BRANCHED NUCLEIC ACIDS
NOVEL PROBES FOR STUDYING PRE mRNA PROCESSING AND TRIPLE HELICAL DNA

Ravinderjit Singh Braich

Department of Chemistry
McGill University
Montreal, Canada.

A Thesis submitted to the Faculty of
Graduate Studies and Research
in partial fulfillment of the requirements of the degree of
Doctor of Philosophy

The domain in which chemical synthesis exercises its creative power is vaster than that of nature herself.

Marcellin Berthelot

ABSTRACT

A general procedure for the solid-phase synthesis of branched oligonucleotide analogs, including some with similar base sequence to the yeast (*Saccharmoyces cerevisiae*) rp51 mRNA, is described using the automated solid-phase phosphite triester approach. Using readily available phosphoramidite reagents, branch-points were introduced by sequential removal of the phosphate (β-cyanoethyl or methyl) and silyl protecting groups without detaching the nascent oligonucleotide from the solid-support. This then allowed the extension from the branch-point in the conventional 3',5', or the unconventional 5',3' directions, thus affording complete control over the base composition and direction of strand around the branch point.

The protocol was modified for the synthesis of msDNA, a DNA/RNA branched chimera. Using a combination of silyl and Fpmp groups for the protection of 2'-hydroxyl functions, a linear oligonucleotide was assembled. Branching was then introduced using the methodology developed for branched DNA.

Utilization of capillary electrophoresis (CE) in the analysis of branched nucleic acids was also touched upon. CE proved to be a very powerful technique in the separation and identification of a mixture of four branched oligonucleotides. This technique, CE, was also seen to be an attractive alternative for following the debranching reaction, thus eliminating the need for radioactively labeled substrates.

The unique structural properties of branched nucleic acids were exploited in investigation of the association properties of bNAs. By synthesizing branch molecules, capable of forming T:A/T triplexes in the parallel (Hoogsteen) or the antiparallel (reverse Hoogsteen) T:A/T motifs allowed, for the first time, a direct comparison of their relative stabilities. Most of the 0.25 kcal mol^{-1}/TAT base triplet difference in stabilities was seen to come from differences in enthalpy, with the parallel (Hoogsteen) base triplet being more stable.

Studies to explore structural effects in the use of a branched adenosine as replacement for nucleotide loops in duplex DNA are also described. Branched oligonucleotides of the type $rA[^{2',5'}dC_ndA_{10}\text{-}5']^{3'}\cdot 5dC_ndT_{10}\text{-}3'$ and $rA[^{2',5'}dC_n3',3'dA_{10}\text{-}5']^{3'}\cdot 5dC_ndT_{10}\text{-}3'$ are shown to form hairpin duplexes with thermal stabilities comparable to or better than that of one with a natural deoxynucleotide loop. Molecular modeling was used to study the conformation of branched hairpin loops. In certain cases branched hairpins adopted conformations similar to the ones seen in naturally occurring super stable RNA hairpins.

The branched nucleic acids (bNAs) also proved to be invaluable tools for studying *in vitro* splicing and the substrate specificity of the yeast debranching enzyme.

RÉSUMÉ

Un protocole général de synthèse d'analogues d'oligonucléotides ramifiés, en phase solide automatisée, est présenté. La synthèse de nucléotides contenant des séquences similaires à celles de la levure (*Saccharmoyces cerevisiae*) rp51 mRNA est aussi rapportée. Les points de ramification ont été introduits par le retrait séquentiel de groupes protecteurs de phosphate (β-cyanure d'éthyle ou de méthyle) ou de silyle, tout en évitant de détacher l'oligonucléotide du support solide. Il est ainsi possible de faire croître l'oligonucléotide à partir du point de ramification selon la direction conventionelle 3',5' ou dans la direction opposée 5',3', ceci permettant un parfait contrôle de la composition en bases et de la direction du brin autour du point de ramification.

Le protocole a été modifié pour la synthèse du msADN, un analogue ramifié d'ADN/ARN. Un oligonucléotide a été assemblé en utilisant la combinaison des groupes silyles et Fpmp pour la protection des fonctions 2'-hydroxyles. La ramification a ensuite été réalisée en utilisant la méthodologie développée pour l'ADN ramifié.

L'électrophorèse capillaire (EC) a aussi été utilisée pour l'analyse d'acides nucléiques ramifiés. L'EC s'est avérée être une technique très efficace pour la séparation et l'identification de mélanges de quatre oligonucléotides ramifiés. L'EC peut aussi être une alternative intéressante au marquage radioactif de substrats pour suivre la réaction de déramification.

Les propriétés structurales inédites des acides nucléiques ont été exploitées pour la caractérisation d'ADN hélicoidal triple. La synthèse de molécules ramifiées capables de former des triplex T:A/T dans le sens parallèle (Hoogsteen) ou anti-parallèle (Hoogsteen inversé), selon le motifs T:A/T, permet de comparer leur stabilité relative. La différence de stabilité est de 0,25 kcal mol^{-1}/TAT. La majeure partie de cette différence s'explique par la différence d'enthalpie, le triplet parallèle (Hoogsteen) étant plus stable.

L'utilisation d'adénosine ramifiée comme remplacement de boucles de nucléotides dans le duplexe d'ADN est aussi décrite. Des oligonucléotides ramifiés de type $rA[^{2',5'}dC_ndA_{10}\text{-}5']^{3'}\text{-}5dC_ndT_{10}\text{-}3'$ et $rA[^{2',5'}dCn3',3'dA_{10}\text{-}5']3'\text{-}5dC_ndT_{10}\text{-}3'$ forment des duplexes en tête d'épingle et ont une stabilité thermique comparable à celle de la boucle de déoxynucléotide existant à l'état naturel. La modélisation moléculaire a aussi été utilisée pour étudier la conformation des boucles en tête d'épingle ramifiées. Dans certains cas, les conformations de ces systèmes sont similaires à celles adoptées par les duplexes d'ADN super stables existant naturellement.

Les acides nucléiques ramifiés (bNAs) s'avèrent être de très bons outils pour étudier le collage *in vitro* et la spécificité de l'enzyme de déramification de la levure vis-à-vis du substrat.

ACKNOWLEDGEMENTS

"You can't throw a stone in the water without making waves. Each new personality you meet effects your life, because you always hear, see or get something from them"

Ayub Ogada

This is perhaps the most difficult task of writing a thesis. So with the words of Ayub Ogada in mind, I would like to acknowledge the people who made this thesis possible and mostly an enjoyable experience.

First and foremost I would like to acknowledge the supervision of Dr. Masad J. Damha. For giving me the encouragement and the opportunity to succeed, I owe my deepest gratitude.

I also want to express my deepest gratitude to Dr. David N. Harpp, for his unconditional generosity, taking in an extra charge, and making the duration of my stay in his lab an enjoyable experience.

I would also like to acknowledge and thank my coworkers past and present for their assistance during the course of my studies.

And a thanks to those who have collaborated during the course of this work. In no particular order, thanks go out to: Dr. Tanya Tadey (McGill University) for teaching me capillary electrophoresis; Dr. A. Dorian and Prof. J. Friesen (Hospital for Sick Kids, Toronto) for pre-mRNA splicing studies; Dr. K Nam, Prof. J. Boeke, and J. Trambley (Johns Hopkins University, School of Medicine,

Baltimore) for the yDBR debranching assays, to Prof. J. Turnbull (Concordia Universtiy) for assistance with CD experiments; Dr. Gadi Borkow (McGill AIDS Centre) for teaching me radioactive labeling of nucleic acids, to Antisar for MALDI-TOF mass spectroscopy.

I would like to thank all the people in lab 240, in no particular order, Andrej, Charlie, Erwin, Pierre and Sylvie, for making my stay a memorable experience.

To Ms. Colette Wabnitz for help in translation of literature and keeping the tea pot boiling! Thanks coco :)

A special thank you to Erwin (Techno B.) Schultz for his jokes and trips to the "House of Higher Learning", but most of all for introducing me to the joy of Guatemalan Java! I owe you one.

Thanks to Ms. Priscilla Abrego for the "Mayan magic".

Last but not the least, A., D., F., K., K., L., M., many thanks!

And to those I might have unintentionally left out, thank you.

Finally, I would like to express my appreciation to the Chemistry Department of McGill University for providing me the opportunity to study here.

TABLE OF CONTENTS

ABBREVIATIONS AND SYMBOLS

A	adenosine
A_{260}	UV absorbance measured at 260 nm
Ac_2O	acetic anhydride
Ad	adenine
AIDS	acquired immunodeficiency syndrome
AP	alkaline phosphatase
APS	ammonium persulfate
Atm	atmospheres
b	broad
BBP	branch binding protein
bDNA	branched DNA
BIS	N,N'-methylene-bis(acrylamide)
bNA	branched nucleic acid
BPB	bromophenol blue
bRNA	branched RNA
BSA	benzenesulfonic acid
Bz	benzoyl
C	cytidine
CE	capillary electrophoresis
CEO-	2-cyanoethoxy (β-cyanoethyl)
CSPDE	calf spleen phosphodiesterase
Cy	cytosine
d	doublet
DCC	1,3-dicyclohexylcarbodiimide
DCE	1,2-dichloroethane
DCM	dichloromethane
dd	doublet of doublets
DEC	1-(3-dimethylaminopropyl)-3-ethylcarbodiimide hydrochloride
DIPEA	*N,N*-diisoproylethylamine
4-DMAP	4-dimethylaminopyridine
DMF	*N,N*-dimethylformamide
DMT	dimethoxytrityl
dN	2'-deoxynucleotides
DNA	2'-deoxyribonucleic acid
DTT	dithiothreitol
E. coli	*Escherichia coli*
EDTA	disodium ethylenediaminetetraacetate dihydrate
esp.	especially
EtOH	ethanol
eq	equivalents
FAB-MAS	fast atom bombardment mass spectrometry
Fpmp	1-(2-fluorophenyl)-4-methoxypiperidine

G	guanosine
Gu	guanine
hDBE	human (HeLa) lariat debranching enzyme
HOAc	glacial acetic acid
HPLC	high performance liquid chromatography
Hz	hertz
i-, *n-*, *t-*	iso, normal, tertiary
J	coupling constant
λ	wavelength
LCAA-CPG	long-chain alkylamine controlled pore glass
Lv_2O	levulinic anhydride
LvOH	levulinic acid
M	molar
m	multiplet
MALDI-TOF	matrix assisted laser desorption ionization-time of flight
max	maximum
MeOH	methanol
min	minute
ml	milliliter
mM	millimolar
μM	micromolar
MMT	monomethoxytrityl
mol	mole
mRNA	messenger RNA
MS	mass spectrum
MS-Cl	2-Mesitylenesulfonyl chloride
MSNT	1-(2-mesitylenesulfonyl)-3-nitro-1,2,4-triazole
NCS	N-chlorosuccinimide
nm	nanometer
NMR	nuclear magnetic resonance
NP1	nuclease P1
NPE	*p*-nitrophenylethyl
OD	optical density
OPC	oligonucleotide purification cartridge
PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
Ph	phenyl
ppm	parts per million
pre	precursor
Pu	purine
Py	pyrimidine
py	pyridine
®	registered trademark
R_f	(TLC mobility) retardation factor
Rf	radio-frequency
rN	ribonucleotide

RNA	ribonucleic acid
Rnase	ribonuclease
RNasin	ribonuclease inhibitor
rRNA	ribosomal RNA
RPC	reversed-phase chromatography
rt	room temperature
s	singlet
S. cervisiae	*Saccharomyces cerevisiae*
SEC	size exclusion chromatography
sec	second
snRNA	small nucleolar RNA
snRNP	small nuclear ribonucleoprotein particles
SVPDE	snake venom phosphodiesterase
t	triplet
T	thymidine
TBAF	tetra-*n*-butylammonium fluoride
TBDMSi	*t*-butyldimethylsilyl
TBE	TRIS/boric acid/EDTA buffer
TCA	trichloroacetic acid
TEAA	triethylamine ammonium acetate
TEMED	N,N,N',N'-tetramethylethylenediamine
Th	thymine
THF	tetrahydrofuran
2,4,6-TIPSCl	2,4,6- triisopropylbenzenesulphonyl chloride
TLC	thin layer chromatography
T_m	thermal melt transition temperature
TREAT.HF	triethylamine trihydrofluoride
TRIS	2-amino-2-(hydroxymethyl)-1,3-propanediol
tRNA	transfer RNA
U	uridine
Ur	uracil
UV-vis	ultraviolet-visible
v/v	volume by volume
w/v	weight by volume
yDBE	yeast lariat debranching enzyme
XC	xylene cyanol
z/m	charge to mass ratio

LIST OF FIGURES

CHAPTER 1:

CHAPTER 2:

CHAPTER 3:

CHAPTER 4:

LIST OF SCHEMES

CHAPTER 1:

CHAPTER 2:

CHAPTER 3:

CHAPTER 4:

LIST OF TABLES

CHAPTER 2:

CHAPTER 3:

I. INTRODUCTION

1.1 IMPORTANCE OF NUCLEIC ACIDS

The chemistry of life on our planet is ultimately based upon the chemical and physical behavior of nucleic acids. Nucleic acids are also thought to have played an important role in the origin and evolution of life. Remarkably, although based on relatively simple monomers, nucleic acids participate in an impressive array of complex biological functions including the storage, replication and transmission of genetic (hereditary) information. DNA was the first chemical found to be capable of carrying genetic information.[1] This was so revolutionary a finding that the authors were very cautious and they made no mention of the word "gene" in their publication. This cautious approach was well founded since up to that time proteins were thought to be carriers of genes and nucleic acids were thought of being just a lowly inert skeleton supporting the proteins, the "real" molecules of life. As Moskovits wrote:

> *"No one was willing to believe that so boring a molecule as DNA, made up as it is from sugar, phosphate, and four simple organic bases, was the selfish emissary chosen to carry so important a message."*[2]

Later RNA, was also found to be capable of carrying and transmission of genetic information.[3,4]

The primary structure of both DNA and RNA is shown in figure **1.1**. Nucleic acids are polymeric molecules composed of nucleotides. A mononucleotide consists of a specific heterocyclic base, a five-membered carbohydrate moiety and a phosphate group. It is the sequence in which different nucleotides are arranged

along the polymer chain that gives each nucleic acid its own distinctive chemical and physical properties. The major heterocyclic bases are five in total and could be divided into two classes. The two bicyclic purines; adenine and guanine, and the three monocyclic pyrimidines; cytosine, uracil and thymine (DNA contains thymine while RNA contains uracil). The carbohydrate moiety in nucleic acids is the ß-D-ribofuranose (for RNA) or its 2-deoxy derivative (for DNA). To differentiate between the carbohydrate and heterocyclic base, the positions of atoms in the sugar rings are primed.

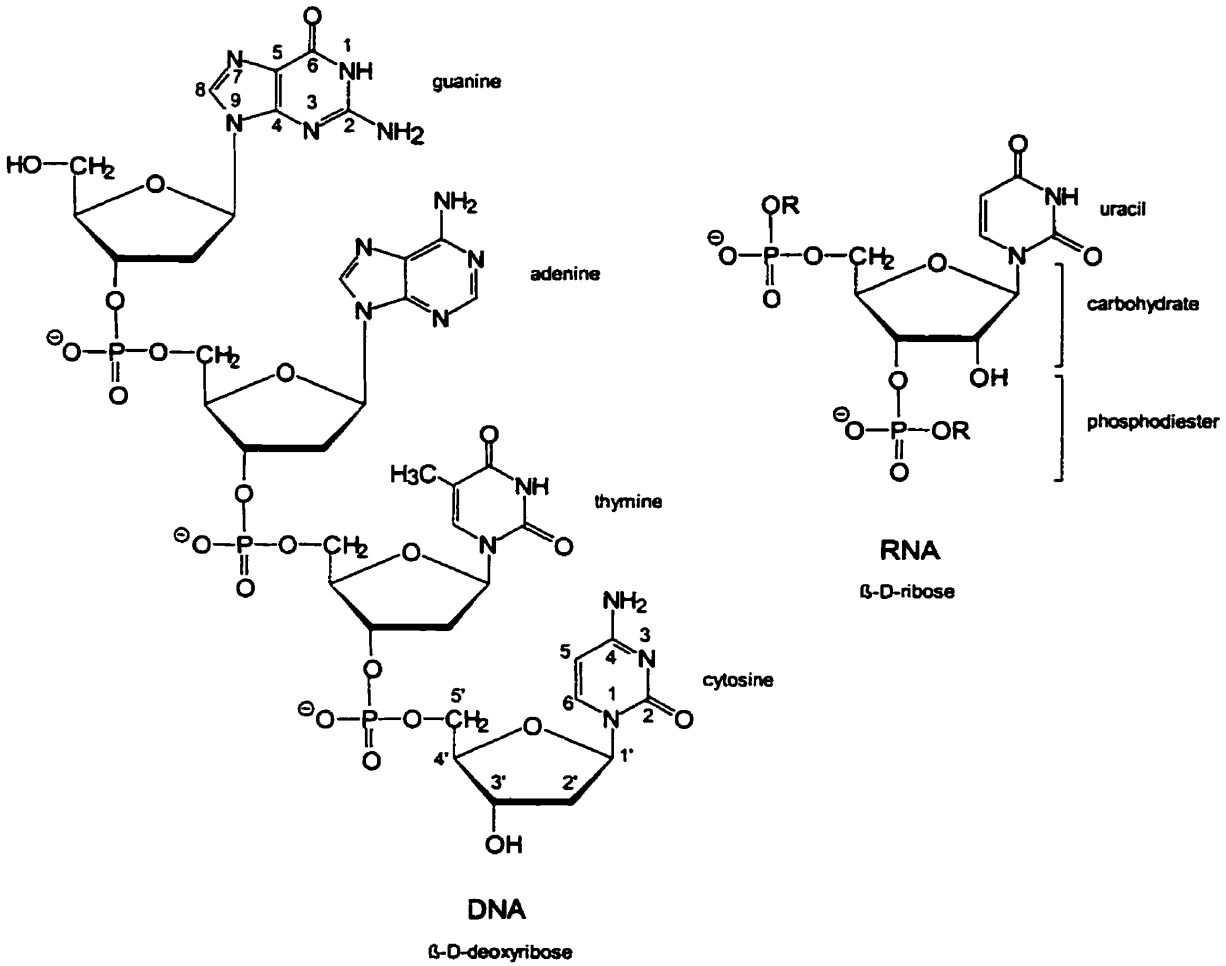


Figure 1.1: Primary structures of 2'-deoxyribonucleic acid (DNA, left) and ribonucleic acid (RNA, right) including numbering for carbohydrate, purine, and pyrimidine moieties.

In both DNA and RNA, the carbohydrate moieties are attached to the heterocyclic bases via a C1'-N9 ß-glycosidic bond for purines and a C1'-N1 ß-glycosidic bond for pyrimidines. The combined heterocyclic base and the carbohydrate moiety is termed a nucleoside. The nucleoside monomers are joined by a phosphodiester linkage between the C3' hydroxyl of one nucleoside and the C5' hydroxyl of the next nucleoside forming the polymeric nature of nucleic acids. All nucleotides in the polymer are linked by similar 3'-5'-phosphodiester linkages thus providing a directionality to the nucleic acids. Although the above represent the basic composition of DNA and RNA, there exists a variety of modified nucleotides in nature.[5]

Although DNA and RNA are very similar in their primary structure, there are some major differences between them. Absence of the 2'-OH makes DNA much more stable than RNA.[6,7] This relative stability of DNA molecules also makes them much better suited for the storage of genetic information. Also DNA is typically double stranded whereas RNA is primarily single stranded. In 1953, Watson and Crick[8] deciphered the three-dimensional structure of DNA and showed that DNA was in fact composed of two linear polynucleotides. The two linear polynucleotides were held together by the ability of heterocyclic bases to recognize their complement through base pairing by hydrogen bonding (H-bonds), adenine pairs with thymine and guanine with cytosine. These polynucleotide chains usually form right handed double helices with the chains oriented antiparallel to each other.

Although DNA contains the genetic information, it in itself is inert, without means of expression. The flow of genetic information begins with the DNA being transcribed to RNA, which is ultimately translated into proteins. For a long time

fidelity of the transmission of genetic information from DNA was thought to be due to formation of hydrogen bonds (H-bonds) in the Watson-Crick base pairings. The H-bonds between Watson Crick base pairs were viewed as the actual conduct of information flow, whereas base-stacking interactions served more of a structural role. Recently Kool's group has cast doubt on this simple theory.[9] Using nucleoside derivatives lacking the ability for hydrogen bonding based on Watson-Crick pairing, they observed fidelity of genetic transmission similar to the wild type cases. Also interesting was the finding that derivatives unable to hydrogen bond worked equally well if it they were part of the template being transcribed or incorporated into the nascent strand. By providing nucleoside analog derivatives through chemical synthesis, they have demonstrated that chemistry can play an important role in elucidating the mechanism involved in this and other similar biochemical processes so important for life on Earth.

Compared to DNA, RNA is unique in that it is capable of at least four distinct biological roles. The main role of RNA appears to be in the transfer of genetic information encoded in DNA. This function is readily carried out by messenger RNAs transcribed from the DNA. In certain viral genomes the storage and transmission of genetic information is in the form of RNAs, which are first "reverse" transcribed to DNA. Secondly, RNAs such as transfer and ribosomal RNA act as structural scaffolds during translation of information from messenger RNAs to protein. Thirdly, RNA has been found to possess catalytic properties. This was initially demonstrated by Cech and coworkers in the self-splicing activity of an intron in the nuclear pre-rRNA of the protozoan Tetrahymena thermophila without the assistance of proteins.[10-12] Additionally, RNA has been shown to function in true catalytic fashion in the RNA processing activity of RNase P[13,14] and the activity of various ribozymes and related structures. That

RNA can carry out reactions other than cleavage of nucleic acids was shown by the discovery of a RNA-mediated insertion of uridine residues in certain mitochondrial genes in kinetoplastid protozoa.[15] Noller[16] and others[17] have suggested that in some lower organisms the catalytic component of the ribosome carrying out the peptide bond formation is actually a RNA molecule. However no further studies have been reported. Fourthly, RNA is a regulatory element in HIV-1 (*e.g.* the "tar" RNA). RNA has also been shown to function as transcription factor in eukaryotes.[18]

Based on the catalytic activities of RNA molecules mentioned above, and using in-vitro molecular evolution techniques, the repertoire of reactions performed by nucleic acids has been expanded. Catalytic nucleic acids are now able to recognize and bind substrates other than nucleic acids and to carry out catalysis of specific reactions.[19-23]

From the ability of RNA to carry out such a diverse array of functions, several investigators have speculated that life was first based on RNA molecules.[24-28] Further evidence supporting this hypothesis is the observation that most biological coenzymes are ribonucleotides or derivatives of ribonucleotides.[29] Transfer of genetic information storage function from RNA to the more stable DNA may have been facilitated by reverse transcriptase, enzymes which catalyze the reverse transcription of RNA into DNA.[30,31] Observation of bacterial reverse transcriptases which use their own RNA for priming the reverse transcription process as in msDNA[32], further lends support to the "RNA world" hypothesis.

1.2 RNA SPLICING

The mechanism by which information stored in DNA is expressed has largely been elucidated. For a long time it was assumed that flow of information from the

genome to the final functional protein product was from a continuous array of coding sequences in the genomic DNA. However, in 1977, Sharp[25] and Roberts[33] independently discovered that protein-coding DNA sequences ("exons") contained interspersed segments which were not present in the final mature RNA. Now it has been well established that almost all eukaryotic RNAs are generated from precursor transcripts (pre-mRNA) containing stretches of nucleotides which must be removed to produce functional mature RNAs. These usually noncoding sequences, termed the "introns" (or intervening sequences), are removed from pre-mRNA during post-transcriptional processing (RNA splicing). Studies on genes coding for tRNA and rRNA have also shown the presence of "introns" in the genes coding for these RNAs.[34]

The number of introns per gene sequence varies from gene to gene and species to species. Initially thought of as "junk DNA", introns are now known to be an important part in the flow of genetic information. Studies by Leverette[35], Tycowski[36], and Bachellerie[37] have shown that in some cases pre-mRNA introns contain the coding sequences for some small nucleolar RNAs (snoRNAs), that may be important in rRNA assembly and maturation. The creation of a human intron-encoded RNA library,[38] database searches for sequences with complementarity to rRNA,[39] and isolation from HeLa cells[40] have led to identification of many encoding introns that play an array of roles in rRNA maturation, *e.g.* antisense snoRNAs.

Based on sequence complementarity between intron-encoded snoRNAs and sites of 2'-O-methylation in rRNAs, Bachellerie and coworkers[37,41] demonstrated that some intron-encoded snoRNAs were directing 2'-O-methylation of rRNAs. Gene deletion[38] and altered sequence complementarity[42] confirmed the role of snoRNA

in methylation of rRNA. More recent work[43,44] has shown snoRNAs to be also involved in formation of pseudouridines. Also recently, Ho and coworkers[45] observed an intron coded protein from *Aspergillus nidulans* assisting in group I intron RNA splicing. Of interest was the observation that the intron encoded protein also showed DNA endonuclease activity[45].

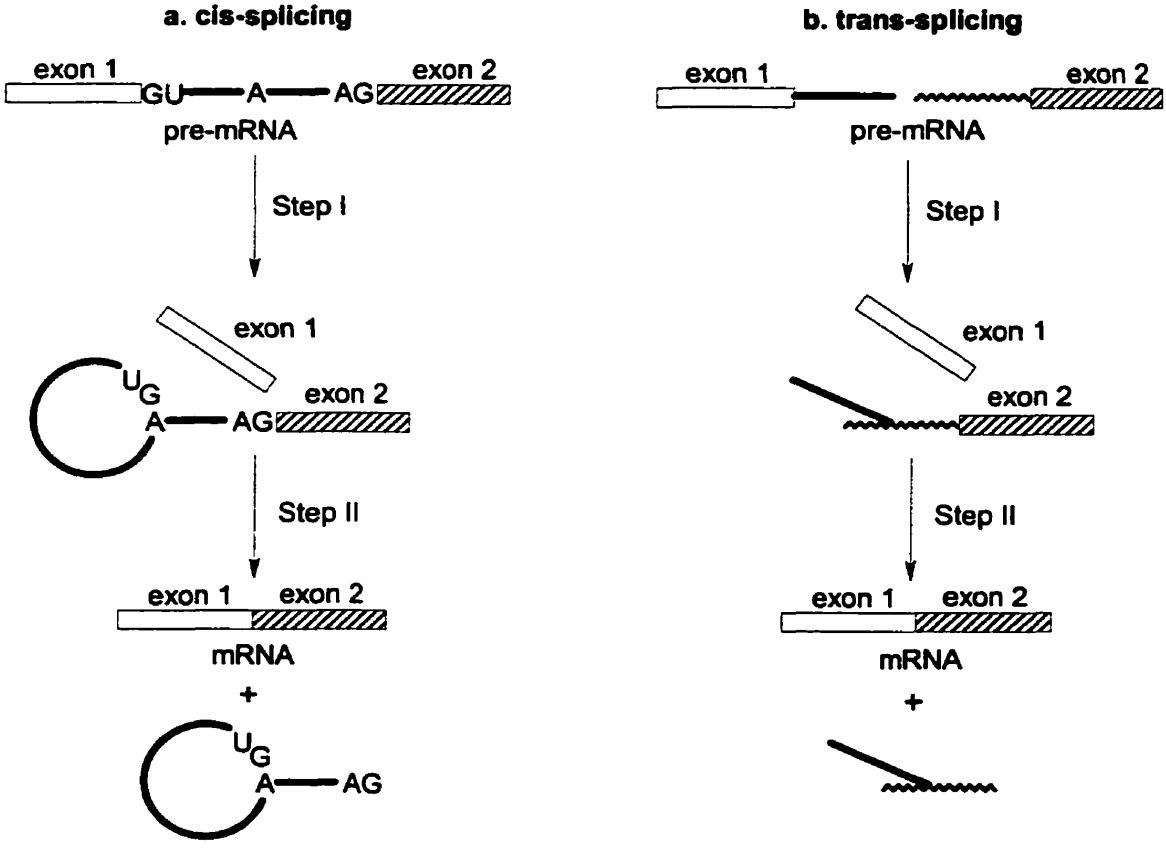
RNA splicing is a complex process which varies based on the type of RNA involved, but the key feature is the excision of introns and ligation of exons to form a mature RNA. A number of genetic diseases have been known to come from defects in RNA splicing, by which abnormal mature RNAs are produced. Recently, it has been reported that aberrant splicing of genes may be implicated in breast[46] and colon[47] cancers. Aberrant splicing has also been indicated in some cases of familial retinoblastoma.[48]

Although no mechanism for retaining the unspliced RNA transcripts in the nucleus is known, splicing of pre-RNAs invariably occurs in the nucleus.[49] RNA transcripts which can not be properly spliced are retained in the nucleus.[50] It is interesting to note that the RNA transcripts which are part of an alternative splicing pattern are readily transported to the cytoplasm even though they contain alternatively spliced introns. It is not clear if splicing of the pre-RNA transcript is a requirement for transport to the cytoplasm. For example, some studies have shown a strict requirement of splicing for transport to cytoplasm,[51,52] while others have shown the absence of such stringencies.[53,54] Also mRNAs from genes that do not contain introns are transported to the cytoplasm without splicing.[55]

From *in vitro* studies on the splicing of adenovirus 2 late major transcripts, Padgett *et al.*[56] postulated a mechanism for the pre-mRNA splicing pathway

involving two transesterification steps, scheme **1.1**. The excised intron is released in a novel "lariat" configuration, *i.e.* a circular RNA with a "tail". Presence of novel vicinal 2',5'- and 3',5'-phosphodiester linkages define the "branchpoint," a key feature of the lariat structure, figure **1.2**. The results of these studies showed that there was a conservation of phosphate groups and phosphodiester energy and thus required no exogenous energy. The two transesterification steps mechanism was deduced from the fact that the released lariat structure has a free terminal 3'-hydroxyl, confirming the formation of a phosphate moiety with a newly formed 3'-5' phosphodiester linkage originating from the 3'-splice site. Further evidence for this mechanism came from stereochemical studies of the cleavage/ligation reaction in pre-mRNA splicing.[57,374]

Scheme 1.1



Using phosphorothioate linkages at the 5'-splice site, Maschhoff and Padgett[58] showed that the first cleavage/ligation step takes place with inversion of configuration at the phosphate. This inversion of configuration was consistent with previous findings obtained for Group I self-splicing and gave strong support to the transesterification model.[59,60] Stereochemical results of a mechanism with separate and sequential cleavage and ligation, similar to pre-tRNA splicing,[57] would likely involve two inversions leading to a net retention of configuration at the phosphate.

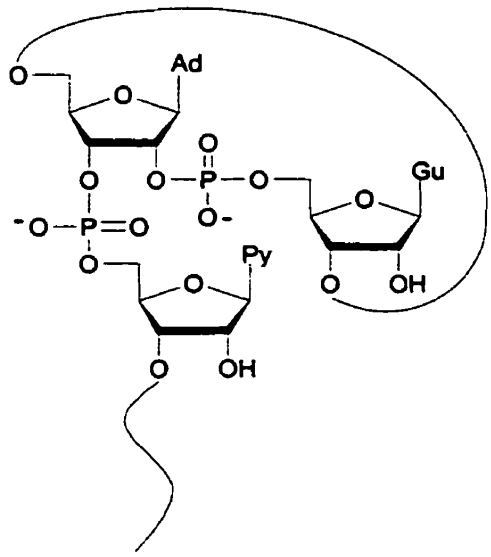


Figure 1.2: Lariat structure formed during cis-splicing of pre-mRNA.

In some cases, such as chloroplasts,[61-63] nematodes,[64,65] and trypanosomes,[66] mature RNA is formed by splicing together exons from two separate pre-mRNA transcripts. This trans-splicing[66] of two separate precursor transcripts, releases the intron in a "forked" structure comparable to the lariat structure formed during cis-splicing. Trans-splicing serves an important role during instances where a number of mRNA share the same 5'-end sequence (or exon). This allows the 5'-

exon to be transcribed from one place instead of each mRNA having its own copy.

Although fidelity of nuclear pre-mRNA splicing is remarkable, low frequency events of exon skipping and scrambling have shown that errors in fidelity of pre-mRNA splicing do occur.[67-69] These errors in splicing are known to result in unusual splice products, such as circular exons and scrambled exons.[67-69] Circular exons (circons), have also been produced artificially in several model systems both *in vivo*[70] and *in vitro*[71] from autocatalytic intron sequences and circular pre-mRNA transcripts. Initially thought to be due to aberrant splicing, circons are now known to play an important biological role in the replication of covalently circular subviral RNAs of some plant viruses.[70,72,73] Continuous replication of these pathogenic RNAs proceed via a "rolling circle" mechanism from the circular template.[74] The newly produced linear strand then undergoes autocatalytic cleavage to form the circular RNA.

The highlight of RNA splicing has been the discovery by Cech *et al.*[10] of autocatalytic splicing ability of a ribosomal RNA from *Tetrahymena thermophila* (a ciliated protozoan). Similar autocatalytic splicing has also been discovered in other rRNAs and mRNAs.[75-79] Based on the requirements for an external nucleotide cofactor, autocatalytic splicing has been classified into two groups; Group I, which requires an external guanosine co-factor, and Group II which require no external co-factor. An evolutionary link between Group I and nuclear pre-mRNA has been suggested based on the similarity of Group II intron with Group I as well as nuclear pre-mRNA introns.[80] The Group II introns are known to undergo autocatalytic splicing[78,79] whereas splicing of nuclear pre-mRNA requires the assembly of various *trans* acting protein factors into a high molecular

weight splicing complex known as the spliceosome, see also section 1.2.1. Production of "lariat" structure and similarities in mechanistic characteristics of pre-mRNA splicing and self-splicing Group II introns has promoted speculation that pre-mRNA splicing may have originated from, or may still rely on, catalytic properties of RNA. Various proteins are thought of serving more of a structural purpose in bringing the pre-mRNA into the proper conformation for the splicing reaction, which might fundamentally be autocatalytic.

1.2.1 Spliceosome

In eukaryotes (multicellular organisms that have well defined nucleus in their cells), splicing of nuclear pre-mRNA into mature mRNA requires the assembly of various factors into the spliceosome complex. Formation of this complex seems to be a prerequisite in specifying the pre-mRNA splicing process.[81-83] The final spliceosome complex consists of a multitude of RNA factors and proteins. The major components of the spliceosome are the small nuclear ribonucleoprotein particles, snRNPs (or "snurps"), comprising of various proteins and five small nucleolar RNAs (snRNAs), labeled U1, U2, U4, U5, and U6 due to their high uracil base content. These snRNAs have chain lengths of up to about 200 nucleotides and, except for U6, have an unusual 2,2,7-trimethyl guanosine cap structure at their 5' ends.[84] The snRNP proteins that are tightly complexed with snRNAs are usually referred to as the core snRNP proteins and possess determinants recognized by anti-Sm antibodies. These core proteins appear to be critical for synthesis of stable snRNAs in yeast and mammals. Other non-snRNP proteins, which are not stably associated with snRNAs, referred to as extrinsic factors, appear to associate transiently with the spliceosomal components during splicing.

Interaction of U1 snRNA with 5'-splice sites (and in some species with 3' splice sites)[85-87] and of U2 snRNA with the branchpoint[88-90] through Watson-Crick base pairing is well established. In fact, recognition of introns by U1 snRNA has been shown to be the crucial step in committing the pre-mRNA to the splicing pathway.[85] Mediated by snRNP-associated proteins, binding of U5 to 5'- and 3'-splice sites has been suggested.[91,92] Recent work of Ast and Weiner[93] suggests that U1 may be helping in directing U5 to the splice site, based on a novel site-specific crosslink between the U1 and U5 snRNAs. No direct Watson-Crick base pairing interactions between U4 or U6 and pre-mRNA has been found, although there is some evidence of a catalytic role for an U4-U6 RNA complex during splicing based on similarity of a conserved domain to the catalytic hammerhead RNA domain.[94] The U6 snRNA has also been implicated in selecting and proofreading the 5'-splice site during splicing.[95]

Recently, Abovich and Rosbash[96] identified a protein factor from the yeast *S. cerevisiae* in the early spliceosome complex designated the branchpoint bridging protein (BBP). Berglund *et al.*[97] have now shown that BBP interacts specifically with the pre-mRNA branchpoint sequence UACUAAC and may contribute to branchpoint recognition and possibly to branchpoint selection during early spliceosome complex formation.

The assembly of the spliceosome complex involves a highly ordered, cooperative, stepwise association of snRNPs with the substrate pre-RNA, requiring the input of energy at almost every step.[98,99] The role of ATP in splicing is seen more to be in assembling the complex and release of products since, as described above, the chemical events of splicing itself require no exogenous energy. Recent demonstration of putative ATP-binding RNA helicase

domains in several essential yeast splicing factors[100,101] and identification of an RNA-dependent ATPase activity in mammalian U5 snRNPs[102] have provided a link between requirement of ATP hydrolysis and conformational changes of the snRNPs during the spliceosome complex formation.

In the spliceosome, the interactions between the snRNPs and precursor RNA not only involve RNA-RNA interactions but probably RNA-protein, and protein-protein interactions as well. Admittedly this description is oversimplified and new studies will undoubtedly shed some light on the mechanism for selecting and matching the various splice sites during pre-mRNA splicing.

1.2.2 Importance of Branched RNA during splicing

First described in 1983 by Wallace and Edmonds,[103] branched RNA has become the quintessential feature of pre-mRNA splicing. Formation of this novel structure, figure **1.2**, has led to speculation as to its possible role during RNA splicing. A number of explanations have been offered as to the role of branch formation during mRNA splicing,[104] but no conclusive answers have been found. The removal of the reactive 5'-end of the intron by branch formation has been suggested to drive the reaction towards exon ligation by preventing the reactive 5'-end of the intron from recombining with the 5'-exon.[11] Based on complementary base pairing between U2 snRNA and the branch region consensus sequence, branchpoint adenosine is suggested to be unpaired or bulged within a helical region.[105] By establishing a base pairing interaction between the branchpoint adenosine and U2 snRNA, a strong inhibition of splicing is seen.[106] These observations suggest that the bulging out at the branchpoint leads to the correct geometry necessary for nucleophilic attack of the adenosine 2'-hydroxyl on the 5'-splice site.[107,108]

In yeast cells, studies with single point mutations at the normal branch-accepting adenosine were seen to completely abolish splicing.[109] Similar mutations at adenosine in higher eukaryotes, while not effecting branch formation, did prevent the second step of splicing[110-112] *i.e.*, the intron-3'-exon cleavage and exon-exon ligation. This appears to serve as a positive control for execution of the second step of splicing reaction. Of interest is the activation of cryptic branchpoints on mutation of branchpoint sequences in higher eukaryotes.[113] Splicing using these alternative branch sites was seen to proceed with no problems. In all these cases an alternative adenosine residue was seen to serve as the branch formation residue.

Recent work by Chattopadhyaya *et al.*[114] has shown that changing the branchpoint adenosine to a guanosine or uridine led to changes in conformation around the branch structure. Thus, it could be speculated that formation of the branch structure effects the first step in splicing while the actual conformation, once formed, effects the second step of the mRNA splicing reaction. Much work remains to be done in order to provide a complete and clearer explanation for formation of the branched RNA during splicing and if it serves any role in the regulation of gene expression and genetic diversity.

1.2.3 Branched RNA metabolism

The normally low abundance of branched RNA relative to that of formed mature mRNA suggests that either the branched RNA structures are being degraded rapidly or that the branch structure is specifically hydrolyzed leading to linear RNA molecules[103]. Thus, the released intron sequence may also be functioning as a signal for recognition by nucleases for degradation and nucleotide recycling.

Stability of branched molecules in the spliceosome complex has been attributed to the binding of some nuclear factors which protect the RNA from nuclease degradation.[91,115-117]

In 1985, Ruskin and Green[118] observed a novel enzymatic activity in the HeLa (mammalian) cell extracts that converted RNA lariats to the corresponding linear RNA molecules, which they termed "debranching activity". This debranching activity was seen to be specific for the 2',5' phosphodiester linkage solely in the branch structure. The other interesting finding of the study was the fact that *in-vitro* synthesized RNA lariats formed *in vitro*, were only seen to undergo cleavage if first deproteinized and then added back to the nuclear or cellular extract.[118] This suggested that the branch structure was protected against degradation from the debranching activity during splicing.

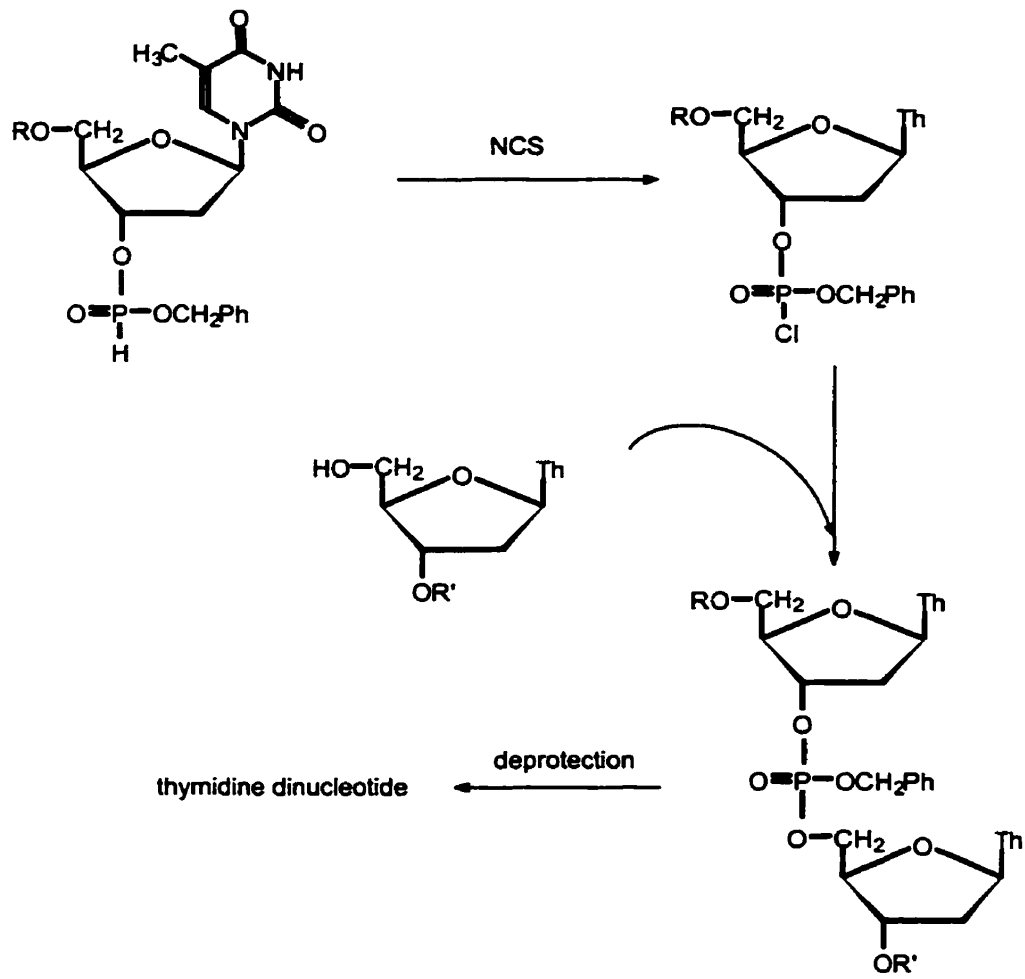
Recently, the gene encoding yeast (*Saccharomyces cerevisiae*) debranching enzyme was isolated and characterized by Chapman and Boeke. They showed that in the absence of debranching activity *in vivo*, the introns are quite stable and accumulate to levels exceeding that of the corresponding mature RNAs. Thus, they concluded that the debranching enzyme carries out a rate limiting step in the metabolic pathway of intron turnover in mammalian cells and in yeast. More recently, Boeke's group[120] has shown that a yeast mutant (*Schizoachharomyce pombe*) lacking the debranching enzyme had severe growth defects. Thus efficient intron lariat degradation seems to be critical for normal cell growth and function. The debranching enzyme is a crucial and so far the only identified component of this intron lariat degradation pathway.

More recently, biosynthesis of intron-encoded small nucleolar RNAs in *Saccharomyces cerevisiae* was shown to be dependent on the RNA lariat-debranching enzyme.[121] In mutants lacking the lariat debranching enzyme gene (*dbr1* mutant) intronic snoRNAs were found to be "trapped" within the host intron lariat.[122] Interestingly, the *dbr1* mutation did not block site specific rRNA methylation targeted by U24 snoRNA, suggesting that at least for U24 snoRNA, snoRNAs in the form of lariats may still be functional.

1.3 CHEMICAL SYNTHESIS OF NUCLEIC ACIDS

Two years after discovery of the helical structure of DNA, Michelson and Todd[123] reported the chemical synthesis of the dinucleotide monophosphate 3',5'-linked, scheme **1.2**. From this very first synthesis of a dinucleotide there has been a great progress in the synthesis of larger oligonucleotides. Over the years, four major approaches to oligonucleotide synthesis have been reported: the phosphotriester, phosphodiester, the phosphite triester, and the phosphonate approach. The four methodologies differ by the type of phosphate ester produced at the end of the condensation reaction and are named based on the internucleotide bond formed.

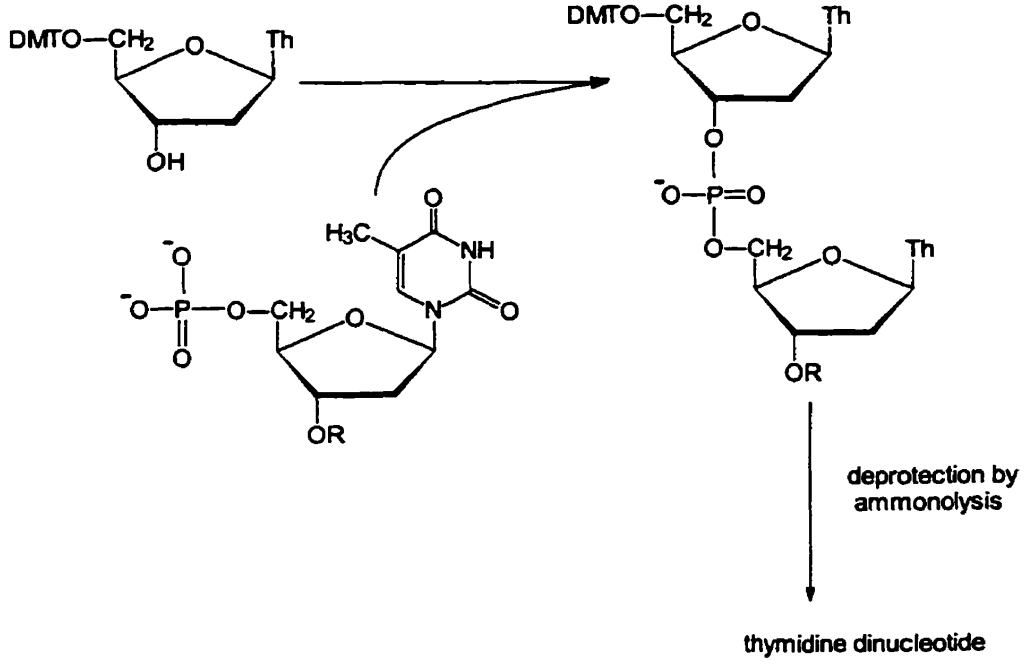
Scheme 1.2



No discussion of oligonucleotide synthesis would be complete without mentioning the pioneering work of Khorana and colleagues in the synthesis of biologically active genes during the late 1960s and early 1970s. Based on carbodiimide-mediated condensation reactions, Khorana introduced the phosphodiester approach to oligonucleotide synthesis.[124] Using activating agents such as N,N-dicyclohexylcarbodiimide (DCC), mesitylenesulfonyl chloride, or 2,4,6-triisopropylbenzenesulfonyl chloride, Khorana was able to carry out the condensation of a 3'-protected nucleoside 5'-monophosphate and a suitably protected nucleoside with a free 3'-hydroxyl function, scheme **1.3**.[124] Extension of the nucleotidic chain was carried out by removal of the 3'-hydroxyl protecting

group and subsequent condensation with a suitably protected nucleoside 5'-monophosphate.[125,126] Using these short synthetic oligonucleotides, the genetic code was deciphered in 1968, some 24 years after identification of DNA as the hereditary information carrier. Oligonucleotides with lengths up to 20-25 were then synthesized by using the "block condensation" of small oligonucleotide chains.[127] Synthesis of even longer chains was accomplished using a combination of chemical and enzymatic methods.[128,129] For example, DNA ligase served to link chemically synthesized shorter chains during the total synthesis of the gene coding for the yeast alanine tRNA[128] and the *E. coli* tyrosine suppressor tRNA.[129]

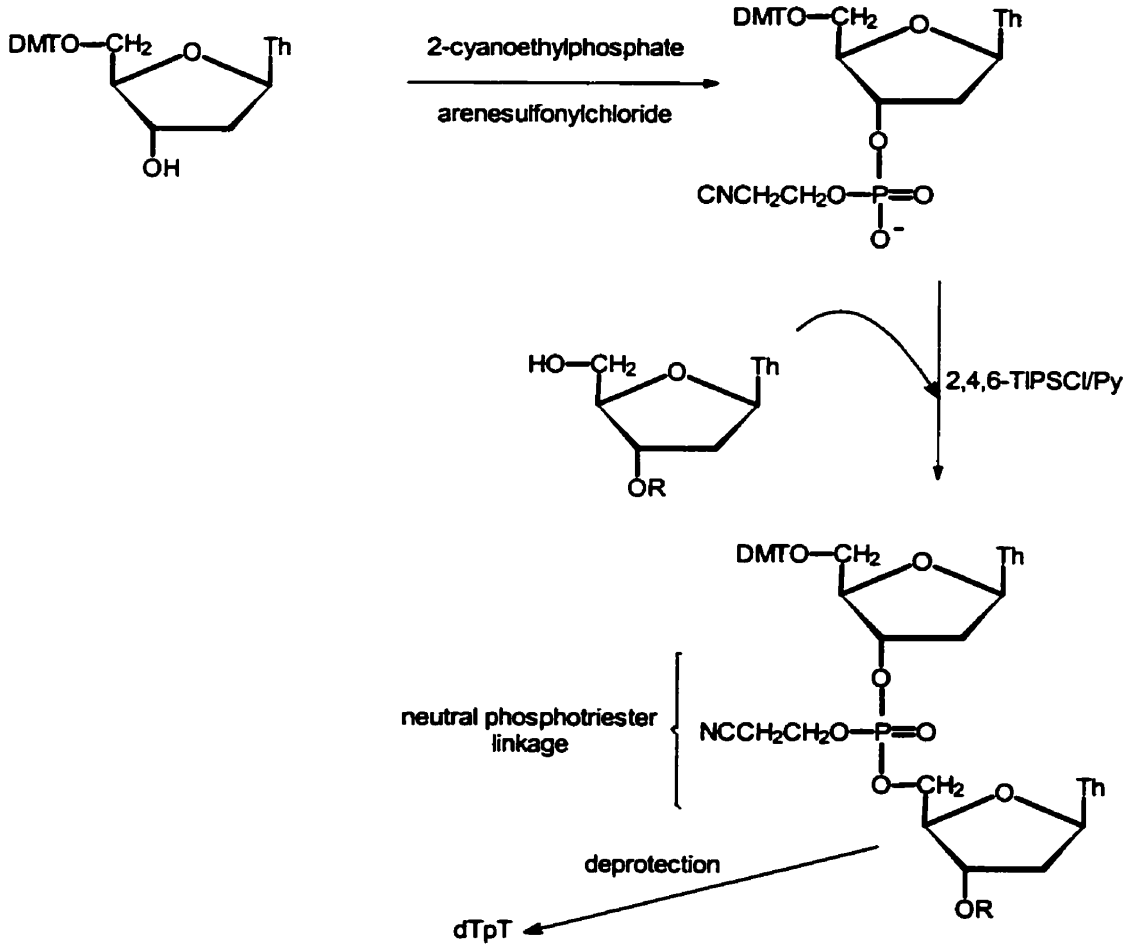
Scheme 1.3



During this period, Khorana also developed trityl group derivatives for the protection of the 5'-hydroxy function of nucleosides along with the benzoyl and isobutyryl groups for the protection of the exocyclic amino function of adenine and guanine, respectively. These protecting groups have become the standard in nucleic acids chemistry and are still employed in modern oligonucleotide synthesis.

