

In compliance with the
Canadian Privacy Legislation
some supporting forms
may have been removed from
this dissertation.

While these forms may be included
in the document page count,
their removal does not represent
any loss of content from the dissertation.

**RNA FOLDING: SYNTHESIS, STRUCTURE AND BIOLOGICAL
PROPERTIES OF HAIRPINS BASED ON
2',5'-LINKED RNA LOOPS**

Rami Nabil Hannoush

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
Doctor of Philosophy

© Copyright by Rami Nabil Hannoush 2002



National Library
of Canada

Bibliothèque nationale
du Canada

Acquisitions and
Bibliographic Services

Acquisitons et
services bibliographiques

395 Wellington Street
Ottawa ON K1A 0N4
Canada

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file *Votre référence*

ISBN: 0-612-88486-4

Our file *Notre référence*

ISBN: 0-612-88486-4

The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

L'auteur conserve la propriété du droit d'auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

Canada

Dedicated to my parents: Nabil and Hayfa Hannoush

With special gratitude to my family: Mazen, Hiba, Mohamad, and Wafa

ABSTRACT

Ribonucleic acids (RNA) are polymers of ribonucleotides linked together by 3',5'-phosphodiester linkages and play a prominent role in protein biosynthesis. Much less common are ribonucleic acids based on 2',5'-phosphodiester linkages (2',5'-RNA). In this study, the synthesis, physical and biological properties of several 2',5'-linked oligoribonucleotides are described.

One aspect of this thesis deals with the synthesis and structural studies of 2',5'-dodecaribonucleotides. Their base sequence was designed to promote intramolecular folding by base pairing leading to the formation of "hairpin loops". The hairpins consist of a tetranucleotide loop and a 4 base-pair duplex (stem). The thermodynamic stability and conformation of these hairpins were assessed by UV, CD and NMR spectroscopy. Melting experiments (UV) revealed that with a few exceptions, hairpins comprised of 2',5'-(UUCG) loops exhibit greater thermodynamic stability (*e.g.* T_m) than the corresponding hairpins with 3',5'-linked loops. The 'extra' stability imparted by the 2',5'-(UUCG) loop was found to be virtually independent of the sugar composition of the stem. For example, the 2',5'-tetraloop stabilizes hairpins whether their duplex stem is based on double-stranded DNA or RNA. In contrast, the 3',5'-tetraloop stabilizes hairpins only when their stem duplex is double-stranded RNA. NMR studies revealed that the 2',5'-tetraloop (UUCG) is self-stabilized by a wobble U-G base pair, extensive base stacking and sugar-base contacts. A more striking feature is the protrusion of the cytosine residue into the solvent without participating in any of the loop stabilizing interactions. This identifies the 2',5'-linked (UUCG) loop as a novel structural motif and provides the first demonstration that 2',5'-RNA can fold back on itself to form a hairpin structure of unusual thermodynamic stability.

The ability of hairpin structures to inhibit human immunodeficiency virus (HIV-1) reverse transcriptase (RT) was also evaluated. Four potent hairpins based on 3',5'- and 2',5'-RNA were able to inhibit the RNase H activity of HIV-1 reverse transcriptase, a key enzyme for the proliferation of the human immunodeficiency virus (HIV-1). The

polymerase activity of HIV-1 RT was not affected by this class of oligonucleotides. The hairpins were identified from a nucleic acid library synthesized *via* diversity-oriented solid-phase synthesis. These compounds represent the first examples of hairpins, 12 nucleotides in length, that inhibit specifically the RNase H activity of HIV-1 RT without affecting its polymerase activity.

Another study in this work dealt with yeast RNase III (Rnt1p), an enzyme implicated in the mechanism of action of RNA interference (RNAi). Through these investigations, it was discovered that Rnt1p cleaves the DNA strand of DNA:RNA hybrids. This was totally unexpected since Rnt1p, like other RNase III enzymes, was thought to be specific only towards the cleavage of double-stranded RNA. These studies also increased our knowledge about the mechanism by which the enzyme cleaves the target RNA (or DNA) strand and suggest that the vicinal 2'-hydroxyl group on the ribose sugar does not participate directly in the cleavage reaction.

Finally, the formation of C-tetrad structures (*i*-motif) was induced through the design and synthesis of extra-stable hairpin loops containing deoxycytidine rich stems. The corresponding hairpins containing ribocytidine folded into duplexes rather than C-tetrad structures. These studies lead to the first detection and identification of antiparallel 2',5'-RNA duplexes based on hemiprotonated C·C⁺ base pairs.

RÉSUMÉ

Les acides ribonucléotides (ARN) sont des polymères de ribonucléotides liés entre eux par des liaisons phosphodiester en position 2' et 5' (2',5'-ARN), qui jouent un rôle prédominant dans la biosynthèse des protéines. Dans cette étude sont décrites la synthèse ainsi que les propriétés physiques et biologiques de différents oligoribonucléotides liés en 2' et 5'.

Un aspect de cette thèse consiste en la synthèse et les études structurales des dodécaribonucléotides liés en position 2' et 5'. La séquence des bases a été choisie de manière à provoquer un repliement intramoléculaire par association des bases qui conduit à la formation de boucles en épingle. Ces épingles consistent en une boucle de quatre nucléotides qui s'associent par complémentarité des bases (tronc). La thermostabilité et la conformation de ces épingles ont été déterminées par spectroscopie UV, CD et RMN. Les expériences melting démontrent à peu d'exceptions près, que les épingles contenant des séquences 2',5'-(UUCG) dans les boucles montrent une plus grande stabilité (Temperature) que les épingles liées en position 3',5' correspondantes. Cette plus grande stabilité imputée aux 2',5' boucles est virtuellement indépendante de la composition du sucre du tronc. Par exemple, la 2',5' tétraboucle stabilise les épingles avec des troncs de doubles hélices composées de deux brins d'ARN ou d'ADN. Inversement, les 3',5' tétraboucles ne stabilisent que les épingles composées de deux brins d'ARN. Des études RMN démontrent que la 2',5'-tétraboucle (UUCG) est auto-stabilisée par un appariement wobble des bases U·G, une forte interaction des bases et un contact entre les sucres et bases. Un autre élément important est l'exclusion du résidu cytosine au niveau du solvant sans aucune participation dans les interactions stabilisantes de la boucle. Ceci désigne la boucle (UUCG) liée en 2',5' comme un nouveau motif structural et démontre pour la première fois que les 2',5'-ARN peuvent se replier sur eux-même pour former une structure épingle avec une inhabituelle stabilité thermodynamique.

La capacité des structures épingles à inhiber la transcriptase inverse du virus de l'immunodéficience humaine (VIH-1) a été étudiée. Quatre potentielles épingles basées

sur des 3',5' et 2',5'-ARN sont inhibitrices de l'activité ARNase H de la transcriptase inverse VIH-1, une enzyme essentielle à la prolifération du virus d'immunodéficience humaine (VIH-1). L'activité polymérase du VIH-1RT n'est pas affectée par cette catégorie d'oligonucléotides. Les épingles ont été identifiées à partir d'une librairie d'acides nucléiques synthétisés au cours d'une synthèse en phase solide avec une diversité orientée. Ces composés représentent le premier exemple d'épingles d'une longueur de 12 nucléotides, qui inhibent spécifiquement l'activité ARNase H du VIH-1RT sans affecter l'activité polymérase.

D'autres études dans ce travail concernent la levure ARNase III (Rnt 1 p), une enzyme impliquée dans le mécanisme d'action lors du phénomène d'interférence de l'ARN (ARNi). Lors de ces investigations il a été découvert que la Rnt 1p clive le brin ADN des hybrides ADN:ARN. Ceci n'était pas envisageable étant donné que la Rnt 1p, comme toutes les autres enzymes ARNase III, était considérée comme spécifique du clivage des doubles brins ARN. Ces études ont aussi permis de mieux comprendre le mécanisme par lequel l'enzyme clive le brin ARN ciblé (ou ADN) et permet de penser que le groupe hydroxyle vicinal en position 2' sur le sucre ribose, ne participe pas directement à la réaction de clivage.

Pour finir, la formation de la structure C-tetrad (motif i) a été introduite grâce au design et à la synthèse de boucles en épingle très stables contenant un tronc riche en déoxycytidine. L'épingle correspondante contenant la ribocytidine, s'arrange en double hélice plutôt qu'en une structure C-tetrad. Ces études conduisent à la première détection et identification d'une double hélice constituée de deux brins d'ARN 2',5' antiparallèles basés sur une paire de bases hémiprotonées (C·C+).

ACKNOWLEDGEMENTS

I would like to express my deepest appreciation to my Ph.D. supervisor Professor Masad J. Damha for giving me the opportunity to conduct research in his laboratory. Looking at the path that I have taken during these past five years, I have come to realize that the decision I took regarding joining his group is one of the best decisions I have made. I am really thankful for his invaluable advice, endless encouragement and most of all, for being the passionate academic mentor. Thank you “Dr. Masad”.

I am very grateful to both NSERC and FCAR for post-graduate scholarships. I am also grateful to the Department of Chemistry at McGill for providing me with graduate assistantships throughout the years. All of these have enabled me to focus diligently on my research without any distractions. I am extremely thankful to NSERC for a post-doctoral fellowship that will allow me to pursue my scientific endeavors.

I would like to thank all my coworkers in Lab 207, past and present, for their help during the course of my work. A special thank you to Dr. Ekaterina Viazovkina and Dr. Mohammad El-Zagheid for their great assistance in solution and solid-phase synthesis. Thanks especially to my colleagues Maria Mangos, Sandra Carriero and Robert Donga for proofreading my thesis. An additional thank you to Benedicte Patureau for translating the abstract. I would also like to thank Daniel Yazbeck with whom I set up the first ^{32}P -labeling experiments in the Damha laboratory; I also thank you for your friendship.

I am indebted to all the scientists that collaborated with me during the course of my studies including: Dr. Kalle Gehring and Dr. Aleksej Denisov for their outstanding expertise in NMR, Dr. Kyung-Lyum Min for his help with RNase H screening, and Dr. Nam Lok for teaching me how to run HIV-1 reverse transcriptase assays. I am also indebted to Dr. Paul Xia for teaching me NMR, Dr. J. Turnbull (Concordia University) for use of the CD instrument, Antisar Hlal for MALDI-TOF MS, and Mr. Nadim Saade for running FAB-MS.

I would especially like to thank faculty members who shared their experiences with me in different ways and helped make my stay in the Otto Maass Chemistry Building most enjoyable. I thank Professor Hanadi Sleiman for her unparalleled enthusiasm about science, and the support and encouragement I received from her since my first year

of undergraduate studies at the American University of Beirut. I thank Professor David Harpp for all those superb discussions about teaching chemistry in the classroom and those morning coffee chats. I would like to thank Professor Patrick Farrell for always keeping up with my progress, for teaching me thermodynamic energy states, and not the least, for sharing his secrets and unparalleled knowledge about wine. A special thanks to Professor Just for giving me the opportunity to run the organic seminars with him. This was a most enjoyable experience for me.

A special thank you to Renee Charron, Chantal Marotte, Carol Brown, Sandra Aerssen and Fay Nurse for their administrative assistance.

Thanks to Valerie Gandubert, Debbie Mitra, and Andrew Corbet for their friendship.

I would like to thank a few special people who have helped me in every possible way since I first arrived to Montréal. The Hachem family (especially Mr. Charif Hachem) for making me feel at home and endless support, and my two great friends, Ziad and Bilal for making the first few months in Montréal very pleasant.

I would also like to thank a few special friends who offered all the support especially in the last period during my writing of the thesis. My best friends Marco Abu Zahr, for hosting me during the most critical time, and Alex Baasiri for being around all the time. I would also like to thank the Bayaa family for all their love. A special thank you to Ziad Bayaa for showing me all the time that life is easy-going.

I would like to thank a special person to my heart, who stood by me all those years. To my fiancé Wafa, I thank you for all your endless love and support throughout the years and for being there for me all the time.

Last but not least, I deeply thank my parents who offered love and support beyond description. I thank my dad for inspiring me with his passionate love for knowledge and for showing me perseverance without which I could not have reached my goals. I thank my mother for giving me all the love and strength in the world to keep me going. I thank my brothers Mazen and Mohamad and my sister Hiba for always making sure that I am doing fine and for reminding me to eat especially when I am busy with my studies. To all of you, a million thanks.

TABLE OF CONTENTS

Dedication	ii
Abstract	iii
Resumé	v
Acknowledgments	vii
Table of Contents	ix
Abbreviations	xv
List of Figures	xviii
List of Tables	xxvii
I GENERAL INTRODUCTION	1
1.1 NUCLEIC ACIDS	1
1.2 SOLID-PHASE OLIGONUCLEOTIDE SYNTHESIS	6
1.3 TECHNIQUES USED IN THE STUDY OF NUCLEIC ACIDS	9
1.3.1 Ultraviolet (UV) Spectroscopy	10
1.3.2 Circular Dichorism (CD) Spectroscopy	11
1.3.3 Gel Electrophoresis	11
1.3.4 Nuclear Magnetic Resonance (NMR) Spectroscopy	12
1.4 RNA ORGANIZATION AND ITS STRUCTURAL ELEMENTS	12
1.5 THERMODYNAMICS OF HAIRPIN FOLDING	22
1.5.1 Effects of Loop Sequence and Composition on Hairpin Stability	22
1.5.2 Unusually Stable Hairpins	23
1.6 RNA INTERFERENCE	25
1.7 HIV-1 REPLICATION AND REVERSE TRANSCRIPTASE	27
1.8 NUCLEIC ACIDS WITH 2',5'-PHOSPHODIESTER LINKAGES	31
1.8.1 Biological Occurrence of 2',5'-Nucleic Acids	31
1.8.2 The Literature of 2',5'-Nucleic Acids	33
1.8.3 Preferred Sugar Puckers of 2',5'-Nucleic Acids	40
1.9 CHRONOLOGICAL SUMMARY OF THE BIOPHYSICAL PROPERTIES OF 2',5'-NUCLEIC ACIDS AND THE DEVELOPMENT OF 2',5'-NUCLEIC ACID CHEMISTRY	42
1.10 THESIS OBJECTIVES	46

II	THERMODYNAMIC AND STRUCTURAL STUDIES OF HAIRPINS CONTAINING 2',5'-RNA LOOPS	48
2.1	BACKGROUND	48
2.2	EXPERIMENTAL DESIGN	50
2.3	HAIRPIN SYNTHESIS AND CHARACTERIZATION	51
2.4	HAIRPIN <i>VERSUS</i> DIMER FORMATION	56
2.5	THERMODYNAMIC CHARACTERIZATION AND DATA ANALYSIS	58
2.6	THERMODYNAMIC PROPERTIES OF RNA HAIRPINS: R <i>VERSUS</i> <u>R</u> LOOPS	61
2.7	THERMODYNAMIC PROPERTIES OF DNA HAIRPINS: <u>DRD</u> <i>VERSUS</i> DRD & DDD HAIRPINS	65
2.8	R AND <u>R</u> LOOPS IN HAIRPINS WITH DD AND RR STEMS	66
2.9	HAIRPINS WITH VARIOUS STEMS: R <i>VERSUS</i> <u>R</u> LOOPS	68
2.10	FORMATION OF 2',5'-RNA:DNA HYBRIDS	68
2.11	GLOBAL HELICAL CONFORMATION OF HAIRPINS	70
2.12	EVIDENCE FOR THE EXISTENCE OF AN ALL 2',5'-RNA HAIRPIN AND COMPARISON TO ALL-DNA AND ALL-RNA HAIRPINS	71
2.13	EFFECT OF C·G LOOP-CLOSING BASE PAIR ON THERMODYNAMIC STABILITY OF HAIRPINS WITH 3',5'- AND 2',5'-RNA (UUCG) LOOPS	73
2.13.1	Background and General Considerations	73
2.13.2	Effects of Substitution of Deoxyribose for Ribose in the Loop-Closing Base Pair of RNA Hairpins	75
2.13.3	Effects of Substitution of Ribose for Deoxyribose in the Loop-Closing Base Pair of Hairpins with DNA Stems	78
2.14	NMR SOLUTION STRUCTURE OF 2',5'-RNA LOOPS	81

III	SELECTIVE INHIBITION OF HIV-1 REVERSE TRANSCRIPTASE RNASE H BY HAIRPIN APTAMERS EVOLVED FROM A LIBRARY CONSTRUCTED VIA DIVERSITY-ORIENTED SYNTHESIS	96
3.1	BACKGROUND	96
3.2	IMPORTANCE OF RNASE H AS A DRUG TARGET	97
3.3	PROJECT OBJECTIVE	100
3.4	DESIGN OF MINI-HAIRPINS AS POTENTIAL INHIBITORS OF HIV-1 RT	101
3.5	LIBRARY GENERATION VIA NUCLEIC ACID COMBINATORIAL SPLIT POOL SYNTHESIS	103
3.6	CONFORMATIONAL DIVERSITY OF LIBRARY MEMBERS	109
3.7	SELECTION OF POTENT HIV-1 RT INHIBITORS	111
3.8	STRUCTURE-ACTIVITY RELATIONSHIPS	117
3.8.1	A Folded Hairpin, But Not Linear Oligomer, is Recognized by HIV-1 RT RNase H	118
3.8.2	Comparison of HIV-Inhibitory Activity of 2',5'- <i>Versus</i> 3',5'-RNA Loops	119
3.8.3	Hairpin Inhibitory Potency is Dependent on Loop Conformation	122
3.8.4	Hairpin Inhibitory Activity is Independent of Thermal Stability	123
3.9	GEL SHIFT MOBILITY ASSAY	125
3.10	INHIBITION OF RNA- AND DNA- DEPENDENT DNA POLYMERASE ACTIVITY OF HIV-1 RT	125
3.11	UV CROSS-LINKING STUDY	129
3.12	STABILITY OF HAIRPIN APTAMERS TOWARDS NUCLEASES	129
3.12.1	Stability Towards Blood Serum	129
3.12.2	Stability Towards Nuclease P1	131
3.13	CONCLUSIONS	137

IV	PROBING THE BINDING SPECIFICITY AND CLEAVAGE MECHANISM OF YEAST RNASE III (RNT1P)	139
4.1	BACKGROUND	139
4.2	SUBSTRATE RECOGNITION AND CLEAVAGE BY RNT1P	141
4.3	PROJECT OVERVIEW AND DESIGN	142
4.4	INVESTIGATION OF RNT1P SUBSTRATE RECOGNITION	143
4.4.1	Thermal and Helical Properties	145
4.4.2	Effect of Stem Composition on Hairpin Binding to Rnt1p	146
4.4.3	Effect of Loop Structure and Base Sequence on Hairpin Binding to Rnt1p	149
4.5	PROBING THE MOLECULAR REQUIREMENTS FOR CLEAVAGE OF HAIRPIN DUPLEXES BY RNT1P	151
4.6	YEAST RNASE III (RNT1P) CLEAVES DNA IN DNA:RNA HYBRIDS	158
4.7	PROPOSED MECHANISM FOR RNA CLEAVAGE BY YEAST RNASE III	162
4.8	OTHER REQUIREMENTS FOR RNT1P-MEDIATED DNA CLEAVAGE	165
4.9	DNA CLEAVAGE IN DNA:RNA HYBRIDS IS SPECIFIC TO EUKARYOTIC, BUT NOT PROKARYOTIC, RNASE III ENZYMES	165
4.10	RNT1P DOES NOT CLEAVE A 2',5'-PHOSPHODIESTER LINKAGE	169
4.11	CONCLUSIONS	172
4.12	BIOLOGICAL SIGNIFICANCE	173
V	ASSOCIATION OF CYTOSINE-RICH HAIRPINS INTO THE I-MOTIF	174
5.1	BACKGROUND	174
5.2	ASSOCIATION OF OLIGONUCLEOTIDES INTO THE I-MOTIF	177
5.3	PROJECT OVERVIEW	180

5.4	HAIRPINS WITH DEOXYCYTIDINE STEMS	180
5.4.1	UV Thermal Melting Studies and Dependence on pH	182
5.4.2	Hairpin <i>Versus</i> Tetraplex Formation at pH 5.0	189
5.4.3	Circular Dichroism Studies	192
5.4.4	Gel Mobility Shift Assay	193
5.4.5	Proposed Structure for the Dimeric i-Motif Complex	199
5.5	HAIRPINS WITH RIBOCYTIDINE STEMS	201
5.5.1	Background	201
5.5.2	Experimental Design	202
5.5.3	UV Thermal Studies	204
5.5.4	Hairpin Duplex <i>Versus</i> Tetraplex Formation at pH 5.0	208
5.5.5	Gel Mobility Shift Assay	210
5.5.6	Structural Characterization of the Duplex via Circular Dichroism Studies	212
5.5.7	Evidence for the Formation of Antiparallel But Not Parallel C·C+ Duplexes	217
5.6	2',5'-RNA ASSOCIATES INTO AN ANTIPARALLEL C·C+ DUPLEX	219
5.7	BIOLOGICAL STUDIES	225
5.7.1	Inhibition of HIV-1 RT RNase H Activity	226
5.7.2	Inhibition of HIV-1 RT RNA- and DNA-Dependent DNA Polymerase Activity	227
5.7.3	Insights into the Nature of the Complex Formed Under Assay Conditions	227
5.8	CONCLUSIONS AND FUTURE PERSPECTIVES	229
VI	CONTRIBUTIONS TO KNOWLEDGE	230
VII	EXPERIMENTAL	235
7.1	GENERAL METHODS	235
7.1.1	General Reagents	235
7.1.2	Chromatography	235
7.1.3	Instruments	236
7.2	OLIGONUCLEOTIDE SYNTHESIS	237
7.2.1	Reagents for Solid Support Derivatization and Nucleoside Loading	237
7.2.2	Derivatization of Solid Support	237
7.2.3	Monomers for Automated Synthesis	238
7.2.4	Automated Oligonucleotide Synthesis	239

7.3	PURIFICATION OF OLIGONUCLEOTIDES	241
7.3.1	Polyacrylamide Gel Electrophoresis (PAGE)	241
7.3.2	Gel Shift Mobility Assays (Native Gels)	242
7.3.3	Visualization of Oligonucleotides	243
7.3.4	Ion-Exchange HPLC	243
7.3.5	Desalting of Oligonucleotides	244
7.4	CHARACTERIZATION OF OLIGONUCLEOTIDES	247
7.4.1	Purity Check	247
7.4.2	MALDI-TOF Mass Spectrometry	247
7.4.3	Hybridization Properties	247
7.5	HAIRPIN STRUCTURE DETERMINATION	251
7.5.1	NMR Samples: Synthesis and Preparation	251
7.5.2	NMR Spectroscopy	252
7.5.3	Structural Modeling	253
7.6	BIOLOGICAL STUDIES	254
7.6.1	Materials	254
7.6.2	Preparation of HIV-1 RT	254
7.6.3	5' End ³² P Labeling Assay of oligonucleotides	255
7.6.4	HIV-1 Reverse Transcriptase RNase H Inhibition Assay	255
7.6.5	<i>E.coli</i> and Human RNase H Inhibition Assays	256
7.6.6	HIV-1 Reverse Transcriptase RNA-Dependent DNA Polymerase Assay	256
7.6.7	HIV-1 Reverse Transcriptase DNA-Dependent DNA Polymerase Assay	257
7.6.8	Gel-Shift Binding Assay	257
7.6.9	UV Cross-Linking Study	258
7.6.10	Stability of Hairpins in Biological Media	258
7.7	RNase III	259
7.7.1	Materials	259
7.7.2	Synthesis of Rntp Hairpin Substrates	260
7.7.3	Gel-Mobility Shift Assay	260
7.7.4	Kinetic Parameters Determination	261
7.7.5	Enzymatic Cleavage Assay	261
	References	262

ABBREVIATIONS AND SYMBOLS

A	adenosine
Å	angstrom
A ₂₆₀	UV absorbance at 260 nm
ACN	acetonitrile
Ade	adenine
AIDS	acquired immunodeficiency syndrome
APS	ammonium persulfate
bp	base pair
BIS	N,N'-methylene-bis(acrylamide)
Bn	benzyl
BPB	bromophenol blue
Bz	benzoyl
C	Celsius
CD	circular dichroism
COSY	correlated spectroscopy
CPG	controlled pore glass
Cyt	cytosine
DCE	1,2-dichloroethane
DCI	4,5-dicyanoimidazole
DIPEA	N,N'-diisopropylethylamine
DMAP	N,N-dimethyl-4-aminopyridine
DMF	N,N-dimethylformamide
DMSO	dimethylsulfoxide
DMT	4,4'-dimethoxytrityl
dN	2'-deoxynucleotide
DNA	2'-deoxyribonucleic acid
DTT	dithiothreitol
ds	double-stranded
dsRBD	double-stranded binding domain
dsRBP	double-stranded RNA binding protein
<i>E.coli</i>	<i>Escherichia coli</i>
EDTA	disodium ethylenediaminetetracetate dihydrate
e.g.	for example
EtOH	ethanol
FAB-MS	fast atom bombardment mass spectroscopy
G	guanosine
ΔG°	standard Gibbs free energy of formation
g	gram
Gua	guanine
ΔH°	standard enthalpy change
HIV-1	human immunodeficiency virus type-1
HOAc	glacial acetic acid
HPLC	high performance liquid chromatography
h	hour(s)

%H	percentage hypochromicity
i.e.	that is
i-Bu	isobutyryl
i-motif	intercalative motif
J	coupling constant
λ	wavelength
LCAA-CPG	long-chain alkylamine controlled pore glass
M	molar
MALDI-TOF	matrix-assisted laser Desorption ionization time of flight
max	maximum
Me	methyl
MeOH	methanol
min	minute(s)
ml	milliliter
mg	milligram
mM	millimolar
μ M	micromolar
MMTr	monomethoxytrityl
mol	mole
mRNA	messenger RNA
MS	mass spectroscopy
MW	molecular weight
NBA	4-nitrobenzyl alcohol
nm	nanometer
NMR	nuclear magnetic resonance
NOE	nuclear Overhauser enhancement
NOESY	Nuclear Overhauser and Exchange Spectroscopy
nt	nucleotide(s)
OD	optical density
Pac1	fission yeast RNase III
PAGE	polyacrylamide gel electrophoresis
ppm	parts per million
pre	precursor
Pu	purine
Py	pyrimidine
R	universal gas constant
rN	ribonucleotide
RNA	ribonucleic acid
RNAi	RNA interference
RNase	ribonuclease
rRNA	ribosomal RNA
Rnt1p	yeast RNase III
RT	reverse transcriptase
ΔS°	standard entropy change
sec	second
snRNA	small nucleolar RNA

T	thymidine
t	tertiary
TBDMS	<i>t</i> -butyldimethylsilyl
TBE	TRIS/boric acid/EDTA buffer
TCA	trichloroacetic acid
TEA	triethylamine
TEMED	N,N,N,N-tetramethylethylenediamine
Thy	thymine
THF	tetrahydrofuran
TLC	thin layer chromatography
T_m	melting temperature
TM	trademark
Tr	trityl
TREATHF	triethylamine-tris(hydrogen fluoride)
TRIS	2-amino-2-(hydroxymethyl)-1,3-propanediol
tRNA	transfer RNA
U	uridine
UUCG	RNA tetraloop of the sequence UUCG with 3',5'-phosphodiester linkages
<u>UUCG</u>	RNA tetraloop of the sequence UUCG with 2',5'-phosphodiester linkages
Ura	uracil
UV	ultraviolet
UV-VIS	ultraviolet-visible
v/v	volume by volume
w/v	weight by volume
XC	xylene cyanol

LIST OF FIGURES

Figure 1.1	DNA autograph by Francis Crick and James Watson.	1
Figure 1.2	Schematic representation of the structure of DNA and RNA and the Watson-Crick pairing complementarity.	2
Figure 1.3	Definitions of torsion angles in nucleotides.	4
Figure 1.4	Structures and average parameters comparing A-form and B-form helical conformations adopted by nucleic acids.	5
Figure 1.5 A	Surface of the CPG solid support derivatized with a nucleoside.	7
Figure 1.5 B	Protecting groups for DNA building blocks.	7
Figure 1.5 C	Protecting groups for the nucleobase exocyclic amine groups.	7
Figure 1.6	Cycle for oligonucleotide solid-phase synthesis according to the phosphate trimer approach	9
Figure 1.7	Schematic diagrams of RNA secondary structure elements.	13
Figure 1.8	G·U wobble pairing motif.	14
Figure 1.9	Secondary structure model of <i>E.coli</i> 16S rRNA.	16
Figure 1.10 A	NMR-derived structure of an unusually stable hairpin.	19
Figure 1.10 B	Structure of the (UUCG) tetraloop.	20
Figure 1.11	Morphology of the UG reverse wobble base pair evident in the (UUCG) tetraloop.	20
Figure 1.12	Unusual backbone conformation and cytosine-phosphate contact observed within (UUCG) tetraloops.	21
Figure 1.13	Proposed mechanism of RNA interference.	26
Figure 1.14	HIV replication cycle showing various classes of antiretroviral drugs as well as their stages of intervention.	29
Figure 1.15	Crystal structure of HIV-1 RT.	30
Figure 1.16	RNase H domain from the crystal structure of HIV-1 RT.	30
Figure 1.17	Chemical structures of (A) 2',5'-DNA and (B) 2',5'-RNA.	31

Figure 1.18	Chemical structure of lariat showing the 2',5'-phosphodiester linkage.	32
Figure 1.19	Chemical structure of 2',5'-linked thioformacetal nucleic acids.	36
Figure 1.20	NMR-derived structure of a 2',5'-DNA duplex.	37
Figure 1.21	Chemical structures of mixed backbone oligonucleotides composed of (A) 2',5'-DNA/3',5'-DNA and (B) 2',5'-RNA/3',5'-DNA.	39
Figure 1.22	Preferred sugar conformations of 2',5'-RNA and 3',5'-RNA.	41
Figure 1.23	NMR-derived structure of a 2',5'-RNA duplex [(GCCGCGGC) ₂].	42
Figure 2.1	Secondary structure and base sequence of the hairpin under study.	50
Figure 2.2	Phosphoramidite monomers used in the solid phase synthesis of (A) 3',5'-RNA and (B) 2',5'-RNA.	51
Figure 2.3	Chemical structures of various coupling reagents used in oligonucleotide solid phase chemistry.	53
Figure 2.4	Preparative polyacrylamide gel electrophoresis (25%) showing the formation of "long-mers" from representative 2',5'-RNA synthesis using 5-ethylthio-1H-tetrazole.	53
Figure 2.5	Analytical polyacrylamide (24%) denaturing gel showing representative purified oligonucleotides.	54
Figure 2.6	1D-NMR spectra of various representative hairpins showing (A) the aromatic [7-8 ppm] and (B) the sugar H1' regions.	55
Figure 2.7	Schematic diagram illustrating the competing dynamic processes between hairpin and dimer states.	56
Figure 2.8	Van't Hoff plots of representative (A) RNA and (B) DNA stem hairpins differing in loop composition (DNA <i>versus</i> RNA <i>versus</i> 2',5'-RNA loops).	57
Figure 2.9	Typical UV melting profile of a sample hairpin showing the ordered (low temperature region) and disordered (high temperature region) states.	60

Figure 2.10 A	Plot of α versus temperature (derived from the UV melting curve).	60
Figure 2.10 B	Van't Hoff plot of $\ln K$ versus $1/T$. The linear line is extrapolated to both ends of the curve.	61
Figure 2.11 A	Normalized melting temperature profiles of RNA hairpins with various loop compositions.	64
Figure 2.11 B	Circular dichroism spectra at 22 °C of RNA hairpins with various loop compositions.	64
Figure 2.12	Normalized melting temperature profiles of DNA hairpin with various loop compositions.	66
Figure 2.13	Relative Gibbs free energies of formation ($\Delta\Delta G^\circ_{37}$) in kcal/mol for hairpins with different stem combinations but with (A) 3',5'-RNA loops and with (B) 2',5'-RNA loops.	70
Figure 2.14	Circular dichroism spectra at 22 °C of DNA and RNA hairpins of various loop compositions: (1) 2',5'-r(UpUpCpGp) loops; (2) 3',5'-r(UpUpCpGp) loops; and (3) d(TTCG) loops.	72
Figure 2.15	CD spectra of DDD, RRR and <u>RRR</u> at 22 °C.	73
Figure 2.16	Thermal melting curves of RNA hairpin duplexes containing deoxynucleotide sugar residues at the C·G loop-closing base pair ($\sim 4.5 \mu\text{M}$).	76
Figure 2.17	Circular dichroic spectra of RNA hairpins differing in the sugar composition of the C·G loop-closing base pair.	77
Figure 2.18	Thermal melting curves of DNA hairpin duplexes containing ribose substitutions at the C·G loop-closing base pair ($\sim 4.5 \mu\text{M}$).	79
Figure 2.19	Circular dichroic spectra of DNA hairpins differing in the sugar composition of the C·G loop-closing base pair.	80
Figure 2.20	Schematic representation of the three hairpin sequences under study.	81
Figure 2.21	Expanded plots of NOESY spectra at 500 MHz.	83
Figure 2.22	Conserved structure of the 2',5'-RNA loop showing U·G wobble base pairing, U6 stacking on U5, and C7 protruding out into the solvent.	84

Figure 2.23	Conserved structure of the 2',5'-RNA loop showing unique NOEs between U5, G8 and G9 residues.	84
Figure 2.24	Conserved structure of the 2',5'-RNA loop showing NOEs of imino G8 with H1'/H4'/H5'(U6).	85
Figure 2.25	U·G Wobble base pairing motif observed in conserved structures of 2',5'-RNA loop.	85
Figure 2.26	Base-base and sugar-base contacts within the conserved 2',5'-RNA loop structure.	86
Figure 2.27	Stereoview of the NMR-derived structure of <u>RRR</u> hairpin.	88
Figure 2.28	Stereoview of the NMR-derived structure of <u>DRD</u> hairpin.	89
Figure 2.29	Structural comparison of the general unique features of (A) 3',5'-linked ^{53,54,55} versus (B) 2',5'-linked (UUCG) loops.	93
Figure 3.1	HIV-1 RT RNase H-mediated degradation of RNA in an RNA:DNA hybrid duplex.	98
Figure 3.2	Representative examples of the three classes of HIV-RT RNase H inhibitors.	99
Figure 3.3	Schematic representation of the mode of inhibition of HIV-1 RT RNase H-mediated degradation of viral RNA by small-molecule hairpin aptamers.	102
Figure 3.4	Secondary structure and base sequence of the hairpin constituting the building core for library generation.	104
Figure 3.5	Protocol for the <i>diversity-generating</i> combinatorial synthesis of a 27-member hairpin library.	105
Figure 3.6	Expedite™ 8909 DNA/RNA synthesizer used in diversity-oriented combinatorial synthesis.	106
Figure 3.7	Van't Hoff plots of representative hairpin library members.	108
Figure 3.8	Conformation spectrum generated by diversity-oriented synthesis.	110
Figure 3.9	Gel autoradiogram illustrating RNase H inhibitory activity of representative hairpin library members.	113

Figure 3.10	Inhibition of HIV-1 RT RNase H activity by representative hairpins with either R or <u>R</u> loops.	115
Figure 3.11	Specific inhibition of <i>in vitro</i> HIV-1 RT RNase H activity.	116
Figure 3.12	Integrity of hairpin aptamers towards RNase H activity.	117
Figure 3.13	3-D Graph showing % Inhibition {at 40 μ M [aptamer]} <i>versus</i> helical conformation for hairpins with various loops and having the same stem compositions.	120
Figure 3.14	3-D Graph showing % Inhibition {at 40 μ M [aptamer]} <i>versus</i> helical conformation showing effects of loop base sequence, point insertions and stem length.	121
Figure 3.15	Variation of inhibitory constant (IC ₅₀) with melting temperature.	124
Figure 3.16	Gel-shift binding assay to determine formation of [HIV-1 RT]:[Hairpin] complex.	126
Figure 3.17	DNA synthesis catalyzed by HIV-RT DNA- or RNA-dependent DNA polymerase activity.	127
Figure 3.18	Inhibition of DNA synthesis catalyzed by (A) DNA-dependent DNA polymerase (B) RNA-dependent DNA polymerase activities of HIV-1 RT.	128
Figure 3.19	UV Cross-linking experiment of hairpin R ₆ RR ₆ with HIV-1 RT.	130
Figure 3.20	%Remaining hairpin after incubation with 0.5 X reticulocyte lysate for 18 hours at 37 °C.	132
Figure 3.21	Plots of absorbance (at 260 nm) <i>versus</i> time of exposure of hairpins to Nuclease P1.	133
Figure 3.22	Denaturing polyacrylamide gel electrophoresis showing a representative digestion pattern by Nuclease P1.	135
Figure 4.1 A	Schematic representation comparing the functional domains between bacteria and yeast RNases III.	141
Figure 4.1 B	Proposed hypothetical model for the role of N-terminal domain in Rnt1p cleavage efficiency.	141

Figure 4.2	Short hairpin substrates (22 nucleotides in length) designed to probe the binding specificity of Rnt1p.	144
Figure 4.3	Normalized melting temperature profiles (at 260 nm) of representative hairpin substrates.	145
Figure 4.4	Circular dichroism spectra at 22 °C of representative hairpins of mixed stem composition.	146
Figure 4.5	Gel shift binding assay of short hairpin substrates (9 bp stems).	147
Figure 4.6	Plot of % shifted hairpin versus Rnt1p concentration as obtained from quantitating the gel bands in Figure 4.5 .	148
Figure 4.7	Schematic representation comparing the binding constant (K_d), thermal stability (T_m) and %hypochromicity (%H) of various RNA hairpins differing only in the loop base-sequence.	150
Figure 4.8	Long hairpin substrates (42 nucleotides in length) designed to probe the cleavage reaction by Rnt1p.	152
Figure 4.9	Gel shift binding assay of long hairpin substrates (19 bp stems).	153
Figure 4.10	Cleavage assay of long hairpins (42-nt) with various stem compositions.	154
Figure 4.11 A	Normalized melting temperature profiles (at 260 nm) of representative long hairpin substrates.	155
Figure 4.11 B	Circular dichroism spectra at 22 °C of representative hairpins of mixed stem composition.	155
Figure 4.12	RNase treatment of D ₁₉ LR ₁₉ , R ₁₉ LD ₁₉ , and R ₁₉ LR ₁₉ substrates in the presence and absence of Rnt1p.	156
Figure 4.13	The product from the cleavage assay of D ₁₉ LR ₁₉ run side by side with a synthetically-made 5-nucleotide DNA fragment of the sequence 5'-ggcgt-3' in the (+) presence of Rnt1p and (-) absence of Rnt1p.	157
Figure 4.14 A	Schematic diagram showing the substrate cleavage site and potential chemical modifications to probe the cleavage reaction.	159

Figure 4.14 B	Long hairpin substrates (42-mer) containing chemically-modified inserts and designed to probe the cleavage reaction by Rnt1p.	160
Figure 4.15	Circular dichroism spectra at 22 °C of hairpins with chemically-modified inserts.	161
Figure 4.16	Gel shift binding assay of hairpin substrates containing chemical modifications at the cleavage site.	161
Figure 4.17	Cleavage assay of hairpins with chemically modified inserts at the cleavage site.	162
Figure 4.18	Proposed mechanism of nucleophilic cleavage of RNA.	163
Figure 4.19	Chimeric DNA/RNA hairpin substrates (42 nucleotides in length) designed to probe the cleavage reaction by Rnt1p.	166
Figure 4.20	(A) Plot of % bound hairpin <i>versus</i> Rnt1p concentration and (B) cleavage assay of various long hairpin DNA/RNA chimeras (42-nt).	167
Figure 4.21	The base sequence of the control all-RNA substrate [U5-Like, 5.18] and its modified DNA:RNA hybrid analogue [^{D/R} U5-Like, 5.19].	168
Figure 4.22	Cleavage assay of U5-like hairpins with various RNases III.	169
Figure 4.23	Schematic representation of the hairpin substrate containing a 2',5'-phosphodiester linkage at the cleavage site (14 nucleotides away from the loop on the 5'-end stem).	170
Figure 4.24	Cleavage assay of the hairpin [R _{19(2'-FU)} LR ₁₉] with a 2',5'-phosphodiester linkage at the cleavage site.	171
Figure 5.1	Schematic representation of a hemiprotonated C·C ⁺ base pair.	175
Figure 5.2 A	Schematic representation of i-motif DNA (C-tetrad).	176
Figure 5.2 B	Stereoview of the NMR-derived structure of an i-motif DNA (C-tetrad).	177

Figure 5.3 A and B	Representative diagrams showing topologies and stoichiometries of various possible i-motif structures.	179
Figure 5.3 C	Representative diagram showing intramolecular i-motif formation.	180
Figure 5.4	UV melting curves of representative deoxycytidine rich oligomers.	183
Figure 5.5	Plots of T_m versus pH for deoxycytidine rich oligomers containing G insertions.	186
Figure 5.6	Reversible UV melting curves comparing the behavior of control dC ₅ and c ₅ Rgc ₄ [5.5] at pH 5.0.	190
Figure 5.7	Reversible UV melting curves comparing behavior of representative deoxycytidine rich oligomers at pH 7.0.	191
Figure 5.8	Van't Hoff plots showing T_m concentration-dependence of various deoxycytidine-rich oligomers with (UUCG) loops.	192
Figure 5.9 A and B	Circular dichroism spectra of dC ₅ and c ₅ Rc ₅ at 5 °C and 65 °C.	194
Figure 5.9 C and D	Circular dichroism spectra of c ₅ Rc ₅ and c ₅ Rgc ₄ at 5 °C and 65 °C.	195
Figure 5.9 E and F	Circular dichroism spectra of c ₅ Rgc ₄ and c ₅ Ugc ₄ at 5 °C and 65 °C.	196
Figure 5.10	Circular dichroism spectra of dC ₅ and c ₅ Rgc ₄ at various pH.	197
Figure 5.11	Gel mobility shift assay (native conditions) of various cytosine-rich oligomers.	198
Figure 5.12	Proposed tetrad structures for i-motif formation by hairpin oligomers.	200
Figure 5.13	UV thermal melting data for ribocytidine-rich oligomers	205
Figure 5.14	Plots of T_m versus pH for ribocytidine oligomers.	206
Figure 5.15	Reversible UV melting behavior of C ₅ RC ₅ [5.10] and C ₁₀ (R)GC ₉ [5.11] at pH 5.0 and 7.0 respectively.	209

Figure 5.16	Van't Hoff plots showing T_m independence of ribocytidine-rich oligomers with (UUCG) loops.	210
Figure 5.17	Gel mobility shift assay (native conditions) of various cytosine-rich oligomers.	211
Figure 5.18	Circular dichroism spectra of ribocytidine-rich oligomers at 5 °C.	213
Figure 5.19 A and B	Circular dichroism spectra of c_9Rgc_8 and $C_{10}(R)GC_9$ at 5 °C and 70 °C.	214
Figure 5.19 C and D	Circular dichroism spectra of $C_{10}(R)GC_9$ and an RNA hairpin duplex at 5 °C and 70 °C.	215
Figure 5.19 E and F	Circular dichroism spectra of DNA and RNA:DNA hairpin duplex controls at 5 °C and 70 °C.	216
Figure 5.20	Circular dichroism spectra of c_9Rgc_8 and $C_{10}(R)GC_9$ at various pH.	218
Figure 5.21	UV melting curve of $C_{10}RGC_9$ [5.13].	220
Figure 5.22	UV melting curves of $C_{10}RGC_9$ as a function of pH.	220
Figure 5.23	Circular dichroism spectra of the all 2',5'-RNA cytidine oligomer at 5 °C.	223
Figure 5.24	Circular dichroism spectra of $C_{10}RGC_9$ at 5 °C and 70 °C.	224
Figure 5.25	Circular dichroism spectra of $C_{10}RGC_9$ at various pH.	224
Figure 5.26	Gel mobility shift assay (native conditions) of $C_{10}RGC_9$.	225
Figure 5.27	UV thermal melting data for representative cytidine oligomers under physiological conditions.	228
Figure 7.1	Analytical polyacrylamide (24%) denaturing gel showing representative purified oligonucleotides.	247
Figure 7.2	Representative anion-exchange HPLC profile for pure sample.	247

LIST OF TABLES

Table 1.1	UV melting temperatures and Gibbs free energies of formation for various unusually stable RNA hairpins	25
Table 1.2	Chronological summary of the biophysical properties of 2',5'-nucleic acids and the development of 2',5'-nucleic acid chemistry	43
Table 2.1	Thermodynamic parameters of hairpins with RNA stems	62
Table 2.2	Thermodynamic parameters of hairpins with DNA stems	67
Table 2.3	Thermodynamic parameters of hairpins with DNA:RNA, DNA:2',5'-RNA, RNA:2',5'-RNA, and 2',5'-RNA:2',5'-RNA stem hybrids	69
Table 2.4	Thermodynamic parameters for RNA-stem hairpins with deoxy insertions in the loop-closing base pair	74
Table 2.5	Thermodynamic parameters for DNA-stem hairpins with ribose insertions in the loop-closing base pair	75
Table 2.6	Thermodynamic parameters of hairpins under NMR study	91
Table 3.1	Diversity-oriented library synthesis: Representative 16 hairpin chimeras obtained by combinatorial split synthesis	107
Table 3.2	Inhibition of RNase H activity of HIV-1 RT by a nucleic-acid hairpin library	112
Table 3.3	Test of RNase H inhibitory activity of short linear oligomers	118
Table 3.4	Relative half-life times of representative hairpins towards Nuclease P1 digestion	134
Table 3.5	Thermally-induced melting data for pseudo hairpin formation	136
Table 4.1	Binding constants (K_d) of various short hairpin substrates towards Rnt1p.	149
Table 4.2	Long hairpin substrates (42-nt) containing chemically-modified inserts at the cleavage site	160
Table 5.1	Sequences of deoxycytidine rich oligomers under study	181

Table 5.2	UV thermal melting data for deoxycytidine rich oligomers	183
Table 5.3	UV melting data of deoxycytidine rich oligomers under study	185
Table 5.4	Base sequence of cytidine-rich oligomers under study	203
Table 5.5	UV melting data of ribocytidine rich oligomers under study	207
Table 5.6	UV melting data for the all-2',5'-RNA cytidine-rich oligomer [5.13]	221
Table 5.7	Inhibitory constants against HIV-1 RT RNase H of cytidine-rich oligomers	226
Table 7.1	Comparison of nucleoside loadings	238
Table 7.2	Reported yields from solid-phase synthesis	246
Table 7.3	MALDI-TOF spectrometry analysis and calculated extinction coefficients	249
Table 7.4	Reported yields for solid-phase synthesis of NMR hairpins	252
Table 7.5	Synthesis yields obtained from solid-phase synthesis of Rnt1p substrates	260

CHAPTER I: GENERAL INTRODUCTION

1.1 NUCLEIC ACIDS^{1,2}

Interdisciplinary science has had a major impact on the development of human culture. Within the second half of the twentieth century, scientific disciplines began to amalgamate, thus bringing forth a new form of 'hybrid' science referred to as "science at the interface." The field of nucleic acids is a fruit of the progressive intermingling of two disciplines, namely chemistry and biology.

In 1953, James Watson and Francis Crick 'gave birth' to the DNA double helix.³ Their monumental discovery marked the beginning of the era of molecular biology. DNA, in its native state, is a right handed double helix composed of two sugar-phosphate strands that are wound around each other on the exterior in an antiparallel orientation (**Figure 1.1**). This structure is stabilized by intrinsic hydrogen bonds, hydrophobic interactions and most of all stacking of base pairs that lie in between the two strands on the interior.

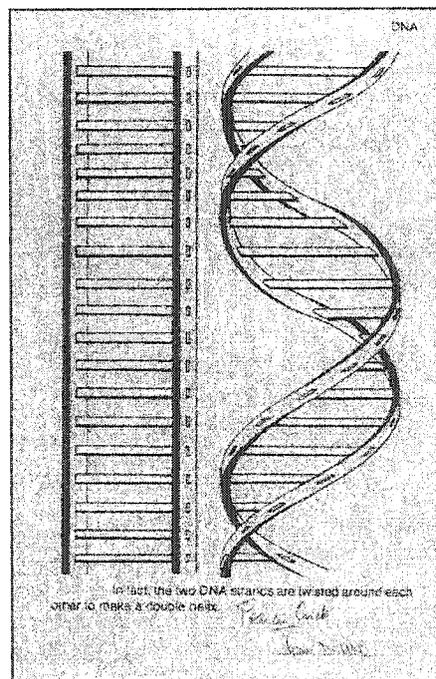


Figure 1.1: DNA autograph by Francis Crick and James Watson.

Base recognition within DNA disseminates via a specific Watson-Crick pairing motif: Adenine pairs with thymine via two hydrogen bonds, while guanine pairs with cytosine via three hydrogen bonds (Figure 1.2). The sequence of the bases within DNA encodes genetic information that constitutes the basis of heredity and evolution. DNA consists of a chain of 2-deoxy-D-ribose repeats, joined together by 3',5'-phosphodiester linkages, carrying one of four possible heterocyclic bases in a β configuration at their sugar 1' positions (Figure 1.2). The combination of a phosphate group, a furanose ring, and a heterocyclic base constitutes what is called a 'nucleotide'. The heterocyclic bases are planar and can be either pyrimidines (monocyclic rings) or purines (fused heterocyclic rings).

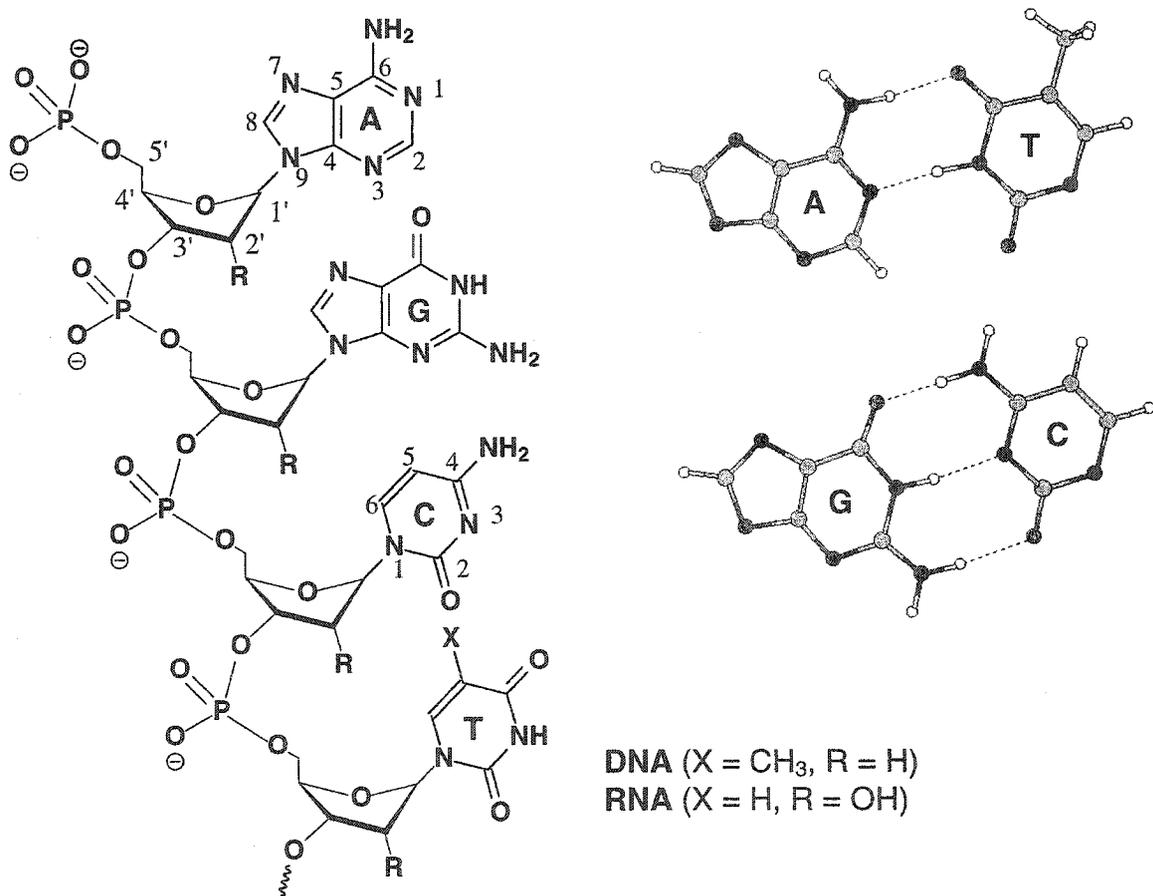


Figure 1.2: Schematic representation of the structure of DNA and RNA and the Watson-Crick pairing complementarity. Nitrogen is shown in blue, oxygen in red, carbon in cyan, and hydrogen in white.

Ribonucleic acid (RNA) is another genetic-carrier of information in cells. It exists mostly in single-stranded form while DNA usually exists in the double stranded form. RNA has the same primary structure as DNA except that the sugar is ribose instead of 2'-deoxyribose. It contains uracil instead of thymine which is found only in DNA. All other bases (adenine, guanine and cytosine) are found in both DNA and RNA. In RNA as well as in DNA, the nucleic acid strand has a chemical orientation or polarity that extends from the 5'-end (with a phosphate group) to the 3'-end (with a free hydroxyl group) or vice versa.

Conformational Features of Nucleotides¹

Glycosidic and backbone torsion angles help identify the intrinsic conformations in a nucleotide unit (**Figure 1.3**).

Sugar pucker is specified by the torsion angle δ . As in the case for cyclopentane, energetic factors do not allow the furanose ring to be planar. It rather adopts an *envelope* conformation in which four atoms lie in the same plane and the fifth is either above or below this plane. Another conformation is the *twist* form in which three adjacent atoms are coplanar and the remaining two adjacent atoms are either above or below the plane of the other three atoms. This 'puckering' is depicted from the major displacement of carbons 2' and 3' relative to the median plane of C1'-O4'-C4' (**Figure 1.3**). In RNA helices the C-3' *endo* pucker conformation of ribose is favored, while sugars in DNA can be in dynamic equilibrium and adopt more than one conformation depending on the helical structure (discussed below).

Base orientation is inferred from the glycosidic torsion angle χ . In the *anti* conformation, the bulk of the heterocyclic base (either the six-membered ring in purines or the O2 in pyrimidines) is pointing away from the sugar ring; in the *syn* conformation the base is rotated by approximately 180° and is positioned over the sugar ring. Typically, bases adopt the *anti* conformation because it is more energetically favored than the *syn* conformation.

The other torsion angles (**Figure 1.3**) specify the folding of the sugar-phosphate backbone.

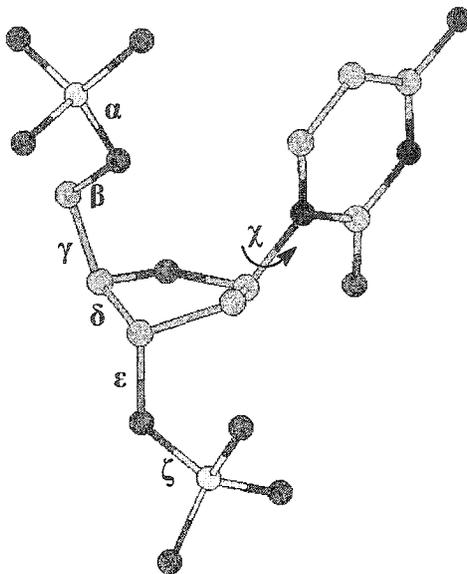
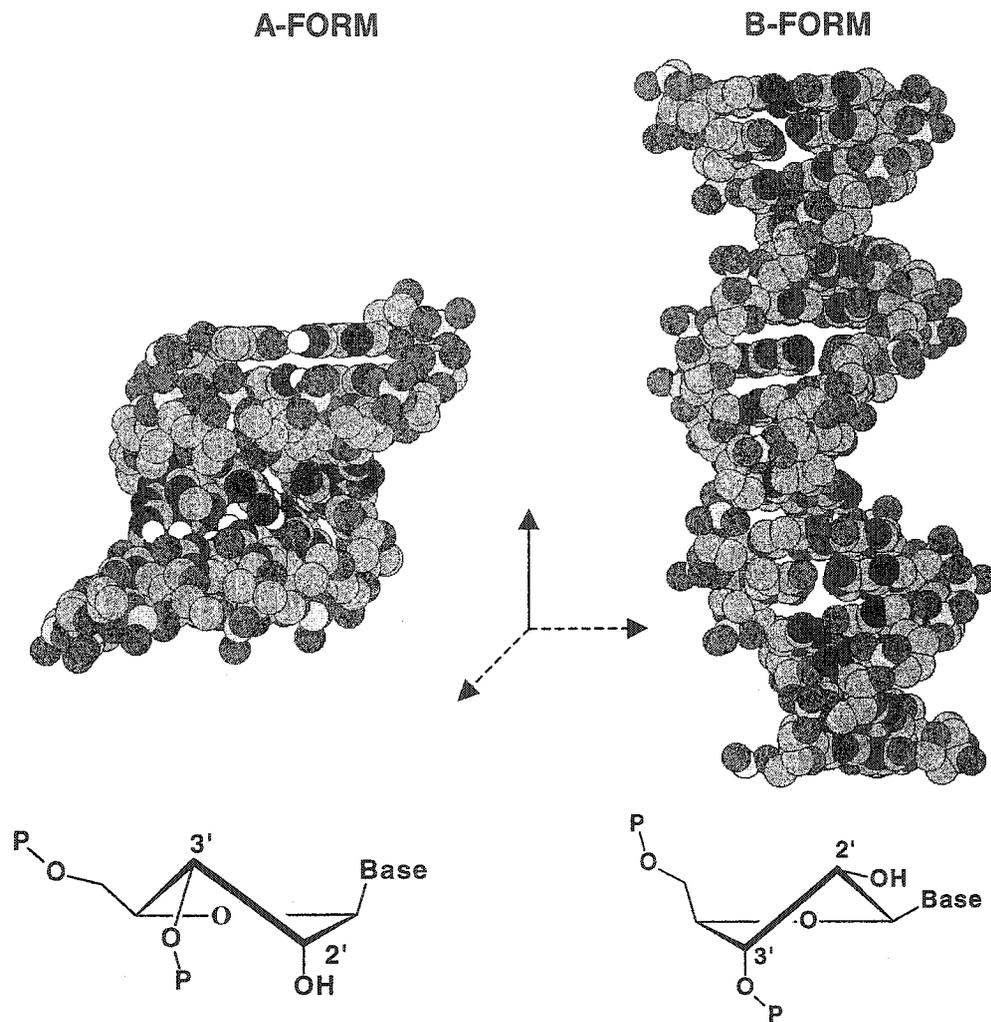


Figure 1.3: Definitions of torsion angles in nucleotides. Nitrogen is in blue, carbon in cyan, oxygen in red, and phosphorus in yellow. Hydrogen atoms are removed for clarity.

The Double Helical Forms of Nucleic Acids¹

X-ray diffraction studies on fibers have demonstrated that there exist at least two conformations for the DNA double helix. At high salt concentration (or low humidity), the duplex assumes the so-called “A-form”, whereas at high humidity (or low salt concentration) the duplex adopts the “B-form” helical conformation. The general characteristics of each type of duplex are outlined in **Figure 1.4**. In both helix forms, the interwinding of the sugar phosphate strands creates spaces on the outside of the helix. These identify two distinct grooves of different dimensions referred to as the major and minor grooves.

A-form DNA exhibits 11 base pairs per helical turn, and the base pairs are tilted with respect to the helical axis. The sugar residues are aligned parallel to the helical axis and adopt the C-3' *endo* sugar pucker. The minor groove is broad and shallow while the major groove is narrow and deep.



Helicity	Right	Right
Residues/turn	11	10.4
Base Tilt	20°	-6°
Rise/base pair (bp)	2.56 Å	3.3-3.4 Å
Twist/base pair	32.7 Å	36 Å
Displacement/base pair	4.5 Å	-0.2 to -0.8 Å
Sugar Pucker	C-3' <i>endo</i>	C-2' <i>endo</i>
Minor Groove	Broad and Shallow	Narrow and Deep
Major Groove	Narrow and deep	Wide and deep
Helix Diameter	25.5 Å	23.7 Å

Figure 1.4: Structures and average parameters comparing A-form and B-form helical conformations adopted by nucleic acids. Adapted from Blackburn, G. M.; Gait, M. J. In *Nucleic Acids in Chemistry and Biology*; Blackburn, G. M., Gait, M. J., Eds.; Oxford University Press: New York, 1990. Nitrogen is shown in blue, oxygen in red, carbon in cyan, and hydrogen in white.

B-form DNA exhibits 10 base pairs per helical turn and they are oriented perpendicular to the helix axis. The sugars adopt the C-2' *endo* pucker. The minor groove is narrow and deep while the major groove is wide and deep.

In general, RNA helices are typically right-handed A-form with 11 base pairs per turn and C-3' *endo* sugars. DNA helices are typically right-handed B-form with 10 base pairs per turn and C-2' *endo* puckers.

RNA:DNA hybrids are intermediate between the A- and B- families.⁴⁻⁶ The sugar residues in the RNA strand adopt a uniform C-3' *endo* pucker repeat, while those in the DNA strand exist as a mixture of conformations that include C-2' *endo*, C-3' *endo* and O-4' *endo* pucker types.^{7,8}

1.2 SOLID-PHASE OLIGONUCLEOTIDE SYNTHESIS

The concept of solid-phase synthesis was initially developed simultaneously by Merrifield⁹ and Letsinger¹⁰ for peptide chemistry, and later on adapted to oligonucleotide synthesis by Letsinger.¹¹ There are four basic principles for this approach: (1) the oligomer is synthesized while covalently attached to an insoluble solid support; (2) excess of soluble protected nucleotides and coupling reagent are employed to drive chain extension to completion; (3) the reaction is done in a single vessel in order to diminish mechanical losses due to solid support manipulation; and (4) the various reactions are standardized and thus can be automated for oligonucleotide chain assembly.

The most popular chemical route for DNA solid-phase synthesis nowadays is the phosphate triester method developed by Letsinger and later modified by Beaucage and Caruthers to what has been known as the phosphoramidite approach.¹²⁻¹⁵ The synthesis starts with the first nucleoside attached to a solid support. The nature of the solid support is essential to the success of synthesis. One type of solid support that has been predominant throughout the last years is controlled pore glass (CPG). It consists of a uniformly prepared glass matrix with pores of defined size. The advantage of this support lies in its ability to retain the same pore size even after exposure to reagents, thus allowing consistency from run to run. Other polystyrene-based supports have been

developed recently and are commercially available. Regardless of the choice of support, the nucleoside is always loaded on the support *via* a linker molecule (Figure 1.5 A).¹⁶

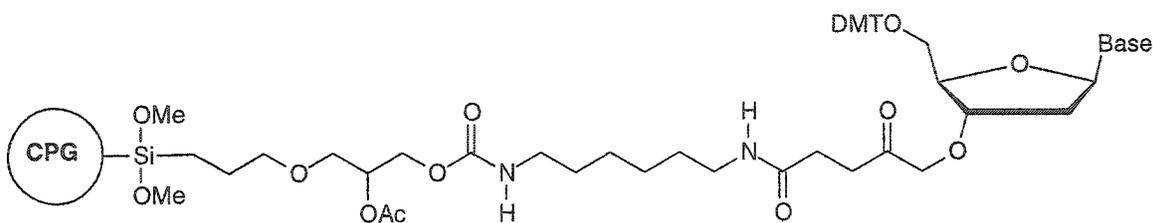


Figure 1.5 A: Surface of the CPG solid support derivatized with a nucleoside.

The protecting groups employed in DNA synthesis can be classified into two types: (1) permanent groups that remain attached to the oligonucleotide throughout synthesis and are removed after completion of chain assembly (base, phosphate, and in the case of RNA, 2'-OH protecting groups), and (2) temporary groups that ensure specificity of a single condensation and are removed immediately thereafter (Figure 1.5 B).^{2,16}

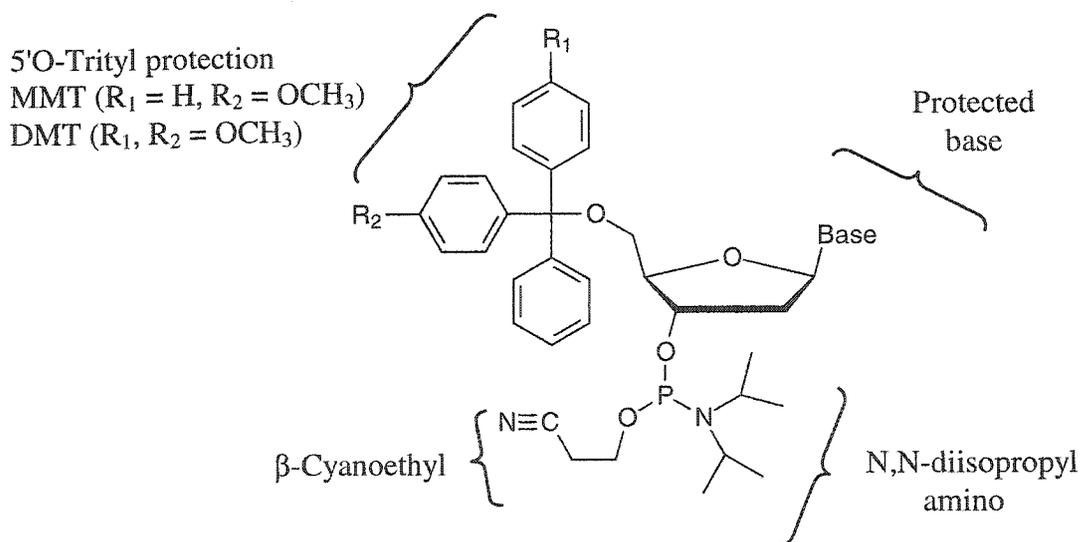


Figure 1.5 B: Protecting groups for DNA building blocks.

Several base protecting groups have been described (permanent protection). The preferred method of nucleobase protection involves acylation of the exocyclic amino

groups of adenine and cytosine (benzoyl groups) and guanine (isobutyryl group) (Figure 1.5 C). The 5'-hydroxyl group is transiently protected with a trityl group (either 4'-monomethoxytrityl [MMT] or more commonly 4',4'-dimethoxytrityl [DMT]) that can be easily removed by mild acid treatment.² The efficiency of deprotection is monitored by spectrophotometry owing to the strong visible absorption of the DMT or MMT groups. Protection of the internucleotidic phosphate bond is accomplished with the β -cyanoethyl group which is removed with mild ammonia treatment. In RNA synthesis, an additional permanent protecting group is required at the 2'-hydroxyl moiety. The most common protecting group remains the *t*-butyl dimethylsilyl (TBDMS) group introduced by Ogilvie and coworkers.^{17,18} This group is acid stable and can be cleaved with a fluoride ion.

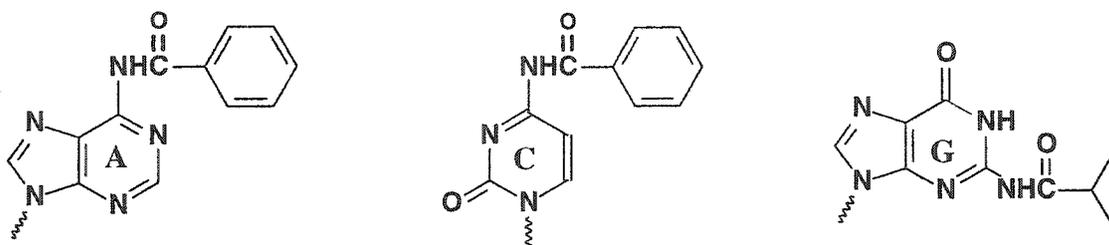


Figure 1.5 C: Protecting groups for the nucleobase exocyclic amine groups.

Solid-phase oligonucleotide synthesis according to the phosphoramidite method is shown in Figure 1.6. Briefly, the first step of the cycle involves cleavage of the 5'-protecting group by acid treatment to afford the free 5'-hydroxyl moiety (“Detritylation”). This group subsequently reacts with a 3'-O-phosphoramidite monomer in the presence of the coupling reagent tetrazole,^{19,20} thus adding an extra unit to the growing chain (“Coupling”). Since the coupling reaction is not quantitative (98-99%), the unreacted 5'-OH groups are acetylated (“Capping”). This aids in isolating the final product by diminishing the side products. Oxidation of the newly formed trivalent phosphite triester is accomplished by the use of a solution of iodine, water and pyridine in tetrahydrofuran to afford the more stable pentavalent phosphate triester.^{12,13} The cycle can be repeated until the desired length of oligonucleotide is obtained. This is followed by cleavage from the support and subsequent base and phosphate deprotection. RNA synthesis requires an additional deprotection step for the 2'-protecting group. In the case of 2'-O-*t*-

butyldimethylsilyl protection, the RNA oligonucleotide is deprotected via fluoride treatment.

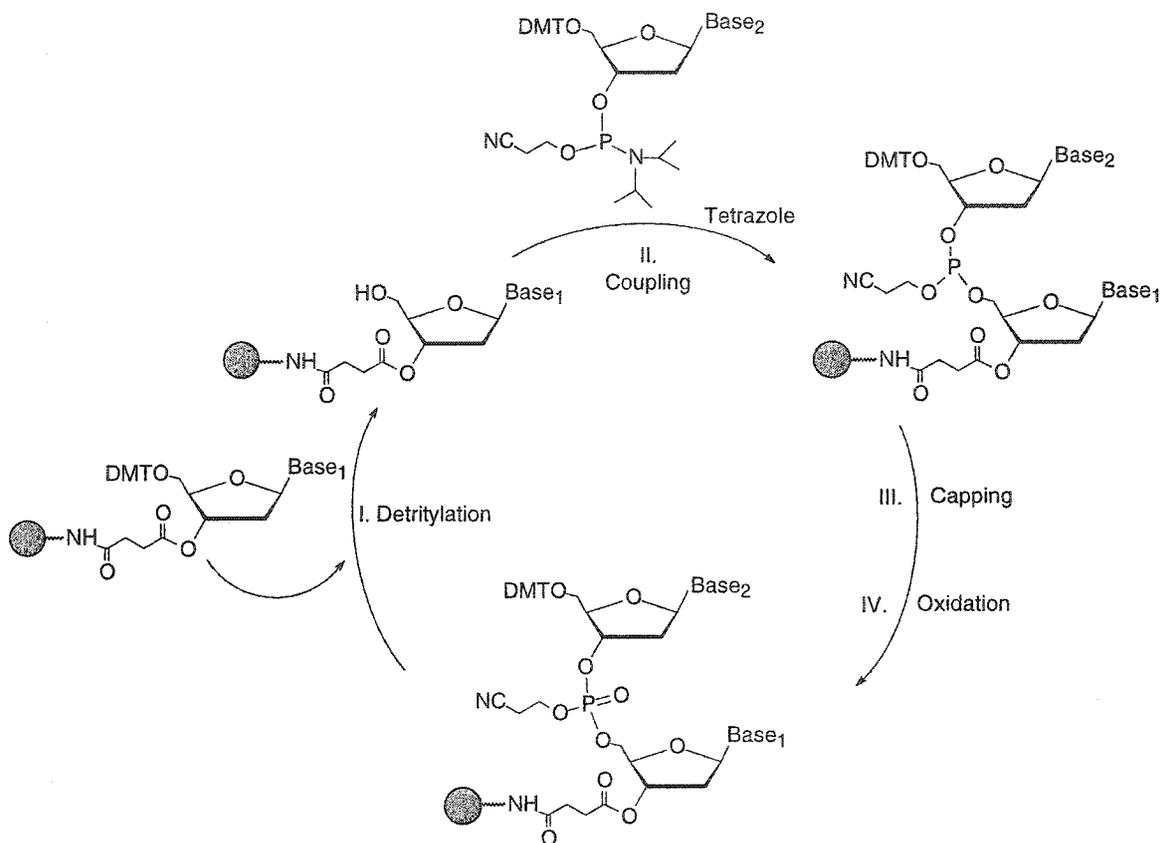


Figure 1.6: Cycle for oligonucleotide solid-phase synthesis according to the phosphate triester approach

1.3 TECHNIQUES USED IN THE STUDY OF NUCLEIC ACIDS²

An array of techniques has been employed in the study of nucleic acid complexes in order to elucidate their hybridization properties, stereochemical characteristics, dynamics and biological features. A few general principles of some of the techniques used in this work are outlined below.

1.3.1 Ultraviolet (UV) Spectroscopy

One of the essential and most commonly used techniques for studying the behavior of nucleic acid complexes is UV spectroscopy. Its versatility as well as utility as a means for deriving characteristic physical parameters for nucleic acid association have been extensively reviewed by Tinoco, Turner, Breslauer and many others.²¹⁻²³ The technique comprises monitoring the thermally-induced transition between nucleic acid native (folded) and denatured (single-stranded) states. The obtained absorbance *versus* temperature profile, called a UV melting curve, yields information about the nature of the thermal denaturation transition for a specific nucleic acid molecule. The midpoint of the transition is referred to as melting temperature, and is designated by T_m . This is the temperature when half of the molecules are in the denatured state and the other half are in the native folded state. As the temperature increases, the ratio of molecules in the denatured *versus* native states increases. Since there is a disparity in absorptivities between the two states, the change in UV absorbance with temperature can effectively monitor the transition process. Whereas other methods such as NMR are capable of monitoring the thermal transition, the convenience of UV spectroscopy lies in the fact that it does not require large amounts of nucleic acids for conducting experiments (1-10 nanomoles of an RNA oligonucleotide).²⁴

The melting profiles of nucleic acid complexes exhibit characteristic increase in their UV absorptions. This phenomenon is associated with unstacking of the stacked bases in the native state with increasing temperature and is referred to as hyperchromicity.²⁵ Alternatively, the reduced absorbance of a nucleic acid relative to its denatured state is referred to as hypochromicity. Hypochromicity is defined as the percent change in absorbance upon denaturation of a nucleic acid species, and is calculated according to the following formula:

$$\text{Hypochromicity (\%)} = 100 (A_D - A_N)/A_D$$

where A_D is the absorbance of the denatured species (at high temperature) and A_N the absorbance of the native species (at low temperature).

The thermodynamic parameters (ΔG° , ΔH° , ΔS°) associated with a thermally-induced transition can be calculated from the UV melting profile.^{21-23,26} This depends on several assumptions of which the most significant is an “all-or-none” or “two-state”

model. This assumes that the nucleic acid species under study exists in only two states, native or single-stranded, and that there are no intermediate states. Furthermore, the dependence of melting temperature on nucleic acid strand concentration reveals the molecularity of the transition process. For instance, melting of hairpin loop structures occurs via a unimolecular process, and thus does not depend on concentration. Melting of a two-strand duplex on the other hand is bimolecular and thus concentration-dependent.^{21-24,27}

1.3.2 Circular Dichroism (CD) Spectroscopy

Circular dichroism (CD) spectroscopy is a very sensitive and qualitative technique for studying the conformation of nucleic acid complexes. It measures the difference in extinction coefficients for right and left circularly polarized light as a function of wavelength. The technique is highly dependent on base sequence and the stacking geometry of the bases within a nucleic acid complex. It can instantly distinguish between helices and unstacked single strands.²⁸ Also, it can differentiate among helical structures and identify A-, B-, and Z-forms. The advantage of this technique is that, like UV spectroscopy, it does not require large amounts of nucleic acids in order to effect their structure characterization. However, unlike NMR, it only provides information about the overall but not detailed conformational features of a nucleic acid complex.

1.3.3 Gel Electrophoresis²⁹

Polyacrylamide gel electrophoresis is an important tool that is commonly used to purify and characterize nucleic acid structures. Denaturing gels are used in order to effect separation of oligonucleotides according to their charge and molecular weight. These gels employ urea, and high temperature (obtained from gel resistance to the applied current) in order to denature any secondary and tertiary structures. Native gels are used to distinguish among multimeric oligonucleotide species mainly according to their structure as well as charge and molecular weight. These are run under similar conditions as the denaturing gels (same percentage cross-linker and total acrylamide concentrations) except that they do not utilize any denaturant (urea) and employ reduced temperature and voltage. Electrophoretic mobility for a given molecule depends on various factors such as

temperature, power level, buffer ionic strength, percentage cross-linking and density of the gel. Tuning of these factors effects better separation of nucleic acids.

1.3.4 Nuclear Magnetic Resonance (NMR) Spectroscopy

Nuclear magnetic resonance (NMR) techniques are widely used to determine the structure of nucleic acid complexes. Considering their momentous contributions towards further understanding of the behavior of nucleic acid structures, this year's Nobel Prize in chemistry was awarded to one of the most prominent protein and nucleic acid NMR spectroscopists (Kurt Wuthrich). NMR measures distances using through-space interactions (nuclear Overhauser effect or NOE) and measures dihedral angles using through-bond interactions (J-coupling).³⁰ The nuclei of interest in NMR of nucleic acids comprise ^1H , ^{31}P , ^{13}C and ^{15}N . Nucleic acids possess two types of protons, exchangeable and non-exchangeable. The latter are bound to carbon atoms, while the former are bound to either nitrogen or oxygen, and exchange rapidly with water (solvent). The exchange rate is affected by several factors such as hydrogen bonding, solvent accessibility, pH and buffer species.³¹ The first step in determining the NMR structure of any nucleic acid consists in assignment of resonances. Protons in nonstandard geometries are grouped by the spin systems (*i.e.*, protons showing sequential J-couplings like those of the ribose sugar), and then by NOEs, which connect the sugars.³² After the resonances have been assigned, the interactions between the nuclei are measured, and consequently converted to distance and angle constraints. Helix geometries (for example, A-type) have characteristic NOEs and J-couplings for proton assignments.^{30,33}

1.4 RNA ORGANIZATION AND ITS STRUCTURAL ELEMENTS

Single-stranded ribonucleic acid structures play a central role in the molecular biology of the cell. Depending on their inherent morphology and the environment conditions they exist in, they can form a plethora of shapes ranging from hairpin loops and bulges to knots, duplexes and overhangs (**Figure 1.7**).²⁷ These constitute primary secondary structural motifs of cellular RNA and display an important role in a variety of biological processes.³⁴ They also facilitate the occurrence of tertiary interactions which help fold the RNA into its final three dimensional structure.³⁵

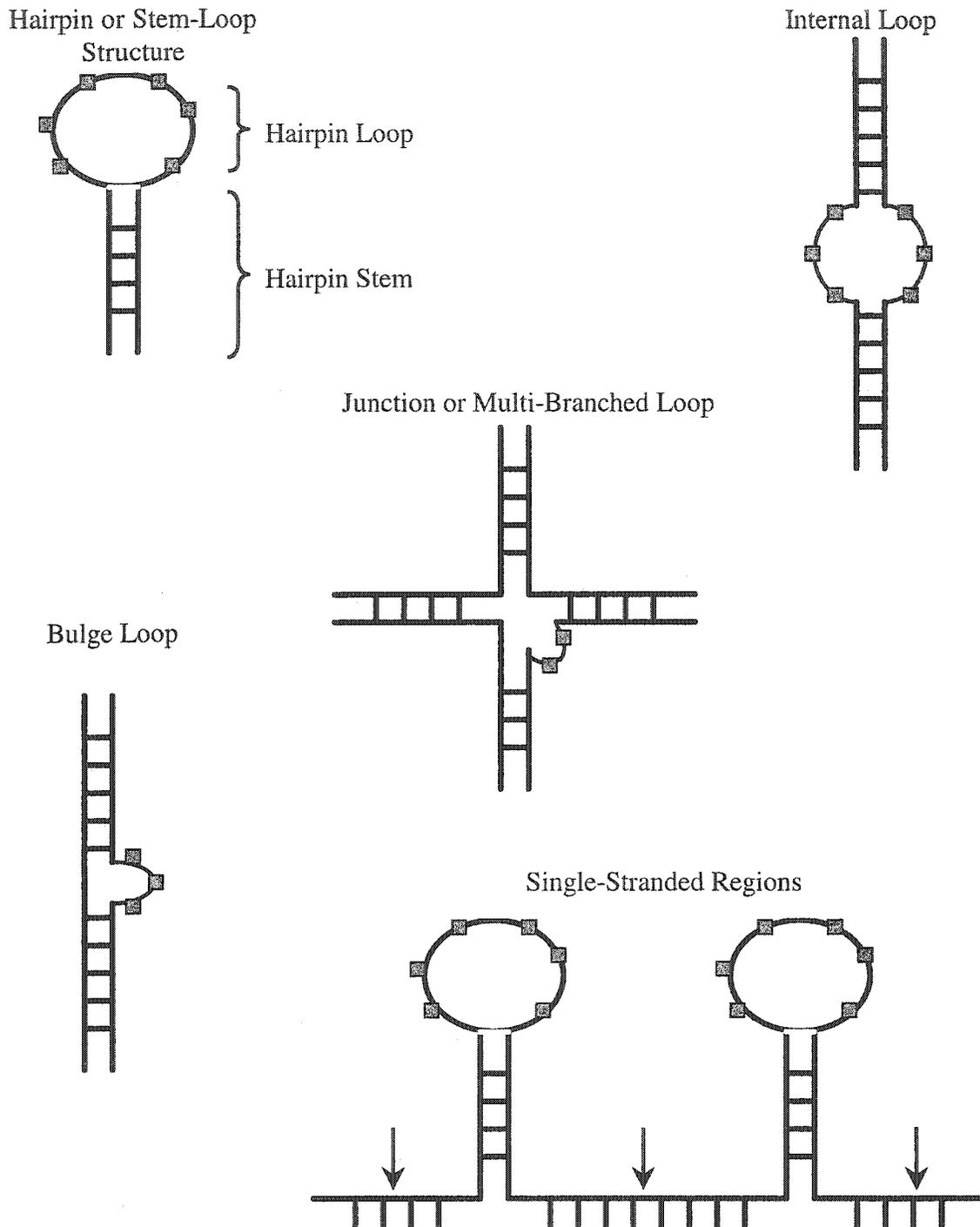


Figure 1.7: Schematic diagrams of RNA secondary structure elements. Single stranded RNA can fold intramolecularly on itself to form Watson-Crick base paired stem structures that are either capped with terminal loops or interrupted with internal loops or bulges.²⁷ Also, single strands joining two or more stem-loop structures can exist along with multi-branch loops or junctions. ■ represents unpaired nucleotides.

Single-stranded RNAs organize in such a way that they fold back on themselves so that complementary sequences in close proximity to each other can base pair and form short double helices. Hence, RNA secondary structure mainly consists of a series of 'hairpin loops' or 'stem-loops' that are separated by sequences of less defined conformations referred to as single strands (**Figure 1.7**). Watson-Crick base pairing is the major driving force aligning RNA strands. This forces non-canonical bases to juxtapose, and possibly pair when found within stabilized hydrogen bonded helical stems. Several of these non-canonical (*i.e.*, non Watson-Crick) pairing motifs have been identified and are particularly essential for ribosomal function since they promote tertiary interactions³⁶ and act as unusual recognition factors for other cellular components.³⁷ Among these are A-G, A-A, and U-A pairs that are highly conserved at certain positions within 16S and 23S ribosomal RNA (rRNA).^{38,39} G-U pairing motifs illustrate another important example,⁴⁰ where G pairs with U in a “wobble” configuration (**Figure 1.8**). Because wobble GUs are so common and fit easily into double-stranded helices, they are no longer referred to as non-canonical but rather Watson-Crick base pairs. Repeats of G-U pairs have been shown to serve as recognition signals for proteins and RNA-RNA interactions.

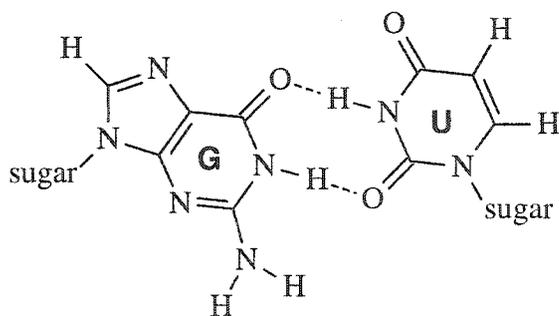


Figure 1.8: G-U wobble pairing motif.

In the RNA double-stranded helical regions, unpaired nucleotides may occur that give rise to bulged bases (**Figure 1.7**). The number of residues forming these bulges may vary from one (thus forming a single base bulge) to many nucleotide units (thus forming bulge loops). Single base bulges can either loop out or intercalate into the helix depending on various factors such as base identity, base sequence of the neighboring nucleotides and temperature. Bulge loops can induce bends and distortions in the helix

and may extend to surrounding helices. This might in turn lead to formation of multi-branched loops or 'junctions' that act as a liaison bringing three or more helices together. The number and conformation of the unpaired nucleotides governs these motions and imposes a profound effect on RNA three-dimensional structure.

Hairpins constitute the dominant secondary structure motif in RNAs and are formed when double helices are bridged by loop moieties, thus resulting in stem-loop structures. The composition and conformation of these loops govern the architecture of RNA folding in the cell. A quick glimpse at ribosomal RNAs (rRNA) reveals that their secondary structures are mainly composed of short Watson-Crick helices (or stems) that are either linked to one another by internal loops, or capped with terminal loops.⁴¹ **Figure 1.9** shows the secondary structure illustrating most of the loop motifs present in bacterial 16S rRNA. Terminal loops are ubiquitous in rRNAs. Although they vary in the number of constituent nucleotide units, they are most frequently present in the form of four nucleotide units or 'tetraloops'.⁴² Certain loop sequences are recurrent and comprise over half the terminal loops in rRNAs. They belong to one of three families of tetraloops, UNCG (where N = any nucleotide), GNRA (where R = purine), and CUNG. The sequence conservation within these loop motifs implies a critical association with ribosomal function. Apart from acting as nucleation sites for three-dimensional folding of large cellular RNA molecules, they are believed to stabilize neighboring stem structures via their inherent thermodynamic stability, and to promote tertiary structure interactions.^{34,43-47} Structural studies have shown that these loops share a highly ordered and unique structure with extensive hydrogen bonding interactions.

For purposes of brevity and conciseness, we will briefly discuss GNRA and CUNG loops while we focus in detail on the (UNCG) family of tetraloops since these constitute the main subject of this thesis.

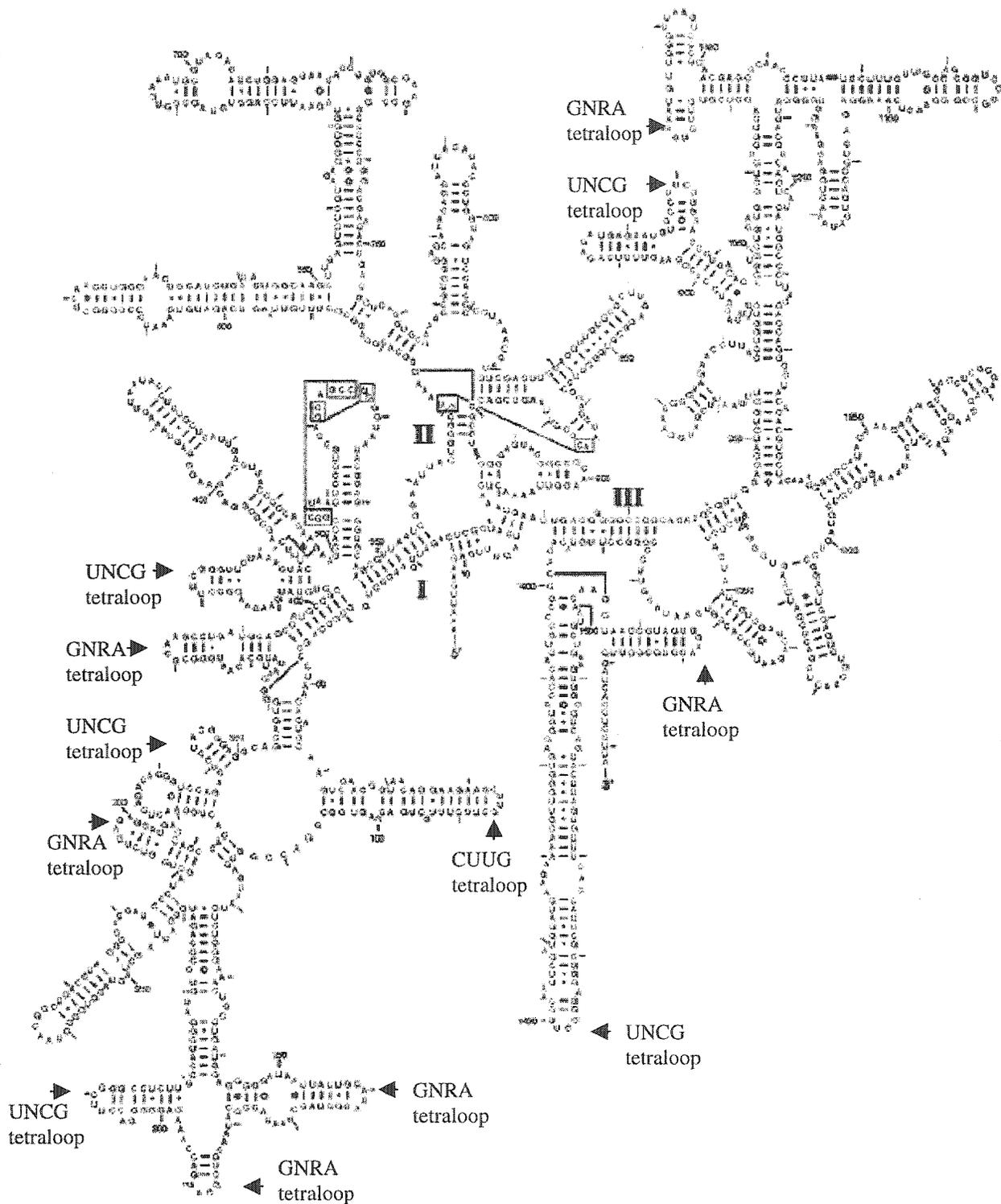


Figure 1.9: Secondary structure model of *E.coli* 16S rRNA. Indicated are GNRA, UNCG and CUUG tetraloops. Also described are various pairing motifs such as G-U, G-A and A-A (denoted by •). Adapted from [Gutell, R.R. In *Ribosomal RNA: Structure, Evolution, Processing, and Function in Protein Biosynthesis*; eds. Zimmermann, R.A. and Dahlberg, A.E.; CRC Press, Boca Raton, FL, 1996, p.111].

GNRA Tetraloops

Phylogenetic studies show that GNRA loops are not closed by a specific base pair.⁴² The solution structure of these loops has been determined by Pardi and coworkers.^{48,49} All loops belonging to this consensus proved, as expected, to exhibit similar conformations. The most striking feature is the base pair formed between the first and last loop residues (*i.e.*, G·A). This helps reduce the distances between the backbones of the two stem strands being capped, and thus allows the remaining two middle nucleotides (N and R) to bridge the stem ends. Base-phosphate interactions and extensive base stacking provide further stabilization to the structure. GNRA loops display significant biological activity. They serve as recognition signals for tertiary interactions with helices containing specific tetraloop receptors.^{43,44} Their conformation has been shown to be similar to that of loops contained within RNAs acting as aptamers that bind AMP.^{50,51}

CUNG Tetraloops

The conformation of CUNG tetraloops is quite different from that of GNRA tetraloops.⁴⁵ The first and last loop residues form a C·G Watson-Crick base pair. The second loop residue (U) fits into the minor groove of the helical stem being capped and interacts with its last base pair. These features help stabilize the loop structure. Phylogenetic studies show that whenever these loops exist in rRNAs, they are always in conjugation with terminal G·C base pairs, hence the consensus sequence G(CUNG)C. This implies that these loops are actually UN diloops surrounded with neighboring specific Watson-Crick base pairs.

UNCG Tetraloops

Similar to GNRA loops, UNCG loops have received considerable attention due to their high abundance in rRNAs. Among all classes of tetraloops, these exhibit the highest thermodynamic stability and are mostly favored by shorter stems.⁵² Phylogenetic analysis reveals that these loops are almost always closed by C·G base pairs [C(UNCG)G].⁴² Their conformation was initially investigated in Tinoco's laboratory in 1990 when studying the sequence 5'-G₁G₂A₃C₄(U₅U₆C₇G₈)G₉U₁₀C₁₁C₁₂-3' (where the brackets

denote the loop region).^{53,54} Five years later, their proposed loop structure was revised using a larger set of NMR-derived restraints.⁵⁵ The NMR solution structure revealed that the RNA helical stem exhibits typical A-form features (helix geometry and dimension, phosphate backbones, etc...) with all ribose rings in the C-3' *endo* pucker forms (**Figure 1.10 A**). In sharp contrast, unusual backbone conformations and intramolecular interactions exist in the loop moiety (**Figure 1.10 B**). The first (U₅) and last (G₈) loop residues stack on top of the helical A-stem. The second loop residue (U₆) is not involved in any of the stabilizing interactions and its base rather projects outside the loop. The most unique feature about these loop motifs is the formation of a reverse-wobble base pair between U₅ and G₈. Strikingly, the N-glycosidic bonds of the U₅-G₈ pair exists in an unusual *anti-syn* conformation and involves both base-base and base-sugar hydrogen bonds (**Figure 1.11**). This makes the interphosphate distance so small that it can be spanned by the middle two residues, U₆ and C₇. Not only that, the unusual *syn* conformation of G₈ induces an unusual backbone conformation. The absence of this base pair impairs thermodynamic stability and abolishes the unique structural features, a phenomenon that has been observed in hairpins with deoxyribose sugars.^{56,57}

An unusual backbone conformation occurs in the linkage between U₅ and U₆ residues where the phosphate C3'-O3'-P-O5' angle assumes a *trans* conformation instead of the commonly adopted *gauche* conformation (**Figure 1.12**).⁵³⁻⁵⁵ Although this arrangement is less energetically favored (loss of *gauche* effect is estimated at 2 kcal/mol), it is partially compensated, among other tertiary interactions, with a hydrogen bond between the exocyclic amino group of C₇ and an oxygen of the turning phosphate linking U₅ to U₆. This in turn promotes extensive base stacking overlap between C₇ and the U₅-G₈ base pair. Other unusual torsion angle values are evident within the loop moiety. For instance, the O5'-C5'-C4'-C3' γ torsion angle [C₇→G₈] is in a *trans* conformation in order to assist in bridging the two stems. This is accomplished with the help of the middle two nucleotide residues (U₆ and C₇) that switch their sugar conformations to the extended C-2' *endo* pucker.

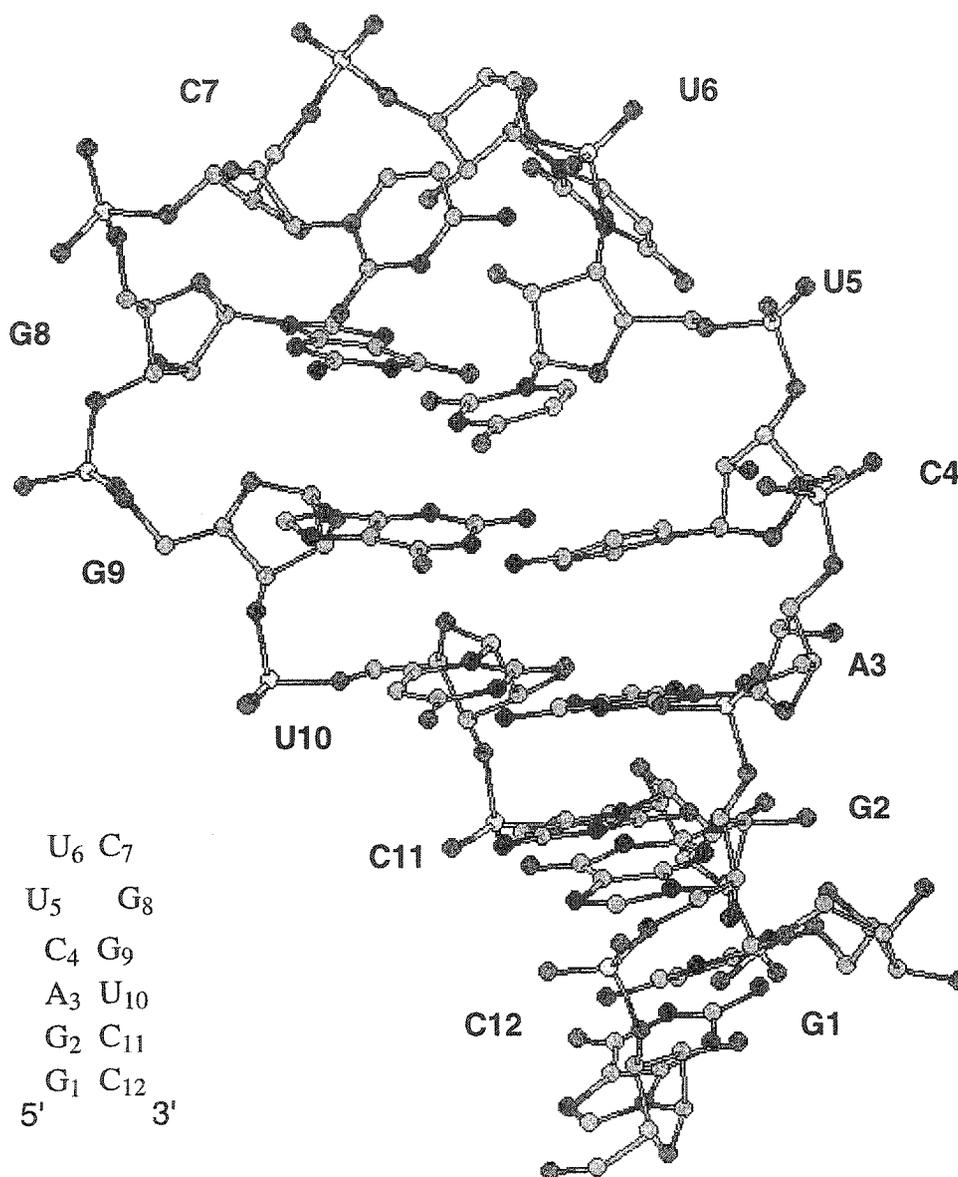


Figure 1.10 A: NMR-derived structure of an unusually stable hairpin. The structure has the sequence 5'-G₁G₂A₃C₄(U₅U₆C₇G₈)G₉U₁₀C₁₁C₁₂-3' and contains the extra-stable (UUCG) loop. The hairpin stem is defined by G₁-C₄ and G₉-C₁₂ residues. Adapted from Allain, F. H.-T.; Varani, G. *J. Mol. Biol.* **1995**, 250, 333 and was processed (for viewing purposes) using HyperChem molecular modeling software (version 6.1). Nitrogen is shown in blue, oxygen in red and carbon in cyan.

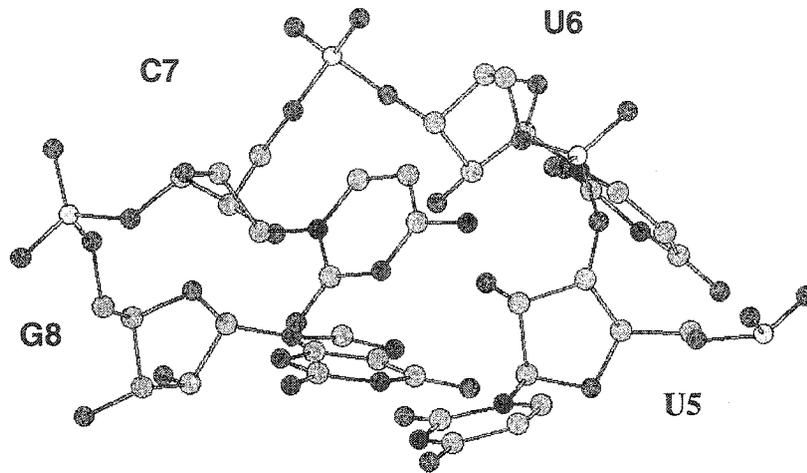


Figure 1.10 B: Structure of the (UUCG) tetraloop.

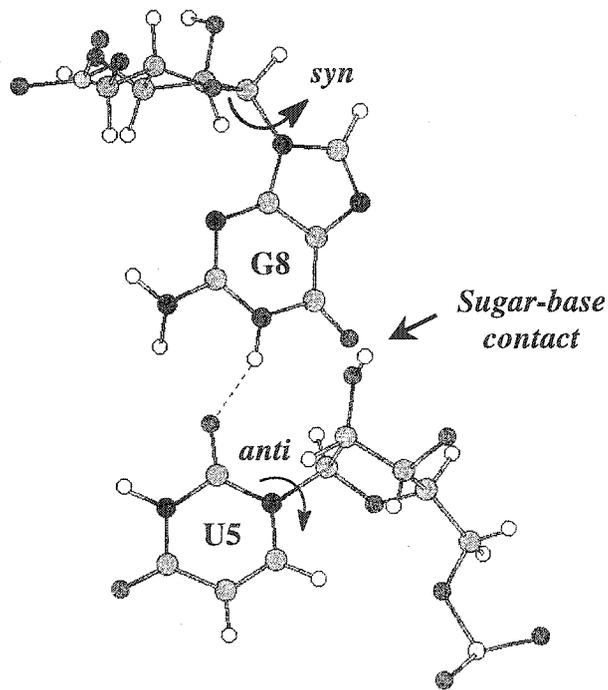


Figure 1.11: Morphology of the UG reverse wobble base pair evident in the (UUCG) tetraloop. The figure shows base-base and base-sugar hydrogen bonds.

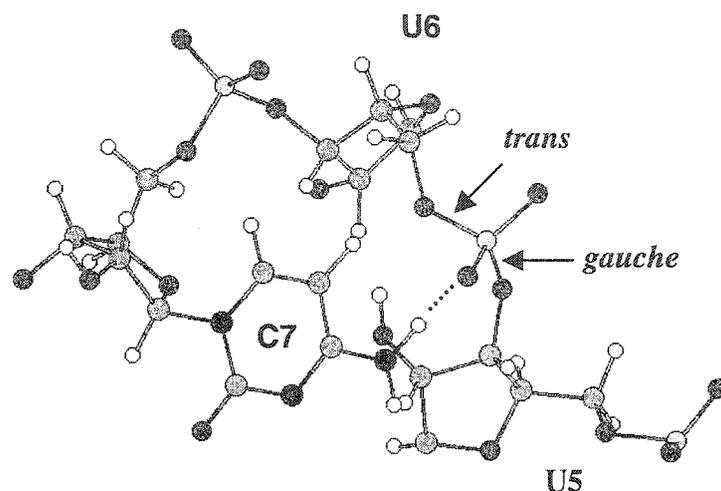


Figure 1.12: Unusual backbone conformation and cytosine-phosphate contact observed within (UUCG) tetraloops. The dashed line shows the electrostatic interaction observed between the exocyclic amine group of C7 and the phosphate oxygen between U5 and U6.

Interestingly, the X-ray crystal structure⁵⁸ of the RNA oligomer $G_1G_2A_3C_4(U_5U_6C_7G_8)G_9U_{10}C_{11}C_{12}-3'$ is different from its NMR solution form.⁵³ A hairpin species is eminent in solution, while a bimolecular duplex with internal bulged loops is the dominant form in the crystal phase. The X-ray diffraction studies showed that the structure is a typical A-form double-stranded helix that exhibits intermolecular interactions involving the 2'-OH groups.⁵⁸ Also, the role of water molecules is essential and crucial in stabilizing the structure through bonding interactions between the mismatch bases.

