DNA COMPLEXES CONTAINING

NOVEL AROMATIC RESIDUES

by

WENZHAO DONG

STEPHEN WOSKI, COMMITTEE CHAIR

PATRICK FRANTOM
EDWIN STEPHENSON
SHANE STREET
RUSSELL TIMKOVICH

A DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy
in the Department of Chemistry
in the Graduate School of
The University of Alabama

TUSCALOOSA, ALABAMA

2016

ABSTRACT

The investigation of DNA complexes containing novel aromatic residues was performed. In the first part of this work, a series of novel nucleosides possessing a C1'-carboxamide linkage between the aryl moiety and the sugar group were successfully introduced into a single strand DNA oligonucleotide. The results of the thermal denaturation studies indicate that the incorporation of the modified nucleosides into DNA complexes destabilizes the DNA duplexes. However, the "bulged" complexes are only slightly destabilized and they are the most stable complexes among all the DNA complexes containing novel aromatic residues. This suggests that the carboxamide motif may be a general method for the insertion of non-natural residues into DNA for applications such as spectroscopic probes.

The second part of this study involves a seven step synthesis of novel aryl C-nucleosides. The aromatic residues are directly linked to the deoxyribose moieties through a carbon-carbon connection instead of the original carbon-nitrogen glycoside bond. Three novel C-nucleosides containing 4-substituted phenyl residues were successfully synthesized by following this synthesis scheme. The isomer problem involved in the multi-step synthesis of aryl C-nucleoside was resolved and as a result, the β-aryl C-2′-deoxynucleoside can be successfully separated from the α-aryl C-2′-deoxynucleoside. The synthesized aryl C-nucleoside can be introduced into a DNA oligonucleotide as a non-natural nucleobase.

The third part of this research was focused on the determination of the structure of DNA oligonucleotide duplexes containing aryl C-nucleoside using 2D NMR techniques and computational methods. 2D NMR experiments including COSY and NOESY were performed,

followed by resonance assignment and structure calculation to construct the preliminary 3D structure of DNA oligonucleotide duplex containing aryl C-nucleoside. Due to the limitation of the obtained restrains from NMR experiment, the study of molecular modeling has been performed to compensate the ambiguous of the preliminary structure. Conflicts between the calculated duplex structure and the data from the NMR experiment were observed, so an alternate possible structure of hairpin was proposed. The results of thermal denaturation study and molecular modeling may indicate that the hairpin structure is more preferred than the duplex structure for the non-natural DNA oligonucleotide containing aryl C-nucleoside.

LIST OF ABBREVIATIONS

1D	one-dimensional
2D	two-dimensional
3D	three-dimensional
A	adenine, adenosine
br	broad
C	cytosine, cytidine
cal	calorie
C−C	carbon-carbon
C−N	carbon-nitrogen
$CDCl_3$	*deuterated* chloroform
cm	centimeter
CNS	crystallography & nuclear magnetic resonance system
CPG	controlled-pore-glass
COSY	correlation spectroscopy
d	doublet
dd	doublet of doublets
ddd	doublet of doublet of doublets
ddI	double deionized
DQF-COSY	double quantum filtered correlation spectroscopy
DMAP	4-dimethylaminopyridine

DMF	*N*,*N*-dimethylformamide
DMT	4,4′-dimethoxytrityl
DMTrCl	4,4′-dimethoxytrityl chloride
DNA	deoxyribonucleic acid
dt	double of triplet
EDTA	ethylenediaminetetraacetic acid
g	gram(s)
G	guanine, guanosine
^{1}H NMR	proton nuclear magnetic resonance
HETCOR	heteronuclear correlation spectroscopy
HIV-1	human immunodeficiency virus type 1
HMDS	hexamethyldisilazane
HPLC	high-performance liquid chromatography
HSV-1	herpes simplex virus type 1
IMP	inosine-5′-monophosphate
K	Kelvin
kcal	kilocalorie
L	liter
LNA	locked nucleic acid
m	multiplet
M	molar
MALDI-TOF	matrix-assisted laser desorption ionization-time of flight
MHz	megahertz

min	minute
mL	milliliter
mmol	millimole
mol	mole
MS	mass spectrometry
ms	millisecond
MWCO	molecular weight cut-off
NAD^+	nicotinamide adenine dinucleotide
NMR	nuclear magnetic resonance
NOE	nuclear Overhauser effect
NOESY	nuclear Overhauser effect spectroscopy
OPC	oligonucleotide purification cartridge
Pd	palladium
PES	sodium phosphate/EDTA/sodium chloride buffer
PNA	peptide nucleic acid
ppm	parts per million
R_f	retention factor (TLC)
RNA	ribonucleic acid
s	singlet
t	triplet
T	thymine, thymidine
TBDMS-Cl	*tert*-butyldimethylsilyl chloride
TEMED	tetramethylethylenediamine

TBE	tris/borate/EDTA buffer
THF	tetrahydrofuran
TLC	thin-layer chromatography
TOCSY	total correlation spectroscopy
TEA	triethylamine
TEAA	triethylammonium acetate
TFA	trifluoroacetic acid
TMS	tetramethylsilane
UV/Vis	ultraviolet/visible
V	volt
Å	angstrom
℃	degree Celsius
ΔH	change in enthalpy
ΔG	change in Gibb's free energy
ΔS	change in entropy
μL	microliter
μM	micromolar

ACKNOWLEDGMENTS

I would like to express my sincere gratitude to my research advisor, Dr. Stephen A. Woski, for the opportunity to work and study in his laboratory and for his encouragement, patient and guidance throughout the completion of my dissertation. I really appreciate Dr.Woski's assistance in all of my research projects. Without my research advisor's advice and suggestion, my projects would not be completed perfectly. I would also like to thank the members of my dissertation committee for taking time to review my research progress.

I would like to thank my parents for their love, encouragement and support all the time during these years. Sincere thanks to my wife, Minli Xing, I could not have reached this point without her love and support. Special thanks to all of my friends for their generous assistance in work and my personal life when I needed it.

I wish to thank the Department of Chemistry of the University of Alabama for letting me fulfill my dream of being a graduate student here and for their funding. I would like to thank all the staffs and faculties in the Department of Chemistry for their assistance and friendly demeanor, especially Dr. Frantom for his patient proofreading for my literature seminar, Dr. Belmore and Dr. Timkovich for their training and help during my NMR experiment and Dr. Liang for her training and assistance in my MS experiments.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

Introduction

1.1 Deoxyribonucleic acid

Deoxyribonucleic acid (DNA) is an important biological molecule in all living cells because it stores the genetic information that is transmitted from one generation to the next. The basic units of DNA are nucleotides, which can be joined together by phosphodiester linkages to form oligodeoxynucleotides and polynucleotides. The structure of DNA is best considered as being composed of two parts: (1) backbone made of alternating phosphodiester groups and sugars (deoxyribose), and (2) nucleobases that adorn the backbone at each sugar position. The nucleobases are flat, aromatic heterocycles, and they include monocyclic pyrimidines (thymine and cytosine) and bicyclic purines (guanine and adenine). The genetic information is encoded in the specific sequence of nucleotide bases.

The three-dimensional structure of DNA duplex was first elucidated by Watson and Crick in 1953.[1] The sugar-phosphate backbone winds around the exterior of the helix, and paired nucleobases are located in the core. The specific pairing of each of the four nucleobases with a single partner (A-T, T-A, C-G, G-C) forms the basis for information storage and transmission. The sugar-phosphate backbone is anionic and hydrophilic. However, the purines and pyrimidines nucleobases are neutral and hydrophobic. The nearly planar ring structure of nucleobases stack in an offset face-to-face fashion along the helical axis, and only the hydrophilic heteroatom substituents on the edges of the nucleobase rings are exposed to the solvent.

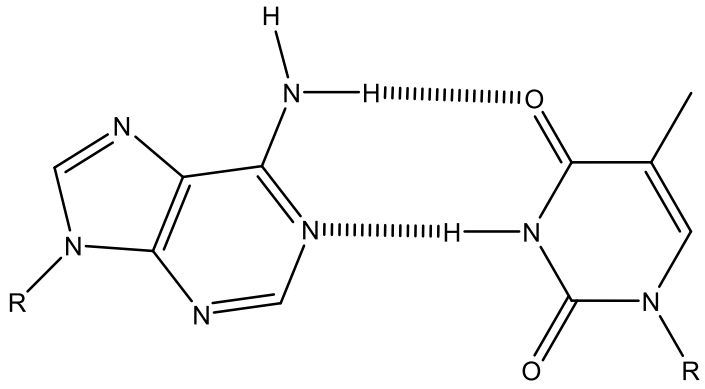
Several different conformations can be found in the DNA double helices depending on the medium condition and the DNA sequence. The predominant DNA structure found in nature is the right-handed B-DNA conformation. Other conformations including right-handed A-DNA and left-handed Z-DNA play only a minor part in nature.[2] The B-DNA conformation is most stable under physiological conditions, and it is the standard point in the study of the properties of DNA. The B-DNA structure has a diameter of about 20 Å and 10.5 base pairs per helical turn. The base pairs in the B-DNA conformation are stacked almost perpendicularly to the central axis with a 3.4 Å interval. Meanwhile, in B-DNA, the deoxyribose ring adopts a C2′-endo conformation, and the glycosyl bond conformation is in an *anti*-configuration. The structure of B-DNA conformation creates two distinct helical grooves including major grooves and minor grooves on the surface of the DNA duplex. These grooves play important roles in the DNA function because they can create special microenvironments for binding by proteins and other ligands.

The structure of a DNA duplex in aqueous medium is maintained by a balance of three non-covalent interactions.[3] Hydrogen bonding and base stacking are energetically favorable and stabilize the structure of DNA duplex. The third force, interstrand electrostatic repulsion between the two negatively charged sugar-phosphate backbone, is energetically unfavorable. However, this repulsive force can be partially ameliorated by the interaction of negatively charged backbone and organic/inorganic cations. Together, these weak forces combine to form the stable helical structure and contribute to the interaction of functional proteins and other ligands with DNA.

Because of the central role it plays in Watson-Crick base pairing, hydrogen bonding has been regarded as the most essential interaction for base pairing specificity and duplex stability (Figure 1.1). Numerous studies have been performed to elucidate the importance of these

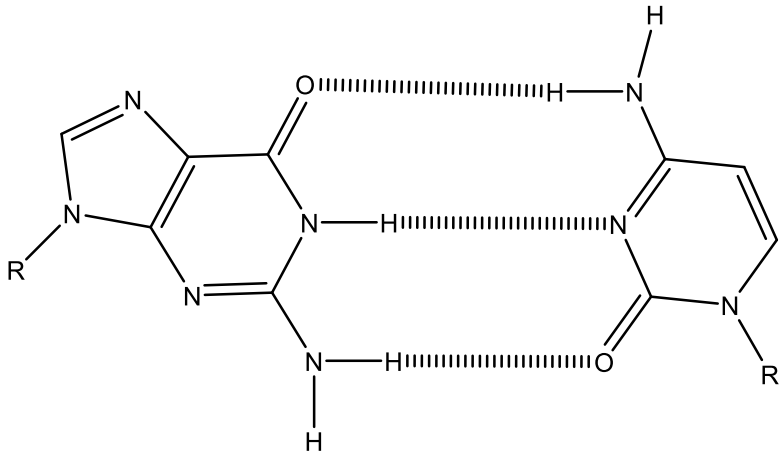
interactions.[4, 5] However, it was frequently unappreciated that the base pair hydrogen bonds merely replace roughly equivalent hydrogen bonds between the bases and water. The energy consumed to break these hydrogen bonds should roughly negate any energy gained by base pairing. Thus, hydrogen bonds cannot be responsible for stabilizing the duplex. Recent research also showed the pairing selectivity and the stability in DNA complexes can be accomplished without the hydrogen bonds, but the existence of hydrogen bonding is considered to be an important factor in the properly orienting the nucleobases within the helical structure.[6,7]

In addition to hydrogen bonding, base stacking interaction is also a potential significant contributor to the stability of DNA complexes. The base stacking interaction is constructed vertically along the helical axis by the offset face-to-face contacts between the flat nucleobases. Several interactions have been implicated in stacking including dispersion (van der Waals) forces, electrostatic interactions (dipole-dipole, quadrupole-quadrupole, and dipole-induced dipole), and solvophobic effects.[8-10] Some of these forces may cooperate to stabilize the overall helical structure of DNA complexes, however, the relative degree of each force contributes to the base stacking is unclear. Kool and coworkers[11] performed one study to focus on the base stacking interaction isolated from the hydrogen bonding interaction. Natural bases or synthesized non-natural bases were placed in dangling positions at the ends of a base paired duplex, so the bases at the ends do not have a pairing partner. Aryl C-nucleoside including benzene, naphthalene, phenanthrene and pyrene were examined. The results showed that the base stacking interaction is relatively strong and these interactions become more favorable with the increased size of the stacking residues, which means the order of base stacking capability was found to be pyrene > phenanthrene > naphthalene > benzene. The authors also believed that the solvophobic effect



Adenine

Thymine



Guanine

Cytosine

R = deoxyribofuranose

Figure 1.1 Hydrogen bonded A-T and G-C base pairs in DNA.

might play the crucial role in determining this order. With the same size, less polar molecule stacks more strongly that one of high polarity. Among the non-polar molecules, surface area is correlated with the stacking ability when excluded from solvent.

1.2 Modification of DNA

As the numerous discoveries of anticancer and antiviral properties of both natural and synthetic nucleosides/oligonucleotides have unfolded, nucleic acid chemistry has received much attention. Various applications of synthetic oligonucleotides for the treatment of different kinds of diseases (especially cancer) have been investigated extensively. In one approach, antisense technology, an oligonucleotide binds to the messenger RNA and blocks the synthesis of protein. Compared with natural oligonucleotides, the modified oligonucleotides can incorporate advantages such as nuclease resistance and cell permeability; these practical pharmacokinetic properties prompt the continuous development of non-natural oligonucleotides. Besides the significant application in medicinal chemistry, the modified oligonucleotides can also be useful tools for the determination of the biophysical properties of nucleic acids. The modifications of oligonucleotides can be divided into three categories: modification of the sugar, modification of the phosphate backbone, and modification of the nucleobases.

Sugar modification

There are several strategies have been used for the modification of sugar moiety. The modification can be achieved by the alteration of the size of the sugar ring including the ring expansion as well as the ring contraction (Figure 1.2 A). Some of these nucleosides have shown

selective and effective antiviral activity. For example, oxetanocin and cyclohexenyl guanine shown potent activity against herpes simplex virus type 1 (HSV-1).[12, 13]

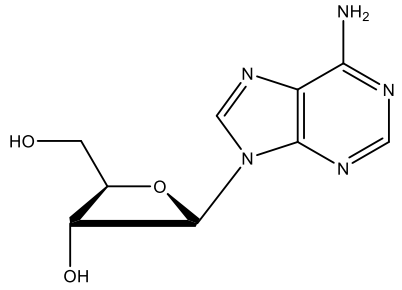
Another modification approach is the replacement of the oxygen with other appropriate atoms such as nitrogen, carbon and sulfur (Figure 1.2 B). For example, 4′-thiothymidine and its derivatives showed a strong antiviral activity with high selectivity in vitro against herpes simplex virus, varicella-zoster virus, and especially against Epstein-Barr virus;[14] C-thymidine also exhibited its significant antiviral feature, as well as its capability to increase the survival time of mice bearing leukemia cells;[15] immucilin is an excellent inhibitor of human purine nucleoside phosphorylase, moreover, immucilin's 5′-monophosphate ester has the potential ability to control the function disorders of T-cell.[16]

In addition, the replacement of 3′-hydroxyl group is also an appealing modification approach (Figure 1.2 C). For instance, both 3′-azido-2′, 3′-dideoxythymidine and 2′,3′-dideoxycytidine show potent antiviral activity against type 1 human immunodeficiency virus (HIV-1) by incorporation into the host DNA strand through the respective 5′- triphosphates.[17]

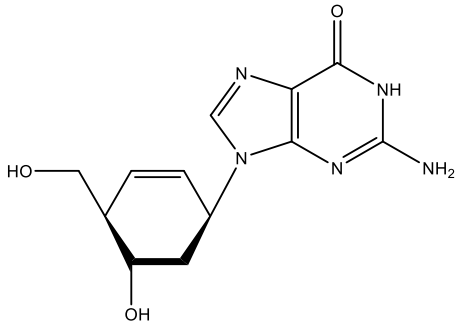
Phosphate backbone modification

Modifications of the backbone can provide a prolonged biological lifetime due to resistance to endogenous nucleases. These modifications can result in more stable duplexes when hybridized to a target nucleic acid. This can be observed with an increased melting temperature (T_m) determined using thermal denaturation method.[18] Peptide nucleic acid (PNA), in which a synthetic peptide backbone consisting of N-(2-aminoethyl)-glycine units linked by amide bonds is used to replace the sugar-phosphate backbone, and locked nucleic acid (LNA), which contain a methylene bridge that connects the 2′-oxygen of ribose with the 4′-carbon, are two well-known

A

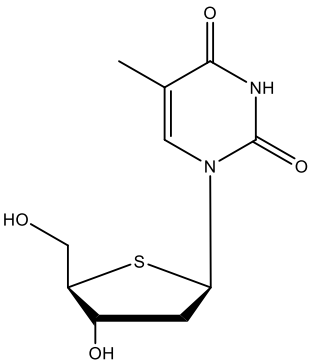


Oxetanocin

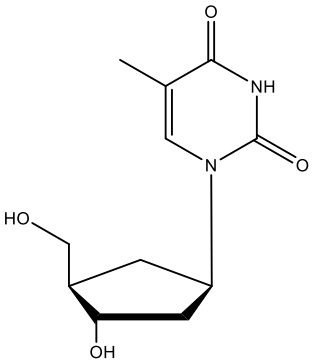


Cyclohexenyl guanine

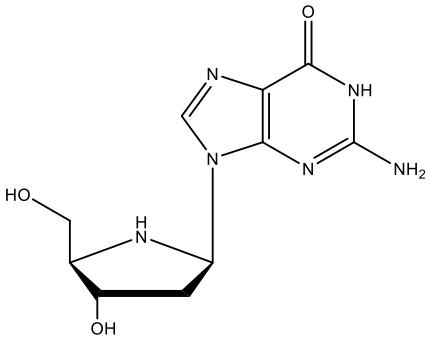
B



4'-Thiothymidine

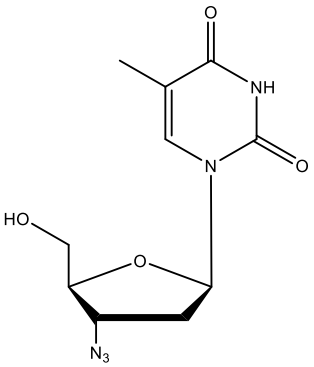


C-Thymidine

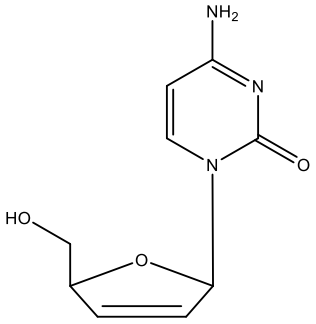


Immucilin

C



3'-azido-2',3'-dideoxythymidine
(AZT)



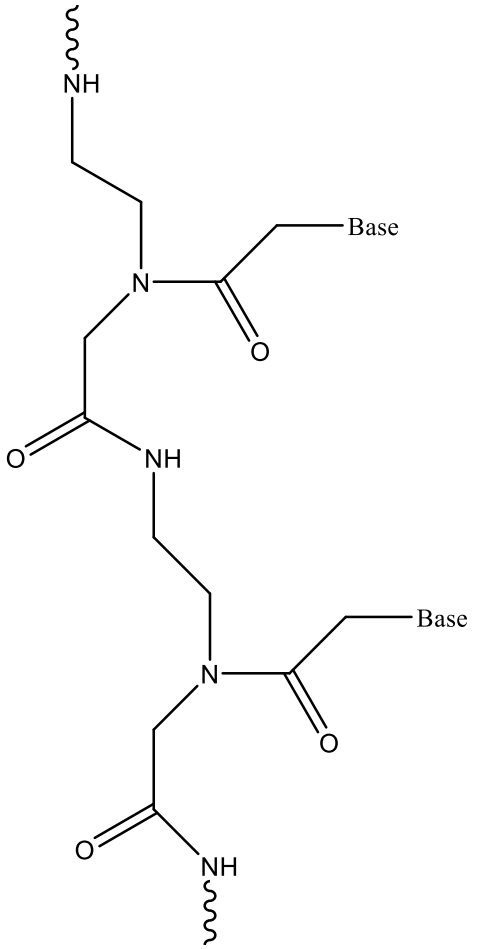
2',3'-dideoxycytidine
(ddC)

Figure 1.2 Structures of sugar modified nucleosides

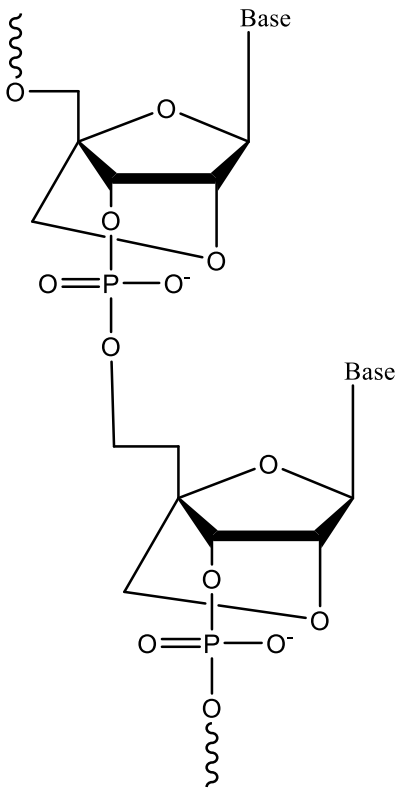
examples of backbone modified oligonucleotides (Figure 1.3).[19, 20]

The sugar-phosphate backbone of DNA possesses negative charges, and the electrostatic repulsion between the two DNA strands destabilizes the DNA structure. In PNA, the replacement of the sugar-phosphate backbone with an uncharged synthetic peptide backbone eliminates the repulsion, enhances the stability of DNA duplex and confers several favorable properties to PNA.[21] An attractive application of PNA is the regulation of the cellular process as an antisense probe. This approach is focused on the inhibition of transcription by binding to specific DNAs to inhibit the formation of messenger RNA; it also can be concentrated on the inhibition of translation of gene by selectively binding to messenger RNA to prevent from the expression of protein. For example, David Corey and co-workers[22] reported the inhibition of expression of the human caveolin 1 protein by PNA was successful. Elimination of negative charges can also be achieved by using neutral linkages. Instead of replacing the whole backbone as in PNA, the neutral linkage removes or neutralizes the natural anionic charges of the phosphodiesters to reduce the repulsion between the two DNA strands. For example, phosphoramidite,[24] methylphosphonate,[25] phosphotriester,[26] and phosphoramidimate[27] can be applied as the neutral or zwitterion ionic linkages and improve the stabilities of the DNA duplex (Figure 1.4).

Unlike PNA, the LNA can be regarded as a modified RNA. It keeps the structure of phosphate backbone but the ribose moiety is modified. The bridge created between 2′-oxygen and 4′-carbon can make the ribose in a locked 3′-endo conformation, which provides several unique properties including increasing the local organization of the phosphate backbone and reducing the conformational flexibility of the ribose.[28] The special structure results the improvement of the hybridization properties of oligonucleotides, reports have been shown that

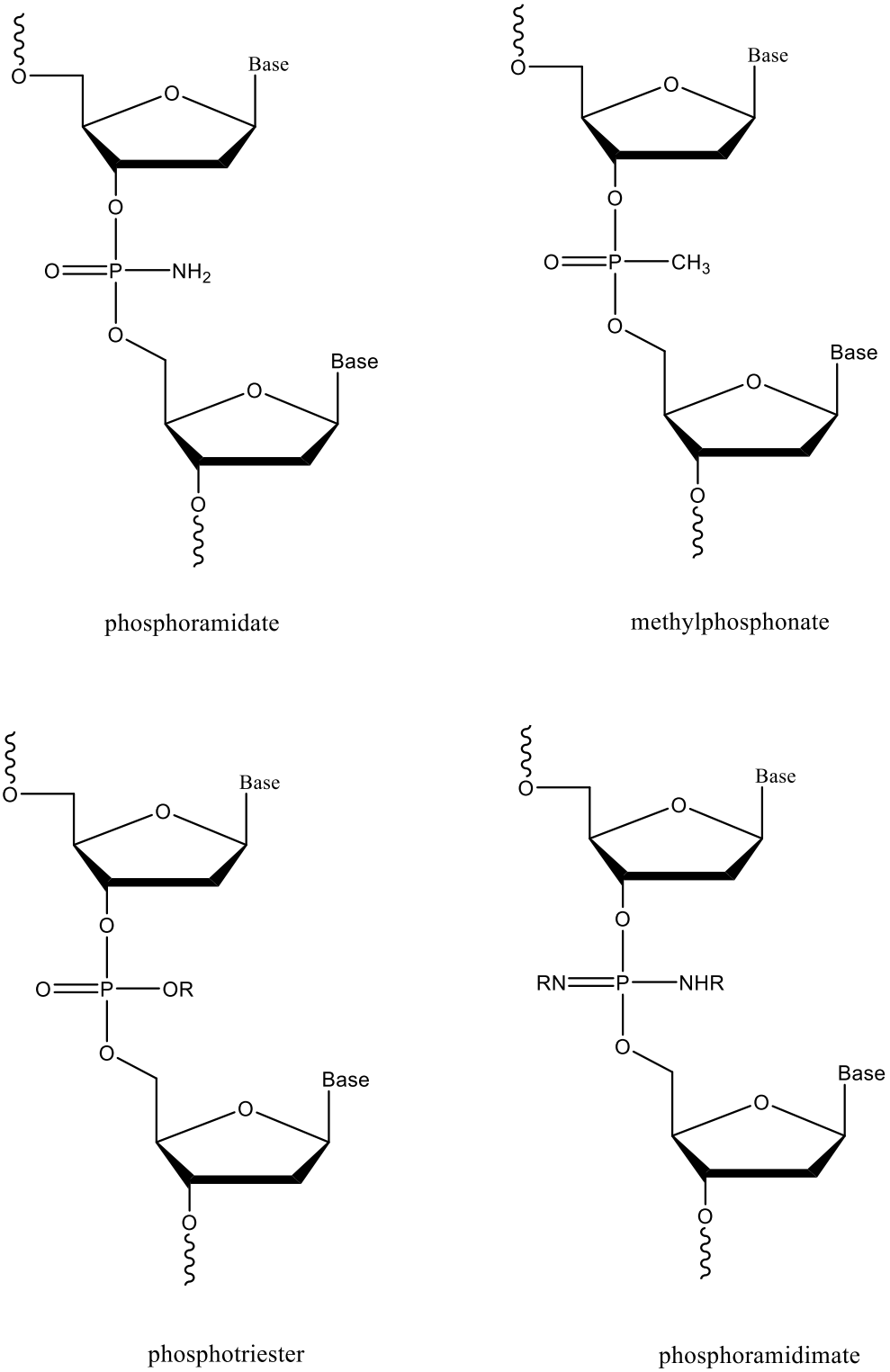


Peptide nucleic acid (PNA)



Locked nucleic acid (LNA)

Figure 1.3 Structures of PNA and LNA



phosphoramidate

methylphosphonate

phosphotriester

phosphoramidimate

Figure 1.4 Structures of neutral linkage DNA analogues

the melting temperature (Tm) can be increased by as much as 10 ℃.[29, 30] Similar to PNA, LNA can bind to the complementary DNA or RNA sequence, control gene expression, and show some useful applications such as the treatment of cancer by inhibiting the activity of telomerase.[31]

Nucleobases modification

Compared with modification of sugar and phosphate backbone, the modification of nucleobases allows synthesized functional groups to be incorporated into the interior of the DNA duplex. Several modified nucleobases have shown valuable properties such as antiviral, antibacterial, and anticancer activity. So the improvement of nucleobases modification is also an attractive approach for the discovery of drug design and development.

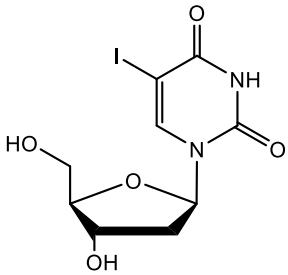
Several strategies are available for the modification of nucleobases. One relatively simple strategy is to modify natural nucleobases. Only one or two groups on the nucleobase are replaced (Figure 1.5 A), and the modification site can be chosen such that it does not affect the hydrogen bonding capability. These modifications retain most of the properties of the natural nucleobases. For example, an iodine atom can be used to replace the methyl group in thymidine to create the idoxuridine, which has showed the application for the treatment of ocular herpes infection.[32] Toyocamycin is another example of nucleobase modification, this compound showed the RNA binding ability and anticancer activity.[33] Another attracting base modification with minor structural transformation is to incorporate an alkyne functional group to the nucleobases (Figure 1.5 A), which can introduce the application of the copper catalyzed alkyne–azide cycloaddition reaction mediated click chemistry on the DNA oligonucleotide. For example, oligonucleotides can be labeled with fluorescent dyes or other reporter groups, DNA can be cyclized and DNA catenanes can be prepared.[34, 35]

A relatively complicated method is the replacement of the whole natural nucleobases (Figure 1.5 B&C). This kind of modification can provide the application in the field of medicinal chemistry, for instance, tiazofurin is composed of a thiazole ring system containing a carboxamide residue (Figure 1.5 B). Tiazofurin is considered an inosine-5′-monophosphate (IMP) dehydrogenase inhibitor and has displayed potential antiviral and anticancer activity.[36] In addition to the chemotherapy application, nucleobase modification can also be applied for the structure study of nucleic acid. In some of the cases, the natural C-N glycoside bond can be replaced with the C-C bond. As discussed above, Kool and coworkers[11] used the aryl C-nucleoside to examine the base stacking interaction in a DNA complex. By utilizing the C-C bond, Coleman and coworkers[37] constructed a coumarin C-nucleoside, and this non-natural base can be incorporated into the DNA complex as a fluorescent probe (Figure 1.5 B). In our research group, the nucleobases were replaced by numerous types of aryl groups containing different substitutions, and the C-N glycoside bond has been replaced by either C-C bond or carboxamide bond (Figure 1.5 C).[38-40] Based on the stability investigation of DNA complexes containing modified nucleobases, one significant result was obtained in our research group that revealed the major forces for DNA base stacking is caused by the dipole-induced dipole interactions.[38]

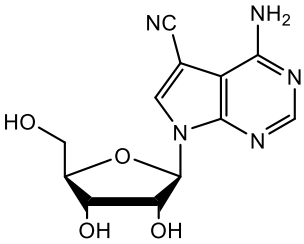
1.3 Summary

The synthesis of novel non-natural nucleobases has been studied in a significant success, and effective synthesis strategies have been discovered and can be used for the synthesis of modified nucleoside monomers.[36-38] As a result, various non-natural nucleosides can be produced and incorporated into DNA oligonucleotides and these modified DNA strands can obtain useful biochemical properties. These modified DNA oligonucleotides can lead to a valuable insight

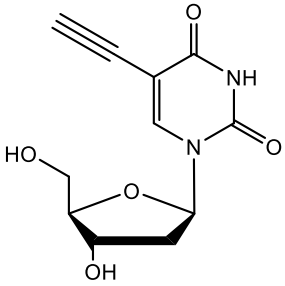
A



Idoxuridine

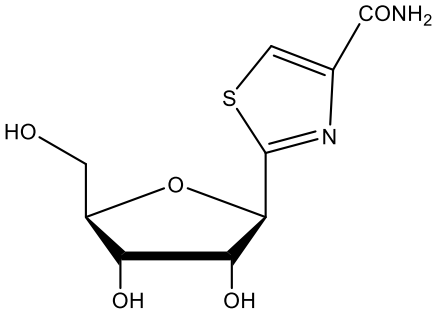


Toyocamycin

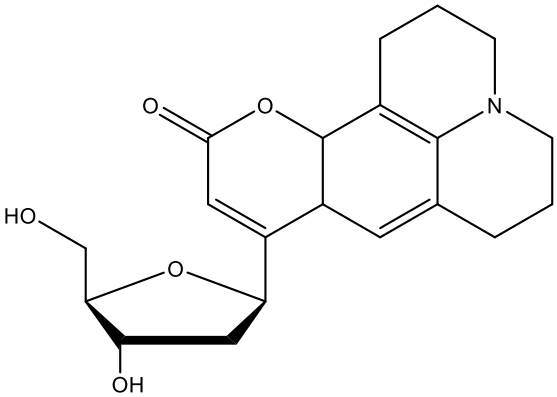


Alkyne

B

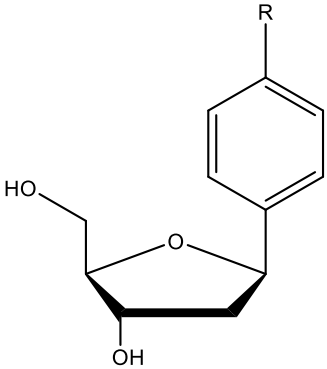


Tiazofurin

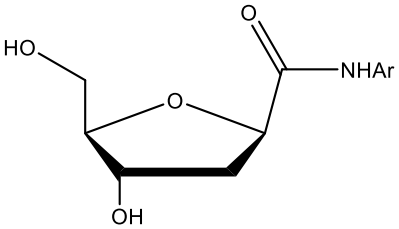


Coumarin

C



Aryl C-nucleoside



Nucleoside containing carboxamide residue

Figure 1.5 Structures of nucleobase modified nucleosides

when examination of hydrogen bonding and base stacking interaction was performed.[35, 37] With the introduced new functional groups, the modified residues have potential applications for the 3D structure determination of nucleic acid and as spectroscopic probes.

In Chapter II, building on the previous research in our laboratory, a series of natural DNA oligonucleotides and modified DNA oligonucleotides containing carboxamide residues were synthesized. To study the stability of DNA duplexes containing modified residues, thermal denaturation experiments were performed and thermodynamic data for hybridization were obtained. Two results are described. The first reveals the role of substitutions that can affect the stability of DNA duplexes; the second, involves the relationship between the thermodynamic parameters and the structure of DNA duplex.

In Chapter III, three modified nucleosides were synthesized through an eight-step reaction scheme. The specific compounds that were made are the phenyl, 4-tolyl, and 4-fluorophenyl aryl C-nucleoside. The synthetic routes and optimization to improve the efficiency are described. Specially, the unexpected observation that a mixture of β and α anomers were made using a literature procedure resulted in the use of a different starting material in the key Heck reaction.

In Chapter IV, the structure of DNA duplex containing novel aryl C-nucleoside residue was studied using 2D NMR techniques and computational methods. The unmodified Dickerson dodecamer was studied first and used as reference for the determination of NMR experimental conditions and parameters. Based on the result of these experiments, optimal conditions were applied for the structure study of non-natural DNA duplex containing aryl C-nucleoside residue. These studies were supplemented by thermal denaturation studies and computer-based structural calculation that can be applied to elucidate the 3D structure of non-natural DNA oligonucleotide.

1.5 References

1. Watson, J.D.; Crick, F.H. *Nature* **1953**, *171*, 737.

2. Blackburn, G M.; Gait, M.J. *Nucleic Acid in Chemistry and Biology*; Oxford University Press: New York, **1996**; Chapter 1, and 2.

3. Anderson, M. L. M. *Nucleic Acid Hybridization*; Bios Scientific Publishers: Oxford, **1999**; Chapter 1.

4. Kool, E. T.; Morales, J. C.; Guckian, K. M. *Angew. Chem. Int. Ed.* **2000**, *39*, 991.

5. Brotsch, C.; Haberli, A.; Leumann, C. J. *Angew. Chem. Int. Ed.* **2001**, *40*, 3012.

6. Schweitzer, B. A.; Kool, E. T. *J. Am. Chem. Soc.* **1995**, *117*, 1863.

7. Matray, T. J.; Kool, E. T. *J. Am. Chem. Soc.* **1998**, *120*, 6191.

8. Wu, Y.; Ogawa, A. K.; Berger, M.; McMinn, D. L.; Schultz, P. G.; Romesberg, F. E. *J. Am. Chem. Soc.* **2000**, *122*, 7621.

9. Honig, B.; Friedman, R. A. *Biopolymers* **1992**, *32*, 145.

10. Sundaralingam, M.; Ponnuswamy, P. K. *Biochemistry* **2004**, *43*, 16467.

11. Guckian, K. M.; Schweitzer, B. A.; Ren, R. X-F.; Sheils, C. J.; Paris, P. L. Tahmassebi D. C.; Kool E.T. *J. Am. Chem. Soc.* **1996**, *118*, 8182.

12. Hoshino, H.; Shimizu, N.; Shimada, N.; Takita, T.; Takuchi, T. *J. Antibiot.* **1978**, *40*, 1077.

13. Wang, J.; Froeyen, M.; Hendrix, C.; Andrie, G.; Snoeck, R.; Clercq, E. D.; Heredewijn, P. *J. Med. Chem.* **2000**, *43*, 736.

14. Coen, N.; Duraffour, S.; Haraguchi, K.; Balzarini, J.; van den Oord, J. J.; Snoeck, R.; Andreia, G. *Antimicrob. Agents Chemother.* **2014**, *58*, 4328.

15. Shealy, Y. F.; O'Dell, C. A.; Shanon, W. M.; Arnett, G. *J. Med. Chem.* **1984**, *27*, 1416.

16. Evans, G. B.; Furneanux, R. H.; Hutchison, T. L.; Kezar, H. S.; Morris, P. E.; Scharman, V. L.; Tyler, P. C. *J. Org. Chem.* **2001**, *66*, 5723.

17. DeClercq, E. *AIDS Res. Hum. Retroviruses* **1992**, *8*, 119.

18. Nulf, C. J.; Corey, D. R. *Nucleic Acids Res.* **2004**, *32*, 3792.

19. Koshkin, A. A.; Singh, S. K.; Nielsen, P.; Rajwanshi, V. K.; Kumar, R.; Meldgaard, M.; Olsen C. E.; Wengel, J. *Tetrahedron* **1998**, *54*, 3607.

20. Nielsen, P. E.; Egholm, M.; Berg, R. H.; Buchardt, O. *Science* **1991**, *254*, 1497.

21. Eriksson, M.; Nielsen, P. E. *Q. Rev. Biophys*. **1996**, *29*, 369.

22. Kaihatsu, K.; Huffman, K. E.; Corey, D. R. *Biochemistry* **2004**, *43*, 14340

23. Shammas, M. A.; Simmons, C. G.; Corey, D. R.; Reis, R. J. S. *Oncogene* **1999**, *18*, 6191.

24. Peyrottes, S.; Vasseur, J.-J.; Imbach, J.-L.; Raynor, B. *Nucleic Acid Res.* **1996**, *24*, 1841.

25. Bec, C. L.; Wickstrom, E. *J. Org. Chem.* **1996**, *38*, 510.

26. Hayakawa, Y.; Hirose, M.; Hayakawa, M.; Noyori, R. *J. Org. Chem.* **1996**, *61*, 925.

27. Fischer, R. W.; Caruthers, M. H. *Tetrahedron Lett.* **1995**, *36*, 6807.

28. Braasch, D. A., Corey, D. R. *Chem. Biol.* **2001**, *8*, 1.

29. Wengel, J. *Acc. Chem. Res.* **1999**, *32*, 301.

30. Braasch, D. A.; Liu, Y.; Corey, D. R. *Nucleic Acids Res.* **2002**, *30*, 5160.

31. Elayadi, A. N.; Braasch, D. A.; Corey, D. R. *Biochemistry* **2002**, *41*, 9973.

32. Humber, D. C.; Mulholland, K. R.; Stoodley, R. J. *J. Chem. Soc. Perkin Trans. I* **1990**, 283

33. Ramasamy, K.; Robins, R. K.; Revankar, G. R. *J. Heterocycl. Chem.* **1988**, *25*, 1043.

34. Kolb H. C.; Finn M. G.; Sharpless K. B. *Angew. Chem. Int. Ed.* **2001**, *40*, 2004.

35. El-Sagheer A. H.; Brown T. *Chem. Soc. Rev.* **2010**, *39*, 1388.

36. Srivastava, P. C.; Pickering, L. B.; Allen, L. B.; Streeter, D. G.; Campbell, M. T.; Witkowski, J. T.; Sidwell, R .K.; Robins, R. K. *J. Med. Chem.* **1977**, *20*, 256.

37. Coleman, R. S.; Madaras, M. L. *J. Org. Chem.* **1998**, *63*, 5700.

38. Liu, C. "Investigation of the stabilities of PNA&DNA and DNA&DNA complex containing novel aromatic residues", Ph.D. Dissertation, University of Alabama, **2006**.

39. Chu, Y. "Synthesis and characterization of non-natural DNA with novel nucleosides", M.S. Thesis, University of Alabama, **2002**.

40. Wichai, U. “Synthesis and investigation of PNA:DNA complexes containing novel aromatic residues”, Ph.D. Dissertation, University of Alabama, **2003**.

CHAPTER II

Synthesis and Stability Studies of DNA Duplexes with 1′-Carboxamide Residues

2.1 Introduction

The introduction of internal labels into DNA structures has received much attention as a result of its numerous applications. A deeper understanding of the molecular interactions involved in normal cellular functions and diseases can be achieved by comparison of the characteristics of non-natural and natural DNA species. Moreover, new functional and reporter groups can be introduced into DNA, which can have potential applications for biotechnology research and stereoselective chemical reactions.[1] For example, an artificial DNAzyme, which was created by combination of a double-stranded DNA as a scaffold and a metal complex used as a catalytic site, can be applied to stereoselective Diels-Alder reaction.[2, 3]

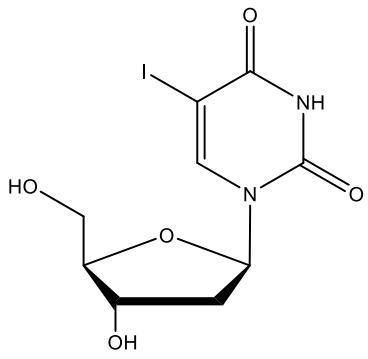
Several strategies are available for incorporation of non-natural functional groups into oligonucleotides, including modifications of the 5′- and 3′-termini, the sugar groups, the phosphodiesters and the nucleobases.[4, 5] It is easiest to modify DNA strands at their termini. Modification of nucleobase sites is generally more synthetically challenging than any other sites. However, unlike modification of the termini, the modification of the sugars, phosphodiesters or nucleobases has the advantage of allowing for incorporation at any site within the DNA strand. In addition, unlike sugar and phosphodiester groups which are located outside the DNA helix, the nucleobases are located in the core of the DNA helical structure. Modification of the nucleobases enables the incorporation of the non-natural functional groups into the core of the

DNA duplex. Therefore, the non-natural functional groups can interact intimately the stacked bases of the DNA helix.