Despite successful utilization of phosphodiester approach to the synthesis of biologically active genes, this strategy suffered from some major inherent drawbacks, including low solubility of reactants and products in organic solvents due to their ionic nature. The ionic nature of reactants and products also required the use of tedious and time consuming ion-exchange chromatography for purification after each coupling step. Also, the internucleotide phosphodiester linkages could act as nucleophiles during subsequent phosphorylation reactions, resulting in lower coupling yields with increasing oligonucleotide lengths. Another problem was the phosphorylation of the phosphodiester linkage leading to pyrophosphate formation which then led to the solvent assisted cleavage of internucleotide linkages.[130]

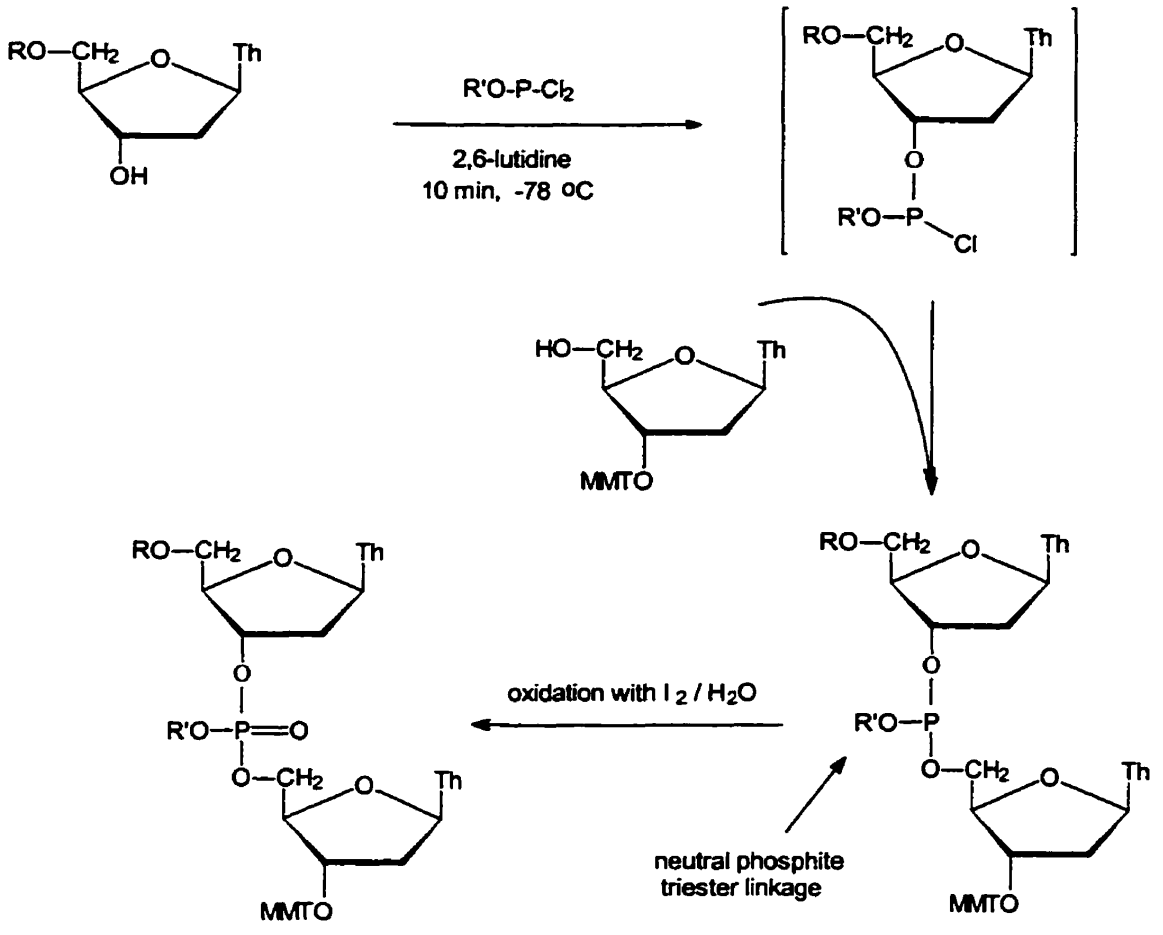
Scheme 1.4



In order to alleviate the problems inherent in the phosphodiester approach, the phosphotriester method of Michelson and Todd was reinvestigated by Letsinger's group.[131-134] The resulting fully protected, neutral molecules were easily adaptable to isolation and purification by standard chromatographic methods, scheme **1.4**. Introduction by Letsinger and coworkers of the ß-cyanoethyl group for phosphate protection led to the renewal of the phosphotriester approach[133] for both DNA and RNA synthesis. This approach remains a popular method for large scale solution-phase synthesis of oligonucleotides.

Further work by Letsinger and colleagues in mid 1970s revolutionized the chemical synthesis of oligonucleotides through the development of the "phosphite triester" methodology,[135-137] scheme **1.5**. This approach took advantage of the high reactivity of chlorophosphites with nucleosides and the rapid oxidation of phosphite triesters to the phosphate triesters by aqueous iodine. Based on the phosphite triester methodology, Ogilvie *et al.*[138] reported the world's first machine-assisted synthesis of an oligodeoxynucleotide. The phosphite coupling suffered from one drawback; the highly reactive chlorophosphite intermediates were moisture sensitive and thus difficult to handle, making this methodology not very easily amenable to automation. Also due to the bifunctional nature of the phosphite reagent, symmetrical 3',3'- and 5',5'-linked products were also obtained, although such dimers usually did not cause any purification problems.

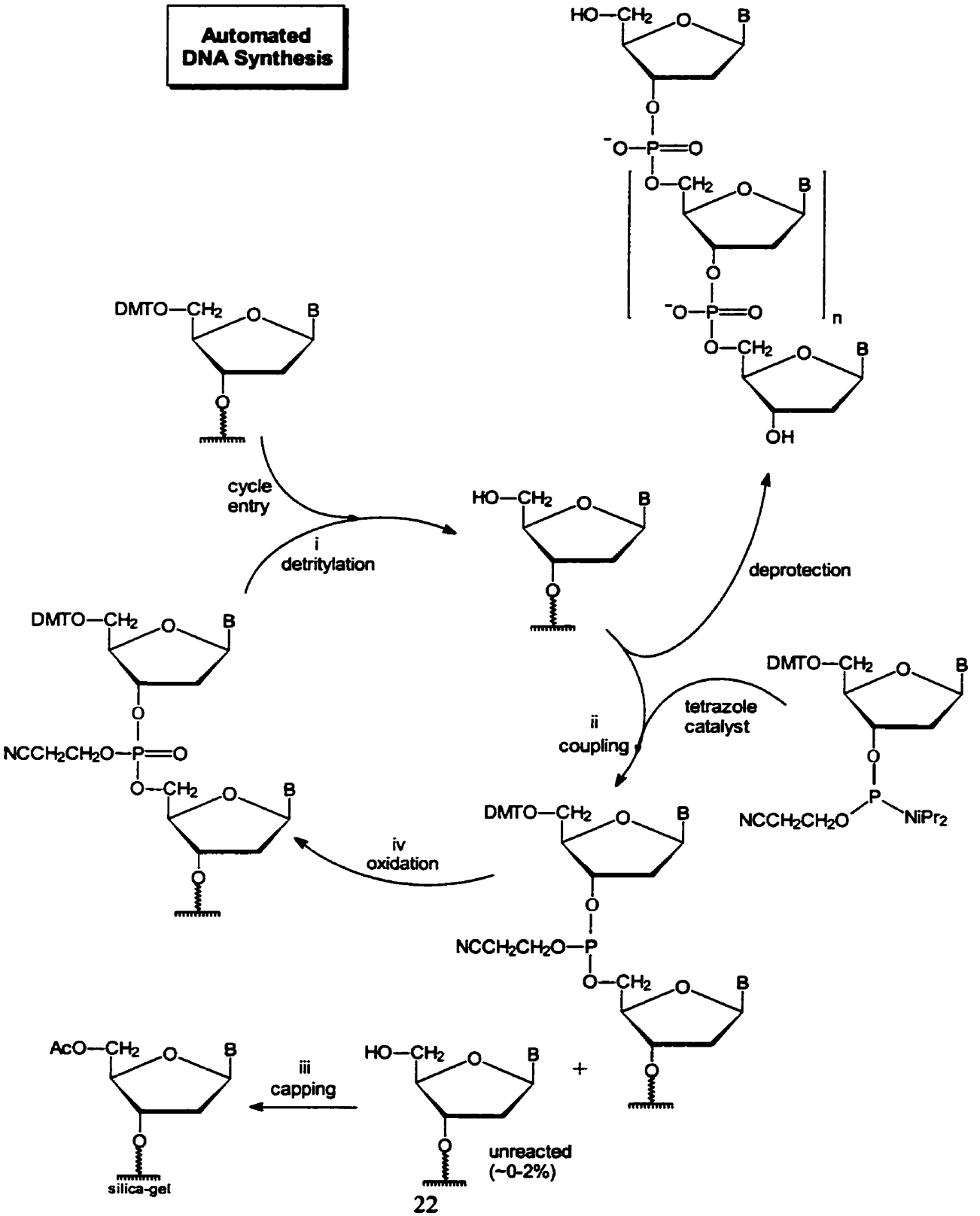
Scheme 1.5



To alleviate the problems associated with chlorophosphite intermediates in automation, Caruthers[139,140] introduced nucleoside N,N-dialkylamino and N-morpholino phosphoramidites as a new class of building blocks for the synthesis of oligonucleotides. These new intermediates were easy to purify, isolable as stable powders and could be stored before use at –20°C for long periods of time. These phosphoramidite reagents were easily activated by weak acids such as tetrazole (pKa ~4.9) and gave near quantitative coupling yields within one minute. The use of these more stable phosphoramidite derivatives, and the

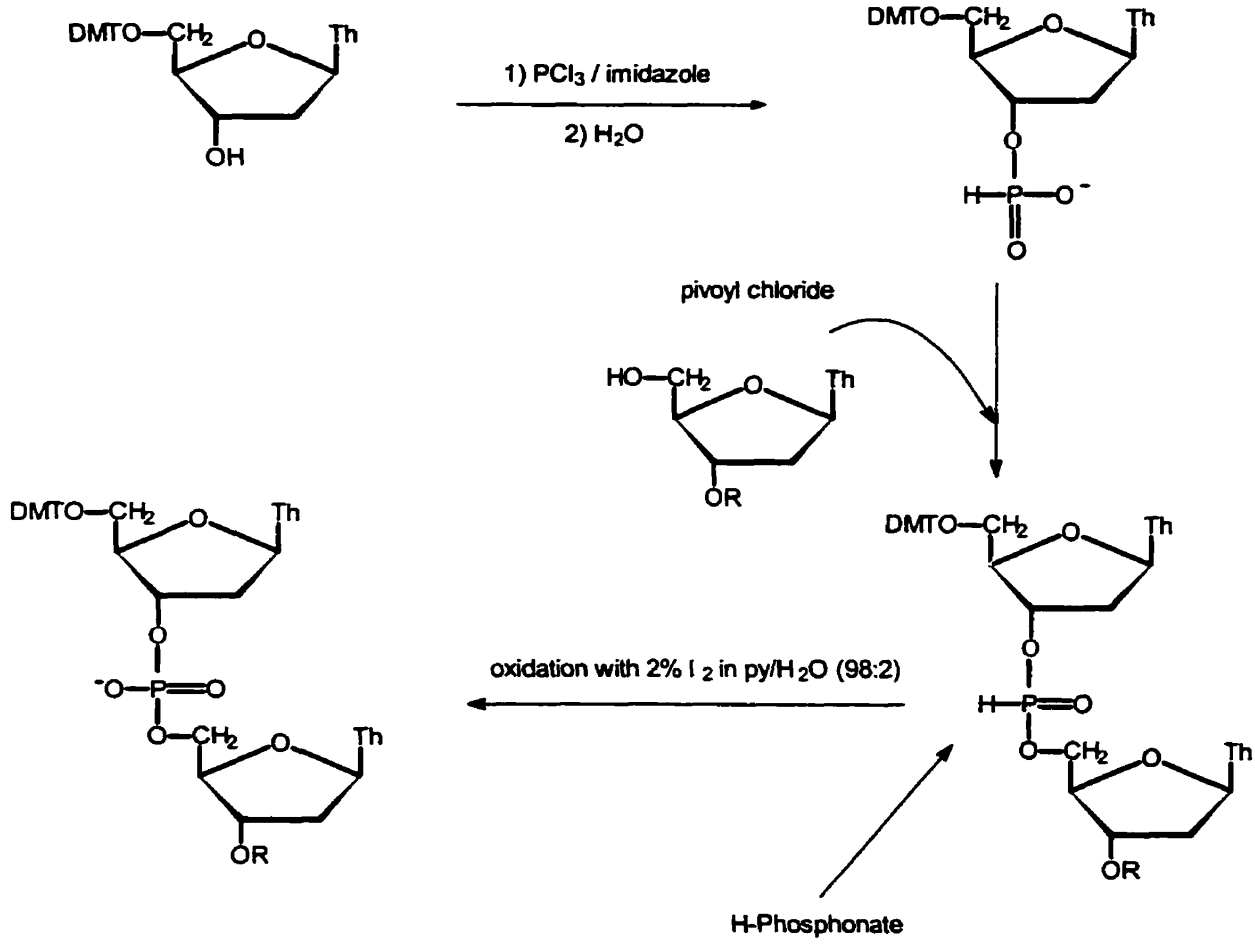
development of silica gel-based solid supports, led to the successful commercialization of automated DNA/RNA synthesizers or "gene machines".[138]

Scheme 1.6



In the mid 1980s, Gregg *et al.*[141] and Froehler *et al.*[142,143] took a different approach to the synthesis of oligonucleotides. Based on the work of Todd and Hall,[144], this approach has come to be known as the "H-phosphonate" approach as a phosphonate internucleotidic linkage is formed after condensation, scheme **1.7**. In this method, a nucleoside 3'-O-hydrogen phosphate was coupled with the free 5'-hydroxyl of a suitably protected nucleoside, leading to a dinucleotide linked *via* a hydrogen phosphonate group. Upon oxidation, the phosphonate linkage was converted to the phosphodiester. The H-phosphonate diester is relatively inert to further phosphitylation and the chain can be extended without prior oxidation. Simultaneous oxidation of all phosphonates at the end of the synthesis allows for a faster synthesis cycle. Also oxidation is subject to general base catalysis and this allows nucleophiles other than water to be substituted during oxidation to give a range of internucleotide modified oligonucleotide analogues, including phosphorothioates and phosphoramidates. The only known drawback of the H-phosphonate is the marginally lower stepwise coupling yields obtained in comparison to the phosphite triester approach.

Scheme 1.7



Presence of the additional 2'-hydroxyl functionality in RNA makes its synthesis much more complicated. Considerable attention has focused on whether the protecting groups for the 2'-OH should be acid or base labile. Work by Ogilvie and co-workers[145,146] led to the development of *t*-butyldimethylsilyl (TBDMS) group for protecting the 2'-hydroxyl during RNA synthesis. Originally described by Corey[147], the TBDMS group is quite stable under acidic and basic conditions and is easily removed under neutral conditions in the presence of fluoride ions. Other 2'-protecting groups, especially the acid labile N-substituted piperdin-4-yl acetal derivatives,[148-150], have been reported during RNA synthesis. Nevertheless, the TBDMS has remained the protecting group of choice in routine RNA synthesis.

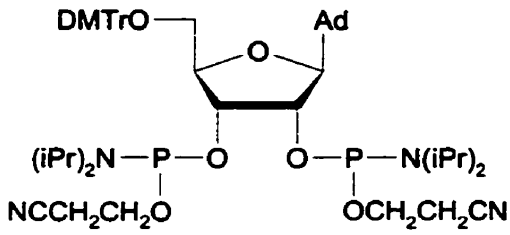
Advances in nucleic acid chemistry have had a tremendous impact on a wide and diverse variety of disciplines. Molecular biology, medicinal and biological chemistries have benefited enormously from the advances made in nucleic acid chemistry. Recently, Adleman[151] using synthetic oligonucleotides and simple DNA manipulations solved a directed Hamiltonian path problem, and suddenly opened up a hot new field in computer science: "Computing With DNA". Although still in its infancy, research on DNA based computing may eventually offer insight into mechanisms that underlie such important biological systems as evolution and the immune system.

1.4 CHEMICAL SYNTHESIS OF BRANCHED NUCLEIC ACIDS

Ever since the identification of branched RNA in 1983, several reports on the synthesis of branched nucleic acids have appeared. Most of these made use of the solution-phase method and demonstrated synthetic difficulties in introducing vicinal 2',5' and 3',5' phosphodiester linkages in a stepwise manner (see chapter 2; section 2.1, for a more detailed review).

Initial work by Damha and coworkers[152,153] is of particular importance since it laid the foundations for the later solid-phase synthesis of branched nucleic acids. Originally Damha and colleagues[152] introduced the bisphosphoramidite branching synthon, **1.1**, for simultaneous introduction of the nucleotides at the 2' and 3' positions in solution. Later on, Damha *et al.*[154,155] extended the bisphosphoramidite approach to the first solid-phase synthesis of branched nucleic acids, resulting in branch structure with identical sequences at the 2'- and 3'- positions.

Caruther's group[156] was the first to report the regiospecific solution-phase synthesis of a natural core as found in the branchpoint of RNA lariats.[103] Recently, Sproat and colleagues[157] have reported the regiospecific solid-phase synthesis of small to medium branch molecules using elaborate synthons.



1.1

1.5 OTHER NATURALLY OCCURING BRANCHED NUCLEIC ACIDS

Interest in the structure and function of nucleic acids has led to the observation of a wide structural diversity in nucleic acids. The so called A-, B-, and Z- forms are well characterized.[158] Unusual structures, such as branched nucleic acids, are discussed below to give the reader a feeling for the variety of natural occurrences of branched, or branched-like nucleic acids beside the well known intron-lariats.

1.5.1 msDNA

Originally observed by Yee *et al.*[159] in 1984, gram negative soil bacteria *Myxococcus xanthus* contains an unusual extrachromosomal element having an unusual branched structure consisting of single stranded DNA covalently joined to RNA. Present in 500 to 700 copies per genome, this peculiar satellite DNA element was designated "multicopy single stranded DNA", (msDNA), figure **1.3**. msDNAs have also been observed to be present in other Gram-negative bacteria and various *Escherichia coli* strains including some clinical isolates.[32,160,161]

Although msDNA from various sources show little or no nucleotide sequence homology, a few structural features have been found to be common among all msDNAs, figure **1.3**. Firstly, they all have a short single-stranded DNA linked at its 5'-end, through a 2',5'-phosphodiester linkage, to a completely conserved guanosine residue in the middle of an RNA species. Secondly, both the DNA and the RNA molecules have stable secondary structures. Thirdly, the last few nucleotides at the 3' ends of the DNA and RNA species form a stable DNA-RNA hybrid. Surprisingly, msDNAs appear to be quite stable in cells, in spite of the fact that the molecule contains single-stranded DNA and RNA regions (figure **1.3**). This stability has been attributed in part to the presence of the branch structure, the formation of DNA-RNA hybrid duplex at the 3' end, and the stem-loop structures present in the single stranded regions.

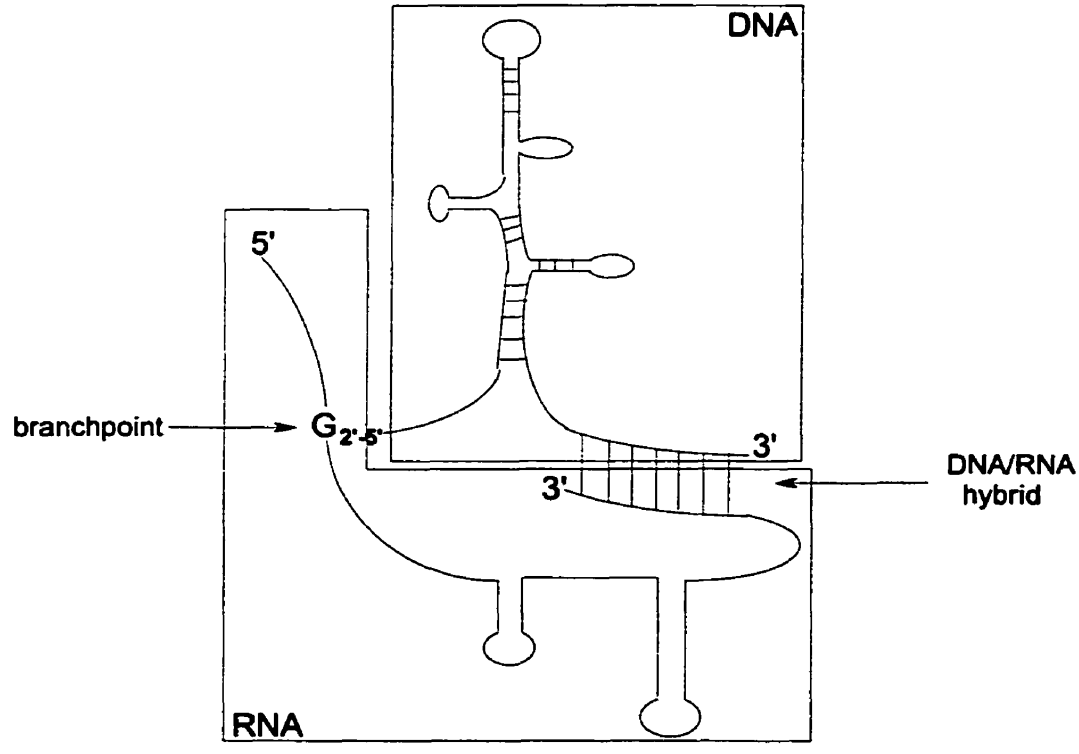


Figure 1.3: msDNA structure

Considerable focus has been directed towards msDNA molecules to elucidate the mechanism of its biosynthetic formation and possible functions they may serve. The known essential components required for msDNA biosynthesis are seen confined to a short (2-3 kb) DNA sequence on the bacterial chromosome, and belong to a single operon, present in only one copy per chromosome.[162] This single locus encodes, in order, the genes for msdRNA (*msr*), msDNA (*msd*), and for a bacterial reverse transcriptase (*ret*), that is structurally similar to the polymerases of retroviruses. The genes are transcribed from a promoter located upstream region of *msr*. The coding regions for msdRNA (*msr*) and msDNA (*msd*) are oriented in opposite directions, overlapping by 5 to 8 bases at their 3' ends. Based on their similarities to other retroelements, genetic elements encoding reverse transcriptase, Temin[163] proposed the term *retron* to describe the genetic element required for msDNA synthesis. Whereas other retroelements such as retroviruses, retrotransposons, and pararetroviruses contain other genes beside the RT, retrons are known to contain only a RT gene and thus could be considered essentially a primitive form of retroelements.

Eukaryotic reverse transcriptase carries out an essential step in the transposition of retroelements through production of single stranded cDNA from an RNA template (see Boeke and Corces[164] for a review). Transposition is the process by which copies of certain genetic sequences are inserted into nonhomologous target sequences and serves as a conduit for the integration of retroviruses into the mammalian genomes, production of infectious retroviruses from proviruses integrated into the genome, and formation of pseudogenes in eukaryotic cells. Although, RT assisted production of single-stranded cDNA from an RNA template serves an essential step in the transposition mechanism of eukaryotes, no such RT produced single-stranded cDNA has been detected in eukaryotes *in vivo*.

On the other hand, stability of msDNA, essentially a RT produced cDNA, lead to their detection in prokaryotes *in vivo*.

Currently two different models are under consideration as possible biosynthetic pathways leading to the synthesis of mature msDNA. In the model put forward by Lease and Yee,[165], a conventionally primed single-stranded DNA molecule is joined with the RNA molecule in a branch ligation reaction taking place several minutes after the synthesis of both molecules. On the other hand, work by Inouye *et al.*[166] demonstrated that RNA but not DNA was responsible for msDNA synthesis. In their model, the 5'-end sequence of the pre-RNA transcript forms a duplex with the 3'-end sequence of the same transcript and serves as a primer as well as template for msDNA synthesis. Addition of a dNTP to the RNA transcript through a 2',5'-phosphodiester linkage at the branchpoint primes the synthesis of the complementary DNA by reverse transcriptase. Mutations at the branchpoint guanosine lead to inhibition of the mature msDNA formation. In both proposed pathways, the initial synthesis product is processed by ribonuclease H (RNase H), leading to the mature msDNA.

Recent work suggests that msDNA may be involved in regulation of mutagenesis.[167-169] However more work needs to be done before a definite answer as to the function of msDNA will emerge.

1.5.2 Forked Nucleic Acids with Noncovalent Branches

Beside RNA intron-lariats and msDNA, nucleic acids with noncovalently linked branches are also known, e.g. those forming during genetic recombination. Genetic recombination involves the interaction of two segments of double stranded DNA to yield new genetic material that may incorporate parts of both

interacting molecules. The mechanism for such recombinations involves branched structures referred to as the Holliday[170] junction intermediates. Typically, Holliday junctions have four-stranded structures in which two strands have been nicked and religated together so that they fuse the two strands together and form a branchpoint, the other two strands remain intact and are not directly involved in branching.

Other known instances of noncovalent branching are the multistem structures. These generally occur as stable structural elements of biologically active nucleic acids and as intermediates in nucleic acid metabolism. Some examples include the 5S, 16S and 23S rRNAs, and the hammerhead ribozymes. For a review of these multistem nucleic acid structures see the report of Lilley and Clegg.[171]

1.6 PROJECT OBJECTIVES

Based on the natural occurrence of branched nucleic acids (bNAs) and the scarcity of methods for their solid-phase synthesis, studies were undertaken for the development of a regiospecific synthesis of bNAs. These studies led to the synthesis of branched oligodeoxynucleotides with different sequences at the 2'- and 3'- positions of the branchpoint. Slight modifications of this method resulted in the first chemical synthesis of a msDNA branch core.

The conformational rigidity imparted to the branched DNA chains, by the branchpoint ribose, will be exploited for inducing the formation of duplex and triplex DNA. Previous work in our research group on branched nucleic acids had shown that they are capable of forming triple-helical structures.[172] C-rich synthetic branched nucleic acids have also been shown by members of the Damha group to form remarkably stable C-tetrad structures.[173] However due to the

synthesis methods in use at the time, the association of branched molecules into duplex structures had not been previously investigated. With a new synthesis method available, such structural studies become possible, and will be undertaken. Furthermore, the stability of branched triple helices containing T:A/T triads will also be analyzed and quantitated.

In the last part of the thesis, the usefulness of synthetic branched oligonucleotides as tools for studying biological processes involving bNAs will be demonstrated. Specifically, small branched NAs will be used to probe the substrate specificity of the yeast debranching enzyme, while some of the larger branched nucleic acids will be used for studying pre-mRNA splicing *in vitro*.

2. RESULTS AND DISCUSSION

NOVEL APPROACHES TO THE SYNTHESIS AND ANALYSIS OF BRANCHED NUCLEIC ACIDS

2.1 INTRODUCTION

Correct gene expression in eukaryotic cells depends on the accurate splicing of messenger ribonucleic acid precursors (pre-mRNA). This process involves the formation of a branched RNA (bRNA) intermediate known as the RNA intron "lariat". The key feature of these branched nucleic acid structures is the presence of vicinal 2',5'- and 3',5'-phosphodiester linkages.[103] The excised intron lariats are then quickly degraded in vivo, with a half life of only a few seconds,[174] and thus are not stable enough to be detected in vivo. Novelty of these structures has raised interest as to their role in regulating precursor mRNA splicing and possible influence on cell viability and growth. In addition, there has been increasing recent interest in synthetic branched DNA and RNA for use in diagnostic applications,[175] as "molecular anchors" for inducing the formation of duplex bRNA:RNA and novel triple-helical DNA,[172,176-180] and as tools for studying branched mRNA splicing.[181] Further advancements in methodologies for the *regiospecific* solid-phase synthesis of branch nucleic acids will serve an important function in supplying material in large quantities for studying bRNA structure and function.

2.2 CHEMICAL SYNTHESIS OF BRANCHED NUCLEIC ACIDS

The chemical synthesis of branched nucleic acids is not trivial. The sequential, regiospecific introduction of phosphate linkages at the vicinal 2'- and 3'-hydroxyls at the branchpoint has been the major source of problems encountered during their

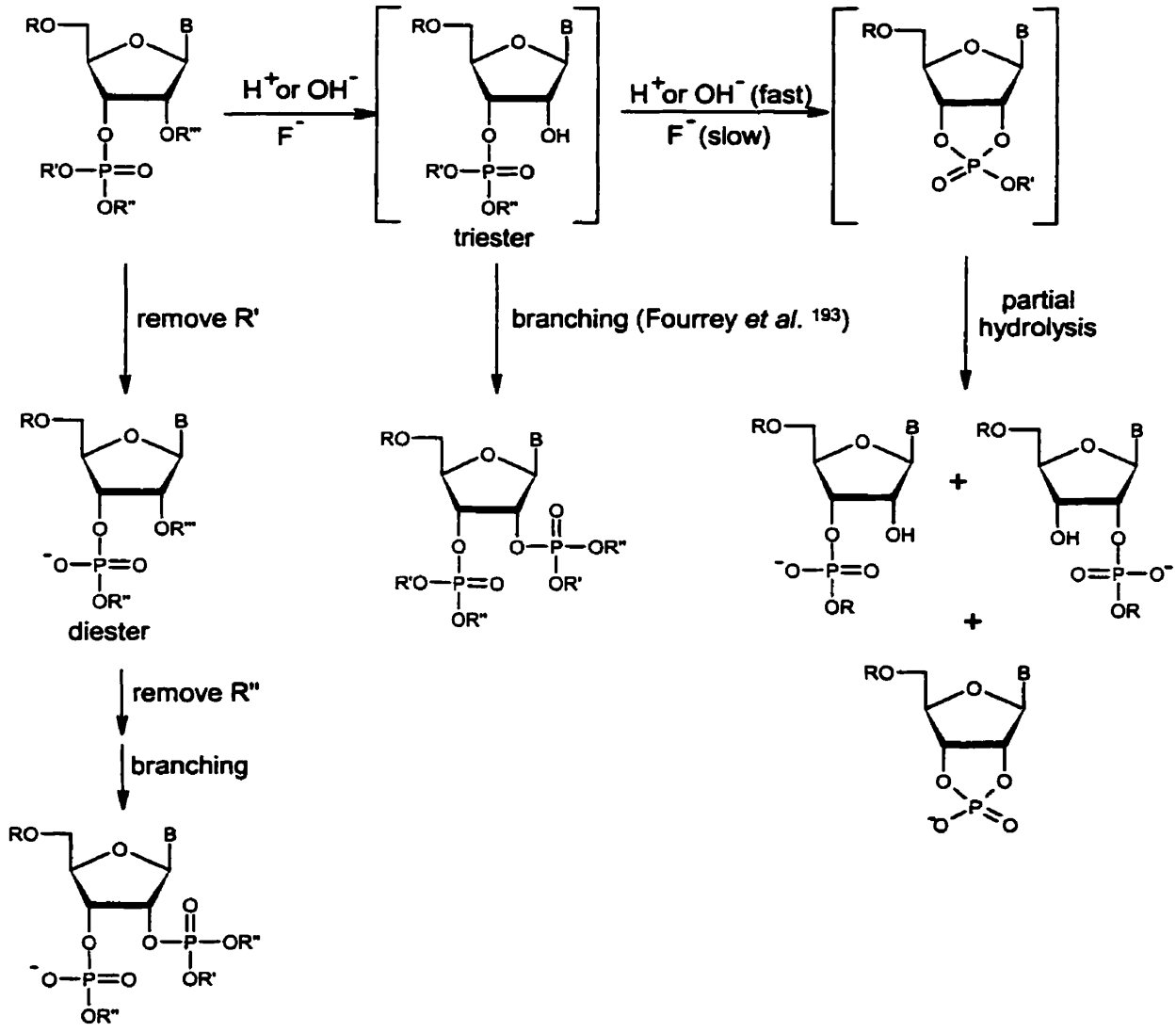
chemical synthesis. It has been demonstrated that the nature of the phosphate group dictates the nucleophilic reactivity of a vicinal hydroxyl function towards the phosphate.[182-185] *Phosphotriester* linkages are known to be very susceptible to nucleophilic attack by the vicinal hydroxyl group under neutral, acidic and basic conditions.[152,182,184-187] This leads to phosphoryl migration and/or chain cleavage during deblocking of the neighboring hydroxyl protecting group, scheme **2.1**. On the other hand, a *phosphodiester* linkage is reasonably stable under conditions used for the removal a variety of hydroxyl protecting groups.[186,188-190] This stability allows for the sequential introduction of phosphate linkages at the 2'- and 3'- hydroxyls of the branchpoint after conversion of the phosphotriester to phosphodiester linkage.

Kierzek *et al.*[186] were the first to exploit the stability of phosphodiester linkages in the solution-phase synthesis of a branched tetramer, $GA(^{2'5'}G)_{3'5'}C$. After assembling a fully protected $A_{3'5'}C$ dimer bearing a TBDMS group on the 2'-position of adenosine, the β-cyanoethyl group of the phosphotriester linkage was removed selectively. The silyl group was then removed without any noticeable phosphoryl migration and/or chain cleavage. Subsequent coupling of the newly free 2'-hydroxyl with a suitably protected guanosine-5'-phosphoramidite gave the branched trimer $A(^{2'5'}G)_{3'5'}C$, which was then extended from the 5'-position to afford the tetramer.

In a similar approach, Sekine and Hata carried out the solution-phase synthesis of the branched trimer $A(^{2'5'}G)_{3'5'}C$.[191] Improving on this methodology, Sekine, Hejkkilä and Hata reported the solution-phase synthesis of branched oligomers up to six nucleotides in length.[192,193] However the overall yields attained were low.

Fourrey *et al.*[194] dealt with the instability of the phosphotriester by carrying out the desilylation of the vicinal hydroxyl function under mild conditions (F^-, 0 °C), which in turn, slowed down the subsequent rapid attack of the released 2'-hydroxyl function on the vicinal phosphotriester linkage. Phosphitylation of the 2'-hydroxyl enabled its subsequent coupling to a suitably protected nucleoside monomer, scheme **2.1**. Overall yields were seen to be very low since complete desilylation could not be achieved without significant chain cleavage.

Scheme 2.1



Huss *et al.*[189,195] used the acid labile 4-methoxytetrahydropyran-2yl (Mthp) protecting group for the 2'-hydroxyl function of the branchpoint to prepare several branched trimers in solution. The Mthp group was removed under mild acidic conditions after conversion of the phosphotriester linkage into the phosphodiester, and branching effected by coupling of a suitably protected nucleoside 5'-phosphoramidite.

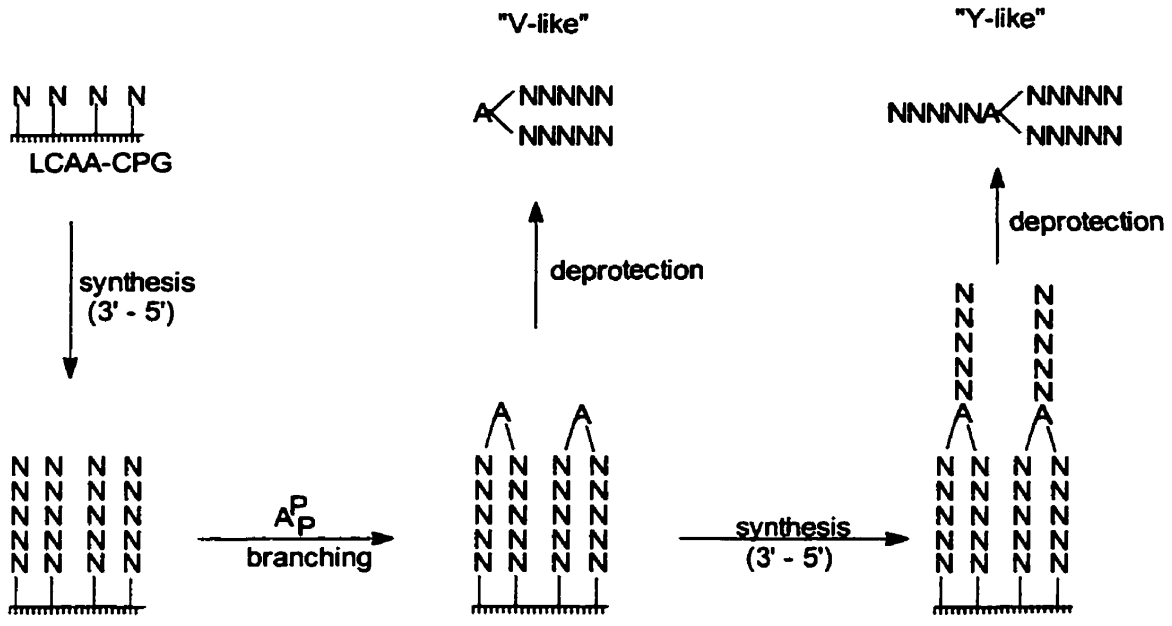
Owing to the involvement of the neighboring hydroxyl group, Huss *et al.*[183] were unable to synthesize branched oligomers using the H-phosphonate chemistry. However, Balgobin[196] reported the successful synthesis of small branched molecules by first oxidizing the H-phosphonate linkage into the phosphodiester before cleavage of the vicinal hydroxyl protecting group. Sund *et al.*[190,197] have reported the synthesis of small cyclic branched RNA oligomers using this H-phosphonate/phosphoramidite approach.

2.2.1 Solid Phase Synthesis

Early work of Damha and coworkers[152,153] laid the foundation for the solid-phase synthesis of branched oligomers. They overcame the problems associated with sequential phosphorylation by simultaneous introduction of both the 2',5'- and 3',5'-phosphotriester linkages. This was achieved by either condensing 5'-MMT-N6-benzoyladenosine with an excess of a suitably protected nucleoside 5'-phosphoramidites or by reaction of a protected adenosine-2',3'-bis(phosphoramidite) derivative with an excess of nucleoside bearing a free 5'-hydroxyl. This methodology was not regiospecific as the synthesis of $A(^{2'5'}X)_{3'5'}Y$ also yielded $A(^{2'5'}X)_{3'5'}X$, $A(^{3'5'}Y)_{3'5'}Y$, and $A(^{2'5'}Y)_{3'5'}X$ which, in the majority of cases, could easily be separated and purified by preparative TLC and/or HPLC.

In 1987, Damha and Zabarylo[154] reported the first solid-phase synthesis of branched oligomers. First, linear oligomers were assembled on the controlled pore glass solid support. Then by reacting adenosine-2',3'-bis(phosphoramidite) derivative, **1.1**, with the free 5'-OH groups of two adjacent support bound linear oligomers, they were able to obtain branched oligomers having identical sequences at adenosine's 2'- and 3'-positions[155], scheme **2.2**.

Scheme 2.2



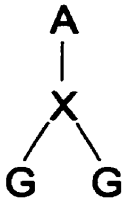
Recently, Sproat and coworkers[157,198,199] have reported on the regiospecific solid-phase synthesis of small to medium sized branched oligoribonucleotides. Although the method is regiospecific, it is laborious and requires an elaborate set of phosphoramidites and branchpoint synthons, thus potentially limiting the ease and speed synthesis of the small branched nucleic acids. Given these limitations, we directed our attention towards alternative strategies for the synthesis of bRNA, particularly, the regiospecific solid-phase synthesis of branched oligonucleotides.

2.3 SYNTHESIS OF SYMMETRIC* BRANCHED TETRARIBONUCLEOTIDES

Preliminary studies by Boeke and Damha's groups[200] on the yeast debranching enzyme (yDBR) have shown that branched tetraribonucleotides are of sufficient size to be substrates for the enzyme (see chapter 4 for more details). Given this, we initially undertook the solid-phase synthesis of small branched oligonucleotides, table **2.1**, systematically replacing the nucleotide at the branchpoint. This would allow us to probe the yeast debranching enzyme in terms of the nature of the branchpoint nucleotide. Thus using the methodology developed by Damha *et al.*[154], the branchpoint nucleotide was introduced by coupling the appropriate nucleoside 2',3'-bis(phosphoramidite) synthons, **2.5** - **2.8**, to a guanosine residue attached to LCAA-CPG solid support.

Table 2.1: Branched tetrameric RNAs synthesized using the convergent methodology of Damha *et al.*

	Sequence
A	$AG(^{2'5'}G)_{3'5'}G$
B	$AU(^{2'5'}G)_{3'5'}G$
C	$AC(^{2'5'}G)_{3'5'}G$

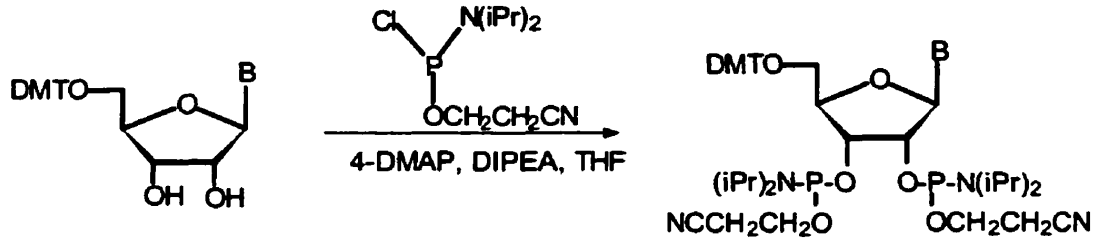


$\equiv\equiv$ $AX(^{2'5'}G)_{3'5'}G$

Scheme **2.3** shows the synthetic scheme for the preparation of the branchpoint bisphosphoramidite synthons **2.5** – **2.8**. The required monomers, **2.1** - **2.4**, are also convenient intermediates in the synthesis of nucleoside 2'-TBDMS-3'-phosphoramidite monomers used in the solid-phase synthesis of RNA.

*symmetric here means that the two strands at the 2'- and 3'- position of the branchpoint are the same.

Scheme 2.3

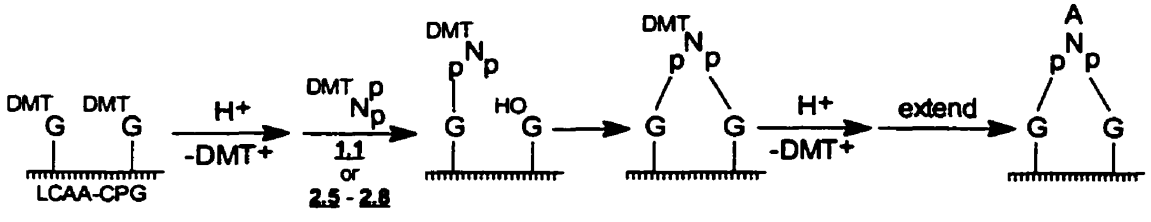


	B
1.0	Ad^{Bz}
2.1	Gu^{ibU}
2.2	Ur
2.3	Cy^{Bz}
2.4	$Gu^{NPE,Bz}$

	B
1.1	Ad^{Bz}
2.5	Gu^{ibU}
2.6	Ur
2.7	Cy^{Bz}
2.8	$Gu^{NPE,Bz}$

Damha and coworkers[155] have shown that the efficiency of the branching reaction is dependent on both the degree of derivatization of the solid support and the solution concentration of the bis(phosphoramidite) reagent 1.1. Thus a solid support with a high degree of substitution gives best results as it ensures an optimal distance between the adjacent 5'-hydroxyl groups for the branching reaction. Whereas deoxynucleoside and ribonucleoside phosphoramidites are commonly used at concentrations of 0.10 M and 0.15 M respectively, bis(phosphoramidite) concentrations of *ca.* 0.02 - 0.03 M were required to form bRNA in good yields. Under these conditions only a fraction of the free 5'-hydroxyl groups become phosphitylated by 1.1, favoring the branching reaction, scheme 2.4.

Scheme 2.4



The adenosine-, **1.1**, cytosine-, **2.7**, and uridine-, **2.6**, 2',3'-bisphosphoramidites were seen to couple to rG (LCAA-CPG) in good yields as determined from quantitation of released trityl cation, scheme **2.4**. After introduction of the branch junction, a one-half reduction in released trityl would indicate a successful branching reaction as two chains are coupled together to one bis(phosphoramidite) monomer. However it has been suggested that for the singly-branched RNAs, a trityl quantitation showing a 50% drop (after branching) usually represents a coupling of 50% of the population of oligomers on the solid-support, 25% monophosphitylation and 25% branching.[181] Low coupling indicates the absence of branch formation. Branch introduction with guanosine 2',3'-bisphosphoramidite, **2.5**, was low as seen from trityl quantitation, (25.4%). However PAGE analysis indicated that the crude product, **A**, was cleaner than the other cases, figure **2.1**. Further analysis showed that branching by **2.5** was very inefficient. Thus the putative AG($^{2'5'}$G)$_{3'5'}$G branched tetramer, **A**, was seen to move similarly to the ^{31}P-AGG linear trimer after ^{32}P labeling at the 5'-end, indicating the absence of branch formation, figure **2.2**. This observation was taken to be consistent with the long-standing problem of side reactions of guanine residues during oligonucleotide synthesis. Further evidence of no branch formation with guanosine bis(phosphoramidite), **2.5**, was obtained from the debranching studies, chapter 4, section 4.2 and figure **2.2**.

Table 2.2: Coupling yields and purification of branched tetramers **A** – **D**.

	Sequence	$\%X^{P}{}_{P}$ coupling	Crude (A_{260})	Purification	
				Loaded (A_{260})	Recovered (A_{260})
A via **2.5**	$AG(^{2'5'}G)_{3'5'}G$	25.4	14.9	10	1.07
A via **2.8**[•]	$AG(^{2'5'}G)_{3'5'}G$	57	18.3	5	0.48
B	$AU(^{2'5'}G)_{3'5'}G$	54	21.6	10	1.37
C	$AC(^{2'5'}G)_{3'5'}G$	58	23.4	10	1.68

[•]**A** synthesized via **2.8** is designated as sequence **D**

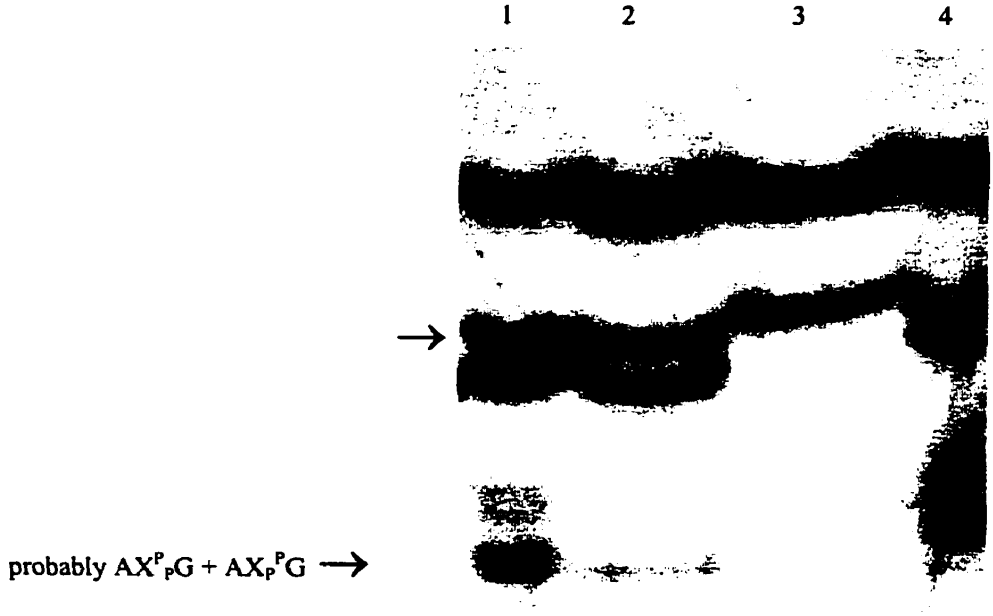


Figure 2.1: PAGE (24%, 7M urea) of crude branched tetramers. *Lane 1*: $AU(^{2'5'}G)_{3'5'}G$, **B**; *lane 2*: $AC(^{2'5'}G)_{3'5'}G$, **C**; *lane 3*: $AG(^{2'5'}G)_{3'5'}G$, **A** (via **2.5**); *lane 4:* $AG(^{2'5'}G)_{3'5'}G$, **D** (**A** via **2.8**). Arrow marks approximate mobility expected for the branched tetramers. The slowest moving band in each lane represent the uncoupled first nucleoside, rG, at the surface of CPG.

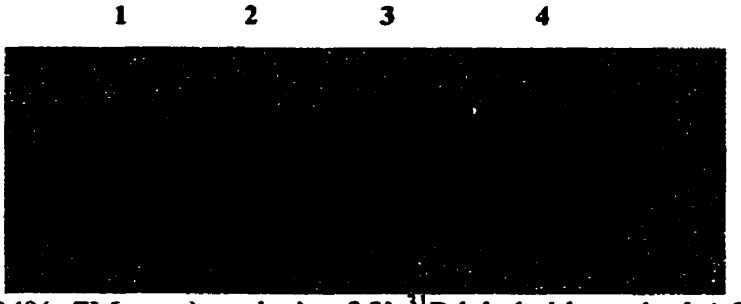


Figure 2.2: PAGE (24%, 7M urea) analysis of 5'-^{31}P labeled branched AG(G)G tetramer, A. *Lanes 1, 2*: putative 5'-^{31}P-AG($^{2'5'}$G)$_{3'5'}$G without (lane 1) and with the yeast debranching enzyme (lane 2); *lanes 3, 4*: linear ^{32}P-AGG trimer without (lane 3) and with the yeast debranching enzyme (lane 4).

