NOTE TO USERS

SYNTHESIS OF DNA CONTAINING SELENIUM FOR NUCLEIC ACID X-RAY CRYSTALLOGRAPHY

by

QUAN DU

A dissertation submitted to the Graduate Faculty in
Biochemistry in partial fulfillment of the requirements
for the degree of Doctor of Philosophy
The City University of New York

2005

INFORMATION TO USERS

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleed-through, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

ProQuest Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

This manuscript has been read and accepted for the Graduate Faculty in Biochemistry in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

04/08/2005
Date

Chair of Examining Committee

April 19, 2005
Date

Executive Officer

Supervisory Committee

The City University of New York

ABSTRACT

SYNTHESIS OF DNA CONTAINING SELENIUM FOR NUCLEIC ACID X-RAY CRYSTALLOGRAPHY

By

Quan Du

Adviser: Professor Zhen Huang

X-ray crystallography is arguably the most powerful experimental tool to determine high-resolution structures of biomacromolecules. In protein crystallography, a popular methodology named multiwavelength anomalous dispersion (MAD) requires the incorporation of selenomethionine during peptide biosynthesis. However, no equivalent methodology was developed for nucleic acids prior to this work.

In our studies, new nucleoside analogues containing selenium functionalities at 2', 3', and 5' positions were chemically synthesized. Some of these modified nucleoside monomers were incorporated site-specifically

to oligonucleotides through phosphoramidite chemistry. Similar to selenomethionine derivatization in protein X-ray crystallography, selenium-derivatized nucleosides may be used for X-ray crystal structure determination of nucleic acids and nucleic acid-protein complexes, by the MAD method.

For the first time, we have successfully demonstrated this new strategy by covalently derivatizing nucleotides with selenium at the 2'-position for phase and structure determination. The successful incorporation and the stability of selenium functional group were verified by NMR, mass spectrometry, X-ray fluorescence and X-ray crystallography. Experimental results show the 2'-deoxy-methylselenouridine residue displays C3'-endo conformation in oligonucleotides, which is characteristic of RNA and A-form DNA double helix. Moreover, this new method provides a rational way to make isomorphous nucleic acid crystals for X-ray crystallography, which was carried out previously by heavy metal soaking.

ACKNOWLEDGEMENTS

I would like to thank my mentor, Professor Zhen Huang, for his guidance during my graduate research. I was inspired by Professor Huang's dedication to scientific research, strong work ethics and rigorous pursuit for scientific excellence. The work achievement in this thesis would not be possible without his enormous effort spent in the past. It was quite an intriguing journey to see how the laboratory expanded and matured during these years. I give my best wishes to Professor Huang and the laboratory in the future.

Many thanks to all my graduate committee members, Professor Lesley Davenport and Professor Richard Magliozzo at Brooklyn College, Professor Xiangpeng Kong at New York University Medical Center, Professor Wilson Xu at Memorial Sloan Kettering Cancer Institute. Professor Davenport, also the Executive Officer of Biochemistry Ph.D. Program at the Graduate Center, offered lots of help and academic advice regarding my graduate studies. Professor Magliozzo kindly provided access to his laboratory resources all the time, and gave selfless help at personal level as well as professional level. Professor Kong offered great help in handling my DNA crystals, and worked out a partial crystal structure of our synthetic DNA oligonucleotide analog. Professor Xu consulted some molecular biology experiments for my early project; in addition, he generously provided me with a number of plasmid constructs and yeast strain.

Many thanks to these chemistry professionals for their constant technical assistance: Professor Herman Zieger maintained and provided trouble-shooting for the NMR facility at Brooklyn College; Dr. Hsin Wang at College of Staten Island, Dr. Michael Blumenstein at Hunter College provided me machine time at their high-field NMR facilities; Dr. Clifford Soll at Hunter College delivered numerous trials of mass spectrometry for my samples.

I owe a debt of gratitude to my parents for my being away so long. My wife made considerable sacrifice and endured much in recent years. I thank them deeply for the support of this work.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

A	Adenine; or Adenosine
Ara-C	Arabinosylcytidine
AZT	3'-azido-3'-deoxythymidine
Ar	Aromatic
Bz	Benzoyl
C	Cytosine; or Cytidine
CD	Circular dichroism
COSY	Correlated spectroscopy
CPG	Controlled pore glass
ddC	2',3'-dideoxycytidine
DMF	N, N'-Dimethylformamide
DMTr	Dimethoxytrityl
DNA	Deoxyribonucleic acid
DNG	Deoxyribonucleic guanidine
ESI-MS	Electrospray ionization mass spectrometry
FRET	Fluorescence resonance energy transfer
FT	Fourier transform

FUDR	5-fluoro-2'-deoxyuridine
G	Guanine; or Guanosine
GOESY	Gradient NOE spectroscopy
HIV	Human immunodeficiency virus
HMQC	Heteronuclear multiple quantum correlation
HNA	Hexitol-nucleic acid
HPLC	High pressure liquid chromatography
I	Inosine
IR	Infrared
LNA	Locked nucleic acid
MAD	Multiwavelength anomalous dispersion
MALDI	Matrix assisted laser desorption/ionization
MIR	Multiple isomorphous replacement
Ms	Mesyl; Methanesulfonyl
MS	Mass spectrometry
NDB	Nucleic Acid Database
NMR	Nuclear magnetic resonance
NOE	Nuclear Overhauser Effect
NOESY	NOE spectroscopy

PCR	Polymerase chain reaction
PDB	Protein Data Bank
PNA	Peptide nucleic acid
RNA	Ribonucleic acid
RNAi	RNA interference
mRNA	Messenger RNA
rRNA	Ribosomal RNA
snoRNA	Small nucleolar RNA
tRNA	Transfer RNA
RNPs	Ribonucleoproteins
RT	Reverse transcriptase
Se	Selenium
T	Thymine; or Thymidine
TBDMS	*tert*-Butyldimethylsilyl
3TC	2',3'-dideoxy-3'-thia-L-cytidine
TEA	Triethylamine
THF	Tetrahydrofuran
TLC	Thin layer chromatography
T_m	Melting temperature

TNA	α-L-Threofuranosyl-(3'→2')-nucleic acid
U_{MeO}	2'-methoxy-uridine
U_{Se}	2'-deoxy-2'-methylselenide uridine
UV/VIS	Ultraviolet/Visible

Chapter 1. Structural Biology of Nucleic Acids

1.1 The 50th anniversary of DNA double helix

> The molecule is so beautiful. Its glory was reflected on Francis and me.
> —— James Watson

On February 28, 1953, Francis Crick walked into the Eagle pub in Cambridge, England, and announced to everyone within hearing distance that he and James Watson had "found the secret of life". At least that's what Watson remembers; Crick's memory is different. The exact words don't matter that much because the fact is, they had done it.(1) Earlier that day, they had built a model of double helix DNA that showed by its very structure how it carries the genetic code and thus the key molecule of heredity, developmental biology and evolution.(2)

They were not the only people interested in solving DNA structure during early 1950s. Maurice Wilkins and Rosalind Franklin at King's College, London, had collected diffraction patterns of two morphologically different DNA types – B form and A form,(3,4) from X-ray crystallography of fibrous DNA. Nobel Laureate Linus Pauling proposed a three-stranded structure for DNA that placed the phosphate backbone in the center of a

helix and the bases facing outwards.(5,6) Pauling's proposal relied on X-ray crystallographs for hints at the molecular level, but depended more heavily on scaled-up models he built by hand. Pauling's DNA model was considered unlikely for electrostatic conflicts, and the double-stranded structure proposed by Watson and Crick was later proved to be correct.

As convincing as this landmark discovery was, there was much left unanswered. Watson and Crick's model only provides an average view of the DNA structure based on limited fiber diffraction data, and is heavily model dependent. In fact, the poor fit of this model with R factor of about 0.80 indicated to crystallographers a structure so erroneous as to be beyond rescue.(7) The possibility of other structures could not be fully ruled out. Some doubt would remain until more definitive structural evidence for the double helix could be produced.(8)

With the fast progress in molecular biology and the development in syntheic methods for nucleic acid, DNA or RNA homopolymers and oligonucleotides with defined sequence became available. The first concrete affirmation of the double-helical geometry came in 1973 by Alexander Rich

and coworkers.(9) The crystal structures of ApU and GpC showed the double helix at atomic resolution. Both structures were of RNA fragments.

In 1979, a left-handed double helix named Z-DNA was determined by X-ray crystallography of d(CGCGCG) by Rich's group,(10,11) and was further confirmed by the structure of d(CGCG) by Richard Dickerson's group. Prior to the characterization of Z-DNA, the fiber diffraction X-ray structures of naturally occurring and synthetic DNA were all right-handed helices, e.g. B-DNA and A-DNA. The discovery of Z-DNA was unusual in that it was not only left-handed, but also the first detailed structure of oligonucleotide determined by X-ray diffraction of single-crystals.

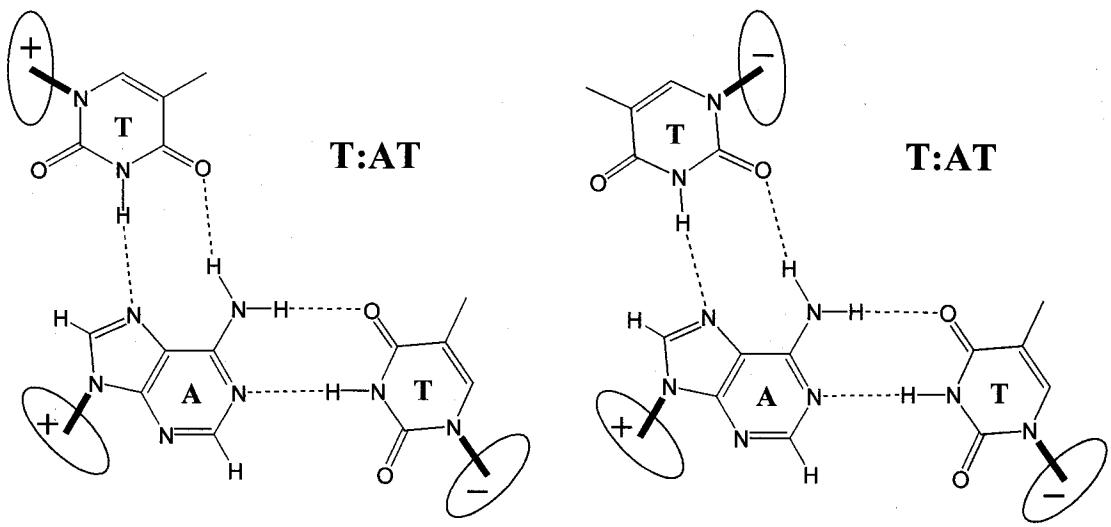
The first single-crystal structures of B-DNA were determined for dodecamer d(CGCGAATTCGCG) by Dickerson's group in 1980.(12,13) Any doubt about the double-helical nature of B-DNA could now be laid aside. Dickerson recalled Francis Crick's visit to his laboratory at Caltech to see their B-DNA structure one day in 1980: They huddled around contoured Plexiglas sheets stacked atop a light box in a dark room, in that pre-computer graphics age. The host pointed out both the similarities to fiber structure for B-DNA, and the differences produced by local base sequence,

which only a single-crystal analysis could show. When the presentation was concluded, Crick smiled and said, "So *that's* what it looks like!"(7) Crick's remark was quite perceptive. Fiber diffraction yields the averaged helix structure, and does not show the effects of base sequence except when that sequence affects the overall averaged structure. Single-crystal structures can reveal the local base-to-base variations in helix structure that are consequence of the particular DNA sequence in question.

The first single-crystal structures of A-DNA fragments were obtained in the early 1980s,(14,15) nearly 30 years after Rosalind Franklin collected X-ray images of fibrous A-DNA. In 1952, Franklin discovered that DNA could assume two forms under the examination of X-ray diffraction. By increasing the humidity, DNA possessing A-form could be further hydrated to form thinner fibers called B-form. Both forms are known as right-handed antiparallel double helices. With single-crystal structures available nowadays, it's easy to see their conformational differences: A-form DNA displays a shallow and wide minor groove, and a deep and narrow major groove compared with B-DNA. This feature reduces the surface hydrogen-bonding area between DNA and the solvent molecules (e.g. water),

especially around the sugar-phosphate backbone.(16) Thus A-DNA has fewer crystalline water and appears dehydrated.

A few years after Linus Pauling's conjecture of triple-stranded DNA, it was discovered that nucleic acids actually could form three-stranded structures.(17,18) Early experiments demonstrated triplex formation from DNA or RNA homopolymers. For example, poly r(A) and poly r(U), besides forming a Watson-Crick type antiparallel duplex, could also form a complex with a stoichiometry of 2:1 [poly r(U)]:[poly r(A)], named as U:AU. The adenine base forms Watson-Crick base pairing with one of the uracil bases. In the mean time, this adenine base forms Hoogsteen base pairing with the other uracil.(19) The Hoogsteen strands may run in parallel or antiparallel direction with each other, leading to parallel or antiparallel triplexes. A synthetic nucleic acid analogue species called peptide nucleic acid (PNA) could result in the formation of triplex as well.(20) When targeted to duplex DNA, they form more stable complex than their DNA counterparts. In addition, strand invasion of PNA generates another kind of triplex with stoichiometry of 2:1 PNA:DNA.(21)



DNA parallel triplex

DNA antiparallel triplex

Figure 1-1. DNA three-stranded helices

The structure formed by guanine-rich DNA segments has been recognized for nearly forty years.(22) The level of interest in G-quadruplex has increased as its functional roles *in vivo* has accumulated. These Guanine-rich segments play significant roles in telomere and centromere sites, as well as in triplet-repeat genetic diseases.(23) Structures determined from NMR and X-ray crystallography display an amazing polymorphism.(24) Quite a number of structural motifs have been observed among G-quadruplexes.

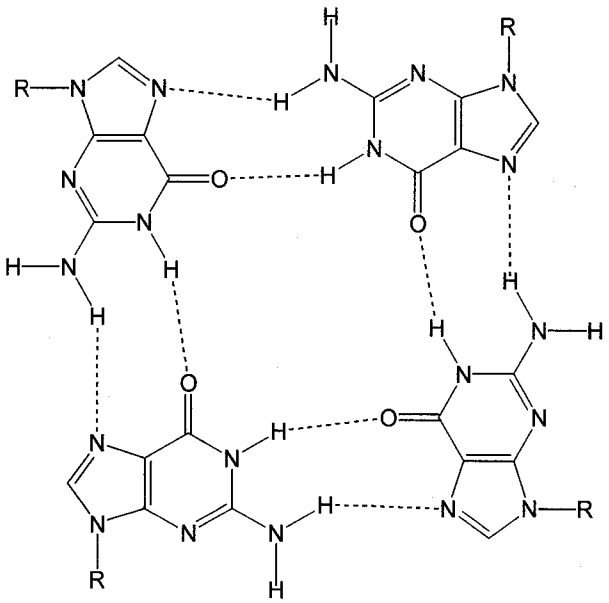


Figure 1-2. Hydrogen bonding pattern of G-quadruplex.

Like deoxyribonucleic acid, RNA oligomers with complementary sequence, chimeric duplexes with mixed RNA/DNA residues, and RNA/DNA hybrid may also form double helices with Watson-Crick base pairs.(25) These structures are mostly found to be A-form. The structures of many important biological RNAs have been determined. The first among them was yeast phenylalanine transfer RNA ($tRNA^{Phe}$) solved by three research groups.(26-28) The discovery of RNA catalytic activity in 1980s revolutionized people's understanding about enzymes, and created more interest about RNA structure. Two variants of the hammerhead ribozyme have their crystal structure determined in early studies.(29,30) Another breakthrough was the crystal structure of the self-splicing P4-P6 domain

(154 nucleotides), of group I intron from Tetrahymena thermophila. (31) Recently, high-resolution X-ray crystal structures of large ribosomal complexes composed of multiple peptide/RNA subunits have been reported. (32,33) Even better, the structure of the entire 70S ribosome was later determined at 5.5 angstrom resolution containing bound mRNA and tRNAs. (34)

1.2 The potential of DNA/RNA derivatives containing selenium for X-ray crystallography

1.2.1 The importance of determining nucleic acid structures

The detailed three-dimensional structures of biomacromolecules, such as DNA, RNA and proteins, can be obtained by X-ray crystallography and NMR. In addition, techniques like UV/VIS spectroscopy, circular dichroism (CD), fluorescence resonance energy transfer (FRET)(35) can provide partial structural information about the molecules under study. Theoretical modeling studies using molecular dynamics (MD) emerged as another useful tool recently for structural biologists.

More than 8,500 sets of structure coordinates were placed in the Protein Data Bank (PDB) since its establishment in 1971.(36) Most of these structures were solved by means of X-ray crystallography, while nearly 15% of the models were derived from NMR spectroscopy. With the surge of structural genomics, many more structures are to be solved in many years to come. For chemists and biologists, these structural studies of proteins, nucleic acids, protein/nucleic acid complexes, and their complexes with small ligands are crucial to the understanding of the molecular interactions

and biological functions. They are also proved critical to the development of medical therapies for human diseases.

Watson and Crick raised their famous model of DNA double helix in 1953. But the study on the DNA structure did not stop there for apparent reasons: It's a model built from X-ray diffraction of DNA fiber. The model can only explain an averaged view (and quite imperfect according to modern crystallography) on double helix, without sequence-specific variation that actually occurs in DNA. Moreover, DNA interacts with various cellular proteins. Structures of these complexes play a key role in furthering our understanding of gene regulation and expression. Structures of DNA-drug complexes reveal the mechanism of DNA-drug binding, and can help with the design of therapeutic drugs.(37)

RNA molecules, on the other hand, perform essential and amazingly versatile functions within cells. Some researchers believe that protein-based life was preceded by a simpler life form based primarily on RNA, which is referred to as the 'RNA world'.(38,39) Although RNA molecules differ little from DNA in chemical terms, they exhibit remarkable conformational diversity, constitute various cellular apparatus (i.e. mRNA, tRNA, rRNA

and snoRNA), and may even carry out enzymatic catalysis by RNAs called ribozymes.(40-42) Recently discovered RNA interference (RNAi) mediates resistance to parasites or pathogens, and regulates the expression of genes.(43) Structures of antisense nucleic acid duplexes provide hints on improving antisense strategy. In a 2001 survey of Nucleic Acid Database (NDB), only less than one hundred structures of protein-RNA complexes were available, compared with over three hundred known protein-DNA complex structures.(44) There is still a great mine of wealth among RNA structures to be discovered.

1.2.2 X-ray crystallography and NMR spectroscopy

Structural models of proteins and nucleic acids in solution can be derived from nuclear magnetic resonance (NMR) spectroscopy. The improvement in high-field magnets, the development of multi-dimensional NMR experiments and methods of analysis, have allowed NMR to become increasingly powerful in structural biology.(45) The spectral data, i.e. intensities of quantitative NOESY (Nuclear Overhauser Effect Spectroscopy), or coupling constants of various kinds of correlated spectroscopy, can eventually establish conformational information of a structure. However, the limitations in current NMR methods have

determined, that most of the protein NMR structures determined so far have less than 150 residues. The primary challenge is that the individual conformational parameters are not precisely defined through NMR restraints, when compared with X-ray crystallography. Second, the biomolecules in water are more or less flexible. This flexibility presents a systematic problem for defining high-resolution structure, since atom distance restraints in NMR are derived from NOE data, which is collected through time-averaged measurement. By nature, it is somehow difficult to have enough NMR data to define a large structure unambiguously.

The most common experimental means of determining three-dimensional structure of a macromolecule is by single-crystal X-ray crystallography. X-ray structures as large as the 70S ribosome has been reported.(34) Many proteins and nucleic acids may form crystals from solution, under slow and controlled precipitation. The crystalline state is an ordered three-dimensional array of unit cells within crystal lattice. A fine crystal can diffract X-ray beam and produce a distinct diffraction pattern on recording film or CCD plate. The dark spots with various intensities in the pattern are generated where the condition for Bragg's law is met. Reflections

on the film show the spacing of the reciprocal lattice, which is inversely proportional to the real lattice within the crystal.

The next transition is to obtain electron density maps from diffraction data. The actual diffractors of the X-ray are the electron clouds of molecules in the crystal. The function of average electron density $\rho(x,y,z)$ in real space can be calculated from structure factors $\boldsymbol{F_{hkl}}$ in reciprocal space by Fourier transform (FT) as follows:

$$\rho(x,y,z) = \frac{1}{V}\sum_{h}\sum_{k}\sum_{l} F_{hkl} e^{-2\pi i(hx+ky+lz)}$$

Information about $\boldsymbol{F_{hkl}}$ can be measured from reflections in X-ray diffraction. A full description of a periodic wave function like $\boldsymbol{F_{hkl}}$ must include amplitude, frequency and phase. Its frequency is that of the X-ray source. The amplitude of $\boldsymbol{F_{hkl}}$ is proportional to the square root of the reflection intensity $\boldsymbol{I_{hkl}}$:

$$I_{hkl} \propto |F_{hkl}|^2$$

However, the phase of $\boldsymbol{F_{hkl}}$ is not directly available from the measurement of reflection intensity $\boldsymbol{I_{hkl}}$. In order to compute $\rho(x,y,z)$ from $\boldsymbol{F_{hkl}}$, we must obtain the phase of each reflection through other ways.

One popular strategy to overcome the phase problem is to prepare one or more heavy-atom derivatives. Crystals of interest are soaked or co-crystallized in solutions of heavy ions or their ionic complexes such as Hg, Pt, or Au. We would expect to have a small number of heavy atoms distributed to identical sites in all unit cells of a crystal. The added heavy atoms have higher atomic numbers (hence more electrons) than these commonly seen elements (i.e., H, C, N, O, P, S) in proteins or nucleic acids, so that they are stronger X-ray diffractors. Subsequently, one or more additional diffraction patterns could be generated from these crystals with added derivatives. An ideal derivative crystal must be isomorphic with the native crystal. At the molecular level, this means the heavy atoms must not disturb its conformation or crystal packing. In practice, it often requires two or more heavy-atom derivatives to produce enough phase estimates for initial elucidation of the phase problem, known as the method of multiple isomorphous replacement (MIR).(46)

Three developments—variable wavelength synchrotron X-rays, cryocrystallography, and the production of proteins containing selenomethionine instead of natural methionine—have recently allowed

rapid exploitation of the multiwavelength anomalous dispersion (MAD) technique in crystallography.(47) For proteins that naturally contain a heavy atom, such as iron (Fe) in a hemoglobin or cytochrome, the native heavy atom provides the source of anomalous dispersion. Other proteins without native heavy atoms can be expressed in cell culture (e.g. *E. coli*) containing exclusively selenomethionine. Selenium (Se), among a number of different heavy atoms used, has the capacity of absorbing X-rays of specified wavelength. Measurement of X-ray absorption near the wavelength of absorption maximum and absorption edge, which are characteristic to selenium, would provide critical phasing information to crystallographers. In this case, diffraction data sets at different wavelengths from only one crystal can be sufficient to solve the protein structure. Remember in MIR phasing, crystals of native sample plus isomorphous crystals of several heavy-atom derivatives are usually required. Thus, MAD phasing method may considerably reduce the work on growing protein crystals, especially for those samples that are difficult to crystallize. Moreover, selenomethionine labeling is built-in covalently to peptide chains, ensuring uniform and stable presence of selenium throughout the cell units of a crystal.

1.2.3 The significance of making selenium-derivatives of oligonucleotides for structural biology

Despite the success of MAD phasing techniques used in protein crystallography, it was not applicable directly to nucleic acids prior to the work reported in the following chapters. For example, DNA-protein complexes have been labeled by selenomethionine on the protein constituent for MAD phasing. But this strategy cannot be applied to other cases where crystals of interest are made of pure nucleic acid constituent.

Our research was focusing on making nucleoside module(s) containing selenium, which may well play the role of selenomethionine in protein crystallography. Upon incorporation of one or more Se-labeled nucleoside units, an oligonucleotide chain will be granted with the capability for MAD phasing as well. This method would liberate DNA and RNA crystals with the new possibility of MAD phasing, an important crystallographic technique from which nucleic acid was previously prohibited. Crystal structures of DNA-protein or RNA-protein complex may also be solved by selenium labeling present on the nucleic acid constituent.

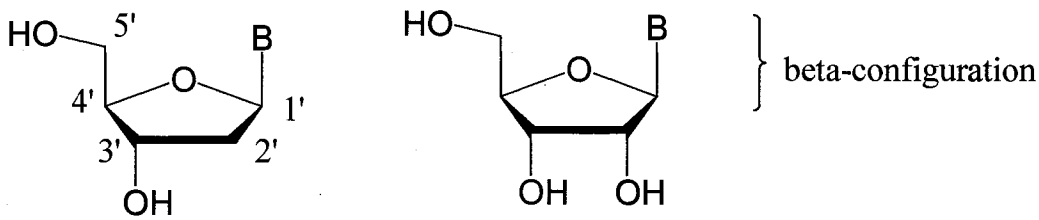
The strategy of labeling nucleic acid also has a practical advantage over the traditional selenomethionine labeling on DNA- or RNA-protein

complexes. Remember that selenoproteins are extracted from cell culture grown in selenomethionine medium. All the methionine residues in the target protein sequence will be replaced by selenomethionine. The number of selenomethionine in a protein is dependent on the peptide sequence. Sometimes, large proteins may contain too many selenomethionine residues, thus too many heavy atoms. This may instead complicate the MAD phasing process. In contrast, our nucleic acid labeling strategy is comprised of chemical incorporation of derivatized nucleotide residue precisely at intended location(s) within a DNA or RNA sequence. Custom design and synthesis allow users control where and how many selenium atoms are added to their interested sequence. A detailed discussion is available in Chapter 4.

1.3 Background in nucleosides, oligonucleotides and their analogues

1.3.1 Structure of nucleosides

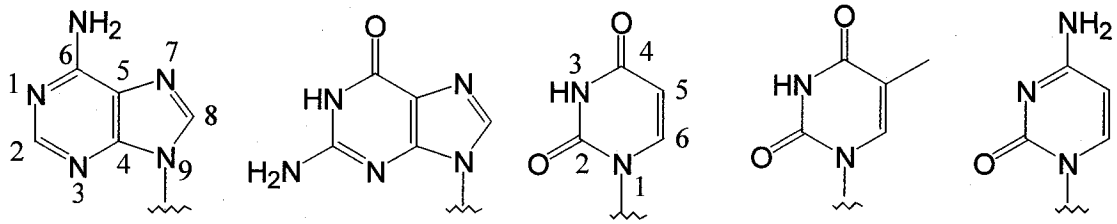
The nucleoside components of DNA and RNA are composed of a sugar moiety, 2'-deoxyribose (2'-deoxy-β-D-ribofuranose) and ribose (β-D-ribofuranose) respectively, linked to a purine or pyrimidine base through a β-*N*-glycosidic bond, through N^9 of the purine and N^1 of the pyrimidine heterocyclic base (Figure 1-3).



1.1, B = A, C, G, T **1.2**, B = A, C, G, U

Figure 1-3. Sugar moiety of DNA and RNA nucleosides.

Other than the common types of nucleosides shown (Figure 1-4), numerous naturally occurring nucleosides have been isolated.(48) Some of these unusual nucleosides play roles in nucleoside metabolism pathway, while the others constitute building blocks in functional nucleic acids. All of these nucleosides invariably have β-configuration at the anomeric center (C-1') of the sugar moiety.

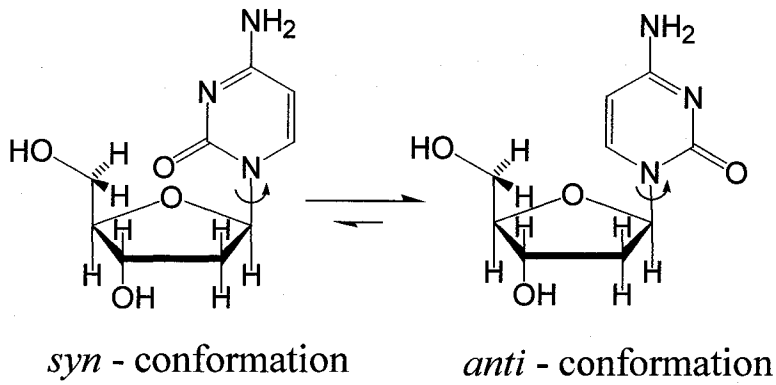


1.3 Adenine **1.4** Guanine **1.5** Cytosine **1.6** Thymidine **1.7** Uracil

Figure 1-4. Common purine and pyrimidine bases in nucleic acid.

1.3.2 Nucleoside conformations

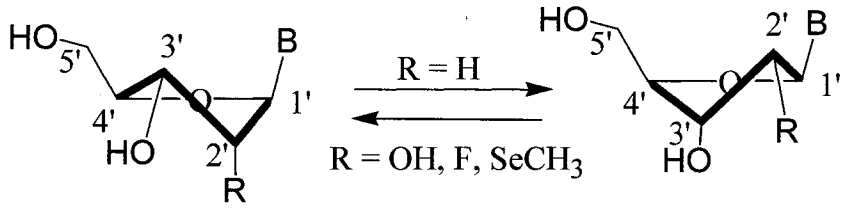
The geometry of the glycosyl link between the base and the ribose ring is determined by the torsion angle χ. The value of χ describes whether the conformation is *anti* or *syn*. For pyrimidine bases, χ is defined by covalent bonds through C2-N1-C1'-O4'. An *anti*-conformation is usually preferred over *syn*-conformation in pyrimidine bases. The latter would give rise to steric interference between the C-2 carbonyl group of the base and the ribose ring (Figure 1-5).(49)



syn - conformation *anti* - conformation

Figure 1-5. Glycosidic bond conformations.

