Design and synthesis of functionalised polyamide/polycarbamate-based DNA analogues and their biophysical evaluation

Thesis submitted to

Savitribai Phule Pune University for the degree of

Doctor of Philosophy

in

Chemistry

By

Tanaya Bose

Research Guide
Dr. Vaijayanti A. Kumar

Division of Organic Chemistry
CSIR-National Chemical Laboratory
Pune -411008

August, 2016

CERTIFICATE

This is to certify that the work incorporated in the thesis entitled **"Design and synthesis of functionalised polyamide/polycarbamate-based DNA analogues and their biophysical evaluation"** submitted by **Tanaya Bose** for the degree of Doctor of Philosophy, was carried out by the candidate under my supervision at the CSIR-National Chemical Laboratory, Pune, 411008, India. Such materials, as has been obtained from other sources, have been duly acknowledged in the thesis.

Date: **Dr. Vaijayanti A. Kumar**

                      (Research Supervisor)

CANDIDATE'S DECLARATION

I hereby declare that the thesis entitled **"Design and synthesis of functionalised polyam-ide/polycarbamate-based DNA analogues and their biophysical evaluation"** submitted for the award of degree of *Doctor of Philosophy* in Chemistry to Savitribai Phule Pune University has not been submitted by me to any other university or Institution. This work was carried out by me at the CSIR-National Chemical Laboratory, Pune, India. Such materials as obtained from other sources have been duly acknowledged in the thesis.

Tanaya Bose, **August, 2016**

National Chemical Laboratory

Pune- 411 008

Dedicated to my mother

"Failures are the pillars of success"

- *imbibed in me by my mother*

Acknowledgement

"Dream, Dream Dream; Dreams transform into thoughts; And thoughts result in action."

A.P.J. Abdul Kalam

Finally, I touch the feet of the Almighty in obeisance for bestowing upon me the courage to face the complexities of life and complete this dissertation successfully and for inculcating in me the dedication and discipline to do whatever I undertake well.

Research becomes most enjoyable with wonderful lab mates around. I was fortunate to have a vivacious blend of talented and unique lab mates who can rise to any occasion in their own different way, in addition to being co-operative & helpful. The association of educative seniors like Dr Madhuri, Dr. Sachin, Dr. Seema, Dr. Namrata, Dr. Venu, Dr. Kiran, Dr. Anjan, and Dr. Manoj helped me soak up the essentials and the trivia. The warmth and cohesive attitude of Amit, Govind, Harsha, Ragini, Manisha, Komal, Aniket, and Atish made workplace fun and lively. Thank you all, for making my period of work fun, enjoyable and sometimes unforgettable. I thank Bhumkar for the laboratory assistance provided to me.

I am grateful to Council of Scientific and Industrial Research, Government of India, for awarding the junior and senior research fellowship and Dr. Sourav pal, former director, and Dr. Ashwini Nangia, Director, National Chemical Laboratory for helping me to carry out my research work by extending all infrastructural facilities and thus facilitating submission of this work in the form of a thesis for the award of Ph. D. degree.

It really requires some serious relaxation outside the laboratory chores to keep up a spirited pace. I was bestowed with a charged bunch of friends and acquaintances that kept my spirits high. My NCL days were made colorful with friendship of Dr. Joyasish, Dr. Krishanu, Dr. Shyamsundar, Dr. Arijit, Dr. Subhadip (DOS), Animesh and Pravin who filled in the gaps of occupational failures/ loneliness with numerous joyous moments and happiness.

Finally, I would like to mention about the special person in my life, Debabrata. This thesis deserved a special "me" which was crafted majorly by his thoughts, words and inspiration.

He's been and will be my wonderful beloved husband, friend and soulmate as we move forward together for eternity.

Tanaya Bose

Contents

Chapter 1

1   **Introduction: Nucleic acids in therapeutics**

Chapter 2

---

2   **Synthesis of polycarbamate analogue of GNA/TNA and their bio-physical studies**

Section A

Section B

Chapter 3

3 **β, γ-bis-substituted PNA with configurational and   conformational switch: binding studies with cDNA/RNA and their cellular uptake**

Section A

Chapter 4

4 Structural and functional evaluation of PNA and modified PNA as G quadruplex structure

Publications and Symposia

Publications

1. "Critical role of select peptides in the loop region of G-rich PNA in the preferred G-quadruplex topology and stability" **Tanaya Bose**, Vaijayanti A. Kumar. (manuscript communicated)

2. "β,γ-bis-substituted PNA with configurational and conformational switch: Preferred binding to cDNA/RNA and cell-uptake" **Tanaya Bose**, Anjan Banerjee, Smita Nahar, Souvik Maiti and Vaijayanti A. Kumar, *Chem. Commun.*, **2015**, 51, 7693-7696.

3. "Simple molecular engineering of glycol nucleic acid: Progression from self-pairing to cross-pairing with cDNA and RNA" **Tanaya Bose** and Vaijayanti A. Kumar, *Bioorg. Med. Chem.*, **2014**, 22, 6227-6232.

Conferences attended

- Presented a Poster in the conference "21st ISCB International Conference (ISCBC-2015)" held during 25-28th February 2015, at Central Drug Research Institute, Lucknow, India.

- Attended the symposium "17th CRSI National Symposium in Chemistry" held during 6-8th February, 2015, at National Chemical Laboratory, Pune, India.

- Presented a Poster in "10th International Symposium on Bio-Organic Chemistry (IS-BOC10)" held on 11-15th January, 2015 at Department of Chemistry, IISER Pune, Pune, India.

- Presented poster in "National Science Day Celebrations 2013" held on 26-27th February, 2013 at National Chemical Laboratory, Pune, India and won the Best Poster Prize in the area of Organic Chemistry

- Oral presentation in "8th Junior-National Organic Symposium Trust (JNOST)" held on 15-17th December, 2012 at IIT Guwahati, Assam, India.

<u>Abbreviations</u>

A	Adenine
ACN	Acetonitrile
Ac$_2$O	Acetic anhydride
aeg	Aminoethylglycine
Ala	Alanine
ap	Antiparallel
aq.	Aqueous
AS-ON	Antisense oligonucleotide
Boc	*N-tert*-butoxycarbonyl
Bz	Benzoyl
C	Cytosine
Calc.	Calculated
Cbz	Benzyloxycarbonyl
CD	Circular Dichroism
CHCA	α-Cyano-4-hydroxycinnamic acid
CPP	cell penetrating peptide
ch	Cyclohexyl
cp	Cyclopentyl
Cys	Cysteine
D-	Dextro-
DCC	Dicyclohexylcarbodiimide
DCM	Dichloromethane
DCU	Dicyclohexyl urea
DEPT	Distortionless enhancement by polarization transfer
DIAD	Diisopropyl azodicarboxylate
DIPCDI	Diisopropyl carbodiimide
DIPEA/DIEA	*N, N*-Diisopropylethylamine
dm	decimeter
DMAP	*N, N*-Dimethyl-4-aminopyridine
DMF	*N, N*-dimethylformamide
DMSO	*N, N*-Dimethyl sulfoxide
DNA	2'-deoxyribonucleic acid
ds	Double stranded
EDTA	Ethylene diamine tetraacetic acid
Et	Ethyl
EtOAc	Ethyl acetate
FDA	Food and Drug Administration
FITC	Fluorescein isothiocyanate
Fmoc	9-Fluorenylmethoxycarbonyl
g	gram

G	Guanine
gly	Glycine
GNA	Glycol Nucleic Acid
h	Hours
HBTU	2-(1H-Benzotriazole-1-yl)-1,1,3,3 tetramethyl-uronum-hexafluoro-phosphate
HIV	Human Immuno Difficiency Virus
HOBt	N-Hydroxybenzotriazole
HPLC	High Performance Liquid Chromatography
HSQC-COSY	Heteronuclear single-quantum correlation spectroscopy Correlation spectroscopy
Hyp	Hydroxy proline
IR	Infrared
K_d	Dissociation Constant
L-	Levo-
HR-MS	High Resolution-Mass Spectrometry
Lys(K)	Lysine
MALDI-TOF	Matrix Assisted Laser Desorption Ionisation-Time of Flight
MBHA	4-Methyl benzhydryl amine
MeOH	Methanol
mg	milligram
MHz	Megahertz
min	minutes
miRNA	microRNA
µL	Microlitre
µM	Micromolar
mL	millilitre
mM	millimolar
mmol	millimoles
m.p.	melting point
mRNA	messenger RNA
MS	Mass spectrometry
MW	Molecular weight
N	Normal
n.d.	not determined
nm	Nanometer
NMR	Nuclear Magnetic Resonance
NOESY	Nuclear Overhauser Effect Spectroscopy
n.t.	no transition
Obsvd.	Observed
p	Parallel
PAGE	Polyacrylamide Gel Electrophoresis
PCR	Polymerase chain reaction

ppm	Parts per million
PNA	Peptide Nucleic Acid
PTSA	*para*-toluene sulphonic acid
R_f	Retention factor
RNA	Ribonucleic Acid
RP	Reverse Phase (-HPLC)
rt	Room temperature
siRNA	small interfering Ribonucleic Acid
SPPS	Solid Phase Peptide Synthesis
ss	Single strand
T	Thymine
TBA	Thrombin binding aptamer
TBAF	Tetrabutylammonium fluoride
TBDPSCl	*tert*-Butyl(chloro)diphenylsilane
TBSCl	*tert*-Butyldimethylsilyl chloride
TBTU	O-(Benzotriazol-1-yl)-*N*,*N*,*N'*,*N'*-tetramethyluronium tetra-fluoroborate
TEA/Et$_3$N	Triethylamine
TFA	Trifluoroacetic acid
TFMSA	Trifluoromethane sulfonic acid
THF	Tetrahydrofuran
Threo	Threonine
TLC	Thin Layer Chromatography
T_m	Melting temperature
TNA	Threose Nucleic Acid
TOCSY	Total Correlation Spectroscopy
TPP	Triphenylphosphine (PPh$_3$)
t_R	Retention time
UV-Vis	Ultraviolet-Visible
Val	Valine

Abstract of Thesis

The thesis entitled "**Design and synthesis of functionalised polyamide/polycarbamate-based DNA analogues and their biophysical evaluation**" is divided into four chapters:

Chapter 1: Introduction: Nucleic acids in therapeutics

Chapter 2: Synthesis of polycarbamate analogues of GNA/TNA and their biophysical Studies

Section A: Synthesis of *R*/*S*-GCNA monomers and oligomers

Section B: Biophysical evaluation of *R*/*S*-GCNA carbamate oligomers

Chapter 3: β, γ-bis-substituted PNA with configurational and conformational switch: binding studies with cDNA/RNA and their cellular uptake

Section A: Synthesis of (*R*, *R*)/(*S*, *S*) β, γ-bis- hydroxymethyl /methoxymethyl *aeg*PNA

Section B: Biophysical evaluation of (*R*, *R*)/(*S*, *S*) β,γ-bis-hydroxymethyl /methoxymethyl substituted *aeg*PNA incorporated oligomers

Section C: Cell uptake experiments

Chapter 4: Structural and functional evaluation of PNA and modified PNA as G- quadruplex structure

Chapter 1: Introduction: Nucleic acids in therapeutics

Drugs consisting of oligonucleotide analogues capable of recognizing RNA through a Watson-Crick base-pairing mechanism, thereby arresting cellular processes at the transcription level are known as antisense agents (AS-ONs). The potential of oligodeoxyribonucleotides to act as antisense agents that inhibit viral replication in cell culture was discovered by Zamecnik and Stephenson in 1978. Efficient methods for gene silencing have been receiving increasing attention in the era of functional genomics, since sequence analysis of the human genome and the genomes of several other organisms revealed numerous genes, whose function is not yet known. As Bennett and Cowsert pointed out in their re-

view article, AS-ONs combine many desired properties such as broad applicability, direct utilization of sequence information, rapid development at low costs, high probability of success and high specificity compared to alternative therapeutic technologies and target validation. However, modification of AS-ONs is necessary for its survival and success in biological environment. Peptide Nucleic Acid (PNA) is one such milestone in the field of antisense research which has found its use in other fields like biosensing, PCR amplification, diagnosis and many more. PNA also has certain setbacks like poor aqueous solubility, low penetration inside the cells and ambiguity in distinguishing between parallel and anti-parallel mode of binding. Modifications to overcome these difficulties have been discussed in this chapter.

G-quadruplexes are higher-order DNA and RNA structures formed from G-rich sequences that are built around tetrads of hydrogen-bonded guanine bases. Potential quadruplex sequences have been identified in G-rich eukaryotic telomeres, and more recently in non-telomeric genomic DNA, e.g. in nuclease-hypersensitive promoter regions. Short PNA sequences are also known to form multimolecular G-quadruplex structures, but much needs to be explored using PNA backbone in the field of quadruplex structures. In this chapter we have also reviewed the PNA forming G-quadruplex.

Chapter 2: Synthesis of polycarbamate analogue of GNA/TNA and their biophysical studies

2A.1 Introduction

TNAs [(L)-α-threofuranosyl oligonucleotides] are an important class of nucleic acid discovered in 1999 by Eschenmoser *et al* where it was observed that the nucleic acid derived from tetrose sugar not only formed stable antiparallel duplexes with themselves but also cross paired with DNA/RNA. Inspired by Eschenmoser's TNA structure, Glycol Nucleic acid (GNA), a structurally simplified oligonucleotide mimic, was designed by Eric Meggers. The homoligomers of GNA were found to pair very well with themselves through Watson-Crick base pairing, but did not cross-pair with either RNA or DNA.
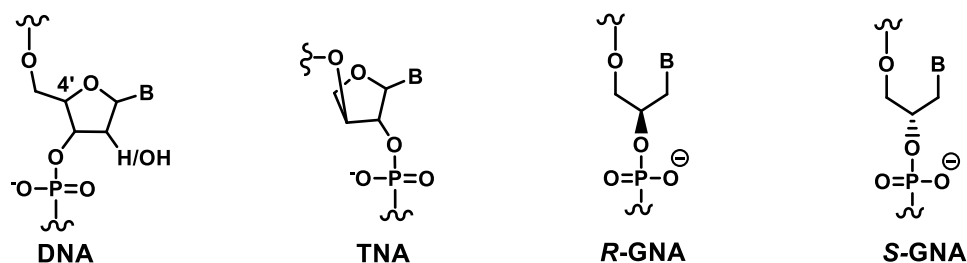
Figure 1: DNA, Threose Nucleic Acid (TNA), Glycol Nucleic Acid (GNA)

The study of carbamate linkages dates back to 1974 but was not fully exploited in the development of antisense oligomers. Earlier, pyrrolidinyl carbamate oligonucleotides were reported by our group. The flexibility in the linker group led to destabilization of the complexes. In another report by our group, the internucleoside polyamide linkages in PNA backbone were replaced by carbamate linkages and the nucleobase linker was retained as in PNA to get polycarbamate nucleic acids (PCNA). Both *R*-PCNA and *S*-PCNA formed more stable duplex with DNA.
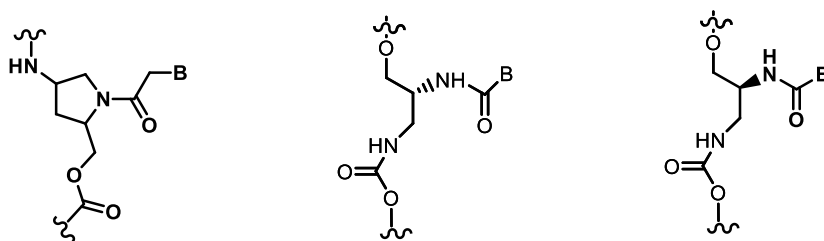
Figure 2: Chemical structures of pyrrolidine carbamate nucleic acid, *R*-PCNA and *S*-PCNA respectively.

2A.2 Rationale

In this piece of work we simplify the PCNA structure based on GNA. The open chain GNA containing phosphodiester linkages could have been highly flexible and probably was unable to reorganize to bind to the DNA/RNA backbone. Our designed monomer comprised of 5-atoms in the backbone as in the case of GNA where the phosphodiester linkage was replaced by carbamate linkage. The carbamate linkage is uncharged like PNA, but more compact compared to phosphodiester linkage in GNA and can be synthesized easily from L-serine.
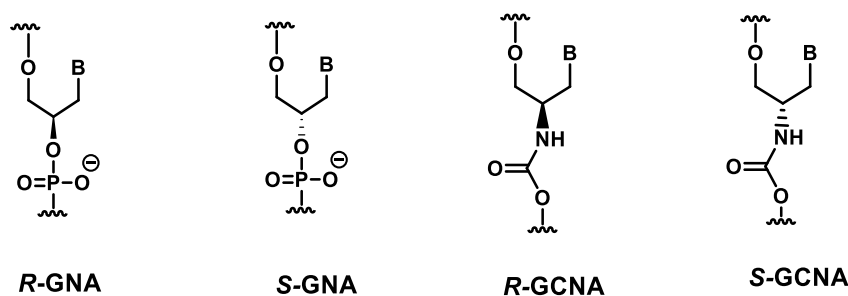
Figure 3: Glycol Nucleic acid (GNA) and our designed Glycol Carbamate Nucleic acid (GCNA)

The synthesis of both the stereochemistries of GCNA monomers was started from L-serine. Boc Strategy was chosen for the synthesis of oligomers.

2A.3 Synthesis of *R*-GCNA monomers for Boc chemistry

Naturally occurring L-serine was used as a starting material to synthesize both the enantiomers of GCNA as described in **Scheme 1** and **Scheme 2**.

.
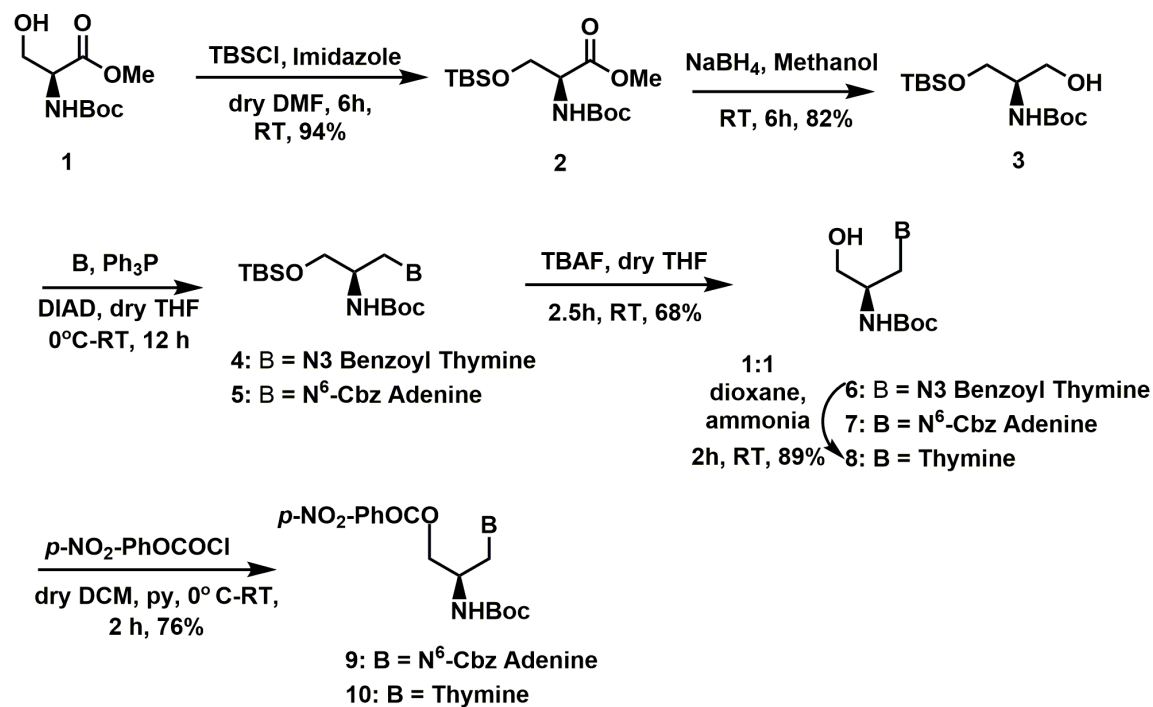
Scheme 1: Synthesis of *R*-GCNA monomer

2A.4 Synthesis of *S*-GCNA monomers
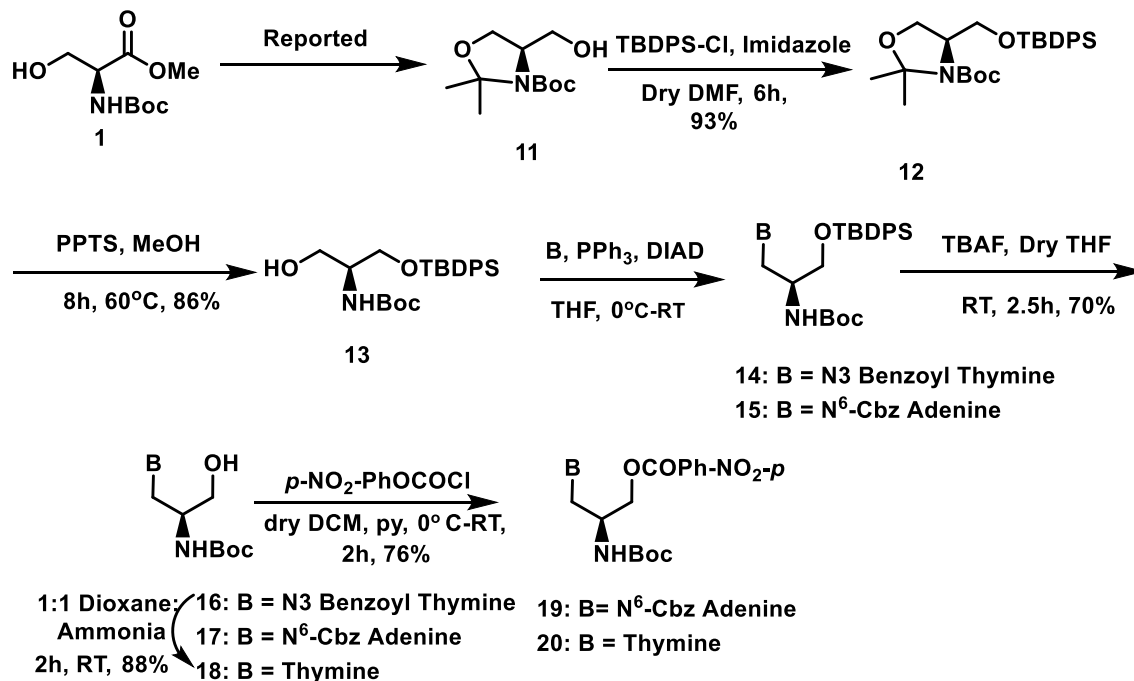
Scheme 2: Synthesis of *S*-GCNA monomer

The optical purity of the monomers **9, 10, 19, 20** was confirmed by chiral HPLC.

2A.5 Solid phase carbamate Synthesis

The Solid Phase carbamate Synthesis of GCNA oligomers started with coupling of first monomer (synthesis starts from *C*-terminus to *N*-terminus) with lysine derivatized MBHA (4-methylbenzhydryl amine) resin. Boc-deprotection followed by coupling with activated GCNA monomer gave carbamate linked dimer unit. Stepwise deprotection and coupling yielded the resin supported oligomers. The oligomers was cleaved from the resin using TFA-TFMSA conditions, purified by HPLC and characterized by MALDI-TOF.

Unlike the standard peptide chemistry, the monomer units do not require any additional activating reagents and also the excess monomer can be recovered at the end of each coupling.

2A.6 UV Job's plot

To find out the stoichiometry of binding of the pyrimidine and mixed-pyrimidine/purine GCNA sequences to their complementary DNA sequences, absorbance at fixed wavelength (λ_{max} 260 nm) was plotted against mole fraction of GCNA. Experiments show decrease in absorbance value when the concentration of *R*-GCNA-1 in the mixture increased from 0-60 to 70%. Further increase in GCNA % resulted in increased absorbance, giving rise to an inflection point on the graph. The inflection point suggested that the polypyrimidine GCNA forms complex with DNA in 2:1 ratio. However, *S*-GCNA-1 could not show any inflection point. Mixed purine and pyrimidine *R*/*S*-GCNA-2 sequences formed GCNA: DNA duplex in 1:1 stoichiometry.

2A.7 UV-melting experiment

The complex formation of GCNA was investigated initially using homothyminyl homochiral sequences comprising (*R*)-and (*S*)-thymine units, namely, *R*-GCNA-1 and *S*-GCNA-1. *S*-GCNA-1 supported the observation of UV Job's plot.

Table 1: UV melting of *R*-GCNA-1 complexes (10mM salt concentration)

Sequences	UV-T_m (°C)	
	DNA-3	**cRNA-3**
R-GCNA-1 tttttttt-Lys	35.3	29.0
S-GCNA-1 tttttttt-Lys	n.t.	32.1

Mixed pu/py sequences *R*-GCNA-2 and *S*-GCNA-2 sequences were then employed for studying their thermal stability with DNA, RNA, mismatch DNA, Parallel DNA, and for self-pairing with GCNA.

Table 2: UV melting of *R/S*- GCNA-2 complexes with cDNA, RNA and complimentary GCNA

Sequences	UV-T_m values º C						
	SS 10 mM	DNA-4 10 mM (100 mM)	*p*-cDNA-5 10 mM	mmDNA-6 10 mM	cRNA-4 10 mm (100 mM)	*R*-GCNA-3 aattaataa-tat-Lys 10 mM(10% DMSO)	*S*-GCNA-3 aattaataa-tat-Lys 10 mM(10% DMSO)
DNA-2: 5'-ATATTATT AATT	n.t.	(23.6)	n.t.	n.t.	(22.2)	n.d.	n.d.
R-GCNA-2: atattattaatt-Lys	n.t.	40.8 (39.6)	26.1	26.4	29.6 (30.0)	>80.0^c	20.1^c
S-GCNA-2:atattattaat t-Lys	20.07	28.3 (27.5)	n.t.	16.5	26.5 (24.6)	21.6^c	>81.0^c

Capital letters denote DNA/RNA sequences, small letters denote *R/S*-GCNA, DNA-4: 5'-AATTAATAATAT, *p*-cDNA-5: 5'-TATAATAATTAA mmDNA-6: 5'-AATTATTAATAT, cRNA-4: 5'-UAUAAUAAUUAA. n.d.: not determined, n.t.: no sigmoidal transition.

2.7 Circular dichroism (CD) spectra

CD studies were carried out for the monomers **9** and **19**.As expected the CD signatures showed mirror image relationship between the enantiomers.
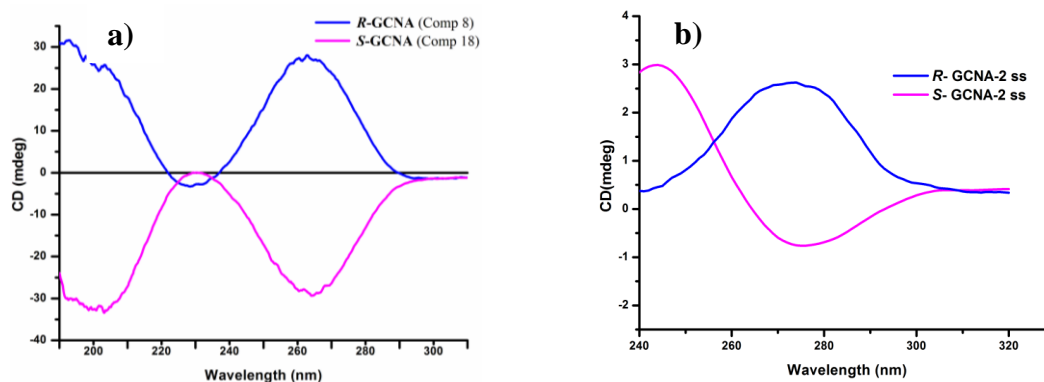
Figure 4: CD spectra of a) *R*-GCNA and *S*-GCNA monomers b) *R*-GCNA-2 ss and *S*-GCNA-2 ss

The CD plots of the single stranded *R*-GCNA-2 ss and isomeric *S*-GCNA-2 ss were oppo-site in cotton effect at ~275 nm, showing differential stacking interactions in enantiomeric single strands.
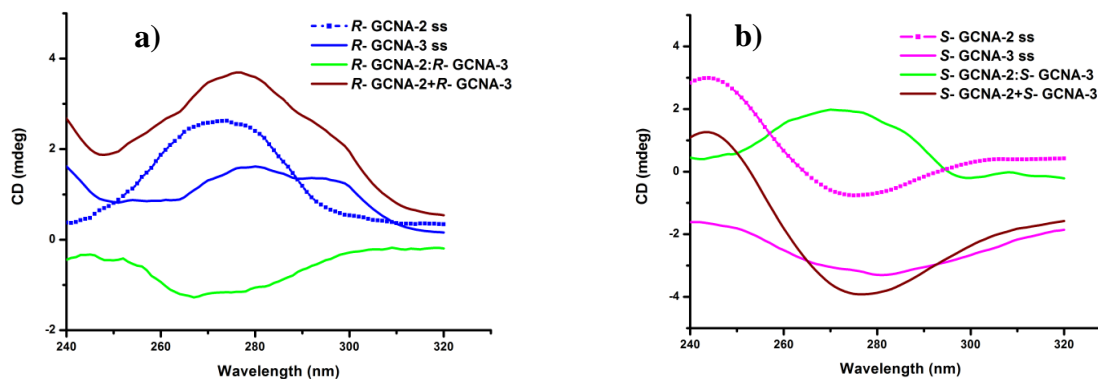
Figure 5: CD spectra of a) *R*-GCNA-2: *R*-GCNA-3 complex b) *S*-GCNA-2: *S*-GCNA-3 complex

Figure 5 shows the CD spectral properties of the enantiomeric duplexes *R/S* GCNA-2: *R/S* GCNA-3.The addition spectra of the individual oligomers were found to be distinctly dif-ferent than that of the complexes, confirming the duplex formation.

2.8 Conclusion

In conclusion, these studies describe the synthesis of chirally homogeneous (*R*)- and (*S*)-Glycol Carbamate Nucleic Acids (GCNA) from naturally occurring L-serine. The chi-rality of the monomers affected the binding properties while cross pairing with complemen-tary DNA and RNA sequences. The highly stable homochiral homogeneous complexes dis-played differently configured duplexes.

Chapter 3: β, γ-bis-substituted PNA with configurational and conformational switch: binding studies with cDNA/RNA and their cellular uptake

3A Synthesis of (*R*, *R*)/(*S*, *S*) β, γ-bis- hydroxymethyl /methoxymethyl *aeg*PNA

3A.1 Introduction

Peptide nucleic acid (PNA) is a promising class of nucleic acid mimic developed in the past two decades in which the naturally occurring sugar phosphodiester backbone is replaced with *N*-(2- aminoethyl) glycine units. PNA can hybridize to complementary DNA or RNA just as the natural counterpart, in accordance with the Watson-Crick base pairing rules, but with higher affinity and sequence specificity. Furthermore, PNA can invade selected sequences of double-stranded DNA (dsDNA). The improvement in thermodynamic stability has been attributed, in part, to the non-ionic neutral PNA backbone devoid of electrostatic repulsion in the backbone. The other contribution may come from counter ion release upon hybridization, as opposed to condensation taking place with DNA and RNA, resulting in an increase in the overall entropy of the system. Structural studies suggested that hydration may play a key role in rigidifying the backbone of PNA upon hybridization to DNA or RNA (or PNA), making it less accommodating to structural mismatches. Besides hybridization properties, another appealing aspect of PNA is enzymatic stability.

Inspite of many appealing features, PNA has some major setbacks as compared to other backbone modifications of DNA. Because of the charge-neutral backbone, PNA is only quite insoluble in water. Furthermore, it has tendency to aggregate and adhere to surfaces and other macromolecules in a nonspecific manner. Due to its achiral backbone, PNA also does not discriminate between parallel and antiparallel mode of binding with DNA. This inherent property posts a considerable technical challenge for the handling and processing of PNA. Several approaches have been attempted to address this concerns, including incorporation of charged amino acid residues, such as lysine at the termini or in the interior part of the oligomer, inclusion of polar groups in the backbone, carboxymethylene bridge, replacement of the original (aminoethyl) glycyl backbone skeleton with a negatively charged scaffold, conjugation of high molecular weight polyethylene glycol (PEG) to one

of the termini, fusion of PNA to DNA to generate a chimeric oligomer and redesign of the backbone architecture. These chemical modifications have led to improvement in solubility, but it is often achieved at the expense of binding affinity and/or sequence specificity, not to mention the requirement of an elaborate synthetic scheme and separation of enantiomers in some cases. Incorporation of cationic residues can improve water solubility, but it can also lead to nonspecific binding due to an increase in charge-charge interaction upon hybridization to DNA or RNA. .

3A.2 Rationale

Literature reports suggested that introduction of hydroxymethyl (Figure **6b**) unit in the γ position of the backbone of PNA (Figure **6a**) induced chirality and thus selectively recognized complementary DNA.
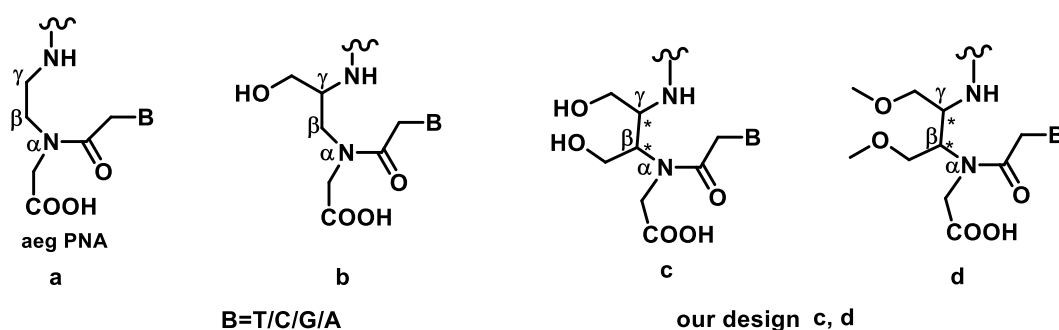
Figure 6: Chemical structure of a) *aeg***PNA, b) γ- hydroxymethyl** *aeg***PNA, c) β, γ-bis- hydroxymethyl /methoxymethyl-substituted** *aeg***PNA**

We envisioned that introduction of bis-hydroxymethyl/ methoxymethyl units on the backbone of PNA would enhance the solubility and make the oligomers viable for cellular uptake. Additionally, introduction of two chiral centres in a PNA unit may discriminate between parallel and antiparallel mode of binding. Thus, we designed bis-hydroxymethyl-PNA monomers (Figure **6c** and **d**).

3A.3 Synthesis of (*R, R/S, S*) β, γ-bis-hydroxymethyl /methoxymethyl-substituted *ae-g*PNA monomers for Boc chemistry

(*R, R/S, S*) β, γ-bis-hydroxymethyl /methoxymethyl-substituted *aeg*PNA monomers **8a₁, 8a₂, 8b₁, 8b₂, 9a₁, 9a₂, 9b₁** and **9b₂** were synthesized from D-tartaric acid and L-tartaric acid respectively following the synthetic steps given below.

2-9	(*R, R*)		(*S, S*)	
	a₁	a₂	b₁	b₂
R	Me	Bn	Me	Bn
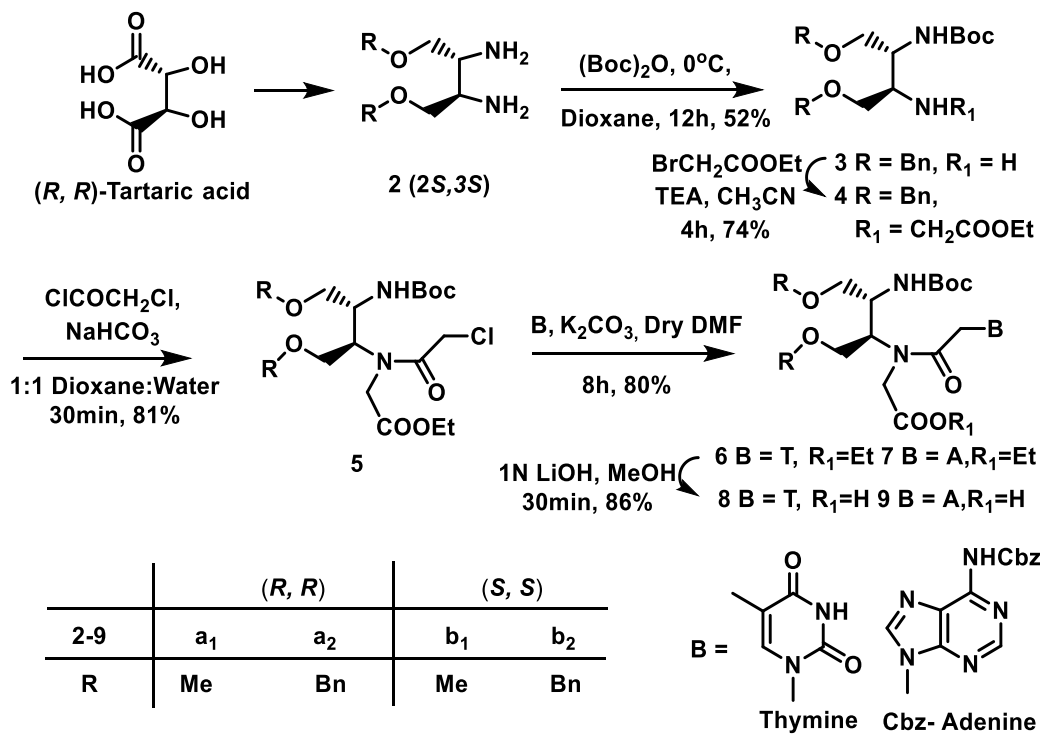
Scheme 3: Synthesis of (*R, R/S, S*) β, γ-bis-benzyloxymethyl/ methoxymethyl-substituted *aeg*PNA monomers

3A.3 Synthesis, characterization and melting experiments of oligomers

The Solid Phase Peptide Synthesis (SPPS) of PNA oligomers were undertaken by coupling of the first monomer with lysine derivatized 4-methylbenzhydryl amine (MBHA) resin using standard protocol.

3B Biophysical evaluation of (*R, R*)/(*S, S*) β,γ-bis hydroxymethyl/methoxymethyl substituted *aeg*PNA incorporated oligomers

3B.1 UV melting experiments

The thermal stability of modified oligomers was monitored by UV with complementary DNA, RNA, parallel complementary DNA and mismatch RNA.

Table 3: UV- melting of of (*R, R/S, S*) β, γ-bis-hydroxymethyl/methoxymethyl modified *aeg*PNA oligomers with *ap*-cDNA, RNA, *p*-cDNA, mmRNA

Sequence Code	*ap*-cDNA 10mM(100mM)	*p*-cDNA 10mM	cRNA 10mM(100mM)	mmRNA 10mM
PNA 1	58.3(53.7)	46.4	63.1(60.9)	46.3
PNA 3a	n.t.(n.t.)	n.t.	n.t.(n.t.)	n.t.
PNA 3b	n.t.(n.t.)	n.t.	n.t.(n.t.)	n.t.
PNA 8a	51.5(49.7)	38.2	60.8(58.5)	47.8
PNA 8b	n.t.(n.t.)	n.t.	n.t.(n.t.)	n.t.
PNA 6a	n.t.(n.t.)	n.t.	n.t.(n.t.)	n.t.
PNA 6b	n.t.(n.t.)	n.t.	n.t.(n.t.)	n.t.
PNA 2	61.5(58.5)	49.5	66.6(62.7)	51.9
PNA 4a	59.2(53.7)	38.5	64.7(59.9)	51.8
PNA 4b	43.6(36.1)	n.t.	51.7(44.3)	42.8
PNA 5a	60.0(54.5)	n.t.	66.3(63.9)	53.9
PNA 5b	n.t.(n.t.)	n.t.	n.t.(n.t.)	n.t.
PNA 10a	61.5(55.1)	n.t.	68.5(63.1)	53.7
PNA 9a	57.6(52.1)	38.5	63.2(61)	50.3
PNA 9b	50.6(47)	36.2	57.3(50.4)	37.8
PNA 7a	58.2(50.7)	38.8	61.7(57.3)	52.0
PNA 7b	38.1(28.7)	n.t.	49.2(41.9)	40.4

cDNA: 5'-CTGAAATCGGTT **Mismatch RNA**: CUGAA*U*UCGGUU
cRNA: 5'-CUGAAAUCGGUU **Parallel DNA**: 5'-TTGGCTAAAGTC

3B.2 Gel electrophoresis mobility assay

Gel electrophoresis mobility assay confirmed the strong binding of sequence **PNA 4a** to its complementary DNA and also that **PNA 3a** does not bind to its complementary DNA.
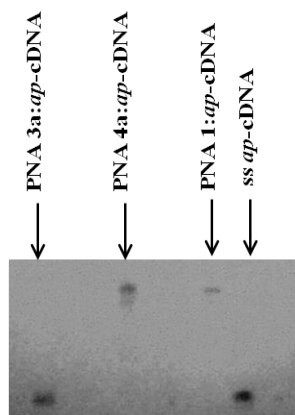
Figure 7: Gel electrophoresis mobility assay

3C Cell uptake experiments

3C.1 Carboxyfluorescein attachment to the PNA sequences

Carboxyfluorescein was attached to the synthesized PNA oligomers PNA 1, 2, 5a, 10a for studying the internalization of the oligomers into cells. To synthesize carboxyfluorescein attached PNA oligomers, couplings were carried out using PNA 1, 2, 5a, 10a on MBHA resin in presence of ten equivalents of 5(6)-carboxyfluorescein, HOBt, DIPCDI (diisopropylcarbodiimide) in DMF overnight. The oligomers were cleaved from solid phase in presence of TFA, TFMSA employing the regular protocol. The crude peptide was purified by semi preparative C18 column.

3C.2 Cellular uptake studies

HCT-116 cells were treated with CF-labeled PNAs at 1µM concentration and incubated for 10h at 37°C. In order to remove the cell surface bound PNAs, the cells were washed with heparin (1mg/ml) before flow cytometric analysis. Thus, the fluorescent posi-

tive cells that were obtained after FACS analysis were indicative of internalized PNAs and not associated with the cell membrane.
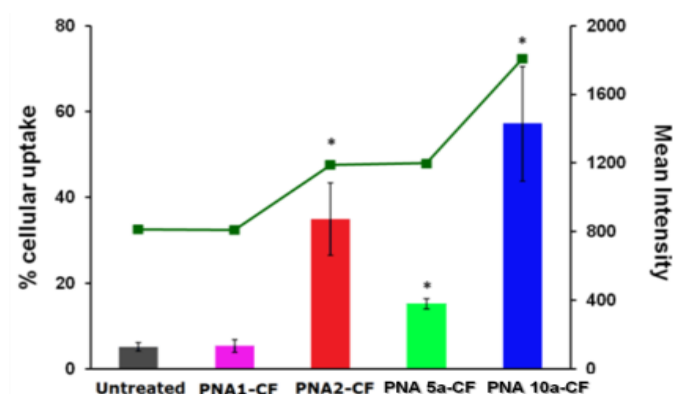
Figure 8: Flow cytometric analysis of CF-conjugated PNA 1, 2, 5a, 10a in HCT-116 cell line

In order to examine the ability of PNA [PNA (1, 2, 5a, 10a)-CF] to enter the cells without the help of transfection agent, we performed flow cytometry. The untreated cells (~5.5%) and unmodified PNA (PNA1-CF) showed ~5.4% fluorescent positive cells. The presence of bis-hydroxymethyl units in the sequence slightly improved the uptake to ~15%. Lysine residues are known to improve the cellular uptake of PNA. The unmodified PNA 2-CF with four lysine residues improved internalization upto~35 %.The bis-hydroxymethyl modified PNA containing 4-lysine units (PNA10a-CF) was most efficiently internalized without the aid of transfection agent and ~ 57% fluorescent positive cells were observed.

3C.2 Conclusion

In conclusion, these studies describe the synthesis of (*S, S*) and (*R, R*) bis-hydroxy-methyl/methoxymethyl monomers from L(+) and D(-) tartaric acid respectively. The chirality of the monomers, substitution on the backbone as well as position of modification affected the binding properties while cross pairing. A simple change in backbone from hydroxyl to methyl in (*R, R*)-bis-methoxymethyl PNA drastically changed the binding of the PNA oligomers with DNA and RNA. The sequence chosen for cell uptake experiment could invade the cells upto 57% in comparison to control (16%).

Chapter 4: Structural evaluation of PNA and modified PNA forming G-quadruplex structure

4.1 Introduction

G-quadruplexes are polyguanylic acid gels formed from stretches of contiguous guanines forming tetramers of planar hydrogen-bonded arrangements. They have been demonstrated to be stabilized by different cations. These include monovalent cations such as Na^+, K^+, Li^+, Cs^+, etc., and also divalent cations such as Ba^{++}, Ca^{++}, Sr^{++} etc. In addition, non-metal cations such as NH_4^+ have also been shown to result in stable G-quadruplexes.
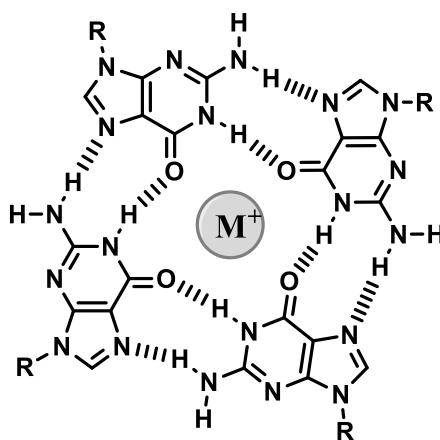
Figure 9: Structure of G-quadruplex

4.2 Rationale

Formation of G-quadruplex structures by Peptide Nucleic Acid has been reported earlier. However, unimolecular G-tertads formed from PNA have not been investigated till date. For our preliminary study we have chosen the 15mer sequence [d(GGTTGGTGTGG TTGG)]. In this quadruplex sequence the thymine bases occupy the loops.

