4-21-2009

Synthesis, Structure and Function Studies of Selenium and Tellurium Derivatized Nucleic Acids

Jia Sheng

SYNTHESIS, STRUCTURE AND FUNCTION STUDIES OF SELENIUM AND TELLURIUM DERIVATIZED NUCLEIC ACIDS

by

JIA SHENG

Under the Direction of Professor Zhen Huang

ABSTRACT

Nucleic acids play important roles in living systems by storing and transferring genetic information and directing protein synthesis. Recently, it was found that nucleic acids can catalyze chemical and biochemical reactions similar to protein enzymes. In addition, they can also serve as drug targets for the treatment of deadly diseases such as AIDS and cancers. As a result, the 3D structure study of nucleic acids and protein-nucleic acids complexes by X-ray crystallography has become one of the most active research areas.

However, the two intrinsic bottlenecks of macromolecule X-ray crystallography, including crystallization and phase determination, have significantly limited its application in study and discovery of the new structures and folds, as well as in exploration of the biological mechanisms. So far, the selenium derivatization (Se-Met) of

proteins and multiple anomalous dispersion (MAD) or single anomalous dispersion (SAD) technology have revolutionized the protein crystallography field by providing a rational solution to solve the phase determination problem. Similarly, it's important and urgent to develop a corresponding methodology for nucleic acid X-ray crystallography.

The work presented here includes two general research directions: the selenium-derivatized nucleic acids (SeNA) and the tellurium-derivatized nucleic acids (TeNA):

1) The SeNA strategy by site-specifically replacing oxygen with selenium at the 2' and 4 positions of thymidine and uridine has been developed. We found that the selenium derivatization at both sites are relatively stable and doesn't cause significant structure perturbations by comparing with their corresponding native counterparts. In addition to the phase determination, the 2'-Se modification can also facilitate crystal growth of many oligonucleotides. Moreover, we have observed colorful DNAs and RNAs with the 4-Se modification for the first time.

2) The TeNA strategy by covalently incorporating tellurium functionalities into different positions of nucleic acids, particularly at the 2' and 5 position of thymidine, has been developed. We have demonstrated the compatibility of the tellurium modification and solid-phase synthesis, as well as the potential application of the tellurium modifications in anti-viral drug synthesis and DNA-damage investigation.

INDEX WORDS: Selenium, Tellurium, Nucleic acids, DNA, RNA, 3D-structure, X-ray crystallography, Anti-virus drug, DNA-damage

SYNTHESIS, STRUCTURE AND FUNCTION STUDIES OF SELENIUM AND TELLURIUM DERIVATIZED NUCLEIC ACIDS

by

JIA SHENG

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

2009

SYNTHESIS, STRUCTURE AND FUNCTION STUDIES OF SELENIUM AND TELLURIUM DERIVATIZED NUCLEIC ACIDS

by

JIA SHENG

Committee Chair: Dr. Zhen Huang

Committee: Dr. Al Baumstark
Dr. David W. Wilson
Dr. Lucjan Strekowski

Electronic Version Approved:

Office of Graduate Studies
College of Arts and Sciences
Georgia State University
May 2009

ACKNOWLEDGEMENTS

All of the work in this dissertation was carried out under the direction of Prof. Zhen Huang. I am deeply indebted to Prof. Zhen Huang, whose valuable suggestions and tireless encouragement helped me during all the time that I carried out this research. His enthusiasm and love for science have been major driving forces through my graduate career. I would like to thank Prof. Al Baumstark, Prof. David W. Wilson and Prof. Lucjan Strekowski who gave me their full support and guidance in performing this research. This dissertation would not have been possible without their guidance. Many people from the faculty, staff and colleagues of the Department of Chemistry have assisted and encouraged me in various ways during my graduate years. I gladly express my gratitude to them for their help, support, interest and valuable discussions. Especially, I am obliged to Dr. Jozef Salon for his help in many ways, to Dr. Abdalla E. A. Hassan for his great help in hands-on teaching and discussions, to Dr. Sekar Chandrasekaran for his help in NMR data collection, to Julianne Caton-Williams, Lina Lin, Sarah Spencer, Rob Abdur and Wen Zhang for their discussions and helps in lab research. I would also like to thank Dr. Jiansheng Jiang, Dr. Anand M. Saxena, and Dr. Alexei Soares at Brookhaven National Laboratory, Dr. Tatsuya Maehigashi at Georgia Institution of Technology, Ying Zhang and Guoxing Fu at Georgia State University for their help in crystal diffraction data collection, processing and structure determination. This work is supported by grants from the NIH (GM069703) and NSF (CHE-0750235, MCB-0517092 and MCB-0824837). Finally, my very special thanks go out to my parents Xiaoniu Sheng and Huijian Hong, my wife Lijun Ni, and my brother Zhen Sheng, his wife Qin Wang, whose patient love and inspiration enabled me to complete this work.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SCHEMES

LIST OF ABBREVIATIONS

BNL	Brookhaven National Laboratory
5-BMT	5-(benzylthio)-1H-tetrazole
CD	Circular dichroism
CPG	Controlled pore glass
CNS	Crystallography & NMR System
DMAP	4-Dimethylaminopyridine
DMF	Dimethylformamide
DMTr	Dimethoxytrityl
DNA	Deoxyribonucleic acid
DIEA	N,N-Diisopropylethylamine
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
ESI	Electrospray ionization
HR-MS	High resolution mass spectrum
$LiAlH_4$	Lithium aluminium hydride
MAD	Multi-wavelength anomalous dispersion
MALDI	Matrix-Assisted Laser Desorption/ionization
MIR	Molecular isomorphorous replacement
MPD	2-Methyl-2,4-Pentanediol
MR	Molecular replacement
MSCl	Mesyl chloride
$NaBH_4$	Sodium borohydride
NMR	Nuclear magnetic resonance
NSLS	National Synchrotron Light Source
PAC	Phenoxyacetyl
PEG	Polyethylene glycol
RNA	Ribonucleic acid
RP-HPLC	Reverse phase high-performance liquid chromatography
SAD	Single-wavelength anomalous dispersion
SeNA	Selenium modified nucleic acids
TBAF	Tetra-*n*-butylammonium fluoride
TBDMS	*tert*-butyldimethylsilyl
TCA	Trichloroacetic acid
TEA	Triethyl amine
TeNA	Tellurium modified nucleic acids
TFA	Trifluoroacetic acid
THF	Tetrahydrofuran
TMS-Im	Trimethylsilyl imidazole
TIBSCl	2,4,6-(Triisopropylbenzene)sulfonyl chloride
TIPSCl	Triisopropylsilyl chloride
Tf	Trifluoromethanesulfonyl
TLC	Thin layer chromatography
UV	Ultraviolet

1. INTRODUCTION

1.1 Nucleic acids and macromolecule X-ray crystallography

Nucleic acids play important roles in cells by storing and transferring genetic information and directing the protein synthesis, as well as catalytic functions. (1) (2) (3) Nucleic acid research helps to better understand diseases and identify novel drug targets for development of therapeutics and the treatment of some deadly world-wide epidemic diseases, such as AIDS and cancers. Instead of the traditional attacking of aberrant proteins, modulating specific gene expressions of diseases and cancers can lead to better treatment. (4) More importantly, nucleic acids themselves can be used as therapeutics by serving as potential antisense, antigene, aptamer, ribozyme and siRNA agents. (5) Research activities in this area require further development of new nucleic acid analogs with improved properties.

3D structural investigations at the atomic level can normally provide useful guidance for functional studies. X-ray crystallography, NMR, electron microscopy and atomic force microscopy are the major methodologies used for macromolecular 3D structure studies. (6) Among them, X-ray crystallography continues to be the major and the most powerful investigation tools in nucleic acid research, which allows us to better understand biological processes at the atomic level, including mechanisms of replication initiation, transcription and translation regulation, DNA-drug complexes, catalytic RNAs, riboswitch function, and nucleic acid-protein interactions. It has also provided fundamental insights into hydration, stereoelectronic effects and conformational preorganization. (7) The nucleic acid research has provided guiding principles for the optimal design of small molecular drugs and the nucleic acid-based therapeutics.

Especially in the past few decades, development in instrumentation, including area detectors, advanced computers, graphical display units and synchrotron radiation, has greatly contributed to the success of X-ray crystallography, which provides obvious advantages over other methodologies.(8)

Although powerful it is, there are still two major challenges in the field of X-ray crystallography. One is crystallization to get high quality of macromolecular crystals, for which, there are no rational approaches so far except the regular trial-and-error screen strategy, because there are so many factors can affect the crystal growth, such as temperature, concentration, buffer conditions and even the noise. The other one is called phase problem, which actually is a more serious bottleneck and has largely slowed down the structure determination of new macromolecules. This problem was induced by the fact that the precise phase angle is not detectable from the diffraction pattern of macromolecules. (9) (10) But phase information is needed to calculate the 3D electron densities of macromolecules structure through Fourier transformation after collecting diffraction data.

Most of the traditional ways to solve this phase problem were based on the existence of an endogenous heavy atom (such as selenomethionine residues in the proteins) or the incorporation of other heavy atoms which would occupying enough sites or space in the molecule to give a clear signal. (11) Actually, all the nowadays popular used methods including multiple and single isomorphous replacement (MIR, SIR), multiple and single isomorphous replacement with anomalous scattering (MIRAS, SIRAS) and multi- and single-wavelength anomalous dispersion (MAD, SAD) require heavy atoms derivatization. (12) So far, besides the powerful “magic seven” derivatives, a useful heavy atom databank has already been founded for protein crystallography. (13)

Comparing with protein phasing, there are fewer strategies for nucleic acids phasing. Most of the heavy atoms soaking and co-crystallization, which usually very powerful for protein crystallography, have been proved to be not very successful in nucleic acid case, probably due to the lack of the large solvent channels in nucleic acid structure, which allows diffusion of metal ions into crystals in most of proteins, as well as the limitation of specific binding sites for ions in nucleic acid. In addition, these bindings could also cause the backbone cleavage of nucleic acids.

People used to believed that halogenated strategy (bromination and iodination) in combination with multi- or single-wavelength anomalous dispersion (MAD or SAD, respectively) are the most convenient and powerful method for the determination of nucleic acid new structures which avoid the requirement of at least three isomorphous crystals to use MIR methodology (multiple isomorphous replacement). (14) However, since the chemical incorporation of halogen atoms is primarily limited to the 5-position of pyrimidine nucleosides due to their stability, this lack of multiple choice of positions greatly limited derivatization freedom to avoid base-stacking disruptions, other structural perturbations, and possible crystallization problems. Moreover, in the majority of cases, halogen derivatives cannot be crystallized under native conditions and the derivative crystals sometimes do not diffract as well as the native ones. In addition, recent reports have indicated that the light sensitivity of halogen derivatives can lead to their decomposition after long-time exposure to even moderate X-ray dose. (15) Therefore, it’s very necessary to develop an alternative and more powerful method for the crystallography structure studies of nucleic acids, this method could also be applied to determine the structure of protein-nucleic acids complexes and some more complicated biomacromolecule systems.

1.2 SeNA: A powerful tool to get the nucleic acids structure

1.2.1 Brief introduction of selenium in chemistry and biochemistry

Selenium (Se), following oxygen and sulfur, is an element located in VIA group of periodic table. Although it was discovered early in 1817 by Jons Jakob Berzeliu, (16) owing to the toxic nature of Se (found in the Se-accumulating plants), the early research interest in this area was generally low. Between the 1930s and mid 1950s, the animal nutritionists and chemists were curious to know the structures and functions of the Se compounds present in those Se-containing plants. (17) However, these studies were not successful and failed to identify any Se compounds, probably due to their instability under the isolation conditions. Since the discoveries of the presence of Se in proteins and key roles that Se played in a number of physiological processes in mid 1950s, there has been a growing interest in Se-related research. (18) (19)

Although it is toxic in large doses, (exceeding the tolerable upper intake level of 400 micrograms per day lead to selenosis), selenium is nowadays regarded an essential trace micronutrient element. (20) In plants, it occurs as a bystander mineral, sometimes in toxic proportions in forage (some plants may accumulate selenium as a defense against being eaten by animals, but other plants such as locoweed require selenium, and their growth indicates the presence of selenium in soil). (21) It is a component of the unusual amino acids selenocysteine and selenomethionine. In humans, selenium functions as cofactor for reduction of antioxidant enzymes such as glutathione peroxidases and certain forms of thioredoxin reductase found in animals and some plants (this enzyme occurs in all living organisms, but not all forms of it in plants require selenium). (22) (23) (24) Selenium also plays a role in the functioning of the thyroid gland by participating as a cofactor for three known thyroid hormone deiodinases.

The further health effects such as cancer, HIV, tuberculosis and diabetes related selenium functions are still under controversial, detailed research are still to be carried out to draw the conclusions. (25) Usually, the daily dietary selenium comes from nuts, cereals, meat, fish and eggs.

As oxygen (atomic radius, 0.73 Å) and sulfur (1.02 Å), selenium (1.16 Å) was also found in several tRNAs besides its presence in proteins due to their close chemical and electronic properties. Se in natural tRNAs is present in naturally occurring 2-selenouridine and 5-[(methylami-no)methyl]-2-selenouridine (mnm5se2U). (26) The Se-atoms at C(2) of the modified uridines are introduced by 2-selenouridine synthase via displacement of the S-atoms in the corresponding 2-thiouridine nucleotides of the tRNAs, and selenophosphate is used as the Se donor. (27) (28) Selenium provides both RNA and protein with many unique biochemical and biological functions, and properties. There are ca. 23 known Se-containing proteins in mammals. (29) In most of these enzymes, Se is present in the form of selenocysteine residue. Selenocysteine (Sec, an analog of cysteine, Cys), referred to as the 21st amino acid, exists naturally in all kingdoms of life. (30) Also owing to its large size and chemical and electronic properties, Se was found in mnm5se2U at the anticodon wobble positions of several bacterial tRNAs, (31) including tRNALys, tRNAGlu, and tRNAGln. The 2-seleno functionality on this modified nucleotide most likely improves the translation accuracy and efficiency. These observations in vivo suggest that the presence of Se can provide natural RNA with many useful and unique properties to better function and survival. As themechanism of the Se modification in tRNAs remains a mystery, understanding of the Se presence and function in protein can provide useful information for studying the Se derivatization and function in RNA. The Se-atom in Sec gives quite different properties

from O and S. (32) (33) The most obvious difference is the lower pKa value of the Sec SeH group (pKa 5.2) compared to the Cys SH (pKa 8.5), and its higher reduction potential. (34) (35) Thus, Sec is a much stronger nucleophile and better reductant than cysteine at physiological pH. These differences are the fundamental reasons why the Sec participation in selenoproteins, particularly in the redox proteins, can largely enhance the redox reaction rates, even up to 1000-fold. (36) (37) Obviously, Se plays an important role at the redox centers of several important redox enzymes in both eukaryotes and prokaryotes. (38) Compared to the extensive studies on the Se-derivatized natural proteins, the study and progress on the Se-derivatized naturally occurring RNAs are rather limited due to the Se-RNA availability and the research challenges. The exciting research and progress in protein study, however, have stimulated the research on syntheses and studies of Se-derivatized naturally occurring and unnatural nucleic acids.

1.2.2 Selenium derivatized protein in X-ray crystallography

Another important property of selenium is that it can serve as a good anomalous scattering center in X-ray diffraction with a special K edge of 0.9795 Á, 12.6578 keV. The anomalous scattering is a phenomenon that was led by the resonance between beams of X-ray waves and electronic transitions from bound atomic orbitals. (39) This effect can be exploited in X-ray crystallographic studies on biological macromolecules by making diffraction measurements at selected wavelengths associated with a particular resonant transition. (40) (41) With this manner, the problem of determining the three dimensional structure of thousands of atoms is reduced to that of initially solving for a few anomalous scattering centers that can then be used as a reference for

developing the entire structure. Also based on that, MAD (multiple wavelength anomalous dispersion) was especially applied to provide a definitive experimental solution for the phase problem. This MAD experiment can be thought of as in situ isomorphous replacements (as MIR) in which physics rather than chemistry are used to effect the change in scattering strength at the site. But unlike MIR, a series of changes in common at a single constellation of sites suffices in general for definitive phase determination because of the phase shifts that accompany the anomalous scattering. (42) The MAD method has potential advantages for accuracy in phase evaluation in that isomorphism is perfect, relative scattering strength and phase power increase with scattering angle, diffraction data can be measured from a single crystal and algebraically exact analysis is possible. To better apply this method, the most important task is finding proper heavy elements as resonance centers. (43)

In principle, an accessible spectral range for MAD experiments can be considered to the window from ~0.3 to ~3.0 Å in wavelength (4 to 40 keV in energy). Obviously, the dominant lighter atoms (H, C, N and 0) in macromolecules are inappreciable at this range, and although sulfur anomalous scattering has been used for protein phasing in special cases, its resonance energy is ordinarily inaccessible for MAD experiments. Either intrinsic metal centers (e.g. Fe or Zn) or introduced heavy atoms (e.g. Hg,Au or Gd) are suitable and they can produce dramatic effects. However, metalloproteins or chemical derivatives are not always at hand. (44) Fortunately, systematic biological incorporation of selenomethionine in place of methionine residues in proteins offers the possibility of generality. Cowie and Cohen (1957) showed that a strain of E. coli made auxotrophic for methionine could grow for 100 generations in selenomethionine,

suggesting that all proteins needed to sustain life in this organism can function well when fully substituted with selenomethionine. (45)

Early experiment results of MAD phasing of seleno achieved by Hendrickson et.al., including selenolanthionine (1985) and selenobiotinyl streptavidin (1989) (46) have demonstrated that selenium is very effective and successful as a center for MAD phasing. In 1990, the ribonuclease H structure was also successfully determined by this selenomethionyl MAD analysis, (47) after that, this selenium derivatization has greatly revolutionized the protein X-ray crystallography field. It is estimated that about two thirds of all new crystal structures of proteins are now determined by this strategy. (48)

1.2.3 Selenium derivatized nucleic acids in X-ray crystallography

Since selenium is such an attractive and powerful functionality for X-ray crystallography in combining with MAD, and also because of the above stated special difficulties of nucleic acid crystallography structure determination (lack of metal binding sites, backbone cleavage etc.), naturally, we decided to introduce selenium into nucleic acids and our group has developed and pioneered this exciting area since year of 1998. Selenium functionalities were successfully introduced into different locations of nucleic acids through chemically and enzymatically synthetic method by atom-specific replacement of oxygen with selenium. Here we defined the selenium modified nucleic acids as SeNA.

This selenium derivatization strategy has great potential in structural and functional studies of nucleic acids. So far, our results suggested that this derivatization can become a better alternative to the conventional Br derivatization in terms of stability during the exposure of crystal sample in X-ray, (49) and comparing the limited sites of

bromine modification, the selenium strategy has more site-options since there are so many oxygen atoms available in nucleic acids. In addition, based on a rudimentary calculation on X-ray phasing power of a selenium scatter, one selenium atom can enable phase determination for DNA or RNA up to 30 nucleotides. (50) Since selenium and oxygen are in the same family VIA in periodic table, it's possible that this selenium derivatization doesn't cause significant structure perturbation when replacing oxygen atoms in nucleic acids. Actually, this novel derivatization methodology has been utilized by several laboratories to study the X-ray crystal structures of DNA and RNA molecules, such as ribozymes, HOMO-DNA and riboswitches. (51) (52) To further facilitate this strategy, enzymatic method has also been developed to prepare selenium-derivatized nucleic acids via DNA polymerization and RNA transcription. (53) The next two sections will briefly summarize the recently achievement and development of SeNA synthesis and applications.

1.3 Chemical synthesis of SeNA

1.3.1 Selenium modification at 5'-position

As proof of principle, the methyl selenide functionality was introduced into the 5'-position of all the A, C, G, T, and U residues. (54) This synthesis was achieved by activating of 5'-hydroxyl groups of 2a with good leaving groups such as Br, Ms and Ts, followed by a SN_2 substitution of selenide reagents in presence of a phase-transfer catalyst. After being transferred to the according phosphoramidites, all these Se-containing building blocks 2b and 2c (in figure 1.1) were successfully incorporated into DNA oligonucleotides through standard solid phase synthesis. It's been proved that the

selenium function was very stable when treated with mild iodine solution in the methylated form (2c), while compound 2b could easily be oxidized into diselenide form.

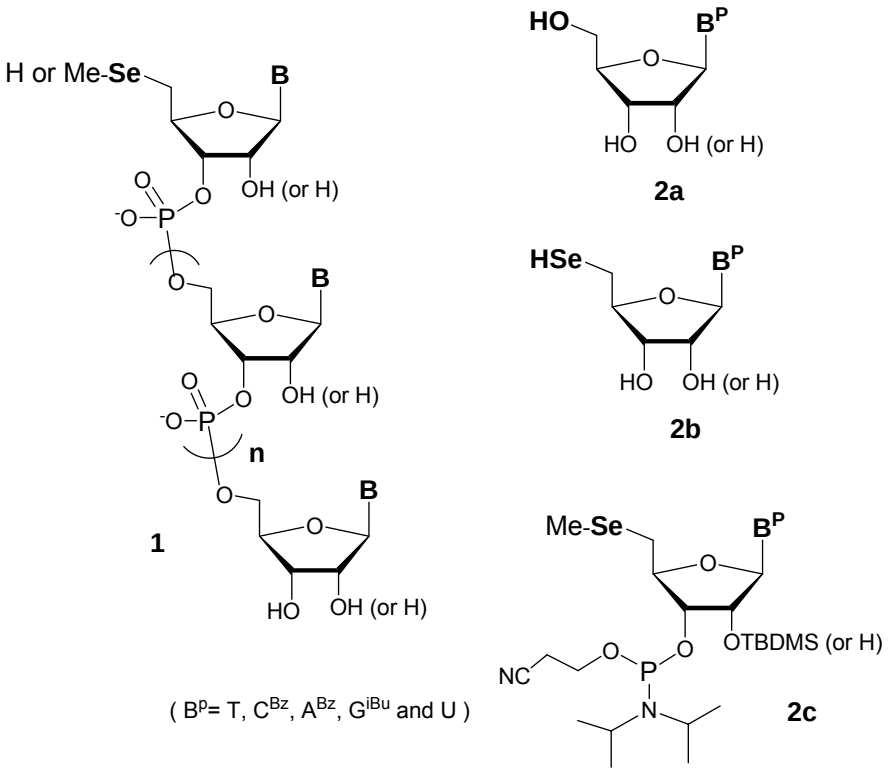
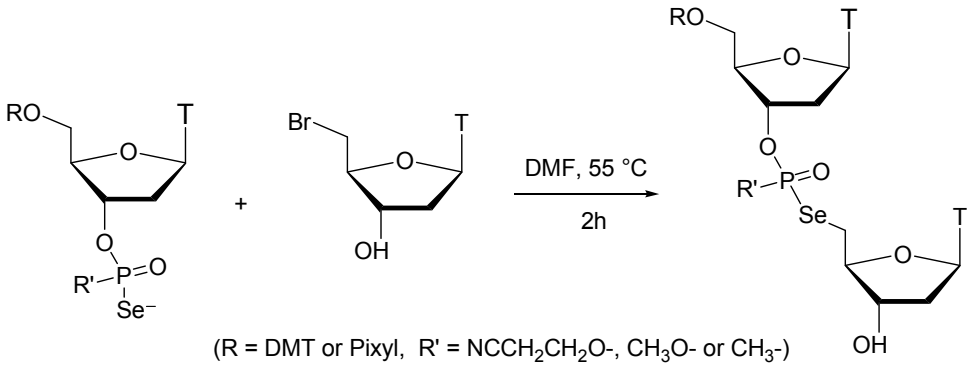


Figure 1.1: 5'-Selenide functionalized nucleic acids. (OTBDMS: tert-butyldimethylsilyloxyl)

It's also worthy pointing out that Stec et al synthesized a selenium modified TT dimer by replacing the 5' bridging oxygen with selenium in 1994 (Scheme 1.1). (14) The purpose of this work was to test the inhibition of this kind of compounds to nucleases and find some anti-virus applications as antisense drugs. As a result, they observed certain cytotoxicity of these selenium compounds instead of anti-virus activity, which however, could still provide a useful tools for molecular and cell biological applications.

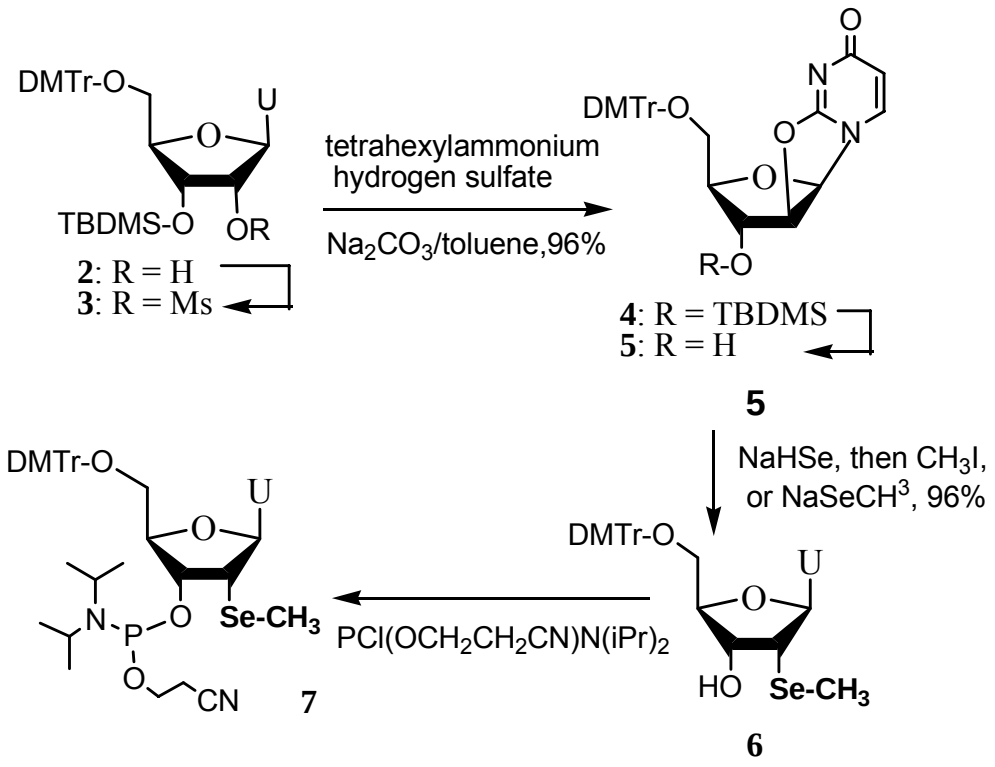


Scheme 1.1: synthesis of 5’-Se-modified TT dimer.

1.3.2 Selenium modification at 2’-position

The successful demonstration of the stability of selenium functionality and its compatibility with solid phase phosphoramidite chemistry promoted us to investigate the further possible positions in nucleic acid for selenium introduction. Then, the synthesis of phosphoramidite and oligonucleotides containing selenium at the 2’-α-position of uridine were achieved by using the 3’,5’-protected-2,2’-anhydrous uridine as intermediate (Scheme 1.2). (55) In this work, the crystal structural of a selenium modified DNA decamer 5’-GCGTA$\underline{U}_{2'\text{-SeMe}}$ACGC-3’ was also investigated at high resolution (1.3 Å, comparing 1.8 Å of the native sample without Se). The diffraction data was successfully phased on the basis of selenium anomalous signal. Figure 1.2 showed the absorption of selenium atom in X-ray fluorescence spectrum and in this experiment, three sets of data were collected for MAD: the inflection point (edge), the absorption maximum (peak) and ca.460 eV above the peak wavelength (remote). This structural data showed the 2’-methylseleno group was directed into the minor groove and this modification can remain the native C3’-endo conformation, which was consist with A-form DNA and RNA molecules. It is also worthy to point out that the data collected at the

peak wavelength is sufficient for SAD phasing when high-resolution diffraction data are collected. The Fourier electron density maps were computed (figure 1.3) based on the refinement of the two selenium positions.



Scheme 1.2: Synthesis of 2'-methylseleno containing uridine phosphoramidite and its incorporation into oligonucleotides.

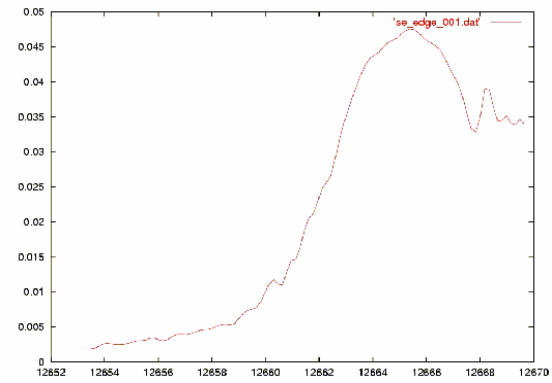


Figure 1.2: X-ray fluorescence spectrum of the decamer crystal. The theoretical value for the Se K edge is 12.6578 keV (0.9795 Å).

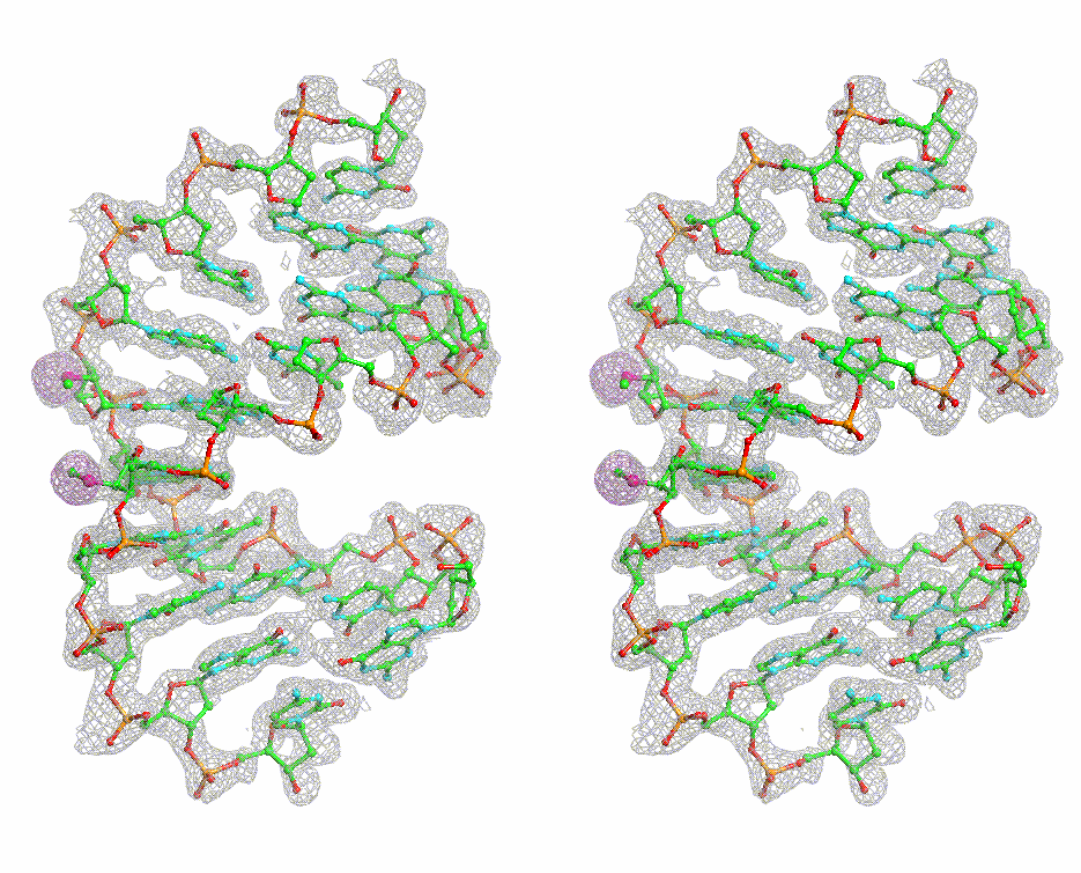
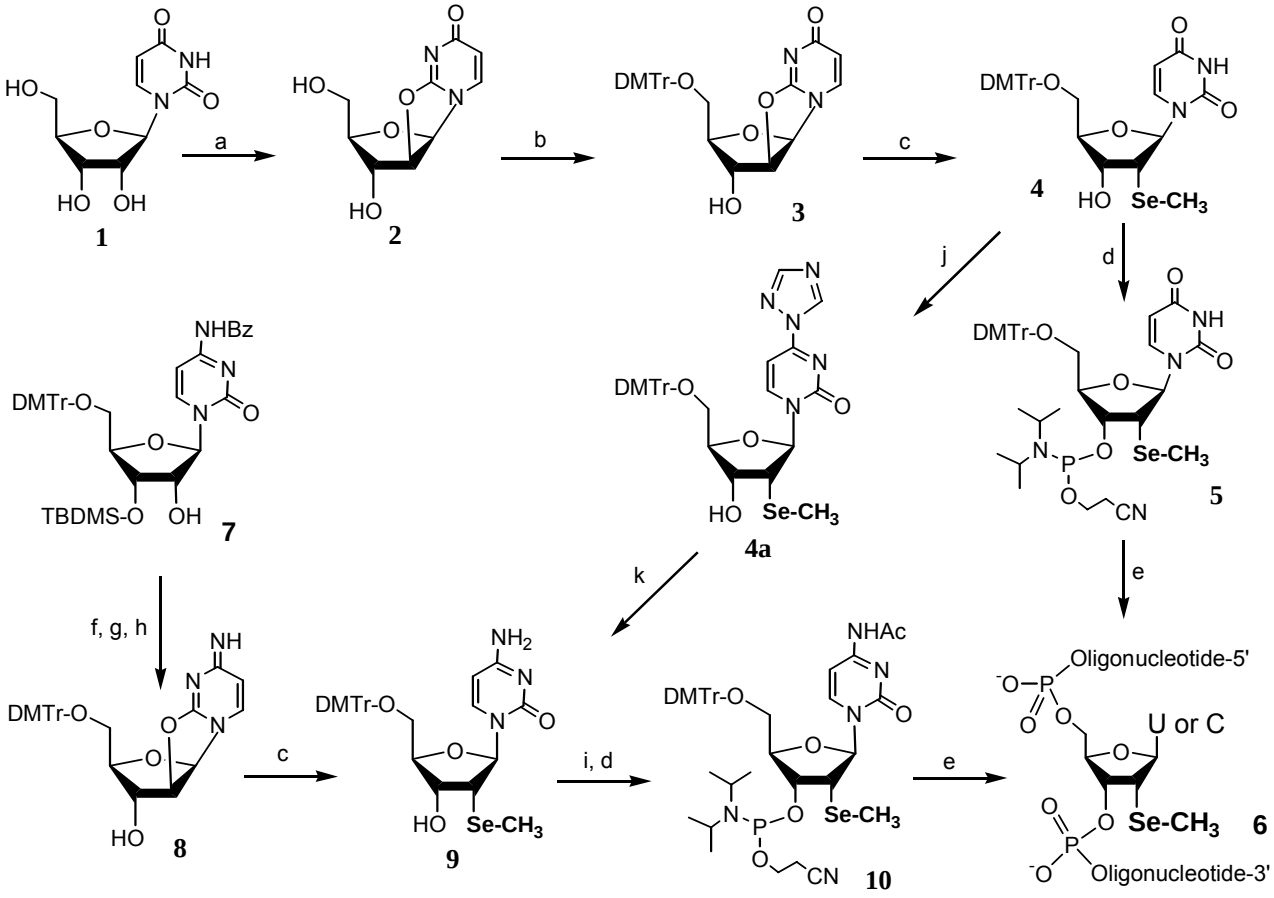


Figure 1.3: Fourier electron density maps for the final structure of the decamer DNA duplex calculated at 1.3 Å resolution with MAD phases. [d(GCGTA)USed(ACGC)]$_2$, DNA atoms are colored green, cyan, red and orange for carbon, nitrogen, oxygen and phosphorus, respectively, and the two selenium atoms per asymmetric unit are highlighted in pink.

Then, a modified synthetic route was developed to introduce selenium functionality into the 2' position for both uridine and cytidine as showed in Scheme 1.3. (56) A new coupling reagent 5-(benzylmerapto)-1H-tetrazole was used to make RNAs including a number of biologically and structurally important sequences with different secondary structures. This reagent could make the coupling yields of the modified nucleosides as high as over 99%. Moreover, the thermaldenaturization and crystallization of a selenium

containing self-complementary octomer GU(2'-SeMe)GTACAC were also studied comparing the native one. As a result, the two derivatized U_{SeMe} residues have been proved to have no significant effect on duplex stability of this A-form DNA. In addition, this oligonucleotide could be easily crystallized in many more screen conditions than native and bromine modified one, although their morphologies seemed the same (Figure 1.4, scaled by 5 fold). The high resolution structure of this sequence containing selenium derivatization was determined via MAD and SAD with a resolution of 1.28 Å.



(a) $(Ph)_2CO_3$, Na_2CO_3, DMF; (b) DMTr-Cl, Py; (c) $NaSeCH_3$, EtOH-THF; (d) 2-cyanoethyl N,N-diisopropyl-chlorophosphoramidite and N,N-diisopropylethylamine in

CH_2Cl_2; (e) synthesis of oligonucleotides on solid phase; (f) Ms-Cl, TEA, THF; (g) toluene/ tetrahexylammonium hydrogen sulfate, Na_2CO_3 (satd); (h) $(Bu)_4N^+ F^-$, THF; (i) TMS-Im, then Ac_2O, TEA and DMAP in THF; (j) TMS-Im, then $POCl_3$-triazole-TEA in CH_3CN; (k) NH_4OH.

Scheme 1.3: Improved synthesis of 2'-methylseleno containing uridine and cystidine phosphoramidite and their incorporation into oligonucleotides.

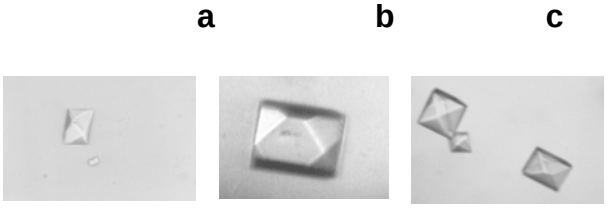
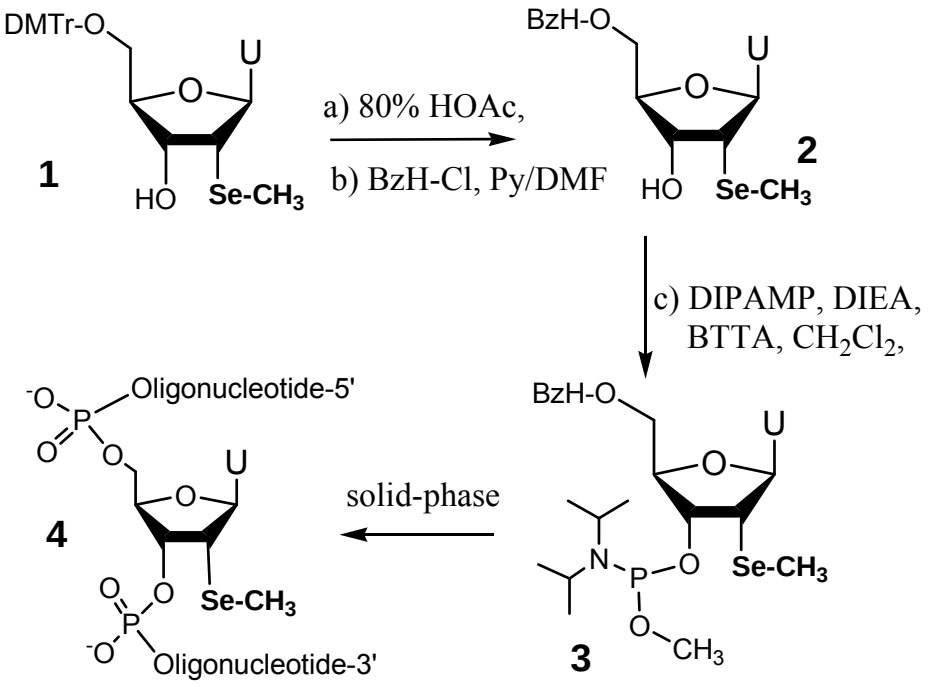


Figure 1.4: Photos of crystals of the native and Se-derivatized octamers. (a) Native-Oct.; (b) Se-Oct.; (c) Se/Br-Oct. Sizes of the crystals are ranged from 0.1x0.1 to 0.4x0.4 mm.

Based on the key intermediate of 2'-methylselenouridine, a 5'-silyl-2'-ACE chemistry was successfully applied for the synthesis of longer RNAs, (57) which usually was achieved by chemical synthesis of short fragments and enzymatic ligation to longer sequences (This part will be also mentioned in the following section). In this synthesis, the benzhydry group was introduced into the 5' of phosphoramidte replacing trityl group (Scheme 1.4) and several DNAs have been examed as model systems for the compatibility with the conditions used in the 5'-sily-2'-ACE chemistry.



Scheme 1.4: Synthesis of 5'-BzH-2'-SeMe-Uridine phosphoramidite and oligonucleotides.

A NMR study of the DNA duplex 5'-ATGGAGCTC-3'& 3'-TACCU_{Se}CGAG was carried to get more insight into the base pairs, which also tested the possible application of selenium label in NMR spectrum assignment and even functional studies. From the partial spectra showed in figure 1.5, it's clear that the chemical shift of hydrogen in the A:U_{Se} pair was obvious upfield-moved comparing the native counterpart (9D-5 Vs 9A-5), which could be further applied to the complicated NMR assignment and more structure determination.

1 2 3 4 5 6 7 8 9

5'-A T G G A G C T C-3'

3'-T A C C U_{Se} C G A G–5'

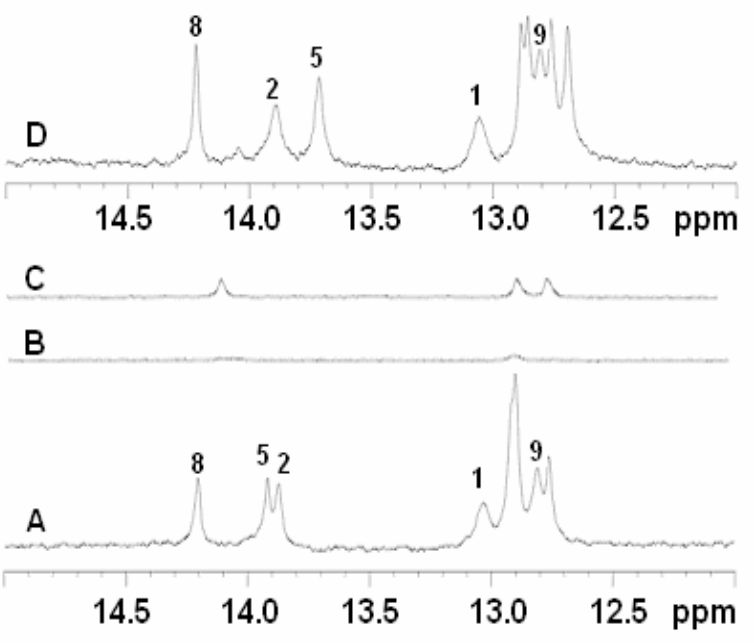
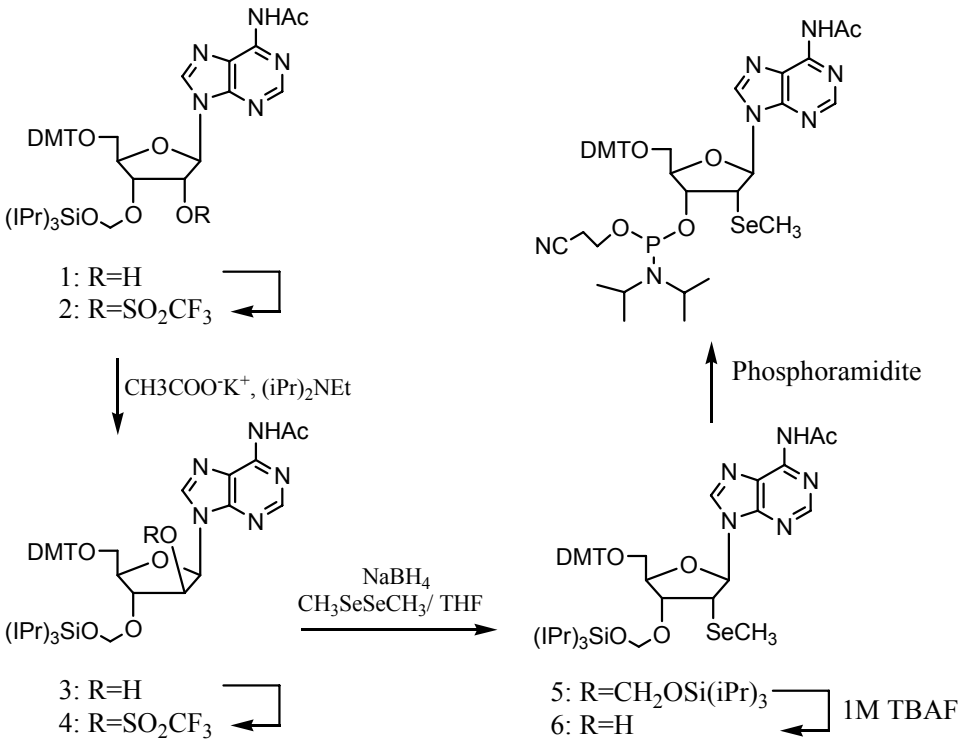


Figure 1.5: NMR study of Se-DNA, the complementary strand, and Se-derivatized DNA duplex at 283 K. (A) unmodified heteroduplex as a control; (B) 5'-GAGCUSeCCAT-3' single strand; (C) 5'-ATGGAGCTC-3' single strand; (D) modified heteroduplex.

All the synthesis stated above used 2,2'-anhydrous nucleosides as intermediates, which was preferred by pyrimidines while not working for purine bases. Therefore, to achieve that, a new strategy was developed by Micura and co-workers: a good leaving group such as Tf (trifluoromethanesulfonyl) was usually introduced into the 2'-β-position, followed by the substitution (SN_2) by selenium. This method has been already proved by the successful synthesis of 2'-methylseleno adenosine (Scheme 1.5) (58) and the following guanosine (Scheme 1.6). (59) So far, all the four 2'-SeMe containing RNA building blocks A C G U phosphoramidites are achieved and actually, they are most stable and convenient to use practically. It's claimed in their work that DTT is necessary to use for reduction of selenium atoms due to the long exposure to the mild iodine

solution (Figure 1.6). In addition, the crystallization of two Se-RNAs (12mer: $CG_{Se}CGAAUUAGCG$ and 16mer: $GCAG_{Se}AGUUAAAUCUGC$) was investigated. Similarly, they observed that there are boarder numbers of buffer conditions than their native counterparts, although the appearance of some thin needle-shaped and microsized crystals was also included. Importantly, a large number of Se-RNAs were found to be crystallized in new unique conditions totally different with the native ones. After that, the structure of the Se-16mer was studied comparing with native one. The superimposition vision of was showed in Figure 1.7a and the electron density map of the region around Se-modified residue G4 was showed in 1.7e. It's worthy to mention that the two local G(syn):A(anti) mispairs in the Se-sample were held differently with native one and the electron density of this mispair is slightly worse than the base pairs closer to the RNA tips (Figure 1.7b), which was a sign of the structure "breathing".

Furthermore, the crystal contact model in the lattices revealed that the RNA helices are packed in the head-to-tail fashion with a few side-to-side. Comparing the figure 1.7c and 1.7d, it was found that the side-to-side interaction model was shifted in the Se-crystal structure, which didn't involve another attached helix structure. Another rather noticeable difference was the orientation of G4 and G20, which faced the opposite strand of the neighboring helix in native RNA, while in Se-RNA faced against that. Moreover, from the structure data showed in 11f, which presented that three selenium atoms could form a equilateral triangle in the center of the packing interactions and therefore placing this nine nucleotides into close proximity, the authors assumed that was also a possible reason that this selenium modification could drive the crystal growth.



Scheme 1.5: Synthesis of the 2'-Se-methyladenosine phosphoramidite 7.

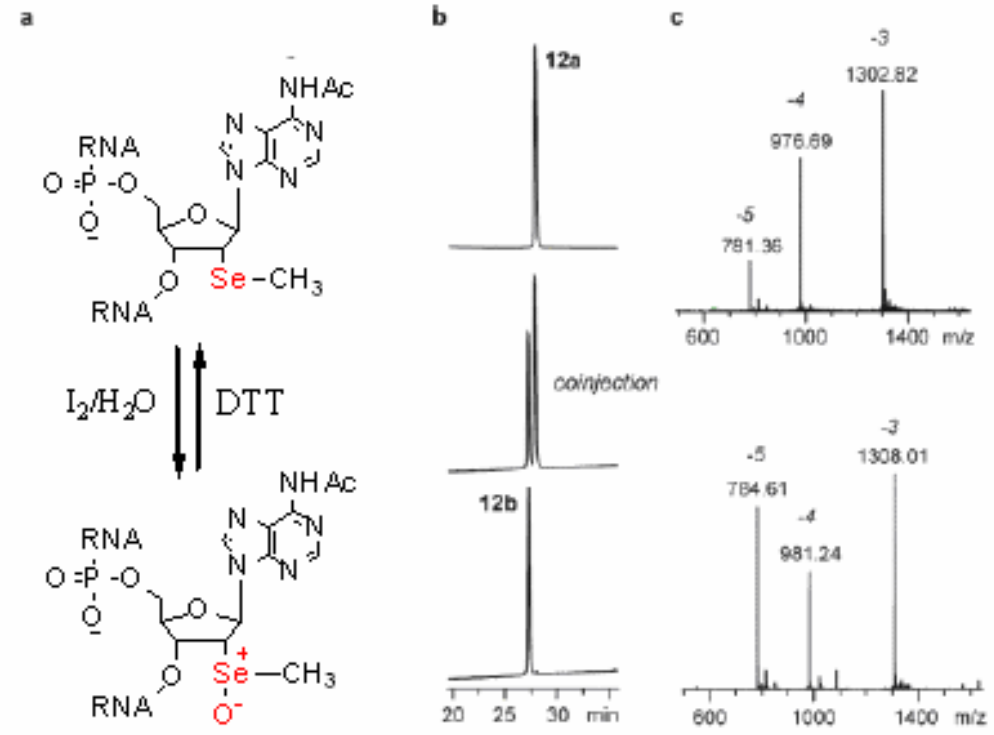
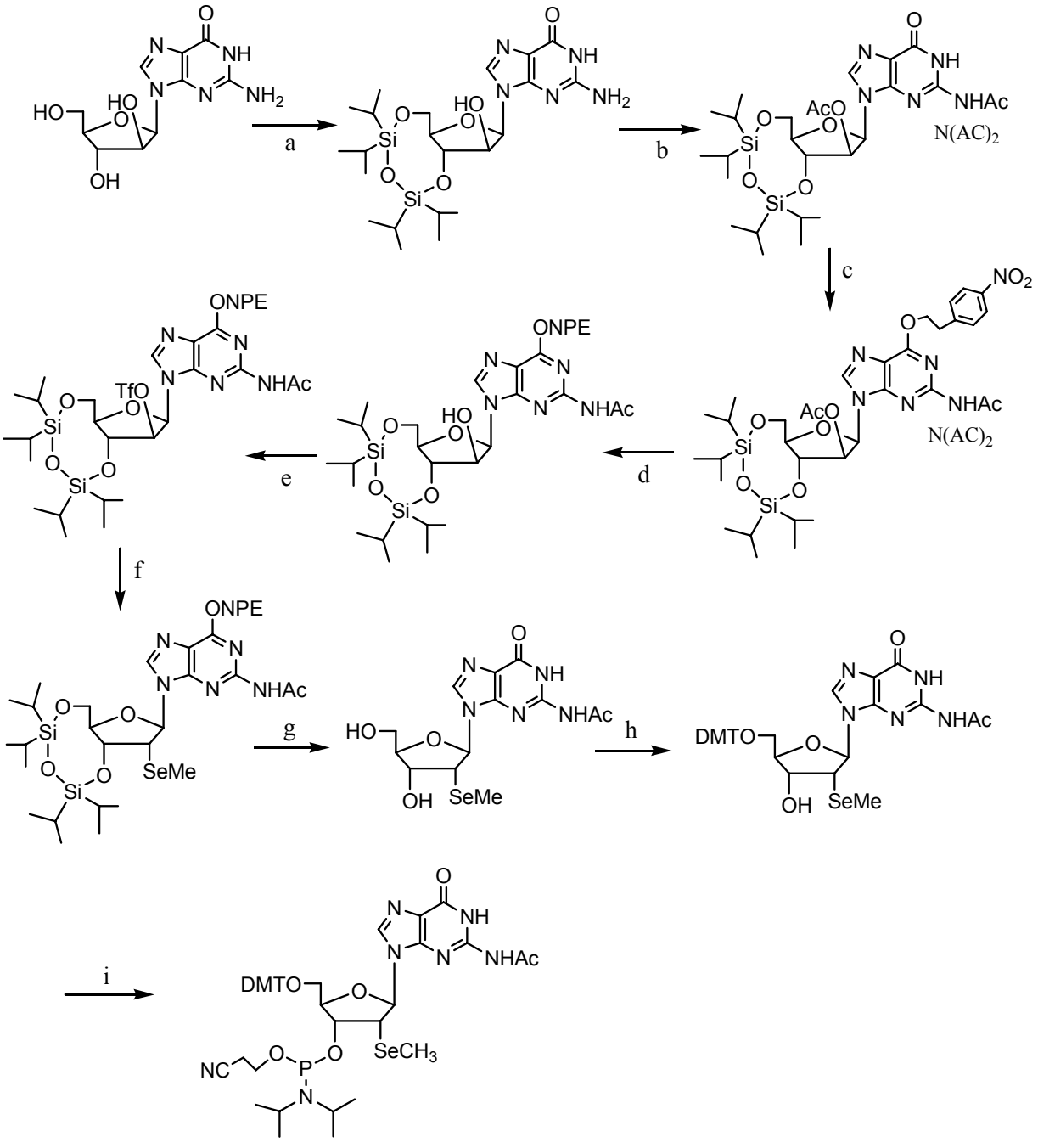


Figure 1.6: Redox behavior of Redox behavior of RNA 12mer. CGSeCGAAUUAGCG. (a) Chemical structures of reduced and oxidized selenium-modified RNA and (b) their HPLC traces and (c) their LC-ESI-MS spectra.



Scheme 1.6: Synthesis of the 2'-Methylseleno Guanosine phosphoramidite.

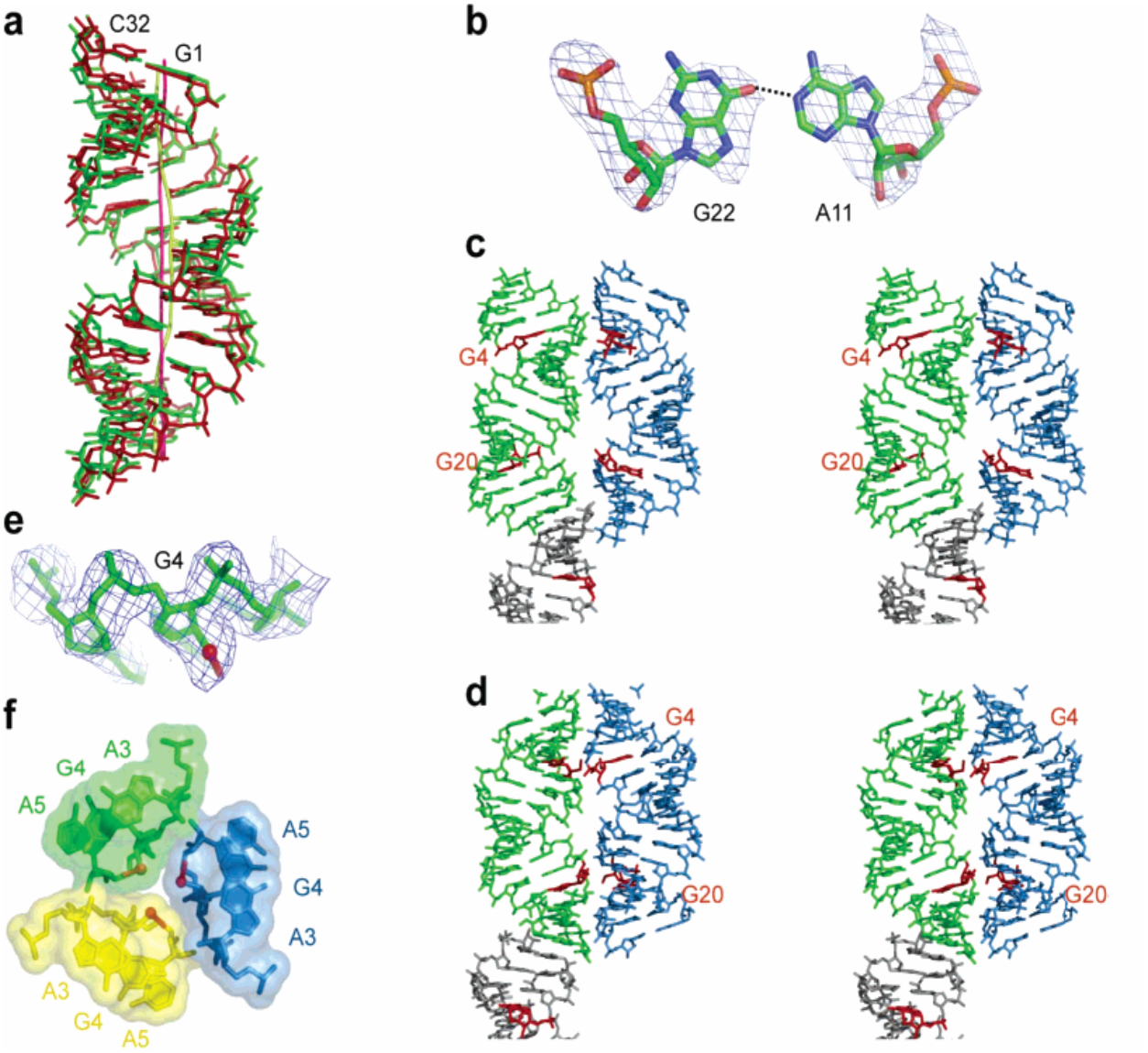


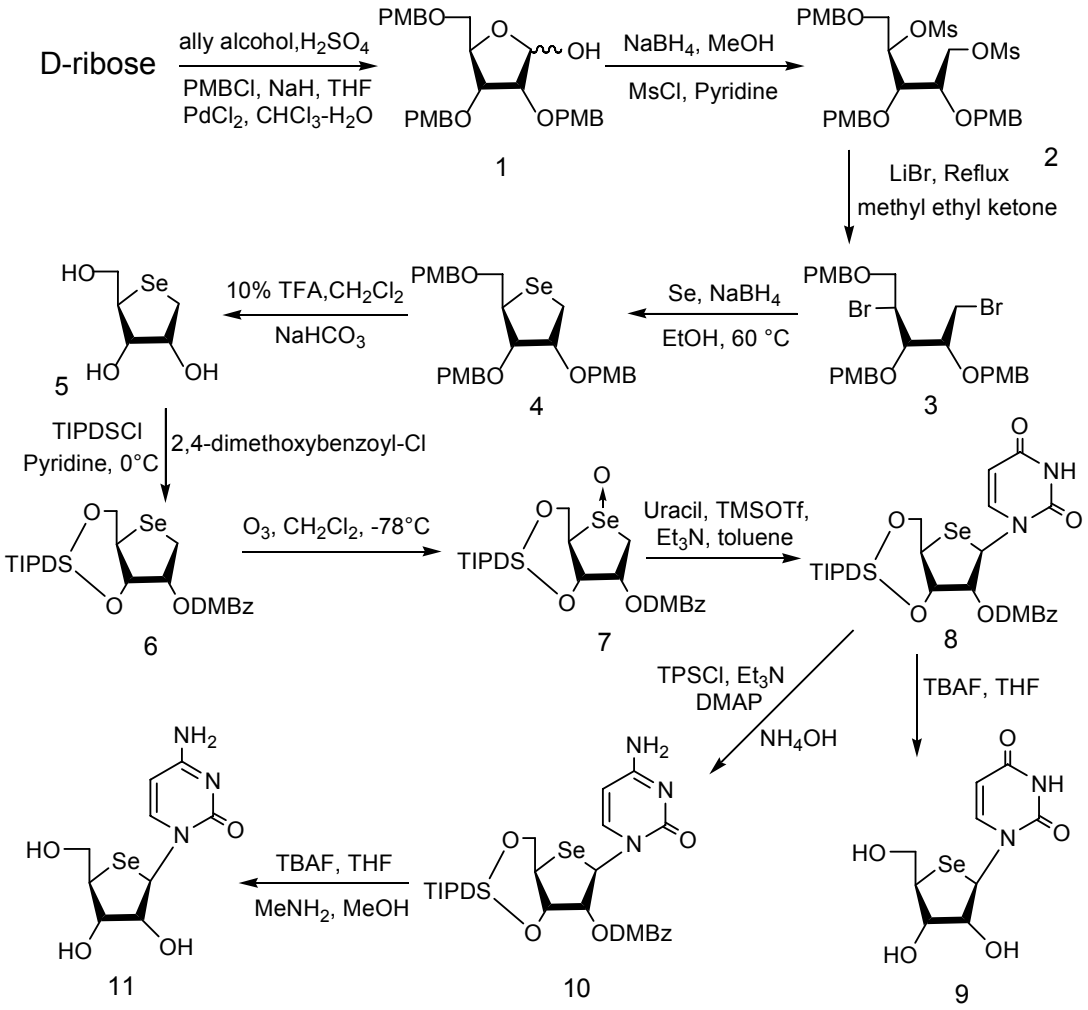
Figure 1.7: Crystal structure of the 2'-methylseleno-modified RNA duplex. (rGCA$\underline{G}_{Se}$A GUUAAAUCUGC)$_2$ (a) its superimposition (green) with the native structure (red). (b) electron density map of G(syn):A+(anti) mispair, hydrogen bond was showed in dash line. (c) and (d) crystal packing of the native and selenium modified sample. (e) electron density map for the Se-G4. (f) Packing contacts between three Se-modified RNAs involving 2'-methylseleno groups (red spheres for Se atoms and sticks for methyl groups).