The puckering of the ribose ring gives rise to two main structural conformations, envelope (E) and twist (T).(50) The puckering is determined by the pseudorotational phase angle *P*, which describes which atoms are displaced out of the plane and the direction of displacement. Both steric effects and the gauche effect influence the conformational preference of the sugar ring.(51) The sugar moieties are generally classified as either Northern (N) or Southern (S).(52) Northern conformation of ribose ring puckering is C3'-*endo*, whereas Southern type puckering is C3'-*exo*. Northern conformation is usually adopted in RNA nucleosides, RNA duplexes and A-form DNA double helices. Southern conformation is more preferred in 2'-deoxyribose nucleosides and B-form DNA double helices (Figure 1-6).



C3'-Endo / C2'-Exo,
Northern (N), A-Form

C3'-Exo / C2'-Endo,
Southern (S), B-Form

Figure 1-6. Pseudorotation of the sugar ring pucker between two conformations: Northern and Southern. Substitution of 2'-α-hydrogen with functional groups, such as –OH, -OCH_3, -F, -$SeCH_3$, favors C3'-*endo* (Northern) conformation.

1.3.3 Nucleoside analogs

The monomeric nucleoside analogs have modifications on the sugar moiety, or heterocyclic base, or both. Some well-known categories are classified according to their structural features, e.g. L-nucleosides, *C*-nucleosides, carbocyclic nucleosides, acyclic nucleosides, abasic nucleosides (Figure 1-7).

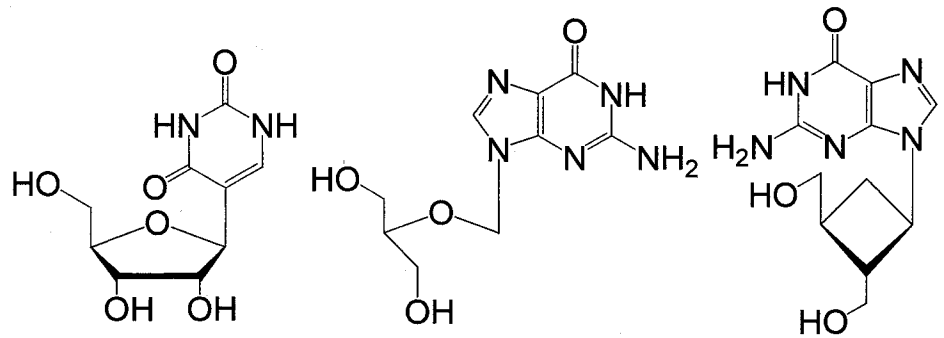


Figure 1-7. Nucleoside analogs.

1.3.4 Biological significance of nucleoside analogs

Nucleoside analogs are of enormous importance in life sciences and medicine. They are an established class of clinically useful agents possessing antiviral and anticancer activity. Nucleoside analogs interact with either specific viral enzymes or with cellular enzymes involved in the biosynthesis of DNA/RNA precursor nucleotides. The therapeutic effects are observed from the enzymes that are inhibited.

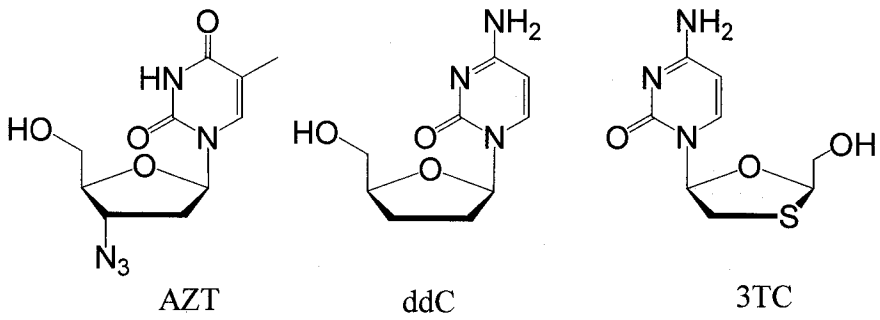
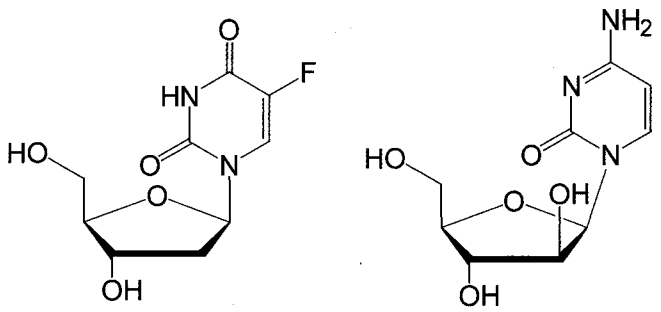


Figure 1-8. Nucleoside derivatives approved for antiviral therapeutics. 3TC possess L-ribose configuration, compared with other nucleoside drugs in D-series, i.e. AZT and ddC.

Great interest has been focused on the development of drugs for AIDS therapies.(53-55) A number of Nucleoside analogs such as 3'-azido-3'-deoxythymidine (AZT) and 2',3'-dideoxycytidine (ddC) are currently used for the treatment of AIDS. These drugs lack C-3' hydroxyl groups, so that incorporation of an analog residue by HIV reverse transcriptase (RT) would cause viral DNA chain termination. Another example is 2',3'-dideoxy-3'-thiacytidine (3TC), a nucleoside derived from L-ribose. Its unnatural L-configuration (enantiomer of the usual D-ribose) proved to be a superior inhibitor of the HIV RT, because of the reduced affinity between L-nucleoside and human cellular enzymes (Figure 1-8).(56)

The biosynthesis of nucleotides is a limiting process in cell proliferation. Some of the enzymes involved in these pathways are highly active in cancer cells. Thus, nucleoside analogs that inhibit nucleotide biosynthesis are of great interest for cancer chemotherapy. 5-fluoro-2'-

deoxyuridine (FUDR), for instance, is an inhibitor of thymidylate synthase and has benn used clinically for the treatment of leukemia and colorectal cancer for over three decades. Arabinosylcytidine (Ara-C), containing unusual β-D-arabinofuranose sugar moiety, was another longtime drug used for treatment of leukemia (Figure 1-9).(57)



FUDR Ara-C

Figure 1-9. Nucleoside analogs that are used for cancer chemotherapy.

1.3.5 Oligonucleotide analogs and applications

Modified nucleosides play a significant role in the development of gene therapies such as antisense and antigene technology. In these strategies, designed synthetic polynucleotides hybridize with target cellular nucleic acid species, and subsequently induce transcriptional arrest, or RNA cleavage.(58) Zamecnik and Stephenson raised the idea of oligonucleotide directed inhibition of gene expression more than two decades ago.(59)

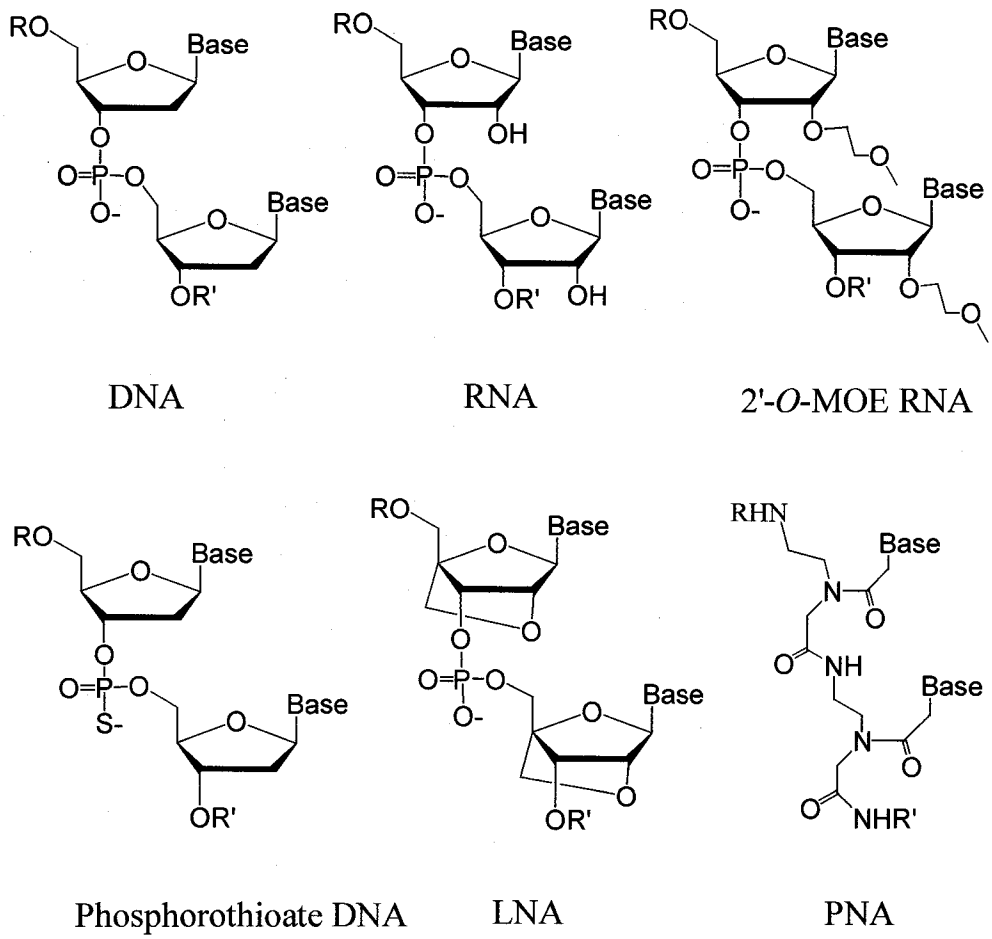


Figure 1-10. DNA, RNA, and modified antisense oligonucleotides.
2'-*O*-MOE RNA = 2'-*O*-(2-Methoxyethyl)-RNA;(60)
LNA = Locked Nucleic Acid;(61-64)
PNA = Peptide Nucleic Acid.(20,65-68)

Today, a phosphorothioate based oligonucleotide called *Vitravene* has been approved by the FDA for treatment of cytomegalovirus (CMV) induced retinitis. About a dozen new antisense oligonucleotides are in various stages of clinical trials. The oligonucleotides intended for antisense therapeutics usually contain chemical modifications in order to improve cell permeability, nuclease resistance and hybridization stability. They may

display a variety of changes within each subunit, including heterocyclic base modifications, pentofuranosyl modifications and phosphodiester linkage modifications (Figure 1-10).(69,70) Common types of antisense oligonucleotides consist of first generation phosphorothioate DNA, second generation 2'-*O*-modified RNA, or "gapmers" made of 2'-O-modified RNA nucleotides at either end of the chain, and a core of unmodified or phosphorothioate DNA, responsible for RNase H activation.

Chapter 2. Chemical Synthesis of Selenium-Derivatized Nucleosides and Oligonucleotides

2.1 Strategies to synthesize nucleic acid analogs

The basic components of nucleic acid: phosphate group, sugar moiety and bases, offer a variety of sites for chemical modification.(71) A huge number of nucleoside, nucleotide and oligonucleotide analogs have been synthesized till today. These syntheses were achieved by either organic chemistry methods, or a combination of organic chemistry and enzyme-catalyzed reactions. (72-74) With the development of biotechnology and green chemistry, an increasing number of biocatalytic reactions have been studied and reported.(75) However, the organic chemistry approach is irreplaceable, and is much dominant by far, in making different nucleic acid analogs for decades.

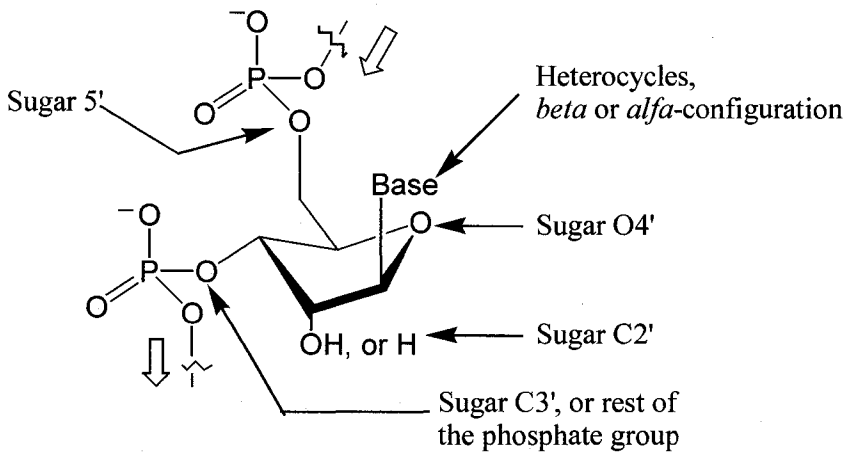


Figure 2-1. Typical sites for chemical modification within the nucleic acid framework.

Most common sites for chemical modification on nucleic acid are: heterocyclic bases, C3' and C5' positions on DNA deoxyribose, C2', C3' and C5' positions on RNA ribose (Figure 2-1).

The synthetic monomeric nucleoside analogs are classified using a number of categories according to their structural features, e.g. L-nucleosides, *C*-nucleosides, carbocyclic nucleosides, acyclic nucleosides, abasic nucleosides.

Synthetic oligonucleotide analogs can be made by site-specific incorporation of modified nucleosides into DNA or RNA chains. Moreover, the 5'→3' phosphodiester linkage between neighboring residues can also be modified in more unconventional manners (Figure 1-10, Figure 2-2).(76) The difference in structural appearance may cause variation in strand-to-strand binding affinity during nucleic acid hybridization, membrane permeability, nuclease resistance and RNase H activation. Those biochemical characters are critical to the success of antisense applications.

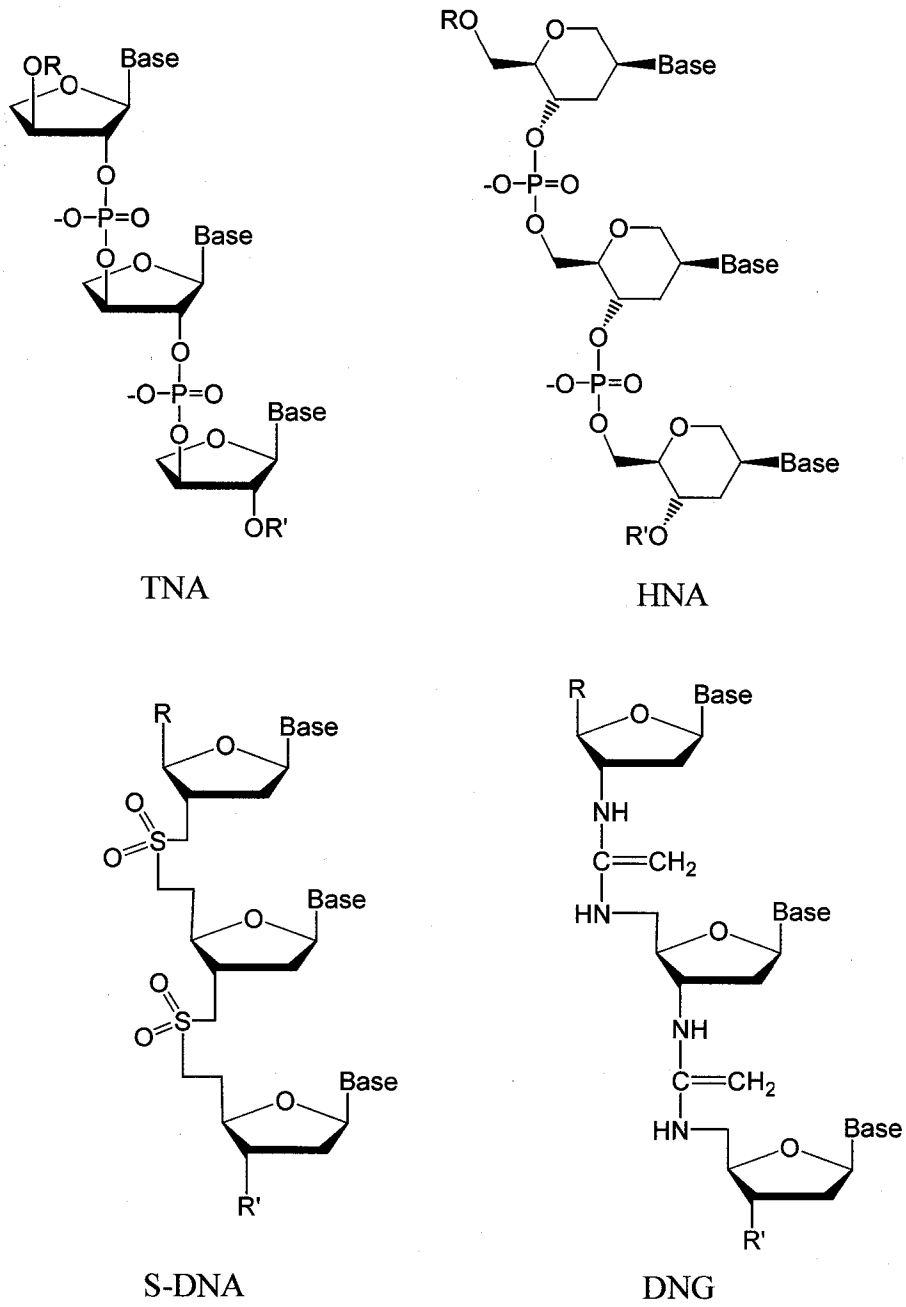


Figure 2-2. Backbone modifications of oligomeric nucleic acids.
TNA: α-L-Threofuranosyl-(3'→2')-NA.(77,78)
HNA: Hexitol-NA.(79,80)
S-DNA: DNA oligonucleotide with electrically neutral dimethylene sulfone linkage.(81,82)
DNG: Deoxyribonucleic guanidine. Contains positively charged backbone.

On the other hand, systematic studies on alternative oligonucleotides provided insight into chemical etiology of nucleic acid: Why does nature choose this particular molecule form as the genetic basis of all life?(83) It was suggested by a number of researchers that different kinds of genetic molecules could have existed during historic Darwinian evolution, such as PNA, and 2'-*O*-methylated RNA.(84,85)

The backbone modification of nucleic acid could be ribose alteration (i.e. HNA), position shift of phosphodiester linkage (i.e. TNA), or the replacement of the entire phosphate group (i.e. S-DNA and DNG). The new backbones could be electrically neutral or even positive, compared with negatively charged natural nucleic acid. Some of the chemically modified oligonucleotide species may well form stable Watson-Crick type hybridization with unmodified RNA or DNA, while some other species cannot.

2.2 Locations of selenium substitution

In our studies, selenium was chosen to derivatize amino acid and proteins in X-ray crystallography for two reasons: selenium can serve as *anomalous dispersion* center during X-ray diffraction; and it's in the same group in periodic table as oxygen. Being in the same periodic group explains that selenium and oxygen have similar chemical characters to some extent. When oxygen is substituted by selenium, the charge remains the same; the original structure of a molecule may not be seriously disturbed. The principle of selenium substitution has proven successful for MAD phasing in protein crystallography.(86,87) Therefore, when this project started, substitution of oxygen by selenium seemed to be a prospective path for modifying nucleic acids as well. Our experiments actually have now proved this principle—selenium derivatization can work for nucleic acid crystallography!

The 2' and 5'- hydroxyl groups on deoxyribose of DNA are good targets for selenium replacement. In addition, the α-hydroxyl group at C2' position of RNA could be substituted as well (Figure 2-3).

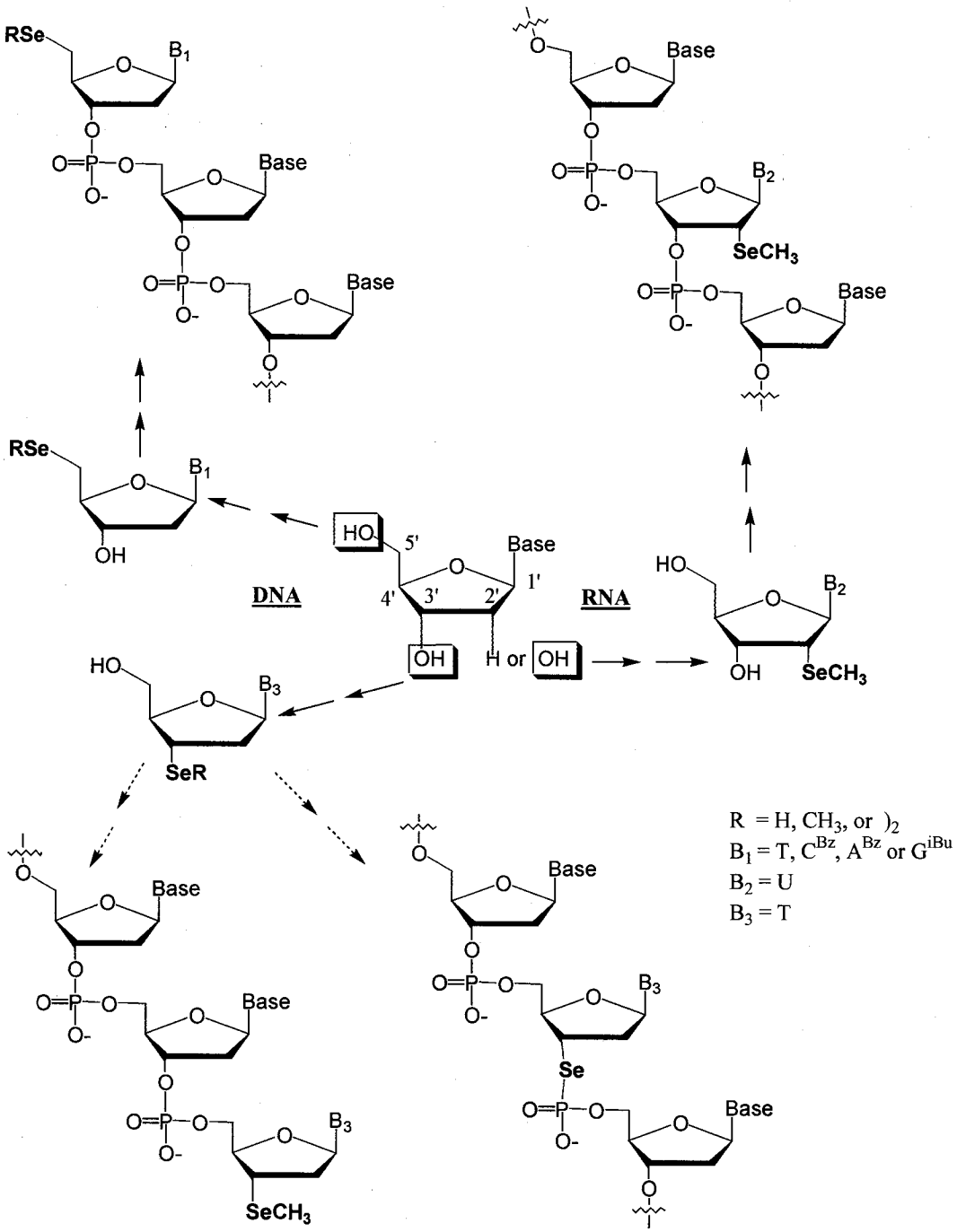


Figure 2-3. Replacement of carbohydrate hydroxyl groups by selenium functional groups at separate sites: C5', C3' of DNA, and C2' of RNA nucleosides. These modified nucleosides may be further incorporated into oligonucleotides.

The first series of selenium-modified nucleosides made during the thesis project have selenium attached to the C5' position on deoxyribose of A, C, G, T (Chapter 3).(88) The –OH group at the C5' position was substituted by $-Se)_2$, or $-SeCH_3$ groups. The reactions are typical SN2 nucleophilic substitutions. The capping of the C5' with a relatively stable group, like $-SeCH_3$, prohibits further nucleotide chain elongation towards the 5' end. As the solid phase oligonucleotide synthesis starts from the 3' terminal and extends toward 5' terminal in one residue-per-cycle fashion, this type of C5' capped nucleoside residue can only be placed at the very end of 5' strand terminus. So the method confines the possible number of selenium tags is one per oligonucleotide strand.

Subsequent selenium labeling was achieved at the C2' position of uridine (U).(89,90) Chemical modification at the C2' position of RNA does not directly interfere with the 5'→3' phosphodiester backbone linkage (where both –OH groups at C5' and C3' participate) in oligonucleotide, so that a C2'-modified uridine could be incorporated to virtually any internal site(s) of a designed DNA or RNA short sequence through solid phase synthesis. Several uridine residues with selenium tag can be inserted within

a newly synthesized strand (Chapter 4). Discussion of the crystallographic structure is available in Chapter 6.

Finally, experiments on selenium labeling at C3' of thymidine (T) were conducted (unpublished results, Chapter 5). One thymidine analog was made with $-SeCH_3$ at the C3'. This compound can potentially label the 3' end of a pre-made oligonucleotide sequence. The attempt to make thymidine analog with selenophosphoramidite group for standard solid phase synthesis is not yet successful.

Chapter 3. Introducing Selenium Functionality to 5'-position of Nucleosides and Oligonucleotides

3.1 Introduction

Nucleic acid strands are polymers of ribonucleotide (RNA) or deoxyribonucleotide (DNA) residues connected by phosphodiester linkage. Oligonucleotide refers to a short polymer of nucleotides with less than a few dozen residues. Oligonucleotides are widely utilized in molecular biology and biochemistry, as PCR primers, diagnostic and therapeutic sequences. Oligonucleotides, or at least their 5' termini, must be chemically synthesized *in vitro*, since all DNA polymerases lack the priming capability.

One appealing feature of synthetic oligonucleotides is the possibility of creating artificial molecules tailored to specific needs, from sequence to structure. Chemical modifications of nucleic acids are introduced mostly on the nucleoside monomer level at first. These chemically modified nucleoside monomers can be incorporated within strands of synthetic oligonucleotides at intended locations.

In order to test the feasibility of selenium-labeled nucleic acid for X-ray crystallography, the 5'-terminal hydroxyl group was chosen as the target site for selenium labeling. The 5'-OH of ribose is the only primary hydroxyl in a nucleoside structure. Generally speaking, it is more reactive as well as more regioselective compared with 2'- and 3'-OH groups, which are both secondary. Substitution of oxygen by selenium at 5'-OH should be easier than the other two –OH positions. Oligonucleotides with selenium labeling at 5'-termini can be synthesized by standard phosphoramidite chemistry.

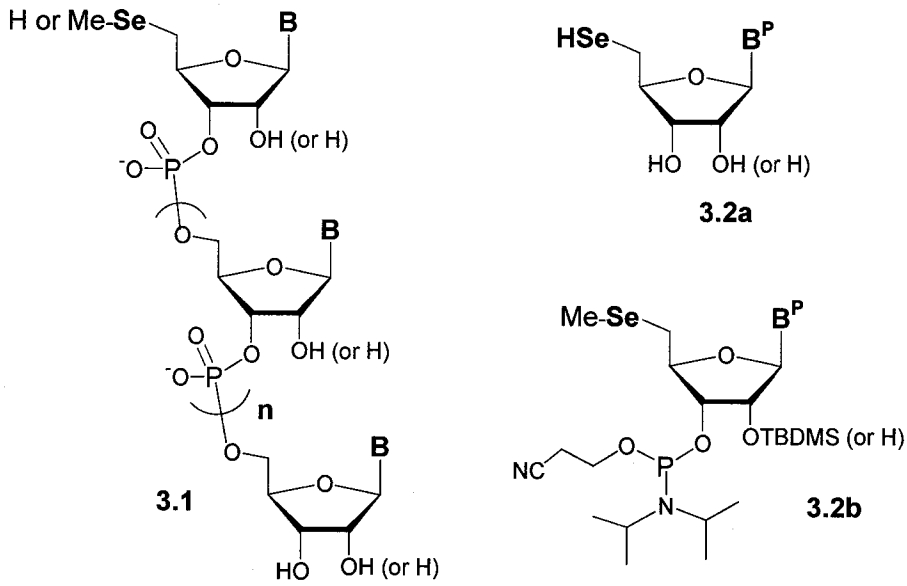


Figure 3-1. Nucleosides and oligonucleotides with selenium labeling at the 5'.

The synthetic method of 5' selenium labeling was developed mainly by colleagues Carrasco and Ginsberg.(88) My experiments contributed to part

of the synthesis and data collection in order to complete this part of the work.

3.2 Experiments and results

Incorporation of the selenium functionality is commonly done by nucleophilic substitution chemistry in ethanol or DMF solvent using sodium selenide made by $NaBH_4$ reduction of selenium metal. However, this conventional approach proved unsatisfactory in the case of acyl protected nucleosides, given the fact that strong base in the nucleophilic substitution induces removal of the acyl protecting groups on nucleobases. To overcome this problem, a two-phase system (H_2O-toluene) was developed by colleagues to incorporate selenium using a phase transfer catalyst. The half time of the nucleophilic reaction was less than 10 min. when Ms-(mesyl) and Br- groups were used as the leaving groups and sodium selenide was used as the nucleophile. As this nucleophilic substitution was fast in the organic phase, the selenide anions transferred into the organic phase did not cause removal of the acyl groups from the nucleobases. For C, A and G nucleosides with acyl protection, the two-phase reactions were conducted at pH 8, which avoids the base deprotection. These reactions assisted by the phase transfer catalyst are fast, easy to workup, and give high yields.