2.3.1 Synthesis of AG($^{2'5'}$G)$_{3'5'}$G via O^6 protected guanosine monomer

Guanine differs from other bases by possessing potential reactive sites at the O^6-, N^7-, and N^1- positions. Substitution[201], sulfonylation[202-204], silylation[203], and phosphitylation[205,206] reactions have been described. As shown by Pon *et al.*[205,206], phosphitylation of guanine O^6 under the phosphite triester synthesis conditions would lead to depletion of the phosphoramidite molecules available for coupling with the growing chain. This may not present a problem since usually a 25 - 30 fold excess of phosphoramidite to 5'-hydroxyl of the growing chain is employed, but the problems with guanine side reactions would be most severe in situations where diluted phosphoramidite solutions are required. Since low concentration bis(phosphoramidite) solutions are used for branching reactions *via* the convergent methodology, any side reactions would significantly effect coupling with the support bound oligomer.

Damha and coworkers[152] reported the necessity of guanosine O^6-protection in preparing branched oligonucleotides bearing guanosine residues at the 2'- and 3'-positions of the branch core. Others have also reported the necessity of protecting the guanosine O^6-position in analogous reactions leading to the synthesis of branch oligonucleotides.[189,195]

The susceptibility of the guanine O^6 to undergo reactions with a variety of reagents led to the realization for the necessity of O^6-protection, and a number of protecting groups have been proposed for protecting the guanine residue. The *p*-nitrophenylethyl[204] group is the most predominantly used.

In order to overcome the problems associated with O^6-phosphitylation of the guanine residue, the O^6-*p*-nitrophenylethyl protected guanosine bis(phosphoramidite), **2.8** was prepared by the route shown in scheme **2.3**. The N^2-benzoyl-O^6-*p*-nitrophenylethyl protected guanosine monomer, **2.4**, provided by Dr. Damha, was phosphitylated with N,N-diisopropylethylamine-2-cyanoethylphosphonamidic chloride under standard conditions leading to the desired O^6-*p*-nitrophenylethyl protected nucleoside **2.8**.

The branched tetramer $AG(^{2'5'}G)_{3'5'}G$, **D**, was then resynthesized, using **2.8** as the branch-introduction synthon. From quantitation of released trityl cation, branch-introduction was seen to be reasonably good (45% vs 50% theoretical). After the usual ammonia/ethanol treatment to cleave the nascent oligomer from the solid-support and the benzoyl groups, removal of the O^6-NPE and 3'-O-silyl groups was effected with a 1M solution of TBAF/THF. The $NEt_3.3HF$ reagent, commonly used to remove silyl groups from oligoribonucleotides, has been shown to be ineffective in deblocking the O^6-NPE group (A. Noronha, unpublished results).

A purified sample of **D** was subjected to enzymatic hydrolysis and the resulting products identified by reversed-phase HPLC. Enzymatic hydrolysis with nuclease P1 and alkaline phosphates yielded inosine and trinucleotide diphosphate $G(^{2'5}G)_{3'5'}G$ in the expected 1:1 ratio (data not shown). In addition, treatment of

AG($^{2'5'}$G)$_{3'5'}$G with a mixture of snake venom phosphodiesterase and alkaline phosphatase afforded inosine and guanosine in the predicted 1:3 ratio, figure **2.3**. Inosine results from the presence of adenosine deaminase, which is present as a contaminant in the enzyme preparation, leading to deamination of adenosine into inosine.

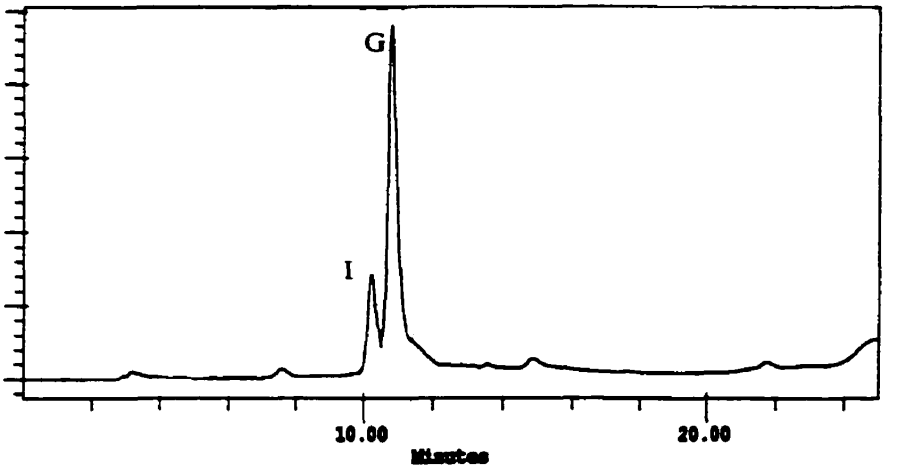


Figure 2.3: HPLC analysis of enzymatic digest for compound AG(G)G. Digestion with a combination of snake venom phosphodiesterase (SVPDE) and alkaline phosphatase (AP). Analysis was carried out at 254 nm using a reversed-phase Whatman Partisil ODS-2 (10 µm, 4.6 x 250 mm); mobile phase solvent A: 20 mM KH_2PO_4 (pH 5.5); solvent B, methanol, gradient 0 – 50% solvent B in 25 minutes, flow rate 1.5 ml/min, 30 °C.

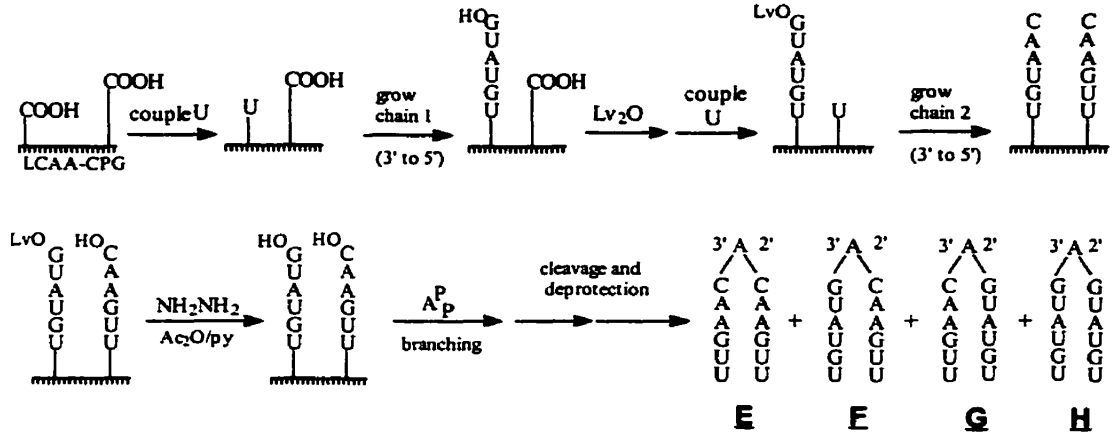
2.4 REGIOSPECIFIC SOLID-PHASE SYNTHESIS OF BRANCHED NUCLEIC ACIDS

2.4.1 Previous work

The solid-phase method of branched nucleic acids, as initially introduced by Damha[154], although elegant and ingenious, is limited in that similar sequences are present at the 2'- and 3'- positions of the branchpoint. To overcome this problem, the Damha research group has been actively looking at developing new methods for the regiospecific synthesis of branched oligonucleotides.

Recently K. Ganeshan,[207,208] a previous member of our research group adopted the original approach of Damha[154] for the synthesis of bRNA of mixed branch composition. The key feature of this methodology was the synthesis of two different linear oligonucleotides on the surface of controlled-pore glass (LCAA-CPG). Thus coupling of these two sequences with a suitably protected adenosine-2',3'-bis(phosphoramidite) (*e.g.*, **1.1**) afforded all four possible species: two "symmetric" V-like molecules , (**E**, and **H**) and two "asymmetric" V-like isomers, (**F**, and **G**), in more or less equal amounts, scheme **2.5**. However separation of the four isomers proved to be very difficult (see below).

Scheme 2.5



Analysis by gel electrophoresis showed the presence of one major band with the expected mobility. The band was purified[207] by gel electrophoresis and extracted with water (37 °C, 16 h). The purified product was then subjected to hydrolysis by enzymatic digestions with nuclease P1/alkaline phosphatase (AP). Analysis of the enzymatically hydrolyzed sample by reversed-phase HPLC indicated the presence of four constituent branched trinucleotides, i.e. $A(^{2'5'}G)_{3'5'}C$, $A(^{2'5'}G)_{3'5'}G$, $A(^{2'5'}C)_{3'5'}G$, and $A(^{2'5'}C)_{3'5'}C$ which were identified by comparison with previously characterized branched trimer samples.

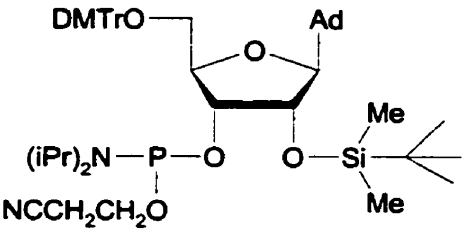
Similarity in the size and structure of the related branched species, **E**, **F**, **G**, and **H** prevented their separation from the mixture. All attempts to separate these compounds by reversed-phase and ion-exchange HPLC or polyacrylamide gel electrophoresis were unsuccessful. However, the four closely related branched molecules could be resolved by capillary electrophoresis.[207,209] As only minuscule amounts of materials are injected into the capillary, fraction collection is usually impractical. However it is clear that the resolving power of capillary electrophoresis makes it an excellent analytical tool in the analysis of synthetic branched oligonucleotides.

2.5 NEW APPROACH TO THE REGIOSPECIFIC SYNTHESIS OF BRANCHED NUCLEIC ACIDS

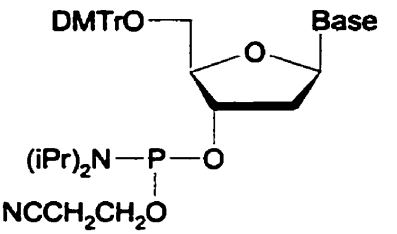
2.5.1 Synthesis of Asymmetric* Branched Deoxyoligonucleotides

The failure to isolate **E**, **F**, **G**, and **H** from the mixture required investigation of alternative methods for the regiospecific synthesis of branched nucleic acids. The overall strategy that we considered is illustrated in scheme **2.6**, and the procedure can be demonstrated by the synthesis of **I**. This oligonucleotide has all conserved nucleobases and mimics the natural occurring sequence found in *S. cerevisiae* lariat introns, table **2.3**. The readily available N^6-benzoyl-5'-dimethoxytrityl-2'-O-*t*-butyldimethylsilyladenosine 3'-O-phosphoramidite (**2.9**) is the key building block since it serves as the branchpoint allows sequential extension in all three directions. The other synthons required are 2'-deoxynucleoside-3'-O- or 5'-O-cyanoethylphosphoramidites **2.10** and **2.11** respectively, which are easily synthesized or can be obtained commercially.

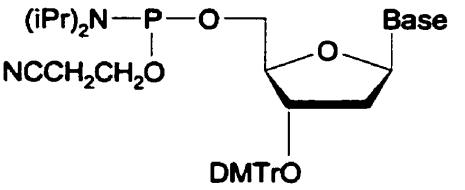
*asymmetric here means that the strands at the 2'- and 3'- positions of the branchpoint have different base compositions.



2.9



2.10



2.11

Long-chain alkylamine controlled pore glass served as the solid support and was derivatized with the 3'-terminal nucleoside as previously described.[210] Thus employing the phosphoramidite method for oligonucleotide synthesis, the synthesis began by assembly of the linear dodecanucleotide 5'-d(TACTA)-***rA***-d(CAAGTT)-3' in the normal 3'-to-5' direction, (scheme **2.6**). Following acetylation of the terminal 5'-hydroxyl group, the support was treated with triethylamine/acetonitrile solution (4:6 v/v, 1.5 h, r.t.) to effect the removal of all cyanoethyl phosphate protecting groups thus providing an oligomer with a intrinsically more stable *phosphodiester* backbone. As discussed in section 2.2, the stepwise introduction of 2',5'- and 3',5'-phosphodiester linkages *via* a *phosphotriester* intermediate is not viable due to the phosphoryl migration/cleavage under neutral, acidic or basic conditions.[152,182,184-187] Therefore conversion of the *phosphotriester* linkage into the more stable phosphodiester linkages is a key feature of this synthetic methodology.

Scheme 2.6

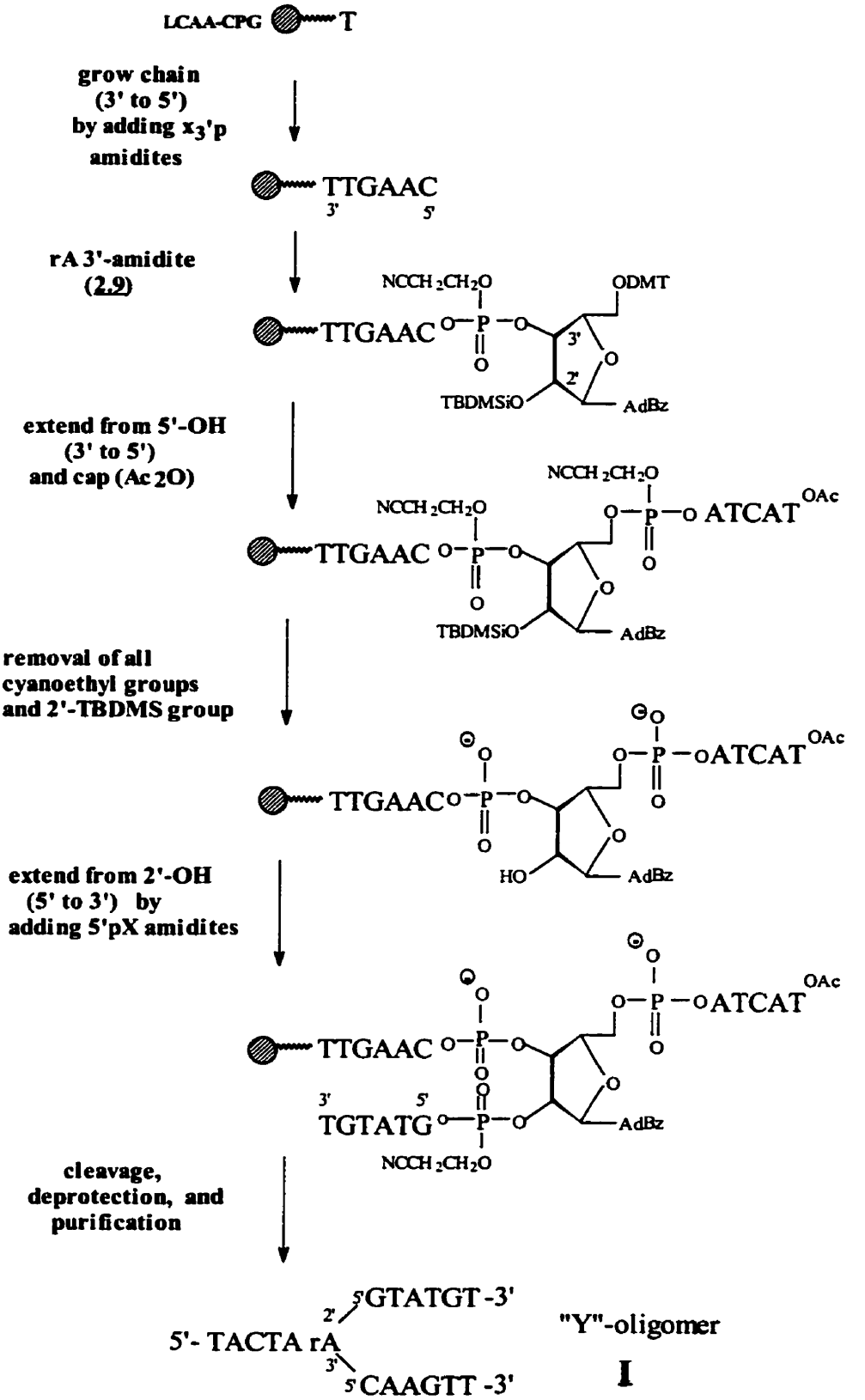

LCAA-CPG
T
grow chain (3' to 5') by adding x3'p amidites
TTGAAC
3'
5'
rA 3'-amidite (2.9)
NCCH2CH2O
ODMT
TTGAAC
TBDMSiO
AdBz
extend from 5'-OH (3' to 5') and cap (Ac2O)
ATCAT
OAc
removal of all cyanoethyl groups and 2'-TBDMS group
HO
extend from 2'-OH (5' to 3') by adding 5'pX amidites
TGTATG
cleavage, deprotection, and purification
5'- TACTA rA
5'GTATGT-3'
5'CAAGTT-3'
"Y"-oligomer
I

Table 2.3: Branched sequences synthesized using the regiospecific method outlined in scheme **2.6**

Sequence	Branched Oligonucleotide
I	$^{5'\text{-HO}}$d(TAC TA) rA$^{2'\text{-}5'}$d(GTA TGT)$^{\text{-}3'}$ $_{3'\text{-}5'}$d(CAA GTT)$^{\text{-}3'}$
J	$^{5'\text{-HO}}$d(TAC TA) rA$^{2'\text{-}5'}$d(I TA TGT)$^{\text{-}3'}$ $_{3'\text{-}5'}$d(CAA GTT)$^{\text{-}3'}$
K	$^{5'\text{-HO}}$rA$^{2'\text{-}5'}$d(TTT TTT TTT T)$^{\text{-}3'}$ $_{3'\text{-}5'}$d(TTT TTT TTT T)$^{\text{-}3'}$
L	$^{5'\text{-HO}}$rA$^{2'\text{-}5'}$d(T$^{3'\text{-}3'}$T$^{5'\text{-}3'}$T TTT TTT T)$^{\text{-}5'}$ $_{3'\text{-}5'}$d(TTT TTT TTT T)$^{\text{-}3'}$
M	$^{5'\text{-HO}}$rA$^{2'\text{-}3'}$d(T$^{5'\text{-}3'}$TT TTT TTT T)$^{\text{-}5'}$ $_{3'\text{-}5'}$d(TTT TTT TTT T)$^{\text{-}3'}$
N	$^{5'\text{-HO}}$rA$^{2'\text{-}3'}$d(A$^{5'\text{-}3'}$AA AAA AAA A)$^{\text{-}5'}$ $_{3'\text{-}5'}$d(TTT TTT TTT T)$^{\text{-}3'}$
O	$^{5'\text{-HO}}$rA$^{2'\text{-}3'}$d(C$^{5'\text{-}3'}$AA AAA AAA A)$^{\text{-}5'}$ $_{3'\text{-}5'}$d(CTT TTT TTT T)$^{\text{-}3'}$
P	$^{5'\text{-HO}}$rA$^{2'\text{-}3'}$d(C$^{5'\text{-}3'}$CA AAA AAA A)$^{\text{-}5'}$ $_{3'\text{-}5'}$d(CCT TTT TTT T)$^{\text{-}3'}$
Q	$^{5'\text{-HO}}$rA$^{2'\text{-}5'}$d(CC$^{3'\text{-}3'}$A$^{5'\text{-}3'}$AA AAA AA)$^{\text{-}5'}$ $_{3'\text{-}5'}$d(CCT TTT TTT T)$^{\text{-}3'}$

An important feature was the use of fluoride ions to effect the removal of the 2'-O-TBDMS without "dissolving" the solid-support (SiO_2) and/or detaching the oligomer from its surface. Model experiments with DMT-dT-(LCAA-CPG) showed that extended treatment with fluoride leads to significant cleavage of the nucleoside from the solid support. Thus it is imperative that desilylation is carried out for no longer than 10 minutes and that the solid support be thoroughly washed with THF and acetonitrile immediately after desilylation.

Removal of the silyl group was carried out by first washing the solid-support attached nascent linear oligomer with THF (30 ml), followed by manually pushing a solution of 1M tetrabutylammonium fluoride/THF (10 min, r.t.). Immediately after the fluoride solution had been pushed through the column, the support was washed with THF (30 ml) and acetonitrile (30 ml). The solid-support was then dried by flushing dry argon through the column before proceeding.

At this point, the 2'-chain (5'-GTA TGT-3') was synthesized in the 5' to 3' direction using commercially available 5'-phosphoramidite derivatives. To force branching at the sterically hindered 2'-hydroxyl group both the concentration and the coupling time of the first 5'-amidite (dG) were tripled to 0.3M and 30 min, respectively. Under these conditions coupling proceeded with 91% efficiency. Lower phosphoramidite concentrations (0.10 M) resulted in significantly lower coupling yields (ca. 60%, e.g., J, table **2.4**). Synthesis under standard conditions was then continued until the desired "Y-mer" I was assembled.

The successful completion of I was followed by the synthesis of eight other V- and Y-shaped oligonucleotides J – Q, table **2.3**. Sequence J mimics a lariat structure in which the conserved 2'-guanine base has been replaced by hypoxanthine. Apart from the use of deoxyinosine as the 2'-residue, the other major difference in the preparation of I and that of J, was the use of adenosine-2'-O-cyanoethylphosphoramidite instead of the regioisomeric 3'-O-cyanoethylphosphoramidite. Thus 5'-d(TACTA)-*rA*$^{2',5'}$d(ITATGT)-3' was first assembled in the conventional 3' to 5' direction, and capped (Ac_2O) at the 5'-hydroxyl. This was followed by removal of the cyanoethyl and 3'-silyl groups and assembly of the 3'-branch, namely d(CAAGTT). All other aspects of branch

assembly, deprotection procedures, and handling remained invariant. Under the conditions used (0.5 M tetrazole, r.t., 30 min), the branch rA-3'-amidite isomer coupled (95%) with greater efficiency than the 2'-amidite isomer (61%). However, if 5-ethylthio-1H-tetrazole is used instead of tetrazole as the acid catalyst, both 2' and 3'-amidite couple with similar (97-99%) efficiency.[211] For the preparation of sequence **K**, both rA-3'-O-*methyl* and *cyanoethyl* protected phosphoramidites were used as the branching synthon, table **2.4**. In the case of methyl phosphate protection, a NEt_3/dioxane/thiophenol (10 ml, 2:2:1 v/v/v) solution, instead of the NEt_3/CH_3CN, was used to cleave the methyl phosphate protecting group at the branchpoint rA prior to 2'-desilylation and 2'-chain assembly. Presumably the thiophenoxide step also removes the cyanoethyl groups attached to the deoxynucleotide residues. Coupling yields at various stages of these syntheses, as determined by the quantitation of released trityl cation, as well as isolated yields of oligomers are given in Table **2.4**.

The branched oligonucleotides were removed from the support and deblocked by the standard treatment with 15M aqueous ammonia/ethanol 4:1 v/v (r.t., 48 h). Removal of the ammoniacal solution furnished the crude products. Initial characterization and purification of oligomers was accomplished by electrophoresis by comparison of mobilities to linear oligomers and to authentic samples of branched oligonucleotides. Figure **2.4b** shows a polyacrylamide gel electrophoretogram of the crude synthesis mixture of **I**.

Table 2.4: Conditions and yields of branched oligonucleotide synthesis

Sequence	Scale	% Coupling Yield[a]			Crude[b]	Purification (A_{260} U)[c]	
	(μmol)	rA[d]	2'5'pX	next	A_{260} U	loaded	Recovered
I	1.0	95	91	99	87	25	7.4
J[e]	0.6	61[e]	60[f]	78	29	20	3.1
K	1.0	100	97	101	131	25	2.8
K	1.0	97[g]	99	111	64	25	6.2
K	1.0	87	115	86	45	20	2.1
M	1.0	97	89	93	85	25	6.0
N	1.0	98	81	83	43	25	2.6
O	1.0	89	91	92	123	20	3.1
P	0.1	97	124	102	10	8	1.4
Q	0.1	97	111	135	9	7	1.9

[a]Coupling yield of the 2',5'-pX residue was calculated relative to coupling of branchpoint A. [b]Total oligomer recovered after cleavage and deprotection. [c]PAGE followed by desalting with C_{18} SEP-PAK® cartridges. Sequences P and Q were purified by the "trityl-on" reversed-phase procedure (OPC purification matrix). [d]Unless otherwise indicated, rA^{Bz}-5'-O-DMT-2'-O-silyl-3'-O-cyanoethylphosphoramidite was used as the branching synthon. [e]rA^{Bz}-5'-O-DMT-3'-O-silyl-2'-O-cyanoethylphosphoramidite. [f]Lower coupling reflects the inadvertent use of lower concentration (0.1M) of the phosphoramidite solution after the branchpoint. [g]rA^{Bz}-5'-O-DMT-2'-O-silyl-3'-O-methylphosphoramidite used.

The branched oligonucleotides were removed from the support and deblocked by the standard treatment with 15M aqueous ammonia/ethanol 4:1 v/v (r.t., 48 h). Removal of the ammoniacal solution furnished the crude products. Initial characterization and purification of oligomers was accomplished by electrophoresis by comparison of mobilities to linear oligomers and to authentic samples of branched oligonucleotides. Figure **2.4b** shows a polyacrylamide gel electrophoretogram of the crude synthesis mixture of I.

Generally, the mobility of oligonucleotides on a polyacrylamide gel subjected to an electric field depends on the net charge and size of the oligomers, and single nucleotide resolution based on size is usually obtained. The pattern of bands revealed in the PAGE analysis arises as a result of incomplete coupling reactions and, in conjunction with quantitation of the released trityl cation, is useful in diagnosing problems arising during the synthesis. Thus PAGE analysis served as

a preliminary step in the characterization of branched oligonucleotides. Analytical polyacrylamide gel electrophoretic analysis of the crude branched octadecamer oligomer **I** indicated that it was present essentially as a single species, and its mobility was comparable to that of a "symmetric" branched oligonucleotide of similar size but different sequence, i.e. TACTAA($^{2'5'}$GTATGT)$_{3'5'}$GTATGT, figure **2.4b**.

With the exception of **P** and **Q**, the oligonucleotides were purified by preparative gel electrophoresis and the slowest moving band excised under UV shadow, extracted with water at 37 °C overnight, and desalted by reversed phase chromatography (Sep-Pak cartridges). Isolated yields of crude and purified oligonucleotides are reported in Table **2.4**.

Oligonucleotide purification cartridge (OPC) protocol is a popular method for purification of linear oligonucleotides having a trityl group at the 5' hydroxyl. Although OPC purification is entirely compatible with branch oligonucleotide purification it is not usually employed during purification of bNAs from convergent synthesis. Problems arise due to elongation after incomplete branching leading to oligonucleotides other than the full length product containing a 5'-trityl. Thus purification by OPC leads to a mixture of products requiring additional purification steps. Purification by OPC also leads to the loss of any diagnostic information that may be gained from the non-trityl containing oligomers, which are lost during OPC purification. This loss of diagnostic information about the syntheses could easily be overcome by PAGE analysis of the syntheses before proceeding with the OPC purification step.

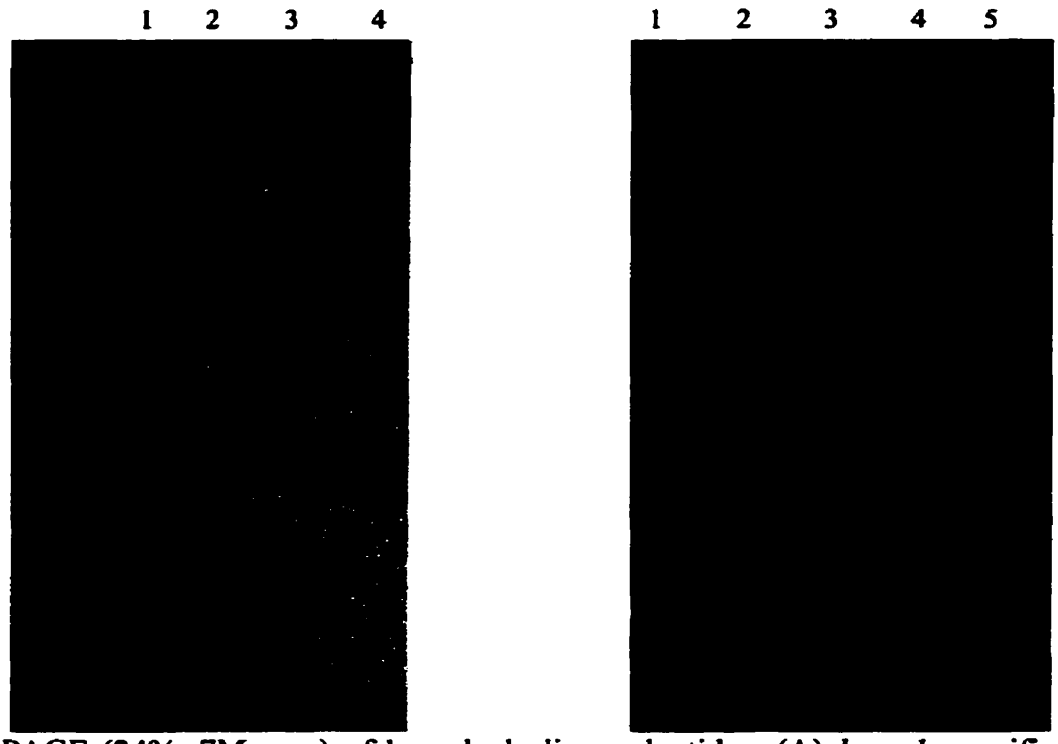


Figure 2.4: PAGE (24%, 7M urea) of branched oligonucleotides. (A) *lane 1*: purified oligomer **K** prepared the method presented in this study; *lane 2*: authentic sample of **K** prepared by a convergent method; *lane 3*: crude sample of **K** prepared by the presented method; *lane 4*: marker dyes bromophenol blue (fast) and xylene cyanol (slow). (B) *lanes 1, 5*: marker dyes; *lane 2*: branched 18-mer TACTAA($^{2'5'}$GTATGT)$_{3'5'}$GTATGT; *lane 3*: branched 18-mer **I**; *lane 4*:linear 12-mer TACTA**rA**$^{2'5'}$GTATGT.

With the divergent synthetic approach outlined here, no molecule other than the full length product is expected to contain the 5'-trityl, and as such is completely amenable to purification by the OPC-protocol. After characterization of the synthesis by PAGE, purification of branched oligomers **P** and **Q** by OPC was seen to give results similar to those obtained from preparative gel electrophoresis, table **2.4**. Therefore the "OPC protocol" is directly applicable to the purification of branched oligonucleotides synthesized with the trityl ON option using this new method. This considerably reduces the time needed for the purification of branched oligonucleotides.

In order to verify sample purity, branched structure, and nucleotide composition, a small sample of **I** was subjected to enzymatic hydrolysis and the resulting products analyzed by reversed-phase HPLC. Hydrolysis of the purified **I** with a mixture of nuclease P1 (NP1) and alkaline phosphatase (AP) generated the branched trinucleoside diphosphate $A(^{2'5}dG)_{3'5'}dC$ and other constituent nucleosides, and were unambiguously identified. By integrating the peaks corresponding to the products obtained, expected ratios were obtained. In addition, total enzymatic hydrolysis of **I** with snake venom phosphodiesterase (SVPDE) and alkaline phosphatase (AP) afforded the various nucleosides in the expected ratios (data not shown).

The nucleotide composition of **I** was further confirmed by subjecting the purified sample to Matrix Assisted Laser Desorption Ionization Time of Flight (MALDI-TOF) mass spectroscopy. Using an ammonium citrate matrix, relative molecular weight of **I** was found to be 5525.0 (calculated 5524.9), thus further confirming the nucleotide composition and chain length of **I**.

Analysis of **K** by analytical gel and capillary electrophoresis also confirmed the presence of a single species, figure **2.5**. Further PAGE and CE analysis proved that **K** produced by this new approach was identical to that obtained *via* the convergent approach. Equivalence of **K** from two synthetic methodologies was further confirmed by the indistinguishability of thermal dissociation of triplexes formed between **K** and dA_{10}, (see chapter 3, section 3.2).

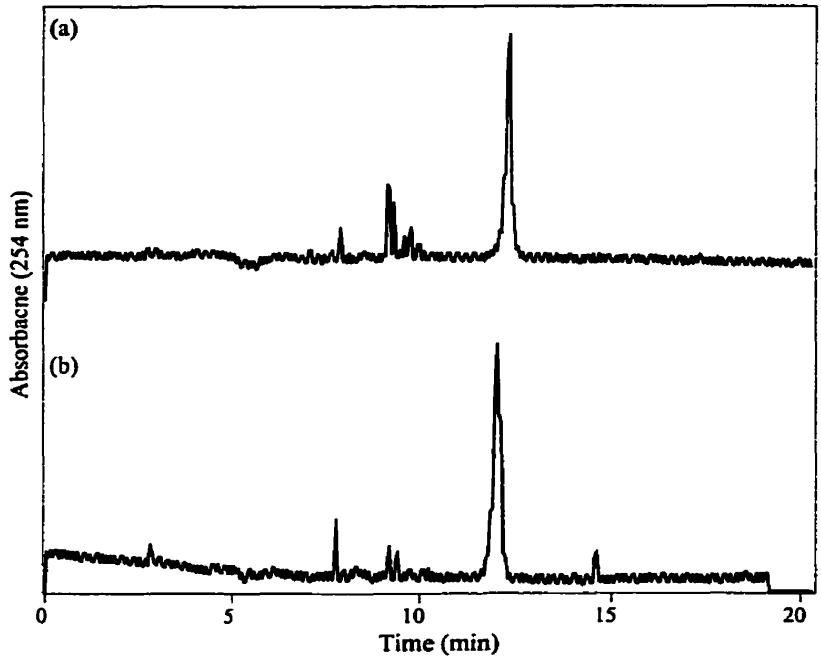


Figure 2.5: Capillary gel electrophoresis chromatograms of branched oligomer **K**. (A) crude synthetic mixture obtained from the convergent method. (B) crude synthetic mixture obtained by the regiospecific method presented here.

2.5.2 Synthesis of Asymmetric Branched Tetraribonucleotides

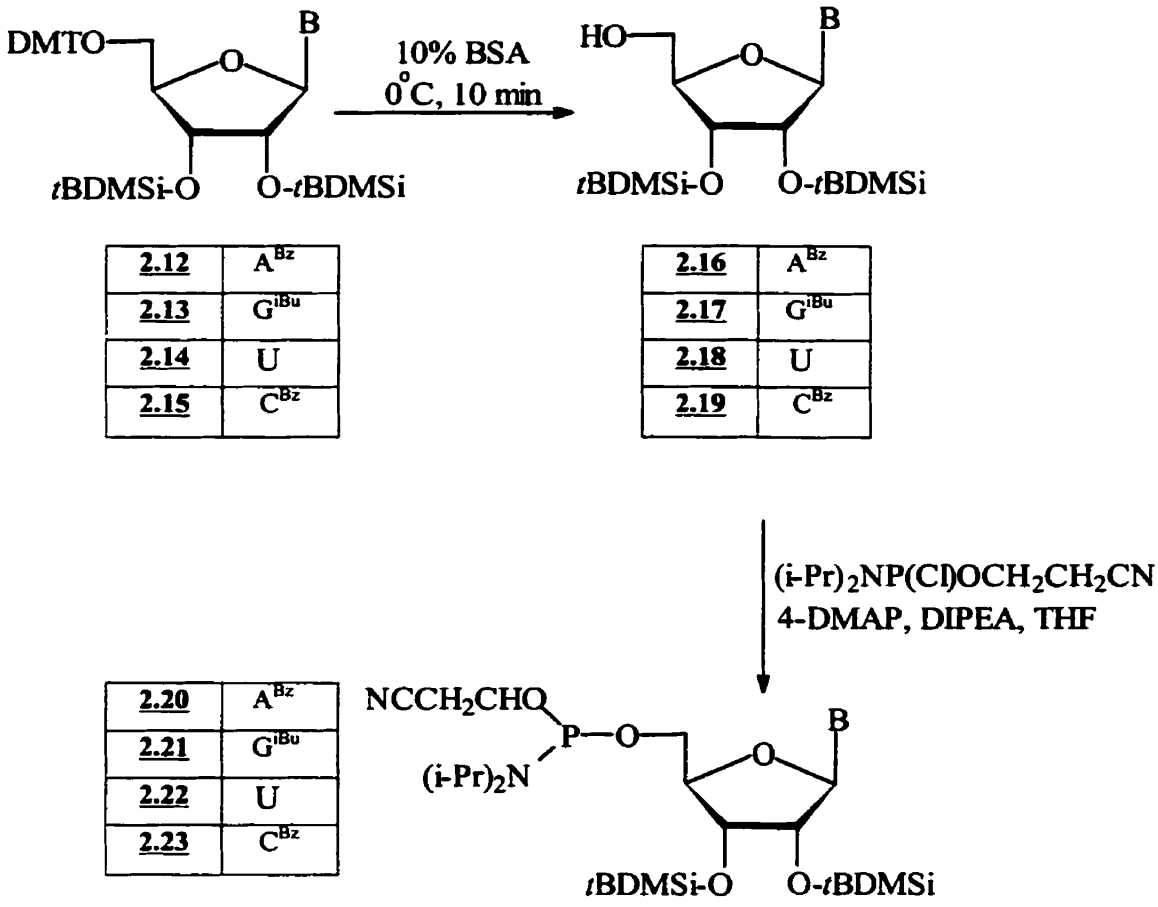
With slight modifications, the approach used for the regiospecific synthesis of branched DNA (previous section), was used for the synthesis of asymmetric branched tetraribonucleotides, schemes **2.8** and **2.9**. These molecules will be used to probe the substrate specificity and structural requirements of the yeast debranching enzyme (section 4.2).

The necessity of removing the silyl protecting group from the branchpoint during the synthesis excluded the possibility of using CPG solid-support derivatized with usual 2'- or 3'- *tert*-butyldimethylsilyl protected monomers. Thus appropriately base protected, 5'-tritylated nucleosides were directly coupled to LCAA-CPG solid support[210], and the remaining free 2'- or 3'- hydroxyl group was then acetylated along with any underivatized sites which could potentially react and

generate by-products during synthesis. The acetylation step was conveniently carried out on the DNA synthesizer just prior to the actual oligonucleotide synthesis. Using this method loadings in the 28 - 35 μmol/g of solid support were obtained. On storage at room temperature, no appreciable loss of the nucleosides from the solid support was seen for up to 12 months.

The preparation of monomers **2.20** – **2.23** is convenient since the fully blocked nucleoside intermediates, **2.12** – **2.15** are obtained as side products during the synthesis of 5'-DMT-2'-*t*-butyldimethylsilylated nucleosides. The trityl protecting group has long been employed in nucleoside chemistry and is easily cleaved under mild acidic conditions. Thus deblocking the 5'-trityl group (10% BSA in CH_2Cl_2) of **2.12** – **2.15** yielded the 5'-hydroxy compounds **2.16** – **2.19** in quantitative yields. These intermediates were then phosphitylated to give the desired nucleoside 5'-phosphoramidite monomers **2.20** – **2.23**, scheme **2.7**.

Scheme 2.7



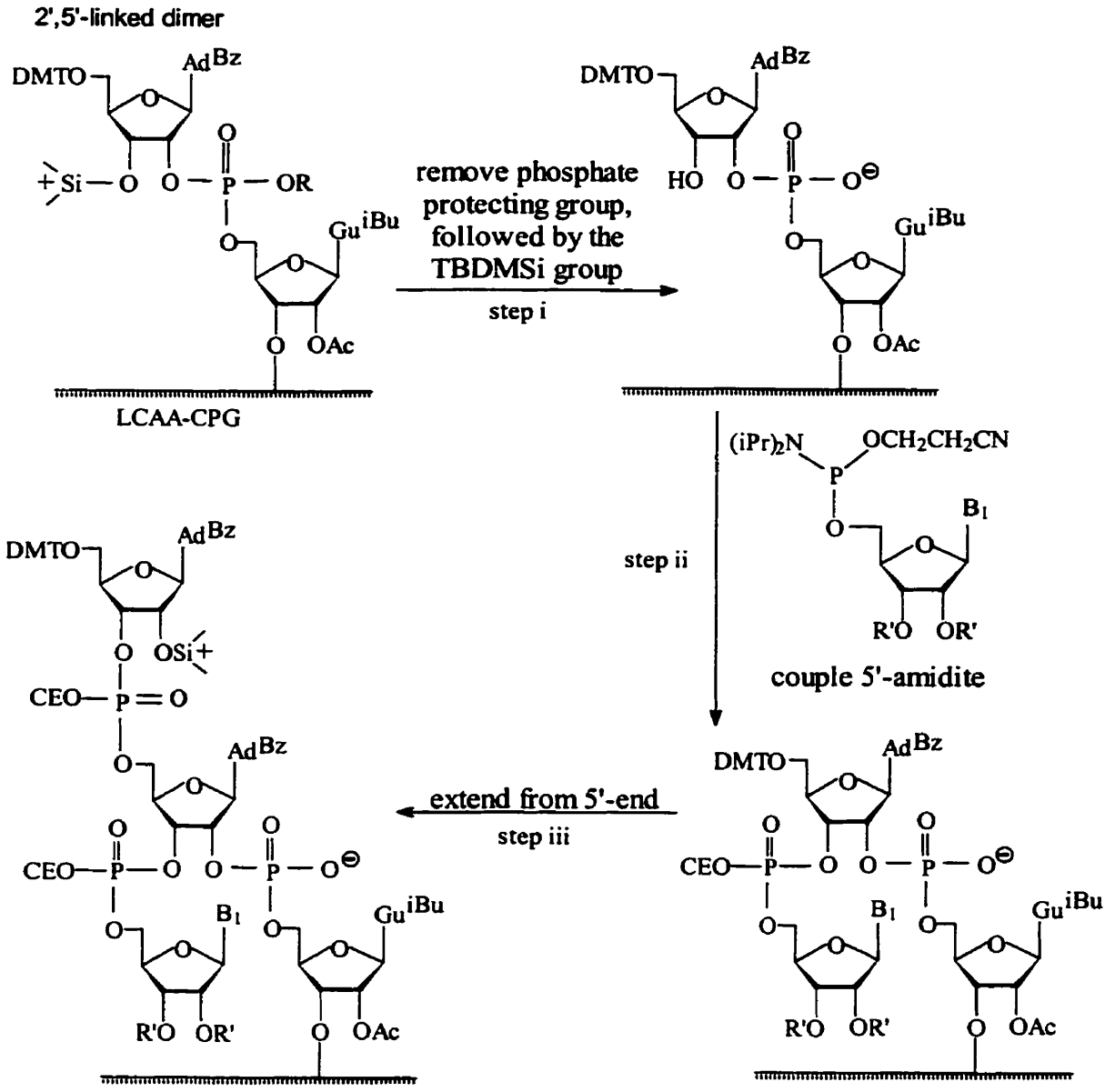
2.12	A^{Bz}
2.13	G^{iBu}
2.14	U
2.15	C^{Bz}

2.16	A^{Bz}
2.17	G^{iBu}
2.18	U
2.19	C^{Bz}

2.20	A^{Bz}
2.21	G^{iBu}
2.22	U
2.23	C^{Bz}

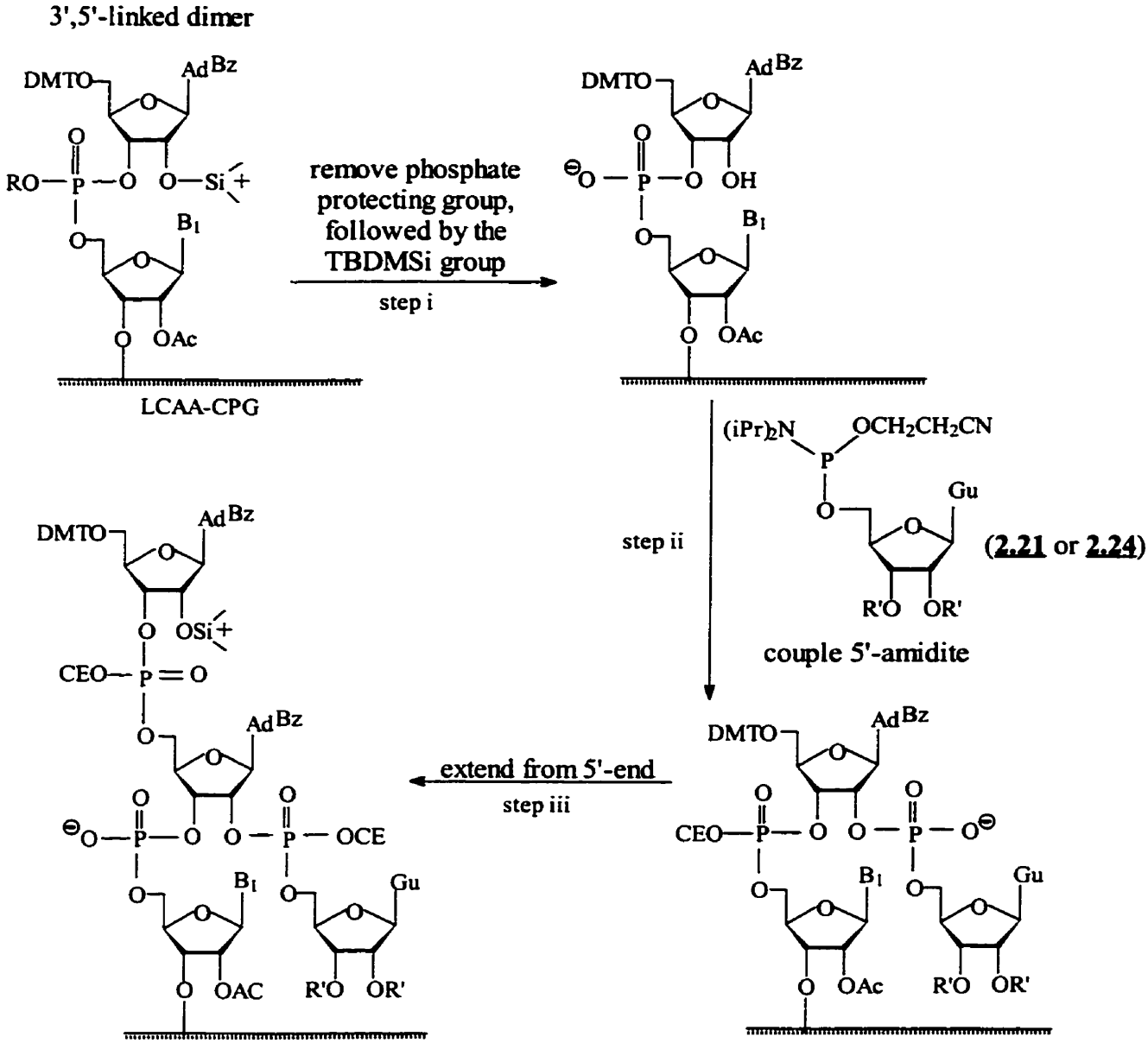
bRNA syntheses started by assembling the needed 2',5'-linked dimer (scheme **2.8**) Or 3',5'-linked dimer (scheme **2.9**) in the normal 3'-to-5' fashion. After partial deprotection to free the 3'(2') hydroxyl (step i), the standard synthesis cycle was slightly modified for coupling of the 2',3'-O-TBDMS protected ribonucleoside 5'-O-(β-cyanoethyl) phosphoramidites. Namely steps for delivery of the detritylating reagent (3% TCA) to the synthesis column were by-passed and coupling time increased from 5 to 30 min. After coupling of the 5'-phosphoramidite (step ii), normal synthesis cycle was resumed and oligomer extended at the 5'-position by one nucleotide. The known symmetric branched[153] tetramer $UA(^{2'5'}U)_{3'5'}U$, U, was first synthesized using the approach in order to establish experimental conditions.

Scheme 2.8



Oligomer	B_1	R	R'
R	Ad^{Bz}	-Me or $-CH_2CH_2CN$	TBDMS
S	Cy^{Bz}	-Me or $-CH_2CH_2CN$	TBDMS
T	Ur	-Me or $-CH_2CH_2CN$	TBDMS

Scheme 2.9



Oligomer	B_1	R	R'
R	A^{Bz}	-Me or -CH_2CH_2CN	TBDMS
S	C^{Bz}	-Me or -CH_2CH_2CN	TBDMS
T	U	-Me or -CH_2CH_2CN	TBDMS

Table 2.5: Branched tetrameric oligonucleotides synthesized using the divergent approach outlined in this study.

Sequence	Branched Tetramer	Route
R	$AA(^{2'5'}G)_{3'5'}A$	Scheme **2.8**
S	$AA(^{2'5'}G)_{3'5'}C$	Scheme **2.8/2.9**
T	$AA(^{2'5'}G)_{3'5'}U$	Scheme **2.8**
U	$UA(^{2'5'}U)_{3'5'}U$	Scheme **2.8**

A typical PAGE analysis of the crude mixture of $UA(^{2'5'}U)_{3'5'}U$, **U** (synthesized *via* scheme **2.8**) is shown in figure **2.6**. The mixture consisted of 4 bands (lane 3). The major, fastest moving band had an electrophoretic mobility identical to that of an authentic[153] sample of tetramer **U** (lane 4) and was tentatively assigned as such. In comparison with the mobility of samples of $A^{2'5'}U$ and $U^{3'5'}A^{2'5'}U$, the slowest moving bands in the crude compound **U** were assigned as A, $A^{2'5'}U$, and $U^{3'5'}A^{2'5'}U$, figure **2.6**. These linear products were the result of incomplete coupling during assembly of the branched tetraribonucleotide. Nevertheless the desired tetramer **U** was the major component of the mixture.

