5-10-2014

Novel Selenium-modified Nucleic Acids For Structural and Functional Studies

Sibo Jiang
Georgia State University

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

Recommended Citation
http://scholarworks.gsu.edu/chemistry_diss/92

This Dissertation is brought to you for free and open access by the Department of Chemistry at ScholarWorks @ Georgia State University. It has been accepted for inclusion in Chemistry Dissertations by an authorized administrator of ScholarWorks @ Georgia State University. For more information, please contact scholarworks@gsu.edu.
ABSTRACT

Nucleic acids, as one of the most important macromolecules in living systems, play critical roles in storing, transferring, regulating genetic information, directing proteins synthesis, and catalysis. Understanding the structure of nucleic acid can bring us valuable information for mechanistic study and for drug discovery as well. Among all experimental methods, X-ray crystallography is the most powerful tool in structural biology study to reveal the 3D structure of macromolecules, which has provided over 80% of the highly detailed structural information to date. However, this great technology comes with two disturbing features, crystallization and phasing. The covalent selenium modification of nucleic acids has been proven to be a powerful tool to address both issues in nucleic acid crystallography. First part of this dissertation focuses on the development of novel selenium-modified nucleic acids (SeNA) for crystallization and phasing of B-form DNA containing structures. The novel 2’-SeMeANA modification is the first
and currently the only selenium modification, which is fully compatible with X-ray crystallographic study of B-form DNA. Since selenium derivatization at 2’-arabino position does not affect the B-type 2’-endo sugar conformation, this strategy is suitable for incorporating selenium into DNA for structural studies of B-DNA, DNA-protein complexes, and DNA-drug complexes.

Specific base pairing is essential to many biological processes, including replication, transcription, and translation. It is crucial to NA (nucleic acid) sequence-based diagnostic and therapeutic applications as well. By utilizing the unique steric and electronic property of selenium, we designed, synthesized the novel 2-Se-U RNA modification, and demonstrated its highly specific base-pairing property by both biophysical and crystallographic methods. Our studies of 2-Se-U-containing RNAs suggest that this single-atom replacement can largely improve base pairing fidelity against U/G wobble pair, without significant impact on U/A pair.

INDEX WORDS: Nucleic acid, Selenium, X-ray crystallography, Phasing, Structure, Function
NOVEL SELENIUM-MODIFIED NUCLEIC ACIDS FOR STRUCTURAL AND FUNCTIONAL STUDIES

by

SIBO JIANG

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

2014
NOVEL SELENIUM-MODIFIED NUCLEIC ACIDS FOR STRUCTURAL AND FUNCTIONAL STUDIES

by

SIBO JIANG

Committee Chair: Zhen Huang

Committee: Binghe Wang
Donald Hamelberg

Electronic Version Approved:

Office of Graduate Studies
College of Arts and Sciences
Georgia State University
May 2014
DEDICATION

This dissertation is dedicated to my parents, Feng Jiang and Hua Lou, who had instilled in me the strength of character that navigates me through each challenge. Thank you for your unconditional love, encouragement, and support.

I further dedicate this dissertation to my brilliant, loving and supportive wife Huiyan Sun. Thank you for being my cheerleader these years. Your patience and love have supported me throughout this journey. And, of course, our lovely son Nathaniel Jiang who has always been supporting and encouraging me even without knowing it.

Also, this dissertation is dedicated to my Grandpa and Aunt. You were not here to experience this time in my life but I know you were always watching over me. Thank you for your unconditional love, inspiration and for wonderful memories. In the memory of Zhiqi Jiang and Shi Jiang. Forever missed and loved.
ACKNOWLEDGEMENTS

Graduate school has been a challenging but rewarding experience and many thanks are owed to the people who have helped me through it.

I would like to express my deepest gratitude to my advisor Dr. Zhen Huang, for his guidance, understanding, patience, providing the lab with an excellent atmosphere for doing research, and most importantly, encouraging students to develop their own interests. Without your guidance, support, and continues encouragement, I would not have been able to complete the dissertation. For everything you’ve done for me, Dr. Huang, I thank you.

I would like to extend a very special thank you to my dissertation committee members, Dr. Binghe Wang and Dr. Donald Hambelberg for their encouragement, insightful comments, constructive advices, and hard questions. My sincere thanks also goes to Dr. Yujun George Zheng, Dr. David Wilson, Dr. Markus Germann, Dr. Dabney Dixon. Their advices, encouragement, and wisdom has given me the foundation necessary to complete this work.

I would also like to thank my lab mates for their support and discussion. Especially, I am very grateful to Dr. Jia Sheng and Dr. Jianhua Gan for their help in solving structures, to Huiyan Sun for her help in RNA synthesis and characterization, to Dr. Jozef Salon for his advice in nucleic acid chemistry and solid-phase synthesis, and to Dr. Julliane Caton-Williams, Dr. Abdalla E.A Hassan, Dr. Lina Lin, Dr. Hehua Liu, Dr. Wen Zhang, Dr. Manindar Kaur, Dr. Lilian Kamau, Dr. Abdur Rob, Dr. Yiyong Yan, Dr. Tian Tian, Dr. Weina Han, Ryan Smith, Jaya Punetha, Daniel Smith, Chuilun Kong, Yifei Wang, Razieh Esmaeili, Travon Haynes, Edwin Ogbonna, James Campbell, Ziyuan Fang, Cen Chen, for their support and discussion.
I would like to thank the financial, academic and technical support from the Department of Chemistry at Georgia State University. In particular, I would like to thank Dr. Siming Wang, Dr. Lifang Wang, and Dr. Yanyi Chen (Johnny) for their help in mass analysis, as well as Dr. Zhenming Du (Jimmy) and Dr. Bin Xu for maintaining the NMR instruments.

I thank the beamline staffs at ALS Beamline8.2.1, ALS Beamline8.2.2, SIBYLS Beamline12.3.1 at Lawrence Berkeley National Laboratory, Beamline7-1 at the Stanford Synchrotron Radiation Lightsource, Beamline X12C at the Brookhaven National Laboratory. Thank you all for your assistance during X-ray data collection.

I would like to give special thanks to Dr. Robert Sweet, Dr. Alex Soares, Dr. Oliver Clarke, Dr. John Schwanof, Dr. Qun Liu, Sonya Kiss, and the RapiData 2012 course at the National Synchrotron Light Source, thank you for teaching me X-ray crystallography.

I would also like to thank Dr. Punit Seth and Dr. Michael Østergaard at the ISIS Pharmaceuticals for providing us LNA modified oligonucleotides for structural study.

Last but not the least, I would like to thank the funding and financial support from the National Science Foundation, the National Institutes of Health, the Georgia Cancer Coalition, and the Molecular Basis of Disease program at Georgia State University.
# TABLE OF CONTENTS

DEDICATION .......................................................................................................................... iv

ACKNOWLEDGEMENTS ......................................................................................................... v

LIST OF TABLES .................................................................................................................... xi

LIST OF FIGURES .................................................................................................................. xii

LIST OF SCHEMES ................................................................................................................ xiv

1 GENERAL INTRODUCTION .................................................................................................. 1

1.1 Nucleic acids and macromolecule X-ray crystallography ................................................. 1

1.1.1 Challenges in Nucleic Acid X-ray Crystallography ....................................................... 2

1.1.2 Phase determination in nucleic acid X-ray crystallography .......................................... 3

1.1.3 Phase determination with selenium modified nucleic acid (SeNA) ......................... 4

2 SYNTHESIS OF 2’-SeMe-ANA FOR X-RAY CRYSTALLOGRAPHIC APPLICATIONS ....... 7

2.1 Introduction ..................................................................................................................... 7

2.1.1 Current available sugar modified SeNA for DNA/RNA structural study .............. 7

  2.1.1.1 Selenium modification at 5’-Position ...................................................................... 7

  2.1.1.2 Selenium modification at 2’-position of pyrimidine ...................... 8

  2.1.1.3 Selenium modification at 2’-Position of Purines ............................................. 12

2.1.2 Summary of current sugar selenium derivatization methods ........................... 13

2.1.3 Limitations of current sugar Se modification methods ..................................... 15

2.2 Experimental Section ..................................................................................................... 17
2.2.1  Synthesis of 2’-SeMeANA-dA phosphoramidite..........................18

2.2.2  Synthesis of 2’-SeANA-dG phosphoramidite.............................23

2.2.3  Synthesis of 2’-SeMeANA-dU derivative.................................28

  2.2.3.1  Attempted synthesis of 2’-SeANA-dU nucleoside..................29

  2.2.3.2  Select N3 protecting groups...........................................32

  2.2.3.3  Improved synthesis of 2’-SeMe-ANA-dU nucleoside.................34

  2.2.3.4  Stability test of 2’-SeMe-ANA-dU nucleoside.......................39

  2.2.3.5  Alternative method of making 2’-SeMe-ANA-dU modified DNA ......40

  2.2.3.6  Synthesis of 2’-SeMe-ANA-dU triphosphate.........................41

2.2.4  Synthesis of 2’-SeMe-ANA modified DNA by solid-phase synthesis......42

2.2.5  Thermal denaturation study of 2’-SeANA containing DNA.............44

2.2.6  Crystallization of 2’-SeMe-ANA modified nucleic acids...............46

2.2.7  Structural study of 2’-SeMe-ANA-dG containing DNA..................47

  2.2.7.1  MAD data collection and phasing.......................................47

  2.2.7.2  High-resolution data collection, structure determination, and analysis49

2.2.8  Structural study of 2’-SeMe-ANA-dG/protein complex..................51

  2.2.8.1  Crystallization ............................................................51

  2.2.8.2  Diffraction data collection and structure refinement...............52

  2.2.8.3  Structure of 2’-SeMe-ANA-dG DNA/Bacillus fragment complex.......53

2.3  Conclusion..................................................................................55

3  HIGH SPECIFICITY OF RNA BASE PAIRING WITH 2-SE URIDINE ...............57
3.1 Introduction ........................................................................................................................................57

3.1.1 U/G wobble base pair in nature .................................................................................................57

3.1.2 Biophysical properties of U/G wobble base pair .......................................................................57

3.1.3 Nucleic acid base modification to increase base pairing specificity ........................................58

   3.1.3.1 Sulfur modification on nucleobases ......................................................................................59

   3.1.3.2 2-Seleno-thymidine ..............................................................................................................60

3.1.4 2-Se-U RNA modification ..........................................................................................................62

3.2 Experimental Section ......................................................................................................................63

3.2.1 Synthesis of 2-Se-Uridine phosphoramidite ...............................................................................64

3.2.2 2-Se-Uridine containing RNA synthesis and purification ..........................................................71

3.2.3 Thermal denaturation study of 2-Se-U RNA .............................................................................72

3.2.4 Crystallization and diffraction data collection ...........................................................................74

3.2.5 Structure determination and refinement .................................................................................76

3.2.6 Crystal structure of 2-Se-U modified RNA .................................................................................77

3.2.7 High fidelity base pairing of 2-Se-U RNA in crystal: Experimental design ..............................78

3.2.8 Crystallization and diffraction data collection of 5’-GUGUAUAC-3’ RNA ...............................80

3.2.9 Structure determination and refinement of 5’-GUGUAUAC-3’ RNA .......................................81

3.2.10 Crystal structure of 5’-GUGUAUAC-3’ RNA ...........................................................................82

3.2.11 Crystallization and diffraction data collection of 5’-GUGUA_{258}^{258}UAC-3’ RNA 83
3.2.12 Structure determination and refinement of 5'-GUGUA$^{25e}$UAC-3’ RNA........84

3.2.13 Crystal structure of 5'-GUGUA$^{25e}$UAC-3’ RNA ........................................85

3.3 Conclusion .............................................................................................................88

REFERENCES .............................................................................................................90

APPENDICES ............................................................................................................99

Nucleic acid mini screen kit (NAM) and formulation ..............................................99

NMR spectrums of key compounds .......................................................................100
LIST OF TABLES

Table 2.1 MALDI-TOF MASS analysis of 2'-SeMe-ANA-dG containing DNA .................... 44
Table 2.2 Thermal denaturation experiment of 2'-SeMeANA-dG modified DNA........ 45
Table 2.3 MAD data collection statistics ................................................................. 48
Table 2.4 Crystallographic data collection and refinement statistics ....................... 54
Table 3.1 MALDI-MASS analysis of 2-Se-U modified RNA ..................................... 72
Table 3.2 Thermal denaturation study of RNA mis-pairing ....................................... 74
Table 3.3 Data collection and refinement statistics of 3S49 ....................................... 76
Table 3.4 Hydrogen bond distances of U/A and 2-Se-U/A base pairs in 1JAH ............ 86
Table 3.5 Data collection and refinement statistics of 1JAB and 1JAH ...................... 88
LIST OF FIGURES

Figure 1.1 Selenium Modified Nucleic Acid (SeNA) ............................................................... 5

Figure 2.1 Crystal structure of 1MA8 ....................................................................................... 10

Figure 2.2 2'-SeMe-T containing DNA structure ................................................................. 12

Figure 2.3 Sugar puckers and DNA conformation ............................................................... 16

Figure 2.4 N-3 protecting groups for compound (18) ............................................................ 32

Figure 2.5 Oxidation of selenium functionality during deprotection .................................... 34

Figure 2.6 2'-SeMe-ANA-dU to 2'-OMe-dU conversion (1H NMR) ....................................... 40

Figure 2.7 HRMS-ESI(-) analysis of 2'-SeMe-ANA-dUTP .................................................... 42

Figure 2.8 DNA solid-phase synthesis .................................................................................... 42

Figure 2.9 DNA crystal in cryo-loop at ALS 8.2.2 ................................................................. 47

Figure 2.10 Fluorescent Scan of crystal .................................................................................. 47

Figure 2.11 Fourier electron density maps provided by MAD phasing .................................. 48

Figure 2.12 Global and local structures of the 2'-SeMe-ANA-dG modified DNA ................. 49

Figure 2.13 Hydration pattern comparison between 2'-SeMe-ANA-dG and dG .................. 51

Figure 2.14 Structure of 2'-SeMe-ANA DNA/protein complex .......................................... 54

Figure 3.1 U/G Wobble pair and codon degeneracy .............................................................. 57

Figure 3.2 Structure of the 2-Se-T-DNA (5' -G\textsuperscript{2'-SeMe}UG\textsuperscript{2-SeTACAC-3'} \textsubscript{2}) ................................................................. 62

Figure 3.3 2-Se-U/A base pair and 2-Se-U/G base pair ....................................................... 63

Figure 3.4 and local structures of the 2-Se-U-containing RNA ............................................ 78

Figure 3.5 The new RNA sequence for high fidelity base pairing study .............................. 79

Figure 3.6 Hypothesized duplex rearrangement after 2-Se-U modification ....................... 80
Figure 3.7 Global and local structures of 5’-GUGUAUAC-3’ RNA ................................. 82
Figure 3.8 5’-GUGUAUAC-3’ and 5’-GUAUAUAC-3’ structures .................................. 83
Figure 3.9 Overall structure of 5’-GUGUA_{2Se}UAC-3’ RNA with electron density map... 85
Figure 3.10 More insight into 5’-GUGUA_{2Se}UAC-3’ RNA structure ............................. 87
LIST OF SCHEMES

Scheme 2.1 Synthesis of 5’-Se-Derivatized Phosphoramidites and Oligonucleotides 8

Scheme 2.2 Synthesis of 2’-Se modified oligonucleotides ........................................ 9

Scheme 2.3 Synthesis of 2’-Se-Derivatized Cytidine Phosphoamidite and DNAs/RNAs 11

Scheme 2.4 Synthesis of 2’-SeMe-purine containing oligonucleotides .......................... 13

Scheme 2.5 Synthesis of 2’-SeMeANA-dA phosphoramidite ........................................ 18

Scheme 2.6 Synthesis of 2’-SeMe-ANA-dG phosphoramidite ........................................ 23

Scheme 2.7 Attempted synthesis of 2’-SeANA-dU nucleoside ........................................ 29

Scheme 2.8 Mechanism of generating compound (19) .................................................. 31

Scheme 2.9 Using benzoyl as N3 protecting group ....................................................... 33

Scheme 2.10 Improved synthesis of 2’-SeMe-ANA-dU nucleoside .................................. 35

Scheme 2.11 2’-SeMe-ANA-dU to 2’-OMe-dU conversion under Ultra Mild condition 39

Scheme 2.12 Synthesis of 2’-SeMe-ANA-dUTP ............................................................ 41

Scheme 3.1 Synthesis of 2-Se-T phosphoramidite and DNAs ........................................ 61

Scheme 3.2 Synthesis of 2-Se-Uridine phosphoramidite ............................................... 64
1 GENERAL INTRODUCTION

1.1 Nucleic acids and macromolecule X-ray crystallography

Nucleic acids, including DNA and RNA, play critical role in storing, transcribing, translating genetic information, regulating gene expression, and serving catalytic functions. It has been six decades since DNA was first identified as the master blueprint of life \(^1\); nucleic acid-related research areas have drawn tremendous attentions from chemistry, biology as well as biomedical-related research. The rapid advances in nucleic acid research are enormous, along with the invention and advance of many new technologies and concepts; bring us more insights into the structures, functions and biomedical applications of this fundamental building block of life.

X-ray crystallography is a technique that allows the determination of the 3D structure of molecules. Macromolecular crystallography has come a long way since the determination of the first protein structures (myoglobin at 6 Å resolution) \(^2\). Nowadays, it has become the method of choice and provided atomic resolution structures of macromolecules that can lead to comprehensive understanding of conformation-related biological activities for nucleic acid research.

The structural details of nucleic acids revealed in X-ray crystallographic experiment are used to explain the mechanistic aspect of many biological processes, including replication, transcription, translation, transcriptional / translational regulations, DNA damaging and repairing, and ribozyme catalysis \(^3\ 4\ 5\ 6\ 7\ 8\ 9\ 10\ 11\). The 3-D structural information of nucleic acid at atomic level is also extremely helpful in structure-base drug design (SBDD) \(^12\ 13\).
1.1.1 Challenges in Nucleic Acid X-ray Crystallography

X-ray crystallography is by far the most successful and widely used method for structural study of bio-macromolecules. However, this technology comes with two disturbing features: crystallization and phasing, both limit the determination of novel macromolecule structures.

Crystallization is usually the rate-limiting step in X-ray crystallographic experiment. The macromolecules that can be studied are of course limited to those that can be crystallized. Numerous methods had been established to provide crystals for diffraction experiment or to offer crystals with better diffraction quality. Unfortunately, these methods are still very much based on trial-and-error screening. Crystallization of a novel macromolecule could take weeks, months, or even years. It also takes tremendous amount of effort to yield crystals with better diffraction quality, which can offer higher resolution data that leads to more structural details.

In a diffraction experiment, the film or CCD detector only serves as photon counter by which researcher can measure the diffraction pattern and intensity of each diffracted beam. All information about the relative phases of different diffractions is lost systematically during the experiment. However, the amplitudes along with the phase information of each diffraction is essential for the solution of structures. The absence of phase information raises a serious problem to the reconstruction of the electron density map from a diffraction experiment. The additional work required to recover phase information is called “The phase problem”.

Generally, there are several approaches to recover the phase information in a diffraction experiment. Molecular Replacement (MR) $^{14}$ method can be used to solve a structure when a good model for a large fraction of the molecule in the crystal is available. The MR mainly relies upon the presence of a previously solved related-structure and it is becoming
more and more useful as the database of solved the structures gets larger and larger. Phase information could also be obtained from experimental data. This type of phasing method is called experimental phasing. For macromolecule crystallography, there are several experimental phasing methods available, including single isomorphous replacement (SIR), multiple isomorphous replacement (MIR), single isomorphous replacement with anomalous scattering (SIRAS), multiple isomorphous replacement with anomalous scattering (MIRAS), single-wavelength anomalous dispersion (SAD), and multi-wavelength anomalous dispersion (MAD)\textsuperscript{15}. These experimental phasing methods entirely depend on the presence of special atoms in crystal. For protein crystals, heavy-atoms can be introduced in to crystal by soaking or co-crystallization methods. For example, a protein with free cysteine thiols can be derivatized by mercury containing compounds; K\textsubscript{2}PtCl\textsubscript{4} can preferentially bind to methionine side chains. Experimental phasing has been largely improved by the combination of multiwavelength anomalous dispersion (MAD) and selenomethionine (SeMet) protein over the past twenty years\textsuperscript{16 17 18 19 20}. Until 2000, MAD phasing had provided about two thirds of all new protein crystal structures\textsuperscript{21}.

1.1.2 Phase determination in nucleic acid X-ray crystallography

In contrast with protein derivatization techniques, the derivatization methods for nucleic acid are quite limited. The heavy-atom soaking and co-crystallization methods for protein had been proven to be much less successful in nucleic acids, due to the lack of specific binding sites for cations and a potential risk of the backbone cleavage by the metal cations. Research also shown that some heavy metal ions used for phase determination of protein could generate unwanted mismatched base pairs in DNA\textsuperscript{22 23 24} and RNA\textsuperscript{25} structures.
The conventional phasing method for nucleic acid structures involves 5-Br derivatives in synthetic nucleic acids. Bromine offers a K-edge of 0.920 Å, which is a convenient wavelength within the range of modern synchrotron radiation. MAD phasing has been successfully achieved on several bromine-containing DNA crystal structures. However, the 5-Br derivative method does not work in many cases. The drawbacks of this methodology include X-ray induced debromination, structural perturbation, and inadequate modification sites (5-BrdU and 5-BrU). In some cases, 5-Br derivatives cannot be crystallized under native conditions. All of these unwanted properties limit the application of 5-Br derivative in novel nucleic acid structures determination.

