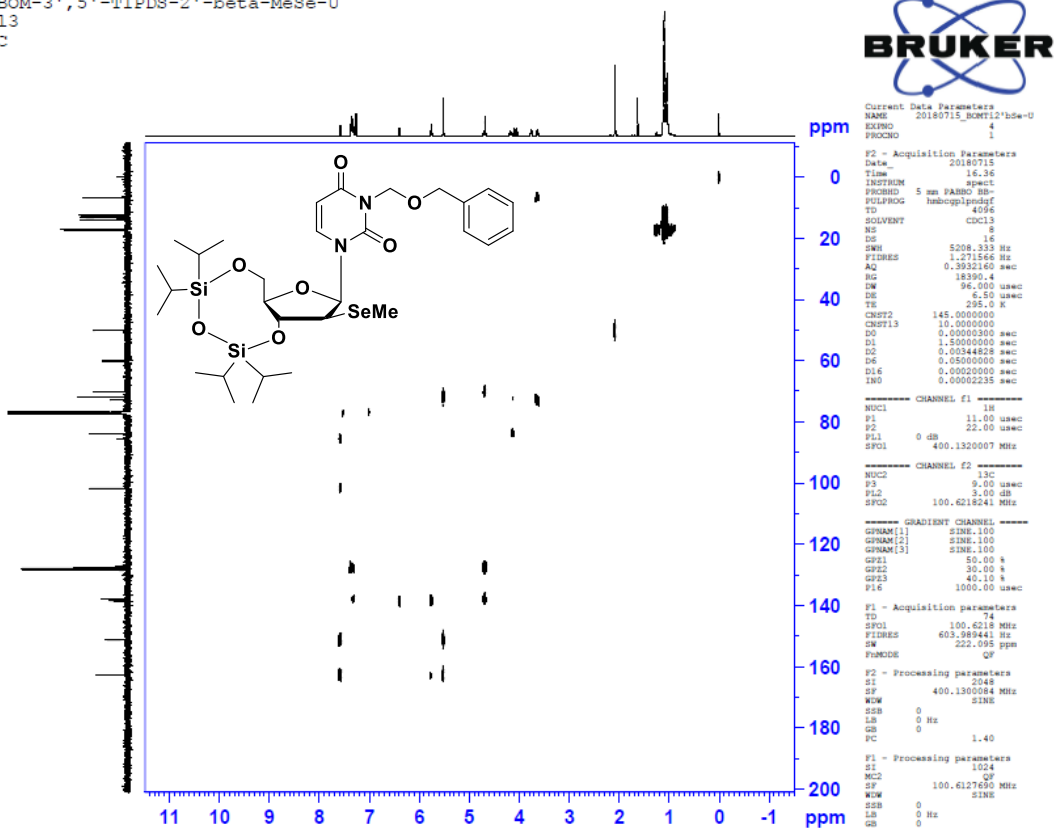
====== GRADIENT CHANNEL ======
GPNAM[1]   SINE.100
GPNAM[2]   SINE.100
GPZ1   80.00 %
GPZ2   20.10 %
P16   1000.00 usec

F1 - Acquisition parameters
TD   126
SFO1   100.6227 MHz
FIDRES   319.437225 Hz
SW   200.000 ppm
FnMODE   Echo-Antiecho

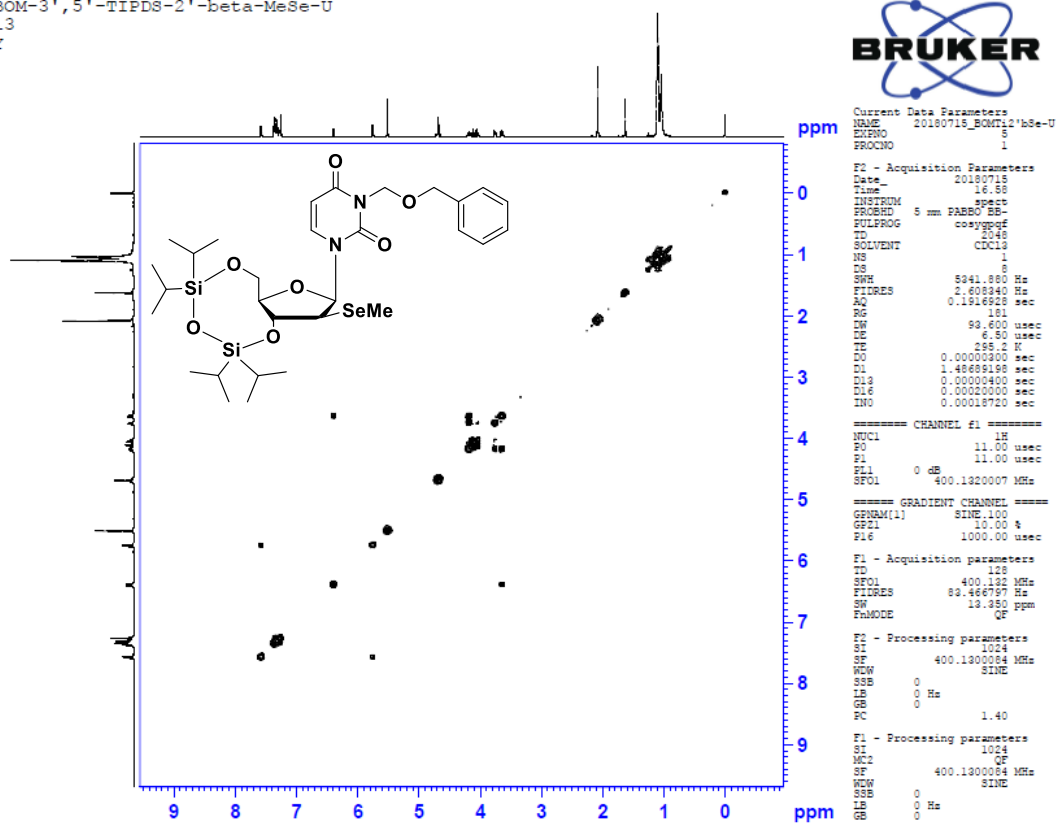F2 - Processing parameters
SI   1024
SF   400.1300078 MHz
WDW   QSINE
SSB   2
LB   0 Hz
GB   0
PC   1.40