Of interest is the apparent reversal of oligonucleotide mobility of $A^{2'5'}U$ versus $U^{3'5'}A^{2'5'}U$. For nucleic acids mobility is usually inversely proportional to size. However for oligonucleotides of smaller sizes, less than 7-mers, this is not the case. For larger molecules (> 7-mer) mobility is based solely on mass as the charge to mass *ratio*; here *z/m* approaches 1 and thus a 10-mer for example move faster than an 11-mer. However, in the case of smaller molecules, *both* charge and mass play a role. $A^{2'5'}U$ with $z/m = 0.5$ moves more slowly, as expected, than $U^{3'5'}A^{2'5'}U$, $z/m \sim 0.66$.[212]

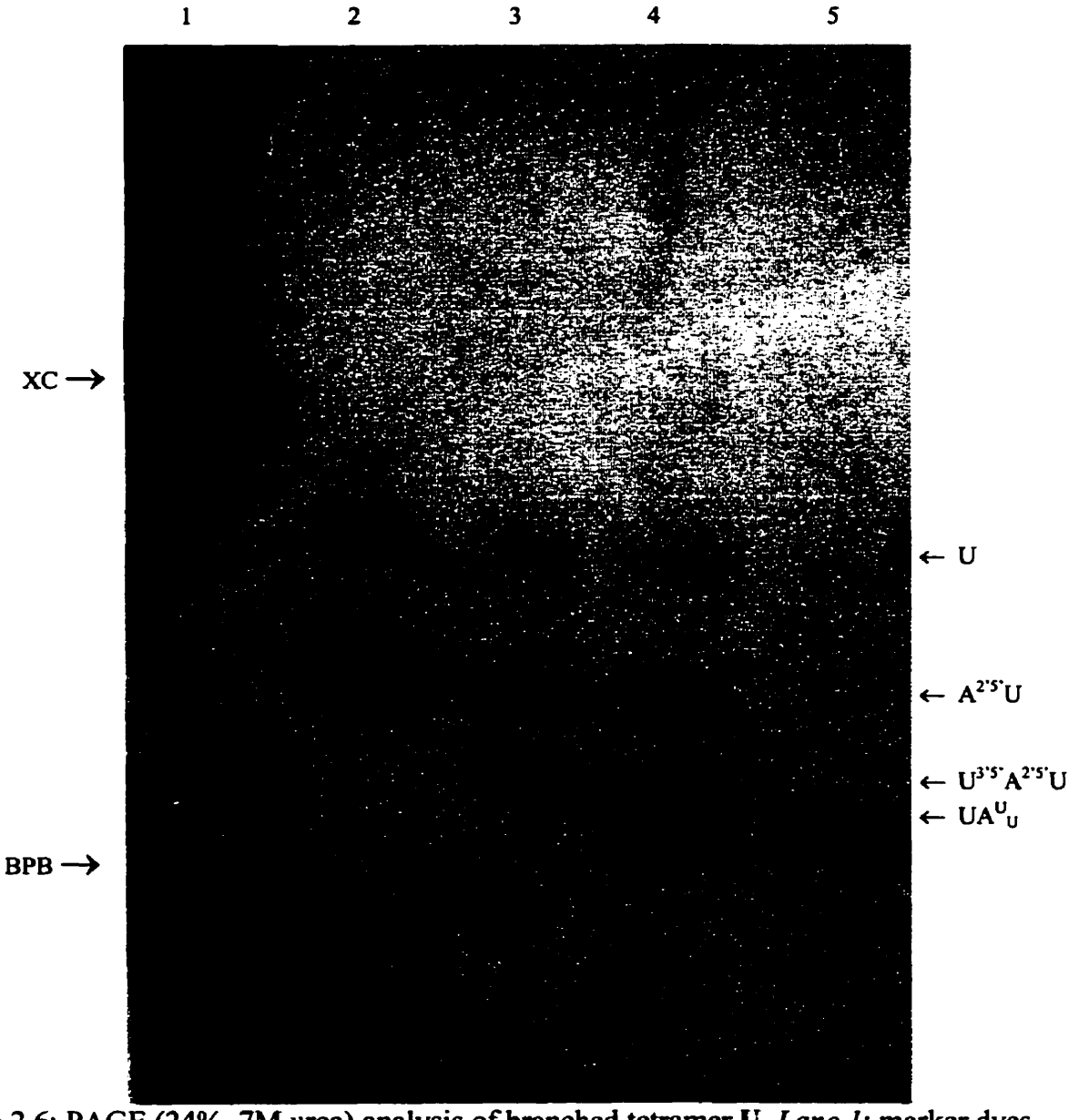


Figure 2.6: PAGE (24%, 7M urea) analysis of branched tetramer U. *Lane 1*: marker dyes, BPB (fast) and XC (slow); *lane 2*: crude of $A^{2'5'}U$; *lane 3*: crude mixture $A^{2'5'}U$ and $U^{3'5'}A^{2'5'}U$; *lane 4*: crude synthetic mixture of $UA(^{2'5'}U)_{3'5'}U$, U; *lane 5*: authentic sample of U previously prepared and characterized[153].

Figure **2.7** shows the electrophoretogram of the crude product obtained from the synthesis of $AA(^{2'5'}G)_{3'5'}C$, S (*via* scheme **2.8**) Lane 1 shows an unseparable mixture containing isomers $AA(^{2'5'}G)_{3'5'}C$ and $AA(^{2'5'}C)_{3'5'}G$ synthesized using a previously reported methodology.[207] The presence of the two isomers in the mixture was unambiguously determined using a combination of enzymatic

hydrolysis and HPLC analysis. Based on the similarity in mobility, of crude **S** and this mixture, the major band (lane 2) was excised, extracted with water and further purified by reverse-phase chromatography, (Sep-Pak cartridge).

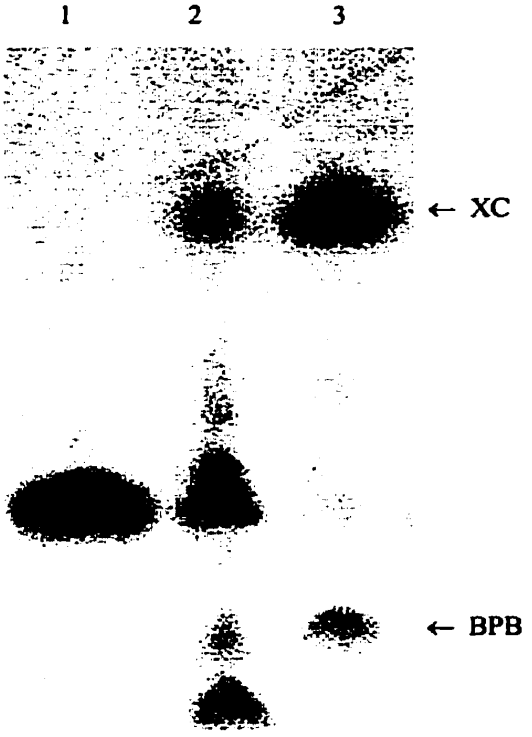


Figure 2.7: PAGE (24%, 7M urea) analysis of branched tetramer **S**. *Lane 1*: mixture of AA($^{2'5'}$G)$_{3'5'}$C and AA($^{2'5'}$C)$_{3'5'}$G, previously prepared and characterized[208]; *lane 2*: crude synthetic mixture of **S**.

Hydrolysis of the purified branched AA($^{2'5'}$G)$_{3'5'}$C tetramer, **S**, with a mixture of nuclease P1 and alkaline phosphatase generated the branched core trinucleoside diphosphate A($^{2'5'}$G)$_{3'5'}$C (R_t min) and inosine (R_t min) in the expected 1:1 ratio, figure **2.8**. The branched core trimer A($^{2'5'}$G)$_{3'5'}$C was identified by HPLC comparison with an authentic sample of the trimer prepared by the established solution-phase methods.[153] Of note, digestion of **S** did not result in the formation of A($^{2'5'}$C)$_{3'5'}$G (R_t min) thus establishing the regioselectivity of our approach. Digestion of **S** with SVP/AP, followed by HPLC analysis afforded the constituent monomers in the expected ratios (I:G:C 2:1:1). MALDI-TOF mass spectrometry

of $AA(^{2'5'}G)_{3'5'}C$ (found M: 1242.2; calculated 1245.2) was also consistent with its assigned structure.

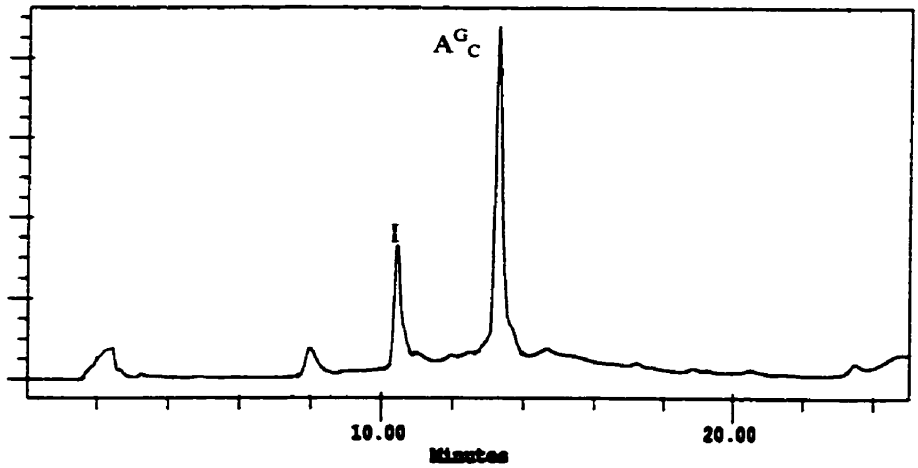


Figure 2.8: HPLC analysis of the enzymatic digestion of S (PAGE purified) with nuclease P1/alkaline phosphatase. Compounds were identified by coinjection with authentic samples of digestion products. Inosine results from the deamination of adenosine by the enzyme adenosine deaminase which is present as a contaminant in the enzyme preparation.

2.5.2.1 β-cyanoethyl vs O-methyl Phosphate Protection

Generality of our synthesis method could be further illustrated by the regiospecific synthesis of branched tetraribonucleotides via a different phosphate protecting group. At certain times it may be advantageous to use the O-methyl group over the β-cyanoethyl phosphate protecting group or vice versa. Except for the different conditions required for the deprotection of the phosphate protecting group, all other steps were similar in both cases. Figure **2.9** shows the PAGE analysis of the crude mixtures of branched RNAs, synthesized under *nonoptamized* conditions, with the O-methyl (lane 2) or the O-(β-cyanoethyl) (lane3) phosphate protecting groups. Both pathways were seen to work, however the route employing the β-cyanoethyl protecting group was seen to give the cleaner product mixture. Temsamani and coworkers[213] have shown that the first few couplings in the solid-phase synthesis of oligonucleotides are usually variable

even when trityl quantitation indicate couplings to be quantitative. Therefore the difference seen between the two syntheses may represent this variability rather than differences between phosphate protecting groups.

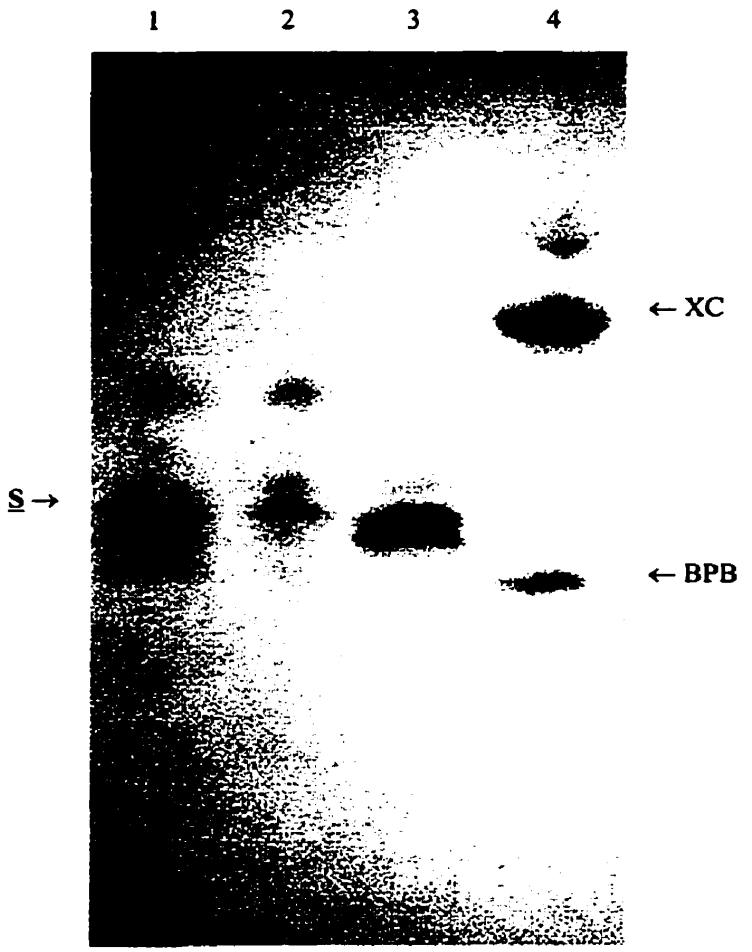


Figure 2.9: PAGE (24%, 7M urea) analysis of branched tetramer S. *Lane4*: marker dyes, BPB (fast) and XC (slow); *lane 1*: synthesis crude of S synthesized with the O-methyl phosphate protection; *lane 2*: synthetic crude of S synthesized with the O-cyanoethyl phosphate protection; *lane 4*: crude synthetic mixture of S synthesized via $A_{3'p5'}C$ (scheme **2.9**) dimer instead of $A^{2'p5'}G$ dimer (scheme **2.8**) (see section 2.5.2.2).

2.5.2.2 Synthesis From the 2',5' vs 3',5' dimers

A systematic study of the sequence specificity of yeast debranching enzyme will inevitably require analogs which may be easy to assemble from a 3', 5'-linked dimer than a 2',5'-linked dimer or vice versa. Easy adaptation of the synthesis

protocol to both situations was illustrated by synthesis of branched tetramers from $A^{2'5'}G$ dimers, section 2.5.2, or using a suitably protected guanosine-5'-phosphoramidite from the $A_{3'5'}X$, (X = A, C, U), linked dimer. On analysis by gel electrophoresis, the crude of synthesis of **S** by starting from $A_{3'5'}C$ dimer showed two major bands of approximately equal intensity, figure **2.9** lane 3. Again this represents nonoptamized conditions and this situation may change as the various factors influencing the synthesis are optimized.

2.5.2.3 Need for Guanine O^6-Protection

As discussed in sections 1.5 (chapter 1) and 2.3, side reactions of the guanine residue at the O^6-position generally present a problem during the synthesis of small asymmetric bRNAs. When branching was carried out by coupling a guanosine 5'-phosphoramidite lacking protection at the O^6-position, **2.21**, none of the expected product was seen as determined from mobility by gel electrophoresis, figure **2.10** However, on introducing branching with an O^6-protected guanosine 5'-phosphoramidite, **2.24**, the situation was seen to change dramatically. Figure **2.10** shows the gel electrophoresis of crude mixtures from the two situations. This implies that O^6-protecction of guanosine is an absolute requirement in the regiospecific solid-phase synthesis of branched RNAs.

2.6 SOLID-PHASE SYNTHESIS OF MSDNA ANALOGUES

The peculiar structure of msDNA, chapter 1; section 1.5.1, has led to a great interest in its formation and the possible biological functions it may serve. A survey of the literature on msDNA showed that no chemical synthesis of msDNA had been reported. Thus its chemical synthesis was undertaken not only to extend applicability of the branched synthetic methodology, but also to obtain msDNA analogs for possible biochemical studies.

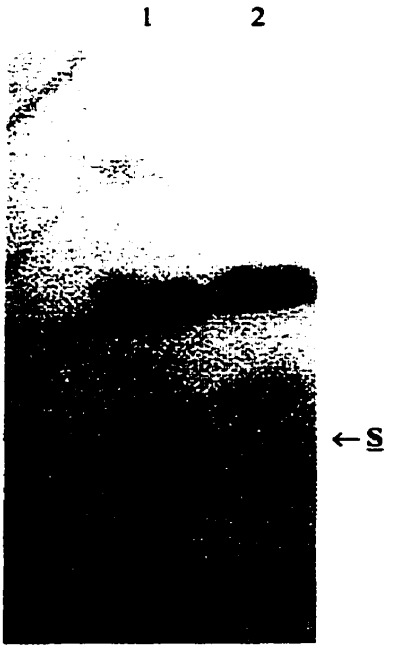
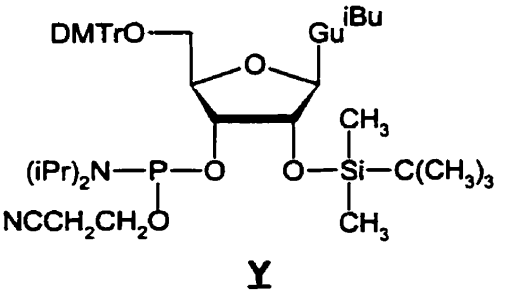
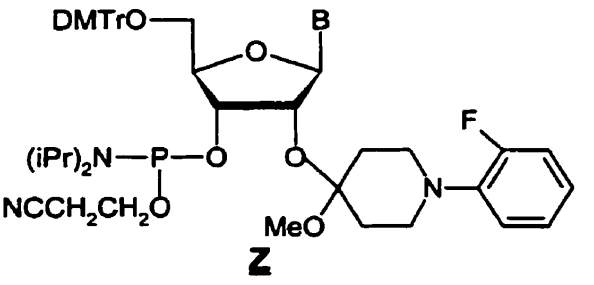


Figure 2.10: PAGE (24%, 7M urea) analysis of branched tetramer S. *Lane 1*: synthesis crude of S synthesized using the O^6-protected guanosine 5'-phosphoramidite, **2.4**; *lane 2*: synthesis of S with guanosine 5'-phosphoramidite lacking the O^6-protection, **2.1**.

msDNA is a branched DNA/RNA chimera and thus its synthesis required two mutually exclusive RNA phosphoramidite monomers Y and Z (see also scheme **2.10**).

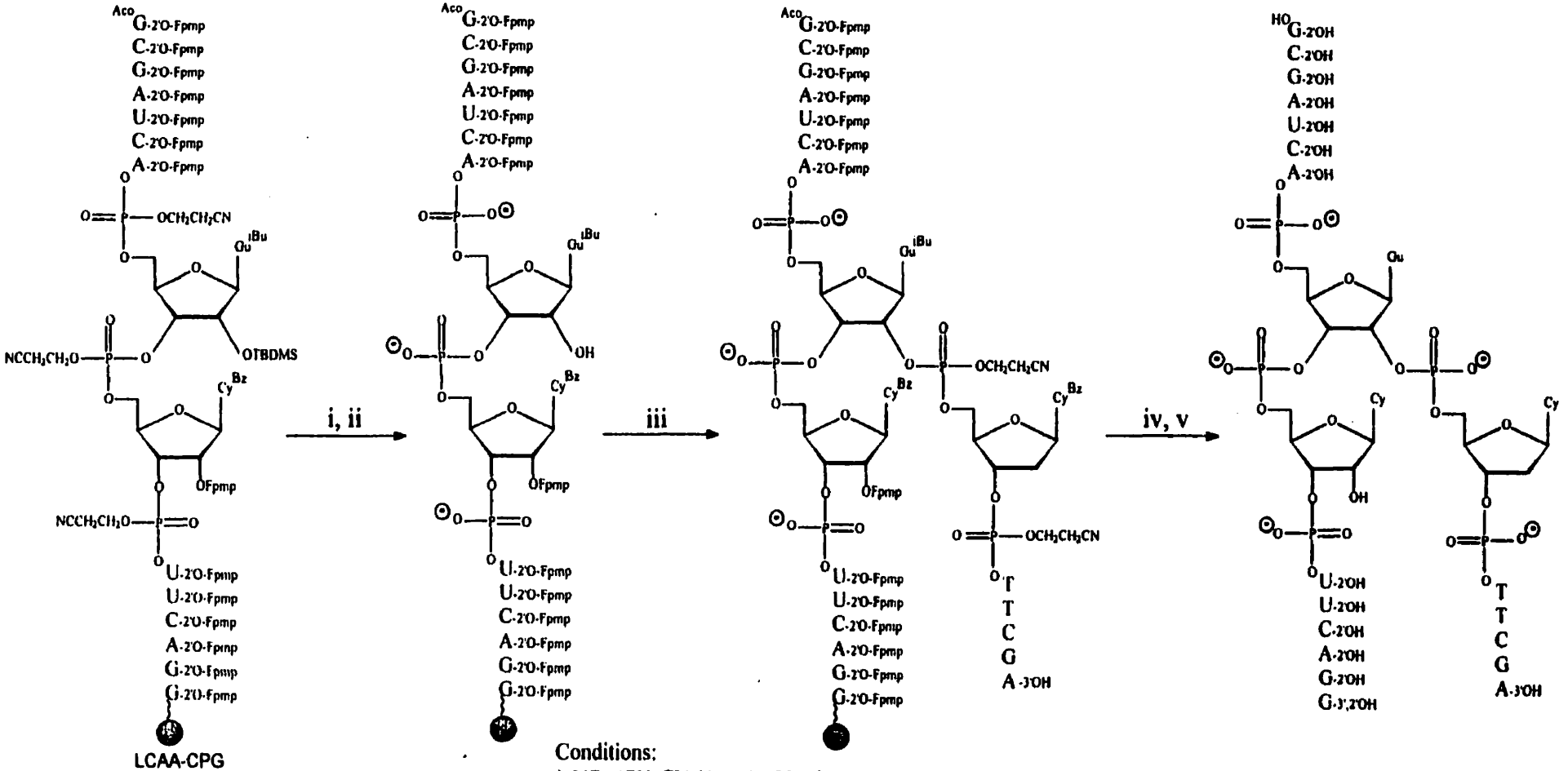


Y



Z

Scheme 2.10


i, ii
iii
iv, v
LCAA-CPG
Conditions:
i. NEt_3/CH_3CN (4:6 v/v, 90 min)
ii. 1 M TBAF/THF (10 min)
iii. extend from 2'OH using 5'pX amidites
iv. NH_3/EtOH (3:1 v/v, 48-60 hr)
v. 0.5 M Tris acetate (pH 3.25, 24 hrs)

Reese *et al.*[150] introduced the 2'-O-[1-(2-fluorophenyl)-4-methoxypiperidin-4-yl] (Fpmp) group for the protection of 2'-hydroxyls during RNA synthesis. For short oligonucleotides, the Fpmp group has been shown to be relatively stable to the conditions employed during RNA synthesis (using monomer **Z**) and easily removed under buffered acidic conditions. Thus for the msDNA, the RNA chain was assembled using a combination of 2'-O-TBDMS (branchpoint guanosine) and 2'-O-Fpmp (rest of the RNA oligomer) monomers **Y** and **Z** (scheme **2.10**).

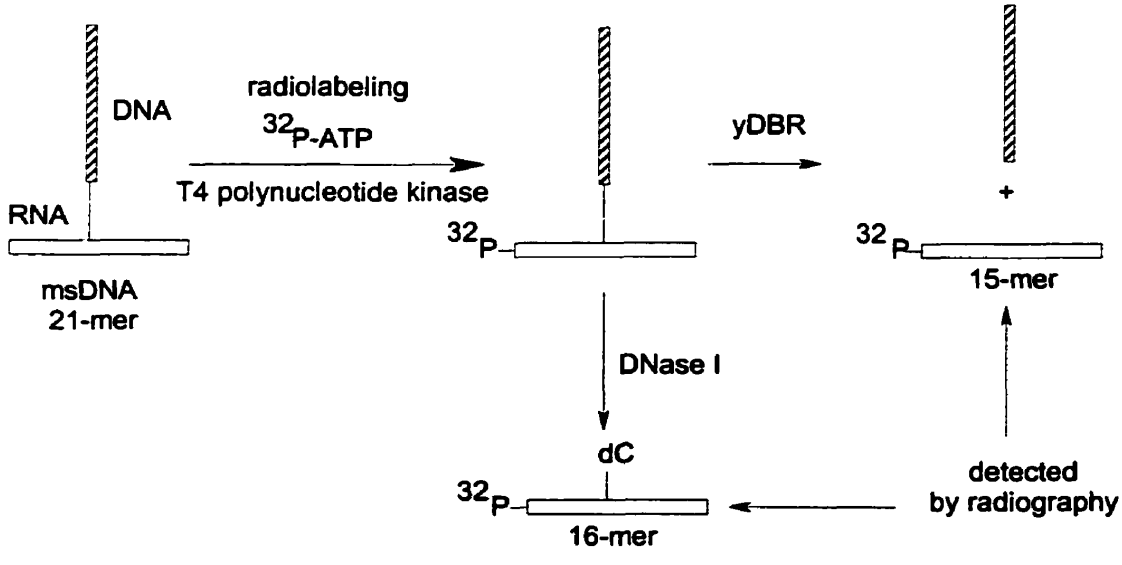
As the Fpmp group becomes unstable on repeated exposure to acidic conditions, slight modifications to the RNA synthesis cycle were required. Namely, as suggested by the supplier, an additional capping step (Ac_2O) was required after the oxidation step. Also the detritylation step was shortened to 90 sec from 120 sec, (chapter 6, section 6.3.2.1). The acid sensitivity has been reported to become a problem when longer RNA oligomers (> 40 mer) are synthesized, due to the repeated removal of 5'-trityl groups under acidic conditions during chain elongation.

After assembly of the RNA chain, the DNA portion of msDNA was synthesized following the same steps as for regiospecific synthesis of branched DNA (section 2.5 and scheme **2.10**). After cleavage and deblocking of the RNA-DNA chimera from the solid-support under standard conditions (4:1 NH_4OH/EtOH, 48 hrs, rt), the crude mixture was quantitated and divided into small portions before proceeding with the cleavage of the 2'-O-Fpmp groups. The 2'-O-Fpmp groups were deblocked under reported buffered acidic conditions known to lead to rapid cleavage of Fpmp group while minimizing any deleterious effect of acidic conditions on the oligonucleotides (e.g. 3',5' to 2',5' isomerization of the phosphodiester linkage).[214]

Initial characterization was by gel electrophoretic mobility in comparison to a linear oligomer of similar size, figure **2.11**. The two major slow moving bands were excised from a preparatory PAGE and extracted with water. The purified material obtained from the "msDNA" band (figure **2.11**) was analyzed by MALDI-TOF mass spectroscopy. MALDI-TOF mass spectrometry of this material corroborated the composition of the expected 21-unit long msDNA [M (Na+): 6656.5; found: 6655.8 ± 6]. [Insufficient material was obtained from the slowest moving band to allow proper characterization.]

Further evidence for the presence of branch structure was gained from the enzymatic hydrolysis with the yeast debranching enzyme (yDBR), chapter 4; section 4.2 and bovine pancreas deoxynuclease (DNase I). In the case of yDBR the DNA part of msDNA was expected to be cleaved at the 2',5'-intenucleotide linkage leaving the RNA intact (scheme **2.11**).

Scheme 2.11



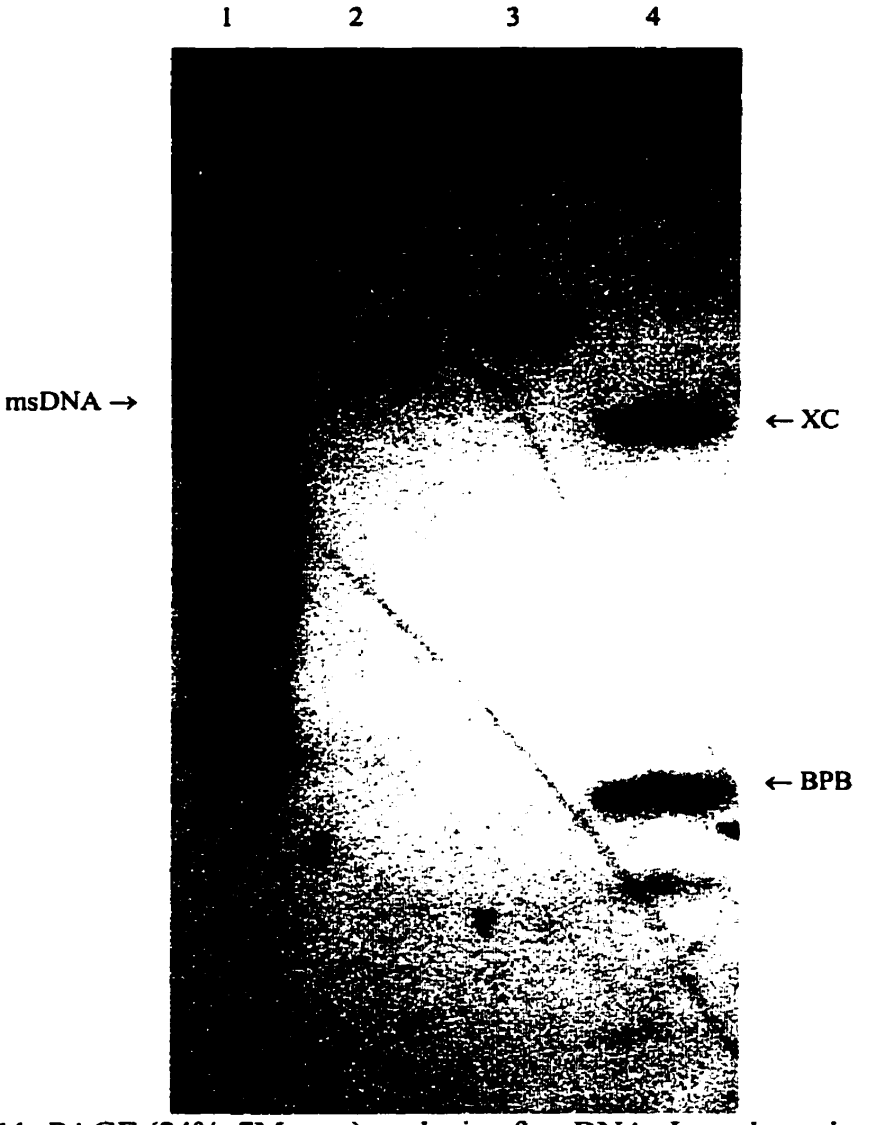


Figure 2.11: PAGE (24%, 7M urea) analysis of msDNA. *Lane 1*: crude synthetic mixture of msDNA; *lane 2*: linear 26-mer DNA; *lane 3*: linear 28-mer DNA; *lane 4*: marker dyes.

Figure **2.12** shows that the labeled msDNA is indeed sensitive to both enzymes and is partially digested to a product that corresponds to the ^{32}P-labeled RNA lacking the 2'-DNA branch. With yDBR the extent of debranching observed (*ca.* 10-20%) was similar to wild-type msDNA and branched RNA structures that have a pyrimidine at the 2'-position. On hydrolysis with DNase I the product migrated one base longer than the linear RNA control, figure **2.12**, *lane 6*. This

indicated that the product contains the 5'-end deoxycytidine residue, which is directly attached to the (DNase I resistant) RNA chain *via* the 2',5'-phosphodiester bond.

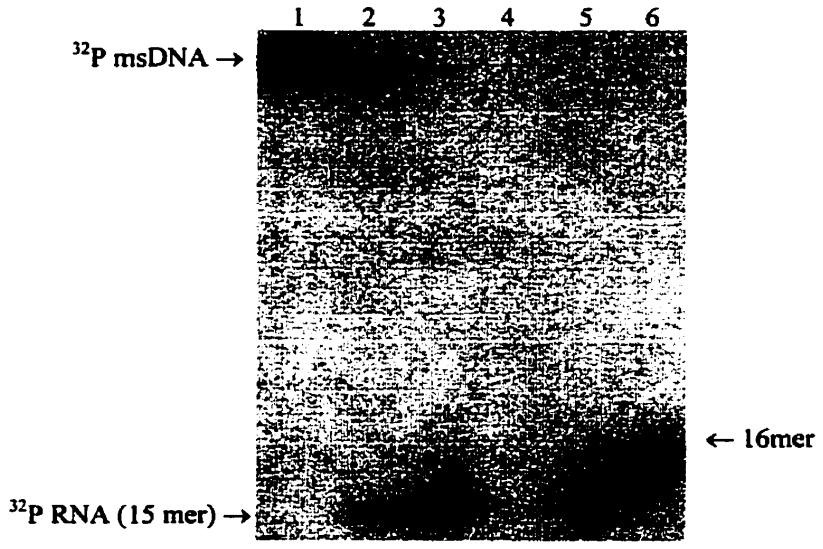


Figure 2.12: Image of PAGE (20%, 8 M urea) analysis of msDNA and its hydrolysis product upon digestion with yDBR and DNase I. *Lane 1*: 5'-^{32}P-labeled msDNA; *lane 2*: msDNA + yDBR, the cleaved DNA chain is not detected since it lacks the ^{32}P 5'-label; *lane 3*: 5'-^{32}P-labeled linear RNA (5'-GCGAUCAGCCUCAGG-3'); *lane 5*: linear RNA + yDBR; *lane 6:* msDNA + DNase I.

2.7 CONCLUSIONS

A new method for the rapid and efficient regiospecific solid-phase synthesis of "V" and "Y" shaped branched oligonucleotides was developed. Slight modifications to this protocol allowed for the first chemical solid-phase synthesis of a msDNA analogue. This method now enables one to independently change the sequences around the branchpoint of bDNA, msDNA, and tetrameric bRNA. Some of the lessons learned in this investigation have already been applied to the synthesis of hyper-branched (dendrimeric) nucleic[373]. The rich and varied set of branched oligonucleotides prepared in this study will be used to probe two important biochemical processes, *i.e.* RNA splicing and debranching (chapter 4).

3. RESULTS AND DISCUSSION

ASSOCIATION PROPERTIES OF BRANCHED NUCLEIC ACIDS: TRIPLE HELIX AND HAIRPIN FORMATION

3.1 INTRODUCTION

The structural diversity of nucleic acids has become increasingly evident in recent years. First described by Felsenfeld and Rich[215] in 1957, nucleic acid triple helices have received renewed interest in recent years for their possible role in the control of gene expression.[216,217]

Nucleic acid triple helices are formed when a short single-stranded oligonucleotide of the appropriate sequence binds to the major groove of a double-stranded homopurine-homopyrimidine complex.[218] Investigation of physical and chemical requirements for triplex formation has differentiated DNA triple helices into two distinct motifs based on the mode of binding and sequence composition of the short single-stranded oligonucleotide (referred to as "third strand").[218] In the 'pyrimidine motif', a *pyrimidine*-rich third strand binds *parallel* to the purine strand of the Watson-Crick duplex through Hoogsteen hydrogen bonds[219], forming T:A*T and C:G*C^+ triplets*.[220-222] This motif is also referred to as the 'parallel motif'. The requirement of protonation of third strand cytosines makes this triple helix motif pH-dependent. To overcome this pH-dependence several reports investigating the triple helix formation with cytosine analogs which do not require protonation have appeared.[223-229]

*T:A or C:G indicate Watson Crick base pairing while A*T or G*C^+ indicate Hoogsteen base pairing.

The second triple helix motif, designated the 'purine' or 'antiparallel' motif, involves the formation of T:A•A, C:G•G, and T:A•T triplets* when the third strand (*purine*-rich) binds in an *anti-parallel* fashion to the Watson-Crick duplex purine strand[230-233] *via* reverse Hoogsteen interactions. Formation of purine triple helix motif often requires high divalent cation concentrations. Furthermore, the usually guanosine-rich third strand is seen to form self-associated structures[234] under conditions employed for triplex formation and thus impedes the formation of purine motif triplexes.[235-237] Guanosine analogues which lack the ability to self-associate but still recognize the C:G base pair have been reported to overcome the problems associated with self-association of guanosine rich sequences.[238]

With some of the sequence restrictions imposed on triple helix formation, recent work has focused on search for ways to stabilize nucleic acid triplexes. Methods such as covalent attachment of intercalating groups,[239-245] cross linking of the strands[246] and use of triple helix binding ligands[244,247-252] to cite a few examples have been used. Damha's group has reported the ability of branched oligonucleotides to recognize and bind single stranded target polypyrimidine strands in the antiparallel (reverse-Hoogsteen) motif.[172] Similarly, Kool and coworkers have investigated the use of circular DNA and RNA molecules for stabilization of parallel-motif triplexes.[253-260]

Several approaches, such as gel-mobility shift,[261] DNase I footprinting,[262] NMR,[221,222] fluorescence,[263] and circular[263] and linear[264] dichroism studies, FTIR,[265-268], molecular modeling[269,270] and photo cross-linking[271], have been described for studying triple helices. However observation of changes in UV-

*here A•A, A•T, or G•G indicate reverse Hoogsteen base pairing.

light absorbance by the nucleic acid complexes during their association or dissociation has been the most widely used technique for following such processes and is commonly applied in the study nucleic acid triplexes. That polynucleotides absorb UV light (λ_{max} *ca.* 260nm) considerably to a lesser extent relative to the sum of the individual constituent nucleotide monomers has long been known and is termed *hypochromism* ("less color"). The difference in UV-light absorbance arises due to the base-stacking interactions of adjacent nucleotides in nucleic acid polymers and is very sensitive to the environment due to changes in conformation. Thus aqueous solutions of single-stranded nucleic acids are known to absorb less UV light at low temperatures than at higher temperatures. Also nucleic acids in a complex absorb less UV light than the sum of the individual constituent single strands due to electronic interactions between the chromophoric bases in the multi-stranded complex.

The difference between the absorption by the nucleic acids in the initial state (A_o) and the final state (A_f) is referred to as the degree of hypochromicity of the original state and is usually expressed as a percentage, where $\Delta H\% = [(A_f - A_o)/A_f]x100$, with the less light absorbing state being considered the initial state. Thus for single stranded nucleic acids, the degree of hypochromicity refers to the difference between the absorbance of the intact polynucleotide, initial state (A_o), and that of constituent mononucleotides, final state (A_f). For nucleic acid complexes, the difference in absorbance of the complex or the initial state (A_o), and the fully uncomplexed final state (A_f) is reported as the degree of hypochromicity.

Thus, nucleic acid complex formation can be investigated by following the changes in the absorbance at 260 nm of a nucleic acid solution as a function of the

variable of interest. Most commonly this involves mixing the constituent strands together in an aqueous buffer, allowing them to anneal at low temperature and then observing UV-absorption as temperature is slowly raised. This results in an absorbance *versus* temperature plot, more commonly referred to as the melting curve or the melt profile. When a solution of fully complexed nucleic acid strands is heated, the ordered helix undergoes unwinding or *denaturation* due to the dissociation of single strands from the complex. As this occurs, the UV absorbance increases, *hyperchromism* ("more color"), compared with the UV absorbance of the completely annealed strands. Association of strands can be monitored by going in the opposite direction, or the reverse of the above technique. By beginning at a high temperature, as the solution cools, the two strands of nucleic acids re-anneal, and UV absorbance is seen to decrease. The "melting temperature" (T_m) is then defined as the temperature at which the half the nucleic acid population in solution exists as complex and the remaining half as single strands. Usually, at the T_m, hyperchromicity is seen to be half maximal.

The melting temperature (T_m) of a short double-stranded oligonucleotides is usually taken as a measure of the double helix stability. The stability of double helix is strongly dependent on chain-length and base sequence and concentration as well as the ionic strength of the solution.

3.2 PARALLEL T:A*T VERSUS ANTIPARALLEL T:A•T TRIPLEXES

Thymidine residues in the third strand can bind with the adenosine residues of the Watson Crick duplex *via* Hoogsteen or the reverse-Hoogsteen hydrogen bonding, figure **3.1**. Thus T:A*T triple-stranded helices can take up structures with the third (T) strand binding in either *parallel* or *antiparallel* orientation with respect to the adenosine strand of the Watson-Crick (T:A) base pair.[270,272,273] The first

studies on these structures examined the properties of the classical dT_{10}:dA_{10}•dT_{10} intermolecular triplex, in which the third (dT_{10}) strand is (Hoogsteen) bonded to the dA_{10} strand in the (more common) parallel orientation.[272] In contrast there have been fewer studies on antiparallel T:A•T triple helices and most of these have used a few T:A•T triplets interdispersed between the more dominant C:G•G triplets. In fact, it had been argued that blocks of antiparallel-T:A•T are not very stable and are difficult to observe.[243,244,274,275]

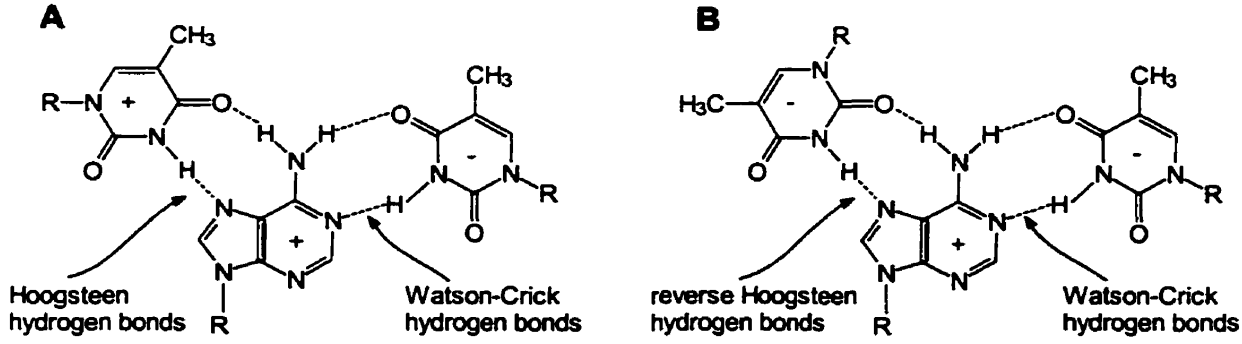
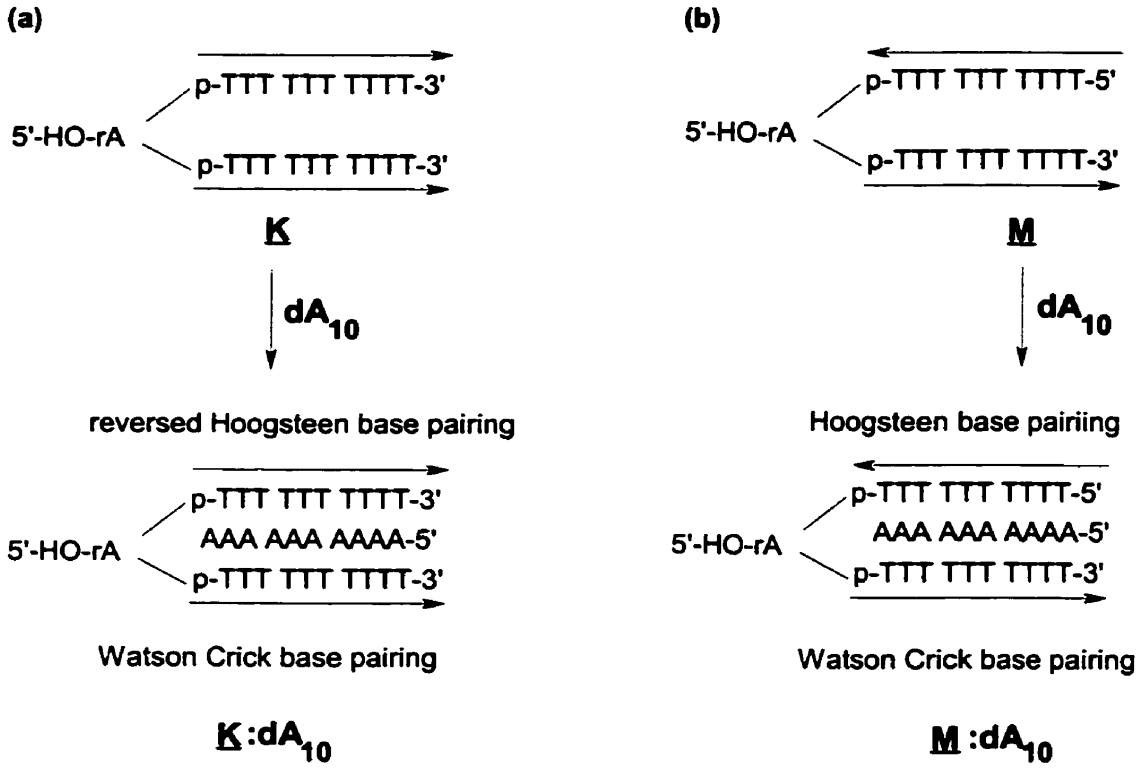


Figure 3.1: Hoogsteen vs reversed Hoogsteen bonding in T:A/T triplets.

Damha's group[172] and Taillandier and co-workers[268] have recently reported the formation of stable antiparallel T:A•T triple-stranded helices with "forced" antiparallel third (T) strand orientation. The French group induced the formation of the antiparallel T:A•T motif by folding back twice the tridecamer dT_{10}-linker-dA_{10}-linker-dT_{10} (linker = ethylene glycol polymer) in NaCl buffer (pH 7). Biphasic melting was readily apparent for this complex, with the first transition reflecting dissociation of the third strand at *ca.* 23-27 °C, followed by a second transition at *ca.* 50 °C reflecting the melting of the dA_{10}-linker-dT_{10} duplex. In the Damha lab, the triplex was obtained by mixing stoichiometric amounts of "V"-oligomer **K** and dA_{10} in magnesium buffer[172] (scheme **3.1**). Here, the central riboadenosine unit served as a rigid "linker" to which the dT_{10} strands are joined through the neighboring 2'- and 3'- oxygen atoms. The geometrical constraints

are such that, when the triplex is formed, both dT strands are parallel and thus *anti-parallel* with respect to the purine strand, scheme **3.1b**. If the third-strand (dT) orientation were inverted from antiparallel to parallel, its binding scheme should also change from reverse-Hoogsteen[272] to Hoogsteen[268]. This effect was investigated using the branched compound **M** (scheme **3.1**). Direct comparison between **M**:dA_{10} and the known complex **K**:dA_{10} would allow us to measure, for the first time, the relative stability of T:A*T and T:A•T triplets.

Scheme 3.1



3.2.1 Association of M and dA_{10}

The association properties of branched nucleic acids **M** and **K** were chosen as model compounds for studying the relative stabilities of parallel and antiparallel T:A/T triple helices. Whereas in compound **K** the dT_{10} strands attached to rA are parallel to each other and have identical polarity (5' → 3'), the two dT_{10} strands in

compound **M** lie antiparallel to each other and thus have opposite polarities (5' → 3' and 3' → 5'), scheme **3.1a**. Thus on formation of a triple helix with the dA_{10} complement strand, the third (dT) strand will bind in the readily observed and well characterized parallel (Hoogsteen) (**M**:dA_{10}) or the less commonly observed antiparallel (reverse Hoogsteen) (**K**:dA_{10}) fashion, scheme **3.1b**.

The helix-coil transition of the two complexes was followed by observing the corresponding UV-spectral changes and results are reported in Table **3.1**. The complex formed between **K** and dA_{10} shows a melting profile similar to as reported by Hudson *et al.*, who have shown that **K**:dA_{10} is a bimolecular triple helix involving both dT_{10} strands of the V-shaped **K** molecule.[172] The melt curve at 260 nm for the **M**:dA_{10} complex also shows a monophasic, cooperative transition at 42.5 °C, figure **3.2a**. However, the *Tm* value for the helix-to-coil transition is 10 °C greater for the **M**:dA_{10} than for **K**:dA_{10} or the dA_{10}:dT_{10} Watson-Crick duplex. The higher *Tm* of **M**:dA_{10} relative to the duplex dA_{10}:dT_{10} suggests the formation of a complex which involves interactions in addition to the expected Watson-Crick base pairing. As the formation of a triple helical complex between **K** and dA_{10} had already been established, it was reasonable to suggest that the molecule **M** was forming a parallel-type triple helix structure by using both dT_{10} 'arms' in binding to the dA_{10} target (scheme **3.1**). The *Tm* differential between **M**:dA_{10} and **K**:dA_{10} thus represents the difference in stabilities between the parallel and antiparallel T/AT triplexes.

The formation of a triple-helical complex between **M** and dA_{10} was confirmed by several experiments. For example the melting of the presumed **M**:dA_{10} triplex complex could be followed independently at 284 nm.[272] At this wavelength the Watson-Crick duplex dA_{10}:T_{10} does not show any significant changes in

absorption during the melting process as the UV spectra of native duplex dA_{10}:dT_{10} shares a (virtual) isosbestic point with the melted dA_{10} + dT_{10} single strands.[276-279] Under the same conditions, however, the triplex does not share the same isosbestic point with the transition of a triplex to duplex + single strand,[272,278-280] or the triplex to single strands.[172,253,279,280] Thus any hyperchromic absorbance changes observed at 284 nm indicate melting of a third strand from the underlying duplex. Indeed, monitoring the dissociation of **M**:dA_{10} complex at 284 nm showed a small cooperative transition of significant hyperchromicity. Under the same conditions the control dA_{10}:dT_{10} duplex showed no observable hyperchromicity at 284 nm on melting, as expected.[172,272,279]

Table 3.1: Thermal melt data for the two triple helix complexes

Complex components	buffer system[a]	*Tm*[b] (°C)	Δ H%[c]	curve shape
$A^{T10\text{-}3'}{}_{T10\text{-}3'}$ (**K**) + dA_{10}	Na^+	30.9	21.5	monophasic
$A^{T10\text{-}5'}{}_{T10\text{-}3'}$ (**M**) + dA_{10}	Na^+	40.0	12.5	monophasic
$A^{T10\text{-}3'}{}_{T10\text{-}3'}$ (**K**) + dA_{10}	Mg^{2+}	32.7	24.6	monophasic
$A^{T10\text{-}5'}{}_{T10\text{-}3'}$ (**M**) + dA_{10}	Mg^{2+}	42.5	14.5	monophasic
$A^{T10\text{-}5'}{}_{T10\text{-}3'}$ (**M**) + 2 dA_{10}	Na^+	41.7	13.5	monophasic
$A^{T10\text{-}5'}{}_{T10\text{-}3'}$ (**M**) + 2 dA_{10}	Mg^{2+}	43.6	13.3	monophasic

[a]*buffer systems*: Na+ = 1 M NaCl, 10 mM Na_2HPO_4, pH 7.0; Mg^{2+} = 50 mM $MgCl_2$, 10 mM TRIS, pH 7.3. [b]T_m's were determined from plots of first derivative of absorbance against temperature and represent average of four-to-six melting curves. [c]ΔH% values were calculated using the formula $[A_f - A_o)/A_f] \times 100$, where A_0 = initial absorbance and A_f = final absorbance.

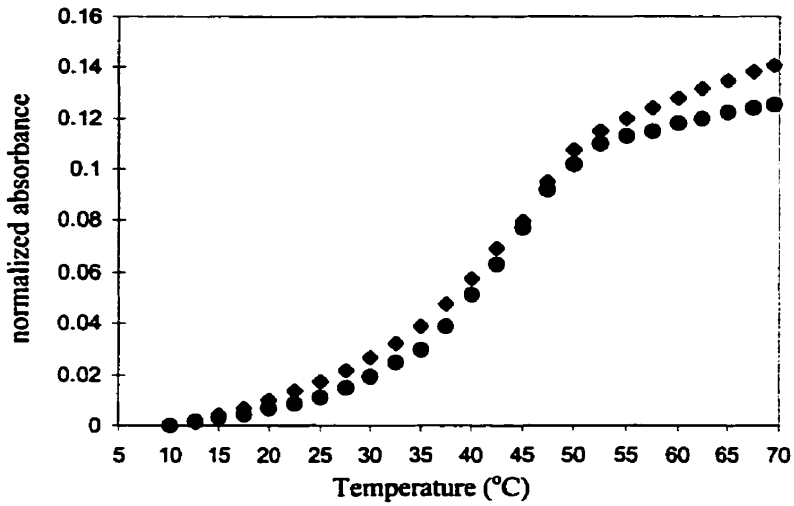


Figure 3.2: Melting profile for the complex formed between **M** and dA_{10}, 1:1 (•) and 1:2 (◆) monitored at 260 nm. Normalized absorbance was calculated by: $[(A_t - A_i)/(A_f)]$ where A_t = absorbance at temperature t; A_i = initial absorbance; A_f = final absorbance. Conditions for T_m determination are 50 mM Mg^{2+}, 10 mM Tris, pH = 7.3.

3.2.1.1 Stoichiometry of Interactions between M and dA_{10}

Upon addition of more dA_{10} to the complex **M**:dA_{10}, no significant change in the thermal denaturation profile was observed (figure **3.2**, table **3.1**), suggesting that **M** and dA_{10} interact in a 1:1 stoichiometry. The UV properties in mixtures of continuous variation[281] were then employed to independently determine the stoichiometry of the interaction between **M** and dA_{10}.