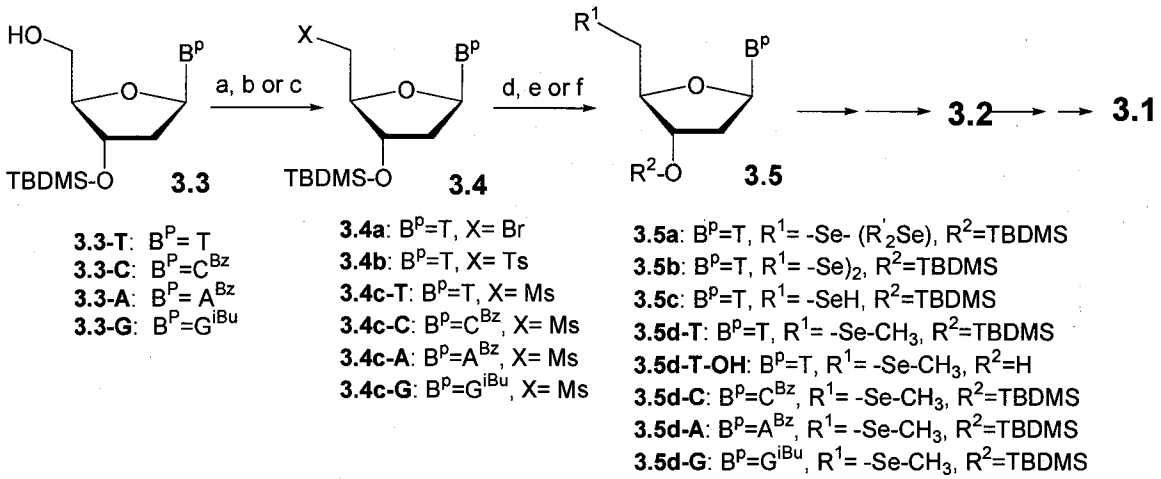


Figure 3-2. (a) CBr_4/Ph_3P/DIAD/THF. (b) Ts-Cl/Py. (c) Ms-Cl/TEA/THF. (d) Na_2Se, (e) Na_2Se_2, or (f) $NaSeCH_3$ plus H_2O/Toluene and Tetrahexylammonium Hydrogen Sulfate. TBDMS: *tert*-butyldimethylsilyl.

My work in this chapter involved the synthesis of **3.3T→ 3.4c-T →** **3.5d-T**, and **3.3A→ 3.4c-A → 3.5d-A**. The 5'-hydroxyl groups of partially protected nucleosides **3.3** (T, A^{Bz}) were activated for nucleophilic substitution with Ms (Mesyl) leaving group. Compounds **3.4c** (T, A^{Bz}) were synthesized by standard procedures.

A two-phase solvent system (H_2O-toluene) and a phase-transfer catalyst (tetrahexyl ammonium hydrogen sulfate) were used to carry out the selenium transfer reaction. The phase-transfer catalyst was used to shuttle the methylselenide anions (CH_3Se^-) from the aqueous phase to the organic phase where the reaction took place. Sodium methyl selenide ($Na^+CH_3Se^-$) was freshly prepared by reducing dimethyl diselenide ($CH_3SeSeCH_3$) with

$NaBH_4$ in anaerobic conditions. The presence of excess CH_3Se^- anion in organic phase replaced 5'-mesyl group of nucleoside (**3.4c-T** or **3.4c-A**) by nucleophilic substitution. After respective synthesis, **3.5d-T** or **3.5d-A** was formed. Thymine base is relatively stable during these reactions without particular precautions. However, reactions related with adenosine must be carried out under neutral or weakly basic condition in order to prevent the hydrolysis of the Benzoyl group under strong base.

Two DNA oligonucleotides were synthesized with 5'-methylseleno-thymidine (**3.2b**) by colleagues: 5'-^{Se}TT-3' and 5'- ^{Se}TGCGCA-3' (**3.1**, n = 0, 4 respectively). Standard phosphoramidite chemistry was used to make the oligos. Their identities and the presence of selenium were proved by mass spectroscopy. Crystallization of self-complimentary sequence 5'-^{Se}TGCGCA-3' has not yet been successful.

3.3 Discussion

Selenium derivatization at the 5'-position of DNA nucleosides and oligonucleotides was achieved. Introduction of methylselenol group at the 5'-position required only a few hours of reaction and gave high synthetic yield (>90%). Reduced selenol group in R-SeH is extremely unstable in air

(unlike –OH group) and will be oxidized very quickly to the dimerized product: R-SeSe-R'. This type of 5' head-to-head conjugate is undesired because the nucleotide chain orientation is reversed within itself. As a result, the feasible way to incorporate selenium at 5' would be capping Se by a stable alkyl group such as $-CH_3$.

So far, no structural result has been obtained for any 5' Se-labeled oligonucleotide. Little is known about how the 5'-$SeCH_3$ group would interact with the rest of the structure in solution or in crystal. The 5'-$SeCH_3$ group is bonded to a primary carbon and could dangle around the C5'-Se bond. The preferred bond angles and the influence on the structural stability may be revealed after further study.

An inherent limitation of this type of 5' derivatization is that only one selenium atom per strand of polynucleotide may be introduced (as shown in compound **3.1**). The selenium tagged monomer must be incorporated during the very last cycle of solid phase synthesis, which extends oligonucleotide in 3'→5' direction. The 5' Se-labeling may serve as a good method for X-ray crystallography of short oligonucleotide (estimated up to 30 nucleotides). However, the shortage of selenium scattering centers could be a problem for

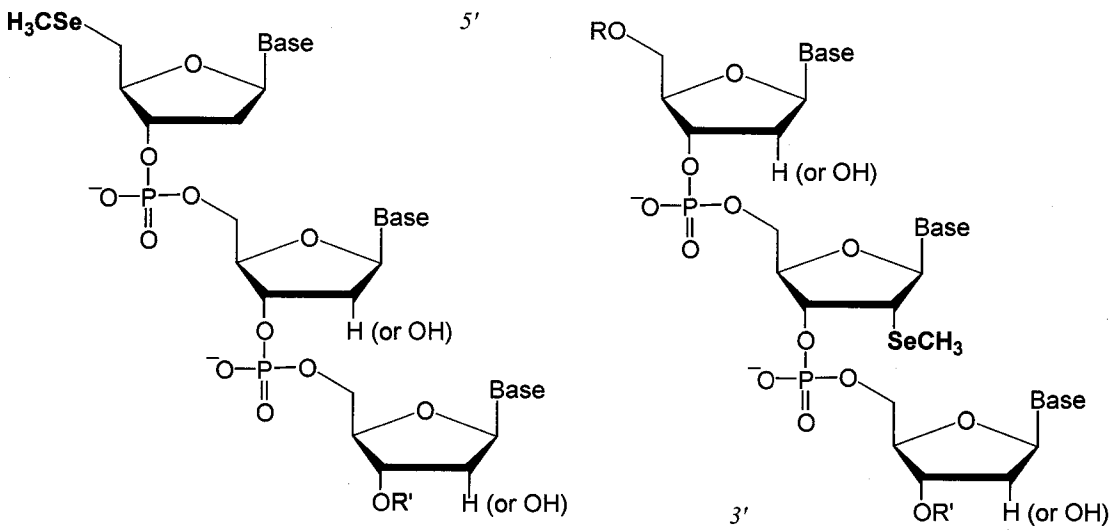
crystallographers when the target molecule is large. In this case, more selenium tags may be necessary in order to keep the selenium number in satisfactory ratio with increasing molecular weight.

For long DNA sequences or possibly DNA-protein complexes, the 5' Se-labeling technique alone is not going to be sufficient because of the size issue. A new kind of Se-labeling method, which allows adding multiple Se to internal sequence, must emerge in order to overcome the limitation.

Chapter 4. Internal derivatization oligonucleotides with selenium for X-ray crystallography using MAD

4.1 Introduction

Successful Se-labeling at the 5'-position of nucleosides and oligonucleotides was demonstrated in Chapter 3. The work provided a method to stably connect the selenium functional group onto the 5'-terminal of a generic oligonucleotide (**4.1**), replacing the typical –OH group. The synthesis of selenium derivatized oligonucleotides was intended for elucidating of their X-ray crystal structures by Multiwavelength Anomalous Dispersion (MAD).



4.1 **4.2**

Figure 4-1. Oligonucleotide structures with 5'-terminal (**4.1**) or internal 2'-α-position (**4.2**) Se-labeling.

Two problems still remain for Se-labeling at the 5'-termini: No crystal structure was concluded from this methodology until now; and no more than one selenium atom is allowed per strand of oligonucleotide. The latter problem will severely limit the maximum length of DNA or RNA for crystallographic study using MAD. A nucleic acid – protein complex will be unquestionably too large for the strategy. This inherent difficulty could only be overcome by alternative labeling strategies that allow multiple Se-labels.

One possible strategy is to incorporate Se at the 2'-position instead of the 5' (**4.2**). Nucleic acids are polymers that use 5' and 3' ribose (or deoxyribose) positions for phosphodiester linkage. Modification at the 2'-position does not directly interfere with chain elongation when compared with 5'-labeling. Theoretically, this will allow Se-labeling at almost any nucleotide position within an oligonucleotide sequence; the number of Se-labels could increase along with the extension of the oligonucleotide. Thus, the 2'-labeling seemed very attractive for our purpose.

A variety of modifications have been made to the ribose 2'-α-hydroxyl group of nucleosides and oligonucleotides by organic synthesis. Typical examples include 2'-*O*-alkyls, 2'-*O*-alkyls with glycol ether

linkages, 2'-F and 2'-*O*-aminoalkyls. In fact, 2'-*O*-methylated RNA is found in mature ribosomal RNA (rRNA) of all organisms. It clearly indicates that 2'-*O*-methylation of RNA is an ancient biological process. The process is carried out by small nucleolar ribonucleoproteins (snoRNPs) in eukaryotes, and by protein enzymes in bacteria. For chemists, the studies were mainly driven by the high demand in antisense technology since the 1980s, where specially designed oligonucleotide analogues were examined as potential therapeutic agents.

A chemical modification of the 2'-OH group usually maintains an RNA like C3'-*endo* conformation in oligonucleotide duplex, also known as the A-form structure. These substitution groups reside in the minor groove of the RNA-antisense oligonucleotide duplexes formed. 2'-modification may provide nuclease resistance for antisense oligonucleotides. Some 2'-modifications result in substantially increased melting temperatures, which indicate stronger duplex hybridization.(71) For example, the 2'-*O*-methoxyethyl (2'-MOE) group leads to an increased Tm around +1.4 ^{o}C per modification.(91)

4.2 Experiments and results

4.2.1 The chemistry of solid-phase oligonucleotide synthesis

The first chemical synthesis of a dinucleotide with 3'-5' phosphodiester linkage was performed by phosphotriester approach in 1955. In the following year, Khorana developed the phosphodiester approach for internucleotide bond formation. The field had revolutionary changes during the 1980s, when Beaucage and Caruthers reported phosphoramidite as a new intermediate for DNA oligonucleotide synthesis. The discovery and subsequent improvement by Beaucage and others founded the principle of solid-phase DNA and RNA oligonucleotide synthesis used today. It finds application in almost all biomedical sciences, such as synthetic gene constructs, DNA sequencing, PCR, diagnosis of genetic diseases, and therapeutic antisense drugs.

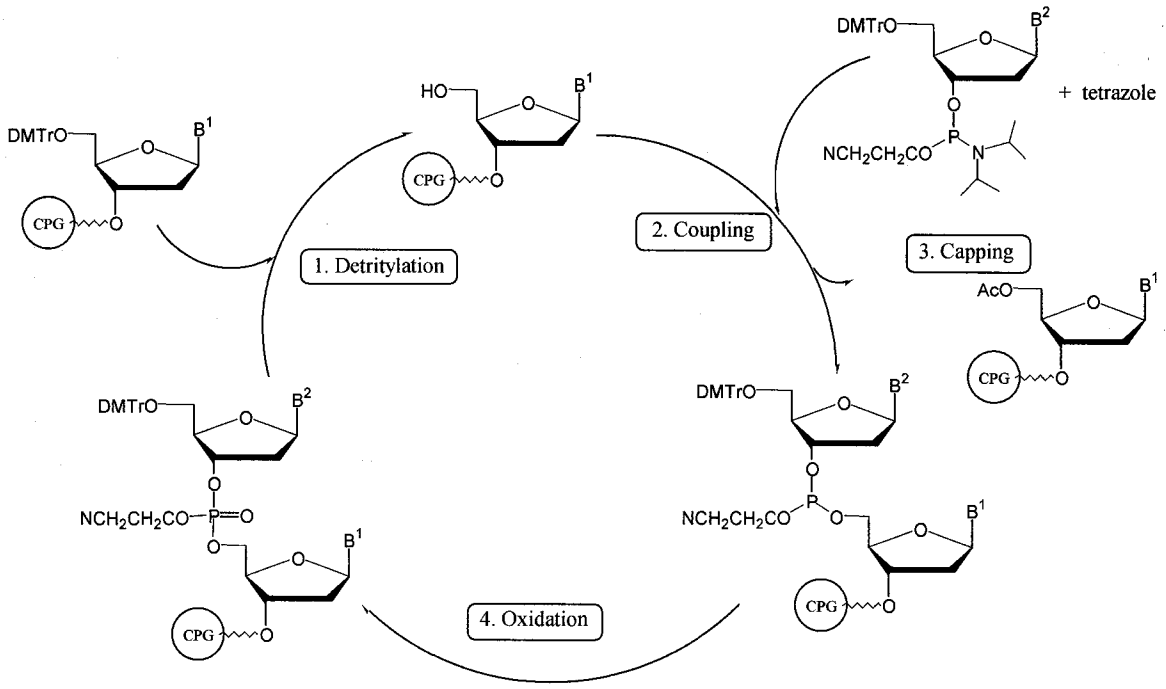


Figure 4-2. Solid-phase synthesis of DNA oligonucleotide by phosphoramidite method.

In standard solid-phase synthesis of DNA oligonucleotide, the 3'-OH group of the first nucleoside is anchored to a CPG (controlled pore glass) solid support through an organic linker. The oligonucleotide synthesis extends from 3' to 5' direction. Each reaction cycle consists of 4 major steps. In step one, the beads are treated with weak acid in order to remove the dimethoxytrityl group (DMTr) from the 5'-end. In step two, Nucleobase and 5'-protected nucleoside phosphoramidite (A, C, G, or T) is activated by tetrazole and added to the immobilized oligonucleotide. The chain grows by nucleophilic attack of the exposed 5'-OH on the activated phosphoramidite functional group. After the coupling step, exposed (unreacted) 5'-OH, albeit

only a tiny percentage, is capped by acetyl group for blocking elongation of wrong sequence. In step four, the phosphite linker is oxidized by iodine to the more stable phosphate.

The oligonucleotide is elongated by one nucleotide after each synthesis cycle. A sequence with "N" nucleotides will require a total of "N-1" synthesis cycles. Today, specialized machines can automatically carry out these operations on preset programs.

4.2.2 Design and synthesis of a uridine analogue with methylselenide at the 2'-position

Among the four common RNA nucleoside monomers (A, C, G, U), uridine was chosen as the initial target for selenium incorporation. The main reason was the workup "simplicity" of the uracil base, when compared with adenine, cytosine, and guanine. The latter three nucleosides all require proper nucleobase protection of the amino groups during routine chemical synthesis.

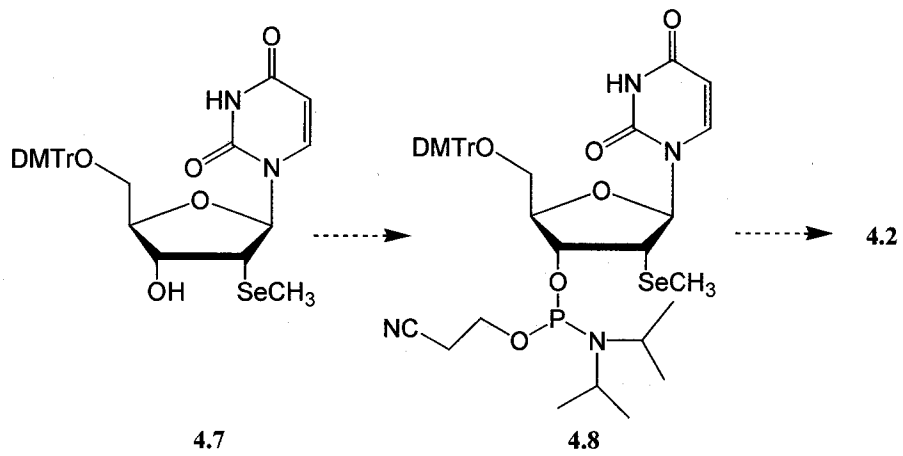


Figure 4-3. Precursor compounds for the synthesis of oligonucleotide **4.2**. DMTr = dimethoxytrityl. *t*-BDMS = *t*-butyldimethylsilyl.

Nucleoside analogue **4.8** is the ideal building block to obtain oligonucleotide **4.2** by solid-phase synthesis. The structure of **4.8** contains a 5'-DMTr and a 3'-phosphoramidite as necessary modules for solid-phase synthesis, a 2'-α-$SeCH_3$ group for stable Se-labeling. **4.8** can be synthesized from **4.7** in one step by treatment of chloro-phosphoramidite.

4.2.3 Early trials and a key intermediate

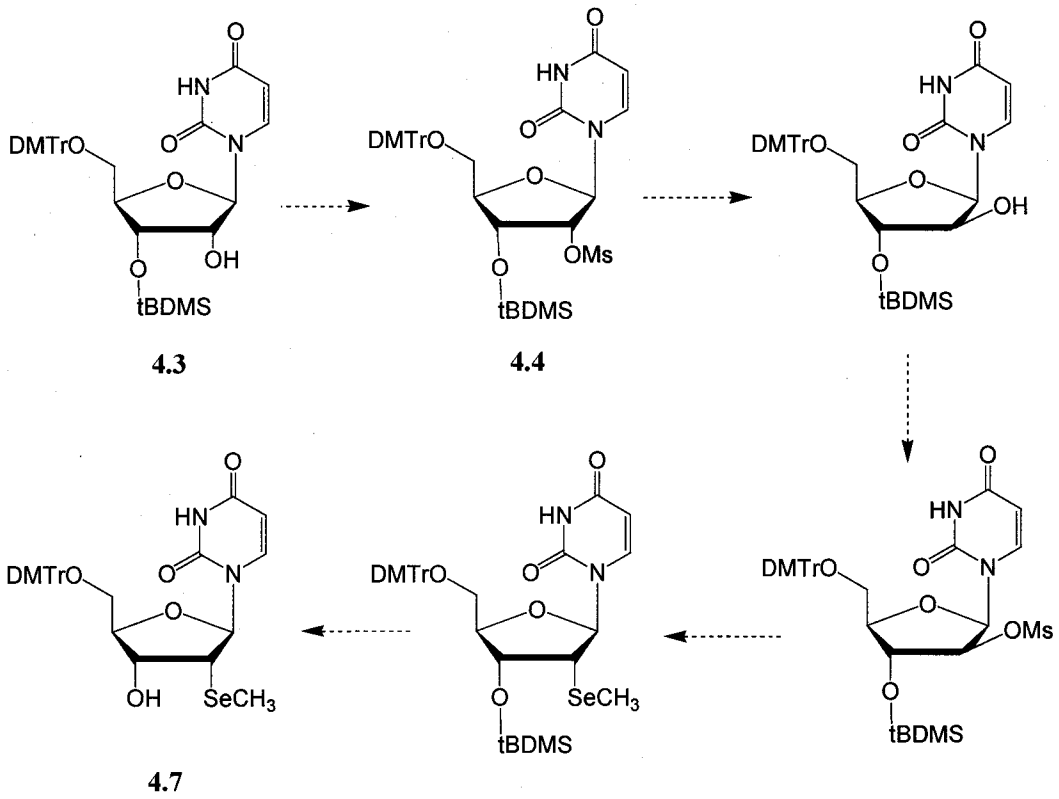


Figure 4-4. Early proposal to synthesize **4.7** from starting material **4.3**. However, the arabinose nucleoside was not eventually made.

The initial plan to synthesize precursor **4.6** began from a partially protected uridine compound **4.3**. After the mesylation of 2'-α-OH group and subsequent base treatment, the conversion of ribose to arabinose was wanted. The 2'-β-OH group on arabinose could then be mesylated again. The product could be susceptible to nucleophilic attack from CH_3Se^- ions from the opposite side of the sugar ring, and finally be converted to 2'-α-$SeCH_3$ group (Figure 4.4).

However, no arabinose product was obtained as anticipated earlier. In fact, the 2'-*O*-mesylated uridine **4.4** was converted to a stable 2,2'-anhydrous compound **4.5** after the base treatment. The presence of compound **4.5** was identified by a collection of experimental data: decrease in molecular weight by 18 (one less H_2O, compared with **4.3**) from mass spectroscopy; and changes in ^{1}H-NMR spectrum, notably the disappearance of the N-3 imino proton signal, and the shift of H-5 and H-6 signals. The weak base used, like Na_2CO_3, does not normally react with the relatively stable uracil. The NMR data clearly indicated that an intramolecular reaction took place, which resulted in significant rearrangement of uracil structure.

4.2.4 The structure and the use of compound 4.5

Compound **4.4** undergoes intramolecular substitution in basic environment (Figure 4.5). The removal of N-3 imino proton was confirmed by loss of its signal in ^{1}H-NMR spectrum. The six-member heterocyclic uracil and the five-member ribose ring are held together by an oxygen bridge between C-2 and C-2'. Much less free rotation is allowed for the glycosidic bond between C-1' and N-1 in **4.5**.

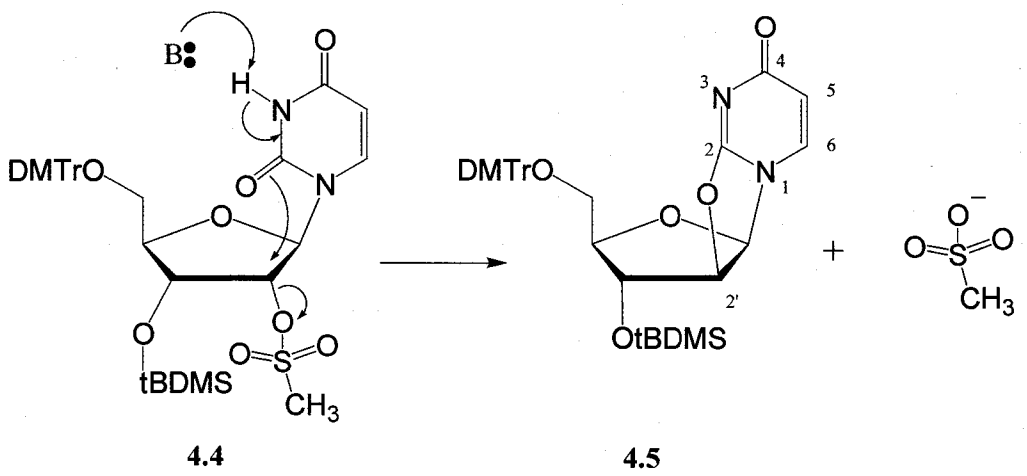


Figure 4-5. Probable mechanism for compound **4.5** formation.

Although the arabinose intermediate was not obtained as planned, the discovery of compound **4.5** can provide a great platform for nucleophilic substitution at the C-2' position. The rigidity of the triple ring provides stereochemical control so that a nucleophile will most likely attack the C-2' from underneath the ribose ring (α-position).

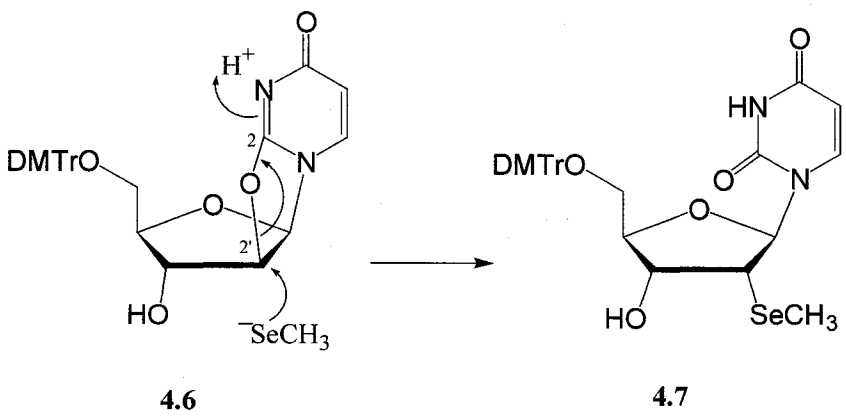


Figure 4-6. Ring opening by nucleophilic attack of methylselenol from 2'-α-position.

Later experiments confirmed the advantage of this 2,2'-anhydrouridine structure. A good nucleophile such as CH_3Se^- attacked C-2' under mild condition in THF solvent, and give stereochemically pure 2'-α-$SeCH_3$ product at high yield. (Figure 4.6) The single bond between C-2' and the oxygen is broken during the ring opening. The unique tricyclic structure in **4.6** is then transformed back to original uracil and ribose in **4.7**.

4.2.5 Synthesis of selenium-labeled uridine phosphoramidite

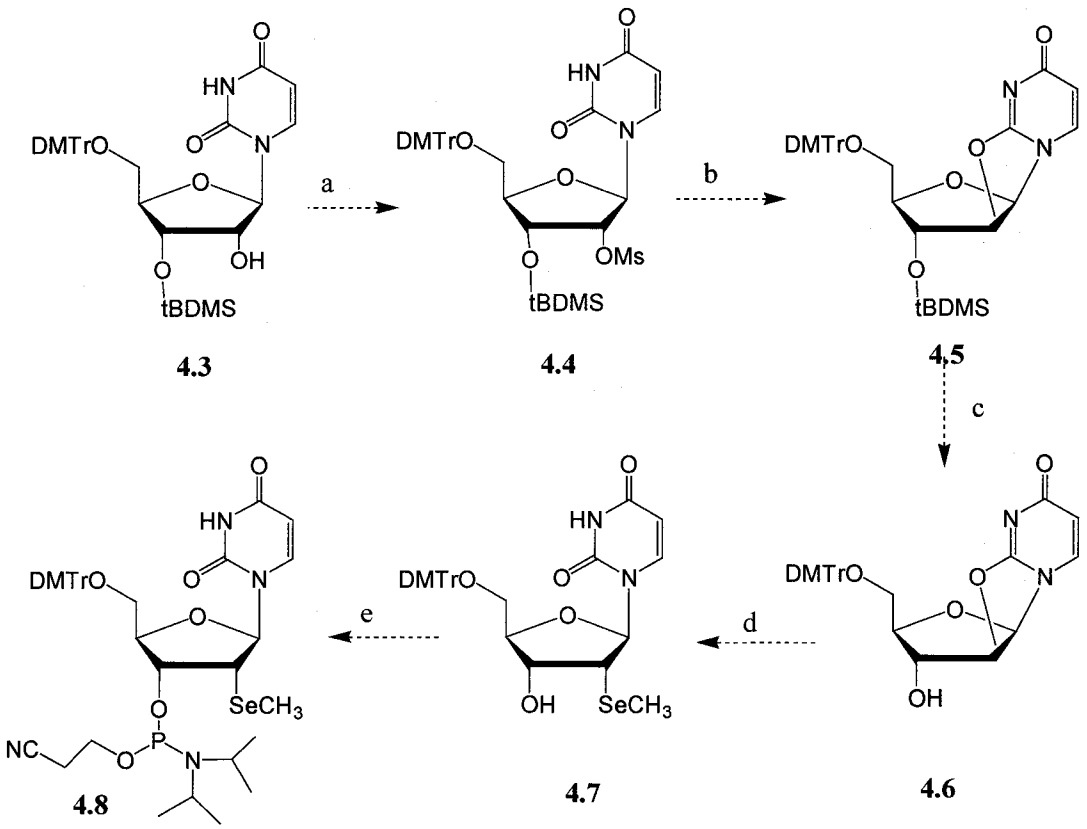


Figure 4-7. Synthesis of uridine phosphoramidite containing selenium at the 2'-position. (a) MsCl/THF/TEA, 95%; (b) Tol/Na_2CO_3 aq/tetrahexylammonium hydrogen sulfate, 96%; (c) $(Bu)_4N^+F^-$/THF, 95%; (d) $NaSeCH_3$/THF, 96%; (e) $PCl(OEtCN)N(iPr)_2$ /acetonitrile, 92%.

The synthesis began with a partially protected uridine **4.3**. After mesylation, the compound **4.4** was mixed in a two-phase solvent (Toluene and Na_2CO_3 aq) with a phase-transfer catalyst. The mesyl groups were fully displaced by the uracil exo-2-oxygen under basic condition in 96% yield. Experiments indicated that the bulky 3'-tBDMS group blocked selenide nucleophiles from attacking the 2'-position of anhydro-uridine **4.5**. So this protecting group was removed by fluoride treatment to give **4.6**. Subsequent addition of reduced sodium methylselenide opened the tricyclic ring on the nucleoside, and **4.7** was obtained in 96% yield. The methyl group on selenium prevents the otherwise easy oxidation of free selenol (-SeH). The selenide nucleophilic reaction was conducted in THF solution, which avoided the ring opening at the nucleobase C-2 position. Highly purified **4.7** was finally converted to selenium-labeled phosphoramidite **4.8** in 92% yield by reaction with 2-cyanoethyl *N*,*N*-diisopropyl chlorophosphoramidite.

For quality assurance, selenium-labeled uridine phosphoramidite **4.8** was synthesized freshly before use. Like all commercial phosphoramidites, this compound is not stable. The compound needs to be scrupulously

purified and dried to avoid decomposition. The prepared product was stored in the freezer sealed under dry argon.

4.2.6 Chemical evidence for selenium incorporation and its conformation

The presence of selenium in compound **4.7** was analyzed and confirmed by MS, NMR (^{77}Se NMR, 1D and 2D NMR) and other experiments. (1) The low and high-resolution mass spectroscopy data exactly matches the value calculated from the molecular formula. The mass spectrum also exhibits characteristic distribution of mass peaks from major Se isotopes, e.g. ^{78}Se, ^{80}Se and ^{82}Se. (2) ^{77}Se NMR spectrum has a singlet peak at –378.0 ppm, which proves that the selenium exists and resides at a single structural location. (3) ^{1}H-NMR and ^{13}C-NMR indicate strong electron shielding effect on a number of nuclei, particularly H-2', C-2', and H, C in -$SeCH_3$. The chemical shifts of H-2' and C-2' shifted significantly to higher field when compared with those values from natural uridine. This result can be interpreted by the bonding of selenium instead of oxygen to the C-2'. Because selenium has lower electronegativity than oxygen, the lone pair valence electrons on selenium tend to shield the other nuclei in vicinity.