The replacement of DNA backbone of TBA by PNA backbone has not yet been studied in literature. Here, we synthesized TBA sequence with PNA and modified PNA backbone. We also substituted the loop region (T_3, T_4 and T_{12}, T_{13}) with modified PNA monomer (bis-hydroxymethyl).
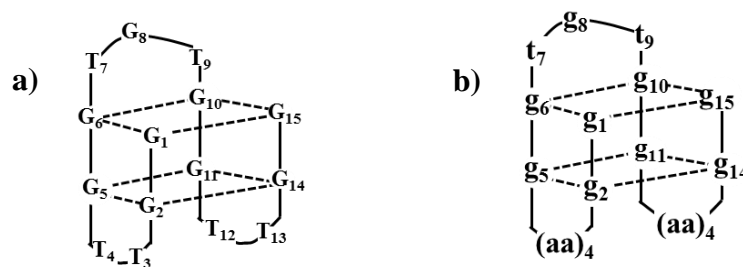
Figure 10: Structure of a) DNA-TBA G-quadruplex, b) amino acid modified PNA-TBA G-quadruplex

We also replaced these loops by amino acid sequences considering that different amino acids sequences adopt different structure in three dimensional spaces. The side chains of the amino acids would be amenable for further utilization to get the desired arrays of the side chain functional groups in three dimensional spaces. Each thymine unit would be replaced by two amino acids to meet the requirement of loop length of the quadruplex. The (R, R)-PNA-β, γ-bis-hydroxymethyl monomers synthesized previously can also be incorporated in the sequence to find out its ability to form G-quadruplex.

4.3 Synthesis of oligonucleotides

The TT loop region of DNA-TBA sequence was modified with modified bis-hydroxymethyl PNA monomers. In addition, several sequences were synthesized to replace the TT loops. The most simple, achiral amino acid glycine was used to study the effect of incorporation of amino acids on quadruplex structure. The TT loop regions were then modified with chiral amino acid L-alanine and D-alanine to study the induced chirality in the quadruplex. The PNA-TBA and modified PNA-TBA oligomers were synthesized using semi-automated microwave assisted peptide synthesizer following regular SPPS protocol.

4.4 UV melting experiments

The stability of the modified G-quadruplexes was followed by the change in the UV absorbance at 295 nm with temperature. The monovalent cations Na^+ and K^+ were necessary for the stability of the quadruplex structure, K^+ being most favored. PNA-TBA was

found to form unimolecular quadruplex. Using modified PNA monomer completely desta-bilized the quadruplex structure.

Table 4: UV melting experiment of unmodified PNA-TBA and DNA-TBA and modified PNA-TBA at 295 nm

Sequence Code	T_m water at 295 nm		T_m NaCl at 295 nm		T_m KCl at 295 nm	
	Heat	**Cool**	**Heat**	**Cool**	**Heat**	**Cool**
DNA-TBA	20.45	20.2	22.6	22.6	50.0	49.7
PNA-TBA	21.2	21.7	31.9	32.3	48.4	48.7
(*R,R*)-PNA-β,γ-OH	n.t.	n.t.	n.t.	n.t.	n.t.	n.t.
PNA-Gly	28.8	19.6	30.8	21.8	34.1	20.7
PNA-L-Ala	27	19.1	26.9	20.7	45.3	35.5
PNA-D-Ala	26.8	20.2	24.8	19.2	41.8	30.2

Modifying the loop region with glycine and alanine changed the structure from unimolecular to multimolecular quadruplex which was evident from the hysteresis observed in melting and annealing curves in cases of modified PNA oligomers when monitored at 295 nm wavelength. It is known in the literature that glycine is the most flexible prochiral amino acid whereas introduction of alanine may increase the propensity of formation of α-helical peptide geometry. The studied quadruplexes with peptides containing glycine and alanine may not have induced the turns in the oligomers required for the unimolecular quadruplex and thus might be responsible for the formation of multimolecular quadruplex.

4.5 Synthesis of PNA-TBA sequences containing hydroxyproline

Proline and its derivatives are known to induce turns in the peptide secondary struc-tures and we therefore replaced the alanine amino acids in the loop region to get a uni-molecular quadruplex of TBA sequence. Thus, hydroxyproline modifications were intro-duced in the loop region to form unimolecular quadruplex . PNA-Val was also synthesized to observe the change on introduction of chiral amino acids like valine and threonine in the loop region of the PNA-TBA sequence.

4.6 UV melting experiments

Table 5: UV melting experiment of *trans*-hydroxyproline modified PNA-TBA oligomers at 295 nm

Sequence Code	T_m water at 295 nm		T_m NaCl at 295 nm		T_m KCl at 295 nm	
	Heat	**Cool**	**Heat**	**Cool**	**Heat**	**Cool**
PNA-L-Hyp	n.t.	n.t.	n.t.	n.t.	35.3	34.9
PNA-D-Hyp	n.t.	n.t.	n.t.	n.t.	34.9	34.3
PNA-(L+D)-Hyp	n.t.	n.t.	29.5	28.9	38.2	38.8
PNA-Val	n.t.	n.t.	24.8	25.4	33.2	33.8

The absence of hysteresis in melting pattern indicated that the introduction of hydroxyproline allows the formation of unimolecular quadruplex instead of multimolecular.

Increase of salt concentration (K^+, Na^+) is known to enhance the stability of G-quadruplex. Hence, next we investigated the role of increased salt concentration in G-quadruplex formation for synthesized PNA G-quadruplexes.

Table 6: UV melting experiment of modified PNA oligomers at different salt concentrations

Sequence Code	T_m at 295nm (50mM NaCl)		T_m at 295nm (100mM NaCl)		T_m at 295nm (150mM NaCl)	
	Heat	**Cool**	**Heat**	**Cool**	**Heat**	**Cool**
DNA-TBA	19.2	19.5	22.6	22.6	22.7	23.3
PNA -TBA	19.1	18.2	31.9	32.3	32.2	32.6
PNA-L-Ala	22.6	19.7	26.9	20.7	28.2	22.9
PNA-(L+D)-Hyp	25.2	24.8	29.5	28.9	33.08	32.2

On increasing the salt concentration there was indeed a rise in the melting temperature. This is indicative of electrostatic contribution to G-quadruplex stability due to the specifically bound Na+ or K+ ions.

It is reported in the literature that increasing the oligomer concentration shows positive change in T_m for multimolecular quadruplex but remains unchanged for unimolecular quadruplex. Therefore, UV-melting experiments were performed at 10μM and 20μM oligomer concentration to determine the influence of oligomer concentration on the sequences.

Table 7: UV melting experiment of modified PNA oligomers at different oligomer concentrations

Sequence Code	5μM	10μM	20μM	Composition
DNA-TBA	50.0	49.6	48.2	unimolecular
PNA-TBA	48.4	46.8	47.8	unimolecular
PNA-Gly	34.1	39.8	65.3	multimolecular
PNA-L-Ala	45.3	48.2	61.6	multimolecular
PNA-D-Ala	41.8	53.8	63.5	multimolecular
PNA-L-Hyp	35.3	32.8	36.9	unimolecular
PNA-D-Hyp	34.9	36.8	38.1	unimolecular
PNA-(L+D)-Hyp	38.2	36.5	34.6	unimolecular
PNA-Val	33.2	32.8	35.6	unimolecular

The PNA-TBA, PNA-L-Hyp, PNA-D-Hyp, PNA-(L+D)-Hyp, PNA-Val sequences did not show any increase in T_m whereas the alanine and glycine modified sequences showed an increase on increasing the oligomer concentration.

A thermal difference spectrum (TDS) is obtained for the quadruplexes by simply recording the ultraviolet absorbance spectra of the unfolded and folded states at temperatures above and below its melting temperature (T_m). The difference between these two spectra is the TDS. The TDS has a specific shape that is unique for each type of nucleic acid structure. Characteristic TDS was observed for PNA and modified PNA oligomers as reported by Jean-Louis Mergny.

A plot of Thermal Difference Spectra Factor further confirms our finding that the quadruplexes having glycine and alanine in the loop region forms parallel quadruplex while unmodified PNA and the PNA-L-Hyp, PNA-D-Hyp, PNA-(L+D)-Hyp, PNA-Val forms antiparallel quadruplex.

4.7 Synthesis of sulphated Oligomers

DNA-TBA is known to bind thrombin in positively charged exosite I or exosite II via hydrogen bonding, electrostatic or nonpolar interactions in the loop regions of the quadruplex structures. Sulphated heparin having highest negative charge density also binds thrombin via electrostatic interactions. The PNA-TBA sequence although formed unimolecular quadruplex structure did not have the charged backbone to interact with the posi-

tively charged binding sites of thrombin. In view of this, we thought of sulphating the hydroxy groups available on the designed modified PNA of the present study. The sulphated oligomers were synthesized post-synthetically in solution phase after the cleavage of peptide from solid support.

$$\text{Ala } \beta-\text{GGGly}\overset{\text{L-}}{\underset{\text{OH}}{\text{Hyp}}}\overset{\text{D-}}{\underset{\text{OH}}{\text{Hyp}}}\text{GlyGGTGTGGGly}\overset{\text{D-}}{\underset{\text{OH}}{\text{Hyp}}}\overset{\text{L-}}{\underset{\text{OH}}{\text{Hyp}}}\text{GlyGG}$$

$$\downarrow \begin{array}{l} \text{DMF.SO}_3, \\ \text{DMF, 6h, RT} \end{array}$$

$$\text{Ala } \beta-\text{GGGly}\overset{\text{L-}}{\underset{\text{OSO}_3^{\ominus}}{\text{Hyp}}}\overset{\text{D-}}{\underset{\text{OSO}_3^{\ominus}}{\text{Hyp}}}\text{GlyGGTGTGGGly}\overset{\text{D-}}{\underset{\text{OSO}_3^{\ominus}}{\text{Hyp}}}\overset{\text{L-}}{\underset{\text{OSO}_3^{\ominus}}{\text{Hyp}}}\text{GlyGG}$$
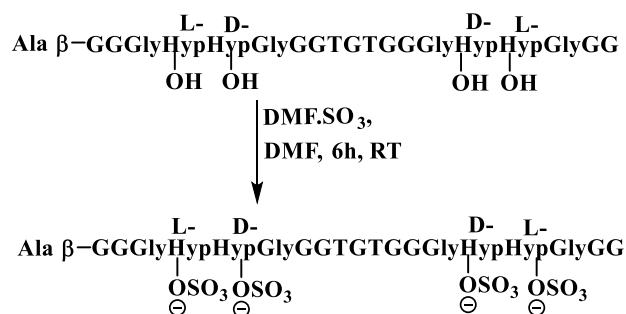
Scheme 4: Sulphation of hydroxyl groups of hydroxy proline in modified PNA oligomers

The synthesized oligomers were purified by HPLC and characterized by MALDI-TOF.

4.8 UV melting of sulphate oligomers

The melting studies showed that the synthesized sulphated oligomers did not form the quadruplex structures.

Table 8: UV melting experiment of sulphate modified PNA oligomers 100mM KCl concentration

Seq code	T_m in 100 mM KCl at 295 nm	
	Heat	**Cool**
PNA-L-Hyp Sulphate	n.t.	n.t.
PNA-D-Hyp Sulphate	n.t.	n.t.
PNA-(L+D)-Hyp Sulphate	n.t.	n.t.
PNA-Val Sulphate	n.t.	n.t.

4.9 Conclusion

- 16-mer PNA and 20-mer modified PNA oligomers were synthesized on solid support using semi-automated microwave peptide synthesizer.
- The presence of peptides in the loop region determines the structure of quadruplex.

Chapter 1

Nucleic acids in therapeutics

1.1 Introduction

"Life" on Earth manifests itself in innumerable diverse forms and complexities in terms of shape, size etc. ranging from the microscopic viruses to gigantic blue whales, from the simple cellular organisms of deep dark ocean-depths to the sunlight-harvesting plants. All life on Earth is based on nucleic acids (DNA and RNA) and proteins, which synchronizes together based on the central dogma of molecular biology[1] (**Figure 1.1**). DNA carries information that is transcribed into RNA, which then acts as a template for protein production. Proteins are cell's workhorses, which serve as important structural elements in tissues. The DNA and RNA are collectively known as nucleic acids.
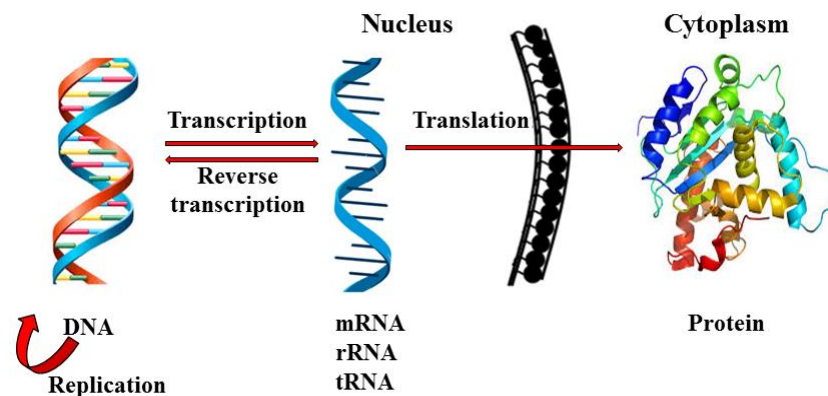
Figure 1.1: Central Dogma of cell displaying the flow of genetic information from DNA to protein

1.2 Nucleic acid structure

1.2.1 Structure of DNA/RNA

Copernican discovery by James Watson and Thomas Crick of the double-helical model of DNA structure[2] marked a milestone in the history of science, which gave rise to modern molecular biology. It was then that disparate findings by other eminent scientists of that era cohered into the complete structure of DNA/RNA, which forms the basis of modern genetic theory. Nature's simplest and elegant molecular recognition event is the base-pairing of nucleic acids, which certifies the storage, transfer and expression of genetic information in living systems.

Nucleic acids have three main components i) Pentose sugar ring (D- Ribose for RNA and 2-deoxyribose in DNA), ii) Nucleobase attached to sugar by β-glycosidic linkage, iii) 3'-OH group of one unit linked to 5'-OH group of neighboring unit through phosphate linkage. The nucleobase may be purine (adenine and guanine) or pyrimidine (cytosine, thymine in case of DNA and uracil in case of RNA).

Purine

Adenine   **Guanine**

Pyrimidine

Cytosine   **Thymine (DNA)**   **Uracil (RNA)**

DNA/RNA (R= H/OH)   **Nucleobases (B)**
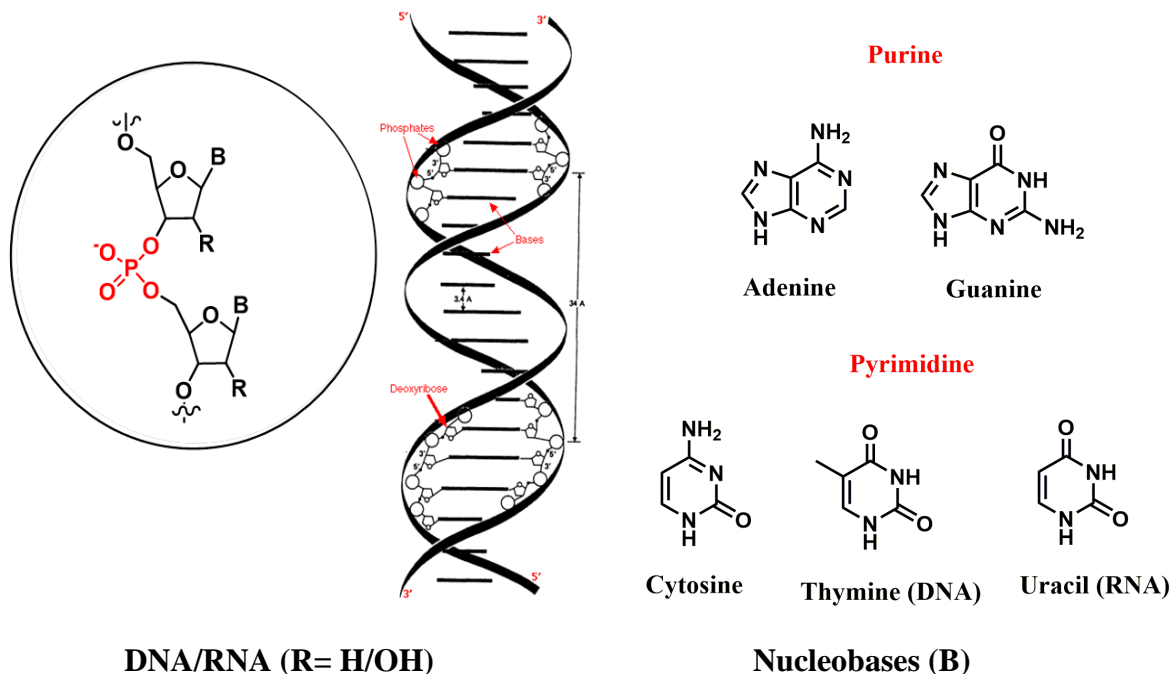
Figure 1.2: **Chemical structure of DNA and RNA**

Some other pseudo-nucleobases viz inosine, hypoxanthine, xanthine, 7-methylguanine are also present in tRNA. The presence or absence of 2'-OH group in DNA or RNA respectively dictates structural difference as well as their chemical behavior. DNA is primarily double stranded structure with two strands in opposite directions i.e. antiparallel to each other. RNA being single stranded is susceptible to formation of bulges, loops, hairpins etc. RNA is also known to function as enzyme and is vulnerable to hydrolysis due to neighboring group participation.

1.2.2 Base pairing *via* Hydrogen bonding

The N-H groups of nucleobases are powerful hydrogen bond donors, while the sp^2-hybridized electron pairs on the oxygen of the carbonyl groups and that on the ring nitrogen are hydrogen bond acceptors. Consequently, adenine forms two hydrogen bonds with thymine/uracil while guanine binds to cytosine by three hydrogen bonds. Such type of base pairing is known as Watson-Crick base pairing[2] (**Figure 1.3**) while other significant types are Hoogsten,[3](**Figure 1.4**) reverse-Hoogsten and Wobble[4] base pairing.
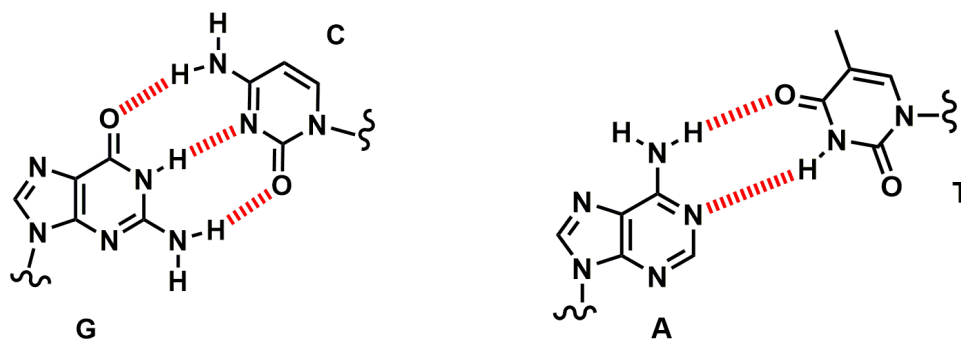
Figure 1.3: Watson-Crick hydrogen- bonding between A:T and G:C base pair

In Hoogsten base pairing, the purine rotates 180º with respect to the helix axis and adopts *syn* conformation.In addition to Watson-Crick (*N*3-*C*4) hydrogen bond of the pyrimidines, the C6-N7 face of purines are also involved in hydrogen bonding with another strand of DNA/RNA in Hoogsten mode. Such type of base pairing permits formation of triple helix.

Figure 1.4: Hoogsten hydrogen- bonding between A:T and G:C base pair

1.2.3 Sugar puckering in nucleos(t)ides

A planar conformation is energetically adverse for the pentafuranose sugar moiety of nucleotide, which results in eclipsing of all the substituents attached to carbon atoms. Consequently, the system relieves itself from the strain by puckering. The effect of non-bonded interactions between substituents at the four ring ca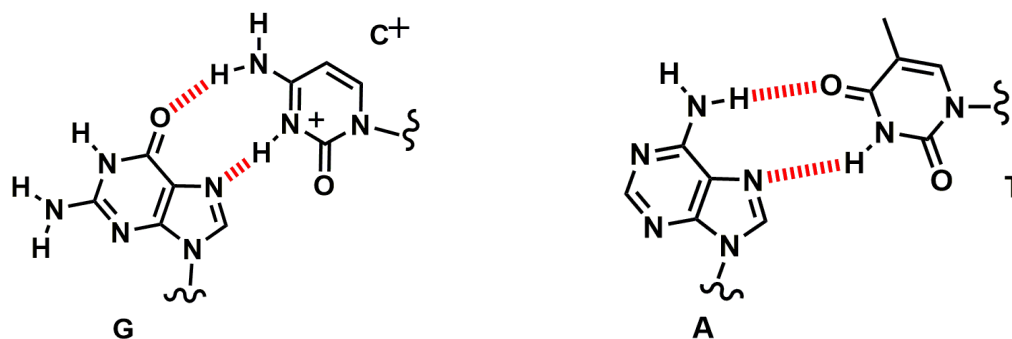rbon atoms gives rise to the ring puckering. The energetically most stable conformation for the ring has all substituents as far apart as possible. This '*puckering*' is described by recognizing the major displacement of the carbons *C2′* and *C3′* from the median plane of *C1′-O4′-C4′*. Thus, if the *endo*-displacement of *C2′* is higher than the *exo*-displacement of C3′, the conformation is called *C2′-endo* and so on for other atoms of the ring.[5] The *endo*-face of the furanose is on the same side as *C5′* and the base; the *exo*-face is on the opposite face to the base (**Figure 1.5**).

C3'-*endo* (N- type)   **C2'-*endo* (S-type)**
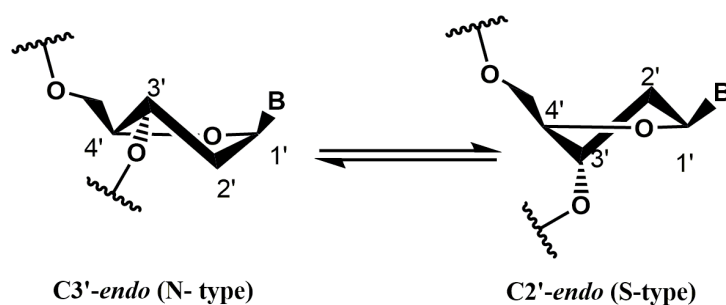
Figure 1.5: **N-type and S-type sugar puckering**

The concept of pseudorotation has been introduced to describe the conformation of ribose and deoxyribose rings in nucleotides with two variable parameters, the pseudorotation phase angle *P* and the puckering amplitude v_{max}.[6] The parameters *P* ($P = (v_4+v_1)-(v_3+v_0)/2v_2$ $(\sin36°+\sin72°)$) and v_{max} ($v_{max} = v_2/\cos P$) depends upon the furanose ring torsion angle v_0 to v_4. Under physiological conditions, the deoxyribofuranosyl sugars in DNA adopt prefer-entially 3′-*exo*, 2′-*endo* twist form (often referred to as *South* or S-type conformation), whereas ribofuranosyl sugars in RNA are in 3′-*endo*, 2′-*exo* twist form (often referred to as *North* or N-type conformation). Interestingly the conformation predominantly adopted by a nucleoside/nucleotide is not always present in a oligonucleotide double helix. In solution, N-type and S-type conformations are in rapid equilibrium and are separated by a low ener-gy barrier (**Figure 1.5**).

1.2.4 Higher order structures of Nucleic acids

Higher order nucleic acid structures such as triplex[7] and quadruplex[8] are formed involving Hoogsten mode of binding. Triplexes are prevalent when pyrimidine or purine bases occupy the major groove of DNA double helix forming Hoogsten base pairs with purines of Watson–Crick base pairs. Quadruplexes are formed from four strands and are mainly of three types depending on the type of nucleobases: i) G-quadruplex ii) i-motif iii) Holliday junction. Holliday junctions[9] (**Figure 1.6**) are four stranded structures, which serves as key intermediate in DNA recombination enabling exchange of genetic information. G quadruplexes are guanine rich sequences forming planar G-quartets[10] (**Figure 1.6**) present in telomere region, while i-motifs [11](**Figure 1.6**) are formed from intercalated cytosine and protonated cytosine of opposite strands of tetraplex.

| Triplex | Holliday junction | G-quadruplex | i-motif |
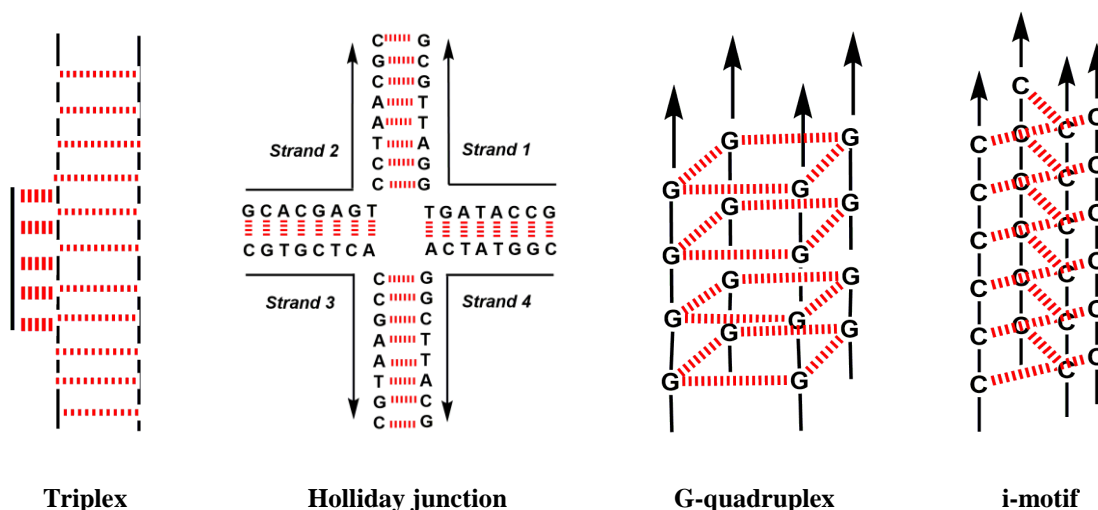
Figure 1.6: Representative figure of higher order structures of nucleic acid

1.3 Antisense approach for therapeutics

The small organic molecules available as drugs in the market targets the protein produced after transcription and translation. A deep structural knowledge of the binding site and the binding forces is required for drug designing of simple organic molecules against classical drug target proteins. Since very little is known about the process of protein folding, discovery of such drugs has various setbacks. In contrast, the nucleotide sequence

in RNA and DNA is universal and the understanding of their structure is facile. Consequently, targeting nucleic acid for therapeutics is enticing. In order for the sequence specific recognition to happen, the drug should contain nucleobases that are fundamental units of nucleic acid recognition. Two innovative strategies are being tested for inhibiting the production of disease related proteins using such sequence specific DNA fragments as gene expression inhibitors. (**Figure 1.7**)
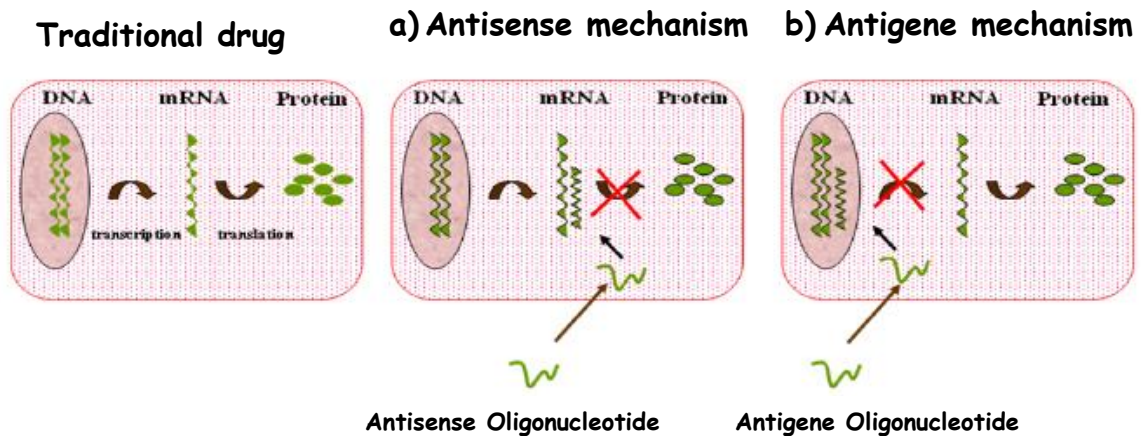

Figure 1.7: Principle of action of antigene and antisense therapy

A. Antisense therapy

Drugs consisting of oligonucleotide analogues capable of recognizing RNA through a Watson-Crick base-pairing mechanism, thereby arresting cellular processes are known as antisense agents (AS-ONs).[12] The basis for antisense method of gene regulation mechanism is to selectively impede the translation process.[13] The sequences of the bases along a messenger RNA (mRNA) molecule spell out the series of amino acids that must be strung together to make a protein. To obtain a molecule, which binds to the sense strand, one must construct a string of nucleotides having complementary `antisense´ sequence. This upon binding to the complementary region on mRNA sterically inhibits the protein synthesis machinery (**Figure 1.7a**). As Bennett and Cowsert pointed out in their review article,[14] AS-ONs combine many desired properties such as broad applicability, direct utilization of sequence information, rapid development at low costs, high probability of success and high

specificity compared to alternative technologies for gene functionalization and target validation.

B. Antigene therapy

Antigene strategy or the triplex approach for controlling gene expression aims to stall the production of undesired protein by selectively inhibiting the transcription of corresponding gene (**Figure 1.7b**). In this method, the oligonucleotide targets the major groove of DNA where it winds around the double helical DNA to form a triplex. Thus the double stranded DNA itself can act as a target for the third strand oligonucleotides or their analogue but the limitation for triplex formation is that it is possible only at homopurine stretches of DNA, since it requires purine to be the central base.

1.3.1 Antisense mechanism

Antisense therapy works by various mechanisms. Oligonucleotides acts by blocking targeted RNA without inducing its degradation, the outcome of which includes modulation of splicing when pre-mRNA is targeted , blocking of mRNA translation or RNA folding and external guide sequence (EGS)-directed mRNA degradation by a tRNA-processing ribozyme, RNase-P. They can also be used to block toxic RNAs that would otherwise seize protein factors at their expanded triplet repeats. The mechanism can be broadly categorized into two categories a) Disruptive antisense approach b) Corrective antisense approach.

1.3.1.1 Disruptive antisense approach

The AS-ONs can disturb the translation processes at mRNA level, thus causing the disruption or down-regulation of the synthesis of disease-causing functional proteins.[15] This is known as disruptive antisense approach and works via various mechanisms.

A. Antisense oligonucleotides can be used to inhibit translation by steric blockage of the ribosomes. They block the mRNA and in turn hinders the process of translation and protein formation. (**Figure 1.8**)

B. A natural enzyme called RNase H is frequently used to degrade the target mRNA. When the antisense oligonucleotides bind to the target mRNA, this enzyme cleaves the mRNA strand of a RNA-DNA hetero duplex thus preventing the translation process. **(Figure 1.8B)**

C. Ribozymes function by binding to the target RNA moiety through base pairing and in-activate it by cleaving the phosphodiester backbone at a specific susceptible end.[16] The various types of reactions performed by ribozymes are based on trans esterification, which includes splicing, oligonucleotide chain extension, RNA ligation, endonuclease action and phosphatase action. The ribozyme action is generated by formation of particular secondary and tertiary structures that create active sites. **(Figure 1.8C)**
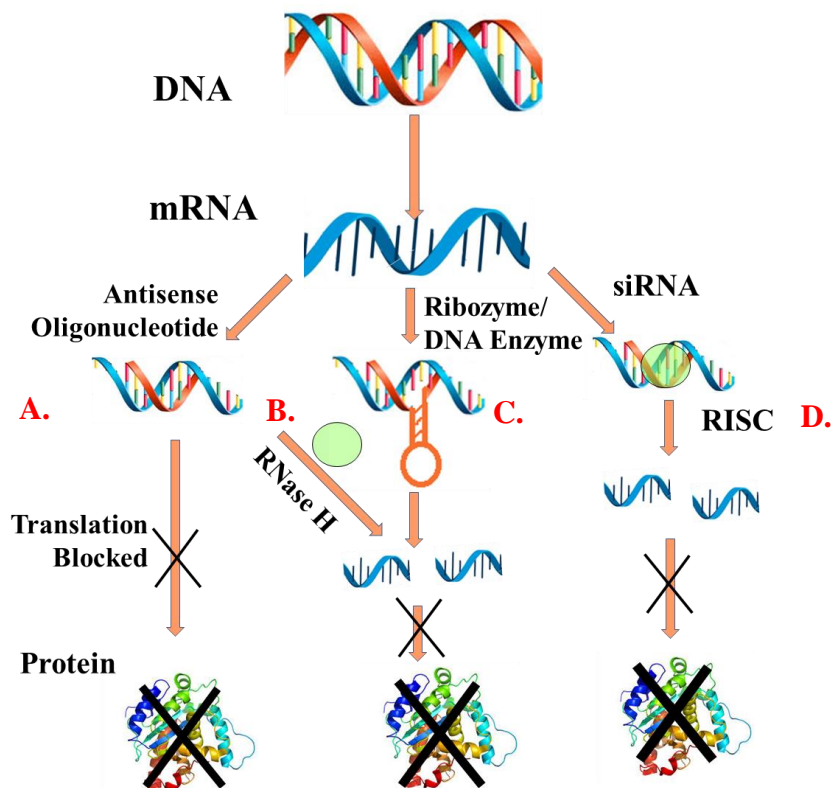

Figure 1.8: Disruptive antisense approach

D. RNA interference[15c] or RNAi is another approach for antisense mechanism. This mechanism works by two different pathways namely a) small interfering RNA (siRNA) b) micro RNA (miRNA).

The mechanism involving small interfering RNA or siRNA targets the mRNA sequence. The siRNA molecules binds to a protein complex termed as RNA-induced silencing complex (RISC), which unwinds the two RNA strands allowing the antisense strand to bind to target RNA.[17] Further RISC also contains an endonuclease activity, which hydrolyses the target RNA.[18] **(Figure 1.8D)**

MicroRNAs (miRNAs)[19] are a class of non-protein-coding RNAs that regulate gene expression post-transcriptionally. They are single-stranded RNAs of ~19-25 nucleotides in length, generated from endogenous hairpin transcript.[20] miRNAs regulate the gene expression by base pairing to partially complementary sites on the target messenger RNAs (mRNAs), usually in the 3'-untranslated region(UTR). Binding of a miRNA to the target mRNA typically leads to translational repression and exonucleolytic mRNA decay, although high complementary targets can be cleaved endonucleolytically. miRNAs have been found to regulate more than 30% of mRNAs and have roles in fundamental processes like development, differentiation, cell proliferation, apoptosis and stress responses. Both loss and gain of miRNA function contribute to cancer development through an array of different mechanisms.[21]

1.3.1.2 Corrective antisense approach

Regulation of RNA processing or RNA splice correction is another efficient mechanism in which oligonucleotides can be utilized to regulate gene expression.[22] The discovery of splicing augured a new era in the study of the molecular biology of eukaryotic gene expression. The molecular machinery that assembles a complimentary copy of *m*RNA from the set of instructions in DNA is not just carbon paper but also a pair of scissors and sticky paste. The separation and pasting together of the exons is all part of a complex editing process called splicing, and is one of the most important stages in the journey from DNA to protein. In this approach, the antisense ONs can help restoring the viable protein production by acting on pre-*m*RNA for splice corrections or to yield *m*RNA that is translated into viable proteins **(Figure 1.9)**. The disruptive antisense effects involve non-antisense and non-stop interaction leading to stimulation of immune response but corrective antisense depends upon specific and stable interaction with *m*RNA for the desired corrective action. A

promising feature of this approach is that in patients, the antisense oligonucleotides restore the correct splicing of pre-mRNA, which remains in its natural chromosomal environment. This precludes the possibility of over-expression or inappropriate expression of mRNA and protein. These oligonucleotides do not remove the mutation and therefore requires periodic administrations. This approach is, thus, more akin to a pharmacological treatment than to gene therapy.
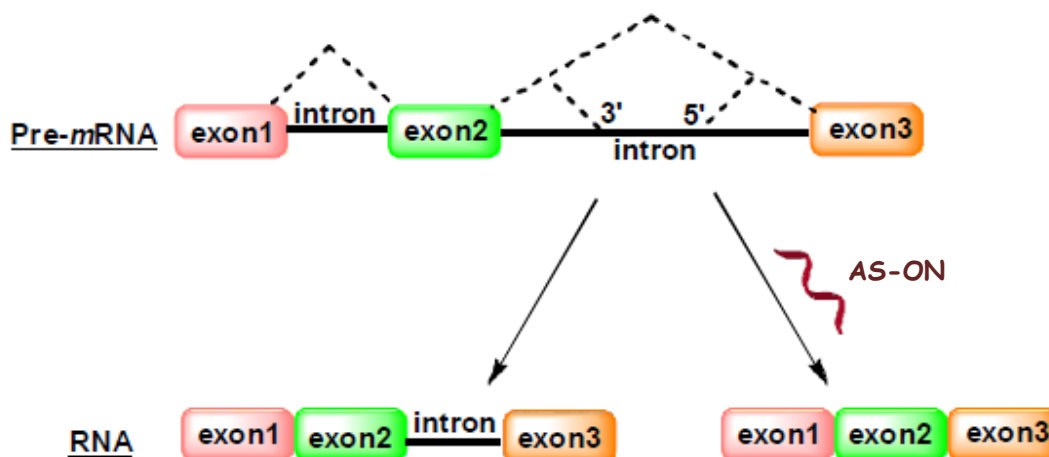

Figure 1.9: Corrective antisense approach

1.3.2 Antisense oligonucleotides from primitive to modern

Chemical modifications of natural DNA/RNA are a prime requisite to augment the properties of oligonucleotides. The key hurdles that need to be addressed are: hybridization to its complementary DNA/RNA, toxicity, potency and nuclease resistance. A number of such chemically modified oilgonucleotides addressing these hindrances are under clinical trial.[23]The first generation antisense oligonucleotides emerged the phosphorothioates[24] in which non-bridging oxygen atom in the phosphodiester linkage was replaced by sulphur atom. They were easy to synthesize and exhibited acceptable pharmacokinetic property. Vitravene[25], a phosphorothioate modified oligonucleotide was the first antisense oiligonucleotide to reach the stage of clinical trial which blocks translation of viral mRNA and confers greater resistance to nuclease degradation. However, these oligomers have a tendency

to induce non-specific effects, through binding to extra cellular and cellular proteins as well as cleavage of non-target mRNA which are only partially complimentary.

The next generation of antisense therapy introduced *C*2' modified analogues. The only chemical difference between DNA and RNA is the 2' substitution in the sugar ring. RNA: RNA duplexes are known to be thermally more stable in comparison to DNA: DNA duplexes. RNA exists in *C*3' *endo* conformation and introduction of electronegative sub-stituent shifts the ribose conformational equilibrium to *C*3' *endo* pucker.[26] Thus 2'-O-Me,[27] 2'-O-methoxyethyl,[28] 2'-F modifications[29] and various other modifications were reported in this era of antisense therapy. These analogues were slightly less toxic and showed higher thermal stability towards complementary RNA. *Mipomersen,* a drug based on 2'- O-methoxymethyl modification of phosphorothioates has been approved by United States FDA for treatment of homozygous familial hypercholesterolemia.[30]
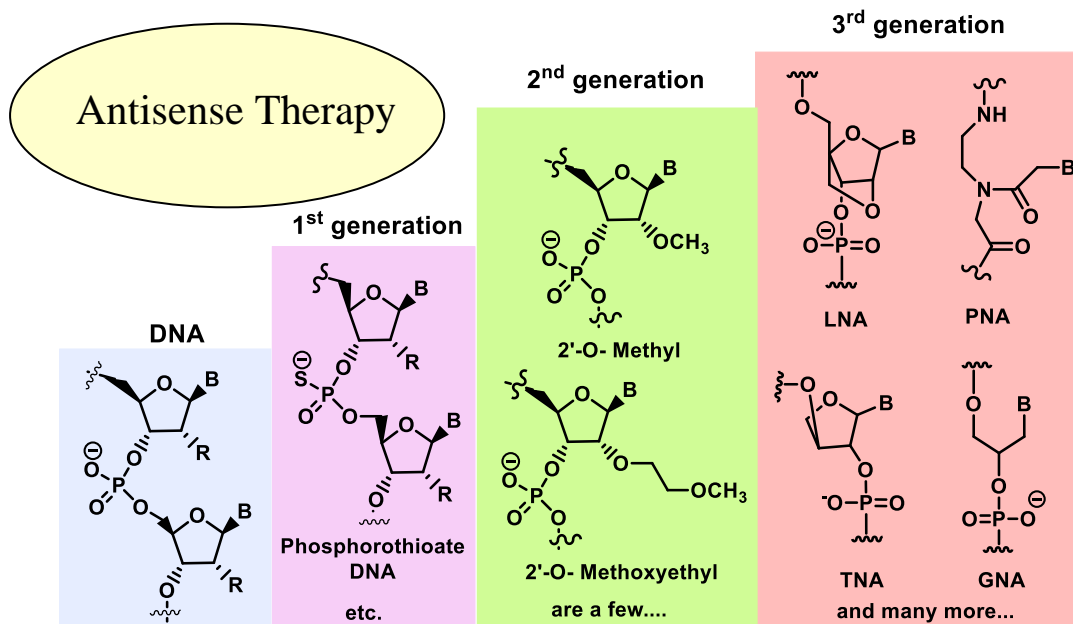

Figure 1.10: Generations of antisense therapy

The next generation of antisense oligonucleotides surfaced other modifications viz Locked Nucleic Acid (LNA),[31] Phosphorodiamidate Morpholino Oligomer (PMO),[32] 2', 4' Bicyclic Nucleic Acid (BNA).[33] Among these, BNA/LNA showed dramatically improved hybridization properties and nuclease resistance. Threose Nucleic Acid (TNA)[34] discovered

by Albert Eschenmoser has a backbone structure composed of repeating threose sugars linked together by phosphodiester bonds. TNA was found to be capable of forming antiparallel duplexes by self-pairing and was able to cross-pair with cDNA and RNA. Henceforth, TNA was believed to be an evolutionary pathway to RNA.[35] Inspired by Eschenmoser's TNA structure Eric Meggers reported Glycol Nucleic Acid (GNA)[36] an open chain analogue of TNA, which formed highly stable antiparallel duplex. One of the distinguished discoveries in the field of antisense therapy during this period was Peptide Nucleic Acid (PNA)[37] where the sugar-phosphate backbone was replaced by neutral, achiral aminoethyl glycyl backbone. They were found to form strong complex with complementary DNA /RNA. Other notable modifications reported in this period of antisense are cyclohexene nucleic acids (CeNA),[38] 2´-Deoxy-2´-fluoro-β-*D*-arabino nucleic acid (FANA)[39] to name a few.

The approval of *Vitravene*, *Mipomersen* and *Macugen* by the FDA has flared up research on antisense-based therapeutics. Furthermore, with persistent promising clinical trials involving these modified oligonucleotides, it can be predicted that more new potent antisense drugs may appear in the near future.

1.4 Charge neutral backbone modifications of DNA

The necessity of synthesizing charge neutral backbone modified DNA rose from the fact that these oligomers will be nuclease resistant as well as will overcome the electronic repulsion emerging due to phosphate backbone.[40] Phosphorodiamidate morpholino oligomer (PMO)[41] (**Figure 1.11A**) are nonionic DNA analogs, in which the ribose is replaced by a morpholino moiety and phosphoroamidate intersubunit linkages are used instead of phosphodiester bonds. PMOs do not activate RNase H, however, if inhibition of gene expression is desired, they should therefore be targeted to the 5´-untranslated region or to the first 25 bases downstream of the start codon to block translation by preventing ribosomes from binding. Because their backbone is uncharged, PMOs are unlikely to form unwanted interactions with nucleic acid-binding proteins. Their target affinities is similar to that of isosequential DNA ONs, but lower than the strength of RNA binding achieved with many of the other modifications.

Another type of charge neutral backbone reported possible, to replace the phosphate backbone of DNA is the carbamate linkage. In 1974, Michael J. Gait *et al* [42] utilized these linkages to form DNA mimics, where a dinucleotide analogue containing the oxyformam ido-linkage, thymidinylformamide-[3'(O) 5' (C)]-5'- deoxythymidine (**Figure 1.11 B**) was synthesized. A number of attempts have been made since then to synthesize different car-bamate analogues, as these linkages are stable under physiological conditions and are re-sistant to nuclease action.