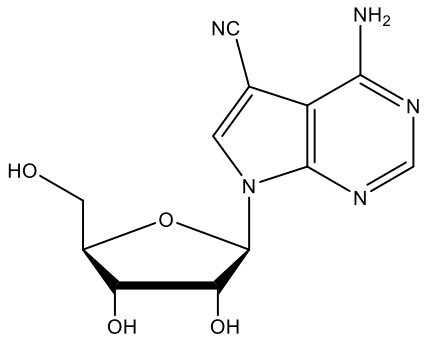
Introduction of modified nucleobases can be accomplished by modification at positions that do not affect the hydrogen bonding such as the 7- and 8-positions of purines and 5-position of pyridines (Figure 2.1 A).[6, 7] These modified residues typically bear strong similarities to the natural bases from which they were derived and retain most of their original properties.[6, 7, 8] Modification can also involve complete replacement of nucleobase by a novel residue, such as 3-nitropyrrole and 5-nitroindole (Figure 2.1 B).[6, 7] These residues were developed as "universal" nucleosides because they cannot form specific hydrogen bonds to any of the four natural bases, but can stack into the DNA duplex. However, in both of these cases, a ring N-H group is present for glycosylation by an activated deoxyribose. The glycosylation with pyridine-like heterocycles is more difficult, and glycosylation with carbocyclic aromatic groups requires very different chemistry. Indeed, the methods to form such C-nucleosides are not general, and researchers will often have to try many reactions with different starting materials to find a successful one.[9-13]

Several years ago, a project was initiated in the Woski group to develop a general method for the introduction of novel functional groups into oligodeoxynucleotides using 2′-deoxyribofuranose-1′-carboxamide residues (Figure 2.2).[14, 15] This design strategy maintains the native structures of 2′-deoxyribofuranosyl ring system and, after oligonucleotide synthesis, the phosphodiester groups. Chemical reactions to form amide bonds are quite abundant, and many of these methods can tolerate a variety of functional groups. Large numbers of substituted amines are commercially available for the production of a great variety of nucleoside derivatives. Finally, the stereochemistry of the amide group and its relatively rigid planar structure partly mimics the planar aromatic bases.

A

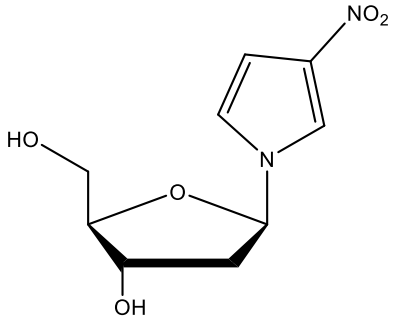


Idoxuridine

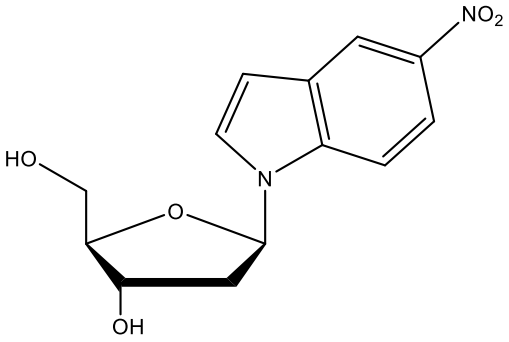


Toyocamycin

B



3-Nitropyrrole



5-Nitroindole

Figure 2.1 The structure of base modified nucleosides. A) 7-position of purines and 5-position of pyridines; B) 3-nitropyrrole and 5-nitroindole.

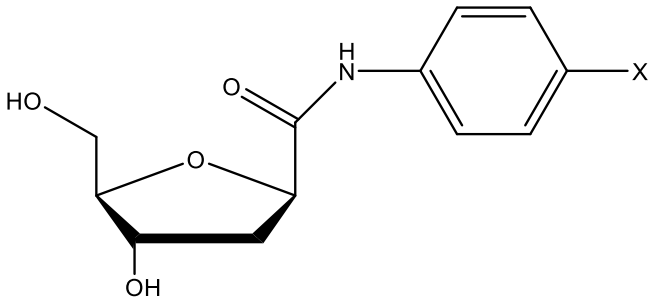


Figure 2.2 Structure of 2′-deoxyribofuranose-1′-carboxamide residue. The X represents the substitution site used in this study.

In previous work, synthesis of carboxamide DNA monomers was accomplished by Dr. Cuiling Liu.[15] Starting with inexpensive D-glucono-1,5-lactone, 8 steps produced a protected 3-deoxy-*ribo*-hexonic acid (Scheme 2.1).[16, 17] Arylamines were coupled to the carboxylic acid using the condensing agent 2-chloro-1-methylpyridinium iodide in the presence of tri-*n*-butylamine.[18] The *tert*-butyldimethylsilyl (TBDMS) protective groups were then removed by treatment with tetra-*n*-butylammonium fluoride (Scheme 2.2). Monomers suitable for solid-phase oligonucleotide synthesis on an automated DNA synthesizer were prepared by sequential protection of the 5′-hydroxyl group with a 4,4′-dimethoxytrityl (DMT) groups and phosphitylation of 3′-hydroxyl group using 2-cyanoethyl diisopropylchlorophosphoramidite (Scheme 2.3).[19]

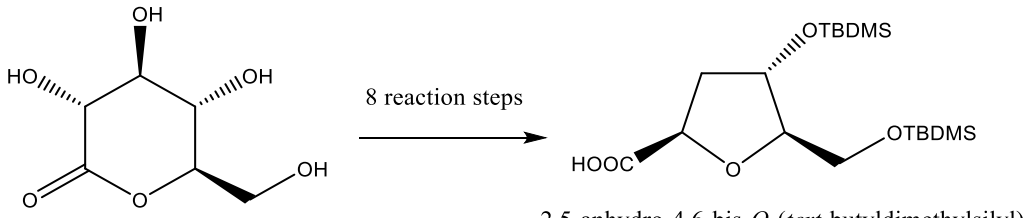
The current work involves a proof-of-concept study aimed at examining the effects of incorporating carboxamide residues into oligonucleotide strands. Four monomers are used: 4-methoxyphenyl, 4-nitrophenyl, 4-cyanophenyl and 4-chlorophenyl groups. These were all converted into corresponding phosphoramidite reagents for solid phase oligodeoxynucleotide synthesis. Thermal denaturation experiments were conducted to determine the effects of these residues on the stabilities of duplexes formed from three different DNA partners. The potential applications of carboxamide oligodeoxynucleotide which include structural studies of DNA, engineering of reporter molecules and functional studies of DNA are further discussed.

2.2 Results and Discussion

Oligonucleotide synthesis and hybridization

Natural oligodeoxynucleotides, non-natural oligodeoxynucleotides containing a stable abasic “d-spacer” residue, and non-natural oligodeoxynucleotides containing carboxamide

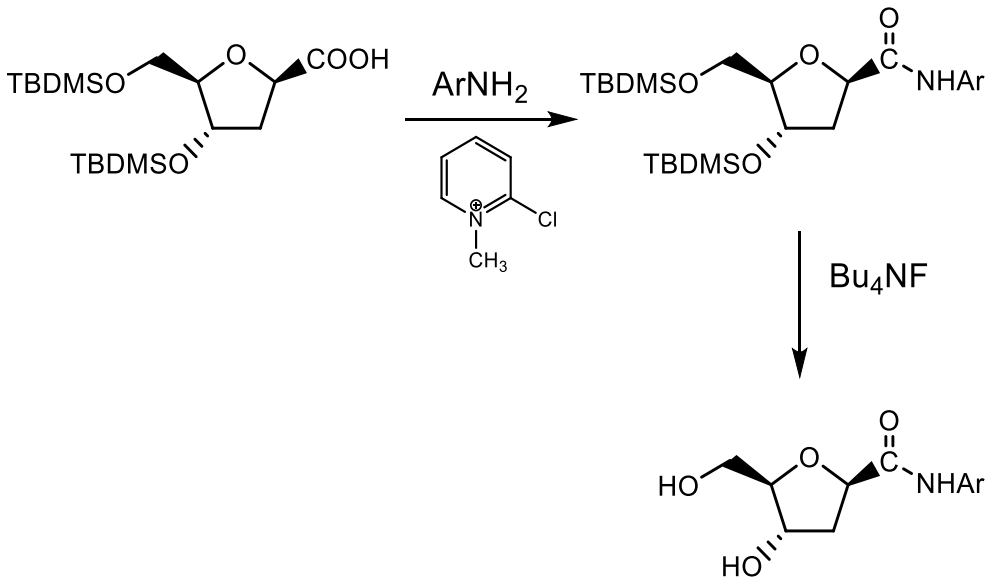
Scheme 2.1[15]



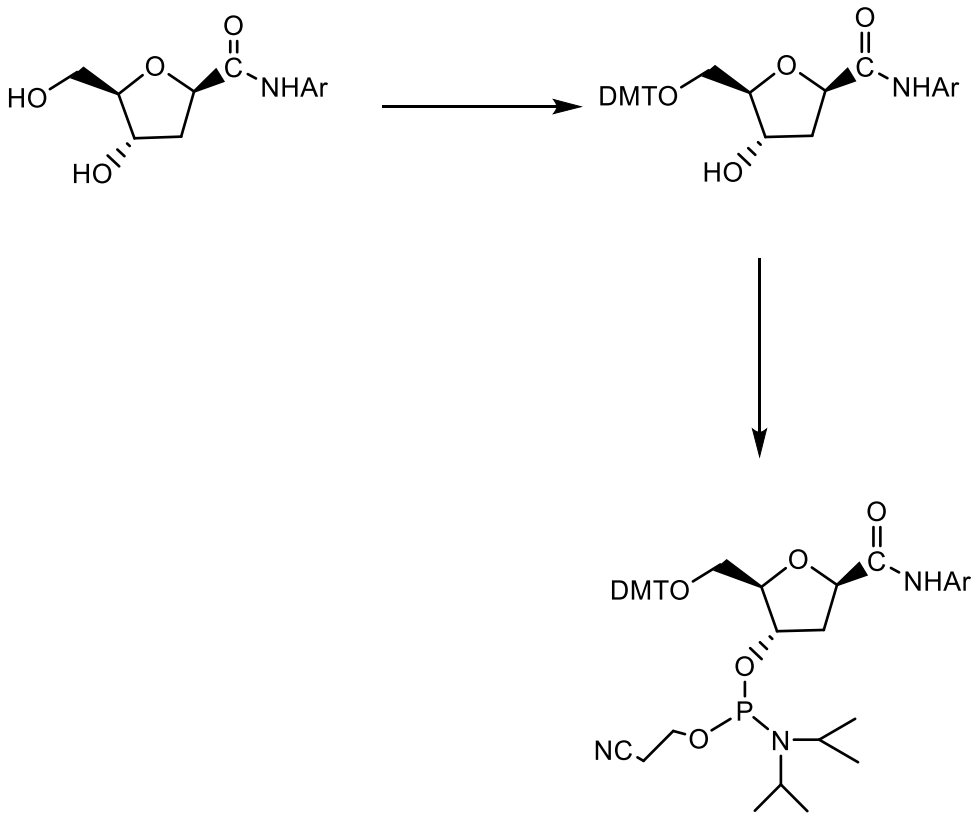
D-glucono-1,5-lactone

2,5-anhydro-4,6-bis-*O*-(*tert*-butyldimethylsilyl)-3-deoxy-D-*ribo*-hexonic acid

Scheme 2.2[15]


TBDMSO
O
COOH
TBDMSO
ArNH2
N
Cl
CH3
TBDMSO
O
O
C
NHAr
TBDMSO
Bu4NF
HO
O
O
C
NHAr
HO

Scheme 2.3[15]



Monomers for DNA synthesis

were synthesized on an Applied Biosystems 391 DNA synthesizer, which applies phosphoramidite chemistry on a solid phase support. Standard DNA synthesis cycles and monomers were used for the natural oligodeoxynucleotides and the abasic residue-containing oligodeoxynucleotides. “Standard” DNA phosphoramidites employ the following scheme to protect the exocyclic amines of the bases: C = *N*-acetyl, A = *N*-benzoyl, G = *N*-isobutyryl.

Preparation of oligodeoxynucleotides containing carboxamide residues require modifications to the standard procedures described above. First, in order to avoid hydrolysis of the carboxamide residues during the standard cleavage and deprotection using concentrated ammonium hydroxide, non-standard base protecting groups were required for adenosine and guanosine residues (phenoxyacetyl group for A and isopropylphenoxyacetyl for G). This allows the employment of an “ultramild” deprotection/cleavage reaction using potassium carbonate in methanol without affecting the novel carboxamide residue. Finally, the synthesis program for the couplings of the carboxamide residues was modified. The time allowed for the coupling reaction was doubled to thirty seconds, and the coupling step was repeated immediate after the first. Increasing the reaction time and “double coupling” were used to prevent low efficiencies that could have complicated the purification.

The crude oligodeoxynucleotides were purified by using denaturing polyacrylamide gel electrophoresis. After excision of the slice of polyacrylamide containing full length DNA, oligonucleotides were electroeluted, desalted, and resuspended in ddI water. The resulting stock solutions were quantitated by using UV absorbance spectroscopy at wavelength of 260 nm.

After successful synthesis and purification of oligodeoxynucleotides, experiments were performed to examine the hybridization properties of carboxamide-containing strand with complementary strand. There are five different types of double helical complexes (Table 2.1).

Complex I is a 14-mer duplex containing a thymine residue at the position complementary to the carboxamide residue. This arrangement forms a mismatch pair between a thymine and the carboxamide residue. Complex II is a 14-mer duplex containing a stable abasic residue positioned opposite to the carboxamide residue. This abasic residue contains a hydrogen atom in place of a nucleobase. Complex III is a duplex pairing a 13-mer oligodeoxynucleotide with the carboxamide-containing 14-mer. Because thirteen residues in the 14-mer strand are complementary to the 13-mer strand, a duplex can form with no partner for the carboxamide. This constitutes an internal one-nucleotide loop called a bulge. The remaining two complexes are controls: complex IV is a perfectly paired 13-mer duplex, and complex V is a 14-mer duplex containing an A-T base pair at the position modified in complexes I-III.

Thermal denaturation studies of DNA duplexes containing carboxamide residues

The molar absorptivities of the DNA bases are reduced in stacked conformations.[16] Thus, upon raising the temperature, it is possible to observe the cooperative transition between the double helix into a pair of unstructured single strands as a sigmoidal trace in the absorbance versus temperature graph. Using software to fit the data can elucidate both the T_m and the thermodynamic parameters ΔH and ΔS for the transition.[20] The T_m is the temperature where half of the double-helical complexes have melted to single strands. ΔH and ΔS are the enthalpy and entropy changes for the transition. Using the equation of state, $\Delta G = \Delta H - T\Delta S$, it is possible to find the free energy of the complex.

Thermal denaturation data for carboxamide-containing duplexes are summarized in Tables 2.2-2.4. The melting temperature of 14-mer control is about 1 ℃ higher than that of 13-mer control due to the additional insertion of an A-T base pair at the modification position. The

Table 2.1 DNA double helical complexes

DNA	Strands
Complex I (X—T)	5′-CGTTCG **X** GACAGCT-3′ 3′-GCAAGC **T** CTGTCGA-5′
Complex II (X—Abasic (θ))	5′-CGTTCG **X** GACAGCT-3′ 3′-GCAAGC θ CTGTCGA-5′
Complex III (Bulged X)	5′-CGTTCG **X** GACAGCT-3′ 3′-GCAAGC—CTGTCGA-5′
Complex IV (13-mer control)	5′-CGTTCG—GACAGCT-3′ 3′-GCAAGC—CTGTCGA-5′
Complex V (14-mer control）	5′-CGTTCG **A** GACAGCT-3′ 3′-GCAAGC **T** CTGTCGA-5′
* X represents carboxamide residue: 4-methoxyphenyl, 4-nitrophenyl, 4-cyanophenyl or 4-chlorophenyl.	

Table 2.2 Thermal denaturation of complexes I containing carboxamide residues paired with thymine

DNA	T_m (℃)	ΔH (kcal/mol)	ΔS (cal/K·mol)	ΔG at 298K (kcal/mol)
4-chlorophenyl	49.0 ± 0.1	-99.0	-281	-15.3
4-nitrophenyl	51.2 ± 0.3	-92.4	-259	-15.2
4-cyanophenyl	50.3 ± 0.4	-105	-299	-15.9
4-methoxyphenyl	47.6 ± 0.4	-94.6	-269	-14.4
13-mer control	57.9 ± 0.4	-110	-307	-18.5
14-mer control	58.7 ± 0.4	-113	-316	-18.8

*The values of ΔH and ΔS are averaged, and the values of ΔG are calculated according to the equation $\Delta G = \Delta H - T\Delta S$.

Table 2.3 Thermal denaturation of the complex II containing carboxamide residues paired with an abasic residue

DNA	T_m (°C)	ΔH (kcal/mol)	ΔS (cal/K·mol)	ΔG at 298K (kcal/mol)
4-chlorophenyl	49.9 ±0.4	-102	-290	-15.6
4-nitrophenyl	52.5 ±0.1	-95.3	-267	-15.7
4-cyanophenyl	52.6 ±0.8	-105	-298	-16.2
4-methoxyphenyl	47.3 ±0.5	-96.6	-275	-14.7
13-mer control	57.9 ± 0.4	-110	-307	-18.5
14-mer control	58.7 ± 0.4	-113	-316	-18.8

*The values of ΔH and ΔS are averaged, and the values of ΔG are calculated according to the equation $\Delta G = \Delta H - T\Delta S$ at temperature of 298 K.

Table 2.4 Thermal denaturation of the bulged complex III containing carboxamide residues

DNA	T_m (°C)	ΔH (kcal/mol)	ΔS (cal/K·mol)	ΔG at 298K (kcal/mol)
4-chlorophenyl	55.5 ±0.4	-100	-279	-16.9
4-nitrophenyl	56.4 ±0.2	-94.4	-260	-16.9
4-cyanophenyl	55.5 ±0.8	-111	-312	-18.0
4-methoxyphenyl	53.0 ±0.2	-102	-287	-16.5
13-mer control	57.9 ± 0.4	-110	-307	-18.5
14-mer control	58.7 ± 0.4	-113	-316	-18.8

*The values of ΔH and ΔS are averaged, and the values of ΔG are calculated according to the equation $\Delta G = \Delta H - T\Delta S$ at temperature of 298 K.

melting temperatures of all the carboxamide-containing duplexes decreased compared to the two control duplexes.

For complex I (thymine residue positioned opposite to the carboxamide residue) the melting temperatures decrease by about 8-11 ℃ compared to the 14-mer control. The stability of the 4-methoxylphenyl complex decreased the most (11.1 ℃), and the 4-nitrophenyl duplex decreased the least (7.5 ℃). It is probable that the carboxamide and thymine groups do not exist in a coplanar arrangement. There are four possibilities to resolve this steric issue: (1) the carboxamide flips out of the helix, (2) the thymine flips out of the helix, or (3) the two residues slide apart from each other, or (4) the two residues mutually intercalate. Flipping the carboxamide out of the helix is the worst option because the carboxamide is the more non-polar piece. At this time, there is nothing to distinguish hypotheses (2), (3) and (4). However, Yongjun Chu (a MS student in the Woski group) examined molecular models of a 4-fluorophenyl•thymine pair. His results suggest that the carboxamide partly intercalates between the thymine and a neighboring cytosine residue.[14] Interestingly, a similar intercalating arrangement was also suggested by Kool and coworkers[21] for a pyrene C-nucleoside when paired with natural bases in a DNA double helix. They observed that the sterically demanding pyrene only moderately destabilized the duplex ($\Delta T_m = -7$ °C), and they proposed that an interleaved arrangement might largely maintain stacking interactions. With the smaller phenylcarboxamide residues, inefficient stacking would tend to destabilize the duplex more, which fit the observations.

For complex II, an abasic residue is positioned opposite to the carboxamide residue (Figure 2.3). The range of melting temperatures is similar to that of seen for complex I: the melting temperature decreased by about 6 ℃ for 4-nitrophenyl and 4-cyanophenyl (the least destabilized) and decreased by about 11 ℃ for the 4-methoxylphenyl complex (the most destabilized).

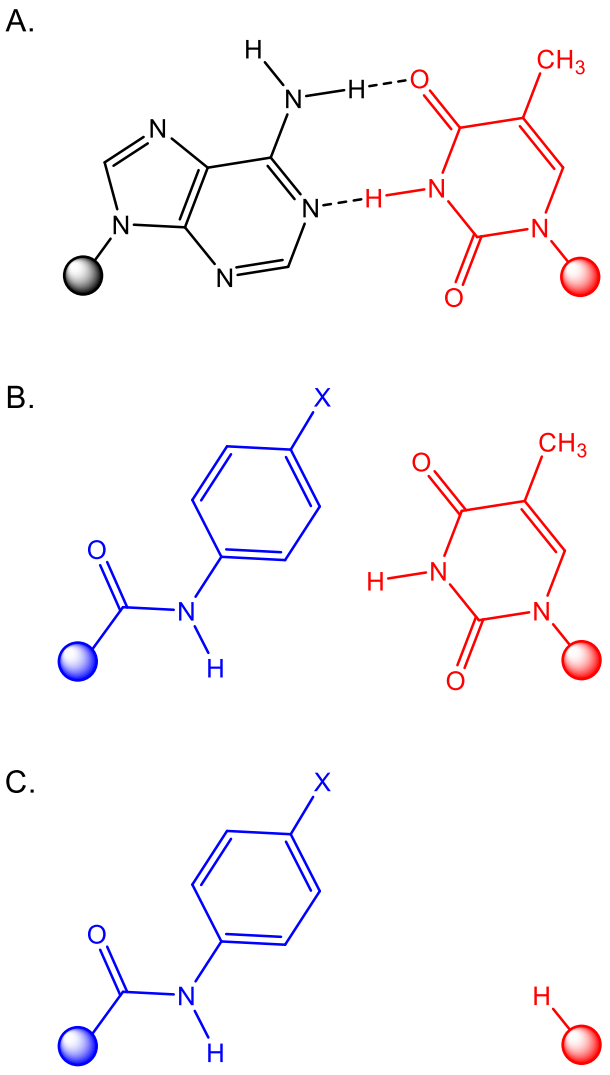


Figure 2.3 Different pairs at the modification site. A) canonical A-T pair at the modification site; B) Carboxamide-T mismatch at the modification site; C) Carboxamide-abasic pair at the modification site.

Interestingly, all of the T_m's increase slightly over that of Complex I except for the 4-methoxyphenyl carboxamide. The results for the carboxamides with electron-withdrawing substituents can be explained by invoking similar structures for the two complexes. The reduced steric demand of the abasic site may stabilize Complex II versus Complex I. Alternatively, the thymine residue may be flipped out of the helix in Complex I and is not present in Complex II. Thus, the thymine contributes nothing to stacking in either case. The modest increase in stabilities seen with loss of the base may reflect a more favorable sugar conformation in the abasic residue compared to the flipped out thymine. The fact that there appears to be an electronic factor involved in the less favorable stacking of the methoxyphenyl carboxamide supports the former hypothesis. If the main source of stabilization between Complexes II and I were the sugar conformation of the abasic residue, the electron demand of the methoxy substituent would not play a role.

For complex III, there is no residue positioned opposite to the carboxamide residue; this kind of DNA complex is the most stable one among the three types of non-natural DNA complexes examined (Figure 2.4). In all cases, the melting temperatures for the bulged complexes were significantly higher than the melting temperatures obtained from the thymine-mismatch complexes and the basic complexes, and the melting temperatures decrease only by 2-5 ℃ compared to 13-mer control. Although the insertion of a carboxamide residue into the double helix has to be sterically accommodated, the favorable intercalation of a carboxamide residue into the double-helical oligodeoxynucleotide increases the stacking interaction which almost compensates the distortion. This kind of intercalation has been observed previously and was reported by Patel and coworkers.[22] Their research indicated that the extra adenosine positioned in single residue loop without pairing partner can stack into the DNA duplex. In

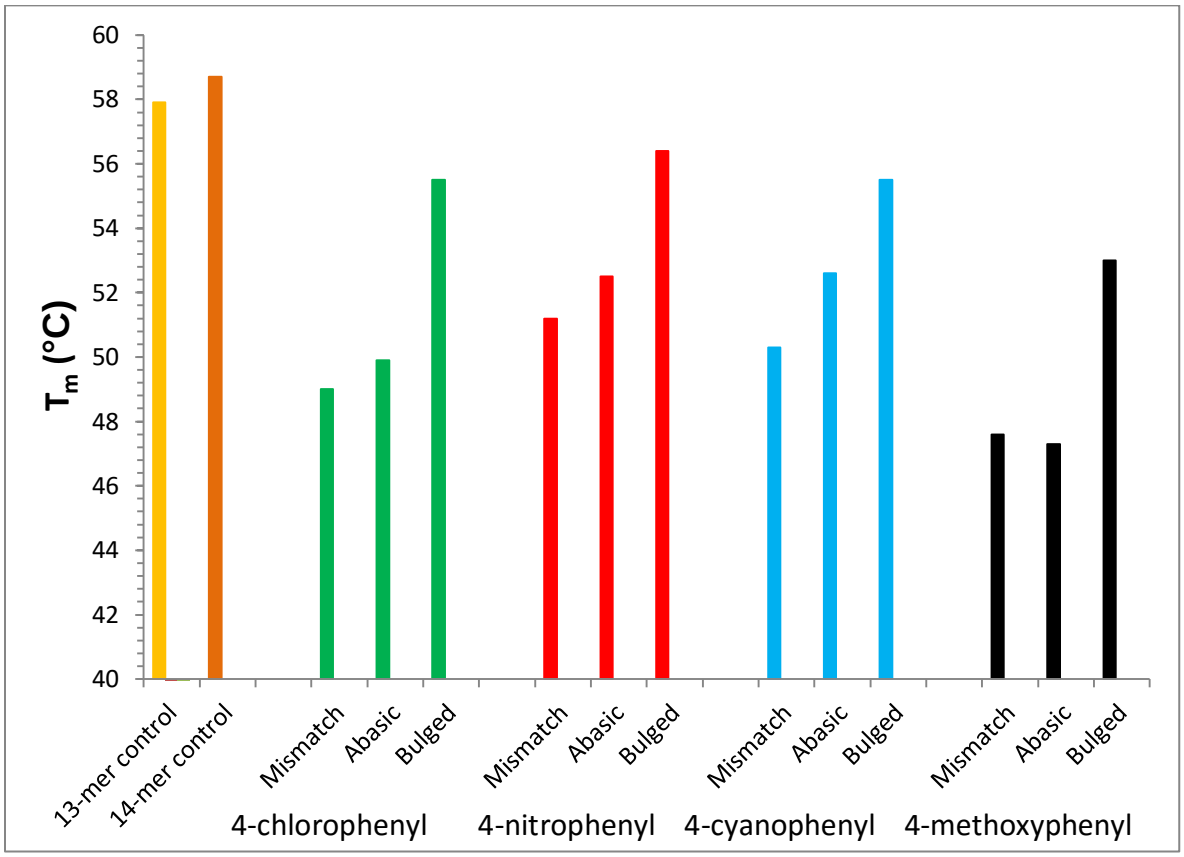


Figure 2.4 Thermal denaturation data for aryl carboxamide nucleosides. Electron-donating substituent consistently showed lowest T_m's.

addition, it has been observed that transcription factors such as TATA-box binding protein can bend DNA by partial intercalation of phenylalanine side chains in the DNA helix.[23] Besides the favorable intercalation that can stabilize the helical structure, the insertion of carboxamide residue may also provide a new hydrogen bond between the carboxamide N-H and the nucleoside O-4′ on the 3′ side of the modification position.[14] The newly formed hydrogen bond can partially overcome the distortion. Overall, the incorporation of the carboxamide as an intercalating "bulge" holds the greatest potential for future applications. Novel residues such as fluorescent and spin labels could potentially be introduced into the core of a DNA target with minimal destabilization.

Among the series of carboxamide residues, the residues containing strong electron withdrawing groups (4-nitrophenyl & 4-cyanophenyl) have the highest melting temperatures while the complexes containing the strongly electron donating 4-methoxyphenyl carboxamide have the lowest melting temperature values (Figure 2.5). A plot of melting temperature (T_m) vs residue Hammett constant (σ_p) shows the relationship between the stability of DNA complexes and the electron demand of the substiuents (Figure 2.6). The Hammett constant is a measure of the electron donating (negative σ_p) or withdrawing (positive σ_p) abilities of aryl substituents. In all three complexes, there is a good linear relationship between the increase of electron-withdrawing chararistic of the substiuent and the stability of DNA complex. These results are in agreement with the aryl stacking model constructed by Siegel and coworkers.[24] Their research has demonstrated that electron-withdrawing ability of functional groups can stabilize the stacking interactions while electron-donating functional groups can destabilize the system.

In addition to the melting temperature values, thermodynamic data also can be obtained from the thermal denaturation data.[20] The free energy change (ΔG) was calculated from ΔH and

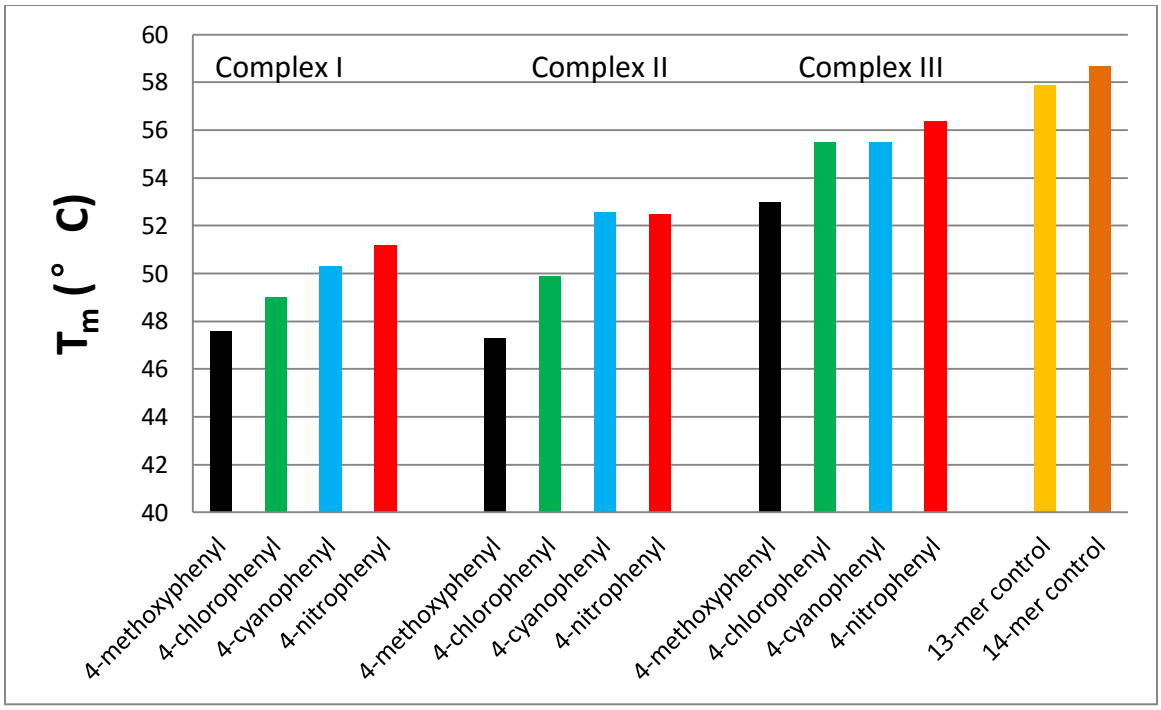


Figure 2.5 Thermal denaturation data for aryl carboxamide nucleosides. Complex I: carboxamide vs T; Complex II: carboxamide vs. abasic; Complex III: bulged carboxamide.

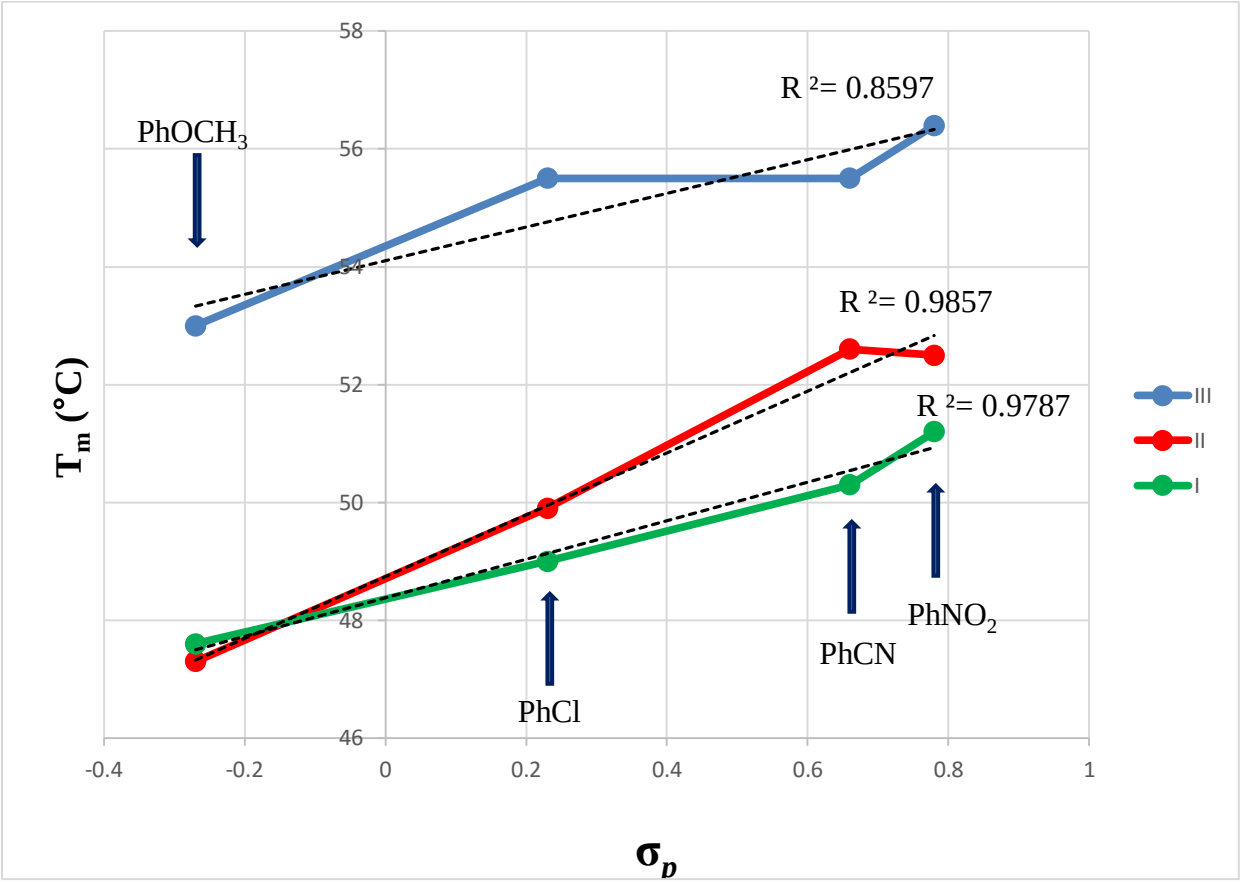


Figure 2.6 Plot of melting temperature (T_m) vs residue Hammett constant (σ_p). From left to right: 4-methoxyphenyl, 4-chlorophenyl, 4-cyanophenyl, 4-nitrophenyl. Green represents Complex I, red represents Complex II and blue represents Complex III.

ΔS at 298K. Consistent with the T_m values, the most stable complex (14-mer control) has the largest ΔG value, whereas the least stable complex (complex containing 4-methoxyphenyl) has the lowest ΔG value. Interestingly, the nitrophenyl carboxamide residues show significantly less favorable enthalpy changes and more favorable entropy changes than the other substituents in all three complexes. This suggests a difference in the stacking interaction of this residue. One possible explanation is that the nitrophenyl carboxamide is able to preorganize the single strand in a conformation favorable for duplex formation. Hybridization to the complementary DNA would then be more entropically favorable. If a favorable stacking interaction is responsible for the preorganization, there would be less of a gain in stacking upon duplex formation, reducing the ΔH of complex formation.

2.3 Conclusion

Deoxyribonucleoside 1′-carboxamides are being developed as a general method for the introduction of novel functional and reporter groups into oligodeoxynucleotides. The design strategy maintains the structure of phosphate groups and 2′-deoxyribofuranosyl ring system, and incorporates nucleobase replacements via an amide linkage at a 1′-carboxylic acid moiety.

The hybridization of the non-natural DNAs with complementary strands where the non-natural residues were paired with a natural nucleobase (a thymine) or with a stable abasic residue was studied. In addition, the hybridization where no complementary residue is used (a formal “bulge”) was also examined. The results of thermal denaturation experiments indicate that variant substitutions can be incorporated into oligonucleotide duplexes. Notably, the “bulged” complexes were most stable, suggesting the intercalation of the non-natural residues into the

DNA helix. This indicates that the carboxamide motif may be a general method for the insertion of non-natural residues into DNA for applications such as spectroscopic probes.

2.4 Experimental

The natural DNA monomers were purchased from Glen Research. Natural oligodeoxynucleotides, non-natural oligodeoxynucleotides containing a stable abasic residue and non-natural oligodeoxynucleotides containing carboxamide residues were all synthesized on an Applied Biosystems 391 DNA synthesizer. All oligonucleotides were prepared in a 0.2 μmol scale with the application of traditional phosphoramidite chemistry.

Standard DNA synthesis cycles and standard base protection groups (adenosine is protected by a benzoyl group, cytosine is protected by an acetyl group, guanosine is protected by an isobutyryl group and thymidine does not need a protective group because it does not have an exocyclic base amine) were used for the natural oligodeoxynucleotides and non-natural oligodeoxynucleotides containing an abasic residue. After the completion of synthesis, the natural oligodeoxynucleotides and oligodeoxynucleotides containing an abasic residue were treated with about 2 mL of concentrated ammonium hydroxide at 55 ℃ overnight to cleave from the controlled-pore glass (CPG) support and remove the protective groups. Ammonium hydroxide was removed by uncapping the tube containing the oligodeoxynucleotides. The crude oligodeoxynucleotides can be dried and obtained by centrifugal evaporation.

For the non-natural oligodeoxynucleotides containing carboxamide residues, several modifications were made in the program of standard DNA synthesis cycles to maximize the coupling efficiencies of the carboxamide monomers, including the coupling step for the carboxamide residues were performed twice and the coupling times were extended to 30 seconds.

In addition, non-standard base protecting groups were used for adenosine and guanosine residues that phenoxyacetyl group was used for adenosine and isopropylphenoxyacetyl was used for guanosine. Non-natural oligodeoxynucleotides containing carboxamide residues were cleaved and deprotected by treatment with 0.05 M K_2CO_3 in methanol at room temperature for about 5 hours. The crude oligodeoxynucleotides can be dried and obtained by centrifugal evaporation.

All the oligonucleotides were purified by electrophoresis through a denaturing polyacrylamide gel. The 20% acrylamide solution (19:1 acrylamide/bisacrylamide) with 7.5 M urea in 1× TBE buffer (pH 8.3, 0.089 M tris, 0.089 M borate and 0.002 M EDTA) was prepared by mixing 500 mL of 40% acrylamide solution, one packet of TBE buffer mix, about 450 g of urea and appropriate ddI water to bring up the final volume to 1 L. Then, the solution was filtered through 0.45 μm nylon membrane. The gel was prepared by addition of TEMED and 10% aq ammonium persulfate to the 20% acrylamide solution, followed by pouring the solution between two glass plates fitted with side spacers and taped along three sides to prevent leakage. A comb is inserted into the top of the gel to provide wells for addition of oligonucleotide solutions. After polymerization is complete, the comb and tape are removed, and the gel is set into a vertical stand. TBE buffer solution (1×) is added to the top and bottom chambers to provide electrical contact between the electrodes and the gel. The positive lead to a high voltage power supply is connected to the bottom electrode and the negative to the top. Running the apparatus at a current of 35 mA for about 30 minutes served to warm the gels, which improves resolution.

The crude oligodeoxynucleotide was dissolved in about 150 μL of loading buffer containing 98% formamide and 2% 0.5 M EDTA. Then, the DNA solution and loading dye mixture containing xylene cyanol FF and bromophenol blue were loaded onto the gel, followed by

running the gel at constant current of 30 mA until the bromophenol blue dye migrated close to the bottom of the gel. The power supply was disconnected, and the electrophoresis apparatus was dismantled. The gel was removed from the gel stands and wrapped in saran wrap. The full length DNA should migrate slowest because it is the longest length species in the mixture. DNA-containing bands were visualized by UV shadowing over a silica TLC plate with a fluorescent indicator, and the appropriate band was excised using a clean razor blade.

The oligodeoxynucleotides were recovered by electroelution in 1× TBE buffer at 200 volts for 15 minutes. The voltage was removed, and the DNA-containing solution was removed from the capture chamber. This procedure was repeated four times to achieve a maximal recovery. The obtained oligodeoxynucleotide solutions were applied to a reverse phase C-18 Sep-Pack column (Waters), washed with ddI water to remove salts, and finally recovered by elution using a 40% acetonitrile aqueous solution. Lyophilization produces purified oligodeoxynucleotides as white powders.

The purified oligodeoxynucleotides were dissolved in ddI water to prepare DNA stock solutions. The concentration of DNA stock solutions was determined by measurement of UV absorbance at wavelength of 260 nm. The extinction coefficient of an oligodeoxynucleotide can be taken as the sum of the monomers approximately. The extinction coefficients for the four natural bases are: A = 15,400, C = 7,300, G = 11,700, T = 8,800 $L \bullet mol^{-1} \bullet cm^{-1}$ respectively.[9] The extinction coefficients for the four non-natural bases were determined by Dr. Cuiling Liu,[11] and the values are: monomer containing arylamine 4-methoxyphenyl = 7,544, 4-nitrophenyl= 2,721, 4-cyanophenyl= 20,695 and 4-chlorophenyl= 7,302 $L \bullet mol^{-1} \bullet cm^{-1}$, respectively.