The continuous variation method[281] involves monitoring the absorbance of a solution having continuously varying proportions of the two interacting stands while maintaining an equimolar total concentration of the strands in solution. The theory behind this approach is that an absorbance 'breakpoint' should be observed at the exact molar ratio of interacting strands in a specific complex. Since the formed complexes are usually hypochromic, due to differential base-stacking interactions in the complexed and uncomplexed states, a minimum absorption occurs where the largest amount of complex is formed.

When **M** was titrated with dA_{10} at 5 °C in the Mg^{2+} buffer, the mixing curve yielded a breakpoint at *ca*~ 0.5 mole fraction in dA_{10}, regardless of the direction of titration for **M** and dA_{10}, figure **3.3**. This result is consistent with a 1:1 **M**:dA_{10} triplex complex involving both dT_{10} strands. However it could also be reasoned that **M**:dA_{10} is a duplex wherein binding of dA_{10} to the first dT_{10} tail sterically inhibits the association of another dA_{10} molecule to the second dT_{10} tail. Similar arguments were also initially presented for the **K**:dA_{10} complex and later shown to be improbable. Further evidence for triplex formation comes from the CD studies described below.

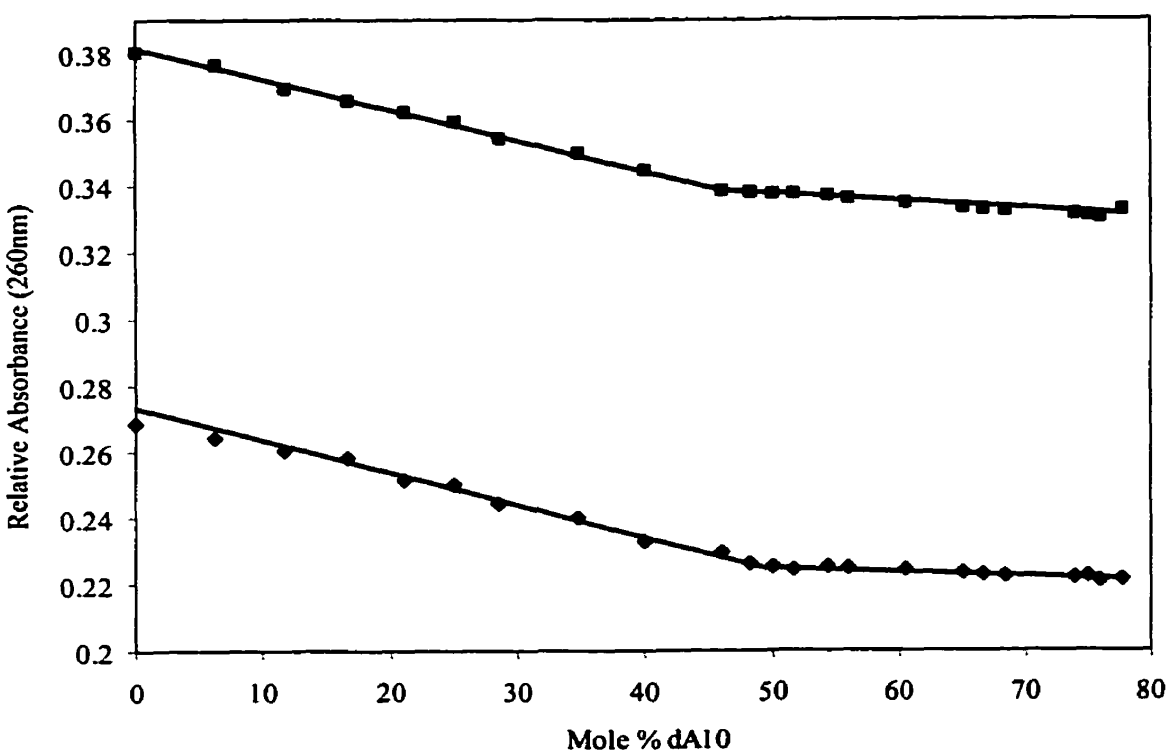


Figure 3.3: Determination of the stoichiometry of interaction between **M** and dA_{10} by the method of continuous variation. Stoichiometry was determined in 50 mM Mg^{2+}, 10 mM Tris, pH = 7.3 (—◆—) or 122 mM K^+, 4.31 mM NaH_2PO_4, 0.8 mM Mg^{2+}, pH 6.61 (—■—). The complexes were allowed to equilibrate for 15 minutes at 5 °C before each reading.

3.2.1.2 Circular Dichroism Studies on the M:dA_{10} Complex

Circular dichroic spectroscopy, (CD spectra) has been used extensively for studying nucleic acids complexes such as DNA:DNA or RNA:RNA duplexes, DNA:RNA hybrids, DNA-protein complexes, binding of small ligands to DNA duplexes, and DNA triple helices. CD spectra contains more information than simple absorption spectra and are thus useful in providing valuable structural information and is highly suitable for distinguishing between different DNA structures. CD experiments measure the differential absorbance of right- and left-handed circularly polarized light over the range of wavelengths where UV-absorption occurs.[282] A region of rapid change with respect to wavelength is the predominant feature of a CD spectra and designated the Cotton effect. Both the sign and the intensity of Cotton effect are sensitive to the small perturbations in the structures of molecules and provide information about the structure.[283]

In 1988, Antao *et al.*[284] observed that the formation of the triple helix $d(CT)_{12}$:$d(AG)_{12}$*$d(C^+T)_{12}$ could be followed by CD measurements. Pilch and coworkers[233,278] also observed differences between the CD spectra of dT_{10}:dA_{10}*dT_{10} triple helix and the corresponding dT_{10}:dA_{10} Watson-Crick duplex. The major difference between the duplex and triplex spectra included an amplitude increase of the negative band at 248 nm and slight amplitude decrease in the positive bands at 259 nm and 284 nm for the triplex. A slight red-shift of the triplex bands was also observed relative to these bands observed in the duplex.

Complex formation between M and dA_{10} was thus studied by CD. In 50 mM Mg^{2+} buffer, at 5 °C, the CD spectra of M is qualitatively similar to the reported dT_{10} spectra. The CD spectra shows a large positive Cotton effect at 280 nm, a cross over at 261 nm which is close to the maximum of the UV absorbance (260

nm), and a weaker negative effect at 246 nm, figure **3.4**. The position of the positive band at 280 nm is between that of reported values for dT_{10} and the linear 11-mer dAT_{10} (279 nm). Complex formation is readily monitored by changes in the position of bands and their amplitudes.

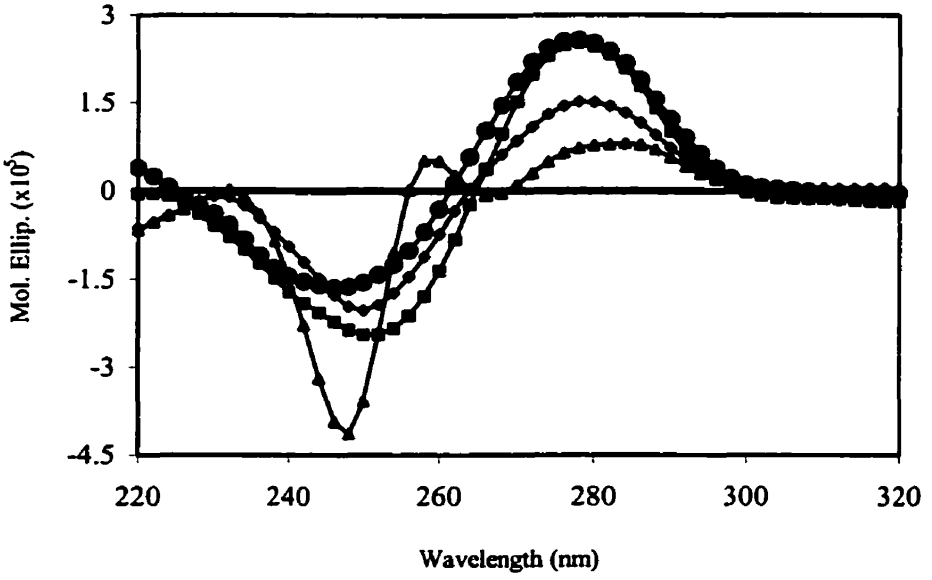


Figure 3.4: Circular ichroism of **M** alone at 5 °C (—■—), 60 °C (—●—), and **M**:dA_{10} complex at 5 °C (—▲—), at 60 °C (—◆—).

The CD spectra of complex formed between the compound **M** and dA_{10} is shown in figure **3.4**. In Mg^{2+} buffer at 5 °C, a temperature where the triple helix complex exists, the CD spectrum of **M**:dA_{10} (1:1 mixture) shows great similarities to the CD spectra of the dT_{10}:A_{10}*dT_{10} triple helix as reported by Pilch *et al.* and the bimolecular triple helix formed by the hairpin loop $d(T_{10}C_4T_{10})$ and dA_{10}, figure **3.5**. In comparison to the CD spectra of dT_{10}:dA_{10} duplex, the **M**:dA_{10} complex CD spectra shows a reduction in amplitude and red-shift of positive bands at 259 nm and 284 nm. The negative band at 247 nm was also red-shifted but its amplitude showed an increase relative to the duplex negative band, figure **3.6**.

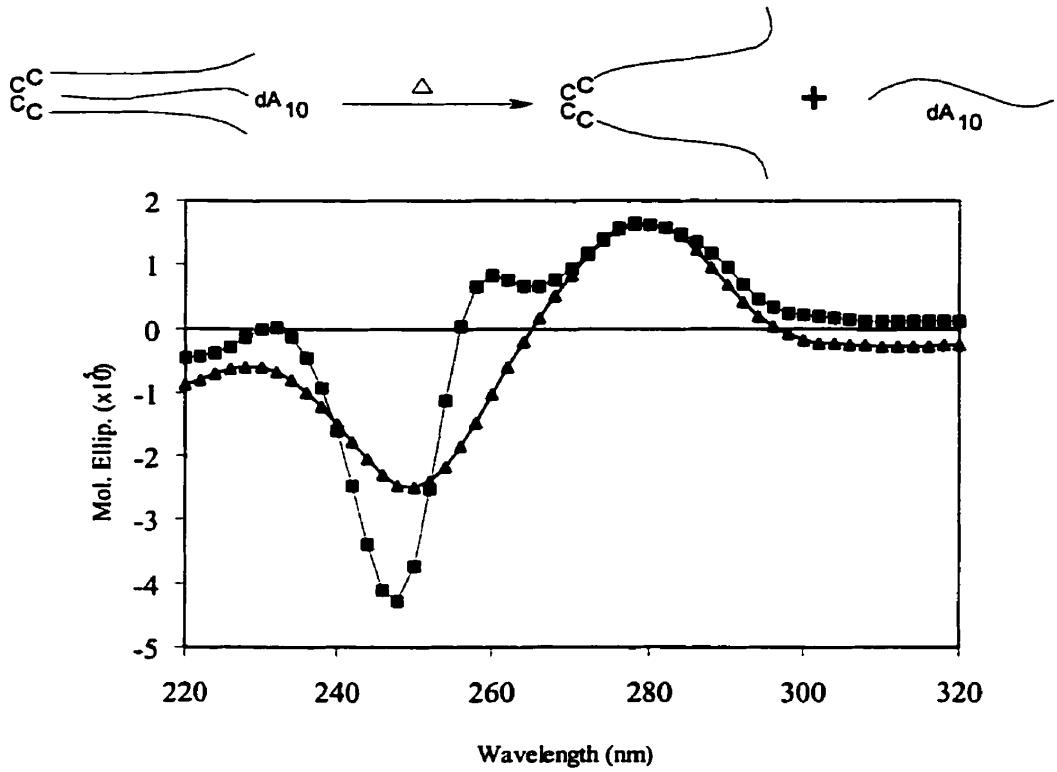


Figure 3.5: Circular dichroism of hairpin loop $d(T_{10}C_4T_{10})$:dA_{10} complex at 5 °C (—■—) and 60 °C (—▲—).

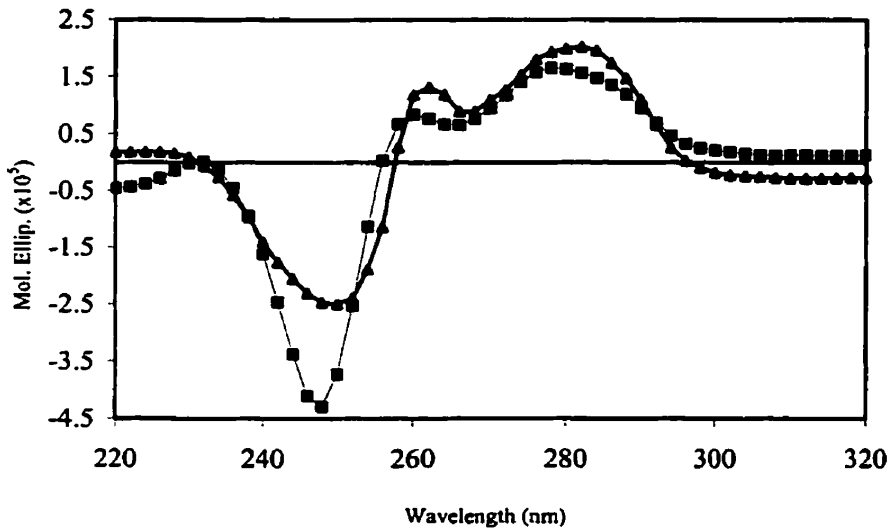


Figure 3.6: Circular dichroism of M:dA_{10} complex (—■—) and dA_{10}:dT_{10} duplex (—▲—), at 5 °C.

On heating, CD spectra of the complex **M**:dA_{10} was seen to undergo significant changes, particularly at temperatures close to the *Tm* of the complex, figure **3.7**. On heating, as the temperature neared the melting temperature, the amplitude of the positive band at 259 nm was reduced until the band totally disappeared at 40 °C. Similarly, the negative band at 247 nm underwent a reduction in amplitude as the temperature increased. On the other hand, the positive band at 284 nm was seen to increase in amplitude as the temperature was raised. All three bands displayed small blue shifts as the temperature increased. A clear isoelliptic point was seen at 253 nm in the temperature dependent CD spectra. At 60 °C, the spectrum is similar to a superimposition of the spectra of the free **M** and dA_{10}. Taken together the CD data supports the formation of a triple helix between **M** and dA_{10}.

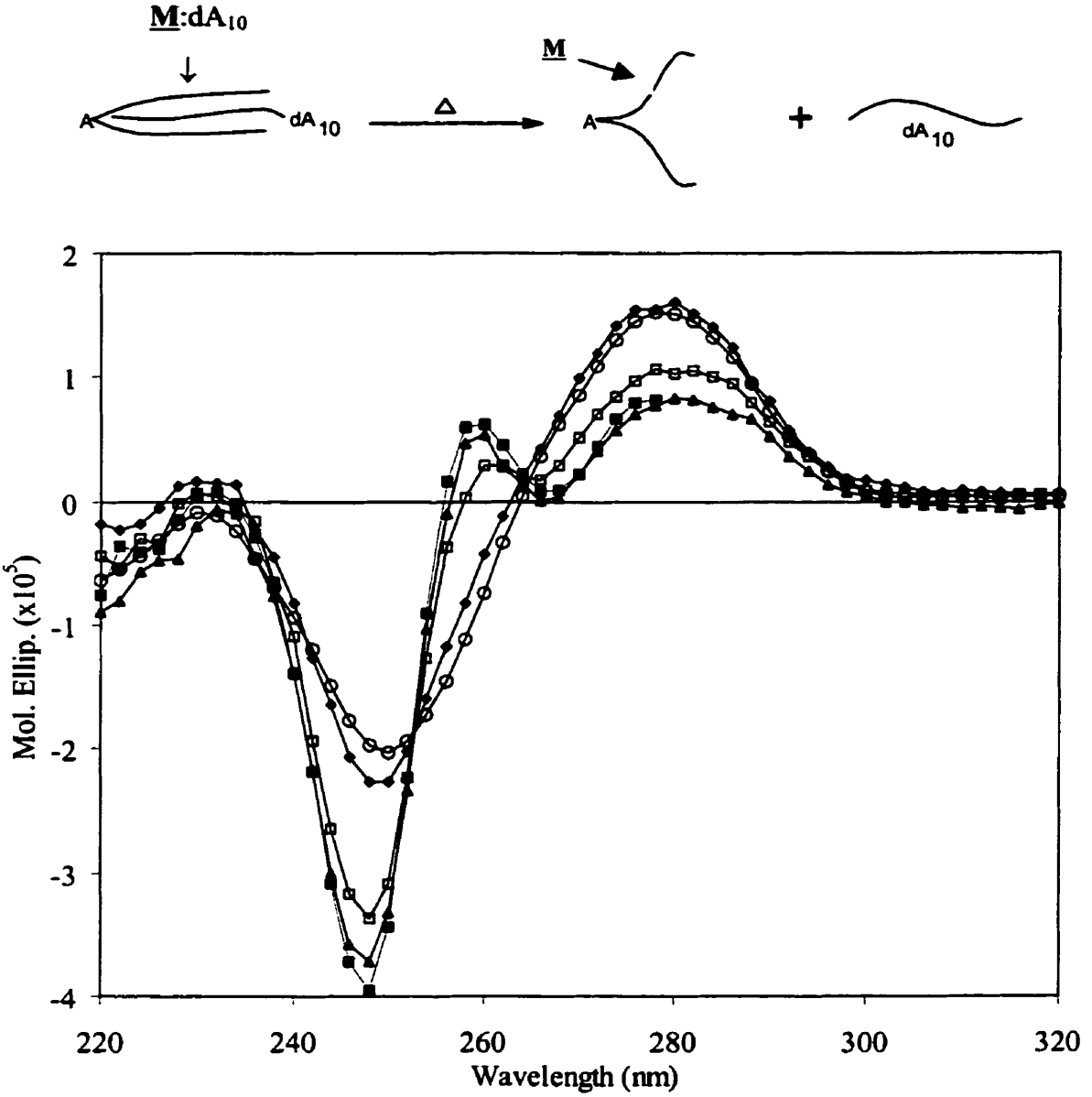


Figure 3.7: Circular dichroism of M:dA_{10} complex at 15 °C (—■—), at 25 °C (—▲—), at 35 °C (——), at 45 °C (—◆—), and at 60 °C (—o—).

3.2.1.3 Melting Temperature Determination From CD Spectra

As the CD band amplitudes and positions show temperature dependence, this information can be used to calculate melting temperatures (*Tm*). Chan and Breslauer[279,285-287] recommended following the temperature induced change in the CD signal at wavelengths corresponding to the maximum change in ellipticity.

In the case of T:A*T triplexes, the negative band at 247 nm and the positive band at 259 nm represent the points of maximum change as shown by temperature dependent CDs in figure **3.7**. Figure **3.8** shows such CD melting curves for the **M**:dA_{10} triple helix complex at 259 nm and 247 nm. CD signal at both wavelengths behave in a manner similar to the UV-absorption melting profiles and show abrupt changes near the *Tm* point. Determination of temperature where the greatest change occurred gave a value close to *Tm* determined from UV-absorbance melt curves (*ca.* 43 °C). Others have also used temperature-dependent changes in CD spectra to follow the association and/or dissociation of nucleic acid complexes.[286] Thus by employing CD experiments one can gain insight not only into the structural aspects but also determine the melting temperatures for dissociation of nucleic acids complexes.

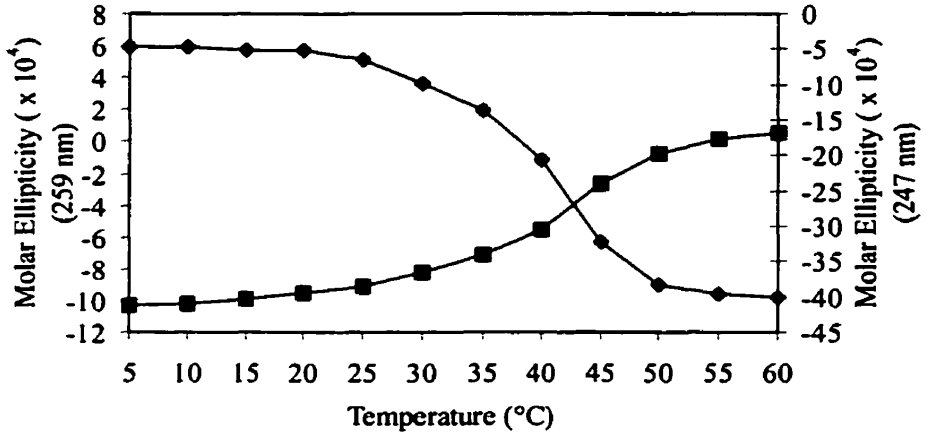


Figure 3.8: Temperature dependence change in molar ellipticity at 259 nm (—◆—), left axis, and at 247 nm (—■—), right axis.

3.2.2 Relative Stability of Parallel versus Antiparallel T:A/T Triplexes

Very little information is currently known about the thermodynamic properties of the anti-parallel T:A•T triplex. The results of molecular dynamic studies[270,288] and a few *ab initio* calculations,[289] comprise the majority of this information. The ability of the isomeric oligomers **M** and **K** to bind to dA_{10} and form, respectively,

parallel and antiparallel triplexes made them ideal tools for studying the relative stabilities of parallel and antiparallel T:A/T complexes.

The thermal stabilities of the complexes **M**:dA_{10} (p-T:A*T) and **K**:dA_{10} (ap-T:A•T) are brought out in figure **3.9**. The melting curves show a single inflection, indicative of cooperative melting and reflecting triplex-to-coil equilibria. The hyperchromicity exhibited by **K**:dA_{10} (ap-T:A•T) on melting is significantly greater than for **M**:dA_{10} (p-T:A*T) (25% versus 14%, table **3.1**), and reflects different stacked conformations between these triplexes. However, the T_m value for the helix-to-coil transition is 11°C greater for **M**:dA_{10} than for **K**:dA_{10} suggesting that the parallel triplex is more stable than the antiparallel triplex. In agreement with this conclusion are the results of thermodynamic analysis described below.

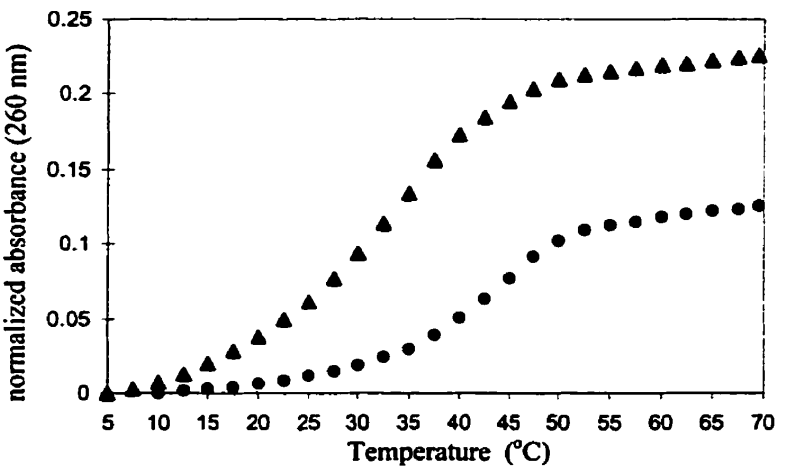


Figure 3.9: Melting profile for the complex formed between **M** and dA_{10}, (—•—) and **K** and dA_{10} (—▲—) monitored at 260 nm. Normalized absorbance was calculated by: $[(A_t - A_i)/(A_f)]$ where A_t = absorbance at temperature t; A_i = initial absorbance; A_f = final absorbance. Conditions for T_m determination are 50 mM $MgCl_2$, 10 mM Tris, pH 7.3.

3.2.2.1 Thermodynamic Parameters for Parallel and Antiparallel T:A/T triplexes

The helix to coil transitions for both triplexes have $1/T_m$ values that decrease linearly with increasing log of total strand concentration, C_{tot}, figure **3.10**. In addition to supporting that these transitions are bimolecular in nature, this linear correlation is consistent with triplex dissociation occurring as "all-or-none" process. Similar one step tsDNA → ssDNA transitions has been observed by Kool and coworkers.[253-259] By employing circular DNAs, they demonstrated the one step melting of pyrimidine (parallel) type triplexes to single strands. Hudson *et al.*[172] have also observed a similar one step melting of triple helix complexes of branched oligonucleotides.

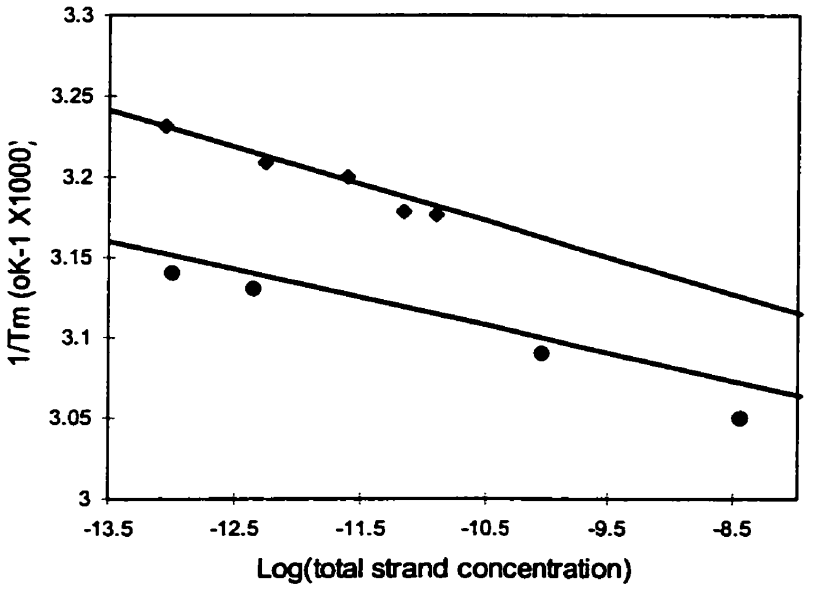


Figure 3.10: Reciprocal melting temperature vs log (total stand concentration) of **K**:dA_{10} (—●—) and **M**:dA_{10} (—◆—) complexes. Conditions for T_m determination were 50 mM $MgCl_2$, 10 mM Tris, pH 7.3.

Several methods for determining the thermodynamics of triple helix formation have been described. Some methods such as affinity cleavage,[290-295] filter binding[296] and fluorescence resonance energy transfer,[297] involve direct measurement of the binding constant K, and corresponding ΔG while the

remaining parameters are then determined from the temperature dependence of *K*. Thermodynamic properties have also been calculated based on spectrophotometric and calorimetric techniques.[233,269,278,289,298-302]

There has been some reports of discrepancy between enthalpy values obtained from direct calorimetric determinations and those determined by van't Hoff analysis, for the dissociation of the third strand.[299,303-305] Plum *et al.*[306] attributed this due to the dissociation not occurring as an all-or-none process. Recent work by others have confirmed that when the dissociation occurs as an all-or-none process, the enthalpy values obtained by the two methods are generally in very good agreement.[300]

Thermodynamic parameters for the triple helix complexes **M**:dA_{10} and **K**:dA_{10} were obtained using two different methods for the analysis of UV-melting curves, and are reported in table **3.2**. From the concentration dependence studies, reciprocal melting temperature vs log (concentration) were constructed, figure **3.10**, and parameters calculated from the slope ($R/\Delta H°$) and y-intercept ($\Delta S° - R \bullet \ln 4$). Based on the methods of Breslauer,[307] calculation of the equilibrium constant from the melt curves, also allowed extraction of thermodynamic properties for the triple helices under investigation. The lack of either the low- or the high-temperature baselines and/or the sloping nature of the baselines usually make the Breslauer's method hard to apply. Melting experiments were thus run such that non-sloping upper and lower baselines of sufficient length were obtained for extraction of meaningful thermodynamic data. The values obtained by both methods showed excellent agreement.

Table 3.2: Calculated thermodynamic parameters for the Hoogsteen and reversed Hoogsteen T:A/T base triplets

Complex	Binding	$\Delta H°$ [a] (kcal/mol)	$\Delta S°$ [a] (cal·deg/mol)	$\Delta G°_{25}$ [b] (kcal/mol)
K:dA_{10}	Reverse Hoogsteen + w/c	-6.8	-19.8	-10.3
M:dA_{10}	Hoogsteen + w/c	-8.3	-23.5	-13.0
dT_{10}:dA_{10}*dT_{10}[278]	Hoogsteen	-2.3	-56	-6.6
dT_{10}:dA_{10}	Watson Crick	-6.0	-171	-8.8

[a]the calculated values are for each T:A/T base triplet calculated from plots of reciprocal melting temperature versus natural logarithm of total stand concentration (section 6.6.1.1). [b]for the complete dissociation of the complex.

A survey of literature showed a range of enthalpy values for triple helices containing all-pyrimidine third strands. This range of values, from 2 to 7.7 kcal mol^{-1} /base triplet results from a number of factors including base composition and solution conditions such as salt concentrations and pH.[299-302,304,306,308] Slightly higher values obtained in this study reflect the dissociation of *both* the Watson-Crick and Hoogsteen (or reverse-Hoogsteen) interactions at the same time and thus represents the total dissociation of the complex. On the other hand, literature reported values are from the dissociation of the third strand from the underlying Watson-Crick duplex. Thus using "Hess's law", when the total dissociation of the triple helix to single strands is taken into account, enthalpy ranges from 8 to 12.1 kcal mol^{-1} /base triplet. From the seminal work of Pilch and coworkers[278] on the parallel dT_{10}:dA_{10}*dT_{10} triple helix, a total enthalpy value of 8.32 kcal/mol of base triplet for the parallel T:A*T triplex can be obtained. A triplex system consisting of dA_{12}-x-dT_{12}-x-dT_{12}, x= hexaethylene glycol linker, gave enthalpy values in the range of 9.0 kcal/mol to 9.9 kcal/mol of base triplet for the parallel T:A*T triplex depending on the salt concentration of the solution.[279] Our result of 8.3 kcal/mol of base triplet for the putative parallel **M**:dA_{10} triplex are in excellent agreement with the value reported by Pilch for

dT_{10}:dA_{10}*dT_{10} under the same conditions (*i.e.* 2.3 + 6.0). This similarity of values obtained could be taken as an indicator of the presence of 10 hydrogen bonded T:A*T base triplets in the branched triple helix complex. The agreement between the ΔH values obtained in this study and those reported by Pilch suggests that branching triplex stability (ΔG) does not derive from enthalpic differences between the initial triplex states and their corresponding final single-stranded states but rather from the entropic consequences of branching, *vide infra*. The considerably lower enthalpy of the antiparallel T:A•T (-6.8 kcal/mol, table **3.2**) may reflect the known lower stability reverse Hoogsteen interactions relative to Hoogsteen interactions.

Calculated entropy values on the other hand, table **3.2**, were significantly lower than the reported range of values for trimolecular triple helices, e.g. dT_{10}:dA_{10}*dT_{10}, but are similar to the values reported for intramolecular triplexes.[299] The range of values reported for trimolecular triplexes, 0.28 to 0.42 cal•deg•mol^{-1}/base triplet, again reflect the various conditions under which the various experiments were performed.[299-302,304,306] Lower entropy values in the branch triplexes indicate the highly structured state of the branched molecules even when they are not in a complex, and/or the reduced molecularity of triplex formation (bimolecular vs trimolecular). Thus on binding with the dA_{10} complement there is a lesser loss in entropy leading to the stabilization of both the parallel and antiparallel branched T:A/T triplexes relative dT_{10}:dA_{10}:dT_{10}.

As expected the enthalpy changes for antiparallel triplex formation were less negative than those for parallel triplex formation, table **3.2**. This enthalpy difference is the major contributor to the observed reduction in stability of the antiparallel triplex **K**:dA_{10} relative to the parallel triplex **M**:dA_{10}. Thus, the

parallel triplex was found to be more stable than the antiparallel triplex, with an average difference in free energy ($\Delta G°_{25}$) and enthalpy ($\Delta H°$) of 0.27 and 1.5 kcal/mol T:A/T base triplet, respectively. These values can only be considered an estimate, since the isomeric "V" oligomers **K** and **M** have some small differences in structure, and it is assumed that 10 hydrogen bonded T:A/T triads are possible in both complexes.[172] With these caveats in mind, these results are nevertheless in agreement with the general understanding of antiparallel T:A•T triplexes that are not very stable and usually hard to observe.

3.3 BRANCHED NUCLEIC ACIDS HAIRPINS

Determination of the three-dimensional structures of nucleic acid sequences, through NMR, X-ray crystallography, and other methods has become a major research endeavor. However at the very high concentrations required for such studies, appearance of conformers different from the native structure complicate the analysis. Thus methods are required for stabilizing these structures as close to their native state as possible. Although cross-linked oligonucleotides have been utilized in modeling high molecular weight DNA[309,310,375,376] and as conformationally stable constructs for molecular recognition studies,[311,312,377,378] several drawbacks are apparent. Cross-linking of oligomers can lead to alternation in the stability of the structures through changes in the hydration.[313,314] Position of cross-links may also impede recognition by ligands of interest.[315] In addition cross-linking can disrupt native geometry.[316]

Discovery of unusually stable nucleic acid 'hairpins' has led to their use in stabilizing DNA and RNA structures and for stabilization of mRNAs and antisense oligonucleotides, (see Varani[317] for a review of "exceptionally stable nucleic acid hairpins"). Based on the structural rigidity observed in model branch

nucleic acids, the use of branched oligonucleotide hairpins for similar purposes was contemplated for the first time.

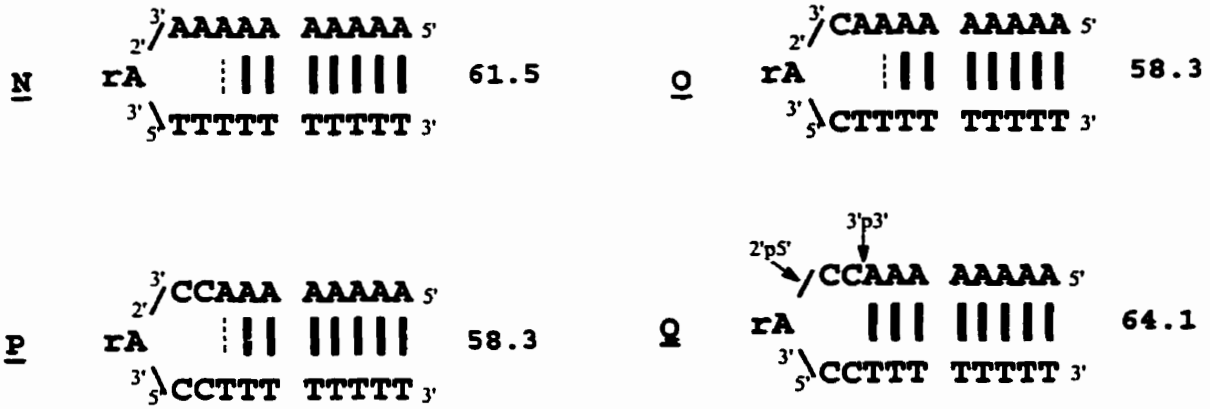
3.3.1 Association Properties of Branched Nucleic Acid Hairpins

Studies on branched nucleic acids have shown the phosphate framework around the branchpoint core to be rigid due to strong base stacking between the branchpoint residue and the 2'-residue.[114,188,318,319] Furthermore the 2',5'- and 3',5'-phosphodiester linkages are oriented nearly parallel with respect to each other[318] so that branched molecules **K** and **M** can be topologically regarded as V-shaped structures. This has led us to consider branch molecules **N** – **Q** as models for hairpin formation (table **3.3**).

As the orientation of the two strands can be controlled independently, (see chapter 2), these molecules are highly attractive candidates for studying duplex structure *via* hairpin formation.[320] Also the effect of branching on the stability of hairpin structures can be determined by comparing the stability of **N**, **O**, **P**, and **Q** with that of the linear control hairpins **1** and **2** (table **3.3**).

Table 3.3: Melting temperatures (°C) of branched and linear hairpin DNA

Branched



Linear

*conditions for T_m determination: 50 mM Mg^{2+}, 10 mM TRIS, pH 7.3 with a final oligomer concentration of 1.1 μM.

Compounds **N** - **Q** exhibited single cooperative melting transitions (T_m) that were nearly independent of oligonucleotide concentrations over a 10 to 50-fold range (50 mM $MgCl_2$, 10 mM Tris, pH 7.3), figure **3.11**. In addition, their T_m values are >28 °C higher than for double-stranded complexes formed from two independent strands e.g., dA_{10} + dT_{10}, T_m 32 °C. These results support the view that branched oligomers can form hydrogen-bonded, base-stacked structures reversibly in solution by folding intramolecularly to a hairpin conformation.

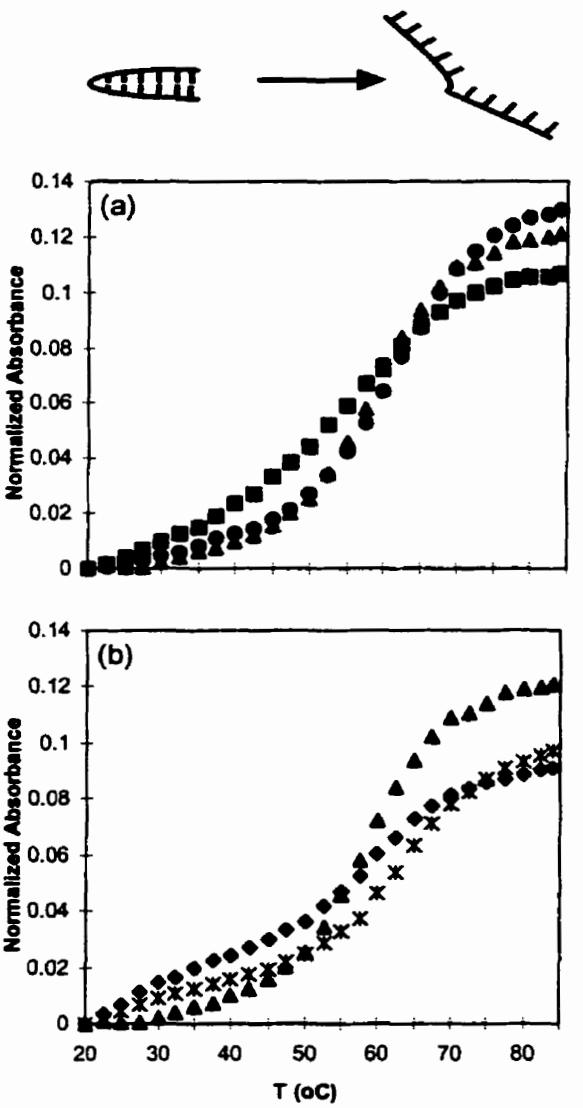


Figure 3.11: Melting curves of branched and hairpin structures in 50 mM Mg^{2+}, 10 mM TRIS, pH 7.3. (a) N (•), O (■), P (✳), Q (◆) and 1 (▲). Oligomer concentration was 1.1 μM.

Molecular modeling performed on hairpins N, O and P, shows that their structure is energetically most favorable when the "loop" is composed of a heptanucleotide (5'...AXX-3',2'rA-3',5'YYT...3') and the "stem" has seven A:T base-paired residues.

In agreement with the view that N, O and P have similar folded structures, the temperature of dissociation of these complexes are comparable (58-61 °C, table 3.3, figure 3.11). In contrast, the energetically optimized structure of compound Q

(the constitutional isomer of **P**) can accommodate an additional A:T base pair, leaving a loop with only five unpaired bases (5'...CC-5',2'rA-3',5'CC...3') that is constrained sterically by the requirements of loop closing. It is noteworthy that the structural features of the loop in **Q** are in excellent agreement with the NMR-derived structure of small branched RNA fragments,[318,319] the most significant being the extensive base-base stacking interactions between the central rA, the 2'5'-dC and the next dC residue (...dC dC 5',2' rA...), figure **3.12, 3.13**. Such stacking interactions are also evident in the minimized structure of **P**, but due to the structural constrains of the vicinal 2',3'/3',5' linkages, the overlap is less extensive. Since, in fact, the T_m value is 6 °C greater for compound **Q** than for **P**, or an unmodified hairpin with identical stem and loop sequences (**1**, table **3.3**. figure **3.11**), one can conclude that both the pentaloop as well as the eight A:T Watson-Crick pairs contributes to stabilization of **Q**. The favorable stabilization derived from the branched pentaloop structure in **Q** is also evident when its thermal dissociation (64.1 °C) is compared to that of a model hairpin with identical stem but with a four-unit linear (CCCC) loop (compound **2**, T_m 60.7 °C), the optimum loop length for normal DNA hairpins.[317] It could be speculated that base stacking interactions within the pentaloop provide the structural basis for the gain of thermal stability observed in **Q** relative to **1** and **2**.[317] The slightly higher T_m observed for **N** relative to **O**, and **P** may be partly because of the more favorable purine-purine (rA2'5'→dA) over purine-pyrimidine (rA2'5'→dC)] stacking interaction.[318]

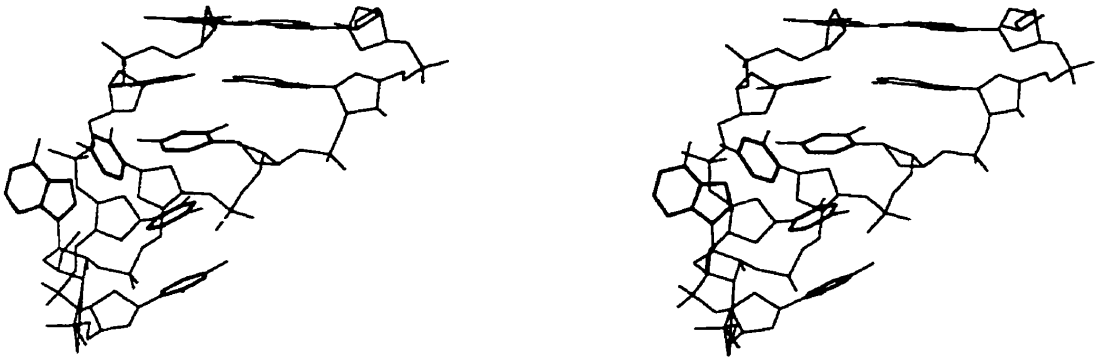


Figure 3.12: Stereoview of an energy-minimized structure of branched hairpin **Q**. Bases in the pentaloop are shown in boldface; only two T:A base pairs (bottom) are shown for clarity. See also figure **3.13**.

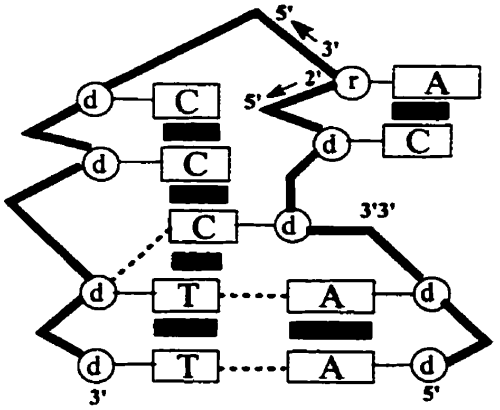


Figure 3.13: Schematic representation of the structural features of stacking interactions of the branched pentaloop structure of hairpin **Q**. Symbols: ribose (r); deoxyribose (d); base stacking (▬); hydrogen bonds (- - -).

3.3.2 Association of Branched Hairpin DNA, N, with dT_{10}

The development of antisense and antigene oligonucleotides as highly specific potential therapeutic agents has stimulated the search for analogs which are more resistant to nucleases[321] and display better stabilities on binding with the target. Most of the reported work has been on modifying the linear nucleic acids with respect to their bases and backbones. Apart from these modifications, several reports of higher order structural modifications have also appeared and have included structures such as hairpins,[322,323] dumbbells,[324] looped,[176,324] and

circular oligonucleotides,[253,255,259,325] to name a few. These structurally modified molecules were seen to have enhanced resistance against degradative nucleases.[322,324] Similar to these structurally modified oligonucleotides, branched single-stranded loops may confer desirable resistance against the same degradative nucleases.[321] Thus preliminary work on the association properties of such branched oligonucleotides was carried out to evaluate their ability to interact with single-stranded target pyrimidine sequences.

Preliminary studies on the interaction of a branched hairpin with single-stranded DNA, namely dT_{10}, with hairpin **N** showed that dT_{10} forms a stable complex with **N** in Tris buffer (pH 7.3, 50 mM $MgCl_2$) as indicated by the presence of two well separated transitions, figure **3.14**. The lower temperature transition (21.5 °C) is nearly coincident with the first transition given by the classical dT_{10}:dA_{10}*dT_{10} triplex (T_m= 19 °C) and corresponds to the process: triplex → dT_{10} + dT_{10}:dA_{10} duplex[272]. The lower temperature transition (21.5 °C) is therefore attributed to dissociation of dT_{10} from the triplex dT_{10}:**N**, whereas the high temperature transition (61.5 °C) correspond to the helix-to-coil transition of duplex **N**, figure **3.11**. This preliminary result demonstrates that a unimolecular branched duplex (hairpin) can interact with a pyrimidine sequence to form a triplex structure. This property may be significant in the development of antisense therapeutic agents that target single stranded RNA via triple helix formation.[326-328]

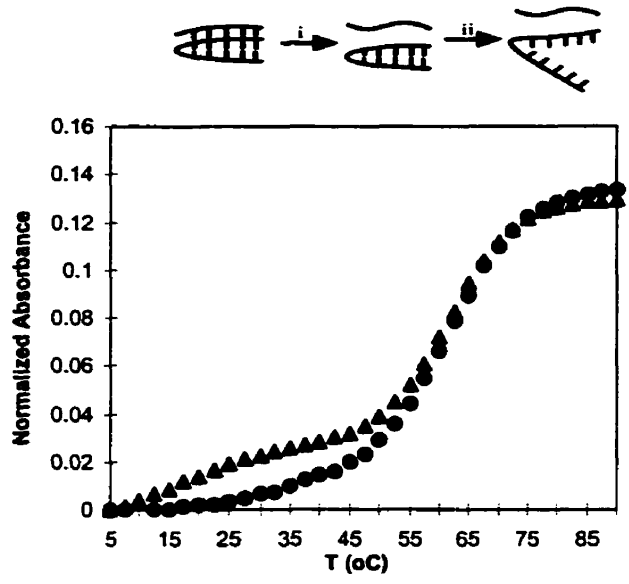


Figure 3.14: Melting curves of N:dT_{10} (▲) and N (●)in 50 mM $MgCl_2$, 10 mM TRIS-HCl buffer, (pH 7.3). The two transitions observed for N:dT_{10} complex correspond to the process i and ii shown in the schematic above.

3.4 FUTURE DIRECTIONS

Parallel-stranded DNA (psDNA) is an interesting and unusual DNA structure. This duplex structure has been shown to form *via* the self-association of hairpins[329] and other intramolecular self-folding structures. Acidic conditions,[330-332] ligand binding,[252] and changes in base sequence, [333,334] are strategies that have been used to induce the formation of parallel DNA. It is envisioned that the entropic advantage offered by the branched oligonucleotides could be exploited to induce parallel DNA formation. If the sequences are chosen properly they can be made to interact with each other by using either the reverse Watson-Crick or the Hoogsteen base pairings.[335] These branched parallel stranded oligonucleotides could then be used in targeting single stranded sequences, thus representing an alternative approach to the antisense strategy.

3.5 CONCLUSIONS

Branched oligonucleotides are capable of forming triple-helix structures by interacting and binding with the appropriate single-stranded targets. This ability was exploited to quantitatively measure the relative stabilities of Hoogsteen versus reverse-Hoogsteen T:A/T triplets. Further extension of this work illustrated the ability of branched nucleic acids induce intramolecular hairpin (duplex) formation. Molecular modeling studies on branched hairpins showed that their loops adopted conformations similar to those previously seen in naturally occurring "super stable" RNA hairpins.[317] Finally, it was demonstrated that branched hairpins interact with pyrimidine-rich oligodeoxynucleotides forming triple helical structures.

4. RESULTS AND DISCUSSION

BIOLOGICAL APPLICATIONS OF bNAs

"Of all sciences that will contribute the most to the understanding of biology, it is chemistry."

Jacob Berzelius

4.1 INTRODUCTON

In 1917 Emil Fisher suggested that chemistry would contribute to the field of genetics, although nothing was known about the molecules responsible for hereditary information flow from generation to generation.[336] Over the years, chemistry has led the way in opening up the field of modern genetic engineering and biotechnology. Part of this success came from the developments in the field of nucleic acids chemistry. Although chemical synthesis of nucleosides was first carried out by Emil Fischer in the early part of the century, for the main purpose of proving the structure of the purine bases, not much happened till Watson and Crick published their work on the structure of the DNA double helix.[8] This, now classic, publication drew the attention of a much wider scientific audience to the importance of nucleic acids because of the biological implications of their structural model; *i.e.* that the DNA base sequence within the cell was all the information needed to carry out the functions of that cell. Development of chemical synthesis of short oligonucleotides of desired sequences then laid the foundation for modern molecular biology.

A survey of the literature clearly shows that the study of nucleic acids has become a broad area spanning many diverse fields of science, and this is undoubtedly due to the developments in the chemical synthesis of oligonucleotides and their analogs. From their initial use in deciphering the genetic code by Khorana and

coworkers,[337] synthetic oligonucleotides have found wide and varied applications in the study of biochemical processes. Recently novel applications for synthetic oligonucleotides have seen to emerge in completely unrelated fields such as computer science.[151,338,339]

The variety of structural features and chemical functional groups found in nucleic acids has enabled them to partake in many biochemical processes. Nucleic acid chemistry has also facilitated the development of detailed understanding of many biologically important processes by providing natural and modified oligonucleotides of desired sequences.

As linear oligonucleotides have been used for investigating the biological processes where they naturally occur, it also makes sense to use the synthetic branched nucleic acids to systematically study those processes where they occur. In particular, synthetic branched nucleic acids should allow a systematic study of RNA splicing and subsequent debranching by the unique and recently discovered debranching enzymes.