Among all the non-ionic backbones of DNA the modification worth special men-tion is Peptide Nucleic Acid (PNA) (**Figure 1.11C**) first introduced by Nielsen *et al* in 1991 where the sugar phosphate backbone has been replaced by polyamide linker composed of *N*-2-aminoethylglycine repeating units, covalently linked to nucleobases through a carbox-ymethyl spacer.[43] Details about Peptide Nucleic Acids (PNAs) have been elaborately dis-cussed in the following sections.
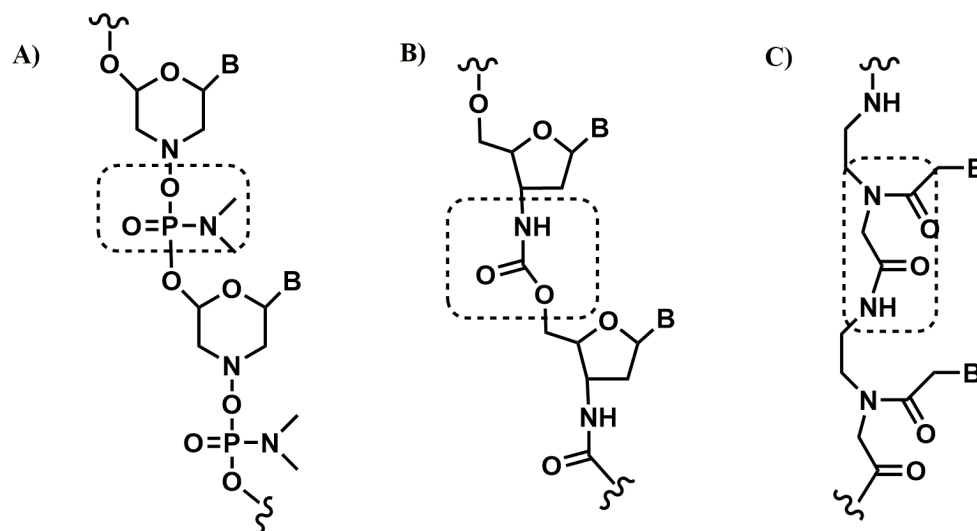

Figure 1.11: A) Phosphorodiamidate modified oligomer (PMO), B) Carbamate modified oligomer, C) Peptide modified oligomer (PNA)

1.5 Peptide Nucleic Acids (PNA)

Peptide Nucleic Acids (PNAs) are DNA analogues first introduced by Nielsen *et al* in 1991[37] where the sugar phosphate backbone has been replaced by polyamide linker

composed of *N*-2-aminoethylglycine repeating units, covalently linked to nucleobases through a carboxymethyl spacer.[43] They bind through Watson-Crick base pairing with their complementary DNA/RNA.[44] The neutral achiral backbone of PNA confers extra stability with complementary DNA/RNA in comparison to the natural counterpart in parallel as well as antiparallel fashion.[45] Due to high hybridization properties,[46] chemical and biostability,[47] PNA emerged as a successful lead for antisense therapy,[48] biosensors,[49] molecular biology,[50] diagnostic markers[51] and PCR amplification.[52]
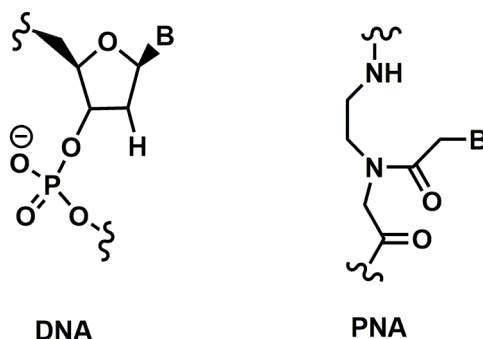

Figure 1.12: Chemical structure of DNA and PNA

1.5.1 Important properties of PNA

PNA has several properties to prove itself as a promising antisense or antigene drug, such as stable and highly sequence specific binding to the complementary *m*RNA or *ds*DNA gene target, high biological and chemical stability. The easy synthetic accessibility of PNA allows further optimization of the structure, especially with regard to bioavailability and pharmacokinetic properties. Thus, it is not surprising that the drug aspects together with the utility as a molecular biology tool of PNA technology is being actively pursued, and the results so far are very encouraging.

1.5.1.1 Duplex formation with complementary oligonucleotides

PNAs were originally conceived as mimics of triple helix forming oligonucleotides designed for sequence specific targeting of double stranded DNA *via* major groove recognition. Nevertheless, later on it was found that PNA is indeed a very potent structural mimic of DNA, capable of forming Watson-Crick base pair dependent double helices with se-

quence complementary to DNA, RNA or PNA.[53] Unlike DNA or other DNA analogues, PNAs do not contain any pentose sugar moieties or phosphate groups. They are depicted as peptides from the *N*-terminus to the *C*-terminus, corresponding to the 5´ to 3´ direction as in DNA. Since all intramolecular distances and the configuration of the nucleobases are similar to those in natural DNA molecules, specific hybridization occurs between PNA and DNA or RNA sequences by hydrogen bonding. Though in DNA: DNA duplexes, the two strands are always in antiparallel orientation (with the 5´-end of one strand opposed to the 3´-end of the other), PNA: DNA adducts can be formed in two different orientations, arbitrarily termed *parallel* and *antiparallel*. Both adducts are formed at room temperature, with the antiparallel orientation showing higher stability. This creates the possibility for PNAs to bind two DNA tracts of opposite sequence. The stability of DNA: DNA hybrids were shown to increase with increasing salt concentration, whereas in case of PNA: DNA duplex the stability remains same.[54] The contrasting effect of ionic strength on duplex formation can be explained by the association of counter ions in case of DNA: DNA duplex formation and by displacement of counter ions in the case of PNA: DNA duplex formation. One of the most important features of the PNA: DNA duplex is that their stability is highly sensitive to the presence of a single mismatched base pair. Thus, PNA probes are very sequence-selective and are superior to DNA probes in recognizing single-base mispairing.

1.5.1.2 Antigene properties

Four modes of binding for sequence-specific targeting of double-stranded DNA by PNA have been identified (**Figure 1.13**). Three of these modes involve invasion of the DNA duplex by PNA strands.

Triplex   Triplex Invasion   Duplex Invasion   Double Duplex Invasion   Tail-Clamp
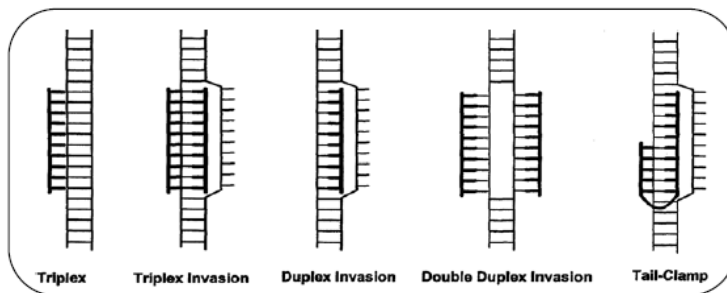
Figure 1.13: Various structural modes of binding of PNA oligomers to sequence complementary targets in *ds*DNA

It is possible either for a single PNA (homopurine) strand to invade ('duplex invasion') via Watson-Crick base pairing, or alternatively, invasion may be accomplished by two pseudo-complementary PNA strands, each of which binds to one of the DNA strands of the target ('double duplex invasion'). These pseudo complementary PNAs contain modified adenine and thymine nucleobases that do not allow stable hybridization between the two sequence complementary PNAs, but do permit good binding to the DNA. In order to obtain efficient binding in double duplex invasion of pseudo-complementary PNAs, the target and thus the PNAs, should contain at least 50% AT (no other sequence constraints), and in the PNA oligomers, all A/T base pairs are substituted with 2,6-diaminopurine/ 2-thiouracil 'base pairs'. This base pair is very unstable due to steric hindrance. Therefore, the two sequence-complementary PNAs will not be able to bind each other, but they bind their DNA complement very well. The third invasion ('triplex invasion') requires a homopurine DNA target and complementary homopyrimidine PNAs that bind the purine DNA strand through combined Watson-Crick and Hoogsteen base paring (**Figure 1.13**) *via* formation of a very stable PNA$_2$:DNA triplex. For most applications, the two PNA strands are connected in a bis-PNA designed such that the one strand is antiparallel (W-C strand) and the other strand is parallel (H-strand) to the DNA target. The most efficient binding at physiological pH is obtained when cytosine in the PNA H-strand is replaced by pseudo-isocytosine, which mimics *N*3-protonated cytosine.

1.5.1.3 Thermal stability of PNA and its hybrid complexes

PNA oligomers have properties favorable to many molecular biological applications. One of the most impressive is the higher thermal stability of PNA/DNA and PNA/RNA duplexes compared with DNA/DNA and DNA/RNA duplexes.[55] This stronger binding is attributed to the lack of charge repulsion between the neutral PNA strand and the DNA or RNA strand. Typically, the melting temperature (T_m) of a 15-mer PNA/DNA duplex is ~70°C, whereas the corresponding DNA/DNA duplex exhibits a T_m of ~55°C (pH 7, 100mmoldm^{-3} NaCl). PNA also forms very stable duplexes with RNA with a similar increase in T_m. As a general trend, T_m of a PNA/DNA duplex is 1°C higher per base pair than T_m of the corresponding DNA/DNA duplex (in 100mmoldm^{-3} NaCl).

1.5.1.4 Greater specificity of interaction

PNA also shows greater specificity in binding to complementary DNA.[56] A PNA/DNA mismatch is more destabilizing than a mismatch in a DNA/DNA duplex. A single mismatch in mixed PNA/DNA 15-mers lowers T_m by 8-20°C (15°C on average). In the corresponding DNA/DNA duplexes a single mismatch lowers T_m by 4-16°C (11°C on average).

1.5.1.5 Strand invasion

PNA oligomers that contain only thymine and cytosine (pyrimidines) often bind with a PNA$_2$: DNA stoichiometry, resulting in PNA: DNA: PNA triplex formation.[57] The triplex comprises a PNA: DNA double helix (formed by Watson–Crick hydrogen bonds) with a second PNA strand lying in the major groove of the duplex (held by Hoogsteen hydrogen bonds). The stability of these triplexes are so great ($T_m > 70$°C for a 10-mer) that 'strand invasion' of DNA: DNA is possible, binding of the PNA results in the formation of a D-loop in the double-stranded DNA, where the PNA displaces one of the DNA strands. This characteristic may potentially be used to manipulate gene expression at the transcriptional level.

1.5.1.6 Stronger binding independent of salt concentration

Another important consequence of the neutral backbone is that the T_m values of PNA/DNA duplexes are practically independent of salt concentration. This is in sharp contrast to the T_m values of DNA/DNA duplexes, which are highly dependent on ionic strength.[54] At low ionic strength, PNA can bind to a target sequence at temperatures where DNA hybridization is strongly inhibited. Hybridization of PNA can also proceed in the absence of Mg^{2+}, a factor that further inhibits DNA/DNA duplex formation. Adjusting the ionic strength can therefore be very useful in designing procedures where competing DNA or RNA is present in the sample or where the nucleic acid being probed contains a high level of secondary structure.

1.5.1.7 Resistance to nucleases and proteases

PNAs with their peptide backbone bearing purine and pyrimidine bases are the molecular species that are not easily recognized by either nucleases or proteases.[58] Therefore, the lifetime of these compounds is extended both *in vivo* and *in vitro*.

1.5.1.8 Insolubility of PNA and cellular uptake

PNAs are charge-neutral compounds and hence have poor water solubility compared with DNA. Neutral PNA molecules have a tendency to aggregate to a degree that is dependent on the sequence of the oligomer. PNA solubility is also related to the length of the oligomer and to the purine/pyrimidine ratio.[54]Some recent modifications, including the incorporation of positively charged lysine residues (carboxyl-terminal or backbone modification in place of glycine), have shown improvements in solubility. Negative charges may also be introduced, especially in PNA/DNA chimeras, which will enhance the water solubility. Poor solubility of PNA and cell wall impermeability are the major roadblocks for them to be used for therapeutic purpose. Thus efficient cellular delivery systems for PNAs are required if these are to be developed into antisense and antigene agents.

1.5.2 Chemical modifications of PNA

Since the discovery of PNA, many modifications has been reported to overcome the shortcomings of PNA i) selectivity between parallel and antiparallel mode of binding ii) improved solubility and cell internalization. Various attempts have been made to conquer these problems. Some of these attempts have been elucidated below.

1.5.2.1 Preorganization through conformational constraints

Contrary to DNA, PNA binds to DNA and RNA in both parallel and antiparallel fashion. This brings in the necessity of modification to selectively bind to a particular sequence in particular orientation. Selectivity between parallel and antiparallel mode of binding can be consummated by preorganization which can be achieved either by

A. inserting substituents at α, β or γ carbon

B. inserting aminoethyl group in cyclic structures

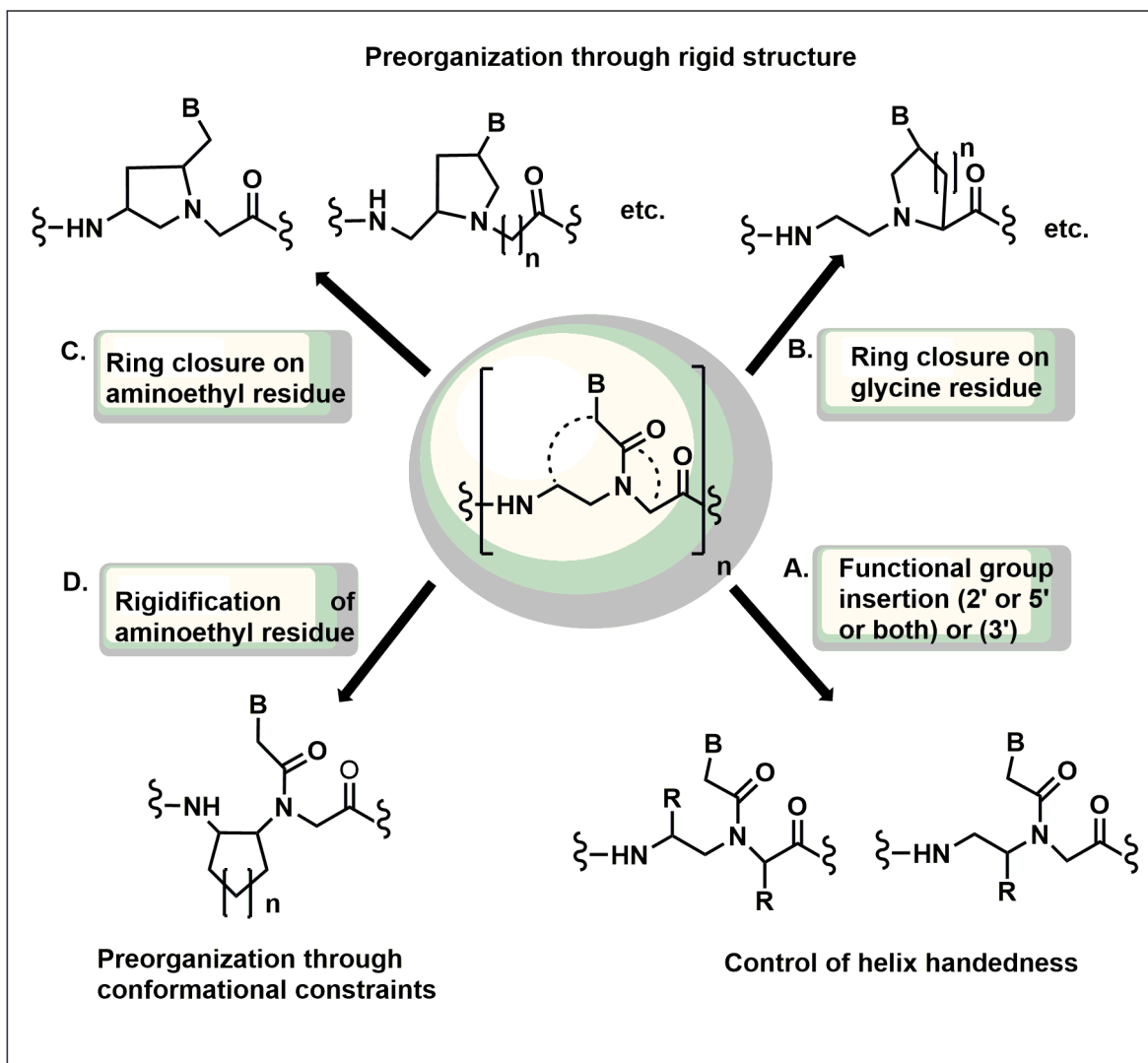C. cyclization of the PNA backbone

Figure 1.14 Strategies for inducing preorganization in monomer of PNA

A. Control of helix handedness

Preorganization of the PNA backbone has been endeavored by insertion of different functional groups at 2' or 5' position or both or at 3' position.[59] These functional groups included hydrophobic, hydrophilic or charged at α, β, γ position. Modifications like *N*-(2-aminoethyl)-β-Alanine,[60] (**Figure 1.15A**) *N*-(3-aminopropyl)-Glycine[61] (**Figure 1.15B**) and eth-

ylene carbonyl linked nucleobases[62](**Figure 1.15C**) were reported to improve the sequence specificity for directional preference. Nevertheless, these subtle alterations proved detrimental to UV-melting temperature. Incorporation of chiral monomers i.e. L/D- alanine[63] retained the hybridization properties though less efficiently with D-alanine being slightly better in comparison to L-alanine (**Figure 1.15D**). Among various other modifications reported, D-lysine (**Figure 1.15E**) exhibited thermal stability as good as that of PNA. The incorporation of chirality on the backbone did induce sequence selectivity of PNA oligomers in hybridization maximizing the effect for D-glutamic acid and D-lysine substitutions. In general, the different substituents caused equal or lower destabilization of PNA: RNA hybrids as compared to PNA: DNA hybrids. Reductive amination was the most used procedure to obtain an α-chiral backbone of PNA from amino acids.[64]

However, on substituting the β position of PNA monomer with alkyl groups[65] having *S*-configuration (derived from L-alanine) (**Figure 1.15F**) showed hybridization with complementary DNA while *R*-isomer (derived from D-alanine) did not. Thus, the stereochemistry of the β-carbon of the PNA backbone was critical to the hybridization ability of PNA and strictly limited to *S*-configuration. This contrasts α-PNAs in which stereochemistry of the α-carbon arising from a methyl group hardly affected the DNA binding ability.
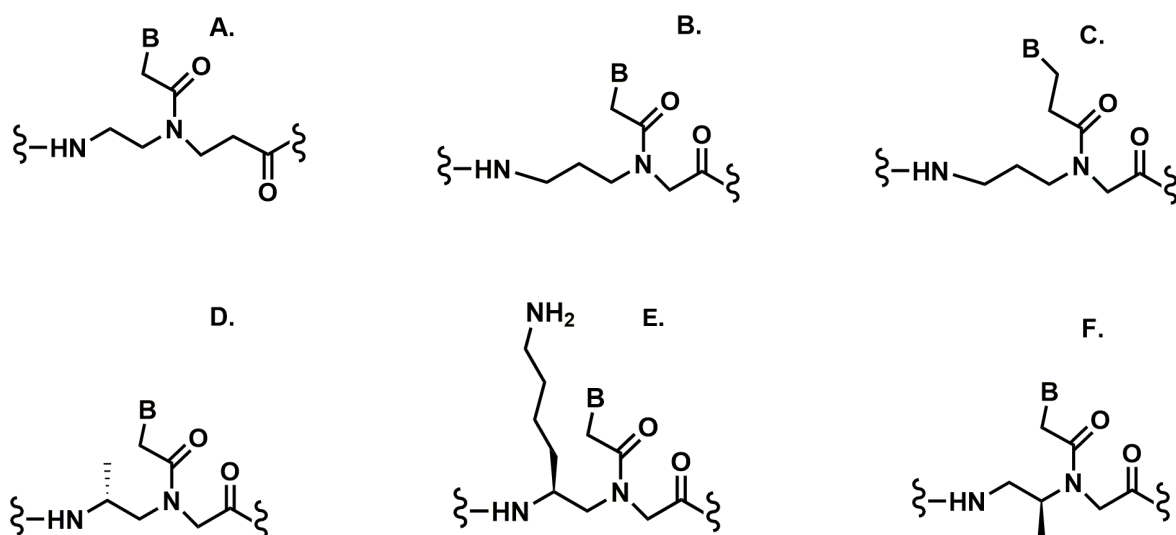
Figure 1.15: PNA modifications to achieve preorganization through control of helix handedness

B. Ring closure on glycine residue

Some of the relatively successful conformationally preorganized modifications[66-68] so far are based on introduction of methylene/ethylene groups to bridge the methylene carbonyl side chain to generate diverse five or six membered nitrogen heterocyclic analogues. The cyclic analogues where the nucleobases are directly attached to the ring have defined nucleobase orientation, overcoming the rotamer problem. It also collateraly introduces chiral centres introducing directional selective binding of PNA with complementary DNA/RNA. Ring closure on glycine residue with five membered ring generated aminoethylprolyl PNA (*aep*PNA),[69](**Figure 1.16A**) capable of forming highly stable triplexes. The aminoethylprolyl-5-one thymine monomers[70] (**Figure 1.16B**) were synthesized and incorporated in *aeg*PNA-T8 backbone at different positions. The *aepone*PNAs showed marvellous hybridization of PNA to DNA triplexes compared to *aeg*PNA. Six membered ring on the glycine residue namely (2*S*, 5*R*) aminoethyl pipecolic PNA (**Figure 1.16**C) was also reported[71] where aminoethyl pipecolic PNA was found to form stable DNA triplexes which is possible only in case of favorable pre-organization of PNA.
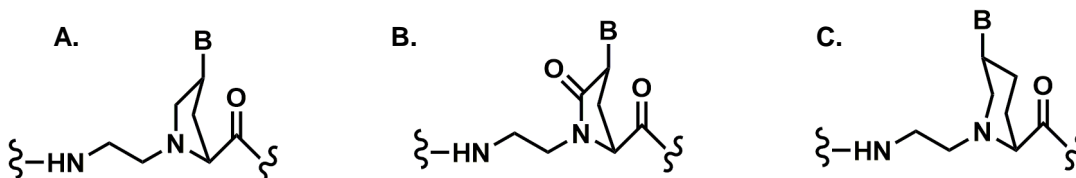
Figure 1.16: PNA modifications with ring closure on glycine residue

C. Ring closure on aminoethyl residue

Ring closure on aminoethyl residue can also impart preorganization of PNA backbone. Number of modifications involving ring closure of methylene carbonyl side chain has been reported till date. A wide variety of such chiral, constrained and structurally organized PNAs have been synthesized from naturally occurring *trans*-4-hydroxy-L-proline. An example of such modification is *N*-(thymin-1-yl-acetyl)-4-aminoproline, all possible monomers were synthesized and the L- *trans*-4-aminoprolyl isomer, was shown to bind to DNA with higher affinity, while the L-*cis* isomer and the D-*trans* which could not adopt

the same spatial arrangement and showed reduced binding preferences.[72] PNAs derived from piperidone[73] were also reported, though the rigidity of the ring and their particular geometry led to a decrease of PNA: DNA duplex stability. However, among all the isomers synthesized (3*R*, 6*R*)-isomer, which is in line with proper group arrangement emerged as the best modification.
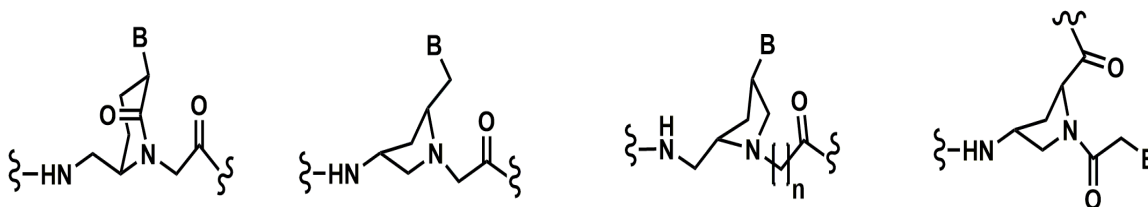

Figure 1.17: **PNA modifications with ring closure on aminoethyl residue**

D. Rigidification of aminoethyl residue

Suitable substitutions may also lead to generation of cyclic structures with 1, 2-cyclohexylamino,[74](**Figure 1.18A, B, C**) 1,2-cyclopentylamino[75](**Figure 1.18D, E**) and spirocyclohexyl[76] rings in monomers. The transcyclopentane (tcyp-PNA) modified PNA structure which is equivalent to freezing the aminoethyl glycyl backbone was also reported to provide high level of mismatch discrimination[75b] and was utilized as a target capture strand to improve detection limit of known DNA detection assay. On the other hand, *cis*-(1*R*, 2*S*) cyclopentyl PNA hybridized to DNA/RNA without much discrimination due to sugar puckering. Certain modifications like introduction of methylene/ethylene groups to bridge the aminoethyl glycyl backbone and methylene carbonyl side chain generated six or five membered ring, which introduced chiral centre as well as concomitantly directed the nucleobase solving the rotamer problem. *Trans* -4-hydroxy-L-proline is a versatile commercially available compound for the synthesis of different analogues of modified PNA viz *N*-(thymin-1-yl-acetyl)-aminoproline[77] all stereoisomers were synthesized. L- *trans* (**Figure 1.18F**) isomer showed enhanced binding affinity with DNA while L-*cis* and D- *trans* showed reduced hybridization performance.
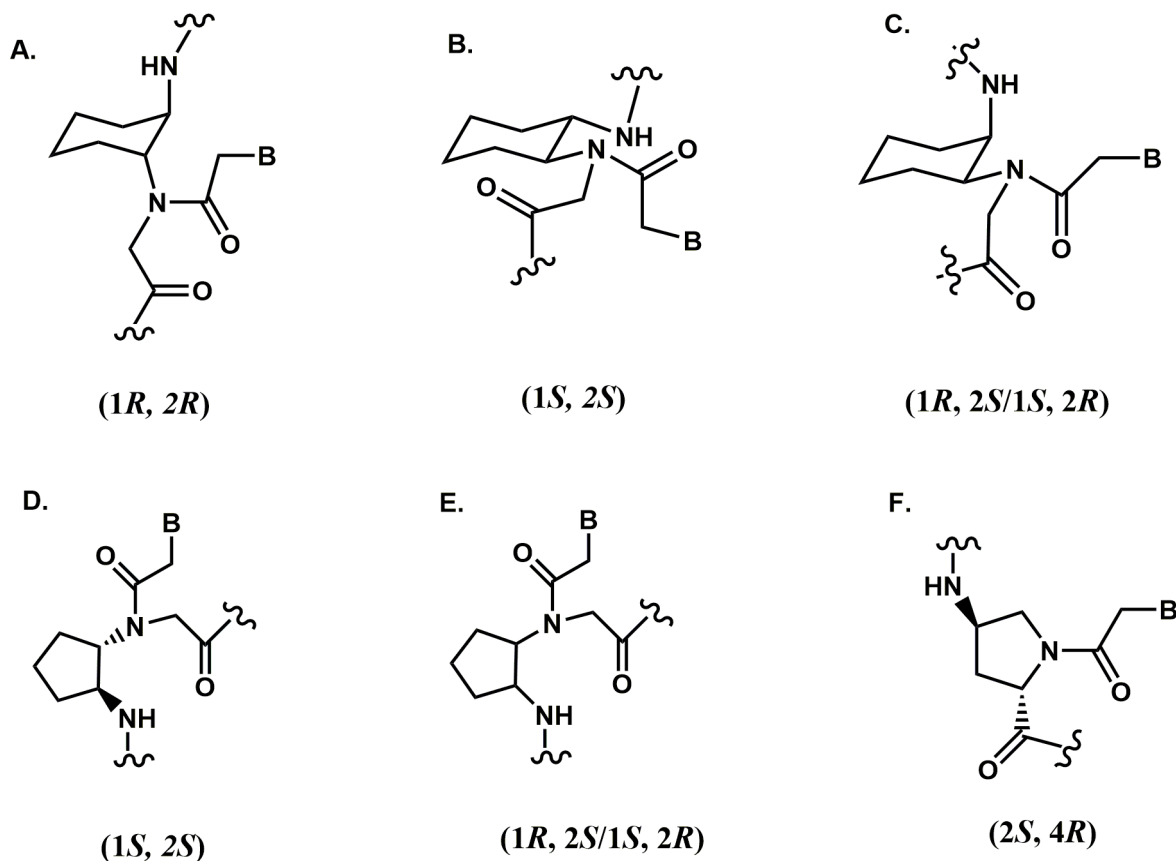
Figure 1.18: 1, 2-cyclohexylamino substituted Peptide Nucleic Acid A) (1*R*, 2*R*) B) (1*S*, 2*S*) C) (1*R*, 2*S* /1*S*, 2*R*); 1, 2-cyclohexylamino substituted Peptide Nucleic Acid D) (1*S*, 2*S*) E) (1*R*, 2*S*/1*S*, 2*R*); F) (2*S*, 4*R*)-amino-N^α-(substituted acetyl) - proline Peptide Nucleic Acid

1.5.2.2 Chemical modifications to enhance cellular uptake of PNA

PNA binds to both DNA and RNA with superior affinity. However, the biggest hurdle which does not allow the usage of PNA for therapeutic use is its poor aqueous solubility and hence poor cellular uptake and endosomal entrapment.[78] Current methods to enhance the cellular uptake of PNA are a) chimera of PNA and DNA b) conjugating cell-penetrating peptides (CPP)[79] c) introducing charged or polar groups on the backbone of PNA.

A. PNA DNA chimera

DNA comprises of a negatively charged phosphate group, which makes it highly soluble in water. Generating a chimera would be synergistic in terms of overpowering the

limitations of PNA as well as DNA. Some types of chimera reported are i)5'-DNA linker X-PNA- *pseudo*-3'[80] ii) *pseudo* 5'-DNA linker X-DNA- 3'[81] iii) *pseudo* 5'-PNA linker X-DNA-3'[76]Synthetic protocols have been developed with protecting groups compatible for carrying out on-line synthesis of both PNA and DNA to generate the chimeras. Several interesting properties were noticed in such covalent hybrids such as co-operative stabilizing effects against proteases and nucleases, enhanced water solubility and duplex/triplex stabilities dependent on the structure of chimerae and the linker.
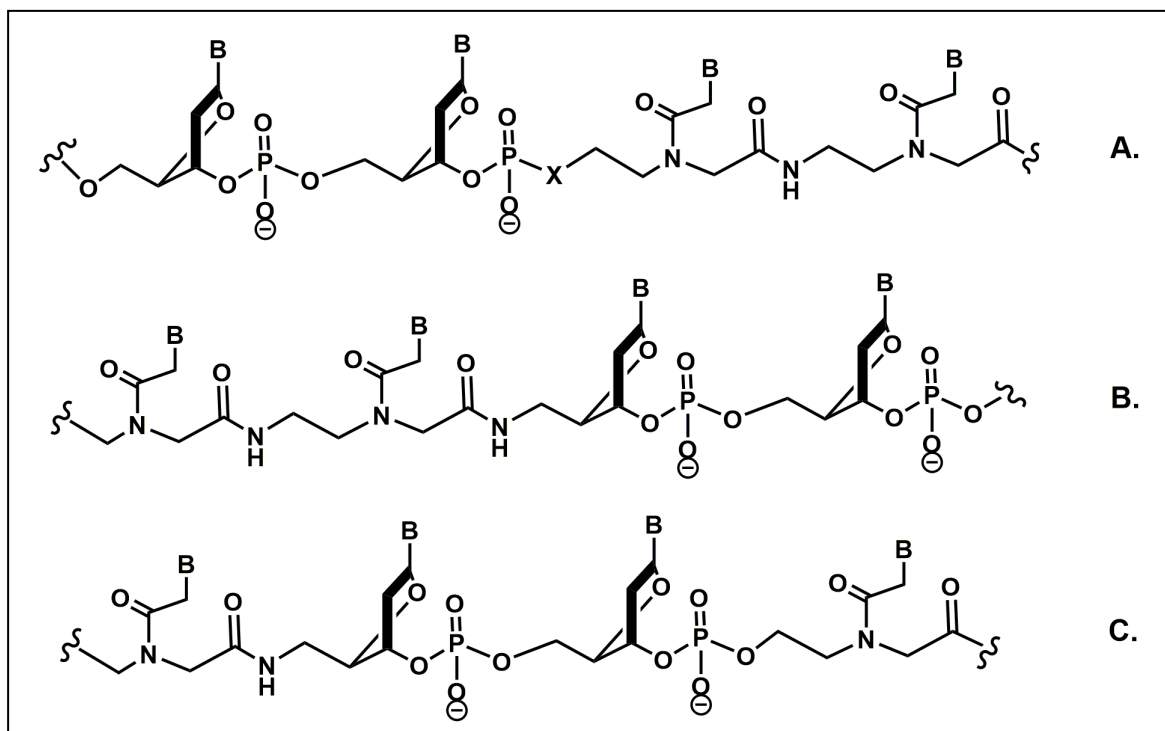
Figure 1.19: PNA DNA chimeras

B. Conjugation of PNA with cell penetrating peptides

Various efforts have been made to equip PNAs with a hydrophilic group that would confer cell permeation. PNAs having neutral backbone do not associate with delivery vehicles based on cationic lipids. Hence, to use lipofectamine (cationic lipid), PNA needs to be hybridized to complementary oligodeoxynucleotides that aids complexation with the lipid.[82] But the most popular approach to enhance cellular delivery has been conjugation of PNA to cell penetrating peptides that deliver the conjugate through endocytotic pathway.[83]

However, the low ability of PNA-CPP conjugates to escape from endosome has been the bottleneck of this approach. Various endosomolytic compounds have been explored, but unfortunately proved to be too toxic for *in vivo* applications.[79a] Conjugates with arginine-rich peptides have shown promising activity in HeLa cells in the absence of endosomolytic agents.[84] However, certain peptides are internalized very efficiently by cells.[85] The remark-able uptake properties of HIV-1 Tat transduction domain is the result of short basic se-quences of (GRKKRRQRRR). They are amphiphatic-helices with high content of basic amino acid residues (Lys, Arg). These peptides are conjugated to one or both the ends of PNA by an amide or disulfide linker.[86] However, CPPs are relatively large peptides which complicates the preparation of PNA-CPP conjugate. The groups Corey[87] and Gait[88] showed that conjugation of PNA with short oligolysine enabled efficient delivery in fibroplast and various cancer cell lines.
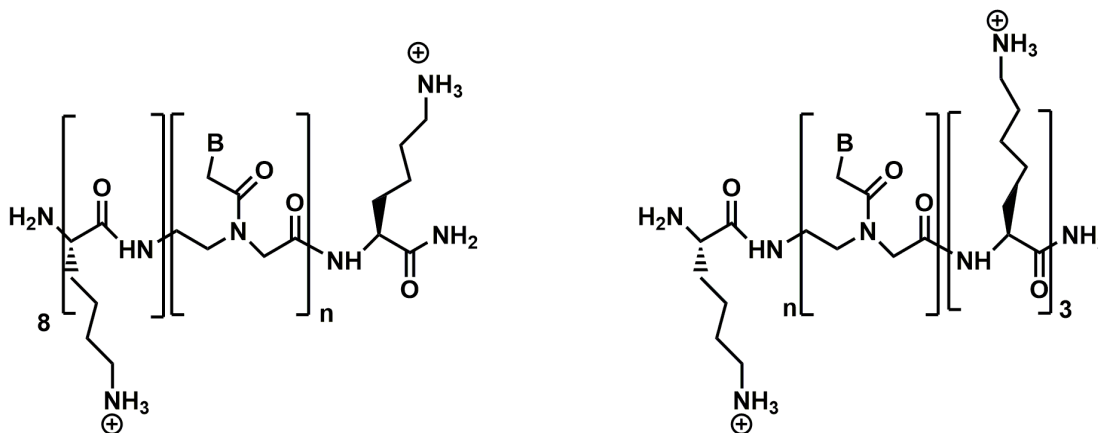
Figure 1.20: Conjugation of PNA with short oligolysine improves cellular uptake

As few as four lysine residues achieved similar efficiency as CPP previously optimized for cellular delivery of PNA.[80] Use of short oligolysine instead of CPP reduces the complexity and effort to deliver the PNA into the cells. Most recently, Gait and coworkers showed that introduction of a terminal Cys residue further increases the cellular uptake of Cys-Lys-PNA-Lys conjugate.[89]

C. Introduction of charged or polar groups on the backbone

Charged or polar functional groups can amplify the solubility of any organic molecule. This idea afforded fusion of positively charged lysine at the termini[90] or in the interior part of the oligomer, inclusion of polar groups in the backbone,[91] replacement of the original (aminoethyl) glycyl backbone skeleton with a negatively charged scaffold[92] or conjugation of biocompatible polyethylene glycol (PEG) to one of the termini[93] rendering solubility of the PNA molecule.[94] Incorporation of cationic residues can enhance hydrophilicity, on the other hand can also lead to nonspecific binding due to an increase in electrostatic interaction upon hybridization to DNA or RNA. Ly and coworkers synthesized guanidine groups at α-[95](**Figure 1.21A**) and γ-position [96](**Figure 1.21B**) of PNA's backbone. The α-guanidine-modified PNA, GPNA derived from D-arginine had higher affinity for complementary DNA[97] and RNA.[98] GPNA was readily assimilated by several cell lines (HCT116, human ES, and HeLa), which was associated with the cationic guanidine modifications. GPNA being less toxic to cells than a PNA-polyarginine inhibited E-cadherin in A549 cells.[99] Englund and Appella reported that the *S*-isomer of γ-modified PNA (derived from the natural L-lysine) (**Figure 1.21C**) could hybridize better with complementary DNA in comparison to the *R*-isomer.[100] Manicardi *et al*[101] used both α- and γ-modified GPNA 15-mers to inhibit microRNA-210 in K562 cells.[97] Mitra and Ganesh recorded identical observation on DNA binding and cellular uptake of α- and γ-aminomethylene PNA (*am*-PNA).[102, 103] The aminomethylene modification boosted PNA binding to DNA, with γ-(*S*)*am*- PNA (**Figure 1.21E**) being significantly superior than α-(*R*)*am*-PNA (**Figure 1.21D**), which, in turn, was better than α-(*S*)*am*-PNA (**Figure 1.21F**).[102]

Polar groups like –OH, -SH, -OSO$_3^{\ominus}$, *mini*PEG were introduced on the backbone of PNA to increase the solubility and pharmacokinetic properties of PNA. While, γ -thiomethyl PNA[104] (**Figure 1.21G**) were utilized for synthesizing long PNA sequences for intracellular activities by native chemical ligation, γ -hydroxymethyl PNA[105] (**Figure 1.21H**) reported by Pensato *et al* successfully invaded the duplex. PNA bearing a sulfate group[106] (**Figure 1.21I**) was reported by Romanelli *et al*, which targeted at making PNAs more like DNA in terms of polarity and charge resulting in destabilized duplex. This destabilization was rationalized by electrostatic repulsion between the negatively charged sulphate of the

modified PNA and phosphate of DNA. The modified PNA was successfully lipofected into human breast-cancer (SKB3) cells exhibited antigene activity against *ErbB2* gene. Minipeg group at the γ position of PNA reported by Sahu *et al*[107] (**Figure 1.21J**) was not only easy to synthesize but also boosted the solubility as well as hybridization properties of PNA.
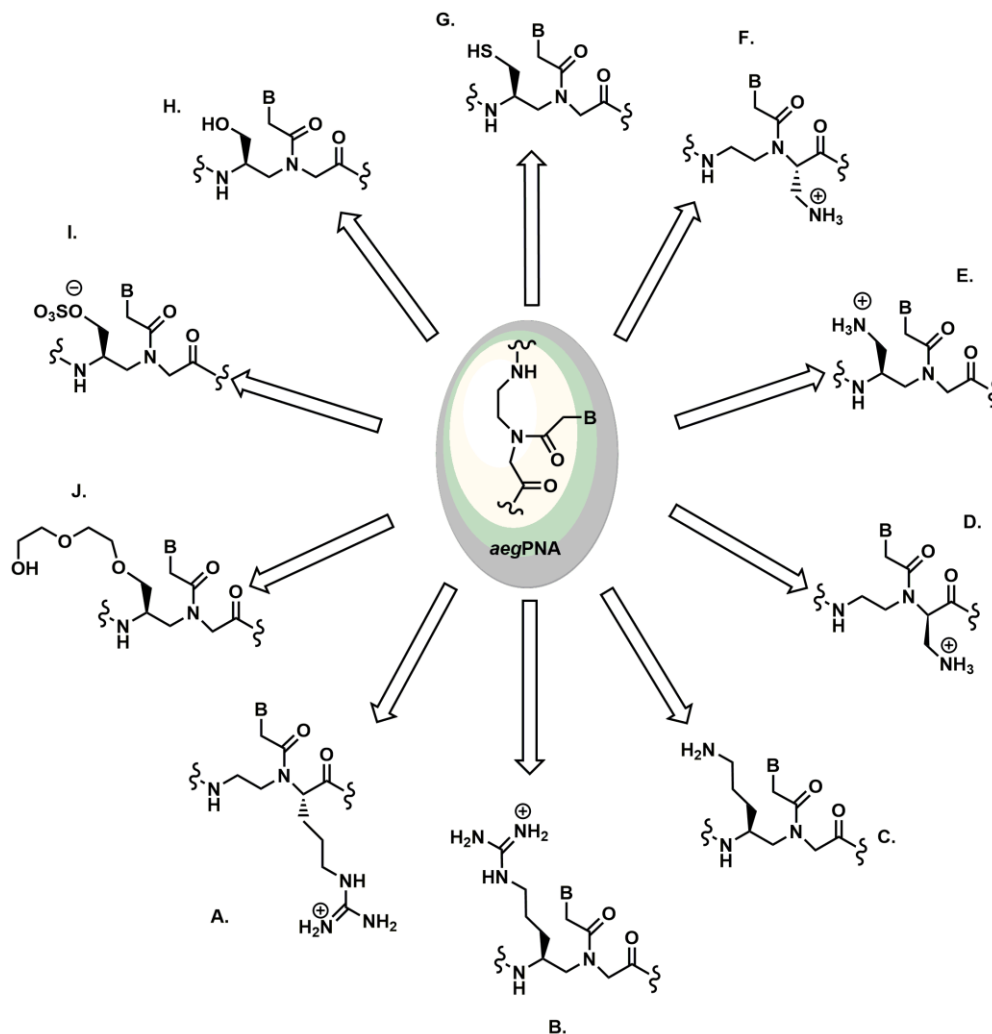

Figure 1.21: Modifications reported to enhance the solubility of PNA

1.5.2.3 Nucleobase modifications

Only a few 'non-Watson-Crick' nucleobases have been studied in 'PNA context' (**Figure 1.22**). However, one of them, pseudoisocytosine, which mimic *N*3-protonated cytosine, is indispensable for pH-independent recognition of DNA guanine in the Hoogsteen mode of triplex forming bis-PNAs,[108] and is routinely used in bis- PNAs for recognition of

double stranded DNA. Another useful substitution is to use diaminopurine in place of adenine as it improves the T_m of the PNA/nucleic acid duplex by up to 4°C per substitution.[109] 2-Aminopurine hydrogen bonds with U and T in reverse Watson-Crick mode and has the advantage of being inherently fluorescent to enable study of kinetic events associated in hybridization. Expanding the repertoire of bases recognized by Hoogsteen bonding beyond the purines, especially thymine, has been an extraordinary challenge because of the steric clash presented by the 5- methyl group projecting from the Hoogsteen face of the nucleobase. The base was rationally designed for recognition of A: T base pair in the major groove and form a stable triad with T in the central position. It demonstrated good mismatch discrimination and bound to thymine better than guanine did. In order to increase the stability of complexes formed by PNA with target nucleic acids, bases that possess a larger surface area (for greater hydrophobic/stacking interactions), make additional H-bonds or that are positively charged have been prepared. Wide varieties of 5-substituted uracils were synthesized and their ability for triplex formation has been studied. The G-clamp base was developed to build in specific, additional bonding interactions with guanine. Unnatural heterocyclic 3-nitropyrrole has been used as potential universal bases in PNA. The synthesis of cyanuryl PNA monomer containing cyanuric acid as the base was achieved by direct *N*-monoalkylation of cyanuric acid with *N*-(2-Boc-aminoethyl)- *N'*-(bromoacetyl)glycyl ethyl ester.[110] The monomer was incorporated as a T-mimic into PNA oligomers and biophysical studies on their triplexes/duplex complexes with complementary DNA oligomers indicated unusual stabilization of PNA: DNA hybrids when the cyanuryl unit was located in the middle of the PNA oligomer.
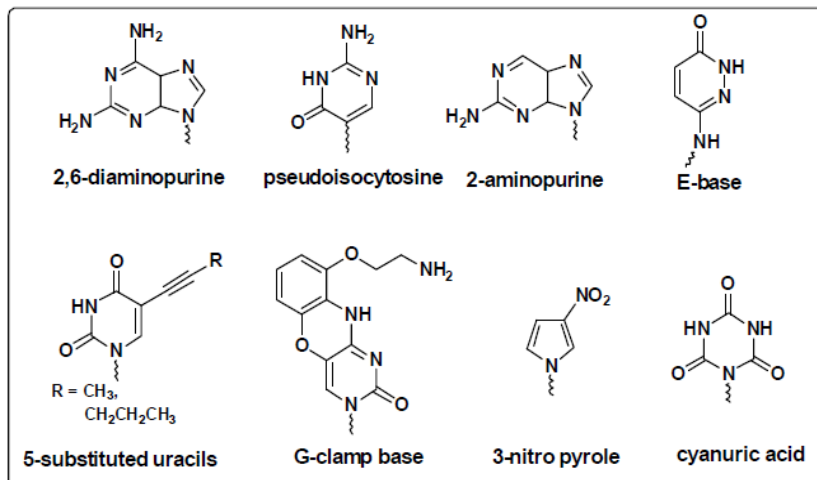
Figure 1.22: Non-standard nucleobases used in PNA oligomers

1.6 Quadruplex forming DNA and their application in therapeutics

The observation of polycrystalline gels by a medical doctor/clinical laboratory scientist Ivar Christian Bang (1869-1918)[111,112] had stirred up the scientific community.[113] Rigorous diffraction studies and UV melting experiments of the polyguanylic acid gels proved that stretches of contiguous guanines formed tetramers of planar hydrogen-bonded arrangements, called G-quadruplexes.
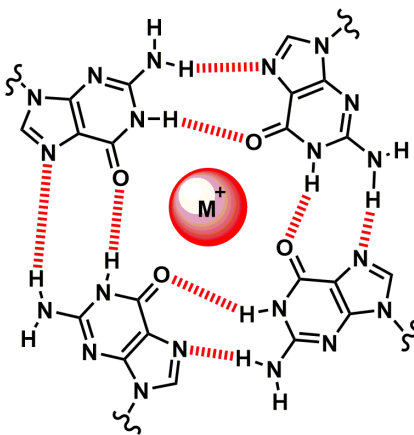
Figure1.23: Proposed arrangement of bases in G-quadruplex

The ability of DNA and RNA G-rich sequences to self-associate in forming novel higher ordered structures received much attention when they were identified to play a ma-

jor role in structural and regulatory maintenance of telomeres present at the end of human chromosomes.[114] Blackburn, Greider and Szostak, were awarded the Nobel prize in 2009 for their work on telomerase and telomere maintenance. G-quadruplexes are sprinkled throughout the human genome[115] and their ability in regulating gene expression has wide scope in the field of quadruplex cancer therapeutics.