F1 - Processing parameters
SI   1024
MC2   echo-antiecho
SF   100.6127690 MHz
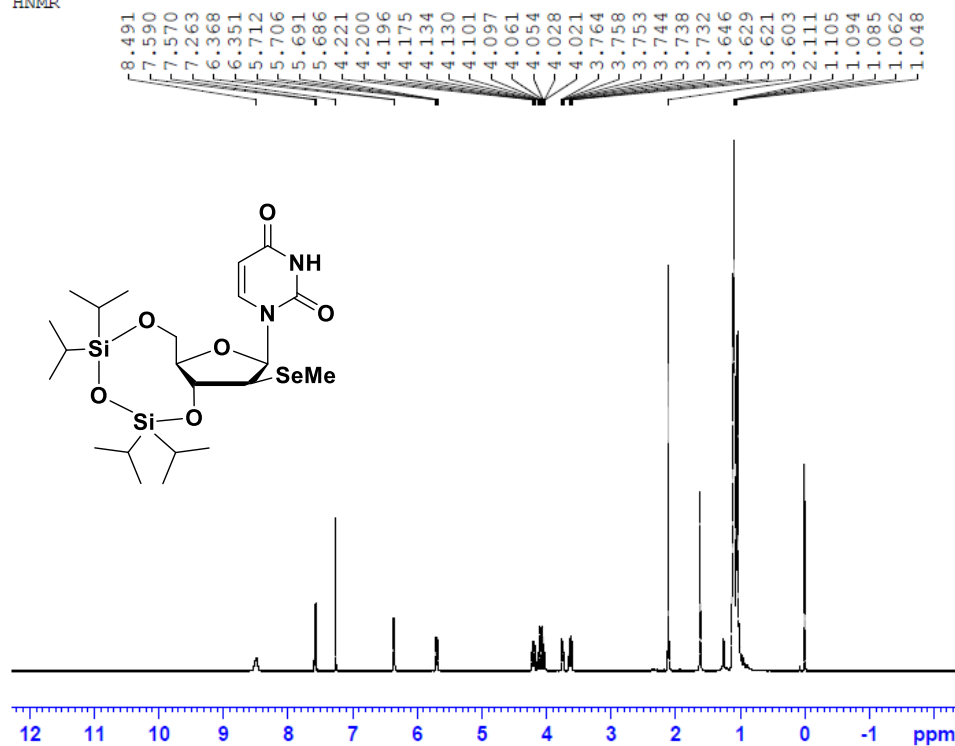WDW   QSINE
SSB   2
LB   0 Hz
GB   0

3N-BOM-3',5'-TIPDS-U
CDC13
HMBC
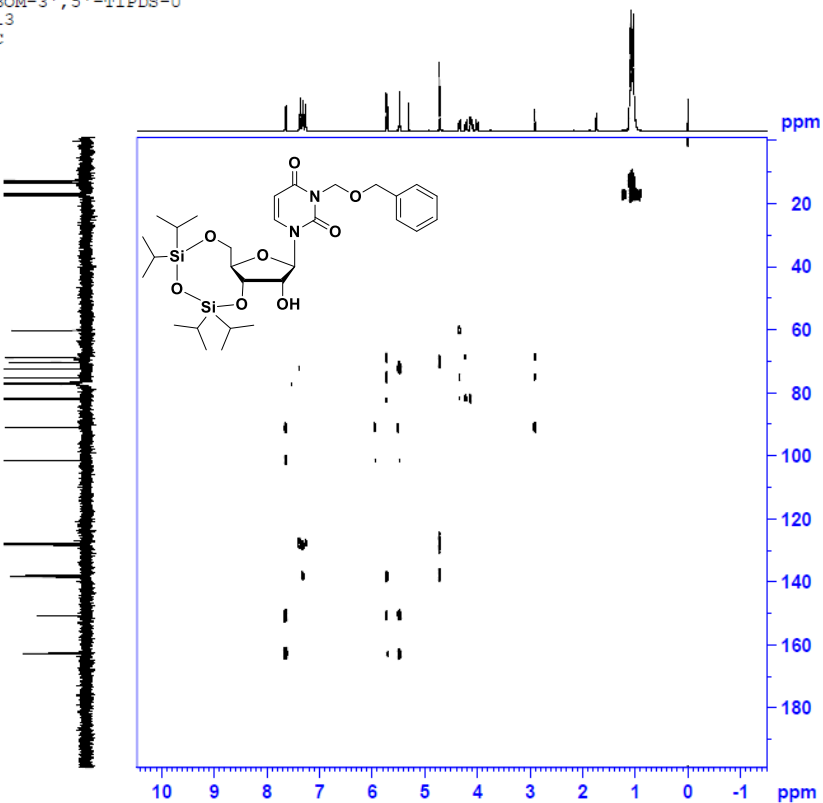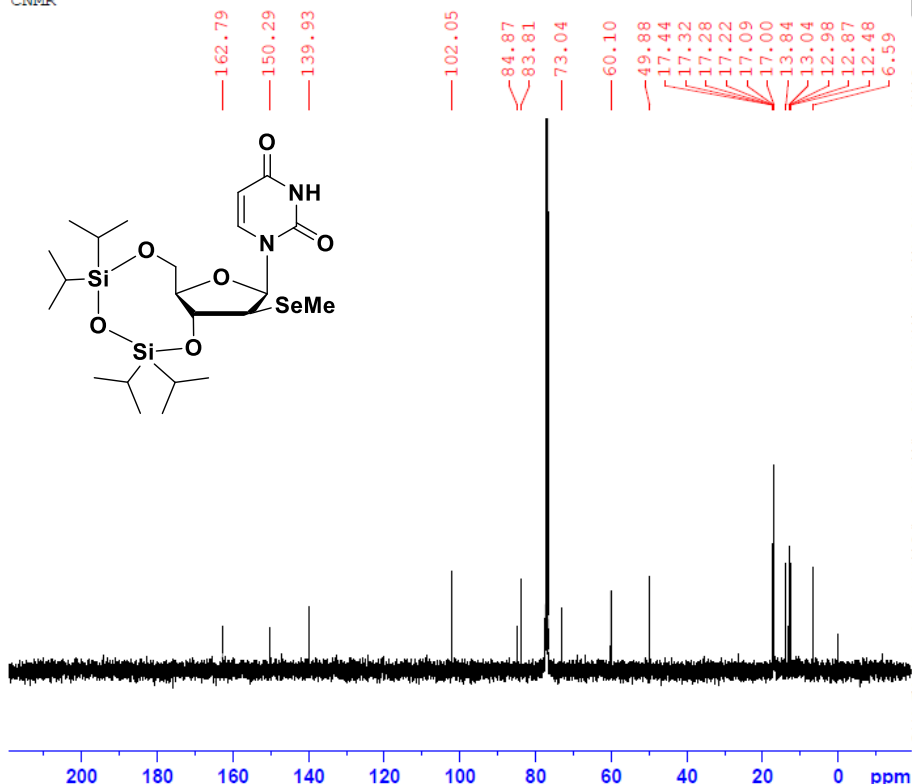
Current Data Parameters
NAME   20180708_BOMTi-U
EXPNO   5
PROCNO   1

