Chemistry Dissertations   Department of Chemistry

12-18-2014

Atom-Specific Modification of Uracil Bases with Selenium for RNA Structure and Function Studies

Huiyan Sun

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

ATOM-SPECIFIC MODIFICATION OF URACIL BASES WITH SELENIUM FOR RNA STRUCTURE AND FUNCTION STUDIES

by

HUIYAN SUN

Under the Direction of Zhen Huang

ABSTRACT

The atom-specific modification has been extensively applied in RNA function and structure investigations, catalysis analysis, mechanism studies, as well as therapeutics discoveries. Selenium-modified uridine (^{Se}U-RNA) is one of the naturally occurring modifications that was discovered in bacterial tRNAs (^{Se}U-RNA) at the wobble position of the anticodon loop. Its exact role in the RNA-RNA interaction, especially during the mRNA decoding is not completely understood but it was proposed that such Se derivatization on tRNAs probably improves the accuracy and efficiency of base-pairing. The wobble base pairs, where U in RNA (or T in DNA) pairs with G instead of A, might compromise the high specificity of the base pairing. The U/G wobble pairing is ubiquitous in RNA, especially in non-coding RNA. To

assist the research exploration, we have hypothesized to discriminate against U/G wobble pair by tailoring the steric and electronic effects at the 2-exo position of uridine base and replacing 2-exo oxygen with a selenium atom. This oxygen replacement with selenium offers a unique chemical strategy to enhance the base pairing specificity at the atomic level. Here, we report the first synthesis of the 2-Se-U-RNAs through synthetic incorporation of 2-Se-uridine (^{Se}U) phosphoramidite as well as enzymatic incorporation of 2-Se-uridine triphosphate. Our biophysical and structural studies of the ^{Se}U-RNAs indicate that this single atom replacement can indeed create a novel U/A base pair with higher specificity than the natural one. We reveal that the ^{Se}U/A pair maintains a structure virtually identical to the native U/A base pair, while discriminating against U/G wobble pair. Moreover, we have demonstrated that the synthesized ^{Se}UTPs (2-Se-UTP and 4-Se-UTP) are stable and recognizable by T7 RNA polymerase. Furthermore, the transcribed ^{Se}U-hammerhead ribozyme has the similar activity as the corresponding native, which suggests usefulness of ^{Se}U-RNAs in function and structure studies of noncoding RNAs, including the Se-tRNAs.

INDEX WORDS: Se-modification, Nucleic acid, Base-pairing, RNA structure, Triphosphate, Hammerhead ribozyme

ATOM-SPECIFIC MODIFICATION OF URACIL BASES WITH SELENIUM FOR RNA

STRUCTURE AND FUNCTION STUDIES

by

HUIYAN SUN

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

ATOM-SPECIFIC MODIFICATION OF URACIL BASES WITH SELENIUM FOR RNA

STRUCTURE AND FUNCTION STUDIES

by

HUIYAN SUN

Committee Chair: Zhen Huang

Committee: Kathryn B. Grant

Gangli Wang

Electronic Version Approved:

Office of Graduate Studies

College of Arts and Sciences

Georgia State University

December 2014

DEDICATION

For my loving parents, who support me unconditionally.

For my handsome husband and beautiful son, who made me a better person.

ACKNOWLEDGEMENTS

Foremost, I would like to express my sincere gratitude to advisor Prof. Zhen Huang for his continues encouragement, support, guidance, and patience in my entire Ph.D study. Without his immense wisdom and knowledge, his devotion to my education, and his encouragement during difficult times of my research, I would not have been able to complete my study. In my five and half years in Prof. Zhen Huang's research laboratory, he always provided us the best study atmosphere and during my entire study, I never heard a negative word from him. I could not have imagined a better advisor and mentor.

I would like to thank my committee member Prof. Kathryn Grant for her guidance for developing my background in nucleic acids research. The fundamental knowledge I have learned from her nucleic acids structure and function class in the first semester of my Ph.D study has greatly facilitated my later years in nucleic acid research and studies. I would like to thank my committee member Prof. Gangli Wang, for his insightful comments and suggestions for my research and my dissertation writing. Also, I would like to thank Prof. Binghe Wang, Prof. Stuart Allison and Prof. Maged Henary for their guidance and kindness in my Ph.D qualifying exam.

I would like to thank all my lab colleagues for their support and discussion over the years. Especially, I would like to give my special thanks to Dr. Abdalla Hassan for my basic organic synthesis training in nucleic acids chemistry, to Dr. Jia Sheng, Dr. Jianhua Gan and Dr. Sibo Jiang for their knowledge and collaboration in X-ray crystallography, to Dr. Jozef Salon for his advice in nucleic acid chemistry, solid-phase synthesis and oligonucleotide purification, to Dr. Julliane Caton-Williams, Dr. Lina Lin and Dr. Manindar Kaur for their guidance in triphosphate synthesis and gel electrophoresis, and to Dr. Sarah Spencer, Dr. Rob Abdur, Dr.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SCHEMES

LIST OF APPENDICES

1 GENERAL INTRODUCTION

1.1 RNA and RNA Modification

The text of this work has been published as "Atom-specific Mutagenesis of RNAs for Structure, Function and Therapeutics Studies", RNA Nanotechnology and Therapeutics, John Wiley & Sons, Inc., 2013, 213-234. I would like to acknowledge Dr. Zhen Huang for his intellectual contribution as a co-author.

Since RNAs are involved in complex biological processes as regulators, their diversities in both function and structure have been greatly appreciated.[1] RNA possesses not only the ability to store genetic information and participate in transcription and translation but also the capacity to adopt well-defined three-dimensional structures, which can be readily adjusted to meet various functional needs.[2] Although the importance of numerous RNAs in catalysis, gene expression, protein binding, and therapeutics[3] has been acknowledged by the entire scientific society, current understanding of RNA function and structure is still limited. Thus, it is not a coincidence that RNA modifications have become the subject of very intensive and active research.[4]

Over a hundred modified nucleoside residues have been discovered in natural RNAs,[5] including simple methylation, isomerization, and single-atom modification. These modifications alter the biophysical and biochemical properties of RNA structure and function. Most of these modifications are found in transfer RNAs (tRNA) despite the fact that the precise roles of many natural modifications remain mysterious. To synthesize RNAs containing modifications on the nucleobases, sugars, and phosphate backbone, both chemical and enzymatic strategies can be pursued.[6] Modified nucleic acids have tremendous potential for functional and structural investigations as well as for drug discovery, especially when equipped with unique properties,

such as enhanced thermal stability,[7] nuclease resistance,[8] and improved bio-availability.

1.2   Atom Specific Modification

Atom-specifically modified or substituted RNAs can offer many unique and novel properties without significant perturbation of three-dimensional structures and structural features of noncoding RNAs and RNA–protein complexes.[9] Hydrogen (H), carbon (C), nitrogen (N), and oxygen (O) are the four fundamental organic elements that establish nucleobases and sugars, while phosphorus (P) exists in the nucleic acid backbone. These five essential elements constitute the frame of nucleic acids. Single-atom replacement (or atom-specific mutagenesis) of nucleic acids substitutes one nucleotide atom with another atom from the same element family (such as O, S, Se, and Te) or an equivalent atom. Atom-specific mutagenesis of RNA provides useful tools to investigate RNA folding, study RNA–RNA and RNA–protein interactions, improve biochemical and biophysical properties of RNAs, facilitate gene delivery in RNA nanotechnology and drug discovery, and explore potential RNA therapeutics.[4, 10]

Among the atom-specific modifications, sulfur and selenium are in the same family with oxygen, thereby sharing similar physical and chemical properties, such as atomic radius (O: 0.73 Å; S: 1.02 Å; Se: 1.16 Å).[11] In principle, every oxygen atom on nucleic acids can be replaced by sulfur or selenium, and in practice, almost all of the oxygen atoms on the nucleobases, sugar, and phosphate backbone have been chemically or enzymatically replaced with sulfur or selenium atoms (Figure 1.2.1). This is a great advantage of chalcogen modification in comparison with halogen modification (except for fluorine), due to their instability as good leaving groups. In general, only the C-5 of pyrimidines, C-8 of purines and C-2 of adenosine are appropriate locations for the halogen substitutions. Moreover, the sulfur and selenium modifications have been discovered in natural RNAs. Incorporation of the S and Se modifications into

oligonucleotides via synthetic and enzymatic methodologies can help uncover the roles of such naturally occurring modifications in order to utilize them in related research area and drug discovery. Furthermore, it is noteworthy to mention that the element tellurium, which belongs to the chalcogen family with oxygen, sulfur and selenium, but has a much larger size (atomic radius: 1.40 Å) and more metallic character, has been introduced into sugar and base moieties of DNA.[12] The tellurium–DNA demonstrated strong topographic and current peaks through STM (Scanning Tunneling Microscope) imaging, which opens a new approach to image nucleic acids and their complexes with proteins and small molecule ligands.
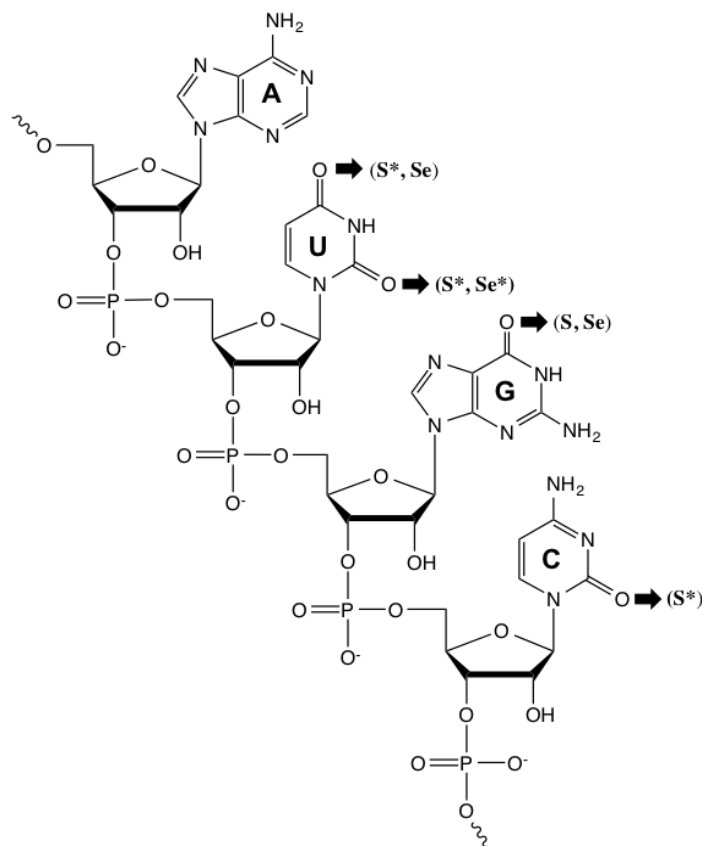
Figure 1.2.1. Sulfur and selenium modifications on RNA nucleobases.
The asterisk (*) indicates naturally occurring compounds.

1.2.1   *Sulfur Modification*

Sulfur is in the same family with oxygen and is one of the essential elements in organisms involved in biological processes. In nature, sulfur-containing nucleobases, including 2-thiouridine (s^2U), 4-thiouridine (s^4U) and 2-thiocytidine (s^2C), are observed and isolated in yeast and *Escherichia coli* tRNAs as minor components (Figure 1.2.2).[13] 2-Thionyl modified uridine was discovered back in the 1960s and often found with additional modifications at the C-5-position. These 2-thiouridine derivatives occur at wobble position 34 of *E. coli* transfer RNA (tRNA[Glu], tRNA[Lys], and tRNA[Gln]) as well as human tRNA[Lys] and are involved in codon–anticodon interaction during protein translation.[14] Biophysical studies showed that the s^2U exhibits better thermostability, compared to the native one.[7, 15] *In vitro* experiment indicated that 2-thiouridine derivatives in tRNA prefer A over G at wobble position 34.[16] An additional study carried out by Ashraf and coworkers shows that the site-specific substitution of 2-thiouridine in tRNA has higher affinity in binding to ribosome, compared to unmodified tRNA despite the modifications on C-5,[17] which has thus highlighted the functional importance of s^2U mutation.
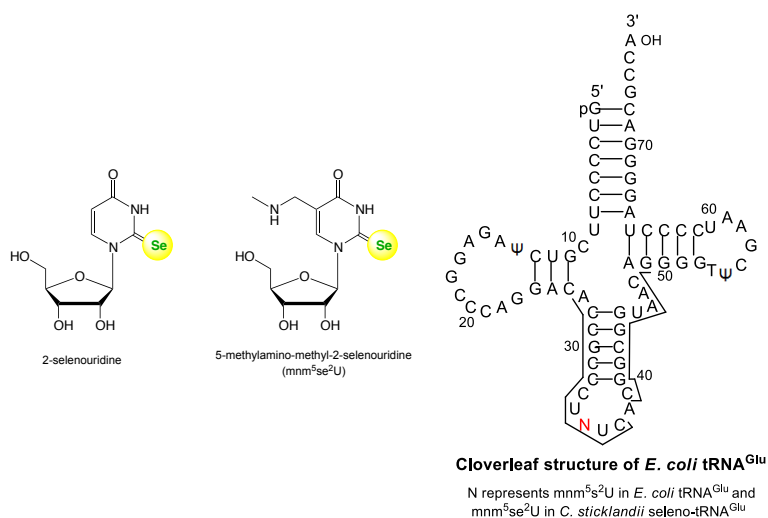
Figure 1.2.2. Cloverleaf structure of *E. coli* tRNA[Glu].
N represents mnm[5]s[2]U in *E.coli* tRNA[Glu] and mnm[5]se[2]U in C. sticklandii seleno-tRNA[Glu].

1.2.2   Selenium Modification

Element selenium belongs to chalcogen group in the periodic table together with oxygen and sulfur. Although selenium shares similar electronic and chemical properties with oxygen and sulfur, their subtle differences determine their distinct applications in biological processes and systems. Similar to sulfur, selenium-modified nucleobases are naturally occurring compounds that exist in many bacterial tRNAs, such as *Escherichia coli*, *Clostridium sticklandii*, *Methanococcus vannielii*, etc..[5] The Se modification is often found at the wobble position (position 34) of anticodon loop, which is essential for mRNA decoding.[18] The seleno nucleobases were identified as 2-selenouridine derivatives with modifications on position C-5, such as 5-aminomethyl, 5-carboxymethylaminomethyl, 5-formyl, and 5-methylaminomethyl functionalities (Figure 1.2.2), decades ago. However, the exact role of selenium at position C-2 is not yet clear. Since it was hypothesized that the 2-Se functionality discourages the U/G wobble pairing but does not affect U/A Watson-Crick base pairing (Figure 2.1.1), the 2-Se functionality is able to improve RNA base pair fidelity, thereby enhancing the accuracy of RNA transcription and translation. The 2-Se-uridine-containing RNA was chemically synthesized by Huang's lab to further explore the functionality of the seleno modification.[9a] Consistent with their hypothesis, our study showed that with the introduction of selenium at the 2-position, the Se–RNA duplex structure is virtually identical to the corresponding native form. The U/G wobble pair was greatly discouraged due to the large size of selenium atom and poor electronegativity, which severely weakened the hydrogen bonding, while the U/A base pair was not significantly affected. Thus, the increased fidelity of U/A base pairing provided new insights into codon–anticodon recognition with the seleno modification at the third codon base. Moreover, the 2-selenouridine-modified hammerhead ribozyme has catalytically activity.[19] The Se-modified thymidine at

position 4[20] and guanosine at position 6[21] in DNA oligonucleotides were also reported recently. In addition, the 4-Se-U RNA has been synthesized in the laboratory recently.[22] In natural RNA, 4-selenouridine was reported earlier in *E. coli* tRNA,[23] and the later studies suggested the 4-Se functionality as a misincorporation.[24] Furthermore, 6-selenoguanine has been applied in anticancer therapeutic studies in comparison with 6-thioguanine, and it showed promising antitumor activity against L1210 lymphomas, L5178Y lymphomas, and Sarcoma 180,[25] while no encouraging result was observed yet in the treatment of solid tumors.[26]

2   HIGHER SPECIFICITY OF RNA BASE PAIRING

2.1   Introduction

The text of this work has been published as "Novel RNA Base Pair with Higher Specificity using Single Selenium Atom", Nucleic Acids Res., 2012, 40, 5171-5179. I would like to acknowledge Dr. Jia Sheng, Dr. Sibo Jiang and Dr. Jianhua Gan for their contribution in RNA structure determination as co-authors, and I would like to acknowledge Dr. Zhen Huang and Dr. Abdalla E. A. Hassan for their intellectual contribution as co-authors.

2.1.1   U/G wobble pair

DNA and RNA are crucial genetic information carriers.[27] The base pairs of DNAs (T/A and C/G) and RNAs (U/A and C/G) need to be highly specific and accurate for the purpose of the precise genetic information storage, replication, transcription and translation. However, the wobble base pairs, where U in RNA (or T in DNA) pairs with G instead of A, may compromise the high specificity of the base pairing. In RNA, especially non-coding RNA, U/G wobble pair (Figure 2.1.1) is ubiquitous[28] and sometimes it has the similar stability as the Watson–Crick U/A pair.[29] U/G wobble pair offers unique structural and thermodynamic features.[28-29] On the one hand, the U/G pairing increases structure and function diversities of RNA.[30] But on the other hand, it may jeopardize the pairing specificity and can cause potential mutations in RNA transcription and protein translation. Codon–anti- codon mismatch or misreading is observed with an error frequency at 10^{-5} or higher, which may affect the accuracy of synthesized proteins.[31] For instance, the first position of the codon–anticodon interaction with wobble mismatch (U/G) was discovered in Escherichia coli (error frequency = 0.1%) with 100-fold higher than the normal error level.[31c] In this mis-incorporation of serine (codon: AGC),[31c] glycine

codon (GGC) in mRNA is recognized by Ser-charged tRNA (anticodon: GCU) instead of Gly-charged tRNA (anticodon: GCC). Similarly, the second position of the codon–anticodon interaction with wobble mismatch (U/G) was also observed, where Lys (codon: AAA) is mis-incorporated instead of normal incorporation of Arg (codon: AGA), with much higher error frequency (5–12%).[32] To avoid the negative impact of the wobble pairing on the level of protein synthesis, the genetic codes with degeneracy are used to deal with the consequence of the wobble pairing. Thus, wobble pairing is often observed at the third codon position through the codon degeneracy to limit errors. However, the codons forming the Watson–Crick pairs with tRNA anticodons are still preferred.[33] Study shows that the third codon position with a Watson–Crick base pair can reduce the frequency of amino acid mis-incorporation by nearly 10-fold, and it is much more accurate than that with a wobble pair for the same amino acid.[34] Nevertheless, the 3-nt genetic codes that accommodate the wobble pairing are used as the most ideal counter-measure at the level of protein synthesis in living organisms.[35] Clearly, on the basis of the chemical principle, this degeneracy strategy properly guarantees the translation accuracy at the protein level by tolerating wobble pairs and silent mutations at the RNA and DNA levels.

U/A and ^{Se}U/A base pairs   U/G and ^{Se}U/G wobble pairs
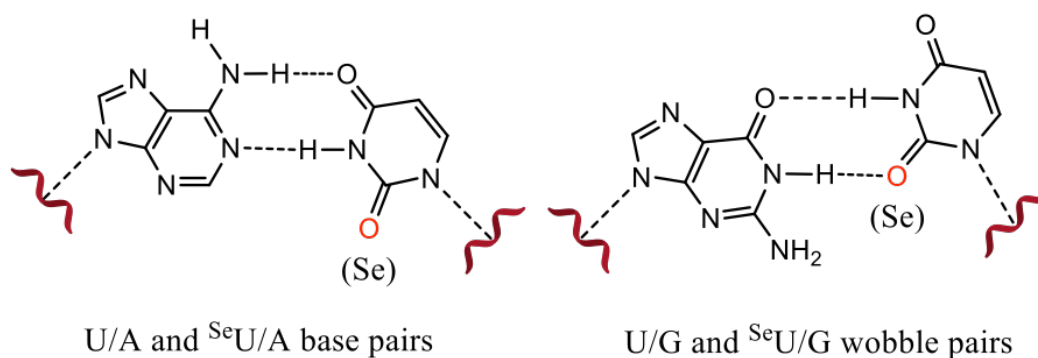
Figure 2.1.1. Native and Se-modified U/A pairs and U/G wobble pairs. Selenium substitution for the oxygen atom was labeled in red.

2.1.2 Sulfur- and selenium-modified wobble pair

Since the 2-exo-oxygen of uridine plays a significant role in U/G wobble pair, we hypothesized that tailoring the steric and electronic effects at this site may discriminate against the wobble pair, enabling the modified U/A base pair with higher specificity. Interestingly, selenium has been discovered in natural tRNAs in the 2-Se-uridine form, i.e. 5-methylaminomethyl-2-selenouridine (mnm^{5}se^{2}U), in the wobble position on the anticodon loop.[18, 24] The function of such selenium modification is not completely clear yet, though it was proposed that such Se derivatization on tRNAs probably improves the accuracy and efficiency of protein translation.[36] Similarly, the corresponding sulfur modification has been observed on natural tRNAs.[37] Sulfur was chemically introduced to the 2-position of uridine.[38] The S-modified U/G pair is slightly less stable than the native U/G pair,[29b] while the ^{S}U/A is more stable over the native U/A pair. Thus, we hypothesized that the 2-oxygen replacement with selenium (^{Se}U, Figure 2.1.1) can destabilize and discriminate against the U/G wobble pair, because the atomic size of selenium (1.16 Å) is larger than that of sulfur (1.02 Å) and oxygen (0.73 Å). Moreover, selenium has the least ability to form a hydrogen bond among O, S and Se, which weakens the hydrogen bond originally formed by the 2-oxygen of the wobble pair. Thus, it is expected that this 2-Se-replacement can largely destabilize U/G pair by generating a steric hindrance against the pair and significantly weakening the hydrogen bond. Furthermore, it is expected that the 2-Se-substitution does not significantly affect the hydrogen bonds within the U/A pair, since the 2-oxygen is not directly involved in the U/A base pairing. Therefore, we decided to incorporate selenium into the 2-position of uridine in RNA, in order to atom-specifically increase the U/A pair specificity and disrupt the U/G wobble.

2.2　General Experimental Section

2.2.1　*Synthesis of 2-Se-uridine phosphoramidite*

Though selenium was incorporated into uridine four decades ago,[39] RNA containing 2-Se-uridine (SeU) hasn't been synthesized because of the synthetic challenge. Recently our laboratory has successfully developed a novel strategy to incorporate the selenium functionality to the 2-position of thymidine in DNA.[40] This successful strategy has encouraged us to introduce the selenium functionality to the 2-position of uridine in RNA. Herein we report the first synthesis of the 2-selenouridine derivatives and RNAs. The synthesis (Scheme 2.2.1) started from the glycosidation of **1** with silylated 2-thiouracil (**3**), followed by benzoyl deprotection and trityl protection of the 5'-hyroxyl group to offer **6**.[41] After methylation of **6** to activate the 2-thio-functionality,[40] NaSeH was used to displace the 2-S-functionality and offer the 2-Se-uridine (**8**) in 85% yield. Following the protections of the 2'-hydroxyl group and the 2-Se-functionality with ICH$_2$CH$_2$CN, the Se-phosphoramidite (**11**) was synthesized by phosphitylation of **10a**.[40, 42] The SeU-phosphoramidite was finally incorporated into RNAs by solid-phase synthesis. The synthesized SeU-RNAs (**12**) were deprotected, purified, and confirmed by HPLC and MS (Figure 2.2.3,Table 2.2.1). For the purpose of comparison, the corresponding S-modified RNA was also synthesized by following the literature[41] and characterized by HPLC and MS analyses (Table 2.2.1, Figure 2.2.3).
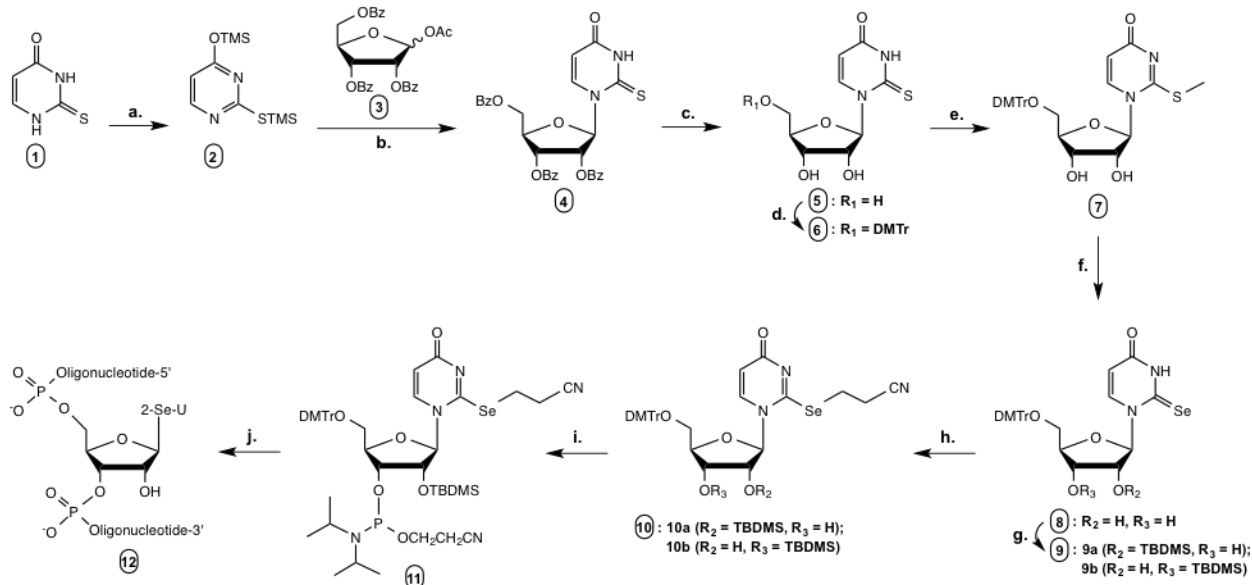
Scheme 2.2.1. Synthesis of 2-Se-Uridine containing RNA.

Reagents and conditions: a) TMS-Cl, HMDS, reflux; b) SnCl$_4$, C$_2$H$_4$Cl$_2$, -20°C; c) NaOCH$_3$, MeOH; d) DMTr-Cl, pyridine, rt. (e) CH3I, DBU, DMF; (f) Se, NaBH4, EtOH; (g) TBDMS-Cl, imidazole, DMF; (h) ICH$_2$CH$_2$CN, (i-Pr)$_2$NEt, CH$_2$Cl$_2$; (i) (i-Pr$_2$N)$_2$P(Cl)OCH$_2$CH$_2$CN, (i-Pr)$_2$NEt, CH$_2$Cl$_2$; (j) solid-phase synthesis.

2.2.1.1 Synthesis of compound 4

1-(2',3',5'-tri-O-benzoyl-beta-D-ribofuranosyl)-2-thiouridine (Compound 4).

4-[(trimethylsilyl)oxy]-2-[(trimethylsilyl)thio]-pyrimidine **2** was synthesized[7, 43] by the silylation of 2-thiouracil **1** (3.81 g, 29.76 mmol) with hexamethyldisilazane (HMDS, 100 mL) and catalytic amount of trimethylsilyl chloride (TMSCl, 0.5 mL) under reflux condition overnight until a clear yellow solution was obtained. The excess of HMDS and TMSCl was evaporated under reduced pressure. 1-O-Acetyl-2,3,5-tri-O-benzoyl-bata-D-ribofuranose **3** (10 g, 19.82 mmol) was dissolved in 1,2-dichloroethane (99 mL), then it was added to the concentrated silylated 2-thiouracil **2**. Tin (IV) chloride (6.9 mL) was subsequently added at -20 °C under nitrogen. The reaction was stirred for 5 hours and poured into a saturated aqueous sodium bicarbonate solution with stirring. After 1 hour, the suspension was extracted with

dichloromethane (6x50 mL). The organic layers were combined and dried over anhydrous magnesium sulfate, followed by filtration and evaporation under reduced pressure. The crude compound was purified by column chromatography (1% methanol in dichloromethane), offering **4** as a white foam (10.5 g, 85% yield). The compound was analyzed by ^{1}H NMR and its chemical shifts were consistent with the known compound[44]. ^{1}H NMR (400 MHz, CDCl$_3$) 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').

2.2.1.2   *Synthesis of compound 5*

2-thio-1-beta-D-ribofuranosylpyrimidine-2,4-dione (Compound 5).

Compound **4** (10.5 g, 18.35 mmol) was dissolved in methanol (92 mL) and sodium methoxide (5.95 g, 0.11 mol) was added to the solution. After stirring for 6 hours, the solution was neutralized by adding DOWEX 50WX8-400 ion-exchange resin (H[+]-form, approximately 110 meq until neutral, monitored by moisturized pH paper) washed with methanol. The mixture was filtered and methanol was evaporated. The residue was suspended with water (100 mL) and extracted with ethyl acetate (2x30 mL). Water layer was lyophilized (or evaporated under reduced pressure) to give deprotected nucleoside **5**. The crude product was monitored by TLC plate (20% methanol in dichloromethane, Rf = 0.4) to confirm the removal of the by-product, methyl benzoate. The crude product was purified by recrystallization in ethanol and gave a white powder. The compound was analyzed by ^{1}H NMR and its chemical shifts were consistent with the known compound (Ref. 1,2). ^{1}H NMR (D$_2$O) δ: 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').

2.2.1.3 Synthesis of compound 6

1-(5'-O-4,4'-dimethoxytrityl-beta-D-ribofuranosyl)-2-thiouridine (Compound 6).

2-thiouridine **5** (3 g, 11.54 mmol) and 4, 4'-dimethoxytrityl chloride (DMTr-Cl, 4.69 g, 13.85 mmol) were dried individually under high vacuum. A solution of DMTr-Cl dissolved in anhydrous pyridine (15 mL) was slowly added to **5** dissolved in anhydrous pyridine (40 mL) under nitrogen gas, at 0 $^{\circ}$C. The mixture was stirred for three hours at room temperature and methanol (5 mL) was then added to the mixture to quench the reaction. Pyridine was evaporated under reduced pressure and co-evaporated with toluene (20 mL) for 3 to 4 times. The residue was dissolved with ethyl acetate (30 mL) and washed with water twice (20 mL). The organic phase was dried over anhydrous magnesium sulfate, filtered and evaporated to dryness. Purification was carried out by flash column chromatography (2% methanol in dichloromethane) pre-equalized by 1% triethylamine in dichloromethane before sample loaded. Compound **6**, a yellow form was obtained (5 g, 80% yield; 31). ^{1}H NMR (CDCl$_3$) δ: 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$_3$) δ: 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), 128.19 (Ar), 127.33 (Ar), 113.48 (Ar), 106.75 (C-5), 94.37 (C-1'), 87.22 (C-Ar$_3$), 83.78 (C-4'), 75.80 (C-2') , 69.17 (C-3')

, 61.21 (C-5'), 55.40 (OCH$_3$); HRMS (ESI-TOF) [M-H$^+$]$^-$ = 561.1718 (calc. 561.1695), Chemical

Formula: C$_{30}$H$_{29}$N$_2$O$_7$S.

2.2.1.4   Synthesis of compound 7

1-(5'-O-4,4'-dimethoxytrityl-beta-D-ribofuranosyl)-2-methylthiouridine (Compound 7).

Dry compound **6** (5 g, 8.89 mmol) was dissolved in dry *N, N*-dimethylformamide (DMF),

followed by addition of iodomethane (5.5 mL, 89 mmol). 1,8-diazabicyclo[5.4.0]undec-7-ene (2

mL, 13.3 mmol) was then added to the reaction mixture at 0 °C. The reaction was monitored by

TLC plate (12% methanol in dichloromethane, Rf = 0.4) and completed in 4 hours. Ethyl acetate

(50 mL) was poured into the mixture and DMF was removed by washing the organic layer with

saturated sodium chloride solution. The organic phase was dried over anhydrous magnesium

sulfate and evaporated under reduced pressure. The residue was purified by flash column

chromatography (10% methanol in dichloromethane) and pure compound **7** was obtained in 95%

yield. ^{1}H NMR (CDCl$_3$) 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-3'), 4.24 (d, J = 2.3 Hz, 1H, H-2'), 3.75 (d, J = 3.1 Hz, 6H, OCH$_3$),

3.42 (m, 2H, H-5'), 3.40 – 3.30 (br, 1H, OH), 2.55 (s, 3H, SCH3); ^{13}C NMR (CDCl$_3$) 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-Ar$_3$), 84.82 (C-4'), 75.24 (C-2'), 71.63 (C-3'), 63.40 (C-5'), 55.40 (OCH$_3$), 15.39

(SCH$_3$); HRMS (ESI-TOF) [M+H$^+$]$^+$ = 577.2003 (calc. 577.2008), Chemical Formula:

C$_{31}$H$_{33}$N$_2$O$_7$S.

2.2.1.5 Synthesis of compound 8

1-(5'-O-4,4'-dimethoxytrityl-beta-D-ribofuranosyl)-2-selenouridine (Compound 8).

A solution of NaSeH was generated by addition of absolute ethanol (50 mL) to selenium (6.2 g, 78 mmol) and sodium borohydride (NaBH$_4$, 4.43 g, 0.117 mol) at 0 ^{o}C. The reaction was completed in two hours and a clear solution was formed. The ethanolic solution was added to compound **7** (4.5 g, 7.80 mmol) and the mixture was stirred for eight hours under argon. The reaction mixture was then concentrated under reduced pressure and ethyl acetate (50 mL) was added to the residue. The organic layer was washed with water several times (5x30 mL), and then dried over anhydrous magnesium sulfate. Purification was performed by flash column chromatography (4% methanol in dichloromethane) and the light yellow compound (**8**) was obtained (85% yield). ^{1}H NMR (CDCl$_3$) 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, OCH$_3$), 3.58 (dd, J = 23.6, 9.2 Hz, 2H, H-5'), 2.97 (br, 1H, OH); ^{13}C NMR (CDCl$_3$) 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-Ar$_3$), 84.41 (C-4'), 76.33 (C-2'), 69.19 (C-3'), 61.20 (C-5'), 55.48 (OCH$_3$). HRMS (ESI-TOF) [M-H$^+$]$^-$ =609.1136 (calc. 609.1140), Chemical Formula: C$_{30}$H$_{29}$N$_2$O$_7$Se; UV (MeOH): λ$_{max}$ = 311 nm (in methanol).

2.2.1.6 Synthesis of compound 9

1-(2'-O-tert-butyldimethylsilyl-5'-O-4,4'-dimethoxytrityl-beta-D-ribofuranosyl)-2-selenouridine (Compound 9a) and 1-(3'-O-tert-butyldimethylsilyl-5'-O-4,4'-dimethoxytrityl-beta-D-ribofuranosyl)-2-selenouridine (Compound 9b).

5'-DMTr-2-selenouridine **8** (0.5 g, 0.82 mmol) was dissolved in dry *N,N*-dimethylformamide, then *tert*-butyldimethylsilyl chloride (TBDMSCl, 0.15 g, 0.98 mmol) and imidazole (0.11 g, 1.64 mmol) were added into the solution under nitrogen gas. The reaction was monitored by TLC plate (15% ethyl acetate in dichloromethane, Rf = 0.8). The mixture was stirred overnight at room temperature and then directly poured into ethyl acetate (20 mL) and washed with water (2x20 mL). The organic layer was dried by anhydrous magnesium sulfate and evaporated under reduced pressure. Two compounds, **9a** and **9b**, were obtained. The two regional isomers (ratio 1:1) were purified together by flash column chromatography (10% ethyl acetate in dichloromethane) and were not further separated. Since it was both challenging and unnecessary to separate each isomer, we decided to move to the next step of synthesis without separation of these two isomers. HR-MS (ESI-TOF, **9a** and **9b**) $[M-H^+]^-$ = 723.1990 (calc. 723.2005). Chemical Formula: $C_{36}H_{43}N_2O_7SeSi$.

2.2.1.7 Synthesis of compound 10

1-(2'-O-tert-butyldimethylsilyl-5'-O-4,4'-dimethoxytrityl-beta-D-ribofuranosyl)-2-cyanoethylselanyluridine (Compound 10a) and 1-(3'-O-tert-butyldimethylsilyl-5'-O-4,4'-dimethoxytrityl-beta-D-ribofuranosyl)-2-cyanoethylselanyluridine (Compound 10b).

The mixture (0.52 g, 0.72 mmol) of **9a** and **9b** was dissolved in dried dichloromethane at 0°C. Iodopropionitrile (0.78 g, 4.31 mmol) was added to the solution, followed by addition of diisopropylethylamine (0.37 mL, 2.15 mmol). The reaction was monitored by TLC plates (30%

Ethyl acetate in dichloromethane). After 4 hrs reaction, the solvent was removed under reduced pressure and the residue was partitioned between ethyl acetate (20 mL) and water (20 mL). The organic phase was dried over anhydrous magnesium sulfate and evaporated into dryness. Two crude products were obtained: 1-(2'-*O*-tert-butyldimethylsilyl-5'-*O*-4,4'-dimethoxytrityl-beta-D-ribofuranosyl)-2-cyanoethylselanyluridine **10a** (Rf = 0.35) and 1-(3'-*O*-tert-butyldimethylsilyl-5'-*O*-4,4'-dimethoxytrityl-beta-D-ribofuranosyl)-2-cyanoethylselanyluridine **10b** (Rf = 0.30). These two compounds can be separated by flash column chromatography (15% ethyl acetate in dichloromethane). **10a** was obtained in 0.228 g (41% yield) and **10b** was obtained in 0.235 g (42% yield). **10a**: ^{1}H NMR (CDCl$_3$) δ: 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, OCH3), 3.54 – 3.34 (m, 4H,H-5', SeCH$_2$C$\underline{H}_2$CN), 3.01 (m, 2H, SeC$\underline{H}_2$CH$_2$CN), 2.91 (s, 1H, OH), 0.94 (s, 9H, SiCMe$_3$), 0.09 (d, 6H, SiMe$_2$); ^{13}C NMR (CDCl$_3$) δ: 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 (OCH3), 25.84 (SiC$\underline{M}$e$_3$), 24.06 (SeCH$_2$C$\underline{H}_2$CN), 18.89 (SeC$\underline{H}_2$CH$_2$CN), 18.14 (Si$\underline{C}$Me$_3$), -4.56 (SiCH$_3$)., -4.91 (SiCH$_3$). HRMS (ESI-TOF) [M+H$^+$]$^+$ = 778.2464 (calc. 778.2427). Chemical Formula: C$_{39}$H$_{48}$N$_3$O$_7$SeSi. **10b**: ^{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$_2$C$\underline{H}_2$CN), 3.22 (d, J = 5.5 Hz, 1H, OH), 3.08 (m, 2H, SeC$\underline{H}_2$CH$_2$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 (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 (SiCMe3), -4.59 (SiCH$_3$), -4.60 (SiCH$_3$). HRMS (ESI-TOF) [M+H$^+$]$^+$ = 778.2401 (calc. 778.2427). Chemical Formula: C$_{39}$H$_{48}$N$_3$O$_7$SeSi.

2.2.1.8   Synthesis of compound 11

1-[2'-O-tert-butyldimethylsilyl-3'-O-(2-cyanoethyl-N,N-diisopropylamino) phosphoramidite-5'-O-(4,4'-dimethoxytrityl-beta-D-ribofuranosyl)]-2-cyanoethylselanyluridine (Compound 11).

Diisopropylethylamine (15.5 mg, 0.12 mmol) and 2-cyanoethyl *N,N*-diisopropylchlorophosphoramidite (26 mg, 0.11 mmol) were added to a solution of 10a (100 mg, 0.10 mmol) in dry dichloromethane (5 mL) at room temperature under nitrogen gas. The mixture was monitored by TLC (15% ethyl acetate in dichloromethane). When the reaction was completed in 4 hrs, rapid Al$_2$O$_3$ column chromatography (dichloromethane as the eluent) was performed to remove the organic salts. The solvent was then evaporated under reduced pressure and the residue was dissolved in 0.5 mL dichloromethane and precipitated in dry hexane under vigorous stirring. The precipitate was collected by filtration, 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,), 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 (SeCH$_2$CH$_2$CN), 117.75 (OCH$_2$CH$_2$CN), 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 (OCH$_2$CH$_2$CN), 55.48 (OCH$_3$), 43.22-43.10 (NCMe$_2$), 29.90 (OCH$_2$CH$_2$CN), 26.12-25.97 (NCMe$_2$), 24.85 (SiCMe$_3$), 24.02 (SeCH$_2$CH$_2$CN), 18.94 (SeCH$_2$CH$_2$CN), 18.37 (SiCMe$_3$),-4.35 (SiCH$_3$), -4.56 (SiCH$_3$); ^{31}P NMR (CDCl$_3$) δ: 148.81, 152.30. HRMS (ESI-TOF) [M+H$^+$]$^+$ = 978.3528 (calc. 978.3505). Chemical Formula: C$_{48}$H$_{65}$N$_5$O$_8$PSeSi.