The central void of G-quadruplexes can accommodate different cations of different radii, which determine the topology, as well as stability of quadruplexes (**Figure 1.23**). Potassium and sodium were found to be optimal in increasing the G-quadruplex stability. The G-stretches are held together by hydrogen-bonding between the Watson-Crick face of each guanine with the Hoogsten face of adjacent guanine creating a cyclic arrangement of four guanines. The tetrads are stacked in a right handed helical motif,[114] with a helical twist of 30° and a diameter of 25Å.

The phosphate backbone generates four grooves, which accommodate well-defined network of water molecules. Unlike duplex DNA/RNA, G-quadruplexes exist when the strands of DNA/RNA are parallel as well as antiparallel. The linkers between the G-runs forms three different types of loops (**Figure 1.24**) i) lateral loops, joining adjacent antiparallel phosphate backbone on the same quadruplex ii) diagonal loop, connecting the opposite antiparallel strands iii) propeller loop, fastening adjacent parallel strand.
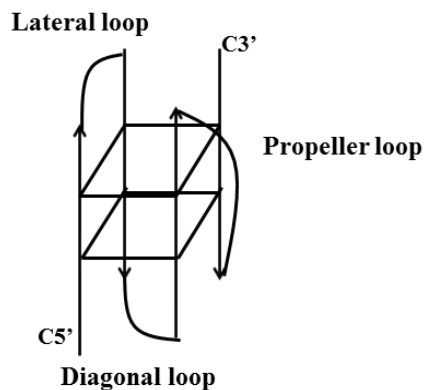
Figure 1.24: G-quadruplex linker orientation

This linker length is instrumental in defining the final topology and stability of the quadruplex, and also determines if the resulting quadruplex will be intra or inter molecular.[116]

Balasubramanian *et al* [117] designed G-rich DNA libraries and used UV-T_m and CD to compare the properties of 21 DNA quadruplex libraries.

1.6.1 Types of quadruplex structures

G-quadruplexes have been reported till date to form more than 100 types of structures. This variety arises from introduction of different cations, salt concentration, loop length, nucleobases in the loop, strand concentration, strand polarity etc. [118] Some of the extensively found varieties of G- quadruplexes are depicted below.
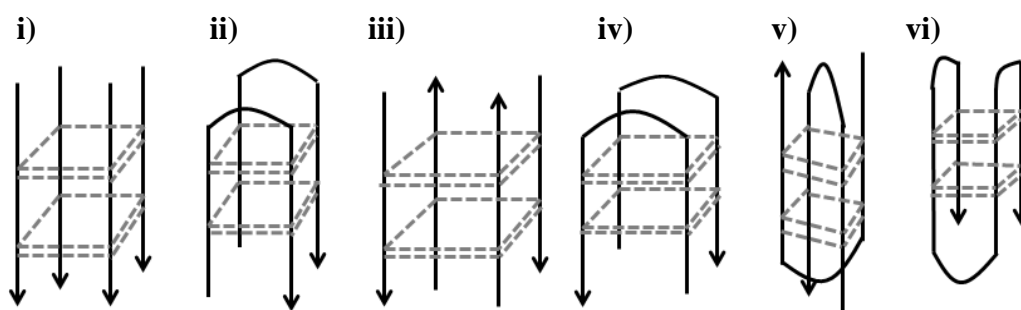
Figure 1.25: Structure of i) tetramolecular parallel quadruplex, ii) bimolecular antiparallel quadruplex with adjacent parallel Strands, iii)tetramolecular antiparallel quadruplex with adjacent antiparallel strands, iv) bimolecular antiparallel quadruplex with adjacent parallel Strands v) Basket type intermolecular quadruplex vi) chair type intramolecular quadruplex

1.6.2 G-quadruplex in aptamers

Aptamers are single-stranded DNA or RNA, peptides or modified nucleic acids proficient in binding to specific targets with selectivity, specificity and affinity folding into specific three-dimensional structures. In 1990 Joyce,[119] Szostak,[120] and Gold[121] independently reported on the development of an in vitro selection and amplification technique for the isolation of oligonucleotides able to bind non-nucleic acid targets with high affinity and specificity. Several of these aptamers were found to be G-rich with an ability to form G-quadruplex structures. The enhanced resistance to nuclease degradation and increased cellular uptake are advantageous in the development of G-quadruplexes as drug candidates.[122] Number of such aptamers have been designed successfully viz HIV protein targets,[123] bovine prion proteins,[124] anticancer targets such as nucleolin,[125] Thrombin Binding

Aptamer (TBA)[126] targeting human α-Thrombin amid which Thrombin Binding Aptamer has been studied widely.

1.7 Tools and techniques for structural studies of nucleic acid complexes

1.7.1 UV-Visible spectroscopy

The absorbance of polynucleotide is the resultant of the sum of absorbances of the individual nucleotides and interaction between them. The interaction between the bases results in decrease of absorbance due to coupling of transition dipoles between neighboring stacked bases[127] and hydrogen bonds between base pairs hindering the involvement of nucleobases in hydrogen bonding.[128] Increasing temperature perturbs this system, inducing a structural transition by causing disruption of hydrogen bonds between the base pairs, diminished stacking between adjacent nucleobases and larger torsional motions in the backbone leading to a loss of secondary and tertiary structure. This is evidenced by an increase in the UV absorption at 260nm, termed as hyperchromicity. The DNA melting can be readily monitored by measuring its absorbance at a wavelength of 260nm. A plot of absorbance *vs.* temperature gives a sigmoidal curve in case of duplexes/triplexes and the midpoint of the transition gives the T_m or melting temperature.

A. Duplex melting

The UV absorbance value at any given temperature is the average of the duplex and single strands according to the "all-or -none" model. Thus, the melting temperature of duplex is a state when half of the strands are in single strand form and half bound to its complimentary DNA. The hyperchromicity observed in duplex melting is about 10-20%.

B. Triplex melting

DNA triplex melting monitored at 260nm wavelength gives a double sigmoidal curve. The first dissociation leads to the triplex (Hoogsteen strand) transition to duplex and the second corresponds to duplex dissociation (Watson-Crick duplex) at higher temperature

into two single strands.[129] The hyperchromicity observed in triplex melting is about 30-35%.

C. Quadruplex melting

Quadruplex meltings can also be monitored at 260nm. G-quartet dissociation records smaller variation (4%) of absorbance. However, the stability of the modified G-quadruplexes can be followed by the change in the UV absorbance at 295nm with increase in temperature.[130] Maximal hyper/hypo chromic shift is observed at this wavelength between folded and unfolded states.

A non-sigmoidal transition with low hyperchromicity is a consequence of non-complexation (non-complementation). In many cases, the transitions are broad and the exact melting temperature is obtained from the peak in the first derivative plots. This technique has provided valuable information regarding complementary in nucleic acid hybrids involving DNA, RNA and PNA.

D. Stoichiometry of binding

The binding stoichiometry of oligonucleotides with complementary DNA/RNA can be determined by UV-titration monitored at 260nm known as Job's plot.[131] The oligonucleotides and DNA are mixed in different molar ratios of 0:100, 10:90, 20:80, 30:70, 40:60, 50:50, 60:40, 70:30, 80:20, 90:10, 100:0 and their absorbances are recorded at the particular molar ratio. Increase in complexation causes reduction in absorbance, reaching minima, beyond which the absorbance keeps on increasing. This inflection point helps in determining the stoichiometry of binding. For example- inflection at 50:50 indicates 1:1 complex while 70:30 indicates 2:1 complex.

1.7.2 Circular Dichroism

Circular Dichroism is a spectroscopic technique used for elucidating the secondary structure of chiral molecules and biomolecules like proteins and nucleic acids.[132] It helps in understanding the structural, kinetic and thermodynamic properties of such macromolecules. Measurements carried out in the visible and ultra-violet region of the elec-

tro-magnetic spectrum monitors electronic transitions, of molecule comprising of chiral chromophores, giving rise to non-zero CD signal. CD is measured in millideg indicated by θ. The ellipticity θ is usually converted to the molar ellipticity for comparison with other data. This is represented by symbol [θ] (with units of degrees.cm squared.per deci-mole).Thus, [θ] = θ / (100 x *c* x *l)* where c is the Molar concentration of the sample (mole/L) and *l* is the pathlength in cm, factor 100 converts the pathlength to metres.

In the nucleic acids, the heterocycle bases are principal chromophores. As these bases are planar, they donot have any intrinsic CD. CD arises from the asymmetry induced by linked sugar group. when the bases are linked together in polynucleotides, they give rise to many degenerate interactions and gain additional characteristics associated with asymmetric feature of secondary structure of nucleic acids.

1.7.3 Fluorescence Spectroscopy

Nucleic acids can be studied by tagging them with fluorophores, which enables them in studying with fluorescence spectroscopy. They can be studied for their delivery in cells, intracellular activity. However, G-rich quadruplexes exhibit significantly higher intrinsic fluorescence at room temperature in comparison to homologous sequence with Guanine. [133, 134] Hence, labeling of such sequence is not required for studying G-quadruplex sequence.

1.7.4 Polyacrylamide Gel Electrophoresis

Polyacrylamide gel electrophoresis (PAGE), describes a technique widely used in biochemistry, forensics, genetics, molecular biology and biotechnology to separate biological macromolecules, usually proteins or nucleic acids, according to their electrophoretic mobility. Mobility is a function of the length, conformation and charge of the molecule. Native PAGE can be used to differentiate molecules according to the their size and shape. Thus, duplexes, triplexes and quadruplexes can be differentiated from their single strand on polyacrylamide gel. Duplexes having higher molecular weight are retained on gel in comparison to the same single strand. G-quadruplexes are retained on polyacrylamide gel due

to its folded structure in comparison to its unfolded structure. Further association of G-quadruplex to dimer or higher multimer can be demonstrated with PAGE.[135]

1.7.5 Nuclear Magnetic Resonance Spectroscopy

NMR spectroscopy is a powerful experimental technique that is useful for confirmation of the formation and topology of complexes of DNA and RNA such as duplexes, triplexes and quadruplexes. It can also be used to study hydrogen bonding between nucleobases. The information obtainable includes sugar pucker characterization, backbone conformations and also local architecture. In case of G-tetrad formation, 1D proton NMR spectra provides characteristic imino, amino and aromatic peaks indicating the presence of Hoogsten and Watson-Crick base-pairing. Further structure elucidation, of course, necessitates the use of more specialized techniques such as NOESY, TOCSY, HSQC-COSY etc. This technique has clearly dominated the structure elucidation research of quadruplex folding and topology.

1.8 Present Work

Chapter 2: **Synthesis of polycarbamate analogue of GNA/TNA and their biophysical Studies**

This chapter is divided into two sections:

Section A: Synthesis of *R/S*-GCNA monomers and oligomers

The synthesis of polycarbamate analogues of GNA/TNA is described in this section. Oligomers containing these modified units were then synthesized on solid phase, purified by HPLC and characterized by MALDI-TOF.

Section B: Biophysical evaluation of *R/S*-GCNA carbamate oligomers

Biophysical experiments were performed to study the effects of these modifications on the stability of complexes with target DNA/RNA.

Chapter 3: **β,γ-bis-substituted PNA with configurational and conformational switch: binding studies with cDNA/RNA and their cellular uptake**

This chapter is divided into three sections:

Section A: Synthesis of (*R*, *R*)/(*S*, *S*) β,γ-bis- hydroxymethyl /methoxymethyl *aeg*PNA

This section describes the rationale of designing and synthesis of (*R, R/S, S*) β,γ-bis-hydroxymethyl /methoxymethyl substituted *aeg*PNA from (D/L) tartaric acid respectively. Oligomers with modified monomers incorporated at desired position were then synthesized on solid phase. The oligomers were purified by HPLC and characterized by MALDI-TOF TOF mass spectrometry.

Section B: Biophysical evaluation of (*R*, *R*) /(*S*, *S*) β,γ-bis- hydroxymethyl /methoxym-ethyl substituted *aeg*PNA incorporated oligomers

This section describes the biophysical evaluation of modified oligomers. The effect of such modification on binding with target DNA and RNA is studied using temperature dependent UV-T_m experiments. Electrophoretic Gel mobility assay was also performed to discriminate between the binding of β,γ-bis-hydroxymethyl /methoxymethyl substituted PNA with cDNA.

Section C: Cell uptake experiments

This section illustrates the ability of the sequences chosen after biophysical evaluation for cell uptake experiments and the results were compared with control PNA.

Chapter 4: **Structural and functional evaluation of PNA and modified PNA as G quadruplex structure**

This part of the work elucidates the synthesis of PNA G-quadruplex and peptide modification of the G-quadruplex sequence. The synthesized oligomers were meticulously studied for their structural properties.

1.9 References

1. F. Crick, *Nature*, 1970, **227**, 561.

2. J. D. Watson and F. H. C. Crick, *Nature*, 1953, **171**, 737.

3. K. Hoogsten, *Acta. Cryst.*, 1963, **65**, 907.

4. F. H. C. Crick, *J. Mol. Biol.*, 1966, **19**, 548.

5. M. J. Sundaralingam,. *Biopolymers*, 1969, **7**, 821.

6. a) W. Saenger, Principles of nucleic acid structure. Springer-Verlag, New York, 1984.

   b) E. Lescrinier, M. Froeyen and P. Herdewijn, *Nucleic Acid Res.*, 2003, **31**, 2975.

7. G. Felsenfeld, D. R. Davies and A. Rich, *J. Am. Chem. Soc.*, 1957**, 79,** 2023**.**

8. T. Simonsson, *Biol. Chem.*, 2001, **382**, 621.

9. B. F. Eichman, M. Oritz-Lombardia, J. Aymami, M. Coll and P. S. Ho, *J. Mol. Bio.*, 2002, **320**, 5, 1037. b) R. Holliday, *Genet. Res.*, 1964, **5**, 282.

10. G. N. Parkinson, M. P. Lee and S. Neidle, *Nature*, 2002, **417**, 876.

11. J. Weil, T. Min, C. Wang, C. Sutherland, N. Sinha and C. Kang, *Acta Crystallogr. D Biol. Crystallogr.*, 1999, **55**, 422.

12. a) B.Weiss, *CRC Press, Boca Raton*, FL, 1997 (ed.). b) C. Stein, *Nat. Med.*, 1995, **1**, 1119.

13. [13] P. C. Zamecnik and M. L. Stephenson, *Proc. Natl. Acad. Sci. U.S.A.*, 1978, **75**, 280.

14. C. F. Bennett and L .M. Cowsert, *Biochim. Biophys. Acta*, 1999, **1489**, 19.

15. a) J. Kurreck, *Eur. J. Biochem.*, 2003, **270**, 1628. b) A. Dallas, A. V. Alexander, *Med Sci Monit*, 2006, **12**, 67. c) J. Kurreck, *Angew. Chem. Int. Ed.*, 2009, **48**, 1378.

16. K. J. Scanlon, *Curr Pharm Biotechnol.*, 2004, **5**, 415.

17. C. Matranga, Y. Tomari, C. Shin, D. P. Bartel and P. D. Zamore, *Cell*, 2005, **123**, 607.

18. J. J. Song, S. K. Smith, G. J. Hannon and L. Joshua-Tor, *Science*, 2004, **305**, 1434.

19. R. C. Lee, R. L. Feinbaum and V. Ambros, *Cell*, 1993, **75**, 843.

20. V. N. Kim, *Nat. Rev. Mol. Cell Biol.*, 2005, **6**, 376.

21. H. Iguchi and T. Ochiya, *Gan To Kagaku Ryoho.*, 2010, **37**, 389.

22. a) Z. Dominski and R. Kole, 1993, **90**, 8673. b) P. Sazani and R. Kole, *J. Clin. Invest.*, 2003, **112**, 481.

23. V. K. Sharma, R. K. Sharma and S. K. Singh, *Med. Chem. Commun.*, 2014, **5**, 1454.

24. a) F. Eckstein, *Antisense Nucleic Acids Drug Dev.*, 2000, **10**, 117. b) E. De Clercq, F. Eckstein and T. C. Merigan, *Science*, 1969, **165**, 1137.

25. M. Matsukura, K. Shinozuka, G. Zon, H. Mitsuya, M. Reitz, J. S. Cohen and S. Broder, *Proc. Natl Acad. Sci. U.S.A.*, 1987, **84**, 7706.

26. W. Guschlbauer and K. Jankowski, *Nucleic Acids Res.*, 1980, **8**, 142.

27. H. Inoue, Y. Hayase, A. Imura, S. Iwai, K. Miura and E. Ohtsuka, *Nucleic Acids Res.*, 1987, **15**, 6131.

28. P. Martin, *Helv. Chim. Acta*, 1995, **78**, 486.

29. A. M. Kawasaki, M. D. Casper, S. M. Freier, E. A. Lesnik, M. C. Zounes, L. L. Cummins, C. Gonzalez and P. D. Cook, *J. Med. Chem.*, 1993, **36**, 831.

30. E. Merki, M. J. Graham and A. E. Mullick, *Circulation*, 2008, **118**, 743.

31. a)J. S. Jespen, M. D. Sorensen and J.wengel, *Oligonucleotides*, 2004, **14**, 130. b) C. F. Bennet and E. E. Swayze, *Annu. Rev. Pharmacol. Toxicol.*, 2010, **50**, 259.

32. C. Hendrix, H. Rosemeyer, I. Verheggen, F. Seela, A. V. Aerschot and P. Herdewijn, *Chem. Eur. J.*, 1997, **3**, 110.

33. a)S. Obika, D. Nanbu, Y. Hari, J. I. Andoh, K. I. Morio, T. Doi and T. Imanishi, *Tetrahedron Lett.*, 1998, **39**, 5401. b) S. Obika, D. Nanbu, Y. Hari, K. I. Morio, Y. In, T. Ishida and T. Imanishi, *Tetrahedron Lett.*, 1997, **38**, 8735.

34. K. U. Schoning, P. Scholz, S. Guntha, R. Krishnamurthy and A. Eschenmoser, *Science*, 2000, **290**, 1347.

35. H. Yu, S. Zhang and J. C. Chaput, *Nature Chemistry*, 2012, **4**, 183.

36. L. Zhang, A. Peritz and E. Meggers, *J. Am. Chem. Soc.*, 2005, **127**, 4174.

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

38. J. Wang, B. Verbeure, I. Luyten, E. Lescrinier, M. Froeyen, C. Hendrix, H. Rosemeyer, F. Seela, A. van Aerschot and P. Herdewijn, *J. Am. Chem. Soc.*, 2000, **122**, 8595.

39. M. J. Damha, C. J. Wilds, A. Noronha, I. Bruckner, G. Borkow, D. Arion and M. A. Parniak, *J. Am. Chem. Soc.*, 1998, **120**, 12976.

40. S. T. Crooke and B. Lebleu, *Antisense Research and Applications*, 1993.

41. J. Summerton and D. Weller, *Antisense Nucleic Acid Drug Dev*, 1997, **7**, 187.

42. M. J. Gait, A. S. Jones and R. T. Walker, *J. Chem. Soc. Perkin Trans.*, 1974, **1**, 1684.

43. M. Egholm, O. Buchardt, P. E. Nielsen and R. H. Berg, *J. Am. Chem. Soc.*, 1992, **114**, 1895.

44. a) S. C. Brown, S. A. Thomson, J. M. Veal and D. G. Davis, *Science*, 1994, **265**, 777. b) M. Eriksson and P. E. Nielsen, *Nat. Struct. Biol.*, 1996, **3**, 410. c) H. Rasmussen, J. S. Kastrup, J. N. Nielsen, J. M. Nielsen and P. E.Nielsen, *Nat. Struct. Biol.*, 1997, **4**, 98.

45. P. E. Nielsen and M. Egholm, *Current Issues Molec. Biol.*, 1999, **1**, 89.

46. D. J. Rose, *Anal. Chem.*, 1993, **65**, 3545.

47. V. V. Demidov, V. N. Potaman, M. D. Frank-Kamenetskii, M. Egholm, O. Buchard, S. H. Sonnichsen and P. E. Nielsen, *Biochem. Pharmacol.*, 1994, **48**, 1310.

48. a) D. Praseuth, M. Grigoriev, A. L. Guieysse, L. L. Pritchard, A. Harel-Bellan, P. E. Nielsen and C. Helene, *Biochim. Biophys. Acta*, 1996, **1309**, 226. b) H. Knudsen and P. E. Nielsen, *Nucleic Acids Res.*, 1996, **24**, 494. c) L. Good and P. E. Nielsen, *Nature Biotechnol.*, 1998, **16**, 355. d) R. W. Taylor, P. F. Chinnery, D. M. Turnbull and R. N. Lightowlers, *Nature Genet.*, 1997, **15**, 212.

49. J. Wang, *Curr. Issues Mol. Biol.*, 1999, **1**, 117.

50. D. Praseuth, M. Grigoriev, A. L. Guieysse, L. L. Pritchard, A. Harel-Bellan, P. E. Nielsen and C. Helene, *Biochim. Biophys. Acta*, 1996, **1309**, 226.

51. H. Kuhn, V. V. Demidov, B. D. Gildea, M. J. Fiandaca, J. M. Coull, M. D. Frank-Kamenetskii, *Antisense Nucleic Acid Drug Dev.*, 2001, **11**, 265.

52. D. B. Demer, E. T. Curry, M. Egholm and A.C. Sozer, *Nucleic Acid Res.*, 1995, **23**, 15, 3050.

53. a) M. Egholm, O. Buchardt, L. Christensen, C. Behrens, S. M. Freier and D. A. Driver, *Nature*, 1993, **365**, 566. b) K. K. Jensen, H. Orum, P. E. Nielsen and B. Norden, *Biochemistry,* 1997, **36**, 5072. c) P. Wittung, P. E. Nielsen, O. Buchardt, M. Egholm and B. Norden, *Nature*, 1994, **368**, 561.

54. S. Tomac, M. Sarkar, T. Ratilainen, P. Wittung, P. E. Nielsen, B. Norden and A. Grae-slund, *J. Am. Chem. Soc.*, 1996, **118**, 5544.

55. a) P. E. Nielsen, *Pure Appl. Chem.*, 1998, **70**, 105. b) B. Hyrup and P. E. Nielsen, *Bioorg. Med. Chem.*, 1996, **4**, 5.

56. V. V. Demidov and M. D. Frank-Kamenetskii, *Trends Biochem. Sci.*, 2004, **29**, 62.

57. L. Betts, J. A. Josey, J. M. Veal and S. R. Jordan, *Science*, 1995, **270**, 1838.

58. V. V. Demidov, V. N. Potaman, M. D. Frank-Kamenetskii, M. Egholm, O. Buchardt and S. H. Sonnichsen, *Biochem. Pharmacol.*, 1994, **48**, 1310.

59. T. Sugiyama and A. Kittaka**, *Molecules,* 2013, **18**, 287.

60. B. Hyrup, M. Egholm, M. Rolland, P. E. Nielsen, R. H. Berg and O. Buchardt, *J. Chem. Soc. Chem. Commun.*, 1993, **5**, 518.

61. B. Hyrup, M. Egholm, P. E. Nielsen, P. Wittung, B. Norden and O. Buchardt, *J. Am. Chem. Soc.*, 1994, **116**, 7964.

62. K. L. Dueholm and P. E. Nielsen, *New J. Chem.*, 1997, **21**, 19.

63. K. L. Dueholm, K.H. Petersen, D. K. Jensen, M. Egholm, P. E. Nielsen and O. Bu-chardt, *Bioorg. Med. Chem. Lett.*, 1994, **4**, 1077.

64. a) G. Haaima, A. Lohse, O. Buchardt, and P. E. Nielsen, *Angew. Chem. Int. Ed. Engl.* 1996, **35**, 1939. b) A. Puschl, S. Sforza, G. Haaima, O. Dahl and P. E.Nielsen, *Tetra-hedron Lett.*, 1998, **39**, 4707. c) P. Gupta, O. Muse and E. Rozners, *Biochemistry*, 2012, **51**, 63.

65. T. Sugiyama, Y. Imamura, Y. Demizu, M. Kurihara, M. Takano and A. Kittaka, *Bioorg. Med.Chem. Lett*., 2011, **21**, 7317.

66. V. A. Kumar, *Eur. J. Org. Chem.*, 2002, **71**, 2021.

67. V. A. Kumar and K. N. Ganesh, *Accounts Chem. Res*., 2005, **38**, 404.

68. V. A. Kumar and K. N. Ganesh, *Curr. Top. Med. Chem.*, 2007, **7**, 715.

69. M. D'Costa, V. A. Kumar and K. N. Ganesh, *J. Org. Chem.*, 2003, **68**, 4439.

70. N. K. Sharma and K.N. Ganesh, *Tetrahedron Lett.*, 2004, **45**, 1403.

71. P. S. Shirude, V. A. Kumar and K.N. Ganesh, *Tetrahedron Lett.*, 2004, **45**, 3085.

72. B. P. Gangamani, V. A. Kumar and K. N. Ganesh, *Tetrahedron*, 1999, **55**, 177.

73. A. Puschl, T. Boesen, T. Tedeschi, O. Dahl and P. E.Nielsen, *J. Chem. Soc.-Perkin Trans. 1*, 2001, 2757.

74. a) P. Lagriffoule, P. Wittung, M. Ericksson, D. K. Jensen, B. Norden, O. Buchardt and P. E. Nielsen, *Chem. –Eur. J.*, 1997, **3**, 912. b) T. Govindaraju, V. A. Kumar and K. N. Ganesh, *Chem. Commun.*, 2004, **23**, 860. c) T. Govindaraju, V. A. Kumar and K. N. Ganesh, *J. Am. Chem. Soc.*, 2005, **127**, 4144. d) T. Govindaraju, R. G. Gonnade, M. M. Bhadbhade, V. A. Kumar and K. N. Ganesh, *Org. Lett.*, 2003, **17**, 3013. e) T. Govindaraju, V. A. Kumar and K. N. Ganesh, *J. Org. Chem.*, 2004, **69**, 1858.

75. a) M. C. Myers, M. A. Witschi, N. V. Larionova, J. M. Frank, R. D. Haynes, T. Hara, A. Grajkowski and D. H. Appella, *Org. Lett.*, 2003, **5**, 2695. b) J. K. Pokorski, M. A. Witschi, B. L. Purnell and D. H. Appella, *J. Am. Chem. Soc.*, 2004, **126**, 15067.

76. a) A. H. Kortz, S. Larsen, O. Buchardt and P. E. Nielsen, *Bioorg. Med. Chem. Lett.*, 1998, **6**, 1983. b) W. Maison, I. Schlemminger, O. Westerhoff and J. Martens, *J. Bioorg. Med. Chem. Lett.*, 1999, **9**, 581.

77. B. P. Gangamani, V. A. Kumar and K. N. Ganesh, *Tetrahedron*, 1996, **52**, 15017.

78. P. E. Nielsen, *Q. Rev. Biophys.*, 2005, **38**, 345.

79. a) G. Cutrona, E. M. Carpaneto, M. Ulivi, S. Roncella, O. Landt, M. Ferrarini and L. C. Boffa, *Nat. Biotech.*, 2000, **18**, 300. b) U. Koppelhus, S. K. Awasthi, V. Zachar, H. Holst, P. Ebbesen and P. E. Nielsen, *Antisense Nucleic Acids Drug Dev.*, 2002, **12**, 51.

80. P.J. Finn, N. J. Gibson, R. Fallon, A. Hamilton and T. Brown, *Nucleic Acids Res.*, 1996, **24**, 3357.

81. F. Bergmann, W. Bannwarth and S. Tam, *Tetrahedron Lett.*, 1995, **36**, 6823.

82. D. A. Braasch and D.R. Corey, *Methods in Mol. Bio.*, 2002, **208**, 211.

83. a) T. Shiraishi and P. E. Nielsen, *Nature Protocols*, 2006, **1**, 633. b) F. S. Hassane, A. F. Saleh, R. Abes, M. J. Gait and B. Lebleu, *Cellular and Molecular Life Sciences*, 2010, **67**, 715.

84. S. Abes, J. J. Turner, G. D. Ivanova, D. Owen, D. Williams, A. Arzumanov, P. Clair, M. J. Gait and B. Lebleu, *Nucleic Acids Res.*, 2007, **35**, 4495.

85. a) D. Derossi, A. H. Joliot, G. Chassaing and A. Prochiantz, *J. Biol. Chem*, 1994, **269**, 10444. b) J. P. Richard, K. Melikov, E. Vives, C.Ramos, B. Verbeure, M. J. Gait, L. V. Chernomordik and B. Lebleu, *J. Biol. Chem.*, 2003, **278**, 585.

86. a) M. Pooga, U. Soomets, M. Hallbrink, A. Valkna, K. Saar, K. Rezaei, U. Kahl, J. Hao, X. Xu, Z. Wiesenfeld-Hallin, T. Hokfelt, T. Bartfai and U. Langel*, Nature Biotechnol.*, 1998, **16**, 857. b) M. Koning, G. Marel and M. Overhand, *Curr. Open. Chem. Biol.*, 2003, **7**, 734.

87. J. X. Hu, M. Matsui, K.T. Gagnon, J. C. Schwartz, S. Gabillet, K. Arar, J. Wu, I. Bezprozvanny and D. R. Corey, *Nat. Biotechnol.*, 2009, **27**. 478.

88. M. M. Fabani and M. J. Gait, *RNA-Publ. RNA Soc.*, 2008, **14**, 336.

89. A. G. Torres, M. M. Fabani, E. Vigorito, D. Williams, N. Al-Obaidi, F. Wojciechowski, R. H. E. Hudson, O. Seitz and M. J. Gait, *Nucleic Acids Res.*, 2012, **40**, 2152.

90. N. P. Boyarskaya, Y. G. Kirillova, E. A. Stotland, D. I. Prokhorov, E. N. Zvonkova, Shvets and V. I. Dokl, *Chem. (Transl. of Dokl. AkadNauk)*, 2006, **408**, 57.

91. a) A. Peyman, E. Uhlmann, K. Wagner, S. Augustin, G. Breipohl, D. W. Will, A. Schafer and H. Wallmeier, *Angew. Chem. Int. Ed. Engl.*, 1996, **35**, 2636. b)V. A. Efimov, M. V. Choob, A. A. Buryakova, A. L. Kalinkina and O.G. Chakhmakhcheva, *Nucleic acids Res.*, 1998, **26**, 566.

92. A. Cattani-Scholz, D. Pedone, F. Blobner, G. M. Abstreiter and L. Andruzzi, *Biomacromolecules*, 2009, **10**, 489.

93. K. H. Petersen, D. K. Jensen, M. Egholm, P. E. Nielsen and O. Buchardt, *Bioorg. Med. Chem. Lett.*, 1995, **5**, 1119.

94. B. Sahu, I. Sacui, S. Rapireddy, K.J. Zanotti, R. Bahal, B. A. Armitage and D. H. Ly, *J. Org. Chem*, 2011, **76**, 5614.

95. P. Zhou, M. Wang, L. Du, G. W. Fisher, A. Waggoner and D. H. Ly, *J. Am. Chem. Soc.*, 2003, **125**, 6878.

96. B. Sahu, V. Chenna, K. L. Lathrop, S. M. Thomas, G. Zon, K. J. Livak and D. H. Ly, *J. Org. Chem.*, 2009, **74**, 1509.

97. P. Zhou, A. Dragulescu-Andrasi, B. Bhattacharya, H. O'Keefe, P. Vatta, J. J. Hyldig-Nielsen and D. H. Ly, *Bioorg. Med. Chem. Lett.*, 2006, **16**, 4931.

98. A. Dragulescu-Andrasi, P. Zhou, G. He and D. H. Ly, *Chem. Commun.*, 2005, **2**, 244.

99. A. Dragulescu-Andrasi, S. Rapireddy, G. He, B. Bhattacharya, J. J. Hyldig-Nielsen, G. Zon and D. H. Ly, *J. Am. Chem. Soc.*, 2006, **128**, 16104.

100. E. A. Englund and D. H. Appella, *Angew. Chem. Int. Ed. Engl.*, 2007, **46**, 1414.

101. A. Manicardi, E. Fabbri, T. Tedeschi, S. Sforza, N. Bianchi, E. Brognara, R. Gambari, R. Marchelli and R. Corradini, *Chembiochem*, **13**, 1327.

102. R. Mitra and K. N. Ganesh, *Chem. Commun.*, 2011, **47**, 1198.

103. R. Mitra and K. N. Ganesh, *J. Org. Chem.*, 2012, **77**, 5696.

104. S. Pensato, M. Saviano, N. Bianchi, M. Borgatti, E. Fabbri, R. Gambari and A. Romanelli, *Bioorg. Chem.*, 2010, **38**, 196.

105. C. Dose and O. Seitz, *Org Lett*, 2005, **7**, 4365.

106. C. Avitabile, L. Moggio, G. Malgieri, D. Capasso, S. Di Gaetano, M. Saviano, C. Pedone and A. Romanelli, *PLoS One*, 2012, **7**, e35774.

107. B. Sahu, I. Sacui, S. Rapireddy, K. J. Zanotti, R. Bahal, B. A. Armitage and D. H. Ly, *J. Org. Chem.*, 2011, **76**, 5614.

108. M. Egholm, L. Christensen, K. Dueholm, O. Buchardt, J. Coull and P.E. Nielsen, *Nucleic Acids Res.*, 1995, **23**, 217.

109. G. Haaima, H. F. Hansen, L. Christensen, O. Dahl and P. E. Nielsen, *Nucleic Acids Res.*, 1997, **25**, 4639.

110. K. N. Ganesh and R. Vysabhattar, *Tetrahedron Lett.*, 2008, **49**, 1314.

111. I. Bang, *Z. Physiol. Chem.*, 1898, **26**, 133.

112. I. Bang, *Z. Biochem*, 1910, **26**, 293.

113. G. Biffi, D. Tannahill, J. McCafferty and S. Balasubramanian, *Nat. Chem.*, 2013, **5**, 182.

114. R. K. Moyzis, J. M. Buckingham, L. S. Cram, M. Dani, L. L. Deaven, M. D. Jones. J. Meyne, R. L. Ratliff and J. R. Wu, *Proc. Natl. Acad. Sci. U. S. A.*, 1998, **85**, 6622.

115. M. Gellert, M. N. Lipsett and D. R. Davies, *Natl. Acad. Sci. U. S. A.*, 1962, **48**, 2013.

116. A. K. Todd, M. Johnston and S. Neidle, *Nucleic Acid Res.*, 2005, **33**, 2901.

117. A. Bugaut and S. Balasubramanian, *Biochemistry*, 2008, **47**, 689.

118. a) S. Burge, G. N. Parkinson, P. Hazel, A. K. Todd and S. Neidle, *Nucleic Acid Res.*, 2006, **34**, 5402. b) G. W. Collie and G. N. Parkinson, *Chem. Soc. Rev.*, 2011, **40**, 5867. c) S. M. Mirkin, *Front Biosci.*, 2008, **13**, 1064.

119. D. L. Robertson and J. F. Joyce, *Nature*, 1990, **344**, 467.

120. A. D. Ellington and J. W. Szostak, *Nature*, 1990, **346**, 818.

121. C. Tuerk and L. Gold, *Science*, 1990, **249**, 505.

122. a) B. Gatto, M. Palumbo and C. Sissi, *Curr. Med. Chem.*, 2009, **16**, 1248. b) P. J Bates, D. A. Laber, D. M. Miller and S. D. Thomas, *J. Trent, Exp. Mol. Pathol.*, 2009, **86**, 151.

123. a) D. J. Schneider, J. Feigon, Z. Hostomsky and L. Gold, *Biochemistry*, 1995, **34**, 9599. b) D. Michalowski, R. Chitima-Matsiga, D. M. Held and D. H. Burke, *Nucleic Acid Res.*, 2008, **36**, 7124. c) M. L. Andreola, F. Pileur, C. Calmels, M. Ventura, L. Tarrago-Litvak, J. J. Toulme and S. Litvak, *Biochemistry*, 2001, **40**, 10087.

124. K. Murakami, F. Nishikawa, K. Noda, T. Yokoyama and S. Nishikawa, *Prion*, 2008, **2**, 73.

125. P. J. Bates, D. A. Laber, D. M. Miller, S. D. Thomas and J. O. Trent, *Exp. Mol. Pathol.*, 2009, **86**, 151.

126. L. C. Bock, L. C. Griffin, J. A. Latham, E. H. Vermaas and J. J. Toole, *Nature*, 1992, **355**, 564.

127. G. M. Blackburn, M. J. Gait, D. Loakes and D. M. Williams, *Nucleic acids in chemistry and biology*, 2006, *RSC publication 3rd edition*.

128. B. Jeremy and J. Tymoczko, 2006, *Biochemistry 6th edition*. W.H. Freeman and Company.

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

130. J. L. Mergny, A.T. Phan and L. Lacroix, *FEBS Letters*, 1998, **435**, 74.

131. a) P. Job, *Ann. Chim. Appl.*, 1928, **9**, 113. b) C. Y. Huang, *Methods in Enzymology*, 1982, **87**, 509.

132. K. E. Vaanholde, J. Brahms and A. Michelson, *J. Mol. Bio.*, 1965, **12**, 726.

133. P. R. Callis, *Annu. Rev. Phys. Chem.*, 1983, **34**, 329.

134. a) M. A. Mendez and V. A. Szalai, *Biopolymers*, 2009, 91, 841. b) F. A. Miannay, A. Banyasz, T. Gustavsson and D. Markovitsi, *J. Phys. Chem.,* 2009, **113**, 11760. c) N. T. Dao, R. Haselsberger, M. E. Michel-Beyerle and A. T. Phan, *FEBS Lett.*, 2011, **585**, 3969.

135. a) H. Martadinata and A. T. Phan, *J. Am. Chem. Soc.*, 2009, **131**, 2570. b) J. Qi and R. H. Shafer, *Biochemistry*, 2007, **46**, 7599. c) G. W. Collie, G. N. Parkinson, S. Neidle, F. Rosu, E. De Pauw and V. Gabelica, *J. Am. Chem. Soc.*, **132**, 9328.

Chapter 2

Synthesis of polycarbamate analogue of GNA/TNA and their biophysical studies

Section A

Synthesis of *R*/*S*-GCNA monomers and oligomers

2A.1 Introduction

Key inputs from Eschenmoser's group in solving the mystery of chemical etiology of nucleic acid structure has established that Watson-Crick base pairing can be supported by sugars in the backbone that differ from the naturally selected ribose/deoxyribose RNA or DNA (**Figure 2.1A**).[1]Amongst several possibilities studied, an interesting example is that of Tetrose Nucleic Acid (TNA) in which the pentose sugar in DNA/RNA is replaced by an atom-edited cyclic form of tetrose sugar (**Figure 2.1B**).[2] TNA was found to be capable of forming antiparallel duplexes by self-pairing and were also able to cross-pair with cDNA and RNA.[3] Inspired by Eschenmoser's TNA structure, Eric Meggers further structurally simplified TNA to an atom economic novel nucleic acid analogue (an acyclic version of TNA), known as Glycol Nucleic Acid (GNA)[4,5] (**Figure 2.1C, D**). Although strong self-pairing of complementary GNA duplexes was observed, only *S*-GNA could form cross-paired duplexes with complementary RNA but with much less stability than the natural DNA: RNA duplex.
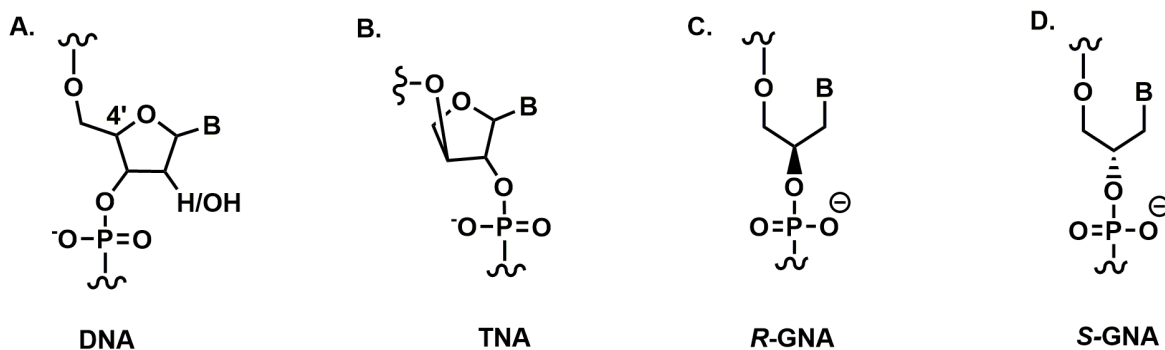
Figure 2.1: DNA, Threose Nucleic Acid (TNA), Glycol Nucleic Acid (GNA)

Other glycerol based nucleic acid analogues in which the base attachment is three bonds away (FNA)[6] (**Figure 2.2A**) or one bond away (isoGNA)[7] (**Figure 2.2B**) compared to GNA were also studied but were found incompatible for self-pairing or cross-pairing. In-

spired by GNA, acyclic threoninol nucleic acids (aTNA)[8] (**Figure 2.2C**) and serinol nucleic acids (SNA)[9] (**Figure 2.2D**) were reported by Asanuma *et al*. They found that SNA as well as aTNA formed stable homo duplex and SNA was able to cross-pair with both DNA and RNA.
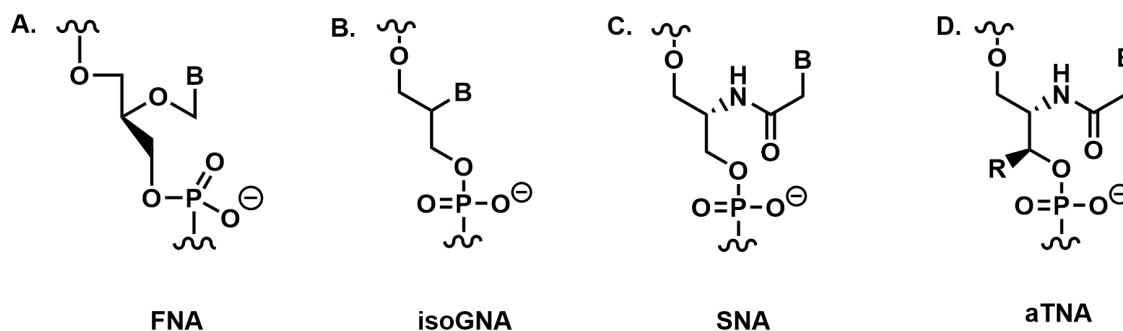
Figure 2.2: Framework of FNA, isoGNA, SNA, aTNA

A number of modifications on the backbone have been reported in literature but a pioneering report was made by Nielsen *et al*[10] which dates back to 1991. In PNA, nucleo-bases are attached to the polyamide backbone through conformationally rigid tertiary acetamide linker.[11] PNA formed a strong and sequence specific binding to both complementary DNA and RNA. Despite having several advantages such as resistance to cellular enzymes[12] (proteases and nucleases), the major limitations confounding its application are, (1) poor aqueous solubility (2) inefficient cellular uptake and (3) ambiguity in DNA/RNA recognition arising from its almost equally facile binding in both parallel and antiparallel orientation with the complementary nucleic acid sequences.[13] The other non-ionic modification reported involves replacement of phosphodiester linkage with carbamate linkage. These linkages were earlier utilized to form DNA mimics by some groups. First reports were made in 1974 by Gait *et al*,[14] (**Figure 2.3A**) where a dinucleotide analogue containing the oxyformamido-linkage, thymidinyl formamido-[3'(*O*) 5'(*C*)]-5'-deoxythymidine was synthesized. A number of attempts were made to synthesize polymers containing the oxyformamido-linkage, but without success. These linkages are stable under physiological conditions and are resistant to nuclease action.[15]

Earlier, pyrrolidinyl carbamate oligonucleotides[16] (**Figure 2.3B**) were reported by our group. The flexibility in the linker group led to destabilization of the complexes. In another

report by our group, the internucleoside polyamide linkages in PNA backbone were re-placed by carbamate linkages and the nucleobase linker was retained as in PNA to get pol-ycarbamate nucleic acids (PCNA). Both *R*-PCNA[17] (**Figure 2.3C**) and *S*-PCNA[18] (**Figure 2.3D**) formed more stable duplex with DNA than RNA.
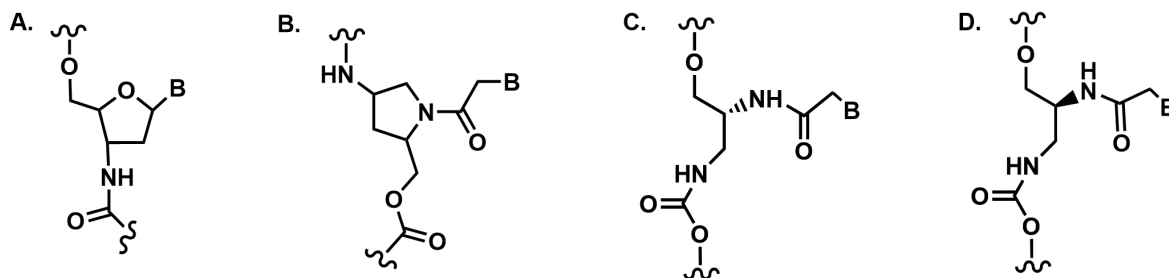
Figure 2.3: Carbamate analogues reported

2A.2 Rationale

In the present work, keeping acyclic GNA backbone unaltered, the phosphodiester linker in GNA was replaced by carbamate, changing secondary hydroxyl group in GNA to amino-group and was named Glycol Carbamate Nucleic Acid (GCNA, **Figure 2.4 D, E**) be-cause of the structural similarity to GNA (**Figure 2.4 B, C**). The proposed atom-edited GCNA backbone was shorter than the acyclic PCNA, both in linker to the nucleobase and the repeating bonds in the backbone. Carbamate linkages are slightly shorter and rigid compared to the phosphodiester linkages but more flexible than the amide linkers.[19, 20] The rigidity in cyclic tetrose sugar-phosphate backbone in TNA (**Figure 2.4 A**) allows self-pairing as well as cross-pairing with either DNA or RNA. X-ray studies have pointed out that the higher stability of TNA: RNA duplexes is due to the reduced P…P distance in cyclic TNA which is compatible for A-form duplex with RNA.[3] While forming duplex with TNA, DNA has to adopt conformation (like RNA) compatible to TNA and hence the stability of TNA:DNA duplexes is less pronounced.[3] We envisaged that the shorter, rigid carbamate linker replacing phosphodiester linkage in acyclic GNA may impart the oligomers structur-al features that would give incremental benefit for cross-pairing with RNA.[11]
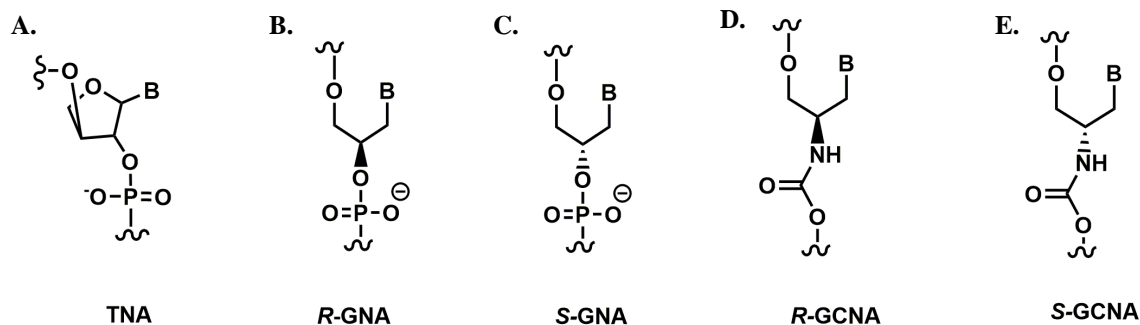
Figure 2.4: Rational design of proposed *R*-GCNA and *S*-GCNA in the present work

Robust RNA pairing is the most important attribute for applications of modified nucleic acids. The charge neutral nature of the oligomers may also show better self-pairing and cross-pairing ability with cDNA/RNA in the absence of charge-charge repulsions.