The thermal denaturation experiments were performed using a Beckman DU 800 UV/Vis spectrophotometer which is fitted with a T_m microcell holder (Peltier) and a computer-driven

temperature controller. Sample solutions for thermal denaturation experiments were comprised of two complementary oligonucleotides in 1X PES buffer (10 mM sodium phosphate, 0.1 mM EDTA and 0.1 mM NaCl at pH = 7) with concentration of each oligodeoxynucleotide was 4 μM. Solutions were heated at 90 ℃ for about 5 minutes, then were slowly cooled down to room temperature to allow duplex structures to anneal. The melting temperature studies were performed in T_m cuvettes with a stopper. Moreover, the sample solutions in the cuvettes were covered by light mineral oil to reduce evaporation of sample solutions during the experiments. Absorbance was monitored at 260 nm, and the temperature in each experiment was increased from 25 ℃ to 75 ℃ at a rate of 0.4 ℃/min with a read interval of 0.2 ℃. Data were exported from the DU 800 system and were analyzed using Meltwin 3.5 software.[20]

2.5 References:

1. Mao, C.; Sun, W.; Seeman, N. C. *Nature* **1997**, *386*, 137.

2. Roelfes, G.; Feringa, B. L. *Angew. Chem. Int. Ed.* **2005**, *4*, 3230.

3. Roelfes, G; Boersma, A. J.; Feringa, B. L. *Chem. Commun.* **2006**, 635.

4. Uhkmann, E.; Peyman, A. *Chem. Rev.* **1990**, *90*, 544.

5. Varma, R. S. *Synlett.* **1993**, 621.

6. Bergstrom, D. E.; Zhang, P. and Johnson, W. T. *Nucleic Acids Res*. **1997**, *25*, 1935.

7. Loakes, D.; Brown, D. M. *Nucleic Acids Res*. **1994**, *22*, 4039.

8. Humber, D. C.; Mulholland, K. R.; Stoodley, R. J. *J. Chem. Soc. Perkin Trans. I.* **1990**, 283.

9. Millican, T. A.; Mock, G. A.; Chauncey, M. A.; Patel, T. P.; Eaton, M. A. W.; Gunning, J.; Gutbush, S. D.; Neidle, S.; Mann, J. *Nucleic Acids Res.* **1984**, *12*, 7435.

10. Wichai, U.; Woski, S. A. *Bioorg. Med. Chem. Lett.* **1998**, *8*, 3465.

11. Wichai, U.; Woski, S. A. *Org. Lett*. **1999**, *1*, 1173.

12. Matsuda. S.; Leconte. A. M.; Romesberg. F. E. *J. AM. CHEM. SOC.* **2007**, *129*, 5551.

13. (a) Ren R. X.-F ; Chaudhuri, N. C.; Paris, P. L.; Rumney, S.; Kool, E. T. *J. Am. Chem. Soc.* **1993**, *118*, 7671; (b) Chaudhuri, N. C.; Kool, E. T. *Tetrahedron Lett.* **1995**, *36*, 1795; (c) Schweitzer, B. A.; Kool, E. T. *J. Org. Chem.* **1994**, *59*, 7238; (d) Chaudhuri, N. C.; Ren R. X.-F ; Kool, E. T. *Synlett* **1997**, 341.

14. Chu, Y. "Synthesis and characterization of non-natural DNA with novel nucleosides", M.S. Thesis, University of Alabama, **2002**.

15. Liu, C. "Investigation of the stabilities of PNA&DNA and DNA&DNA complex containing novel aromatic residues", Ph.D. Thesis, University of Alabama, **2006**.

16. Pedersen, C. *Carbohydrate Res.* **1999**, *315*, 192.

17. Hart, T. W.; Metcalfe, D. A.; Scheinmann, F. *J. Chem. Soc. Chem. Comm.* **1979**, 156.

18. Blad, E.; Saigo, K.; Mukaiyama, T. *Chem. Lett.***1975**, 1163.

19. Hakimelahi,G.; Proba, Z. A.; Oglivie, K. K. *Can. J. Chem.* **1982**, *60*, 1106.

20. McDowell, J. A.; Turner. D. H. *Biochemistry* **1996**, *35*, 14077.

21. Matray, T.; Kool, E. T. *J. Chem. Soc.* **1998**, *120*, 6191.

22. Kalnik, M. W.; Norman, D. G.; Swann P. F.; Patel, D. J. *J. Biol. Chem.* **1989**, *264*, 3702.

23. Nikilov, D. B.; Chen, H.; Halay, E. D.; Hoffmann, A.; Roeder, R. G.; Burley, S. K. *Proc. Natl. Acad. Sci.* **1996**, *93*, 4862.

24. Cozzi, F.; Siegel, J. S. *Pure Appl. Chem.* **1995***, 67*, 683.

CHAPTER III

Synthesis of Novel Aryl C-Nucleoside

3.1 Introduction

Noncovalent interactions including hydrogen bonding and base stacking play a significant role in the function and structure of DNA duplex. Hydrogen bonding is considered to be an important factor for the hybridization specificity and stability of DNA duplex; meanwhile, the π-π stacking interactions are crucial for the stability of DNA duplexes. Modification of the natural bases of DNA can facilitate the understanding of how these factors affect the stability of DNA duplex. These base modifications can be achieved by changing the hydrogen bonding, the shape and the size of the aromatic system, or replacing the entire DNA bases. For example, the preparation of hydrophobic isosteres of the natural bases has been reported by Kool and co-workers.[1] These aromatic base replacements were designed to sterically mimic the natural bases without using hydrogen bonding oxygen and nitrogen atoms (Figure 3.1).[1] Modeling the 6-membered pyrimidine rings posed a problem because the presence of a nitrogen able to form the usual N—C glycosidic bond to deoxyribose would either possess a positive charge (like the pyridinium ring in NAD^+) or aromaticity would have to be lost. The third alternative is to replace the nitrogen with a carbon atom, producing an aryl C-nucleoside.

Aryl C-nucleosides have aromatic residues attached to the deoxyribose moieties through a carbon-carbon connection instead of the original carbon-nitrogen bond in the natural nucleosides. Compared with the natural DNA bases, the aryl C-nucleoside has some unique properties. First,

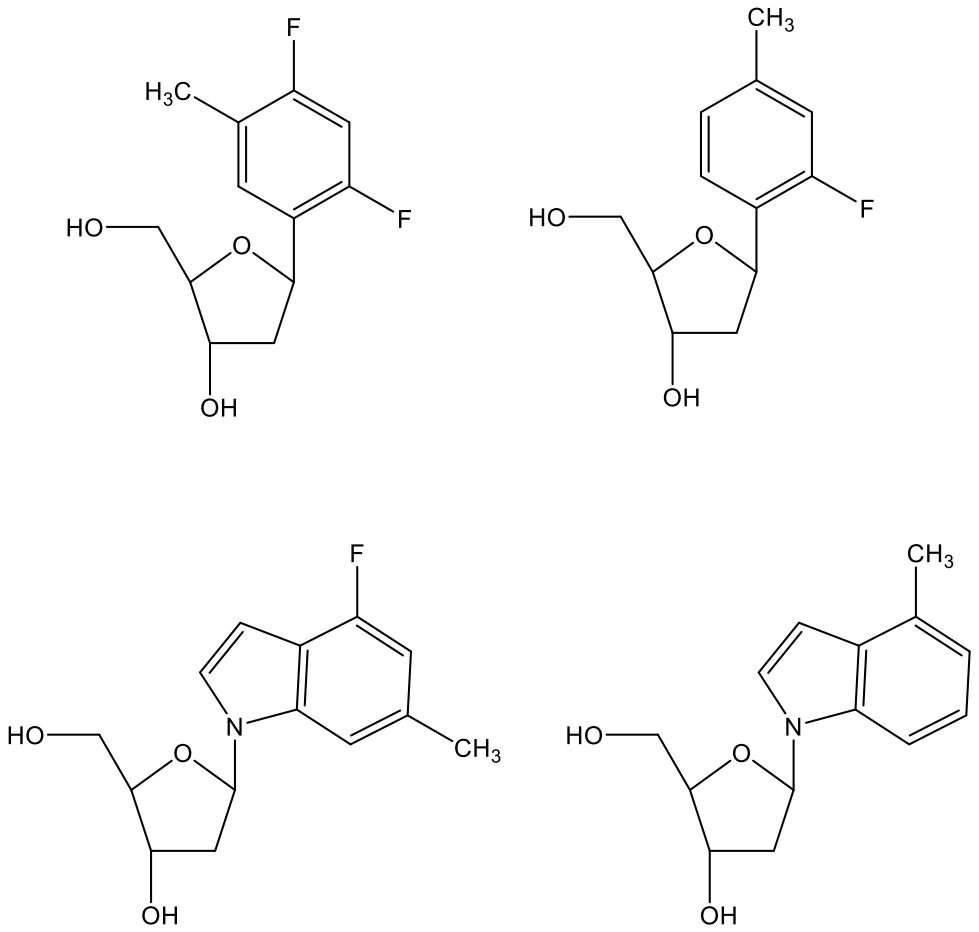


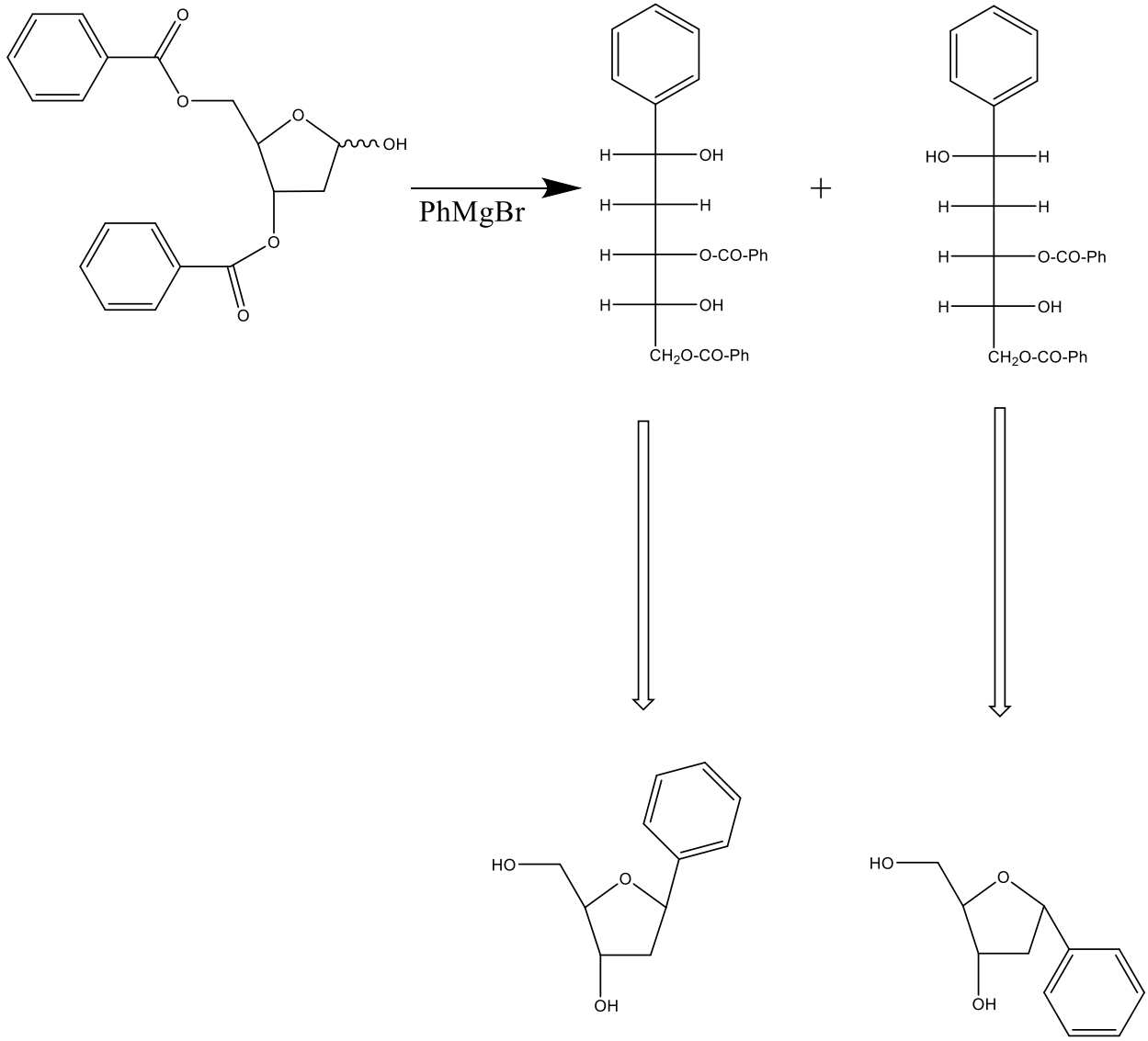
Figure 3.1 Structures of Kool's hydrophobic isosteres of the natural bases.[1]

due to the C-C bond in aryl C-nucleoside, these compounds lack the anomeric effect at the C1′ position which renders them stable to loss of the base under strongly acid conditions and resistant to the degradation by enzymes. Second, many of the aryl C-nucleosides are incapable of forming hydrogen bonds, which means they should be relatively non-selective in pairing with the four natural DNA bases.

As a result of these special characteristics, aryl C-nucleosides have numerous applications in the field of DNA research. Some of the aryl C-nucleosides have shown significant biological properties such as antibacterial, antiviral and antitumorigenic activities.[2, 3] C-nucleosides containing highly conjugated functional groups can be used as fluorescent probes to study the structure and dynamic properties of nucleic acids or the nucleic acid-protein complexes.[4] Finally, the modified aryl C-nucleoside can be applied to studying the stability and internal interaction of DNA complexes.[1] In order to examine the structural effects of the introduction of aryl C-nucleosides into DNA double helical structures, it was necessary to prepare the protected phosphoramidite reagents of the targeted C-nucleosides.

There are several strategies for the synthesis of aryl C-nucleosides. The most common method for the creation of the C-C bond is to attack the sugar C1′ site using the nucleophilic functional groups, and organometallic species are frequently used reagents for this purpose. For example, Millican and coworkers[5] described a full preparation of 1,2-dideoxy-1-β-phenyl-D-ribofuranose where the key C—C bond was formed by reaction of phenylmagnesium bromide with the open chain aldehyde of 2-deoxyribose (Scheme 3.1). Subsequent ring closure resulted in a mixture of isomers at C1′; separation was achieved by multiple recrystallatizations with the expected loss in yield.

Scheme 3.1[5]

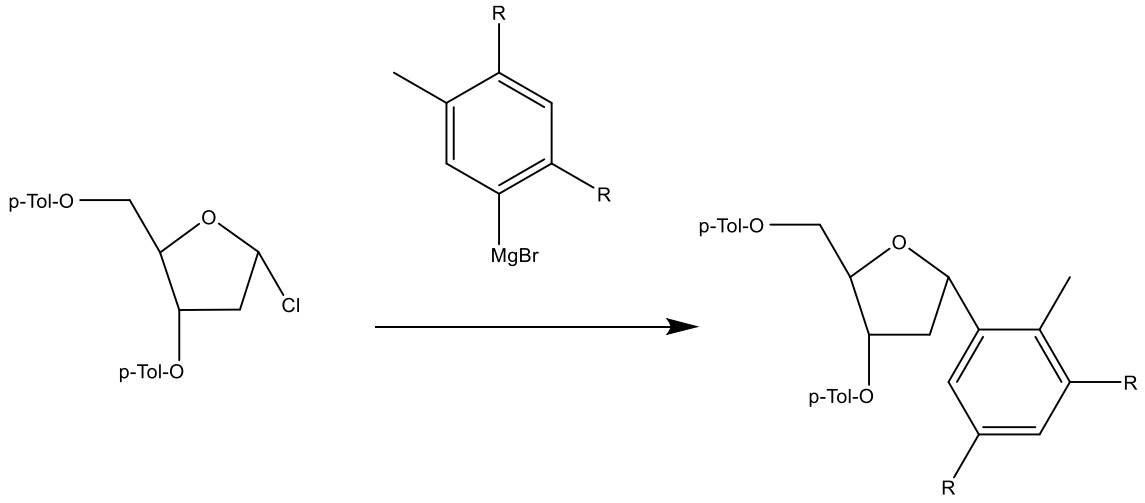

O
O
O
OH
O
O
PhMgBr
H
OH
H
H
H
O-CO-Ph
H
OH
CH2O-CO-Ph
+
HO
H
H
H
H
O-CO-Ph
H
OH
CH2O-CO-Ph
HO
O
OH
HO
O
OH

In Kool's research,[1c] Grignard reagents were used as nucleophiles in substitutions of a protected 1-α-chloro-2-deoxyribofuranose (Scheme 3.2). Unfortunately, the predominant product was the aryl C-nucleoside α-anomers (unfortunately, until a crystal structure was solved, these were mistakenly assigned as the β-anomers). The nucleophilic attack was also performed on the glycosyl halides using arylmagnesium, arylzinc or arylcadmium reagents, but the produced product was a mixture of isomers (Scheme 3.3).[1a, 1b, 1d] Treatment with benzenesulfonic acid was found to equilibrate the mixture to favor the β-anomer.[1a] However, separation of the anomers can be difficult and the yields are modest.

Dr. Wichai in the Woski group used aryllithium addition/reduction reaction to prepare aryl C-nucleosides.[6, 7 ,8] There are two useful strategies mentioned in Wichai's research. The first route utilizes the addition of phenyllithium to a protected 2-deoxyribonolactone, and followed by selective reduction with triethylsilane/boron trifluoride etherate to produce the β-anomer (Scheme 3.4). Use of the disiloxane-protected 2-deoxyribonolactone as the precursor in these reactions somewhat improved the yields of the reactions at the cost of stereoselectivity (β:α = 10:1, Scheme 3.5).[7] The application of this methodology to the synthesis of numerous aryl C-nucleosides is hindered by an extreme sensitivity of the reaction yields to substitution on the aryllithium reagent.[7] Also, no substitute to the use of aryllithium reagents was found, so sensitive functional groups cannot be present on either reagent.

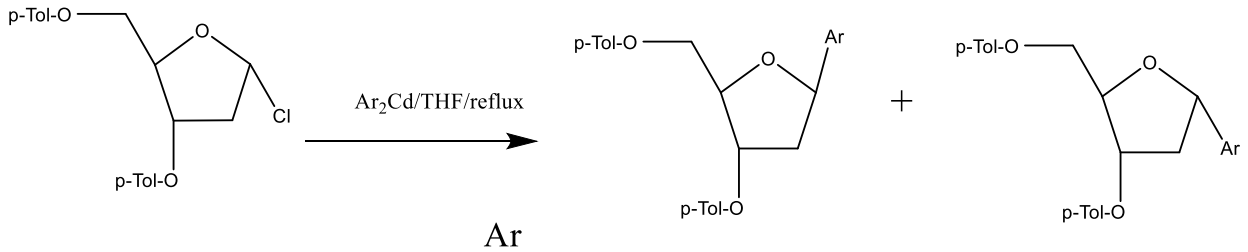
As discussed above, these synthesis routes are limited by the problematic control of the sterechemical outcome of the addition reaction or the limited functional groups that can be used on the substrates. The synthetic approach used in this dissertation for the preparation of aryl C-nucleoside was first developed by Daves and co-workers[9] and utilizes a palladium-catalyzed

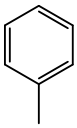
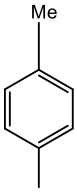
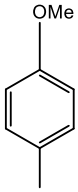
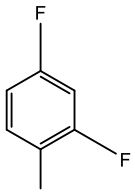
Scheme 3.2[1c]



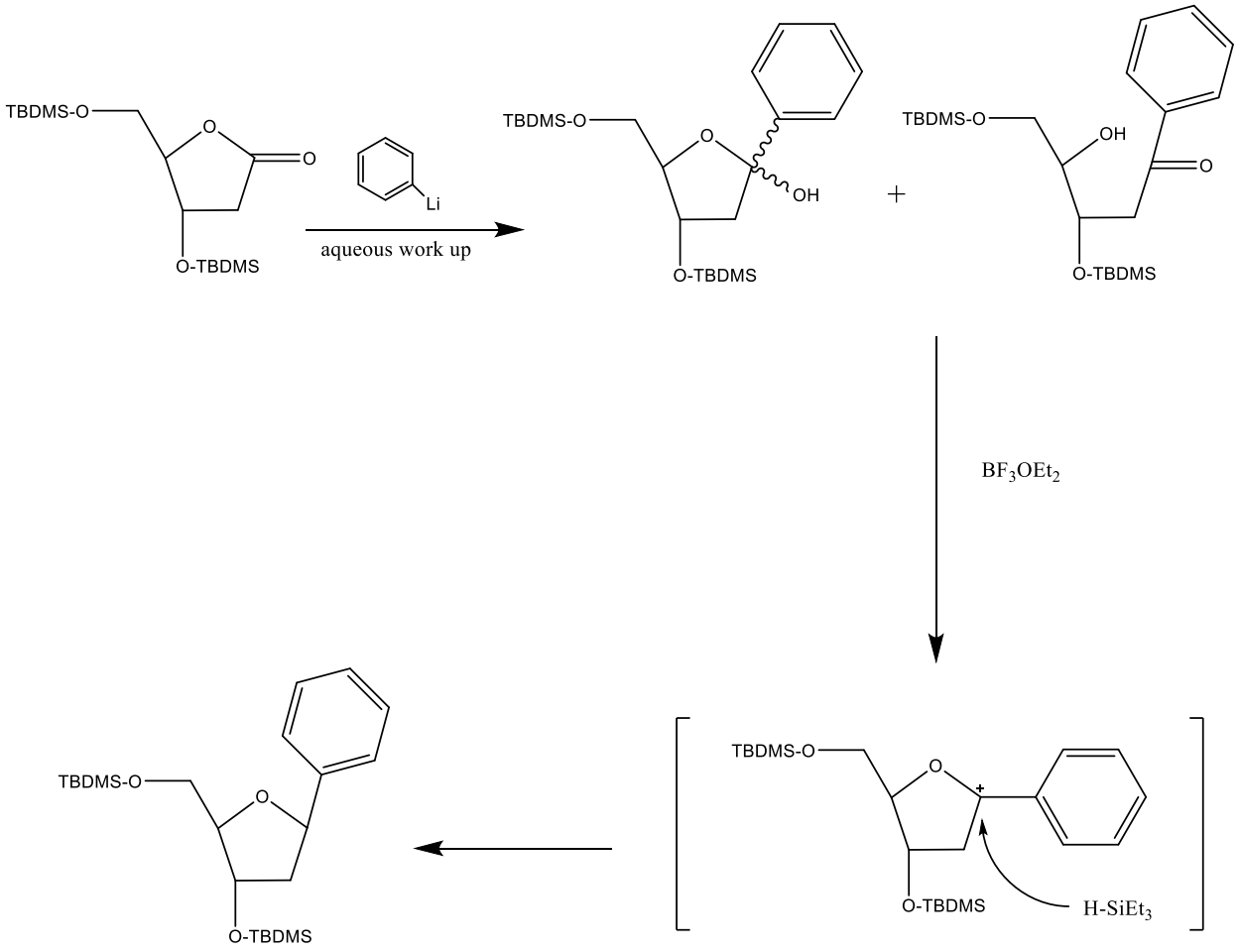
R = F or CH_3

Scheme 3.3[1b]

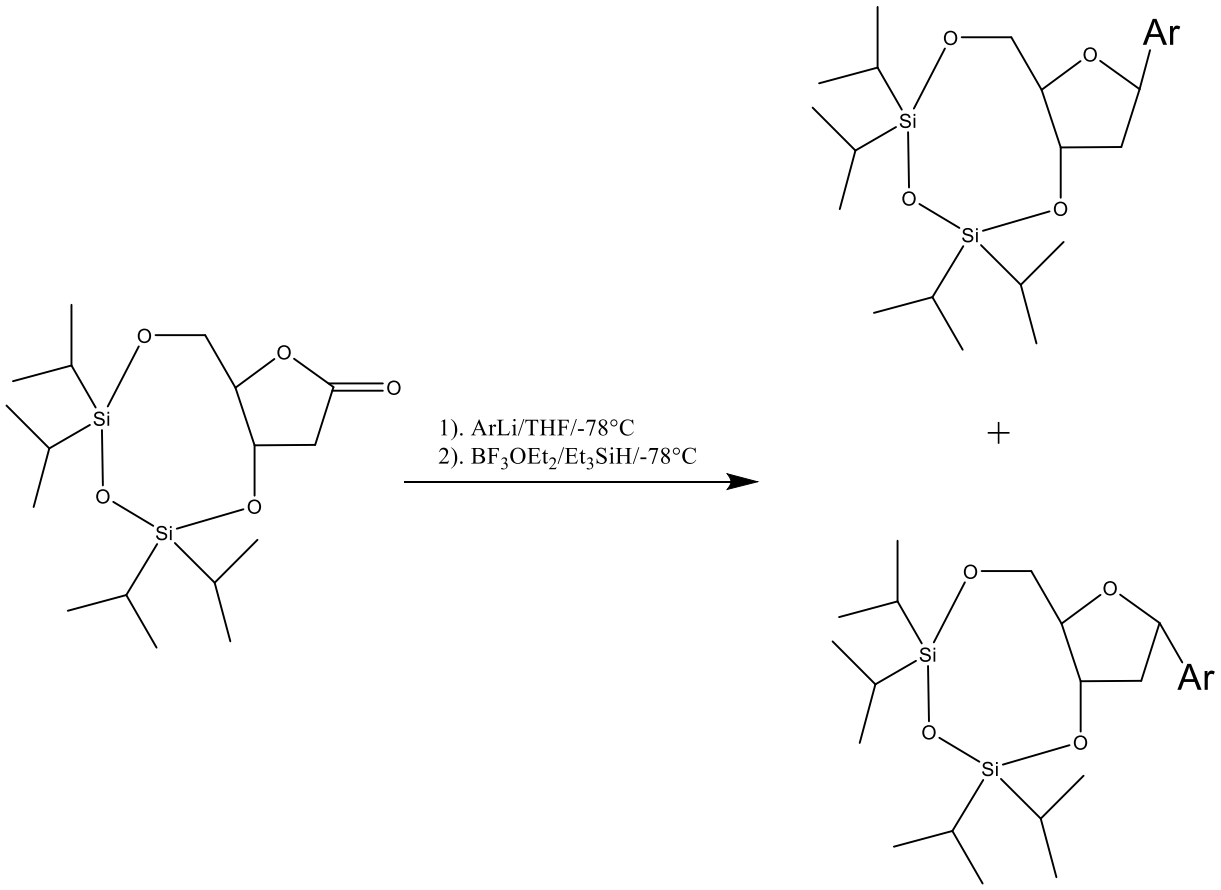


Ar			
a		80%	20%
b		74%	26%
c		83%	17%
d		79%	21%

Scheme 3.4[6]


TBDMS-O
O
O
O-TBDMS
Li
aqueous work up
TBDMS-O
O
OH
O-TBDMS
+
TBDMS-O
OH
O
O-TBDMS
BF3OEt2
TBDMS-O
O
+
O-TBDMS
H-SiEt3
TBDMS-O
O
O-TBDMS

Scheme 3.5[7]



Heck coupling of *para*-substituted iodobenzenes to furanoid glycals (Scheme 3.6, a detailed scheme is shown in results and discussion). This reaction is stereospecific and only the β-anomer can be produced. In addition, Romesberg and coworkers[10] reported that high yields of around 80% can be obtained from this reaction. The synthesis of aryl C-nucleosides ready to be incorporated into synthetic oligodeoxynucleotide strands requires a seven-step eight-reaction sequence.

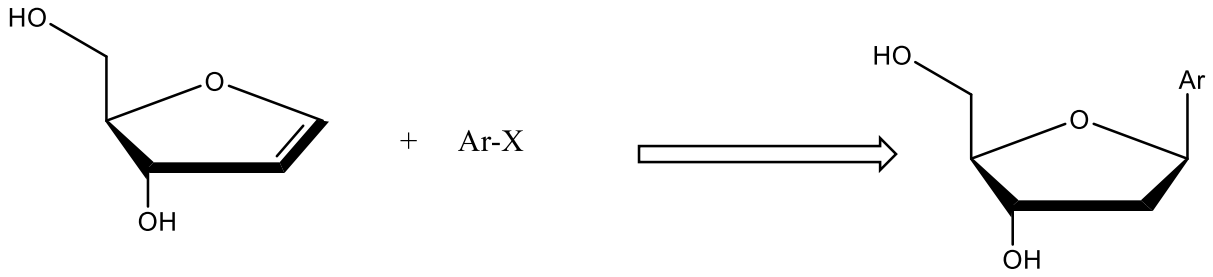
3.2 Results and Discussion

Synthesis of novel DNA monomers

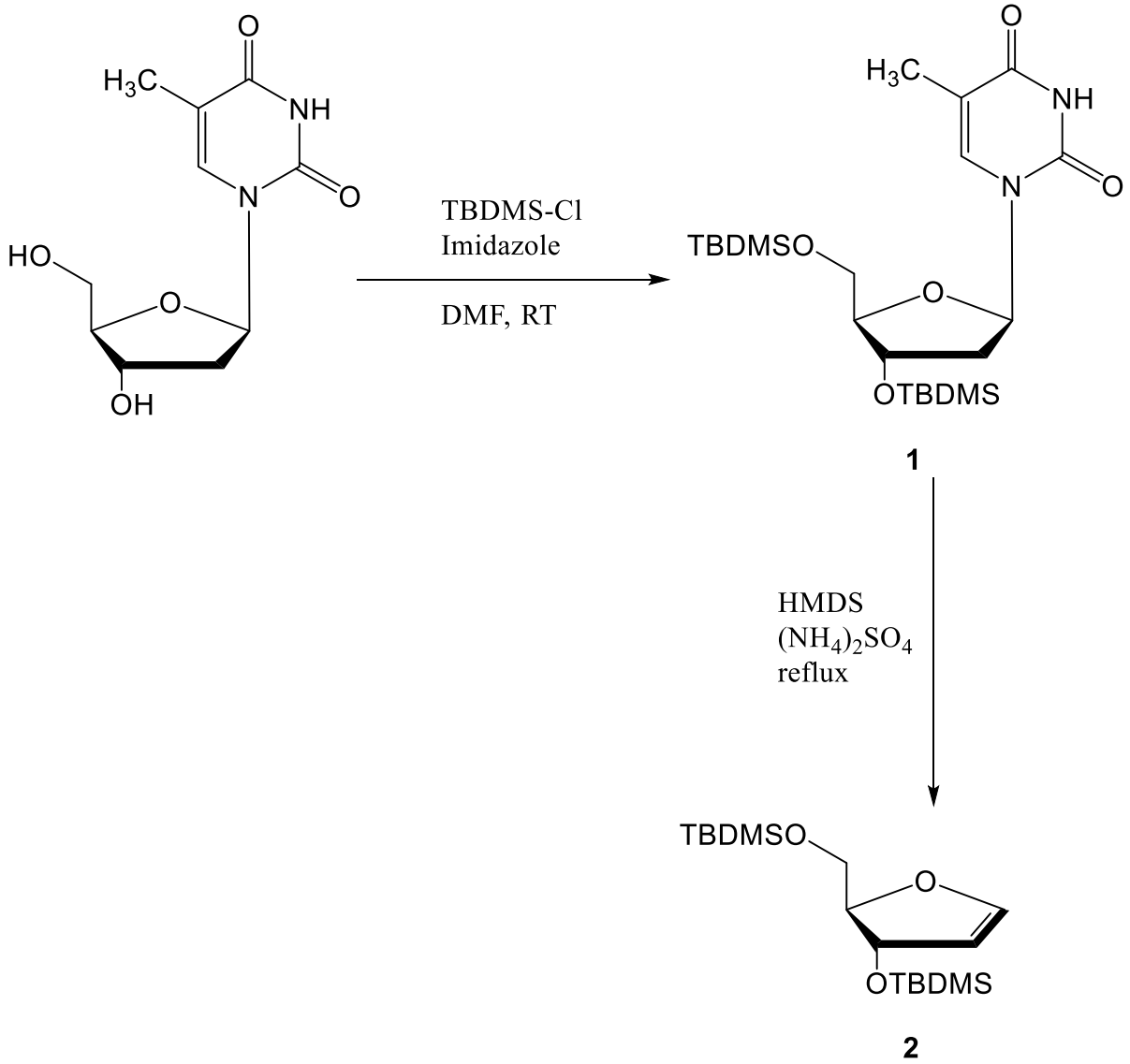
The synthesis scheme starts with the synthesis of 3′-and 5′-silyl protected furanoid glycal from thymidine (Scheme 3.7). First, the hydroxyl functional groups of thymidine were protected by reacting with *tert*-butyldimethylsilyl chloride (TBDMS-Cl) in the presence of imidazole in DMF.[11] The product, 3′,5′-bis-*O*-(*tert*-butyldimethylsilyl) thymidine (**1**), was then depyrimidinated by refluxing in excess hexamethyldisilazane (HMDS) and ammonium sulfate. With loss of the thymine base, a double bond was formed between C'1 and C'2. The fully protected furanoid glycal **2** was produced in a good yield.[12]

The original thought was to use glycal **2** for the Heck reaction to produce the keto C-ribosides (Scheme 3.8). There are literature precedents for the use of this glycal in such reactions and the reported result showed this reaction should be stereospecific.[10] However, in my hands, this coupling reaction produced a mixture of isomers at C1′ and the ratio between α-product and β-product is around 1:1 based on the NMR analysis. In order to maintain the native β (up) orientation of the nucleobases, the aryl group should also be in the β configuration. Various method were tried to separate the α and β products, but there was no effective way to completely separate the β-product from the α-product.

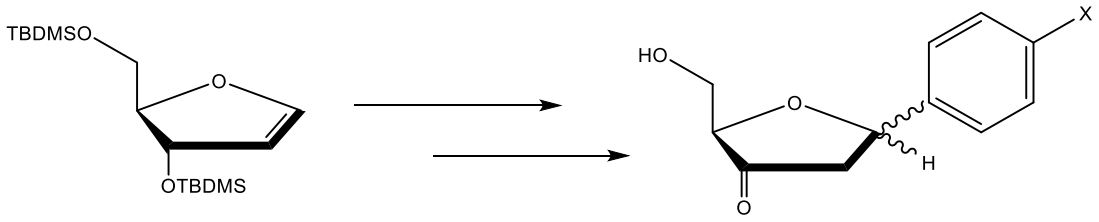
Scheme 3.6



Scheme 3.7


O
H3C
NH
N
O
HO
O
OH
TBDMS-Cl
Imidazole
DMF, RT
TBDMSO
OTBDMS
1
HMDS
(NH4)2SO4
reflux
2

Scheme 3.8



2

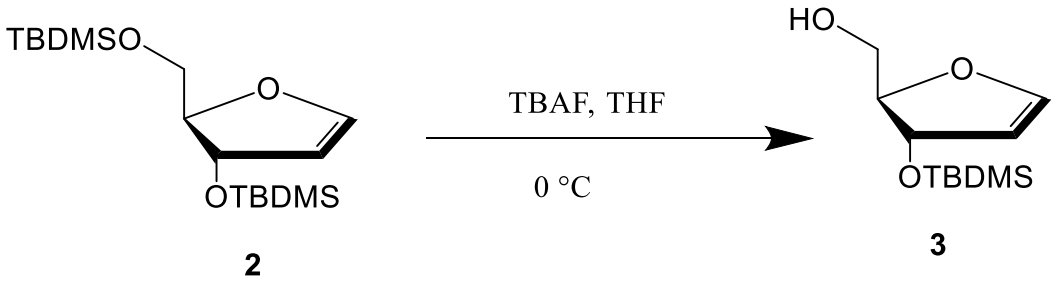
X can be H, F or methyl

In order to improve the stereoselectivity of the Heck reaction, it was decided to use the glycal with a free 5′-hydroxyl. The selective desilylation of the primary silyl ether of **2** was accomplished successfully by treatment with one equivalent of tetra-*n*-butylammonium fluoride in THF at 0 ℃ for 2 hours (Scheme 3.9).[13] The free 5′-hydroxyl resulted in a stereospecific Heck reaction, and only the desired β-product was obtained. The different stereochemical outcome may cause by steric effects: removing the bulky 5′-TBDMS protective group, the 4-substituted iodobenzenes can attack the double bond from above more easily. Alternatively, the free hydroxyl group may direct the palladium attack from the β-face by coordination of the metal center.

With the furanoid glycal **3** in hand, the synthesis of 2′-deoxy C-ribosides started with the coupling reaction of furanoid glycal and *para*-substituted iodobenzenes by using the palladium-mediated method (Scheme 3.10).[14] The optimal reaction condition was 1 equivalent of glycal, 1 equivalent of *para*-substituted iodobenzene, 10 mol% of palladium acetate, 2 equivalents of tri-*n*-butylamine and 1 equivalent of sodium acetate in DMF. The reaction was run for 6 hours at room temperature. Without isolation, the remaining silyl ether was removed by treatment with acetic acid and tetra-*n*-butylammonium fluoride.

The yields of this coupling procedure were not high, especially for the fluoro-substituted iodobenzenes (yield between 20% and 30%). Some reaction variables such as reaction time and temperature were examined in order to optimize the yields. Increasing the reaction time from 6 hours to 18 hours or overnight, and reaction temperature from room temperature to 70 ℃ did not improve the coupling yields. The fluoro-substituted iodobenzene always produced the lowest yields, while the iodotoluene and iodobenzene resulted in higher yields. This points to an electronic effect in this reaction where electron-withdrawing groups such as fluorine reduce

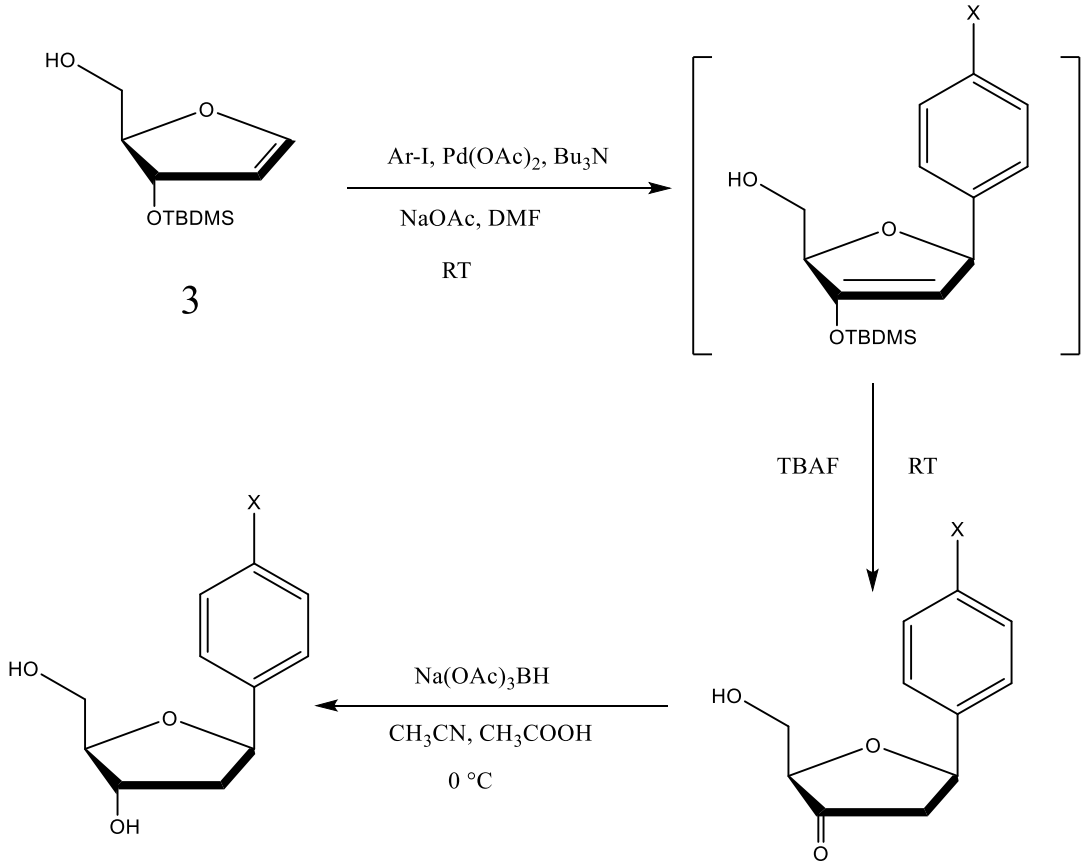
Scheme 3.9



coupling yields. But most importantly, there was no evidence of the generation of the α isomer. After the formation of 2′-deoxy-3′-keto C-ribosides (**4**), stereospecific reduction of the ketone group using sodium triacetoxyborohydride (Scheme 3.10) was performed to obtain the desired aryl C-nucleoside.[15] The reaction was run for 30 minutes at 0 ℃ using 1:1 acetonitrile/acetic acid as the solvent.

DNA monomers suitable for oligonucleotide synthesis were prepared starting from the C-nucleosides **5** (Scheme 3.11). First, protection of the 5′ -hydroxyl functional group as its dimethoxytrityl ether was accomplished by using DMTrCl in the presence of DMAP or silver nitrate.[11, 16] This reaction used pyridine as solvent and was run for 20 hours at room temperature. The DMT-ether products proved to be very sensitive to hydrolysis even in mildly acidic condition, including during the column chromatography using silica gel. In order to avoid this situation, a weak base, triethylamine (TEA), was added into the elution solvent, and the silica column was pre-equilibrated with the TEA-containing elution solvent. The highest yield (59%) for this 5'-hydroxyl protection reaction was obtained by using 5% TEA in the elution solvent. After isolation of the DMT-protected C-nucleosides, the 3′-hydroxyl functional group was ready for the phosphitylation. This step was accomplished by the reaction of DMT ether **6** and 2-cyanoethyl-*N*,*N*-diisopropylchlorophosphoramidite in the presence of TEA.[16] Dichloromethane was used as solvent and it took 1 hour at room temperature to finish the reaction. The elution solvent used for silica column chromatography also contained TEA to avoid the hydrolysis of DMT ether as discussed above. The yields of phosphitylation varied from 35% to 85%; using freshly opened 2-cyanoethyl-*N*,*N*-diisopropylchlorophosphoramidite was key for obtaining higher yields.

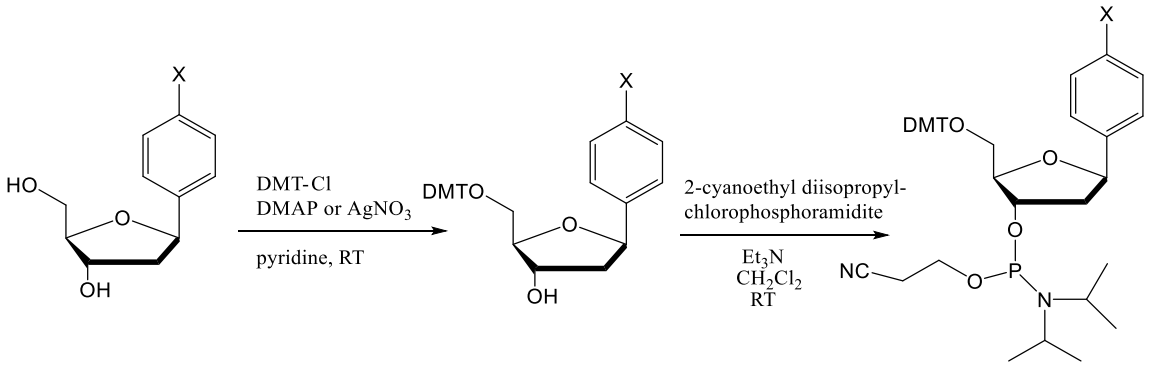
Scheme 3.10



5a: X = H
5b: X = F
5c: X = CH_3

4a: X = H
4b: X = F
4c: X = CH_3

Scheme 3.11



5a: X = H
5b: X = F
5c: X = CH_3

6a: X = H
6b: X = F
6c: X = CH_3

7a: X = H
7b: X = F
7c: X = CH_3

3.3 Conclusion

In summary, three C-nucleosides containing 4-substituted phenyl residues were successfully synthesized by following a seven-step reaction scheme. Unexpectedly, a loss of stereoselectivity occurred when performing the Heck reaction with the 3′,5′-bis-TBDMS ether. Efforts to separate the diastereomers proved unsuccessful. However, the reaction with the 3′-TBDMS 5′-hydroxyl glycal showed no signs of the unwanted α-isomer. By adding an extra deprotection step, the yield of β-aryl C-2′-deoxynucleoside increased by about 30%.