Recently, several 2'-slenium modified residues were applied into the structure study of Diels-Alder ribozyme, a unique 49-nucleotide ribozyme, which could accelerate the Diels-Alder cyclization reactions with high enentioselectivity. The well solved structures in both unbound and bound states with reaction product have provided more detail insight into its catalytic pocket, the substrate recognition, binding and catalytic mechanism of steroselective carbon-carbon bond formation. (52)

1.3.3 Selenium modification at 4'-position

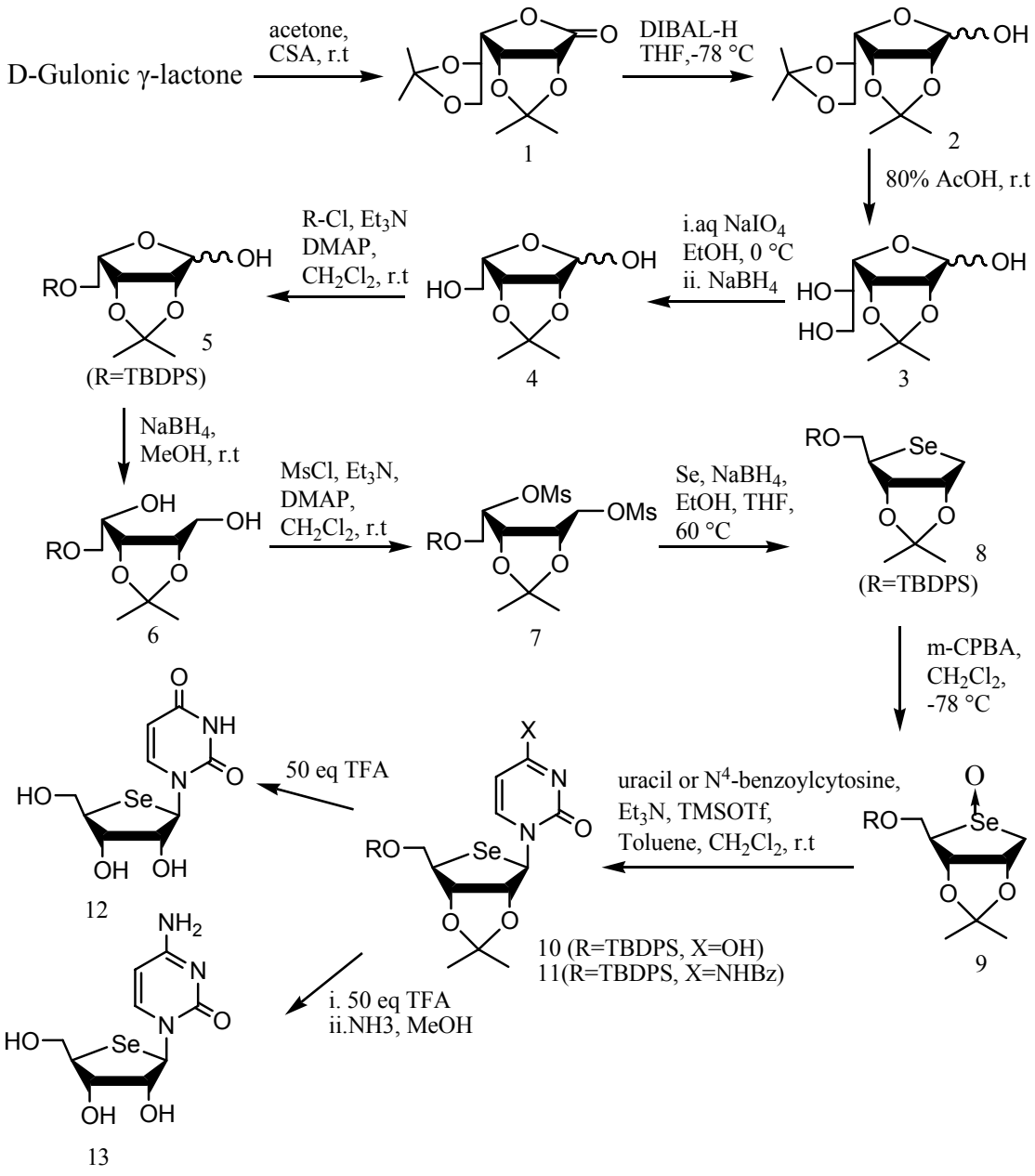
The replacement of the sugar ring (4'-) oxygen of nucleosides and nucleic acids by sulfur has been the subject of numerous studies, especially the application to the modification of aptamers and siRNA, as well as gene regulation. (60) The 4'-S-DNA can be amplified by PCR and can direct transcription in mammalian cells. In case of 2'-F-ANA, the sulfur substitution can cause a conformational switch which leads to interesting properties. The synthesis of the corresponding selenium counterparts has been one of the most challenging projects and will also have interesting biological properties besides the structural applications, including seeking new nucleoside based anti-virus templates and building blocks for antisense oligonucleotides synthesis.

The first 4'-Selenouridine was synthesized by Matsuda et.al (61) via the stereoselective seleno-Pummerer reaction between a protected selenoxide 7 (Scheme 1.7) and a persilylated uracil, which could also be transferred to 4'-Selenocytidine derivative 11 through the activation and aminolation of position 4. The coupling constants ($J_{1',2'}$=8.6 Hz and $J_{3',4'}$=1.7 Hz) showed a conformational preference for the S type (2'-endo) conformer.



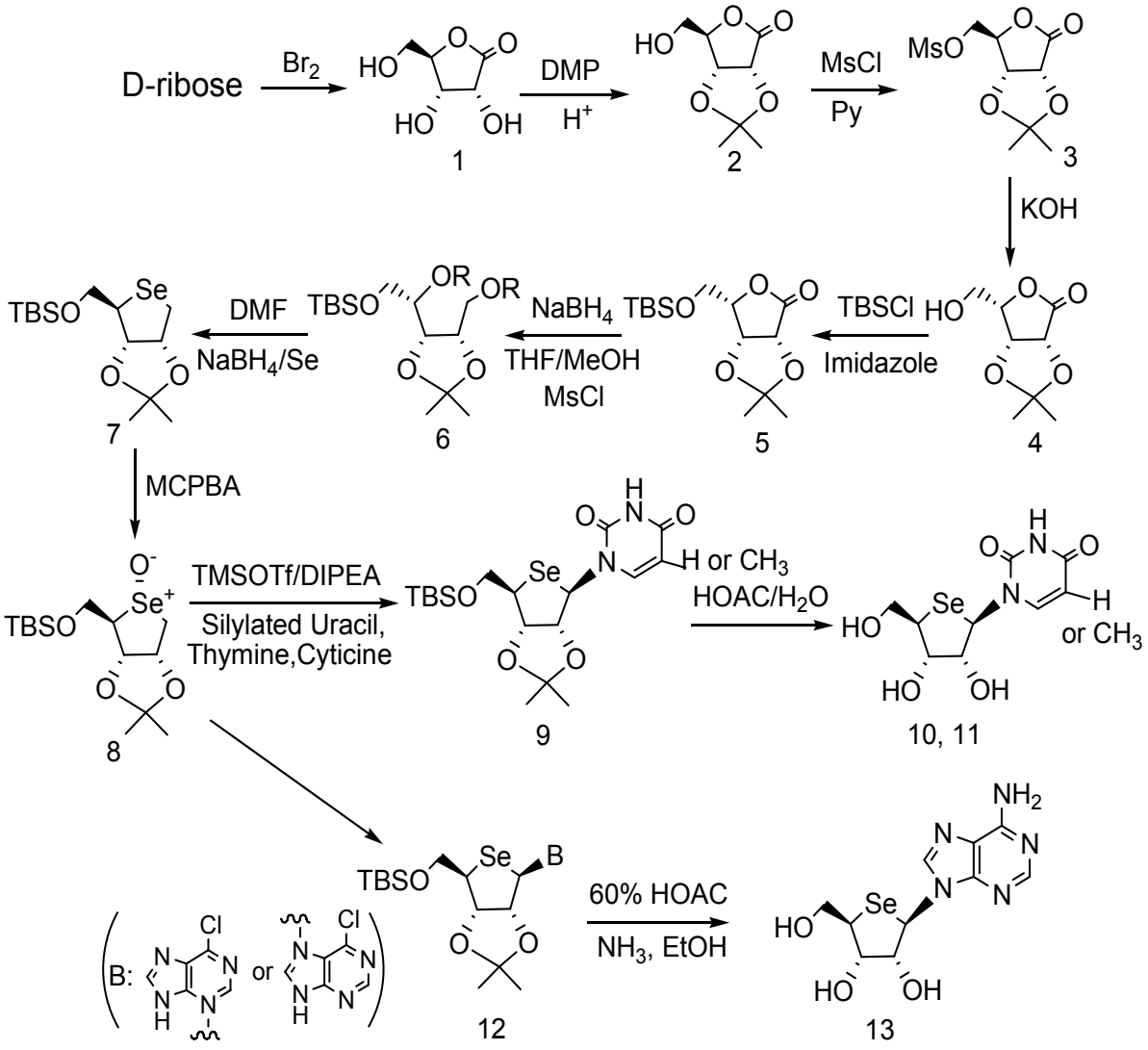
Scheme 1.7: Synthesis of 4'-Selenouridine 9 and cytidine 11.

Similar synthetic strategy was also applied by Jeong et al. to synthesize 4'-Selenouridine and cytidine. (62) Their key intermediate 4'-selenoxide 8, which was synthesized from D-gulonic-γ-lactone instead of D-ribose, as showed in scheme 1.8. In this work, the X-ray crystal structure of this 4'-Se-uridine was obtained and the following conformational study by coupling constants and NOE were also carried on, which further indicated that the replacement of the ribose ring oxygen by selenium causes a conformational change from a C3'-endo to a C2'-endo pucker in the solution state.



Scheme 1.8: Synthesis of the 4'-selenouridine 12 and 4'-selenocytidine 13.

The following work by Pinto et.al(63) further applied the Pummerer rearrangement of a selenoxide as the key step to build not only pyrimidines U, T, C, but also the purine nucleosides with high β-stereoselectivity, as shown in scheme 1.9. This breakthrough could provide a useful way to finish the synthesis of all the 4'-Selenium containing nucleosides building blocks.



Scheme 1.9: Synthesis of the 4'-selenouridine, thymidine, cytidine and adenosine derivatives.

Since it has been proved that these 4'-selenonucleosides adopt a southern, DNA-like conformations, it's very interesting to further explore this conformational preference in their oligonucleotides forms. With all the building blocks in hand, Pinto and Damha et.al carried out the biophysical characterization of oligonucleotides containing this 4'-Selenium functionalities. (64) In this work, the 5-mehtyl-4'-Se-Uridine (4'-Se-rT) was

converted into the according phosphoramidite and several DNAs and RNAs containing this residue were synthesized through solid phase synthesis, although the coupling yield is low (10-40%). Their thermal denaturalization and CD-studies indicated that this 4'-Se-rT insertion in an oligonucleotide behaved more like RNA in terms of both thermal binding affinity and hairpin structure affection, instead of the DNA-like conformation preference in its nucleoside form, although it also caused less base stacking. The further crystallography studies are underway.

1.3.4 Selenium modification at phosphate backbone

Although powerful the above stated 2'-methylseleno and 4'-oxygen replacement by selenium derivatization strategies are, these modifications could greatly induce the sugar pucker into 3'-endo form, which was different with the duplex conformation in solution. Therefore, this strategy has much better application for A-form DNAs and usual RNAs. To further develop the advantage that selenium could replace all kinds of oxygen atoms in nucleic acid, Egli and his co-workers have successfully applied the strategy of PSe-DNAs into this field and proved its feasibility. (65) PSe-DNAs are backbone modified nucleic acid analogues where one of the nonbridging oxygen atoms of the phosphate group is replaced by a selenium atom (Figure 1.8). This analog was developed initially for antisense purpose (66) and it was found to have diminished binding and antisense activity relative to the parent phosphodiester oligomers and DNA phosphorothioates (PS-DNA), as well as more cytotoxicity. But these biological properties didn't affect its application in X-ray crystallography, since it has been proved to be stable for several months under crystallization conditions.

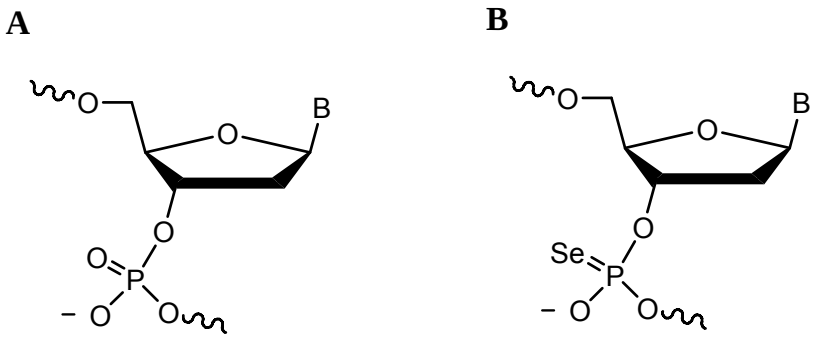


Figure 1.8: Comparison of the phosphate and phosphoroselenoate linkages.

The synthesis of this PSe-DNAs could be achieved by the modification of the oxidation step in the process of standard solid phase oligonucleotides synthesis, using potassium selenocyanate (KSeCN) in acetonitrile/triethylamine as oxidizing reagent instead of iodine solution. The selenium introduction was quantitative when the modification positions are located close to the 5'-end of target sequences, and the produced diastereomeric mixtures of PSe-DNA could be separated very well either by reversed-phase, SAX, or anion-exchange HPLC. However, more part of selenium will be oxidized when moving the modified position more inside, which was confirmed by MALDI TOF MS-Spec, presumably due to a reaction of the oxidation reagent. Moreover, earlier stability study about this derivatization concluded that the half-life of selenium was about one month and there was 10% lost in a week. But for crystallography purpose, some loss of selenium was still acceptable, the remained of which was still powerful enough to generate the anomalous signals.

Using this modification, the crystallization of a Z-form DNA d(CpSeGCGCG)$_2$ was studied by one of the diastereomers. The diffraction quality crystals usually appeared in one week, and as expected, the presence of two selenium atoms in one asymmetric unit (because it's complementary sequence) resulted in a very strong anomalous signal.

As usual, the fraction data were collected at three wavelengths (inflection point, peak and reference). Figure 1.9 and 1.10 showed the calculated Fourier electron density maps and the comparison of selenium structure and native one respectively, which showed that these two geometries are very similar, although actually the PSe duplex is rotated a little bit. From the structure data, it's also clear that both selenium atoms are attaching to water molecules by hydrogen bonds.

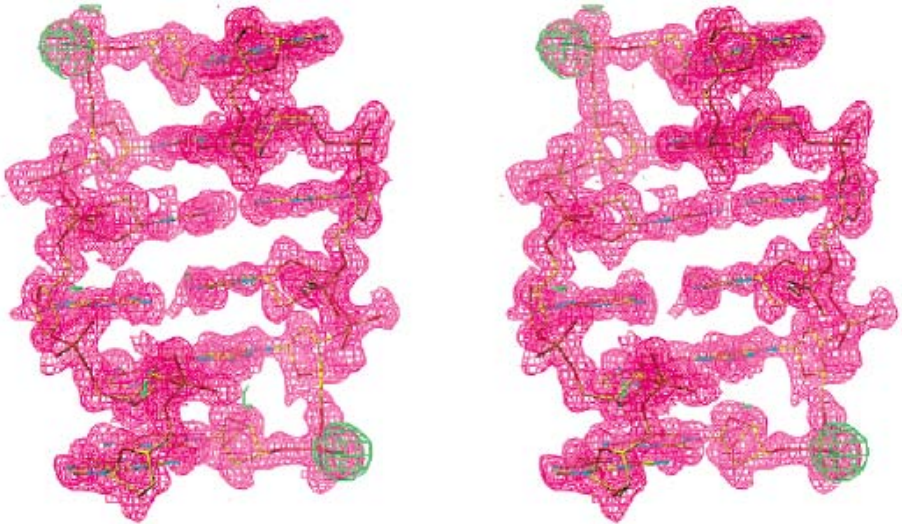


Figure 1.9: Fourier electron density map of the PSe-DNA duplex.

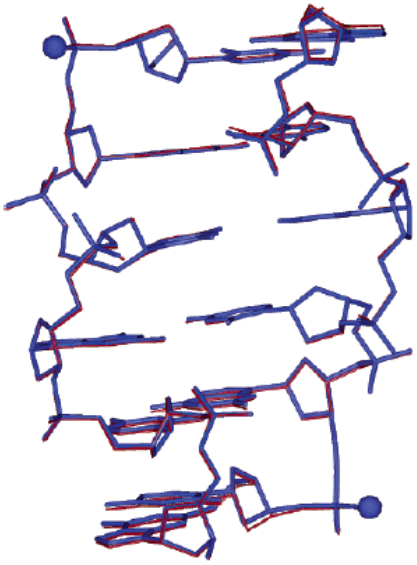


Figure 1.10: Superimposition of the PSe Z-DNA and native one ZDF001. Selenium atoms were represented by blue balls.

This successful prelude study of PSe-DNA encouraged them to further apply this strategy into the structure study of a homo-DNA $[dd(CGAATTCG)]_2$, the oligonucleotides with the standard 2'-deoxyribofuranose replaced by the 2',3'-dideoxyglucopyranose, as showed in figure 1.11, (67) which could be used to answer the question: Why nature uses exclusively pentoses as the building blocks for the nucleic acids, but never hexoses? By the later 1990, they have collected several preliminary data for this sequence, but the structure remained not resolved due to the phase problem; all the attempts either using molecular replacement (MR), multiple isomorphous replacement (MIR) or multiwavelength anomalous dispersion (MAD) methods had failed. In this work, one of the PSe-isomer was used and the final crystal structure was determined in high resolution (figure 1.12), moreover, more detail insights regarding the role of backbone-base inclination and interstrand stacking in pairing selectivity were also investigated based on the structural data. This also represented the first example of an unknown structure to be solved using selenium derivatization strategy.

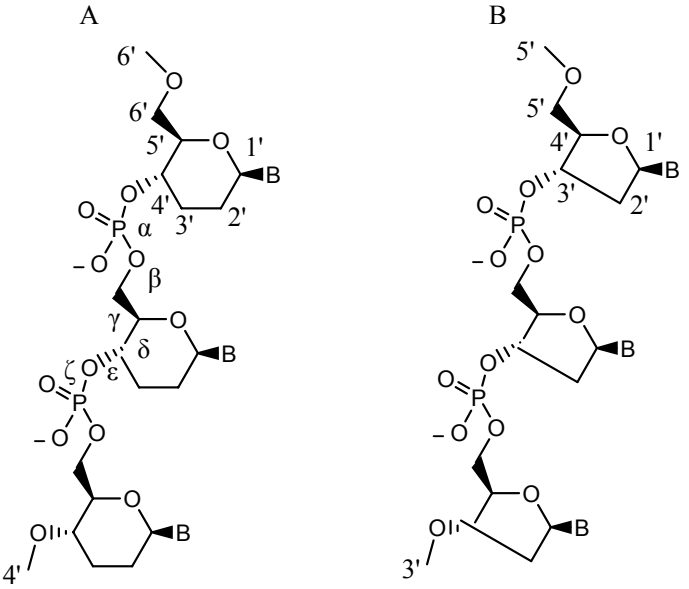


Figure 1.11: Homo DNA and regular DNA.

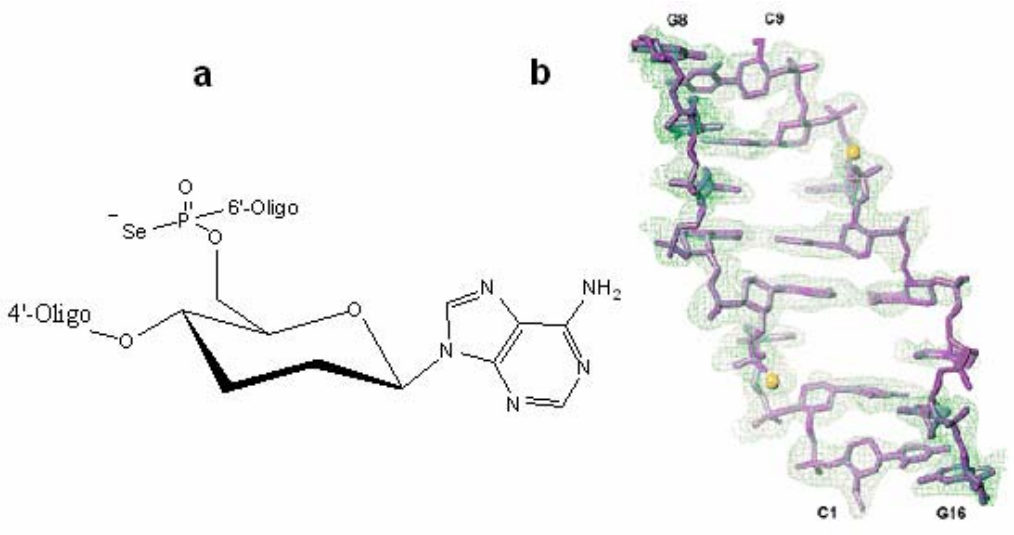


Figure 1.12: Structure and electron density of PSe-homo-DNA. (a) Structure of phosphoroselenoate homo-DNA. (b) Electron density of PSe-homo-DNA home-DNA $[dd(CGAATTCG)]_2$, superimposed on the final structure. Selenium atoms are shown as yellow spheres.

Besides the selenocyanate type of reagent, there were actually several other alternative selenium transfer reagents developed for the synthesis of PSe-nucleoside, such as $SePPh_3$, BTSe (Stawinski's benzothioselenol-3-one), and the organometallic reagent $(iPrC_5H_4)_2TiSe_5$. Similar as KSeCN, the working mechanism of introducing selenium was still based on the modified oxidation step of solid phase synthesis. (68) It's worthy to mention that in the work of $(iPrC_5H_4)_2TiSe_5$, a metal binding study was carried out for the sequence of $d(GGAATGTC_{Se}TGTCG)$ using ^{31}P NMR. As showed in figure 1.13, two set of metal titrations were compared, as a result, 5 equivalents of Cd^{2+} could cause an obvious 4.0 ppm upfield shift of PSe signal, while there was no significant shift even 20 eqiv of Mg^{2+} was added to the sample, which indicates that the phosphoroselenoate could preferentially bind Cd^{2+} rather than Mg^{2+} ions and this selectivity could be very useful for the future study.

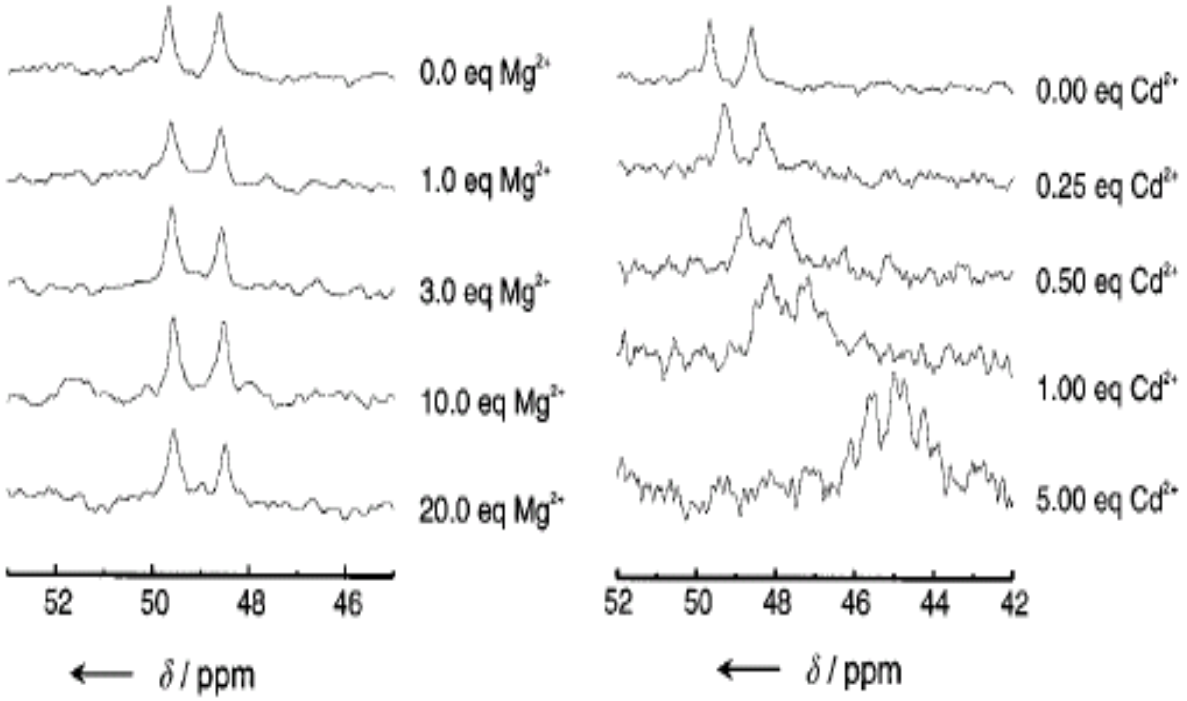
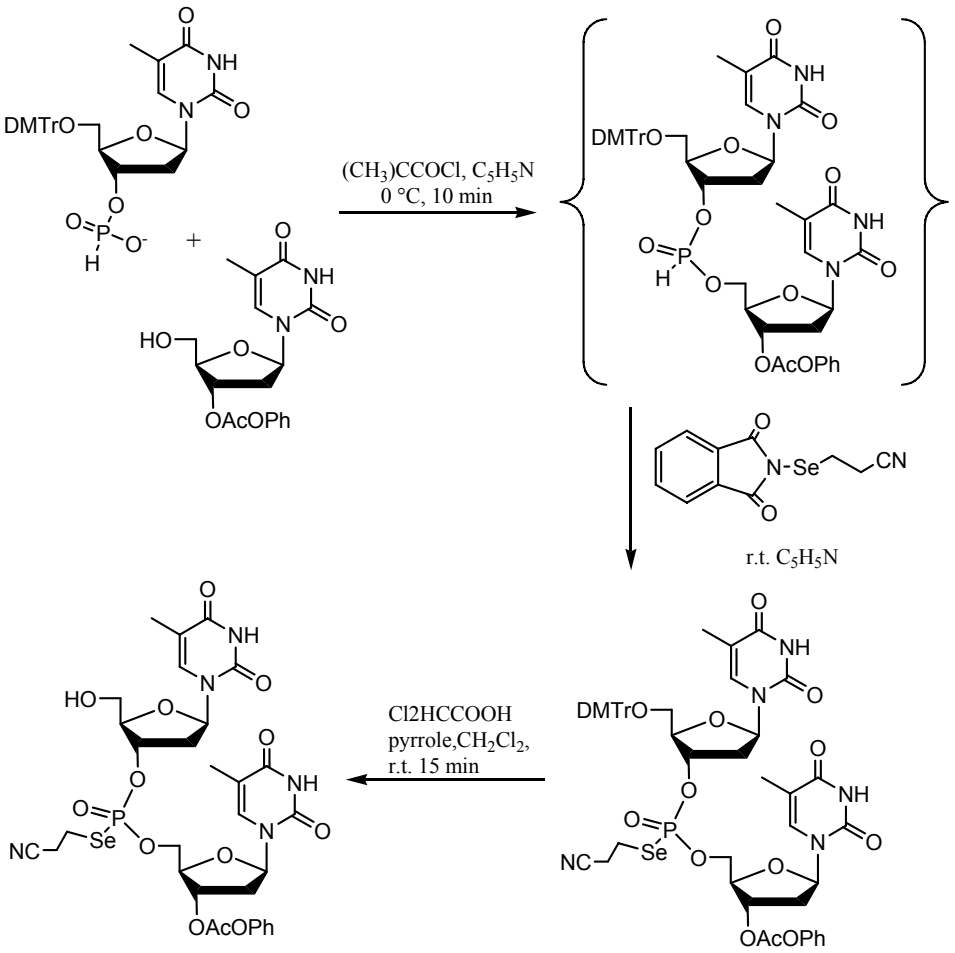


Figure 1.13: ^{31}P NMR spectra of the Se-modified 13mer DNA. (GGAATGTC$_{Se}$TGTCG). The plot on the right shows a shift of the phosphoroselenoate signals upon addition of Cd^{2+} ions (304 MHz), whereas the plot on the left shows no shift of the signals upon addition of Mg^{2+} ions (202 MHz).

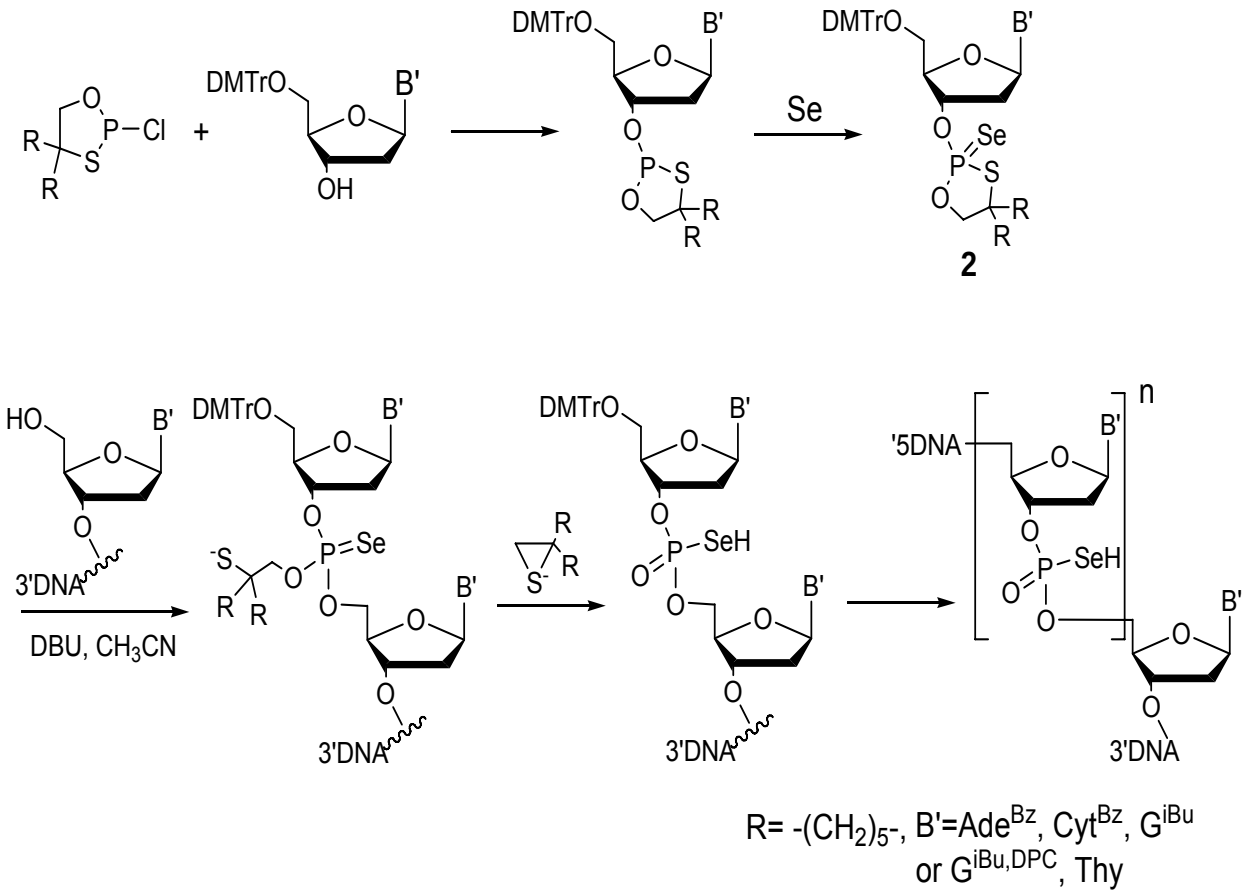
Recently, another new selenium transfer reagent Se-Cyanoethylphthalimide was developed for the further high efficient synthesis of oligonucleotide phosphoroselenoate (Scheme 1.10). (69) Inspired by the similar sulfur protection group, the selenium atom in this work was protected by a cyanoethyl group and a phthalimide function, which could prevent the selenium oxidation in early stages, and be removed during the base treatment of clued nucleic acids oligonucleotides before purifications. As an example, a selenium containing dimer was synthesized, as well as the deprotection, purification and characterization.



Scheme 1.10: Synthesis of Se-(2-Cyanoethyl) Phosphoroselenoate Triester.

Although powerful they are, the serious drawback of abovementioned reagents is the lose of stereoselective control of R*p* Vs S*p* for the phosphorus center, which usually can't be achieved using enzymatic methods. Since most of the biomolecules the oligonucleotides interact with are chiral and exist in single stereochemical forms, each diasteromer of a P-chiral oligonucleotide may interact with them in a slightly different way. Therefore, in the anti-sense application point of view, it's necessary to take the stereoselectivity into consideration to assure their most favorable biological activity, although it has been proved that both Sp and Rp diastereomers of TTPαSe could be accepted by DNA polymerase. To address this issue, Stec et.al (70) has developed an

oxathiaphospholane based novel approach which originally designed for the stereodefined phosphorothioate oligonucleotides. (71) As showed in scheme 1.11, the diastereomerically pure seleno-oxathiaphospholane monomer intermediate 2 was obtained in satisfactory yield, which could be inserted into any other part of oligonucleotides.



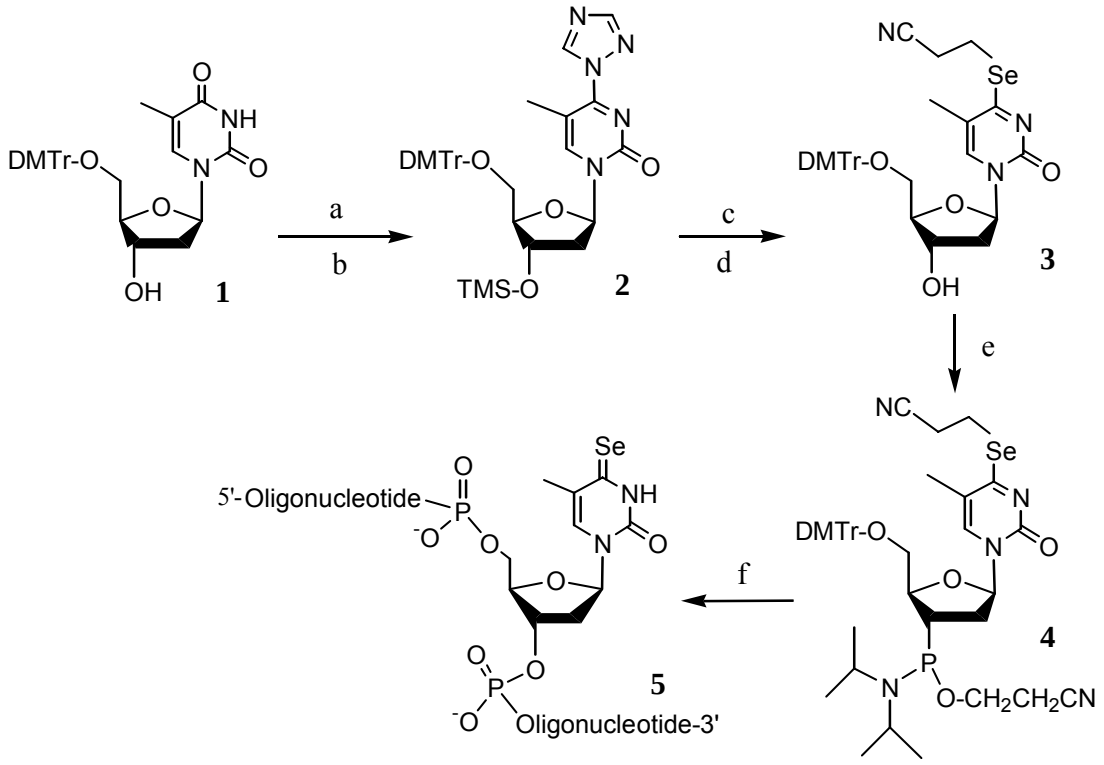
Scheme 1.11: Stereocontrolled selenium backbone modification.

The other part of backbone modification in triphosphate with selenium will be discussed in the section of enzymatic synthesis of SeNA.

1.3.5 Selenium modification in nucleobases

Very recently, we achieved the incorporation of selenium functions into the 4-position on the thymidine base for Se base pair and crystal structure studies. (72) (The project will be detailed descripted in Chapter 6). Since the replacement of oxygen with sulfur on nucleobases has been proved to be able to enhance the DNA duplex base paring selectivity, replication efficiency and fidelity, thus providing much more insights on the DNA duplex at the atomic level, (73) (74) (75) we assumed that besides facilitating the phase problem in X-ray crystallography, selenium could also be used as a molecular probe to track more insight information to the duplex structures. In the sulfur derivatization work, it was found that the 2-thio thymidine causes the slight stabilization of the T-A paring and DNA duplex, while the 4-thio thymidine causes the slight destabilization of them. So, it's interesting to check what the effect of selenium, actually, it's the first time that selenium base pairs in nucleic acids are studied, although the synthesis of Se containing nucleobase and nucleoside was finished several decades ago. (76) The incorporation of this modified residue into real nucleic acids is still challenging. Furthermore, some specific pyrimidines in natural tRNAs have been derivatized by Se on the nucleobased and this Se function has been barely understood, although it was suggested that the Se substitution might be involved in the tRNA anticodon. (77) So, this work could also offer a powerful tool to do the functionality study.

The synthesis of target phosporamidite and oligonucleotide is showed in scheme 1.12. The position 4 was activated by a triazoyl group, followed by the substitution of selenium functionality, which was generated by the reduction of di(2-cyanoethyl) diselenide with $NaBH_4$, to generate the key intermediate 3. Then the according phosphoramidte building block was made and incorporated into oligonucleotides.



Scheme 1.12: Synthesis of 4-Se-thymidine phosphoramidite (4) and oligonucleotides containing 4-Se-T (5). (a) TMS-Im and CH_3CN; (b) triazole-$POCl_3$-Et_3N; (c) $(NCCH_2CH_2Se)_2/NaBH_4$, EtOH; (d) 10% Et_3N in MeOH; (e) 2-cyanoethyl N,N-diisopropylchlorophosphor-amidite and N,N- diisopropylethylamine in CH_2Cl_2; (f) the solid phase synthesis.

After proving the compatibility of this derivatization with solid phase synthesis, we have synthesized several oligonucleiotides and the thermal stability experiments was carried out using a duplex system of 5'-ATGGT$_{Se}$GCTC-3' & 3'-TACCACGAG-5' as example. As a result, the Tm results showed the Se containing duplex (Tm=38.6) worked very well comparing the native one (Tm=39.2). The following crystallography study also proved that there is no obvious global structural perturbation observed comparing the native structure (Figure 1.14A), although the modified base has shifted a little bit, which indicated that the DNA is flexible and has the ability to accommodate the

bigger atom such as selenium (Figure 1.14B). In addition, the density map of base pair clearly showed (Figure 1.14C) that the thymidine 4-selenium atom forms a hydrogen bond (Se…H-N) with the adenosine 6-amino group in DNA duplex. In this work, a 2'-SeMe-uridine was used to facilitate the crystal growth. This trick could also have potential application to some other new sequences not easy to be crystallized. The following more detailed functional research about this base modified nucleic acid, as well as the elucidation of this kind of modification in tRNAs, are still ongoing. It's also worthy to point out that this selenium modification in the base could generate yellow colored DNA, which could provide a useful probe for nucleic acids researches.

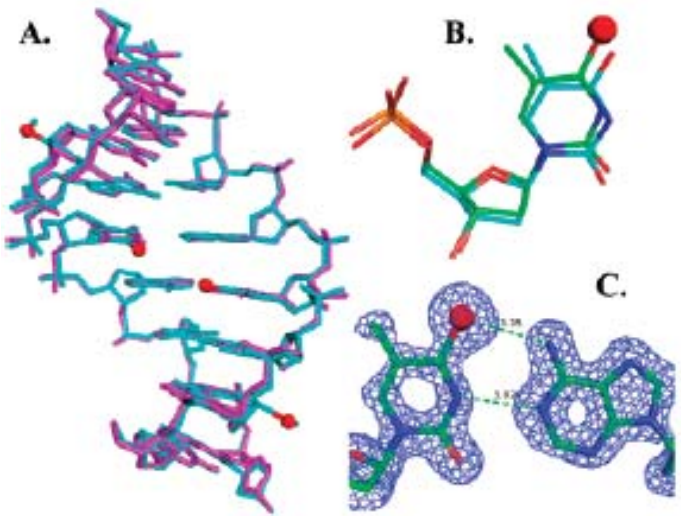


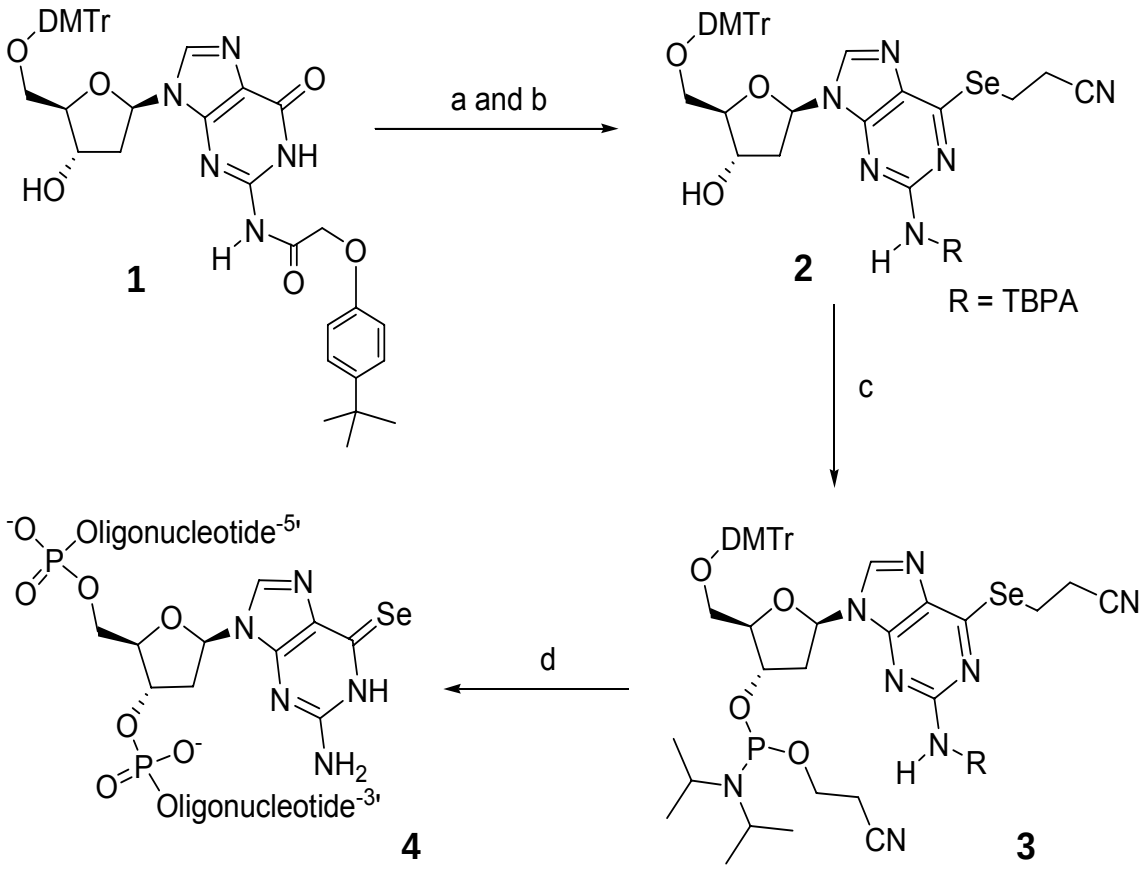
Figure 1.14: The global and local structures of the 4-Se-T DNA (5'-GdU_{Se}GT^{Se}ACAC-3'). (A) superimposition of modified (2NSK, in cyan) over native (1DNS, in pink). (B) Comparison of Se (green) and native (Cyan) local T4 structures. (C) Electron density of T4-A5 Se base pair.

Following this work and continuing our investigation of nucleic acid base pairing and stacking via atom-specific mutagenesis and crystallography, we have also

synthesized for the first time the 6-Se-deoxyguanosine phosphoramidite and its containing oligonucleotides, scheme 1.13. (78) Despite the synthesis of the 2'-deoxy-6-selenoguanosine and its derivatives several decades ago, (79) (76) as well as the successful incorporation of 6-S-derivative into nucleic acids, (80) the synthesis of nucleic acids containing the Se-guanine remained a challenge. Similarly, the synthesis of the key 6-Se-G intermediate was achieved by the selective activation of guanosine 6-position by TIBS group (2,4,6-triisopropylbenzene-1-sulfonyl), followed by the selenium substitution. Again, the corresponding phosphoramidite was proved to be compatible well with the solid phase synthesis and several DNA oligonucleotides containing this functionality were successfully synthesized, purified and characterized. Similar as 4-Se-T work, all these kinds of DNAs are found to be yellow colored with a new UV absorption at 360 nm, from which the extinction coefficient was determined at both 260 and 360 nm ($\varepsilon_{360}^{Se_G}$ as 2.3 x 10^4 $M^{-1}cm^{-1}$ and $\varepsilon_{260}^{Se_G}$ as 5.3 x 10^3 $M^{-1}cm^{-1}$).

Moreover, we crystallized a ternary complex of the Se-G-DNA, RNA and RNase H, and determined its X-ray crystal structure. This is the first structure determination of a protein-nucleic acid complex on the basis of Se-derivatized nucleic acids and MAD phasing. It was revealed by this work (as shown in figure 1.15) that the global nucleic acid structures of the native and Se-modified duplexes are very similar; and the SeG and C can also form a base pair similar to the native G-C pair by selenium mediation of hydrogen bond (Se…H-N). The GC base pair could accommodate the large selenium atom by shifting approximately 0.3 Å, which also explained the decrease of Tm by this kind of modification (including sulfur). It's worthy pointing out that this novel trial provide an alternative and easier tool to solve the structure of a protein-nucleic acid complex via Se-nucleic acids instead of Se-Met methodology. Besides the structural application, this

yellow Se-G DNAs may also have great potentials in construction of colored DNA nanoscale devices and structures as well as in nucleic acid diagnosis.



Scheme 1.13: Synthesis of the 6-(2-cyanoethyl)seleno deoxyguanosine phosphoramidite (3) and oligonucleotides containing 6-Se-G (4). Reagents and conditions: a) TIBS, DMAP, TEA, CH_2Cl_2, room temperature; b) diselenide, $NaBH_4$/EtOH, -5 °C, 80% yield in two steps; c) phosphoramidite, BTT, CH_2Cl_2,, 75% yield; d) solid-phase synthesis. TIBS = 2,4,6-triisopropylbenzene-1-sulfonyl chloride, DMAP=4-dimethylaminopyridine, diselenide = $(NCCH_2CH_2Se)_2$, phosphoramidite = 2-cyanoethyl tetraisopropyl-phosphorodiamidite, BTT = 5-(benzylthio)-1H-tetrazole.

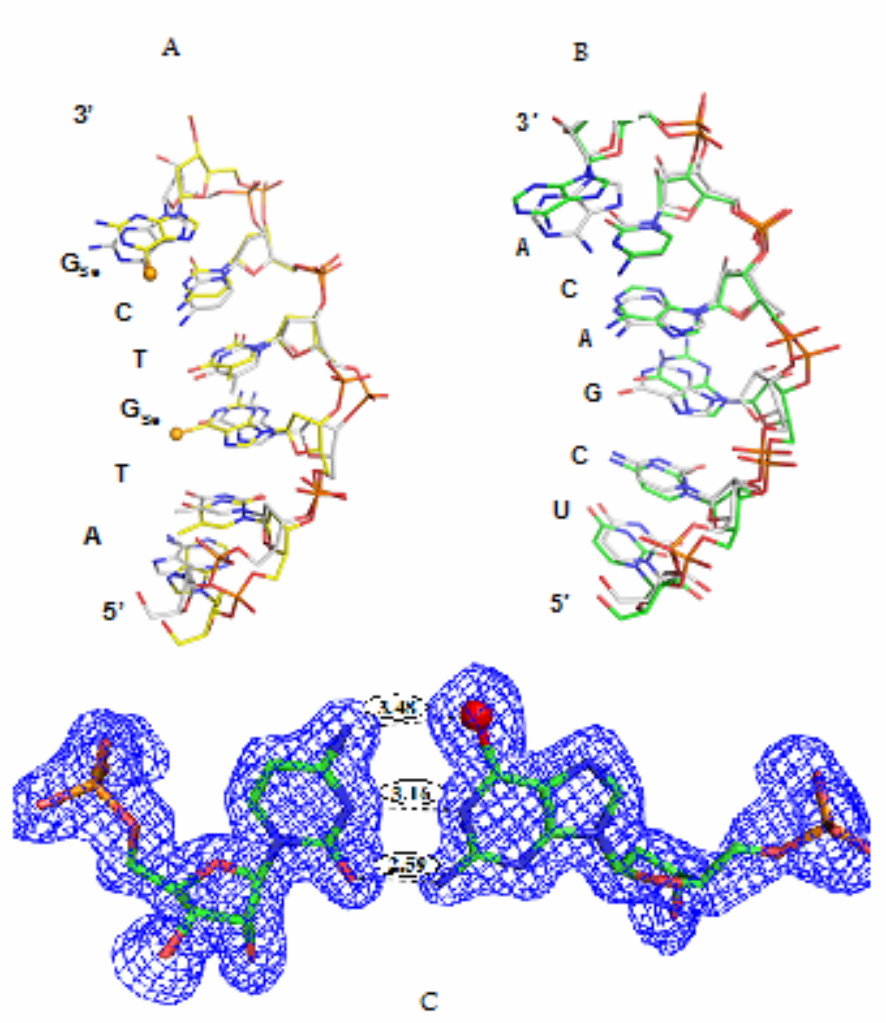


Figure 1.15: The superimposed global and local structures of the 6-Se-G-modified (2R7Y) and native (2G8U) DNA/RNA duplexes. (5'-ATGTCG-p-3'/5'-UCGACA-3') of the nucleic acid-protein complex. The balls represent selenium atoms in the Se-derivatized DNA (5'-AT-SeG-TC-SeG-p-3'). (A) The structure of the Se-DNA sequence (2R7Y, in yellow) is superimposed over the corresponding native (2G8U, in grey); (B) The structure of the RNA sequence (2R7Y, in green) is superimposed over the corresponding native (2G8U, in grey); (C) The Se-G3/C5 base pair (2R7Y) with the experimental electron density shows three hydrogen bonds (exo-6-Se/exo-4-NH2, 1-NH/N3, and exo-2-NH2/exo-2-O) with bond lengths in 3.48, 3.16 and 2.59 Å, respectively.

1.4 Enzymatic synthesis of SeNA

The chemical methods to obtain selenium derivatized nucleic acids at all different positions were based on the Se modified phosphoramidites synthesis and automatic or semiautomatic solid phase synthesis, which was very suited to short oligonucleotides, while not powerful for the preparation of longer ones. To achieve that, a nucleoside triphosphate analogue containing selenium functionalities at different positions could be synthesized and incorporated into DNAs and RNAs enzymatically.

As trial, the α-nonbridging selenium (α-Se-TTP) was synthesized in our lab, as showed in figure 1.16. (56) Since polymerases can conduct the polymerization reaction steroselectively, the enzymatic approach will generate diastereomerically pure PSe-nucleic acids, which provided another advantage over chemical counterpart. So, it's not necessary to separate the diastereomers.

A designed primer (21nt) and template (55nt) (figure 1.17A) were used in this demo-system and it's clearly showed from a time-course experiment (figure 1.17B) that the full-length DNAs were obtained from all primer extension reactions including those with both diastereomers of PSe in quantitative yield. It was also found that this two diastereomers could be recognized equally by the Klenow fragment of DNA polymerase I and even as well as the native TTP. To further confirm that the formation of DNAs were not due to the oxidation, a exonuclease digestion experiments was carried out using Exonuclease III since it has been reported that the modification of the nucleotide backbone could prevent this process. From the result of figure 1.17C, an obvious resistance was observed comparing the native TTP, which indicated the real selenium incorporation. Moreover, the different digestion resistance patterns of the two diastereomers were observed from the gel data, which probably due to the different

binding modes of nuclease at the initial and the stationary phases. As expected, when the selenium modified DNAs were oxidized to normal TTP, no resistance was observed.

This successful demonstration of the principle and the fact that large-scale preparation could be achieved using a larger volume and higher concentration of primer/template have paved the way to enzymatic synthesis of long DNAs and RNAs with selenium derivatizations on the backbones without further separation of the resulting diastereomers. To further widen its application, we developed the PSe-RNAs by in vitro transcriptions using T7 RNA polymerase and α-Se-ATP as selenium building block in the same primer/template system (Figure 1.18). (81) Surprisingly, one of the PSe-diastereomers (we define it as ATPαSe I) was found to be even more efficient than native ATP, while the other one was not the right substrate. There was no full-length product formed when using ATPαSe II and actually, more data showed that this non-working ATPαSe was neither a substrate nor an inhibitor (Figure 1.18b). Therefore, it's not necessary to separate the mixture in this transcript experiment.

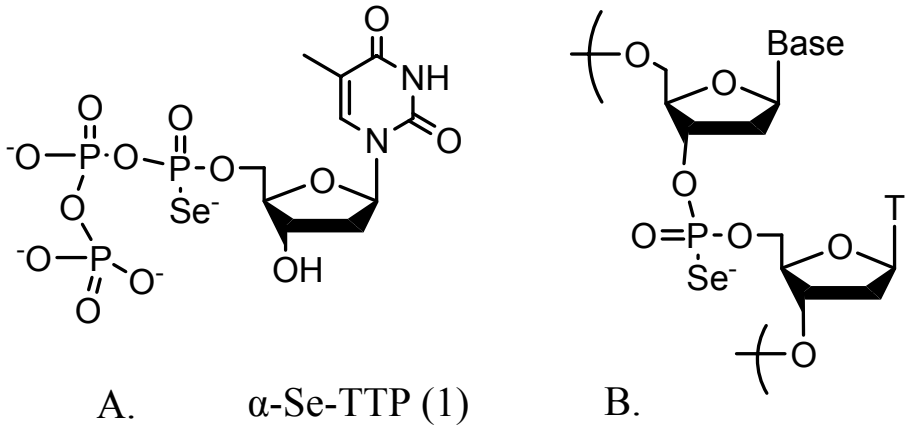


Figure 1.16: Structure of thymidine 5'-(α-P-seleno)triphosphate (A) and phosphoroselenoate (PSe) DNA (B).

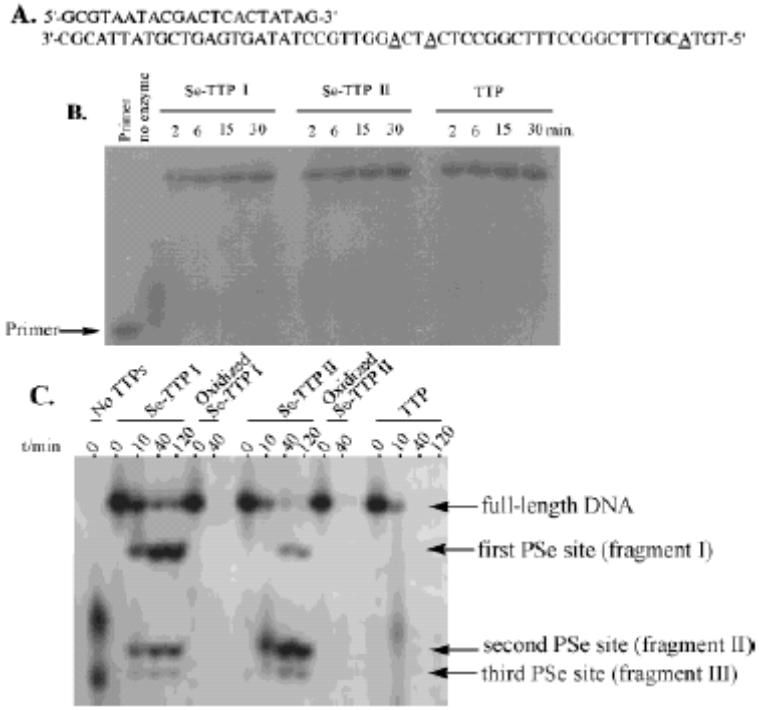


Figure 1.17: Enzymatic incorporation of backbone selenium into DNAs. (A) Primer and template sequences. (B) Time-course primer extension using R-Se-TTP I and II (Sp and Rp) on 19% polyacrylamide gel. (C) Exonuclease III digestion after primer extension.

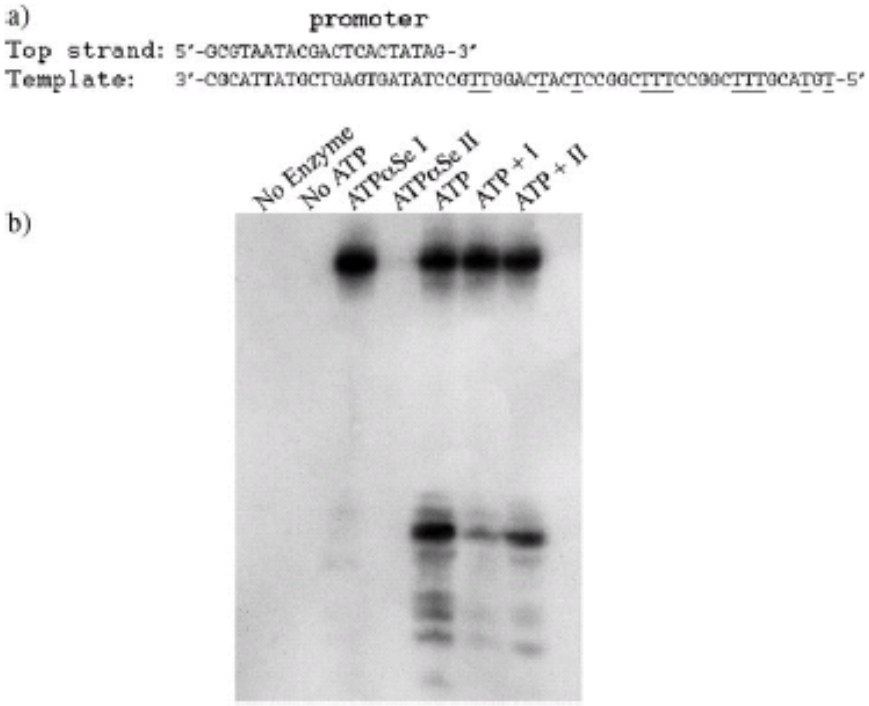


Figure 1.18: Enzymatic incorporation of ATPaSe into RNA by T7 RNA polymerase. a) The DNA sequences of the top strand and template. b) Gel electrophoresis analysis of the transcription mixture after 1 h of incubation at 37.8C.

Again, to further confirm the real incorporation of selenium, the RNA transcript products were subjected to the digestion with snake venom phosphodiesterase I, which is an exonuclease that degrades both DNA and RNA in 3' to 5' direction. The result (figure 1.19) indicated that the formed phosphoroselenoate RNA can indeed resist the enzymatic digestion, which was four to five times slower than that of the corresponding native one.