1.1.3 Phase determination with selenium modified nucleic acid (SeNA)

In 2001, Huang, Egli and co-workers raised a new concept of using a combination of selenium modified nucleic acid (SeNA) and multi-wavelength anomalous dispersion (MAD) to facilitate structure determination of nucleic acid. In their proof of principle, selenium was introduced into nucleic acid chemically through solid-phase synthesis to address phasing issue in nucleic acid crystallography. Both selenium and oxygen are in the Family VIA on the periodic table and thus they share similar chemical and electronic properties, which allows for atom-specific substitution of oxygen with selenium in nucleic acid structures.
Figure 1.1 Selenium Modified Nucleic Acid (SeNA)

Compare with bromine (K-edge 0.920 Å), selenium exhibits a theoretical K-edge of 0.979 Å. In terms of the availability of wavelength, structural biology labs have been routinely using Se K-edge for SeMet MAD/SAD phasing of protein structure. Also, selenium offers comparable anomalous signals compared with bromine\(^3\), which indicates that there’s no need to increase the number of modification sites when switch from 5-Br derivative method to the SeNA approach. Since oxygen is one of the key elements in nucleic acid, selenium can be chemically or enzymatically introduced into nucleic acids structures by replacing oxygen atoms without causing significant destabilization effect. (Shown in Figure 1.1)

Moreover, research has shown that SeNA offers superior stability under x-ray irradiation\(^2\). The photo-stability of SeNA is extremely important for its application in a MAD experiment, since MAD phasing requires multiple datasets from same crystal collected at different wavelengths. Improved stability under x-ray irradiation could save researchers lots of time and headache during sample preparation and data collection process. More importantly, we also observed from current available crystal structures that selenium modifications does not cause significant structural perturbation. Plus, selenium modified nucleic acids (SeNA) are much easier to prepare, purify, and store than SeMet proteins and 5-Br-U modified nucleic acids as well.
In this dissertation, I would like to mainly focus on the development and application of novel selenium modified nucleic acids (SeNA) for structural and functional studies of nucleic acids.
2 SYNTHESIS OF 2’-SeMe-ANA FOR X-RAY CRYSTALLOGRAPHIC APPLICATIONS

2.1 Introduction

2.1.1 Current available sugar modified SeNA for DNA/RNA structural study

2.1.1.1 Selenium modification at 5’-Position

In 1998, Dr. Huang’s lab initiated the project of synthesizing selenium-modified nucleic acids (SeNA) for crystal structure determination in x-ray crystallography. Selenium was first introduced to 5’-positions of A, T, C, G and U to facilitate phase determination via MAD (Multiplewavelength Anomalous Dispersion) phasing \(^{31}\). The incorporation of selenium functionality was done in a two-phase system (water-toluene) using a phase transfer catalyst (tetrahexyl-ammonium hydrogen sulfate). The 5’-hydroxy groups of nucleosides (A, T, C, G and U) were activated by Br-, Ms- or Ts- followed by nucleophilic substitution reaction using sodium selenide or methyl selenide as nucleophile. The resulting 5’-MeSe labeled nucleoside was converted to corresponding phosphoramidite conveniently by a one step phosphorylation reaction and then integrated into oligonucleotides by solid-phase synthesis (Scheme 2.1). In this study, selenium functionality showed outstanding stability under various conditions involved in solid-phase synthesis, including moderate acidic condition, strong base and oxidant. The desired 5’-SeMe DNA can be easily purified by either HPLC or gel electrophoresis to achieve over 95% purity. The selenium modification at 5’ terminal position demonstrated various advantages of selenium derivatization of nucleic acids, including compatibility with phosphoramidite chemistry and solid-phase automate DNA/RNA synthesis, outstanding stability under various conditions, easy handling and purification procedures.
Scheme 2.1 Synthesis of of 5’-Se-Derivatized Phosphoramidites and Oligonucleotides  
Reagents and conditions: a) CBr₄, Ph₃P, Diisopropyl azodicarboxylate (DIAD), THF.  
b) Ts-Cl, Pyridine. c) MsCl, Et₃N, THF. d) NaSeMe. e) Et₃N 3HF. f) PCl(OCH₂CH₂CN)N(iPr)₂.  
g) Solid-phase synthesis  
2.1.1.2  Selenium modification at 2’-position of pyrimidine  

The application of 5’-SeMe modification is quite limited since it is only applicable for  
modifying 5’ terminal positions of oligonucleotides. In order to further expand the application  
of SeNA for structure determination of nucleic acids, it was of great importance to develop a  
method to integrate selenium to internal positions of DNAs/RNAs. Through the collaboration  
between Huang and Egli lab, 2’-Se modified oligonucleotide was successfully prepared  
chemically (shown in Scheme 2.2) and its potential usefulness in x-ray crystallography study  
of nucleic acid structure was demonstrated as well. Detailed synthetic procedure was also  
described in Nature Protocol.
Scheme 2.2 Synthesis of 2’-Se modified oligonucleotides

Reagents and conditions: a) MsCl, THF, Et$_3$N; 95%. b) Toluene, tetrahexylammonium hydrogen sulfate, Na$_2$CO$_3$ (sat.). c) (Bu)$_4$NF; 95%. d) NaHSe, then MeI, or NaSeMe; 96%. e) PCl(OCH$_2$CH$_2$CN)N(iPr)$_2$; 92%. f) Solid-phase synthesis.

The diffraction data of 2’-Se-dU modified DNA has shown that 2’-Se modification could provide excellent anomalous signal for SAD/MAD phasing. The crystal of the 2’-Se-dU modified decamer (5’-GCGTAY$_{2’}$SeMeACGC-3’)$_2$ diffracted to a maximum resolution of 1.3 Å. In-depth study of crystal structure containing 2’-SeMe modification (PDB ID: 1MA8) revealed that the 2’-SeMe groups locate in the minor groove of DNA (crystal structure shown in Figure 2.1). More importantly, the selenium modified uridine derivatives retain the RNA-like C3’-endo sugar pucker in the A-form DNA duplex. A direct comparison between 2’-SeMe-dU DNA (PDB ID: 1Z7I) and conventional 5-Br-dU modified DNA (PDB ID: 2H05) reveals that the 2’-SeMe-dU DNA offers higher resolution (1.28 Å vs 1.8 Å), faster crystal growth, and less local structure perturbation.
Figure 2.1 Crystal structure of 1MA8 with 1.3 Å resolution. (5’-GCgTAU2’SeMeACGC-3’)$_2$

In order to allow large-scale synthesis of 2’-SeMe-dU at low cost, the synthetic route was later optimized by converting uridine to 2, 2’-anhydouridine followed by a ring opening reaction using sodium methylselenide as nucleophile \(^{35}\). 2’-MeSe-dC was also successfully synthesized based on an amination step in which C4 carbonyl of 5’-O-DMTr-2’-MeSe-Uridine was activated by 1,2,4-triazole or 2,4,6-triisopropylbenzenesulfonyl chloride followed by amination reaction in aqueous ammonium hydroxide \(^{36,37}\). (Shown in Scheme 2.3) Enzymatic preparation of 2’-MeSe-U/C modified RNA also showed excellent compatibility with T4 RNA ligase \(^{37}\), which allows the production of longer RNA segment with biological activity for structure and function study. This work provides access for the preparation of long selenium-modified RNA sequences, which is not accessible by conventional solid-phase synthesis. 2’-
SeMe-U modified RNA has been successfully used in several structural studies\textsuperscript{38,39} and played a crucial role in the structure determination of the Diels-Alder ribozyme.\textsuperscript{40}

Scheme 2.3 Synthesis of 2'-Se-Derivatized Cytidine Phosphoamidite and DNAs/RNAs

Reagents and conditions: a) R\textsubscript{1} : TMS-Im, then POCl\textsubscript{3}, 1H-1,2,4-triazole, Et\textsubscript{3}N in CH\textsubscript{3}CN; R\textsubscript{2} : 2,4,6-triisopropylbenzenesulfonyl chloride, Et\textsubscript{3}N, DMAP in CH\textsubscript{2}Cl\textsubscript{2} b) aqueous NH\textsubscript{4}OH. c) TMS-Im, then Ac\textsubscript{2}O, TEA and DMAP in THF. d) 2-cyanoethyl N,N-diisopropyl chlorophosphoramidite, N,N-diisopropylethylamine in CH\textsubscript{2}Cl\textsubscript{2}. (e) Solid-phase synthesis.

By using similar strategy described by Carrasco\textsuperscript{35}, 2'-SeMe-thymidine phosphoramidite was synthesized in which 2,2'-anhydrothymidine served as the key intermediate for introducing selenium functionality \textsuperscript{41}. In this work, 2'-SeMe-T modification was successfully incorporated into DNAs and RNAs (tRNA ΨC hairpin loop) with high yields. The structure of 2'-SeMe-T containing DNA DNA (5'-GT\textsubscript{2}Se GTACAC-3')\textsubscript{2} was determined at 1.4 Å resolution (shown in Figure 2.2) and was virtually identical to corresponding native structure. Moreover, the thermal
denaturation experiment suggests that 2'-SeMe-T modified DNAs/RNAs offer similar duplex stability to corresponding native ones.

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]

\[2\]
Scheme 2.4 Synthesis of 2'-SeMe-purine containing oligonucleotides

With the combination of chemical solid-phase synthesis and enzymatic ligation, 2'-SeMe modification can be incorporated into RNA with up to 100 nucleotides. Study done by Prof. Micura lab demonstrated an alternative method of preparing long RNA with 2'-SeMe modification for structural study purpose. In this work, 2'-SeMe CTP/UTP/ATP/GTP was synthesized and efficiently incorporated into long RNA with mutant T7 RNA polymerase. This opens up an alternative access for the preparation of long selenium-modified RNA sequences, which is not accessible by conventional solid-phase synthesis.

2.1.2 Summary of current sugar selenium derivatization methods

After years of comprehensive study, 2'-SeMe modification has became a convenient method for selenium derivitization of nucleic acid for x-ray crystallographic studies. The 2'-SeMe phosphoramidites can be conveniently prepared and incorporated into oligonucleotides with high yield; the purification of modified oligos is also quite straightforward. For several reasons, the 2'-MeSe labeling has greater potential over conventional 5-Br derivatives for MAD/SAD phasing.
First of all, 2’-SeMe modification shows superior photo stability over 5-Br pyrimidine 29. UV induced 5-Br-RNA/protein cross-linking 48, X-ray induced debromination during MAD data collection 28 was reported by Gott, et al. and Ennifar, et al. respectively. The photo stability of macromolecule is more critical in SAD/MAD data collection than in regular single-wavelength experiment. For instance, in MAD data collection, three data sets at different energies will be taken while each of them needs about twice much data compare to native data collection because of the average redundancy loss (caused by loss of symmetry in diffraction pattern). This results in approximately six times longer exposure time in a three wavelengths MAD experiment than in a single-wavelength experiment. Under this circumstance, the excellent stability 2’-MeSe modifications can save crystallographer precious crystals and valuable beamline time.

Plus, compared to 5-Br derivatives, 2’-SeMe modified oligonucleotides can resist much stronger basic condition even in ammonium hydroxide at 65°C. Thanks to the outstanding chemical stability of 2’-SeMe modification, the synthesis, purification and handling of 2’-MeSe containing oligonucleotides is more convenient than 5-Br pyrimidine containing oligonucleotides.

Despite the stability issue, 5-Br oligo causes more perturbation in overall and local structures, which mainly caused by disruption or alternation of base stacking 33 and drastic change in hydration pattern in major groove 30. Additionally, 2’-MeSe modified DNA/RNA can be readily prepared with all five nucleobases that gives 2’-SeMe more diversified modification sites over 5-Br pyrimidine analogs.
2.1.3 Limitations of current sugar Se modification methods

$2'-\text{SeMe}$ modification has been the most successful approach to facilitated structure determination of A-form DNA and RNA structures. Studies have shown that this RNA-like $2'-\text{SeMe}$ modification favors C3'-endo sugar pucker in crystal structure. Thus, C2'-ribo position selenium modification is only applicable to A-Form DNAs and RNAs. The backbone phosphoroselenoate DNA/RNA was also developed for MAD phasing by Prof. Egli lab. A P-Se containing left handed Z-DNA structure was reported with 1.1 Å resolution. Currently, no selenium-derivative has been reported with the compatibility for B-DNA’s $2'$-endo sugar pucker.

B-form as most common DNA conformation at physiological conditions, is an ideal model for study of protein-DNA, drug-DNA interaction. The wider major groove as well as the elongated helix, provides greater accessibility for protein and drugs. This brings our attention to expand the application of selenium modification to B-form DNA to facilitate structure-based study.

As one of the critical parameters of nucleic acid structures, sugar conformation plays an important role in maintaining overall shape of DNA duplex; $2'$-endo sugar is preferred in B-form DNA, 3'-endo sugar pucker appears in A-form DNA dominantly. (Shown in Figure 2.3) This allows researchers to control the DNA conformation by tuning sugar pucker with modified nucleosides. Over decades, hundreds of novel oligonucleotides modifications were discovered for antisense therapeutics, especially at $2'$-position. Modifications at the $2'$ position have proved to be critical in modifying the property of oligonucleotide. Conventional $2'$-F, $2'$-OMe or $2'$-LNA modifications cause the sugar to adopt to A-type 3'-endo conformation, which leads to an increase in A-form duplex stability. In contrast, very few attempts were made to drive the
sugar conformation to 2'-endo for the purpose of stabilizing B-form duplexes. Due to the synthetic difficulty, research related to 2'-arabinosyl analogs was still quite limited. Until late 90s, 2'-deoxy-2'-fluoroarabino-thymidine (2’-FANA-T) modified oligonucleotides was successfully synthesized and characterized. The 2’-FANA modification showed increased duplex stability in solution\textsuperscript{58,59} as well as conformational preorganization\textsuperscript{60} in crystal structure. Interestingly, the 2’-F-ANA modification adopts an unusual O4’-endo sugar conformation instead of DNA-like B-type conformation.

Figure 2.3 Sugar pucker and DNA conformation

Based on previous study where 2’-SeMe-modification can radically speed up the crystallization process of A-form DNA. We have hypothesized that 2’-SeMe arabinosyl (2’-SeMe-ANA) modification may facilitate pre-organization of 2’-endo sugar pucker, which could lead to successful crystallization. To demonstrate the proof of principle, we selectively incorporate
selenium atom to the 2′-arabino-position of deoxyguanosine in oligodeoxynucleotides to achieve both crystallization and phasing of B-form DNA.

The first synthesis of the 2′-selenomethyl-arabino nucleic acids (2′-SeMe-ANA), corresponding phosphoramidites and 2′-SeMe-ANA containing DNAs is presented in this chapter. Moreover, by taking advantage of Se labeling, convenient structure determination of B-DNA by MAD phasing was also demonstrated, the crystal structure of selenium-derivatized Dickerson Drew Dodecamer was determined at 1.49 Å resolution.

2.2 Experimental Section

1H, 13C and 31P NMR data were acquired on a Bruker 400MHz NMR instrument. Chemical shifts (δ values) and coupling constants (J values) are given in ppm and Hz, respectively, using TMS (1H NMR) and solvents (13C NMR) as internal standards. Reagents were purchased from Sigma/VWR/Chem Genes and used without further purification. All reactions were set up under argon atmosphere. Analytical/preparative thin layer chromatography was carried out on 250u/1000u silica gel 60 coated TLC with F-254 as indicator (Dynamic Adsorbents). Flash chromatography purification was performed on silica gel 60 (230-400mesh) (Dynamic Adsorbents) with gradient elution.
2.2.1 Synthesis of 2’-SeMeANA-dA phosphoramidite

2’-O-[(trifluoromethyl)sulfonyl]-3',5'-O-(1,1,3,3-tetraisopropyldisiloxane-1,3-diyl)adenosine (3) Adenosine (1) 3.52 g (13.16 mmol) was co-evaporated with anhydrous pyridine twice (10 ml each time) and suspended in 50 ml dry pyridine under argon, and the mixture was cooled in an ice bath with stirring. To the suspension, 4.36 g of 1,3-dichloro-1,1,3,3-tetraisopropyl-disiloxane (13.82 mmol, 1.05 eq) was added drop wise. The flask was removed from ice bath and stirred for 3 h at room temperature. After TLC showed reaction completion, 5 ml methanol was added and the reaction mixture was evaporated and co-evaporated with toluene 5 ml x 3 under vacuum. The residue was dissolved in CH₂Cl₂ 150 ml, washed with water 75 ml x 2, brine 75 ml x 2, dry over anhydrous Na₂SO₄. The drying agent was filtered off and solution was evaporated to dryness to afford crude product (2) as white foam.

Crude product (2) was dried under high vacuum, and then it can be directly used for the next step without further purification. Crude product (2) together with 4-dimethylaminopyridine
4.82 g (39.48 mmol, 3 eq) were dried under high vacuum and dissolved in 50 ml anhydrous dichloromethane under argon. The flask was chilled in ice bath, trifluoromethanesulfonyl chloride 1.68 ml (2.66 g, 15.79 mmol, 1.2 eq) was added drop wise. The reaction was kept stirring at 0 °C for 15 min (desired product is a less polar spot on TLC), and then the yellowish color solution was diluted by adding 100 ml CH₂Cl₂. The organic solution was washed with saturated sodium bicarbonate aqueous solution, dried over anhydrous Na₂SO₄ and evaporated. The crude product was purified by flash chromatography on SiO₂. (CH₂Cl₂/MeOH, 100/0-99/1, v/v). Yield pure product (3) 7.1 g (84.1% in two steps) as colorless solid.

**Compound (2)** ¹H NMR (400 MHz, CDCl₃): δ 8.30 (s, 1H, 8-H), 7.98 (s, 1H, 2-H), 5.99 (s, 1H, 1'-H), 5.89 (br, 2H, NH₂), 5.08 (dd, 1H, 3'-H), 4.60 (d, J = 5.3 Hz, 1H, 2'-H), 4.10 (dd, 3H, 4'-H, 5'-H, 5''-H), 3.61 (br, 2'-OH), 1.23 – 0.96 (m, 28H, 2 x (CH₃)₂CHSi).

**Compound (3)** ¹H NMR (400 MHz, CDCl₃): δ 8.27 (s, 1H, 8-H), 7.96 (s, 1H, 2-H), 6.11 (s, 1H, 1'-H), 5.81 (d, J = 4.8 Hz, 1H, 2'-H), 5.80 - 5.71 (br, 2H, NH₂), 5.30 (dd, J = 9.1, 4.8 Hz, 1H, 3'-H), 4.19 - 4.07 (m, 3H, 4'-H, 5'-H, 5''-H), 1.21 – 0.96 (m, 28H, 2 x (CH₃)₂CHSi).

**3',5'-O-(1,1,3,3-tetraisopropylidisiloxane-1,3-diyl)-2'-deoxy-2'-selenomethylarabinosine (4)** Sodium borohydride 1.24 g (32.73 mmol, 3 eq) was placed a sealed 250 ml round bottom flask, dried under high vacuum for 10 min, purged with dry argon twice to deplete oxygen. To the flask, 100 ml dry dioxane was injected, followed by adding 3.1 ml of CH₃SeSeCH₃ (32.73 mmol, 3 eq, d = 1.987 g / ml). The yellowish color suspension was stirred at room temperature and 10 ml of anhydrous ethanol was added drop wise. The result solution was stirred at room temperature for 1.5 h. To this slightly yellow solution, staring material (3) 7 g (10.91 mmol) in dry dioxane 75 ml was injected. The reaction mixture was heat to 45 °C for 30
min. After TLC indicated reaction completion (Rf value of desired product was 0.1 smaller than that of the starting material CH$_2$Cl$_2$/MeOH, 95/5, v/v), the solution was concentrated to thick oil under reduced pressure and redissolved in 100 ml CH$_2$Cl$_2$. The organic solution was washed by water 100 ml X 2, brine 100 ml X 2, dried over Na$_2$SO$_4$, evaporated to dryness to afford crude product. The crude substance was then purified by column chromatography on SiO$_2$ (CH$_2$Cl$_2$/MeOH, 100/0-98/2, v/v). The result pure compound (4) was obtained (5.2 g, 81%) as white foam. $^1$H NMR (400 MHz, CDCl$_3$) $\delta$ 8.34 (s, 1H, 8-H), 8.07 (s, 1H, 2-H), 6.50 (d, $J$ = 7.1 Hz, 1H, 1'-H), 6.24 (s, 2H), 4.74 (t, 1H, 3'-H), 4.20 (dd, $J$ = 12.9, 3.2 Hz, 1H, 5'-H), 4.08 (dd, $J$ = 12.8, 2.6 Hz, 1H, 5''-H), 3.93 – 3.84 (m, 1H, 4'-H), 3.78 (dd, $J$ = 9.6, 7.2 Hz, 1H, 2'-H), 1.95 (s, 3H, SeCH$_3$), 1.32 – 0.83 (m, 28H, 2 x (CH$_3$_2)2CHSi). $^{13}$C NMR (101 MHz, CDCl$_3$) $\delta$ 155.88 (C+6), 153.06 (C-2), 149.96 (C-4), 139.20 (C-8), 119.61(C-5), 84.59 (C-1'), 84.26 (C-4'), 75.14 (C-3'), 61.36 (C-5'), 49.95 (C-2'), 17.67, 17.60, 17.55, 17.51, 17.31, 17.29, 17.25, 17.17, 13.89, 13.20, 13.17, 12.76 (2 x (CH$_3$_2)CHSi), 5.80 (SeCH$_3$). HRMS (ESI): calc. for C$_{23}$H$_{42}$N$_5$O$_4$SeSi$_2$ [M+H]$^+$: 588.1941, found 588.1938.

3',5'-O-(Tetraisopropyldisiloxane-1,3-diyl)-6-N,N-Dibenzoyl-2'-deoxy-2'-selenomethyl-arabinoadenosine (5) Starting material (4) 5 g (8.52 mmol) was placed in a 100 ml round bottom flask, dried under high vacuum and dissolved in 25 ml dry pyridine. Benzoyl chloride 3.96 ml (34.82 mmol, 4 eq) was added drop wise under ice bath. After the addition, ice bath was removed; the reaction mixture was stirred at room temperature. The reaction mixture was evaporated to dryness under reduced pressure, co-evaporated with toluene 5 ml X 2 and dissolved in 100 ml CH$_2$Cl$_2$. The solution was washed with water 100 ml X 2, brine 100 ml X 2, dried over Na$_2$SO$_4$ and evaporated. The crude product was purified by flash chromatography.
(SiO₂, CH₂Cl₂/MeOH, 100/0-99.5,0.5 v/v) to afford pure compound (5) 5.7 g (84%) as white foam. ¹H NMR (400 MHz, CDCl₃): δ 8.65 (s, 1H, 8-H), 8.27 (s, 1H, 2-H), 7.84 (d, J = 7.6 Hz, 4H, Bz), 7.45 (m, 2H, Bz), 7.32 (t, J = 7.5 Hz, 4H, Bz), 6.52 (d, J = 7.1 Hz, 1H, 1'-H), 4.70 (t, J = 8.6 Hz, 1H, 3'-H), 4.16 (dd, J = 12.7, 3.2 Hz, 1H, 5'-H), 4.07 (d, d, J = 11.4 Hz, 1H, 5''-H), 3.94 – 3.83 (m, 1H, 4'-H), 3.77 (t, J = 8.4 Hz, 1H, 2'-H), 1.83 (s, 3H, SeCH₃), 1.23 – 0.93 (m, 28H, 2 x (CH₃)₂CHSi).