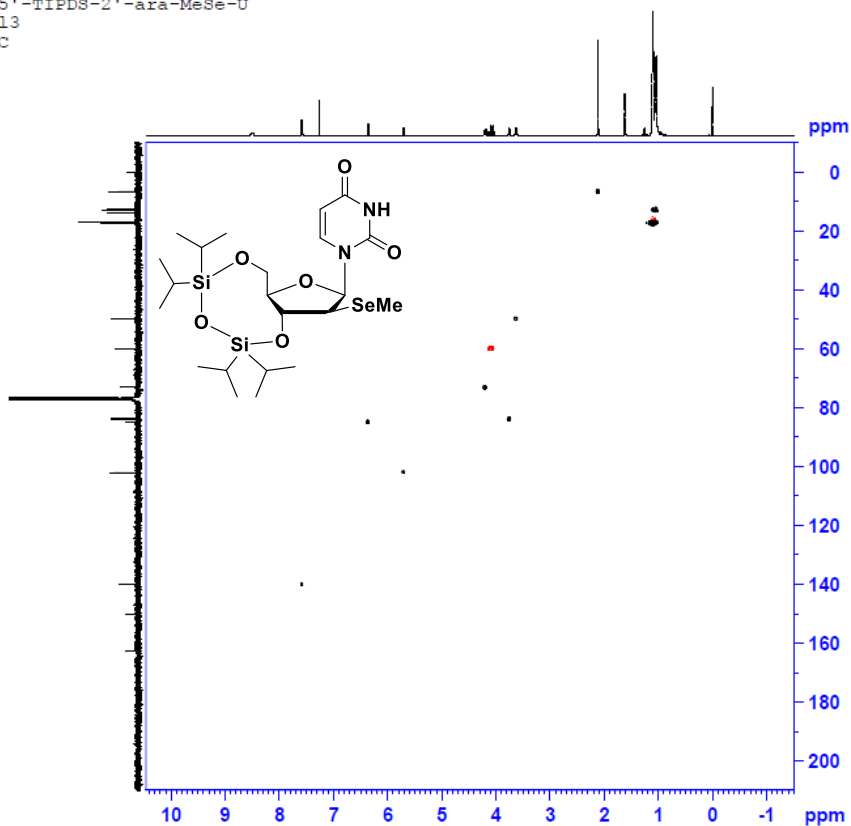F2 - Acquisition Parameters
Date_   20180708
Time_   11.24
INSTRUM   spect
PROBHD 5 mm PABBO BB-
PULPROG hmbcgplpndqf
TD   4096
SOLVENT   CDC13
NS   8
DS   16
SWH   4789.272 Hz
FIDRES   1.169256 Hz
AQ   0.4276224 sec
RG   16384
DW   104.400 usec
DE   6.50 usec
TE   295.3 K
CNST2   145.0000000
CNST13   10.0000000
D0   0.00000300 sec
D1   1.50000000 sec
D2   0.00344828 sec
D6   0.05000000 sec
D16   0.00020000 sec
IN0   0.00002485 sec

======= CHANNEL f1 ========
NUC1   1H
P1   11.00 usec
P2   22.00 usec
PL1   0 dB
SFO1   400.1318006 MHz

======= CHANNEL f2 ========
NUC2   13C
P3   9.00 usec
PL2   3.00 dB
SFO2   100.6227297 MHz

====== GRADIENT CHANNEL ======
GPNAM[1]   SINE.100
GPNAM[2]   SINE.100
GPNAM[3]   SINE.100
GPZ1   50.00 %
GPZ2   30.00 %
GPZ3   40.10 %
P16   1000.00 usec

F1 - Acquisition parameters
TD   64
SFO1   100.6227 MHz
FIDRES   628.892090 Hz
SW   200.000 ppm
FnMODE   QF

F2 - Processing parameters
SI   2048
SF   400.1300078 MHz
WDW   SINE
SSB   0
LB   0 Hz
GB   0
PC   1.40

F1 - Processing parameters
SI   1024
MC2   QF
SF   100.6127690 MHz
WDW   SINE
SSB   0
LB   0 Hz
GB   0