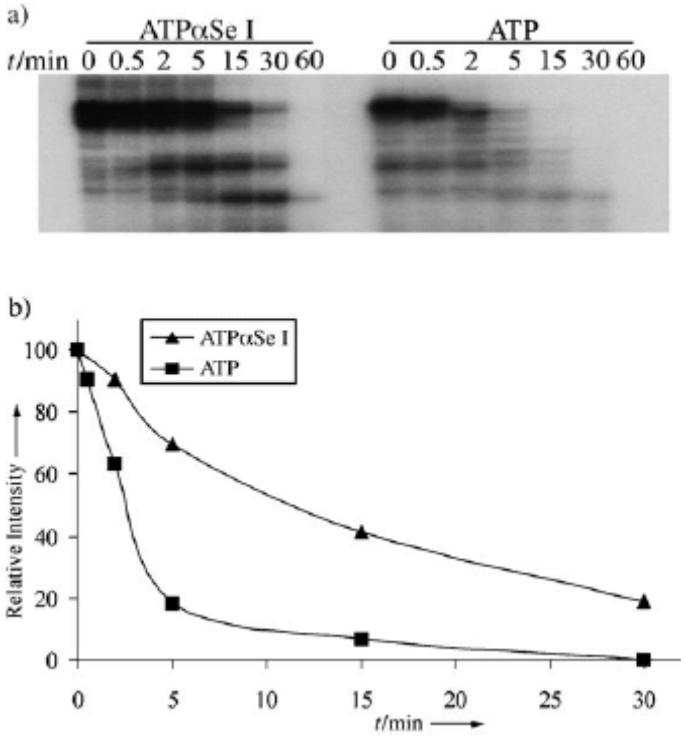


Figure 1.19: Time-course enzymatic digestion of phosphoroselenoate RNA with snake venom phosphodiesterase I. (a) gel electrophoresis autoradiography; (b) the digestion Vs time plot.

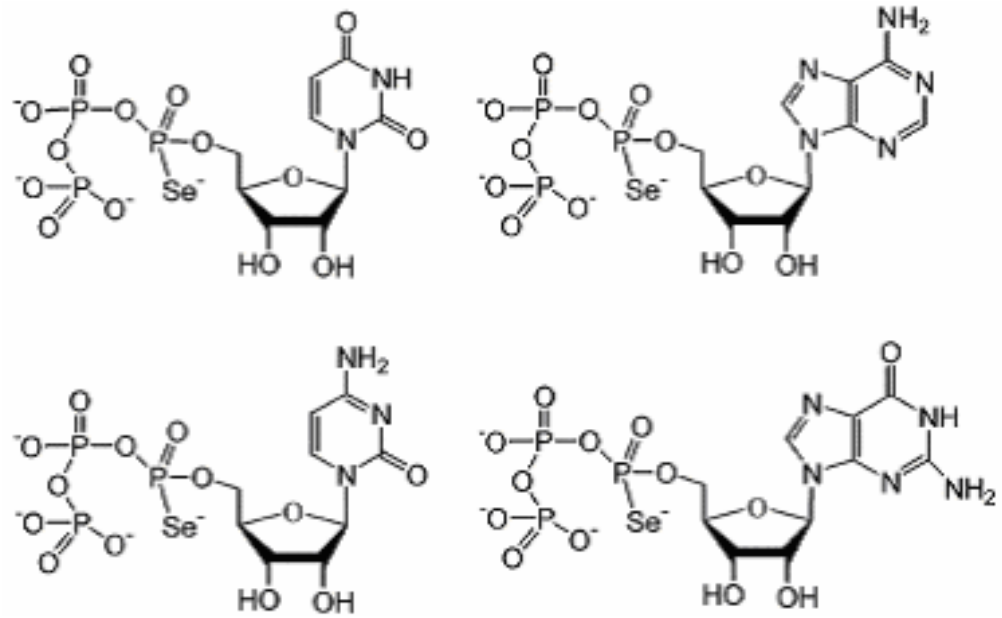


Figure 1.20: NTPaSe structures.

Then, all the four selenium modified RNA nucleoside triphosphate building blocks (Adenosine, Cytidine, Guanosine and Uridine triphosphats, figure 1.20) were synthesized and were incorporated into a desired hammerhead ribozyme as a model system to test the in vitro transcriptional incorporation and properties of these selenium containing-residues in a relatively larger biological system. (82) As showed in figure 1.21b, the results were consistent with our previous experiments, NTPαSe I was a good substrate as the native NTP, while interestingly, the NTPαSe II was neither a substrate nor an inhibitor for the enzyme, which indicated that it is not necessary to isolate this pair of mixture for enzymatic incorporation. The following time-course experiment using NTPαSe I was carried out to compare their incorporation efficiency as substrates with the natural NTPs. It's found that the transcription yields are at the same levels as that of natural ones (figure 1.21c).

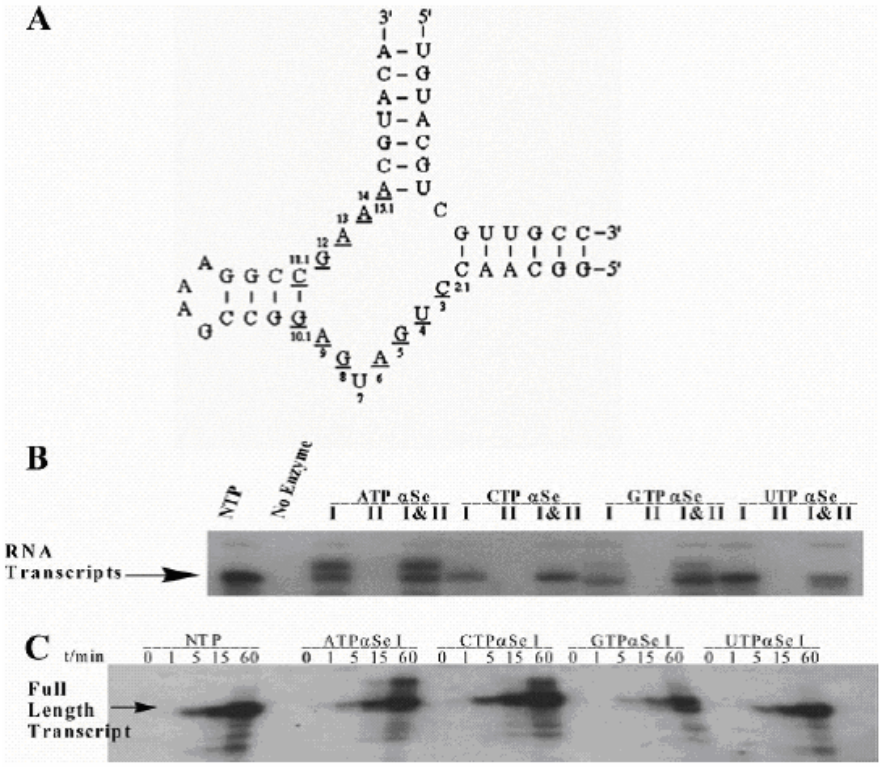


Figure 1.21: Enzymatic incorporation of selenium functionality into hammerhead ribozyme. (A) Secondary structure of the hammerhead ribozymes; the underlined sequences are highly conserved. (B) Transcription with native NTP, ATPRSe I and II, CTPRSe I and II, GTPRSe I and II, and UTPRSe I and II. (C) Time-course experiment of the transcription with native NTP, ATPRSe I, CTPRSe I, GTPRSe I, and UTPRSe I.

To study the activities of these Se-hammerhead ribozymes, the catalysis efficiency was carried out using a kinased 33nt RNA substrate in presence of metal ions such as Mg^{2+}, as showed in figure 1.22. Interestingly, the hammerhead ribozymes transcribed with UTPαSe I and CTPαSe I have the same level of efficiency as the native one, while the GTPαSe transcribed was only 30% of native activity, and for the ATPαSe one, there was only a trace amount product, although it could be rescued by nearly 200-fold when Mg^{2+} was replaced by Mn^{2+} (Figure 1.22c). The reason behind this was maybe because

the 5'-phosphate oxygen atoms could be involved in the catalysis process, which then provided a possible method to investigate more detail insight for the catalytic mechanism of ribozymes.

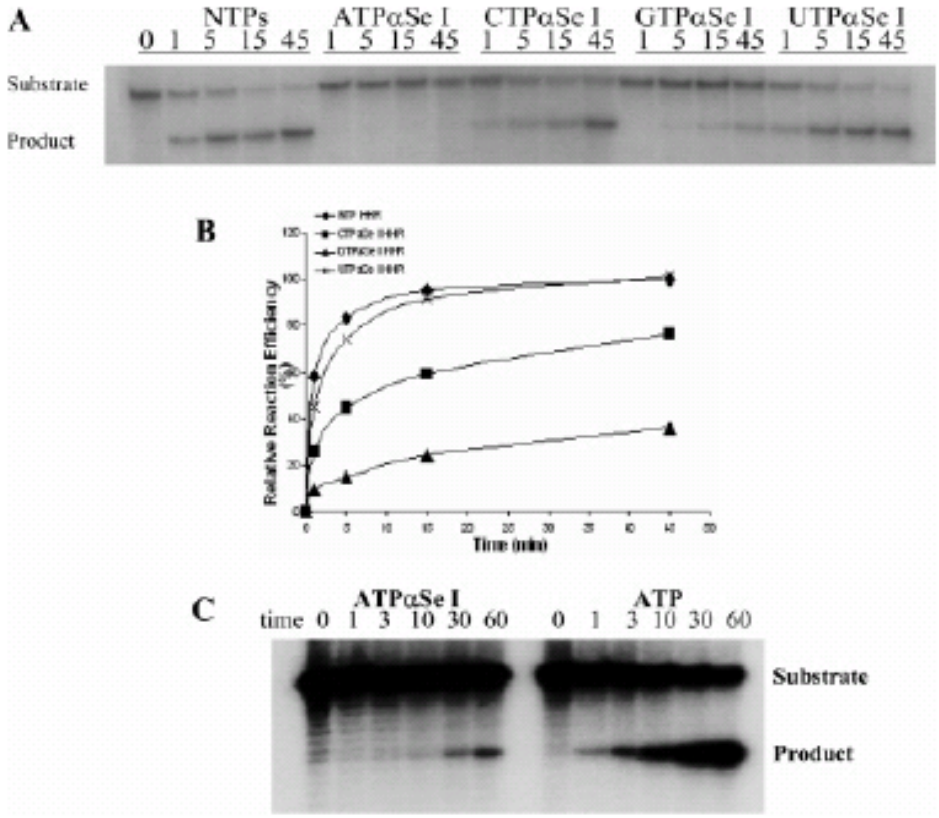
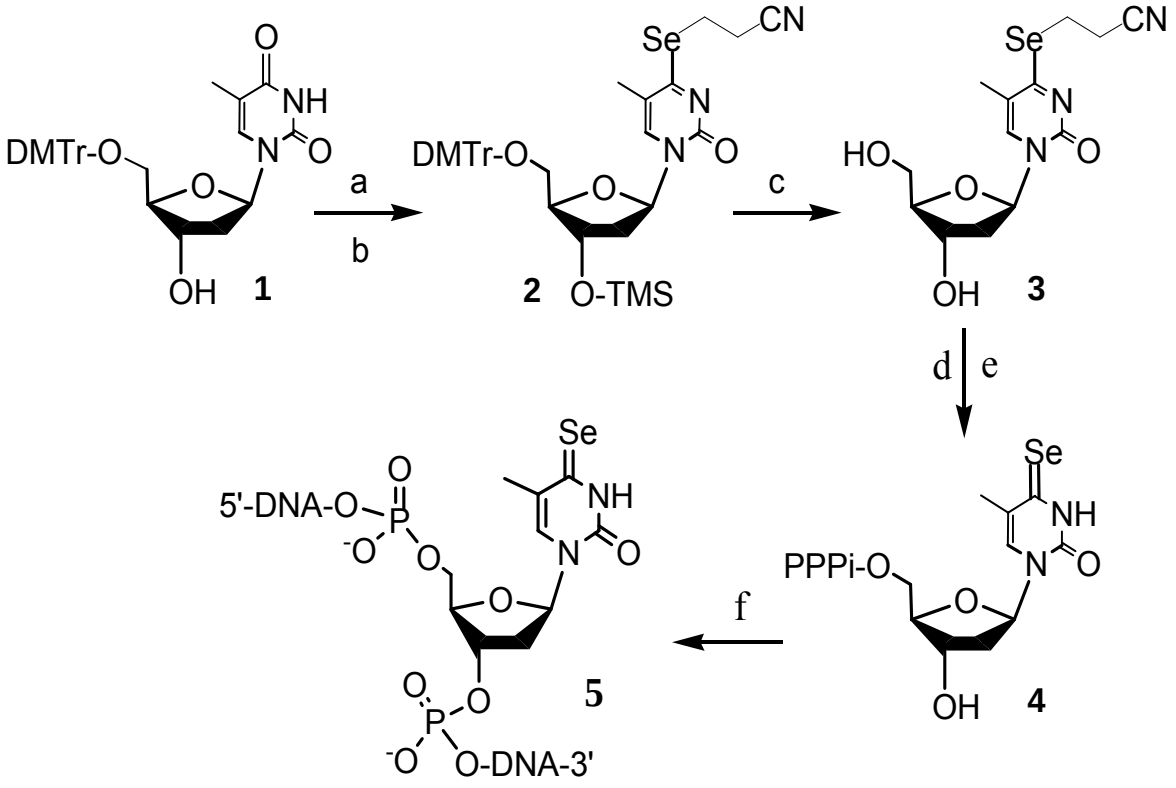


Figure 1.22: Catalysis and analysis of the modified and native hammerhead ribozymes. (A) Time-course ribozyme digestion of the RNA substrate. (B) Plot of the catalytic experimental results, relative to the activity of the native ribozyme. (C) Time-course Mn^{2+} rescue experiment of the ribozymes transcribed with ATPαSe and ATP.

Inspired by the previous results that the replacement of thymidine 4-oxygen with selenium didn't affect the DNA duplex stability and the hydrogen bond between A and T, very recently, we also synthesized a base selenium modified TTP and tried to incorporate it into DNA by DNA polymerase (Scheme 1.14). (83) As a result, this base

Se-triphosphate is proved a good substrate for DNA polymerases. Using a short DNA template (T1), SeTTP was incorporated into a short DNA primer (Figure 1.23A) and confirmed by MS analysis (Figure 1.23B). Using a longer DNA template, our time-course experiment (Figure 1.24A) has shown that SeTTP polymerization efficiency is at the same level of native TTP (Figure 1.24B). Both Klenow and Klenow exo- can incorporate SeTTP with high efficiency.



Scheme 1.14: Chemical synthesis of SeTTP (4) and enzymatic synthesis of SeT-containing DNAs (5). (a) TMS-Im; 1,2,4-triazole-$POCl_3$-Et_3N; (b) $(NCCH_2CH_2Se)_2/NaBH_4$, EtOH; (c) 80% acetic acid; (d) $POCl_3$, Me_3PO_4; tributylamine (TBA), pyrophosphate, DMF; H_2O; (e) 0.05 M K_2CO_3; (f) template-dependent DNA polymerization.

A.

B.

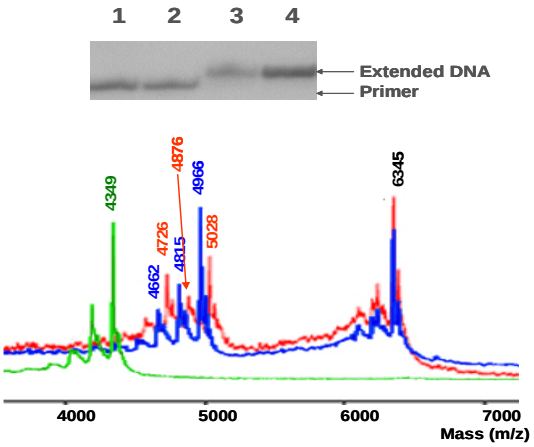


Figure 1.23: Incorporation of ^{Se}TTP or TTP by Klenow exo(-) on DNA template. (A): all lanes contain P1 and T1. Lane 1: no enzyme; 2: no TTP; 3: ^{Se}TTP and enzyme; 4: TTP and enzyme; (B): MS spectra of P1 (green color, $[M-H]^-$: 4349, calc. 4349.8), the T-extended DNA (O-15mer, blue, 4662, 4815 and 4966), and the ^{Se}T-extended DNA (Se-15mer, red, 4726, 4876 and 5028).

A.

B.

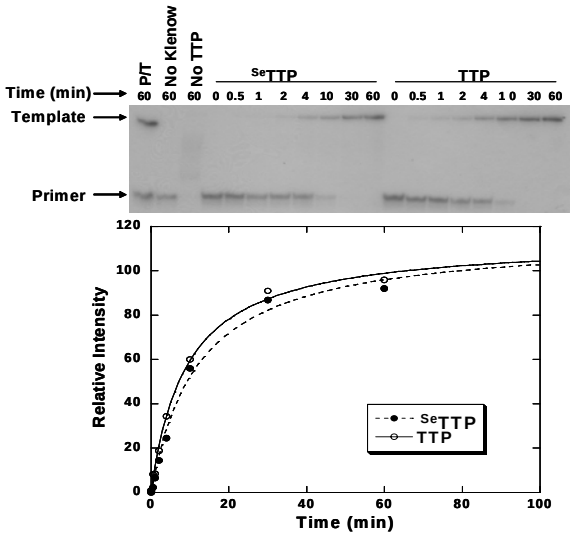


Figure 1.24: The time-course experiment of incorporating TTP and ^{Se}TTP into DNA. Use a DNA primer (P2, 5'-GCGTAATACGACTCACTATAG-3') and Klenow on a DNA template (T2, 3'-CGCATTATGCTGAGTGATATCCG-TTGGACTACTCCGGCTTTCCGG CTTTGCATGT-5'). (A): the gel electrophoresis; (B): the plot of the TTP and ^{Se}TTP incorporation into DNA.

Besides the direct incorporation of modified residues by enzymes, there is another way to make relative large selenium modified nucleic acids, using a comprehensive enzymatic ligation to fuse two or more Se-containing fragments together. As the starting work of this strategy, Micura's lab has successfully achieved the selenium derivatization for a rat spliceosomal U6 snRNA stem-loop motif using T4 RNA ligase, (84) which possesses the ability to join the 3'-hydroxy and the 5'-phosphate termini of individual ribonucleic acid strands. In this work, 2'-SeMe-U and C were used as selenium building blocks and four potential sites were investigated for the ligation yield as indicate by arrows in figure 1.25a. Finally, some criteria were founded about the "donors (providing the 5'-phosphate terminus)" and "accepters", based on which position 4 was selected with optimal sites of selenium derivatization compared (figure 1.24b).

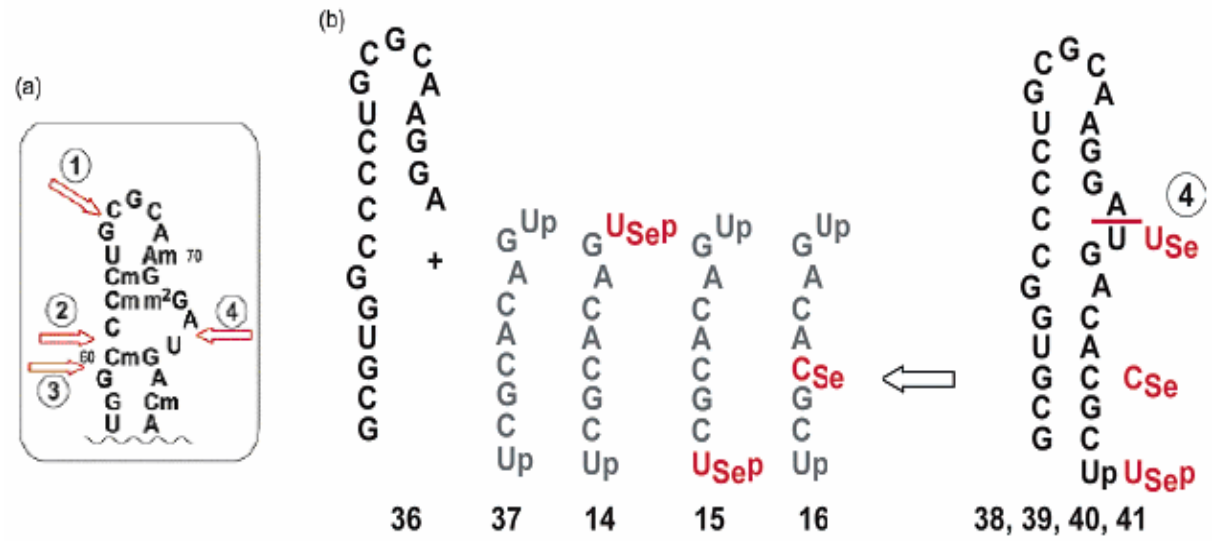


Figure 1.25: Target stem-loop motif of rat spliceosomal U6 snRNA (a) and the conception of the ligation experiments with T4 RNA ligase in position 4 (b).

Although the target RNA was only about 30 nucleotides in length, their experimental results have showed that selenium-modified RNAs are stable throughout enzymatic ligation conditions and have approved the potential application in making longer nucleotides containing one or multiple selenium modified residues. Later on, (58) they further selected the aptamer domain of the adenine deaminase (add) adenine-riboswitch as candidate and synthesized the relevant selenium modified sequences based on single-stranded ligation via T4 RNA ligase and on splinted ligation via T4 DNA ligase respectively. As showed in figure 1.26, they considered two potential ligation sites for the 71 nt RNA as indicated by arrows based on the standards they founded previously.

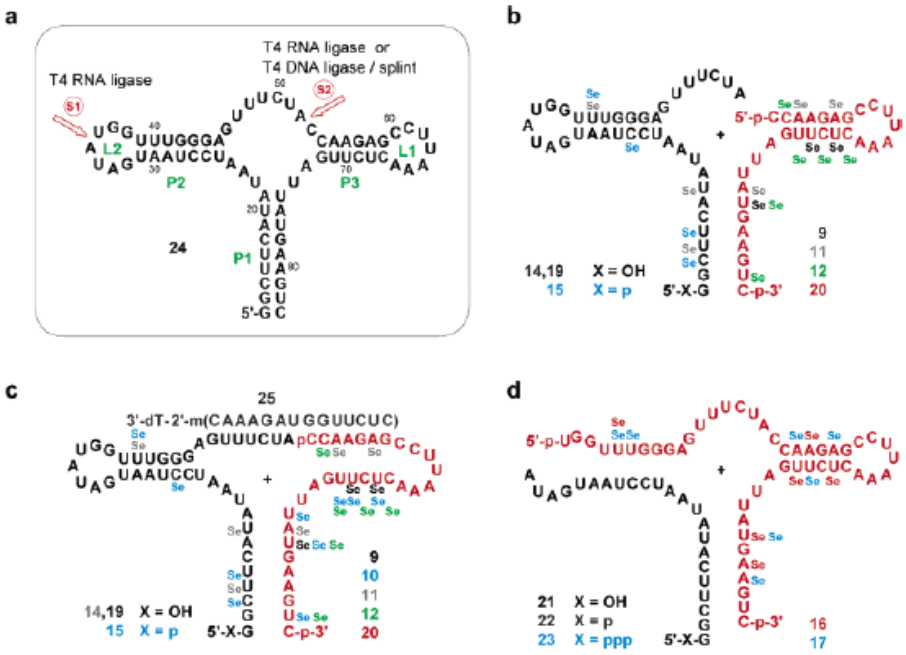


Figure 1.26: Enzymatic ligation of Se-modified RNA. (a) adenine riboswitch (add) target sequence and potential sites S1 and S2 for ligation with T4 RNA ligase and T4 DNA ligase, respectively. (b) T4 RNA ligations of 31 nt donor (5'-phosphate) and 40 nt acceptor (3'-OH) strands (for 3-10 Se-labels); (c) T4 DNA ligations of 31 nt donor and 40

nt acceptor strands with a 15 nt 2'-O-methyl RNA splint (for 3-10 Se-labels); (d) T4 RNA ligations of 48 nt donor and 23 nt acceptor strands (for 6-7 Se-labels).

1.5 Tellurium derivatized nucleic acids (TeNA)

Comparing the other chalcogen elements oxygen, sulfur and selenium which play unique roles in the functional groups of living systems, tellurium seems much quiet. So far, no similar structural and biological functions have yet been discovered in the case of the heavier tellurium. The underlying reasons can be sought in its more metallic properties which are expressed by a distinct weakening of covalent bonds with hydrogen and carbon atoms when compared with sulfur and selenium, but no further conclusive information about tellurium in living systems is available due to the lack of proper model. (85)

The first successful attempts to incorporate tellurium into proteins were made with Te-resistant fungi grown in the presence of tellurite on a sulfur-free medium, and the results were supportive for a biosynthesis of tellurium-containing amino acids. (86) More recently, telluromethionine (TeM) was found to replace Met residues in dihydrofolate reductase in high percentage (40%) using a methionine-auxotrophic E. coli strain and supplying synthetic telluromethionine to the medium. (87) Optimization of the selective pressure incorporation method finally allowed for quantitative TeM bioincorporation into a series of proteins, and the related X-ray crystallographic analysis confirmed the stability of tellurium function and the isomorphous character of these protein variants. (88) Comparing selenium, tellurium atom is not a good anomalous center for MAD application, however, the electron density of the tellurium atom is sufficient to generate clear signals in the isomorphous and anomalous difference Patterson maps at the

commonly used CuKα wavelength, which makes the synchrotron facility is not necessary to collect X-ray diffraction data of the Te-derivatized protein crystals. (89)

Unfortunately, the facile oxidation of telluromethionine was also found to present a serious drawback for the use of this methionine analogue for bioexpression of heavy atom variants of proteins, as well as the further application in protein structure determination, since oxidized forms were difficult to refold from inclusion bodies or to crystallize. (90) In particular, surface-located and, in a more pronounced manner, solvent-exposed TeM residues were found to oxidize readily and thus to generate electron densities of difficult assignment and local non-isomorphism. In terms of structure stability in solution, the per-telluromethionine-variant of annexin V was less stable than the related wild-type protein and the selenomethionine-variant. (91)

However, the success of introducing selenium into nucleic acids and the great developing of tellurium chemistry has inspired us to further explore the potential application of tellurium functionality in nucleic acid structure determination through X-ray crystallography or even cryo-EM by covalent incorporation of organic telluride into nucleic acid. Only if the isomorphism is confirmed as proof of principle, the limitation of tellurium oxidation wouldn’t cause drawback for nucleic acids since there are much less refolding issues comparing proteins.

1.6 The objectives and contributions of this dissertation

The objectives of this research are to further develop more SeNA and TeNA building blocks as powerful tools for much wider and convenient applications nucleic acids structure determination, as well as explore other functions of selenium and tellurium in nucleic acids. This research focuses on four key objectives:

A. Synthesizing the 2'-SeMe-thymidine building block

The selenium modified ribothymidine at 2'-position was obtained especially for some tRNA structure and function study. Meanwhile, this work have completed the set of 2'-Se-modification building blocks for all natural residues of nucleic acids, making the 2'-selenium labels more available. In addition, we also found that this 2'-modification strategy could greatly facilitate the crystal growth for many oligonucleotides sequences comparing their native counterparts.

B. Investigating base modified SeNA at 4 position of thymidine

This is the first trial of SeNA produce on the base. To our surprise, besides the structural application, this base SeNA could generate yellow DNAs with a special UV-absorption. The following structure and hydrogen bonding studies confirmed the isomorphism of this derivatization and provide a powerful tool for hydrogen bonding, base stacking studies.

C. Proof of TeNA principle by 2'-Te modified DNA

In this work, the tellurium function was incorporated into the 2'-position of the ribose in several DNA oligonucleotides for the first time. Although this derivatization showed some structure perturbation in terms of DNA duplex stability probably due to the modification location is in the minor groove, and no crystal could be growing in this work, it still could provide a potential tool to investigate the DNA backbone cleavage and base elimination damage caused by heavy atoms such as Te. In addition, a novel addition-elimination reaction based method was developed for making d4T analogs, a FDA proved anti-HIV drug.

D. Proof of TeNA principle by 5-Te modified DNA

Due to the structure perturbation caused by the 2'-Te modification, we also carried out the derivatization of tellurium at 5 position of nucleobase since it locates in the DNA major groove where provide more space for the accommodation of bigger atoms. In this work, a crystal structure of first Te-DNA was resolved in high resolution and the structure isomorphism of Te-structure at this position was also confirmed.

2. METHODS AND MATERIALS

This chapter summarizes some common used materials, methods and instruments through all this dissertation research, the detailed, special and additional method information will be provided in each chapter.

2.1 Synthesis of Se and Te modified nucleosides

Most solvents and reagents were purchased from Sigma, Fluka, or Aldrich and used without purification unless mentioned otherwise. The dimethyl ditelluride reagent was purchased from Organometallic Inc. Triethylamine (TEA) was dried over KOH (s) and distilled under argon. When necessary, solid reagents were dried under high vacuum. Reactions with compounds sensitive to air or moisture were performed under argon. Solvent mixtures are indicated as volume/volume ratios. Thin layer chromatography (TLC) was run on Merck 60 F254 plates (0.25 mm thick; Rf values in the text are for the title products), and visualized under UV-light or by a Ce-Mo staining solution (phosphomolybdate, 25 g; $Ce(SO_4)_2 \bullet 4H_2O$, 10 g; H_2SO_4, 60 mL, conc.; H_2O, 940 mL) with heating. Preparative TLC was performed using Merck 60 F254 pre-coated plates (1000 mesh, 2 mm thick). Flash chromatography was performed using Fluka silica gel 60 (mesh size 0.040-0.063 mm) using a silica gel:crude compound weight ratio of ca. 30:1. ^{1}H, ^{13}C and ^{31}P-NMR spectra were recorded using Bruker-300 or 400 (300 or 400 MHz). All chemical shifts (δ) are in ppm relative to tetramethylsilane and all coupling constants (*J*) are in Hz. High resolution mass spectrum (HRMS) analysis was performed at Georgia State University Mass-Spec center and Scripps Center for Mass Spectrometry, California.

2.2 Synthesis of phosphoramidites containing Se and Te

The common two phosphoramidite reagents: 2-cyanoethyl N,N-diisopropylchloro-phosphoramidite and 2-cyanoethyl N,N,N,N-tetraisopropylphosphorodiamidite were purchased from Chemgen. Inc., the 5-(benzylthio)-1H-tetrazole (5-BMT) is used as catalyst when the later one was used. The equivalence of phosphoramidite reagent is usually from 1.5 to 2.5 by monitoring the reactions. The access reagent could be removed by pentane or hexane precipitation. Usually, a flash neutral alumina column is needed to removed the organic salts which otherwise could affect the following solid phase synthesis.

2.3 Design and synthesis of oligonucleotides sequences containing Se and Te

2.3.1 PDB and NDB database

Most of the sequences applied for crystallography structure studies are referred from the protein and nucleic acids data base (PDB and NBD), as initial models to study the structure derivatization isomorphism. Their web address is: http://www.pdb.org/ and http://ndbserver.rutgers.edu/. The positions of modified residues are randomly selected and usually cover 5', middle and 3' regions of a sequence.

2.3.2 Solid phase synthesis

Both native and the modified oligonucleotides are synthesized through solid phase synthesis in the ABI392 DNA/RNA Synthesizer (Figure 2.1). The samples are usually synthesized in a 1.0 μmol scale, in several cases it could be 0.2 and 10 μmol. The general solid phase synthesis procedure is showed in figure 2.2.

Figure 2.1: ABI 392 Solid Phase Synthesizer.

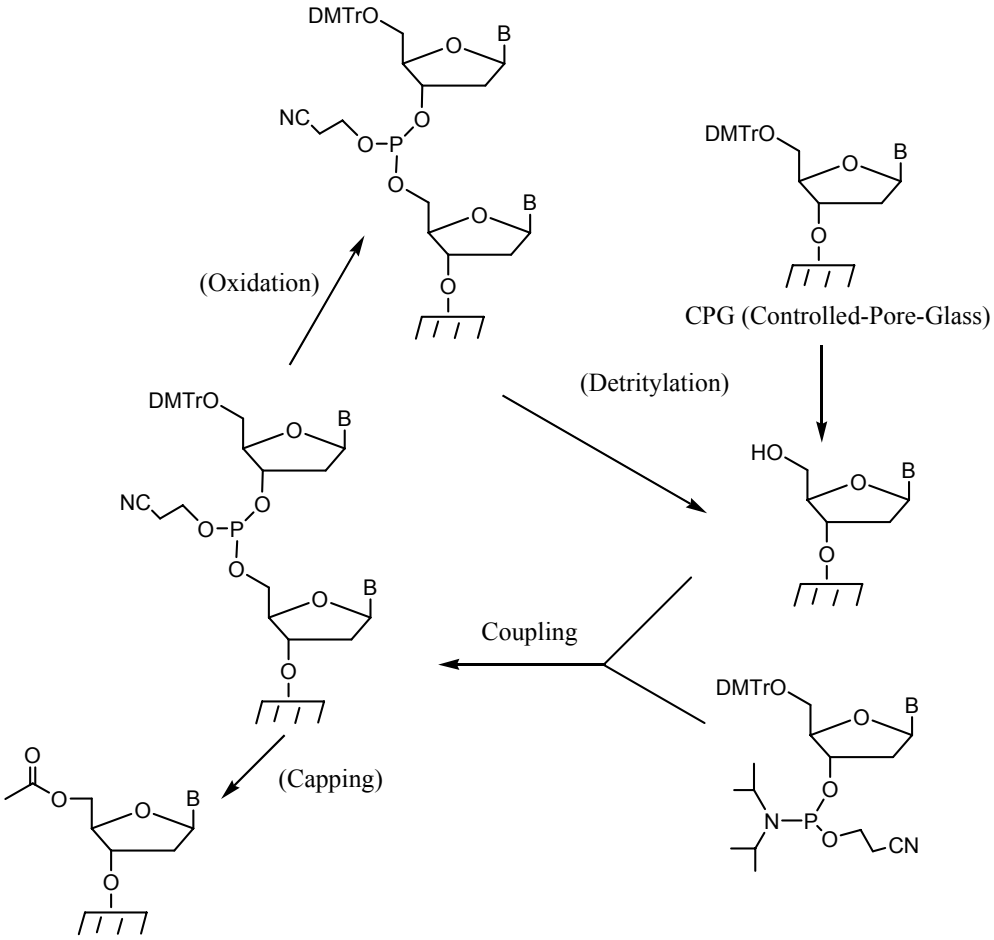


Figure 2.2: DNA solid phase synthesis procedure.

This procedure is performed from 3’ to 5’ of a sequence on control pore glass (CPG-500 or 1000) immobilized with the appropriate nucleoside through a succinate linker. All the reagents are purchased from Glen Research. In DNA works, regular DNA phosphoramidites (dA-CE, dG-CE, Ac-dC-CE, and dT-CE, 0.1 M in acetonitrile) are used for most of synthesis, when base selenium modification is involved, the ultra-mild phosphoramidites (Pac-dA-CE, Ac-dC-CE and iPr-Pac-dG-CE, also 0.1 M in acetonitrile) are used which provide a mild basic deprotection (0.05 M K_2CO_3 in methanol instead of ammonium hydroxide). In RNA works, 2’-TOM protected phosphoramidites are used for most of synthesis, when base selenium modification is involved, the ultra-mild 2’-TBDMS protected phosphoramidits are used (the concentration of both types of reagent is 0.1M in acetonitrile). Coupling steps are carried out using a 5-(benzylmercapto)-1H-tetrazole (5-BMT) solution (0.25 M) as activator in acetonitrile with 25 seconds coupling time. A 10% of THF solution of acetic anhydride, along with 16% MeIm in THF serves as capping reagents. 3% trichloroacetic acid in methylene chloride and 0.02 m iodine in THF/pyridine/H_2O are used as detritylation and oxidation reagents respectively in the last two steps. In this work, all the oligonucleotides were prepared with DMTr-on form, which means the last trityl group is reserved for the following purification purpose.

2.4 Oligonucleotides purification and characterization

After synthesis, the DNA oligonucleotides are cleaved from the solid support and fully deprotected by aqueous ammonia (conc.) treatment for 14 hr at 55°C. For the ultramild synthesized sample associated with base Se modification in this work, the cleavage step is finished by using 0.05 M K_2CO_3 in methanol for 8 hours at room temperature. Similarly, the 2’-O-TOM protected RNA oligonucleotides were deprotected

following recommendations from Glen Research with minor modifications. Briefly, 1 mL of a methylamine solution [prepared by mixing 40% aqueous methylamine with 33% ethanolic methylamine (Fluka) in a 1:1 ration] was added to an eppendorf tube containing the RNA resin (1 μmol), followed by incubation for 6 hr at 35°C or overnight at room temperature. The supernatant was evenly transferred into two sterile 1.5 mL-eppendorf tubes. After these tubes were chilled at -20°C for 10 min, the solvents were completely evaporated on speed vacuum. Each portion was treated with 25 equivalents of tetrabutylammonium fluoride [0.25 mL (for 0.5 μmol RNA 20mer), 1.0 M in THF, Fluka] and incubated for 6 hr at 35°C. Following evaporation of THF on speed vacuum, the two portions were combined in 1.0 mL of 1.0 M Tris-HCl buffer (RNase-free, pH 7.5), incubated with shaking overnight at 25°C, and filtrated through a 0.2 μm pore-sized filter. The 5'-DMTr deprotection of both DNA and RNA oligonucleotides was performed in a 2% trichloroacetic acid solution (from 10% w/w, 0.9 M in water) for 1.5 min, followed by neutralization to pH 7.0 with a freshly made aqueous solution of triethylamine (1.1 M) and petroleum ether extraction to remove DMTr-OH.

2.4.1 High Performance Liquid Chromatography (HPLC)

The DNA and RNA oligonucleotides were analyzed and purified by reverse phase high performance liquid chromatography (RP-HPLC) both DMTr-on and DMTr-off. Purification was carried out using a 21.2x250 mm Zorbax, RX-C8 column at a flow rate of 6 mL/min. Buffer A consisted of 20 mM triethylammonium acetate (TEAAc, pH 7.1, RNase-free water), while buffer B containing 50% aqueous acetonitrile in buffer A. Similarly, analysis was performed on a Zorbax SB-C18 column (4.6x250 mm) at a flow of 1.0 mL/min using the same buffer system. The DMTr-on oligonucleotides were

eluded with up to 100% buffer B in 20 min in a linear gradient, while the DMTr-off oligonucleotides were eluted with up to 70% of buffer B in a linear gradient in the same period of time. The collected fractions were lyophilized; the purified compounds were re-dissolved in RNase-free water. The pH was adjusted to 7.0 after the final purification of the Se-oligonucleotides without the DMTr group.

2.4.2 Mass spectrometry

The ESI-MS was performed using a Waters Micromass Q-TOF micro instrument. The data were acquired in both positive and negative ion modes by syringe pump infusion of the sample solutions at a flow rate of 7 μL/min. The MALDI mass spectrometry analysis was performed on an Applied Biosystems 4800 plus MALDI TOF/TOF analyzer mass spectrometer (Framingham, MA). This instrument is equipped with a Diode-pumped Nd:YAG laser at 355-nm. The mass spectra were acquired as an average of 3 shots with a same laser intensity attenuator setting (4002 arbitrary unit) for all samples. All spectra were recorded on a biomass spectrometer in linear-negative mode with delayed extraction. 3-hydroxypicolinic acid (3-HPA)/diammonium citrate (9:1) in water was used as the matrix. A 10:1 matrix:sample mixture was prepared, applied to the sample plate as 1 μL drops, and allowed to air dry before the analysis. 25 kV were applied as the acceleration voltage. A mass range of 1000–10000 was scanned. Each spectrum was summed from multiple spectra at different spots. Proteins, such as insulin, thioredoxin, and apomyoglobin, were used as external standards. This work was performed by Dr. Siming Wang, Dr. Lupei Du and Yanyi Chen at Mass Spectrometry Center of Georgia State University. In addition, the Scrippts center of mass spectrometry can also provide similar service.

2.5 Nucleic acids stability and thermaldenaturization (Tm) studies

Solutions of the duplex DNAs (2 μM) were prepared by dissolving the DNAs and RNAs in a buffer containing NaCl (50 mM), sodium phosphate (10 mM, pH 7.2), EDTA (0.1 mM) and $MgCl_2$ (10 mM) . The solutions were then heated to 95 °C for 2 min, cooled slowly to room temperature, and stored at 4 °C overnight before measurement. Prior to thermal denaturation, helium was bubbled through the samples. Denaturation curves were acquired at 260 nm by heating and cooling between 5 to 80 °C for four times in a rate of 0.5 °C/min using Cary-300 UV-Visible Spectrometer equipped with a Peltier Temperature Controller. The data were analyzed in accordance with the convention of Puglisi and Tinoco. (92)

2.6 Circular dichroism spectroscopy

The CD spectra of samples were recorded in a Jasco-810 spectropolarimeter at 25 ºC. The DNA spectrum were measured in a 1-mm pathlength cell with a UV-absorption at around 0.5. After subtracting the spectrum of blank solvent, all sample spectra were obtained as the average of at least five scans from 200 to 500 nm with a scan rate of 100 nm/min.

2.7 Crystallization of modified and non-modified nucleic acids

The most popular used hanging drop vapor diffusion technique is applied for the crystallization of all the oligonucleotides in this work. And all the experiments are carried out in greased VDX 24-well plate. The principle of vapor diffusion is showed in figure 2.3. A drop composed of a mixture of sample and reagent is placed in vapor equilibration with a liquid reservoir of reagent. Typically the drop contains a lower

reagent concentration than the reservoir. To achieve equilibrium, water vapor leaves the drop and eventually ends up in the reservoir. As water leaves the drop, the sample undergoes an increase in relative supersaturation and crystallization out of the solution. Equilibration is reached when the reagent concentration in the drop is approximately the same as that in the reservoir.

In this work, the purified oligonucleotides (1 mM) were heated to 70°C for 2 minute, and cooled slowly to room temperature before hanging drop crystal setting. Both native buffer (from the original literatures) and Nucleic Acid Mini Screen Kit (Hampton Research) were applied to screen the crystallization conditions at different temperatures.

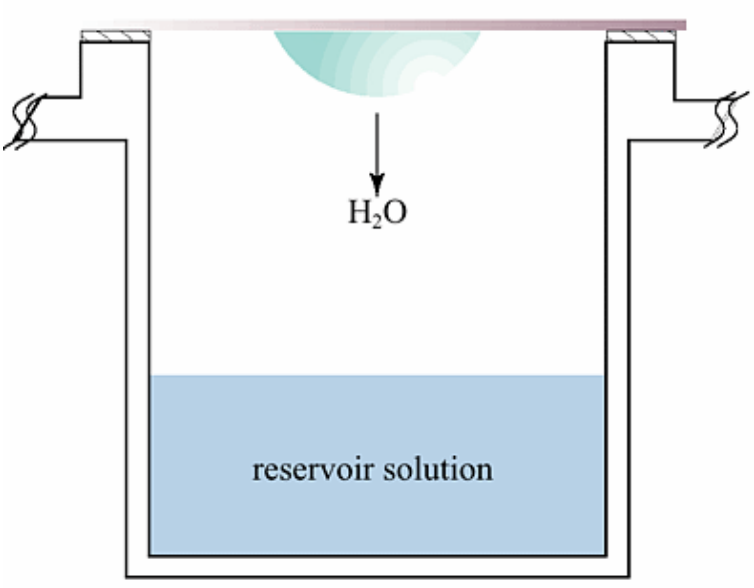


Figure 2.3: Hanging drop vapor diffusion for crystallization.

2.8 Crystal growth monitor

All the crystal growth was monitored by Olympus BX-60 optical microscope equipped with Olympus U-CMAD3-3.3 MPX digital camera, which is used to take the

crystal pictures in high resolution. The software Rincon from Imagingplanet, Inc. was applied to measure and calculate the crystal parameters.

2.9 Crystal data collection and structure determination

All the diffraction data were collected in beamline X-12B and X-12C of NSLS (National Synchrotron Light Source) in Brookhaven National Laboratory. The obtained crystals were mounted through Magnetic Crystal Cap with different sizes of pins on the top. 30% glycerol, PEG 400, 35% MPD, perfluoropolyether, Paratone-N or the mixture of mother liquid with certain amount of MPD were normally used and screened as cryoprotectants before the crystals were freezing in liquid nitrogen. The data collection was taken under the liquid nitrogen stream at 99°K either by automounter system or physically on-site. Usually, a number of crystals were scanned to find the one with strong anomalous scattering at the K-edge absorption of selenium. The distance of the detector to the crystals was adjusted depending of the diffraction, usually set to 150 mm. The wavelengths of 0.9795 Å was chosen for selenium SAD phasing, while 0.9792 Å, 0.9795 Å and 0.9400 Å were chosen for selenium MAD phasing collection. The crystals were exposed for 10 or 15 seconds per image with an oscillation angle of one degree, and a total of 180~360 images were taken for each data set based on different space groups. All the data were processed using HKL2000 and DENZO/SCALEPACK. (93)

The crystal structure of this Se and Te modified oligonucleotides were solved by molecular replacement, followed by the refinement of selenium and tellurium atom positions in CNS. (94) The refinement protocol includes simulated annealing, positional refinement, restrained B-factor refinement, and bulk solvent correction. Besides the stereo-chemical topology and geometrical restrain parameters of DNA/RNA,(95) the

topologies and parameters for modified dU with selenium and tellurium were constructed and applied. After several cycles of refinement and the model rebuilding, a number of highly ordered waters were added. Finally, the occupancies of Se and Te were refined. Cross-validation with a 5-10% test set was monitored during the refinement. (96) The σA-weighted maps of the (2m|Fo| - D|Fc|) and the difference (m|Fo| - D|Fc|) maps were computed and used throughout the model building. (97)

3. SYNTHESIS AND STRUCTURE STUDY OF THE 2'-SeMe-THYMIDINE CONTAINING OLIGONUCLEOTIDES

(This work has been published on Organic Letters, 2007, 9 (5): 749-752, Jiansheng Jiang, Jozef Salon and Zhen Huang are listed as co-authors.)

3.1 Introduction

As stated in the introduction part, SeNA is built by chemical replacement of oxygen atoms in nucleic acids with selenium. Therefore, there are multiple positions available in different residues. Besides the criteria of structure isomorphism, synthetic accessibility and stability are the most important factors need to be firstly considered to design the right applicable SeNA molecules. The previous results from our and other research groups have confirmed the advantage of selenium modification at 2' position of nucleoribose, which actually, this is also one of the very early starting points of this young research area. The methylated selenium function is stable enough to tolerate not only the normal solid phase synthesis including strong acid and base treatment, but also the long time high energy X-ray exposure, more important, although locating in the minor groove of DNA, this group doesn't cause significant global and local structure perturbations. Because of that, all this series of 2'-SeMe modified A, G, C, U building blocks have already been developed recently and especially, the 2'-Se-uridine-CE is now commercially available from Glen research as a novel phosphoramidite. The easily accessibility and stability make this 2'-SeMe derivatization widely used in the determination of some more complicated nucleic acids structures, especially in the RNA research area, since this is more like a RNA analog.

Considering that ribothymidine is also a common modification in maturized tRNAs, (98) it's worthy developing the synthesis of this 2'-SeMe modified thymidine as an

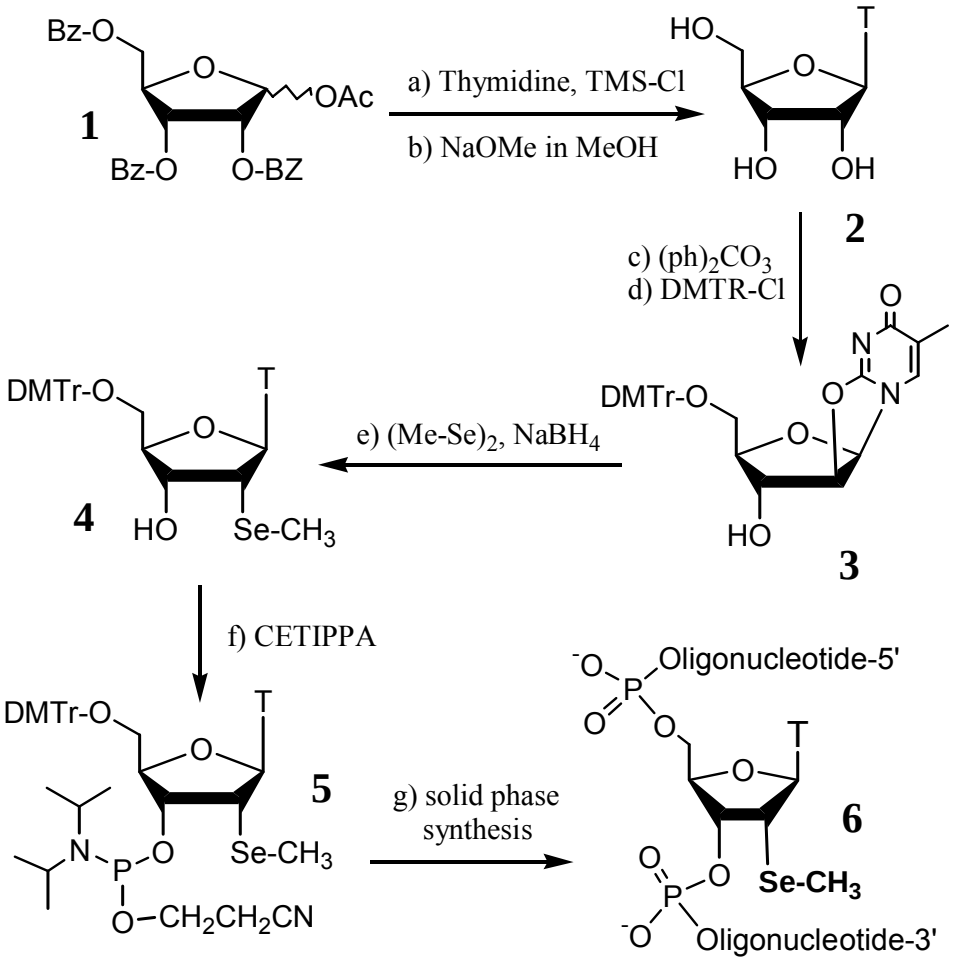
analog for both thymidine and ribothymidine in DNA and RNA researches. (99) (100) It can also serve as the supplementary building blocks to complete this 2'-Se modification family. In this chapter, the first synthesis of 2'-SeMe-thymidine phosphoramidite and its containing DNA and RNA oligonucleotides are described, as well as a crystallography structure study based on this derivatization.

3.2 Results and discussion

3.2.1 Synthesis of 2'-SeMe-thymidine and the phosphoramidite

As showed in scheme 3.1, the synthesis of the 2'-methylseleno-ribothymidine derivative (4) started from the glycosylation of the protected D-ribose (1) with thymine under the Vorburggen conditions. Due to the 2'-neiboring group effect, this glycosylation generated exclusively the protected β-ribothymidine derivative, followed by the NaOMe treatment to give fully-deprotected ribothymidine (2) in 88% yield over two steps. After the cyclization of 2 in DMF, the 2,2'-anhydrothymidine was selectively protected with DMTr group at the 5'-position to give 3 in 87% yield in two steps. Unlike the synthesis of the 2'-methylseleno-uridine, we initially encountered problems of low yield and long reaction time in the introduction of the selenium functionality into anhydroribothymidine 3. We attempted to increase the reaction yield by using different reducing reagents, such as $LiAlH_4$ and DIBAH to form the methyl selenide nucleophile via the reduction of dimethyl diselenide. Unfortunately, we always observed the reduction of the nucleobase, which leaded to lower yields. This problem was later overcome by using a weaker reducing reagent ($NaBH_4$) at a higher temperature (50 ºC) in pure EtOH, which gave high yield in three hours. Finally, the 2'-Se-ribothymidine (4) was converted to the corresponding phosphoramidite (5) in 90% yield, in dry CH_2Cl_2 using 2-cyanoethyl

N,N,N,N-tetraisopropyl phosphoramidite and 5-benzylmercapto-tetrazole (5-BMP, as catalyst). The detailed synthesis steps and the characterization of each intermediate are listed in the following.



Scheme 3.1: Synthesis of 5'-DMTr-2'-MeSe-thymidine phosphoramidite (5) and oligonucleotides containing this Se derivatization (6).

5-Methyluridine or ribothymidine (2)

Thymine (3.80 g, 30.23 mmol) and 1-O-acetyl-2,3,5-tri-O-benzoylribose (12.7 g, 25.4 mmol) were suspended in dry acetonitrile (350 mL), followed by adding trimethylsilyl

chloride (3.2 mL, 25.9 mmol), hexamethyldisilazane (5.3 mL, 25.4 mmol), and tin chloride (2.97 mL, 25.4 mmol), then the suspension was heated to reflux for 1.5 h. The solution was concentrated in vacuum (approximately 20 mL) and dissolved in methylene chloride (100 mL). The organic solution was washed with water (20 mL), saturated sodium bicarbonate (2 X 20 mL), and brine (20 mL) , dried over $MgSO_4$ and concentrated to give 2',3',5'-tri-O-benzoyl-5-methyluridine as a white solid (14.36 g). ^{1}H-NMR (400 MHz, $CDCl_3$) δ:1.61 (s, 3H, 5-CH_3), 4.72-4.78 (m, 1H, H-2'), 4.68 and 4.92 (2x dd, J_1 =2.4 Hz, J_2 = 12 Hz, 2H, H-5'), 5.78 (t, J =6.0 Hz, 1H, H-3'), 4.92-4.97 (m, 1H, H-4'), 6.43 (d, J =6.2 Hz, 1H, H-1'), 7.18-8.14 (m, 16H, aromatic H and H-6). The white solid (14.36 g, 25.2 mmol), without further purification, was dissolved in methanol (350 mL). Sodium methoxide (8.16 g, 151.2 mmol) was then added, and the reaction mixture was stirred at room temperature overnight. The solution was neutralized with Dowex 50 X 8-4200 ion-exchange resin (approximately 5 g, monitored by wet pH paper), concentrated, and dissolved in water (100 mL), which then be lyophilized after extracting with ethyl ether (2 X 150 mL). The lyophilized sample was recrystallized from absolute ethanol to give compound **2** as a white solid (5.76 g, 88% in two steps). ^{1}H-NMR (400 MHz, D_2O) δ:1.78 (s, 3H, 5-CH_3), 3.72 and 3.83 (2 X dd, J_1 =4.0 Hz, J_2 = 12.8 Hz, 2H, H-5'), 4.02 (m, 1H, H-4'), 4.15 (t, J =2.6 Hz, 1H, H-3'), 4.25 (t, J =2.5 Hz, 1H, H-2'), 5.82 (d, J =6.0 Hz, 1H, H-1'), 7.60 (s, 1H, H-6). ^{1}H-NMR spectrum is identical to the literature (18).

2,2'-Anhydro[1-(β-D-arabinofuranosyl)-5-methyluridine]

Diphenylcarbonate (7.0 g, 0.033 mol) and sodium bicarbonate (0.16 g, 1.8 mmol) were added to the DMF (25 mL) solution of 5-methyluridine (5.6 g, 0.022 mol). The mixture was heated to reflux, and carbon dioxide was generated. After 1.5 hour, the darkened

solution was concentrated to approximately 4 mL under reduced pressure. The black residue was poured into diethylether (200 mL) under vigorous stirring, and a brown gum product was formed. The ether was decanted and the residue was dissolved in a minimum amount of methanol. The crude product was precipitated again with fresh ether (150 mL) to yield a stiff light brown gum. The ether was decanted and the gum was dried under high vacuum over night to give a solid that was easily crushed to give a light tan powder (4.4 g, 85% crude yield). The material could be used directly in the following step without further purification. ^{1}H-NMR (400 MHz, d-DMSO) δ: 1.79 (s, 3H, 5-CH_3), 3.12-3.25 (m, 2H, H-5'), 4.05-4.11 (m, 1H, H-4'), 4.35-4.41 (m, 1H, H-3'), 5.19 (d, *J* = 5.6 Hz, 1H, H-2'), 6.29 (d, *J* =6.2 Hz, 1H, H-1'), 7.76 (s, 1H, H-6). ^{1}H-NMR spectrum is identical to the literature (19).

2,2'-Anhydro-1-[2'-deoxy-5'-O-(4,4-dimethoxytrityl)-β-D-arabinofuranosyl-5-ethyluridine]: The starting material, 2,2'-anhydro[1-(β-D-arabinofuranosyl)-5-methyluridine] (4.0 g non-purified after the cyclization step), was co-evaporated with dry pyridine (3 X 5 mL) and then dissolved in dry pyridine (25 mL). A first half of dimethoxytrityl chloride (2.36 g, 6.95 mmol) was added and the solution was stirred at room temperature for one hour. Then a second half of dimethoxytrityl chloride was added and the reaction was completed after stirred for additional one hour, followed by quenching with methanol (4 mL). The solution was evaporated and co-evaporated with CH_3CN (2 X 5 mL). The residue was then dissolved in $CHCl_3$ (40 mL) and the solution was washed with saturated sodium bicarbonateand (2 X 15 mL) and saturated sodium chloride (2 X 15 mL). The organic phase was dried over Na_2SO_4 (s) and evaporated after the filtration of the drying agent. The obtained crude product (6.9 g) was purified on a silica gel column equilibrated with methylene chloride, and the pure product was eluted with a methanol/CH_2Cl_2 gradient

(methanol in CH_2Cl_2, 0-3%). The fractions containing the product were combined, evaporated and dried on high vacuum overnight to yield the pure product (**3**, 7.80 g) in 87% yield (calculated from ribothymidine). ^{1}H-NMR (400 MHz, $CDCl_3$) δ: 1.92 (s, 3H, 5-CH_3), 2.94-3.09 (m, 2H, H-5'), 3.79 (s, 6H, 2x CH_3O), 4.36-4.42 (m, 1H, H-4'), 4.49-4.51 (m, 1H, H-3'), 5.26 (dd, J_1 =1.4 Hz, J_2 = 5.8 Hz, 1H, H-2'), 6.15 (d, J=6.0 Hz, 1H, H-1'), 6.76-6.80 (m, 4H, aromatic), 7.17 -7.35 (m, 10H, H-6 and 9 aromatic protons). ^{13}C-NMR ($CDCl_3$) δ: 55.23 (OCH_3), 63.01 (C-5'), 75.63 (C-3'), 86.23 (Ar-C) 87.41 (C-4'), 89.09 (C-2'), 90.40 (C-1') , 118.75 (C-5), 113.21, 126.95, 127.90, 129.84, 135.45, 144.37, 158.50, 159.27 (Ar-C), 131.22 (C-6), 159.27 (C-2), 172.86 (C-4). HRMS (ESI-TOF): molecular formula, $C_{31}H_{30}N_2O_7$; $[M+H]^+$:543.2051 (calc.543.2053).

2'-methylseleno-5'-O-(4,4'-dimethoxytrityl)-5-methyluridine (4)

$NaBH_4$ (0.56 g, 14.72 mmol) was placed in a 50 ml round ask and suspended in dry THF (15 mL) under vigorous stirring. Dimethyl diselenide ($CH_3SeSeCH_3$, 1.47 ml, 14.72 mmol) was slowly injected in an ice-water bath under dry argon. Anhydrous ethanol (2 ml) was added dropwisely until gas bubbles started in the yellow mixture. The reaction mixture turned white or light yellow after 30 min, then a solution of **3** (4.0 g, 7.36 mmol) in THF (8 ml) was injected into the flask. The ice-water bath was removed after the addition, and the reaction was heated at 50 ºC and monitored by TLC (5% CH_3OH in CH_2Cl_2, product Rf = 0.47). The reaction was completed in 3 h, and the solvent was removed under reduced pressure. The residue was dissolved in methylene dichloride (20 mL) and the solution was washed with NaCl solution (sat., 3 X 10 mL). The organic layer was dried over anhydrous $MgSO_4$, followed by the filtration and solvent evaporation. The residue was then purified on a silica gel column (equilibrated with CH_2Cl_2) and the column was eluted with a methanol/methylene chloride gradient

(CH_3OH in CH_2Cl_2, 0-3%) to afford the pure foamy product (**4**, 4.24 g) in 82 % yield. ^{1}H-NMR (400 MHz, $CDCl_3$) δ: 1.45 (s, 3H, 5-CH_3), 2.12 (s, 3H, CH_3Se), 2.82-2.87 (m, 1H, OH), 3.39-3.60 (dd, J_1 =2.0 Hz, J_2 = 10.6Hz, 2H, H-5'), 3.62-3.66 (dd, J_1 =5.2 Hz, J_2 = 8.8 Hz , 1H, H-2'), 3.82 (s, 6H, CH_3O), 4.23-4.25 (m, 1H, H-4'), 4.38-4.40 (dd, J_1 =2.0 Hz, J_2 = 5.2 Hz, 1H, H-3'), 6.26 (d, J = 8.8 Hz, 1H, H-5), 6.81-6.88 (m, 4H, aromatic), 7.25-7.43 (m, 9H, aromatic), 7.62 (d, J = 1.2 Hz, 1H, H-6), 8.62 (br, 1H, NH). ^{13}C-NMR ($CDCl_3$) δ: 4.69 ($SeCH_3$), 11.69 (5-CH_3), 50.17 (C-2'), 55.30 (OCH_3), 63.78 (C-5'), 72.28 (C-3'), 84.52 (C-4'), 87.08 (C-1'), 87.25 (Ar-C), 111.79 (C-5), 113.33, 127.31, 128.14, 130.11, 135.12 (Ar-C), 135.19 (C-6), 144.18 (Ar-C), 150.46 (C-2), 158.82 (Ar-C), 163.49 (C-4). HRMS (ESI-TOF): molecular formula, $C_{32}H_{34}N_2O_7Se$; $[M-H]^-$: 637.1446 (calc.637.1453).