¹³C NMR (101 MHz, CDCl₃) δ 172.30 (C=O), 153.01 (C+6), 152.14 (C+2), 151.94 (C+4), 143.71 (C+8), 134.26, 132.96, 129.55, 128.72 (Bz), 127.88 (C-5), 84.99 (C+1'), 84.35 (C-4'), 75.29 (C-3'), 61.49 (C-5'), 50.04 (C-2'), 17.60, 17.56, 17.52, 17.44, 17.28, 17.19, 17.12, 13.80, 13.14, 13.07, 12.70 (2 x (CH₃)₂CHSi), 5.82 (SeCH₃). HRMS (ESI): calc. for C₃₇H₅₀N₅O₆SeSi₂ [M+H]+: 796.2465, found 796.2437.

**6-N,N-Dibenzoyl-2'-deoxy-2'-selenomethylarabinoadenosine (6)**

Compound (5) 5.7 g was dried on high vacuum and dissolved in 50 ml dry THF. 0.83ml (7.2 mmol, 1 eq) of Et₃N.3HF was added. The solution was heat to 40 °C for 15 min. The solvent was evaporated under reduced pressure and residue was dissolved in 100 ml ethyl acetate, washed with water 100 ml X 2, brine 100 ml X 2, dried over Na₂SO₄, evaporated to dryness. The crude product was then purified by column chromatography (SiO₂, CH₂Cl₂/MeOH, 100/0-96/4, v/v). 2.9 g (73%) pure compound (6) was obtained as white foam. ¹H NMR (400 MHz, CDCl₃) δ 8.73 (s, 1H, 8-H), 8.63 (s, 1H, 2-H), 7.81 (d, J = 7.6 Hz, 4H, Bz), 7.45 (t, J = 7.3 Hz, 2H, Bz), 7.32 (t, J = 7.6 Hz, 4H, Bz), 6.54 (d, J = 7.0 Hz, 1H, 1'-H), 5.08 – 4.68 (br, 1H, OH), 4.50 (t, J = 8.8Hz, 1H, 3'-H), 3.84-3.61 (m, 4H, 5'-H, 5''-H, 4'-H, OH), 1.71 (s, 3H, SeCH₃). ¹³C NMR (101 MHz, CDCl₃) δ 172.53 (C=O), 152.83 (C-6) 152.14 (C-2), 151.52 (C-4), 144.87 (C-8), 133.83, 133.36, 129.50, 128.92 (Bz), 127.17 (C-5), 86.03 (C-1'), 85.04 (C-4'), 72.43 (C-3'), 59.51 (C-5'), 49.78 (C-2'), 5.13

5'-O-(4,4'-dimethoxytrityl)-6-N,N-Dibenzyoyl-2'-deoxy-2'-selenomethylarabinoadenosine (7) Compound (6) 1.35 g (2.44 mmol) was co-evaporated with dry pyridine 5 ml X 2 and then dissolved in pyridine 15 ml. The flask was then placed in ice bath. Dimethoxytrityl chloride 1g (2.93 mmol, 1.2 eq) was dried under vacuum, dissolved in 15 ml dry pyridine and injected to the reaction flask. The ice bath was removed after the addition. The reaction mixture was stirred in room temperature for 2 h. The reaction was quenched by adding 1 ml of methanol. The reaction mixture was evaporated under reduced pressure and co-evaporated with toluene 5 ml X 2. The residue was dissolved in 100 ml CH₂Cl₂, washed with NaHCO₃ 100 ml X 2, dried over Na₂SO₄ and evaporated. The result crude product was purified by column chromatography on SiO₂ (CH₂Cl₂/MeOH/Et₃N, 99/0/1-98.5/0.5/1, v/v/v). Yield: 1.71 g (82%) of compound (7) as white foam. 1H NMR (400 MHz, CD₂Cl₂) δ 8.57 (s, 1H, 8-H), 8.31 (s, 1H, 2-H), 7.82 (d, J = 7.7 Hz, 4H, Ar), 7.56 – 7.17 (m, 15H, Ar), 6.82 (d, J = 7.6 Hz, 4H, Ar), 6.64 (d, J = 6.8 Hz, 1H, 1'-H), 4.42 (d, J = 7.8 Hz, J = 7.6 Hz, 1H, 3'-H), 4.09 (m, 1H, 4'-H), 3.74 (s, 7H, 2 X OCH₃, 2'-H), 3.49 (m, 2H, 5'-H,5''-H), 1.72 (s, 3H, SeCH₃). ¹³C NMR (101 MHz, CD₂Cl₂) δ 172.71 (C=O), 159.21 (Ar), 153.42 (C-6), 152.47 (C-2), 152.08 (C-4), 145.16 (Ar), 144.51 (C-8), 136.29, 136.15, 134.62, 133.44, 130.62, 130.58, 129.81, 129.18, 128.69, 128.38 (Ar), 128.24 (CPh₃), 127.38 (C-5), 113.65 (Ar), 87.07 (C-1’), 86.12 (C-4’), 84.29 (Ar), 75.89 (C-3’), 63.62 (C-5’), 55.71 (2 X OCH₃), 50.85 (C-2’), 5.65 (SeCH₃).
5'-O-(4,4'-dimethoxytrityl)-6-N,N-Dibenzoyl-2'-deoxy-2'-' selenomethyralabinoadenosine-3'-(2-Cyanoethyl)-N,N-diisopropyl]-phosphoramidite (8)

Under argon, compound (7) 854 mg (1 mmol) was dried and dissolved in dry CH₂Cl₂ 15 ml with N,N-diisopropylethylamine 0.52 ml (3 mmol, 3 eq). 2-cyanoethyl N,N-diisopropylchlorophosphoramidite 355 mg (1.5 mmol, 1.5 eq) was added slowly and the solution was stirred at room temperature for 2 h. The reaction mixture was evaporated to dryness then purified by column chromatography on SiO₂ (CH₂Cl₂/MeOH/Et₃N, 99/0/1-98.5/0.5/1, v/v/v). A mixture of diastereoisomers (8) was obtained as white foam. Yield: 740 mg (70%). HRMS (ESI): calc. for C₅₅H₆₀N₈O₈PSe [M+H]⁺: 1056.3328, found: 1056.3309; calc. for C₅₅H₆₀N₈O₈PSe [M+Na]⁺: 1078.3147, found: 1078.2892.

2.2.2 Synthesis of 2’-SeANA-dG phosphoramidite

Scheme 2.6 Synthesis of 2’-SeMe-ANA-dG phosphoramidite
3',5'-O-(1,1,3,3-Tetraisopropyldisiloxane-1,3-diyl)-N2-isobutyrylguanosine (10) 9-[(β-D-Ribofuranosyl] guanine (9) 4.28 g (15.10 mmol) was placed in a 250 mL round bottom flask, dried over high vacuum and co-evaporated with anhydrous pyridine 2 x 15mL. Anhydrous DMF 80 mL, anhydrous pyridine 80 mL was added under argon. After the mixture was cooled in ice bath, 1, 3- dichloro-1, 1, 3, 3 -tetraisopropylsiloxane (TIPS-Cl2) 5 g (15.85 mmol, 1.05 eq) was added drop wise with stirring. The reaction was then warmed up to room temperature and allowed to proceed over night. The reaction mixture was then cooled in ice bath, trimethsilyl chloride (TMS-Cl) 11.50 mL (90.6 mmol, 6eq) was added slowly to the reaction mixture. The reaction flask was removed from ice bath and stirred in room temperature for 3 h. The reaction flask was chilled by ice bath again, isobutryl chloride (iBu-Cl) 1.90 mL (18.12 mmol, 1.2 eq) was added drop wise, and the reaction mixture was stirred in room temperature for additional 2 h. The reaction solution was then poured into 1 L ice-cold saturated aqueous NaHCO3 solution with vigorous stirring. The resulted white precipitant was filtered, washed with water 2 x 200 mL and dissolved in CH2Cl2 200 mL. The organic solution was washed by water two more times (200 mL each), dried over anhydrous MgSO4 and filtered. CF3COOH (TFA) 2.31 mL (30.20 mmol, 2 eq) was dissolved in 25 mL dioxane and added to the CH2Cl2 solution. The mixture was stirred at room temperature for 2h, and then extracted by saturated aqueous NaHCO3 solution (3 x 200mL). The organic layer was collected, dried over anhydrous MgSO4 and evaporated to dryness. The crude product was then purified by flash chromatography on silica gel to afford compound (10) as white form. Yield: 90%. (CH2Cl2/MeOH, 100/0 -98/2, v/v) Yield: 4.95 g, 55 % over four steps. 1H NMR (400 MHz, CDCl3) δ 12.33 (s, 1H, H-N1), 10.98 (s, 1H, H-N-iBu), 7.94 (s, 1H, H-C8), 5.70 (s, 1H, H-C1’), 4.39 (dd, 1H, H-C3’), 4.25 (d, 1H, H-C2’), 4.12 (m, 2H, H-C4’, H-
C5’a), 3.97 (d, 1H, H-C5’b), 2.93 (pentet, 1H, iBu), 1.2 (t, 6H, iBu), 1.15-0.82 (m, 28H, iPr). $^{13}$C NMR (101 MHz, CDCl$_3$) δ 180.19 (C=O), 156.07 (C6), 148.42 (C2), 148.08 (C4), 136.63 (C8), 121.01 (C5), 89.02 (C1’), 81.57 (C4’), 75.23 (C2’), 69.39 (C3’), 60.43 (C5’), 35.99 (iBu), 19.04 (iBu), 19.01 (iBu), 17.43, 17.27, 17.21, 17.01, 16.91, 16.78, 13.35, 12.92, 12.84, 12.47 (iPr).

2'-O-[(Trifluoromethyl)sulfonyl]-3’,5’-O-(1,1,3,3-Tetraisopropyldisiloxane-1,3-diyl)-N2-isobutryrylguanosine (11) Starting material (10) 4.8 g (8.05 mmol), 4-dimethyaminopyridine 2.95 g (24.15 mmol, 3 eq) was placed in a round bottom flask, dried over high vacuum and then dissolved in 75 mL anhydrous CH$_2$Cl$_2$. After the reaction flask was cooled by ice bath, CF$_3$SO$_2$Cl (1.3 mL, 2.03 g, 12.08 mmol) was added drop wise. The reaction was stirred at 0°C for 1 h, and then diluted by addition of 100 mL CH$_2$Cl$_2$; the solution was washed by saturated aqueous NaHCO$_3$ solution (100 mL X 3), dried over anhydrous MgSO$_4$, filtered and evaporated to dryness. The crude product was purified by flash chromatography on silica gel. (CH$_2$Cl$_2$/MeOH, 100/0 - 98.5/1.5, v/v) Yield: 4.33g 74 %. $^1$H NMR (400 MHz, CDCl$_3$) δ 12.15 (s, 1H, H-N1), 8.90 (s, 1H, H-N-iBu), 7.96 (s, 1H, H-C8), 6.09 (s, 1H, H-C1’), 5.32 (d, J = 4.2 Hz, 1H, H-C2’), 4.69 (dd, J = 9.2, 4.2Hz), 1H, H-3’), 4.26 (d, J = 13.6 Hz, 1H, H-C5’a), 4.10 (d, J = 9.2 Hz, 1H, H-C4’) 4.03 (d, J = 13.6 Hz, 1H, H-C5’b), 2.72 (pentet, 1H, iBu), 1.24 (m, 6H, iBu) 1.18-0.99 (m, 28H, iPr). $^{13}$C NMR (101 MHz, CDCl$_3$) δ 178.91 (C=), 155.49 (C6), 148.19 (C2), 147.29 (C4), 136.01 (C8), 122.00 (C-5), 116.97 (CF$_3$), 88.28 (C1’), 86.43 (C2’), 81.62 (C4’), 67.49 (C3’), 59.28 (C5’), 36.47, 19.06, 19.00 (iBu), 17.49, 17.35, 16.82, 13.39, 13.03, 12.95, 12.86 (iPr). HRMS (ESI): calc. for C$_{27}$H$_{43}$F$_3$N$_5$O$_9$SSi [M-H]$^-$: 726.2293, found: 726.2272.
3′,5′-O-{1,1,3,3-Tetraisopropyldisiloxane-1,3-diyl}-N2-isobutyryl-2′-deoxy-2′-selenomethylarabinoguanosine (12) NaBH₄ 851 mg (22.52 mmol, 4eq) was placed a sealed 250 mL round bottom flask, dried under high vacuum for 10 min, purge with dry argon twice to deplete oxygen. To the flask, 70 mL dry dioxane was injected, followed by the addition of 1.60 mL of CH₃SeSeCH₃ (3.18 g, 16.89 mmol, 3eq). The yellowish color suspension was stirred at room temperature and 7 mL of anhydrous ethanol was added drop wise. The result solution was stirred at room temperature for 3 h. To this yellow solution, staring material (11) 4.1 g (5.63 mmol) in dry dioxane 20 mL was injected. The reaction mixture was heat to 45 °C for 30 min. After TLC indicated reaction completion, the solution was concentrated to thick oil under reduced pressure and re-dissolved in 100 mL CH₂Cl₂. The organic solution was washed by water 100 mL X 2, brine 100mL X 2, dried over MgSO₄, evaporated to dryness to afford crude product. The crude substance was then purified by column chromatography on silica gel (CH₂Cl₂/MeOH, 100/0-98/2, v/v). The result pure compound (12) was obtained 5.2 g (Yield: 81 %) as white foam. ¹H NMR (400 MHz, CDCl₃) δ 12.38 (s, 1H, H-N1), 10.96 (s, 1H, H-N-iBu), 7.91 (s, 1H, H-C8), 6.17 (d, J = 6.2 Hz, 1H, H-C1’), 4.37 (m, 1H, H-C3’), 4.09-3.99 (m, 2H, H-C4’, H-C5’a), 3.78 (d, H-C5’b), 3.68 (m, 1H, H-C2’), 2.96 (m, 1H, iBu), 1.88 (s, 3H, SeCH₃), 1.19-0.99 (m, 34H, iBu, iPr). ¹³C NMR (101 MHz, CDCl₃) δ 180.49 (C=O), 156.27 (C6), 149.16 (C4), 148.25 (C2), 136.95 (C8), 120.25 (C5), 84.08 (C1’), 83.83 (C4’), 73.05 (C3’), 60.46 (C5’), 49.45 (C2’), 35.96, 19.20, 19.11(iBu), 17.51, 17.38, 17.34, 17.30, 17.10, 17.02, 16.77, 13.79, 13.03, 12.97, 12.55(iPr), 5.53 (SeCH₃). HRMS (ESI): calc. for C₂₇H₄₆N₅O₆SeSi₂ [M-H]⁺: 672.2133, found: 672.2152.
N2-isobutyryl-2′-deoxy-2′-selenomethylarabinoguanosine (13) Compound (12) 5 g (7.43 mmol) was dried under high vacuum and dissolved in 75 mL dry THF, and then treated with 3HF.Et$_3$N 1.20 g, 1.21 mL (7.43 mmol, 1 eq) at 40 °C for 15 min. The solvent was evaporated; the residue was purified by flash column chromatography on silica gel to afford pure compound (13) 2.3 g (Yield: 72 %) as white solid. $^1$H NMR (400 MHz, DMSO) δ 12.09 (s, 1H, H-N1), 11.73 (s, 1H, H-N-iBu), 8.21 (s, 1H, H-C8), 6.27 (d, $J = 7.0$ Hz, 1H, H-C1′), 5.81 (s, 1H, OH), 5.17 (s, 1H, OH), 4.36 (m, 1H, H-C3′), 3.78-3.70 (m, 4H, H-C2′, H-C4′, H-C5′ab), 2.77 (m, 1H, iBu), 1.88 (s, 3H, SeCH$_3$), 1.11 (d, $J = 6.6$ Hz, 6H, iBu) $^{13}$C NMR (101 MHz, DMSO) δ 180.17 (C=O), 154.87 (C6), 148.43 (C4), 148.06 (C2), 137.76 (C8), 119.44 (C5), 85.00 (C1′), 83.99 (C4′), 73.14 (C3′), 59.51 (C5′), 48.72 (C2′), 34.75, 18.90, 18.85 (iBu), 3.98 (SeCH$_3$). HRMS (ESI): calc. for C$_{15}$H$_{22}$N$_5$O$_5$Se [M+H$^+$]: 432.0769, found: 432.0768.

5′-O-(4,4′-dimethoxytrityl)-N2-isobutyryl-2′-deoxy-2′-selenomethylarabinoguanosine (14) Compound (13) 2.0 g (4.65 mmol) was dried over high vacuum, co-evaporated with anhydrous pyridine 10mL X 2 and suspended in 50 mL dry pyridine. The white suspension was cooled by ice bath then treated with dimethoxytrityl chloride (DMTrCl) 1.97 g (5.12 mmol, 1.1 eq). The reaction was allowed to warm up to room temperature and stirred overnight. Solvent was removed under vacuum; orange color residue was dissolved in CH$_2$Cl$_2$, washed with water 100mL x 2, brine 100mL x 2, dried over MgSO$_4$ and evaporated to dryness. The crude product was purified by flash column chromatography, in which 1% Et$_3$N was added to the eluent to maintain basic environment. (CH$_2$Cl$_2$/MeOH, 100/0-95/5, v/v) 3.1 g of pure compound (14) was obtained as yellow solid. $^1$H NMR (400 MHz, CD$_2$Cl$_2$): δ 12.02 (s, 1H, H-N1), 8.93 (s, 1H, H-N-iBu), 7.71 (s, 1H, H-C8), 7.32-7.09 (m, 9H, trityl-H), 6.68 (d, 4H, trityl-H), 6.19 (d, $J = 7.3$ Hz, 1H, H-C1′),
4.56 (m, 1H H-C3’), 3.93 (m, 1H, H-C4’), 3.65 (s, 6H, 2×OCH3), 3.55 (m, 2H, H-C2’, H-C5’a), 3.40 (m, 1H, H-C5’b), 2.61 (pentet, 1H, iBu), 1.78 (s, 3H, SeCH3), 1.21 - 1.13 (m, 6H, iBu). 13C NMR (101 MHz, CDCl3): δ 179.88 (C=O), 158.59 (trityl), 155.87 (C6), 148.31 (C2), 147.77 (C4), 144.63 (trityl), 138.43 (C8), 136.02, 135.95, 130.22, 129.28, 128.43, 128.00, 127.86, 126.96 (trityl), 120.90 (C5), 113.16 (trityl), 86.68 (CPh3), 86.07 (C1’), 84.40 (C4’), 76.28 (C3’), 63.82 (C5’), 55.32 (2×OCH3), 49.13 (C2’), 36.47, 19.24, 18.95 (iBu), 5.11 (SeCH3). HRMS (ESI): calc. for C_{36}H_{40}N_{5}O_{7}Se [M+H]^+: 734.2103, found: 734.2093.

5’-O-(4,4’-dimethoxytrityl)-N2-isobutyryl-2’-deoxy-2’-selenomethylarabinoguanosine

3’-[(2-Cyanoethyl)-N,N-diisopropyl]-phosphoramidite (15) Compound (14) (0.4 g, 0.54 mmol) was dissolved in a 5ml anhydrous dichloromethane under argon, diisopropylethyl amine (0.28ml, 3eq) was added. 2-Cyanoethyl N,N-diisopropylchlorophosphoramidite (0.165 mg, 6.97 mmol, 1.3 equiv) was slowly added and the solution was stirred at room temperature for 2 h. The crude product was purified by column chromatography on SiO2 (CH2Cl2/MeOH/ Et3N, 98/0/2-97.5/0.5/2 v/v/v). Yield: 417 mg of 7 as colorless foam (88%). 31P NMR (162MHz, CD2Cl2): δ 149.72, 149.69. HRMS (ESI): calc. for C_{45}H_{57}N_{7}O_{8}PSe [M+H]^+: 934.3171, found: 934.3171. Using N,N,N,N-tetraisopropyl phosphoramidite could be an alternative phosphorylation method.

2.2.3 Synthesis of 2’-SeMeANA-dU derivative

Initially, the method used for synthesizing 2’-SeMeANA-dA and 2’-SeMeANA-dG was directly applied to 2’-SeMeANA-dU synthesis where 2’-OH was converted to a leaving group then displaced by methylselenide nucleophile. However, this synthetic route could only lead to a mixture of 2,2’-anhydro uridine derivative and a 2’-SeMe-deoxyuridine derivative. (Shown in Scheme 2.7)
2.2.3.1 Attempted synthesis of 2′-SeANA-dU nucleoside

3′,5′-O-(1,1,3,3-Tetraisopropyldisiloxane-1,3-diyl)-uridine (17) Uridine (16) 2 g (8.19 mmol) was placed in a 50 mL round bottom flask, dried over high vacuum and co-evaporated with anhydrous pyridine 2 x 5mL. Anhydrous pyridine 30 mL was added under argon to dissolve compound (16). After the mixture was cooled in ice bath, 1, 3- dichloro-1, 1, 3, 3- -tetraisopropyldisiloxane (TIPSICl$_2$) 2.75 ml (2.71 g, 8.60 mmol, 1.05 eq, d = 0.986 g/ml) was added drop wise with stirring. The reaction was then warmed up to room temperature and allowed to proceed for 3 hrs. The reaction solution quenched with 5 ml of brine, and then concentrated to an oil under reduced pressure. The residue was then redissolved in 100 ml of DCM and washed with brine 50 ml x 2, water 50 ml x 2, dried over anhydrous MgSO$_4$, filtered, evaporated to dryness. The resulting sticky foam is then pumped on high vacuum to give a white foam (17) in near quantitative yield. Compound (17) was used in next step without further purification.