4.2 STUDIES ON THE YEAST DEBRANCHING ENZYME*

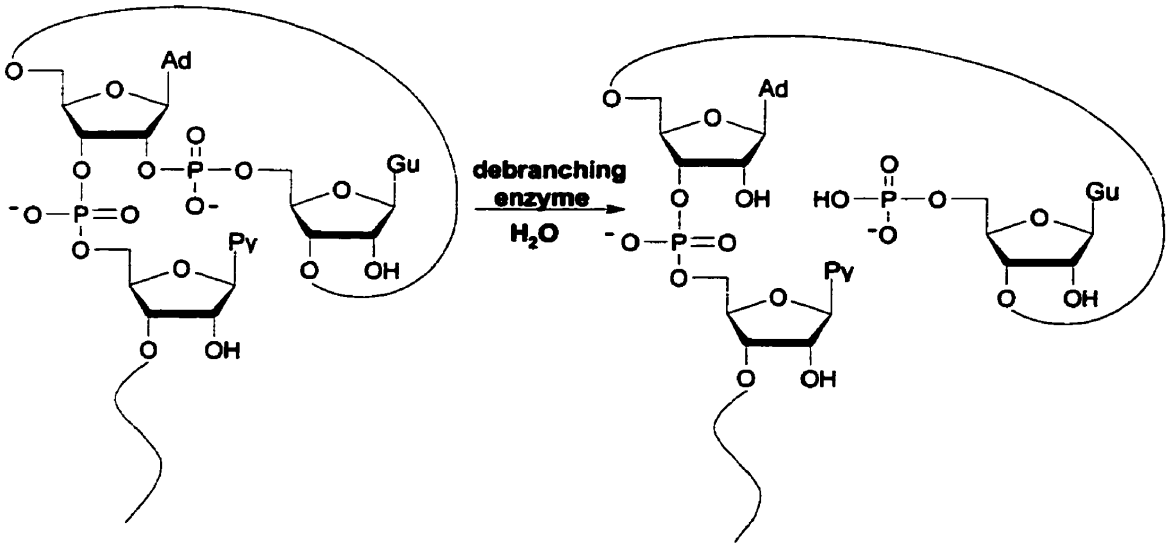
4.2.1 Biological Properties of Debranching Enzymes

A novel enzymatic activity that specifically hydrolyzes the nuclease resistant branch RNA was first identified by Ruskin and Green in 1985 in HeLa cell extracts.[118] Later designated the "RNA lariat debranching enzyme", this polypeptide is one component in the metabolic pathway of intron turnover in

* the debranching enzyme in the context of this thesis refers to the enzymatic activity that carries out the rate-limiting step in the degradation of RNA intron lariats and is different from the enzymatic system that carries out the debranching of glycogen. Components of this glycogen debranching system are also usually referred to as the debranching enzyme.

mammalian cells[118] and in yeast.[119,340] This intron-lariat debranching activity was shown to be separate from the pre-mRNA splicing factors. One of its roles is to selectively cleave the 2',5'-phosphodiester linkages at the branchpoint of excised intron lariats thus converting them into linear molecules, scheme **4.1**. This linearization of intron-lariats then allows for their further degradation during RNA catabolism.

Scheme 4.1



Remarkably, debranching enzymes are not nonspecific 2',5'- or 3',5'-phosphodiesterases since linear 2',5'- or 3',5'-linked oligonucleotides are not hydrolyzed by these enzymes.[118,119,341,342] Figure **4.1** illustrates this specificity of the yeast debranching enzyme (yDBR) for a branched molecule, namely the conversion of ^{32}P-XX($^{2'5'}$Y)$_{3'5'}$X to ^{32}P-XX$_{3'5'}$X and Y. This substrate specificity differentiates the debranching enzymes from other non-specific phosphodiesterases such as snake venom phosphodiesterase[343] and the interferon-inducible 2',5'-phosphodiesterases in mammals.[344-346]

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18

Figure 4.1: Substrate specificity of yeast debranching enzyme. Lane numbers at the bottom with the time of incubation in minutes at the top; o/n = overnight. *Lanes 1-8*: linear $CA^{2'5'}G$; *lane 10*: linear $CA_{3'5'}G$ control; *lanes 11-18*: branched tetramer $CA(^{2'5'}G)_{3'5'}G$. No yDBR was added to samples in lanes 1, 2, 10, 11, and 12.

The gene encoding the yeast debranching enzyme (*dbr1*) was recently identified from the yeast *S. cerevisiae*[119] and purified to *ca.* 90% homogeneity by its over expression in *E. coli.*[200] The *dbr1* gene encodes for a single protein of 406 amino acids, which carries the debranching activity. Although the amino acid sequence shows no similarity to any known protein sequences, a 268 amino acid region of it has been shown to be about 60% similar (12% identical) to the yeast glycogen phosphorylase.[119] Recent database searches for sequence homologs to *dbr1* has led to the identification and cloning of similar genes from *C. elegans* and *S. pombe.*[120] The proteins they encode show 33% (*C. Elegans*) or 40 % (*S. pombe*) sequence homology to the yDBR enzyme, but are completely functional as they could reconstitute debranching activity and normal intron degradation to a mutant *S. cerevisiae* strain lacking the yDBR.

There exists some information as to some of the biochemical requirements of the HeLa cell debranching enzyme,[58,118,340,341,347] however little investigative work had been reported on the yeast debranching enzyme. Using synthetic branched oligonucleotides, Sproat *et al.*[157] reported the initiation of studies with the HeLa

debranching enzyme in 1994, but their results were scarce. Independently, and in collaboration with the Boeke's group, our laboratory conducted studies to probe the substrate specificity of the yeast debranching enzyme.

Initial work of Chapman and Boeke implied a possible relationship between the yeast debranching enzyme and retrotransposon Ty1 mobility.[119] Initial studies on disrupting the yeast debranching enzyme gene (*dbr1*) showed accumulation of intact intron lariats in levels exceeding that of the corresponding mRNAs but little effect on the growth rate of such mutant cells. Thus it was concluded that debranching function was not essential for viability. However more recent work from Boeke's group has shown that a similar mutation in *Schizoachharomyce pombe* had severe growth defects.[120] This suggests that efficient intron RNA degradation in this organism is critical for cell growth and the debranching enzyme plays a crucial role in the lariat degradation pathway.

The yDBR, from *S. cerevisiae*, has been shown to debranch a variety of natural and synthetic branched nucleic acid substrates including self-splicing group II intron-lariats, and small synthetic branched oligonucleotides.[200] Preliminary studies had also shown that small branched oligonucleotides of the type 5'-X-rY[($^{2'-5'}$N)$_{3'-5'}$N] (where X, Y and N are variable nucleotides) were sufficient for studying the biochemical requirements of yDBR.

An area of this thesis work is to further our understanding of the substrate specificity of the yDBR. Debranching assays presented here were generally performed by workers in Dr. Jef Boeke's laboratory (The Johns Hopkins University) and some by the author using synthetic substrates prepared during the course of this work.

Initial work of Hudson[181] had shown that debranching reactions were impossible to follow on unlabelled substrates using HPLC or the commonly used UV-shadowing of analytical polyacrylamide gels. Thus the 5'-ends of synthetic branched oligonucleotides were labeled first with ^{32}P-phosphates in the presence of T4 polynucleotide kinase and the debranching were then followed by autoradiography (see for example **4.1**). However branch molecules lacking the 5'-residues are not substrates of T4 polynucleotide kinase and different approaches are required for following their degradation in the presence of yDBR.

4.2.2 Effect of Nucleotide at the 5'-position on debranching by yDBR

Figure **4.2** shows the debranching of 5'-rNA($^{2'5'}$G)$_{3'5'}$G (N = G, U)* with the expected debranching product linear trimers 5'-NAG-3'. Since the tetramers with a guanosine residue at both the 2'- and 3'- positions were being debranched very quickly, an enzyme preparation of known lower activity was used for debranching reactions, thus allowing to detect any small differences in the rate of hydrolysis between the substrates. The time course of the debranching by yDBR showed that GA($^{2'5'}$G)$_{3'5'}$G was being cleaved slightly more efficiently relative to UA($^{2'5'}$G)$_{3'5'}$G. Similar results were obtained from parallel studies in which AA($^{2'5'}$G)$_{3'5'}$G was debranched more efficiently than CA($^{2'5'}$G)$_{3'5'}$G. These results are qualitative but show that yDBR has only a small preference as to the nature of nucleotide present on the 5'- position of the branchpoint. This notion is supported by the observation that branched oligonucleotides lacking any nucleotides at the 5'-position can undergo efficient debranching, see chapter 2; section 2.3. This again would suggest that the 5'- residue does not impose an

* the branched tetramers 5'-NA($^{2'5'}$G)$_{3'5'}$G (N= A or C) were synthesized by Hudson, but all the debranching reactions were carried out at the same time under the same conditions.

overly strict requirement on recognition and efficient debranching by the yeast debranching enzyme.

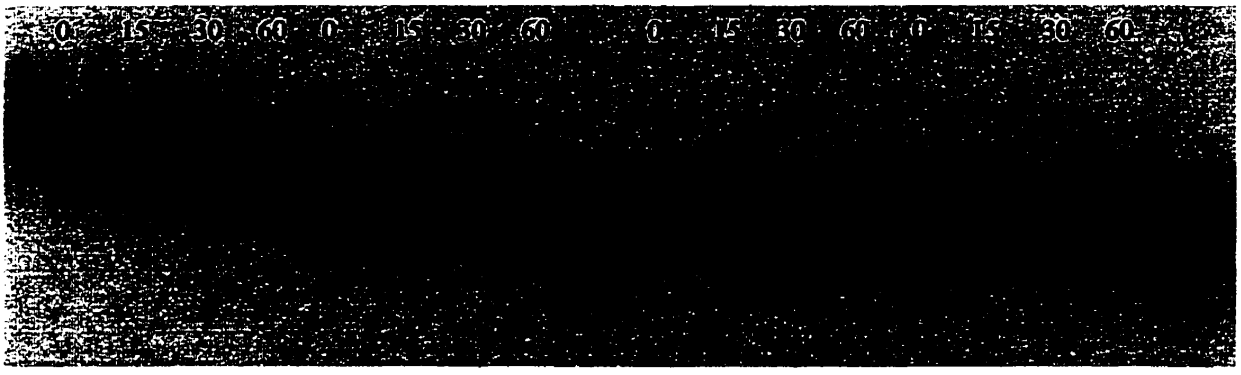


Figure 4.2: Image of autoradiogram for assay of substrate recognition by yeast debranching enzyme done with 5'-rNA($(2'5')$G)$_{3'5'}$G series of branched tetramers where N is the following ribonucleotides: *lanes 1-4*: N = G; *lanes 5-8*: N = A; *lane 9*: linear trimer AAG; *lanes 10-13*: N = U; *lanes 14-17*: N = C; *lane 18*: linear trimer rCAG. Incubation times are at top in minutes. (Note: In comparison to figure 4.3, yDBR preparation of lower activity was used in these incubations.)

4.2.3 Effect of Branchpoint Nucleotide on debranching by yDBR

Some preliminary results from varying the branchpoint nucleotide are presented in figure 4.3. A<u>A</u>($(2'5')$G)$_{3'5'}$G , a substrate having a branchpoint <u>adenosine residue</u>, as found in nature, underwent complete cleavage. On the other hand substrates having the pyrimidine nucleotides at the branchpoint showed little (uridine) or no cleavage (cytosine) under the same conditions. Based on these qualitative results, yDBR can be suggested to have a preference for purines over pyrimidines at the branchpoint.

Figure 4.3: Image of autoradiogram for assay of substrate recognition by yeast debranching enzyme done with 5'rAN($^{2'5'}$G)$_{3'5'}$G series of branched tetramers where N is the following branchpoint ribonucleotides: *lanes 1 and 2*: N = C; *lane 3*: linear trimer ACG; *lanes 4 and 5*: N = U; *lane 6*: linear trimer AUG; *lanes 7 and 8*: N = A; *lane 9*: linear trimer AAG. At the top, + or - indicates presence (+) or the absence (-) of yDBR during the incubation.

4.2.4 Effect of Nucleotides at the 2'- and 3'- positions on debranching by yDBR

Results from debranching assays of tetranucleotides containing identical residues at the 2'- and 3'- position *i.e.*, 5'-XA$^{N}_{N}$ (N= A, G, C, U), indicated that yDBR shows a strong preference for purines at the 2'- and 3'- positions. These results were confirmed by the debranching of the isomeric mixture of branch molecules **E** – **H** (scheme **2.5**, p. 45).[200] Similar results have also been obtained on debranching studies carried out on certain mutant msDNAs.[200] This indicates that branched oligonucleotides with cytosines at both the 2'- and 3'- positions or just the 2'- positions are poor substrates for the yDBR.

From the debranching studies reported here, yDBR was seen to have a general preference for 2'-purines over 2'-pyrimidines. Whether this is due to some aspect of yDBR or due to some structural difference between the substrates remains an open question. It also remains to be seen how the debranching

enzymes from other sources, e.g. HeLa cells, behave in respect to their substrate specificity

From the above discussion it can be seen that the yeast debranching enzyme is able to recognize a wide array of substrates for debranching. Future studies, currently underway, will involve making more radical changes at and around the branch core to get a better understanding of the factors involved in recognition and debranching by yDBR.

4.3 STUDIES ON PRE-mRNA SPLICING

As indicated in introduction (section **1.2**) most, if not all, RNAs are synthesized as precursor molecules which undergo a number of post-transcriptional reactions to generate the final mature, functional form of the RNA. The maturation process starts by separation of cotranscribed RNAs such as the various rRNAs, multimeric tRNAs, and mRNA-tRNA dual function transcripts. This is usually followed by formation of mature 5' and 3' ends by removal of precursor specific sequences. Excision of introns from precursors that contain intron sequences, then takes place in the cytoplasm. For some RNAs, base or sugar modifications, or nucleotide additions are also required before the RNA is fully functional. All of these types of reactions, carried out on a diverse number of RNA molecules, have made RNA maturation, especially the excision of introns, a much studied process. Although pre-mRNA splicing is a common and important feature of gene expression in eukaryotes, the molecular mechanisms by which it occurs is not yet fully understood.[348]

With all the recent work directed towards pre-mRNA splicing, a number of generalities about this reaction have started to emerge. It is now well established

that pre-mRNA splicing is regulated by highly conserved sequences at the exon-intron boundaries[349,350] and the branch site.[351] Whereas in yeast there is a strict requirement of the branchpoint sequence UACUAAC, in mammalian systems mutations at the branchpoint adenosine (A) has been seen to lead to activation of cryptic branchpoints.[113] Recently, other sites, especially in the introns, have also emerged as playing important roles during splicing by binding to a multitude of splicing factors. Some exonic sequences are also known to play important role in selection of 5' and 3' splice sites. Coulter and coworkers[352] have recently used the *in vitro* evolution techniques "SELEX" to identify exonic sequences that enhanced the inclusion of exons containing those sequences into the mature product.

Based on the analysis for complementary sequences and base mismatch mutation studies, most of the small nucleolar RNA (snRNA) have been shown to bind to pre-mRNA at one time or the other during splicing. Some of these RNA-RNA interactions underlie many aspects of substrate recognition, juxtaposition of reaction partners and the catalysis. Accordingly, binding of U1-snRNA to the 5'-splice site has been shown to be the crucial step in committing the pre-mRNA to the splicing pathway.[353,354] Interactions between several splicing factors and the 5'-splice site have been well described. In contrast, recognition of the 3' splice site was poorly understood till very recently. According to the work published by Wu and Green, 3' splice site selection occurs subsequent to the first step, scheme **1.1** (page 8).[355]

As to which 5'- exon ligates with which 3'-exon, Sharp[356] initially proposed a scanning mechanism for selection of the 3' splice site. According to this model 3'-splice site selection is carried out by a snRNP complex diffusing laterally along

the RNA until it comes in contact with a second splice site specific snRNP complex. However, observations from trans-splicing reactions (scheme **1.1**) led to the "jumping" model of 3' splice site selection. In this case, the interaction of the two snRNP complexes occur independently of the RNA between splice sites through a three-dimensional diffusion mechanism. Recent work by Anderson and Moore[357] has lent further support to this three dimensional diffusion mechanism. Their work noted that human spliceosome does not require covalent attachment of a 3' splice site to the branchpoint for exon ligation to take place.

Branchpoint selection was initially thought to be due to an adenosine residue being bulged out of an RNA helix formed by the U2-snRNA:pre-mRNA base pairing, figure **4.4**.[105] The unusual conformation of the bulged out branchpoint adenosine was proposed to represent important signaling information in the splicing pathway. Recently Pascolo and Sérapin[358] has shown this not to be the case. By carrying out base mutation studies, they observed that, in yeast, branchpoint selection occurs in multiple steps and the nature of branchpoint is recognized in the absence of U2-snRNA, during commitment complex formation. It is only afterwards that base pairing with U2-snRNA constrains the adenosine residue into a bulge conformation. Thus it could be speculated that bulging configuration plays a role in the chemistry of the splicing reaction but not in the recognition of branchpoint.

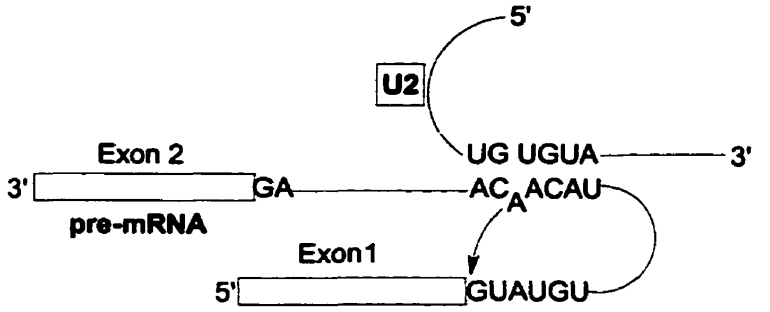


Figure 4.4: U2-snRNA:pre-mRNA interactions showing the "bulged" branchpoint adenosine cleaving exon 1 (5'-splice site).

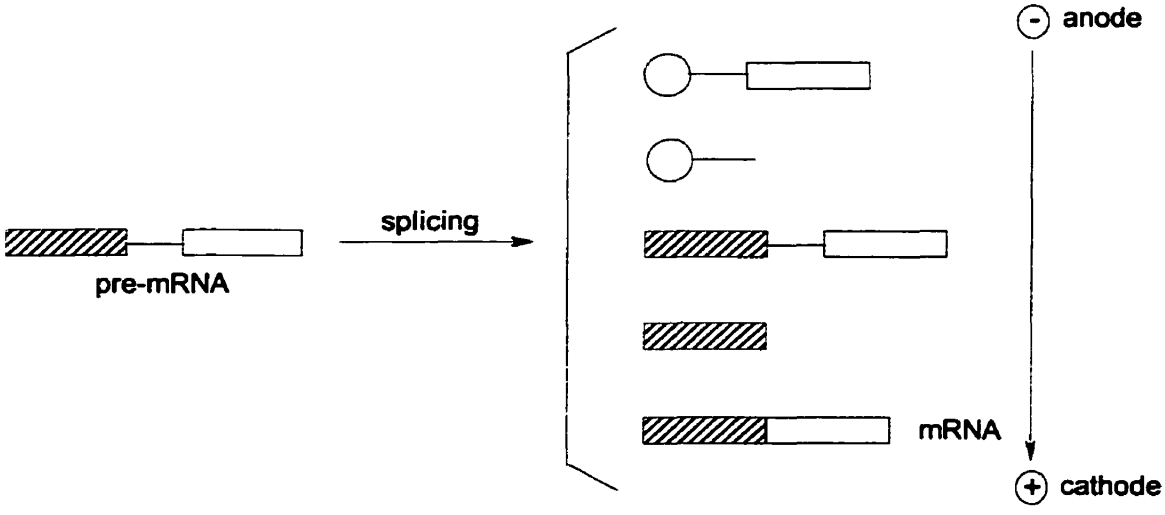
Splicing reaction has been investigated using natural and modified synthetic oligonucleotides. Use of phosphorothioates has made important contributions to the uncovering of the stereochemical course of the splicing reaction.[58,359,360] Linear oligonucleotides have been used for separation and purification of splicing factors. Accordingly Hudson[181], a previous member of our research group, initiated investigation of the recognition of synthetic branched oligonucleotides by a yeast nuclear extract capable of performing *in vitro* pre-mRNA splicing (unpublished results). These preliminary results showed inhibition of splicing by some well-defined branched molecules and suggested sequestering of some speculative branch-recognizing factor(s) by the synthetic branched molecules thus preventing splicing. This merited a further investigation of the importance of sugar, sequence, and branchpoint nucleotide in the recognition of pre-mRNA by branch-recognition factor(s). The splicing reactions using synthetic branched nucleic acids prepared during the present study were carried out Dr. Dorian Anglin in Dr. Jef Friesen's laboratory (U. of Toronto). Some results from these studies are presented below.

4.3.1 *In vitro* pre-mRNA Splicing Assays

In the mid 1980's *in vitro* splicing systems were developed and a general scheme of the pre-mRNA splicing reaction was proposed by a number of laboratories. It

was observed that *in vitro*, splicing goes through two distinctive and separate steps. The first step involves a simultaneous cleavage of the 5'-exon/intron splice site and the branch formation at an adenosine in the intron. The subsequent second step leads to the cleavage at the 3'-intron/exon splice site and ligation of the exons, see Chapter 1, scheme **1.1** (page 8). By allowing the *in vitro* reaction to proceed only for a short time period, initial substrate pre-mRNA, lariat containing intermediates and mature RNA are easily separated and identified by gel electrophoresis and autoradiography. During gel electrophoresis, lariat containing intermediates show abnormal mobility relative to linear molecules of similar size and base composition. Scheme **4.2** illustrates the relative mobility of various intermediates and products present in the reaction mixture of the two exon pre-mRNA splicing used in these studies, shortly after initiation of splicing.

Scheme 4.2



Figures **4.5**, **4.6**, **4.7**, and **4.8** show the results of *in vitro* splicing reactions done in the presence of various concentrations of synthetic branched and linear oligonucleotides. The extent of splicing was measured by quantitating the mature mRNA band as a % of total radioactivity in the lane.

By comparing the effects caused by synthetic "Y-shaped" RNA and DNA molecules it became clear that they both inhibit the formation of the mature mRNA product, figure **4.5**. However oligoribonucleotides (Y-RNA) were seen to lead to inhibition at smaller concentrations as compared to the branched oligodeoxynucleotides (Y-DNA) under similar conditions. Similar results were also seen when the inhibition caused by "V-shaped" (V-RNA) is compared to that caused by "V-shaped" (V-DNA), figure **4.6**. It is known that in the spliceosome there are a number of RNA-binding proteins that have one or more RNA-recognition motifs, and it is conceivable that one or more of these RNA-binding proteins are the putative branchpoint recognition factor(s) (see below). Difference of RNA-binding domains to RNA and DNA could explain the differential inhibition of splicing seen by the synthetic RNA and DNA branched oligonucleotides. If the branch core recognition factor has an RNA binding domain, it maybe demonstrating a lower affinity for DNA, therefore leading to the observed differential in splicing inhibition.

In a recent study Kim *et al.*[361] reported some interesting observations on changing certain ribonucleotide residues in U6 snRNA to deoxynucleotides. They noted that these substitution often led to inhibition of pre-mRNA splicing. Interestingly however, substitution with 2'-O-methyl ribonucleotide derivatives at the same places showed no difference in splicing efficiencies from the wild type U6 snRNA. This suggests that the ribose sugar may be involved in interactions with splicing factors which are unable to recognize and/or bind 2' -deoxynucleotides. Again this supports the view, that in the spliceosome complex, there are certain factors that show a preference for RNA over DNA.

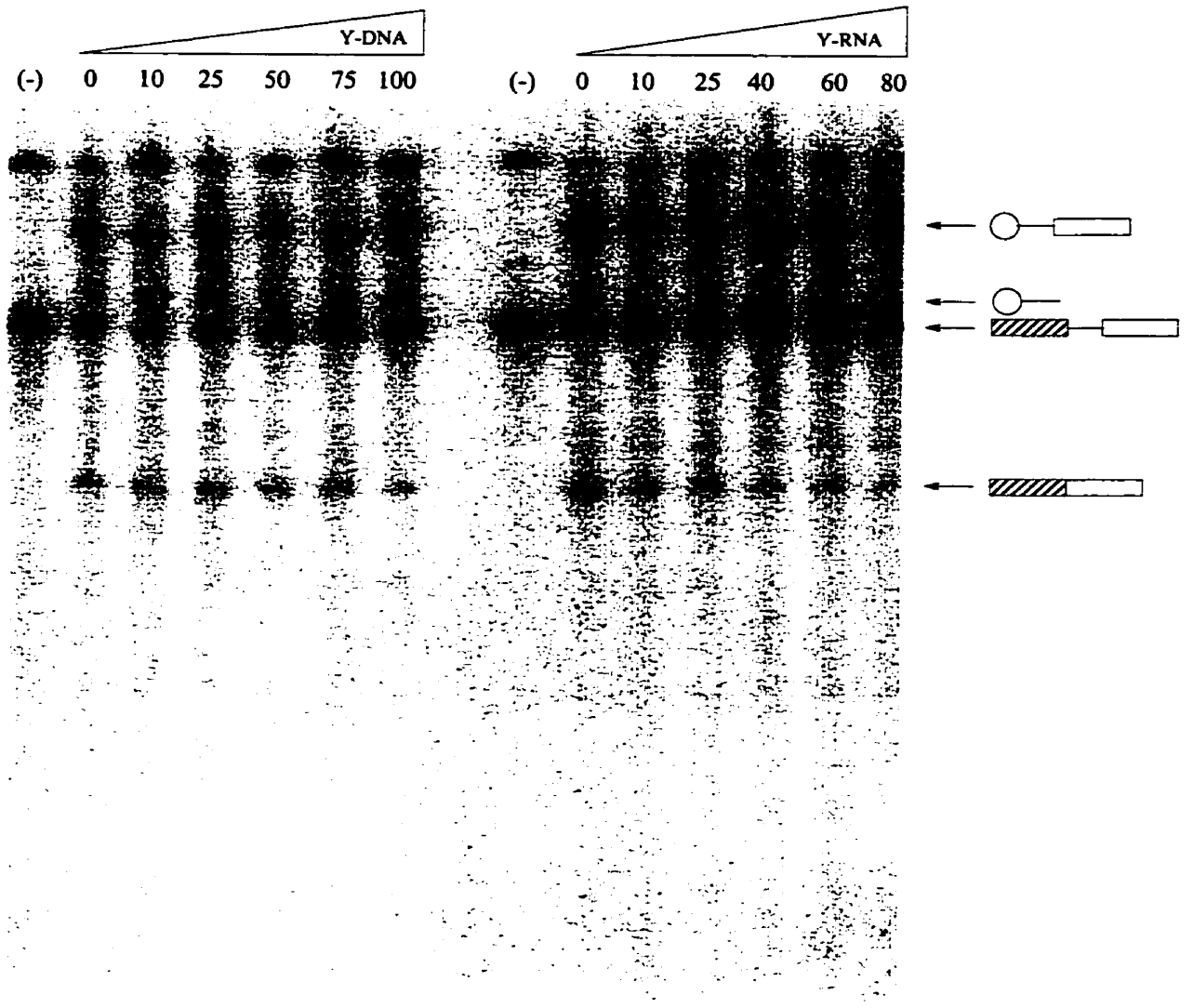


Figure 4.5: Image of autoradiogram of PAGE analysis of *in-vitro* pre-mRNA splicing reactions in the presence of synthetic oligonucleotides. Looking at the effect of Y-type branched synthetic (a) Y-DNA, TACTAA($^{2'5'}$GTATGT)$_{3'5'}$GTATGT; and (b) Y-RNA, UACUAA($^{2'5'}$GUAUGU)$_{3'5'}$GUAUGU oligonucleotides. The concentration of synthetic oligonucleotide in the reaction are given at the top in μM. Lanes marked (-) are the negative controls. As can be seen from the ratio of mature product to pre-mRNA, it takes smaller amounts of Y-RNA to abolish splicing.

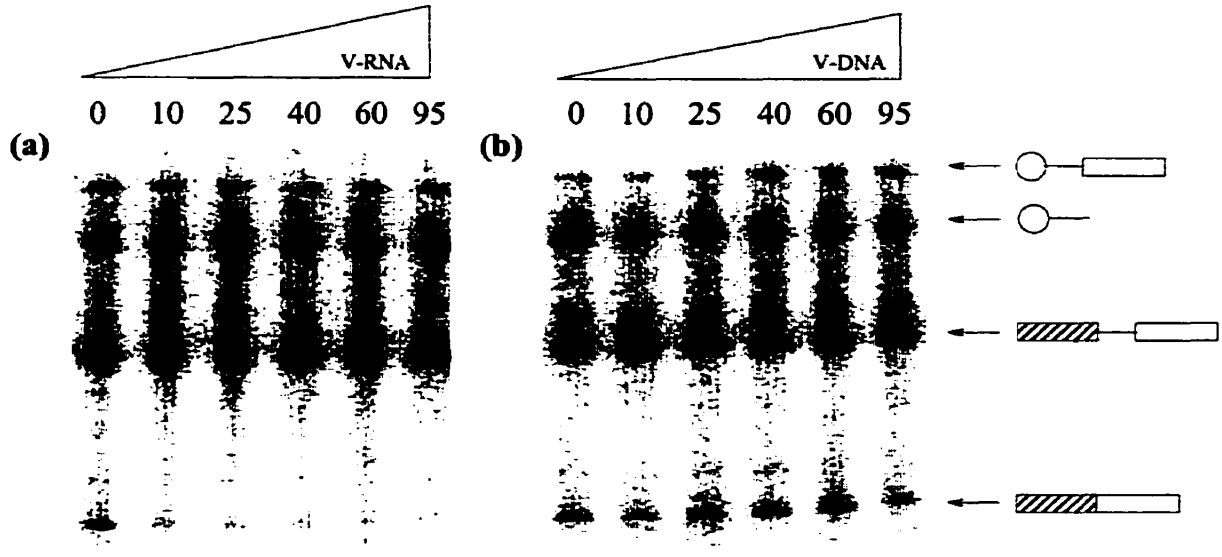


Figure 4.6: Image of autoradiogram of PAGE analysis of *in-vitro* pre-mRNA splicing reactions in the presence of synthetic oligonucleotides. Looking at the effect of V-type branched synthetic (a) V-RNA, A($^{2'5'}$GUAUGU)$_{3'5'}$GUAUGU; and (b) V-DNA, A($^{2'5'}$GTATGT)$_{3'5'}$GTATGT oligonucleotides. The concentration of synthetic oligonucleotide in the reaction are given at the top in μM.

Hudson[181] had reported the requirement for a well-defined branchpoint for splicing inhibition by branched oligothymidylates. Figure **4.7** reports the results of splicing in the presence of branched oligoribonucleotides with or without the 5'-tail. The Y-shaped oligonucleotide is seen to be much better at inhibiting splicing. The Y branched oligonucleotide *completely* prevents the pre-mRNA splicing even at low (10 μM) concentrations. On the other hand only *ca.* 50% reduction in the mature mRNA (with respect to the reaction lacking any branched

NAs) was seen even when the concentration of the V-RNA molecule was in 9 fold excess of the Y-RNA concentration that completely abolished splicing. Similar results were also obtained with V- versus Y-DNA molecules (data not shown). These results are interesting in that the only difference between the two branched oligonucleotides is the presence or absence of the sequence 5' to the branch. It seems that branch recognition elements involves not only the branch core but may be the part of the 5'-sequence as well. Rain and Legrain have recently suggested that the nucleotides proceeding the conserved UACUAAC branch region sequence contributes to the branchpoint recognition process. The differential inhibition shown by "V" and "Y" oligonucleotides could also be due to their slightly different three dimensional structures. The branchpoint residue (A) is known to be involved in intramolecular base-base stacking interactions with the nucleotide at the 2'-position.[318,319] It was noted that the presence of nucleotides on the 5'-position of the branchpoint led to a reduction in the 2'-5' base-stacking interactions. Sund *et al.*[319] speculated that this serves to distort the normal "A" -type helix structure in "Y"-like molecules. As no nucleotides at the 5'-end of "V"-shaped molecules are present no such distortion is expected. Thus the recognition by the branchpoint recognition factors may not only involve base sequences but also the three dimensional structure of the branch core.

Another interesting result appeared to be the inhibition of splicing by *linear* RNA molecules at very high concentrations, figure **4.8.** Hudson had also seen a similar effect, whereby some inhibition of splicing was seen with high concentrations of small synthetic linear oligonucleotides. This may represent nonspecific effects, although more studies are needed to confirm this.

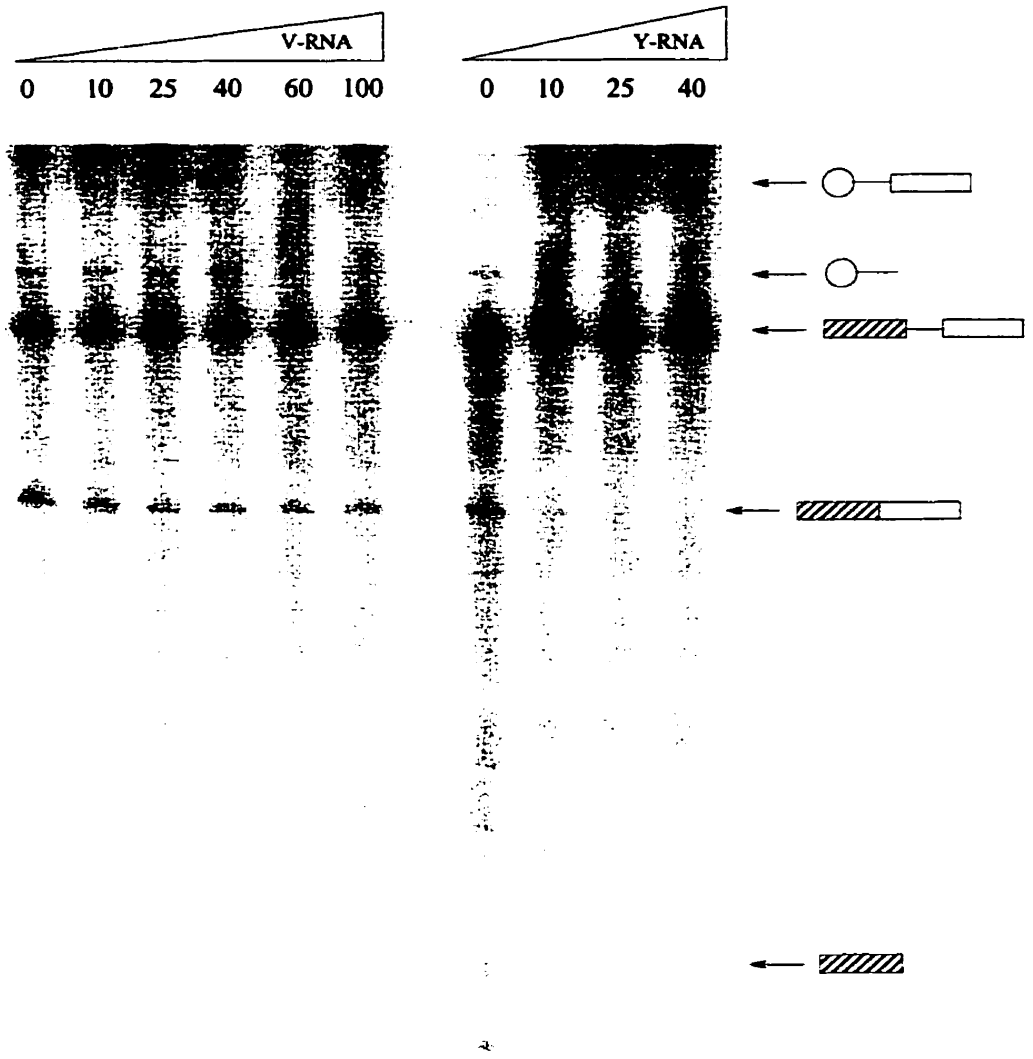


Figure 4.7: Image of autoradiogram of PAGE analysis of *in-vitro* pre-mRNA splicing reactions in the presence of synthetic oligonucleotides. Specifically looking at the effect of V-RNA, A($^{2'5'}$GUAUGU)$_{3'5'}$GUAUGU, and Y-RNA, UACUAA($^{2'5'}$GUAUGU)$_{3'5'}$GUAUGU, on splicing. The concentration of synthetic oligonucleotide in the reaction are given at the top in μM.

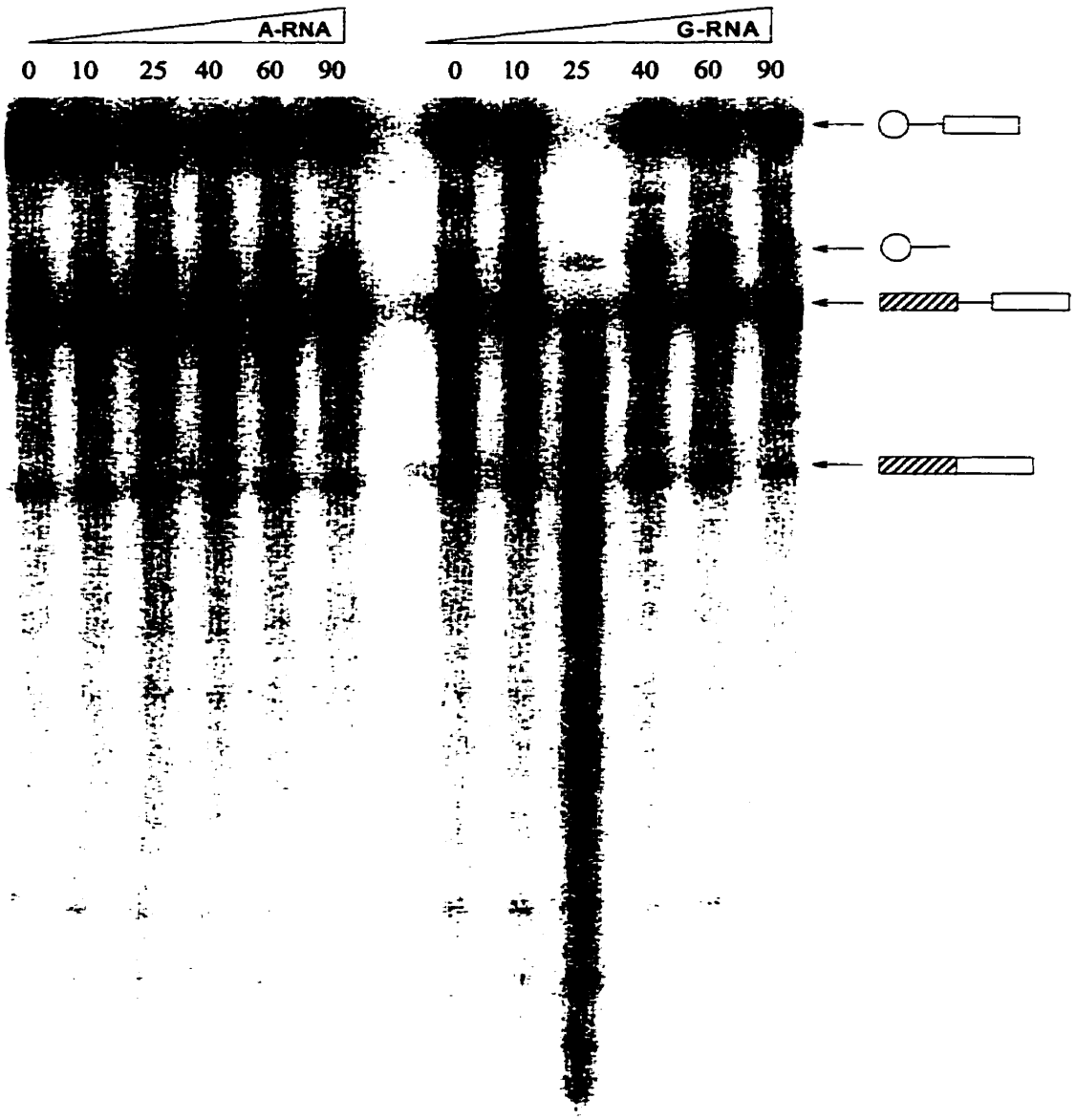


Figure 4.8: Image of autoradiogram of PAGE analysis of *in-vitro* pre-mRNA splicing reactions in the presence of synthetic oligonucleotides. Looking at the effect linear RNAs. Concentration of synthetic oligonucleotides given at the top in μM. A-RNA: AUG GAU UCU GAU AUG UUC UA and G-RNA: AUG GAU UCU GGU AUG UUC UA.

A lot of speculation prevails as to the identity of the branch-recognition trans-acting factor(s), and no definite answers exist. It has been known for some time that U2-snRNP is associated with some auxiliary factors (U2AFs) which facilitate U2-snRNP's interactions with the branchpoint sequence.[362] In particular

U2AF65 has been seen to bind at the polypyrimidine tract near the branchpoint and may influence branchpoint selection[362]. In 1997 Abovich and Rosbash[96] identified a protein factor in the committed-complex that was involved in protein-protein interactions that bridge the 5'- and 3'- splice-site ends of the intron, before U2snRNP addition to the complex in yeast. More recent work has shown the *branchpoint bridging protein* (BBP) to be interacting specifically with the pre-mRNA branchpoint sequence[97]. Based on sequence homology and similar functional domains, Berglund *et al.*[97] have also identified a previously known protein, splicing factor 1 (SF1), in the mammalian spliceosome complex as having branchpoint sequence recognition properties. Whereas the yeast BBP showed strong sequence specificity for the conserved branchpoint sequence, specificity of the mammalian BBP (mBBP) binding was reported to be weaker. Other interesting features included the issue of affinity, which was unusually low as compared with some other RNA binding proteins. This suggests that mBBP interacts with other splicing factors that tether them closer to the branchpoint. On the other hand, it could also be argued that mBBP recognizes fully formed branches and that linear sequence binding is just a residual property. Further studies are required to fully understand the binding of BBP and other similar splicing factors.

Weaker affinity of mBBP to the branchpoint sequence may help in explaining the splicing inhibitions seen with linear synthetic oligonucleotides A-RNA and G-RNA at high concentrations (figure **4.8**). If other branchpoint recognition factors also have low affinities for the linear sequence, higher concentrations will be required for sequestering them with linear oligonucleotides. On the other hand their binding affinity may be higher when the branch has formed and thus the higher inhibition effect of branched oligonucleotides.

These preliminary results indicate that it will be worthwhile to pursue splicing studies with synthetic branch oligonucleotides to further understanding of the factors involved in branchpoint recognition and selection. Furthermore these synthetic oligonucleotides may also prove to be invaluable in separation and purification of some of these branchpoint recognition factors through affinity-selection studies.

4.4 CONCLUSIONS

The results reported here further attest to the applicability of small synthetic branched oligonucleotides in systematically studying the biochemical requirements of the yeast debranching enzyme (yDBR) and *in vitro* pre-mRNA splicing. From the results reported in this thesis, in conjunction with the results produced by former students in the Damha research group, it becomes apparent that yDBR shows a strong preference for 2'-purines over 2'-pyrimidines. Furthermore identity of the nucleotide at the 5'-position of the branchpoint has little influence over the kinetics of yDBR debranching.

During *in vitro* pre-mRNA splicing, RNA oligonucleotides were observed to be better inhibitors of splicing reactions than their DNA counterparts. It was also seen that a fully-formed branch, with extensions in all three direction, *i.e.* Y-shaped, was a better inhibitor than V-shaped molecules. Further studies can be expected to eventually lead to a better understanding of mRNA biosynthesis, from pre-mRNA processing to final metabolism, and synthetic oligonucleotides can be expected to play an important role in the process.

5. CONCLUSION

5. CONTRIBUTION TO KNOWLEDGE

A general procedure for the solid-phase regiospecific synthesis of branched oligonucleotides (bNA) analogues using readily available phosphoramidite reagents has been developed. The method is based on the well established solid-phase synthesis of linear DNA and RNA oligonucleotides. The key feature of this method is use of phosphoramidite chemistry to assemble linear oligonucleotide sequences and sequential removal of the phosphate (β-cyanoethyl or O-methyl) and silyl protecting groups without detaching the nascent oligonucleotide from the solid support. Conversion of the phosphate backbone into the more stable phosphodiester linkages allows for removal of the vicinal 2'-O-*t*-butyldimethylsilyl protecting group without cleavage or isomerization at the branchpoint. This method allows for the formation of branched oligonucleotides with sequences of arbitrary base composition, length and orientation around the branchpoint junction, including a 21-unit branched DNA/RNA chimera, 5'-r-GCGAUCAG[2'd-*CTTCGA*]r-CCUCAGG, similar to the msDNA molecule isolated from *Myxococcus xanthus*. The synthetic method has general applicability for constructing branched oligonucleotides of varied base composition and size as shown by the synthesis of various branched molecules in this study.

Studies to explore structural effects in the use of a branched adenosine as replacement for nucleotide loops in duplex and triplex DNA were also described. The synthetic methodology developed during the course of this study, allowed the synthesis of branched molecules with the potential to form intramolecular

duplexes. Formation of these duplexes was studied *via* UV-thermal melt profiles and showed intramolecular interactions between a range of concentrations. Branched oligonucleotides of the type $rA[^{2',3'}dC_ndA_{10}\text{-}5']_{3',5'}dC_ndT_{10}\text{-}3'$ and $rA[^{2',5'}dC_n3',3'dA_{10}\text{-}5']_{3',5'}dC_nT_{10}\text{-}3'$ form hairpin duplexes with comparable or better thermal stability to one with a natural deoxynucleotide loop. Molecular modeling on some these branched oligonucleotide hairpins showed them to have a similar conformation around the branchpoint as found in lariat introns some "super stable" RNA hairpins. The $5'\text{-}rA(^{[2'3']}dA^{[5'3']}dA_9\text{-}5')_{[3'5']}T_{10}\text{-}3'$ branched hairpin was seen to target a dT_{10} strand through triple helix formation with the same stability as seen for an intermolecular parallel TAT triple helix.

In the presence of dA_{10} and Mg^{2+} buffer, the branched oligomers $rA[^{2',5'}dT_{10}\text{-}3']^{3',5'}dT_{10}\text{-}3'$ and $rA[^{2',3'}dT_{10}\text{-}5']^{3',5'}dT_{10}$ -3' were seen to form triplexes conforming, respectively, to the "antiparallel" (reverse-Hoogsteen) and "parallel" (Hoogsteen) T/AT motifs. The parallel triplex was found to be more stable than the antiparallel triplex, with an average difference in free energy of 0.27 kcal mol^{-1} /TAT base triplet. Most of this difference in stability was found to be due the difference in enthalpy of the two types of TAT triads, with the parallel TAT triads being 1.5 kcal mol^{-1} /TAT base triplet more stable than its counterpart. This is consistent with the general understanding that antiparallel py:pu*py triplexes that are usually hard to observe.

Synthetic branched nucleic acids served as very useful biological probes. From the studies on the yeast debranching enzyme, using small synthetic branched molecule, some useful information regarding the substrate specificity of the enzyme came into light. Preliminary results from the *in vitro* pre-mRNA splicing studies carried out in collaboration with the Friesen group showed inhibition by

synthetic oligonucleotides. A basis for differential inhibition by the various molecules was contemplated, on the basis of currently known and speculated putative splicing factors which might be serving in branch recognition during the splicing process. This work establishes the need and utilization of synthetic branched nucleic acids in probing biologically important processes.

5.2 PUBLICATIONS

As a direct result of the studies described herein, the following publications have recently appeared, or will be published in the future:

(1) M. J. Damha and R. S. Braich, "Synthesis of Branched DNA/RNA Chimera Similar to the msDNA molecule of *Myxococcus xanthus*", *Tetrahedron Lett.*, 1998, 39, 3907-3910.

(2) R. S. Braich and M.J. Damha, "Regiospecific Solid-Phase Synthesis of Branched Oligonucleotides. Effect of Vicinal 2',5' (or 2',3'-) and 3',5'-Phosphodiester Linkages on the Formation of Hairpin DNA", *Bioconjugate Chemistry*, 1997, 8(3), 370-377.

(3) K. Ganeshan, T. Tadey, K. Nam, R. Braich, W. C. Purdy, J. D. Boeke, and M. J. Damha, "Novel Approaches to the Synthesis and Analysis of Branched RNA", *Nucleosides & Nucleotides*, 1995, 14(3-5), 1009-1013.

(4) R. S. Braich, J. Chen and M.J. Damha, "Solid-phase synthesis of small branched oligoribonucleotides for probing the substrate specificity of the yeast debranching enzyme", (see section 2.5), manuscript in preparation.

(5) R. S. Braich, A. Noronha and M. J. Damha "Relative Stabilities of Hoogsteen versus reverse Hoogsteen Triple Helices Containing all T/AT base triplets", (see section 3.2), manuscript in preparation.

Publications resulting from collaborations during the course of this work but not reported in this thesis:

(6) P.M. Macdonald, M.J. Damha, K. Ganeshan, R. Braich, and S.V. Zabarylo, "Phosphorous 31 solid state NMR characterization of oligonucleotides covalently bound to a solid support", *Nucleic Acids Res.*, 1996, 24(15), 2868-2876.

Presentations based on the studies described in this thesis:

(1) R. S. Braich and M. J. Damha, Effect of Vicinal 2',5'-(or 2',3'-) and 3',5'-Phosphodiester Linkages on the Formation of Hairpin DNA, 10th Conversation in Bimolecular Stereodynamics, State University of New York, Albany, NY, USA, 6/1997.

(2) K. Ganeshan, K. Nam, R. H. E. Hudson, T. Tadey, R. Braich, K.B. Chapman, W. Purdy, J. D. Boeke and M. J. Damha, Synthesis and Biological Applications of Branched RNA: Probing Small Nuclear RNA-Branched RNA Interactions, 5th Ontario-Quebec Minisymposium, University of Toronto, Toronto, Ontario, 10/1994.

(3) K. Ganeshan, K. Nam, R. H. E. Hudson, T. Tadey, R. Braich, K.B. Chapman, W. Purdy, J. D. Boeke and M. J. Damha, Synthesis and Biological Applications of Branched RNA: Probing Small Nuclear RNA-Branched RNA Interactions and the Substrate Specificity of the Yeast Debranching Enzyme, Eleventh International Roundtable: Nucleosides, Nucleotides and Their Biological Applications, Katholieke Univesiteit Leuven, Belgium, 09/1994.

6. EXPERIMENTAL

6.1 GENERAL METHODS

6.1.1 Reagents and chemicals

In general, solvents were dried and distilled, before use according to established procedures. Tetrahydrofuran (THF) was distilled from sodium and benzophenone ketyl under nitrogen, as needed. Dichloromethane and hexanes (BDH Toronto, ON) were distilled from P_2O_5 or calcium hydride (BDH) and stored over activated (400 °C) 4Å molecular sieves. Acetonitrile (BDH) was first dried by refluxing over P_2O_5 and then continuously refluxed over, and distilled from, calcium hydride. N,N-Dimethylformamide (BDH) was dried by shaking with KOH, followed by distillation under reduced pressure, and stored over activated (400 °C) 4Å molecular sieves. All traces of acidity, from ethyl acetate, were removed by washing with 5% $NaHCO_3$, and drying over anhydrous sodium sulfate. Methanol was distilled from magnesium. Pyridine (BDH) was first dried by refluxing over KOH, followed by distillation from BaO. Chloroform (BDH) was shaken with water, dried with $CaCl_2$, and distilled from P_2O_5.

Tetra-*n*-butylammonium fluoride (TBAF), 1M solution in THF, triethylamine-trihydrofluoride, 1,3-dicyclohexylcarbodiimide (DCC), 1-(3-dimethylaminopropyl)-3-ethylcarbodiimide hydrochloride (DEC), N-methylimidazole (N-MeIm), (methacryloxypropyl) trimethoxysilane (MAPS) (Aldrich), and DNA synthesis grade tetrazole (Dalton) were used as received.

Other reagents such as glacial acetic acid, acetic anhydride, anhydrous sodium sulfate, sodium chloride, triethylamine, aqueous ammonia, 1,2-dichloroethane,

iodine, trichloroacetic acid obtained from BDH were of analytical grade and used without further purification.

Deuterated solvents, acetone-D6, CD_3CN, $CDCl_3$, and dimethylsulfoxide-D6 were obtained from Isotec, Inc. (Miamisburg).