Comparison of UNCG and GNRA Tetraloops

The most common sequences within the UNCG and GNRA classes are (UUCG) and (GCAA) tetraloops. These have been studied extensively by NMR and have been shown to exhibit extensive stacking interactions and sequence-specific base-phosphate contacts.^{48,53,54} Their overall fold and appearances are alike. An extra base pair forms between the first and last loop nucleotides, thus promoting a 'diloop' structure. The only difference between these loops lies in their intrinsic dynamic states. The (UUCG) loop is highly rigid and its two middle nucleotides are frozen in the C-2' *endo* pucker conformation.^{53,54} In sharp contrast, the middle nucleotide residues in GNRA loops

generally adopt flexible sugar conformations that are in a dynamic equilibrium between the C-2' *endo* and C-3' *endo* conformers.⁴⁸ This may explain the enhanced biological activity displayed by GNRA tetraloops relative to (UNCG). The former loops exhibit tertiary contacts and act as binding sites for certain proteins.^{46,47,59-61}

1.5 THERMODYNAMICS OF HAIRPIN FOLDING

1.5.1 Effects of Loop Sequence and Composition on Hairpin Stability

Contrary to common belief, the unpaired bases in the loop contribute significantly to hairpin thermodynamic stability. Their sequence and base composition controls the behavior and folding of single stranded nucleic acids.⁶² With the advent of Differential Scanning Calorimetry (DSC) and Ultraviolet Spectroscopy (UV) techniques as well as computer software facilitating collection and processing of data, many hairpins loops of variable sizes and compositions have been extensively investigated. Hairpin loops containing 4-5 nucleotide residues were found to be the most stable.⁶³ Although loops found in tRNAs generally contain 6-7 nucleotide residues, those found in rRNAs are smaller in size and generally comprise 4-5 nucleotide units.^{1,42} Studies conducted on RNA hairpins revealed that upon varying the size (3-9) and composition of homopolymeric loops uniformly made of unpaired adenines, cytidines, or uracils, the most stable were loops composed of 4-5 units.⁶³ A similar trend was observed in DNA hairpins where 4 to 5 nucleotide-unit loops proved to be more stable relative to larger or smaller loops.^{64,65} A close inspection of the enthalpic values of hairpin formation reveals that enthalpy decreases with increasing loop size from 5 to 7 nucleotide residues. In sharp contrast, the thermal melting values (T_m) decrease with increasing loop size. This is an outcome of an unfavorable entropy of formation and comes at the expense of the enthalpic energy of association. The latter is 5-10 kcal/mol greater than that expected from the stem contribution, indicative of the presence of *intraloop* stacking interactions.⁶⁶ However, these are not enough to win over the *entropic* cost and thus justifies the drop observed in melting temperatures with increasing loop size.

Loop base sequence has also been shown to affect hairpin thermodynamic stability, though not to the same extent as changing loop size. Among DNA hairpins

comprising homopolymeric tetraloops, the tetra-thymidylate loop exhibits the highest thermal stability.⁶² The order of stability observed is $dT_4 > dC_4 > dG_4 > dA_4$, and the difference in melting temperature between the most and least stable loops is 5-6 °C. On the other hand, RNA hairpin loops displayed a different order of stability. The tetra-uridylate loop C(UUUU)G maintained the lead, followed by the less stable adenylate [C(AAAA)G] and lastly cytidylate [C(CCCC)G] loops.⁶³ The order of stability in free energy of association for the various loops parallels that obtained via calorimetric methods.⁶² The slight differences in thermal stabilities among the various loops are enthalpic in origin.

1.5.2 Unusually Stable Hairpins

Unusual thermodynamic properties of RNA hairpin structures containing either UNCG or GNRA loops were detected in rRNAs. Tuerk and coworkers reported that RNA sequences containing the general consensus (UUCG) prevented reverse-transcriptase from reading through, probably due to their high unusual thermal stability.⁶⁷ Pioneering studies by Tinoco and coworkers aimed at a detailed examination of the contribution of each of the loop nucleotides to the thermodynamic stability of hairpins with UNCG and GNRA tetraloops.^{68,69} Any RNA hairpin that displays a significantly higher stability than the corresponding r(UUUU) hairpin was considered *unusually stable*.^{68,69} To recall, the r(UUUU) loop is the most stable among the series of homopolymeric loops.⁶³ A hairpin containing a C(UUCG)G RNA loop and a 3 base pair stem was significantly more stable (ΔT_m ca. 20 °C) than the corresponding hairpin [G(UUUU)C] with purportedly normal thermodynamic stability,⁶⁸ and thus was considered *unusually stable*. The corresponding DNA hairpin with d[C(TTCG)G] loop had normal stability comparable to those with homopyrimidine loops (*i.e.*, d[C(TTTT)G]). NMR structural studies revealed that the RNA (UUCG) loop exhibits a highly stable and compact conformation whereas its DNA counterpart is very flexible and lacks well-defined interactions within the loop.^{53,54,57} In sharp contrast, both r(GAAA) and d(GAAA) loops were *unusually stable* compared to other DNA hairpin loops.⁶⁸ In general, a stable RNA hairpin sequence has a stable DNA hairpin counterpart. The only exception that violates the rule is the UUCG and TTCG loops, where the RNA loop is significantly more stable.^{34,68}

The thermodynamic properties of (UUCG) loops have been studied extensively due to their high abundance in 16S and 23S rRNAs.⁴² Base mutation studies conducted within the r(UUCG) loop seemed to agree with its NMR-derived structure. It was found that an RNA hairpin with C(UUCG)G loop [$T_m = 71.7$ °C] was more stable thermally than the corresponding hairpin with C(UUUG)G loop [$T_m = 64.0$ °C], which in turn was significantly more stable than that with a C(UUUU)G loop [$T_m = 60.4$ °C] (**Table 1.1**).⁶⁸ The C(UACG)G loop is extra stable and of the same stability as the wild-type C(UUCG)G loop. This may explain why r(UACG) loops occur in rRNAs with almost the same frequency as r(UUCG) loops. These observed patterns of stability correlate well with the specific interactions observed in the NMR-derived structure.^{53,54} The second loop residue can be mutated to any base without impairing thermodynamic stability. The slightest mutations (such as changing cytosine to uracil) in the third loop residue cause *ca.* 2 kcal/mol loss in free energy of formation (ΔT_m *ca.* 9 °C; **Table 1.1**), in agreement with the proposed cytosine-phosphate interaction observed in NMR.⁵⁴ Likewise, the last loop residue (G) cannot be replaced by any base. Even the slightest modifications such as replacing guanine with inosine (the exocyclic amino group of G is changed to hydrogen) significantly destabilizes the structure by 8 °C.⁵⁶ Switching the loop-closing base pair (C·G to G·C) causes a loss of 2.3 kcal/mol in free energy (ΔT_m *ca.* 10 °C; **Table 1.1**), indicative of the direct effect of the loop-closing base pair on the stem and end bases of the loop.⁶⁸ However, the G(UUCG)C loop was still more stable than the G(UUUG)C and G(UUUU)C loops. This suggests that the (UUCG) loop is capable of stabilizing hairpins with either C·G or G·C loop-closing base pairs, though a higher thermodynamic preference is given to those with C·G base pairs.⁶⁸ This might explain the frequency of their occurrence in rRNAs. Furthermore, the 5' to 3' directionality of the loop bases is fundamental to the formation of the unusually stable structure. For instance, the C(GCUU)G loop exhibits a reduced thermal stability relative to the C(UUCG)G loop (ΔT_m *ca.* 9 °C; **Table 1.1**).⁶⁸ The G(GCUU)C loop, unlike the C(UUCG)G loop, is not unusually stable.

Table 1.1: UV melting temperatures and Gibbs free energies of formation for various unusually stable RNA hairpins

RNA Loop Sequence	T_m (°C)	ΔG°_{37} (kcal/mol)
C(UUCG)G	71.7	-5.7
C(UACG)G	69.3	-5.2
C(UUUG)G	64.0	-3.8
C(UUUU)G	60.4	-3.0
G(UUCG)C	60.1	-3.4
G(UUUG)C	51.1	-1.7
G(UUUU)C	51.5	-1.7
C(GCUU)G	62.2	-3.2
G(GCUU)C	52.3	-1.7
C(GAAA)G	65.9	-4.2

Adapted from Antao, V. P.; Lai, S. Y.; Tinoco, I., Jr. *Nucleic Acids Res.* **1991**, *19*, 5901-5905. Bold letters represent base mutations in the (UUCG) loop. The RNA hairpin sequence is: 5'-GGAN(NNNN)NUCC-3' where the N residues represent the hexameric loop region (*i.e.*, loop + closing base pair).

The studies reviewed herein about the thermodynamics of (UUCG) loop formation clearly demonstrate that these motifs are highly sensitive to loop base sequence, loop composition, loop directionality, and their surrounding environment (identity of loop-closing base pair). This implies the slightest intrinsic modifications would be expected to perturb drastically their unique structure and unusual stability.

1.6 RNA INTERFERENCE

RNA interference or post-transcriptional gene silencing is the process by which double-stranded RNA (dsRNA) directs sequence specific degradation of mRNA in animal and plant cells (for reviews see references⁷⁰⁻⁷³). This mechanism regulates the expression of protein-coding genes and has been employed as a means to manipulate gene expression and to probe gene function.⁷¹ One of the proposed models for the RNAi pathway in *Drosophila* embryo lysates and S2 cells involves a four-step mechanism which is initiated by an ATP-dependent processive cleavage of long dsRNA (**Figure 1.13**). This very first step is executed by the enzyme Dicer, a member of the RNase III

family of dsRNA-specific endonucleases, which yields 21-23 nucleotide (nt) fragments called small interfering RNAs (siRNAs).⁷⁴⁻⁷⁷ Like the products of other RNase III enzymes, siRNA duplexes contain 5' phosphate and 3' hydroxyl termini, and two unpaired nucleotides on their 3'-ends.⁷⁸ These siRNA duplexes are then incorporated into a protein complex that is not yet active to mediate RNAi.⁷⁹ The unwinding of the siRNA duplex (ATP-dependent process) transforms the complex into an active RNA-induced silencing complex (RISC*), which can recognize and subsequently cleave a target RNA complementary to the guide strand of the siRNA.⁷⁹⁻⁸¹

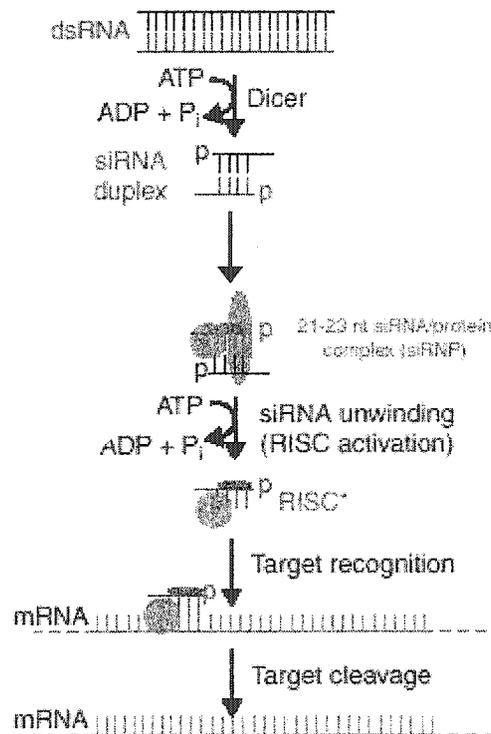


Figure 1.13: Proposed mechanism of RNA interference. Adapted from Hutvagner, C.; Zamore, P. D. *Curr. Opin. Genes. Dev.* **2002**, *12*, 225-232

Through this mechanism, it has been estimated that in *Drosophila* embryos, ~35 dsRNA molecules can silence a target RNA thought to be present at > 1000 copies per cell.⁸² Very recently, the mechanism of RNA interference has been used to modulate HIV-1 replication. Synthetic siRNA duplexes were targeted to various regions of the HIV-1 genome thereby promoting specific degradation of genomic HIV-1 RNA, and thus inhibiting early and late steps in HIV-1 replication in human cell lines.⁸³ Other hairpin RNAs that mimic siRNA have also been shown to target genes for silencing, and their

mechanism of silencing occurs through mRNA degradation and not translational repression.⁸⁴ Finally, RNA interference has been even employed as a means to facilitate functional genomics whose main goal is to determine the function of genes expressed (or protein encoded) by an organism. For instance, the so-called “exon-specific RNAi” involves selectively degrading specific alternatively spliced mRNA isoforms by treating the cells with dsRNA corresponding to an alternatively spliced exon.⁸⁵ This provides a tool for assessing the function of proteins synthesized from alternatively spliced RNAs.

1.7 HIV-1 REPLICATION AND REVERSE TRANSCRIPTASE

For over forty years, the AIDS epidemic has been assimilating worldwide leading to deterioration of the human immune defensive systems. In 1983, human immunodeficiency virus type-1 (HIV-1) was isolated and identified as the main retrovirus that causes AIDS.⁸⁶ Around the same period, a second virus (HIV-2) was isolated and proved to be by far less pathogenic relative to HIV-1, thus leading to AIDS very slowly.^{87,88} This has prompted the search for new therapeutic agents that can arrest the replication of the causative virus (HIV-1) and hence, a complete understanding the biochemical replicative cycle of HIV-1 became imperative.

A close inspection of the HIV-1 replication process reveals that the retrovirus undergoes several stages before it regenerates into new viruses. HIV-1 infection is initiated by binding of a virion to the cell surface membrane via specific recognition receptors [CD4 and chemokine co-receptors].⁸⁹ Entry into the cytosol via receptor-mediated endocytosis followed by subsequent uncoating of the virion nucleocapsid releases the genomic viral RNA (**Figure 1.14**). This sets the stage for reverse transcription, a process that transforms viral RNA to double-stranded DNA (or proviral DNA). In this process, viral RNA is reverse transcribed to DNA by the RNA-dependent DNA polymerase (RDDP) activity of reverse transcriptase (RT). This yields a viral RNA:DNA hybrid duplex that is immediately recognized by the RNase H activity of reverse transcriptase which degrades the viral RNA strand. The intact released DNA is transcribed to a double-helical viral DNA through the DNA-dependent DNA polymerase activity (DDDP) of reverse transcriptase. This proviral DNA helix is transferred into the nucleus of the cell (nuclear import) where it gets integrated into the host cell genome by

the viral-encoded integrase. The host cell recognizes and treats the integrated provirus as one of its own genetic sequences and thus initiates transcription to yield viral mRNAs. These are in turn translated to polyproteins that are cleaved by the viral protease. Finally, the newly manufactured virion particles are assembled and, after maturation, bud outside the cell in the form of a mature virus that is ready to attack other cells.

Therapeutic agents can be designed to specifically intervene with various stages of the HIV replication cycle and thus arrest viral proliferation (**Figure 1.14**). Some compounds have made it through to the clinic, while others are still in early phases of development. Since reverse transcriptase plays a pivotal role in the life cycle of HIV-1, it became one of the most eminent targets of choice for anti-HIV therapy (for reviews see references 90, 91 and 92). Reverse transcriptase is a multifunctional enzyme that displays both RNA- and DNA-dependent polymerase activities as well as a ribonuclease H (RNase H) activity that cleaves the RNA strand in RNA:DNA hybrids. More specifically, HIV-1 RT is a heterodimeric enzyme possessing 66 and 51 kDa subunits, referred to as p66 and p51 respectively (**Figure 1.15**). Both DNA polymerase and RNase H enzymatic activities are situated in the p66 polypeptide domain. The RNase H domain is located in the C-terminal region of the protein while the DNA polymerase domain is in the N-terminal portion. X-ray diffraction studies have shown that the enzyme has an asymmetric structure.⁹³ In other words, the polymerase domain of the p66 subunit exhibits a large cleft in sharp contrast to the p51 subunit of identical sequence which shows no such cleft and a rather different conformation. The p66 subunit folds into five separate subdomains. Four of these correspond to the DNA polymerase domain and are similar to a right hand, thus their naming as fingers, palm, thumb and connection. The latter lies between the three other DNA polymerase and the RNaseH subdomains.⁹³

The fifth region within the p66 peptide chain comprises the RNase H domain.⁹⁴ The X-ray crystal structure of the HIV-1 RT RNase H domain was solved before that of the whole polymerase domain.⁹⁵ A comparison between the separate domain of HIV-1 RNase H and *Escherichia coli* RNase H reveals a striking similarity in their three-dimensional structure.

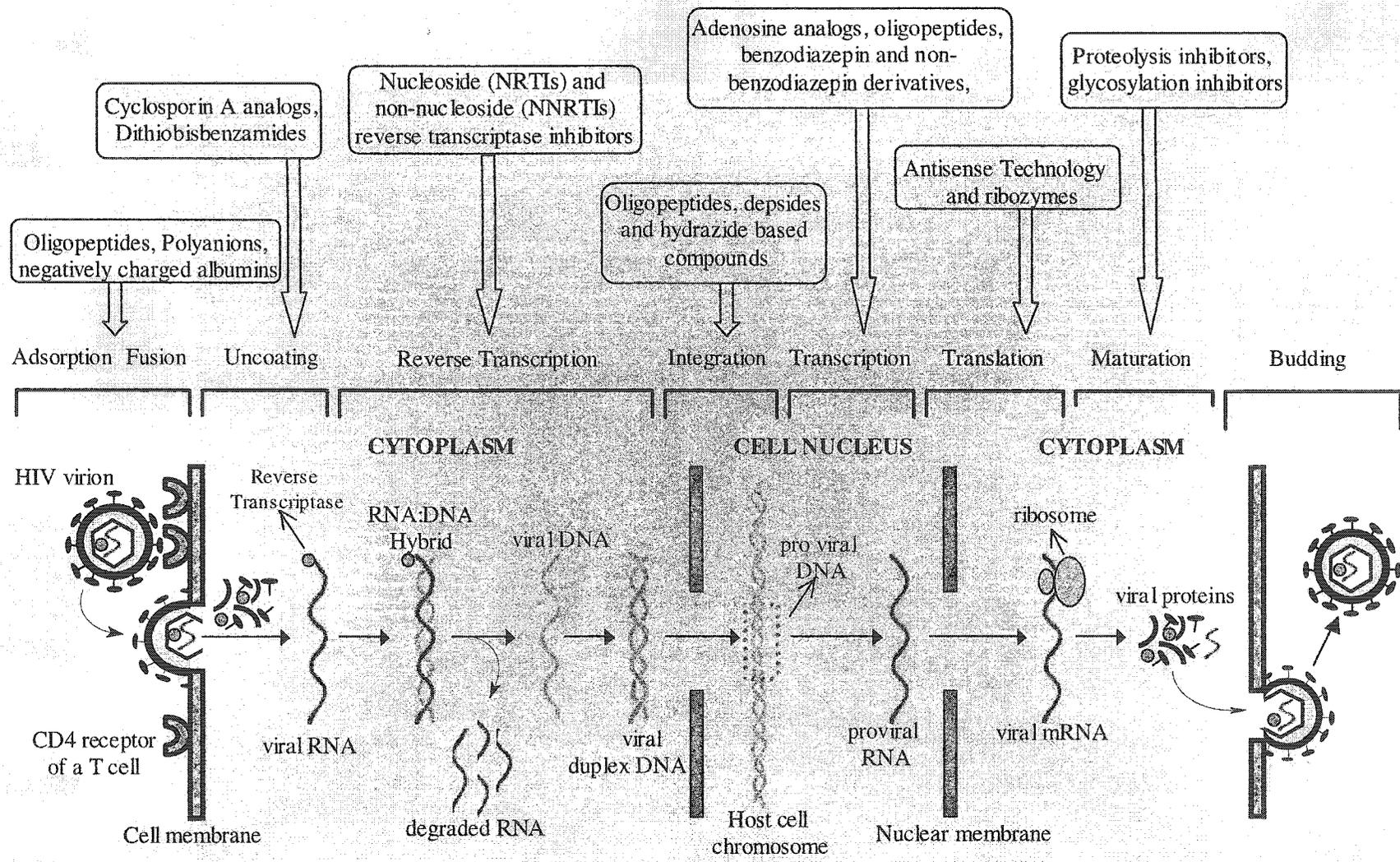


Figure 1.14: HIV replication cycle showing various classes of antiretroviral drugs as well as their stages of intervention.

Both fold into a five-stranded mixed β sheet surrounded by asymmetrically distributed α helices (Figure 1.16) Additionally, the key amino acid residues at the active sites are alike. However, the degree of homology in the sequence of amino acids between the two enzymes is very low. RNase H activity necessitates the presence of divalent cations. In fact, the crystal structure has shown that two divalent ions (Mg^{2+}) bind to the active site of the protein and participate in the cleavage mechanism of the RNA strand.

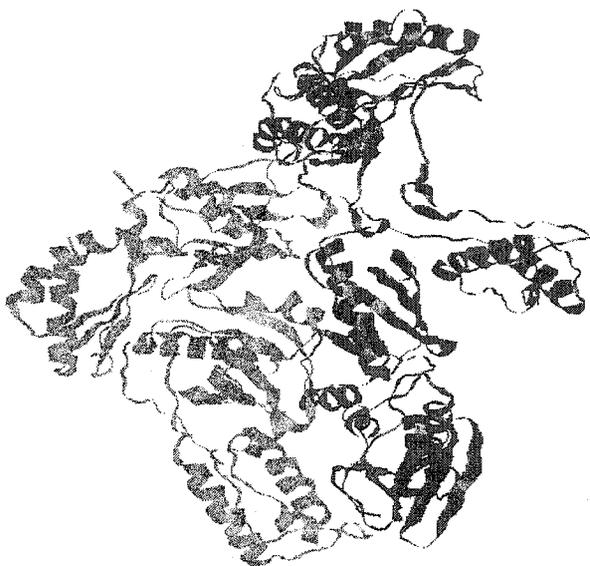


Figure 1.15: Crystal structure of HIV-1 RT. Adapted from [*J. Med. Chem.* 1996, 39, 1589].



Figure 1.16: RNase H domain from the crystal structure of HIV-1 RT. Adapted from [*J. Med. Chem.* 1996, 39, 1589].

1.8 NUCLEIC ACIDS WITH 2',5'-PHOSPHODIESTER LINKAGES

1.8.1 Biological Occurrence of 2',5'-Nucleic Acids

One of the most intriguing inquiries in the history of RNA evolution is concerned with why nature chose the 3',5'-phosphodiester linkage (over the 2',5'-linkage) as a carrier of genetic information. Although 2',5'-linked nucleic acids are the most analogous chemical mimic to 3',5'-RNA (**Figure 1.17**), they were harshly rejected by nature as the sugar-phosphate backbone of choice during the course of evolution. At first, scientists ascribed this to the inability of 2',5'-linked nucleic acids to assemble via Watson-Crick base pairing and thus their reluctance to form double helical structures.⁹⁶⁻⁹⁸ This theory prevailed for quite some time before major breakthroughs in oligonucleotide solid-phase synthesis emerged,^{12-15,17-18} thus providing researchers with unlimited tools to explore in further detail the properties of a wide variety of diverse oligonucleotide mimics, among which are the 2',5'-nucleic acids.

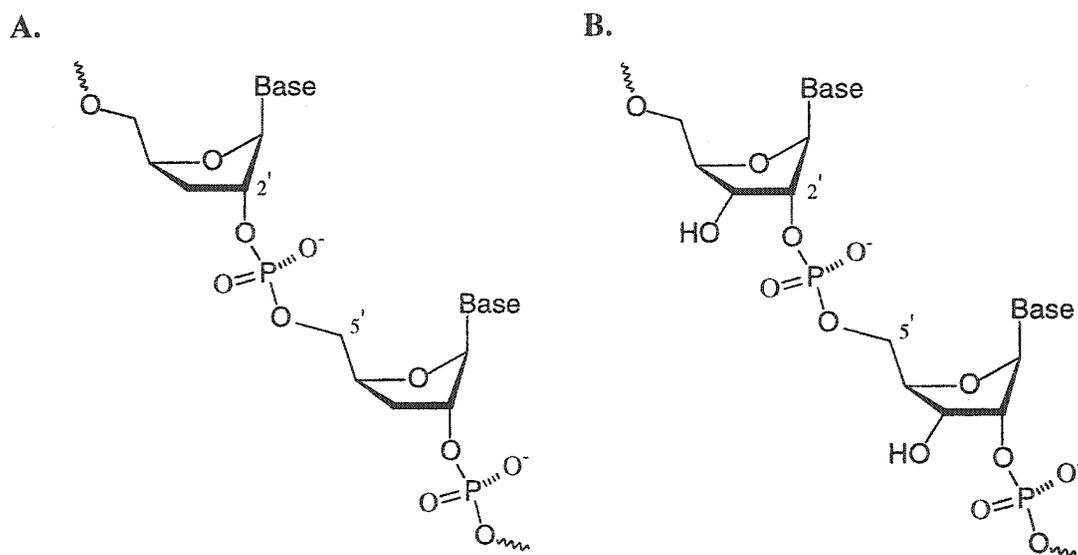


Figure 1.17: Chemical structures of (A) 2',5'-DNA and (B) 2',5'-RNA.

Contrary to common beliefs, 2',5'-linkages do occur naturally in living cells. However, they are by far less abundant than the more commonly occurring 3',5'-linkages. 2',5'-Linked nucleic acids have been reported to occur during RNA splicing as well as in interferon treated cells.⁹⁹⁻¹⁰¹ In RNA splicing, the newly formed 'lariats' ("circular RNAs with a tail") contain a 2',5'-phosphodiester bond at their nucleotide branch point (**Figure**

1.8.2 The Literature of 2',5'-Nucleic Acids

The earliest studies on 2',5'-nucleic acids were conducted in the late 1960s shortly after the monumental discovery of the DNA double helix. In 1967, short fragments of 2',5'-oligoadenylates were hybridized to poly rU, and the complexes observed were found to be less stable thermally than those formed with the 3',5'-oligoadenylate counterparts.¹⁰⁶ Around the same time period (1968), 2',5'-phosphodiester linkages were shown to be cleaved, like 3',5'-linkages, under basic conditions due to intrasugar nucleophilic attack of the 3'-OH on the vicinal phosphodiester bond.¹⁰⁷ However, unlike 3',5'-linkages, 2',5'-linkages were more resistant against some hydrolytic nucleases. Shortly after (1969-1970), the overlap between bases in single stranded 2',5'-RNA was reported, via NMR, to be more extensive than that in 3',5'-RNA.^{108,109} In 1971 and 1974, the Orgel laboratory published pioneering studies aimed at simulating experimentally the prebiotic template-directed non-enzymatic oligomerization of ribonucleotides.^{110,111} Their experiments yielded oligoribonucleotides containing predominantly the less commonly occurring 2',5'-phosphodiester bond. For example, ethylenediamine-catalyzed coupling of adenosine 2',5'-cyclic phosphate on a poly(U) template led to predominant formation (97%) of a 2',5'-phosphodiester bond within the dimer produced.¹¹⁰ In 1976, a ground-breaking study by Usher attempted to address, from the point of view of hydrolytic stability, the issue of evolutionary selective advantage for the natural 3',5'-bond over its 2',5'-isomer.¹¹² These studies show that under basic buffer conditions the 2',5'-phosphodiester bond is by far less hydrolytically stable than the 3',5'-bond. For example, when present in duplex helical forms, the ratio of the rates of cleavage of one 2',5'-bond to one 3',5'-bond is 900:1.¹¹² It might be argued that the 2',5'-bond is more exposed to hydrolytic cleavage while the 3',5'-bond is protected against basic hydrolysis within the context of helical duplexes.¹¹³ This, along with the inability of 2',5'-nucleic acids to form stable double helical structures,⁹⁶ would explain its unsuitability as a carrier of genetic information. Theoretical studies in 1986 unveiled different predictions about the possibility of helix formation by 2',5'-nucleic acids.^{114,115} Between 1977 and 1991, many X-ray diffraction and NMR studies were conducted on dimers and trimers containing 2',5'-linkages.^{97,116-120} These culminated in conflicting findings about the preferred sugar pucker adopted by the 2',5'-nucleic acids both in solution and in the crystal.

Between 1967 and 1991, only dimers and trimers of 2',5'-nucleic acids were studied due to the very limited accessibility of longer 2',5'-RNA fragments. However, with the development and expansion of solid-phase synthesis techniques in the mid 1980s, important works describing 2',5'-linked oligoribonucleotides came into sight soon after thus allowing detailed characterization of their hybridization and physicochemical properties. In 1991, the Damha laboratory demonstrated, through UV studies, the ability of 2',5'-linked decariboadenylic acid [2',5'-rA₁₀] to bind to poly rU via Watson-Crick complementary base pairing.¹²¹ The formed complex exhibited a lower melting temperature (26.5 °C) relative to the complex formed by the 3',5'-RNA analogue, *i.e.* poly rU:3',5'-rA₁₀ (30.0 °C). The same authors also reported that 2',5'-linked homopolymers [rA₁₀, rC₁₀, rU₁₀] were more stable than the corresponding 3',5'-RNA counterparts towards hydrolytic degradation by various nucleases such as nuclease T2, nuclease S1, nuclease P1 and calf spleen phosphodiesterase.¹²¹ Although not thermally very stable, its association into a double helical structure alienates to the possibility of existence of a prebiotic self-replicating 2',5'-RNA system during the course of evolution (the RNA world).

One year later (1992), three major reports on the properties of 2',5'-nucleic acids emerged. Turner and coworkers showed for the first time that 2',5' oligoribonucleotides [2',5'-RNA] can form complementary Watson-Crick double helices with antiparallel strands.¹²² These duplexes were by far less thermally stable than those formed by 3',5'-oligoribonucleotides [3',5'-RNA] due to less favorable enthalpy of association.¹²² About the same time, Breslow and Switzer independently demonstrated the ability of 2',5'-DNA to associate via antiparallel Watson-Crick base pairing under high salt conditions.^{123,124} Breslow also showed that insertion of a few 2',5'-linkages within 3',5'-DNA duplexes was always destabilizing, while Switzer demonstrated the weaker thermal stability of duplex 2',5'-DNA relative to duplex 3',5'-DNA. All these studies suggested that the rather poor thermal stability displayed by 2',5'-duplexes (whether 2',5'-RNA or 2',5'-DNA) might have possibly acted as a negative selective pressure that favored the predominance of 3',5'-nucleic acids during the final stages of evolutionary selection. However, all these studies did not assess the hybridization of a 2',5'-nucleic acid with a complementary 3',5'-nucleic acid strand.

A more recent publication (1993) by Damha and coworkers reported the formation of 2',5'-RNA:RNA hybrid duplexes.¹²⁵ Their studies show that 2',5'-RNA exhibited binding selectivity for RNA over single stranded DNA. This was in sharp contrast to 3',5'-RNA which could form stable duplexes with both complementary RNA and DNA single strands. They also showed that contiguous 2',5'-RNA substitutions within duplexes made entirely of 3',5'-RNA always resulted in thermally destabilizing these duplexes and in enhancing selectivity for complementary RNA over single stranded DNA. Additionally, 2',5'-rA₁₀ was capable of forming a triplex structure with poly rU but not with poly dT. However, 3',5'-rA₁₀ formed triplexes with both poly dT and poly rU of approximately the same stability. Chimeric 2',5'/3',5' riboadenylate did form a triplex with poly dT, indicative of the inability of 2',5'-RNA alone to associate with complementary 3',5'-DNA. These findings suggested that, in principle, hybrids of 2',5'-RNA and 3',5'-RNA may have existed during the early stages of evolution. In the same year (1993), similar studies conducted on 2',5'-DNA were reported by Breslauer and coworkers.¹²⁶ They showed that at high ionic strength 2',5'-dA₁₆ and 2',5'-dT₁₆ were able to form triplex structures exhibiting higher thermal stability, lower charge density and different conformational features than those formed by the corresponding 3',5'-strands.¹²⁶ The 2',5'-triplex helical structures are more 'open' and dispersed than those made of 3',5'-strands.^{126,127} The 3',5'-strands were able to form either a duplex or triplex depending on salt conditions. On the contrary, 2',5'-single strands did not form any duplex, *i.e.*, either a triplex or no complex was observed. The 2',5'-triplex directly denatured to single strands without passing through a duplex intermediate state. The lack of stable duplex association in the case of 2',5'-single strands, the authors claim,¹²⁶ gives the 3',5'-linked deoxy nucleic acids an evolutionary advantage as a genomic material.

In 1994, Switzer and coworkers reported on the ability of 2',5'-DNA comprising guanine and cytosine residues to associate at low ionic strength.¹²⁸ This established that 2',5'-DNA might have existed as an alternative ancestral genetic material comprising a four-component genetic code. According to their hypothesis, this form of genetic material (*i.e.*, 2',5'-nucleic acids) was later on abolished and taken over by the 3',5'-nucleic acids *via* a process known as "genetic takeover".¹²⁸ In this process, assuming a given self-replicating system, there arises a metabolism that leads to the synthesis of another species

such as the 3',5'-linked nucleic acids which in turn take over all metabolic functions as well as the original genetic material of the self-replicating system.¹²⁹ Further support of this hypothesis stemmed from evidence provided by Damha and co-workers a year earlier which alienates towards hybrid formation between 2',5' and 3',5' linked nucleic acids.¹²⁵ This hybrid duplex could have possibly been formed in the intermediary stages during transformation from the primitive ancestral 2',5'-linked nucleic acid to the present day 3',5'-linked genetic material. In the same year (1994), oligodeoxynucleotides containing thioformacetal linkages as neutral achiral replacements for the 2',5'-phosphodiester linkage were described by Matteucci and coworkers (**Figure 1.19**), and shown to exhibit binding selectivity to single stranded RNA over single stranded DNA.¹³⁰ Like the 2',5'-phosphodiester linkages, substituting few 2',5'-thioformacetal linkages within DNA helices resulted in destabilizing such helices.

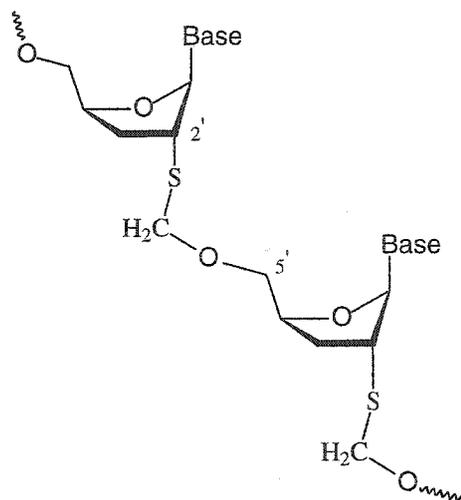


Figure 1.19: Chemical structure of 2',5'-linked thioformacetal nucleic acids.

In 1995, a breakthrough was witnessed in the structural features of 2',5'-nucleic acid complexes. Molecular modeling studies conducted by Lalitha and Yathindra showed that 2',5'-RNA duplexes exist with predominant C-2' *endo* pucker repeats.¹³¹ They depict that this was more favored than the normal C-3' *endo* pucker (adopted by 3',5'-RNA) because of the steric interference with the phosphodiester linkages imposed by the 3'-OH group when present within the C-3' *endo* conformation.¹³¹ On the other hand, Switzer and Wang reported, via NMR studies, that 2',5'-DNA duplexes exist in A-type conformation with the sugar residues adopting both C-2' and C-3' *endo* puckers.¹³² The duplex major

groove was narrow and deep as observed in natural A-DNA, while its minor groove was not as broad due to the fact that the two antiparallel backbones are brought closer together by virtue of the presence of the 2',5'-linkages (Figure 1.20).¹³² In the same year, a 2',5'-linked octamer of deoxyadenylic acid [2',5'-dA₈] and its phosphorothioate analogue were reported to selectively associate with poly rU (and not poly dT) to form duplexes of the same order of thermal stability as the 3',5'-dA₈ : poly rU duplex counterpart.¹³³ Also, reverse transcriptase was shown to read through an RNA template bearing one 2',5'-linkage.¹³⁴

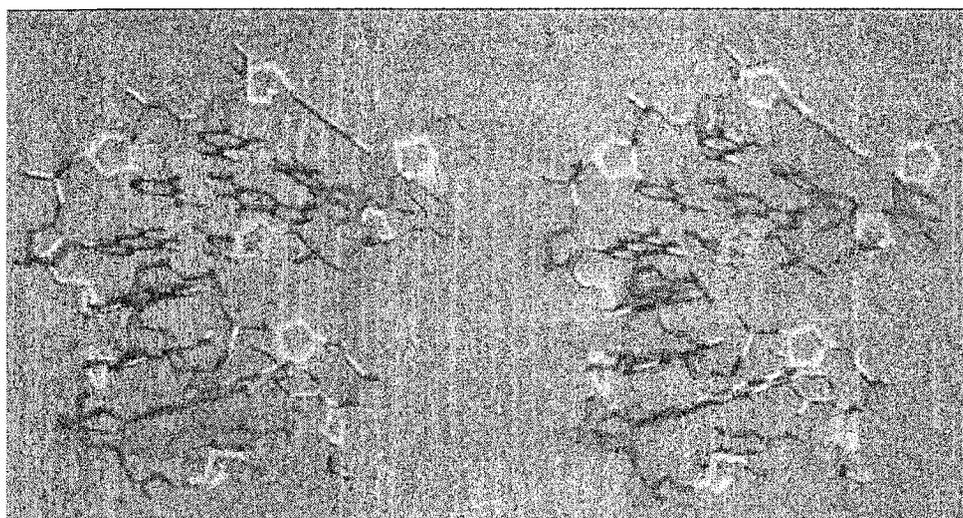


Figure 1.20: NMR-derived structure of a 2',5'-DNA duplex. The sequence is 5'-CGGCGCCG-2' and the sugars are colored according to their pucker (yellow for C-3' endo pucker and red for C-2' endo pucker). Adapted from [Robinson, H.; Jung, K.-E.; Switzer, C.; Wang, A. H.-J. *J. Am. Chem. Soc.* **1995**, *117*, 837-838].

In 1996, five years after the original findings of Damha and coworkers on the binding selectivity of 2',5'-RNA, Breslow and Switzer independently demonstrated that 2',5'-DNA, like 2',5'-RNA, exhibited binding selectivity towards complementary 3',5'-RNA but not 3',5'-DNA.^{135,136} The 2',5'-DNA:RNA hybrid was less stable than the normal DNA:RNA hybrid although they share the same global conformation as judged from circular dichroism studies (A-type helix).^{135,136} In spite of this, 2',5'-DNA:RNA hybrids are not substrates for RNase H, *i.e.* the enzyme does not degrade the RNA strand in the 2',5'-DNA:RNA hybrid.¹³⁶ Also, the susceptibility of 2',5'-DNA towards various nucleases was assessed and the results clearly showed enhanced resistance of 2',5'-DNA

to cellular exo and endonucleases relative to 3',5'-DNA. In the same year, Ozaki and coworkers reported that 2',5'-RNA forms triplex structures that are less thermally stable than the corresponding 3',5'-triplex counterparts.¹³⁷

In 1997, several reports dealt mainly with mixed backbone oligonucleotides comprising DNA/2',5'-DNA¹³⁸ or DNA/2',5'-RNA¹³⁹ chimeras (**Figure 1.21**). 2',5'-Linked phosphorothioate oligodeoxynucleotides exhibited less non-specific binding to human cellular proteins relative to their 3',5'-linked analogues.¹³⁸ Also, hybrids of 2',5'-DNA with RNA did not elicit RNase H activity. However, a 2',5'-DNA-[DNA]-2',5'-DNA construct containing a central DNA gap (7 nt in length) was capable of activating RNase H and causing degradation of the RNA strand.¹³⁸ Other mixed backbone oligonucleotides composed of DNA and containing 2',5'-RNA insertions were also reported.¹³⁹ As the number of 2',5'-RNA inserts increased within the DNA strands, their binding affinity to complementary single stranded DNA decreased. These results confirmed what previously has been reported about binding selectivity of 2',5'-RNA towards single stranded RNA.¹²⁵ Phosphorothioated DNA-2',5'-DNA constructs exhibited less non-specific binding to proteins relative to unmodified DNA constructs, thus resulting in less immune stimulation.¹³⁹ In a different setting, Switzer and coworkers reported the activity of 2',5'-RNA in template-directed oligomerization of mononucleotides.¹⁴⁰ This supports the notion that 2',5'-RNA might have contributed to the evolution of the present-day nucleic acids.

Emerging in 1998 was the first report describing the preparation and hybridization properties of 2',5'-oligoribonucleotides (2',5'-RNA) of mixed base composition (containing all four bases).¹⁴¹ The order of thermal stability observed was as follows: RNA:RNA > DNA:DNA \approx DNA:RNA > RNA:2',5'-RNA > 2',5'-RNA:2',5'-RNA \gg DNA:2',5'-RNA (undetected). Furthermore, RNA:2',5'-RNA hybrids proved not to elicit RNase H activity.

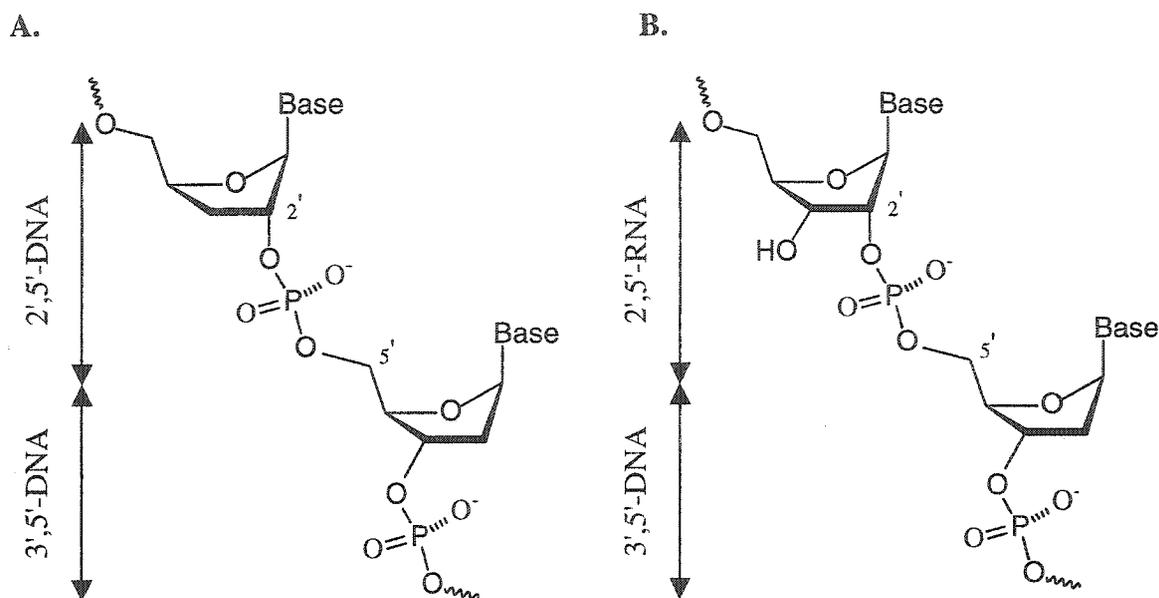


Figure 1.21: Chemical structures of mixed backbone oligonucleotides composed of (A) 2',5'-DNA/3',5'-DNA and (B) 2',5'-RNA/3',5'-DNA.

However, they were still able to inhibit the RNase H-mediated cleavage of RNA in RNA:DNA hybrid substrates.¹⁴¹ CD spectra showed that RNA:2',5'-RNA hybrids adopt A-type helical conformations similar to those of RNA:DNA hybrids, although molecular modeling experiments showed that their interstrand phosphate-phosphate distances are smaller by 1 Å than those of normal duplexes. They also observed that the structure of the RNA strand does not change significantly upon hybridization to complementary RNA or 2',5'-RNA even though the latter induces structural changes to the helical structure.¹⁴¹ In the same year, the structure of 2',5'-RNA double helices was suggested by Yathindra and Premraj through molecular modeling studies. Here, 2',5'-RNA duplexes [polyA:polyU] were built taking into account the crystal structure of 2',5'-ApU 'dinucleotide' repeats. The obtained helices exhibited alternating C-2' *endo* and C-3' *endo* sugar repeats.¹⁴² The 2',5'-linkages induced wider major groove dimensions, an observation in agreement with what has been previously reported for 2',5'-DNA helices.¹³² This leads to an increase in slide and displacement along the individual strands within the helix as well as to an increase in the adjacent interphosphate distance. Even more striking was the hydrogen bonding interaction between the anionic oxygen of the phosphodiester bond and the sugar 3'-hydroxyl group as a stabilizing factor in 2',5'-RNA helices.¹⁴²

Homopyrimidine 2',5'-linked RNA has been shown to recognize and bind to nucleic acid double helices.¹⁴³ Also, 2',5'-linked decauridylylate serves as a template for the synthesis of oligadenylates from activated 2',5'-(pA)₂.¹⁴⁴ In 1999, the catalytic activity of RNA containing a few 2',5'-linkages was explored. Namely, the domains I, II and III within a hammerhead ribozyme have been replaced by 2',5'-linkages without impairing catalytic activity.¹⁴⁵

In 2001, the first NMR-derived structure of 2',5'-RNA duplexes was reported (discussed below). In 2002, the structural basis for the unusual properties of 2',5'-nucleic acids was proposed via molecular modeling studies.¹⁴⁶ Furthermore, 2',5'-RNA was shown to fold, like 3',5'-RNA, into *hairpins* or 'stem-loop' structures with unusual thermal stability.^{147,148} The 2',5'-RNA loop adopted a novel uniquely folded structure that was different from that of the 3',5'-RNA counterpart (discussed in chapter 2). Also, 2',5'-ribocytidine nucleic acids were shown to associate into antiparallel stranded duplexes with hemiprotonated C·C+ base pairs (discussed in chapter 5).

1.8.3 Preferred Sugar Puckers of 2',5'-Nucleic Acids

With the development of molecular modeling and NMR techniques, researchers have started to understand the structural properties of 2',5'-nucleic acids. While the hybridization properties of 2',5'-nucleic acids are becoming clearer, their conformational properties remain poorly understood. Understanding the folding pattern of 2',5'-RNA is important not only for understanding the evolutionary selection of 3',5'-links over 2',5'-links but also the factors that may stabilize 2',5'-nucleic acid structures.

A close inspection of the morphology of nucleoside 2',5'- *versus* 3',5'-phosphates reveals that their major structural difference lies in the C2'-C3' bond extension, *i.e.*, seven *versus* six bonds separating the phosphate groups respectively (**Figure 1.17**). This "additional" bond separating the phosphate moieties imposes variations in the conformational features and preferences between the two isomeric nucleotide forms and induces a switch in the dimensions of their individual pucker repeats. In late 1990s, the Yathindra laboratory has set out to determine the structure of 2',5'-nucleic acids. This lab highlighted the nature of these differences from the point of view of nucleotide shape and dimensions.¹⁴² Their studies show that in a C-2' *endo* pucker mode, the C2'-O2' bond is in

pseudo axial orientation, while the C3'-O3' bond is pseudo equatorial. Axial orientations tend to disperse the phosphate groups away from each other while equatorial orientation tends to bring them close to one another. The proximity of the intranucleotide phosphate groups defines sugar pucker mode whether 'compact' or 'extended' (**Figure 1.22**). For example, a C-2' *endo* sugar pucker would be 'extended' in 3',5'-linked nucleic acids [P-P = 7Å], but 'compact' for the 2',5'-linked isomer [P-P = 5.9Å]. On the contrary, a C-3' *endo* pucker would be 'compact' in 3',5'-nucleotide diphosphates [P-P = 5.9Å], but 'extended' in the 2',5'-isomeric nucleotides [P-P = 7.5Å]. As evident from these studies, the authors propose that phosphodiester bond isomerization [between vicinal C-2' and C-3' positions] necessitates a switch in the ribose pucker preferences in 2',5'- and 3',5'-linked nucleic acids.¹⁴² Three years later (2001), this hypothesis was confirmed by the same laboratory when studying 2',5'-RNA via high-resolution NMR. They showed that 2',5'-RNA adopts A-type duplex structures with repeating C-2' *endo* sugar units (**Figure 1.23**).¹⁴⁹ This is in sharp contrast to the isomeric 3',5'-RNA which adopts A-type duplexes with repeating C-3' *endo* puckers.¹

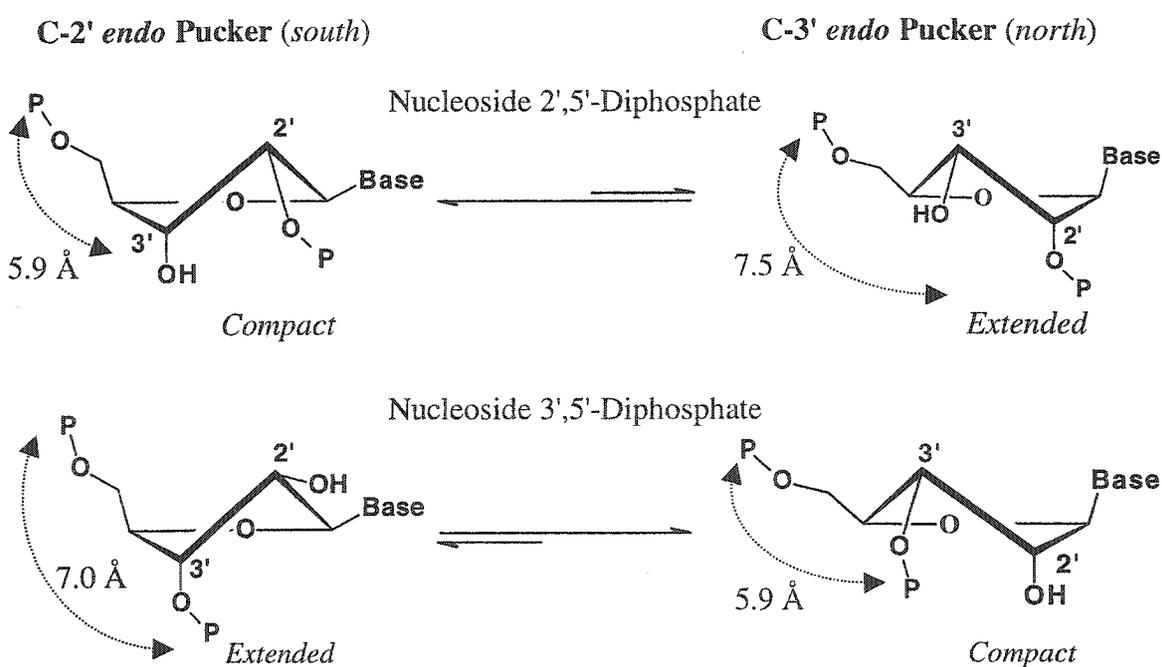


Figure 1.22: Preferred sugar conformations of 2',5'-RNA and 3',5'-RNA.¹⁴²

In the same way, NMR studies conducted on trimers of 2',5'-deoxyadenylic acid have shown that they favor C-3' *endo* pucker,¹¹⁹ a conformation again in sharp contrast to the isomeric 3',5'-DNA (which more commonly adopts the C-2' *endo* pucker). Moreover, 3'-deoxyuridine¹⁵⁰ and 3'-deoxyadenosine¹⁵¹ puckers exist predominantly in C-3' *endo* conformation. However, more recent NMR studies on 2',5'-DNA duplexes (referred to as isoDNA) indicate that they adopt A-type global conformation in solution with alternating C-2' *endo* and C-3' *endo* sugar repeats.¹³² This is in agreement with previous reports describing the crystal structure of 2',5'-linked RNA dimers and showing that their sugar puckers adopt alternating C-2' *endo* and C-3' *endo* conformations.^{97,117,152,153}

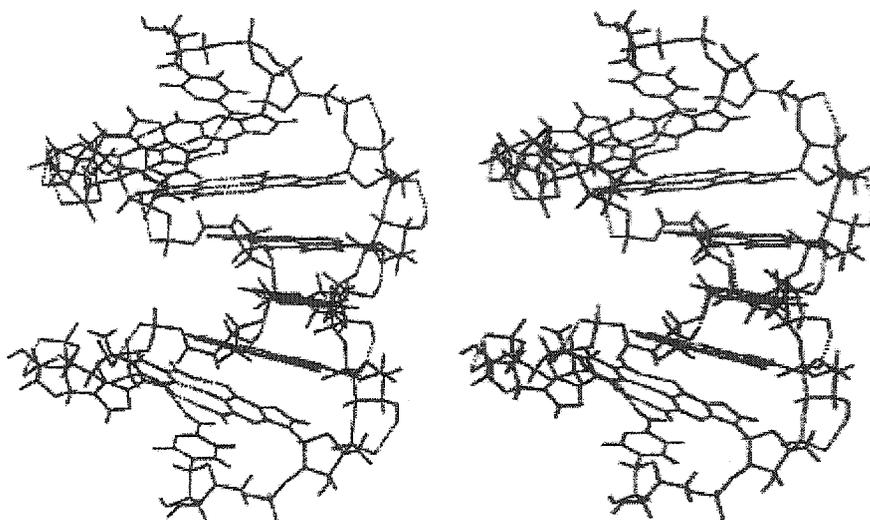


Figure 1.23: NMR-derived structure of a 2',5'-RNA duplex [(GCCGCGGC)₂]. Adapted from [Premraj, B. J.; Patel, P. K.; Kandimalla, E. R.; Agrawal, S.; Hosur, R. V.; Yathindra, N. *Biochem. Biophys. Res. Comm.* **2001**, 283, 537-543].

1.9 CHRONOLOGICAL SUMMARY OF THE BIOPHYSICAL PROPERTIES OF 2',5'-NUCLEIC ACIDS AND THE DEVELOPMENT OF 2',5'-NUCLEIC ACID CHEMISTRY

Over the years, numerous reports have been published on 2',5'-nucleic acids. **Table 1.2** compiles findings on this topic summarizing the chronological development of the biophysical and conformational properties of 2',5'-nucleic acids. Additionally, the table includes updated published reports on recent advancements in their chemical synthesis as well as their therapeutic uses. The year of publication, nature of discovery, and chief investigators are also listed.

Year	Scientific Finding	Chief Scientist
1967	Triple helix structures formed between 2',5'-oligoadenylates [A ₈] and poly rU are less thermally stable than those formed with 3',5'-oligoadenylates ($\Delta T_m = 6^\circ\text{C}$).	Michelson
1968	Dimers and trimers with a 2',5'-phosphodiester linkage are cleaved, like the 3',5'-linkage, under basic conditions. However, the 2',5'- linkage is more stable than the 3',5'-linkage to some hydrolytic nucleases.	Westheimer
1969	Crystal structure: 2',5'-RNA [A ⁺ pU dimer]	Shefter
1969	Base overlap in single stranded 2',5'-RNA is more extensive than in 3',5'-RNA	Ts'o
1970	Both 2',5'-r(ApA) and 3',5'-r(ApA) bind to poly rU with the same stoichiometry [1A:2U]	
1971	Predominant formation of 2',5'-bonds in template-directed non-enzymatic polymerization of ribonucleotides	Orgel /Renz
1972	Helix constraints force 2',5'-bond formation in prebiotic template-directed non-enzymatic ribonucleotide synthesis	Usher
1974	Coupling of adenosine 2',5'-cyclic phosphate on a poly(U) template leads to predominant 2',5'-bond formation	Orgel
1976	The 2',5'-phosphodiester linkage is less hydrolytically stable than the 3',5'-linkage in alkaline media (in double helices)	Usher
1977	Crystal structure of 2',5'-ApC RNase S complex reveals a preference for C-3' <i>endo</i> pucker	Wodak
1978	Inability of 2',5'-RNA to form double helical structures	Sarma
1981	Crystal structure: 2',5'-RNA [8 methyl GpC+ dimer]	Hamada
1981	NMR studies show that dimers of 2',5'-RNA exhibit predominant C-2' <i>endo</i> conformation	Altona
1982	Crystal structure: 2',5'-RNA [ApC+ dimer]	Parthasarathy
1983	2',5'-dA ₃ exhibits predominant C-3' <i>endo</i> pucker in solution	Altona
1986	2',5'-RNA exhibits better pre-organization than the 3',5'-RNA counterparts.	Anukanth
1986	2',5'-RNA exhibits better pre-organization than 3',5'-RNA	Srinivasan,
1988	2',5'-RNA dimers exist in dynamic equilibrium (C-2' <i>endo</i> ↔ C-3' <i>endo</i>) with predominant C-2' <i>endo</i> conformation.	Ogilvie

Year	Scientific Finding	Chief Scientist
1991	2',5'-rA ₁₀ forms a stable complex with poly(rU) whose melting temperature is lower than that of 3',5'-rA ₁₀ and poly(rU)	Damha
1992	Insertion of 2',5'-linkages within duplexes made of 3',5'-DNA always destabilize such helices. Duplex 2',5'-DNA is very weak and occurs only at high salt concentrations	Breslow
1992	Self-association of 2',5'-linked DNA structures with reduced stability relative to their 3',5'-linked isomers	Switzer
1992	Self-association of 2',5'-linked RNA structures with reduced stability relative to their 3',5'-linked isomers	Turner
1993	Binding selectivity of 2',5'-linked RNA to complementary single stranded RNA and not DNA Triplex formation of 2',5'-A ₁₀ with 3',5'-U ₁₀ and not 3',5'-dT ₁₀	Damha
1993	Crystal structure: 2',5'-RNA [ApA dimer]	Krishnan
1993	2',5'-DNA [A ₁₆ :2T ₁₆] forms a triplex structure of higher thermal stability than the isomeric 3',5'-triplex	Breslauer
1993	2',5'-linked oligoadenylates act as inhibitors of HIV-1 reverse transcriptase	Suhadolnik
1994	2',5'-linked oligoadenylates block DNA-topoisomerase-I in HIV infected cells	Muller
1994	2',5'-DNA comprising guanine and cytosine residues can associate into double helices even at low ionic strength	Switzer
1994	2',5'-thioformacetal DNA exhibits binding selectivity to single stranded RNA over DNA. Introducing few 2',5'-thioformacetal linkages within DNA helices resulted in destabilization.	Matteuci
1995	2',5'-RNA duplexes exist with predominant C-2' <i>endo</i> pucker	Yathindra
1995	2',5'-DNA duplexes adopt A-type helical conformation with alternating C-2' <i>endo</i> and C-3' <i>endo</i> sugar repeats	Switzer/Wang
1995	2',5'-dA ₈ selectively associate with poly U (and not poly T) to form duplexes of the same order of thermal stability as the 3',5'-dA ₈ : poly U duplex counterpart	Alul
1995	Reverse transcriptase reads through an RNA template bearing one 2',5'-linkage	Szostak

Year	Scientific Finding	Chief Scientist
1996	2',5'-RNA forms a stable triplex [A ₇ :2U ₇] of lower thermal stability than the corresponding 3',5'-triplex 2',5'-rA ₇ :2rU ₇ < 3',5'-rA ₇ :2rU ₇	Sawai
1996	Selective complex formation of 2',5'-DNA with single stranded RNA over single stranded DNA 2',5'-DNA:RNA hybrids are not substrates for RNase H	Breslow
1996	2',5'-DNA hybridizes to complementary RNA but not DNA. 2',5'-DNA:RNA hybrids and 2',5'-DNA duplexes adopt A-form helical conformations similar to that of RNA duplexes, as judged from CD studies.	Switzer
1997	2',5'-RNA acts as a template that directs oligomerization of mononucleotides, though less efficiently than 3',5'-RNA.	Switzer
1997	2',5'-RNA inserted within a 3',5'-DNA strand decreases its binding affinity to complementary target DNA, and does not evoke RNase H activity when bound to complementary RNA Phosphorothioate mixed backbone oligonucleotides (containing 2',5'-RNA and 3',5'-DNA) exhibit less non-specific binding to proteins than the normal phosphodiester linkages	Kandimalla
1997	2',5'-DNA (comprised of C,G, and T bases) forms a complex with complementary RNA (but not DNA) that does not elicit RNase H activity 2',5'-linked phosphorothioate oligodeoxynucleotides exhibited less non-specific binding with human cellular proteins relative to their 3',5'-linked analogues 2',5'-DNA containing a 7 nucleotide stretch of the isomeric 3',5'-DNA (gapmer) elicits RNase H activity	Bhan
1998	Homopyrimidine 2',5'-RNA recognizes nucleic acid double helices	Damha
1998	-First report describing switch in sugar pucker preferences between 2',5'- and 3',5'- nucleic acids. -Concept of 'compact' and 'extended' nucleotide repeats introduced for 2',5'-nucleic acids. -Nucleoside 2',5'-diphosphates adopt C-2' <i>endo</i> sugar pucker. -Molecular modeling studies of 2',5'-RNA duplexes, based on crystal structures of 2',5'-dimers, show alternating C-2' <i>endo</i> and C-3' <i>endo</i> sugar puckers	Yathindra

Year	Scientific Finding	Chief Scientist
1998	-First report describing the properties of 2',5'-RNA with mixed base composition (containing all four bases) -RNA:2',5'-RNA hybrid duplexes, though exhibiting similar CD conformational features to RNA:DNA hybrids, fail to elicit RNase H activity	Damha
1998	2',5'-linked decauridylate serves as a template for the ligation of activated 2',5'-(pA) ₂	Sawai
1999	2',5'-linkages incorporated within the domains I, II and III of a hammerhead ribozyme do not impair catalytic activity	Burlina
2001	2',5'-RNA duplexes adopt global A-type conformation with dominant C-2' <i>endo</i> sugar puckers	Yathindra
2001	2',5'-RNA folds intramolecularly to form hairpin structures of unusual stability	This Thesis
2002	Structural basis for the hybridization and biological properties of 2',5'-nucleic acids	Yathindra
2002	2',5'-RNA loops exhibit a uniquely folded structure different from that adopted by the 3',5'-RNA counterparts	This Thesis
2002	2',5'-RNA loops recognize HIV-1 RT and inhibit its RNase H activity when conjugated with the proper stems	This Thesis
2002	2',5'-ribocytidine nucleic acids associate into anti parallel stranded duplexes with hemiprotonated C-C+ base pairs	This Thesis
2002	A 2',5'-phosphodiester linkage is highly resistant to cleavage by yeast RNase III (Rnt1p)	This Thesis

1.10 THESIS OBJECTIVES

The Damha laboratory has had a long research interest in the physicochemical and biophysical properties of 2',5'-linked ribonucleic acids (2',5'-RNA). Several publications on the stacking interactions of 2',5'-RNA have appeared during the past 40 years. For instance, NMR structural studies suggested that there is more extensive base stacking and overlap between the bases in single stranded 2',5'-RNA than in 3',5'-RNA.^{108,109} More recent theoretical studies have also indicated that single stranded 2',5'-RNA exhibits better pre-organization, *i.e.*, fewer low energy conformations, than 3',5'-RNA.^{114,115} In light of these findings, and taking into perspective that base stacking is the central

criterion governing the folding and stability of hairpin loops, we reasoned that replacing the 3',5'-linkages with 2',5'-linkages would enhance the stability of these hairpin loops and promote better folding. Chapter II investigates the ability of 2',5'-linked RNA to fold into unusually stable loops. It also describes the effects of incorporating 2',5'-RNA linkages within hairpin loop and/or stem regions on the thermodynamic and conformational properties of diverse hairpin structures. Namely, a comparison between the ability of 3',5'- and 2',5'-RNA (UUCG) loops to stabilize various stem constructs is presented. The role of the sugar residues of the loop-closing base pair in both DNA and RNA hairpins is also assessed. Furthermore, the three-dimensional structure of the 2',5'-RNA loop, solved via high resolution NMR, identifies this loop as a novel structural motif.

Chapter III describes the design, evolution and identification of *hairpin aptamers* from a nucleic acid library constructed via *diversity-oriented* solid-phase synthesis, and their evaluation as specific inhibitors of HIV-1 reverse transcriptase RNase H activity.

Very recently, our laboratory interest has shifted to studying RNA interference mechanisms. In Chapter IV, we explore via the use of chemically-modified *hairpins* or 'stem-loop' structures the binding specificity and cleavage mechanism of yeast RNase III (Rnt1p), a double-stranded RNA-specific enzyme that cleaves the phosphodiester linkages at various positions on each side of RNA duplexes.¹⁵⁴ This key enzyme has been implicated in cellular RNA *interference* and thus understanding its substrate specificity as well as its mechanism of action are crucial.

Finally in Chapter V, the use of 3',5'- and 2',5'-RNA (UUCG) loops as a means to enhancing the stability and promoting the folding of C-tetraplexes is explored.

CHAPTER II: THERMODYNAMIC AND STRUCTURAL STUDIES OF HAIRPINS CONTAINING 2',5'-RNA LOOPS

2.1 BACKGROUND

RNA hairpins are prominent structural motifs that are associated with a wide variety of biological functions. They are of immense importance for recognition of RNA by proteins¹⁵⁵⁻¹⁵⁷ as well as for folding⁶⁷ and stabilizing RNA *in vivo* (for reviews, see³⁴). Hairpins having the tetranucleotide loop sequence UUCG are ubiquitous in ribosomal RNA.⁴² Tuerk *et al.*⁶⁷ found that messenger RNAs with the sequence 5'...C(UUCG)G...3' prevented reverse transcriptase from reading through, and that hairpin loops with this particular sequence are significantly more stable relative to those with homopolymeric tetraloops. Pioneering studies by Tinoco and co-workers revealed that the tetraloop is stabilized by extended base stacking, a reverse wobble U₅(anti):G₈(syn) base pair, and an electrostatic interaction between C₇ and the phosphate group linking U₅ and U₆ (**Figure 1.10 A**).⁵³⁻⁵⁵ The NMR structure revealed that the U₆ residue is not involved in any specific contact with other bases in the loop, and protrudes out into the solvent. These observations are consistent with chemical modification experiments in that the U₆ residue could be mutated to any base without loss of stability.⁶⁸ Furthermore, mutation of C₇ to a uracil residue destabilized the hairpin by 1.5 kcal/mol,⁶⁸ consistent with the proposed specific C₇-phosphate contact.

Slight perturbations in the sugar-phosphate backbone have a profound effect on hairpin stability. For example, Sakata⁵⁶ and Tinoco^{57,68} demonstrated that stability drops substantially when the loop riboses are replaced by 2'-deoxyriboses. Thus, 5'-UGAGC(UUCG)GCUC-3' forms a more stable folded structure compared to 5'-UGAGC(uucg)GCUC-3', 5'-ggac(ttcg)gtcc-3', or 5'-ggac(uucg)gucc-3'[where the small letters represent deoxyribose residues].^{56,57,68} Although there is general agreement that the loop 2'-hydroxyl groups impart stability, the molecular basis for this stabilization remains unclear. Molecular dynamics studies¹⁵⁸ revealed favorable electrostatic interactions between the 2'-hydroxyl group and O3', which would rigidify both the stem and loop regions; this 'pre-organization' would make the riboses more compact and hence more

stable than deoxyribose-based loops. Specific tertiary interactions involving 2'-hydroxyl groups have also been proposed.¹⁵⁹

Very little is known about structural RNA motifs containing 2',5'-linked internucleotide linkages. In fact, the hybridization properties of single stranded 2',5'-DNA^{135,136} and 2',5'-RNA^{121,122,125,141} are only beginning to be explored. As discussed in chapter I, these studies have shown that both single and multiple 2',5'-phosphodiester substitutions within RNA and DNA helices result in lower melting temperatures of duplexes. In addition, RNA:2',5'-RNA hybrids and pure 2',5'-RNA duplexes have lower melting temperatures than RNA duplexes, whereas DNA:2',5'-RNA hybrids are too unstable to be observed. Braich and Damha recently showed that branched pentaloops of the type $rA^*-[d(CC)]_2$ (where rA^* is joined to both $d(CC)$ units via *vicinal* 2',5' and 3',5'-phosphodiester linkages) induce the formation of DNA hairpins that are slightly thermally more stable than those with the native $d(CCCC)$ loop.¹⁶⁰ The domains I, II and III within a hammerhead ribozyme have been replaced by 2',5'-linkages without impairing catalytic activity.¹⁴⁵ To our knowledge, contiguous 2',5'-substitutions and the impact that these have on the structure and thermodynamic stability of RNA tetraloops have not been characterized.

In this chapter, we explore the effects of multiple 2',5'-internucleotide substitutions in the loop region of RNA hairpins. Specifically, we wished to learn if 2',5'-linked RNA (“R”), like the native RNA (“R”), can adopt unusually stable tetraloop structures, and if so, whether the same sequences are extra stable. Remarkably, we find that 2',5'-linked RNA (UUCG) tetraloops show a pattern of relative stability similar to that observed for the native RNA hairpins with identical base sequence. We also find that the stability of both loops is highly dependent on the nature of the $C_4:G_9$ loop-closing base pair and how they stack in the overall context of the stem. Unlike the native system, the 2',5'-loop analog (UUCG) remains extra stable with various stem hybrids whether the stem is either 2',5'-RNA:RNA, 2',5'-RNA:DNA, or 2',5'-RNA:2',5'-RNA. These observations not only advance our knowledge on the molecular origin of tetraloop stabilization, but also have implications for the use of 2',5'-RNA as novel structural motifs (*e.g.* RNA-based “aptamers” as discussed in chapter III).

2.2 EXPERIMENTAL DESIGN

The compounds studied in this work are hairpins containing the conserved 5' (UUCG) 3' loop sequence (Figure 2.1). The stem is defined by C₁-G₄ and G₉-C₁₂ residues, with C₄:G₉ constituting the loop closing base pair. We have inserted 2',5'-linkages at various positions in the stem and/or loop regions and studied their effect on hairpin thermodynamic stability. Thus, all oligomers described herein share the same stem and loop sequence, but differ in the sugar composition and/or the connectivity of phosphodiester bonds. The base sequence is identical to that reported by Tinoco and co-workers,⁶⁸ thus providing a good model system with which the thermodynamic effects can readily be obtained. For example, hairpins RRR, RDR, and RRR (where "R" represents a 2',5'-linked RNA segment, "R" represents a 3',5'-linked RNA segment, and "D" represents a 3',5'-linked DNA segment) were designed to test the effect of the tetraloop structure on the stability of RNA hairpins. Similarly, compounds DRD, DRD, and DDD were designed to assess the stabilizing effect of the various tetraloops on duplex DNA. Moreover, comparison of an all-DNA, all-RNA, and all-2',5'-RNA hairpins provides a direct means to compare the properties of 2',5'-RNA with the more common nucleic acids.

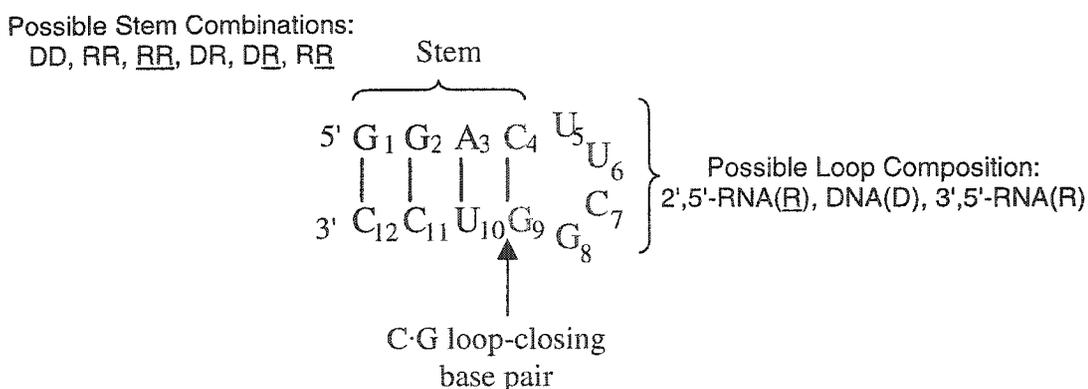


Figure 2.1: Secondary structure and base sequence of the hairpin under study. The structure consists of a loop (U₅U₆C₇G₈), and a stem region [defined by G₁-C₄ and G₉-C₁₂ residues] that includes the "loop-closing base pair" C-G. The stem can be either DNA:DNA "DD", RNA:RNA "RR", 2',5'-RNA:2',5'-RNA "RR", DNA:RNA "DR", DNA:2',5'-RNA "DR", or RNA:2',5'-RNA "RR", while the loop was modified with DNA "D", RNA "R" and 2',5'-RNA "R" residues.

In our base-sequence nomenclature, DNA residues are represented by small letters, and 3',5'-linked ribonucleic acid (RNA) residues are represented by capital letters, while 2',5'-linked ribonucleic acid (2',5'-RNA) residues are represented by capital underlined letters. For example, *gu* would refer to $dG_{3'p5'} dU_{3'p}$, *GU* would refer to $rG_{3'p5'} rU_{3'p}$ while *GU* would denote a $rG_{2'p5'} rU_{2'p}$ fragment. The UUCG loop in the hairpin GGAC(UUCG)GUCC refers to the 2',5'-linked RNA tetraloop, *i.e.*, $U_{2'p5'}U_{2'p5'}C_{2'p5'}G_{2'p}$.

2.3 HAIRPIN SYNTHESIS AND CHARACTERIZATION

Hairpins were synthesized on an Applied Biosystems (381A) or an Expedite™ 8909 PE synthesizer following slight modifications of standard phosphoramidite chemistry.^{14,161,162} The standard 5'-*O*-dimethoxytrityl-2'-*O*-*tert*-butyldimethylsilylribonucleoside-3'-*O*-(2-cyanoethyl)*N,N'*-diisopropyl phosphoramidite monomers were used for normal RNA synthesis while the corresponding ribonucleoside 2'-*O*-phosphoramidites were used for 2',5'-RNA synthesis (**Figure 2.2**).

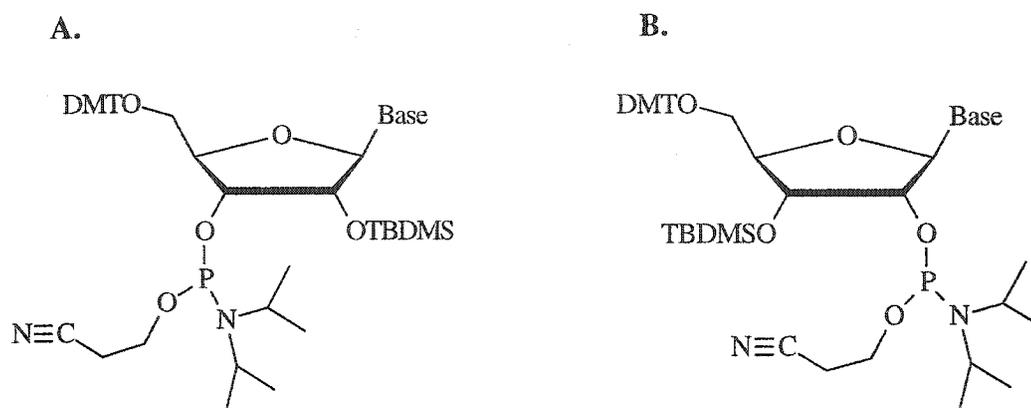


Figure 2.2: Phosphoramidite monomers used in the solid phase synthesis of (A) 3',5'-RNA and (B) 2',5'-RNA. TBDMS denotes *t*-butyl dimethylsilyl protecting group. DMT denotes 4',4'-dimethoxytrityl protecting group. RNA synthesis proceeded through the phosphite triester approach as shown in **Figure 1.6**. The base moiety can be any of adenine, guanine, cytosine, or uracil.

The monomer coupling times and their concentration for 2',5'-RNA synthesis were the same as those used for 3',5'-RNA synthesis (see experimental **Section 7.2.4**). All other conditions for 2',5'-RNA solid phase synthesis were the same as those for 3',5'-RNA

synthesis, except for the coupling reaction. 1H-tetrazole (0.5 M in acetonitrile), one of the most commonly used coupling activators¹⁹ which typically gives excellent yields in 3',5'-RNA synthesis (>99%), failed to produce good coupling yields with 2',5'-RNA coupling (80-90%). To recall, 1H-tetrazole has a dual function during the activation step in solid phase oligonucleotide synthesis.^{19,20} First, it acts as a weak acid that protonates the amino group of the phosphoramidite moiety, an activation step that is fast and reversible. Second, it acts as a nucleophile that displaces the diisopropylamino group, a rate-determining step that results in the formation of a very reactive intermediate ready for coupling with the next monomer. Hence, switching to a slightly more acidic activator would help speed up the protonation step and thus activation. We resorted to one of the commercially available and commonly used alternative activators, 5-ethylthio-1H-tetrazole (pK_a 4.28) (Figure 2.3).¹⁶³⁻¹⁶⁵ Despite the fact that it improved the yields for 2',5'-RNA synthesis as previously observed by members of our group,¹⁴¹ the emergence of new side products referred to as “long mers” was very prominent on the denaturing gels.¹⁶⁶ We observed the formation at least one or more complex species (slower moving than the desired band) by gel electrophoresis (Figure 2.4). At the time, the chemical discovery group at NeXstar Technology Products Inc., a company specializing in the discovery, manufacturing and marketing of oligonucleotide-based drugs, reported the discovery of 4,5-dicyanoimidazole (DCI) as a better alternative coupling reagent to tetrazole.¹⁶⁷ DCI is not a new compound, in fact its first synthesis dates back to 1949 (US patent 2,534,331, 1949). However, its use in the solid phase synthesis of DNA and RNA was first reported in 1995 by Picken and co-workers.¹⁶⁷ DCI (pK_a 5.2) is less acidic but more nucleophilic than tetrazole (pK_a 4.8, similar to that of acetic acid),¹⁶⁸ and thus speeds up the rate-determining step (it has been reported to double the coupling rate compared to tetrazole¹⁶⁷). We have tried DCI and found it the best activating reagent so far for the synthesis of 2',5'-RNA as well as 2',5'/3',5'-RNA chimeras. It improved the overall coupling efficiency significantly (>98%) resulting in easier separation of the desired compound. It also improved the isolated yields drastically compared to 1H-tetrazole and 5-ethylthio-1H-tetrazole.

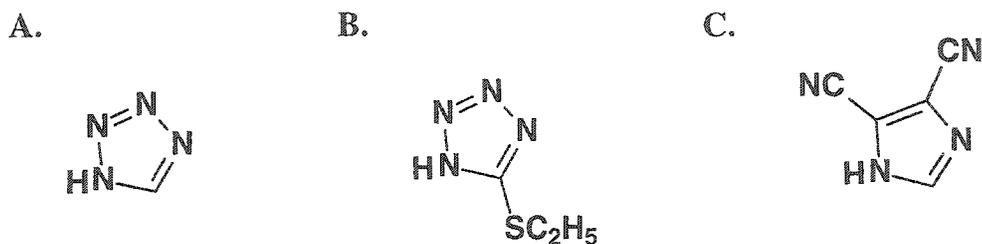


Figure 2.3: Chemical structures of various coupling reagents used in oligonucleotide solid phase chemistry. (A) 1H-tetrazole, the most commonly used coupling reagent; (B) 5-ethylthio-1H-tetrazole, a more acidic coupling reagent than tetrazole that helps improve the coupling efficiency; (C) 4,5-dicyanoimidazole, a more nucleophilic coupling reagent than tetrazole.

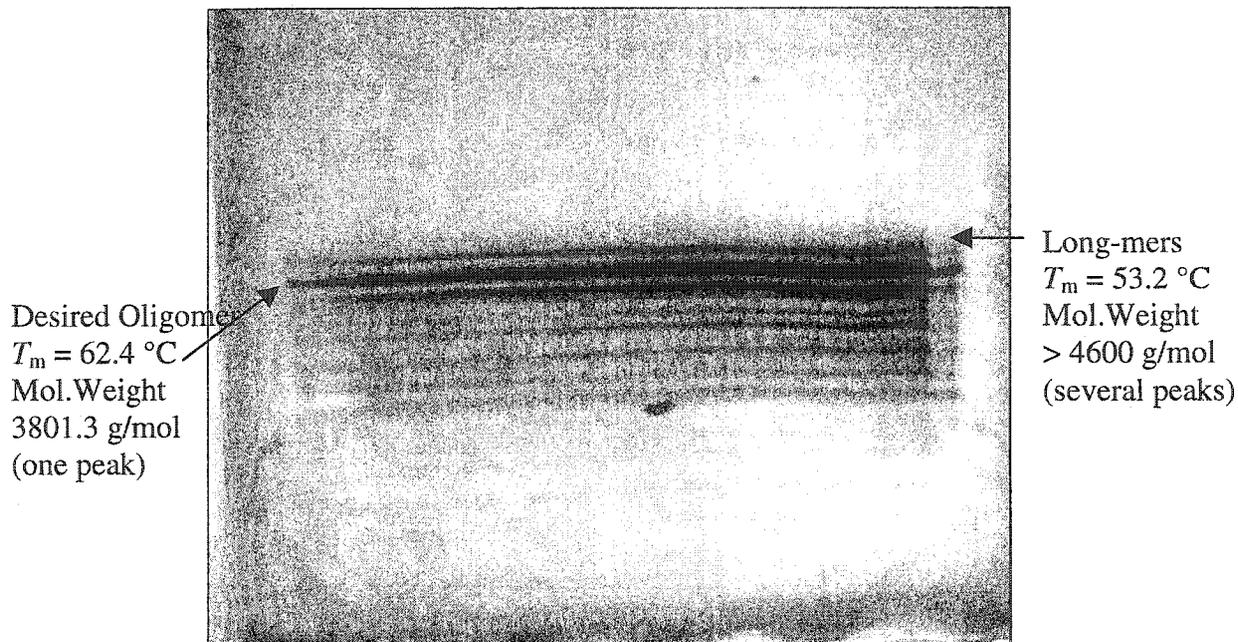


Figure 2.4: Preparative polyacrylamide gel electrophoresis (25%) showing the formation of “long-mers” from representative 2',5'-RNA synthesis using 5-ethylthio-1H-tetrazole. These side products were excised from the gel and their molecular weights were determined by MALDI-TOF mass spectrometry. Also, they were shown to melt at lower temperatures than the full-length desired oligomer (product).

Following synthesis, the oligomers were deprotected and purified by either denaturing polyacrylamide gel electrophoresis (24%, 7 M urea) or ion-exchange HPLC. Desalting was done by size-exclusion chromatography (Sephadex G-25) or reversed-phase column chromatography (Sep-Pak cartridges - see experimental **Section 7.3**). The purity of all sequences was assessed using 24% analytical gels (**Figure 2.5**) or analytical HPLC (see experimental **Section 7.4**) and the structures were confirmed by MALDI-TOF mass spectrometry (see **Table 7.3** - experimental **Section 7.4.2**). The integrity of some representative hairpins was also qualitatively verified by 1D-NMR (**Figure 2.6**). The sum of the peak integrations of the residual sugar H1' signals along with cytosine/uracil H5 signals correspond to 19 protons in total [5.2-6.0 ppm], in agreement with the total number of protons in this region predicted for an RNA-based oligomer of the sequence [GGAC(UUCG)GUCC]. Moreover, the total integration of the peaks in the aromatic region [7.0-8.0 ppm] corresponding to the doublet signals of cytosine/uracil H6 protons, H2/H8 of A, and H8 of G residues indicate a total number of 13 protons, again in agreement with the specific base-sequence under study.

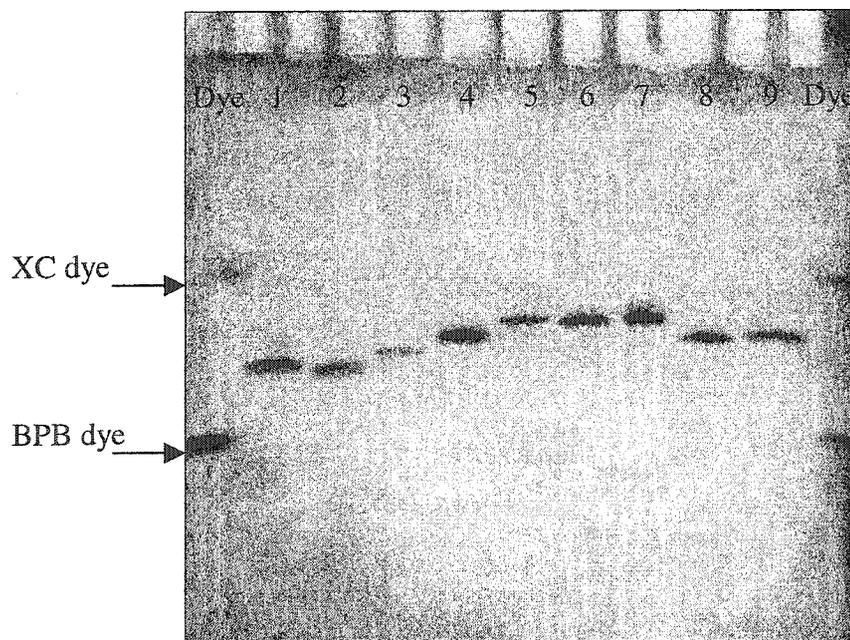


Figure 2.5: Analytical polyacrylamide (24%) denaturing gel showing representative purified oligonucleotides. Lane 1: DDD; Lane 2: DRD; Lane 3: DRD; Lane 4: RRD; Lane 5: DRR; Lane 6: RRR; Lane 7: RRR; Lane 8: RRR; Lane 9: RRR.

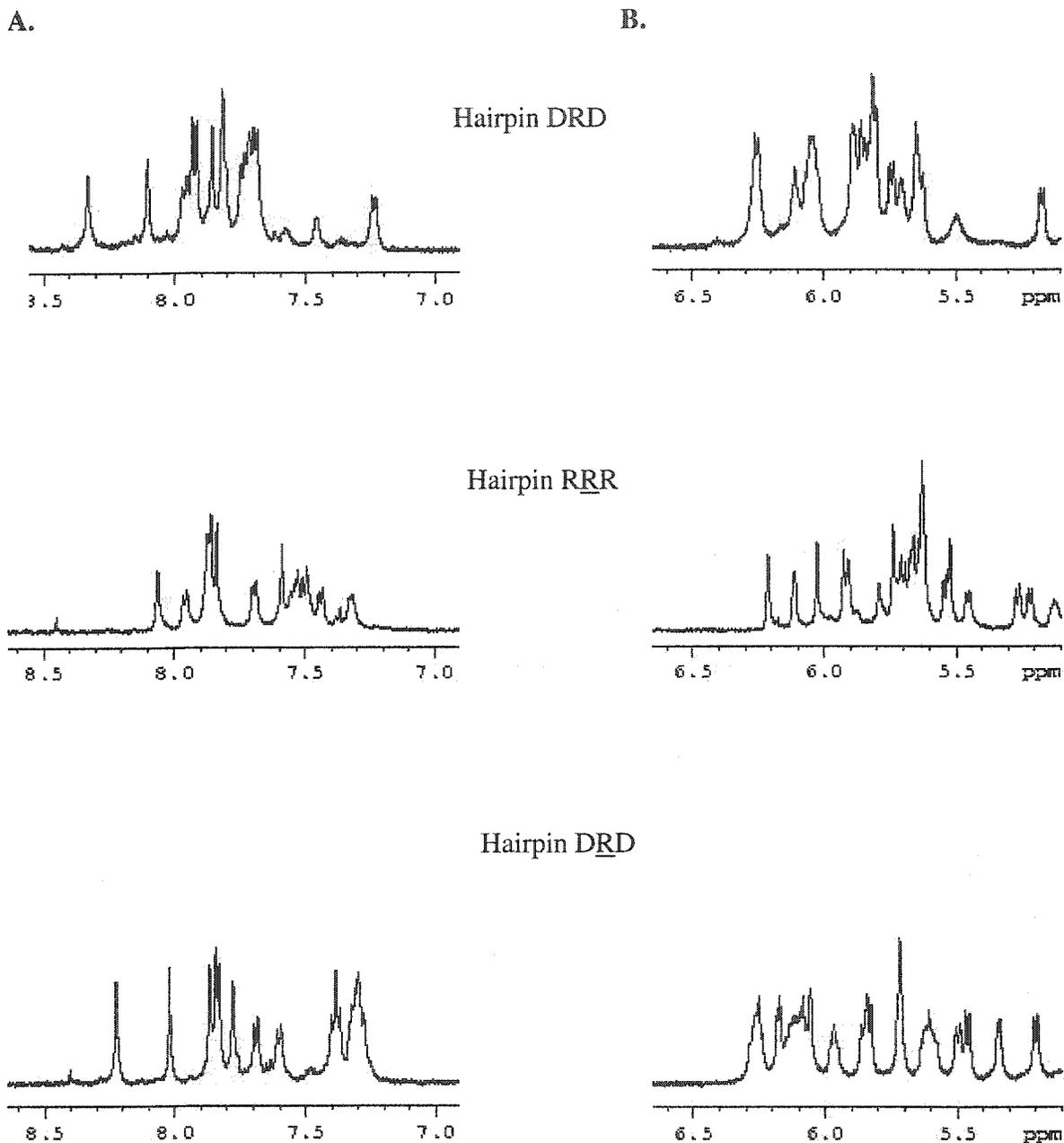


Figure 2.6: 1D-NMR spectra of various representative hairpins showing (A) the aromatic [7-8 ppm] and (B) the sugar H1' regions. The integrity of hairpin sequence was checked by peak area integration. According to the hairpin sequence, twelve H1' protons + 7 H5 (C,U) should be observed, in agreement with the overall peak integration value. In the aromatic region, 7 H6 +H2 + H8 corresponding to a total of 9 protons, again in agreement with the total peak area.

2.4 HAIRPIN VERSUS DIMER FORMATION

During helix formation, a single stranded self-complementary oligonucleotide can undergo one of two competing dynamic processes in solution (Figure 2.7). The first (path a) is *unimolecular* and involves self-association of one single strand on its own to form a *hairpin* species. The second (path b) is *bimolecular* where two single strands hybridize to form a *duplex* species. The relative amounts of hairpin and duplex species depend on solution conditions, with the duplex being favored at high salt concentrations and the hairpin at low salt concentrations.²³ Sometimes, both can form simultaneously in solution (at intermediate salt concentrations), leading to biphasic behavior in UV melting curves. For a unimolecular process, the melting temperature (T_m) is not dependent on oligonucleotide concentration.

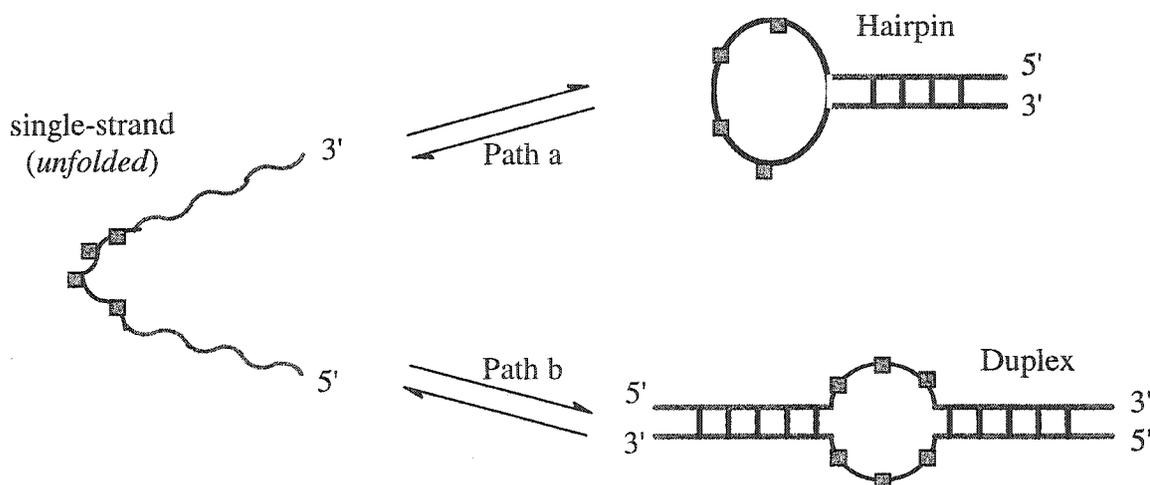


Figure 2.7: Schematic diagram illustrating the competing dynamic processes between hairpin and dimer states. The blue and red regions represent complementary base pairs, while the black region contains unpaired nucleotide residues.

To ensure that only *hairpin* species formed in solution, the synthesized oligonucleotide samples were dissolved in appropriate buffer [0.01 M Na_2HPO_4 , 0.1 mM Na_2EDTA , pH 7.0] and melted over at least a 30-fold range in nucleic acid concentration with no detectable change in melting temperature (Figure 2.8). This concentration-independent behavior of T_m confirms a *unimolecular* process; in other words, the equilibrium transition of a hairpin to a single stranded species in solution under the specified buffer conditions.

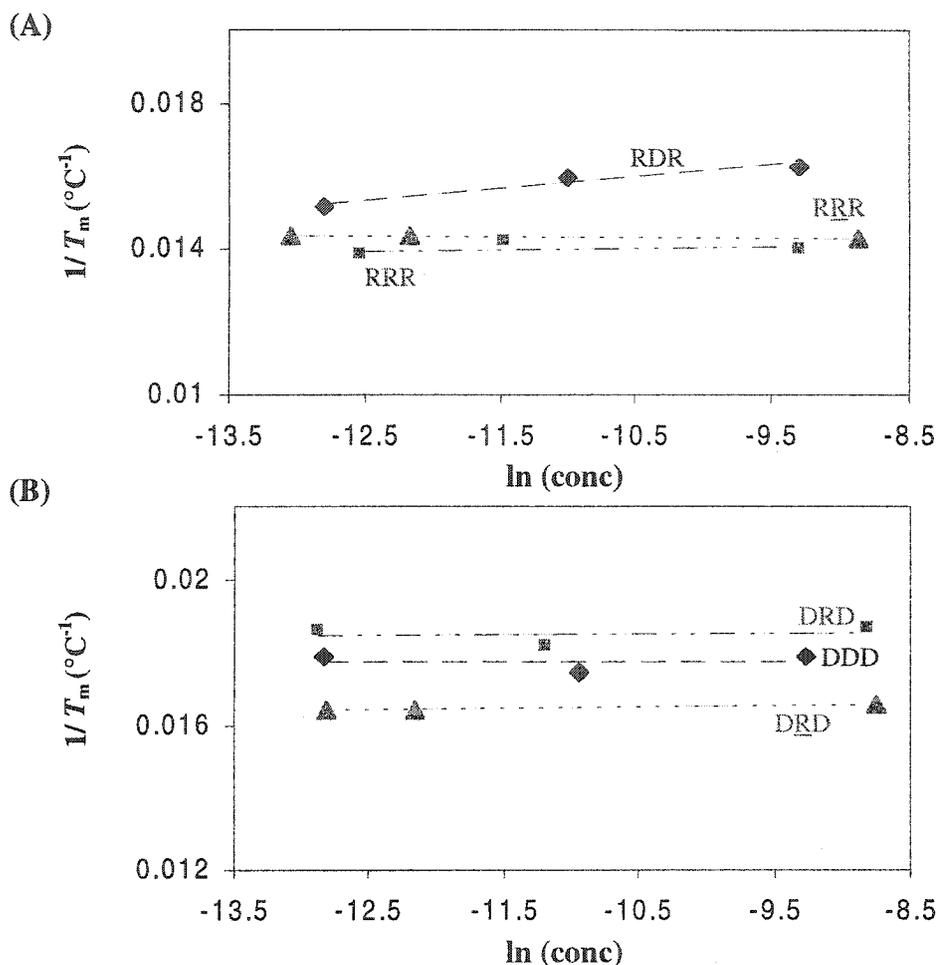


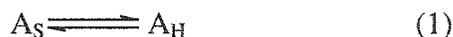
Figure 2.8: Van't Hoff plots of representative (A) RNA and (B) DNA stem hairpins differing in loop composition (DNA *versus* RNA *versus* 2',5'-RNA loops). The spectra show concentration independence over a 50-fold range, thus confirming a *unimolecular* transition.

The thermally induced melting profile for any nucleic acid species correlates with its helix-to-coil transition and can be obtained by monitoring the increase in UV absorbance (at an appropriate wavelength) with increase in temperature. Accordingly, all hairpin samples ($\sim 4.5 \mu\text{M}$) under scrutiny were studied by UV absorbance and their thermally induced melting curves showed unimolecular cooperative and reversible profiles. They were heated to 90°C at $0.5^{\circ}\text{C}/\text{min}$, then allowed to cool to 5°C at the same rate, then instantly heated again to 90°C (data not shown). No hysteresis was observed, that is, all hairpin samples were able to reform with the same melting

temperatures, indicating fast kinetics of association (no difference in the value of T_m of association and T_m of re-association).

2.5 THERMODYNAMIC CHARACTERIZATION AND DATA ANALYSIS

The thermodynamic parameters (ΔH° , ΔS° , ΔG°) for hairpin formation were derived from the UV absorbance melting profile of each individual hairpin. The unimolecular equilibrium under study can be represented in the form of the following reaction,



where A_S is the denatured *single-stranded* species and A_H is the base paired *hairpin* species. **Figure 2.9** represents a typical UV melting curve of a self-complementary nucleic acid (hairpin). At low temperatures (below the thermal transition), the oligomer is believed to be in the *hairpin* ordered state (A_H). At high temperatures, the species in solution is believed to be the unfolded *single-stranded* state (A_S). Hence, when the thermodynamics of association of two different hairpins are analyzed, they are compared relative to a common final state in which the denatured single-strands of both hairpins are assumed to be thermodynamically degenerate.²³ Assuming an all-or-none (two state) model,²¹ the melting temperatures and thermodynamic parameters for hairpin formation (*unimolecular* process) were computed following standard published literature procedures.^{21,22} The melting temperature (denoted by T_m , and defined as the temperature at which 50% of the duplex is in the dissociated state) was calculated according to the base line method.²¹ This was achieved by constructing linear least squares lines for the associated and dissociated regions of the melting curve and then extrapolating to both ends of the graph (**Figure 2.9**). If we define an equilibrium property α as the fraction of molecules in the hairpin state, then α can be calculated at each temperature, and the melting profile can consequently be converted to a plot of (α) versus temperature (**Figure 2.10 A**). Once this is constructed, the T_m value can be obtained by interpolating to $\alpha = 0.5$ (**Figure 2.10 A**). A line was drawn tangent to the linear part where the melting transition occurs.

The transition equilibrium constant, K_{eq} , can be defined in terms of α and calculated at various temperatures according to the following equation:

$$K_{eq} = \frac{[\text{Product}]}{[\text{Reactant}]} = \frac{[A_H]}{[A_S]} = \frac{(\alpha)}{(1-\alpha)} \quad (2)$$

Using the van't Hoff relationship,

$$\partial \ln K / \partial (1/T) = -\Delta H^\circ / R \quad (3)$$

where R is the gas constant and T the temperature in Kelvin units, a plot of $\ln K$ versus $1/T$ was generated (Figure 2.10 B), and the slope and intercept of the calculated line yielded the standard enthalpy (ΔH°) and standard entropy (ΔS°) of hairpin formation respectively, from which the standard Gibbs free energy at 37 °C (ΔG°_{37}) was calculated using the following equations:

$$\Delta G^\circ = \Delta H^\circ - T\Delta S^\circ \quad (4)$$

$$\Delta G^\circ = -RT \ln K_{eq} \quad (5)$$

In order to validate our calculations, ΔH° was sometimes calculated, via another method, from the slope of the line drawn tangent to the α versus T curve (Figure 2.10 A) according to the equation:

$$\Delta H^\circ = 4 R T_m^2 \frac{[\partial \alpha]}{[\partial T]} \quad (6)$$

All UV melting measurements were carried out at least five independent times. In each run, a newly prepared sample solution was used. The thermodynamic parameters were calculated following the above procedure, and then all values from the separate measurements (for the same thermodynamic property) were averaged. The reported error limits in the thermodynamic data represent standard deviations with all values weighted equally.¹⁶⁹ They vary within acceptable standard limits^{68,69,170} and are *within* $\pm 7.5\%$ for ΔH° , $\pm 7.5\%$ for ΔS° , and ± 0.2 kcal/mol for ΔG° . The error in T_m is within $\pm 1^\circ\text{C}$.

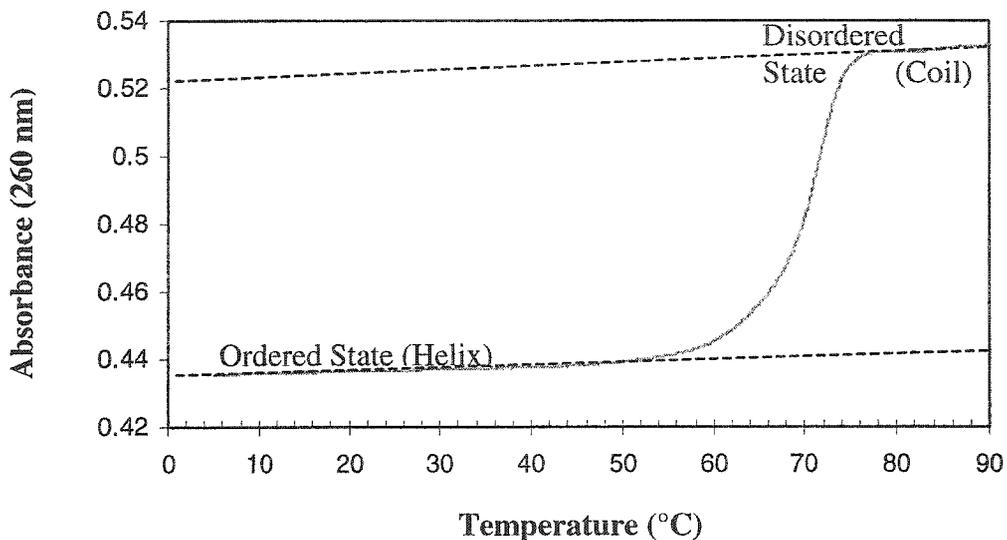


Figure 2.9: Typical UV melting profile of a sample hairpin showing the ordered (low temperature region) and disordered (high temperature region) states. In the first step towards calculating thermodynamic parameters, linear least squares lines are constructed along the ordered and disordered regions and extrapolated to both ends of the curve.

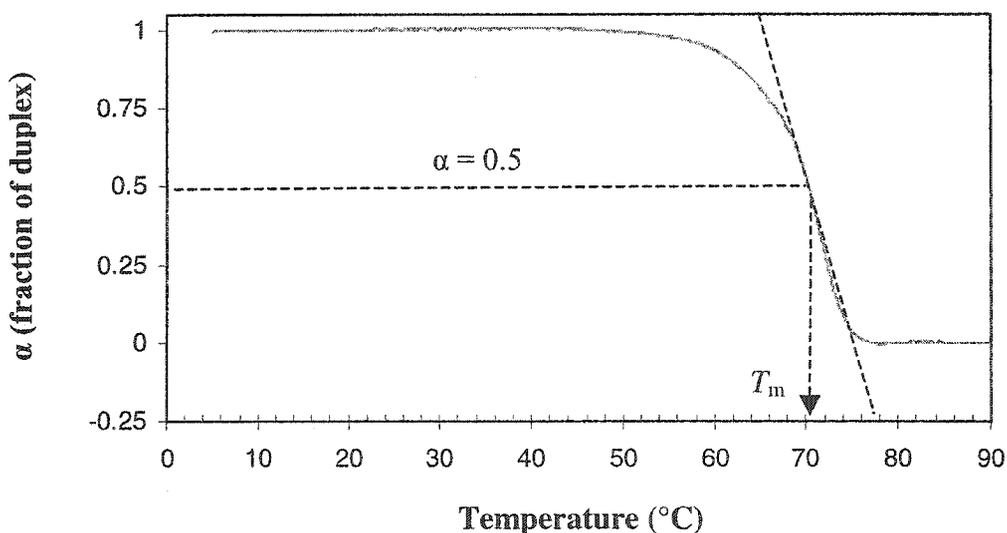


Figure 2.10 A: Plot of α versus temperature (derived from the UV melting curve). α is defined as the fraction of single strands in the *hairpin* state. T_m is the temperature at which the thermal transition is half complete and is referred to as melting temperature. The slope of the line drawn tangent to the thermal transition yields ΔH° according to the van't Hoff equation (6).

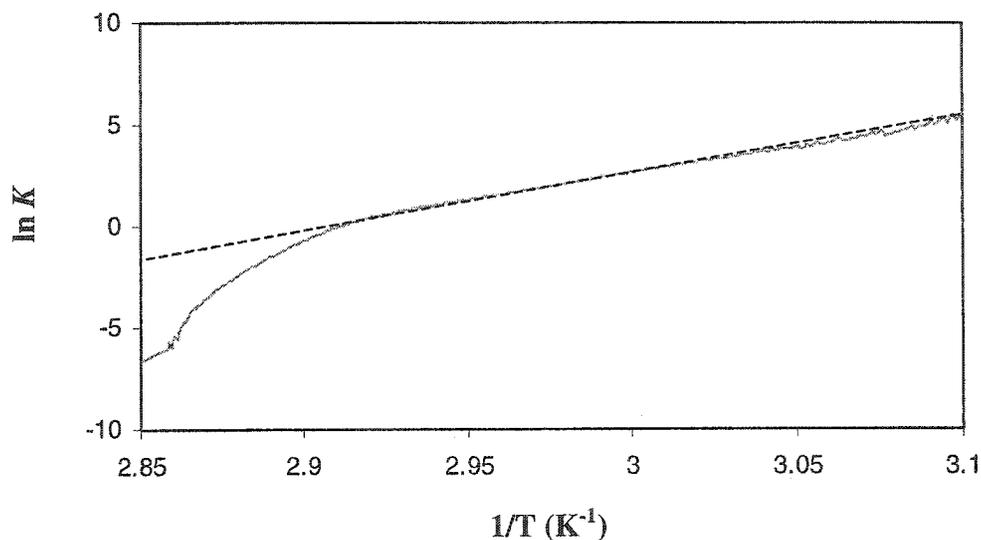


Figure 2.10 B: Van't Hoff plot of $\ln K$ versus $1/T$. The linear line is extrapolated to both ends of the curve. The slope and intercept of the calculated line yielded the standard enthalpy (ΔH°) and standard entropy (ΔS°) of hairpin formation respectively according to equations (3), (4) and (5). ΔG° was calculated accordingly [equation (4)].