3.4 Experimental

General

All solvents and reagents were obtained from commercial sources and were used without purification. All moisture- and air-sensitive reactions were carried out under a nitrogen atmosphere. Reactions were monitored by thin-layer chromatography (TLC) using commercially available silica gel plates (MF 254 from Agela Technologies, aluminum back). After development, TLC plates were visualized under ultraviolet (UV) light. Column chromatography was performed by using silica gel 60 Å 40-63μm from Sorbent Technologies.

^{1}H NMR spectra of products from the organic synthesis were recorded at 360 MHz using a Bruker AVANCE-360 spectrometer. Chemical shifts are reported in ppm from an internal reference, tetramethylsilane (TMS). Multiplicities are denoted as s (singlet), d (doublet), t (triplet), m (multiplet), dd (doublet of doublets), ddd (doublet of doublet of doublets), br (broad). Mass spectra were recorded using a VG AutoSpec Mass Spectrometer operating in EI mode.

Synthesis of furanoid glycal

3′, 5′-Bis-O-(tert-butyldimethylsilyl) thymidine (1).[11] Thymidine (2.14 g, 8.86 mmol) and DMF (40 mL) were introduced into a dry flask under nitrogen atmosphere protection, and the mixture was stirred until the thymidine was dissolved completely. Then imidazole (1.44 g, 21.2 mmol) was added into the reaction solution followed by *tert*–butyldimethylsilyl chloride (TBDMS-Cl). After 24 h, the mixture was poured into 100 mL of water and was extracted with diethyl ether (3×60 mL). The combined organic extracts were washed with aqueous sodium bicarbonate (100 mL) and water (100 mL). The organic phase was dried over anhydrous sodium sulfate and evaporated under reduced pressure to give a crude product as clear oil or white solid. The crude product was purified by column chromatography (hexanes:ethyl acetate = 2:1) to yield a pure white solid (3.96 g, yield = 95.2%): R_f = 0.61 (2:1 hexanes/ethyl acetate); ^{1}H NMR ($CDCl_3$) δ 8.43 (s, 1H), 7.52 (d, 1H), 6.38 (dd, 1H), 4.42 (m, 1H), 3.94 (dd, 1H), 3.88 (dd, 1H), 3.78 (dd,1H), 2.31 (ddd,1H), 2.05 (m, 1H), 2.01 (d, 3H), 0.92, 0.89 (2s, 18H), 0.11, 0.08 (2s, 6H).

1,4-Anhydro-3,5-bis-*O*-(*tert*-butyldimethylsilyl)-2-deoxy-D-erythro-pent-1-enitol (2).[12]

3′,5′-Bis-*O*-(*tert*-butyldimethylsilyl)thymidine (2.11 g, 4.48 mmol) and hexamethyldisilazane (20 mL) were introduced into a dry flask under nitrogen atmosphere protection, and the mixture was stirred and heated until the solid was dissolved completely. After the addition of ammonium sulfate (1.20 g, 9.24 mmol), the solution was refluxed for 5 hours. The solution was then poured into 50 mL of water, and the mixture extracted with cyclohexane (3×50 mL). The organic extracts were combined and washed with aqueous sodium bicarbonate (50 mL) and water (50 mL). The organic phase was dried over anhydrous sodium sulfate and evaporated under reduced pressure to give the crude product as yellow oil. The crude product was purified by column

chromatography (7:1 hexanes/ether) to yield a pure light liquid (1.02 g, yield = 66.2%): R_f = 0.92 (7:1 hexanes/ether); ^{1}H NMR ($CDCl_3$) δ 6.52 (d, 1H), 5.05 (t, 1H), 4.92 (t, 1H), 4.32 (dt, 1H), 3.73 (dd,1H), 3.54 (dd,1H), 0.95,0.94 (2s, 18H), 0.14 (1s, 6H), 0.13, 0.12 (2s, 6H).

1,4-Anhydro-3-*O*-(*tert*-butyldimethylsilyl)-2-deoxy-D-erythro-pent-1-enitol (3).[13] 1,4-Anhydro-3,5-bis-*O*-(*tert*-butyldimethylsilyl)-2-deoxy-D-erythro-pent-1-enitol (0.57 g, 1.66 mmol) and THF (15 mL) were added into a dry flask under nitrogen atmosphere protection. The solution was cooled to 0 ℃ using an ice bath. Then tetra-*n*-butylammonium fluoride (1M solution in THF, 1.6 mL) was added dropwise via syringe. The reaction solution was stirred at 0 ℃ for 2 h. After the completion of reaction, the solvent was evaporated under reduced pressure to give a crude product as yellow oil. The crude product was purified by column chromatography (95:5 dichloromethane/methanol) to yield the pure product as a light yellow oil (0.35 g, yield = 93%): R_f = 0.75 (95:5 dichloromethane/methanol); ^{1}H NMR ($CDCl_3$) δ 6.52 (d, 1H), 5.08 (t, 1H), 4.84 (t, 1H), 4.37 (dt, 1H), 3.71 (dd,1H), 3.63 (dd,1H), 2.31 (br, 1H), 0.92 (1s, 9H), 0.11(1s, 6H).

Synthesis of 2'-deoxy-3'-keto C-ribosides

(2R, 5R)-Dihydro-2-(hydroxymethyl)-5-phenylfuran-3(2H)-one (4a).[14] Furanoid glycal **3** (0.20 g, 0.88 mmol), iodobenzene (0.179 g, 0.88 mmol), sodium acetate (68 mg, 0.88 mmol), tributylamine (44 μL, 0.17 mmol) and palladium acetate (20. mg, 0.088 mmol) were dissolved in 10 mL of DMF in a dry flask under nitrogen atmosphere protection. The mixture was stirred for 6 hours at room temperature. Then acetic acid (100 μL, 1.65 mmol) and tetra-*n*-butylammonium fluoride (1M solution in THF, 0.88 mL, 0.88 mmol) were added. The reaction mixture was stirred for an additional 30 minutes. Then the mixture solution was poured into 50 mL of water

and extracted with 1:1 ether/ethyl acetate (3×50 mL). The organic extracts were combined and washed with 50 mL of aqueous sodium bicarbonate and 50 mL of water. The organic phase was dried over anhydrous sodium sulfate and evaporated under reduced pressure to give a crude product as dark orange oil. The crude product was purified by column chromatography (1:2 hexanes/ethyl acetate) to yield the pure product as light yellow oil (0.0842 g, yield = 49.8%): R_f = 0.82 (1:2 hexanes/ethyl acetate); ^{1}H NMR ($CDCl_3$) δ 7.42 (m, 5H), 5.22 (dd, 1H), 4.08 (t, 1H), 3.98 (d, 2H), 2.91 (dd,1H), 2.58 (dd,1H).

(2R,5R)-5-(4-Fluorophenyl)-dihydro-2-(hydroxymethyl)furan-3(2H)-one(4b).[14] Furanoid glycal **3** (0.20 g, 0.88 mmol), 1-fluoro-4-iodobenzene (0.195 g, 0.88 mmol), sodium acetate (68 mg, 0.88mmol), tributylamine (44 μL, 0.172 mmol) and palladium acetate (20 mg, 0.088mmol) were dissolved in 10 mL of DMF in a dry flask under nitrogen atmosphere protection. The mixture was reacted for 6 hours at room temperature. Then acetic acid (100 μL, 1.65 mmol) and tetra-*n*-butylammonium fluoride (1M solution in THF, 0.88 mL, 0.88 mmol) were added. The reaction mixture was stirred for an additional 30 minutes. Then the mixture solution was poured into 50 mL of water and extracted with 1:1 ether/ethyl acetate (3×50 mL). The organic extracts were combined and washed with 50 mL of aqueous sodium bicarbonate and 50 mL of water. The organic phase was dried over anhydrous sodium sulfate overnight and then evaporated under reduced pressure to give a crude product as dark orange oil. The crude product was purified by column chromatography (1:1 hexanes/ethyl acetate) to yield the pure product as light yellow oil (0.0431 g, yield = 23.7%): R_f = 0.72 (1:1 hexanes/ethyl acetate); ^{1}H NMR ($CDCl_3$) δ 7.42 (m, 4H), 5.21 (dd, 1H), 4.05 (t, 1H), 3.93 (d, 2H), 2.88 (dd,1H), 2.56 (dd,1H).

(2R, 5R)-Dihydro-2(hydroxymethyl)-5-*p*-methylphenylfuran-3(2H)-one (4c).[14] Furanoid glycal **3** (0.25 g, 1.1 mmol), 4-iodotoluene (0.26 g, 1.1 mmol), sodium acetate (95 mg, 1.2 mmol), tributylamine (45 μL, 0.19 mmol) and palladium acetate (25 mg, 0.11 mmol) were dissolved in 10 mL of DMF in a dry flask under nitrogen atmosphere protection. The mixture was reacted for 6 hours at room temperature. Then acetic acid (110. μL, 1.81 mmol) and tetra-*n*-butylammonium fluoride (1M solution in THF, 1.1 mL, 1.1 mmol) were added. The reaction mixture was stirred for an additional 30 minutes. Then the mixture solution was poured into 50 mL of water and extracted with 3×50 mL of 1:1 ether/ethyl acetate. The organic extracts were combined and washed with 50 mL of aqueous sodium bicarbonate and 50 mL of water. The organic phase was dried over anhydrous sodium sulfate overnight and then evaporated under reduced pressure to give a crude product as dark orange oil. The crude product was purified by column chromatography (1:1 hexanes/ethyl acetate) to yield the pure product as light yellow oil (0.107 g, yield = 46.5%): R_f = 0.56 (1:1 hexanes/ethyl acetate); ^{1}H NMR ($CDCl_3$) δ 7.39 (m, 4H), 5.20 (dd, 1H), 4.07 (t, 1H), 3.95 (d, 2H), 2.88 (dd,1H), 2.55 (dd,1H), 2.39 (s, 3H).

Synthesis of 2'-deoxy C-ribosides

1, 2-Dideoxy-1-β-phenyl-D-ribofuranose (5a).[15] 2'-Deoxy-3′-keto C-ribose **4a** (0.0763 g, 0.39 mmol) was dissolved in 10 mL of acetonitrile and 10 mL of glacial acetic acid in a dry flask under nitrogen atmosphere protection. The reaction mixture was cooled to 0 ℃ using an ice bath, then sodium triacetoxyborohydride (0.1 g, 0.5 mmol) was added. The reaction solution was stirred at 0 ℃ for 30 minutes. After the completion of reaction, the solvent was poured into 50 mL of aqueous sodium bicarbonate solution and extracted with 3×50 mL of ethyl acetate. The combined organic phases were dried over anhydrous sodium sulfate and evaporated under

reduced pressure to give a crude product as light yellow oil/solid. The crude product was purified by column chromatography (23:2 dichloromethane/methanol) to yield the pure product as a white solid (0.031 g, yield = 40.3%): R_f = 0.36 (23:2 dichloromethane/methanol); ^{1}H NMR ($CDCl_3$) δ 7.39-7.30 (m, 5H), 5.20 (dd, 1H), 4.47 (t, 1H), 4.05 (dt, 1H), 3.86 (dd, 1H), 3.77 (dd, 1H), 2.28 (ddd,1H), 2.08 (ddd,1H), 1.98 (br, 1H), 1.63 (br, 1H).

1,2-Dideoxy-1-β-(4-fluorophenyl)-D-ribofuranose (5b).[15] 2'-Deoxy-3′-keto C-ribose **4b** (0.02 g, 0.1 mmol) was dissolved in 5 mL of acetonitrile and 5 mL of glacial acetic acid in a dry flask under nitrogen atmosphere protection. The reaction mixture was cooled to 0 ℃ using an ice bath, then sodium triacetoxyborohydride (0.033 g, 0.15 mmol) was added. The reaction solution was stirred at 0 ℃ for 30 minutes. After the completion of reaction, the solvent was poured into 30 mL of aqueous sodium bicarbonate solution and extracted with 3×30 mL of ethyl acetate. The organic phase was dried over anhydrous sodium sulfate overnight and then evaporated under reduced pressure to give a crude product as light yellow oil/solid. The crude product was purified by column chromatography (9:1 dichloromethane/methanol) to yield the pure product as a white solid (0.0068 g, yield = 34.1%): R_f = 0.53 (9:1 dichloromethane/methanol); ^{1}H NMR ($CDCl_3$) δ 7.48-7.29 (m, 4H), 5.26 (dd, 1H), 4.51 (t, 1H), 4.08 (dt, 1H), 3.82 (dd, 1H), 3.78 (dd, 1H), 2.29 (ddd,1H), 2.08 (ddd,1H), 1.75 (br, 1H), 1.21 (br, 1H).

1,2-Dideoxy-1-β-(4-methylphenyl)-D-ribofuranose (5c).[15] 2'-Deoxy-3′-keto C-ribose **4c**(0.1 g, 0.5 mmol) was dissolved in 5 mL of acetonitrile and 5 mL of glacial acetic acid in a dry flask under nitrogen atmosphere protection. The reaction mixture was cooled to 0 ℃ using an ice bath, then sodium triacetoxyborohydride (0.15 g, 0.7 mmol) was added. The reaction solution was

stirred at 0 ℃ for 30 minutes. After the completion of reaction, the solvent was poured into 30 mL of aqueous sodium bicarbonate solution and extracted with 3×30 mL of ethyl acetate. The organic phase was dried over anhydrous sodium sulfate overnight and then evaporated under reduced pressure to give a crude product as light yellow oil/solid. The crude product was purified by column chromatography (9:1 dichloromethane/methanol) to yield the pure product as a white solid (0.0834 g, yield = 82.7%): R_f = 0.51 (9:1 dichloromethane/methanol); ^{1}H NMR ($CDCl_3$) δ 7.45-7.27 (m, 4H), 5.19 (dd, 1H), 4.48 (t, 1H), 4.02 (dt, 1H), 3.80 (dd, 1H), 3.78 (dd, 1H), 2.41 (s, 3H), 2.23 (ddd,1H), 2.07 (ddd,1H), 1.84 (br, 1H).

Synthesis of 5-O-(dimethoxytrityl)-2-deoxy C-ribosides.

5-*O*-(Dimethoxytrityl)-1,2-dideoxy-1-β-phenyl-D-ribofuranose (6a).[16] To a solution of **5a** (0.02 g, 0.1 mmol) in 1 mL of anhydrous pyridine, DMAP (3 mg, 0.02 mmol), DMTrCl (45 mg, 0.13 mmol) and triethylamine (20 μL, 0.13 mmol) were added. The mixture solution was stirred for 20 hours at room temperature. The reaction was quenched by adding 5 mL of methanol. The solution was evaporated under reduced pressure and then 40 mL of water was poured into the flask, followed by extraction with 3×50 mL of ethyl acetate. The organic extracts were combined and dried over anhydrous sodium sulfate overnight and then evaporated under reduced pressure to give a crude product as light yellow solid. The crude product was purified by column chromatography (3:1 hexanes/ethyl acetate with 3% TEA) to yield the pure product as a white solid (0.018 g, yield = 33.3%): R_f = 0.56 (3:1 hexanes/ethyl acetate with 3% TEA); ^{1}H NMR ($CDCl_3$) δ 7.49-7.15 (m, 14H), 6.84 (d, 4H), 5.20 (dd, 1H), 4.45 (m, 1H), 4.08 (m, 1H), 3.81 (s, 6H), 3.39 (dd, 1H), 3.28 (dd, 1H), 2.27 (ddd,1H), 2.05 (ddd,1H).

5-*O*-(Dimethoxytrityl)-1,2-dideoxy-1-β-(4-fluorophenyl)-D-ribofuranose (6b).[16] To a solution of **5b** (0.006 g, 0.03 mmol) in 1 mL of anhydrous pyridine, DMAP (1 mg, 0.007 mmol), DMTrCl (18 mg, 0.05 mmol) and triethylamine (8 μL, 0.05 mmol) were added. The mixture solution was stirred for 20 hours at room temperature. The reaction was quenched by adding 5 mL of methanol. The solution was evaporated under reduced pressure and then 40 mL of water was poured into the flask, followed by extraction with 3×50 mL of ethyl acetate. The organic extracts were combined and dried over anhydrous sodium sulfate overnight and then evaporated under reduced pressure to give a crude product as light yellow solid. The crude product was purified by column chromatography (2:1 hexanes/ethyl acetate with 3% TEA) to yield the pure product as a white solid (0.0044 g, yield = 28.2%): R_f = 0.61 (2:1 hexanes/ethyl acetate with 3% TEA); ^{1}H NMR ($CDCl_3$) δ 7.49-6.83 (d, 17H), 5.16 (dd, 1H), 4.44 (m, 1H), 4.07 (m, 1H), 3.81 (s, 6H), 3.37 (dd, 1H), 3.28 (dd, 1H), 2.26 (ddd,1H), 2.04 (ddd,1H).

5-*O*-(Dimethoxytrityl)-1,2-dideoxy-1-β-(4-methylphenyl)-D-ribofuranose (6c).[16] To a solution of **5c** (0.08 g, 0.4 mmol) in 5 mL of anhydrous pyridine, DMTrCl (0.22g, 0.64 mmol) and $AgNO_3$ (0.11g, 0.66 mmol) were added. The mixture solution was stirred for 3-4 hours at room temperature. Then additional DMTrCl (0.1 g, 0.3 mmol) and $AgNO_3$ (0.05 g, 0.33 mmol) were added, and the reaction was left for overnight. The reaction was quenched by adding 30 mL of saturated sodium bicarbonate solution, followed by extraction with 3×30 mL of dichloromethane. The organic extracts were combined and dried over anhydrous sodium sulfate overnight and then evaporated under reduced pressure. The crude product was purified by column chromatography (7:3 hexanes/ethyl acetate with 5% TEA) to yield the pure product as a white foam (0.1207 g, yield = 59.2%): R_f = 0.43 (7:3 hexanes/ethyl acetate with 5% TEA); ^{1}H NMR ($CDCl_3$) δ 7.51-

6.85 (d, 17H), 5.13 (dd, 1H), 4.42 (m, 1H), 4.07 (m, 1H), 3.85 (s, 6H), 3.42 (dd, 1H), 3.28 (dd, 1H), 2.36(s, 3H), 2.23 (ddd,1H), 2.04 (ddd,1H),

Synthesis of phosphoramidites

3-*O*-(2-Cyanoethyl-*N*,*N*-diisopropylphosphoramidite)-5-*O*-(dimethoxytrityl)-1,2-dideoxy-1-β-phenyl-D-ribofuranose(7a).[16] To a solution of **6a** (0.01 g, 0.021 mmol) in 2 mL of dichloromethane, triethylamine (8.7 μL, 0.063 mmol) and 2-cyanoethyl *N*,*N*-diisopropylchloro-phosphoramidite (9.3 μL, 0.042 mmol) were added. The mixture solution was stirred for 1 hour at room temperature. Then the reaction mixture was diluted with 20 mL of dichloromethane and washed sequentially with 50 mL of aqueous sodium bicarbonate and 50 mL of aqueous sodium chloride. The organic portion was dried over anhydrous sodium sulfate overnight and then evaporated under reduced pressure to give a crude product as light yellow solid. The crude product was purified by column chromatography (2:1 hexanes/ethyl acetate with 5% TEA) to yield the pure product as a white solid (0.0097 g, yield = 66.4%): R_f = 0.78 (2:1 hexanes/ethyl acetate with 5% TEA); ^{1}H NMR ($CDCl_3$) δ 7.49-7.15 (m, 14H), 6.82 (m 4H), 5.15 (dd, 1H), 4.49 (m, 1H), 4.28 (m, 1H), 3.84-3.55 (m, 4H), 3.80 (s, 6H), 3.33 (dd, 1H), 3.29 (dd, 1H), 2.69 (m, 2H), 2.52 (m, 2H), 2.26 (m,1H), 2.04 (m,1H), 1.43-1.15 (m, 12H).

3-*O*-(2-Cyanoethyl-*N*, *N*-diisopropylphosphoramidite)-5-*O*-(dimethoxytrityl)-1,2-dideoxy-1-β-(4-fluorophenyl-D-ribofuranose (7b). [16] To a solution of **6b** (0.01 g, 0.021 mmol) in 2 mL of dichloromethane, triethylamine (8.9 μL, 0.064 mmol) and 2-cyanoethyl *N*,*N*-diisopropylchlorophosphoramidite (9.6 μL, 0.043 mmol) were added. The mixture solution was stirred for 1 hour at room temperature. Then the reaction solution was diluted with 20 mL of

dichloromethane and washed sequentially with 50 mL of aqueous sodium bicarbonate and 50 mL of aqueous sodium chloride. The organic portion was dried over anhydrous sodium sulfate overnight and then evaporated under reduced pressure to give a crude product as light yellow solid. The crude product was purified by column chromatography (3:1 hexanes/ethyl acetate with 3% TEA) to yield the pure product as a white solid (0.0054 g, yield = 37.8%): R_f = 0.72 (hexanes:ethyl acetate = 3:1 and 3% TEA); ^{1}H NMR ($CDCl_3$) δ 7.49-6.82 (m, 17H), 5.20 (dd, 1H), 4.49 (m, 1H), 4.23 (m, 1H), 3.85-3.51 (m, 4H), 3.81 (s, 6H), 3.31 (dd, 1H), 3.28 (dd, 1H), 2.73 (m, 2H), 2.51 (m, 2H), 2.27 (m,1H), 2.05 (m,1H), 1.45-1.17 (m, 12H).

3-*O*-(2-Cyanoethyl-*N*,*N*-diisopropylphosphoramidite)-5-*O*-(dimethoxytrityl)-1,2-dideoxy-1-β-(4-methylphenyl-D-ribofuranose (7c).[16] To a solution of **6c** (0.1 g, 0.2 mmol) in 5 mL of dichloromethane, triethylamine (82 μL, 0.58 mmol) and 2-cyanoethyl *N*,*N*-diisopropylchloro-phosphoramidite (88 μL, 0.4 mmol) were added. The mixture solution was stirred for 2 hours at room temperature. Then the reaction solution was diluted with 20 mL of dichloromethane and washed sequentially with 50 mL of aqueous sodium bicarbonate and 50 mL of aqueous sodium chloride. The organic portion was dried over anhydrous sodium sulfate overnight and then evaporated under reduced pressure to give a crude product as yellow solid. The crude product was purified by column chromatography (3:1 hexanes/ethyl acetate with 6% TEA) to yield the pure product as a white foam (0.1225 g, yield = 86.3%): R_f = 0.62 (3:1 hexanes/ethyl acetate with 6% TEA); ^{1}H NMR ($CDCl_3$) δ 7.48-6.80 (m, 17H), 5.21 (dd, 1H), 4.46 (m, 1H), 4.21 (m, 1H), 3.87-3.52 (m, 4H), 3.80 (s, 6H), 3.33 (dd, 1H), 3.29 (dd, 1H), 2.76 (m, 2H), 2.48 (m, 2H), 2.33 (s, 3H), 2.24 (m,1H), 2.04 (m,1H), 1.51-1.16 (m, 12H).

Synthesis of 2'-deoxy-3'-keto C-ribosides using bisTBDMS glycal (Scheme 3.8)

(2R, 5R)-Dihydro-2-(hydroxymethyl)-5-phenylfuran-3(2H)-one(4a).[10] Palladium acetate (4.5 mg, 0.02 mmol) and triphenylarsine (12.2 mg, 0.04 mmol) were dissolved in 4 mL of DMF in a dry flask under nitrogen atmosphere protection and stirred for 20 min. Then 1,4-anhydro-3,5-bis-*O*-(*tert*-butyldimethylsilyl)-2-deoxy-D-erythro-pent-1-enitol **2** (70 mg, 0.2 mmol), iodobenzene (51 mg, 0.25 mmol) and tributylamine (74 μL, 0.31 mmol) in DMF (3 mL) were added. The mixture was reacted under nitrogen for 18 hours at 70 ℃. Then the mixture was cooled to 0 ℃ using ice bath and tetra-*n*-butylammonium fluoride (1M solution in THF, 0.5 mL, 0.5 mmol) were added. The reaction mixture was stirred for an additional 1 hour. Then the mixture solution was filtered and extracted with ether/ethyl acetate (1:1, 3×50 mL) and saturated $NaHCO_3$ solution (50 mL). The organic phase was dried by anhydrous sodium sulfate and then evaporated under reduced pressure to give a crude product as light yellow oil. The crude product was purified by column chromatography (hexanes:ethyl acetate = 1:1) to yield the product as light yellow oil (6.4 mg, yield = 16.7%): R_f = 0.52 (hexanes:ethyl acetate = 1:1); ^{1}H NMR ($CDCl_3$) δ 7.41-7.57 (m, 10H), 5.61 (dd, 1H, α-product), 5.24 (dd, 1H, β-product), 4.28 (t, 1H, α-product), 4.11 (t, 1H, β-product), 4.02 (d, 2H, α-product), 3.99 (d, 2H, β-product), 3.04 (dd,1H, α-product), 2.96 (dd,1H, β-product), 2.75 (dd,1H, α-product) 2.60 (dd,1H, β-product).

(2R, 5R)-5-(4-Fluorophenyl)-dihydro-2-(hydroxymethyl)furan-3(2H)-one(4b).[10] Palladium acetate (6.5 mg, 0.029 mmol) and triphenylarsine (17.7 mg, 0.06 mmol) were dissolved in 3 mL of DMF in a dry flask under nitrogen atmosphere protection and stirred for 20 min. Then 1,4-anhydro-3,5-bis-*O*-(*tert*-butyldimethylsilyl)-2-deoxy-D-erythro-pent-1-enitol **2** (100 mg, 0.291 mmol), 1-fluoro-4-iodobenzene (80 mg, 0.36 mmol) and tributylamine (104 μL, 0.44 mmol) in

DMF (2 mL) were added. The mixture was reacted under nitrogen for 18 hours at 70 ℃. Then the mixture was cooled to 0 ℃ using ice bath and tetra-*n*-butylammonium fluoride (1M solution in THF, 0.59 mL, 0.59 mmol) were added. The reaction mixture was stirred for an additional 1 hour. Then the mixture solution was filtered and extracted with ether/ethyl acetate (1:1, 3x50 mL) and saturated $NaHCO_3$ solution (50 mL). The organic phase was dried by anhydrous sodium sulfate and then evaporated under reduced pressure to give a crude product as light yellow oil. The crude product was purified by column chromatography (hexanes:ethyl acetate = 1:1) to yield the product as light yellow oil (10 mg, yield = 16.4%): R_f = 0.58 (hexanes:ethyl acetate = 1:1); ^{1}H NMR ($CDCl_3$) δ 7.40-7.55 (m, 8H), 5.59 (dd, 1H, α-product), 5.23 (dd, 1H, β-product), 4.26 (t, 1H, α-product), 4.10 (t, 1H, β-product), 4.01 (d, 2H, α-product), 4.00 (d, 2H, β-product), 3.03(dd,1H, α-product), 2.98 (dd,1H, β-product), 2.71 (dd,1H, α-product) 2.58 (dd,1H, β-product).

3.6 References

1. (a) Ren R. X.-F ; Chaudhuri, N. C.; Paris, P. L.; Rumney IV, S.; Kool, E. T. *J. Am. Chem. Soc.* **1996**, *118*, 7671; (b) Chaudhuri, N. C.; Kool, E. T. *Tetrahedron Lett.* **1995**, *36*, 1795; (c) Schweitzer, B. A.; Kool, E. T. *J. Org. Chem.* **1994**, *59*, 7238; (d) Chaudhuri, N. C.; Ren R. X.-F ; Kool, E. T. *Synlett* **1997**, 341; (e) Schweitzer, B. A.; Kool, E. T. *J. Am. Chem. Soc.* **1996**, *118*, 931

2. Srivasta, P. C.; Robin, R. K. *J. Med. Chem.* **1981**, *24*, 1172.

3. Petier, C. R.; Revankar, G. R.; Dalley, N. K.; George, R. D.; McKernan, P. A.; Hamil, R. L.; Robin, R. K. *J. Med. Chem.* **1986**, *29*, 268.

4. Strassler, C.; Davis, N.E.; Kool, E. T. *Helv. Chim. Acta* **1999**, *82*, 2160.

5. Millican, T. A.; Mock, G. A.; Chauncey, M. A.; Patel, T. P.; Eaton, M. A. W.; Gunning, J.; Gutbush, S. D.; Neidle, S.; Mann, J. *Nucleic Acids Res.* **1984**, *12*, 7435.

6. Wichai, U.; Woski, S. A. *Bioorg. Med. Chem. Lett.* **1998**, *8*, 3465.

7. Wichai, U.; Woski, S. A. *Org. Lett*. **1999**, *1*, 1173.

8. Wichai, U. "Synthesis and investigation of PNA:DNA complexes containing novel aromatic residues", Ph.D. Dissertation, University of Alabama, **2003**.

9. (a) Arai, I.; Daves Jr., G. D. *J. Org. Chem.* **1978**, *43*, 4110; (b) Kwok, D.–I.; Farr, R. N.; Daves Jr., G. D. *J. Org. Chem.* **1991**, *56*, 3711; (c) Zhang, H.; Brakta, M.; Daves Jr., G. D. *Nucleosides Nucleotides* **1995**, *14*, 105.

10. Matsuda. S.; Leconte. A. M.; Romesberg. F. E. *J. Am. Chem. Soc.* **2007**, *129*, 5551.

11. Larsen, E.; Sofan, M. A.; Pedersen, E. B.; etc. *Synthesis* **1994**,1037.

12. Cameron, M. A.; Cush, S. B.; Hammer, R. P. *J. Org. Chem.* **1997**, *62*, 9065.

13. Coleman, R. S.; Madaras, M. L. *J. Org. Chem.* **1998**, *63*, 5700.

14. Farr, R. N.; Outten, R. A.; Cheng. J. C.; Daves Jr, G. D. *Organometallics* **1990**, *9*, 3151.

15. Evans, D. A.; Chapamn, K. T.; Carreira, E. M. *J. Am. Chem. Soc.* **1988**, *110*, 3560.

16. Hakimelahi, G.; Proba, Z. A.; Oglivie, K. K. *Can. J. Chem.* **1982**, *60*, 1106.

CHAPTER IV

Structure Determination of Oligodeoxynucleotide Duplex Containing Novel Residue Using 2D NMR

4.1 Introduction

The structural determination of oligodeoxynucleotide duplexes is an important tool to investigate interactions between an individual residue and its nearby or opposite DNA nucleobases. Several strategies including X-ray crystallography, electron microscopy,[1] and NMR spectroscopy[2-4] can be applied to determine the structure of oligodeoxynucleotides. Because electron microscopy has relatively low resolution and X-ray crystallography requires the crystallization of the DNA duplex, NMR spectroscopy is often preferred to probe the structures of oligodeoxynucleotide duplexes. Importantly, this method gives information about the structure as it exists in solution.

The most important consideration for the NMR techniques is size of biomolecule. The size limitation for the routine NMR techniques is up to 25 kDa for protein and 100 nucleotides for DNA/RNA, and this limitation is mainly caused by the peak overlap issue and fast relaxation time of the larger biomolecule sample.[2-4] Compared with smaller biomolecules, more peaks are be observed in the NMR spectra for the larger biomolecule, and some of these peaks can overlap with each other. This issue can make the resonance assignments much more difficult and the accurate analysis of the NMR spectra becomes unavailable. In addition, broader peaks will be shown in the NMR spectra and decrease the resolution of the NMR spectra for larger

biomolecules, because larger biomolecules tumble more slowly in solution and have a faster relaxation of transverse magnetization.

The NMR technique has been proved to be a successful technique for the determination of the structure of DNA duplex containing non-natural residue as long as the length of DNA oligonucleotide is short and its size is relatively small. For example, in 2012 Kanaori and coworkers[5] reported the solution structure of a DNA undecamer duplex containing the unusual ring-deaminated nucleobase, oxanine. Another naturally occurring lesion, (5′S)-5′,8-cyclo-2'-deoxyadenosine, was recently studied using NMR by de los Santos and coworkers.[6] Oligonucleotide NMR methods with modified residues have also been used to probe DNA-ligand interactions. Markiewicz and coworkers[7] prepared a cytidine analog alkylated at N-4 with spermine. Determination of the solution structure was touted as shedding light on the DNA binding preferences of this biologically important polyamine. Moreover, NMR techniques can not only resolve DNA duplex structures, but have also been reported in the examination of an indoloquinoline nucleoside incorporated into a DNA probe designed to bind to a double-helical DNA target as the triple-helical complex.[8]

The essential steps in the study of DNA structure by NMR are to perform a number of 2D experiments including correlation spectroscopy (COSY), total correlation spectroscopy (TOCSY) and nuclear Overhauser effect spectroscopy (NOESY), and assign the proton chemical shifts to the particular protons in the molecule.[5, 9] COSY provides the information on through-bond connectivities in order to identify the protons that are coupled to each other. TOCSY is similar to COSY but, in addition to information protons that are coupled directly, cross peaks are observed for protons that are connected by chains of couplings. This is a useful technique to examine the large interconnected spin couplings. NOESY techniques provide information on the through-

space connectivities of nuclei rather than through-bond interactions. Combined with the COSY and TOCSY data, NOESY allows for the identification of coupled protons that are spatially close.[10, 11] These data allow for the assignment of chemical shifts to specific protons within the duplex structure. Following this, the structure model of DNA can be constructed by using computational calculation based on the input data obtained from NMR experiments. Violation analysis and iterative structural calculation are necessary to achieve a family of converged structures.

4.2 Results and Discussion

Synthesis and purification of the oligodeoxynucleotides for NMR studies

The sequence chosen for this study was the classic self-complementary 12-mer duplex introduced by Dickerson and coworkers in the early 1980's.[8] Solution of the crystal structure of this duplex provided the first high-resolution structural information for DNA. This sequence, 5′-CGCGAATTCGCG-3′, has since been used in numerous structural studies of DNA both in crystal and solution phases.

Our Dickerson dodecamer oligodeoxynucleotide was synthesized on an Applied Biosystems 391 DNA synthesizer utilizing traditional phosphoramidite chemistry. The DNA was prepared in a 1 μmol scale with the standard DNA synthesis cycles to obtain a necessary concentration for the NMR experiments. A "trityl on" end procedure was used, meaning that the 5′-terminal DMT group was not removed. This group is not removed by the ammonium hydroxide treatment used to deprotect the bases and the phosphodiesters and cleave the oligodeoxynucleotide from the solid support. The presence of the DMT group greatly increases the hydrophobicity of the DNA strand, allowing for its retention on reverse-phase chromatographic materials. A two-step

purification scheme was used to produce high purity DNA: (1) use of a reverse-phase cartridge removed most failure sequences and small molecule impurities (such as products from removal of the base protecting groups); (2) denaturing acrylamide electrophoresis removes any remaining failure sequences. Finally, the purified oligodeoxynucleotide was characterized by MALDI-TOF mass spectrometry and quantitated by the measurement of the UV absorbance at 260 nm.

The oligodeoxynucleotide containing a tolyl C-nucleoside residue in position 5 of the dodecamer (5′-CGCGXATTCGCG-3′, X = tolyl nucleoside) was prepared in a 2.5 μmol scale. Several modifications to the procedure outlined above were made in order to maximize the yield of DNA: the coupling time for the modified base was increased from 15 to 600 seconds and cartridge purification was omitted to avoid losses for marginal gains in purity. Dialysis was used as an additional purification step to remove tris cations present from electrophoresis buffers.

1D and 2D NMR experiments of the unmodified Dickerson dodecamer

In order to determine the optimal conditions and parameters for the 1D and 2D NMR experiments, NMR experiments were performed first using the unmodified Dickerson dodecamer. These experiments provide a reference point because the Dickerson dodecamer has been well-studied: NMR spectra were obtained 1D and 2D experiments including COSY and NOESY have been performed, and the chemical shifts of all protons have been assigned.[5-8, 13, 14]

1D and 2D NMR experiments of the Dickerson dodecamer were performed at 600 MHz. The 1D D_2O spectrum of the non-exchangeable protons of the Dickerson dodecamer duplex mostly provided results consistent with those of Hare et al.[13] The chemical shift assignments of the aromatic region are shown in Figure 4.1. The full 1D spectrum is not shown here because the peak regions except the aromatic region are crowded and the peak overlap is a significant

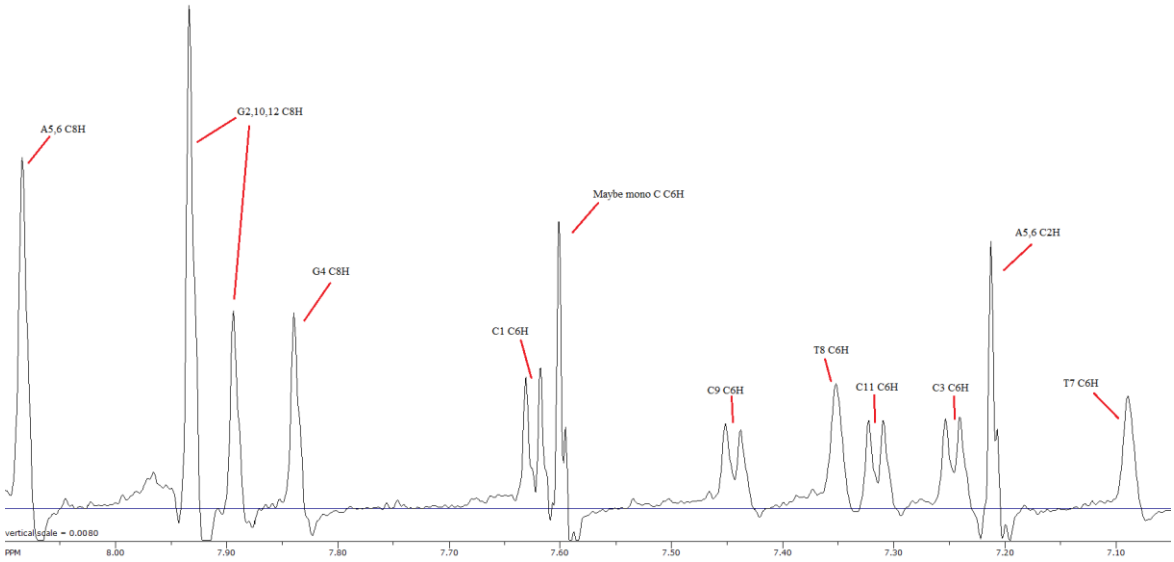


Figure 4.1 Aromatic region of the 1D NMR spectrum of the natural Dickerson dodecamer.

issue for the resonance assignment. The spectrum pattern comparison between these crowded regions from the experiment and Hare's work is challenging. One unanticipated peak may belong to a monomer deoxycytidine; this may be caused minor degradation of the 5′-terminal residue during freeze-thaw cycles used to degas the sample.[15] This 1D spectrum provided useful information for the full peak assignment in 2D COSY and NOESY spectra.

The 2D COSY spectrum of the unmodified dodecamer is shown in Figure 4.2. The most important peak regions in this figure the aromatic region and correspond to peaks from the pyrimidines (C and T). The two weak peaks at the top of the spectrum are from the four-bond coupling between methyl protons and C6 protons of the two thymine residues. The four strong peaks in the bottom of the spectrum come from the coupling between the C5 protons and the C6 protons of the four cytosines. Thus, the four cytosines and two thymines can be identified in the COSY spectrum. The existence of Tris was discovered based on the strong peak at about 3.7 ppm in the COSY spectrum, and this indicated that the Tris should be removed by dialysis prior to further experiments with the non-natural oligodeoxynucleotide.

The 2D NOESY spectrum is most important for the assignment of peaks. Any protons that experience a nuclear Overhouser effect (NOE) and show a cross peak in the NOESY spectrum are spatially close. There are several useful methods to assign the peaks in the NOESY spectrum, and two major ways are described below.[13] First, an NOE signal is normally observed between the C5 methyl group of thymine and C6 or C8 position of the n−1 base. For example, in the dodecamer sequence, the methyl from T8 shows an NOE connectivity to C6 proton from T7, and the methyl from T7 shows an NOE connectivity to C8 proton from A6 (Figure 4.3). Second, NOE signals should be observed between purine C8H or pyrimidine C6H and the C1′H of its own sugar and the (n-1) sugar. In the dodecamer, the C6 proton of T7 should show an NOE

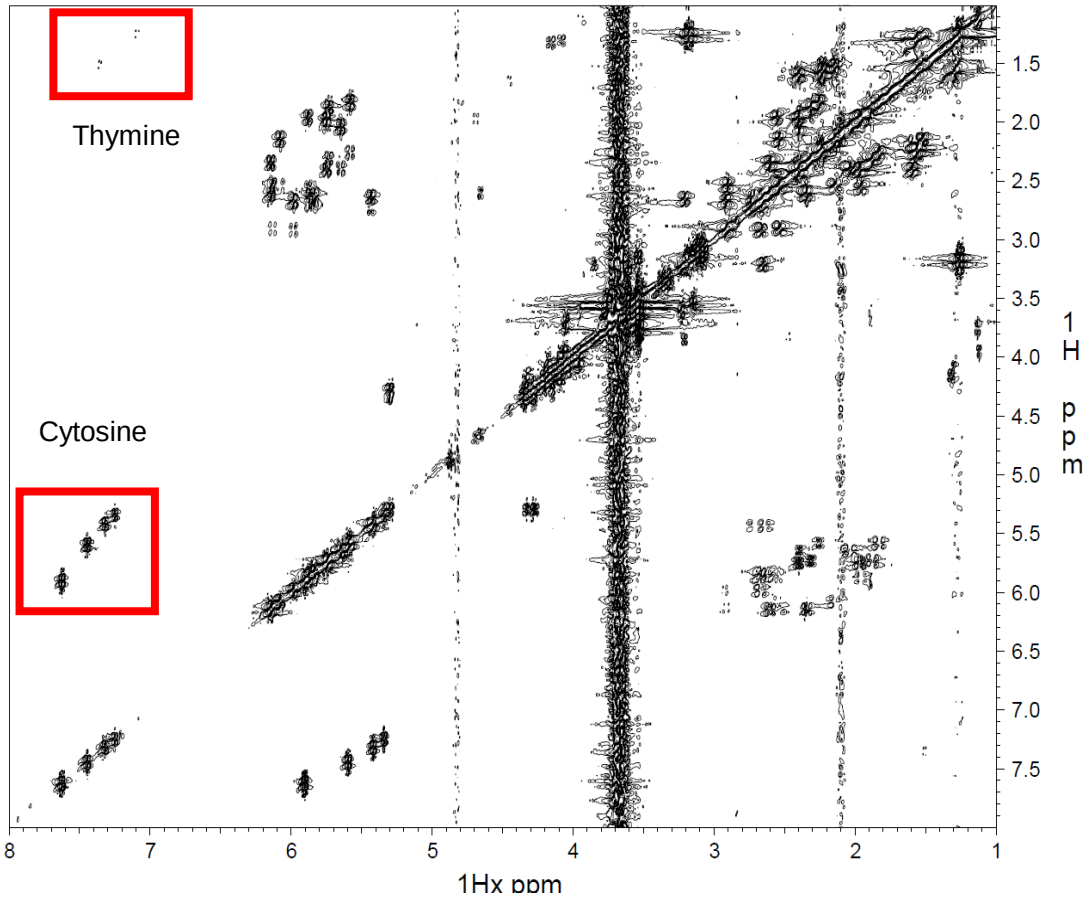


Figure 4.2 2D COSY spectrum of natural Dickerson dodecamer.

connectivity to the C1′H of its own sugar and the sugar from A6 (blue lines in Figure 4.4). Based on these notions, one of the critical regions for the peak assignment-the cross peaks between the C8H, C6H region and the C1′H, C5H region-can be assigned (Figure 4.5).