6.1.2 CHROMATOGRAPHY

Thin-layer chromatography (TLC) was performed on Merck Kieselgel 60 F_{254} plastic- or aluminum-backed (0.2 mm thickness) analytical silica gel sheets (EM Science, Gibbstown, NJ). Nucleosides and derivatives could be visualized on the thin layer plates by illuminating with a UV light source (Mineralite, output *ca.* 254 nm), or for trityl containing compounds, by exposure to HCl vapor.

Regular and flash column chromatography[363] was performed on Merck Kieselgel 60 Silica gel (40-60μ particles, EM Science), using 20-25 g silica gel per g of crude compound.

6.1.3 Instruments

UV Spectra. Ultraviolet spectra were recorded on a Varian-Cary 1 UV/Vis spectrophotometer. For protected nucleoside derivatives, 95% ethanol was the solvent of choice. The thermal denaturation studies were also conducted on the Varian Cary 1 spectrophotometer equipped with a thermostated cell holder.

NMR Spectra. For characterization of new compounds, ^{1}H and ^{31}P nuclear magnetic resonance (NMR) spectra were obtained on a Varian XL-300 or a Varian Unity 500 Spectrometer. Spectra were collected at ambient temperature, and all chemical shifts are reported as downfield from tetramethylsilane (TMS)

for ^{1}H, and from 85% H_3PO_4 (external capillary) for ^{31}P, in parts per million (ppm) scale. Peak assignments were made, in some cases, with the aid of COSY and/or homonuclear decoupling experiments.

FAB-Mass Spectrometry. Fast atom bombardment mass spectra were collected using a Kratos MS25RFA high resolution mass spectrometer with nitrobenzyl alcohol (NBA) matrix.

MALDI-TOF Mass Spectrometry. Matrix assisted laser desorption ionization mass spectroscopy was performed on a Kratos Kompact MALDI III mass spectrometer.

Circular Dichroism Spectroscopy. Circular dichroic spectra were collected on a JASCO J-710 spectropolarimeter. Hellma fused quartz (165-QS) cells were used, and the temperature was controlled by an external constant temperature circulating bath (Neslab RTE-111).

The data were collected on a personal computer with the use of manufacturer supplied software. In general, spectra were recorded between 200 and 350 nm, with a scan speed of 100 nm/min and sampling wavelength of 0.2 nm. For each sample 15 repetitive scans were recorded and averaged. All CD spectra were recorded from oligonucleotide at concentration *ca.* 10 μM each stand in a buffer of 10 mM Tris, 50 mM $MgCl_2$, pH 7.28. Before data acquisition at the appropriate temperature, samples were allowed to equilibrate for at least 10 min.

Before CD calculations, the data were smoothed on a personal computer using the Windows software, J-700 for Windows Standard Analysis ver 1.00 (JASCO Inc).

Appropriately smoothed data were then exported to a spreadsheet package (Excel 6.0/97) for presentation.

6.2 OLIGONUCLEOTIDE SYNTHESIS

6.2.1 Reagents

Unprotected 2'-deoxy and ribonucleosides were obtained from Aldrich or Dalton Chemical Laboratories Inc. (Toronto, ON). Benzoyl chloride, 4-dimethylaminopyridine, levulinic acid (LvOH), and *t*-butyldimethylsilyl chloride (TBDMS-Cl) were purchased from Aldrich. Bis(4-methoxyphenyl)phenylmethyl chloride (dimethoxytrityl chloride), N,N-diisopropyl-2-cyanoethylphosphonamidic chloride, and N,N-diisopropylmethylphosphonamidic chloride were obtained from Dalton. The benzoyl, trityl and silyl-protected nucleosides were prepared according to established procedures.[364]

Nucleoside 3'-phosphoramidites were purchased from Dalton Chemical Laboratories. N^6-benzoyl-5'-O-dimethoxytrityladenosine-2',3'-bis((2-cyanoethyl)-N,N-diisopropylphosphoramidite) **1.1**, was prepared according to the procedure established by Damha and Ogilvie.[152,153] 2'-O-Fpmp protected ribonucleoside phosphoramidites for the synthesis of msDNA were purchased from Cruachem (Dulles, VA) and required a different deprotection protocol, *vide infra*.

For solid-phase synthesis of oligonucleotides, long-chain alkylamine controlled-pore glass support (LCAA-CPG, 500 Å pore size), was obtained from Dalton Laboratories, and derivatized according to the procedure described by Damha.[210]

After derivatization, nucleoside loadings were determined by assaying the dimethoxytrityl cation release from known amounts of derivatized solid support.

Anhydrous acetonitrile prepared by a two step method (*vide supra*) was used for dissolving nucleoside phosphoramidites and intermediate washing steps. Reagents for capping of the undesired reactive sites were prepared as follow: Cap A; 10% (v/v) acetic anhydride, 10% (v/v), 2,4,6 collidine in THF, Cap B; 16% (w/v) N-methylimidazole in THF. For detritylation of the 5'-dimethoxytrityl protecting group a 3% (w/v) solution of trichloroacetic acid in dichloroethane was used. Activation of nucleoside phosphoramidites was achieved with a 0.5 M tetrazole solution in acetonitrile. After coupling of the nucleoside phosphoramidite and capping the undesired reactive sites, the newly synthesized phosphite triester linkage was oxidized, with a 0.1 M iodine in THF/pyridine/water (25:20:2 v/v/v) solution, to a phosphotriester linkage.

6.2.2 Solid Phase Oligonucleotide Synthesis

Solid phase automated oligonucleotide synthesis was carried out on an Applied Biosystems 381A DNA synthesizer (ABI) using a modified 1 μmole scale pulsed-delivery cycle. The major modifications to the standard 1 μmole scale pulsed-delivery cycle, as supplied by ABI, were as follow: (a) *phosphoramidite coupling:* "wait" step was extended to 300 sec for deoxyribonucleoside phosphoramidites and 600 sec for ribonucleoside phosphoramidites. A 900-sec coupling step was used for branching 2',3'-bisphosphoramidite monomers, (b) *capping:* acetylation of the undesired reactive sites was accomplished by a 20-second delivery of "cap A" and "cap B" to column followed by a 30 second "wait" step, (c) *oxidation:* a 30 second delivery of iodine solution to column followed by a 2- second "wait"

step, (d) *detritylation:* a 120-second delivery of TCA/dichloroethane solution to column for DMT deblocking and 180-second delivery for MMT deblocking.

Prior to oligonucleotide chain assembly, the derivatized solid support was treated with the capping reagents, by use of the ABI supplied capping cycle, to block the undesired reactive sites. This step is necessary, since during CPG storage some of the reactive sites may become unmasked and oligonucleotide assembly can take place at these reactive sites, as shown by Temsamani[213], leading to n-1 type oligonucleotides. Nucleoside phosphoramidite reagents were dissolved in anhydrous acetonitrile to a final concentration of 0.1M for 2'-deoxyribo- and 0.15M for ribonucleoside phosphoramidites. The branching nucleoside-2',3'-bisphosphoramidite monomers in the non-regiospecific synthesis were used at 0.03M concentrations. Efficiency of the coupling steps during oligonucleotide synthesis was evaluated by assaying the dimethoxytrityl or monomethoxytrityl cation released during the trichloroacetic acid treatment step.

6.2.3 Solid Phase Synthesis of Branched Oligonucleotides

6.2.3.1 Non Regiospecific Approach

All non-regiospecific syntheses of branched oligonucleotides were conducted on the 1 μmole scale. The branching was carried out by use of the nucleoside-2',3'-bisphosphoramidite monomers, (**1.1**, **2.5** – **2.8**), under the protocol reported by Damha *et al.*[155]

6.2.3.2 Regiospecific Approach

The syntheses were conducted on a 1 μmole scale. After synthesis of 3',5'-linked linear oligomer, the branchpoint monomer was introduced as ribonucleoside-2'(3')-O-TBDMS-3'(2')-phosphoramidite. At this step synthesis was interrupted and the 5'-end of the oligomer was blocked with the capping reagents employing the manufacture supplied capping cycle. After capping the 5'-end of the nucleotide, column was removed from the synthesizer, support was thoroughly washed with anhydrous acetonitrile, and dried by flushing argon through the column.

For regiospecific synthesis, the phosphate protecting group and the 2'(3')-O-TBDMS group from the branchpoint nucleoside must be removed prior to extension from the 2'(3')-OH of the branching monomer. Using dual luer-type disposable syringes a 10 ml solution of 2:4:1 (v/v/v) triethylamine/dioxane/thiophenol for methyl protected phosphate (or a 10 ml solution of 4:6 triethylamine/acetonitrile for β-cyanoethyl protected phosphate), was washed over the beads over the course of 90 minutes. Under these conditions the methyl (β-cyanoethyl) phosphate protecting group was cleaved without any affect on the succinyl linkage (to the support) or the heterocyclic base or base protecting groups. Once the phosphate backbone had been converted into phosphodiester linkages, removal of the 2'(3')-O- TBDMS group was conducted by washing the CPG beads with a 1 ml solution of 1 M TBAF in tetrahydrofuran for no more than 10 minutes. Treatment of 5'-DMT-nucleoside derivatized CPG with 1 M TBAF/THF showed that extended washing led to a loss of trityl release, however within the first 7 minutes a loss less than 5% was seen. Beads were washed extensively with ethanol. The column was replaced on the synthesizer and solid support was extensively washed by delivery of acetonitrile to the

column with alternating "reverse flush" for several minutes. After thorough washing the synthesis was restarted except that coupling of the first phosphoramidite at the 2'(3')-position required the use of 0.3 M phosphoramidite solutions and longer coupling time (30 min).[177,180] The inefficient coupling at the 2'(3')-OH probably reflected the steric hindrance around the 2'(3')-OH and the lower reactivity of a secondary hydroxyl compared to the primary hydroxyl of the 5'-position. Deprotection and purification of oligonucleotides is described in sections 6.3.2 and 6.4.2.

6.2.4 Solid Phase Synthesis of Tetrameric bRNA

6.2.4.1 Synthesis of $XY(^{2'5'}Z)_{3'5'}Z$ Type bRNAs

For yDBR debranching studies, two types of small tetrameric branched oligonucleotides were assembled. For branched tetrameric oligonucleotides of the type $XY(^{2'5'}Z)_{3'5'}Z$, where both the 2'- and 3'- linked nucleotides at the branchpoint are the same, synthesis was carried out by using the bis(phosphoramidite) reagent as outlined in chapter 6, section 6.2.3.1.

6.2.4.2 Regiospecific Synthesis of $AA(^{2'5'}G)_{3'5'}X$ Type bRNAs

Synthesis of regioisomeric small branched oligonucleotides was carried out using the regiospecific approach, chapter 6, section 6.2.3.2, with certain modifications. The necessity of removing the silyl protecting group from the branchpoint during the synthesis excluded the possibility of using CPG solid-support derivatized with the usual 2'(3')-*tert*-butyldimethylsilyl protected monomers. So CPG derivatization was carried out with appropriately base protected, 5'-tritylated nucleosides, and the remaining free hydroxyl group, 2'(3'), acetylated along with

any underivatized sites, just prior to the actual oligonucleotide synthesis. Using this method, loadings in the 28 - 35 μmol/g of solid support were obtained.

After trityl on synthesis of $A^{2'5'}G$ (or $A_{3'5'}X$), the phosphate and the 3'(2')-hydroxyl were deprotected as outlined in regiospecific approach, chapter 6; section 6.2.3.2. The synthesis was restarted at cycle entry, bypassing TCA delivery to column. After coupling the appropriately protected 5'-ribophosphoramidite to the 3'-(2'-) branchpoint hydroxyl, the normal cycle was resumed and nascent branched oligomer extended by one nucleotide at the 5'-position. The branching point being so close to the surface of CPG coupling of the 5'-phosphoramidites were not very high, leading to lower recovery yields of the branched tetrameric bRNAs as compared to the bisphosphoramidite approach. Temsamani has shown that the coupling reaction is not consistent during synthesis of oligonucleotides, and coupling efficiencies for the first few phosphoramidites close to the CPG surface are low probably due to steric hindrance.[213] Deprotection of these branched oligomers is described in section 6.3.2.

6.2.5 Solid Phase Synthesis of msDNA

Due to the RNA/DNA chimeric nature of msDNA, two mutually exclusive 2'-hydroxyl protecting groups for the RNA phosphoramidites were required. After synthesis of the RNA component, synthesis was interrupted and a small portion of the nascent oligomer bearing CPG beads was removed. The 5'-OH was capped through acetylation by running the manufacturer supplied capping cycle on the synthesizer. At this point removal of all ß-cyanoethyl phosphate protecting groups was effected with triethylamine/acetonitrile (10 ml, 4:6 v/v, 90 min rt), chapter 6; section 6.2.3.2. Following the procedure outlined in chapter 6; section

6.2.3.2, the branchpoint 2'-hydroxyl group was then freed by pushing a 1 M TBAF/THF solution through the synthesis column, followed by extensive washing with THF (30 ml) and CH_3CN (30 ml). Once removal of the silyl group from the branchpoint was effected, the column was replaced back on the synthesizer and solid support extensively washed by delivery of acetonitrile to the column with alternating "reverse flush" for several minutes. After thorough washing the synthesis was restarted using 2'-deoxy-3'-DMT-5'-phosphoramidites. As previously reported, to force the branching at the sterically hindered 2'-hydroxyl group, both the concentration and the coupling time of the first 5'-phosphoramidite (dC) were tripled to 0.3 M and 30 min respectively[320]. Deprotection and purification of msDNA is described in section 6.3.2.1.

6.3 DEBLOCKING OF SYNTHETIC OLIGONUCLEOTIDES

6.3.1. General Considerations

The water used in deprotection and subsequent storage of oligonucleotides was double distilled, and sterilized by treatment with diethyl pyrocarbonate (DEP, Aldrich) to form a 2% solution. The DEP treated water was autoclaved at 120°C for 1h using an All American Electric Steam Sterilize (Model No. 25X, Wisconsin Aluminum Co., Inc., Manitowoc, WI). Sodium azide, to a final concentration of *ca* 0.001%, was added to inhibit microbial growth. All plastic- and silanized glassware was also sterilized by autoclaving at 120°C for 1 h in the presence of diethyl pyrocarbonate. Gloves were worn during all handling of deblocked oligonucleotides to minimize enzymatic degradation of free oligonucleotides by nucleases present on skin. Analytical reagent grade ammonia, absolute ethanol and triethylamine-tris(hydrogen fluoride) (TREAT HF, Aldrich) were used as received.

For the cleavage of 2'-O-Fpmp protecting groups, 0.5M Tris-acetate (pH 3.25) and 2M Tris-acetate (pH 6.69) buffers were made in double distilled sterilized water. Buffers were filtered through a 2μm filter and stored at -20°C prior to use.

6.3.2 Deblocking of Synthetic Oligonucleotides

Oligonucleotide containing LCAA-CPG beads were exposed to 3:1 (v/v) 29% aqueous NH_3/EtOH at room temperature for 48-60 hrs to effect the release of the oligonucleotides from solid support and to remove exocyclic amino blocking groups. For ribose containing oligonucleotides, removal of the TBDMSi group required an additional 6-12 hrs of TREAT HF treatment (100 μl TREAT HF for 1 μmole scale synthesis). After TREAT HF treatment, the reaction was quenched with 1 ml of sterile water and liquid conveniently removed *via* lyophilization.

6.3.2.1 Deblocking of msDNA

The msDNA containing LCAA-CPG beads were exposed to 3:1 (v/v) 29% aqueous NH_3/EtOH at room temperature for 48-60 hrs to effect the release of the oligonucleotides from solid support and to remove exocyclic amino blocking groups. After lyophilizing the partially deprotected msDNA, the sample was redissolved in 1 ml of sterile water. The crude yield of 2'-protected material was measured to be *ca.* 90 A_{260} units. From the crude stock solution a 40 A_{260} unit sample was taken and lyophilized to dryness. To the dry crude material was added 250 μl of 0.5 M Tris acetate (pH 3.25) buffer. The slightly turbid solution was kept at room temperature for 18 hrs. At this point 50 μl of 2 M Tris acetate (pH 6.69) buffer were added and solution rigorously mixed for 1 minute. The samples were then centrifuged at 14,000 rpm for 5 minutes and supernatant transferred to a new 1.5 ml micro tube. 900 μl of pre-cooled (to -70 °C) absolute

ethanol were added, with thorough mixing, to the supernatant and samples cooled to -70°C for 35 min. The mixture was then centrifuged for 5 minutes at 14,000 rpm and supernatant carefully removed. The remaining solid pellet was resuspended in 450 µl of cold absolute ethanol and cooled to -70°C. After 35 minutes samples were centrifuged for 5 minutes at 14,000 rpm and supernatant discarded. The residue was lyophilized to dryness and redisolved in 200µl of absolute ethanol and solution lyophilized to dryness, yielding 92.8 A_{260} units of crude material. A similar method was used for the deprotection of RNA oligomer recovered from the msDNA synthesis, before branch induction. The silyl protecting group cleavage from the branchpoint guanosine required an extra 8 hr step with 25 µl of $NEt_3 3HF$ as discussed above. Purification, of msDNA, by 24% polyacrylamide gel electrophoresis (7 M urea)[365] afforded 50 nmol of msDNA (5% yield).

6.4 PURIFICATION OF OLIGONUCLEOTIDES

6.4.1 Reagents

Electrophoresis grade acrylamide, *N,N'*-methylenebisacrylamide (BIS), ammonium persulfate (APS), *N,N,N',N'*-tetramethylethylenediamine (TEMED), bromophenol blue (BPB) and xylene cyanol (XC) were obtained from Bio Rad. Other electrophoresis regents such as boric acid, formamide and disodium ethylene-diaminetetraacetate dihydrate (EDTA, BDH), tris(hydroxymethylaminomethane) (TRIS) (Aldrich) and urea (Caledon) were obtained and used as such.

Tritylated-oligonucleotide purification cartridge (TOPC) resin (Dalton Labs), Sephadex G-25F (Pharmacia) and C_{18} SEP-PAK® cartridges (Waters Associates, MA, USA) were obtained and used according to the suppliers' specifications.

6.4.2 Polyacrylamide Gel Electrophoresis (PAGE)

Purification of crude oligonucleotides was carried out by employing vertical slab gel polyacrylamide gel electrophoresis.[366] Predominantly, detritylated oligonucleotides were purified by denaturing polyacrylamide gel electrophoresis in order to avoid any intra/inter-molecular interaction interference. For the most part, analytical and preparative gel electrophoresis were performed using 24% (w/v) acrylamide (0.75 mm and 1.5 mm thickness, respectively). TBE buffer (89mM TRIS/boric acid, 2.5 mM EDTA, pH 8.3) was used as the running buffer for both denaturing (7M urea) and native (no urea). Denaturing PAGE employed 8:2 deionized formamide/10X TBE as the loading buffer. For use as loading buffer, formamide was deionized by stirring over mixed bed ion-exchange resin AG 501-X8 (Bio Rad). For native polyacrylamide gel electrophoresis 8:2 sucrose (40% (w/v) in water)/10X TBE was employed as loading buffer.

Before loading the samples, gels were prerun for 15-20 minutes at 5 V/cm in order to remove any catalysts and UV-absorbing material.[366] Samples were dissolved in 10µl or 50-100µl of the appropriate loading buffer for analytical and preparative PAGE respectively. After loading the samples, gels were run at 5 V/Cm for 20-30 minutes and then at *ca.* 10 V/cm for the reminder of the running time (3-5 h).

Following electrophoresis, the gels were wrapped with Saran Wrap®, placed over a fluorescent TLC plate and illuminated by a hand held UV lamp (254 nm). Such

illuminated gels were photographed using Polaroid PolaPan® 4 x 5" Instant Sheet Film (#52, medium contrast, ISO 400/27°C; f4.5, 16 second) through a Kodak Wratten gelatin filter (#58 green).

Pure Oligonucleotides were obtained from preparative PAGE by excising the appropriate bands from gels and eluting the oligomers by the crush and soak method. In general, gel pieces were crushed to a fine mesh and incubated at 37°C for 24-36 h. Gel pieces were then removed by filtering through 2.0μm filter cartridges. Such obtained oligonucleotides invariably had low molecular weight impurities and salts and were "desalted" *via* reversed-phase chromatography or size exclusion chromatography, *vide infra*.

6.4.3 Purification of "trityl-on" Oligonucleotides

Oligodeoxynucleotides that were synthesized with the "trityl-on" option contained a 5'-O, or in the case of branched DNA a 3'-O (chapter 2; section 2.5.1), trityl protecting group after standard deprotection. Such oligonucleotides were purified by trityl-selective reversed-phase chromatography on tritylated-oligonucleotide purification cartridge resin (TOPC™) (Dalton Chemical Labs) using supplier's specifications. After elution of pure detritylated oligonucleotides from the TOPC, samples were lyophilized and stored at -20°C. Oligonucleotide quality was then checked by analytical PAGE. Recently this procedure has been extended to purification of 5'-O-trityl-on oligoribonucleotides.[367]

6.4.4 Desalting of Oligonucleotides

6.4.4.1 Size Exclusion Chromatography

The SEC matrix, Sephadex G-25 fine, was hydrated overnight and autoclaved (120 °C, 1h). SEC was run in sterile, 10 ml disposable syringe barrels (Becton

Dickinson & Co., Fraklin Lakes, NJ) plugged with silanized glass wool (Chromatographic Specialties Inc., Brockville, ON). Autoclaved water was used to hydrate the SEC matrix, pack the column and elute the oligonucleotides. The eluted oligonucleotides were collected in *ca.* 1 ml fractions and the amount of oligonucleotides in each was quantitated by UV absorbance spectrophotometery and stored at -20 °C prior to further use. Fractions which showed the presence of salts were pooled together and further desalted by reversed-phase chromatography on C_{18} SEP-PAK® cartridges.

6.4.4.2 Reversed-Phase Chromatography

For desalting by revered phase chromatography, sample obtained from preparative PAGE was dissolved in 250 µl of sterile water. SEP-PAK® cartridge was attached to a 10 ml sterile syringe and flushed with methanol (5 ml, HPLC grade) followed by methanol/water (1:1, 5 ml) and sterile water (10 ml). Oligonucleotide sample was loaded at a rate of 1-2 drop per 10 seconds collecting the eluent. The cartridge was then flushed with sterile water (10 ml). Finally the oligonucleotide was eluted with 50% methanol (1 ml) followed by methanol (1 ml). Pure desalted oligonucleotides usually eluted in the first fraction.

On rare occasions when the above method gave unsatisfactory desalting, a second reverse-phase chromatography following the method of Damha and Ogilvie[365] was carried out. After flushing the SEP-PAK® cartridge with methanol (10 ml, HPLC grade) and sterile water (10 ml), oligonucleotide sample from above, dissolved in 50 mM aqueous triethylammonium acetate (TEAA, 1 ml) was loaded on the cartridge collecting eluent (A). The cartridge was then flushed with 50 mM TEAA (3 ml) collecting eluent (B). Pure desalted oligomer was then eluted from the cartridge with 100 mM TEAA/methanol (1:1, 10 ml) collecting five 0.5

to 1.5 ml fractions. The oligomer usually eluted in the first 2 to 3 TEAA/methanol fractions. These fractions were lyophilized, combined in 1 ml of sterile water and amount of oligonucleotide quantitated by UV absorbance. When fractions A and B showed absorption maxima at 260 nm, these fractions were combined, lyophilized and repurified on SEP-PAK® cartridge.

6.5 OLIGONUCLEOTIDE CHARACTERIZATION

6.5.1 General

Enzymes for the characterization of oligonucleotides were purchased from Boehringer Mannheim (Quebec). Incubation buffers were prepared using sterilized deionized water, filtered through a sterile 0.2 μm-pore filter (Acrodisc®, Gelman Sciences Inc., Rexdale, ON), and stored at -20 °C prior to use.

6.5.2 Enzymatic Hydrolysis of Oligonucleotides

Alkaline phosphatase (AP), from calf intestine, was obtained as a suspension in 3.2 M $(NH_4)SO_4$; 1 mM $MgCl_2$; 0.1 mM $ZnCl_2$ (pH 7.0). Calf spleen phosphodiesterase (CSPDE) was obtained as a suspension in 3.2 M $(NH_4)SO_4$ (pH 6.0). Lyophilized Nuclease P1 (NP1), from *Penicillium citrinum*, was dissolved in 30 mM NH_4OAc (pH 5.3) to a concentration of 1 mg/ml (300 units/ml). Snake venom phosphodiesterase (SVPDE), from *Crotalus durissus*, was obtained as a solution in 50% (v/v) glycerol (pH 6.0).

The CSPDE incubations were performed in 0.5 M $(NH_4)SO_4$ (pH 6.0). NP1/AP incubations took place in 0.1 M TRIS-HCl; 1 mM $ZnCl_4$ (pH 7.2). SVPDE/AP incubations were carried out inn 50 mM TRIS-HCl; 10 mM $MgCl_2$ (pH 8).

Typically, 0.3 A_{260} units of oligonucleotides were dissolved in the appropriate amount of buffer and incubations carried out at 37 °C for 4 hours after addition of the appropriate enzyme. Digestions with CSPDE were carried out in 10 µl of buffer with 0.004 U (2 µl) of enzyme. NP1/AP digests were carried out in µl of buffer employing 0.6 U (2 µl) of NP1 and 9 U (1 µl) of AP. SVPDE digestions studies required µl of buffer to which were added 0.002 U (1 µl) of SVPDE and 9 U (1 µl) of AP. After the incubation period, digested samples were lyophilized and redissolved in 12 - 15 µl of sterile water before analysis by HPLC.

6.5.3 High Performance Liquid Chromatography

HPLC analyses of the crude oligonucleotides and the enzyme digestions of pure oligonucleotides were performed on a Waters 480 instrument equipped with dual 501 pumps, UK6 injector, and a 480 Tunable UV detector, with the gradient being controlled by a 600E gradient controller and solvent delivery system. Analysis under the reversed-phase conditions were carried out at 254 nm using a reversed-phase Whatman Partisil ODS-2 (10 µm, 4.6 x 250 mm, Chromatographic Specialties); mobile phase solvent A: 20 mM KH_2PO_4 (pH 5.5); solvent B, methanol, gradient 0 - 50% solvent B in 25 minutes, flow rate 1.5 ml/min, 30 °C. Analysis under the anion exchange were also performed at 254 nm with a Water Protein Pak DEAD 5PW column (7.5 x 75 mm); mobile phase: water; solvent A, 1 M $NaClO_4$, gradient 0 - 20% solvent A in 30 minutes.

Prior to analysis, samples were centrifuged at 14,000 x g for 10 - 15 seconds. Typically injections of 10 µL (0.1- 0.3 A260 units) were made and found to give good resolution for analysis. Peak areas and previously reported extinction coefficients values at 254 nm were used to calculate relative base compositions of enzymatically hydrolyzed samples.

6.5.4 Capillary Electrophoresis

The capillary electrophoresis system was constructed at McGill[368] and used a Bertan Model 230R power supply (Bertan Assoc., Hicksville, NY, USA) and an Isco CV^4 detector (Isco, Lincoln, NE, USA). The output of the power supply was connected to the buffer reservoir *via* platinum electrodes (Bioanalytical Systems, West Lafayette, IN, USA). Capillary electrophoresis were performed using the protocol of Hjerten[369] as modified by Tadey[368]. The capillary was first conditioned using a 15 min wash with 1 N NaOH, followed by a 5 min rinse with water. The internal surface of the capillary was then derivatized with (methacryloxypropyl) trimethoxysilane (MAPS) by drawing a solution containing 30 µL MAPS in 1 ml of acetic acid-water (1:1 v/v) into the capillary. After 1 hour, solution containing 9 % (w/v) acrylamide, 1% (v/v) TEMED and 1% (W/V) APS in 10 ml 1x TBE was introduced into the capillary *via* a syringe. Where otherwise noted, after polymerization (1 - 2 hrs), the capillary was cut for a separation length of 35 cm, and a total capillary length of 55 cm.

Samples were analyzed under an applied voltage of 15 kV. The samples were injected electrophoretically by dipping the cathodic end of the capillary into oligomer solutions of 0.1-0.5 A_{260} U/mL of deionized water, and applying a voltage of 9 kV. The injection times were varied from 5 sec to 3 min depending on sample concentration. The running buffer contained 0.9% (w/v) polyacrylamide in 1x TBE. Electrophoretograms were collected on a personal computer using the Waters Maxima 820 (Millipore, Milford, MA, USA) chromatography software, interfaced through a Waters System Interface Module. The data was imported into Grams/32® Spectral Notebase™ data analysis program (Galactic Ind. Corp., ver 4.01) for further analysis and presentation.

6.6 HYBRIDIZATION STUDIES

6.6.1 UV-Thermal Melt Studies

Thermal denaturation studies ("melt curves") were performed on a Varian-Cary 1 UV/Vis spectrophotometer (Varian, Victoria, Australia) equipped with a thermostated Peltier type cell holder, interfaced with a personal computer. Hellma QS-1.000 - 104 cells fitted with friction fit Teflon stoppers to prevent evaporation of buffers were used to hold the samples. Data was collected on a personal computer, using the software provided by Varian (Cary13e ver 3.00, Varian, Mulgrave, Australia). Amount of oligonucleotides was calculated applying the nearest neighbor approximation to the UV-absorbance of single-strands[370]. Normally, fixed mole ratios of appropriate strands were mixed together, dried and re-dissolved in appropriate buffer for hybridization studies. Usually solutions were *ca.* 1.08 μM in each strand. Hybridization buffers were 50 mM $MgCl_2$, 10 mM Tris, pH = 7.3 or 1 M NaCl, 10 mM Na_2PO_4, pH = 7.0, in both cases the pH was adjusted with HCl. Complexed oligomers were heated to 85 °C for 10-20 min and then cooled slowly to the starting temperature over a period of 50 - 60 min. The absorbance at 260 nm (or 284 nm) was measured at 0.5 °C intervals while the temperature was ramped at a rate of 0.5 °C/min. Each sample was run at least twice at the wavelength of interest. T_m's were calculated, using the Easytherm 1.0 program provided by Varian, from the first derivative plots of absorbance versus temperature, or by using the plot method of Breslauer[371]. Tm's calculated by both methods were within 0.5 °C of each other. Data were also transferred to a spreadsheet package (Excel 6.0) for further analysis and presentation. For comparison purposes, the curves are presented in normalized absorbance, normalized absorbance = $(A_t)/(A_{final} - A_{initial})$, where A_t is the absorbance at temperature t. This type of plot allows for comparison on a

more equal bases than the method of Kibler-Herzog[372], $(A_t - A_{initial})/(A_{final} - A_{initial})$, which gives the change in absorbance scaled between 0 ($A_t = A_{initial}$) and 1 ($A_t = A_{final}$) and doesn't differentiate between plots of differing hyperchromicity. Hyperchromicity (%H) values of the transitions are reported as percent increase in absorbance at 260 nm with respect to the final absorbance $\{(A_{final} - A_{initial})/A_{final}\}$ as described by Puglisi and Tinoco[370].

6.6.1.1 Calculation of Thermodynamic Data

Thermodynamic parameters of complexation were extracted from the melt curves, via reciprocal melt temperature vs log (concentration) plots assuming a two-state (all-or-none) model, i.e., that the monophonic helix-coil transition of the $\underline{3}$:dA_{10} and $\underline{5}$:dA_{10} triplexes corresponds to the equilibrium reaction:

$$[\text{branch oligomer: } dA_{10}\,] \rightarrow \text{branch oligomer} + dA_{10}$$

Thermodynamic data presented in table **3.2** were calculated from plots of reciprocal melting temperature, $1/T_m$, versus the natural logarithm of total strand concentration, $\ln C_T$ (1-100 μM concentration range), according to

$$1/T_m = (R/\Delta H^\circ)\ln C_T + (\Delta S^\circ - R \ln 4)/\Delta H^\circ \qquad \text{(eq. 1)}$$

The slope of this plot can be used to calculate ΔH°, and the intercept permits calculation of ΔS°. These two thermodynamic parameters can then be used to calculate ΔG° by application of the Gibb's free energy equation,

$\Delta G^\circ = \Delta H^\circ - T\Delta S^\circ$.

Thermodynamic parameters were also obtained from the alpha plot method of Breslauer.[307] With the transition studied being bimolecular, the molar fraction a is related to the changes in enthalpy (ΔH°) and entropy (ΔS°) by the relationship,[307]

$$K = 2/C_T(1-\alpha)^2 = \exp(-\Delta H^\circ/RT + \Delta^\circ S/R) \qquad \text{(eq. 2)}$$

where K is the equilibrium constant of the reaction, α is the fraction of strands in the triplex state (obtained from the melt curve), R is the gas constant, and C_T the total strand concentration (i.e., conc. of **1** + conc. target) present in the UV cuvette. K values were then calculated at each temperature (eq. 1, with $0.15 < \alpha < 0.85$) and the thermodynamic parameters calculated from the lnK vs $1/T_m$ plot, which increased linearly with increasing $1/T_m$ values. The standard enthalpy change is related to the slope of the curve ($-\Delta H^\circ/R$) and the entropy term to the y-intercept ($\Delta S^\circ/R$). The change in free energy of association, or ΔG°, was calculated at 25 °C. These values were in excellent agreement with the ΔG° values derived from eq. 1 above. All plots were analyzed by linear regression.

6.6.2 Job Plots

The stoichiometry of interactions between **M** and dA_{10} were determined by using the continuous variation method of Job[281]. Titrations were performed in Hellma QS-1.000 - 104 cells fitted with friction fit Teflon stoppers. Absorbance readings were carried out at 10 °C on a Varian-Cary 1 UV/Vis spectrophotometer (Varian, Victoria, Australia) equipped with a thermostated Peltier type cell holder. After each addition of complement strand the system was allowed to equilibrate at 10 °C for 15 minutes before taking the absorbance reading. The stoichiometry studies were conducted in 50 mM $MgCl_2$; 10 mM Tris (pH = 7.3) or potassium buffer.

6.6.3 Molecular Modeling

These studies were performed in the AMBER force field developed for nucleic acids (HyperChem 3.0, Hypercube). The cutoff function (switched: inner = 10 Å and outer = 14 Å) and the RMS gradient was kept constant at 10^{-5} kcal/mol Å. The branched pentaloops (3'-dCC5',2'-rA-3',5'dCC-3' and 5'-dCC3',2'-rA-3',5'dCC-3' for **P** and **Q**, respectively) were first minimized individually. These were then capped to A/T duplexes built in the classical B-form prior to further minimization. The most energetically favored structure of **Q** is shown in figure **3.13**.

6.7 MOLECULAR BIOLOGY

6.7.1 Substrate Specificity of yDBR

Preliminary investigation on the substrate specificity of yDBR using synthetic branched oligonucleotides was carried out by Dr. K. Nam and Prof. J. Boeke of The Johns Hopkins University (Baltimore, MD). Some of the these yDBR debranching assays were also carried out by J. Trembley in the same lab.

6.7.2 Debranching of $AA(^{2'5'}G)_{3'5'}C$ and msDNA

In order to confirm the presence of branch core in the regiospecifically synthesized branch oligonucleotides, they were subjected to hydrolysis by the yeast debranching enzyme (generously provided by Dr. J. Boeke, The Johns Hopkins University). The debranching studies were carried out by the author in Dr. M. Parniak's lab (McGill AIDS center) with the assistance of Dr. I. Burkner.

6.7.2.1 5'-End Labeling with ^{32}P

The branched samples were 5'-end labeled with T4 polynucleotide kinase (MBI Fermentas Inc., Flamborough, ON) using [γ-^{32}p]ATP (Amershem). Briefly, the labeling was carried out in the 10x *reaction buffer:* 500 mM Tris-HCl (pH 7.6);

100 mM $MgCl_2$; 50 mM DTT; 1 mM spermidine, and 1 mM EDTA provided with the enzyme. Approximately 0.05 A_{260} units of lyophilized oligonucleotide was brought up in 5 μl of 10x buffer, *ca.* 20 μCi (2 μl) of [γ-^{32}p]ATP, 10 U (1 μl) of T4 polynucleotide kinase (supplied in a buffer of 20 mM Tris-HCl (pH 7.5); 25 mM KCl; 0.1 mM EDTA, 2 mM DTT and 50% glycerol with balance (42 μl) water to a total volume of 50 μl.

The reaction mixture was incubated at room temperature for 45 minutes after which the reaction mixtures were lyophilized and taken up in 5 μl of loading buffer containing bromophenol blue and xylene cyanol dyes in 80% formamide. The labeled oligonucleotides were then repurified by PAGE (20%, 8 M urea). The labeled oligonucleotides were extracted from the gel into water and desalted by size exclusion chromatography using Sephadex G-25. Thus purified branched oligonucleotides were subjected to the debranching enzyme under the following conditions: *ca.* 1-10 fmol labeled molecule in a total volume of 50 μl containing 20 mM HEPES/KOH (pH 7.6), 125 mM KCl, 0.5 mM $MgCl_2$, 1 mM DTT, and 10% glycerol at 30 °C for 4 hours. The reaction was stopped by lyophilizing to dryness. Mixture was then redissolved in 5 μl of loading buffer, analyzed by PAGE (20%, 8 M urea), and visualized by autoradiography (24 hrs).

6.7.3 Pre-mRNA Splicing Studies

All pre-mRNA splicing studies were conducted by Dr. Dorian Anglin in the lab of Prof. J. Friesen at Hospital for Sick Kids (Toronto, ON).

6.8 MONOMER PREPARATION FOR SYNTHESIS OF OLIGONUCLEOTIDES

6.8.1 Convergent Approach

N^6-benzoyl-5'-O-dimethoxytrityladenosine-2',3'-O-bis(N,N-diisopropyl)-2-cyanoethylphosphoramidite): 1.1

This branching synthon was prepared according to the literature procedure.[153]

N^2-isobutryl-5'-O-monomethoxytritylriboguanosine-2',3'-O-bis(N,N-diisopropyl)-2-cyanoethylphosphoramidite): 2.5

To a stirred THF (5 ml) solution of N^4-benzoyl-5'-O-monomethoxytritylguanosine **2.1** (0.43 mmol, 268 mg), and diisopropylethylamine (3.8 mmol, 745 µl) was added dropwise N,N-diisopropyl-2-cyanoe*thylphosphonamidic chloride (2.58 mmol, 576 µl) at room temperature. A precipitate appeared after 15 minutes. After stirring for 20 hrs at room temperature, TLC analysis (ether/CH_2Cl_2 1:2) showed complete consumption of starting material. Ethyl acetate (50 ml, pre washed with 5 % $NaHCO_3$) was added and solution washed with saturated brine (5 x 50 ml). The organic layer was dried over anhydrous Na_2SO_4 and solvent removed under vacuum to afford the crude product as an oil. Initial purification by silica gel column chromatography, elution with 50:44:6 CH_2Cl_2/hexane/triethylamine, still gave the purified product as an oil. The material was then dissolved in CH_2Cl_2 (20 ml) and precipitated from cold hexanes (200 ml) to afford a white foam in 35% yield (150 mg). TLC: ether/CH_2Cl_2 1:1, R_f = 0.20, 0.38, 0.47. UV (95% EtOH) λmax = 216, 234, 275 nm, λmin = 223, 246. ^{31}P-NMR ($CDCl_3$)

5'-O-dimethoxytrityluridine-2',3'-O-bis(N,N-diisopropyl)-2-cyanoethylphosphoramidite): 2.6

To a stirred THF (5 ml) solution of 5'-O-dimethoxytrityluridine **2.2** (1.82 mmol, 995 mg), 4-(dimethylamino) pyridine (0.39 mmol, 44 mg), and diisopropylethylamine (12.9 mmol, 2.48 ml) was added dropwise N,N-diisopropyl-2-cyanothylphosphonamidic chloride (4.4 mmol, 980 μl) at room temperature. A precipitate appeared after 10 minutes. After being stirred for 4 hrs, ethyl acetate (50 ml, pre washed with 5 % $NaHCO_3$) was added and reaction mixture worked up as for **2.5**. The crude product was purified by silica gel chromatography by elution with 50:44:6 CH_2Cl_2/hexane/triethylamine to obtain a white foam. The pure product was obtained in 58% yield (1 g). TLC: ether/CH_2Cl_2 1:1, R_f = 0.29, 0.35. UV (95% EtOH) λmax = 232, 263 nm, λmin = 224, 253. ^{31}P-NMR ($CDCl_3$)

N^4-benzoyl-5'-O-dimethoxytritylcytosine-2',3'-O-bis(N,N-diisopropyl)-2-cyanoethylphosphoramidite): 2.7

To a stirred THF (5 ml) solution of N^4-benzoyl-5'-O-dimethoxytritylcytosine **2.3** (1.54 mmol, 1 g), 4-(dimethylamino) pyridine (0.33 mmol, 44 mg), and diisopropylethylamine (12 mmol, 2.3 ml) was added dropwise N,N-diisopropyl-2-cyanothylphosphonamidic chloride (3.7 mmol, 825 μl) at room temperature. A precipitate appeared after 10 minutes. After being stirred for 4 hrs, ethyl acetate (50 ml, pre washed with 5 % $NaHCO_3$) was added and the reaction mixture worked up as for **2.5**, to afford a slightly yellow foam in 80% yield (1.3 g). TLC: ether/CH_2Cl_2 1:1, R_f = 0.45, 0.53. UV (95% EtOH) λmax = 232, 256 nm, λmin = 248. ^{31}P-NMR ($CDCl_3$)

N^2-benzoyl-O^6-*p*-nitrophenylethyl-5'-O-monomethoxytritylriboguanosine-2',3'-O-bis(N,N-diisopropyl)-2-cyanoethylphosphoramidite): 2.8

To a stirred anhydrous THF (3 ml) solution of pure N^2-benzoyl-O^6-*p*-nitrophenylethyl-5'-O-monomethoxytritylriboguanosine **2.4** (0.35 mmol, 290 mg) and diisopropylethylamine (3.25 mmol, 0.566 ml) was added, dropwise over a period of 90 sec, N,N-diisopropyl-2-cyanoethylphosphonamidic chloride (0.91 mmol, 203 μl) at room temperature. The solution turned slightly yellow and a precipitate appeared after 10 minutes. After stirring for 20 hrs, the reaction mixture was worked up and as for **2.5**. Compound **2.8** was purified as a slightly yellow foam in 73% yield (310 mg) by silica gel column chromatography (eluent: CH_2Cl_2/hexanes/triethylamine, 50:44:6). TLC: dichloromethane/ether 1:1, R_f = 0.67, 0.5. UV(EtOH) λmax = 234, 276 nm, λmin= 248 nm.

6.8.2 Regiospecific Approach

6.8.2.1 Synthesis of 5'-OH monomers

N^6-benzoyl-2',3'-di-*t*-butyldimethylsilyladenosine: 2.16

To a sample of N^6-benzoyl-5'-O-dimethoxytrityl-2',3'-di-*t*-butyldimethylsilyladenosine, **2.12**, (3.4 mmol, 3.1g) was added trifluoroacetic acid (25 ml, 3% in CH_2Cl_2). After stirring for 10 minutes at 0 °C, TLC analysis (5% MeOH in CH_2Cl_2) showed the complete consumption of the starting material. Reaction was neutralized by addition of 20 ml of 5 % $NaHCO_3$. After extraction with 5 % $NaHCO_3$, (20 ml), the organic layer was dried over anhydrous Na_2SO_4 and evaporated under vacuum to yield a white foam. The crude product was taken up in 25 ml of CH_2Cl_2 and precipitated in cold hexanes (450 ml) to afford

2.16 as a white powder in 65% yield (1.3 g). TLC: 5% MeOH/CH_2Cl_2, R_f = 0.64. UV (95 % EtOH) λmax = 280 nm, λmin = 256 nm.

N^2-isobutryl-2',3'-di-*t*-butyldimethylsilylguanosine: 2.17

To a solution of N^2-isobutryl-5'-O-dimethoxytrityl-2',3'-di-*t*-butyldimethylsilylguanosine, **2.13**, (2.83 mmol, 2.5 g) in dichloromethane (5 ml) was added trifluoroacetic acid (20 ml, 3% in dichloromethane). After stirring for 10 minutes at 0 °C, reaction was poured into 5% $NaHCO_3$ (50 ml). After extraction, the organic layer was dried over anhydrous Na_2SO_4 and evaporated under vacuum to yield a bright orange foam. The crude product was purified by silica gel chromatography by elution with a gradient of 0 - 5 % methanol in dichloromethane. The pure product was obtained as a white foam in 61% yield (977 mg). TLC: EtOAc/CH_2Cl_2 1:2, R_f = 0.82. UV (95% EtOH) λmax = 270 nm, λmin = 242 nm.

2',3'-di-*t*-butyldimethylsilyluridine: 2.18

To a solution of 5'-O-dimethoxytrityl-2',3'-di-*t*-butyldimethylsilyuridine, **2.14**, (2.58 mmol, 2 g) in dichloromethane (5ml) was added trifluoroacetic acid (20 ml, 3% in dichloromethane). After stirring for 10 minutes at 0 °C, reaction mixture was poured into 5% $NaHCO_3$ (50 ml). After extraction, the organic layer was dried over anhydrous Na_2SO_4 and evaporated under vacuum to yield a bright orange foam. The crude product was purified by silica gel chromatography by elution with a gradient of 0 - 5 % methanol in dichloromethane. The pure product was obtained as a white foam in 81% yield (977 mg).

N^4-benzoyl-2',3'-di-*t*-butyldimethylsilylcytosine: 2.19

To a solution of N^4-benzoyl-5'-O-monomethoxytrityl-2',3'-di-*t*-butyldimethylsilylcytosine, **2.15**, (1.9 mmol, 1.6 mg) in dichloromethane (5 ml)

was added trifluoroacetic acid (20 ml, 3% in dichloromethane). After stirring for 15 minutes at 0 °C, TLC analysis (5% MeOH in CH_2Cl_2) showed detritylation to be nearly complete. Reaction was neutralized by addition of 20 ml of 5 % $NaHCO_3$. After extraction, the organic layer was dried over anhydrous Na_2SO_4 and evaporated under vacuum to yield a yellow oil. The crude product was purified by silica gel chromatography by elution with a gradient of 0 -2 % methanol in dichloromethane. The pure product was obtained as a white foam in 68% yield (750 mg). TLC: 5% MeOH/CH_2Cl_2, R_f = 0.61. UV (95% EtOH) λmax = 257 nm, λmin = 249 nm.

6.8.2.2 Synthesis of 5'^{p}X phosphoramidites

N^6-benzoyl-2',3'-O-di-*t*-butyldimethylsilyladenosine-5'-O-N,N-diisopropylamino-2-cyanoethylphosphoramidite: 2.20

To a stirred THF (5 ml) solution of N^6-benzoyl-2',3'-O-di-*t*-butyldimethylsilyladenosine **2.16** (2.6 mmol, 1.5 g), 4-(dimethylamino)pyridine (0.16 mmol, 21 mg), and diisopropylethylamine (6.2 mmol, 1.1 ml) was added dropwise N,N-diisopropyl-2-cyanoethylphosphonamidic chloride (1.8 mmol, 400 µl). After 5 minutes a white precipitate appeared. After 4 hours of stirring, TLC analysis (ether/CH_2Cl_2 1:1) showed complete consumption of starting material. The reaction mixture was poured into ethyl acetate (100 ml, prewashed with 5% $NaHCO_3$). The solution was washed with saturated brine (5 x 100 ml). The organic phase was dried (anhydrous Na_2SO_4) and the solvent removed under vacuum to afford the title compound **2.20** as a white foam in 96% yield (2 g). TLC: ether/CH2Cl2 1:1, R_f = 0.50, 0.70. UV(95% EtOH) λmax = 273 nm, λmin = 252 nm. ^{31}P-NMR (acetone-D6, 500 MHz) 148.5, 148.72.

N^2-isobutryl-2',3'-O-di-*t*-butyldimethylsilylguanosine-5'-O-N,N-diisopropylamino-2-cyanoethylphosphoramidite: 2.21

To a stirred THF (5 ml) solution of N^2-isobutryl-2',3'-O-di-*t*-butyldimethylsilylguanosine **2.17** (1.18 mmol, 685 mg) and triethylamine (11.8 mmol, 2 ml) was added dropwise N,N-diisopropyl-2-cyanoethylphosphonamidic chloride (3.54 mmol, 687 µl). After 10 minutes a white precipitate appeared. After being stirred for 24 hours, the reaction mixture was worked up as for **2.20** affording the crude material as an oil. The crude material was purified by silica gel chromatography by elution with 50:44:6 CH_2Cl_2/hexanes/triethylamine. After evaporation of the pooled fractions, the title compound **2.21** was obtained as a white foam in 53% yield (500 mg). TLC: ether/CH_2Cl_2 1:1, $R_f = 0.65$. UV(95% EtOH) $\lambda max = 276$ nm, $\lambda min = 249$ nm. ^{31}P-NMR (acetone-D6, 500 MHz) 149.2, 149.7.

2',3'-O-di-*t*-butyldimethylsilyluridine-5'-O-N,N-diisopropylamino-2-cyanoethylphosphoramidite : 2.22

To a stirred THF (5 ml) solution of 2',3'-O-di-*t*-butyldimethylsilyluridine **2.18** (1.67 mmol, 750 mg), 4-(dimethylamino)pyridine (0.170 mmol, 21 mg), and diisopropylethylamine (6.4 mmol, 1.1 ml) was added dropwise N,N-diisopropyl-2-cyanoethylphosphonamidic chloride (1.85 mmol, 410 µl). After 5 minutes a white precipitate appeared. After stirring for 4 hours, ethyl acetate (25 ml, prewashed with 5% $NaHCO_3$) was added and the reaction mixture worked up as for **2.20**. The title compound **2.22** was obtained as a white foam in 75% yield (850 mg). TLC: ether/CH_2Cl_2 1:1, $R_f = 0.51, 0.55$. UV (95% EtOH) $\lambda max = 257$ nm, $\lambda min = 250$ nm. ^{31}P-NMR (acetone-D6, 500 MHz) 149.05, 149.43.

N^4-benzoyl-2',3'-O-di-*t*-butyldimethylsilylcytosine-5'-O-N,N-diisopropylamino-2-cyanoethylphosphoramidite: 2.23

To a stirred THF (2.5 ml) solution of N^4-benzoyl-2',3'-O-di-*t*-butyldimethylsilylcytosine **2.19** (0.69 mmol, 400 mg), 4-(dimethylamino)pyridine (0.076 mmol, 10 mg), and diisopropylethylamine (2.8 mmol, 500 µl) was added dropwise N,N-diisopropyl-2-cyanoethylphosphonamidic chloride (0.83 mmol, 160 µl). After 15 minutes a white precipitate appeared. After stirring for 4 hours, ethyl acetate (25 ml, prewashed with 5% $NaHCO_3$) was added and the reaction mixture worked up as for **2.20**. The title compound **2.23** was obtained as a white foam in 90% yield (500 mg). TLC: ether/CH_2Cl_2 1:1, R_f = 0.50. UV (95% EtOH) λmax = 257 nm, λmin = 250 nm. ^{31}P-NMR (acetone-D6, 500 MHz) 149.1, 149.5.