2.6 THERMODYNAMIC PROPERTIES OF RNA HAIRPINS: R *VERSUS* R LOOPS

The key properties of RNA hairpins containing 2',5'-RNA tetraloops are brought out in **Table 2.1**. The melting curves show a single cooperative transition (T_m) that is independent of oligonucleotide concentration (data not shown) over at least a 30 fold concentration range [0.01 M Na_2HPO_4 , 0.1 mM Na_2EDTA , pH 7.0]. This result supports the view that these oligonucleotides form hydrogen-bonded, base-stacked structures reversibly in solution by folding *intramolecularly* to a hairpin conformation.

The top section of **Table 2.1** shows the T_m and thermodynamic parameters of RNA hairpins differing only in the loop composition. As mentioned earlier, their base-sequence is identical to that reported by Tinoco and co-workers,⁶⁸ thus providing a good model system with which the thermodynamic effects can readily be obtained. The calculated ΔG°_{37} and T_m values we measured for the C(UUCG)G hairpin [*i.e.*, RRR; **2.1**] are very similar to what has been obtained by Tinoco.

Table 2.1: Thermodynamic parameters of hairpins with RNA stems

Designation	Code	5'- Hairpin -2'/3'	T_m (°C)	% H	ΔH° (kcal/mol)	ΔS° (e.u.)	ΔG°_{37} (kcal/mol)
<u>2.1</u>	RRR	GGAC(UUCG)GUCC	71.8	8.5	-53.4	-154.8	-5.4
<u>2.2</u>	RRR	GGAC(UUCG)GUCC	69.3	9.6	-55.6	-162.1	-5.3
<u>2.3</u>	RDR	GGAC(uucg)GUCC	63.4	7.3	-47.3	-140.1	-3.7
<u>2.4</u>	RR ¹ R	GGAC(UACG)GUCC	62.3	8.1	-45.5	-135.6	-3.4
<u>2.5</u>	RR ² R	GGAC(UUUG)GUCC	67.9	8.9	-53.5	-158.2	-5.1
<u>2.6</u>	RR ³ R	GGAC(UUUU)GUCC	60.5	7.3	-46.7	-139.9	-3.3
<u>2.7</u>	RR <u>U</u> R	GGAC(UUCG)GUCC	60.2	7.3	-52.0	-155.8	-3.7
<u>2.8</u>	R <u>C</u> RR	GGAC(UUCG)GUCC	62.6	8.2	-52.1	-155.0	-4.0
<u>2.9</u>	RR <u>G</u> R	GGAC(UUCG)GUCC	58.0	6.4	-45.9	-138.7	-2.9
<u>2.10</u>	R <u>C</u> RR	GGAC(UUCG)GUCC	61.5	8.2	-47.2	-141.0	-3.5

Measurements were made in 0.01 M Na₂HPO₄, 0.1 mM Na₂EDTA buffer, pH 7.00 ± 0.02; oligonucleotide concentration ~ 4.5 μM. Values represent the average of at least five independent measurements. Errors in thermodynamic parameters are within ± 7.5% for ΔH° and ΔS° , and ± 0.20 kcal/mol for ΔG°_{37} . Error in T_m is within ± 1 °C. For a more accurate calculation, ΔG°_{37} was determined from ΔH° and ΔS° before rounding off. Capital letters represent RNA residues; underlined letters are 2',5'-RNA residues (*e.g.* UC = U_{2p5'}C_{2p}); DNA residues are represented by small letters; bold letters are base mutations in the loop. Percentage hypochromicity (%H) was calculated from UV absorbances of the hairpin (A_0) and fully denatured species (A_f) using the following equation: %H = ($A_f - A_0$) / A_f .

Its melting temperature is *ca.* 20 °C higher than a hairpin of purportedly normal thermodynamic stability, such as the hairpin with the C(UUUU)G loop.⁶⁸ **Table 2.1** also shows that RDR [**2.3**] had intermediate stability whereas RRR [*i.e.*, C(UUCG)G loop; **2.2**] is unusually stable.

A comparison between their thermally-induced melting profiles is shown in **Figure 2.11 A**. To determine if the apparent stabilization imparted by the 2',5'-tetraloop was dependent on loop sequence, we synthesized hairpins containing mutations within the loop (**Section 2, Table 2.1**). We found that the hairpin with the C(UUCG)G loop (RRR; $T_m = 69.3$ °C) was of comparable stability to that of the RR²R [**2.5**] hairpin with the C(UUUG)G loop ($T_m = 67.9$ °C), which in turn was significantly more stable than RR³R [**2.6**] hairpin with the C(UUUU)G loop ($T_m = 60.5$ °C). This does not parallel perfectly well the effects observed by Tinoco for the native RNA system,⁶⁸ and raises the interesting possibility that the 2',5'-loop bears a global structural motif that is different from that of the native 3',5'-loop. These studies manifest that the C7 residue does not significantly participate in the loop stabilizing interactions, a phenomenon in sharp contrast to the native loop structure where U6 is the residue that does not contribute to thermal stability. As shown in **Figure 2.11 B**, the negative CD band at *ca.* 220 nm was reduced for the RRR hairpin in comparison to the same band in the spectrum of RRR; however the overall CD spectrum of RRR is very similar to that of RRR indicating that they share conformational features. In spite of that and remarkably, the somewhat large decrease in T_m on changing the loop from C(UUCG)G to C(UACG)G is not in agreement with this notion (RRR *versus* RR¹R; $\Delta\Delta G_{37}^\circ = 1.9$ kcal/mol), since according to the native loop structure, U6 residue is not involved in any interaction and can be mutated to any base without loss of hairpin stability.⁶⁸ These findings indicate that the second loop residue (U6) in the 2',5'-RNA loop contributes significantly to the *intramolecular* loop stabilizing interactions and yet, the 2',5'-RNA loop still retains a high degree of stability compared to 3',5'-RNA loops. Clearly, high-field NMR analysis (discussed at the end of this chapter) and/or crystallographic work on an RRR hairpin will be required in order to elucidate the finer structural features of 2',5'-loops that are responsible for the apparent anomalous behavior with this system.

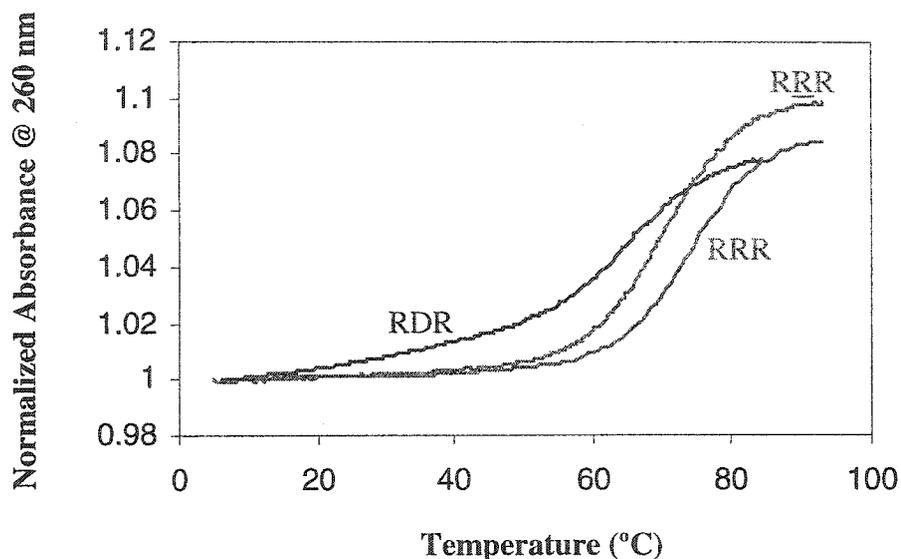


Figure 2.11 A: Normalized melting temperature profiles of RNA hairpins with various loop compositions. Measurements were done in 0.01 M Na_2HPO_4 , 0.1 mM Na_2EDTA buffer, pH 7.00 ± 0.02 at an oligonucleotide concentration $\sim 4.5 \mu\text{M}$.

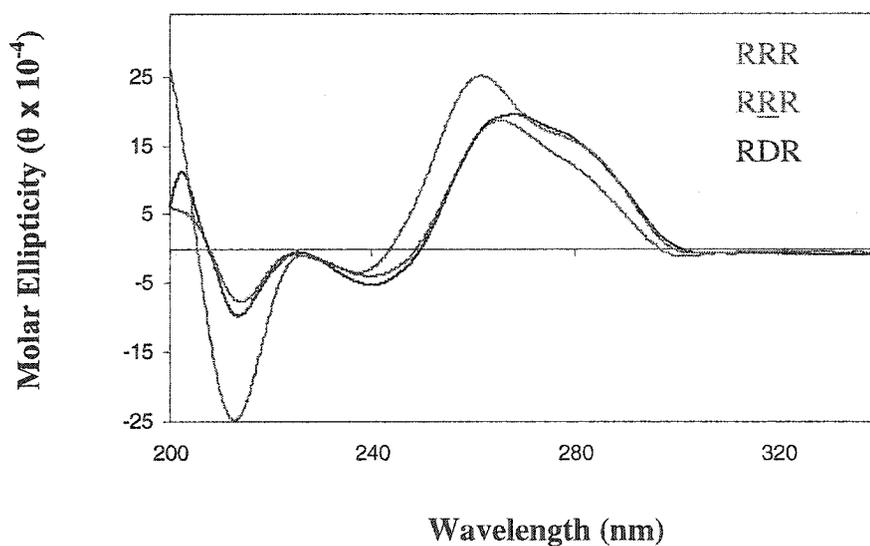


Figure 2.11 B: Circular dichroism spectra at 22 °C of RNA hairpins with various loop compositions. All measurements were done in 0.01 M Na_2HPO_4 , 0.1 mM Na_2EDTA buffer, at pH 7.00 ± 0.02 . Molar ellipticities were normalized to strand concentration. Oligonucleotide concentration $\sim 4.5 \mu\text{M}$.

To further assess the effect of 2',5'-substitutions on loop stability, we also prepared hairpins containing one 2',5'-linkage at, or near, the loop region, namely C(UUCG)G [2.7], C(UUCG)G [2.8], and C(UUCG)G [2.9] (Section 3, Table 2.1). The T_m data for the ...C(UUCG)G... hairpin (RRUR; $T_m = 60.2$ °C) indicate that a major disturbance of the native loop structure occurs as a result of a single modification (*i.e.*, U/G wobble pairing is disrupted). Also, we noted a 7-12 °C drop in T_m and a 1.4-2.5 kcal/mol increase in free energy at 37 °C when the 2',5'-phosphodiester linkage is placed outside the tetraloop sequence [*i.e.*, C(UUCG)G & C(UUCG)G versus C(UUCG)G]. In these cases, differences in stacking of the loop-closing base pair may be responsible for the drop in stability caused by introducing the 2',5'-linkage. No destabilization is observed for C(UUCG)G when all 2',5'-linkages are placed *within* the loop region.

Finally, we assessed the stability of a hairpin containing five consecutive 2',5'-internucleotide linkages in the loop region, *i.e.*, 5'-GGAC(UUCG)GUCC-3' [2.10] (Table 2.1). The large decrease in T_m on adding an extra 2',5'-linkage near the loop indicates a major disturbance of the overall structure.

2.7 THERMODYNAMIC PROPERTIES OF DNA HAIRPINS: DRD VERSUS DRD & DDD HAIRPINS

The stabilization conferred by 2',5'-loops is not limited to hairpins containing RNA stems. Figure 2.12 shows that the T_m value for the helix-to-coil transition is greater for DRD than for DDD. A more significant observation is that DRD [2.15] is more stable than the DRD [2.14] regioisomer by *ca.* 1.0 kcal/mol (Table 2.2; $\Delta T_m = 6.8$ °C). This dramatic decrease in thermodynamic stability reflects better base pairing and stacking interactions when the loop is 2',5'-linked rather than 3',5'-linked. In fact, the T_m value of ggac(UUCG)gtcc (DRD) is virtually the same as that of ggac(UUUU)gtcc [2.12], ggac(uucg)gtcc [2.13] and ggac(tttt)gtcc [2.11], and thus we conclude that this particular hairpin (DRD) is not extra stable.

Interestingly, the loop of DRD displays sequence dependence in a manner that is similar to that observed for RRR. Changing the loop from ggac(UUCG)gtcc to ggac(UUUG)gtcc [2.17] does not affect hairpin stability (Table 2.2). This suggests that the cytosine loop residue in DRD is not involved in any stabilizing interaction, a trend

similar to that seen in the loop of RRR. Moreover, mutation of guanine G₈ to a uracil (*i.e.*, DR³D; [2.18]) destabilized the hairpin by *ca.* 1.0 kcal/mol which may be ascribed, at this moment, to the loss of wobble base pairing and/or base stacking. Similarly, hairpin ggac(UUCG)gtcc is destabilized upon changing the second loop residue to ggac(UACG)gtcc [2.16] by *ca.* 0.7 kcal/mol and parallels the trend seen previously for RRR *versus* RR¹R (Table 2.1). These observations evidently show that U₆ and G₈ residues participate in the loop stabilizing interactions while C₇ does not, and alludes to the possibility of a conserved global 2',5'-RNA loop structure in DNA and RNA stem hairpins. NMR studies described in section 2.14 shed light into the overall conformation and specific tertiary interactions of the 2',5'-loop.

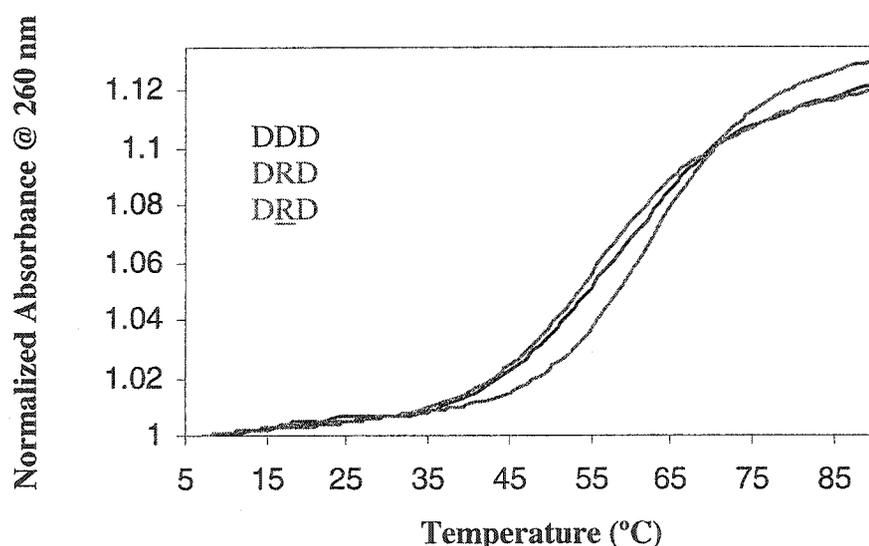


Figure 2.12: Normalized melting temperature profiles of DNA hairpin with various loop compositions. Measurements were done in 0.01 M Na₂HPO₄, 0.1 mM Na₂EDTA buffer, pH 7.00 ± 0.02 at an oligonucleotide concentration ~ 4.5 μM.

2.8 R AND R LOOPS IN HAIRPINS WITH DD AND RR STEMS

It is worth noting that while the 3',5'-RNA loop behaves like the DNA loop in the hairpin with DD stem, the 2',5'-RNA loop imparts remarkable stabilization (DRD and DDD *versus* DRD, Table 2.2). On the other hand, both UUCG and UUCG show the same degree of stability in the RR stem hairpin (RRR *versus* RRR). This suggests that the overall hairpin stability is determined by an interplay of both stem and loop structures as

well as the degree of interdependence exhibited by the nucleobases upon stacking in these different contexts. The stem has an important role in accommodating the loop structure and determining the influence of the loop on the overall hairpin stability. Likewise, the structure of the loop stabilizes the loop-closing base pairs next to it leading to an overall stabilization in the stem and the hairpin structure.

Table 2.2: Thermodynamic parameters of hairpins with DNA stems

Designation	Code	5'- Hairpin -2'/3'	T_m (°C)	% H	ΔH° (kcal/mol)	ΔS° (e.u.)	ΔG°_{37} (kcal/mol)
<u>2.11</u>	DTD	ggac(tttt)gtcc	54.7	9.6	-37.9	-115.7	-2.0
<u>2.12</u>	DUD	ggac(UUUU)gtcc	52.8	7.1	-38.7	-118.8	-1.9
<u>2.13</u>	DDD	ggac(uucg)gtcc	56.2	11.3	-36.6	-111.1	-2.1
<u>2.14</u>	DRD	ggac(UUCG)gtcc	54.6	11.5	-36.0	-109.8	-1.9
<u>2.15</u>	<u>DRD</u>	ggac(<u>UUCG</u>)gtcc	61.4	12.6	-39.9	-119.4	-2.9
<u>2.16</u>	<u>DR</u> ¹ D	ggac(<u>UACG</u>)gtcc	56.7	10.7	-37.4	-113.5	-2.2
<u>2.17</u>	<u>DR</u> ² D	ggac(<u>UUUG</u>)gtcc	62.0	9.8	-42.6	-127.0	-3.2
<u>2.18</u>	<u>DR</u> ³ D	ggac(<u>UUUU</u>)gtcc	54.5	9.7	-40.7	-124.2	-2.2

Measurements were made in 0.01 M Na₂HPO₄, 0.1 mM Na₂EDTA buffer, pH 7.00 ± 0.02; oligonucleotide concentration ~ 4.5 μM. Values represent the average of at least five independent measurements. Errors in thermodynamic parameters are within ± 7.5% for ΔH° and ΔS° , and ± 0.20 kcal/mol for ΔG°_{37} . Error in T_m is within ± 1 °C. For a more accurate calculation, ΔG°_{37} was determined from ΔH° and ΔS° before rounding off. Capital letters represent RNA residues; underlined letters are 2',5'-RNA residues (*e.g.* UC = U_{2'p5'}C_{2'p}); DNA residues are represented by small letters; bold letters are base mutations in the loop. Percentage hypochromicity (%H) was calculated from UV absorbances of the hairpin (A_0) and fully denatured species (A_f) using the following equation: %H = ($A_f - A_0$) / A_f .

2.9 HAIRPINS WITH VARIOUS STEMS: R *VERSUS* R LOOPS

Hairpins with the stems DR, RD, RR, RR and containing UUCG and UUCG tetraloops were also synthesized (Table 2.3). In this study, hairpin TRT [2.31] served as negative control that does not fold into a hairpin species. Interestingly, the degree of stabilization provided by 2',5'-RNA tetraloops is strongly dependent on the sugar composition of the helical stem. For example, stabilization by 2',5'-RNA tetraloops is largest for the least stable stems, *i.e.* RR and RD, but almost negligible for the most stable stems (RR and RD). This is evident from Figure 2.13, which compares the relative Gibbs free energy of association of hairpins ($\Delta\Delta G_{37}^{\circ}$). Hence, dashed lines with a negative slope represent cases where stabilization of a 2',5'-tetraloop is greater relative to a 3',5'-linked tetraloop. It is clear that UUCG stabilizes a hairpin with either RD or RR stems better than UUCG. For example, replacing the 3',5'-internucleotide linkages in RRR with 2',5'-linkages (to obtain RRR) leads to a gain of 10 °C in melting temperature, or a $\Delta\Delta G_{37}^{\circ}$ of -1.5 kcal/mol in free energy (Table 2.3). Another example is the case of RRR [2.28] which is more thermodynamically stable than RRR ($\Delta\Delta G_{37}^{\circ} = -0.7$ kcal/mol).

Substitution of 2',5'-RNA in the loop regions of various hairpins does not seem to have any significant effect on the conformational freedom change (ΔS°) associated with hairpin formation (Table 2.3), suggesting that the thermodynamic stabilization brought about by 2',5'-RNA loops is mainly a consequence of a more favorable enthalpy change (ΔH°) of association. This is probably reflective of better stacking and pairing interactions in hairpins containing 2',5'-RNA loops compared to those with native 3',5'-loops.

2.10 FORMATION OF 2',5'-RNA:DNA HYBRIDS

We have observed for the first time the formation of a stable 2',5'-RNA:DNA hybrid by linking the hybrid's strands to the UUCG loop (RRD [2.22]; Table 2.3). RD duplexes, which are not stable enough to form in an intermolecular complex,¹⁴¹ were stable at room temperature ($T_m = 52.8$ °C; $\Delta G_{37}^{\circ} = -1.9$ kcal/mol). The CD spectrum shows their structure is a mixture of both the pure RNA and DNA forms,⁴ with a greater similarity to the B-type family (Figure 2.14 E). The 2',5'-RNA loop shows superior ability in stabilizing RD hybrids relative to the native 3',5'-RNA loop.

Table 2.3: Thermodynamic parameters of hairpins with DNA:RNA, DNA:2',5'-RNA, RNA:2',5'-RNA, and 2',5'-RNA:2',5'-RNA stem hybrids

Designation	Code	5'- Hairpin -2'/3'	T_m (°C)	% H	ΔH° (kcal/mol)	ΔS° (e.u.)	ΔG°_{37} (kcal/mol)
<u>2.19</u>	DRR	ggac(UUCG)GUCC	56.5	7.4	-41.1	-124.7	-2.4
<u>2.20</u>	DRR	ggac(UUCG)GUCC	56.7	8.0	-41.0	-124.4	-2.5
<u>2.21</u>	RRD	<u>GGAC</u> (UUCG)gtcc	48.1	3.1	-38.9	-121.1	-1.4
<u>2.22</u>	RRD	<u>GGAC</u> (UUCG)gtcc	52.8	4.6	-39.5	-121.3	-1.9
<u>2.23</u>	DRR	ggac(UUCG)GUCC	24.1	5.8	-28.7	-96.6	+1.2 (-0.2) ^a
<u>2.24</u>	DRR	ggac(UUCG)GUCC	30.2	7.0	30.7	-101.1	+0.6 (-0.8) ^a
<u>2.25</u>	RRR	<u>GGAC</u> (UUCG)GUCC	62.6	5.3	-44.2	-131.8	-3.4
<u>2.26</u>	RRR	<u>GGAC</u> (UUCG)GUCC	62.4	4.0	-44.7	-133.2	-3.4
<u>2.27</u>	RRR	<u>GGAC</u> (UUCG)GUCC	54.1	7.1	-42.5	-122.2	-2.2
<u>2.28</u>	RRR	<u>GGAC</u> (UUCG)GUCC	58.1	7.4	-45.5	-137.2	-2.9
<u>2.29</u>	RRR	<u>GGAC</u> (UUCG)GUCC	45.2 ^b	3.9	-30.8	-98.2	-0.8
<u>2.30</u>	RRR	<u>GGAC</u> (UUCG)GUCC	54.8	2.2	-42.1	-128.5	-2.3
<u>2.31</u>	TRT	tttt(UUCG)tttt	- ^c	-	-	-	-

Measurements were made in 0.01 M Na₂HPO₄, 0.1 mM Na₂EDTA buffer, pH 7.00 ± 0.02; oligonucleotide concentration ~ 4.5 μM. Values represent the average of at least five independent measurements. Error in T_m is within ± 1 °C. Errors in thermodynamic parameters are within ± 7.5% for ΔH° and ΔS° , and ± 0.20 kcal/mol for ΔG°_{37} . For a more accurate calculation, ΔG°_{37} was determined from ΔH° and ΔS° before rounding off. Capital letters represent RNA residues; underlined letters are 2',5'-RNA residues (*e.g.* UC = U_{2p5'}C_{2p}); DNA residues are represented by small letters; bold letters are base mutations in the loop. Percentage hypochromicity (%H) was calculated from UV absorbances of the hairpin (A_0) and fully denatured species (A_f) using the following equation: %H = $(A_f - A_0) / A_f$.

^a Values in parenthesis represent ΔG° calculated at 22 °C. ^b Very broad transition. ^c No melting transition observed.

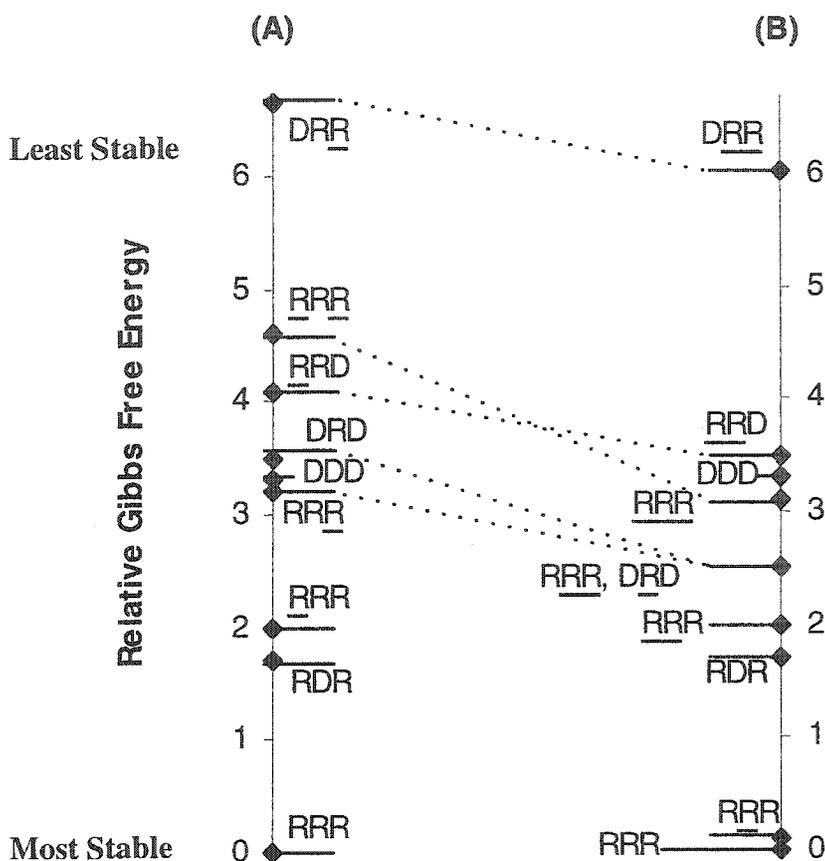


Figure 2.13: Relative Gibbs free energies of formation ($\Delta\Delta G_{37}^{\circ}$) in kcal/mol for hairpins with different stem combinations but with (A) 3',5'-RNA loops and with (B) 2',5'-RNA loops. The control hairpin sequences (DDD, RDR and RRR) are also shown for comparison purposes. The dashed lines represent the drop in Gibbs free energy upon substitution of 2',5'-RNA in the loop for 3',5'-RNA. All values are referenced relative to the RRR hairpin exhibiting the lowest (most negative) free energy value.

2.11 GLOBAL HELICAL CONFORMATION OF HAIRPINS

The CD spectra show that incorporation of 2',5'-RNA in the loop region of hairpins with various stems does not significantly perturb the overall structure. However, there is a notable increase in the negative CD band at *ca.* 212 nm, and the long wavelength CD peak becomes a mixture of both pure RNA and pure DNA peaks (**Figure 2.14**). The increase in the depth of the CD peak at *ca.* 212 nm is typical of A-like duplexes. DRD [2.15] shows a B-form global conformation (**Figure 2.14 A**) but with an increase in the short wavelength negative CD peak compared to DDD [2.13] and DRD [2.14] and it shows a positive CD band at *ca.* 279 nm that is centered between that of DRD (*ca.* 277 nm) and DDD (*ca.* 282 nm). In a similar fashion, the CD spectrum of RRR

[2.2] shows a similarity to the classic A-form global conformation, but exhibits a substantial increase in the negative CD peak at *ca.* 213 nm relative to that of the wild-type RRR hairpin and its 262 nm CD peak is a mixture of both pure A- and B- form bands with more A-character (Figure 2.14 B). Even the slightest insertions would perturb the overall hairpin conformation. For instance, R_cRR [2.8] and RR_cR [2.9] display a major perturbation in their long wavelength CD peak relative to RRR although they still maintain an overall A-like structure (Figure 2.14 H). RRR [2.28] exhibits an enhanced CD peak at *ca.* 212 nm relative to RRR [2.27] and both are mixtures with similarity to the A-form type (Figure 2.14 D). In a similar fashion, both RRR [2.25] and RRR [2.26] display an intermediate conformation (“mixture”) between the A- and B-type (Figure 2.14 C).

2.12 EVIDENCE FOR THE EXISTENCE OF AN ALL 2',5'-RNA HAIRPIN AND COMPARISON TO ALL-DNA AND ALL-RNA HAIRPINS

The thermodynamic stability of RRR [2.30; Table 2.3] falls within the same order of magnitude as that of DDD [2.13], with RRR [2.1] being the most stable (RRR > RRR ~ DDD). The all 2',5'-RNA hairpin exhibits an entropy of formation that is intermediate between that of RRR and DDD. The hypochromicity data suggest that the DNA hairpin exhibits the best stacking interactions followed by the all RNA hairpin and last by the all 2',5'-RNA hairpin.

The global conformations of RRR and DDD exhibit characteristic A-form and B-form features⁴ respectively. The fully modified 2',5'-RNA hairpin [RRR] does not fall into either a pure A-form or B-form pattern. The negative CD peak at *ca.* 216 nm resembles that of pure RNA, while the long wavelength positive CD band is a mixture of both DNA and RNA peaks with more B-like character (Figure 2.15 B). Based on the above, we conclude that the fully modified regioisomeric 2',5'-RNA hairpin does not behave like the 3',5'-RNA counterpart.

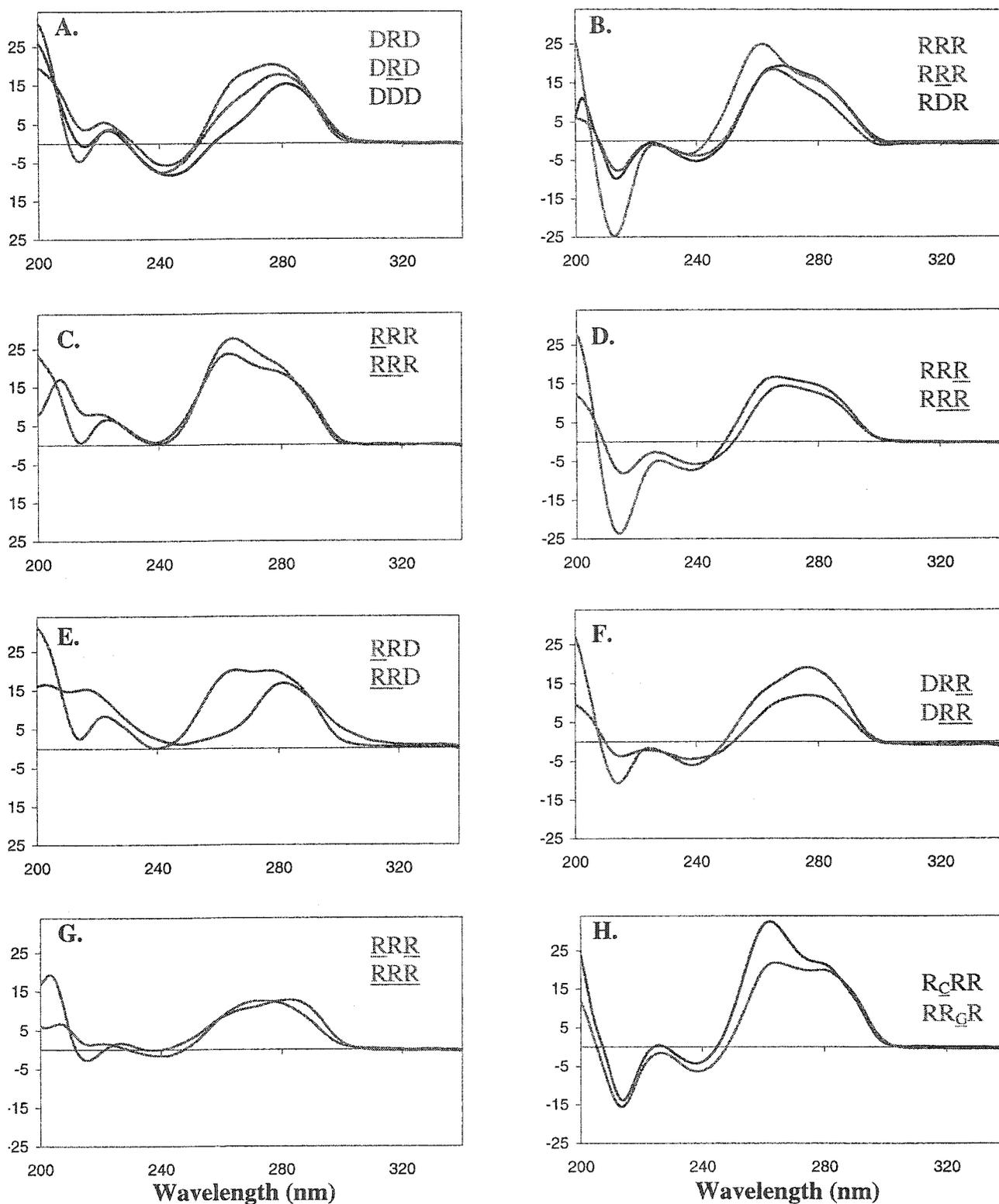


Figure 2.14: Circular dichroism spectra at 22 °C of DNA and RNA hairpins of various loop compositions: (1) 2',5'-r(UpUpCpGp) loops; (2) 3',5'-r(UpUpCpGp) loops; and (3) d(TTCG) loops. The CD spectra of DRR and DRR were recorded at 5 °C. All measurements were done in 0.01 M Na₂HPO₄, 0.1 mM Na₂EDTA buffer, at pH 7.00 ± 0.02. Molar ellipticities were normalized to strand concentration.

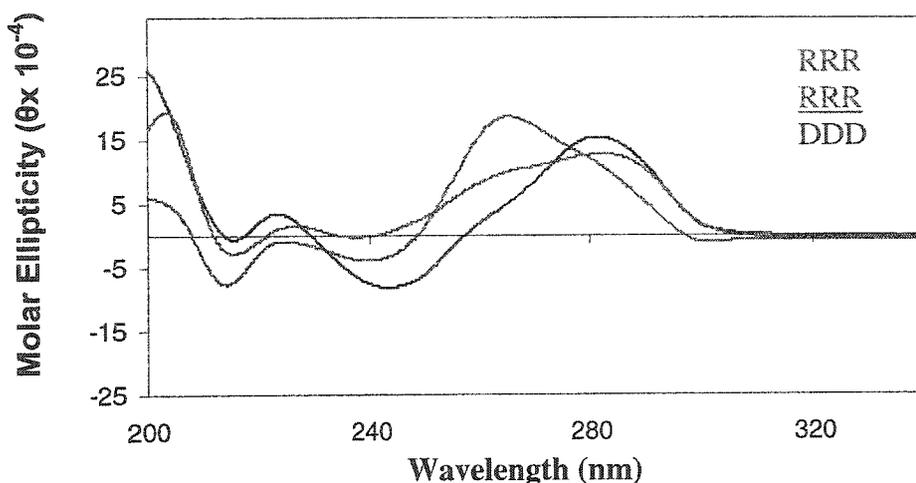


Figure 2.15: CD spectra of DDD, RRR and RRR at 22 °C.

2.13 EFFECT OF C·G LOOP-CLOSING BASE PAIR ON THERMODYNAMIC STABILITY OF HAIRPINS WITH 3',5'- AND 2',5'-RNA (UUCG) LOOPS

2.13.1 Background and General Considerations

The first base pair in the stem located next to the tetraloop moiety is referred to as the loop-closing base pair (for an example, see **Figure 2.2**). Its neighboring existence to both loop and stem nucleotides makes it critical in its contribution to helical base stacking and thus overall thermodynamic hairpin stability. Nature has carefully chosen the identity of the loop closing base pairs present in RNA structures. For example, C·G, G·C or G·U base pairs are the most highly conserved and occur very frequently in ribosomal RNAs.⁴² In a study conducted by Turner and coworkers,¹⁷⁰ they have shown that these base pairs are among the most stable, and thus this might explain their frequent occurrence. Earlier in this chapter (**Sections 2.6 & 2.7**), we showed that both GGAC(UUCG)GUCC [RRR] and ggac(UUCG)gtcc [DRD] were 'extra-stable', and while GGAC(UUCG)GUCC [RRR] is extra-stable,⁶⁸ the hairpin ggac(UUCG)gtcc [DRD] is not. Based on these observations, we reason here that switching of ribose to deoxyribose at the closing base-pair residues, *i.e.*, ...C(UUCG)G... to ...c(UUCG)g... , might in part be responsible for the lack of extra-stability observed in the DRD system. To test this hypothesis and to investigate the role of the sugar residue, we synthesized and studied two new sets of hairpins. The first (**Table 2.4**) was designed to assess the effect of sugar substitutions in hairpins having

'ribonucleotide' character. For instance, hairpins RRR [2.1] and R_cR_gR [GGAc(UUCG)gUCC] [2.32] share the same stem and loop sequence, but differ only in the sugar composition at the C-G closing base pair. The second set (Table 2.5) was designed to probe the effects of deoxyribose to ribose substitution in 'DNA-like' hairpin structures. To our knowledge, this represents the first study assessing the effect of sugar substitution at the closing base pair.

Table 2.4: Thermodynamic parameters for RNA-stem hairpins with deoxy insertions in the loop-closing base pair

Designation	Code	5'- Hairpin -2/3'	T_m (°C)	% H	ΔH° (kcal/mol)	ΔS° (e.u.)	ΔG°_{37} (kcal/mol)
<u>2.1</u>	RRR	GGAC(UUCG)GUCC	71.8	8.5	-53.4	-154.8	-5.41
<u>2.32</u>	R _c R _g R	GGAc(UUCG)gUCC	60.0	6.6	-51.9	-155.5	-3.60
<u>2.33</u>	R _c RR	GGAc(UUCG)GUCC	67.3	9.1	-56.1	-164.0	-5.02
<u>2.34</u>	RR _g R	GGAC(UUCG)gUCC	66.6	8.8	-56.7	-166.6	-4.98
<u>2.2</u>	RRR	GGAC(UUCG)GUCC	69.3	9.6	-55.6	-162.1	-5.33
<u>2.35</u>	R _c R _g R	GGAc(UUCG)gUCC	57.6	6.3	-47.7	-143.9	-3.06

Measurements were made in 0.01 M Na₂HPO₄, and 0.1 mM Na₂EDTA, pH 7.0; oligonucleotide concentration ~ 4.5 μM. Values represent the average of at least five independent measurements. Error in T_m is within ± 1 °C. Errors in thermodynamic parameters are within ± 7.5% for ΔH° and ΔS° , and ± 0.20 kcal/mol for ΔG°_{37} . For a more accurate calculation, ΔG°_{37} was calculated from ΔH° and ΔS° before rounding off and extra significant figures are given in the values of ΔG°_{37} . Capital letters represent RNA residues; underlined letters are 2',5'-RNA residues (*e.g.* UC = U_{2'p5'}C_{2'p}); bold letters are sugar mutations in the loop-closing base pair. Percentage hypochromicity (%H) was calculated from UV absorbances of the hairpin (A_0) and fully denatured species (A_f) using the following equation: %H = ($A_f - A_0$) / A_f .

Table 2.5: Thermodynamic parameters for DNA-stem hairpins with ribose insertions in the loop-closing base pair

Designation	Code	5'- Hairpin -2'/3'	T_m (°C)	% H	ΔH° (kcal/mol)	ΔS° (e.u.)	ΔG°_{37} (kcal/mol)
<u>2.14</u>	DRD	ggac(UUCG)gtcc	54.6	11.5	-36.0	-109.8	-1.9
<u>2.36</u>	D _c R _G D	ggaC(UUCG)Gtcc	59.5	7.5	-45.4	-136.6	-3.08
<u>2.12</u>	DUD	ggac(UUUU)gtcc	52.8	7.1	-38.7	-118.8	-1.9
<u>2.37</u>	D _c U _G D	ggaC(UUUU)Gtcc	51.6	9.8	-38.6	-118.9	-1.74
<u>2.15</u>	<u>DRD</u>	ggac(<u>UUCG</u>)gtcc	61.4	12.6	-39.9	-119.4	-1.83
<u>2.38</u>	D _c <u>R</u> _G D	ggaC(UUCG)Gtcc	52.3	7.0	-39.0	-120.0	-1.83
<u>2.39</u>	D _c <u>R</u> _G D	gga <u>C</u> (UUCG)Gtcc	57.0	7.0	-41.1	-124.5	-2.50

Measurements were made in 0.01 M Na₂HPO₄, and 0.1 mM Na₂EDTA, pH 7.0; oligonucleotide concentration ~ 4.5 μM. Values represent the average of at least five independent measurements. Error in T_m is within ± 1 °C. Errors in thermodynamic parameters are within ± 7.5% for ΔH° and ΔS° , and ± 0.20 kcal/mol for ΔG°_{37} . For a more accurate calculation, ΔG°_{37} was calculated from ΔH° and ΔS° before rounding off and extra significant figures are given in the values of ΔG°_{37} . Capital letters represent RNA residues; underlined letters are 2',5'-RNA residues (*e.g.* UC = U_{2p5'}C_{2p}); bold letters are sugar mutations in the loop-closing base pair. Percentage hypochromicity (%H) was calculated from UV absorbances of the hairpin (A_0) and fully denatured species (A_f) using the following equation: %H = ($A_f - A_0$) / A_f .

2.13.2 Effects of Substitution of Deoxyribose for Ribose in the Loop-Closing Base Pair of RNA Hairpins

The wild-type RNA hairpin [*i.e.*, RRR; 2.1] exhibits higher thermodynamic stability compared to R_cR_GR, *i.e.*, all-RNA hairpin [2.32; Table 2.4] containing the deoxyribonucleotides at the loop-closing positions [...dC(UUCG)dG...]. In fact, the incorporation of deoxyriboses at the closing base pair results in a substantial drop in melting temperature (ΔT_m) of 11.8 °C (Figure 2.16), with a corresponding increase in ΔG°_{37} of 1.81 kcal/mol. A similar destabilizing effect was observed when the loop-closing C·G pair in RRR was replaced by dC·dG (compare “RRR” with “R_cR_GR”, Table 2.4 & Figure 2.16). In this case, $\Delta T_m = 11.7^\circ\text{C}$; $\Delta\Delta G^\circ_{37} = -2.27$ kcal/mol. This suggests

that ribose \rightarrow deoxyribose substitutions in the C·G loop-closing base pair destabilize RNA hairpins containing either 3',5'- or 2',5'-RNA loops. The thermodynamic data presented in **Table 2.4** suggest that the destabilization is mainly enthalpic in origin. For the case of the well-studied wild-type RRR hairpin, it was interesting to assess the destabilizing contribution of each deoxyribose in the C·G base pair. Towards this end, we designed the hairpins R_cRR [2.33] and RR_gR [2.34] with ...dC(UUCG)G... and ...C(UUCG)dG... sequences, respectively. **Table 2.4** shows that each substitution, whether it be at C or G, contributes equally to the loss in hairpin thermodynamic stability ($\Delta T_m \sim 4\text{-}5$ °C drop/deoxy insertion). The effects are nearly additive, suggesting that *both* riboses of the loop-closing base pair contribute significantly to the stability of this particular RNA hairpin sequence. Consistent with this notion is the observation that each single deoxy substitution in the C·G base pair does not significantly affect the hypochromicity (%H) of the hairpin. However, replacing *both* riboses with deoxyriboses causes a perturbation in the degree of base stacking interactions within hairpins comprised of 3',5'- and 2',5'-loops as reflected in the drop in %H values (compare “RRR” and “ R_cR_gR ”; “RRR” and “ R_cR_gR ”; **Table 2.4**).

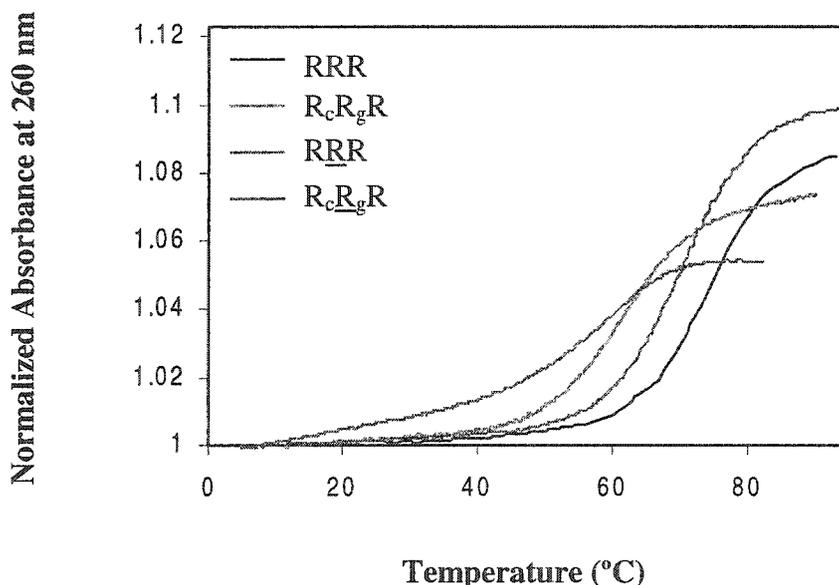


Figure 2.16: Thermal melting curves of RNA hairpin duplexes containing deoxynucleotide sugar residues at the C·G loop-closing base pair (~ 4.5 μM). Buffer: 0.01 M Na_2HPO_4 , 0.1 mM Na_2EDTA buffer, pH 7.0. See **Table 2.4** for base sequences.

CD spectral analysis provides evidence that incorporation of deoxyribose at the loop-closing positions has a significant influence on the conformation of the hairpin structure. **Figure 2.17** compares the CD spectra of RRR and \underline{RRR} to those of hairpins with deoxy substitutions. The CD spectra of the RRR and \underline{RRR} hairpin duplexes are characterized by a strong positive Cotton effect at *ca.* 265 nm and a negative CD band at 213 nm. These features are characteristic of A-form helices.¹ After incorporation of two deoxyriboses at the loop-closing positions, the spectral features change significantly. Note the large decrease in the intensity of the negative CD band, and the shift of the positive CD band towards longer wavelengths. The two deoxyribonucleotide residues incorporated within RRR and \underline{RRR} hairpins seem to induce a conformational change of the hairpin structure in the direction of the 'DNA-like' hairpins. For example, note the resemblance of the R_cR_gR and D_cR_gD spectra (**Figures 2.17** and **2.19**).

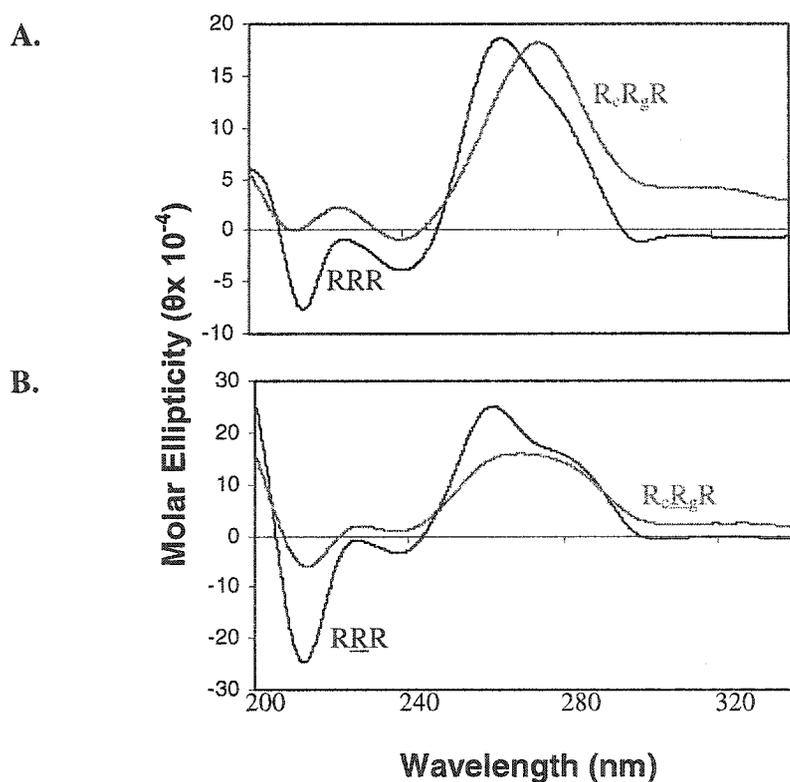


Figure 2.17: Circular dichroic spectra of RNA hairpins differing in the sugar composition of the C-G loop-closing base pair. See **Table 2.4** for base sequence. Measurements were carried at 22 °C in 0.01 M Na_2HPO_4 , 0.1 mM Na_2EDTA buffer (pH 7.0). Molar ellipticities were normalized to strand concentration.

In summary, our results indicate that nearest neighbor interactions between the loop-closing base pair and the loop nucleotides, and the loop-closing base pair and stem structure contribute to the overall hairpin thermodynamic stability.

2.13.3 Effects of Substitution of Ribose for Deoxyribose in the Loop-Closing Base Pair of Hairpins with DNA Stems

Another key observation in this study is that ggaC(UUCG)Gtcc (or “D_CR_GD”) is significantly more stable (**Figure 2.18**) than the corresponding DRD hairpin containing only 4 ribose residues ($\Delta T_m = +5$ °C; $\Delta\Delta G_{37}^\circ = -1.2$ kcal/mol; **Table 2.5**). This shows that the sugars of *both* the loop and closing base pair (‘hexaloop’) play a fundamental influence on the stability of extra-stable RNA hairpin structures. This suggests that the sugar pucker (and thus base orientation) of the closing base pair has to be in uniformity with the overall stacking of the loop residues, thus dictating and organizing both the loop and stem helical structures. The ‘extra-stability’ of D_CR_GD is also sequence specific since mutation of the hexaloop to ...C(UUUU)G... [**2.37**] abolishes the stabilizing effect (see “D_CU_GD”, **Table 2.5** & **Figure 2.18**). In fact, D_CU_GD does not show increased thermodynamic stability relative to DUD [**2.12**], DTD [**2.11**], and DDD [**2.13**]. This supports what has been previously observed by Tinoco and co-workers where the contribution of non-extra stable loops to hairpin stability is the same whether these loops are comprised of DNA or RNA units.⁶⁸

2',5'-Linked ‘tetraloops’ provide ‘extra-stability’ to hairpins with DNA stems. As shown in **Table 2.5**, DRD is much more stable than DRD ($\Delta T_m = +7$ °C), DTD ($\Delta T_m = +7$ °C), or a hairpin with a 2',5'-homopolymeric loop [ggac(UUUU)gtcc] ($\Delta T_m = +8$ °C). Clearly, in this case, UUCG and UUCG respond differently to the same perturbations made within the sugar residues of the loop-closing base pairs. This might have been anticipated since our previous findings based on base mutations suggested that UUCG and UUCG likely adopt different conformations.

The CD profiles of DRD and D_CR_GD, as well as DRD and D_CR_GD (**Figure 2.19 A&B**) indicate that the perturbation of the overall hairpin conformation in this case is not as dramatic as that observed previously for RNA-stem hairpins. For instance, the CD

spectra of DRD and D_CR_GD are similar in shape (Figure 2.19 C), the major difference being the amplitude of the positive peaks.

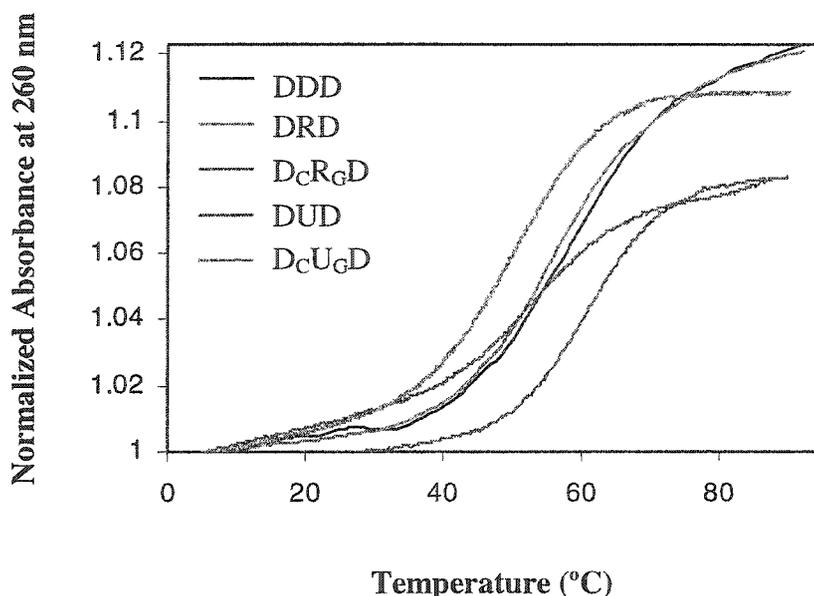


Figure 2.18: Thermal melting curves of DNA hairpin duplexes containing ribose substitutions at the C-G loop-closing base pair ($\sim 4.5 \mu\text{M}$). Buffer: 0.01 M Na₂HPO₄, 0.1 mM Na₂EDTA buffer, pH 7.0. See Table 2.5 for base sequences.

In the end, we designed the hairpin D_CR_GD [2.39] comprised of a 2',5'-RNA pentaloop and having ribose units at the closing base pair position, *i.e.*, 5'-ggaC(UUCG)Gtcc-3'. The introduction of an additional 2',5'-linkage near the 5'-end of the loop increases thermodynamic stability by *ca.* 0.7 kcal/mol compared to 5'-ggaC(UUCG)Gtcc-3' (D_CR_GD, Table 2.5).

Based on the above, we can deduce that the compatibility in overall stacking between the loop residues on one hand and the closing base pair within the context of the stem on the other hand, are critical determinants for imparting extra stability in hairpin structures. Switching the sugar pucker from ribose to deoxyribose at the loop-closing base pair or vice versa likely changes base stacking and thus affects the global hairpin structure. This may explain why the 3',5'-r(UUCG) loop is structurally unique and unusually stable within the context of RNA stems, but not DNA stems. All in all, these results broaden our understanding of the behavior of highly stable (UUCG) hairpin loops

and how they respond to various perturbations. They also add new dimensions and insights in the design of stable loop motifs as means for stabilizing antisense drugs.

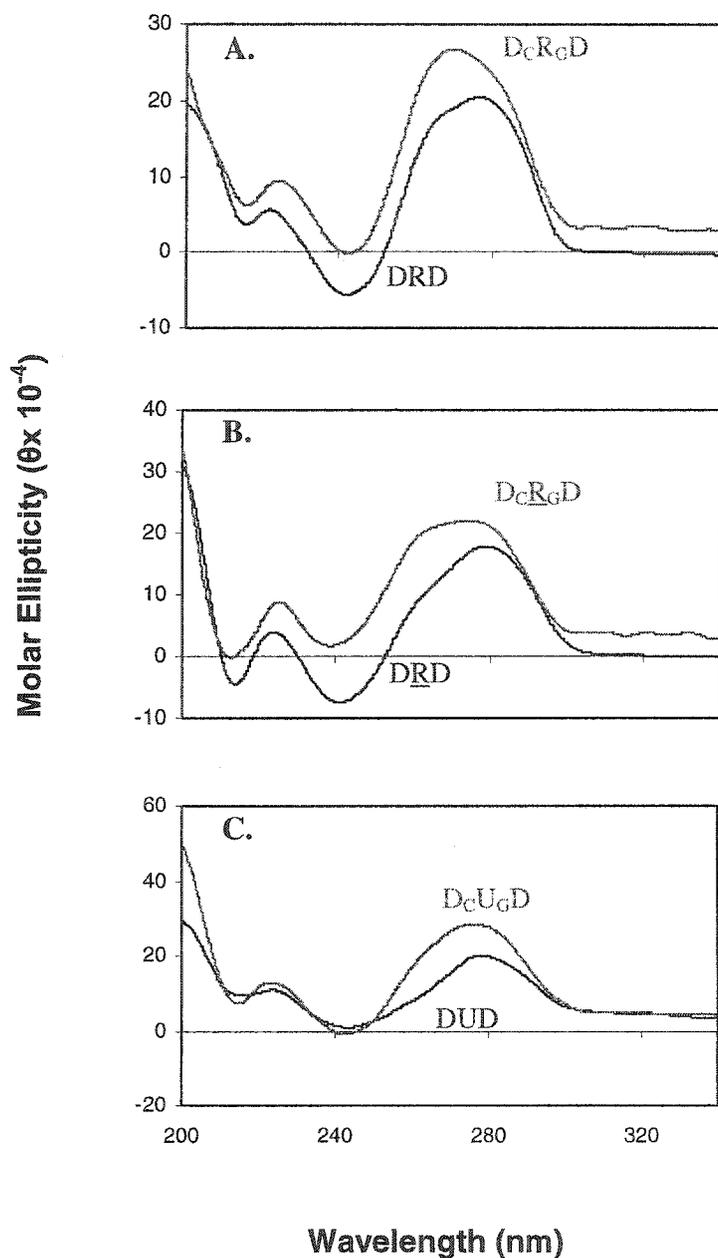


Figure 2.19: Circular dichroic spectra of DNA hairpins differing in the sugar composition of the C-G loop-closing base pair. See Table 2.5 for base sequence. Measurements were carried at 22 °C in 0.01 M Na₂HPO₄, 0.1 mM Na₂EDTA buffer (pH 7.0). Molar ellipticities were normalized to strand concentration.

2.14 NMR SOLUTION STRUCTURE OF 2',5'-RNA LOOPS

In order to provide structure-based reasoning for the exceptional stability of hairpins with 2',5'-linked loops and elucidate their general features, we studied the three dimensional structure of the 2',5'-(UUCG) RNA loop within the context of RNA and DNA stem hairpins by using high-resolution NMR. Towards this end, the following three hairpin sequences were considered: (1) hairpin DRD, (2) hairpin RRR, and (3) hairpin DRD, which was designed as a control in order to compare with DRD and to study the conformation of the native 3',5'-loop structure within the context of DNA stems.

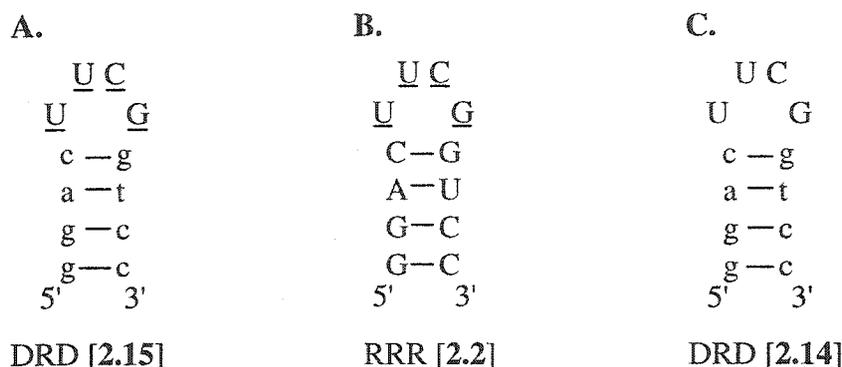


Figure 2.20: Schematic representation of the three hairpin sequences under study. Capital letters represent RNA residues. Small letters represent DNA residues. Capital underlined letters represent 2',5'-RNA residues.

We synthesized these hairpins on a *large* scale (4 μ mol) to obtain quantities sufficient for conducting NMR experiments. The hairpins were purified by preparative ion-exchange HPLC and then desalted on a reversed-phase column (Sep Pak). The overall isolated yields as well as % purity are given in experimental **Section 7.5**.

Earlier in this chapter, melting studies indicated that the above hairpins form unimolecular hypochromic structures with distinct thermodynamic properties. RRR is of the same order of stability as the native RRR hairpin. Remarkably, DRD is thermodynamically more stable than DRD hairpin. Not only that but hairpin DRD is not unusually stable and its stability compares to that of a hairpin with a homopolymeric loop [*i.e.*, DUD], while DRD is distinctly unusually stable. We wished to understand the basis behind such unusual stability and to learn about the folding pattern of 2',5'-RNA loops.

In collaboration with Prof. Kalle Gehring and Dr. Alexei Denisov [Biochemistry Department, McGill University], the structures of the various hairpins were determined from interproton distances and scalar couplings derived from NMR (see experimental **Section 7.5**). The assignment of hairpin non-exchangeable proton resonances in NOESY spectra (acquired at various mixing times) was carried out in the standard manner employed for right-handed DNA duplexes (**Figure 2.21**).³⁰ The assignment of stem imino-protons was made from NOESY spectra in H₂O/D₂O using the crosspeaks of complementary base pairs which include NH(G) with NH₂(C) or NH(T,U10) with H2(A3) (**Figure 2.21 C**). Signals of NH₂(C) were easily identified from their strong crosspeaks with H5(C) of the same nucleotide residue.

The 2',5'-RNA loops in both RRR and DRD hairpins share similar chemical shifts and NOE spectral parameters (**Figure 2.21**) demonstrating that they adopt a common, unusually unique structure that is distinct from that of the native 3',5'-RNA loop. Remarkably, the NMR-restrained molecular modeling shows that the structure of the 2',5'-RNA loop is conserved in both hairpins with DNA & RNA stem duplexes. It is stabilized by a wobble U5·G8 base pair, with both base residues in *anti* conformation, extensive base stacking as well as sugar–base contacts (**Figure 2.22**). The U5·G8 base pair stacks on top of the C4·G9 loop-closing base pair and thus appears as a continuation of the stem. Strong NOE was detected between G8 and U5 imino protons, and medium-weak NOEs of U5 and G8 imino signals with G9 imino, as well as imino G8 with H1'/H4'/H5'(U6) were evident (**Figures 2.21 C, 2.23 and 2.24**). These confirm that U5 and G8 loop residues form a wobble base pair (**Figure 2.25**). Unlike the native loop structure,^{34,53,54} G8 is in *anti* conformation as exemplified by the NOEs between its imino signals and those of G9. Furthermore, the aromatic to sugar H1' NOE, typical of helical strands, shows that uracil U6 base stacks on top of U5 in the 2',5'-RNA loop. However, despite all structural similarity, the 2',5'-RNA loops in hairpins DRD and RRR still show a minor difference which manifests itself in the G8 residue of hairpin RRR sometimes existing in the *syn*-conformation. Indeed, the intraresidual H8-H1' and H3'-H8 (or H2'-H8) NOE crosspeaks of G8 are equal and of strong intensity at any mixing time at low temperatures (**Figure 2.21 B**), which cannot be realized for any single loop conformation.

This fact indicates that there exists some degree of mobility in G8-residue of the 2',5'-RNA (UUCG) loop in the context of RRR hairpin.

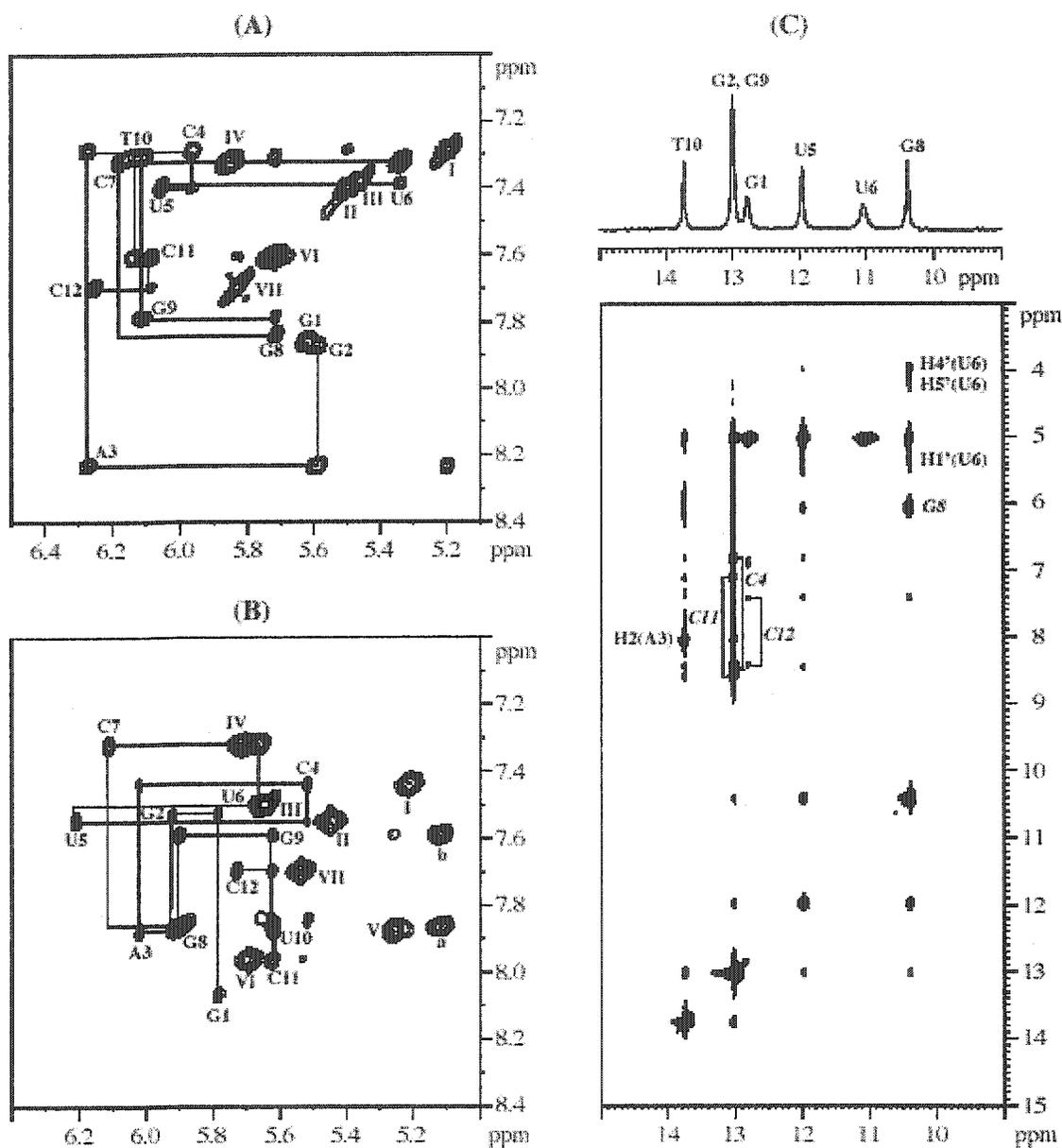


Figure 2.21: Expanded plots of NOESY spectra at 500 MHz. Panels (A) and (B) - hairpins DRD and RRR in D_2O , $15^\circ C$, mixing time $t_m=200$ ms, the assignment of oligonucleotide protons are showed by solid lines and nucleotide name with number; I-VII - H5-H6(C4,U5,U6,C7,U10,C11,C12) crosspeaks respectively; the letter marks a and b - H2'-H8(G8) and H2'(G8)-H8(G9) crosspeaks; panel (C) - hairpin DRD in H_2O/D_2O , $5^\circ C$, with 1D spectrum of imino protons at the top; crosspeaks for amino protons of cytidines and G8 are labeled by marks in italic font.

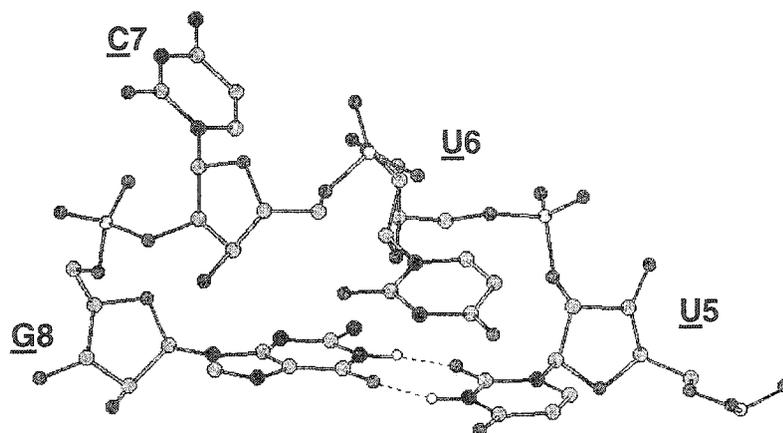


Figure 2.22: Conserved structure of the 2',5'-RNA loop showing U-G wobble base pairing, U6 stacking on U5, and C7 protruding out into the solvent.

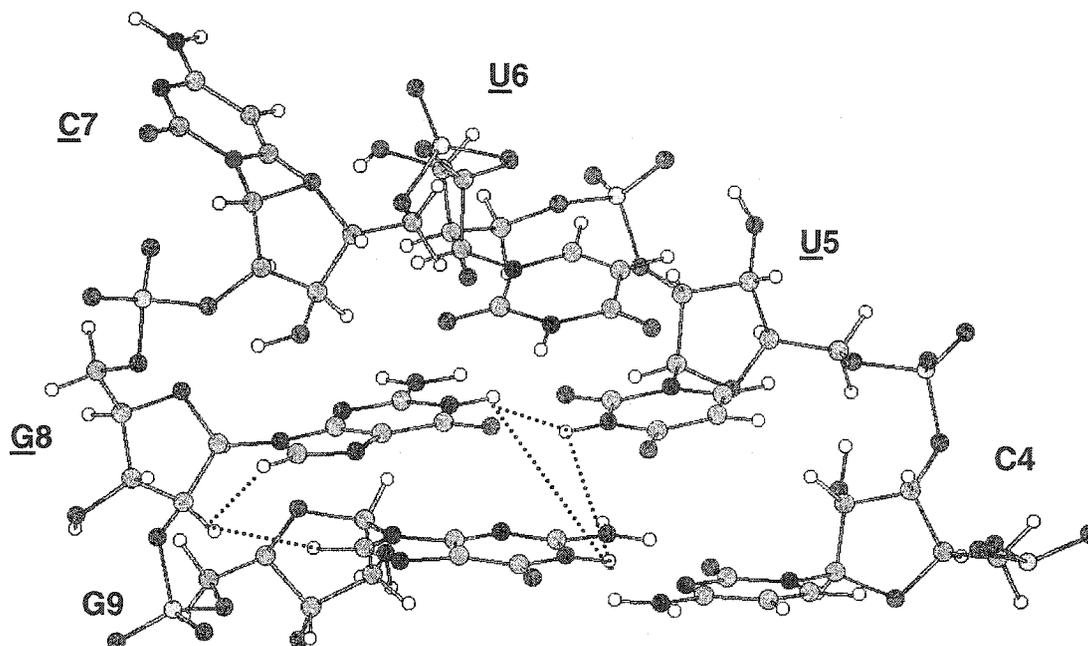


Figure 2.23: Conserved structure of the 2',5'-RNA loop showing unique NOEs between U5, G8 and G9 residues. These confirm that U5 and G8 base pair and that they stack on top of the C4-G9 stem base pair. Dashed lines represent NOE contacts observed in NMR.

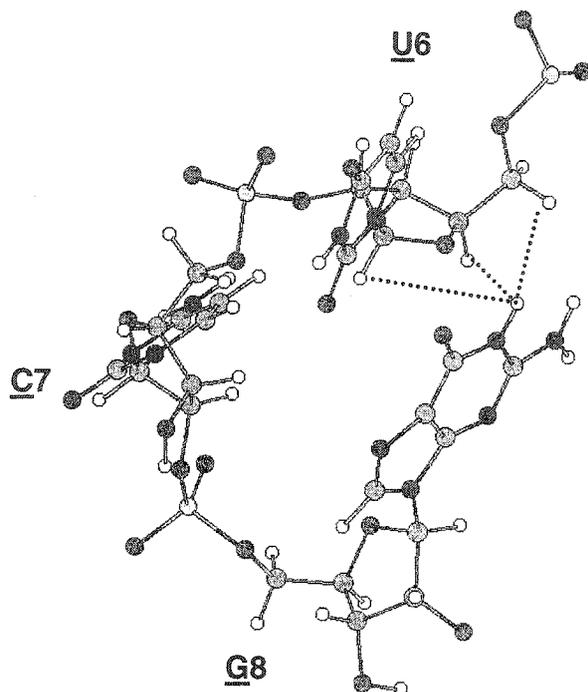


Figure 2.24: Conserved structure of the 2',5'-RNA loop showing NOEs of imino G8 with H1'/H4'/H5'(U6). Dashed lines represent NOE contacts observed in NMR.

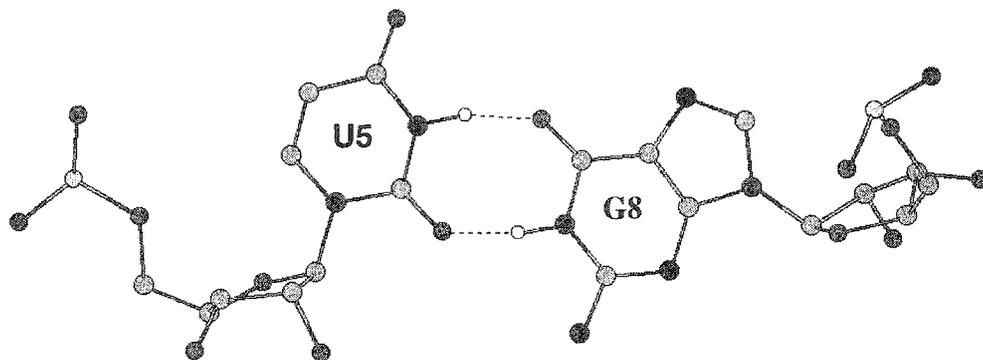


Figure 2.25: U-G Wobble base pairing motif observed in conserved structures of 2',5'-RNA loop. Dashed lines represent hydrogen bonding.

Both hairpins RRR and DRD have two nucleotides in the loop, U6 and C7. Both nucleotides adopt the C-2' *endo* sugar conformation (south). For 2',5'-RNA, this pucker is the so-called “compact” form.^{142,149} This is surprisingly unexpected because as seen in other RNA structures the middle loop nucleotide sugar residues always tend to exist in extended pucker in order to bridge the stem.^{38,171} In this case, and by virtue of the position of the phosphodiester bond (*i.e.*, 2',5'-linked), the C-2' *endo* conformation is not the “extended” form for 2',5'-RNA but rather the “compact” form. However, the 2',5'-RNA loop adjusts in its own unique way by switching the puckers of U5 and G8 to predominantly the “extended” form, *i.e.*, C-3' *endo* (north). Thus, instead of extending the middle two residues in the loop (as is the case for 3',5'-RNA), 2',5'-RNA loop extends the stem arms (U5 and G8 sugars) while keeping the middle two sugar residues (U6 and C7) in the compact form. Other stabilizing interactions are evident within the loop such as strong NOEs from the cytidine (C7) H3' sugar proton to U6 sugar proton H1' and G8 base proton H8. Also, the aromatic proton H6 of C7 shows NOE with U6 sugar proton H1' suggesting that extensive base-base and base-sugar interactions exist within the loop (Figure 2.26).

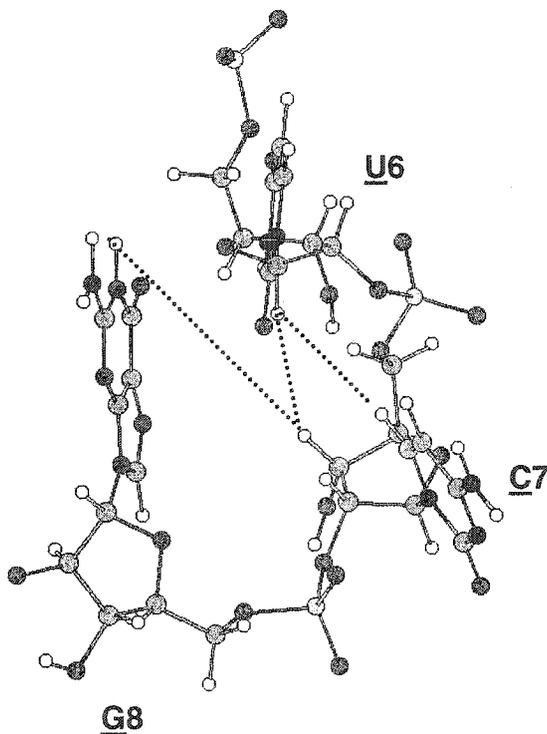


Figure 2.26: Base-base and sugar-base contacts within the conserved 2',5'-RNA loop structure. Dashed lines represent NOE contacts observed in NMR

A stereoview of the energy minimized structures of hairpin RRR and DRD is presented in **Figures 2.27** and **2.28** respectively. Both loop structures are very similar in general, but C7 base is farther from the G8 residue in hairpin RRR. This probably affects G8 and makes it slightly more mobile in hairpin RRR compared to the same residue in hairpin DRD. The minor groove width of the RNA stem is ~ 2 Å bigger in comparison with the DNA stem. As expected, the minor groove width and helical parameters of the RNA stem are closer to A-type DNA, while those of the DNA stem hairpin are closer to B-type DNA.

The proposed conformation for the 2',5'-RNA loop is in strong agreement with the pattern of chemical base modification (described earlier in this chapter) and correlates with the thermal melting data. Changing the loop from C(UUCG)G to C(UACG)G in RRR results in significant loss of thermal stability ($\Delta T_m = 7^\circ\text{C}$; $\Delta\Delta G^\circ_{37} = 1.9$ kcal/mol). This is in agreement with the NMR structure and can be ascribed to the loss of the electrostatic interactions between U6 and other loop residues, all of which contribute significantly to the *unusual* loop folding. However, changing C(UUCG)G ($T_m = 69.3$ °C) to C(UUUG)G ($T_m = 68.0$ °C) did not significantly affect the thermal stability of the hairpin structure. This again is in agreement with our proposed notion that C7 protrudes out into the solvent and hence does not participate in the loop stabilizing interactions. Mutating G8 residue to a uracil [to yield the homopolymeric C(UUUU)G loop, $T_m = 60.5$ °C] destabilized strongly the native 2',5'-RNA loop structure due to loss of U5-G8 wobble base pairing. On the other hand and in parallel with these studies, the same base mutation experiments were conducted within the 2',5'-RNA loop sequence but in the context of the DNA stem, *i.e.*, DRD. The T_m results show a conserved 2',5'-RNA loop motif independent of the identity of hairpin stem, again in total agreement with the NMR data. Changing the loop from ggac(UUCG)gtcc [$T_m = 61.4$ °C] to ggac(UUUG)gtcc [$T_m = 62.0$ °C] did not affect hairpin stability suggesting the same trend [as in RRR] about the folding pattern of the cytosine C7 loop residue in DRD. Mutating guanine G8 to a uracil (*i.e.*, to obtain 5'-ggac(UUUU)gtcc-3') destabilized the hairpin by *ca.* 1.0 kcal/mol which can be ascribed to the loss of wobble base pairing and/or base stacking.

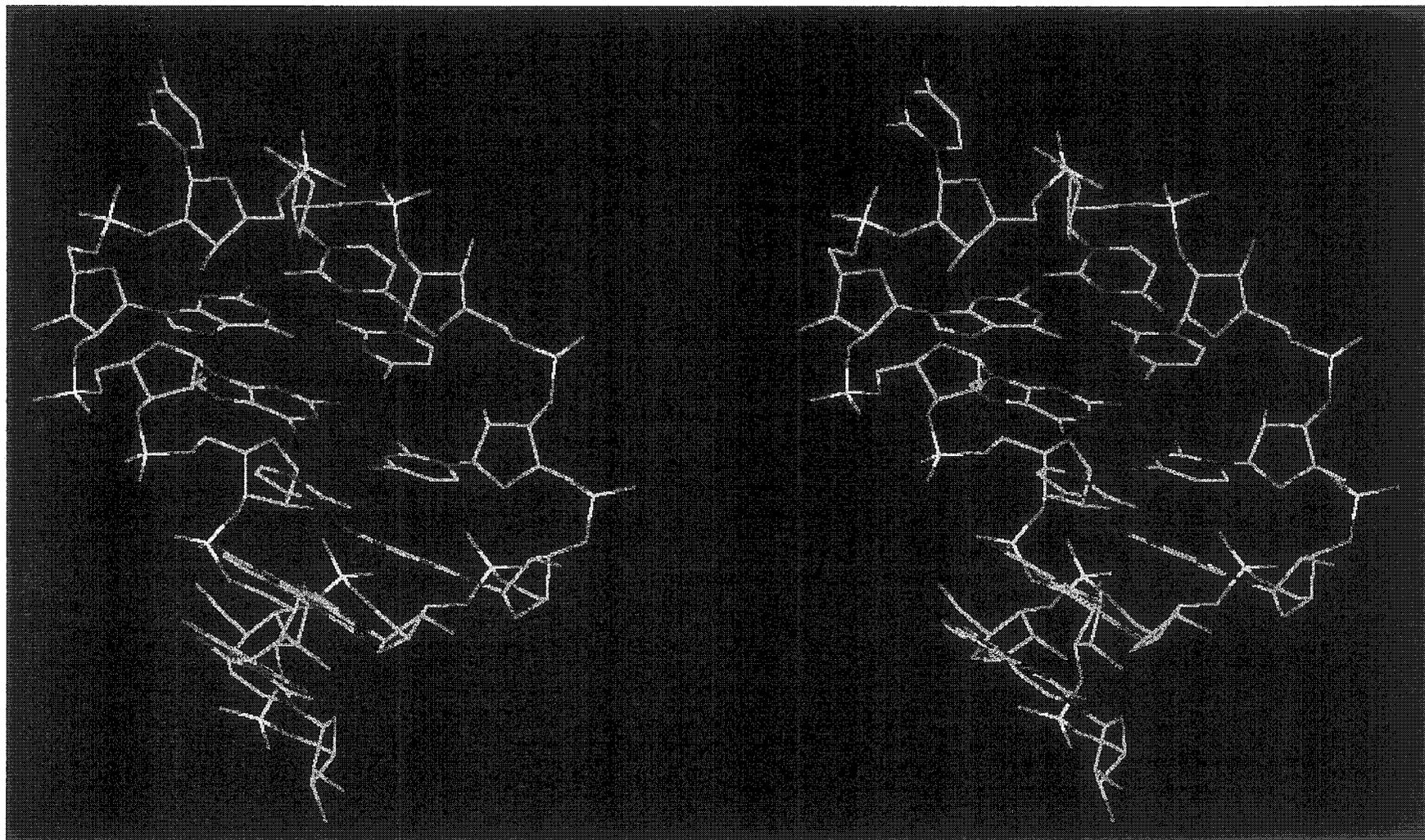


Figure 2.27: Stereoview of the NMR-derived structure of RRR hairpin.

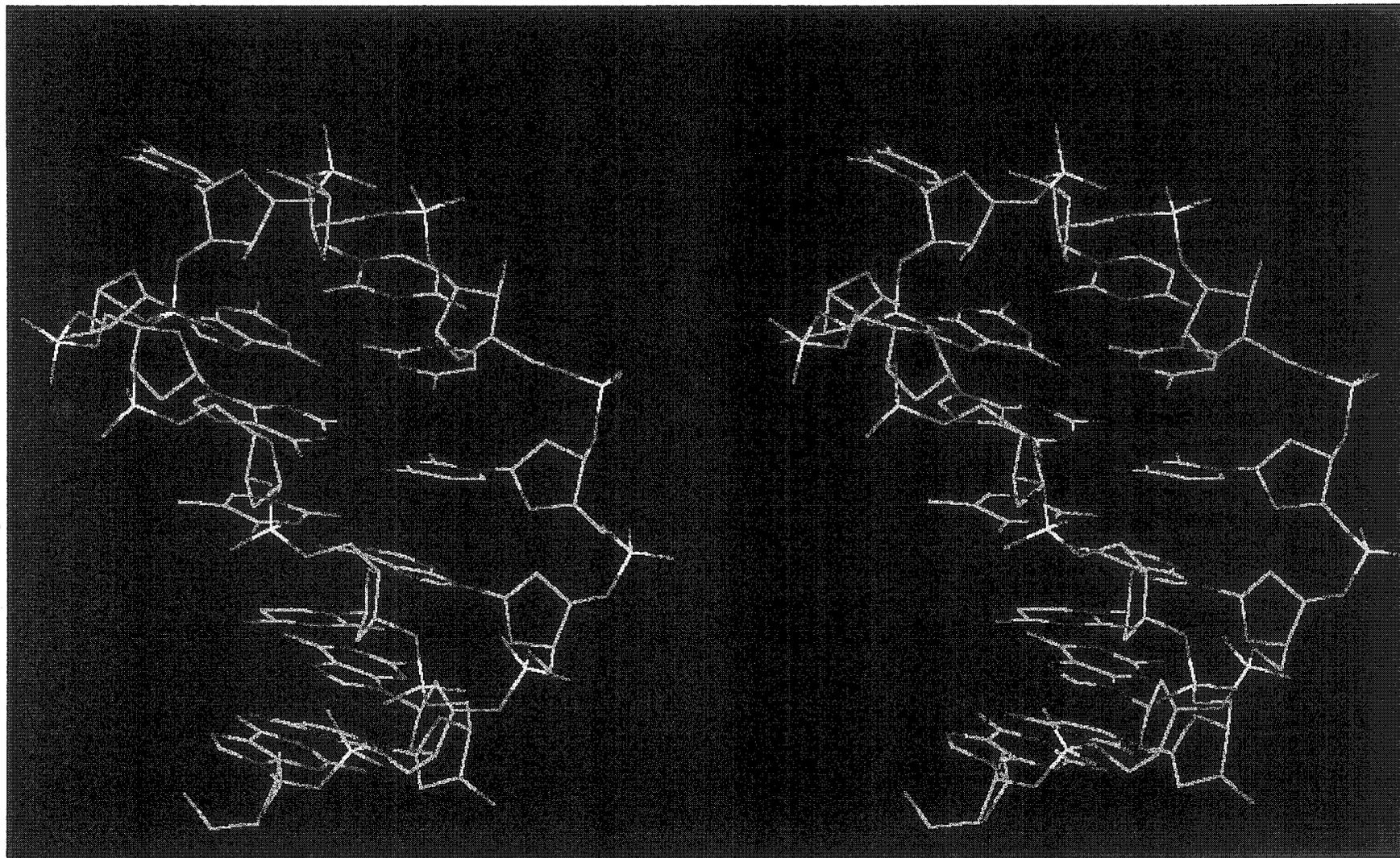


Figure 2.28: Stereoview of the NMR-derived structure of DRD hairpin.

Similarly, hairpin ggac(UUCG)gtcc [$T_m = 61.4$ °C] was destabilized upon changing the second loop residue to ggac(UACG)gtcc [$T_m = 56.7$ °C] by *ca.* 0.7 kcal/mol and thus parallels the trends seen for 2',5'-RNA loop in RNA stem hairpins. Based on these observations, we proposed earlier in this chapter that U5 and G8 residues participate in the loop stabilizing interactions. Now our NMR findings rigidly prove this hypothesis and clearly support the existence of a U5-G8 wobble base pair with both base residues in the *anti* conformation.

A detailed comparison of the thermodynamic parameters for hairpin formation shows that hairpin DRD is significantly more stable than DRD (**Table 2.6**). In order to understand the structural basis for this observation, we studied DRD by NMR. Our data show that the 3',5'-RNA loop in DRD exhibits the same NOEs as those observed for the same loop within the context of RNA stems [*i.e.*, the native hairpin RRR studied by Tinocco and Varani].^{34,53,54} This indicates that the loops have many features in common and we propose that their structures are very similar when found within the context of either DNA or RNA stems. This still does not provide an explanation as to why DRD and RRR are 'extra-stable' but DRD is not. A closer inspection of the sugar J couplings reveals that the DNA sugar stem residues in DRD are very flexible and exist in dynamic equilibrium between the north and south pucker forms. In sharp contrast, the DNA stem in DRD has typical *B*-form features with its sugar residues highly pre-organized in rigid south pucker (*C*-2' *endo*) characteristic of *B*-DNA.

The above NMR findings undoubtedly prove that the 2',5'-RNA loop extra stabilizes RNA and DNA hairpin stems unlike the native 3',5'-RNA loop which is incapable of stabilizing DNA stems. Furthermore, they provide molecular basis for the discrepancies previously observed in the standard free energies of hairpin formation. DRD exhibits a lower standard Gibbs free energy of formation than DRD ($\Delta\Delta G^\circ_{37} = 1.0$ kcal/mol; **2.14** and **2.15**; **Table 2.6**). Inspection of the thermodynamic parameters reveals that this energy preference is a major outcome of stabilization in enthalpy (but not entropy) of hairpin formation, the result of which is indicative of better stacking and pairing interactions. The NMR findings further support this view and provide solid evidence by showing that the stem sugar residues of DRD interchange conformations rapidly. This 'loose' nature of the sugar rings affects helical stacking and electrostatic

interactions and in turn leads to destabilization. In the same way, NMR confirms the thermodynamic parameters observed for RRR versus RRR. Although the 2',5'-RNA loop adopts a unique conformation that is different from that of the 3',5'-RNA loop, RRR still displays typical A-form helical features with all sugar residues fixed in the C-3' *endo* conformation, a behavior similar to that observed for RRR.^{53,54} This explains in part the similarity in the values of the thermodynamic parameters obtained for RRR and RRR hairpin formation (Table 2.6). The flexibility of G8 residue in RRR, observed *via* NMR, might cause a loss in energy (relative to RRR). However, this energy loss is probably compensated by the formation of a wobble base pair, and this is why it does not reflect in the overall standard entropy of formation.

Table 2.6: Thermodynamic parameters of hairpins under NMR study

Designation	Code	5'- Hairpin -2'/3'	T_m (°C)	% H	ΔH° (kcal/mol)	ΔS° (e.u.)	ΔG°_{37} (kcal/mol)
<u>2.1</u>	RRR	GGAC(UUCG)GUCC	71.8	8.5	-53.4	-154.8	-5.4
<u>2.2</u>	<u>RRR</u>	GGAC(UUCG)GUCC	69.3	9.6	-55.6	-162.1	-5.3
<u>2.13</u>	DDD	ggac(uucg)gtcc	56.2	11.3	-36.6	-111.1	-2.1
<u>2.14</u>	DRD	ggac(UUCG)gtcc	54.6	11.5	-36.0	-109.8	-1.9
<u>2.15</u>	<u>DRD</u>	ggac(UUCG)gtcc	61.4	12.6	-39.9	-119.4	-2.9

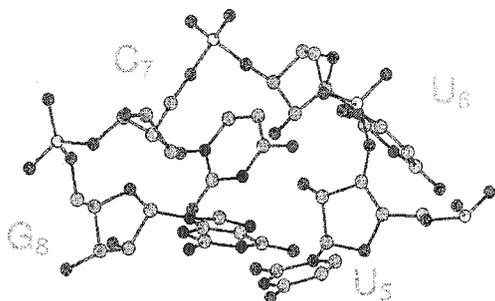
Measurements were made in 0.01 M Na₂HPO₄, and 0.1 mM Na₂EDTA, pH 7.0; oligonucleotide concentration ~ 4.5 μM. Values represent the average of at least five independent measurements. Error in T_m is within ± 1 °C. Errors in thermodynamic parameters are within ± 7.5% for ΔH° and ΔS° , and ± 0.20 kcal/mol for ΔG°_{37} . For a more accurate calculation, ΔG°_{37} was calculated from ΔH° and ΔS° before rounding off and extra significant figures are given in the values of ΔG°_{37} . Capital letters represent RNA residues; underlined letters are 2',5'-RNA residues (*e.g.* UC = U_{2'p5'}C_{2'p}); bold letters are sugar mutations in the loop-closing base pair. Percentage hypochromicity (%H) was calculated from UV absorbances of the hairpin (A_0) and fully denatured species (A_f) using the following equation: %H = ($A_f - A_0$) / A_f .

All in all, the 2',5'-RNA (UUCG) loop displays a uniquely folded and well-ordered structure that is distinct from that of its native 3',5'-linked counterpart.^{53,54,55} **Figure 2.29** highlights the major conformational differences between the two loops. The unusual interactions found within the 2',5'-RNA loop contribute to its unusual stability and can be summarized as follows :

- (1) Both U5 and G8 imino protons have NOEs to amino and imino protons of the C4-G9 base pair, confirming that U5 and G8 form a base pair. The finding that both residues are in *anti* conformation suggests the existence of wobble base pairing between U5 and G8.
- (2) Stacking is continued into the loop from the stem in both DRD and RRR. U5 stacks on top of C4, and G8 (*anti*) stacks on top of G9, thus conserving A-like geometry in the case of RRR.
- (3) The hairpin loop is composed of only two nucleotide residues (U6 and C7), both being in compact C-2' *endo* sugar pucker. The other two residues (U5 and G8) are predominantly in the extended C-3' *endo* conformation. This is in sharp contrast to what has been observed in the native 3',5'-RNA, where the middle two residues (U6 and C7) are the only ones that extend to bridge the stem.^{53,54}
- (4) C7 protrudes out into the solvent and does not participate in any of the loop stabilizing interactions.
- (5) Uracil U6 base stacks on top of U5. This helps stabilize the loop structure and enhances its contribution to the overall thermodynamic stability.

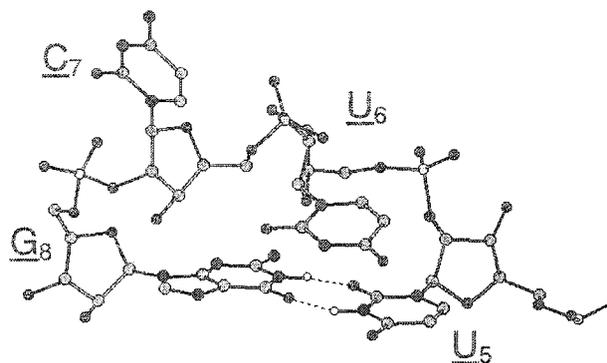
The unique folding pattern of the 2',5'-RNA (UUCG) loop along with its inherent stacking interactions gives it a compact shape that embodies unusual thermodynamic stability. Not only that, its unusual structure is recognized by HIV-1 reverse transcriptase (RT) and can act as an inhibitor of RNase H activity of HIV-1 RT when present within the appropriate hairpin stem (discussed in chapter III). The results described in this chapter not only broaden our understanding of 2',5'-RNA tertiary folding but also of native RNA.

A. 3',5'-RNA



- 1) Stem is A-form. Sugar residues are C-3' *endo* (North pucker)
- 2) U₅ and G₈ form reverse *wobble* base pair [U₅(*anti*) : G₈(*syn*)]
- 3) G₈ is *syn*
- 4) U₆ projects outside the loop
- 5) U₆ stacks on top of the C₇ sugar
- 6) The loop is rigid
- 7) The C₇ amino group forms a contact with the phosphodiester linkage between U₅ and U₆
- 8) U₅ and G₈ (North Pucker) – Compact form for 3',5'-RNA
- 9) U₆ and C₇ (South pucker) – Extended form for 3',5'-RNA

B. 2',5'-RNA



- 1) Stem is A-form. Sugar residues are C-3' *endo* (North pucker)
- 2) U₅ and G₈ form *wobble* base pair [U₅(*anti*) : G₈(*anti*)]
- 3) G₈ is mostly *anti*. Good stacking between C₄-G₉ and U₅-G₈
- 4) C₇ projects outside the loop
- 5) U₆ stacks on top of U₅ base
- 6) The loop is rigid; residue G₈ exhibits mobility in RNA stem duplexes
- 7) H3'(C₇)-H8(G₈); H3'(C₇)-H1(U₆); H1'(U₆) –H6(C₇) contacts
- 8) U₅ and G₈ (North pucker) – Extended form for 2',5'-RNA
- 9) U₆ and C₇ (South pucker) – Compact form for 2',5'-RNA

Figure 2.29: Structural comparison of the general unique features of (A) 3',5'-linked^{53,54,55} versus (B) 2',5'-linked (UUCG) loops.

2.15 CONCLUSIONS

All in all, we have shown that the 2',5'-linked UUCG tetraloop, like the regioisomeric 3',5'-UUCG counterpart, can form superstable hairpin structures of comparable thermodynamic stabilities. We also show that the stability imparted by 2',5'-RNA loops is dependent on base sequence, a property that is shared with the regioisomeric 3',5'-RNA loops. Not only that, UUCG is virtually less dependent on stem

composition and thus exerts higher stability in hairpins with certain stem constructs such as DNA:DNA, 2',5'-RNA:2',5'-RNA, DNA:2',5'-RNA, and RNA:2',5'-RNA compared to UUCG. The UUCG loop exhibits the same order of stability as UUCG only in hairpins with RNA:RNA and DNA:RNA stems. As a result, the relative stabilities of hairpins with a 2',5'-linked tetraloop, *e.g.* ggac(UUCG)gtcc ($T_m = 61.4$ °C), are often superior to those with RNA tetraloops, *e.g.* ggac(UUCG)gtcc ($T_m = 54.6$ °C). In fact, it has been possible to observe the formation of a 2',5'-RNA:DNA hybrid duplex by linking the hybrid's strands to a (UUCG) loop. These duplexes (RD), which are not stable enough to form in an intermolecular complex,¹⁴¹ were stable at room temperature ($T_m \sim 50$ °C). Thus, 2',5'-loops have potentially important implications in the study of nucleic acid complexes where structural data is not yet available. The positive effects of 2',5'-linkages on hairpin formation may be important for the design of novel nucleic acid enzymes¹⁴⁵ as well as antisense agents. These findings also support the hypothesis that 2',5'-RNA (or RNA with both 2',5' and 3',5'-linkages) may have been accessible during early RNA evolution.¹⁴⁰

We have investigated the effect of switching ribose to deoxyribose at the loop-closing base pair of extra-stable RNA and DNA hairpins. We showed that the two riboses at the closing nucleotides (C4 and G9) play a key role in the stabilization of 'extra-stable' RNA hairpins. Incorporation of the superstable 3',5'- or 2',5'-RNA (UUCG) tetraloop within the context of DNA stems, described here for the first time, shows how the compatibility between stem and loop structures modulates overall thermodynamic hairpin stability.

High resolution NMR in combination with molecular modeling experiments, done in collaboration with Prof. Kalle Gehring and Dr. Alexei Denisov of the Biochemistry Department at McGill University, were conducted in order to resolve the unique structure of the 2',5'-RNA tetraloops. The 2',5'-RNA loop in both hairpins with DNA & RNA stems [*i.e.*, RRR and DRD] adopts a conserved unique structure that is distinct from that of the native 3',5'-RNA loop. It is stabilized by a wobble U5·G8 base pair, with both nucleotide residues in *anti* conformation, as well as extensive base stacking and sugar – base contacts. The U5·G8 base pair stacks on top of the C4·G9 loop-closing base pair and thus appears as a continuation of the stem. Unlike the native loop structure, G8 is in *anti* conformation. Moreover, uracil U6 base stacks on top of U5, while C7 protrudes out into

the solvent and does not participate in any of the stabilizing interactions. The proposed loop structure is in agreement with chemical modification experiments. These findings clearly identify the 2',5'-RNA loop as a novel structural motif.

CHAPTER III: SELECTIVE INHIBITION OF HIV-1 REVERSE TRANSCRIPTASE RNASE H BY HAIRPIN APTAMERS EVOLVED FROM A LIBRARY CONSTRUCTED VIA DIVERSITY-ORIENTED SYNTHESIS

3.1 BACKGROUND

Acquired immunodeficiency syndrome (AIDS) is one of the most lethal diseases for which no complete cure has been identified so far. Basic discovery research has attributed the cause of AIDS to a single-stranded RNA virus (retrovirus) referred to as human immunodeficiency virus (HIV).^{172,173} Two genetically distinct subtypes, HIV-1 and HIV-2,^{174,175} have been recognized to date, with the former being identified as the main causative of the disease.

Reverse transcriptase (RT) is the essential enzyme necessary for HIV genomic replication.¹⁷⁶⁻¹⁷⁸ It possesses a multi-activity which comprises an RNA- and DNA-dependent DNA polymerase as well as a ribonuclease H (RNase H). These activities enable the enzyme to reverse transcribe viral RNA to double stranded DNA, hence fundamentally making it one of the most challenging central drug targets in anti-retroviral therapy.¹⁷⁹ Most RT drugs developed to date target mainly its DNA polymerase activity but not RNase H, and they fall into either one of two classes: the first based on nucleoside inhibitors (NRTIs) which inhibit viral replication by acting as chain terminators of DNA synthesis, while the second is more structurally diverse and based on non-nucleoside inhibitors (NNRTIs). The third and rather more recently developed class of HIV-RT inhibitors comprises oligonucleotide constructs (ONRTI). For example, a 28-mer phosphorothioate oligodeoxycytidylate [Sd(C)₂₈] has been shown to bind to HIV-RT and inhibit its DNA polymerase activity.¹⁸⁰ The majority of oligonucleotides acting as effective RT inhibitors have been developed through Systematic Evolution of Ligands by Exponential Enrichment or what is simply referred to as SELEX.¹⁸¹ This technique emerged in 1990 as a promising powerful tool for generating, from randomized nucleic acid pools, oligonucleotide ligands with appropriate desired functions towards specific proteins.^{181,182} The resulting oligonucleotides are referred to as “aptamers”.¹⁸² For instance, RNA pseudoknots composed of at least 24 nucleotide units have been identified by SELEX and were shown to inhibit DNA synthesis catalyzed by HIV-reverse

transcriptase.¹⁸³ Most of these oligonucleotide-based inhibitors have only been selected to target the DNA polymerase activity of HIV-RT without much emphasis on its RNase H activity.

To date, among developed nucleoside, non-nucleoside and oligonucleotide-based inhibitors, surprisingly very few inhibitors have been reported that target selectively the RNase H-mediated activity of HIV-RT, and among these very few are oligonucleotide based. In the next section, we discuss inhibition of RNase H as a valuable therapeutic tool for control of viral gene expression.

3.2 IMPORTANCE OF RNASE H AS A DRUG TARGET

The RNase H activity of HIV-RT is central to retroviral replication. Studies have revealed substantial drops in levels of virus proliferation upon point mutations induced in the RNase H domain of reverse transcriptase.¹⁸⁴ This makes RNase H an attractive target for designing anti-retroviral drugs aimed at suppression of harmful (viral) protein synthesis. RNase H degrades the RNA strand in an RNA:DNA hybrid duplex (**Figure 3.1**), regardless of the identity of the RNA whether it be viral, bacterial, fungal, or tumor-causing mRNA. This allows the DNA strand to disengage and carry on the genetic information for further replication.

There are potentially many strategic inhibitory pathways of RNase H cleavage. Drug ligands can be carefully designed to act in a variety of modes such as blocking the enzyme's catalytic active site, or obstructing the binding to the RNA:DNA hybrid substrate by competitive inhibition. A marine natural product, illimaquinone, has been shown to inhibit non-specifically the RNase H activities of HIV-1 RT, HIV-2 RT, MLV-RT and *E.coli*,¹⁸⁵ whereas 3'-azidothymidylate 5-monophosphate proved to be a potent inhibitor of HIV-1 RNase H with an IC₅₀ of 50 μM (**Figure 3.2 A**).^{186,187} Aptamers, 81 nucleotides long, based on G-quartet motifs were developed using SELEX as inhibitors of HIV-1 RNase H with IC₅₀'s in the nM range.¹⁸⁸ However, these were not solely RNase H selective as they still demonstrated an ability to inhibit the enzyme's DNA polymerase activity.