2.2.2 Solid phase synthesis of the 2-Se-functionalized RNAs
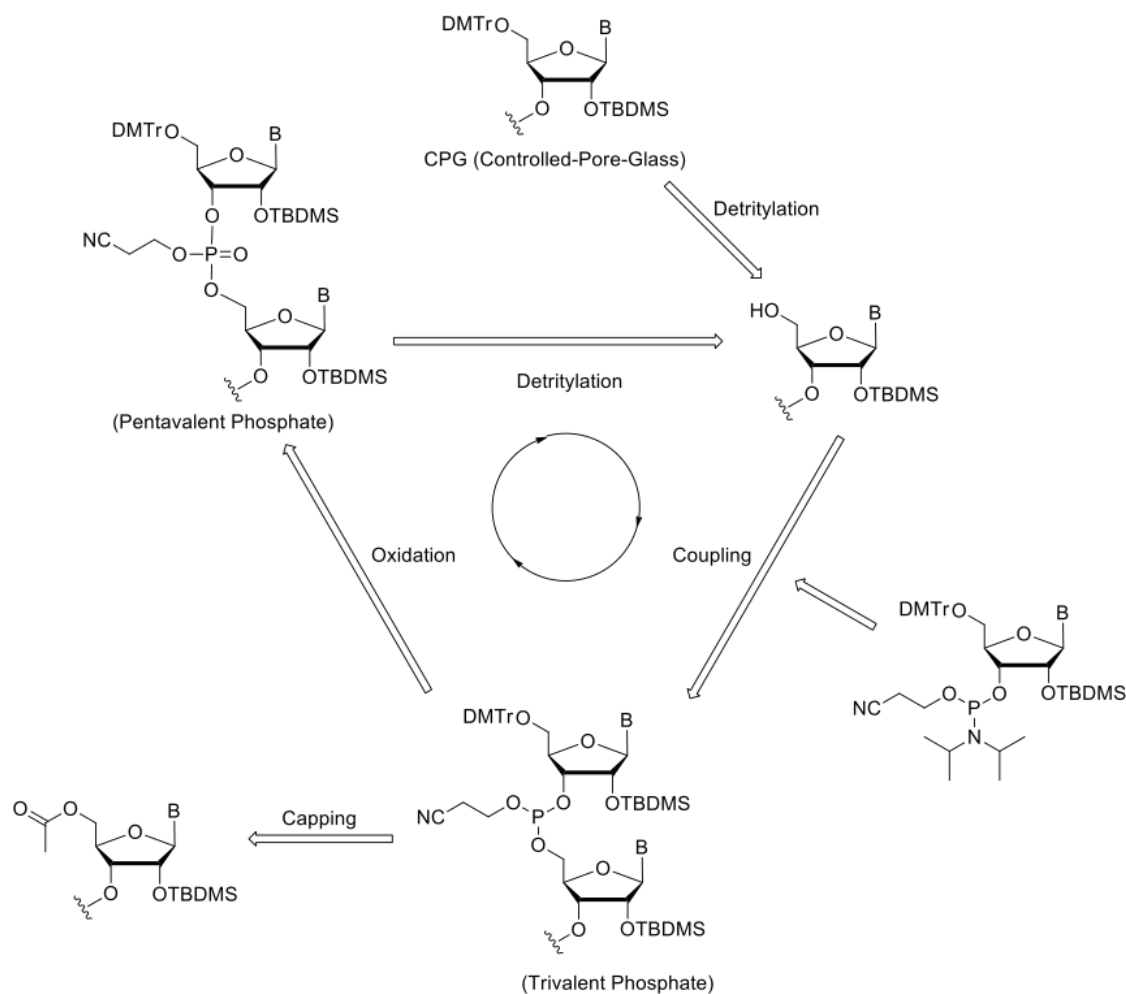
Figure 2.2.1. Oligonucleotide solid phase synthesis cycle.

ABI3400 DNA/RNA Synthesizer was used for all the RNA oligonucleotides synthesis (1.0 μmol scale). All the non-modified nucleoside phosphoramidite reagents used were ultra-mild (Glen Research). The synthetic cycle is a stepwise addition of nucleoside phosphoramidite to the 5' side of the nucleotide chain. It starts from detritylation of 5'-DMTr of the solid support bounded oligonucleotide to free the hydroxyl group. Then the nucleoside phosphoramidite is delivered to couple with the solid support bounded oligonucleotide in a solution of azole catalyst. After coupling, the unreacted 5'-OH group is blocked with a capping mixture to prevent further

elongation reactions. To stabilize the phosphite trimester linkage, the oxidation step is carried out to transform it to a pentavalent phosphate with iodine, and then the solid support bounded oligonucleotide ready for the new addition of next nucleobase (Figure 2.2.1).

RNA oligonucleotides were synthesized in DMTr-on form, cleaved from the beads and deprotected by the treatment of 0.05 M K_2CO_3 methanol solution for 10 hours at room temperature. After evaporating the solution to dryness, the 2'-TBDMS deprotection was performed in TBAF (0.5 mL, 1 M) for 14 hours at room temperature. Then the RNAs were treated with 1 M Tris-HCl buffer (0.5 mL, pH 7.5) for 5 min, followed by concentrating to 0.5 mL and desalting using G-25 Sephadex column. The 5'-DMTr deprotection was then performed using Glen-Pek RNA column, followed by desalting using Sep-Pak Vas column.

2.2.3   pH titration curve of 2-selenouridine
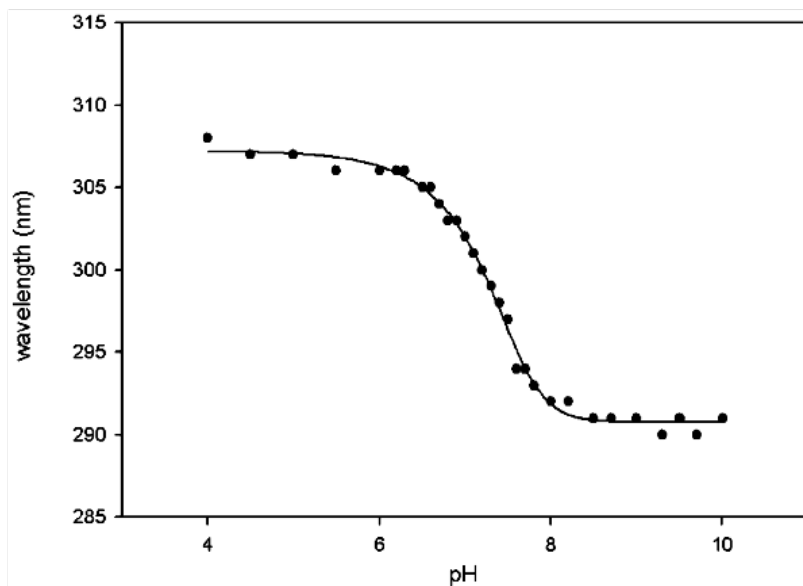
Figure 2.2.2. Plot of wavelength (nm) versus pH for 2-selenouridine nucleoside.

2-Selenouridine was prepared through detritylation of 1-(5'-O-4,4'-dimethoxytrityl-b-D-

ribofuranosyl)-2-selenour idine (8) by acid treatment. The 2-selenouridine solutions were adjusted to desired pH values in the buffer of 50 mM Na_2HPO_4 at room temperature. The UV–Vis spectra were recorded every 0.1 pH unit between pH 6–8 and every 0.2-0.5 pH unit between pH 4–6 and pH 8–10. The pH of each solution was measured before and after its UV–Vis spectrum collection and the error was within ±0.02 pH unit. The titration data was plotted and shown in Figure 2.2.2.

2.2.4   HPLC analysis and purification

Entry	Oligonucleotide	Molecular Formula	Measured (calc.) m/z
1	5'-GUAUA^{Se}UAC-3'	$C_{76}H_{94}N_{29}O_{53}P_7Se$	$[M+H^+]^+ = 2558.7$ (2558.5)
2	5'-AUCACC^{Se}UCCUUA-3'	$C_{111}H_{141}N_{38}O_{82}P_{11}Se$	$[M+H^+]^+ = 3740.3$ (3740.2)
3	5'-AAUGC^{Se}UGCACUG-3'	$C_{114}H_{142}N_{45}O_{81}P_{11}Se$	$[M+H^+]^+ = 3859.4$ (3859.3)
4	5'-AUCACC^{S}UCCUUA-3'	$C_{111}H_{141}N_{38}O_{82}P_{11}S$	$[M^+]^+ = 3692.5$ (3692.3)

Table 2.2.1. MALDI-TOF MS of 2-Se-U RNAs

The RNA oligonucleotides were analyzed and purified by reversed-phase high performance liquid chromatography (RP-HPLC), flow rate 6 mL/min [Buffer A: 20 mM triethylammonium acetate (TEAAc, pH 7.1) in water; buffer B: 20 mM TEAAc (pH 7.1) in 50% acetonitrile]. The HPLC analysis was performed with a linear gradient from buffer A to 100% buffer B in 20 min. Native RNAs were purchased from Integrated DNA Technologies. The concentrations of the native, S- and Se-modified RNAs were adjusted to 1.0 mM in water. The S- and Se-RNA samples were characterized by MALDI-TOF MS (Table 2.2.1) and HPLC (Figure 2.2.3).
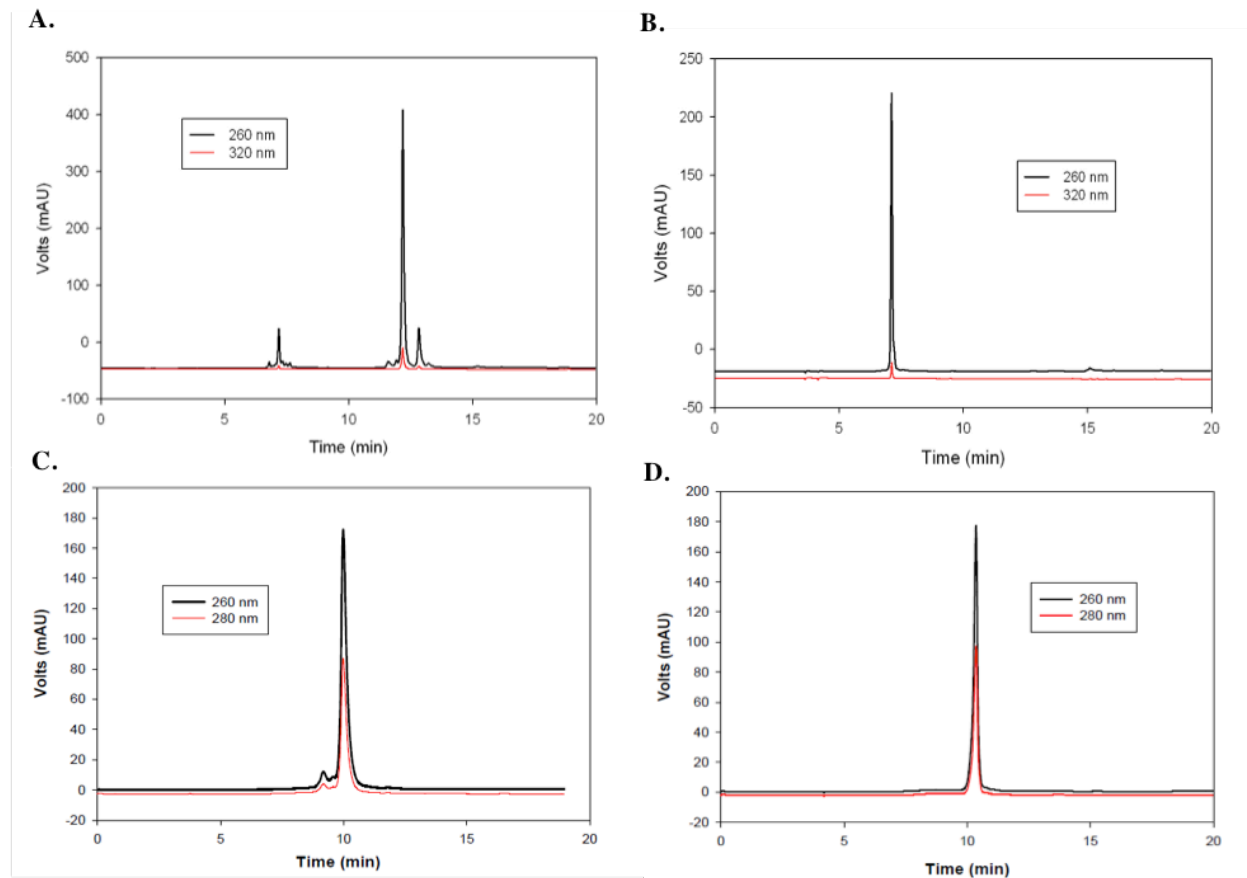
Figure 2.2.3. HPLC analysis and purification of 2-S-U and 2-Se-U Modified RNAs.
(A) The HPLC analysis profile of crude DMTr-on Se-RNA (5'-rAUCACC^{Se}UCCUUA-3') after cleavage from solid support and deprotection steps. The DMTr-on Se-RNA retention time was 12.2 min. (B) The HPLC analysis profile of pure DMTr-off Se-RNA (5'-rAUCACC^{Se}UCCUUA-3') with same gradient and buffer. The DMTr-off Se-RNA retention time was 7.1 min. (C) The HPLC analysis profile of pure DMTr-off Native-RNA (5'-rAUCACCUCCUUA-3'). The DMTr-off Native-RNA retention time was 10.0 min. (D) The HPLC analysis profile of pure DMTr-off S-RNA (5'-rAUCACCSUCCUUA-3') with same gradient and buffer. The DMTr-off S-RNA retention time was 10.4 min. Samples were eluted with a linear gradient from buffer A (20 mM triethylammonium acetate, pH 7.1) to 70% buffer B (50% acetonitrile, 20 mM triethylammonium acetate, pH 7.1) in 10 min, to 100% buffer B in 12 min and continuous 100% buffer B to 20 min.

2.2.5   Thermodenaturization of duplex RNAs

Entry	Sequences	Base Pairs	Tm (°C)
1	I: 5'-rAUCACC**U**CCUUA-3'		
2	I + 3'-rUAGUGG**A**GGAAU-5'	U/A	62.8
3	I + 3'-rUAGUGG**G**GGAAU -5'	U/G	62.5
4	I + 3'-rUAGUGG**C**GGAAU -5'	U/C	50.6
5	I + 3'-rUAGUGG**U**GGAAU-5'	U/U	48.8
6	II: 5'-rAUCACC^S**U**CCUUA-3'		
7	II + 3'-rUAGUGG-**A**-GGAAU-5'	^{S}U/A	65.5
8	II + 3'-rUAGUGG-**G**-GGAAU-5'	^{S}U/G	60.0
9	II + 3'-rUAGUGG-**C**-GGAAU-5'	^{S}U/C	51.2
10	II + 3'-rUAGUGG-**U**-GGAAU-5'	^{S}U/U	59.7
11	III: 5'-rAUCACC^{Se}**U**CCUUA-3'		
12	III + 3'-rUAGUGG-**A**-GGAAU-5'	^{Se}U/A	65.8
13	III + 3'-rUAGUGG-**G**-GGAAU-5'	^{Se}U/G	58.5
14	III + 3'-rUAGUGG-**C**-GGAAU-5'	^{Se}U/C	50.3
15	III + 3'-rUAGUGG-**U**-GGAAU-5'	^{Se}U/U	57.3

Table 2.2.2. Melting temperatures (T_m) of the native, S- and Se-modified RNA duplexes.

Entry	Sequences	Base pairs	Tm (ºC)	ΔTm$_1$	ΔTm$_2$
1	IV: 5'-rAAUGC**U**GCACUG-3'				
2	IV + 3'-rUUACG**A**CGUGAC-5'	U/A	64.1	-	
3	IV + 3'-rUUACG**G**CGUGAC-5'	U/G	59.4	-4.7	
4	IV + 3'-rUUACG**C**CGUGAC-5'	U/C	52.0	-12.1	
5	IV + 3'-rUUACG**U**CGUGAC-5'	U/U	51.5	-12.6	
6	V: 5'-rAAUGC^{Se}**U**GCACUG-3'				
7	V + 3'-rUUACG-**A**-CGUGAC-5'	^{Se}U/A	66.5		-
8	V + 3'-rUUACG-**G**-CGUGAC-5'	^{Se}U/G	55.5		-11.0
9	V + 3'-rUUACG-**C**-CGUGAC-5'	^{Se}U/C	51.9		-14.6
10	V + 3'-rUUACG-**U**-CGUGAC-5'	^{Se}U/U	58.0		-8.5

Table 2.2.3. Melting temperatures of native and 2-Se-U RNA modified duplexes (5'-rAAUGCUGCACUG -3').
ΔTm$_1$ refers to the Tm difference between the native U/A pair and the other mis-pairs (U/G, U/C and U/U), and ΔTm$_2$ refers to the Tm difference between the ^{Se}U/A pair and the other modified mis-pairs (^{Se}U/G, ^{Se}U/C and ^{Se}U/U).

UV-melting temperatures (Tm) of the native, S- and Se- modified duplexes with match and mismatch sequences are shown in Table 2.2.2, Table 2.2.3 and Figure 2.2.6. Tm of the Se-RNA duplex containing the ^{Se}U/A Watson-Crick pair was 2.4 or 3.0 ºC higher than those of the corresponding duplexes containing native U/A pair (Table 2.2.2, Figure 2.2.6). Comparing with native U/G, the ^{Se}U/G pair is approximately 4 ºC less stable than the native formation. While the ^{Se}U/C mis-pair is slightly less stable than the U/G pair, suggesting that ^{Se}U discourages the ^{Se}U/G pair native U/C mis-pair, the ^{Se}U/U mis-pair is more stable than the native U/U mis-pair. The higher stability may be attributed to the higher acidity of the imino group (3-NH) of ^{Se}U [pKa = 7.29 ± 0.02, Figure 2.2.2, compared to that of the native uridine (pKa = 9.18 ± 0.02)[45]], which may promote U/U inter- action via hydrogen bond. In addition, considering a selenium atom is 0.43 Å larger in atomic radius than an oxygen atom, the 2-Se atom may strengthen the stacking

interaction between ^{Se}U and its 3'-nucleobase (Figure 2.2.4).
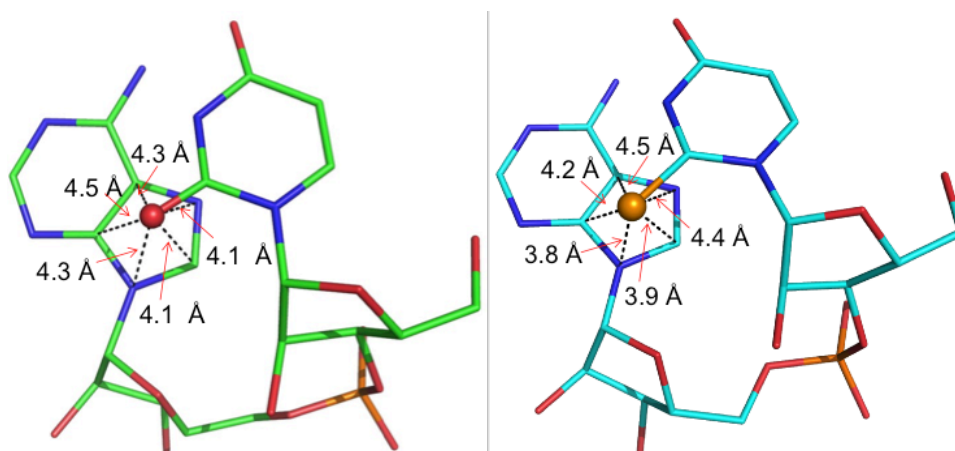
Figure 2.2.4. Local structures of the native RNA and ^{Se}U-containing RNA r[5'-GUAUA(^{Se}U)AC-3']$_2$ with a resolution of 2.3 Å.
The 2-position atom (oxygen or selenium) of U6 stacks with A7. (A) The native 5'-U6-A7-3' local structure; (B) the native 5'-^{Se}U6-A7-3' local structure. A selenium atom is 0.43 Å larger in atomic radius than an oxygen atom.

When directly comparing the Watson-Crick base pairs (U/A and ^{Se}U/A) with their own corresponding mis-pairs, it is clear that ^{Se}U/A pair has the balanced discrimination against all mis-pairs, with the Tm differences of ^{Se}U/G (7.3 °C in Figure 2.2.5 and 11.0 °C in Table 2.2.3), ^{Se}U/C (15.5 °C), and ^{Se}U/U (8.5 °C). On the other hand, the native U/A pair has poor discrimination against U/G wobble pair (the Tm differences: 0.3 °C in Figure 2.2.5 and 4.7 °C in Table 2.2.3), while maintaining fine discrimination against U/C (12.2 °C) and U/U (14 °C) pairs. Therefore, in general, ^{Se}U/A has higher base pair fidelity than the native U/A pair. When comparing the ^{Se}U/A with the corresponding ^{S}U/A pair, the same statement is also true (Figure 2.2.5). Furthermore, the Tm difference (8.5 °C) of ^{Se}U/A and ^{Se}U/U is bigger than the Tm difference (5.8 °C) of ^{S}U/A and ^{S}U/U, thus the ^{Se}U/A can better discriminate against ^{Se}U/U mis-pair than the ^{S}U/A against ^{S}U/U mis-pair. In general, ^{Se}U/A can better discriminate against all

corresponding mis-pairs than [S]U/A, thereby [Se]U/A offering higher base pair fidelity than [S]U/A.
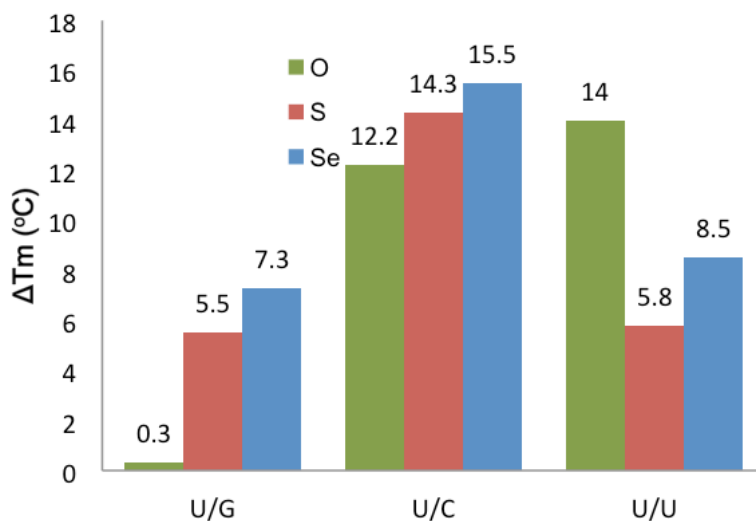
Figure 2.2.5. Differences of melting temperatures (Tm) of the native, S- and Se-modified U/A pairs and their corresponding mis-pairs.
O (white bar) refers to the Tm difference between the native U/A pair and the other mis-pairs (U/G, U/C and U/U); S (grey bar) refers to the Tm difference between the [S]U/A pair and the other modified mis-pairs ([S]U/G, [S]U/C and [S]U/U); Se (black bar) refers to the Tm difference between the [Se]U/A pair and the other modified mis-pairs ([Se]U/G, [Se]U/C and [Se]U/U).

As hypothesized, the 2-Se-functionality on uridine can indeed largely increase the base pairing specificity of RNA by discriminating against U/G wobble pairing. The Tm differences between the native U/A pair and U/G wobble pair were relatively small (4.7 °C in Figure 2 and 0.3 °C in Figure 2.2.6). The small Tm differences indicate possible changes between U/A and U/G pairs without a significant decrease in duplex stability. This is consistent with the ubiquitous presence of U/G wobble pair in RNAs, which diversifies the structure and function of RNAs, especially non-coding RNAs. Such small thermostability difference between native U/A pair and U/G wobble pair has been previously observed in the literature.[29a] Interestingly, the Tm differences between the [Se]U/A and [Se]U/G pairs were significant, such as 7.3 °C (*vs* 0.3 °C in the native) in Figure 2 and 11 °C (*vs* 4.7 °C in the native) in Table 2.2.3. The [Se]U modification in

RNA duplexes directly decreases the thermal stability of the U/G wobble pair by 4.0 ºC (Table 2.2.2) and 3.9 ºC (Table 2.2.3). This experimental observation reveals that the U/G wobble pair is greatly discriminated by incorporating a selenium atom to the 2 position of uridine. The strong discrimination against U/G pair is mainly attributed to the selenium disruption of the hydrogen bond formed by the 2-oxygen (Figure 2.1.1) and to the steric effect of the bulky selenium atom at the 2-position. Clearly, our results indicate that the 2-Se-modification on uridine significantly increases the high specificity of the U/A base pair.
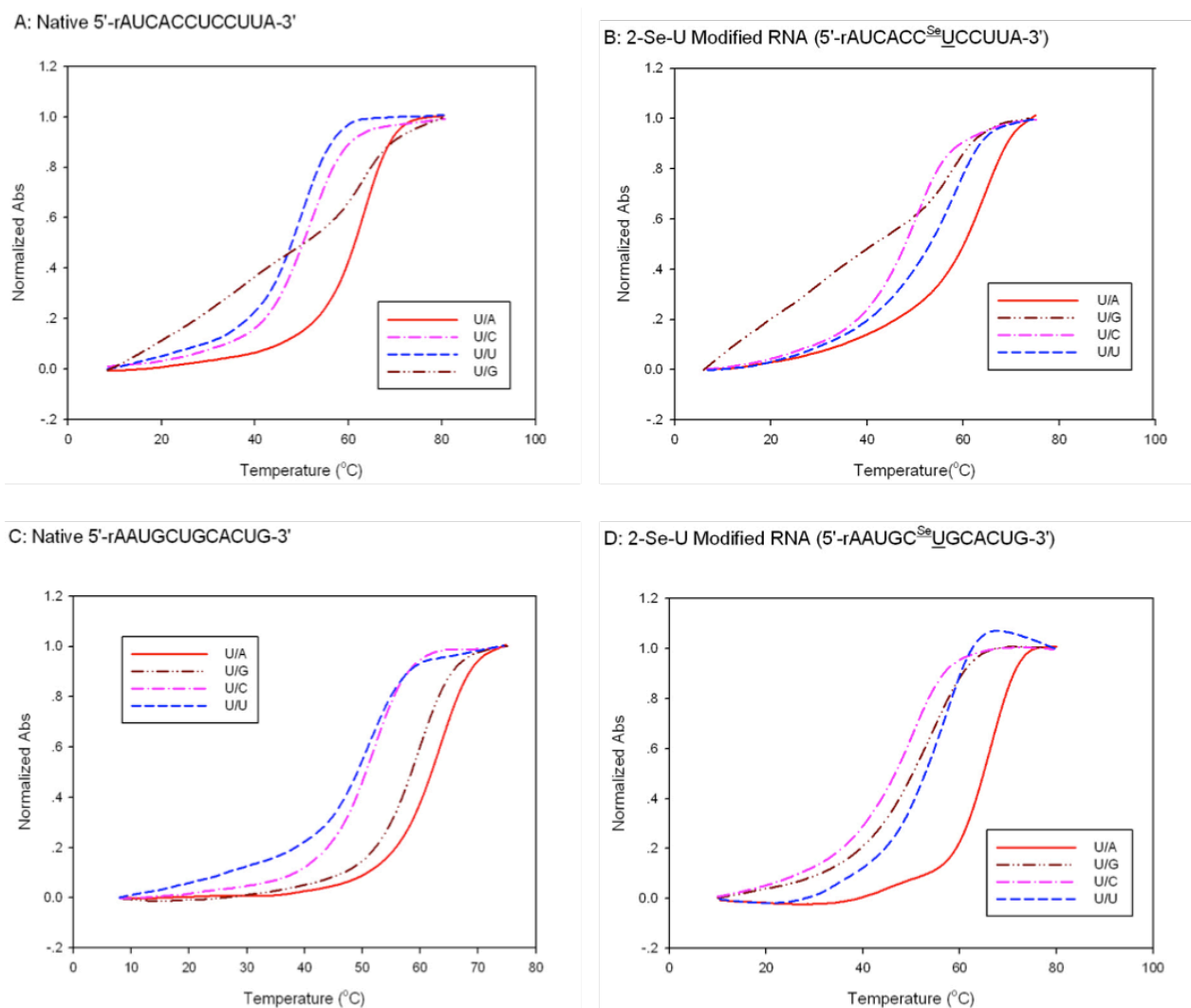
Figure 2.2.6. Normalized UV-melting curves of RNA duplexes.
(A) Native RNA (5'-rAUCACCUCCUUA-3') paired with matched and mismatched strands; (B) Modified RNA (5'-rAUCACC^{Se}UCCUUA-3') with matched and mismatched strands. (C) Native RNA (5'-rAAUGCUGCACUG-3') paired with matched and mismatched strands; (D) Modified RNA (5'-rAAUGC^{Se}UGCACUG-3') with matched and mismatched strands.

2.2.6 *Crystallization and data collection of Se-RNA*

Perfluoropolyether was used as a cryoprotectant during the crystal mounting, and data collection was taken under the liquid nitrogen stream at 99°K. The Se-RNA crystal data were collected at beam line X12B and X12C in NSLS of Brookhaven National Laboratory. A number

of crystals were screened to identify the one with strong anomalous scattering at the K-edge absorption of selenium. The distance of the detector to the crystals was set to 150 mm. The wavelength of 0.9795 Å was chosen for selenium SAD phasing. The crystals were exposed for 10 or 15 seconds per image with one degree oscillation, and a total of 180 images were taken for each data set. All the data were processed using HKL2000 and DENZO/SCALEPACK.[46]

Structure (PDB ID)	GUAUA-^{Se}U-AC (3S49)
Data collection	
Space group	R32
Cell dimensions: a, b, c (Å),	47.095, 47.095, 424.655,
α, β, γ (°C)	90, 90, 120
90, 90, 120	
Resolution range (Å) (last shell)	50.0–2.28 (2.37–2.28)
Unique reflections	9870 (959)
Completeness (%)	99.8 (99.4)
R_{merge} (%)	7.7 (23.5)
$I/\sigma(I)$	21.0 (3.9)
Redundancy	18.8 (10.7)
Refinement	
Resolution range (Å)	30.0–2.3
R_{work}(%)	21.4
R_{free} (%)	26.9
Number of reflections	8206
Number of atoms	
Nucleic acid (double)	1162
Heavy atoms and ion	7 Se
Water	71
RMS deviations	
Bond length (Å)	0.007
Bond angle	1.169

Table 2.2.4. Data collection and refinement statistics of 2-Se-U-RNA 8mer

The crystal structure study of the ^{Se}U-RNA [5'-rGUAUA-^{Se}U-AC-3'] is consistent with the biophysical results of ^{Se}U/A pairing. Similar to the native, the Se-RNA crystal is also in rhombohedral space group R32. The Se-RNA structure, determined at 2.3 Å resolution, is virtually identical to the native one[47] (at 2.2 resolution, Figure 2.2.7). Interestingly, the Se-RNA crystals grew much faster than the native ones. In six days, the Se-RNA formed diffraction-quality crystals in decent sizes (approximately 0.05 x 0.05 mm), while the corresponding native

did not crystallize in 3–4 weeks under the same conditions. Moreover, the Se-RNA crystals could form in broader buffer conditions (12 out of 24 conditions in Hampton buffers) than the corresponding native (2 out of 24 conditions). This observation of faster crystal growth of the Se-RNA is consistent with the Se-facilitated duplex stability. As shown in Figure 2.2.7A, there are seven self-complementary RNA molecules in a unit cell, and the overall shape of the duplexes is almost linear (approximately 8° inclination to the screw axes). Although this assembling pattern results in the discontinued backbones and grooves, the duplexes stack on top of each other in a head-to-tail fashion, and a peudo-fiber is formed. The data collection and structure refinement statistics are summarized in Table 2.2.4.
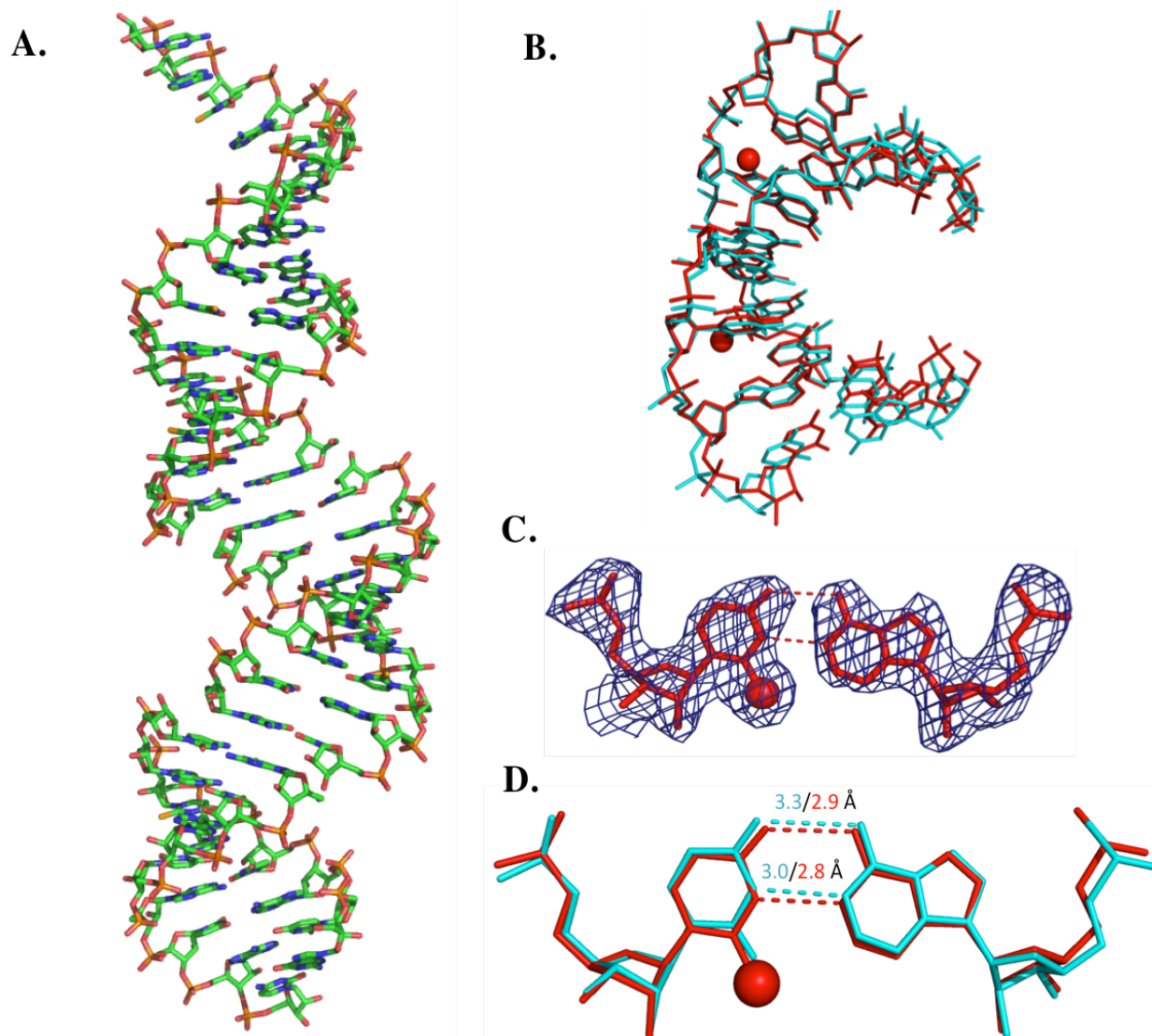
Figure 2.2.7. Global and local structures of the ^{Se}U-containing RNA r[5'-GUAUA(^{Se}U)AC-3']$_2$ with a resolution of 2.3 Å.

(A) The overall structure of duplex. (B) The superimpose comparison of one ^{Se}U-RNA duplex (red; PDB ID: 3S49) with its native counterpart [5'-r(GUAUAUA)-dC-3']$_2$ (cyan; PDB ID: 246D) with a RMSD value 0.55. The two red balls represent the selenium atoms. (C) The experimental electron density of ^{Se}U6/A11 base pair with s=1.0. (D) The superimpose comparison of the local base pair ^{Se}U6/A11 (red) and the native U6/A11 (cyan). The numbers indicate the distance between the corresponding atoms.

Since 2-exo-oxygen of uridine is not involved in the hydrogen bond interactions of U/A pairing, it's expected that the U/A pair will accommodate the larger selenium atom at this

position (Figure 2.1.1 and Figure 2.2.7C). The Se-modification also leads to the acidity increase of the 3-imino group (NH) in the 2-Se-uridine, which strengthens the hydrogen bond between N3 of U6 and N1 of A11. Indeed, after the selenium modification, the U/A hydrogen bond length between N3 of U6 and N1 of A11 is shortened from the native distance (3 Å) to the Se-modified distance (2.81 Å). Moreover, after the Se-modification (Figure 2.2.7D), the U/A hydrogen bond length between O4 of U6 and N6 of A11 decreases by 0.47 Å from the native distance (3.39 Å) to the Se-modified distance (2.92 Å). The shortened H-bond lengths indicate stronger H-bonds, which may explain the increase of duplex stability after the Se-modification. On the contrary, the distance between Se2 of U6 and C2 of A11 in the Se-modified duplex is slightly increased. This distance increase is likely due to a steric effect. This steric clash at the position 2 of the Se-uridine can be a driving force to increase [Se]U/A pair specificity. Consistent with our biophysical study, our structure study has indicated that the selenium bulkiness at the uridine 2-postion discourages the U/G wobble pairing. Moreover, due to the electronic effect of a selenium atom, the inability of a Se atom to form a stable hydrogen bond is another main factor responsible for the discrimination against U/G wobble pair.

2.3   Study of U/G Wobble Pair

2.3.1   U/G wobble pair experimental design and crystallization

Base on our biophysical studies, we were able to confirm the modified 2-Se-uridine is able to maintain the U/A base pair stability meanwhile destabilizing U/G wobble pair. Our 2.3Å resolution crystal structure of [Se]U/A RNA base pair further prove that the selenium atom did not disrupt the overall structure of the RNA but slightly shorten the local hydrogen bonding of the [Se]U/A pair and enhance the stacking of the [Se]U/A pair with neighboring base pairs. To further explore the [Se]U•G wobble pair in the real structure and for better comparison, we choose the

same self-complimentary sequence but switch the A to a G against [Se]U for the study. The newly designed sequence (5'-GUGUAUAC-3') should form two U•G wobble pairs in the eight bases double helix region (Figure 2.3.1). Both the native RNA sequence (5'-GUGUAUAC-3') and Se-modified RNA sequence (5'-GUGUA[Se]UAC-3') were synthesized through solid phase synthesis and purified by HPLC for crystal growth. The integrity of the Se-modified RNA is approved by mass spec analysis, Chemical Formula: $C_{76}H_{95}N_{29}O_{54}P_7Se$, $[M+H^+]^+$: 2574.3, observed 2574.4 (Figure 2.3.1).

<div align="center">

5'- G U A U A U A C -3'   5'- G U G U A U A C -3'
3'- C A U A U A U G -5'   3'- C A U A U G U G -5'

</div>

Figure 2.3.1. Design of RNA duplexes.
Left: 5'-GUAUAUAC-3' with all Watson-crick base pairs. Right: 5'-GUGUAUAC-3' with two U•G wobble pairs. Base pairs of interests were labeled in red.

The purified RNAs was adjusted to 1 mM concentration and was annealed with itself by heating up to 80°C and then was slowly cool to room temperature. The 24 screening buffers were from Hampton Research Nucleic Acid Mini screen kit (APPENDICES). The RNA was mix with screening buffer at 1:1 ratio hanging against 35% MPD (2-methyl-2,4-pentanediol). Among 24 conditions, the native sequence (5'-GUGUAUAC-3') was crystallized in four conditions (No. 10, 15, 17, 19) with in 24 hours while Se-modified sequence (5'-GUGUA[Se]UAC-3') was crystallized in thirteen conditions (No. 2, 3, 4, 6, 8, 10, 11, 12, 15, 17, 19, 20, 21) with in 24 hours. The native crystals perform rod-shaped while the Se-modified crystals appear long needle-shaped (Figure 2.3.2).
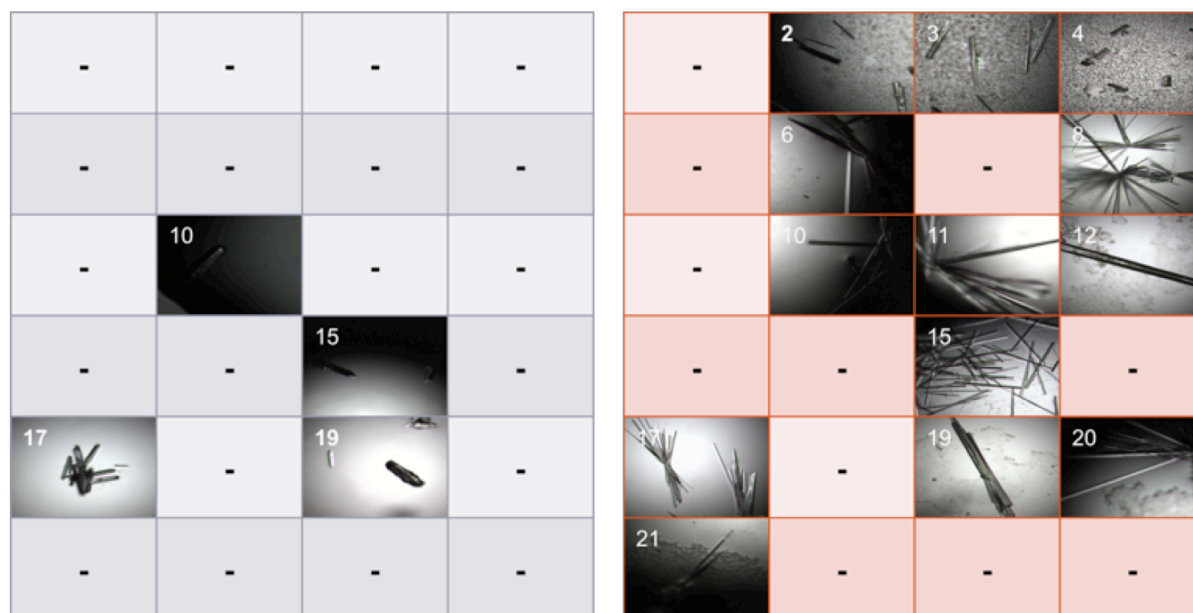
Figure 2.3.2. Crystal growth comparison with native and Se-modified RNA in 24 hours. Left: Native RNA crystal (5'-GUGUAUAC-3') with four buffer conditions. Right: Se-modified RNA (5'-GUGUA^{Se}UAC-3') with thirteen buffer conditions.