Scheme 2.7 Attempted synthesis of 2′-SeANA-dU nucleoside

2′-O-Methylsulfonyl-3′,5′-O-(1,1,3,3-Tetraisopropyldisiloxane-1,3-diyl)-uridine (18) Compound (17) 650 mg (1.34 mmol) was placed in a 10 ml round bottom flask, dried over high vacuum, and dissolved in 5 ml anhydrous DCM. Under ice bath, triethylamine 0.56 ml ( 3 eq, 3.02 mmol, d = 0.726 g /ml) was added to the solution, followed by methanesulfonyl chloride
0.21 ml (2 eq, 2.68 mmol, d = 1.48 g / ml). The reaction was allowed to proceed for 30 min at room temperature, and then quenched by adding 2 ml of brine. The mixture was poured in to 75 ml water. The water layer was extracted with DCM 50 ml x 3. Organic extracts are then pooled and dried over anhydrous MgSO₄, filtered, and evaporated to dryness. The crude substance was then purified by column chromatography on silica gel (CH₂Cl₂/MeOH, 100/0-98/2, v/v). ¹H NMR (400 MHz, CDCl₃): δ 10.40 (s, 1H), 7.81 (d, J = 8.2 Hz, 1H), 5.81 (s, 1H), 5.72 (d, J = 8.1 Hz, 1H), 5.00 (d, J = 4.5 Hz, 1H), 4.41 – 4.11 (m, 2H), 4.04 (dd, 1H), 3.96 (dd, 1H), 3.26 (s, 3H), 1.05 – 0.93 (m, 28H). ¹³C NMR (101 MHz, CDCl₃): δ 163.70, 150.54, 138.76, 102.45, 88.56, 82.74, 81.98, 66.61, 59.02, 39.28, 17.41, 17.34, 17.25, 17.20, 16.89, 16.81, 16.76, 13.43, 12.92, 12.83, 12.54.

NaBH₄ 53 mg ( 1.42 mmol, 4eq) was placed a sealed 10 mL round bottom flask, dried under high vacuum for 10 min, purge with dry argon twice to deplete oxygen. To the flask, 3 mL dry dioxane was injected, followed by the addition of 0.1 mL of CH₃SeSeCH₃ ( 200 mg, 1.06 mmol, 3 eq, d = 1.987 g / ml ). The yellowish color suspension was stirred at room temperature and 0.5 mL of anhydrous ethanol was added drop wise. The result solution was stirred at room temperature for 30 min. To this yellow solution, starting material (18) 200 mg ( 0.354 mmol) in dry dioxane 1 mL was injected. The reaction mixture was heat to 45 °C for 30 min. TLC indicates formation of two major products: one with more polarity (19) and one with less polarity (20). The solution was concentrated to thick oil under reduced pressure and redissolved in 10 mL CH₂Cl₂. The organic solution was washed by water 10 mL X 2, brine 10mL X 2, dried over MgSO₄, evaporated to dryness to afford crude product. The crude substance was then purified by
column chromatography on silica gel (CH$_2$Cl$_2$/MeOH, 100/0-98/2, v/v) to give pure compound (19) and compound (20).

Base on the NMR and MASS data, compound (19) and compound (20) were identified as 3',5'-O-(1,1,3,3-Tetraisopropyldisiloxane-1,3-diyl)-2,2'-anhydrouridine (19), and 3',5'-O-(1,1,3,3-Tetraisopropyldisiloxane-1,3-diyl)-2'-selenomethyl-deoxyuridine (20). Below is a proposed mechanism of this reaction. (Scheme 2.8) Due to the high acidity of N-3 proton (p$_{ka}$ 9.3), it can be easily deprotonated by the methyselenide nucleophile. Followed by electron transfer, a 2, 2’ cyclic compound can form predominately by an intermolecular cyclization reaction to gave compound (19). The 2,2'-anhydro ring on compound (19) can be further opened by the methylselenide on the ribo position of sugar. As a result, compound (20) was formed as the second major product in this reaction.

**Scheme 2.8 Mechanism of generating compound (19)**

In summary, the presence of 2 carbonyl oxygen on uracil ring promotes the formation of a 2,2'-anhydouridine derivative during the reaction. The methylselenide nucleophile can only open the cyclic ring from the ribo position of the sugar to give 2'-SeMe-deoxyuridine derivative as the final product.
2.2.3.2 Select N3 protecting groups

In order to avoid the intramolecular cyclization reaction during the process which compromises the stereochemistry control on 2’-position. A protecting group on N-3 position is necessary since the deprotonation of N-3 initializes the intramolecular cyclization reaction. Several N-protecting groups were considered and tested. (Shown in Figure 2.4)

Figure 2.4 N-3 protecting groups for compound (18)

The protection group must be relatively stable to survive the strong nucleophilic condition, and it should also be removable after the reaction without causing damage to the 2’-SeMe-ANA analog.

Benzoyl protecting groups was commonly used as N protection groups in oligonucleotide synthesis (N4 in cytidine and N6 in adenosine). It considered to be stable against all conditions involves in solid-phase synthesis, and can be removed by a base, most often with aqueous ammonia or methylamine. However, our data has shown that benzoyl protection group is very unstable during the nucleophilic reaction. A mixture of compound (18), compound (19), and compound (20) was obtained after the reaction, (Shown in Scheme 2.9) which indicate that the benzoyl protecting groups was removed by methylselenide nucleophile even before the substitution reaction on 2’-position.
Scheme 2.9 Using benzoyl as N3 protecting group

Benzyl is the most stable protection groups in our candidates; it is widely used in sugar and nucleoside chemistry and it can be readily removed by hydrogenolysis. However, its superior stability causes trouble in the deprotection step. The hydrogenolysis didn’t work on N-benzyl protected-SeMe-uridine derivatives. Probably it is because of the selenium functionally “poisoned” the catalyst during the hydrogenation reaction.

PMB (p-methoxybenzylonl) protecting group is more labile than benzyl; oxidation reagent like cerium ammonium nitrate or DDQ can oxidatively cleave it. Catalytic hydrogenation and acid treatment are also alternative choices for PMB deprotection. The PMB protecting groups works very well during the nucleicphilic reaction. The N3-PMB-2’-SeMe-arabinouridine derivative was obtained successfully. However, during the oxidative deprotection reaction, a strong polar compound was obtained, which was later confirmed to be the N3-PMB-2’-Se(OH)2Me-arabinouridine derivative by Mass spectrometry.(Shown in Figure 2.5) The oxidative deprotection process damages the selenium functionality rather than efficient removal of PMB.
Figure 2.5 Oxidation of selenium functionality during deprotection

Boc (tert-Butyloxycarbonyl) group is commonly used in solid phase polypeptide synthesis. It can be removed by concentrated strong acid. (Such as CF$_3$COOH) Our study has shown that Boc protecting group can be easily introduced to N3 position through a 2’-O-TMS transient protection with high yield. More importantly, it can survive the strong nucleophilic condition, and can be readily removed by a TFA treatment. Below is the modified synthetic route for 2’-SeMe-ANA-dU nucleoside using Boc as the N3 protecting group.

2.2.3.3 Improved synthesis of 2’-SeMe-ANA-dU nucleoside

2’-O-Trimethylsilyl-3’,5’-O-(1,1,3,3-Tetraisopropylsiloxane-1,3-diyl)-N3-(tert-Butyloxycarbonyl)-uridine (21) Uridine (16) 2.63 g (10.78 mmol, MW: 244.20) was co-evaporated with anhydrous pyridine twice (5 ml each time), suspended in 50ml anhydrous pyridine under
argon, then the mixture was cooled in an ice bath. To the suspension, 3.40 g of 1,3-dichloro-1,1,3,3-tetraisopropyl-disiloxane (10.78 mmol, 1.00 eq, MW: 315.43) was added drop wise. The flask was removed from ice bath when finished adding and stirred for 3h at room temperature. After TLC indicates reaction completion (Rf of product: 0.45, dichloromethane - methanol, 95/5, v/v), 4.1 mL (3.5 g, 32.32 mmol, 3eq, MW: 108.64, d: 0.856 g/ml) of trimethylsilyl chloride was added drop wise to the reaction mixture. The reaction was stirred for 15 min at room temperature and then poured into 2L of cold saturated sodium bicarbonate solution with vigorous stirring. The oily residue was collected and dissolved in CH$_2$Cl$_2$ 150 ml, washed with water 75 ml x 2, brine 75 ml x 2, dry over anhydrous Na$_2$SO$_4$. The drying agent was filtered and the solution was evaporated to dryness to afford crude product (21) as white foam.

![Scheme 2.10 Improved synthesis of 2'-SeMe-ANA-dU nucleoside](image)

**Scheme 2.10 Improved synthesis of 2'-SeMe-ANA-dU nucleoside**

**3',5'-O-(1,1,3,3-Tetraisopropylsiloxane-1,3-diyl)-N3-(tert-Butoxycarbonyl)-uridine**

(22) Di-tert-butyl dicarbonate 3.53 g (16.17 mmol, 1.5eq, MW: 218.25), DMAP 130 mg (0.1 eq, MW: 122.17) was added the round-bottom flask containing crude product (2) and dried under
high vacuum. Anhydrous pyridine 50 ml was added to the flask and the reaction was stirred at room temperature until the TLC indicated completion of reaction. ( Rf of product : 0.62, ethyl acetate – hexane, 15/85, v/v). The pale yellow solution was evaporated to dryness and dissolved in 200 ml solvent mixture ( dichloromethane - dioxane, 3/1, v/v). 4.1 g of p-toluenesulfonic acid monohydrate (2eq, MW: 190.22) was added to the solution, the solution was stirred for 15 min at room temperature and poured into 600 ml of saturated sodium bicarbonate solution. The water layer was extracted with 100 ml dichloromethane twice. The organic layers were combined, dried over anhydrous MgSO₄ and evaporated to dryness. The crude product was purified by flash chromatography on SiO₂. (CH₂Cl₂-MeOH, 100/0-99/1, v/v). Yield pure product (21) 7.1g (84 % in two steps) as colorless solid. HRMS (ESI): calc. for C₂₆H₄₇N₂O₉Si₂ [M+H]⁺: 587.2820, found: 587.2838.

2'-O-Tosyl-3',5'-O-\{1,1,3,3-Tetraisopropylsiloxane-1,3-diyl\}-N3-\{tert-Butoxycarbonyl\}-uridine (23) Compound (22) 4g (6.8 mmol), and DMAP 883 mg (6.8 mmol, 1 eq) was placed in a 100 ml round bottom flask, dried under high vacuum, and dissolved in 60 ml anhydrous DCM. To the solution, Et₃N 2.85 ml (2.06 g, 20.4 mmol, 3 eq, d = 0.7255 g / ml ) was added, followed by drop wise addition of 4-toluenesulfonyl chloride 1.69 g ( 8.84 mmol, 1.3 eq ) in 3 ml DCM. The reaction was allowed to stir at room temperature for 30 min, and then diluted by adding 50 ml of DCM. The organic solution was washed by water 50 mL X 2, brine 50mL X 2, dried over MgSO₄, evaporated to dryness to afford crude product. The crude substance was then purified by column chromatography on silica gel (Hexane/Ethyl Acetate, 100/0 - 90/10, v/v) to afford compound (23) as white solid (Yield: 91 %).
3',5'-O-(1,1,3,3-Tetraisopropyldisiloxane-1,3-diyl)-N3-(tert-Butoxycarbonyl)-2'-SeMe-arabinouridine (24) NaBH₄ 673 mg (17.82 mmol, 6eq) was placed a sealed 100 mL round bottom flask, dried under high vacuum for 10 min, purge with dry argon twice to deplete oxygen. To the flask, 30 mL dry dioxane was injected, followed by the addition of 1.1 mL of CH₃SeSeCH₃ (2.2 g, 11.88 mmol, 4eq, d = 1.987 g / ml). The yellowish color suspension was stirred at room temperature and 5 mL of anhydrous ethanol was added drop wise. The result solution was stirred at room temperature for 30 min. To this yellow solution, starting material (22) 2.2 g (2.97 mmol) in dry dioxane 10 mL was injected. The reaction mixture was stirred at room temperature until all the starting material is converted to a less polar compound. The solution was concentrated to thick oil under reduced pressure and redissolved in 100 mL CH₂Cl₂. The organic solution was washed by water 50 mL X 2, brine 50mL X 2, dried over MgSO₄, evaporated to dryness to afford crude product. The crude substance was then purified by column chromatography on silica gel (Hexane/Ethyl Acetate, 100/0-90/10, v/v) to give result compound (23) as white foam. (Yield: 56 %) ¹H NMR (400 MHz, CDCl₃): δ 7.82 (d, J = 14.2 Hz, 1H), 6.01 (d, J = 14.2 Hz, 1H), 5.64 (d, J = 7.3 Hz, 1H), 4.69 (dd, J = 7.2, 2.8 Hz, 1H), 4.45 (s, 1H), 4.03 (dd, J = 17.6, 7.3 Hz, 1H), 3.69 (dd, J = 11.3, 7.9 Hz, 1H), 2.26 (s, 3H), 1.64 – 1.22 (s, 9H), 1.06 (m, 28H). ¹³C NMR (101 MHz, CDCl₃): δ 189.16, 161.30, 148.86, 147.88, 131.61, 128.62, 110.97, 106.34, 85.40, 83.53, 82.42, 78.76, 68.87, 64.10, 62.30, 32.17, 26.97, 25.37, 16.45, 16.30, 15.93, 12.64, 12.39, 12.04, 11.65, 10.83, 3.96, 0.76. HRMS (ESI): calc. for C₂₇H₄₈N₂NaO₈SeSi₂ [M+Na⁺]: 687.2012, found: 687.1992.
3’,5’-O-(1,1,3,3-Tetraisopropyldisiloxane-1,3-diyl)-2’-SeMe-arabinouridine (25)

Compound (24) 500 mg (0.73 mmol) was placed in a 25 ml round bottom flask. 50% TFA in DCM 3 ml was added to the flask. The mixture was stirred for 10 min at room temperature, and then diluted by adding 20 ml DCM. The organic solution was washed by 100 ml saturated aqueous sodium bicarbonate solution twice, dried over MgSO₄, and evaporated to dryness. The crude was purified by flash column chromatography on silica gel (Hexane/Ethyl Acetate, 100/0 – 80/20, v/v) to afford pure compound (25) as white foam (Yield: 90 %). ¹H NMR (400 MHz, CDCl₃) δ 7.84 (d, J = 14.1 Hz, 1H), 6.30 (s, 1H), 5.91 (d, J = 14.1 Hz, 1H), 5.67 (d, J = 7.3 Hz, 1H), 4.23 (dd, J = 16.7, 6.4 Hz, 2H), 4.18 – 4.01 (m, 1H), 4.01 – 3.73 (m, 2H), 2.29 (s, 3H), 1.1 – 0.98 (m, 28H). ¹³C NMR (101 MHz, CDCl₃) δ 189.08, 153.99, 132.37, 108.86, 84.59, 82.36, 79.00, 61.33, 61.00, 16.41, 16.34, 16.25, 16.20, 15.95, 15.91, 12.41, 12.19, 11.75, 11.57, 3.83. HRMS (ESI): calc. for C₂₂H₃₉N₂O₆SeSi₂ [M+H]+: 563.1512, found: 563.1505.

2’-SeMe-arabinouridine (26) Compound (25) 300 mg (0.53 mmol) was dissolved in 5 mL dry THF, and then treated with 86 uL of 3HF.Et₃N (85 mg, 0.53 mmol, 1 eq) at 40 °C for 15 min. The solvent was evaporated; the residue was purified by flash column chromatography on silica gel (DCM/MeOH, 100/0 - 95/5, v/v) to afford pure compound (26) as white solid (Yield: 85 %). ¹H NMR (400 MHz, MeOD) δ 7.78 (d, J = 14.0 Hz, 1H), 5.97 (m, 2H), 4.13 (m, 3H), 3.58 – 3.41 (m, 2H), 2.25 (s, 3H). ¹³C NMR (101 MHz, MeOD): δ 193.16, 154.99, 132.49, 107.87, 87.36, 86.85, 75.20, 61.67, 60.62, 3.11. HRMS (ESI): calc. for C₁₀H₁₃N₂O₅Se [M-H]−: 320.9999, found: 320.9984.
2.2.3.4 Stability test of 2’-SeMe-ANA-dU nucleoside

There is one last challenge of making 2’-SeMe-ANA-dU labeled oligonucleotide: Stability. Those labels which are current available for DNA/RNA solid phase synthesis have one common property - they must be stable to the strong basic conditions required for oligonucleotide deprotection. The conventional oligonucleotide deprotection condition was carried out using ammonium hydroxide overnight at 55 °C or AMA (ammonium hydroxide/methylamine, 1:1, v/v) 5 min at 65 °C. Our 2’-SeMe-ANA-dU nucleoside decomposed rapidly in ammonium hydroxide even at room temperature.

![Scheme](image)

**Scheme 2.11** 2’-SeMe-ANA-dU to 2’-OMe-dU conversion under Ultra Mild condition

By using a combination of Pac-dA, Ac-dC, iPr-Pac-dG monomers, and phenoxyacetic anhydride as capping solution (“Ultra Mild reagents”) in solid-phase synthesis, the deprotection of oligonucleotide can be accomplished with 0.05 M potassium carbonate in methanol at room temperature. However, even at “Ultra Mild” condition, 2’-SeMe-ANA-dU nucleoside can be converted to 2’-OMe-dU with 30 minutes. (Shown in **Scheme 2.11**) The conversion was confirmed by $^1$H NMR (shown in **Figure 2.6**) and ESI-Mass as well (Appendix??).
Figure 2.6 2'-SeMe-ANA-dU to 2'-OMe-dU conversion (1H NMR)

Unfortunately, after carefully studied the stability of 2'-SeMe-ANA-dU derivatives in various oligonucleotide deprotection conditions, we have to draw a conclusion that 2'-SeMe-ANA-dU analog is not compatible with DNA/RNA solid-phase synthesis.

2.2.3.5 Alternative method of making 2'-SeMe-ANA-dU modified DNA

Instead of incorporating modified nucleoside into DNA by solid-phase synthesis, enzymatic DNA synthesis could be an alternative method to make modified DNA. In an enzymatic DNA polymerization reaction, polymerase utilized triphosphate as building blocks to synthesize DNA at physiological condition. This method represents a versatile platform for the facile preparation of modified nucleic acid at mild condition. In order enzymatically incorporate 2'-SeMe-ANA-dU analog into DNA, corresponding 2'-SeMe-ANA-dU triphosphate must be prepared.
2.2.3.6 Synthesis of 2'-SeMe-ANA-dU triphosphate

Scheme 2.12 Synthesis of 2'-SeMe-ANA-dUTP

The 2'-SeMe-ANA-dU nucleoside (26) (20 mg) was dried under high vacuum overnight, dissolved in trimethyl phosphate (0.4 mL), and stirred in ice bath. A separate flask with proton-sponge (2 eq, 55 mg) in trimethyl phosphate (0.3 mL) was injected into solution of compound (26) at 0 °C. After 3-5 min, a diluted phosphorus oxychloride (POCl₃, 10-fold diluted in trimethyl phosphate, 1.5 eq, 0.09 mL) was added into the reaction dropwise. The reaction was tracked by TLC plate and completed in 1 to 2 hr. Tributylammonium pyrophosphate (dissolved in 0.2 ml tributylamine and 0.4 ml DMF) was then quickly injected into reaction flash. After vigorously stirred for 5 min, the reaction was quenched with triethylammonium bicarbonate (1 M, 3 mL) and stirred for another 1 hr in room temperature to obtain compound (27). Crude reaction solution was precipitated in ethanol/NaCl (3 M NaCl, 0.5 mL; ethanol 14.5 mL), followed by 1 hr freeze in -80°C and centrifugation for 15 min at 12,000 rpm. The pellet was re-dissolved in water, analyzed and purified by RP-HPLC. Compound (27) HRMS (ESI): calc. for C₁₀H₁₆N₂O₁₄P₃Se [M-H]⁻: 560.8980, found: 560.8979. (Shown in Figure 2.7)

The resulting 2'-SeMe-ANA-dUTP can be incorporated into DNA by enzymatic polymerization reaction.
Figure 2.7 HRMS-ESI(−) analysis of 2′-SeMe-ANA-dUTP

2.2.4 Synthesis of 2′-SeMe-ANA modified DNA by solid-phase synthesis

Solid-phase DNA synthesis was performed on an ABI3400 DNA/RNA synthesizer with 1.0 μmol scale. 2′-MeSe-Ara-G phosphoramidite was prepared as 0.1 M concentration solution in anhydrous acetonitrile prior of use. Coupling was carried out with 0.3 M 5-
Benzylmercaptotetrazole (5-BMT) in acetonitrile as activator. The coupling time for 2'-MeSe-Ara-G phosphoramidite was set to 300 s. All native phosphoramidite and solution for DNA automate synthesis were purchased from Glen Research: Deblocing Solution: 3% CCl₃COOH in CH₂Cl₂; Capping Mix A: THF/AC₂O (9:1); Capping Mix B: 16% 1-Methylimidazole in THF/pyridine; Oxidation Solution: 0.02 M I₂/THF/Pyridine/H₂O. All the DNAs were prepared in DMTr-on mode.

After synthesis, Se-DNA was cleaved from the CPG beads and fully deprotected by treating with AMA solution (40% methylamine: concentrated ammonium hydroxide, 1:1, v/v) at 65°C for 15 min. The volatile amines were evaporated on a speed-vac concentrator. After removal of CPG-Beads by centrifugation, the oligonucleotides were purified by reversed-phase high performance liquid chromatography (RP-HPLC) twice, with both DMTr-on and DMTr-off.

The HPLC purification was performed on a Welchrom XB-C18 column (21.1 x 250 mm) with buffer A (20 mM TEAAc, pH 7.1) and buffer B (20 mM TEAAc, pH 7.1, in 50% ethanol). The pump was set with a flow rate of 6 ml/min and a linear gradient of 0% buffer B to 100% buffer B over 20 min and stay at 100% buffer B for 10 min. The monitor was set to 260 nm wavelength, DMTr-on product was collected at ~25 min, frozen, and lyophilized. The lyophilized Se-DNAs with DMTr-on were treated with 30% trifluoroacetic acid solution (final concentration: 0.3% w/w) for 2 min, followed by the addition of TEAAc buffer (2 M) to adjust the pH to neutral. The Se-DNAs with DMTr-off were purified by RP-HPLC again; the retention time was about 15 min, followed by lyophilization. The purified DNAs were re-dissolved in water and analyzed by HPLC, kinase gel and MALDI-MS to confirm the quality. (MALDI-MS analysis of 2'-SeMe-ANA-dG modified DNAs is shown in Table 2.1.)
2.2.5 Thermal denaturation study of 2′-SeANA containing DNA

The melting temperatures of the MeSeANA-G derivatized duplexes as well as the native duplexes were determined on Cary 300 UV-Vis spectrometer equipped with a six-sample cell changer and a temperature controller. The sample annealed by heated to 85 °C for 2 min and cooling down slowly to 5 °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 °C/min and data interval of 0.5 °C. DNA duplexes samples were prepared in buffer containing 100 mM NaCl, 10 mM NaH₂PO₄-Na₂HPO₄ (pH 6.5), 2 mM MgCl₂, with 1 μM DNA duplex. Melting data are shown in Table 2.2.