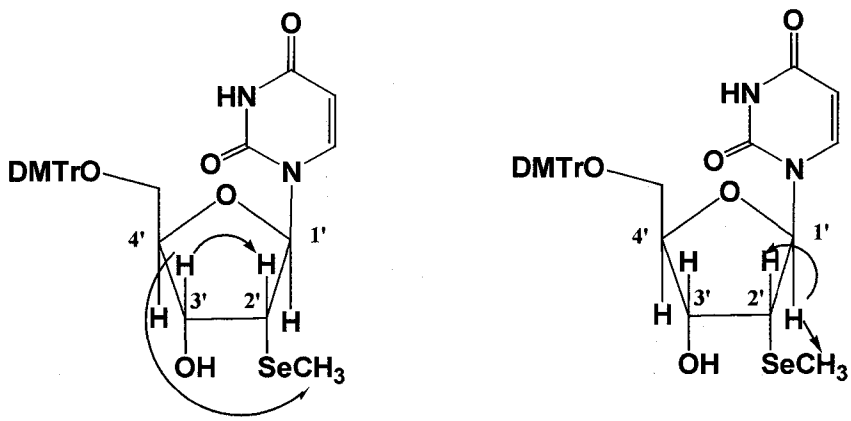


Figure 4-8. Spatial distances result in different NOE enhancement in compound **4.7**. According to the configuration shown above, **(a)** H-3' is closer to H-2' than to methyl-H. **(b)** H-1' is closer to methyl-H than to H-2'. NOE spectra are not shown here.

NOE (Nuclear Overhauser Effect) NMR can look into the spatial coupling between nuclei in close proximity. It's based on the fact that NOE decreases very quickly when the coupling distance increases.(92) The NOE difference spectra provided a simple yet clear picture about the 2'-α-conformation of methylselenide in compound **4.7**. In one NOE experiment, irradiation (saturation) of the H-3' proton signal led to relatively small enhancement in $-SeCH_3$ signal intensity. Meanwhile, there is a large enhancement in H-2' signal. The data suggest H-3' is closer to H-2' than to methyl-H. So H-2' must be on the β-position, just like H-3', to satisfy the notion. (Figure 4-8 a)

In another NOE experiment, irradiation of H-1' signal led to large enhancement in $-SeCH_3$ signal. Very small enhancement in H-2' signal was observed at the same time. Like the first experiment pointed out, these phenomena can only be explained by placing H-2' on the β-position and methylselenide on the α-position. (Figure 4-8 b)

4.2.7 Synthesis of selenium-labeled oligonucleotides

Purified selenium-containing uridine phosphoramidite **4.8** was dissolved in anhydrous acetonitrile to 0.1M concentration. The vial was loaded onto automated DNA synthesizer and kept under argon. For DNA oligonucleotide synthesis, four other standard phosphoramidite nucleosides (A, C, G, T) are also loaded on designated ports of the machine. These nucleobases are pre-capped by standard protecting groups: N-benzoyl group for A and C, and N-isobutyryl group for G. In cases of RNA synthesis, the 2'-OH of A, C, G and U monomers must be protected additionally by groups like TBDMS (*t*-butyldimethylsilyl) or TOM (triisopropylsilyloxymethyl).

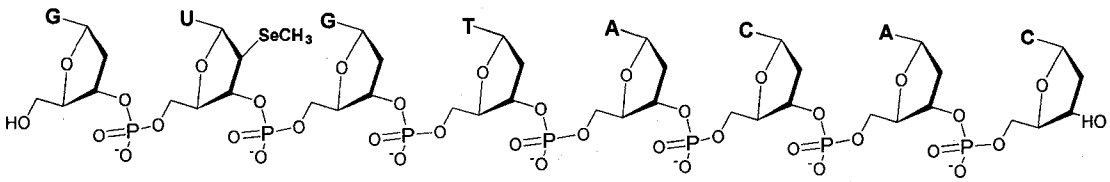


Figure 4-9. Structural formula of a DNA octamer in the study: 5'-GU$_{Se}$GTACAC-3'.

In my experiments, two selenium-derivatized DNA oligonucleotides were synthesized: an octamer 5'-GU$_{Se}$GTACAC-3' and a decamer 5'-GCGTAU$_{Se}$ACGC-3'. U$_{Se}$ is the uridine residue containing methylselenide at the 2'-position.

The machine program for the synthesis of these oligonucleotides largely followed standard protocol. Trityl-On mode was selected. The incubation time for each coupling is 25 seconds typically. The time for U$_{Se}$ incorporation step was elongated up to 10 minutes to ensure maximum coupling yield. After the completion of each solid-phase synthesis, the cartridge containing CPG-linked oligonucleotides was removed from the machine. The beads were taken out and treated with 28% aqueous ammonia for 12 hours at 50 °C. Ammonia treatment removes all nucleobase protecting groups and cleaves the oligonucleotides from chemical linkers on CPG. Afterward, ammonia was evaporated using a SpeedVac. The DNA was finally dissolved in water.

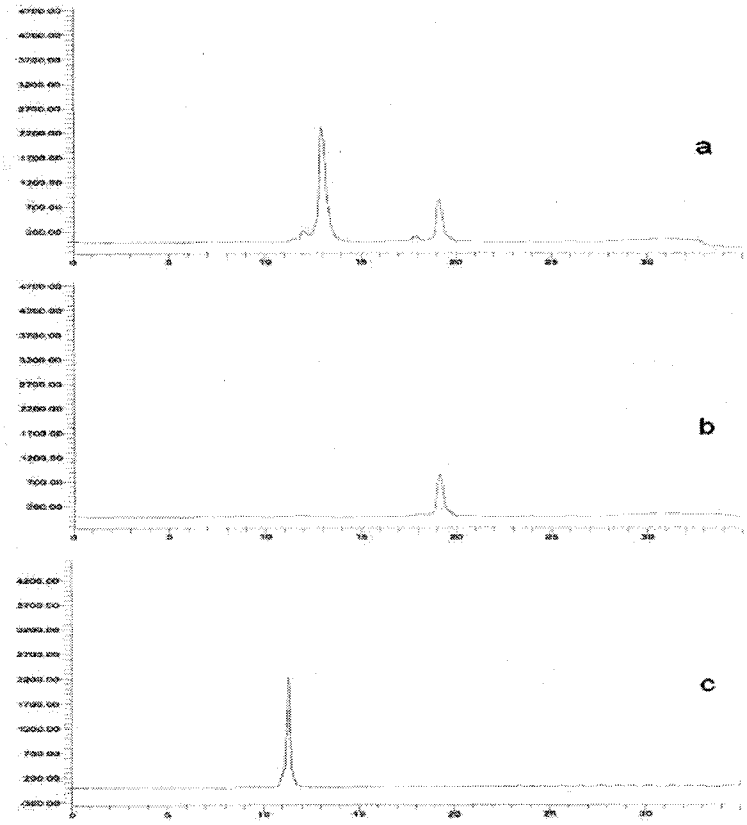


Figure 4-10. Reverse-phase HPLC purification profiles of DNA oligonucleotides. (a) First run with trityl-on. The peak eluted at 19 min contained the full-length DNA. (b) Re-run of the purified full-length component. (c) Detritylated DNA product.

The DNA oligonucleotides were purified by reverse-phase HPLC, using ammonium bicarbonate buffered water and acetonitrile as solvents. During the first run, the full-length oligonucleotide moved slower than other DNA impurities because of the presence of hydrophobic 5'-DMTr group (Figure 4-10 a, b).

After the full-length DNA was separated and concentrated, trityl groups were cleaved by trifluoroacetic acid and later extracted away by hexane. DNA with free 5'-OH is much more polar than with 5'-DMTr group, thus it should migrate faster in reverse-phase HPLC (Figure 4-11). The detritylated DNA was once again purified by reverse-phase HPLC. A single absorption peak at 11 min was observed by UV absorbance (Figure 4-10 c). Its retention time was much shorter than the predecessor's 19 min, as we anticipated.

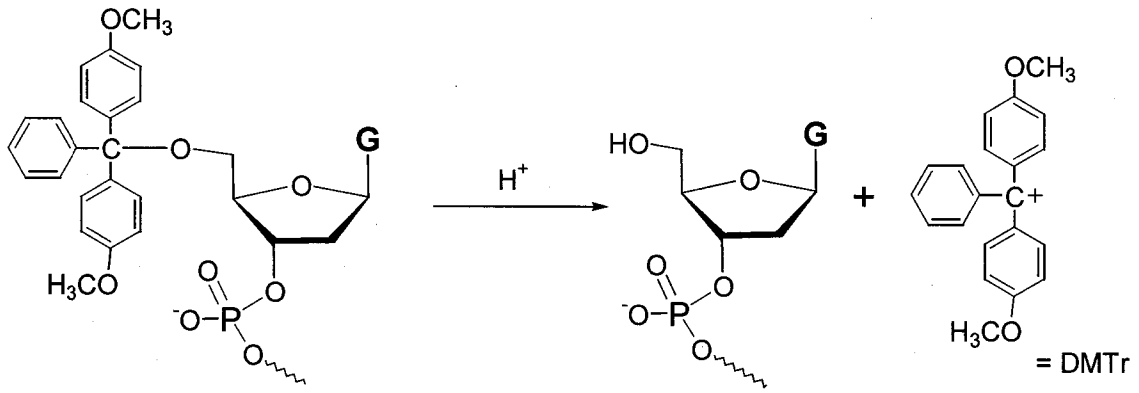


Figure 4-11. The availability of 5'-DMTr group affects the overall polarity of oligonucleotides.

The DNA concentrations were calculated from UV absorption at 260 nm. The estimated overall yield for 10 μmol scale octamer synthesis was

15% in the first trial. The current yield is greatly improved due to better compound purity.

The constituents of selenium-DNA oligonucleotides were confirmed by electrospray mass spectrometry. Molecular mass peaks from MS experiments matched those calculated from formula. The negative ion mode MS spectrum of the octamer, for example, displayed m/z of a set of anions carrying charges from 2 to 5 (Figure 4-12).

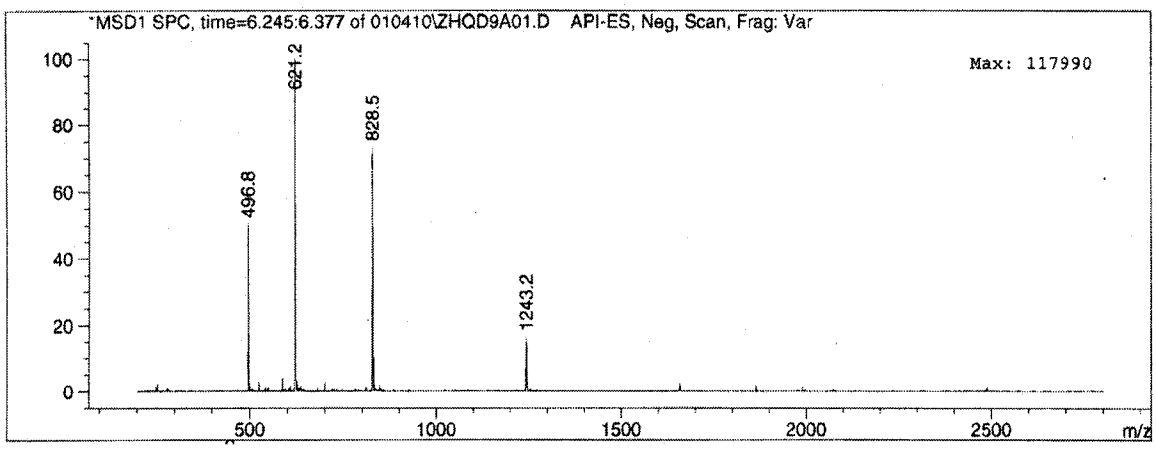


Figure 4-12. LC-ESI-MS of octamer d($GU_{Se}GTACAC$) (formula: $C_{78}H_{99}N_{30}O_{46}P_7Se$), Mt=2489.63. Measured (expected) m/z: $[M\text{-}2H^+]^{2-}$ =1243.2 (1243.8); $[M\text{-}3H^+]^{3-}$= 828.5 (828.8); $[M\text{-}4H^+]^{4-}$ = 621.2 (621.4); $[M\text{-}5H^+]^{5-}$ = 496.8 (496.9).

4.3 Improvements on synthesis of nucleic acid analogues

The initial synthesis of selenouridine phosphoramidite **4.8** involved six compounds and five steps, starting from compound **4.3** (Figure 4-4). This synthesis route was reported in our recent publications. Each of the five steps gave a yield of 95% or better. The high quality of the selenouridine phosphoramidite was verified in successful solid-phase oligonucleotide synthesis. However, the catalog price of the starting material **4.3** is quite expensive. The early reactions were carried out only in small scales. The cost would rise up quickly if we need to incorporate multiple selenium tags to variety of nucleic acid sequences.

4.3.1 Two alternative approaches to make intermediate compound 4.6 from uridine.

One approach was to make tricyclic uridine directly from native uridine. The Conversion of one kilograms scale uridine (**4.9**) to 2, 2'-*O*-anhydrouridine (**4.10**) was reported in literature. Following this lead, **4.6** can be easily made by one tritylation reaction on **4.10** (Figure 4-13). The rest of the steps leading to **4.8** are kept as the same.

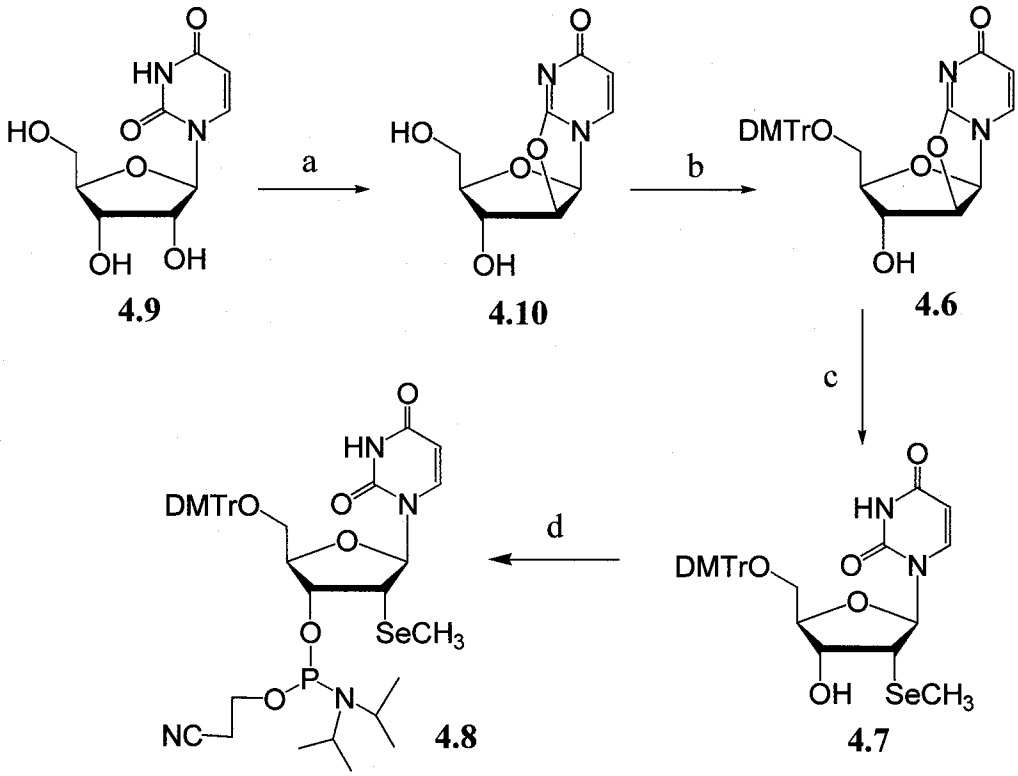


Figure 4-13. Adopted new approach to synthesize phosphoramidite. (a) diphenyl carbonate/DMF/heat, 65%; (b) DMTr-Cl/TEA/Pyridine, 91%; (c) $NaSeCH_3$/THF, 96%; (d) $PCl(OEtCN)N(iPr)_2$ /acetonitrile, 92%.

This new strategy is a nice improvement to our method reported previously. It cuts down the material cost significantly for large scale synthesis. Moreover, the overall procedure is simplified by reducing the reaction steps from five to four. Batches of selenophosphoramidite were made by the new strategy. The high quality was also maintained after tests. So the protocol became the standard procedure of the laboratory.

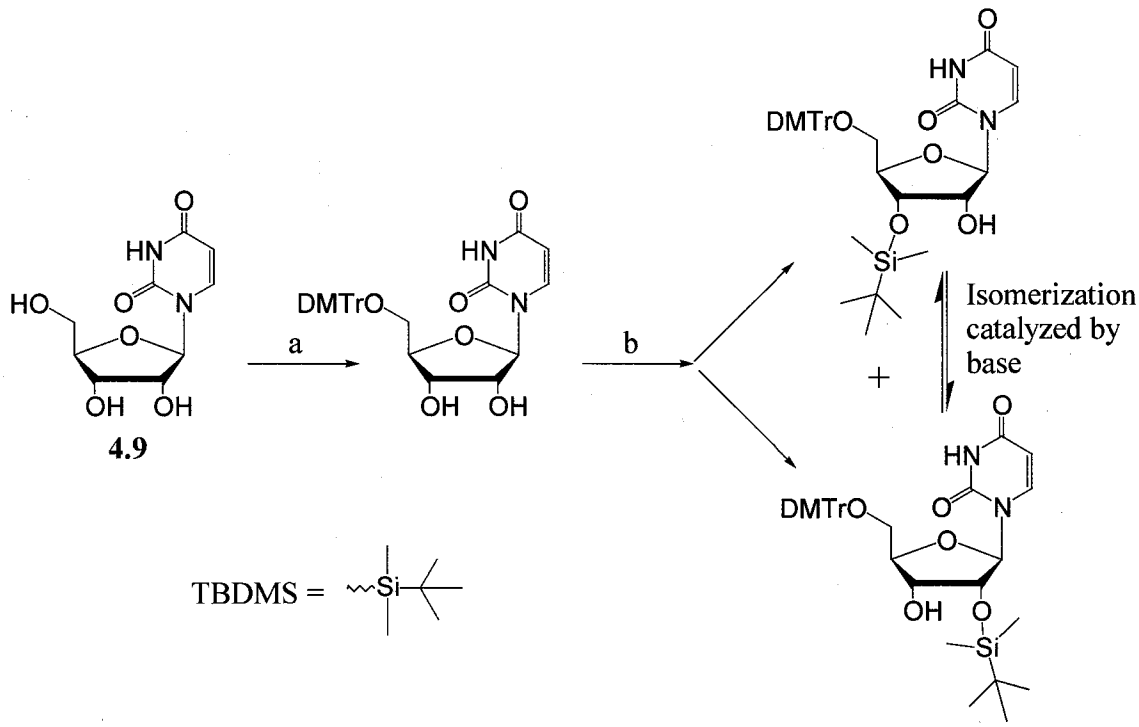


Figure 4-14. Migration of TBDMS group between 2', 3'-diol. (a) DMTr-Cl/ TEA /Pyridine; (b) TBDMS-Cl/imidazole/DMF.

Another approach to start with uridine is: Tritylate the 5'-OH group of uridine; then use ~1 equivalent of TBDMS-Cl to protect one of the 2' and 3'-OH, preferably 3'-OH. In practice, two isomeric products were obtained. The 2'-*O*-silylated uridine, and the desired 3'-*O*-silylated uridine were separated by flash chromatography. The persistent problem here was the migration potential of silyl group between 2' and 3'-OH, which seriously complicated the purification process (Figure 4-14). The phenomenon was thoroughly discussed by Greene and Wuts.(93) This second approach did not survive mainly because of the low yield and the longer process.

4.3.2 Oxidation study on oligonucleotides with selenium labeling

The uridine analogue containing 2'-methylselenide group (U_{Se}) was incorporated into a number of different oligonucleotide sequences through solid phase synthesis, in which selected thymidine or uridine residue was replaced by our special uridine analogue. During the HPLC purification of these oligonucleotides, a closely migrating tiny band (<2% intensity) was discovered. It was later thought to be an oxidized DNA product, based on results in mass spectrometry. This information led to a conjecture that the selenium atoms within the DNA were partially oxidized by I_2, the oxidant used during every cycle of DNA/RNA solid-phase synthesis.

In the case of DNA octamer 5'-d($GU_{Se}GTACAC$)-3' (formula = $C_{78}H_{99}N_{30}O_{46}P_7Se_1$), the highest mass peak in its MS spectrum is at 2489, as a result of isotope distributions of all elements. The MS data (Table 4-1) showed a mass increase of (1251.4–1243.4)×2=16.0, very likely to be a single oxygen, in the minor product. This minor product also moved slightly faster in reverse-phase HPLC, which indicates a higher polarity than the major product. The observation is consistent with our speculation that the polarity of -$SeOCH_3$ is greater than -$SeCH_3$.

DNA products	MS particle identity	Calculated m/z	Observed m/z
Major, less polar	$[DNA - 2\ H^+]^{2-}$	1243.5	1243.4
Minor, more polar	$[DNA + Oxygen - 2\ H^+]^{2-}$	1251.5	1251.4

Table 4-1. Data collected from negative ion mode LC-ESI-MS experiments. Major and minor DNA products were discerned by HPLC prior to MS.

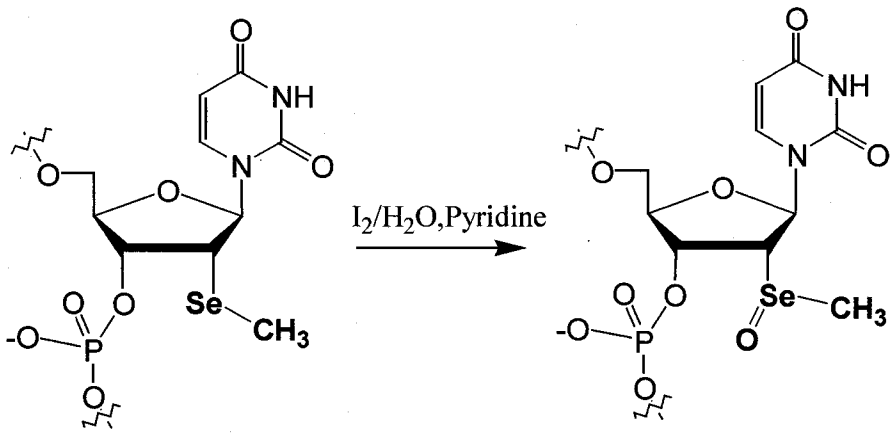


Figure 4-15. A small percentage of methylselenide may be oxidized to methylselenoxide by I_2 treatment during oligonucleotide synthesis.

The solid-phase oligonucleotide synthesis starts from the 3' end and extends one nucleotide per cycle, to the 5' direction. I_2 solution is applied at the end of each cycle for 20 seconds to oxidize phosphite linkage to phosphate. The I_2 treatment is mandatory because phosphite is not stable enough to start the next cycle. During this step, a small percentage of methylseleno groups could be oxidized to methylselenoxide (Figure 4-15). One possible way to minimize the oxidation effect is to put the U_{Se} as close as possible to the 5' end of the sequence, so that selenium will experience fewer additional synthesis cycles, thus shorter exposure to I_2. In general, the

selenium oxidation effect by standard solid-phase synthesis is minimal. In addition, non-oxidized oligonucleotides can be isolated through HPLC.

It was suggested by crystallographers that oxidized selenium might actually be beneficial for solving X-ray structures. So I tested two common oxidants, H_2O_2 and I_2, to study the oxidation of selenium in DNA octamer 5'-d($GU_{Se}GTACAC$)-3'mentioned earlier.

	MS partical Identity	**Calculated m/z**	**Observed m/z**
Product 1	DNA + Oxygen – $2H^+$	1251.5	1251.3
Product 2	G*p*Use*p*+2Oxygen – H^+	763	762.1

Table 4-2. Negative ion mode LC-ESI-MS results from H_2O_2 treated seleno-DNA. Most mass peaks could not be rationally interpreted.

Extended I_2 treatment of DNA solution (from hours to 2 weeks) did increase the oxidation yield up to ~50%. With 100 equivalents of H_2O_2 in DNA aqueous solution for 30minutes, part of the DNA octamer was oxidized to the same methylselenoxide product, according to MS data. The majority of DNA was possibly fragmented by H_2O_2 and was very hard to analyze. One of the fragments could be 5'-d(G*p*$\underline{U}^*_{Se}$*p*)-3', whose calculated mass resembles one of the measured mass peaks (Table 4-2). Here, the $\underline{U}^*_{Se}$ refers to methylselenone with 2 divalent oxygen bonded to each selenium (Figure 4-16). This result may suggest that selenium was initially oxidized to

methylselenone by H_2O_2. Then in the following event, methylselenone group caused chain breakage right between the $U^*_{Se}2$ and the G3.

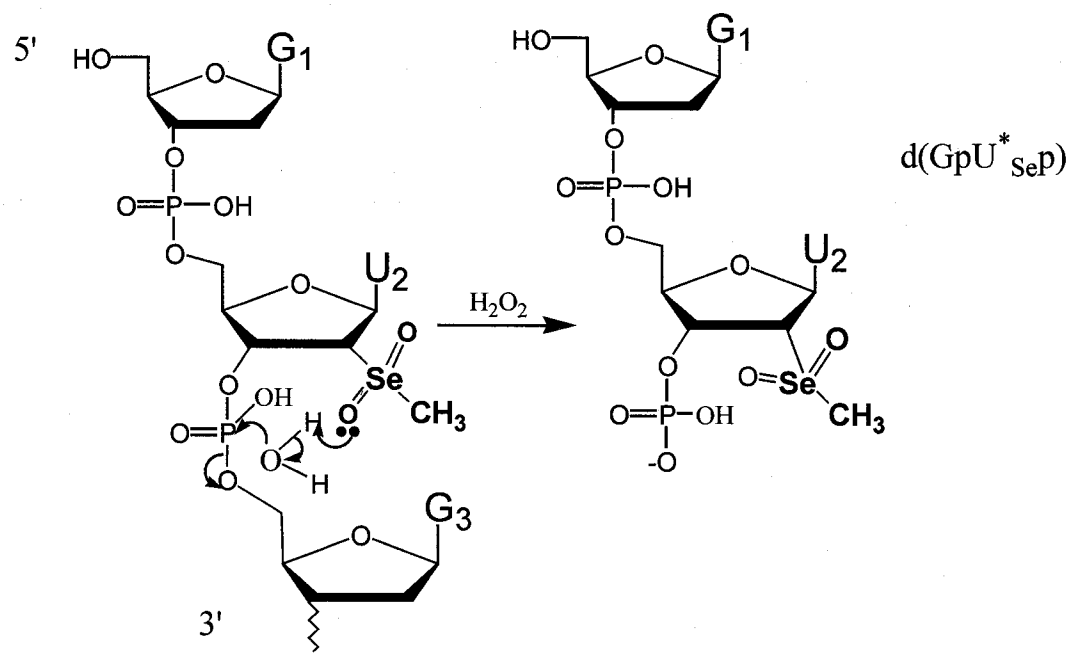
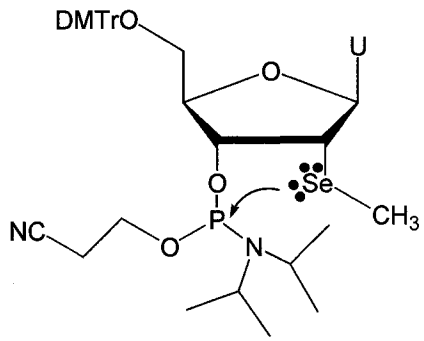


Figure 4-16. Possible mechanism for the DNA chain breakage during H_2O_2 oxidation. Some of the selenium were possibly oxidized to $-SeO_2CH_3$.

Neither I_2 nor H_2O_2 produced ideal oxidized DNA in those experiments. The milder I_2 could not completely convert all selenium to selenoxide, but did retain the rest of DNA structure. H_2O_2 may have largely destroyed full length DNA, and it was possibly too strong for our purpose.

4.3.3 Improvement on oligonucleotide synthesis

Early Se-labeled oligonucleotide synthesis used standard 1H-tetrazole as coupling catalyst. The low yields that were encountered in many trials of synthesis suggest the yield-limiting step could be the incorporation of modified uridine, despite prolonged coupling time. In early trials, the coupling yield for each U_{Se} incorporation was about 30% from trityl analysis, much lower than the >99% yields from native DNA phosphoramidites. The dinucleotide coupling proceeds because of the nucleophilic attack on phosphorus by 5'-OH group from anchored DNA chain. The lone-pair electrons on Se at the 2'-position may shield the electrophilic phosphorus center nearby, thus increase its stability and decrease the reactivity (Figure 4-17).



4.8

Figure 4-17. Shielding around phosphorus center by lone-pair electrons from Se.

It was found that other tetrazole derivatives with electron-donating group at C-5 showed better yields for less reactive phosphoramidites.(94) In

other words, derivatives with smaller pKa values are stronger than classic 1H-tetrazole as coupling catalysts. In our later syntheses, 5-benzyl-thio tetrazole was tried. It gave significantly higher yield for U_{Se} coupling. The yield of RNA oligonucleotide synthesis was also improved by adopting the new catalyst, presumably by similar mechanism.

Through the efforts of the lab colleagues, the current coupling yield for U_{Se} has reached the same level as those for ordinary DNA phosphoramidite. This was largely contributed by the improved purification of U_{Se} phosphoramidite.