Based on these analyses, the proton chemical shift assignments for the aromatic region of the unmodified dodecamer duplex (5′-CGCGAATTCGCG-3′)$_2$ is given in Table 4.1. These results are in excellent agreement with previously published data.[13] Experimental data including the chemical shifts of the assigned aromatic protons and the pattern of NMR spectra, which is more importantly, remains consistent with the data reported by Hare et al.[13] As a result, recording the spectra in 10 mM sodium phosphate buffer at pH = 7 was deemed optimal. However, the presence of the Tris in the sample solution complicated the analysis because some of the peaks from sugar (C4′H) were overlapped by the Tris peak in COSY and NOESY. Steps were added to remove the Tris in all subsequent experiments.

1D and 2D NMR experiments of Dickerson dodecamer containing tolyl C-nucleosides

Residues of the dodecamer duplex in this study were labeled and numbered as follows and this residue numbering format will be referred to throughout the text.

	1	2	3	4	5	6	7	8	9	10	11	12	
5′ –	C	G	C	G	X	A	T	T	C	G	C	G	– 3′
3′ –	G	C	G	C	T	T	A	X	G	C	G	C	– 5′
	12	11	10	9	8	7	6	5	4	3	2	1	

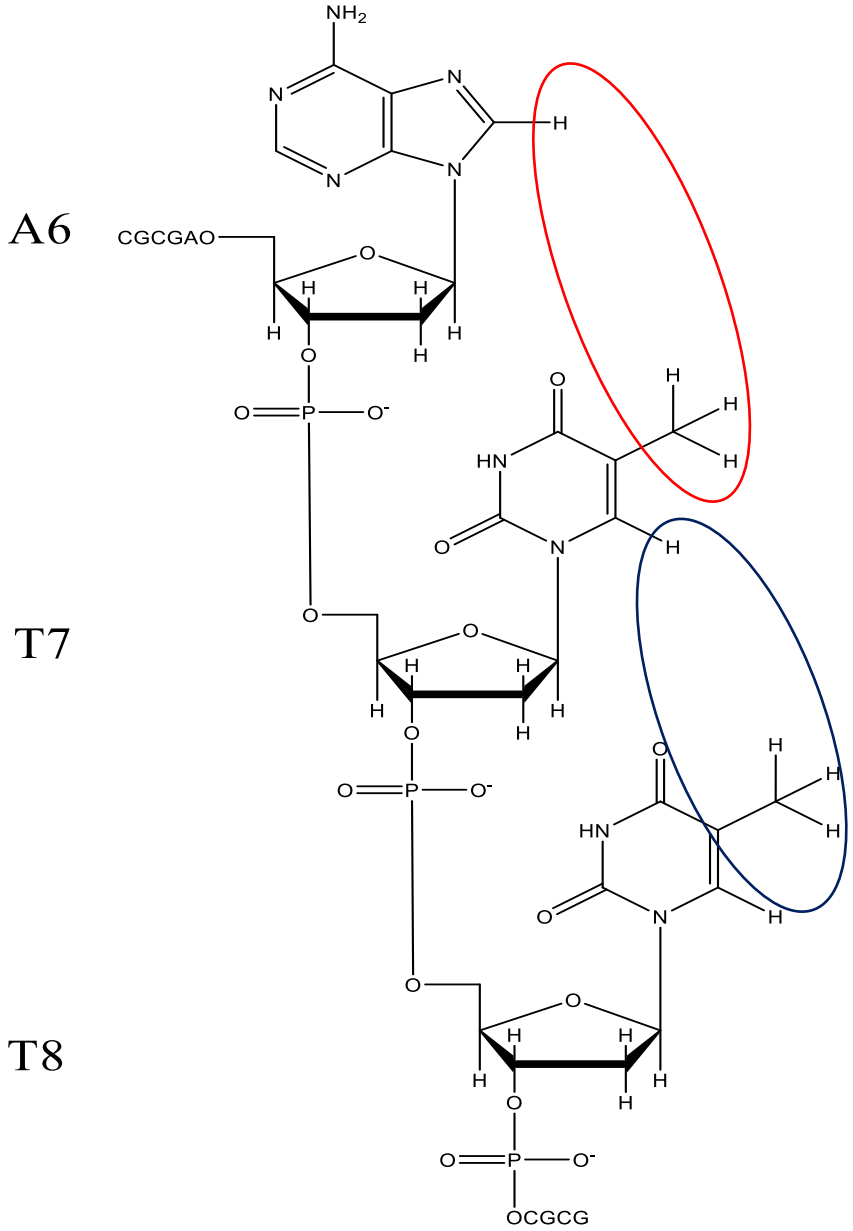


Figure 4.3 The NOE signals formed between C5 methyl group of thymine and C6 or C8 position of the n-1 base.

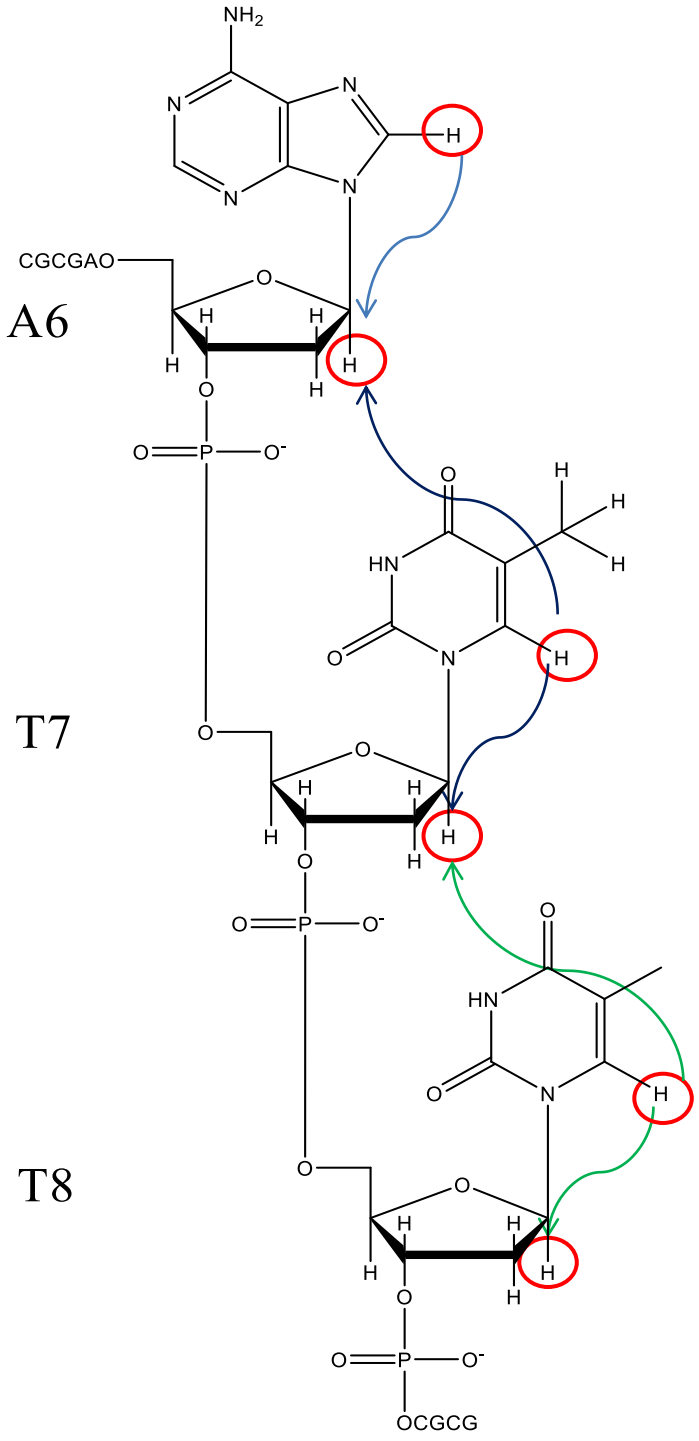


Figure 4.4 The NOE signal formed between the purine C8H or pyrimidine C6H and the C1′H of its own sugar and (n-1) sugar.

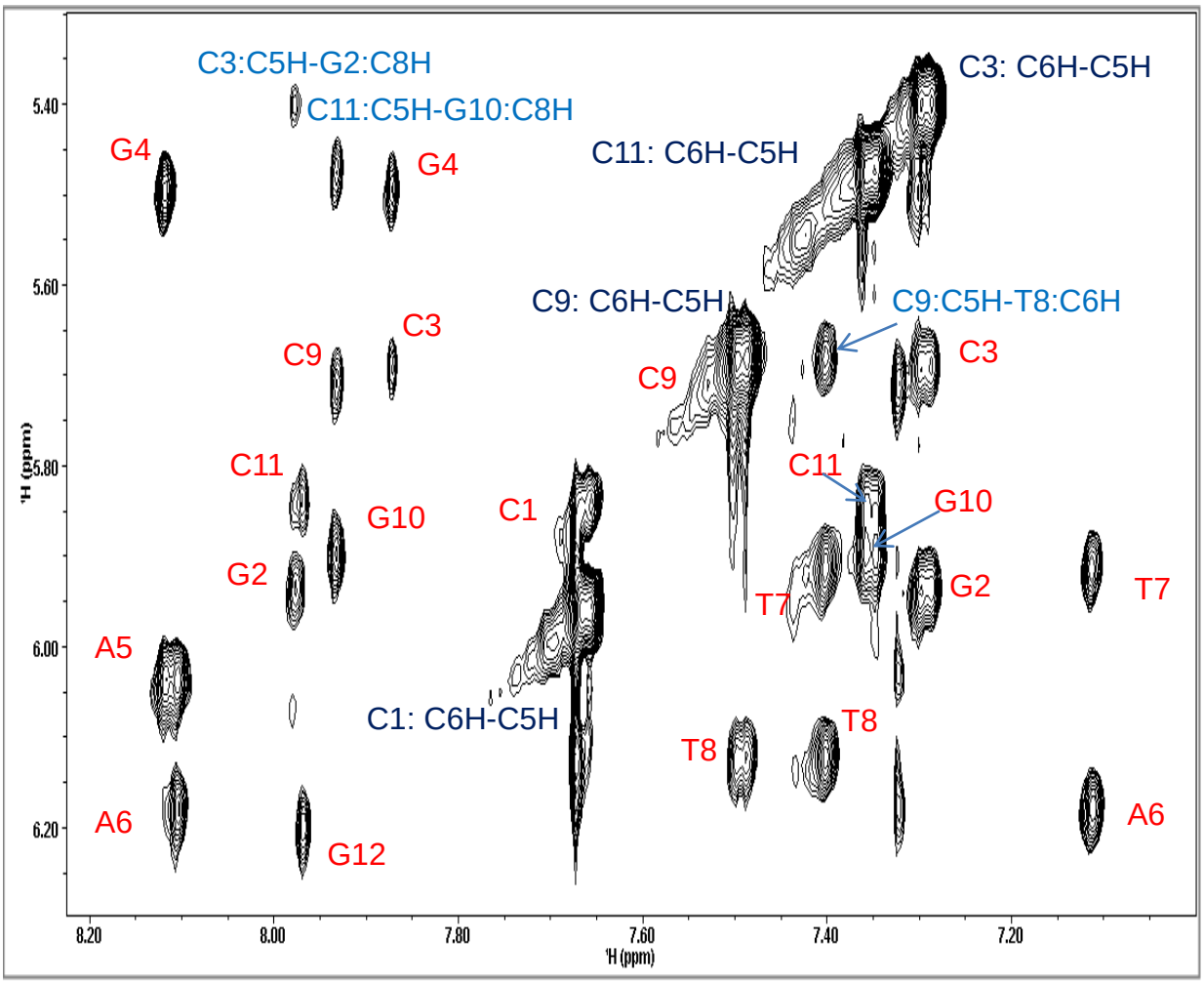


Figure 4.5 Expanded NOESY contour plot of the cross peaks between the C8H, C6H region and the C1'H, C5H region.

Table 4.1 Chemical shifts of the aromatic protons of the unmodified Dickerson dodecamer

Residues	C8H	C6H	C5H	CH_3
C1		7.66	5.96	
G2	7.97			
C3		7.28	5.40	
G4	7.86			
A5	8.12			
A6	8.11			
T7		7.11		1.31
T8		7.40		1.55
C9		7.51	5.67	
G10	7.93			
C11		7.35	5.46	
G12	7.96			

1D and 2D NMR experiments of the Dickerson dodecamer containing a tolyl (4-methylphenyl) C-nucleoside at the 5-position were performed as described above. Because the sequence is otherwise self-complementary, the double helical complex should contain two non-natural residues. The aromatic region of the 1D proton NMR spectrum recorded in D_2O for the tolyl C-nucleoside-containing Dickerson dodecamer is shown in Figure 4.6. As expected, the chemical shifts of some of the resonances change. In particular, the peaks of two cytosines are almost overlapped with each other at 7.55 ppm, and the peaks of adenine (C2H) and one of the four cytosines are overlapped completely at 7.14 ppm. Meanwhile, the AA′BB′ pattern from the *para*-disubstituted phenyl moiety is prominently present at 6.52 ppm and 6.65 ppm. This 1D spectrum provides useful information for the proton assignment of this dodecamer duplex, which includes identifying the peaks of the four cytosines (doublets with chemical shifts ≥ 7 ppm), identifying the ring proton resonances of the tolyl group (which allows for the assignment of all of the tolyl protons in the COSY spectrum), and providing the necessary information for subsequent assignment of the chemical shifts of adenine and thymine residues. Because these last residues are nearest to the tolyl residues, it was anticipated that the chemical shifts would not correspond closely to those in the unmodified dodecamer.

The 2D COSY spectrum for the tolyl dodecamer is shown in Figure 4.7. As mentioned previously, the most important peak regions belong to the weak four-bond coupling between methyl protons and C6 protons of the two thymine residues and the strong coupling between the C5 protons and the C6 protons of the four cytosines. Moreover, one additional weak peak can be found at the thymine region, this peak belongs to the weak four-bond coupling between methyl protons and C3 proton of the tolyl residue (Figure 4.8, marked with rectangle); one additional strong peak also can be found at the cytosine region, this peak belong to the strong coupling

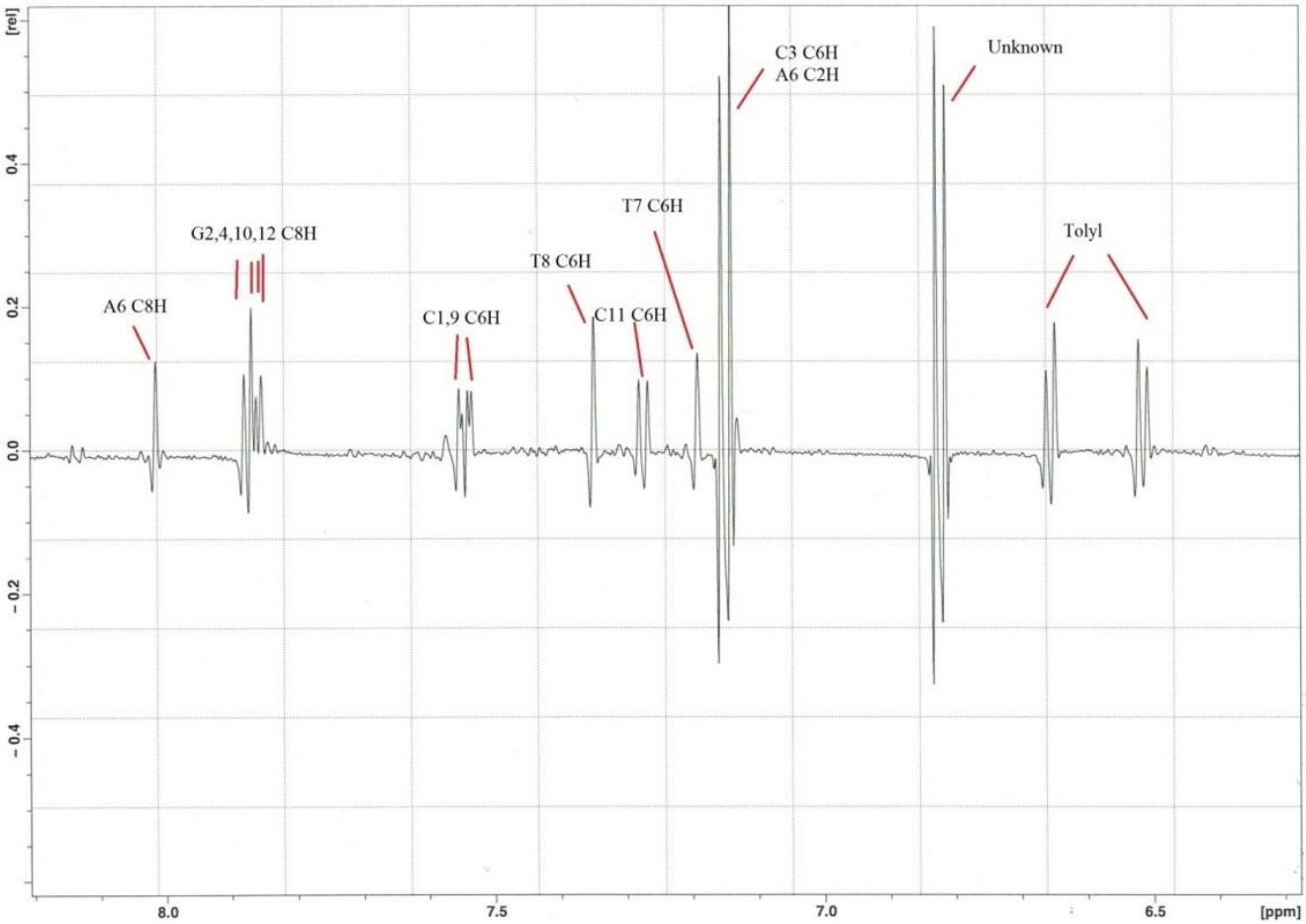


Figure 4.6 The aromatic region of the 1D proton NMR spectrum of the tolyl C-nucleoside-containing Dickerson dodecamer.

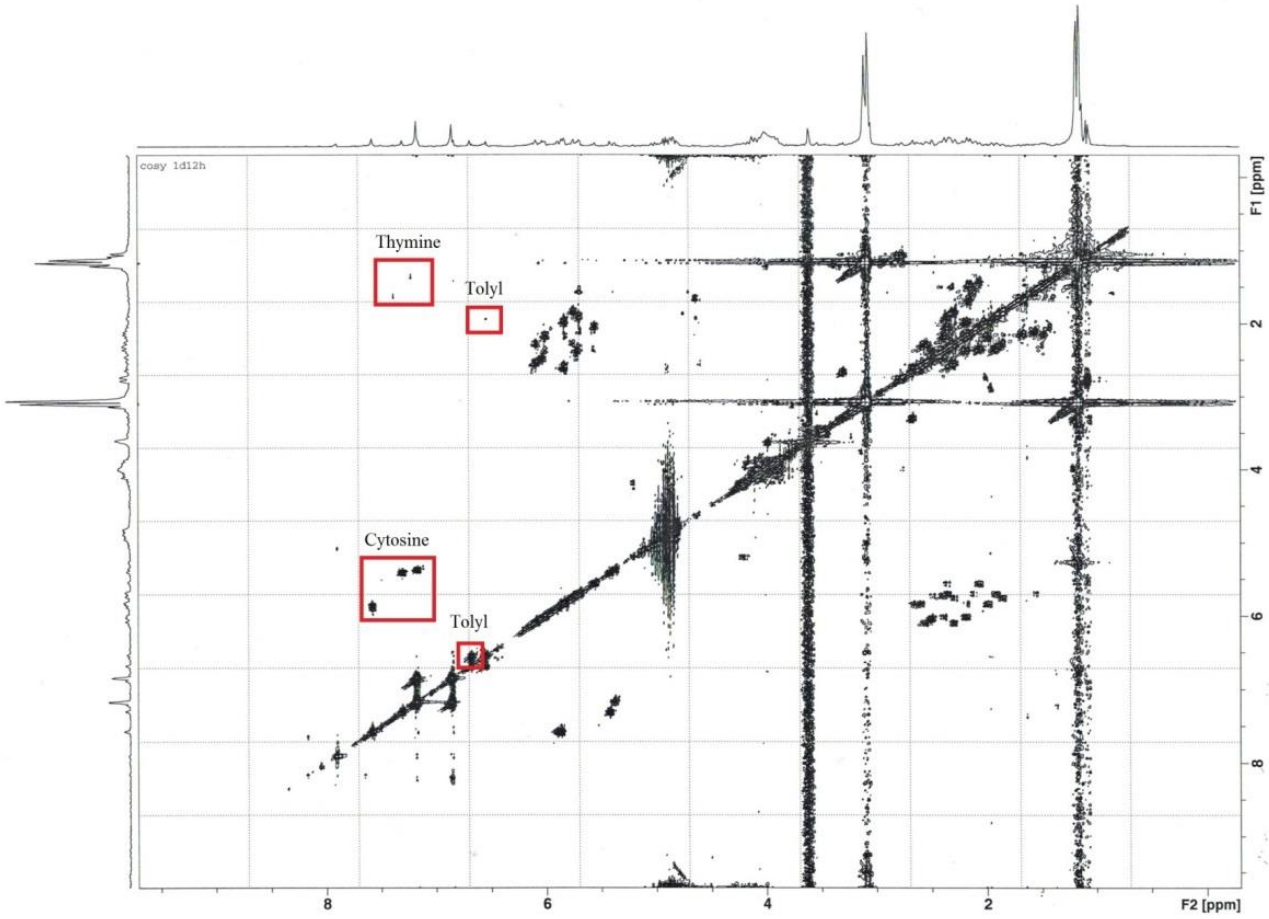


Figure 4.7 2D COSY spectrum of tolyl C-nucleoside-containing Dickerson dodecamer. One weak peak and one strong peak belonging to the tolyl residue are labeled in this spectrum.

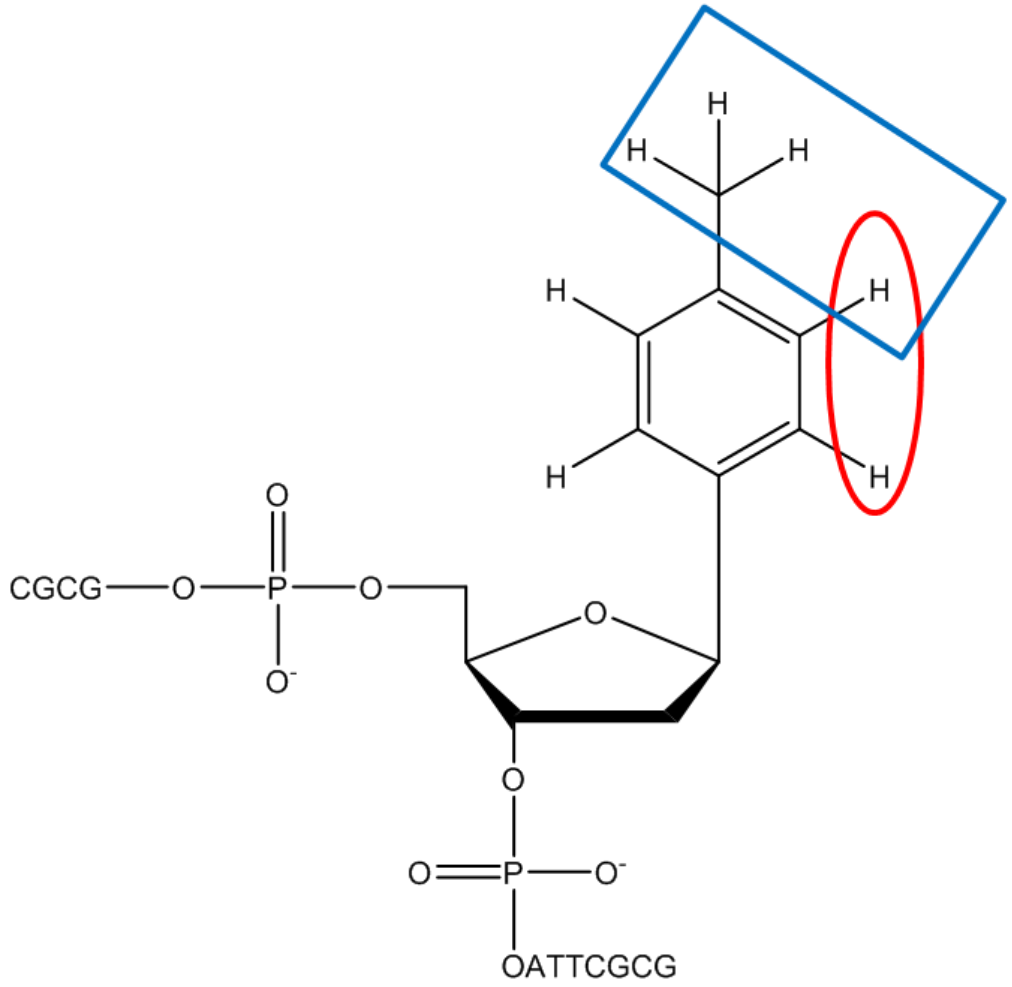


Figure 4.8 Structure of tolyl residue. The weak COSY signal comes from the protons marked in blue rectangle; strong COSY signal comes from the protons marked in red oval.

between the C2 proton and the C3 proton of the tolyl residue (Figure 4.8, marked with circle). Because there is only one tolyl residue in the sequence, the chemical shifts for these protons can be assigned directly from the COSY spectrum: C2 (6) H: 6.65 ppm, C3 (5) H: 6.52 ppm, and CH_3: 1.96 ppm.

The full peak assignment was based on the 2D NOESY spectrum (Figure 4.9). As described in the peak assignment of Dickerson dodecamer, two major methods were used for the peak assignment of aromatic protons and C1′ protons. Once these protons have been assigned, COSY and NOESY were analyzed together to assign all of the remaining sugar protons. C1′ protons are scalar coupled to C2′ (β) and C2″ (α) protons as well as the C3′ protons can be scalar coupled to C2′ (β) and C2″ (α) protons, and these cross peaks can be shown in the COSY spectrum (Figure 4.7). Both the C2′ protons and C2″ protons can be strongly coupled to their C1′ protons, however, any given C2″ proton is closer to its C1′ proton than is C2′ proton, hence a much stronger NOE cross peak can be formed in the NOESY spectrum. This allows the C2″ protons to be distinguished from C2′ protons (Figure 4.10, marked with circle). Moreover, NOE signals between C6/C8 proton and C2″ proton of its neighboring (n-1) sugar can provide an internal check. Then the C3′ protons can be assigned by two methods: the scalar coupling between C3′H and C2″H in the COSY spectrum or the NOE between C3′H and C2′H (Figure 4.10 marked with rectangle). Similar procedures were performed to identify the peaks belonging to C4′ protons. The C5′ and C5″ protons cannot be assigned completely because of the similar chemical shifts of the C5′/C5″ protons and the overlap with the C4′H region (Table 4.2).

The 2D NOESY spectrum shows the proton coupling through space, which means the intensity of peak can reflect the distance between the two coupled protons. The NOE peak intensity is inversely proportional to r^6 (r is the distance between the two protons) and the general

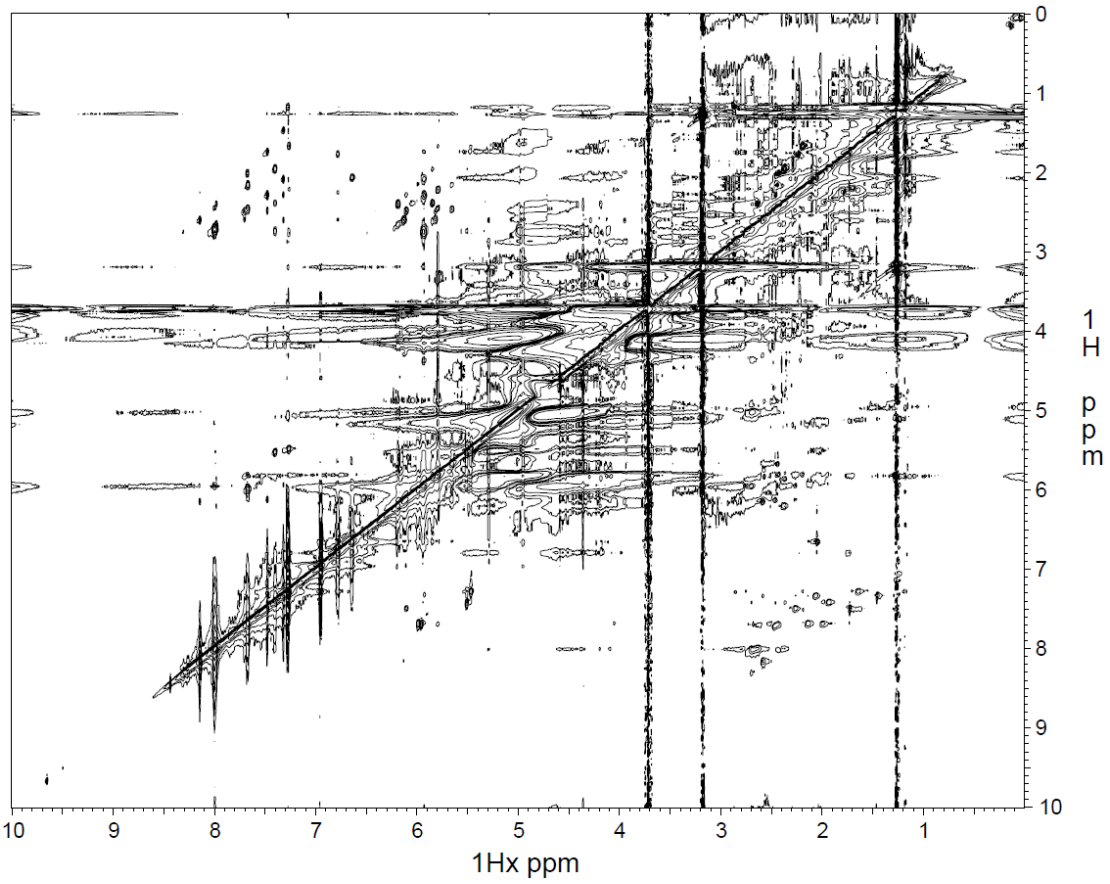


Figure 4.9 NOE spectrum of non-natural Dickerson dodecamer

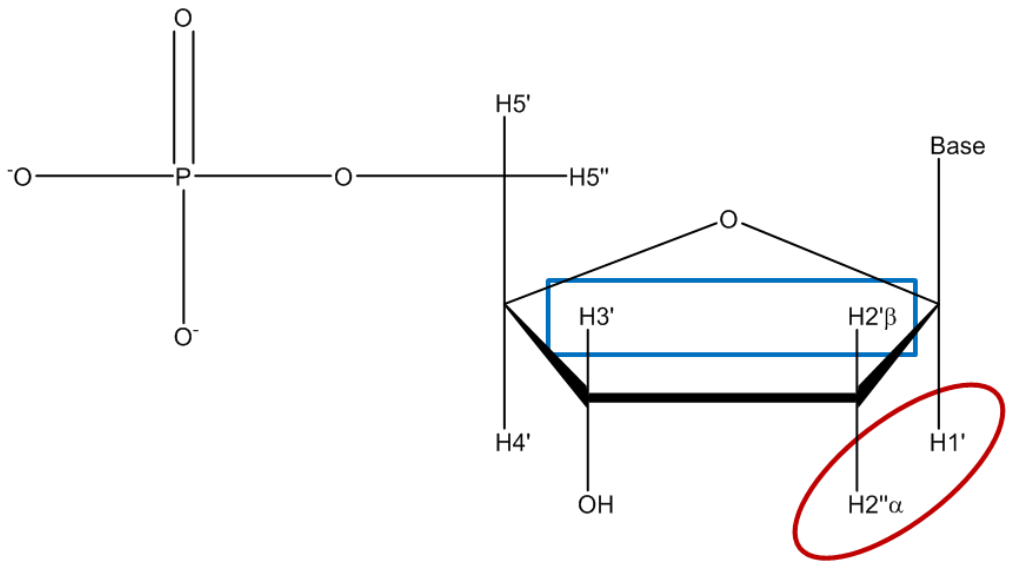


Figure 4.10 Structure of sugar and NOE difference.

Table 4.2 Chemical shifts of the protons of the tolyl dodecamer

Residues	C8H	C6H	C5H	CH_3	C1′H	C2′H	C2″H	C3′H[a]	C4′H[a]
C1		7.67	5.98		5.82	2.00	2.45	4.74	4.01
G2	8.01				5.95	2.02	2.74	4.71	4.11
C3		7.27	5.50		5.81	1.65	2.20	4.81	4.13
G4	8.00				6.13	2.32	2.46	4.92	4.03
X5		6.78	6.64	2.06	5.92	1.77	2.23	4.74	4.06
A6	8.14				6.15	2.26	2.61	4.75	4.29
T7		7.32		1.47	5.93	2.09	2.28	4.73	4.15
T8		7.48		1.75	6.12	2.26	2.48	4.73	4.06
C9		7.66	5.81		5.66	2.15	2.44	4.79	4.18
G10	7.99				5.95	2.16	2.70	4.99	4.12
C11		7.40	5.52		5.85	1.95	2.38	4.86	4.19
G12	7.98				6.20	2.39	2.65	4.70	4.08

a. Some of the assignments of C3′H and C4′H are ambiguous due to the weak intensity or overlapping.

measurement is described as follows: 1) strong interactions (1.5-2.7 Å); 2) medium interactions (1.5-3.3 Å); 3) weak interactions (1.5-4.0 Å); 4) and very weak interactions (1.5-5.5 Å).

Compared with the Dickerson dodecamer, the non-natural dodecamer contains an unnatural tolyl nucleobase. The insertion of the unnatural DNA nucleobase can distort the DNA helical structure because of the mismatches between tolyl residues and their complementary thymines. Due to the position of tolyl (position 5) and thymine (position 8) opposite to the tolyl residue, the sites surrounding these two residues are distorted dramatically. This distortion can be shown in the NOESY spectrum. For example, there is a strong cross peak between C5 methyl group of the thymine 7 and C8 proton of the adenine 6 in natural Dickerson dodecamer, however, in the tolyl dodecamer, the intensity of cross peak between C5 methyl group of the thymine 7 and C8 proton of the adenine 6 is weaker (Figure 4.11, marked in rectangle, 1 and 1′). Similarly, the coupling between C6 proton of thymine 7 and C5 methyl group of thymine 8 is strong in natural Dickerson dodecamer, but is not detected in the spectrum of non-natural Dickerson dodecamer (Figure 4.11 marked in circle, 4 and 4’). The remaining strong cross peaks belong to the intramolecular NOE signals that cannot be affected by the distortion: between methyl protons and C6 proton of the thymine 7 (2 and 2’), between methyl protons and C6 proton of the thymine 8, and between C6 proton and C2’ proton of cytosine 3, respectively (3 and 3’).

The distortion can also be clearly seen during the peak assignment of NOE signals between the purine C8H or pyrimidine C6H and the C1′H of their own and (n-1) sugars (Figure 4.12). For instance, the strong NOE signal between C1′ proton of thymine 7 and C6 proton of thymine 8 shown in Figure 4.4 and Figure 4.5 are not observed for the non-natural Dickerson dodecamer (marked in circle, 1 and 1′, in Figure 4.12). The strong NOE signal between C1′ proton of adenine 6 and C6 proton of thymine 7 shown in Figure 4.4 and Figure 4.5 becomes much weaker

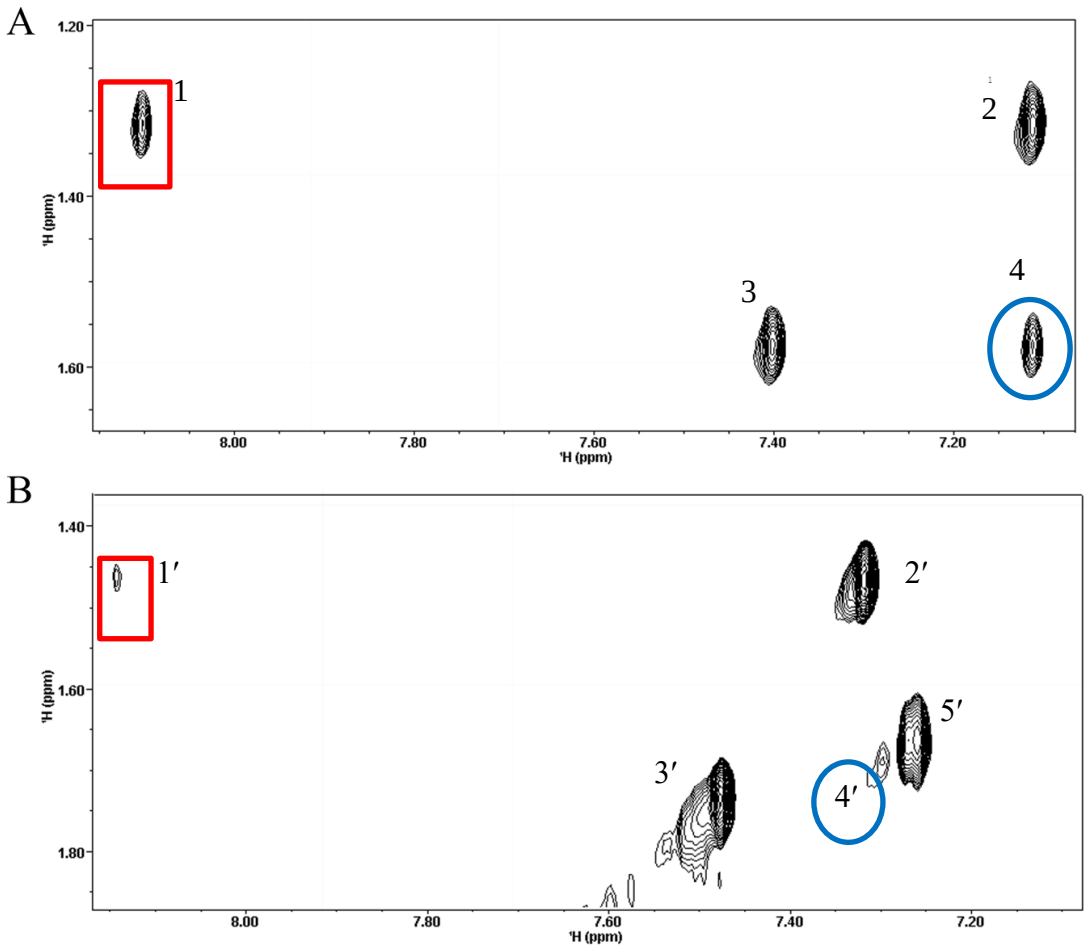


Figure 4.11 NOE signal between C5 methyl group of thymine and C6 or C8 position of the n-1 base. The top spectrum belongs to the natural Dickerson dodecamer, the bottom spectrum belongs to the non-natural Dickerson dodecamer. 1 and 1′ are the couplings between T7 CH_3 and A6 C8H; 2 and 2′ are the couplings between T7 C6H and T7 CH_3; 3 and 3′ are the couplings between T8 C6H and T8 CH_3; 4 and 4′ are the couplings between T7 CH3 and T8 CH_3; 5′ is the coupling between C3 C6H and C3 C2′H (this coupling does not appear in the top spectrum).