### Table 2.1 MALDI-TOF MASS analysis of 2′-SeMe-ANA-dG containing DNA

<table>
<thead>
<tr>
<th>DNA sequences</th>
<th>Calc. [M+H]⁺</th>
<th>Observed</th>
</tr>
</thead>
<tbody>
<tr>
<td>5′-CGCGAATTCCGCG-3′</td>
<td>3740.4</td>
<td>3740.6</td>
</tr>
<tr>
<td>5′-CGCGAATTCCGCG-3′</td>
<td>3740.4</td>
<td>3740.6</td>
</tr>
<tr>
<td>5′-CCAGGCCTGG-3′</td>
<td>3123.0</td>
<td>3123.0</td>
</tr>
<tr>
<td>5′-CCGTTACCCGG-3′</td>
<td>3123.0</td>
<td>3123.2</td>
</tr>
<tr>
<td>5′-ATGGTGCTC-3′</td>
<td>2823.8</td>
<td>2823.5</td>
</tr>
</tbody>
</table>
The UV melting experiment was carried out to study the impact of 2’-SeMe-ANA-dG modification on duplex thermal stability. For this study, two modified Dickerson-Drew dodecamer were synthesized with 2’-SeMe-ANA-dG modification close to either 3’-end or 5’-end. The data suggests about 15 °C drop in melting temperature when modification was close to 3’-end of the oligo. When the modification site was close to 5’-end, an ~10 °C decrease of melting temperature was also observed. The duplex destabilization effect was also observed on other 2’-SeMe-ANA-dG modified duplexes. Our data clearly showed that the modification brings 4 – 7 °C drop on melting temperature. This result is opposite to Ronsenburg’s observation where 2’-F-ANA-T modification could increase the thermal stability of B-form DNA \(^{59}\). This significant differences on melting temperatures between 2’-SeMe-ANA-G and 2’-F-ANA-
T modified sequences could be explained by the formation of a favorable inter-residual pseudo-hydrogen bond (F...H8) in 2’-F-ANA-T containing sequences.

2.2.6 Crystallization of 2’-SeMe-ANA modified nucleic acids

In order to further prove our hypothesis that 2’-SeMe-ANA modification is compatible with B-form DNA crystallization, we did a crystal growth study with the well-studied Dickerson-Drew dodecamer sequence (5’-CGCGAATTCGCG-3’). Twenty-four crystallization conditions were initially screened for 2’-SeMe-ANA-dG modified Dickerson-Drew dodecamer (5’-CGCG_{Se}AATTCGCG-3’) using hanging drop vapor diffusion method. Crystallization buffers used for initial screening are from a commercial crystal screening kit. (Nucleic Acid Mini screen, Hampton Research, CA) Most crystals appeared within two weeks after setup crystallization. All initial hits contain 10% v/v MPD, 40mM Na cacodylate, 12mM spermine tetra-HCl and a combination of NaCl, KCl as well as MgCl$_2$ salt.

Crystals suitable for diffraction experiments were obtained under the following conditions: 2uL of a 1 mM dodecamer solution (single strand) were mixed with 2ul buffer solution (10% 2-methylpentane-2,4-diol (MPD), 40 mM sodium cacodylate pH 7.0, 12 mM spermine tetrahydrochloride, 80 mM sodium chloride, and 20 mM magnesium chloride) and equilibrated against 500 uL of a 35% (v/v) MPD reservoir solution. Crystals for diffraction data collection were mounted in a nylon loop and shock-frozen in liquid nitrogen. (Shown in Figure 2.9)
Figure 2.9 DNA crystal in cryo-loop at ALS 8.2.2

2.2.7 Structural study of 2’-SeMe-ANA-dG containing DNA

2.2.7.1 MAD data collection and phasing

Figure 2.10 Fluorescent Scan of crystal

Multiple wavelength data was collected on SIBYLS (Structurally Integrated Biology for Life Sciences) beam line 12.3.1 of the ALS (Advanced Light Source) at the Lawrence Berkeley National Laboratory. The unit cell of the selected crystal was determined as a) 24.17 Å, b) 40.28 Å, and c) 66.14 Å with orthorhombic space group P2_12_12_1. The precise locations of the inflection point and peak for the selenium absorption edge were determined from X-ray fluorescence spectra (Shown in Figure 2.10). Data at three wavelengths (Peak: 0.9794 Å, Inflection: 0.9796, Remote: 0.9686 Å) from a single crystal were collected to a maximum resolution of 1.59 Å were
integrated and scaled with the programs HKL-2000 and DENZO/SCALEPACK \(^{62}\). The statistics of these three datasets are shown in Table 2.3.

### Table 2.3 MAD data collection statistics

<table>
<thead>
<tr>
<th>MAD Data Collection</th>
<th>Peak</th>
<th>Inflection</th>
<th>Remote</th>
</tr>
</thead>
<tbody>
<tr>
<td>Space group</td>
<td></td>
<td>P 21 21 21</td>
<td></td>
</tr>
<tr>
<td>Cell Dimensions</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a, b, c (Å)</td>
<td></td>
<td>24.17, 40.28, 66.14</td>
<td>90.00, 90.00, 90.00</td>
</tr>
<tr>
<td>α, β, γ (°)</td>
<td></td>
<td>90.00, 90.00, 90.00</td>
<td></td>
</tr>
<tr>
<td>Wavelength Å</td>
<td>0.9794</td>
<td>0.9796</td>
<td>0.9686</td>
</tr>
<tr>
<td>Resolution range Å</td>
<td>50-1.74 (1.77-1.74)</td>
<td>50-1.59 (1.62-1.59)</td>
<td>50-1.60 (1.63-1.60)</td>
</tr>
<tr>
<td>Rmerge %</td>
<td>12.5 (35.2)</td>
<td>8.2 (46.4)</td>
<td>8.2 (45.2)</td>
</tr>
<tr>
<td>Completeness %</td>
<td>97.3 (100)</td>
<td>99.7 (99.5)</td>
<td>99.8 (100)</td>
</tr>
<tr>
<td>I/σ(I)</td>
<td>38.80 (7.88)</td>
<td>46.72 (4.9)</td>
<td>50.2 (5.67)</td>
</tr>
<tr>
<td>Redundancy</td>
<td>6.5 (5.7)</td>
<td>6.3 (5.3)</td>
<td>6.7 (5.9)</td>
</tr>
</tbody>
</table>

MAD phasing was carried out using SHELXC/D/E program \(^{63}\). The data files were prepared by SHELXC. SHELXD then searched for 2 selenium atoms using 95 attempts. The initial phases were improved by density modification using SHELXE with 60% solvent content. The computed Fourier electron density maps were visualized with the program COOT \(^{64}\). (Shown in Figure 2.11) A clear density of a duplex DNA was shown in the density map (Figure 2.11 A), and clear electron density on 2'-arabino position of modified-dG was also observed (Figure 2.11 B). A complete turn of B-form DNA model was build based on the electron density map.

![A.](image1)

![B.](image2)

Figure 2.11 Fourier electron density maps provided by MAD phasing
2.2.7.2 High-resolution data collection, structure determination, and analysis

High-resolution data of 2'-SeMe-ANA-dG modified Dickerson Drew Dodecamer was collected at beam line 8.2.2 of the ALS (Advanced Light Source) at the Lawrence Berkeley National Laboratory. The unit cell of the selected crystal was determined as a) 25.78 Å, b) 39.90 Å, and c) 65.69 Å with orthorhombic space group P2_12_12_1. The data was collected to a maximum resolution of 1.49 Å, integrated and scaled with the programs HKL2000 and DENZO/SCALEPACK. The statistics of the high-resolution datasets are shown in Table 2.4 (4KW0).

Figure 2.12 Global and local structures of the 2'-SeMe-ANA-dG modified DNA DickersonDrew dodecomer (5'-CGCGAATTCCCG-3')_2 at 1.49 Å resolution. The orange balls represent the selenium atoms.

(A) The overall superimpose comparison of 2'-SeMe-ara-G modified (5'-CGCGAATTCCCG-3')_2 (Red; PDB ID: 4KW0) with its native structure (Blue; PDB ID: 355D) with a RMSD value of 0.229
(B) The local superimpose comparison of 2′-SeMe-ara-G modified (5′-CGCGAATTCCG-3′), with its native (Blue; PDB ID: 355D)

(C) The (2Fo-Fc) electron-density map of the base pair with 2′-SeMe-ANA-dG modification.

The model generated from MAD phasing was used as the initial model for refinement. The refinement was carried out with COOT \textsuperscript{64} and REFMAC 5.5.0109 \textsuperscript{65} with CCP4 \textsuperscript{66}, setting aside 5% of the total reflections for calculating the R-free \textsuperscript{67}. 89 water molecules and a magnesium hexahydrate ion were included to the duplex structure. The final R-work and R-free are 20.3 and 24.4%, respectively. An example of the final electron density is shown in Figure 2.12 and refinement parameters are listed in Table 2.4 (4KW0). Structure factor and final coordinate have been deposited in the Protein Data Bank (PDB ID: 4KW0).

The structure of 2′-SeMe-ANA-dG containing DNA (PDB ID: 4KW0, in red, Figure 2.12) is superimposable over the native B-DNA structure (PDB ID: 4KW0, in blue), which has the same orthorhombic space group (P2\textsubscript{1}2\textsubscript{1}2\textsubscript{1}). Both selenium modified and native structures have an magnesium hexahydrate ion in the major groove. The overall r.m.s.d. 2′-SeMe-ANA-dG modified DNA over the native is 0.229 Å. 2′-SeMe-ANA-dG/T base pair is also superimposable over native G/T base pair (shown in Figure 2.12), with a r.m.s.d. of 0.026 Å. The selenium functionality locates in the major groove of duplex. The 2′-SeMe-ANA-dG modification causes no structural perturbation when compared with its native form.
Figure 2.13 Hydration pattern comparison between 2’-SeMe-ANA-dG and dG
(A) 2’-SeMe-ANA-dG local structure (PDB ID: 4KW0) Cyan sphere and dash lines represent water molecule and hydrogen bond respectively.
(B) dG local structure (PDB ID: 246D) Red sphere and dash lines represent water molecules and hydrogen bond respectively. C. Superimposed local structures.

The impact of 2’-SeMe-ANA-dG modification on duplex hydration was also studied. We chose an atomic resolution crystal structure of Dickerson Drew Dodecamer (PDB ID: 346D) to make a direct compare. (Shown in Figure 2.13) In 436D structure, 4 water molecules locate around dG22/2’ with a distance of 3.59, 3.60, 3.71 and 4.38 Å, respectively. (Shown in Figure 2.13B) In the superimposed structure (shown in Figure 2.13C), it is quite clear that the 2’-SeMe functionality repulsed the water molecule away and causes minor hydration change in the major groove of duplex. However, in our case, the structural perturbation brought by major groove hydration change is minimum.

2.2.8 Structural study of 2’-SeMe-ANA-dG/protein complex

2.2.8.1 Crystallization

The large fragment of the *Bacillus Stearothermophilus* DNA polymerase protein was prepared by Dr. Jianhua Gan at Prof. Huang’s laboratory. 2’-SeMe-ANA-dG modified Dickerson Drew Dodecamer (5’-CG\_\_\_CGAATTGCG-3’) was annealed by heating the DNA solution to 90°C
for 1 min and then allowing it to cool slowly to 25°C. The resulting Se-DNA duplex was mixed with the protein (final concentration of 5 mg/ml) in a 1.5:1 duplex DNA to protein molar ratio in a buffer containing 100mM NaCl, 20mM Tris-HCl, pH 7.4. Initial crystallization screening of DNA/protein complex was achieved by using commercial crystal screening kits by sitting drop vapor diffusion method (Index HT, Natrix HT from Hampton Research, CA. Detailed buffer conditions of each kit are available in Appendix) During the crystallization screening process, Protein/DNA solution was mixed with crystallization buffer at 1:1 ratio (0.3uL+0.3uL) and equilbrate against 100uL of the same crystallization buffer. Crystals appeared within two weeks after setup crystallization.

Initial hits from the screening were repeated and further optimized to acquire crystals with better diffraction quality. Crystal for data collection was prepared with a crystallization buffer contains 25% PEG3350, 0.2 M ammonium sulfate, and 0.1 M Bis-Tris, pH 6.5 (Natrix 7B) at 25°C. Prior to data collection, Crystals were flash frozen in liquid nitrogen using the respective reservoir solution and 25% glycerol as a cryoprotectant.

2.2.8.2 Diffraction data collection and structure refinement

Diffraction data were collected from crystals of the 2'-SeMe-ANA-dG DNA/protein complex on beamline 8.2.2 of the ALS (Advanced Light Source) at the Lawrence Berkeley National Laboratory. A number of crystals were scanned in order to find those with better diffraction quality. X-ray data were collected under a liquid-nitrogen stream at 99 K. Each crystal was exposed for 2 second per image with 1° rotation, and a total of 180 images were obtained. The unit cell of the selected crystal was determined as a) 86.77 Å, b) 93.97 Å, and c)
106.04 Å with orthorhombic space group $P_2_1_2_1_2$. The data was collected to a maximum resolution of 2.20 Å. The dataset was integrated and scaled with the programs HKL2000 and DENZO/SCALEPACK. The structure was solved by molecular replacement method using Phaser within CCP4. The protein part of the 4DSL structure was used as the searching model. The resulting model was refined using REFMAC5.0.0109 within CCP4. The DNA/RNA duplex was modeled into the structure by using Coot. Metal ions, sulfate ions, glycerol molecules, and water molecules were added either automatically or manually using Coot. The final R-work and R-free are 17.6% and 22.8%, respectively. Data-collection, and structure-refinement statistics are given in Table 2.4.

2.2.8.3 Structure of 2′-SeMe-ANA-dG DNA/Bacillus fragment complex

The crystal structure of fragment DNA polymerase I from Bacillus stearotherophilus (Residues 298 – 876) with DNA modified with selenium at the 2′-arabino position of guanosine was determined at 2.20 Å resolution. Structure factor and final coordinate have been deposited in the Protein Data Bank (PDB ID: 4O0I). The quality of the electron density maps allowed for unambiguous identification of the nucleotide sequence along the DNA chain. The electron density of 11 nucleotides DNA duplex was clearly showed in $2Fo-Fc$ map. The density of last base pair was missing due to disorder. One 2′-SeMe-ANA-dG modification site was identified close to the active center. From the electron density map, 2′-SeMe functionality can undoubtedly recognized (shown in Figure 2.14).
Figure 2.14 Structure of 2’-SeMe-ANA DNA/protein complex
(A) The crystal structure of fragment DNA polymerase I from Bacillus stearotherophilus (Residues 298 – 876) with 2’-SeMe-ANA-dG modified DNA at 2.20 Å resolution
(B) Electron density map of 2’-SeMe-ANA-dG modification site in complex structure with $\sigma = 1$

<table>
<thead>
<tr>
<th>Crystalgraphic data collection and refinement statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Crystals</strong></td>
</tr>
<tr>
<td><strong>4KW0</strong></td>
</tr>
<tr>
<td><strong>Space group</strong> P 21 21 21</td>
</tr>
<tr>
<td><strong>Cell Dimensions</strong></td>
</tr>
<tr>
<td>$a$, $b$, $c$ (Å) 25.78, 39.90, 65.69</td>
</tr>
<tr>
<td>$\alpha$, $\beta$, $\gamma$ (°) 90.00, 90.00, 90.00</td>
</tr>
<tr>
<td>Wavelength (Å) 0.969</td>
</tr>
<tr>
<td>X-ray source ALS 8.2.2</td>
</tr>
<tr>
<td>Resolution range (Å) 15.0-1.49 (1.54-1.49)</td>
</tr>
<tr>
<td>No. of unique reflections 10484</td>
</tr>
<tr>
<td>Completeness (%) 90.56 (88.45)</td>
</tr>
<tr>
<td>$R_{merge}$ (%) 7.5 (30.5)</td>
</tr>
<tr>
<td>Mean $I$/sigma(I) 14.75 (2.93)</td>
</tr>
<tr>
<td>Redundancy 5.1 (3.8)</td>
</tr>
<tr>
<td>Resolution (Å) 15.0-1.49</td>
</tr>
<tr>
<td>$R_{work}$ 0.2030 (0.2044)</td>
</tr>
<tr>
<td>$R_{free}$ 0.2440 (0.2400)</td>
</tr>
<tr>
<td>Number of non-hydrogen atoms 580</td>
</tr>
<tr>
<td>macromolecules 490</td>
</tr>
<tr>
<td>water 89</td>
</tr>
<tr>
<td>Protein/NA residues 24</td>
</tr>
<tr>
<td>RMS(bonds) 0.026</td>
</tr>
<tr>
<td>RMS(angles) 2.31</td>
</tr>
<tr>
<td>Average B-factor 25.8</td>
</tr>
<tr>
<td>macromolecules 48.8</td>
</tr>
<tr>
<td>ligands 24.7</td>
</tr>
<tr>
<td>solvent 34.6</td>
</tr>
<tr>
<td><strong>4O0I</strong></td>
</tr>
<tr>
<td><strong>Space group</strong> P 21 21 21</td>
</tr>
<tr>
<td><strong>Cell Dimensions</strong></td>
</tr>
<tr>
<td>$a$, $b$, $c$ (Å) 86.77, 93.97, 106.04</td>
</tr>
<tr>
<td>$\alpha$, $\beta$, $\gamma$ (°) 90.00, 90.00, 90.00</td>
</tr>
<tr>
<td>Wavelength (Å) 1</td>
</tr>
<tr>
<td>X-ray source ALS 8.2.2</td>
</tr>
<tr>
<td>Resolution range (Å) 30.0-2.2 (2.28-2.2)</td>
</tr>
<tr>
<td>No. of unique reflections 42209</td>
</tr>
<tr>
<td>Completeness (%) 94.83 (75.57)</td>
</tr>
<tr>
<td>$R_{merge}$ (%) 8.1 (37)</td>
</tr>
<tr>
<td>Mean $I$/sigma(I) 12.94 (2.83)</td>
</tr>
<tr>
<td>Redundancy 5.6 (4.1)</td>
</tr>
<tr>
<td>Resolution (Å) 30.0-2.2</td>
</tr>
<tr>
<td>$R_{work}$ 0.1760 (0.2669)</td>
</tr>
<tr>
<td>$R_{free}$ 0.2280 (0.3758)</td>
</tr>
</tbody>
</table>
2.3 Conclusion

In this paper, we demonstrate a new selenium detrivatization method to address crystallization, phasing and high-resolution structure determination of B-form DNA and DNA/protein complexes. 2’-SeMe-ANA-dG phosphoramidite along with 2’-SeMe-ANA-dA phosphoramidite, and 2’-SeMe-ANA-dU triphosphosphate were successfully synthesized, which enables both chemical or enzymatic incorporation of 2’-SeMe-ANA modification into DNA. We also demonstrate the solid-phase synthesis, deprotection, and purification of 2’-SeMe-ANA-dG modified oligonucleotides. Despite the fact that our thermal dynamic studies of 2’-SeMe-ANA-dG containing DNAs showed decreased duplex stability, the modified DNA can be readily crystalized under broad crystallization conditions within short period of time. More importantly, the crystals obtained could diffract to high resolution (1.49 Å). As a prove of principle, we also present the convenient structure determination via MAD phasing of B-form DNA by 2’-SeMe-ANA labeling. The global and local structure of 2’-SeMe-ANA-dG modified DNA showed minimal perturbation when compare to native one. Remarkably, the 2’-SeMe-ANA modification could adopt B-DNA’s 2’ endo sugar pucker in duplex without any drastic geometry change. Furthermore, we successfully co-crystallized and solved the structure of 2’-SeMe-ANA modified DNA with a fragment of DNA polymerase I from Bacillus stearotherophilus. The modification site was very well accommodated in the protein/DNA complex structure, which indicates its excellent compatibility protein/DNA complex structure study.

By integrating easy accessibility, crystallization compatibility, convenient SAD/MAD phasing, negligible structure perturbation, this 2’-SeMe-ANA-G modification could be further
exploited in structure determination of B-form DNA, quadruplex, DNA nanostructure as well as DNA/drug, DNA/protein complexes
3 HIGH SPECIFICITY OF RNA BASE PAIRING WITH 2-SE URIDINE

3.1 Introduction

3.1.1 U/G wobble base pair in nature

The Watson-Crick base pair system is essential for biological information flow. Base-pair recognition passes genetic information through replication, transcription, and translation to protein. In the mRNA decoding process, a tRNA recognizes more than one codon by non-Watson Crick base pairing with the third base of a codon, which allows 64 codon combinations to be translated by a much less number of tRNAs. Four decades ago, Francis Crick raised the wobble hypothesis to explain this degeneracy of genetic code. In his paper, he hypothesized a possible formation of G/U base pairs between the first base of the anticodon and the third base of the codon results. His prediction was later confirmed by crystal structure of yeast phenylalanine tRNA.

![U/G wobble pair and codon degeneracy](image)

Figure 3.1 U/G Wobble pair and codon degeneracy

3.1.2 Biophysical properties of U/G wobble base pair

The G-U wobble base pair is a fundamental unit of RNA secondary structure in almost every class of RNAs in nature. It plays important roles in RNA folding, providing recognition signal and RNA catalysis. The G/U wobble pairs present a unique arrangement of
hydrogen bond donors/acceptor pattern. The carbonyl oxygen on 2 position of uracil is not involved in hydrogen bonding when paired with A, but forms hydrogen bonds with the imino proton on guanine when paired with G. Another interesting feature of U/G wobble pair is its superior thermo stability compares to Watson-Crick base pairs, which allows it to substitute Watson-Crick base pairs in duplex.