4.4 Discussion

Novel uridine analogues containing methylselenide at the 2'-α-position were synthesized. DNA or RNA oligonucleotides can be covalently labeled with U_{Se} by phosphoramidite chemistry. Different DNA oligonucleotides were prepared with one U_{Se} insertion on each strand. They were later crystallized and examined by MAD X-ray crystallography.

This 2'-Se-labeling technique is a significant improvement over the 5'-Se-labeling. The crucial difference is that 2'-labeled U_{Se} can be incorporated to 5' and any internal nucleotide positions of an

oligonucleotide. It is now feasible to add multiple Se-tags to one sequence in order to facilitate X-ray crystallography of large structures.

In structures that were revealed in X-ray crystallographic studies, the 2'-$SeCH_3$ groups lay in the minor groove of A-form DNA duplexes and they do not intrude normal structures of double helices. The melting temperature of the modified duplex was slightly increased compared with native counterpart. The stability of the strand to strand hybridization is evidently maintained after Se-labeling. These are very positive results toward the future application of 2'-Se-labeled uridine.

Chapter 5. Incorporation of Selenium to 3'-positions of DNA Nucleosides

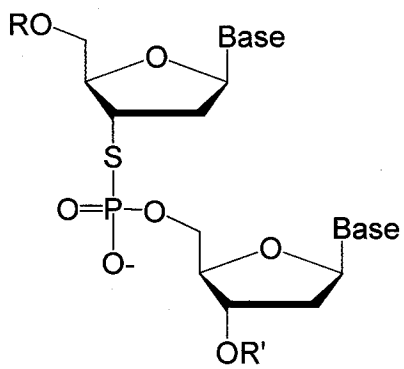
5.1 Introduction

5.1.1 Modification at the 3'-deoxyribose of nucleic acids

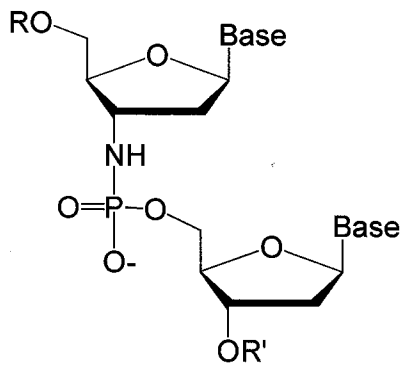
The work discussed in prior chapters (Chapter 3, and 4) has demonstrated incorporation of selenium functional groups to 5' or 2'-position of nucleosides and oligonucleotides. In the case of 2' selenium incorporation, a 2,2'-anhydro-uridine compound was found to be a satisfactory intermediate for two reasons: 1) the arrangement of chemical bonds makes 2'-carbon (C-2') susceptible to substitution from nucleophile such as methylselenide; 2) the tricyclic ring geometry sustains a rigid conformation around C-2', so that a nucleophile would attack exclusively from 2'-α-position in terms of stereochemistry.

These results suggest that a similar mechanism may also be applied to the 3'-α-position of nucleic acid. The most intriguing aspect of introducing a non-oxygen element to this position is: the change of 3'-5' phosphodiester backbone in natural DNA and RNA. Two good examples (Figure 5-1) of 3' modification are the 3'-S-phosphorothiolate linkage made by Cosstick

group,(95) and the N3'→P5' phosphoramidate linkage made by Gryaznov group.(96) Like our prior work on the 2' and 5', chemical modification at the 3' has the advantage of preserving duplex formation and Watson-Crick base-pairing. On the other hand, making an oligonucleotide out of 3' modified nucleoside analogue is relatively challenging: the phosphate diester linkage is disrupted, and the traditional phosphoramidite chemistry may not be directly applied here. The limited species of available modifications on C3'-*O*-P also indicates the difficulty in obtaining those products.



3'-S-phosphorothiolate linkage



N3'-P5' phosphoramidate linkage

Figure 5-1. Oligonucleotides with modified 3'-bridging groups.

Study on a RNA dinucleotide IspU (I = inosine, U = uridine) with 3'-S-phosphorothiolate linkage found the 3'-sulfur substitution made RNA even more sensitive to base hydrolysis than native counterpart.(97) The

linkage is also a substrate for enzymes like SVPD (snake venom phosphodiesterase) and ribonuclease T2. Long sequences with up to nine N3'→P5' phosphoramidate linkers were synthesized, according to literature.(98) The N3'→P5' phosphoramidate linkage was found to increase melting temperature by 0.7 °C and 2.7 °C per modification when hybridized to single-stranded DNA and RNA, respectively. The N3'→P5' phosphoramidate linkage is also resistant to nucleases.

5.1.2 Organic chemistry: 2' or 3'

During the synthesis of compound **4.5**, it was found that only 2,2'-*O*-anhydro-uridine is formed as the product; no 2,3'-*O*-anhydro-uridine formation was observed. This can be explained in part by the geometry of the A-form sugar pucker of uridine: the 2'-exo conformation provides the favorable angle and distance clearance for O-2 nucleophilic attack on C2' from behind. In addition, the thermodynamics favors the formation of five-member-ring over six-member-ring (Figure 5-2). There is an obvious disadvantage for the formation of 2,3'-O-anhydro-uridine, which requires more stringent reaction conditions.

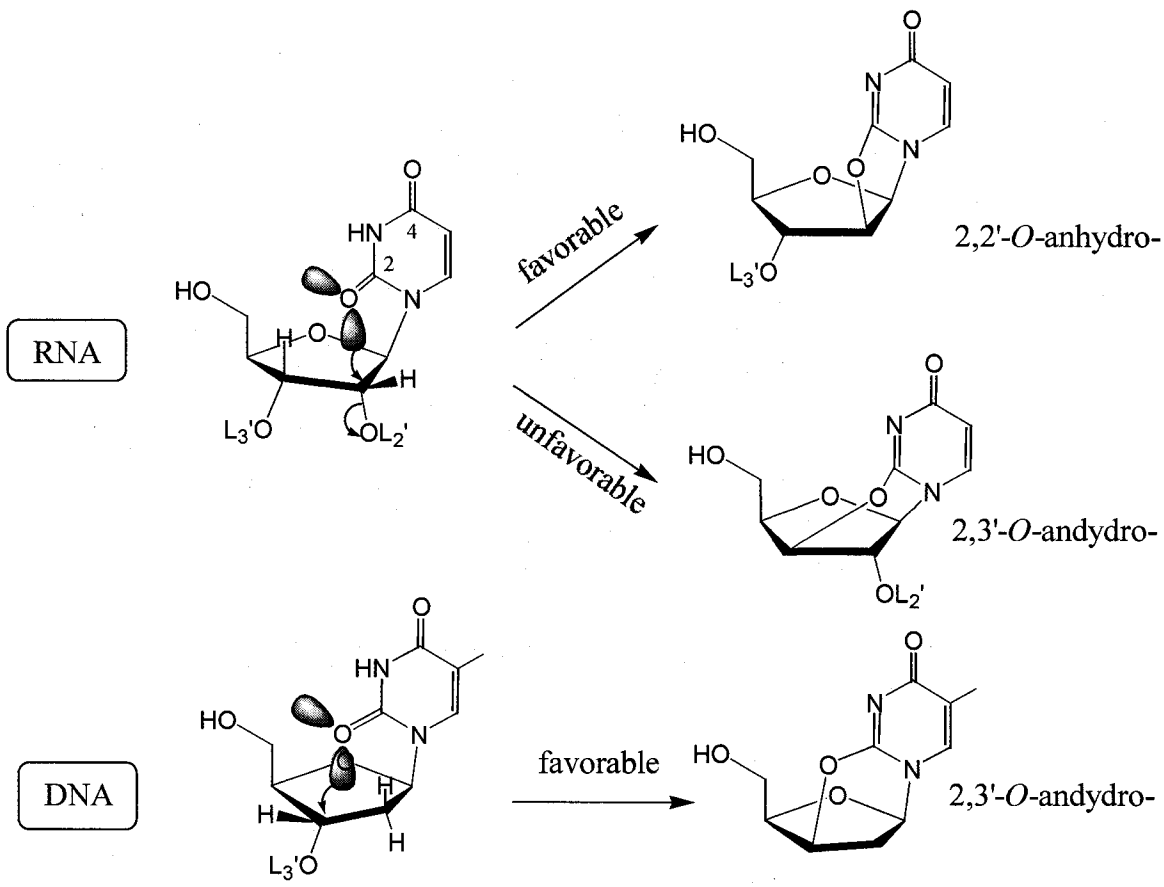


Figure 5-2. Protected uridine after weak base (Na_2CO_3) treatment. The preference of sugar pucker geometry led to the actual 2,2'-O-anhydro product (favorable) in RNA. The 2,3'-O-anhydro product is allowed in DNA. L_2' and L_3' are imaginary leaving groups.

In order to obtain 2,3'-*O*-anhydro-compound with high yield, changes must be made to the starting material as well as the reaction condition. The competition from the formation of 2,2'-*O*-anhydro-uridine can be avoided by choosing deoxyribonucleosides (dC, T), which lack the 2'-OH group.

5.1.3 Goals of chemical synthesis

Thymidine was chosen as the platform to do 3' selenium incorporation: thymidine has the desired pyrimidine base for the formation

of tricyclic intermediate; and thymidine does not need nucleobase protection during regular work up required in cytidine.

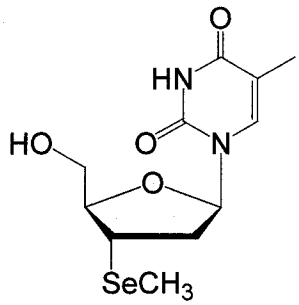


Figure 5-3. Target compound **5.6**.

One target novel compound to be synthesized is the 3'-methylseleno-thymidine **5.6**. In compound **5.6**, the 3'-OH of thymidine is replaced by a stable methylselenide "cap". The cap does not allow addition of phosphoramidite, or various chemical linkers. As a result, this residue is not ideal for typical 3'→5' DNA solid-phase synthesis.

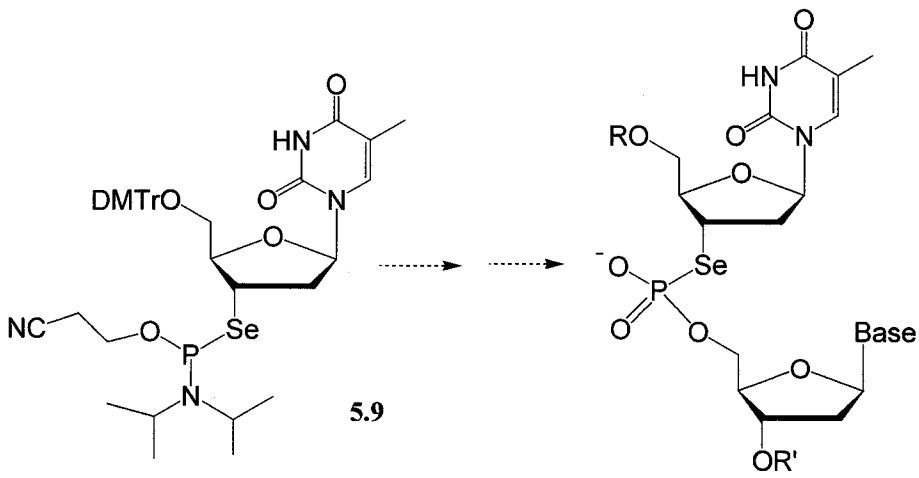


Figure 5-4. Target compound 5.9, and the DNA linkage after solid-phase synthesis.

Another novel compound to be synthesized is the thymidine 3'-selenophosphoramidite **5.9**. The design of the molecule mimics regular phosphoramidite but the 3' bridging oxygen is converted to selenium. The intention of making the molecule is to incorporate Se-modified backbone into DNA by solid-phase synthesis. Like the Se-modification made on the 2' of uridine, this new strategy will probably allow multiple Se insertion when necessary.

5.2 Experiments and results

5.2.1 Synthesis of compound 5.6

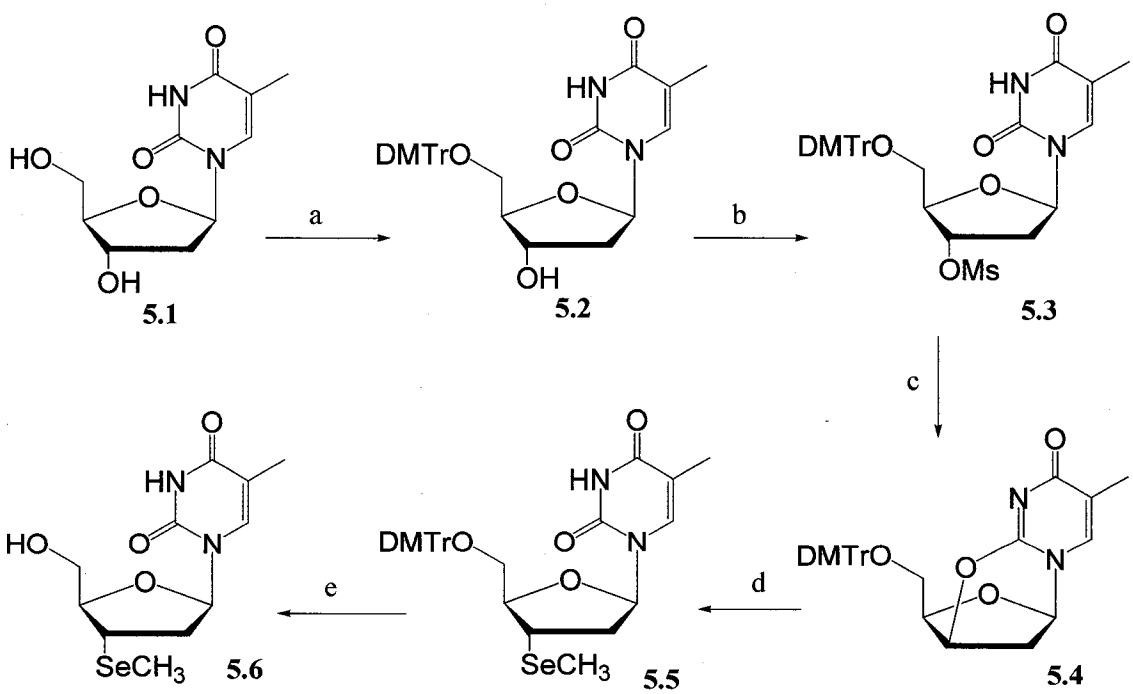


Figure 5-5. The synthesis of compound **5.6**. a) DMTrCl/Pyridine; b) MsCl/TEA/THF; c) 1eq NaOH (aq)/EtOH, reflux; d) $NaSeCH_3$/THF; e) 80% HOAc (aq).

Selective dimethoxytrityl (DMTr) protection of the 5'-OH on thymidine **5.1** was carried out by standard procedures. In the following step, the free 3'-OH was protected by mesyl group (Figure 5-5).

Mesyl (methanesulfonyl) group was found to be a good leaving group during the formation of 2,2'-O-anhydro-uridine. Addition of a weak base, i.e. $NaHCO_3$, Na_2CO_3; a H_2O/toluene two phase solvent; and room temperature were sufficient to convert all uridine reactant in a matter of hours. But when the 3', 5' double protected thymidine **5.3** was left in similar conditions, no detectable conversion took place. Finally, a protocol modified from literature was adopted. Compound **5.3** was diluted to 10mM in ethanol/H_2O containing 1.05 equivalent of NaOH, and was refluxed for 2.5 hours. 95% yield was achieved for compound **5.4**.

It is now clear that the formation of compound **5.4** is far more difficult than 2,2'-anhydro-uridine **4.5**: 1) a strong base was used; 2) one phase solvent instead of two; 3) reaction was refluxed at high temperature. The observation from experiments is consistent with structural analysis mentioned earlier.

The presence of water is crucial to the success of this reaction, even though it is less than 10% (V/V) in the solvent mixture. A probable explanation is that water is essential for better solvation of NaOH.

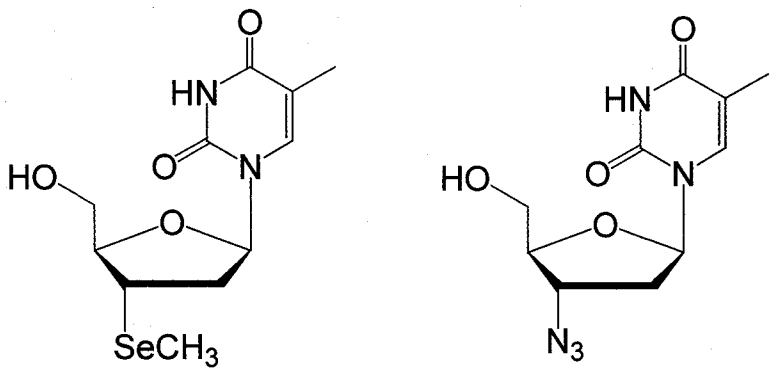
The reaction between **5.4** and reduced methylselenide occurred under similar condition used in Chapter 4. Nucleophilic attack by methylselenide from α-position of C-3' opens the 2,3'-O-anhydrous tricyclic structure. Thymine structure is restored and the methylselenide group attaches exclusively on 3'-α-position in product **5.5**. However, conversion of most reactant to **5.5** may take days, which is much slower than opening 2,2'-O-anhydrous ring structure.

Finally, the 5'-DMTr group on **5.5** was deprotected in acid and **5.6** was obtained.

5.2.2 Possible applications of compound 5.6

Compound 5.6 has similar structural features like many approved antiretroviral (e.g. anti-HIV) drugs. AZT was one of the first approved anti-HIV drugs and is still among the few that are currently used in clinically treatment. Viral reverse-transcriptase (RT) has no proofreading function

hence high error rate. RT may mistakenly incorporate AZT triphosphate in host cells to viral DNA. The result is the termination of viral DNA elongation because of the 3' azide cap. A few other nucleoside-based antiviral drugs work using the same mechanisms.(54)



5.6 **AZT**

Figure 5-6. Structures of **5.6** and AZT.

Since some HIV strains have developed resistance against known drugs, compound **5.6** has its potential for further therapeutic testing. Long before their use on viral treatment, nucleoside analogues were applied to anticancer and antibiotic therapies since the 1950s. Moreover, a number of small selenium compounds were found to inhibit different cancers and many more are being tested as drug candidates. Compound 5.6 has the combination of those two features in one structure.

5.2.3 Synthesis of 5.8 and 5.9

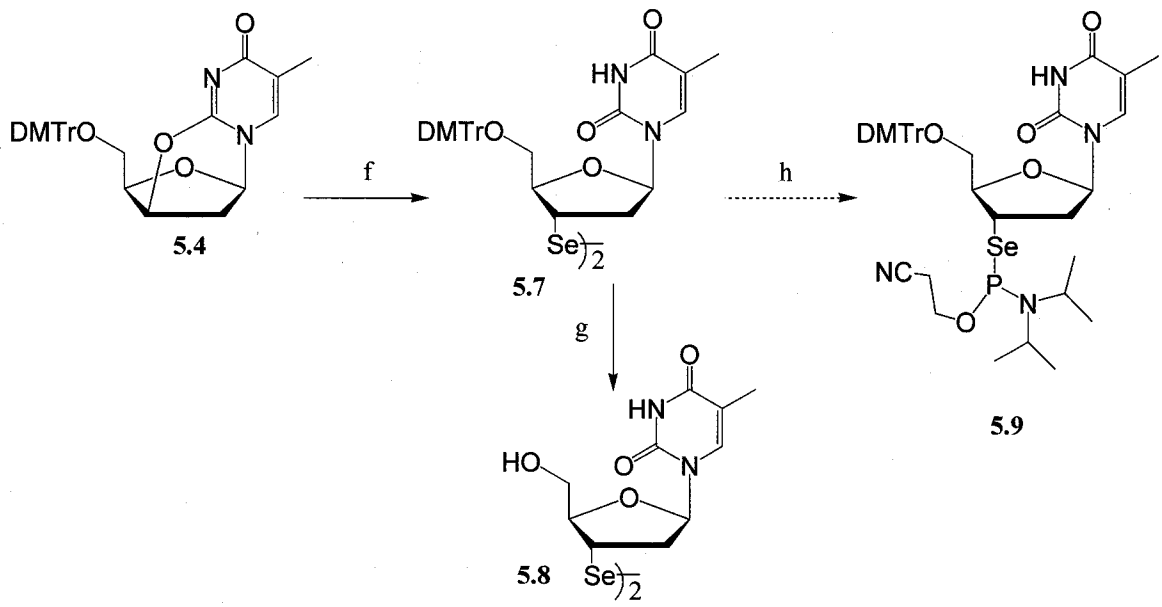


Figure 5-7. Synthesis of compound **5.8** and **5.9**.

Intermediate **5.4** provides a standard platform for introducing different nucleophiles to 3'-α-position of thymidine. Diselenide (-Se-Se-) is a good candidate because it may help to generate a novel selenophosphoramidite group.

Sodium diselenide was prepared following existing protocol. Se metal was reduce by $NaBH_4$ in 1:1 volume of THF and ethanol. Compound **5.4** was dissolved in the solution under dry nitrogen. The system was refluxed overnight and a few different products were generated. Desired 3'-seleno-dimer **5.7** was the major product and was purified by silica gel

chromatography. It has a distinct bright yellow color in ambient light. The side products migrated very close to **5.7** during chromatography, only a small amount of clean product (~20mg) was obtained from each silica gel plate.

5.8 was produced by acid treatment of **5.7**. Synthesis of **5.8** was first published in 2001 from a totally different approach.(99) The method presented here has noticeable advantage because of its simplicity and higher yield. In the published result, compound **5.8** showed cytostatic activity in colon and breast cancer cells.

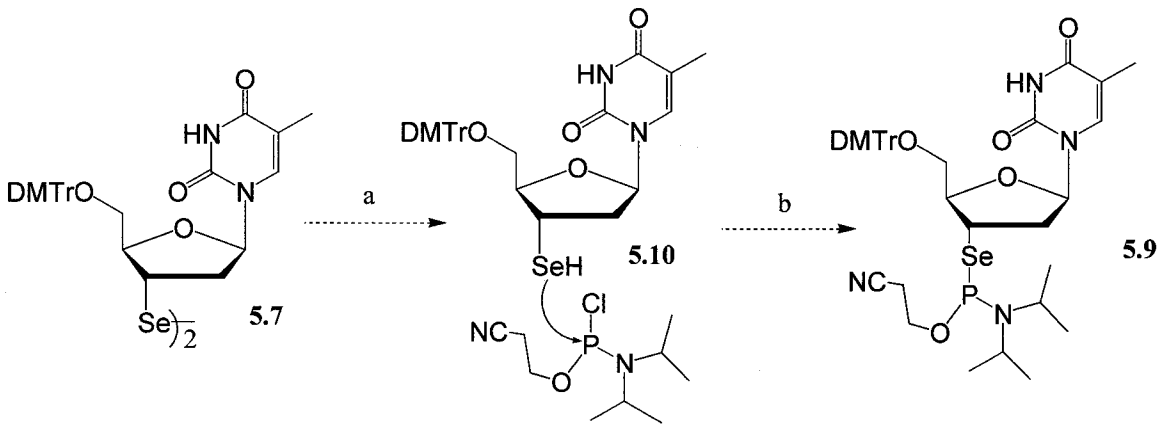


Figure 5-8. Proposed procedure to make **5.9**. (a) Reduction of diselenide; (b) Coupling between selenol and chlorophosphoramidite.

Samples of purified **5.7** were also used to search for synthetic condition for proposed compound **5.9**. Diselenide (-Se-Se-) can be reduced to selenol (-SeH) by $NaBH_4$. The major challenge here is, free selenol is extremely unstable in air and it'll be oxidized quickly by oxygen. It's

impossible for us to purify and obtain clean **5.10** using currently available facility, because air exposure is inevitable. Theoretically, selenol is a better nucleophile than hydroxyl so that the Se-P coupling should not be a problem. All in all, the key issue is to find a way to do reduction and subsequent coupling in strictly anaerobic condition, preferably in one sealed vial.

Reduction was carried out in anhydrous THF or acetonitrile by $NaBH_4$. It was tried in argon atmosphere initially. After a few minutes of gentle stirring, the light yellow color of **5.7** faded to colorless due to the reduction of diselenide. Amine base and chlorophosphoramidite were injected sequentially to the sealed flask. Products of lower polarity were purified by hexane precipitation followed by silica gel chromatography. ^{1}H-NMR seemed to support the formation of **5.9** for the appearance of new peaks at high field, although integration numbers were off. A series examination of mass spectroscopy somehow failed, to confirm **5.9** or gave any reasonable clues. Later experiments were conducted in vacuum-sealed rotary evaporator. But no positive improvement was obtained.

5.2.4 Trials of making dinucleotide with 3'-*Se*-phosphoroselenolate linkage

The unconfirmed compound **5.9** was used in solid-phase synthesis in order to make a TT dinucleotide with 3'-*Se*-phosphoroselenolate linkage. The synthesis was conducted manually in eppendorf tubes instead of using a column/synthesizer machine.

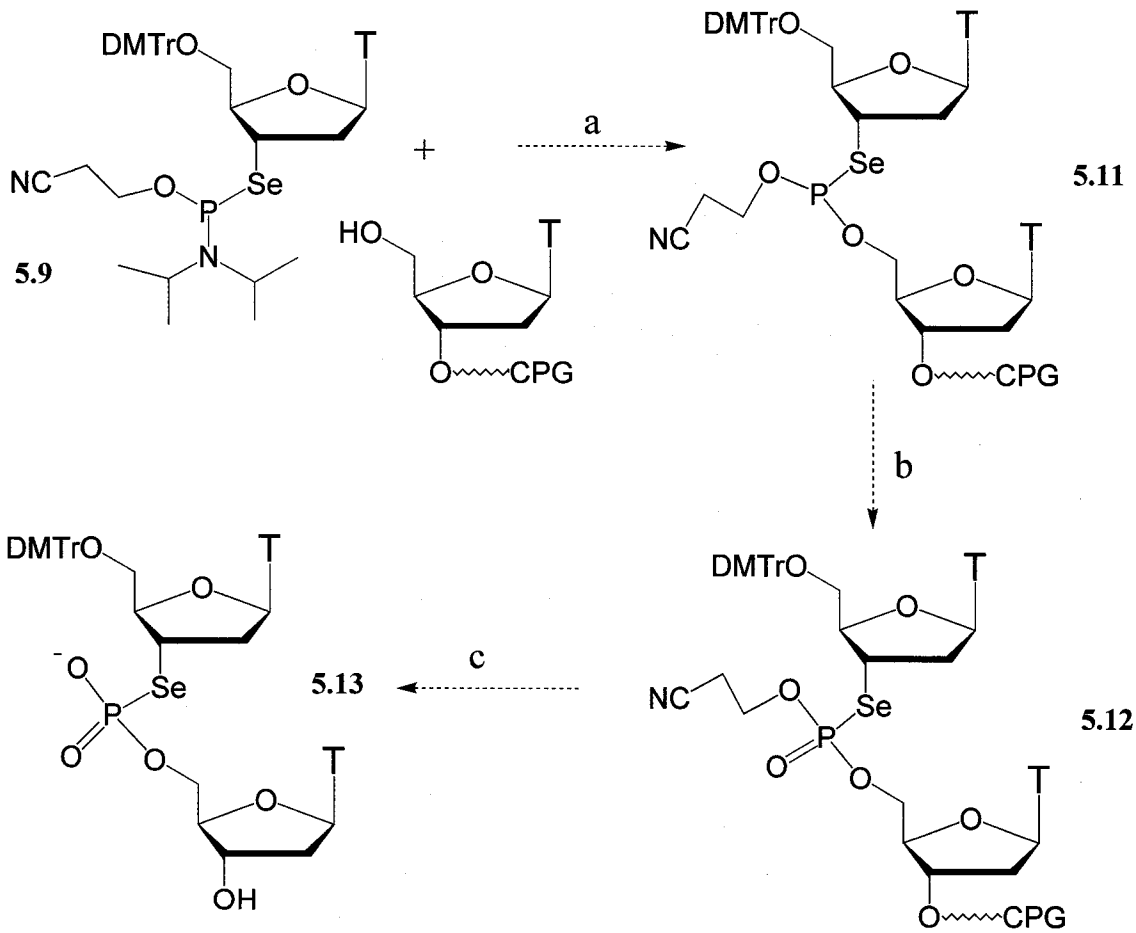


Figure 5-9. Proposed synthesis of dinucleotide from **5.9**. (a) 5-ethylthio-tetrzole/acetonitrile; (b) O_2; (c) NH_3/H_2O.

Conditions have followed standard DNA solid-phase synthesis unless specified. In an eppendorf tube, commercial CPG beads with 3'-anchored thymidine were deprotected with dichloroacetic acid and then washed thoroughly with anhydrous acetonitrile. **5.9** was premixed with 1H-tetrazole in acetonitrile to a 0.1M final concentration. This solution was pipetted to the beads at 50:1 equivalent ratio. The tube was gently shaken for 10 minutes. Afterward, the beads were washed thoroughly with acetonitrile.

A small portion of the beads containing presumed **5.11** were taken out as a sample. The sample showed orange color after it was stained by trityl indicator solution. This result suggested the coupling reaction was successful.

Phosphite is generally oxidized to phosphate by I_2 in DNA solid-phase synthesis. After **5.11** was treated with I_2 solution, another portion of the beads were taken out as a sample. After trityl staining, this second sample did not show any visible orange color. It suggested that the linker between dinucleotide **5.11** was probably cleaved so that the 5'-trityl was lost from the beads.

Since I_2 did not prove satisfactory for oxidation purposes, we switched to a milder oxidant: oxygen gas. Beads containing **5.11** in acetonitrile was bubbled by air for 30 minutes in order to produce **5.12**. Sample beads showed presence of trityl group by staining method. Finally, the dinucleotide was cleaved from CPG solid support by 28% NH_4OH.