3N-BOM-3',5'-TIPDS-U
CDC13
COSY

3N-BOM-3',5'-TIPDS-2'-Ms-U
CDC13
HNMR

3N-BOM-3',5'-TIPDS-2'-Ms-U
CDC13
CNMR

162.39
150.69
137.81
137.30
128.33
127.77
127.65
101.90
88.92
82.86
81.96
72.45
70.32
66.63
58.98
39.29
17.42
17.37
17.25
17.23
16.94
16.86
16.81
16.78
13.56
12.88
12.85
12.59

3N-BOM-3',5'-TIPDS-2'-Ms-U
CDC13
HSQC

3N-BOM-3',5'-TIPDS-2'-Ms-U
CDCl3
COSY

3N-BOM-3',5'-TIPDS-2'-beta-MeSe-U
CDCl3
HNMR

3N-BOM-3',5'-TIPDS-2'-beta-MeSe-U
CDCl3
CNMR

Peak labels (top spectrum): 162.70, 151.29, 138.50, 137.86, 128.58, 128.30, 127.76, 127.66, 127.00, 101.70, 85.46, 83.79, 72.86, 71.81, 70.16, 60.07, 50.13, 17.46, 17.34, 17.30, 17.24, 17.10, 17.08, 17.02, 13.85, 12.98, 12.87, 12.46, 6.70