3'-O-(2-cyanoethyl-N,N-diisopropylphosphoramidite)-2'-methylseleno-5'-O(4,4'-dimeth-oxytrityl)-5-methyluridine (5)

The starting material **4** (2.4 g , 3.76 mmol) and the catalyst 5-Benzylthio-1H-Tetrazole (0.06 g, 0.3 mmol) were placed in a 100 ml round flask and dried on high vacuum over night. Dry CH_2Cl_2 (45 ml) was added followed by adding 2-cyanoethyl N,N,N,N-tetraisopropyl phosphane (1.36 g, 4.5 mmol) and a few drops of N,N-diisopropylethylamine. The reaction mixture was stirred at 0°C under dry argon for 30 min, followed by stirring for 2 h at room temperature until most of the starting material was converted. The reaction mixture was then quenched with $NaHCO_3$ (10 ml, sat.) by stirring for 15 min, and extracted with CH_2Cl_2 (3 X 10 ml). The combined organic layer was washed with NaCl (3 X 10 ml, sat.) and dried over anhydrous $MgSO_4$ for 15 min, followed by the filtration and solvent evaporation. The crude product was re-dissolved in CH_2Cl_2 (3 ml), and precipitated in petroleum ether (200 ml) under vigorous stirring,

generating white powder. The petroleum ether solution was then decanted carefully. The crude product was dissolved in CH_2Cl_2 and loaded on a column packed with aluminum oxide (activated, neutral, 150 mesh, 58 Å) and equilibrated with CH_2Cl_2. The column was then eluded with a gradient solution from CH_2Cl_2 to 30% of EtOAc in CH_2Cl_2 containing 1% triethylamine. The product fractions were collected based on the UV light and evaporated under the reduced pressure. The purified compound was re-dissolved in 3 ml of CH_2Cl_2, followed by the precipitation again in petroleum ether (200 mL, as indicated above) to yield a white foamy product **5** as the diastereomer mixture (2.6 g) in 83 % yield. The compound was dried on the high vacuum for over night before being used in solid phase synthesis. ^{1}H-NMR (400 MHz, $CDCl_3$) δ: 0.89-1.42 (m, 24H, 8xCH_3-ipr), 1.40 (2x s, 6H, 2x 5-CH_3), 2.10 (2x s, 6H, 2x CH_3Se), 2.42 and 2.69 (2x t, *J* =7.8 Hz, 4H, 2x CH_2-CN), 3.48-3.72 (m, 12H, 4x CH-ipr, C$\underline{H_2}$-CH_2-CN, 2x H-2', 2x 2H-5'), 3.82 (s, 12H, 4x CH_3O), 3.90-4.05 (m, 2H, C$\underline{H_2}$-CH_2-CN), 4.22 and 4.31 (2x m, 2H, 2x H-4'), 4.58-4.72 (m, 2H, 2x H-3'), 6.39 and 6.42 (d and d, *J* =8.6 Hz, 2H, 2x H-1'), 6.79-6.92 (m, 8H, aromatic protrons), 7.20-7.45 (m, 18H, aromatic protons), 7.59 and 7.62 (2x s, 2H, H-6), 8.18-8.31 (br, 2H, 2x NH).$^{13}\underline{\text{C-NMR}}$ ($CDCl_3$) δ: 4.57 and 4.71 ($SeCH_3$), 11.53 and 11.59 (5-CH_3), 20.11, 20.18, 20.46, 20.53, 22.50, 24.38, 24.45, 24.58, 24.66, 24.77,24.85, 43.23, 43.30, 43.35, 43.42, 46.36, 46.43, 47.38, 47.43, 50.36 (C-2', CH_2-CH_2-CN and C-ipr), 55.31 (OCH_3), 63.28, 63.33 (C-5'), 74.91 and 75.09 (C-3'), 84.97 (C-4'), 87.29 (Ar-C), 89.38 and 89.25 (C-1'), 111.69 (C-5), 113.46, 117.29, 126.99, 127.81, 128.54, 130.21, 130.42, 135.22, 135.40 (Ar-C), 135.44 (C-6), 144.10 (Ar-C), 150.37 (C-2), 158.85 (Ar-C), 163.34 (C-4). HRMS (ESI-TOF): molecular formula, $C_{41}H_{51}N_4O_8PSe$; $[M+H]^+$:839.2675 (calc.839.2682).

3.2.2 Incorporation of Se-functionality into oligunucleotides

Similar to solid-phase oligonucleotide synthesis using the non-modified thymidine and 2'-SeMe-uridine phosphoramidites, this novel 2'-Se-ribothymidine phosphoramidite was incorporated into several DNA oligonucleotides under the standard conditions (here we still use 0.10 M as the original concentration for all the phosphoramidite building blocks, although as low as 0.05 M concentrations are still working) in the presence of 5-BMP as the coupling catalyst. This Se-phosphoramidite could also be incorporated into RNAs under the same conditions, using a longer coupling time (usually 10 to 12 mins instead of 25s for DNA). As we reported previously, the 2'-selenium functionality was very stable under the mild I_2 treatment (20 mM, 20 seconds) for the phosphite oxidation to the according phosphate at each cycle. The synthesized oligonucleotides were cleaved off the solid support (CPG: control pored glass) and fully deprotected with concentrated ammonia at 55 ºC overnight, followed by the reversed-phase HPLC purification twice: DMTr-on and DMTr-off purification. A typical HPLC analysis profile of the Se-DNAs with the DMTr groups on is shown in Figure 3.1A. The HPLC analysis indicated that the Se-modified and non-modified phosphoramidites have the same incorporation efficiency and that the total formation of the short abortive oligonucleotides is less than 5%, which revealed a high coupling yield per synthetic cycle (higher than 99%), including the Se-T phosphoramidite incorporation. A representative reverse phase HPLC purification profile of the DMTr-off Se-DNAs is shown in Figure 3.1B, which used to further approve the purity of oligonucleotides. All of the purified Se-oligonucleotides were also confirmed by MALDI-TOF MS analysis. A typical MS spectrum of the Se-DNAs is shown in Figure 3.2, and all MALDI-MS data are collected in Table 3.1.

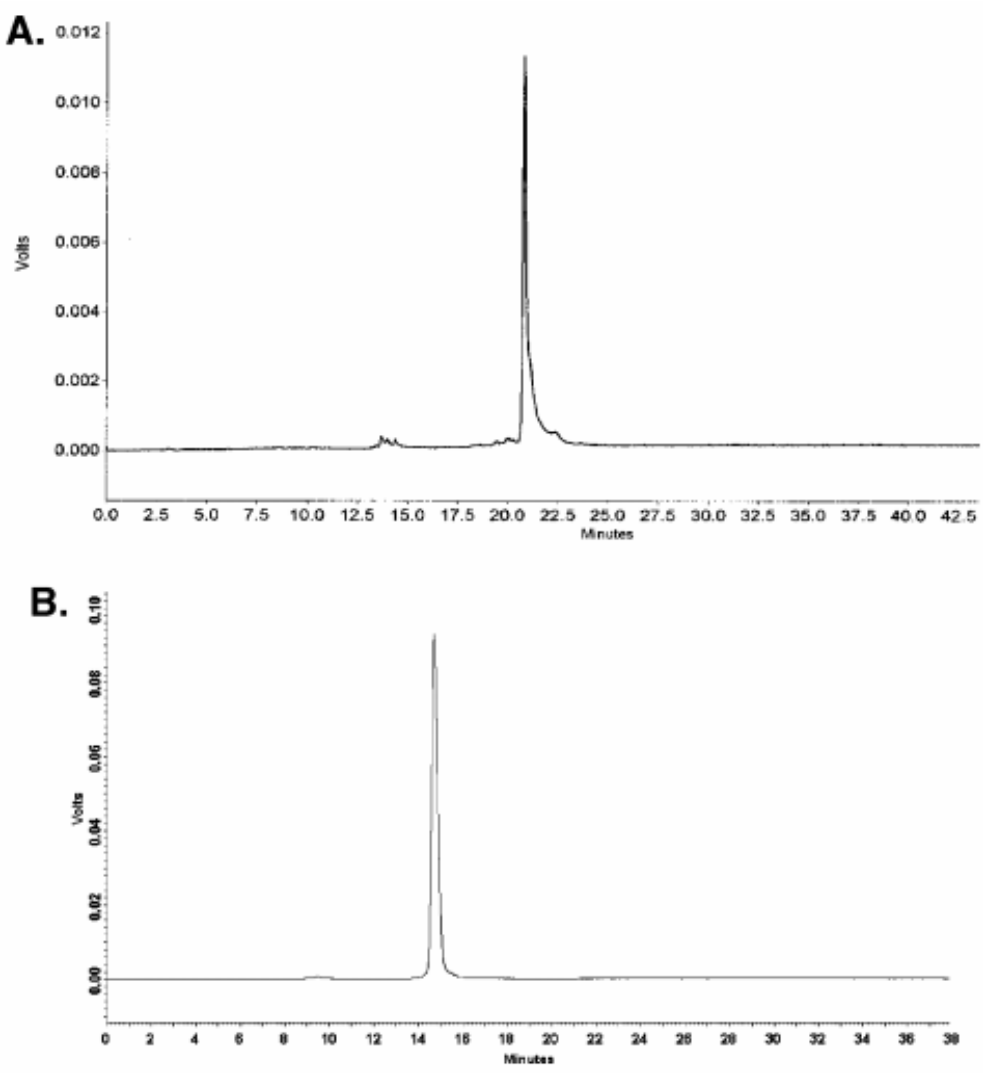


Figure 3.1: Reversed-phase HPLC analysis and purification of the Se-DNAs. (A) the HPLC analysis profile of the crude DMTr-on Se-DNA (5'-DMTr-GTSeGTACAC-3') after cleavage from the solid support and deprotection. The HPLC analysis was performed on a Zorbax SB-C18 column (4.6 x 250 mm) with a linear gradient from buffer A (20 mM triethylammonium acetate, pH 7.1) to 70% buffer B (50% acetonitrile, 20 mM triethyl ammonium acetate, pH 7.1) in 20 minutes; the Se-DNA retention time was 20.6 min. (B) the HPLC purification profile of the DMTr-off Se-DNA (5'-GTSeGTACAC-3'). The sample was eluted on a Zorbax SB-C18 column (21 x 250 mm; flow rate: 6 mL/min) with the same gradient and buffers; its retention time is 14.8 min.

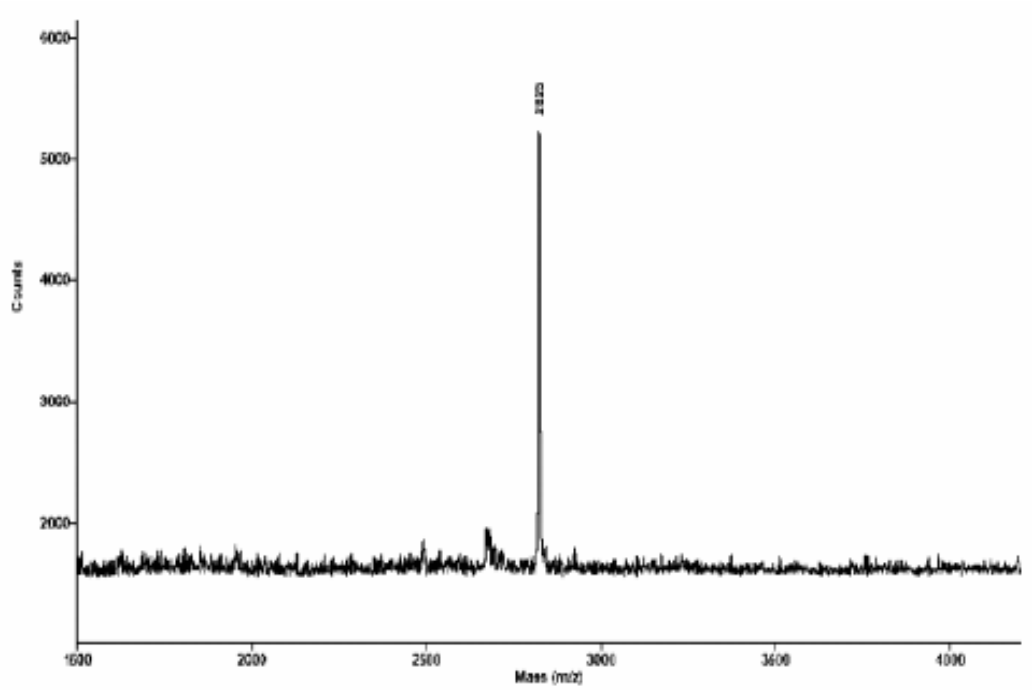


Figure 3.2: MALDI-TOF MS analysis of the DMTr-off Se-DNA (ATGGTSeGCTC). Molecular formula: $C_{89}H_{114}N_{32}O_{54}P_8Se$; FW: 2822.8; found $[M+H]^+$: 2823; calculated [M+H]+: 2823.8.

Table 3.1: MALDI-TOF analysis of 2'-SeMe containing DNA and RNA oligonucleotides.

Se-Sequences	Formula	MS (Calcd.) $[M+H]^+$
DNA-1: $GT_{Se}GTACAC$	$C_{79}H_{101}N_{30}O_{46}P_7Se$	2503 (2503.5)
DNA-2: $ATGGT_{Se}GCTC$	$C_{89}H_{114}N_{32}O_{54}P_8Se$	2823 (2823.8)
DNA-3: $GGAT_{Se}GGGCG$	$C_{91}H_{112}N_{42}O_{51}P_8Se$	2938 (2937.9)
DNA-4: $GCGT_{Se}ATACGC$	$C_{98}H_{125}N_{38}O_{58}P_9Se$	3122 (3122.0)
RNA-1: $GGT_{Se}AUUGCGGUACC$	$C_{135}H_{169}N_{52}O_{97}P_{13}Se$	4555(4554.7)
RNA-2: $UGAGCT_{Se}UCGGCUC$	$C_{124}H_{157}N_{45}O_{91}P_{12}Se$	4186(4185.5)
RNA-3: $CUGUGT_{Se}UCGAUCCACAG$	$C_{162}H_{204}N_{60}O_{118}P_{16}Se$	5455 (5455.2)

3.2.3 Thermal denaturalization study

To compare the thermostability of the Se-oligonucleotides with their native ones, the UV melting experiment was carried out. The samples (2 mM DNA or RNA duplexes or hairpin loop), dissolved in the buffer of 50 mM NaCl, 10 mM NaH_2PO_4-Na_2HPO_4 (pH 6.5), 0.1 mM EDTA, and 10 mM $MgCl_2$, were heated to 85 °C and allowed to cool to room temperature slowly. The oligonucleotide samples were then stored at 5 °C for several hours before the UV melting measurements were carried out. The typical curves and all the melting temperatures are presented in Figure 3.3 and Table 3.2 respectively.

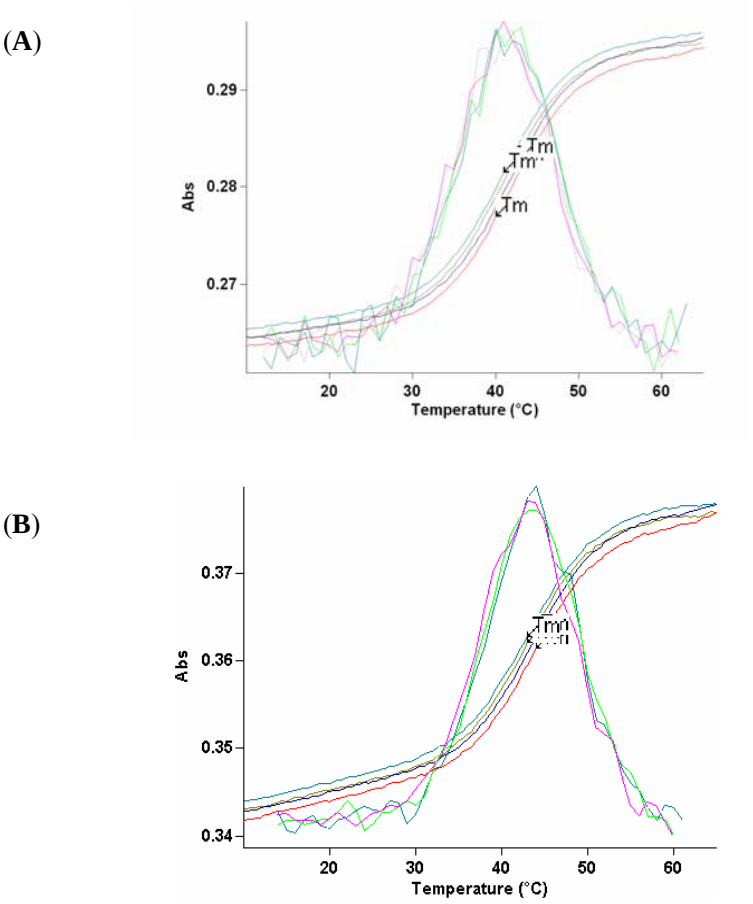


Figure 3.3: Melting curves of native and Se-9mer DNA duplex. (A) Native DNA duplex 5'-ATGGTGCTC-3' & 3'-TACCACGAG-5' (Tm= 42.6 ± 1.4 ºC). (B) Se-DNA Duplex 5'-ATGGT$_{Se}$GCTC-3' & 3'-TACCACGAG-5' (Tm= 42.9 ± 1.1 ºC)

Table 3.2: UV Melting Temperatures of the 2'-Se-RNAs and DNAs.

entry	DNA/RNA pairs	melting temp (°C, native)
	RNA	
a	5′-GGT$_{Se}$AUUGCGGUACC-3′ 3′-CCA--UAACGCCAUGG-5′	53.5 (54.2)
b	5′-UGAGCT$_{Se}$UCGGCUC-3′ 3′-ACUCGA--AGCCGAG-5′	47.2 (47.8)
c	5′-CUGUGT$_{Se}$UCGAUCCACAG-3′ (the tRNA TΨC hairpin loop)	51.7 (52.4)
	DNA	
d	5′-ATGGT$_{Se}$GCTC-3′ 3′-TACCA--CGAG-5′	42.9 (42.6)
e	5′-GGAT$_{Se}$GGGCG-3′ 3′-CCTA--CCCGC-5′	40.1 (40.2)

From these data, it could be concluded that there were no significant differences between the melting temperatures of the 2-Se-oligonucleotide duplexes and the native ones, which indicates a very minor changes of DNA duplex stability comparing the native counterparts, in terms of both DNA and RNA cases. In addition, from the fact that the Se-modified and native tRNA TΨC loops have almost the same melting temperatures, we could draw preliminary conclusion that this selenium derivatization could be potentially applied as a powerful tool to investigate the mechanism of TΨC loops containing tRNAs, since it's still very hard due to the lack of structure information. As the UV melting study cannot provide detailed information on the structure change, NMR and crystallography will be needed for further structure characterizations.

3.2.4 CD-measurement of 2'-SeMe-T modified G-quardruplex

To further explore the potential application of this 2'-Se-thymidine building block, collaborating with Dr. Xiang Zhou's group in Wuhan University, China, we modified the different bridge thymidine residues in human telomeric G-quardruplex sequence 5'-$(TTAGGG)_4$-3' with this functionality and carried out the crystallography studies of these sequences in presence of porphyrin type small compound which could induce the formation of G-quartet structure. (101) Figure 3.4 shows the typical MALDI-TOF spectrum of 2'-SeMe modified 24mer, and figure 3.5 shows the preliminary CD results without and with porphyrin compound, which indicate the similar property as native ones. The further crystallization part of this project is still on going.

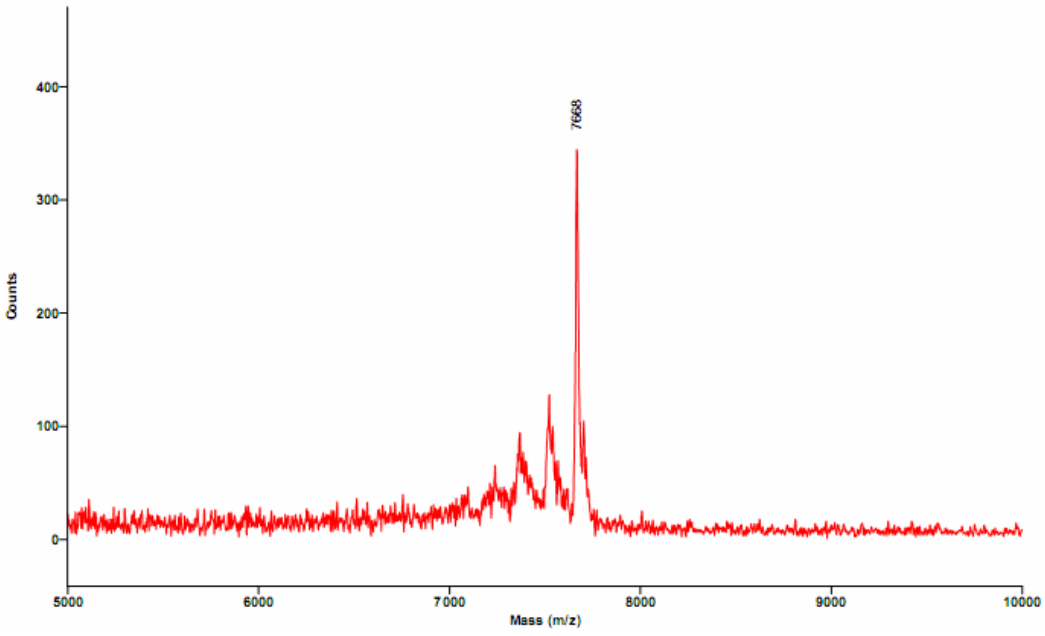


Figure 3.4: MALDI-TOF mass-spectrum of selenium modified huamn telomeric G-quardruplex sequence. 5'-TTAGGGT(2'-SeMe)TAGGGTTAGGGTTAGGG-3', $C_{241}H_{299}N_{96}O_{146}P_{23}Se$, Calculated mass: 7667.9, Observed: 7668.

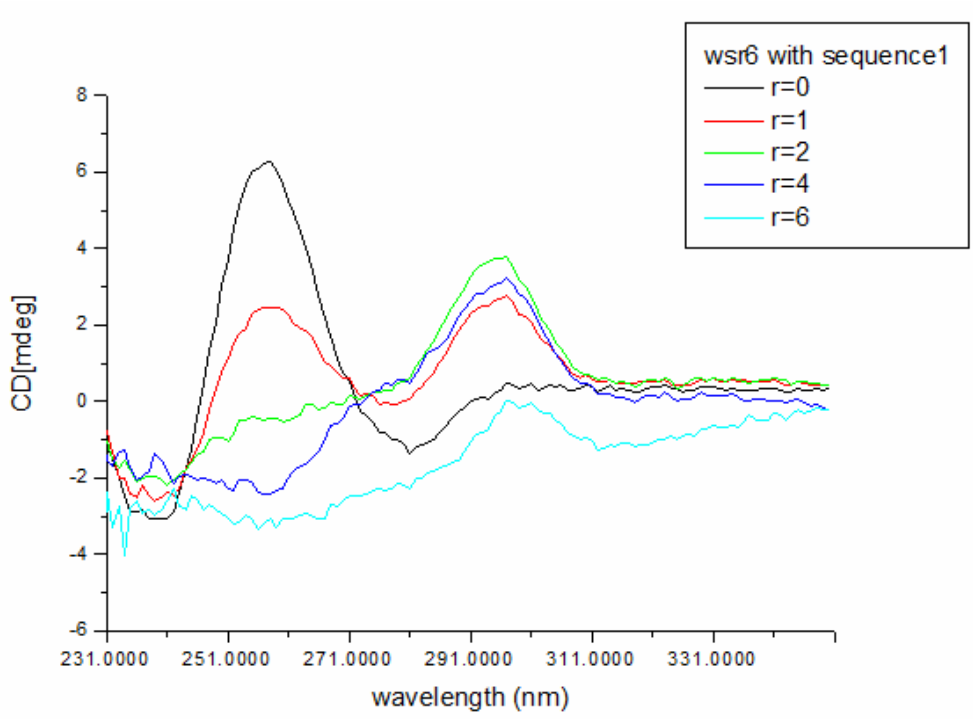


Figure 3.5: CD-spectrum of Se-Huamn telomeric G-quardruplex sequence with different equivalence of porphyrin compound. The peaks at 295 and 295 nm indicate the formation of anti-parallel structure G-quartet.

3.2.5 Crystallization and structure studies

With pure oligonucleotides on hand, we carried out the crystallography studies for this modification work. As a result, we found that the Se-DNA (G$\underline{T}_{Se}$GTACAC, self-complementary) crystallized in 2-3 weeks in the native buffer, where the native DNA (GTGTACAC) crystallized over 2 months. However, when the Nucleic Acid Mini Screen kit (Hampton Research, with 24 diversified crystallization buffers) was screened, to our pleasant surprise, we found that the Se-DNA crystallized overnight in all the 24 buffer conditions overnight. In contrast, the native DNA did not crystallize at all over many weeks in these Hampton buffers. It was also reported that the two RNAs (12-mer and 16-mer) containing the 2-Se-guanosines crystallized, respectively, in 17 and 33 buffers,

whereas their corresponding native RNAs crystallized in 15 and 24 buffers, individually. Though the crystallization conditions of these Se-RNAs expanded slightly, it appears that these two Se-RNAs, which crystallized mostly as thin needle crystals, behaved similarly to many Se-Met-derivatized proteins that required the fine tuning of their native crystallization conditions. By contrast, the majority of these Se-DNA crystals grown under the new buffers was larger than the native DNA crystals grown under the native buffer conditions, and these Se-DNA crystals also diffracted well. Our results reveal that the 2-Se derivatization of thymidine facilitated the crystal packing and growth and appeared to help expand the crystallization conditions.

Finally, the structure of the Se-DNA crystal, grown under the buffer No.7 (the detailed composition of the screen buffers is listed in the appendix 1) was determined at 1.40 Å resolution via the Se SAD phasing and refinement. The data collection parameters, phasing and refinement statistics are listed in table 3.3. It's clearly observed that this oligonucleotide displays a 2'-exo sugar pucker, which is consistant with the A form DNA and RNA molecules. And the selenium function locates in the minor groove of this duplex. When the 2'-Se-T DNA was superimposed with naitve counterpart (they have the same space group), it was found that no virtually structure perturbations was observed (single strand, double strand and local residue comparison in Figure 3.6A, C and D), which indicates that this Se derivatization does not alter the native structure significantly. In addition, figure 3.6B clearly showed the electron density map of selenium functionality (Se were represented by red balls), this 100% occupancy also showed the stability of this SeMe group in X-ray exposure. Furthermore, comparing the resolution of native DNA (2.0 Å), this higher structure resolution allows observation of ordered 43 water molecules and other structure features.

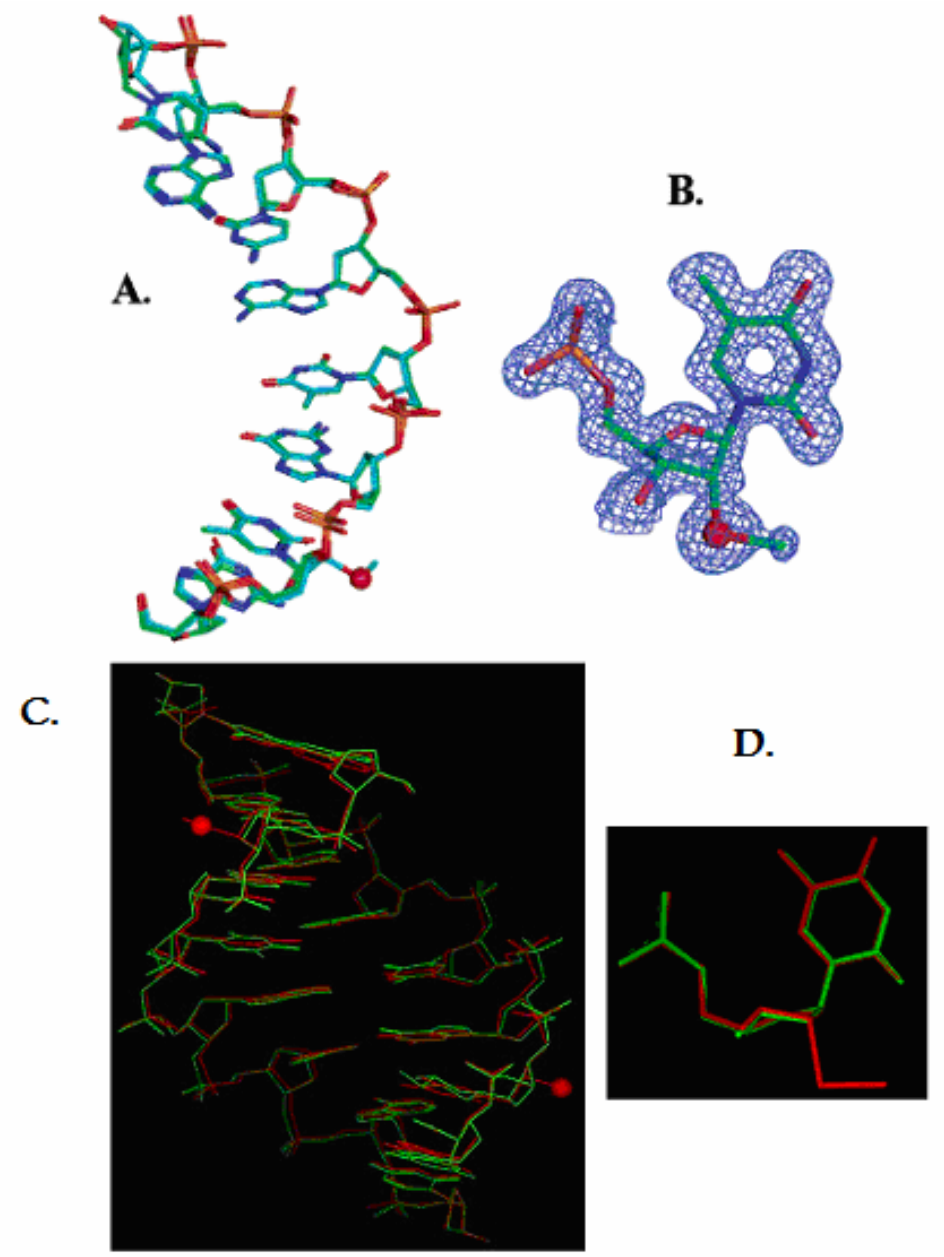


Figure 3.6: Various functionalities modified DNA crystal structures. (A) Superimposed structure comparison of the native DNA (GTGTACAC, 1DNS in green) and the Se-derivatized DNA (GTSeGTACAC, 2HC7 in cyan). (B) The structure model and the electron density map of 2'-Se-ribothymidine residue in the structure. (C) Superimposed duplex structure of native and selenium DNA. (D) Superimposed local residues.

Table 3.3: X-ray data collection, phasing and refinement statistics for 2HC7. Two data sets were collected: the data collected at selenium K-edge (0.9793 Å) was used for SAD phasing, and the data collected at 1.100 Å was used for the refinement.

Data Collection	λ= 0.9793 Å	λ= 1.100 Å
Resolution range, Å (last shell)	40.0-1.50 (1.55-1.50)	40.0 - 1.40 (1.45 - 1.40)
Unique reflections	3786 (349)	4577 (492)
Completeness, %	99.7 (99.7)	98.8 (92.3)
R_{merge} ,%	6.4 (29.7)	3.6 (34.3)
$\langle I/\sigma(I)\rangle$	13.5 (4.3)	13.1 (2.1)
Redundancy	13.3 (13.6)	12.7 (10.0)
R-Cullis	0.352	
Phasing power	4.053	
Figure of merit	0.421	

Refinement	
Resolution range, Å (last shell)	18.72 - 1.40 (1.49 - 1.40)
Number of reflections	4252 (492)
R_{work} , %	17.0 (22.1)
R_{free} , %	18.7 (24.4)
Number of atoms	
Nucleic Acid (single)	162
Heavy atom	1 (Se)
Water	43
R.m.s. deviations	
Bond length, Å	0.015
Bond angle,	1.9
Average B-factors, $Å^2$	
All atoms	18.5
Wilson plot	16.1
Overall anisotropic B-values	
B11/B22/B33	-0.86/-0.86/1.71
Bulk solvent correction	
Solvent density, $e/Å^3$	0.34
B-factors, $Å^2$	45.1
Coordinates error (c.-v.), 5Å	
Esd. from Luzzatt plot, Å	0.15
Esd. from SIGMAA, Å	0.08

3.3 Conclusion of this chapter

In conclusion, we have successfully synthesized the 2'-SeMe-thymidine phosphoramidite as an analogue of both thymidine and ribothymidine phosphoramidites for the solid phase oligonucleotide synthesis. Subsequently, we have incorporated it into DNAs and RNAs with over 99% coupling yield. Our thermal denaturing study of the Se-modified DNAs and RNAs also reveals that this Se derivatization did not significantly change the stability of the structures of the DNA and RNA duplexes and the tRNA TΨC loop. Furthermore, our X-ray crystal structure study with a Se-modified DNA octamer (GTSeGTACAC) shows that the Se-derivatized DNA structure (1.40 Å resolution) is virtually identical to the native structure (2.0 Å resolution) with the same space group. More importantly, we observed that this Se modification dramatically expanded the crystallization conditions of the Se-DNA and largely facilitated crystal growth while retaining the high diffraction quality. In addition to DNA investigations, this 2'-SeMe-thymidine can also serve as a ribothymidine analogue for the structure and function studies of tRNAs and rRNAs. As the Se derivatization can facilitate the phase determination and crystallization, it will significantly impact the structure and function studies of nucleic acids and their protein complexes using X-ray crystallography. In addition, we are in the process of further investigating the mechanism of the crystal growth facilitated by this Se derivatization.

4 ADVANTAGES OF SELENIUM OVER BROMINE STRATEGY: STRUCTURE COMPARISON OF THE 2'-SE AND 5-BR DERIVATIZED NUCLEIC ACIDS

(This work has been published on Nucleic Acids Research, 2007, 35 (2): 477-485, Jiansheng Jiang, Nicolas Carrasco and Zhen Huang are listed as co-authors.)

4.1 Introduction

With the increasing requirement to obtain detailed insight into the structure-function relationships of DNA-drug complexes, functional RNAs (such as catalytic RNAs and snoRNAs), and RNA-protein and DNA-protein interactions, X-ray crystallography application in this field has been much popular and important, accordingly, the improvement of this technology itself especially solving the phase problem in nucleic acids, which actually has been the limiting factor that largely slowed down the structural determination of new structures, has become more and more important and urgent. As stated in the introduction part, the conventional approaches for DNA and RNA derivatization for phase information, such as heavy-atom soaking and co-crystallization, have proved to be much more difficult for nucleic acids than for proteins, probably because nucleic acids often lack specific binding sites for metal ions. As a result, the halogen such as bromine (K edge, 0.920 Å) derivatization is used as the conventional method to derivatize DNAs *via* 5-bromo-deoxyuridine (thymidine analog) and RNAs *via* 5-bromo-uridine in Multiwavelength Anomalous Dispersion (MAD) phasing. The derivatives containing the bromine derivatization are relatively stable because of formation of vinyl halides. However, it is difficult to introduce bromine to other positions because of the chemical nature of the halogen. For instance, as bromide is a good leaving group, bromine placed on the ribose 2'-position can be displaced by the

pyrimidine exo-2-oxygen, the purine 3-nitrogen, or the near-by nucleotide phosphate groups. Though the 8-bromo-purine nucleosides can been easily prepared, (102) the 8-Br-substituted purine bases tend to adopt a syn-conformation, (103) which greatly perturbs local structure. Therefore, in the conventional halogen derivatization, the halogens are primarily limited to the 5-position of uracil (sometimes cytosine), which places the halogen in the major groove. As the major groove of A-form dsDNA or dsRNA is narrow and deep, the halogen derivatization may perturb these structures. The halogen derivatization can also cause base-stacking disruption, hydration pattern change, and other structural perturbations. Probably that's also the reason that the halogen derivatized nucleic acid usually are hard to be crystallized. In addition, halogen derivatives, such as the bromine derivatives, are light sensitive, and long-time exposure to X-ray or UV sources may cause decomposition. (104) These factors could have caused a lot of trouble to solve the nucleic acid structures in many cases.

The revolutionized success of selenomethionine derivatization method in protein X-ray crystallography was also applied in nucleic acid. Indirect derivatization method of RNA using selenomethionine-labeled U1A for phase and structure determination was also successfully demonstrated though it is labor-intensive. More important, the direct derivatization of DNAs and RNAs with selenium has been demonstrated by our preliminary results. So far, this novel derivatization methodology has been utilized in X-ray crystal structure studies of RNA and DNA molecules by several laboratories, including structure analysis and determination of ribozymes and riboswitches. As the synthesis and purification of SeNAs are much easier than those of proteins, nucleic acid-protein complexes may also be derivatized by SeNAs instead of their protein

counterparts. Therefore, this atom-specific derivatization appears to have great potential in structural and functional studies of nucleic acids and their protein complexes.

It has been stated in the previous chapter that the 2'-Se is proved to be the most completed and stable type of Se-derivatization so far, in order to further investigate this SeNA strategy, we carried out the systematic structural comparison between the Se and the conventional Br derivation using the model DNA (5'-GTGTACAC-3'). It is worth to mention that the Se derivatization is placed on the sugar ring and located in the minor groove of the duplex, while the Br derivatization occurs to the base and is located in the major groove. In addition, the crystal growth facilitation by this 2'-Se derivatization was further demonstrated in this work.

4.2 Results and discussion

4.2.1 Preparation of DNA crystals containing Se and Br

For the simplicity and cost of synthesis, 2'-SeMe-dU instead of 2'-SeMe-dT is used to replace T2 in the self-complementary DNA (GTGTACAC, Figure 4.1A). In order to directly compare with the conventional Br derivatization strategy, we also synthesized this DNA model with the Br derivatization at the T4 position (Figure 4.1B), in the presence and absence of the Se derivatization at the T2 position, the building block of 5-Br-2'-deoxyuridine phosphoramidite is commercial available from ChemGens, Inc. The native DNA (5'-GTGTACAC-3'), the Se-DNA (5'-GdUSeGTACAC-3'), the Br-DNA (5'-GTGdUBrACAC-3'), and the Se-Br-DNA (5'-GdUSeGdUBrAC-A-C-3') were synthesized on solid phase DNA synthesizer following the standard procedures and purified by HPLC. The purified oligonucleotides were analyzed by MS and HPLC to confirm their accuracy and purity.

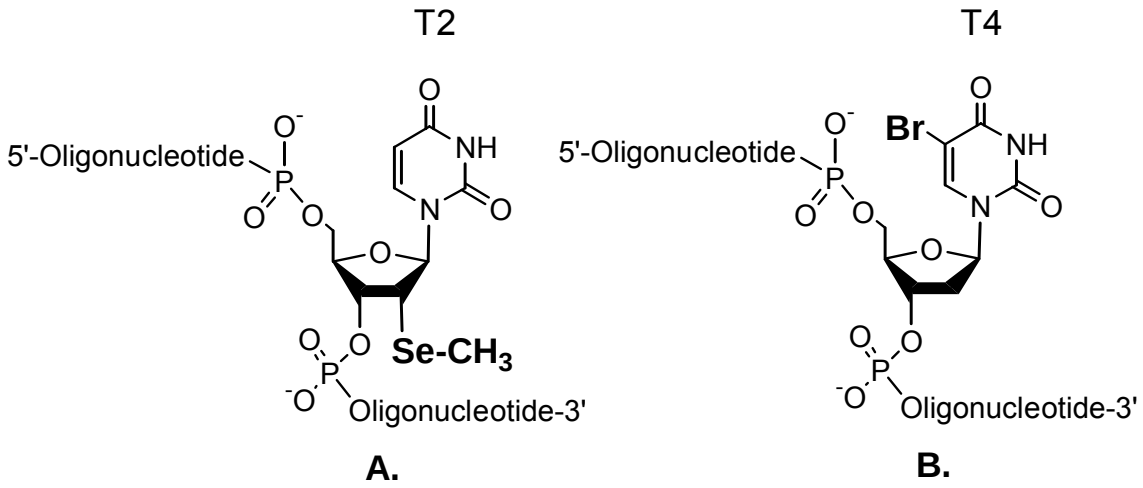


Figure 4.1: Structures of the 2'-Se modification and 5-Br modification of DNA.

4.2.2 Crystal parameters comparison of native, Se and Br derivatized DNAs

Though the Se-DNA crystallized in the native buffer, (105) crystallization conditions were still screened using the Nucleic Acid Mini Screen kit (Hampton Research) (Appendix). l µL of the oligonucleotide solutions (1 mM) was typically used in each of these screens with hanging drop at 25°C. In the cases of the Se-DNA and the Se-Br-DNA, their crystals with diffraction quality were identified in many identical buffer conditions (typical crystal pictures are showed in figure 4.2). Crystals of the Se-DNA (5'-GdUSeGTA-CAC-3'), grown in buffer #7 of the kit (10% v/v MPD, 40 mM sodium cacodylate pH 6.0, 12 mM spermine tetra-HCl, 80 mM potassium chloride and 20 mM magnessium chloride), were identified to give the highest diffraction resolution, while crystals of the Se-Br-DNA (5'-GdUSeG-dUBrACAC-3'), grown in buffer #9 (10% v/v MPD, 40 mM sodium cacodylate pH 6.0, 12 mM spermine tetra-HCl, 80 mM sodium chloride, 12 mM potassium chloride, and 20 mM magnessium chloride), gave highest resolution. The Br-DNA crystal was grown under the native conditions (33,34, 40% v/v MPD, 4 mM spermine and 8 mM magnessium chloride). The best Br-DNA crystal was

obtained after a week under the lower temperature (15°C). The size of these crystals generating good diffraction is approximately 0.1x0.1x0.1 mm. The Se-DNA and Br-DNA data were collected at beam line X12C, and the Se-Br-DNA data sets were collected at X25, X29 and X12 in NSLS of Brookhaven National Laboratory. A number of crystals were scanned to find the one with strong anomalous scattering at the K-edge absorption of selenium and/or bromine. The distance of the detector to the crystals was set to 90 mm. The chosen wavelengths for selenium and bromine MAD are listed in Table 4.1 & 4.2. The crystals were exposed for 10 or 15 seconds per image after one degree rotation, and a total of 180 images were taken for each data set. The additional reference data sets were collected at 1.10 Å wavelength and were used for the final structure refinement. All data were processed using HKL2000 and DENZO/SCALEPACK. The positions of the selenium and bromine atoms were identified by the heavy atom search scripts provided by CNS. Following the refinement of the selenium and bromine positions, the experimental phases were calculated and extended using the MAD or SAD procedures in CNS. The initial phased maps were then improved by the solvent flattening and density modification procedure. The model of single strand of DNA was manually built into the electron density map using O, and then the refinement was carried out. The refinement protocol includes simulated annealing, positional refinement, restrained B-factor refinement, and bulk solvent correction. The stereo-chemical topology and geometrical restrain parameters of DNA/RNA have been applied. The topologies and parameters for modified dU with selenium (UMS) and dU with bromine (BRU) were constructed and applied. After several cycles of refinement and the model rebuilding, a number of highly ordered waters were added. Final, the occupancies of selenium and bromine were refined. Cross-validation with a 5-10% test

set was monitored during the refinement. The σA-weighted maps of the (2m|Fo| - D|Fc|) and the difference (m|Fo| - D|Fc|) density maps were computed and used throughout the model building.

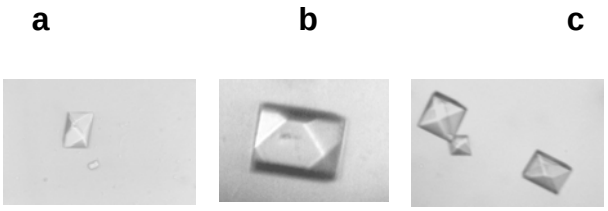


Figure 4.2: Photos of crystals of the native and derivatized octamers. (a) Native-Oct.; (b) Se-Oct.; (c) Se/Br-Oct. Sizes of the crystals ranged from 0.1x0.1 to 0.4x0.4 mm.

4.2.3 Crystallization facilitation by Se-derivatization

We performed the crystallization screening to study the crystallizability of these modified oligonucleotides, as a result, we found that the SeMe functionality containing DNA could formed quality crystals in all of the buffers of the Hampton kit overnight (approximate size: 0.1x0.1x0.1 mm), and the Se-Br-DNA also formed good quality of crystals overnight in the same buffers, though the crystal sizes were smaller (approximate 0.05x0.05x0.05 mm). However, both the Br-DNA and the native DNA did not grow crystals in any buffers of the kit (24 buffers) over several months. While the native DNA and the Br-DNA crystallize in the native conditions in a few weeks. (106) It is also worth to mention that these two Se-derivatized DNAs can also crystallize in the native conditions.

Table 4.1: Summary of data collection and phasing statistics for Se-DNA (1Z7I).

	Selenium K-edge		
	peak	inflection	remote
Wavelength, Å	0.9790	0.9794	0.9400
Resolution range, Å (last shell)	50.0-1.4 (1.49-1.4)	50.0-1.4 (1.49-1.4)	50.0-1.4 (1.49-1.4)
Unique reflections	4478 (441)	4425 (420)	4396 (402)
Completeness, %	97.6 (99.1)	96.3 (94.0)	95.2 (90.0)
R_{merge} ,%	6.9 (40.2)	7.2 (44.0)	7.3 (45.0)
$\langle I/\sigma(I)\rangle$	14.0 (3.8)	13.9 (3.9)	13.2 (3.5)
Redundancy	13.0 (10.5)	11.5 (9.7)	10.5 (9.7)
R-Cullis (Friedel)	0.513 (0.330)	0.371 (0.279)	- (0.428)
Phasing power λ_3/λ_1 (Friedel)	2.026 (4.345)	3.155 (5.021)	- (3.405)
Figure of merit (Friedel)	0.426 (0.508)	0.502 (0.539)	- (0.380)
Overall figure of merit		0.817	

$R_{merge}=\Sigma|I-\langle I\rangle|/\Sigma I$

Table 4.2: Data collection and phasing statistics for Se-Br-DNA (2DLJ).

	Selenium K-edge			Bromine K-edge		
	Peak	inflection	Remote	peak	inflection	remote
Wavelength, Å	0.9790	0.9794	0.9400	0.9196	0.9199	0.9400
Resolution range, Å (last shell)	50.0-1.80 (1.86-1.80)	50.0-1.80 (1.86-1.80)	50.0-1.80 (1.86-1.80)	50.0-1.80 (1.86-1.80)	50.0-1.80 (1.86-1.80)	50.0-1.80 (1.86-1.80)
Unique reflections	2283 (222)	2284 (219)	2288 (220)	2290 (220)	2292 (220)	2288 (220)
Completeness, %	99.8 (100.0)	99.8 (100.0)	99.9(100.0)	99.8 (100.0)	99.9 (100.0)	99.9(100.0)
R_{merge} , %	7.8 (36.7)	7.2 (31.8)	6.7 (32.9)	7.9 (39.5)	7.9 (41.0)	7.8 (40.4)
$I/\sigma(I)$	11.4 (3.1)	12.2 (3.7)	12.1 (3.7)	11.4 (3.1)	11.1 (2.6)	11.1 (2.7)
Redundancy	12.8 (12.8)	12.9 (13.0)	12.9 (13.0)	12.9 (13.1)	12.9 (13.0)	12.9 (13.0)
R-Cullis (Friedel)	0.709 (0.455)	0.469 (0.359)	- (0.510)	0.686 (0.417)	0.587 (0.407)	- (0.482)
Phasing power (Friedel)	1.221 (3.223)	2.365 (4.043)	- (2.826)	1.205 (4.002)	1.646 (4.053)	- (3.395)
Figure of merit (Friedel)	0.321 (0.429)	0.425 (0.477)	- (0.311)	0.324 (0.465)	0.385 (0.478)	- (0.356)
Overall figure of merit		0.731			0.761	

Table 4.3: Data collection and refinement statistics in Se-DNA, Se-Br-DNA and Br-DNA structures.

Structure (PDB ID)	Se-DNA (1Z7I)	Se-Br-DNA (2DLJ)	Se-Br-DNA (2GPX)	Br-DNA (2H05)
Data collection				
Space Group	$P4_32_12$	$P4_32_12$	$P4_32_12$	$P4_32_12$
Cell dimensions: *a*=*b*, *c* (Å),	42.076, 23.935	42.369, 23.744	42.252, 24.133	42.048, 24.270
Resolution range, Å (last shell)	50.00 - 1.28 (1.36 - 1.28)	50.00 - 1.50 (1.55 - 1.50)	50.00 – 1.60 (1.66 – 1.60)	50.00 – 1.80 (1.86-1.80)
Unique reflections	6325 (552)	3777 (354)	3161 (297)	2248 (203)
Completeness, %	95.3 (70.0)	99.7 (99.2)	99.7 (99.3)	99.4 (95.8)
R_{merge} , %	8.0 (32.8)	5.1 (27.8)	8.4 (49.9)	4.5 (32.3)
$I/\sigma(I)$	14.8 (4.3)	14.4 (4.9)	13.1 (4.2)	12.1 (2.7)
Redundancy	12.4 (10.7)	12.3 (9.1)	12.4 (9.0)	12.1 (7.1)
Refinement				
Resolution range, Å (last shell)	20.81 - 1.28 (1.36 - 1.28)	18.16 - 1.50 (1.59 - 1.50)	18.90 – 1.60 (1.70 – 1.60)	18.81 – 1.80 (1.91-1.80)
R_{work} , %	17.7 (23.7)	18.3 (18.9)	18.7 (21.3)	21.3 (27.3)
R_{free} , %	20.0 (23.1)	22.8 (22.7)	23.8 (24.4)	24.8 (41.3)
Number of reflections	5275 (552)	3597 (473)	2909 (289)	2036 (247)
Number of atoms				
Nucleic Acid (single)	161	160	160	160
Heavy Atoms and Ion	1 Se	1 Se, 1 Br	1 Se, 1 Br, 1 Ba^{2+}	1 Br
Water	44	43	31	33
R.m.s. deviations				
Bond length, Å	0.012	0.011	0.010	0.009
Bond angle,	1.8	1.6	1.8	1.9
Average B-factors, $Å^2$				
All atoms	15.0	18.8	25.9	32.5
Wilson plot	14.9	18.2	23.6	29.8
Overall anisotropic B-values				
B11/B22/B33	0.70/0.70/-1.41	-0.29/-0.29/0.58	0.31/0.31/-0.62	2.20/2.20/-4.41
Bulk solvent correction				
Solvent density, $e/Å^3$	0.30	0.32	0.34	0.32
B-factors, $Å^2$	22.2	34.5	40.05	41.04
Coordinates error (c.-v.), 5Å				
Esd. from Luzzatt plot, Å	0.15	0.19	0.22	0.23
Esd. from SIGMAA, Å	0.10	0.05	0.12	0.20

To further demonstrate this interesting observation, we have synthesized and purified four other native DNAs (41,42, PDB ID: 395D & 1QPH) and their selenium derivatives (Se-DNAs): 5'-GdU$_{Se}$ACGTAC-3', 5'-GTAC-GCGdU$_{Se}$AC-3', 5'-CCGdU$_{Se}$ACGTACGG-3', and 5'-GA-CCACGTGGdU$_{Se}$C-3'. The Se derivatization was incorporated into different regions of these DNA molecules, including the 5', middle, and 3' regions. These self-complementary DNA molecules were screened using the Hampton kit. High quality single crystals (average size: 0.1x0.1x0.1 mm) of these Se-DNA sequences were formed in three days under many different buffer conditions: 5'-GdU$_{Se}$ACGTAC-3' in buffer #9, #12, #13, #22 and #23, 5'-GTACGCGdU$_{Se}$AC-3' in buffer #8, #10, #12, #17, #19 and #24, 5'-CCGdU$_{Se}$ACGTACGG-3' in buffer #11, #12 and #21, and 5'-GACCACGTGGdU$_{Se}$C-3' in buffer #18, #22 and #23. The corresponding native DNA sequences, however, did not crystallize over two weeks in any of these 24 buffers. It seems that it is quite challenging to crystallize the native molecules. In summary, these observations suggest that these Se-derivatized DNAs can crystallize in broader buffer conditions and have better crystallizability than these native and Br-derivatized DNAs.

Our observations indicate that the 2'-selenium derivatization may facilitate the crystal growth, which is probably assisted by the better packing of the Se-derivatized DNA. As the bulky 2'-Se modification may be able to lock the sugar pucker into the 2'-exo conformation, the rigidity of the Se-derivatized DNAs may facilitate molecular packing. Furthermore, the structure studies of the Se, Br modified DNAs, grown under different buffer conditions, indicated that the different buffers may not change the crystal packing and the space groups. The more results will be discussed in chapter 5.

4.2.4 MAD & SAD phasing

As one Se atom should enable the phase determination for DNAs and RNAs with up to 30 nucleotides, one selenium or one bromine atom in the octamer DNA is sufficient for MAD or SAD phasing. As shown in Table 4.1, we have used the Se-DNA data only up to 1.4 Å for MAD and SAD phasing. The overall figures of merit (FOM) of the initial phases for both Se-DNA MAD and SAD data sets were 0.816 and 0.442, respectively. This produced an excellent electron density map showing a single strand of DNA after solvent flattening and density modification. The reference data were used as the amplitudes of the structure factors and combined with the phase probabilities from MAD/SAD, and the phases were extended from 1.4 Å to 1.28 Å for the refinement of Se-DNA. The MAD data sets for the Se-Br-DNA were available to up to 1.8 Å resolution. The data at the selenium K-edge and the bromine K-edge were processed separately. As shown in Table 5, the overall FOMs of the initial phases were 0.731 and 0.761 respectively from the selenium and bromine K-edge data sets. Both produced excellent electron density maps clearly showing the DNA strands. The experienced phases were extended to 1.5 Å for the refinement of the Se-Br-DNA structures. Likewise, the Br-DNA structure was determined.

These Se and/or Br modified DNA structures have the same tetragonal space group P43212, as the native structure. These duplex models are generated by applying the symmetry operator (y, x, -z) of P43212. Figure 4.3 shows the electron density maps with the duplex models of the Se and Br derivatized DNA structures. The resulting structures of the 2'-Se-DNA (1Z7I), the Se-Br-DNA (2DLJ), and the Br-DNA (2H05) were refined at 1.28, 1.5, and 1.8 Å resolution, respectively. The second Se-Br-DNA structure

(2GPX), containing a barium ion, was refined at 1.6 Å resolution; its electron density map (identical to 2DLJ) is not shown. Figure 4.4 illustrates the details of the electron density maps of the modified dUs in the structures. Figure 4.3A, 4.3B and 4.3C represent the structures of the 2'-Se-dU in the Se-DNA, 5-Br-dU in the Se-Br-DNA and 5-Br-dU in the Br-DNA, respectively.

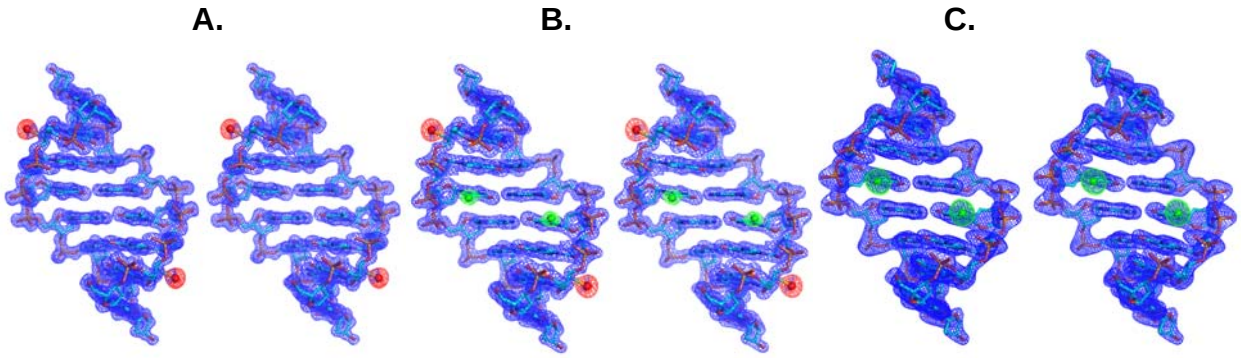


Figure 4.3: Electron density maps and models of the derivatized DNA duplexes (stereo view). Red and green balls represent Se and Br respectively. Contours are at the 1.2σ level. (A). the structure of the Se-DNA ($GdU_{Se}GTACAC$) at 1.28 Å resolution (1Z7I). (B). the structure of the Se-Br-DNA ($GdU_{Se}GdU_{Br}ACAC$) at 1.5 Å resolution (2DLJ). (C). the structure of the Br-DNA ($GTGdU_{Br}ACAC$) at 1.8 Å resolution (2H05).

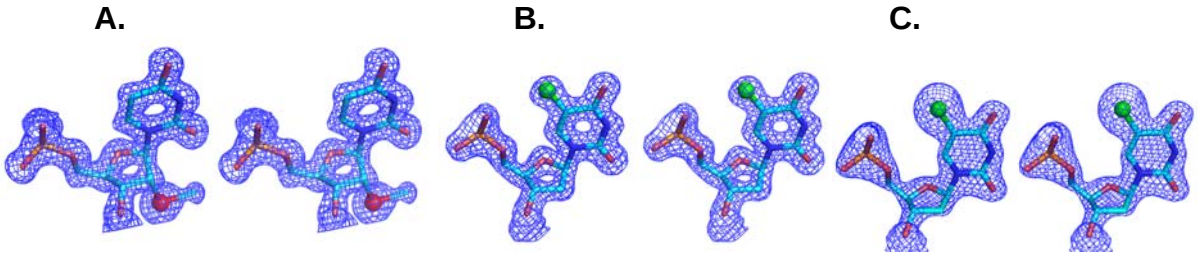


Figure 4.4: Electron density maps and models of the Se and Br modified nucleotides in the derivatized DNAs. (Stereo view) Red and green balls represent Se and Br, respectively. Contours are at the 1.2σ level. (A). the structure of the 2'-Se-dU in

d(GdU_{Se}GTACAC) at 1.28Å resolution (1Z7I). (B). the structure of the 5-Br-dU in d(GdU_{Se}GdU_{Br}ACAC) at 1.50Å resolution (2DLJ). (C). the structure of the 5-Br-dU in d(GTGdU_{Br}ACAC) at 1.80Å resolution (2H05).

4.2.5 Structures of Se and Br derivatized DNAs

As showed in figure 4.5, the Se-DNA structure (1Z7I, in blue, Figure 4.5A) is superimposable over the native A-form DNA structure (1DNS, in pink), which has the same tetragonal space group ($P4_32_12$). The overall r.m.s.d. of the Se-DNA over the native is low (0.328 Å) and mainly contributed by the first and the last bases, and the main structure is considered as the same as the native, although the spermine molecule bound to the native was not observed in the derivatized structure. The superimposed local dU_{Se} structure of the Se-DNA (1Z7I, in blue) over T2 of the native structure (1DNS, in pink) is shown in Figure 4.4B. The derivatized local structure of the Se-DNA is also superimposable over that of the second native (34, 1D78, in gray, Figure 4.5B), which has a different space group ($P6_122$). Therefore, the selenium modification causes no local perturbations when it is compared with both native forms. In addition, superimposing the dU_{Se} in this Se-DNA structure (space group $P4_32_12$, Figure 4.5C) with the dU_{Se} in the reported DNA structure (23, GCGTAdU_{Se}A-CGC, 1MA8 in yellow, space group $P2_12_12_1$) indicates no difference, suggesting that the Se-derivatized local structure is stable and independent from its sequence environment.

Figure 4.5D and 4.5E show the superimposed Br-DNA (2H05, in green) and Se-Br-DNA (2DLJ in cyan) over the native (1DNS, in pink), respectively. We observed the local perturbations in both Br-derivatized DNA structures, thought the global structures

were very close to the native structure with the same crystal form (1DNS, 33). The overall r.m.s.d. values of the Se-Br-DNA and the Br-DNA over the native are 0.713 Å and 0.642 Å, respectively.