2A.3 Synthesis of monomer units

The naturally occurring L-Serine was versatile starting material for the synthesis of *R*- monomer. However, *S*- monomer could also be synthesized by simple manipulation of the protecting groups. The hydroxyl group in L- Serine could be easily implemented for activation with *p*-nitrophenyl chloroformate to obtain the activated carbonate monomer in case of *R*- while for *S*- the acid could be reduced to alcohol to achieve hydroxyl functionality.

2A.3.1 Synthesis of *R*- monomer units

The *R*- monomer units were easily synthesized starting from naturally occurring L-serine as depicted in **Scheme 2.1.** The hydroxy group in *N*-(*tert*-Butoxycarbonyl)-L-serine methyl ester was protected as silyl ether to get compound **2** using TBS-Cl. Reduction of the ester group in protected L-serine derivative **2**, produced alcohol **3** using $NaBH_4$ in MeOH. Mitsunobu reaction was employed for the attachment of protected nucleobases (*N*3-benzoyl thymine and N^{6}-Cbz adenine) to give the monomer precursors **4** and **5**. The TBS group was deprotected using TBAF in THF to get the *R*-monomeric units **6, 7**. Compound **6** was debenzoylated using 1:1 dioxane: aq. ammonia to yield **8**. Monomers **7** and **8** were then ac-

tivated using *p*- nitro-phenyloxycarbonyl chloride yielding **9** and **10**, respectively in good overall yield.
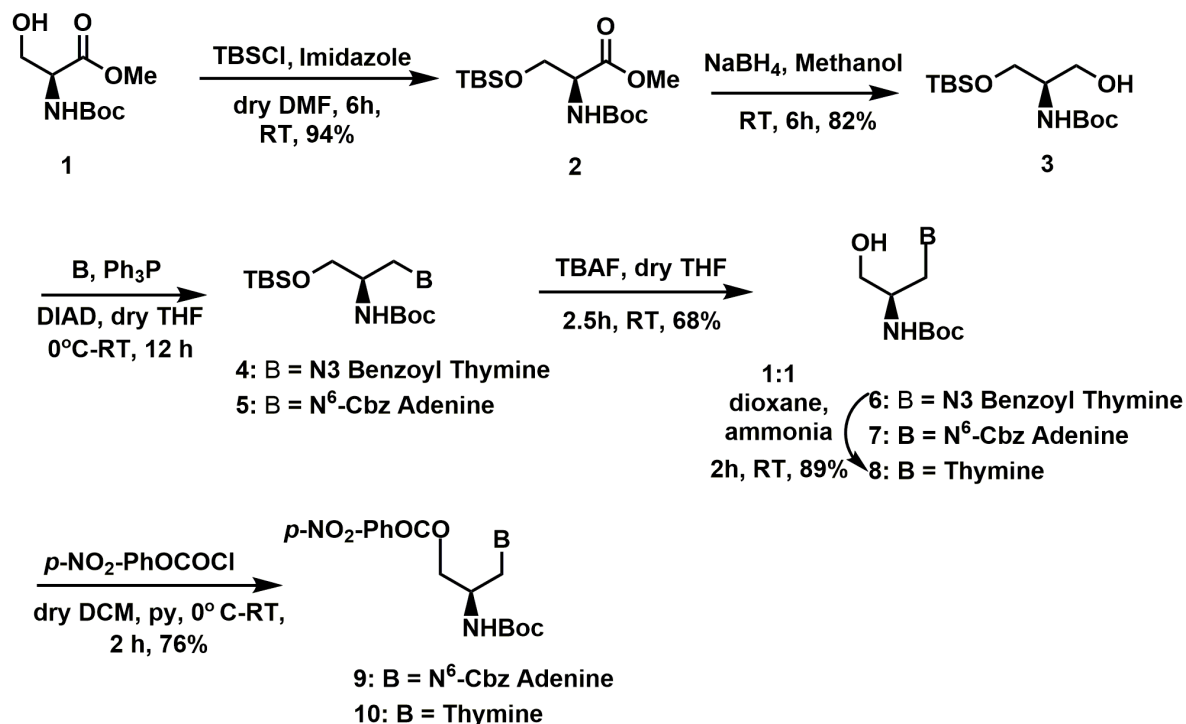
Scheme 2.1: Synthesis of *R*-GCNA monomer units

2A.3.2 Synthesis of *S*- monomer units

Alternatively, the (*S*)-enantiomers were obtained from Garner's alcohol (Compound **11**), which was synthesized from L-Serine derivative **1** according to literature reports.[21] Compound **12** was obtained by TBDPS protection of Garner's alcohol. Mitsunobu reaction with nucleobases (*N*3-benzoyl thymine and *N*[6-]Cbz adenine) yielded compounds **14** and **15**, respectively. Removal of TBDPS group was achieved using TBAF in THF to give *S*-monomer units **16, 17**. Debenzoylation of **16** gave **18** in quantitative yield. Monomers **17, 18** were activated using *p*- nitro-phenyloxycarbonyl chloride to yield **19** and **20,** respectively. The activated monomers **9, 10** and **19, 20** could be directly used for solid phase oligo-carbamate synthesis. All the new compounds were adequately characterized by ^{1}H, ^{13}C NMR and HR-MS analysis.
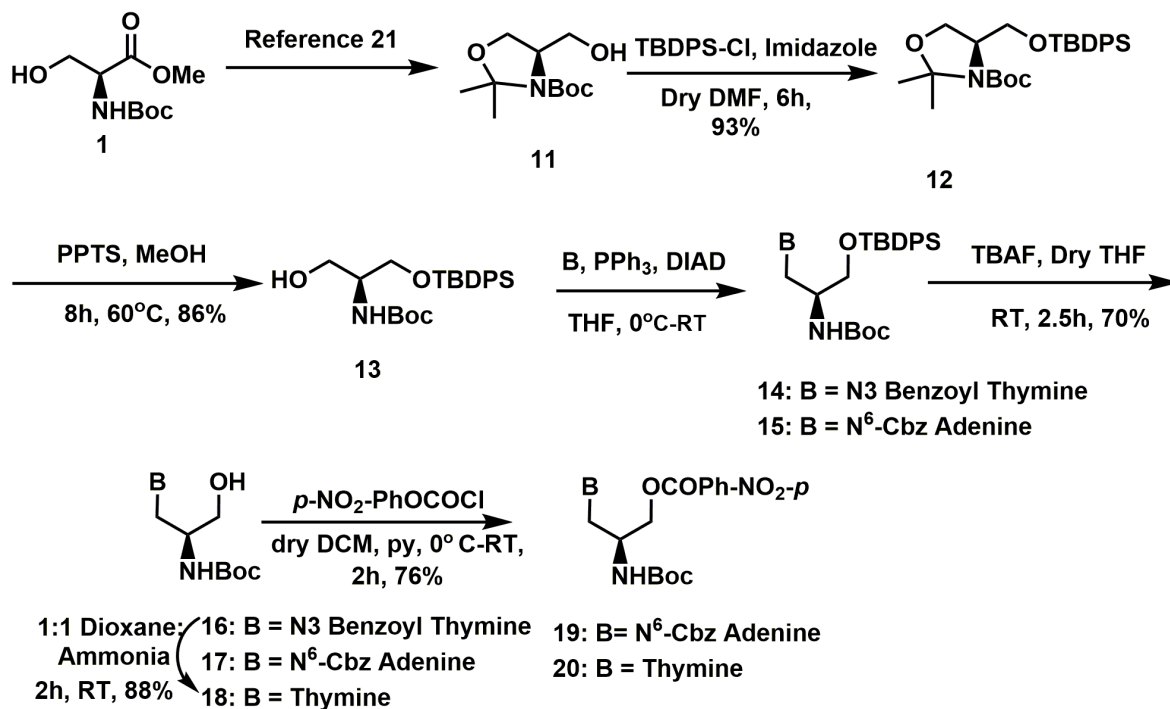
Scheme 2.2: Synthesis of *S*-GCNA monomer units

Optical purity of compounds **7, 17** and **8, 18** was ascertained by Chiral HPLC. Chiral HPLC was accomplished on Kromasil 5-Amycoat (4.6x250mm) column in mobile phase isopropyl alcohol: Petroleum ether (50:50).

2A.4 Synthesis, purification and characterization of oligomers

2A.4.1 Solid Phase Carbamate Synthesis (SPCS)

2A.4.1.1 General principles of Solid Phase Synthesis

In disparity to the solution phase method, the solid phase peptide synthesis strategy implemented by Merrifield,[22] offers great advantage. In this method, the *C*- terminal is attached to an insoluble matrix such as polystyrene beads having reactive functional groups, which also act as a permanent protection for the carboxylic acid (**Figure 2.5**). The next N^{α} – protected amino acid is coupled to the resin bound amino acid either by using an active pentafluorophenyl (pfp) or 3-hydroxy-2,3-dihydro-4-oxobenzotriazole (Dhbt) ester or by an in situ activation with carbodiimide reagents along with HOBt, to drive the reaction to completion (>95%). The unreacted reagents are then washed out and the deprotection, cou-

pling reactions and washing cycles are repeated until the desired peptide is achieved. The need to purify the coupling at every step is obviated. Finally, the resin bound peptide and the side chain protecting groups are cleaved in one step. Solid phase carbamate synthesis proposed in the present studies has a major advantage over solid phase peptide synthesis. As the Boc-amino alcohol is already activated by *p*- nitrophenylchloroformate, expensive coupling reagents are not required for the synthesis of oligomers, therefore excess mono-mers can be recovered at the end of each coupling.

Solid phase peptide/carbamate synthesis can be carried out either via Boc or Fmoc chemistry. The difference between the two synthesis lies in choice of protected monomers employed for solid phase. Boc and Fmoc protecting groups of amine are orthogonal to each other i.e. while Boc group is deprotected under acidic conditions, Fmoc group requires basic condition for deprotection. The choice of resin therefore depends on the chemistry of solid phase synthesis.
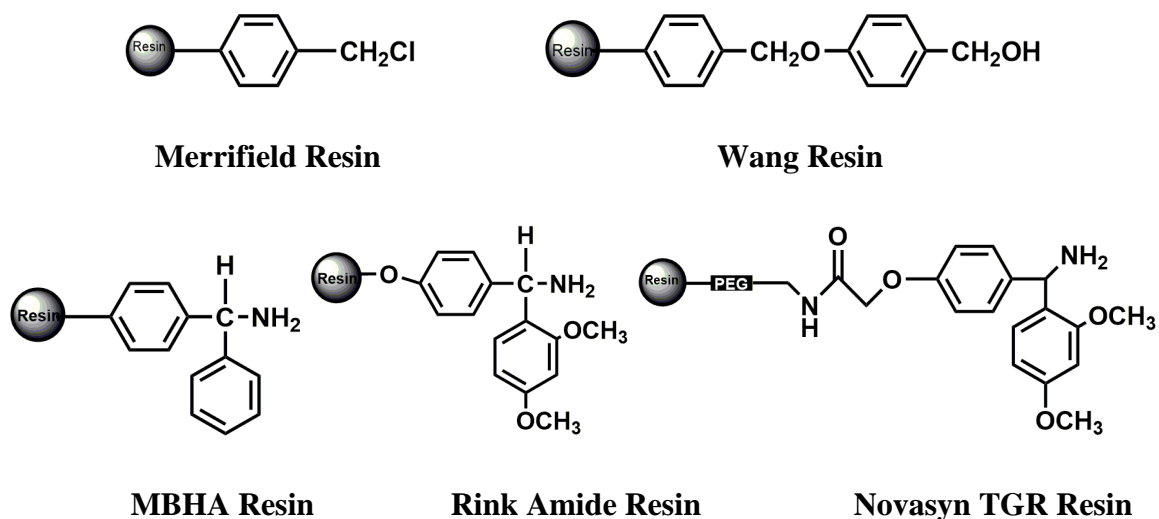
Figure 2.5: Representative structures of resin in SPPS/SPCS

Cleavage of peptides from resins like Merrifield, MBHA requires harsh conditions of TFMSA, TFA, while Wang, Rink amide and Novasyn TGR requires only TFA for cleavage. Oligomers susceptible for degradation under harsh conditions can therefore be synthesized via Fmoc chemistry. Structures of reactive functional groups of some of the commonly used resins are shown above (**Figure 2.5**).

The advantages of solid phase synthesis are (i) all the reactions are performed in a single vessel minimizing the loss due to transfer, (ii) large excess of monomer component can be used resulting in high coupling efficiency (iii) avoids purification step after each coupling reaction .These properties of solid phase synthesis makes the process of synthesis of oligomers easier and faster in comparison to solution phase synthesis.

2A.4.1.2 Functionalization and picric acid estimation of the MBHA [(4-methyl benzhydryl) amine] Resin

The loading value of MBHA resin was reduced to 0.3mmol/g, by coupling diprotected L-lysine carboxylic group (N^{α}-amino group protected with Boc and N^{ω}- amino group by Cl-CBz) to the free amines on the resin using coupling reagents followed by capping with acetic anhydride.[23] Then the final loading value of the functionalized resin was estimated using picric acid.[24]

L-lysine loaded MBHA resin was taken in two different solid phase vessels and their exact weight was recorded. They were then swelled in DCM for half hour to expose the free amines. The Boc protecting group was deprotected using TFA and the trifluoroacetate salts of amine were neutralized with 5% DIPEA in DCM. To the resin 0.1M picric acid solution in DCM was added. After 10 mins, the solution was flushed out from solid phase flask and the procedure was repeated again. The resin was washed properly with DCM. The resins were then washed with 5% DIPEA in DCM when the picric acid eluted as amine salt. This process was repeated again and all the washings were collected in a volumetric flask of 10mL and the volume adjusted to 10 mL. This entire process was repeated for another flask containing weighed amount of resin. The absorbances of the picrate solutions were recorded at 358nm. Loading value of the MBHA resin was calculated from the absorbance applying the formula given below.

$$\text{Loading Value} = \frac{Observed\ Absorbance\ x\ Dilution\ factor}{Molar\ extinction\ coefficient}$$
(Forthe resin)

2A.4.1.3 Kaiser's Test

Kaiser's test[25] was used to monitor the *t*-Boc-deprotection and coupling steps in the solid phase synthesis. A few beads of the resin from solid phase flask were taken in a test tube. To the test tube, 3-4 drops of each of the following solutions (A, B, C) were added.

A. 0.7mg of KCN in 1mL of water and 49mL of pyridine.

B. 1.0 g of ninhydrin in 20 mL of ethanol.

C. 40 g of phenol in 20 mL of n-butanol.

The test tubes were heated at 110°C for 5minutes and the colour of the beads was noted. A blue colour on the beads, which slowly comes into solution, indicated successful deprotection, while colorless beads and the solution confirmed the completion of the coupling reaction.

2A.4.1.4 Cleavage of the GCNA oligomers from the solid support

The resin bound GCNA oligomer (5mg) was kept in an ice bath with thioanisole (10 µL) and 1, 2-ethanedithiol (4µL) for 10 min. TFA (80µL) was added and shaken manually and kept for another 10min. TFMSA (8µL) was added and and the mixture was allowed to stand for 2h. The reaction mixture was filtered through a sintered funnel. The residue was washed with TFA (3x2mL) and the combined filtrate and washings were evaporated under vacuum. The residue was precipitated using dry diethyl ether and centrifuged. The diethyl ether was decanted and the solid part was redissolved in 1:1 ACN: Water.

2A.4.2 Reverse Phase-HPLC

The crude GCNAs were purified on a semipreparative C18 column attached to a Waters HPLC system. A gradient elution method contained A = 5% Acetonitrile in water + 0.1% trifluoroacetic acid and B = 50% Acetonitrile in water + 0.1% trifluoroacetic acid (A to B = 100% in 20 min with a flow rate of 1.5 mL/min), and the eluent was monitored at 260 nm. The purity of the oligomers was further assessed by an RP-C18 analytical HPLC column. The purity of the purified oligomers were found to be >98%.

2A.4.3 MALDI-TOF mass Spectrometry

Literature reports the analysis of PNA oligomers by MALDI-TOF mass spectrometry in which several matrices have been explored, viz. Sinapinic acid (3, 5-dimethoxy-4-hydroxycinnamic acid),[26] CHCA (cyano-4-hydroxycinnamic acid)[27] and DHB (2,5-dihydroxybenzoic acid). Out of these, CHCA was found to give the best signal to noise ratio.

For all the MALDI-TOF spectra recorded for the *R*/*S*-GCNA oligomers reported in this chapter, CHCA (α-Cyano-4-hydroxycinnamic acid) was used as the matrix. The MALDI-TOF spectras were recorded on AB SCIEX 5800 MALDI TOF TOF instrument.

Table 1: GCNA sequences synthesized and characterized

Code	Sequences	HPLC t_R (min)	MALDI-TOF mass	
			Calcd.	Obsvd.
R-GCNA-1	tttttttt-Lys	13.1	1946.81	1951.19
S-GCNA-1	tttttttt-Lys	13.0	1946.81	1952.84
R-GCNA-2	atattattaatt-Lys	11.3	2891.07	2918.26(M+Na)
S-GCNA-2	atattattaatt-Lys	11.2	2891.07	2918.14(M+Na)
R-GCNA-3	aattaataatat-Lys	9.4	2909.10	2929.2053(M+Na)
S-GCNA-3	aattaataatat-Lys	9.3	2909.10	2929.22(M+Na)

t =Thymine, a=Adenine monomers of *R*/*S*-GCNA as indicated in the code

2A.5 Summary

§ A method was developed successfully for the synthesis of chirally homogeneous (*R*)- and (*S*)- Glycol Carbamate Nucleic Acids (GCNA) from naturally occurring L-serine. The carbamate oligomers were synthesized by solid phase synthesis using these activated carbonates.

§ Solid phase synthesis of oligocarbamates did not require any expensive coupling reagents like TBTU, HBTU/HOBt which made synthesis of these oligomers economical. Moreover the excess unreacted activated monomers were easily recovered during solid phase synthesis.

- The simplicity of the proposed backbone with minimum synthetic efforts to access both the isomeric GCNA forms from natural L-serine in this work is worth a mention.

2A.6 Experimental

(*S*)-methyl 2-((*tert*-butoxycarbonyl)amino)-3-((*tert*-butyldimethylsilyl)oxy) propanoate (2): (*S*)-methyl 2-((*tert*-butoxycarbonyl)amino)-3-hydroxypropanoate (2g, 9.12mmol), imidazole (1.9g, 27.36mmol), *tert*-butyl dimethylsilyl chloride (1.6g, 10.95mmol) were dissolved in minimum amount of dry DMF and stirred for 6h. On completion of the reaction the DMF was removed under vacuum and redissolved in EtOAc. The organic layer was washed with water (3 x 50mL) and then brine (2 x 30mL). The organic layer was kept over anhydrous Na_2SO_4 and evaporated to dryness to get the crude compound. The compound was purified by column chromatography (5% EtOAc in petroleum ether) to yield compound **2** (2.9g, 96%) as colorless oil. $[\alpha]_D^{20}$ +14.7° (c 0.03, $CHCl_3$); ^1H NMR(200 MHz, $CDCl_3$): δ (ppm) 0.02-0.04(d, 6H, J=2.65Hz, Si-(C$\underline{H}_3$)$_2$), 0.086(s, 9H, s, 9H, Si-C(C$\underline{H}_3$)$_3$), 1.46(s, 9H, *t*Boc), 3.74(s, 3H, -COOC$\underline{H}_3$), 3.76-3.85 (dd, 1H, J=3.16 Hz, J=6.94 Hz, -OC$\underline{H}_2$-C), 4.01-4.07(dd, 1H, J=2.65 Hz, J= 7.32 Hz, -OC$\underline{H}_2$-C), 4.32-4.38 (m, 1H, -C$\underline{H}$(CO)NH(CH$_2$)), 5.33-5.37 (s, 1H, br s,-N$\underline{H}$Boc); ^{13}C NMR (200 MHz, $CDCl_3$) δ (ppm): -5.52 Si-(C$\underline{H}_3$)$_2$, 18.20 (Si-C(C$\underline{H}_3$)$_3$), 25.71(*t*Boc), 28.34(Si-$\underline{C}$(CH$_3$)$_3$), 52.28(-COO$\underline{C}$H$_3$), 55.60(-O$\underline{C}$H$_2$-C), 63.79(-$\underline{C}$H(CO)NH(CH$_2$), 79.93(*t*Boc), 155.49(-$\underline{C}$O *t*Boc), 171.32 (-$\underline{C}$OOCH$_3$); HRMS calcd. for $C_{15}H_{31}O_5NNaSi$: 356.1864, Observed mass: 356.1860
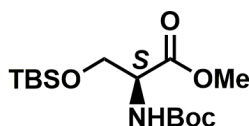
(*R*)-*tert*-butyl (1-((*tert*-butyldimethylsilyl) oxy)-3-hydroxypropan-2-yl)carbamate (3)
A solution of (*S*)-methyl 2-((*tert*-butoxycarbonyl) amino)-3-((*tert*-butyldimethylsilyl) oxy) propanoate **2** (29g, 87.1 mmol) in methanol (250mL) was cooled to 0°C in an ice bath. Solid $NaBH_4$ (4.9g, 130.63mmol) was added in portions for a period of 30mins. The reaction mixture was stirred for a period of 3 h and finally quenched using NH_4Cl solution till pH was neutral. Methanol was removed under reduced pressure and the residue was extracted with EtOAc (4x100mL), washed with water and brine. The organic layer was dried over anhydrous Na_2SO_4 and solvent was removed
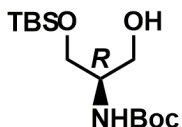
under reduced pressure. The residue was purified by column chromatography (20% EtOAc in petroleum ether) affording compound **3** (24.1g, 83%) as colorless oil: $[\alpha]_D^{20}$ +16.2° (c 0.021, CHCl$_3$); ^{1}H NMR(200 MHz, CDCl$_3$): δ (ppm) 0.08(s, 6H, Si-(C$\underline{H}_3$)$_2$), 0.9 (s, 9H, Si-C(C$\underline{H}_3$)$_3$), 1.45 (s, 9H, tBoc), 3.61–3.87 (m, 5H 2-CH$_2$, 1-CH, 2-CH$_2$), 5.15(br s, OH); ^{13}C NMR (200 MHz, CDCl$_3$) δ (ppm): -5.56 (Si-(C$\underline{H}_3$)$_2$), 18.21 (Si-$\underline{C}$(CH$_3$)$_3$), 25.84 (Si-C(C$\underline{H}_3$)$_3$), 28.38 (O-C(C$\underline{H}_3$)$_3$), 52.54(C$\underline{H}$), 64.06(C$\underline{H}_2$), 79.64(O-$\underline{C}$(CH$_3$)$_3$), 156.06 (NH$\underline{C}$OO-); HRMS calcd for C$_{14}$H$_{31}$NO$_4$Si Na: 328.1911 , Observed mass: 328.1915

(*R*)-*tert*-butyl(1-(3-benzoyl–thyminyl-)-3-((*tert*-butyldimethylsilyl)oxy) propan-2-yl) carbamate (4)
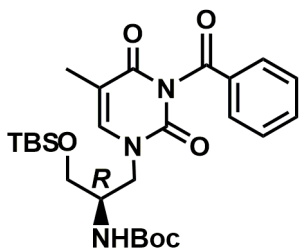
*N*3- benzoylthymine (0.9g, 3.92mmol) and triphenyl phosphine (1 g, 3.92mmol) were dissolved in 40mL dry THF and the solution was cooled to 0 °C. At this temp compound **3** (1g, 3.27mmol) dissolved in 10mL dry THF, was added to the stirred solution followed by dropwise addition of DIAD (1mL, 4.91mmol). The solution was gradually allowed to reach the room temperature and stirring was continued overnight at room temperature. The solvent was removed under pressure and the residue was extracted with EtOAc (4x100mL), washed with water and brine. The organic layer was dried over anhydrous Na$_2$SO$_4$ and solvent was removed under reduced pressure. Compound **4** was used as such for the next reaction without further purification.

(*R*)-*tert*-butyl(1-(6-benzyloxycarbonyladeninyl-)-3-((*tert*-butyl dimethyl silyl) oxy)propan-2-yl)carbamate (5)

N^6- benzyloxycarbonyladenine (1.1 g, 3.93 mmol) and triphenyl phosphine (1 g, 3.92 mmol) were dissolved in 40mL dry THF and the solution was cooled to 0°C. At this temp. compound **3** (1 g, 3.27mmol) dissolved in 10mL dry THF was added to the stirred solution followed by dropwise addition of DIAD (1mL, 4.91 mmol). The solution was gradually allowed to reach the room temperature and stirring was continued overnight at room temperature. The solvent was removed under pressure and the residue was extracted with EtOAc (4x100mL), washed with water and brine. The organic

layer was dried over anhydrous Na_2SO_4 and solvent was removed under reduced pressure. Compound **5** was used as such for the next reaction without further purification.

(*R*)-*tert*-butyl (1-(3-benzoyl-thyminyl)-3-hydroxypropan-2-yl) carbamate (6)

(*S*)-*tert*-butyl (1-(3-benzoyl-thyminyl)-3-hydroxypropan-2-yl) carbamate (16)
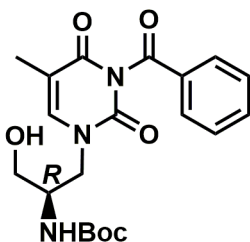
0.5g (0.97mmol) of the compound **4** was dissolved in 3 mL of dry THF. To this TBAF (1.5mL of 1M TBAF, 1.45mmol) was added slowly and with constant stirring. The reaction was allowed to stir for 2.5h. THF was removed under vacuum and the residue was extracted with DCM (2x100mL), dried over Na_2SO_4. The solvent was evaporated under vacuum and the crude compound was purified by column chromatography (2% MeOH in DCM) furnishing compound **6, 16** (0.45g, 68% from **3** to **6/13** to **16**) as colorless solid. $[\alpha]_D^{20}$ +95.2(c 0.015, MeOH) for 5a; $[\alpha]_D^{20}$ -95.3(c 0.015, MeOH) for 5b; ^{1}H NMR(200 MHz, CD_3OD) δ (ppm): 1.33 (s, 9H, *t*Boc), 1.80 (s, 3H, =C(C<u>H</u>3)), 3.23-3.50 (m, 3H, CH_2, CH), 3.92-4.17(m, 2H, CH_2), 7.41-7.48 (m, =C<u>H</u>N, Ph para to CO), 7.59-7.66 (t, Ph meta to CO, J=7.2 Hz, J=7.58 Hz), 8.01-8.05 (d, Ph ortho to CO, J=7.57 Hz); ^{13}C NMR (50 MHz, CD_3OD) δ (ppm): 12.42(=C(<u>C</u>H3)), 28.83(O-C(<u>C</u>H3)3), 51.76(N-<u>C</u>H2-CHN), 52.55 (N-CH2-<u>C</u>HN), 62.94 (O-<u>C</u>H-CHN), 80.42(O-<u>C</u>(CH3)3), 110.17(=<u>C</u>(CH3), 130.28 (Ph-ortho to CO), 131.93 (Ph-meta to CO), 133.11(Ph to CO), 136.27(Ph para to CO), 144.34(<u>C</u>HN), 151.60(N-<u>C</u>O-N), 158.09 (NH<u>C</u>OO-), 165.43(C-<u>C</u>O-N), 170.56 (N-<u>C</u>O-Ph); HRMS calculated for $C_{20}H_{25}O_6N_3Na$: 426.1636, Observed mass: 426.1634

(*R*)-*tert*-butyl (1-(6-benzyloxycarbonyl adeninyl -)-3-hydroxypropan-2-yl) carbamate (7)

(*S*)-*tert*-butyl (1-(6-benzyloxycarbonyl adeninyl -)-3-hydroxypropan-2-yl) carbamate (17)

0.6 g (1.08mmol) of the compound **5** and **15** was dissolved in 3mL of dry THF. To this TBAF (1.6mL of 1M TBAF, 1.62 mmol) was added slowly and with constant stirring. The reaction was allowed to stir for 2.5h. THF was removed under vacuum and the residue was extracted with DCM (2x100mL), dried over Na_2SO_4. The solvent was evaporated under vacuum and the crude compound was purified by column chromatography (2% MeOH in
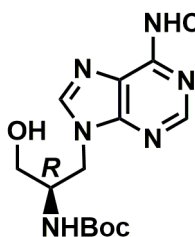
DCM) furnishing compound **7** and **17** (0.59 g, 73% from **5** to **7** and **15** to **17**) as colorless solid; $[\alpha]_D^{20}$ +35.3 (c 0.0102,MeOH) for 6a; $[\alpha]_D^{20}$ -35.3 (c 0.0103,MeOH) for 6b; ^{1}H NMR(200 MHz, CD$_3$OD): δ (ppm) 1.21 (s, 9H, *t*Boc), 3.57-3.62(m,2H,C$\underline{H_2}$), 4.04-4.08(m, 1H,C$\underline{H}$), 4.27-4.29(dd,1H,CH$_2$),4.50-4.57(dd,1H,CH$_2$),5.20 (s, 2H, OC$\underline{H_2}$Ph), 7.26-7.38 (m, 5H, Ph), 8.14 (s, 1H, N-C$\underline{H}$=N), 8.50 (s, 1H, N=C$\underline{H}$-N); ^{13}C NMR (125 MHz, DMSO-*d6*) δ (ppm): 27.86(O-C($\underline{C}$H$_3$)$_3$), 52.02(N-CH$_2$-$\underline{C}$HN), 62.79(N-CH$_2$-$\underline{C}$HN), 67.34(N-$\underline{C}$H$_2$-CHN), 68.74(N-$\underline{C}$H$_2$-CHN), 77.77(O-$\underline{C}$(CH$_3$)$_3$), 123.02(-N=$\underline{C}$(C)-N) ,127.47(-OCH$_2$$\underline{Ph}$ ortho to CH$_2$), 128.00(-OCH$_2$$\underline{Ph}$ para to CH$_2$), 128.58 (-OCH$_2$$\underline{Ph}$ meta to CH$_2$), 149.22(-N=C(N)-C), 151.24(-N$\underline{C}$H=C), 154.27(-NH$\underline{C}$OOCH$_2$Ph), 154.85(-NH$\underline{C}$OOC*t*-Bu); HRMS calculated for C$_{21}$H$_{26}$N$_6$O$_5$Na: 465.1857, Observed mass: 465.1857

(*R*)-*tert*-butyl 1-thyminyl-3-hydroxypropan-2-yl carbamate (8)
(*S*)-*tert*-butyl 1-thyminyl-3-hydroxypropan-2-yl carbamate (18)
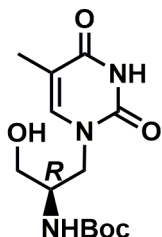
0.89g (2.21mmol) of **6**, **16** was treated with 20mL 1:1 ammonia: dioxane and allowed to stir for 3h. The solvent was then removed completely and the compound was purified by column chromatography (4% MeOH in DCM) to yield **8**, **18** as pure white solid (0.63g, 85%): $[\alpha]_D^{20}$ +42.48 (c 0.025, MeOH) for **8**; $[\alpha]_D^{20}$ -42.6 (c 0.03, MeOH) for **18**; ^{1}H NMR(200 MHz, CD$_3$OD)δ (ppm): 1.37 (s, 9H, *t*Boc), 1.86 (s, 3H, =C(C$\underline{H_3}$), 3.51-3.59 (m, 3H, CH$_2$, CH), 3.96-4.12(m, 2H, CH$_2$), 7.35(s, =C$\underline{H}$N) ; ^{13}C NMR (50 MHz, CD$_3$OD) δ (ppm): 12.41(=C($\underline{C}$H3)), 28.70(O-C($\underline{C}$H$_3$)$_3$), 51.17(N-$\underline{C}$H$_2$-CHN), 52.33 (N-CH$_2$-$\underline{C}$HN), 62.95(O-$\underline{C}$H$_2$-CHN), 80.38(O-$\underline{C}$(CH$_3$)$_3$), 110.48(=$\underline{C}$(CH3)), 143.92($\underline{C}$HN), 153.23(N-$\underline{C}$O-N), 157.94 (NH$\underline{C}$OO-),167.07(C-$\underline{C}$O-N); HRMS calculated for C$_{13}$H$_{21}$O$_5$N$_3$ Na: 322.1370, Observed mass: 322.1373

(*R*)-*tert*-butyl (1-(6-benzyloxycarbonyl adeninyl -)-3-(((4-nitrophenoxy) carbonyl) oxy) propan-2-yl) carbamate (9)
(*S*)-*tert*-butyl (1-(6-benzyloxycarbonyl adeninyl -)-3-(((4-nitrophenoxy) carbonyl) oxy) propan-2-yl) carbamate (19)

Compound **7**, **17** (1g, 2.26mmol) was taken in dry DCM (10mL) and cooled to 0 ℃ in an ice bath. Dry pyridine (0.8 mL, 6.78 mmol), dry triethylamine(0.2mL,1.12mmol) was

added and allowed to stir for 10–15min. *p*- nitrophenyl chloroformate (1.14g, 5.64mmol)
dissolved in DCM (3mL) was added and the reaction was allowed to stir at room temperature for 2h. After completion of the reaction, the reaction mixture was concentrated and the required compound was purified by column chromatography (70% EtOAc in petroleum
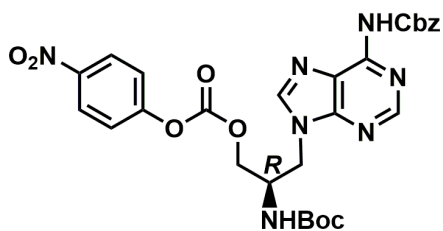
ether) to get **9, 19** (1.04 g, 76%): $[\alpha]_D^{20}$ +48.4 (c 0.006, CHCl$_3$) for **9**; $[\alpha]_D^{20}$ -48.4 (c 0.006, CHCl$_3$) for **19** ; ^{1}H NMR(400 MHz, CDCl$_3$) δ (ppm): 1.26(s, 9H, *t*Boc), 3.56-3.74(m, 1H,-N-CH$_2$-CH), 4.39-4.71(m, 4H,-CH$_2$-CHN,-OCH$_2$-CHN, -N-CH$_2$-CH), 5.32(s, 2H,-OCH$_2$Ph), 7.30-7.39(m, 7H, -OCH$_2$Ph ,-OCOO-Ph ortho to –COO), 8.20-8.29(m, 4H, 2 -N=CN, Ph meta to –OCOO); ^{13}C NMR (125 MHz, CD$_3$OCD$_3$) δ (ppm): 29.70(O-C(CH$_3$)$_3$), 49.75(N-CH$_2$-CHN), 65.35(N-CH$_2$-CHN), 68.07(-OCH$_2$Ph), 69.97(O-CH$_2$-CHN), 81.3(O-C(CH$_3$)$_3$), 122.35(Ph ortho to OCO), 125.94(-N-C(C)=C), 126.24(Ph meta to OCO), 126.24(-OCH$_2$Ph ortho to CH$_2$), 127.43 (-OCH$_2$Ph para to CH$_2$), 128.58(-OCH$_2$Ph meta to CH$_2$), 134.27(-OCH$_2$Ph para to OCO), 139.29(-N=CH-N), 141.02(Ph para to OCO), 145.74(-N-C(N)=C), 146.20(-N=C(C)N), 149.03(-NCH=C), 152.37(-OCOO) ,154.74(-NHCOOCH$_2$Ph), 155.13(-NHCOOC*t*-Bu), 162.69 (Ph-OCO);HRMS calculated for C$_{28}$H$_{30}$O$_9$N$_7$: 608.2100, Observed mass: 608.2088

(R)-tert-butyl 1-thyminyl-3-((((4-nitrophenoxy) carbonyl) oxy) propan-2-yl) carbamate (10)

(S)-tert-butyl 1-thyminyl-3-((((4-nitrophenoxy) carbonyl) oxy) propan-2-yl) carbamate (20)

Compound **8**, **18** (1g, 3.34mmol) was taken in dry DCM (10mL) and cooled to 0℃ in an ice bath. Dry pyridine (0.8mL, 10.02mmol) was added and allowed to stir for 10–15min. *p*-nitrophenyl chloroformate (1.7g, 8.35mmol) dissolved in DCM (3mL) was added and the reaction was allowed to stir at room temperature for 2h. After completion of the reaction, the reaction mixture was concentrated and the required compound was purified by column chromatography (80% EtOAc in petroleum ether) to get **10, 20** (1.2g, 76%): $[\alpha]_D^{20}$ +54.9 (c 0.005,CHCl$_3$) for **10**; $[\alpha]_D^{20}$ -54.75 (c 0.0046,CHCl$_3$) for **20**; ^{1}H NMR(200 MHz, CDCl$_3$): δ (ppm) 1.43(s, 9H, *t*Boc), 1.92(s, 3H, =C(CH$_3$), 3.86-4.40(m, 5H, 2CH$_2$, CH), 5.15(br s,
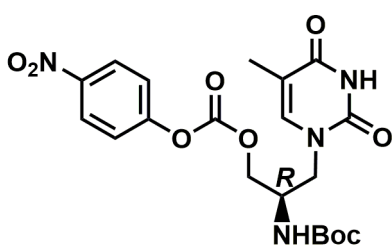
OH), 7.06(s, 1H, =CHN) 7.39-7.43(d, 2H, Ph ortho to OCO, J=8.71Hz), 8.28-8.32(d, 2H, Ph meta to OCO, J=8.84Hz), 8.76(s, 1H, CO-N$\underline{H}$-CO) ; ^{13}C NMR (125 MHz, CD$_3$OCD$_3$) δ (ppm): 12.36(=C($\underline{C}$H3)), 28.46(O-C($\underline{C}$H$_3$)$_3$), 49.47 (N-$\underline{C}$H$_2$-CH-N), 49.69(N-CH$_2$-$\underline{C}$HN), 69.22(O-$\underline{C}$H$_2$-CHN),79.53(O-$\underline{C}$(CH$_3$)$_3$), 109.66(=$\underline{C}$(CH3), 123.14 (Ph ortho to OCO), 126.07 (Ph meta to OCO), 126.07 ($\underline{C}$HN), 142.17(Ph para to OCO), 146.44(N-$\underline{C}$O-N),152.12(O-$\underline{C}$O-O), 153.12(NH$\underline{C}$OO-), 156.66(Ph-OCO), 164.79(-C-$\underline{C}$O-N-); HRMS calculated for C$_{20}$H$_{24}$N$_4$O$_9$Na: 487.1435, Observed mass: 487.1433

(*S*)-*tert*-butyl 4-(((*tert*-butyldiphenylsilyl) oxy) methyl) -2,2-dimethyloxazolidine-3-carboxylate (12)
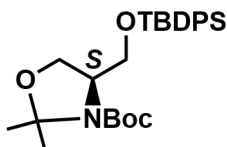
Compound **11** (18.5g, 80.04mmol) was dissolved in minimum amount of dry DMF. To it imidazole (16.3g, 240.11mmol) and TBDPSCl (24.7mL, 96.05mmol) was added along with catalytic amount of DMAP and allowed to stir for 6h at room temperature. After completion of the reaction, DMF was removed under vacuum and the crude compound was redissolved in ethyl acetate. The organic layer was washed with water (3x100mL) and then brine (2x50mL). The organic layer was kept over anhydrous Na$_2$SO$_4$ and evaporated to dryness. The crude compound was purified by column chromatography (5% EtOAc in petroleum ether) to get compound **12** (34.6g, 92%): [α]$_D^{20}$ -21.11(c .027, CHCl$_3$); ^{1}H NMR (200 MHz, CDCl$_3$): δ (ppm) 1.05 (s, 9H, -SiC(C$\underline{H}_3$)$_3$), 1.31(s, 6H, -C(CH$_3$)$_2$, 1.47-1.51 (m, 9H, *t*Boc, rotameric mixture), 3.51-4.24 (m, 5H, 2C$\underline{H}_2$, C$\underline{H}$), 7.38-7.40 (m, 6H, 2 -Si-Ph meta, para), 7.65-7.67 (m, 4H, 2 -Si-Ph ortho); ^{13}C NMR(200 MHz, CDCl$_3$): δ (ppm) 19.28 (-SiC($\underline{C}$H$_3$)$_3$, 23.17(C($\underline{C}$H$_3$)$_2$) , 26.85 (-OC($\underline{C}$H$_3$), 28.37 (-Si$\underline{C}$(CH$_3$)$_3$, 58.28(-CH) ,62.98(-$\underline{C}$H$_2$OH), 65.21 (-$\underline{C}$H$_2$OSi), 79.64 (-O$\underline{C}$(CH$_3$)$_3$, rotameric mixture), 93.94 (-O$\underline{C}$(CH$_3$), 127.73 (-Si-Ph meta) ,129.72 (-Si-Ph ortho),133.51 (-Si-Ph para), 135.53 (-Si-Ph($\underline{C}$)) ,151.80 (O$\underline{C}$OC(CH$_3$)$_3$, rotameric mixture); LCMS calculated for C$_{27}$H$_{39}$O$_4$NSi: 469.26, Observed mass: 469.16

(*S*)-*tert*-butyl (1-((*tert*-butyldiphenylsilyl) oxy)-3-hydroxypropan-2-yl)carbamate (13)

Compound **12** (5g, 10.66mmol) was dissolved in methanol. To it catalytic amount of PTSA was added and the mixture was allowed to stir for 2h. The reaction mixture was quenched with K_2CO_3. The reaction mixture was filtered and the solvent was removed under vacuum. The residue was dissolved in EtOAc and the organic layer was washed with water (3x100mL) and brine (2x100mL). The organic layer was kept over Na_2SO_4 and the solvent was removed to get the crude compound. The compound was purified by column chromatography (30% EtOAc in petroleum ether) to get **13** (3.5g, 77%): $[\alpha]_D^{20}$ -5.86 (c 0.024, MeOH); ^1H NMR(400 MHz, CDCl3): δ (ppm) 1.07 (s, 9H, -SiC(C$\underline{H}_3$)$_3$), 1.43 (s, 9H, *t*Boc), 3.67-3.79 (m, 5H, 2C$\underline{H}_2$, C$\underline{H}$), 5.15 (br s, 1H, -O$\underline{H}$), 7.37-7.39 (m, 6H, 2 -Si-Ph meta, para), 7.64-7.66 (m, 4H, 2 -Si-Ph ortho); ^{13}C NMR(400 MHz, CDCl3): δ (ppm) 18.5 (-SiC($\underline{C}$H$_3$)$_3$, 26.3 (-OC($\underline{C}$H$_3$), 27.65(-Si$\underline{C}$(CH$_3$)$_3$), 52.5(-CH), 62.03(-$\underline{C}$H$_2$OH), 62.91(-$\underline{C}$H$_2$OSi), 78.79(-O$\underline{C}$(CH$_3$)$_3$), 127.09(2 -Si-Ph meta), 129.14(-Si-Ph ortho), 132.29(-Si-Ph para), 134.79(-Si-Ph($\underline{C}$)), 155.37(O$\underline{C}$OC(CH$_3$)$_3$), HRMS calcd for $C_{24}H_{35}NO_4SiNa$: 452.2228, observed mass: 452.2223
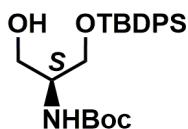
(*S*)-*tert*-butyl(1-(3-benzoyl–thyminyl-)-3-((*tert*-butyldimethylsilyl)oxy) propan -2-yl) carbamate (14)
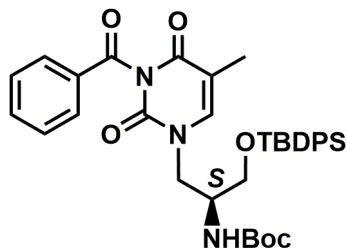
*N*3- benzoylthymine (0.9 g, 3.92mmol) and triphenyl phosphine (1 g, 3.92mmol) were dissolved in 40mL dry THF and the solution was cooled to 0 °C. At this temp compound **13** (1g, 3.27mmol) was dissolved in 10mL dry THF was added to the stirred solution followed by dropwise addition of DIAD (1 mL, 4.91mmol). The solution was gradually allowed to reach the room temperature and stirring was continued overnight at room temperature. The solvent was removed under pressure and the residue was extracted with EtOAc (4x100mL), washed with water and brine. The organic layer was dried over anhydrous Na_2SO_4 and solvent was removed under reduced pressure. Compound **14** was used as such for the next reaction without further purification.