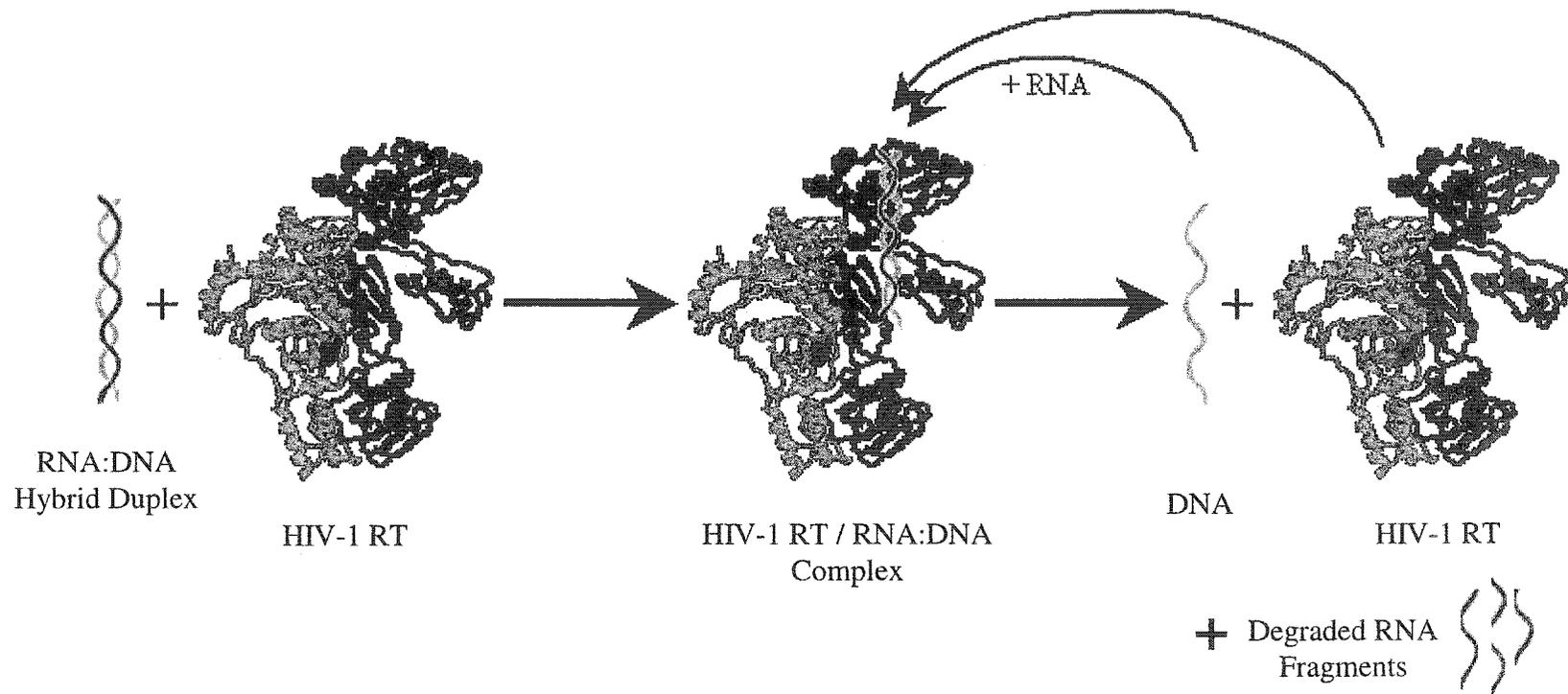


Figure 3.1: HIV-1 RT RNase H-mediated degradation of RNA in an RNA:DNA hybrid duplex. This enzymatic catalytic process allows the DNA strand to disengage and carry on the genetic information for further replication.

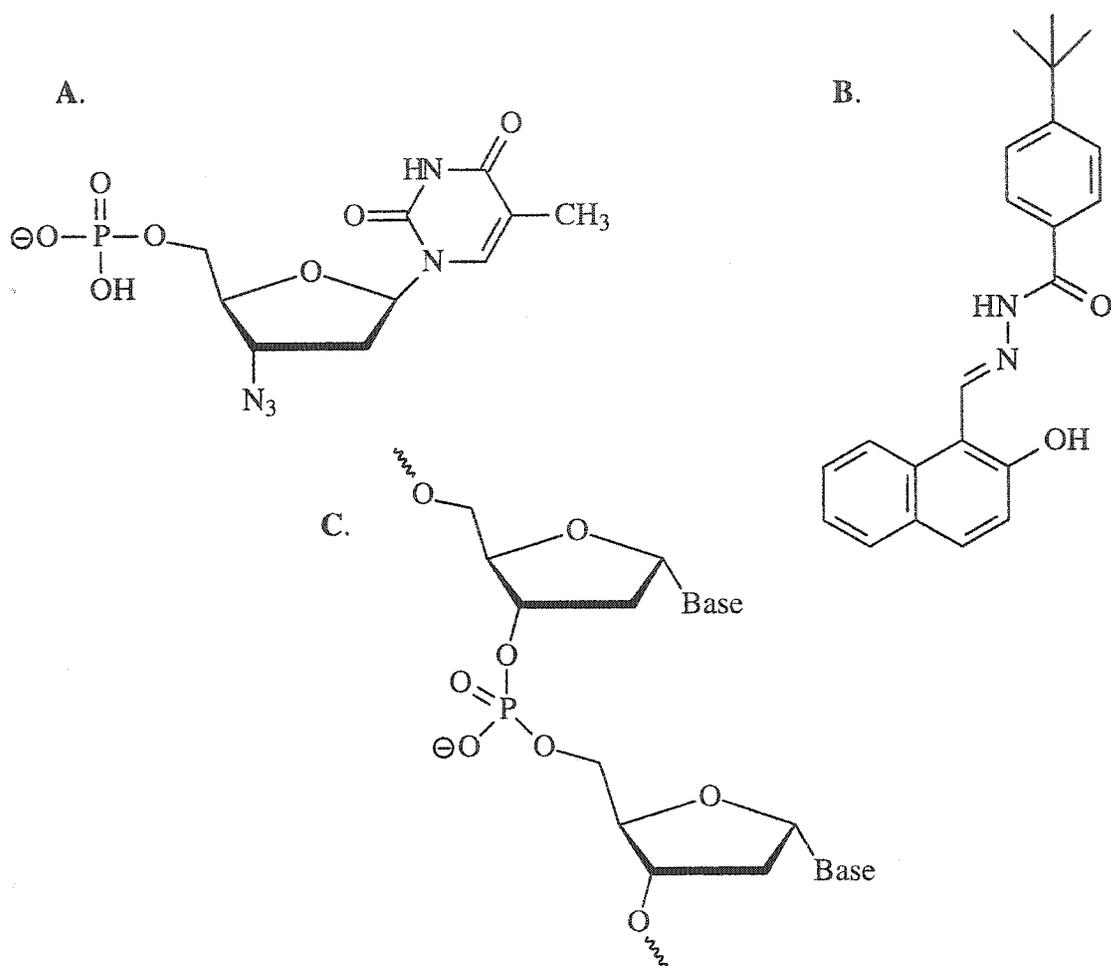


Figure 3.2: Representative examples of the three classes of HIV-RT RNase H inhibitors. (A) 3'-Azido-3'-deoxythymidylate 5'-monophosphate [NRTI]. (B) *N*-(4-*tert*-butylbenzoyl)2-hydroxy-1-naphthaldehyde hydrazone [NNRTI]. (C) Basic skeleton of an α -oligonucleotide which acts as an RNase H inhibitor when hybridized to a complementary RNA strand [ONRTI].

On the other hand, drugs aimed at inhibiting HIV-RT-mediated DNA synthesis were shown to inhibit RNase H activity as well, though at concentrations far higher than those required for inhibition of DNA polymerase activity. For instance, a product isolated from *Juglans mandshurica* has recently been shown to inhibit both DNA polymerase activity ($IC_{50} = 40$ nM) and RNase H activity ($IC_{50} = 39$ μ M) of HIV-1 RT.¹⁸⁹ Parniak and co-workers¹⁹⁰ have shown that *N*-(4-*tert*-butylbenzoyl)2-hydroxy-1-naphthaldehyde hydrazone, an inhibitor of DNA polymerase activity, is also capable of inhibiting RNase H activity with an IC_{50} of 3.5 μ M (Figure 3.2 B). Phosphorothioate oligonucleotides have been shown to inhibit both RNase H and DNA polymerase activities of HIV-1

RT.¹⁹¹ Analogues based on naphthalenesulfonic acid were also reported to exhibit non-selective inhibition of both DNA polymerase and RNase H catalytic functions of HIV-1 RT, with IC₅₀'s in the 15-28 μ M range for the most potent compounds.¹⁹²

Other modified oligonucleotide analogues, when hybridized to complementary RNA, can still bind to the enzyme without evoking its RNase H activity. These can be used to displace RNA:DNA hybrid duplexes thus culminating in competitive inhibition of the RNA strand cleavage. Examples of these include 2',5'-linked RNA,¹⁴¹ 2'-modified analogues such as 2'-O-alkyl, neutral linkages such as phosphoramidates and methylphosphonates, or α -oligonucleotide derivatives containing inverted configuration at the glycosidic (C1') position (**Figure 3.2 C**).¹⁹³⁻¹⁹⁶ Interestingly, RNA duplexes can inhibit the RNase H activity of HIV-1 RT *via* competing against the natural RNA:DNA substrates.^{141,197}

3.3 PROJECT OBJECTIVE

The search for nucleic acid aptamers has become a major area of research focus for the *de novo* design of antiviral drugs. Aptamers have emerged as a new class of therapeutics that rival antibodies even in diagnostics.^{198,199} Our aim is to develop new inhibitors of RNase H activity of HIV-1 RT *in vitro*. As mentioned above, inhibition of RT RNase H activity can potentially disrupt the viral replication cycle. Our endeavor in targeting selectively the RNase H activity of HIV-1 RT stemmed from a long research interest of the Damha laboratory in the discovery of antisense therapeutics. This emerging technology, or “antisense” therapy, has the potential to knock out or suppress completely undesired protein production via the use of oligonucleotides called antisense agents. These are short synthetic nucleic acids designed to bind to mRNA and induce its cleavage *via* invoking RNase H activity. This methodology has been used to suppress viral protein production.

Instead of using antisense agents that target viral mRNA and induce RNase H degradation of the viral mRNA target, our approach utilizes synthetic hairpin aptamers aimed at inhibiting directly the RNase H-associated activity of HIV-1 RT. The idea is to block RNA degradation thus causing RNA:DNA hybrid arrest at this stage of retroviral replication (**Figure 3.3**). In order to achieve this, we constructed a library of hairpin

aptamers with the purpose of screening them for their ability to inhibit HIV-1 RT RNase H activity.

In this chapter, we describe the combinatorial solid-phase synthesis of 48 oligomeric hairpins, 12 nucleotides in length, bearing the same base-sequence but differing in the sugar-phosphate backbone. We also depict the identification of 4 potent molecules from this “library” of compounds that efficiently inhibit RNase H-mediated activity of HIV-1 RT in the 7-30 μM range.

3.4 DESIGN OF MINI-HAIRPINS AS POTENTIAL INHIBITORS OF HIV-1 RT

A close inspection of the molecular requirements for binding of nucleic acids to RNase H reveals a dominant conformational preference,^{141,197} which constitutes the basis of our rationale in designing the hairpin library. Previous studies have shown that recognition of duplexes by the RNase H catalytic site necessitates that they be in A-form geometry (rather than B-form) in order to achieve high binding affinity.¹⁹⁷ Damha and co-workers¹⁴¹ have recently demonstrated that 18 base pair duplexes composed of either RNA:RNA, RNA:DNA or RNA:2',5'-RNA (all in A-form) are able to bind to HIV-1 RT RNase H and inhibit its ability to degrade RNA in RNA:DNA hybrids. In contrast, DNA:DNA duplexes (B-form) of the same sequence failed to show any degree of inhibition. The disadvantage of these molecules to act as potential anti-HIV therapeutics is their rather large molecular weights (~ 11 kD), and thus their unsuitability as drug molecules. Also, the bimolecular nature of these duplexes does not guarantee that the molecule would remain in its duplex form once it enters the cell membrane. The preference in the clinic is for low-molecular weight drug molecules since these are more facile to make in large quantities and easier to manipulate for cellular uptake. We reasoned that short hairpin duplexes, when folded in the right conformation, may potentially bind to HIV-1 RT and inhibit its activity. To facilitate folding, we resorted to hairpin structures containing 'super-stable' (UUCG) tetraloops and stems composed of only 4 base pairs (~ 3.6 kD). Such four base pair duplexes would not form without the existence of the tetraloops.

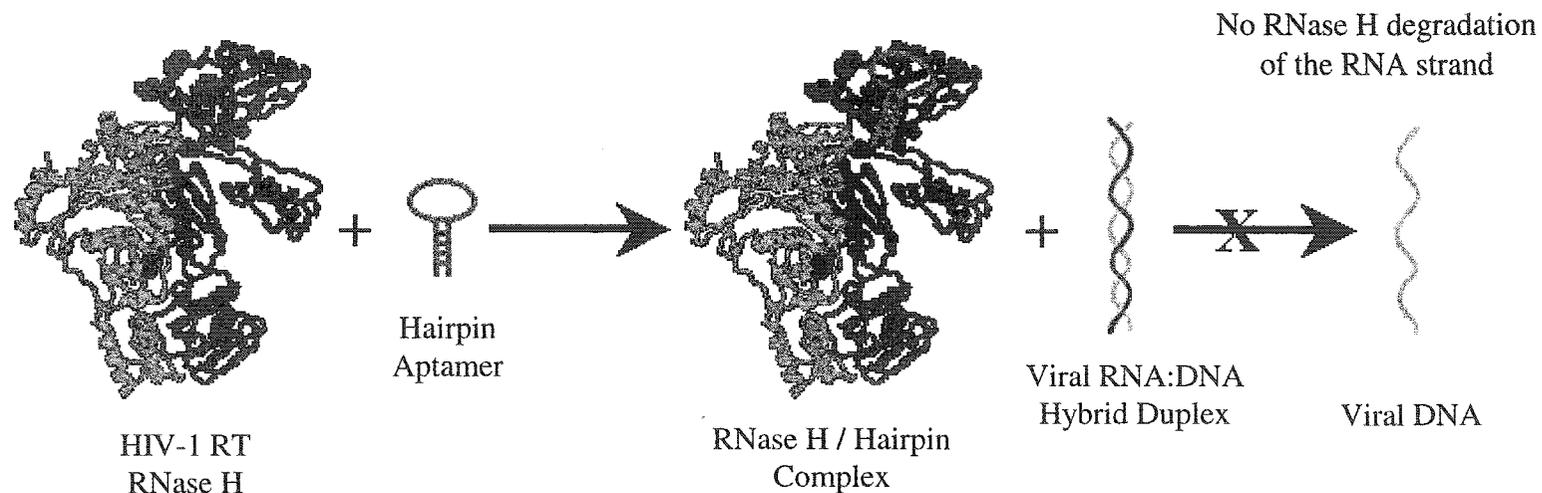


Figure 3.3: Schematic representation of the mode of inhibition of HIV-1 RT RNase H-mediated degradation of viral RNA by small-molecule hairpin aptamers. The hairpin can inhibit RNase H activity in one of several ways: 1) It can bind to the catalytic site, thus blocking enzymatic activity, 2) it can bind to the substrate binding domain and inhibit activity by physical blockage (competitive inhibition), or 3) it can bind to a different site on the enzyme which alters the binding domain or catalytic site, thus inhibiting RNase H activity (non competitive inhibition). All of these plausible mechanisms would lead to hybrid arrest of the viral RNA:DNA duplex, thus preventing further replication.

Hairpins containing various stem combinations spanning the entire conformation spectrum [A-form, A-Like, B-Like, B-Form] were synthesized. For this, we designed a versatile synthetic methodology that provides access to a conformationally-diverse hairpin library. Unlike *in vitro* selection of nucleic acid ligands (SELEX), which starts from arbitrary randomized nucleic acid pools, we adopted a *diversity-generating* synthetic approach as a means of building the hairpin library.

3.5 LIBRARY GENERATION VIA NUCLEIC ACID COMBINATORIAL SPLIT POOL SYNTHESIS

General Considerations

The hairpin sequence [5'-GGACUUCGGUCC-3'] was kept the same throughout library synthesis. This particularly short hairpin with a UUCG loop was chosen because it offers a potential for structural and conformational diversity while still retaining a high thermodynamic stability owing to the presence of the unusually stable loop. Our preliminary testing showed that this hairpin possesses inhibitory activity against HIV-1 RT RNase H. The UUCG loop is a rigid structure displaying a uniquely folded conformation,⁵³⁻⁵⁵ and this may play a role in assisting recognition by HIV-1 RT. Variations in sugar compositions and/or phosphodiester linkages were introduced in the stem and loop regions of this hairpin. Even the slightest modifications would be expected to perturb the overall structure, thus offering access to a broad range of conformationally diverse molecules.

Library Nomenclature

The compounds generated in the library are hairpins (**Figure 3.4**) containing the base sequence: 5'- GGAC (UUCG) GUCC-3'. The [GGAC] segment is referred to as the 5'-stem, and [GUCC] as the 3'-stem, while the (UUCG) fragment constitutes the hairpin loop. In our library nomenclature, the letters “D”, “R”, and “R” refer to DNA, RNA, and 2',5'-linked RNA segments respectively (**Figure 3.5**). Thus, an oligomer designated as RRR would refer to an all-RNA hairpin of the above sequence, while DRR for example

represents a hairpin of the same sequence containing a 2',5'-linked RNA loop and a DNA:RNA hybrid stem. The DNA residues are located at the 5'-stem [GGAC], whereas the RNA residues are located at the 3'-stem [GUCC].

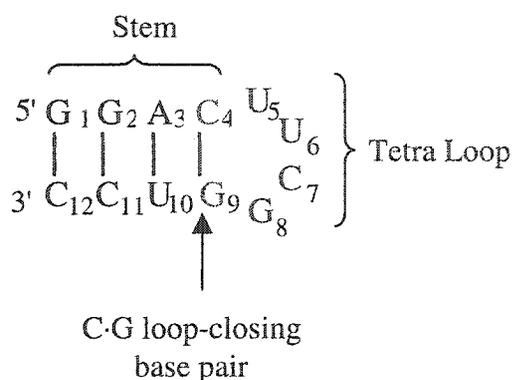


Figure 3.4: Secondary structure and base sequence of the hairpin constituting the building core for library generation. The structure consists of a loop (U₅U₆C₇G₈), and a stem region [defined by G₁-C₄ and G₉-C₁₂ residues] that includes the “loop-closing base pair” C-G. In *diversity-generating* library synthesis, the stem can be either DNA:DNA “DD”, RNA:RNA “RR”, 2',5'-RNA:2',5'-RNA “RR”, DNA:RNA “DR”, DNA:2',5'-RNA “DR”, or RNA:2',5'-RNA “RR”, while the loop was modified with DNA “D”, RNA “R” and 2',5'-RNA “R” residues.

Library Synthesis

Syntheses were achieved on an ExpediteTM DNA/RNA synthesizer containing nine monomer reservoirs (Figure 3.6). Our library synthesis took advantage of the presence of only three base-identities in each of the 5'-stem, 3'-stem, and loop regions. This would save on the number of times the monomer bottles had to be changed on the synthesizer and allowed only a one-time bottle change procedure throughout the whole synthesis of the 45-member library. DNA, RNA and 2',5'-RNA monomers of each G, C and U(or T) were first installed at once on the gene machine, thus allowing structural *diversity-oriented* synthesis of 5'-GUCC-3' (3'-stem) (Figure 3.5). The loop was then synthesized via parallel combinatorial synthesis (Figure 3.5). After this, the monomer bottles containing U (or T) were replaced by A (DNA, RNA, and 2',5'-RNA synthesis) and the synthesis of the 5'-GGAC-3' segment (5'-stem) was continued to generate 27 hairpin molecules (Figure 3.5).

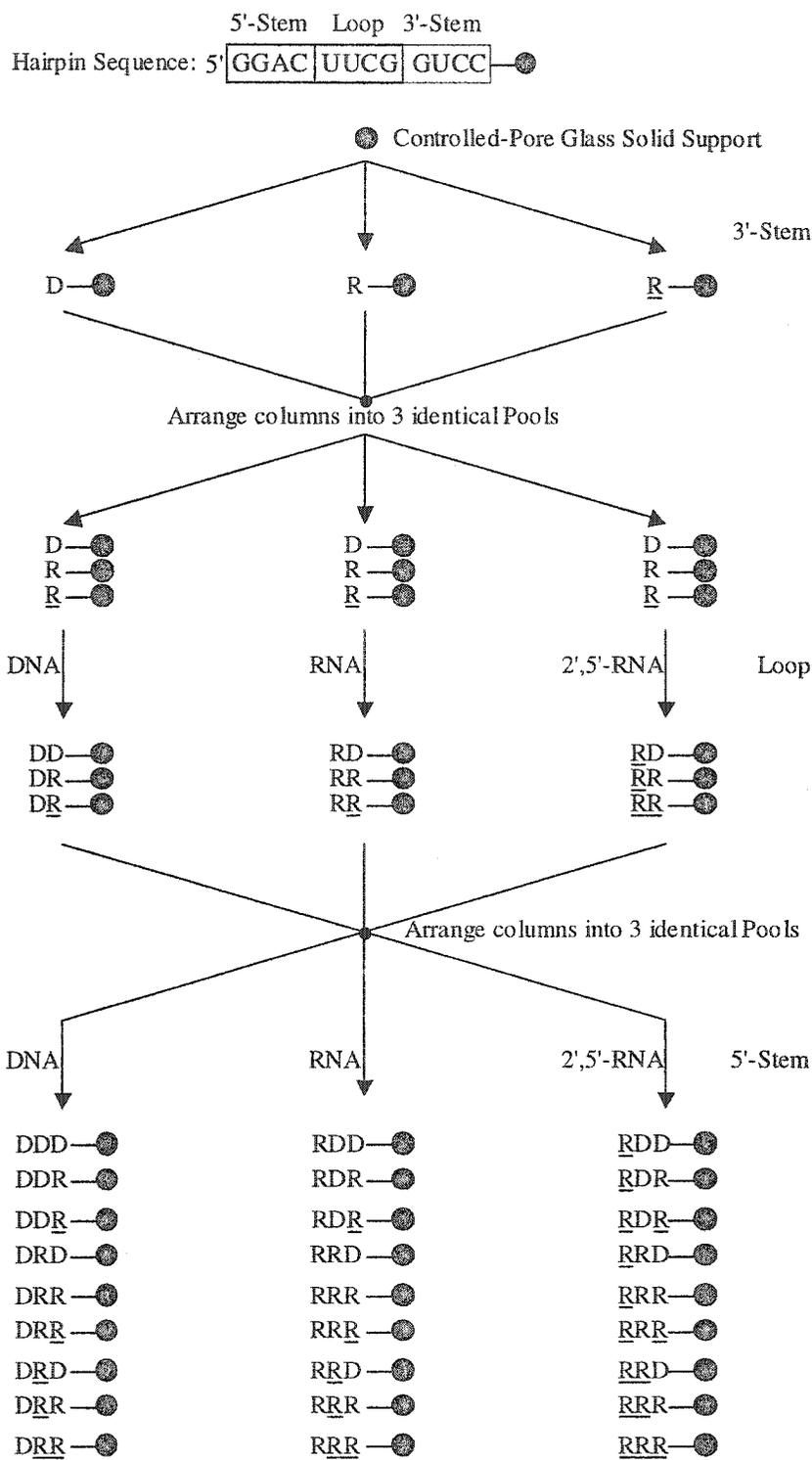


Figure 3.5: Protocol for the *diversity-generating* combinatorial synthesis of a 27-member hairpin library. The synthesis proceeds from the 3'- to the 5'-end. Starting with the 3'-stem, [GUCC] is synthesized as DNA, RNA or 2',5'-RNA fragment via parallel combinatorial synthesis. The columns (containing CPG-support) are then equally arranged *without mixing* into three identical equivalent pools. Subsequent combinatorial synthesis would yield the loop region in again three different morphologies (DNA, RNA, and 2',5'-RNA). The same approach (arranging into pools followed by combinatorial synthesis) installs the 5'-stem, thus yielding 27 hairpin members.

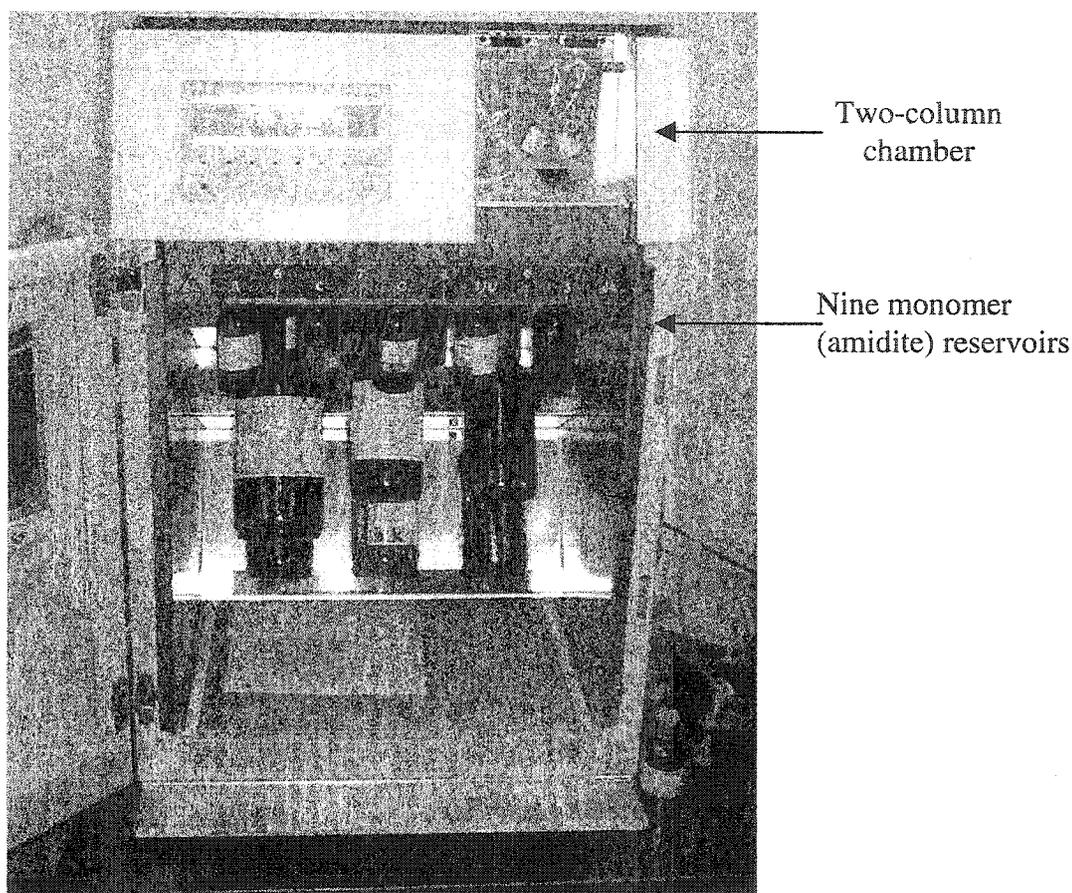


Figure 3.6: Expedite™ 8909 DNA/RNA synthesizer used in diversity-oriented combinatorial synthesis. The advantage of this machine is that it can accommodate nine monomer bottles, and the capacity to execute two syntheses simultaneously.

By following the same *diversity-generating* combinatorial approach, we were able to synthesize 18 other molecules incorporating modifications at specific sites in the hairpin sequence (Table 3.1). The library was prepared as a single copy (one CPG-column per library member) on a 1 μ mol scale using standard phosphoramidite chemistry with slight modifications (see experimental Section 7.2.4), and the individual oligomers were cleaved off the solid support, purified by gel electrophoresis or ion-exchange HPLC, and subsequently characterized by MALDI-TOF mass spectrometry (for a list of representative values, see experimental Section 7.4.2). The overall isolated yields obtained from library solid-phase synthesis were 20-30%. The formation of *hairpin* species for all library members was verified by the T_m independence of oligonucleotide

concentration over at least a 30-fold concentration range (Figure 3.7), thus confirming a *unimolecular* folding process for individual library members in solution (0.01 M Na₂HPO₄, 0.1 mM Na₂EDTA, pH 7.0).

Table 3.1: Diversity-oriented library synthesis: Representative 16 hairpin chimeras obtained by combinatorial split synthesis

5'- Sequence -3'	Code	5'- Sequence -3'	Code
5'- GGAC(UUCG)GUCC —●	R _C RR —●	5'- ggaC(UUCG)Gtcc —●	D _C R _G D —●
5'- GGAC(UUCG)GUCC —●	RR _G R —●	5'- ggaC(UUCG)Gtcc —●	D _C R _G D —●
5'- GGAC(UUCG)GUCC —●	RR _U R —●	5'- ggaC(UUCG)Gtcc —●	D _C R _G D —●
5'- GGAC(UUCG)GUCC —●	R _C RR —●	5'- ggaC(UUUU)Gtcc —●	D _C U _G D —●
5'- GGAc(UUCG)GUCC —●	R _c RR —●	5'- ggac(UACG)gtcc —●	DR ¹ D —●
5'- GGAC(UUCG)gUCC —●	RR _g R —●	5'- ggac(UUUG)gtcc —●	DR ² D —●
5'- GGAc(UUCG)gUCC —●	R _c R _g R —●	5'- ggac(UUUU)gtcc —●	DR ³ D —●
5'- GGAc(UUCG)gUCC —●	R _c R _g R —●	5'- tttt (UUCG) tttt —●	TRT —●

Capital letters represent RNA residues; underlined letters are 2',5'-RNA residues (*e.g.* UC = U_{2p5}C_{2p}); DNA residues are represented by small letters; bold letters represent specific point mutations.

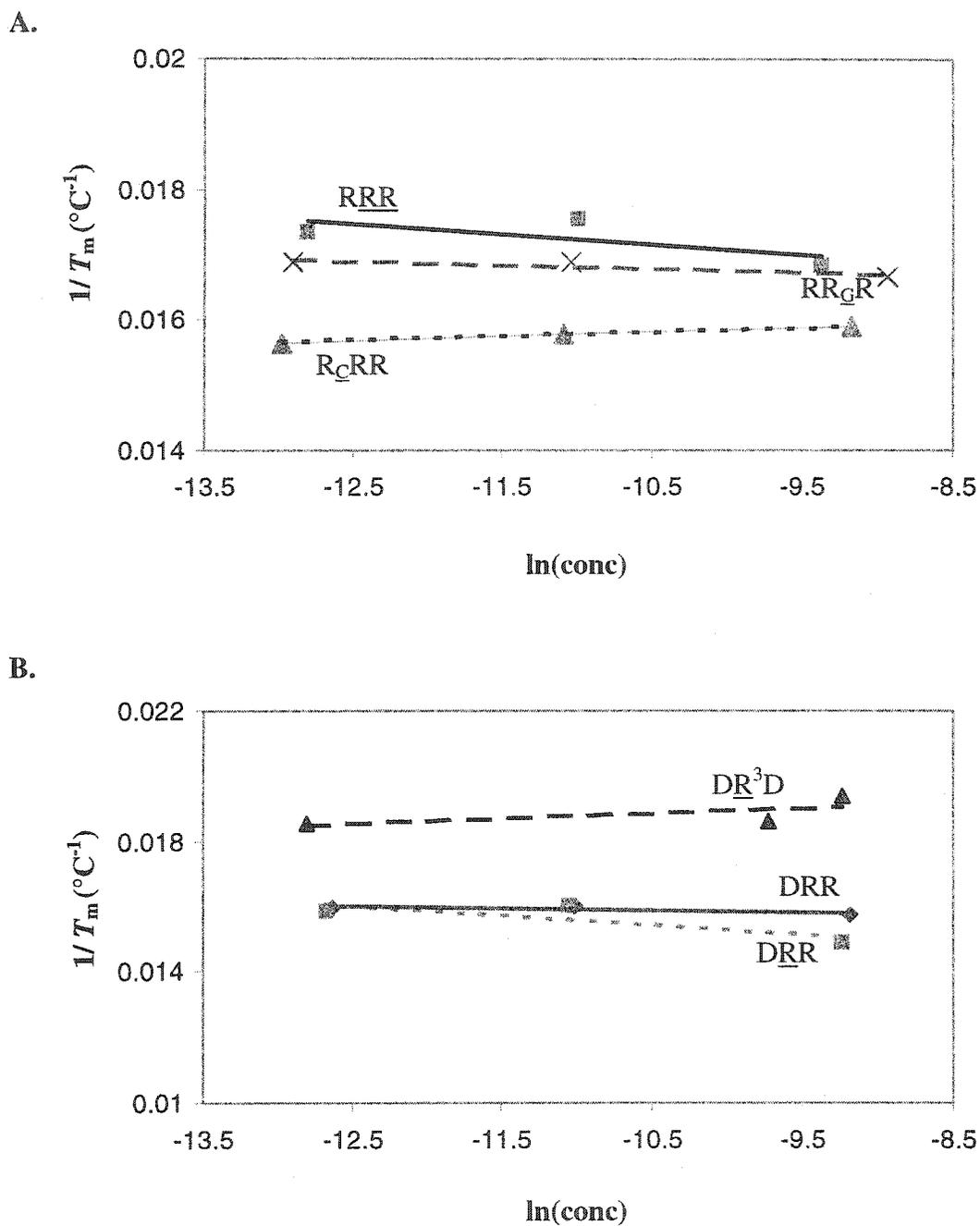


Figure 3.7: Van't Hoff plots of representative hairpin library members. The spectra show concentration independence over a 50-fold range, consistent with a *unimolecular* transition process. UV thermal melting measurements were recorded at 260 nm.

3.6 CONFORMATIONAL DIVERSITY OF LIBRARY MEMBERS

The global conformations of all library members were qualitatively studied by circular dichroism spectroscopy and as expected, they span a wide spectrum of conformational diversity ranging from typical A- or B-forms to intermediate hybrid structures (see **Figure 3.8** for representative examples). Based on the characteristics of the CD spectra, the conformation of each hairpin molecule was analyzed and subsequently denoted as either typical A-form or B-form, A-like or B-like, or I-form (“intermediate-form”). Typical A-form spectra were identified from the characteristic strong positive CD band at *ca.* 265 nm and the negative CD peak at *ca.* 212 nm.¹ Typical B-form spectra were identified from their red-shifted CD band at *ca.* 278 nm.¹ Other structures exhibited strong positive CD bands that were mixtures of the 265 and 278 nm bands.⁴ Depending on the extent of their resemblance to the typical A- or B-spectra, they were classified as either A-like or B-like. In cases when this could not be inferred (*i.e.*, no prominent resemblance for either A- or B-form), the structures were classified as belonging to the I-form. For example, the hairpin RRR exhibits typical A-form conformation, while the global conformation of DDD is typical B-form. In contrast, hairpins RRR, DRD and DRD have different characteristics and their conformations fall in between with A-like, I-form and B-like features, respectively. Furthermore, we have studied these three hairpins by high-resolution NMR (Chapter II), thus providing a good model system to test our conformation scale derived from circular dichroism. As described earlier in **Section 2.14**, we found that the DNA stem residues in DRD are in dynamic equilibrium between the C-2'*endo* and C-3'*endo* puckers and the rU₆ and rC₇ loop residues are locked in C-2'*endo* pucker, whereas the sugars of the ribo U₅·G₈ base pair adopt the C-3'*endo* sugar pucker. This confirms an I-form global conformation for the DRD hairpin. On the other hand, the DNA stem of DRD exhibits typical B-form features with 100% C-2'*endo* sugar repeats, while one of its loop residues is in C-3'*endo* pucker, thus confirming a B-like overall structure. These results clearly support our qualitative conformational designation of each of the library hairpin members using CD spectroscopy (**Figure 3.8**).

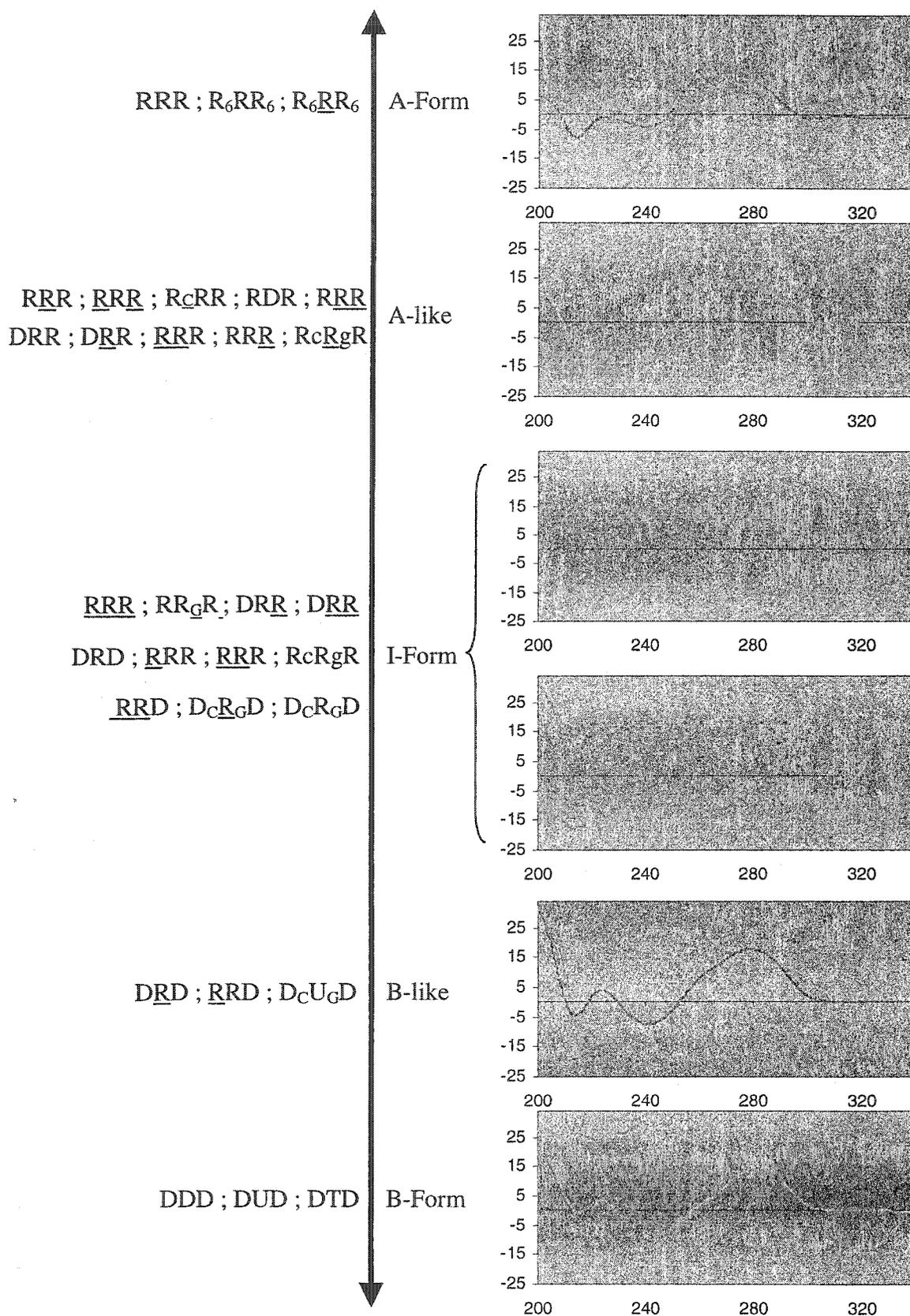


Figure 3.8: Conformation spectrum generated by diversity-oriented synthesis.

3.7 SELECTION OF POTENT HIV-1 RT INHIBITORS

We have classified the library members into six different hairpin classes. **Table 3.2** enlists all library members along with their inhibitory constants and thermal melting data. The first class (entries 1-4) comprises DNA hairpins having DNA “D”, RNA “R” and 2',5'-RNA “R” loops. The second class (entries 5-7) was designed to test the effect of loop base sequence on the inhibitory properties of DRD hairpins. The third class (entries 8-10) is similar to the first class, but contains RNA residues in the stem region. The fourth class (entries 11-22) comprises hairpins that are derivatives of the first or third class that contain one or two sugar-phosphate backbone modifications while keeping the base sequence unchanged. The fifth class (entries 23-38) was designed to test the effects of the stem (DD, RR, RR, DR, DR, or RR) on the inhibition of RNase H activity. Finally, the sixth class (entries 39 and 40) was designed to test the effect of stem length on inhibition of RNase H activity.

In collaboration with Dr. Kyung-Lyum Min from our research group, we screened all hairpin molecules for their ability to act as potential inhibitors of HIV-1 RT RNase H activity. The inhibition assay comprised a 5'-[³²P]-labeled RNA oligonucleotide (18-mer) that was annealed to a complementary unlabeled DNA strand. The formed 5'-[³²P]-RNA:DNA hybrid duplex was then incubated with HIV-1 RT in either the absence or presence of variable amounts of hairpins at 37 °C (**Figure 3.9**). The ability of various hairpins to inhibit HIV-1 RT RNase H-mediated degradation of the 5'-³²P-labeled RNA strand in the RNA:DNA hybrid was measured by gel densitometric analysis as judged from the disappearance of the full-length RNA substrate and/or the appearance of the smaller degradation products. The IC₅₀ value, defined as the hairpin concentration required to inhibit 50% of RNase H-mediated RNA degradation in the RNA:DNA hybrid, was calculated from plots of the residual undegraded 5'-[³²P]-RNA *versus* hairpin concentration (**Figure 3.9**).

The degree of inhibition varied with loop and stem compositions. Hairpins composed of DNA loops were not able to inhibit RNase H activity regardless of hairpin stem composition (**Table 3.2**). For example, hairpins DDD, RDR, DDR, and RDR, all containing DNA residues in the loop, showed no inhibition of HIV-1 RT RNase H mediated degradation of RNA in the RNA:DNA hybrid.

Table 3.2: Inhibition of RNase H activity of HIV-1 RT by a nucleic-acid hairpin library^a

Entry	Code	5'- Sequence -2'/3' ^b	IC ₅₀ (μM)	T _m (°C)
1	DDD	ggac(uucg)gtcc	-	56.2
2	DTD	ggac(tttt)gtcc	-	54.7
3	DRD	ggac(UUCG)gtcc	-	54.6
4	DRD	ggac(UUCG)gtcc	96	61.4
5	DR ¹ D	ggac(UACG)gtcc	-	56.7
6	DR ² D	ggac(UUUG)gtcc	69.1	62.0
7	DR ³ D	ggac(UUUU)gtcc	97.2	54.5
8	RDR	GGAC(uucg)GUCC	-	63.4
9	RRR	GGAC(UUCG)GUCC	25.8	71.8
10	RRR	GGAC(UUCG)GUCC	68.9	69.3
11	R _c R _g R	GGAc(UUCG)gUCC	>100	60.0
12	R _c R _g R	GGAc(UUCG)gUCC	39.4	57.6
13	R _c RR	GGAc(UUCG)GUCC	-	67.3
14	RR _g R	GGAC(UUCG)gUCC	46	66.6
15	RR _U R	GGAC(UUCG)GUCC	50.1	60.2
16	R _c RR	GGAc(UUCG)GUCC	>100	62.6
17	RR _G R	GGAC(UUCG)GUCC	98	58.0
18	R _c RR	GGAc(UUCG)GUCC	-	61.5
19	D _c R _G D	ggaC(UUCG)Gtcc	-	59.5
20	D _c U _G D	ggaC(UUUU)Gtcc	-	51.6
21	D _c R _G D	ggaC(UUCG)Gtcc	-	52.3
22	D _c R _G D	ggaC(UUCG)Gtcc	-	57.0
23	DDR	ggac(uucg)GUCC	-	n.d.
24	DRR	ggac(UUCG)GUCC	-	56.5
25	DRR	ggac(UUCG)GUCC	-	56.7
26	RDD	GGAC(uucg)gtcc	-	n.d.
27	RRD	GGAC(UUCG)gtcc	-	48.1
28	RRD	GGAC(UUCG)gtcc	47	52.8
29	DRR	ggac(UUCG)GUCC	>100	24.1
30	DRR	ggac(UUCG)GUCC	71.4	30.2
31	RDR	GGAC(uucg)GUCC	-	n.d.
32	RRR	GGAC(UUCG)GUCC	-	62.6
33	RRR	GGAC(UUCG)GUCC	58.9	62.4
34	RRR	GGAC(UUCG)GUCC	-	54.1
35	RRR	GGAC(UUCG)GUCC	42.8	58.1
36	RRR	GGAC(UUCG)GUCC	26.2	45.2
37	RRR	GGAC(UUCG)GUCC	88.5	54.8
38	TRT	tttt(UUCG)tttt	-	-
39	R ₆ RR ₆	GUGGAC(UUCG)GUCCAC	7.8	n.d.
40	R ₆ RR ₆	GUGGAC(UUCG)GUCCAC	29.7	n.d.

^a IC₅₀ is the hairpin concentration required to inhibit 50% RNase H activity of HIV-1 RT and was determined as described in Materials and Methods. Values represent the average of 2 to 3 independent measurements. Errors in IC₅₀ values represent standard deviations and were within ± 3 μM. ^b Capital letters represent RNA residues; underlined letters are 2',5'-RNA residues (e.g. UC = U_{2p5}C_{2p}); DNA residues are represented by small letters; bold letters represent specific point mutations in loop base sequence (entries 5-7) or sugar/phosphodiester composition (entries 11-22). (n.d. = T_m was not determined)

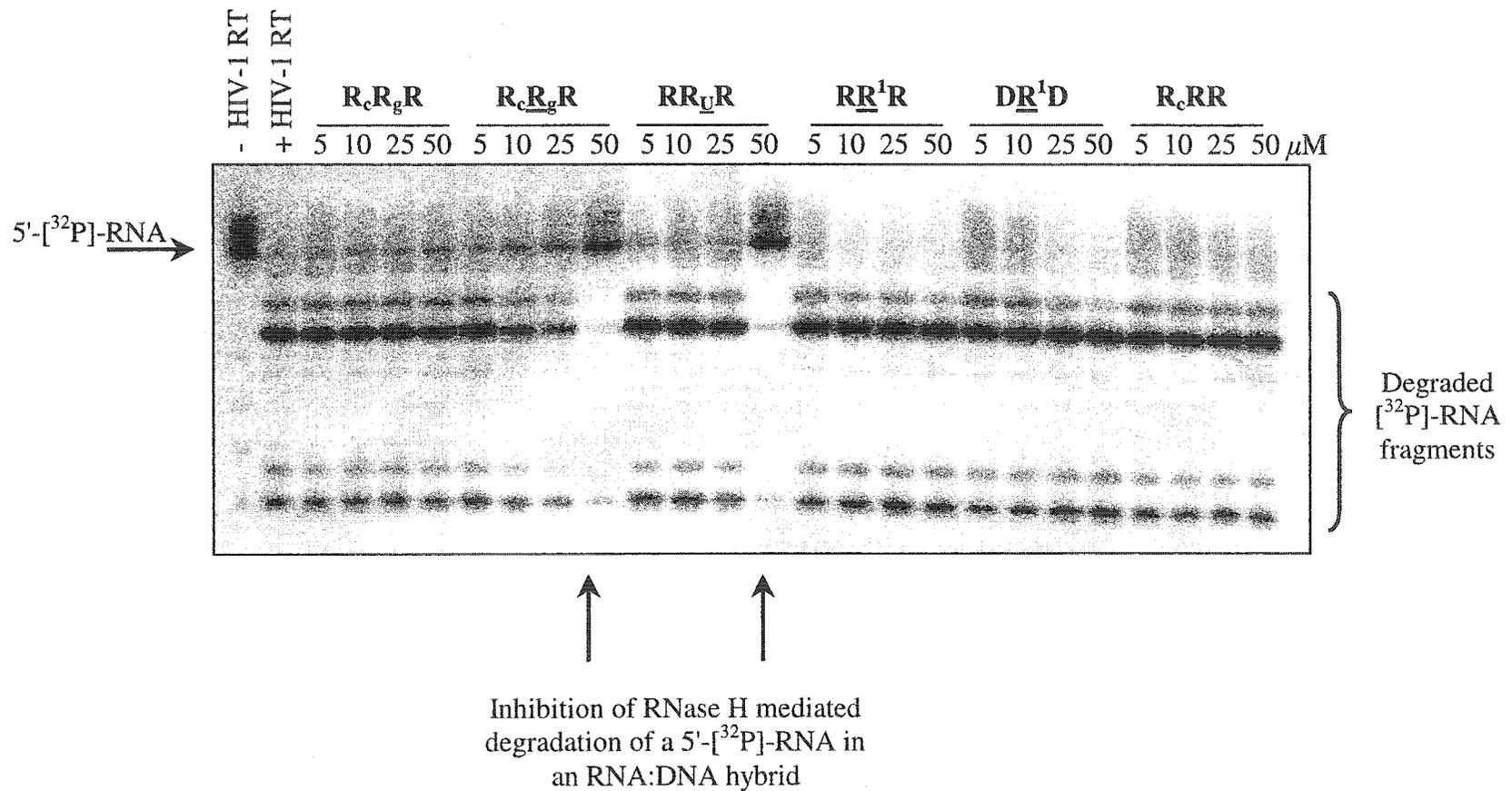


Figure 3.9: Gel autoradiogram illustrating RNase H inhibitory activity of representative hairpin library members. A [³²P]-RNA:DNA hybrid was incubated with HIV-1 RT in the presence of varying amounts of hairpins (0-50 μM). The base sequence of RR¹R is: 5'-GGAC(UACG)GUCC-3'.

On the other hand, hairpins composed of either R or R loops showed various degrees of inhibition depending on hairpin stem composition (**Figure 3.10**). IC₅₀ values were in the 7.8 – 100 μM range. As shown in **Table 3.2**, the hairpin RRR was a potent inhibitor of HIV-1 RT RNase H activity with an IC₅₀ of 25.8 μM. Replacing the loop with 2',5'-RNA [RRR] resulted in an increase in IC₅₀ to 68.9 μM. Among members of the fourth class, the most potent was R_cR_gR [*i.e.*, 5'-GGAc(UUCG)gUCC-3'; IC₅₀ = 39.4 μM]. Of note, RRR was the most potent among members of the fifth class with an IC₅₀ similar to that of RRR (~ 26 μM). Hairpins RRD and RRR (entries 28 and 35) were also excellent inhibitors of HIV-1 RT RNase H activity with almost similar IC₅₀ values (~ 45 μM). On the other hand, the corresponding hairpins with 3',5'-RNA loops, *i.e.*, RRD and RRR, did not show any degree of inhibition of HIV-1 RT RNase H activity.

Increasing stem length resulted in significant increase in inhibitory activity. The hairpin R₆RR₆ (with 6 base pairs in the stem, IC₅₀ = 7.8 μM) was approximately three times more potent than RRR, while R₆RR₆ (IC₅₀ ~ 30 μM) was two times more potent than RRR [IC₅₀ ~ 69 μM]. This led to identifying the RNA hairpin R₆RR₆ as the most potent inhibitor among all library members against HIV-1 RT RNase H activity.

In order to verify specificity of inhibition and to validate the potential of these hairpins as therapeutic agents, we have tested their ability to inhibit human RNase H (type II) as well as bacterial RNase H. Both hairpins R₆RR₆ and RRR showed no degree of inhibition with either human RNase H or bacterial RNase H (**Figure 3.11**), suggesting a remarkable specific effect towards the retroviral enzyme (HIV-1 RT).

To exclude the possibility that the hairpin itself was acting as a substrate for the enzyme (which in turn is responsible for the decrease in the observed rate of RNase H degradation of the natural RNA:DNA substrate), the most potent hairpins (R₆RR₆, RRR, RRR, R_cR_gR) were 5'-[³²P]-labeled and then incubated with HIV-1 RT in the absence of RNA:DNA substrate. For example, after 60 minutes of incubation at 37 °C, we observed no degradation whatsoever of hairpin RRR, indicating that it is not a substrate of HIV-1 RT (*e.g.*, see **Figure 3.12**). Under the same conditions, the RNA:DNA hybrid control was completely cleaved within 20 minutes in the absence of hairpin RRR. The hairpins R₆RR₆, RRR, and R_cR_gR were also not substrates of the enzyme. Additionally, hairpins DRR and DRR were also tested and unexpectedly proved not to be substrates for HIV-1

RT RNase H (data not shown). These hairpins, though not inhibitors of RNase H activity (Table 3.2), were considered due to the appealing nature of their hybrid stem which resembles that of the RNA:DNA substrates recognized by the enzyme.

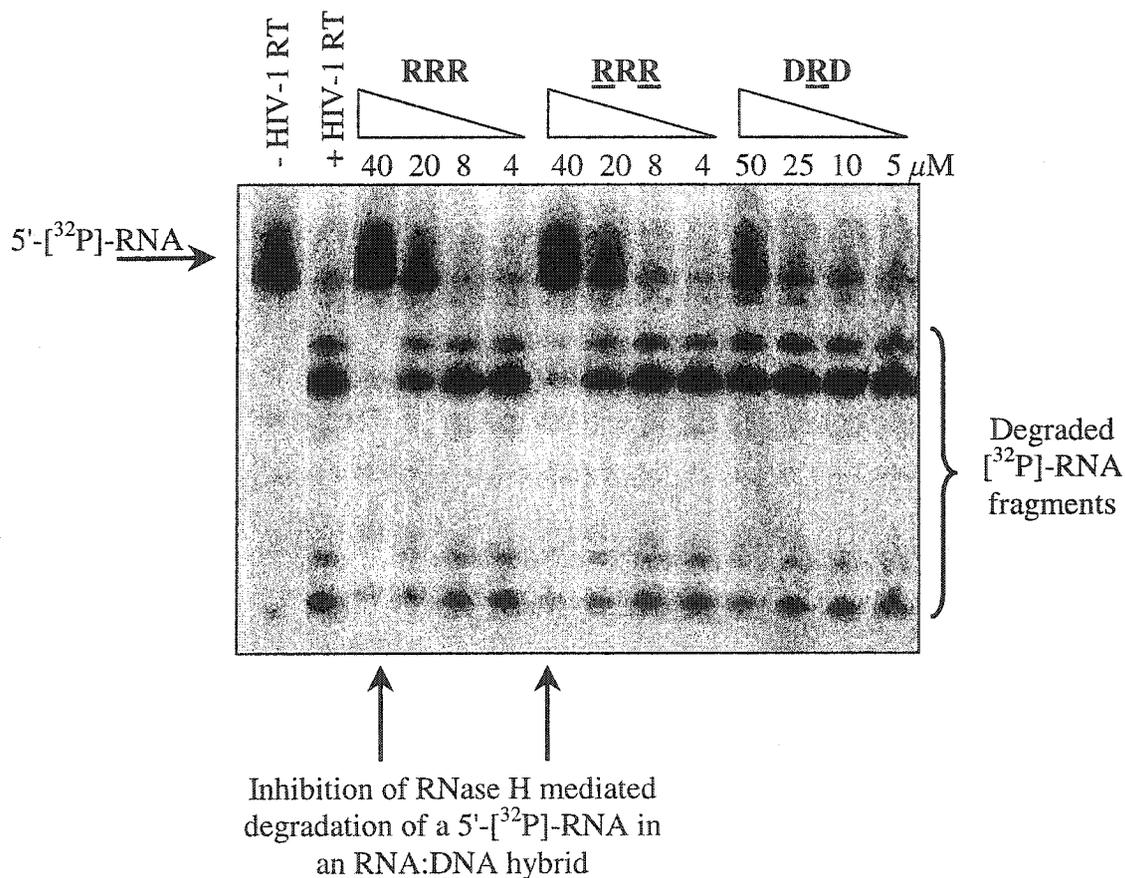


Figure 3.10: Inhibition of HIV-1 RT RNase H activity by representative hairpins with either R or R loops. A [³²P]-RNA:DNA hybrid was incubated with HIV-1 RT in the presence of varying amounts of hairpins. For RRR and RRR, concentration was varied over a 0-40 μM range, while for DRD it was over a 0-50 μM range.

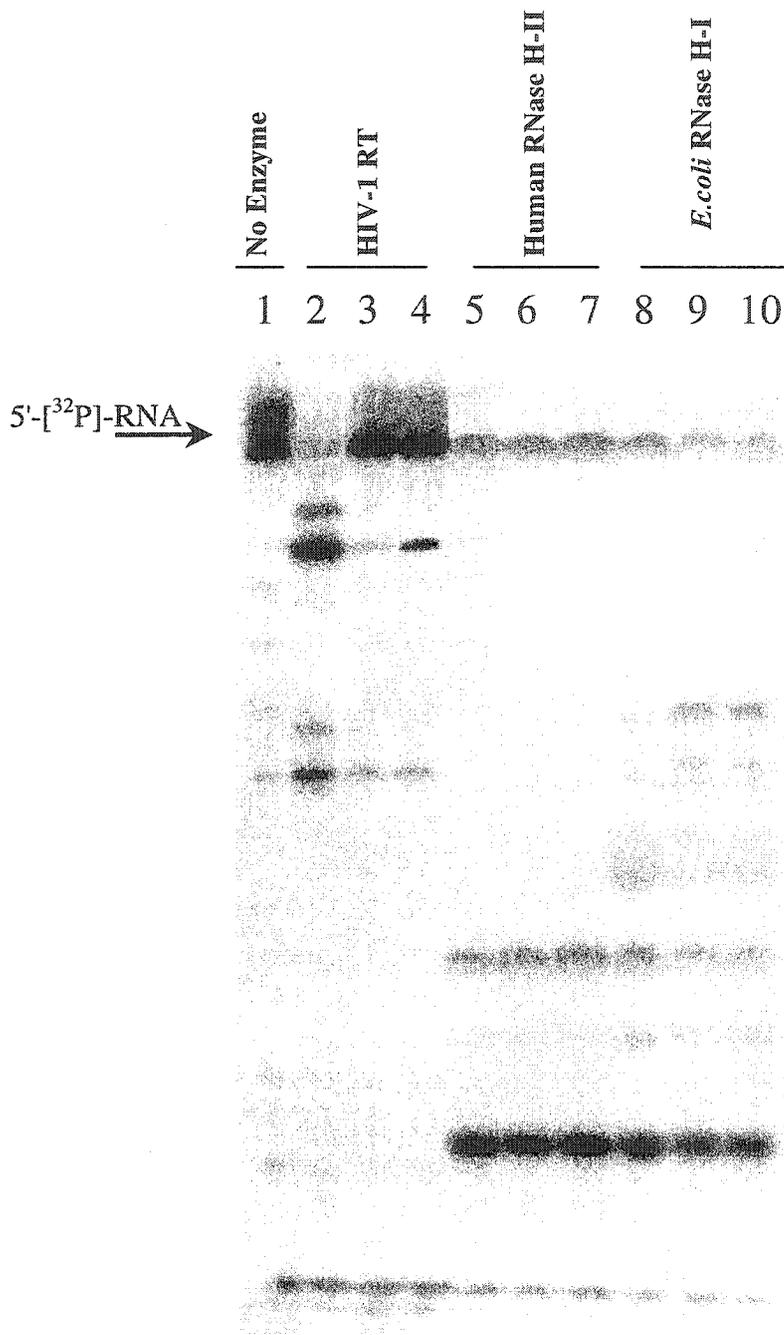


Figure 3.11: Specific inhibition of *in vitro* HIV-1 RT RNase H activity. Hairpins R₆RR₆ and RRR were tested for their ability to inhibit Human RNase H (type II) and bacterial *E. coli* RNase H. The experiments measured the RNase H-mediated decrease in the amount of the 18-mer RNA of a 5'-[³²P]-labeled RNA:DNA duplex substrate in the absence or presence of unlabeled hairpin aptamer; lane 1, control (no enzyme); lane 2, + RT, no hairpin; lane 3, + RT, + R₆RR₆ hairpin; lane 4, + RT, + RRR hairpin; lane 5, + human RNase H, no hairpin; lane 6, + human RNase H, + R₆RR₆ hairpin; lane 7, + human RNase H, + RRR hairpin; lane 8, + *E. coli*, no hairpin; lane 9, + *E. coli*, + R₆RR₆ hairpin; lane 10, + *E. coli*, + RRR hairpin.

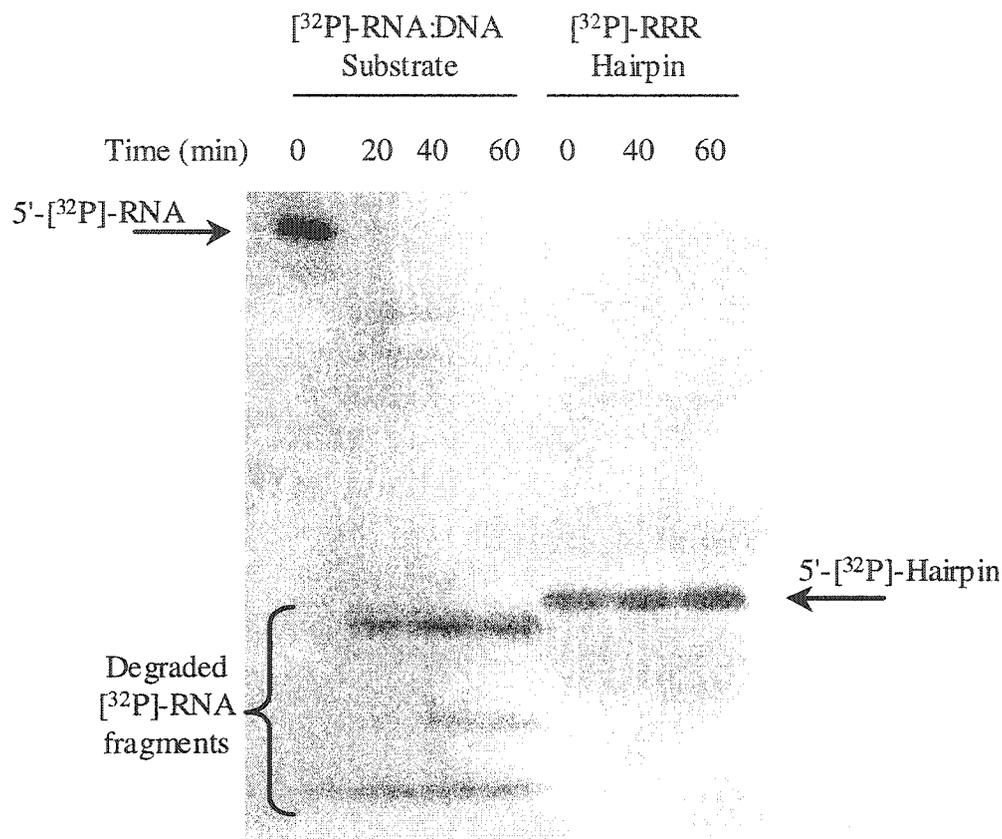


Figure 3.12: Integrity of hairpin aptamers towards RNase H activity. A representative potent hairpin aptamer was incubated with the enzyme in the absence of the natural RNA:DNA substrate. The gel shows no enzyme degradation of the hairpin aptamer, excluding the possibility that this might have been a reason for the observed decreased rate of degradation of the natural substrate in the inhibition assays performed earlier. The 5'- $[^{32}\text{P}]$ -RNA strand in the RNA:DNA substrate, in the absence of hairpin inhibitor, is completely degraded within 20 min.

3.8 STRUCTURE-ACTIVITY RELATIONSHIPS

From the data presented in **Table 3.2**, many correlations can be drawn that help to elucidate the nature of the interaction between HIV-1 RT and the hairpin aptamers. These are exemplified in the hairpin secondary structure, the unusual loop conformation as a recognition motif, thermal stability, stem base sequence as well as its helical properties, and are all discussed in the following sections.

3.8.1 A Folded Hairpin, But Not Linear Oligomer, is Recognized by HIV-1 RT RNase H

As a first step to dissect the requirements for molecular recognition between HIV-1 RT and the potent aptamers in this study, we tested the specificity of RNase H towards hairpin secondary structure. In this attempt, we synthesized linear oligomers and assayed them for their ability to inhibit the enzyme's RNase H activity (Table 3.3). Oligomer TRT (Table 3.2, entry 38), 12 nucleotides in length and containing the RNA (UUCG) loop, does not fold into a secondary structure and is not recognized as an aptamer of HIV-1 RT RNase H. The 8-nucleotide linear sequences 5'-UUCGGUCC-3' and 5'-UUCGGUCC-3', lacking the 5'-stem, do not inhibit HIV-1 RT RNase H activity (data not shown). Additionally, [5'-GGAC-3'], [5'-GUCC-3'], or a 1:1 mixture of both does not show any inhibitory activity. In a similar fashion, a mixture of [5'-GGAC-3'] and either 5'-UUCGGUCC-3' or 5'-UUCGGUCC-3' failed to inhibit RNase H activity (data not shown). These studies suggest that the folded secondary structure (helix) is an important criterion for recognition of hairpin substrates by HIV-1 RT RNase H.

Table 3.3: Test of RNase H inhibitory activity of short linear oligomers

Entry	5'-Sequence-3'	Code	Designation	Mixture	IC ₅₀ (μM)
1	GGAC	5'-R	<u>3.1</u>	5'-R + R-3'	n.i
2	GUCC	R-3'	<u>3.2</u>	5'-D + R-3'	n.i.
3	ggac	5'-D	<u>3.3</u>	5'-R + RR	n.i.
4	(UUCG)GUCC	RR	<u>3.4</u>	5'-R + <u>RR</u>	n.i.
5	(<u>UUCG</u>)GUCC	<u>RR</u>	<u>3.5</u>	5'-D + RR	n.i.
6				5'-D + <u>RR</u>	n.i.

n.i. = no inhibition; IC₅₀ is the oligomer concentration required to inhibit 50% RNase H activity of HIV-1 RT. Measurements were repeated twice. Capital letters represent RNA residues; underlined letters are 2',5'-RNA residues (*e.g.* UC = U_{2p5'}C_{2p}); DNA residues are represented by small letters. The appropriate oligomers were mixed at different molar concentrations, heated to 95 °C, and then left to anneal at 4 °C for 72 h. The inhibition assay was run as described in experimental Section 7.6.6.

3.8.2 Comparison of HIV-Inhibitory Activity of 2',5'- Versus 3',5'-RNA Loops

Among the short hairpins (4 base pair stem), the following order of activity was observed: RRR, RRR > RRD, RRR > RRR >>> RDR, DDD. This indicates that the nature of the loop and the stem has a great influence on the inhibitory activity of the hairpins.

The unusual thermodynamic stability displayed by hairpins containing 2',5'- or 3',5'-RNA tetraloops and their uniquely folded structure might indirectly suggest a distinct structure – activity relationship. In order to investigate this and gain insight into the biological properties of our hairpins, we compared the inhibitory potencies of hairpins with identical stems but differing in their loop compositions (D, R, and R). We also tried to correlate the helical conformation of hairpins with their inhibitory activity. **Figures 3.13** and **3.14** represent a three-dimensional plot of the %inhibition of HIV-1 RT RNase H *versus* helical conformation for the various hairpin aptamers. The graph of **Figure 3.13** shows that the activity is more prevalent for hairpins with 2',5'-RNA loops, although the hairpin pool with 3',5'-RNA loops contains the most potent inhibitors (*e.g.*, RRR and RRR). Hairpins with DNA loops regardless of their stem composition are completely inactive. In general, inhibitory potency seems to increase in the direction of A-form helicity, with the hairpins exhibiting B-type conformations showing no inhibitory activity at all. For hairpins with 3',5'-RNA loops, the data suggest a strong correlation between helical conformation and inhibitory activity. Such dependency is less obvious for 2',5'-RNA loops. This change in activity can be due to several factors such as changes in overall hairpin structural-conformational features imposed by the nature of the loop.

Tertiary interactions with the enzyme may require conformational changes in the loop. The flexibility displayed by 2',5'-RNA loops may be the factor that allows such changes and is thus responsible for the enhanced activity observed with various stem hybrids. In sharp contrast, the 3',5'-RNA loop is very rigid and very few examples of UNCG protein recognition have been found to date. This can be further understood when taking into consideration the behavior of other uniquely folded loops. Consider for example GNRA loops. These display a folding pattern similar to that of UNCG, *i.e.*, an extra base pair forms within the single-stranded nucleotides thus leaving a loop of only two unpaired bases that is sterically constrained by the requirements for loop closing.

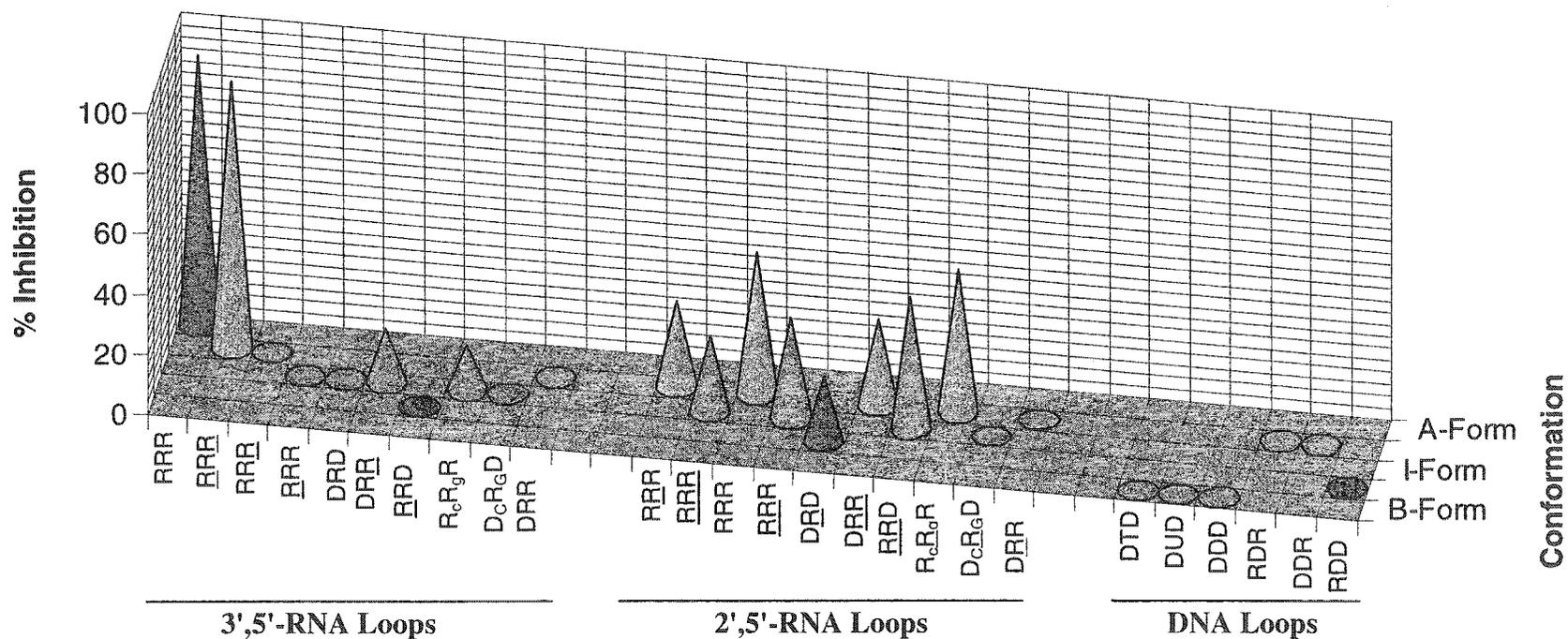


Figure 3.13: 3-D Graph showing % Inhibition {at 40 μ M [aptamer]} versus helical conformation for hairpins with various loops and having the same stem compositions. More pronounced inhibitory activity against HIV-1 RT RNase H is observed for the 2',5'-RNA loop case, though at IC₅₀'s not enough to be the most potent among all library members. Hairpins containing DNA loops, regardless of their stem composition, do not show any degree of inhibition. The color codes correspond to those used in Figure 3.9: A-Form (pink), A-like (cyan), I-Form (green), B-like (brick-red), B-Form (violet).

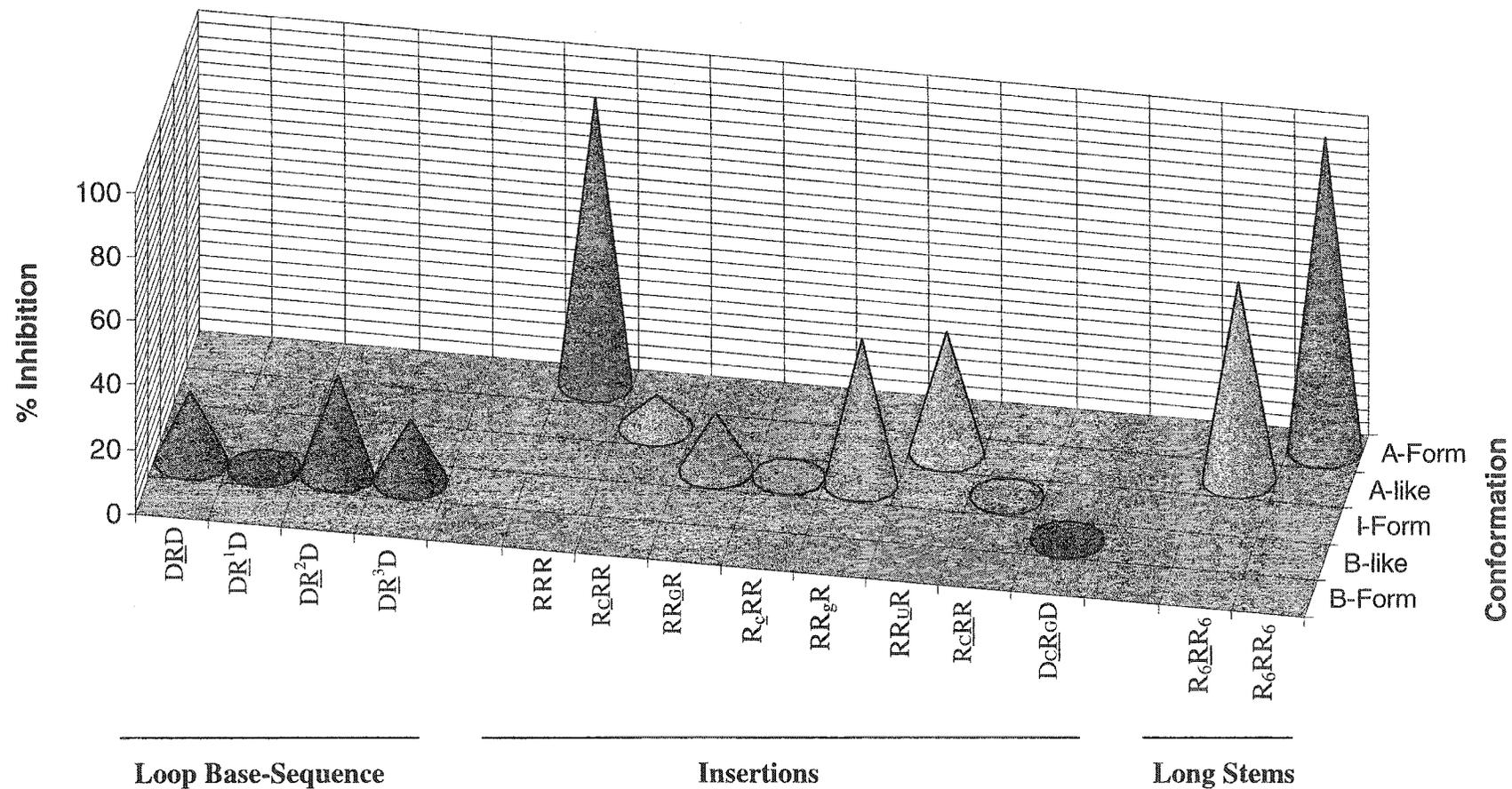


Figure 3.14: 3-D Graph showing % Inhibition {at 40 μ M [aptamer]} versus helical conformation showing effects of loop base sequence, point insertions and stem length. The color codes correspond to those used in **Figure 3.9**: A-Form (pink), A-like (cyan), I-Form (green), B-like (brick-red), B-Form (violet).

However, the major distinction from UNCG loops is a greater degree of flexibility in the sugar conformations of the middle two loop residues. This is why GNRA but not UNCG loops have often been recognized as motifs for tertiary interactions^{46,47} and as sites for protein-binding.⁵⁹⁻⁶¹ Considering the higher degree of flexibility evident (*via* NMR) in the sugar conformations of the 2',5'-RNA loop, the same factor may be operating here.

3.8.3 Hairpin Inhibitory Potency is Dependent on Loop Conformation

The above findings indicate that HIV-1 RT identifies and recognizes the unusually folded loop structure as a signal for binding to its substrates. DNA loops are not extra-stable,^{57,68} and thus are not recognized by the enzyme. Further evidence comes from base mutation experiments conducted within the 2',5'-RNA loop sequence. These show that introducing a single base mutation within the loop [UUCG to UACG] of RNA hairpins completely abolishes inhibitory activity (data not shown). As reported earlier in the NMR studies of chapter II, this mutation is expected to disrupt loop conformation because its position is central to *intraloop* tertiary folding propensity. On the other hand, one may be tempted to argue that the drop in thermal stability might be responsible for the observed loss in inhibitory potency. This is not true in the present case because our findings (discussed in **Section 3.8.4**) demonstrate that there exists no correlation between inhibitory potency and hairpin thermal stability.

Taking into account the DNA hairpins with (UUCG) loops (second class, **Table 3.2**), we observe that replacing the second loop residue U₆ with A [to obtain (UACG)] abolishes inhibitory activity [compare DRD and DR¹D; entries 4 and 5], in agreement with the trend observed above for RNA hairpins. Mutating (UUCG) to (UUUG) does not increase IC₅₀. On the contrary, it enhances the hairpin inhibitory potency [compare DRD and DR²D]. Furthermore, mutating G₈ to U disrupts biological activity [(UUUG) *versus* (UUUU)]. These studies obviously identify the 2',5'-RNA loop as an important recognition motif by virtue of its unusually folded structure. Disruption of the loop sequence (therefore conformation) is deleterious for hairpin recognition. Other trends that further support this notion can be detected among members of the fourth class containing point insertions at various positions in the hairpin (**Table 3.2**). For instance, replacement of one of the loop-closing base pairs (C₄) with deoxy residue abolishes activity [R_cRR].

Replacement of the other residue G₉ compromises activity though not to the same extent as that of C₄ (Figure 3.14). Depending on the nature of the modification, mutations introduced near the loop region affect its tertiary folding pattern. This might explain the dramatic loss in inhibitory potency of R_CRR relative to RRR and of R_CRR compared to RRR.

All in all, the above results unambiguously point towards the uniquely folded loop structure as a recognition motif that plays a key role in hairpin/(HIV-1 RT) interaction.

3.8.4 Hairpin Inhibitory Activity is Independent of Thermal Stability

A three-dimensional plot of % inhibition *versus* melting temperature for various hairpin oligomers was constructed (Figure 3.15). It is clear that T_m and % inhibition do not go in parallel. For example, hairpin R_CR_gR is more thermally stable than R_CR_gR (ΔT_m *ca.* 2.4 °C), yet the latter has a lower IC₅₀ value. Hairpins RRUR, R_CRR and RRGR are of almost the same order of thermal stability; however, RRUR shows good inhibitory potency while R_CRR and RRGR do not. RRR is significantly more thermally stable than RRR, yet it is less biologically active. In some cases, inhibitory potency goes in parallel with melting temperature (for example, DRR *versus* DRR and RRR *versus* RRR), but this does not imply a general trend.

This independence of thermal stability displayed by hairpin inhibitory activity is not surprising since the same trend has been observed in other tertiary interactions. For example, thermodynamic stability is not important in the folding of the group I self-splicing intron core as evident by changing the conserved third and fourth nucleotide residues in GAAA (A to G), or when the loop is replaced by the more rigid (UUCG) loop.⁴⁷ Rather, it is the ability of the loop nucleotides to participate in tertiary contacts that is important.⁴⁷ Another example where loop sequence-specificity but not thermodynamic stability is responsible for tertiary interaction lies in the folding of group I intron ribozymes (L9 loop).⁴⁶ All of these examples, among others, support the results observed here for the nature of the inhibitory activity of hairpin aptamers. They further confirm the above hypothesis about the loop conformation being an important recognition factor in interaction with HIV-1 RT.

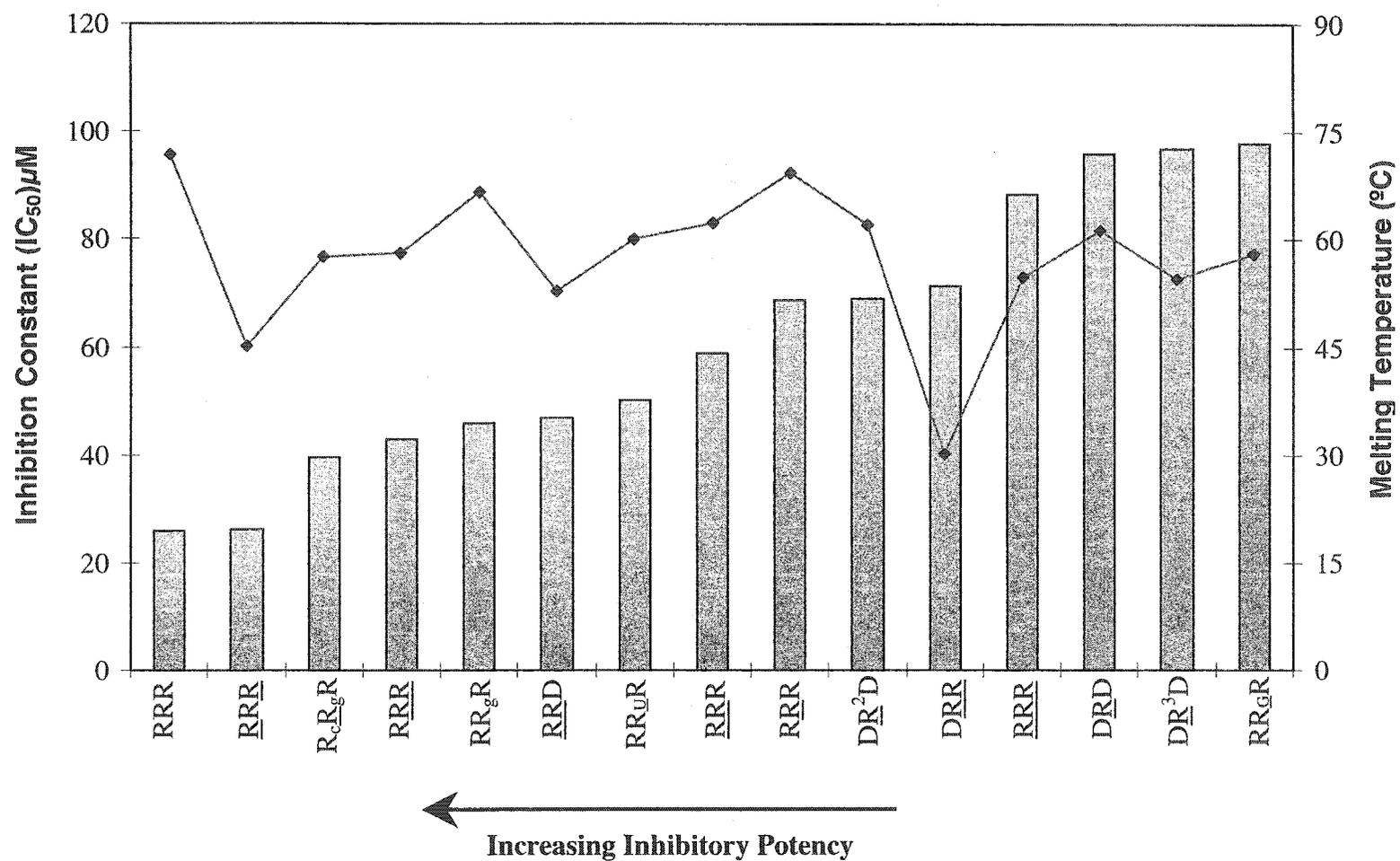


Figure 3.15: Variation of inhibitory constant (IC_{50}) with melting temperature. The hairpin aptamers are shown in order of decreasing potency (increasing IC_{50}). Bars represent IC_{50} plot, while the curve line represents T_m plot.

In summary, these findings suggest a structure-activity relationship that is evident among members of the hairpin library. The unusually folded loop structure is an important determinant governing inhibition of HIV-1 RT RNase H activity, along with other factors that come into play such as stem conformation. In this system, a minimum requirement to attain maximum inhibitory activity is a folded A-form stem bridged by an extra-stable (UUCG) loop.

3.9 GEL –SHIFT MOBILITY ASSAY

In order to detect complex formation between the hairpin aptamer and HIV-1 RT, we conducted gel-shift binding assays under native conditions. Generally, if an enzyme-substrate complex is formed, it exhibits a slower electrophoretic mobility relative to the unbound substrate. The binding of HIV-1 RT to 5'-[³²P]-labeled hairpin was monitored in the presence of increasing enzyme concentration. We show that as the concentration of RT is increased, the amount of R₆RR₆ bound to the enzyme increases in agreement with a decrease in the amount of free hairpin (**Figure 3.16**). These results further support the view that the hairpin oligomer is acting as an aptamer of the enzyme. Hence, inhibition is due to direct binding/interaction between the hairpin and the enzyme, and does not result from other non-specific mechanisms of action. Whether this binding occurs at the RNase H or polymerase domain of HIV-1 RT is elaborated later in this chapter (**Section 3.11**).

3.10 INHIBITION OF RNA- AND DNA- DEPENDENT DNA POLYMERASE ACTIVITY OF HIV-1 RT

All library compounds were screened for their ability to inhibit DNA synthesis catalyzed by HIV-1 RT. A (RNA-template)-primer was used as the natural substrate in RNA-dependent DNA polymerization assays, while a (DNA-template)-primer complex was used in the DNA-dependent DNA polymerization assays. All template and primer sequences were synthesized using standard phosphoramidite solid phase protocols (see experimental **Section 7.2**). The template sequence, from the PBS region of HIV-1, is: 5'-AUC UCU AGC AGA GGC GCC CGA ACA GGG ACA-3', while the primer is an 18-mer DNA oligomer of the following sequence: 5'- TGT CCC TGT TCG GGC GCC -3'.

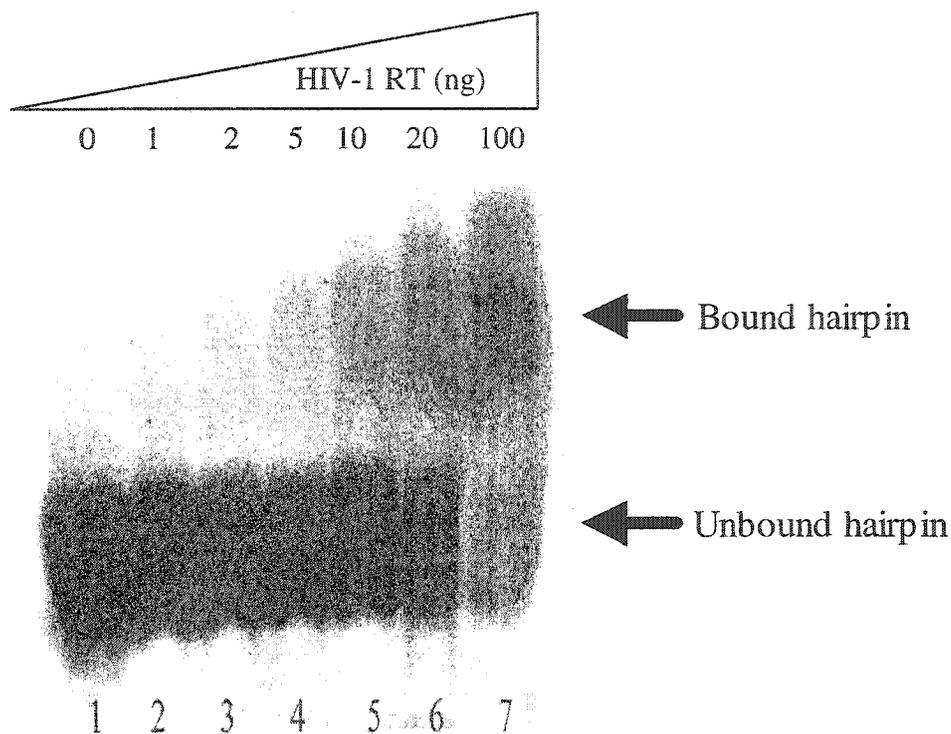


Figure 3.16: Gel-shift binding assay to determine formation of [HIV-1 RT]:[Hairpin] complex. Binding of HIV-1 RT to [^{32}P]-labeled- R_6RR_6 was carried out in the presence of increasing concentration of the p66/p51 RT. The lanes in the scanned autoradiogram of a 6% acrylamide gel are as follows: lane 1, control lane with no enzyme present; lanes 2-7, binding of hairpin to HIV-1 RT [at amount of 1 ng (lane 2), 2 ng (lane 3), 5 ng (lane 4), 10 ng (lane 5), 20 ng (lane 6) 100 ng (lane 7)].

The DNA primer is [^{32}P]-labeled at its 5'-end and then annealed to the template target. In the reaction assay, HIV-1 RT recognizes the formed template-primer duplex, and synthesizes a complementary DNA strand in the presence of various nucleoside triphosphates (**Figure 3.17**). The newly synthesized DNA strand exhibits a slower electrophoretic mobility relative to the primer. Small drug molecules such as NNRTIs can halt DNA synthesis catalyzed by HIV-1 RT.

Our results show that none of the library hairpin members were able to inhibit, at various oligomer concentrations, the activity of the RNA-dependent or DNA-dependent DNA polymerase of HIV-1 RT. Boosting hairpin concentration to 80 μM did not promote inhibition. **Figure 3.18** shows that in the presence of various hairpin oligomers, the gel

band corresponding to the newly synthesized DNA strand does not disappear or even decrease in intensity. Under the same assay conditions, the non-nucleoside inhibitor *N*-(4-*tert*-butylbenzoyl)2-hydroxy-1-naphtaldehyde hydrazone¹⁹⁰ (positive control; **Figure 3.2**) proved to efficiently arrest DNA synthesis catalyzed by HIV-1 RT (data not shown).

All in all, the above results show that R₆RR₆ [IC₅₀ = 7.8 μM] and other hairpins [RRR, RRR, R₆RR₆; IC₅₀'s ~ 26-30 μM] are indeed aptamers that function as inhibitors targeted specifically to the RNase H domain of HIV-1 reverse transcriptase without affecting its polymerase activity.

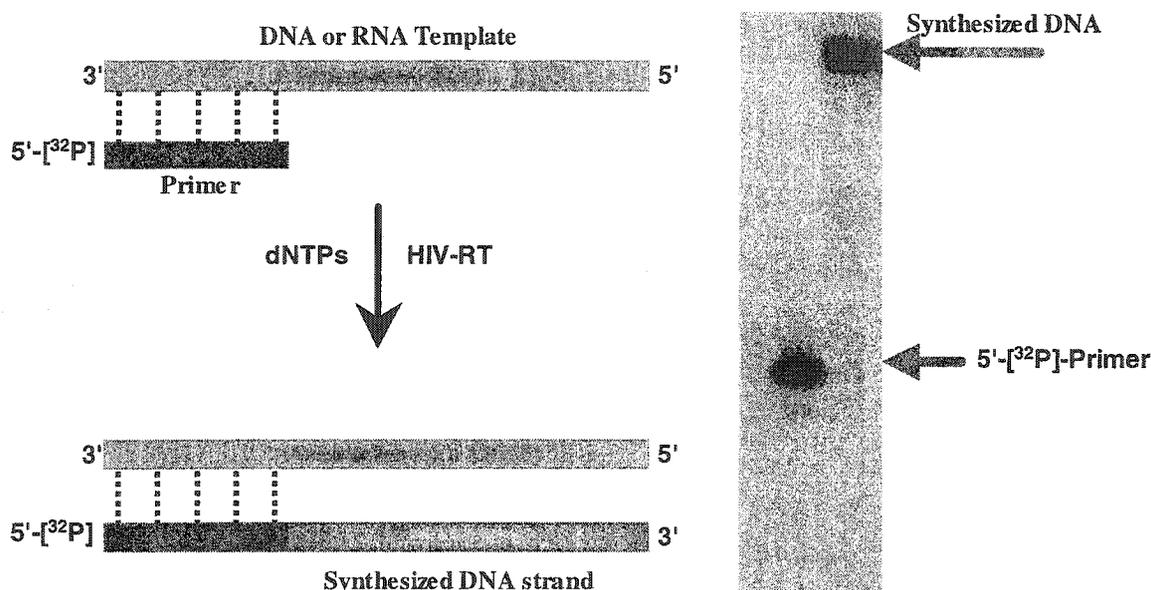


Figure 3.17: DNA synthesis catalyzed by HIV-RT DNA- or RNA-dependent DNA polymerase activity. The 5'-[³²P]-DNA primer hybridizes to a complementary DNA or RNA template, and then HIV-RT catalyzes the synthesis of the DNA strand (5'→3' direction) in the presence of 2'-deoxynucleoside 5'-triphosphates (dNTPs). The newly formed DNA runs slower than the primer on 20% denaturing polyacrylamide gel.

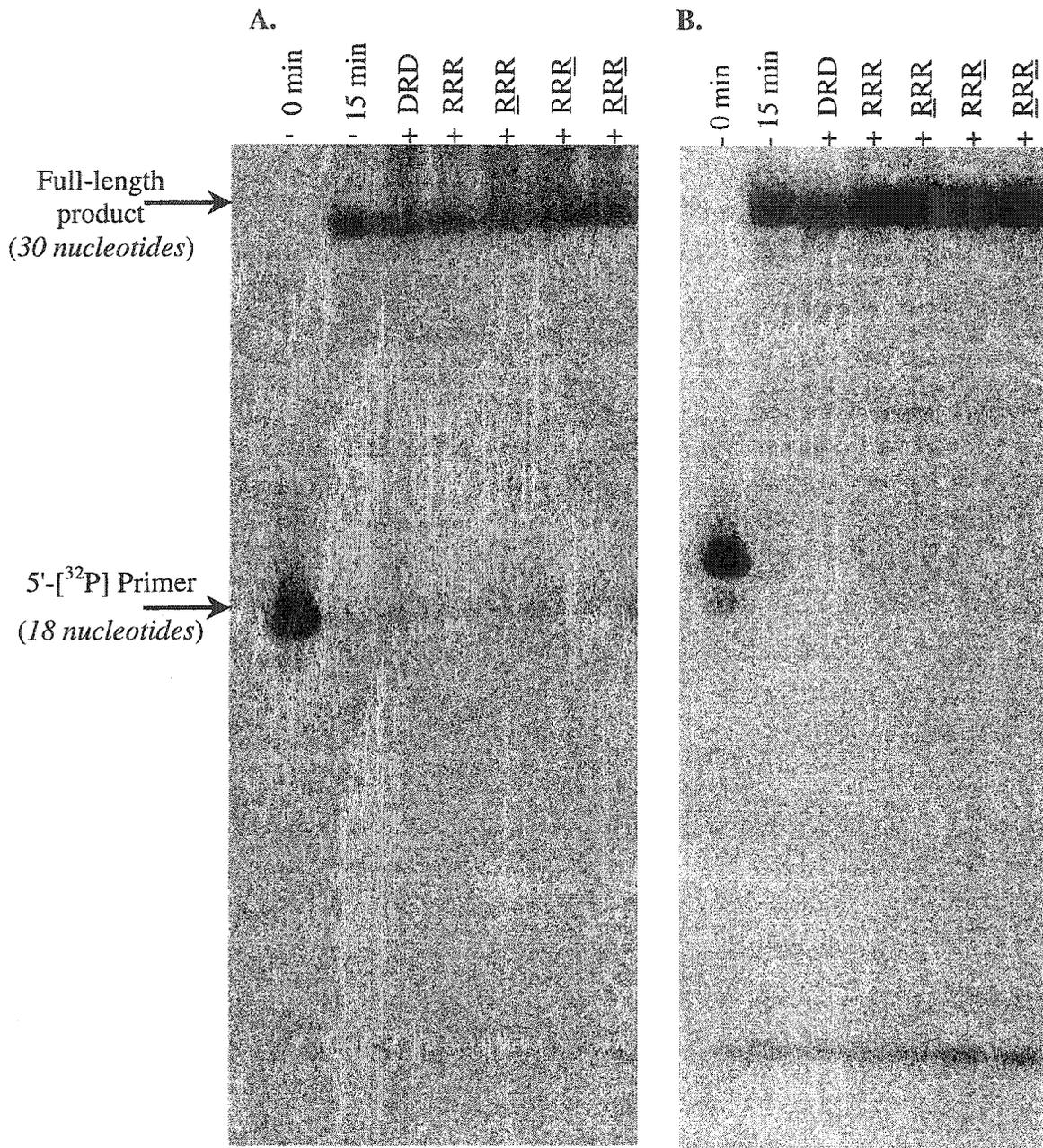


Figure 3.18: Inhibition of DNA synthesis catalyzed by (A) DNA-dependent DNA polymerase (B) RNA-dependent DNA polymerase activities of HIV-1 RT. Unlabeled hairpin [80 μ M] was pre-incubated with RT for 20 min at room temperature prior to initiating the polymerization reaction by adding the proper DNA or RNA template hybridized to a 5'-[32 P]-labeled primer in the presence of dNTPs (see experimental Section 7.6.6). The various lanes show full-length DNA products synthesized after 15 minutes in the absence (-) and presence (+) of unlabeled hairpins.

3.11 UV CROSS-LINKING STUDY

To confirm that binding of R₆RR₆ to HIV-1 RT is selective towards the RNase H domain, we conducted UV cross-linking studies between this hairpin (most potent from library) and either HIV-1 RT (p66/p51 heterodimer with both RNase H and polymerase domains) or p51/p51 homodimer (only polymerase domain without RNase H domain). Ultraviolet light can induce nucleic acid-protein photocrosslinks, thus providing a reliable method for detecting the binding domain within the protein.²⁴ Our results, as detected from denaturing gel electrophoresis, demonstrate that hairpin R₆RR₆ for example does not bind to p51/p51 homodimer, but binds with high affinity to p66/p51 HIV-1 RT heterodimer (**Figure 3.19**). These findings clearly establish that R₆RR₆ does not bind the DNA polymerase domain of HIV-1 RT but binds selectively to the RNase H domain. This is further supported by the results described in **Section 3.10** which show that a series of hairpins do not inhibit the DNA polymerase activity of HIV-1 RT even at a concentration of 80 μ M. Hence, hairpin binding to RT is mediated by RNase H, and whether this binding is competitive (direct binding with the RNase H substrate active site) or allosteric (binding to a non-substrate active site) has yet to be determined.

3.12 STABILITY OF HAIRPIN APTAMERS TOWARDS NUCLEASES

3.12.1 Stability Towards Blood Serum

In order to evaluate the stability of hairpins in biological media, the nuclease resistance of some potent hairpin inhibitors was assessed. The hairpins were treated with rabbit reticulocyte lysate [at 0.5 X dilution, *i.e.*, equal volume of water added to extract] containing various sources of *endo*- and *exo*- nucleases. Under these conditions, the hairpins remained fully intact and did not show any degree of degradation even after 6 h (data not shown). It was not until after a nine-hour incubation period that a slight degradation of the various hairpin inhibitors was detected. Under the same conditions, a control linear oligomer of the same length was more than 50% degraded (data not shown). This indicates that the hairpin molecules are highly resistant towards nucleases present in serum. This prompted us to further assess their biological stability in reticulocyte lysate but this time without any dilution and for a longer incubation time.

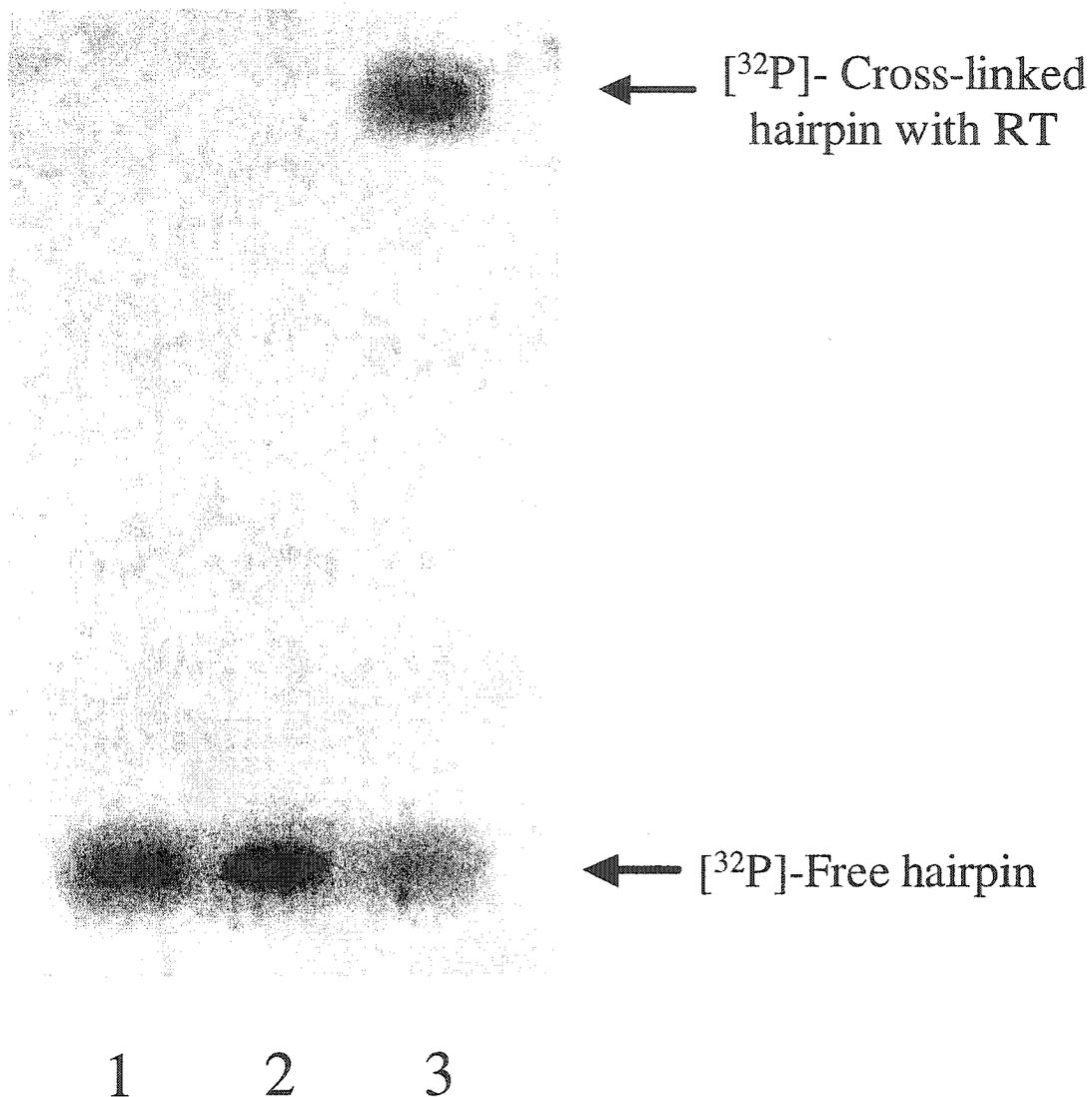


Figure 3.19: UV Cross-linking experiment of hairpin R₆RR₆ with HIV-1 RT. [³²P]-labeled hairpin was incubated with RT and then UV-irradiated on ice for 15 min (see experimental). The complexes were analyzed on a 12.5 % SDS-PAGE gel; lane 1, hairpin R₆RR₆ alone; lane 2, p51/p51 HIV-1 RT homodimer (only polymerase domain - No RNase H domain) and hairpin R₆RR₆; lane 3, p66/p51 HIV-1 RT heterodimer (with both RNase H and polymerase domains) and hairpin R₆RR₆.

The % of remaining substrates was compared among various hairpins after being suspended in blood serum for 18 hours at 0.5 X and 1 X dilution (**Figure 3.20**). All hairpins show discrete biostability in blood serum compared to a control single-stranded oligonucleotide. We found that the presence of 2',5'-RNA in the loop induces increased nuclease resistance of various hairpin aptamers as evident by % remaining substrate for example of RRD versus RRD (**Figure 3.20**). Also, hairpin RRR [with a 2',5'-RNA loop] shows increased nuclease resistance towards blood serum compared to RRR and RDR [with 3',5'-RNA and DNA loops respectively]. This shows that the 2',5'-RNA loop not only increases the thermal stability of hairpins, but also offers increased resistance towards nucleases in biological media. It is important to note that the hairpin nuclease resistance does not parallel its thermal stability. For instance, hairpins RRR and RRR are of approximately the same order of thermal stability, yet RRR is more nuclease resistant (**Figure 3.20**).

3.12.2 Stability Towards Nuclease P1

The stability of various hairpins towards Nuclease P1 was also assessed using a combination of UV spectroscopy and gel electrophoresis techniques. Nuclease P1 is an endonuclease that cleaves specifically single stranded DNA and RNA, although it can hydrolyze double stranded duplexes albeit slowly.^{200,201} The hairpins (6 μ M) were suspended in appropriate buffer [30mM NaOAc, pH = 5.3] and then the reaction was initiated by the addition of Nuclease P1. The increase in UV absorbance with time was monitored at 260 nm. After 150-180 minutes, the absorbance of all hairpins leveled off indicating that no more degradation was taking place (**Figure 3.21**). The relative half-life time of various hairpins was calculated relative to the hairpin showing least resistance, *i.e.*, hairpin DD'D among DNA hairpins and RDR among RNA hairpins] (**Table 3.4**). Again similar to the trend seen when suspended in reticulocyte lysate, the hairpins containing a 2',5'-RNA loop show increased resistance towards nuclease P1 relative to hairpins with 3',5'-RNA loops.