The U/G pairing increases structure and function diversities of RNA and plays critical roles in numerous biological processes. However, the formation of U/G mismatches compromises the high specificity of base pairing, thus reduces the accuracy of the sequence-dependent recognition. For hybridization probes and primers that bind to certain region of genome for diagnostic and therapeutic applications, highly specific base-pairing is extremely important. Developing new methods to provide better sequence recognitions is not just a scientific challenge, but also of practical importance.

3.1.3 Nucleic acid base modification to increase base pairing specificity

By simply replacing the oxygen atoms on nucleobases with other chalcogen elements (such as S and Se) could lead to numerous interesting properties. For example, 2-thiothymidine could be applied to a synthetic biology system to improve replication accuracy. Benner’s lab has reported that the fidelity of PCR reaction involving artificial base pairs (isoC:isoG) could be significantly improved, simply by replacing TTP with 2-thioTTP in PCR reaction. Enhanced base pairing and replication efficiency of thiothymidines was also reported by Kool’s lab. The large chalcogen atoms on nucleobases could be used to sterically alter the fidelity of base pairing. Introducing base modified nucleotides into oligonucleogides by rational design could help to
enhance hybridization selectivity on certain region of the sequence. This property could be further exploited in nucleic acid sequenced-based diagnostic, therapeutic applications and design of nucleic acid nanostructures as well.

3.1.3.1 Sulfur modification on nucleobases

Sulfur modified DNA and RNA is one important category in atom-specific modification of nucleic acids. Like selenium, sulfur belongs to chalcogen family but possesses smaller atomic radius (S: 1.02 Å) compares to selenium (Se: 1.16 Å). Sulfur modified nucleic acids especially sulfur modified non-coding RNAs exist in nature and involved in numerous biological processes. In nature, the thiolations reaction takes place at nucleobases specifically by substitute oxygen atom with sulfur. In E. coli and yeast, the resulting thiolated nucleic acid analogs include 2-thiouridine (s2U), 4-thiouridine (s4U) and 2-thiocytidine (s2C). Moreover, the 2-thiouridine is often located in wobble position 34 of tRNA with various modifications on position 5 and engages in decoding process during protein translation.

To further investigate the functions and roles of sulfur-modified nucleic acids, both synthetic and enzymatic methods have been developed for these natural products. With this advantage, series of research works were carried out including biophysical, biological studies as well as structural studies. The results indicate that thio-modification at position 2 of uridine can enhance the thermostability of the U/A or T/A base pair compare to native one and in vitro experiments the s2U derivatives show preference toward A instead of G during decoding process. The synthesis 4-thiolated nucleic acids are available through solid-phase synthesis. Compare to the uridine with modification on position 2, the thermo stability of 4-thiouridine decreases with no surprise for the season that the atom
on position 4 of U in RNA or T in DNA forms hydrogen bond with A, and with a larger atom like sulfur will certainly increases the distance of the bond thus weaken the overall stability of the modified oligonucleotide. 

### 3.1.3.2 2-Seleno-thymidine

High base paring fidelity is crucial during replication to maintain the genetic integrity. However, the well-known Watson-Crick base pair is not sufficient, it’s been estimated that over 1% mis-incorporation happens during replication. Even though DNA repair mechanism can help pass the genetic information through more accurately, it is still of great importance to optimize the current DNA replication system to offer higher accuracy. As mentioned above, 2-S-T could help to improve base pairing fidelity and replication efficiency, this encouraged us to further investigate the base pair fidelity of 2-Se-T modification. It is worth pointing out that 2-Se derivatized nucleoside had been synthesized for over three decades, but never been incorporated into DNA due to synthetic difficulty. With the assist of phosphoramidite chemistry and solid-phase synthesis, we were able to incorporate 2-Se-T modification into oligonucleotides that can bring base pair recognition and crystallographic study into duplex level.
Scheme 3.1 Synthesis of 2-Se-T phosphoramidite and DNAs.
Reagents and conditions: (a) DMTr-Cl, Pyridine, DMAP, rt; (b) DBU, DMF, CH$_3$I; (c) Se, NaBH$_4$, EtOH; (d) \( \text{CH}_2\text{CH}_2\text{CN, } \text{i-Pr}_2\text{NEt, CH}_2\text{Cl}_2 \); (e) \( (\text{i-Pr}_2\text{N})_2\text{P(Cl)OCH}_2\text{CH}_2\text{CN, (i-Pr)}_2\text{NEt, CH}_2\text{Cl}_2 \); (f) Solid-phase synthesis.

The novel synthesis of 2-Se-T utilized 2-methylthio-T as an active intermediate, followed by selenizing reaction with NaSeH to give 2-seleno-T (shown in Scheme 3.1). To ensure the stability of selenium functionality during solid phase synthesis, cyanoethy groups was attached to selenium to prevent possible deselenization in solid-phase synthesis cycle. Biophysical experiment of 2-Se-T DNAs showed largely increased base pair specificity by destabilizing T/G, T/C pairs. Although 2-exo position is not involved Watson-Crick base pair hydrogen bonding, the weaker hydrogen bonding ability and steric effect of selenium atom could lead to discourage of undesired T/G wobble and T/C pairs. The crystallographic data also suggests that 2-Se-T offers higher basing pairing specificity over native T. This result was also supported by computational data$^{93}$. 
Figure 3.2 Structure of the 2-Se-T-DNA (5'-G\textsuperscript{2-SeMe}UG\textsuperscript{2-SeMe}TACAC-3')\textsubscript{2}

(A). The superimposed comparison of the Se-DNA duplex (3HGD, in green) with its native counterpart (1D78, in cyan). The red balls represent the selenium atoms.

(B). The superimposed comparison of the local $\text{Se}_{2}\text{T4/A5}$ (in green) and native T4/A5 (in cyan) base pairs.

(C). The experimental electron density map of the SeT4/A5 base pair.

3.1.4 2-Se-U RNA modification

The 2-exo-oxygen of uridine serves as a hydrogen bond acceptor in U/G wobble pair, which allows us to tune the base pairing specificity by tailoring the steric and electronic properties at 2-position of uridine. As mentioned before, 2-thiouridine ($s^2\text{U}$) and its C-5 derivatives are found in tRNAs especially on the wobble position. Similar to 2-thio-RNA derivatives, selenium modified RNA are nature-occurring compounds in bacteria tRNAs, such as Escherichia coli, Clostridium sticklandii and etc.\textsuperscript{94} This type of modification was first found at the “wobble” position of the anticodon of tRNA\textsuperscript{95} and has been identified as 5-methylaminometnyl-2-selenouridine (mmm5Se2U)\textsuperscript{96}. Even though, it has been proposed that
this type of selenium derivatized RNA may help to improve the translation efficiency and accuracy, the detailed mechanism is still unclear.

This chapter of my dissertation focuses on synthesis and biophysical studies of 2-Se-U RNA.

![Chemical structures of 2-Se-U/A and 2-Se-U/G base pairs](image)

**Figure 3.3** 2-Se-U/A base pair and 2-Se-U/G base pair

### 3.2 Experimental Section

1H, 13C and 31P NMR data were acquired on a Bruker 400MHz NMR instrument. Chemical shifts (δ values) and coupling constants (J values) are given in ppm and Hz, respectively, using TMS (1H NMR) and solvents (13C NMR) as internal standards. Reagents were purchased from Sigma/VWR/ Chem Genes and used without further purification. All reactions were set up under argon atmosphere. Analytical/preparative thin layer chromatography was carried out on 250u/1000u silica gel 60 coated TLC with F-254 as indicator (Dynamic Adsorbents). Flash chromatography purification was performed on silica gel 60 (230-400mesh) (Dynamic Adsorbents) with gradient elution.
3.2.1 Synthesis of 2-Se-Uridine phosphoramidite

The synthesis of 2-Se-U phosphoramidite was achieved by using a strategy similar to 2-Se-T synthesis\(^\text{90}\). First, we introduced 2-thio uracil to ribose by a glycosylation reaction. The benzoyl protecting groups on 2', 3' and 5'-hydroxyl groups were removed by sodium methoxide in methanol. Dimethoxytrityl protecting groups was introduced to 5'-position. Then, the thio carbonyl was activated by alkylation, and followed by treating with freshly prepared sodium selenide to give 2-Se U as clean product. Then TBDMS was introduced to 2 or 3' hydroxyl group. To avoid being oxidized during the oxidation step in solid synthesis cycle, selenocarbonyl is protected by cyanoethy group. The 2'-protected isomer was isolated and then phosphorylated to give 2-Se-Uridine phosphoramidite as the final product. Synthetic route is shown in Scheme 3.2. The 2-Se-U nucleotide could be incorporated into RNA by solid-phase RNA synthesis.

![Scheme 3.2 Synthesis of 2-Se-Uridine phosphoramidite](image-url)
Reagents and conditions: a) TMS-Cl, HMDS, reflux; b) 1-O-Acetyl-2,3,5-tri-O-benzoyl-beta-D-ribofuranose, SnCl₄, C₂H₄Cl₂, -20°C; c) NaOCH₃, MeOH; d) DMTr-Cl, Pyridine, rt; e) ICH₃, DBU, DMF; f) Se, NaBH₄, EtOH; g) TBDMS-Cl, Imidazole, DMF; h) ICH₂CH₂CN, (i-Pr)₂NEt, CH₂Cl₂; i) (i-Pr₂)₂P(Cl)OCH₂CH₂CN, (i-Pr)₂NEt, CH₂Cl₂

2',3',5'-tri-O-benzoyl-2-thiouridine (28) 2-thiouracil 5.1 g was suspended in hexamethyldisilazane (HMDS, 20eq, 100 mL) with catalytic amount of trimethylsilyl chloride (TMSCl, 0.5 mL). The reaction was refluxed for overnight until a clear yellow solution appeared. After the reaction, HMDS was removed with reduced pressure and the residue was cooled to room temperature and ready for glycosylation reaction. 1-O-Acetyl-2,3,5-tri-O-benzoyl-beta-D-ribofuranose, 10 g, 19.82 mmol was dissolved in 1,2-dichloroethane and added to compound 3 at -20°C under nitrogen, then Tin(IV) chloride (6.9 mL) was added dropwise. The reaction was stirred for 5 hours in room temperature, then washed with a saturated sodium bicarbonate solution and was extracted with dichloromethane. The organic solution was pooled and dried with anhydrous MgSO₄ and evaporated dryness to give crude compound (28) as light yellow form. The crude was later purified by column chromatography silica gel to give the pure compound (28) as white foam. Yield: 84%. ¹H NMR (400 MHz, CDCl₃) δ 10.30 (s, 1H, NH), 8.13 – 8.05 (m, 2H, Ar), 8.05 – 7.97 (m, 2H, Ar), 7.96 – 7.89 (m, 2H, Ar), 7.70 (d, J = 8.2 Hz, 1H, H-6), 7.65 – 7.49 (m, 5H, Ar), 7.43 – 7.35 (m, 4H, Ar), 7.28 (d, J = 4.7 Hz, 1H, H-1'), 5.84 – 5.76 (m, 3H, H-5, H-2',H-3'), 4.88 (dd, J = 12.6, 2.5 Hz, 1H, H-5'), 4.78 (dt, J = 5.3, 2.8 Hz, 1H, H-4'), 4.69 (dd, J = 12.6, 3.1 Hz, 1H, H-5'').

2-thio-uridine (29) Sodium methoxide (6 g, 0.11 mol) and compound (28) (10.5 g, 18.35 mmol) was suspended in methanol 100 ml at room temperature for 6 hours. The reaction mixture was later neutralized with ion-exchange resin (DOWEX 50WX8-400) to pH 7, filtered
and evaporated to dryness. Then the residue was dissolved in water and extracted with dichloromethane. The water layer was separated and evaporated to give the crude compound (29), and the pure compound (29) was obtained by recrystallization in ethanol. Yield: 97%. $^1$H NMR (400 MHz, D$_2$O) $\delta$ 8.15 (d, $J$ = 8.2 Hz, 1H, H-6), 6.64 (d, $J$ = 2.4 Hz, 1H, H-1'), 6.20 (d, $J$ = 8.2 Hz, 1H, H-5), H-4' overlap with HOD signal, 4.41 (m, 1H, H-3'), 4.22 (d, $J$ = 2.4 Hz, H-2'), 4.04 - 3.89 (m, 2H, H-5').

5'-O-4,4'-Dimethoxytrityl-2-thiouridine (30) 2-thiouridine (29) 3 g (11.5 mmol) and DMTr-Cl (4, 4'-dimethoxytrityl chloride, 4.7 g, 13.8 mmol) was weighted and dried under vacuum for 2 hours. A solution of DMTr-Cl in anhydrous pyridine (10-15 mL) was slowly added into a solution of 2-thiouridine in anhydrous pyridine at 0°C. The mixture was allowed to warm up to room temperature and stirred for 5 hours. The reaction was monitored by TLC plate (5% metanol in dichloromethane, Rf = 0.4). The reaction was quenched by adding 5 mL methanol and evaporated to dryness under reduced pressure. The residue was dissolved in dichloromethane and the organic layer was washed with water three times, and then dried over anhydrous MgSO$_4$. The organic solvent was evaporated under reduced pressure to obtain the crude compound (30). The purification of crude compound was performed on flash column chromatography on silica gel using a gradient of methanol in dichloromethane (the column was pre-equilibrated with 1% triethylamine in DCM). Yield 87%. $^1$H NMR (CDCl$_3$) $\delta$: 11.02 (s, 1H, NH), 8.23 (d, $J$ = 8.2 Hz, 1H, H-6), 7.44 – 7.13 (m, 9H, Ar), 6.83 (m, 4H, Ar), 6.43 (s, 1H, H-1'), 5.55 (d, $J$ = 8.2 Hz, 1H, H-5), 4.50 (m, 1H, H-4'), 4.43 (m, 1H, H-3'), 4.35 (s, 1H, OH), 4.19 (d, $J$ = 7.4 Hz, 1H, H-2'), 3.75 (d, $J$ = 1.7 Hz, 6H, OCH3), 3.56 (dd, $J$ = 20.8, 9.5 Hz, 2H, H-5'), 3.41 (br, 1H, OH); $^{13}$C NMR (CDCl$_3$) $\delta$: 175.32 (C-2), 160.70 (C-4), 158.79 (Ar), 144.43 (Ar), 141.08 (C-6), 135.37 (Ar),
135.15 (Ar), 130.30 (Ar), 130.22 (Ar), 128.25 (Ar), 127.33 (Ar), 113.48 (Ar), 106.75 (C-5), 94.37 (C-1'), 87.22 (C-Ar3), 83.78 (C-4'), 75.80 (C-2'), 69.17 (C-3'), 61.21 (C-5'), 55.40 (OCH3); HRMS (ESI): calc. for C30H28N2O7S [M+H]+ = 561.1695, found: 561.1718.

5'-O-4,4'-Dimethoxytrityl-2-methylthiouridine (31) Dry compound (30) (5 g, 8.89 mmol) was dissolved in dry N, N-dimethylformamide (DMF) with iodomethane (5.5 mL, 89 mmol). 1,8-Diazabicyclo[5.4.0]undec-7-ene (DBU, 2 mL, 13.3 mmol) was then added to the reaction mixture at 0°C. The reaction was monitored by TLC (12% Methanol in dichloromethane, Rf = 0.4) and was completed in 4 hours. Ethyl acetate (100 mL) was poured into the mixture and organic layer was washed three times with saturated sodium chloride solution. The organic solution was dried over anhydrous magnesium sulfate, and evaporated under reduced pressure. The residue was purified by flash column chromatography on silica gel (0/100 – 1/9, methanol/dichloromethane, v/v) to give pure compound (31) in 95% yield. 1H NMR (CDCl3) δ: 7.87 (d, J = 7.7 Hz, 1H, H-6), 7.44 – 7.20 (m, 9H, Ar), 6.85 (m, 4H, Ar), 6.11 (br, 1H, OH), 5.88 (d, J = 6.0 Hz, 1H, H-1'), 5.54 (d, J = 7.7 Hz, 1H, H-5), 4.63 (m, 1H, H-4'), 4.44 (m, 1H, H-2'), 4.24 (d, J = 2.3 Hz, 1H, H-2'), 3.75 (d, J = 3.1 Hz, 6H, OCH3), 3.42 (m, 2H, H-5'), 3.40 – 3.30 (br, 1H, OH), 2.55 (s, 3H, SCH3); 13C NMR (CDCl3) δ: 169.19 (C-4), 164.36 (C-2), 158.88 (Ar), 144.49 (Ar), 140.13 (C-6), 135.37 (Ar), 135.22 (Ar), 130.41 (Ar), 130.28 (Ar), 128.32 (Ar), 128.28 (Ar), 127.29 (Ar), 113.54 (Ar), 108.92 (C-5), 91.95 (C-1'), 87.35 (C-Ar3), 84.82 (C-4'), 75.24 (C-2'), 71.63 (C-3'), 63.40 (C-5'), 55.40 (OCH3), 15.39 (SCH3); HRMS (ESI): calc. for C31H34N2O7S [M+H]+ = 577.2008, found: 577.2003.

5'-O-4,4'-Dimethoxytrityl-2-selenouridine (32) Sodium borohydride (NaBH₄, 4.4 g, 0.12 mol) and selenium (6.2 g, 78 mmol) were dried under high vacuum, anhydrous ethanol 100 ml
was added at 0°C to generate an NaSeH solution. The reaction was allowed to stir for 2 hours until a clear solution was observed. The solution was added into a flask contains compound (31) under argon gas and the reaction mixture was stirred for eight hours. The ethanol was evaporated under reduced pressure and the residue was re-dissolved in ethyl acetate. The organic layer was washed with saturated sodium chloride solution for several times until a clear yellow solution was obtained. The resulting solution was dried by anhydrous MgSO₄ and evaporated to dryness under reduced pressure. The crude compound (32) was purified by flash column chromatography on silica gel (0/100 – 5/95, methanol/dichloromethane, v/v) to give pure compound (32) as light yellow foam in 85% yield. ¹H NMR (CDCl₃) δ: 10.95 (s, 1H, NH), 8.24 (d, J = 8.2 Hz, 1H, H-6), 7.44 – 7.19 (m, 9H, Ar), 6.84 (m, 4H, Ar), 6.48 (s, 1H, H-1’), 5.66 (d, J = 8.1 Hz, 1H, H-5), 4.48 (m, 2H, H-4’,H-3’), 4.22 (m, 1H, H-2’), 3.89 (s, 1H, OH), 3.79 (s, 6H, OCH₃), 3.58 (dd, J = 23.6, 9.2 Hz, 2H, H-5’), 2.97 (br, 1H, OH); ¹³C NMR (CDCl₃) δ: 175.74 (C-2), 159.21 (C-4), 158.98 (Ar), 158.94 (Ar), 144.45 (Ar), 140.82 (C-6), 135.38 (Ar), 135.18 (Ar), 130.35 (Ar), 130.27 (Ar), 128.30 (Ar), 128.28 (Ar), 127.45 (Ar), 113.58 (Ar), 108.37 (C-5), 96.86 (C-1’), 87.38 (C-Ar₃), 84.41 (C-4’), 76.33 (C-2’), 69.19 (C-3’), 61.20 (C-5’), 55.48 (OCH₃). HRMS (ESI): calc. for C₃₀H₂₈N₂O₅Se [M+H]⁺: 609.1140, found: 609.1136. UV (MeOH): λmax = 311 nm.

**Compound 33a and 33b** Compound (32) (0.5 g, 0.8 mmol) was dried under vacuum for 2 hours and dissolved in dry N, N-dimethylformamide (DMF). Tert-butyldimethylsilyl chloride (0.15 g, 1 mmol) and imidazole (0.11 g, 1.64 mmol) was added into the reaction solution. The reaction ran overnight in room temperature and was monitored by TLC (15% ethyl acetate in dichloromethane). The reaction mixture was poured into ethyl acetate (50 mL). The organic solution was washed with water 2 x 100 ml, dried over anhydrous MgSO₄, filtered and
evaporated under reduce pressure. The mixture of crude compound (33a) and compound (33b) was purified by flash column chromatography on silica gel without separation of two isomers. (0/100 – 30/70, Ethyl Acetate/dichloromethane, v/v) HRMS (ESI) calc. for C_{36}H_{42}N_{2}O_{7}SeSi [M-H]^- : 723.2005, found: 723.1990.

2'-O-tert-butylidemethylsilyl-5'-O-4,4'-dimethoxytrityl-2-cyanoethylselenouridine (34a)

Anhydrous dichloromethane was added into the mixture of compound (33a) and compound (33b) at 0°C, then iodopropionitrile (0.78 g, 4.31 mmol) and diisopropylethylamine (0.37 mL, 2.15 mmol) were added into the reaction solution. After 4 hours, excess dichloromethane was removed under reduced pressure and the residue was re-dissolved in ethyl acetate. The organic phase was washed with water twice, dried over anhydrous MgSO_4 and evaporated to dryness. The 2'-O-TBDMS and 3'-O-TBDMS isomers was separated by column chromatography with a gradient of ethyl acetate in dichloromethane. The pure compound (34a) has a Rf of 0.35 and compound (34b) has a Rf of 0.30 in 30/70, ethyl acetate/dichloromethane, v/v.