(*S*)-*tert*-butyl(1-(6-benzyloxycarbonyladeninyl-)-3-((*tert*-butyldimethyl silyl) oxy)propan-2-yl)carbamate (15)

N^6- benzyloxycarbonyladenine (1.1g , 3.93mmol) and triphenyl phosphine (1 g, 3.92mmol) were dissolved in 40mL dry THF and the solution was cooled to 0°C. At this temp compound **13** (1g, 3.27mmol) was dissolved in 10mL dry THF was added to the stirred solution followed by dropwise addition of DIAD (1mL, 4.91mmol). The solution was gradually allowed to reach the room temperature and stirring was continued overnight at room temperature. The solvent was removed under pressure and the residue was extracted with EtOAc (4x100 mL), washed with water and brine. The organic layer was dried over anhydrous Na_2SO_4 and solvent was removed under reduced pressure. Compound **15** was used as such for the next reaction without further purification.

2A.7 Appendix

....

¹H NMR
Compound 2

¹³C NMR
Compound 2

DEPT
Compound 2

¹H NMR
Compound 3

¹³C NMR
Compound 3

DEPT
Compound 3

¹H NMR
OTBDPS
Compound 12

¹³C NMR
Compound 12

DEPT
Compound 12

¹H NMR
Compound 6, 16

¹³C NMR
Compound 6, 16

DEPT
Compound 6, 16

¹H NMR
Compound 7, 17

¹³C NMR
Compound 7, 17

DEPT
Compound 7, 17

¹H NMR

Compound 8, 18

Methanol D4

7.35
4.88
4.12
4.10
4.03
4.01
3.98
3.96
3.59
3.56
3.53
3.51
3.32
1.86
1.37

Methanol D4

0.99   2.00   2.97   2.98   9.00

7.5   7.0   6.5   6.0   5.5   5.0   4.5   4.0   3.5   3.0   2.5   2.0   1.5   1.0

¹³C NMR
Compound 8, 18

Methanol d4

167.07
157.94
153.23
143.92
110.48
80.38
62.95
52.33
49.87
49.47
49.04
48.19
28.70
12.41

170 160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10 0

DEPT
Compound 8, 18

143.92
62.94
52.32
51.12
28.70
12.40

150 140 130 120 110 100 90 80 70 60 50 40 30 20 10

Compound 2
Expected mass: 356.1864
Observed mass: 356.1860

Compound 3
Expected mass: 328.1915
Observed mass: 328.1911

F:\DATA\May-2013\21\t   5/21/2013 1:00:39 PM

t #11 RT: 0.17 AV: 1 NL: 2.12E6
T: {0,0} + c ESI corona sid=80.00 det=1600.00 Full ms [100.00-2000.00]

Compound 12
Expected mass: 469.26
Observed mass: 469.16

469.16

205.33

D:\Data\tbdps-alc   8/26/2013 12:53:03 PM

tbdps-alc #1380 RT: 6.15 AV: 1 NL: 1.94E9
T: FTMS + p ESI Full ms [100.00-700.00]

452.2223
R=53107
$C_{24}H_{35}O_4NNaSi$ = 452.2228
-1.0309 ppm

Compound 13
Expected mass: 452.2223
Observed mass: 452.2228

430.2404
R=53707
$C_{24}H_{36}O_4NSi$ = 430.2408
-1.0579 ppm

426.0928
R=30700

436.2482
R=46500

449.2142
R=51407

468.1954
R=47707

Compound 6, 16
Expected mass: 426.1636
Observed mass: 426.1634

Compound 7, 17
Expected mass: 465.1857
Observed mass: 426.1857

D:\Data\d-t-mon 8/26/2013 12:08:13 PM

d-t-mon #843 RT: 3.75 AV: 1 NL: 7.26E8
T: FTMS + p ESI Full ms [100.00-700.00]

322.1370
R=62007
$C_{13}H_{21}O_5N_3Na = 322.1373$
-1.0819 ppm

305.1567
R=62107

328.0446
R=58902

333.9882
R=33600

338.1112
R=57802

Compound 8, 18
Expected mass: 322.1373
Observed mass: 322.1370

Relative Abundance

m/z

D:\Data\l-a-carb 8/26/2013 1:26:37 PM

l-a-carb #924 RT: 4.12 AV: 1 NL: 1.43E9
T: FTMS + p ESI Full ms [100.00-700.00]

608.2088
R=45107
$C_{28}H_{30}O_9N_7 = 608.2100$
-1.8918 ppm

Compound 9, 19
Expected mass: 608.2100
Observed mass: 608.2088

598.2048
R=44007
$C_{28}H_{29}O_7N_7Na = 598.2021$
4.5228 ppm

616.3079
R=44607

622.2247
R=44407

630.1912
R=44307
$C_{28}H_{29}O_9N_7Na = 630.1919$
-1.1655 ppm

Relative Abundance

m/z

Compound 10, 20
Expected mass: 487.1435
Observed mass: 487.1433

Chiral HPLC analysis of compound 7, 17

a) Chiral HPLC of racemic compound
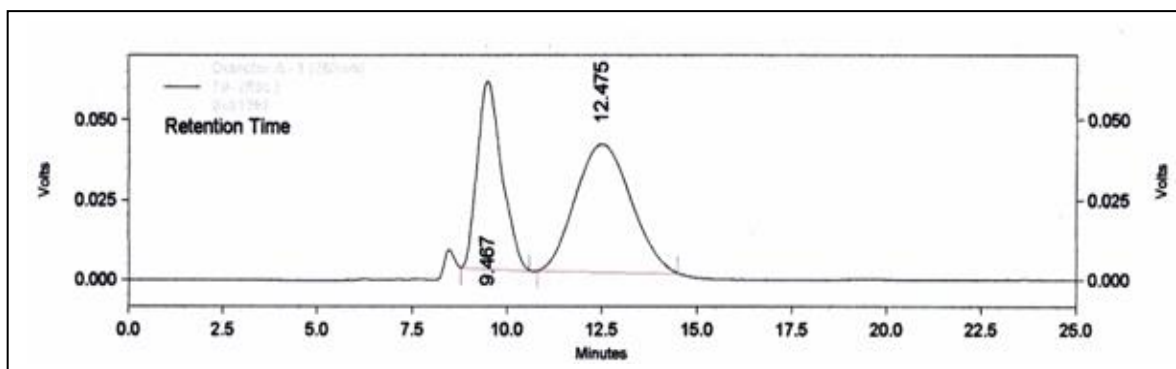
b) Chiral HPLC of **7**
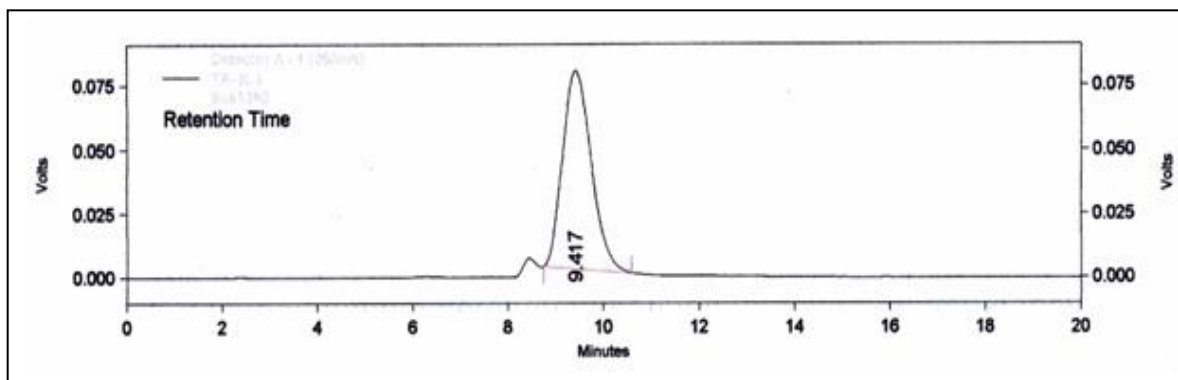
c) Chiral HPLC of **17**
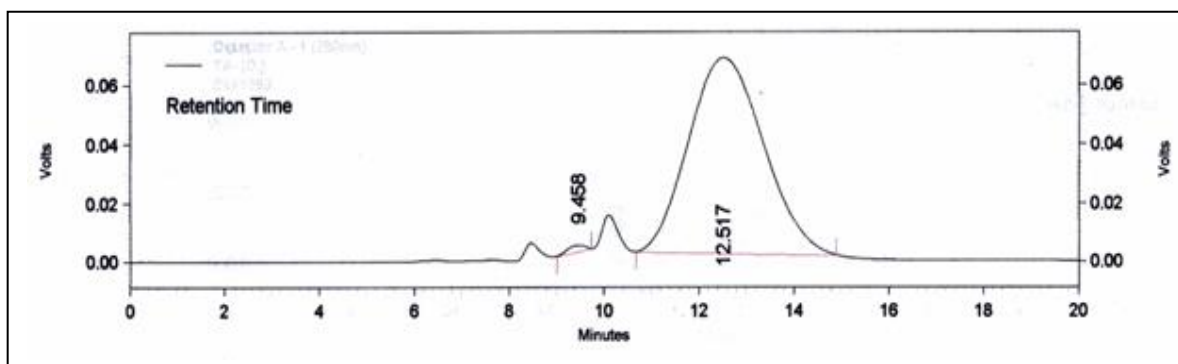
Chiral HPLC analysis of compound 8, 18

a) Chiral HPLC of racemic compound
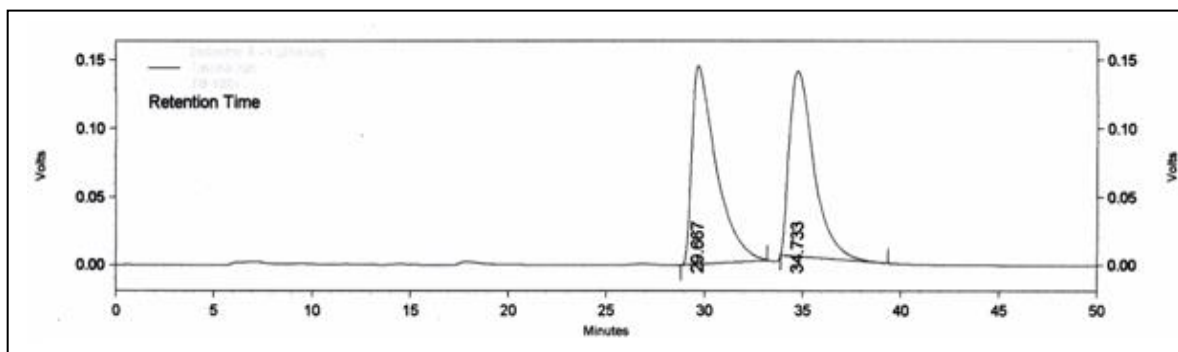
b) Chiral HPLC of **Compound 8**
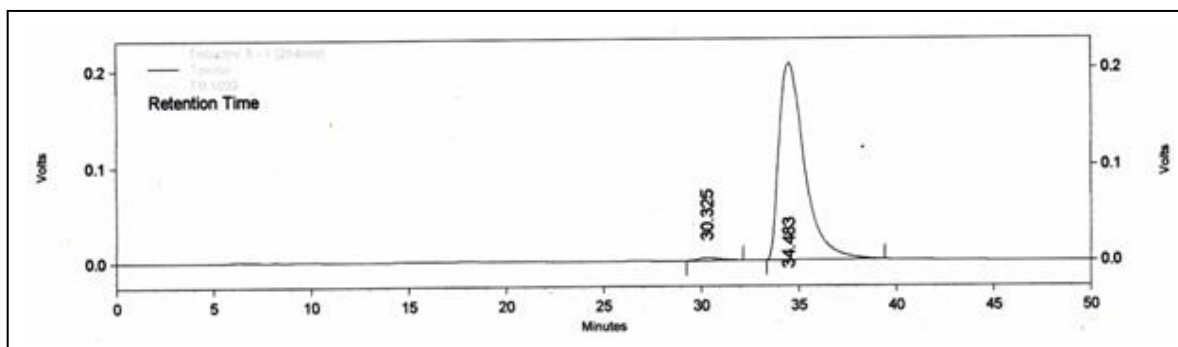
c) Chiral HPLC of **Compound 18**
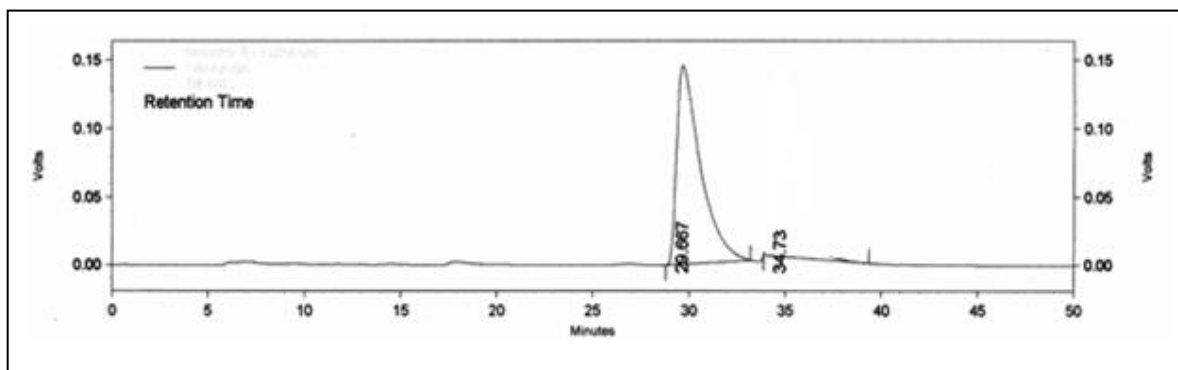
Fig HPLC chromatogram of purified ***R*- GCNA-1** oligomer
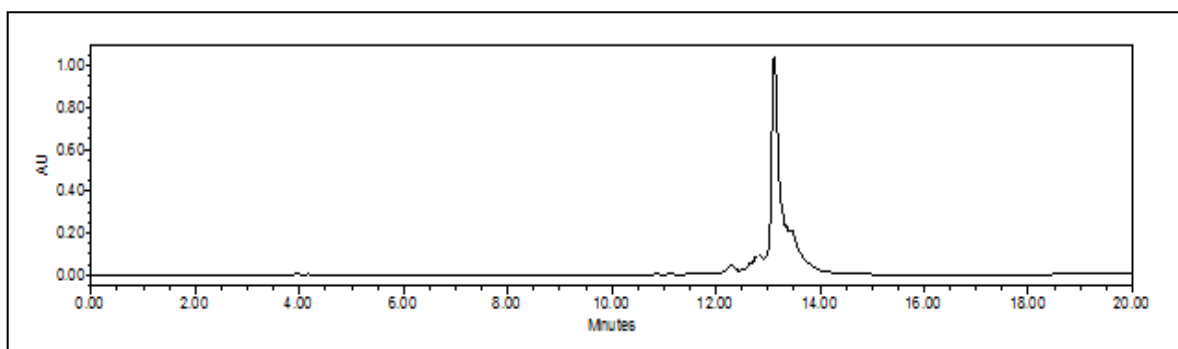
Fig HPLC chromatogram of purified **S- GCNA-1** oligomer
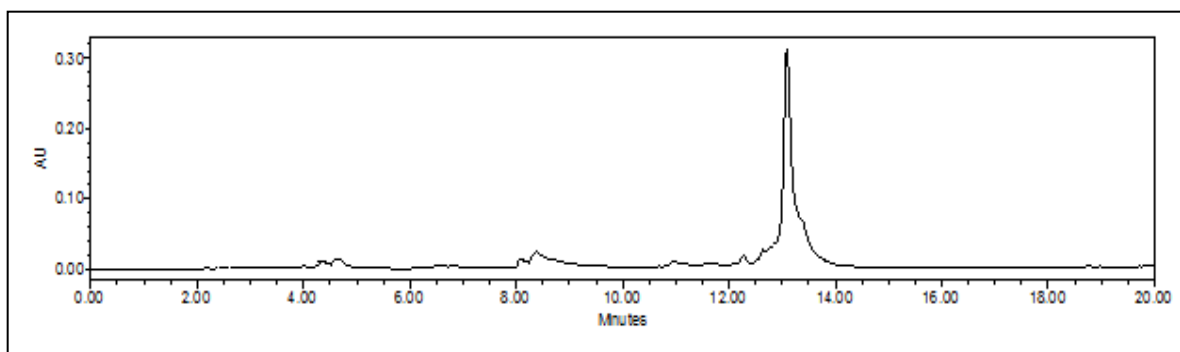
Fig HPLC chromatogram of purified **R- GCNA -2** oligomer
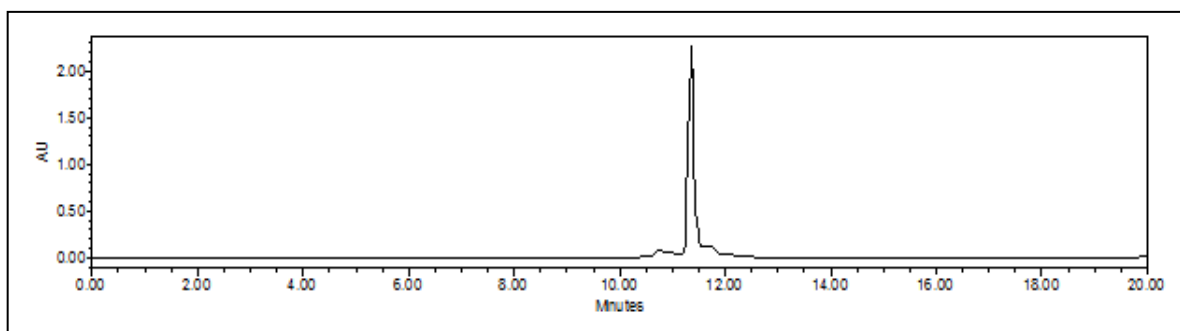
Fig HPLC chromatogram of purified **S- GCNA -2** oligomer
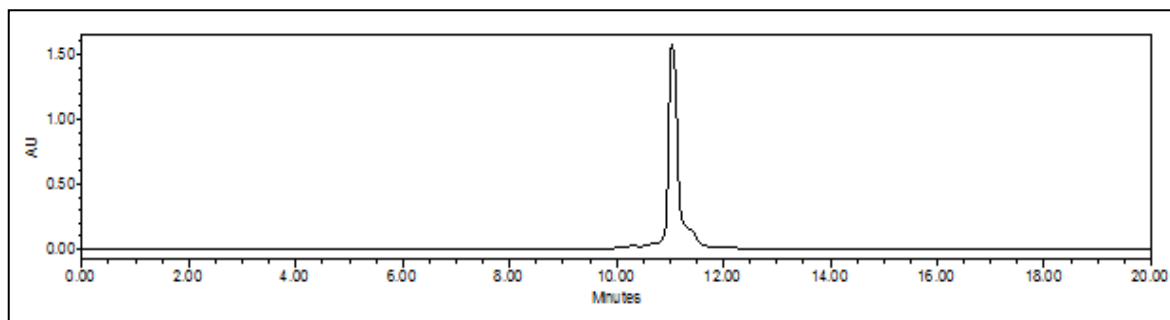
Fig HPLC chromatogram of purified *R*- **GCNA -3** oligome
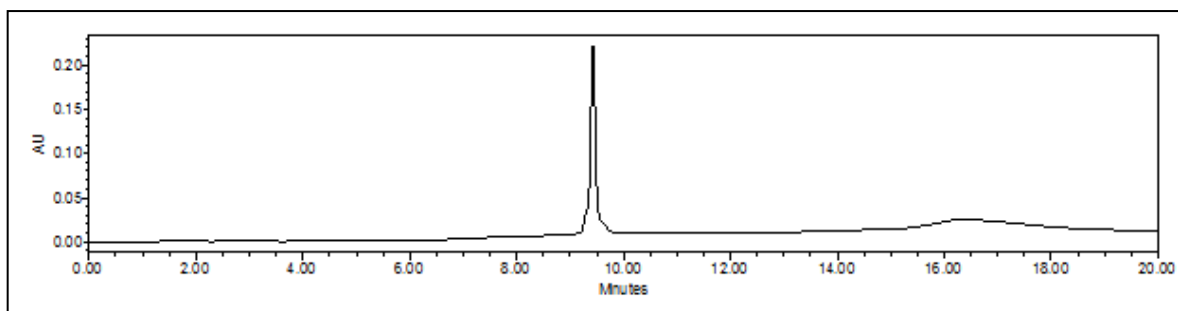
Fig HPLC chromatogram of purified *S*- **GCNA- 3** oligomer

R- **GCNA-1**
Expected MW: 1946.81
Observed MW: 1946.45

S- GCNA-2
Expected MW: 2891.07
Observed MW: 2894.94

R- GCNA-3
Expected MW: 2909.10
Observed MW: 2929.21(M+Na)

TOF/TOF™ Reflector Spec #1=>BC=>AdvBC(32,0.5,0.1)[BP = 2928.3, 1971]

S- GCNA-3
Expected MW: 2909.10
Observed MW: 2929.21(M+Na)
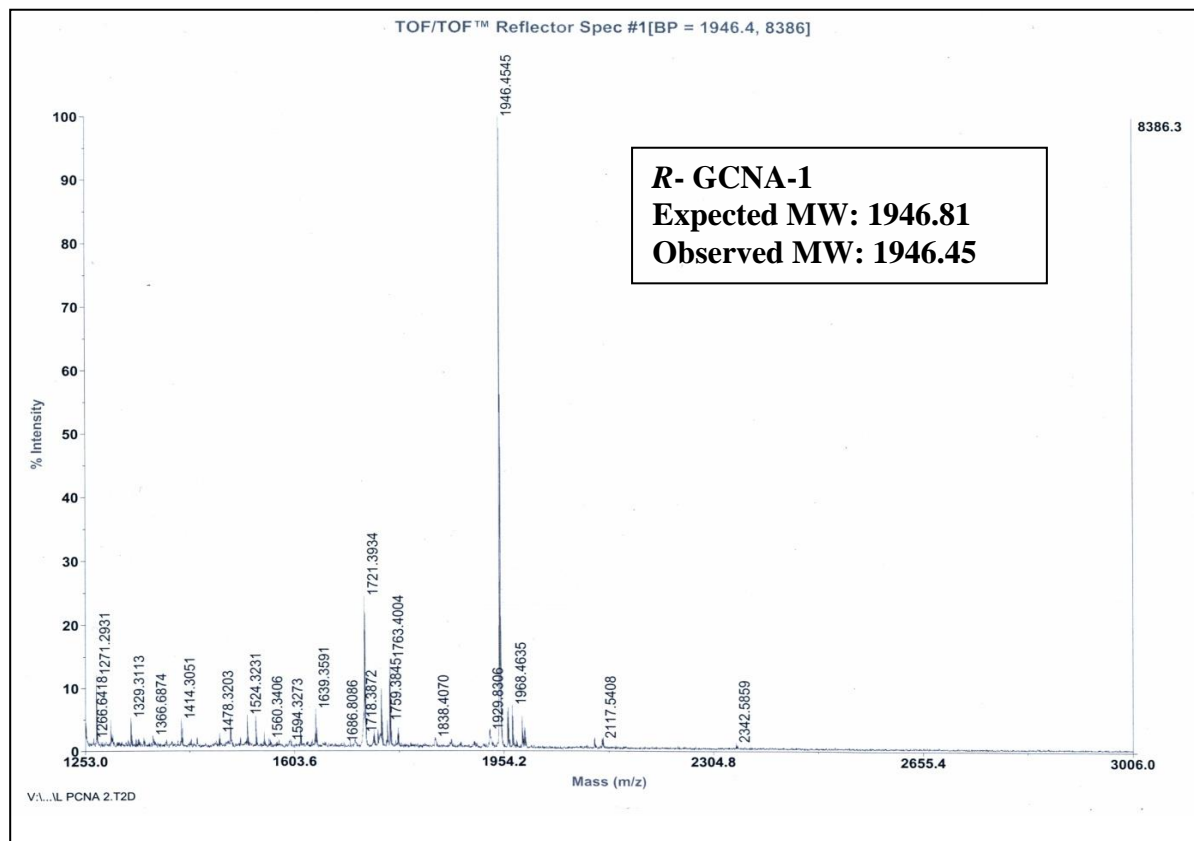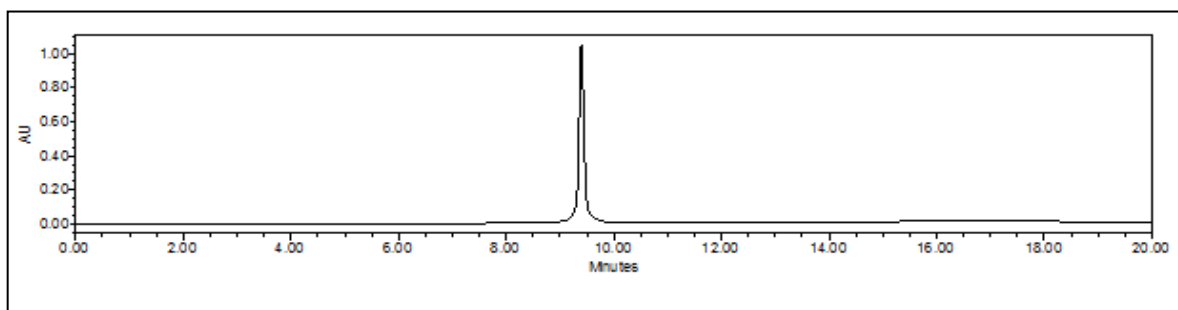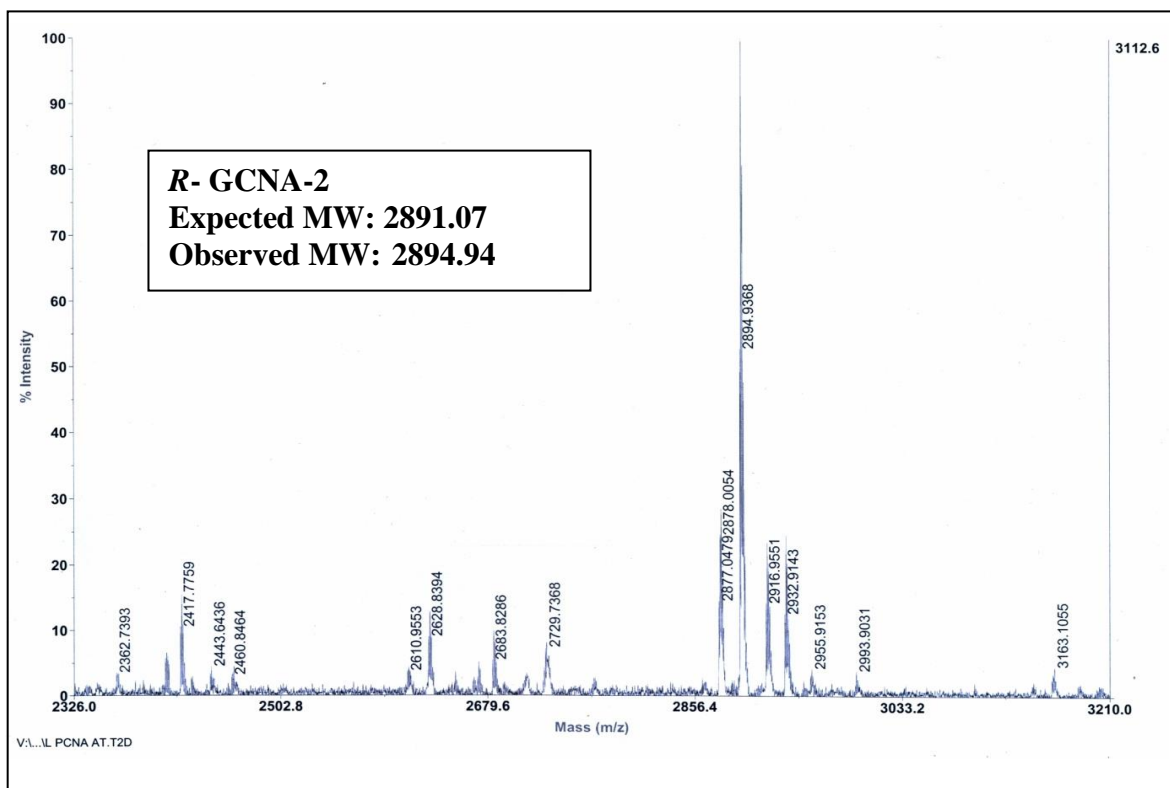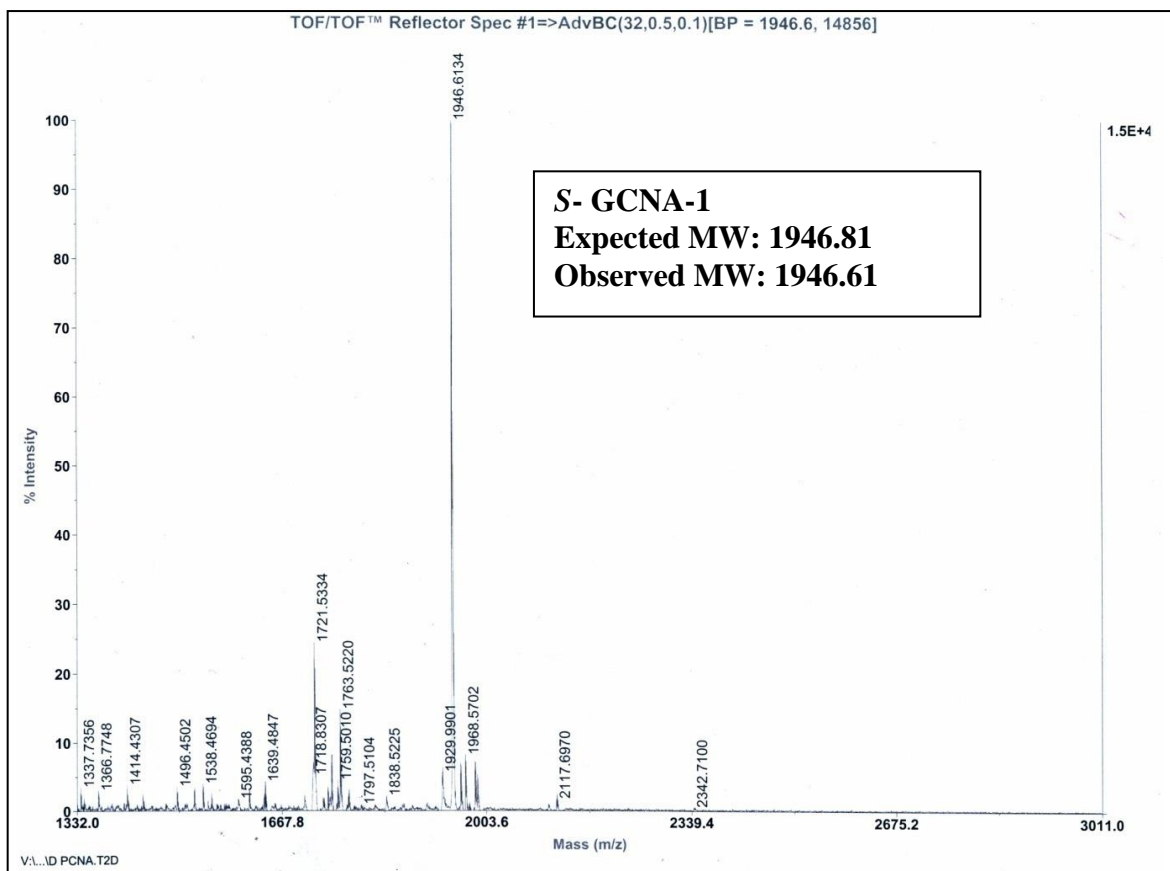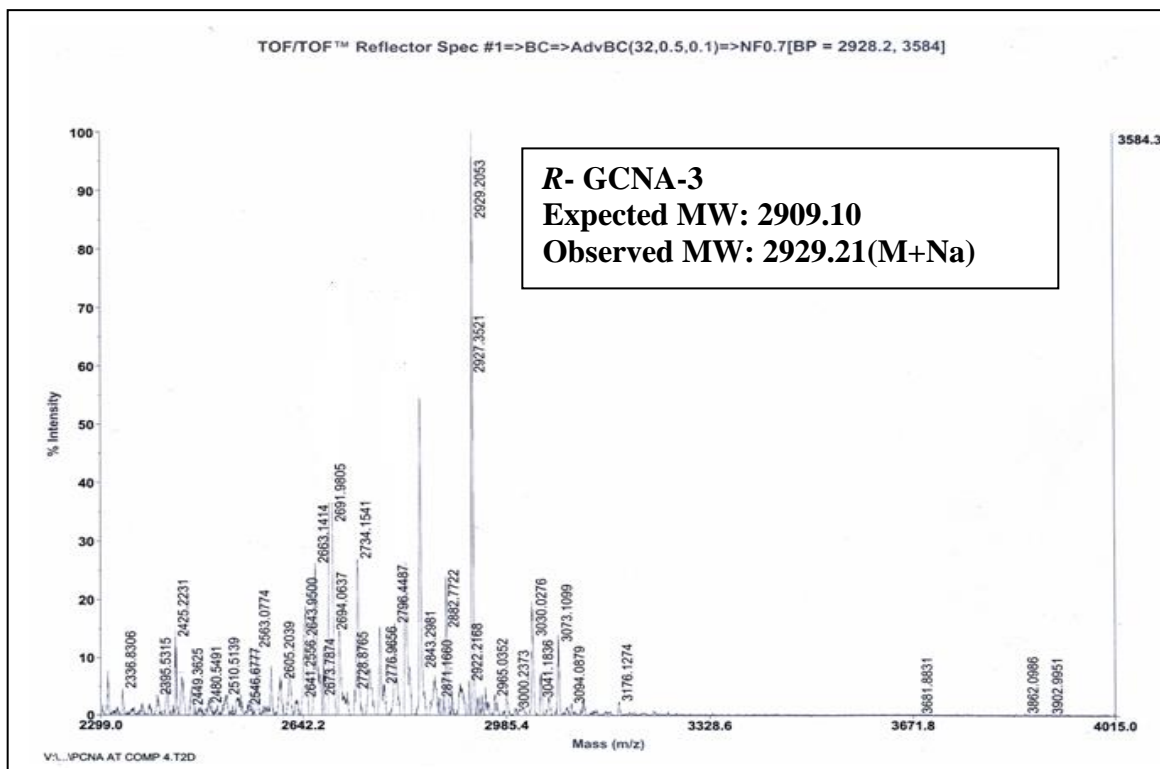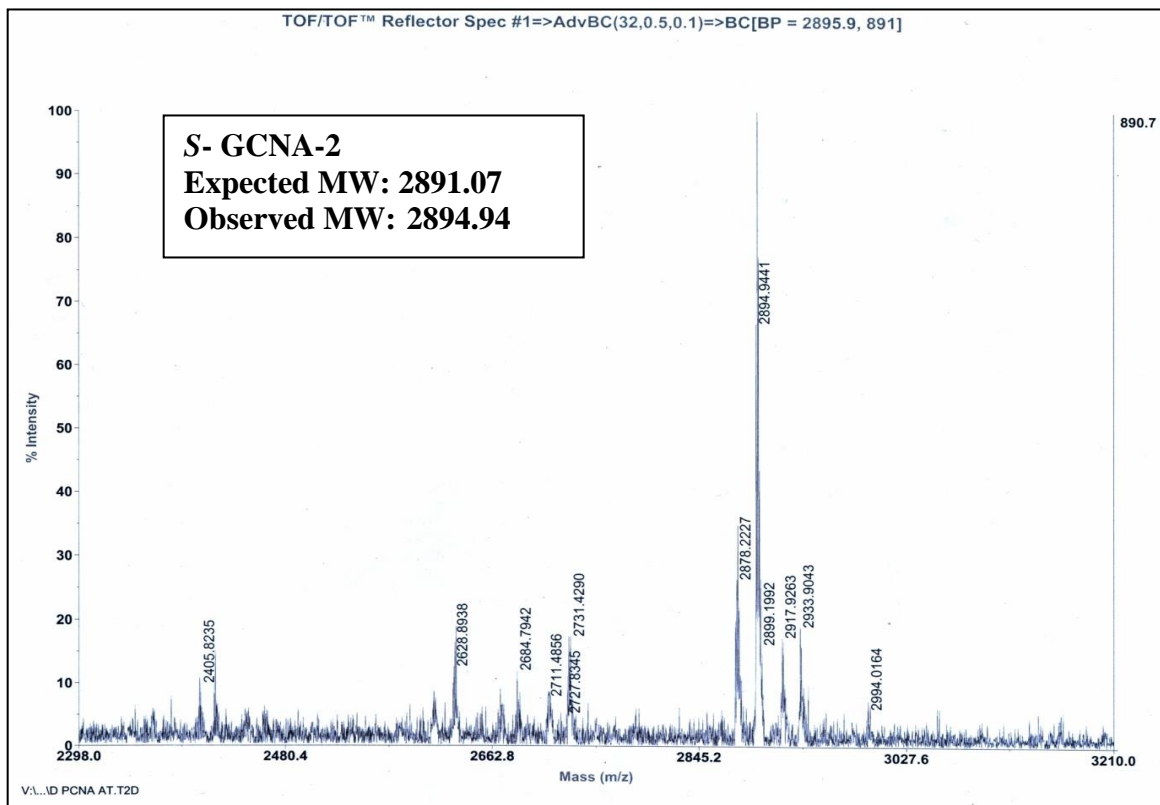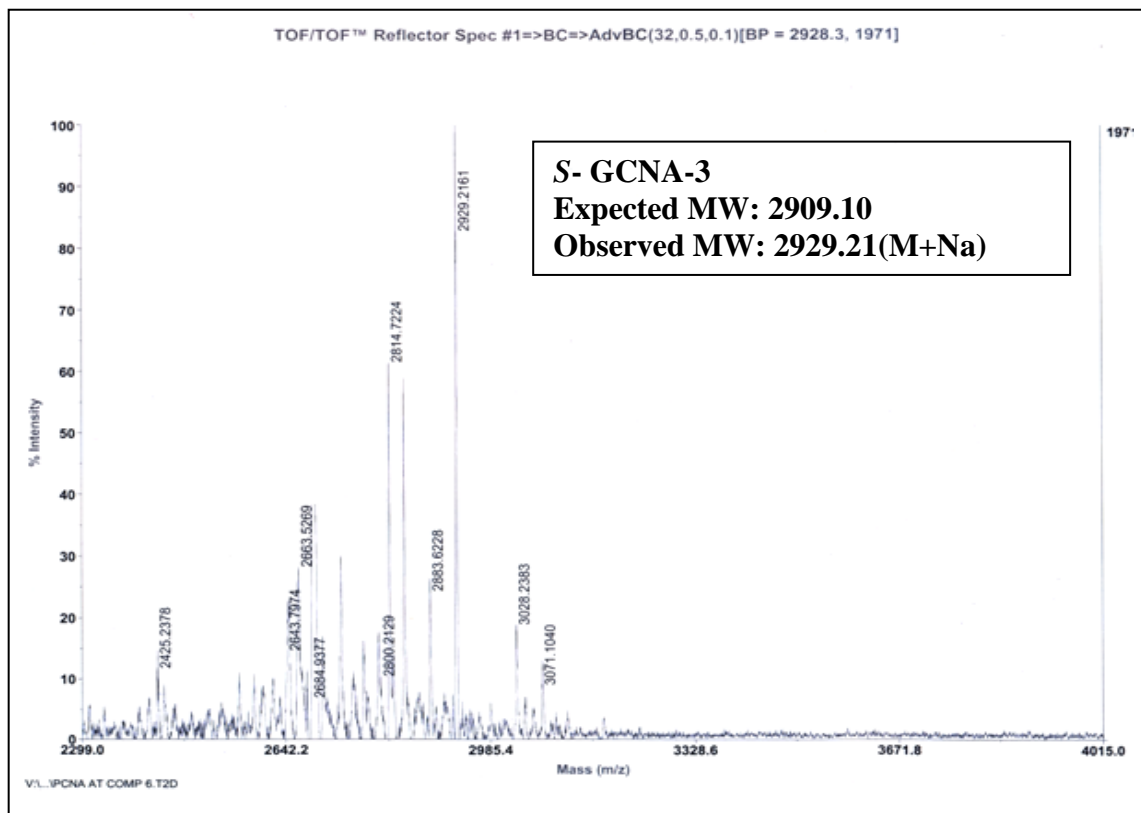
Section B

Biophysical evaluation of *R*/*S*-GCNA carbamate oligomers

2B.1 Synthesis of complementary oligonucleotides

        The complementary, parallel complementary and mismatch DNA oligomers were synthesized on Bioautomation Mer-Made 4 synthesizer using standard β-cyanoethyl phosphoramidite chemistry. The oligomers were synthesized on polystyrene solid support, followed by ammonia treatment. The purity of the oligomers was ascertained by RP HPLC on a C18 column to be more than 98% and was used without further purification in the bio-physical studies of GCNA. The complementary RNA oligonucleotides were obtained commercially.

List of DNA sequences

DNA-1: 5' TTTTTTTT 3' (Control to GCNA-1)

DNA-2: 5' ATATTATTAATT 3' (Control to GCNA-2)

DNA-3: 5' GC AAA AAA AA CG 3' (complementary (*ap*) to GCNA-1)

DNA-4: 5' AATTAATAATAT3' (complementary (*ap*) to GCNA-2)

DNA-5: 5' TATAATAATTAA 3' (complementary (*p*) to GCNA-2)

DNA-6: 5' AATTA**T**TAATAT 3' (**T: T** mismatch to GCNA-2)

List of RNA sequences

RNA-1: 5' GC AAA AAA AA CG 3' (complementary to GCNA-1)

RNA-4: 5' AAUUAAUAAUAU3' (complementary to GCNA-2)

2B.2 Binding Stoichiometry: UV Job's plot

UV Job's plot[28] was done to determine the binding stoichiometry of polypyrimidine and mixed purine/pyrimidine GCNA-1, 2. In this method, the stoichiometry of the interacting strands could be obtained from the mixing curves, in which the absorbance at a given wavelength is plotted as a function of the mole fraction of each strand which determines the stoichiometry of the complex information.[29] Various stoichiometric mixtures of GCNA and complementary DNA were made with relative molar ratios of strands 0:100, 10:90, 20:80, 30:70, 40:60, 50:50, 60:40, 70:30, 80:20, 90:10, 100:0, all at the same total strand concentration of 2 μM in sodium phosphate buffer, 10 mM NaCl, 0.1mM EDTA (10 mM, pH 7.2). The samples with the individual strands were annealed and UV absorbance was recorded.

Job's plot analysis showed that the homopyrimidine-thymine octamer sequence, *R*-GCNA-1, forms a distinct 2:1 complex with cDNA-3 d(5'-GCAAAAAAAACG-3' while, *S*-GCNA-1 did not exhibit distinct binding stoichiometry.

However, complex formation of pyrimidine/purine-mixed base sequences *R*-GCNA-2 and *S*-GCNA-2 with cDNA-4 in 1:1 ratio could be ascertained using Job's plot.
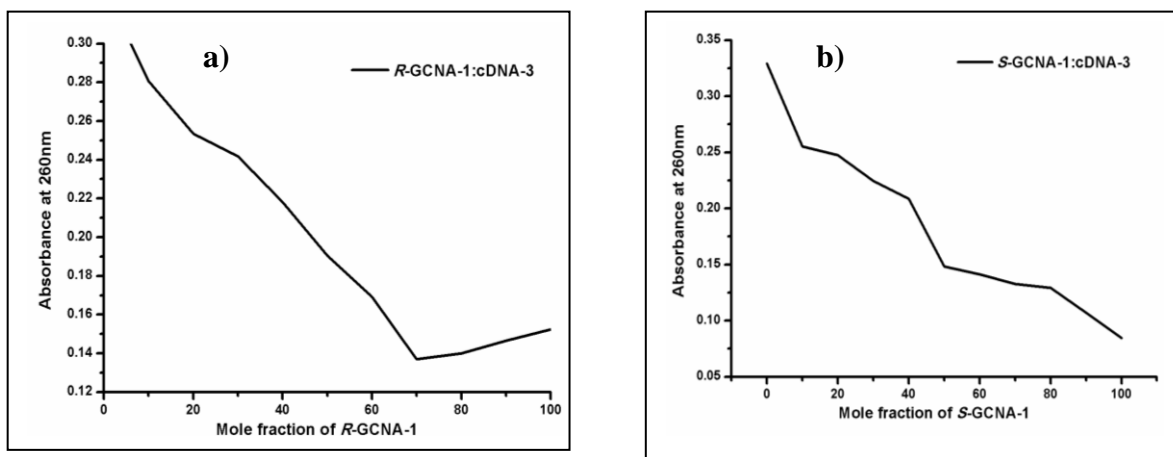
Figure 2.6: UV- absorbance (at 260 nm) of a) *R*- GCNA-1: cDNA-3 b) *S*- GCNA-1: cDNA-3
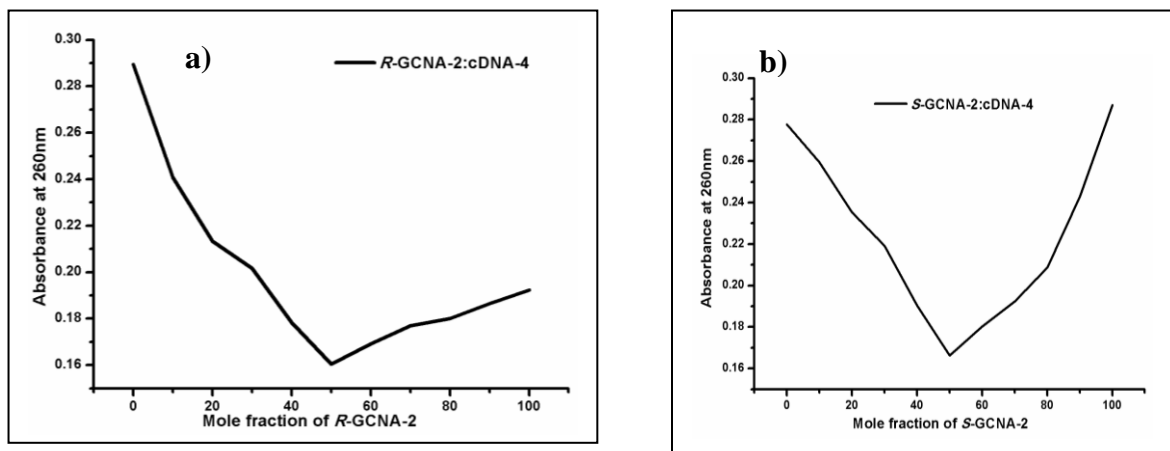
Figure 2.7: UV-absorbance (at 260 nm) of a) *R*- GCNA-2: cDNA-4 b) *S*- GCNA-2: cDNA-4

2B.3 UV melting Experiment

UV melting experiment has provided valuable information regarding complementary interactions in nucleic acid hybridizations involving DNA, RNA and the modified complementary oligomers. This technique has also been used for studying thermal stability of uncharged DNA mimics such as PNA and PCNA.