2.3.2   Crystal structure of Native 5'-GUGUAUAC-3' RNA
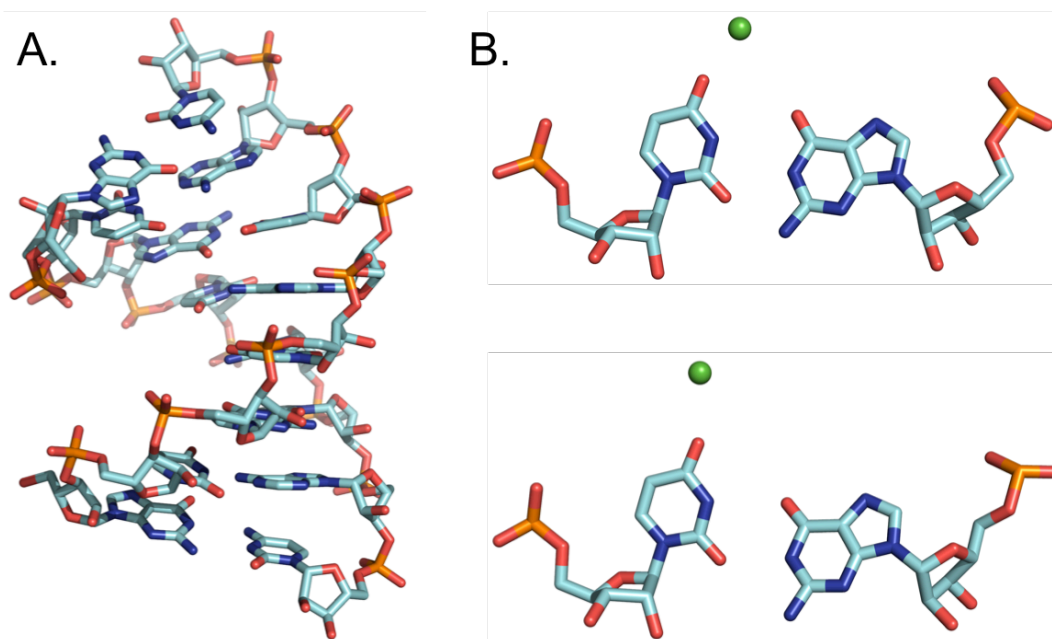
Figure 2.3.3. Global and local structures of native RNA 5'-GUGUAUAC-3' A. Global structures of native RNA 5'-GUGUAUAC-3'; B. Wobble pairs G3/U14 and U6/G11 with strontium ion at C4 position of uridine.

The native 5'-GUGUAUAC-3' structure was determined at 2.2 Å resolution with molecular replacement (PDB ID: 1JAB), the data was collected at beam line BL8.2.2 of the ALS (Advanced Light Source) at the Lawrence Berkeley National Laboratory. The structure of the A-form RNA shows a perfect 8-base duplex with two U•G wobble pairs buried in the middle. The structure was deposited in PDB (1JAB). From the structure, the two U•G pairs (G3•U14 and U6•G11) were both stabilized by strontium ions at C-4 position carbonyl of uridine (Figure 2.3.3). The U•G wobble pairs were also compared with U/A Watson-crick base pairs by global structures superimpose of the both RNA sequences (Figure 2.3.4). The structures were almost identical and no significant perturbation observed.
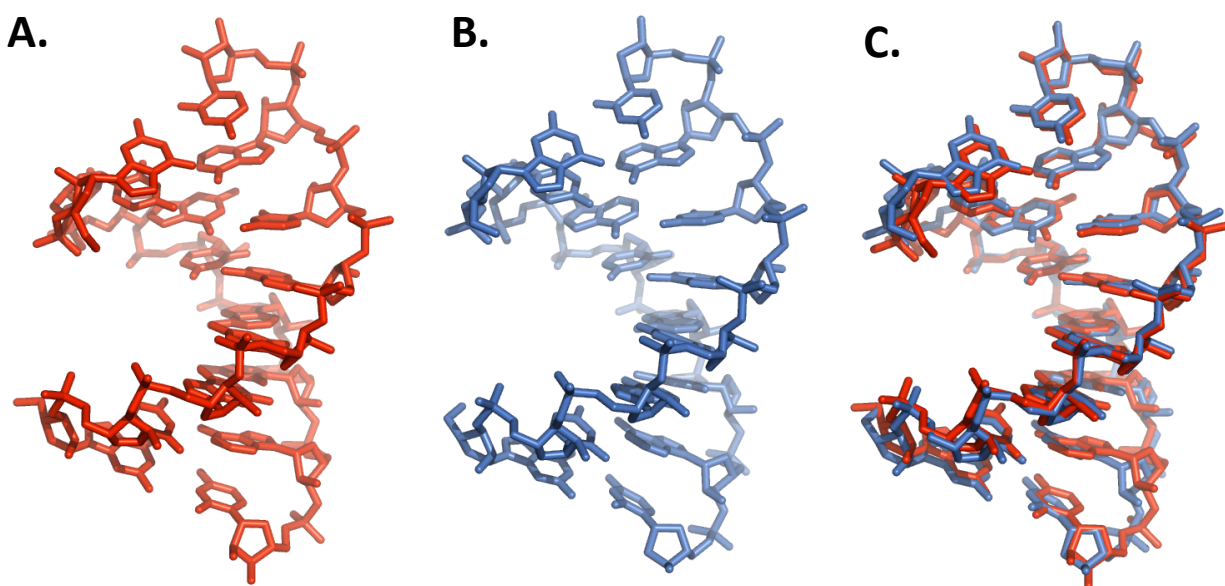
Figure 2.3.4. Native RNA 5'-GUGUAUAC-3' and 5'-GUAUAUAC-3' structures.
A. Native RNA structure (5'-GUAUAUAC-3') PDB: 246D with all watson-crick base pairs; B. Native RNA structure (5'-GUGUAUAC-3') PDB: 1JAB with two wobble base pairs; C. Superimposed of 1JAB and 246D structures.

2.3.3 Crystal structure of 5'-GUGUA^{Se}UAC-3' RNA
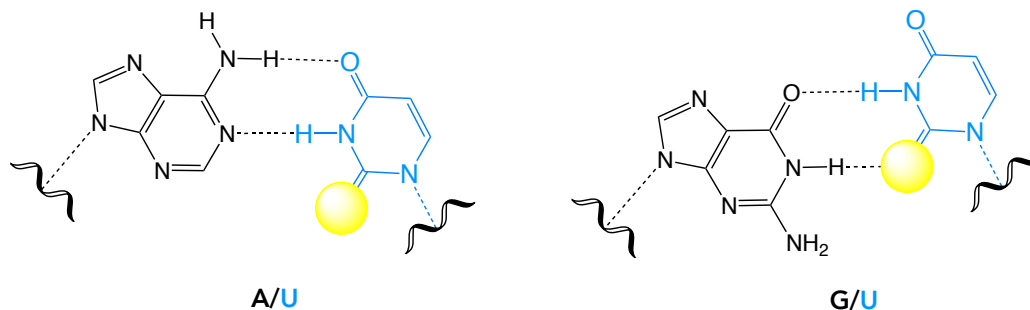
Figure 2.3.5. Hydrogen bonding pattern of ^{2Se}U/A base pair and ^{2Se}U•G wobble pair.

The actual structure of the Se-modified was not as expect as the native structure (5'-GUGUAUAC-3'). However, the selenium atom at C2 of uridine generates a great hindrance when pairing with G to form a wobble pair (Figure 2.3.5), thus RNA refuses to form an 8-based double-strand helix. Instead, each RNA duplex is formed with 6 base pairs and a two nucleotides overhang on each end (Figure 2.3.6). In this particular situation, the Se-modified uridine forms ^{Se}U/A watson-crick base pair to avoid the ^{Se}U•G wobble pair but rather to form two native U•G wobble pair at each of the overhang end with neighboring duplex, which made it altogether two U•G wobble pairs and two ^{Se}U/A watson-crick base pairs in each 8-based unit. This observation has strongly confirmed our previous hypothesis again that when the uridines at wobble base pair positions are replaced by ^{Se}U, the U•G wobble pair was strongly discouraged.
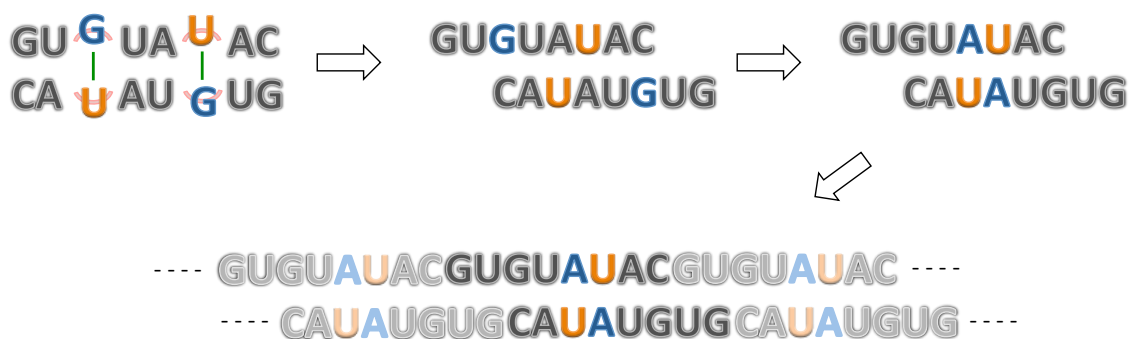
Figure 2.3.6. The ^{Se}U-RNA (5'-GUGUA^{Se}UAC-3') crystal structure formation.

The 5'-GUGUA<u>^{Se}U</u>AC-3' structure was determined at 1.5 Å resolution with molecular replacement (PDB ID: 1JAH). The data was collected at beam line BL8.2.2 of the ALS (Advanced Light Source) at the Lawrence Berkeley National Laboratory. This high-resolution structure offers the most detailed ^{2Se}U containing RNA structure to date. In each duplex pair, there are two selenium atoms in the minor groove of A-form RNA, with distance of 3.69 Å, 3.93 Å, and 3.93 Å respectively (Figure 2.3.7). The structure indicated a three RNA duplex bundle in the asymmetric unit. The intermolecular contacts of the overhand U•G interactions between duplexes probably drive the packing of crystals.
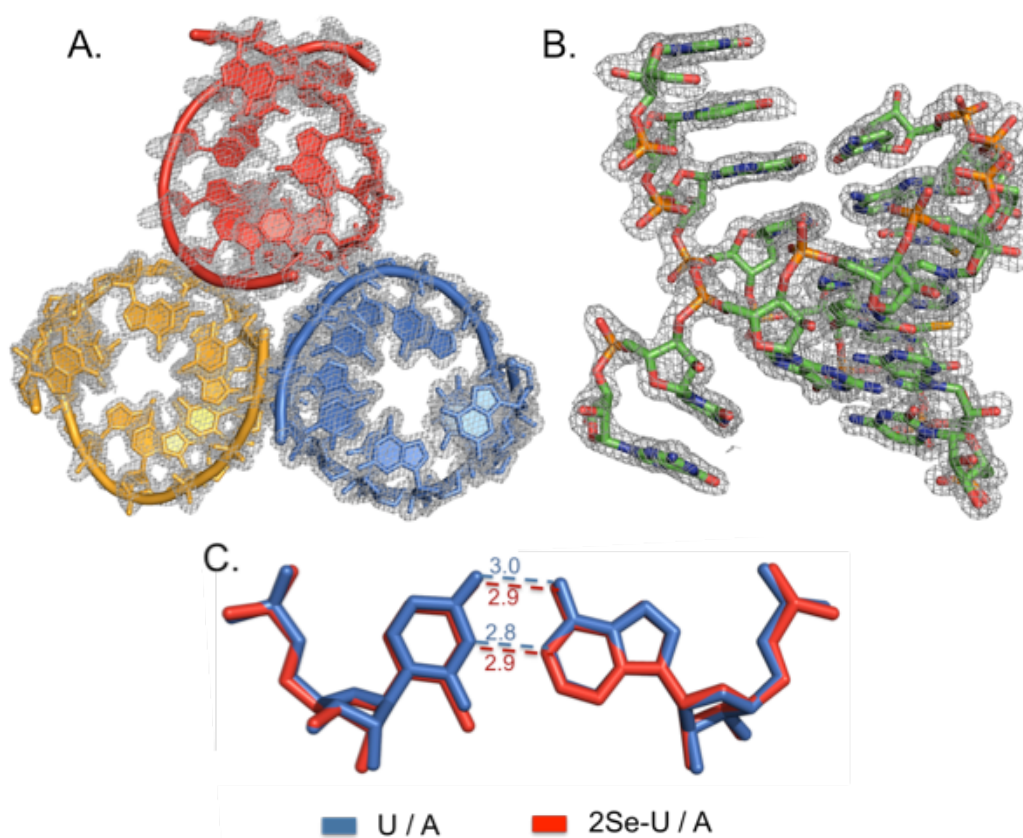
Figure 2.3.7. Structure of 5'-GUGUA<u>^{2Se}U</u>AC-3' RNA with electron density map.
A. Three duplexes bundle in asymmetric unit; B. RNA duplex with two nucleotides overhang on each end; C. ^{2Se}U/A and U/A base pairs superimposed with each other and the measurements of the hydrogen bonds.

3   2-SELENOURIDINE TRIPHOSPHATE SYNTHESIS AND SE-RNA TRANSCRIPTION

3.1   Introduction

The text of this work has been published as "2-Selenouridine Triphosphate Synthesis and Se-RNA Transcription", RNA, 2013, 19, 1309-1314. I would like to acknowledge Dr. Sibo Jiang, Dr. Julianne Caton-Williams, Dr. Hehua Liu and Dr. Zhen Huang for their intellectual contribution as co-authors.

3.1.1   RNA modification

RNA is involved in numerous biological processes, such as genetic storage, transcription, translation, and regulation.[2a, 27b] Moreover, RNA can fold into well-defined three-dimensional structures to interact with proteins and catalyze biochemical reactions.[1b, 2b] The appreciation for the uniqueness of RNAs, especially non- coding RNAs for their structure and function diversities, has increased extensively in the past decade. However, the functional understanding of these complicated macromolecules is often limited. The functional understanding of many natural modifications of the RNAs is even less. Thus, studying RNA natural modifications has become a very active research area in order to better understand biophysical and chemical properties of RNAs (such as tRNA and rRNA). So far, >100 RNA modifications have been discovered in nature,[5] and many of them are frequently dis- covered in tRNA. 2-Selenouridine ($^{2-Se}$U or ^{Se}U) is one of naturally occurring nucleosides and exists at the wobble position of the anticodon loop in various bacterial tRNAs (Escherichia coli, Methanococcus vannielii, Clostridium sticklandii, etc.).[5, 18] This Se-modification might play a critical role in the mRNA decoding process. It was hypothesized that the 2-Se-modification may enhance the accuracy and

efficiency of protein translation.[9a, 48]

3.1.2 Selenium in X-ray crystallography

Moreover, another advantage of selenium modification in nucleic acid research is its assistance in addressing phase issue in X-ray crystallography via multiwavelength anomalous dispersion (MAD) or single-wavelength anomalous dispersion (SAD). Heavy atoms, such as selenium (Se) and bromine (Br), are suitable as anomalous scattering centers, which have been extensively applied in protein and nucleic acid crystallography. Encouraged by the successful selenium-assisted MAD phasing,[49] we have pioneered and established nucleic acid X-ray crystallography with selenium derivatization.[9b, 11, 50] Among the synthesized Se-derivatives, 2-selenouridine is stable and the only one found in nature so far. Furthermore, the single oxygen atom substitution with selenium at the exo-2 position doesn't interfere with the hydrogen bonding in the Watson-Crick U/A base pair, thereby preserving the base-pairing function and structure.[9a] Therefore, the 2-selenouridine synthesis and its incorporation into RNAs may largely facilitate both structure and function investigations.

3.1.3 Methods for Se-RNA synthesis

Generally, there are two strategies to synthesize the Se-derivatized RNAs: solid-phase synthesis, and transcription. The first method offers the site-specific incorporation of the Se-nucleoside. However, it is limited to relatively short RNAs (up to 50 nt) for large-scale synthesis. In addition, it requires multiple steps in deprotection and purification. The 2-selenouridine chemical incorporation into RNAs has been achieved via solid-phase synthesis.[9a] Our biophysical studies have shown that the 2-Se-modification discriminates against a U/G mismatch (wobble pair), while pre- serving the native U/A pair. This result indicates that ^{Se}U can largely improve

the RNA base-pairing specificity and the RNA–RNA interaction fidelity. This result has encouraged us to incorporate the Se-modification into RNA by in vitro transcription, in order to further investigate the function and structure of the [Se]U-RNAs, such as the [Se]U-containing tRNAs. This enzymatic method can allow synthesis of longer RNAs (>50 nt) in a large quantity (multiple milligrams). Multiple selenium atoms can also be conveniently incorporated into RNA under the mild conditions. As a matter of fact, the transcription strategy with T7 RNA polymerase is favored by most molecular and structural biologists. Herein, we report the first synthesis of 2-selenouridine triphosphate ([Se]UTP) and the enzymatic incorporation of [Se]UTP into noncoding RNAs. The active and mutant hammerhead ribozymes (Figure 3.1.1) were successfully transcribed and examined with [Se]UTP. The transcribed [Se]U-hammerhead ribozyme is active, suggesting that the [Se]U- RNAs are useful in both function and structure studies of noncoding RNAs.
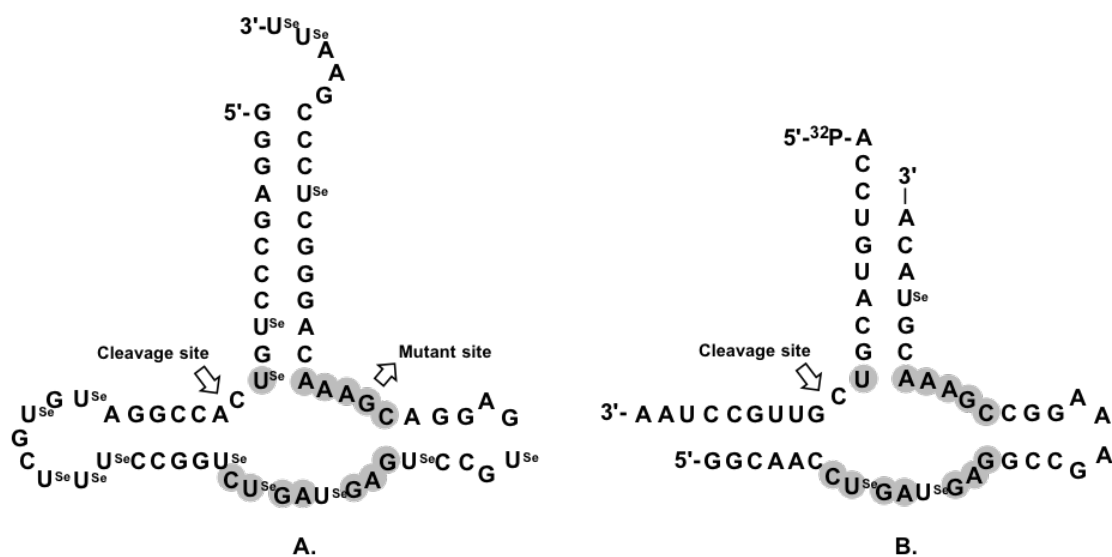
Figure 3.1.1. Hammerhead ribozymes.
(A) Secondary structure of the self-cleaving [Se]U-hammerhead ribozymes, including the wild type (WHR) and crippled mutant (MHR). The mutant and cleavage sites are indicated by arrows. Highly conserved bases are highlighted in gray. (B) Secondary structure of the non-self- cleaving [Se]U-hammerhead ribozyme and its 5'-^{32}P-labeled RNA substrate.

3.2   General Experiment Section

3.2.1   2-Se-uridine triphosphate synthesis

In order to minimize by-product formation, the Se-nucleobase modifications are normally protected during chemical synthesis.[9a, 20-21, 51] Since the 2-seleno-modification on uridine is naturally occurring, we decided to directly explore its compatibility with chemical synthesis. We were pleasantly surprised that 2-seleno-uridine, without protection, can be directly converted to the corresponding triphosphate. Thus, the synthesis (Scheme 3.2.1) of [Se]UTP (3) started from deprotection of the 5'-DMTr group of the Se-uridine derivative 1[9a] under an acidic condition. Then, 2-Se-uridine (2) was converted to [Se]UTP (3) via a one-pot synthesis: sequential treatments with phosphorus oxychloride (POCl$_3$), pyrophosphate, and bicarbonate.[51a, 52]
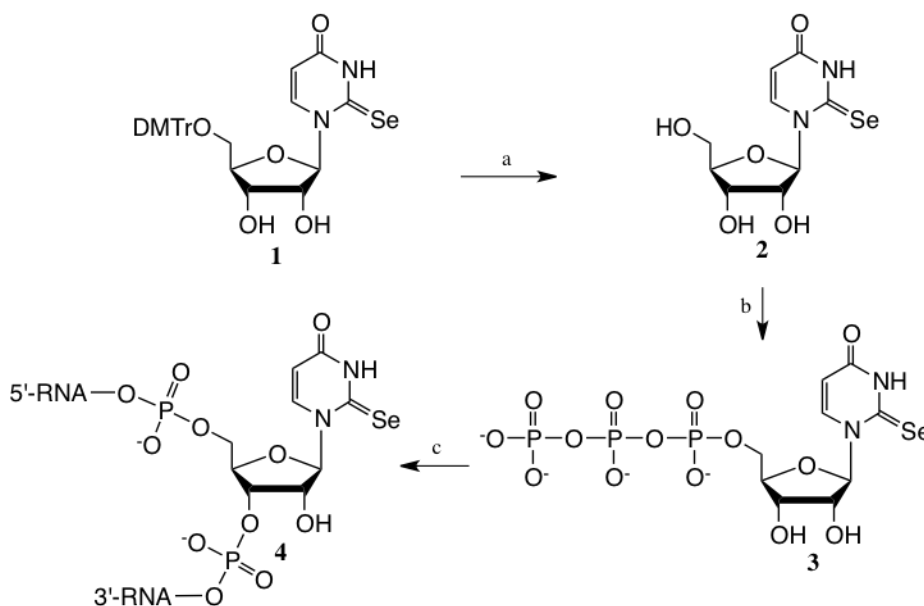
Scheme 3.2.1. Chemical synthesis of [Se]UTP and transcription of [Se]U-containing RNA. Reagents and conditions: (a) 4% trifluoroacetic acid; (b) POCl$_3$, Me$_3$PO$_4$; (tri-n-butyl)amine, pyrophosphate, N, N-dimethyl-formamide; the H$_2$O hydrolysis; (c) RNA transcription.

3.2.1.1   Synthesis of 2-Se-uridine

Trifluoroacetic acid (11 mg) was added to 5'-DMTr-2-Se-uridine (Scheme 3.2.1, step 1, 305 mg, 0.5 mmol)[9a] (Sun et al. 2012) in dichloromethane (5 mL). The solution was heated at 40°C for 30 min, followed by adding methanol (0.2 mL). The reaction was stirred vigorously for another 1 h to obtain a light yellow precipitate product (Scheme 3.2.1, step 2). The precipitate was recovered by centrifugation or filtration; the yield of step 2 was almost quantitative.

3.2.1.2   Synthesis of 2-Se-uridine triphosphate

2-Se-uridine (Scheme 3.2.1, step 2, 20 mg) was weighed and dried in a flask under high vacuum overnight, followed by injecting trimethyl phosphate (0.4 mL) to dissolve it and then stirring the flask in an ice bath. A solution of proton-sponge (55 mg, 2 eq) in trimethyl phosphate (0.3 mL) was injected into the solution of step 2 at 0°C. After 3 min stirring, phosphorus oxychloride ($POCl_3$; 9 μL, 1.5 eq) diluted in trimethyl phosphate (90 μL) was dropwise added into the solution of step 2 at 0°C. The reaction was completed in 1.5 h (monitored on TLC). Tributylammonium pyrophosphate (64 mg, 2 eq., dissolved in 0.2 mL tributylamine and 0.4 mL DMF) was then quickly injected into the reaction. After vigorously stirring for 5 min, the reaction was quenched with triethylammonium bicarbonate (1 M, 3 mL) and stirred for another 1 h at the room temperature to obtain compound 3. To the reaction solution, NaCl (3 M NaCl, 0.5 mL) was added, followed by adding ethanol (14.5 mL) and freezing the suspension at −80°C for 1 h to precipitate the crude product. Compound 3 was recovered by centrifugation for 25 min at 14,000 rpm. The pellet was redissolved in water and analyzed by HPLC. [Se]UTP (step 3) was purified by HPLC. The identity of [Se]UTP as a triethylammonium salt was confirmed by NMR (^{1}H-, ^{13}C-, and ^{31}P-NMR) and mass analyses. ^{1}H-NMR (400 MHz; D_2O) δ: 8.22 (d, J = 8.1 Hz, 1H, H-6), 6.79 (d, J = 3.0 Hz, 1H, H-1'), 6.38 (d, J = 8.1 Hz, 1H, H-5), 4.54–4.33 (m, 5H, H-2',

3', 4', 5'), 3.21 (d, J = 7.3 Hz, CH_2 of triethylammonium), 1.28 (t, J = 7.3 Hz, CH_3 of triethylammonium). ^{13}C-NMR (100 MHz; D_2O) δ: 174.7 (s, C-4), 161.7 (s, C-2), 140.9 (s, C-6), 107.6 (s, C-5), 94.6 (s, C-1'), 82.2 (d, J = 9.1 Hz, C-4'), 74.0 (s, C-2'), 67.3 (s, C-3'), 62.9 (d, J = 5.1 Hz, C-5'), 45.6 (s, CH_2 of triethylammonium), 7.2 (s, CH_3 of triethylammonium). ^{31}P-NMR (162 MHz; D_2O) δ: −7.4 (d, J = 19.7 Hz, α-P), −11.3 (d, J = 19.6 Hz, γ-P), −22.1 (t, J = 19.6 Hz, β-P). HRMS (ESI-TOF) $[M-H^+]^-$ = 546.8812 (calc. 546.8829)

3.2.1.3 *Purification and analysis of 2-Se-uridine triphosphate*

The maximal UV absorbance of native uridine triphosphate is 260 nm, while that of the ^{Se}U-triphosphate is 307 nm. In the HPLC analysis, both the native and selenium-modified UTPs were monitored under two wavelengths (260 and 307 nm). The synthesized ^{Se}UTP was purified by HPLC (Ultimate XB-C18, 250 mm×21.2 mm, 10 μm) with a gradient of 100% buffer A (20 mM triethylammonium acetate in water) to 25% buffer B (20 mM triethylammonium acetate in 50% acetonitrile and 50% water) for 20 min. The HPLC analysis was performed (Ultimate XB-C18, 250 mm × 4.6 mm, 5 μm) with a gradient from 100% buffer A (20 mM triethylammonium acetate in water) to 40% buffer B (20 mM triethylammonium acetate in 50% acetonitrile and 50% water) for 15 min. The HPLC and UV profiles are shown in Figure 3.2.1. The retention times of the native UTP and ^{Se}UTP were 11.2 and 14.1 min, respectively.
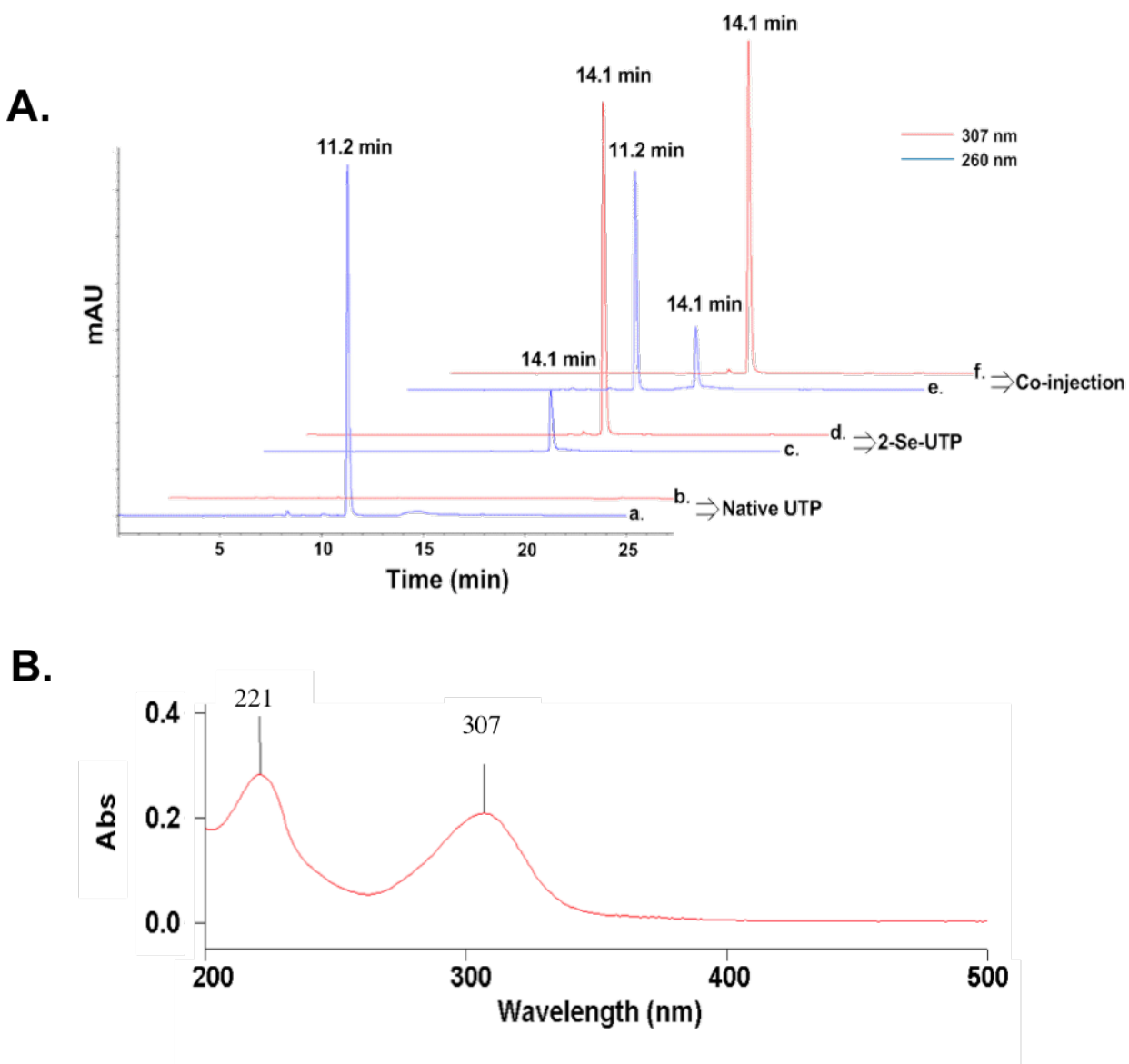
Figure 3.2.1. HPLC and UV analyses of ^{Se}UTP.
A) HPLC profiles: a) Native UTP monitored at 260 nm (retention time: 11.2 min); b) Native UTP monitored at 307 nm; c) ^{Se}UTP monitored at 260 nm (retention time: 14.1 min); d) ^{Se}UTP monitored at 307 nm (retention time: 14.1 min); e) co-injection of both native UTP and ^{Se}UTP monitored at 260 nm (retention time: 11.2 min and 14.1 min); f) co-injection of both native UTP and ^{Se}UTP monitored at 307 nm (retention time: 14.1 min). B) UV-spectrum of ^{Se}UTP (l_{max}= 307 nm).

3.2.2   Transcription of RNAs

Transcription experiment was carried out following standard procedures from manufacturer Epicentre (AmpliScribe™ T7-Flash™ Transcription Kit). ATP [α-^{32}P] was use as radioactive labeling material in this experiment. For each reaction (5 μL), final concentration of 0.5 mM NTP (A, U, G, C in the transcription of native RNAs and A, ^{Se}U, G, C in the transcription of Se-modified RNAs), 50 ng/μL linearized DNA plasmid template, 10 mM DTT, 1×transcription buffer for T7 RNA polymerase and 0.5 μL of T7 RNA polymerase (10 U) were added into reaction tube with RNase-free water to adjust total reaction volume to 5 μL. In transcription efficiency (time-course) experiment, a gel loading dye (5 μL) with 100 mM EDTA was used to terminate reaction at each time point with additional heating (75°C for 30 min). Later the experimental result was visualized via denaturing urea PAGE gel (15%) and autoradiography. Two templates practiced in transcription experiments were double-stranded DNA prepared by PCR. The translated RNAs are mutant hammerhead ribozyme (MHHR) with sequence   of   5'-GGGAGCCCUGUCACCGGAUGUGCUUUCCGGUCUGAUGAGUCCGUGAGGACAAAAC AGGGCUCCCGAAUU-3' (Figure 3.1.1) and wild-type hammerhead ribozyme (WHHR) with sequence   of   5'-GGGAGCCCUGUCACCGGAUGUGCUUUCCGGUCUGAUGAGUCCGUGAGGACGAAAC AGGGCUCCCGAAUU-3' (Figure 3.1.1).[53]

3.2.2.1   Transcription with native NTPs

All native NTPs, the transcription buffer, and T7 RNA polymerase used in our transcription experiments were purchased from Epicentre. The templates of the

wild-type and mutant hammerhead ribozymes were from the linearized plasmids (Lin et al. 2011a). The native RNAs were transcribed with the transcription protocol (final concentration) in RNA polymerase buffer (40 mM Tris-HCl, 6 mM MgCl$_2$, 2 mM spermidine, pH 7.9), DTT (10 mM), ATP, UTP, CTP, and GTP (0.5 mM each NTP), DNA template (non-self-cleaving hammerhead ribozyme: 1 μM dsDNA template [55 nt]; self-cleaving hammerhead ribozyme [mutant and wild-type]: 50 ng/μL linearized plasmid), T7 RNA polymerase (2 units/μL, Epicentre), and RNase-free water to adjust to the final volume (e.g., 20 μL). The transcription reaction was incubated for 1 h at 37°C.

3.2.2.2 *Transcription and analysis of Se-RNAs*

The Se-RNAs were transcribed with the transcription protocol (final concentration) in RNA polymerase buffer (40 mM Tris-HCl, 12 mM MgCl$_2$, 2 mM spermidine, pH 7.5), DTT (10 mM), ATP, ^{Se}UTP, CTP, and GTP (0.5 mM each NTP), DNA template [non-self-cleaving hammerhead ribozyme: 1 μM dsDNA template (55 nt); self-cleaving hammerhead ribozyme (mutant and wild-type): 50 ng/μL linearized plasmid], T7 RNA polymerase (4 units/μL, Epicentre), and RNase-free water to adjust to final volume (e.g., 20 μL). The transcription reaction was incubated for 3 h at 37°C. To examine the ^{Se}UTP compatibility with RNA polymerase in transcription, the linearized plasmid templates for the wild-type hammerhead-ribozyme (WHR) and the crippled mutant hammerhead-ribozyme (MHR) were used (Figure 3.1.1) for the ^{Se}U-RNA transcription. As expected, ^{Se}UTP was recognized by T7 RNA polymerase (Figure 3.2.2A). Moreover, the mutant ^{Se}U-ribozyme (69-nt; containing 15 selenium atoms) was pre- pared via RNA transcription, and the integrity of the ^{Se}U-ribozyme (^{Se}U-MHR) was confirmed by MS analysis (Figure 3.2.2C).
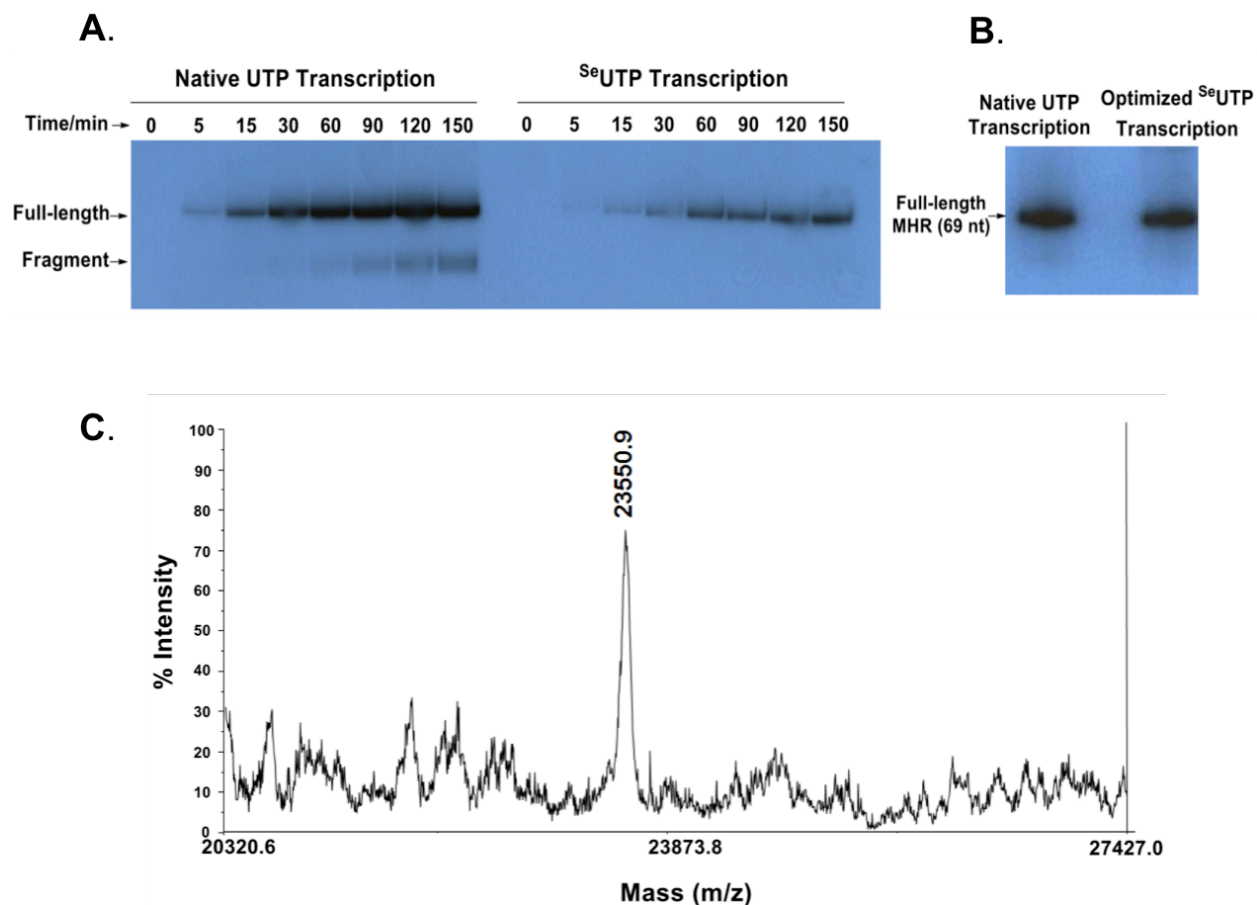
Figure 3.2.2. The ^{Se}U-ribozymetranscription with ^{Se}UTP and T7 RNA polymerase. (A) The auto-radiography gel image of in vitro transcription; (left) transcription of the native RNA (the crippled mutant hammerhead-ribozyme: MHR) with all native NTPs; the minor faster-moving band is the self-cleaved product (fragment); (right) transcription of the ^{Se}U-MHR with ^{Se}UTP and other native NTPs. (B) Optimized Se-RNA transcription (~85% yield compared to the corresponding native RNA transcription). Transcription conditions are listed in Table 3.2.1. C) MALDI-TOF MS analysis of the ^{Se}U-MHR (molecular formula: $C_{657}H_{817}N_{264}O_{476}P_{71}Se_{15}$); matrix: 3-hydroxypicolinic acid (3HPA, molecular formula: $C_6H_5NO_3$); mass of ^{Se}U-MHR and matrix observed: 23550.9 (calc. 23551.4).

3.2.2.3   Optimization of ^{Se}U-RNA transcription

To maximize the transcription yield, condition optimizations have been performed. The linearized plasmid of the mutant hammerhead ribozyme (Figure 3.1.1) was used as the template, which incorporates 15 ^{Se}Us into the ribozyme. The transcription buffers with various pH values were first examined, since the acidity of the imino group (3-NH) of ^{Se}U is higher than that of the

native U.[9a] The pH values of the transcription buffer (40 mM Tris base or sodium phosphate, 6 mM MgCl$_2$, 2 mM spermidine, and 10 mM DTT) were adjusted. The Se-RNA transcription was examined under eight pH values (pH 5.5, 6.0, 6.5, 7.0, 7.5, 8.0, 8.5, and 9.0) and indicated that pH 7.5 was optimal for the Se-RNA transcription (Figure 3.2.3). The pH of the standard transcription buffer is 7.9.
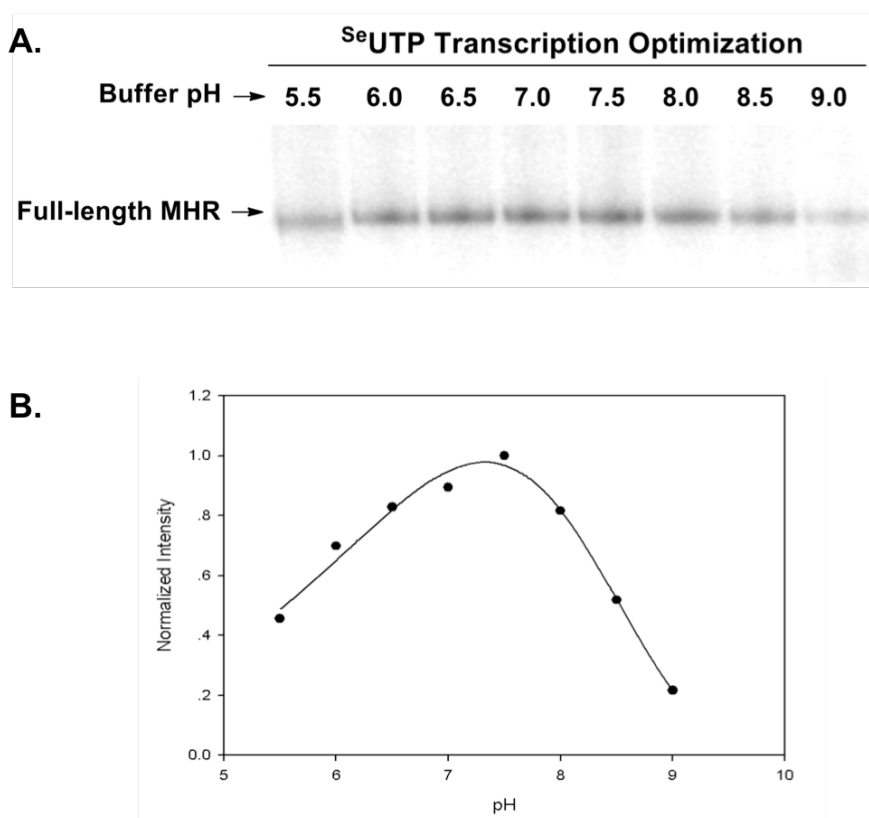
Figure 3.2.3. Experimental results of transcription optimizations with [Se]UTP. (A) Optimization of the transcription buffer pH (5.5–9.0). pH 7.5 is optimal for the Se-RNA transcription, while the pH of the standard transcription buffer is 7.9. (B) Data analysis of the pH optimization.