% Remaining Hairpin

Hairpin Aptamer	0.5 X	1 X	T_m (°C)
RRR	36	27	71.8
<u>RRR</u>	48	33	69.3
RDR	26	15	63.4
<u>RRD</u>	25	16	48.1
<u>RRD</u>	91	78	52.8
<u>RRR</u>	14	9	45.2
<u>RRR</u>	16	12	54.8
TRT	7	7	-

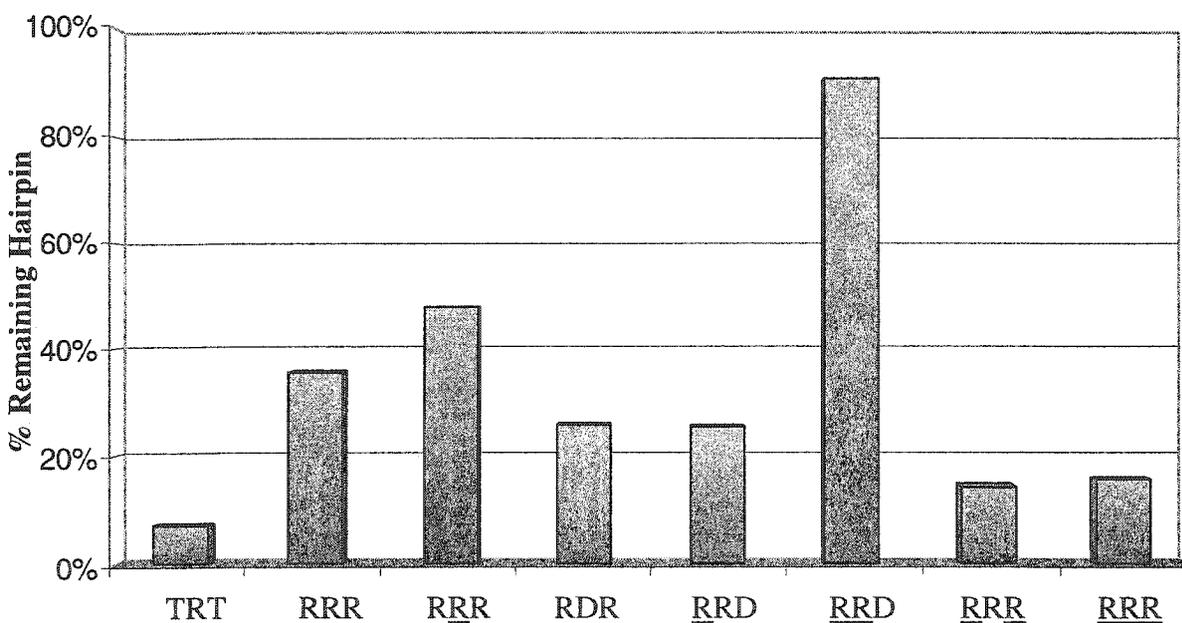
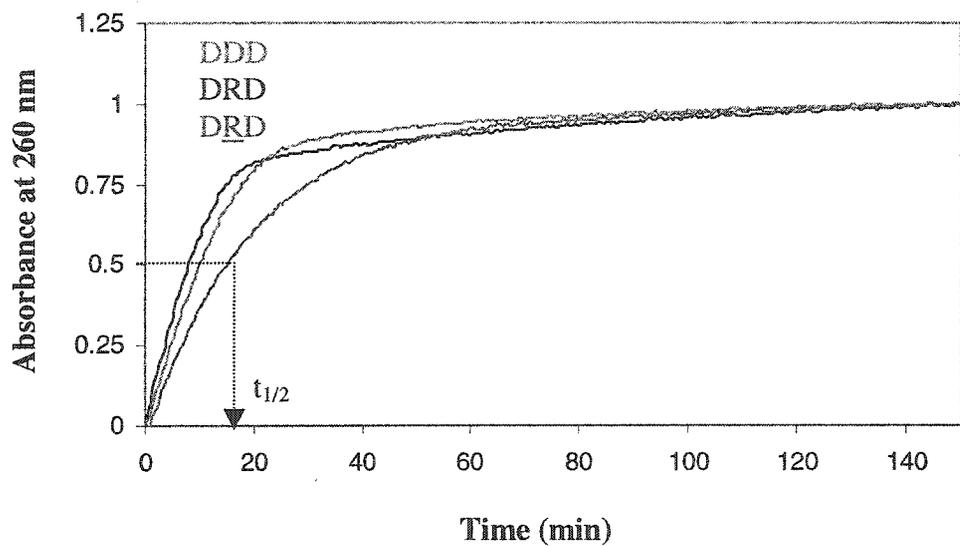


Figure 3.20: %Remaining hairpin after incubation with 0.5 X reticulocyte lysate for 18 hours at 37 °C. The table shows the different values at 0.5 X and 1 X dilution conditions. The 2',5'-RNA loop makes the hairpins more resistant in biological media against blood serum nucleases. TRT does not form any hairpin structure and serves as a control. Equal amounts of 5'-[³²P]-labeled hairpins were prepared in separate tubes containing 60 mM Tris (pH 7.8), 60 mM KCl, 5 mM MgCl₂, and 5 mM K₂HPO₄. After addition of rabbit reticulocytes lysate, the reaction mixtures were incubated at 37 °C for 18 h. Analysis was done on 16 % denaturing polyacrylamide gels.

A.



B.

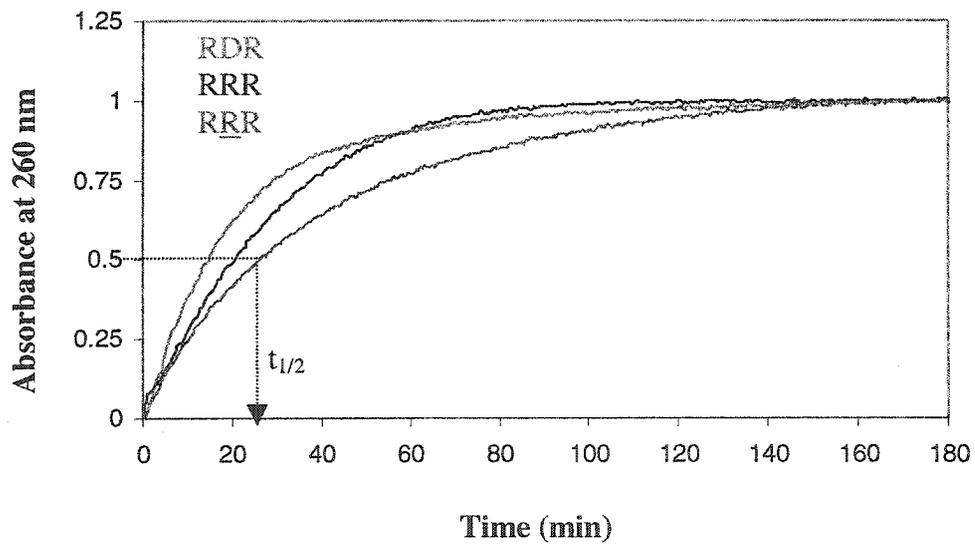


Figure 3.21: Plots of absorbance (at 260 nm) *versus* time of exposure of hairpins to Nuclease P1. The time necessary to cause 50% degradation of the hairpin structure ($t_{1/2}$) was calculated at absorbance = 0.5. The assay was carried out as described in **Section 7.6.10**.

Table 3.4: Relative half-life times of representative hairpins towards Nuclease P1 digestion

Entry	Code	Hairpin	$t_{1/2}$ (min)	T_{rel}
1	<u>2.11</u>	DD'D	8.0	1.0
2	<u>2.13</u>	DDD	9.8	1.2
3	<u>2.14</u>	DRD	8.0	1.0
4	<u>2.15</u>	DRD	16.1	2.0
5	<u>2.1</u>	RRR	21.2	1.4
6	<u>2.2</u>	RRR	27.2	1.9
7	<u>2.3</u>	RDR	14.7	1.0

The hairpins [6 μ M] were incubated with Nuclease P1 (0.001 Units) in 30 mM NaOAc buffer, pH = 5.3 and their UV absorbance monitored at 260 nm for 150 min. The half-life time ($t_{1/2}$) was calculated from the obtained curve at absorbance = 0.5. T_{rel} is the relative half-life time value compared to the fastest degrading (least nuclease stable) hairpin oligomer in each set [DD'D in the first set and RDR in the second set]. DD'D is a control of the following sequence: 5'-ggacttcggtcc-3'. Values were obtained from three different measurements. Errors in values of $t_{1/2}$ are within ± 1 min.

In conjunction with UV absorbance spectroscopy, nuclease resistance was evaluated using gel electrophoresis. The former technique gives a qualitative idea about the relative hairpin stabilities, but does not yield specific details about the degradation pattern. Thus, we complemented the experiments with gel electrophoresis. Accordingly, the same hairpins under investigation were 5'-[32 P] labeled, and their degradation pattern was monitored by denaturing gel electrophoresis. The results confirm the increased nuclease resistance induced by 2',5'-RNA substitution in the loop, in agreement with the UV spectroscopy results. However, a closer look at the degradation pattern reveals an interesting distribution of products. Nuclease P1 (endonuclease) seems to cut only at specific positions in the hairpin sequence, namely in the middle of the loop [at 6th nucleotide position] and the loop closing base pair [4th or 3rd nucleotide position] (**Figure 3.22**). No other degradation products (or “ladder” of products) were observed as would be expected if strand dissociation occurred after the first cleavage reaction.

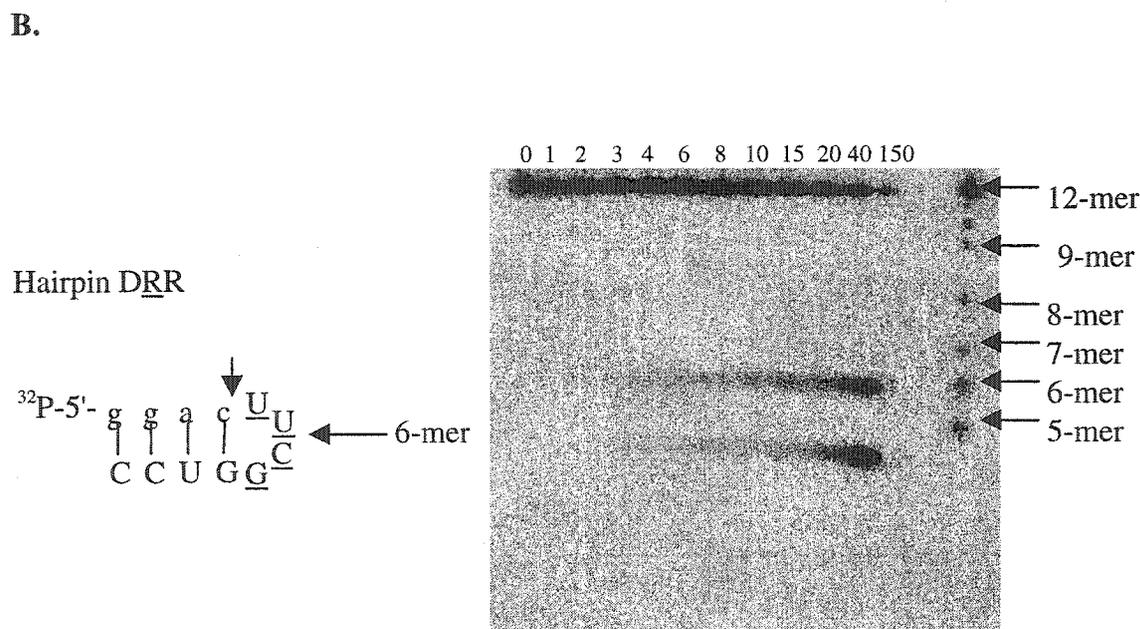
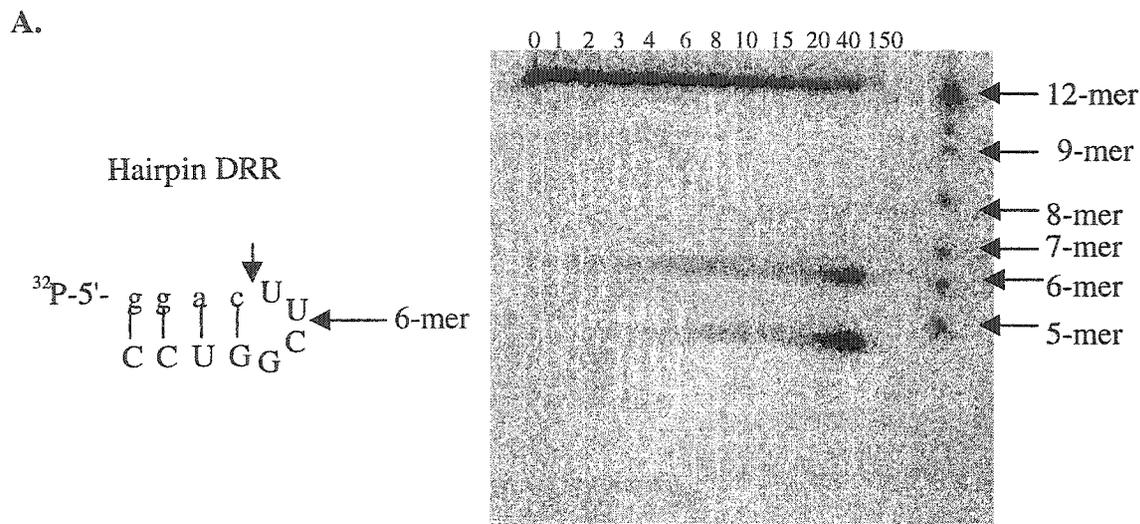


Figure 3.22: Denaturing polyacrylamide gel electrophoresis showing a representative digestion pattern by Nuclease P1. A sequencing ladder of the hairpin is included as a marker on the right gel lane. The [^{32}P]-labeled hairpin was dissolved in 30 mM NaOAc buffer [pH = 5.30 \pm 0.02] to a final concentration of 6 μM . The reaction was initiated by addition of Nuclease P1 (0.0001 Units) and was incubated at 37 $^{\circ}\text{C}$ for 150 min. 8 μl aliquots were taken out at various time intervals, and were loaded onto a 16 % (w/v) polyacrylamide denaturing gel (7 M urea) and then exposed on an X-ray film. The individual lanes were quantitated using the software UN-SCAN-IT by measuring the amount of remaining *versus* degraded hairpin.

This raised the interesting possibility that the oligonucleotide might have maintained its folded structure after the first enzymatic cut was introduced or that the two fragments produced might have re-associated to form the hairpin structure. To test this intriguing possibility, we synthesized various oligomers, 4 and 8 nucleotide units in length, corresponding to the observed degradation products (Table 3.5). Incubation of the appropriate fragments [*e.g.*, GGAC + (UUCG)GUCC] at various concentrations and under different buffer conditions did not lead to duplex formation, thus eliminating the possibility of hairpin reformation after enzyme hydrolysis of the scissile phosphodiester linkage.

Table 3.5: Thermally-induced melting data for pseudo hairpin formation

Entry	5'-Sequence-3'	Code	Designation	Mixture	T_m (°C)
1	GGAC	5'-R	<u>3.1</u>	5'-R + R-3'	n.d.
2	GUCC	R-3'	<u>3.2</u>	5'-D + R-3'	n.d.
3	ggac	5'-D	<u>3.3</u>	5'-R + RR	n.d.
4	(UUCG)GUCC	RR	<u>3.4</u>	5'-R + <u>RR</u>	n.d.
5	(UUCG)GUCC	<u>RR</u>	<u>3.5</u>	5'-D + RR	n.d.
				5'-D + <u>RR</u>	n.d.

n.d. = not detected. The base sequences of the synthetic oligonucleotides under investigation are shown in the left-hand section of the table. No melting transition was observed under various oligomer concentration and buffer conditions. The final oligomer concentrations were varied {5–30 μ M} and the possibility of their association was tested at both low and high ionic strength conditions. The buffer was comprised of 0.01 M Na₂HPO₄, 0.1 mM Na₂EDTA buffer, pH 7.00 \pm 0.02 and contained either no NaCl (low ionic strength) or 1M NaCl (high ionic strength).

Another reason as to why we conducted the above studies was to test whether tagging an extra stable loop at the terminus of a 4-nucleotide single-strand can induce its duplex formation with a complementary oligomer via electrostatic neighboring group assistance (in this case, loop folding). Such a 4 base pair duplex would never form on its own in solution. Unfortunately, we did not investigate in this study the possibility of other combinations (different than those mentioned in Table 3.5) that might improve binding such as adding a C-G closing base pair to promote loop folding [*i.e.*, GGA +

C(UUCG)GUCC], moving to longer stems [*e.g.*, $X_nGGA + C(UUCG)GUCCY_n$, where X & Y are base complementary and n is the number of residues added], or even adding another base pair to stabilize the loop [*i.e.*, $X_nGG + AC(UUCG)GUCCY_n$] and so on.

In summary, the findings presented in this section prove that the hairpin aptamers are highly resistant to nucleases. This might be due to their highly stacked and relatively inaccessible structures, and thus suggests that they can be employed as tools to stabilize antisense oligonucleotides or mRNAs.²⁰²⁻²⁰⁵ Antisense oligonucleotide degradation predominantly occurs via 3'-5' exonuclease activity. Hence, tagging a highly stable loop structure to the 3'-end of the antisense oligomer would protect it from further exonuclease cleavage activity. These hairpin loops can be used in conjunction with other strategies used for stabilizing antisense oligonucleotides.

3.13 CONCLUSIONS

To our knowledge, there have been very few reports in the literature of selective oligonucleotide-based inhibitors of RNase H activity of HIV-1 RT. Most inhibitors are aimed at its DNA polymerase activity. In this study, we discovered the first examples of short oligonucleotide hairpins composed of only 4 base pairs that selectively inhibit RNase H of HIV-1 RT without affecting the enzyme's polymerase activity. These hairpins evolved from a nucleic acid library constructed in a *diversity-generating* combinatorial approach. By varying the sugar content (DNA vs. RNA), phosphodiester linkage (2',5'- vs. 3',5'-) and stem length, we were able to introduce various degrees of structural and conformational diversity within the various hairpin members in the library.

Forty-five hairpin molecules were synthesized and consequently screened, in collaboration with Dr. Kyung-Lyum Min from the Damha Laboratory, for their ability to inhibit the activity of HIV-1 RT RNase H in degrading the RNA strand in RNA:DNA hybrids, a key step in the retroviral replication cycle. Four hairpin aptamers were identified as potent inhibitors in the low μM range. The hairpin R_6RR_6 is the most potent inhibitor with an $IC_{50} = 7.8 \mu\text{M}$, while RRR , \underline{RRR} and $R_6\underline{RR}_6$ displayed IC_{50} 's in the ~ 26 - $30 \mu\text{M}$ range. To shed light into the mechanism of inhibition, we conducted gel shift assays which offered evidence of a complex formation between the hairpin inhibitor and the RT enzyme. Furthermore, the hairpin oligomers did not interfere with either RNA-

dependent or DNA- dependent DNA synthesis catalyzed by HIV-1 RT. This was further supported by UV cross-linking studies that showed that there was no association between the most potent hairpin aptamer and the polymerase p51/p51 homodimer of HIV-1 RT lacking the RNase H domain. Additionally, the potent hairpins had no effect on cellular RNase H (Human RNase H). In the end, the hairpins were evaluated for their ability to resist degradation by cellular nucleases and they proved to be highly nuclease resistant.

All in all, the above results suggest that these mini hairpins can potentially be used as selective tools for probing viral gene expression by shutting down selectively RNase H activity without interfering with the enzyme's DNA polymerase activity. In addition, they can act as probes for distinguishing between RNase H-induced and hybridization arrest (translation arrest) antisense activity. Not only that, they have great potential as therapeutic agents for inhibiting HIV-replication. These can potentially be used in conjunction with other antiretroviral drugs (“combination therapy”) in order to interfere with HIV-1 life cycle and its ability to reproduce.

At the moment, docking experiments as well as mechanistic kinetic studies are being pursued in order to determine the mode of aptamer inhibition (competitive *versus* non-competitive) and the molecular interactions involved in recognition. Cell culture studies are also in progress in order to assess the HIV inhibitory activity of these hairpins.

CHAPTER IV: PROBING THE BINDING SPECIFICITY AND CLEAVAGE MECHANISM OF YEAST RNASE III (RNT1P)

4.1 BACKGROUND

RNA interference (RNAi) is a key gene silencing mechanism in which eukaryotic cells, confronted with double-stranded RNA, are triggered to destroy their own mRNAs that bear sequence homology to the double strand.^{70,72,206} Processing of premature RNA is one of the most indispensable events regulating gene expression and cellular growth.²⁰⁷ During protein manufacturing, nascent RNA is transformed to mature RNA inside the cell. Among various enzymes involved in this process is a ribonuclease (RNase) III, a double-stranded RNA (dsRNA)-specific enzyme that cleaves the phosphodiester linkages at various positions on each side of the RNA duplex.¹⁵⁴ This enzyme has been implicated in the processing of pre-ribosomal RNA (pre-rRNA), pre-transfer RNA (pre-tRNA), and many messenger RNAs (mRNAs) in bacteria.^{208,209} Three homologues of RNase III have so far been identified in eukaryotic cells, namely *Saccharomyces cerevisiae* (Rnt1p)²¹⁰⁻²¹², *Schizosaccharomyces pombe* (Pac1),²¹³⁻²¹⁶ and human RNase III.²¹⁷ Other homologues have also been reported in the mouse and worm.^{215,218} Both Pac1 and Rnt1p have been shown to play a chief role in processing other cellular RNAs *in vitro* and *in vivo*. Examples of these include pre-rRNA,^{210,219} small nuclear RNAs (snRNAs),^{216,220,221} and small nucleolar RNAs (snoRNAs),^{218,222,223} all of which imply a critical association with cellular division and growth.

Members of the RNase III family distinctly recognize their duplex RNA substrates and cleave them at specific sites. They are found in all kingdoms with the exception of archaeobacteria that possess an analogous enzyme with different structure and substrate specificity.²²⁴ All members have in common two functionally independent subdomains: a C-terminal dsRNA-binding domain (dsRBD) and a nuclease domain.^{210,211,225,226} Structural studies have shown that the dsRBD within the RNase III family of enzymes resembles that of several other dsRNA binding proteins such as PKR [an RNA-dependent protein kinase that mediates cellular immune response] and the RNA localization protein Staufin,²²⁷⁻²³¹ implying a highly conserved structure among the family of dsRNA binding proteins (dsRBPs). The nuclease domain, on the other hand, is

mainly associated with RNA cleavage and thus plays a central role in catalysis.^{211,232} Many mutations have been introduced within this subdomain in an attempt to help elucidate the molecular requirements for RNA cleavage.^{154,226} The structures of the *Escherichia coli* RNase III dsRBD²²⁵ and nuclease domain²³³ have been determined recently. Its dsRBD folds into a compact structure that is conserved in all dsRBPs examined to date.^{229,230} Its nuclease domain forms a dimer with a “Ball and Socket” junction that accommodates the dsRNA helix.²³³ The presence of two catalytically active centers at each end of the dimer explains cleavage of duplex RNA by the enzyme.

Despite sharing the above common features, members of the RNase III family exhibit structural differences; these are emphasized in the disparity among their substrate recognition requirements. An example of this would be *Saccharomyces cerevisiae* RNase III (Rnt1p) which shares classical structural features of the bacterial enzyme but not its mode of substrate selection.^{210,234} Its dsRBD bears 25% identity to *E.coli* RNase III and 31% identity to fission yeast *pac1* but exclusively displays a 32-amino acid extension at the C-terminus. Its nuclease domain shares the same size and charged amino acid clusters with bacterial RNase III and *pac1*. However unlike bacterial RNase III, Rnt1p additionally possesses a characteristic N-terminal domain that constitutes 36% of the total Rnt1 protein (**Figure 4.1 A**).^{211,212} Rnt1p functions as a dimer in solution. A pioneering study by Abou Elela and coworkers shows that the N-terminal domain bears no homology to the *pac1* N-terminal domain.²¹¹ Its function is mainly as an auxiliary entity that ensures efficient RNA cleavage by mediating both inter- and intramolecular interactions (**Figure 4.1 B**). Deletion of the N-terminal domain results in reduction of level of processing of the 25S rRNA 3'-end by *ca.* 30% *in vivo* and slows growth by 35-40%.²¹¹ On the other hand, the dsRBD is essential and sufficient for RNA binding, while the nuclease domain is only required for efficient RNA cleavage.

Although the individual functions of each subdomain have been determined, the intrinsic mechanism of RNA recognition and cleavage by the RNase III family still remains vague and hence needs further investigation.

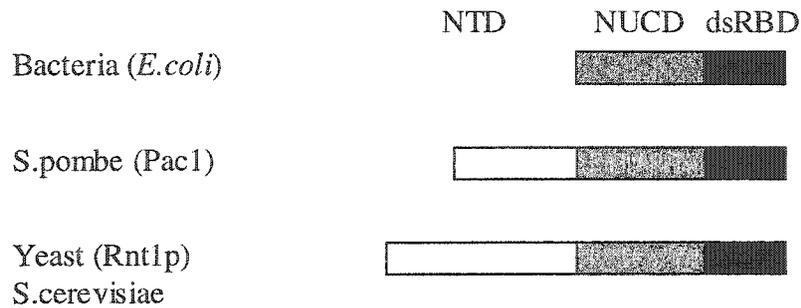


Figure 4.1 A: Schematic representation comparing the functional domains between bacteria and yeast RNases III. All members of the RNase III family possess a double-stranded RNA binding domain (dsRNA) and a nuclease domain (NUCD). The N-terminal domain (NTD) is exclusive to eukaryotic members of the RNase III family.

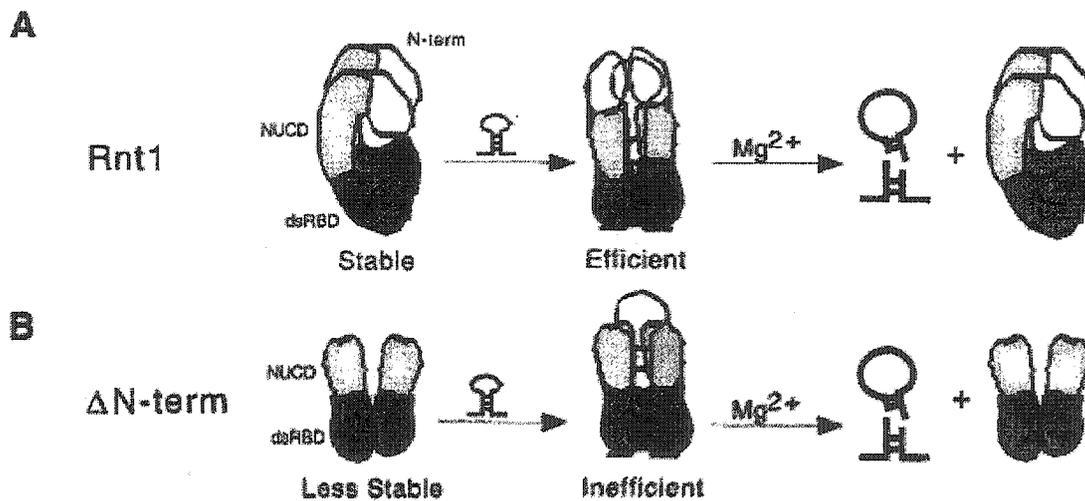


Figure 4.1 B: Proposed hypothetical model for the role of N-terminal domain in Rnt1p cleavage efficiency. Rnt1p without the N-terminal domain is represented by Δ N-term. Intermolecular interaction mediated by the N-terminal domain stabilizes the protein complex on the RNA. Divalent cation (Mg^{2+}) is needed to effect cleavage. Adapted from Lamontagne, B.; Tremblay, A.; Abou Elela, S. *Mol. Cell. Biol.* **2000**, *20*, 1104-1115.

4.2 SUBSTRATE RECOGNITION AND CLEAVAGE BY RNT1P

As mentioned earlier, Rnt1p is involved in the processing of the 25S rRNA and many other snRNAs and snoRNAs.^{210,219-223} A survey of the natural substrates recognized by Rnt1p reveals the recurrence of a terminal tetraloop motif with the general consensus AGNN (where N represents any nucleotide base residue) that directs cleavage 14-16 base

pairs away.^{221,235,236} Introducing base mutations within this tetraloop can affect the cleavage rate as well as the binding affinity to Rnt1p, suggesting that the loop is needed for substrate recognition. For example, changing AGNN to GUNN can significantly affect the rate of RNA cleavage at varying salt concentrations.²³⁶ In a similar fashion, mutating AGNN to GANN decreases both the RNA binding affinity and its cleavage rate,²³⁵ suggestive of a critical role played by the AGNN loop motif in Rnt1p substrate recognition and processing. This is further supported by the 3-6 fold reduction in binding affinity observed for UGNN relative to AGNN loop.²³⁷ Other studies have also reported complex formation between the dsRBD of Rnt1p and a AGNN tetraloop.²³⁶ Substrate binding to Rnt1p is completely impaired when the loop is replaced by GAAA.²³⁷ It remains unclear whether the dsRNA binding domain and/or the sequence of non-conserved adjacent amino acids within the C-terminus of Rnt1p is involved in loop recognition.^{211,236}

Apart from its role as a recognition motif, the loop seems to help the enzyme identify the site of cleavage, *i.e.*, it acts as a structural moiety that the enzyme uses in order to map its cleavage site. In agreement with this notion is the fact that Rnt1p consistently cleaves at a well-defined distance [14-16 nucleotides] from the terminal loop.²³⁵

The base-sequence of the stem plays an essential role in recognition and cleavage of RNA duplexes by Rnt1p.^{210,237} Rnt1p interacts with the dsRNA stem in addition to its primary interaction with the terminal loop. Abou Elela and co-workers have shown that the dsRNA sequence adjacent to the tetraloop regulates Rnt1p cleavage by interfering with RNA binding. Sequences surrounding the cleavage site also directly influence the cleavage efficiency without affecting binding (Abou Elela, unpublished results). Their studies show that a minimum of 5 base pair stem in conjunction with a AGNN loop is sufficient for recognition by Rnt1p, while the minimum cleavage distance is 14 base pairs away from the loop.

4.3 PROJECT OVERVIEW AND DESIGN

rRNA, snRNAs, and snoRNAs are specifically cleaved by Rnt1p, the yeast orthologue of RNase III, at stems capped with a terminal AGNN tetraloop. However, the

mechanism of RNA recognition and cleavage by the enzyme remains poorly understood and needs further exploration. For instance, it is unclear whether the type of helical stem conformation or the loop region is the sole factor governing recognition or whether it is an interdependence of both loop and stem conformations. Understanding Rnt1p selectivity may help explain how other dsRBPs select their substrates.

In this chapter, we explore the molecular requirements for duplex recognition by Rnt1p. Towards this end, we synthesized various DNA/RNA chimeric substrates and investigated their binding affinity to Rnt1p. These studies shed light on the role of stem conformation and loop composition. We also wished to investigate the ability of different hairpin analogues to invoke Rnt1p cleavage. Insights into the cleavage reaction were explored by introducing chemical modifications at specific nucleotide residues near the predicted cleavage site. Comparing the effects of modified substrates relative to their unmodified counterparts allowed us to assess the important factors contributing to the cleavage of the scissile phosphodiester linkage.

4.4 INVESTIGATION OF RNT1P SUBSTRATE RECOGNITION

The molecular requirements for substrate binding to Rnt1p were investigated by using chemically modified oligonucleotides. All oligomers were designed to contain the same base-sequence, but differed in the sugar composition and/or phosphodiester linkages. The control RNA sequence under study [5'-GGC GUC AUG (AGUC) CAU GGC GCC-3'] [4.1] is self-complementary and folds on itself to form a hairpin structure (Figure 4.2).²³⁷ It is designed to contain the terminal AGNN tetraloop and the first 5 base pairs present in Rnt1p substrates at the 3'end of U5 snRNA.²²⁰ Four additional G-C base pairs were added at the end of the stem in order to ensure thermal stability. Considering that this hairpin substrate has 9 base pairs in the stem, it is expected to bind to Rnt1p, but is not supposed to be cleaved by the enzyme since its stem length is less than 14 base pairs.²³⁷

Three modified hairpin analogues were designed to test the effect of loop and stem helical structures on recognition by Rnt1p (Figure 4.2). In the first case, the RNA stem of the control R₉LR₉ [4.1] was replaced with RNA:DNA and DNA:DNA duplexes to obtain R₉LD₉ [4.2] and D₉LD₉ [4.3] respectively. In the second case, the 3',5'-RNA

loop was replaced with the 2',5'-RNA isomer to yield R₉LR₉ [4.4]. The thermal, conformational and binding properties of these hairpins were studied and compared to the well-characterized all-RNA hairpin [R₉LR₉].²³⁷

T_m	71.6 °C	56.9 °C	60.6 °C	70.6 °C
%H	13.6 %	13.6 %	13.4 %	13.9 %
	G U A C	G U A C	G U A C	<u>G</u> U <u>A</u> C
	G-C	G-c	g-c	G-C
	U-A	U-a	u-a	U-A
	A-U	A-u	a-u	A-U
	C-G	C-g	c-g	C-G
	U-G	U-g	u-g	U-G
	G-C	G-c	g-c	G-C
	C-G	C-g	c-g	C-G
	G-C	G-c	g-c	G-C
	G-C	G-c	g-c	G-C
	5' 3'	5' 3'	5' 3'	5' 3'
	R ₉ LR ₉ All RNA <u>4.1</u>	R ₉ LD ₉ RNA:DNA stem <u>4.2</u>	D ₉ LD ₉ DNA:DNA stem <u>4.3</u>	R ₉ <u>L</u> R ₉ 2',5'-RNA loop <u>4.4</u>

Figure 4.2: Short hairpin substrates (22 nucleotides in length) designed to probe the binding specificity of Rnt1p. The stem region consists of nine base pairs and the loop contains the tetranucleotide AGUC. All substrates contain the same base sequence but differ in the sugar composition and /or position of phosphodiester bond. R₉LR₉ is an all-RNA hairpin with identical stem-loop structure to the 3'end of U5 snRNA and serves as a positive control. R₉LR₉ contains a 2',5'-linked RNA loop, R₉LD₉ has an RNA:DNA hybrid stem with an RNA loop, while D₉LD₉ contains the RNA (AGUC) loop with a DNA stem duplex. Capital letters represent RNA residues, small letters represent DNA residues, and capital underlined letters represent 2',5'-RNA residues. UV melting temperature (T_m) of the various hairpin substrates along with their %hypochromicity (%H) are given. Measurements were conducted in 0.01 M Na₂HPO₄, 0.1 mM Na₂EDTA buffer, pH 7.00 ± 0.02 at an oligonucleotide concentration ~ 4.5 μM. Values represent the average of three independent measurements. Error in T_m is ± 0.5 °C. The melting curves show a single cooperative and completely reversible transition. %Hypochromicity was calculated from UV absorbances of the hairpin (A_0) and fully denatured species (A_f) according to the following equation: %H = ($A_f - A_0$) / A_f .

4.4.1 Thermal and Helical Properties

The goal behind these studies was to investigate the effect of hairpin stability and conformation on the binding affinity to Rnt1p.

The thermally induced UV melting profiles of the individual hairpins are shown in **Figure 4.3**. All hairpins exhibit *monophasic* and completely reversible melting behaviors. Insertion of 2',5'-linkages within the (AGUC) tetraloop does not impair thermal stability (**Figure 4.2**; compare R₉LR₉ [4.1] & R₉LR₉ [4.4]). In fact, the percentage hypochromicity (%H) characteristic of each hairpin is not affected, indicative of comparable stacking interactions within the overall helical structure. This result is in agreement with what we previously observed for 2',5'-RNA (UUCG) tetraloops [Chapter III]. The hairpins D₉LD₉ [4.3] and R₉LD₉ [4.2] were of lower thermal stability relative to R₉LR₉, with R₉LD₉ being the least stable ($T_m = 56.9$ °C; **Figure 4.2**).

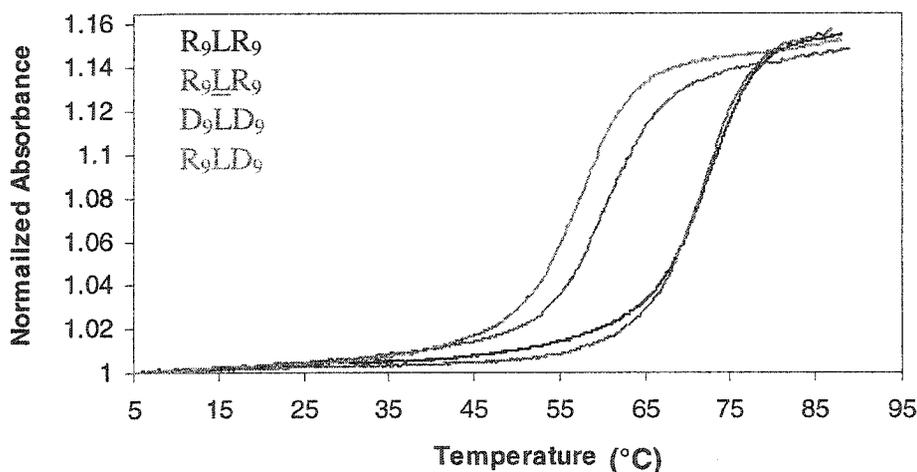


Figure 4.3: Normalized melting temperature profiles (at 260 nm) of representative hairpin substrates. Measurements were done in 0.01 M Na₂HPO₄, 0.1 mM Na₂EDTA buffer, pH 7.00 ± 0.02 at an oligonucleotide concentration ~ 4.5 μM.

To assess the relative global conformations of the various hairpins under study, we resorted to circular dichroism (CD) spectroscopy. The CD spectra of R₉LR₉ & R₉LR₉ are essentially similar [A-form], suggesting that they share common overall helical features (**Figure 4.4**). However, there is increase in the depth of the CD peak at *ca.* 212 nm for the 2',5'-RNA-loop-hairpin [4.4] compared to the native counterpart. This again

parallels the CD findings reported in chapter II about 2',5'-RNA (UUCG) loops. As expected, D₉LD₉, with a DNA duplex stem, exhibits typical B-form features, while R₉LD₉ with an RNA:DNA hybrid stem shows an overall A-like CD spectrum similar to R₉LR₉.

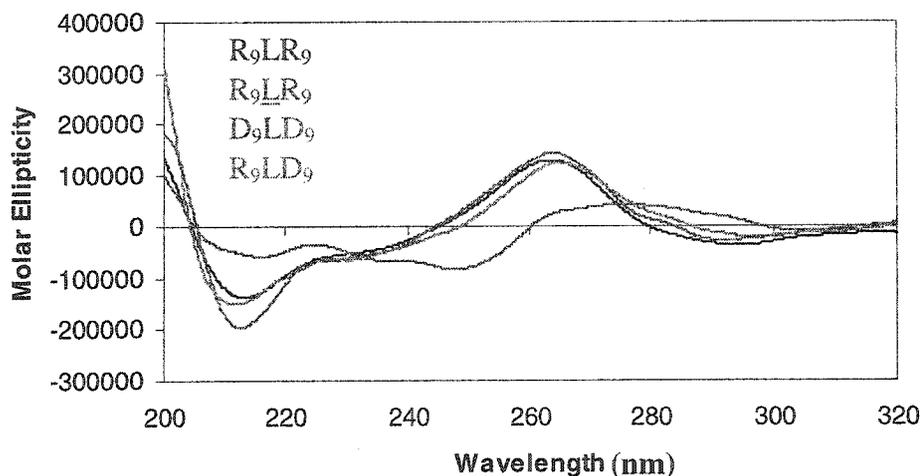


Figure 4.4: Circular dichroism spectra at 22 °C of representative hairpins of mixed stem composition. All measurements were done in 0.01 M Na₂HPO₄, 0.1 mM Na₂EDTA buffer, at pH 7.00 ± 0.02. Molar ellipticities were normalized to strand concentration.

4.4.2 Effect of Stem Composition on Hairpin Binding to Rnt1p

In collaboration with Professor Sherif Abou Elela and Bruno Lamontagne of the Département de Microbiologie et Infectiologie, University of Sherbrooke, the ability of various modified hairpins to act as substrates that bind Rnt1p was tested using gel electrophoresis. The design and synthesis of sugar-modified hairpin substrates were done at McGill, while all biological assays were executed at the University of Sherbrooke.

The individual hairpins were [³²P]-5'-end labeled, incubated in the absence of divalent ions [Mg²⁺] with varying amounts of Rnt1p for 10 min on ice, and then run on native (4%) polyacrylamide gels (**Figure 4.5**). The newly formed band that represents RNA/protein complex migrated slower than the free RNA. The protein concentration was consequently plotted against the percentage of bound hairpin (**Figure 4.6**) from which the binding constant (K_d) value was calculated. Was determined from three independent measurements (data sets). No major RNA duplexes or other complexes could be detected

for any substrate in the absence of enzyme suggesting that substrate folding is unimolecular (*hairpin* species rather than *duplex*).

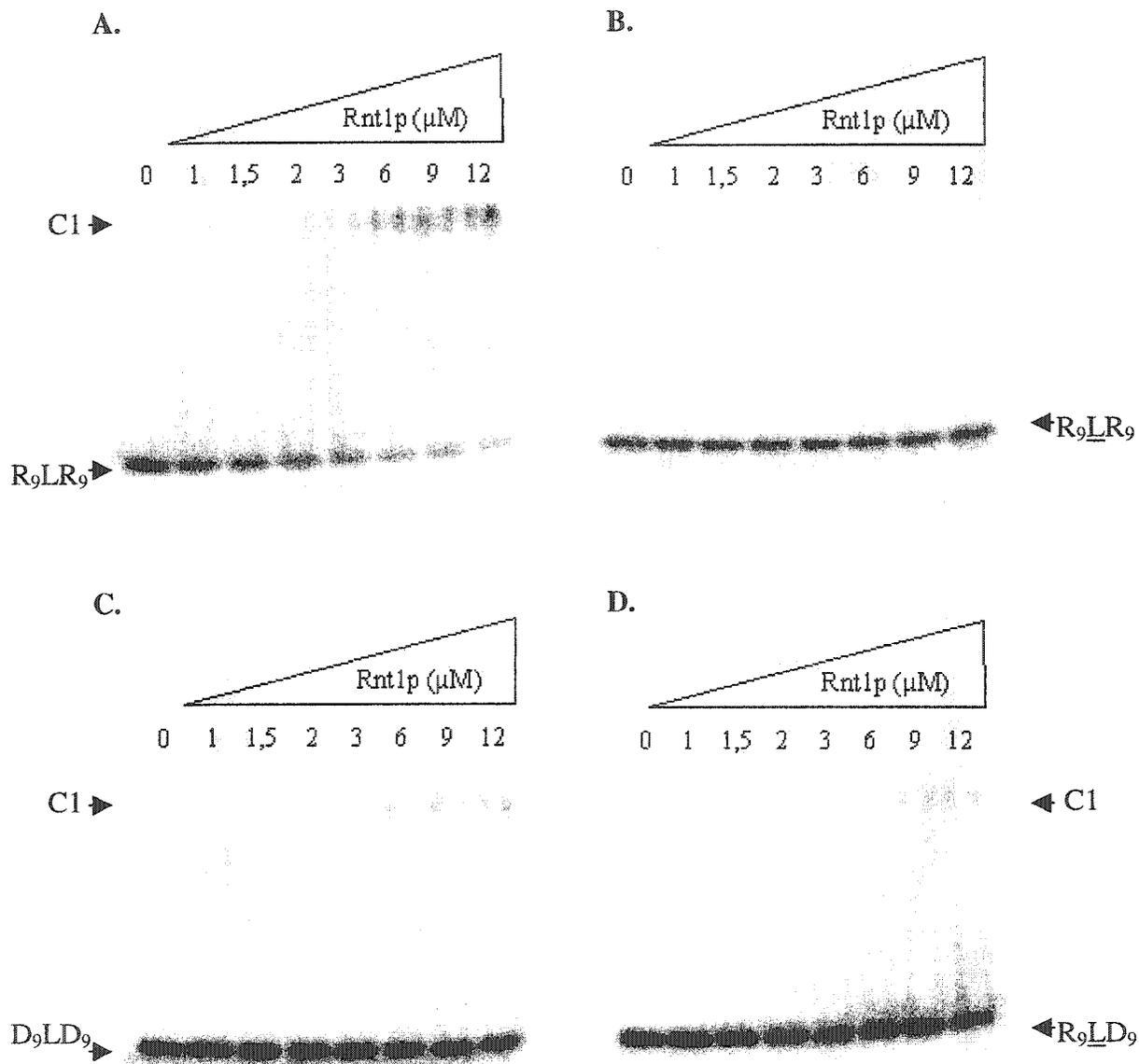


Figure 4.5: Gel shift binding assay of short hairpin substrates (9 bp stems). Increasing concentrations of Rnt1p [0-12 μM, shown on top] were incubated with 2 fmol of 5'-[³²P]-labeled hairpin in 150 mM KCl. The gel pictures show complex formation (C1) between the labeled hairpin and Rnt1p. The positions of the bound (shifted) and unbound RNA are indicated.

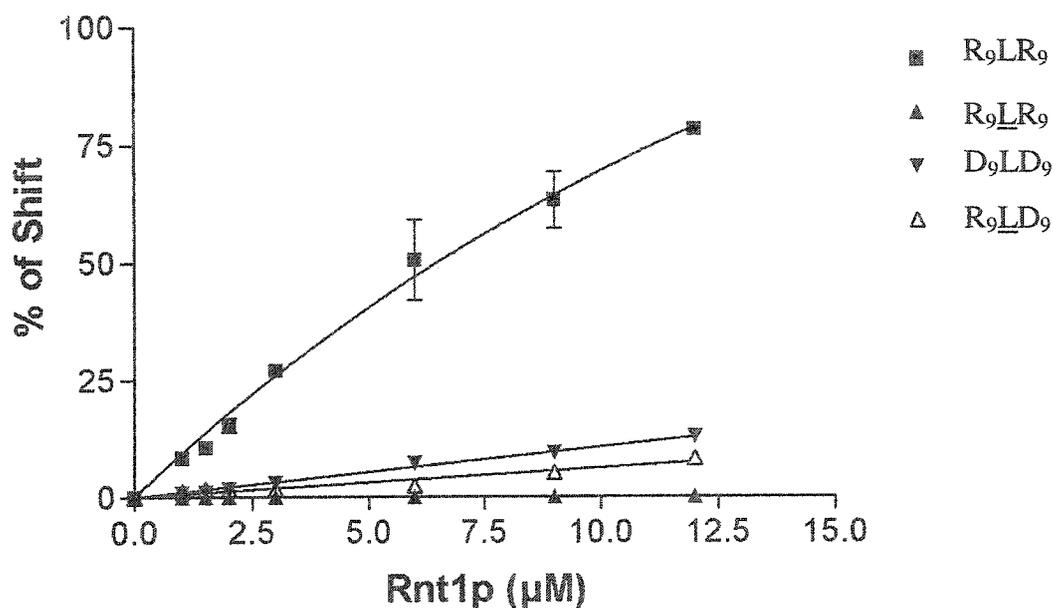


Figure 4.6: Plot of % shifted hairpin versus Rnt1p concentration as obtained from quantitating the gel bands in **Figure 4.5**.

Our results clearly show that hairpins with DNA:DNA and RNA:DNA stem duplexes bind weakly to Rnt1p [$K_d > 15 \mu\text{M}$] (**Table 4.1**). This is surprising in view that Rnt1p is a double-stranded RNA specific enzyme. In sharp contrast, the all-RNA hairpin **[4.1]** (positive control) binds to Rnt1p with high affinity [$K_d > 6 \mu\text{M}$]. The weak binding of the DNA duplex stem [*i.e.*, D₉LD₉] might possibly be due to the drastic differences in helical conformation (A- to B-form as evident from CD studies) compared to the all-RNA hairpin. Although the A-like helical structure of the RNA:DNA hybrid is similar to that of R₉LR₉ (as judged from CD), there exist subtle differences between pure RNA duplexes and RNA:DNA hybrid structures.⁴⁻⁶ In hybrids, the sugar residues in the RNA strand adopt a C-3' *endo* pucker, while those in the DNA strand adopt a mixture of C-2' *endo*, C-3' *endo*, and C-4' *endo* pucker forms.^{7,8} These might contribute to the weak binding affinity observed for R₉LD₉. In contrast, RNA duplexes exhibit uniform C-3' *endo* sugar repeats. All of the above results point in one direction which shows that the stem helical conformation is an important factor governing recognition of small substrates (9 base pair stems) by Rnt1p. A compact A-form helix with repeating C3'-*endo* sugar units is needed to achieve better binding. However, it is important not to confuse

this observation with longer stems. In this study, the small substrates (9 base pair stems) would be expected to interact only with the enzyme's dsRBD due to their short length. In sharp contrast, longer substrates would be expected to bind to dsRBD but also would exhibit other interactions with the enzyme's N-terminal domain that would affect binding and thus would be expected to show different behavior owing to the nature of the additional molecular interactions.²¹¹

Table 4.1: Binding constants (K_d) of various short hairpin substrates towards Rnt1p.

Code	Designation	K_d (μM)
<u>4.1</u>	R ₉ LR ₉	6
<u>4.2</u>	R ₉ LD ₉	>15
<u>4.3</u>	D ₉ LD ₉	>15
<u>4.4</u>	R ₉ <u>L</u> R ₉	-

Measurements were determined from three different assays. Errors in values of K_d are within $\pm 0.2 \mu\text{M}$.

4.4.3 Effect of Loop Structure and Base Sequence on Hairpin Binding to Rnt1p

Next, we investigated the role that the loop plays in substrate recognition. Inserting 2',5'-linkages within the loop [*i.e.*, R₉LR₉ where L = A_{2p5}C_{2p5}U_{2p5}C_{2p}] completely abolishes binding to Rnt1p (**Figure 4.5 B**). Though it does not thermally destabilize the hairpin or disrupt its overall helical structure (as judged by CD in **Figure 4.4**), introducing 2',5'-linkages in the loop would be expected to shift the location of the phosphodiester linkages and affect its intrinsic interactions with Rnt1p. This is further supported by our studies in chapter II which show that the 2',5'-linked (UUCG) RNA loop adopts a different folded structure from the 3',5'-RNA analogue. Given that the same may be operating here for the case of (AGUC) loop, these studies evidently show that loop conformation is an important determinant of substrate binding to Rnt1p.

We also studied hairpins containing base mutations in the loop (**Figure 4.7**). As mentioned earlier, the wild-type AGUC belongs to the AGNN class of tetraloops where N represents any nucleotide base residue. Thus, mutating the A or G residues should in

principle affect binding while changing any of the N residues should not. Indeed, changing the UC residues to AA [to obtain AGAA; [4.7](#)] does not impair binding.

K_d	6 μ M	6 μ M	-	6 μ M	-
T_m	71.6 °C	70.7 °C	71.4 °C	72.6 °C	73.2 °C
%H	13.6%	10.9%	11.8%	12.1%	15.2%

G U	G U	C U	G A	A A
A C	C C	A C	A A	G A
G-C	G-C	G-C	G-C	G-C
U-A	U-A	U-A	U-A	U-A
A-U	A-U	A-U	A-U	A-U
C-G	C-G	C-G	C-G	C-G
U-G	U-G	U-G	U-G	U-G
G-C	G-C	G-C	G-C	G-C
C-G	C-G	C-G	C-G	C-G
G-C	G-C	G-C	G-C	G-C
G-C	G-C	G-C	G-C	G-C
5' 3'	5' 3'	5' 3'	5' 3'	5' 3'
4.1	4.5	4.6	4.7	4.8

Figure 4.7: Schematic representation comparing the binding constant (K_d), thermal stability (T_m) and %hypochromicity (%H) of various RNA hairpins differing only in the loop base-sequence. Bold letters represent base mutations in the wild type (AGUC) loop. UV thermal melting measurements were made in 0.01 M Na_2HPO_4 , 0.1 mM Na_2EDTA buffer, pH 7.00 ± 0.02 at an oligonucleotide concentration of 4.5 μM . Values represent the average of three independent measurements. Error in T_m values is ± 0.5 °C. The melting curves show a single cooperative and completely reversible transition. %Hypochromicity was calculated from UV absorbances of the hairpin (A_0) and fully denatured species (A_f) according to the following equation: $\%H = (A_f - A_0) / A_f$. Errors in values of K_d are within ± 0.2 μM .

Furthermore, mutating the second G residue to a C [thus obtaining ACUC; [4.6](#)] abolishes the binding affinity to Rnt1p, in the same way as changing the AG residues to GA does {compare AGAA [4.7](#) to GAAA [4.8](#)}. Surprisingly, mutating the first A residue to a C residue (CGUC [4.5](#)) does not impair the binding affinity to Rnt1p. This is not in agreement with the proposed notion about the existence of an AGNN motif governing the molecular interaction between the loop and Rnt1p. The CGUC could potentially adopt a folded structure very similar to AGUC which probably explains why it is recognized by

the enzyme. NMR studies will have to be carried out in order to clarify this anomalous behavior.

Together, the above results show that the loop residues [base sequence and composition] and loop conformation is an essential element in recognition of hairpin substrates by Rnt1p enzyme. In short substrates (9 base pair stems), this appears to work in conjunction with the stem helical requirements discussed previously. This suggests that Rnt1p is a very specific enzyme with respect to both loop and stem architectures.

4.5 PROBING THE MOLECULAR REQUIREMENTS FOR CLEAVAGE OF HAIRPIN DUPLEXES BY RNT1P

In order to probe and understand Rnt1p cleavage mechanism, we have synthesized various hairpins [42-nucleotides in length] containing the (AGUC) loop. These hairpins were similar to those studied earlier (Section 4.4.2) except that they were of longer stem length (19 base pairs) in order to qualify for potential cleavage by the enzyme (Figure 4.8). They consisted of either RNA or DNA duplexes or DNA/RNA hybrids in the stem. The loop composition and base-sequence was always kept the same. For example, the stem of hairpin R₁₉LD₁₉ [4.10] contains RNA and DNA residues at the 5'-end and 3'-end, respectively. D₁₉LD₁₉ [4.12] is a hairpin with RNA (AGUC) loop, but having DNA stem duplex.

In our preliminary experiments, all hairpin structures (Figure 4.9) were checked for their ability to bind Rnt1p *via* gel-shift mobility assays. Our results show that all substrates can bind efficiently to Rnt1p with good K_d values (Table in Figure 4.8). R₁₉LR₁₉ binds most efficiently with a K_d of 0.67 μ M. The DNA duplex counterpart [D₁₉LD₁₉] binds less efficiently ($K_d = 0.96 \mu$ M), while the hybrid substrates exhibit the weakest binding ($K_d \sim 2.20$ - 2.40μ M). We then shifted our attention to testing the cleavage of these substrates by Rnt1p. They were incubated in the presence of Mg²⁺ and the product mixture was then run on denaturing polyacrylamide gels. As shown in Figure 4.10, hairpins R₁₉LD₁₉ and D₁₉LD₁₉ are not cleaved by Rnt1p. However, hairpin D₁₉LR₁₉ [4.11] is cleaved by Rnt1p. Surprisingly, the cleavage occurs at the DNA portion of the stem (5'-end of the stem). This finding is remarkable and provides the first example of an RNA-dependent DNA cleavage.

K_d	0.67 μ M	2.20 μ M	2.40 μ M	0.96 μ M
T_m	71.6 °C	56.9 °C	60.6 °C	70.6 °C
%H	13.6 %	13.6 %	13.4 %	13.9 %

G U	G U	G U	G U
A C	A C	A C	A C
G-C	G-c	g-C	g-c
U-A	U-a	t-A	t-a
A-U	A-t	a-U	a-t
C-G	C-g	c-G	c-g
U-G	U-g	t-G	t-g
G-C	G-c	g-C	g-c
U-A	U-a	t-A	t-a
A-U	A-t	a-U	a-t
C-G	C-g	c-G	c-g
U-G	U-g	t-G	t-g
G-C	G-c	g-C	g-c
U-A	U-a	t-A	t-a
A-U	A-t	a-U	a-t
C-G	C-g	c-G	c-g
▶ U-G	U-g	▶ t-G	t-g
G-C	G-c	g-C	g-c
C-G	C-g	c-G	c-c
G-C	G-c	g-C	g-c
G-C	G-c	g-C	g-c
5' 3'	5' 3'	5' 3'	5' 3'
R ₁₉ LR ₁₉	R ₁₉ LD ₁₉	D ₁₉ LR ₁₉	D ₁₉ LD ₁₉
All RNA	RNA:DNA stem	DNA:RNA stem	DNA:DNA stem
<u>4.9</u>	<u>4.10</u>	<u>4.11</u>	<u>4.12</u>

Figure 4.8: Long hairpin substrates (42 nucleotides in length) designed to probe the cleavage reaction by Rnt1p. They consist of 19 bp in the stem and contain the tetranucleotide RNA loop sequence (AGUC). The sequence of loop and the first 5 base pairs adjacent to it represent the terminal stem-loop structure at the 3'end of U5 snRNA. R₁₉LR₁₉ is an all-RNA hairpin that serves as a positive control and is cleaved by Rnt1p fourteen nucleotides away from the loop on the 5'-end of the stem [the arrow shows where cleavage takes place]. R₁₉LD₁₉ has an RNA:DNA hybrid stem with an RNA loop, while D₉LD₉ contains the RNA (AGUC) loop with a DNA stem duplex. Capital letters represent RNA residues, small letters represent DNA residues, and capital underlined letters represent 2',5'-RNA residues. UV thermal melting measurements were conducted in 0.01 M Na₂HPO₄, 0.1 mM Na₂EDTA buffer, pH 7.00 ± 0.02 at an oligonucleotide concentration of 4.5 μ M. Values represent the average of three independent measurements. Error in T_m values is ± 0.5 °C. The melting curves show a single cooperative and completely reversible transition. Errors in values of K_d are within ± 0.2 μ M.

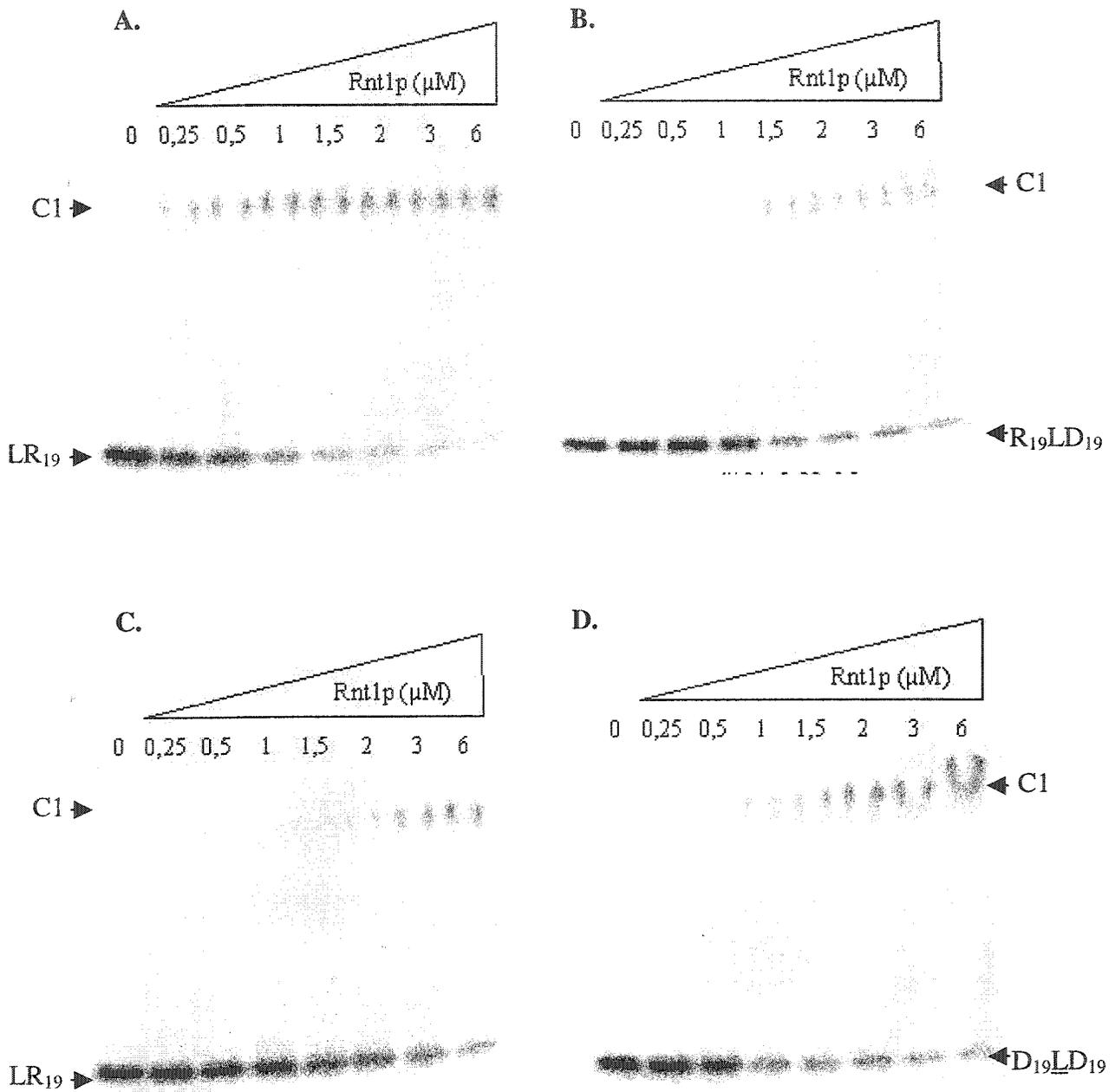


Figure 4.9: Gel shift binding assay of long hairpin substrates (19 bp stems). Increasing concentration of Rnt1p [0-6 μM, shown on top] was incubated with 2 fmol of 5'-[³²P]-labeled hairpin in 150 mM KCl. The gel pictures show complex formation (C1) between the labeled hairpin and the Rnt1p enzyme. The positions of the bound (shifted) and unbound RNA are indicated.

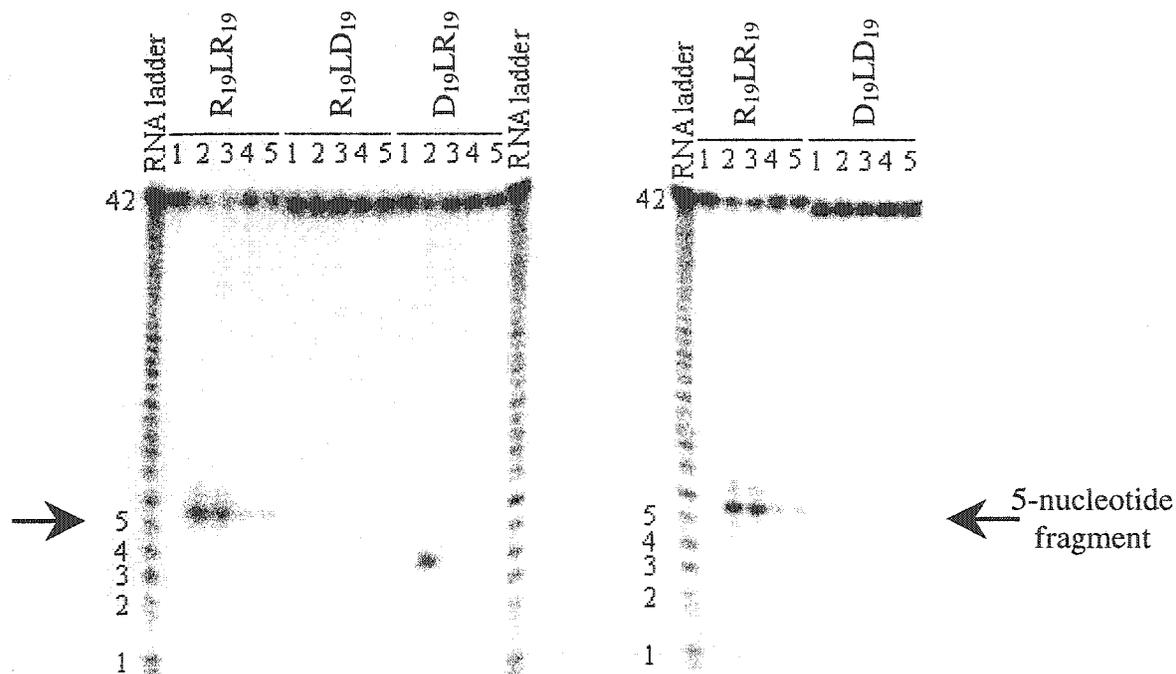


Figure 4.10: Cleavage assay of long hairpins (42-nt) with various stem compositions. The different hairpin substrates were incubated with Rnt1p in the presence of Mg^{2+} and either (1) nothing, (2) 10 mM KCl in protein excess, (3) 150 mM KCl in protein excess, (4) 10 mM KCl in hairpin excess, or (5) 150 mM in hairpin excess. The reaction products were then fractionated on a 20% denaturing polyacrylamide gel and the bands visualized using Instant Imager. The cleavage products are indicated with arrows. A 42-nt RNA marker, obtained by alkaline hydrolysis of $R_{19}LR_{19}$, was run as control at the sides.

A closer inspection of the CD spectral characteristics of each hairpin shows that while there exist subtle differences in the overall conformation of $D_{19}LR_{19}$ and $R_{19}LD_{19}$, both point towards A-like helices (**Figure 4.11**). Yet, $D_{19}LR_{19}$ is cleaved at its 5'-end while $R_{19}LD_{19}$ is not.

To confirm the identity of the substrates, we subjected both $D_{19}LR_{19}$ and $R_{19}LD_{19}$ to cellular RNases in the presence or absence of Rnt1p (**Figure 4.12**). The cleaved five-nucleotide fragment from $D_{19}LR_{19}$ resists nuclease degradation, thus confirming that it is of DNA morphology. In sharp contrast, the cleaved 5-nucleotide fragment from $R_{19}LR_{19}$ is completely degraded upon RNase treatment as would be expected for RNA. Also, important to note is the faster gel mobility of the five-nucleotide fragment produced from $D_{19}LR_{19}$ cleavage relative to that from $R_{19}LR_{19}$, in agreement with the well-known relative mobilities of DNA and RNA.

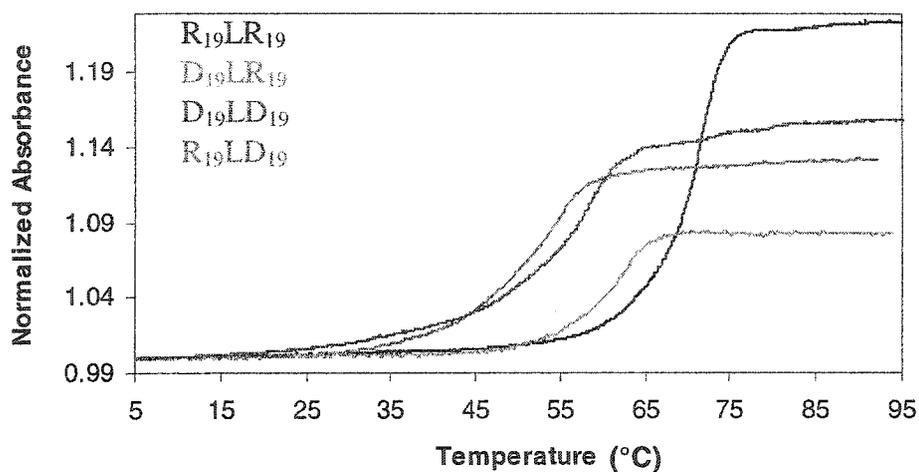


Figure 4.11 A: Normalized melting temperature profiles (at 260 nm) of representative long hairpin substrates. Measurements were done in 0.01 M Na_2HPO_4 , 0.1 mM Na_2EDTA buffer, pH 7.00 ± 0.02 at an oligonucleotide concentration $\sim 4.5 \mu\text{M}$.

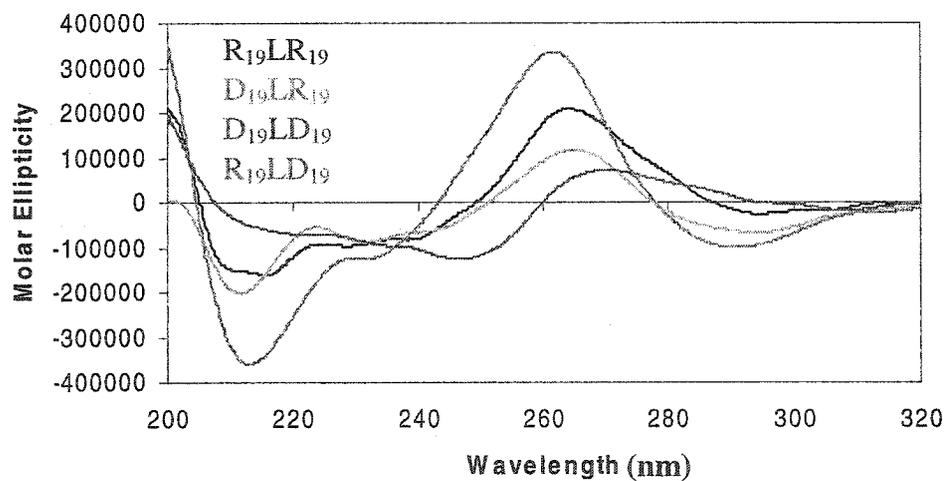


Figure 4.11 B: Circular dichroism spectra at 22 °C of representative hairpins of mixed stem composition. All measurements were done in 0.01 M Na_2HPO_4 , 0.1 mM Na_2EDTA buffer, at pH 7.00 ± 0.02 . Molar ellipticities were normalized to strand concentration.

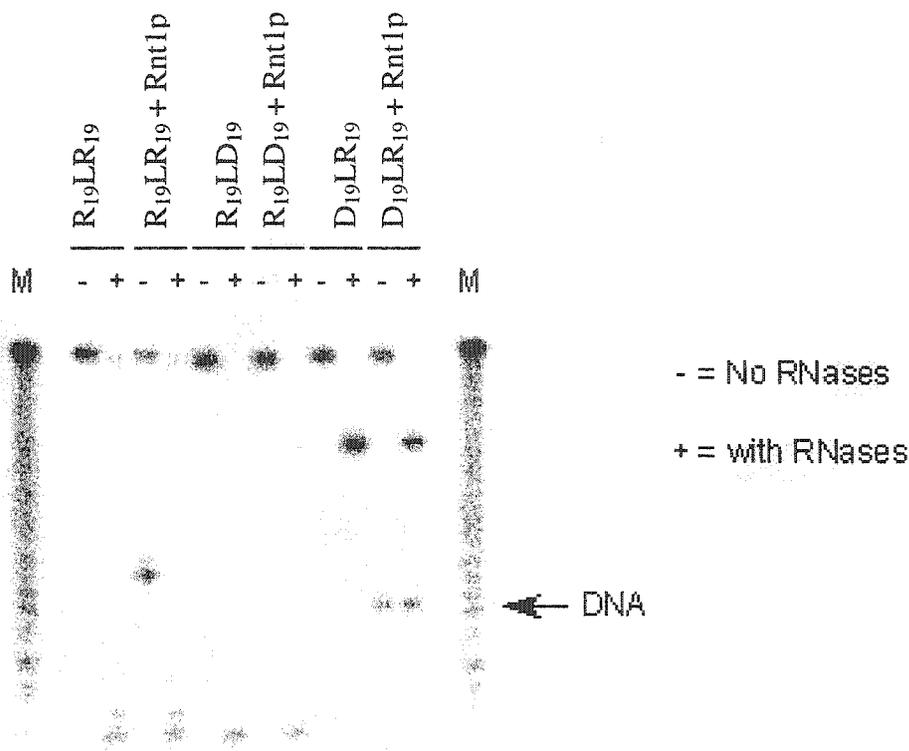


Figure 4.12: RNase treatment of $D_{19}LR_{19}$, $R_{19}LD_{19}$, and $R_{19}LR_{19}$ substrates in the presence and absence of Rnt1p. The cleavage assays were conducted under the same exact conditions as those mentioned in **Figure 4.9** except that RNases were added to the reaction mixture in order to effect RNA degradation. (-) No RNases. (+) with RNases. M denotes a marker obtained by alkaline hydrolysis of the original hairpin substrates under study.

To further confirm that the cleaved fragment is really DNA and not RNA, we ran a synthetically-made DNA oligomer of the sequence 5'-ggcgt-3', corresponding to the same sequence of the proposed five-nucleotide degradation product. **Figure 4.13** shows that both oligomers travel side by side with exactly the same gel mobility, confirming that the cleaved portion from $D_{19}LR_{19}$ is evidently DNA.

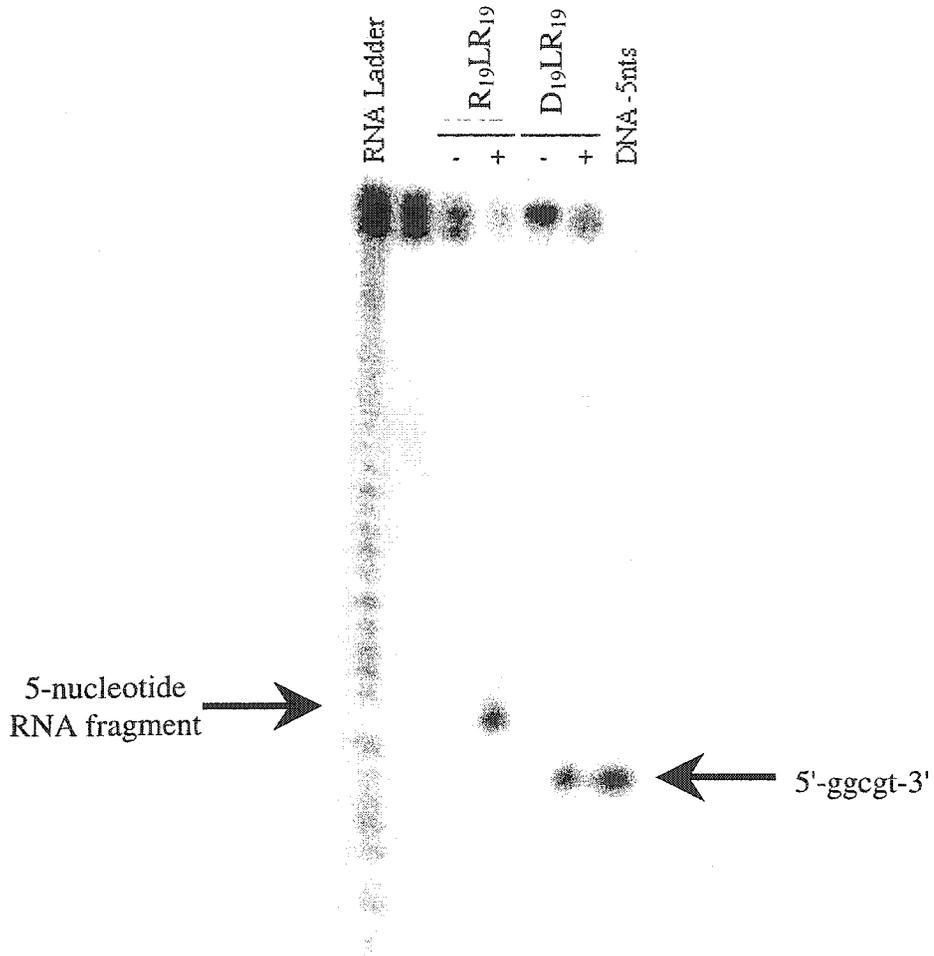


Figure 4.13: The product from the cleavage assay of D₁₉LR₁₉ run side by side with a synthetically-made 5-nucleotide DNA fragment of the sequence 5'-ggcgt-3' in the (+) presence of Rnt1p and (-) absence of Rnt1p. The RNA ladder was created via alkaline hydrolysis of the 5'-[³²P]-labeled hairpin under basic conditions (see **Section 7.6.10b** in experimental).

4.6 YEAST RNASE III (RNT1P) CLEAVES DNA IN DNA:RNA HYBRIDS

To confirm the above observations and shed light on the mechanism of Rnt1p cleavage, the 2'-hydroxyl moiety in the furanose ring located at the 3'-side of the scissile phosphodiester linkage (residue U5) was modified (**Figure 4.14 A**). Namely, the uridine residue at position 5 was replaced with 2'-deoxy uridine (dU) and 2'-O-methyl uridine (2'-OMe U).

These modified hairpins R_{19(dU)}LR₁₉ [**4.13**] and R_{19(2'-OMe)}LR₁₉ [**4.14**] (**Figure 4.14 B**) were slightly less thermally stable than R₁₉LR₁₉. A drop in T_m of *ca.* 3 °C and *ca.* 4.2 °C was caused by the dU and 2'-OMe U inserts, respectively (**Table 4.3**). Furthermore, both insertions had significant effects on the hypochromicity (**Table 4.3**), suggesting a major disturbance in base-base stacking interactions within the duplex. The overall CD spectra of all three hairpins [R₁₉LR₁₉, R_{19(dU)}LR₁₉ & R_{19(2'-OMe)}LR₁₉] look similar and indicate that their conformations fall within the A-type family. **Figure 4.15** shows a representative CD spectrum comparing R_{19(2'-OMe)}LR₁₉ with R₁₉LR₁₉.

To verify hairpin binding to Rnt1p, gel mobility shift assays were conducted in the presence of varying amounts of enzyme (**Figure 4.16**). The gel pictures show that R_{19(2'-OMe)}LR₁₉ binds to Rnt1p slightly better than R_{19(dU)}LR₁₉ [K_d of 0.20 μ M *versus* 0.27 μ M; **Table 4.3**].

As far as cleavage is concerned, R_{19(dU)}LR₁₉ is significantly cleaved by Rnt1p at the usual position [*i.e.*, fourteen nucleotides away from the tetraloop], despite the fact that the dU residue lacks a 2'-hydroxyl group. In a similar fashion, R_{19(2'-OMe)}LR₁₉ [with a 2'-OMe insertion] is cleaved by Rnt1p, though to a lesser extent than R_{19(dU)}LR₁₉ (**Figure 4.17**).

These results offer further evidence (to **Section 4.5**) that Rnt1p is able to cleave the DNA strand of a DNA:RNA hybrid and that the enzyme does not utilize the 2'-hydroxyl moiety in order to effect cleavage of duplex RNA substrates.

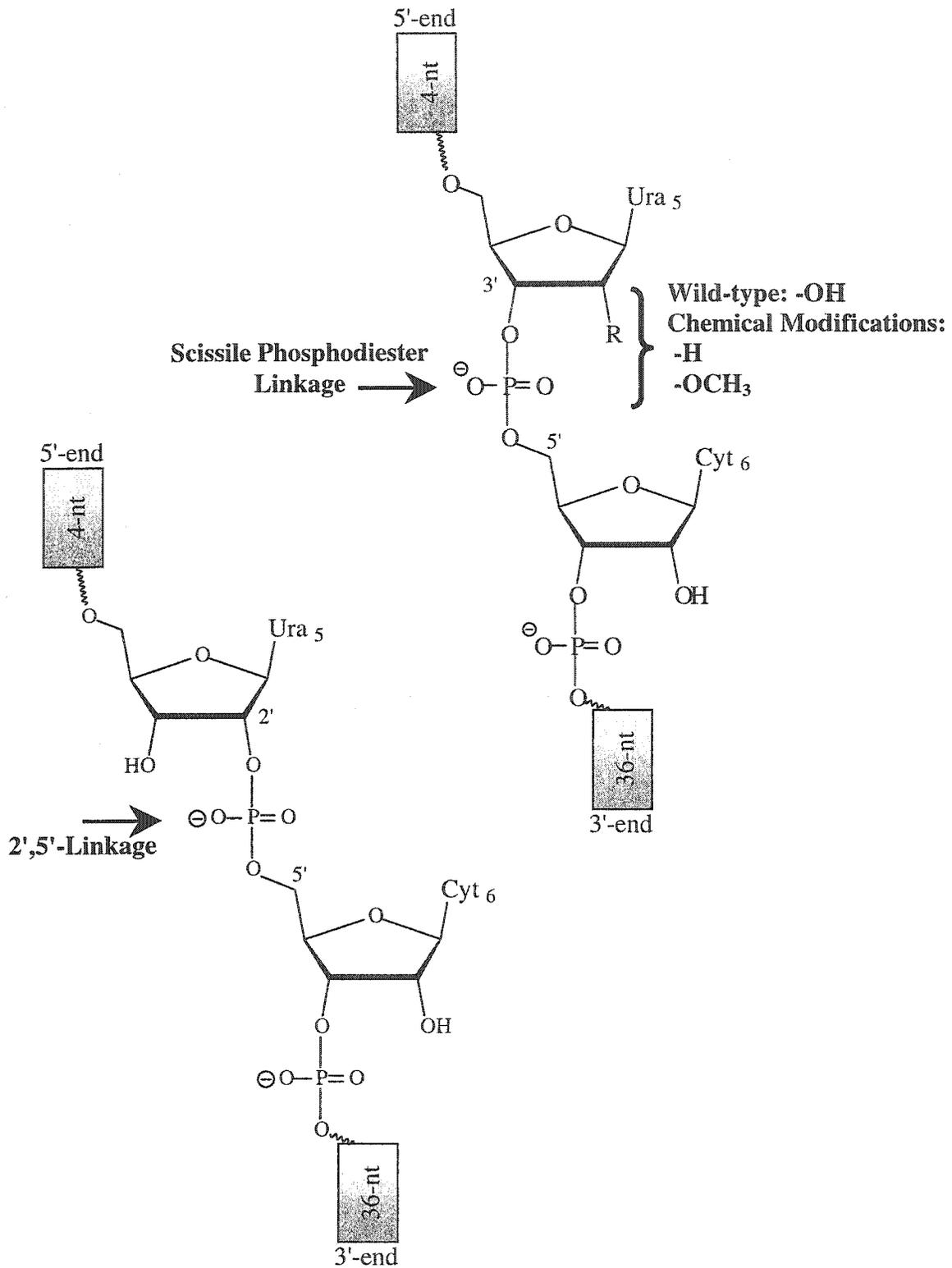


Figure 4.14 A: Schematic diagram showing the substrate cleavage site and potential chemical modifications to probe the cleavage reaction.

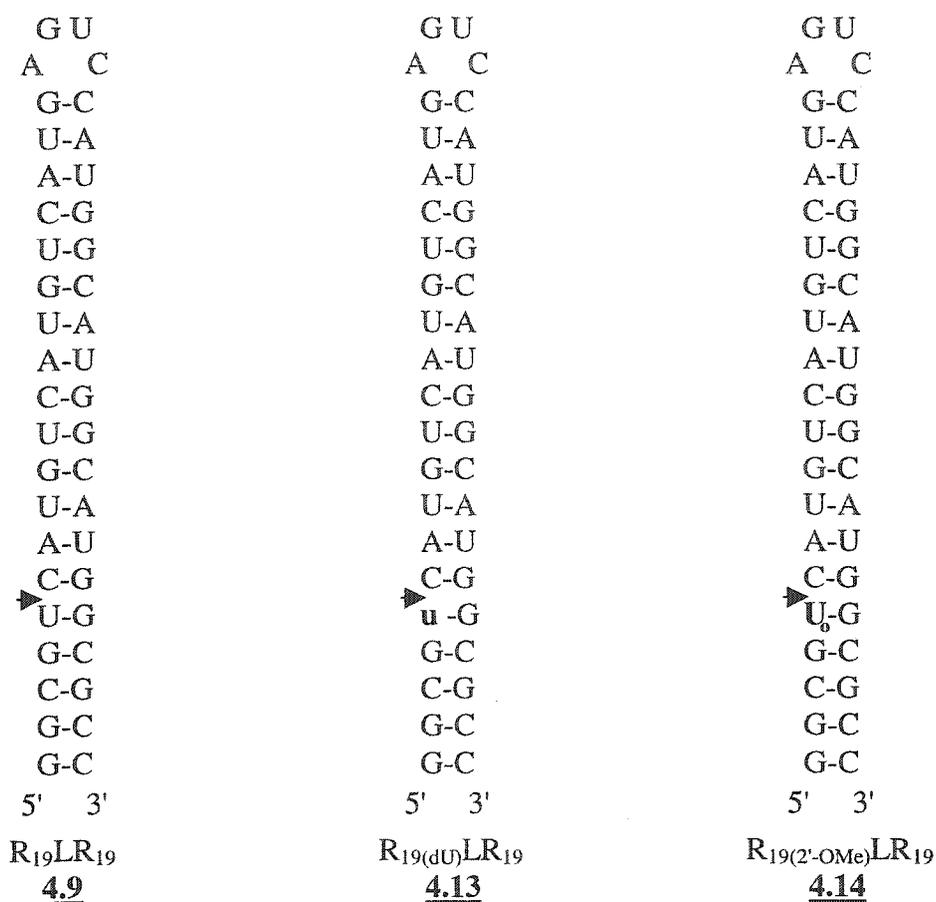


Figure 4.14 B: Long hairpin substrates (42-mer) containing chemically-modified inserts and designed to probe the cleavage reaction by Rnt1p. The letter **u** denotes a 2'-deoxyuridine residue, while **U_o** denotes a 2'-OMe uridine. Capital letters represent RNA residues. The arrow indicates scissile bond where cleavage takes place.

Table 4.3: Long hairpin substrates (42-nt) containing chemically-modified inserts at the cleavage site

Code	Designation	K_d (μM) ^a	T_m ($^{\circ}\text{C}$) ^b	%H
<u>4.9</u>	R ₁₉ LR ₁₉	0.67	69.8	18.3
<u>4.13</u>	R _{19(dU)} LR ₁₉	0.27	66.6	3.0
<u>4.14</u>	R _{19(2'-OMe)} LR ₁₉	0.20	65.7	8.5
<u>4.19</u>	R _{19(2'-rU)} LR ₁₉	0.95	62.9	14.6

^b UV thermal melting measurements were made in 0.01 M Na₂HPO₄, 0.1 mM Na₂EDTA buffer, pH 7.00 ± 0.02 at an oligonucleotide concentration of 4.5 μM . Values represent the average of three independent measurements. Error in T_m values is ± 0.5 $^{\circ}\text{C}$. The melting curves show a single cooperative and completely reversible transition. Errors in values of K_d are within ± 0.09 μM .

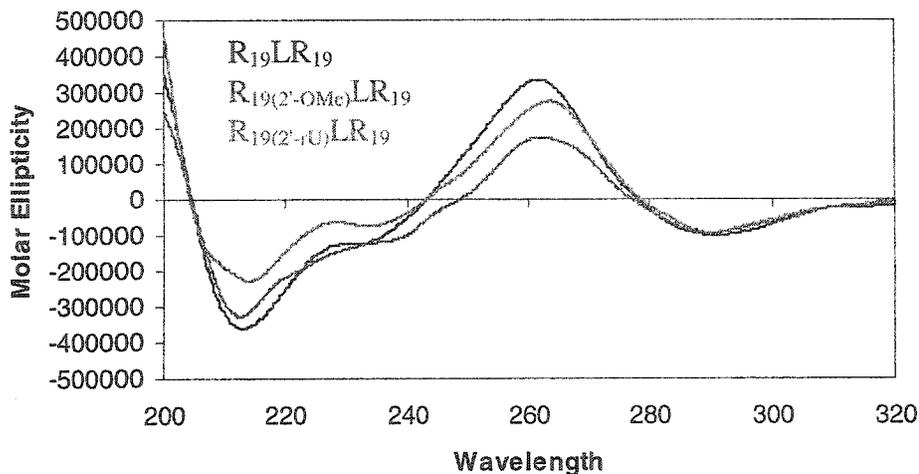


Figure 4.15: Circular dichroism spectra at 22 °C of hairpins with chemically-modified inserts. All measurements were done in 0.01 M Na₂HPO₄, 0.1 mM Na₂EDTA buffer, at pH 7.00 ± 0.02. Molar ellipticities were normalized to strand concentration

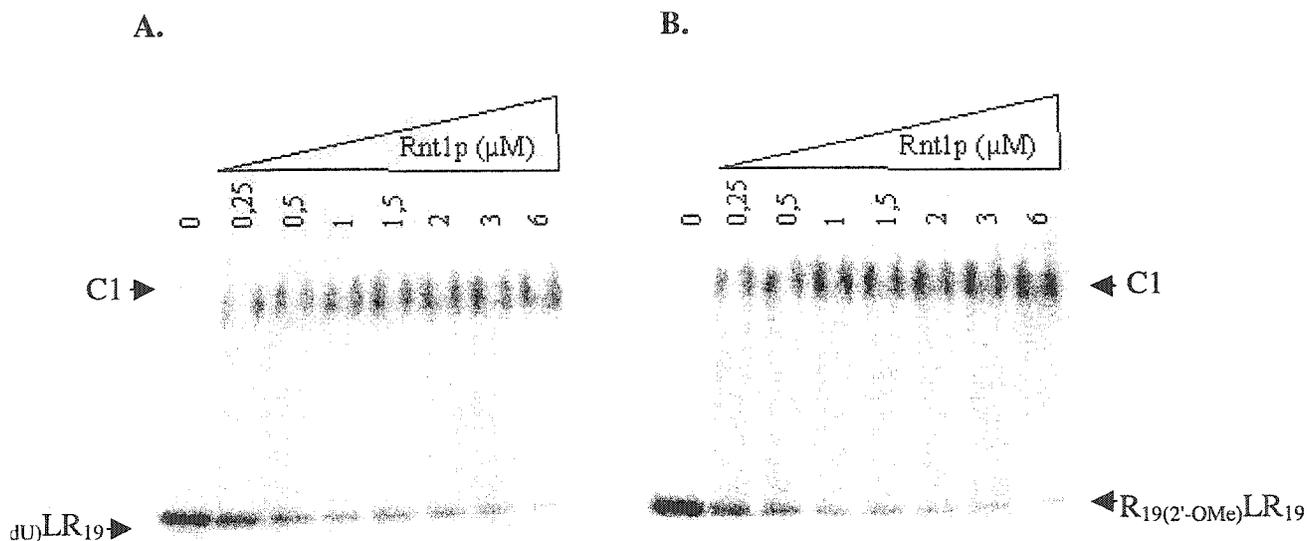


Figure 4.16: Gel shift binding assay of hairpin substrates containing chemical modifications at the cleavage site. Increasing concentration of Rnt1p [0-6 μM, shown on top] was incubated with 2 fmol of 5'-[³²P]-labeled hairpin in 150 mM KCl. The gel pictures show complex formation (C1) between the labeled hairpin and the Rnt1p enzyme. The positions of the bound (shifted) and unbound RNA are indicated.

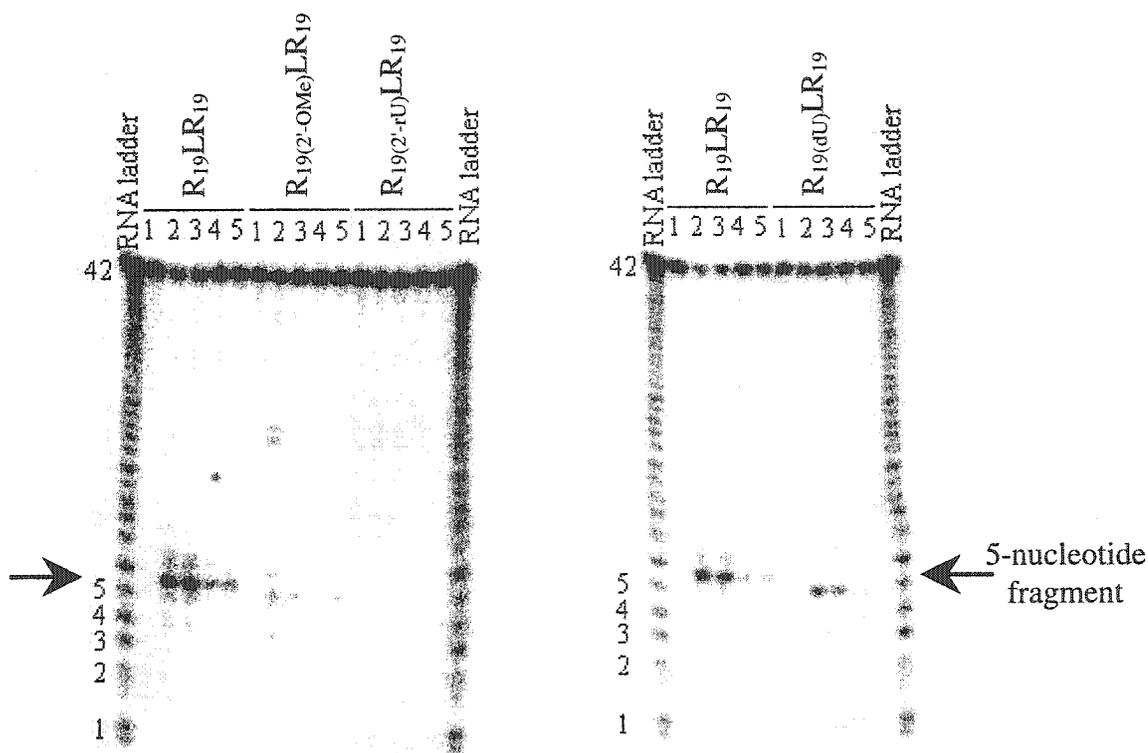


Figure 4.17: Cleavage assay of hairpins with chemically modified inserts at the cleavage site. The hairpin substrates were incubated with Rnt1p in the presence of Mg^{2+} and either (1) nothing, (2) 10 mM KCl in protein excess, (3) 150 mM KCl in protein excess, (4) 10 mM KCl in hairpin excess, or (5) 150 mM in hairpin excess. The reaction products were then fractionated on a 20% denaturing polyacrylamide gel and the bands visualized using Instant Imager. The cleavage products are indicated with arrows. A 42-nt RNA marker, obtained by alkaline hydrolysis of $R_{19}LR_{19}$, was run as control at the sides.

4.7 PROPOSED MECHANISM FOR RNA CLEAVAGE BY YEAST RNASE III

The above results shed light into the mechanism of Rnt1p processing of double stranded RNA and suggest that the 2'-OH group may not be directly involved in the cleavage reaction. A quick review of the literature involving RNA cleavage mechanisms reveals two major proposed models (**Figure 4.18**).²³⁸⁻²⁴¹ In both models, the role of the metal ion is greatly emphasized. A metal ion can coordinate a water molecule, and thus drastically lowers the pK_a value of the bound water.^{238,240} In the first model, the coordinated water (or hydroxide) molecule acting as a nucleophile can directly attack the phosphodiester moiety²³⁹ resulting in a 5'-phosphate monoester and a 3'-OH group. In the second proposed pathway (B), the same coordinated water (or hydroxide) molecule can act as a base that deprotonates the 2'-OH group and the newly generated 2'-alkoxide ion

can then attack the neighboring scissile phosphodiester linkage. This yields a 2',3'-cyclic phosphate and a free 5'-OH. Moreover, the metal ion, in either pathway, can assume a different role by directly coordinating (as a Lewis acid) to the anionic oxygen of the phosphodiester linkage via electrostatic interaction. This causes charge neutralization and consequently reduces the build-up of electrostatic repulsion between the incoming anionic nucleophile and the phosphate electrophile. By stabilizing the rate determining transition state, the activation barrier for the reaction is reduced and thus the nucleophilic attack is more easily achieved.²⁴¹ It is important to note that the metal ion can assume one or more of numerous catalytic roles within these two general cleavage mechanisms.

Taking into account the two models proposed above combined with our findings, one is tempted to conclude that the 2'-OH group plays a secondary role, if any, in the cleavage of RNA by yeast RNase III. In other words, if deletion of the 2'-OH group (to obtain deoxyuridine) does not impair nucleophilic cleavage of the scissile phosphate group 5'..UpC..3', one can rule out mechanism B (**Figure 4.18**) in which *intramolecular* cleavage is mediated by the 2'-OH group. This raises the interesting possibility that cleavage of RNA by yeast RNase III proceeds via mechanism A, whereby a coordinated water (or hydroxide) molecule is the key nucleophilic species in the cleavage reaction (*intermolecular* reaction). Whether the metal ion coordinates to the water molecule and/or the anionic oxygen of the phosphate group remains unknown. Additionally, experimental data show that masking the 2'-OH group with a bulky -CH₃ moiety slows down the cleavage rate relative to the deoxy case. We envision that the incoming coordinated water molecule might be attacking along the same side of the 2'-OH group (from the minor groove). The presence of a bulky methyl group poses a steric barrier along the trajectory of the incoming nucleophile, which might explain the decrease in the observed cleavage rate.

Considering the mechanism of action of other RNase enzymes such as RNase H, the enzyme cleaves the RNA strand in RNA:DNA hybrids to produce 5'-phosphate-monoester fragments (5'pRNA). The 2'-OH group is not directly involved in the cleavage reaction (*i.e.*, it does not produce a 2',3'-cyclic phosphate), but assists by binding to the water molecule that makes the nucleophilic cleavage at the scissile phosphate group (Pathway C, **Figure 4.18**).²⁴²

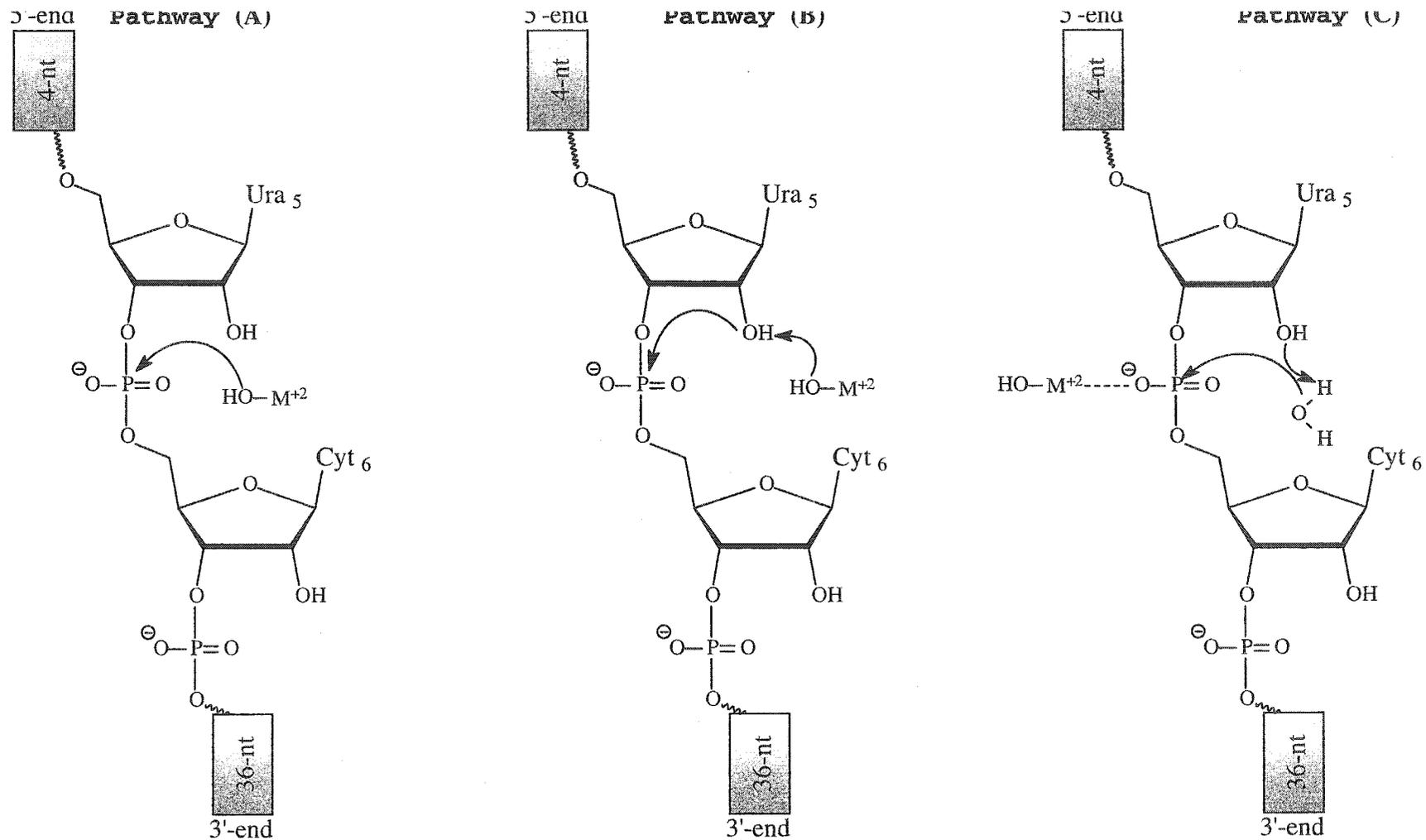


Figure 4.18: Proposed mechanism of nucleophilic cleavage of RNA. In pathway (A), the coordinated water (or hydroxide) molecule can act as a nucleophile that attacks directly the phosphodiester linkage, or in pathway (B) as a base that deprotonates the 2'-OH group to an alkoxide ion attacking intramolecularly the neighboring phosphate moiety. Pathway (C) involves RNase H cleavage mechanism and is the same as pathway (A) [*i.e.*, direct attack on the phosphate bond] except that the metal ion assumes a catalytic electrostatic role by coordinating to the anionic oxygen of the phosphate group.

This is unlikely in yeast RNase III since removal of the 2'-OH does not impair scissile phosphate hydrolysis. Otherwise, it should have affected the rate of nucleophilic cleavage, an effect that we did not detect.

We conclude that RNA cleavage by yeast RNase III likely involves an intermolecular *hydrolysis* reaction where the incoming nucleophile is a coordinated water (or hydroxide) molecule that attacks the scissile phosphodiester moiety.

4.8 OTHER REQUIREMENTS FOR RNT1P-MEDIATED DNA CLEAVAGE

The observation that there was no cleavage observed for R₁₉LD₁₉ [4.10] yet dominantly evident for D₁₉LR₁₉ [4.11] is quite intriguing since the RNA strand [in R₁₉LD₁₉] was expected to be cleaved. It suggests that the enzyme's cleavage activity is governed by the composition of the hairpin's 3'-end-stem. This led us to synthesize a hairpin containing five RNA and fourteen DNA residues at its 3'-terminus (D₁₉LR₅D₁₄ [4.15]; Figure 4.19). When incubated with Rnt1p in the presence of divalent magnesium ions, this hairpin shows absolutely no cleavage. Furthermore, binding to Rnt1p is drastically reduced relative to D₁₉LR₁₉. In a similar fashion, introducing five RNA residues upstream of the loop [D₁₄R₅LR₅D₁₄; 4.16], in the hope of increasing the binding affinity, did not promote cleavage (Figure 4.20). These results taken together with the fact that D₁₉LD₁₉ is not cleaved yet it binds to the enzyme suggest that the presence of RNA in the 3'-side of the stem is a necessary requirement to effect cleavage at the opposite DNA strand (5'-side of the stem).

4.9 DNA CLEAVAGE IN DNA:RNA HYBRIDS IS SPECIFIC TO EUKARYOTIC, BUT NOT PROKARYOTIC, RNASE III ENZYMES

Next we determined whether the ability of RNase III to cleave DNA in DNA:RNA hybrids is conserved among all family members of this enzyme. As a control sequence, we selected an RNA hairpin that is cleaved by all RNase III homologues, namely bacterial RNase III (*E.coli*), yeast RNase III (Rnt1p), and RNase III (Pac1). Its stem-loop sequence is identical to that found in the 3'-region of U5 snRNA (Figure 4.21). We also synthesized the DNA:RNA hybrid analogue of this hairpin and studied its ability to invoke RNase III cleavage by the various enzymes (Figure 4.21).