3N-BOM-3',5'-TIPDS-2'-beta-MeSe-U
CDCl3
HSQC

3N-BOM-3',5'-TIPDS-2'-beta-MeSe-U
CDCl3
HMBC

3N-BOM-3',5'-TIPDS-2'-beta-MeSe-U
CDCl3
COSY

3',5'-TIPDS-2'-ara-MeSe-U
CDCl3
HNMR

3',5'-TIPDS-2'-ara-MeSe-U
CDCl3
CNMR

3',5'-TIPDS-2'-ara-MeSe-U
CDC13
HSQC

3',5'-TIPDS-2'-ara-MeSe-U
CDC13
HMBC

3',5'-TIPDS-2'-ara-MeSe-U
CDC13
COSY

TOF MS ES-

in MeOH+0..5%NH3 588.8971 as ITSD
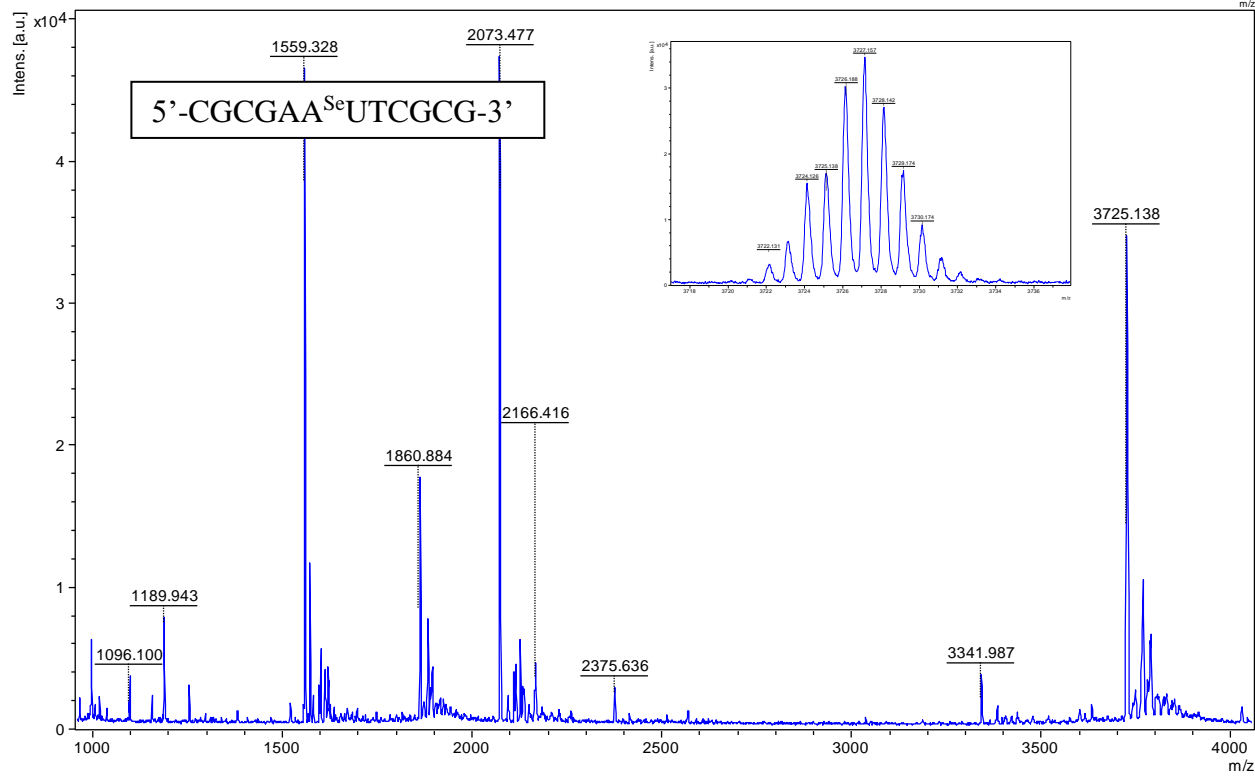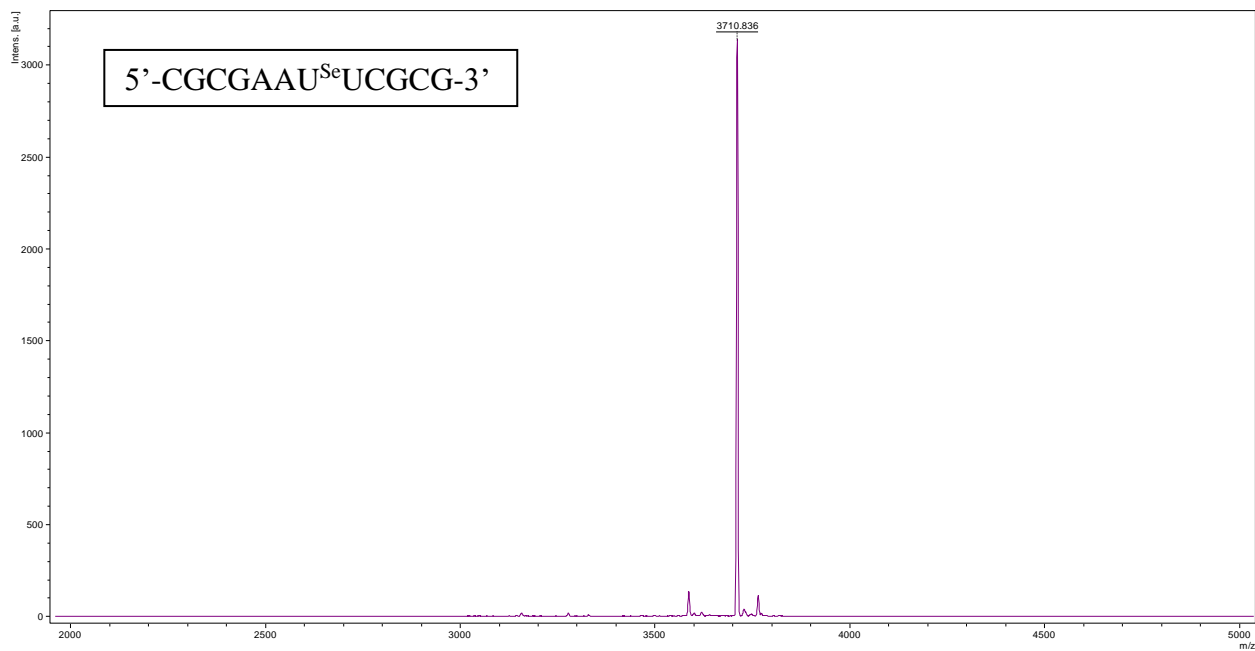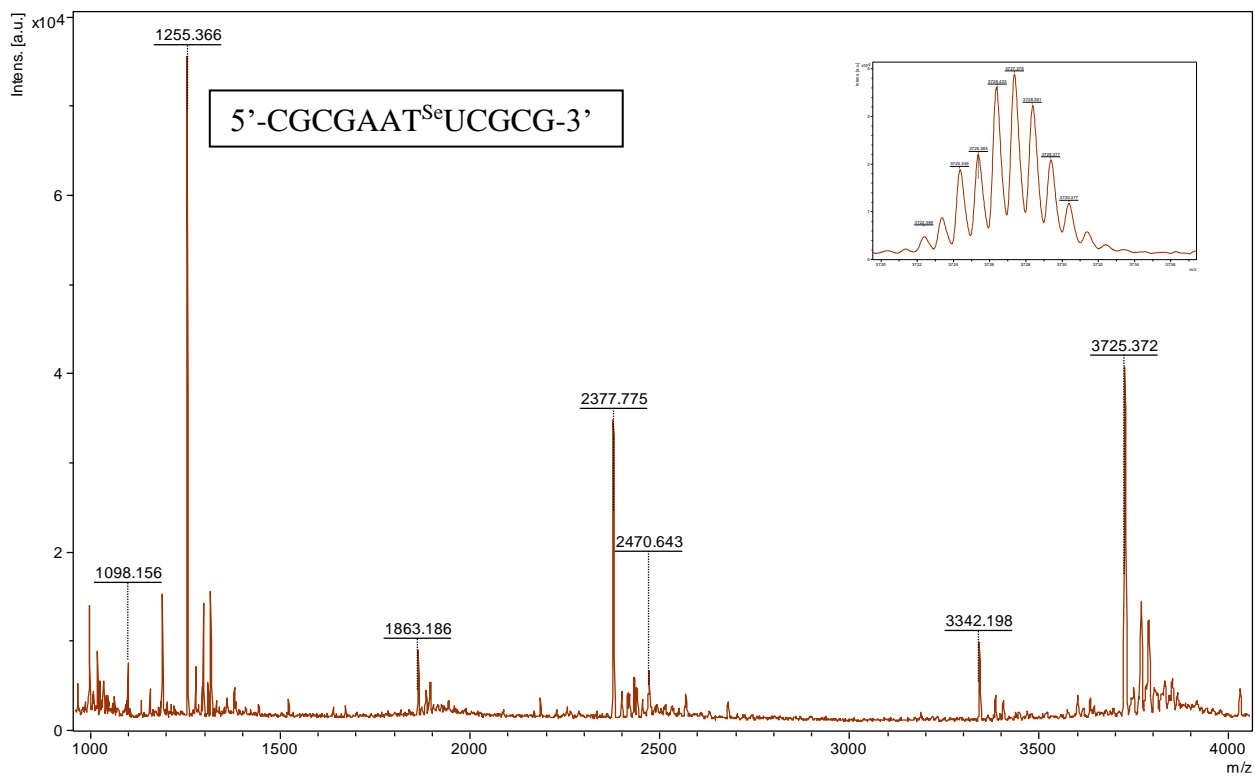CEN_2SE_1_ACCUMASS_ESINEG_HUANG_10022018_1 34 (0.634) AM (Cen,4, 80.00, Ht,5000.0,588.90,0.80); Cm (28:37)

DMTR_BSE_U_ESINEG_HUANG_070819_2 192 (3.579) AM (Cen,4, 80.00, Ht,5000.0,452.92,0.80); Cm (190:192)
TOF MS ES-

TOF MS ES+
1.50e4

685.1212

723.0782

724.0768

725.0906

100.9746

116.9483

178.9804

303.0882

655.1369

764.1539

m/z

5'-CGCGAA^{Se}UUCGCG-3'