Mg^{2+} concentration in the transcription buffer was also examined by varying it from 4 to 12 mM. As the increased MgCl$_2$ concentration yielded higher transcription yield (Figure 3.2.4A), 12 mM MgCl$_2$ was chosen for the Se-RNA transcription. Other components, such as spermidine

(from 2–8 mM) and ^{Se}UTP (from 0.5 to 1.5 mM), were also examined for the transcription optimization. However, we found that increases of the concentrations of these components slightly decreased the transcription yield (Figure 3.2.4B).

Table 3.2.1. Optimized conditions for the Se-RNA transcription

	pH	Mg^{2+}	UTP concentration	T7 polymerase	Transcription time
Native condition	7.9	6 mM	0.5 mM	10 units	1h
Se-modified condition	7.5	12 mM	0.5 mM	20 units	3h

Moreover, a higher quantity of T7 RNA polymerase can increase the Se-RNA transcription yield (Figure 3.2.4C). Finally, after combining these optimized conditions (Table 3.2.1), we could increase the yield of the ^{Se}U-RNA transcription up to 85% of the corresponding native RNA (Figure 3.2.2B), and these conditions have been used to transcribe various ^{Se}U-RNA.

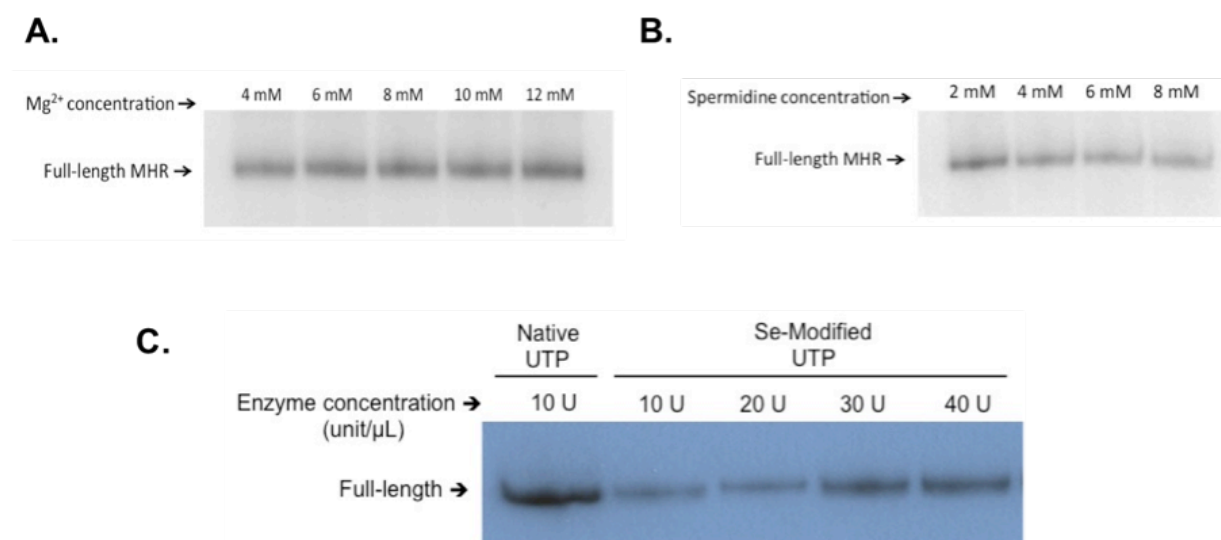**Figure 3.2.4.** Experimental results of transcription optimizations with ^{Se}UTP.
A). Optimization with increasing MgCl$_2$ concentrations. Standard transcription buffer contains 6 mM MgCl$_2$. B). Optimization with increasing spermidine concentrations. Standard transcription buffer contains 2 mM spermidine. C). Optimization with T7 RNA polymerase quantity (10 to 40 unit/μL).

3.2.3 Catalytic activity analysis of the Se-RNAs

3.2.3.1 Self-cleaving wild-type hammerhead ribozyme

To examine hammerhead ribozyme activity with regard of time, we use the wild-type template to transcript both native and Se-modified ribozyme in a self-cleavage manner since the wild-type template directs active ribozyme synthesis while the ribozyme cleaves itself spontaneously (Figure 3.2.5A). The transcription reaction was carried out in standard T7 reaction buffer (containing 6 mM $MgCl_2$) and the result clearly indicated that under standard transcription condition the native hammerhead ribozyme cleavage itself completely while Se-modified ribozyme is quite active but the cleavage is not complete. Later, we examined the selenium modified ribozyme activity in a transcription solution with increased $MgCl_2$ concentration to 10 mM (ref). Under this condition, Se-modified ribozyme gives a complete and efficient activity (Figure 3.2.5B).
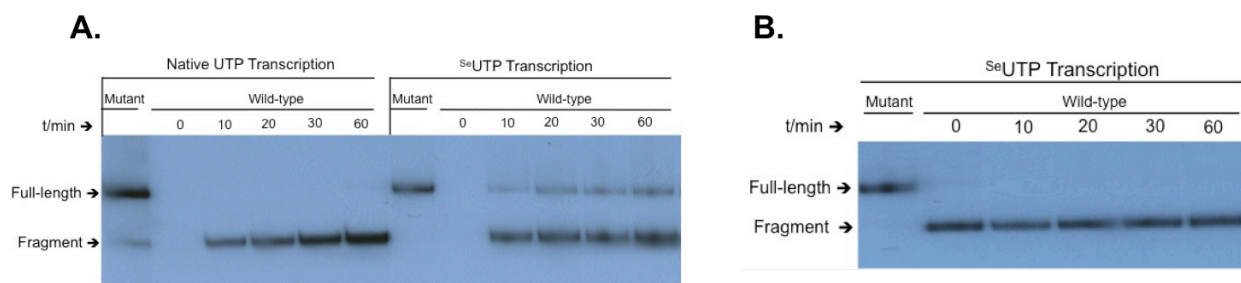
Figure 3.2.5. Wild-type native and Se-modified ribozyme transcription with self-cleavage activity during synthesis.
A) The experiment is carried out under standard transcription buffer condition (40 mM Tris-HCl, 6 mM $MgCl_2$, 10 mM NaCl, 2 mM spermidine and 10 mM DTT) and a mutant native and Se-modified ribozyme was used as comparison respectively. B) Wild-type selenium modified ribozyme transcription with a standard transcription buffer of 10 mM $MgCl_2$. A Se-modified mutant ribozyme is used as comparison.

3.2.3.2 *Non-self-cleaving hammerhead ribozyme*

The non-self-cleaving hammerhead ribozyme (5'-GGCA-ACCUGA UGAGGCCGAAAGGCCGAAACGUACA-3') (Figure 3.1.1) for the catalytic experiments was transcribed following the standard procedures described above. The DNA template used for this transcription was a 55-nt dsDNA (5'-TGTACGTTTCGGCCTTTCGGCCTCATCAGGTTGCCTATAGTGAGTCGTATTACGC-3' and its complementary sequence). After the transcription, the native and Se-modified ribozymes were purified and adjusted to the same concentration (monitored by UV). The RNA substrate (20 nt, 5'-ACCUGUACGUCGUUGCCUAA-3') (Figure 3.1.1) chemically synthesized by solid-phase synthesis was kinased with γ-^{32}P-ATP at the 5' end for the ribozyme digestion. The digestion was performed in the buffer (10 mM Tris-HCl, 10 mM MgCl$_2$, pH 7.6) and with 5'-^{32}P-labeled RNA substrate (final concentration: 50 μM) at 27° C. Aliquots (10 μL each) were taken at the time intervals (0, 5, 10, 30, 90, and 150 min), and each was mixed with EDTA (5 μL, 50 mM) dissolved in a saturated urea solution (aqueous) to quench the digestion. The 5'-labeled RNA substrate was digested to the 9-nt fragment and the 5'-^{32}P-RNA fragment (11 nt). The ^{32}P-labeled RNA allowed monitoring the substrate digestion via gel electrophoresis and autoradiography. The time-course results of the ribozyme digestion are shown in Figure 3.2.6.
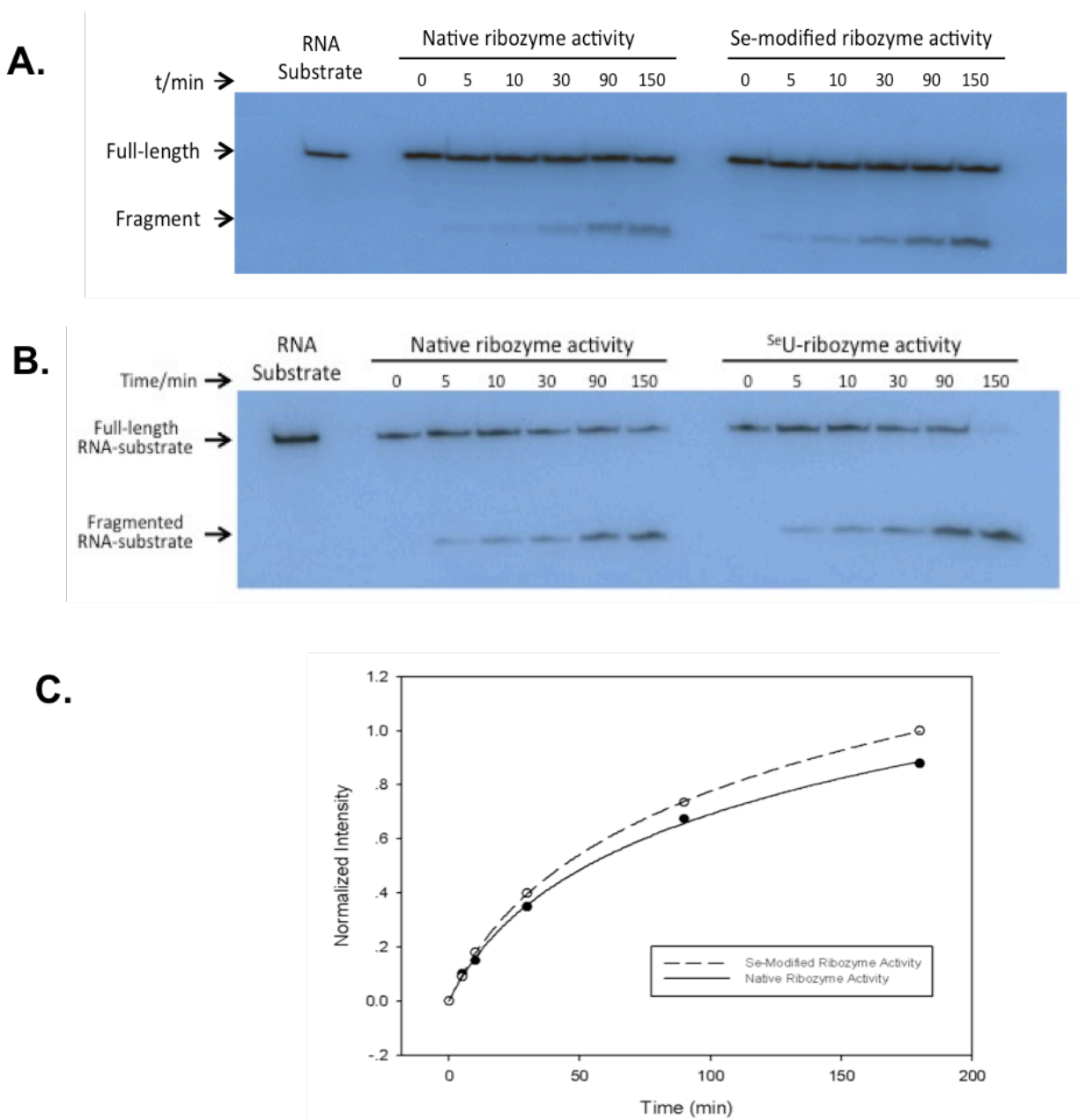
Figure 3.2.6. The catalytic activity of the Se-modified ribozyme.
(A) The time-course experiment of the 5'-^{32}P-RNA substrate digested with the non-self-cleaving native and Se-modified hammerhead ribozymes under the same conditions. The experiment was carried out at room temperature, with 10 mM Mg^{2+}, in the ribozyme buffer. (B) Time-course experiment of (A) with different 5'-^{32}P-RNA substrate concentration. (C) Plot of the ^{Se}U-ribozyme catalysis (dashed line) compared with the corresponding native (solid line). The cleavages of the RNA substrate by the native and Se-modified ribozymes (y-axis) were normalized via comparison to the substrate cleavage by the native ribozyme at 150 min (defined as 1.0).

3.2.4   Thermostability of the [Se]U-RNA

To examine the thermostability of the [Se]U-RNA, we designed a short Se-RNA (trimer: 5'-U[Se]UU-3') for this study. This Se-RNA was chemically synthesized by solid-phase synthesis and purified.[9a] We heated the Se-RNA continually at 70°C for a few hours and monitored it by HPLC at both 260 and 307 nm, since the 2-selenium modification has a unique UV-absorption at 307 nm, while the native nucleotides absorb strongly at 260 nm. The HPLC analysis was performed (Ultimate XB-C18, 250 mm × 4.6 mm, 5 μm) with a gradient from 100% buffer A (20 mM triethylammonium acetate in water) to 40% buffer B (20 mM trie- thylammonium acetate in 50% acetonitrile and 50% water) for 15 min. No significant decomposition was observed over 4-h heating at 70°C (Figure 3.2.7), indicating that this Se-modification is relatively stable.
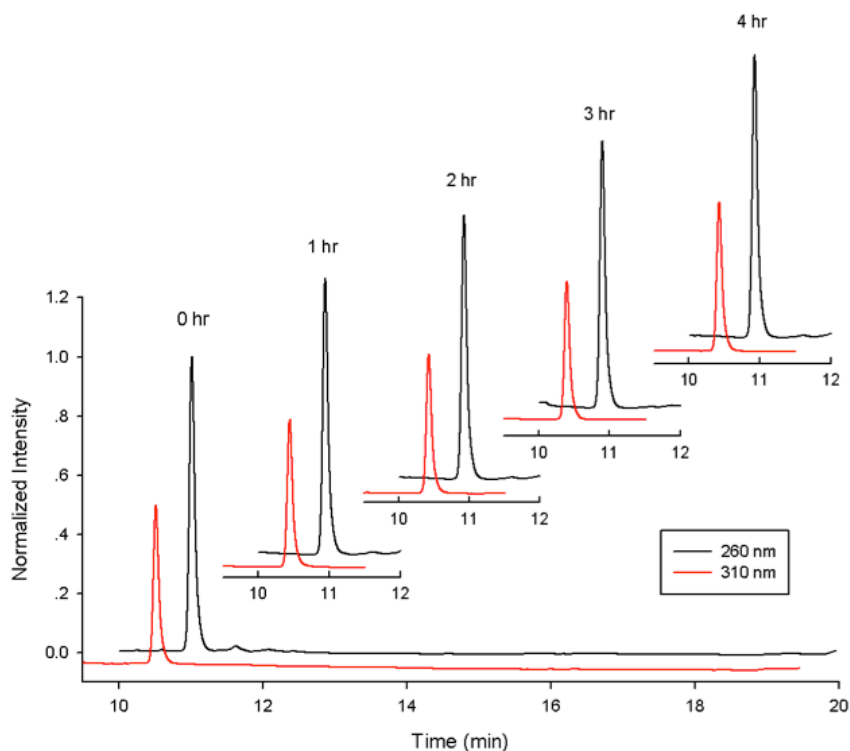

Figure 3.2.7. Thermostability study of [Se]U-RNA.
5'-U[Se]UU-3' was heated at 70°C for several hours. HPLC was monitored at both 260 and 307 nm (retention time: 10.9 min).

4 SYNTHESIS AND TRANSCRIPTION OF COLORED 4-SELENOURIDINE TRIPHOSPHATE WITH A SINGLE ATOM SUBSTITUTION

4.1 Introduction

4.1.1 Se-modified RNA in nature

RNA is essential biological molecule that performs critical functions in genetic information storage, transcription, protein synthesis and regulation. [2a, 27b] The uniqueness of RNA is greatly appreciated by the scientific society for its diversified structures and functions and its extensive applications in nucleic acids-protein studies as well as therapeutic discoveries. [1b, 2b] Although RNA research areas are very activate worldwide, the comprehensive structure and function of this biomolecule are not fully understood due to its complexity and often times due to current technique limitations. Therefore, enormous artificial RNA modifications have been developed to improve their chemical properties, to diversify their functionality and to increase their stability and fidelity. There are over one hundred naturally occurring RNA modifications have been discovered to date.[5] Most of the modifications exist in tRNA including a selenium-modified nucleobase - 2-selenouridine. This selenium-modified uridine occurs at the wobble position of the anticodon loop in several bacterial tRNAs (*Escherichia coli, Clostridium sticklandii, Methanococcus vannielii, etc*.),[5, 18] and its functionality has been fully characterized recently by Sun *et al*. via chemical synthesis.[9a] The experimental date has demonstrated that the Se-modification enhanced base pair fidelity by stabilizing the U/A base pair meanwhile discouraging the U/G mismatch without causing significant perturbation to the RNA structure. Moreover, the 2-Se-uridine triphosphate is recognizable by polymerase and Se-ribozyme is active.
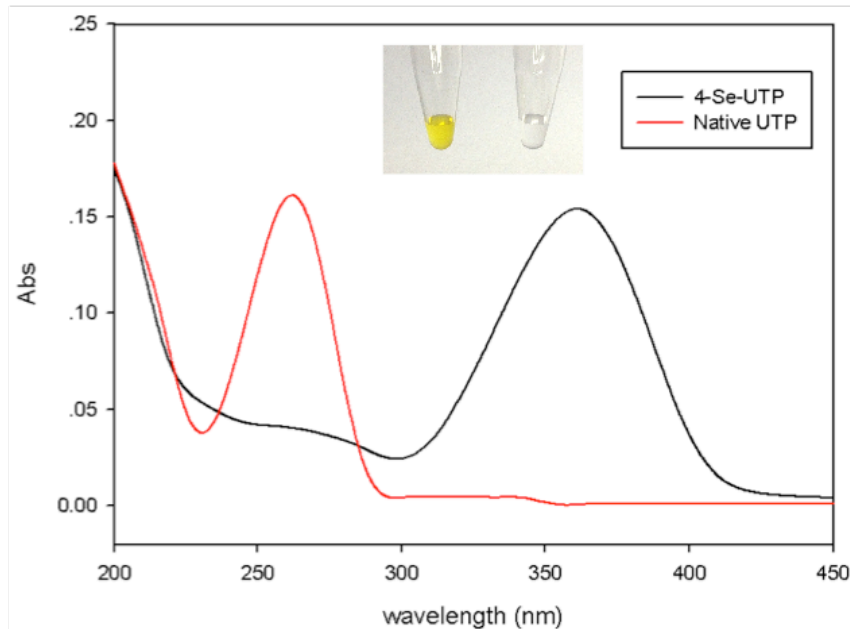
4.1.2 4-Selenouridine

Figure 4.1.1. The UV spectrum of native UTP and [4Se]UTP.
Native (red, λ_{max} = 260 nm) and colored [4Se]UTP (black. λ_{max} = 365 nm). Inset: left: UTP (colorless); right: [4Se]UTP (yellow).

Our research lab has previously replaced the oxygen atom at position 4 of thymidine with selenium.[51a] 4-selenium uridine nucleoside has been synthesized over decades ago; however, it has been incorporated into RNA oligonucleotide only last year through solid phase synthesis.[22] The enzymatic recognition of 4-seleno-uridne is unknown due to synthetic challenges of triphosphates. Compare with 2-selenouridine, the 4-selenouridine and possesses a unique yellow color with a UV absorption of 365 nm (Figure 4.1.1). This property is extremely useful for RNA visualization, detection, as well as spectroscopic study and crystallography of RNAs and protein-RNA complexes and interactions. Compare to other bulky molecules for RNA visualization, our new method only replace a single atom of the nucleobase to achieve such advancement. In addition, heavy atom such as selenium is a suitable anomalous scattering center for multi-wavelength anomalous dispersion (MAD) or single-wavelength anomalous dispersion (SAD) in

protein and nucleic acid crystallography. Thus, this C=Se functionality provides another advantage of seleno-modified nucleic acid research in X-ray crystallography.

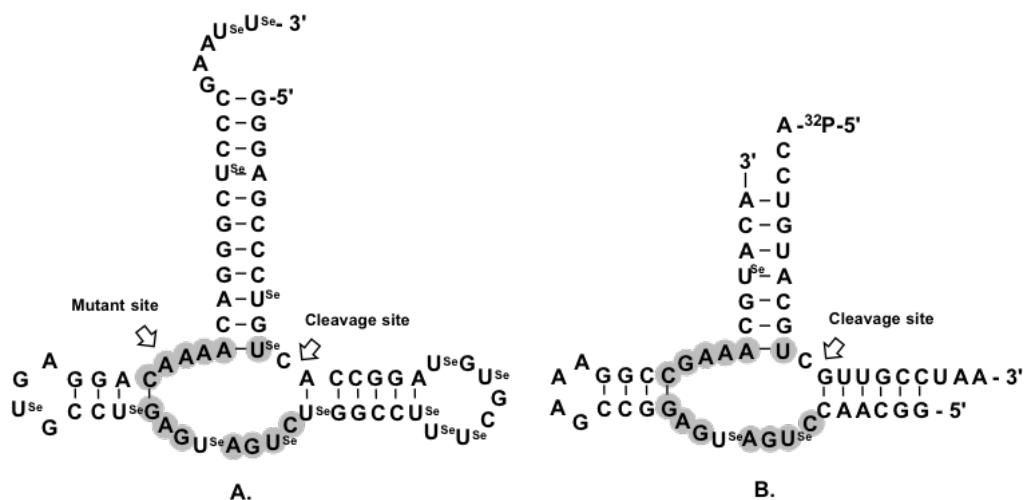### 4.1.3 ^{Se}U-RNA synthesis

Figure 4.1.2. Hammerhead ribozyme.
A) Secondary structure of the non self-cleaving (mutant) ^{Se}U-hammerhead ribozymes, including the wild-type (WHR) and crippled mutant (MHR). The mutant site and cleavage site are indicated by arrows. Highly conserved bases are highlighted in grey. B) Secondary structure of active ^{Se}U-hammerhead ribozyme with 5'-^{32}P-labled RNA substrate.

The two strategies to synthesize the Se-modified RNAs include solid-phase synthesis and transcription. The solid-phase synthesis method utilizes site-specific incorporation of the Se-nucleoside phosphoramidite. This method is applied to large-scale synthesis but limited to relatively short RNAs (up to 50 nt.) due to technical issues. In addition, it requires multiple steps in deprotection and purification. The chemical incorporation of 4-selenouridine into RNAs has been achieved only last year.[22] The enzymatic method on the other hand can allow synthesis of longer RNAs (>50 nt.) in a large quantity (multiple milligrams). In addition, this method can easily achieve multiple selenium atoms incorporation into RNA under the mild conditions. In order to incorporate the 4-Se-uridine into RNA to further investigate the function and structure of

the [Se]U-RNAs by in vitro transcription, 4-selenouridine triphosphate need to be synthesized first. Herein we report the first synthesis of 4-selenouridine triphosphate ([Se]UTP) and the enzymatic incorporation of [Se]UTP into non-coding RNAs. Both active and mutant hammerhead ribozymes (Figure 4.1.2) were successfully transcribed and examined with [Se]UTP. The transcribed SeU-hammerhead ribozyme is active, suggesting that the [Se]U-RNAs are useful in both function and structure studies of non-coding RNAs.

4.2   General Experiment Section

4.2.1   4-Selenouridine triphosphate synthesis

Although the synthesis of 4-Se-uridine nucleoside has been achieved many years ago,[54] it was only recently incorporated into RNA oligonucleotide by our laboratory[22] through 4-Se-uridine phosphoramidite synthesis. This solid-phase synthetic method is able to incorporate modified nucleobase into specific site of RNA with up to 50 nucleotides long. To obtain a longer modified RNA with an efficient and mild approach, herein we report the first synthesis of 4-selenouridine triphosphate and its incorporation into longer RNA via *in vitro* transcription. To minimize the by-product formation, we protect the selenium atom during the chemical synthesis with cyanoethyl group.[9a, 20-22, 51] Thus the synthesis (Scheme 4.2.1) of [4Se]UTP (**5**) started with the activation of the commercial available uridine-nucleoside **1** with 2,4,6-triisopropylbenzenesulfonyl chloride (TIBS-Cl) and 4-dimethylaminopyridine (DMAP) in tetrahydrofuran (THF) at position 4. Then, without purification, a sodium selenide solution ($NCCH_2CH_2SeNa$) generated by di(2-cyanoethyl) diselenide [$(NCCH_2CH_2Se)_2$] and $NaBH_4$ in ethanol was slowly injected into reaction to obtain compound **2**.[20] Compound **2** was deprotected

by triethylamine trihydrofluoride in THF at 40°C to obtain nucleoside **3**. Compound **3** was treated with 4% trifluoroacetic acid in dichloromethane to achieve nucleoside **4**. Via a one-pot synthesis, compound **4** was converted to protected 4-selenouridine triphosphate ([4SeCH2CH2CN]UTP, compound **5**) by treating with phosphorus oxychloride (POCl$_3$), pyrophosphate, and bicarbonate sequentially.[19, 51a, 52]
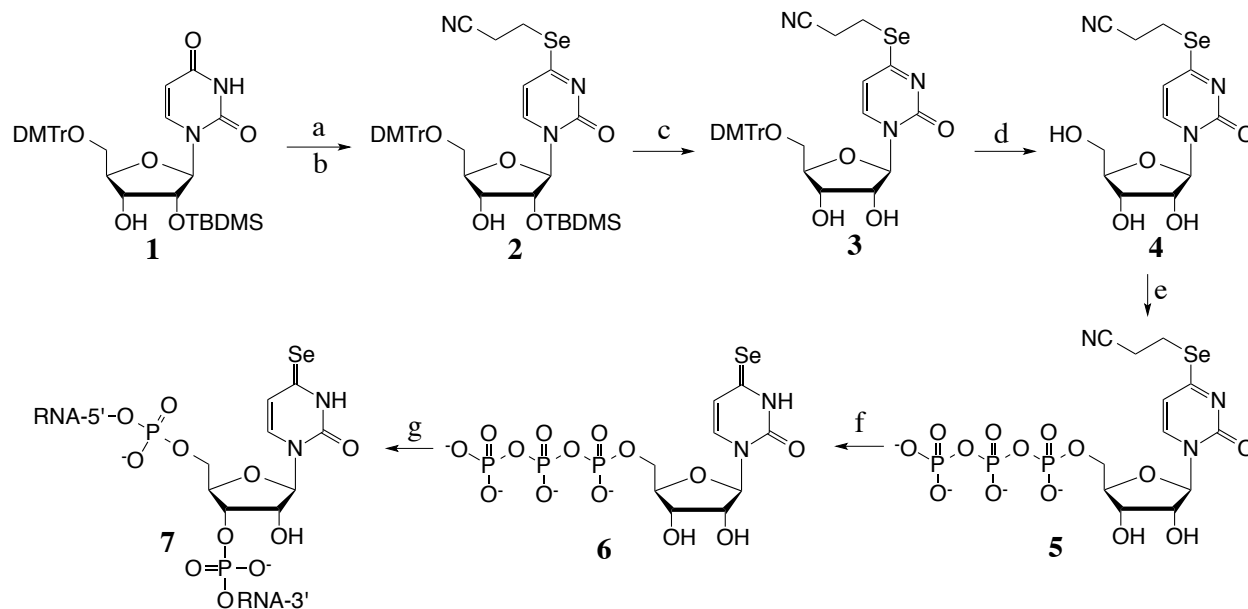
Scheme 4.2.1. Chemical synthesis of [4Se]UTP (5) and transcription of [4Se]U-containing RNA. a) TIBSCl, DMAP, THF; b) (NCCH$_2$CH$_2$Se)$_2$, NaBH$_4$, EtOH; c) Triethylamine trihydrofluoride, THF, 40°C; d) 4% trifluoroacetic acid, CH$_2$Cl$_2$; e) POCl$_3$, Me$_3$PO$_4$; (tri-n-butyl)amine, pyrophosphate, N, N-dimethylformamide; the H$_2$O hydrolysis; f) K$_2$CO$_3$ (0.05 M) in methanol; g) RNA transcription.

4.2.1.1 Synthesis of 4-selenouridine

The starting material **1** (0.2 g, 0.37 mmol) and catalytic amount of 4,4'-dimethylaminopyridine (DMAP, 3 mg) was dissolved in anhydrous THF under argon, followed by adding diisopropylethylamine (DIPEA, 1.85 mmol) and the reaction was stirred at room temperature. Then a solution of 2,4,6-trisopropylbenzenessulfonyl chloride (TIPCl, 0.56 mmol) pre-dissolved in THF was added into reaction dropwisely. The reaction was stirred for 1 hour

and monitored by TLC plate (5% methanol in dichloromethane). Without further purification, the reaction mixture was slowly added into a clear solution of sodium selenide ($NCCH_2CH_2SeNa$) pre-generated by adding ethanol (EtOH) into sodium borohydride ($NaBH_4$, 2.2 mmol) and di(2-cyanoethyl) diselenide [$(NCCH_2CH_2Se)_2$, 1.83 mmol]. The reaction was stirred for another 1 hour and monitored by TLC plate (5% methanol in dichloromethane, Rf = 0.6). After the reaction was complete, the crude compound **2** was dissolved in ethyl acetate and wash with saturated sodium chloride solution. The organic layer was then separated, dried over magnesium sulfate and evaporated into dryness. The compound was purified by flash column chromatography to obtain a pure slight yellow foam compound.[55] Then the pure compound **2** (0.1 g, 0.15 mmol) was dissolved in anhydrous THF and triethylamine trihydrofluoride (0.3 mmol) was added into reaction at 40ºC. The reaction was stirred for 2 hours and monitored by TLC plate (7% methanol in dichloromethane). Once the reaction was complete, the crude reaction mixture was dried under reduced pressure and re-dissolved in dichloromethane. Later, 4% trifluoroacetic acid was added drop-wisely into the reaction until the pH reach 4. After the reaction was complete, methanol (0.2 mL) was injected into the mixture and the organic layer was washed by water twice, then isolated, dried over magnesium sulfate and evaporated under reduced pressure. The crude compound **4** was purified by flash column chromatography and characterized NMR (^{1}H- and ^{13}C-NMR) and ESI-TOF analyses. Compound **4**: ^{1}H NMR (400 MHz, $CDCl_3$) δ: 8.09 (d, J = 7.0 Hz, 1H, H-6), 7.42 – 7.17 (m, 10H, aromatic and N-H), 6.83 (m, 4H, aromatic), 6.09 (d, J = 7.0 Hz, 1H, H-5), 5.83 (d, J = 2.2 Hz, 1H, H-1′), 5.71 – 5.51 (br, 1H, 2'-OH), 4.53 – 4.25 (m, 3H, H-3′, H-4′, 3′-OH), 3.80 (s, 6H, OMe), 3.58 – 3.41 (m, 3H, H-5′, H-2′), 3.37 (t, J = 6.7 Hz, 2H, C$\underline{H}_2$-CN), 2.97 (t, J = 6.6 Hz, 2H, C$\underline{H}_2$-Se). 13C NMR (101 MHz, CDCl3) δ: 174.71 (C-4), 153.73 (C-2), 139.23 (C-6) 157.70,143.20, 139.23, 134.30, 134.10,

129.08, 127.09, 127.03, 126.17, 112.33 (Ar), 117.84 (CN), 105.79 (C-5), 91.95 (C-1′), 86.04 (C-Ar), 84.10 (C-4′), 75.33 (C-2′), 69.62 (C-3′), 61.11 (C-5'), 54.30 (OCH$_3$), 19.61 (CH$_2$CH$_2$CN), 17.91(CH$_2$-CN). Compound **3**: ^{1}H NMR (400 MHz, MeOD) δ: 8.32 (d, J = 7.0 Hz, 1H, H-5), 6.56 (d, J = 7.0 Hz, 1H, H-6), 5.80 (s, 1H, H-1′), 4.14 (m, 3H, H-2′, H-3′, H-4′), 3.98 (dd, 1H, H5′), 3.79 (dd, 1H, H5″), 3.38 (m, 2H, CH$_2$-CN), 3.03 (m, 2H, CH$_2$-Se). ^{13}C NMR (101 MHz, MeOD) δ: 175.15 (C-4), 153.57 (C-2), 139.94 (C-6), 117.85 (CN), 105.83 (C-5), 91.03 (C-1′), 83.54 (C-4′), 74.16 (C-2′), 67.39 (C-3′), 58.83 (C-5′), 19.19 (CH$_2$CH$_2$CN), 17.36 (CH$_2$-CN).

4.2.1.2 Synthesis of 4-selenouridine triphosphate

Protected 4-Se-uridine nucleoside (**4**, 20 mg), tributylammonium pyrophosphate (2 eq.) and proton-sponge (2 eq.) were weighted and dried in individual flasks under high vacuum for 3 hours and then filled with argon gas. Trimethyl phosphate (0.4 mL) was added into the flask that containing compound **4** and the flask was stirred in an ice bath. Later a solution of proton-sponge dissolved in trimethyl phosphate (0.3 mL) was injected into the solution of **4** at 0 °C. After 10 min stirring, a pre-diluted phosphorus oxychloride (POCl$_3$; 9 mL, 2 eq.) in trimethyl phosphate (90 mL) was added dropwisely into reaction mixture at 0 °C. The reaction was monitored by TLC plate (isopropanol: ammonium hydroxide: water; v 5:3:2) and was completed in 2 hours. Then tributylammonium pyrophosphate (2 eq., dissolved in 0.2 mL tributylamine and 0.4 mL DMF) was fast injected into the reaction mixture at 0°C and allows the reaction vigorously to stir for 5 min. Later the reaction was quenched with water (3 mL) and stirred for another 1 hr at the room temperature and monitored by TLC plate. To obtain crude compound **5**, a sodium chloride (NaCl) solution (3 M NaCl, 0.5 mL) was added to the reaction flask, followed by adding pure ethanol (14.5 mL) and freezing the suspension at -80 °C for 1 hr to allow the crude product to

precipitate. Crude compound **5** was recovered by centrifugation for 20 min at 14,000 rpm in a falcon tube (50 mL). The ^{4SeCH2CH2CN}UTP (**5**) pellet was re-dissolved in water, then analyzed and purified by HPLC. Pure ^{4SeCH2CH2CN}UTP (**5**) was later characterized by NMR (^{1}H-, ^{13}C- and ^{31}P-NMR) and ESI-TOF analyses. ^{1}H-NMR. The deprotection of compound **5** was carried out in a potassium carbonate (K_2CO_3) solution in methanol (6 eq.) at room temperature and monitored by TLC plate. After the reaction was complete, a NaCl/EtOH precipitation (described previously) was performed to obtain the compound **6** (strong yellow color). Later the pure ^{4Se}UTP (**6**) was characterized by NMR (^{1}H-, ^{13}C- and ^{31}P-NMR) and ESI-TOF analyses. Compound **5**: ^{1}H NMR (400 MHz, D_2O) δ 8.11 (d, J = 7.1 Hz, 1H, H-5), 6.85 (d, J = 7.1 Hz, 1H, H-6), 5.85 (s, 1H, H-1′), 4.45 – 3.97 (m, 5H, H-2′, H-3′, H-4′, H-5′), 3.34 (t, J = 6.8 Hz, 2H, C$\underline{H}_2$-CN), 3.12 – 2.92 (m, 2H, C$\underline{H}_2$-Se). ^{13}C NMR (101 MHz, MeOD) δ: 177.04 (C-4), 154.40 (C-2), 139.75 (C-6), 119.56 (CN), 107.60 (C-5), 89.56 (C-1′), 81.73 (C-4′), 73.77 (C-2′), 67.34 (C-3′), 63.08 (C-5′), 18.81 ($\underline{C}H_2CH_2CN$), 17.31 ($\underline{C}H_2$-CN). ^{31}P NMR (162 MHz, D2O) δ -10.65 (d, J = 19.8 Hz), -11.51 (d, J = 20.1 Hz), -23.28 (t, J = 20.0 Hz). Compound **6**: ^{1}H NMR (400 MHz, D2O) δ 7.85 (d, J = 7.4 Hz, 1H, H-5), 6.87 (d, J = 7.4 Hz, 1H, H-6), 5.81 (d, J = 3.7 Hz, 1H, H-1′), 4.31 – 4.20 (m, 5H , H-2′, H-3′, H-4′, H-5′). 13C NMR (101 MHz, D_2O) δ 191.58 (C-4), 154.15 (C-2), 134.45 (C-6), 117.10 (C-5), 88.96 (C-1′), 81.69 (C-4′), 73.31 (C-2′), 67.90 (C-3′), 56.43 (C-5′). ^{31}P NMR (162 MHz, D2O) δ -9.99 (d, J = 19.3 Hz), -11.54 (d, J = 20.2 Hz), -23.15 (t, J = 20.1 Hz).

4.2.1.3 HPLC and UV analyses of ^{4SeCH2CH2CN}UTP and ^{4Se}UTP

The crude ^{4SeCH2CH2CN}UTP (**4**) was precipitated from reaction mixture after synthesis and directly purified by reverse-phase high-performance liquid chromatography (RP-HPLC). The purified ^{4SeCH2CH2CN}UTP was characterized by NMR (^{1}H, ^{13}C and ^{31}P), MS, HPLC and UV (Figure

4.1.1). For the preservative purposes, the major potion of 4-selenouridine-triphosphate is kept in the protected form (compound **4**). Before transcription, ^{4SeCH2CH2CN}UTP (**4**) was treated with K_2CO_3 (0.05 M in methanol) and was later precipitated with NaCl (3M) and ethanol, the pure ^{4Se}UTP is obtained. The maximal UV absorbance of native uridine triphosphate is 260 nm, the maximal UV absorbance of the ^{4Se}U-triphosphate is 306 nm and the maximal UV absorbance of the ^{4Se}U-triphosphate is 365 nm and the compound itself is strong yellow color (Figure 4.1.1). In the HPLC analysis, the native and selenium-modified UTPs were monitored with three wavelengths (260, 310 and 360 nm with a buffer gradient of 100% buffer A (20 mM triethylammonium acetate in water) to 25% buffer B (20 mM triethylammonium acetate in 50% acetonitrile and 50% water) in 20 min. The HPLC and UV profiles are shown in Figure 4.1.1 and Figure 4.2.1. The retention times of the native UTP, ^{4SeCH2CH2CN}UTP and ^{4Se}UTP were 11.3, 17.8 and 14.7 min, respectively.
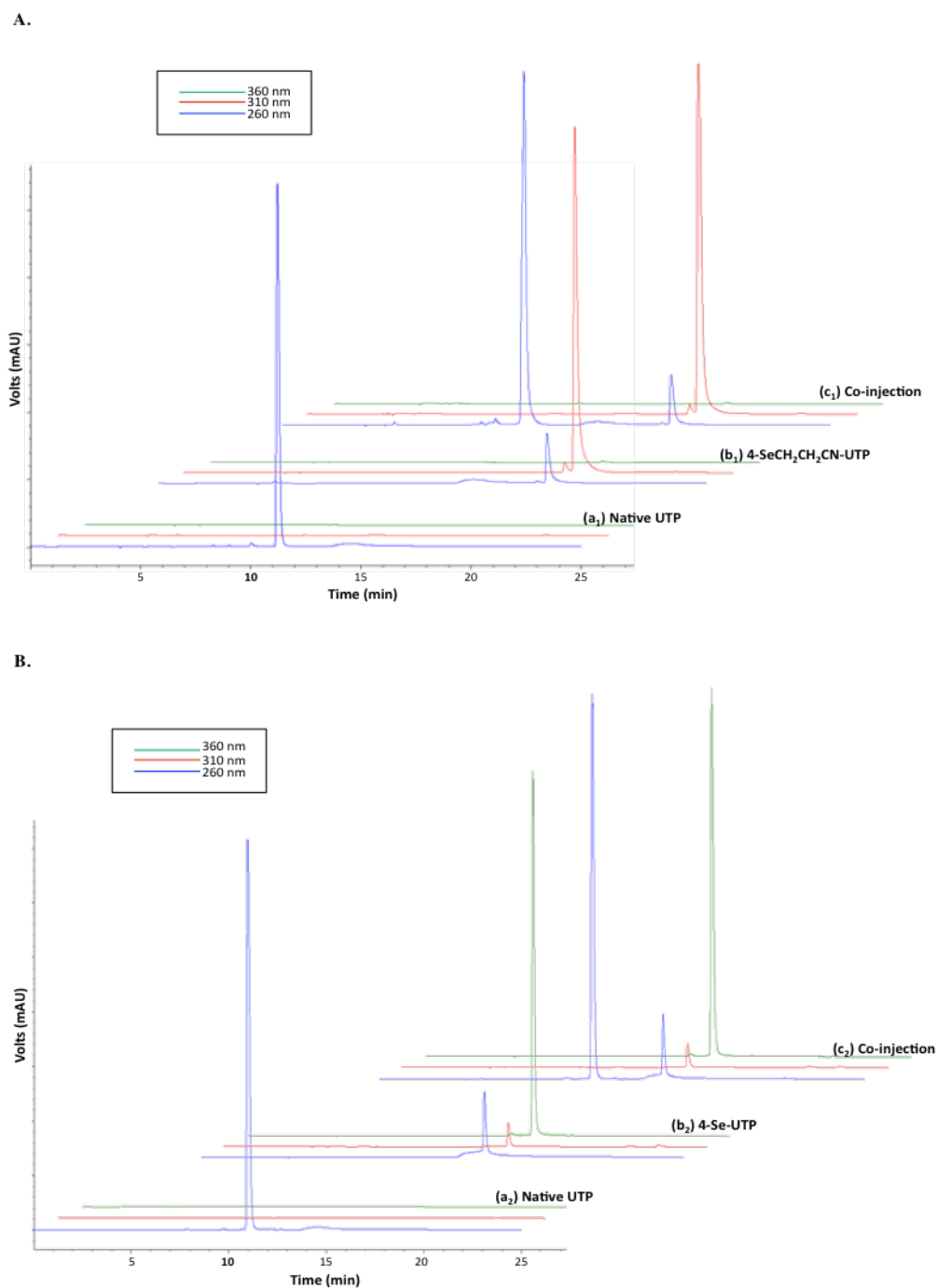
Figure 4.2.1. HPLC analyses of ^{4SeCH2CH2CN}UTP and ^{4Se}UTP at multiwavelength (260 nm, blue; 310 nm, red; 360 nm, green).
A) a_1: Native UTP (retention time: 11.3 min); b_1: ^{4SeCH2CH2CN}UTP (retention time: 17.8 min); c_1: co-injection of both native UTP and ^{4SeCH2CH2CN}UTP (retention time: 11.3 min and 17.8 min). B) a_2: Native UTP (retention time: 11.3 min); b_2: ^{4Se}UTP (retention time: 14.7 min); c_2: co-injection of both native UTP and ^{4Se}UTP (retention time: 11.3 min and 14.7 min).