**Compound (34a):** \(^1\)H NMR (CDCl\(_3\)) \(\delta\): 7.96 (d, J = 7.7 Hz, 1H, H-6), 7.53 – 7.11 (m, 9H, Ar), 6.85 (m, 4H, Ar), 5.71 (d, J = 7.7 Hz, 1H, H-5), 5.60 (d, J = 6.5 Hz, 1H, H-1'), 4.61 – 4.49 (m, 1H, H-4'), 4.31 (m, 2H, H-3', H-2'), 3.80 (s, 6H, OCH\(_3\)), 3.54 – 3.34 (m, 4H,H-5', SeCH\(_2\)CH\(_2\)CN), 3.01 (m, 2H, SeCH\(_2\)CH\(_2\)CN), 2.91 (s, 1H, OH), 0.94 (s, 9H, SiCMe\(_3\)), 0.09 (d, 6H, SiMe\(_2\)); \(^{13}\)C NMR (CDCl\(_3\)) \(\delta\): 167.70 (C-4), 159.04 (C-2), 158.51 (Ar), 144.17 (Ar), 139.19 (C-6), 134.91 (Ar), 134.74 (Ar), 130.26 (Ar), 130.17 (Ar), 128.30 (Ar), 128.11 (Ar), 127.58 (Ar), 118.78 (CN), 113.58 (Ar), 110.61 (C-5), 93.13 (C-1'), 87.82 (C-Ar\(_3\)), 85.39 (C-4'), 77.27 (C-2'), 72.40 (C-3'), 63.82 (C-5'), 55.44 (OCH\(_3\)), 25.84 (SiCMe\(_3\)), 24.06 (SeCH\(_2\)CH\(_2\)CN), 18.89 (SeCH\(_2\)CH\(_2\)CN), 18.14 (SiCMe\(_3\)), -4.56 (SiCH\(_3\)),-4.91 (SiCH\(_3\)). HRMS (ESI): calc. for: C\(_{39}\)H\(_{49}\)N\(_3\)O\(_7\)SeSi [M+H]^+: 778.2427, found: 778.2464.
Compound (34b): $^1$H NMR (CDCl$_3$) δ: 8.09 (d, $J = 7.7$ Hz, 1H, H-6), 7.32 (m, 9H, Ar), 6.88 (m, 4H, Ar), 5.75 (d, $J = 7.7$ Hz, 1H, H-5), 5.65 (d, $J = 3.7$ Hz, 1H, H-1’), 4.46 (m, 1H, H-2’), 4.21 (dd, $J = 9.3$, 5.1 Hz, 1H, H-4’), 4.18 – 4.08 (m, 1H, H-3’), 3.83 (s, 6H, OCH$_3$), 3.70 (m, 1H, H-5’), 3.55 (m, 1H, H-5’), 3.41 (m, 2H, SeCH$_2$CH$_2$CN), 3.22 (d, $J = 5.5$ Hz, 1H, OH), 3.08 (m, 2H, SeC$_2$H$_2$CN), 0.90 (s, 9H, SiCMe$_3$), 0.11 (d, 6H, SiMe$_2$).

$^{13}$C NMR (CDCl$_3$) δ: 167.90 (C+4), 159.05 (C+2), 157.82 (C-Ar), 143.99 (Ar), 138.96 (C-6), 135.02 (Ar), 134.89 (Ar), 130.35 (Ar), 130.34 (Ar), 128.37 (Ar), 128.29 (Ar), 127.58 (Ar), 118.95 (CN), 113.56 (Ar), 113.53 (Ar), 110.49 (C-5), 93.84 (C-1’), 87.57 (C-Ar$_3$), 84.72 (C-4’), 76.17 (C-2’), 71.05 (C-3’), 61.71 (C-5’), 55.48 (OCH$_3$), 25.82 (SiCMe$_3$), 24.12 (SeCH$_2$CH$_2$CN), 18.92 (SeCH$_2$CH$_2$CN), 18.17 (SiCMe$_3$), -4.59 (SiCH$_3$), -4.60 (SiCH$_3$).

HRMS (ESI): calc. C$_{39}$H$_{49}$N$_9$O$_7$SeSi [M+H]$^+$: 778.2427, found: 778.2401.

2’-O-tert-butylidimethylsilyl-5’-O-4,4’-dimethoxytrityl-2-cyanoethylselenouridine-3’-

[(2-Cyanoethyl)-N,N-diisopropyl]-phosphoramidite (35) Diisopropylethylamine (15.5 mg, 0.12 mmol) and 2-Cyanoethyl N,N-diisopropylchlorophosphoramide (26 mg, 0.11 mmol) was added to a solution of 10a (100 mg, 0.10 mmol) in anhydrous dichloromathane (5 mL) at room temperature under nitrogen gas. The mixture was stirred for 4 hrs in room temperature and fast run though Al$_2$O$_3$ column then precipitated by dry hexane. The precipitate was filtered, dried under reduced pressure and directly used for solid phase synthesis. $^1$H NMR (CDCl$_3$) δ: 7.95 (d, $J = 7.7$ Hz, 1H, H-6), 7.30 (m, 9H, Ar), 6.84 (m,4H, Ar), 5.82 (d, $J = 7.6$ Hz, 1H, H-1’), 5.63 (d, $J = 7.7$ Hz, 1H, H-5), 4.68 – 4.43 (m, 1H, H-4’), 4.43 – 4.29 (m, 1H, H-1), 4.24 (s, 1H), 3.98 (dd, $J = 15.6$, 8.4 Hz, 2H), 3.80 (s, 6H, OCH3), 3.63 (dd, $J = 19.8$, 7.3 Hz, 4H), 3.42 (dd, $J = 19.3$, 8.7 Hz, 5H), 3.01 (s, 2H), 2.70 (d, $J = 5.8$ Hz, 2H), 1.20 (d, $J = 6.7$ Hz, 18H), 1.08 (d, $J = 6.5$ Hz, 6H), 0.92 (s, 13H), 0.14 – 0.02 (m, 9H); $^{13}$C NMR (CDCl$_3$) δ: 167.83 (C-4), 159.04 (C-2), 158.71 (Ar), 144.16
(Ar), 139.02 (Ar), 134.96 (Ar), 134.74 (Ar), 130.21 (Ar), 130.14 (Ar), 128.38 (Ar), 128.04 (Ar), 127.58 (Ar), 118.86 (SeCH2CH2CN), 117.75 (OCH2CH2CN), 113.67 (Ar), 110.69 (C-5), 92.50 (C-1’) , 87.92 (C-Ar3), 85.43 (C-4’), 77.15 (C-2’), 72.81 (C-3’), 63.72 (C-5’), 59.39 (OCH2CH2CN), 55.48 (OCH3), 43.22 (NCH2), 43.10 (NCMe2); 31P NMR (CDCl3) δ: 148.81, 152.30.


3.2.2 2-Se-Uridine containing RNA synthesis and purification

Solid-phase RNA synthesis was performed on an ABI3400 DNA/RNA synthesizer with 1.0 µmol scale. 2-Se-U phosphoramidite was prepared as 0.1 M concentration solution in anhydrous acetonitrile prior of use. Coupling was carried out with 0.3 M 5-Benzylmercaptotetrazole (5-BMT) in acetonitrile as activator. The coupling time for both native and modified phosphoramidites are set to 600 seconds. Reagents for RNA automate synthesis were purchased from Glen Research: Deblocking Solution: 3% CCl3COOH in CH2Cl2; Capping Mix A: THF/Pyridine/Pac2, OCapping Mix B: 10 % 1-Methylimidazole in THF/pyridine; Oxidation Solution: 0.02 M I2/THF/Pyridine/H2O. All the DNAs were prepared in DMTr-on mode. After synthesis, 2-Se-U RNA was cleaved from the CPG beads and fully deprotected by treating with 0.05 M potassium carbonate solution in methanol for 10 hours at room temperature. After removal of CPG-Beads by centrifugation, methanol was evaporated on a speed-vac concentrator. The 2’-TBDMS deprotection was performed in 0.5 ml of 1 M TBAF (Tetrabutylammonium fluoride) for 14 hours at room temperature. After the 2’ deprotection, the reaction was quenched by adding 0.5 ml of 1 M Tris-HCl pH 7.5 buffer. The resulting mixture was desalted by G-25 Sephadex column, followed by RP-HPLC purification.
The RP-HPLC purification was performed on a Welchrom XB-C18 column (21.1 x 250 mm) with buffer A (20 mM TEAAc, pH 7.1) and buffer B (20 mM TEAAc, pH 7.1, in 50 % acetonitrile). The pump was set with a flow rate of 6 ml/min and a linear gradient of 0 % buffer B to 100 % buffer B over 20min and stay at 100% buffer B for 10min. The monitor was set to 260 nm wavelength, DMTr-on product was collected at around 25 min, frozen, and lyophilized. The final detritylation and desalting step was performed on Glen-Pak RNA column.

The purified 2-Se-U RNAs were re-dissolved in water and analyzed by HPLC, and MALDI-MS to confirm the quality. (MALDI-MS analysis of 2-Se-U modified RNAs are shown in Table 3.1)

<table>
<thead>
<tr>
<th>DNA sequences</th>
<th>Calc. [M+H]^+</th>
<th>Observed</th>
</tr>
</thead>
<tbody>
<tr>
<td>5’-rGUAAUAU-SeAC-3’</td>
<td>2558.7</td>
<td>2558.5</td>
</tr>
<tr>
<td>5’-rAUCACCUS-SeCUCUA-3’</td>
<td>3740.3</td>
<td>3740.2</td>
</tr>
<tr>
<td>5’-rAAUGCU-SeGCACUG-3’</td>
<td>3859.4</td>
<td>3859.3</td>
</tr>
</tbody>
</table>

3.2.3 Thermal denaturation study of 2-Se-U RNA

The UV-melting temperature studies were performed on a Cary 300 UV–Vis Spectrophotometer with a six-sample cell changer and a temperature controller. The RNA samples (2 mM RNA duplexes) were dissolved in buffer of 150 mM NaCl, 2 mM MgCl₂ and 10 mM Na₂HPO₄–NaH₂PO₄ (pH 6.8). The samples were prepared by heating to 80°C for 2 min and cooling down to room temperature slowly. Both denaturing and annealing curves were acquired at 260 nm with 1cm path length at heating or cooling rates of 0.5 °C/min with data interval of 0.5 °C.

In our thermal denaturation study, the melting temperature (Tₘ) of RNA duplexes were measured with native U, 2-S-U or 2-Se-U paired against all four possible nucleobases on a
complementary strand. (Shown in Table 3.2) The melting temperatures of native U paired with A/G/C/U in duplex serve as our control. The $T_m$ of the native U/A pair is only 0.3 °C higher than the U/G wobble pair’s $T_m$, which indicates that the superior thermo stability of U/G wobble pair allows it to possibly substitute Watson Crick U/A base pair in RNA duplex. This is consistent with literature reported observation. In contrast, the $T_m$ of native U/U (-14 °C) and native U/C (-12.2 °C) pairs are much lower than native Watson Crick U/A pair’s $T_m$. For the melting data of 2-S-U modified RNA duplexes, 2-S-U/A containing RNA duplex presents a 2.7 °C higher melting temperature than U/A containing duplex, while 2-S-U/G containing duplex has a 2.0 °C lower melting temperature than native U/G Wobble base pair containing duplex. The 2-Se-U/A base pair increases the stability of U/A base pair by 3.0 °C. More importantly, the 2-Se-U modification further destabilized the Wobble pair; the melting temperature of 2-Se-U/G containing duplex is 4 °C lower than that of native U/G containing duplex. Our thermal denaturation data suggests that the 2-Se-U modification can stabilize the U/A base pair (+3.0 °C), meanwhile destabilize the U/G mismatch (-4.0 °C). The contribution from both effects increases the U/A – U/G melting temperature difference by 7 °C.
Table 3.2 Thermal denaturation study of RNA mis-pairing with Native, 2-S-U RNA, and 2-Se-U RNA

<table>
<thead>
<tr>
<th>Entry</th>
<th>Sequences</th>
<th>Base Pairs</th>
<th>Tm (°C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5’-AUUCAUCCCUUA-3’</td>
<td>U/A</td>
<td>62.8</td>
</tr>
<tr>
<td>2</td>
<td>5’-AUUCCCUUA-3’</td>
<td>U/A</td>
<td>62.8</td>
</tr>
<tr>
<td>3</td>
<td>5’-AUUCCCUUA-3’</td>
<td>U/G</td>
<td>62.5</td>
</tr>
<tr>
<td>4</td>
<td>5’-AUUCCCUUA-3’</td>
<td>U/C</td>
<td>50.6</td>
</tr>
<tr>
<td>5</td>
<td>5’-AUUCCCUUA-3’</td>
<td>U/U</td>
<td>48.8</td>
</tr>
<tr>
<td>6</td>
<td>5’-AUUCCCUUA-3’</td>
<td>U/A</td>
<td>5.3</td>
</tr>
<tr>
<td>7</td>
<td>5’-AUUCCCUUA-3’</td>
<td>U/G</td>
<td>5.3</td>
</tr>
<tr>
<td>8</td>
<td>5’-AUUCCCUUA-3’</td>
<td>U/C</td>
<td>5.3</td>
</tr>
<tr>
<td>9</td>
<td>5’-AUUCCCUUA-3’</td>
<td>U/U</td>
<td>5.3</td>
</tr>
<tr>
<td>10</td>
<td>5’-AUUCCCUUA-3’</td>
<td>U/A</td>
<td>5.3</td>
</tr>
<tr>
<td>11</td>
<td>5’-AUUCCCUUA-3’</td>
<td>U/G</td>
<td>5.3</td>
</tr>
<tr>
<td>12</td>
<td>5’-AUUCCCUUA-3’</td>
<td>U/C</td>
<td>5.3</td>
</tr>
<tr>
<td>13</td>
<td>5’-AUUCCCUUA-3’</td>
<td>U/U</td>
<td>5.3</td>
</tr>
</tbody>
</table>

Conditions: 2 μM RNA duplexes, 150 mM NaCl, 2 mM MgCl₂ and 10 mM Na₂HPO₄–NaH₂PO₄ (pH 6.8)

This UV denaturing experiment data reveals that the U/G Wobble base pair can be greatly discriminated against U/A Watson Crick base pair by 2-Se-U modification. The strong discrimination against U/G pair may mainly attribute to the disruption of the hydrogen bond formation between U-2 oxygen and G-1 N-H.

3.2.4 Crystallization and diffraction data collection

The purified self-complementary RNA 5’-GUAUA₅ᵗᵉᵗₜ₅ᵗᵉᵗₜ₅ᵗᵉᵗ₄₋UA₄₋₅’ 1 mM was annealed by heating to 80 °C for 2 min, and slowly cooling down to room temperature. Crystallization screening was performed using Nucleic Acid Mini screen kit from Hampton Research with 1 uL of RNA plus 1 uL of crystallization buffer per drop equilibrate against 500 uL of 35% MPD, v/v.
Crystallizations were set up at 25 °C using hanging drop vapor diffusion method. Within two weeks, crystals appeared in 12 out of 24 screening conditions. The crystal was mounted, shock-frozen in liquid nitrogen with perfluoropolyether as cryo protectant.

The diffraction data of 2-Se-U RNA crystal was collected at beam line X12C of the NSLS (National Synchrotron Light Source) at the Brookhaven National Laboratory. A number of crystals were screened to find the crystal with best diffraction quality. The crystal offers highest resolution was grown in drops contain 5% MPD, 20 mM sodium cacodylate pH 7.0, 6 mM spermine tetra-HCl, 40 mM potassium chloride, and 0.25 mM RNA duplex. The diffraction data was collection at 0.9795 Å wavelength for 10 or 15 seconds per frame with 1° oscillation angle. The unit cell of the selected crystal was determined as a) 47.1 Å, b) 47.1 Å, and c) 424.65 Å with rhombohedral space group R32. The data was collected to a maximum resolution of 2.28 Å, integrated and scaled with the programs HKL2000 and DENZO/SCALEPACK. The statistics of the collected data set are shown in Table 3.3.
3.2.5 Structure determination and refinement

The structure of 2-Se-U modified RNA (5'-GUAAU\textsuperscript{Se}UAC-3')\textsubscript{2} was solved by molecular replacement method using Phaser\textsuperscript{68} within CCP4\textsuperscript{66}. The native RNA (5'-GUAAUAC-3')\textsubscript{2} structure (PDB ID: 246D) was used as the searching model. The resulting model was refined using REFMAC5.5.0109\textsuperscript{65} within CCP4\textsuperscript{66} and Coot\textsuperscript{64}. The topologies and parameters for 2-Se-U derivative (US) were constructed and applied. After several cycles of refinement, water molecules were added either automatically or manually using Coot\textsuperscript{64}. All refinements were performed by setting aside 5% of the total reflections for R-free calculation\textsuperscript{67}. The final R-work and R-free are 21.4 % and 26.9 %, respectively. Data-collection, and structure-refinement statistics are given in Table 3.3. Structure factor and final coordinate have been deposited in the Protein Data Bank (PDB ID: 3S49).
3.2.6 Crystal structure of 2-Se-U modified RNA

As shown in Figure 3.4A, there are seven single strands RNA in the ASU (asymmetric unit); 6 of them form three duplexes. The duplexes stack on top of each other, and form a pseudo duplex structure in ASU. In the structure (shown in Figure 3.4C), 2-Se-U modification forms 2-Se-U/A base pair just like Watson Crick U/A base pair. Based on the over structure, 2-Se-U modification was accommodated very well into the minor groove of RNA duplex. And, it is quite obvious that the duplex part of 3S49 structure is quite similar to the native one (PDB ID: 246D); the superimposed 3S49 and 246D duplex structures (shown in Figure 3.4B) and the superimposed the 2-Se-U/A and U/A base pair structures (shown in Figure 3.4D) doesn’t show significant structural perturbation. Since the U-2 carbonyl oxygen is not involved in the hydrogen bonding of U/A, it is expected that 2-Se-U modification should be accommodated in the duplex structure without disrupting the U/A base pair.

We also observed slightly shortened hydrogen bonds distance between 2-Se-U and A, which may be the contribution from both steric and electronic effect of selenium atom on 2 position of uracil. This minor change on hydrogen bond distance between 2-Se-U and A may explain the increased stability of 2-Se-U/A base pairing. However, this conclusion is still questionable base on our RNA structure due to the resolution limit (2.3 Å), a high-resolution structure is necessary to prove our hypothesis.
3.2.7 High fidelity base pairing of 2-Se-U RNA in crystal: Experimental design

Our thermal denaturation data of 2-Se-U RNA suggests that the 2-seleno-uridine modification could discriminate U/G wobble pair against U/A pair in solution by increasing the melting temperature of U/A base pair while destabilizing U/G wobble pair. The 2.3 Å resolution crystal structure of 2-Se-U RNA undoubtedly shows the formation of 2-Se-U/A base pairs without disruption of local structure and overall structure as well. The slightly shortened
hydrogen bonds distance between 2-Se-U and A may indicate stronger hydrogen bond interaction. However, detailed structural-based explanation of high fidelity base pairing of 2-Se-U is still not clear. We are very much interested in investigating the high fidelity base pairing of 2-Se-U base on crystal structure.

For this study, we designed an new oligoribonucleotide sequence r[5’-GUGUAUAC-3’], which can form duplex in two distinctive ways. In one way, this RNA sequence can form duplex with 8 base pairs including two U/G wobble base pair in duplex region. In the other way, this sequence can also form a duplex with 6 base pairs with 2 nucleotides overhang on 5’ ends in the absence of U/G wobble pair. (Shown in Figure 3.5)

![Figure 3.5 The new RNA sequence for high fidelity base pairing study](image)

The underlined uridine in r[5’-GUGUAUAC-3’] could be modified by 2-Se-U derivative. Since 2-Se-U modifications could discourage the formation of U/G wobble base pair, we hypothesized that the duplex with 6 Watson Crick base pairs and 2 Wobble base pairs can be rearranged to a duplex with 6 Watson Crick base pairs only when the uridines at wobble base pair positions are replaced by 2-Se-U. (Shown in Figure 3.6)
3.2.8 Crystallization and diffraction data collection of 5’-GUGUAUAC-3’ RNA

The purified native RNA 5’-GUGUAUAC-3’ 1 mM was annealed by heating to 80 °C for 2 min, and slowly cooling down to room temperature. Crystallization screening was performed using Nucleic Acid Mini screen kit from Hampton Research with 1 uL of RNA plus 1 uL of crystallization buffer per drop equilibrate against 500 uL of 35% 2-Methyl-2,4-pentanediol, v/v. Crystallizations were set up at 25 °C using hanging drop vapor diffusion method. Crystals appeared overnight after set up the crystallization. The crystal for data collection was prepared by mixing equal volume of 1mM RNA solution with 10% 2-Methyl-2,4-pentanediol v/v, 0.040 M sodium cacodylate trihydrate pH 7.0, 0.012 M spermine tetrahydrochloride, 0.04M Lithium chloride, 0.08 M strontium chloride hexahydrate, 0.020 M magnesium chloride hexahydrate ( #22, Nucleic Acid Mini Screen) and equilibrating the droplet against 500 uL of 35% 2-Methyl-2,4-pentanediol, v/v. The crystal was mounted, shock-frozen in liquid nitrogen with 25 % 2-Methyl-2,4-pentanediol as cryo protectant.

The diffraction data of 5’-GUGUAUAC-3’ RNA crystal was collected at beam line BL8.2.2 of the ALS (Advanced Light Source) at the Lawrence Berkeley National Laboratory. A number of crystals were screened to find the crystal with best diffraction quality. The diffraction data was

Figure 3.6 Hypothesized duplex rearrangement after 2-Se-U modification
collection at 0.97895 Å wavelength at 100 K for 2 seconds per frame with 1° oscillation angle. The unit cell of the selected crystal was determined as a) 41.68 Å, b) 41.68 Å, and c) 60.01 Å with rhombohedral space group R3. The data was collected to a maximum resolution of 2.23 Å, integrated and scaled with the programs HKL2000 and DENZO/SCALEPACK. The statistics of the collected data set are shown in Table 3.5.

3.2.9 Structure determination and refinement of 5′-GUGUAUC-3′ RNA

The structure of of 5′-GUGUAUC-3′ was solved by molecular replacement method using Phaser within CCP4. The self-complementary RNA (5′-GUAUAUC-3′) structure (PDB ID: 246D) was used as the searching model. The found solution was refined using REFMAC5.0109 within CCP4 and Coot. After several cycles of refinement, the underlined A in 246D was mutated to G, and strontium cations were added to the RNA structure manually using Coot. All refinements were performed by setting aside 5% of the total reflections for R-free calculation. The final R-work and R-free are 15.2 % and 22.9 %, respectively. Data-collection, and structure-refinement statistics are given in Table 3.5. Structure factor and final coordinate have been deposited in the Protein Data Bank (PDB ID: 1JAB).
3.2.10 Crystal structure of 5’-GUGUAUAC-3’ RNA

The (2Fo-Fc) electron density of overall structure, G3/U14, and U6/G11 wobble base pairs at 2.2 Å resolution with σ = 1.0. (PDB ID: 1JAB)

The native 5’-GUGUAUAC-3’ was determined with 2.2 Å resolution. As expected, the crystal structure clearly showed formation of 8 base pair A-form RNA duplex with 2 U/G wobble pairs. (Shown in Figure 3.7) Like other U/G wobble base pair structures, in 1JAB structure, the steep electrostatic gradient from C4 carbonyl oxygen also stabilized the strontium ions that bind to the major groove of G·U pairs. The native U/G wobble pairs are quite stable, and they can be accommodated in a duplex structure without causing significant change in duplex overall structure. Figure 3.8 shows that in the presences of 2’ U/G wobble
pairs, the 5'-'GUGUAUAC-3' (1JAB) structure is very much similar to the 5'-'GUAAUAC-3' (246D) structure.