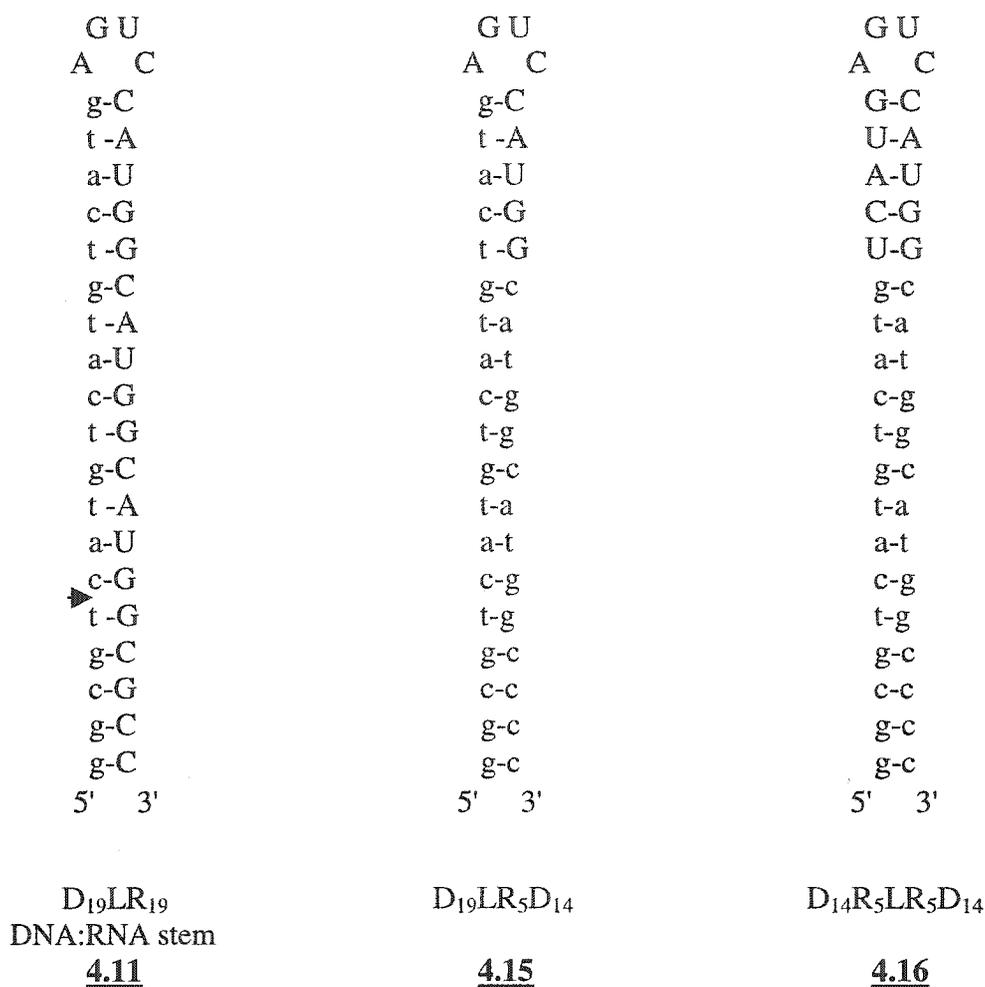


Figure 4.19: Chimeric DNA/RNA hairpin substrates (42 nucleotides in length) designed to probe the cleavage reaction by Rnt1p. They consist of nineteen base pairs in the stem and contain the tetranucleotide RNA loop sequence (AGUC). All substrates contain the same base sequence but differ in the sugar composition of the stem. The arrow indicates where cleavage takes place. Capital letters represent RNA residues and small letters represent DNA residues.

The gel autoradiogram in **Figure 4.22** shows that DNA cleavage is specific to eukaryotic, but not prokaryotic, RNase III enzymes. While the U5-like hairpin [4.17] is efficiently cleaved by all three enzymes, the DNA:RNA analogue [4.18] is cleaved only by the yeast RNase III (Rnt1p) and Pac1 enzymes. In case of Rnt1p, cleavage occurs at the DNA portion of the 5'-end stem, consistent with the DNA cleavage previously observed for D₁₉LR₁₉ [4.11]. Pac1 cleaves at a different position located at the DNA/RNA junction (neighboring the loop), thus producing a sixteen-nucleotide fragment. Remarkably, *E.coli* RNase III does not show any cleavage activity at all,

suggesting that there exist differences in the RNA cleavage mechanism and substrate requirements among these enzymes.²⁰⁹

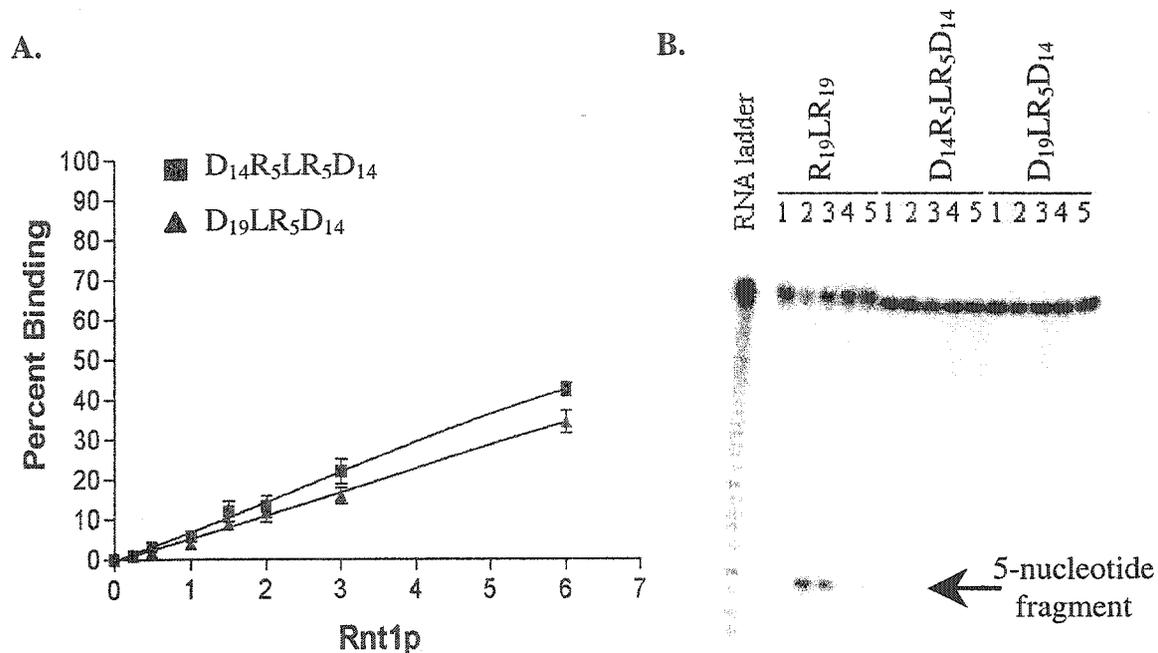
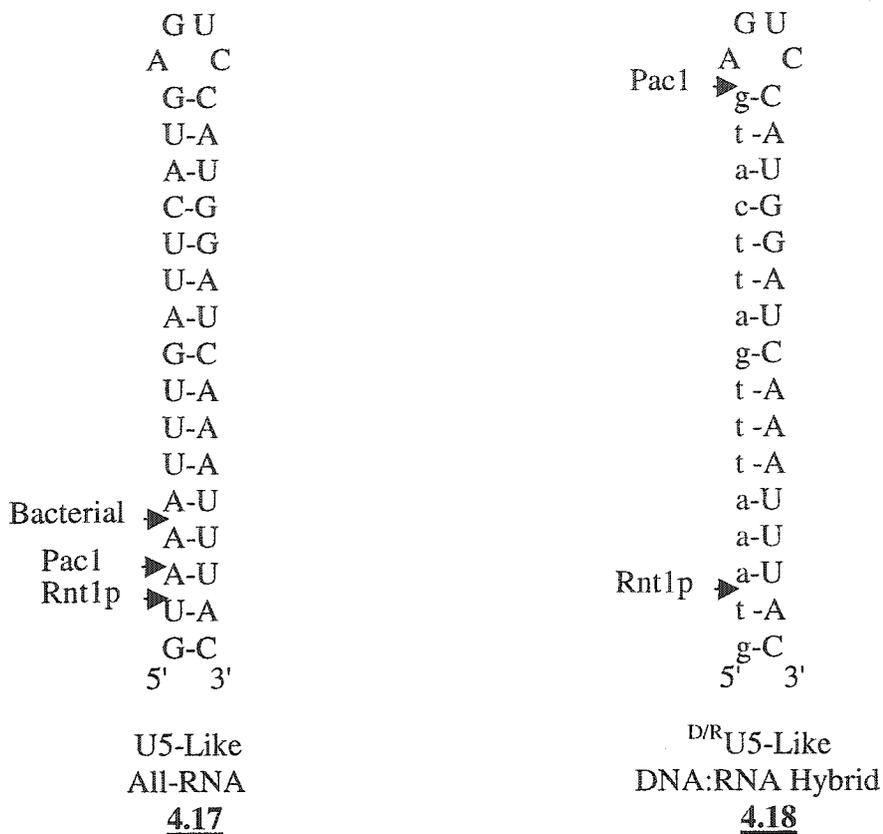


Figure 4.20: (A) Plot of % bound hairpin *versus* Rnt1p concentration and (B) cleavage assay of various long hairpin DNA/RNA chimeras (42-nt). The hairpin substrates $D_{14}R_5LR_5D_{14}$ and $D_{19}LR_5D_{14}$ were incubated with Rnt1p in the presence of Mg^{2+} and either (1) nothing, (2) 10 mM KCl in protein excess, (3) 150 mM KCl in protein excess, (4) 10 mM KCl in hairpin excess, or (5) 150 mM in hairpin excess. The reaction products were then fractionated on a 20% denaturing polyacrylamide gel and the bands visualized using Instant Imager. The cleavage products are indicated with arrows. A 42-nt RNA marker, obtained by alkaline hydrolysis of $R_{19}LR_{19}$, was run as control at the sides.



Code	Designation	T_m (°C)	%H
<u>4.17</u>	U5-Like	53.3	19.4
<u>4.18</u>	^{D/R} U5-Like	46.1	15.2

Figure 4.21: The base sequence of the control all-RNA substrate [U5-Like, **5.18**] and its modified DNA:RNA hybrid analogue [^{D/R}U5-Like, **5.19**]. The table compares their thermal stability (T_m), % hypochromicity (%H), and binding constant (K_d) towards yeast RNase III (Rnt1p), Pac1, and *E.coli* RNase III (Bacterial). The arrows indicate the cleavage sites by each RNase III enzyme.

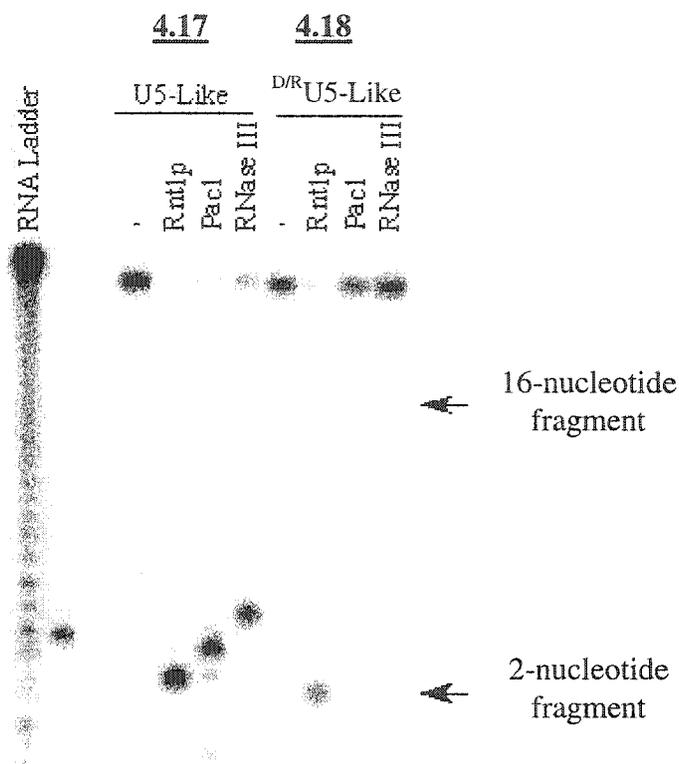
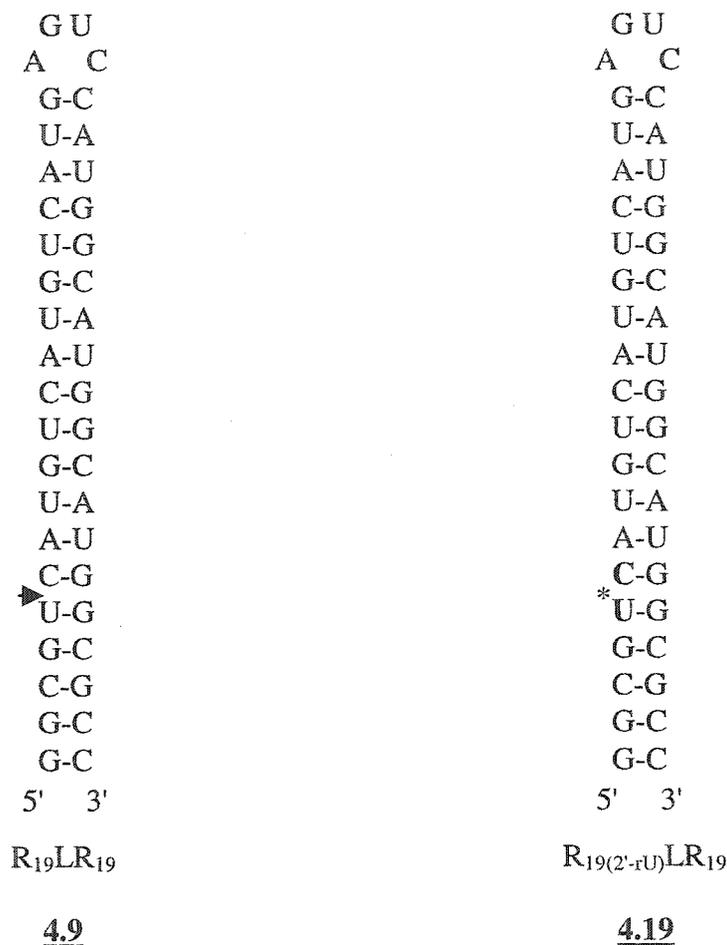


Figure 4.22: Cleavage assay of U5-like hairpins with various RNases III. The hairpin substrates were incubated with either yeast RNase III (Rnt1p), Pac1, or *E.coli* RNase III (bacterial). Rnt1p and bacterial RNase III were at 0,16 μM while Pac1 was at 0,32 μM in final enzyme concentration. The reactions were conducted in 10 mM KCl in the presence of Mg^{2+} . The reaction products were then fractionated on a 20% denaturing polyacrylamide gel. The cleavage products are indicated with arrows. A 42-nt RNA marker, obtained by alkaline hydrolysis of $\text{R}_{19}\text{LR}_{19}$, was run as control at the sides.

4.10 RNT1P DOES NOT CLEAVE A 2',5'-PHOSPHODIESTER LINKAGE

Another site for potential modification is the scissile phosphodiester bond itself. To gain further insight into the mechanism of cleavage and substrate specificity of Rnt1p, we have inserted a 2',5'-linkage fourteen nucleotides away from the loop at the exact location where cleavage occurs (**Figure 4.14A**). The presence of a single 2',5'-phosphodiester linkage within the 19 base pair stem helix drastically decreases the thermal stability of the hairpin [$\Delta T_m = \text{ca. } 7 \text{ }^\circ\text{C}$; **Figure 4.23**]. However, it does not disrupt the hairpin conformation which maintains an A-form helical structure as

evidenced by CD (Figure 4.15). Although the hairpin binding affinity towards Rnt1p is not compromised, the 2',5'-phosphodiester linkage is completely resistant to Rnt1p RNA processing (Figure 4.24).



Code	Designation	K_d (μ M)	T_m ($^{\circ}$ C)	%H
<u>4.9</u>	R ₁₉ LR ₁₉	0.67	69.8	18.3
<u>4.19</u>	R _{19(2'-rU)} LR ₁₉	0.95	62.9	14.6

Figure 4.23: Schematic representation of the hairpin substrate containing a 2',5'-phosphodiester linkage at the cleavage site (14 nucleotides away from the loop on the 5'-end stem). The control all-RNA substrate [R₁₉LR₁₉, **5.18**] is also shown for comparison purposes. The table compares their thermal stability (T_m), % hypochromicity (%H), and binding constant (K_d) towards yeast RNase III (Rnt1p). U*C = U_{2'p5'}C. The arrow indicates cleavage site on the wild-type RNA hairpin.

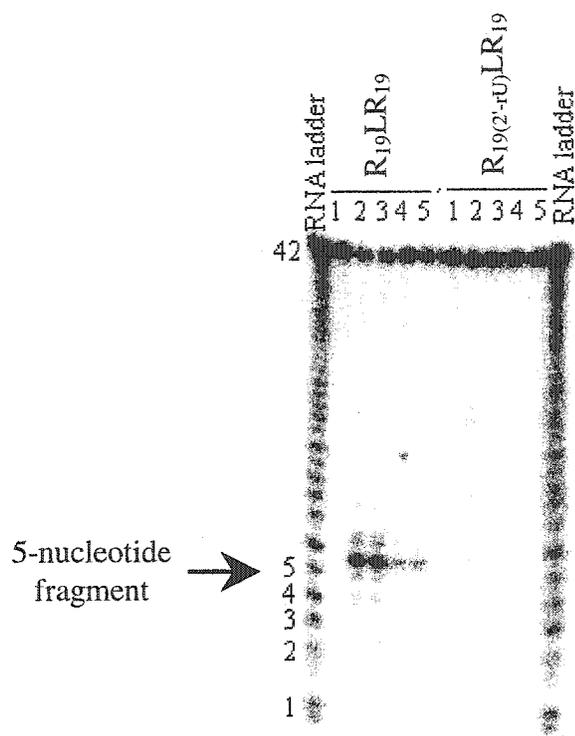


Figure 4.24: Cleavage assay of the hairpin $[R_{19(2'-rU)}LR_{19}]$ with a 2',5'-phosphodiester linkage at the cleavage site. The unmodified hairpin $R_{19}LR_{19}$ is shown for comparison purposes. The hairpins were incubated with Rnt1 protein in the presence of Mg^{2+} and either (1) nothing, (2) 10 mM KCl in protein excess, (3) 150 mM KCl in protein excess, (4) 10 mM KCl in hairpin excess, or (5) 150 mM in hairpin excess. The reaction products were then fractionated on a 20% denaturing polyacrylamide gel and the bands visualized using Instant Imager. The cleavage products are indicated with arrows. A 42-nt RNA marker, obtained by alkaline hydrolysis of $R_{19}LR_{19}$, was run as control at the sides.

These results somewhat contrast previous studies showing that a 2',5'-linkage within an RNA duplex is cleaved, by the vicinal alkoxide, 900 times faster than the corresponding 3',5'-linkage.¹¹² The authors ascribe this to the proper positioning of the 3'-OH group which is readily aligned for intramolecular nucleophilic attack at the phosphate moiety. We observe the opposite trend when the duplexes are exposed to Rnt1p, *i.e.*, the 2',5'-linkage is by far more hydrolytically stable than the 3',5'-linkage. This is consistent with our proposal that RNase III-mediated RNA cleavage occurs via an *intermolecular* rather an *intramolecular* cleavage mechanism. In other words, the vicinal hydroxyl group on the ribose sugar does not participate directly in the cleavage reaction.

4.11 CONCLUSIONS

We synthesized hairpins composed of a 9 bp stem and a loop of the general consensus sequence (AGNN). The properties of these hairpins were explored in an attempt to elucidate the molecular requirements for substrate binding to yeast RNase III (Rnt1p). Stems composed of RNA:DNA hybrids or DNA duplexes bound very weakly to Rnt1p [R₉LD₉ & D₉LD₉]. The effect of loop on hairpin binding to Rnt1p was also explored. A hairpin [R₉LR₉] containing an RNA duplex with a 2',5'-RNA (AGUC) loop did not bind at all. Sequence mutation studies conducted within the (AGUC) loop revealed no significant effect on thermal stability. In fact, the loops (AGAA), (ACUC), (GAAA), and (CGUC) exhibited the same order of thermal stability. However, while (ACUC) and (GAAA) loops failed to bind to Rnt1p, (CGUC) was surprisingly recognized by the enzyme. Overall, these studies showed that the stem composition and loop base sequence as well as their conformation are important determinants of substrate recognition by Rnt1p.

We also synthesized 19 bp hairpins containing the (AGUC) loop in order to test their ability to invoke cleavage by yeast RNase III. Rnt1p did not cleave DNA duplexes. However, DNA:RNA hybrids were cleaved by Rnt1p depending on the location of the RNA residues in the hairpin stem. Hairpin D₁₉LR₁₉ was cleaved while R₁₉LD₁₉ was not, although both displayed the same overall conformation as judged from CD experiments. Interestingly, yeast RNase III cleaves the DNA portion of hairpin D₁₉LR₁₉ [4.11]. DNA cleavage is dependent on the RNA content of the complementary strand. DNA cleavage in DNA:RNA hybrids is specific to eukaryotic but not prokaryotic RNase III enzymes. These findings demonstrate that there exist differences in the RNA cleavage mechanism and substrate requirements among the various RNase III enzymes. To gain further insight into the mechanism of cleavage by Rnt1p, we synthesized the substrates R₁₉(dU)LR₁₉ and R₁₉(2'-OMe)LR₁₉, containing 2'-deoxy uridine and 2'-OMe uridine modifications, respectively, at the cleavage site. Both substrates were cleaved by Rnt1p, which prompted us to conclude that the 2'-OH group does not participate in the nucleophilic attack on the vicinal 3',5'-phosphodiester linkage (5'..UpC..3'). Rnt1p is incapable of cleaving a 2',5'-phosphodiester linkage. We propose that the cleavage reaction takes place through an *intermolecular* rather than *intramolecular* mechanism. The incoming

nucleophile is a coordinated water (or hydroxide) molecule that attacks the scissile phosphodiester bond directly.

4.12 BIOLOGICAL SIGNIFICANCE

The simple demonstration that a dsRNA specific protein binds a DNA duplex and cleaves DNA within a DNA/RNA hybrid is quite stunning because it defies our view of dsRNA binding proteins and their behavior. Rnt1p, the yeast orthologue of the bacterial RNase III and a kin of Dicer the mammalian gene involved in RNAi, cleaves DNA within a DNA/RNA duplex in a very efficient and specific manner. Comparing the activities of the three different RNase III enzymes including Rnt1p, bacterial RNaseIII, and fission yeast *pac1* shows that both eukaryotic enzymes cleave the DNA/RNA hybrid while Rnt1p is the only enzyme that cleaves within the DNA strand. To the best of our knowledge, this is the first demonstration of DNA cleavage by an RNase especially a member of the dsRNA-specific RNase III family. This observation opens a completely new window of investigation since this demonstrated RNA-dependent DNA cleavage may lead to the discovery of new biological functions of dsRNA-specific proteins. The results also allude to the possible existence of a “mirror image” of the RNase H activity (which cleaves the RNA strand of DNA/RNA hybrids). Recognition of DNA/RNA hybrid by a Dicer orthologue underlines the possibility that a modified RNA could be used to target Dicer activity. In addition, using Dicer for the purpose of cleaving duplex DNA may now be considered a possibility (in progress). Ongoing experiments have also shown that DNA/RNA hybrids can be cleaved in cell extracts by this enzyme, a strong indication of the *in vivo* utility of our discovery.

CHAPTER V: ASSOCIATION OF CYTOSINE-RICH HAIRPINS INTO THE I-MOTIF

5.1 BACKGROUND

In the early 1960s, it was proposed that chains of polycytidylic acid would associate into parallel duplex structures at acidic pH (low pH).²⁴³ Since then, conflicting reports based on X-ray diffraction and NMR studies have emerged in an attempt to explain the nature of the complex formed by polyribocytidylic acid.²⁴⁴⁻²⁴⁶ Studies conducted on oligomers containing deoxycytosine²⁴⁷⁻²⁵⁰ seemed to coincide in opinion about the existence of a parallel stranded duplex or multimeric structure containing cytosine-protonated cytosine (C·C⁺) base pairs (**Figure 5.1**). It was not until 1993 that this enigma was solved when the first detailed NMR structure of cytosine rich sequences was reported by Gueron and coworkers.²⁵¹ They show that the oligomer d(TCCCCC) forms a tetrameric four-stranded DNA structure held together by (C·C⁺) base pairs. This tetrad is folded in a symmetrical way in which two parallel-stranded duplexes *intercalate* in opposite orientation with one another (**Figure 5.2**). Hence the name *i*-motif or *i*-DNA.²⁵¹ One year later, the X-ray structure of d(CCCC) was reported by the Rich laboratory and proved to be in agreement with the proposed *i*-motif.²⁵²

A comparison of the NMR solution and X-ray structures published to date on sequences forming tetrad DNA reveals similar helical and structural properties.²⁵¹⁻²⁵⁵ The tetrad is stabilized by a network of inter residual-sugar hydrogen-bonding and van der Waals interactions as well as electrostatic contacts between the anionic phosphodiester bonds and the hemiprotonated C·C⁺ base pairs.^{251,256,257} Furthermore, base stacking interactions between C·C⁺ base pairs tend to stabilize the tetrad structure.²⁵¹ Because *i*-DNA is composed of four strands, two minor and two major grooves exist.²⁵² In the pioneering NMR study by Gueron, long range NOE crosspeaks between sugar H1' protons (H1'-H1') were recognized, which suggested a close proximity of the furanose moieties within the tetrad structure.²⁵¹ Such NOEs were not observed in any other nucleic acid structure, and thus were in later studies assumed, among other characteristics, to be one of the standard criteria that distinguish *i*-motif structures.

Cytosine-Protonated Cytosine Base pair
C·C+

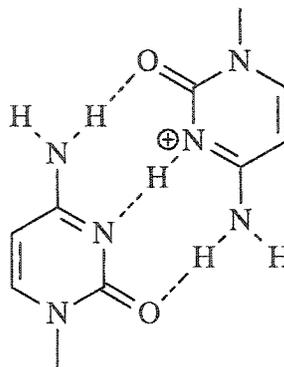


Figure 5.1: Schematic representation of a hemiprotonated C·C⁺ base pair. The base stacking interaction involves the exocyclic amino and carbonyl groups.

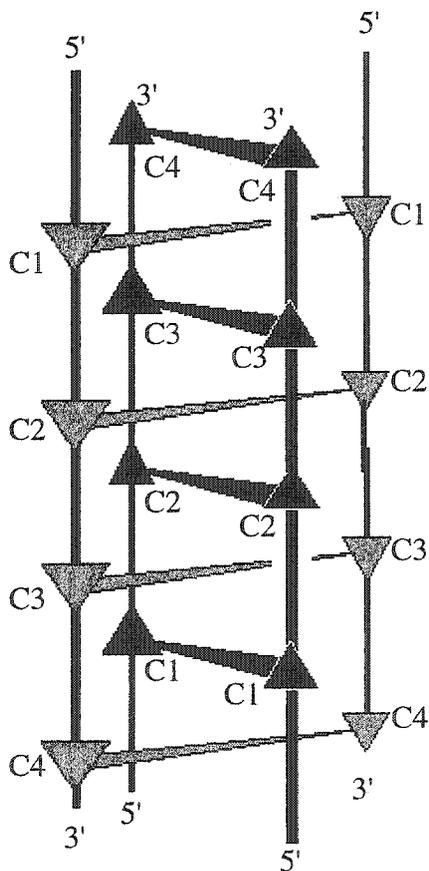


Figure 5.2 A: Schematic representation of i-motif DNA (C-tetrad). The structure is highly symmetrical and involves C·C⁺ base pairing. It consists of two intercalating parallel stranded duplexes in opposite orientation to each other.

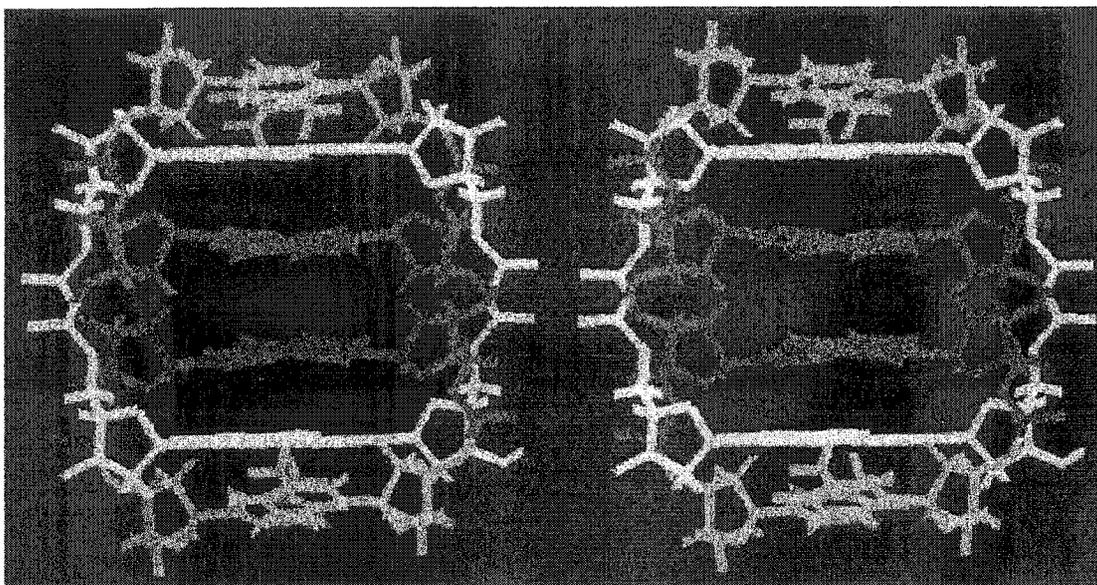
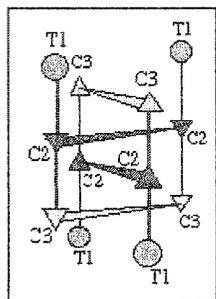


Figure 5.2 A: Stereoview of the NMR-derived structure of an i-motif DNA (C-tetrad). The stoichiometry of the structure is $d(\text{CCT})_4$. Adapted from Leroy, J.-L.; Gueron, M. *Structure* **1995**, *3*, 101-120.

Among the distinctive features present within the i-motif is the unusual C-C+ interaction. This base pairing motif was first elucidated by Rich and Hartman through IR and UV studies conducted on polyribocytidylic acid with hemiprotonated cytosine base pairs.²⁵⁸ The structure under investigation exhibited two different pK_a 's, one at 3.0 and the other at 5.7. In both, the uptake of one proton per cytosine base pair was observed, thus allowing the formation of a third hydrogen bond which stabilized the interaction between the base pairing cytosine residues (**Figure 5.1**). Addition of a second proton impaired stability since it obstructed the formation of the third hydrogen bond. The authors also observed that the stability of the duplex decreased at high salt

concentrations, which they ascribed at the time to interference of the increased ionic strength with the magnitude of electrostatic interaction between the C·C+ base pair and the anionic phosphodiester linkage.²⁵⁸ Consistent with this notion, studies conducted by Guschlbauer revealed that the complex formed by polyribocytidylic acid was dependent on solution pH, with an apparent polymorphic behavior observed in the pH margin 4.0 to 5.0 upon thermal denaturation.²⁵⁹ An inspection of the reported pK_a values reveals that cytosine has a pK_a of 4.3.²⁶⁰ This implies that during thermally-induced complex dissociation, deprotonation occurs at pH's higher than 4.3 while protonation of unprotonated cytosines takes place at pH's lower than 4.3. The two studies of Guschlbauer²⁵⁹ on the one hand and of Rich and Hartman²⁵⁸ on the other hand, show that high salt concentrations destabilize the complex formed by polyribocytidylic acid at pH's higher than 4.5, but stabilize it at pH's lower than 4.5, the results of which are dependent on the form of the complex at various pH's.

5.2 ASSOCIATION OF OLIGONUCLEOTIDES INTO THE I-MOTIF

Tetramolecular Association

The i-motif structure can be formed from one, two or four oligomers depending on the base sequence and architecture of the associating strands. For example, four strands of cytidine can come together to form one of two possible i-motif cores (**Figure 5.3**). These differ in the intercalation topologies in which the terminal C·C+ base pair can be present at either the 3'-end or the 5'-end of the stretch (structures 3'E or 5'E respectively, **Figure 5.3**). Structures bearing different intercalation topologies may still exhibit similar stabilities as well as a variety of other internal motions. Each of the oligomers $d(A_2C_4)$ ²⁵⁵ and $d(C_4T)$ ²⁶¹ for example have been shown to form two tetrameric i-motif structures with 3'E and 5'E topologies. In $d(A_2C_4)$, like in any other C-tetrads, the four cytosine strands associate into two parallel-stranded duplexes with intercalated hemiprotonated C·C+ base pairs, thus forming four helical grooves (two broad and flat major grooves and two narrow minor grooves). The only difference from other C-tetrads lies in the unique arrangement of the terminal adenines and their contribution to the overall tetrad stacking.²⁵⁵

Bimolecular Association

A single stranded cytosine-rich oligomer can also interact bimolecularly to form a dimeric i-motif structure (Figure 5.3B). For instance, the oligomer d(TC₃GT₃C₂A) with a [GTTT] loop intercalates in a head-to-head fashion,²⁵³ while the oligomer d(5mCCT₃AC₂) with a [TTTA] loop forms a dimeric i-motif structure in a head-to-tail pattern.²⁶² In both examples, the overall helical features such as geometry of the base pairs and orientation of the strands are similar to those typical of i-motif tetrads. The differences rather lie in the structure of the loops and their intrinsic tertiary interactions. Furthermore, Damha and coworkers have shown that joining the 5'-ends of two pentadeoxycytidine stretches by a branching riboadenosine (rA) linker promotes the formation of a tetrad structure of high thermal stability.^{263,264} This arrangement causes the orientation of the dC strands to be parallel and forces the formation of C·C+ duplex that associates bimolecularly into the i-motif.

Unimolecular Association

In the third case, a single strand can fold intramolecularly to form a monomeric i-motif structure. An illustration of this is the sequence d(5mCCT₃CCT₃ACCT₃CC).²⁵⁴ The [TTTA] loop spans across one of the wide major grooves. The i-motif core is well defined and extended by a T·T pair and a A·T pair that independently stack next to the C·C+ base pairs.

From these three different i-motif structures, one can deduce that the choice of the linker plays a central role in defining the folding pattern and the molecularity of association. The intercalation and loop topologies highly depend on the nature and base sequence of the loop, a moiety that affects the degree of symmetry within the tetrad and which may in turn promote formation of other possible multimeric structures.^{265,266}

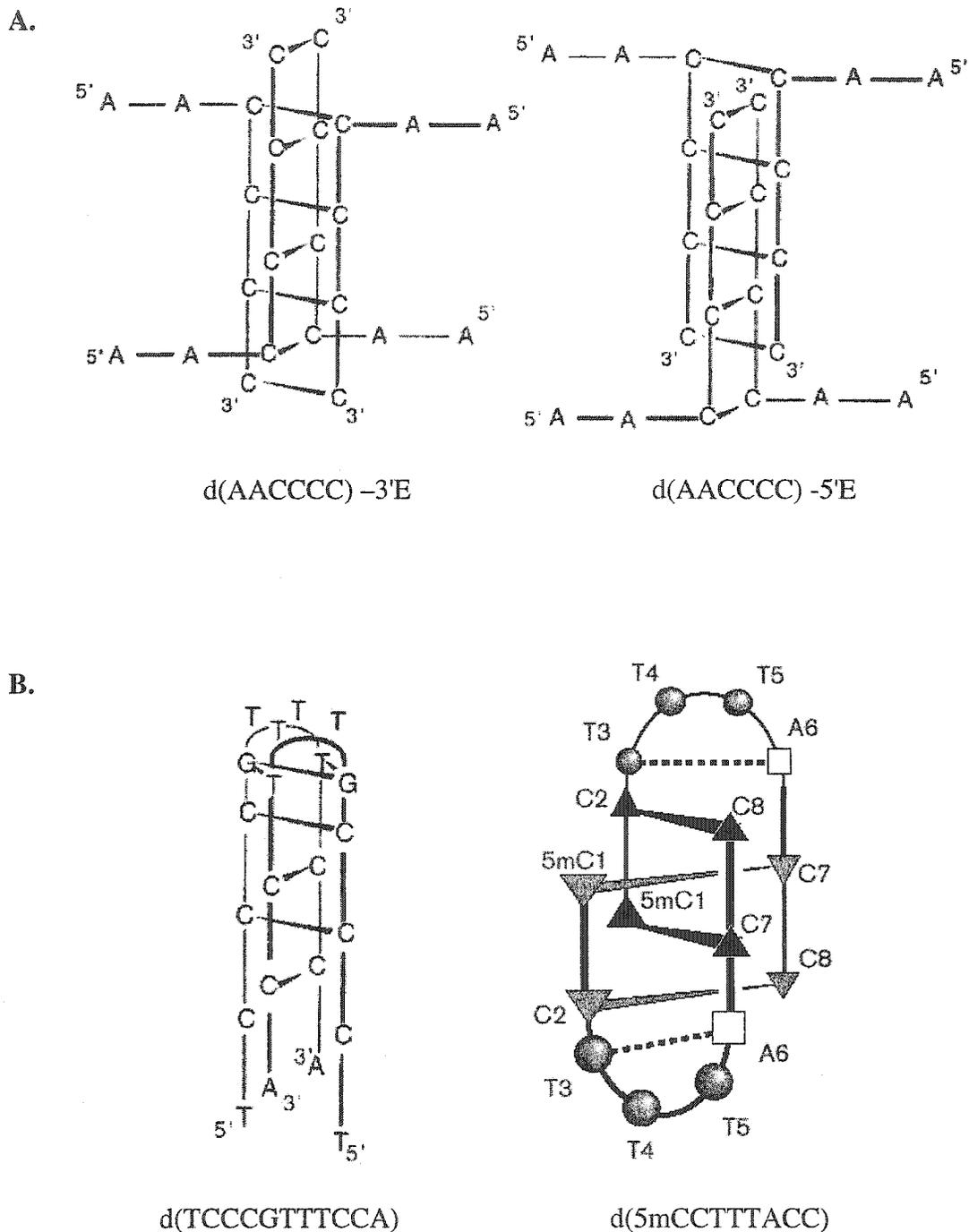


Figure 5.3 A and B: Representative diagrams showing topologies and stoichiometries of various possible i-motif structures. In (A), four strands come together to form an i-motif which can be in 3'E and/or 5'E topologies. In (B), two hairpins can associate intermolecularly to form a dimeric i-motif structure exhibiting 3'E and/or 5'E topology. The linker can be a loop or a branching point (to form a branched oligomer). The two molecules arrange in either head-to-head or head-to-tail.

C.

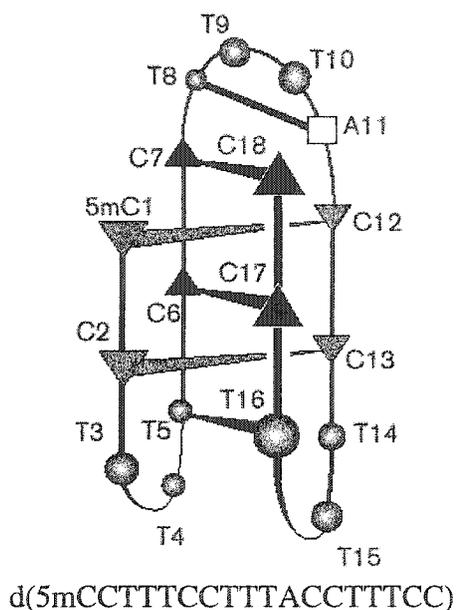


Figure 5.3 C: Representative diagram showing intramolecular *i*-motif formation. A single stranded cytidine-rich oligomer folds back on itself to form a C-tetrad structure.

5.3 PROJECT OVERVIEW

In this chapter, we designed C-rich oligonucleotides containing extra stable loops as a means to facilitate the formation of *i*-motif structures. Accordingly, we investigated the possibility of formation of dimeric C-tetrads from hairpins with both 2',5'- and 3',5'-linked (UUCG) loops. We also explored the ability of 3',5'-RNA to associate into duplexes that are held together by hemiprotonated C·C⁺ base pairs. Moreover, the ability of 2',5'-RNA to form C·C⁺ base pairs, in particular *i*-motif structures, was investigated by a combination of UV and CD spectroscopy as well as gel electrophoresis.

5.4 HAIRPINS WITH DEOXYCYTIDINE STEMS

Hairpins containing uniform deoxycytidine stems (5 base pairs) and incorporating either the 3',5'- or 2',5'-RNA (UUCG) loop were synthesized and studied for their ability to associate into the *i*-motif (Table 5.1). In some cases, a C·G loop-closing base pair was introduced in the hairpin sequence in order to promote folding. Hairpin $c_5\underline{R}c_5$, for example, contains a 2',5'-RNA loop and a pentadeoxycytidine stem. Hairpin $c_5\underline{R}gc_4$ is the same as $c_5\underline{R}c_5$ except that the deoxyC residue downstream from the loop is replaced with deoxyG. We envisioned that the dG residue would act as a nucleation site that assists in

structure folding by promoting C·G base pair formation. The “V”-shape oligomer [dV-5] has been shown to form a branched i-motif structure,²⁶³ and is included in this study for comparison purposes. The TRT oligomer contains deoxythymidylate stem and serves as a negative control.

Table 5.1: Sequences of deoxycytidine rich oligomers under study

Code	Designation	5'- Oligomer Sequence -3'
<u>5.1</u>	dC ₃	5'- ccc -3'
<u>5.2</u>	dC ₅	5'- ccccc -3'
<u>5.3</u>	c ₅ <u>R</u> c ₅	5'- ccccc(<u>UUCG</u>)cccc -3'
<u>5.4</u>	c ₅ Rc ₅	5'- ccccc(UUCG)cccc -3'
<u>5.5</u>	c ₅ <u>R</u> gc ₄	5'- ccccc(<u>UUCG</u>)gcccc -3'
<u>5.6</u>	c ₅ Rgc ₄	5'- ccccc(UUCG)gcccc -3'
<u>5.7</u>	c ₄ CRGc ₄	5'- ccccC(UUCG)Gcccc -3'
<u>5.8</u>	c ₅ <u>U</u> gc ₄	5'- ccccc(<u>UUUU</u>)gcccc -3'
<u>5.9</u>	dV-5	$\begin{array}{l} \text{rA} \begin{array}{l} \nearrow 2',5' \text{ ccccc -3'} \\ \searrow 3',5' \text{ ccccc -3'} \end{array} \end{array}$
<u>2.31</u>	TRT	5'- tttt(UUCG)tttt -3'

Underlined nucleotides represent 2',5'-RNA residues (*e.g.*, UC = U_{2p5}C_{2p}). Capital letters represent 3',5'-RNA residues, while small letters represent DNA residues. Section I represents the control sequences dC₃ and dC₅. Section II represents deoxycytidine oligomers with 3',5'- and 2',5'-(UUCG) loops. Sections III and IV contain a CG loop-closing base pair. The reference molecule dV-5 [5.9] is taken from [Robidoux *et al. J. Biomol. Struct. Dyn.* **1997**, *15*, 517-527], and is shown here for continuity and comparison purposes of the work. The TRT oligomer serves as a negative control and exists in a single-stranded state (it does not associate into any hairpin species).

5.4.1 UV Thermal Melting Studies and Dependence on pH

As a first step towards studying the ability of various hairpins to associate into the i-motif, they were investigated by UV spectroscopy for their ability to form hypochromic structures under different buffer conditions. Pentadeoxycytidilic acid (dC₅) was used as a control oligomer, thus providing a good model system to compare the thermal properties of our molecules. This oligomer is very similar to the well-characterized dC₄²⁵² and TdC₅ systems.²⁵¹

Figure 5.4 shows the UV melting curves of representative hairpins c₅Rc₅ [5.3] and c₅Rgc₄ [5.5] along with dC₅ [5.2] in 50 mM sodium citrate buffer at pH 4.6. Table 5.2 presents their thermal melting data. It is clear that the complexes formed by oligomers with loop moieties are of higher thermal stability relative to those formed by dC₅. Also, there is no effect observed on the magnitude of hypochromicity suggesting the existence of similar C·C+ base pair stacking interactions within these complexes. This is in sharp contrast to the dV5 complex exhibiting reduced hypochromicity consistent with the presence of fewer C·C+ base pairs in the branched complex.²⁶⁷ The close proximity of the nucleosides adjacent to the rA branchpoint obstructs them from efficient base pairing, thus lowering hypochromicity.^{160,267} Also, the degree of thermal stabilization brought about by the (UUCG) loop moiety is of higher order relative to that brought about by the branching rA linker (compare ΔT_m ; Table 5.2). The enhanced thermal stability observed by c₅Rc₅ [5.3] and c₅Rgc₄ [5.5] can be attributed to the existence of the loop which may play a role in strand preorganization by helping the folding process and/or reducing the molecularity of the association reaction. It is important to note that the complex observed by these oligomers is not exclusively a result of loop folding and/or stacking, but most probably an outcome of C·C+ base pairing. This can be said in view of the fact that the oligomer TRT [2.31] with a uniform thymidylate stem does not display any melting transition.

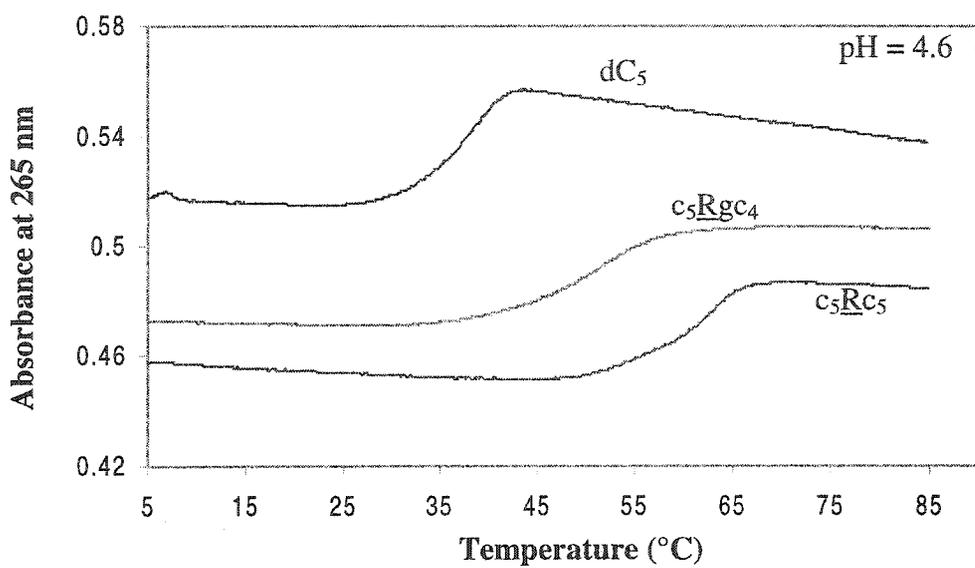


Figure 5.4: UV melting curves of representative deoxycytidine rich oligomers. Measurements were conducted at 265 nm in 50 mM sodium citrate buffer, pH 4.6. Oligomer concentrations were 2.3 μM for c_5Rc_5 [5.3] and c_5Rgc_4 [5.5], and 4.6 μM for dC_5 .

Table 5.2: UV thermal melting data for deoxycytidine rich oligomers

Code	Designation	T_m ($^{\circ}\text{C}$)	%H	ΔT_m ($^{\circ}\text{C}$)
<u>5.2</u>	dC_5	36.8	8.4	0
<u>5.3</u>	c_5Rc_5	60.1	8.4	+ 23.3
<u>5.5</u>	c_5Rgc_4	49.6	8.3	+ 12.8
<u>5.9</u>	dV-5	-	-	+ 3
<u>2.31</u>	TRT	-	-	-

Measurements were recorded at 265 nm in 50 mM sodium citrate, pH = 4.60 ± 0.02 . Oligomers c_5Rc_5 and c_5Rgc_4 were 2.3 μM while dC_5 was 4.6 μM in total concentration. All samples exhibited monophasic and cooperative melting curves. The values represent the average of at least two independent measurements. The melting temperature (T_m) was calculated according to the base line method. The error in melting temperature is ± 0.5 $^{\circ}\text{C}$.

Melting Temperature (T_m) Dependence on pH

Next, we wanted to confirm the presence of C·C⁺ base pairing. Towards this end, we studied the UV melting behavior of the various oligomers as a function of pH. Since pK_a of cytosine is 4.3 and since C·C⁺ pairing involves protonation of N3 (Figure 5.1), varying the pH would be expected to affect C·C⁺ base pair formation and thus helical stability (reflected in the value of T_m). Accordingly, we monitored the melting behavior of the oligomers in 10 mM acetate buffer (pH 3.5 – 5.5) and in 10 mM cacodylate buffer (6.0 – 7.0). All samples were incubated for one week at 4 °C before the melting temperature experiments. Table 5.3 gives the T_m and %H values as a function of pH and Figure 5.5 plots the dependence of T_m on pH. All complexes show a distribution of T_m values centered around an optimum pH, *i.e.*, the stability is reduced at pH's lower or higher than the optimum. This supports the existence of hemiprotonated C·C⁺ base pairs evident in i-motif structures.²⁶⁸ The maximum thermal stability for the linear dC₅ complex is observed at *ca.* pH 4.5. The optimum stability of the oligomers with a (UUCG) loop is slightly shifted and occurs at pH 4.5-5.0.

c₅Rc₅ [5.3] and c₅Rc₅ [5.4] are more thermally stable than dC₅ across the entire pH range (for a representative example, see Table 5.3). They even form stable complexes at neutral pH, while dC₅ does not. For instance at pH 4.5, there is significant stabilization upon linking the deoxycytidine strands with either a 2',5'-RNA loop ($\Delta T_m = 25$ °C) or a 3',5'-RNA loop ($\Delta T_m = 13$ °C). Additionally at any pH, c₅Rc₅ with a 2',5'-RNA loop has a higher thermal stability than c₅Rc₅, suggesting that the 2',5'-RNA loop accommodates better the deoxycytidylate stems.

Hairpins with a C·G pair at the loop

A similar trend is observed for c₅Rgc₄ and c₅Rgc₄ where both exhibit a higher thermal stability than dC₅, with the complex formed by the 2',5'-RNA loop displaying the highest stability (Table 5.3 and Figure 5.5). At pH 3.5-4.0, dC₅ and c₅Rgc₄ are of the same order of stability, while c₅Rgc₄ and c₅Rgc₄ are equally stable at pH 5.5-6.0. At neutral pH, a complex is detected with the c₅Rgc₄ showing high stability. dC₅ does not form any complex at pH>5.0 (Table 5.3 and Figure 5.5).

Table 5.3: UV melting data of deoxycytidine rich oligomers under study

pH	dC ₅ [5.2]		c ₅ Rgc ₄ [5.5]		c ₅ Rc ₅ [5.4]		c ₅ Rgc ₄ [5.6]		c ₄ CRGc ₄ [5.7]		c ₅ Ugc ₄ [5.8]	
	T _m (°C)	%H	T _m (°C)	%H	T _m (°C)	%H	T _m (°C)	%H	T _m (°C)	%H	T _m (°C)	%H
3.5	23.2	9.49	26.6	8.7	28.0	8.1	21.2	6.0	20.0	4.4	26.0	7.4
4.0	33.3	12.3	38.2	9.6	39.8	10.8	34.3	5.7	33.0	5.0	40.2	8.9
4.5	35.3	12.2	43.4	9.7	49.0	6.0	39.5	6.3	39.1	5.1	46.7	8.6
5.0	33.6	9.5	43.6	9.1	51.0	6.7	40.6	4.0	39.5	6.6	47.5	8.8
5.5	28.8	11.5	39.7	8.8	48.7	5.9	39.3	3.9	36.6	4.9	45.2	7.6
6.0	-	-	34.5	8.3	43.13	6.0	33.8	3.2	34.4	2.8	40.9	7.4
6.5	-	-	30.2	5.8	36.1	3.3	26.9	1.4	28.3	2.1	36.2	3.5
7.0	-	-	24.4	3.2	29.3	4.5	14.7	1.2	22.4	2.4	28.0	0.8

The control sequence dC₅ is shown for comparison purposes. All measurements were recorded at 265 nm in either 10 mM sodium acetate (pH 3.5-5.5) or 10 mM sodium cacodylate buffer (pH 5.5-7.0). Oligomer concentrations were 2.3 μM except for dC₁₀ (4.6 μM). The UV melting temperatures were computed according to the base line method. All values represent the average of two independent measurements. The error in melting temperature is ± 0.5 °C. Percentage hypochromicity (%H) was calculated from UV absorbances at low (A₀) and high (A_f) temperatures according to the following equation: %H = (A_f - A₀) / A_f.

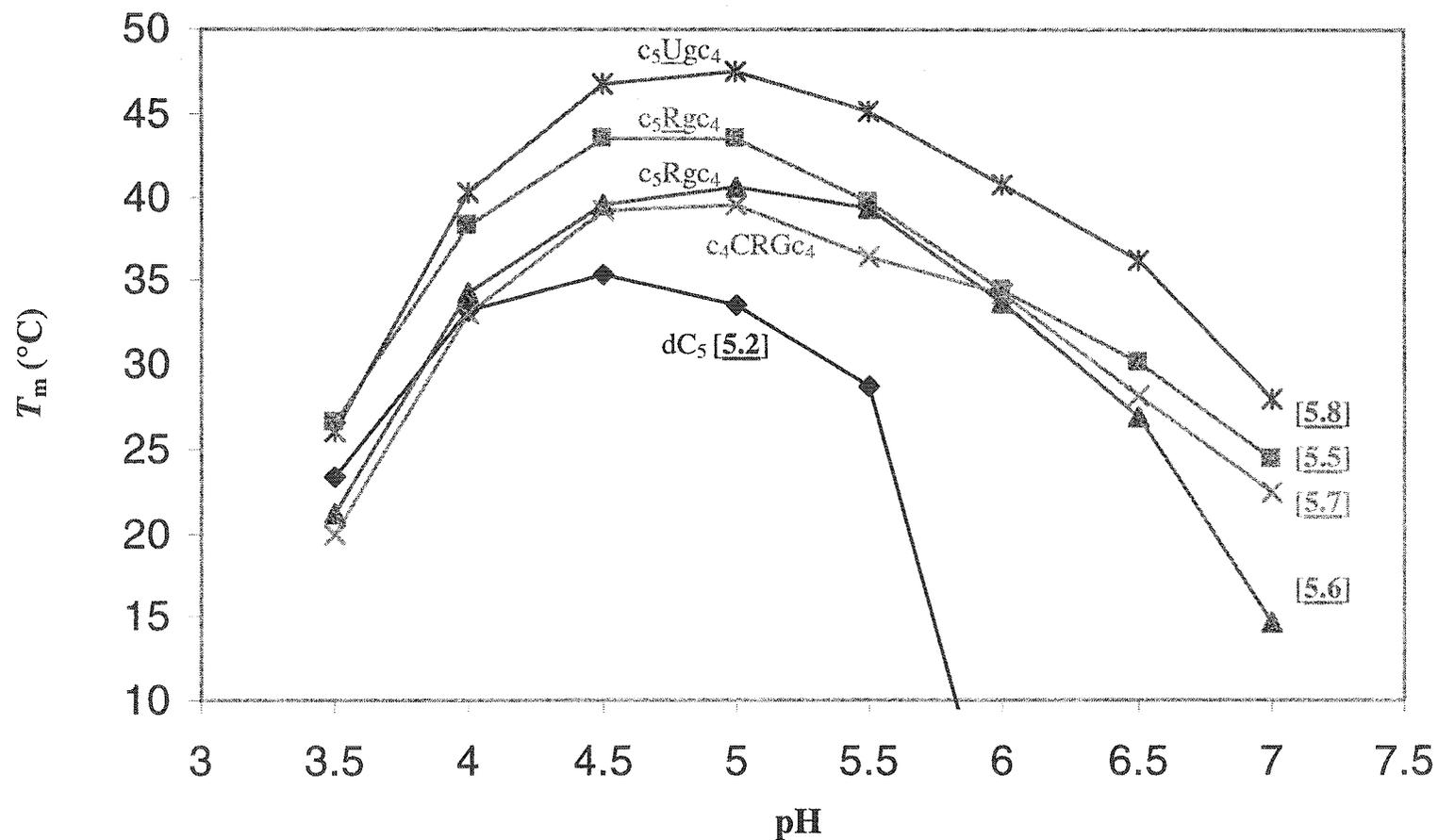


Figure 5.5: Plots of T_m versus pH for deoxycytidine rich oligomers containing G insertions. The optimum T_m for dC₅ [5.2] occurs at pH 4.5, while those of c₅Rgc₄ [5.5], c₅Rgc₄ [5.6], c₄CRGc₄ [5.7], and c₅Ugc₄ [5.8] occur at pH 4.5-5.0. dC₅ was 4.6 μ M in concentration, while all other oligomers were 2.3 μ M. All measurements were recorded at 265 nm in either 10 mM sodium acetate (pH 3.5-5.5) or 10 mM sodium cacodylate buffer (pH 5.5-7.0).

Changing the deoxy C·G loop closing base pair to riboses does not significantly increase the thermal stability of the complex containing the 3',5'-RNA loop. c_4CRGc_4 [5.7] and c_5Rgc_4 [5.6] show the same order of stability across the pH range under scrutiny. c_5Rgc_4 is more stable than c_4CRGc_4 at only pH 5.5 and 7.0 (Table 5.3 and Figure 5.5).

In comparison with c_5Rc_5 and c_5Rc_5 , inserting a G residue as a means to inducing better folding propensity inflicts both positive and negative effects. The advantage is that it rules out the chance for formation of other multimeric species as judged from the thermal melting curves (only a monophasic transition is detected after 48 hr). On the contrary, it unexpectedly lowers the thermal stability, though keeping it higher than that of dC_5 . The formation of a C·G base pair closing the loop may constrict flexibility and number of degrees of freedom needed to attain optimum complex arrangement. One might argue that other factors may come into play such as changes in sugar pucker mode.

Effect of Loop Base Sequence

In order to test the effect of loop base-sequence, the (UUCG) loop of the most stable complex in this series was mutated to (UUUU) and the thermally-induced melting behavior was monitored with changing pH. The complex formed by c_5Ugc_4 [5.8] exhibits higher T_m values than that formed by c_5Rgc_4 [5.5] at $pH > 3.5$, with the most pronounced differences being at pH 5.5-6.5 ($\Delta T_m = 5.5 - 6.4$ °C). It is remarkable that the complex forms under neutral conditions with a T_m of *ca.* 28 °C, suggesting that a more flexible loop might accommodate better the deoxycytidylic stems.

Comparison to the Branched dV-5 Tetrad

The stability of the linear and branched dV-5 complexes is of the same order of magnitude over the pH range 3.5 to 5.5, with the latter complex being more stable at pH 4.5 ($\Delta T_m = +3$ °C).²⁶³ At $pH > 6.0$, the melting temperature of the branched complex is less than 25 °C, culminating with a T_m value of *ca.* 16 °C at pH 6.8. In sharp contrast, all complexes formed by oligomers containing the (UUCG) loop exhibit higher thermal stabilities within the same pH range. The presence of the loop moiety might be responsible for complex formation at $pH > 5.5$ via shifting the equilibrium between

unprotonated and protonated cytosines.²⁶³ It is important to note that branched complexes with longer stems (*e.g.* dV-7) would increase thermal stability relative to dV-5 due to invoking a higher number of C·C+ base pairs.²⁶³

The Loop Motif as a Means to Stabilize the i-Motif

It is apparent from the above studies that the presence of a loop moiety within the oligomer sequence enhances the stability of the formed tetrad complex. This is manifested in the drastic increase in T_m values at various pH (Table 5.3). For instance at pH 4.5, both c_5Rgc_4 [5.5] and c_5Ugc_4 [5.8] exhibit significantly higher thermal stabilities relative to dC_5 (ΔT_m 's of 8 °C and 11 °C respectively). At pH 5.5, the oligomers c_5Rgc_4 [5.5] and c_5Rgc_4 [5.6] stabilize the complex by *ca.* 10 °C relative to dC_5 . Even more, stable complex formation is detected at neutral pH as evident from the oligomers c_5Rgc_4 [5.5], c_5Rgc_4 [5.6], c_4CRGc_4 [5.7] and c_5Ugc_4 [5.8].

One interpretation concerning complex formation at neutral pH has to deal with preorganization as well as molecularity of the reaction. Entropy considerations predict that a tetrad is more easily formed from two hairpin strands than from 4 single strands. In other words, the initial state of the bimolecular association is more organized, and it takes less entropic energy to attain the final tetrad state compared to the tetramolecular association. This was evident from the reversible thermally-induced melting curves (discussed below) which show faster kinetics of association for the hairpin oligomers relative to the single strands.

Kinetics of Association

The kinetics of complex formation was studied by heating the oligomer samples at 0.5 °C/min (dissociation) followed directly by cooling at the same rate (association). At pH 5.0, linear dC_5 showed hysteresis (*i.e.*, it was not able to reassociate instantly upon cooling) indicative of slow kinetics of formation for this complex. After 15 min equilibration time at 5 °C, the UV melting curve yields a very weak biphasic transition, again confirming that association of dC_5 into the i-motif is a very slow process (Figure 5.6).^{268,269} At the same pH, the UV melting profiles for oligomer c_5Rgc_4 [5.5] show hysteresis [the T_m 's of dissociation and reassociation do not coincide] typical of i-tetrad

behavior.^{268,269} However, the kinetics of formation are much faster than that observed for dC₅ since a complex that melts at the same T_m value (43.5 °C) is detected after 15 min of wait time (**Figure 5.6**). The same behavior is observed for all other oligomers containing a (UUCG) loop. Even though the UV melting profiles of c₅Rc₅ [**5.4**], c₅Rgc₄ [**5.6**], and c₅Ugc₄ [**5.8**] show hysteresis between heating and cooling, they are still able to form the same complex at a very fast rate (data not shown). One interpretation consistent with this behavior is the reduced molecularity of the process brought about by the presence of the loop moiety, which in turn induces faster kinetics of formation relative to the linear oligomer dC₅.

The same reversible melting studies were conducted at pH 7.0. All oligomers {c₅Rc₅ [**5.4**], c₅Rgc₄ [**5.5**], c₅Rgc₄ [**5.6**], and c₅Ugc₄ [**5.8**]} showed hysteresis in their UV melting behavior, indicative of slow kinetics of association consistent with C-tetrad formation.^{263,268,269} After a 15 min wait step, no complex re-formed (**Figure 5.7**). This behavior shows that complex formation at pH 7.0 is by far slower than at pH 5.0. The linear dC₅ molecule was not studied in this case since it did not form any complex at pH 7.0 (**Figure 5.5**).

5.4.2 Hairpin Versus Tetraplex Formation at pH 5.0

In an attempt to resolve the nature of the observed monophasic thermal transitions, the melting temperatures were analyzed over a broad range of oligonucleotide concentration (~30-fold). To recall, the T_m is independent of oligonucleotide concentration in an *intramolecular* association process (hairpin formation).²¹⁻²³ On the contrary, a concentration-dependent melting behavior is observed in an *intermolecular* process. Accordingly, a plot of the natural logarithm of total strand concentration (ln[conc]) *versus* the reciprocal melting temperature ($1/T_m$) was constructed over a range of 30-fold in oligomer concentration. **Figure 5.8** reveals concentration dependence of the UV melting temperature, as judged from the negative slope of the best fit line consistent with other studies involving intermolecular processes.²⁷⁰ The most drastic change in T_m value with concentration is seen for c₅Rgc₄ [**5.5**] in which T_m changes from 43.6 °C to 51.4 °C when concentration increases from 4.2 μM to 133 μM.

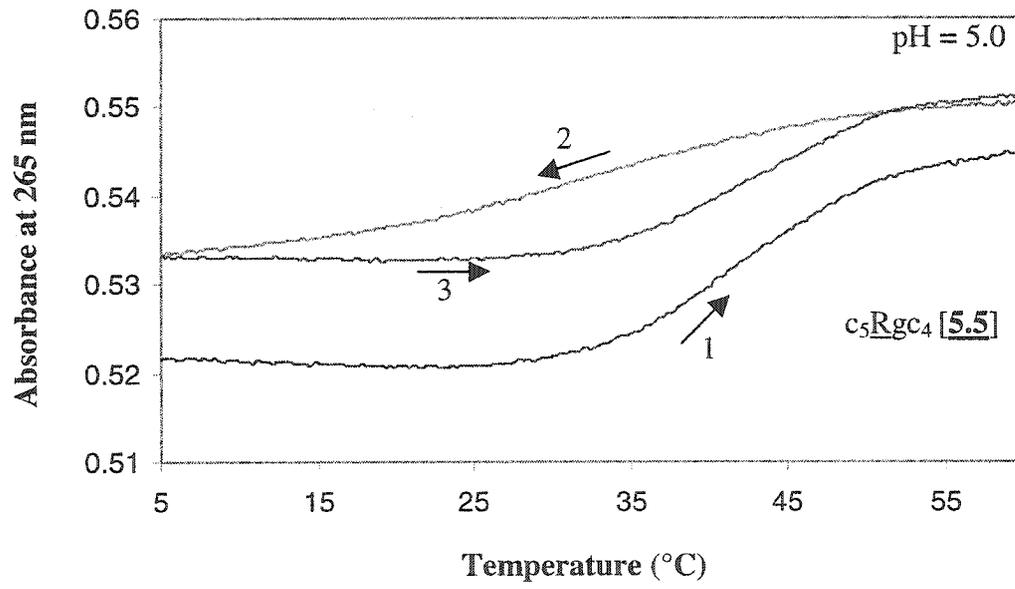
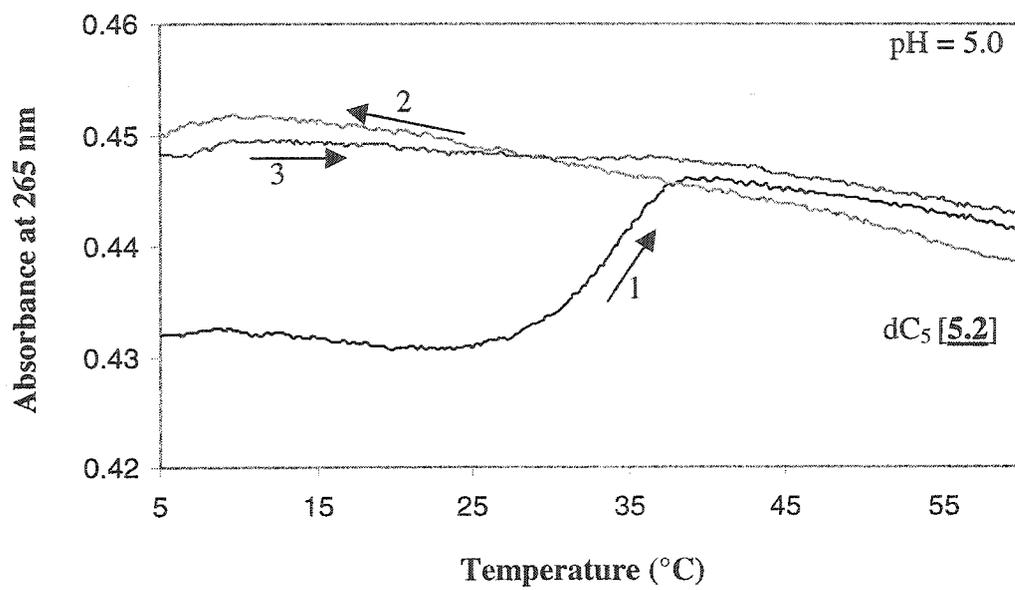


Figure 5.6: Reversible UV melting curves comparing the behavior of control dC₅ and c₅Rgc₄ [5.5] at pH 5.0. Measurements were done in 10 mM acetate buffer, pH 5.0 at an oligonucleotide concentration of 4.6 μM for dC₅, and 2.3 μM for c₅Rgc₄. The solutions were heated to 85 °C at a rate of 0.5 C/min, then cooled to 5 °C at the same rate. They were equilibrated for 15 min at 5 °C before the final heating step to 85 °C was started again.

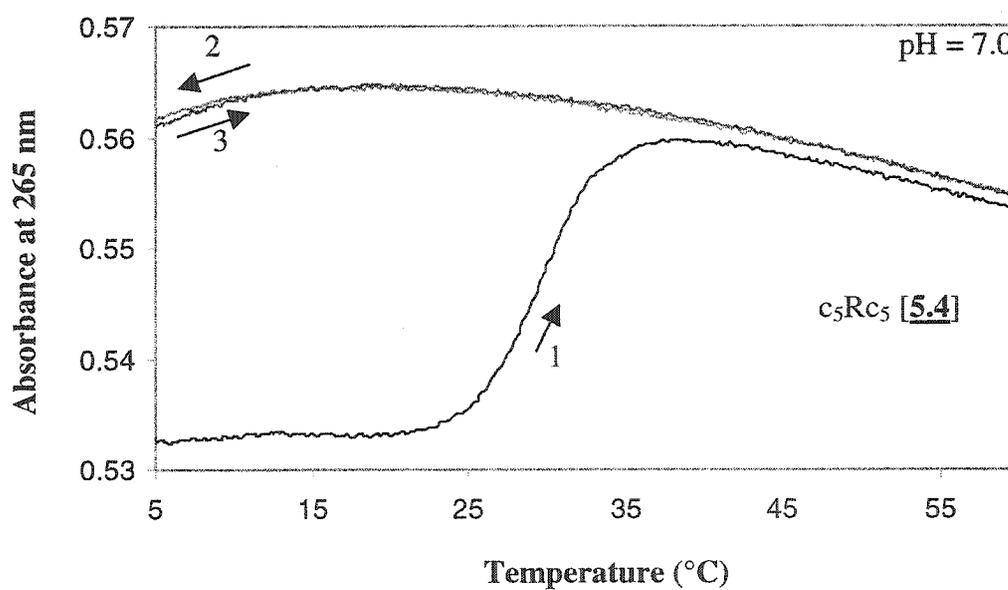
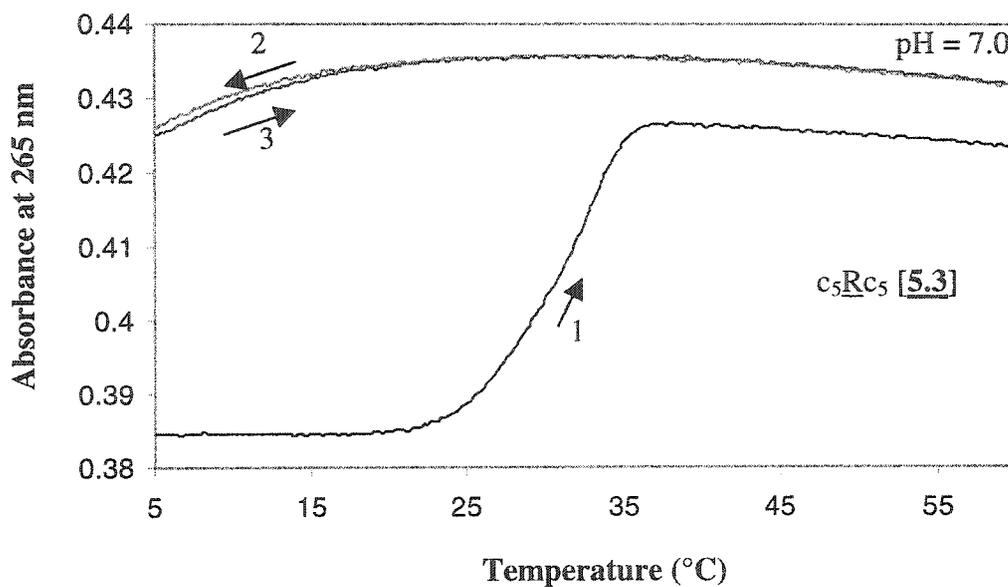


Figure 5.7: Reversible UV melting curves comparing behavior of representative deoxycytidine rich oligomers at pH 7.0. Measurements were done in 10 mM cacodylate buffer, pH 7.0 at an oligonucleotide concentration of $2.3 \mu\text{M}$. The oligomers c_5Rc_5 [5.3] and solutions were heated to 85°C at a rate of $0.5^\circ\text{C}/\text{min}$, then cooled to 5°C at the same rate. They were equilibrated for 15 min at 5°C before the final heating step to 85°C was started again.

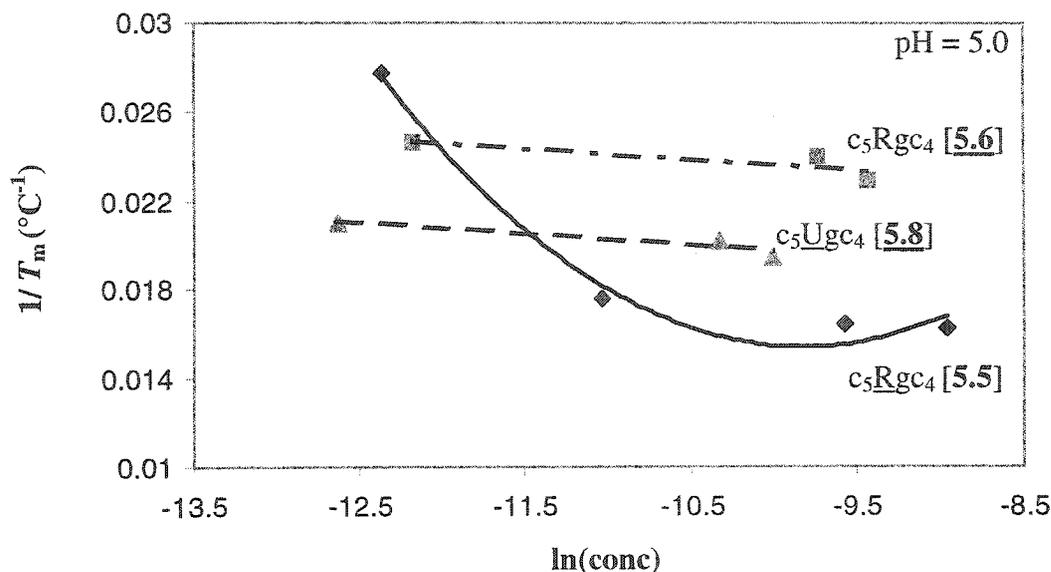


Figure 5.8: Van't Hoff plots showing T_m concentration-dependence of various deoxycytidine-rich oligomers with (UUCG) loops. All measurements were done in 10 mM sodium acetate buffer at pH 5.0.

In a similar fashion but to a lesser extent, the T_m increases by 3 °C for the complex formed by c_5Rgc_4 [5.6] when concentration is varied between 5.1 and 81 μM . The oligomer c_5Ugc_4 [5.8] displays a ΔT_m of 4 °C over a 15-fold change in concentration. The oligomer c_5Rgc_4 surprisingly displays a non-linear behavior that we can not explain at the moment.

These data clearly confirm that the observed complexes formed by deoxycytidine-rich hairpins associate via an *intermolecular* pathway.

5.4.3 Circular Dichroism Studies

The CD spectrum characteristic of an i-motif displays a strong positive band centered at *ca.* 288 nm that decreases in magnitude and becomes blue shifted as the complex is destabilized.^{271,272} Additionally, there is a negative band centered at *ca.* 267 nm that disappears upon thermal denaturation (for example, see **Figure 5.9A**), indicative of dissociation of protonated C·C⁺ base pairs in i-motif DNA.²⁷¹ At 5 °C, all deoxycytidine oligomers exhibit a strong positive CD band at *ca.* 287 nm, and a strong negative CD peak at *ca.* 267 nm (**Figure 5.9**). Heating to 65 °C causes the positive CD peak to decrease in intensity and become blue-shifted, while the 267 nm band disappears,

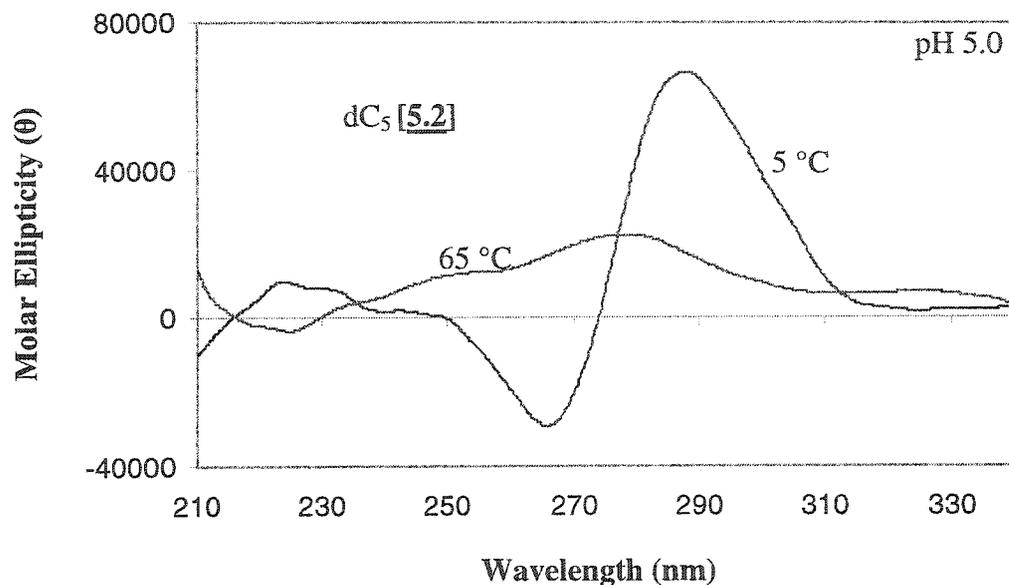
in agreement with the behavior typical of i-motifs.²⁷¹ Additionally, the CD spectra of these oligomers are influenced by pH of the solution. Their positive CD bands at 287 nm decrease in amplitude and become blue-shifted with a gradual loss of the 265 nm band as deviation from the optimum pH occurs, a trend that is similar to that observed in melting temperature dependence on pH. The complexes formed by these oligomers are most favored at pH 4.5-5.0, as judged from the highest amplitude displayed by the 287 nm CD peak (**Figure 5.10**). These results strongly support the existence of a bimolecular i-motif formed from two hairpin species with hemiprotonated C·C+ base pairs.

5.4.4 Gel Mobility Shift Assay

Gel electrophoresis differentiates between oligonucleotides according to their mass/charge ratio.²⁹ Accordingly, the migration of all deoxycytidine-containing oligomers was monitored relative to three control markers. Two are linear oligonucleotides, 30-nucleotides in length, of the same base sequence but different composition (RNA and DNA). The third marker is an RNA hairpin molecule 16 nucleotides in length and containing a (UUCG) tetraloop.

At pH 4.2, the oligomers c_5Rc_5 [**5.4**], c_5Rgc_4 [**5.6**], and c_4CRGc_4 [**5.7**] migrated with almost the same electrophoretic mobility as the 30-nucleotide DNA marker, but much slower than the 16-nucleotide RNA hairpin marker (**Figure 5.11**). This is in agreement with a 'size' of 28 nucleotides, and consistent with the formation of a dimeric tetrad species (*e.g.*, 2 x [c_5Rc_5]). In sharp contrast, the ribocytidine-rich oligomer C_5RC_5 [**5.10**], discussed below in **Section 5.5**, migrated with almost the same rate as the 16-nucleotide hairpin marker, but much faster than the 30-nucleotide linear DNA oligonucleotide. This is consistent with the formation of a hairpin duplex of the 'size' 14 nucleotides (1 x [C_5RC_5]). The oligomers c_5Rc_5 [**5.3**] and c_5Rgc_4 [**5.5**] also showed complexes that migrate similar to the 30-nucleotide DNA marker, indicative of dimeric tetrad formation. However at the gel concentrations, they also showed other faster migrating bands, indicative of formation of other multimeric species.

A.



B.

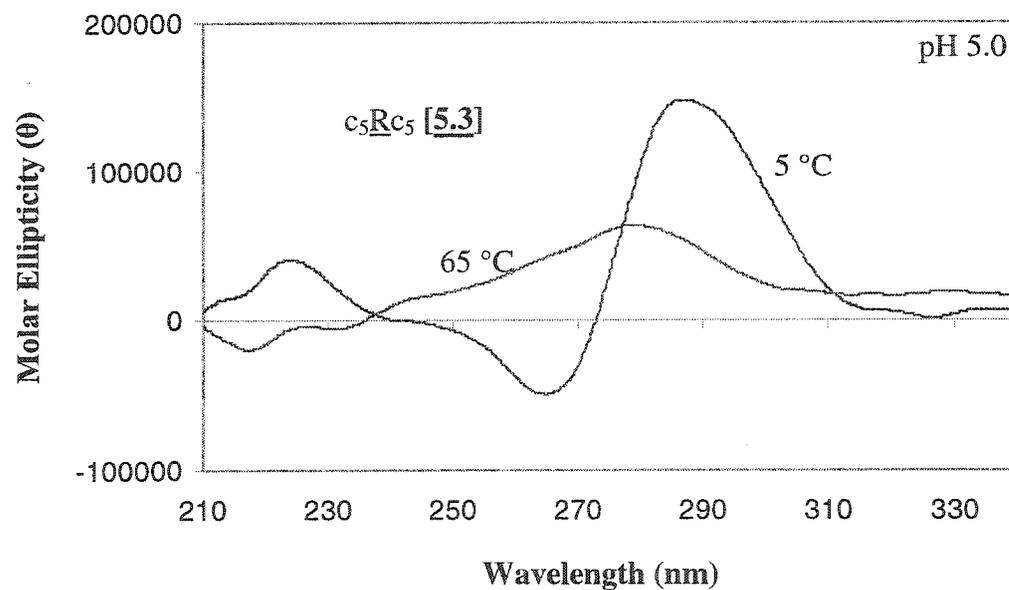
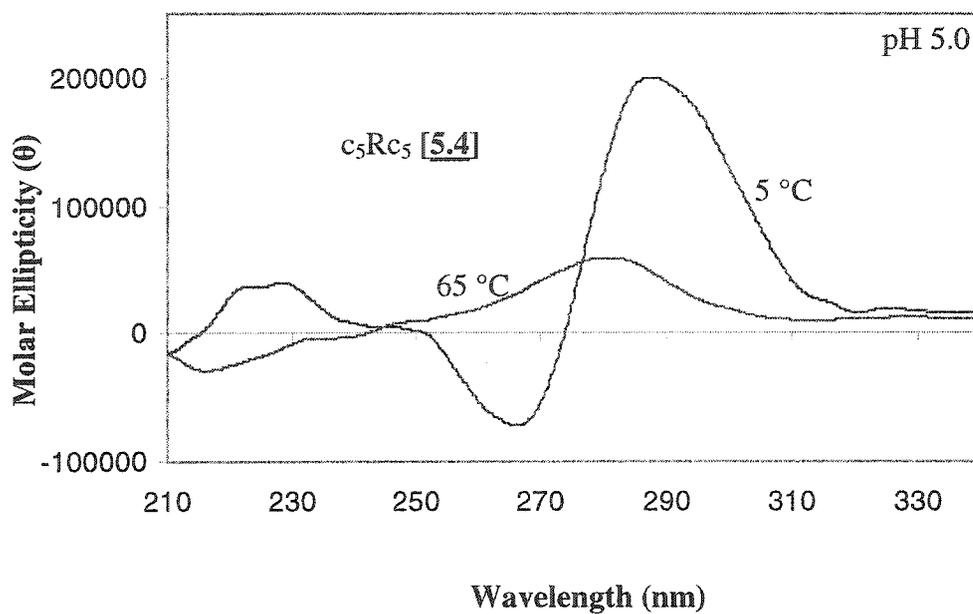


Figure 5.9 A and B: Circular dichroism spectra of dC₅ and c₅Rc₅ at 5 °C and 65 °C. All measurements were done in 10 mM sodium acetate buffer, pH 5.0. Oligomer concentrations were 2.3 μM for c₅Rc₅, and 4.6 μM for dC₅. Molar ellipticities were normalized to strand concentration.

C.



D.

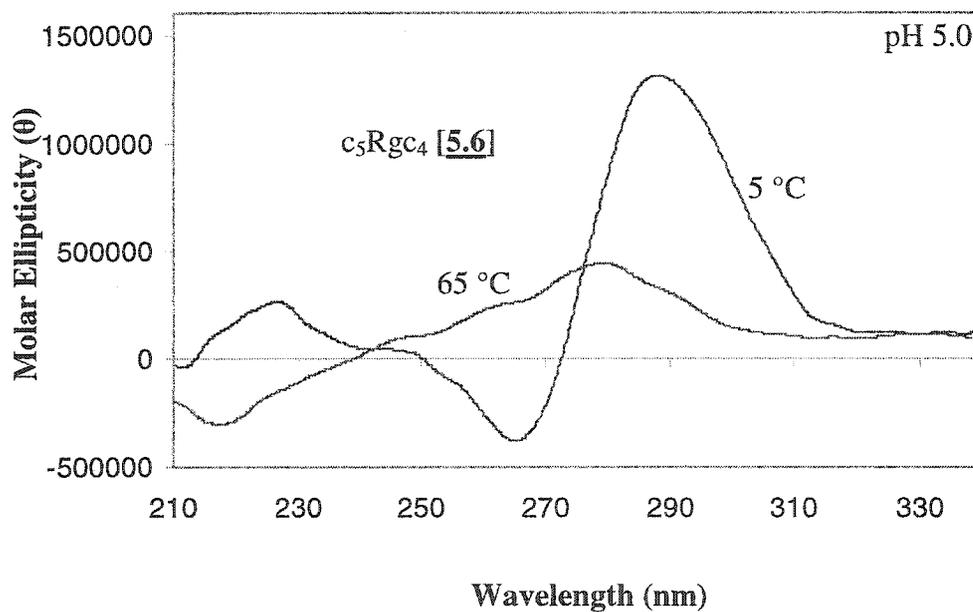
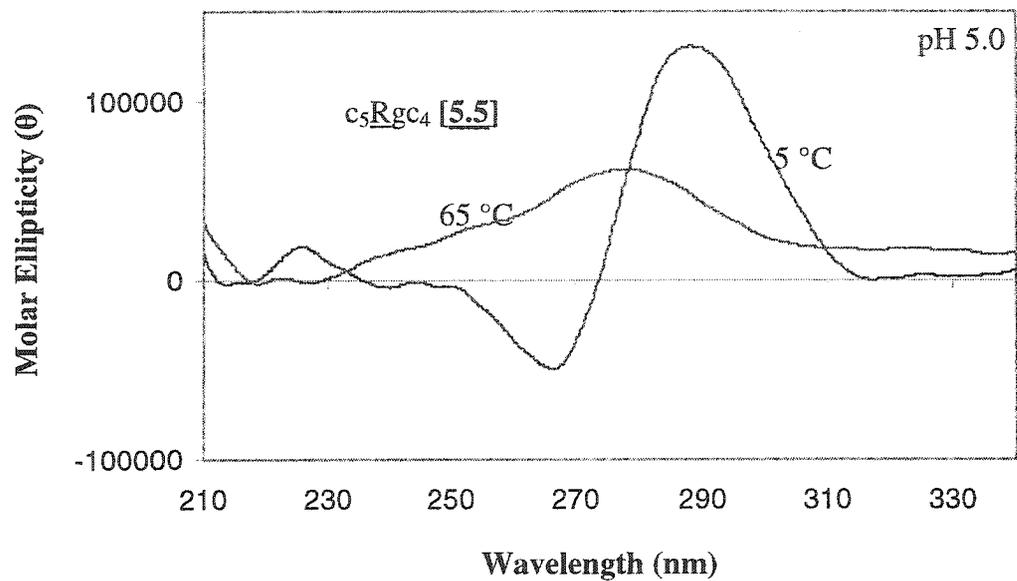


Figure 5.9 C and D: Circular dichroism spectra of c_5Rc_5 and c_5Rgc_4 at 5 °C and 65 °C. All measurements were done in 10 mM sodium acetate buffer, pH 5.0. Oligomer concentrations were 2.3 μ M. Molar ellipticities were normalized to strand concentration.

E.



F.

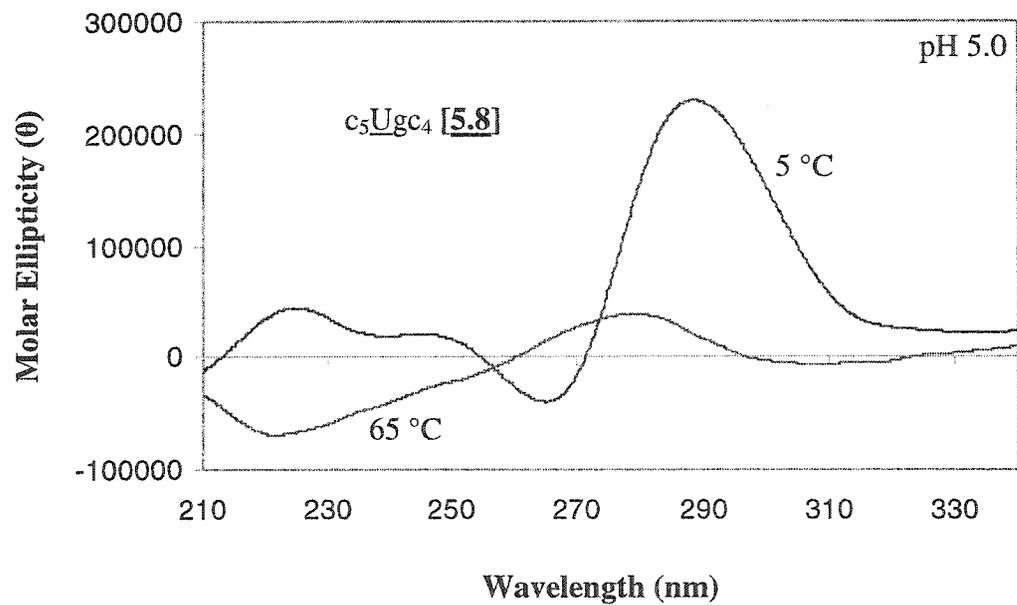
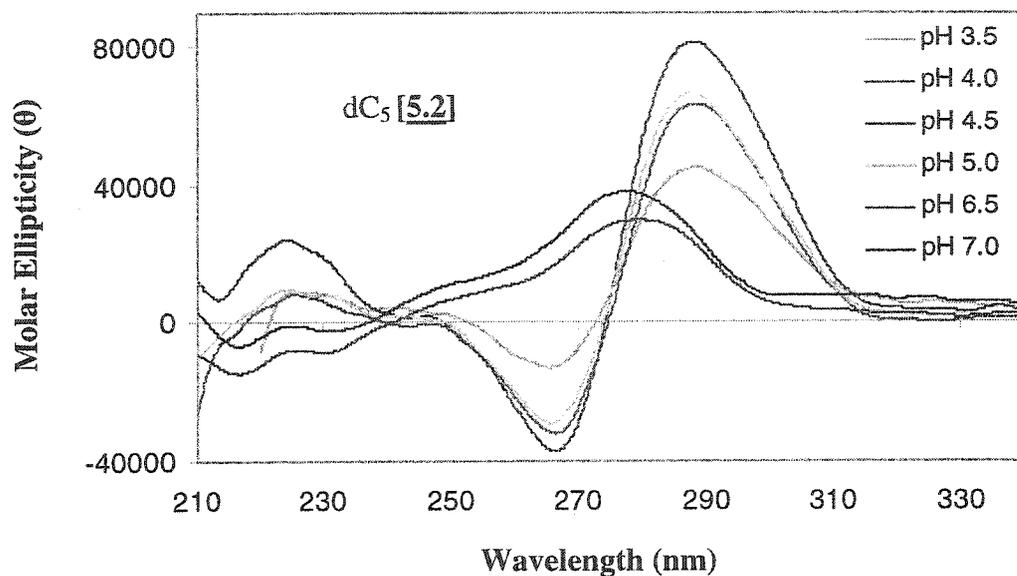


Figure 5.9 E and F: Circular dichroism spectra of c_5Rgc_4 and c_5Ugc_4 at 5 °C and 65 °C. All measurements were done in 10 mM sodium acetate buffer, pH 5.0. Oligomer concentrations were 2.3 μ M. Molar ellipticities were normalized to strand concentration.

A.



B.

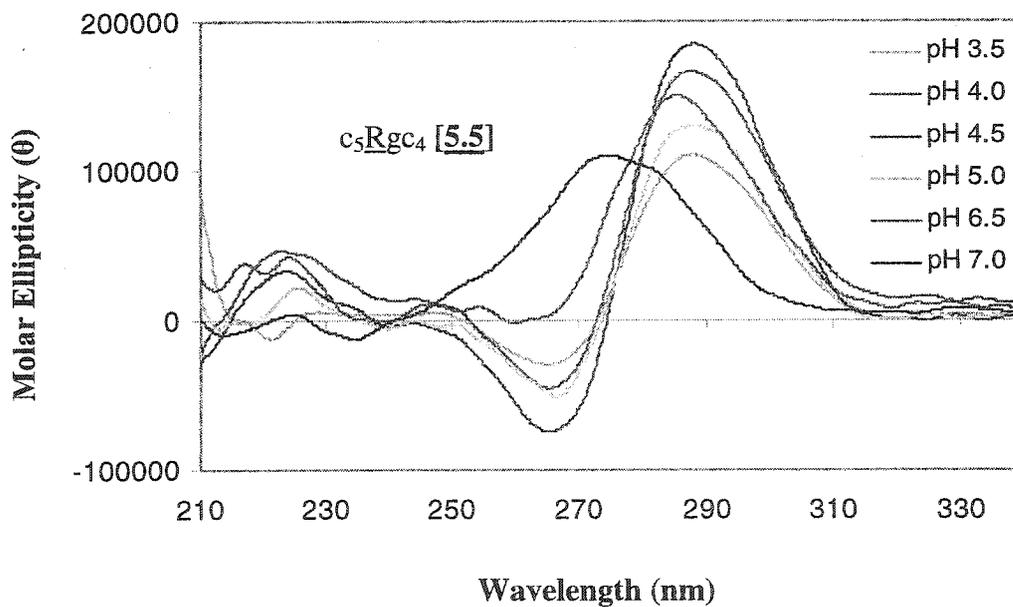


Figure 5.10 A: Circular dichroism spectra of dC₅ and c₅Rgc₄ at various pH. All measurements were done at 5 °C in either 10 mM sodium acetate buffer (pH 3.5 - 5.5) or 10 mM sodium cacodylate buffer (pH 5.5 - 7.0). Oligomer concentrations were 2.3 μM for c₅Rgc₄, and 4.6 μM for dC₅. Molar ellipticities were normalized to strand concentration at all pH values.

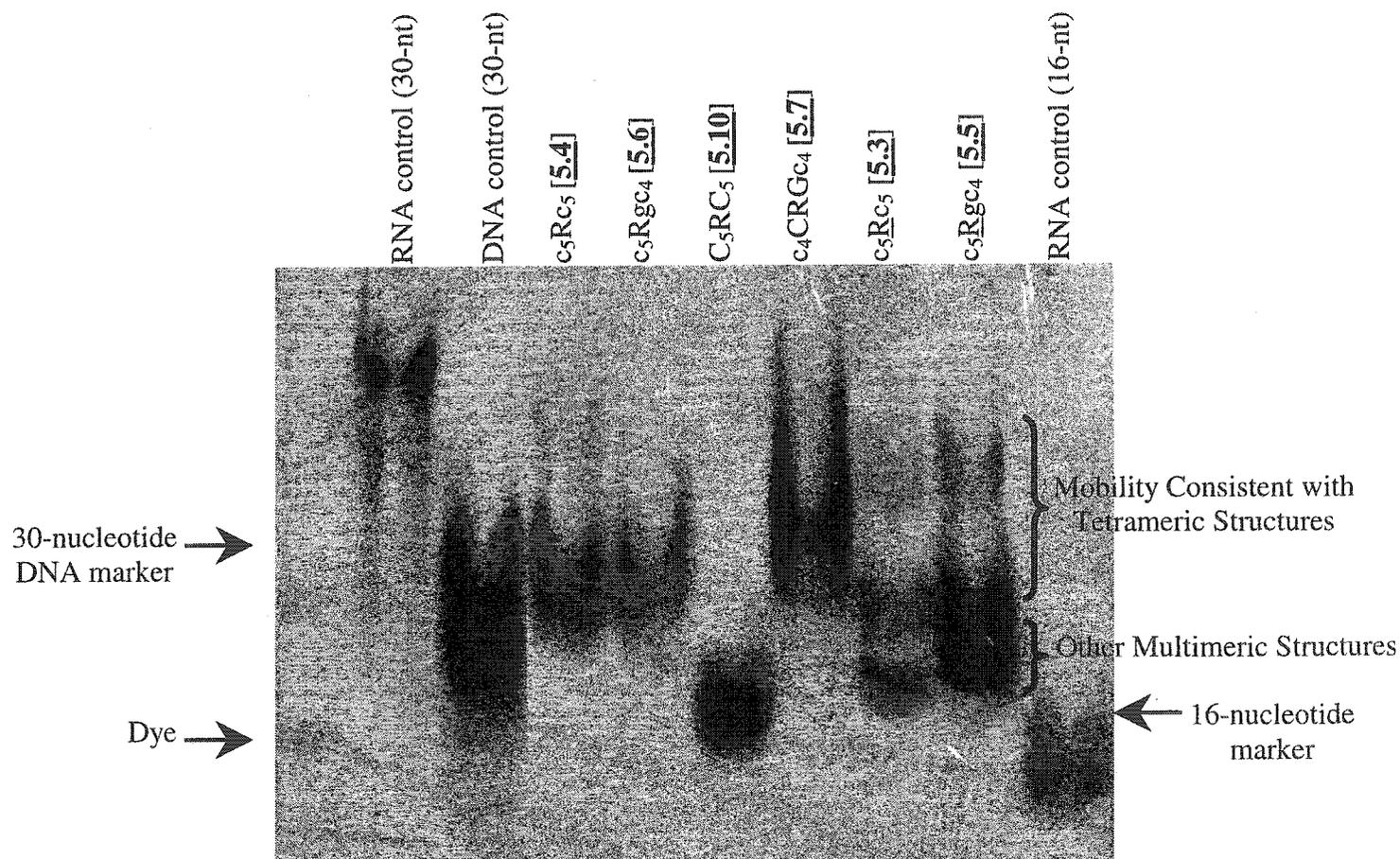


Figure 5.11: Gel mobility shift assay (native conditions) of various cytosine-rich oligomers. Samples (~0.35 ODU) were dissolved in 10 μ l loading buffer [30% sucrose in 10 x TAE buffer, pH = 4.2], heated to 95 $^{\circ}$ C and then cooled down slowly to room temperature and finally incubated at 4 $^{\circ}$ C for a week prior to analysis. The samples were loaded on a 16 % acrylamide gel (no urea added) and the gels were run at low voltage (100 – 150 V) for 5-6 h in the fridge (4 $^{\circ}$ C) at pH 4.2. The sequence of the DNA control is: 5'- ggc gtc atg tca tgt catg agtc catg gca tgg cat ggc gcc-3', while that of the RNA control is: 5'-GGC GUC AUG AGUC CAU GGC GCC-3'.

5.4.5 Proposed Structure for the Dimeric i-Motif Complex

The above experiments evidently point towards the formation of a dimeric i-motif complex. In principle, two hairpins can stack anti parallel to one another (head-to-head or head-to-tail) to yield an intercalated hemiprotonated C·C⁺ structure (Figure 5.12). In the head-to-tail configuration, the two (UUCG) loops are situated on opposite ends of the tetrad structure. This arrangement is similar to what has been previously observed in the structure of d(5mCCTTTTAC₂) with a [TTTA] loop where TA pairing motifs were detected.²⁶² In a different way, the deoxycytidine hairpins can display a head-to-head arrangement where the two loops are in close proximity to one another. In each of these proposed models, C5 can possibly base pair with G10 and U6 can pair with G9, thus contributing to stabilization of the structure (Figure 5.12). The base pairs that form at the loops can be parallel to the adjacent C·C⁺ base pair, a trend that has been observed in dimeric and monomeric i-motif structures. Furthermore, in the head-to-head arrangement, the formation of an intermolecular U·G·G·U motif is also possible. In fact, T·G·G·T motifs have been observed in the crystal structure of the oligomer d(TC₃GT₃C₂A) that intercalates in a head-to-head fashion.²⁵³

In each of the proposed arrangements, two different topologies can exist. A close inspection of the native gels (Figure 5.11) reveals that only one tetrameric form is dominant for the oligomers c₅Rc₅, c₅Rgc₄, and c₄CRGc₄. However, oligomers c₅Rc₅ and c₅Rgc₄ form multimeric species, the result of which may be ascribed to structures bearing different topologies or different arrangements. In Figure 5.12, we represent only the 3'E topology for each configuration, while the 5'E can also exist. Since it is not possible to determine from the present data which configuration is dominantly favored by these hairpin structures, NMR studies will have to be pursued in the future to solve the structure of this highly stable i-motif complex.

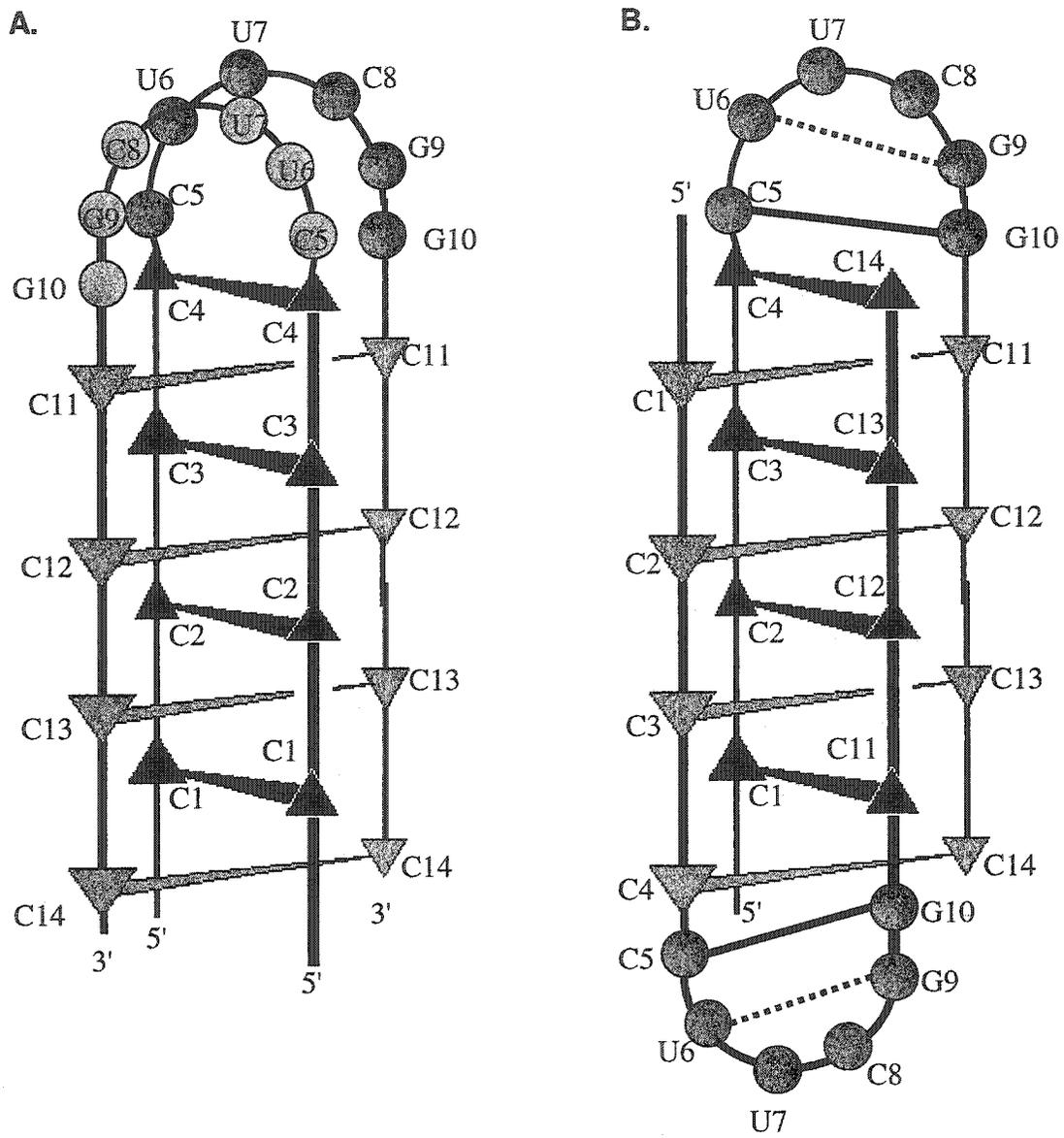


Figure 5.12: Proposed tetrad structures for i-motif formation by hairpin oligomers.