Mass spectroscopy showed a variety of m/z signals, which were difficult to interpret. No major mass peak exactly matched anticipated value for **5.13**. Furthermore, no Se isotopic distribution pattern was observed in mass spectra.

5.3 Discussion

A novel compound **5.6** was successfully synthesized. It will be screened in further studies as potential antiviral or anticancer agent. Even though **5.6** is labeled with selenium, it is not ideal for labeling an oligonucleotide: obviously the 3'-$SeCH_3$ group will block the typical 3'→5' solid-phase synthesis.

In fact, unorthodox methods will be able to allow the coupling of **5.6** to the 3'-termini of DNA oligonucleotides. For instance, special protocols allow 5'→3' solid-phase synthesis. After the 5'-OH of **5.6** is activated by regular phosphoramidite, this selenium-tagged nucleoside can be added to the 3'-end of a synthetic DNA strand. The method can be practiced on ordinary synthesizer machines, although it requires the purchase of specialized columns and DNA phosphoramidite agents. The scenario has not been tested in experiment, but it is highly plausible.

The quality of compound **5.9** was not yet proved. The trial synthesis of a dinucleotide **5.13** did not produce convincing or even rational mass spectra. On the contrary, trityl staining indicated positive coupling between phosphoramidite and the beads. The result suggests two possibilities: 1) the projected product **5.13** was not created, or it had low abundance so that mass spectroscopy did not pick up the anticipated signal; 2) **5.13** or one of its precursors (**5.11**, **5.12**) was created but it was not stable enough during oxidation, deprotection, storage, or even the process of mass spectrometry.

In future studies, the synthetic method of **5.9** will be improved to ensure the purity and consistency. Doing the reactions in a sealed nitrogen box may help to prevent the oxidation of intermediate and improve the yield.

The synthesis of the TT dinucleotide containing a 3'-*Se*-phosphoroselenolate linkage may also be improved. One concern during the past experiment is that manual washing by a pipette may not be sufficient to remove all leftover reagents (e.g. dichloroacetic acid, I_2) that are trapped inside the glass beads. Any of the leftover reagents from the previous synthetic step will affect the following step. Conducting the synthesis on synthesizer machine will waste some precious compound but it may handle the cleaning job better by pumping rapid streams of acetonitrile back and forth.

Another issue is the effect of I_2 oxidation on the dinucleotide linkage. Coupling was observed from trityl staining but the projected product **5.12** did not seem to survive after I_2 treatment. Also, the effectiveness of air oxidation needs to be examined in further detail. Since the linker is entirely novel, there is currently no safe way to ensure every actual step goes as anticipated.

Compound **5.8** is a potential drug for anticancer studies, according to literature.(99) The work shown above has presented an alternative and improved method to do the synthesis. The new method involves fewer steps and higher overall yield, and does not use highly toxic organic-selenium compounds.

Chapter 6. X-ray Crystal Structures of Selenium-derivatized Oligonucleotides: Proof of Principle

6.1 Introduction

2'-methylseleno-uridine (U_{Se}) was synthesized and incorporated into oligonucleotides by a novel method discussed in Chapter 4. Previous to the selenium story, covalent derivatization of nucleic acids for crystallographic phasing was limited to incorporation of 5-halogen-2'-deoxyuridine or 5-halogen-2'-deoxycytidine.

Bromine exhibits a 0.920 A theoretical K edge and MAD phasing has been performed on bromine-containing DNA crystals. In addition, numerous DNA and RNA crystal structures have been determined by multiple isomorphous replacement (MIR), using brominated fragments either alone or in combination with additional derivatives. However, applications of either MAD or MIR with bromination are often not successful in practice. Disruption or alteration of base-stacking and other structural perturbations caused by a bulky Br at the 5-position of pyrimidines may hamper crystallization efforts. In the majority of cases, halogen derivatives cannot be crystallized under native conditions. Moreover, derivative crystals sometimes do not diffract as well as the native ones. A further potential

problem associated with halogen derivatives is that the modified nucleotides are light sensitive. Thus, long-time exposure of halognated oligonucleotides to X-ray or UV sources may cause decomposition.

Selenium and oxygen both belong to Group 6A in the periodic table. In principle, Se can selectively replace oxygen atoms within nucleotide structure, including 2'-, 3'-, 5'-ribose oxygens, the furan oxygen, oxygens within phosphate linkage, and oxygens on the nucleobases. Theoretically, the X-ray anomalous signals of Br and Se are comparable to each other, which indicates that similar numbers of Br or Se atoms need to be present in longer nucleic acids for MAD phasing.

Se has previously been incorporated into DNA phosphate at bridging and non-bridging positions. 1) The replacement of 5'-bridging-oxygen by Se was done in Kool's group: a procedure that involves joining two short DNA strands together by a selenophosphate tail in one strand and an iodide head in the other strand. This method allows only one Se insertion per DNA, and the preparation of the two precursor strands was not simple. 2) Se was also used to replace one of the non-bridging oxygen atoms and subquently form the so-called phosphoroselenoate. The syntheses of phosphoroselenoates are usually carried out by customized phosphoramidite chemistry, where

selenium-transfer agent replaces the standard I_2 solution during the oxidation step. The use of phosphoramidite chemistry in 2) is a significant plus when compared with 1), because many laboratories with solid-phase synthesis facility could easily make custom phosphoroselenoate oligonucleotides. However, phosphoroselenoate linkage suffers from decomposition with a half-life around 30 days.

U_{Se} derivatization enjoys the advantage of having the stable modification, as well as utilizing phosphoramidite chemistry. Methylselenide is an especially poor leaving group under most circumstances, so that it is very unlikely to observe $-SeCH_3$ substitution by another functional group during organic synthesis. Furthermore, no photo-lesion or other decomposition was observed after normal handling and storage. Lots of the experimental evidence suggested that U_{Se} is a superior candidate for MAD phasing and X-ray crystallography.

6.2 Materials and methods

6.2.1 Oligonucleotide synthesis and purification

Two DNA oligonucleotides were synthesized and purified separately by methods in Chapter 4: octamer $GU_{Se}GTACAC$ and decamer $GCGTAU_{Se}ACGC$. Standard 3'-CE deoxyphosphoramidites of A, C, G, T were used for regular bases. The phosphoramidite of U_{Se} was prepared by freshly before use. The coupling time for U_{Se} nucleotide was extended to 10 minutes to ensure high yield (we use standard coupling time nowadays, after improvement in compound quality). The syntheses of oligonucleotides were ended in trityl-on mode. After ammonia deprotection, the full-length DNA oligonucleotides were purified by reversed phase HPLC. The concentrated DNA solutions were detritylated by acid treatment. Detritylated DNA oligonucleotides were purified by RP HPLC again to give the final products.

6.2.2 UV thermal denaturation studies

Both DNA have self-complementary sequences. Two strands of the same specimen may hybridize with each other under ideal conditions. The DNA decamer was examined for thermal stability in Egli's group and the result was compared with the ones from the native sequence and the

sequence made with U_{MeO} (= 2'-methoxy-U). 1 μM oligonucleotides were dissolved in a buffer consisting of 90 mM sodium chloride, 10 mM sodium phosphate, and 1 mM EDTA (pH = 7.0). Molar extinction coefficients were calculated by the nearest neighbor approximation. The solutions were heated to 85 °C for 15 min, cooled slowly to room temperature, and stored at 4 °C overnight. Melting curves were measured at 260 nm at a heating rate of 0.5 °C/min.

6.2.3 Crystallizations

Crystals of the Se-derivatized octamer and decamer were grown separately, by the hanging drop method and the Nucleic Acid Mini Screen™ crystallization kit with 24 unique precipitant buffers (Hampton Research Inc.).

Purified DNA samples were dissolved in 2 mM aqueous solutions. Crystallization screens were carried out in 24-well VDX plates (Hampton Research): one buffer condition per well (reservoir). Matching cover slides were cleaned and siliconized individually. The reservoir rim was greased before use. 1 mL of 35% MPD (2-methyl-2,4-pentanediol) solution was pipeted to every reservoir as the dehydrant. Small volumes (1 or 2 μL) of

DNA solution and a buffer were mixed together on each slide. Afterward, the slides were carefully flipped and seated on designated reservoirs. In crystallization attempts on both DNA samples, single crystals were obtained in less than one week from multiple buffer conditions. Single crystals for X-ray data collection were grown from optimized condition that used equal volumes (2 μL each) of DNA solution and buffer.

6.3 Results and discussion

6.3.1 Hybridization stability of DNA oligonucleotides containing U_{Se}

UV melting temperatures (T_m) were measured to examine the effect of incorporation of U_{Se} residue on the overall thermodynamic stability of the DNA decamer duplex $[d(GCGTAU_{Se}ACGC)]_2$. Egli's group measured the T_m of the Se-derivatized DNA along with the native duplex $[d(GCGTAUACGC)]_2$ and a duplex $[d(GCGTA)U_{MeO}d(ACGC)]_2$ that contains a 2'-methoxy-uridine (U_{MeO}).

Oligonucleotide	T_m (^oC)
GCGTAUACGC	58.2
$GCGTAU_{MeO}ACGC$	63.3
$GCGTAU_{Se}ACGC$	58.8

Table 6-1. Melting temperatures (T_m) of the native, 2'-methoxy- (U_{MeO}) and 2'-methylseleno-modified (U_{Se}) DNA oligonucleotides.

Incorporation of one U_{Se} residue per strand has virtually no effect on the thermal stability of the decamer duplex relative to the native control (Table 6-1). The T_m of $[d(GCGTAU_{Se}ACGC)]_2$ is raised by approximately 0.3 °C from the addition of a single methylseleno groups. The result clearly indicates that the hybridization strength of the duplex is maintained (in fact, it's increased slightly) after the chemical modification. This is a positive aspect of U_{Se} application in crystallography, because its incorporation will not destabilize the hybridization between two complementary DNA strands.

It is well known that 2'-methoxy substitution in DNA leads to increased stability of duplexes between this modified DNA and a native RNA relative to the parent DNA:RNA hybrid. That increase in T_m is about 1.5 °C per substitution. In the current case, the T_m increase of a DNA:DNA hybrid from 2'-methoxy substitution is about 2.5 °C per substitution. Apparently, selenium cannot counter all the stability increase that is brought by oxygen at the 2'-position.

6.3.2 Incomplete X-ray crystal structure of DNA octamer solved by MIR

A partially completed X-ray crystal structure of DNA octamer duplex $[d(GU_{Se}GTACAC)]_2$ was determined by method of molecular replacement

by Professor X.P. Kong. The structure shows a characteristic A-form double helix with canonical Watson-Crick base pairing. Electron density maps confirmed the presence of the selenium group (again!) at the 2'-α-position of uridine residue. The 2'-$SeCH_3$ groups are directed into the minor groove with antiperiplanar orientation around the C2'-Se bond, without clashing surrounding structures in vicinity. Despite the increased physical size, the methylseleno group seems to fit nicely into the minor groove of the A-form duplex.

The 1.4 Angstrom resolution structure of native octamer duplex $[d(GTGTACAC)]_2$ was solved by X-ray crystallography previously. Interestingly, the duplex adopted A-form structure even with the pure DNA constituent. This could be the result of crystallization condition, nucleotide sequence, or both.

The only difference in formula between the two octamers is inside the nucleotide number 2: the thymine base is replaced by a uracil; and the 2'-α-H is substituted by methylseleno group, in modified DNA. The conversion of thymine to uracil should not pose much disparity in base pairing. The real spotlight was focused on the effect from the addition of 2'-$SeCH_3$ group.

Encouragingly, the overall structure of derivatized duplex matches very well to the native duplex.

The comparison of the X-ray crystal structures of two DNA octamer duplexes satisfied the anticipation that the 2'-$SeCH_3$ will not cause clashes or vast changes to the structure. However, the crystallography process did not use the MAD technique, which was the real intention of Se-derivatization. The feasibility was later tested on the crystals of Se-derivatized DNA decamer.

6.3.3 MAD structure determination

In order to select optimal wavelengths for MAD data collection, X-ray fluorescence spectrum was recorded from a $[d(GCGTAU_{Se}ACGC)]_2$ decamer duplex crystal. Compared with the K edge X-ray spectrum of a protein crystal, the spectrum of modified DNA crystal exhibits a somewhat broader peak and the maximum is shifted to higher energy by about 5 eV.

Datasets at three wavelengths to a maximum resolution of 1.3 Angstrom were collected from a single crystal. Comparison of the intensities of the Friedel pairs reveals a striking effect of two Se atoms per

crystallographic asymmetric unit. The magnitude of the anomalous differences is undoubtedly due in part to the presence of one Se atom per DNA strand. This is approximately comparable to a frequency of one Se-Met residue per 20 amino acids in a protein. We need to bear in mind that the occupancy of selenium sites in the U_{Se}-modified DNA is 100% due to the nature of tandem solid-phase synthesis. However, the replacement of Met by Se-Met during protein biosynthesis will not proceed with 100% efficiency. A further difference is generated by the flexibility of the Se-Met side chain in proteins and the 2'-$SeCH_3$ group in DNA double helices. In most cases (except termini and bulges), the sugar moieties in duplexes are well ordered in their crystal structures. As expected, both 2'-$SeCH_3$ groups in the decamer crystal structure are ordered and all atoms were refined with crystallographic occupancies of 1.0.(90)

Initially, the locations of selenium atoms were determined from MAD density maps, relying on all three data sets (peak, inflection point and reference). In addition, Egli's group also confirmed that data collected at the peak wavelength alone was sufficient to allow successful phasing of the DNA decamer (single wavelength anomalous dispersion, SAD).

6.3.4 Structure of $[d(GCGTAU_{Se}ACGC)]_2$

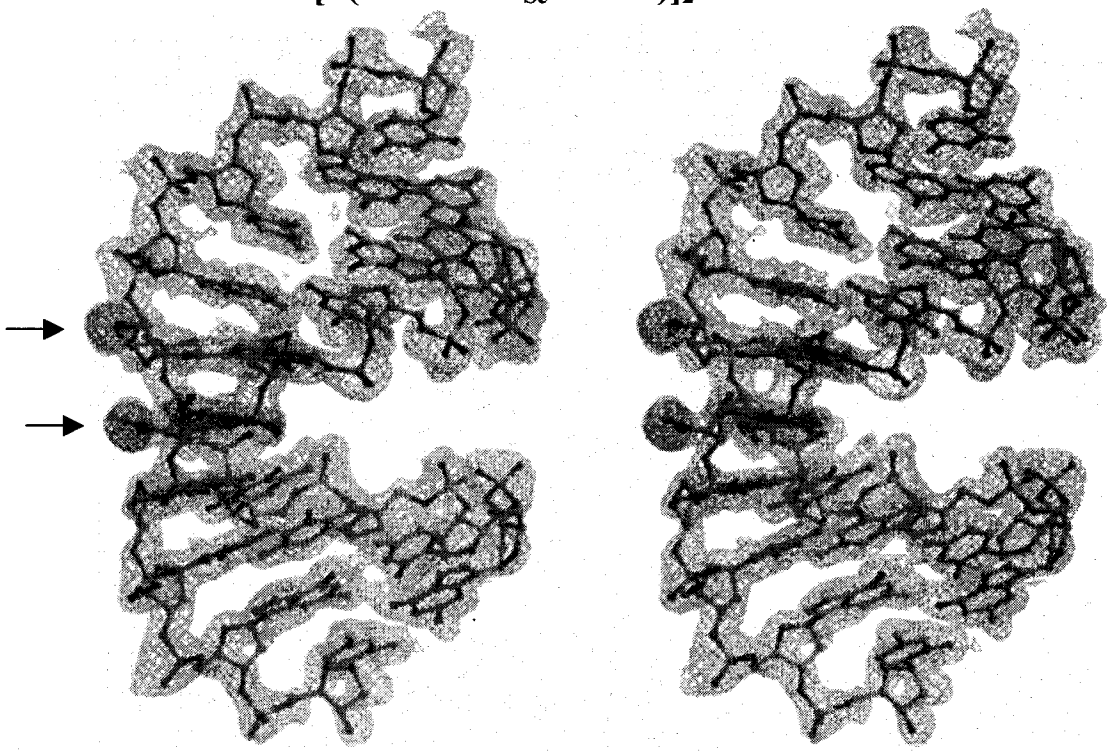


Figure 6-1. Stereo diagram depicting electron density calculated at 1.3 angstrom resolution with MAD phases and superimposed on the final structure of the decamer duplex. The two Se atoms per asymmetric unit are pointed by the arrows.

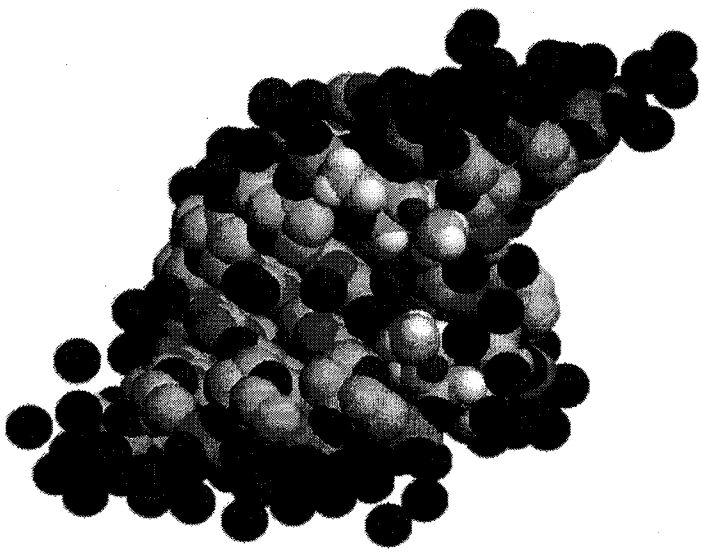


Figure 6-2. Structure of the decamer duplex $[d(GCGTAU_{Se}ACGC)]_2$ in spacefilling display, showing the minor groove.

The crystal structure (Figure 6-1) of the octamer DNA duplex $[d(GCGTAU_{Se}ACGC)]_2$ was deposited in the Protein Data Bank (1MA8) and the Nucleic Acid Database (AD0027). The structure is isomorphous to that of the $[d(GCGTA)T_{MeO}d(ACGC)]_2$ (T_{MeO} = 2'-methoxy-3'-methylenephosphonate thymidine). In U_{Se}-modified DNA duplex, the two selenium atoms are separated by 9.1 angstrom across the minor groove. The $A5pT_{MeO}(U_{Se})6$:$A15pT_{MeO}(U_{Se})16$ base pair steps of the two duplexes exhibit virtually identical conformations. The result clearly indicates that 2'-α-$SeCH_3$ substitution does not distort the A-form geometry of double helices. Even with the two bulky 2-$SeCH_3$ groups located closely in adjacent base pairs, they are accommodated nicely within minor groove and no structural distortion is generated (Figure 6-2).

All sugars in the decamer adopt C3'-endo pucker and the pseudorotation phase angles (*P*) for the sugars of $U_{Se}6$ and $U_{Se}16$ are 13° and 20°, respectively. The 2'-$SeCH_3$ and the 2'-OCH_3 both adopt antiperiplanar orientation in their crystal structures. The C3'-C2'-Se-CH_3 torsion angles for $U_{Se}6$ and $U_{Se}16$ are −175° and −172°. The C3'-C2'-O-CH_3 torsion angles for $T_{MeO}6$ and $T_{MeO}16$ are −171° and −175°, respectively. The bond lengths of

C2'-Se and Se-CH_3 are approximately 2.0 angstrom each. By comparison, the bond lengths of C2'-O and O-CH_3 are shorter by 0.6 angstrom (Figure 6-3).

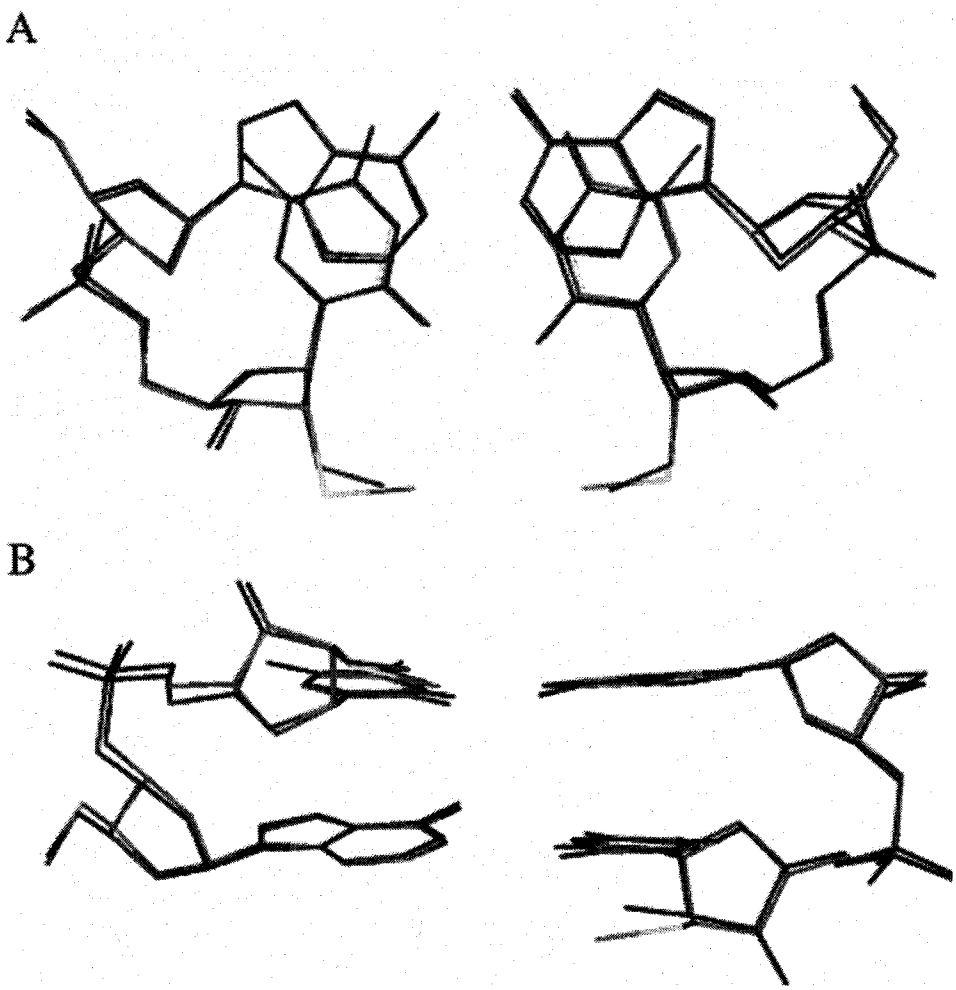


Figure 6-3. Superposition of A5pU_{Se}6:A15pU_{Se}16 and A5pT_{MeO}6:A15pT_{MeO}16 base pair steps in respective duplexes.

Selenium is more hydrophobic than oxygen and sulfur. But the higher hydrophobicity is of little consequence in terms of the number of water molecules surrounding the C2' substituents. Introduction of selenium does not appear to affect minor groove hydration to a significant extent.

6.4 Conclusion

Covalent incorporation of Se into DNA and RNA oligonucleotides at the 2'-position in combination with MAD or SAD opens up a new way to nucleic acid crystallography and potentially to that of protein-nucleic acid complexes. The approach provides a general, feasible, and rational way to derivatize nucleic acids for X-ray crystallography.

Compared with the traditional technique of heavy metal soaking or co-crystallization, the covalent derivatization by Se simplifies the crystallization process and is highly reliable. It does not have the stability problems associated with halogen modifications or phosphoroselenoates. Moreover, the conformation of 2'-α-$SeCH_3$ is unambiguous while phosphoroselenoate products usually come as diastereoisomers.

Available crystal structures of 2'-Se-derivatized DNA so far are A-form duplexes. This is satisfactory for molecules that are intended to be A-form, as in most RNA:RNA duplexes, RNA:DNA duplexes, etc. The preference between A-form and B-form structure is a comprehensive result of the crystallization condition, nucleotide sequence and chemical modifications. Based on earlier studies on 2'-methoxy (2'-OCH_3) nucleotides as well as many other C-2' modified nucleotides, the U_{Se} is very likely to adopt the same C3'-endo structure (A-form). The U_{Se} residue in a DNA sequence may serve a leader that induces conformational tendency to A-form in the entire duplex. Thus, the U_{Se} substitution could be undesirable for DNA duplexes that are intended to possess B-form structures.

APPENDIX

Basic ^{1}H and ^{13}C NMR spectra were collected on Bruker 250 MHz machine at Brooklyn College. ^{77}Se, NOE, gs-HMQC NMR spectra were collected on Varian 600 MHz machine at College of Staten Island. Some experiments including ^{1}H, ^{31}P and ^{1}H-^{1}H COSY NMR were performed on a JOEL 500 MHz and a 300 MHz machine at Hunter College. Low- resolution ESI-MS data were collected at mass spectrometry facility of Hunter Collge. High-resolution mass spectrometry data were obtained from MS service at the Scripps Research Institute. UV spectra were recorded on a Shimadzu 240 UV/VIS spectrometer. IR spectra were recorded on Perkin Elmer 781 spectrometer. Solid-phase DNA oligonucleotide syntheses were conducted on an Applied Biosystems 392 model.

Most organic solvents were purchased from Fluka, Sigma-Aldrich and used without purification unless mentioned otherwise. Solid reagents for organic syntheses were dried under high vacuum. Reactions that are sensitive to oxygen or moisture were performed under dry argon or dry nitrogen. Bands on thin layer chromatography (TLC) were visualized under short-wavelength UV light. The presence of trityl group (DMTr-) was examined by a Ce-Mo staining solution (phosphomolybdate 25 g, $Ce(SO_4)_2 \bullet$

$4H_2O$ 10 g, conc. H_2SO_4 60 mL, H_2O 940 mL). Preparative TLC was carried out on thick (2 mm) pre-coated plates. Flash chromatography was done using silica gel G60.

Chapter 3.

N^6-Benzoyl-1-[(2R, 4S, 5R)-4-*tert*-butyldimethylsilyloxy-5-methanesulfonyl-methyl-tetrahydrofuran-2-yl]-adenine (3.4c-A).

3.3-A (54.0 mg, 0.115 mmol) was placed in 10-mL flask and dried on high vacuum for 2 hours. THF (2.3 mL) and TEA (47 μL, 0.345 mmol, 3 eq) were then added, and the flask was placed on an ice-water bath and kept under dry argon. Methanesulfonyl chloride (13 μL, 0.17 mmol, 1.5 eq) was added and the reaction was completed in 15min. (silica gel TLC in 5% $MeOH/CH_2Cl_2$, **3.4c-A** $R_f = 0.35$, **3.3-A** $R_f = 0.30$). MeOH (1mL) was added to consume the excess reagent and the reaction was stirred for another 15 minutes. The solvents were removed by rotary evaporation at 40^0C; the residue was dissolved in EtOAc (15 mL), and the solution was filtered. The filtrate was then evaporated, and the residue was purified on silica gel G60 column (gradient, from 0 to 5% $MeOH/CH_2Cl_2$). The collected fractions were evaporated under reduced pressure and dried on high vacuum overnight. A colorless foamy product (**3.4c-A**) was obtained (60 mg, 98% yield).

^{1}H-NMR ($CDCl_3$). δ: 0.13 [s, 6H, $(CH_3)_2Si$], 0.92 (s, 9H, t-Bu), 2.45-2.55 (m, 2H, 3'-H), 2.98 (s, 3H, CH_3SO_3), 4.20-4.26 (m, 1H, 5'-H), 4.40-4.52 (m, 2H, 5"-H), 4.70-4.78 (m, 1H, 4'-H), 6.45-6.53 (t, J = 6.6Hz, 1H, 2'-H), 7.49-7.65 (m, 3H, Ar), 8.01-8.08 (m, 2H, Ar), 8.20 (s, 1H, 8-H), 8.80 (s, 1H, 2-H), 8.95-9.20 (b, 1H, NH, exchangeable with D_2O).

ESI-MS (positive ion mode): $[M + H]^+$, 548.0. $[M + Na]^+$, 570.0. Calculated m/z = 547.0.

N^6-Benzoyl-1-[(2R,4S,5R)-4-*tert*-butyldimethylsilyloxy-5-methselenomethyl-tetrahydrofuran-2-yl]-adenine (3.5d-A).

$NaBH_4$ (20.0 mg, 0.525 mmol) was placed in 10mL-round flask under nitrogen. Water (1.5 mL) and dimethyl diselenide (17.0 μL, 0.175 mmol) were sequentially and slowly injected into the flask. Vigorous stirring helped to dissolve dimethyl diselenide completely, forming a colorless homogeneous solution after 5 –10 minutes; the pH of the solution was higher than 11. Since high pH caused the hydrolysis of the protecting benzoyl group on adenine base, and pH 7.0 or lower made the following selenide substitution very slow, the pH was adjusted to 8.0 by adding dilute HCl dropwise. **3.4c-A** (19.2mg, 0.0350mmol) and tetrahexylammonium hydrogen sulfate (0.5 mg) dissolved in toluene (0.7 mL) were then added to the sodium methylselenide solution (pH 8.0) described above, and the two-phase mixture (toluene and water) was stirred under nitrogen. The reaction

was complete after 5 hours, forming **3.5d-A** (silica gel TLC, 5% MeOH/CH_2Cl_2, R_f = 0.52). Longer reaction time caused slow hydrolysis of the benzoyl group (the hydrolyzed product R_f = 0.33 on TLC, 5% MeOH/CH_2Cl_2). The organic phase was removed and the aqueous phase was extracted twice with EtOAc; the combined organic phase was washed twice with saturated NaCl solution. The solvents were removed by rotary evaporation under reduced pressure at 40°C. The crude product was dissolved in CH_2Cl_2 and loaded on a silica gel TLC plate (5% MeOH/CH_2Cl_2). Colorless product was recovered from this purification (18.2 mg, 95% yield).