![Figure 3.8 5'-GUGUAUAC-3' and 5'-GUAAUAC-3' structures](image)

A. 1JAB structure (5'-GUGUAUAC-3')
B. 246D structure (5'-GUAAUAC-3')
C. Superimposed 1JAB and 246D structure

3.2.11 Crystallization and diffraction data collection of 5'-GUGUA$^{2Se}$UAC-3' RNA

The purified native RNA 5'-GUGUA$^{2Se}$UAC-3' 1 mM was annealed by heating to 80 °C for 2 min, and slowly cooling down to room temperature. Crystallization screening was performed using Nucleic Acid Mini screen kit from Hampton Research with 1 uL of RNA plus 1 uL of crystallization buffer per drop equilibrate against 500 uL of 35% 2-Methyl-2,4-pentanediol, v/v. Crystallizations were set up at 25 °C using hanging drop vapor diffusion method. Crystals appeared overnight after set up the crystallization. The crystal for data collection was prepared by mixing equal volume of 1mM RNA solution with 10% 2-Methyl-2,4-pentanediol v/v, 0.040 M sodium cacodylate trihydrate pH 7.0, 0.012 M spermine tetrahydrochloride, 0.04M potassium chloride, and 0.020 M magnesium chloride hexahydrate and equilibrating the droplet against
500 μL of 35% 2-Methyl-2,4-pentanediol, v/v. The crystal was mounted, shock-frozen in liquid nitrogen with 25% 2-Methyl-2,4-pentanediol as cryo protectant.

The diffraction data of 5’-GUGUA^{2Se}UAC-3’ RNA crystal was measured at beam line BL8.2.2 of the ALS (Advanced Light Source) at the Lawrence Berkeley National Laboratory. A number of crystals were screened to find the crystal with best diffraction quality. The diffraction data was collected at 0.9795 Å wavelength at 100 K for 1 seconds per frame with 1° oscillation angle. The unit cell of the selected crystal was determined as a) 46.44 Å, b) 46.44 Å, and c) 12.61 Å with hexagonal space group P3_121. The data was collected to a maximum resolution of 1.50 Å, integrated and scaled with the programs HKL2000 and DENZO/SCALEPACK. The statistics of the collected data set are shown in Table 3.4.

### 3.2.12 Structure determination and refinement of 5’-GUGUA^{2Se}UAC-3’ RNA

The Matthew's coefficient 99 of the hexagonal 5’-GUGUA^{2Se}UAC-3’ crystal indicated that the ASU most likely contained 3 copies of 8 mer RNA duplexes. And duplexes from the 246D were chosen to serve as the search model. The molecular replacement was performed with Phaser 68 within CCP4 66 package. The correct solution contains 3 copies of RNA duplexes. Interestingly, the initial structure solution indicated a three RNA duplex bundle in ASU. And each RNA duplex is formed with 6 base pairs and a two nucleotides overhang on each end. The molecular model was updated and was refined using REFMAC5.5.0109 65 within CCP4 66 and Coot 64. After several cycles of refinement, the topologies and parameters for 2-Se-U derivative (US) were constructed and applied, water molecules were added either automatically or manually using Coot 64. All refinements were performed by setting aside 5% of the total reflections for R-free calculation 67. The final R-work and R-free are 20.8 % and 23.8 %,
respectively. Data-collection, and structure-refinement statistics are given in Table 3.5. Structure factor and final coordinate have been deposited in the Protein Data Bank (PDB ID: 1JAH).

3.2.13 Crystal structure of 5’-GUGUA\textsubscript{2Se}UAC-3’ RNA

![Figure 3.9](image)

**Figure 3.9 Overall structure of 5’-GUGUA\textsubscript{2Se}UAC-3’ RNA with electron density map (2Fo - Fc electron density map contoured at 1\sigma)**

A. Three duplexes bundle structure in asymmetric unit
B. Duplex view with two nucleotides overhang on each end.

The crystal structure of the 2-Se-U modified 5’-GUGUA\textsubscript{2Se}UAC-3’ RNA duplex is completely different from the corresponding native RNA structure (PDB ID: 1JAB). First of all, 3 duplexes forms a bundle in ASU instead of one duplex per ASU (Shown in Figure 3.9A). Each of the three duplexes formed by two 5’-GUGUA\textsubscript{2Se}UAC-3’ with six hydrogen bonds, left two nucleotides overhang on each end. (Shown in Figure 3.9B) This observation proved our previous hypothesis that when the uridines at wobble base pair positions are replaced by 2-Se-U, the eight base pairs duplex could rearrange to a “partial duplex” with six base pairs only.
The two nucleotides overhang 5’-GU-XXXXX-3’ forms two U/G wobble base pairs with the UG on the adjacent 3’-XXXXX-UG-5’. The UG/GU interactions between duplexes create intermolecular contacts involved in packing of crystals. A network of three duplexes bundles was brought together by the UG/GU interaction.

<table>
<thead>
<tr>
<th>Duplex</th>
<th>Base pair</th>
<th>O4…N6 (Å)</th>
<th>N3…N1 (Å)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A/B</td>
<td>U4/A15</td>
<td>2.98</td>
<td>2.79</td>
</tr>
<tr>
<td></td>
<td>A5/2-SeU14</td>
<td>2.96</td>
<td>2.95</td>
</tr>
<tr>
<td></td>
<td>2-SeU6/A13</td>
<td>2.78</td>
<td>2.99</td>
</tr>
<tr>
<td></td>
<td>A7/U12</td>
<td>2.95</td>
<td>2.83</td>
</tr>
<tr>
<td>C/D</td>
<td>U4/A15</td>
<td>2.98</td>
<td>2.84</td>
</tr>
<tr>
<td></td>
<td>A5/2-SeU14</td>
<td>2.81</td>
<td>2.89</td>
</tr>
<tr>
<td></td>
<td>2-SeU6/A13</td>
<td>2.94</td>
<td>2.96</td>
</tr>
<tr>
<td></td>
<td>A7/U12</td>
<td>3.06</td>
<td>2.85</td>
</tr>
<tr>
<td>E/F</td>
<td>U4/A15</td>
<td>2.97</td>
<td>2.83</td>
</tr>
<tr>
<td></td>
<td>A5/2-SeU14</td>
<td>2.9</td>
<td>2.96</td>
</tr>
<tr>
<td></td>
<td>2-SeU6/A13</td>
<td>2.92</td>
<td>2.94</td>
</tr>
<tr>
<td></td>
<td>A7/U12</td>
<td>3</td>
<td>2.85</td>
</tr>
</tbody>
</table>

The high resolution structure of 5’-GUGUA^{2Se}UAC-3’ (1.5 Å) offers the most detailed 2-Se-U containing RNA structure to date. In each duplex, two selenium atoms locate in the minor groove of duplex, with distance of 3.69 Å, 3.93 Å, and 3.93 Å respectively. Remarkably, a hydrogen bond distance change between U/A and 2-Se-U/A base pairs in crystal was observed. **Table 3.4** shows the hydrogen bond distance between O4 and N6, N3 and N1 in all U/A and 2-Se-U/A base pairs in 1JAH structure. The averaged hydrogen bond distance between O4 and N6 in all native U/A base pairs is 2.99 Å; the averaged distance of O4…N6 of all 2-Se-U/A base pairs is 2.88 Å. For the second hydrogen bond distance, the averaged O4…O6 distance in native U/A pairs is 2.83 Å; the average distance of N3…N1 in all 2-Se-U/A pairs is 2.95 Å. In summary, compare to the hydrogen bond on native U/A base pairs, the 2-Se-U modification shorten the
O4...N6 hydrogen bond distance by ~ 0.1 Å shorter while lengthen the N3...N1 hydrogen bond by ~ 0.1 Å. (Hydrogen bond distance data in 1JAH shown in Table 3.4)

**Figure 3.10 More insight into 5’-GUGUA^{2Se}UAC-3’ RNA structure**  
(2Fo - Fc electron density map contoured at 1σ)  
A. 5’-GUGUA^{2Se}UAC-3’ structure from the minor groove side of view  
B. High resolution structure of 2-Se-U/A base pair in 1JAH (Hydrogen bond distances shown here are averaged distance of all 2-Se-U/A base pairs in 1JAH)  
C. Superimposed structures of 2-Se-U/A and U/A base pairs
Table 3.5 Data collection and refinement statistics of 1JAB and 1JAH

<table>
<thead>
<tr>
<th></th>
<th>1JAB</th>
<th>1JAH</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Data Collection</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Cell Dimensions</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a, b, c (Å)</td>
<td>41.68, 41.68, 60.01</td>
<td>46.44, 46.44, 120.61</td>
</tr>
<tr>
<td>α, β, γ (°)</td>
<td>90.00, 90.00, 120.00</td>
<td>90.00, 90.00, 120.00</td>
</tr>
<tr>
<td>Wavelength (Å)</td>
<td>0.9789</td>
<td>0.9795</td>
</tr>
<tr>
<td>X-ray source</td>
<td>ALS 8.2.2</td>
<td>ALS 8.2.2</td>
</tr>
<tr>
<td>Resolution range (Å)</td>
<td>30.93-2.23 (2.31-2.23)</td>
<td>40.22-1.50 (1.55-1.50)</td>
</tr>
<tr>
<td>No. of unique reflections</td>
<td>1723 (162)</td>
<td>24253 (1916)</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>91.02 (90.00)</td>
<td>97.3 (78.5)</td>
</tr>
<tr>
<td>Rmerge (%)</td>
<td>6.6 (33.9)</td>
<td>8.0 (35.1)</td>
</tr>
<tr>
<td>Mean I/σ(I)</td>
<td>10.5 (1.98)</td>
<td>20.39 (2.4)</td>
</tr>
<tr>
<td>Redundancy</td>
<td>2.2 (2.0)</td>
<td>7.3 (3.2)</td>
</tr>
</tbody>
</table>

| **Refinement** |               |               |
| Resolution (Å)  | 30.93-2.23    | 40.22-1.50    |
| Rwork (%)       | 15.20 (34.98) | 20.80 (29.12) |
| Rfree (%)       | 22.90 (73.39) | 23.80 (30.06) |
| Number of non-hydrogen atoms | 339  | 1147 |
| Nucleotides     | 16            | 48            |
| RMS ( bond )    | 0.011         | 0.037         |
| RMS ( angles )  | 2.19          | 2.58          |
| Average B-factor | 33           | 22.6          |
| macromolecules  | 32.4          | 21.6          |
| solvent        | 34.6          | 29.1          |

### 3.3 Conclusion

The Watson-Crick base pairing system is essential for genetic information storage, replication, transcription and translation. The U/G wobble pairs are ubiquitous in RNA, especially in non-coding RNA. On the basis of the four-letter genetic code of RNA (A, C, G and U), we have developed 2-Se-uridine by replacing the 2-exo oxygen on uridine with selenium to improve the accuracy of RNA base pairing.¹⁰⁰

In summary, we have first demonstrated the synthesis of 2-SeU-phosphoramidite and 2-Se-U modified RNA. Our biophysical and crystallographic studies on the 2-Se-U-RNAs indicate...
that the 2-Se-U modification could facilitate high fidelity base pairing in solution and crystal as well. The 2-Se-U modifications can significantly discriminate U/G wobble base pair without significant impact on U/A pair. Moreover, by utilizing the anomalous dispersion property of selenium, the 2-Se-modification can serve as a useful tool in X-ray crystal structure studies of RNAs and their protein complexes. The high specificity base-pairing property of 2-Se-U RNA enables better preservation of genetic information at RNA level, which could open new research opportunities for further investigating fidelity in transcription, reverse transcription, and translation. Furthermore, our discovery may provide a unique chemical strategy to enhance RNA base pair recognition, minimize the off-target effects in sequence-based diagnostics and therapeutics, and design novel nucleic acid nanostructures.
REFERENCES


d[ACGTACG(5-BrU)]2 at 1.46 A and 100 K. *Acta crystallographica. Section D, Biological crystallography* 1999, 55 (Pt 4), 729-35.


61. Nielsen, J.; Taagaard, M.; Marugg, J. E.; van Boom, J. H.; Dahl, O., Application of 2-cyanoethyl N,N,N'-tetraisopropylphosphorodiamidite for in situ preparation of


## APPENDICES

### Nucleic acid mini screen kit (NAM) and formulation

<table>
<thead>
<tr>
<th>#</th>
<th>Precipitant 1</th>
<th>Buffer</th>
<th>pH</th>
<th>Polyamine</th>
<th>Additive</th>
<th>Concentration</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>(+/-) 2-Methyl-2,4-pentanediol</td>
<td>40mM Sodium cacodylate</td>
<td>5.5</td>
<td>20mM Hexammine cobalt(III) chloride</td>
<td>Magnesium chloride</td>
<td>20mM Hexahydrate</td>
</tr>
<tr>
<td>2</td>
<td>(+/-) 2-Methyl-2,4-pentanediol</td>
<td>40mM Sodium cacodylate</td>
<td>5.5</td>
<td>20mM Hexammine cobalt(III) chloride</td>
<td>Magnesium chloride</td>
<td>80mM Sodium chloride</td>
</tr>
<tr>
<td>3</td>
<td>(+/-) 2-Methyl-2,4-pentanediol</td>
<td>40mM Sodium cacodylate</td>
<td>5.5</td>
<td>20mM Hexammine cobalt(III) chloride</td>
<td>Magnesium chloride</td>
<td>80mM Potassium chloride</td>
</tr>
<tr>
<td>4</td>
<td>(+/-) 2-Methyl-2,4-pentanediol</td>
<td>40mM Sodium cacodylate</td>
<td>5.5</td>
<td>20mM Hexammine cobalt(III) chloride</td>
<td>Magnesium chloride</td>
<td>80mM Potassium chloride</td>
</tr>
<tr>
<td>5</td>
<td>(+/-) 2-Methyl-2,4-pentanediol</td>
<td>40mM Sodium cacodylate</td>
<td>6.0</td>
<td>12mM Spermine tetrahydrochloride</td>
<td>Magnesium chloride</td>
<td>80mM Potassium chloride</td>
</tr>
<tr>
<td>6</td>
<td>(+/-) 2-Methyl-2,4-pentanediol</td>
<td>40mM Sodium cacodylate</td>
<td>6.0</td>
<td>12mM Spermine tetrahydrochloride</td>
<td>Magnesium chloride</td>
<td>80mM Potassium chloride</td>
</tr>
<tr>
<td>7</td>
<td>(+/-) 2-Methyl-2,4-pentanediol</td>
<td>40mM Sodium cacodylate</td>
<td>6.0</td>
<td>12mM Spermine tetrahydrochloride</td>
<td>Magnesium chloride</td>
<td>80mM Potassium chloride</td>
</tr>
<tr>
<td>8</td>
<td>(+/-) 2-Methyl-2,4-pentanediol</td>
<td>40mM Sodium cacodylate</td>
<td>6.0</td>
<td>12mM Spermine tetrahydrochloride</td>
<td>Magnesium chloride</td>
<td>80mM Potassium chloride</td>
</tr>
<tr>
<td>9</td>
<td>(+/-) 2-Methyl-2,4-pentanediol</td>
<td>40mM Sodium cacodylate</td>
<td>6.0</td>
<td>12mM Spermine tetrahydrochloride</td>
<td>Magnesium chloride</td>
<td>80mM Potassium chloride</td>
</tr>
<tr>
<td>10</td>
<td>(+/-) 2-Methyl-2,4-pentanediol</td>
<td>40mM Sodium cacodylate</td>
<td>6.0</td>
<td>12mM Spermine tetrahydrochloride</td>
<td>Magnesium chloride</td>
<td>80mM Potassium chloride</td>
</tr>
<tr>
<td>11</td>
<td>(+/-) 2-Methyl-2,4-pentanediol</td>
<td>40mM Sodium cacodylate</td>
<td>6.0</td>
<td>12mM Spermine tetrahydrochloride</td>
<td>Magnesium chloride</td>
<td>80mM Potassium chloride</td>
</tr>
<tr>
<td>12</td>
<td>(+/-) 2-Methyl-2,4-pentanediol</td>
<td>40mM Sodium cacodylate</td>
<td>6.0</td>
<td>12mM Spermine tetrahydrochloride</td>
<td>Magnesium chloride</td>
<td>80mM Potassium chloride</td>
</tr>
<tr>
<td>13</td>
<td>(+/-) 2-Methyl-2,4-pentanediol</td>
<td>40mM Sodium cacodylate</td>
<td>6.0</td>
<td>12mM Spermine tetrahydrochloride</td>
<td>Magnesium chloride</td>
<td>80mM Potassium chloride</td>
</tr>
<tr>
<td>14</td>
<td>(+/-) 2-Methyl-2,4-pentanediol</td>
<td>40mM Sodium cacodylate</td>
<td>7.0</td>
<td>12mM Spermine tetrahydrochloride</td>
<td>Magnesium chloride</td>
<td>80mM Potassium chloride</td>
</tr>
<tr>
<td>15</td>
<td>(+/-) 2-Methyl-2,4-pentanediol</td>
<td>40mM Sodium cacodylate</td>
<td>7.0</td>
<td>12mM Spermine tetrahydrochloride</td>
<td>Magnesium chloride</td>
<td>80mM Potassium chloride</td>
</tr>
<tr>
<td>16</td>
<td>(+/-) 2-Methyl-2,4-pentanediol</td>
<td>40mM Sodium cacodylate</td>
<td>7.0</td>
<td>12mM Spermine tetrahydrochloride</td>
<td>Magnesium chloride</td>
<td>80mM Potassium chloride</td>
</tr>
<tr>
<td>17</td>
<td>(+/-) 2-Methyl-2,4-pentanediol</td>
<td>40mM Sodium cacodylate</td>
<td>7.0</td>
<td>12mM Spermine tetrahydrochloride</td>
<td>Magnesium chloride</td>
<td>80mM Potassium chloride</td>
</tr>
<tr>
<td>18</td>
<td>(+/-) 2-Methyl-2,4-pentanediol</td>
<td>40mM Sodium cacodylate</td>
<td>7.0</td>
<td>12mM Spermine tetrahydrochloride</td>
<td>Magnesium chloride</td>
<td>80mM Potassium chloride</td>
</tr>
<tr>
<td>19</td>
<td>(+/-) 2-Methyl-2,4-pentanediol</td>
<td>40mM Sodium cacodylate</td>
<td>7.0</td>
<td>12mM Spermine tetrahydrochloride</td>
<td>Magnesium chloride</td>
<td>80mM Potassium chloride</td>
</tr>
<tr>
<td>20</td>
<td>(+/-) 2-Methyl-2,4-pentanediol</td>
<td>40mM Sodium cacodylate</td>
<td>7.0</td>
<td>12mM Spermine tetrahydrochloride</td>
<td>Magnesium chloride</td>
<td>80mM Potassium chloride</td>
</tr>
<tr>
<td>21</td>
<td>(+/-) 2-Methyl-2,4-pentanediol</td>
<td>40mM Sodium cacodylate</td>
<td>7.0</td>
<td>12mM Spermine tetrahydrochloride</td>
<td>Magnesium chloride</td>
<td>80mM Potassium chloride</td>
</tr>
<tr>
<td>22</td>
<td>(+/-) 2-Methyl-2,4-pentanediol</td>
<td>40mM Sodium cacodylate</td>
<td>7.0</td>
<td>12mM Spermine tetrahydrochloride</td>
<td>Magnesium chloride</td>
<td>80mM Potassium chloride</td>
</tr>
<tr>
<td>23</td>
<td>(+/-) 2-Methyl-2,4-pentanediol</td>
<td>40mM Sodium cacodylate</td>
<td>7.0</td>
<td>12mM Spermine tetrahydrochloride</td>
<td>Magnesium chloride</td>
<td>80mM Potassium chloride</td>
</tr>
<tr>
<td>24</td>
<td>(+/-) 2-Methyl-2,4-pentanediol</td>
<td>40mM Sodium cacodylate</td>
<td>7.0</td>
<td>12mM Spermine tetrahydrochloride</td>
<td>Magnesium chloride</td>
<td>80mM Potassium chloride</td>
</tr>
</tbody>
</table>
NMR spectrums of key compounds
$^1$H NMR (400 MHz, CDCl$_3$+D$_2$O)
$^1$H NMR (400 MHz, CDCl₃)
$^1$H NMR (400 MHz, CDCl$_3$)
$^{1}$H NMR (101 MHz, CDCl$_3$)
$^1$H NMR (400 MHz, CDCl$_3$)
$^1$H NMR (400 MHz, DMSO)
$^1$HNMR (400 MHz, MeOD)
$^{13}$C NMR (101 MHz, MeOD)

![Chemical Structure](image)
$^1$H NMR (400 MHz, CD$_3$Cl$_2$)
$^1$H NMR (400 MHz, CDCl$_3$)
$^1$H NMR (400 MHz, MeOD)
$^1$H NMR (400 MHz, CDCl$_3$)
$^{13}$C NMR (101 MHz, CDCl$_3$)
$^1$H NMR (400 MHz, CDCl$_3$)
$^1$HNMR (101 MHz, MeOD)

![Chemical Structure]

- 2.15
- 3.67
- 4.08
- 4.53
- 4.83
- 5.62
- 6.17
- 6.72
- 7.20
- 7.69

**Chemical Shifts:**
- 2.15 ppm
- 3.67 ppm
- 4.08 ppm
- 4.53 ppm
- 4.83 ppm
- 5.62 ppm
- 6.17 ppm
- 6.72 ppm
- 7.20 ppm
- 7.69 ppm

**Field Strength:**
- 0 ppm
- 10 ppm
- 20 ppm
- 30 ppm
- 40 ppm
- 50 ppm
- 60 ppm
- 70 ppm
- 80 ppm
- 90 ppm
- 100 ppm
- 110 ppm
- 120 ppm
- 130 ppm
- 140 ppm
- 150 ppm
- 160 ppm

**Horizontal Axis:** Field (ppm)

**Vertical Axis:** Intensity