4.2.2 4-Se-RNAs transcription
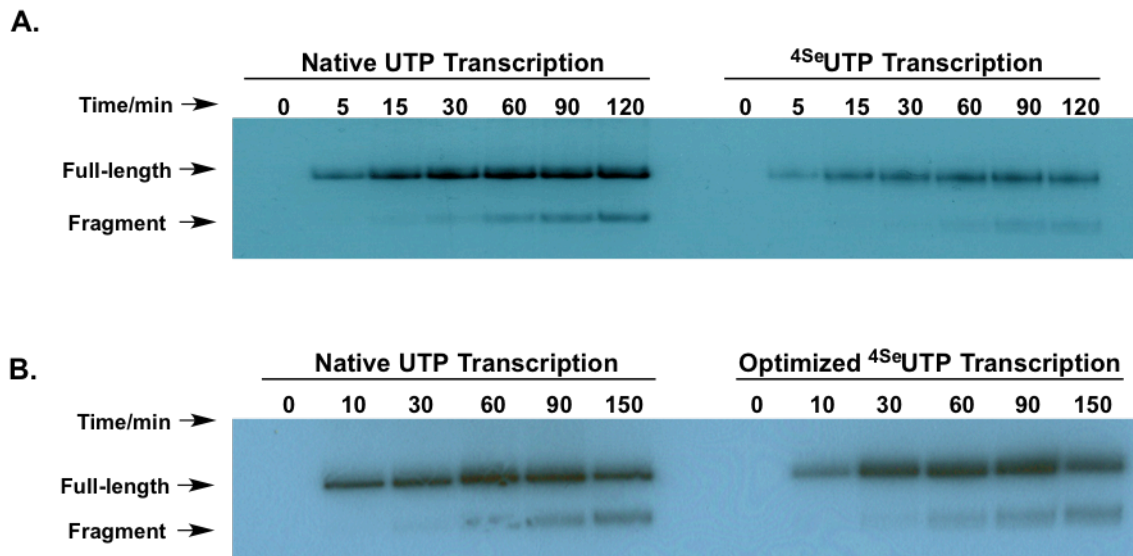
A.

B.

Figure 4.2.2. The ribozyme time course experiments.
A) The auto-radiography gel imagine of *in vitro* transcription with both native and ^{4Se}U-RNA under same experimental conditions; (left): transcription of the native RNA (mutant hammerhead-ribozyme: MHR) with all native NTPs; the minor faster-moving band is the self-cleaved product (fragment); (right): transcription of the ^{4Se}U-MHR with ^{Se}UTP and other native NTPs. B) *In vitro* transcription of ^{4Se}U-RNA under optimized conditions; (left): transcription of the native RNA with all native NTPs under standard condition; (right): transcription of the ^{4Se}U-MHR with ^{Se}UTP and other native NTPs under optimized conditions.

To examine the ^{4Se}UTP compatibility with RNA polymerase in transcription, the linearized plasmid templates for the crippled mutant hammerhead-ribozyme (MHR) were used (Figure 4.1.2). The transcription result shows that T7 RNA polymerase can recognize ^{4Se}UTP (Figure 4.2.2A) and the transcript RNA contains 15 selenium atoms incorporation. Under the same experimental conditions, the transcription of ^{4Se}UTP yields less product compare to the native one. To increase the ^{4Se}U-RNA transcription yield, series of optimization experiments were carried out including buffer pH adjustment and ^{4Se}UTP concentration alternation (Figure 4.2.5). From the result, we have observed that with higher ^{4Se}U-RNA concentration (4 times higher than native UTP) at pH 7.5, the transcription yield of Se-modified RNA is comparable to

the corresponding native RNA. The time-course experiments of both the native and ^{4Se}U-modified mutant ribozymes were performed using the mutant hammerhead-ribozyme template. The experiments are carried out under the same transcription condition as well as optimized transcription conditions for comparison (Figure 4.2.2). Although the transcript hammerhead ribozyme are mutant, partially self-cleaved fragment was still observed (minor faster-moving band). Detailed experimental condition was discussed in materials and methods. This result indicates that in the enzymatic catalysis, ^{4Se}UTP does not cause significant interference.

4.2.2.1 *Transcription analysis of the 4-Se-RNAs*

The transcription experiment was carried out by following the standard procedures from the manufacturer, Epicentre (AmpliScribe™ T7-Flash™ Transcription Kit). α-^{32}P-ATP was used as the radioactive labeling material for transcription experiments. Each transcription reaction (5 µL) contained ATP, CTP, GTP and UTP (0.5 mM each) or ^{4Se}UTP (2 mM for optimization), linearized plasmid DNA template (50 ng/µL), DTT (10 mM), transcription buffer (1x) for T7 RNA polymerase, T7 RNA polymerase (10 U), and RNase-free water. At each time point of the time-course experiments, a gel loading dye (5 µL) containing 100 mM EDTA was added to quench the reaction, later the experiment was analyzed by denaturing PAGE (15% gel) and autoradiography. The translated RNAs were MHR (Figure 4.1.2). To conform the integrity of modified RNA, we transcribed an active self-cleavage RNA by using a linearized plasmid template of wild-type hammerhead ribozyme (WHR, Figure 4.1.2). This transcribed RNA contained 13 selenium atoms and unlike the MHR RNA, this WHR provides a clean cleaved RNA product (Figure 4.2.3A), the integrity of WHR was confirmed by MALDI-TOF MS analysis (Figure 4.2.3B).
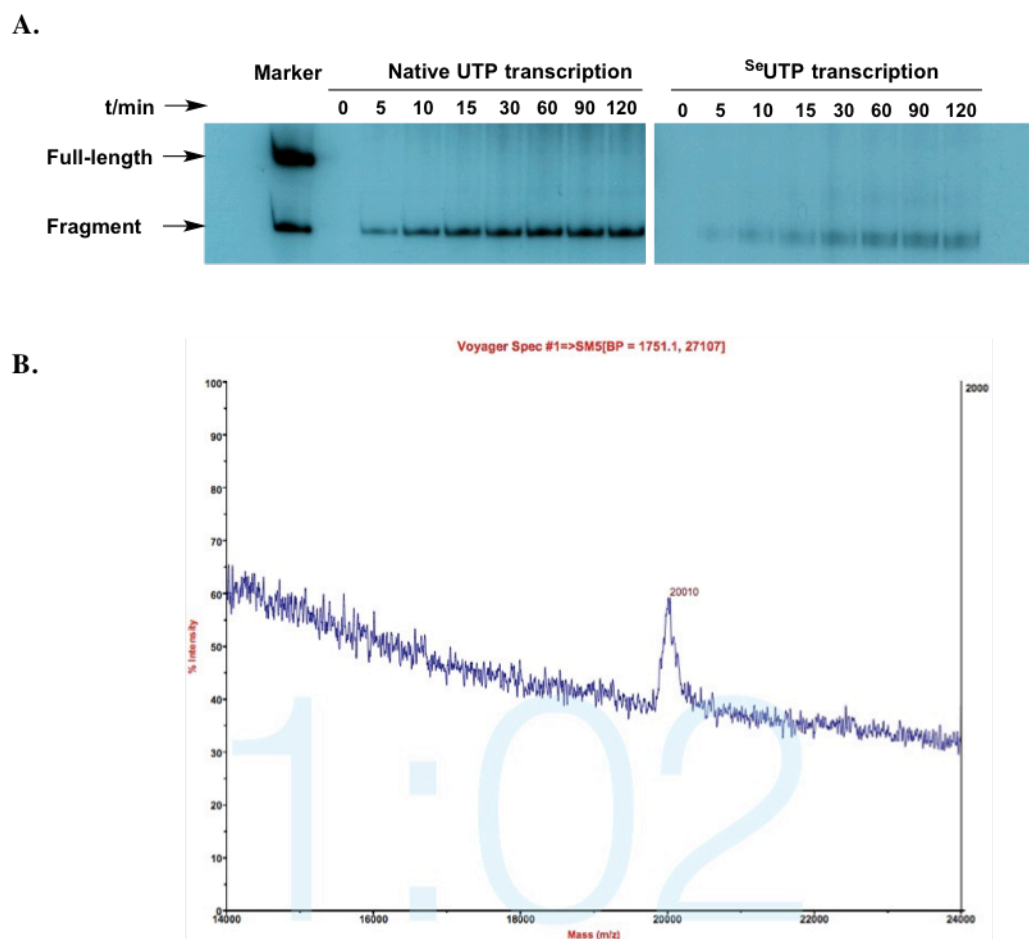
A.

B.

Figure 4.2.3. Wild-type ribozyme transcription.
(A)Wild-type native and 4-Se-modified ribozyme transcription with self-cleavage activity during synthesis. (B) MALDI-TOF MS analysis of the ^{Se}U-WHR (molecular formula: $C_{543}H_{672}N_{218}O_{385}P_{56}Se_{13}$); matrix: 3-hydroxypicolinic acid (3HPA, molecular formula: $C_6H_5NO_3$); mass of SeU-WHR and six matrix observed: 20010 (calc. 20010.6).

4.2.2.2 pH titration curve of 4-selenouridine

The 4-selenouridine solutions were adjusted to desired pH values in the buffer of 50 mM Na_2HPO_4 at room temperature. The UV–Vis spectra were recorded every 0.1 pH unit between pH 6–8 and every 0.2-0.5 pH unit between pH 4–6 and pH 8–10. The pH of each solution was measured before and after its UV–Vis spectrum collection and the error was within ±0.02 pH unit. The titration data was plotted and shown in Figure 4.2.4.
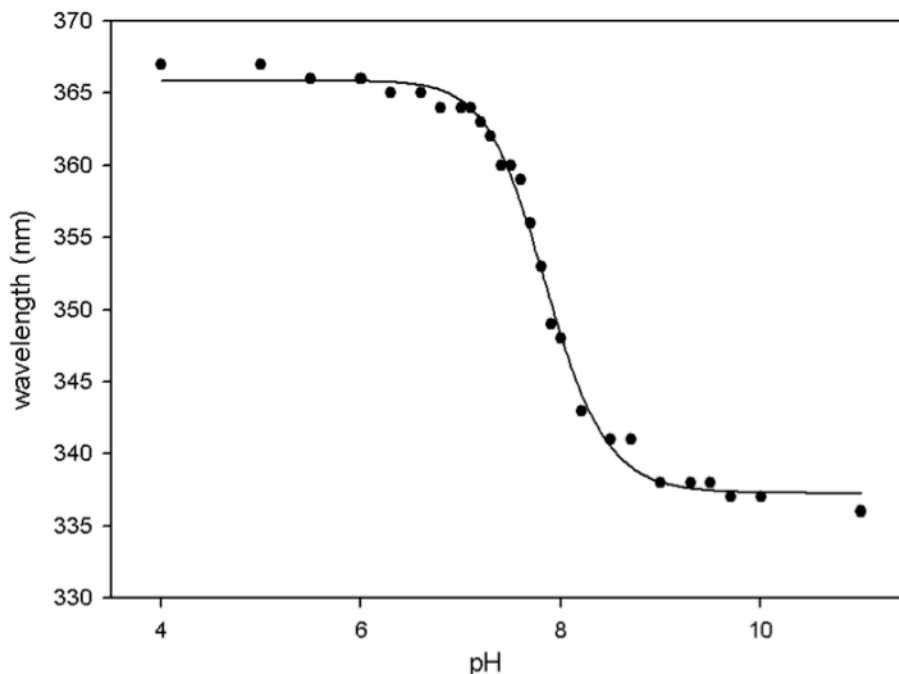
Figure 4.2.4. Plot of wavelength (nm) versus pH for 4-selenouridine triphosphate. The fitted titration curve yields the pK_a value (7.85±0.02).

4.2.2.3 Transcription optimization of ^{4Se}U-RNA

With standard transcription condition, the transcription yield of ^{4Se}U-RNA is lower than native RNA (Figure 4.2.2A). To reach the native transcription level, optimization experiments were carried out under different conditions. In the optimization experiments, we chose the linearized plasmid template to transcibe mutant hammerhead ribozyme (Figure 4.1.2A) that incorporates fifteen ^{4Se}UTPs. We examined the transcription with different buffer pH values first since the acidity of the imino group (3-NH) of ^{4Se}UTP varies with the selenium modification (pK_a = 7.85, Figure 4.2.4). Eight transcription buffer pH were tested (pH 5.5, 6.0, 6.5, 7.0, 7.5, 8.0, 8.5 and 9.0), the buffer pH was adjusted by adding concentrated HCl into the solution, the buffer solution also includes 40 mM tris base, 6 mM $MgCl_2$, 2 mM spermidine and 10 mM DTT. The best yield comes from buffer pH 7.5 (Figure 4.2.5A), the standard native buffer condition is pH

7.9. Under this optimized condition (buffer pH 7.5), higher concentrations of ^{4Se}UTP were also examined and the transcription yield reached native transcription level when four times concentration of ^{4Se}UTP was applied (Figure 4.2.5B). Later, this optimized condition with transcription buffer pH 7.5 and higher ^{4Se}UTP concentration (4X) was applied to optimized transcription time course experiment (Figure 4.2.2B) and catalytic activity studies (Figure 4.2.6).
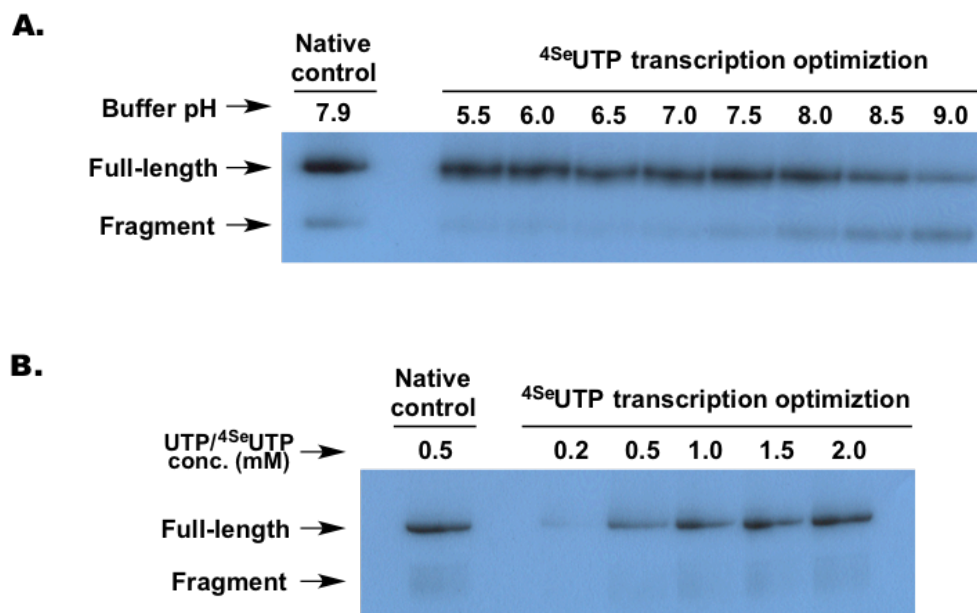

Figure 4.2.5. Experimental results of transcription optimizations with ^{4Se}UTP. A) Optimization with different transcription buffer pH (5.5 to 9.0), standard transcription buffer is pH 7.9. B) Optimization with different ^{4Se}UTP concentration (0.2 mM to 2.0 mM), UTP concentration in native control is 0.5 mM.

4.2.2.4 Catalytic activity analysis of the Se-RNAs

The active hammerhead ribozymes with the sequence of 5'-GGCAACCUGAUGAGGCCGAAAGGCCGAAACGUACA-3' (Figure 4.1.2) is used in the catalytic experiments. The transcription of both native and 4-Se-RNA used in this experiment containing ATP, CTP, GTP and UTP (2.5 mM each) or ^{4Se}UTP (10 mM for optimization), linearized plasmid DNA template (50 ng/μL), DTT (10 mM), transcription buffer (1x) for T7

RNA polymerase, T7 RNA polymerase (20 U), and RNase-free water. The template used in this transcription was a 55-nt long DNA duplex synthesized through solid-phase synthesis (5'-TGTACGTTTCGGCCTTTCGGCCTCATCAGGTTGCCTATAGTGAGTCGTATTACGC-3' and its complimentary sequence). The transcription reaction was performed at 37°C for 2 hours. After transcription, the ribozymes were isolated from the template and primer. To compare catalysis activity, both the native and 4-Se-ribozymes were adjusted to same concentration (measured by UV). For ribozyme digestion experiment, a RNA substrate (5'-ACCUGUACGUCGUUGCCUAA-3') was synthesized through solid-phase synthesis (Figure 4.1.2) and was purified. In order to monitor the transcribed ribozyme catalytic activity, the substrate was kinased with γ-^{32}P-ATP at 5' end by T4 polynucleotide kinase and the result was observed by gel analysis and autoradiography. After ribozyme digestion, the RNA substrate was cleaved and two fragments were obtained (11-nt and 9-nt in length respectively), but only the 5'-^{32}P-labled end (11-nt) was visible by autoradiography. The digestion time-course analysis was shown in Figure 4.2.6.
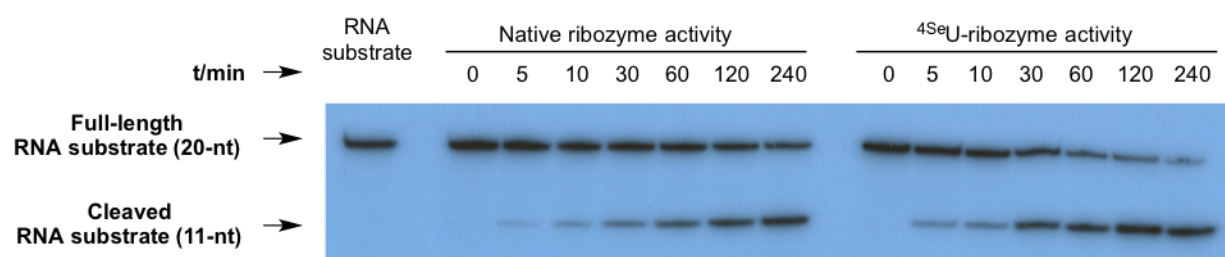
Figure 4.2.6. The transcription of the wild-type native and ^{4Se}U-modified ribozymes. Time-course of a 5'-^{32}P-labled RNA substrate digestion by non self-cleaving native and ^{4Se}U-modified ribozymes under same enzyme and substrate concentration. The experiments were carried out at room temperature with 10 mM Mg^{2+} concentration. The pure RNA substrate is used as the comparing marker.

5 CONCLUSIONS

In summary, the atom-specific mutagenesis has been extensively applied in RNA function and structure investigations, catalysis analysis, mechanism studies, as well as therapeutics discoveries. The great advantage of the single-atom replacements is that they may not only drastically improve beneficial properties of RNAs, such as thermostability and nuclease resistance, but also preserve RNA structure integrity without significant alteration. The atom-specific modifications have indeed become a very convenient and practical strategy in the fundamental research of nucleic acids, including structural and functional studies and drug development. The selenium modifications in nucleic acids focus on the facilitation of crystallization and phasing in X-ray crystallography for structure determination of nucleic acids, nucleic acid–protein complexes, and nucleic acids complexed with small molecules as well as metal ions. In addition to the crystal structure study, the selenium derivatization can facilitate function studies, drug discoveries, and material investigations.

We have first synthesized the ^{Se}U-phosphoramidite, ^{Se}U-triphosphate (^{2Se}UTP and ^{4Se}UTP) as well as ^{Se}U-RNAs. Our biophysical and structural studies on the ^{Se}U-RNAs indicate that the native and Se-modified structures are virtually identical. The 2-Se-modification can largely discriminate against the U/G wobble pair without significant impact on U/A pair, thereby providing a unique chemical strategy to further enhance base pair fidelity. The Se-modification will also provide a useful tool in X-ray crystal structure studies of RNAs and their protein complexes. Moreover, we have demonstrated that the synthesized ^{Se}UTPs (^{2Se}UTP and ^{4Se}UTP) are stable and recognizable by T7 RNA polymerase. Under the optimized conditions, the transcription yield of ^{Se}U-RNA can reach up to 85% of the corresponding native RNA. Furthermore, the transcribed ^{Se}U-hammerhead ribozyme has the similar activity as the

corresponding native, which suggests usefulness of ^{Se}U-RNAs in function and structure studies of noncoding RNAs, including the Se-tRNAs. The atom-specific mutagenesis with selenium opens a new research avenue for investigating base-pair recognition, fidelity and RNA modification. This novel base pair (^{Se}U/A) with higher specificity likely enables better preservation of genetic information at the RNA level.

REFERENCES

1.   (a) Sun, H.; Huang, Z., *Atom-specific mutagenesis of RNAs for structure, function and therapeutics studies*. John Wiley & Sons, Inc.: 2013; Vol. 11; (b) Schmeing, T. M.; Ramakrishnan, V., What recent ribosome structures have revealed about the mechanism of translation. *Nature* **2009,** *461* (7268), 1234-42; (c) Wahl, M. C.; Will, C. L.; Luhrmann, R., The spliceosome: design principles of a dynamic RNP machine. *Cell* **2009,** *136* (4), 701-18.

2.   (a) Serganov, A.; Patel, D. J., Ribozymes, riboswitches and beyond: regulation of gene expression without proteins. *Nature reviews. Genetics* **2007,** *8* (10), 776-90; (b) Ponting, C. P.; Oliver, P. L.; Reik, W., Evolution and functions of long noncoding RNAs. *Cell* **2009,** *136* (4), 629-41.

3.   Drude, I.; Dombos, V.; Vauleon, S.; Muller, S., Drugs made of RNA: development and application of engineered RNAs for gene therapy. *Mini reviews in medicinal chemistry* **2007,** *7* (9), 912-31.

4.   Wachowius, F.; Hobartner, C., Chemical RNA modifications for studies of RNA structure and dynamics. *Chembiochem : a European journal of chemical biology* **2010,** *11* (4), 469-80.

5.   Dunin-Horkawicz, S.; Czerwoniec, A.; Gajda, M. J.; Feder, M.; Grosjean, H.; Bujnicki, J. M., MODOMICS: a database of RNA modification pathways. *Nucleic acids research* **2006,** *34* (Database issue), D145-9.

6.   Muller, S.; Wolf, J.; Ivanov, S., Current strategies for the synthesis of RNA. *Current Organic Synthesis* **2004,** *1* (3), 293-307.

7.   Kumar, R. K.; Davis, D. R., Synthesis and studies on the effect of 2-thiouridine and 4-thiouridine on sugar conformation and RNA duplex stability. *Nucleic acids research* **1997,** *25* (6), 1272-80.

8.   (a) Pieken, W. A.; Olsen, D. B.; Benseler, F.; Aurup, H.; Eckstein, F., Kinetic characterization of ribonuclease-resistant 2'-modified hammerhead ribozymes. *Science* **1991,** *253* (5017), 314-317; (b) Cummins, L. L.; Owens, S. R.; Risen, L. M.; Lesnik, E. A.; Freier, S. M.; McGee, D.; Guinosso, C. J.; Cook, P. D., Characterization of fully 2'-modified oligoribonucleotide hetero- and homoduplex hybridization and nuclease sensitivity. *Nucleic acids research* **1995,** *23* (11), 2019-24.

9.   (a) Sun, H.; Sheng, J.; Hassan, A. E.; Jiang, S.; Gan, J.; Huang, Z., Novel RNA base pair with higher specificity using single selenium atom. *Nucleic acids research* **2012,** *40* (11), 5171-9; (b) Lin, L.; Sheng, J.; Huang, Z., Nucleic acid X-ray crystallography via direct selenium derivatization. *Chemical Society reviews* **2011,** *40* (9), 4591-602.

10.   (a) Cornish, P. V.; Ha, T., A survey of single-molecule techniques in chemical biology. *ACS chemical biology* **2007,** *2* (1), 53-61; (b) Motorin, Y.; Helm, M., RNA nucleotide methylation. *Wiley interdisciplinary reviews. RNA* **2011,** *2* (5), 611-31; (c) Rublack, N.; Nguyen, H.; Appel, B.; Springstubbe, D.; Strohbach, D.; Müller, S., Synthesis of specifically modified oligonucleotides for application in structural and functional analysis of RNA. *Journal of nucleic acids* **2011,** *2011*.

11.   Caton-Williams, J.; Huang, Z., Biochemistry of selenium-derivatized naturally occurring and unnatural nucleic acids. *Chemistry & biodiversity* **2008,** *5* (3), 396-407.

12.   (a) Sheng, J.; Hassan, A. E.; Huang, Z., New telluride-mediated elimination for novel synthesis of 2',3'-didehydro-2',3'-dideoxynucleosides. *The Journal of organic chemistry* **2008,** *73*

(10), 3725-9; (b) Sheng, J.; Hassan, A. E.; Huang, Z., Synthesis of the first tellurium-derivatized oligonucleotides for structural and functional studies. *Chemistry* **2009,** *15* (39), 10210-6; (c) Sheng, J.; Hassan, A. E.; Zhang, W.; Zhou, J.; Xu, B.; Soares, A. S.; Huang, Z., Synthesis, structure and imaging of oligodeoxyribonucleotides with tellurium-nucleobase derivatization. *Nucleic acids research* **2011,** *39* (9), 3962-71.

13.   Carbon, J.; David, H.; Studier, M. H., Thiobases in Escherchia coli Transfer RNA: 2-Thiocytosine and 5-Methylaminomethyl-2-thiouracil. *Science* **1968,** *161* (3846), 1146-7.

14.   Sprinzl, M.; Horn, C.; Brown, M.; Ioudovitch, A.; Steinberg, S., Compilation of tRNA sequences and sequences of tRNA genes. *Nucleic acids research* **1998,** *26* (1), 148-53.

15.   Houssier, C.; Degee, P.; Nicoghosian, K.; Grosjean, H., Effect of uridine dethiolation in the anticodon triplet of tRNA(Glu) on its association with tRNA(Phe). *Journal of biomolecular structure & dynamics* **1988,** *5* (6), 1259-66.

16.   Agris, P. F.; Soll, D.; Seno, T., Biological function of 2-thiouridine in Escherichia coli glutamic acid transfer ribonucleic acid. *Biochemistry* **1973,** *12* (22), 4331-7.

17.   Ashraf, S. S.; Sochacka, E.; Cain, R.; Guenther, R.; Malkiewicz, A.; Agris, P. F., Single atom modification (O-->S) of tRNA confers ribosome binding. *Rna* **1999,** *5* (2), 188-94.

18.   Ching, W. M.; Alzner-DeWeerd, B.; Stadtman, T. C., A selenium-containing nucleoside at the first position of the anticodon in seleno-tRNAGlu from Clostridium sticklandii. *Proceedings of the National Academy of Sciences of the United States of America* **1985,** *82* (2), 347-50.

19.   Sun, H.; Jiang, S.; Caton-Williams, J.; Liu, H.; Huang, Z., 2-Selenouridine triphosphate synthesis and Se-RNA transcription. *Rna* **2013,** *19* (9), 1309-14.

20.   Salon, J.; Sheng, J.; Jiang, J.; Chen, G.; Caton-Williams, J.; Huang, Z., Oxygen replacement with selenium at the thymidine 4-position for the Se base pairing and crystal structure studies. *Journal of the American Chemical Society* **2007,** *129* (16), 4862-3.

21.   Salon, J.; Jiang, J.; Sheng, J.; Gerlits, O. O.; Huang, Z., Derivatization of DNAs with selenium at 6-position of guanine for function and crystal structure studies. *Nucleic acids research* **2008,** *36* (22), 7009-18.

22.   Sheng, J.; Gan, J.; Soares, A. S.; Salon, J.; Huang, Z., Structural insights of non-canonical U*U pair and Hoogsteen interaction probed with Se atom. *Nucleic acids research* **2013,** *41* (22), 10476-87.

23.   Hoffman, J. L.; McConnell, K. P., The presence of 4-selenouridine in Escherichia coli tRNA. *Biochimica et biophysica acta* **1974,** *366* (1), 109-13.

24.   Wittwer, A. J.; Tsai, L.; Ching, W. M.; Stadtman, T. C., Identification and synthesis of a naturally occurring selenonucleoside in bacterial tRNAs: 5-[(methylamino)methyl]-2-selenouridine. *Biochemistry* **1984,** *23* (20), 4650-5.

25.   (a) Ross, A. F.; Agarwal, K. C.; Chu, S. H.; Parks, R. E., Jr., Studies on the biochemical actions of 6-selenoguanine and 6-selenoguanosine. *Biochemical pharmacology* **1973,** *22* (2), 141-54; (b) Melvin, J. B.; Haight, T. H.; Leduc, E. H., Cytological effects of sulfur and selenium purine analogues on two transplantable hepatomas and on normal renewing cells in mice. *Cancer research* **1984,** *44* (7), 2794-8.

26.   Griffin, A. C., Role of selenium in the chemoprevention of cancer. *Advances in cancer research* **1979,** *29*, 419-42.

27.   (a) Watson, J. D.; Crick, F. H., Genetical implications of the structure of deoxyribonucleic acid. *Nature* **1953,** *171* (4361), 964-7; (b) Watson, J. D., Involvement of RNA in the synthesis of proteins. *Science* **1963,** *140* (3562), 17-26.

28.   Juhling, F.; Morl, M.; Hartmann, R. K.; Sprinzl, M.; Stadler, P. F.; Putz, J., tRNAdb 2009: compilation of tRNA sequences and tRNA genes. *Nucleic acids research* **2009,** *37* (Database issue), D159-62.

29.   (a) Freier, S. M.; Kierzek, R.; Caruthers, M. H.; Neilson, T.; Turner, D. H., Free energy contributions of G.U and other terminal mismatches to helix stability. *Biochemistry* **1986,** *25* (11), 3209-13; (b) Testa, S. M.; Disney, M. D.; Turner, D. H.; Kierzek, R., Thermodynamics of RNA-RNA duplexes with 2- or 4-thiouridines: implications for antisense design and targeting a group I intron. *Biochemistry* **1999,** *38* (50), 16655-62.

30.   Varani, G.; McClain, W. H., The G x U wobble base pair. A fundamental building block of RNA structure crucial to RNA function in diverse biological systems. *EMBO reports* **2000,** *1* (1), 18-23.

31.   (a) Bouadloun, F.; Donner, D.; Kurland, C. G., Codon-specific missense errors in vivo. *The EMBO journal* **1983,** *2* (8), 1351-6; (b) Stansfield, I.; Jones, K. M.; Herbert, P.; Lewendon, A.; Shaw, W. V.; Tuite, M. F., Missense translation errors in Saccharomyces cerevisiae. *Journal of molecular biology* **1998,** *282* (1), 13-24; (c) Toth, M. J.; Murgola, E. J.; Schimmel, P., Evidence for a unique first position codon-anticodon mismatch in vivo. *Journal of molecular biology* **1988,** *201* (2), 451-4.

32.   Seetharam, R.; Heeren, R. A.; Wong, E. Y.; Braford, S. R.; Klein, B. K.; Aykent, S.; Kotts, C. E.; Mathis, K. J.; Bishop, B. F.; Jennings, M. J.; et al., Mistranslation in IGF-1 during

over-expression of the protein in Escherichia coli using a synthetic gene containing low frequency codons. *Biochemical and biophysical research communications* **1988,** *155* (1), 518-23.

33.   (a) Ikemura, T., Correlation between the abundance of yeast transfer RNAs and the occurrence of the respective codons in protein genes. Differences in synonymous codon choice patterns of yeast and Escherichia coli with reference to the abundance of isoaccepting transfer RNAs. *Journal of molecular biology* **1982,** *158* (4), 573-97; (b) Bernardi, G.; Bernardi, G., Codon usage and genome composition. *Journal of molecular evolution* **1985,** *22* (4), 363-5.

34.   Parker, J., Errors and alternatives in reading the universal genetic code. *Microbiological reviews* **1989,** *53* (3), 273-98.

35.   Crick, F. H., Codon--anticodon pairing: the wobble hypothesis. *Journal of molecular biology* **1966,** *19* (2), 548-55.

36.   Lim, V. I., Analysis of action of wobble nucleoside modifications on codon-anticodon pairing within the ribosome. *Journal of molecular biology* **1994,** *240* (1), 8-19.

37.   Agris, P. F., Wobble position modified nucleosides evolved to select transfer RNA codon recognition: a modified-wobble hypothesis. *Biochimie* **1991,** *73* (11), 1345-9.

38.   (a) Dunkel, M.; P Dan Cook, O. L. A., Synthesis of novel c-2 substituted pyrimidine nucleoside analogs. *Journal of Heterocyclic Chemistry* **1993**; (b) Seio, K.; Sasami, T.; Tawarada, R.; Sekine, M., Synthesis of 2'-O-methyl-RNAs incorporating a 3-deazaguanine, and UV melting and computational studies on its hybridization properties. *Nucleic acids research* **2006,** *34* (16), 4324-34.

39.   (a) Wise, D.; Townsend, L., Synthesis of the selenopyrimidine nucleosides 2 seleno and 4 selenouridine. *Journal of Heterocyclic Chemistry* **1972,** *9* (6), 1461-1462; (b) Shiue, C. Y.; Chu, S. H., FACILE SYNTHESIS OF 1-BETA-D-ARABINOFURANOSYL-2-SELENOURACIL

AND 1-BETA-D-ARABINOFURANOSYL-4-SELENOURACIL AND RELATED

COMPOUNDS. *Journal of Organic Chemistry* **1975,** *40* (20), 2971-2974.

40. Hassan, A.; Sheng, J.; Zhang, W.; Huang, Z., High fidelity of base pairing by 2-

selenothymidine in DNA. *Journal of the American Chemical Society* **2010,** *132* (7), 2120-2121.

41. Kumar, R. K.; Davis, D. R., Synthesis and studies on the effect of 2-thiouridine and 4-

thiouridine on sugar conformation and RNA duplex stability. *Nucleic Acids Research* **1997,** *25*

(6), 1272-1280.

42. (a) Salon, J.; Jiang, J. S.; Sheng, J.; Gerlits, O. O.; Huang, Z., Derivatization of DNAs

with selenium at 6-position of guanine for function and crystal structure studies. *Nucleic Acids*

Research **2008,** *36* (22), 7009-7018; (b) Salon, J.; Sheng, J.; Jiang, J. S.; Chen, G. X.; Caton-

Williams, J.; Huang, Z., Oxygen replacement with selenium at the thymidine 4-position for the

Se base pairing and crystal structure studies. *Journal of the American Chemical Society* **2007,**

129 (16), 4862-+.

43. Vorbrüggen, H.; Strehlke, P., Nucleosidsynthesen, VII. Eine einfache Synthese von 2‐

Thiopyrimidin‐nucleosiden. *Chemische Berichte* **1973,** *106* (9), 3039-3061.

44. Vorbrüggen, H.; Strehlke, P., Nucleosidsynthesen, VII. Eine einfache Synthese von 2

Thiopyrimidin nucleosiden. *Chem. Ber.* **1973,** *106* (9), 3039-3061.

45. Knobloch, B.; Da Costa, C. P.; Linert, W.; Sigel, H., Stability constants of metal ion

complexes formed with N3-deprotonated uridine in aqueous solution. *Inorganic chemistry*

communications **2003,** *6* (1), 90-93.

46. Zbyszek Otwinowski, W. M., Processing of X-ray diffraction data collected in oscillation

mode. *METHODS IN ENZYMOLOGY* **1997**.

47.   Wahl, M. C.; Ban, C.; Sekharudu, C.; Ramakrishnan, B.; Sundaralingam, M., Structure of the purine-pyrimidine alternating RNA double helix, r(GUAUAUA)d(C), with a 3'-terminal deoxy residue. *Acta crystallographica. Section D, Biological crystallography* **1996,** *52* (Pt 4), 655-67.

48.   Lim, V. I.; Curran, J. F., Analysis of codon:anticodon interactions within the ribosome provides new insights into codon reading and the genetic code structure. *Rna* **2001,** *7* (7), 942-57.

49.   (a) Hendrickson, W. A.; Horton, J. R.; LeMaster, D. M., Selenomethionyl proteins produced for analysis by multiwavelength anomalous diffraction (MAD): a vehicle for direct determination of three-dimensional structure. *The EMBO journal* **1990,** *9* (5), 1665-72; (b) Ferre-D'Amare, A. R.; Zhou, K.; Doudna, J. A., Crystal structure of a hepatitis delta virus ribozyme. *Nature* **1998,** *395* (6702), 567-74.

50.   (a) Carrasco, N.; Ginsburg, D.; Du, Q.; Huang, Z., Synthesis of selenium-derivatized nucleosides and oligonucleotides for X-ray crystallography. *Nucleosides, nucleotides & nucleic acids* **2001,** *20* (9), 1723-34; (b) Sheng, J.; Huang, Z., Selenium derivatization of nucleic acids for X-ray crystal-structure and function studies. *Chemistry & biodiversity* **2010,** *7* (4), 753-85.

51.   (a) Caton-Williams, J.; Huang, Z., Synthesis and DNA-polymerase incorporation of colored 4-selenothymidine triphosphate for polymerase recognition and DNA visualization. *Angewandte Chemie* **2008,** *47* (9), 1723-5; (b) Hassan, A. E.; Sheng, J.; Zhang, W.; Huang, Z., High fidelity of base pairing by 2-selenothymidine in DNA. *Journal of the American Chemical Society* **2010,** *132* (7), 2120-1.

52.   Yoshikawa, M.; Kato, T.; Takenishi, T., A novel method for phosphorylation of nucleosides to 5'-nucleotides. *Tetrahedron letters* **1967,** *50*, 5065-8.

53.   Lin, L.; Caton-Williams, J.; Kaur, M.; Patino, A. M.; Sheng, J.; Punetha, J.; Huang, Z., Facile synthesis of nucleoside 5'-(alpha-P-seleno)-triphosphates and phosphoroselenoate RNA transcription. *Rna* **2011,** *17* (10), 1932-8.

54.   Wise, D.; Townsend, L., Synthesis of the selenopyrimidine nucleosides 2 seleno and 4 selenouridine. HETERO CORPORATION PO BOX 993, ODESSA, FL 33556-0993: 1972; Vol. 9, pp 1461-1462.

55.   Brunger, A. T.; Adams, P. D.; Clore, G. M.; DeLano, W. L.; Gros, P.; Grosse-Kunstleve, R. W.; Jiang, J. S.; Kuszewski, J.; Nilges, M.; Pannu, N. S.; Read, R. J.; Rice, L. M.; Simonson, T.; Warren, G. L., Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta crystallographica. Section D, Biological crystallography* **1998,** *54* (Pt 5), 905-21.

APPENDICES

APPENDIX 1. Nucleic Acid Mini Screen contains twenty-four unique reagents from Hampton Research.