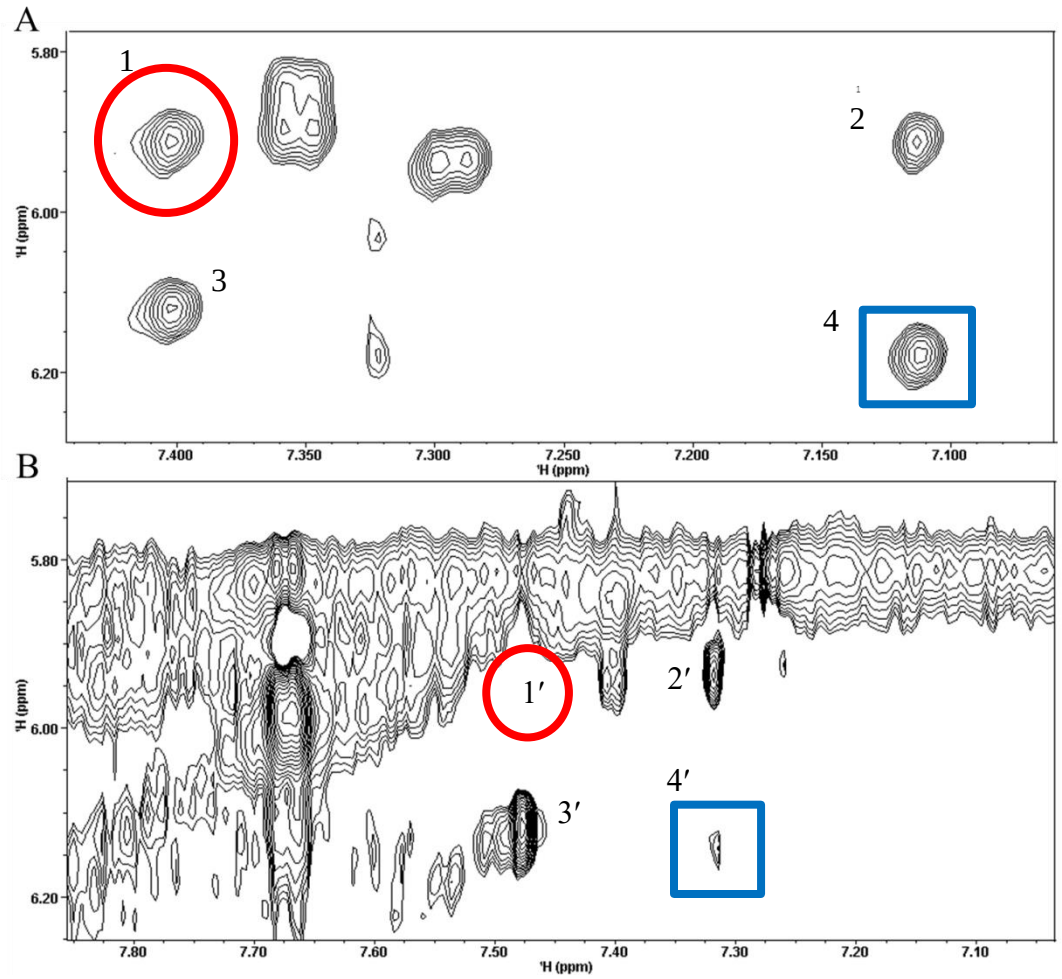


Figure 4.12 NOE signal between the purine C8H or pyrimidine C6H and the C1′H of its own sugar and (n-1) sugar. The top spectrum belongs to the natural Dickerson dodecamer, the bottom spectrum belongs to the non-natural Dickerson dodecamer. 1 and 1′ are the couplings between T7 C1′H and T8 C6H; 2 and 2′ are the couplings between T7 C6H and T7 C1′H; 3 and 3′ are the couplings between T8 C6H and T8 C1′H; 4 and 4′ are the couplings between A6 C1′H and T7 C6H.

In non-natural Dickerson dodecamer (marked in rectangle, 4 and 4′, in Figure 4.12).Intramolecular NOE signals (2, 2′, 3 and 3′) can be used as reference peaks to identify the missing or weaker peaks. Several other missing or weaker cross peaks can be discovered by comparing the NOESY spectra of the original and tolyl Dickerson dodecamer.

Although it is more difficult for the peak assignment based on the connection in the sequential NOE signal, the weaker or absent couplings can provide useful information related to the structural changes of non-natural Dickerson dodecamer. The analysis of the NOE spectra between the tolyl modified and natural duplexes suggested that introduction of tolyl modification does influence the overall structure. In order to obtain a more detailed insight into the structural perturbation induced by the incorporation of tolyl residue, the computational methods were employed as described below.

Structural calculation of Dickerson dodecamer containing aryl C-nucleoside

Currently, a preliminary struture of the tolyl-containing dodecamer was calculated by using 117 distance restraints, 28 hydrogen bond restraints and 99 dihedral angle restraints. This structure still contains many violations of dihedral angle restrains and distance restraints, but this preliminary structure is the best structure obtained from calculation using CNS and shown in Figure 4.13. Deoxyribose sugar pucker restraints determined from the J-coupling constants of DQF-COSY cross-peaks and backbone dihedral angles were estimated based on J-coupling constants of C3′H–P obtained from the ^{1}H – ^{31}P HETCOR NMR experiment that can greatly improve the quality of a solution structure.[16, 17] However, due to the limited resolution of DQF-COSY spectrum and the lack of ^{1}H – ^{31}P HETCOR probe for the 600 MHz instrument at the University of Alabama, the torsion angle restrains were estimated from previous work (angle

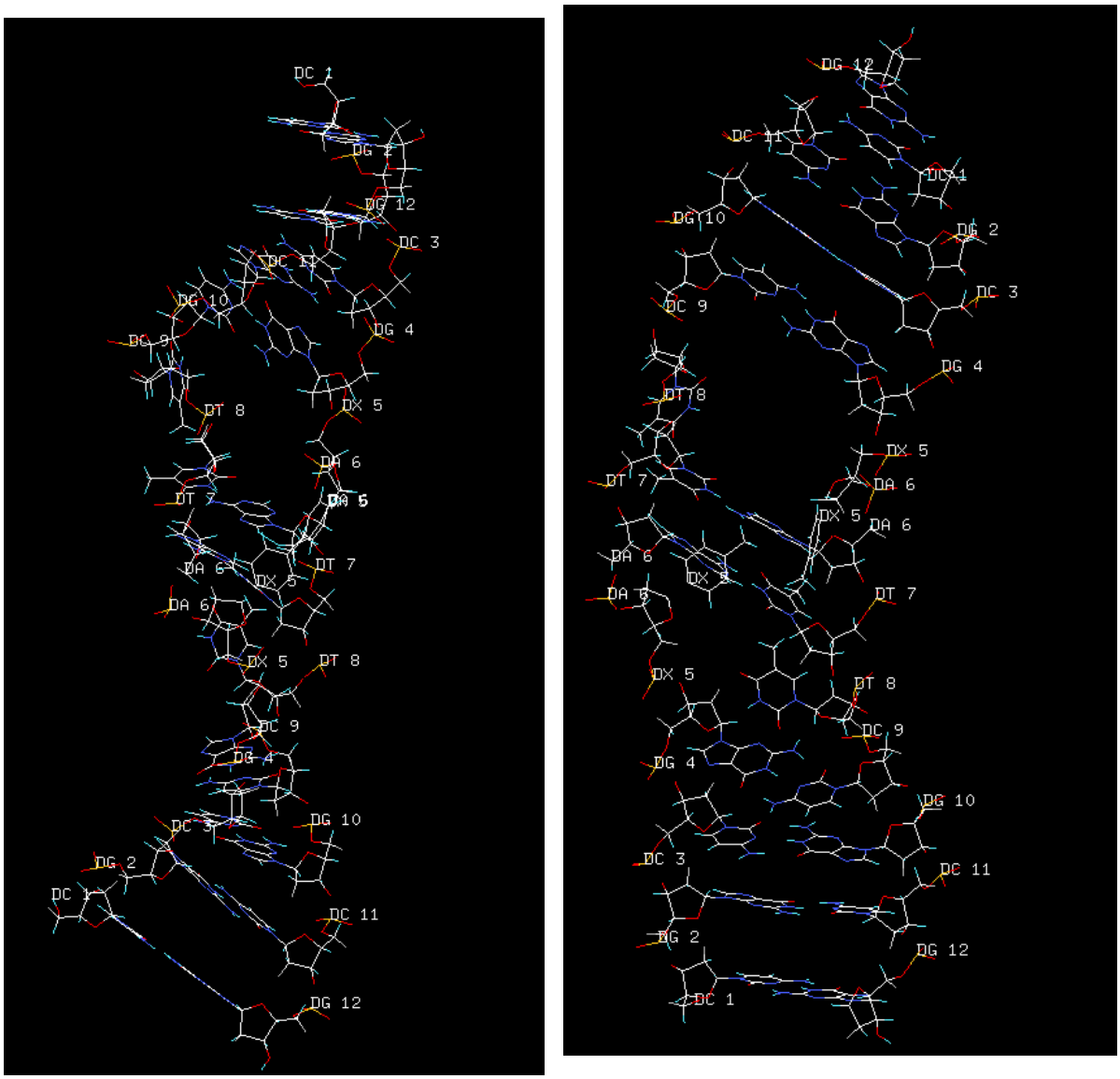


Figure 4.13 3D structure of Dickerson dodecamer containing aryl C-nucleoside calculated by CNS. The left figure is the side-view and the right figure is the front view.

restrains for non-natural base X were not added).[18, 19] The incorporation of these constraints seemed justified based upon the similarity of proton chemical shifts and COSY and NOESY spectra seen with the natural and the non-natural Dickerson dodecamers in this work. Denisov and co-workers[14] report that the J-coupling constants from COSY and ^{1}H – ^{31}P HETCOR were similar between the natural and non-natural Dickerson dodecamers except at the modified site.

Based on the examination of this preliminary structure, the structure possesses a double stranded structure with Watson-Crick base pairing except where the tolyl locates opposite to T8. Neither the tolyl residue nor T8 appear to be stacked inside the duplex. Both residues appear to be turned parallel to the helical axis. In addition, there are discontinuities in the base stacking along the structure. The disappearance of NOE signals between T7 and T8 and between X5 and G4 are likely responsible for this highly unusual structure. The distance between X5 and A6 shown in the preliminary structure is large, and this sequential NOE signal should be disconnected, but NOE connections between X5 C6H and A6 C2″H and between A6 C8H and X5 C2′H are observed. In addition, violations were observed at the non-natural base X during the structure calculation, which means the structure around X is highly uncertain and errors may be present. Lack of restraints makes violation analysis and iterative structural calculation difficult, and, as a result, the converged structure family cannot be achieved. In addition, the abnormal orientations of T8 and X5 might indicate the presence of an alternate structure.

Due to the limitations of the obtained restraints from NMR experiment and in order to get a more accurate 3D structure, the study of molecular modelling has been performed to complement the ambiguous preliminary strcuture. This molecular modeling calculation was carried out using the Amber forcefield of the Hyperchem 8.0 software package, and the calculation did not consider the effects from the solvent and counter ions (Figure 4.14A). The result shows the tolyl

residue and its complementary base both stacking inside of the duplex. Clearly the tolyl residue that points directly to the thymine cannot be accommodated in a coplanar fashion, but the tolyl residue intercalates into the space between the thymine (T8) and the neighboring guanine (G9). The significant distortion not only disrupts the hydrogen bonds between A6 and T7, but also prompts a new hydrogen bond formed between A6 and T8.

To support features of this calculated structure, the structure calculation was also carried out in Prof. David Dixon's group. The initial model used for calculations was the natural Dickerson dodecamer using NMR spectroscopy reported by Dr. Bax and co-workers[19] (PDB number 1DUF). The A5 was replaced by X5 manually, and the computational modeling was carried out using the Gaussian 09 package.[20] The geometry optimization was carried out using PM6 semiempirical calculations with dispersion corrected D3 version of Grimme's dispersion with Becke-Johnson damping (Figure 4.14 B).[21] This structure does not show significant differences compared with the structure calculated by Hyperchem. The tolyl residue and T8 are still stacked inside the duplex. The tolyl residue cannot be accomodated in a coplanar fashion but twisted about 30 °due to the mismatch.

However, conflicts between the calculated structure and the data from the NOE experiment were observed. For example, compared with the natural Dickerson dodecamer, NOE signals disappear between C9 and T8 as well as between T8 and T7 in the non-natural Dickerson dodecamer (see information of cross peak assignment in appendix). This indicates that T8 may be extending to the outside of the structure because it seems unlikely that the sequential connectivities be broken completely between T8 and its neighboring nucleobases if T8 locates inside of the duplex.

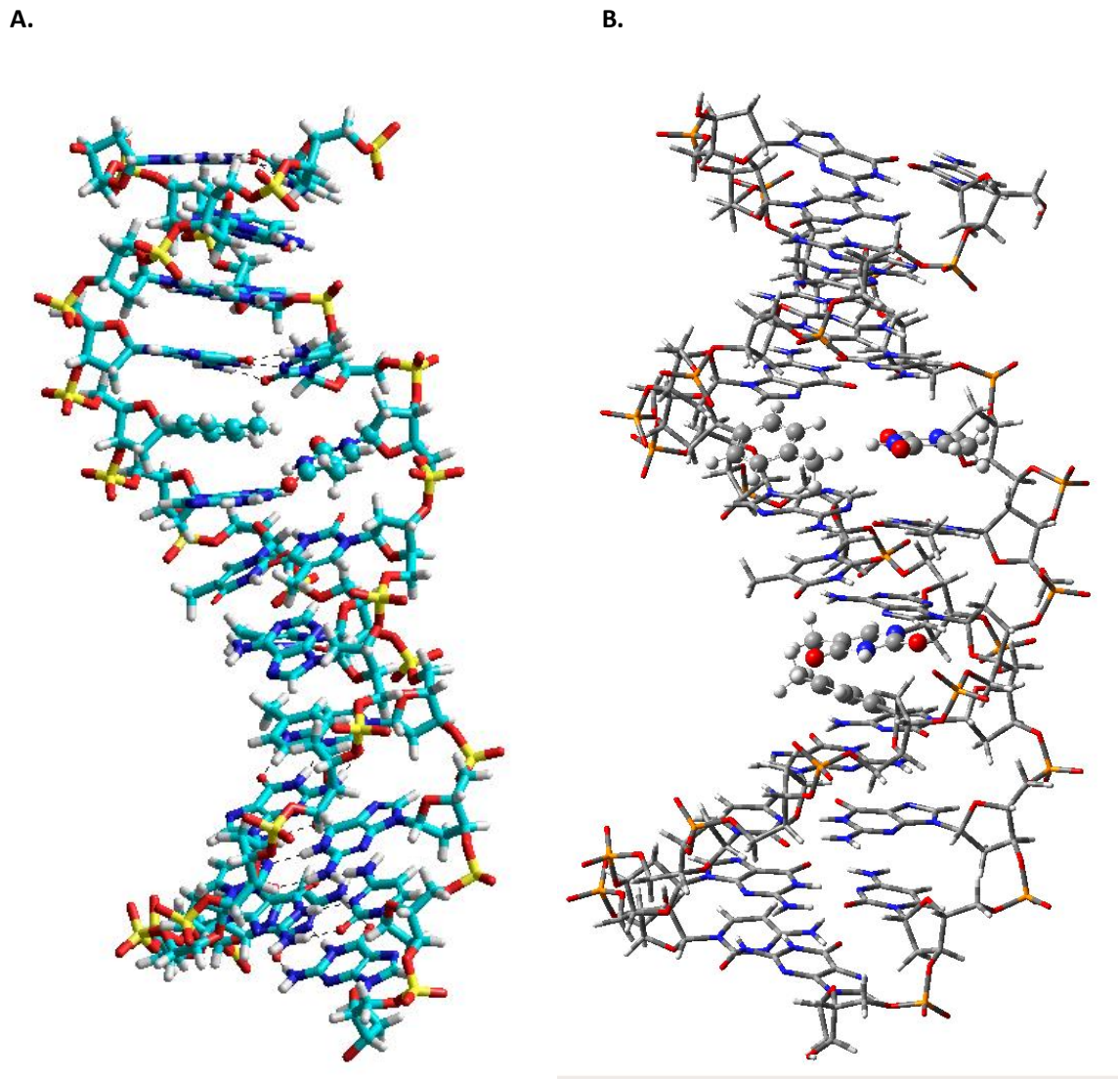
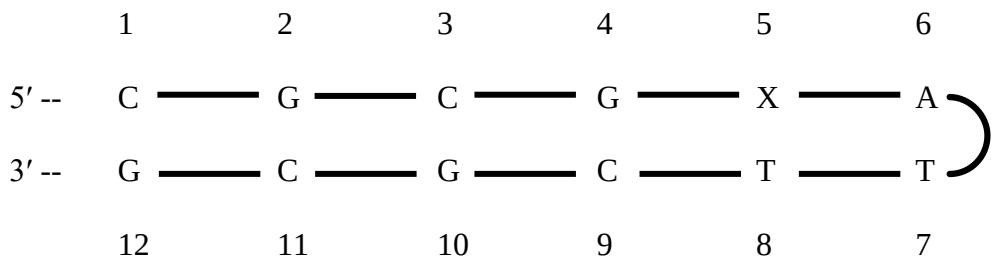


Figure 4.14 Computational method calculated model of 3D structure of Dickerson dodecamer containing aryl C-nucleoside. A is the structure of the non-natural Dickerson dodecamer calculated by Hyperchem. B is the structure of the non-natural Dickerson dodecamer calculated in Prof. Dixon's group.

Thermal denaturation study of non-natural Dickerson dodecamer containing tolyl residue and the calculation of hairpin structure.

Due to the special sequence of Dickerson dodecamer, the single-stranded tolyl dodecamer may form a hairpin structure as follows:



Although the duplex form of Dickerson dodecamer dominates in the reported literature,[14, 16-19] incorporation of the non-natural tolyl residue introduces a mismatch pair between the tolyl residue and thymine which might destabilize the stability of DNA duplex. It is entirely possible that hairpin structure becomes the major structural form because the hairpin is more stable than the duplex. Notably, Kool and coworkers[22] have reported that hydrophobic non-natural nucleoside residues can enhance the stabilities of hairpin structures.

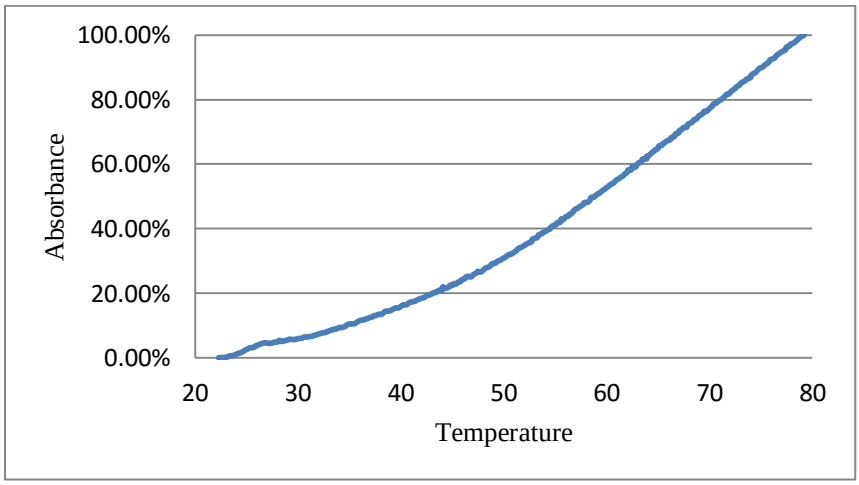
The structure difference between hairpin and duplex can be distinguished by thermal denaturation experiments. While the T_m of an intramolecular hairpin duplex is independent of concentration, the T_m is strongly influenced by oligonucleotide concentration when two oligonucleotides form a duplex.[23, 24] By varying the concentration of oligodeoxynucleotide during the thermal denaturation study, the form, duplex or hairpin, can be determined. Thermal denaturation experiments were performed for the tolyl Dickerson dodecamers (Figure 4.15). The calculated melting temperatures are all around 64 ℃. The formation of hairpin structure is

independent of the concentration of oligonucleotide because the single-stranded DNA does not need to seek its complementary strand to form a duplex.

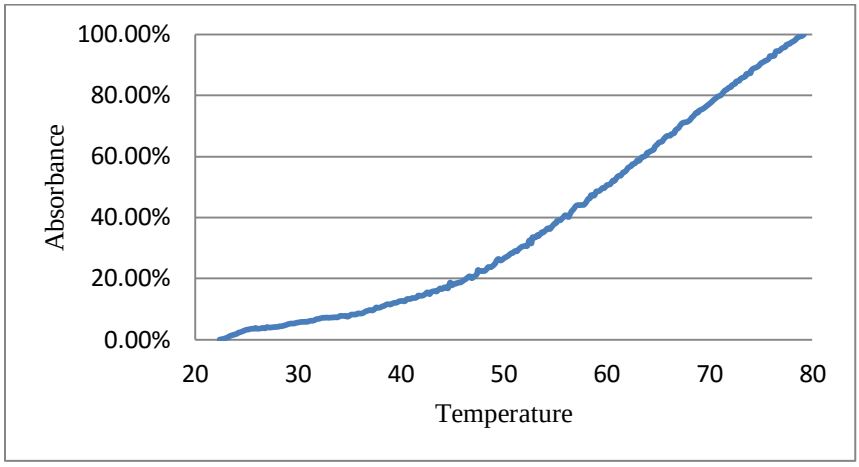
The molecular modelling calculation was performed again to construct the hairpin structure using Hyperchem 8.0 software package (Figure 4.16). The result shows the tolyl residue can be stacked inside the hairpin structure, but the thymine (T8) locates opposite to tolyl residue is in a 'flipped out' form, which means the T8 has been pushed away from the inside of the duplex into an exposed configuration. This special structure of hairpin may explain the discordant data obtained from the NOE spectrum of tolyl dodecamer. The measured distance between C1′ proton of T7 and C6 proton of T8 is 7.57 Å in a hairpin configuration, which is out of the NOE range, so there should not be a cross peak in the NOE spectrum. The sequential connectivity also breaks at the position between T8 and C9 (see information in appendix), matching the long distance between these two nucleobases shown in the calculated structure.

The tetraloop (XATT) hypothesized above is not unusual, similar structures can be found not only in RNAs, but also in single-stranded DNAs that form unimolecular hairpins.[24-26] For these hairpin structures containing the tetraloop structure, the common feature include both one or more nucleobases in the loop flipped out and a disconnection in the sequential NOE signal in the loop, especially at the site around the flipped out base. However, because the tolyl residue stacks inside the hairpin structure and might be expected to show interactions with its neighboring bases, the lack of NOEs between G4 and X5 does not support the hairpin model. While the hairpin may not be the "right" structure, the fact that the hairpin begins to explain some of the discrepancies in the NOE data for the tolyl dodecamer suggests that such a structure must be considered as a candidate for future attempts to elucidate the solution structure of this self-complementary duplex.

a. 1 μM



b. 4 μM



c. 10 μM

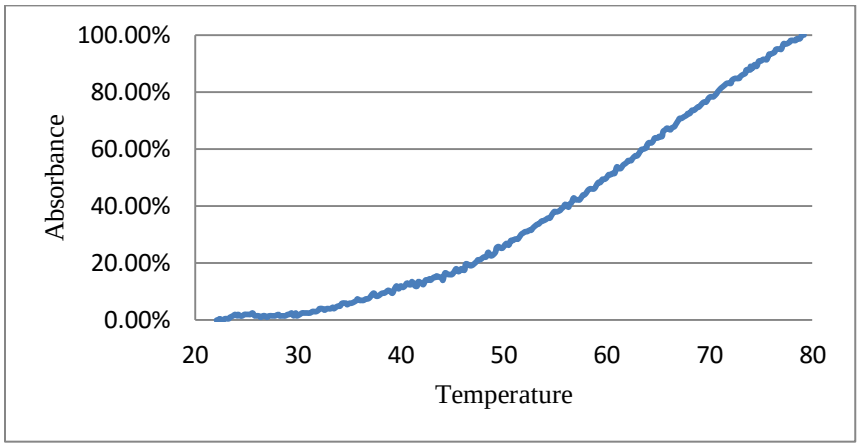


Figure 4.15 Melting curves of tolyl Dickerson dodecamer at three different concentrations.

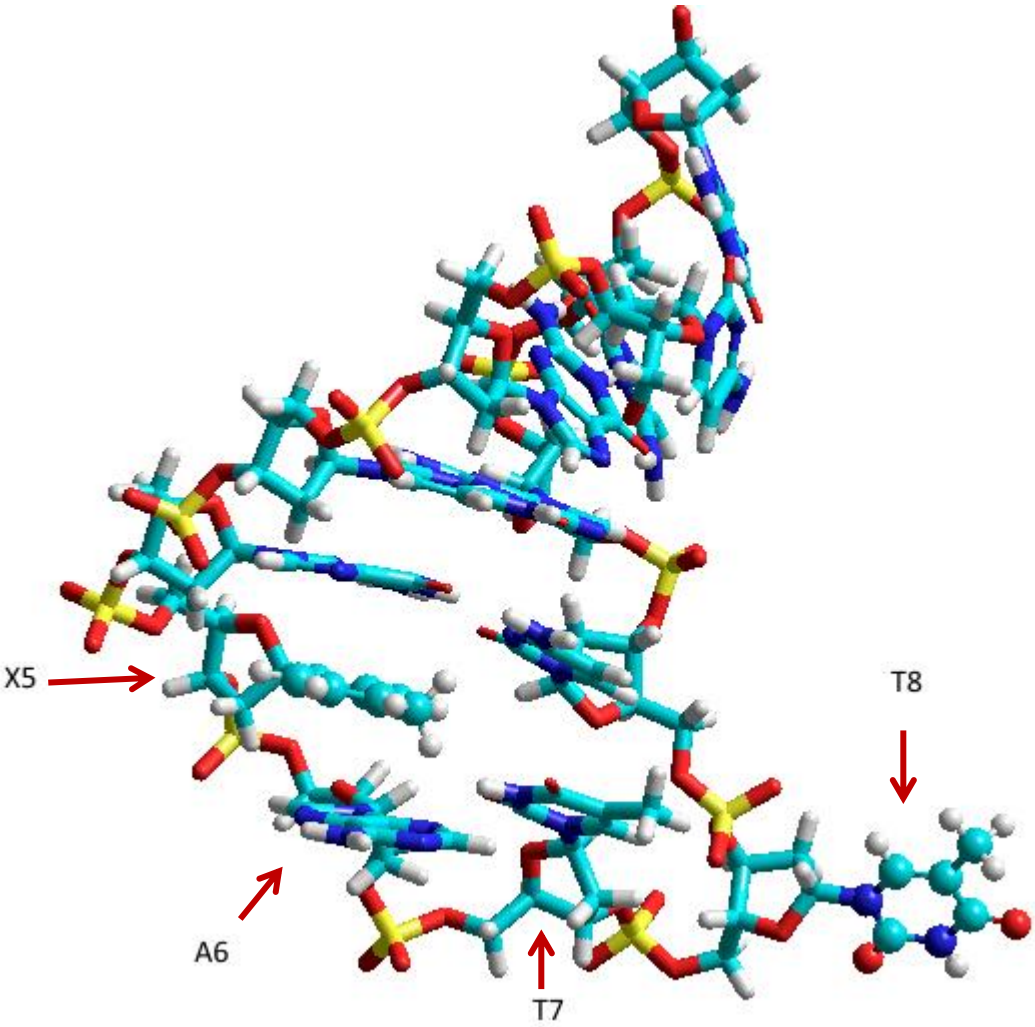


Figure 4.16 The Hyperchem-calculated hairpin structure of the tolyl dodecamer DNA.

4.3 Conclusion

The study of Dickerson dodecamer and its tolyl C-nucleoside "mutant" was performed. Several important key differences between the expected structure and the NMR data show that the real structure must significantly differ. Computational analysis of the non-natural tolyl dodecamer duplex also does not fit the data well. However, a unimolecular hairpin duplex begins to address some of the discrepancies and cannot yet be ruled out.

4.4 Experimental

General

DNA oligomers were assembled on an Applied Biosystems 391 DNA synthesizer employing phosphoramidite chemistry. The crude oligodeoxynucleotides were purified by oligonucleotide purification cartridge (OPC), preparative gel electrophoresis on denaturing 20% polyacrylamide gels or ion-exchange HPLC (Clarity 10u Oligo-wax from Phenomenex). The gel electrophoresis and subsequent electroelution were performed using 1X TBE buffer (pH 8.3, 0.089 M tris, 0.089 M borate and 0.002 M EDTA). The quantity of oligodeoxynucleotide was measured by the absorbance at 260 nm using a Beckman DU 800 UV/VIS spectrophotometer. The oligodeoxynucleotides were characterized by MALDI-TOF mass spectrometry on a Bruker Ultraflex mass spectrometer (using 2′,4′,6′-trihydroxyacetophenone monohydrate as the matrix).

General procedure for solid phase DNA synthesis and purification

Synthesis of unmodified dodecamer DNA

The unmodified Dickerson dodecamer DNA (5′-CGCGAATTCGCG-3′) was synthesized on an Applied Biosystems 391 DNA synthesizer. The oligonucleotide was prepared in a 1 μmol

scale using the standard 1 μmol synthesis cycle with the "trityl-on" end procedure. After the completion of synthesis, the oligonucleotide was treated with about 2 mL of concentrated ammonium hydroxide at 55 ℃ overnight to cleave from the controlled-pore-glass (CPG) support and remove the protective groups. The resulting mixture was allowed to cool to room temperature, and the CPG was removed by centrifugation.

The crude oligonucleotide solution was purified using a commercial reverse-phase oligonucleotide purification cartridge (OPC, Glen Research) first, followed by electrophoresis through a denaturing polyacrylamide gel. For the purification by OPC, first, 5 mL of acetonitrile was passed through the column, then 5 mL of 2 M TEAA. The ammonia solution containing the oligonucleotide was diluted with an equal volume of water. The diluted solution was passed through the OPC at a rate of about 1 drop per second, and the eluate was collected. The eluate was passed through the OPC a second time. Next, 5 mL of 1.5 M ammonium hydroxide followed by 10 mL of water was passed through the OPC. A syringe barrel was then charged with 5 mL of 3% TFA in water. This was attached to the OPC and a portion of TFA passed through to begin detritylation. After standing for 5 minutes, the remainder of the TFA solution was passed through the column, followed by 10 mL of water through the OP. The purified, detritylated oligonucleotide was eluted by slowly passing 1 mL of 20% acetonitrile through the OPC. The eluate was dried by centrifugal evaporation for the second part of purification.

Denaturing polyacrylamide gel electrophoresis purification, electroelution, desalting and lyophilization of the DNA was performed as was described in Chapter 2.

Synthesis of tolyl dodecamer DNA

A modified Dickerson dodecamer DNA containing a tolyl C-nucleoside residue at position 5 (5′-CGCG**X**ATTCGCG-3′) was synthesized on an Applied Biosystems 391 DNA synthesizer. In order to ensure that sufficient DNA was prepared to allow for NMR experiments, the oligodeoxynucleotide was prepared on a 2.5 μmol scale with a modified 1 μM DNA synthesis cycle. To accommodate the additional scale, the delivery time of base and tetrazole (step 6 to step 12 of function 19 and 90) was extended by 1 additional second per step, the delivery of base and tetrazole (function 19 and 90) was repeated 4 times instead of 3 times as original, and the wait time for coupling step was extended from 15 seconds to 30 seconds for natural nucleobases and 600 seconds for unnatural nucleobases. The cycle ending method was set as trityl-off in order to deprotect the 5′-terminus. After the completion of synthesis, cleavage and deprotection were performed using 0.05 M K_2CO_3 in methanol at room temperature overnight. The crude DNA was obtained by centrifugal evaporation.

Denaturing polyacrylamide gel electrophoresis purification, electroelution, desalting and lyophilization of the DNA was performed as was described in Chapter 2.

The presence of tris in the NMR spectra indicated that this cation, which is present from the electrophoresis and electroelution steps, was present. Most likely, this is the counterion of the DNA, and it was removed by dialysis against a sodium phosphate buffer. The DNA sample was dissolved in 0.5 mL of 10 mM sodium phosphate buffer and dialyzed against 200 volumes of 10 mM phosphate buffer for 6-18 hours using 500 MWCO membrane. The dialysis procedure was repeated for four times. The sodium form of the oligodeoxynucleotide was lyophilized and stored at -25 ℃.

NMR spectroscopy

An NMR sample of the unmodified Dickerson dodecamer was prepared by dissolving the lyophilized DNA in 0.5 mL of 10 mM phosphate buffer (pH = 7.0) in 100% D_2O (99.9%, Cambridge Isotope Laboratories, Inc.) or 90% H_2O/10% D_2O to a concentration of about 1 mM.

1D and 2D NMR spectra including COSY, TOCSY and NOESY were obtained on a Bruker AVANCE-600 with standard pulse program at temperature of 298 K. 1D experiments were performed using the pulse program zgesgp, COSY experiments were performed using the pulse program cosydfesgpph, and NOESY experiments were performed using the pulse program noesyphpr.[27-29] A recycle delay of 1.7 s was used for the COSY and 2.2 s for the NOESY spectra. The mixing time of 300 ms was used for the NOESY spectra. All data was analyzed and processed by NMRDraw and NMRPipe software package. Chemical shifts were referenced by the water peak.

Structural calculation using CNS

Structure calculation was performed on a LINUX-based computer using the CentOS operating system using CNS version 1.3. The calculation started with two extended linear single oligodeoxynucleotide strands, and then randomly selected initial atomic trajectories were assigned. The simulated annealing structure calculations were performed using a set of hydrogen bonds restrains, base planarity restrains, the unambiguous NOE restraints and the angle restrains of the phosphate backbone and the sugar ring as conformational restraints.[30,31]

Thermal denaturation study

The thermal denaturation experiments were performed by using a Beckman DU 800 UV/Vis spectrophotometer which is fitted with a T_m microcell holder (Peltier) and a computer-driven temperature controller. Sample solutions for thermal denaturation experiments were comprised of two complementary oligonucleotides in 1X PES buffer (10 mM sodium phosphate, 0.1 mM EDTA and 0.1 mM NaCl at pH = 7) with concentration of each oligodeoxynucleotide was 1, 4, or 10 μM. solutions were heated at 90 ℃ for about 5 minutes, then were slowly cooled down to room temperature to allow duplex structures to anneal. The melting temperature studies were performed in T_m cuvettes with a stopper. Moreover, the sample solutions in the cuvettes were covered by light mineral oil to reduce evaporation of sample solutions during the experiments. Absorbance was monitored at 260 nm, and the temperature in each experiment was increased from 20 to 80 ℃ at a rate of 0.4 ℃/min with a read interval of 0.2 ℃.

4.5 References:

1. Egli, M. *Curr. Protoc. Nucleic Acid Chem*. **2010** Chapter 7: Unit–7.13.

2. Flinders, J.; Dieckmann, T. *Prog. Nucl. Magn. Reson. Spectrosc.* **2006**, *48*, 137.

3. Zidek, L.; Stefl, R.; Sklenar, V. *Curr. Opin. Struct. Biol.* **2001**, *11*, 275.

4. Kwan, A. H.; Mobli, M.; Gooley, P. R.; King, G. F.; Mackay, J. P. *FEBS J.* **2011**, *278*, 687.

5. Pack, S. P.; Morimoto, H.; Makino, K.; Tajima, K.; Kanaori, K. *Nucleic Acids Res.* **2012**, *40*, 1841.

6. Zaliznyak, T.; Lukin M.; de los Santos, C. *Chem. Res. Toxicol.* **2012**, *25*, 2103.

7. Brzezinska, J.; Gdaniec, Z.; Popenda, L.; Markiewicz, W. T. *Biochim. Biophys. Acta,* **2014**, *1840*, 1163.

8. Eick, A.; Krause, F. R.; Weisz, K. *Bioconjugate Chem.* **2012**, *23*, 1127.

9. Yang, D.; Gao, Y.; Robinson, H.; Marel, G.A.; Boom, J. H.; Wang, A. H. *Biochemistry* **1993**, *32*, 8672.

10. Aue, W. P.; Bartholdi, E.; Ernst, R. R. *J. Chem. Phys.***1976**, *64*, 2229.

11. Jeener, J.; Meier, B. H.; Bachmann, P.; Ernst, R. R. *J. Chem. Phys.***1979**, *71*, 4546.

12. Drew, H. R.; Wing, R. M.; Takano, T.; Broka, C.; Tanaka, S.; Itakura, K.; Dickerson, R. E. *Proc. Natl. Acad. Sci.* **1981**, *78*, 2179; Dickerson, R. E. and Drew, H. R. *J. Mol. Biol.* **1981**, *149*, 761.

13. Hare, D. R.; Wemmer, D. E.; Chou, S .H.; Drobny, G. *J. Mol. Biol.* **1983**, *171*, 319.

14. Denisov, A. Y.; Zamaratski, E. V.; Maltseva, T. V.; Sandstrom, A.; Bekiroglu, S.; Altmann, K. H.; Egli, M.; Chattopadhyaya, J. *J. Biomol. Struct. Dyn.* **1998**, *16*, 547.

15. Davis, D.; O'Brien, E.P.; Bentzley, C. M. *Anal. Chem.* **2000**, *72*, 5092.

16. SantaLucia, J.; Allawi, T. *Nucleic Acids Res.* **1998**, *26*, 4925.

17. Markiewicz, W. T.; Brzezinska, J.; Gdaniec, Z.; Popenda, L. *Biochim. Biophys. Acta.* **2014**, 1163.

18. Sklent, V.; Bax, A. *J. Am. Chem. Soc.* **1987**, *109*, 1525.

19. Bax, A.; Lerner, L. *J. Magn. Reson.***1988**, 79, 429.

20. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A. Gaussian 09, Revision C.01, Gaussian, Inc., Wallingford CT: 2009.

21. Grimme, S.; Ehrlich, S. and Goerigk, L. *J. Comp. Chem.* **2011,** *32* 1456.

22. Ren, X.-F.; Schweitzer, B. A.; Sheils, C. J.; Kool, E. T. *Angew. Chem. Int. Ed.* **1996**, *35*, 743.

23. Owczarzy, R.; Vallone, P.M.; Gallo, F. J.; Paner, T. M.; Lane, M. J.; Benight, A. S. *Biopolymers* **1997**, *44*, 217.

24. Zhao, Q., Huang, H., Nagaswamy, U., Xia, Y., Gao, X., Fox, G. *Biopolymers* **2012**, *97*, 617.

25. Santini, G. P.; Cognet, J. A.; Xu, D.; Singarapu, K. K.; Herve du Penhoat, C. *J. Phys. Chem. B* **2009**, *113*, 6881.

26. Ghosh, M.; Vinay Kumar, N.; Varshney, U., Chary, K. V. *Nucleic Acids Res.* **2000**, *28*, 1906.

27. Hwang, T. L.; Shaka, A. J. *J Magn Reson Ser A* **1995**, *112*, 275.

28. Derome, A.; Williamson, M. *J. Magn. Reson.* **1990**, *88*, 177.

29. Shaka, A. J.; Lee, C. J.; Pines, A. *J. Magn. Reson.* **1988**, *77*, 274.

30. Wishart, D.S.; Sykes, B. D.; Richards, F. M. *Biochemistry* **1992**, *31*, 1647.

31. Brünger, A. T.; Adams, P. D.; Rice, L. M. *Structure* **1997**, *5*, 325.

CHAPTER V

Conclusions and Future Research

Four carboxamide residues were successfully incorporated into the DNA oligonucleotides. These single stranded non-natural DNA oligonucleotides were then hybridized to complementary natural DNA oligonucleotides to form the double helical complexes. The thermal denaturation studies showed that the incorporation of these non-natural residues into DNA complexes resulted in destabilizing the DNA duplex. However, the "bulged" complexes only moderately destabilized the DNA double helical, and they are the most stable complexes among three non-natural DNA complexes.

These results suggest that bulged complexes may be excellent routes to introduce non-natural residues into DNA structures. Practical applications include the use of spectroscopic labels (fluorescent, spin labels, etc.) conjugated at a carboxyamide site within a probe molecule. Selective introduction of the probe into the stacked base pairs would allow for sensitive detection of the formation of complex. In addition, these labeled carboxamide complexes would be useful probes to examine the interactions of DNA with proteins and other ligands.

Three novel aryl C-nucleosides containing 4-substituted phenyl residues were successfully synthesized by following a seven step reaction scheme. The 5′-protective group needs to be removed so that the palladium-mediated Heck reaction of iodoaromatics and deoxyribose glycal produces only the β-form product. As long as the production of non-natural DNA monomer is successful and effective, the incorporation of any kinds of aryl C-nucleoside into DNA oligonucleotide is available.

The synthesized tolyl C-nucleoside was introduced into a DNA oligonucleotide as a non-natural nucleobase. Structure calculation based on the NMR techniques was performed and a preliminary structure was obtained. Due to limited number of restraints obtained from NMR experiment, computational methods were applied to construct the molecular model and to propose a more accurate 3D DNA structure together with the NMR data. There are two possibilities of the DNA configuration, which includes duplex configuration and hairpin configuration. The results of thermal denaturation studies, molecular modeling and NMR data supported that the hairpin configuration should be the preferable structure for the non-natural DNA oligonucleotide containing tolyl residue. This result indicated the insertion of the non-natural tolyl residue destabilized the stability of DNA duplex, and as a result, a hairpin becomes the preferred structure because it is more stable than the duplex. In addition, a flipped out form of thymine residue revealed that the structure of hairpin contains a tetraloop. By performing the structure calculation, the structure changes caused by the incorporation of non-natural nucleobase can be discovered.

In order to obtain a real and accurate 3D structure of the non-natural oligonucleotide containing tolyl residue, the future research should be focused on the refinement of the preliminary structure obtained from NMR calculation. Higher concentration of non-natural DNA oligonucleotide and supplemental experiments such as ^{1}H–^{31}P HETCOR may be helpful to find more torsion angle restrains and NOE restrains.

To further study the preference of duplex configuration or hairpin configuration in a non-natural self-complementary DNA oligonucleotide, non-natural nucleobase can be introduced into different DNA sequences with self-complementary feature. The simplest way to determine whether the configuration is duplex or hairpin is to perform the thermal denaturation study. It is

also encouraged to position the non-natural nucleobase into different site in the sequence (the sites in the center of the sequence preferred), and study the structural changes caused by different modification sites.

The future research also can be focused on the structure calculation of DNA duplex containing carboxamide residue using NMR techniques. Together with the results obtained from the thermal denaturation studies described in Chapter 2, the structure and property of DNA oligonucleotide containing carboxamide residue, especially the bulged complex, can be fully investigated.