2B.3.1 UV-T_m studies of *R/S*-GCNA-1 with cDNA-3/cRNA-3

UV meltings were also recorded for glycol carbamate modified oligomers. *R*-GCNA-1: cDNA-3 complex showed UV melting temperature (UV-T_m) of 35.3°C. The complex of *R*-GCNA-1 with cRNA-3 r(5'-GCAAAAAAAACG-3') was found to be comparatively less stable (UV-T_m = 29.0°C). On contrary, *S*-GCNA-1 did not form complex with cDNA-3 which was also clearly evident from the Job's plot. However, *S*-GCNA-1 showed a UV-T_m of 32.1°C with cRNA-3. DNA-1 was found neither to bind with cDNA-3 or cRNA-3.

Table 2: UV-T_m values for homothymidine (*R/S*)-GCNA-1 at 10mM salt concentration

	UV-Tm (°C)	
	cDNA-3	**cRNA-3**
R-**GCNA-1**	35.3	29.0
S-**GCNA-1**	n.t.	32.1
DNA-1	n.t.	n.t.
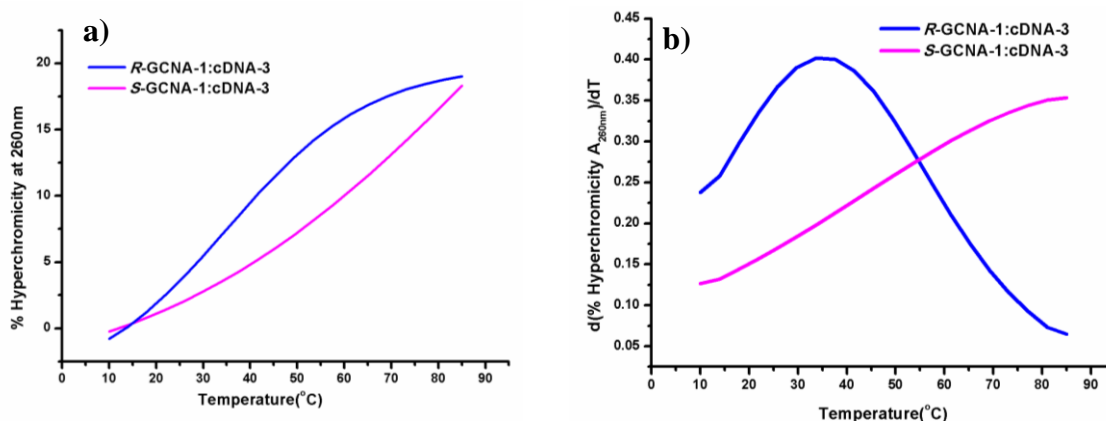
Figure 2.8: UV-thermal melting profiles of a) (*R/S*)-GCNA-1:cDNA-3 , first derivative curves for the sigmoidal melting curves of b) (*R/S*)-GCNA-1:cDNA-3
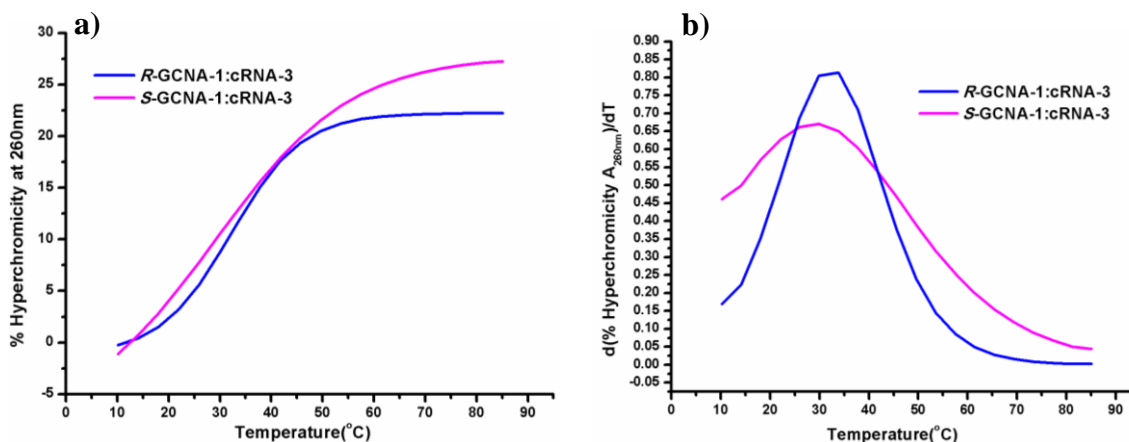
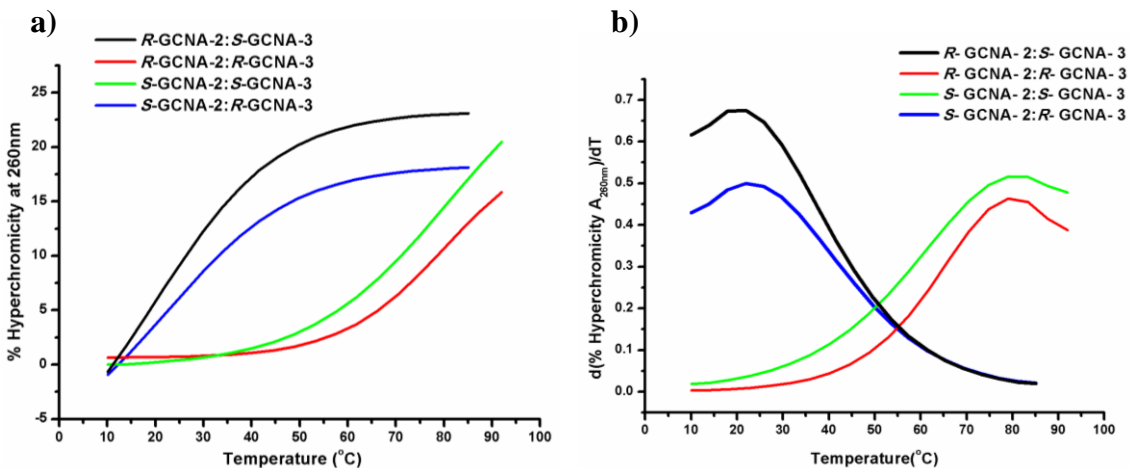**Figure 2.9: UV-thermal melting profiles of a) (*R/S*)-GCNA-1:cRNA-3, first derivative curves for the sigmoidal melting curves of b) (*R/S*)-GCNA-1:cRNA-3 at 10mM salt concentration.**

2B.3.2 UV-T_m studies of *R/S*-GCNA-2 with cDNA-4/cRNA-4

The duplex formation of pyrimidine/purine-mixed base sequences *R*-GCNA-2 and *S*-GCNA-2 with cDNA-4 and cRNA-4 were investigated. The transition in the case of *R*-GCNA-2: cDNA-4 was found to be sigmoidal with 20% hyperchromicity. The results shown in **Table 3** indicate the preference of *R*-GCNA-2 to bind complementary nucleic acids as compared to the binding of carbamate oligomer with *S*-chirality, i.e. *S*-GCNA-2. Further, the *R*-GCNA-2 prefers complexation with cDNA-4, better than that with complementary cRNA-4, whereas the *S*-GCNA-2 binds with almost equal strength to both complementary DNA/RNA but with less T_m compared to the *R*-stereochemistry of the oligomer. Compared to the duplex UV-T_m values of natural DNA: cDNA or DNA: cRNA (DNA-2:cDNA-4 and DNA-2:cRNA-4) the cross-pairing of *R*-GCNA-2 with cDNA-4 and cRNA-4 shows ΔT_m of +16°C and +7°C respectively. This difference is somewhat less for *S*-GCNA-2 (+5-6ºC). Although, the stereocentre in *R*-GCNA-2 cannot be directly related to the GNA/*S*-TNA stereocentre, its preference while forming cross-pairing to the natural nucleic acids is noteworthy. Although, unlike TNA, where binding with DNA was found to be less compared to that with RNA, the open-chain carbamate GCNA prefers binding to DNA better. This was also observed in our previous studies with PCNA[18] where the nucleobase was four bonds away from the backbone. Fortunately, because of the flexibility of the acyclic structure in the present *R/S*-GCNA could mould itself to cross-pair with RNA also, with some cost in binding strength. Despite this, the complexes of GCNA with RNA are still with

higher strength than those of DNA: cRNA duplexes. The single base mismatch studies were carried out to establish proper W-C base pairing in cross-paired duplexes

Table 3: UV-T_m values for mixed pu/py (R/S)-GCNA-2 at 10mM salt concentration

|	UV-T_m (°C)	
	cDNA-4	**RNA-4**
R-GCNA-2 atattattaatt-Lys	40.8	29.6
S-GCNA-2 atattattaatt-Lys	28.3	26.5
DNA-2 5'ATATTATTAATT 3'	n.t.	n.t.
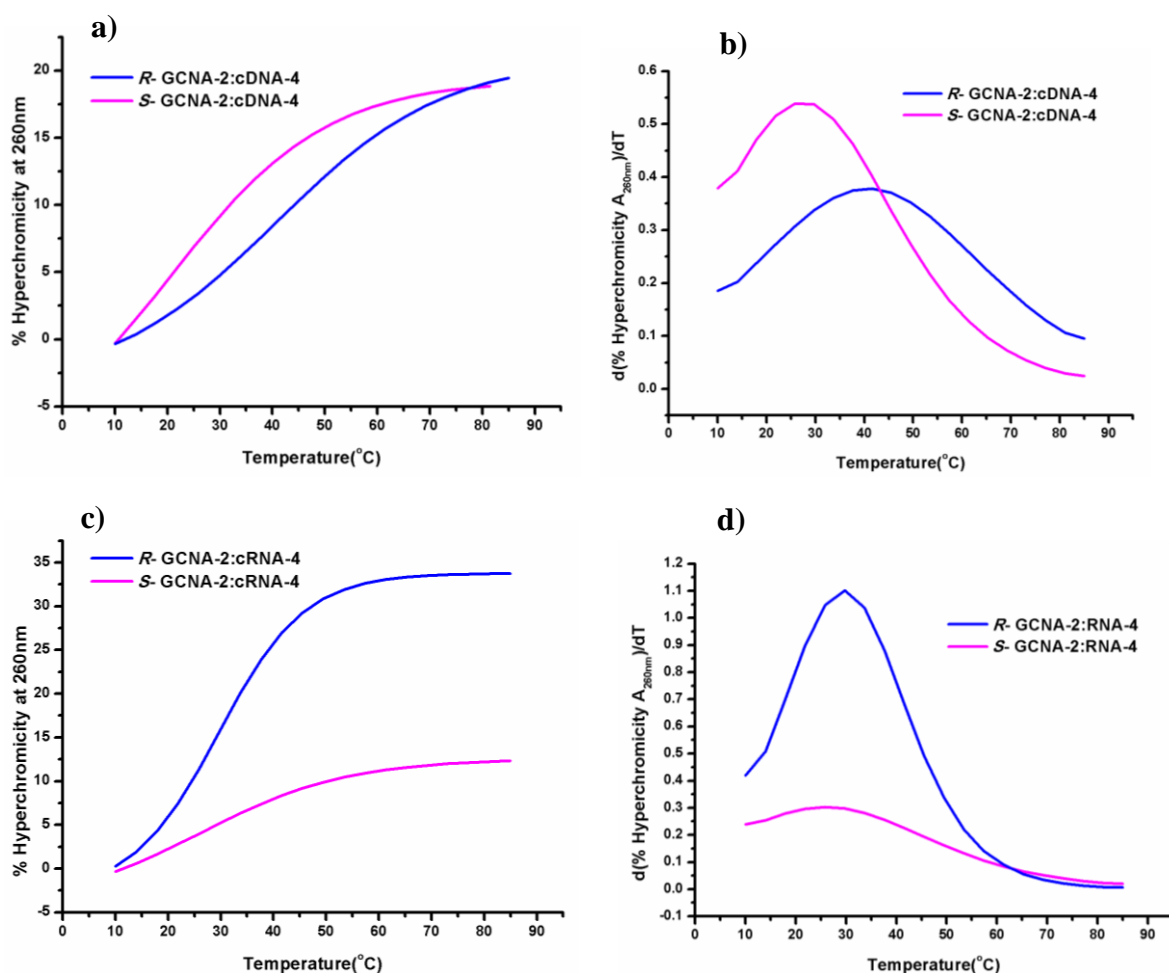
Figure 2.10: UV-Thermal melting profiles. Sigmoidal curves for a) (R/S)-GCNA-2:cDNA-4 c) (R/S)-GCNA-2:cRNA-4, first derivative curves for the sigmoidal melting curves of b) (R/S)-GCNA-2:cDNA-4 d) (R/S)-GCNA-2:cRNA-4 at 10mM salt concentration.

Table 4: UV-T_m values for mixed pu/py (*R/S*)-GCNA-2 at 100mM salt concentration

	UV-T_m (°C)	
	cDNA-4	**cRNA-4**
R-**GCNA-2:**atattattaatt-Lys	39.6	30.0
S-**GCNA-2:**atattattaatt-Lys	27.5	26.5
DNA-2: 5'ATATTATTAATT 3'	23.6	24.6
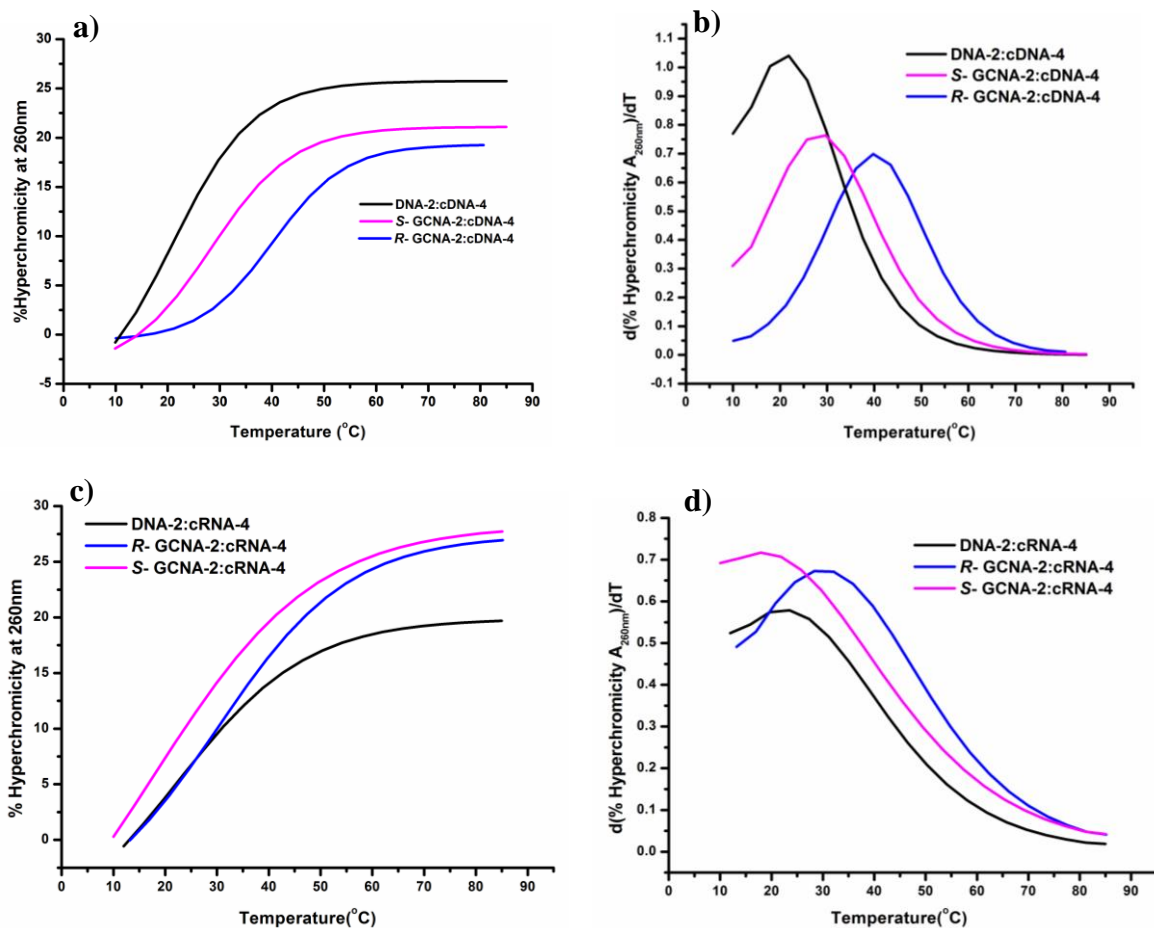
Figure 2.11: UV-Thermal melting profiles. Sigmoidal curves for a) (*R/S*)-GCNA-2:cDNA-4 c) (*R/S*)-GCNA-2:cRNA-4, first derivative curves for the sigmoidal melting curves of b) (*R/S*)-GCNA- 2:cDNA-4 d) (*R/S*)-GCNA- 2:cRNA-4 at 100mM salt concentration

Table 5: UV-T_m values for mixed pu/py (*R/S*)-GCNA-2 at 2µM strand concentration

	UV-T_m (°C)	
	cDNA-4	**cRNA-4**
R-**GCNA-2:** atattattaatt-Lys	41.5	30.3
S-**GCNA-2:** atattattaatt-Lys	27	28.5
DNA-2: ATATTATTAATT	22.5	21.8
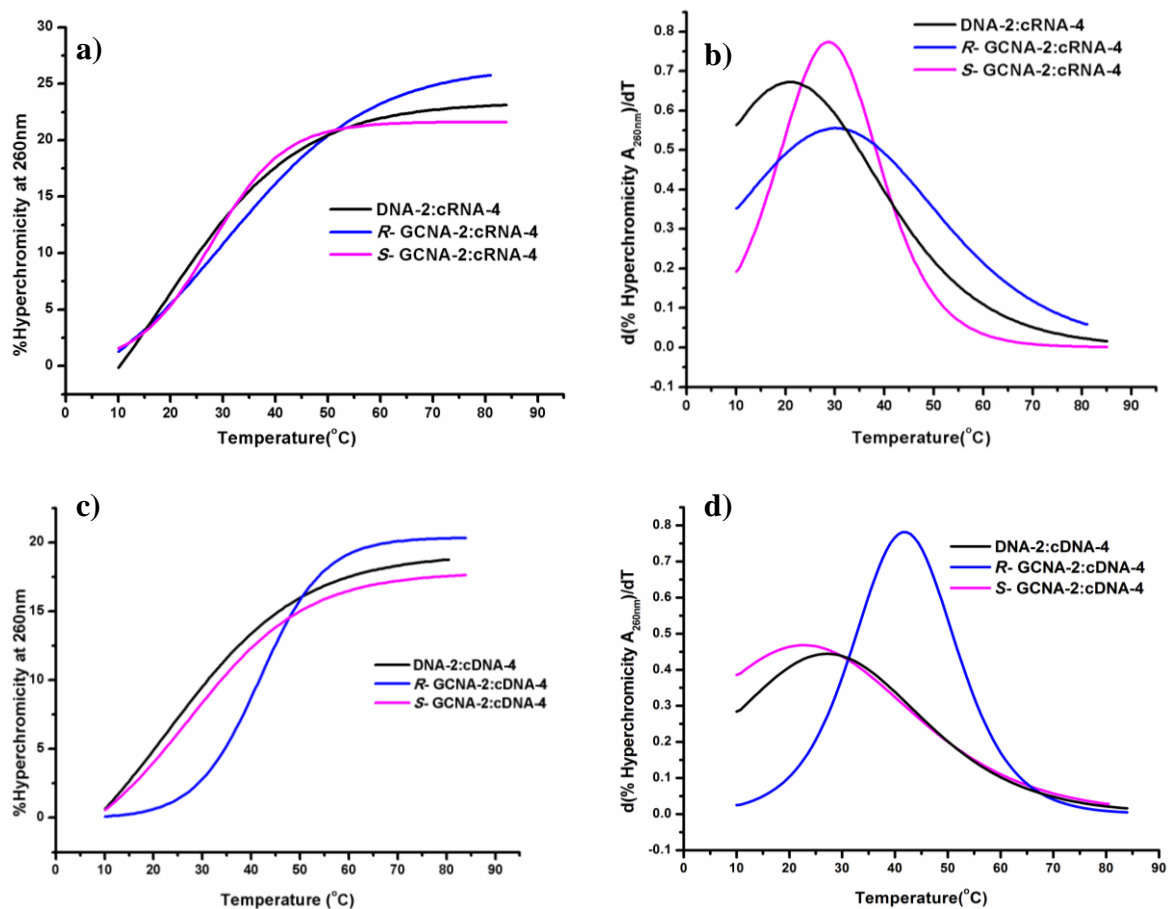
Figure 2.12: UV-Thermal melting profiles. Sigmoidal curves for a) (*R/S*)-GCNA-2:cDNA-4 c) (*R/S*)-GCNA-2:cRNA-4, first derivative curves for the sigmoidal melting curves of b) (*R/S*)-GCNA-2:cDNA-4 d) (*R/S*)-GCNA-2:cRNA-4 at 100mM salt concentration

We also studied the stability of the complex with complementary RNA, single base mismatched DNA and with DNA in parallel mode assuming N→C direction corresponding to 5'-3'direction in natural oligomers. The chirality of the sequences could be important determinant of the direction of binding. The melting temperatures were derived from the first derivative of the plot. The single base mismatch studies were carried out to establish proper W-C base pairing in cross-paired duplexes. A single mismatch in the dodecamer resulted in 11-14℃ reduction in UV-T_m values for *R*-GCNA-2: mmDNA-6 and *S*-GCNA-2: mmDNA-6 duplexes (**Table 6**).

Table 6: Single strand meltings of (*R/S*)-GCNA-2 with *p*-cDNA-4, mmDNA-6

	UV-T_m (°C)	
	p-cDNA-4	mmDNA-6
R-GCNA-2:atattattaatt-Lys	26.1	26.4
S-GCNA-2:atattattaatt-Lys	n.t.	16.5
DNA-2: 5'ATATTATTAATT 3'	n.t.	n.t.

a)

b)

c)

d)
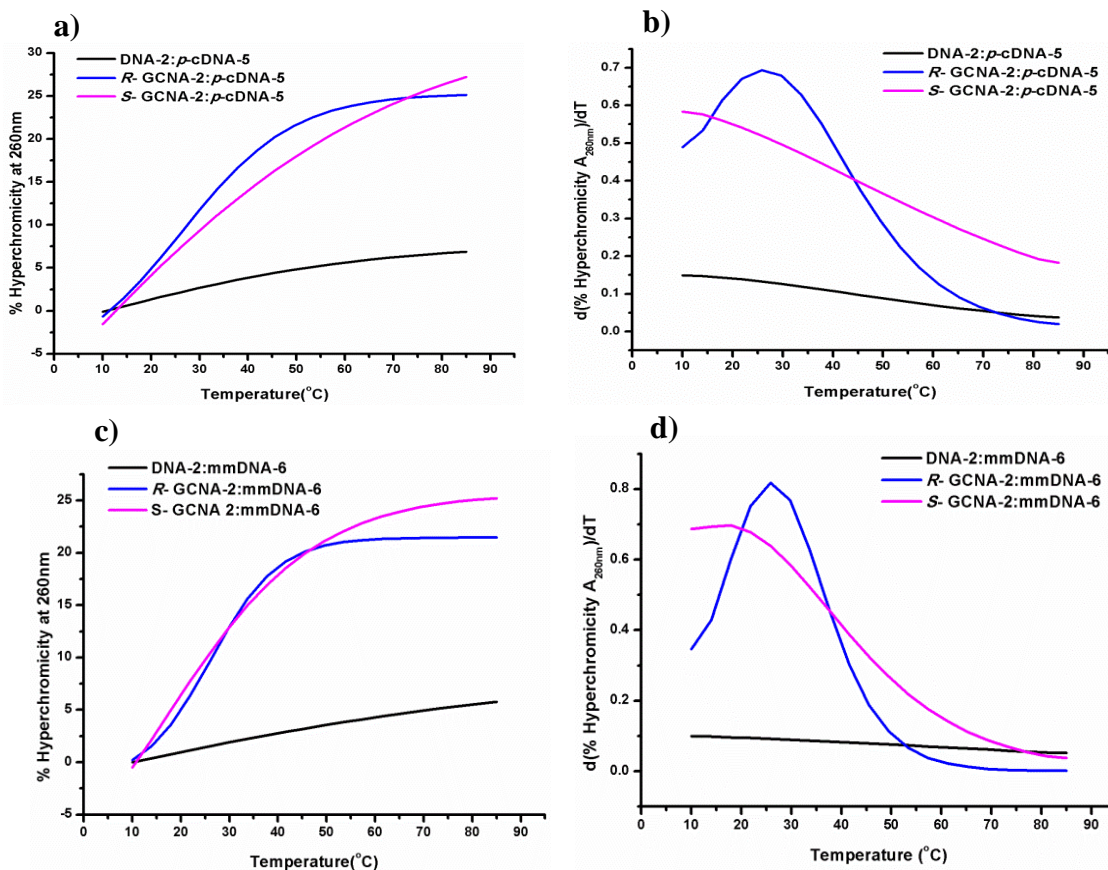
Figure 2.13: UV-Thermal melting profiles. Sigmoidal curves for a) (*R/S*)-GCNA-2:*p*-cDNA-5 c) (*R/S*)-GCNA-2: mmDNA-6, first derivative curves for the sigmoidal melting curves of b) (*R/S*)-GCNA-2:*p*-cDNA-5 d) (*R/S*)-GCNA-2: mmDNA-6 at 100mM salt concentration

Investigations of binding properties of self-pairing of these simplified GCNA oligomers were then taken up. The antiparallel complementary sequences *R*-GCNA-3 and *S*-GCNA-3 were synthesized. Analogous to GNA, as expected, these oligomers specifically exhibited strong binding only to complementary sequences with same chirality. *R*-GCNA-2:*R*-GCNA-3 and *S*-GCNA-2:*S*-GCNA-3 complexes were stable upto ~80-81°C, whereas the *S*-GCNA-2: *R*-GCNA-3 and *R*-GCNA-2: *S*-GCNA-3 exhibited UV-T_m values only for that

observed for the individual single *S*-GCNA strands. The higher binding strength of the self-paired dodecamer GCNA-2: GCNA-3, was found to be further augmented in the absence of negative charge repulsions in GCNA.

Table 7: UV-T_m values for (*R/S*)-GCNA-2 with (*R/S*)-GCNA-3 at 10 mM NaCl and 10% DMSO

Sequences	UV-T_m (°C)	
	***R*-GCNA-3** aattaataatat-Lys	*S*-GCNA-3 aattaataatat-Lys
***R*- GCNA-2** atattattaatt-Lys	>80.0	20.1
***S*- GCNA-2** atattattaatt-Lys	21.6	>81.0

Figure 2.14: UV-Thermal melting profiles. Sigmoidal curves for a) (*R/S*)-GCNA- 2: (*R/S*)-GCNA- 3, first derivative curves for the sigmoidal melting curves of b) (*R/S*)-GCNA- 2: (*R/S*)-GCNA- 3 in 10% DMSO at 10mM salt concentration

Table 8: Single strand meltings of (*R/S*)-GCNA-2, (*R/S*)-GCNA-3

Sequences	*R*- GCNA-2 ss	*S*- GCNA-2 ss	*R*- GCNA-3 ss	*S*- GCNA-3 ss
UV-T_m (°C)	n.t.	20.07	n.t.	20.07
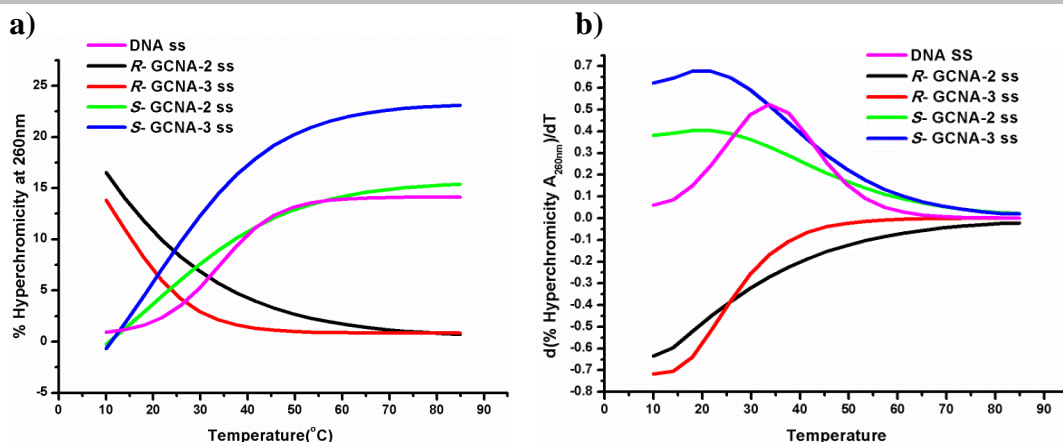
Figure 2.15: UV-Thermal melting profiles. Sigmoidal curves for a) (*R/S*)-GCNA- 2 ss and (*R/S*)-GCNA-3 ss, first derivative curves for the sigmoidal melting curves of b) (*R/S*)-GCNA- 2 ss and (*R/S*)-GCNA-3 ss at 100mM salt concentration

2B.4 Circular Dichroism Studies

Circular Dichroism is a spectroscopic technique used extensively to study chiral molecules specially large biomolecules.This spectrum is observed not only due to chirality of the molecule but also due its three dimensional structure. The crests and troughs observed on recording CD spectra is known as Cotton effect. Enantiomers are known to show opposite cotton effect.

CD spectra of compound **8, 18** were recorded at room temperature. The CD spectra of the enantiomeric monomers exhibited opposite cotton effect of equal magnitude at 275nm (**Figure 2.15**).
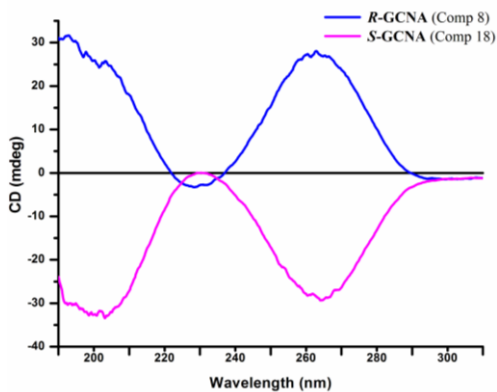
Figure 2.16: CD spectra of enantiomeric monomers 8, 18

The CD plots of the single stranded *R*-GCNA-2 and isomeric *S*-GCNA-2 were also opposite in cotton effect at ~275 nm, showing differential stacking interactions in single strands. However, the spectra were not exact mirror images as observed for the *R/S*-monomers. This could be because of the presence of L-lysine at *C*-terminus of the sequences.
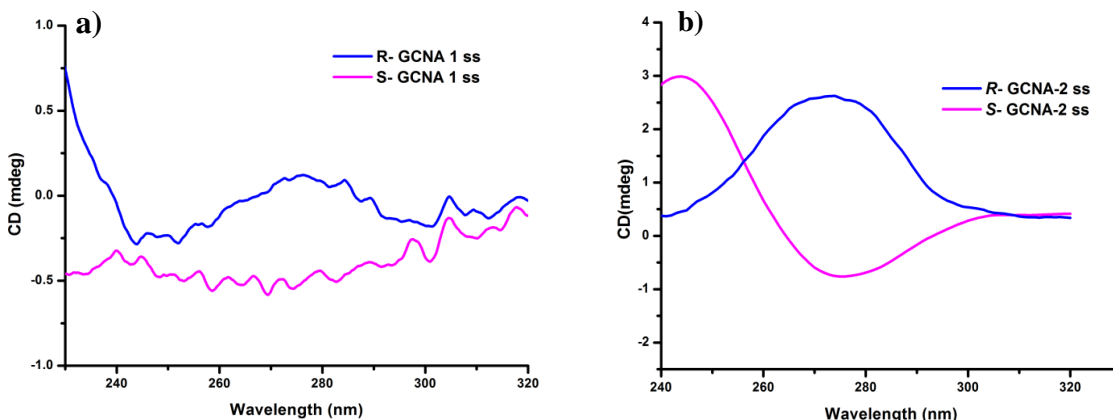
Figure 2.17: CD spectra of a) *R*-GCNA-1 ss, *S*-GCNA-1 ss, **b)** *R*-GCNA-2 ss, *S*-GCNA-2 ss. **Experiments were done in buffer containing 10mM sodium phosphate, 10mM sodium chloride, pH 7.2, using 5µM concentration of strands at 10°C.**

The CD studies were then undertaken for supplementing the formation of the highly stable self-paired duplexes of mixed pu/py sequences. The representative CD curves corresponding to the *R*-stereochemistry are shown in **Figure 2.17**.
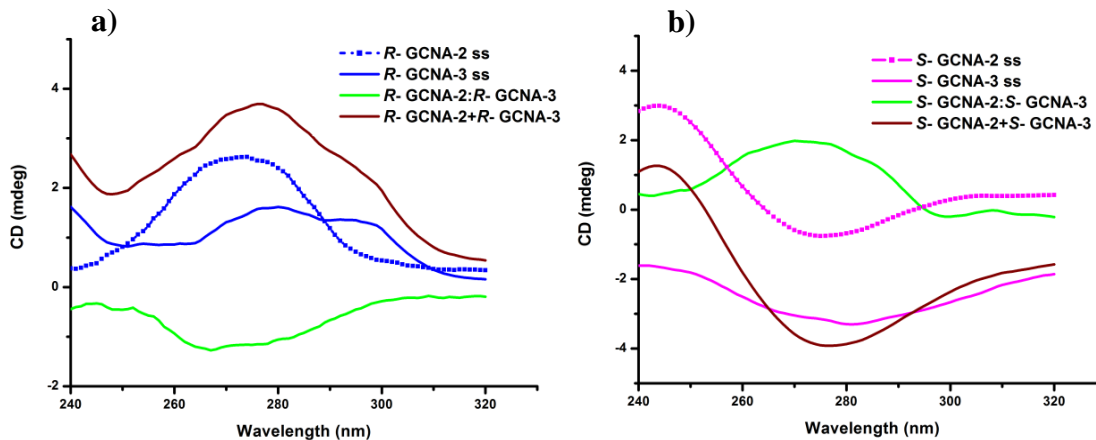
Figure 2.18: Representative CD plots of a) *R*-GCNA-2 ss, *R*-GCNA-3 ss, the addition spectrum of (*R*-GCNA-2 ss+*R*-GCNA-3 ss) and complex *R*-GCNA-2:*R*-GCNA-3 **b)** *S*-GCNA-2 ss, *S*-GCNA-3 ss, the addition spectrum of (*S*-GCNA-2 ss+*S*-GCNA-3 ss) and complex *S*-GCNA-2:*S*-GCNA-3. **Experiments were done in buffer containing 10mM sodium phosphate, 10mM sodium chloride, pH 7.2, using 5µM concentration of each of the two complementary strands and 5µM concentration each of the single strands at 10°C.**

The addition spectrum of the two is same as the individual oligomers. Interestingly, the unique feature observed is that the CD of the duplex *R*-GCNA-2:*R*-GCNA-3 is reversed and CD maximum signal is observed at 280 nm. Opposite trend was observed for the *S*-GCNA-2 and *S*-GCNA-3 single strands and the duplex *S*-GCNA-2: *S*-GCNA-3. The complete reversal of the CD on duplex formation suggests differentially base-stacked structures in single strands and the duplexes. The addition spectra of the individual oligomers were found to be distinctly different than that of the complexes confirming the duplex formation.

2B.5 Conclusion

- The formation of highly stable self-paired and moderately stable cross-paired complexes with natural nucleic acids is the unique feature of this backbone for artificial nucleic acid. The chirality of the monomers affected the binding properties while cross pairing.

- The highly stable homochiral homogeneous complexes displayed differently configured duplexes. *R*-GCNA gave incremental benefit for RNA binding which could be of high value for the applications of this chemically evolved backbone in nucleic acids chemistry[18] and structural DNA nanotechnology.[30]

2B.6 References

1. A. Eschenmoser, *Science,* 1999, **284**, 2118.

2. K. U.Schoning, P. Scholz, S. Guntha, X. Wu, R. Krishnamurthy and A. Eschenmoser, *Science,* 2000, **290**, 1347.

3. P. S. Pallan, C. J. Wilds, Z. Wawrzak, R. Krishnamurthy, A. Eschenmoser and M. Egli, *Angew. Chem. Int. Ed.*, 2003, **42**, 5893.

4. L. L. Zhang, A. Peritz and E. Meggers, *J. Am. Chem. Soc.*, 2005, **127**, 4174.

5. E. Meggers and L. Zhang, *Acc. Chem. Res.*, 2010, **43**, 1092.

6. K. C. Schneider and S. A. Benner, *J. Am. Chem. Soc.,* 1990, **112**, 453.

7. P. Karri, V. Punna, K. Kim and R. Krishnamurthy, *Angew. Chem. Int. Ed.*, 2013, **52**, 5840.

8. a) H. Asanuma, T. Toda, K. Murayama, X. Liang and H. Kashida, *J. Am. Chem. Soc.,* 2010, **132**, 14702. b) K. Murayama, Y. Tanaka, T. Toda, H. Kashida and H. Asanuma, *Chem. Eur. J.*, 2013, **19**, 14151.

9. H. Kashida, K. Murayama, T. Toda and H. Asanuma, *Angew. Chem. Int. Ed.,* 2011, **50**, 1285.

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

11. P. E. Nielsen, *Acc. Chem. Res.*, 1999, **32***,* 624.

12. V. V. Demidov, V. N. Potaman, M. D. Frank-Kamenetskii, M. Egholm, O. Buchardt and S. H. Sonnichsen, *Biochem. Pharmacol.*, 1994, **48**, 1310.

13. P. E. Nielsen and M. Egholm, *Current Issues Molec. Biol.*, 1999, **1**, 89.

14. M. J. Gait, A. S. Jones and R. T. Walker, *J. Chem. Soc. Perkin Trans.*, 1974, **1**, 1684.

15. W. S. Mungall and J. K. Kaiser, *J. Org. Chem.*, 1977, **42**, 703.

16. Meena and V. A. Kumar, *Bioorg. Med. Chem.*, 2003, **11**, 3393.

17. V. Madhuri and V. A. Kumar, *Org. Biomol. Chem.*, 2010, **8**, 3734.

18. V. Kotikam, M. Fernandes and V. A. Kumar, *Phys. Chem. Chem. Phys.*, 2012, **14**, 15003.

19. E. P. Stirchak, J. E. Summerton and D. D. Weller, *J. Org. Chem.*, 1987, **52**, 4202.

20. a) M. J. Gait, A. S. Jones and R. T. Walker, *Journal of the Chemical Society. Perkin Trans.* 1, 1974, **1684**. b) J. M. Coull, D. V. Carlson and H. L. Weith, *Tetrahedron Lett.*, 1987, **28**, 745. d) E. P. Stirchak, J. E. Summerton and D. D. Weller, *Nucleic Acids Research,* 1989, **17**, 6129. e) M. Prhavc, E. A. Lesnik, V. Mohan and M. Manoharan, *Tetrahedron Lett.,* 2001, **42**, 8777.

21. F. Hu, Y. H. Zhang and Z. J. Yao, *Tetrahedron Lett.*, 2007, **48**, 3511.

22. R. B. Merrifield, *J. Am. Chem. Soc.*, 1963, **85**, 2149.

23. K. L. Dueholm, M. Egholm, C. Behrens, L. Christensen, H. F. Hensen, T. Vulpius, K. H. Peterson, R. H. Berg, P. E. Nielsen and O. Buchardt, *J. Org. Chem.*, 1994, **59**, 5767.

24. B. F. Gisin, *Anal Chim Acta*, 1972, **58**, 248.

25. E. Kaiser, R. L. Colescott, C. D. Bossinger and P. I. Cook, *Anal. Biochem.*, 1970, **34**, 595.

26. V. N. Soyfer and V. N. Potaman, Triple Helical Nucleic Acids, Eds 1996, SpringlerVerlag, New York.

27. R. C. Beavis and B. T. Chait, *Rapid Commun. Mass Spectrom.*, 1989, **3**, 436.

28. P. Job, *Ann. Chim. Appl.*, 1928, **9**, 113.

29. M. Egholm, O. Buchardt, L. Christensen, C. Behrens, S. M. Freier, D. A. Driver, R. H. Berg, S. K. Kim, B. Norden and P. E. Nielsen, *Nature*, 1993, **365**, 566-568.

30. a) R. S. Zhang, E. O. McCullum and J. C. Chaput, *J. Am. Chem. Soc.*, 2008, **130**, 5846. b) C. Lin, Y. Liu and H. Yan, *Biochemistry*, 2009, **48**, 1663. c) P. S. Ng and D. E. Bergstrom, *Nano Lett*, 2005, **5**, 107.

Chapter 3

β,γ-bis-substituted PNA with
configurational and conformational
switch: binding studies with
cDNA/RNA and their cellular uptake

Section A

Synthesis of (*R, R*)/(*S, S*) β,γ-Bis-Hydroxymethyl/Methoxymethyl substituted *aeg*PNA

3A.1 Introduction

The aminoethyl glycyl peptide nucleic acids (*aeg*PNA), designed by Nielsen *et al*, as synthetic nucleic acid mimics, are known since their discovery in 1991 for their superior properties such as strong and sequence specific binding to both DNA and RNA.[1-8] *aeg*PNA consist of an amide linked backbone formed by repeating *N*-(2-aminoethyl) glycyl units, carrying adenine, guanine, cytosine and thymine nucleobases *via* methylene carbonyl linkages (**Figure 3.1**).[4-6] This leads to complete replacement of the anionic sugar-phosphate-backbone of DNA. *aeg*PNA acts as a mimic of natural DNA/RNA and exhibits strong sequence specific binding with complementary oligonucleotide sequences following Watson-Crick base pairing rules but falls short of water solubility, orientational specificity of binding and cellular uptake.[1-3] Additionally, PNA aggregates and adheres to the surface of macromolecules in non-specific manner[9,10]. In the last two decades, several analogues of PNA have been synthesized, in the interest to develop new entities for the oligonucleotide (ON) based therapeutic strategies.[2] The desired additional attributes that are required for PNA for such applications are sufficient aqueous solubility,[7] efficient cellular uptake[8] and the ease of synthesis of modified PNAs, while maintaining the specificity of binding to complementary RNA sequences. The substitutions of *aeg*PNA, at α, β or γ positions (**Figure 3.1**) have been studied earlier to fulfill the requirement of above mentioned additional attributes.
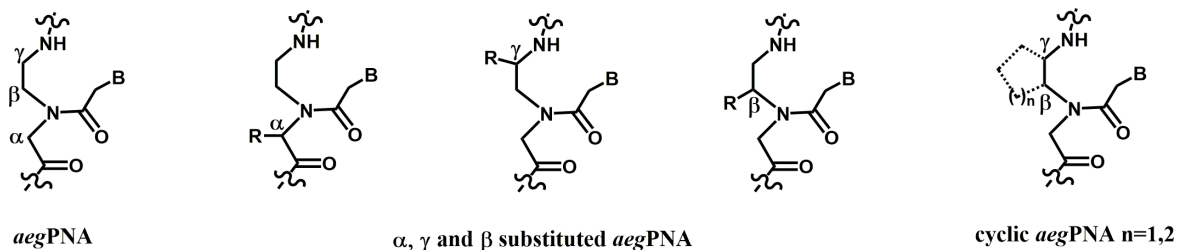
Figure 3.1: *aeg*PNA and reported modifications at α, β and γ position

The use of D/L-amino acids instead of glycine resulted in α-PNA[11-13] and α-GPNA.[14-15] Substitution at γ-position in aminoethyl segment of *aeg*PNA generated γ–PNA[16] and γ–GPNA[17].