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

APPENDIX 2. ^{1}H NMR spectra of 1-(5'-*O*-4,4'-dimethoxytrityl-beta-*D*-ribofuranosyl)-2-thiouridine **6**.
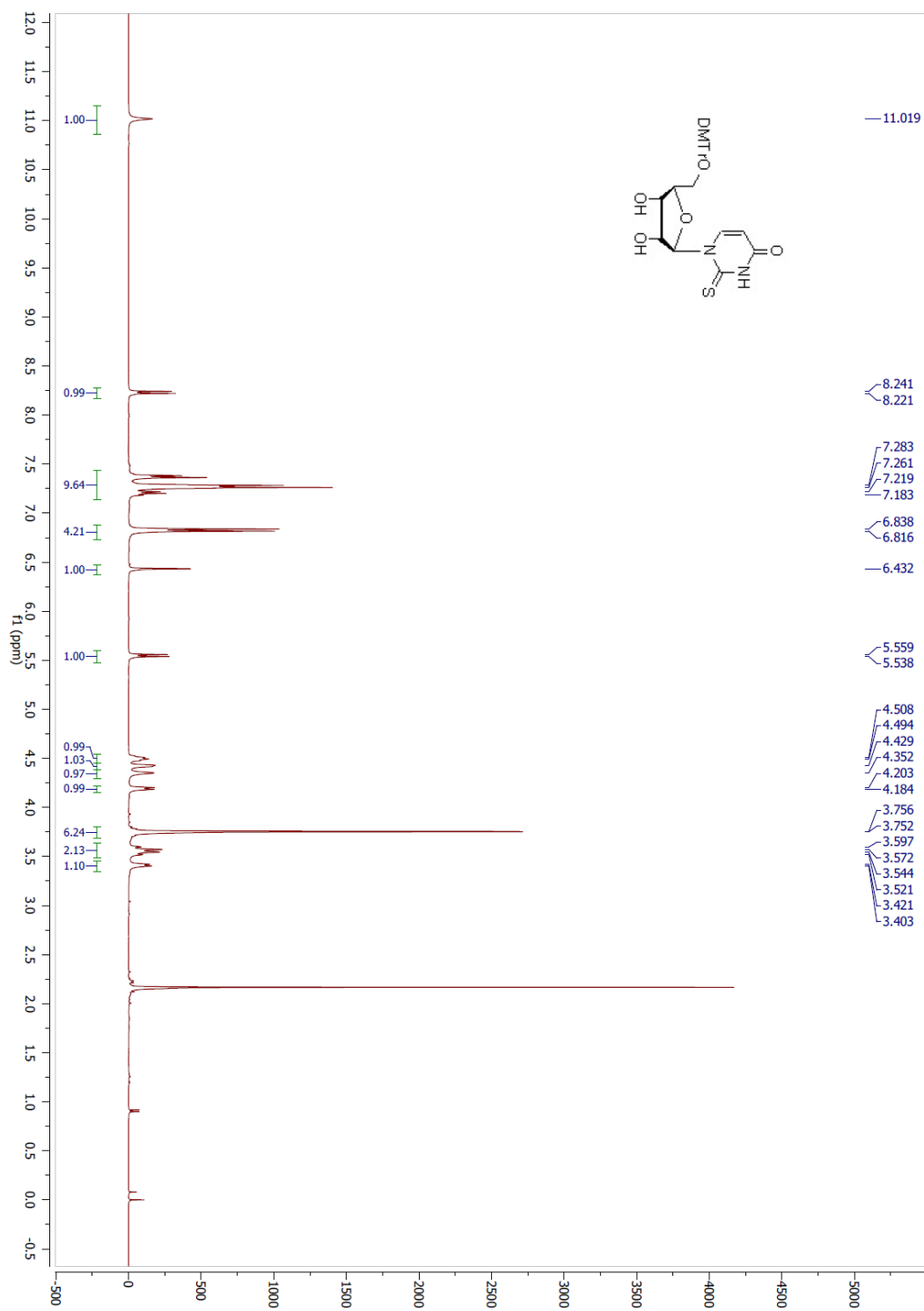
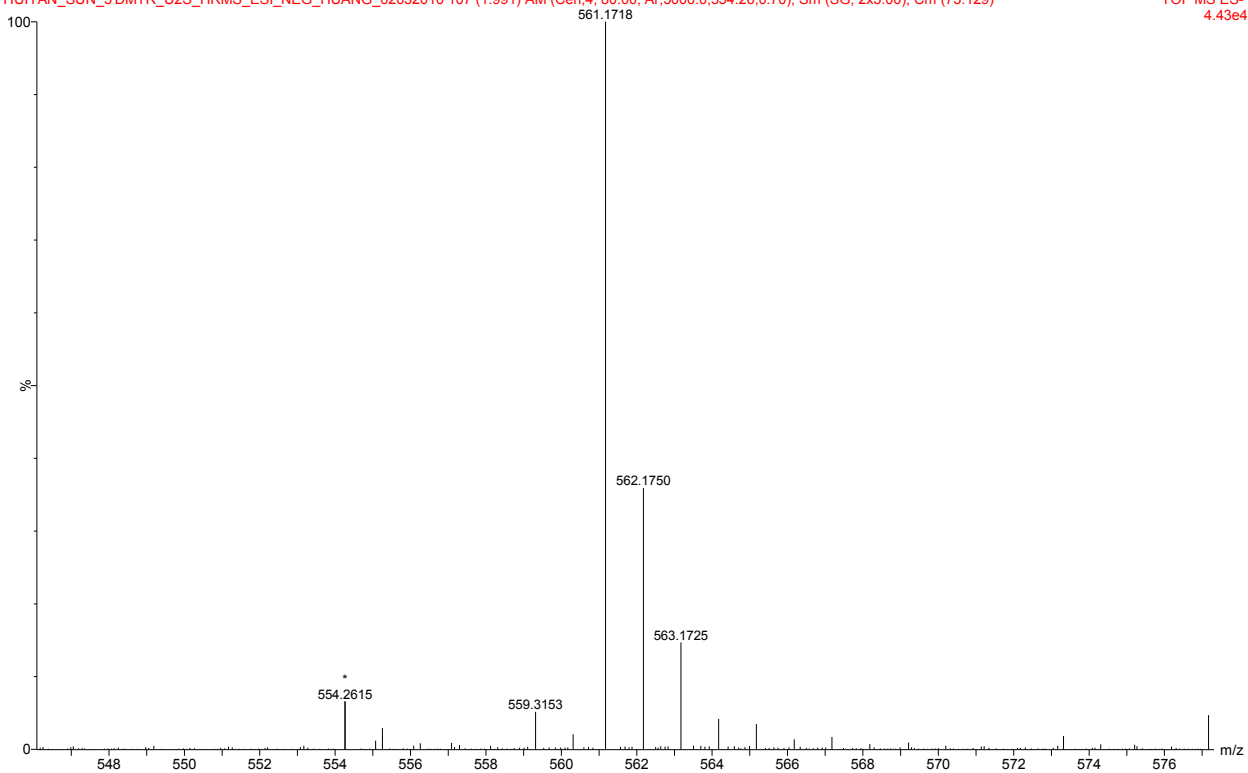APPENDIX 3. ^{13}C NMR spectra of 1-(5'-*O*-4,4'-dimethoxytrityl-beta-*D*-ribofuranosyl)-2-thiouridine **6**.
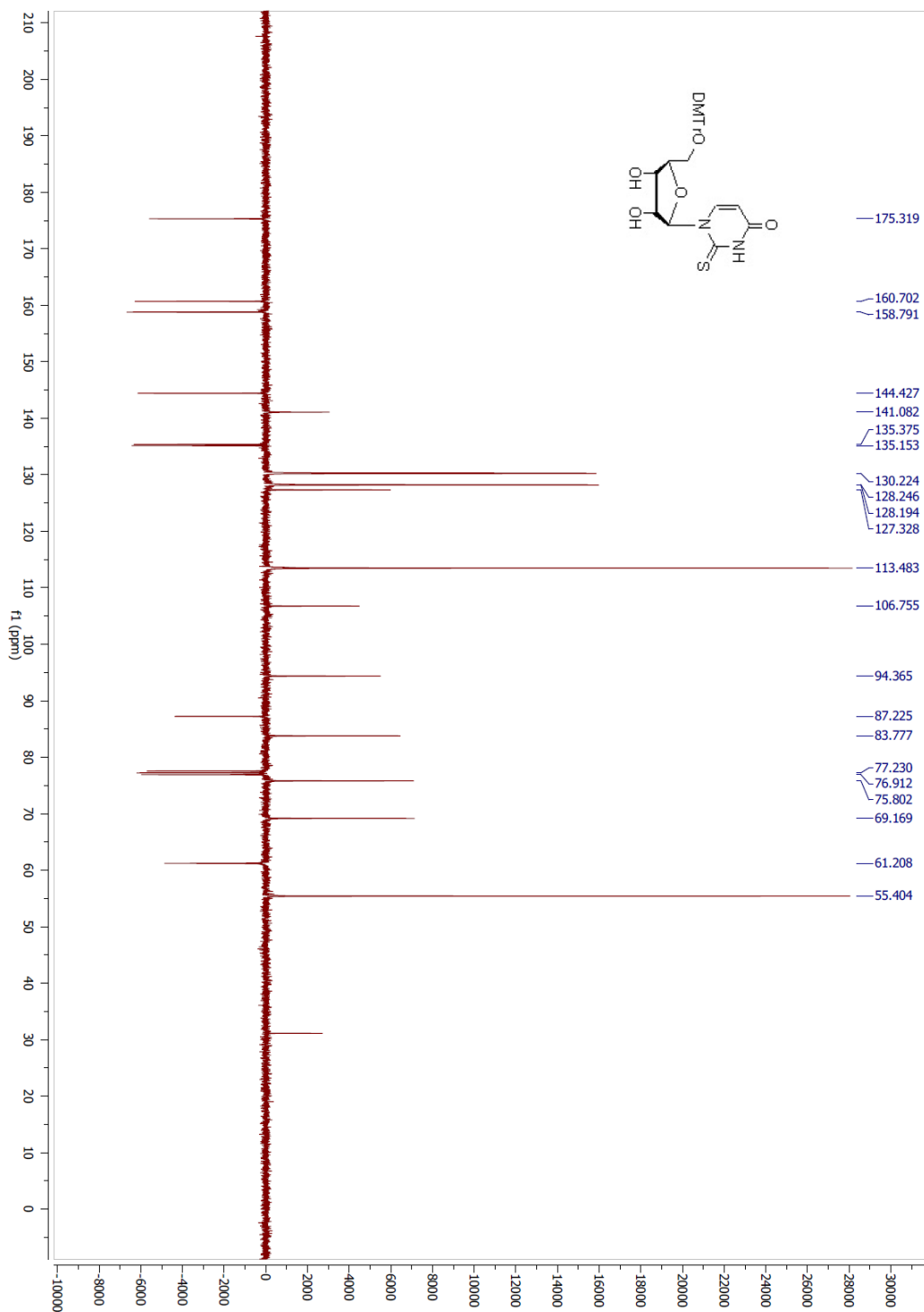
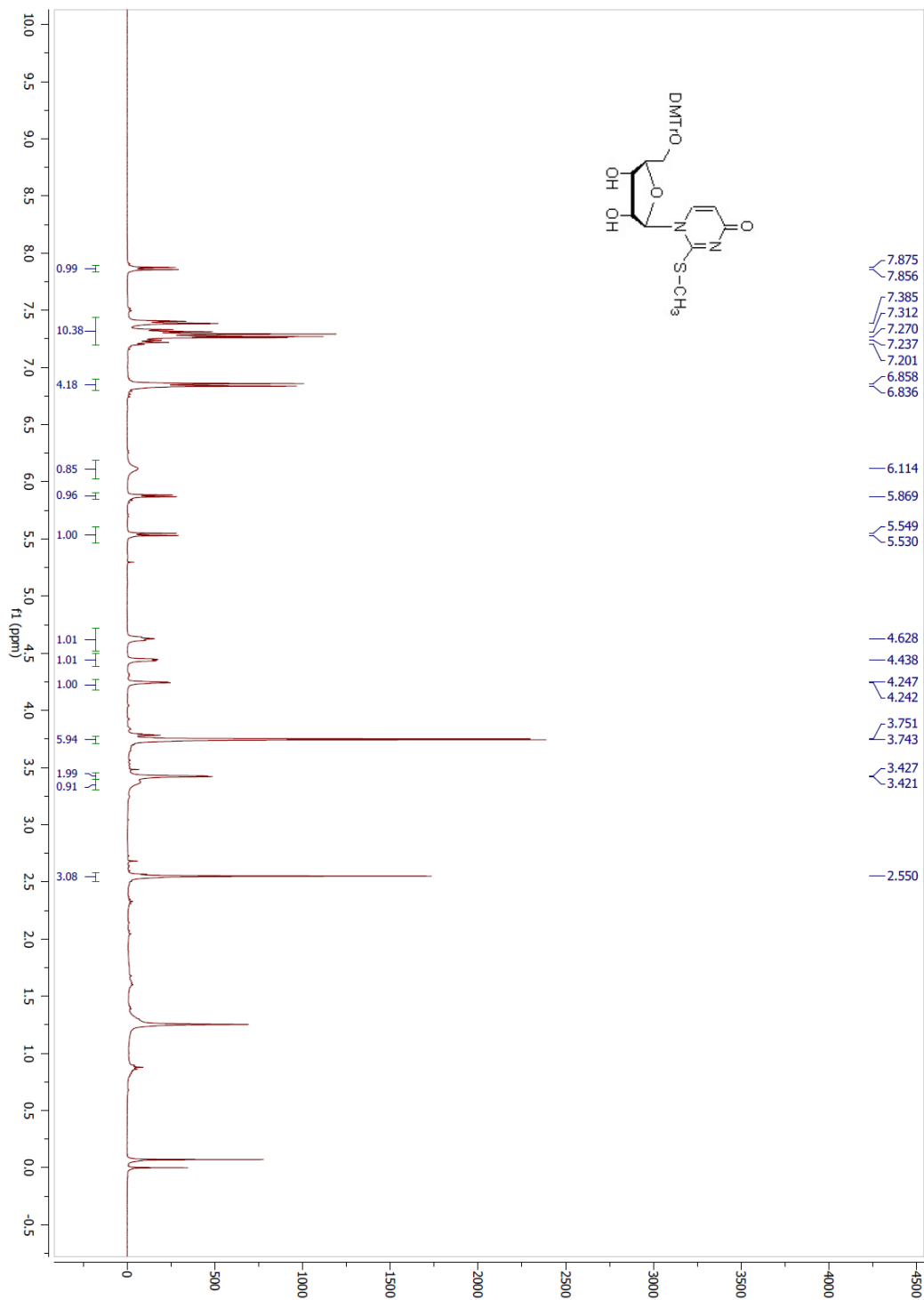APPENDIX 4. HRMS (ESI-TOF) of 1-(5'-*O*-4,4'-dimethoxytrityl-beta-*D*-ribofuranosyl)-2-thiouridine
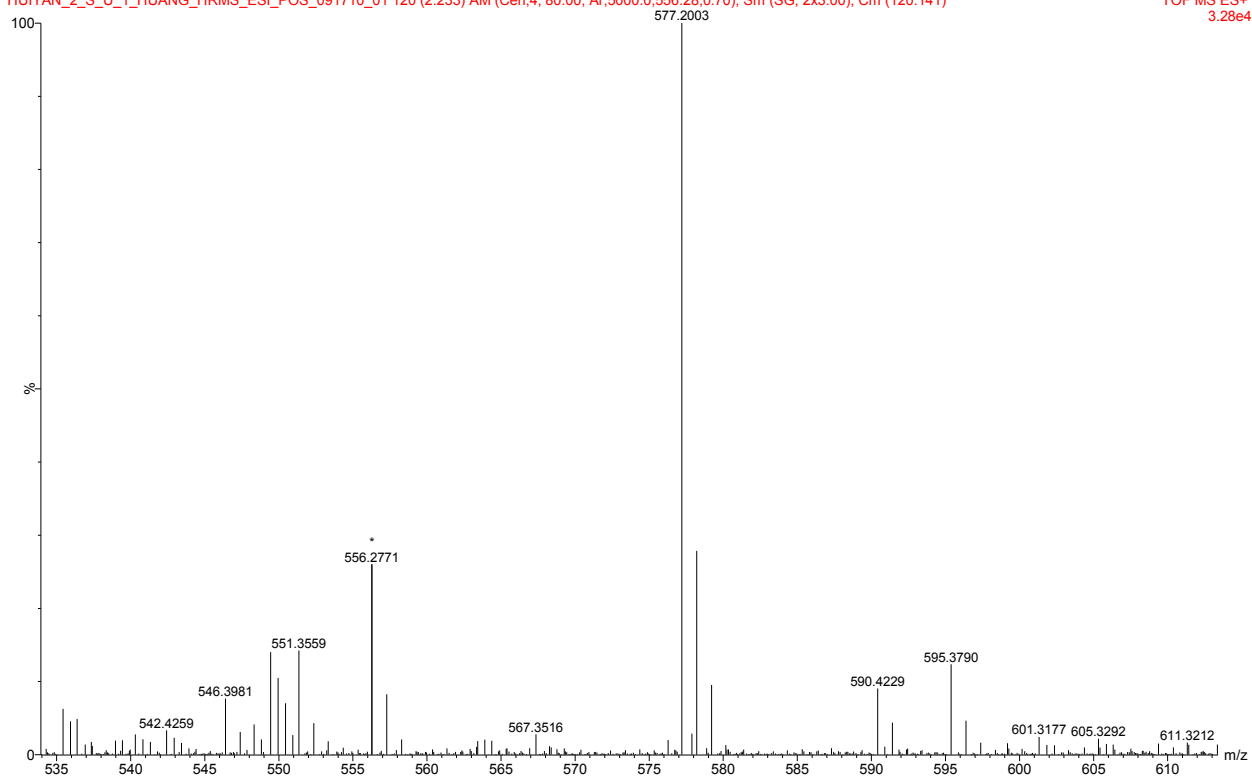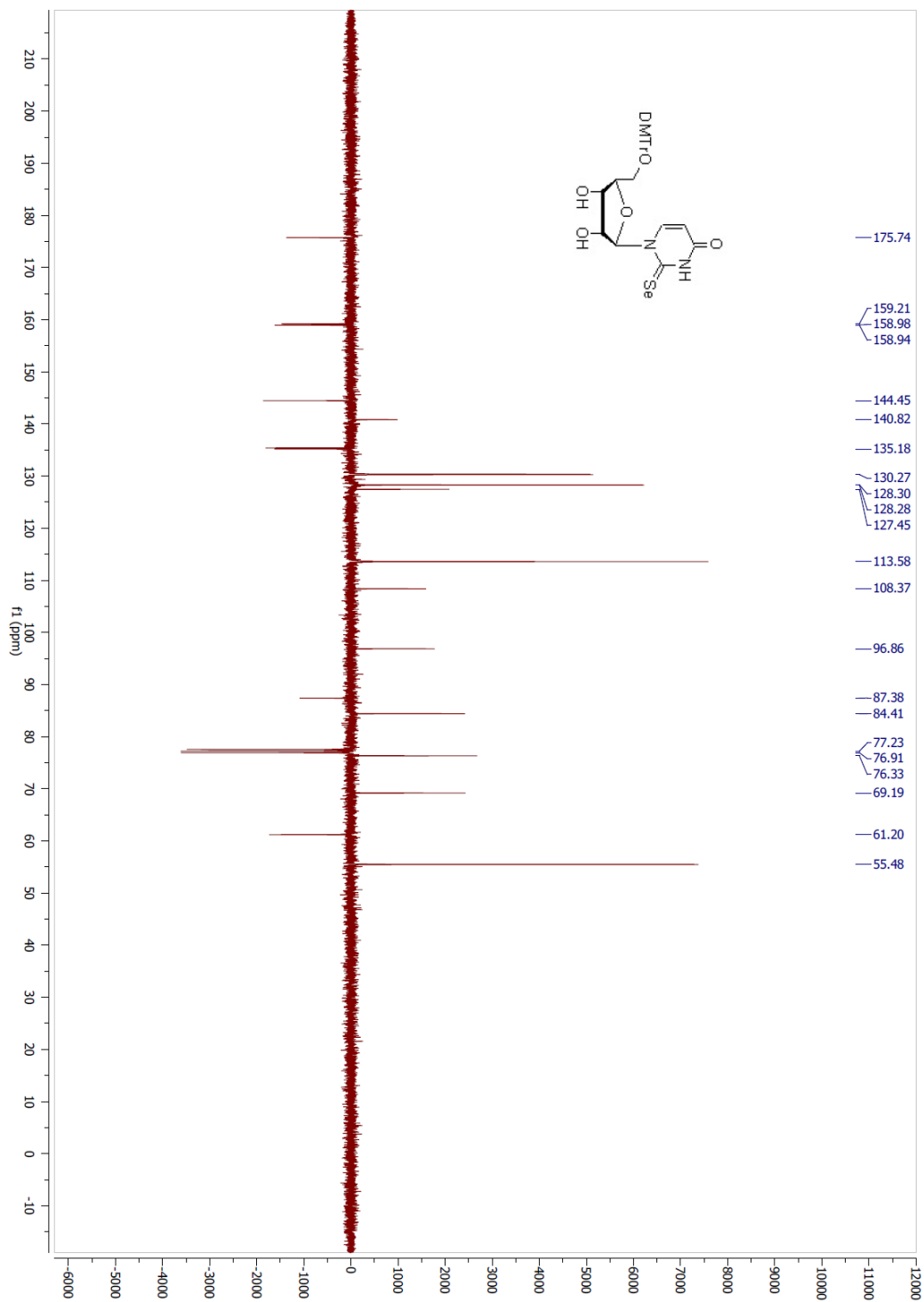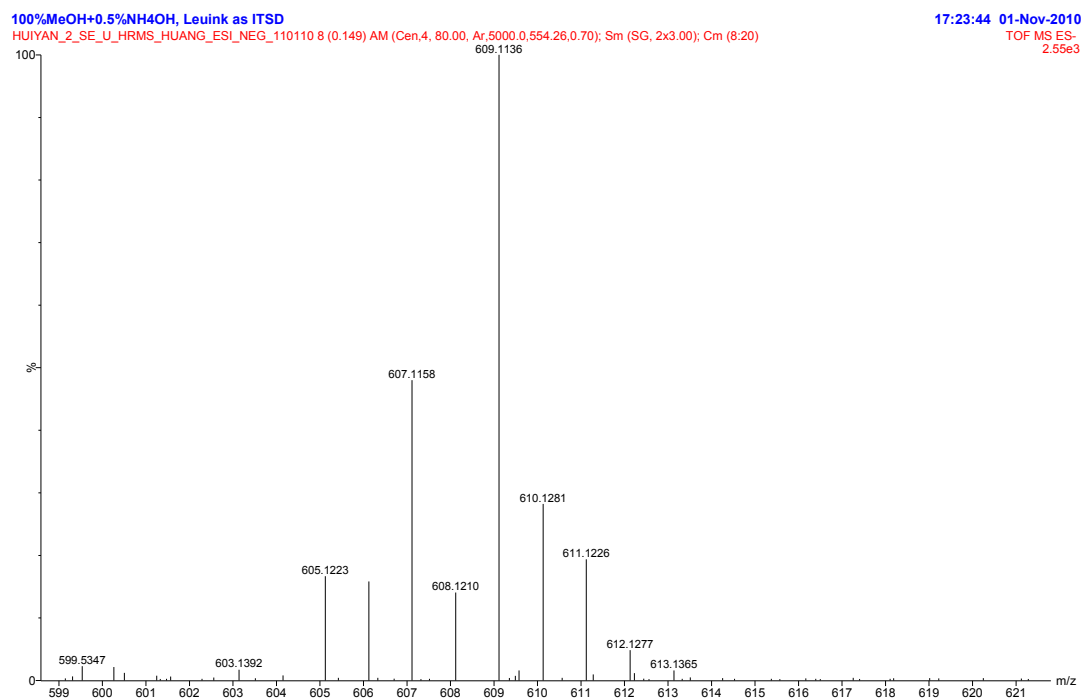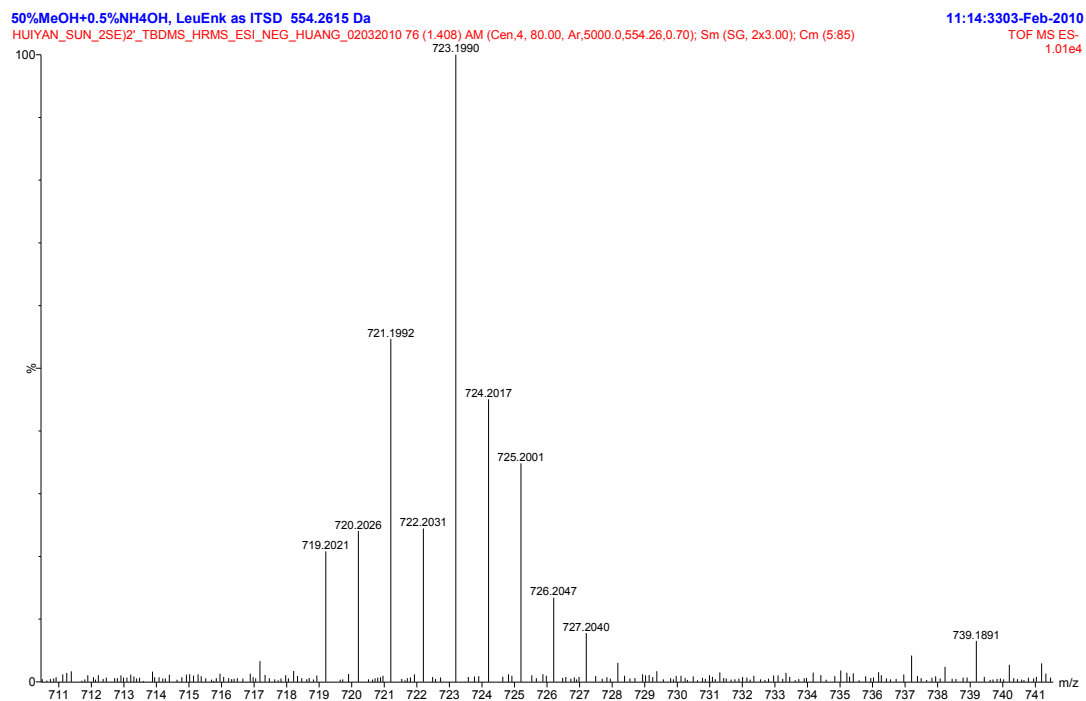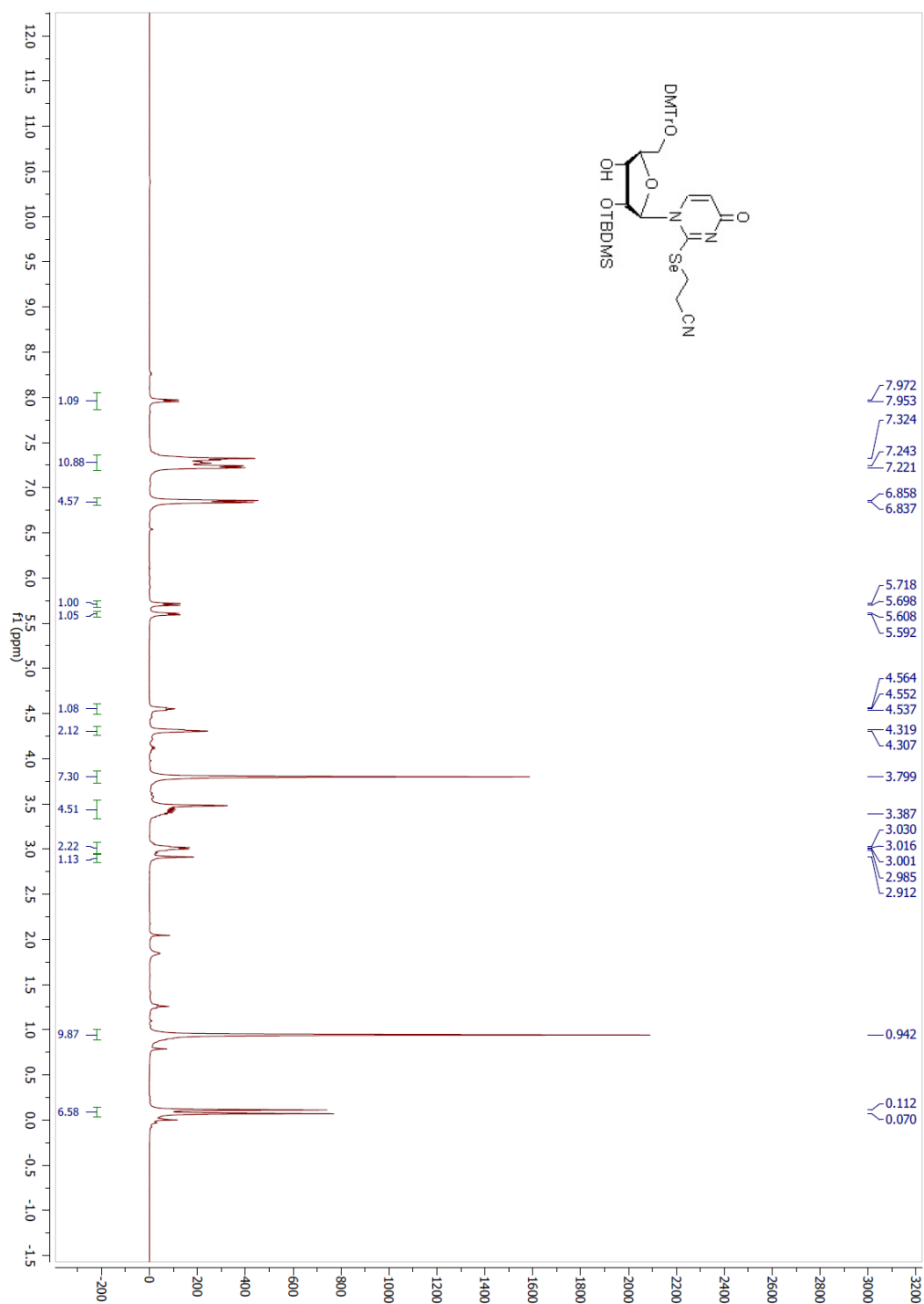
6.

APPENDIX 5. ^{1}H NMR spectra of 1-(5'-*O*-4,4'-dimethoxytrityl-beta-*D*-ribofuranosyl)-2-

methylthiouridine **7.**

APPENDIX 6. ^{13}C NMR spectra of 1-(5'-*O*-4,4'-dimethoxytrityl-beta-*D*-ribofuranosyl)-2-methylthiouridine **7.**
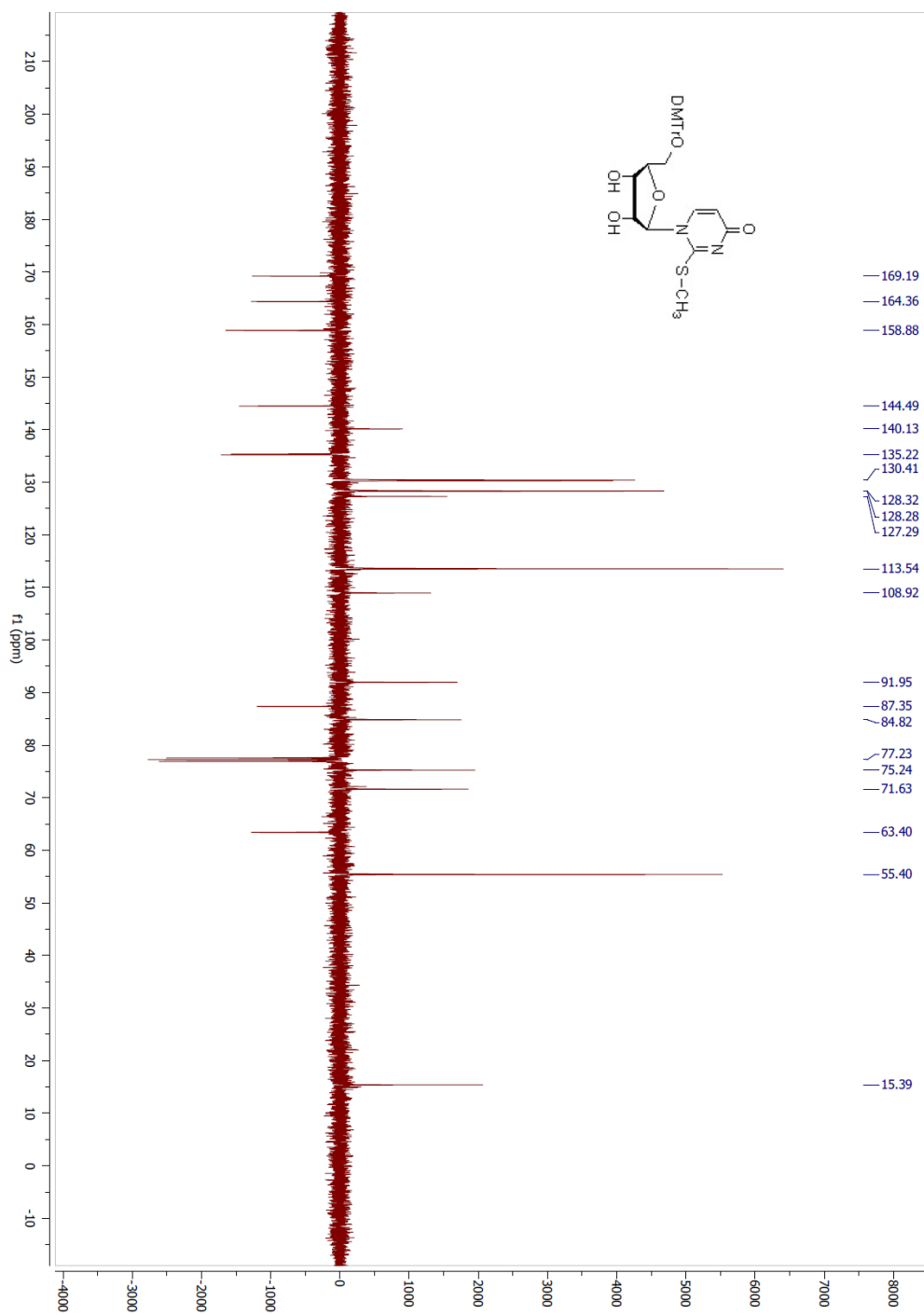
APPENDIX 7. HRMS (ESI-TOF) of 1-(5'-*O*-4,4'-dimethoxytrityl-beta-*D*-ribofuranosyl)-2-methylthiouridine compound **7**.

APPENDIX 8. ¹H NMR spectra of 1-(5'-*O*-4,4'-dimethoxytrityl-beta-*D*-ribofuranosyl)-2-selenouridine

8.
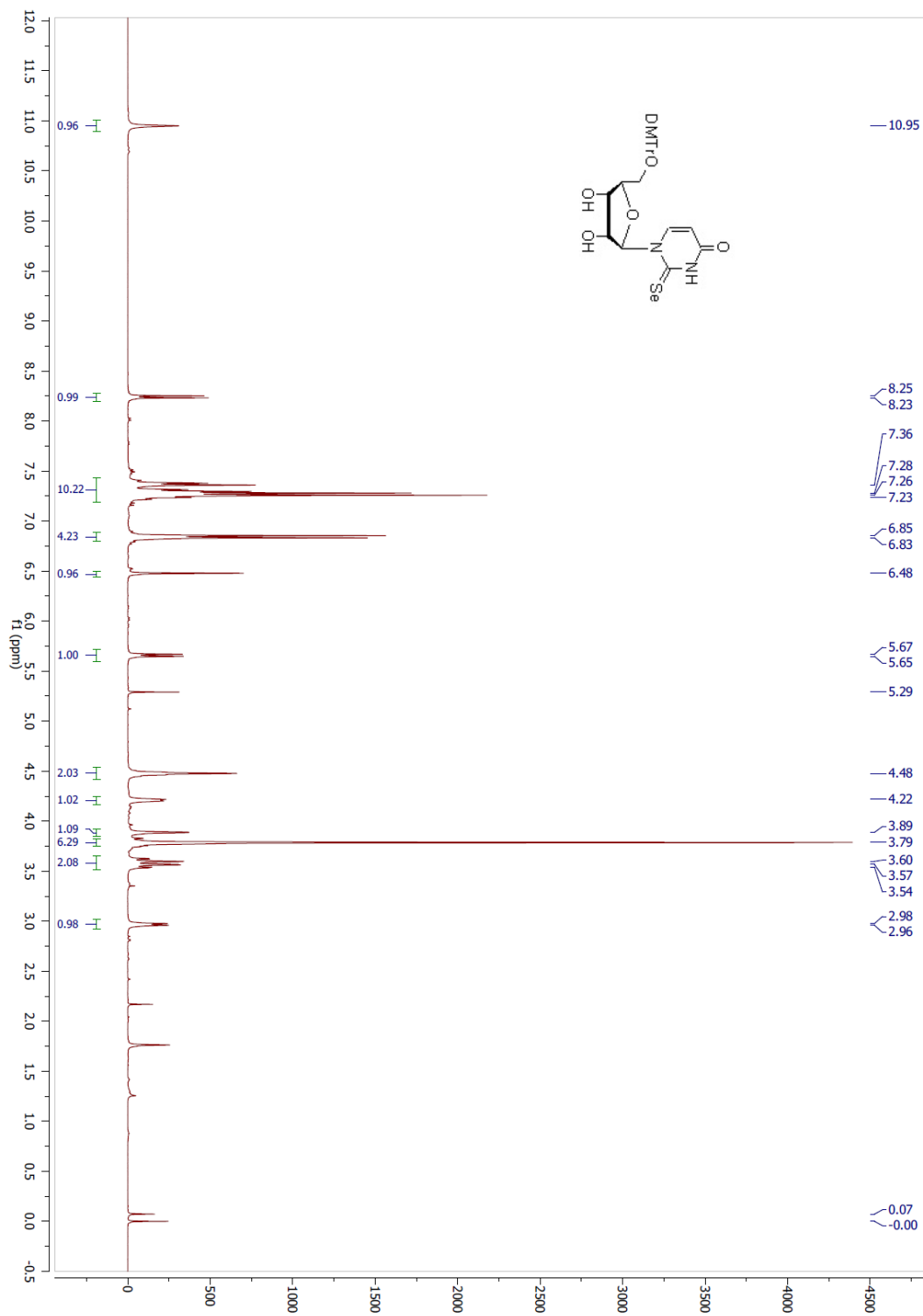
APPENDIX 9. ¹³C NMR spectra of 1-(5'-*O*-4,4'-dimethoxytrityl-beta-*D*-ribofuranosyl)-2-selenouridine

8.

APPENDIX 10. HRMS (ESI-TOF) of 1-(5'-*O*-4,4'-dimethoxytrityl-beta-*D*-ribofuranosyl)-2-selenouridine.

APPENDIX 11. HRMS (ESI-TOF) of compound **9a** and **9b.**

APPENDIX 12. ^{1}H NMR spectra of 1-(2'-*O*-tert-butyldimethylsilyl-5'-*O*-4,4'-dimethoxytrityl-beta-*D*-ribofuranosyl)-2-cyanoethylselanyluridine **10a**.

APPENDIX 13. ^{13}C NMR spectra of 1-(2'-*O*-tert-butyldimethylsilyl-5'-*O*-4,4'-dimethoxytrityl-beta-*D*-ribofuranosyl)-2-cyanoethylselanyluridine **10a**.

APPENDIX 14. ^{1}H NMR spectra of 1-(3'-*O*-tert-butyldimethylsilyl-5'-*O*-4,4'-dimethoxytrityl-beta-*D*-ribofuranosyl)-2-cyanoethylselanyluridine **10b**.