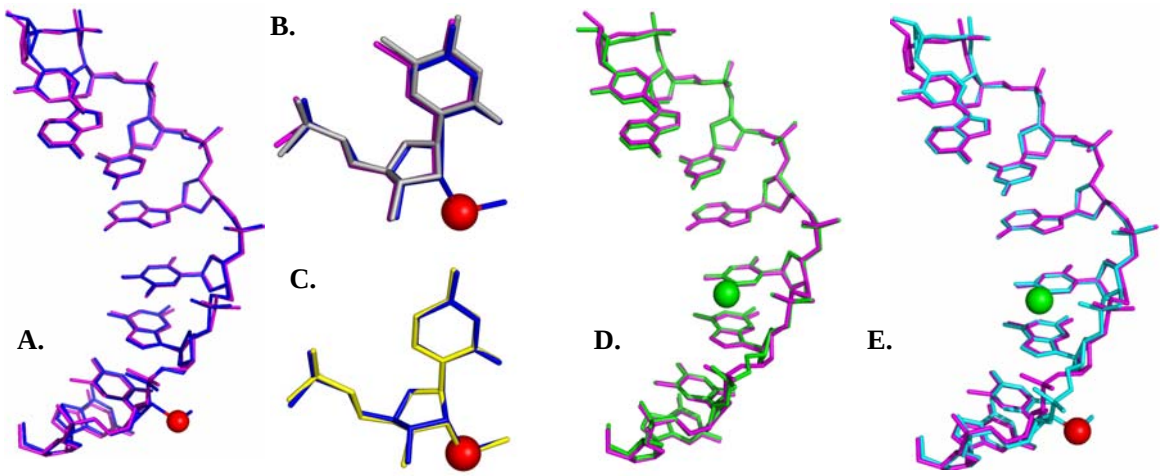


Figure 4.5: Superimposed comparison of the native and the derivatized DNA structures. (red and green balls represent Se and Br, respectivly). (A) the comparison of the Se-DNA (GdU_{Se}GTACAC, 1Z7I in blue) and the native (GTGTACAC, 1DNS in pink). (B) the comparison of the Se-DNA dU_{se} (1Z7I in blue) and the two native T2 (1DNS in pink and 1D78 in gray). (C) the comparison of the dU_{se} of the Se-DNA (1Z7I in blue) and the reported DNA (GCGTAdU_{Se}ACGC, 1MA8 in yellow). (D) the comparison of the Br-DNA (GTGdU_{Br}ACAC, 2H05 in green) and the native (1DNS in pink). (E) the comparison of the Se-Br-DNA ($GdU_{Se}GdU_{Br}$ACAC, 2DLJ in cyan) and the native (1DNS in pink).

In addition, the comparison of the Br-DNA dU_{Br} with the T4 of the native DNA (1DNS) and the Se-DNA indicated that the Br incorporation could alter the hydration pattern (Figure 4.6A and 4.6B) and the water networking (Figure 4.7) at the derivatized site, though the sugar pucker of the dU_{Br} was almost not changed from that of the native (Figure 4.6A). The alterations at the dG3 site of the Br-DNA (Figure 4.6C), including the

dG sugar pucker change and approximately 107 degree rotation about its $C_{4'}$-$C_{5'}$ bond, were also observed. This rotation caused large local perturbations on the dG3, with its $C_{5'}$ and P deviations by 1.30 and 0.63 Å, respectively. The $C_{2'}$ of the Br-DNA T2 is also deviated by 0.60 Å (Figure 4.6D), which also created the flattened sugar pucker as the dG3 (Figure 4.6C). Unsurprisingly, the same perturbations were also observed with the Se-Br-derivatized DNA (2DLJ) while the perturbations were not observed with the Se-derivatized DNA (1Z7I). These perturbations caused by the Br derivatization may explain why some Br-derivatized DNAs have been difficult to crystallize even after successful crystallization of their natives. Our observation may serve as a reminder in the bromine derivatization.

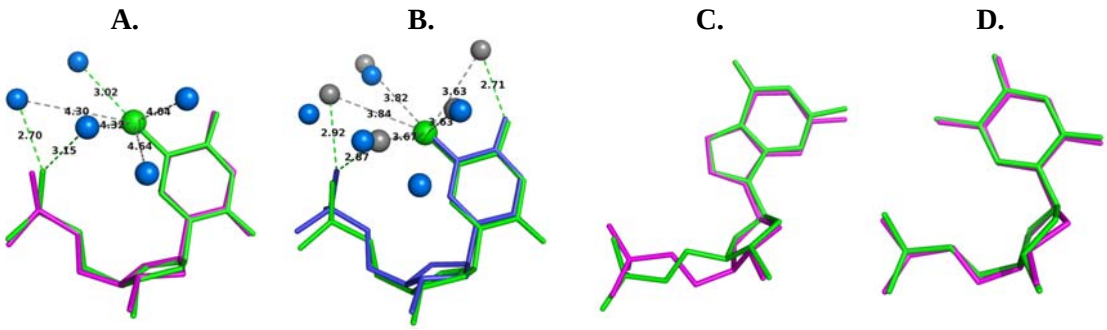


Figure 4.6: The superimposed comparison of the native and Br-DNA local structures. (green ball represents Br). The green and gray dash lines represent hydrogen and non-hydrogen bonds, respectively. (A) the comparison of the dU_{Br} of the Br-DNA ($GTGdU_{Br}ACAC$, 2H05 in green, and water molecules in_blue balls) and the T4 of the native DNA (GTGTACAC, 1DNS in pink). (B) the comparison of the dU_{Br} of the Br-DNA (2H05 in green, and its water molecules in_blue balls) and the T4 of the Se-DNA ($GdU_{Se}GTACAC$, 1Z7I in blue, its water molecules in gray). (C) the comparison of the dG3 of the Br-DNA (in green) and the native (1DNS in pink). (D) the comparison of the T2 of the Br-DNA (in green) and the native (1DNS in pink).

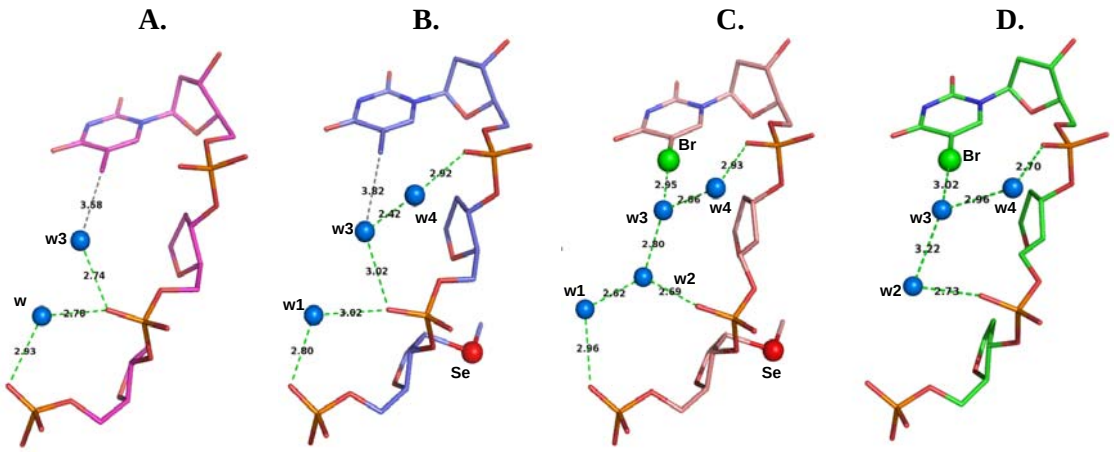


Figure 4.7: The bromine hydration pattern and the local structures. Water molecules, Se, and Br are in blue, red, and green balls, respectively. The green and gray dash lines represent hydrogen and non-hydrogen bonds, respectively. (A) the TGT backbone structure of the native DNA (GTGTACAC, 1D78). (B) the dUSeGT backbone structure of the Se-DNA (GdUSeGTACAC, 1Z7I). (C) the dUSeGdUBr backbone structure of the Se-Br-DNA structure (GdUSeGdUBrACAC, 2GPX). (D) the TGdUBr backbone structure of the Br-DNA (GTGdUBrACAC, 2H05).

4.2.6 Local purterbation and hydration pattern comparison

The native d(GTGTACAC) structure in tetragonal space group (1DNS) did not include any water molecules in the refined structure, while the other native structure in hexagonal space group (1D78), determined at 1.4 Å resolution, has included 37 waters. The TGT local structures of the tetragonal native DNA (GTGTACAC), the hexagonal native, and the Se-DNA are identical (Figure 4.5A, 4.7A and 4.7B). Instead of comparing with those of the tetragonal native structure due to the absence of the water molecules, we compared the water networking and local backbone structure of the Se-DNA structure with those of the hexagonal native structure. The water molecules were found

forming hydrogen bonds to the oxygen atoms of the 5'-phosphate groups of TGT in the native and Se-derivatized structures (Figure 4.7A and 4.7B). In the hexagonal native structure (1D78, Figure 4.7A), two water molecules (W1 and W3) formed hydrogen bonds (the green dash lines) to the oxygen atoms of the T2 and dG3 5'-phosphate groups. In the Se-DNA structure (1Z7I, Figure 4.7B), three water molecules (W1, W3 and W4) formed the hydrogen bonds to each other and to the oxygen atoms of the dUSe, dG3 and T4 5'-phosphate groups, forming the same water networking as the native. In the Se-Br-DNA structure (2GPX in Figure 4.7C), four water molecules (W1, W2, W3 and W4) were found in the region, connecting with each other and the oxygen atoms of the dUSe, dG3 and T4 5'-phosphate groups. These four water molecules formed a different water networking from those of the native and Se-DNA structures. Similar to the Se-Br-DNA structure (Figure 4.7C), the same water networking (W2, W3 and W4) was found in the Br-DNA (GTGdUBrACAC, 2H05, Figure 4.7D). Interestingly, in both the Br-DNA and Se-Br-DNA structures, the W2 water insertion, which can probably facilitate the 5'-phosphate perturbation and the C4'-C5' bond rotation of the dG3, was observed. This inserted water (W2) formed the hydrogen bonds with W1, the dG3 5'-phosphate oxygen, and W3, which was pushed closer to the dUBr bromine atom in both structures. In both the Br-DNA structure (Figure 4.7D) and the Se-Br-DNA structure (Figure 4.7C), we observed that ordered W3 molecules are very close to the bromine atom (3.02 Å and 2.95 Å, respectively), which suggests a hydrogen bond formation. This hydrogen bond may weaken the Br-C5 bond, which probably is one reason of the debromination reported previously.

Our experimental results indicate that though it does not cause the change in the sugar pucker of the dUBr, the Br derivatization placed in major groove caused significant conformation changes of the sugar puckers in the dG3 and T2 and the large changes of the hydration pattern and water molecule networking, which may lead to the local perturbation. Unlike the bromine derivatization, the 2'-selenium derivatization is placed in the minor groove, and the hydration alteration in the minor groove was not observed, probably due to disorder of the water molecules near the 2'-position of T2. No ordered water molecules near the 2'-positions in the Se-derivatized and native structures have been observed. Probably, the hydration alteration caused by the 2'-selenium modification is insignificant.

4.3 Conclusion of this chapter

To facilitate X-ray crystal structure studies of nucleic acids, we are in a process of developing selenium derivatization strategies by synthesizing SeNA *via* atom-specific replacement of oxygen with selenium and studying SeNA structures by comparing them with the structure-known natives and the conventional bromine derivatization. By using both MAD and SAD phasing, we have demonstrated the X-ray crystal structure determination of the 2'-Se-derivtized DNA at high resolution. The Se-derivatized DNA structure (1.28 Å resolution) is identical to the native structure with the same crystal form (33, 2.0 Å resolution). Besides the demonstration of the formation of these Se-derivatized crystals with high diffraction quality and in broader buffer conditions, we also demonstrated that these Se-derivatizated DNAs crystallized faster than these native and Br-derivatized DNAs. Furthermore, by comparing with the conventional bromine

derivative, which retained the same global structure as the native, we found that the bromine derivatization caused the local perturbations, such as the backbone rotation, the water networking disruption, and the hydrate pattern alteration. Our experimental results indicate that the selenium derivatization may assist the studies of the nucleic acid X-ray crystal structures in two major ways: phasing and crystal growth with high quality, which leads to structure determination with high resolution. These experimental results suggest that the selenium derivatization can be an alternative to the conventional bromine derivatization. This Se derivatization strategy *via* the atom-specific substitution will significantly facilitate X-ray crystal structure studies of nucleic acids and their protein complexes.

5. CRYSTALLIZATION FACILITATION OF DNA OLIGONCULEOTIDES BY 2'-SELENIUM FUNCTIONALITY

5.1 Introduction

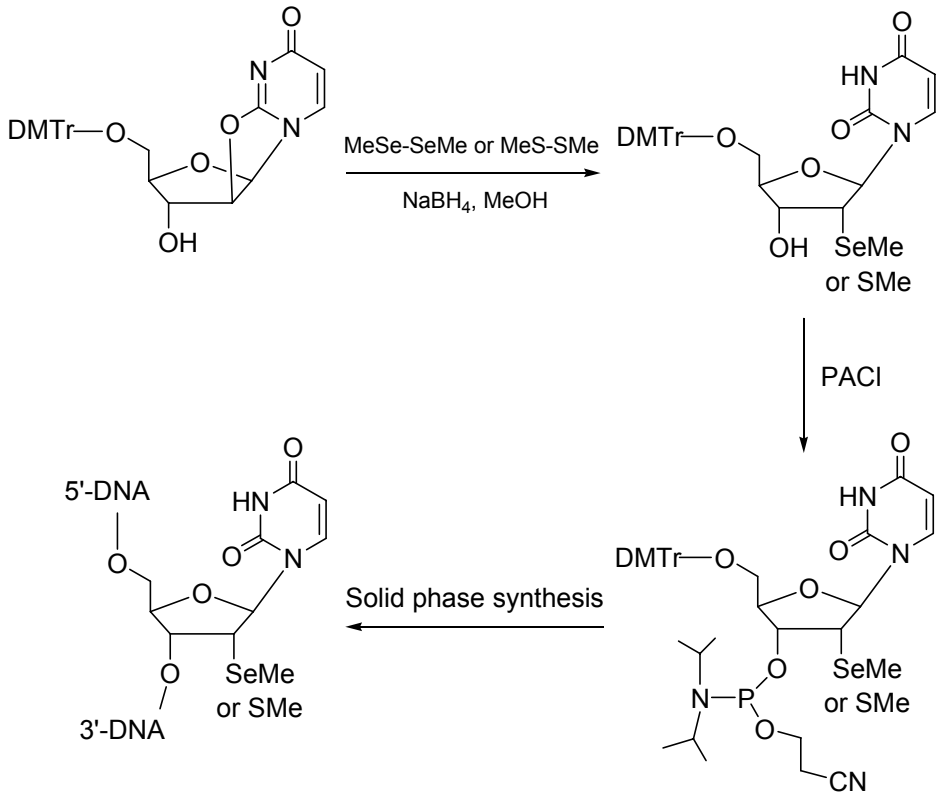
Since the discovery of dramatic facilitation of selenium function modification at 2'-position of ribose to the growth of oligonucleotide crystal, the further effort has been devoted to explore the mechanism of this phenomenom. In addition, it was amazing to find that in certain DNA sequences with 2'-SeMe modification and certain concentrations, the crystal growth could happen in a few minutes with a fixed shape of single or double crystals, therefore, it is possible to real-time monitor the crystal growth by a normal digital camera through microscopes, as well as the calculation of time dependent crystal volumes, which could provide the direct information about the macromolecule packing rate, although the nucleation process is still a mystery.

Inspired by this exciting result, in this section, we systematically incorporated a series of functional group at 2' position, such as –F, -OH, -OMe, -SMe, -H and -SeMe, into a well studied oligonucleotide model sequence 5'-GU*GTACAC-3', where the U is the modified residue. By screening various combination, the same condition for single crysal growth of all these modified DNA was selected. Then, under the same concentration, their time-course crystal growth or the molecular packing was compared. Besides this thermaldynamic study, the following NMR study of the modified nucleosides and mechanical simulation calculation of the modified oligomers are still on going by the writing of this dissertation. The results, together with our current studies, will be summarized and published later.

5.2 Results and discussion

5.2.1 Preparation of DNA samples containing various functionalities at 2'-position

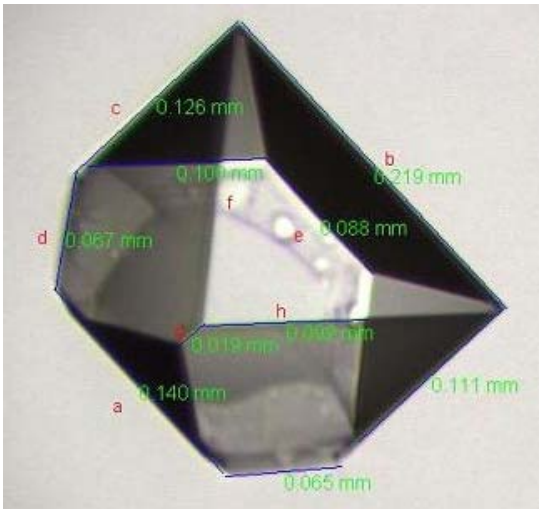
The phosphoramidite building blocks containing 2'-SeMe and 2'-SMe were synthesized through the nucleophilic ring-opening addition of 2,2'-anhydrouridine by either SeMe or SMe group which are generated by the reduction of commercial available reagent dimethyl diselenide and dimethyl disulfide, followed by the standard procedure of making phosphoramidites, as showed in scheme 5.1. All the other building blocks including 2'-OMe, F, OH and H phosphoramidites are purchased from ChemGenes Inc. All the modified octermer were synthesized and purified by the standard DNA synthesis procedures.



Scheme 5.1: Synthesis of oligonucleotides containing 2'-SeMe or 2'-SMe.

5.2.2 Time-course crystals growth measurement of 2'-Se-DNAs

The 2'-SeMe modified model was applied to screen the best crystallization conditions in terms of buffers, temperatures, concentrations of sample and precipitant, more importantly, crystal shape repeatability. Finally, the number seven of the Hampton nucleic acid mini-screen buffer was found to be the best condition. (see appendix A) (buffer consist of: 10% v / v MPD, 40 mM Na Cacodylate pH 6.0, 12 mM Spermine tetra-HCl, 80 mM Sodium Chloride and 20 mM Magnesium Chloride). By setting the concentration of sample and precipitant to 0.25 mM and 35% MPD respectively, the time course growth of single crystal was achieved. The shape of crystal was shown in figure 5.1. All the lengthes and angles was measured with calibrated high resolution digital camera Olympus-3.3MPX. Based on a real model we built, a volume calculation of this crystal was figured out.



$$V= \ 0.7581*c*d*(a+b)/2 + c*(b+c)*[f2-(c2+(b-e)2)/4]1/2/3$$

$$+ a*(c+g) *[h2-(a2+(c-g)2)/4]1/2/3$$

Figure 5.1: Crystal shape, parameters and volume calculation of 2'-SeMe-8mer.

With this information in hand, a time-course volume increasement was then recorded and calculated during the crytal growing. Since it has been known by our crystallography study that one unit cell contains two molecules of DNA duplex, and the unit cell volume is also known, so, although it's not very accurate in terms of the absolute number, we are still able to calculate the rough number of DNA duplex molecules in a fixed shape of crystal. With this guidance, a time coursed molecule packing curve for this 2'-SeMe-8mer DNA crystal was obtained, as showed in figure 5.2.

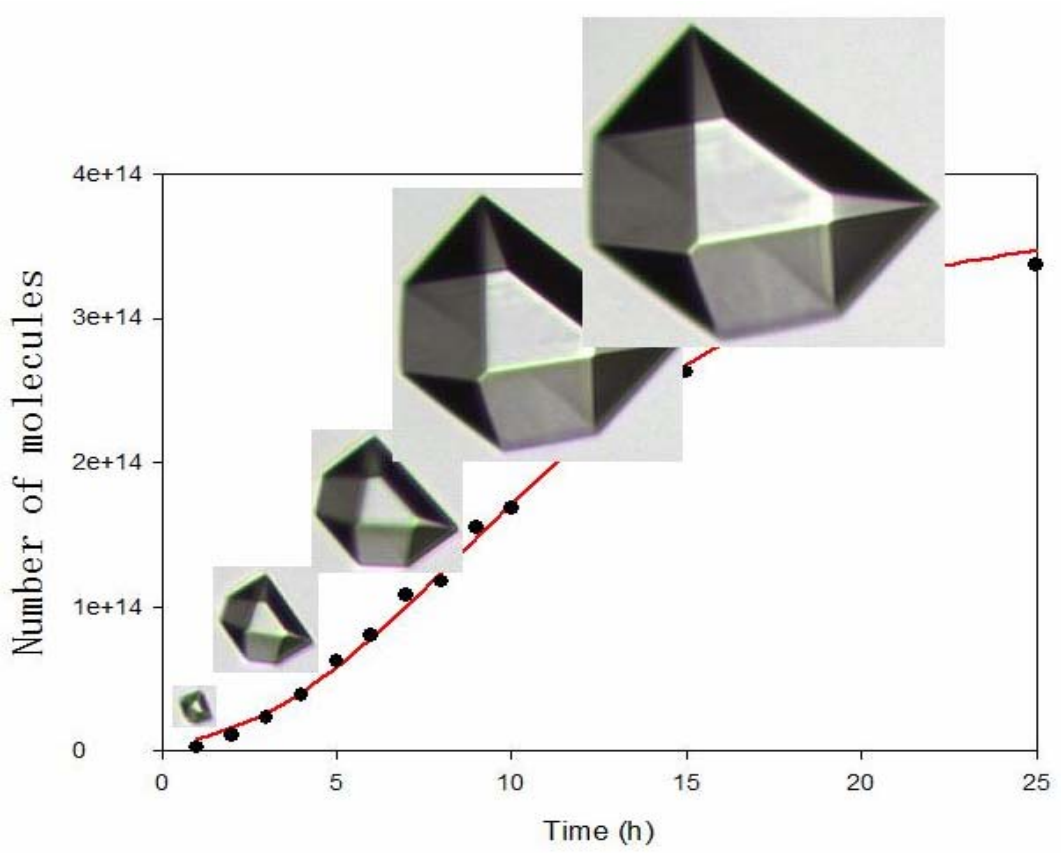


Figure 5.2: Time-course molecule packing curve for the 2'-SeMe-8mer DNA. The picutures on the curve represents the according real size of the time.

5.2.3 Comparison of crystal growth of DNAs containing various functionalities

Then, a comparison of this crystal growth facilitation effect among these 2'-modifications was carried out. Fortunitely, the single crystals with the same shapes could be obtained in the same concentration and crystal growing conditions as selenium derivatization, except the 2'-F, which need much higher concentrations to be crystallized. As example, the molecule packing curves of 2'-SMe and 2'-OMe were placed together with a long time-course (up to 100 hours), as showed in figure 5.3. The 2'-OH and 2'-H growth is too slow in this time scale comparing the aboved three samples, therefore is hard to compare here. For each sample, the crystal growing rate and quality under the 24 screen conditions were further compared in table 5.1.

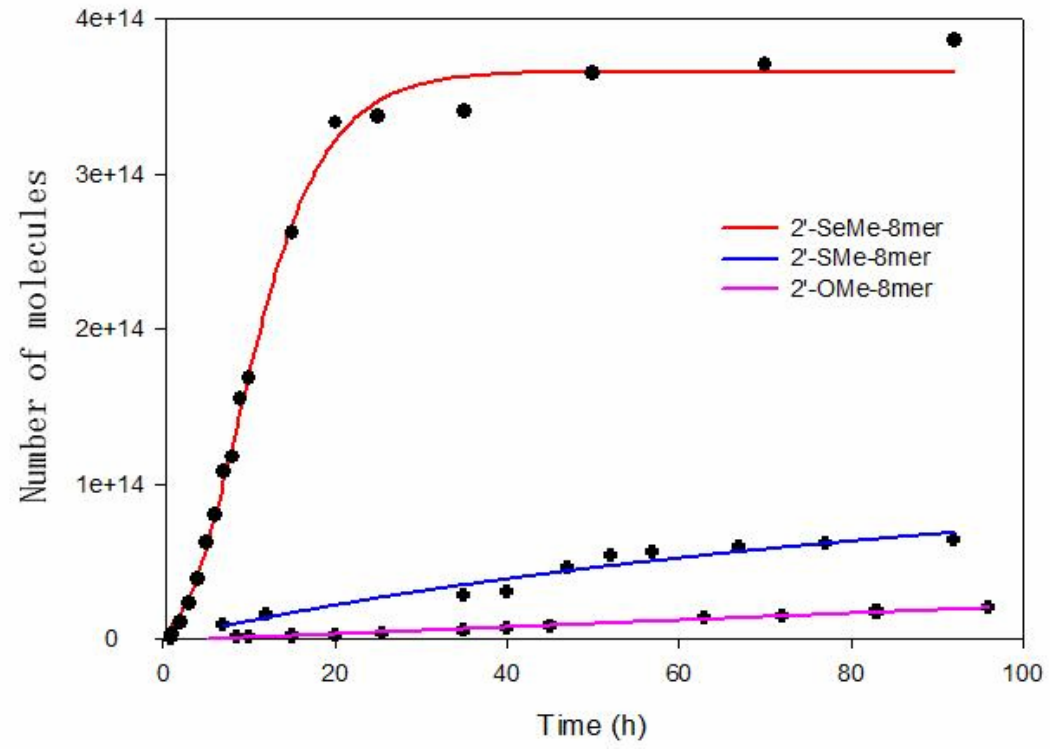


Figure 5.3: Time-course molecule packing comparison of DNA-8mer containing 2'-SeMe, 2'-SMe and 2'-OMe. (Represented by red, blue and pink curves respectively).

Table 5.1: Crystal growth comparison for DNA-8mer with various modifications.

DNA	Working buffers	Best ones	Appear time (ave)	Max-size time (ave)	One crystal rate
H	0	0	None	None	None
F	17 (3,5,67,8,9,10, 11,13,14,15,16, 17,18,19,21,23)	7, 14, 16	Two weeks	N/A in 4 Month	2/96 (3 plate)
OH	11 (4,5,6,7,9,10,11, 15,22,23,24)	5,7,10,23	7 days	7 days	3/120 (5 plate)
OMe	13 (4,6,7,8,10,11, 12,14,16,18,19, 22,23)	6,7,10,11 16	7 days	200 h	1/120 (5 plate)
SMe	15 (2,3,4,5,7,8,9, 10,11,13,15,18, 19,21,23)	7,10, 13, 19	3 days	70 h	2/144 (6 plate)
SeMe	All the 24	5,6,7,9,11 13,18,20, 22,24	8 min	22 h	10/120 (5 plate)

* Conditions: 0.25mM DNA, 35% MPD, r.t.

* Explanation: Working buffers mean the buffers from Hampton Mini-Screen kid which could grow the crystal successfully; Best ones means the buffers which can grow less

than five crystals; The appear time was averaged by certain numbers of experiment in the same condition (0.25 mM).

The results in the aboved table clearly showed that comparing the native one, all these 2' groups showed a certain function of facilitating crystal growth of this 8mer DNA. While comparing the other groups, this function is much stronger, in all the aspects, including the buffer conditioin diversity, number of best crystallization conditions, the first crystal appear time, the maximum-size reach time and the one crystal possibilities.

It's also interesting to point out that the crystal shapes of these DNA samples could be very diverse with different buffer conditions, concentrations and temperatures. The figure 5.4 summerized the most oftenly appeared crystal shapes during this work.

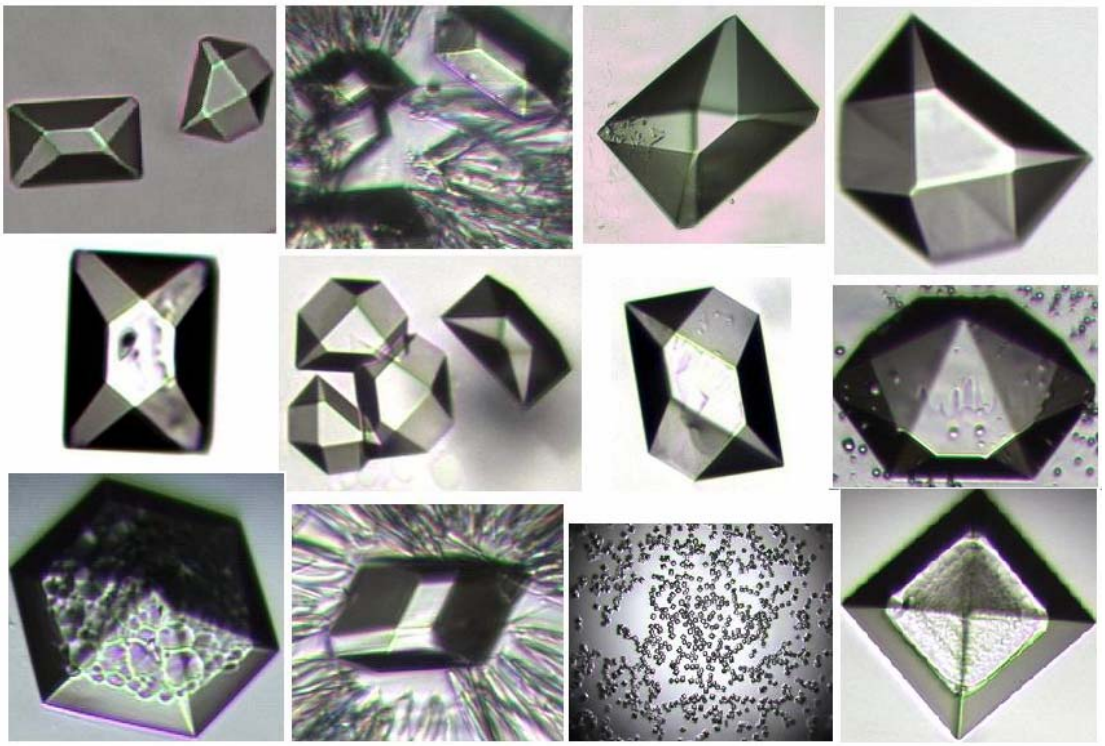


Figure 5.4: The diversity of DNA crystal shapes in this work.

5.3 Conclusion of this chapter

In this work, it was found that the modifications at the 2’ position of deoxyribouridine could facilitate the crystal growth of a DNA-8mer model 5’-GU*GUACAC-3’, and comparing other groups such -SMe, -OMe, -OH, and -F, methylseleno group has much stronger effort, which could make crystals grow in less than ten minutes. By optimizing the crystal growing conditions, we were able to obtain the single crystal with a fixed shape for each modified DNA sample. A crystal volume calculation was figured out based on the measurement of crystal parameters, which was further applied to the quantitative calculation of time-course molecule packing patterns with the support of our previous crystallography results.

As far as we know, this is the first dynamic study of macromolecular crystal packing, although the mechanism is still not clear. We assume that this selenium modification at 2’ position of the ribose could lock the sugar pucker into 2’-exo form in a right extent which provides a perfect scafford for the next molecular packing, while the other modifications generates more dynamic equilibrium for the ribose between the 2’-exo and 2’-endo form. Although the nucleation process is hard and impossible to monitor with current conditions, the fact that this 2’-SeMe-8mer DNA crystal appears in an amazingly short time indicates that the original molecules should have a relatively rigid shape and low entropy value, which makes the starting packing very accurate, fast and highly reproducible. When a certain size is reached, the surface of the crystals can serve as a good template for the self-assembly of more freely wandering molecules into the organized model. Further researches such as NMR and mechanic modeling studies are needed to explore the detailed mechanism of this crystal facilitation.

6. SYNTHESIS OF BASE MODIFIED 4-SELENIUM-THYMIDINE AND 4-SELENIUM-URIDINE FOR STRUCTURE STUDIES OF DNA AND RNA OLIGONUCLEOTIDES

6.1 Introduction

Besides the structural application of SeNA strategy, we are also seeking their potential functional studies for both DNA and RNA. This chapter presents our work about introducing selenium function into the nucleobase, including 4-position of thymidine and uridine.

Nucleic acids participate in many important biological functions in living systems, including the genetic information storage, the gene expression, and catalysis. (1) (2) The T-A and C-G base-pairing and stacking allow the formation of the stable DNA duplex for genetic information storage, transcription and replication. (107) (3) The replacement of oxygen on the nucleobases with sulfur(108) (109) has provided more insights on the DNA duplex stability, recognition and replication at the atomic level. (110) (75) The 2-thio thymidine causes the slight stabilization of the T-A pare and DNA duplex, while the 4-thio thymidine causes the slight destabilization of them. (73) The recent studies on these sulfur modifications have revealed the enhanced base-pare selectivity, and the replication efficiency and fidelity, especially with the 2-thio thymidine probably due to discouraging the undesired T-G wobble base-pairing by the larger sulfur atom at 2-position of the thymine.

As selenium (atomic radius, 1.16 Å), another element member of Family VIA in the Periodic Table, is much larger than oxygen (0.73 Å), the replacement of O with Se on the nucleobases of nucleic acids will most likely provide more advantages over the

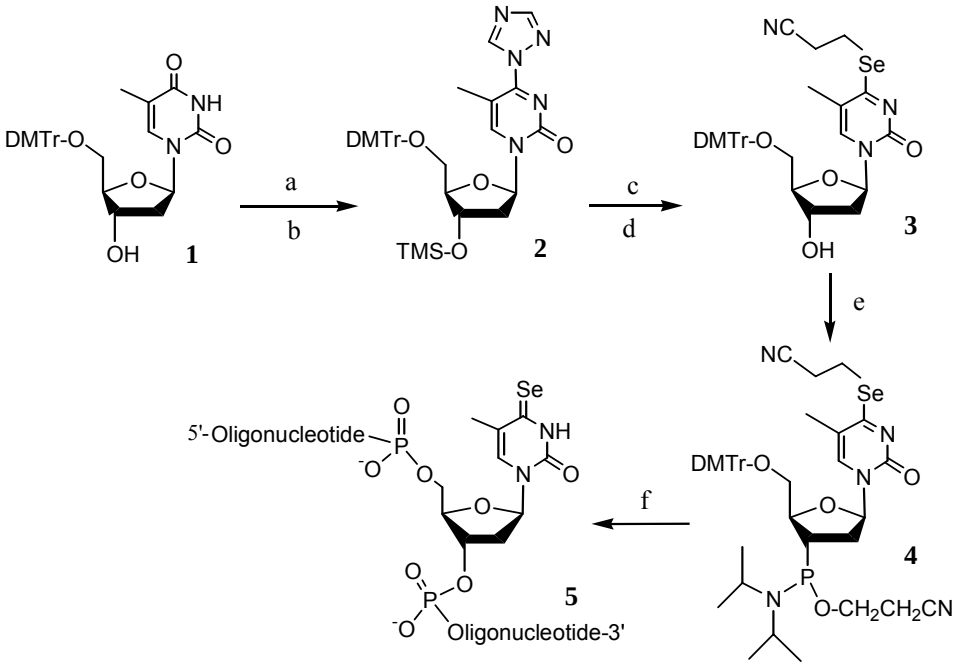
natives and reveal more insights on the base-pairing selectivity, the duplex stability and recognition, and the replication efficiency and fidelity. Combining the function of anormalous dispersion center, we believe that this nucleobase Se-derivatization can provide unique opportunities over the Se-derivatized nucleic acids (SeNA) with Se on other positions for the structure study. In addition, specific pyrimidines in natural tRNAs have been derivatized by Se on the nucleobases, and the Se function has been barely understood though it was suggested that the Se substitution might involved in the tRNA anticodon. (77) Furthermore, the discoveries on the functional importance of small RNAs and specialized RNA domains in many biological processes have increased dramatically over the past decade, so, the further investigation of phenomena such as RNA interference and RNA-protein interactions for gene regulation requires advanced research on RNA base pair recognition and stability based on hydrogen bonding, stacking interaction and size-shape impacts, which is also accompanied by a growing demand for X-ray structure analysis of RNAs. Despite the synthesis of the Se-containing nucleobases and nucleosides over decades ago, (76) (111) synthesis of the nucleic acids containing the Se-nucleobases remained a challenge.

6.2 4-Se-thymidine containing DNA

6.2.1 Synthesis of 4-Se-thymidine and its phorsphoramidite

As showed in figure 6.1, to replace the 4-oxygen with selenium, we activated the partially protected thymidine derivative (1) at the 4-position via the formation of triazolide 2 in two reactions (87% yield). We also developed di(2-cyanoethyl) diselenide reagent for the Se introduction. (112) After reducing the diselenide reagent to the corresponding

selenol with $NaBH_4$, the protected Se functionality was introduced to the 4-position of 2 by displacing the 4-triazolyl activating group, achieving the simultaneous Se incorporation and protection for the Se-T intermediate (3). Due to the undesired removal of the 2-cyanoethyl group on Se while removing the 3'-TMS group by the fluoride or base treatment, we developed a mild condition to remove the 3'-TMS group by using 10% triethylamine in methano (81% yield). Alternatively, the 3'-TMS group could also firstly be removed by the fluoride treatment, followed by the Se incorporation (85% yield in two reactions). Finally, Se-thymidine derivative 3 was converted to the corresponding phosphoramidite 4 in 92% yield. The detailed experiment and characterization for each compound were listed in the following.



Scheme 6.1: Synthesis of 4-Se-thymidine phosphoramidite (4) and oligonucleotides containing 4-Se-T (5). (a) TMS-Im and CH_3CN; (b) triazole-$POCl_3$-Et_3N; (c) $(NCCH_2CH_2Se)_2/NaBH_4$, EtOH; (d) 10% Et_3N in MeOH; (e) 2-cyanoethyl *N,N*-diisopropylchlorophosphor-amidite and *N,N*-diisopropylethylamine in CH_2Cl_2; (f) the solid phase synthesis.

4-(1,2,4-triazol-1-yl)-5'-O-(4,4'-dimethoxytriphenylmethyl)-3'-O-trimethylsilylthymidine(2). Phosphorus oxychloride (0.92 mL, 10 mmol) was added to a solution of 1,2,4-triazole (2.77 g, 40 mmol) in dry acetonitrile (32 mL) under argon, and the reaction was stirred for 45 min at room temperature. Dry triethylamine (11.2 mL, 79 mmol) was then injected into the reaction and the reaction was stirred for another 45 min. Without exposing to air, the suspension was filtered directly into a round bottle flask (100 mL) containing compound **1** (1.82 g, 3.35 mmol) reacted with 1-(trimethylsilyl)imidazole (0.98 ml, 6.67 mmol) in dry acetonitrile (15 mL) under argon. The reaction mixture was allowed to run for 1 hr at room temperature. The formation of N^4-triazolide **2** was revealed by a fluorescent spot on TLC (5% MeOH in CH_2Cl_2, product R_f = 0.25). After the reaction was complete, the reaction mixture was poured into water (150 mL), and extracted with ethylacetate (3x 80 mL). The combined organic layer was washed with saturated NaCl (100 mL) and dried over $MgSO_4$ before evaporation. After loading the crude product on a silica gel column equilibrated with hexanes/CH_2Cl_2 (1:1), the column was eluted with hexanes/CH_2Cl_2 (1:1), pure CH_2Cl_2, and then with a methanol/CH_2Cl_2 gradient (0, 0.5% and 1% methanol in CH_2Cl_2). After solvent evaporation and dry on high vacuum, the pure compound (**2**) was obtained as a white foam product (1.94 g, 87% yield). ^{1}H-NMR (300 MHz, $CDCl_3$) δ: 0.61 [s, 9H, $(CH_3)_3Si$], 1.98 (s, 3H, CH_3), 2.26-2.35 and 2.63-2.71 (2x m, 2H, H-2'), 3.30-3.60 (2x dd, *J* = 2.7, 11.1 Hz, 2H, H-5'), 3.79 (s, 6H, 2xCH_3O), 4.07-4.10 (m, 1H, H-3'), 4.46-4.52 (q, *J* = 4.8 Hz, 1H, H-4'), 6.30 (t, *J* = 6.3 Hz, 1H, H-1'), 6.84 (dd, *J* = 1.5, 9.0 Hz, 4H, aromatic), 7.22-7.42 (m, 9H, aromatic), 8.09 (s, 1H, H-6), 8.42 (s, 1H, H-Im), 9.29 (s, 1H, H-Im). ^{13}C-NMR (100 MHz, $CDCl_3$) δ: -0.06 (CH_3-Si), 16.54 (CH_3), 42.20 (C-2'), 55.26 (OCH_3), 62.12 (C-5'), 70.58 (C-3'), 86.88, 135.31,

135.33, 144.23, 158.16 (Ar-C), 87.16 (C-4'), 87.65 (C-1'), 105.71 (C-5), 113.29, 113.31, 127.14, 128.02, 128.13, 130.06, 130.07 (Ar-CH), 146.70 (C-6), 153.36 (CH-Im), 153.95 (C-2), 158.79 (C-4). HRMS (ESI-TOF): molecular formula, $C_{36}H_{41}N_5O_6Si$; $[M+Na]^+$: 690.2724 (calc. 690.2718).

4-(2-cyanoethyl)seleno-5'-O-(4,4'-dimethoxytriphenylmethyl) thymidine (3).

The $NaBH_4$ suspension (250 mg $NaBH_4$ in 5 mL of EtOH) was injected into a flask containing di(2-cyanoethyl) diselenide [$(NCCH_2CH_2Se)_2$, 0.45 mL, d= 1.8 g/mL, 3.0 mmol] and ethanol (8 mL) on an ice bath and under argon. After injection of the $NaBH_4$ suspension for 15 minutes, the ice bath was removed. The yellow color of the diselenide disappeared in approximately 15 min, giving an almost colorless solution of sodium selenide ($NCCH_2CH_2SeNa$). The solution of **2** (1.34 g, 2 mmol) in THF (4 mL) was injected to the solution of sodium selenide. After the selenium incorporation was completed in 15 min. (monitored on TLC, 5% MeOH in CH_2Cl_2, product R_f = 0.58), water (100 mL) was added to the reaction flask. The solution was adjusted to pH 7-8 using CH_3COOH (10%), and was then extracted with ethyl acetate (3 x 100 mL). The organic phases were combined, washed with NaCl (sat., 100 mL), dried over $MgSO_4$ (s) for 30 min., and evaporated to minimum volume under reduced pressure. Deprotecting solution (5 mL, 10% triethylamine in methanol) was added to the crude product, and the reaction was stirred for 4 hours to completely remove the 3'-TMS group (monitored on TLC, 5% MeOH in CH_2Cl_2, product R_f = 0.46). The deprotecting solution was evaporated under reduced pressure. The crude product was then dissolved in methylene chloride (5 mL) and purified on a silica gel column equilibrated with hexanes/methylene chloride (1:1). The column was eluded with a gradient of methylene

chloride (CH_2Cl_2, 0.5%, 1%, and 2% MeOH in CH_2Cl_2, 500 mL each). After solvent evaporation and dry on high vacuum, the pure compound (**3**) was obtained as a white foamy product (1.07 g, 81% yield). ^{1}H-NMR (400 MHz, $CDCl_3$) δ: 1.62 (s, 3H, CH_3), 2.27-2.35 and 2.71-2.77 (2x m, 2H, H-2'), 3.00 (dd, J = 6.5 and 6.7 Hz, 2H, CH_2-CN), 3.32 (br, 1H, OH), 3.37-3.41 (m+t, J = 6.7 Hz, 3H, 1H-5', CH_2-Se), 3.52 (dd, $J_{5'-5'}$ = 10.6 Hz, $J_{5'-4'}$ = 3.0 Hz, 1H, 1H-5'), 3.81 (s, 6H, 2x OCH_3), 4.17-4.22 (m, 1H, H-3'), 4.57-4.62 (m, 1H, H-4'), 6.34 (t, $J_{1'-2'}$ = 6.3 Hz, 1H, H-1'), 6.85 (dd, J = 8.8 Hz, 4H, aromatic), 7.23-7.42 (m, 9H, aromatic), 7.91 (s, 1H, H-6). ^{13}C-NMR (100 MHz, $CDCl_3$) δ: 14.30 (CH_3), 18.73 ($\underline{C}H_2$-CN), 20.53 ($\underline{C}H_2$-CH_2-CN), 42.17 (C-2'), 55.29 (OCH_3), 63.18 (C-5'), 71.71 (C-3'), 86.90, 135.37, 135.41, 144.36, 158.73 (Ar-C), 86.73 (C-4'), 87.01 (C-1'), 113.31, 127.17, 128.03, 128.10, 130.07 (Ar-CH), 114.38 (C-5), 119.00 (CN), 137.56 (C-6), 153.35 (C-2), 177.20 (C-4). HRMS (MALDI-TOF): molecular formula, $C_{34}H_{35}N_3O_6Se$; $[M+Na]^+$: 684.1583 (calc. 684.1583).

Alternative approach: Et_3N•3HF (7.5 mL, 1 M in THF, 7.5 mmol) was added to the solution of **2** (4.98 g, 7.47 mmol) in THF (10 mL). The reaction was completed after stirring under argon for 15 minutes (monitored by TLC). The solvent of the reaction was removed by rotary evaporation and the crude intermediate was then dried on high vacuum. The $NaBH_4$ suspension (933 mg $NaBH_4$ in 10 mL of EtOH) was injected into a flask containing di(2-cyanoethyl) diselenide [$(NCCH_2CH_2Se)_2$, 1.66 mL, d= 1.8 g/mL, 11.2 mmol, 3 eq.] and ethanol (10 mL) on an ice bath and under argon. After adding the $NaBH_4$ suspension and stirring it for 15 minutes, the ice bath was removed. The brown-yellow color of the diselenide disappeared in approximately 15 min, giving an almost colorless solution of sodium selenide ($NCCH_2CH_2SeNa$). The sodium selenide solution

was then injected into the flask containing the crude intermediate under argon. After the selenium incorporation reaction was completed in 45~60 minutes (monitored on TLC, 5% MeOH in CH_2Cl_2, product R_f = 0.58), CH_2Cl_2 (100 mL) was added to the reaction flask, followed by addition of NaCl solution (100 mL, saturated) and addition of CH_3COOH (5%) to adjust the pH to 7-8. After removing the organic phase, the aqueous phase was extracted again with CH_2Cl_2 (2 x 100 mL). The organic phases were combined, washed with NaCl (sat., 50 mL), dried over $MgSO_4$ (s) for 25 min., and evaporated to minimum volume under reduced pressure. The crude product was dissolved in CH_2Cl_2 (5 mL) and purified on a silica gel column equilibrated with CH_2Cl_2 and eluded with a gradient (CH_2Cl_2, 0.5%, 1.0%, 2.0% MeOH in CH_2Cl_2, 400 mL each). After solvent evaporation and dry on high vacuum, the pure compound (**3**) was obtained as a white foamy product (4.20 g, 85% yield).

4-(2-cyanoethyl)seleno-5'-O-(4,4'-dimethoxytriphenylmethyl)-thymidine3'-O-(2-cyanoethyl) diisopropylamino phosphoramidite (4).

To the flask (25 mL) containing **3** (453 mg, 0.68 mmol) under argon, dry methylene chloride (2.5 mL), *N*,*N*-diisopropylethylamine (0.17 mL, 1.03 mmol, 1.5 eq.), and 2-cyanoethyl *N*,*N*-diisopropyl-chlorophosphoramidite (195 mg, 0.83 mmol, 1.2 eq.) were added sequentially. The reaction mixture was stirred at -10 °C in an ice-salt bath under argon for 10 minutes, followed by removal of the bath. The reaction was completed in 45 minutes at room temperature (indicated by TLC, 5% MeOH in CH_2Cl_2, product R_f = 0.63 and 0.68), generating a mixture of two diastereisomers. The reaction was then quenched with $NaHCO_3$ (2 mL, sat.), stirred for 5 min, and the product was then extracted with CH_2Cl_2 (3 x 5 mL). The combined organic layer was washed with NaCl

(10 mL, sat.) and dried over $MgSO_4$ (s) for 20 min, followed by filtration. The solvent was then evaporated under reduced pressure and the crude product was re-dissolved in CH_2Cl_2 (2 mL). This solution was added drop-wise to petroleum ether (100 mL) under vigorous stirring, generating a white precipitate. The petroleum ether solution was decanted. The crude product was re-dissolved again in CH_2Cl_2 (2 mL) and then loaded into an Al_2O_3 column (neutral) that was equilibrated with CH_2Cl_2/Hexanes (1:1). The column was eluded with a gradient of methylene chloride and ethyl acetate [CH_2Cl_2 to CH_2Cl_2/EtOAc (7:3)]. After solvent evaporation and dry over high vacuum, the pure **4** (538 mg) was obtained as a white foamy product (92% yield). ^{1}H-NMR (400 MHz, $CDCl_3$, two sets of signals from a mixture of two diastereomers) δ: 0.85-1.20 (m, 24H, 8x CH_3-ipr), 1.55 and 1.57 (2x s, 6H, 2x CH_3), 2.30-2.38 and 2.70-2.82 (2x m, 4H, 2x H-2'), 2.34 and 2.64 (2x t, J = 6.4 Hz, 4H, 2x O-CH_2-$\underline{CH_2}$-CN), 3.00-3.04 (m, 4H, 2x Se-CH_2-$\underline{CH_2}$-CN), 3.32-3.44 (m, 6H, 2x H-5', 2x Se-CH_2), 3.52-3.64 (m, 8H, 4x CH-ipr, 2x O-$\underline{CH_2}$-CH_2-CN), 3.73-3.84 (m, 2H, 2x H-5'), 3.82 and 3.83 (2x s, 12H, 4x OCH_3), 4.18-4.24 (m, 2H, 2x H-3'), 4.63-4.72 (m, 2H, 2x H-4'), 6.28 and 6.32 (2x t, $J_{1'-2'}$ = 6.3 Hz, 2H, 2x H-1'), 6.83-3.88 (m, 8H, aromatic), 7.27-7.43 (m, 18H, aromatic), 7.82 and 7.94 (2x s, 2H, 2x H-6). ^{13}C-NMR (100 MHz, $CDCl_3$, two sets of signals from a mixture of two diastereomers) δ: 14.18 (CH_3), 18.78 (Se-CH_2-$\underline{C}H_2$-CN), 20.15 and 20.23 (O-CH_2-$\underline{C}H_2$-CN), 20.43 (Se-$\underline{C}H_2$-CH_2-CN), 24.48, 24.52, 24.55, 24.59, 24.67 (CH_3-ipr), 40.85, 40.89 and 41.17, 41.19 (C-2'), 43.17, 43.25, 43.29, 43.38 (CH-ipr), 55.27 and 55.31 (OCH_3), 58.15 and 58.34 (O-$\underline{C}H_2$-CH_2-CN), 62.30 and 62.65 (C-5'), 72.31 and 72.87 (C-3'), 85.68 and 85.86 (C-4'), 86.74 and 86.76 (C-1'), 86.87, 135.28, 135.31, 144.25, 158.77 (Ar-C), 113.27, 127.18, 127.23, 128.00, 128.18, 128.26, 130.13, 130.19, 130.21 (Ar-CH),

114.04 and 114.08 (C-5), 117.38 and 117.53 (O-CH_2-CH_2-CN), 119.03 (Se-CH_2-CH_2-CN), 137.41 and 137.44 (C-6), 153.10 (C-2), 176.98 (C-4). HRMS (MALDI-TOF): molecular formula, $C_{43}H_{52}N_5O_7PSe$; $[M+H]^+$: 862.2836 (calc. 862.2842).

6.2.2 Preparation of DNA oligonucleotides containing 4-Se-T

The 4-Se-T phosphoramidite was examined on solid-phase nucleotide synthesis. As expected, it is compatible with the coupling reaction, acetylation, I_2 oxidation, and trichloroacetic acid treatment to remove the 5'-DMTr group, without causing any deselenization. Unfortunately, the concentrated ammonia treatment, which is commonly used for deprotecting the nucleobases and the 2-cyanoethyl groups from the backbone, will cause the deselenization probably due to the good electrophile of the positioin 4 in the thymidine base under high pH values caused by the selenium functionality. To prevent this deselenization, we found the deprotecting condition by using 0.05 M of potassium carbonate solution in methanol, which could remove the 2-cyanothyl group on Se in two hours. The oligonucleotides synthesized using the phosphoramidites containing the ultramild nucleobase deprotection groups can also be fully deprotected for eight hours or overnight with the same potassium carbonate solution. Then the solution was neutralized with 1M HCl and ready to be purified. The synthesized Se-DNAs were purified and analyzed by HPLC and MS (Figure 6.1 and Table 6.1). A typical crude Se-DNA HPLC profile, which is almost identical to that of the corresponding native, as well as the according MALDI-TOF mass spectrum is shown in Figure 6.2. Less than 2% of short oligonucleotides were formed in this Se-decamer synthesis, indicating the high yield of the Se-thymidine coupling (99%).

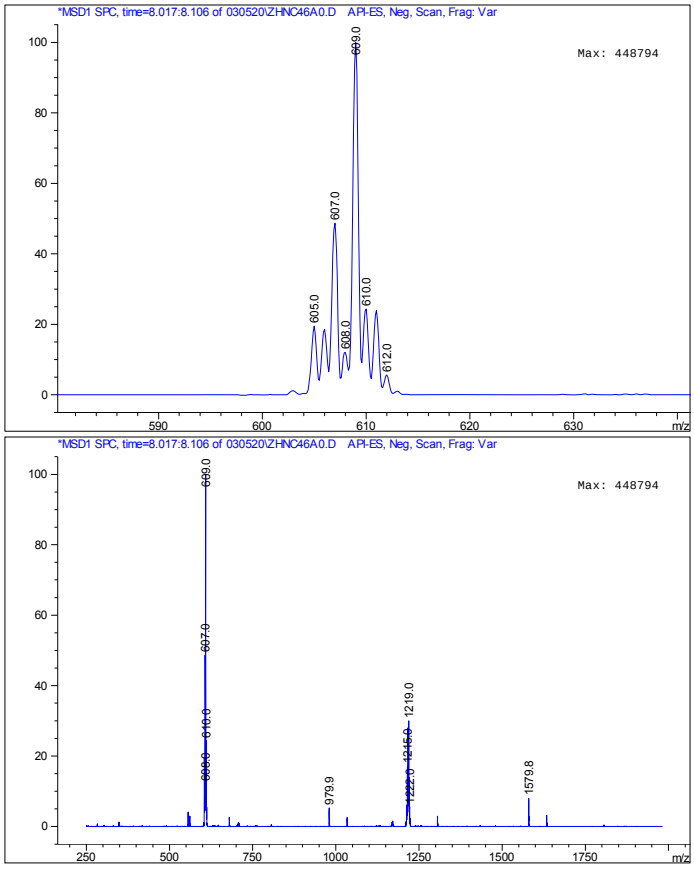
Table 6.1: MS analysis of the DNAs Containing 4-Se-T.

entry	Se oligonucleotides	measured (calcd) *m/z*
1	5′-^{Se}TT-3′ $C_{20}H_{17}N_4O_{12}PSe$: FW 610.1	$[M - H^+]^-$: 609.0 (609.1)
2	5′-DMTr-^{Se}T^{Se}TT-3′ $C_{51}H_{58}N_6O_{19}P_2Se_2$: FW 1280.1	$[M - H^+]^-$: 1279.0 (1279.1)
3	5′-T^{Se}TTT-3′ $C_{40}H_{53}N_8O_{25}P_3Se$: FW 1218.1	$[M + H^+]^+$: 1219.0 (1219.1)
4	5′-TT^{Se}TTT-3′ $C_{50}H_{66}N_{10}O_{32}P_4Se$: FW 1522.2	$[M + H^+]^+$: 1523.0 (1523.2)
5	5′-G^{Se}TGTACAC-3′ $C_{78}H_{99}N_{30}O_{45}P_7Se$: FW 2473	$[M + H^+]^+$: 2476 (2474)
6	5′-ATGG^{Se}TGCTC-3′ $C_{88}H_{112}N_{32}O_{53}P_8Se$: FW 2792.4	$[M + H^+]^+$: 2792.5 (2793.4)
7	5′-GCG^{Se}TATACGC-3′ $C_{97}H_{123}N_{38}O_{57}P_9Se$: FW 3091.0	$[M + H^+]^+$: 3091.6 (3092.0)

6.2.3 Study of 4-Se-T DNA with yellow color

More excitingly about this work, we have discovered for the first time the yellow colored 4-Se-thymidine with special UV-absorption at 369 nm (Figure 6.3) by just altering a single atom on native thymidine (colorless). This visualization of DNA is of a lot of interest and importance to many biochemical, biological and medical applications, (113) (114) such as gene expression analysis, human disease-and-pathogen detection and even the crystal screening of nucleic acid-protein complex. This function of DNA visualization is usually achieved by intercalated dye molecules (115) (116) and specific DNA probes tagged with fluorescent moieties. (117) (118) However, these intercalated dye molecules and tethered fluorescent moieties are often bulky, which can cause perturbation of DNA duplex structures. So, this atom-specific selenium modification could have potential to be applied as a novel DNA probes for visualization without obvious structure perturbation.

A: HO-^{Se}T-T-OH



B: DMTr-^{Se}T-^{Se}T-T-OH

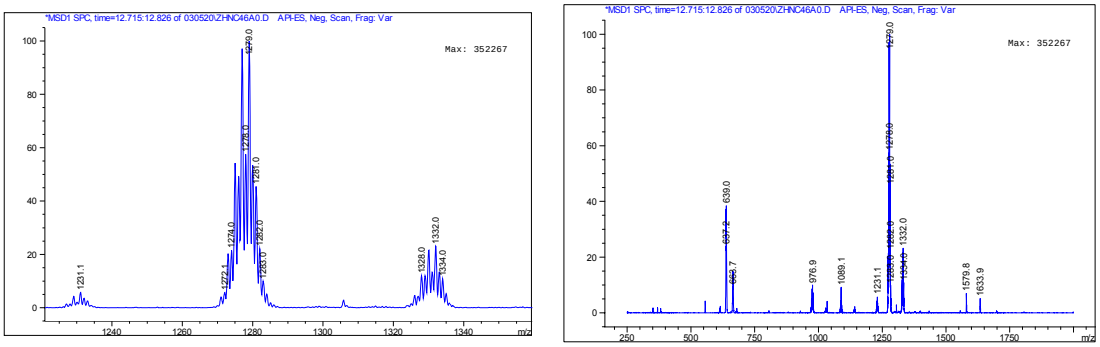


Figure 6.1: LC-MS analysis of SeNAs. (A): 5'-^{Se}TT-3'; Molecular Formular: $C_{20}H_{17}N_4O_{12}PSe$; $[M-H]^-$: 609.0 (calcd: 609.1). (B): 5'-DMTr-^{Se}T^{Se}TT-3'; Molecular Formular: $C_{51}H_{58}N_6O_{19}P_2Se_2$; $[M-H]^-$: 1279.0 (calcd: 1279.1)

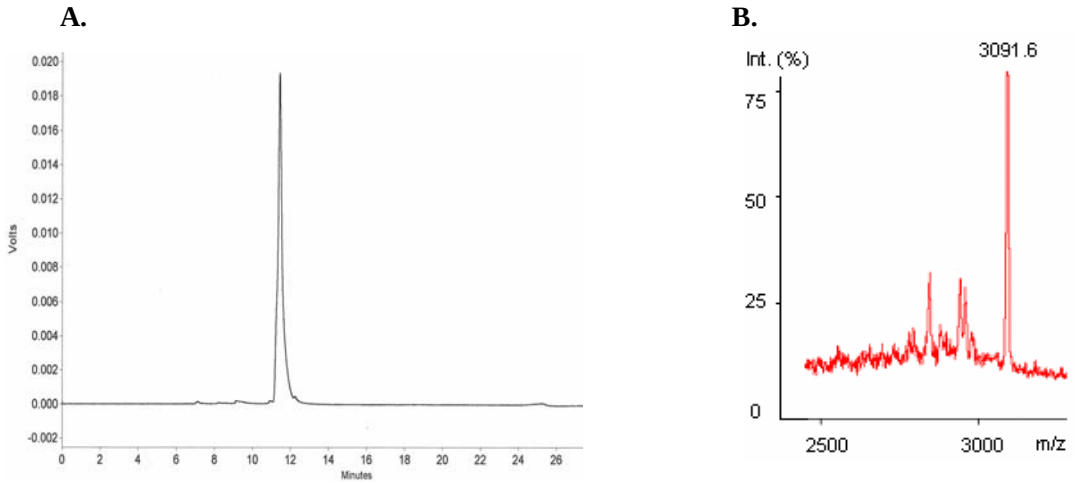


Figure 6.2: HPLC and MS analysis of purified 5'-GCG(^{Se}T)ATACGC-3'. (A) HPLC analysis: the sample was analyzed on a Zorbax SB-C18 column (4.6 x 250), eluted (1ml/min) with a linear gradient from buffer A (20mM triethylammonium acetate, pH 7.1) to 100% buffer B (50% acetonitrile, 20 mM triethylammonium acetate, pH 7.1) in 30 min. Its retention time is 11.52 min. (B). MALDI-MS analysis: Molecular Formular: $C_{97}H_{123}N_{38}O_{57}P_9Se$; $[M+H]^+$: 3091.6 (calcd: 3092.0).