REFERENCES

1) Avery, T. O.; Macleod, C.; McCarty, M. *J. Exp. Med.* **1944**, *79*, 137.

2) Moskovits, M. *Science and Society: The John C. Polanyi Nobel Laureates Lectures*; House of Anansi Press Ltd.: Concord, Ontario, Canada, 1995.

3) Gierer, A.; Schramm, G. *Nature* **1956**, *177*, 702.

4) Loeb, T.; Zinder, N. D. *Proc. Natl. Acad. Sci. U.S.A.* **1961**, *47*, 282-289.

5) Limbach, O. A.; Crain, P. F.; McCloskey, J. A. *Nucleic Acids Res.* **1994**, *22*, 2183.

6) Westheimer, F. H. *Acc. Chem. Res.* **1968**, *1*, 70.

7) Westheimer, F. H. *Science* **1987**, *235*, 1173.

8) Watson, J. D.; Crick, F. H. C. *Nature* **1953**, *171*, 737.

9) Moran, S.; Ren, R. X.-F.; Kool, E. T. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 10506.

10) Cech, T. R.; Zaug, A. J.; Grabowsky, P. J. *Cell* **1981**, *27*, 487.

11) Kruger, K.; Grabowski, P. J.; Zaug, A. J.; Sands, J.; Gottschling, D. E.; Cech, T. R. *Cell* **1982**, *31*, 147.

12) Zaug, A. J.; Cech, T. R. *Science* **1986**, *231*, 470.

13) Guerrier-Takada, C.; Gardiner, K.; Marsh, T.; Pace, N.; Altman, S. *Cell* **1983**, *35*, 849.

14) Altman, S. *Angew. Chem. Int. Ed. Engl.* **1990**, *29*, 749.

15) Blum, B.; Stum, N. R.; Simpson, A. M.; Simpson, L. *Cell* **1991**, *65*, 543.

16) Noller, H. F.; Hoffarth, V.; Zimniak, L. *Science* **1992**, *256*, 1416.

17) Piccirilli, J. A.; McConnell, T. S.; Zaug, A. J.; Noller, H. F.; Cech, T. R. *Science* **1992**, *256*, 1420.

18) Young, L. S. *Science* **1991**, *252*, 542-546.

19) Huang, F.; Yarus, M. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 8965.

20) Huang, F.; Yarus, M. *Biochemistry* **1997**, *36*, 6557-6563.

21) Illangasekare, M.; Sanchez, G.; Nickles, T.; Yarus, M. *Science* **1995**, *267*, 643.

22) Robertson, D. L.; Joyce, G. F. *Nature* **1990**, *344*, 467.

23) Prudent, J. R.; Uno, T.; Schultz, P. G. *Science* **1994**, *264*, 1924.

24) Pace, N. R.; Marsh, T. L. *Orig Life* **1985**, *16*, 97.

25) Sharp, P. A. *Cell* **1985**, *42*, 397.

26) Lewin, R. *Science* **1986**, *231*, 545.

27) Cech, T. R. *Proc. Natl. Acad. Sci. U.S.A.* **1986**, *83*, 4360.

28) Gilbert, W. *Nature* **1986**, *319*, 618.

29) Joyce, G. F. *Nature* **1989**, *338*, 217.

30) Baltimore, D. *Nature* **1970**, *226*, 1209.

31) Temin, H.; Mitzutani, S. *Nature* **1970**, *2206*, 1211.

32) Lampson, B. C.; Sun, J.; Hsu, M. Y.; Vallejo-Ramirez, J.; Inouye, S.; Inouye, M. *Science* **1989**, *243*, 1033.

33) Chow, L. T.; Gelinas, R. E.; Broker, T. R.; Roberts, R. J. *Cell* **1977**, *12*, 1.

34) Lewin, B. *Genes II*; John Wiley & Sons: Toronto, ON, Canada, 1985.

35) Leverette, R. D.; Andrews, M. T.; Maxwell, E. S. *Cell* **1992**, *71*, 398.

36) Tycowski, K. T.; Shu, M. D.; Steitz, J. A. *Genes Dev.* **1993**, *7*, 1176.

37) Bachellerie, J. P.; Nicoloso, M.; Qu, L. H.; Michot, B.; Caizergues-Ferrer, M.; Cavaillé, J.; Renalier, M. H. *Biochem. Cell Biol.* **1995**, *73*, 835.

38) Kiss-lászló, Z.; Henry, Y.; Bachellerie, J. P.; Caizergues-Ferrer, M.; Kiss, T. *Cell* **1996**, *85*, 1077.

39) Nicoloso, M.; Qu, L. H.; Michot, B.; Bachellerie, J. P. *J. Mol. Bio.* **1996**, *260*, 178.

40) Tycowski, K. T.; Shu, M. D.; Steitz, J. A. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 14480.

41) Bachellerie, J. P.; Michot, B.; Nicoloco, M.; Balakin, A.; Ni, J.; Fournier, M. J. *Trends Biochem. Sci.* **1995**, *20*, 261.

42) Cavaillé, J.; Nicoloso, M.; Bachellerie, J. P. *NaturE* **1996**, *383*, 732.

43) Ni, J.; Tien, A. L.; Fournier, M. J. *Cell* **1995**, *89*, 569.

44) Ganot, P.; Bortolin, M. L.; Kiss, T. *Cell* **1997**, *89*, 799.

45) Ho, Y.; Kim, S. J.; Waring, R. B. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 8994.

46) Lee, M. P.; Feinberg, A. P. *Cancer Res.* **1997**, *57*, 3131.

47) Goodison, S.; Yoshida, K.; Sugino, T.; Woodman, A.; Gorham, H.; Bolodeoku, J.; Kaufmann, M.; Tarin, D. *Cancer Res.* **1997**, *57*, 3140.

48) Schubert, E. L.; String, L. C.; Hansen, M. F. *Hum. Genet.* **1997**, *100*, 557.

49) Nevins, J. R. *J. Mol. Bio.* **1979**, *130*, 493.

50) Khoury, G.; Gruss, P.; Dhar, R.; Lai, C. J. *Cell* **1979**, *18*, 85.

51) Gruss, P.; Lai, C. J.; Dhar, R.; Khoury, G. *Proc. Natl. Acad. Sci. U.S.A.* **1979**, *76*, 4317.

52) Gruss, P.; Khoury, G. *Nature* **1980**, *286*, 634.

53) Gething, M. J.; Sambrook, J. *Nature* **1981**, *293*, 620.

54) Treisman, R.; Novak, U.; Favaloro, J.; Kamen, R. *Nature* **1981**, *292*, 595.

55) Fu, L.; Suen, C. K.; Waseem, A.; White, K. N. *Biochem. Mol. Biol. Int.* **1997**, *42*, 329.

56) Padgett, R. A.; Konarska, M. M.; Grabowski, P. J.; Hardy, S. F.; Sharp, P. A. *Science* **1984**, *225*, 898.

57) Zillmann, M.; Gorovsky, M. A.; Phizicky, E. M. *Mol. Cell Bio.* **1991**, *11*, 5410.

58) Maschhoff, K. L.; Padgett, R. A. *Nucleic Acids Res.* **1993**, *21*, 5456.

59) Rajagopal, J.; Doudna, J. A.; Szostak, J. W. *Science* **1989**, *244*, 692.

60) McSwiggen, J. A.; Cech, T. R. *Science* **1989**, *244*, 679.

61) Choquet, Y.; Scheneider, M.; Dron, M.; Bennoun, P. *EMBO J* **1987**, *6*, 2185.

62) Koller, B.; Fromm, H.; Galun, E.; Edelman, M. *Cell* **1987**, *48*, 111.

63) Zaita, N.; Torozawa, K.; Shinozaki, K.; Sugiura, M. *FEBS Lett.* **1987**, *210*, 153.

64) Breathnach, R.; Chambon, P. *Annu. Rev. Biochem.* **1981**, *50*, 349.

65) Breathnanch, R.; Benoist, C.; O'Hare, K.; Gannon, F.; Chambon, P. *Proc. Natl. Acad. Sci. U.S.A.* **1978**, *75*, 4853.

66) Laird, P. W. *Trends in Genet.* **1989**, *5*, 204.

67) Cocquerelle, C.; Daubersies, P.; Majerus, M. A.; Kerckaert, J. P.; Bailleul, B. *EMBO J.* **1992**, *11*, 1095.

68) Cocquerelle, C.; Mascrez, B.; Hetuin, D.; Bailleul, B. *FASEB J* **1993**, *7*, 155.

69) Capel, B.; Swain, A.; Nicolis, S.; Walter, M.; Koopman, P.; Goodfello, P.; Lovell-Badge, P. *Cell* **1993**, *73*, 1019.

70) Wang, K. S.; Choo, Q. L.; Weiner, A. J.; Ou, J. H.; Najarian, R. C.; Thayer, R. M.; Mullenbach, G. T.; Denniston, K. J.; Gerin, J. L.; Houghton, M. *Nature* **1986**, *323*, 508.

71) Schindewolf, C.; Domdey, H. *Nucleic Acids Res.* **1995**, *23*, 1133.

72) Gross, H. J.; Domdey, H.; Lossow, C.; Jank, P.; Raba, M.; Albert, H.; Sänger, H. L. *Nature* **1978**, *273*, 203.

73) Sogo, J. M.; Schneider, I. R. *Virology* **1982**, *117*, 401.

74) Branch, A. D.; Levine, B. J.; Robertson, H. D. *Semin. Virol.* **1990**, *1*, 143.

75) Waring, R. B.; Scazzocchio, C.; Brown, T. A.; Davies, R. W. *J. Mol. Bio.* **1983**, *167*, 595.

76) Michel, F.; Dujon, B. *EMBO J.* **1983**, *2*, 33.

77) Michel, F.; Umesono, K.; Ozeki, H. *Gene* **1989**, *82*, 5.

78) Van Der Veen, R.; Arnberg, A. C.; Van der Horst, G.; Bonen, L.; Tabak, H. F.; Petrillo, L. A.; Tabor, J. H.; Jarrell, K. A.; Cheng, H. L. *Gene* **1989**, *82*, 213.

79) Kuck, U.; Godehardt, I.; Schmidt, U. *Nucleic Acids Res.* **1990**, *18*, 2691.

80) Cech, T. R. *Sci. Am.* **1986**, 64.

81) Fredenwey, D.; Keller, W. *Cell* **1985**, *42*, 355.

82) Grabowski, P. J.; Seiler, S. R.; Sharp, P. A. *Cell* **1985**, *42*, 345.

83) Brody, E.; Abelson, J. *Science* **1985**, *228*, 963.

84) Bringmann, P.; Rinke, J.; Appel, B.; Reuter, R.; Luhrmann, R. *EMBO J.* **1983**, *2*, 1129.

85) Rosbash, M.; Séraphin, B. *Trends Bio. Sci.* **1991**, *16*, 187.

86) Reich, C. I.; VanHoy, R. W.; Porter, G. L.; Wise, J. A. *Cell* **1992**, *69*, 1159.

87) Séraphin, B.; Kandels-Lewis, S. *Cell* **1993**, *73*, 803.

88) Parker, R.; Muhlrad, D.; Deshler, J. O.; Taylor, N.; Rosi, J. J. I *Gene Regulation: Biology of Antisense RNA and DNA*; Erikson, R. P. and Izant, J. G., Eds.; Raven Press, Ltd.: New York, NY USA, 1992, pp 55.

89) Wu, J.; Manley, J. L. *Genes Dev.* **1989**, *3*, 1553.

90) Zhuang, Y.; Weiner, A. M. *Genes Dev.* **1989**, *3*, 1545.

91) Chabot, B.; Black, D. L.; LeMaster, D. M.; Steitz, J. A. *Science* **1985**, *230*, 1344.

92) Newman, A. J.; Norman, C. *Cell* **1992**, *68*, 743.

93) Ast, G.; Weiner, A. M. *RNA* **1997**, *3*, 371.

94) Wang, J.-H.; Cedergren, R.; Nadal-Ginard, B. *Science* **1994**, *263*, 77.

95) Kandels-Lewis, S.; Séraphin *Science* **1993**, *262*, 2035.

96) Abovich, N.; Rosbash, M. *Cell* **1997**, *89*, 403.

97) Berglund, J. A.; Chua, K.; Abovich, N.; Reed, R.; Rosbash, M. *Cell* **1997**, *89*, 781.

98) Bindereif, A.; Gree, M. R. *EMBO J.* **1987**, *6*, 2415.

99) Ruby, S. W.; Abelson, J. *Science* **1988**, *242*, 1028.

100) Moore, M. J.; Query, C. C.; Sharp, P. A. In *The RNA World*; Gestelan, R. F., Atkins, J. F., Eds.; CSH Laboratory Press: New York, NY, USA, 1993.

101) Wassarman, D. A.; Steitz, J. A. *Nature* **1991**, *349*, 463.

102) Laggerbauer, B.; Lauber, J.; Lührmann, R. *Nucleic Acids Res.* **1996**, *24*, 868.

103) Wallace, J. C.; Edmonds, M. *Proc. Natl. Acad. Sci. U.S.A.* **1983**, *80*, 950.

104) Keller, W. *Cell* **1984**, *39*, 423.

105) Guthrie, C. *Amer. Zool.* **1989**, *29*, 557.

106) Schmelzer, C.; Schweyen, R. *Cell* **1986**, *46*, 557.

107) Parker, R.; Siliciano, P. G.; Guthrie, C. *Cell* **1987**, *49*, 229.

108) Madhani, H. D.; Guthrie, C. *Cell* **1992**, *71*, 803.

109) Langford, C. J.; Klintz, F. J.; Donath, C.; Gallwitz, D. *Cell* **1984**, *36*.

110) Reed, R.; Maniatis, T. *Cell* **1985**, *41*, 95.

111) Vijayraghan, U. *EMBO J.* **1986**, *5*, 1683.

112) Hornig, H.; Abei, M.; Weissmann, C. *Nature* **1986**, *324*, 589.

113) Ruskin, B.; Green, J. M.; Green, M. R. *Cell* **1985**, *41*, 833.

114) Remaud, G.; Balgobin, N.; Glemarec, C.; Chattopadhaya, J. *Tetrahedron* **1989**, *45*, 1537.

115) Mount, S. M.; Pettersson, I.; Hinterberger, M.; Karmas, A.; Steitz, J. A. *Cell* **1983**, *33*, 509.

116) Black, D. L.; Chabot, B.; Steitz, J. A. *Cell* **1985**, *42*, 737.

117) Ruskin, B.; Green, M. R. *Cell* **1985**, *43*, 131.

118) Ruskin, B.; Green, M. R. *Science* **1985**, *229*, 135.

119) Chapman, K. B.; Boeke, J. D. *Cell* **1991**, *65*, 483.

120) Nam, K.; Lee, G.; Trambley, J.; Devine, S. E.; Boeke, J. D. *Mol. Cell Bio.* **1997**, *17*, 809.

121) Petfalski, E.; Dandeker, T.; Henry, Y.; Tollervey, D. *Mol. Cell Bio.* **1998**, *118*, 1181.

122) Ooi, S. L.; Samarsky, D. A.; Fournier, M. J.; Boeke, J. D. *RNA* **1998**, *4*, 1096.

123) Michelson, A. M.; Todd, A. R. *J. Chem. Soc. Part 3* **1955**, 2632.

124) Gilham, P. T.; Khorana, H. G. *J. Am. Chem. Soc.* **1958**, *80*, 6216.

125) Jacob, T. M.; Khorana, H. G. *J. Am. Chem. Soc.* **1965**, *87*, 368.

126) Agrawal, K. L.; Yamazaki, A.; Cashion, P. J.; Khorana, H. G. *Angew. Chem. Int. Ed. Engl.* **1972**, *11*, 451.

127) Hackman, J.; Khorana, H. G. *J. Am. Chem. Soc.* **1969**, *91*, 2749.

128) Agrawal, K. L.; Buchi, H.; Caruthers, M. H.; Gupta, N.; Khorana, H. G.; Kleppe, K.; Kumar, A.; Ohtsuka, E.; RajBhandary, U. L.; van de Sande, J. H.; Sgaramella, V.; Weber, H.; Yamada, T. *Nature* **1970**, *227*, 27.

129) Khorana, H. G.; Agrawal, K. L.; Besmer, P.; Buchi, H.; Carithers, M. H.; Cashion, P. J.; Fridkin, M.; Jay, E.; Kleppe, K.; Kleppe, R.; Kumar, A.; Loewen, P. C.; Miller, R. C.; Minamoto, K.; Panet, A.; RajBhandary, U. L.; Ramamoorthy, B.; Sekiya, T.; Takeya, T.; van de Sande, J. H. *J. Biol. Chem.* **1979**, *251*, 565.

130) Reese, C. B. *Tetrahedron* **1978**, *34*, 3143.

131) Letsinger, R. L.; Mahadevan, V. *J. Am. Chem. Soc.* **1965**, *87*, 3526.

132) Letsinger, R. L.; Mahadevan, V. *J. Am. Chem. Soc.* **1966**, *88*, 5319.

133) Letsinger, R. L.; Ogilvie, K. K. *J. Am. Chem. Soc.* **1967**, *89*, 4801.

134) Letsinger, R. L.; Ogilvie, K. K. *J. Am. Chem. Soc.* **1969**, *91*, 3350.

135) Letsinger, R. L.; Finnan, J. L.; Heavner, G. A.; Lunsford, W. B. *J. Am. Chem. Soc.* **1975**, *97*, 3278-3279.

136) Letsinger, R. L.; Heavner, G. A. *Tetrahedron Letters* **1975**, 147.

137) Letsinger, R. L.; Lunsford, W. B. *J. Am. Chem. Soc.* **1976**, *98*, 3655.

138) Alvaredo-Urbina, G.; Sathe, G. M.; Liu, W. C.; Gillen, M. F.; Duck, P. D.; Bender, R.; Ogilvie, K. K. *Science* **1981**, *214*, 270.

139) Beaucage, S. L.; Caruthers, M. L. *Tetrahedron Lett.* **1981**, *22*, 1859.

140) McBride, L. J.; Caruther, M. H. *Tetrahedron Lett.* **1983**, *24*, 245.

141) Gregg *Tetrahedron Lett.* **1986**, *27*, 4051.

142) Froehler, B. C.; Matteucci, M. D. *Tetrahedron Lett.* **1986**, *27*, 469.

143) Froehler, B. C.; Ng, P. G.; Mattescci, M. D. *Nucleic Acids Res.* **1986**, *14*, 5399.

144) Hall, R. H.; Todd, A. R.; Webb, R. F. *J. Chem. Soc.* **1957**, 3291.

145) Ogilvie, K. K.; Sadana, K. L.; Thompson, E. A.; Quillam, M. A.; Westmore, J. B. *Tetrahedron Lett.* **1974**, *15*, 2861.

146) Ogilvie, K. K.; Beaucage, S. L.; Schiffman, A. L.; Theriault, N. Y.; Sadana, K. L. *Can. J. Chem.* **1978**, *56*, 2678.

147) Corey, E. J.; Venkateswarlu, A. *J. Am. Chem. Soc.* **1972**, *94*, 6190.

148) Reese, C. B.; Saffhill, R.; Sulston, J. E. *J. Am. Chem. Soc.* **1967**, *89*, 3366.

149) Reese, C. B.; Serafinowska, H. T.; Zappia, G. *Tetrahedron Lett.* **1986**, *27*, 2291.

150) Reese, C. B.; Thompson, E. A. *J. Chem. Soc., Perkin Trans. I* **1988**, 2881.

151) Adleman, L. M. *Science* **1994**, *266*, 1021.

152) Damha, M. J.; Pon, R. T.; Ogilvie, K. K. *Tetrahedron Lett.* **1985**, *26*, 4839.

153) Damha, M. J.; Ogilvie, K. K. *J. Org. Chem.* **1988**, *53*, 3710.

154) Damha, M. J.; Zabarylo, S. V. *Tetrahedron Lett.* **1989**, *30*, 6295.

155) Damha, M. J.; Ganeshan, K.; Hudson, R. H. E.; Zabarylo, S. V. *Nucleic Acids Res.* **1992**, *200*, 6565.

156) Kierzek, R.; Kapp, D. W.; Edmonds, M.; Caruthers, M. H. *Nucleic Acids Res.* **1986**, *14*, 4751.

157) Sproat, B. S.; Beijer, B.; Grøtli, M.; Ryder, U.; Morand, K. L.; Lamond, A. I. *J. Chem. Soc., Perkin Trans. I* **1994**, 419-430.

158) Sanger, W. *Principles of nucleic acid structure*; Springer Verlag: New York, **1984**.

159) Yee, T.; Furuichi, T.; Inouye, S.; Inouye, M. *Cell* **1984**, *38*, 203.

160) Lim, D.; Maas, W. K. *Cell* **1989**, *56*, 891.

161) Rice, S. A.; Bieber, J.; Chun, J.-Y.; Stacey, G.; Lampson, B. C. *J. Bacteriol.* **1993**, *175*, 4250.

162) Inouye, M.; Inouye, S. *Ann. Rev. Microbiol.* **1991**, *45*, 163.

163) Temin, H. M. *Nature* **1989**, *339*, 254.

164) Boeke, J. D.; Corces, V. G. *Annu. Rev. Microbiol.* **1989**, *43*, 403.

165) Lease, R. A.; Yee, T. *J. Biol. Chem.* **1991**, *266*, 14497.

166) Lampson, B. C.; Inouye, M.; Inouye, S. *Cell* **1989**, *56*, 701.

167) Maas, W. K.; Wang, C.; Lima, T.; Zubay, G.; Lim, D. *Mol. Microbiol.* **1994**, *14*, 437.

168) Mao, J. R.; Inouye, S.; Inouye, M. *FEMS Microbiol. Lett.* **1996**, *144*, 109.

169) Jeong, D. W.; Kim, K.; Lim, D. B. *Mol. Cells* **1997**, *7*, 347.

170) Holliday, R. *Genet. Res.* **1964**, *5*, 282.

171) Lilley, D. M. J.; Clegg, R. M. *Annu. Rev. Biophys. Biomol. Struct.* **1993**, *22*, 299.

172) Hudson, R. H. E.; Uddin, A. H.; Damha, M. J. *J. Am. Chem. Soc.* **1995**, *117*, 12470.

173) Robidoux, S.; Damha, M. J.; Gehring, K.; Klinck, R. *J. Biomol. Struct. Dyn.* **1997**, *14*, 804.

174) Sharp, P. A.; Konarska, M. M.; Grabowski, P. J.; Lamond, A. I.; Marciniak, R.; Seiler, S. R. *Cold Spring Harbor Symp. Quant. Biol.* **1987**, *LII*, 277.

175) Urdea, M.; Horn, T.; Anderson, M.; Runnings, J.; Hamren, S.; Ahle, D.; Chang, C.-A. *Nucleic Acids Res. Sym. Ser.* **1991**, *24*, 197.

176) Azhayeva, E.; Azhayev, A.; Guzaev, A.; Hovinen, J.; Lönnberg, H. *Nucleic Acids Res.* **1995**, *23*, 1170.

177) Brandenburg, G.; Petersen, G. V.; Rasmussen, K.; Wengel, J. *Bioorg. Med. Chem. Lett.* **1995**, *5*, 791.

178) Hudson, R. H. E.; Damha, M. J. *Nucleic Acids Res. Sym. Ser.* **1993**, *29*, 97.

179) Hudson, R. H. E.; Ganeshan, K.; Damha, M. J. Branched Nucleic Acids: Synthesis and Biological Applications. In *Carbohydrate Modifications In Antisense Research*; Sanghvi, Y. S. and Cook, P. D., Eds.; ACS Symposium Series No. 580; The American Chemical Society: Washington, DC., 1994; pp 133.

180) von Büren, M.; Petersen, G. V.; Rasmussen, K.; Brandenburg, G.; Wengel, J.; Kirpekar, F. *Tetrahedron* **1995**, *51*, 8491.

181) Hudson, R. H. E. Ph.D. Thesis; University of Toronto, Toronto, ON, Canada, **1995**.

182) Brown, D. M.; Magrath, D. I.; Todd, A. R. *J. Chem. Soc.* **1955**, 4496.

183) Huss, S.; Gosselin, G.; Stawinski, S.; Strömberg, R.; Imbach, J.-L. *Nucleosides & Nucleotides* **1988**, *7*, 321.

184) Reese, C. B.; Skone, P. A. *Nucleic Acids Res.* **1985**, *13*, 5215.

185) Pathak, T.; Chattopadhyaya, J. *Acta Chem. Scand.* **1985**, *B39*, 799.

186) Kierzek, R.; Kopp, D. W.; Edmons, M.; Caruthers, M. H. *Nucleic Acids Res.* **1986**, *14*, 4751.

187) De Rooij, J. F. M.; Willie-Hazeleger, G.; Burgers, P. M. J.; van Boom, J. H. *Nucleic Acids Res.* **1979**, *6*, 2237.

188) Remaud, G.; Vial, J.-M.; Nyilas, A.; Balgobin, N.; Chattopadhyaya, J. *Tetrahedron* **1987**, *43*, 947.

189) Huss, S.; Gosselin, G.; Imbach, J.-L. *Tetrahedron Lett.* **1987**, *28*, 415.

190) Sund, C.; Agback, P.; Chattopadhyaya, J. *Tetrahedron* **1991**, *47*, 9659.

191) Sekine, M.; Hata, T. *J. Am. Chem. Soc.* **1985**, *107*, 5813.

192) Sekine, M.; Heikkilä, J.; Hata, T. *Bull. Chem Soc. Jpn.* **1991**, *64*, 588.

193) Sekine, M.; Heikkilä, J.; Hata, T. *Tetrahedron Lett.* **1987**, *28*, 5691.

194) Fourrey, J. L.; Varenne, J.; Fontaine, C.; Guittet, E.; Yang, Z. W. *Tetrahedron Lett.* **1987**, *28*, 1769.

195) Huss, S.; Gosselin, G.; Imbach, J.-L. *J. Org. Chem.* **1988**, *53*, 499.

196) Balgobin, N.; Földesi, A.; Remaud, G.; Chattopadhyaya, J. *Tetrahedron* **1988**, *44*, 6929.

197) Sund, C.; Agback, P.; Chattopadhyaya, J. *Tetrahedron* **1993**, *49*, 649.

198) Grøtli, M.; Sproat, B. S. *J. Chem. Soc., Chem. Commun.* **1995**, 495.

199) Grøtli, M.; Eritja, R.; Sproat, B. S. *Tetrahedron* **1997**, *53*, 11317.

200) Nam, K.; Hudson, R. H. E.; Chapman, K. B.; Ganeshan, K.; Damha, M. J.; Boeke, J. D. *The J. Biol. Chem.* **1994**, *269*, 20613.

201) Reese, C. B.; Ubasawa, A. *Tetrahedron Lett.* **1980**, *21*, 2265.

202) Daskalov, H. P.; Sekine, M.; Hata, T. *Tetrahedron Lett.* **1980**, *21*, 3899.

203) Daskalov, H. P.; Sekine, M.; Hata, T. *Bull. Chem. Soc. Jap.* **1981**, *54*, 3076.

204) Gaffney, B. L.; Jones, R. A. *Tetrahedron Lett.* **1982**, *23*, 2257.

205) Pon, R. T.; Damha, M. J.; Ogilvie, K. K. *Tetrahedron Lett.* **1985**, *26*, 2525.

206) Pon, R. T.; Damha, M. J. *Nucleic Acids Res.* **1985**, *13*, 6447.

207) Ganeshan, K.; Tadey, T.; Nam, K.; Braich, R. S.; Purdy, W. C.; Boeke, J. D.; Damha, M. J. *Nucleosides & Nucleotides* **1995**, *14*, 1009.

208) Ganeshan, K. Ph.D. Thesis, University of Toronto, Toronto, ON, Canada, 1995.

209) Andrus, A. In *Methods in Molecular Biology: Protocols for Oligonucleotide Conjugates;* Agrawal, S. Ed.; Humana Press Inc.: Totowa, NJ, USA, 1994; Vol 26.

210) Damha, M. J.; Giannaris, P. A.; Zabarylo, S. V. *Nucleic Acids Res.* **1990**, *18*, 3813.

211) Wasner, M.; Arion, D.; Damha, M. J. *Biochemistry* **1998**, *37*, 7478.

212) Doudna, J. A.; Szostak, J. W.; Rich, A.; Usman, N. *J. Org. Chem.* **1990**, *55*, 5547.

213) Temsamani, J.; Kubert, M.; Agrawal, S. *Nucleic Acids Res.* **1995**, *23*, 1841.

214) Capaldi, D. C.; Reese, C. B. *Nucleic Acids Res.* **1994**, *22*, 2209.

215) Felsenfeld, G.; Davies, D. R.; Rich, A. *J. Am. Chem. Soc.* **1957**, *79*, 2023.

216) Stroble, S. A.; Doucette-Stamm, L. A.; Riba, L.; Housman, D. E.; Dervan, P. B. *Science* **1991**, *254*, 1639.

217) Strobel, S. A.; Dervan, P. B. *Methods Enzymol.* **1992**, *216*, 309.

218) Radhakrishnan, I.; Patel, D. J. *Biochemistry* **1994**, *33*, 11405.

219) Hoogsteen, K. *Acta Crystallogr.* **1963**, *16*, 907.

220) Mosher, H. E.; Dervan, P. B. *Science* **1987**, *238*, 645.

221) de los Santos, C.; Rosen, M.; Patel, D. *Biochemistry* **1989**, *28*, 7282.

222) Rajagopal, P.; Feigon, J. *Nature* **1989**, *339*, 637.

223) Ono, A.; Ts'o, P. O. P.; Kan, L. S. *J. Org. Chem.* **1992**, *57*, 3225.

224) Krawczyk, S. H.; Milligan, J. F.; Wadwani, S.; Moulds, C.; Froehler, B. C.; Matteucci, M. D. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 3761.

225) Jetter, M. C.; Hobbs, F. W. *Biochemistry* **1993**, *32*, 3249.

226) Huang, C. Y.; Miller, P. S. *J. Am. Chem. Soc.* **1993**, *115*, 10456.

227) Xiang, G.; Soussou, W.; McLaughlin, L. W. *J. Am. Chem. Soc.* **1994**, *116*, 11155.

228) Berressem, R.; Engels, J. W. *Nucleic Acids Res.* **1995**, *23*, 3465.

229) Koshlap, K. M.; Schultze, P.; Brunar, H.; Dervan, P. B.; Feigon, J. *Biochemistry* **1997**, *36*, 2659.

230) Cooney, M.; Czernuszewicz, G.; Postel, E. H.; Flint, S. J.; Hogan, M. E. *Science* **1988**, *241*, 456.

231) Beal, P. A.; Dervan, P. B. *Science* **1991**, *251*, 1360.

232) Durland, R. H.; Kessler, D. J.; Gunnell, S.; Duvic, M.; Pettit, B. M.; Hogan, M. E. *Biochemistry* **1991**, *30*, 9246.

233) Pilch, D. S.; Levenson, C.; Shafer, R. H. *Biochemistry* **1991**, *30*, 6081.

234) Sen, D.; Gilbert, W. *Nature* **1988**, *334*, 364.

235) Cheng, A.-J.; Van Dyke, M. W. *Nucleic Acids Res.* **1993**, *21*, 5630.

236) Milligan, J. F.; Krawczyk, S. H.; Wadwani, S.; Matteucci, M. D. *Nucleic Acids Res.* **1993**, *21*, 327.

237) Olivas, W. M.; Maher III, L. J. *Biochemistry* **1995**, *34*, 278.

238) Olivas, W. M.; Maher III, L. J. *Nucleic Acids Res.* **1995**, *23*, 1936.

239) Hélène, C. *Br. J. Cancer* **1989**, *60*, 157.

240) Sun, J. S.; François, J. C.; Montenay-Garestier, T.; Behomara, T.; Roig, V.; Chassignol, M.; Thuong, N. T.; Hélène, C. *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86*, 9198.

241) Givannangeli, C.; Montenay-Gerestier, T.; Thuong, N. T.; Hélène, C. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 8631.

242) Thuong, N. T.; Hélène, C. *Angew. Chem. Int. Ed. Engl.* **1993**, *32*, 666.

243) Cassidy, S. A.; Strekowski, L.; Wilson, W. D.; Fox, K. R. *Biochemistry* **1994**, *33*, 15338.

244) Fox, K. R. *Nucleic Acids Res.* **1994**, *22*, 2016.

245) Silver, G. C.; Sun, J.-S.; Nguyen, C. H.; Boutorine, A. S.; Bisagni, E.; Hélène, C. *J. Am. Chem. Soc.* **1997**, *119*, 263.

246) Chaudhuri, N. C.; Kool, E. T. *J. Am. Chem. Soc.* **1995**, *17*, 10434.

247) Mergny, J.-L.; Collier, D.; Rougée, M.; Montenay-Garestier, T.; Hélène, C. *Nucleic Acids Res.* **1991**, *19*, 1521.

248 Mergny, J. L.; Duval-Valentin, G.; Nguyen, C. H.; Perrouault, L.; Faucon, B.; Rougee, M.; Montenay-Gerestier, T.; Bisagni, E.; Hélène, C. *Science* **1992**, *256*, 1681.

249) Pilch, D. S.; Waring, M. J.; Sun, J.-S.; Rougée, M.; Nguyen, C.-H.; Bisagni, E.; Montenay-Garestier, T.; Hélène, C. *J. Mol. Bio.* **1993**, *232*, 926.

250) Pilch, D. S.; Martin, M.-T.; Nguyen, C. H.; Sun, J.-S.; Bisagni, E.; Montenay-Garestier, T.; Hélène, C. *J. Am. Chem. Soc.* **1993**, *115*, 9942.

251) Escudé, C.; Nguyen, C. H.; Mergny, J.-L.; Sun, J.-S.; Bisagni, E.; Montenay-Garestier, T.; Hélène, C. *J. Am. Chem. Soc.* **1995**, *117*, 10212.

252) Escudé, C.; Mohammadi, S.; Sun, J.-S.; Nguyen, C.-H.; Bisagni, E.; Liquier, J.; Taillandier, E.; Garestier, T.; Hélène, C. *Chem. Biol.* **1996**, *3*, 57.

253) Prakash, G.; Kool, E. T. *J. Chem. Soc., Chem. Commun.* **1991**, 1161.

254) Kool, E. T. *J. Am. Chem. Soc.* **1991**, *1113*, 6265.

255) Prakash, G.; Kool, E. T. *J. Am. Chem. Soc.* **1992**, *114*, 3523.

256) Rubin, E.; McKee, T. L.; Kool, E. T. *J. Am. Chem. Soc.* **1993**, *115*, 360.

257) Wang, S.; Kool, E. T. *Nucleic Acids Res.* **1994**, *22*, 2326.

258) Rubin, E.; Kool, E. T. *Angew. Chem. Int. Ed. Engl.* **1994**, *33*, 1004.

259) Wang, S.; Kool, E. T. *Nucleic Acids Res.* **1995**, *23*, 1157.

260) Vo, T.; Wang, S.; Kool, E. T. *Nucleic Acids Res.* **1995**, *23*, 2937.

261) Shea, R. G.; Ng, P.; Bischofberger, N. *Nucleic Acids Res.* **1990**, *18*, 4859.

262) Maher III, L. J.; Wold, B.; Dervan, P. B. *Science* **1989**, *245*, 725.

263) Lee, S. J.; Johnson, D. A.; Morgan, R. A. *Nucleic Acids Res.* **1979**, *6*, 3073.

264) Tuite, E.; Norde, B. *J. Chem. Soc., Chem. Commun.* **1995**, 53.

265) Liquier, J.; Coffinier, P.; Firon, M.; Taillandier, E. *J. Biomol. Struct. Dyn.* **1991**, *9*, 437.

266) Taillandier, E.; Liquier, J. In *Methods Enzymol.* **1992**, *211*, 307.

267) Ouali, M.; Letellier, R.; Adnet, E.; Liquier, J.; Sun, J.; Lavery, R.; Taillandier, E. *Biochemistry* **1993**, *32*, 2098.

268) Dagneaux, C.; Liquier, J.; Taillandier, E. *Biochemistry* **1995**, *34*, 14815.

269) Escudé, C.; François, J.-C.; Sun, J.-S.; Ott, G.; Sprinlz, M.; Monetnay-Garestier, T.; Hélène, C. *Nucleic Acids Res.* **1993**, *21*, 5547.

270) Kiran, M. R.; Bansal, M. *Ind. J. Biochem. Biophys.* **1995**, *32*, 391.

271) Geselowitz, D. A.; Neumann, R. D. *Bioconjugate Chem.* **1995**, *6*, 502.

272) Pilch, D. S.; Levenson, C.; Shafer, R. H. *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 1942.

273) Cheng, Y. K.; Pettitt, B. M. *J. Am. Chem. Soc.* **1992**, *114*, 4465.

274) Washbrook, E.; Fox, K. R. *Nucleic Acids Res.* **1994**, *19*, 3977.

275) Chandler, S. P.; Strekowski, L.; Wilson, W. D.; Fox, K. R. *Biochemistry* **1995**, *34*, 7234.

276) Riley, M.; Maling, B.; Chamberlin, M. J. *J. Mol. Bio.* **1966**, *20*, 359.

277) Blake, R. D.; Massoulie, J.; Fresco, J. R. *J. Mol. Bio.* **1967**, *30*, 291.

278) Pilch, D. S.; Brousseau, R.; Shafer, R. H. *Nucleic Acids Res.* **1990**, *18*, 5743.

279) Durland, M.; Peloille, N. T.; Thuong, N. T.; Maurizot, J. C. *Biochemistry* **1992**, *31*, 9197.

280) Rumney IV, S.; Kool, E. T. *J. Am. Chem. Soc.* **1995**, *117*, 5635.

281) Job, P. *Ann. Chim. (Paris)* **1928**, *9*, 113.

282) Harada, N.; Nakanishi, K. In *Circular Dichroic Spectroscopy; Exciton Coupling in Organic Stereochemistry;* University Science Books: New York, NY, USA, 1983.

283) Steely, J. R.; Gray, D. M.; Ratliff, R. L. *Nucleic Acids Res.* **1986**, *14*, 10071.

284) Antao, V. P.; Gray, D. M.; Ratliff, R. L. *Nucleic Acids Res.* **1988**, *16*, 719.

285) Chan, S. S.; Breslauer, K. J. *Biochemistry* **1990**, *29*, 6161.

286) Callahan, D. E.; Trapane, T. L.; Miller, P. S.; Ts'o, P. O. P.; Kan, L.-S. *Bichemistry* **1991**, *30*, 1650.

287) Völker, J.; Osborne, S. E.; Glick, G. D.; Breslauer, K. J. *Biochemistry* **1997**, *36*, 756.

288) Sekharudu, C. Y.; Yatindra, N.; Sundaralingam, M. In *Structural Biology; The State of the Art. Proceedings of the Eighth Conversation;* Samra, R. H. and Samra, M. H., Eds.; Adenine Press: Guilderland, NY, 1994; *2*, 113.

289) Jiang, S.-P.; Jerinigan, R. L.; Ting, K.-L.; Syi, J.-L.; Raghunathan, G. *J. Biomol. Struct. Dyn.* **1994**, *12*, 383.

290) Singleton, S. F.; Dervan, P. B. *Biochemistry* **1992**, *31*, 10095.

291) Singleton, S. F.; Dervan, P. B. *J. Am. Chem. Soc.* **1992**, *114*, 6957.

292) Singleton, S. F.; Dervan, P. B. *Biochemistry* **1993**, *32*, 13171.

293) Singleton, S. F.; Dervan, P. B. *J. Am. Chem. Soc.* **1994**, *116*, 10376.

294) Best, G. C.; Dervan, P. B. *J. Am. Chem. Soc.* **19954**, *117*, 1187.

295) Greenberg, W. A.; Dervan, P. B. *J. Am. Chem. Soc.* **1995**, *117*, 5016.

296) Shindo, H.; Torigoe, H.; Sarai, A. *Bichemistry* **1993**, *32*, 8963.

297) Yang, M.; Ghosh, S. S.; Millar, D. P. *Biochemistry* **1994**, *33*, 15329.

298) Mergny, J. L.; Sun, J.-S.; Rougée, M.; Montenay-Garestier, T.; Barcelo, F.; Chomilier, J.; Hélène, C. *Biochemistry* **1991**, *30*, 9791.

299) Völker, J.; Botes, D. P.; Lindsey, G. G.; Klump, H. H. *J. Am. Chem. Soc.* **1993**, *230*, 1278.

300) Scaria, P. V.; Shafer, R. H. *Biochemistry* **1996**, *35*, 10985.

301) Manzini, G.; Xodo, L. E.; Gasparotto, D.; Quadrifoglio, F.; van der Marel, G. A.; van Boom, J. H. *J. Mol. Bio.* **1990**, *213*, 833.

302) Xodo, L. E.; Manzini, G.; Quadrifoglio, F. *Nucleic Acids Res.* **1990**, *18*, 3557.

303) Plum, G. E.; Park, Y. W.; Singleton, S. F.; Dervan, P. B.; Breslauer, K. J. *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 9436.

304) Plum, G. E.; Breslauer, K. J. *J. Mol. Bio.* **1995**, *248*, 679.

305) Naghibi, H.; Tamura, A.; Sturtevant, J. M. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 5597.

306) Plum, G. E.; Pilch, D. S.; Singleton, S. F.; Breslauer, K. J. *Annu. Rev. Biophys. Biomol. Struct.* **1995**, *24*, 319.

307) Marky, L. A.; Breslauer, K. J. *Bioploymers* **1987**, *26*, 1601-1620.

308) Soyer, V. N.; Potman, V. N. In *Triple Helical Nucleic Acids;* Springer-Verlag: New York, NY, USA, 1995.

309) Erie, D.; Sinha, N.; Olson, W.; Jones, R.; Breslauer, K. J. *Biochemistry* **1987**, *26*, 7150.

310) Doktyca, M. J.; Goldstein, R. F.; Paner, T. M.; Gallo, F. J.; Benight, A. S. *Bioploymers* **1992**, *32*, 849.

311) Ma, M. Y.-X.; McCallum, K.; Climie, S. C.; Kuperman, R.; Lin, W. C.; Summer-Smith, M.; Barnett, R. W. *Nucleic Acids Res.* **1993**, *21*, 2585.

312) Stevens, S. Y.; Swanson, P. C.; Voss Jr., E. W.; Glick, G. D. *J. Am. Chem. Soc.* **1993**, *115*, 1585.

313) Subramanian, P. S.; Ravishanker, G.; Beveridge, D. L. *Proc. Natl. Acad. Sci. U.S.A.* **1988**, *85*, 1836.

314) Subramanian, P. S.; Swaminathan, S.; Beverideg, D. L. *J. Biomol. Struct. Dyn.* **1990**, *7*, 1161.

315) Neidle, S. *FEBS Lett.* **1992**, *298*, 97.

316) Gao, H.; Chidambaram, N.; Chen, B. C.; Pelham, D. E.; Patel, R.; Yang, M.; Zhou, L.; Cook, A. F.; Cohen, J. S. *Bioconjugate Chem.***1994**, *5*, 445.

317) Varani, G. *Annu. Rev. Biophys. Biomolec. Struct.* **1995**, *24*, 379.

318) Damha, M. J.; Ogilvie, K. K. *Biochemistry* **1988**, *27*, 9403.

319) Sund, C.; Agback, P.; Koole, L. H.; Sandström, A.; Chattopadhyaya, J. *Tetrahedron* **1992**, *48*, 695.

320) Braich, R. S.; Damha, M. J. *Bioconjugate Chem.* **1997**, *8*, 370.

321) Tang, J. Y.; Temsamani, J.; Agrawal, S. *Nucleic Acids Res.* **1993**, *21*, 2729.

322) Hirao, I.; Kawai, G.; Yoshizawa, S.; Yoshifumi, N.; Yoshiharu, I.; Kimitsuna, W. *Nucleic Acids Res.* **1994**, *22*, 576.

323) Davison, A.; Leach, D. R. F. *Nucleic Acids Res.* **1994**, *22*, 4361.

324) Gao, H.; Yang, M.; Cook, A. F. *Nucleic Acids Res.* **1995**, *23*, 285.

325) Gao, H.; Yang, M.; Patel, R.; Cook, A. F. *Nucleic Acids Res.* **1995**, *23*, 2025-2029.

326) Giovannangeli, C.; Monteray-Garestier, T.; Rougée, M.; Chassignol, M.; Thuong, N. T.; Hélène, C. *J. Am. Chem. Soc.* **1991**, *113*, 7775.

327) Salunkhe, M.; Wu, T.; Letsinger, R. L. *J. Am. Chem. Soc.* **1992**, *114*, 8768.

328) Wang, S.; Kool, E. T. *J. Am. Chem. Soc.* **1994**, *116*, 8857.

329) van de Sande, J. H.; Ramsing, N. B.; Germann, M. W.; Elhorst, W.; Kalisch, B. W.; Kitzing, E. V.; Pon, R. T.; Clegg, R. C.; Jovin, T. M. *Science* **1988**, *241*, 551.

330) Rich, A.; Davies, D. R.; Crick, F. H.; Watson, J. D. *J. Mol. Bio.* **1961**, *3*, 71.

331) Westhof, E.; Sundaralingam, M. *Proc. Natl. Acad. Sci. U.S.A.* **1980**, *77*, 1852.

332) Sarma, M. H.; Gupta, G.; Sarma, R. H. *FEBS Lett.* **1986**, *205*, 223.

333) Ramsing, N. B.; Jovin, T. M. *Nucleic Acids Res.* **1988**, *16*, 6659.

334) Liu, K.; Miles, H. T.; Frazier, J.; Sasisekharan, V. *Biochemistry* **1993**, *32*, 11802.

335) Kandimalla, E. R.; Agrawal, S. *J. Am. Chem. Soc.* **1995**, *117*, 6416.

336) Smith, M. In *Science and Scoiety: The John C. Polanyi Nobel Laureates Lectures;*; House of Anansi Press Ltd.: Concord, Ontario, Canada, 1995.

337) Nishimura, S.; Jacob, T. M.; Khorana, H. G. *Proc. Natl. Acad. Sci. U.S.A.* **1964**, 1494.

338) Lipton, R. J. *Science* **1995**, *268*, 542.

339) Guarnieri, F.; Fliss, M.; Bancroft, C. *Science* **1997**, *273*, 220.

340) Jacquier, A.; Rosbach, M. *Proc. Natl. Acad. Sci. U.S.A.* **1986**, *83*, 5835.

341) Arenas, J.; Hurwitz, J. *J. Biol. Chem.* **1987**, *262*, 4274.

342) Krämer, A.; Keller, W. *EMBO J.* **1985**, *4*, 3571.

343) Cobianchi, F.; Wilson, S. H. *Guide to Molecular Cloning Techniques*; Academic Press, Inc.: Orlando, Florida, USA, 1987, *152*.

344) Choongeun, L.; Suhadolnik, R. J. *Biochemistry* **1985**, *24*, 551.

345) Torrence, P. F.; Brozda, D.; Alster, D.; Charubala, R.; Pfleiderer, W. *J. Biol. Chem.* **1988**, *263*, 1131.

346) Ruskin, B.; Green, M. R. *Methods in Enzymology* **1990**, *181*, 180.

347) Fouser, L. A.; Friesen, J. D. *Cell* **1986**, *45*, 81.

348) Sharp, P. A. *Angew. Chem. Int. Ed. Engl.* **1994**, *33*, 1229.

349) Green, M. R. *Methods Enzymol.* **1986**, *20*, 671.

350) Padgett, R. A.; Grabowski, P. J.; Konarska, M. M.; Seiler, S.; Sharp, P. A. *Annu. Rev. Biochem.* **1986**, *55*, 1119.

351) Parker, R.; Guthrie, C. *Cell* **1985**, *41*, 107.

352) Coulter, L. R.; Landree, M. A.; Cooper, T. A. *Mol. Cell. Biol.* **1997**, *17*, 2143.

353) Zhuang, Y.; Weiner, A. M. *Cell* **1986**, *46*, 827.

354) Legrain, P.; Séraphin, B.; Rosbash, M. *Mol. Cell Bio.* **1987**, *8*, 3755.

355) Wu, S.; Green, M. R. *EMBO Journal* **1997**, *16*, 4421.

356) Sharp, P. A. *Cell* **1981**, *23*, 643.

357) Anderson, K.; Moore, M. J. *Science* **1997**, *276*, 1712.

358) Pascolo, E.; Séraphin, B. *Mol. Cell. Biol.* **1997**, *17*, 3469.

359) Maschhoff, K. L.; Padgett, R. A. *Nucleic Acids Reseacrh* **1992**, *20*, 1949.

360) Suh, E.; Waring, R. B. *Nucleic Acids Res.* **1992**, *20*, 6303.

361) Kim, C. H.; Ryan, D. E.; Marciniec, T.; Abelson, J. *EMBO J.* **1997**, *16*, 2119.

362) Zamore, P. D.; Green, M. R. *EMBO J.* **1991**, *10*, 207.

363) Still, W. C.; Kahn, M.; .Mitra, A. *J. Org. Chem.* **1978**, *43*, 2923.

364) Ogilvie, K. K.; Usman, N.; Nicoghosian, K.; Cedergren, R. J. *Proc. Natl. Acad. Sci. U.S.A.* **1988**, *85*, 5764.

365) Damha, M. J.; Ogilvie, K. K. In *Oligoribonucleotide Synthesis*; Agrawal, S., Ed.; The Humana Press Inc.: Totowa, NJ, 1993, pp 81.

366) Grierson, D. In *Gel Electrophoresis of RNA*; Rickwood, D. and Hames, B. D., Eds.; IRL Press: Washington DC, 1982, pp 1.

367) Mullah, B.; Andrus, A. *Nucleosides & Nucleotides* **1996**, *15*, 419.

368) Tadey, T. Ph.D. Thesis; McGIll University, Montreal, PQ, Canada, 1994.

369) Hjertén, S. *J. Chromatogr.* **1985**, *347*, 191.

370) Puglisi, J. D.; Tinoco Jr., I. *Methods Enzymol.* **1989**, *180*, 304.

371) Breslauer, K. J. In *Protocols for Oligonucleotide Conjugates;* Agrawal, S. Ed.; The Humana Press Inc.: Totowa, NJ, 1994, *26*, 347.

372) Kibler-Herzog, L.; Zon, G.; Whittier, G.; Shaikh, M.; Wilson, W. D. *Anti-Cancer Drug Design* **1993**, *8*, 65.

373) Robidoux, S. Ph.D. Thesis; McGill University, Montreal, PQ, Canada, **1999**.

374) Moore, M .J; Sharp, P. A. *Nature* **1993**, *365*, 364.

375) Ferentz, A. E.; Verdine, G. L. *J. Am. Chem. Soc.* **1991**, *113*, 4000.

376) Wolfe, S. A.; Verdine, G. L. *J. Am. Chem. Soc.* **1993**, *115*, 12585.

377) Erlanson, D. A.; Chen, L.; Verdine, G. L. *J. Am. Chem. Soc.* **1993**, *115*, 12583.

378) Wolfe, S. A.; Ferentz, A. E.; Grantcharova, V.; Churchill, M. E. A.; Verdine, G. L. *Chem. Biol.* **1995**, *2*, 213.