APPENDIX 15. ^{13}C NMR spectra of 1-(3'-*O*-tert-butyldimethylsilyl-5'-*O*-4,4'-dimethoxytrityl-beta-*D*-ribofuranosyl)-2-cyanoethylselanyluridine **10b**.
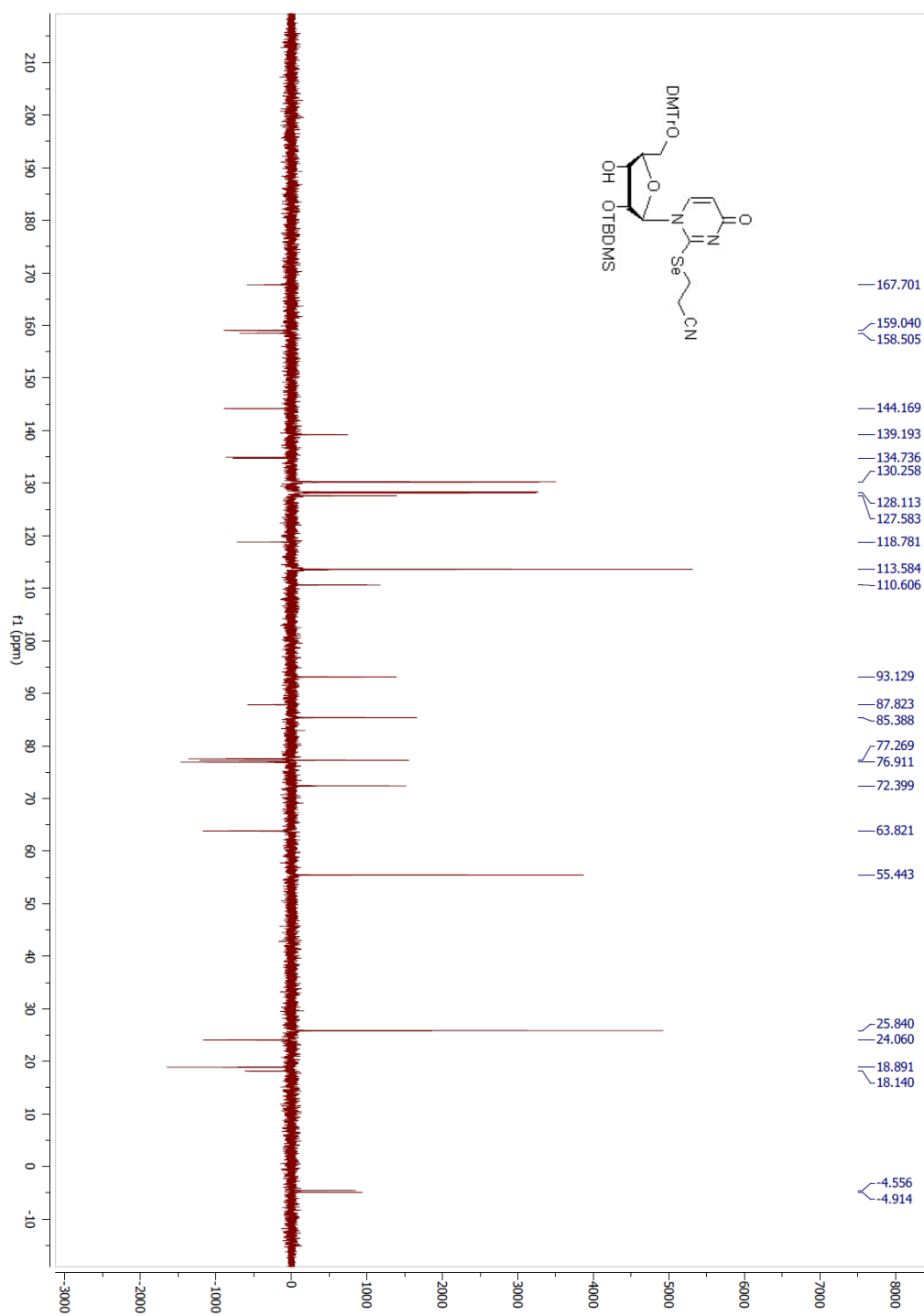
APPENDIX 16. HRMS (ESI-TOF) of 1-(2'-*O*-tert-butyldimethylsilyl-5'-*O*-4,4'-dimethoxytrityl-beta-*D*-ribofuranosyl)-2-cyanoethylselanyluridine **10a**.
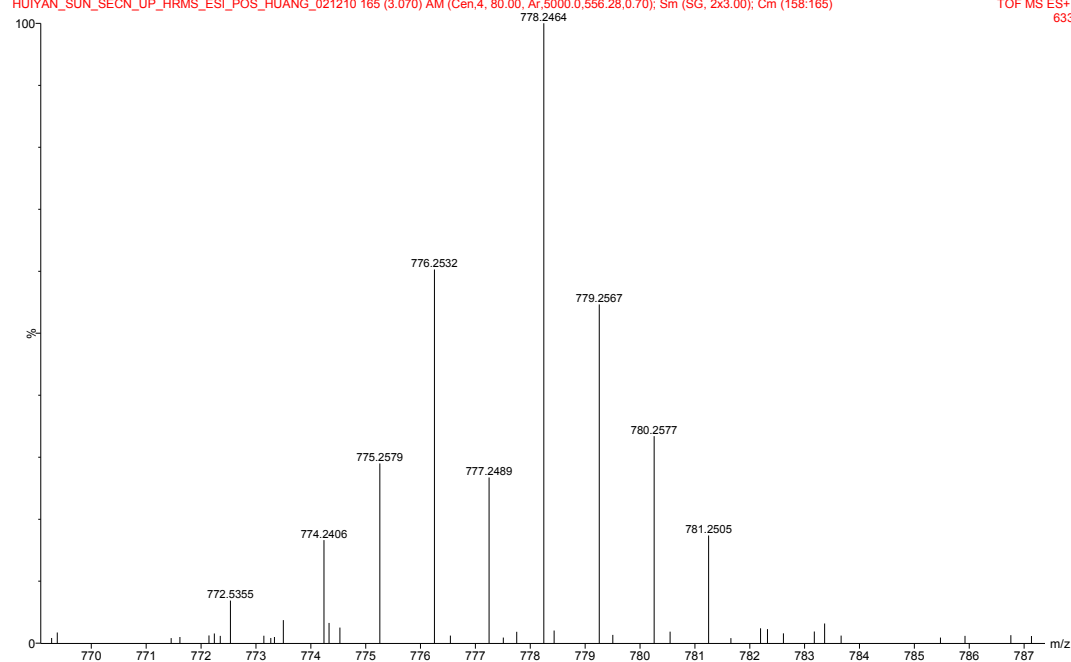
APPENDIX 17. HRMS (ESI-TOF) of 1-(3'-*O*-tert-butyldimethylsilyl-5'-*O*-4,4'-dimethoxytrityl-beta-*D*-ribofuranosyl)-2-cyanoethylselanyluridine **10b**.
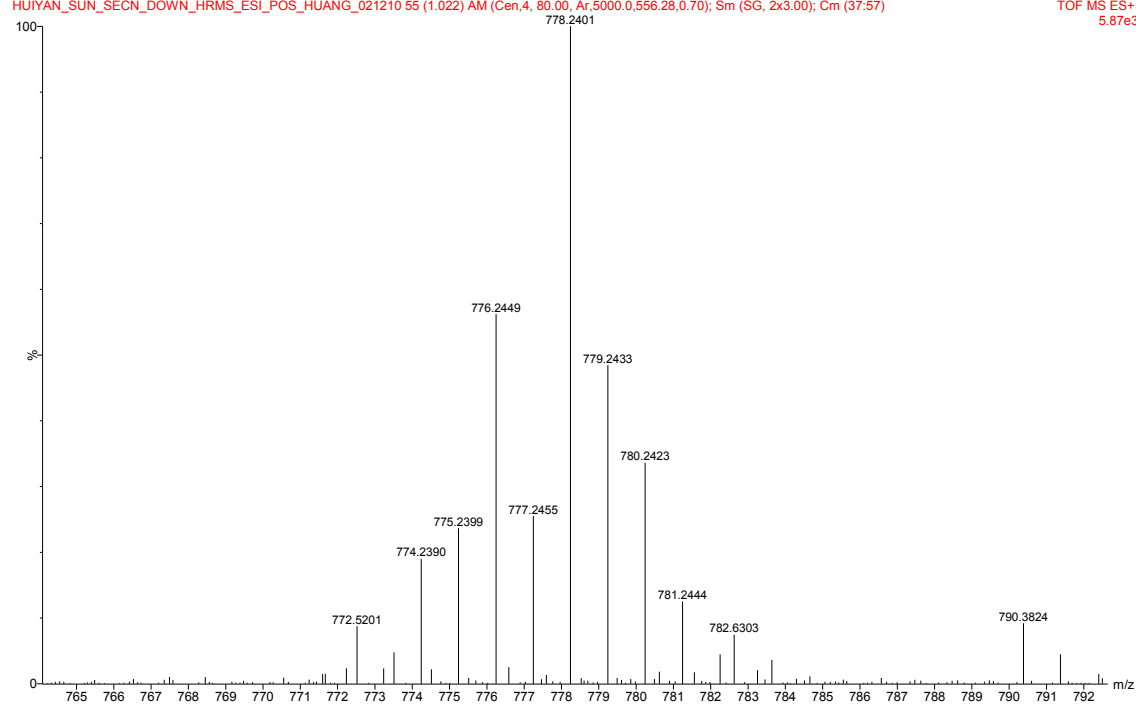
APPENDIX 18. ¹H NMR spectra of 1-[2'-*O*-tert-butyldimethylsilyl-3'-*O*-(2-cyanoethyl-*N*,*N*-diisopropylamino) phosphoramidite-5'-*O*-(4,4'-dimethoxytrityl-beta-*D*-ribofuranosyl)]-2-cyanoethylselanyluridine **11**.
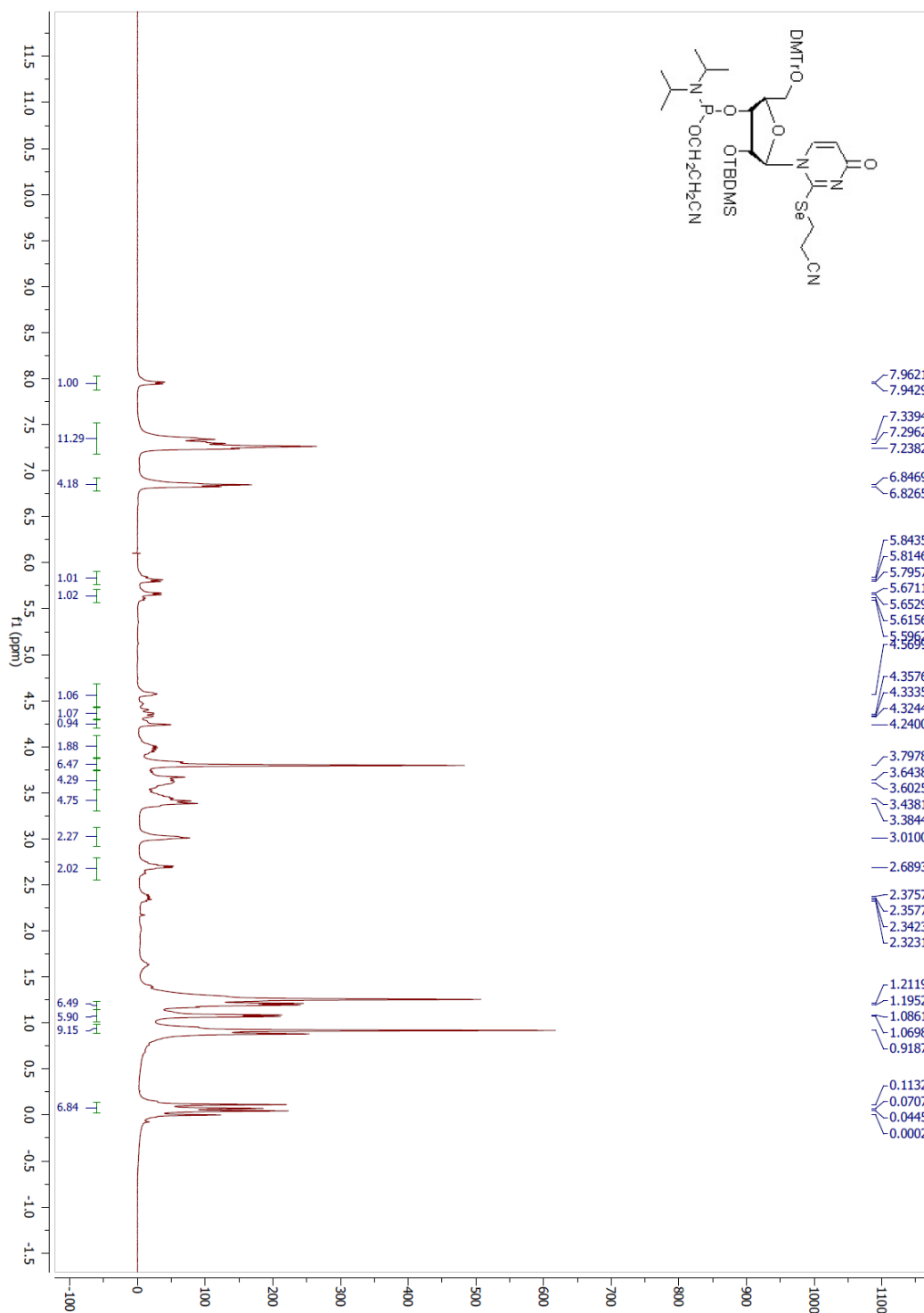
APPENDIX 19. ¹³C NMR spectra of 1-[2'-*O*-tert-butyldimethylsilyl-3'-*O*-(2-cyanoethyl-*N,N*-diisopropylamino) phosphoramidite-5'-*O*-(4,4'-dimethoxytrityl-beta-*D*-ribofuranosyl)]-2-cyanoethylselanyluridine **11**.
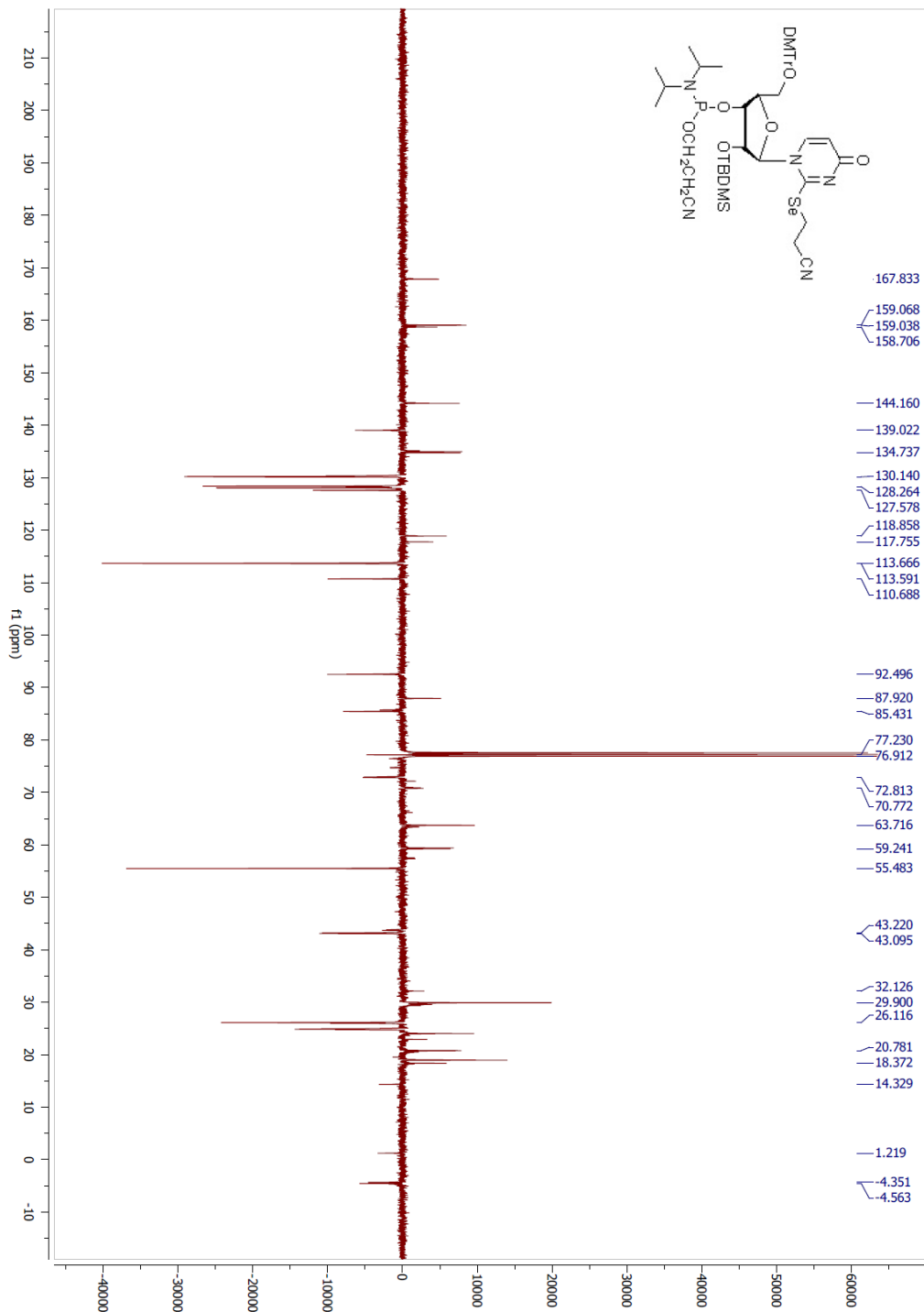
APPENDIX 20. ^{31}P NMR spectra of 1-[2'-*O*-tert-butyldimethylsilyl-3'-*O*-(2-cyanoethyl-*N*,*N*-diisopropylamino) phosphoramidite-5'-*O*-(4,4'-dimethoxytrityl-beta-*D*-ribofuranosyl)]-2-cyanoethylselanyluridine **11**.
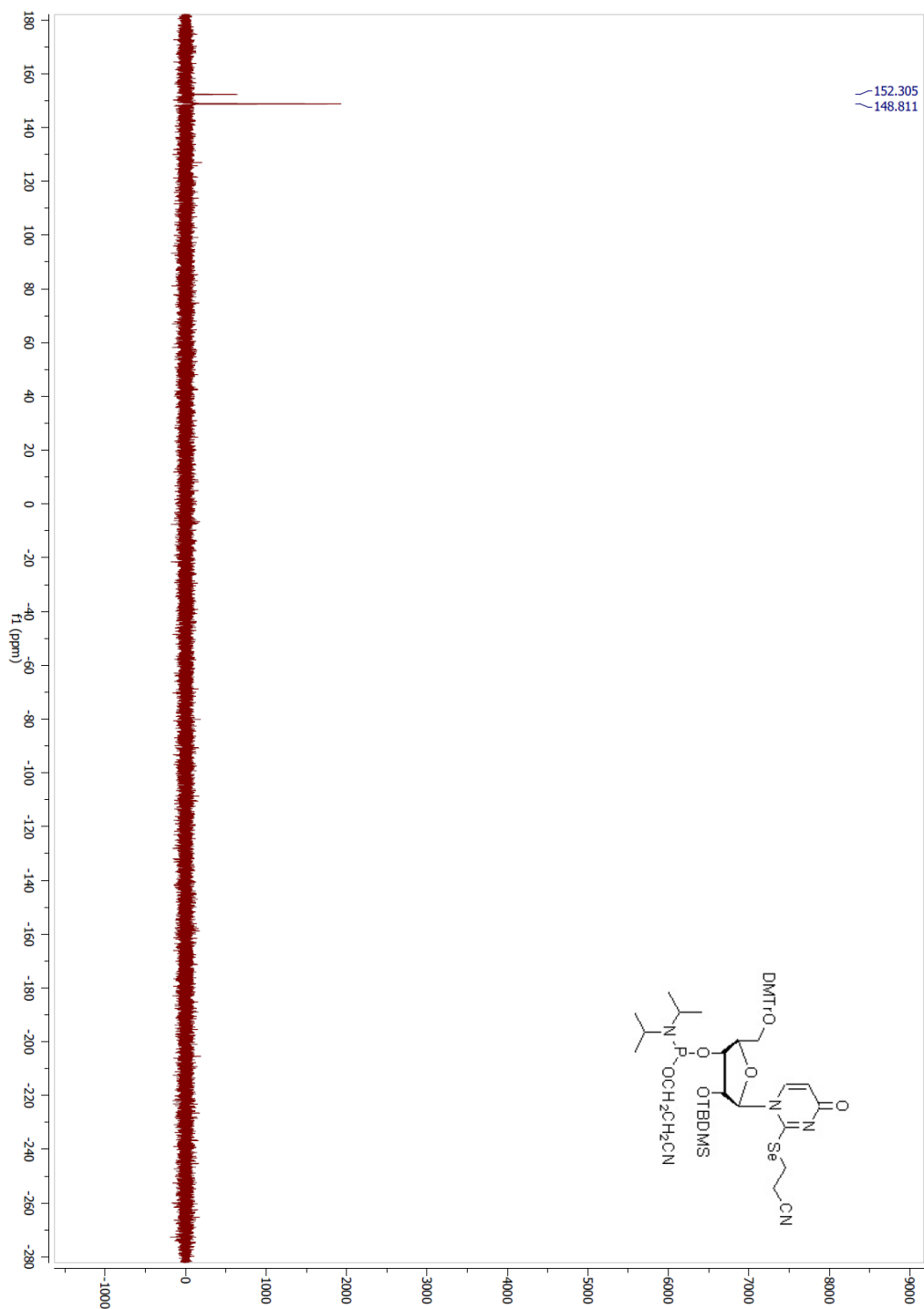
APPENDIX 21. HRMS (ESI-TOF) of 1-[2'-*O*-tert-butyldimethylsilyl-3'-*O*-(2-cyanoethyl-*N*,*N*-diisopropylamino) phosphoramidite-5'-*O*-(4,4'-dimethoxytrityl-beta-*D*-ribofuranosyl)]-2-cyanoethylselanyluridine **11**.
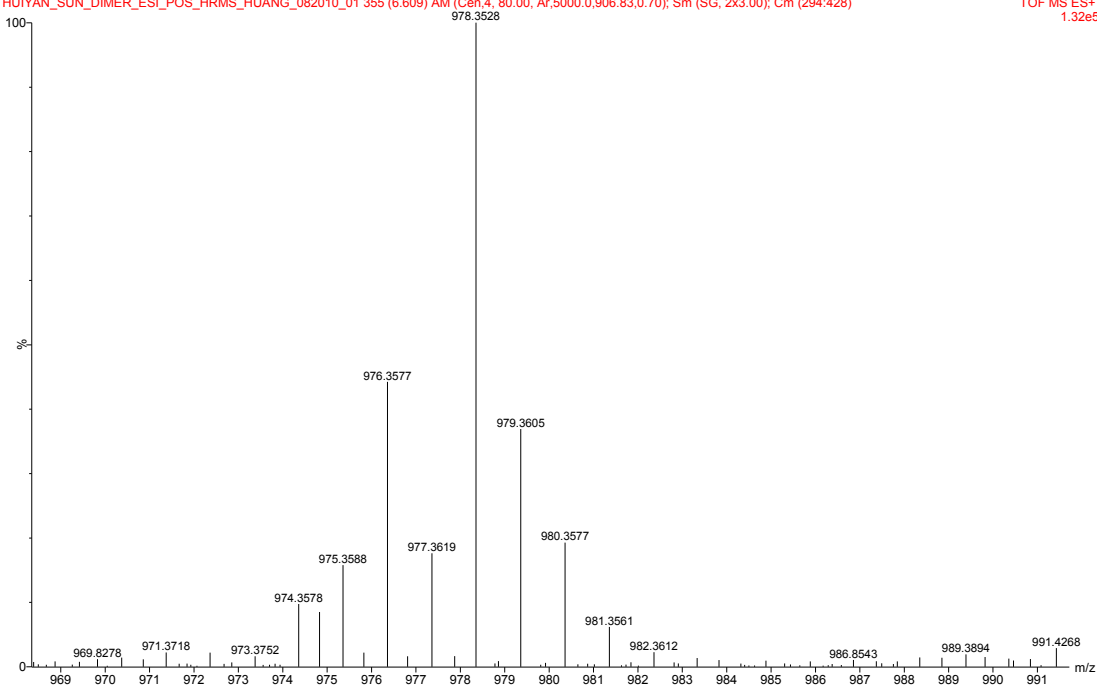
APPENDIX 22. MALDI-TOF MS of 2-Se-U 12mer (5'-AUCACC^Se^UCCUUA-3') [M+H^+^]^+^ = 3740.3 (calc. 3740.2).
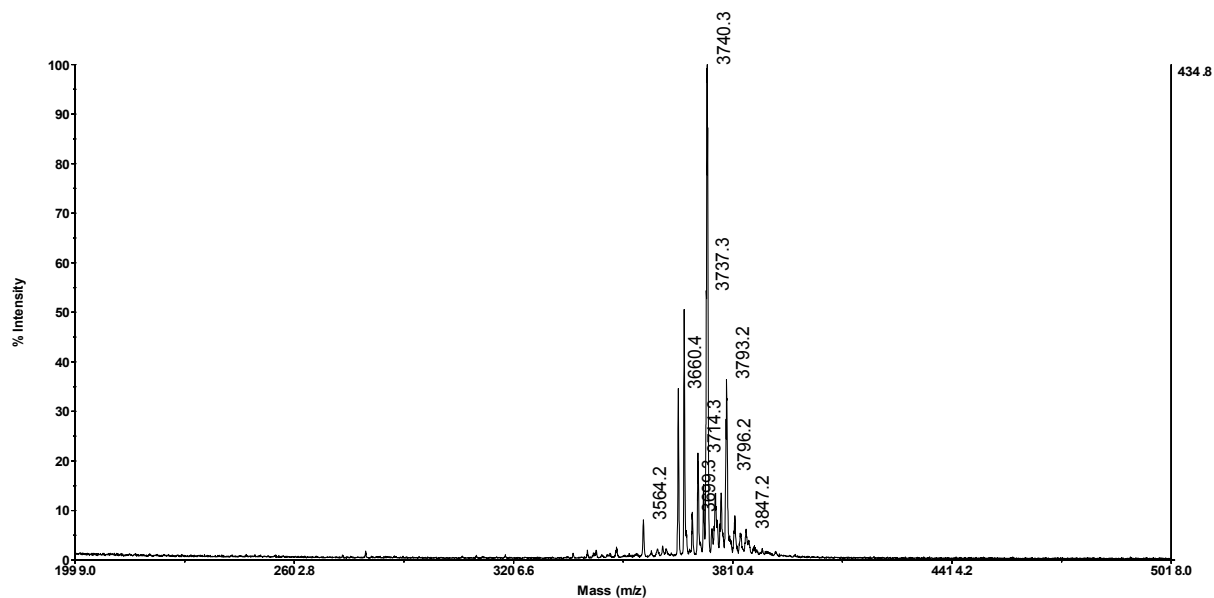
APPENDIX 23. MALDI-TOF MS of 2-Se-U 12mer (5'-AAUGC^{Se}UGCACUG-3') [M+H$^+$]$^+$ = 3859.4 (calc. 3859.3).
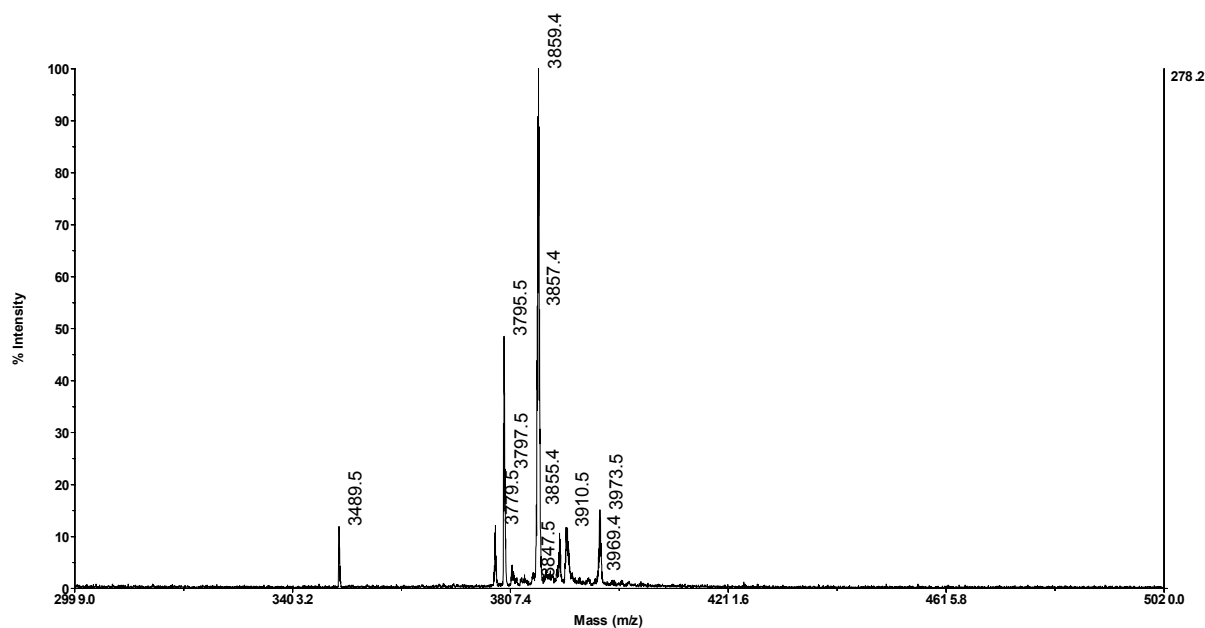
APPENDIX 24. MALDI-TOF MS of 2-Se-U 8mer (5'-GUAUA^{Se}UAC-3') [M+H$^+$]$^+$ = 2558.7 (calc. 2558.5).
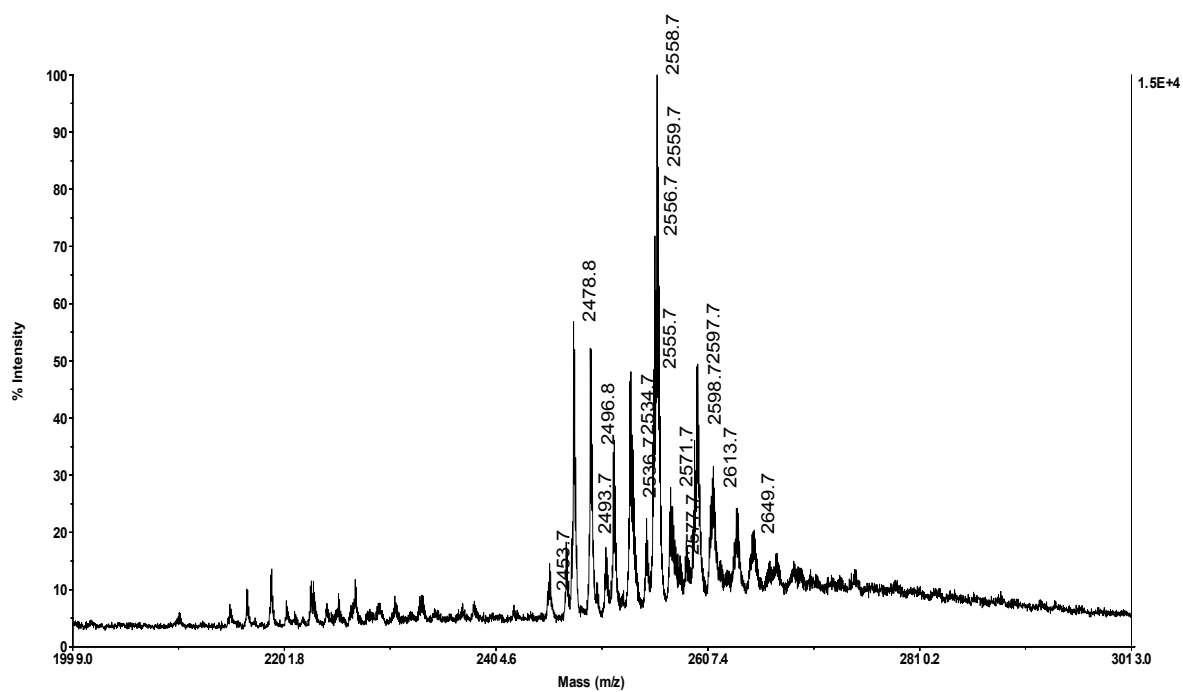
APPENDIX 25. ^{1}H-NMR of ^{2Se}UTP (with Na$^+$ and triethylammonium as counter ions).
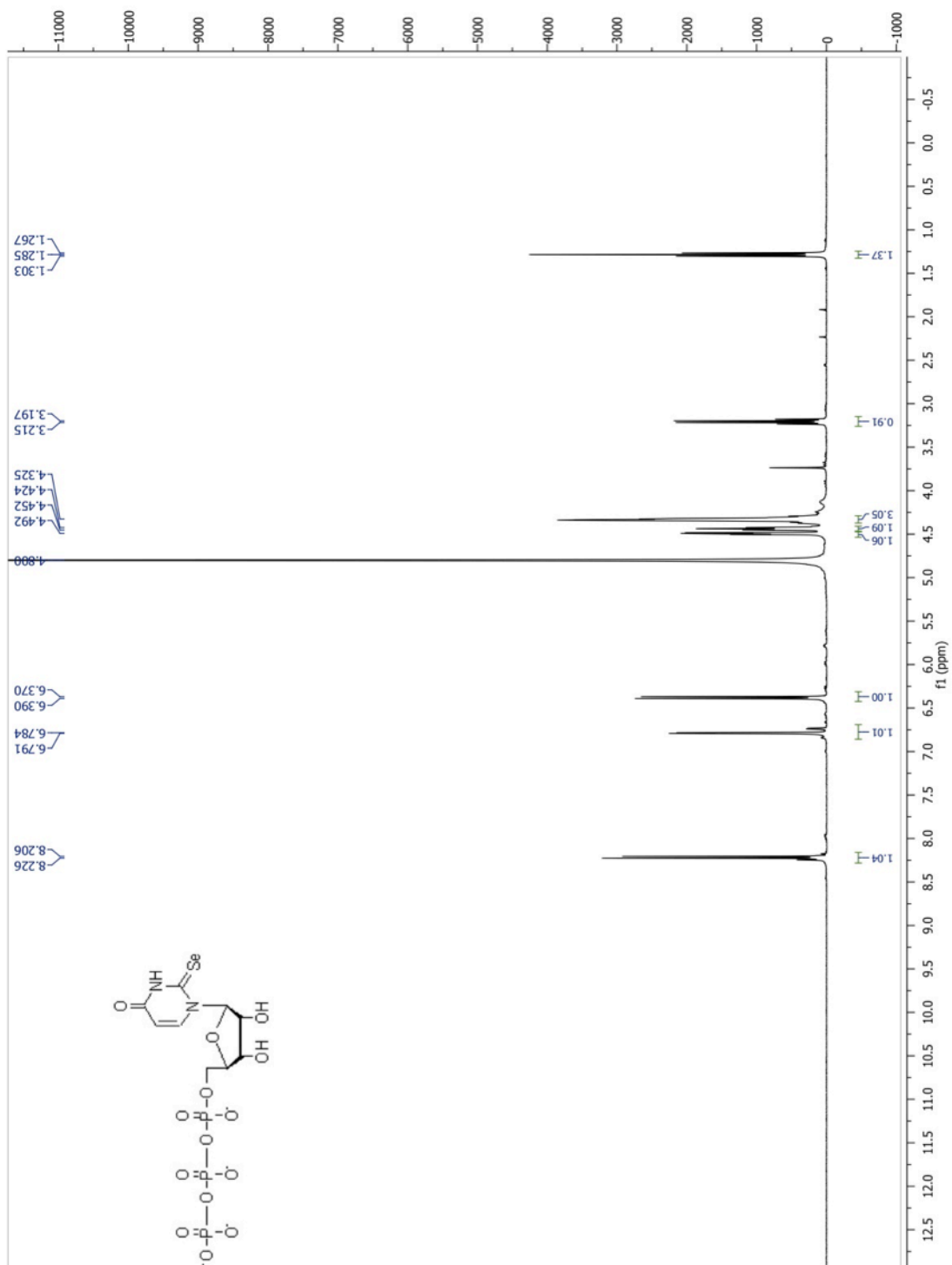
APPENDIX 26. ^{13}C-NMR of ^{2Se}UTP (with Na$^+$ and triethylammonium as counter ions).
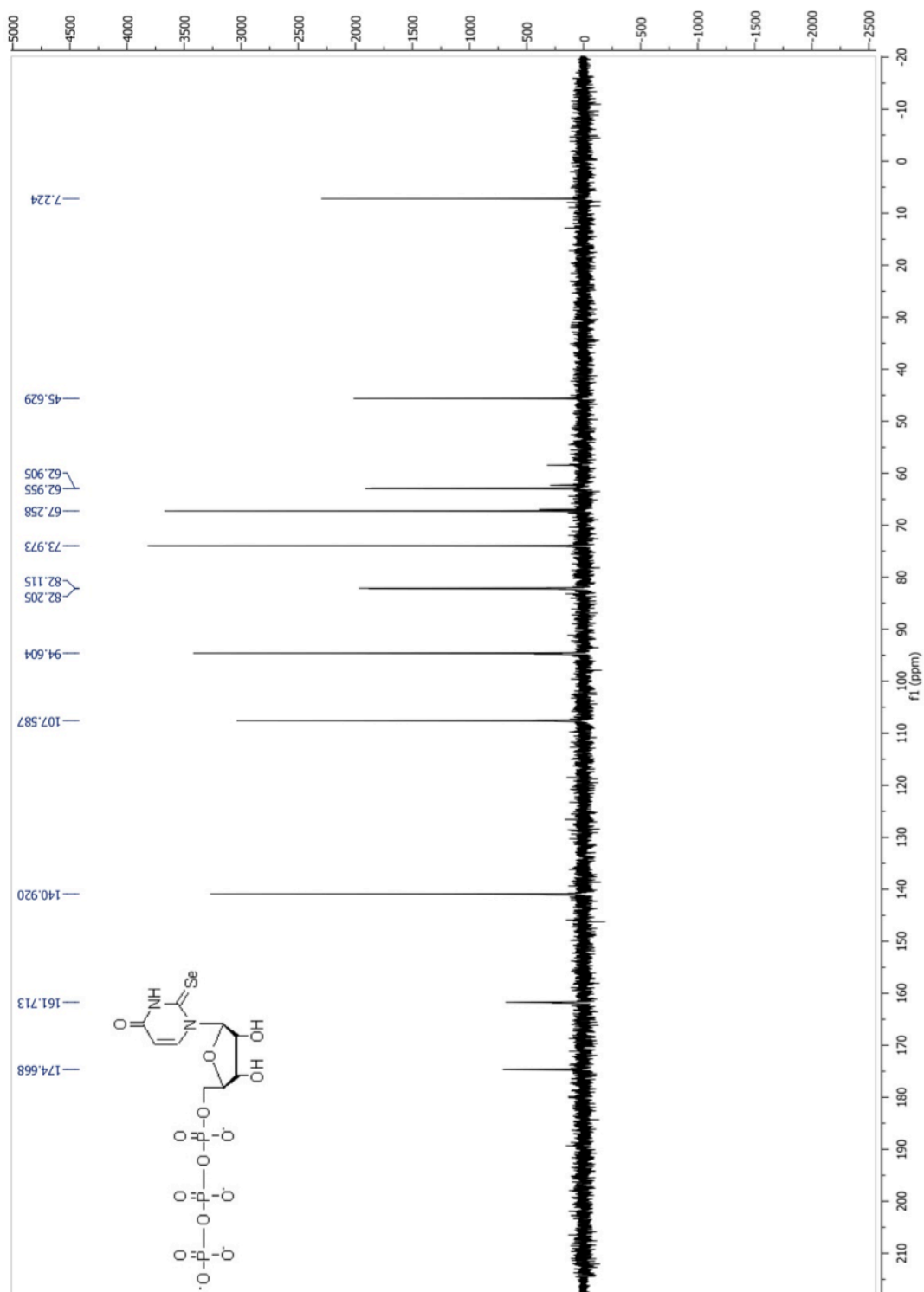
APPENDIX 27. ^{31}P-NMR of ^{2Se}UTP (with Na$^+$ and triethylammonium as counter ions).
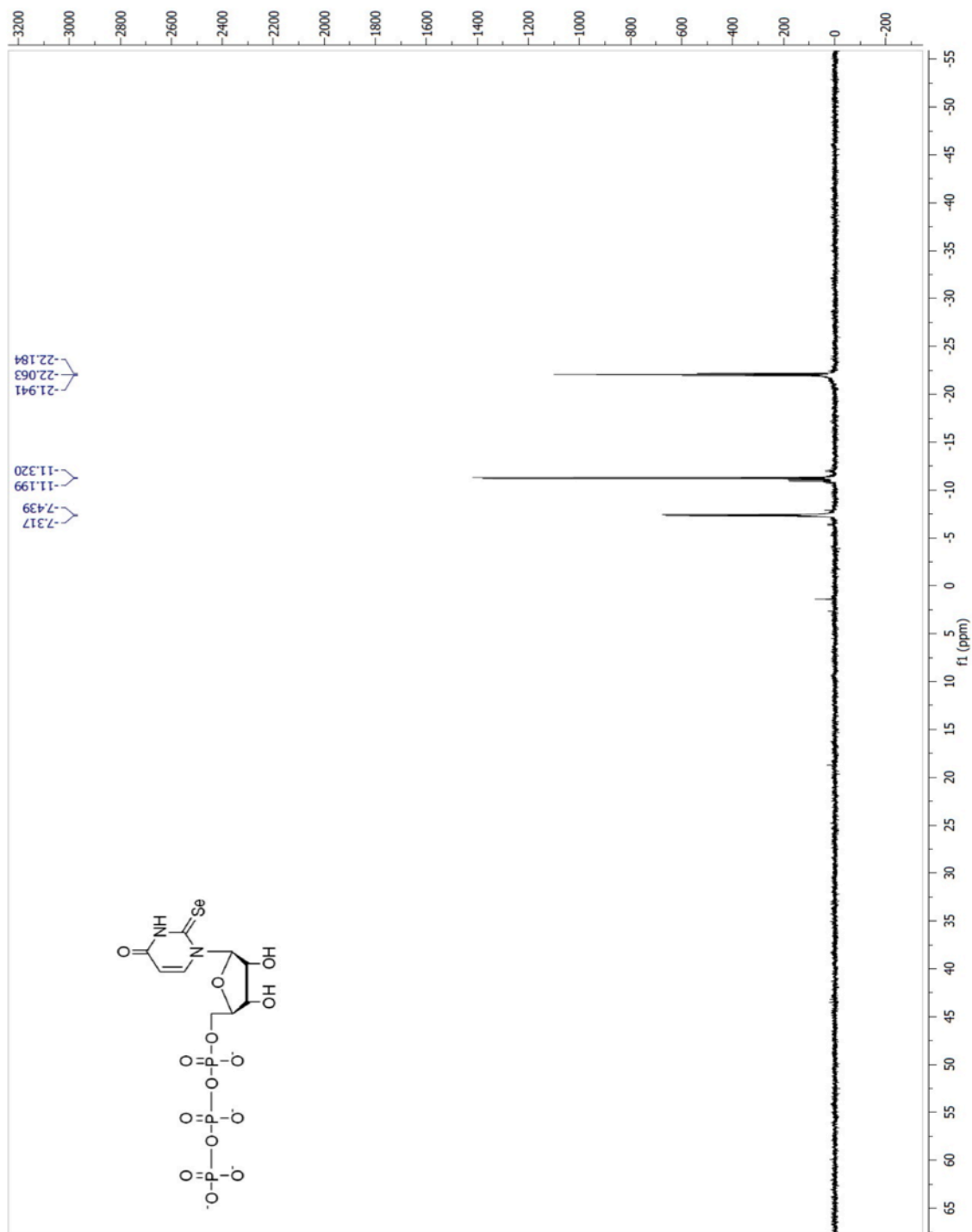
APPENDIX 28. Mass spectrum of ^{2Se}UTP. Molecular formula: $C_9H_{14}N_2O_{14}P_3Se^-$. HRMS (ESI-TOF): [M-H$^+$]$^-$ = 546.8812 (calc. 546.8829).
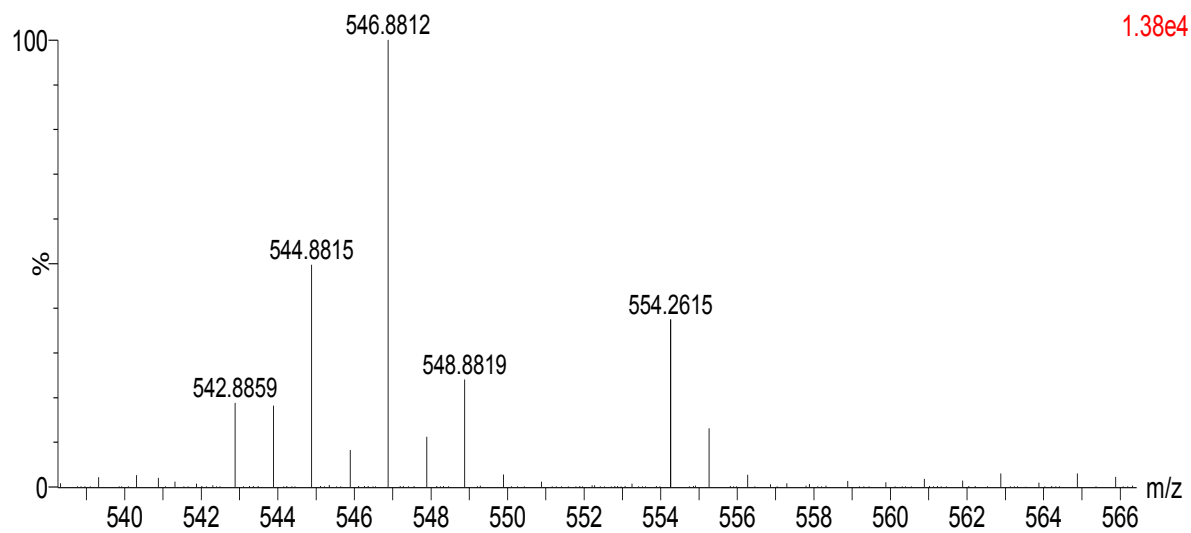
APPENDIX 29. ^{1}H NMR spectra of 5'-O-4,4'-dimethoxytrity -4-cyanoethylselanyluridine.
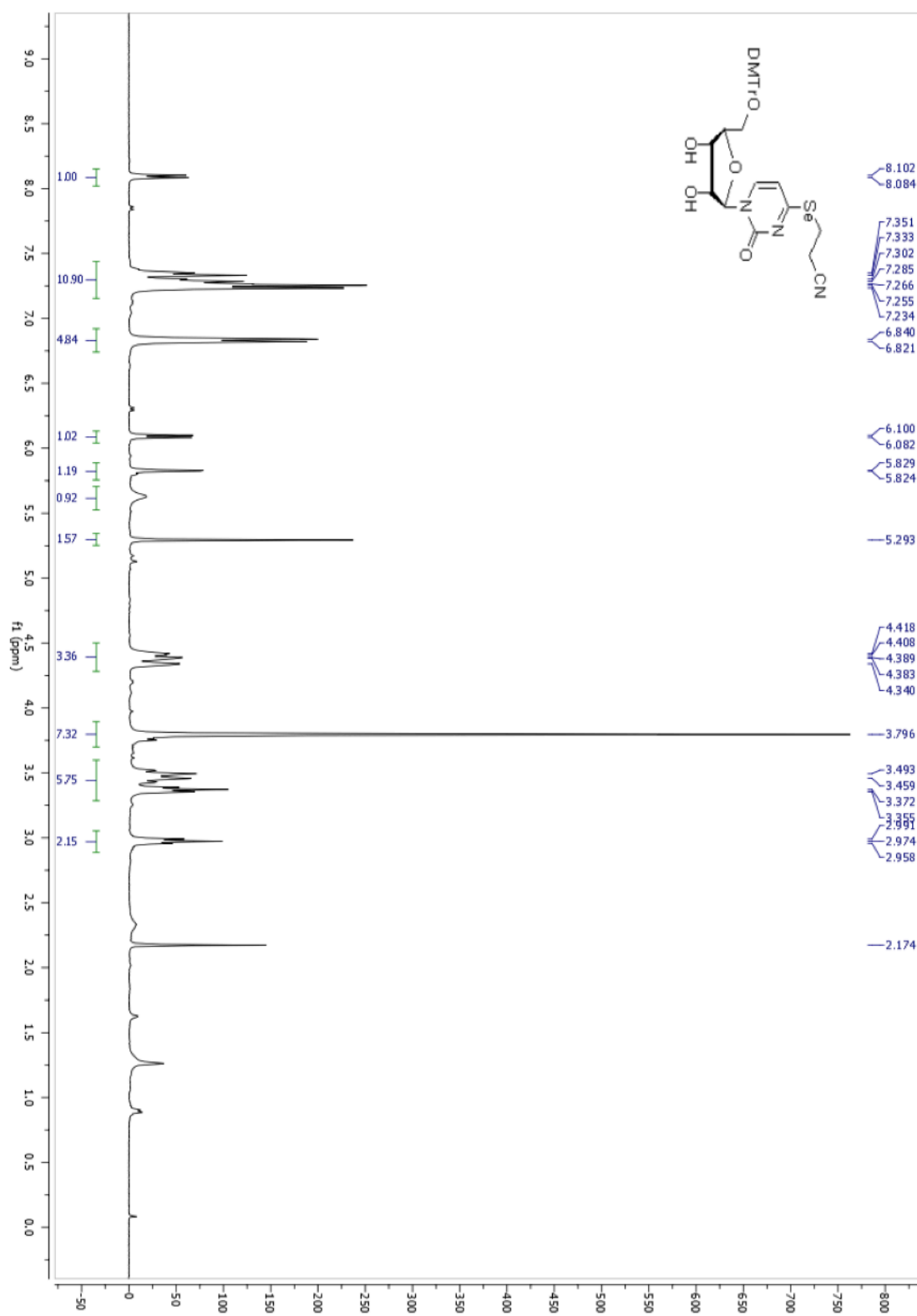
APPENDIX 30. ^{13}C NMR spectra of 5'-*O*-4,4'-dimethoxytrity -4-cyanoethylselanyluridine
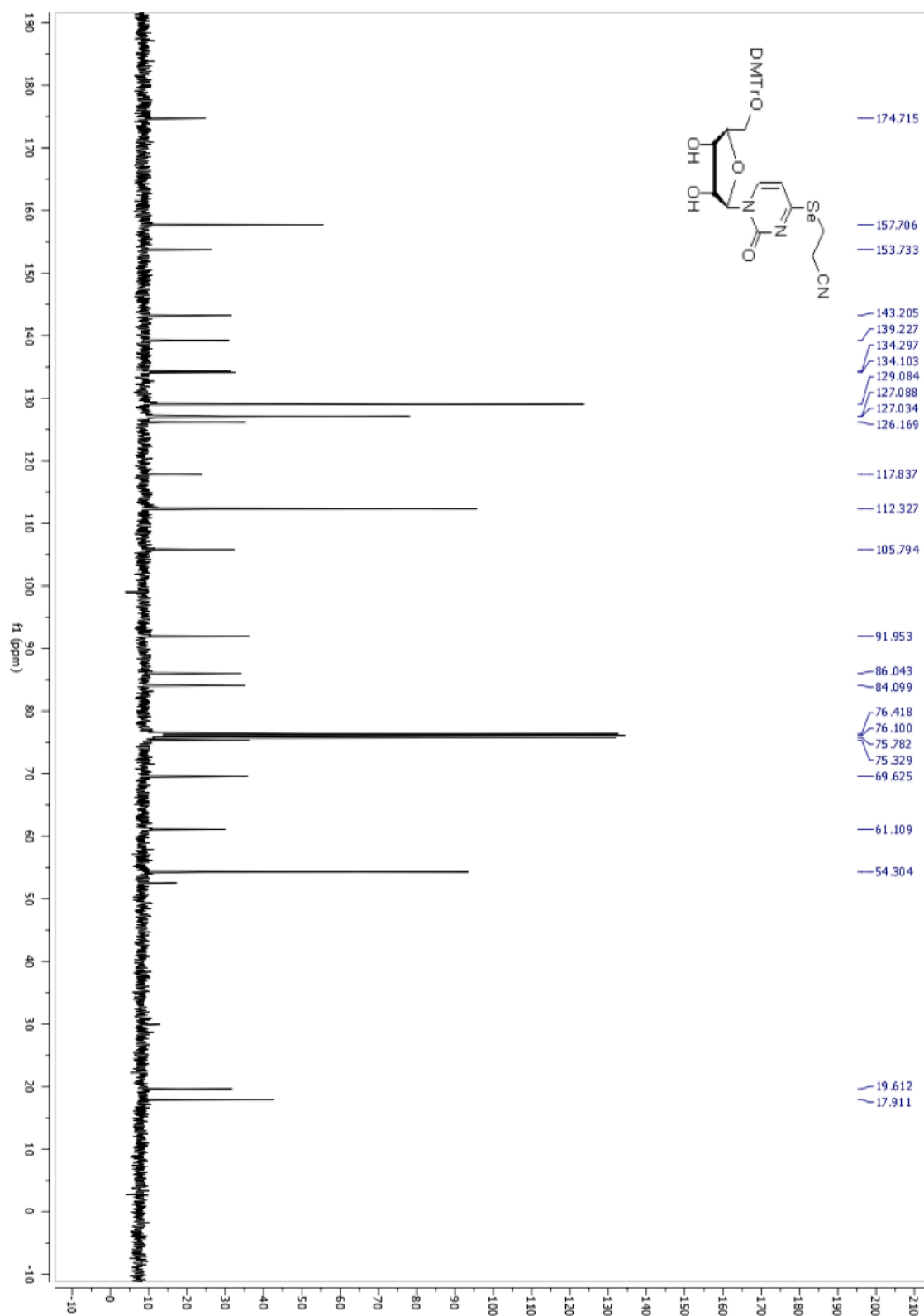
APPENDIX 31. ^{1}H NMR spectra of 4-cyanoethylselanyluridine.
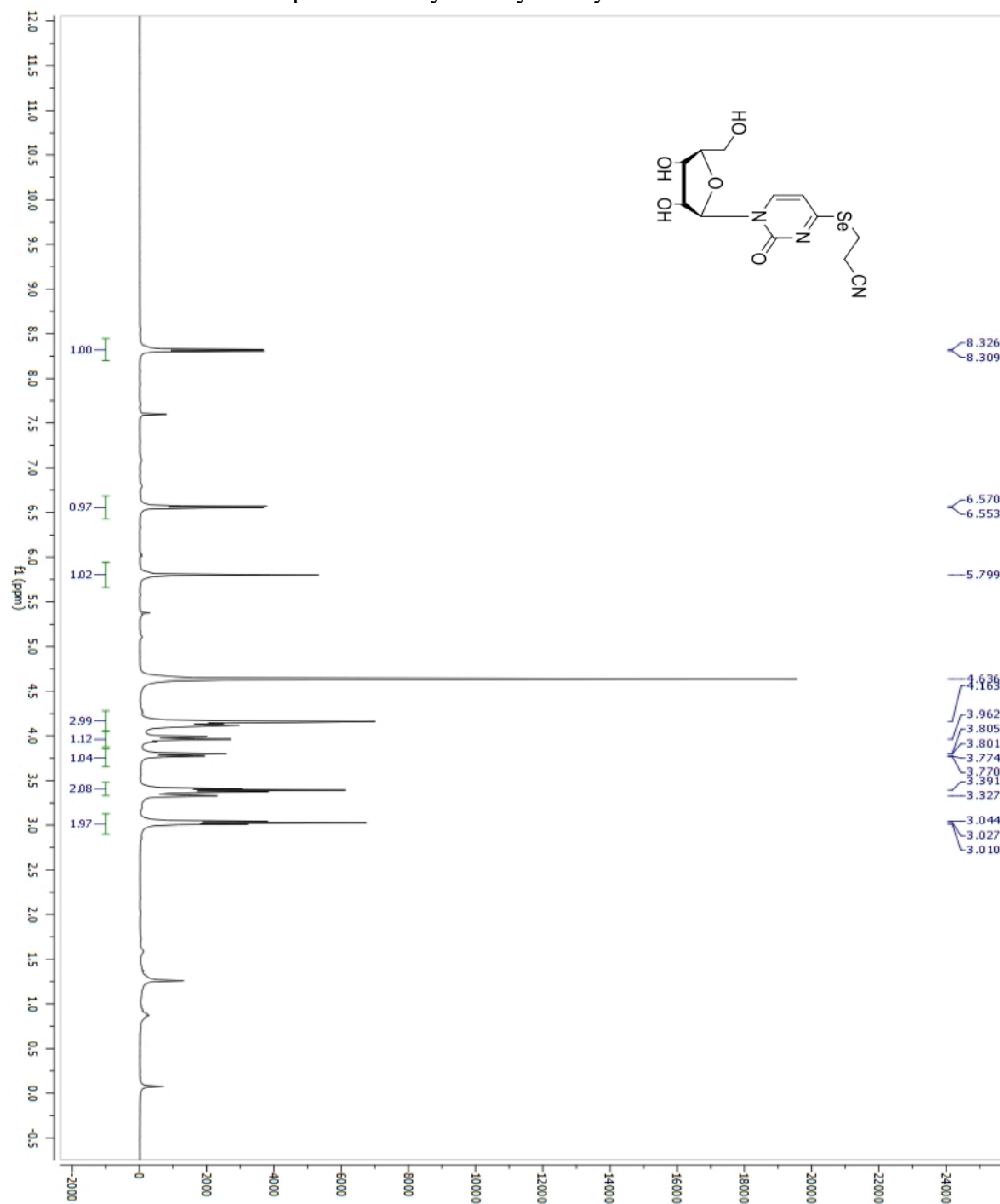
APPENDIX 32. ^{13}C NMR spectra of 4-cyanoethylselanyluridine.
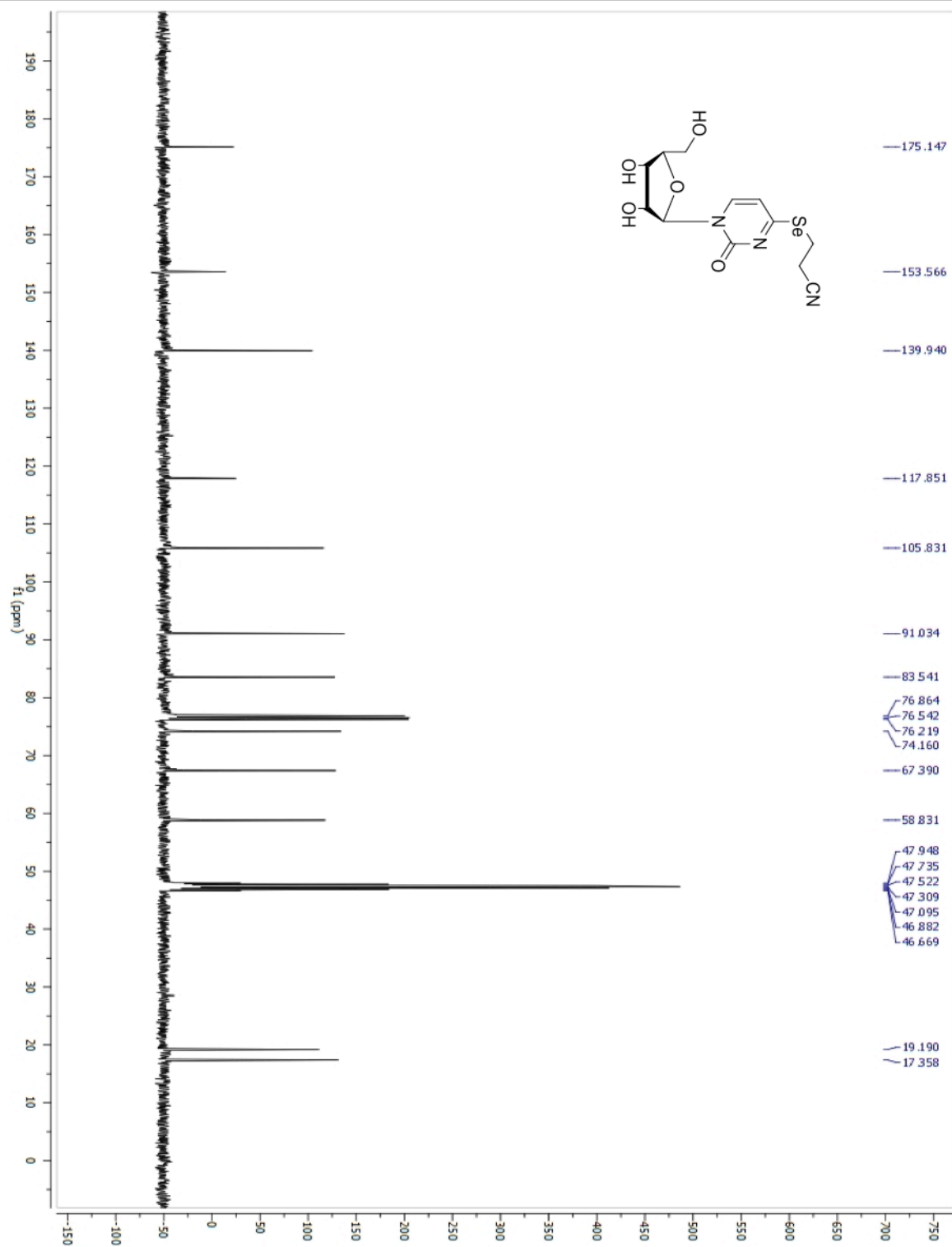
APPENDIX 33. ¹H-NMR of ⁴ˢᵉᶜᴴ²ᶜᴴ²ᶜᴺUTP (with Na⁺ and triethylammonium as counter ions).
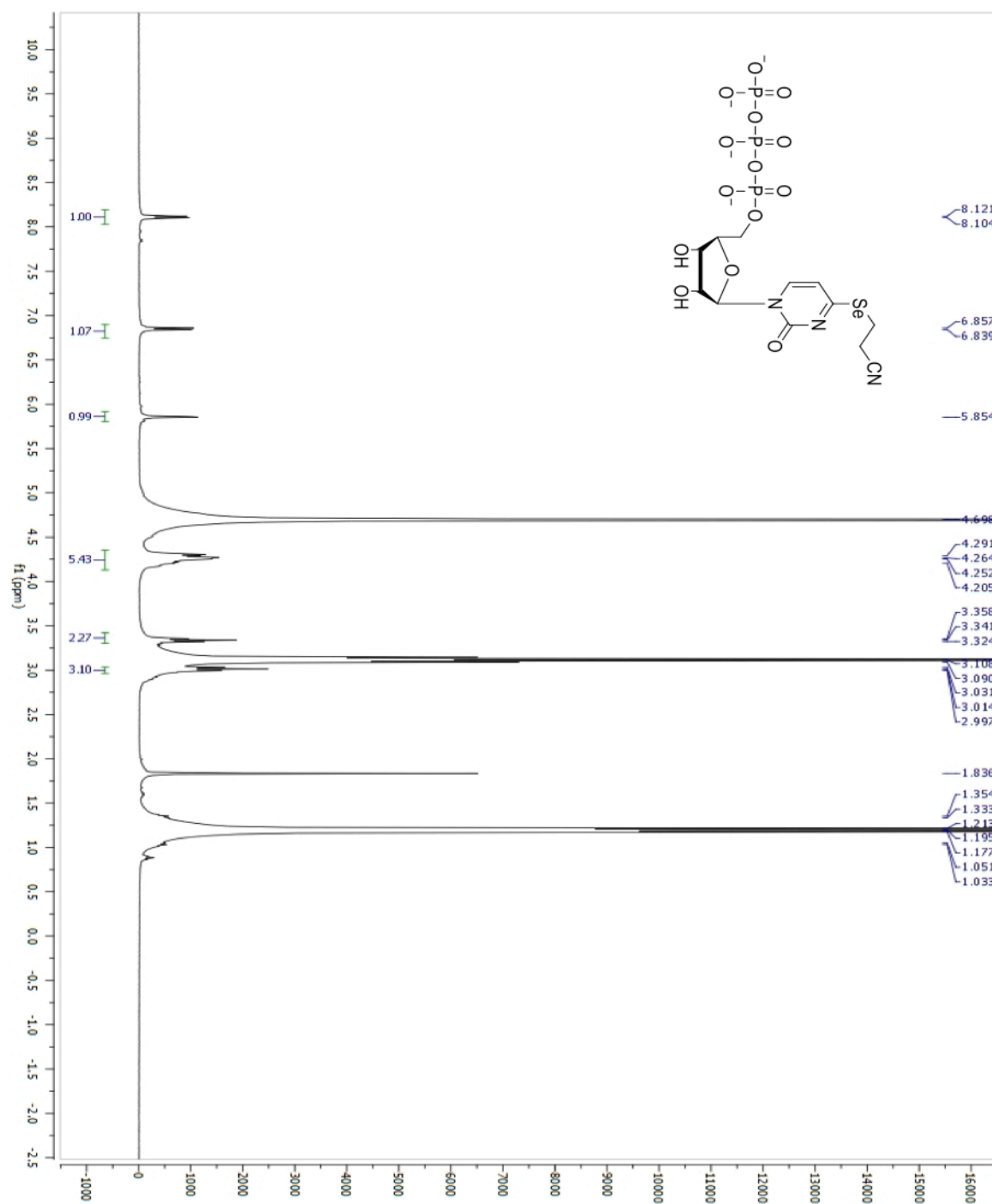
APPENDIX 34. ^{13}C-NMR of ^{4SeCH2CH2CN}UTP (with Na$^+$ and triethylammonium as counter ions).
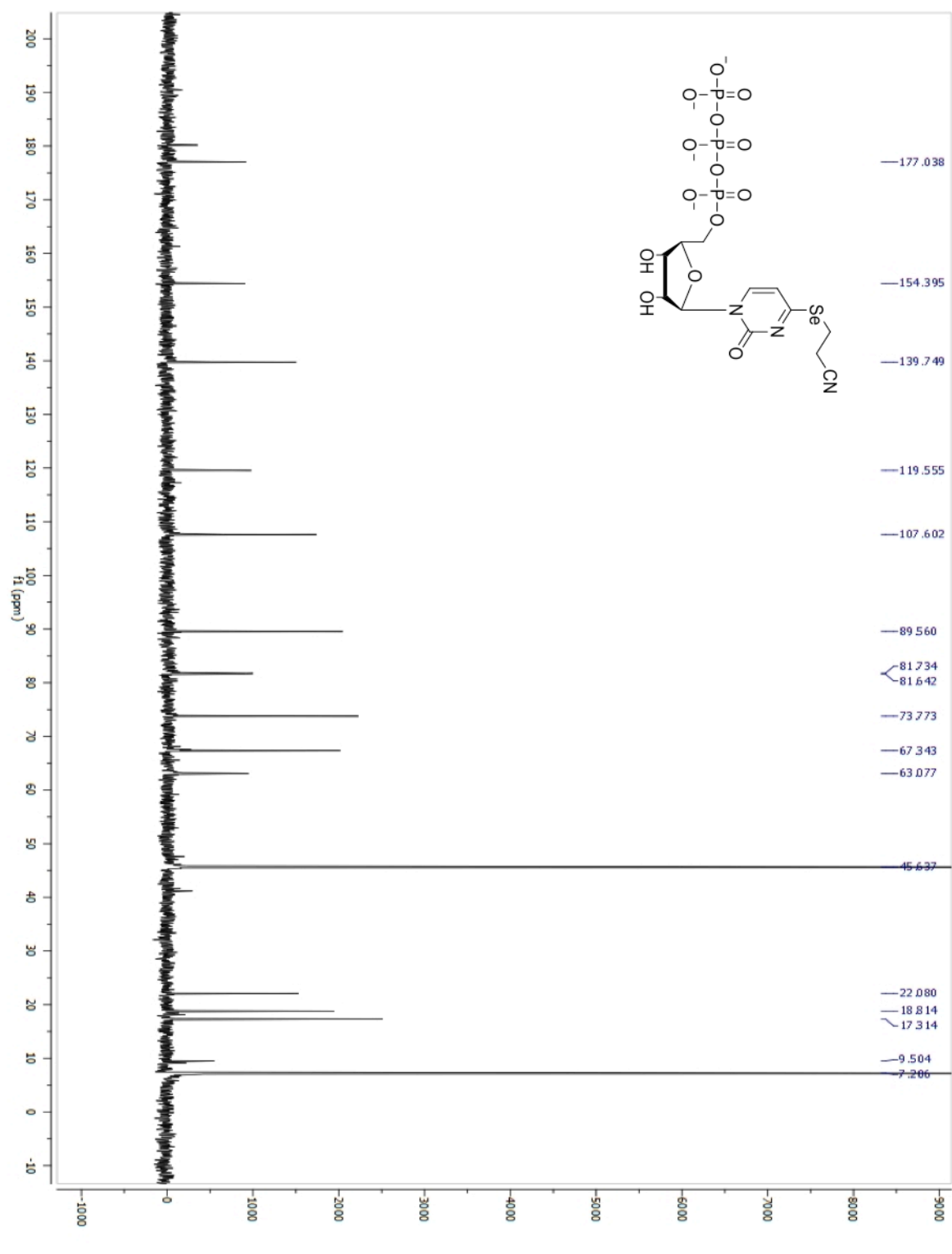
APPENDIX 35. ^{31}P-NMR of ^{4SeCH2CH2CN}UTP (with Na$^+$ and triethylammonium as counter ions).
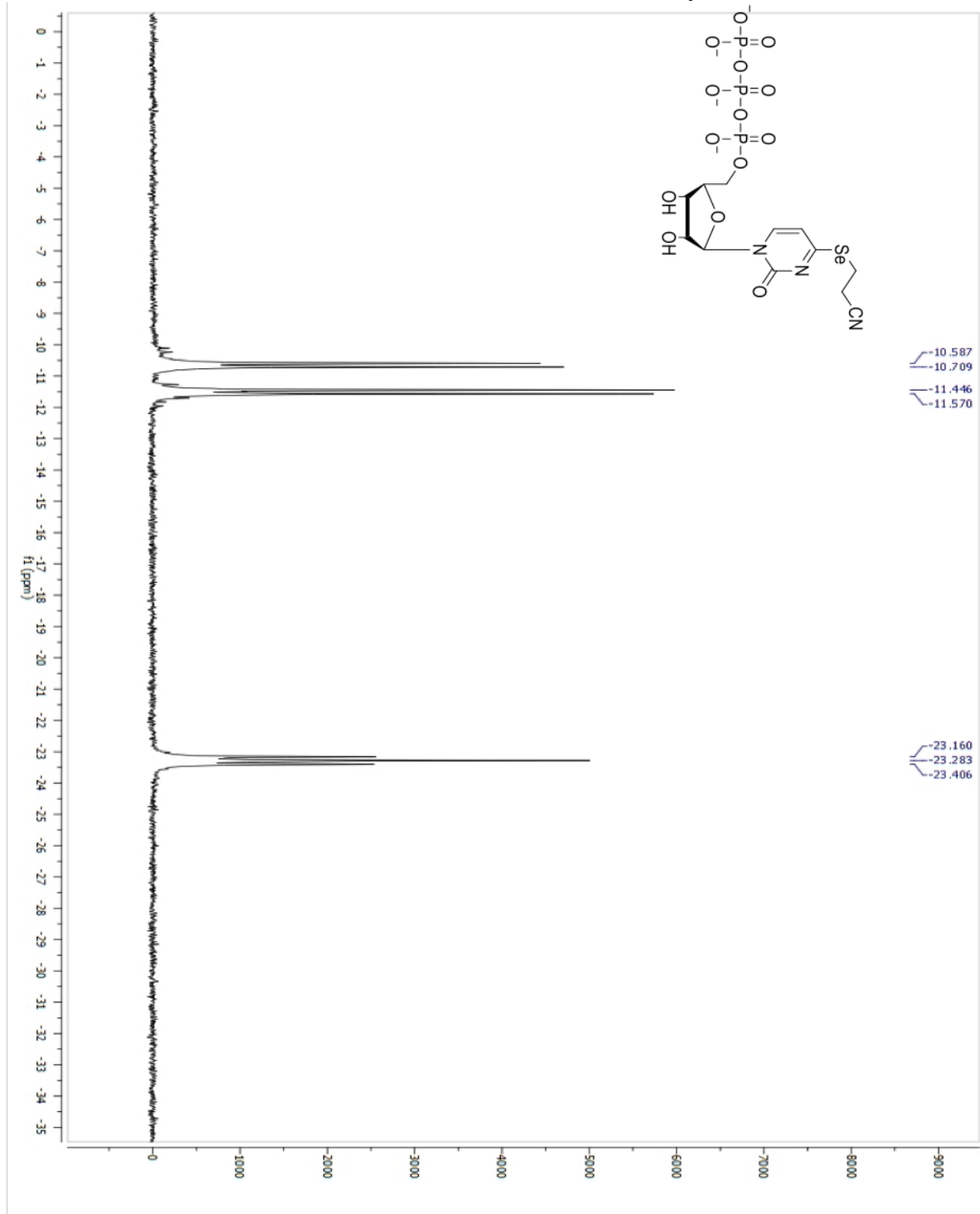
APPENDIX 36. ¹H-NMR of ⁴ˢᵉUTP (with Na⁺ and triethylammonium as counter ions).
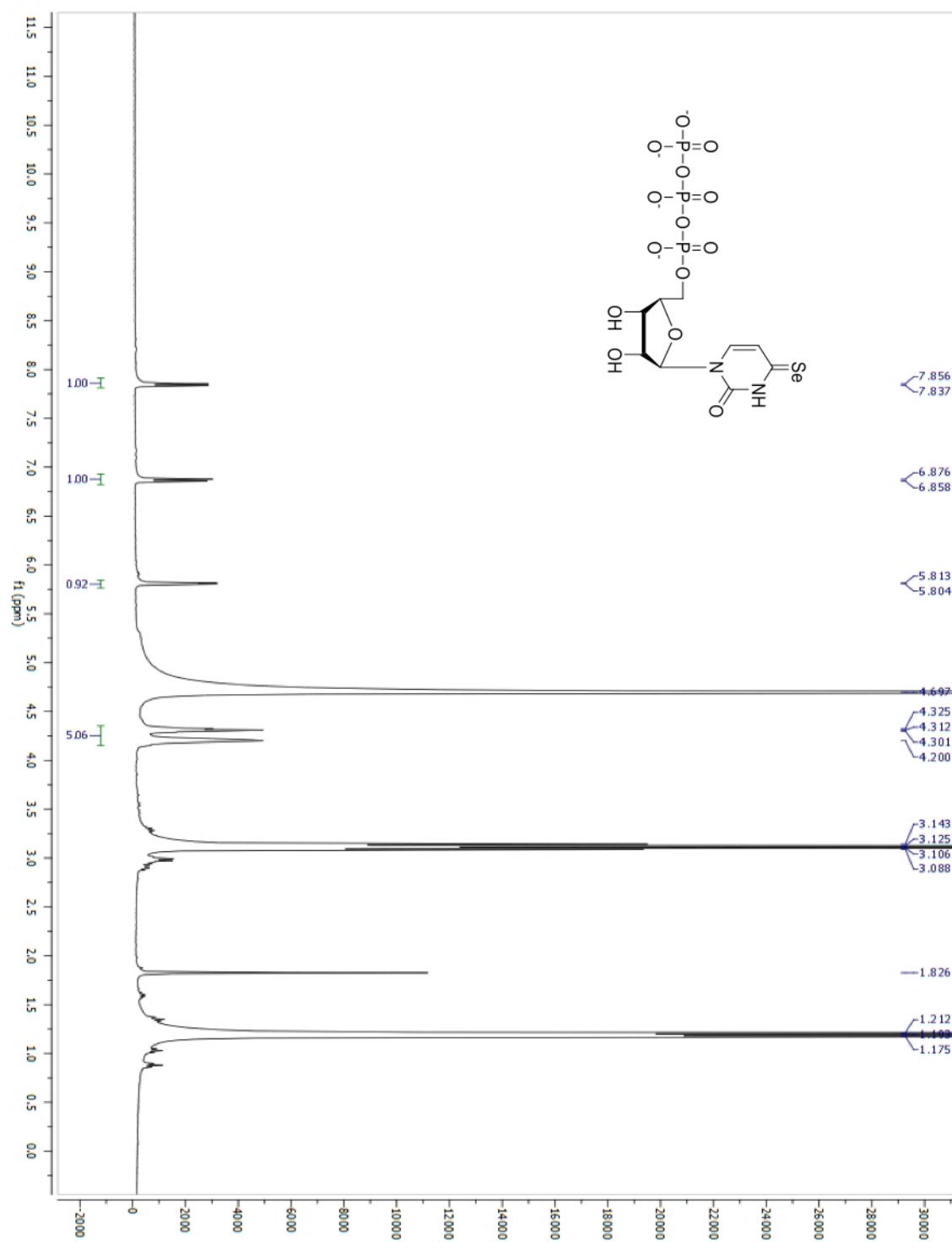
APPENDIX 37. ^{13}C-NMR of ^{4Se}UTP (with Na$^+$ and triethylammonium as counter ions).
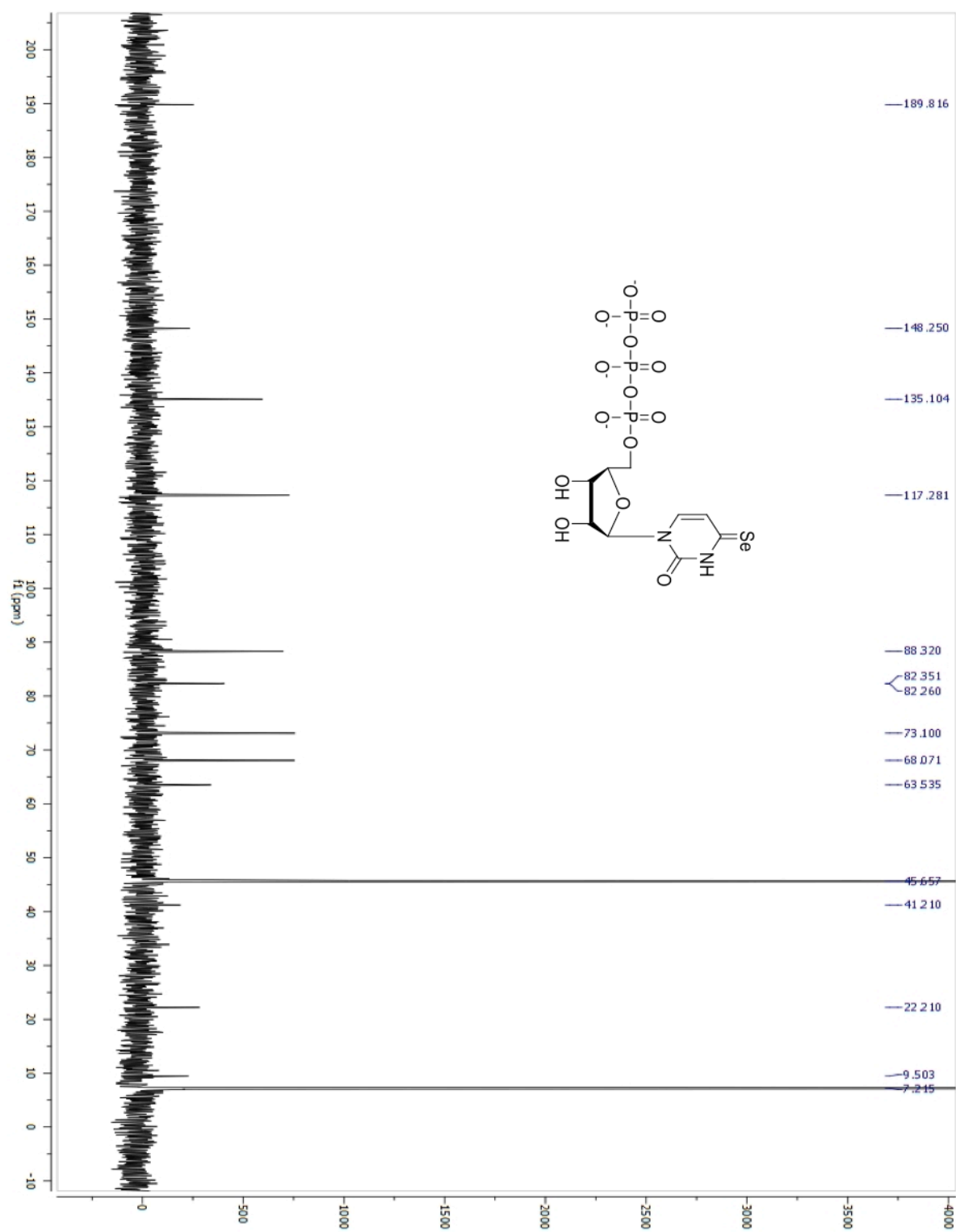
APPENDIX 38. ^{31}P-NMR of ^{4Se}UTP (with Na$^+$ and triethylammonium as counter ions).
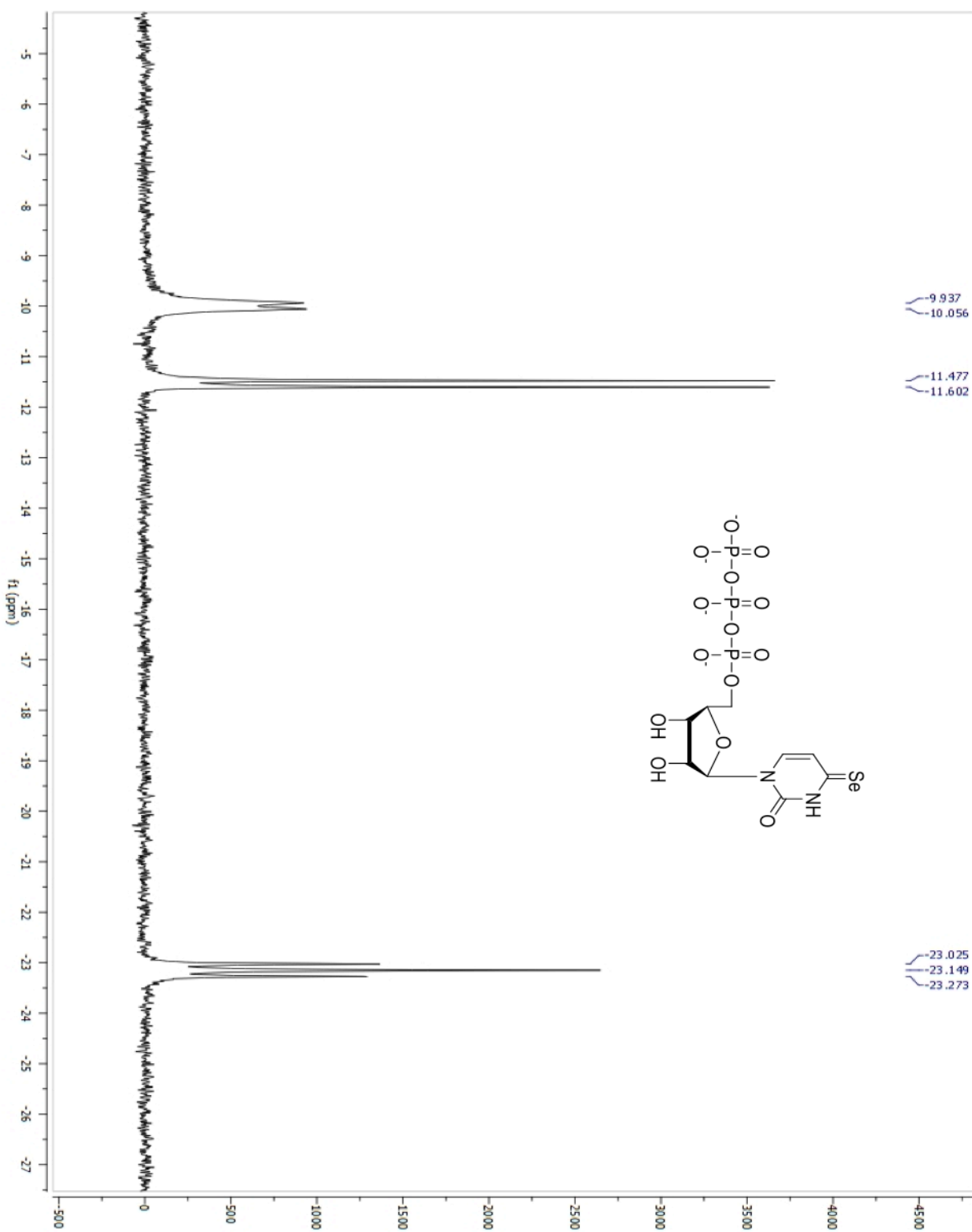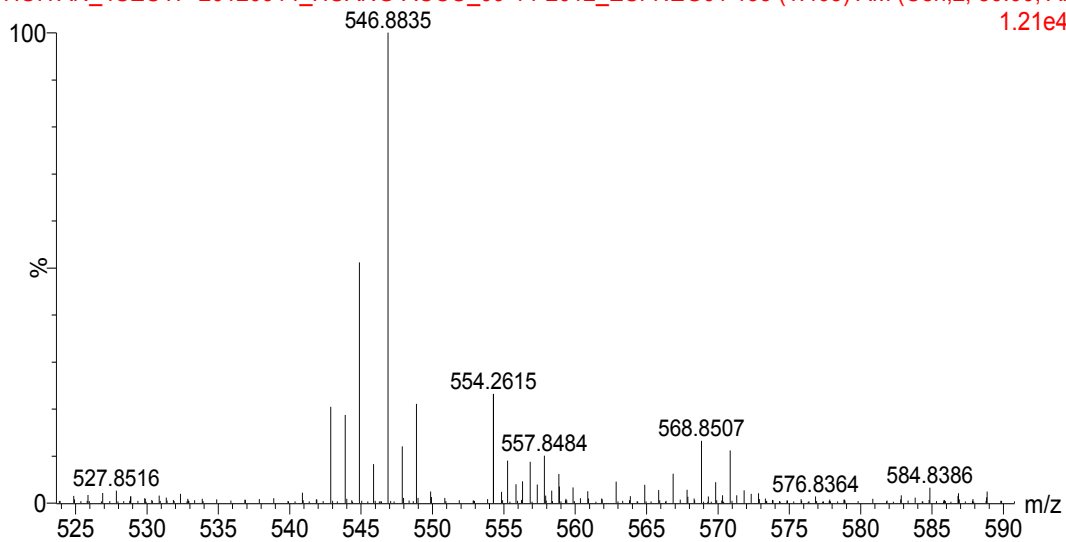
15:33:45 17-Aug-2012