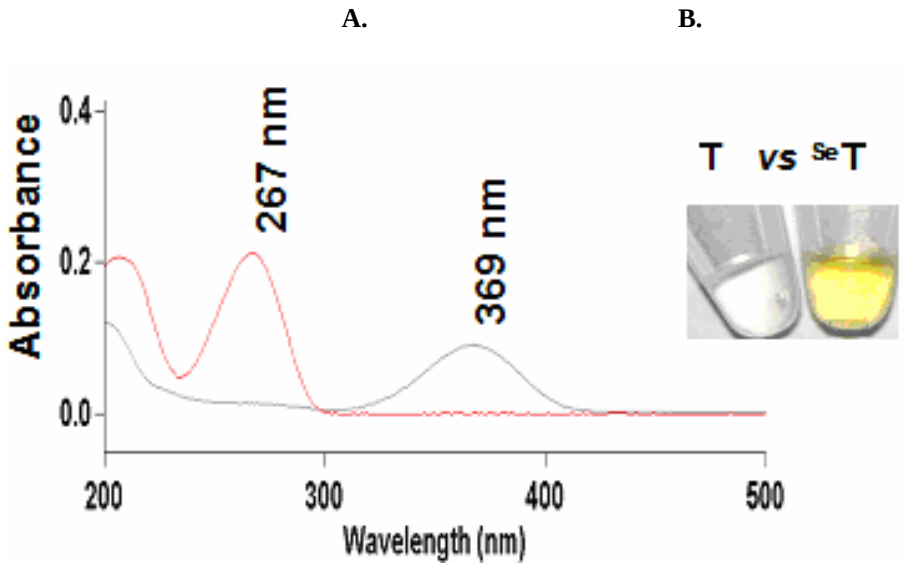


Figure 6.3: UV spectra and color of natural thymidine and 4-Se-thymidine TTP. (A): UV spectrum of thymidine (red, λ_{max}= 267 nm) and UV spectrum of 4-Se-thymidine (gray, λ_{max}= 369 nm); (B): thymidine (on left, colorless) and 4-Se-thymidine (on right, yellow).

6.2.4 Thermal stability and denaturalization study

Besides that, the thermal stability and thermal denaturalization studies of this selenium modification containing DNA was carried out. Figure 6.4 indicated that the selenium functionality of the research model is stable at 60 ^{o}C over three hours with less than 5% was decomposed. Considering that the stability in air is required in most crystallization experiments, we have also monitored the stability of the Se-derivatized DNA for a week in dark. The results showed in figure 6.5 indicated that the Se-DNA is quite stable for long time.

Based on the previous results, the UV-melting temperatures were measured for the duplex system of 5'-ATGG^{Se}TGCTC-3' and 5'-GAGCACCAT-3'. Figure 6.6 is the original Tm curves which repeated for four times. It clearly showed that the derivatized Tm of Se-DNA (38.6 ^{o}C) is very close to that of the corresponding native duplex (39.2 ^{o}C), suggesting no significant perturbation caused by the this modification on thymine.

6.2.5 Crystallization and structure determination of 4-Se-T- DNA

After confirming the structure isomorphism by Tm study, we have also successfully crystallized the 4-Se-T-containing DNA (5'-G-dU_{Se}-G-^{Se}T-A-C-A-C-3', self-complementary), where the dU_{Se} (2'-Se-dU) was used to facilitate the crystal growth based on our previous results. Though the Se functionality is relatively stable in air, 10 mM DTT was still used in the crystallization buffers to prevent the oxidative deselenization of 4-Se-thymidine. The Se-DNA crystallized in three days in buffer # 8, #17, and #21 of the Hampton kit). The structure of the Se-DNA crystal, grown under buffer #17, was determined at 1.50 Å resolution via Se MAD phasing.

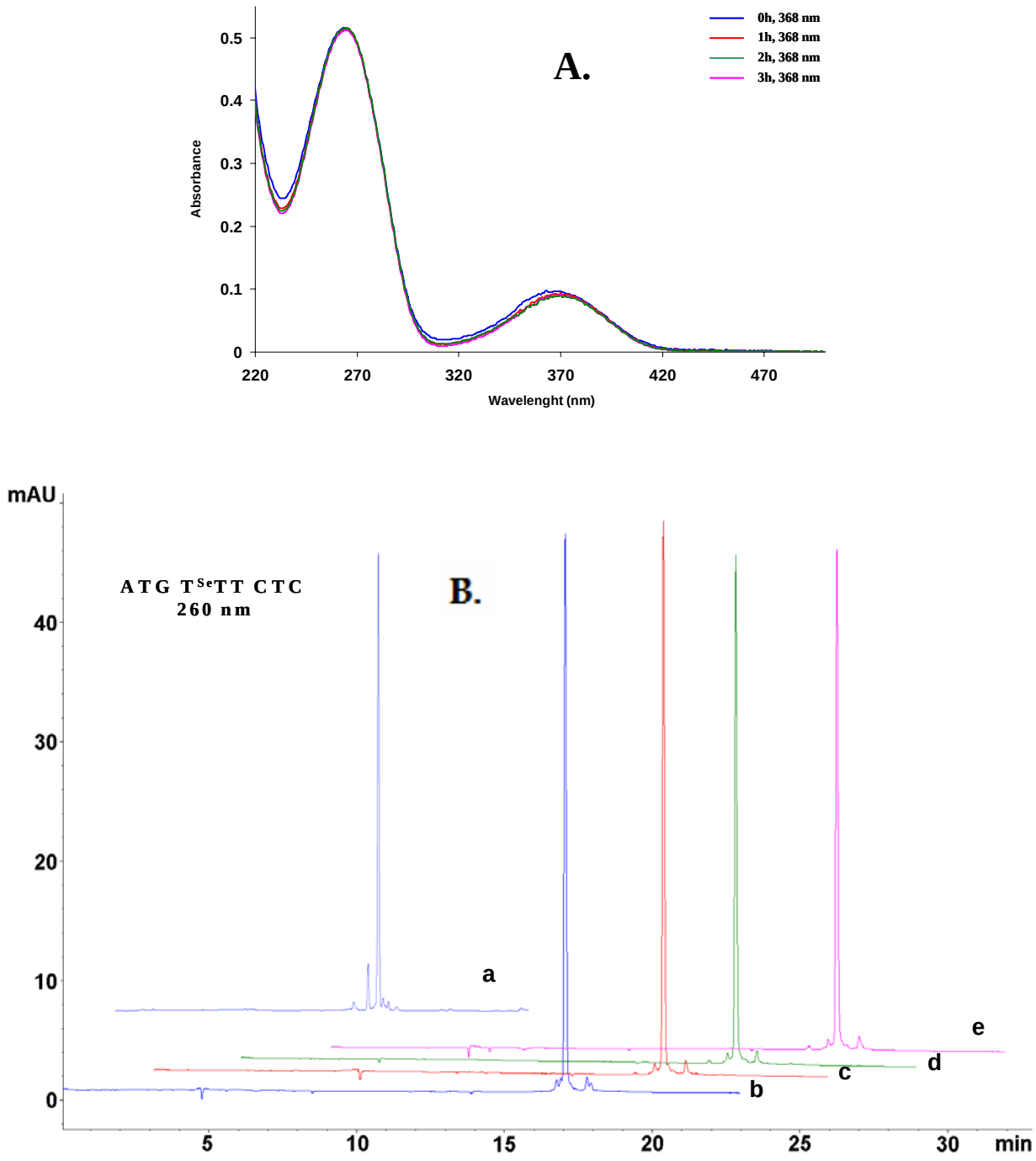


Figure 6.4: UV thermostability study of the SeNA DNA-9-mer (5'-ATGT-^{Se}T-TCTC-3'). Smaple heated at 60 ^{o}C for 3 hours in buffer (5 mM, NaH_2PO_4-Na_2HPO_4, pH 7.5). A. UV analysis over 3 hr. B. HPLC analysis: a) 0 hr, monitored at 260 nm by HPLC; b) 0 hr; monitored at 368 nm by HPLC; c) 1 hr, monitored at 368 nm by HPLC; d) 2 hr, monitored at 368 nm by HPLC; e) 3 hr, monitored at 368 nm by HPLC.

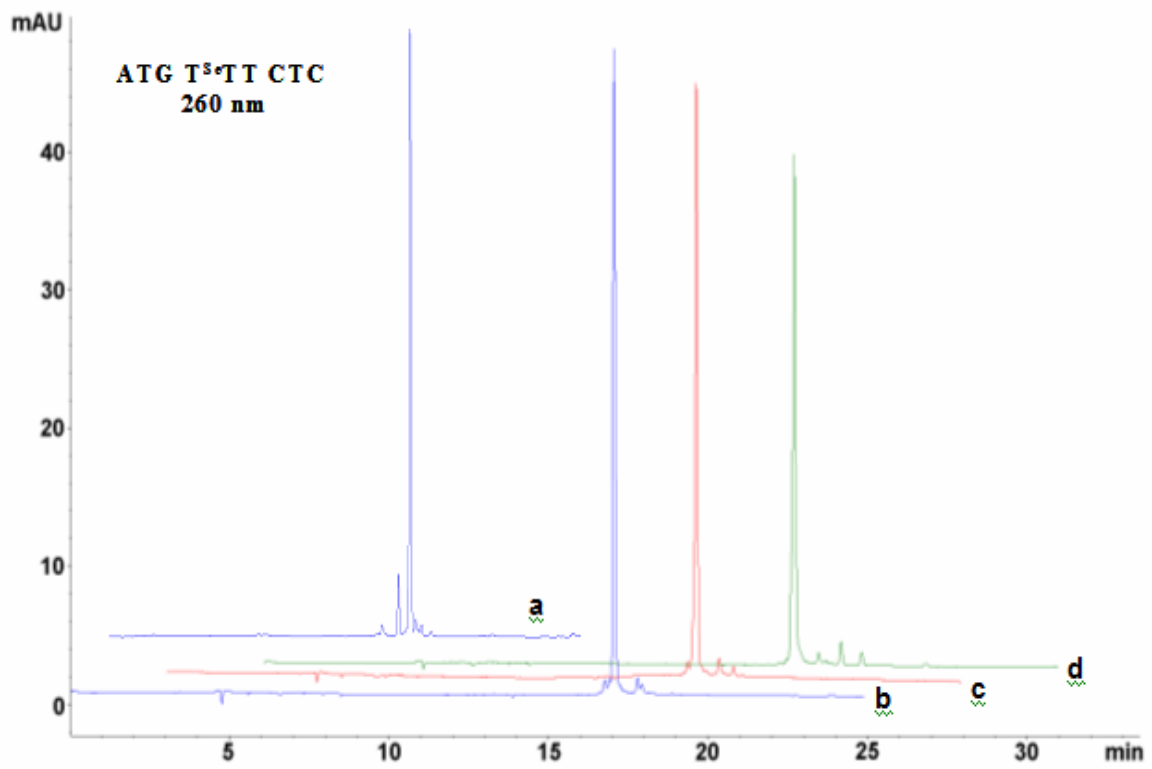


Figure 6.5: Stability of the SeNA DNA-9-mer (5'-ATGT-SeT-TCTC-3') in air. Samples were stored in buffer (5 mM, NaH2PO4-Na2HPO4, pH 7.5). a) 0 day, monitored at 260 nm by HPLC; b) 0 days; monitored at 368 nm by HPLC; c) 3 days, monitored at 368 nm by HPLC; d) 7 days, monitored at 368 nm by HPLC.

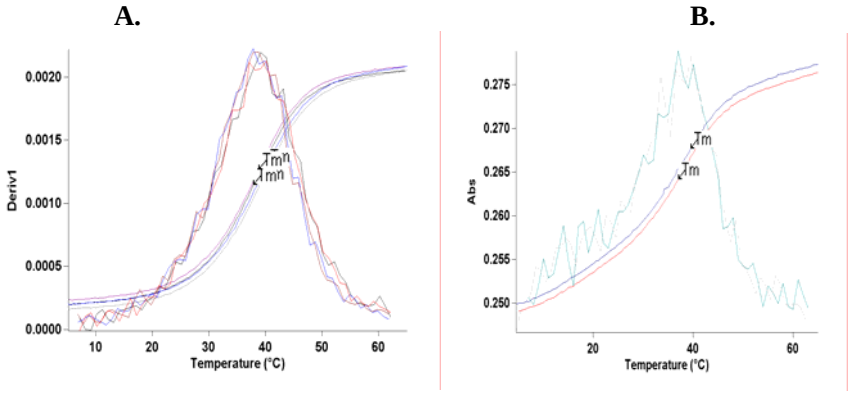


Figure 6.6: UV melting curves of native and Se-DNAs. (A) The native DNA duplex: 5'-ATGGTGCTC-3' & 5'-GAGCACCAT-3' (Tm= 39.2 ºC). (B) The Se-DNA duplex: 5'-ATGG-SeT-GCTC-3' & 5'-GAGCACCAT-3' (Tm= 38.6 ºC).

The data collection, phasing and refinement statistics of the determined Se-DNA structure (2NSK) are listed here in the Table 6.2. The positions of the selenium were identified by CNS heavy atoms search function (crystallography & NMR structure). Following the refinement of the selenium positions, the experimental phases were calculated and extended using the SAD or MAD (single- or multiple- wavelength anomalous dispersion) procedures in CNS. The initial phased maps were then improved by the solvent flattening and density modification procedure. The model of single strand of DNA was manually built into the electron density map using program O, and then the following refinement was carried out. The refinement protocol includes simulated annealing, positional refinement, restrained B-factor refinement, and bulk solvent correction. The stereo-chemical topology and geometrical restrain parameters of DNA/RNA, as well as the water, ions parameters from CNS default files have been applied for the normal DNA residues. The topologies and parameters for Se-modified dU (UMS) and dT (T4S) were built based on the parameters of selenium and carbon bonds and angles, then applied in the refinement. After several cycles of refinement and the model rebuilding, a number of highly ordered waters were picked up and added to the structure before the final runs of refinement. Finally, the occupancies of selenium were adjusted. A Cross-validation with a 5-10% test set was monitored during each refinement. The σA-weighted maps of the (2m|Fo| - D|Fc|) and the difference (m|Fo| - D|Fc|) density maps were used in model building. The final R value is blow 20% (19.5%) and the R free is 21.2%. The helical parameters of 4-Se-T DNA (2NSK) and the comparison with the native DNA structure (1DNS) are compiled using CURVES version 5.3 and listed in Table 6.3.

Table 6.2: X-ray data collection, phasing, refinement and model statistics for the Se DNA (5'-G-dU_{Se}-G-^{Se}T-A-C-A-C-3', 2NSK).

Data collection and phasing	peak	inflection	remote	reference
Wavelength, Å	0.9795	0.9798	0.9400	1.100
Resolution range, Å	40.0-1.5	40.0-1.5	40.0-1.5	40.0-1.5
(last shell)	(1.55-1.5)	(1.55-1.5)	(1.55-1.5)	(1.55-1.5)
Unique reflections	3736	3741	3743	3718
	(357)	(352)	(354)	(352)
Completeness, %	99.1	99.2	99.2	98.3
	(100.0)	(100.0)	(100.0)	(100.0)
R_{merge} ,%	8.8 (27.6)	6.3 (26.8)	6.3 (27.9)	4.8 (27.2)
$<I/\sigma(I)>$	15.3 (6.8)	14.9 (6.3)	14.6 (5.9)	14.7 (5.6)
Redundancy	13.0	13.1	13.2	12.7
	(13.6)	(13.6)	(13.7)	(12.2)
R-Cullis (Friedel)	0.485	0.371	(0.428)	
	(0.306)	(0.279)		
Phasing power (Friedel)	2.31	4.37	(4.79)	
	(4.98)	(6.24)		
Figure of merit (Friedel)	0.473	0.578	(0.438)	
	(0.548)	(0.584)		
Overall figure of merit		0.882		

Refinement and model	
Space group	P 4_3 2_1 2
Unit Cell	a=b= 42.31 Å, c= 23.84 Å
Resolution range, Å (last shell)	15.82 - 1.50 (1.59 - 1.50)
Number of reflections	3531 (492)
R_{work} , %	19.5 (19.5)
R_{free} , %	21.2 (19.9)
Nucleic Acid (single strand)	160
Heavy atom	2 (Se)
Water	36
Bond length R.M.S.D, Å	0.011
Bond angle R.M.S.D	1.8
Average B-factors, $Å^2$	
All atoms	24.0
Wilson plot	17.4
Overall anisotropic B-values	
B11/B22/B33	-1.51/-1.51/3.02
Solvent density, e/$Å^3$	0.34
B-factors of solvent, $Å^2$	47.6
Esd. from Luzzatt plot, Å	0.16
Esd. from SIGMAA, Å	0.08

Table 6.3: Comparison of helical parameters of 4-Se-T DNA (2NSK) and native (1DNS).

	Se-DNA (2NSK)	Native DNA (1DNS)
Mean Twist (^{o})	32.4	32.6
Mean Rise (Å)	2.7	2.9
Roll (^{o})	1.4	0.7
Inclination (^{o})	12.7	11.7
X-displacement (Å)	-4.3	-4.3
Sugar pucker	C2'-exo	C2'-exo
Minor groove		
Mean width (Å)	9.5	9.7
Mean depth (Å)	1.5	1.4
Mean helical diameter (Å)	19.3	19.2

The final solved 4-Se-T DNA structure was superimposable over the native structure (2.0 Å resolution) with the same tetragonal space group P43212, as showed in Figure 6.7A, which indicates that the oxygen replacement by selenium at this position does not cause the significant structure perturbation. Interestingly, the large Se atom is observed to be accommodated by a slight shift (Fig. 6.7B), revealing the flexibility of the DNA duplex structure. Furthermore, the atomic distance (3.02 Å) between the thymine N3 and the adenine N1 indicates a hydrogen bond formation (Fig. 6.7C). In addition, considering that the Se atomic radius is 0.43 Å larger than that of O, and that the distance between T4 exo-O4 and A5 exo-N6 in the native structure is 2.87 Å, the distance (3.35 Å) between the exo-Se4 and exo-N6 indicates a Se-mediated longer hydrogen bond formation. The observation of the longer Se-hydrogen bond (usual hydrogen bond length: 2.8-3.2 Å) in the crystal structure is consistent with the 4-Se-DNA UV-melting study which has a little bit decreased Tm.

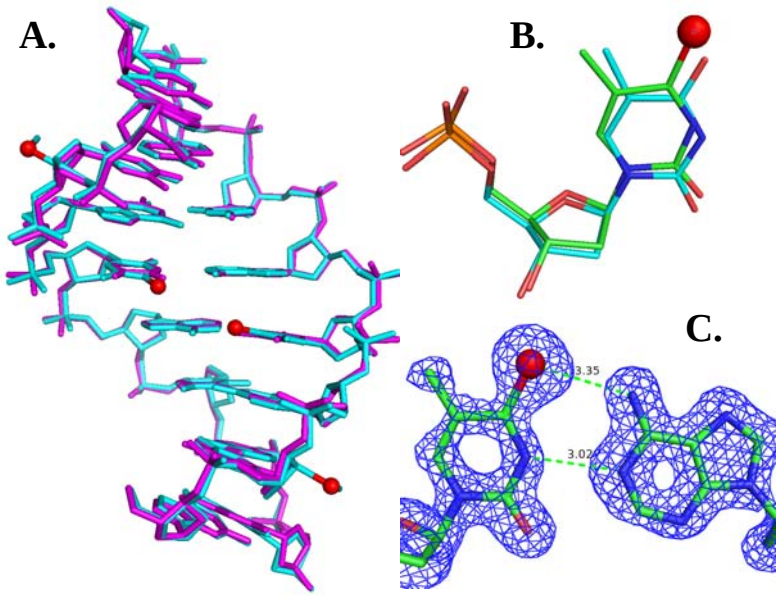


Figure 6.7: The global and local structures of the 4-Se-T DNA [(5'-G-dU_{Se}-G-^{Se}T-ACAC-3')$_2$]. (A) The duplex structure of the modified DNA (2NSK, in cyan) is superimposed over the native (1DNS, in pink). (B) The comparison of the modified (in green) and native (in cyan) local T4 structures. (C) The Se-base pair of T4-A5 with the experimental electron density.

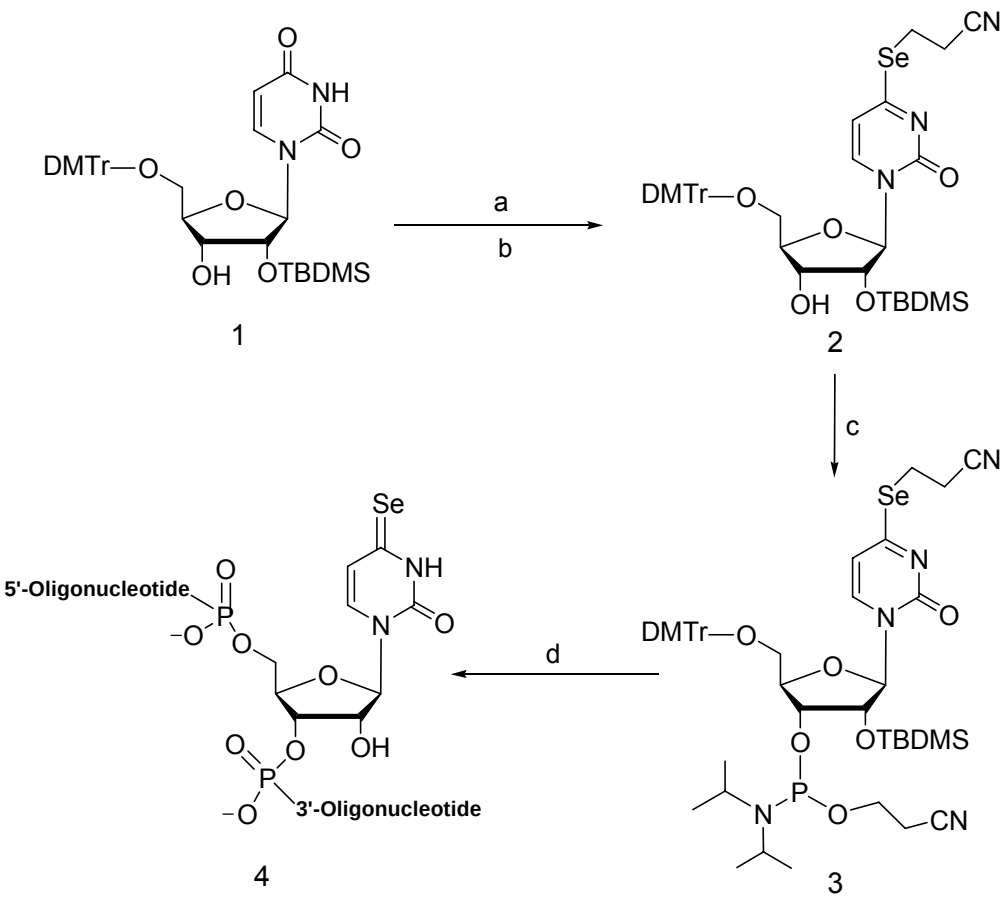
6.3 4-Se-uridine containing RNA

6.3.1 Synthesis of 4-Se-uridine and the phorsphoramidite

Following the above stated work, the selenium replacing oxygen at the thymidine 4-position have provided an additional useful methodology for DNA base pair and duplex stability studies besides the sulfur modification on the nucleobases. It was also clearly revealed by crystallography studies that DNA is flexible enough to accommodate a large atom and the selenium atom can also mediate the hydrogen bond formation with adenosine while not causing obvious structure perturbation. Furthermore, it was found in our 4-Se-thymidine work that this atom-specific mutagenesis could generate fairly stable yellow colored DNAs with a special UV-absorption, which could have great

potentials to visualize nucleic acids molecule. Therefore, to further explore this novel strategy, it's very necessary and interesting to explore the selenium modified RNA version in the nucleobase for RNA and RNA-protein complex structural and functional studies. It's being more urgent to develop this kind of novel tool in nowadays since the small RNA has been widely investigated as RNA interference and gene regulators. (119) Here, we also would like to report the first synthesis of the 4-Se-uridine phosphoramidite, its incorporation into oligonucleotides, thermostability and the ongoing crystal structure studies of the RNAs containing 4-Se-uridine modification.

As showed in scheme 6.2, our synthesis started from the partially protected 2'-TBDMS-5'-trityl-uridine 1 considering the following compatibility of RNA solid phase synthesis. Considering the presence of bulky TBDMS group at the 2' positioin, we simplified here the synthesis route by using a relatively more chemoselective TIBS (2,4,6-Triisopropylbenzenesulfonyl) group to activate position 4, in stead of the previous trizaole group, thus avoiding the protection and deprotection steps for the 3'-hydroxyl group. Without purifying this activated intermediate, the selenium functionality could be introduced directly by the substitution of TIBS group at position 4 with sodium 2-cyanoethylselenide in good yield, which was generated by the reduction of di-(2-cyanoethyl) diselenide with $NaBH_4$ in ethanol. This selenium functionality has been proved to be well compatible with solid phase synthesis and can be removed by week base treatment. Finally, the Se-uridine derivative 2 was converted to the corresponding phosphoramidite 3 in a satisfied yield, although the reaction took much longer time comparing the thymidine due to the TBDMS group. The experimental detail and characterization of these two compounds are listed as following.



Scheme 6.2: Concise Synthesis of 4-Se-Uridine Phosphoramidite (3) and 4-Se-U-RNAs. Reagents and conditions: (a) TIBSCl, DMAP, CH_2Cl_2, r.t; (b) $(NCCH_2CH_2Se)_2/NaBH_4$, EtOH, r.t; (c) 2-cyanoethyl N,N-diisopropylchlorophosphoramidite and N,N-diisopropylethylamine in CH_2Cl_2, 0 °C then r.t; (d) the solid phase synthesis. TIBSCl: 2,4,6-(Triisopropylbenzene)sulfonyl chloride.

4-(2-cyanoethyl)seleno-5'-O-(4,4'-dimethoxytriphenylmethyl) thymidine (2).

To a dry dichloromethane solution of the starting material compound 1 and 4,4'-dimethylamino-pyridine (DMAP) under argon, was dropwisely added the dichloromethane solution of 2,4,6-Trisopropyl-Benzenessulfonyl chloride, followed by several drops of triethyl amine. The system was stirred for one hour before this reaction is finished (monitor by TLC, 5% methanol in dichloromethane). At the same time, the

$NaBH_4$ suspension (250 mg $NaBH_4$ in 5 mL of EtOH) was injected into a flask containing di(2-cyanoethyl) diselenide [$(NCCH_2CH_2Se)_2$, 0.45 mL, d= 1.8 g/mL, 3.0 mmol] and ethanol (8 mL) on an ice bath and under argon. The yellow color of the diselenide disappeared in approximately 15 min, giving an almost colorless solution of sodium selenide ($NCCH_2CH_2SeNa$). Then, the solution of compound **1** (1.34 g, 2 mmol) was slowly injected into this solution. After the selenium incorporation was completed in 45 min, (monitored on TLC, 5% MeOH in CH_2Cl_2, product R_f = 0.68), water (100 mL) was added to the reaction flask. The solution was adjusted to pH 7-8 using CH_3COOH (10%), and was then extracted with ethyl acetate (3 x 100 mL). The organic phases were combined, washed with NaCl (sat., 100 mL), dried over $MgSO_4$ (s) for 30 min., and evaporated to minimum volume under reduced pressure. The crude product was then dissolved in methylene chloride (5 mL) and purified on a silica gel column equilibrated with hexanes/methylene chloride (1:1). The column was eluded with a gradient of methylene chloride (CH_2Cl_2, 0.5%, 1%, and 2% MeOH in CH_2Cl_2, 500 mL each). After solvent evaporation and dry on high vacuum, the pure compound (**2**) was obtained as a slightly yellow foamy product (1.07 g, 81% yield). ^{1}H-NMR (400 MHz, $CDCl_3$) δ: 0.21 (s, 3H, CH_3), 0.38 (s, 3H, CH_3), 0.95 (s, 6H, 2x CH_3), 2.31-2.37 (m, 1H, H-2'), 3.00 (dd, *J* = 6.5 and 6.7 Hz, 2H, CH_2-Se), 3.37-3.41 (m, 2H, CH_2-CN), 3.50-3.52 (m, 2H, 1H-5'), 3.81 (s, 6H, 2x OCH_3), 4.17-4.22 (m, 1H, H-3'), 4.31 (s, 1H, 3'-OH), 4.40-4.50 (m, 1H, H-4'), 5.78 (s, 1H, H-1'), 5.90 (d, 1H, *J*=6.8 Hz, H-5), 6.8-6.90 (m, 4H, aromatic), 7.20-7.46 (m, 9H, aromatic), 8.31 (d, 1H, *J*=6.8 Hz, H-6). ^{13}C-NMR (100 MHz, $CDCl_3$) δ: -4.30, -4.40 (CH_3), 18.09 ($\underline{C}H_2$-CN), 19.04 ($\underline{C}H_2$-CH_2-CN), 20.45 (($CH_3)_2\underline{C}$(t-Bu)), 25.87 (CH_3), 55.30 (OCH_3), 68.67 (C-3'), 76.42 (C-2'), 83.14 (C-4'), 91.03 (C-1'), 106.0 (C-5), 118.8 (CN),

113.3, 127.1, 128.0, 128.2, 130.1, 135.0, 135.3, 144.2, 158.73 (Ar-C), 140.4 (C-6), 153.3 (C-2), 175.0 (C-4). HRMS (MALDI-TOF): molecular formula, $C_{39}H_{46}N_3O_7PSeSi$; $[M+H]^+$: 776.2267 (calc. 776.2270).

4-(2-cyanoethyl)seleno-5'-O-(4,4'-dimethoxytriphenylmethyl)-thymidine3'-O-(2-cyanoethyl) diisopropylamino phosphoramidite (3).

To the flask (25 mL) containing **2** (453 mg, 0.68 mmol) under argon, dry methylene chloride (2.5 mL), *N*,*N*-diisopropylethylamine (0.17 mL, 1.03 mmol, 1.5 eq.), and 2-cyanoethyl *N*,*N*-diisopropyl-chlorophosphoramidite (195 mg, 0.83 mmol, 1.2 eq.) were added sequentially. The reaction mixture was stirred at -10 °C in an ice-salt bath under argon for 10 minutes, followed by removal of the bath. The reaction was completed in 45 minutes at room temperature (indicated by TLC, 5% MeOH in CH_2Cl_2, product R_f = 0.63 and 0.68), generating a mixture of two diastereisomers. The reaction was then quenched with $NaHCO_3$ (2 mL, sat.), stirred for 5 min, and the product was then extracted with CH_2Cl_2 (3 x 5 mL). The combined organic layer was washed with NaCl (10 mL, sat.) and dried over $MgSO_4$ (s) for 20 min, followed by filtration. The solvent was then evaporated under reduced pressure and the crude product was re-dissolved in CH_2Cl_2 (2 mL). This solution was added drop-wise to petroleum ether (100 mL) under vigorous stirring, generating a white precipitate. The petroleum ether solution was decanted. The crude product was re-dissolved again in CH_2Cl_2 (2 mL) and then loaded into an Al_2O_3 column (neutral) that was equilibrated with CH_2Cl_2/Hexanes (1:1). The column was eluded with a gradient of methylene chloride and ethyl acetate [CH_2Cl_2 to CH_2Cl_2/EtOAc (7:3)]. After solvent evaporation and dry over high vacuum, the pure **3** (538 mg) was obtained as a white foamy product (92% yield). ^{1}H-NMR (400 MHz,

$CDCl_3$, two sets of signals from a mixture of two diastereomers) δ: 0.2-0.4 (m,12H, 4 x CH_3), 0.85-1.20 (m, 36H, 8x CH_3-ipr and 4x Si(CH_3)), 2.30-2.38 and 2.70-2.82 (2x m, 4H, 2x H-2'), 2.34 and 2.64 (2x t, *J* = 6.4 Hz, 4H, 2x O-CH_2-$\underline{CH_2}$-CN), 3.00-3.04 (m, 4H, 2x Se-CH_2-$\underline{CH_2}$-CN), 3.32-3.44 (m, 6H, 2x H-5', 2x Se-CH_2), 3.52-3.64 (m, 8H, 4x CH-ipr, 2x O-$\underline{CH_2}$-CH_2-CN), 3.73-3.84 (m, 2H, 2x H-5'), 3.82 and 3.83 (2x s, 12H, 4x OCH_3), 4.12-4.35 (m, 2H, 2x H-3'), 4.43-4.48 (m, 2H, 2x H-4'), 5.70-5.90 (m, 4H, 2x H-5 and 2x H-1'), 6.83-6.88 (m, 8H, aromatic), 7.27-7.43 (m, 18H, aromatic), 8.30 and 8.39 (2x s, 2H, 2x H-6). HRMS (MALDI-TOF): molecular formula, $C_{48}H_{64}N_5O_8PSeSi$; $[M+H]^+$: 978.3512 (calc. 978.3505.).

6.3.2 Preparation of RNA oligonucleotides containing 4-Se-U

Then this phosphoramidite was examined on the RNA solid-phase synthesis, which usually is less facile and need more coupling time than DNA synthesis due to the additional sterically protected 2'-hydroxyl group. In this work, the ultramild CE phosphoramidites protected with widely used (tert-butyl)-dimethylsilyl (TBDMS) group were selected due to the instability of selenium functionality under strong base cleavage, as well as 5-benzylmercapto-1H-tetrazole as the activating reagent. As expected, this modified residue is compatible with the long coupling reaction (12 min), I_2 oxidation and trichloroacetic acid treatment to remove 5'-DMTr group, without causing deselenization. When the oligonucleotides contain many guanosine residues, phenoxyacetic anhydride (Pac_2O), instead of acetic anhydride is used in the capping step to avoid the G acetylation, which is difficult to remove under mild K_2CO_3 condition. All the RNA oligonucleotides were synthesized as DMTr-on form. After synthesis, they were cleaved

and partially deprotected by 0.05 M K_2CO_3 solution in methanol for 8h. The solution was neutralized with 1M HCl and evaporated to dryness, then all the 2'-TBDMS groups were removed by 1M TBAF solution in THF for overnight. After desalting using Sephadex-G25, the pure RNA oligoes (MALDI-MS listed in Table 6.4) were obtained by RP-HPLC purification, followed by the detritylation and precipitation. A typical MS analysis is shown in Figure 6.8. Similar as 4-Se-Thymidine, these base modified RNAs also present yellow color with the same UV absorption at 368 nm.

Table 6.4: MALDI-MS analysis of RNAs containg 4-Se-U.

Entry	Selenium oligonucleotides	Measured (calcd) m/z
1	5'-G$\underline{U}^{(4\text{-Se})}$GUACAC-3'	$[M+H]^+$: 2573.3
2	5'-GUG$\underline{U}^{(4\text{-Se})}$ACAC-3'	$[M+H]^+$: 2573.3
3	5'-CGCGAA$\underline{U}^{(4\text{-Se})}$UCGCG-3' C114H143N46O81P11Se: FW 3873.3	$[M+H]^+$: 3873.3 (3874.3)
4	5'-CGCGAAU$\underline{U}^{(4\text{-Se})}$CGCG-3' C114H143N46O81P11Se: FW 3873.3	$[M+H]^+$: 3873.3 (3874.3)
5	5'-U$\underline{U}^{(4\text{-Se})}$AUAUAUAUAUAA-3' C133H162N49O95P13Se: FW 4448.6	$[M+H]^+$: 4449.7 (4449.6)
6	5'-AA$\underline{U}^{(4\text{-Se})}A\underline{U}^{(2'\text{-Se})}$AUAUAUAUU-3' C134H164N49O94P13Se2: FW 4525.6	$[M+H]^+$: 4526.4 (4526.6)
7	5'-GG$\underline{U}^{(4\text{-Se})}$AUUGCGGUACC-3' C133H165N52O97P13Se: FW 4525.7	$[M+H]^+$: 4526.4 (4526.7)

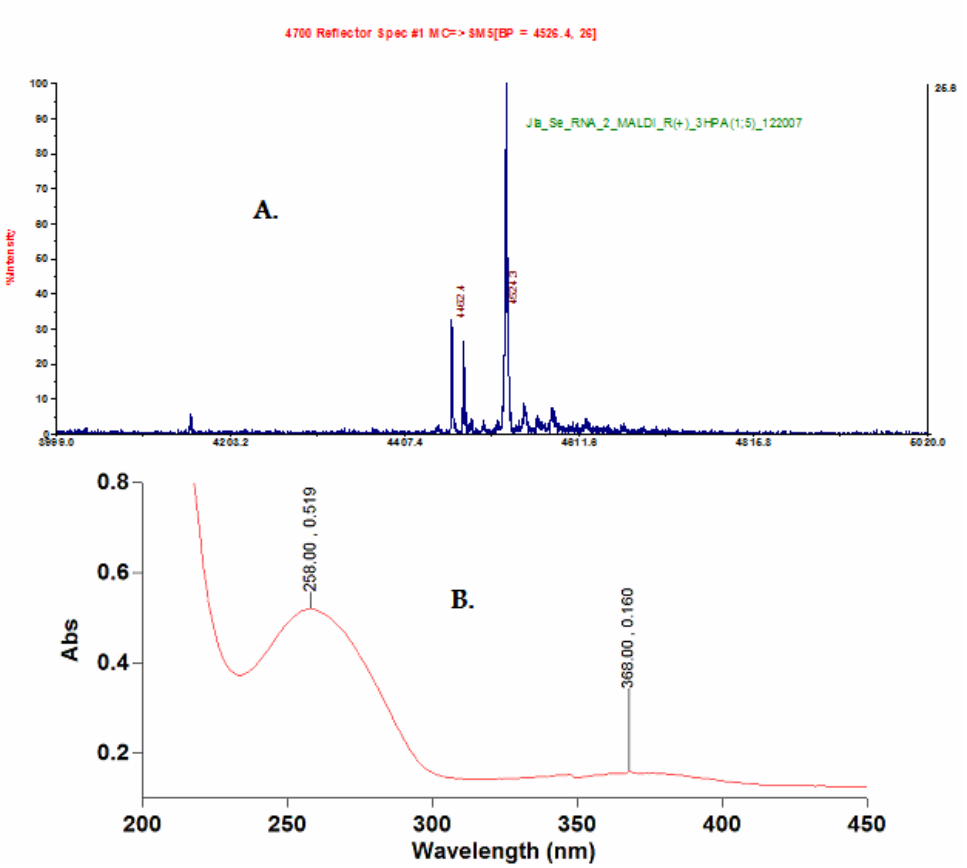
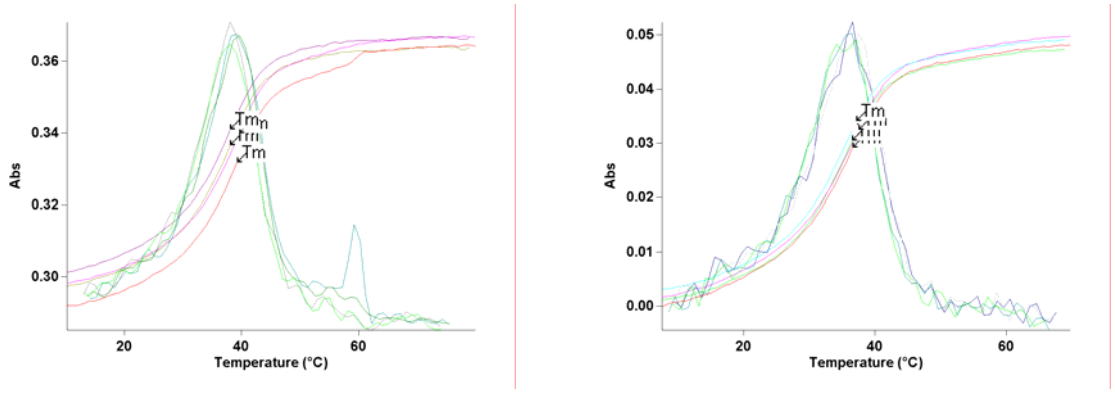


Figure 6.8: MALDI-TOF mass spectrum of 4-Se-U-RNA and its UV-spectrum. 5'-rGG U(4-Se)AUUGCGGUACC-3', $C_{133}H_{165}N_{52}O_{97}P_{13}Se$, [M+H]+: 4526.4 (Calcul. 4526.7).

6.3.3 Thermal denaturalization study

Again, the thermal denaturalization study was also carried out using a RNA 14mer as model. Similar as 4-Se-T-DNA, comparing the native counterpart, there is no obvious difference was observed for the 4-Se-U containing RNA, as showed in figure 6.9, which indicates no obvious structure perturbation.



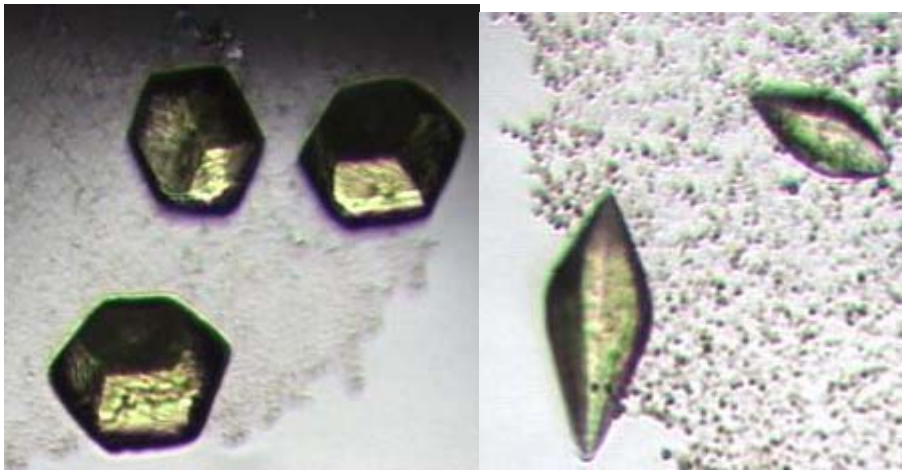
A: 5'-rUUAUAUAUAUAA-3'
Tm=38.0 ±0.6 ºC

B: 5'-rUUAU(4-Se)AUAUAUAA-3'
Tm=37.3 ±0.6 ºC

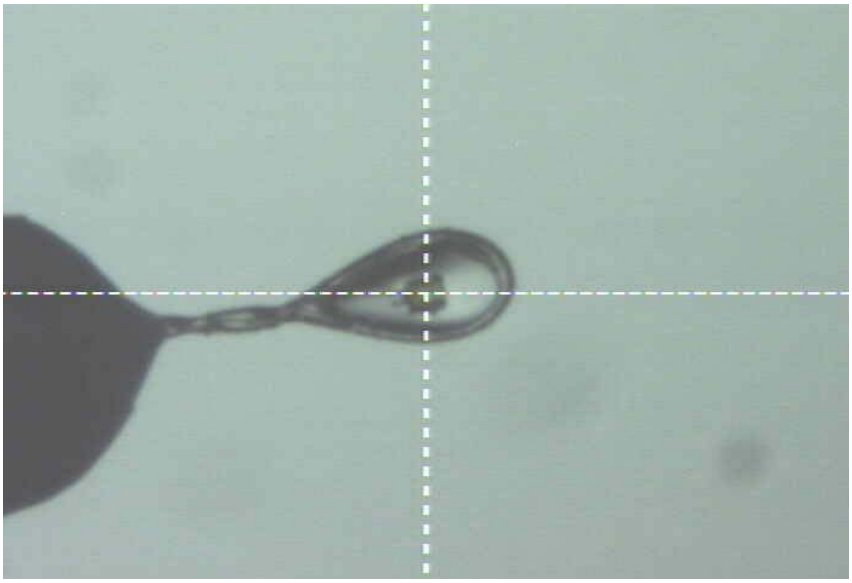
Figure 6.9: Tm studis of Se-U 14mer RNA comparing with native RNA.

6.3.4 Crystallization and structure studies

After the preliminary confirming of the structural isomorphism by Tm study with several RNA models, we have also successfully crystallized two RNA sequences containing this modified 4-Se-U residue (5'-rGUGU^{Se}ACAC-3' and 5'-rUUAU(4-Se)AUAUAUAUAA-3') through nucleic acids mini-screen conditions and Natrix buffer (also from Hampton Res.) within a week. Although the crystal qualities is not as perfect as the our previous ones (indicated by the rough surfaces of the RNA crystals comparing the shining and flatter ones), we still collected several data sets from beamline X-12B of brookhaven national laboratory with acceptable resolution (from 2.5 to 2.9 Å), the typical crystal pictures were showed in figure 6.10. Although the structure determination work hasn't been done till the writing of this dissertation, the obvious yellow color of these types of crystal gives a signal that it can be potentially used as a probe of detecting RNA-protein complex.



(A) (B)



(C)

Figure 6.10: Crystals photos of 4-Se-U containing RNAs (A, B) and the X-ray exposure of crystal (C).

6.4 Conclusion of this chapter

In summary, we have synthesized the 4-Se thymidine and 4-Se uridine phosphoramidite, and incorporated them into DNA and RNA oligonucleotides with

quantitative yields. The Se substitution on the nucleobase is relatively stable under the elevated temperature. By the UV-melting and crystal structure studies, we have further demonstrated that the O substitution with Se on the nucleobase does not cause the significant duplex structure perturbation (the RNA case is still to be confirmed). The crystal structure study also reveals the accommodation of the large Se atom on the thymine by the DNA duplex and the formation of the Se-mediated hydrogen bond within the T-A base pair. This work will stimulate the studies on the Se substitution on other nucleobases and positions, opening a new avenue for further exploring the base-paring selectivity governed by the hydrogen bond and base shape, the duplex structure and flexibility, the DNA replication efficiency and fidelity, and the polymerase recognition of the modified nucleobases, as well as the RNA structural and functional research. Moreover, the Se derivatization on the nucleobases will largely facilitate X-ray crystal screen and structure determination of protein-nucleic acids complexes.

7. SYNTHESIS OF 2'-TELLURIUM MODIFIED URIDINE AND THYMIDINE FOR NUCLEIC ACID STRUCTURE AND FUNCTION STUDIES

7.1 Introduction

The increasing application of selenium as anormalous scattering center in both protein and nucleic acids X-ray crytallogarphy studies have also gained our attention to its downstair neighbour of element periodic table: Tellurium.

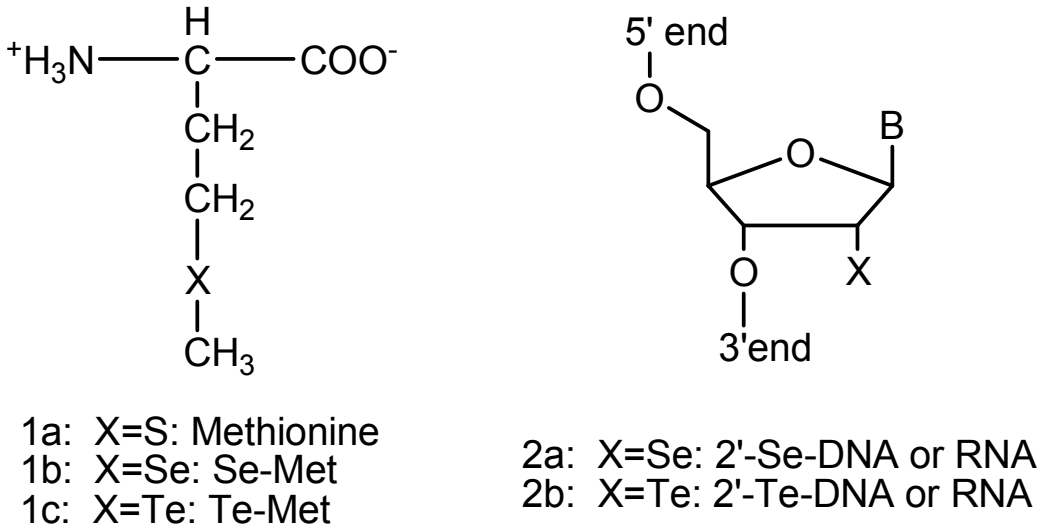
As a chalcogen element, tellurium follows sulfur and selenium in elemental group VI of the periodic table. Tellurium has a larger covalent radius (1.35 Å), compared with selenium (1.17 Å), sulfur (1.02 Å) and oxygen (0.73 Å) (their van der Waals radius is 2.06, 1.90, 1.80 and 1.52 Å respectively). Probably due to its matallic property and weak covalent bonds with carbon and hydrogen, so far, no tellurium functionality has been discovered in any natural biological molecules. Usually, it's used in alloys with other metals to improve the strength and durability, and to decrease the corrosive action of acids; tellurium is also a p-type semiconductor that shows a greater conductivity in certain directions which depends on atomic alignment, so, it has a wider application in electronic and optical industry. (120)

In 1989, tellurium was successfully incorporated into proteins through Te-resistant fungi grown in the presence of tellurite on a sulfur-free medium. Later, the Te-methionine derivatization for protein structure determination was investigated and developed as a similar exploration of Se-Met strategy. In principle, the electronic and atomic properties of tellurium are ideal to generate clear signals in the isomorphous and anomalous difference Patterson maps using just the common in-house X-ray diffraction

facility (Cu, Kα wavelength: 1.54 Å). Therefore, a synchrotron facility is not necessary to collect X-ray diffraction data of the Te-derivatized protein crystals. This Te-derivatization strategy is a promising alternative to the conventional multiple isomorphous replacements (MIR) by the soaking procedures. With the guidance of this direction, Boles and co-workers reported in 1994 the expression of telluromethionine (TeMet) dihydrofolate reductase, and the Te-incorporation technique was further optimized to efficiently introduce TeMet into several proteins. Furthermore, the Te-protein stability under the X-ray exposure and the isomorphism of Te-proteins were also confirmed. At the same time, more and more tellurium chemistry has also been developed.

However, the facile oxidation of telluromethionine was also found to present a serious drawback for the further use of this methionine analogue for bioexpression of heavy atom variants of proteins, since oxidized forms were difficult to refold from inclusion bodies or to crystallize. In particular, surface-located and, in a more pronounced manner, solvent-exposed Tem residues were found to oxidize readily and thus to generate electron densities of difficult assignment and local non-isomorphism. Nonetheless, telluromethionine represents a heavy atom Met analogue well suited for isomorphous replacements in proteins. Although tellurium has a lower electronegativity and larger coordination-sphere radius than sulfur and selenium and the C–Te bonds are larger than Se–C and S–C (2.4 Å comparing with 1.95 and 1.82 Å), the plasticity of proteins still could allow for accommodation of the tellurium atoms without significantly affecting the native structures. In terms of structure stability in solution, it was found that the per-telluromethionine-variant of annexin V was less stable than the related wild-type protein and the selenomethionine-variant.

Comparing the unstability of TeMet in protein, it's possible to make the tellurium centers more stable in nucleic acids, and our success in selenium-derivatization of nucleic acids also naturally encourages us to take advantage of tellurium and develop tellurium-derivatized nucleic acids (TeNA) for structure and function studies. We initiated our projects aiming at overcoming the tellurium stability issues and their compatibility with nucleic acid solid phase synthesis. As the proof of principle, this chapter will report, the first synthesis of 2'-Te derivatized uridine, thymidine, their according phosphoramidites and oligonucleotides containing tellurium functionalities (Figure 7.1).



1a: X=S: Methionine
1b: X=Se: Se-Met
1c: X=Te: Te-Met

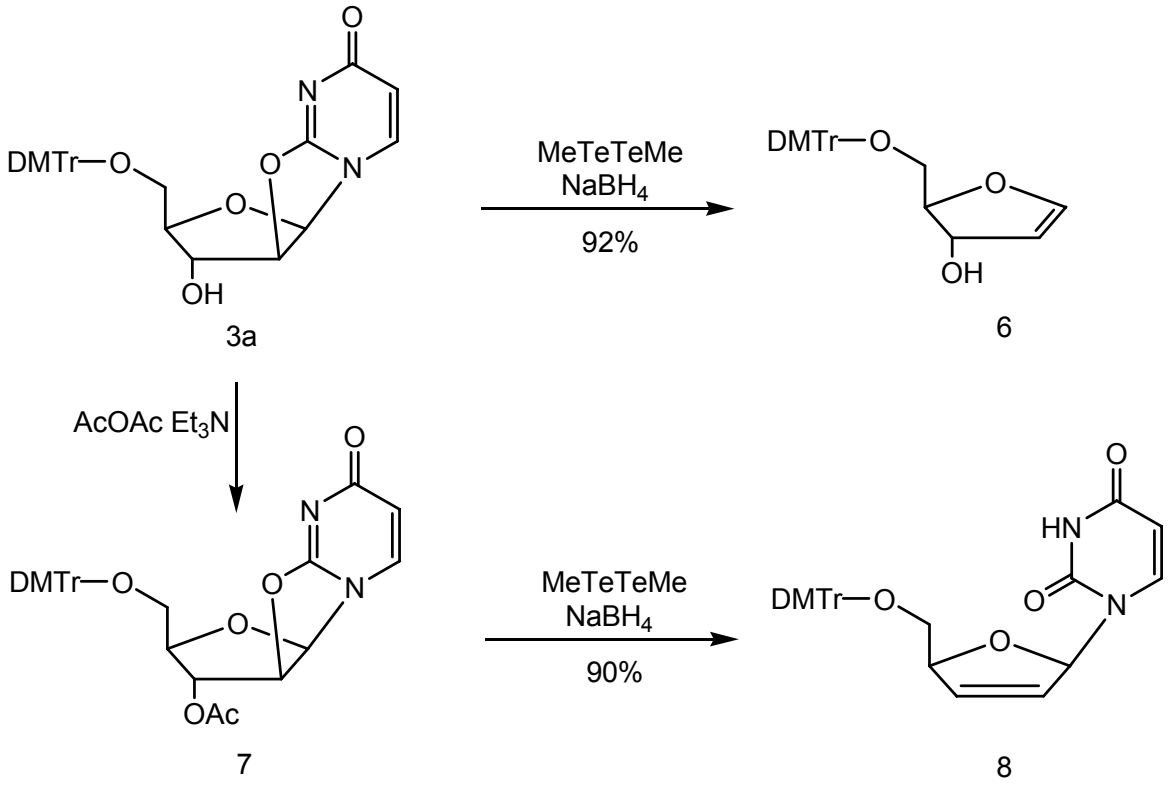
2a: X=Se: 2'-Se-DNA or RNA
2b: X=Te: 2'-Te-DNA or RNA

Figure 7.1: Selenium and tellurium derivatized methionine and oligonucleotide.

7.2 Novel tellurium induced addition-elimination for anti-HIV drug d4T synthesis

Our previous studies of the nucleic acids derivatized with selenium at 2'-position of uridine and thymidine have revealed that the 2'-Se-modification does not cause

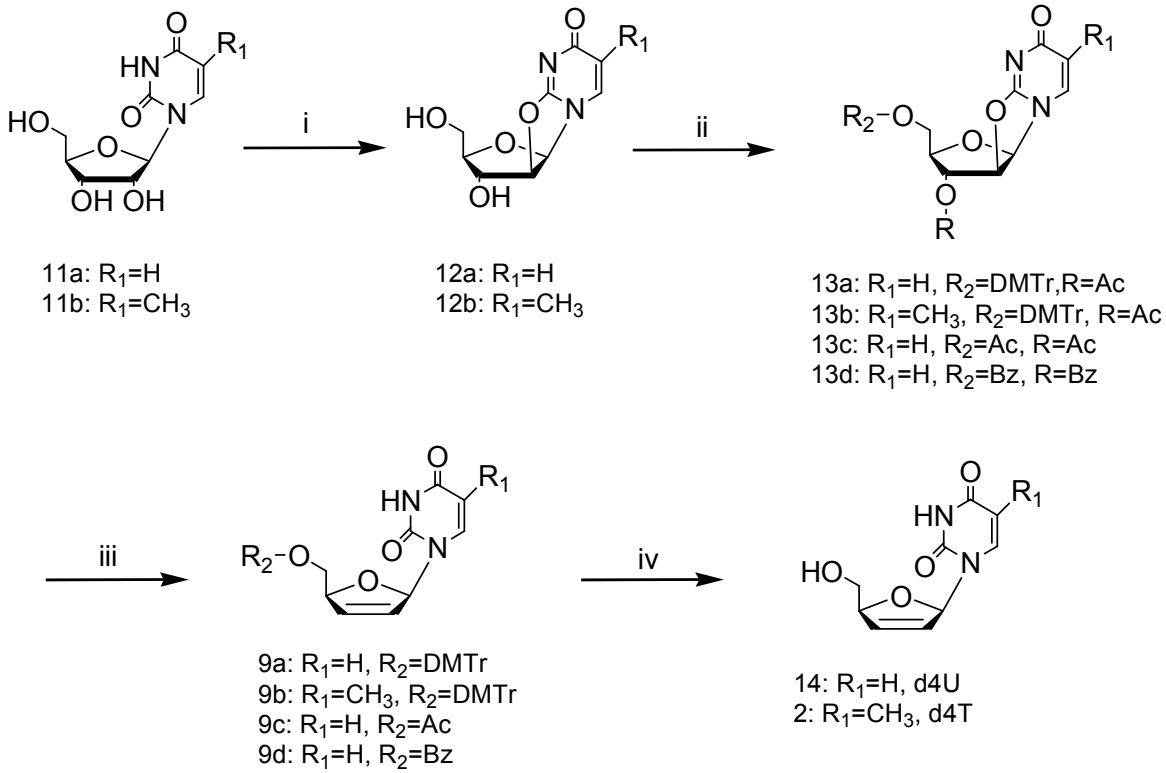
significant perturbation of DNA and RNA structures. Thus, the 2' position became our starting point to synthesize Te-functionalized DNA. As showed in Scheme 7.1, the 5'-DMTr-2,2'-anhydrouridine was chosen as the starting materials which were synthesized previously. Similar as selenium work, the tellurium functionalities were designed to be introduced by a SN_2 reaction of methyl telluride ion, which was generated through the reduction of dimethyl ditelluride by sodium borohydride. But to our surprise, instead of the target compound, there was only the base elimination product **6** formed. Then, the acetylation step was carried out to protect the 3'-hydroxyl group (compound **7**), as a result, 5'-DMTr-d4U (**8**) was the major product in dimethyl ditelluride experiment, which could be further detritylated to d4U, an analog of anti-HIV drug.



Scheme 7.1: Elimination reactions resulted from the presence of MeTeTeMe and $NaBH_4$.

Moreover, it was also found that catalytic amount (10%) of dimethyl ditelluride was enough to finish this elimination conversion, without which there was only hydrolyzed products observed. However, the amount of $NaBH_4$ seemed very important to the reaction, in absence of $NaBH_4$, there was no reaction happened, which meant that the effective part of the tellurium reagent was the reduced tellurium ion form; on the other hand, the more it was used, the more hydrolyzed compounds formed, thus the lower the yields were, probably due to the resulted strong basic solution. Finally, the same amount of sodium borohydride as dimethyl ditelluride (usually 0.5 equiv) was found a good condition to obtain both acceptable reaction rates and yields.

Inspired by that, several attempts were carried out to check the universe of this strategy. A general synthetic route was described in scheme 7.2, starting from uridine (11a) and ribothymidine (11b), which were cyclized to the according 2,2'-anhydronucleosides before the different protecting groups (including benzyl, trityl, acetyl and their combinations) were introduced into 3' and 5' positions to generate compounds 13a-f. The further treatment of dimethyl ditelluride and $NaBH_4$ in the mixed solution of THF with ethanol at 50 °C for 3-5 hours could generated the according 5'-protected intermediate 9a-d in moderate to good yields (69-89%) (Table 7.1), which gave the final compounds d4U and d4T by the following deprotection steps. Since the 2',3'-dideoxynucleosides (ddNs) and 2',3'-didehydro-2',3'-dideoxynucleosides (d4Ns) are FDA-approved anti-HIV drugs, our accidently discovered novel telluride-mediated high yield elimination reaction has potential to simplify the conventional synthesis of these compound and further reduce the cost of AIDS treatment and will also facilitate development of novel d4N and ddN analogs.



Scheme 7.2: The novel synthetic route of d4U and d4T with different derivatives. (i) diphenolcarbonate, DMF, 100 °C; 12a: 85%. 12b: 83%. (ii) 13a, 13b: Dimethoxytrityl chloride, acetic anhydride, Py, r.t. 87%. 13c: acetic anhydride, py, r.t. 89%. 13d: benzoyl chloride, py, 75%. (iii) dimethyl ditelluride, $NaBH_4$, THF:EtOH=7:1, 50 °C; 9a: 87%. 9b: 90%. 9c: 83%. 9d: 79%. (iv) Dowex 50 W for 9a, 9b, or NH_4OH, CH_3OH for 9c and 9d, 90-92%.

Table 7.1: Synthesis of d4U and d4T derivatives.