APPENDIX

Thermal Denaturation Data and Selected 1D ^{1}H NMR and 2D NMR Spectra

Natural DNA duplex

14-mer control (3'-GCAAGC T CTGTCGA-5') - (5'-CGTTCG A GACAGCT-3')				
	Tm(℃)	ΔH(kcal/mol)	ΔS(cal/K*mol)	ΔG(kcal/mol) at 298K
Run 1	58.56	-115.7	-322.7	-19.5
Run 2	58.42	-116.4	-325.1	-19.6
Run 3	59.12	-108.3	-299.9	-18.9
Average	58.70	-113.5	-315.9	-19.3
13-mer control (3'-GCAAGC - CTGTCGA-5') - (5'-CGTTCG - GACAGCT-3')				
	Tm(℃)	ΔH(kcal/mol)	ΔS(cal/K*mol)	ΔG(kcal/mol) at 298K
Run 1	58.31	-121.3	-340.0	-20.02
Run 2	57.85	-111.0	-309.3	-18.84
Run 3	57.57	-98.19	-270.8	-17.49
Average	57.91	-110.2	-306.7	-18.78

DNA duplex containing 1'-carboxamide residue of 4-chlorophenyl

14-mer (3'-GCAAGC T CTGTCGA-5') - Cl (5'-CGTTCG Cl GACAGCT-3')				
	Tm(℃)	ΔH(kcal/mol)	ΔS(cal/K*mol)	ΔG(kcal/mol) at 298K
Run 1	49.00	-98.76	-280.5	-15.17
Run 2	49.11	-95.57	-270.5	-14.97
Run 3	48.91	-102.8	-293.0	-15.45
Average	49.01	-99.03	-281.3	-15.20
13-mer (3'-GCAAGC-CTGTCGA-5') - Cl (5'-CGTTCG Cl GACAGCT-3')				
	Tm(℃)	ΔH(kcal/mol)	ΔS(cal/K*mol)	ΔG(kcal/mol) at 298K
Run 1	55.84	-104.1	-290.5	-17.58
Run 2	55.33	-93.4	-258.4	-16.45
Run 3	55.34	-102.8	-287.0	-17.31
Average	55.50	-100.1	-278.6	-17.11
Abasic (3'-GCAAGC Θ CTGTCGA-5') - Cl (5'-CGTTCG Cl GACAGCT-3')				
	Tm(℃)	ΔH(kcal/mol)	ΔS(cal/K*mol)	ΔG(kcal/mol) at 298K
Run 1	50.04	-102.5	-291.0	-15.75
Run 2	50.01	-95.5	-269.6	-15.21
Run 3	49.49	-108.3	-309.6	-16.04
Average	49.85	-102.1	-290.1	-15.67

DNA duplex containing 1'-carboxamide residue of 4-nitrophenyl

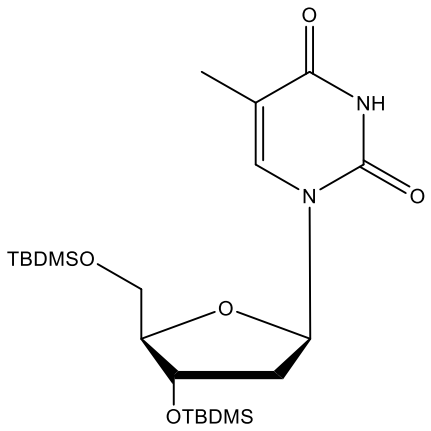
14-mer (3'-GCAAGC T CTGTCGA-5') - NO_2 (5'-CGTTCG NO_2 GACAGCT-3')				
	Tm(℃)	ΔH(kcal/mol)	ΔS(cal/K*mol)	ΔG(kcal/mol) at 298K
Run 1	51.04	-84.05	-233.2	-14.56
Run 2	51.07	-100.8	-284.8	-15.92
Run 3	51.43	-92.28	-258.2	-15.33
Average	51.18	-92.37	-258.7	-15.27
13-mer (3'-GCAAGC-CTGTCGA-5') - NO_2 (5'-CGTTCG NO_2 GACAGCT-3')				
	Tm(℃)	ΔH(kcal/mol)	ΔS(cal/K*mol)	ΔG(kcal/mol) at 298K
Run 1	56.50	-89.57	-245.7	-16.37
Run 2	56.25	-99.62	-276.3	-17.27
Run 3	56.51	-93.89	-258.7	-16.79
Average	56.42	-94.36	-260.2	-16.81
Abasic (3'-GCAAGC Θ CTGTCGA-5') - NO_2 (5'-CGTTCG NO_2 GACAGCT-3')				
	Tm(℃)	ΔH(kcal/mol)	ΔS(cal/K*mol)	ΔG(kcal/mol) at 298K
Run 1	52.51	-89.19	-247.8	-15.35
Run 2	52.51	-104.6	-295.1	-16.65
Run 3	52.47	-92.07	-256.7	-15.58
Average	52.50	-95.29	-266.5	-15.86

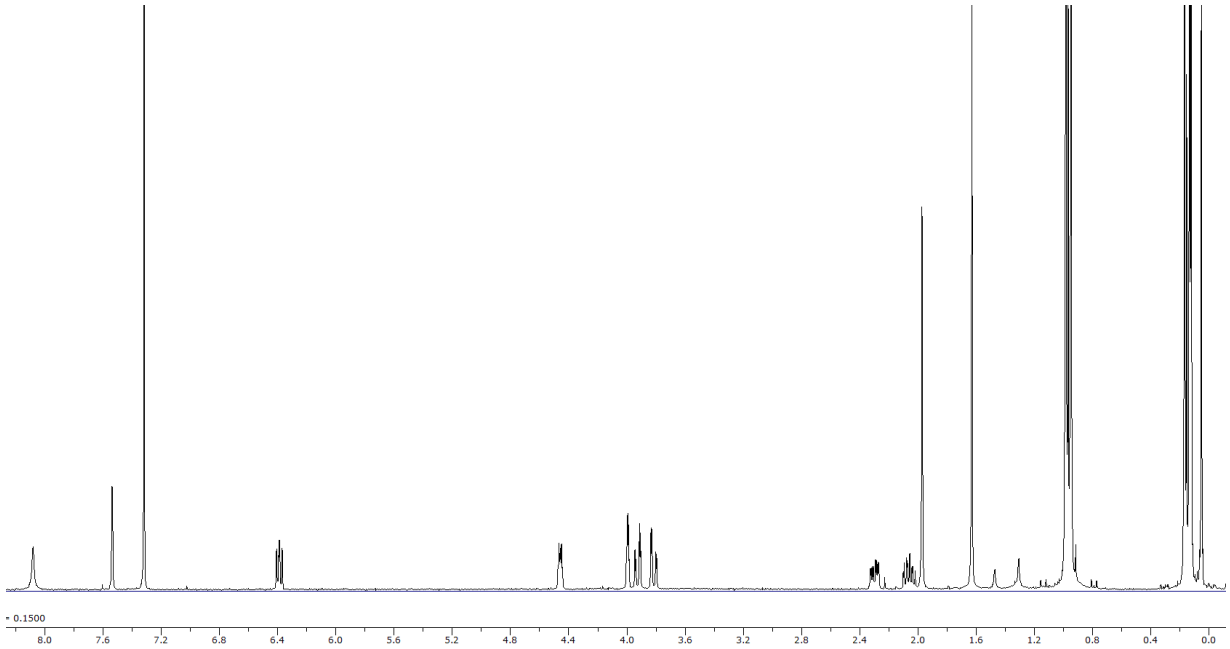
DNA duplex containing 1'-carboxamide residue of 4-cyanophenyl

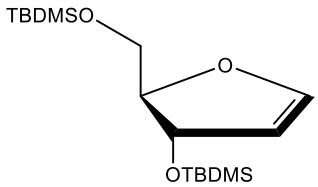
14-mer (3'-GCAAGC T CTGTCGA-5') - CN (5'-CGTTCG CN GACAGCT-3')				
	Tm(℃)	ΔH(kcal/mol)	ΔS(cal/K*mol)	ΔG(kcal/mol) at 298K
Run 1	50.02	-105.9	-301.5	-16.01
Run 2	50.22	-99.45	-281.5	-15.57
Run 3	50.79	-105.6	-299.8	-16.22
Run 4	50.13	-109.3	-312.0	-16.32
Average	50.29	-105.0	-298.7	-16.03
13-mer (3'-GCAAGC-CTGTCGA-5') - CN (5'-CGTTCG CN GACAGCT-3')				
	Tm(℃)	ΔH(kcal/mol)	ΔS(cal/K*mol)	ΔG(kcal/mol) at 298K
Run 1	55.00	-104.5	-292.3	-17.37
Run 2	56.29	-104.3	-290.4	-17.72
Run 3	55.29	-120.2	-339.8	-18.90
Run 4	55.43	-115.1	-324.2	-18.48
Average	55.50	-111.0	-311.7	-18.12
Abasic (3'-GCAAGC Θ CTGTCGA-5') - CN (5'-CGTTCG CN GACAGCT-3')				
	Tm(℃)	ΔH(kcal/mol)	ΔS(cal/K*mol)	ΔG(kcal/mol) at 298K
Run 1	53.46	-104.2	-293.0	-16.90
Run 2	52.16	-89.73	-249.8	-15.30
Run 3	52.72	-117.1	-333.3	-17.78
Run 4	51.99	-110.8	-314.7	-17.02
Average	52.58	-105.5	-297.7	-16.75

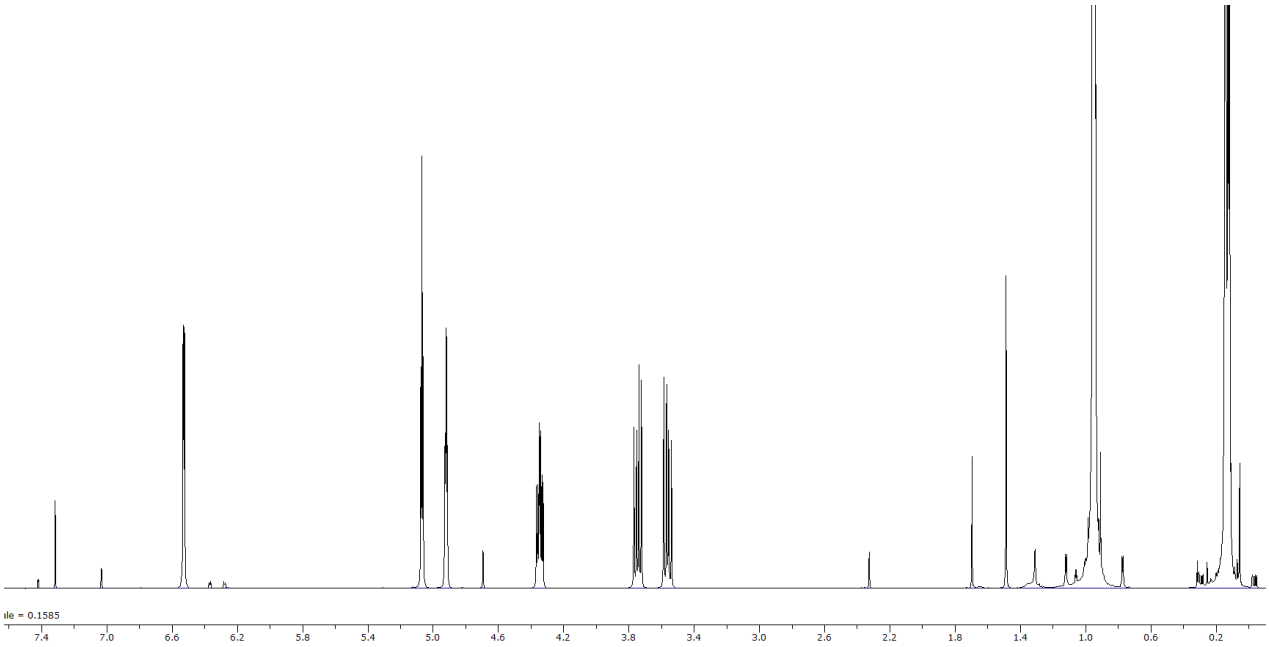
DNA duplex containing 1'-carboxamide residue of 4-methoxyphenyl

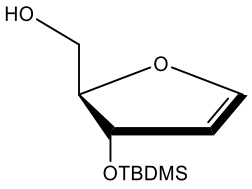
14-mer (3'-GCAAGC T CTGTCGA-5') – OCH_3 (5'-CGTTCG OCH_3 GACAGCT-3')				
	Tm(℃)	ΔH(kcal/mol)	ΔS(cal/K*mol)	ΔG(kcal/mol) at 298K
Run 1	47.40	-98.69	-281.8	-14.71
Run 2	47.52	-87.72	-247.5	-13.97
Run 3	46.87	-97.48	-278.5	-14.48
Average	47.26	-94.63	-269.3	-14.39
13-mer (3'-GCAAGC-CTGTCGA-5') - OCH_3 (5'-CGTTCG OCH_3 GACAGCT-3')				
	Tm(℃)	ΔH(kcal/mol)	ΔS(cal/K*mol)	ΔG(kcal/mol) at 298K
Run 1	52.97	-104.6	-294.6	-16.79
Run 2	53.27	-94.12	-262.3	-15.97
Run 3	52.85	-107.3	-303.1	-16.99
Average	53.03	-102.0	-286.7	-16.58
Abasic (3'-GCAAGC Θ CTGTCGA-5') - OCH_3 (5'-CGTTCG OCH_3 GACAGCT-3')				
	Tm(℃)	ΔH(kcal/mol)	ΔS(cal/K*mol)	ΔG(kcal/mol) at 298K
Run 1	47.32	-99.30	-283.8	-14.73
Run 2	47.69	-90.11	-254.8	-14.19
Run 3	46.77	-100.3	-287.5	-14.64
Average	47.26	-96.58	-275.4	-14.52

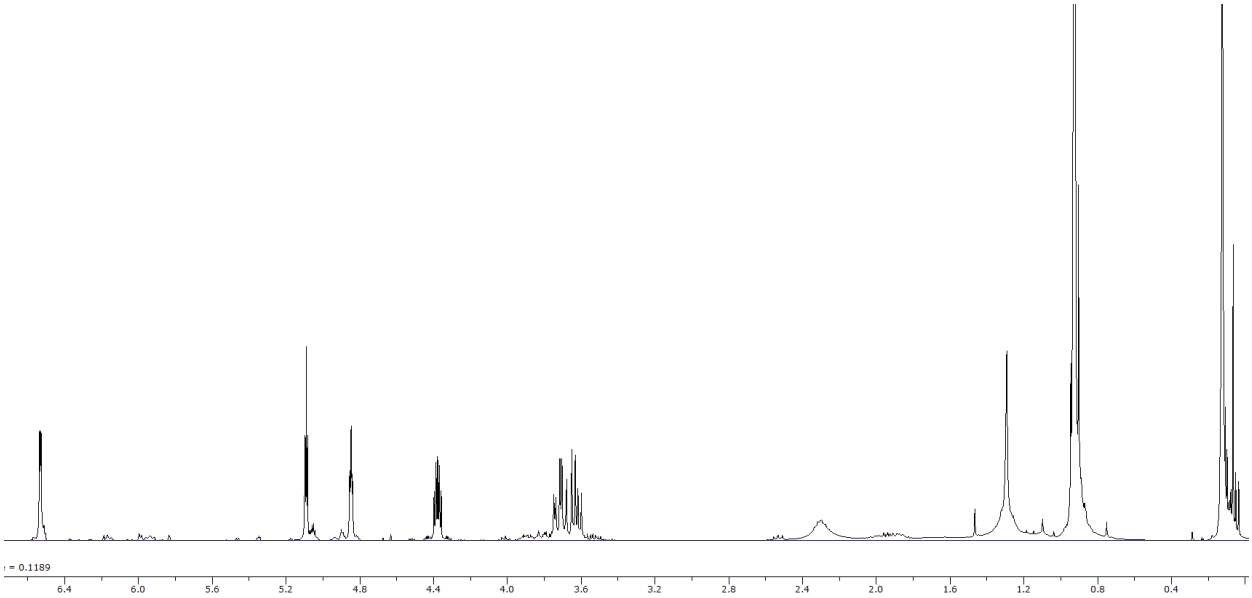

O
NH
N
O
TBDMSO
O
OTBDMS

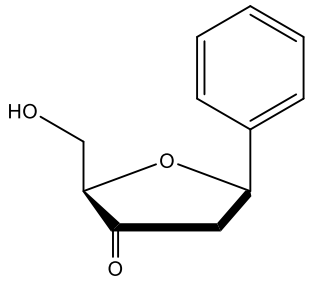

0.1500
8.0
7.6
7.2
6.8
6.4
6.0
5.6
5.2
4.8
4.4
4.0
3.6
3.2
2.8
2.4
2.0
1.6
1.2
0.8
0.4
0.0

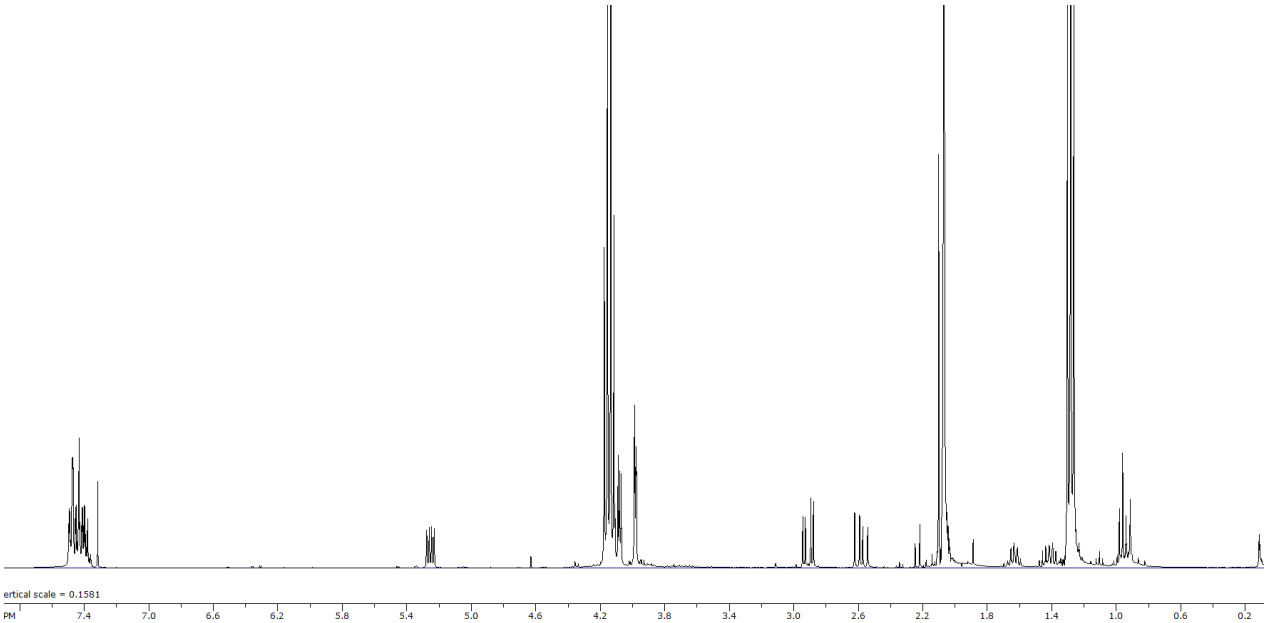

TBDMSO
O
OTBDMS

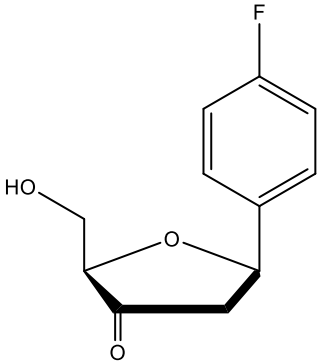

le = 0.1585
7.4
7.0
6.6
6.2
5.8
5.4
5.0
4.6
4.2
3.8
3.4
3.0
2.6
2.2
1.8
1.4
1.0
0.6
0.2

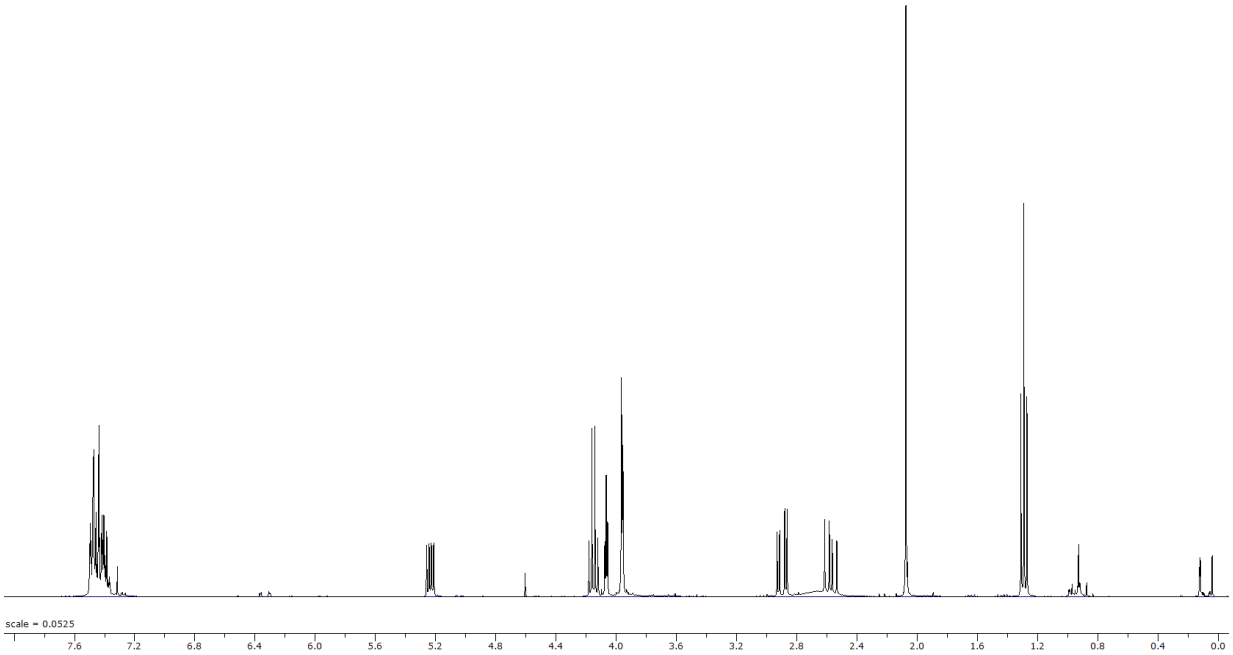

HO
O
OTBDMS

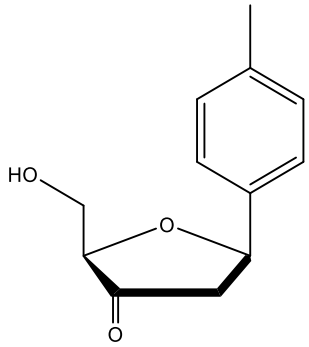

= 0.1189
6.4
6.0
5.6
5.2
4.8
4.4
4.0
3.6
3.2
2.8
2.4
2.0
1.6
1.2
0.8
0.4

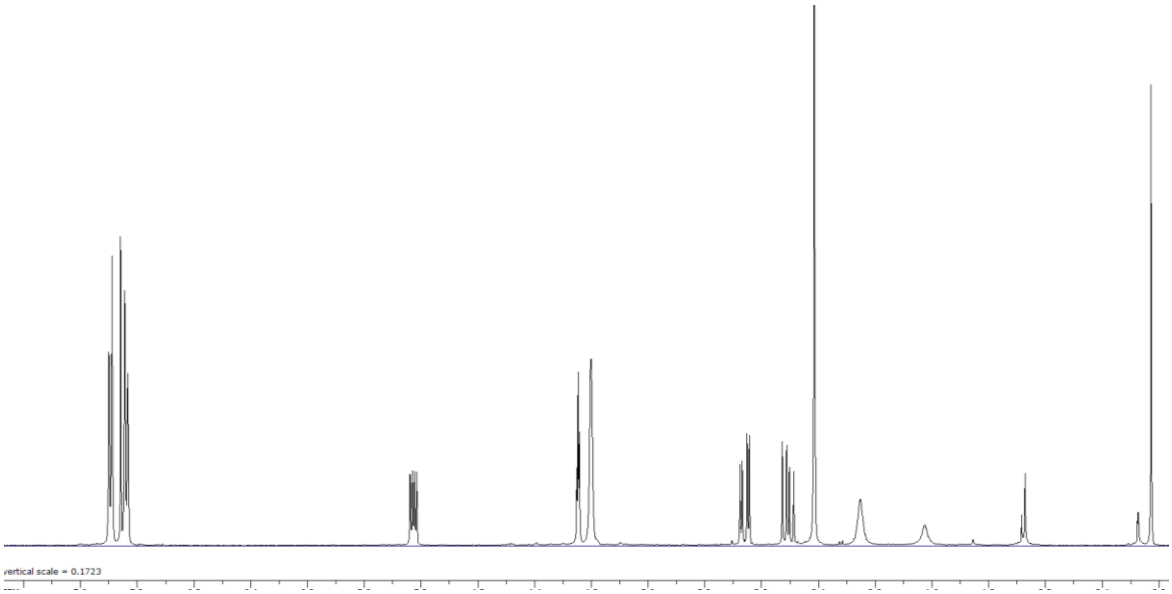

HO
O
O

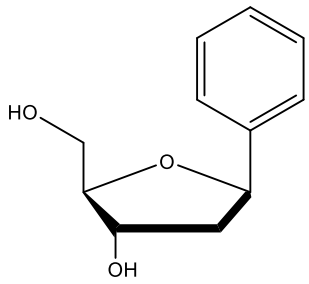

ertical scale = 0.1581
PM
7.4
7.0
6.6
6.2
5.8
5.4
5.0
4.6
4.2
3.8
3.4
3.0
2.6
2.2
1.8
1.4
1.0
0.6
0.2

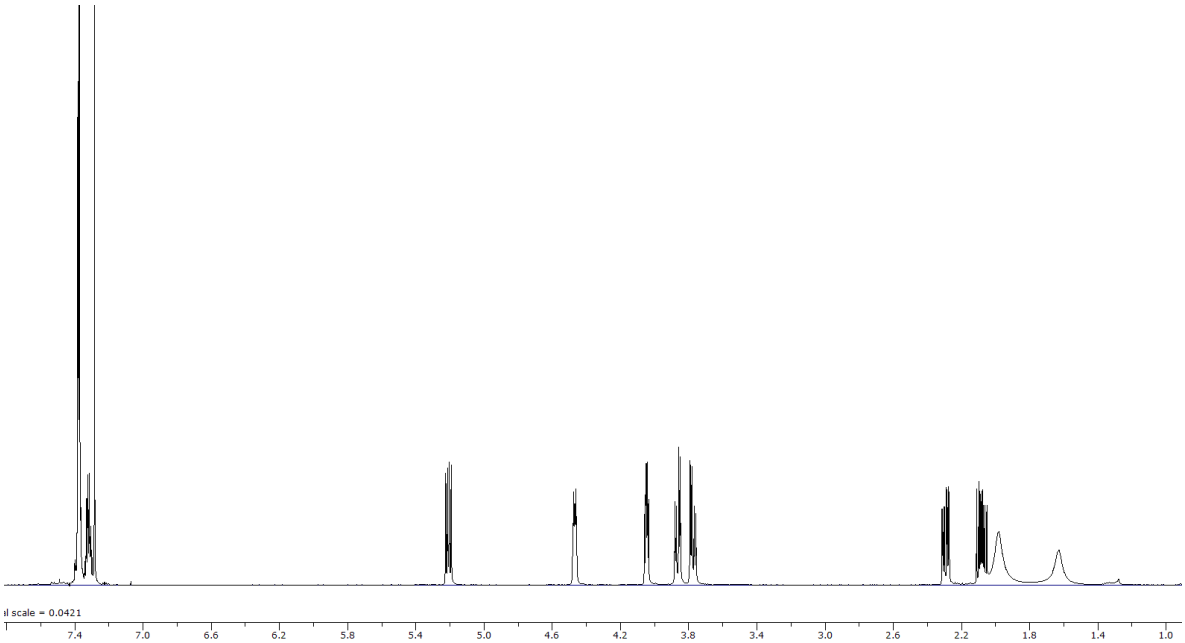

F
HO
O
O

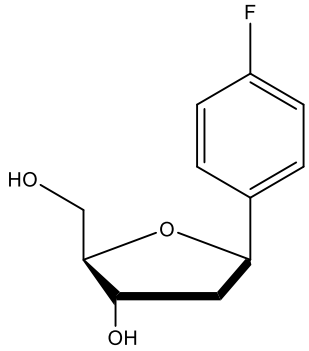

scale = 0.0525
7.6
7.2
6.8
6.4
6.0
5.6
5.2
4.8
4.4
4.0
3.6
3.2
2.8
2.4
2.0
1.6
1.2
0.8
0.4
0.0

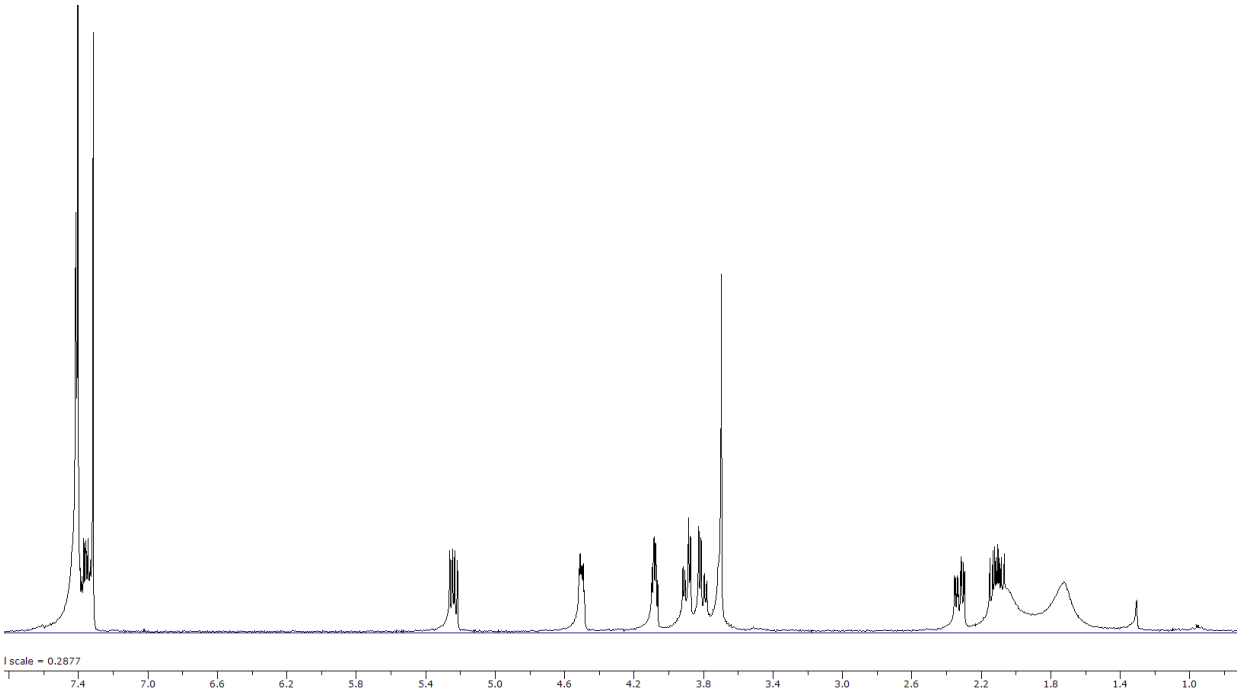

HO
O
O

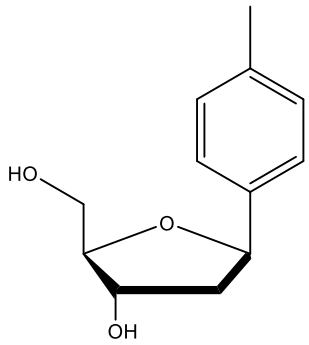

vertical scale = 0.1723

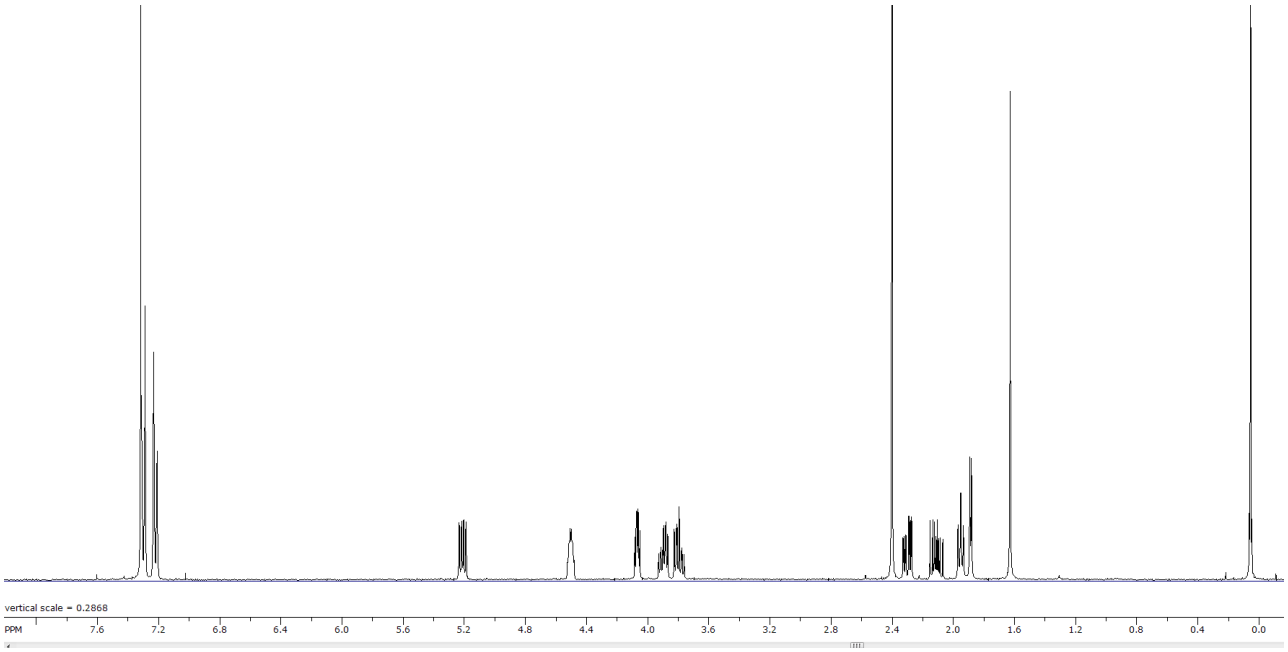

HO
O
OH

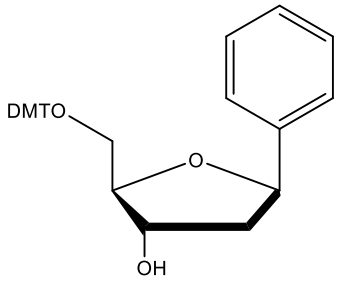

al scale = 0.0421
7.4
7.0
6.6
6.2
5.8
5.4
5.0
4.6
4.2
3.8
3.4
3.0
2.6
2.2
1.8
1.4
1.0

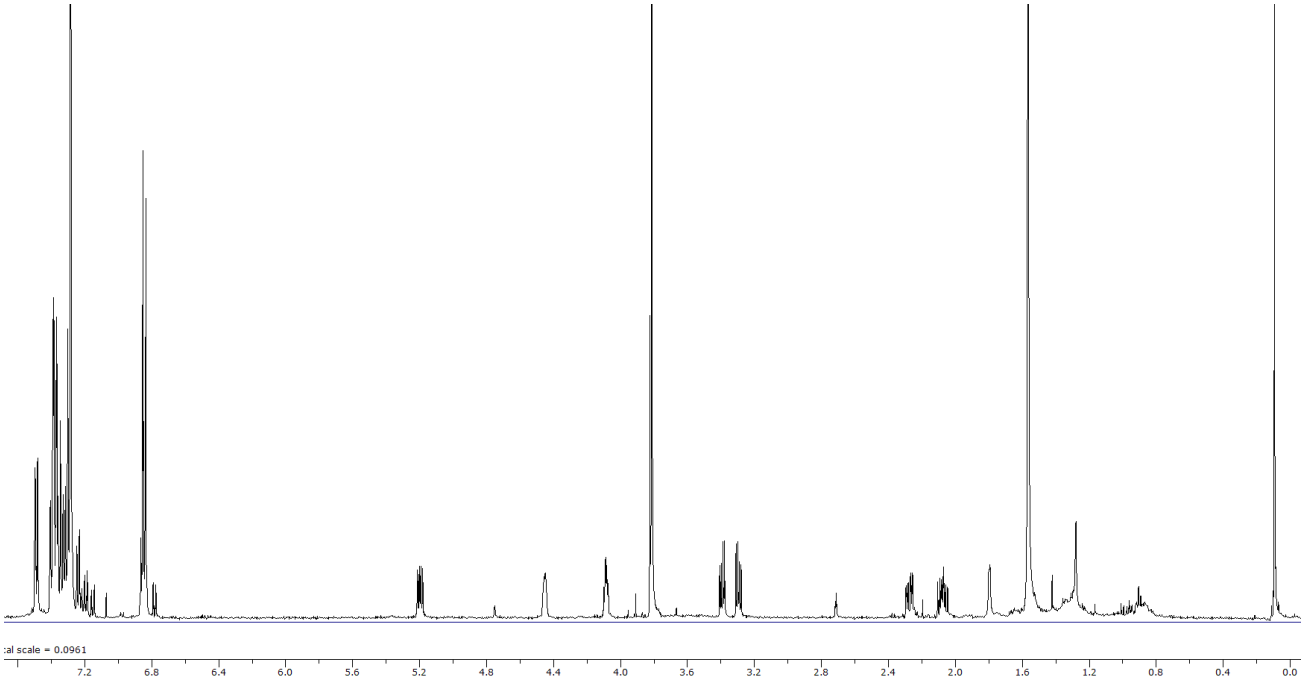

F
HO
O
OH

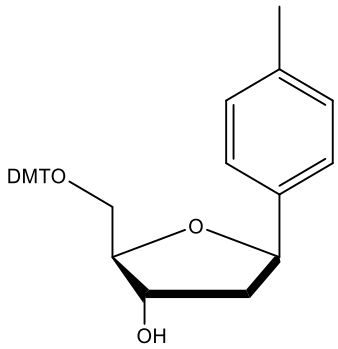

I scale = 0.2877
7.4
7.0
6.6
6.2
5.8
5.4
5.0
4.6
4.2
3.8
3.4
3.0
2.6
2.2
1.8
1.4
1.0

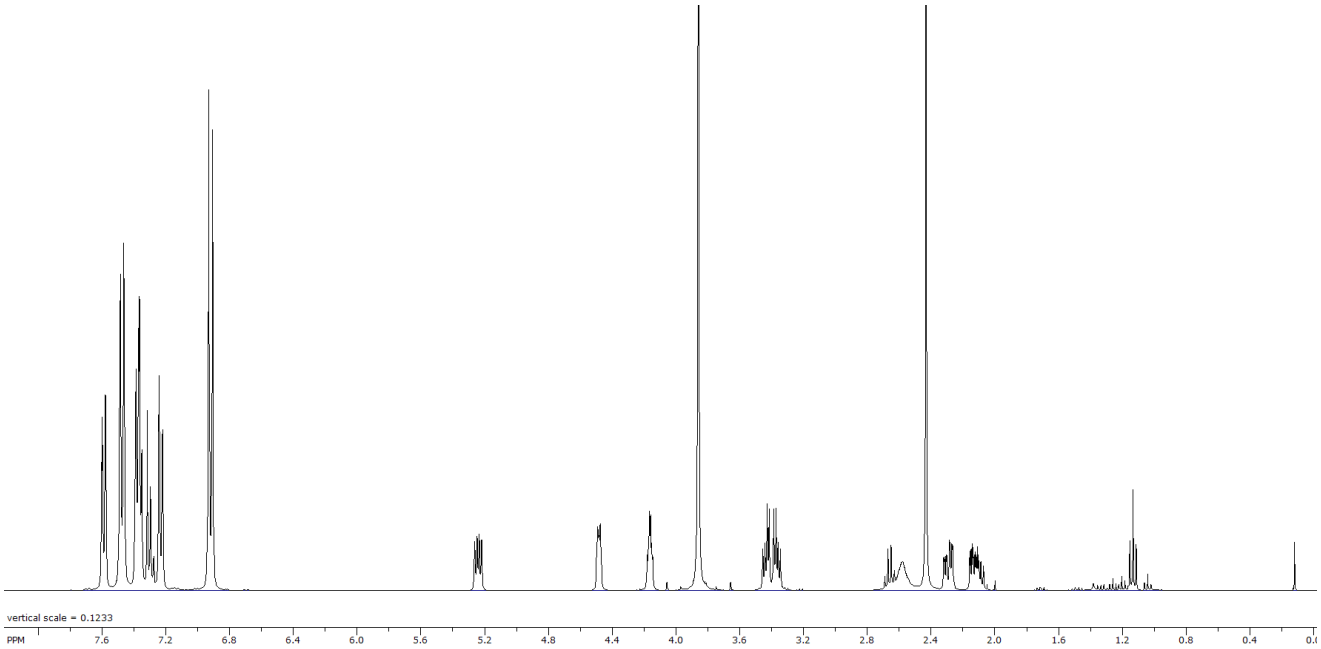

HO
O
OH

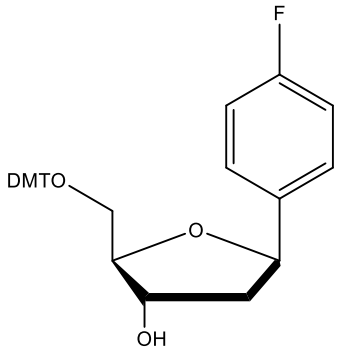

vertical scale = 0.2868
PPM
7.6
7.2
6.8
6.4
6.0
5.6
5.2
4.8
4.4
4.0
3.6
3.2
2.8
2.4
2.0
1.6
1.2
0.8
0.4
0.0

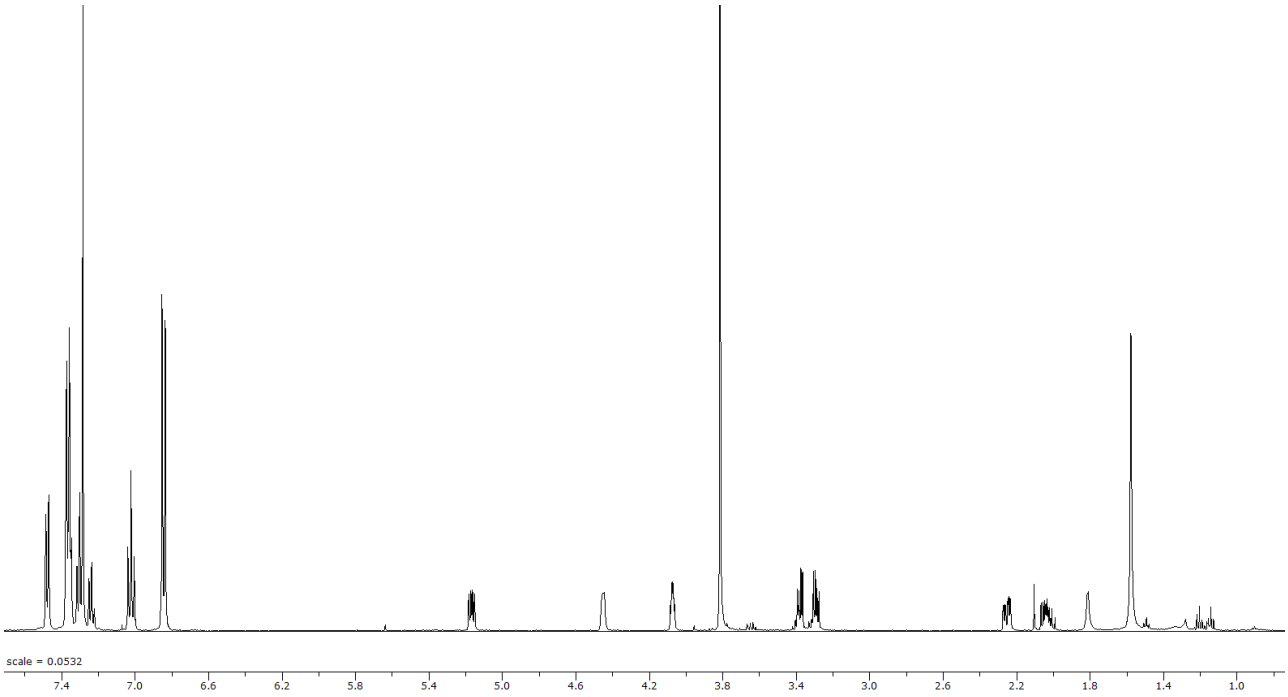

DMTO
O
OH

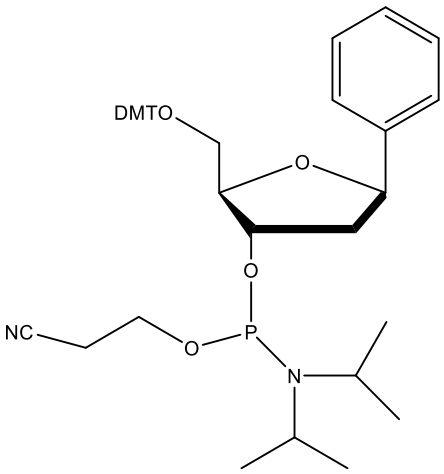

:al scale = 0.0961
7.2
6.8
6.4
6.0
5.6
5.2
4.8
4.4
4.0
3.6
3.2
2.8
2.4
2.0
1.6
1.2
0.8
0.4
0.0