^{1}H-NMR ($CDCl_3$). δ(ppm): 0.12 [s, 6H, $(CH_3)_2Si$], 0.92 (s, 9H, t-Bu), 2.02 (s, 3H, CH_3Se), 2.43-2.53 (m, 1H, 3'-H), 2.78-3.00 (m, 3H, 3'-H and 5"-H), 4.15-4.23 (m, 1H, 5'-H), 4.55-4.63 (m, 1H, 4'-H), 6.42-6.50 (t, J = 6.6Hz, 1H, 2'-H), 7.48-7.57 (m, 2H, Ar), 7.57-7.65 (m, 1H, Ar), 8.00-8.07 (m, 2H, Ar), 8.27 (s, 1H, 8-H), 8.80 (s, 1H, 2-H), 8.94-9.00 (b, 1H, NH, exchangeable by D_2O).

The molecular weight of **3.5d-A** ($C_{24}H_{33}N_5O_2SiSe$) is 547 with adjustment for ^{80}Se isotope [average atomic weight of Se is 79, including 76 (9%), 77 (7%), 78 (23%), 80 (49%), 82 (9.2%)]. The major molecular peaks (measured by ESI-MS) are: 546 $[M(^{78}Se) + H]^+$, 548 $[M(^{80}Se) + H]^+$, 550 $[M(^{82}Se) + H]^+$, 568 $[M(^{78}Se) + Na]^+$, 570 $[M(^{80}Se) + Na]^+$, 572 $[M(^{82}Se) + Na]^+$.

1-[(2R,4S,5R)-4-*tert*-butyldimethylsilyloxy-5-methselenomethyl-tetrahydro-furan-2-yl]-thymidine (3.5d-T).

$NaBH_4$ (63 mg, 1.65 mmol) was dissolved in 1.5 ml water and sealed in a 10-mL flask under nitrogen. Dimethyl diselenide (54 μL, 0.55 mmol) was slowly injected into the flask. Addition of 0.2 mL of ethanol with vigorous stirring helped to dissolve dimethyl diselenide completely; a colorless homogeneous solution was formed in 5 minutes. The solution of **3.4c-T** (47.7mg, 0.110mmol) and tetrahexylammonium hydrogen sulfate (1 mg) in toluene (1.1 mL) was then added to the sodium methyl selenide solution described above. The two-phase mixture (toluene and water) was stirred under nitrogen. The reaction was complete in 3 hr. as indicated silica gel TLC (5% MeOH/CH_2Cl_2, R_f = 0.40). The organic phase was removed, the aqueous phase was extracted twice with EtOAc (10 mL each time), and the combined organic phase was washed twice with saturated NaCl solution (10 mL each time). The solvents were removed by rotary evaporation under reduced pressure at 40°C. The crude product was purified on TLC (5% MeOH/CH_2Cl_2). The pure product (**3.5d-T**) was dried on high vacuum overnight to afford 44.2 mg (93% yield).

^{1}H-NMR ($CDCl_3$) δ: 0.05 [(s, 6H, $(CH_3)_2Si$], 0.88 (s, 9H, t-Bu), 1.94 (s, 3H, 5-CH_3), 2.08 (s, 3H, CH_3-Se), 2.05-2.18 and 2.28-2.40 (2m, 2H, 3'-H), 2.78-2.93 (m, 2H, 5''-H), 4.02-4.09 (m, 1H, 5'-H), 4.28-4.38 (m, 1H, 4'-H), 6.26 (t, J = 6.5 Hz, 1H, 2'-H), 7.42 (s, 1H, 6-H), 8.95-9.06 (b, 1H, NH, exchangeable in D_2O).

^{13}C-NMR ($CDCl_3$) δ: 5.66 (CH_3-Se), 12.55 (5-CH_3), 17.68 (CH_3-Si), 25.68 [($\underline{C}H_3)_3C$], 27.63 (C-5''), 40.63 (C-3'), 73.94 (C-5'), 84.57 (C-4'), 85.60 (C-2'), 111.11 (C-5), 135.59 (C-6), 150.18 (C-2), 163.69 (C-4).

The molecular weight ($C_{17}H_{30}N_2O_4SiSe$) is 434 for ^{80}Se isotope. Because of binding of H or Na ions, two sets of molecular peaks are observed, in which all selenium isotopic peaks [Se 76 (9%), 77 (7%), 78 (23%), 80 (49%), 82 (9.2%)] are also observed: 431 [M (^{76}Se) + H]$^+$, 432 [M (^{77}Se) + H]$^+$, 433 [M (^{78}Se) + H]$^+$, 435 [M (^{80}Se) + H]$^+$, 437 [M (^{82}Se) + H]$^+$, 453 [M (^{76}Se) + Na]$^+$, 454 [M (^{77}Se) + Na]$^+$, 455 [M (^{78}Se) + Na]$^+$, 457 [M (^{80}Se) + Na]$^+$, and 459 [M (^{82}Se) + Na]$^+$.

IR (KBr): 3157, 3034, 2960, 2940, 2862, 2363, 1698, 1470, 1426, 1370, 1276, 1198, 1120, 1089, 1049, 827, 777, 682, 621 cm^{-1}.

UV (in acetonitrile): 265.2 nm.

Chapter 4.

2'-mesyl-3'-tert-butyldimethylsilyloxy-5'-dimethoxytrityl-uridine (4.4)

Partially protected uridine **4.3** (0.429 g, 0.65 mmol) was placed in a 25-mL round flask and dissolved in dry THF (6.5 mL, 0.1 M). In ice bath, TEA

(269 μL, 1.95 mmol) and methanesulfonyl chloride (76 μL, 0.98 mmol) were then added. The reaction mixture was stirred under argon for 20 min. After the reaction was complete, MeOH (0.5 mL) was added to quench the extra reagent. The stirring continued for another 15 min. The solvents were removed under reduced pressure. The crude product was purified by flash chromatography on a silica gel column (CH_3OH/CH_2Cl_2; 0% to 2% gradient) to give **4.4** (0.455 g, 95%) as white foam.

^{1}H-NMR (CD_3OD) δ (ppm): 0.08 and 0.21 [s, s, 6H, $(CH_3)_2Si$], 0.92 [m, 9H, $(CH_3)_3CSi$], 3.37 (s, 3H, CH_3SO_2), 3.50-3.54 (m, 2H, H-5'), 3.88 (s, 6H, 2×CH_3O), 4.20-4.25 (m, 1H, H-4'), 4.63-4.67 (m, 1H, H-3'), 5.35-5.42 (m, 1H, H-2'), 5.41 (d, 1H, H-5, J = 8.1 Hz), 6.19 (d, 1H, H-1', J = 2.7 Hz), 6.92-6.98 (m, 4H, Ar-H), 7.37-7.56 (m, 9H, Ar-H), 8.21 (d, 1H, H-6, J = 8.2 Hz)

^{13}C-NMR (CD_3OD) δ (ppm): 19.23 (Si-CH_3), 26.72 [SiC($\underline{C}H_3)_3$], 39.57 (mesyl CH_3), 56.28 (OCH_3), 62.01 (C-5'), 71.14 (C-3'), 82.52 (C-4'), 85.11 (C-2'), 89.06 (C-1'), 103.53 (C-5), 114.80, 132.04, 136.70, 145.95, 160.94 (Ar-C), 142.40 (C-6), 152.52 (C-2), 166.31 (C-4).

IR (KBr): 3450, 3068, 3030, 2950, 2839, 1702, 1610, 1519, 1460, 1380, 1256, 1188, 1110, 1054, 1010, 933, 904, 858, 782, 763, 715, 561 cm^{-1}.

UV (in acetonitrile), λ_{max} = 236.2, 268.6 nm.

FAB-HRMS: $C_{37}H_{47}N_2O_{10}SiS$ ($M+H^+$), 739.2719. Calculated m/z = 739.2721.

2,2'-anhydro-1-(2'-deoxy-3'-tert-butyldimethylsilyloxy-5'-O-dimethoxytrityl-β-D-ribofuranosyl)-uracil (4.5)

Compound **4.4** (0.407 g, 0.551 mmol) and tetrahexylammonium hydrogen sulfate (25 mg, 0.1 eq) were dissolved in toluene (11.0 mL, 0.05 M). A saturated Na_2CO_3 solution (11 mL) was added and the suspension was vigorously stirred at room temperature for 3 hrs. After the completion of the reaction, the suspension was extracted three times with ethyl acetate. The organic phase solutions were combined, washed with saturated NaCl solution, and dried over anhydrous $MgSO_4$. After filtration, the solvents were evaporated under reduced pressure. The crude product was purified on flash chromatography (CH_3OH/CH_2Cl_2; 0% to 3% gradient) to afford **4.5** (0.339 g, 96%) as white foam.

^{1}H-NMR ($CDCl_3$) δ (ppm): 0.08 and 0.14 [s, s, 6H, $(CH_3)_2Si$], 0.88 [m, 9H, $(CH_3)_3CSi$], 2.98-3.14 (m, 2H, H-5'), 3.82 (s, 6H, 2×CH_3O), 4.22-4.29 (m, 1H, H-4'), 4.48-4.52 (m, 1H, H-3'), 5.03-5.09 (m, 1H, H-2'), 5.98 (d, 1H, H-5, J = 7.6 Hz), 6.10 (d, 1H, H-1', J = 5.7 Hz), 6.72-6.84 (m, 4H, Ar-H), 7.18-7.36 (m, 10H, 9Ar-H, H-6).

^{13}C-NMR ($CDCl_3$) δ (ppm): 18.22 (Si-CH_3), 25.94 [SiC($\underline{C}H_3)_3$], 55.61 (OCH_3), 62.51 (C-5'), 77.17 (C-3'), 86.95 (Ar-C), 87.90 (C-4'), 89.22 (C-

2'), 90.04 (C-1'), 110.87 (C-5), 113.62, 127.38, 128.22, 130.13, 134.54, 135.56, 144.50, 158.98 (Ar-C), 137.40 (C-6), 159.51 (C-2), 171.68 (C-4).

IR (KBr): 3450, 3035, 2930, 2860, 1670, 1530, 1505, 1460, 1250, 1190, 1085, 1060, 820, 790, 770, 710, 610 cm^{-1}.

UV (in acetonitrile), λ_{max} = 233.2, 281.0 nm.

FAB-HRMS: $C_{36}H_{43}N_2O_7Si$ ($M+H^+$), 643.2838. Calculated m/z = 643.2839.

2,2'-anhydro-1-(2'-deoxy-5'-O-dimethoxytrityl-β-D-ribofuranosyl)-uracil (4.6)

Compound **4.5** (0.140 g, 0.218 mmol) was placed in 10-mL round flask and dissolved in 1.6 mL THF. A 1 M solution of tert-butylammonium fluoride in THF (0.43 mL) was then injected. The reaction mixture was stirred at room temperature for 1 hr. The solvent was evaporated under reduced pressure. The crude product was purified by flash chromatography (CH_3OH/CH_2Cl_2; 0% to 3% gradient) to afford **4.6** (0.109 g, 95%) as white foam.

^{1}H-NMR ($CDCl_3$) δ (ppm): 2.99-3.18 (m, 2H, H-5'), 3.69 (s, 6H, 2×CH_3O), 4.29-4.36 (m, 1H, H-4'), 4.42-4.46 (m, 1H, H-3'), 5.20-5.25 (m, 1H, H-2'), 5.92 (d, 1H, H-5, *J* = 7.5 Hz), 6.07 (d, 1H, H-1', *J* = 5.7 Hz), 6.98-6.79 (m, 4H, Ar-H), 7.11-7.30 (m, 10H, 9Ar-H, H-6).

^{13}C-NMR ($CDCl_3$) δ (ppm): 55.20 (OCH_3), 62.96 (C-5'), 75.60 (C-3'), 86.16 (Ar-C), 87.64 (C-4'), 89.42 (C-2'), 90.31 (C-1'), 109.64 (C-5), 113.18,

126.91, 127.91, 129.83, 135.40, 144.39, 158.45 (Ar-C), 135.93 (C-6), 159.75 (C-2), 172.69 (C-4).

IR (KBr): 3400, 3030, 2920, 2850, 1670, 1520, 1510, 1490, 1460, 1260, 1190, 1095, 1055, 820, 770, 705, 580 cm^{-1}.

UV (in acetonitrile), λ_{max} = 233.8, 281.2 nm.

FAB-HRMS: $C_{30}H_{29}N_2O_7$ ($M+H^+$), 529.1976. Calculated m/z = 529.1974.

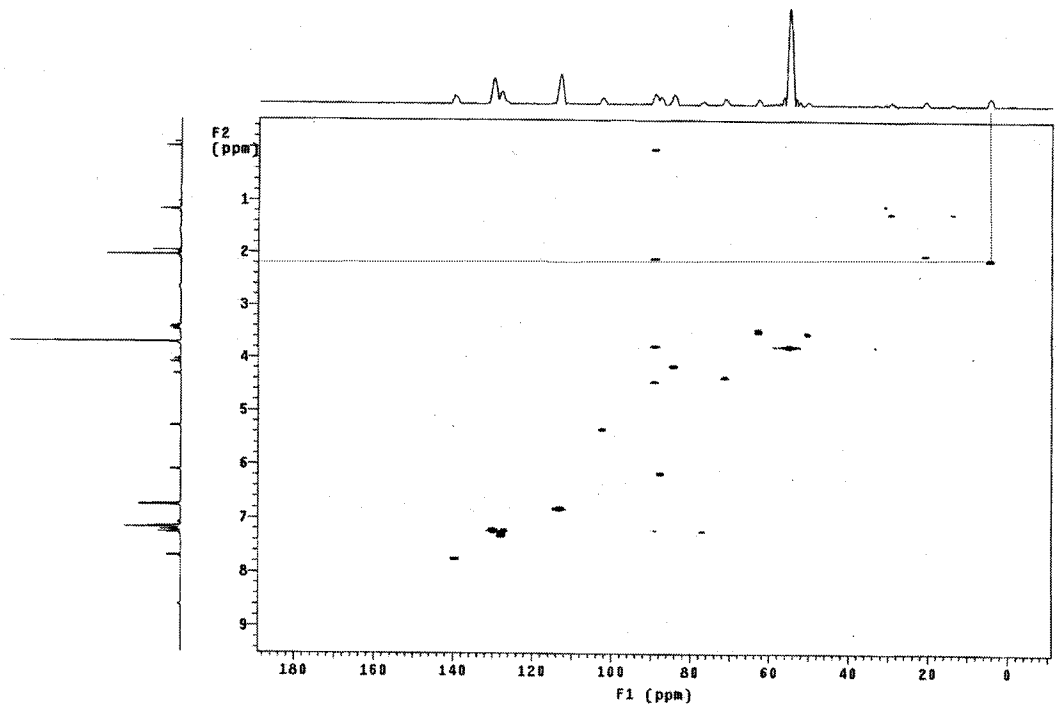
2'-methylseleno-5'-dimethoxytrityl-uridine (4.7)

$NaBH_4$ (0.170 g, 4.5 mmol) was placed in 25-mL round flask, dried on high vacuum for 15 min to deplete oxygen, and suspended in 9 mL dry THF. Dimethyl diselenide (150 μL, 1.5 mmol) was slowly injected to this suspension, followed by dropwise addition of anhydrous ethanol until gas bubbles started to occur. After the reaction mixture turned from yellow to colorless (usually in 10 min), the solution was injected to **4.6** (0.160 g, 0.30 mmol) dissolved in THF (6 mL). The reaction was stirred under argon at room temperature for 3 hrs. The solvent was evaporated under reduced pressure. The residue was treated with 5 mL water, and then extracted by ethyl acetate (3×10 mL). The organic phases were combined, washed with saturated NaCl solution, and dried over anhydrous $MgSO_4$. The mixture was filtrated and the ethyl acetate was evaporated under reduced pressure. The

crude product was purified by flash chromatography (CH_3OH/CH_2Cl_2; 0% to 3% gradient) to afford **4.7** (0.179 g, 96%) as white foam.

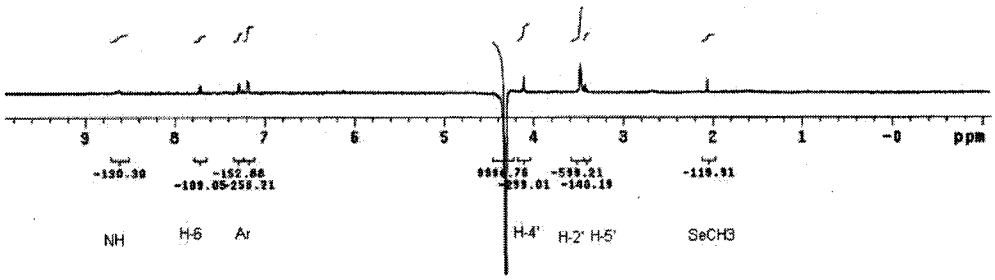
^{1}H-NMR ($CDCl_3$) δ (ppm): 2.08 (s, 3H, CH_3Se), 3.42-3.47 (m, 2H, H-5'), 3.49-3.54 (m, 1H, H-2'), 3.78 (s, 6H, 2×CH_3O), 4.14-4.18 (m, 1H, H-4'), 4.34-4.39 (m, 1H, H-3'), 5.36 (d, 1H, H-5, J = 8.0 Hz), 6.19 (d, 1H, H-1', J = 3.3 Hz), 6.86-6.92 (m, 4H, Ar-H), 7.19-7.38 (m, 9H, Ar-H), 7.77 (d, 1H, H-6, J = 7.8 Hz), 8.48 (br, 1H, NH). Note the $SeCH_3$ signal at 2.08 ppm, and the significant shift of H-2' signal to higher field (smaller ppm value). The results are caused by the electron-shielding effect of selenium on protons two covalent bonds away.

^{13}C-NMR ($CDCl_3$) δ (ppm): 5.18 ($SeCH_3$), 51.12 (C-2'), 55.63 (OCH_3), 62.80 (C-5'), 71.90 (C-3'), 84.84 (C-4'), 87.66 (Ar-C), 87.98 (C-1'), 103.05 (C-5), 113.69, 128.43, 130.42, 135.50, 144.36, 158.55 (Ar-C), 139.45 (C-6), 150.55 (C-2), 163.03 (C-4). Note the $SeCH_3$ signal at unusual 5.18 ppm, and the appearance of C-2' signal at higher field. The results are caused by the electron-shielding effect of selenium on its bonded carbon nuclei.



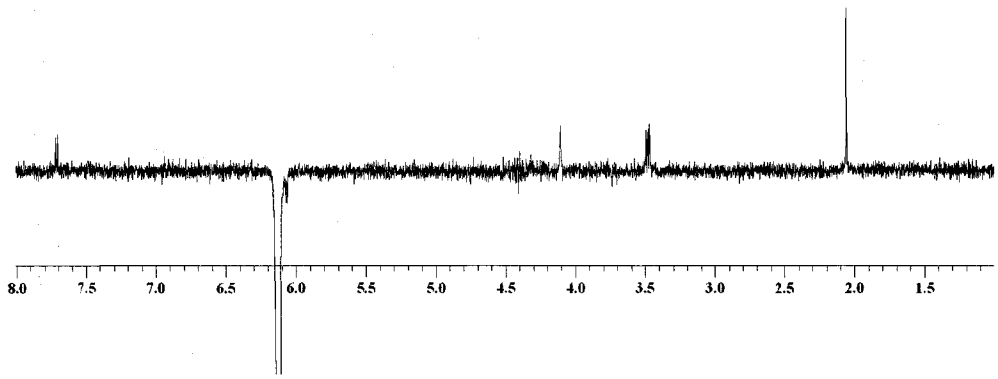
The spectrum above shows the gs-HMQC NMR result of **4.7**. The indicated cross peak (F1 = 5.18, F2 = 2.08 ppm) shows the correlation between ^{1}H and ^{13}C nuclei in –Se$\underline{CH_3}$ group.

GOESY experiments:

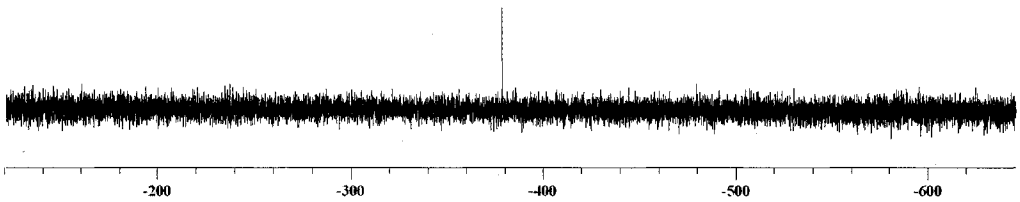


Irradiation at 4.36 ppm (H-3', β-H) gives NOE enhancement at 2.08 (CH_3Se, 1.2%), 3.42-3.47 (H-5', 1.2%), 3.49-3.54 (H-2', 6.2%), 4.14-4.18 (H-4', 2.9%), 7.19 (Ar-H, 2.5%), 7.28 (Ar-H, 1.5%), 7.77 (H-6, 1.1%). Stronger NOE enhancement on H-2' signal and weaker enhancement on H-4' (α-H)

suggest H-2' is located at cis-position from H-3'. In other words, H-2' is also a β-hydrogen, while $SeCH_3$ must be located at α-position.



Irradiation at 6.19 ppm (H-1', α-H) gives NOE enhancement at 2.08 (CH_3Se, 1.7%), 3.49-3.54 (H-2', 1.5%), 4.14-4.18 (H-4', 1.0%), 7.77 (H-6, 0.9%). Compared with the 6.2% enhancement on H-2' signal by H-3' irradiation, H-1' irradiation only generated 1.5% enhancement. This result confirms H-2' is at trans-position from H-1', i.e. H-2' is a β-hydrogen.



Se-NMR ($CDCl_3$) δ (ppm): −244.8 ppm (s, 1Se), corrected for reference dibenzyl diselenide at 133.25 ppm. The spectrum shows δ = −378.0 ppm against reference dibenzyl diselenide at zero ppm.

IR (KBr): 3450, 3080, 3030, 2940, 1705, 1610, 1520, 1460, 1390, 1245, 1190, 1090, 1045, 850, 782, 710, 590 cm^{-1}.

UV (in acetonitrile), λ_{max} = 236.2, 273.4 nm.

ESI-MS (positive ion mode): [M (^{80}Se) + NH_4^+] = 641.2; calculated m/z = 641.0. [M (^{78}Se) + NH_4^+] = 639.1; calculated m/z = 639.0.

FAB-HRMS: $C_{31}H_{32}N_2O_7Se$ ($M+H^+$), 624.1376. Calculated m/z = 624.1374.

2'-methylseleno-3'-(2-cyanoethyl-N,N-diisopropylphosphoramidite)-5'-dimethoxytrityl-uridine (4.8)

To a solution of 4.7 (0.320 g, 0.51 mmol) in anhydrous CH_2Cl_2 (5.1 mL) under dry argon, N, N-diisopropylethylamine (0.25 mL, 1.5 mmol) and 2-cyanoethyl N, N-diisopropylchlorophosphoramidite (0.23 mL, 1.0 mmol) were slowly added. The mixture was stirred at room temperature for 1 hr. Reaction completion was indicated by TLC [in CH_3OH/CH_2Cl_2 (1:19)]. Two diastereotopic products were produced at about 1:1 mole ratio. The reaction was quenched by saturated $NaHCO_3$ solution (2 mL), followed by CH_2Cl_2 extraction (3×5 mL). The combined organic layer was washed with saturated brine (10 mL) and dried over anhydrous $MgSO_4$, followed by filtration. The solvent was evaporated under reduced pressure. The crude product was re-dissolved in 2 mL CH_2Cl_2. This solution was added dropwise to petroleum ether (100 mL) under vigorous stirring. A white precipitate was gradually formed on the beaker wall, while the ether phase became clear after 15 min. After pouring out the ether, the product was dissolved in CH_2Cl_2 (2 mL), and

the precipitation process was repeated. The final product was dried on high vacuum overnight to yield white foam (0.386 g, 92%).

UV (in acetonitrile), λ_{max} = 236.6, 267.8 nm.

IR (KBr): 3450, 3070, 3030, 2930, 2850, 1715, 1610, 1505, 1480, 1390, 1250, 1180, 1090, 1045, 785, 715, 580 cm^{-1}.

FAB-HRMS: $C_{40}H_{50}N_4O_8PSe$ ($M+H^+$), 825.2538. Calculated m/z = 825.2532.

Chapter 5.

2'-methanesufonyl-5'-dimethoxytrityl-thymidine (5.3)

Compound **5.2** (1.44g, 2.65 mmol) was dried in flask on high vacuum for 30 min. THF (26.5 mL) and TEA (1.1 mL, 7.95 mmol) were injected to the flask. The mixture was stirred under dry nitrogen at 0 °C for 5min, then methanesulfonyl chloride (247 μL, 3.18 mmol) was slowly injected. The mixture was stirred at room temperature for 30min. The reaction was quenched by adding 1 mL methanol. The solvent was evaporated under reduced pressure. The product was re-dissolved in ethyl acetate and the salts were filtrated away. The solvent of eluent was evaporated under reduced pressure. The collected residue was purified by flash chromatography

(CH_3OH/CH_2Cl_2; 0% to 5% gradient). The product was dried under high vacuum to afford white foam (1.63 g, 99% yield).

^1H NMR ($CDCl_3$) δ(ppm): 1.45 (3H, s, CH_3 on thymine base), 2.39-2.80 (2H, m, H-2', H-2"), 3.03 (3H, s, SO_2CH_3), 3.46-3.54 (2H, m, H-5'), 3.80 (6H, s, OCH_3), 4.08-4.16 (1H, q, H-4'), 5.38 (1H, d, H-1', J = 5.9 Hz), 6.41-6.47 (1H, m, H-3'), 6.83-6.86, 7.24-7.38 (13H, m, Ar-H), 7.55 (1H, s, H-6), 8.95 (1H, br, NH). The assignment was assisted by additional ^1H-^1H COSY experiment (spectrum not shown).

^{13}C-NMR ($CDCl_3$) δ (ppm): 11.89 (CH_3 on base), 38.58 (SO_2CH_3), 38.77 (C-2'), 55.40 (2×OCH_3), 63.07 (C-5'), 78.98 (C-Ph_3), 83.92 (C-1'), 84.41 (C-3'), 87.52 (C-4'), 111.93 (C-5), 113.52, 127.46, 128.18, 130.15 (Ar), 135.11 (C-6), 144.14 (Ar), 150.47 (C-2), 158.98 (Ar), 163.68 (C-4)

$C_{32}H_{34}N_2O_9S$, exact mass = 622.1985. Expected ($M + Na^+$) = 645.1877. ESI-MS: observed ($M + Na^+$) = 645.

5'-dimethoxytrityl-2,3'-anhydro-thymidine (5.4)

Compound **5.3** (1.63g, 2.62 mmol) was dissolved in 262 mL of ethanol containing 13.75 mL 0.2N NaOH (2.75 mmol, 1.05 eq) and refluxed for 2.5 hrs. The solvent was evaporated and the residue was extracted with EtOAc and distilled water three times. The organic phase was collected. After evaporation of EtOAc, the residue was purified by flash chromatography

(CH_3OH/CH_2Cl_2; 0% to 10% gradient). The product was dried on high vacuum to afford a fine white powder (1.31 g, 95% yield).

^1H NMR ($CDCl_3$), δ(ppm): 1.90 (3H, s, CH_3 on thymine), 2.34-2.38, 2.63-2.68 (2H, m, H-2', H-2"), 3.30-3.33 (2H, m, H5'), 3.77 (6H, s, OCH_3), 4.24-4.25 (1H, m, H-4'), 5.11 (1H, s, H-3'), 5.47 (1H, d, H-1', J = 3.75Hz), 6.77-6.82, 7.17-7.43 (13H, m, Ar-H), 6.93 (1H, s, H-6). The assignment was assisted by ^1H-^1H COSY experiment.

^{13}C NMR ($CDCl_3$), δ(ppm): 13.42 (CH_3 on thymine), 33.62 (C-2'), 55.30 (2×OCH_3), 62.21 (C-5'), 76.91 (C-3'), 84.57 (C-1'), 86.76 (C-Ph_3), 87.57 (C-4'), 113.25 (Ar), 118.23 (C-5), 126.92, 127.92, 130.05, 135.36 (Ar), 135.54 (C-6), 144.53 (Ar), 153.41 (C-2), 158.60 (Ar), 171.80 (C-4).

$C_{31}H_{30}N_2O_6$, exact mass = 526.2104. Expected ($M + H^+$) = 527.2177, ($M + Na^+$) = 549.1996. ESI-MS: observed ($M + H^+$) = 527, ($M + Na^+$) = 549. High resolution MALDI-FTMS: observed ($M + Na^+$) = 549.1969.