3708.371

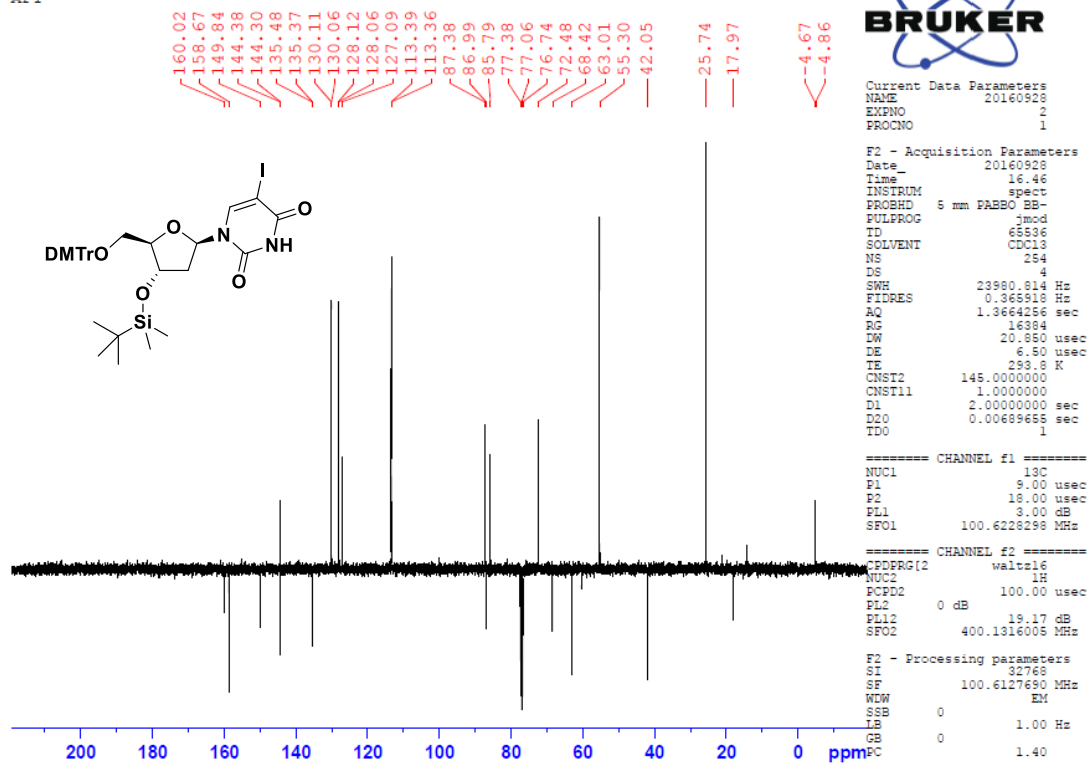2074.327

m/z

5'-CGCGAAU^{Se}UCGCG-3'
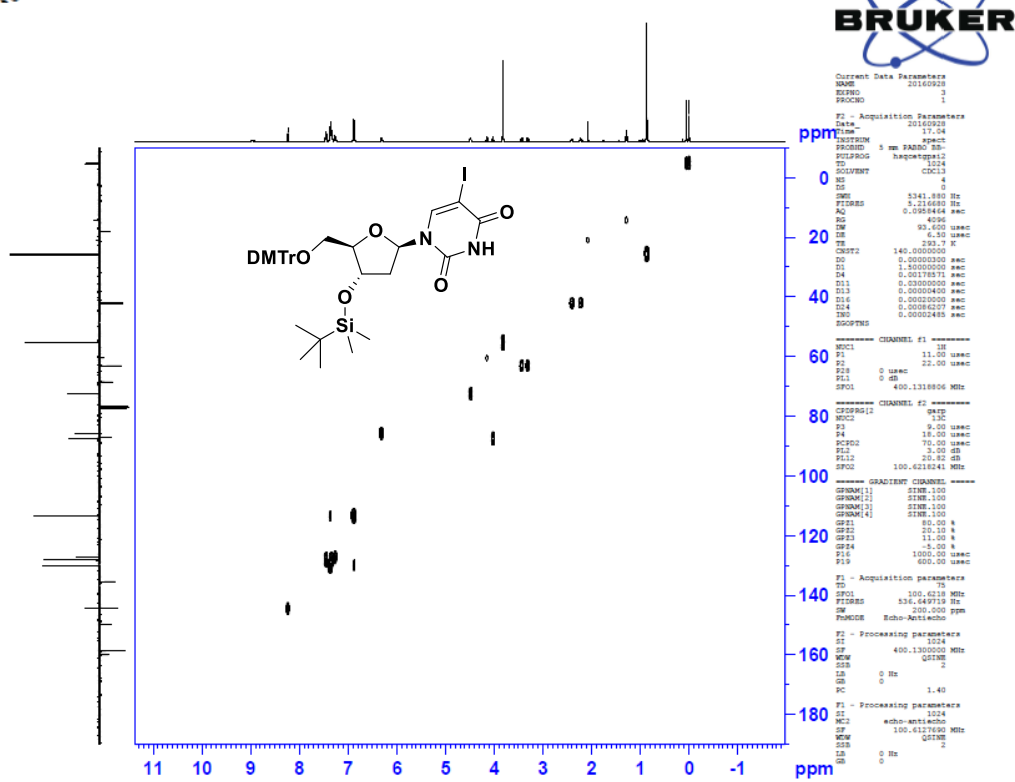
5'-CGCGAA^{Se}UTCGCG-3'
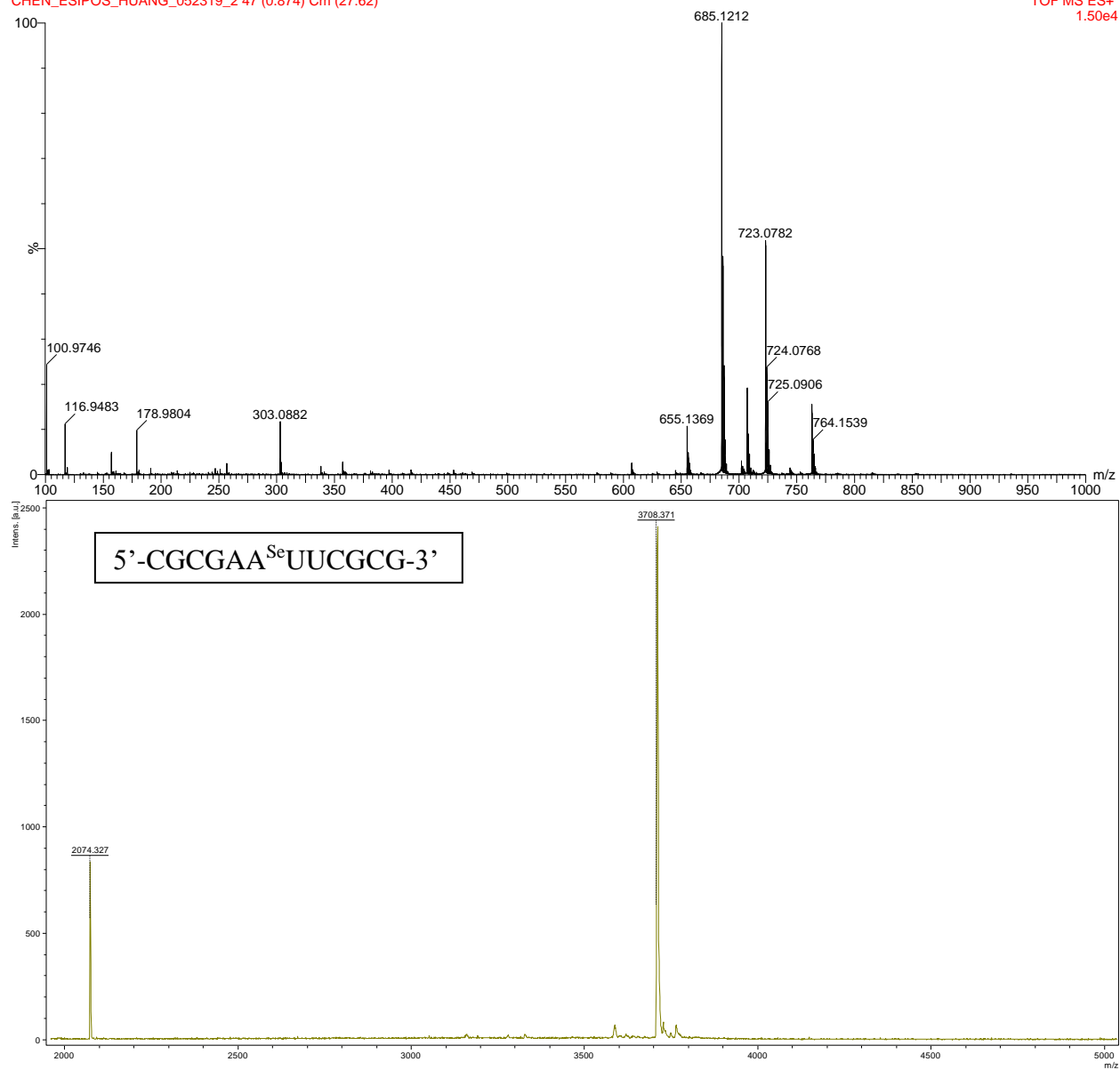