HUIYAN_4SEUTP-D1_HUANG-ACCU_08-17-2012_ESI-NEG01 148 (1.581) AM (Cen,2, 80.00, Ar,5000.0

1.03e3
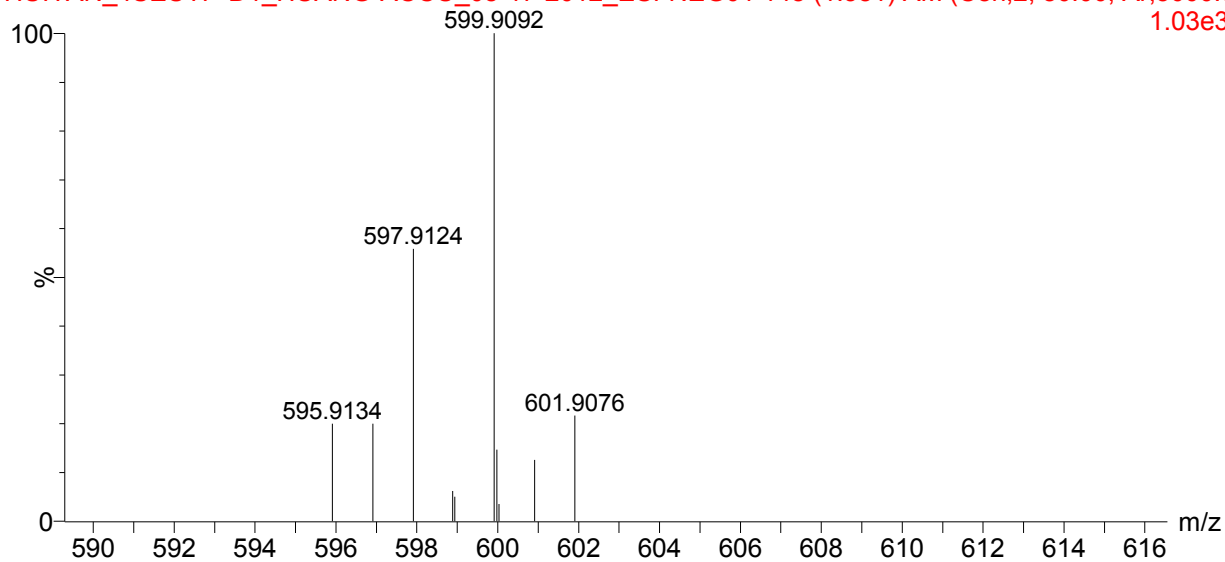
APPENDIX 39. Mass spectrum of ^{4SeCH2CH2CN}UTP. HRMS (ESI-TOF): $[M-H^+]^- = 599.9092$ (calc. 599.9094).

20x in 50%MeOH+0.5%NH4OH

14:57:48 14-Sep-2012

HUIYAN_4SEUTP-20120914_HUANG-ACCU_09-14-2012_ESI-NEG01 133 (1.405) AM (Cen,2, 80.00, Ar

1.21e4

APPENDIX 40. Mass spectrum of ^{4Se}UTP. HRMS (ESI-TOF): $[M-H^+]^- = 546.8835$ (calc. 546.8829).