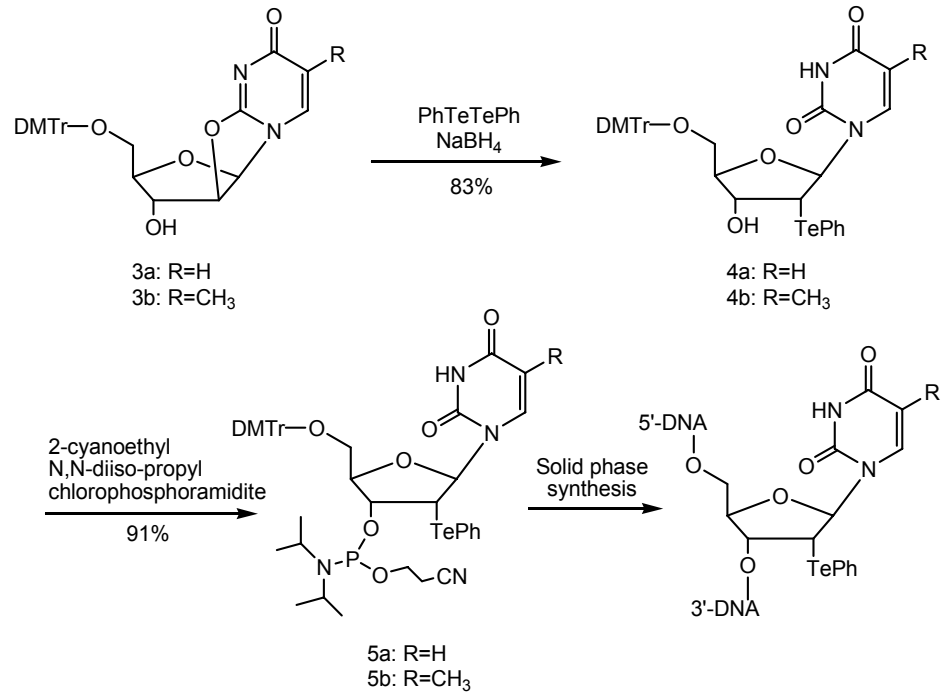
Substrates	R_1	R_2	R	Products	Yield (%)
13a	H	DMTr	Ac	9a	90
13b	CH_3	DMTr	Ac	9b	85
13c	H	Ac	Ac	9c	80
13d	H	Bz	Bz	9d	69
13e	H	DMTr	DMTr	None	None
13f	H	DMTr	MOM	None	None
13g	H	DMTr	H	10	92

* DMTr: dimethoxytrityl; MOM: methoxymethyl.

7.3 Preparation of 2'-Te functionalized nucleosides and DNAs

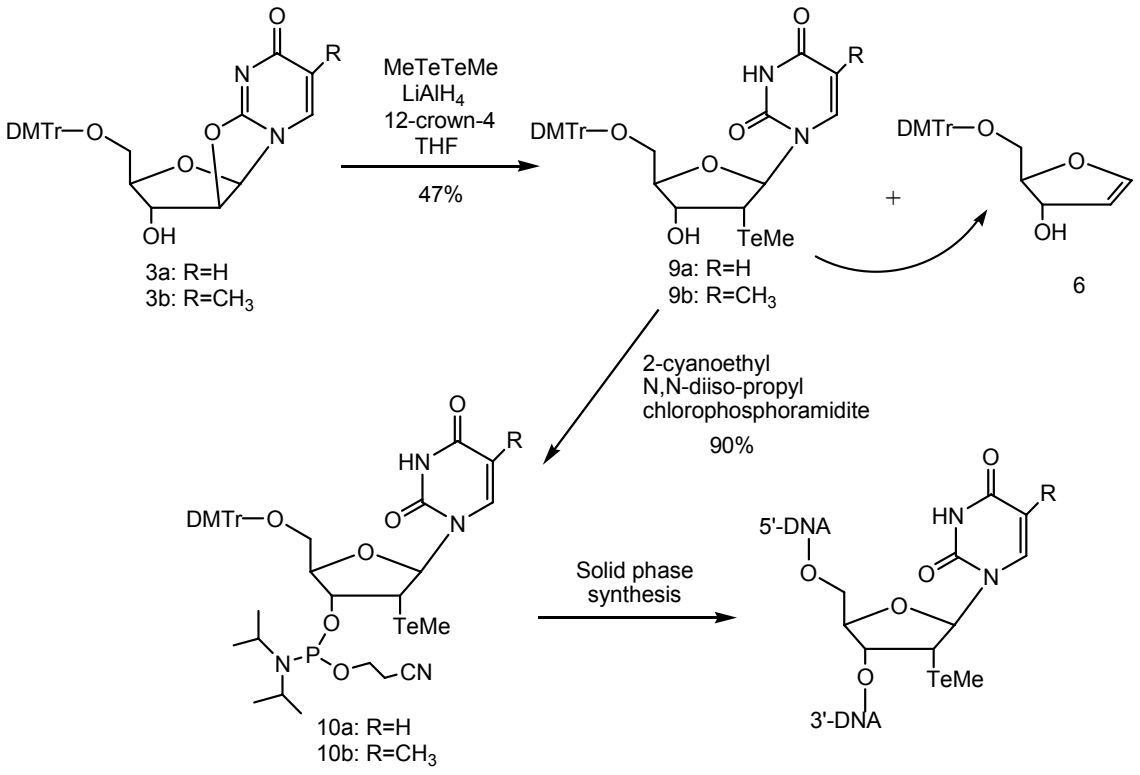
7.3.1 Synthesis of 2'-Te -U, T and the phosphoramidites

As stated above, due to the early synthetic challenges caused by the 1',2' and 2',3' elimination using dimethylditelluride, we decided turn our attention to pursue the Te-incorporation with diphenylditelluride first (Scheme 7.3), the electron withdrawing ability of which could also make the tellurium center more stable, preventing being oxidized. It turned out that the diphenylditelluride reagent is much easier to be reduced with sodium borohydride in dry THF at room temperature. The ring opening reaction underwent well at 50 °C with a a satisfied yield, indicating that the tellurium nucleophile center is still electron-rich enough to attack the 2'-position from the α face. The generated key intermediates 4a and 4b were successfully converted to the corresponding phosphoramidites 5a and 5b in high yields by following the standard protocol.



Scheme 7.3: Synthesis of 2'-TePh-5'-DMTr-uridine (5a), thymidine phosphoramidite.

At the same time, considering that the possible potential structure perturbation of the duplex structures caused by the bulky phenyl group at the minor groove, we still made more efforts targeting the dimethylditelluride after the success of phenyl one. As stated above, since it is difficult to reduce dimethylditelluride efficiently with certain amount of sodium borohydride even at an elevated temperature, which probably is due to the relatively higher instability of methyl telluride moity, a stronger reducing reagent $LiAlH_4$ was applied in this work (Scheme 7.4). However, only trace amount of the desired compound (9) was observed initially, while the base elimination compound was still the major by-product. In order to minimize its formation and enhance the methyltelluride reactivity, we decided to run the telluride incorporation at a lower temperature (0 °C) in presence of crown ether (12-crown-4) to chelate lithium ions. Fortunately but not sufficiently, we were able to obtain the desired product in 47% yield, while the 1',2'-eliminated product was still formed as the major by-product, we also played with the various temperature, concentration, solvents and the reaction time, but no further optimized conditions were successful identified so far. The characteristic chemical shift of the Te-methyl group was observed at -21.2 ppm in ^{13}C-NMR spectrum, which indicated the real incorporation of the methyl telluride group. Moreover, the tellurium isotope distributions in the HR-MS spectra also provided the supplemental proof of the TeMe existance. The isolated key intermediates 9a and 9b were then successfully converted to the corresponding phosphoramidites 10a and 10b in high yields by following the standard protocol and they were found to be stable to the basic conditions and not to disturb the reactivity of the 3'-hydroxyl group. All the detailed experimental information and compound identification are listed as following.



Scheme 7.4: Synthesis of 2'-TeMe-5'-DMTr-uridine (10a), thymidine phosphoramidites (10b) and the 2'-TeMe functionalized DNAs.

2'-deoxy-2'-phenyltellanyl-5'-O-(4,4'-dimethoxytrityl)-uridine (4a) and 2'-deoxy-2'-phenyltellanyl-5'-O-(4,4'-dimethoxytrityl)-5-methyluridine (4b):

To a stirred suspension of $NaBH_4$ (6.2 mg, 0.15 mmol) in anhydrous THF (5 mL), under argon at 0 °C, the THF solution of diphenylditelluride (0.2 g, 0.5 mmol in 5 mL) was added, followed by several drops of dry ethanol until bubble formed and the solution turned colorless. To this solution the starting material **3a** and **3b** (0.285 g and 0.292g separately, 0.5 mmol, dissolved in 5 mL of THF) was added and the reaction was slowly warmed up to room temperature and then into 50 ºC for three hours, which monitored by TLC. The solvent was then evaporated. The residue was dissolved in CH_2Cl_2 (20 mL) and washed with water (3 x 20 mL). The CH_2Cl_2 solution was dried ($MgSO_4$),

concentrated and the crude product was purified by silica gel column chromatography with 0-3% methanol in CH_2Cl_2 to give compound **4a** and **4b** as slight yellow solid (**4a**: 310 mg, 80% yield, **4b**: 292mg, 78% yield). **4a**: ^{1}H NMR (400 MHz, $CDCl_3$, 25 ºC): *δ*=3.45-3.46 (m, 2H), 3.82 (s, 3H), 3.92-3.95 (m, 1H), 4.24 (m, 1H), 4.54-4.57 (m, 1H), 5.12 (d, *J* = 8 Hz, 1H), 6.63 (d, *J* = 9.2 Hz, 1H), 6.81-6.86 (m, 4H), 7.19-7.37 (m, 12H), 7.45 (d, *J* = 8 Hz, 1H), 7.82 (m, 2H); ^{13}C NMR (100 MHz, $CDCl_3$, 25 ºC): *δ*=36.9, 55.3, 63.9, 85.5, 87.2, 91.6, 102.5, 109.6, 113.3, 127.2, 127.8, 128.1, 128.1, 128.7, 129.5, 130.1, 135.2, 140.2, 144.2, 150.2, 158.7, 162.7; HRMS: *m/z* : calcd for $C_{36}H_{34}N_2O_7TeNa$: 759.1326, found: 759.1316 $[M+Na]^+$. **4b**: ^{1}H NMR (400 MHz, $CDCl_3$, 25 ºC): *δ*=1.21 (s, 3H), 3.35-3.52 (m, 2H), 3.81 (s, 6H), 3.98-4.01 (m, 1H), 4.22 (m, 1H), 4.55 (m, 1H), 6.69 (d, *J* = 10 Hz, 1H), 6.81-6.84 (m, 4H), 7.16-7.35 (m, 15H), 7.83 (m, 2H), 8.17 (b, 1H); ^{13}C NMR (100 MHz, $CDCl_3$, 25 ºC): *δ*=11.5, 36.8, 55.3, 63.9, 85.2, 87.2, 91.0, 109.5, 111.3, 113.3, 127.2, 128.0, 128.1, 128.6, 129.6, 130.1, 135.0,135.2, 140.3, 144.2, 150.4, 158.8, 163.1; HRMS: *m/z* : calcd for $C_{37}H_{36}N_2O_7TeNa$: 773.1477, found: 773.1475 $[M+Na]^+$.

2'-phenyltellanyl-5'-O-(4,4'-dimethoxytriphenylmethyl)-uridine 3'-O-(2-cyanoeth yl) diisopropylamino phosphoramidite (5a) and 2'-phenyltellanyl-5'-O-(4,4'-dimethoxytri phenylmethyl)-5-methyluridine 3'-O-(2-cyanoethyl) diisopropylamino phosphoramidite:

To the flask (25 mL) containing **4a** or **4b** (500 mg and 510 mg seperately, 0.68 mmol) under argon, dry methylene chloride (2.5 mL), N,N-diisopropylethylamine (0.17 mL, 1.03 mmol, 1.5 eq.), and 2-cyanoethyl N,N-diisopropyl-chlorophosphoramidite (195 mg, 0.83 mmol, 1.2 eq.) were added sequentially. The reaction mixture was stirred at -10 °C in an ice-salt bath under argon for 10 minutes, followed by removal of the bath. The reaction

was completed in 45 minutes at room temperature (indicated by TLC, 5% MeOH in CH_2Cl_2), generating a mixture of two diastereisomers. The reaction was then quenched with $NaHCO_3$ (2 mL, sat.), stirred for 5 min, and the product was then extracted with CH_2Cl_2 (3 x 5 mL). The combined organic layer was washed with NaCl (10 mL, sat.) and dried over $MgSO_4$ (s) for 20 min, followed by filtration. The solvent was then evaporated under reduced pressure and the crude product was re-dissolved in CH_2Cl_2 (2 mL). This solution was added drop-wise to petroleum ether (200 mL) under vigorous stirring, generating a white precipitate. The petroleum ether solution was decanted carefully. The crude product was re-dissolved again in CH_2Cl_2 (2 mL) and then loaded into an Al_2O_3 column (neutral) that was equilibrated with CH_2Cl_2/Hexanes (1:1). The column was eluded with a gradient of CH_2Cl_2 to CH_2Cl_2/EtOAc (7:3). After solvent evaporation and dry over high vacuum, the pure **5a** and **5b** (570 mg) were obtained as a white foamy products (88-90% yield). **5a**: ^{1}H NMR (400 MHz, $CDCl_3$, 25 ºC): δ=1.13-1.35 (m, 24H), 2.37 and 2.68 (2x t, *J*=6.4Hz, 4H), 3.38-3.69 (m, 12H), 3.83 (s, 12H), 3.90-4.03 (m, 2H), 4.32 and 4.38 (2x m, 2H), 4.67-4.82 (m, 2H), 4.95 and 5.00 (2x d, *J*=8 Hz, 2H), 6.73 and 6.75 (2x d, *J*=5.8 Hz, 2H) , 6.83-6.86 (m, 8H), 7.16-7.37 (m, 26H), 7.76 (br, 2H), 7.81-7.85 (m, 4H); ^{13}C NMR (100 MHz, $CDCl_3$, 25 ºC): δ=20.13, 20.20, 20.44, 20.51, 24.50, 24.61, 24.69, 25.23, 33.85, 34.43, 43.36, 43.40, 43.49, 43.52 , 55.29, 57.68, 57.87, 59.04, 59.23, 63.58, 63.73, 74.91 and 75.09, 84.97, 87.29, 89.38, 111.69, 113.46, 117.29, 126.99, 127.81, 128.54, 130.21, 130.42, 135.22, 135.40, 135.44, 144.10, 150.37, 158.85, 163.34; ^{31}P NMR (160 MHz, $CDCl_3$, 25 ºC): δ=148.6, 149.2; HRMS: *m/z* : calcd for $C_{45}H_{52}N_4O_8PTe$: 937.2579, found: 937.2578 $[M+H]^+$. **5b**: ^{1}H NMR (400 MHz, $CDCl_3$, 25 ºC): δ=1.10-1.42 (m, 24H), 2.34 and 2.66 (2x t, *J*=6.4Hz, 4H), 3.35-3.69

(m, 12H), 3.83 (s, 12H), 3.92-4.05 (m, 2H), 4.31 and 4.37 (2x m, 2H), 4.65-4.82 (m, 2H), 6.71 and 6.74 (2x d, *J*=5.8 Hz, 2H) , 6.84-6.86 (m, 8H), 7.16-7.37 (m, 26H), 7.77 (br, 2H), 7.80-7.84 (m, 4H); ^{13}C NMR (100 MHz, $CDCl_3$, 25 ºC): *δ*=11.69, 20.12, 20.22, 20.43, 20.53, 24.49, 24.60, 24.67, 25.24, 33.85, 34.43, 43.37, 43.40, 43.49, 43.53 , 55.27, 57.67, 57.88, 59.05, 59.25, 63.57, 63.74, 74.92 and 75.09, 84.97, 87.27, 89.38, 111.68, 113.45, 117.28, 126.98, 127.81, 128.54, 130.20, 130.39, 135.24, 135.38, 135.42, 144.11, 150.35, 158.87, 163.36; ^{31}P NMR (160 MHz, $CDCl_3$, 25 ºC): *δ*=148.4, 149.3; HRMS: *m/z* : calcd for $C_{46}H_{53}N_4O_8PTeNa$: 973.2561, found: 973.2560 $[M+Na]^+$.

(R)-5-(4,4'-dimethoxytrityloxymethyl)-2,3-dihydrofuran-4-ol (6):

To a stirred suspension of $NaBH_4$ (12 mg, 0.3 mmol) in dry THF (5 mL), under argon, dimethyl ditelluride (0.05 mL, 0.3 mmol) was added, followed by several drops of dry ethanol until bubble formed. The suspension was heated to 50 °C, followed by the addition of THF solution of starting material **3a** (320 mg, 0.6 mmol). The mixture was heated for three hours at this temperature under argon. The solvent was then evaporated and the residue was dissolved in CH_2Cl_2 (20 mL) and washed with water (3 x 20 mL). The CH_2Cl_2 layer was dried $MgSO_4$, concentrated and the residue was purified by silica gel column chromatography with pure CH_2Cl_2 to give compound **6** as white solid (230 mg, 92% yield). ^{1}H NMR (400 MHz, $CDCl_3$, 25 ºC): *δ*=3.18-3.22 (m, 2H), 3.81 (s, 3H), 4.44-4.49 (m, 1H), 4.79-4.81 (m, 1H), 5.20-5.21 (m, 1H), 6.62-6.63 (m, 1H), 6.84-6.86 (m, 4H), 7.23-7.46 (m, 9H,), 7.85 (d, *J* = 7.6 Hz, 1H); ^{13}C NMR (100 MHz, $CDCl_3$, 25 ºC):*δ*=55.2, 63.7, 76.2, 86.0, 88.3, 103.3, 113.1, 126.8, 127.8, 128.1, 130.1, 136.0, 144.8, 150.3, 158.5; HRMS: *m/z* : calcd for $C_{26}H_{25}O_5$: 417.1702, found: 417.1708 $[M-H]^-$.

2,2'-Anhydro-1-[2'-deoxy-3'-acetyl-5'-O-(4,4-dimethoxytrityl)-β-D-arabino furanosyl]-uracil (7):

To a solution of **3a** (320 mg, 0.6 mmol) in THF, acetic anhydride (0.19 mL, 2 mmol) was added, followed by several drops of triethylamine. The mixture was stirred at 50 degree for further 45 minutes (monitored by TLC, 5% MeOH in CH_2Cl_2) before quenched by methanol (4 mL). The solvent were removed under reduced pressure and the residue was re-dissolved in CH_2Cl_2 (40 mL). The suspension was washed with sodium bicarbonate (sat., 2 x 15 mL) and sat. brine (2 x 15 mL). The organic layer was dried over $MgSO_4$ (s), concentrated under reduced pressure, and the resulting residue was subjected to silica gel chromatography (0-5% MeOH in CH_2Cl_2) to give **7** (291 mg, 85 % yield) as white solids. **7**: ^{1}H NMR (400 MHz, $CDCl_3$, 25 ºC): *δ*=2.14 (s, 3H), 2.99-3.06 (m, 2H), 3.81 (s, 6H), 4.45 (m, 1H), 5.30-5.32 (m, 1H), 5.40 (m, 1H), 5.86 (d, *J* = 7.6 Hz, 1H), 6.27 (d, *J* = 5.6 Hz, 1H), 6.80-6.83 (m, 4H), 7.21-7.35 (m, 10H); ^{13}C NMR (100 MHz, $CDCl_3$, 25 ºC): *δ*=20.7, 55.3, 62.6, 77.0, 85.8, 86.3, 86.6, 90.4, 110.2, 113.3, 127.1, 128.0, 129.8, 135.2, 144.1, 158.6, 134.5, 159.1, 169.4, 171.2; HRMS: *m/z* : calcd for $C_{32}H_{31}N_2O_8$: 571.2080, found 571.2080 $[M+H]^+$.

5'-O-(4,4-dimethoxytrityl)-2',3'-didehydro-2',3'-dideoxyuridine (8):

To a stirred suspension of $NaBH_4$ (12 mg, 0.3 mmol) in anhydrous THF (5 mL), under argon, dimethyl ditelluride (0.05 mL, 0.3 mmol) was added, followed by several drops of dry ethanol until bubbles were formed. The suspension was heated to 50 °C, and the THF solution of the starting material **7**, (170 mg, 3 mmol) was added. The reaction completed in 3 hours, which monitored by TLC. Then the solvents were evaporated and the residue was dissolved in CH_2Cl_2 and washed with water. The organic solution was

dried over $MgSO_4$ and concentrated. The crude product was purified by silica gel column chromatography with 0-3% methanol in CH_2Cl_2 to give 90% yields of **8**. ^1H NMR (400 MHz, $CDCl_3$, 25 ºC): *δ*= 3.47-3.48 (m, 2H), 3.82 (s, 3H), 4.97-4.98 (m, 1H), 5.06 (d, 1H, *J* = 7.6 Hz), 5.89-5.91 (m, 1H), 6.35-6.37 (m, 1H), 6.84-6.86 (m, 4H), 7.05 (d, *J* = 2.0 Hz, 1H), 7.27-7.38 (m, 9H), 7.85 (d, *J* = 7.6 Hz, 1H); ^{13}C NMR (100 MHz, $CDCl_3$, 25 ºC): *δ*=55.4, 64.2, 86.1, 86.9, 89.6, 102.2, 113.2, 127.1, 127.8, 127.4, 129.1, 130.20, 150.6, 127.1, 134.6, 141.4, 150.6, 158.6, 163.2; HRMS: *m/z* : calcd for $C_{30}H_{27}N_2O_6$: 511.1869, found 511.1868 $[M-H]^-$.

2'-deoxy-2'-methyltellanyl-5'-O-(4,4'-dimethoxytrityl)-uridine (9a) and 2'-deoxy-2'-methyltellanyl-5'-O-(4,4'-dimethoxytrityl)-5-methyluridine (9b):

To a THF solution of dimethyl ditelluride ($CH_3TeTeCH_3$, 0.2 ml, 1.1 mmol, in 10 mL of THF) located in an ice bath under argon, was added 1M THF solution of $LiAlH_4$ (0.55 mmol) in 5 min. After the solution turned to slightly yellow, 12-crown-4 (0.6 mmol, 0.1 mL) was added and the mixture was stirred at 0 °C for further 20 min before the starting material **3a** and **3b** (0.57g and 0.58 g respectively in THF, 1.0 mmol) was added dropwisely. The reaction was monitored by TLC (5% CH_3OH in CH_2Cl_2) and it was found that the yield reached maximum in 4-5 hours before it dropped again by the generation of compound **6**. The reaction was quenched by adding 10 mL of saturated sodium chloride, followed by adding 10 mL of CH_2Cl_2, the organic solvent was further washed, dried over anhydrous $MgSO_4$ and removed under reduced pressure. The residue was then purified by a silica gel column (equilibrated with CH_2Cl_2) eluted with a methanol/methylene chloride gradient (CH_3OH in CH_2Cl_2, 0-3%) to afford the pure foamy product **9a** and **9b** in 40%-47% yield. **9a**: ^1H NMR (400 MHz, $CDCl_3$, 25 ºC):

δ=2.0 (s, 3H), 2.5 (s, 1H), 3.52 (dd, *J*1=2.8 Hz, *J*2=7.2Hz, 2H), 3.69 (t, *J*= 4 Hz, 1H), 3.82 (s, 6H), 4.21-4.23 (m, 1H), 4.31-4.36 (m, 1H), 5.40 (d, *J*=8.4 Hz, 1H), 6.33 (d, *J*= 8.4Hz, 1H) , 6.86-6.88 (m, 4H), 7.26-7.39 (m, 9H), 7.77 (d, *J*=8.4 Hz, 1H), 8.38 (br, 1H); ^{13}C NMR (100 MHz, $CDCl_3$, 25 ºC): *δ*=-21.2, 33.35, 55.29, 63.37, 73.75, 84.51, 87.29, 89.90, 102.73, 113.35, 127.28, 128.07, 130.08, 135.12, 139.77, 144.20, 150.23, 158.79, 162.64; HRMS: *m/z* : calcd for $C_{31}H_{31}N_2O_7Te$: 673.1194, found: 673.1204 $[M-H]^-$. **9b**: ^{1}H NMR (400 MHz, $CDCl_3$, 25 ºC): *δ*=1.43 (s, 3H), 1.99 (s, 3H), 3.47 (dd, *J*1=1.6 Hz, *J*2= 9.6Hz, 2H), 3.70-3.74 (m, 1H), 3.82 (s, 6H), 4.24-4.25 (m, 1H), 4.36-4.37 (m, 1H), 6.37 (d, *J*=9.2 Hz, 1H), 6.85-6.87 (m, 4H), 7.26-7.40 (m, 9H), 7.60 (s, 1H), 8.03 (br, 1H); ^{13}C NMR (100 MHz, $CDCl_3$, 25 ºC): *δ*=-20.8, 11.71, 33.11, 55.29, 63.86, 74.62, 84.64, 87.19, 89.90, 111.75, 113.32, 127.29, 128.09, 129.14, 135.20, 135.16, 144.21, 150.48, 158.82, 163.41; HRMS: *m/z* : calcd for $C_{32}H_{33}N_2O_7Te$: 687.1350, found: 687.1354 $[M-H]^-$.

2'-methyltellanyl-5'-O-(4,4'-dimethoxytriphenylmethyl)-uridine 3'-O-(2-cyanoethyl) diisopropylamino phosphoramidite (10a) and 2'-methyltellanyl-5'-O-(4,4'-dimethoxy triphenylmethyl)-5-methyluridine-3'-O-(2-cyanoethyl) diisopropylamino phosphoramidite: To the flask (25 mL) containing **9a** or **9b** (230 mg and 234 mg seperately, 0.34 mmol) under argon, dry methylene chloride (2.5 mL), N,N-diisopropylethylamine (0.1 mL, 0.51 mmol, 1.5 eq.), and 2-cyanoethyl N,N-diisopropyl-chlorophosphoramidite (120 mg, 0.51 mmol, 1.5 eq.) were added sequentially. The reaction mixture was stirred at -10 °C in an ice-salt bath under argon for 10 minutes, followed by removal of the bath. The reaction was completed in 45 minutes at room temperature (indicated by TLC, 5% MeOH in CH_2Cl_2), generating a mixture of two diastereisomers. The reaction was then quenched with $NaHCO_3$ (2 mL, sat.), stirred for 5 min, and the product was then extracted with

CH_2Cl_2 (3 x 5 mL). The combined organic layer was washed with NaCl (10 mL, sat.) and dried over $MgSO_4$ (s) for 20 min, followed by filtration. The solvent was then evaporated under reduced pressure and the crude product was re-dissolved in CH_2Cl_2 (2 mL). This solution was added drop-wise to petroleum ether (200 mL) under vigorous stirring, generating white precipitate. The petroleum ether solution was decanted. The crude product was re-dissolved again in CH_2Cl_2 (2 mL) and then loaded into an Al_2O_3 column (neutral) that was equilibrated with CH_2Cl_2/Hexanes (1:1). The column was eluded with a gradient of methylene chloride and ethyl acetate [CH_2Cl_2 to CH_2Cl_2/EtOAc (7:3)]. After solvent evaporation and dry over high vacuum, the white **10a** and **10b** was obtained as a white foamy product (88-90% yield). **10a**: ^{1}H NMR (400 MHz, $CDCl_3$, 25 ºC): δ=0.85-1.38 (m, 24H), 1.98 and 1.99 (2x s, 6H), 2.41 and 2.67 (2x t, *J*=7.6Hz, 4H), 3.42-3.69 (m, 12H), 3.83 (s, 12H), 3.89-4.05 (m, 4H), 4.22 and 4.30 (2x m, 2H), 4.57-4.77 (m, 2H), 6.43 and 6.55 (2x d, *J*=8.6 Hz, 2H) , 6.80-6.92 (m, 8H), 7.16-7.45 (m, 18), 7.78 and 7.79 (s, 2H); ^{13}C NMR ($CDCl_3$, 25 ºC): δ=-21.32, 19.16, 19.66, 20.88, 24.81, 24.44, 43.30, 43.46, 46.34, 47.53, 51.47 and 51.85, 55.62, 57.03, 58.39, 62.52, 73.36, 73.67, 84.57, 87.35, 88.12, 103.39, 113.42, 117.28, 127.21, 127.94, 128.36, 130.19, 130.25, 135.17, 135.38, 139.53, 144.20, 150.42, 158.80, 163.19. ^{31}P NMR (160 MHz, $CDCl_3$, 25 ºC): δ=148.5, 149.2; ESI-TOF: m/z calcd for $C_{40}H_{48}N_4O_8PTe$: 873.2272, found 873.2264. **10b**: ^{1}H NMR (400 MHz, $CDCl_3$, 25 ºC): δ= 0.89-1.42 (m, 24H), 1.40 (2x s, 6H), 2.10 (2x s, 6H), 2.42 and 2.69 (2x t, *J* =7.8 Hz, 4H), 3.48 -3.72 (m, 12H), 3.82 (s, 12H), 3.90-4.05 (m, 2H), 4.22 and 4.31 (2x m, 2H), 4.58-4.72 (m, 2H), 6.42 (d, *J* =8.6 Hz, 2H), 6.79-6.92 (m, 8H), 7.20-7.45 (m, 18H), 7.62 (2x s, 2H), 8.25 (br, 2H); ^{13}C NMR (100 MHz, $CDCl_3$, 25 ºC): δ=-21.32, 4.58, 4.72, 20.13, 20.16, 20.48, 20.55, 22.52, 24.39,

24.47, 24.59, 24.68, 24.76,24.83, 43.21, 43.32, 43.34, 43.41, 46.35, 46.42, 47.37, 47.42, 50.36, 55.32, 63.29, 63.33, 74.90, 75.09, 84.99, 87.28, 89.36, 111.70, 113.44, 117.29, 126.99, 127.83, 128.55, 130.20, 130.41, 135.23, 135.41, 135.44, 144.11, 150.38, 158.86, 163.35. ^{31}P NMR (160 MHz, $CDCl_3$, 25 ºC): δ=148.6, 149.5; ESI-TOF: m/z calcd for $C_{41}H_{52}N_4O_8PTe$: 889.2585, found 889.2587.

7.3.2 Incorporation of Te-functionality into oligonucleotides

Both 2'-TeMe and 2'-TePh modified phosphoramidites were incorporated into oligonucleotides by solid-phase synthesis in over 95% coupling yield, using 5-(benzylmercapto)-1H-tetrazole (5-BMT) as the activator. We found that the bulky 2'-Te-moieties didn't significantly affect the coupling reaction, and these 2'-Te-modified phosphoramidites could achieve almost quantitative coupling even with a standard 25 seconds of coupling time. The Te-oligonucleotides were synthesized via the DMTr-on mode using ultra-mild phosphoramidites and cleaved with K_2CO_3 at room temperature for 8 hours. To reduce the partially oxidized Te-DNAs, a diborane treatment was performed right after the deprotection step. As an example, the HPLC analysis profile of a crude 2'-TePh-T DNA with DMTr-on is showed in Figure 7.2, from which we determined the Te-coupling yield as over 95%. After the HPLC purification with DMTr-on, the DMTr group was removed by treatment with 3% trichloroacetic acid, and the oligonucleotides were purified again by HPLC, followed by MALDI-TOF MS analysis to confirm the integrety of these Te-DNAs (Table 7.2 and Figure 7.3). To our pleasant surprise, both methyltelluride and phenyltelluride functionalities are stable during the deprotection and purification.

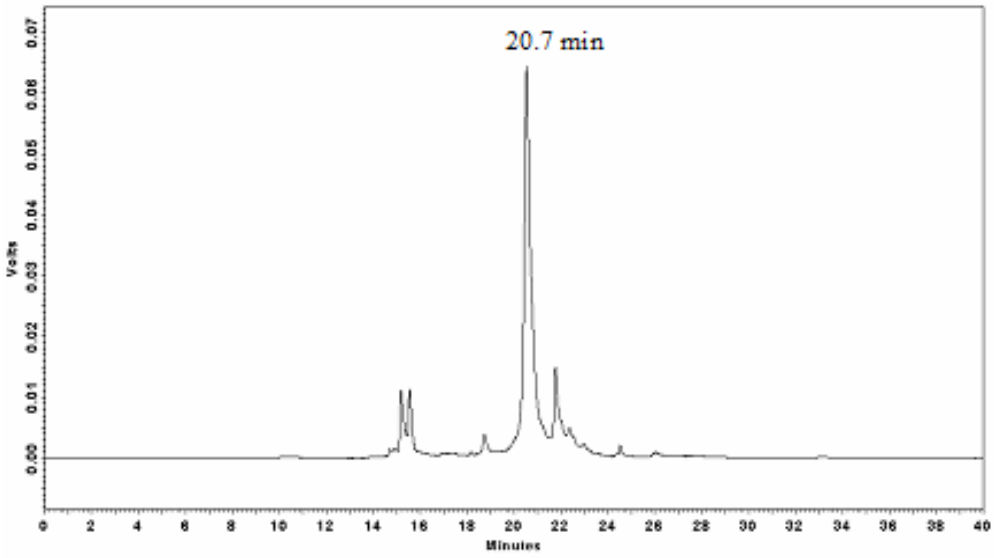


Figure 7.2: HPLC analysis of the Te-DNA: 5'-DMTr-G(2'-TePh-U)GTACAC-3'. This analysis was performed on a Zorbax SB-C18 column (4.6 x 250 mm) with a linear gradient from buffer A (20 mM TEAAc: triethylammonium acetate, pH 7.1) to 100% buffer B (50% acetonitrile in 20 mM TEAAc) in 20 min.; the retention time of the full-length Te-DNA is 20.7 min.

Table 7.2: MALDI-TOF analysis of 2'-Te modified oligonucleotides.

DNA sequences	Formula	MS (M-H) (Calcd.)
G<u>U</u>(2'-TeMe)GTACAC	$C_{78}H_{99}N_{30}O_{46}P_7Te$	2536.1 (2536.2)
ATGG<u>U</u>(2'-TeMe)GCTC	$C_{88}H_{112}N_{32}O_{54}P_8Te$	2856.3 (2856.4)
G<u>U</u>(2'-TePh)GTACAC	$C_{83}H_{101}N_{30}O_{46}P_7Te$	2599.4 (2599.3)
ATGG<u>U</u>(2'-TePh)GCTC	$C_{93}H_{114}N_{32}O_{54}P_8Te$	2919.0 (2918.5)
GCG<u>U</u>(2'-TePh)ATACGC	$C_{102}H_{125}N_{38}O_{58}P_9Te$	3218.0 (3216.7)
G<u>T</u>(2'-TePh)GTACAC	$C_{84}H_{103}N_{30}O_{46}P_7Te$	2611.8 (2612.3)
ATGG<u>T</u>(2'-TePh)GCTC	$C_{94}H_{116}N_{32}O_{54}P_8Te$	2933.3 (2932.5)
CTd<u>U</u>(2'-TePh)CTTGTCCG	$C_{111}H_{140}N_{32}O_{69}P_{10}Te$	3464.5 (3465.4)
CTd<u>U</u>(2'-TeMe)CTTGTCCG	$C_{106}H_{138}N_{32}O_{69}P_{10}Te$	3400.5 (3401.8)

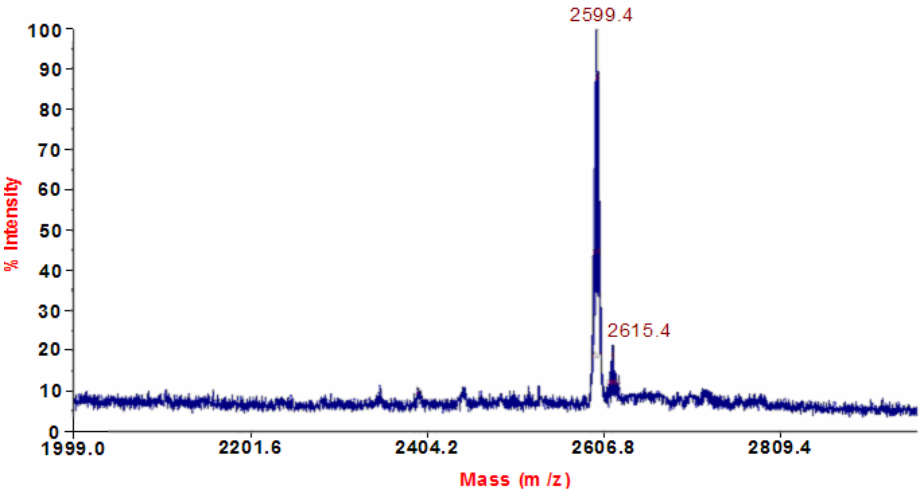
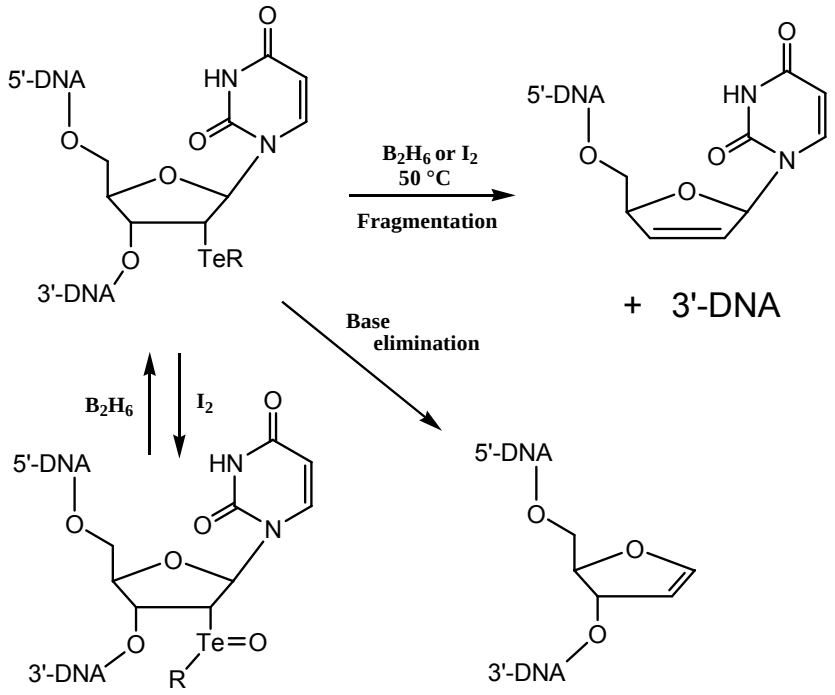


Figure 7.3: MALDI-TOF mass of Te-oligonucleotide 5'-GU(2'-TePh)GTACAC-3'. The peak at m/z with M+16 showed the possible oxidation of tellurium atom.

It's noteworthy that approximately 5% of the Te-DNAs are oxidized to the telluoxides after the solid-phase synthesis. Fortunately, the telluoxides can be reduced back to the tellurides. The detailed investigation of the purified Te-DNAs by HPLC analysis has revealed that the telluride DNA can be oxidized to the telluoxide DNA by the I_2 condition, and the telluoxide can be reduced back to the telluride by B_2H_6 (Scheme 7.5 and Figure 7.4). This provides a useful approach to recover the oxidized telluride, and allows further investigation of the redox property of tellurium functionalities. Interestingly, under heating at 50 °C, we have observed a site-specific cleavage of the 2'-TePh-DNA at the modification site in the presence of either B_2H_6 or I_2 (Figure 7.5). This result is also consistent with our previous observation of the 2',3' elimination of the Te-nucleosides. This site-specific fragmentation provides a good model for DNA damage study. Furthermore, in a 2'-TeMe modified 9mer DNA [5'-ATGG(2'-TeMe-U)GCTC-3'], the base elimination was also observed (Figure 7.6). This Te-DNA research can facilitate investigation of the DNA cleavage and nucleobase damage.



Scheme 7.5: Redox and fragmentation of DNA oligonucleotides containing 2'-Te functionalities.

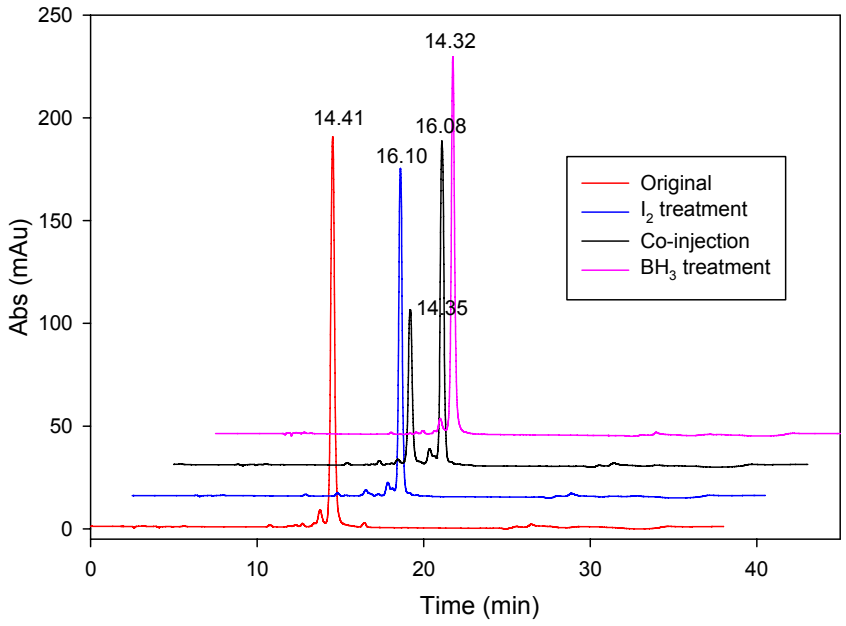


Figure 7.4: Redox behavior of 2'-TePh-8mer GU(2'-TePh)GTACAC with B_2H_6/I_2 system. Red: original profile; blue: I_2 treatment; black: co-injection; purple: re-treated with B_2H_6.

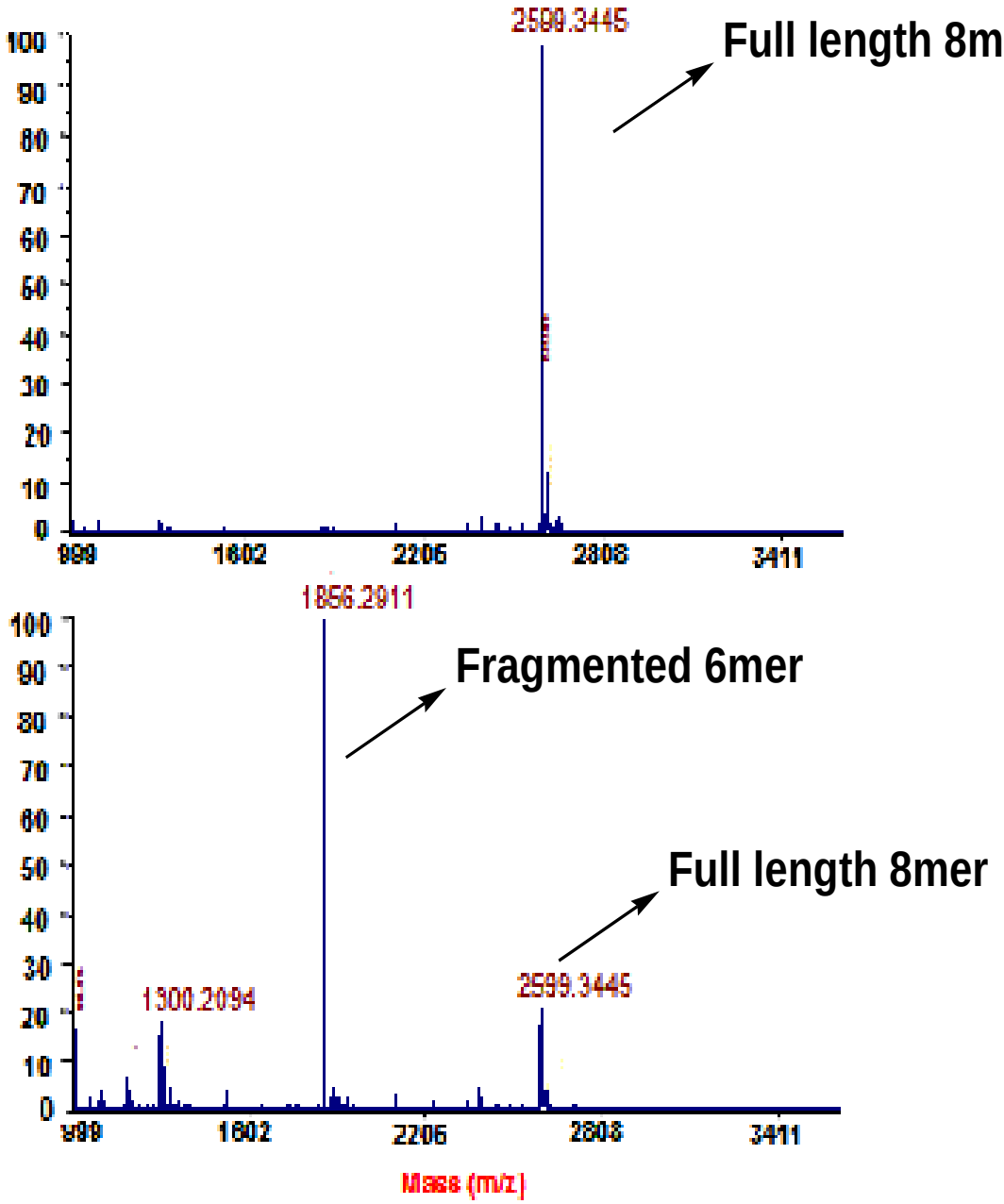
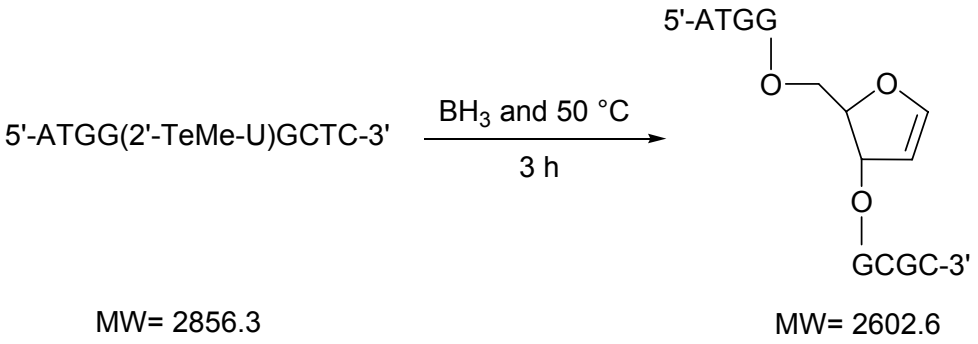


Figure 7.5: Thermo-cleavage of 2'-TePh modified DNA 8mer 5'-GU(2'-TePh)GTACAC-3'. Above: the original MALDI-TOF spectrum, molecular weight is 2600 $(M+H)^+$; Bottom: MALDI-TOF spectrum of fragmentated DNA, after heating with either I_2 or B_2H_6 at 50 degree, the fragmentated 6mer represents 5'-p-GTACAC-3', Fw=1856.



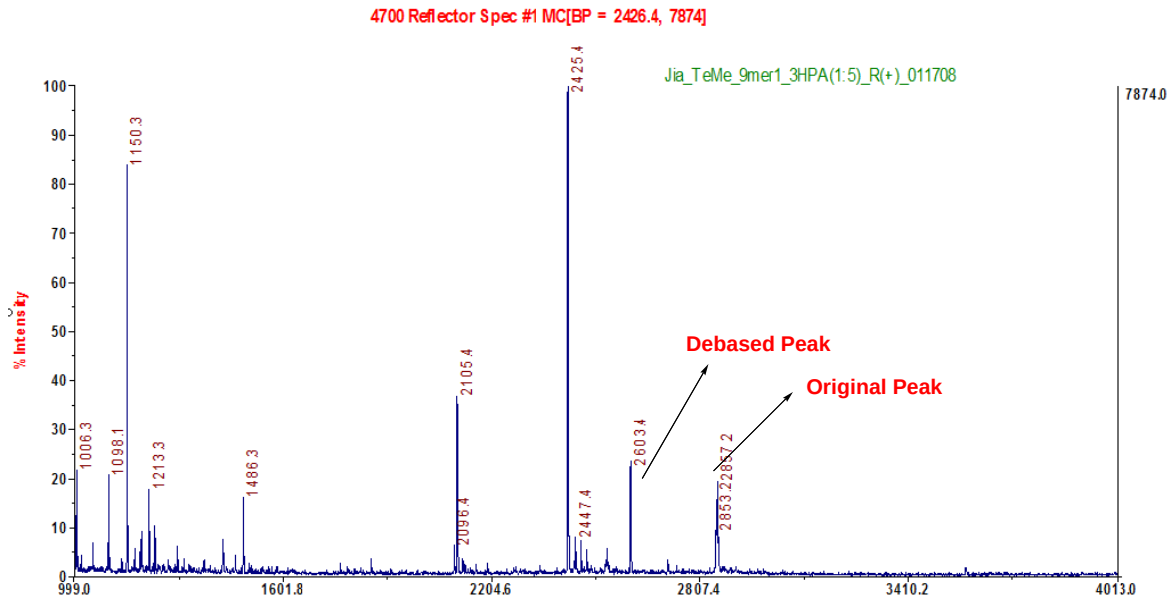


Figure 7.6: MALDI-TOF analysis of the reductive base elimination of the 2'-TeMe modified sequence 5'-ATGG(2'-TeMe-U)GCTC-3'.

7.4 Thermal denaturization studies of 2'-Te containing DNAs

Due to the size of Te atom, we anticipate that the incorporation of the Te-functionalities at the 2' position perturbs the DNA duplexes. In addition, we expect that the 2'-Te-Ph moiety causes more perturbation than the 2'-Te-Me. Thus, we have investigated the Te-modified duplex stability by measuring their UV-melting temperatures. In the case of a self-complementary duplex [5'-G(2'-TeX-U)GTACAC-3']$_2$,

where the double Te-modifications are introduced to a short DNA duplex, the melting temperature dropped dramatically (Figure 7.7). In the case of the 2'-TePh, the Tm decreased by 15 ºC while the 2'-TeMe by 10 ºC from the native duplex. It suggests that short duplex of the self-complimentary DNA can not accommodate these two bulky modifications well.

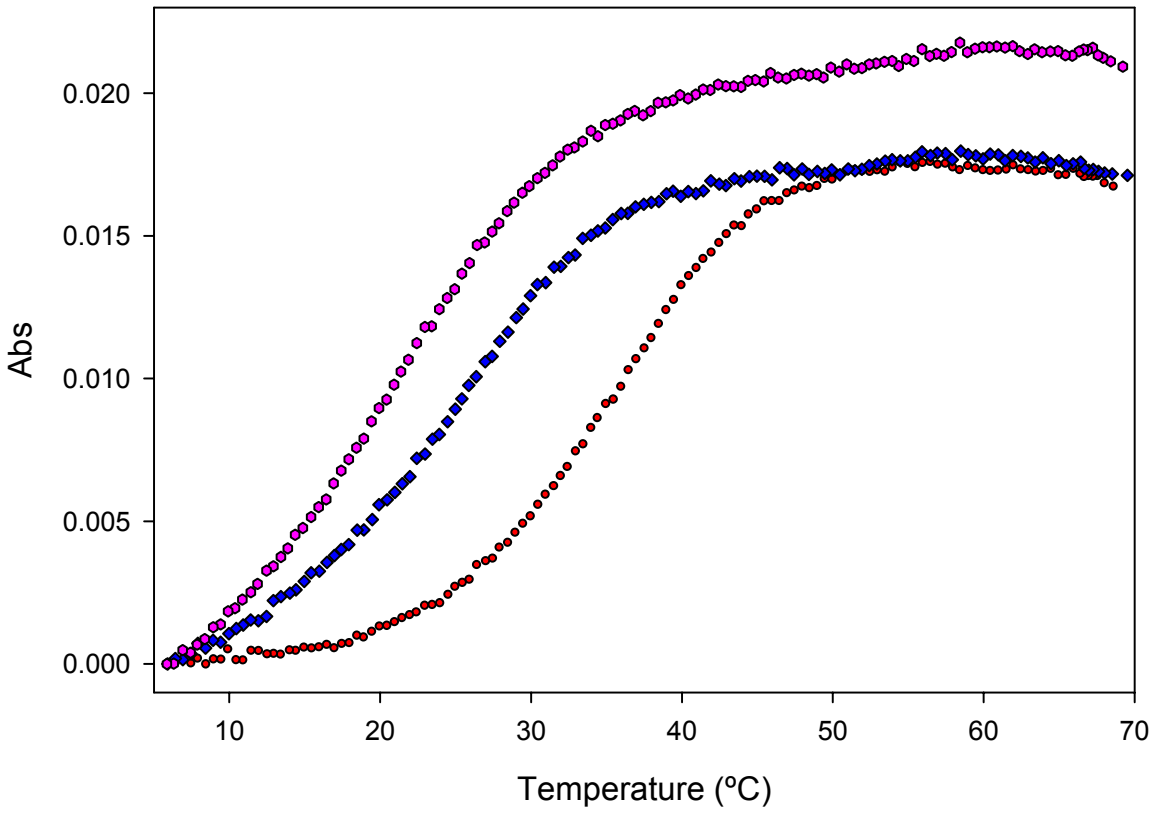


Figure 7.7: Thermaldenaturation curves of 2'-TeX modified self-complementary sequence 5'- GU(2'-TePh)GTACAC -3' (X=Me or Ph). The samples (1 mM duplex) were heated from 6 to 70 ºC with a rate of 0.5 ºC/min. The red, blue and purple curves represent the native, 2'-TeMe and 2'-TePh duplex respectively. The averaged (four times experiments) Tm are: 39.4ºC(purple), 28.6ºC(Blue), 23.4ºC(Pink) respectively.

When introducing the Te-functionalities to a non-self-complementary duplex [5'-C(2'-TeX-U)TCTTGTCCG-3' and 5'-CGGACAAGA-AG-3'], it was observed that the melting temperatures of the native, 2'-TeMe and 2'-TePh duplexes were 44.0, 40.7 and 36.8 ºC, respectively. It is noteworthy that when the TeMe moiety is introduced close to a terminal position, the Te-modification is relatively well accommodated (Figure 7.8). Since the termini are usually dynamic and spacious, our result suggests that the dynamics and space allow a better accommodation. Therefore, the perturbation is most likely a local event, and this bulky Te-modification may be useful in investigating the formation of DNA or RNA loop, bulge, and other local conformations.

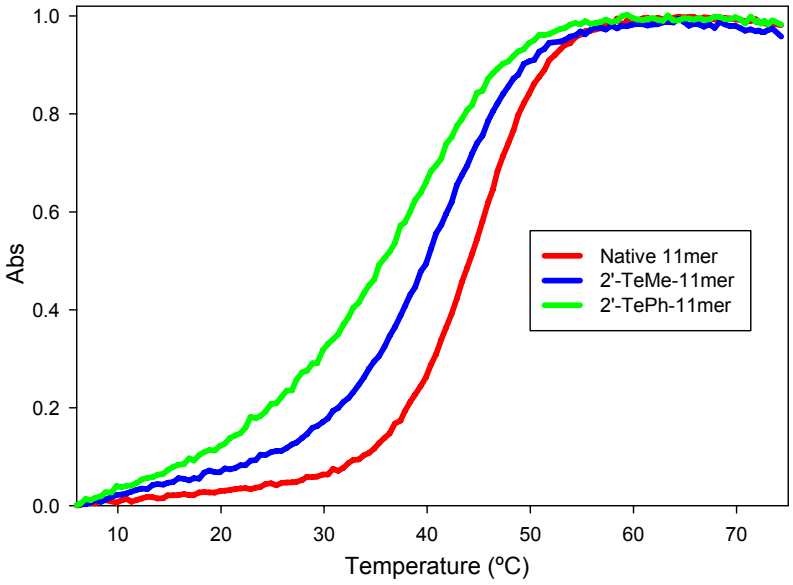


Figure 7.8: Normalized thermaldenaturation curves of 11mer pair 5'-CU(2'-TeX)TCTTGTCCG-3'&3'-CGGACAAGAAG-5'. The samples (1 mM duplex) were heated from 6 to 75 ºC with a rate of 0.5 ºC/min. The red, blue and green curves represent the native, 2'-TeMe and 2'-TePh duplex respectively. The averaged (four times experiments) Tm are: 44.0 ± 0.1 ºC, 40.7 ± 0.1 ºC and 36.8 ± 0.2 ºC respectively.

7.5 Conclusion of this chapter

In summary, we have synthesized for the first time the Te-derivatized nucleoside phosphoramidites and incorporated the Te-functionalities into oligonucleotides with a high coupling yield. Several 2'-Te-modified oligonucleotides were synthesized and characterized by HPLC and MS. We have demonstrated that both of the 2'-TeMe and 2'-TePh functionalities are compatible with the solid-phase synthesis, deprotection and purification. In addition, we revealed the redox property of the Te-functionalities and the interchange of the telluride and telluoxide DNAs by a redox reaction. We also found that the 2'-Te-modified DNA can be reductively or oxidatively cleaved, and the DNA nucleobase can be eliminated via the 1',2' elimination. This system may be useful in studying the DNA cleavage and nucleobase damage. Furthermore, in our UV-melting study of the 2'-Te-modified DNA duplexes, as expected, we observed the Tm decrease of the Te-DNA duplexes. And probably that's also the reason why we haven't succeeded to obtain the crystal for this derivatization using the self-complementary sequence model. But since the Tm decrease is dependent on the size and location of the Te-modifications, this derivatization strategy is useful in probing DNA and RNA local conformations and topological structures. Moreover, this Te-derivatization of nucleic acids has great potentials in nucleic acid X-ray crystallography as well as in structural and functional analysis of nucleic acid-protein large complexes with electron microscope. In the end, a novel tellurium mediated elimination reaction was discovered by accident during this process, which could provide another new application of tellurium reagents in medicinal research area.

8. SYNTHESIS OF 5-TELLURIUM MODIFIED URIDINE FOR NUCLEIC ACID STRUCTURE AND FUNCTION STUDIES

8.1 Introduction

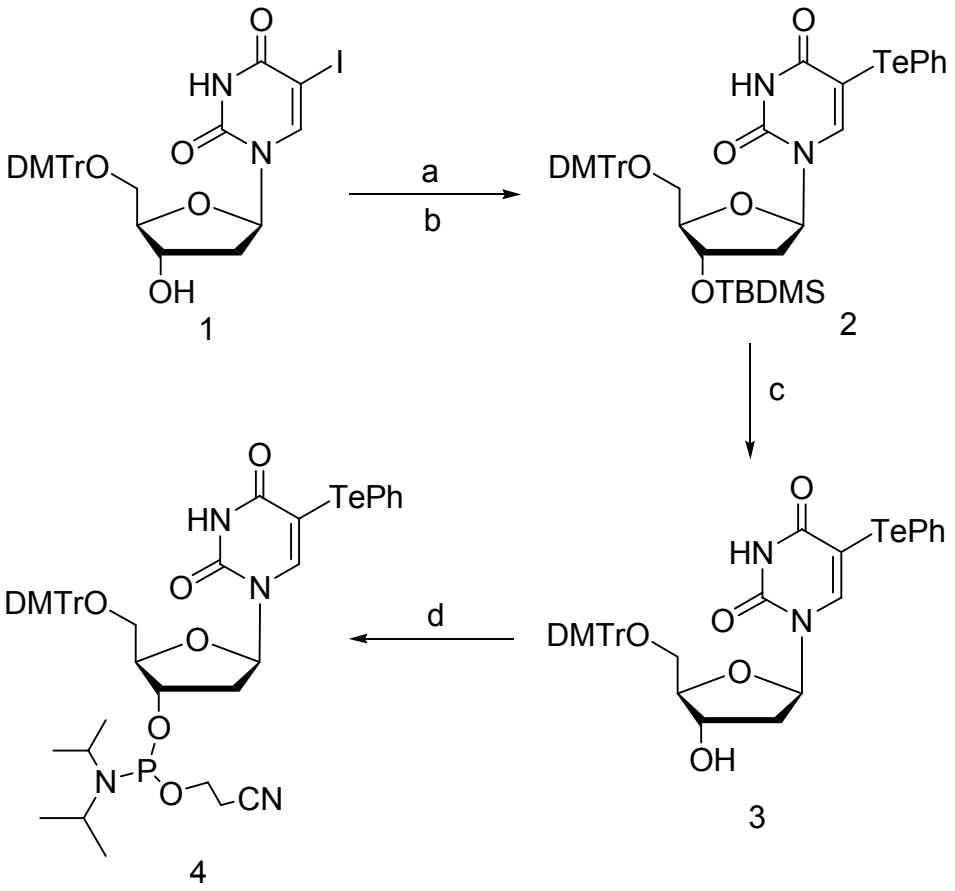
To further take the potential advantage of tellurium functionality as promising alternative to the usual soaking procedure of heavy-atoms incorporation for classic multiple isomorphous replacement (MIR) while avoiding the use of synchrotron light source which is not always convenient and available, we continue our TeNA (tellurium modifife nucleic acids) project by exploring the Te-derivatization at different positions. As stated in the previous chapter, although the 2'-tellurium (either phenyl or methyl) functionalities could cause some structure perturbations in some oligonucleotide duplexes especially self-complementary ones, probably due to the large size of tellurium center, the compatibility of tellurium chemistry with solid phase synthesis is well proved, as well as the stability of tellurium functions. In addition, although we could not get the crystal for this 2'-Te modified DNAs, the stability and structural isomorphorism of TeMet-proteins have been confirmed as far comparing the native proteins. So, we thought it's still very necessary to get the more detailed structure information about the effect of tellurium functionality to nucleic acid structures.