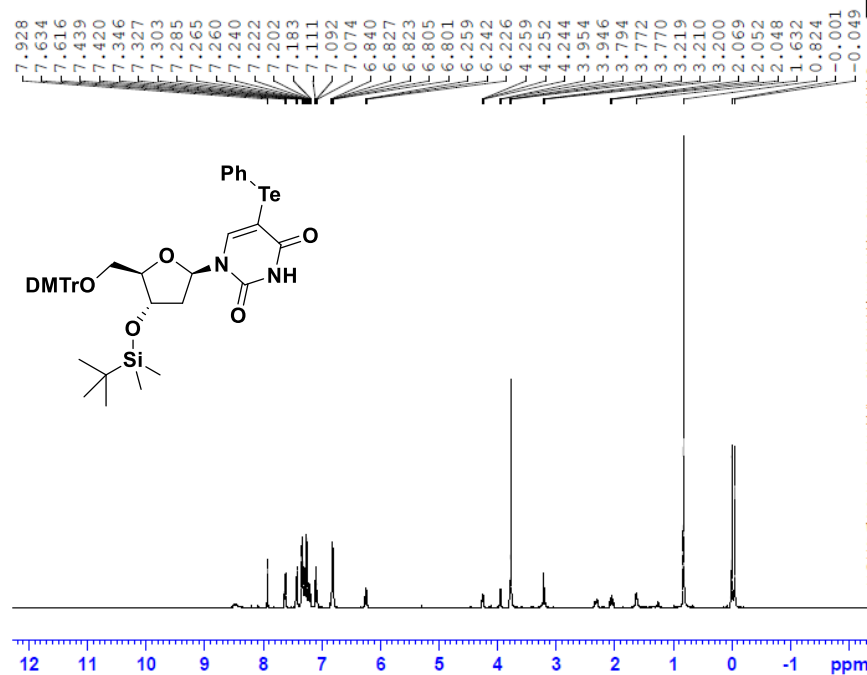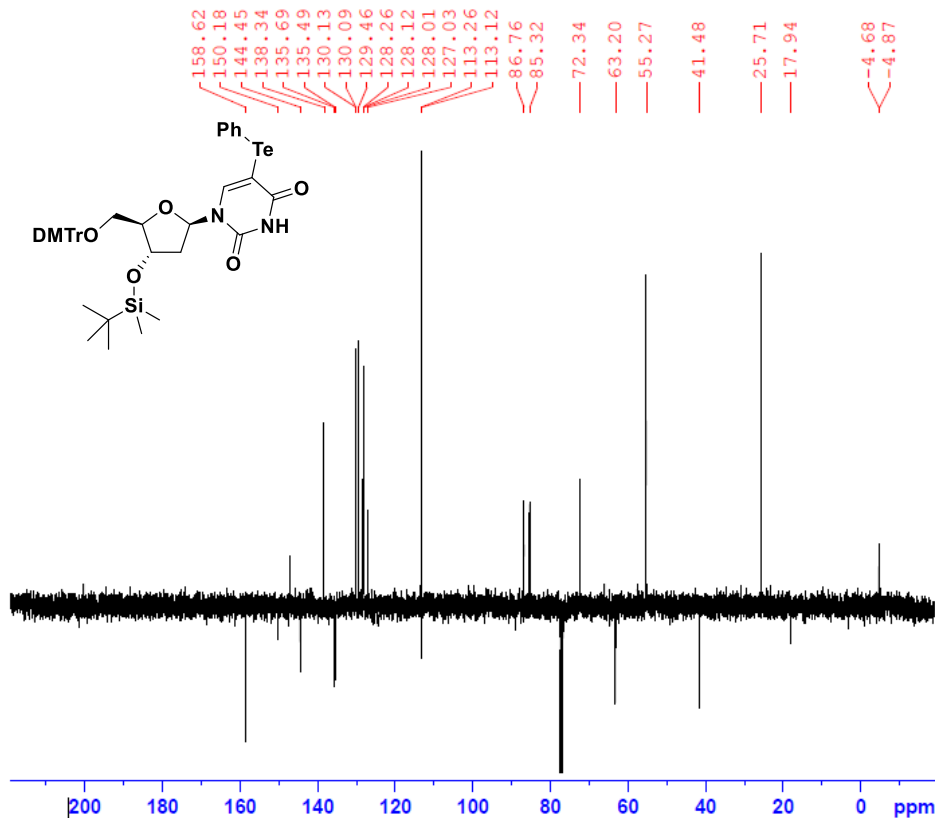Appendix D.4 NMR and MS spectra for 5-PhTe thymidine modification
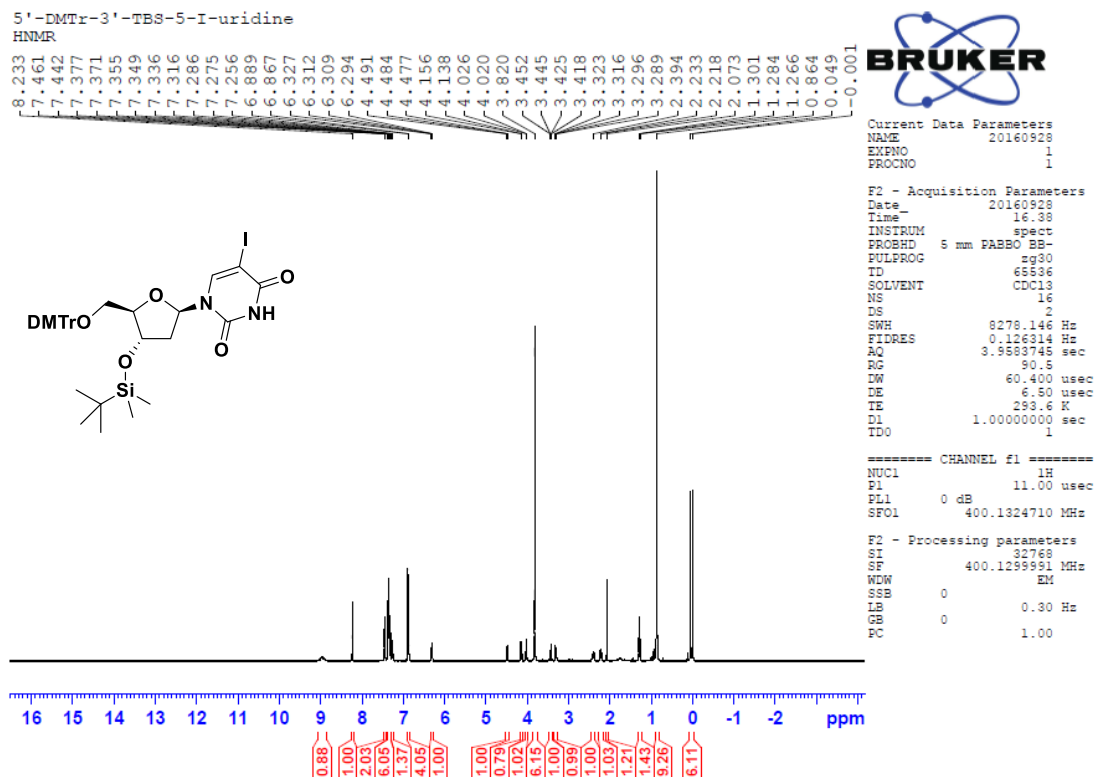
5'-DMTr-3'-TBS-5-I-uridine
APT

5'-DMTr-3'-TBS-5-I-uridine
HSQC

5'-DMTr-2'-TBS-5-PhTe-uridine
HNMR

5'-DMTr-2'-TBS-5-PhTe-uridine
APT

5'-DMTr-2'-TBS-5-PhTe-uridine
HSQC

5'-GTG^{5PhSe}TACAC-3'-SH

5'-G^{5PhSe}TGTACAC-3'-SH

2794.130

2349.454

5'-G^{5PhSe}TG^{5PhSe}TACAC-3'-SH

2935.036

1018.879