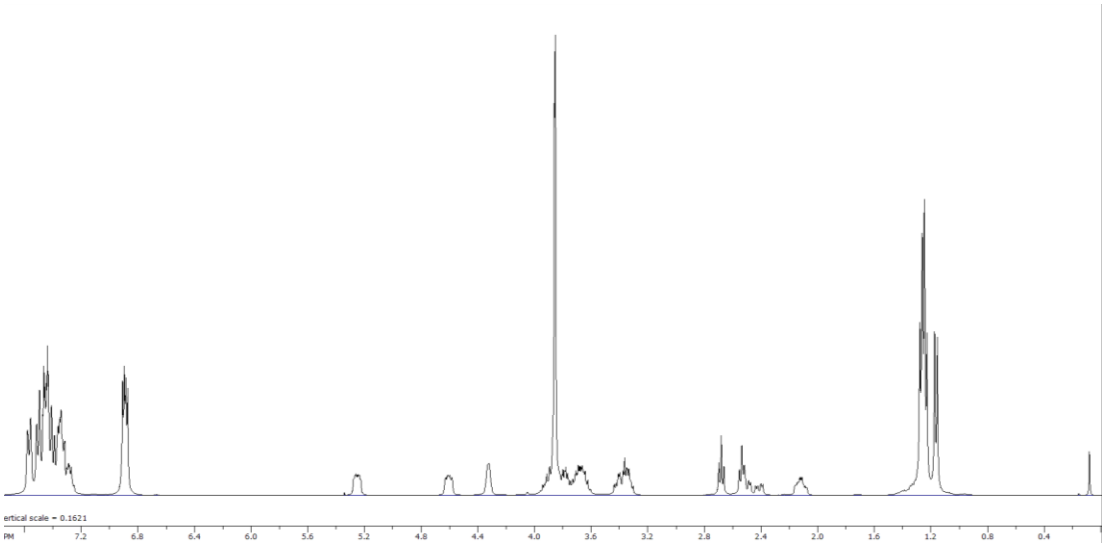

DMTO
O
OH

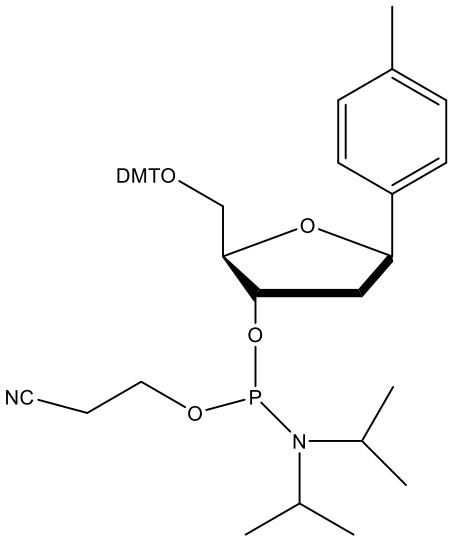

vertical scale = 0.1233
PPM
7.6
7.2
6.8
6.4
6.0
5.6
5.2
4.8
4.4
4.0
3.6
3.2
2.8
2.4
2.0
1.6
1.2
0.8
0.4

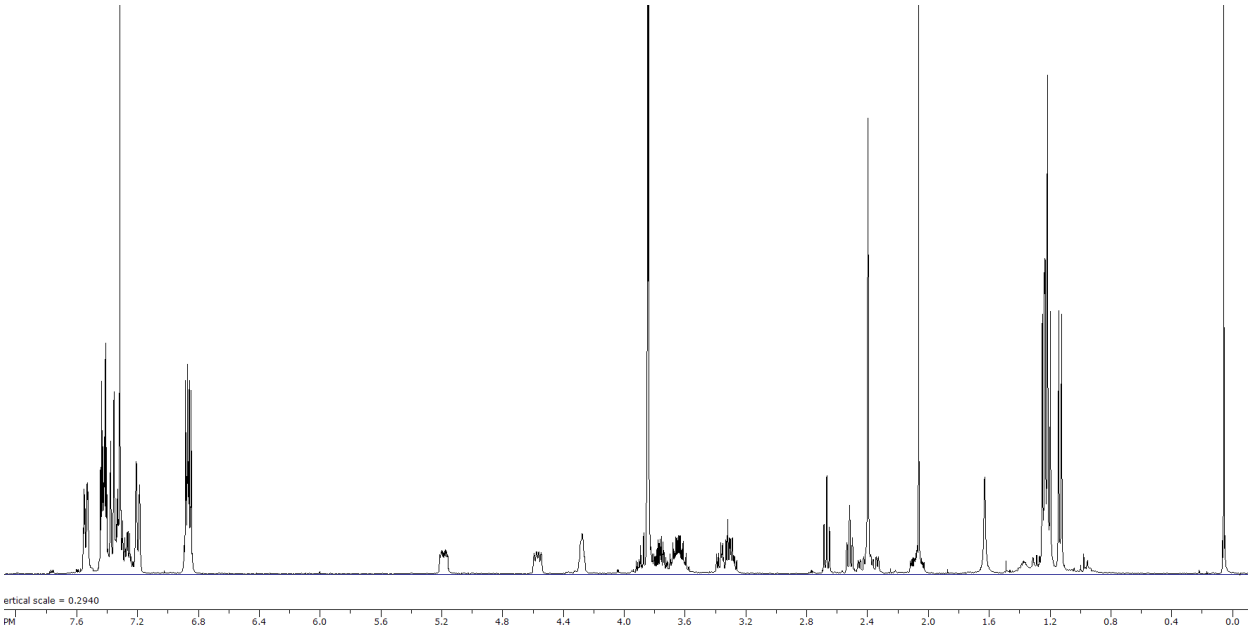

F
DMTO
O
OH

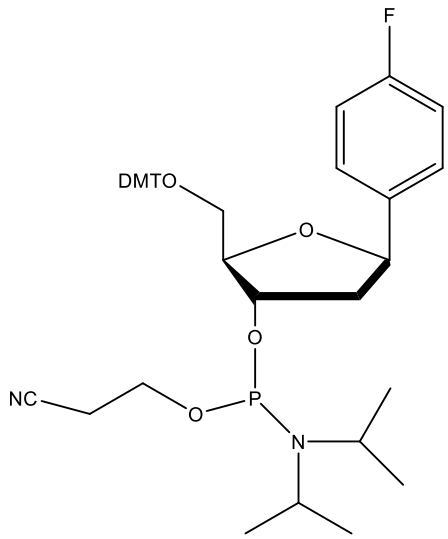

scale = 0.0532
7.4
7.0
6.6
6.2
5.8
5.4
5.0
4.6
4.2
3.8
3.4
3.0
2.6
2.2
1.8
1.4
1.0

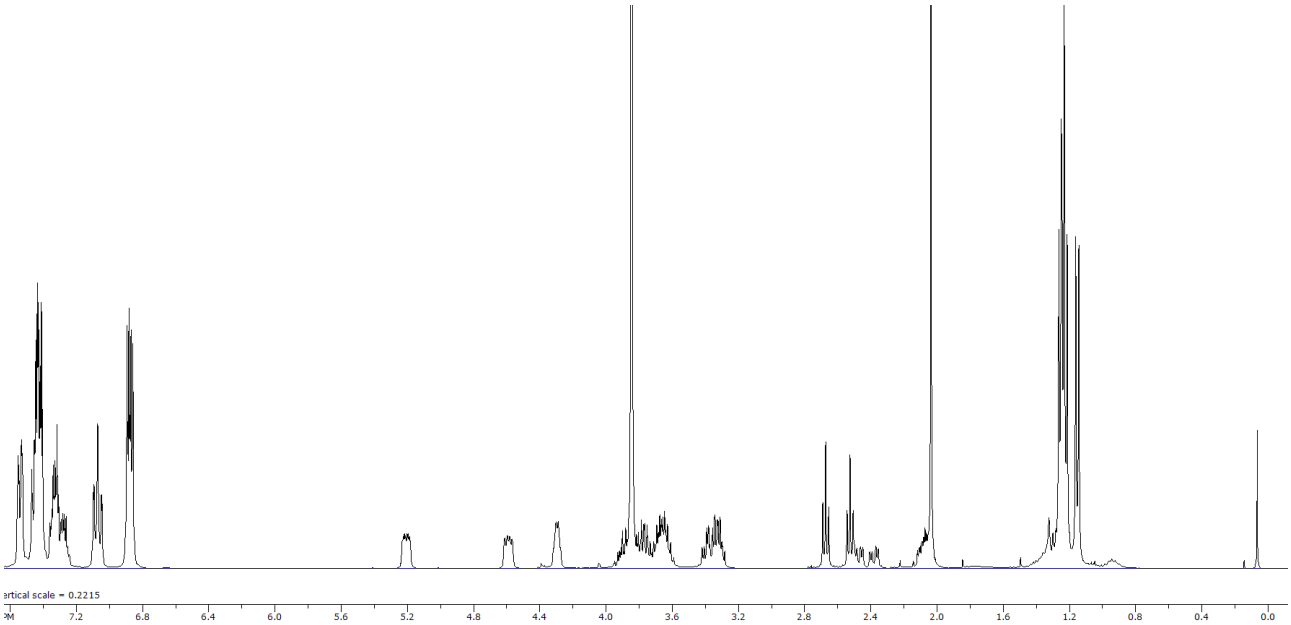

DMTO
O
O
NC
O
P
N


ertical scale = 0.1621
PM
7.2
6.8
6.4
6.0
5.6
5.2
4.8
4.4
4.0
3.6
3.2
2.8
2.4
2.0
1.6
1.2
0.8
0.4

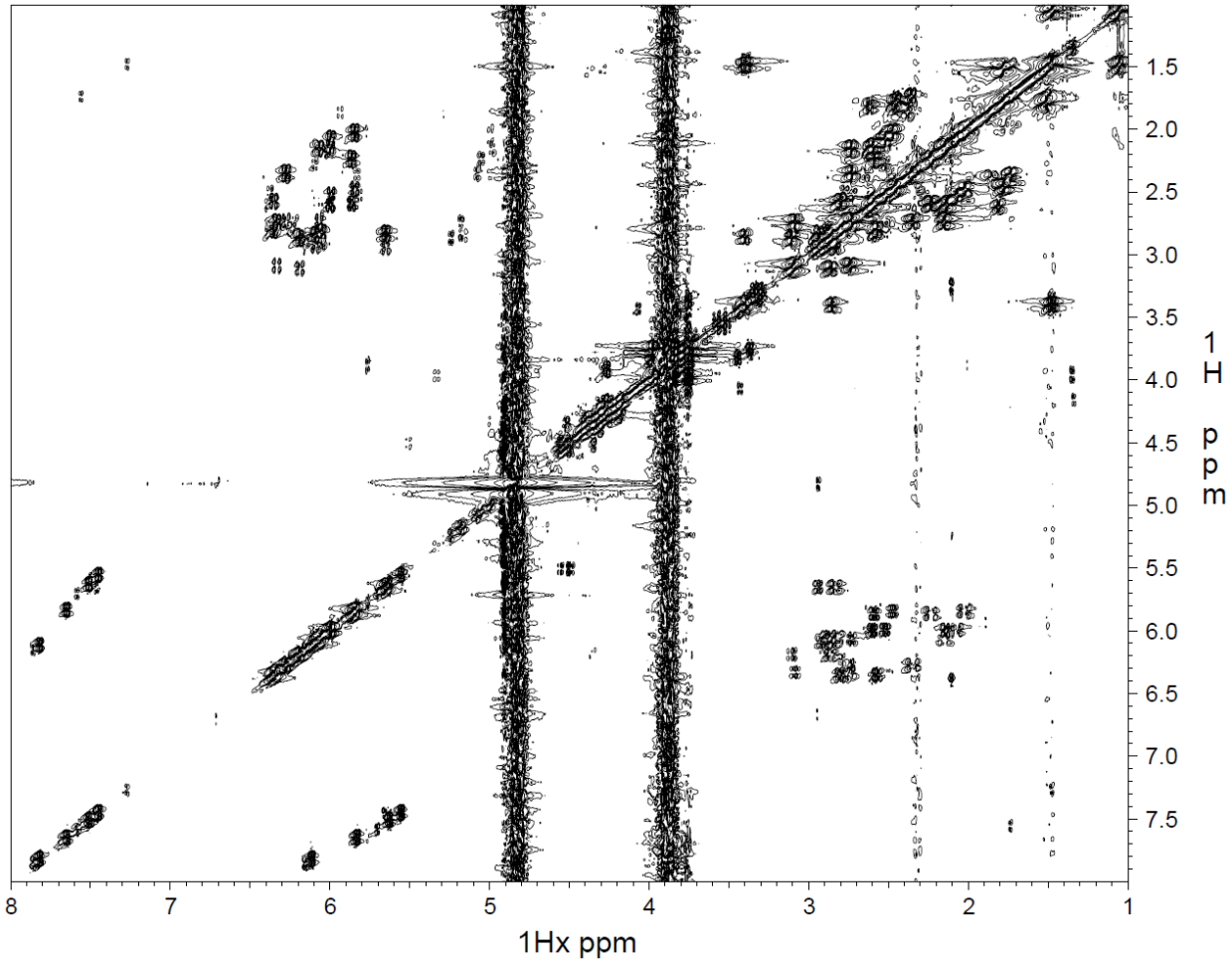

DMTO
O
O
NC
O
P
N


ertical scale = 0.2940
PM
7.6
7.2
6.8
6.4
6.0
5.6
5.2
4.8
4.4
4.0
3.6
3.2
2.8
2.4
2.0
1.6
1.2
0.8
0.4
0.0

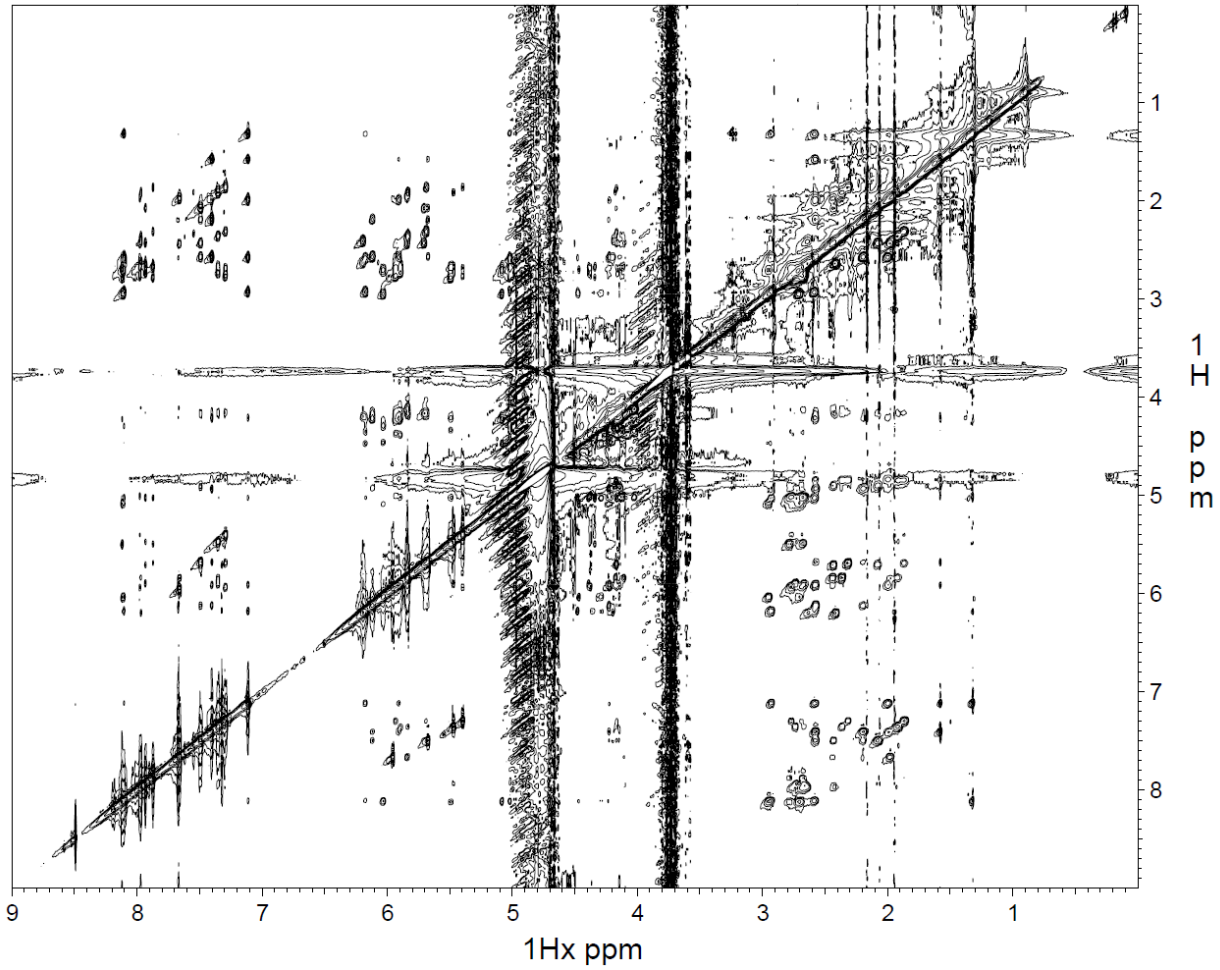

F
DMTO
O
O
NC
O
P
N


ertical scale = 0.2215
PM
7.2
6.8
6.4
6.0
5.6
5.2
4.8
4.4
4.0
3.6
3.2
2.8
2.4
2.0
1.6
1.2
0.8
0.4
0.0

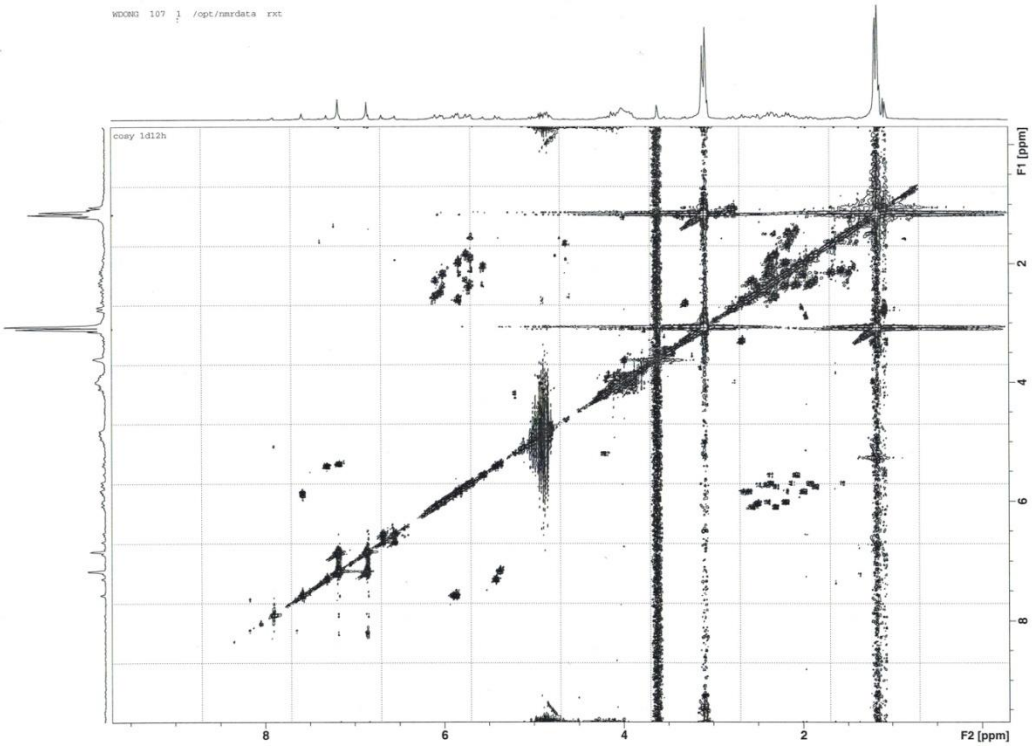
2D COSY of Dickerson dodecamer in 100% D_2O



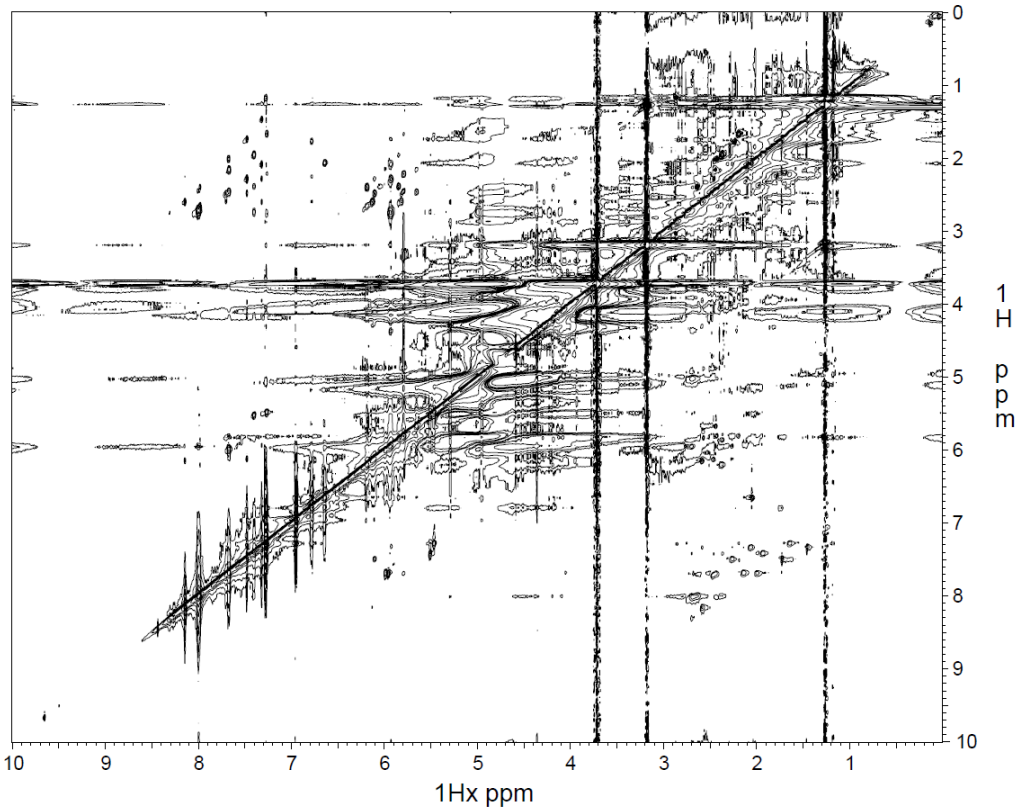
2D NOESY of Dickerson dodecamer in 100% D_2O



2D COSY of modified Dickerson dodecamer in 100% D_2O



2D NOESY of modified Dickerson dodecamer in 100% D_2O



NOE peak assignment of modified Dickerson dodecamer

Number	H1	H1’	Intensity
0	8.00987	5.94611	2.0023
1	8.01038	5.81253	0.7825
2	8.01034	5.48650	1.0637
3	8.00028	6.13142	0.8976
4	7.99257	5.52060	1.1052
5	7.99176	5.95626	1.9445
6	7.99174	5.66664	0.7761
7	7.98566	6.20540	1.2848
8	7.98475	5.85948	0.6661
9	8.14621	6.14443	2.1886
10	7.48052	6.11955	3.9602
11	7.67125	5.98209	8.6778
12	7.32220	5.94019	2.0365
13	7.39920	5.85702	1.3991
14	7.70515	5.82680	1.3737
15	7.67163	5.81006	1.9784
16	7.41002	5.52463	5.5907
17	7.39902	5.52059	5.5794
18	7.27372	5.48382	5.1648
19	7.32207	6.14208	0.7210
20	7.27602	5.92851	0.4149
21	7.66579	5.66754	0.7622
22	7.67744	5.66822	0.5349
23	7.40006	5.94974	1.0876
24	7.25902	5.81151	1.3388
25	7.25968	5.93166	0.4990
26	7.27582	5.81231	1.2159
27	7.26378	2.76768	1.9205
28	7.39908	2.74003	2.6065
29	8.00958	2.71059	8.6402
30	7.99270	2.72307	9.5304
31	8.14622	2.59822	6.7934
32	7.70452	2.58733	2.0333
33	7.32176	2.59761	2.9975
34	6.77978	2.59402	1.3482
35	7.70470	2.48730	3.8359
36	7.48048	2.48352	3.2507
37	7.67484	2.46091	4.1981
38	8.00972	2.44849	1.5843
39	7.98451	2.39360	3.0267
40	7.39899	2.38636	2.3237
41	7.48055	2.28293	7.5539

42	7.32164	2.28163	3.0072
43	8.14645	2.25277	1.0084
44	6.77953	2.24945	1.4837
45	7.70447	2.20874	1.0442
46	7.26379	2.21253	1.6663
47	7.67725	2.15901	5.2081
48	7.66687	2.15888	5.5814
49	7.39633	2.16091	0.6788
50	7.36889	2.12982	0.7013
51	7.32192	2.08299	6.0162
52	6.77852	2.06233	0.9256
53	6.63998	2.06184	5.9716
54	8.01059	2.01088	0.9093
55	7.68292	2.01205	4.2111
56	7.67202	2.01130	4.4108
57	7.98405	1.95466	0.7470
58	7.39935	1.95377	5.4583
59	6.77989	1.76720	2.7206
60	7.48061	1.73850	7.6045
61	7.27323	1.66620	3.3533
62	8.14503	1.46550	0.7337
63	7.32184	1.46713	6.7940
64	7.31863	1.71848	0.4721
65	8.14552	1.77043	0.5861
66	7.99175	2.15606	0.3705
67	5.94241	2.74736	7.4423
68	5.51846	2.73834	1.2352
69	6.19276	2.65605	3.2238
70	6.13583	2.60021	6.7462
71	5.81994	2.59352	3.9556
72	6.10954	2.48545	3.9801
73	5.79903	2.45249	3.1648
74	5.65930	2.46067	3.5600
75	6.19241	2.39451	6.4109
76	5.85164	2.38777	4.1132
77	5.93408	2.28692	3.9978
78	6.10961	2.28259	1.7041
79	5.79604	2.21594	4.1357
80	5.65901	2.16124	1.4511
81	5.93334	2.08644	2.1502
82	5.85092	1.95636	1.5007
83	5.79594	1.66930	1.5565
84	6.02033	2.37768	0.7854
85	6.01091	2.82850	0.5197
86	6.01095	2.61412	0.5185

87	6.01072	2.36344	0.7682
88	5.99896	2.48072	0.5922
89	5.98952	2.62443	0.6053
90	5.98965	2.34159	0.7330
91	5.98919	2.15219	0.9941
92	5.98753	2.48081	0.6169
93	5.97775	2.35831	0.5671
94	5.79690	2.01472	0.7258
95	5.93372	1.74295	0.6721
96	4.13186	2.21413	1.0108
97	4.15070	2.29026	0.8313
98	4.19488	2.39149	1.8389
99	4.06081	2.22750	0.7800
100	4.06049	2.47247	0.9010
101	4.01322	2.45584	0.8304
102	4.29149	2.59822	1.7577
103	4.11555	2.72636	1.5266
104	4.08207	2.72028	1.3968
105	4.02657	2.47030	0.6174
106	6.78	6.65	8.4280
107	7.68	5.81	1.9744

5'- C1 G2 C3 G4 X5 A6 T7 T8 C9 G10 C11 G12 - 3'
3'- G12 C11 G10 C9 T9 T8 A6 X5 G4 C3 G2 C1 - 5'
*** means overlap or ambiguous**

H1 assignment	Number	H1	H1'	Intensity
C1 C6H	11	7.67125	5.98209(C1 C5H)	8.6778
	107	7.68	5.81(C1 C1'H)	1.9744
	37	7.67484	2.46091(C1 C2"H)	4.1981
	56	7.67202	2.01130(C1 C2'H)	4.4108
	55*	7.68292	2.01205(G2 C2'H)	4.2111
G2 C8H	1	8.01038	5.81253(C1 C1'H)	0.7825
	54	8.01059	2.01088(G2 C2'H)	0.9093
	0	8.00987	5.94611(G2 C1'H)	2.0023
	29	8.00958	2.71059(G2 C2"H)	8.6402
	38	8.00972	2.44849(C1 C2"H)	1.5843
	2	8.01034	5.48650(C3 C5H)	1.0637
C3 C6H	18	7.27372	5.48382(C3 C5H)	5.1648
	26	7.27582	5.81231(C3 C1'H)	1.2159
	20	7.27602	5.92851(G2 C1'H)	0.4149
	27	7.26378	2.767689(G2 C2"H)	1.9205
	61	7.27323	1.66620(C3 C2'H)	3.3533
	46	7.26379	2.21253(C3 C2"H)	1.6663
G4 C8H	3	8.00028	6.13142(G4 C1'H)	0.8976
	39	7.98451	2.39360(G4 C2'H)	3.0267
	38*	8.00972	2.44849(G4 C2"H)	1.5843
X5 C6H	106	6.78	6.65(X5 C5H)	8.4280
	34	6.77978	2.59402(A6 C2"H)	1.3482

	44	6.77953	2.24945(X5 C2”H)	1.4837
	52	6.77852	2.06233(X5 CH3)	0.9256
	59	6.77989	1.76720(X5 C2’H)	2.7206
X5 C5H	53	6.63998	2.06184(X5 CH3)	5.9716
A6 C8H	9	8.14621	6.14443(A6 C1’H)	2.1886
	31	8.14622	2.59822(A6 C2”H)	6.7934
	43	8.14645	2.25277(A6 C2’H)	1.0084
	62	8.14503	1.46550(T7 CH3)	0.7337
	65	8.14552	1.77043(X5 C2’H)	0.5861
T7 C6H	12	7.32220	5.94019(T7 C1’H)	2.0365
	19	7.32207	6.14208(A6 C1’H)	0.7210
	33	7.32176	2.59761(A6 C2”H)	2.9975
	42	7.32164	2.28163(T7 C2”H)	3.0072
	64	7.31863	1.71848(T8 CH3)	0.4721
	51	7.32192	2.08299(T7 C2’H)	6.0162
	63	7.32184	1.46713(T7 CH3)	6.7940
T8 C6H	10	7.48052	6.11955(T8 C1’H)	3.9602
	36	7.48048	2.48352(T8 C2”H)	3.2507
	41*	7.48055	2.28293(T8 C2’H), (T7 C2”H)	7.5539
	60	7.48061	1.73850(T8 CH3)	7.6045
C9 C6H	15	7.67163	5.81006(C9 C5H)	1.9784
	21	7.66579	5.66754(C9 C1’H)	0.7622
	37*	7.67484	2.46091(C9 C2”H)	4.1981
	48	7.66687	2.15888(C9 C2’H)	5.5814
	47*	7.67725	2.15901（G10 C2’H）	5.2081

G10 C8H	5	7.99176	5.95626(G10 C1'H)	1.9445
	6	7.99174	5.66664(C9 C1'H)	0.7761
	30	7.99270	2.72307(G10 C2"H)	9.5304
	66	7.99175	2.15606(G10 C2'H)	0.3705
	4	7.99257	5.52060(C11 C5H)	1.1052
C11 C6H	16	7.41002	5.52463(C11 C5H)	5.5907
	23	7.40006	5.94974(G10 C1'H)	1.0876
	13	7.39920	5.85702(C11 C1'H)	1.3991
	28	7.39908	2.74003(G10 C2"H)	2.6065
	40	7.39899	2.38636(C11 C2"H)	2.3237
	58	7.39935	1.95377(C11 C2'H)	5.4583
	49*	7.39633	2.16091(G10 C2'H)	0.6788
G12	7	7.98566	6.20540(G12 C1'H)	1.2848
	8	7.98475	5.85948(C11 C1'H)	0.6661
	39	7.98451	2.39360(G12 C2'H),(C11 C2"H)	3.0267
	57*	7.98405	1.95466(C11 C2'H)	0.7470

Number	H1	H1’	intensity
25	7.25968 C3 C6H	5.93166 G10 C1’H(the complementary strand)	0.4990
67	5.94241 G2 C1’H	2.74736 G2 C2’’H	7.4423
68	5.51846 C11 C5H OR C3 C5H	2.73834 G2 C2’’H (the complementary strand)	1.2352
69	6.19276 G12 C1’H	2.65605 G12 C2’’H	3.2238
70	6.13583 A6 C1’H	2.60021 A6 C2’’H	6.7462
72	6.10954 T8 C1’H	2.48545 T8 C2’’H	3.9801
73	5.79903 C1 C1’H	2.45249 C1 C2’’H	3.1648
74	5.65930 C9 C1’H	2.46067 C9 C2’’H	3.5600
75	6.19241 G12 C1’H	2.39451 G12 C2’H	6.4109
76	5.85164 C11 C1’H	2.38777 C11 C2’’H	4.1132
77	5.93408 T7 C1’H	2.28692 T7 C2’’H	3.9978
78	6.10961 T8 C1’H	2.28259 T7 C2’’H	1.7041
79	5.79604 C3 C1’H	2.21594 C3 C2’’H	4.1357
80	5.65901 C9 C1’H	2.16124 C9 C2’H	1.4511

81	5.93334 T7 C1'H	2.08644 T7 C2'H	2.1502
82	5.85092 C11 C1'H	1.95636 C11 C2'H	1.5007
83	5.79594 C3 C1'H	1.66930 C3 C2'H	1.5565
95	5.93372 X5 C1'H	1.74295 X5 C2'H	0.6721
105	4.02657 G4 C4'H	2.47030 G4 C2''H	0.6174
104	4.08207 G12 C4'H	2.72028 G2 C2''H (the complementary strand)	1.3968
103	4.11555 G2 C4'H	2.72636 G2 C2''H	1.5266
102	4.29149 A6 C4'H	2.59822 A6 C2''H	1.7577
101	4.01322 C1 C4'H	2.45584 C1 C2''H	0.8304
99	4.06081 X5 C4'H	2.22750 X5 C2''H	0.7800
100	4.06049 X5 C4'H	2.47247 G4 C2''H or T8 C2''H(the complementary strand	0.9010
98	4.19488 C11 C4'H	2.39149 C11 C2''H	1.8389
97	4.15070 T7 CH'4	2.29026 T7 C2''H	0.8313
96	4.13186 C3 C4'H	2.21413 C3 C2''H	1.0108

94	5.79690 C3 C1'H	2.01472 G2 C2''H	0.7258
71	5.81994 C1 C1'H	2.59352 G12 C2''H(the complementary strand)	3.9556
14*	7.70515 C9 C6H	5.82680 C3 C1'H	1.3737
32*	7.70452 C9 C6H	2.58733 A6 C2''H	2.0333
35*	7.70470 C9 C6H	2.48730 G4 C2''H	3.8359
45*	7.70447 C9 C6H	2.20874 C3 C2''H	1.0442
50*	7.36889 A6 C2H	2.12982 T7 C2'H	0.7013
88	5.99896 G10 C1' H	2.48072 C9 C2''H	0.5922
89	5.98952 T7 C1' H	2.62443 A6 C2''H	0.6053
90	5.98965 G10 C1' H	2.34159 C11 C2''H	0.7330
91	5.98919 G10 C1' H	2.15219 G10 C2''H	0.9941
92	5.98753 T7 C1' H	2.48081 T8 C2''H	0.6169
93	5.97775 X5 C1''H	2.35831 G4 C2'H	0.5671
84	6.02033 C1 C5H	2.37768 C11 C2''H(the complementary strand)	0.7854
86	6.01095	2.61412	0.5185

	C1 C5H	G12 C2’’H(the complementary strand)	
87	6.01072 C1 C5H	2.36344 G12 C2’H(the complementary strand)	0.7682
85	6.01091 C1 C5H	2.82850 G2 C2’’H	0.5197
24*	7.25902 C3 C6H	5.81151 C9 C5H (the complementary strand)	1.3388
17*	7.39902 C11 C6H	5.52059 C3 C5H(the complementary strand)	5.5794
47*	7.67725 C9 C6H	2.15901 G10 C2’H	5.2081
55*	7.68292 C1 C6H	2.01205 G2 C2’H	4.2111

Comparison of the NOE's between purine C8H or pyrimidine C6H and the C1'H of its own sugar and (n-1) sugar

Non-natural duplex			**Natural duplex**		
H1	H1'	Intensity	H1	H1'	Intensity
C1 C6H	C1 C5H	S	C1 C6H	C1 C5H	S
	C1 C1'H	W		C1 C1'H	W
G2 C8H	C1 C1'H	VW	G2 C8H	C1 C1'H	MISSING
	G2 C1'H	W		G2 C1'H	W
	C3 C5H	VW		C3 C5H	VW
C3 C6H	C3 C5H	M	C3 C6H	C3 C5H	S
	C3 C1'H	W		C3 C1'H	W
	G2 C1'H	VW		G2 C1'H	W
G4 C8H	G4 C1'H	VW	G4 C8H	G4 C1'H	W
	C3 C1'H	MISSING		C3 C1'H	W
X5 C6H	X5 C5H	S	A5 C8H	A5 C1'H	W
	X5 CH3	W		G4 C1'H	W
	X5 C1'H	MISSING			
	G4 C1'H	MISSING			
X5 C5H	X5 CH3	M			

A6 C8H	A6 C1’H	W	A6 C8H	A6 C1’H	W
	X5 C1’H	MISSING		A5 C1’H	W
T7 C6H	T7 C1’H	W	T7 C6H	T7 C1’H	W
	A6 C1’H	VW		A6 C1’H	W
T8 C6H	T8 C1’H	M	T8 C6H	T8 C1’H	W
	C9 C5H	MISSING		C9 C5H	W
	T7 C1’H	MISSING		T7 C1’H	W
C9 C6H	C9 C5H	M	C9 C6H	C9 C5H	S
	C9 C1’H	W		C9 C1’H	W
	T8 C1’H	MISSING		T8 C1’H	W
G10 C8H	G10 C1’H	W	G10 C8H	G10 C1’H	W
	C9 C1’H	W		C9 C1’H	W
	C11 C5H	W		C11 C5H	W
C11 C6H	C11 C5H	S	C11 C6H	C11 C5H	S
	G10 C1’H	W		G10 C1’H	W
	C11 C1’H	W		C11 C1’H	W
G12	G12 C1’H	W	G12	G12 C1’H	W

	C11 C1’H	W		C11 C1’H	W
S:strong, M: medium, W:weak, VW: very weak					

Comparison of the NOE's between purine C8H or pyrimidine C6H and the C2'H or C2''H of its own sugar and n-1 sugar

Non-natural duplex			**Natural duplex**		
H1	H1'	Intensity	H1	H1'	Intensity
C1 C6H	C1 C2'H	M	C1 C6H	C1 C2'H	S
	C1 C2''H	M		C1 C2''H	M
G2 C8H	G2 C2'H	W	G2 C8H	G2 C2'H	S
	G2 C2''H	S		G2 C2''H	M
	C1 C2''H	W		C1 C2''H	W
C3 C6H	C3 C2'H	M	C3 C6H	C3 C2'H	S
	C3 C2''H	M		C3 C2''H	M
	G2 C2''H	M		G2 C2''H	M
G4 C8H	G4 C2'H	M	G4 C8H	G4 C2'H	S
	G4 C2''H	M		G4 C2''H	M
	C3 C2''H	MISSING		C3 C2''H	W
	C3 C2'H	MISSING		C3 C2'H	W
X5 C6H	X5 C2'H	M	A5 C8H	A5 C2'H	S
	X5 C2''H	M		A5 C2''H	S
	G4 C2''H	MISSING		G4 C2''H	S
	A6 C2''H	M		A6 C2''H	DON'T HAVE

A6 C8H	A6 C2’H	W	A6 C8H	A6 C2’H	S
	A6 C2’’H	S		A6 C2’’H	S
	T7 CH3	W		T7 CH3	S
	X5 C2’’H	MISSING		A5 C2’’H	MISSING
	X5 C2’H	VW		A5 C2’H	DON’T HAVE
T7 C6H	T7 C2’H	S	T7 C6H	T7 C2’H	S
	T7 C2’’H	M		T7 C2’’H	S
	T7 CH3	S		T7 CH3	S
	A6 C2’’H	M		A6 C2’’H	S
	T8 CH3	VW		T8 CH3	M
T8 C6H	T8 C2’’H	M	T8 C6H	T8 C2’H	S
	T8 C2’H, T7 C2’’H (OVERLAP)	S		T8 C2’’H, T7 C2’’H (OVERLAP)	S
	T8 CH3	S		T8 CH3	S
	T7 C2’H	MISSING		T7 C2’H	M
C9 C6H	C9 C2’H	S	C9 C6H	C9 C2’H	S
	C9 C2’’H	S		C9 C2’’H	M
	T8 C2’’H	MISSING		T8 C2’’H	M
	T8 C2’H	MISSING		T8 C2’H	W

G10 C8H	G10 C2’’H	S	G10 C8H	G10 C2’’H, G10 C2’H (OVERLAP)	S
	G10 C2’H	VW		C9 C2’H	W
	C9 C2’H	MISSING		C9 C2’’H	W
	C9 C2’’H	MISSING			
C11 C6H	C11 C2’H	S	C11 C6H	C11 C2’H	S
	C11 C2’’H	M		C11 C2’’H	M
	G10 C2’’H	M		G10 C2’’H	M
	G10 C2’H	W		G10 C2’H	MISSING
G12	G12 C2’’H	MISSING	G12	G12 C2’’H	S
	G12 C2’H, C11 C2’’H (OVERLAP)	M		G12 C2’H, C11 C2’’H (OVERLAP)	M
	C11 C2’H	W		C11 C2’H	W
C2’’H is the alpha proton and C2’H is the beta proton S:strong, M: medium, W:weak, VW: very weak					