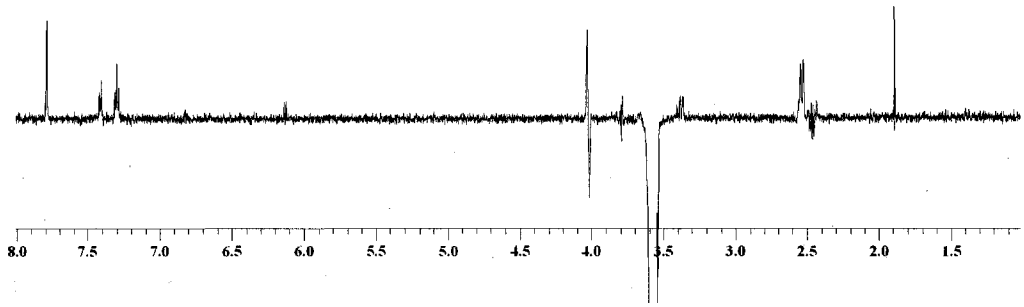
3'-methylseleno-5'-dimethoxytrityl-thymidine (5.5)

Compound **5.4** (53 mg, 0.1 mmol) was dried on high vacuum together with $NaBH_4$ powder (45.4 mg, 1.2mmol) for 10 min. THF (10.6 mL) was injected to the reaction flask and the suspension was stirred at room temperature under dry nitrogen. A brown liquid of dimethyl diselenide (39 μL, 0.4 mmol) was carefully added to the flask. *Caution: dimethyl diselenide is toxic and its odor spreads easily. Leave used syringes and contaminated gloves*

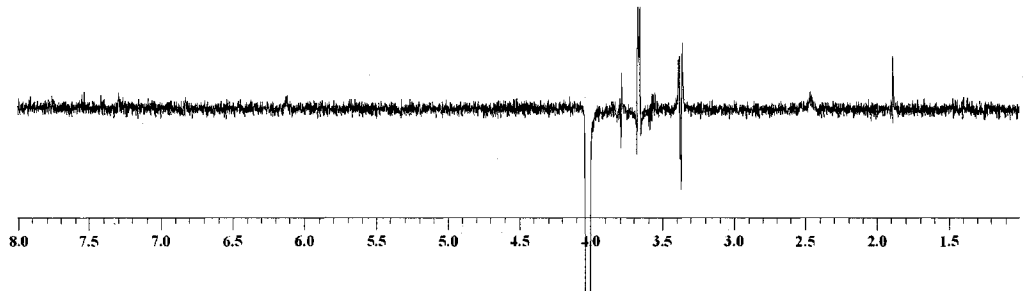
under the fume hood. Afterward, inject small volumes of ethanol drop wise until gas effervescence was generated. Wait until the solution turns colorless, otherwise, repeat ethanol injection. The mixture was stirred at room temperature for 72 hrs and then the solvent was evaporated under reduced pressure. Dissolve the product in methylene chloride and wash the organic layer with saturated NaCl solution 3 times. The collected organic layer was dried with anhydrous $MgSO_4$ and filtrated. The solvent was evaporated under reduced pressure and the residue was placed on silica gel chromatography. The elution solvent was 0-8% gradient of methanol in methylene chloride with 2 drops of additional TEA. The product band was collected and the solvent was evaporated. The residue was dried on high vacuum to afford white foam (58.8 mg, 94%).

^{1}H NMR ($CDCl_3$), δ(ppm): 1.40 (3H, s, CH_3 on thymine), 1.89 (3H, s, $SeCH_3$), 2.46-2.55 (2H, m, H-2', H-2"), 3.35-3.41, 3.66-3.70 (2H, m, H-5'), 3.56-3.60 (1H, m, H-3'), 3.79 (6H, s, OCH_3), 4.01-4.05 (1H, m, H-4'), 6.15 (1H, q, H-1', J = 3.34 Hz, 6.52 Hz), 6.82-6.86, 7.24-7.44 (13H, m, Ar-H), 7.80 (1H, s, H-6), 9.33 (1H, s, NH). The significant shift of H-3' signal to higher field is caused by the bonding of -$SeCH_3$ to C-3'.

GOESY experiments:



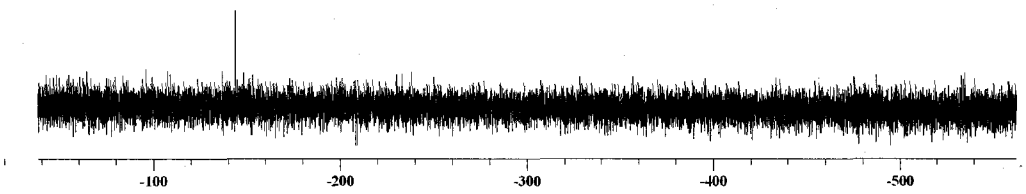
Irradiation of H-3' gives NOE enhancement at 1.89 ppm ($SeC\underline{H}_3$, 2.2%), 2.53 ppm (H-2', 1.2%), 7.29 and 7.40 ppm (Ar-H, increase not recorded), 7.79 ppm (H-6, 2.1%). NOE observed between H-3' and H-6 strongly suggests that H-3' is on β-position, which is close enough to H-6 when the glycosyl bond angle (χ) is in favorable *anti*-conformation.



Irradiation of H-4' gives significant NOE enhancement at 1.89 ppm ($SeC\underline{H}_3$, increase not recorded), 3.66 ppm (1H, H-5' or H-5", 2.9%). Absence of NOE between H-3' and H-4' also confirmed the previous result.

^{13}C NMR ($CDCl_3$), δ(ppm): 4.76 ppm ($Se\underline{C}H_3$), 12.96 ppm ($\underline{C}H_3$ on thymine), 32.30 ppm (C-3'), 40.73 ppm (C-2'), 54.19 ($2\times OCH_3$), 62.20 ppm (C-5'), 83.95 ppm ($C\text{-}Ph_3$), 85.66 ppm (C-1'), 86.74 ppm (C-4'), 112.09 (C-5), 114.61, 126.73, 128.92, 131.43 (Ar), 135.38 (C-6), 144.31 (Ar), 150.42 (C-2), 158.80 (Ar), 164.70 (C-4). Hydrogen bonded ^{13}C-assignment was confirmed by additional gs-HMQC experiment (spectrum not shown). The

significant shift of C-3' signal to the higher field is caused by the selenium that is directly bonded to.



^{77}Se NMR ($CDCl_3$), δ(ppm): −144.37 (s, $\underline{Se}CH_3$). Chemical shift was referenced against dibenzyl diselenide at 133.25 ppm.

$C_{32}H_{34}N_2O_6Se$, exact mass = 622.1582. Expected $(M - H)^-$ = 621.1509, $(M + Na^+)$ = 645.1474. ESI-MS: observed $(M - H)^-$ = 621, $(M + Na^+)$ = 645. High resolution MALDI-FTMS: observed $(M + Na^+)$ = 645.1460.

3'-deoxy-methylseleno-thymidine (5.6)

Compound **5.5** was treated with 80% acetic acid solution and stirred at room temperature for 30min. The trityl-deprotection reaction was near completion by TLC monitoring. After neutralization with NaOH solution, the solvents were evaporated under reduced pressure. The crude product was purified on preparative TLC plates with 8% methanol/methylene chloride. The collected product was dried under high vacuum overnight to give a colorless oil compound.

^{1}H NMR ($CDCl_3$), δ(ppm): 1.90 (3H, s, CH_3 on thymine), 2.07 (3H, s, $SeCH_3$), 2.46-2.56 (2H, m, H-2', H-2"), 3.44-3.55 (1H, m, H-3'), 3.86-4.10 (2H, m, H-5'), 3.96-4.00 (1H, m, H-4'), 6.10 (1H, q, H-1', J = 3.78 Hz, 6.89 Hz), 7.59 (1H, s, H-6).

^{13}C NMR ($CDCl_3$), δ(ppm): 3.52 ($SeCH_3$), 12.69 (CH_3 on thymine), 32.48 (C-3'), 40.14 (C-2'), 61.18 (C-5'), 85.66 (C-1'), 86.71 (C-4'), 110.82 (C-5), 136.65 (C-6), 150.48 (C-2), 164.04 (C-4).

Exact mass for $C_{11}H_{16}N_2O_4Se$ = 320.02775.
LC-ESI-MS positive ion experiment: Expected ion peaks: $[M+H]^+$ =321.0 and $[2M+H]^+$ = 641.0. Observed ion peaks: $[M+H]^+$ =321.0 and $[2M+H]^+$ = 640.9 (data from Hunter College).
Expected $(M - H)^-$ = 319.0202, $(M + Na^+)$ = 343.0167. ESI-MS: observed $(M - H)^-$ = 319, $(M + Na^+)$ = 343. High resolution MALDI-FTMS: observed $(M + Na^+)$ = 343.0167 (data from TSRI).

5'-dimethoxytrityl-3'-deoxy-thymidine 3'-3'' diselenide dimer (5.7)

Selenium powder (120mg, 1.52mmol), and $NaBH_4$ powder (29mg, 0.76mmol) were mixed in a flask and dried on high vacuum for 10min. Add 15mL of THF to the flask and stir vigorously at room temperature under dry nitrogen. Add 15mL of anhydrous ethanol slowly to the mixture and stir for 30min. A dark red solution was formed, with some extra selenium solid remaining. Open the flask and add compound **5.4** (80mg, 0.152mmol) immediately. Quickly cover the flask with a glass condenser sealed under

nitrogen balloon. Heat the flask under reflux overnight. After the flask is cooled to room temperature, bubble air into the liquid for 2 hrs. The solvent was evaporated under reduced pressure. The product was re-dissolved in ethyl acetate and centrifuged at low speed (~1000 G) to remove solid impurities. The solvent in supernatant was evaporated. The residue was placed onto two 1mm thick 20×20cm TLC plates for purification. Solvent was 10% methanol/methylene chloride. The silica powder containing a major band (this product has a bright yellow color in ambient light, while the other bands do not show any color on TLC) was collected. The compound was dissolved in ethyl acetate and went through filtration. After solvent evaporation, the product dried on high vacuum, giving a yellow oily substance (74.6mg, 81%).

^{1}H NMR ($CDCl_3$), δ(ppm): 1.47 (3H, s, CH_3 on thymine), 2.50-2.54 (2H, m, H-2'), 3.36-3.60 (2H, m, H-5'), 3.74 (1H, H-3'), 3.77 (6H, s, OCH_3), 3.89-3.93 (1H, q, H-4', J = 7.3 Hz), 6.16 (1H, t, H-1', J = 6.05 Hz), 6.81-6.84, 7.22-7.68 (13H, m, Ar-H), 7.72 (1H, s, H-6), 9.27 (1H, s, NH). Proton integration numbers are normalized to half of the molecule.

^{13}C NMR ($CDCl_3$), δ(ppm): 12.12 (CH_3 on thymine), 35.71 (C-3'), 40.65 (C-2'), 55.38 (2×OCH_3), 62.18 (C-5'), 84.95 (C-Ph_3), 86.48 (C-1'), 87.00 (C-4'), 111.05 (C-5), 113.44, 127.90, 128.64, 130.21 (Ar), 135.44 (C-6), 144.35 (Ar), 150.56 (C-2), 158.75 (Ar), 164.01 (C-4).

^{77}Se NMR ($CDCl_3$) was carried out, but no obvious peak was observed in the scanning frequency range of experimental trials.

$C_{62}H_{62}N_4O_{12}Se_2$, exact mass = 1214.2694. Expected ($M + Na^+$) = 1237.2587. ESI-MS: observed ($M + Na^+$) = 1237. High resolution MALDI-FTMS: observed ($M + Na^+$) = 1237.2625.

3'-deoxy-thymidine 3'-3'' diselenide dimer (5.8)

Detritylation procedure may refer to this reaction: **5.5** → **5.6**.

Exact mass for $C_{20}H_{26}N_4O_8Se_2$ = 610.0081.
LC-ESI-MS positive ion experiment: Expect ion peak ($M+H^+$) = 611.0; measured ion peak ($M+H^+$) =611.0 (data from Hunter College).
Expected $(M + Cl)^-$ = 644.9775, ($M + Na^+$) = 632.9973. ESI-MS: observed $(M + Cl)^-$ = 645, ($M + Na^+$) = 633. High resolution MALDI-FTMS: observed ($M + Na^+$) = 632.9959 (data from TSRI).

^1H NMR ($CDCl_3$), δ(ppm): 1.90 (3H, s, CH_3 on thymine), 2.45-2.56 (2H, m, H-2', H-2"), 3.84-4.03 (2H, m, H-5'), 3.68 (1H, H-3'), 3.84-3.88 (1H, q, H-4'), 6.11 (1H, t, H-1'), 7.54 (1H, s, H-6). Proton integration numbers are normalized to half of the molecule.

^{13}C NMR ($CDCl_3$), δ(ppm): 12.01 (CH_3 on thymine), 35.79 (C-3'), 40.56 (C-2'), 61.78 (C-5'), 86.64 (C-1'), 86.74 (C-4'), 111.23 (C-5), 136.27 (C-6), 150.52 (C-2), 164.04 (C-4).

Bibliography

(1) Gibbs, N. The Secret of Life. In *Time*, 2003; pp 42-45.

(2) Watson, J. D.; Crick, F. H. C. Molecular structure of nucleic acids: A structure for deoxyribose nucleic acid. *Nature* **1953**, *171*, 737-738.

(3) Wilkins, M. H. F.; Stokes, A. R.; Wilson, H. R. Molecular structure of deoxypentose nucleic acids. *Nature* **1953**, *171*, 738-740.

(4) Franklin, R. E.; Gosling, R. G. Molecular configuration in sodium thymonucleate. *Nature* **1953**, *171*, 740-741.

(5) Pauling, L.; Corey, R. B. *Proc. Natl. Acad. Sci. USA* **1953**, *39*, 84.

(6) Pauling, L.; Corey, R. B. *Nature* **1953**, *171*, 346.

(7) Neidle, S. *Oxford Handbook of Nucleic Acid Structure*; Oxford University Press, 1999.

(8) Berman, H. M. Crystal studies of B-DNA: The answers and the questions. *Biopoly.* **1997**, *44*, 23-44.

(9) Rosenberg, J. M.; Seeman, N. C.; Kim, J. J. P.; Suddath, F. L.; Nicholas, H. B. et al. Double helix at atomic resolution. *Nature* **1973**, *243*, 150-154.

(10) Wang, A. H.-J.; Quigley, G. J.; Kolpak, F. J.; Crawford, J. L.; van Boom, J. H. et al. Molecular structure of a left-handed double helical DNA fragment at atomic resolution. *Nature* **1979**, *282*, 680-686.

(11) Wang, A. H.-J.; Quigley, G. J.; Kolpak, F. J.; van der Marcel, G.; van Boom, J. H. et al. Left-handed double helical DNA: variations in the backbone conformation. *Science* **1981**, *211*, 171-176.

(12) Wing, R. M.; Drew, H. R.; Takano, T.; Broka, C.; Tanaka, S. et al. Crystal structure analysis of a complete turn of B-DNA. *Nature* **1980**, *287*, 755-758.

(13) Drew, H. R.; Takano, T.; Tanaka, S.; Itakura, K.; Dickerson, R. E. Crystal structure analysis of a complete turn of B-DNA. *Nature* **1980**, *286*, 567-573.

(14) Shakked, D.; Rabinovich, D.; Cruse, W. B. T.; Egert, E.; Kennard, O. et al. *Proc. R. Soc. (London) B* **1981**, *213*, 479.

(15) Wang, A. H.-J.; Fujii, S.; van Boom, J. H.; Rich, A. Molecular structure of the octamer d(G-G-C-C-G-G-C-C): modified A-DNA. *Proc. Natl. Acad. Sci. USA* **1982**, *79*, 3968-3972.

(16) Wahl, M. C.; Sundaralingam, M. Crystal structures of A-DNA duplexes. *Biopoly.* **1997**, *44*, 45-63.

(17) Felsenfeld, G.; Davies, D. R.; Rich, A. Formation of a three-stranded polynucleotide molecule. *J. Am. Chem. Soc.* **1957**, *79*, 2023-2024.

(18) Felsenfeld, G.; Rich, A. *Biochim. Biophys. Acta* **1957**, *26*, 457.

(19) Hoogsteen, K. *Acta Cryst.* **1959**, *12*, 822.

(20) Nielsen, P. E.; Egholm, M.; Buchardt, O. Sequence-selective recognition of DNA by strand displacement with a thymine-substituted polyamide. *Science* **1991**, *254*, 1497-1500.

(21) Betts, L.; Josey, J. A.; Veal, J. M.; Jordan, S. R. A nucleic acid triple helix formed by a peptide nucleic acid-DNA complex. *Science* **1995**, *270*, 1838.

(22) Gellert, M.; Lipsett, M. N.; Davies, D. R. *Proc. Natl. Acad. Sci. USA* **1962**, *48*, 2013-2018.

(23) Blackburn, E. H. Telomeres: no end in sight. *Cell* **1994**, *77*, 621-623.

(24) Keniry, M. A. Quadruplex structures in nucleic acids. *Biopoly.* **2001**, *56*, 123-146.

(25) Holbrook, S. R.; Kim, S.-H. RNA Crystallography. *Biopoly.* **1997**, *44*, 3-21.

(26) Stout, C. D.; Mizuno, H.; Rubin, J.; Brennan, T.; Rao, S. T. et al. Atomic coordinates and molecular conformation of yeast phenylalanyl tRNA. An independent investigation. *Nucleic Acids Res.* **1976**, *3*, 1111-1123.

(27) Hingerty, B.; Brown, R. S.; Jack, A. Further refinement of the structure of yeast tRNA-Phe. *J. Mol. Biol.* **1978**, *124*, 523-534.

(28) Sussman, J. L.; Holbrook, S. R.; Warrant, R. W.; Church, G. M.; Kim, S.-H. Crystal structure of yeast phenylalanine transfer RNA. I. Crystallographic refinement. *J. Mol. Biol.* **1978**, *123*, 607-630.

(29) Pley, H. W.; Flaherty, K. M.; McKay, D. B. Three-dimensional structure of a hammerhead ribozyme. *Nature* **1994**, *372*, 68-74.

(30) Scott, W. G.; Finch, J. T.; Klug, A. The crystal structure of an all-RNA hammerhead ribozyme: a proposed mechanism for RNA catalytic cleavage. *Cell* **1995**, *81*, 991-1002.

(31) Cate, J. H.; Gooding, A. R.; Podell, E.; Zhou, K.; Golden, B. L. et al. Crystal Structure of a Group I Ribozyme Domain: Principles of RNA Packing. *Science* **1996**, *273*, 1678-1685.

(32) Nissen, P.; Hansen, J.; Ban, N.; Moore, P. B.; Steitz, T. A. The Structural Basis of Ribosome Activity in Peptide Bond Synthesis. *Science* **2000**, *289*, 920.

(33) Ban, N.; Nissen, P.; Hansen, J.; Moore, P. B.; Steitz, T. A. The Complete Atomic Structure of the Large Ribosomal Subunit at 2.4 Å Resolution. *Science* **2000**, *289*, 905-920.

(34) Yusupov, M. M.; Yusupova, G. Z.; Baucom, A.; Lieberman, K.; Earnest, T. N. et al. Crystal Structure of the Ribosome at 5.5 Å Resolution. *Science* **2001**, *292*, 883-896.

(35) Haugland, R. P. *Handbook of Fluorescenct Probes and Research Products*; 9th ed.; Molecular Probes, 2002.

(36) Berman, H. M. The past and future of structure databases. *Curr. Opin. Biotechnol.* **1999**, *10*, 76-80.

(37) Neidle, S. Crystallographic insights into DNA minor groove recognition by drugs. *Biopoly.* **1997**, *44*, 105-121.

(38) Joyce, G. F. The antiquity of RNA-based evolution. *Nature* **2002**, *418*, 214-221.

(39) Gesteland, R. F.; Cech, T. R.; Atkins, J. F. *The RNA World*; 2nd ed.; Cold Spring Harbor Laboratory Press, 1999.

(40) Doudna, J. A.; Cech, T. R. The chemical repertoire of natural ribozymes. *Nature* **2002**, *418*, 222-228.

(41) Kruger, K. Autoexcision and autocyclization of the ribosomal RNA intervening sequence of Tetrahymena. *Cell* **1982**, *31*, 147-157.

(42) Guerrier-Tkada, C.; Gardiner, K.; Marsh, T.; Pace, N.; Altman, S. The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. *Cell* **1983**, *35*, 849-857.

(43) Hannon, G. J. RNA interference. *Nature* **2002**, *418*, 244-250.

(44) Jones, S.; Daley, D.; Luscombe, N. M.; Berman, H. M.; Thornton, J. M. Protein-RNA interactions: a structural analysis. *Nucleic Acids Res.* **2001**, *29*, 943-954.

(45) Wuthrich, K. *NMR of Proteins and Nucleic Acids*; John Wiley & Sons: New York, 1986.

(46) Rhodes, G. *Crystallography Made Crystal Clear*; 2nd ed.; Academic Press, 2000.

(47) Hendrickson, W. A.; Ogata, C. M. Phase determination from multiwavelength anomalous diffraction measurements. *Methods Enzymol.* **1997**, *276*, 494-523.

(48) Simons, C. *Nucleoside Mimetics: Their Chemistry and Biological Properties*; Gordon and Breach Science Publishers, 2001.

(49) Nelson, D. L.; Cox, M. M. *Lehninger Principles of Biochemistry*; 3rd ed.; Worth Publishers, 2000.

(50) Marquez, V. E.; Siddiqui, M. A.; Ezzitouni, A.; Russ, P.; JianyingWang et al. Nucleosides with a twist. Can fixed forms of sugar ring pucker influence biological activity in nucleosides and oligonucleotides? *J. Med. Chem.* **1996**, *39*, 3739-3747.

(51) Altona, C.; Sundaralingam, M. Conformational analysis of the sugar ring in nucleosides and nucleotides. Improved method for the interpretation of proton magnetic resonance coupling constants. *J. Am. Chem. Soc.* **1973**, *95*, 2333-2344.

(52) Altona, C.; Sundaralingam, M. Conformational analysis of the sugar ring in nucleosides and nucleotides. New description using the concept of pseudorotation. *J. Am. Chem. Soc.* **1972**, *94*, 8205-8212.

(53) Pathak, T. Azidonucleosides: Synthesis, Reactions, and Biological Properties. *Chem. Rev.* **1992**, *102*, 1623-1667.

(54) Richman, D. D. HIV chemotherapy. *Nature* **2001**, *410*, 995-1001.

(55) Huryn, D. M.; Okabe, M. AIDS-Driven Nucleoside Chemistry. *Chem. Rev.* **1992**, *92*, 1745-1768.

(56) Herdewijn, P. Structural requirements for antiviral activity in nucleosides. *Drug Discovery Today* **1997**, *2*, 235-242.

(57) Montgomery, J. A. Studies on the biologic activity of purine and pyrimidine analogs. *Med. Res. Rev.* **1982**, *2*, 271-308.

(58) Crooke, S. T. *Antisense Drug Technology: Principles, Strategies, and Applications*; Marcel Dekker, Inc: New York, Basel, 2001.

(59) Zamecnik, P. C.; Stephenson, M. L. *Proc. Natl. Acad. Sci. USA* **1978**, *75*, 280-284.

(60) Manoharan, M. 2'-Carbohydrate modification in antisense oligonucleotide therapy: importance of conformation, configuration and conjugation. *Biochim. Biophys. Acta* **1999**, *1489*, 117-130.

(61) Koshkin, A. A.; Rajwanshi, V. K.; Wengel, J. *Tetrahedron Lett.* **1998**, *39*, 4381.

(62) Koshkin, A. A.; Singh, S. K.; Nielsen, P.; Rajwanshi, V. K.; Kumar, R. et al. *Tetrahedron* **1998**, *54*, 3607.

(63) Obika, S.; Nanbu, D.; Hari, Y.; Andoh, J. I.; Morio, K. I. et al. Stability and structural features of the duplexes containing the nucleoside analogues with a fixed N-type conformation: 2'-O, 4'-C-methyleneribonucleosides. *Tetrahedron Lett.* **1998**, *39*, 5401-5404.

(64) Braasch, D. A.; Corey, D. R. Locked nucleic acid (LNA): fine tuning the recognition of DNA and RNA. *Chem. Biol.* **2001**, *8*, 1-7.

(65) Larsen, H. J.; Bentin, T.; Nielsen, P. E. Antisense properties of peptide nucleic acid. *Biochim. Biophys. Acta* **1999**, *1489*, 159-166.

(66) Nielsen, P. E. Peptide nucleic acids as therapeutic agents. *Curr. Opin. Struct. Biol.* **1999**, *9*, 353-357.

(67) Nielsen, P. E. Applications of peptide nucleic acids. *Curr. Opin. Biotechnol.* **1999**, *10*, 71-75.

(68) Nielsen, P. E. Peptide nucleic acid: a versatile tool in genetic diagnostics and molecular biology. *Curr. Opin. Biotechnol.* **2001**, *12*, 16-20.

(69) Cook, P. D. *Antisense Medicinal Chemistry*; Springer-Verlag: Berlin, 1998.

(70) Cook, P. D. Making drugs out of oligonucleotides: a brief review and perspective. *Nucleosides, Nucleotides & Nucleic Acids* **1999**, *18*, 1141-1162.

(71) Egli, M. Towards the structure-based design of oligonucleotide therapeutics. *Advan. Enzyme Regul.* **1998**, *38*, 181-203.

(72) Hui, Y.; Chen, Y. *Chemistry and Life Sciences*; Chemical Industry Press: Beijing, China, 1991.

(73) Earnshaw, D. J.; Gait, M. J. Modified oligoribonucleotides as site-specific probes of RNA structure and function. *Biopoly.* **1998**, *48*, 39-55.

(74) Verma, S.; Eckstein, F. Modified oligonucleotides: synthesis and strategy for users. *Annu. Rev. Biochem.* **1998**, *67*, 99-134.

(75) Ferrero, M.; Gotor, V. Biocatalytic selective modifications of conventional nucleosides, carbocyclic nucleosides, and C-nucleosides. *Chem. Rev.* **2000**, *100*, 4319-4347.

(76) Micklefield, J. Backbone modification of nucleic acids: Synthesis, structure and therapeutic applications. *Curr. Med. Chem.* **2001**, *8*, 1157-1179.

(77) Schoning, K. U.; Scholz, P.; Guntha, S.; Wu, X.; Krishnamurthy, R. et al. Chemical etiology of nucleic acid structure: The a-Threofuranosyl-(3'-2') oligonucleotide system. *Science* **2000**, *290*, 1347-1351.

(78) Schoning, K.-U.; Scholz, P.; Wu, X.; Guntha, S.; Delgado, G. et al. The a-L-Threofuranosyl-(3'-2')-oligonucleotide system (TNA): Synthesis and pairing properties. *Helv. Chim. Acta* **2002**, *85*, 4111-4153.

(79) Hendrix, C.; Rosemeyer, H.; Verheggen, I.; Seela, F.; Van Aerschot, A. et al. *Chem. Eur. J.* **1997**, *3*, 110.

(80) Kerremans, L.; Schepers, G.; Rozenski, J.; Busson, R.; Aerschot, A. V. et al. Hybridization between "six-membered" nucleic acids: RNA as a universal information system. *Org. Lett.* **2001**, *3*, 4129-4132.

(81) Huang, Z.; Schneider, C. K.; Benner, S. A. Building blocks for oligonucleotide analogues with dimethylene sulfide, sulfoxide, and sulfone groups replacing phosphodiester linkages. *J. Org. Chem.* **1991**, *56*, 3869-3882.

(82) Huang, Z.; Benner, S. A. Oligodeoxyribonucleotide analogues with bridging dimethylene sulfide, sulfoxide, and sulfone groups. Toward a second-generation model of nucleic acid structure. *J. Org. Chem.* **2002**, *67*, 3996-4013.

(83) Eschenmoser, A. Chemical etiology of nucleic acid structure. *Science* **1999**, *284*, 2118-2124.

(84) Nelson, K. E.; Levy, M.; Miller, S. L. Peptide nucleic acids rather than RNA may have been the first genetic code. *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 3868-3871.

(85) Poole, A.; Penny, D.; Sjoberg, B.-M. Methyl-RNA: an evolutionary bridge between RNA and DNA? *Chem. Biol.* **2000**, *7*, 207-216.

(86) Hendrickson, W. A.; Pahler, A.; Smith, J. L.; Satow, Y.; Merritt, E. A. et al. Crystal structure of core streptavidin determined from multiwavelength anomalous diffraction of synchrotron radiation. *Proc. Natl. Acad. Sci. USA* **1989**, *86*, 2190-2194.

(87) Yang, W.; Hendrickson, W. A.; Crouch, R. J.; Satow, Y. Structure of ribonuclease H phased at 2A resolution by MAD analysis of the selenomethionyl protein. *Science* **1989**, *249*, 1398-1405.

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

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

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

(91) Freier, S. M.; Altmann, K.-H. The ups and downs of nucleic acid duplex stability: stucture-stability studies on chemically modified DNA:RNA duplexes. *Nucleic Acids Res.* **1997**, *25*, 4429-4443.

(92) Friebolin, H. *Basic One- and Two-Dimensional NMR Spectroscopy*; 3rd ed.; WILEY-VCH, 1998.

(93) Greene, T. W.; Wuts, P. G. M. *Protective Groups in Organic Synthesis*; 3rd ed.; John Wiley & Sons, 1999.

(94) Moriguchi, T.; Yanagi, T.; Kunimori, M.; Wada, T.; Sekine, M. Synthesis and property of aminoacylamido-AMP: Chemical optimization for the construction of an N-acyl phosphoramidate linkage. *J. Org. Chem.* **2000**, *65*, 8229-8238.

(95) Cosstick, R.; Vyle, J. S. Solid phase synthesis of oligonucleotides containing 3'-thiothymidine. *Tetrahedron Lett.* **1989**, *30*, 4693-4696.

(96) Gryaznov, S. M.; Chen, J.-K. Oligodeoxyribonucleotide N3'-P5' phosphoramidates: synthesis and hybridization properties. *J. Am. Chem. Soc.* **1994**, *116*, 3143-3144.

(97) Weinstein, L. B.; Earnshaw, D. J.; Cosstick, R.; Cech, T. R. Synthesis and characterization of an RNA dinucleotide containing a 3'-S-phosphorothiolate linkage. *J. Am. Chem. Soc.* **1996**, *118*.

(98) Gryaznov, S. M.; Winter, H. RNA mimetics: oligonucleotide N3'-P5' phosphoramidates. *Nucleic Acids Res.* **1998**, *26*, 4160-4167.

(99) Belostotskii, A. M.; Keren-Yeshuah, H.; Lexner, J.; Hassner, A. New 3'-deoxythymidines bearing a nucleophilic 3'-substituent. *Nucleosides, Nucleotides & Nucleic Acids* **2001**, *20*, 93-101.