5.5 HAIRPINS WITH RIBOCYTIDINE STEMS

5.5.1 Background

In 1963, Langridge and Rich suggested, through X-ray diffraction studies, that polyribocytidine [poly (rC)] associates into hemiprotonated parallel duplexes at pH 5.5.²⁴³ Further studies have detected a hypochromic structure of unknown nature for oligo(rC)²⁷³ and poly(rC)^{258,259,274} under acidic conditions. The idea that poly(rC) forms a duplex prevailed for thirteen years until Arnott and coworkers reported a huge discrepancy with regards to the poly(rC) structure.²⁴⁴ Contradictory to prior beliefs, their findings proved that these homopolymers instead form structured single strands. Additionally, poly (rC) was shown via NMR to exhibit a left-handed single stranded helix at neutral pH.²⁴⁶ The view that ribocytidine homopolymers adopted hemiprotonated structures was evident, but it was uncertain whether these structures were single strands, duplexes or tetrads. RNA was shown to be incapable of forming the i-motif.²⁷⁵ Replacing deoxycytidine with ribocytidine within pyrimidine oligonucleotides significantly destabilized the i-motif and led to the formation of triple helices instead. A recent study by Damha and coworkers attributed this inferior stability to the presence of the 2'-hydroxyl groups that sterically impeded the formation of stable RNA i-motifs.²⁶⁴ Introduction of an arabinose moiety (the 2'-epimer of ribose), whose 2'-hydroxyl group is supposed to point into the wide groove of the i-motif, did not impair thermal stability. Indeed, the authors were able to detect C-tetrad formation in a branched oligomer composed of fully modified arabinose residues, and the properties of this complex were very similar to that formed by the deoxycytidine counterpart (i-motif). Further evidence came from structural studies conducted on linear d(TC₅) oligomers.²⁶⁹ Single or double ribose insertions within this sequence had strong destabilizing effects. The most dramatic effect was observed when two ribose residues were placed in close proximity to one another (adjacent sugar residues), indicative of a steric clash between the 2'-OH groups and in agreement with the earlier findings of Damha and coworkers. Furthermore, r(UC₅) was not able to form any tetramer at temperatures higher than 5 °C as judged from UV melting studies.²⁶⁹ However, its structured form, based on gel filtration experiments, was

reported to involve a two-strand stoichiometry (*i.e.*, duplex) instead of the four-strand stoichiometry expected for i-motif formation.²⁷⁵

The first detailed structure of an RNA i-motif has recently been reported by Leroy and coworkers.²⁷⁶ The solution structure was observed at $-3\text{ }^{\circ}\text{C}$ via NMR, indicative of its high thermal instability. Despite bearing much resemblance to i-DNA, there exist modest differences between the two motifs. The oligomer r(UC₅) associates into two i-motif tetrads that differ in their intercalation topologies. The core of the major species resembles that of i-DNA, except that the intercalation topology avoids one of the six 2'-OH/2'-OH repulsive contacts anticipated for maximal intercalation. The sugar residues are in the normal C-3' *endo* pucker form. This, along with the orientation of the hemiprotonated C·C+ base pairs, causes widening of the narrow groove in order to reduce the repulsive interactions between the hydroxyl groups. This is in sharp contrast to the d(TC₅) oligomer which displays only one intercalated form.²⁵¹

All of the above studies point in one direction that suggests the reluctance of RNA to form stable i-motif structures at moderate temperatures.

5.5.2 Experimental Design

The inability of oligoribonucleotides to form stable C·C+ complexes hindered the detailed structural characterization of the nature of these complexes. In this study, we were able to induce stable complex formation by linking poly ribocytidine to an 'extra-stable' C(UUCG)G loop (Chapter II). **Table 5.4** lists the cytosine-rich oligomers under study. C₁₀(R)GC₉ [**5.11**] represents an all-RNA oligomer, while C₁₀(R)GC₉ [**5.12**] is the same as **5.11** but with a 2',5'-RNA loop. The characterization of all complexes formed by ribocytidine, presented here for the first time, proceeds through UV spectroscopy, circular dichroism, and gel electrophoresis experiments. Branched oligomers^{263,264} containing vicinal 3',5'- or 2',5'-phosphodiester linkages at a branchpoint adenosine were also included in this study for comparison purposes.

Since a stable well-studied oligoribocytidine control sequence is unavailable, we used instead an oligodeoxycytidine analogue as a model for comparison to the oligomers under investigation. Specifically, decadeoxycytidylic acid [dC₁₀] and c₉Rgc₈ were employed as the controls (**5.14** and **5.15**; **Table 5.4**).

All oligomers, including the controls, were synthesized on solid-phase using phosphoramidite chemistry (see experimental **Section 7.2**). They were purified by denaturing gel electrophoresis (25% acrylamide), and the structures were confirmed by MALDI-TOF mass spectrometry. The purity of all oligomers was >94% as judged from analytical gel electrophoresis.

Table 5.4: Base sequence of cytidine-rich oligomers under study

Code	Designation	5'- Oligomer Sequence -3'
<u>5.10</u>	C ₅ RC ₅	5'- CCCCC(UUCG)CCCC -3'
<u>5.11</u>	C ₁₀ (R)GC ₉	5'- C ₉ [C(UUCG)G] C ₉ -3'
<u>5.12</u>	C ₁₀ (<u>R</u>)GC ₉	5'- C ₉ [C(UUCG)G] rC ₉ -3'
<u>5.13</u>	<u>C</u> ₁₀ RGC ₉	5'- C ₉ [<u>C</u> (UUCG) <u>G</u>] rC ₉ -3'
<u>5.14</u>	c ₉ Rgc ₈	5'- c ₉ (<u>UUCG</u>) g c ₈ -3'
<u>5.15</u>	dC ₁₀	5'- ccccccccc -3'
<u>5.16</u>	rV-5	$\begin{array}{l} \text{rA} \begin{array}{l} \nearrow 2'-5' \\ \searrow 3'-5' \end{array} \text{C}_5 \quad -3' \\ \text{C}_5 \quad -3' \end{array}$
<u>5.17</u>	rV-10	$\begin{array}{l} \text{rA} \begin{array}{l} \nearrow 2'-5' \\ \searrow 3'-5' \end{array} \text{C}_{10} \quad -3' \\ \text{C}_{10} \quad -3' \end{array}$
<u>5.18</u>	2'-5'rV-5	$\begin{array}{l} \text{rA} \begin{array}{l} \nearrow 2'-5' \\ \searrow 3'-5' \end{array} \underline{\text{C}}_5 \quad -3' \\ \underline{\text{C}}_5 \quad -3' \end{array}$
<u>5.19</u>	dV-10	$\begin{array}{l} \text{rA} \begin{array}{l} \nearrow 2'-5' \\ \searrow 3'-5' \end{array} \text{c}_{10} \quad -3' \\ \text{c}_{10} \quad -3' \end{array}$

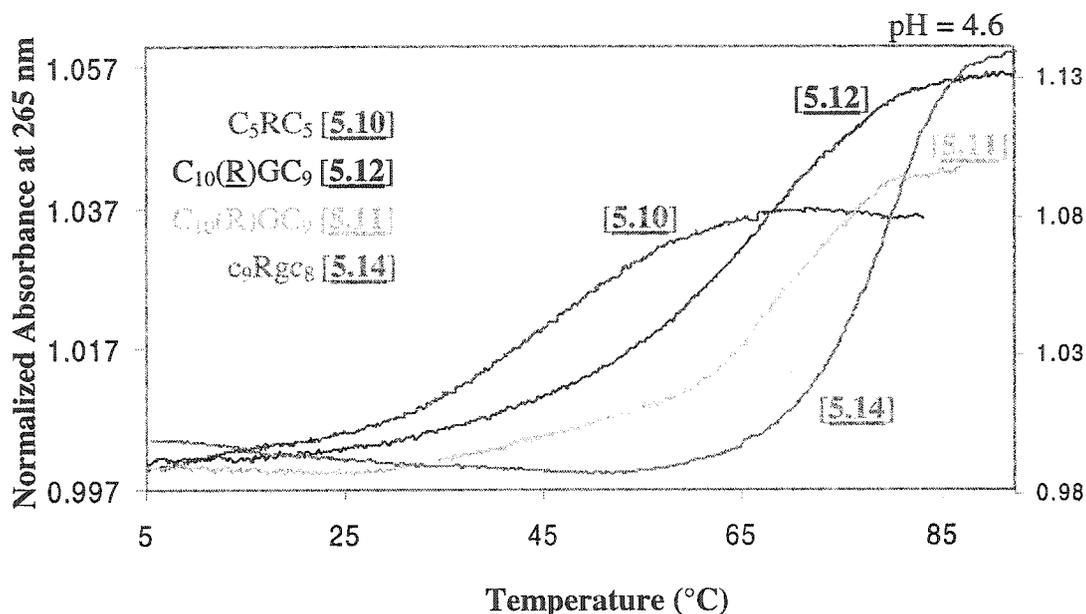
Underlined letters represent 2',5'-linked RNA residues (e.g., UC = U_{2p5}C_{2p}). Capital letters represent 3',5'-RNA residues, while small letters represent DNA residues. The reference branched molecules (5.16 - 5.19) are taken from [Robidoux *et al. J. Biomol. Struct. Dyn.* **1997**, *15*, 517-527; Robidoux & Damha *J. Biomol. Struct. Dyn.* **1997**, *15*, 529-535], and are shown here for continuity and comparison purposes of the work.

5.5.3 UV Thermal Studies

As a preliminary step to check possible complex formation by the ribocytidine-rich oligomers containing the (UUCG) loop, the thermally induced melting profile of C_5RC_5 [5.10] was monitored in 50 mM sodium citrate at pH 4.6. Surprisingly, a monophasic behavior (T_m ca. 43.5 °C) with a weak hypochromic transition was observed (Figure 5.13). Increasing the stem length [$C_{10}(R)GC_9$ [5.11] & $C_{10}(R)GC_9$ [5.12]] resulted in further stabilization of the observed complex. $C_{10}(R)GC_9$ and $C_{10}(R)GC_9$ exhibited remarkably high T_m values of 69.0 °C and 65.1 °C, respectively. These oligomers exhibited reduced hypochromicity compared to the control c_9Rgc_8 complex ($\Delta\%H = 8-9\%$). In sharp contrast, the branched compounds rV-5 and rV-10 did not form any complex under the same conditions.²⁶⁴ To test whether complex formation involved C·C+ base pairing, we studied the thermal stability of the above oligomers at different pH.

T_m as a Function of pH

The thermal stabilities of the complexes formed by C_5RC_5 [5.10], $C_{10}(R)GC_9$ [5.11] and $C_{10}(R)GC_9$ [5.12] are pH-dependent (Figure 5.14). The maximum melting temperature occurs at a pH value of ~ 4.5 similar to the pK_a value of N3-protonated cytosine. $C_{10}(R)GC_9$ exhibits the highest thermal stability ($T_m = 71.9$ °C), followed by $C_{10}(R)GC_9$ ($T_m = 66.4$ °C) and C_5RC_5 ($T_m = 41.3$ °C). Table 5.5 compares the calculated T_m and %H values for the different complexes observed at various pH. In the 5.0 to 7.0 range, the T_m dropped as a function of pH for all oligomers including the deoxycytidine controls. In a similar fashion, the drop was observed at pH < 4.5. This behavior has been recurrent in both polyribo- and polydeoxyribonucleotides containing cytosine stretches,^{274,277} and thus confirms protonation of the cytosines within the oligomers under investigation. It is important to note that the drop in T_m for the ribocytidine oligomers elevates at pH > 6.0, a behavior in sharp contrast to that of the deoxycytidine counterparts (Figure 5.14).



Code	Designation	T_m (°C)	% H
<u>5.10</u>	C ₅ RC ₅	43.5	3.5
<u>5.11</u>	C ₁₀ (R)GC ₉	69.0	4.2
<u>5.12</u>	C ₁₀ (R)GC ₉	65.1	5.2
<u>5.14</u>	c ₉ Rgc ₈	76.5	13.2
<u>2.31</u>	TRT	-	-

Figure 5.13: UV thermal melting data for ribocytidine-rich oligomers. Measurements were recorded at 265 nm in 50 mM sodium citrate, pH = 4.60 ± 0.02. The right-hand vertical axis represents the absorbance for control c₉Rgc₈ [5.14], while the left-hand vertical axis represents the absorbance for all other oligomers under study. All oligomers were 2.3 μM in total concentration. All samples exhibited monophasic and cooperative melting curves. The values represent the average of at least two independent measurements. The melting temperature (T_m) was calculated according to the base line method. The error in melting temperature is ± 0.5 °C.

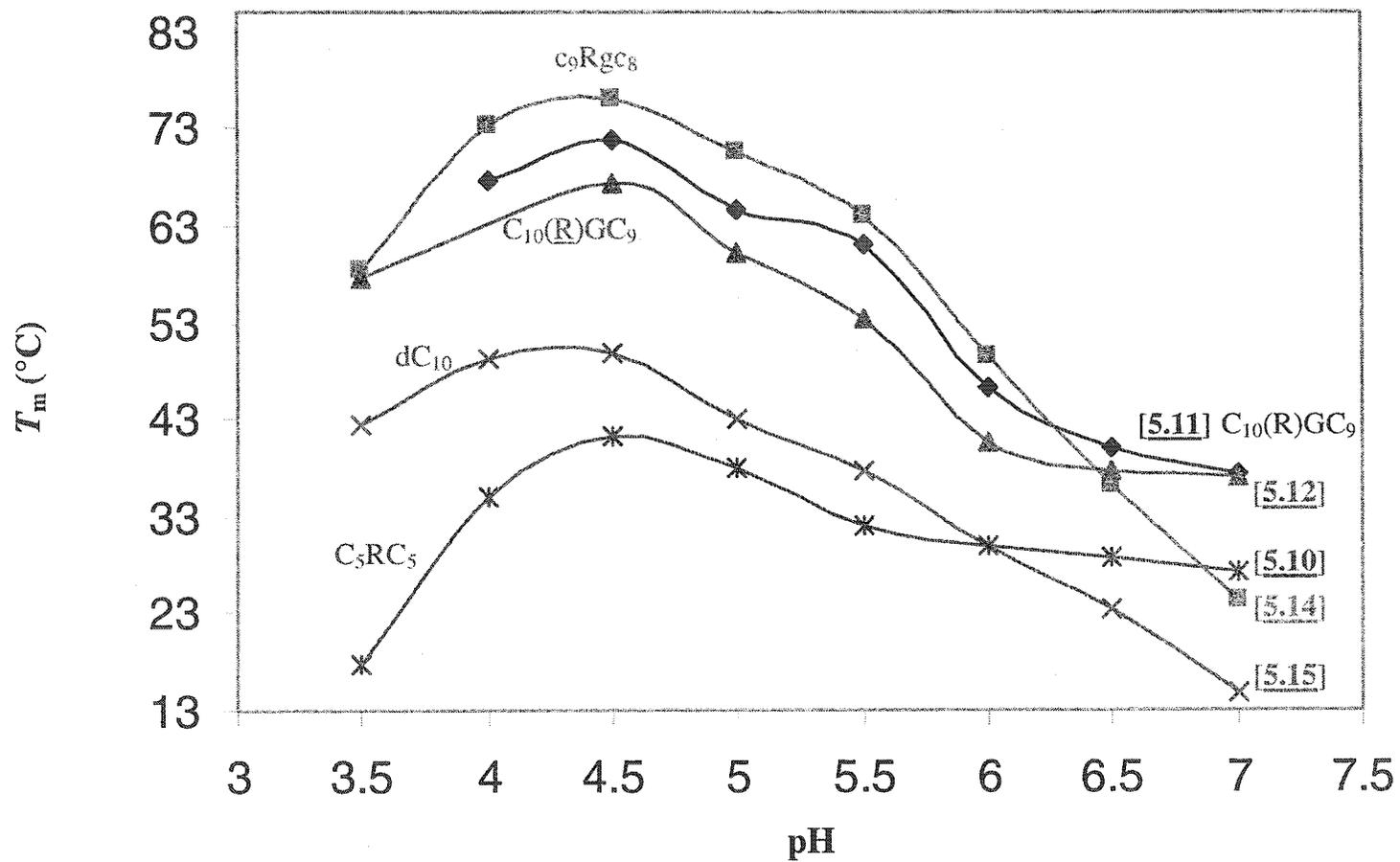


Figure 5.14: Plots of T_m versus pH for ribocytidine oligomers. The control deoxycytidine oligomers are also shown. dC_{10} was $4.6 \mu\text{M}$ in concentration, while all other oligomers were $2.3 \mu\text{M}$. All measurements were recorded at 265 nm in either 10 mM sodium acetate (pH 3.5-5.5) or 10 mM sodium cacodylate buffer (pH 5.5-7.0).

Table 5.5: UV melting data of ribocytidine rich oligomers under study

pH	dC ₁₀ [5.15]		c ₉ Rgc ₈ [5.14]		C ₁₀ (R)GC ₉ [5.11]		C ₁₀ (R)GC ₉ [5.12]		C ₅ RC ₅ [5.10]	
	T _m (°C)	%H	T _m (°C)	%H	T _m (°C)	%H	T _m (°C)	%H	T _m (°C)	%H
3.5	42.5	8.4	58.5	9.7	n.c.	-	57.7	2.0	17.8	2.4
4.0	49.3	9.4	73.5	9.9	67.7	7.3	n.c.	n.c.	34.9	2.3
4.5	50.0	8.8	76.1	10.1	71.9	10.4	67.5	6.8	41.3	2.4
5.0	43.0	11.4	70.7	11.4	64.8	8.3	60.4	7.4	38.0	3.9
5.5	37.8	10.7	64.2	9.6	61.3	8.1	53.5	4.9	32.2	3.4
6.0	30.1	11.1	49.5	9.8	46.4	7.5	40.6	4.9	30.0	2.8
6.5	23.4	8.5	36.2	9.2	40.1	4.1	37.6	3.3	28.9	2.0
7.0	14.9	2.8	24.3	9.3	37.3	3.9	37.0	4.6	27.3	3.2

n.c. = not calculated due to a broad melting transition. The control sequences dC₁₀ and T7 are shown for comparison purposes. All measurements were recorded at 265 nm in either 10 mM sodium acetate (pH 3.5-5.5) or 10 mM sodium cacodylate buffer (pH 5.5-7.0). Oligomer concentrations were 2.3 μM except for dC₁₀ (4.6 μM). The UV melting temperatures were computed according to the base line method. All values represent the average of two independent measurements. The error in melting temperature is ± 0.5 °C. Percentage hypochromicity (%H) was calculated from UV absorbances at low (A₀) and high (A_f) temperatures according to the following equation: %H = (A_f - A₀) / A_f.

Kinetics of Association

At pH 5.0, all ribocytidine-containing oligomers {C₅RC₅ [5.10], C₁₀(R)GC₉ [5.11], & C₁₀(R)GC₉ [5.12]} did not show any hysteresis in their UV melting profiles. Furthermore, they were able to re-form a complex within 15 min wait time (Figure 5.15). These results confirm very fast kinetics of association. Even more intriguing are the control oligomers c₉Rgc₈ [5.14] and dC₁₀ [5.15] containing deoxycytidine residues and exhibiting similar behavior in their UV melting profiles, *i.e.*, no hysteresis observed and very fast kinetics for complex re-association. This prompted us to examine these oligomers at a different pH. Indeed, at pH 7.0, the control oligomers c₉Rgc₈ and dC₁₀ showed non-superimposable UV melting curves (hysteresis), consistent with C-tetrad formation.^{263,268,269} In contrast, the ribocytidine oligomers persistently showed no hysteresis at pH 7.0, inconsistent with C-tetrad formation (Figure 5.15). Even more, they were still able to reform with the same melting temperature at this pH. These results confirm very fast kinetics of association for the ribocytidine-containing oligomers C₅RC₅, C₁₀(R)GC₉, and C₁₀(R)GC₉, and demonstrate that the complex formed by these molecules is not likely an i-motif.

5.5.4 Hairpin Duplex Versus Tetraplex Formation at pH 5.0

Decadeoxycytidylic acid [dC₁₀] and c₉Rgc₈ [5.14] demonstrated the most dramatic change in melting temperature across a range of varying oligomer concentration in solution (Figure 5.15). The T_m value of dC₁₀ changes by *ca.* 15 °C over a 38-fold increase in oligomer concentration, whereas that of c₉Rgc₈ changes by *ca.* 7 °C over a 30-fold concentration range. All hairpins with ribocytidine stems show concentration-independent melting behavior, indicative of a *unimolecular* process of complex formation. The T_m value of C₁₀(R)GC₉ [5.12] remains constant over a 30-fold increase in concentration. A similar trend is recurrent in C₁₀(R)GC₉ [5.11] and even in the smaller ribocytidine stems C₅RC₅ [5.10], confirming that the association process for the complex formed by these ribocytidine oligomers involves a unimolecular transition, *i.e.*, a *hairpin* species at pH 5.0 (Figure 5.16). This suggests an intramolecular folding of the single strand to form an antiparallel C·C+ *duplex* stem.

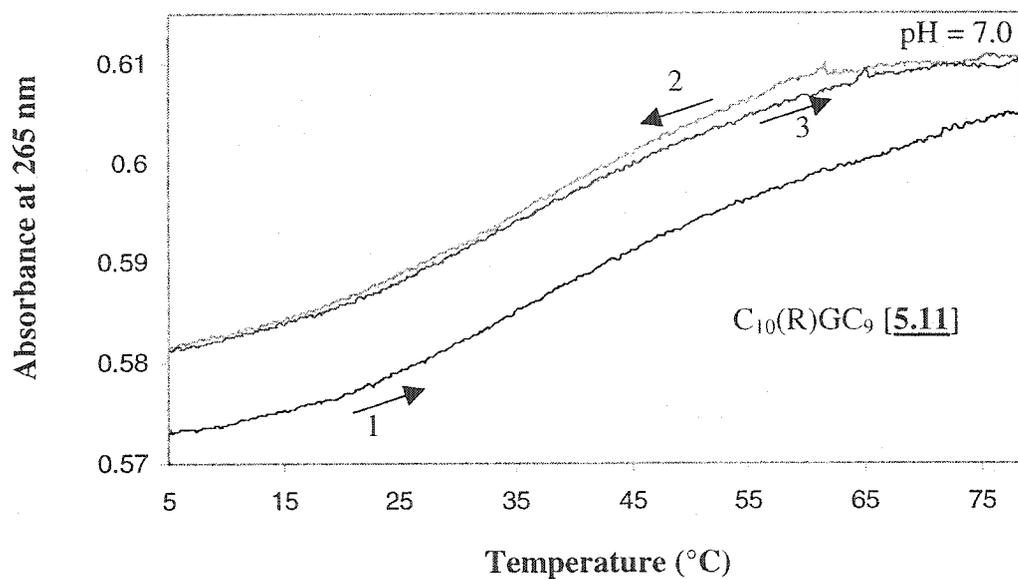
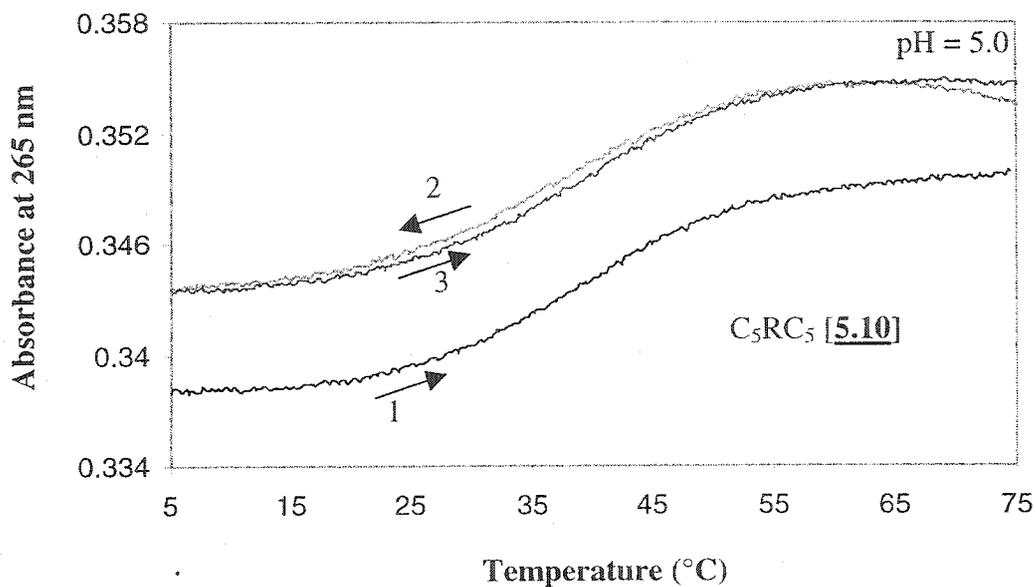


Figure 5.15: Reversible UV melting behavior of C_5RC_5 [5.10] and $C_{10}(R)GC_9$ [5.11] at pH 5.0 and 7.0 respectively. Measurements were done in either 10 mM acetate buffer, pH 5.0 or 10 mM sodium cacodylate, pH 7.0. Oligonucleotide concentration was $2.3 \mu\text{M}$. The solutions were heated to 85°C at a rate of $0.5^\circ\text{C}/\text{min}$, then cooled to 5°C at the same rate. They were equilibrated for 15 min at 5°C before the final heating step to 85°C was started again.

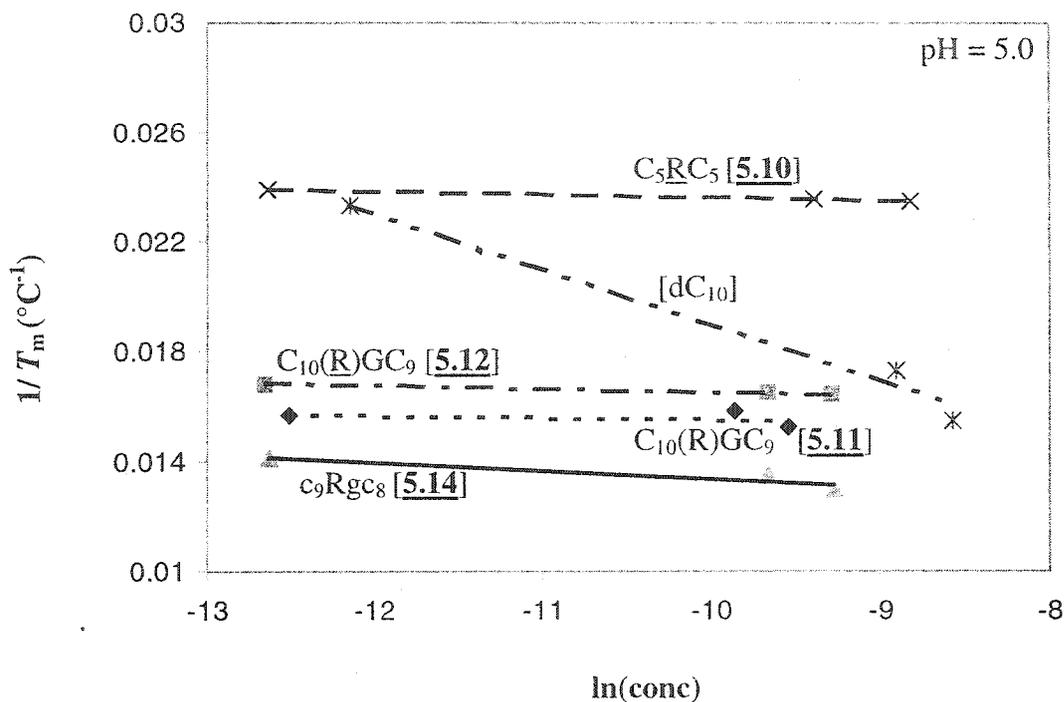


Figure 5.16: Van't Hoff plots showing T_m independence of ribocytidine-rich oligomers with (UUCG) loops. The control deoxycytidine oligomers are shown as controls. All measurements were done in 10 mM sodium acetate buffer at pH 5.0.

5.5.5 Gel Mobility Shift Assay

Native gels were conducted at pH 4.2 in order to confirm the formation of duplex structure by the ribocytidine-rich oligomers. **Figure 5.17** shows the relative electrophoretic mobility of these oligomers along with the control sequence dC_{10} . Two hairpin markers of mixed base sequence were run, one 22 nucleotides (RNA) and the other 42 nucleotides (DNA) in length. The migration of dC_{10} is similar to that of the 42-nucleotide DNA marker, consistent with C-tetrad formation for this oligomer. In sharp contrast, the oligomers $C_{10}(R)GC_9$ and $C_{10}(R)GC_9$ exhibit electrophoretic migration similar to that of the 22-nucleotide RNA marker. In a similar manner, the complex formed by C_5RC_5 [5.10] exhibits a similar gel mobility to that of a 16-nucleotide hairpin marker. This strongly suggests that these oligomers associate into duplex structures and rules out the possibility of any C-tetrad formation.

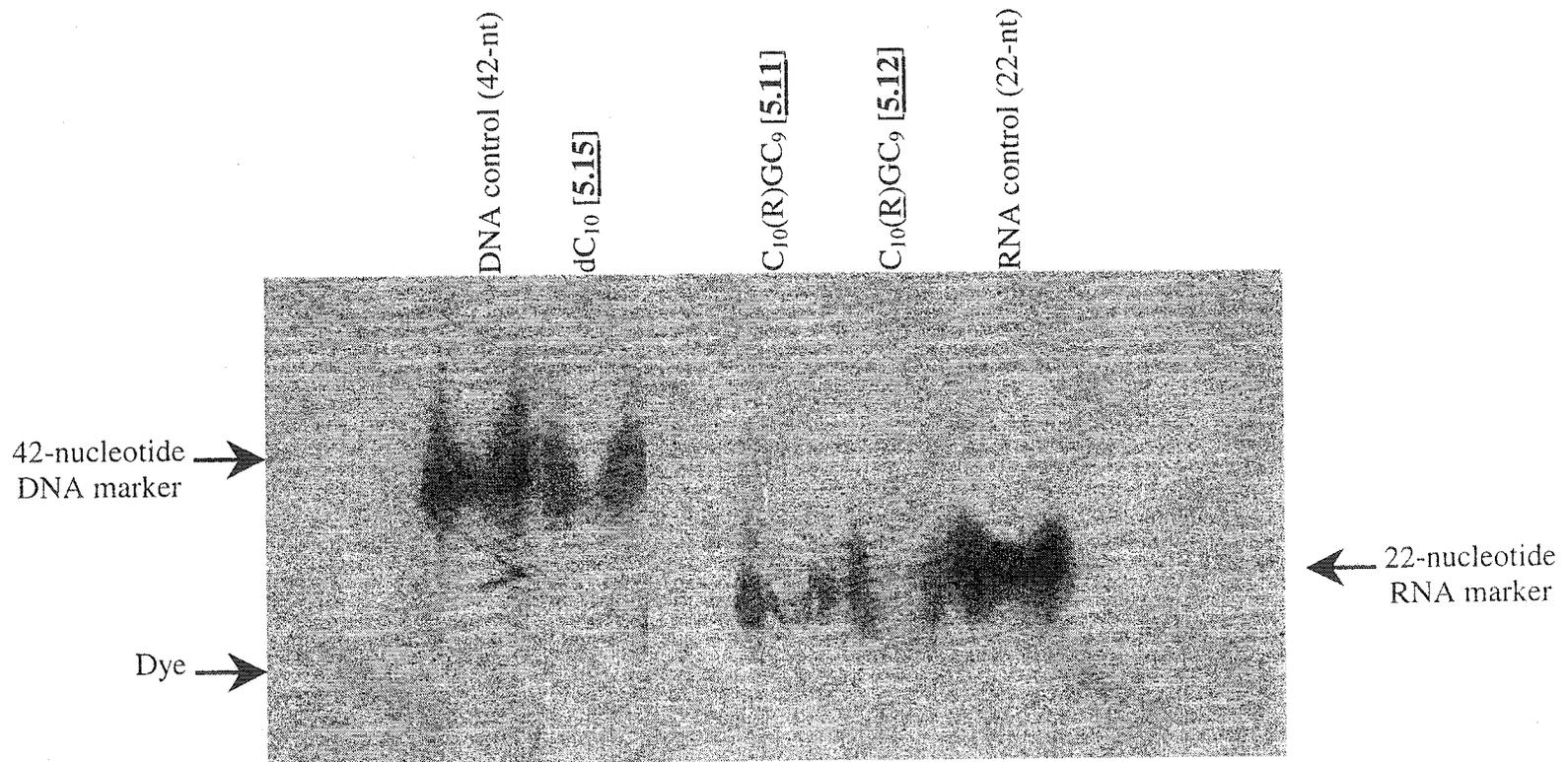


Figure 5.17: Gel mobility shift assay (native conditions) of various cytosine-rich oligomers. Samples (~0.35 ODU) were dissolved in 10 μ l loading buffer [30% sucrose in 10 x TAE buffer, pH = 4.2], heated to 95 °C and then cooled down slowly to room temperature and finally incubated at 4 °C for a week prior to analysis. The samples were loaded on a 16 % acrylamide gel (no urea added) and the gels were run at low voltage (100 – 150 V) for 5-6 h in the fridge (4 °C) at pH 4.2. The sequence of the DNA control is: 5'- ggc gtc atg tca tgt catg agtc catg gca tgg cat ggc gcc-3', while that of the RNA control is: 5'-GGC GUC AUG AGUC CAU GGC GCC-3'.

5.5.6 Structural Characterization of the Duplex via Circular Dichroism Studies

The high thermal stability of the complexes formed by ribocytidine oligomers allowed us to characterize their conformational features at various pHs via CD spectroscopy. As expected, C_5RC_5 [5.10], $C_{10}(R)GC_9$ [5.11], and $C_{10}(R)GC_9$ [5.12] do not display CD spectral properties characteristic of i-motif formation, *i.e.*, neither a positive CD band at *ca.* 288 nm nor a negative CD peak at *ca.* 267 nm is detected.^{271,272} The CD profiles at pH 5.0 exhibit a strong positive CD band at *ca.* 278 nm and a characteristic negative CD peak at *ca.* 212 nm. These features are in sharp contrast to those typical of i-motif structures. **Figure 5.18 A** compares the CD spectra of $C_{10}(R)GC_9$ and $C_{10}(R)GC_9$ to that of the deoxycytidine control c_9Rgc_8 [5.14] which forms an i-motif structure under these conditions.

A close inspection of the CD spectra of the ribocytidine-rich oligomers under study reveals a slight global similarity to A-form helices.^{4,8} Normally, A-form RNA duplexes display a characteristic negative CD band at *ca.* 212 nm and a strong positive CD peaks at *ca.* 264 nm. RNA:DNA hybrids display “A-like” CD signatures but their positive CD bands are slightly red-shifted relative to the A-form helices. In this study, the ribocytidine-rich oligomers $C_{10}(R)GC_9$ [5.11], and $C_{10}(R)GC_9$ [5.12] exhibit negative CD bands at *ca.* 212 nm, while their positive CD peaks occur at *ca.* 278 nm. **Figure 7.18 B** compares their CD spectra (*e.g.*, $C_{10}(R)GC_9$) with those of typical RNA, DNA and RNA:DNA duplexes of mixed base sequence and of the same length. One can deduce that the spectra of $C_{10}(R)GC_9$ [5.11], and $C_{10}(R)GC_9$ [5.12] fall as mixtures of both A- and B-form helices, with global similarity to the A-family. Their long wavelength CD peak is red-shifted (in the direction of DNA helices) and appears as a mixture of both A- and B- peaks. Their 212 nm is typical of A-helices.

Heating to 70 °C led to the disappearance of the 212 nm band. Furthermore, this was accompanied by a decrease in the intensity of the long wavelength CD peak (**Figure 5.19**), a behavior that is typical of A-form duplex structures. The same trend was observed for the RNA and RNA:DNA duplex controls (the 212 nm band of the RNA duplex control did not disappear completely due to its partial denaturation as a result of its high thermal stability).

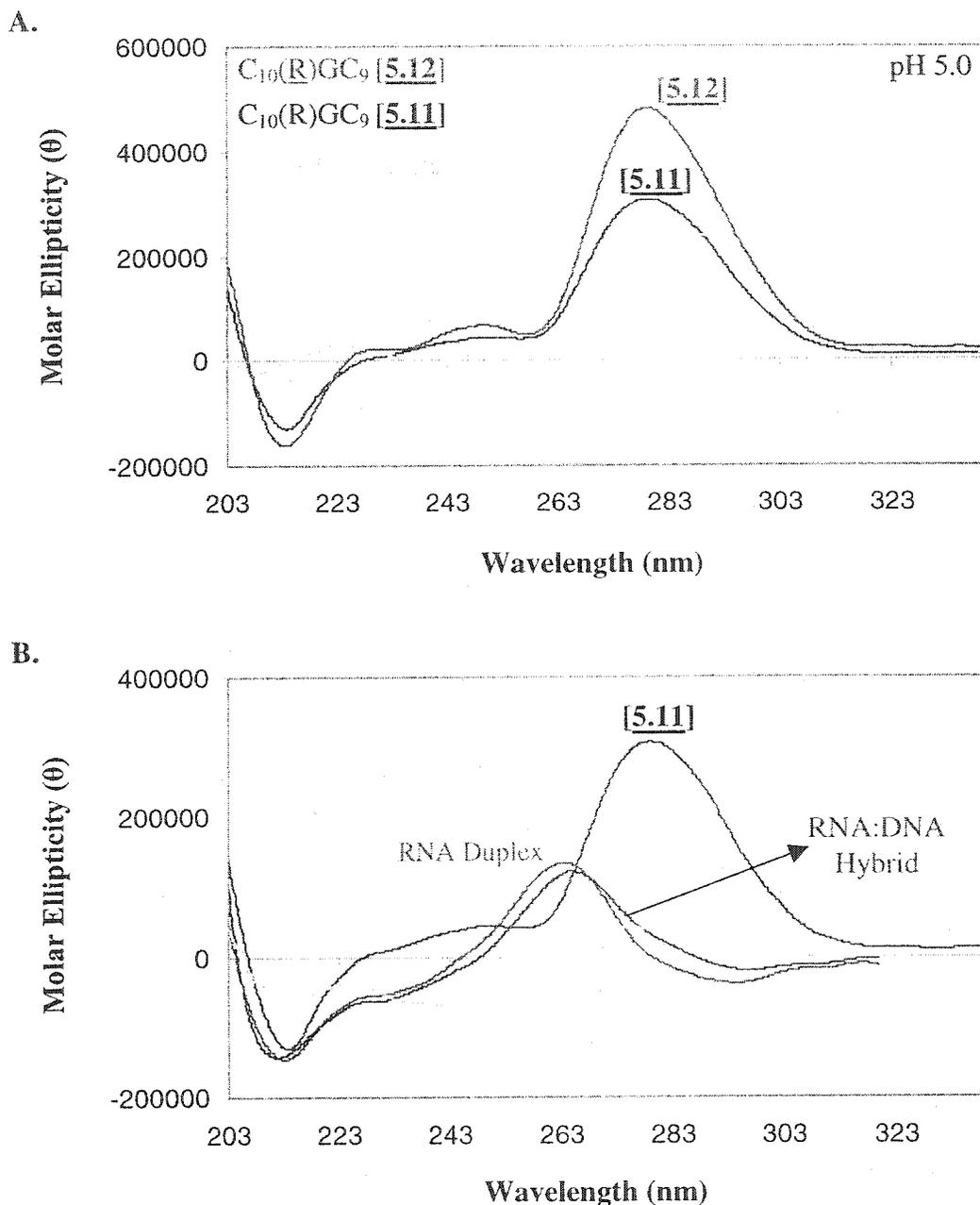
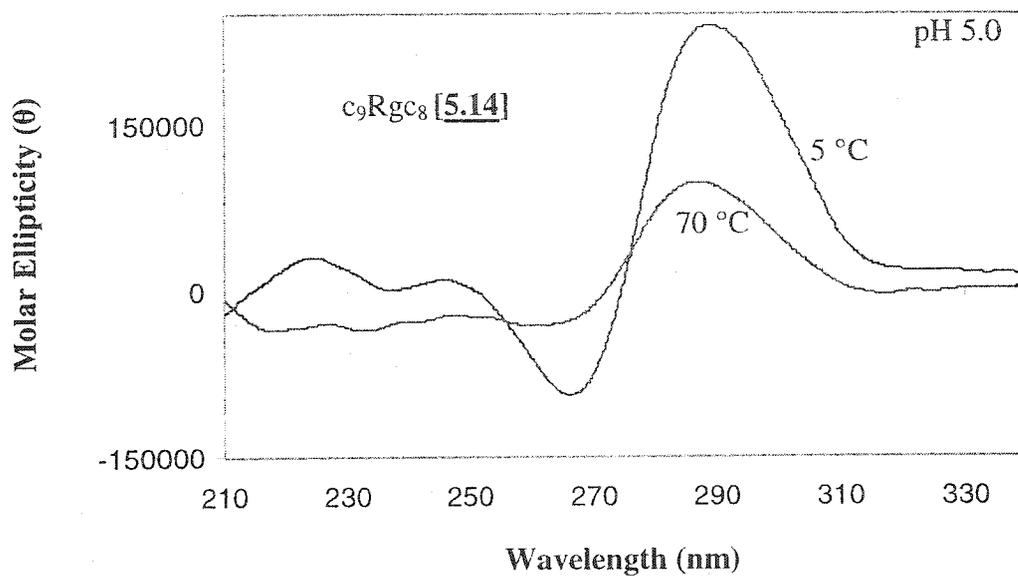


Figure 5.18: Circular dichroism spectra of ribocytidine-rich oligomers at 5 °C. (A) compares the CD spectra of $C_{10}(R)GC_9$ [5.11] and $C_{10}(R)GC_9$ [5.12] relative to c_9Rgc_8 [5.14]. (B) compares the CD spectrum of $C_{10}(R)GC_9$ relative to control RNA, DNA, and RNA:DNA hairpin duplexes. The base sequence of the controls was the same: 5'-GGC GUC AUG (AGUC) CAU GGC GCC-3'. Measurements for [5.11], [5.12], and [5.14] were done in 10 mM sodium acetate buffer, pH 5.0 (oligomer concentration 2.3 μ M). Measurements for the control sequences were done in 0.01 M Na_2HPO_4 , 0.1 mM Na_2EDTA buffer, pH 7.0 (oligomer concentration 4.5 μ M). All molar ellipticities were normalized to strand concentration and background buffer was subtracted from all scans.

A.



B.

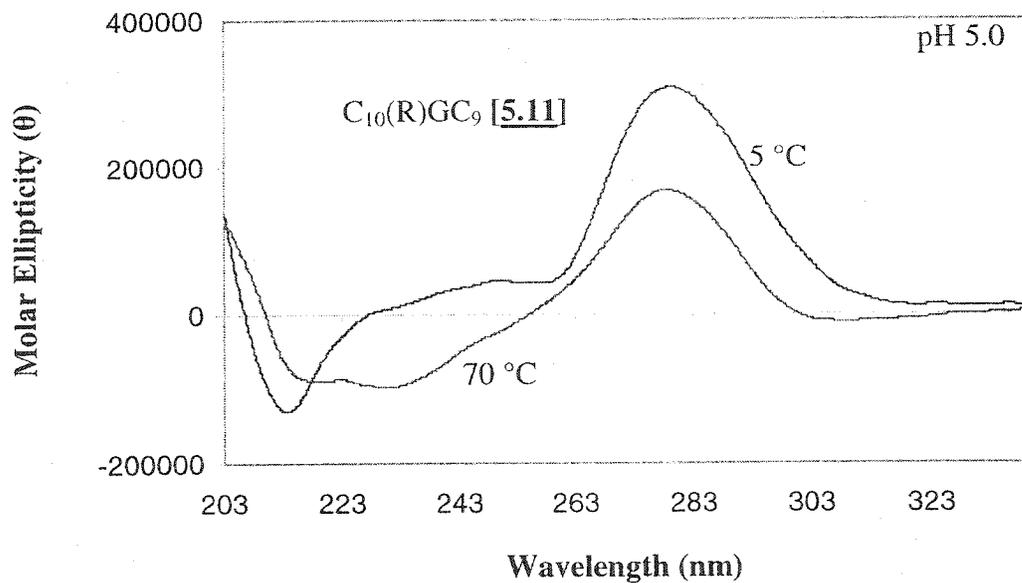
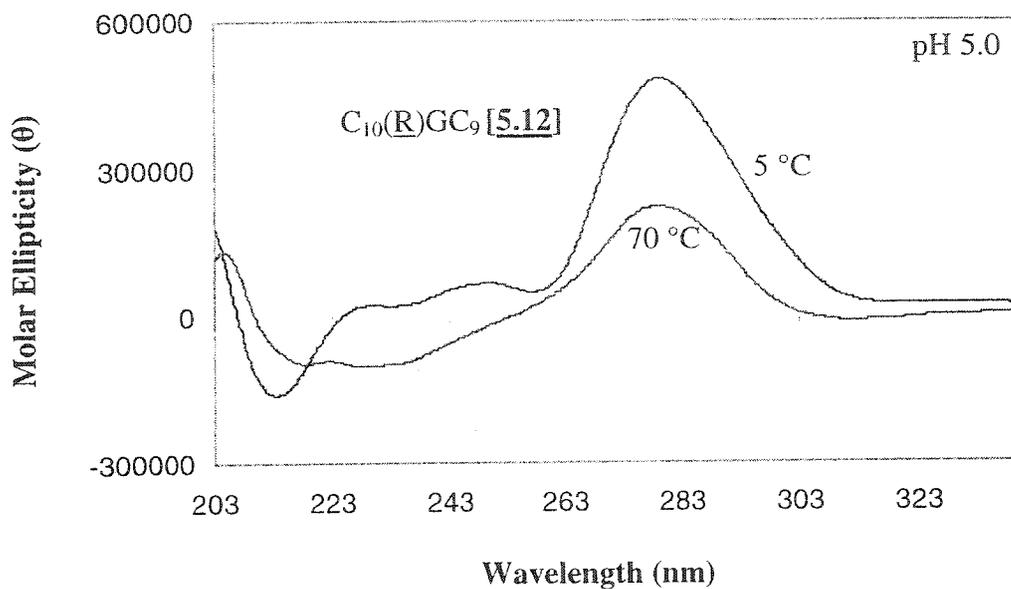


Figure 5.19 A and B: Circular dichroism spectra of c_9Rgc_8 and $C_{10}(R)GC_9$ at 5 °C and 70 °C. All measurements were done in 10 mM sodium acetate buffer, pH 5.0. Oligomer concentrations were 2.3 μ M. Molar ellipticities were normalized to strand concentration.

C.



D.

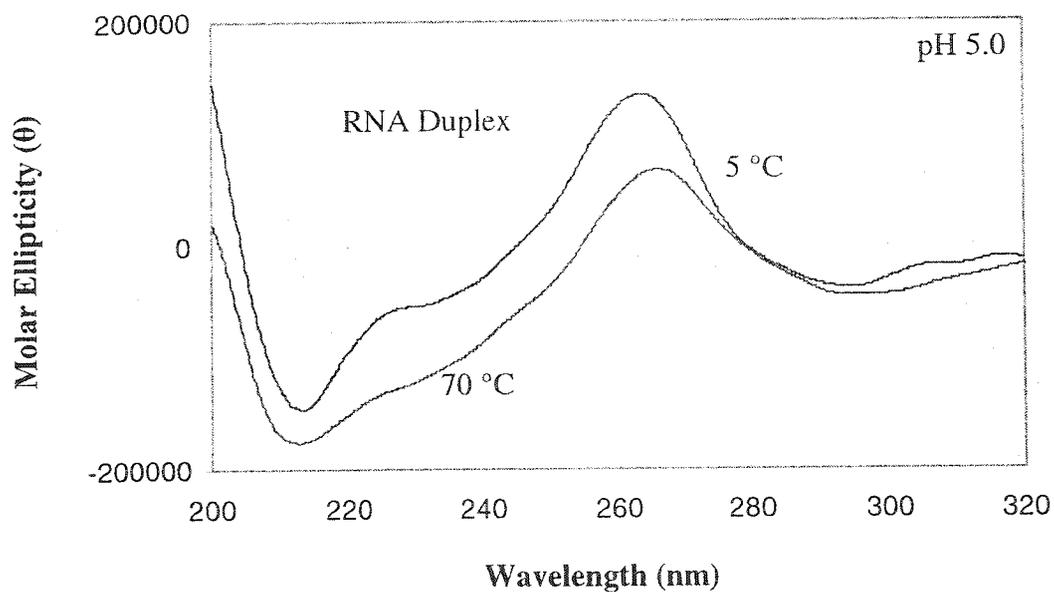
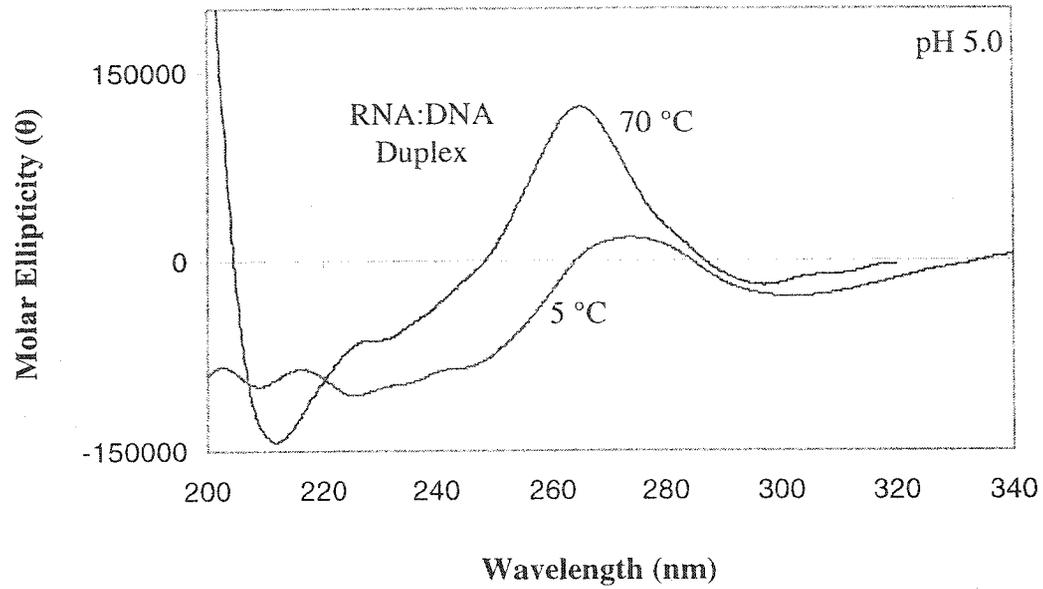


Figure 5.19 C and D: Circular dichroism spectra of $C_{10}(\underline{R})GC_9$ and an RNA hairpin duplex at 5 °C and 70 °C. Measurements for $C_{10}(\underline{R})GC_9$ were done in 10 mM sodium acetate buffer, pH 5.0 at an oligomer concentration of 2.3 μM . Measurements for the RNA hairpin duplex were done in 0.01 M Na_2HPO_4 , 0.1 mM Na_2EDTA buffer, pH 7.0 (oligomer concentration 4.5 μM). All molar ellipticities were normalized to strand concentration and the background buffer was subtracted from all scans.

E.



F.

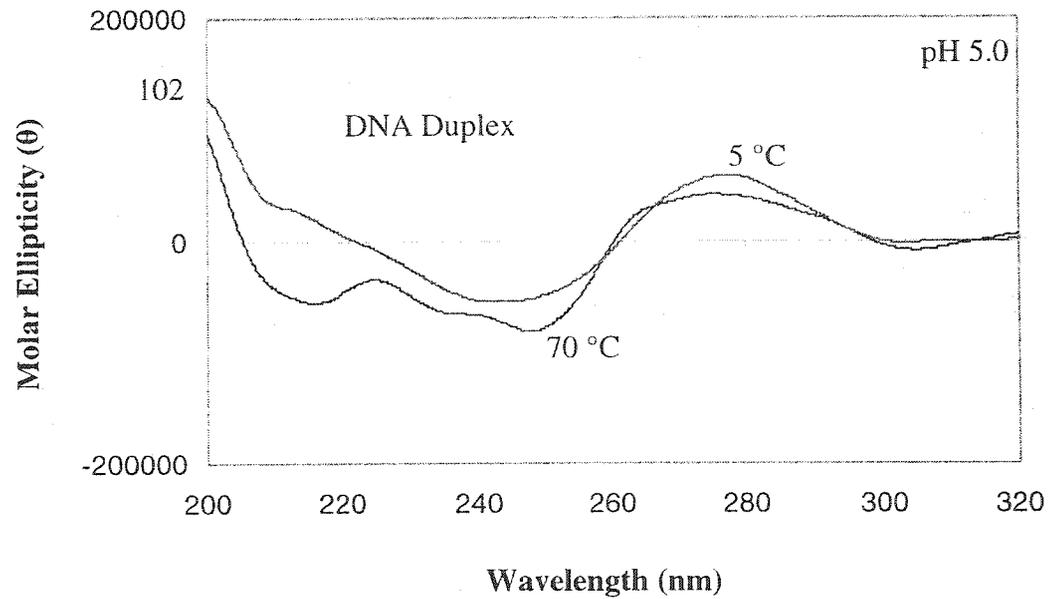


Figure 5.19 E and F: Circular dichroism spectra of DNA and RNA:DNA hairpin duplex controls at 5 °C and 70 °C. Measurements were done in 0.01 M Na_2HPO_4 , 0.1 mM Na_2EDTA buffer, pH 7.0 at an oligomer concentration of 4.5 μM . Molar ellipticities were normalized to strand concentration and the background buffer was subtracted from all scans.

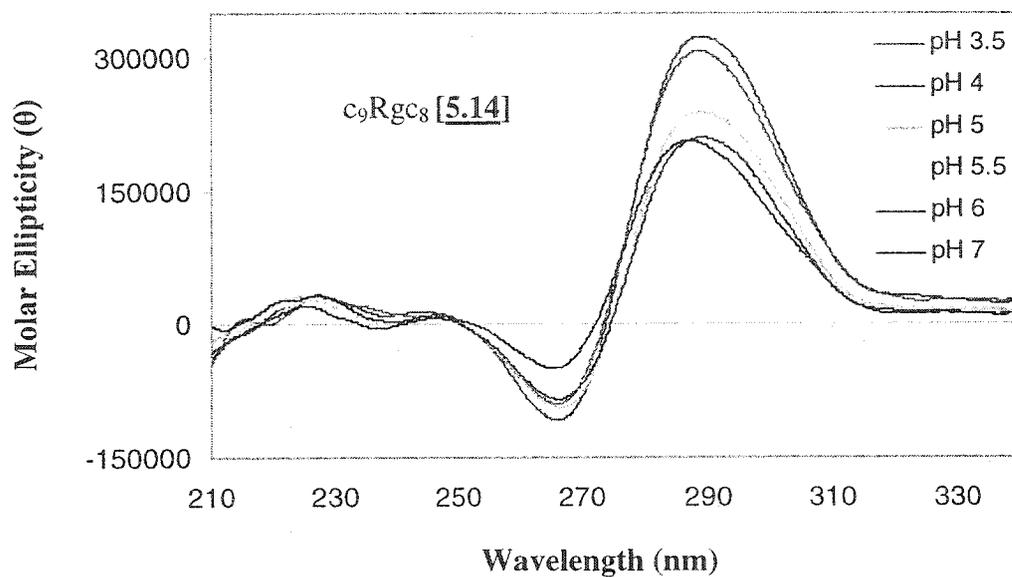
The CD spectra of the ribocytidine-rich oligomers displayed pH dependency in a manner that is different from that of deoxycytidine oligomers. The spectra of the former maintain their overall characteristics across the entire pH range under scrutiny, with slight differences in the amplitude of the CD peaks. For example, the CD spectra of C₁₀(R)GC₉ recorded at pH 4.0 – 7.0 reveal that there is a blue-shift and a gradual increase in the intensity of the of long wavelength CD peak with increasing pH, with the 212 nm band remaining almost unchanged (**Figure 5.20**). In sharp contrast, the control deoxycytidine oligomers c₉Rgc₈ [5.14] and dC₁₀ [5.15] displayed CD spectral features characteristic of i-motif formation. With deviations from the optimum pH, a decrease in the intensity of their long wavelength CD band accompanied with a gradual loss of their 265 nm band was observed (**Figure 5.20** shows c₉Rgc₈ as an example). This behavior was not detected for either C₅RC₅ [5.10], C₁₀(R)GC₉ [5.11] or C₁₀(R)GC₉ [5.12].

In summary, the above studies rule out the possibility of i-motif formation by ribocytidine-rich oligomers. Furthermore, it raises the interesting possibility for formation of antiparallel A-like duplexes instead. Although the negative 265 nm CD peak characteristic of C·C+ base pairing does not appear in the CD spectra of these oligomers, the behavior exhibited by their UV melting curves with varying pH suggests that complex formation is dependent on acidity of solution, a phenomenon which involves C·C+ base pairing. To recall, the oligomer TRT [2.31] with deoxythymidine stem residues does not associate into any complex under any of the above conditions, thus confirming that the complex observed by the ribocytidine-rich oligomers is mainly due to C·C+ base pairing interactions.

5.5.7 Evidence for the Formation of Antiparallel But Not Parallel C·C+ Duplexes

The Damha Laboratory has demonstrated the inability of 3',5'-RNA to form parallel stranded duplexes with hemiprotonated C·C+ base pairs. Their studies could not detect any hypochromic structure for the branched rV-5 oligomer [5.16] at pH 5.0 as judged from UV melting experiments.²⁶⁴ Increasing chain length to obtain rV-10 [5.17] still did not induce the formation of any complex.

A.



B.

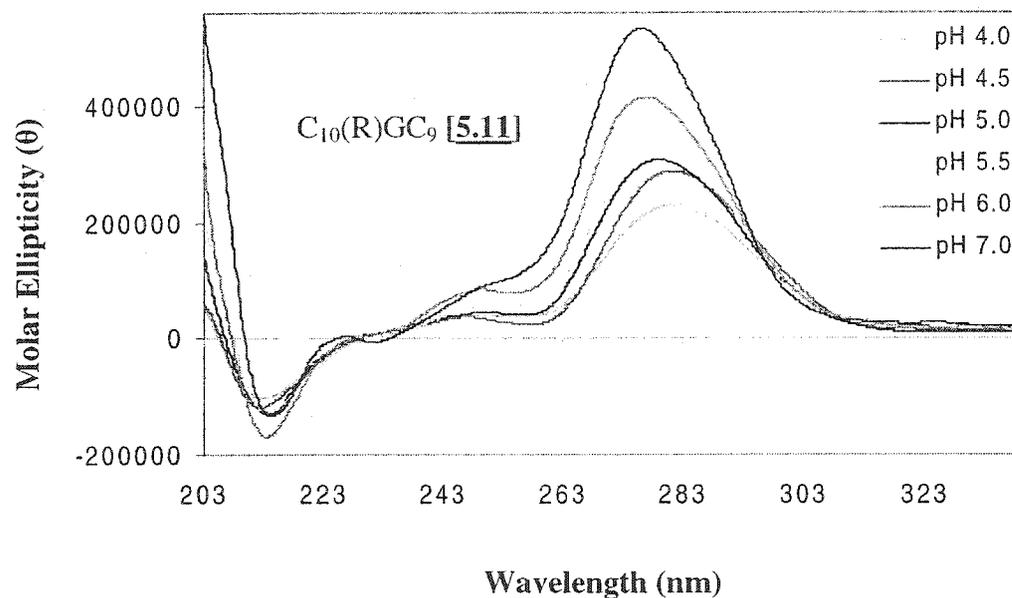


Figure 5.20: Circular dichroism spectra of c_9Rgc_8 and $C_{10}(R)GC_9$ at various pH. All measurements were done at 5 °C in either 10 mM sodium acetate buffer (pH 3.5 - 5.5) or 10 mM sodium cacodylate buffer (pH 5.5 - 7.0). Oligomer concentrations were 2.3 μ M. Molar ellipticities were normalized to strand concentration at all pH values.

As mentioned earlier, the presence of the branching rA linker forces the strands to be in a parallel orientation. This is in sharp contrast to our studies in which the ribocytidine strands exist in an antiparallel orientation by virtue of the presence of the loop moiety.

Based on the above experiments, the association of ribocytidine-rich oligomers into antiparallel duplex structures with hemiprotonated C·C⁺ base pairs is highly possible. As a first step towards supporting this hypothesis, we showed that the transition melting temperature was pH dependent, indicative of cytosine protonation. Furthermore, the reversible melting curves showed no hysteresis and suggested fast kinetics of complex association. The UV melting temperature displayed an independent behavior of oligomer concentration over a 30-fold range, thus confirming a *unimolecular* transition. This was further corroborated by native gel electrophoresis. The complexes formed by ribocytidine-rich oligomers migrated at the same rate as that of a control duplex RNA marker with the same length. Additionally, CD studies confirmed the presence of a complex that is mixture of both A- and B-form conformations with more similarity to the A-form family. All these results taken together clearly illustrate that ribocytidine-rich oligomers with UUCG tetraloops form antiparallel A-like duplex stems with hemiprotonated C·C⁺ base pairs.

5.6 2',5'-RNA ASSOCIATES INTO AN ANTIPARALLEL C·C⁺ DUPLEX

Depending on base-sequence, 2',5'-linked RNA has been shown to associate into duplex as well as triplex (but not i-motif) structures of lower thermal stability relative to the 3',5'-RNA regioisomer.^{121,122,125,141,264} Robidoux and Damha recently reported that 2',5'-rV5 [5.18] does not associate into any complex at pH 5.0, indicative of the inability of cytidylic 2',5'-RNA to form either a tetrameric i-motif or a parallel stranded duplex.²⁶⁴ To understand this phenomenon and explore further the tendency of cytosine-rich 2',5'-RNA to associate into other structures, we designed and studied the all-2',5'-RNA oligomer C₁₀RGC₉ (5.13; Table 5.4). The two ribocytidine stretches are linked by a 2',5'-RNA (UUCG) loop which sets the two stem strands in an antiparallel orientation.

In 50 mM sodium citrate buffer (pH 4.6), C₁₀RGC₉ surprisingly forms a complex whose UV melting profile (at 265 nm) yields a monophasic transition with a 1.3 %

hypochromic change and a T_m of *ca.* 64.8 °C (Figure 5.21). To confirm that cytosine protonation is involved in complex formation, the melting temperature was monitored across a 3.5 –7.0 pH range. Figure 5.22 shows the melting curves as a function of pH. The maximum T_m was obtained at a pH close to pK_a of cytosine. The complex could not be detected at pH <4.6 (Table 5.6).

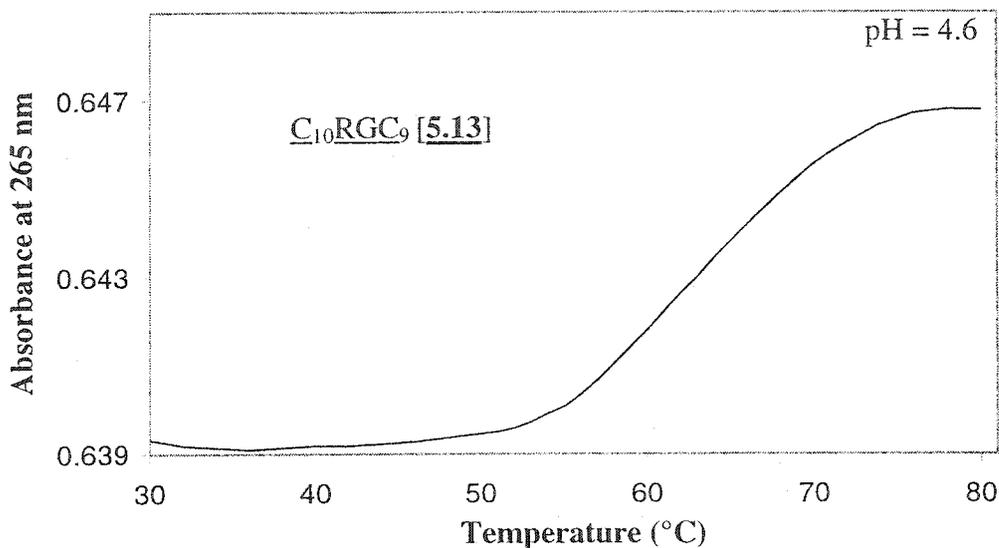


Figure 5.21: UV melting curve of $C_{10}RGC_9$ [5.13]. Measurements were conducted at 265 nm in 50 mM sodium citrate buffer, pH 4.6. Oligomer concentration was 2.3 μ M.

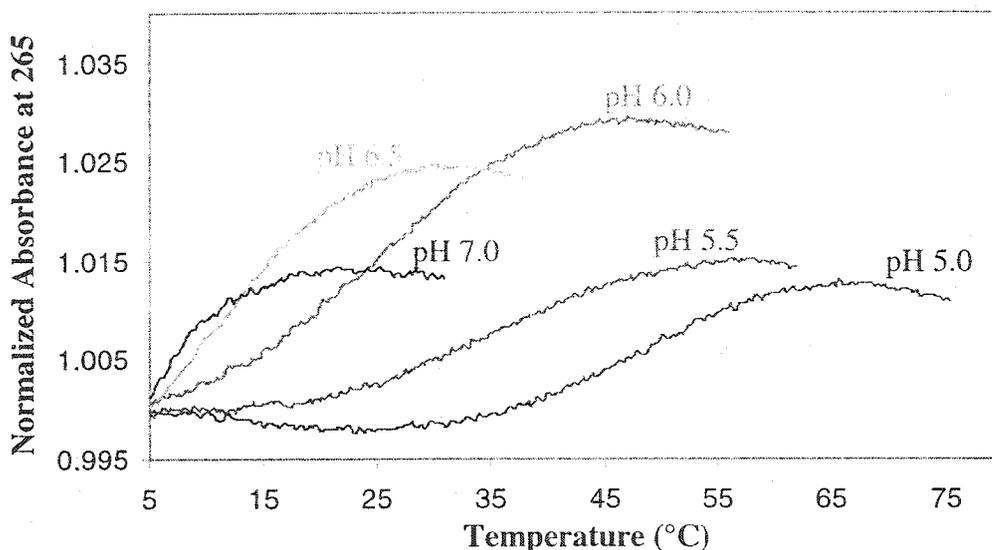


Figure 5.22: UV melting curves of $C_{10}RGC_9$ as a function of pH. All measurements were recorded at 265 nm in either 10 mM sodium acetate (pH 3.5-5.5) or 10 mM sodium cacodylate buffer (pH 5.5-7.0). Oligomer concentrations were 2.3 μ M.

Table 5.6: UV melting data for the all-2',5'-RNA cytidine-rich oligomer [5.13]

pH	<u>C₁₀RGC₉</u> [5.13]	
	<i>T_m</i> (°C)	%H
3.5	-	-
4.0	-	-
4.5	n.c.	n.c.
5.0	44.8	1.5
5.5	34.8	1.5
6.0	27.3	2.8
6.5	12.9	2.4
7.0	< 9.0	n.c.

n.c. = not calculated due to broad transition. All measurements were recorded at 265 nm in either 10 mM sodium acetate (pH 3.5-5.5) or 10 mM sodium cacodylate buffer (pH 5.5-7.0). Oligomer concentrations were 2.3 μ M. The UV melting temperatures were computed according to the base line method. All values represent the average of two independent measurements. The error in melting temperature is ± 0.5 °C. Percentage hypochromicity (%H) was calculated from UV absorbances at low (A_0) and high (A_f) temperatures according to the following equation: $\%H = (A_f - A_0) / A_f$. (n.c. = not calculated).

Thermal melting studies conducted over a 30-fold increase in oligomer concentration revealed a concentration-independent behavior, indicative of a unimolecular association process for the complex formed by C₁₀RGC₉ (data not shown). The pH value did not have an effect on the kinetics of intramolecular folding as evident from the reversible melting curves. At pH 5.0 or 7.0, the complex formed by C₁₀RGC₉ did not show any hysteresis, and the melting curves were completely cooperative and reversible (data not shown). This behavior is in sharp contrast to that observed for i-motif structures and is also supportive of fast kinetic of association.

Circular dichroism studies clearly demonstrate that the complex (pH 5.0) exhibits a strong positive CD peak at *ca.* 281 nm and a hump at *ca.* 224 nm (**Figure 5.23 A**). No band was observed at 288 nm, thus eliminating the possibility of i-motif formation by C₁₀RGC₉. This is in sharp contrast to the CD spectrum of control c₉Rgc₈ showing CD peaks at 288 nm and at 265 nm, indicative of i-motif formation. Moreover, the spectrum does not look like that of the RNA isomers, C₁₀(R)GC₉ and C₁₀(R)GC₉. In an attempt to understand the unique CD signature displayed by C₁₀RGC₉, we ran two control oligomers, an RNA and DNA duplex. A comparison of the obtained spectra reveals that the CD spectrum of C₁₀RGC₉ is very similar to that of the DNA (but not RNA) duplex, though not purely B-form (**Figure 5.23 B**). The hump at 224 nm coincides with that of the DNA duplex, while the CD band at 281 nm exhibits crossovers similar to those of the DNA-duplex 278 nm band. Furthermore, heating to 70 °C results in disappearance of the 224 nm hump (**Figure 5.24**), a behavior similar to that observed upon heating of the DNA duplex (**Figure 5.19 F**). The CD profiles were dependent on pH. As pH was increased, the long wavelength CD peak became blue-shifted (to *ca.* 276 nm at pH 7.0) while the 224 nm hump was not affected (**Figure 5.25**). It is important to note that the intensity of the 281 nm band did not follow a consistent decline upon pH increase, a behavior that is different from that of C₁₀(R)GC₉ [**5.11**], and C₁₀(R)GC₉ [**5.12**].

The above observations were further supported by native gel electrophoresis conducted at pH 5.0. The complex formed by C₁₀RGC₉ at pH 5.0 migrates very close to that of a 22-nucleotide RNA marker (**Figure 5.26**). This strongly suggests that the transition seen for the complex formed by C₁₀RGC₉ is unimolecular and thus implies intramolecular folding of a single strand into an antiparallel duplex structure with hemiprotonated C·C+ base pairs.

All in all, the above findings reveal, for the first time, the ability of 2',5'-RNA to associate into anti parallel stranded duplexes with hemiprotonated C·C+ base pairs, a behavior that is similar to that of the regioisomeric 3',5'-RNA. However, the overall conformations adopted by each of these duplexes are different.

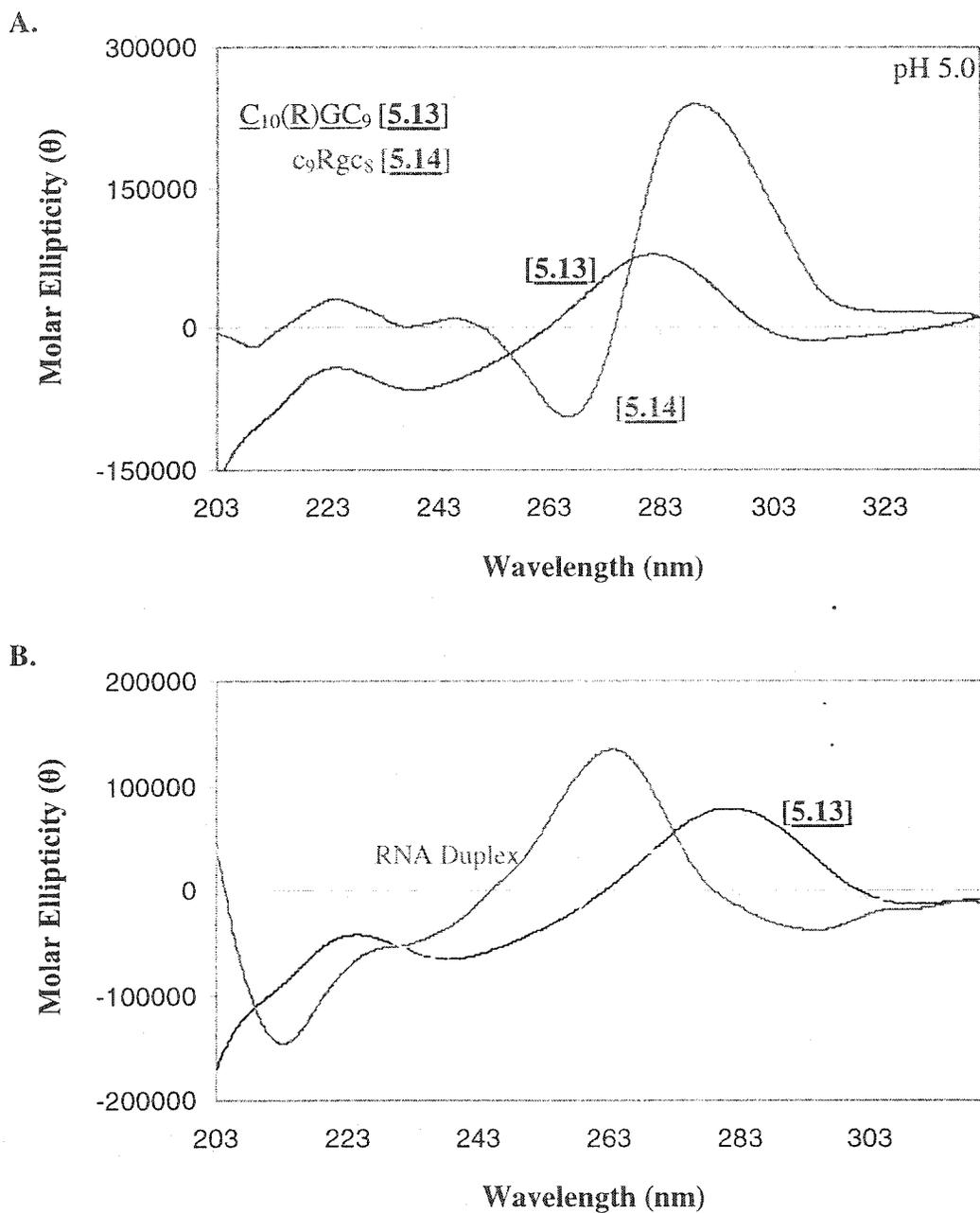


Figure 5.23: Circular dichroism spectra of the all 2',5'-RNA cytidine oligomer at 5 °C. (A) compares the CD spectrum of $C_{10}(R)GC_9$ [5.13] relative to c_9Rgc_8 [5.14]. (B) compares the CD spectrum of $C_{10}(R)GC_9$ relative to control RNA and DNA duplexes. The base sequence of the controls was the same: 5'-GGC GUC AUG AGUC CAU GGC GCC-3'. Measurements for [5.13] were done in 10 mM sodium acetate buffer, pH 5.0 (oligomer concentration 2.3 μ M). Measurements for the control sequences were done in 0.01 M Na_2HPO_4 , 0.1 mM Na_2EDTA buffer, pH 7.0 (oligomer concentration 4.5 μ M). All molar ellipticities were normalized to strand concentration and background buffer was subtracted from all scans.

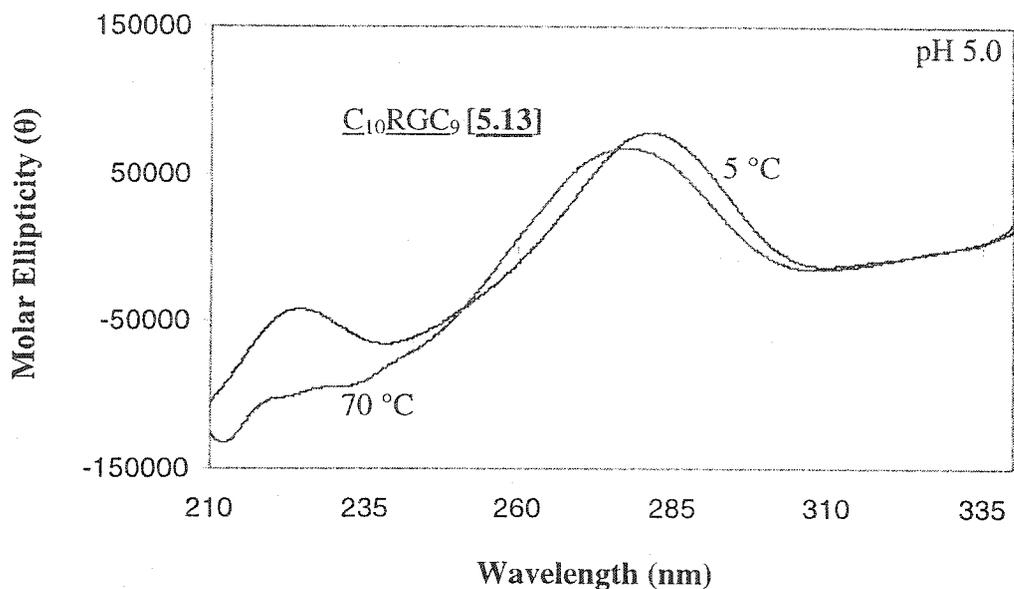


Figure 5.24: Circular dichroism spectra of $C_{10}RGC_9$ at 5 °C and 70 °C. All measurements were done in 10 mM sodium acetate buffer, pH 5.0. Oligomer concentrations were 2.3 μ M. Molar ellipticities were normalized to strand concentration.

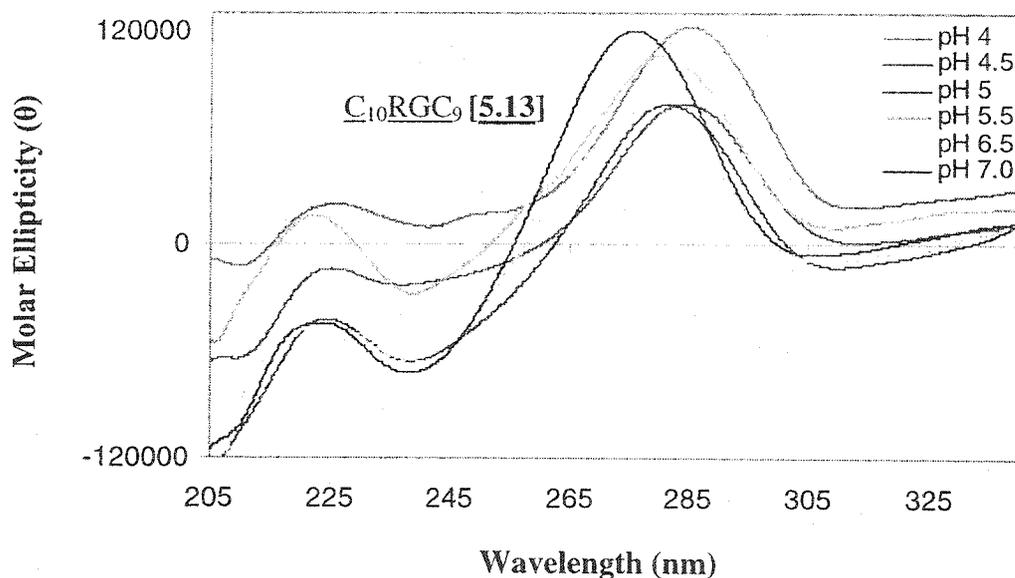


Figure 5.25: Circular dichroism spectra of $C_{10}RGC_9$ at various pH. All measurements were done at 5 °C in either 10 mM sodium acetate buffer (pH 3.5 - 5.5) or 10 mM sodium cacodylate buffer (pH 5.5 - 7.0). Oligomer concentrations were 2.3 μ M. Molar ellipticities were normalized to strand concentration at all pH values.

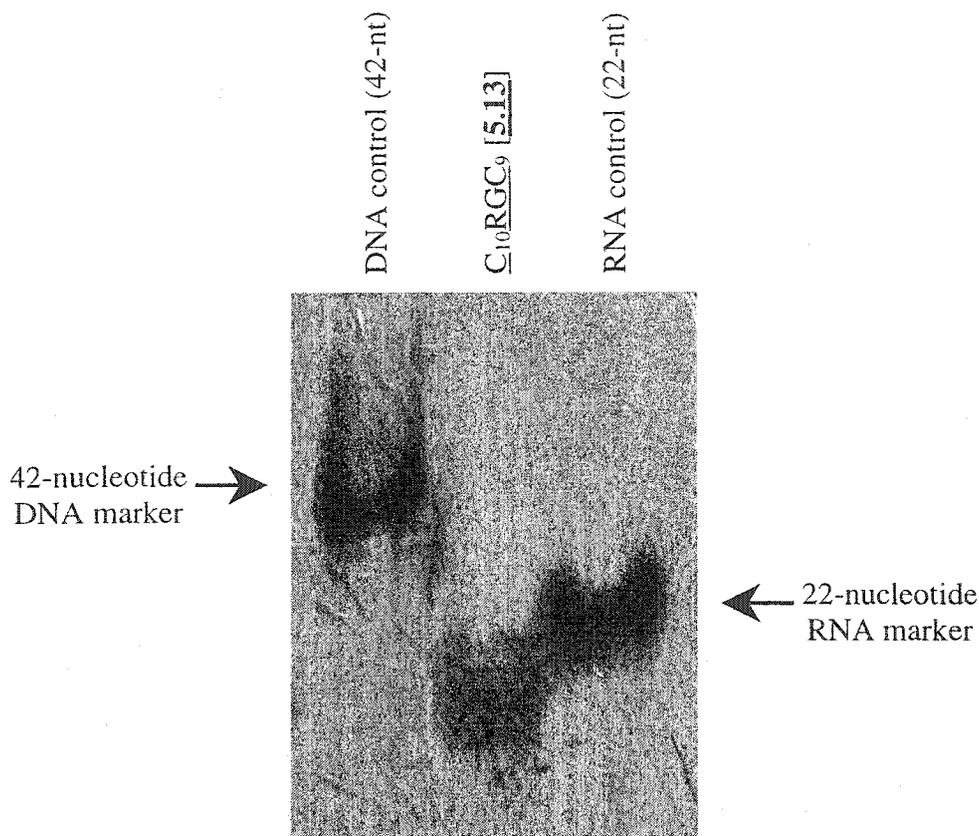


Figure 5.26: Gel mobility shift assay (native conditions) of C₁₀RGC₉. Samples (~0.35 ODU) were dissolved in 10 μ l loading buffer [30% sucrose in 10 x TAE buffer, pH = 5.0], heated to 95 °C and then cooled down slowly to room temperature and finally incubated at 4 °C for a week prior to analysis. The samples were loaded on a 16 % acrylamide gel (no urea added) and the gels were run at low voltage (100 – 150 V) for 10-12 h in the fridge (4 °C) at pH 5.0. The sequence of the DNA control is: 5'- ggc gtc atg tea tgt catg agtc catg gca tgg cat ggc gcc-3', while that of the RNA control is: 5'-GGC GUC AUG AGUC CAU GGC GCC-3'.

5.7 BIOLOGICAL STUDIES

We have shown earlier (chapter III) that the superstable (UUCG) loop motif is a key motif for recognition of hairpin oligomers by HIV-1 RT. The most potent inhibitors of HIV-1 RT were hairpins that adopted the A-helical form. Our present findings that ribocytidine-rich oligomers containing a (UUCG) loop adopt an overall A-like helical structure, led us to investigate them as potential inhibitors of HIV-1 RT RNase H activity. Phosphorothioate deoxycytidine oligonucleotides have been reported to inhibit the HIV-1

RT DNA polymerase activity.¹⁸⁰ However, the utility of ribocytidine-rich oligomers to act as RNase H inhibitors has not been explored.

5.7.1 Inhibition of HIV-1 RT RNase H Activity

We tested the ability of cytosine-rich oligomers to inhibit RNase H-mediated degradation of the RNA strand of RNA:DNA hybrid substrates. The test oligomers were incubated with HIV-1 RT in the presence of a [³²P]-RNA:DNA hybrid [50 mM Tris/HCl, 60 mM KCl, 2.5 mM MgCl₂, pH 7.8]. The results show that the ribocytidine-rich oligomers were more potent than the deoxycytidine-rich oligomers at inhibiting the activity of RNase H, *i.e.*, RNA cleavage in the RNA:DNA hybrid. **Table 5.7** gives the IC₅₀ values for the oligomers under study. The most potent among the molecules tested were C₁₀(R)GC₉ and C₁₀(R)GC₉, with IC₅₀'s 29.5 μM and 32.0 μM respectively. Inhibitory activity decreased upon decreasing stem length [C₅RC₅], a trend in agreement with that observed earlier (chapter III) for other duplex structures. The deoxycytidine hairpins were less potent and exhibited higher IC₅₀'s [for example see c₅Rc₅ and c₅Rgc₄; **Table 5.7**]. Surprisingly, mutating the deoxyribose to ribose moieties at the loop-closing base pair significantly enhanced inhibitory potency to the extent that the IC₅₀ value became comparable to that of C₁₀(R)GC₉ and C₁₀(R)GC₉. This is consistent with the hypothesis put forward in chapter III. Replacement of deoxyribose with ribose at the loop-closing base pair may optimize the base stacking interactions within the loop, which in turn stabilize the helical stem geometry and make it well defined.

Table 5.7: Inhibitory constants against HIV-1 RT RNase H of cytidine-rich oligomers.

Code	Designation	IC ₅₀ (μM)
<u>5.3</u>	c ₅ <u>R</u> c ₅	79.5
<u>5.5</u>	c ₅ <u>R</u> gc ₄	96.0
<u>5.7</u>	c ₄ CRGc ₄	37.5
<u>5.10</u>	C ₅ RC ₅	62.5
<u>5.11</u>	C ₁₀ (R)GC ₉	29.5
<u>5.12</u>	C ₁₀ (<u>R</u>)GC ₉	32.0

Values were determined from two different independent data sets. Errors in IC₅₀ are within ± 3 μM.

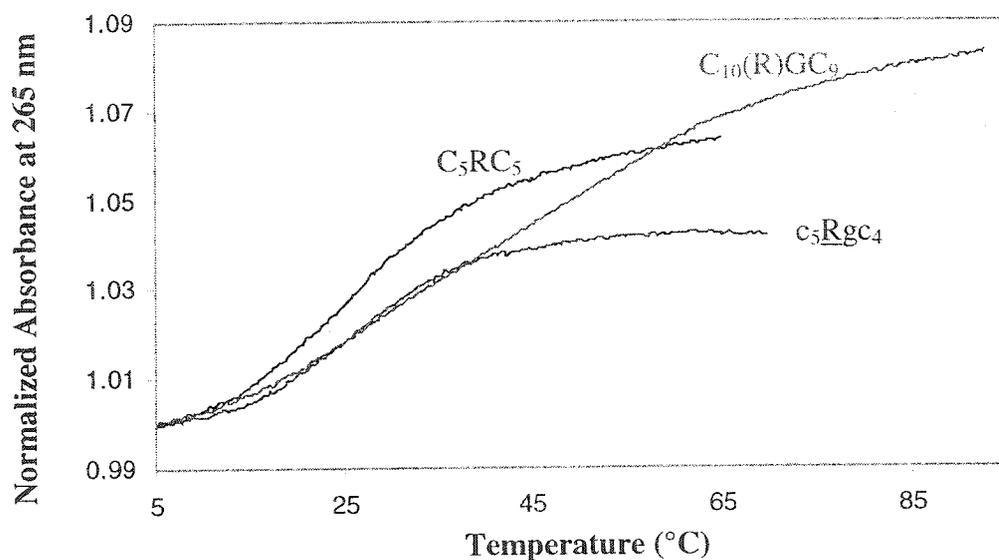
5.7.2 Inhibition of HIV-1 RT RNA- and DNA-Dependent DNA Polymerase Activity

Similar to the studies conducted in chapter III, we explored the ability of some representative cytosine-rich oligomers to inhibit DNA synthesis catalyzed by HIV-1 RT. The various oligomers were tested against both RNA- and DNA-dependent polymerase activities of HIV-1 RT. A 5'-[³²P] labeled primer was annealed to either RNA or DNA complementary sequences and was added, along with dNTPs, to a solution containing HIV-1 RT preincubated with the oligomer under investigation [50 mM Tris/HCl, 60 mM KCl, 2.5 mM MgCl₂, pH 7.8] (for a schematic sketch of the assay, see **Figure 3.17**). The oligomers under scrutiny were not able to inhibit the HIV-1 RT DNA polymerase activity at 50 μM concentration (data not shown). The DNA polymerization products, obtained from both RNA and DNA templates, were the same in the absence or presence of cytidine-rich oligomers.

The biological studies presented herein are still at a preliminary stage. Gel shift mobility assays in conjunction with UV cross-linking experiments are in progress in order to resolve the nature of interaction between the enzyme and the cytidine-rich aptamers as well as to identify the aptamer binding domain on the HIV-1 RT. Also, docking experiments are in progress in order to shed light into the molecular basis governing the interaction between the enzyme and the C-rich aptamers.

5.7.3 Insights into the Nature of the Complex Formed Under Assay Conditions

In order to gain insight into the nature of the active species adopted by the cytidine-rich oligomers, we conducted UV thermal melting studies under the same conditions as those employed in the inhibition assays [50 mM Tris/HCl, 60 mM KCl, 2.5 mM MgCl₂, pH 7.8]. Surprisingly, a complex that melts above room temperature is detected under these conditions (for representative examples, see **Figure 5.27**). This is remarkable in view of the fact that the complexes are detected at slightly basic conditions (pH 7.8). Obviously, more studies need to be undertaken in order to characterize the active form of the cytidine-rich oligomers under these simulated physiological conditions.



Code	Designation	T_m	%H
<u>5.3</u>	c_5Rc_5	20.0	1.3
<u>5.5</u>	c_5Rgc_4	27.8	4.1
<u>5.10</u>	C_5RC_5	30.0	6.2
<u>5.11</u>	$C_{10}(R)GC_9$	43.1	7.7
<u>5.12</u>	$C_{10}(\underline{R})GC_9$	42.9	7.6

Figure 5.27: UV thermal melting data for representative cytidine oligomers under physiological conditions. Measurements were recorded at 265 nm in 50 mM Tris/HCl, 60 mM KCl, 2.5 mM MgCl₂ buffer, pH 7.8. All oligomers were 2.3 μ M in total concentration. All samples exhibited monophasic and cooperative melting curves. The values represent the average of at least two independent measurements. The melting temperature (T_m) was calculated according to the base line method. The error in melting temperature is ± 0.5 $^{\circ}$ C.

5.8 CONCLUSIONS AND FUTURE PERSPECTIVES

We have shown that deoxycytidine oligomers with (UUCG) loops can associate into dimeric i-motif structures with hemiprotonated C·C⁺ base pairs. These structures displayed pH-dependent UV melting profiles and CD spectra. The incorporation of either 3',5'- or 2',5'-RNA loops within deoxycytidine oligomers significantly enhanced the thermal stability of the i-motif. It also gave rise to stable C-tetrad structures at neutral pH. The loop motif enhanced the kinetics of complex formation by reducing the molecularity of the process. At neutral pH, the denaturation and renaturation curves were not superimposable, consistent with the behavior of cytosine-rich oligodeoxynucleotides and in agreement with i-motif formation.^{263,268,278} A model was proposed for the i-motif structure formed by these oligomers. Furthermore, the nature and base sequence of the loop were shown to be important criteria governing the formation of a variety of possible multimeric i-motif structures.

We have investigated the possibility of complex formation by ribocytidine-rich oligomers. By linking the ribocytidine stretches with (UUCG) loops, it was possible to show, for the first time, that 3',5'- and 2',5'-RNA can form antiparallel stranded duplexes held together by hemiprotonated C·C⁺ base pairs. These duplexes displayed pH-dependent UV and CD profiles and were stable at neutral pH. Their reversible melting curves showed no hysteresis and indicated very fast kinetics of association. They also proved to be potent inhibitors that are specific towards the RNase H but not DNA polymerase activity of HIV-1 RT.

All in all, the data obtained with looped deoxycytidine-rich oligonucleotides demonstrated considerable stabilization of the i-motif structure. The results reported herein may be useful in future design of new probes capable of forming highly stable C-tetrads. To date, several cytidine-rich sequences have been identified in DNA chromosomes. An example is the sequence d(TC₃GT₃C₂A) found in the centromere of some human chromosomes and which has been shown to form a dimeric i-motif tetrad in solution.²⁵³ It constitutes part of a region of the CENP-B box, a centromere that acts as the binding site for centromere protein B.²⁵³ The tetrads described here could find use as

affinity ligands to 'fish' proteins that recognize and bind specifically to cytosine-rich i-motif structures.

CHAPTER VI: CONTRIBUTIONS TO KNOWLEDGE

REMARKABLE STABILITY OF HAIRPINS CONTAINING 2',5'-RNA LOOPS

Hairpins that differ in the connectivity of phosphodiester linkages (3',5'- versus 2',5'-linkages) were synthesized on solid-phase and subsequently characterized via UV, CD, HPLC, gel electrophoresis, NMR, and MALDI-TOF techniques. Specifically, we constructed hairpins containing one of six stem duplex combinations, *i.e.*, DNA:DNA, RNA:RNA, DNA:RNA, 2',5'-RNA:RNA, 2',5'-RNA:DNA and 2',5'-RNA:2',5'-RNA, and one of three tetraloop compositions, *i.e.*, 2',5'-RNA, RNA, and DNA. All hairpins contained the conserved and well-studied loop sequence 5'-...C(UUCG)G...-3'. We showed that the 2',5'-linked loop C(UUCG)G, *i.e.*, ...C_{3'p5'}U_{2'p5'}U_{2'p5'}C_{2'p5'}G_{2'p5'}G_{3'p5'} ..., like its 'normal' RNA counterpart, forms an unusually stable tetraloop structure. We also showed that the stability imparted by 2',5'-RNA loops is dependent on base sequence, a property that is shared with the regioisomeric 3',5'-RNA loops. Remarkably, we find that the stability of the UUCG tetraloop is virtually independent of the hairpin stem composition, whereas the native RNA tetraloop exerts extra stability only when the stem is duplex RNA. As a result, the relative stabilities of hairpins with a 2',5'-linked tetraloop, *e.g.* ggac(UUCG)gtcc ($T_m = 61.4$ °C), are often superior to those with RNA tetraloops, *e.g.* ggac(UUCG)gtcc ($T_m = 54.6$ °C). In fact, it has been possible to observe the formation of a 2',5'-RNA:DNA hybrid duplex by linking the hybrid's strands to a (UUCG) loop. These duplexes, which are not stable enough to form in an intermolecular complex, were stable at room temperature ($T_m \sim 50$ °C). Also, an all-2',5'-RNA oligomer was shown to fold intramolecularly into a hairpin structure that had lower thermal stability relative to the corresponding all-RNA and all-DNA hairpins. The 2',5'-RNA hairpin loop exhibited a global conformation that was intermediate between the A- and B-forms.

These studies provided the first demonstration that hairpins containing 2',5'-linked RNA loops exhibit extra-stability and adopt a unique 3D-structure that is different from that of the native RNA structure.

SOLUTION NMR STRUCTURE OF 2',5'-RNA (UUCG) LOOP: A NOVEL STRUCTURAL MOTIF

NMR structural studies carried out in collaboration with Prof. K. Gehring and Dr. A. Denisov (Department of Biochemistry, McGill University) revealed the molecular basis for the unusual stability displayed by 2',5'-RNA loops. The 2',5'-UUCG loop is stabilized by a wobble U-G base pair, extensive base stacking and sugar – base contacts, and its cytosine residue protrudes into the solvent without participation in any of the stabilizing interactions. This identifies the 2',5'-RNA loop as a novel structural motif.

EVOLUTION OF A NUCLEIC ACID HAIRPIN LIBRARY VIA DIVERSITY-ORIENTED SYNTHESIS

Forty-five hairpin molecules based on DNA, RNA and 2',5'-RNA were synthesized via a diversity-oriented synthesis. This methodology provided access to an array of hairpin molecules possessing distinct conformational, structural and biological features.

DEVELOPMENT OF POTENT HAIRPIN INHIBITORS TARGETED SELECTIVELY AGAINST HIV-1 RT RNASE H

Screening assays were set up in collaboration with Dr. Kyung-Lyum Min from our research group in order to test the biological activity of hairpin members of our library. Four potent hairpin aptamers containing 3',5'- or 2',5'-RNA stems were able to selectively inhibit HIV-1 RT RNase H, an enzyme essential to HIV-1 replication. Inhibition is specific towards HIV-1 RT RNase H but not human RNase H. The aptamers do not interfere with the polymerase activity of HIV-1 RT. Gel-shift assays combined with UV cross-linking studies show that the inhibitors bind specifically to the RNase H domain of HIV-1 RT. The inhibition was shown to be strongly dependent on the unique conformation (folding) of the tetraloop, with hairpins bearing a 2',5'-RNA being more potent than the 3',5'-linked counterpart. Also, we show that the inhibition is dependent on helical conformation rather than thermal stability. Hairpin structures are significantly more stable towards hydrolysis by endo and exonucleases found in biological media. These investigations bring into light a new class of nucleic acid aptamers and provide the

first examples of *mini* hairpins (4 base pairs) that inhibit specifically the RNase H activity of HIV-1 RT without interfering with its polymerase activity.

YEAST RNASE III (Rnt1p) CLEAVES DNA IN DNA:RNA HYBRIDS

In collaboration with Prof. Sherif Abou Elela (University of Sherbrooke, Quebec), yeast RNase III was found to cleave the DNA strand of DNA:RNA hybrids. The cleavage is efficient, specific, and can only be directed by an RNA tetraloop. Studies conducted with sugar-modified oligonucleotides revealed that the 2'-hydroxyl group vicinal to the scissile phosphodiester linkage is not directly involved in the mechanism of phosphodiester internucleotide cleavage. This RNA-dependent DNA cleavage was shown to be specific to eukaryotic RNase III enzymes. Furthermore, a 2',5'-phosphodiester linkage was not cleaved by Rnt1p.

These studies shed light into the mechanism of action of dsRNA binding proteins. Specifically, the studies described revealed the very first example of an RNA-dependent DNA cleavage catalyzed by RNase III, which prior to these investigations was believed to be specific towards the hydrolysis of an RNA duplex only (*i.e.*, RNA-dependent RNA cleavage).

POSSIBLE TETRAPLEX FORMATION OF CYTOSINE-RICH HAIRPINS INCORPORATING 3',5'- and 2',5'-RNA LOOPS

2'-Deoxycytidine-rich hairpins based on 3',5'- or 2',5'-RNA (UUCG) loops were shown to associate into the so-called i-motif (or C-tetrad). The structures exhibited high thermal stabilities and were shown to display very fast kinetics of association in a pH-dependent manner.

The corresponding ribocytidine hairpins (with 3',5'- or 2',5'-UUCG loops) were shown for the first time to associate into antiparallel duplexes displaying hemiprotonated C-C⁺ base pairs. The formation of these duplexes is pH-dependent and their conformations are mixtures of both A- and B-type helices.

PUBLICATIONS AND INVENTION DISCLOSURES

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

- Hannoush, R.N., Min, K.-L., Damha, M.J. "Inhibition of HIV Reverse Transcriptase by Nucleic Acid Mini Hairpins", US Provisional patent, filed January 6, 2003.
- Hannoush, R.N., Denisov, A.Y., Gehring, K. and Damha, M.J. "Structure of 2',5'-Linked Tetraribonucleotide Loops: A Novel RNA Motif", Submitted to Nucleosides and Nucleotides
- Hannoush, R.N. and Damha, M.J. "Remarkable Stability of Hairpins Containing 2',5'-Linked RNA Loops", The Journal of the American Chemical Society, 2001, 123, 12368 -12374.
- Hannoush, R.N. and Damha, M.J. "Extra Stable 2',5'-Linked RNA Loops", Nucleosides and Nucleotides, 2001, 20, 1201-1204.
- Hannoush, R.N. and Damha, M.J. "Structure and Thermodynamics of Hairpins containing 2',5'-Linked RNA Loops", Journal of Biomolecular Structure and Dynamics, 2001, 18, 906.
- Hannoush, R.N., Lamontagne, B., Damha, M.J. and Abou Elela, S. "Yeast RNase III Cleaves DNA in DNA/RNA Hybrids", in preparation.
- Hannoush, R.N., Min, K.L. and Damha, M.J. "Mini Hairpins: A New Class of Aptamers that Selectively Inhibits HIV-1 Reverse Transcriptase RNase H", in preparation.
- Denisov, A.Y., Hannoush, R.N., Gehring, K. and Damha, M.J. "Solution NMR Structure of Unusually Stable RNA Hairpins Composed of 2',5'-Linked r(UUCG) Loops", in preparation.
- Hannoush, R.N. and Damha, M.J., "Stabilization of DNA C-Tetrads by 2',5'-RNA loops", in preparation.

CONFERENCE PRESENTATIONS AND INVITED LECTURES

- Hannoush, R.N., Denisov, A.Y., Gehring, K. and Damha, M.J., "A Novel RNA Structural Motif Based on 2',5'-Linked r(UUCG) Loops", 5th Graduate Chemistry and Biochemistry Conference, Montreal, Quebec, September 27-28, 2002.
- Hannoush, R.N., Denisov, A.Y., Gehring, K. and Damha, M.J., "Solution NMR Structure of Unusually Stable RNA Hairpins Composed of 2',5'-Linked r(UUCG) Loops", XV International Roundtable on Nucleosides, Nucleotides and Nucleic Acids, Leuven, Belgium, September 10-14, 2002.

- Hannoush, R.N., "2',5'-Linked RNA: A Brother or Cousin to RNA?", Harvard University, Cambridge, MA, USA, May 17, 2002.
- Hannoush, R.N., "Unusually Stable 2',5'-RNA Loops and Possible Tetraplex Formation", The Scripps Research Institute, La Jolla, CA, USA, December 13, 2001.
- Hannoush, R.N., "2',5'-Linked RNA: A Brother or Cousin to RNA?", Universite' de Sherbrooke, Ribo Club Meeting, Sherbrooke, Quebec, November 5, 2001.
- Hannoush, R.N., Ganeshan, K. and Damha, M.J., "Synthesis of Circular Nucleic Acids: Towards Understanding DNA/RNA Lariat Structures", 4th Graduate Chemistry and Biochemistry Conference, Montreal, Quebec, September 21, 2001. (**Best Poster Presentation**)
- Hannoush, R.N. and Damha, M.J., "Structure and Thermodynamics of Hairpins containing 2',5'-Linked RNA Loops", Albany 2001: The 12th Conversation in Biomolecular Stereodynamics: "The First Conversation of the Twenty First Century", State University of New York, Albany, June 19-23, 2001.
- Hannoush, R.N. and Damha, M.J., "Thermodynamics and Nuclease Stability of RNA Hairpin (UUCG) Tetraloops", McGill Organic Seminars, McGill University, Montreal, Quebec, June 3, 2001.
- Hannoush, R.N. and Damha, M.J., "Effect of 2',5'-Linked RNA Substitutions on the Thermodynamic and Nuclease Stability of RNA Hairpin (UUCG) Tetraloops", 84th Canadian Society for Chemistry Conference and Exhibition, Montreal, Quebec, May 23-27, 2001. (**Best Oral Presentation**)
- Hannoush, R.N. and Damha, M.J., "Super Stable 2',5'-RNA Tetra Loops", XIV International Roundtable on Nucleosides, Nucleotides and Their Biological Applications, San Francisco, California, September 10-14, 2000. (**Best Paper Presentation**)
- Hannoush, R.N. and Damha, M.J., "A Thermodynamic Study of Chimeric 2',5'/3',5' RNA/DNA Hairpins", 3rd Concordia Graduate Chemistry and Biochemistry Conference, Montreal, Quebec, March 17, 2000. (**Best Oral Presentation**)
- Hannoush, R.N. and Damha, M.J., "2',5'-Linked Oligoribonucleotides", 10th Quebec/Ontario Minisymposium in Synthetic and Bioorganic Chemistry, Saint-Sauver-des-Monts, Quebec, November 5-7, 1999.

CHAPTER VII: EXPERIMENTAL METHODS

7.1 GENERAL METHODS

7.1.1 General Reagents

The following reagents were purchased from Aldrich Chemical Company (Milwaukee, WI) and used as obtained: 4-dimethylaminopyridine (4-DMAP), sodium acetate, triethylamine tris(hydrofluoride) [TREAT.HF], sodium cacodylate, lithium perchlorate, sodium hydroxide, tetrabutylammonium fluoride, pentachlorophenol, and iodine. Acetic anhydride, disodium hydrogen phosphate, 1,2-dichloroethane, and analytical reagent grade hydrochloric acid, glacial acetic acid and sulfuric acid were all obtained from Fisher Scientific (Fairlawn, NJ). Magnesium chloride, potassium chloride, sodium chloride, triethylamine, and ethylenediamine tetraacetate (EDTA) were purchased from BDH and used as received. The following were purchased from Caledon: n-butanol, n-propanol, low-water acetonitrile, and trifluoroacetic acid. Other reagents used include: TRIS (Bio-Rad), anhydrous ethanol (Commercial Alcohols Inc.), and acetone (ACP).

Most solvents were distilled under nitrogen atmosphere or reduced pressure (for high boiling point solvents). Acetonitrile was dried over phosphorous pentoxide and then refluxed over calcium hydride prior to use. Dichloromethane was dried over CaCl_2 and then refluxed under nitrogen over CaH_2 . Pyridine (Caledon Laboratories Ltd., Georgetown, ON), collidine, N-methylimidazole, N,N-diisopropylethylamine (Aldrich), N,N-dimethylformamide (DMF) (BDH) were dried by fractional distillation over CaH_2 and were then stored over 4 Å molecular sieves.