Considering that the space limitation of the 2'-position which located in the minor groove of DNA, we would like to carried out the trial to the 5-position of the deoxyuridine base which located in the major groove instead. Hopefully, the derivatization of this position will cause much less structure perturbations. This chapter will present the related work, as well as our structure studies.

8.2 Preparation of 5-TePh functionalized nucleosides and DNAs

8.2.1 Synthesis of 5-TePh-uridine and the phosphoramidites

The synthesis of 5-TePh-uridine phosphoramidite is showed in scheme 8.1, starting from the silylation of commercially available 5-Iodo-2-deoxyuridine derivative 1 in a quantitative yield. After purification, the treatment of this intermediate with NaH/n-BuLi followed by adding diphenylditelluride $(PhTe)_2$ gave an inseparable mixture of the desired 5-PhTe derivative (major) and the corresponding 6-PhTe derivatives, along with the reduction byproduct 2'-TBDMS-5'-DMTr-deoxyuridine. After several failure trials such as re-crystallization and slow column chromatography to separate the isomers, we switched the starting compound to the 3',5'-di-O-silyl derivatives, but unfortuniterly, this 5-isomer was still difficult to be isolated from the corresponding 6-regio-isomer. Finally, the concentrations of the reactant were adjusted, based on the 3'-O-silyl-5'-DMTr derivative (Scheme 8.1). As a result, it was found that when the concentration of reactant was increased to as high as 0.15 M or 0.18 M of THF solution and the reaction was performed at the lower temperature, the 6-TePh isomeric compound could indeed be completely eliminated, which produced a mixture of 5-TePh derivative and the corresponding reductive byproduct deoxyuridine. This mixture could then be separated by careful column purification in acceptable yield. With this key intermediate in hand, the deprotection of compound 2 with TBAF/THF was carried out, as a result, the precursor 3 was generated quantitatively, which is ready to be transferred into the phosphoramidite for solid phase synthesis. The final phosphoramidite compounds were obtained and purified by following the standard procedures.



Scheme 8.1: Synthesis of 5-TePh derivatized 2'-deoxyuridine phosphoramidite.
(a) TBDMSCl, Pyridine. (b) NaH, nBuLi, PhTeTePh, THF. (c) Et_3N•3HF, THF. (d) 2-Cyanoethyl N, N-diisopropyl-chlorophosphoramidite and N,N-diisopropyl ethylamine, CH_2Cl_2.

8.2.2 Incorporation of 5-Te-functionality into oligonucleotides

Considering that the tellurium center could be a good electrophile, ultra-mild phosphoramidite reagents was used to avoid the ammonium treatment. As expected, this Te-phosphoramidite is well compatible with solid phase synthesis conditions and thus has been incorporated into several designed DNA sequences. After K_2CO_3 (0.05 M in MeOH) cleavage and deprotection, the crude DNA solution was checked by analytical

RP-HPLC. From the percentage of full length DNA in figure 9.1, the real coupling yield could be claimed as over 98%. The samples were then purified in a preparative RP-HPLC system as trityl-on form. After treated with 3% trichloroacidic acid for two minutes, the solution was neutralized with triethylamine and extracted with petroleum ether, ready for the second purification as trityl-off form.

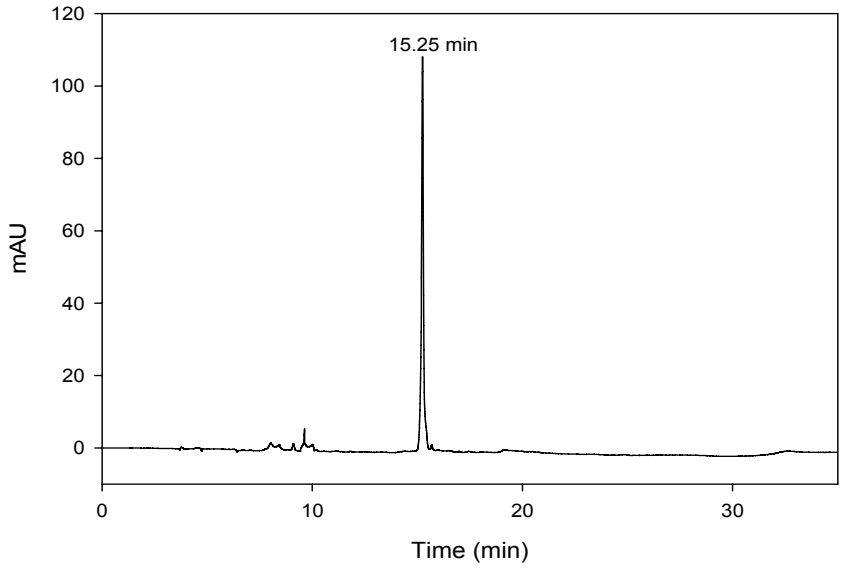


Figure 8.1: HPLC analysis of 5-TePh-9mer: 5'-ATGGdU(5-TePh)GCTC-3'. Buffer A: 20 mM of TEAAC, Buffer B: 50% ACN in buffer A. Gradient: 5%~ 50% B in 10min, up to 100% B in 15 min, keep 100% B until 25min, then go back to 0% in 30 min. The full length sample retention time is 15.25 min.

All the tellurium modified DNA oligonucleotides were characterized by MALDI-TOF. From an example showed in figure 8.2, we can tell that the tellurium function is also fairy stable with the oxidation environment including iodine treatments and air exposure during synthesis and purification, except a tiny amount of oxidized form was observed in the mass spectrum. As demonstration, several tellurium modified DNA

oligoes were listed in table 8.1. It's also worthy to mention that the thermal stability test was also carried out by heating this DNA 9mer at 80 degree for 8 hr, and from the mass spectrum, we didn't see any fragmentation or decomposition.

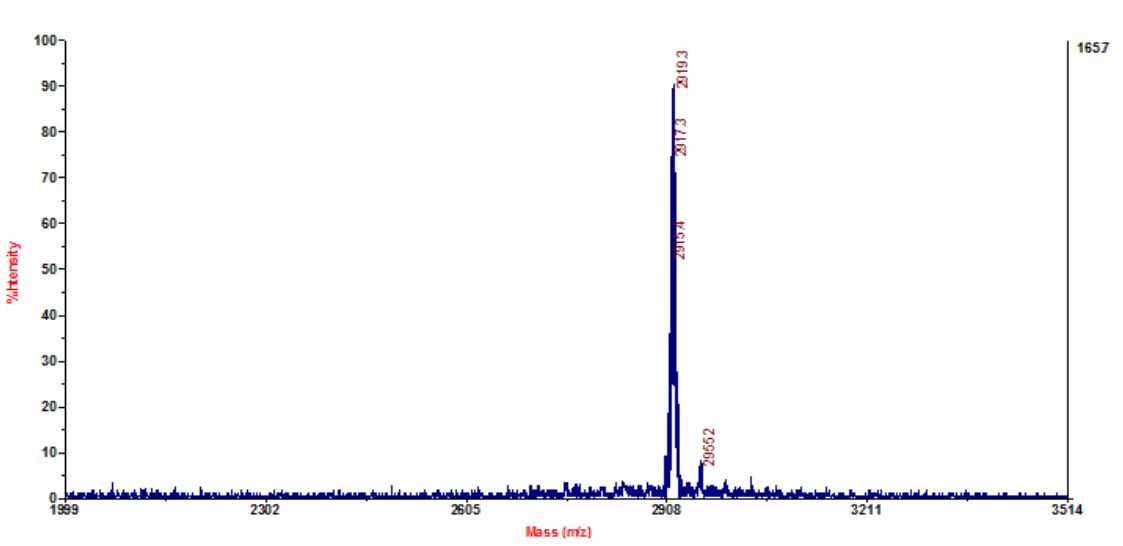


Figure 8.2: MALDI-TOF mass spectrum of 5-TePh-9mer: 5'-ATGGdU(5-TePh)GCTC-3'. $C_{93}H_{114}N_{32}O_{54}P_8Te$, calculated: 2919.5, observed: 2919.3

Table 8.1: MALDI-TOF analysis of 5-TedU containing DNA oligonucleotides.

DNA sequences	Formula	MS (Calcd.)
ATGGdU(5-TePh)GCTC	$C_{93}H_{114}N_{32}O_{54}P_8Te$	2919.3 (2920.4)[a]
CTCCCAdU(5-TePh)CC	$C_{89}H_{113}N_{27}O_{53}P_8Te$	2984.4 (2785.4)[a]
CTdU(5-TePh)CTTGTCCG	$C_{111}H_{140}N_{32}O_{69}P_{10}Te$	3464.5 (3465.4)[a]
GCGdU(5-TePh)ATACGC	$C_{102}H_{125}N_{38}O_{58}P_9Te$	3218.0 (3216.7)[b]
GTGdU(5-TePh)ACAC	$C_{83}H_{101}N_{30}O_{46}P_7Te$	2599.4 (2600.3)[a]
GdU(2'-SeMe)GdU(5-TePh)ACAC	$C_{83}H_{101}N_{30}O_{46}P_7SeTe$	2679.2 (2678.3)[b]

a: negative charged (M-H); b: positive charged (M+H).

8.3 Thermal denaturalization study

After confirming the thermal stability of Te-DNA oligoes, we have further conducted the duplex thermodenaturaization studies. As original thought, the bulky phenoyltelluride group shouldn't affect the duplex global structures since it's located in the much wider major groove. Our experimental data also confirmed this assumption, which showed in figure 8.3 as an example in a DNA 9mer duplex model. There was no obvious Tm decrease observed, instead, a little bit increase was observed, probably due to the enhancement of base stacking between the Te-dU and the neighbours. Similarly, the other two DNA models were also applied, which showed the consistent results.

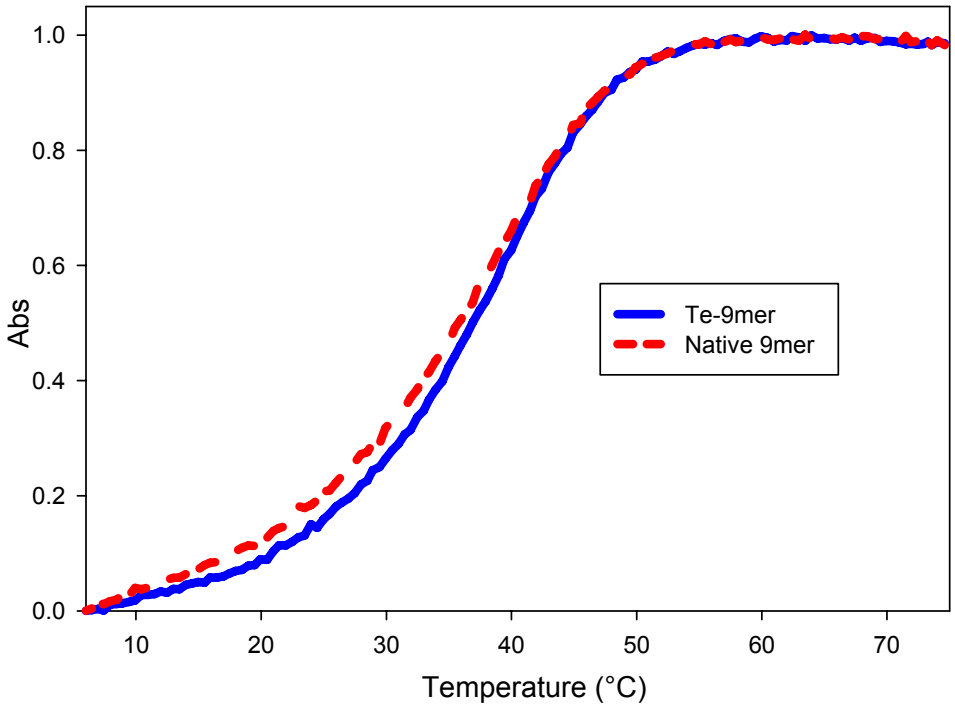


Figure 8.3: Normalized thermaldenaturization curves of DNA 9mer duplex: 5'-ATGGU*GCTC-3'&5'-GAGCACCAT-3'. Red desh curve represents native pair, Tm=39.8 ± 0.3 °C; while blue line represents the Te-pair, Tm=40.3 ± 0.2 °C.

8.4 Crystallization and data collection of tellurium derivatized DNA

To further confirm that this 5-TePh modification didn't cause obvious structure perturbations, we have also tried to get the crystal structures for them. Initially, none of the sequence crystallized for a long time although more than 100 buffer conditions were applied to screen. Later on, this problem was solved by using our 2'-Se modification strategy we discovered previously, which could greatly facilitate the crystal growth. This both Se-Te-modified DNA (5'-G$\underline{U}^{2'\text{-SeMe}}G\underline{U}^{5\text{-TePh}}$ACAC-3') was designed, synthesized and purified. Before setting crystallization, the sample was annealed by heating at 70 degree for 3 min and cooling down to room temperature slowly. As a result, all the 24 conditions of the Hampton Mini-Screening kit were working well to generate high quality of crystals with reasonable sizes and shapes within two days (Figure 8.4A and 8.4B as an example). By diffraction screening, buffer #7 of the kit (10% v/v MPD, 40 mM sodium cacodylate pH 6.0, 12 mM spermine tetra-HCl, 80 mM potassium chloride and 20 mM magnessium chloride) was identified to give the highest resolution. Besides this, several good data sets were also collected for further refine purpose. The data collection and refine parameters are listed in table 8.2 and 8.3.

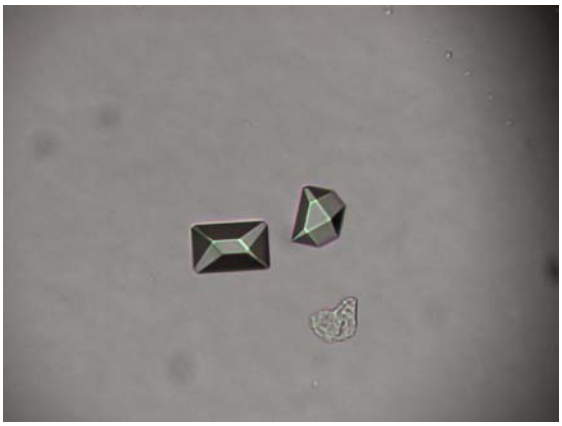


Figure 8.4 (A): Photos of Se-Te derivatized DNA octomer crystals.

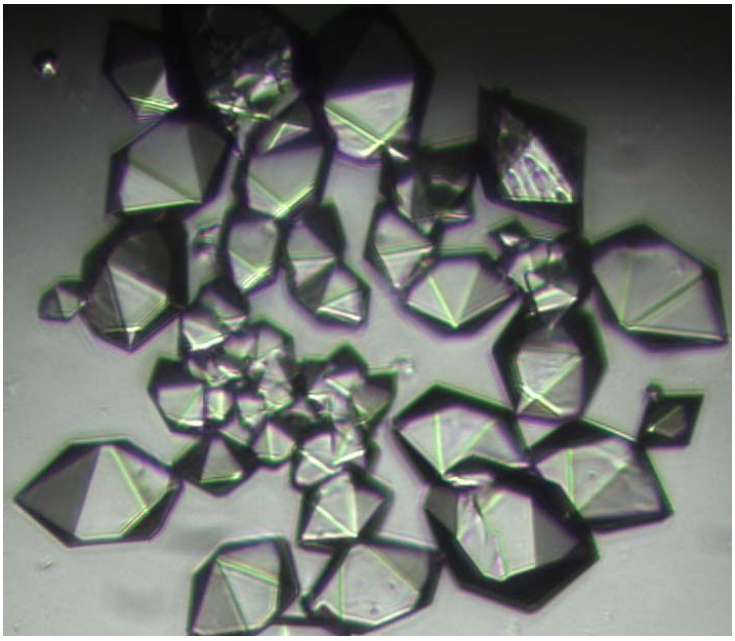


Figure 8.4 (B): Photos of Se-Te derivatized DNA octomer crystals.

Table 8.2: X-ray data collection: The data collected at selenium K-edge.

Data Collection	$\lambda = 0.9795$ Å
Resolution range, Å (last shell)	18.89-1.50 (1.55-1.50)
Unique reflections	3659
Completeness, %	94.4
R_{merge} ,%	7.5 (42.7)
$\langle I/\sigma(I)\rangle$	14.9 (2.4)
Redundancy	8.63 (5.30)
Reduced ChiSquared	0.93 (0.42)

$R_{merge}=\Sigma|I-\langle I\rangle|/\Sigma I$

Table 8.3: Refinement statistics in Se-Te-8mer.

Refinement	
Resolution range, Å (last shell)	25-1.50 (1.55 - 1.5)
Number of reflections	6859 (424)
R_{work} , %	19.3
R_{free} , %	21.5
Nucleic Acid (single)	162
Heavy atom	2 (Se, Te)
Water	28
Bond length R.M.S.D, Å	0.010
Bond angle R. M.S.D,	1.73
Average B-factors, $Å^2$	
All atoms	22.3
Wilson plot	19.6
Overall anisotropic B-values	
B11/B22/B33	0.6/0.6/-1.2
Solvent density correction, $e/Å^3$	0.44
Solvent B-factors, $Å^2$	59.6
Coordinates error (c.-v.), 5Å	
Esd. from Luzzatt plot, Å	0.17
Esd. from SIGMAA, Å	0.15

8.5 X-ray crystallography of the first Te-containing DNA

This Se-Te-8mer DNA structure have the same tetragonal space group $P4_32_12$ as the native one (33). Figure 8.5 showed the electron density maps of the Te-dU residue, as well as the base pair between this residue and the A5, which indicates that this modification doesn't affect the duplex base pair. To our surprise, there was no phenyl group present in this structure, which was also consistent with the other sets of high-quality data we have collected. In addition, although the density map of tellurium atom is very obvious (much bigger than oxygen), the Te occupancy was only about 40%. Since our thermal stability study has showed there was no decomposition happened during the annealing heating, we therefore tested its X-ray stability, as a result (figure 8.6), our MALDI-TOF spectrum showed that both phenyl and phenyl telluride groups could partially fell off during the exposition. We also reduced the exposure time to as short as 5s, but the same results were obtained. Although low occupancy it was, we thought the tellurium signal should be still strong enough for the structure determination purpose. Furthermore, the absence of phenyl group could also have the positive affect to reduce the potential structure perturbation due to the big group accommodation. We also observed there was a reserved water molecule attached closely with tellurium center, which might be another factor to facilitate the elimination of phenyl group during the high energy x-ray exposure. In terms of structure perturbation, when this Se-Te-DNA structure was superimposed over the native one (Figure 8.7a), the main global structure was considered the same, although a small content of backbone torsion was observed, probably because the accommodation of much bigger tellurium atom could slightly affect the base packing, as well as the hydration pattern. The overall r.m.s.d value is

0.407 Å, which was mainly contributed by this slight shift and the last base. The superimposed local Te-dU residue over native counterparts was also shown in figure 8.7b, which could indicate the reason of the backbone shift.

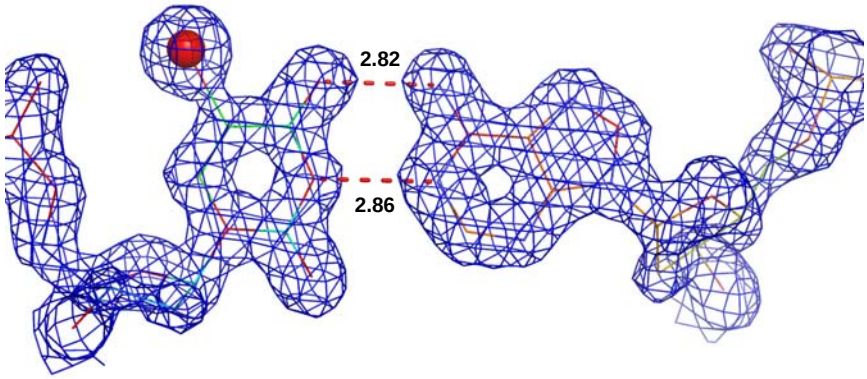


Figure 8.5: Electron density maps of Te-residue TTE and the base pair of T4-A5 in derivatized DNA. Contours are at 2.0σ level, the red ball represents the Te atom.

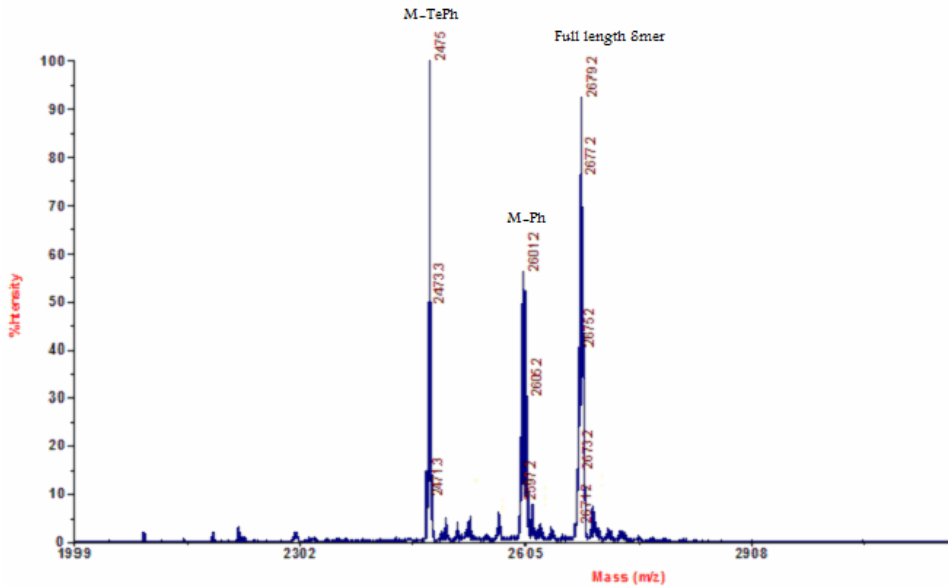


Figure 8.6: MALDI-TOF mass of Se-Te-DNA 8mer: GU(2'-SeMe)GdU(5-TePh)ACAC after irradiating with X-ray (positive mode). $C_{83}H_{101}N_{30}O_{46}P_7SeTe$, exact mass is 2678.16. MS 2601.2 represents removal of phenyl group [M-77], while 2475.3 represents removal of TePh group.

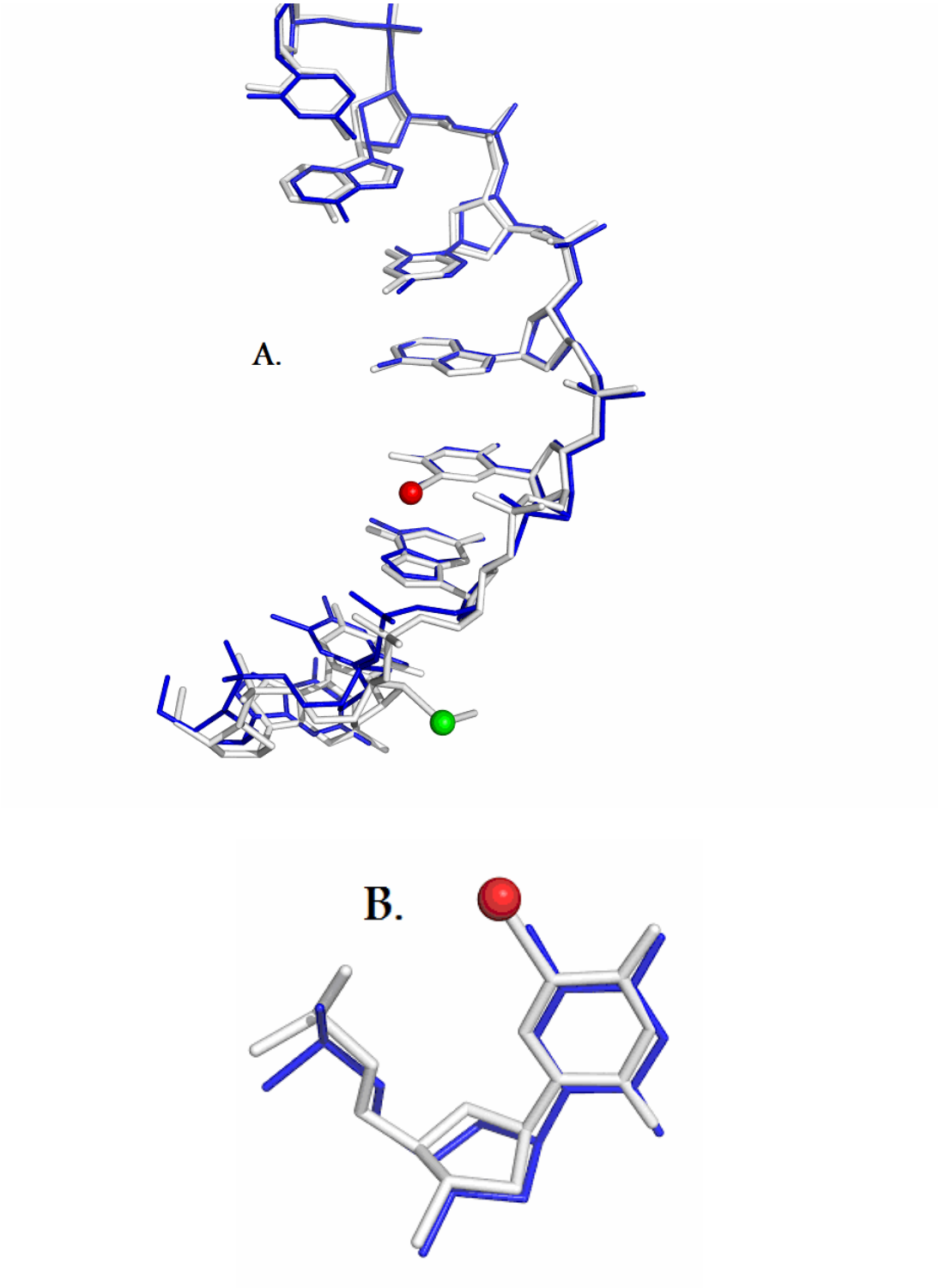


Figure 8.7: Superimposed comparison of the native and tellurium derivatized DNA structures. (1DNS and 3FA1). The red and green balls mean Te and Se respectively). (A) Global comparison; (B) Local comparison of Te-residue TTE and native T4.

8.6 Conclusion of this chapter

In summary, we have successfully developed a novel tellurium modification strategy in our process of developing new heavy atom derivatization methodologies for nucleic acids structure determination besides selenium. The 5-TePh-uridine phosphoramidite was synthesized and incorporated into several DNA oligonuelctotides through solid phase synthesizer in high yields. Our thermaldenaturation studies indicated that this tellurium modification on the nuclei-base doesn't affect the DNA duplex stabilities. The x-ray crystallography structure of a Te-DNA was determined to further demonstrate the presence of tellurium atom and the structure isomorphism of this modification comparing with native one. Following the considerable interests in the bioincorporation of tellurium methionine analogs into proteins, our site specifically chemical incorporation of tellurium functionality into nucleic acids is expected to offer similar advantages, which using tellurium to provide clear signal in both isomorphous and anomalous difference Patterson maps at the commonly used CuKα wavelengths, thus avoiding the synchrotron light source requirement. In addition, since tellurium is much more metallic comparing O, S, and Se, this method could also have potential to be applied in microscopy detection such as AFM or cryo-EM. Furthermore, our Te-DNA could also provide a model for the research of tellurium toxicity in macromolecule system.

PUBLICATIONS AND MANUSCRIPTS IN PREPARATION

1. **Jia Sheng,** Abodalla E.A. Hassan and Zhen Huang*, “First Chemical Incorporation of Tellurium Functions into Nucleic Acids for Structural and Functional Studies: Proof of Principle”. (Manuscript submitted).

2. **Jia Sheng#,** Abodalla E.A. Hassan# (equal contribution) Wen Zhang, and Zhen Huang*, “Chemical Incorporation of Tellurium Function into Nucleic Acids: A Promising Novel Apporach for X-ray Crystallography Structure Studies of Nucleic Acids”. (Manuscript submitted).

3. **Jia Sheng** and Zhen Huang*, “Oxygen Replacement with Selenium at the Uridine 4-position for Functional and Crystal Structure Studies”. (Manuscript in preparation).

4. **Jia Sheng** and Zhen Huang*, " Selenium Derivatization of Nucleic Acids: A Powerful Method for Structural and Functional Investigation", (Invited Review, in preparation).

5. **Jia Sheng,** Abodalla E.A. Hassan and Zhen Huang*, “Telluride-Mediated Elimination and Novel Synthesis of 2’,3’-Didehyro-2’,3’-dideoxynucleosides (d4Ns)”, **The Journal of Organic Chemistry**, 2008; 73(10); 3725-3729.

6. **Jia Sheng** and Zhen Huang*, "Selenium Derivatization of Nucleic Acids for Phase Determination in Nucleic Acid X-ray Crystallography", **International Journal of Molecular Sciences**, 2008, 9, 258-271.

7. **Jia Sheng** and Zhen Huang*, "Synthesis of a 4-Se-Thymidine Phosphoramidite and Its Incorporation into Oligonucleotides for Structure and Function Studies", **Current Protocols in Nucleic Acid Chemistry**, 2008, 32: 1.19.1-1.19.13.

8. **Jia Sheng,** Jiansheng Jiang, Jozef Salon, and Zhen Huang*, “Synthesis of a 2’-Se-Thymidine Phosphoramidite and Its Incorporation into Oligonucleotides for Crystal Structure Study”, **Organic Letters**, 2007, 9, 749-752.

9. Zhen Huang and **Jia Sheng**, Title: “Telluride-Mediated Elimination and Novel Synthesis of 2’,3’-Didehyro-2’,3’-dideoxynucleosides (d4Ns)”, 2008, **Patent submitted**.

10. Jozef Salon, **Jia Sheng**, Jiansheng Jiang, Guexiong Chen, Julianne Caton-Williams, Zhen Huang*, “Oxygen Replacement with Selenium at the Thymidine 4-position for the Se-Base-Paring and Crystal Structure Studies”, **Journal of American Chemical** Society, 2007, 129, 4862-4863.

11. Lina Lin, **Jia Sheng**, Razin K. Momin, Quan Du, Zhen Huang*, “Facile synthesis and anti-tumor cell activity of Se-containing nucleosides”, Nucleosides, Nucleotides **and Nucleic Acids**, 2009, 28, 56-66.

12. Jiansheng Jiang, **Jia Sheng**, Nicolas Carrasco, and Zhen Huang*, “Selenium Derivatization of Nucleic Acids for Crystallography”, **Nucleic Acids Research,** 2007, 35, 477-485.

13. Jozef Salon, Jiansheng Jiang, **Jia Sheng,** Oksana O. Gerlits, Zhen Huang*, "Derivatization of DNAs with Selenium at 6-Position of Guanine for Function and Crystal Structure Studies", **Nucleic Acids Research**, 2008, 36, 7009-7018.

APPENDIX: Nucleic Acids Mini Screen (NAM)

- **Application and Features:**

Crystallization screen for nucleic acid fragments; Can be used to screen a matrix of pH, polyamine and ions; Uses MPD as the primary precipitant.

- **Description:**

The Nucleic Acid Mini Screen is an efficient screen formulated to assist in the determination of preliminary crystallization conditions of nucleic acid fragments. The formulation is based upon the publication “A Highly Efficient 24 Condition Matrix for the Crystallization of Nucleic Acid Fragments” where the preliminary crystallization conditions of 35 nucleic acids were determined utilizing this formulation. Samples include DNA-Drug complexes, C-Tetrad and G-Quartet Motifs, RNA oligomers, and others. By using 1 to 4 mM oligonucleotide stock concentration, the screen requires less than 100 µl of sample. The screen is typically performed at 4°C although room temperature incubations can also be performed. To evaluate the effect of equilibration kinetics as well as initial and final sample concentrations, the screens are typically performed using the hanging or sitting drop vapor diffusion method with two drops on a slide (1 µl + 2 µl and 2 µl + 2 µl), side by side. The 24 well format of the mini screen provides for a fast setup and uses small amounts of sample. This makes it possible to cost-effectively screen many sequences as well as variations of a particular sequence. The composition of the Nucleic Acid Mini Screen allows one to apply the formulation to other nucleic acids such as deoxy- and ribozymes, pseudoknots, and tRNAs. Each Nucleic Acid Mini Screen kit contains 24 unique reagents, 1.0 milliliter each plus a

250 milliliter volume of dehydrant (35% v/v MPD). All solutions are formulated using ultra-pure water and are sterile filtered. The detailed compositions are listed as table A1.

Figure A1: Picture of Nucleic Acid Mini Screen from Hampton Research.

Table A1: Crystal buffer conditions of NAM.

Entry[a]	40 mM Sodium Cacodylate	Monovalent Ion	Divalent Ion or others
1	pH 5.5	None	20 mM $MgCl_2$
2	pH 5.5	80 mM NaCl	20 mM $MgCl_2$
3	pH 5.5	12 mM NaCl, 80 mM KCl	None
4	pH 5.5	40 mM LiCl	20 mM $MgCl_2$
5	pH 6.0	80 mM KCl	20 mM $MgCl_2$
6	pH 6.0	80 mM KCl	None
7	pH 6.0	80 mM NaCl	20 mM $MgCl_2$
8	pH 6.0	80 mM NaCl	None

9	pH 6.0	80 mM NaCl, 12 mM KCl	20 mM $MgCl_2$
10	pH 6.0	12 mM NaCl, 80 mM KCl	None
11	pH 6.0	80 mM NaCl	20 mM $BaCl_2$
12	pH 6.0	80 mM KCl	20 mM $BaCl_2$
13	pH 6.0	none	80 mM $SrCl_2$
14	pH 7.0	80 mM KCl	20 mM $MgCl_2$
15	pH 7.0	80 mM KCl	None
16	pH 7.0	80 mM NaCl	20 mM $MgCl_2$
17	pH 7.0	80 mM NaCl	None
18	pH 7.0	80 mM NaCl, 12 mM KCl	20 mM $MgCl_2$
19	pH 7.0	12 mM NaCl, 80 mM KCl	None
20	pH 7.0	80 mM NaCl	20 mM $BaCl_2$
21	pH 7.0	80 mM KCl	20 mM $BaCl_2$
22	pH 7.0	40 mM LiCl	80 mM $SrCl_2$ 20 mM $MgCl_2$
23	pH 7.0	40 mM LiCl	80 mM $SrCl_2$
24	pH 7.0	None	80 mM $SrCl_2$ 20 mM $MgCl_2$

a: #1-4 buffers contain 10% (v/v) MPD, and 20 mM cobalt hexamine chloride [$Co(NH_3)_6Cl_3$]. #5-24 buffers contain 10% (v/v) MPD, and 12 mM Spermine tetra-HCl. Entry numbers are the same as the Hampton kit number (Nucleic Acid Mini Screen Kit, http://www.hamptonresearch.com/).

REFERENCES

1. Storz, G. (2002) An expanding universe of noncoding RNAs. *Science*, **296**, 1260-1263.
2. Eddy, S.R. (2001) Non-coding RNA genes and the modern RNA world. *Nat Rev Genet*, **2**, 919-929.
3. Blount, K.F. and Uhlenbeck, O.C. (2005) The structure-function dilemma of the hammerhead ribozyme. *Annu Rev Biophys Biomol Struct*, **34**, 415-440.
4. Chan, J.H., Lim, S. and Wong, W.S. (2006) Antisense oligonucleotides: from design to therapeutic application. *Clin Exp Pharmacol Physiol*, **33**, 533-540.
5. Dallas, A. and Vlassov, A.V. (2006) RNAi: a novel antisense technology and its therapeutic potential. *Med Sci Monit*, **12**, RA67-74.
6. Boehr, D.D., Dyson, H.J. and Wright, P.E. (2006) An NMR perspective on enzyme dynamics. *Chem Rev*, **106**, 3055-3079.
7. Egli, M. (2004) Nucleic acid crystallography: current progress. *Curr Opin Chem Biol*, **8**, 580-591.
8. Egli, M. and Pallan, P.S. (2007) Insights from crystallographic studies into the structural and pairing properties of nucleic acid analogs and chemically modified DNA and RNA oligonucleotides. *Annu Rev Biophys Biomol Struct*, **36**, 281-305.
9. Ferre-D'Amare, A.R., Zhou, K. and Doudna, J.A. (1998) Crystal structure of a hepatitis delta virus ribozyme. *Nature*, **395**, 567-574.
10. Yang, W., Hendrickson, W.A., Crouch, R.J. and Satow, Y. (1990) Structure of ribonuclease H phased at 2 A resolution by MAD analysis of the selenomethionyl protein. *Science*, **249**, 1398-1405.

11. Xiong, Y. and Sundaralingam, M. (2000) Crystal structure of a DNA.RNA hybrid duplex with a polypurine RNA r(gaagaagag) and a complementary polypyrimidine DNA d(CTCTTCTTC). *Nucleic Acids Res*, **28**, 2171-2176.

12. Boggon, T.J. and Shapiro, L. (2000) Screening for phasing atoms in protein crystallography. *Structure*, **8**, R143-149.

13. Garman, E. and Murray, J.W. (2003) Heavy-atom derivatization. *Acta Crystallogr D Biol Crystallogr*, **59**, 1903-1913.

14. Shah, K., Wu, H. and Rana, T.M. (1994) Synthesis of uridine phosphoramidite analogs: reagents for site-specific incorporation of photoreactive sites into RNA sequences. *Bioconjug Chem*, **5**, 508-512.

15. Ennifar, E., Carpentier, P., Ferrer, J.L., Walter, P. and Dumas, P. (2002) X-ray-induced debromination of nucleic acids at the Br K absorption edge and implications for MAD phasing. *Acta Crystallogr D Biol Crystallogr*, **58**, 1262-1268.

16. Berzelius, J.J., M., L.d. and M., B.a. (1817) Berthollet sur deux metaux nouveaus. *Ann. Chim. Phys. Ser.*, **2**, 199-121.

17. Stadtman, T.C. (1974) Selenium biochemistry. *Science*, **183**, 915-922.

18. Stadtman, T.C. (1996) Selenocysteine. *Annu Rev Biochem*, **65**, 83-100.

19. Stadtman, T.C. (2000) Selenium biochemistry. Mammalian selenoenzymes. *Ann N Y Acad Sci*, **899**, 399-402.

20. Zhang, J., Wang, X. and Xu, T. (2008) Elemental selenium at nano size (Nano-Se) as a potential chemopreventive agent with reduced risk of selenium toxicity: comparison with se-methylselenocysteine in mice. *Toxicol Sci*, **101**, 22-31.

21. Gladyshev, V.N. and Hatfield, D.L. (1999) Selenocysteine-containing proteins in

mammals. *J Biomed Sci*, **6**, 151-160.

22. Hatfield, D.L. and Gladyshev, V.N. (2002) How selenium has altered our understanding of the genetic code. *Mol Cell Biol*, **22**, 3565-3576.
23. Zinoni, F., Birkmann, A., Stadtman, T.C. and Bock, A. (1986) Nucleotide sequence and expression of the selenocysteine-containing polypeptide of formate dehydrogenase (formate-hydrogen-lyase-linked) from Escherichia coli. *Proc Natl Acad Sci U S A*, **83**, 4650-4654.
24. Bock, A., Forchhammer, K., Heider, J., Leinfelder, W., Sawers, G., Veprek, B. and Zinoni, F. (1991) Selenocysteine: the 21st amino acid. *Mol Microbiol*, **5**, 515-520.
25. Leinfelder, W., Forchhammer, K., Zinoni, F., Sawers, G., Mandrand-Berthelot, M.A. and Bock, A. (1988) Escherichia coli genes whose products are involved in selenium metabolism. *J Bacteriol*, **170**, 540-546.
26. Hoffman, J.L. and McConnell, K.P. (1974) The presence of 4-selenouridine in Escherichia coli tRNA. *Biochim Biophys Acta*, **366**, 109-113.
27. Veres, Z., Tsai, L., Scholz, T.D., Politino, M., Balaban, R.S. and Stadtman, T.C. (1992) Synthesis of 5-methylaminomethyl-2-selenouridine in tRNAs: 31P NMR studies show the labile selenium donor synthesized by the selD gene product contains selenium bonded to phosphorus. *Proc Natl Acad Sci U S A*, **89**, 2975-2979.
28. Veres, Z. and Stadtman, T.C. (1994) A purified selenophosphate-dependent enzyme from Salmonella typhimurium catalyzes the replacement of sulfur in 2-thiouridine residues in tRNAs with selenium. *Proc Natl Acad Sci U S A*, **91**, 8092-8096.

29. Mihara, H., Kato, S., Lacourciere, G.M., Stadtman, T.C., Kennedy, R.A., Kurihara, T., Tokumoto, U., Takahashi, Y. and Esaki, N. (2002) The iscS gene is essential for the biosynthesis of 2-selenouridine in tRNA and the selenocysteine-containing formate dehydrogenase H. *Proc Natl Acad Sci U S A*, **99**, 6679-6683.

30. Ching, W.M., Tsai, L. and Wittwer, A.J. (1985) Selenium-containing transfer RNAs. *Curr Top Cell Regul*, **27**, 497-507.

31. Wolfe, M.D., Ahmed, F., Lacourciere, G.M., Lauhon, C.T., Stadtman, T.C. and Larson, T.J. (2004) Functional diversity of the rhodanese homology domain: the Escherichia coli ybbB gene encodes a selenophosphate-dependent tRNA 2-selenouridine synthase. *J Biol Chem*, **279**, 1801-1809.

32. Metanis, N., Keinan, E. and Dawson, P.E. (2006) Synthetic seleno-glutaredoxin 3 analogues are highly reducing oxidoreductases with enhanced catalytic efficiency. *J Am Chem Soc*, **128**, 16684-16691.

33. Starks, C.M., Francois, J.A., MacArthur, K.M., Heard, B.Z. and Kappock, T.J. (2007) Atomic-resolution crystal structure of thioredoxin from the acidophilic bacterium Acetobacter aceti. *Protein Sci*, **16**, 92-98.

34. Ip, C. and Ganther, H.E. (1992) Comparison of selenium and sulfur analogs in cancer prevention. *Carcinogenesis*, **13**, 1167-1170.

35. Allmang, C. and Krol, A. (2006) Selenoprotein synthesis: UGA does not end the story. *Biochimie*, **88**, 1561-1571.

36. Zhong, L., Arner, E.S. and Holmgren, A. (2000) Structure and mechanism of mammalian thioredoxin reductase: the active site is a redox-active

selenolthiol/selenenylsulfide formed from the conserved cysteine-selenocysteine sequence. *Proc Natl Acad Sci U S A*, **97**, 5854-5859.

37. Kim, H.Y. and Gladyshev, V.N. (2005) Different catalytic mechanisms in mammalian selenocysteine- and cysteine-containing methionine-R-sulfoxide reductases. *PLoS Biol*, **3**, e375.

38. Aboul-Fadl, T. (2005) Selenium derivatives as cancer preventive agents. *Curr Med Chem Anticancer Agents*, **5**, 637-652.

39. Hendrickson, W.A., Smith, J.L. and Sheriff, S. (1985) Direct phase determination based on anomalous scattering. *Methods Enzymol*, **115**, 41-55.

40. Hendrickson, W.A., Pahler, A., Smith, J.L., Satow, Y., Merritt, E.A. and Phizackerley, R.P. (1989) Crystal structure of core streptavidin determined from multiwavelength anomalous diffraction of synchrotron radiation. *Proc Natl Acad Sci U S A*, **86**, 2190-2194.

41. Hendrickson, W.A., Smith, J.L., Phizackerley, R.P. and Merritt, E.A. (1988) Crystallographic structure analysis of lamprey hemoglobin from anomalous dispersion of synchrotron radiation. *Proteins*, **4**, 77-88.

42. Hendrickson, W.A. (1985) Stereochemically restrained refinement of macromolecular structures. *Methods Enzymol*, **115**, 252-270.

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

44. Hendrickson, W.A. (2000) Synchrotron crystallography. *Trends Biochem Sci*, **25**, 637-643.

45. Cowie, D.B. and Cohen, G.N. (1957) Biosynthesis by Escherichia coli of active altered proteins containing selenium instead of sulfur. *Biochim Biophys Acta*, **26**, 252-261.

46. Hendrickson, W.A., Horton, J.R., Murthy, H.M., Pahler, A. and Smith, J.L. (1989) Multiwavelength anomalous diffraction as a direct phasing vehicle in macromolecular crystallography. *Basic Life Sci*, **51**, 317-324.

47. Hendrickson, W.A., Horton, J.R. and LeMaster, D.M. (1990) Selenomethionyl proteins produced for analysis by multiwavelength anomalous diffraction (MAD): a vehicle for direct determination of three-dimensional structure. *EMBO J*, **9**, 1665-1672.

48. Deacon, A.M. and Ealick, S.E. (1999) Selenium-based MAD phasing: setting the sites on larger structures. *Structure*, **7**, R161-166.

49. Jiang, J., Sheng, J., Carrasco, N. and Huang, Z. (2007) Selenium derivatization of nucleic acids for crystallography. *Nucleic Acids Res*, **35**, 477-485.

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

51. Egli, M.P., P. S.; Pattanayek, R.; Wilds, C. J.; Lubini, P.; Minasov, G.; Dobler, M.; Leumann, C. J.; Eschenmoser, A. (2006) Crystal Structure of Homo-DNA and Nature's Choice of Pentose over Hexose in the Genetic System. *J. Am. Chem. Soc.*, **128**, 10847-10856.

52. Serganov, A., Keiper, S., Malinina, L., Tereshko, V., Skripkin, E., Hobartner, C., Polonskaia, A., Phan, A.T., Wombacher, R., Micura, R. *et al.* (2005) Structural basis for Diels-Alder ribozyme-catalyzed carbon-carbon bond formation. *Nat Struct Mol Biol*, **12**, 218-224.

53. Carrasco, N. and Huang, Z. (2004) Enzymatic synthesis of phosphoroselenoate DNA using thymidine 5'-(alpha-P-seleno)triphosphate and DNA polymerase for X-ray crystallography via MAD. *J Am Chem Soc*, **126**, 448-449.

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

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

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

57. Salon, J., Chen, G., Portilla, Y., Germann, M.W. and Huang, Z. (2005) Synthesis of a 2'-Se-uridine phosphoramidite and its incorporation into oligonucleotides for structural study. *Org Lett*, **7**, 5645-5648.

58. Hobartner, C., Rieder, R., Kreutz, C., Puffer, B., Lang, K., Polonskaia, A., Serganov, A. and Micura, R. (2005) Syntheses of RNAs with up to 100

nucleotides containing site-specific 2'-methylseleno labels for use in X-ray crystallography. *J Am Chem Soc*, **127**, 12035-12045.

59. Moroder, H., Kreutz, C., Lang, K., Serganov, A. and Micura, R. (2006) Synthesis, oxidation behavior, crystallization and structure of 2'-methylseleno guanosine containing RNAs. *J Am Chem Soc*, **128**, 9909-9918.

60. Dande, P., Prakash, T.P., Sioufi, N., Gaus, H., Jarres, R., Berdeja, A., Swayze, E.E., Griffey, R.H. and Bhat, B. (2006) Improving RNA interference in mammalian cells by 4'-thio-modified small interfering RNA (siRNA): effect on siRNA activity and nuclease stability when used in combination with 2'-O-alkyl modifications. *J Med Chem*, **49**, 1624-1634.

61. Inagaki, Y., Minakawa, N. and Matsuda, A. (2007) Synthesis of 4'-selenoribonucleosides. *Nucleic Acids Symp Ser (Oxf)*, 139-140.

62. Jeong, L.S., Tosh, D.K., Kim, H.O., Wang, T., Hou, X., Yun, H.S., Kwon, Y., Lee, S.K., Choi, J. and Zhao, L.X. (2008) First synthesis of 4'-selenonucleosides showing unusual Southern conformation. *Org Lett*, **10**, 209-212.

63. Jayakanthan, K., Johnston, B.D. and Pinto, B.M. (2008) Stereoselective synthesis of 4'-selenonucleosides using the Pummerer glycosylation reaction. *Carbohydr Res*, **343**, 1790-1800.

64. Watts, J.K., Johnston, B.D., Jayakanthan, K., Wahba, A.S., Pinto, B.M. and Damha, M.J. (2008) Synthesis and biophysical characterization of oligonucleotides containing a 4'-selenonucleotide. *J Am Chem Soc*, **130**, 8578-8579.

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

66. Mori, K.B., C.; Cazenave, C.; Marsukura, M.; Subasinghe, C.; Cohen, and J. S.; Broder, S.T., J. J.; Stein, C. A. (1989) Phosphoroselenoate oligodeoxynucleotides: synthesis, physico-chemical characterization, anti-sense inhibitory properties and anti-HIV activity. *Nucleic Acids Res.* , **17**, 8207-8219.

67. Egli, M., Lubini, P. and Pallan, P.S. (2007) The long and winding road to the structure of homo-DNA. *Chem Soc Rev*, **36**, 31-45.

68. Holloway, G.A., Pavot, C., Scaringe, S.A., Lu, Y. and Rauchfuss, T.B. (2002) An organometallic route to oligonucleotides containing phosphoroselenoate. *Chembiochem*, **3**, 1061-1065.

69. Tram, K., Wang, X. and Yan, H. (2007) Facile synthesis of oligonucleotide phosphoroselenoates. *Org Lett*, **9**, 5103-5106.

70. Guga, P., Maciaszek, A. and Stec, W.J. (2005) Oxathiaphospholane approach to the synthesis of oligodeoxyribonucleotides containing stereodefined internucleotide phosphoroselenoate function. *Org Lett*, **7**, 3901-3904.

71. Stec, W.J., Karwowski, B., Boczkowska, M., Guga, P., Koziołkiewicz, M., Sochacki, M., Wieczorek, M.W. and Błaszczyk, J. (1998) Deoxyribonucleoside 3‘-O-(2-Thio- and 2-Oxo-“spiro”-4,4-pentamethylene-1,3,2-oxathiaphospholane)s: Monomers for Stereocontrolled Synthesis of Oligo(deoxyribonucleoside phosphorothioate)s and Chimeric PS/PO Oligonucleotides. *J. Am. Chem. Soc.,*, **120**, 7156–7167.

72. Salon, J., Sheng, J., Jiang, J., Chen, G., Caton-Williams, J. and Huang, Z. (2007) Oxygen replacement with selenium at the thymidine 4-position for the Se base pairing and crystal structure studies. *J Am Chem Soc*, **129**, 4862-4863.

73. Sintim, H.O. and Kool, E.T. (2006) Enhanced base pairing and replication efficiency of thiothymidines, expanded-size variants of thymidine. *J Am Chem Soc*, **128**, 396-397.

74. Sintim, H.O. and Kool, E.T. (2006) Remarkable sensitivity to DNA base shape in the DNA polymerase active site. *Angew Chem Int Ed Engl*, **45**, 1974-1979.

75. Sismour, A.M. and Benner, S.A. (2005) The use of thymidine analogs to improve the replication of an extra DNA base pair: a synthetic biological system. *Nucleic Acids Res*, **33**, 5640-5646.

76. Milne, G.H. and Townsend, L.B. (1974) Synthesis and antitumor activity of alpha- and beta-2'-deoxy-6-selenoguanosine and certain related derivatives. *J Med Chem*, **17**, 263-268.

77. Chen, C.S. and Stadtman, T.C. (1980) Selenium-containing tRNAs from Clostridium sticklandii: cochromatography of one species with L-prolyl-tRNA. *Proc Natl Acad Sci U S A*, **77**, 1403-1407.

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

79. Mautner, H.G., Chu, S.H., Jaffe, J.J. and Sartorelli, A.C. (1963) The Synthesis and Antineoplastic Properties of Selenoguanine, Selenocytosine and Related Compounds. *J Med Chem*, **6**, 36-39.

80. Kadokura, M., Wada, T., Seio, K. and Sekine, M. (2000) Synthesis of 4-thiouridine, 6-thioinosine, and 6-thioguanosine 3',5'-O-bisphosphates as donor molecules for RNA ligation and their application to the synthesis of photoactivatable TMG-capped U1 snRNA fragments. *J Org Chem*, **65**, 5104-5113.

81. Carrasco, N., Caton-Williams, J., Brandt, G., Wang, S. and Huang, Z. (2006) Efficient enzymatic synthesis of phosphoroselenoate RNA by using adenosine 5'-(alpha-P-seleno)triphosphate. *Angew Chem Int Ed Engl*, **45**, 94-97.

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

83. Caton-Williams, J. and Huang, Z. (2008) Synthesis and DNA-polymerase incorporation of colored 4-selenothymidine triphosphate for polymerase recognition and DNA visualization. *Angew Chem Int Ed Engl*, **47**, 1723-1725.

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

85. Moroder, L. (2005) Isosteric replacement of sulfur with other chalcogens in peptides and proteins. *J Pept Sci*, **11**, 187-214.

86. Ramadan, S.E., Razak, A.A., Ragab, A.M. and el-Meleigy, M. (1989) Incorporation of tellurium into amino acids and proteins in a tellurium-tolerant fungi. *Biol Trace Elem Res*, **20**, 225-232.

87. Boles, J.O., Lewinski, K., Kunkle, M., Odom, J.D., Dunlap, B., Lebioda, L. and Hatada, M. (1994) Bio-incorporation of telluromethionine into buried residues of dihydrofolate reductase. *Nat Struct Biol*, **1**, 283-284.

88. Budisa, N., Karnbrock, W., Steinbacher, S., Humm, A., Prade, L., Neuefeind, T., Moroder, L. and Huber, R. (1997) Bioincorporation of telluromethionine into proteins: a promising new approach for X-ray structure analysis of proteins. *J Mol Biol*, **270**, 616-623.

89. Budisa, N., Steipe, B., Demange, P., Eckerskorn, C., Kellermann, J. and Huber, R. (1995) High-level biosynthetic substitution of methionine in proteins by its analogs 2-aminohexanoic acid, selenomethionine, telluromethionine and ethionine in Escherichia coli. *Eur J Biochem*, **230**, 788-796.

90. Farina, M., Soares, F.A., Zeni, G., Souza, D.O. and Rocha, J.B. (2004) Additive pro-oxidative effects of methylmercury and ebselen in liver from suckling rat pups. *Toxicol Lett*, **146**, 227-235.

91. Karnbrock, W., Weyher, E., Budisa, N., Huber, R. and Moroder, L. (1996) A New Efficient Synthesis of Acetyltelluro- and Acetylselenomethionine and Their Use in the Biosynthesis of Heavy-Atom Protein Analogs. *J. Am. Chem. Soc.*, **118**, 913-914.

92. Puglisi, J.D. and Tinoco, I., Jr. (1989) Absorbance melting curves of RNA. *Methods Enzymol*, **180**, 304-325.

93. Otwinowski, Z. and Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Meth. Enzymol.*, **276**, 307-326.

94. 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. *et al.* (1998) Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr D Biol Crystallogr*, **54**, 905-921.

95. Parkinson, G., Vojtechovsky, J., Clowney, L., Brunger, A.T. and Berman, H.M. (1996) New parameters for the refinement of nucleic acid-containing structures. *Acta Crystallogr D Biol Crystallogr*, **52**, 57-64.

96. Brunger, A.T. (1992) Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature*, **355**, 472-475.

97. Read, R.J. (1986) Improved Fourier coefficients for maps using phases from partial structures with errors. *Acta Cryst.*, **A42**, 140-149.

98. Becker, H.F., Motorin, Y., Florentz, C., Giege, R. and Grosjean, H. (1998) Pseudouridine and ribothymidine formation in the tRNA-like domain of turnip yellow mosaic virus RNA. *Nucleic Acids Res*, **26**, 3991-3997.

99. Sprinzl, M. and Vassilenko, K.S. (2005) Compilation of tRNA sequences and sequences of tRNA genes. *Nucleic Acids Res*, **33**, D139-140.

100. McCloskey, J.A. and Rozenski, J. (2005) The Small Subunit rRNA Modification Database. *Nucleic Acids Res*, **33**, D135-138.

101. Parkinson, G.N., Lee, M.P. and Neidle, S. (2002) Crystal structure of parallel quadruplexes from human telomeric DNA. *Nature*, **417**, 876-880.

102. Ikehara, M., Uesugi, S. and Kaneko, M. (1978) In Townsend, L. B. a. T., R.S. (ed.), *Nucleic Acid Chemistry*. Wiley, NY, Vol. 2, pp. 837-841.

103. Tavale, S.S. and Sobell, H.M. (1970) Crystal and molecular structure of 8-bromoguanosine and 8-bromoadenosine, two purine nucleosides in the syn conformation. *J Mol Biol*, **48**, 109-123.

104. Willis, M.C., Hicke, B.J., Uhlenbeck, O.C., Cech, T.R. and Koch, T.H. (1993) Photocrosslinking of 5-iodouracil-substituted RNA and DNA to proteins. *Science*, **262**, 1255-1257.

105. Thota, N., Li, X.H., Bingman, C. and Sundaralingam, M. (1993) High-resolution refinement of the hexagonal A-DNA octamer d(GTGTACAC) at 1.4 A. *Acta Crystallogr D Biol Crystallogr*, **49**, 282-291.

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

107. Latham, M.P., Brown, D.J., McCallum, S.A. and Pardi, A. (2005) NMR methods for studying the structure and dynamics of RNA. *Chembiochem*, **6**, 1492-1505.

108. Lezius, A.G. and Scheit, K.H. (1967) Enzymatic synthesis of DNA with 4-thio-thymidine triphosphate as substitute for dTTP. *Eur J Biochem*, **3**, 85-94.

109. Sprinzl, M., Scheit, K.H. and Cramer, F. (1973) Preparation in vitro of a 2-thiocytidine-containing yeast tRNA Phe -A 73 -C 74 -S 2 C 75 -A 76 and its interaction wiith p-hydroxymercuribenzoate. *Eur J Biochem*, **34**, 306-310.

110. Kutyavin, I.V., Rhinehart, R.L., Lukhtanov, E.A., Gorn, V.V., Meyer, R.B., Jr. and Gamper, H.B., Jr. (1996) Oligonucleotides containing 2-aminoadenine and 2-thiothymine act as selectively binding complementary agents. *Biochemistry*, **35**, 11170-11176.

111. Shiue, C.Y. and Chu, S.H. (1975) A facile synthesis of 1-beta-D-arabinofuranosyl-2-seleno- and -4-selenouracil and related compounds. *J Org Chem*, **40**, 2971-2974.

112. Logan, G., Igunbor, C., Chen, G.-X., Davis, H., Simon, A., Salon, J. and Huang, Z. (2006) A Novel and Simple Strategy for Incorporation, Protection, and Deprotection of Selenium Functionality. *Synlett*, **10**, 1554.

113. Dueymes, C., Decout, J.L., Peltie, P. and Fontecave, M. (2002) Fluorescent deazaflavin-oligonucleotide probes for selective detection of DNA. *Angew Chem Int Ed Engl*, **41**, 486-489.

114. Kim, H.K., Liu, J., Li, J., Nagraj, N., Li, M., Pavot, C.M. and Lu, Y. (2007) Metal-dependent global folding and activity of the 8-17 DNAzyme studied by fluorescence resonance energy transfer. *J Am Chem Soc*, **129**, 6896-6902.

115. Bunkenborg, J., Gadjev, N.I., Deligeorgiev, T. and Jacobsen, J.P. (2000) Concerted intercalation and minor groove recognition of DNA by a homodimeric thiazole orange dye. *Bioconjug Chem*, **11**, 861-867.

116. Karunakaran, V., Perez Lustres, J.L., Zhao, L., Ernsting, N.P. and Seitz, O. (2006) Large dynamic Stokes shift of DNA intercalation dye Thiazole Orange has contribution from a high-frequency mode. *J Am Chem Soc*, **128**, 2954-2962.

117. Elghanian, R., Storhoff, J.J., Mucic, R.C., Letsinger, R.L. and Mirkin, C.A. (1997) Selective colorimetric detection of polynucleotides based on the distance-dependent optical properties of gold nanoparticles. *Science*, **277**, 1078-1081.

118. Marti, A.A., Puckett, C.A., Dyer, J., Stevens, N., Jockusch, S., Ju, J., Barton, J.K. and Turro, N.J. (2007) Inorganic-organic hybrid luminescent binary probe for

DNA detection based on spin-forbidden resonance energy transfer. *J Am Chem Soc*, **129**, 8680-8681.

119. Ni, Z., Kim, E.D., Ha, M., Lackey, E., Liu, J., Zhang, Y., Sun, Q. and Chen, Z.J. (2009) Altered circadian rhythms regulate growth vigour in hybrids and allopolyploids. *Nature*, **457**, 327-331.

120. Nishiuchi, K., Kitaura, H., Yamada, N. and Akahira, N. (1998) Dual-Layer Optical Disk with Te–O–Pd Phase-Change Film. *Jpn. J. Appl. Phys.*, **37**, 2163-2167.