7.1.2 Chromatography

Various compounds (for example, purchased phosphoramidites) were checked for their purity prior to use by performing thin-layer chromatography (TLC). The compounds were spotted on a TLC plate (Merck Kieselgel 60 F-254 aluminium-back analytical silica gel sheet, 0.2 mm thickness, EM Science, Gibbstwon, NJ) and eluted with an appropriate solvent. They were then visualized by a UV illuminator (Mineralite, emission wavelength

ca. 254 nm) and stained with trifluoroacetic acid vapors to check the integrity of the trityl groups.

7.1.3 Instruments

a. UV Spectroscopy

UV-VIS spectra and UV- thermal curves were recorded on a Varian CARY 1 UV-VIS spectrophotometer (Varian: Mulgrave, Victoria, Australia) equipped with a 6x6 multi-cell holder and a Peltier temperature controller. Data acquisition and analysis were done using Cary WinUV version 2 software (Varian Ltd.).

b. Circular Dichorism Spectroscopy

CD spectra were acquired on a JASCO J710 spectropolarimeter equipped with an external constant-temperature NESLAB RTE-111 circulating bath for temperature control. Fused quartz cells (Hellma, 165-QS) were used. The data were processed on a PC computer using WinTM based software supplied by the manufacturer (JASCO, Inc.).

c. NMR Spectroscopy

The NMR spectra were recorded on a Bruker DRX-500 spectrometer equipped with a ¹H/¹³C/³¹P triple resonance (x, y, z) gradient probe operating at 500.13 MHz of proton resonance. The proton chemical shifts were measured relative to internal DSS, and the phosphorus resonances were indirectly referenced to 85% H₃PO₄.²⁷⁹

d. Matrix-Assisted Laser Desorption/Ionization Time of Flight (MALDI-TOF) Mass Spectrometry

MALDI-TOF spectra were recorded on a Kratos Kompact-III TOF instrument operated in either the positive reflecton or negative linear mode. The laser output was 6 mW at a wavelength of 337 nm (3 ns pulse width, 100 mm diameter spot).

e. FAB-Mass Spectrometry

Fast atom bombardment mass spectra were collected on a Kratos MS25RFA high-resolution mass spectrometer by the McGill University Analytical Services. The matrix used was p-nitrobenzyl alcohol (NBA).

f. High Performance Liquid Chromatography

Oligomer purification was done on a Waters HPLC instrument equipped with a W600E multi-solvent delivery system and a M486 tunable absorbance detector. The instrument was interfaced to a computer running "Millenium 2000" software manager via a Waters Bus LAC/E Card. The anion exchange column, purchased from Waters, was a Protein Pak DEAE-5PW with dimensions (7.5 mm x 7.5 cm). All solvents and solutions used were filtered through a 0.22 μm filter and degassed with helium prior to use.

7.2 OLIGONUCLEOTIDE SYNTHESIS

7.2.1 Reagents for Solid Support Derivatization and Nucleoside Loading

All 5'-*O*-Dimethoxytrityl nucleosides (ribo and 2'-deoxy) were either synthesized or purchased from Dalton Chemical Laboratories (DCL) or ChemGenes Corp. (Watham, MA). Long-chain alkylamine controlled-pore glass (LCAA-CPG, 500 Å pore size, density 0.4 g/ml) was obtained from DCL and further derivatized as described below. *O*-(7-Azabenzotriazol-1-yl)-*N,N,N',N'*-tetramethyluronium hexafluorophosphate (HATU) was purchased from Fluka.

7.2.2 Derivatization of Solid Support

Long-chain alkylamine controlled-pore glass (LCAA-CPG, 500 Å pore size, density 0.4 g/ml, DCL) was derivatized according to one of two published procedures. The first method was developed by Damha *et al.*^{280,281} and it consists in activation of LCAA-CPG with 5% TCA in 1,2-dichloroethane (w/v) followed by succinylation and subsequent coupling with the appropriate 5'-protected nucleoside. The second method was recently developed by Pon *et al.*²⁸² and it utilizes HATU as the coupling catalyst. The latter is faster and yields higher loadings compared to the first method (Table 7.1).

Nucleoside loadings on the solid support from both methods were determined by spectrophotometric quantitation of the released dimethoxytrityl cation upon acid treatment at 504 nm ($\epsilon = 76000 \text{ M}^{-1} \cdot \text{cm}^{-1}$). The support was dried *in vacuo* 24 h prior to usage and was then loaded into an empty Teflon column with replaceable filters (ABI) held firmly by aluminum seals (Pierce), and consequently installed on the synthesizer.

Table 7.1: Comparison of nucleoside loadings

Nucleoside Coupled	Average Nucleoside Loading	
	Method 1 ($\mu\text{mol/g}$)	Method 2 ($\mu\text{mol/g}$)
dC	25.0	74.4
rC	33	57.6
dT	35	85
rU	33	55

The values represent the average of at least three independent trials. Errors are within $\pm 7 \mu\text{mol/g}$. The amount of CPG used in the coupling reaction is 250 mg. The various nucleosides were loaded on the solid support according to either the method of Damha (method 1) or Pon (method 2). The amount of loading was determined from spectrophotometric quantitation of the released trityl cation upon treatment with 3% trichloroacetic acid. The unit of loading is in $\mu\text{mol/g}$ (nucleoside/CPG).

7.2.3 Monomers for Automated Synthesis

5'-*O*-Dimethoxytrityl-3'-*O*-(2-cyanoethyl)-*N,N'*-diisopropylphosphoramidite monomers for normal DNA and RNA synthesis were purchased from Dalton Chemical Laboratories (Ontario, Canada). The corresponding ribonucleoside 2'-*O*-phosphoramidites (2',5'-RNA synthesis) were obtained from ChemGenes Corp. (Watham, MA). These reagents were stored at $-20 \text{ }^\circ\text{C}$ and dried *in vacuo* (under P_2O_5) for 24 h prior to use.

7.2.4 Automated Oligonucleotide Synthesis

a. Reagents for Automated Synthesis and Oligomer Deprotection

Reagents for solid-phase oligonucleotide synthesis were purchased from Applied Biosystems Inc. (Foster City, CA) or prepared as described below. DNA synthesis grade of each 1H-tetrazole (DCL), 5-ethylthio-1H-tetrazole (Aldrich) and 4,5-dicyanoimidazole (Glen Research Inc.) was used as purchased without further purification. Reagent grade acetic anhydride (Ac_2O), trichloroacetic acid (TCA), 1,2-dichloroethane (DCE), iodine, aqueous ammonia (BDH), and $\text{NEt}_3 \cdot 3\text{HF}$ (TREAT HF) were used as received. Anhydrous acetonitrile (Caledon) was pre-dried by distillation over P_2O_5 (BDH) and then refluxed over calcium hydride under nitrogen before use. Tetrahydrofuran (Fisher) was refluxed over calcium hydride, filtered and then distilled over sodium (Aldrich)/benzophenone (Aldrich) until the violet color persisted. N-methyl imidazole and collidine were dried over calcium hydride under nitrogen before use.

Phosphoramidite reagents were prepared by dissolving the desired phosphoramidite in dry, freshly distilled acetonitrile which was retrieved from the collection bulb of a continuous reflux still via a syringe fitted with a stainless steel needle and introduced to the amidite bottle (previously purged with nitrogen) through a septum-sealed stopcock.

Double distilled water (Millipore Purification System, Mississauga, Ontario) was sterilized by treatment with diethyl pyrocarbonate (Aldrich) and subsequent autoclaving (1.5 h, 121 °C, 10 atm). All glassware and apparatus used in oligonucleotide handling were sterilized in the same way in order to eliminate the possibility of any enzymes that can degrade the oligomers.

b. Oligomer Assembly

Oligomers were synthesized on Applied Biosystems (381A) or PerSeptive Biosystems Expedite™ 8909 synthesizers. All oligomers were prepared on a 1 μmol scale using LCAA-controlled pore glass (500Å pore size) derivatized with the appropriate protected nucleoside as described above. Prior to oligonucleotide assembly, the derivatized support was treated with the capping reagents [acetic anhydride/N-

methylimidazole/DMAP] using the standard capping cycle²⁸¹ in order to block undesired reactive sites and was then washed extensively with anhydrous acetonitrile.

Assembly of sequences on the 381A model was carried out according to the following procedures:

- (1) Detritylation: 3% (w/v) Trichloroacetic acid in dichloroethane delivered in 120 sec (+40 sec burst) steps. The eluate from this step was collected and the absorbance measured at 504 nm for DMT⁺ ($\epsilon = 76000 \text{ M}^{-1} \cdot \text{cm}^{-1}$) in order to determine coupling efficiencies for the individual additions.
- (2) Phosphoramidite coupling: *Coupling times*, 10 min (RNA), 120 sec (DNA); extended coupling times were used for riboguanosine phosphoramidite (15 min) and deoxyguanosine phosphoramidite (180 sec); *phosphoramidite concentration* in anhydrous acetonitrile, ribonucleoside 2'-(or 3'-) phosphoramidites (0.15-0.17 M), deoxynucleoside phosphoramidites (0.1M); *activator solution*, 1H-tetrazole (0.1 M) for 3',5'-RNA assembly, and 5-ethylthio-1H-tetrazole (0.65 M) for 2',5'-RNA assembly^{164,165}. Recently, we have started to use 4,5-dicyanoimidazole (1 M) which proved to be an even better coupling reagent for 2',5'-RNA synthesis.
- (3) Capping: Acetic anhydride/2,4,6-collidine/THF 1:1:8 (v/v/v, Cap A), and 1-methyl-1H-imidazole/THF 16:84 (v/v, Cap B), delivered in 34 sec + 90 sec wait steps.
- (4) Oxidation: 0.1 M Iodine in THF/Pyridine/H₂O 25:20:2 (v/v/v), delivered in 20 sec + 20 sec wait steps. These conditions gave coupling yields ranging from to 98% to 99.5%.

Assembly of sequences on the ExpediteTM 8909 synthesizer was carried out using standard DNA/RNA protocols supplied by the manufacturer (PerSeptive Biosystems, Inc.) with the following modifications (same reagent concentrations were used as those employed for the 381A model):

- (1) Deblocking: Trichloroacetic acid solution delivered in 10 pulses + 50 pulses during 60 sec. Qualitative evaluation of the trityl portions was done by using the Expedite trityl monitor.
- (2) Coupling: RNA (15 pulses, 759 sec), DNA (7 pulses, 96 sec); extended coupling times were used for riboguanosine phosphoramidite (900 sec) and deoxyguanosine phosphoramidite (120 sec); *phosphoramidite concentration* in anhydrous acetonitrile,

- ribonucleoside 2'-(or 3')phosphoramidites (50.6-58.0 mM), deoxynucleoside phosphoramidites (56.0–66.0 mM).
- (3) Capping: Caps A and B delivered in 13 pulses during 30 sec for RNA protocol, and in 8 pulses during 15 sec for DNA protocol.
 - (4) Oxidizing: Iodine solution delivered in 20 pulses for RNA protocol, and in 15 pulses for DNA protocol.
 - (5) Capping: Caps delivered in 7 pulses followed by acetonitrile wash -'Wash A'- delivered in 45 pulses. Under these conditions, coupling yields were 98% - 99%.

c. Deprotection of oligonucleotides

Following chain assembly, the column was extensively washed with anhydrous acetonitrile and dried with argon or helium. The CPG support was taken out of the column and split in half into two portions (each ~ 0.5 μmol). Each portion was treated with aqueous ammonia/ethanol (3:1 v/v, 1ml) either for 24 – 48 h at room temperature or for 16 h at 55 °C to effect cleavage of the oligomer from solid support and removal of the cyanoethyl phosphate protecting groups as well as cleavage of the base protecting groups. After centrifugation, the supernatant was collected and the solid support was washed with 3 x 1000 μl ethanol. The supernatant and ethanol washings were combined and cooled to –20 °C for at least 1 h before they were centrifuged to dryness. For RNA containing oligonucleotides, an additional step was performed in which the pellet obtained was treated with neat $\text{NEt}_3 \cdot 3\text{HF}$ [triethylaminetris(hydrogenfluoride)]²⁸³ (100 μl) at room temperature for 48 h²⁸⁴ in order to effect removal of the silyl protecting groups. The reaction was then quenched by addition of deionized double distilled water. The resulting solution was lyophilized to dryness under vacuum (Savant Industries Speed-Vac[®]) and subsequently purified by either PAGE or HPLC as described below.

7.3 PURIFICATION OF OLIGONUCLEOTIDES

7.3.1 Polyacrylamide Gel Electrophoresis (PAGE)

The crude oligomers were purified by denaturing polyacrylamide gel electrophoresis. The set up consisted of a vertical slab gel electrophoresis unit SE 600

(Hoefer Scientific Instruments) and a power supply unit Model 2000/200 (Bio Rad). The glass plates used were 18 x 16 cm and the spacers were either 0.75 mm or 1.5 mm thick. Electrophoresis grade Acrylamide, N,N'-methylene-bis(acrylamide) [BIS], ammonium persulfate [APS], tris(hydroxymethylamino)methane [TRIS], bromophenol blue [BPB], xylene cyanol [XC] and N,N,N',N'-tetramethylethylenediamine [TEMED] were all purchased from Bio-Rad. Other electrophoresis reagents used were: disodium ethylenediaminetetracetate dihydrate [EDTA] and formamide (BDH), boric acid (Baker Chemicals), sodium acetate (Anachemia), urea (Fisher Biotech) and sucrose (Aldrich).

Denaturing gels were used for oligomer purification and were normally made of 24%(w/v) acrylamide, 5% (w/v) BIS and 7M urea in dd water. Urea (denaturing agent) was added in order to ensure complete disruption of hydrogen bonds, thus avoiding secondary structure formation which complicates the purification process. After the gel solution was prepared, it was degassed for 20 min and then TEMED (30-35 μ l) and 10% APS (w/v, 250 μ l) were added in order to initiate polymerization (1 h). The gel thickness was either 0.75 mm or 1.5 mm (depending on the spacer used) for analytical and preparative scales respectively. The running buffer was 1 x TBE buffer (89 mM TRIS/boric acid, 2.5 mM EDTA, pH 8.3)²⁸⁵. Deionized formamide was prepared by stirring over mixed bed ion-exchange resin (Bio Rad AG 501-X8) and was then used to prepare the loading buffer (denaturing) which consisted of 8:2 formamide / 10 x TBE (v/v). The amount of material to be loaded on the gel was quantitated by UV absorbance at 260 nm and then evaporated to dryness. Typically, 0.5 ODU of oligomer was used for analytical gels while 30 ODU for preparative gels. The samples were dissolved in the loading buffer, inserted on top of the polymerized gels and run at 600-800 V for 4-5 h. A mixture of 2% BPB and XC (w/v in 10 x TBE) was loaded on a separate lane as tracking dye. The desired bands were excised from the gel, crushed and soaked in 6-8 ml dd water overnight. The solution was then filtered using Millex-HV 0.45 μ m filter unit (Millipore) and then evaporated to yield a pellet that was desalted as described below.

7.3.2 Gel Shift Mobility Assays (Native Gels)

These gels were conducted under native conditions (no denaturing agent added) in order to allow detection of nucleic acid complexes. Native gels were conducted at pH 4.2.

The running buffer was 1 x TAE (89 mM TRIS/acetate, 2.5 mM EDTA, pH = 4.2 adjusted with HCl) instead of the 1 x TBE buffer used in denaturing gels. The loading buffer is 30% sucrose in 10 x TAE buffer, pH = 4.2. The dried oligonucleotide samples (~0.35 ODU) were dissolved in 10 μ l loading buffer, heated to 95 °C and then cooled down slowly to room temperature and finally incubated at 4 °C for a week. The samples were loaded on a gel that consisted of 16 % acrylamide (no urea added). The gels were run at low voltage (100 – 150 V) for 5-6 h in the fridge (4 °C) and visualized as discussed below.

7.3.3 Visualization of Oligonucleotides

Once the fastest moving dye (BPB) has reached the lower end of the gel electrophoresis apparatus, the current was stopped, the glass plates removed and the gels covered with Saran Wrap[®]. Then they were photographed over a fluorescent plate illuminated by a UV lamp source (Mineralight, model UVG-54) using a Camera, Kodak Wratten gelatin filter (#58) and a Polaroid PolaPan 4" x 5" Instant Sheet Film (#52, medium contrast, ISO 400/21°C). The camera was set at f 4.5 with an exposure time of 16 sec.

In some instances, gels were better visualized by using Stains-All²⁸⁶ (Bio-Rad) according to manufacturer's protocol. The solution was made of (1-ethyl-2-(3-[1-ethylnaphtol(1,2-d)-thiazolin-2-ylidene]-2-methylpropenyl)naphthol(1,2-d)-thiazolium bromide). During photography, the gel was placed on a white background, the camera filter removed and the same film with different camera settings (f 11, 30 sec) was used.

7.3.4 Ion-Exchange HPLC

Oligonucleotides were purified by anion-exchange HPLC (Protein Pak DEAE-5PW column –Waters; 7.5 mm x 7.5 cm) using a linear gradient of 0 – 20% NaClO₄ or LiClO₄ in H₂O with a flow rate 1 ml/min at 55 °C over a period of 60 min. For analytical injections, 1 ODU of oligomer was dissolved in 20 μ l dd H₂O. Between 30-60 ODU of oligomer in 300 μ l H₂O were injected on a preparative scale depending on integrity of the crude synthesis. Prior to injection, the sample was heated to 70 °C for 2 min in order to denature any coiled structure. The detector was set at 260 nm for analytical injections and

at 290 nm for preparative injections in order to avoid saturation. After the peak of interest was collected in a tube, it was evaporated to dryness under vacuum. Subsequent desalting was done as described below.

7.3.5 Desalting of Oligonucleotides

The oligomers were purified from low molecular weight impurities and salts by size exclusion chromatography (SEC) using Sephadex G-25 Fine (Pharmacia Biotech). The SEC matrix was swelled in doubly distilled deionized water for at least one hour and then autoclaved (120 °C, 10 atm, 1.5 h) in an All American Electric Steam Sterilizer – Model No. 25X (Wisconsin Aluminium Co., Inc., Manitowoc, WI). The SEC matrix was loaded onto an appropriately sized sterile disposable syringe (Becton Dickinson & Co., Franklin Lakes, NJ) plugged with silanized glass wool at the tip in order to hold the matrix and prevent its loss. The SEC matrix was washed extensively with doubly distilled deionized sterilized water (~15 ml) in order to hydrate it and allow good packing of the column. Generally, the volume of SEC matrix used was 10 times the volume of the oligonucleotide sample. Typically, 30 ODU/1ml were loaded on a 10 ml SEC matrix and eluted with doubly distilled sterile deionized water. The recovery yields were very good (70-80 %). The eluate was collected in small tubes in 1-1.5 ml portions, and each was individually quantitated by UV absorbance spectrophotometry ($\lambda = 260$ nm). The tubes containing oligomers were combined to make a stock solution that was stored at -20 °C.

Oligomers purified by HPLC were also desalted using two other different methods. Those that were purified on HPLC using LiClO₄ gradient were desalted by precipitation from propanol. The collected fraction of the desired oligomer from HPLC was evaporated to dryness. The obtained pellet was dissolved in 1 ml of propanol and stored for at least 5-6 h at -20 °C. Once the oligonucleotide precipitated out of solution, the tube was centrifuged at 14000 rpm/min for 5 min. The solution was decanted and the obtained precipitate was successively washed with ethanol (2 x 100 ml) and dried. Recovery yields from this desalting procedure were fairly high (90-95%). The other method employed reverse phase chromatography in order to effect separation of the salts and low molecular weight impurities by using Sep-Pak Column connected to a 10 ml sterile disposable syringe. The column was extensively washed with water (10 ml)

followed with acetonitrile (8 ml) in order to remove any fines or particulates on the column. It was then equilibrated with water (15 ml) and the oligonucleotide sample (dissolved in water) was loaded on the column via pushing by a syringe. Subsequent washing with 10 ml of water eluted the salts followed by washing with a 1:1 acetonitrile/water (v/v, 5 ml) to afford the desalted desired oligomer which was collected in a tube, dried under vacuum and quantitated by UV- absorbance spectrophotometry at 260 nm.

An optical density unit (ODU) is defined as the number of absorbance units of a sample in 1 ml water quantitated at 260 nm. The crude yield represents the total number of optical density units obtained from synthesis. Actual yield is the total number of optical density units obtained after purification and final desalting of the oligonucleotide. The following table gives the isolated total yields of representative hairpins synthesized *via* solid-phase phosphoramidite chemistry on a 1 μ mol scale.

Table 7.2: Reported yields from solid-phase synthesis

Entry	Code	Hairpin 5'→2'/3'direction	Crude Yield (ODU)	Isolated Yield (ODU)	Purification Technique
1	DRD	ggac(UUCG)gtcc	90.2	23.8	Gel/Seph.
2	RRR	GGAC(UUCG)GUCC	61.3	18.6	HPLC/ppt.
3	<u>RRR</u>	<u>GGAC</u> (UUCG)GUCC	67.7	16.9	Gel/Seph.
4	RR <u>R</u>	GGAC(UUCG) <u>GUCC</u>	63.6	8.34	HPLC/ppt.
5	<u>RRR</u>	<u>GGAC</u> (UUCG) <u>GUCC</u>	53.5	17.2	Gel/Seph.
6	<u>RRD</u>	<u>GGAC</u> (UUCG)gtcc	63.8	19.86	Gel/Seph.
7	<u>DRD</u>	ggac(<u>UUCG</u>)gtcc	47.1	19.4	HPLC/ppt.
8	<u>DRR</u>	ggac(UUCG) <u>GUCC</u>	52.8	17.5	Gel/Seph.
9	<u>RRR</u>	<u>GGAC</u> (UUCG) <u>GUCC</u>	56.0	9.8	Gel/Seph.
10	<u>RRR</u>	<u>GGAC</u> (<u>UUCG</u>)GUCC	58.0	13.9	HPLC/ppt.
11	<u>DRR</u>	ggac(<u>UUCG</u>)GUCC	49.1	12.4	Gel/Seph.
12	<u>RRD</u>	<u>GGAC</u> (<u>UUCG</u>)gtcc	66.1	14.3	Gel/Seph.
13	<u>RRR</u>	<u>GGAC</u> (<u>UUCG</u>)GUCC	61.7	9.8	Gel/Seph.
14	<u>RRR</u>	<u>GGAC</u> (<u>UUCG</u>)GUCC	63.7	12.9	Gel/Seph.
15	R _C RR	GGAC(<u>UUCG</u>)GUCC	61.4	9.3	Gel/Seph.
16	RR _G R	GGAC(UUCG) <u>GUCC</u>	65.0	11.4	Gel/Seph.
17	TRT	tttt(UUCG)tttt	60.7	12.3	Gel/Seph.
18	RR ¹ R	GGAC(<u>UACG</u>)GUCC	53.6	4.4	HPLC/ppt.
19	RR ² R	GGAC(<u>UUUG</u>)GUCC	51.3	3.7	HPLC/ppt.
20	RR ³ R	GGAC(<u>UUUU</u>)GUCC	34.2	2.7	HPLC/ppt.
21	DR ¹ D	ggac(<u>UACG</u>)gtcc	45.8	6.1	Gel/Seph.
22	DR ² D	ggac(<u>UUUG</u>)gtcc	79.0	10.5	HPLC/ppt.
23	DR ³ D	ggac(<u>UUUU</u>)gtcc	86.0	6.6	HPLC/ppt.
24	DDD	ggac(uucg)gtcc	88.8	19.2	Gel/Seph.
25	RDR	GGAC(uucg)GUCC	80.8	5.5	Gel/Seph.
26	DTD	ggac(tttt)gtcc	90.5	22.3	Gel/Seph.
27	DUD	ggac(UUUU)gtcc	64.1	15.6	Gel/Seph.
28	DRR	ggac(UUCG)GUCC	60.0	12.4	Gel/Seph.
29	<u>DRR</u>	ggac(<u>UUCG</u>)GUCC	58.9	10.1	Gel/Seph.
30	R _C RR	GGAC(<u>UUCG</u>)GUCC	55.6	14.0	Gel/Seph.
31	RR _U R	GGAC(<u>UUCG</u>)GUCC	65.8	19.1	Gel/Seph.

All oligonucleotides were synthesized on a 1 μ mol scale. Crude yield represents the total amount of oligonucleotide obtained after solid-phase synthesis. The isolated yield represents that total amount of oligonucleotide obtained after purification/desalting. "Seph." denotes Sephadex; "ppt." refers to propanol precipitation.

7.4 CHARACTERIZATION OF OLIGONUCLEOTIDES

7.4.1 Purity Check

Prior to any studies, the purity of all purified desalted oligonucleotides was checked by using either analytical 24% denaturing acrylamide gels or analytical ion-exchange HPLC, and was found to be > 90% in most cases.

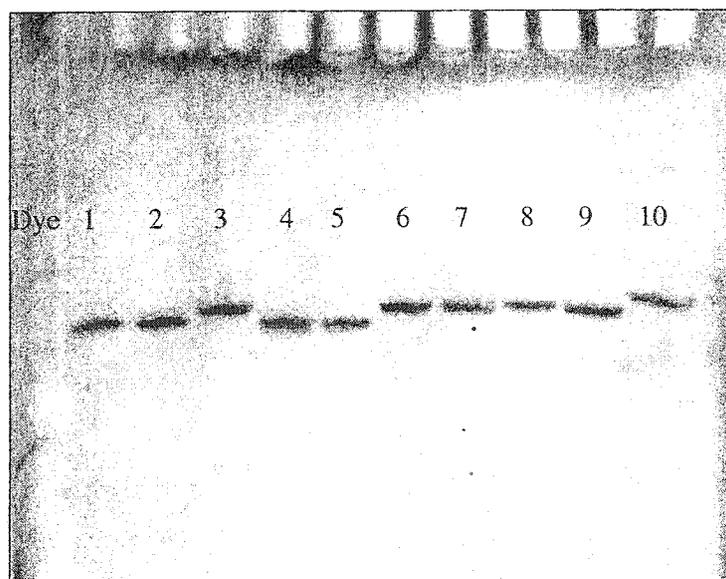


Figure 7.1: Analytical polyacrylamide (24%) denaturing gel showing representative purified oligonucleotides. Lane 1: RRR; Lane 2: RRR; Lane 3: DRR; Lane 4: RRD; Lane 5: RRR; Lane 6: RRR; Lane 7: R_CRR; Lane 8: RR_CR; Lane 9: TRT; Lane 10: C₅RC₅.

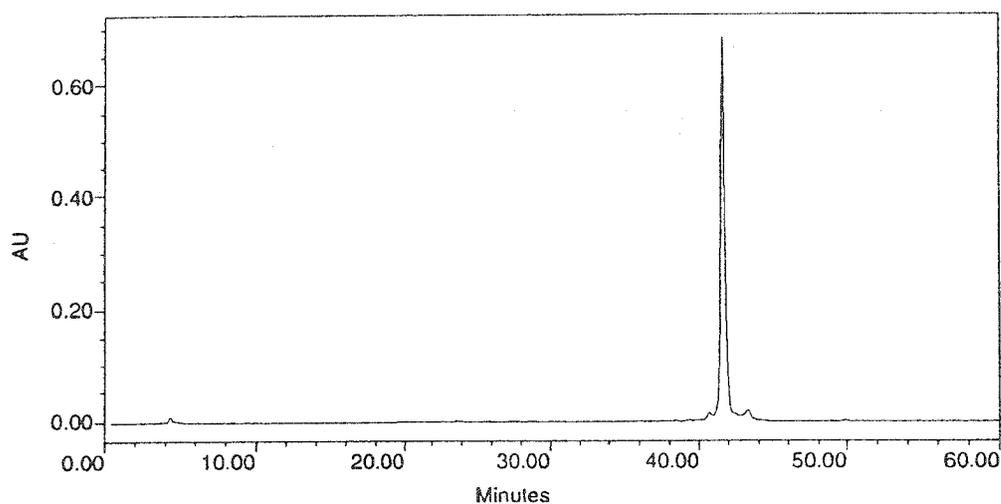


Figure 7.2: Representative anion-exchange HPLC profile for pure sample. Purification was done at 55 °C with a flow rate of 1 ml / min and a linear gradient of 0 – 23% NaClO₄ in H₂O.

7.4.2 MALDI-TOF Mass Spectrometry

The molecular weights of the purified oligonucleotides were confirmed by MALDI-TOF mass spectrometry (for representative examples, see **Table 7.3**). The matrix used was 6-aza-2-thiothymine²⁸⁷ (Aldrich) at a concentration of 10 mg/ml in 20 mM ammonium citrate (Fluka) (1:1 acetonitrile/water, v/v) buffer. The machine was run in either the positive reflectron or negative linear mode. This gave correct molecular weight signals for the desired oligonucleotides with excellent signal to noise ratios.

Typically, 20 μ M solution of purified oligomer in water was prepared, from which 2 μ l was pipetted into a tube containing 2 μ l of matrix. After this solution was vortexed, 1 μ l was pipetted, applied to a metal plate and subsequently air-dried. A brief pulse of nitrogen laser was directed to the sample in order to ionize it prior to analysis by the Kratos Kompact instrument.

7.4.3 Hybridization Properties

a. UV –Melting Studies

UV thermal denaturation curves were acquired on a Varian CARY 1 UV-VIS spectrophotometer equipped with a multi-cell holder and a Peltier temperature controller. All thermal measurements were conducted in 0.01 M Na₂HPO₄, 0.1 mM Na₂EDTA buffer, pH 7.00 \pm 0.02. Absorbance versus temperature spectra were collected at 260 nm over a range from 5 $^{\circ}$ C to 90 $^{\circ}$ C with 0.1 $^{\circ}$ C increments and a heating rate of 0.5 $^{\circ}$ C/min. Samples (\sim 4.5 μ M) were annealed by heating rapidly to 95 $^{\circ}$ C for 10-15 min, followed by cooling slowly to room temperature, and cooling (5 $^{\circ}$ C) overnight. Before data acquisition, oligonucleotides were degassed in an ultrasound bath for 3 min and consequently equilibrated at 5 $^{\circ}$ C for at least 10 min prior to the melting run. Single-strand molar extinction coefficients were calculated from those of mononucleotides and dinucleotides using the nearest-neighbor approximation method.²¹ 2',5'-RNA, and RNA:2',5'-RNA chimeras were assumed to have the same molar extinction coefficient as RNA. Single-strand concentration was determined from UV absorbance at high temperature. Percentage hypochromicity (%H) was calculated from UV absorbances of the hairpin and fully denatured species using the following equation: %H = (A_f - A₀) / A_f.

Table 7.3: MALDI-TOF spectrometry analysis and calculated extinction coefficients

Entry	Code	Calculated $\epsilon \times 10^{-4}$ ($M^{-1} \cdot cm^{-1}$)	Calculated Mass (g/mol)	Observed Mass (g/mol)	Observed Molecular Ion
1	DRD	11.448	3682.4	3682.1	(M-3H+2Li) ⁻
2	RRR	11.595	3810.2	3810.6	(M-3H+4Li) ⁺
3	<u>RRR</u>	11.595	3802.3	3802.3	(M-4H+3Li) ⁻
4	<u>RRR</u>	11.595	3792.4	3792.8	(M+Li) ⁺
5	<u>RRR</u>	11.595	3792.4	3792.7	(M+Li) ⁺
6	<u>RRD</u>	11.448	3756.5	3755.5	(M-2H+Na) ⁻
7	<u>DRD</u>	11.448	3688.3	3686.9	(M-4H+3Li) ⁻
8	<u>DRR</u>	11.453	3748.4	3748.8	(M-3H+Na+Li) ⁻
9	<u>RRR</u>	11.595	3806.4	3804.3	(M-2H+Na) ⁻
10	<u>RRR</u>	11.595	3806.4	3806.2	(M-2H+Na) ⁻
11	<u>DRR</u>	11.453	3742.5	3740.0	(M-2H+Na) ⁻
12	<u>RRD</u>	11.448	3752.3	3752.1	(M-4H+3Li) ⁻
13	<u>RRR</u>	11.595	3802.3	3801.3	(M-4H+3Li) ⁻
14	<u>RRR</u>	11.595	3784.5	3786.7	(M-H) ⁻
15	R _C RR	11.595	3802.3	3801.3	(M-4H+3Li) ⁻
16	RR _G R	11.595	3802.3	3802.3	(M-4H+3Li) ⁻
17	TRT	10.690	3650.2	3648.5	(M-4H+3Li) ⁻
18	<u>RR</u> ¹ R	11.918	3807.5	3807.6	(M-H) ⁻
19	<u>RR</u> ² R	11.778	3803.2	3803.3	(M-4H+3Li) ⁻
20	<u>RR</u> ³ R	11.709	3764.2	3763.3	(M-4H+3Li) ⁻
21	<u>DR</u> ¹ D	11.771	3693.6	3694.8	(M-H) ⁻
22	<u>DR</u> ² D	11.631	3689.3	3689.8	(M-4H+3Li) ⁻
23	<u>DR</u> ³ D	11.562	3650.3	3648.5	(M-4H+3Li) ⁻
24	DDD	11.448	3618.4	3621.8	(M-3H+2Li) ⁻
25	RDR	11.595	3736.4	3736.6	(M-3H+2Na) ⁻
26	DTD	11.124	3624.6	3626.5	(M-H) ⁻
27	DUD	11.562	3632.5	3632.2	(M-H) ⁻
28	DRR	11.453	3720.5	3720.6	(M-H) ⁻
29	<u>DRR</u>	11.453	3720.5	3720.6	(M-H) ⁻
30	R _C RR	11.595	3784.5	3783.5	(M-H) ⁻
31	RR _U R	11.595	3784.5	3785.4	(M-H) ⁻

Molar Extinction coefficients (ϵ) for the various oligonucleotides were calculated using the nearest-neighbor approximation method.²¹ 2',5'-RNA, and RNA:2',5'-RNA chimeras were assumed to have the same molar extinction coefficient as RNA.

b. Concentration Studies (van't Hoff Plots)

In order to study the behavior of self-complementary oligonucleotide species in solution (unimolecular vs. bimolecular), van't Hoff plots ($\ln[\text{conc}]$ versus $1/T_m$) were constructed. Consequently, the change in melting temperature was monitored as oligomer concentration was varied over at least a 30-fold range. Quartz cells (Hellma) with various path lengths (1 mm – 10 mm) and volume capacities were used.

c. Calculation of Thermodynamic Parameters and Error Analysis

Melting temperatures and thermodynamic parameters for hairpin formation were computed using Cary WinUV version 2 software (Varian Ltd.). Absorbance versus temperature profiles were fit to a two-state (all-or-none) model²¹ and the software was adapted for a unimolecular transition. Sloping baselines was achieved by constructing linear least squares lines for associated and dissociated parts and extrapolating to both ends of the melting curve. Consequently, a plot of the fraction of single strands in the hairpin state (α) versus temperature was constructed and used to calculate the T_m value by interpolating to $\alpha = 0.5$. The transition equilibrium constant, K , was calculated at various temperatures using the equation:

$$K = \alpha / (1 - \alpha) \quad (1)$$

A van't Hoff plot of $\ln K$ versus $1/T$ was generated, and the slope and intercept of the calculated line yielded the enthalpy (ΔH°) and entropy (ΔS°) of hairpin formation, respectively, from which Gibbs free energy at 37 °C (ΔG°_{37}) was calculated according to the equation:

$$\ln K = -[\Delta H^\circ/R]1/T + \Delta S^\circ/R \quad (2)$$

where R is the gas constant, and T the temperature in Kelvin units. The data obtained represent the average of at least five independent measurements for each hairpin. To ensure that only hairpin species formed in solution, the samples were melted over at least 30-fold nucleic acid concentration with no detectable change in melting temperature, thus confirming an *intramolecular* transition.

The calculated error limits in the thermodynamic data represent standard deviations with all values weighted equally.¹⁶⁹ The variations in T_m values for all hairpin

structures are within $\pm 1^\circ\text{C}$. The variations in ΔH° and ΔS° are within $\pm 7.5\%$ while those for ΔG°_{37} are within ± 0.2 kcal/mol.

d. Circular Dichroism Studies

CD spectra were measured on a JASCO J710 spectropolarimeter at ambient temperature (unless otherwise indicated) using fused quartz cells (Hellma, 165-QS). The temperature was controlled by an external constant temperature NESLAB RTE-111 circulating bath. Each spectrum represented the average of five 220-340 nm scans obtained at a rate of 0.5 nm/min (bandwidth: 1nm; sampling wavelength: 0.2 nm). The buffer used was 0.01 M Na_2HPO_4 , 0.1 mM Na_2EDTA , pH 7.00 ± 0.02 . The data were processed on a PC computer using WindowsTM based software supplied by the manufacturer (JASCO, Inc.). The obtained CD spectra were normalized by subtraction of the background scan with buffer. Taking the known oligonucleotide concentration into account, the normalized spectra were converted to molar ellipticities.

7.5 HAIRPIN STRUCTURE DETERMINATION

7.5.1 NMR Samples: Synthesis, Purification and Preparation

The hairpin samples under NMR study (DRD, RRR, and DRD) were synthesized as described above (see oligonucleotide synthesis) using an Applied Biosystems (381A) synthesizer on a 4 μmol scale and utilizing LCAA-controlled pore glass (500 \AA) as solid support. The monomer coupling times and concentrations were the same as those described above. After deprotection and evaporation, the oligomers were purified by preparative anion-exchange HPLC (Protein Pak DEAE-5PW column - Waters, 22.5 mm x 150 mm) using a linear gradient of 0 – 20 % LiClO_4 in H_2O (1M) with a flow rate 5 ml/min at 55 $^\circ\text{C}$. The oligonucleotides were then desalted by using reversed-phase chromatography on a Sep-Pak cartridge²⁸¹ and evaporated to dryness to yield pure DRD, RRR and DRD (> 95% purity). **Table 7.4** presents the isolated yields along with % purity and HPLC retention times.

The samples were then dissolved in 0.3 ml of 100% D_2O or 9:1 $\text{H}_2\text{O}/\text{D}_2\text{O}$, v/v (for imino-proton spectra) to a final concentration of ~ 1.5 mM. The solutions contained 0.1 mM EDTA sodium salt, and the final pH was adjusted to 7.0 with 100 mM NaOH.

7.5.2 NMR Spectroscopy

These experiments (along with molecular modeling) were run in collaboration with Prof. Kalle Gehring and Dr. Alexei Denisov from the McGill Biochemistry Department. NOESY experiments were performed in D₂O at 15 °C for all three hairpins, and additionally at 25 °C for hairpin DRD using mixing times t_m of 70, 200 and 300 ms. The volumes of crosspeaks of NOESY spectra were calculated using XWINNMR (Bruker). NOESY experiments in 9:1 H₂O/D₂O (v/v) were performed at 5 °C with a mixing time of 180 ms. DQF-COSY spectra were collected with phosphorus decoupling (final data size of 8K x 2K points). Proton MLEV-17 TOCSY experiments were performed with a mixing time of 84 ms. H,C-correlation HMQC spectra were recorded using GARP heteronuclear decoupling ($^1J_{CH} = 180$ Hz). Inverse H,P-HetCOSY spectra were collected with final spectral sizes of 2K x 1K data points.

Table 7.4: Reported yields for solid-phase synthesis of NMR hairpins

Entry	Code	Scale of Synthesis	Crude Yield (ODU)	Isolated Yield (ODU)	% Recovery	Purity
1	DRD	4 μ mol	350	130	37	96%
2	RRR	4 μ mol	270	130	48	95%
3	DRD	2 μ mol	170	60	35	98%

All NMR samples were purified by ion-exchange HPLC using a linear gradient of 0-20 % LiClO₄ (1 M) for 1 h at 60 °C. The retention times for all desired peaks corresponding to full-length product were *ca.* 52-53 min. The obtained solutions were dried *in vacuo* and then desalted by reversed-phase chromatography using a 1:1 acetonitrile/water solvent system on Sep-Pak cartridges. Crude yield represents the total amount of crude deprotected oligonucleotides obtained from solid-phase synthesis. Isolated yield is the final amount of pure oligonucleotide after HPLC purification and desalting. % Recovery refers to the % oligonucleotide amount recovered from the total crude synthesis after HPLC purification and desalting, and was calculated according to the equation:

$$\% \text{Recovery} = \{ [\text{Isolated Yield}] / [\text{Crude Yield}] \} \times 100.$$

7.5.3 Structural Modeling

The starting coordinates of all hairpins under study were generated using Sybyl 6.5 software (Tripos Inc.) from both *A*- and *B*-type DNA structures. The X-PLOR 3.843 package²⁸⁸ with standard nucleic acid force field was used for hairpin molecular modeling.

At the initial stage, 100 starting structures with 'randomized' loop (half with canonical *A*-type and half with *B*-type hairpin stem) were generated by molecular dynamics without experimental constraints. Subsequent stages of simulated annealing with NOE-distance and torsion angle constraints were similar to those described previously.^{54,57} Global fold was reached by restrained simulated annealing at 5000 K, and the 20 most energetically favored structures with minimal number of structural violations were selected at this stage. Subsequently, the following gentle refinement was accomplished from these structures by molecular dynamics simulation of 12 ps (5 ps at 1000 K, 4 ps of cooling to 300 K and then 3 ps at 300 K). During the last stage, NOE-distance and hydrogen bond force constants were gradually built up to final values of 30-40 kcal/(mol Å²) and the backbone torsion angle constants to 60 kcal/(mol rad²). A distance-dependent dielectric constant was used to mimic the solvent. Final ensemble of 10 best structures (including structures which converged from both *A*- and *B*-type starting models) have been deposited into the Protein Data Bank (PDB) under ID codes "1ME0" for hairpin DRD and "1ME1" for hairpin RRR.

Distance restraints were derived from NOESY spectra at different mixing times by crosspeak volume integration, using the r^{-6} distance relationship³⁰ and average crosspeak volume values for calibration of H5-H6 in cytidines ($r=2.46$ Å) and Me-H6 in thymidines ($r=2.70$ Å). The distance constraints were given with 10% of lower and 15% of upper bounds. Sugar pucker was determined by the PSEUROT 3B program²⁸⁹ from vicinal coupling constants. The five ν_0 - ν_4 torsion angles for hairpin sugars were constrained (with $\pm 10^\circ$ bounds) according to *south* or *north* sugar conformations determined from J-couplings. Backbone torsion angle constraints were set in hairpins DRD and RRR. The β torsion angles were constrained using the information about H5'/H5''-P and H4'-P crosspeaks in H,P-HetCOSY spectra. The β angles were found to be in the *trans* conformation ($180 \pm 60^\circ$) as determined by symmetry and the relatively low

intensity of the H5'/H5''-P crosspeaks as well as detectable $^4J_{H4'-P}$ W-pathway coupling constants.^{1,30,290} The γ angles were constrained ($60\pm 40^\circ$) using the sums of $J_{H4'H5'}$ and $J_{H4'H5''}$ which were available from phosphorus decoupled DQF-COSY spectra, and the NOE H1'/H6/H8-H4' crosspeak linewidths.²⁹⁰ The ϵ angles have been estimated from the vicinal $^3J_{H3'-P}$ or $^3J_{H2'-P}$ coupling constants. These coupling constants lie in the range of 7-9 Hz for all RNA stem and loop nucleotides, and ϵ angles (or C3'-C2-'O2'-P torsions for 2'5'-RNA) were constrained to $240\pm 50^\circ$ from the Karplus equation.²⁹⁰ In contrast, the small $^3J_{H3'-P}$ values observed for the DNA stem (3-5 Hz) suggest a ϵ value of $170\pm 50^\circ$. Finally, glycosidic angles were not constrained in structure calculations but were fixed indirectly in *anti*-conformation by *intranucleotide* aromatic-sugar distance constraints.

7.6 HIV-1 RT BIOLOGICAL STUDIES

7.6.1 Materials

E.coli RNase H, gamma-ATP and ^{32}P labeling kit were all purchased from Amersham Pharmacia. The RNA and DNA strands utilized in RNase H inhibition assays were synthesized using standard phosphoramidite chemistry protocols (see oligomer assembly – **Section 7.2.4b**). Similarly, the RNA and DNA templates as well as the primer (from PBS region of HIV-1) utilized in the polymerase assays were also synthesized in the lab via the same method (see assay for sequence). Rabbit reticulocytes lysate was purchased from Promega. Nuclease PI from *Penicillium citrinum* was purchased from Amersham Pharmacia.

The experiments described herein were performed in collaboration with Dr. Kyung-Lyum Min from our research group.

7.6.2 Preparation of HIV-1 RT

The p66- and p51-kDa subunits of HIV-1 RT were cloned into a pBAD/HisB prokaryotic expression vector (Invitrogen) between the XhoI and HindIII sites. RT p66/p51 heterodimers and p51/p51 homodimers were purified as described in the literature.²⁹¹

7.6.3 5'-End [³²P]-Labeling Assay of Oligonucleotides

100 pmol of oligonucleotide was dissolved in 5 μ l of 10 x T4 PNK Buffer (0.5 M Tris-HCl, pH 7.6, 100 mM MgCl₂, 100 mM 2-mercaptoethanol) to which 10 pmol of Gamma-ATP and 9 units (3-6 μ l, diluted with 1 x T4 PNK buffer) of T4 Polynucleotide Kinase (PNK) were added. The assay volume was adjusted to 50 μ l in total volume with dd water. After the contents were mixed well and centrifuged briefly, they were incubated at 37 °C for 1.5 h. The reaction was terminated by heating to 65 °C for 5 min followed by evaporation (under vacuum) until dryness.

Just prior to loading onto a gel, the dried sample was dissolved in 16 μ l of loading buffer and subsequently heated to 100 °C for about 10 min to effect denaturation of any secondary structure. The gel was run at 2000 V for 3 h. After removal of the glass plates, the major and slowest moving band was excised from the gel and crushed and soaked in 0.5 ml of dd water overnight at 37 °C. The sample was then desalted by loading on a NapTM-5 column containing Sephadex[®] G-25 Medium (0.9 x 2.8 cm, Pharmacia Biotech) and elution with 1 ml dd water. The obtained solution was lyophilized to dryness and subsequently quantitated for radioactive content by using a Bioscan/QC 2000 counter (Amersham).

7.6.4 HIV-1 Reverse Transcriptase RNase H Inhibition Assay

This assay measured the ability of various hairpin oligomers to inhibit the RNase H-mediated degradation of a 5'-[³²P]-labeled RNA:DNA 18 base pair substrate. The sequence of the 5'-[³²P]-RNA used is: 5'-GAU CUG AGC CUG GGA GCU-3', and was prepared by the transfer of ³²P from [γ -³²P] ATP in a reaction catalyzed by bacteriophage T4 polynucleotide kinase. This 5'-[³²P]-RNA oligonucleotide was annealed to a complementary unlabeled DNA strand of the sequence: 5'-AGC TCC CAG GCT CAG ATC-3' to form the [³²P]-RNA/DNA hybrid duplex substrate. In separate microtubes, aliquots of 10 μ l of 50 mM Tris/HCl (pH 7.8, 37 °C), containing 60 mM KCl and 2.5 mM MgCl₂, 1.5 nM p51/p66 RT were mixed with variable amounts of hairpins and incubated at 37 °C for 15 min. The reactions were initiated by addition of [³²P]-RNA/DNA hybrid duplex substrate (final concentration 50 nM), and the assay tubes were then incubated at

37 °C. After 15 min, the reactions were stopped by the addition of an equal volume of denaturing gel loading buffer (98% deionized formamide, 10 mM EDTA, 1 mg/ml bromophenol blue and 1 mg/ml xylene cyanol). The reaction products were denatured by heating at 100 °C for 5 min, and were then loaded onto a 16% polyacrylamide sequencing gel containing 7 M urea and the products resolved by electrophoresis. The resolved reaction products were then visualized by autoradiography and the extent of RNase H inhibition was determined quantitatively by densitometric analysis as judged from the disappearance of the full-length RNA substrate and/or the appearance of the smaller degradation products. In this experiment, the bands corresponding to the undegraded 18 nucleotide 5'-[³²P]-RNA were quantified by densitometry, using the software UN-SCAN-IT (Silk Scientific, Orem, UT). The IC₅₀ values for hairpin inhibition of HIV-1 RT associated RNase H activity were calculated from plots of the residual undegraded 5'-[³²P]-RNA *versus* oligomer hairpin concentration.

7.6.5 *E.coli* and Human RNase H Inhibition Assays

The hairpins were tested for their ability to inhibit either *E.coli* or Human (mammalian) RNase H (II). The inhibition assays employed the same conditions as those used in the HIV-1 reverse transcriptase assay, except that *E.coli* or Human RNase H (II) enzymes were used instead of the HIV-1 reverse transcriptase.

7.6.6 HIV-1 Reverse Transcriptase RNA-Dependent DNA Polymerase Assay

This assay measured the ability of various hairpins to inhibit the HIV-1 RT-mediated synthesis of DNA from a 5'-[³²P]-labeled DNA primer:RNA template duplex substrate in the presence of various dNTPs. The template and primer oligomers were synthesized as described above using standard phosphoramidite solid phase protocols. The sequence of the RNA template employed in the assay is: 5'- AUC UCU AGC AGA GGC GCC CGA ACA GGG ACA -3'. The 5'-end labeled synthetic primer, 5'- TGT CCC TGT TCG GGC GCC-3', was annealed to a 3-fold molar excess of unlabeled RNA template. The reaction was performed in 50 mM Tris/HCl (pH 7.8), 60 mM KCl, and 2.5 mM MgCl₂. The enzyme was preincubated with hairpin oligomer for 20 min at room temperature prior to initiating the polymerization reaction with the template-primer and

dNTPs (each at a final concentration of 200 nM). All reactions, each at a final volume of 10 ml, were incubated at 37 °C for 15 min. The reactions were stopped by addition of an equal volume of sequencing dye (98% deionized formamide, 10 mM EDTA, 1 mg/ml bromophenol blue and 1 mg/ml xylene cyanol), followed by heating to 100°C for 5 min before loading onto a 16 % polyacrylamide denaturing gels and subsequent densitometric analysis using the software UN-SCAN-IT (Silk Scientific, Orem, UT).

7.6.7 HIV-1 Reverse Transcriptase DNA-Dependent DNA Polymerase Assay

This experiment was designed to test the ability of various hairpins to inhibit the HIV-1 RT-mediated synthesis of DNA from a 5'-[³²P]-labeled DNA primer:DNA template duplex substrate in the presence of various dNTPs. The assay employed the same conditions used above for the RNA-dependent DNA polymerase assay. The only difference is that a DNA template was used instead of the RNA template. The sequence of the DNA template is: 5'- ATC TCT AGC AGA GGC GCC CGA ACA GGG ACA-3'. This template was synthesized in our lab as described above (**Section 7.2.4**) using standard phosphoramidite solid-phase protocols.

7.6.8 Gel-Shift Binding Assay

The binding of various hairpins to HIV-1 RT was measured using electrophoretic-mobility shift assay (native conditions). A 5'-[³²P]-labeled hairpin was heated to 95 °C and then left to anneal slowly at room temperature for one hour. The binding reaction assays consisted of 50 mM Tris/HCl, pH 7.8, 50 mM KCl, 10 % glycerol, various amounts of HIV-1 RT and 0.3 pmol of the labeled annealed hairpin in a final volume of 25 μ l. The reaction mixtures were incubated for 20 min at room temperature, and then analyzed on a 6%-polyacrylamide (non-denaturing) gel in 25 mM Tris/glycine/1 mM EDTA, pH 8.0 at 160 V for about 1 hr. The gel autoradiograph was exposed to an X-ray film. Complex formation between the [³²P]-labeled hairpin and HIV-1 RT was detected (with increasing enzyme concentration) from the apparent retarded electrophoretic mobility of the hairpin as a result of its association with the enzyme.

7.6.9 UV Cross-Linking Study

HIV-1 RT was incubated with ³²P-labeled hairpin in 50 mM Tris (pH 7.8), 50 mM KCl, and 5 mM Mg²⁺ for 20 min at room temperature. Mixtures were then irradiated with a UV transilluminator (254 nm) on ice for 15 min and analyzed by electrophoresis on a 12.5 % SDS-polyacrylamide gel (denaturing conditions).

7.6.10 Stability of Hairpins in Biological Media

a. Stability in Blood Serum

To measure the degradation of various inhibitors by intracellular nucleases, the individual hairpin aptamers were 5'-³²P-end labeled. They were then dissolved in 60 mM Tris (pH 7.8), 60 mM KCl, 5 mM MgCl₂, and 5 mM K₂HPO₄ buffer. After the addition of rabbit reticulocytes lysate (Promega), the reaction mixture was incubated for 18 h at 37 °C. The degradation assay was stopped by the addition of an equal volume of loading dye [98% deionized formamide, 10 mM EDTA, 1 mg/ml bromophenol blue and 1 mg/ml xylene cyanol] followed by heating to 95 °C. They were then loaded onto 16 % (w/v) polyacrylamide denaturing gels (7 M urea) and then exposed on an X-ray film. The individual lanes were quantitated using the software UN-SCAN-IT by measuring the amount of remaining hairpin after 18 h.

b. Stability Towards Nuclease P1

The stability of various hairpin aptamers towards Nuclease P1 (*Penicillium citrinum*) was assessed either by UV spectroscopy or gel electrophoresis. In both techniques, the obtained Nuclease P1 enzyme was diluted with 1 ml of 30 mM sodium acetate buffer, pH 5.3.

Enzymatic degradation studies were conducted on a Varian CARY 1 UV-VIS spectrophotometer equipped with a multi-cell holder and a Peltier temperature controller. Hairpin samples were dissolved in 30 mM NaOAc buffer [pH = 5.30 ± 0.02] to a final concentration of 6 μM [750 μl in total volume], and then annealed by heating rapidly to 95 °C for 10-15 min, followed by cooling slowly to room temperature. Before data acquisition, oligonucleotides were degassed in an ultrasound bath for 3 min and

consequently equilibrated at 37 °C for at least 10 min prior to the enzymatic run. The reaction was initiated by the addition of Nuclease P1 (0.001 Units), and the UV absorbance of the hairpin was monitored with time over a range of 180 min at 260 nm with 0.5 min increments and an average time of 0.1 sec [signal band width = 1nm]. All runs were corrected by taking into account the time elapsed from the moment the enzyme was added until the moment the run was started. The nuclease resistance of various hairpins was analyzed by comparing their half-life time ($t_{1/2}$) values defined as the time at which 50% degradation of the hairpin occurs. $t_{1/2}$ was calculated by interpolating to $A = 0.5$ in the absorbance versus time plot.

In gel electrophoresis, the individual hairpin aptamers were 5'-end-labeled with Gamma-ATP. They were then dissolved in 30 mM NaOAc buffer [pH = 5.30 ± 0.02] to a final concentration of 6 μ M [750 μ l in total volume]. The reaction was initiated by addition of Nuclease P1 (0.0001 Units) and was incubated at 37 °C for 150 min. Aliquots (8 μ l) were taken out at various time intervals, and the reaction was stopped immediately by instant addition of an equal volume of loading dye [98% deionized formamide, 10 mM EDTA, 1 mg/ml bromophenol blue and 1 mg/ml xylene cyanol] followed by heating to 95 °C. The aliquots were loaded onto a 16 % (w/v) polyacrylamide denaturing gel (7 M urea) and then exposed on an X-ray film. The individual lanes were quantitated using the software UN-SCAN-IT by measuring the amount of remaining versus degraded hairpin.

A sequencing ladder, created via alkaline hydrolysis, was run as a marker on the gel. The 5'-³²P-labeled hairpin was dissolved in sodium carbonate buffer, pH 9.0 and incubated for 35 min at 95 °C. The reaction was stopped by the addition of an equal volume of dye loading buffer.

7.7 RNASE III BIOLOGICAL STUDIES

7.7.1 Materials

The experiments described in chapter IV were performed in collaboration with Bruno Lamontagne and Prof. Sherif AbouElela of the Département de Microbiologie et Infectiologie at Université Sherbrooke. Recombinant Rnt1p was produced in bacteria and purified as described previously.²¹² All RNA substrates used in the enzymatic assays

were 5'-end labeled with T4 polynucleotide kinase (New England Biolabs) using [γ - 32 P] ATP and subsequently purified by gel electrophoresis (see **Section 7.6.3**).

7.7.3 Synthesis of Rnt1p Hairpin Substrates

All hairpin substrates used in this study were synthesized on solid-phase according to the phosphate trimer method. The crude and pure synthesis yields are reported below.

Table 7.5: Synthesis yields obtained from solid-phase synthesis of Rnt1p substrates

Code	Designation	Synthesis Scale (μ mol)	Obtained Crude (ODU)	Isolated Pure (ODU)	Purification Technique
4.1	R ₉ L R ₉	0.2	19.4	0.4	Gel/Seph.
4.2	R ₉ L D ₉	1.0	30.2	9.2	Gel/Seph.
4.3	D ₉ L D ₉	1.0	47.9	11.9	Gel/Seph.
4.4	R ₉ L R ₉	1.0	100	17.9	Gel/Seph.
4.10	R ₁₉ L D ₁₉	1.0	46.5	11.4	HPLC/ppt.
4.11	D ₁₉ L R ₁₉	0.5	65.7	0.5	Gel/Seph.
4.12	D ₁₉ L D ₁₉	1.0	65.8	9.2	HPLC/ppt.
4.13	R ₁₉ (dU) L R ₁₉	0.5	65.2	0.4	Gel/Seph.
4.14	R ₁₉ (2'-OMe) L R ₁₉	0.5	42.6	0.4	Gel/Seph.
4.15	D ₁₉ LR ₅ D ₁₄	0.5	47.8	1.4	Gel/Seph.
4.16	D ₁₄ R ₅ LR ₅ D ₁₄	0.5	44.3	1.0	Gel/Seph.
4.18	^{D/R} U5-Like	0.5	52.6	6.9	Gel/Seph.
4.19	R ₁₉ (2'-rU) L R ₁₉	0.5	46.6	0.82	Gel/Seph.

The crude represents the total amount of oligonucleotide obtained after solid-phase synthesis. The isolated yield represents that total amount of oligonucleotide obtained after purification and desalting. "Seph." refers to Sephadex while "ppt." refers to propanol precipitation.

7.7.3 Gel-Mobility Shift Assay

These assays were done in order to detect complex formation between Rnt1p and various substrates in the absence of divalent metal ions. The experiments were performed using 2 fmol of radiolabelled RNA in 20 μ l buffer [20% (v/v) glycerol, 30 mM Tris (pH 7.5), 150 mM KCl, 5 mM spermidine, 0.1 mM Dithiothreitol (DTT) and 0.1 mM EDTA]

for 10 min on ice. For each experiment, various amounts of protein were used ranging from 0.25 to 12 μM . The reactions were fractionated on 4% (w/v) non-denaturing polyacrylamide gels (1/80 bisacrylamide / acrylamide) at 0.5 V/cm² and 4 °C.

7.7.4 Kinetic Parameters Determination

All calculations used in this study were done using Graph Pad Prism 3.0 program (GraphPad Software, CA).

7.7.5 Enzymatic Cleavage Assay

For the in vitro cleavage, 0.4 – 2.8 pmol of radiolabeled hairpin substrate was incubated in the presence of 0.2 pmol of Rnt1p for 20 min at 30°C in 20 μl of reaction buffer [30 mM Tris (pH 7.5), 5 mM spermidine, 20 mM MgCl₂, 0.1 mM DTT, 0.1 mM EDTA pH 7.5]. The reactions were stopped by addition of stop buffer (20 mM EDTA pH 7.5, 0.1% bromophenol blue, and 0.05% xylene cyanol in formamide) and directly loaded on a 20% denaturing polyacrylamide gel. The cleavage rate was quantified using Instant Imager (Packard, Meriden, CT).

References

- (1) Saenger, W. In *Principles of Nucleic Acids Structure*; Cantor, C.R. ed.; Springer-Verlag: New York, NY, USA, 1984.
- (2) Blackburn, G. M.; Gait, M. J. In *Nucleic Acids in Chemistry and Biology*; Blackburn, G. M., Gait, M. J., Eds.; Oxford University Press: New York, 1990.
- (3) Watson, J. D.; Crick, F. *Nature* **1953**, *171*, 737 and 964.
- (4) Ratmeyer, L.; Vinayak, R.; Zhong, Y. Y.; Zon, G.; Wilson, W. D. *Biochemistry* **1994**, *33*, 5298-5304.
- (5) Wang, A. C.; Kim, S. G.; Flynn, P. F.; Chou, S.-H.; Orban, J.; Reid, B. R. *Biochemistry* **1992**, *31*, 3940-3946.
- (6) Hung, S.-H.; Yu, Q.; Gray, D. M.; Ratliff, R. L. *Nucleic Acids Res.* **1994**, *22*, 4326-4334.
- (7) Salazar, M.; Champoux, J. J.; Reid, B. R. *Biochemistry* **1993**, *32*, 739-744.
- (8) Salazar, M.; Federoff, O. Y.; Reid, B. R. *Biochemistry* **1996**, *35*, 8126-8135.
- (9) Merrifield, R. B. *J. Am. Chem. Soc.* **1963**, *85*, 2149-2154.
- (10) Letsinger, R. L.; Kornet, M. J.; Mahadevan, V.; Jerina, D. M. *J. Am. Chem. Soc.* **1964**, *86*, 5163-5165.
- (11) Letsinger, R. L.; Mahadevan, V. *J. Am. Chem. Soc.* **1965**, *87*, 3526-3227.
- (12) Letsinger, R. L.; Finnan, J. L.; Heavner, G. A.; Lunsford, W. B. *J. Am. Chem. Soc.* **1975**, *97*, 3278.
- (13) Letsinger, R. L.; Lunsford, W. B. *J. Am. Chem. Soc.* **1976**, *98*, 3655.
- (14) Beaucage, S. L.; Caruthers, M. H. *Tetrahedron Lett.* **1981**, *22*, 1859-1862.
- (15) Caruthers, M. H.; Barone, A. D.; Beaucage, S. L.; Dodds, D. R.; Fisher, E. E.; McBride, L. J.; Matteucci, M. D.; Stabinsky, Z.; Tang, J. Y. *Methods Enzymol.* **1987**, *157*, 287.
- (16) Gait, M. J. In *Oligonucleotide Synthesis: A practical approach*; Gait, M. J., Ed.; IRL Press: Oxford, 1984; pp 1-22.
- (17) Ogilvie, K. K.; Thompson, E. A.; Quilliam, M. A.; Westmore, J. B. *Tetrahedron Lett.* **1974**, *15*, 2865.
- (18) Wu, T.; Ogilvie, K. K.; Pon, R. T. *Nucleic Acids Research* **1989**, *17*, 3501.
- (19) Dahl, O.; Nielsen, J.; Dahl, B. H. *Nucleic Acids.* **1987**, *15*, 1729-1742.

- (20) Berner, S.; Muehlegger, K.; Seliger, H. *Nucleic Acids Res.* **1989**, *17*, 853-865.
- (21) Puglisi, J. D.; Tinoco, I., Jr. *Methods Enzymol.* **1989**, *180*, 304-332.
- (22) Breslauer, K. J. In *Methods in Molecular Biology: Protocols for Oligonucleotide Conjugates*; Agrawal, S., Ed.; Humana Press Inc.: Totawa, NJ, 1994; Vol. 26, pp 347-372.
- (23) Plum, G. E.; Breslauer, K. J.; Roberts, R. In *Comprehensive Natural Product Chemistry*; pp 15-53.
- (24) Jaeger, J. A.; SantaLucia, J., Jr.; Tinoco, I., Jr. *Annu. Rev. Biochem.* **1993**, *62*, 255-287.
- (25) Tinoco, I., Jr. *J. Am. Chem. Soc.* **1960**, *82*, 4785-4790.
- (26) Petersheim, M.; Turner, D. H. *Biochemistry* **1983**, *22*, 256-263.
- (27) Chastain, M.; Tinoco, I., Jr. *Prog. Nucleic Acids Res. Mol. Biol.* **1991**, *41*, 131-177.
- (28) Gray, D. M.; Liu, J.-J.; Ratliff, R. L.; Allen, F. S. *Biopolymers* **1981**, *20*, 1337-1382.
- (29) Riesner, D.; Henco, K.; Steger, G. *Adv. Electrophor.* **1991**, *4*, 151-250.
- (30) Wuthrich, K. *NMR of Proteins and Nucleic Acids*; John Wiley & Sons: New York, 1986.
- (31) Leroy, J. L.; Bolo, N.; Figueroa, N.; Plateau, P.; Gueron, M. *J. Biomol. Struct. Dyn.* **1985**, *2*, 915-939.
- (32) Varani, G.; Tinoco, I., Jr. *Q. Rev. Biophys.* **1991**, *24*, 479-532.
- (33) Varani, G.; Wimberly, B.; Tinoco, I., Jr. *Biochemistry* **1989**, *28*, 7760-7772.
- (34) Varani, G. *Annu. Rev. Biophys. Biomol. Struct.* **1995**, *24*, 379-404.
- (35) Pleij, C. W. *Trends Biotechnol.* **1990**, *15*, 143-147.
- (36) Michel, F.; Westhof, E. *J. Mol. Biol.* **1990**, *216*, 585.
- (37) Zwieb, C. *J. Biol. Chem.* **1992**, *267*, 15650.
- (38) Wimberly, B.; Varani, G.; Tinoco, I., Jr. *Biochemistry* **1993**, *32*, 1078-1087.
- (39) Gautheret, D.; Konings, D.; Gutell, R. R. *J. Mol. Biol.* **1994**, *242*, 1.
- (40) Gautheret, D.; Konings, D.; Gutell, R. R. *RNA* **1995**, *1*, 807.
- (41) Gregory, S. T.; O'Connor, M.; Dahlberg, A. E. In; pp 189-204.
- (42) Woese, C. R.; Winker, S.; Gutell, R. R. *Proc. Natl. Acad. Sci. USA.* **1990**, *87*, 8467-8471.
- (43) Costa, M.; Michel, F. *EMBO J.* **1995**, *14*, 1276.

- (44) Cate, J. H.; Gooding, A. R.; Podell, E.; Zhou, K.; Golden, B. L.; Kundrot, C. E.; Cech, T. R.; Doudna, J. A. *Science* **1996**, *273*, 1678.
- (45) Jucker, F. M.; Pardi, A. *Biochemistry* **1995**, *34*, 14416-14427.
- (46) Jaeger, L.; Michel, F.; Westhof, E. *J. Mol. Biol.* **1994**, *236*, 1271-1276.
- (47) Murphy, F. L.; Cech, T. R. *J. Mol. Biol.* **1994**, *236*, 362-364.
- (48) Heus, H. A.; Pardi, A. *Science* **1991**, *253*, 191.
- (49) Jucker, F. M.; Heus, H. A.; Yip, P. F.; Moors, E. H. M.; Pardi, A. *J. Mol. Biol.* **1996**, *264*, 968.
- (50) Jiang, F.; Kumar, R. A.; Jones, R. A.; Patel, D. J. *Nature* **1996**, *382*, 183.
- (51) Dieckmann, T.; Suzuki, E.; Nakamura, G. K.; Feigon, J. *RNA* **1996**, *2*, 628.
- (52) Wolters, J. *Nucleic Acids Res.* **1992**, *20*, 1843-1850.
- (53) Cheong, C.; Varani, G.; Tinoco, I., Jr. *Nature* **1990**, *346*, 680-682.
- (54) Varani, G.; Cheong, C.; Tinoco, I., Jr. *Biochemistry* **1991**, *30*, 3280-3289.
- (55) Allain, F. H.-T.; Varani, G. *J. Mol. Biol.* **1995**, *250*, 333.
- (56) Sakata, T.; Hiroaki, H.; Oda, Y.; Tanaka, T.; Ikerhara, M.; Uesugi, S. *Nucleic Acids Res.* **1990**, *18*, 3831-3839.
- (57) James, J. K.; Tinoco, I., Jr. *Nucleic Acids Res.* **1993**, *21*, 3287-3293.
- (58) Holbrook, S. R.; Cheong, C.; Tinoco, I., Jr.; Kim, S.-H. *Nature* **1991**, *353*, 579-581.
- (59) Endo, Y.; Gluck, A.; Wool, I. G. *J. Mol. Biol.* **1991**, *221*, 193-207.
- (60) Gluck, A.; Endo, Y.; Wool, I. G. *J. Mol. Biol.* **1992**, *226*, 411-424.
- (61) Zwieb, C. *Eur. J. Biochem.* **1994**, *222*, 885-890.
- (62) Senior, M. M.; Jones, R. A.; Breslauer, K. J. *Proc. Natl. Acad. Sci. U.S.A.* **1988**, *85*, 6242.
- (63) Groebe, D. R.; Uhlenbeck, O. C. *Nucleic Acids Res.* **1988**, *16*, 11725-11735.
- (64) Haasnoot, C. A.; De Bruin, S. H.; Hilbers, C. W.; van der Marcel, G. A.; van Boom, J. H. *J. Biosci.* **1985**, *8*, 767-780.
- (65) Blommers, M. J.; Walters, J. A.; Haasnoot, C. A.; Aelen, J. M.; van der Marcel, G. A.; van Boom, J. H.; Hilbers, C. W. *Biochemistry* **1989**, *28*, 7491-7498.
- (66) Haasnoot, C. A.; Hilbers, C. W.; van der Marcel, G. A.; van Boom, J. H.; Singh, U. *C. J. Biomol. Struct. Dyn.* **1986**, *25*, 843-857.

- (67) Tuerk, C.; Gauss, P.; Thermes, C.; Groebe, D. R.; Gayle, M.; Guild, N.; Stormo, G.; D'Aubenton-Carafa, Y.; Uhlenbeck, O. C.; Tinoco, I., Jr.; Brody, E. N.; Gold, L. *Proc. Natl. Acad. Sci. U.S.A.* **1988**, *85*, 1364-1368.
- (68) Antao, V. P.; Lai, S. Y.; Tinoco, I., Jr. *Nucleic Acids Res.* **1991**, *19*, 5901-5905.
- (69) Antao, V. P.; Tinoco, I., Jr. *Nucleic Acids Res.* **1992**, *20*, 819-824.
- (70) Hutvagner, C.; Zamore, P. D. *Curr. Opin. Genes. Dev.* **2002**, *12*, 225-232.
- (71) Hannon, G. J. *Nature* **2002**, *418*, 244-251.
- (72) Hunter, C. P. *Curr. Biol.* **2000**, *10*, R137-R140.
- (73) Sharp, P. A. *Genes Dev.* **1999**, *13*, 139-141.
- (74) Elbashir, S. M.; Harborth, J.; Lendeckel, W.; Yalcin, A.; Weber, K.; Tuschl, T. *Nature* **2001**, *411*, 494-498.
- (75) Bernstein, E.; Caudy, A. A.; Hammond, S. M.; Hannon, G. J. *Nature* **2001**, *409*, 363-366.
- (76) Zamore, P. D.; Tuschl, T.; Sharp, P. A.; Bartel, D. *Cell* **2000**, *101*, 25-33.
- (77) Hamilton, A. J.; Baulcombe, D. C. *Science* **1999**, *286*, 950-952.
- (78) Elbashir, S. M.; Lendeckel, W.; Tuschl, T. *Genes Dev.* **2001**, *15*, 188-200.
- (79) Nykanen, A.; Haley, B.; Zamore, P. D. *Cell* **2001**, *107*, 309-321.
- (80) Hammond, S. M.; Bernstein, E.; Beach, D.; Hannon, G. J. *Nature* **2000**, *404*, 293-296.
- (81) Hammond, S. M.; Boettcher, S.; Caudy, A.; Kobayashi, R.; Hannon, G. J. *Science* **2001**, *293*, 1146-1150.
- (82) Kennerdell, J. R.; Carthew, R. W. *Cell* **1998**, *95*, 1017-1026.
- (83) Jacque, J.-M.; Triques, K.; Stevenson, M. *Nature* **2002**, *418*, 435-438.
- (84) McManus, M. T.; Petersen, C. P.; Haines, B. B.; Chen, J.; Sharp, P. A. *RNA* **2002**, *8*, 842-850.
- (85) Celotto, A. M.; Graveley, B. R. *RNA* **2002**, *8*, 718-724.
- (86) Barre'-Sinoussi, F.; Cherman, J. C.; Rey, F.; Nugeyre, M. T.; Chamaret, S.; Gruest, J.; Dauguet, C.; Axler-Blin, C.; Vezinet-Brun, F.; Rouzioux, C.; Rozenbaum, W.; Montagnier, L. *Science* **1983**, *220*, 868.

- (87) Clavel, F.; Guetard, D.; Brun-Vezinet, F.; Chamaret, S.; Rey, M. A.; Santos Ferreira, M. O.; Laurent, A. G.; Dauguet, C.; Katlama, C.; Rouzioux, C.; Klatzman, D.; Champalimaud, J. L.; Montagnier, L. *Science* **1986**, 233, 343.
- (88) Whittle, H.; Morris, J.; Todd, J.; Corrah, T.; Sabally, S.; Bangali, J.; Ngom, P. T.; Rolfe, M.; Wilkins, A. *1994 AIDS*, 8, 1617.
- (89) Alfrano, M.; Poli, G. *Curr. Pharm. Des.* **2001**, 7, 993.
- (90) Litvak, S. *Retroviral reverse transcriptases*; R.G. Landes Publishing Co.: Austin, USA/Springer Verlag. Heidelberg. Allemagne, 1996.
- (91) Arts, E. J.; Wainberg, M. A. *Advances in Virus Res.* **1996**, 46, 97.
- (92) Hsu, M.; Wainberg, M. A. *J. Human Virol.* **2000**, 3, 16.
- (93) Kohlstaedt, L. A.; Wang, J.; Friedman, J. M.; Rice, P. A.; Steitz, T. A. *Science* **1992**, 256, 1783.
- (94) Hansen, J.; Schultze, T.; Moelling, H. *J. Biol. Chem.* **1987**, 262, 12393.
- (95) Davies, J. F.; Hostomska, Z.; Hostomsky, Z.; Jordon, S. R.; Mathews, D. A. *Science* **1991**, 252, 88.
- (96) Dhingra, M. M.; Sarma, R. H. *Nature* **1978**, 272, 798-801.
- (97) Parthasarathy, R.; Malik, M.; Friday, S. M. *Proc. Natl. Acad. Sci. USA* **1982**, 79, 7292-7296.
- (98) Sarma, R. H.; Dhingra, M. M. In *Conformation in Biology*; Srinivasan, R., Sarma, R. H., Eds.; Adenine Press: NY, 1982; pp 259-265.
- (99) Padget, R. A.; Kornarska, M. M.; Grabowski, P. J.; Hardy, S. H.; Sharp, P. A. *Science* **1984**, 225, 898-903.
- (100) Kerr, I. M.; Brown, R. E. *Proc. Natl. Acad. Sci. USA* **1978**, 75, 256-260.
- (101) Lesiak, K.; Imai, J.; Smith, G. F.; Torrence, P. F. *J. Biol. Chem.* **1983**, 258, 13082-13088.
- (102) Cailla, H.; LeBorne De Kaouel, C.; Roux, D.; Delage, M.; Marti, J. *Proc. Natl. Acad. Sci. USA* **1982**, 79, 4742-4746.
- (103) Kitade, Y.; Alster, D. K.; Pabuccuoglu, A.; Torrence, P. F. *Bioorg. Chem.* **1991**, 19, 283-299.
- (104) Farrell, P. J. *Proc. Natl. Acad. Sci. USA* **1978**, 75, 5893-5897.
- (105) Silverman, R. H. *Proc. Natl. Acad. Sci. USA* **1993**, 90, 1300-1304.

- (106) Michelson, A. M.; Monny, C. *Biochim. Biophys. Acta* **1967**, *149*, 107-126.
- (107) Westheimer, F. H. *Acc. Chem. Res.* **1968**, *1*, 70.
- (108) Ts'o, P. O. P.; Kondo, N. S.; Schwizer, M. P.; Hollis, D. P. *Biochemistry* **1969**, *8*, 997-1029.
- (109) Kondo, N. S.; Holmes, H. M.; Stempel, L. M.; Ts'o, P. O. P. *Biochemistry* **1970**, *9*, 3479-3499.
- (110) Renz, M.; Lohrmann, R.; Orgel, L. E. *Biochim. Biophys. Acta* **1971**, *240*, 463-471.
- (111) Orgel, L. E.; Lohrmann, R. *Acc. Chem. Res.* **1974**, *7*, 368-377.
- (112) Usher, D. A.; McHale, A. H. *Proc. Natl. Acad. Sci. USA* **1976**, *73*, 1149-1153.
- (113) Usher, D. A. *Nature New Biol.* **1972**, *235*, 207-208.
- (114) Srinivasan, A. R.; Olson, W. K. *Nucleic Acids Res.* **1986**, *14*, 5461-5478.
- (115) Anukanth, A.; Ponnuswamy, P. K. *Biopolymers* **1986**, *25*, 729-752.
- (116) Wodak, S. Y.; Liu, M. Y.; Wyckoff, H. W. *J. Mol. Biol.* **1977**, *116*, 855-875.
- (117) Hamada, K.; Honda, I.; Fujiwara, T.; Tomita, K. *Nucleic Acids Res. Symp. Ser.* **1981**, *10*, 137-140.
- (118) Doornbos, J.; Den Hartog, J. A. J.; Van Boom, J. H.; Altona, C. *Eur. J. Biochem.* **1981**, *116*, 403-412.
- (119) Doornbos, J.; Charubala, R.; Pfeider, W.; Altona, C. *Nucleic Acids Res.* **1983**, *11*, 4569-4582.
- (120) Krishnan, R.; Seshadri, T. P.; Viswamitra, M. A. *Nucleic Acids Res.* **1991**, *19*, 379-384.
- (121) Damha, M. J.; Giannaris, P. A.; Khan, N. *Nucleic Acids Res. Symp. Ser.* **1991**, *24*, 290.
- (122) Kierzek, R.; He, L.; Turner, D. H. *Nucleic Acids Res.* **1992**, *20*, 1685-1690.
- (123) Dougherty, J. P.; Rizzo, C. J.; Breslow, R. *J. Am. Chem. Soc.* **1992**, *114*, 6254-6255.
- (124) Hashimoto, H.; Switzer, C. *J. Am. Chem. Soc.* **1992**, *114*, 6255-6256.
- (125) Giannaris, P. A.; Damha, M. J. *Nucleic Acids Res.* **1993**, *21*, 4742-4749.
- (126) Jin, R.; Chapman, W. H.; Srinivasan, A. R.; Olson, W.; Breslow, R.; Breslauer, K. *J. Proc. Natl. Acad. Sci. USA* **1993**, *90*, 10568-10572.
- (127) Srinivasan, A. R.; Olson, W. K. *J. Biomol. Struct. Dyn.* **1987**, *4*, 895-938.

- (128) Jung, K.-E.; Switzer, C. *J. Am. Chem. Soc.* **1994**, *116*, 6059-6061.
- (129) Joyce, G. F.; Swartz, A. W.; Miller, S. L.; Orgel, L. E. *Proc. Natl. Acad. Sci. USA* **1987**, *84*, 4398.
- (130) Pudlo, J. S.; Cao, X.; Swaminathan, S.; Matteucci, M. D. *Tetrahedron Lett.* **1994**, *35*, 9315-9318.
- (131) Lalitha, V.; Yathindra, N. *Curr. Sci.* **1995**, *68*, 68-75.
- (132) Robinson, H.; Jung, K.-E.; Switzer, C.; Wang, A. H.-J. *J. Am. Chem. Soc.* **1995**, *117*, 837-838.
- (133) Alul, R.; Hoke, G. D. *Antisense Res. And Development* **1995**, *5*, 3-11.
- (134) Lorsch, J. R.; Bartel, D. P.; Szostak, J. W. *Nucleic Acids Res.* **1995**, *23*, 2811-2814.
- (135) Prakash, T. P.; Jung, K.-E.; Switzer, C. *Chem. Commun.* **1996**, 1793-1794.
- (136) Sheppard, T. L.; Breslow, R. C. *J. Am. Chem. Soc.* **1996**, *118*, 9810-9811.
- (137) Sawai, H.; Seki, J.; Ozaki, H. *J. Biomol. Struct. Dyn.* **1996**, *13*, 1043-1051.
- (138) Bhan, P.; Bhan, A.; Hong, M.; Hartwell, J. G.; Saunders, J. M.; Hoke, G. D. *Nucleic Acids Res.* **1997**, *25*, 3310-3317.
- (139) Kandimalla, E. R.; Manning, A.; Zhao, Q.; Shaw, D. R.; Byrn, R. A.; Sasisekaran, V.; Agrawal, S. *Nucleic Acids Res.* **1997**, *25*, 370-378.
- (140) Prakash, T. P.; Roberts, C.; Switzer, C. *Angew. Chem. Int. Ed. Engl.* **1997**, *36*, 1522-1523.
- (141) Wasner, M.; Arion, D.; Borkow, G.; Noronha, A.; Uddin, A. H.; Parniak, M. A.; Damha, M. J. *Biochemistry* **1998**, *37*, 7478-7486.
- (142) Premraj, B. J.; Yathindra, N. *J. Biomol. Struct. Dyn.* **1998**, *16*, 313-328.
- (143) Damha, M. J.; Noronha, A. *Nucleic Acids Res.* **1998**, *26*, 5152-5156.
- (144) Sawai, H.; Totsuka, S.; Yamamoto, K.; Ozaki, H. *Nucleic Acids Res.* **1998**, *26*, 2995-3000.
- (145) Burlina, F.; Fourrey, J.-L.; Lefort, V.; Favre, A. *Tetrahedron Lett.* **1999**, *40*, 4559-4562.
- (146) Premraj, B. J.; Raja, S.; Yathindra, N. *Biophys. Chem.* **2002**, *95*, 253-272.
- (147) Hannoush, R. N.; Damha, M. J. *J. Am. Chem. Soc.* **2001**, *123*, 12368-12374.
- (148) Hannoush, R. N.; Damha, M. J. *Nucleosides Nucleotides* **2001**, *20*, 1201-1204.

- (149) Premraj, B. J.; Patel, P. K.; Kandimalla, E. R.; Agrawal, S.; Hosur, R. V.; Yathindra, N. *Biochem. Biophys. Res. Comm* **2001**, *283*, 537-543.
- (150) Westhof, E.; Plach, H.; Cuno, I.; Ludemann, H.-D. *Acta Cryst.* **1977**, *C53*, 1694-1696.
- (151) Radwan, M. M.; Wilson, H. R. *Acta Cryst.* **1980**, *B26*, 2185-2187.
- (152) Shefter, E.; Barlow, M.; Sparks, R. A.; Trueblood, K. N. *Acta Cryst.* **1969**, *B25*, 895-909.
- (153) Krishnan, R.; Seshadri, T. P. *J. Biomol. Struct. Dyn.* **1993**, *10*, 727-745.
- (154) Nicholson, A. W. *Prog. Nucleic Acid Res. Mol. Biol.* **1996**, *52*, 1-65.
- (155) Heaphy, S.; Dingwall, C.; Ernberg, I.; Gait, M. J.; Green, S. M.; Karn, J.; Lowe, A. D.; Singh, M.; Skinner, M. A. *Cell* **1990**, *60*, 685-693.
- (156) Malin, M. H.; Tiley, C. S.; McCarn, D. F.; Rusche, J. R.; Hauber, J.; Cullen, B. R. *Cell* **1990**, *60*, 675-683.
- (157) Liao, X.; Brennwald, P.; Wise, J. A. *Proc. Natl. Acad. Sci. USA* **1989**, *86*, 4137-4141.
- (158) Singh, S. B.; Kollman, P. A. *Biophys. J.* **1996**, *70*, 1940-1948.
- (159) Williams, D. J.; Hall, K. B. *J. Mol. Biol.* **2000**, *297*, 251-265.
- (160) Braich, R. S.; Damha, M. J. *Bioconjug. Chem.* **1997**, *8*, 370-377.
- (161) McBride, L. J.; Caruthers, M. H. *Tetrahedron Lett.* **1983**, *24*, 245.
- (162) Sinha, D. N.; Biernat, J.; McManus, J.; Koester, H. *Nucleic Acids Res.* **1984**, *11*, 4539-4557.
- (163) Vinayak, R.; Colonna, F.; Tsou, D.; Mullah, B.; Andrus, A.; Sproat, B. *Nucleic Acids Symp. Ser.* **1994**, *31*, 165-166.
- (164) Sproat, B.; Colonna, F.; Mullah, B.; Tsou, D.; Andrus, A.; Hampel, A.; Vinayak, R. *Nucleosides Nucleotides* **1995**, *14*, 255-273.
- (165) Wincott, F.; DiRenzo, A.; Shaffer, C.; Grimm, S.; Tracz, D.; Workman, C.; Sweedler, D.; Gonzalez, C.; Scaringe, S.; Usman, N. *Nucleic Acids Research* **1995**, *23*, 2677-2684.
- (166) Krotz, A. H.; Klopchin, P. G.; Walker, K. L.; Srivatsa, G. S.; Cole, D. L.; Ravikumar, V. T. *Tetrahedron Lett.* **1997**, *38*, 3875-3878.

- (167) Vargeese, C.; Carter, J.; Yegge, J.; Krivjansky, S.; Settle, A.; Kropp, E.; Peterson, K.; Pieken, W. *Nucleic Acids Res.* **1995**, *26*, 1046-1050.
- (168) Terent'ev, A. P.; Vinogradova, E. V.; Chetverikov, V. P.; Dashkevich, S. N. *The Chemistry of Heterocyclic Compounds*; Interscience: New York, NY, 1970; Vol. 6, p.146.
- (169) Meyer, S. L. *Data Analysis for Scientists and Engineers*; Wiley & Sons: New York, 1975.
- (170) Serra, M. J.; Lyttle, M. H.; Axenson, T. J.; Schadt, C. A.; Turner, D. H. *Nucleic Acids Res.* **1993**, *21*, 3845-3849.
- (171) Puglisi, J. D.; Wyatt, J. R.; Tinoco, L., Jr. *Biochemistry* **1990**, *29*, 4215-4226.
- (172) Coffin, J.; Hasse, A.; Levy, J. A.; Montagnier, L.; Oroszlan, S.; Teich, N.; Temin, H.; Toyoshima, K.; Varmus, H.; Vogt, P.; Weiss, R. *Science* **1986**, *232*, 697.
- (173) Gallo, R. C.; Montagnier, L. *Sci. Am.* **1988**, *259*, 40.
- (174) Clavel, F.; Guyader, M.; Guetard, D.; Salle, M.; Montagnier, L.; Alizon, M. *Nature* **1986**, *324*, 691.
- (175) Guyader, M.; Emerman, M.; Sonigo, P.; Clavel, F.; Montagnier, L.; Alizon, M. *Nature* **1987**, *326*, 662.
- (176) DeClerq, E. *J. Med. Chem.* **1986**, *29*, 1561-1569.
- (177) Krug, M. S.; Berger, S. L. *Biochemistry* **1991**, *30*, 10614-10623.
- (178) Kedar, P. S.; Abbotts, J.; Kovacs, T.; Lesiak, K.; Torrence, P.; Wilson, S. H. *Biochemistry* **1990**, *29*, 3603-3611.
- (179) Gilboa, E.; Mitra, S. W.; Goff, S.; Baltimore, D. *Cell* **1979**, *18*, 93-100.
- (180) Majumdar, C.; Stein, C. A.; Cohen, J. S.; Broder, S.; Wilson, S. H. *Biochemistry* **1989**, *28*, 1340-1346.
- (181) Tuerk, C.; Gold, L. *Science* **1990**, *249*, 505-510.
- (182) Ellington, A. D.; Szostak, J. W. *Nature* **1990**, *346*, 818-822.
- (183) Tuerk, C.; MacDougall, S.; Gold, L. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 6988-6992.
- (184) Schatz, O.; Cromme, F. V.; Naas, T.; Lindemann, D.; Mous, J.; LeGrice, S. F. J. In *Gene Regulation and AIDS*; Papas, T. S., Ed.; Portfolio Publishing Co.: Houston, 1990.
- (185) Loya, S.; Hizi, A. *J. Biol. Chem.* **1993**, *268*, 9323-9328.

- (186) Tan, C. K.; Civil, R.; Mian, A. M.; So, A. G.; Downey, K. M. *Biochemistry* **1991**, *30*, 4831-4835.
- (187) Zhan, X.; Tan, C. K.; Scott, W. A.; Mian, A. M.; Downey, K. M.; So, A. G. *Biochemistry* **1994**, *33*, 1366-1372.
- (188) Andreola, M.-L.; Pileur, F.; Calmels, C.; Ventura, M.; Tarrago-Litvak, L.; Toulme', J.-J.; Litvak, S. *Biochemistry* **2001**, *40*, 10087-10094.
- (189) Min, B. S.; Nakamura, N.; Miyashiro, H.; Kim, Y. H.; Hattori, M. *Chem. Pharm. Bull.* **2000**, *48*, 194-200.
- (190) Borkow, G.; Fletcher, R. S.; Barnard, J.; Arion, D.; Motakis, D.; Dmitrienko, G. I.; Parniak, M. A. *Biochemistry* **1997**, *36*, 3179-3185.
- (191) ElDirani-Diab, R.; Sarih-Cottin, L.; Delord, B.; Dumon, B.; Moreau, S.; Toulme, J. J.; Fleury, H.; Litvak, S. *Antimicrob. Agents Chemother.* **1997**, *41*, 2141-2148.
- (192) Mohan, P.; Loya, S.; Avidan, O.; Verma, S.; Dhindsa, G. S.; Wong, M. F.; Huang, P. P.; Yashiro, M.; Baba, M.; Hizi, A. *J. Med. Chem.* **1994**, *37*, 2513-2519.
- (193) Boiziau, C.; Larrouy, B.; Brian, S.; Toulme', J.-J. *Nucleic Acids Res.* **1995**, *23*, 64-71.
- (194) Monia, B. P.; Lesnik, E. A.; Gonzalez, C.; Lima, W. F.; McGee, D.; Guinosso, C. J.; Kawasaki, A. M.; Cook, P. D.; Freier, S. M. *J. Biol. Chem.* **1993**, *268*, 14514-14522.
- (195) Morvan, F.; Rayner, B.; Imbach, J. L. *Anti-Cancer Drug Design* **1991**, *6*, 521-529.
- (196) Boiziau, C.; Debart, F.; Rayner, B.; Imbach, J.-L.; Toulme', J.-J. *FEBS Letters* **1995**, *361*, 41-45.
- (197) Lima, W. F.; Crooke, S. T. *Biochemistry* **1997**, *36*, 390-398.
- (198) Jayasena, S. D. *Clinical Chem.* **1999**, *45*, 1628-1650.
- (199) Brody, E. N.; Gold, L. *Rev. Mol. Biotech.* **2000**, *74*, 5-13.
- (200) Allen, R. J. L. *Biochem. J.* **1946**, *34*, 858-865.
- (201) Furuichi, Y.; Miura, K. *Nature* **1975**, *253*, 374.
- (202) Adams, C. C.; Stern, D. B. *Nucleic Acids Res.* **1990**, *18*, 6003-6010.
- (203) Hirao, I.; Yoshikawa, S.; Miura, K.-I. *FEBS Lett.* **1993**, *321*, 169-172.
- (204) Khan, I. M.; Coulson, J. M. *Nucleic Acids Res.* **1993**, *21*, 2957-2958.
- (205) Tang, J. Y.; Temsamani, J.; Agrawal, S. *Nucleic Acids Res.* **1993**, *21*, 2729-2735.

- (206) Fire, A.; Xu, S.; Montgomery, M. K.; Kostas, S. A.; Driver, S. E.; Mello, C. C. *Nature* **1998**, *391*, 806-811.
- (207) Gitelman, D. R.; Apirion, D. *Biochem. Biophys. Res. Commun.* **1980**, *96*, 1063-1070.
- (208) Court, D. *RNA processing and degradation by RNase III*; Academic Press, Inc.: New York, N.Y., 1993.
- (209) Nicholson, A. W. *FEMS Microbiol. Rev.* **1999**, *23*, 371-390.
- (210) Abou Elela, S.; Igel, H.; Ares, M., Jr. *Cell* **1996**, *85*, 115-124.
- (211) Lamontagne, B.; Tremblay, A.; Abou Elela, S. *Mol. Cell. Biol.* **2000**, *20*, 1104-1115.
- (212) Lamontagne, B.; Abou Elela, S. *Methods Enzym.* **2001**, *342*, 159-167.
- (213) Xu, H. P.; Riggs, M.; Rodgers, L.; Wigler, M. *Nucleic Acids Res.* **1990**, *18*, 5304.
- (214) Iino, Y.; Sugimoto, A.; Yamamamoto, M. *EMBO J.* **1991**, *10*, 221-226.
- (215) Rotonodo, G.; Friendewey, D. *Nucleic Acids Res.* **1996**, *24*, 2377-2386.
- (216) Zhou, D.; Friendewey, D.; Lobo Ruppert, S. M. *RNA* **1999**, *5*, 1083-1098.
- (217) Wu, H.; Xu, H.; Miraglia, L. J.; Crooke, S. T. *J. Biol. Chem.* **2000**, *275*, 36957-36965.
- (218) Chanfreau, G.; Rotonodo, G.; Legrain, P.; Jacquier, A. *EMBO J.* **1998**, *17*, 3726-3737.
- (219) Kufel, J.; Dichtl, B.; Tollervy, D. *RNA* **1999**, *5*, 909-917.
- (220) Chanfreau, G.; Abou Elela, S.; Ares, M., Jr.; Guthrie, C. *Genes Dev.* **1997**, *11*, 2741-2751.
- (221) Abou Elela, S.; Ares, M. *EMBO J.* **1998**, *17*, 3738-3746.
- (222) Chanfreau, G.; Legrain, P.; Jacquier, A. *J. Mol. Biol.* **1998**, *284*, 975-988.
- (223) Qu, L. H.; Henras, A.; Lu, Y. J.; Zhou, H.; Zhou, W. X.; Zhu, Y. Q.; Zhao, J.; Henry, Y.; Caizergues-Ferrer, M.; Bachellerie, J. P. *Mol. Cell. Biol.* **1999**, *19*, 1144-1158.
- (224) Durovic, P.; Dennis, P. P. *Mol. Microbiol.* **1994**, *13*, 229-242.
- (225) Kharrat, A.; Macias, M. J.; Gibson, T. J.; Nilges, M.; Pastore, A. *EMBO J.* **1995**, *14*, 3572-3584.
- (226) Dasgupta, S.; Fernandez, L.; Kameyama, T.; Inada, Y.; Nakamura, Y.; Pappas, A.; Court, D. L. *Mol. Microbiol.* **1998**, *28*, 629-640.

- (227) Robertson, H. D.; Mathews, M. B. *Biochimie* **1996**, *78*, 909-914.
- (228) Clemens, M. J. *Int. J. Biochem. Cell. Biol.* **1997**, *29*, 945-949.
- (229) Naduri, S.; Carpick, B. W.; Yang, Y.; Williams, B. R.; Qin, J. *EMBO J.* **1998**, *17*, 5458-5465.
- (230) Bycroft, M.; Grunert, S.; Murzin, A. G.; Proctor, M.; St Johnston, D. *EMBO J.* **1995**, *14*, 3563-3571.
- (231) Roegiers, F.; Jan, Y. N. *Trends Cell. Biol.* **2000**, *10*, 220-224.
- (232) Sun, W.; Jun, E.; Nicholson, A. W. *Biochemistry* **2001**, *40*, 14976-14984.
- (233) Blaszczyk, J.; Tropea, J. E.; Bubunencko, M.; Routzahn, K. M.; Waugh, D. S.; Court, D. L.; Ji, X. *Structure (Camb)* **2001**, *9*, 1225-1236.
- (234) Lamontagne, B.; Larose, S.; Boulanger, J.; Elela, S. A. *Curr. Issues Mol. Biol.* **2001**, *3*, 71-78.
- (235) Chanfreau, G.; Buckle, M.; Jacquier, A. *Proc. Natl. Acad. Sci. USA.* **2000**, *97*, 3142-3147.
- (236) Nagel, R.; Ares, M., Jr. *RNA* **2000**, *6*, 1142-1156.
- (237) Lebars, I.; Lamontagne, B.; Yoshizawa, S.; Abou Elela, S.; Fourmy, D. *EMBO J.* **2001**, *20*, 7250-7258.
- (238) Shiiba, T.; Komiyama, M. *Nucleic Acids Symp. Ser.* **1992**, *27*, 39-40.
- (239) Bashkin, J. K.; Jenkins, L. A. *Comments Inorg. Chem.* **1994**, *16*, 77-93.
- (240) Yeh, G. C.; Beatty, A. M.; Bashkin, J. K. *Inorg. Chem.* **1996**, *35*, 3828-3835.
- (241) Oivanen, M.; Kuusela, S.; Lonngberg, H. *Chem. Rev.* **1998**, *98*, 961-990.
- (242) Yazbeck, D. R.; Min, K.-L.; Damha, M. J. *Nucleic Acids Res.* **2002**, *30*, 1-11.
- (243) Langridge, R.; Rich, A. *Nature* **1963**, *198*, 725-728.
- (244) Arnott, S.; Chandrasekaran, R.; Leslie, G. W. *J. Mol. Biol.* **1976**, *106*, 735-748.
- (245) Weshof, E.; Sundaralingam, M. *Proc. Natl. Acad. Sci. USA.* **1980**, *77*, 1852-1856.
- (246) Broido, M. S.; Kearns, D. R. *J. Am. Chem. Soc.* **1982**, *104*, 5207-5216.
- (247) Lyamichev, V. I.; Mirkin, S. M.; Danilevskaya, O. N.; Voloshin, O. N.; Balatskaya, S. V.; Dobrynin, V. N.; Filippov, S. A.; Frank-Kamenetskii, M. D. *Nature* **1989**, *339*, 634-637.
- (248) Ahmed, S.; Henderson, E. *Nucleic Acids Res.* **1992**, *20*, 507-511.
- (249) Luo, J.; Sarma, M. H.; Yuan, R.-D.; Sarma, R. H. *FEBS* **1992**, *306*, 223-228.

- (250) Robinson, H.; van der Marcel, G. A.; van Boom, J. H.; Wang, A. H.-J. *Biochemistry* **1992**, *31*, 10510-10517.
- (251) Gehring, K.; Leroy, J. L.; Gueron, M. *Nature* **1993**, *363*, 561-565.
- (252) Chen, L.; Cai, L.; Zhang, X.; Rich, A. *Biochemistry* **1994**, *33*, 13540-13546.
- (253) Gallego, J.; Chou, S. H.; Reid, B. R. *J. Mol. Biol.* **1997**, *273*, 840-856.
- (254) Han, X. G.; Leroy, J. L.; Gueron, M. *J. Mol. Biol.* **1998**, *278*, 949-965.
- (255) Cai, L.; Chen, L. Q.; Raghavan, S.; Ratliff, R.; Moyzis, R.; Rich, A. *Nucleic Acids Res.* **1998**, *26*, 4696-4705.
- (256) Berger, I.; Egli, M.; Rich, A. *Proc. Natl. Acad. Sci. USA.* **1996**, *93*, 12116-12121.
- (257) Steiner, T. *Chem. Comm.* **1997**, 727-734.
- (258) Hartman, K. A., Jr; Rich, A. *J. Am. Chem. Soc.* **1965**, *87*, 2033-2038.
- (259) Guschlbauer, W. *Proc. Natl. Acad. Sci. USA.* **1967**, *57*, 1441-1448.
- (260) Saenger, W. In *Principles of Nucleic Acid Structure*; Springer-Verlag:, 1984.
- (261) Kanaori, K.; Maeda, A.; Kanehara, H.; Tajima, K.; Makino, K. *Biochemistry* **1998**, *37*, 12979-12986.
- (262) Nonin, S.; Phan, A. T.; Leroy, J.-L. *Structure* **1997**, *5*, 1231-1246.
- (263) Robidoux, S.; Klinck, R.; Gehring, K.; Damha, M. J. *J. Biomol. Struct. Dyn.* **1997**, *15*, 517-527.
- (264) Robidoux, S.; Damha, M. J. *J. Biomol. Struct. Dyn.* **1997**, *15*, 529-535.
- (265) Phan, A. T.; Leroy, J.-L. *J. Biomol. Struct. Dyn.* **2000**, *17*, 377-383.
- (266) Geuron, M.; Leroy, J.-L. *Curr. Opin. Struct. Biol.* **2000**, *10*, 326-331.
- (267) Robidoux, S. Ph.D. Thesis; McGill University: Montreal, 1999.
- (268) Mergny, J.-L.; Lacroix, L.; Han, X.; Leroy, J. L.; Helene, C. *J. Am. Chem. Soc.* **1995**, *117*, 8887-8898.
- (269) Collin, D.; Gehring, K. *J. Am. Chem. Soc.* **1998**, *120*, 4069-4072.
- (270) Morse, S. E.; Draper, D. E. *Nucleic Acids Res.* **1995**, *23*, 302-306.
- (271) Manzini, G.; Yathindra, N.; Xodo, L. E. *Nucleic Acids Res.* **1994**, *22*, 4634-4640.
- (272) Kanehara, H.; Mizuguchi, M.; Tajima, K.; Kanaori, K.; Makino, K. *Biochemistry* **1997**, *36*, 1790-1797.
- (273) Brahms, J.; Maurizot, J. C.; Michelson, A. M. *J. Mol. Biol.* **1967**, *25*, 465-480.
- (274) Akirimisi, E. O.; Sander, C.; Ts'o, P. O. P. *J. Mol. Biol.* **1963**, *2*, 340-344.

- (275) Lacroix, L.; Mergny, J.-L.; Leroy, J.-L.; Helene, C. *Biochemistry* **1996**, *35*, 8715-8722.
- (276) Snoussi, K.; Nonin-Lecomte, S.; Leroy, J.-L. *J. Mol. Biol.* **2001**, *309*, 139-153.
- (277) Inman, R. B. *J. Mol. Biol.* **1964**, *9*, 624-637.
- (278) Leroy, J. L.; Gueron, M.; Mergny, J.-L.; Helene, C. *Nucleic Acids Res* **1994**, *22*, 1600-1606.
- (279) Trempe, J.-F.; Wilds, C. J.; Denisov, A. Y.; Pon, R. T.; Damha, M. J.; Gehring, K. *J. Am. Chem. Soc.* **2001**, *123*, 4896-4903.
- (280) Damha, M. J.; Giannaris, P. A.; Zabarylo, S. V. *Nucleic Acids Research* **1990**, *18*, 3813-3821.
- (281) Damha, M. J.; Ogilvie, K. K. In *Methods in Molecular Biology: Protocols for Oligonucleotides and Analogues: Synthesis and Properties*; Agrawal, S., Ed.; The Humana Press: Totawa, NJ, 1993; Vol. 20, pp 81-114.
- (282) Pon, R. T.; Yu, S.; Sanghvi, Y. S. *Bioconjugate Chem.* **1999**, *10*, 1051-1057.
- (283) McClinton, M. A. *Aldrichimica Acta* **1995**, *28*, 31.
- (284) Gasparutto, D.; Livache, T.; Bazin, J.; Duplaa, A.-M.; Guy, A.; Khorlin, A.; Molko, D.; Roget, A.; Teoule, R. *Nucleic Acids Research* **1992**, *20*, 5159.
- (285) Grierson, D. In *Electrophoresis of Nucleic Acids, A Practical Approach*; Rickwood, D., Hanes, B. D., Eds.; IRL press Ltd.: Oxford, England, 1982.
- (286) Dahlberg, A. E. *J.Mol.Biol.* **1969**, *41*, 139.
- (287) Lecchi, P.; Le, H. M. T.; Pannell, L. K. *Nucleic Acids Research* **1995**, *23*, 1276.
- (288) Brunger, A. T. *X-PLOR 3.1, A System for X-Ray Crystallograpy and NMR*; Yale University Press: New Haven, CT, 1992.
- (289) de Leeuw, F. A. A. M.; Altona, C. *Quantum Chemistry Program Exchange, no 463: PSEUDOROT 3B*; Indiana University: Bloomington, IN, 1983.
- (290) Kim, S. G.; Lin, L. J.; Reid, B. R. *Biochemistry* **1992**, *31*, 3564-3574.
- (291) Fletcher, R. S.; Holleshak, G.; Nagy, E.; Arion, D.; Borkow, G.; Gu, Z.; Wainberg, M. A.; Parniak, M. A. *Protein Expression Purif* **1996**, *7*, 27-32.