Increased binding strength was observed in the case of positively charged PNAs depending upon the stereochemistry in the backbone and also the position of the substitution. The L-γ-GPNA exhibited higher T_m, whereas D-α-GPNA exhibited higher sequence selectivity.[18] Marchelli *et al* reported that γ–Lys PNA formed a stable duplex with DNA in comparison to α–Lys PNA. The change of stereochemistry had higher impact in γ–Lys PNA than α–Lys PNA.[19,20] Substitution of PNA at both α- and γ- positions were also explored that gave excellent improvement in the desired properties of *aeg*PNA.[21,22] The β-substitution was recently studied and was found to be acceptable in PNA:DNA/RNA duplex structures depending upon stereochemistry[23] The β, γ substituted cyclic-PNA containing both five-[24-26] and six-membered rings[27-30] (**Figure 3.1e**) showed good results in terms of directional selectivity of binding and discrimination in binding to DNA versus RNA complementary sequences. These rigid cyclic structures need to be synthesized from racemic compounds involving tedious resolution step in most cases.[27] Also the hydrocarbon side chains of cyclic PNA add to the undesired solubility setbacks.

3A.2 Rationale and objectives of the present work

β, γ-bis-substitution of PNA without involving cyclic structures can be synthesized from natural chiral pool. The side chains are amenable for further manipulations of different substitutions which could have favorable properties.
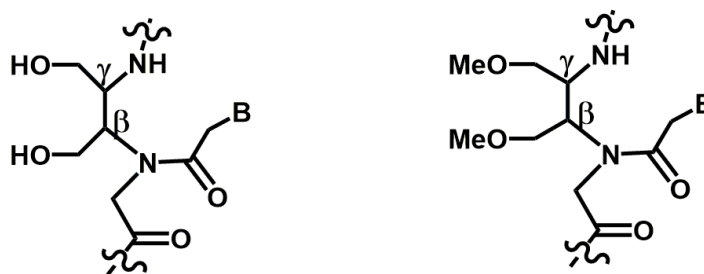
(*R, R/S, S*)-β, γ - bis - hydroxymethyl and methoxymethyl substituted *aeg*PNA

Figure 3.2: Proposed modifications of *aeg*PNA

Having β, γ-bis-hydroxymethyl/methoxymethyl substitution on the backbone of PNA can conquer the solubility problems and being chiral, may discriminate between parallel and antiparallel mode of binding.

3A.3 Synthesis of β, γ bis-hydroxymethyl/methoxymethyl *aeg*PNA monomers

To get the PNA with β, γ-bis-methoxymethyl substitutions, the synthesis of protected PNA monomers with (*S, S*) - stereocentre was accomplished from L (*R, R*)-tartaric acid (**Scheme 3.1**). L-tartaric acid was converted to diester and diacetonide. The diester was reduced to diol according to literature reports. The hydroxy groups were methylated employing methyl iodide to yield **2b₁**. The acetonide group was deprotected by refluxing compound in methanol in presence of PTSA yielding compound **3b₁**. The free hydroxy groups were then mesylated (**4b₁**) and converted to their corresponding azide **5b₁**. The bis-*O*-Me C_2-symmetric vicinal diamine compound **6b₁** was derived from the azide.[31] Mono-Boc-protection of the vicinal diamine **6b₁** gave compound **7b₁** in about 52% yield. This was further monoalkylated using ethylbromo acetate to get **8b₁**. Compound **8b₁** was monoacylated using choroacetyl chloride to get **9b₁** in good yield. This intermediate was further used to alkylate thymine and Cbz-adenine nucleobases to get the ester forms **10b₁** and **11b₁** respectively. Hydrolysis of **10b₁** and **11b₁** yielded the protected thymine and adenine modified PNA monomers (2*S*, 3*S*)-T^{OMe} **12b₁** and (2*S*, 3*S*)-A^{OMe} **13b₁** respectively.
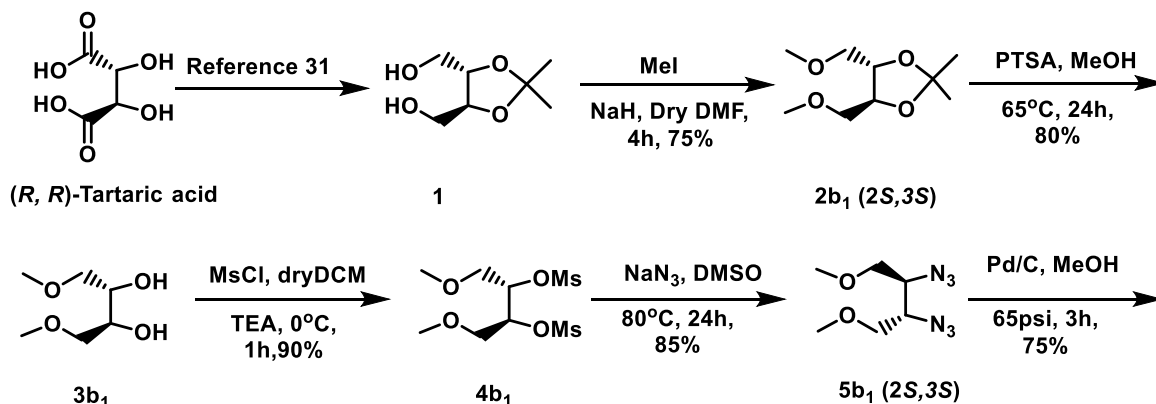
Scheme 3.1: Synthesis of (*S, S*) β,γ-substituted PNA monomers

Continued…..

…….continued
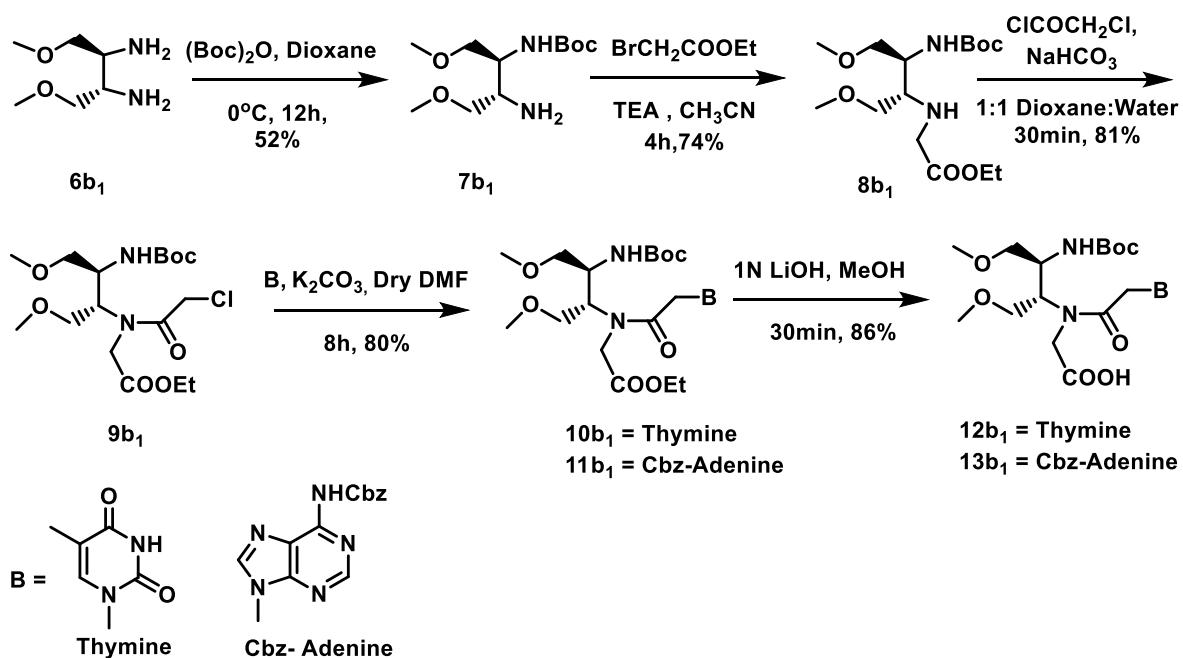
Scheme 3.1: Synthesis of (*S*, *S*) β,γ-substituted PNA monomers

The synthesis of (2*R*, 3*R*)-T^{OMe} **12a$_1$** and (2*R*, 3*R*)-A^{OMe} **13a$_1$** was accomplished using same set of reactions from D-tartaric acid[32] as drafted below (**Scheme 3.2**)
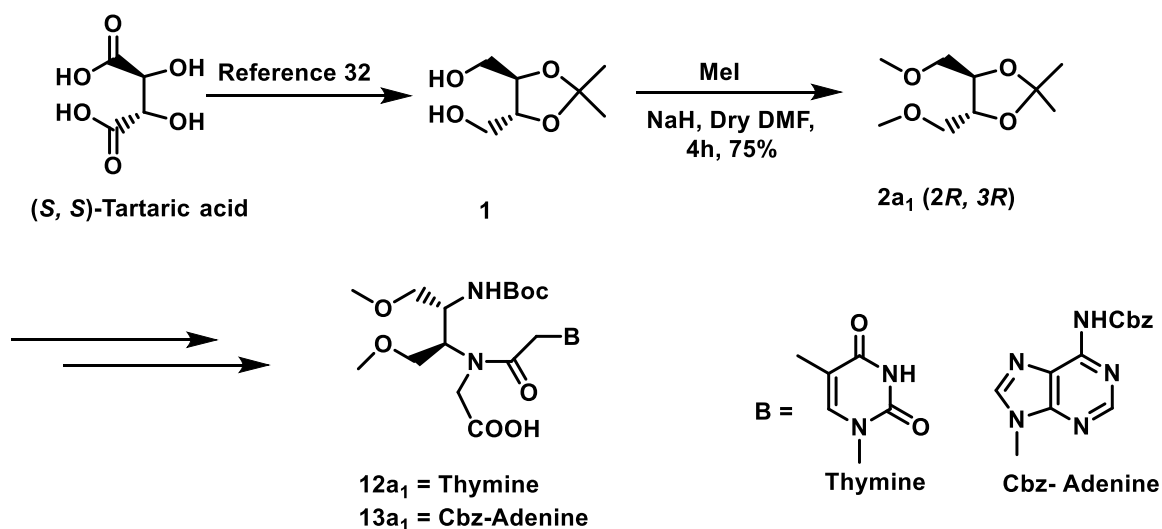
Scheme 3.2: Synthesis of (*R*, *R*) β,γ-substituted PNA monomers

To get the PNA with β, γ-bis-hydroxymethyl substitutions, the synthesis of β, γ- bis- benzyloxymethyl substituted thymine and adenine PNA monomers (2S, 3S)-T^{OBn} **12b$_2$**/ (2S, 3S)-A^{OBn} **13b$_2$** was accomplished from L tartaric acid (**Scheme 3.3**). The free hydroxy groups of compound **1** were benzylated using benzyl bromide in presence of NaH in dry DMF to generate compound **2b$_2$**. Monomers **12b$_2$** and **13b$_2$** were then synthesized from **2b$_2$** using the same set of reactions as described for *O*-methoxymethyl protected monomers.
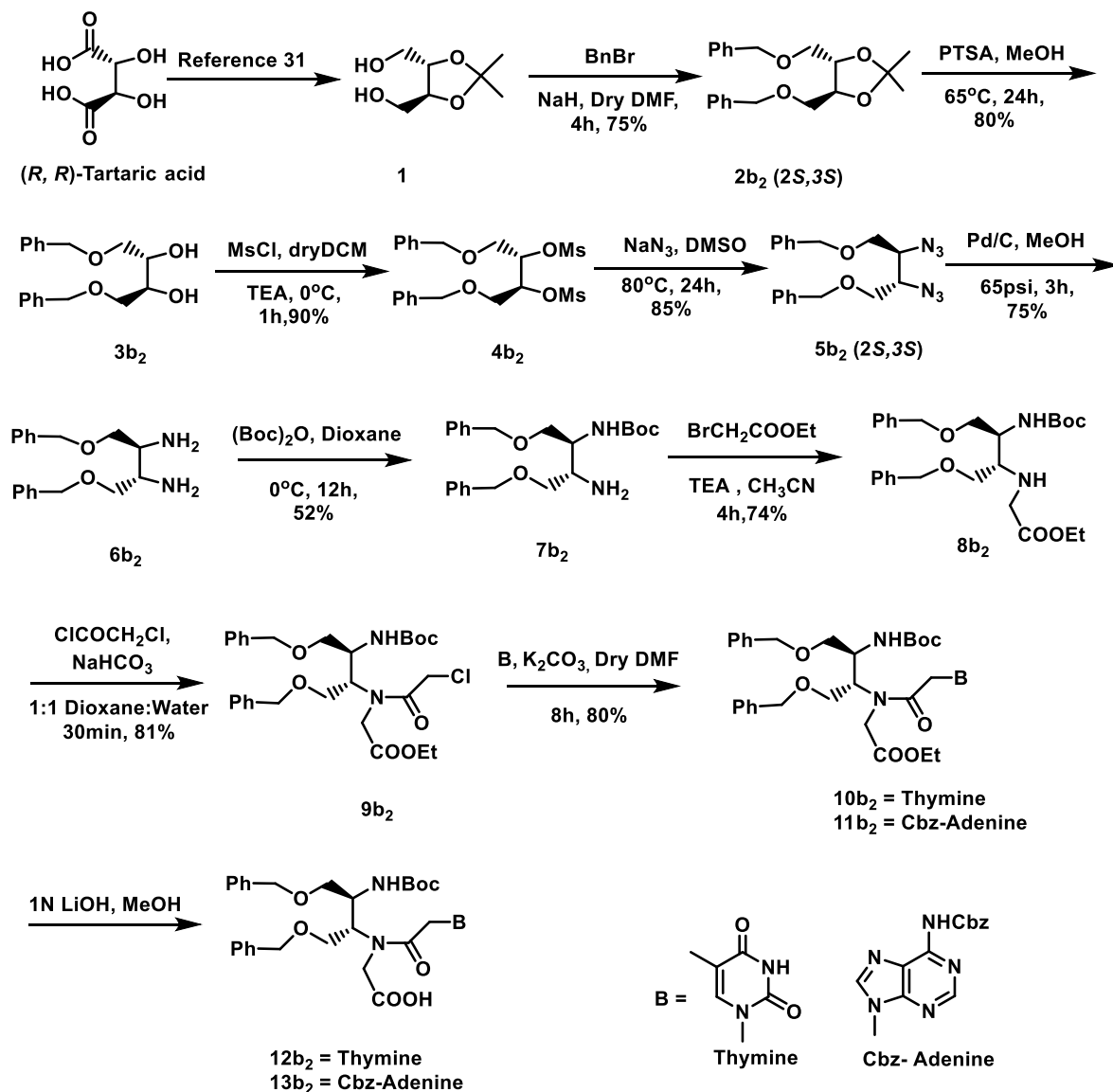
Scheme 3.3: Synthesis of (*S, S*) β,γ-substituted PNA monomers

$(2R, 3R)$-T^{OBn} **12a$_2$**/ $(2S, 3S)$-A^{OBn} **13a$_2$** were synthesized starting from D-tartaric acid following the same synthetic route as above.
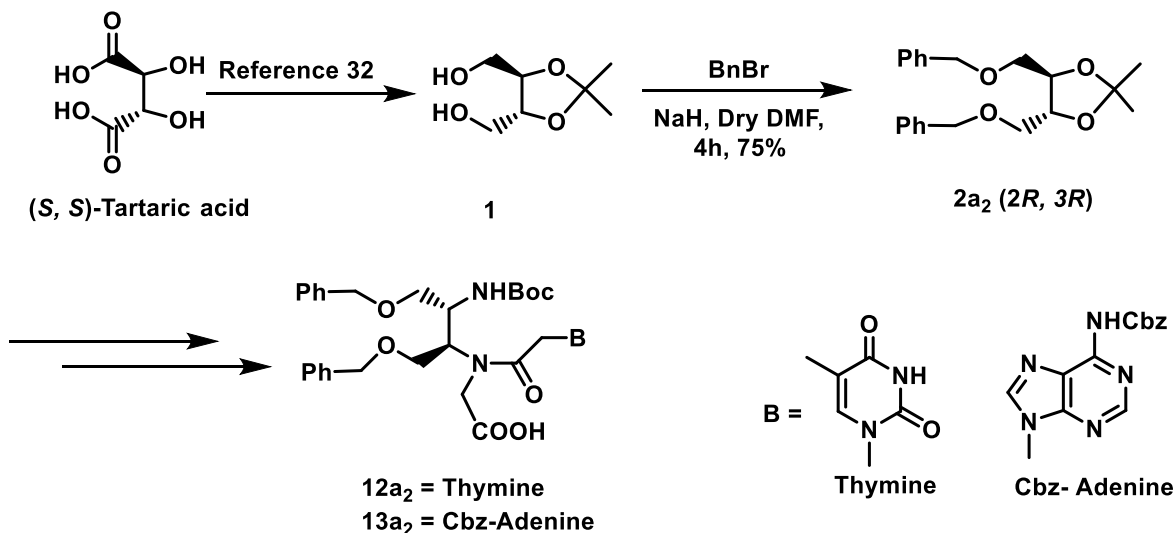
Scheme 3.4: Synthesis of (R, R) β,γ-substituted PNA monomers

Use of bis-OBn protected monomers at defined positions during the synthesis of PNA sequences allowed us the access to the bis-hydroxymethyl substitution in PNA as the OBn protection was cleanly removed[33] at the final deprotection step. The two hydroxymethyl groups were expected to largely improve the aqueous solubility of PNA.[34] The compounds were all characterized by ^{1}H, ^{13}C NMR, HRMS and optical rotation analysis. The chiral purity of the products was also ascertained by chiral HPLC of the enantiomeric pair **10a$_1$/10b$_1$**.

3A.4 Solid phase PNA synthesis, purification and MALDI-TOF characterization of oligomers

The monomers synthesized were assimilated in oligomers at the desired positions by solid phase peptide synthesis through Boc chemistry on L-Lysine derivatized MBHA resin. Couplings were performed in presence of PNA monomers (modified and unmodified), HOBt, TBTU and DIPEA in DMF to get the oligomer of required length. Other procedures like deprotection, cleavage, Kaiser Test were performed following the protocol documented in *Chapter 2*. The positions of modification were chosen at the centre as well as the terminus to gain complete knowledge of the role of modifications in the sequence. Four lysines were also incorporated at the *N*-terminus (PNA 1 and PNA 10a) of the oligo-

mers to further enhance the solubility. The oligomers were then purified by RP-HPLC and characterized by MALDI-TOF.

Table 1: HPLC purification of PNA 1, 2, 3, 4, 5, 6, 7, 8, 9 and 10 and their MALDI-TOF characterization

Seq Code	Sequences	HPLC t_R (min)	MALDI	
			Calc.	Obsvd.
PNA 1	aaccgatttcag-K	10.9	3381.4	3384.09
PNA 2	K_4-aaccgatttcag-K	10.9	3893.9	3895.98
PNA 3a	aaccga$^{(R,R)}T^{OH}$ttcag-K	10.3	3439.4	3478.6(M+K)
PNA 3b	aaccga$^{(S,S)}T^{OH}$ttcag-K	10.3	3439.4	3477.9(M+K)
PNA 4a	aaccga$^{(R,R)}T^{OMe}$ttcag-K	11.7	3467.5	3468.1
PNA 4b	aaccga$^{(S,S)}T^{OMe}$ttcag-K	11.8	3467.5	3469.1
PNA 5a	aaccga$^{(R,R)}T^{OMe}$t$^{(R,R)}T^{OMe}$CAG-K	12.6	3555.5	3558.2
PNA 5b	aaccga$^{(S,S)}T^{OMe}$t$^{(S,S)}T^{OMe}$cag-K	12.5	3555.5	3597.3(M+K)
PNA 6a	aaccg$^{(R,R)}A^{OH}$tttcag-K	10.9	3439.4	3464.9(M+Na)
PNA 6b	aaccg$^{(S,S)}A^{OH}$tttcag-K	10.8	3439.4	3478.2(M+K)
PNA 7a	aaccg$^{(R,R)}A^{OMe}$tttcag-K	11.4	3467.5	3491.8(M+Na)
PNA 7b	aaccg$^{(S,S)}A^{OMe}$tttcag-K	11.3	3467.5	3490.7(M+Na)
PNA 8a	$^{(R,R)}A^{OH}$$^{(R,R)}A^{OH}$CCGATTTCAG-K	10.9	3499.4	3505.0
PNA 8b	$^{(S,S)}A^{OH}$$^{(S,S)}A^{OH}$CCGATTTCAG-K	10.9	3499.4	3497.5
PNA 9a	$^{(R,R)}A^{OMe}$$^{(R,R)}A^{OMe}$ccgatttcag-K	11.6	3555.5	3598.0(M+K)
PNA 9b	$^{(S,S)}A^{OMe}$$^{(S,S)}A^{OMe}$ccgatttcag-K	11.6	3555.5	3593.4(M+K)
PNA 10a	K_4-accga$^{(R,R)}T^{OMe}$t$^{(R,R)}T^{OMe}$cag-K	12.5	4067.9	4104.7(M+K)

lower case letters denote *aeg*PNA, uppercase letters denote modified units with bis-hydroxymethyl (^{OH}), bis-methoxymethyl (^{OMe}) substitutions and (*R, R*), (*S, S*) denote the stereochemistry at β, γ positions

3A.5 Summary

- This section of the work describes the synthesis of (*S, S*) and (*R, R*) bis-hydroxymethyl/methoxymethyl monomers from L (+) and D (-) Tartaric Acid respectively.

   { The synthesized monomers were successfully incorporated in oligomers at definite positions and were characterized

3A.6 Experimental

(4S, 5S)-4, 5-bis (methoxymethyl)-2, 2-dimethyl-1, 3-dioxolane (2b$_1$)

Sodium hydride (50% suspension in mineral oil) (2.2g, 92.5mmol) was stirred in anhydrous DMF at 0°C to 5°C.To it (4S, 5S)-2, 2 - dimethyl-1, 3- dioxolane-4,5-diyl dimethanol (3g, 18.5mmol) was added and allowed to stir vigorously for 10mins followed by methyl iodide (3.5mL, 55.5mmol) .The solution was allowed to gradually attain room temperature and stir for 4h. After completion of the reaction DMF was removed under vacuum and the residue redissolved in ethyl acetate. The organic layer was washed with water (3x100mL) and then brine (2x50mL). The organic layer was dried over anhydrous Na$_2$SO$_4$ and evaporated to dryness. The crude compound was purified by column chromatography (10% EtOAc in petroleum ether) providing compound **2b$_1$**. (2.6g, 74%).[α]$_D^{20}$ +8.13(c 1.2, CHCl$_3$); ^{1}H NMR(200 MHz, CDCl$_3$): δ (ppm) 1.43 (s, 6H), 3.41(s, 6H), 3.51-3.54(m, 4H), 3.94-3.99(m, 2H); ^{13}C NMR (50 MHz, CDCl$_3$): δ (ppm) 26.92, 59.41, 23.36, 109.69; HRMS calcd. for C$_9$H$_{18}$O$_4$Na: 213.1097,Observed mass: 213.1101.

(4R, 5R)-4, 5-bis (methoxymethyl)-2, 2-dimethyl-1, 3-dioxolane (2a$_1$)

[α]$_D^{20}$ -8.25(c 1.4, CHCl$_3$)
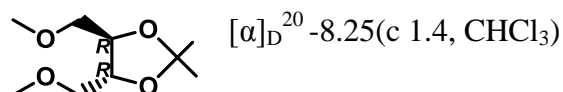
(4S, 5S)-2, 2-dimethyl-4, 5-bis (benzyloxy)-1, 3-dioxolane (2b$_2$)

Sodium hydride(50% suspension in mineral oil) (2.2g, 92.5mmol) was stirred in anhydrous DMF at 0°C to 5°C.To it (4S, 5S)-2, 2-dimethyl-1, 3-dioxolane-4, 5-diyl dimethanol (3g, 18.5mmol) was added and allowed to stir vigorously for 10mins followed by benzyl bromide (5.3mL, 44.4mmol) .The solution was allowed to gradually attain room temperature and stir for 4 h. After completion of the reaction DMF was removed under vacuum and redissolved in ethyl acetate. The organic layer was washed with water (3x100mL) and then brine (2x50mL). The organic layer was kept over anhydrous Na$_2$SO$_4$ and evaporated to dryness. The crude compound was purified by column chromatography (5% EtOAc in petroleum ether) to get compound

2b$_2$. (5.2g, 83%); $[\alpha]_D^{20}$ -7.4(c 2.3, CHCl$_3$); +8.0(c 3, CHCl$_3$) for **2b$_2$**,^{1}H NMR (200 MHz, CDCl$_3$): δ (ppm) 1.43 (s, 6H), 3.58-3.60(m, 4H), 4.02-4.05(m, 2H), 4.55(s,4H), 7.3(s,10H); ^{13}C NMR (50 MHz, CDCl$_3$): δ (ppm) 27.01, 70.64, 73.47, 77.47, 109.62, 127.60, 128.00, 128.34, 137.96; HRMS calcd for C$_{21}$H$_{26}$O$_4$Na: 365.1723, Observed mass:365.1731

(4*R*, 5*R*)-2, 2-dimethyl-4, 5-bis (benzyloxy)-1, 3-dioxolane (2a$_2$)

$[\alpha]_D^{20}$ +8.0(c 3, CHCl$_3$)
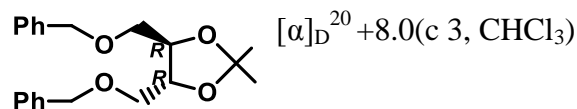
(4*S*, 5*S*)-1, 4-dimethoxybutane-2, 3-diol (3b$_1$)

Compound **2b$_1$** (3.6g, 18.9mmol) was dissolved in methanol. To it catalytic amount of PTSA and H$_2$O (1mL, 56.7mmol) was added and the mixture was allowed to stir for 24h. The reaction mixture was quenched with K$_2$CO$_3$. The reaction mixture was filtered and the solvent was removed under vacuum. The residue was dissolved in EtOAc and the organic layer was washed with water (3x100mL) and brine (2x100mL). The organic layer was kept over Na$_2$SO$_4$ and the solvent was removed to get the crude compound. The compound was purified by column chromatography (80% EtOAc in petroleum ether) to get **3b$_1$**.(2.2g, 78%); $[\alpha]_D^{20}$ -3.7 (c 1.6,CHCl$_3$); ^{1}H NMR (200 MHz, CDCl$_3$): δ (ppm) 2.94 (br s, 2H), 3.4(s, 6H), 3.51-3.54(m, 4H), 3.8-3.84(m,2H); ^{13}C NMR (50 MHz, CDCl$_3$): δ (ppm) 59.3, 70.46, 74.54; HRMS calcd for C$_6$H$_{14}$O$_4$Na: 173.0784, Observed mass: 173.0789.

(4*R*, 5*R*)-1, 4-dimethoxybutane-2, 3-diol (3a$_1$)

$[\alpha]_D^{20}$ +4.2 (1.5, CHCl$_3$)
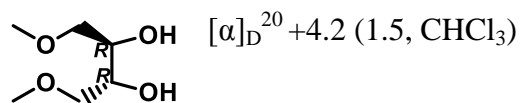
(4*S*, 5*S*)-1, 4-bis (benzyloxy) butane-2, 3-diol (3b$_2$)

Compound **2b$_2$** (8g, 23.3mmol) was dissolved in methanol. To it catalytic amount of PTSA and H$_2$O (1.7mL, 93.2mmol) was added and the mixture was allowed to stir for 24h. The reaction mixture was quenched with K$_2$CO$_3$. The reaction mixture was filtered and the solvent was removed under vacuum. The residue was dissolved in EtOAc and the organic layer was washed with water (3 x100mL) and brine (2x100mL). The organic layer was kept over Na$_2$SO$_4$ and the solvent

was removed to get the crude compound. The compound was purified by column chromatography (30% EtOAc in petroleum ether) to get **3b$_2$** (5.6g, 79%); $[\alpha]_D^{20}$ -6.3(c 5, CHCl$_3$); ^{1}H NMR (200 MHz, CDCl$_3$): δ (ppm) 2.89(br s, 2H), 3.57-3.61(m, 4H), 3.84-3.88(m, 2H), 4.54(s, 4H), 7.32(s, 10H); ^{13}C NMR (50 MHz, CDCl$_3$): δ (ppm) 70.54, 71.95, 73.57, 127.78, 127.85, 128.48, 137.69; HRMS calcd for C$_{18}$H$_{22}$O$_4$Na: 325.1410, Observed mass: 325.1418.

(4R, 5R)-1, 4-bis (benzyloxy) butane-2, 3-diol (3a$_2$)

$[\alpha]_D^{20}$ +7.2(c 4.8, CHCl$_3$)
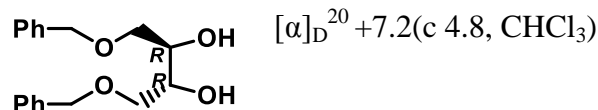
(4S, 5S)-1, 4-dimethoxybutane-2, 3-diyl dimethanesulfonate (4b$_1$)

 Methanesulfonyl Chloride (2.8mL, 15.9mmol) was added dropwise to a cooled solution at 0°C of **3b$_1$** (1g, 6.7mmol) and TEA (1.2mL, 19.9mmol) in dry DCM. After 1h the reaction mixture was washed with sat. NaHCO$_3$ (2x50mL) followed by brine (2x50mL).The solvent was removed and the crude compound was purified by column chromatography (30% EtOAc in Pet ether) to get white crystalline compound **4b$_1$** (1.8g, 91%) $[\alpha]_D^{20}$ -26.72(c 1.4, CHCl$_3$) for **4b$_1$**; ^{1}H NMR (200 MHz, CDCl$_3$): δ (ppm) 3.13(s, 6H), 3.41(s, 6H), 3.69-3.71(m, 4H), 4.92-4.95(m, 2H); ^{13}C NMR (50 MHz, CDCl$_3$): 38.71, 59.25, 71.18, 78.73; HRMS calcd for C$_8$H$_{18}$O$_8$NaS$_2$: 329.0335, Observed mass: 329.0347.

(4R, 5R)-1, 4-dimethoxybutane-2, 3-diyl dimethanesulfonate (4a$_1$)

 $[\alpha]_D^{20}$ +25.6(c 1.6, CHCl$_3$)
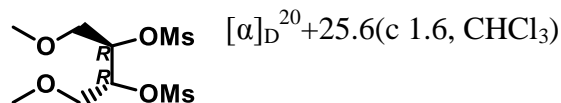
(4S, 5S)-1, 4-bis (benzyloxy) butane-2, 3-diyl dimethanesulfonate (4b$_2$)

 Compound **3b$_2$** (2.7g, 8.9mmol) was dissolved in 10mL of dry DCM at 0°C. At this temperature TEA (3mL, 21.4mmol) was added and allowed to stir for 5mins. To the reaction mixture Methanesulfonyl chloride(2.1mL, 26.8mmol) was added drop wise and was allowed to stir for 1h.The reaction mixture was washed with sat. NaHCO$_3$ (2x50mL) followed by brine(2x50mL).The solvent was removed

and the crude compound was purified by column chromatography (20% EtOAc in Pet ether) to get white crystalline compound **4b$_2$** (7.7g, 93%).$[\alpha]_D^{20}$ -10.7(c 4.4, CHCl$_3$); ^{1}H NMR (200 MHz, CDCl$_3$): δ (ppm) 3.02(s,6H), 3.74-3.76(m,4H), 4.42(d,2H, *J*=11.6 Hz), 4.53(d,2H, *J*=11.6Hz), 4.97-5.00 (m,2H),7.30-7.34(m,10H); ^{13}C NMR (50 MHz, CDCl$_3$): δ (ppm) 38.75, 68.64,73.63, 78.74, 128.05, 128.15, 128.58, 136.98; HRMS calcd for C$_{20}$H$_{26}$O$_8$NaS$_2$: 481.0961, Observed mass: 481.0977

(4*R*, 5*R*)-1, 4-bis (benzyloxy) butane-2, 3-diyl dimethanesulfonate (4a$_2$)

$[\alpha]_D^{20}$ +11.3(c 5.2, CHCl$_3$)
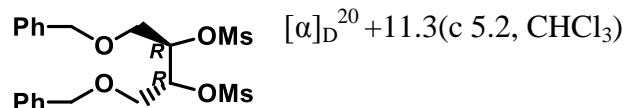
(4*S*, 5*S*)-2, 3-diazido-1, 4-dimethoxybutane (5b$_1$)

A mixture of dimesylate **4b$_1$**(2g, 6.5mmol) and sodium azide (4.2g, 65.3mmol) in DMSO (20mL) was stirred at 80°C for 24h,cooled to rt, and white suspension was diluted with brine. The water layer was extracted with ethyl acetate (3x50mL) and the combined organic extracts were dried over Na$_2$SO$_4$.Concentration of organic layer in vacuum followed by column chromatography on silica gel (10% EtOAc in Pet ether) produced compound **5b$_1$** as yellow liquid (1.2g, 89%); $[\alpha]_D^{20}$ -40.58(c 1.4, CHCl$_3$), ^{1}H NMR (200 MHz, CDCl$_3$): δ (ppm) 3.41(s, 6H), 3.59-3.69(m, 6H); ^{13}C NMR (50 MHz, CDCl$_3$): δ (ppm) 59.15, 60.97, 72.11.

(4*R*, 5*R*)-2, 3-diazido-1, 4-dimethoxybutane (5a$_1$)

$[\alpha]_D^{20}$ +41.2 (c 2.0, CHCl$_3$)
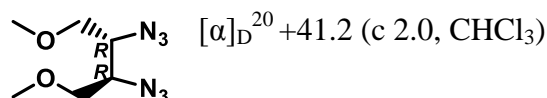
(4*S*, 5*S*)- (((2, 3-diazidobutane-1, 4 diyl) bis(oxy)) bis(methylene))dibenzene (5b$_2$)

Compound **4b$_2$** (3g, 6.5mmol) was dissolved in DMSO (30mL). To it sodium azide (4.2g, 65.3mmol) was added and the reaction mixture was heated to 80°C for 1d. The reaction mixture was cooled to rt, and white suspension was diluted with brine. The water layer was extracted with ethyl acetate (3x50mL) and the combined organic extracts were dried over Na$_2$SO$_4$. The organic layer was concentrated, which was purified by column chromatography (5% EtOAc in Pet ether) to afford **5b$_2$** (2g, 86%) as yellow oil. $[\alpha]_D^{20}$ -43(c 0.7, CHCl$_3$); ^{1}H NMR (200 MHz,

CDCl$_3$): δ (ppm) 3.63-3.72(m, 6H), 4.53(s, 4H), 7.32(s, 10H); ^{13}C NMR (50 MHz, CDCl$_3$): δ (ppm) 61.1, 69.71, 73.66, 127.88, 128.08, 128.64, 137.49; HRMS calcd for C$_{18}$H$_{20}$O$_2$N$_6$Na: 375.1540, Observed mass: 375.1537

(4R, 5R)-(((2,3-diazidobutane-1,4-diyl)bis(oxy))bis(methylene))dibenzene (5a$_2$)

[α]$_D$^{20} +44.4(c 1.0, CHCl$_3$)
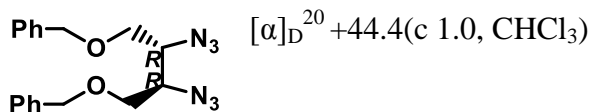
(4S, 5S)-1, 4-dimethoxybutane-2, 3-diamine (6b$_1$)

(4R, 5R)-1, 4-dimethoxybutane-2, 3-diamine (6a$_1$)
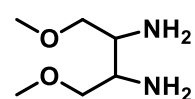
To a solution of azide (S, S)/(R, R) (500mg, 2.9mmol) in methanol was added 10% Pd/C (50mg). The mixture was hydrogenated at room temperature under 50psi pressures for 4h. The catalyst was then removed by filtration over celite. The filtrate was collected and the solvent was removed under vacuum to furnish a colorless liquid which eventually turned brown. The crude compound was carried forward for the next reaction without further purification.

(4S, 5S)-1, 4-bis (benzyloxy) butane-2, 3-diamine (6b$_2$)

(4R, 5R)-1, 4-bis (benzyloxy) butane-2, 3-diamine (6a$_2$)
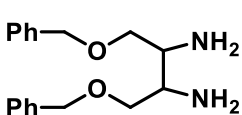
To a solution of azide (300mg, 0.9mmol) in THF (12mL), PPh$_3$ (892mg, 3.4mmol) and few drops of water was added. The solution was heated to 60°C for 1h cooled down to rt. The mixture was diluted with ether (10mL) and the organic phase was extracted with AcOH (10% in H$_2$O).The combined aqueous extract was washed with ether and lyophilized to give amine. The crude compound was used as such for the next reaction without further purification.

(4S, 5S)-*tert*-butyl (3-amino-1,4-dimethoxybutan-2-yl)carbamate (7b$_1$)
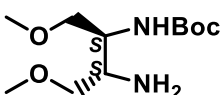
A solution of di-*tertiary* butyl carbonate (1.5g, 7.0mmol) diluted in dioxane (250mL) was added over a period of 5h to the solution of diamine(1.3g, 8.8mmol) in dioxane(75mL) at 0°C.The mixture was allowed to stir for 12h at rt. The solvent was removed under vacuum and water was added to it when di-*tert*-butyl carbamate compound separated out. The aqueous layer was extracted with 10% MeOH in

DCM (10x50mL). The organic layer was collected and solvent removed to get **7b$_1$** (1.4g, 52%). $[\alpha]_D^{20}$ -7.84 (c 1.02, CHCl$_3$); ^{1}H NMR (200 MHz, CDCl$_3$): δ (ppm) 1.55(s, 9H), 3.25-3.47(m, 5H), 3.35(s, 6H), 3.62-3.65(m, 1H), 5.31-5.35 (br s); ^{13}C NMR (50 MHz, CDCl$_3$): δ (ppm) 28.31, 50.79, 51.27, 58.85, 73.48, 75.03, 79.16, 155.83; HRMS calcd for C$_{11}$H$_{25}$N$_2$O$_4$: 249.1809, Observed mass: 249.1815.

(4R, 5R)-*tert*-butyl (3-amino-1,4-dimethoxybutan-2-yl) carbamate (7a$_1$)

$[\alpha]_D^{20}$ +7.1 (c 1, CHCl$_3$)
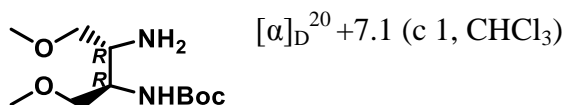
(4S, 5S)-*tert*-butyl (3-amino-1,4-bis(benzyloxy)butan-2-yl)carbamate (7b$_2$)

A solution of di-*tertiary* butyl carbonate (0.99g, 4.5mmol) diluted in dioxane (250mL) was added over a period of 5h to the solution of diamine(1.7g, 5.7mmol) in dioxane(75mL).The mixture was allowed to stir for 12h at rt. The solvent was removed under vacuum and water was added to it when di-*tert*-butyl carbamate compound separated out. The aqueous layer was extracted with 8% MeOH in DCM (10x50mL). The organic layer was collected and solvent removed to get **7b$_2$** (1.3g, 55%). $[\alpha]_D^{20}$-9.2(c 1.4,CHCl$_3$); ^{1}H NMR (500 MHz, CDCl$_3$): δ (ppm) 1.43(s, 9H), 3.45-3.64(m, 6H), 3.78(m, 1H), 4.49(s, 4H), 5.26-5.30(br s, 1H), 7.31(s, 10H); ^{13}C NMR (125 MHz, CDCl$_3$): δ (ppm) 28.34, 51.10, 61.59, 69.12, 70.22, 73.23, 79.98, 127.87, 127.94, 128.44, 137.79, 156.70; HRMS calcd for C$_{23}$H$_{33}$N$_2$O$_4$: 401.2435, Observed mass: 401.2435

(4R, 5R)-*tert*-butyl (3-amino-1, 4-bis(benzyloxy)butan-2-yl)carbamate (7a$_2$)

$[\alpha]_D^{20}$ +9.4 (c 1.0, CHCl$_3$)
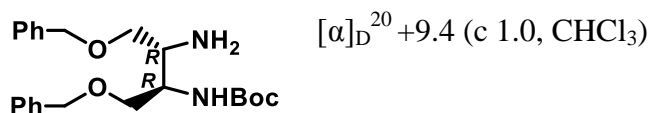
(4S, 5S)-ethyl 2-((3-((*tert*-butoxycarbonyl)amino)-1,4-dimethoxybutan-2-yl)amino) acetate (8b$_1$)
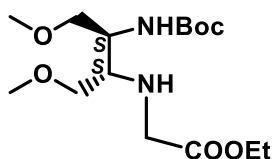
Compound **7b$_1$** (3.2g, 12.9mmol) was dissolved in 20mL dry ACN and TEA (3.6mL, 25.8mL) was added with stirring. The reaction mixture was cooled to 0°C and ethyl bromoacetate (1.7mL,

15.5mmol) diluted in 10mL dry ACN was added drop wise. The reaction was stirred at rt for 4h after which solvent was removed under vacuo and the residue was diluted with EtOAc. The organic layer was washed with NaHCO$_3$ (2x50mL), dried over anhydrous Na$_2$SO$_4$ and concentrated to get the crude compound. The crude compound was purified by column chromatography (20% EtOAc in pet ether) to yield Compound **8b$_1$** as colorless liquid (3.2g, 74%); $[\alpha]_D^{20}$ -6(c 2.0, CHCl$_3$);^{1}H NMR (200 MHz, CDCl$_3$): δ (ppm) 1.28-1.31(t,3H,*J*=7.2Hz), 1.44(s, 9H), 2.27(br s, 1H), 2.94-3.01(m, 1H), 3.33(s, 3H), 3.34(s, 3H), 3.36-3.48(m, 6H), 3.77-3.81(m, 1H),4.13-4.24 (q, 2H, *J*=7.2Hz); ^{13}C NMR (50 MHz, CDCl$_3$): δ (ppm) 14.11, 28.28, 49.82, 57.79, 58.79, 60.62, 72.42, 73.33, 79.07, 155.71, 172.53; HRMS calcd for C$_{15}$H$_{31}$N$_2$O$_6$: 335.2177, Observed mass: 335.2183

(4*R*, 5*R*)-ethyl 2-((3-((*tert*-butoxycarbonyl)amino)-1,4-dimethoxybutan-2-yl)amino) acetate (8a$_1$)

$[\alpha]_D^{20}$ +6.7(c 1.7, CHCl$_3$)
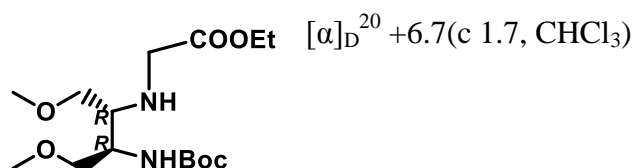
(4*S*, 5*S*)-ethyl 2-((1,4-bis(benzyloxy)-3-((*tert*-butoxycarbonyl)amino)butan-2-yl)amino) acetate (8b$_2$)
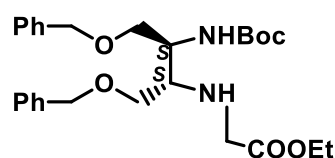
Compound **7b$_2$** (0.6g, 2.4mmol) was dissolved in 5mL dry ACN and TEA (0.7mL, 4.8mmol) was added with stirring. The reaction mixture was cooled to 0°C and ethyl bromoacetate (0.3mL, 2.9mmol) diluted in 1mL dry ACN was added drop wise. The reaction was stirred at rt for 4h after which solvent was removed under vacuo and the residue was diluted with EtOAc. The organic layer was washed with NaHCO$_3$ (2x30mL), dried over anhydrous Na$_2$SO$_4$ and concentrated to get the crude compound. The crude compound was purified by column chromatography (15% EtOAc in pet ether) to yield Compound **8b$_2$** as colorless liquid (0.82g, 71%). $[\alpha]_D^{20}$ -10 (c 3.2, CHCl$_3$);^{1}H NMR (200 MHz, CDCl$_3$): δ (ppm) 1.21-1.28 (t,3H,*J*=7.2Hz), 1.43(s, 9H), 2.19(br s, 1H), 3.06-3.10 (m, 1H), 3.45-3.59 (m, 6H),3.86-3.90(m, 1H), 4.1-4.2 (q, 2H, *J*=7.2Hz), 4.47-4.54 (m, 4H), 7.31(s, 10H); ^{13}C NMR (125 MHz, CDCl$_3$): δ (ppm) 14.20, 28.41, 49.97, 57.91, 60.75, 70.09, 71.04, 73.33,

79.20, 127.7, 127.74, 128.38, 138.12, 155.81, 172.56; HRMS calcd for $C_{27}H_{39}N_2O_6$: 487.2803, Observed mass: 487.2806

(4*R*, 5*R*)-ethyl 2-((1,4-bis(benzyloxy)-3-((*tert*-butoxycarbonyl)amino)butan-2-yl) amino) acetate (8a₂)
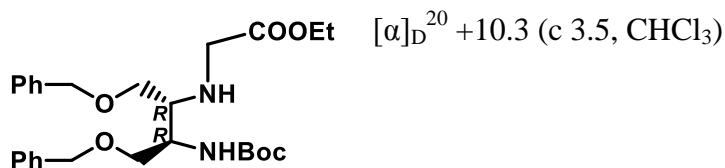

$[\alpha]_D^{20}$ +10.3 (c 3.5, CHCl₃)

(4*S*, 5*S*)-ethyl 2-(*N*-(3-((*tert*-butoxycarbonyl)amino)-1,4-dimethoxybutan-2-yl)-2-chloroacetamido)acetate (9b₁)
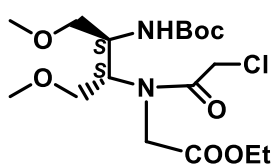
Chloroacetyl chloride (1.4mL, 17.9mmol) was added in 2-3 portions to the solution of compound **8b₁** (1.2g, 3.6mmol) and NaHCO₃ (3g, 35.9mmol) in 20mL dioxane, water (1:1) at 0°C. The pH of the reaction mixture was maintained between 8-9. The mixture was allowed to stir for 1/2h. After completion of the reaction the dioxane was removed under vacuum and the water was extracted with ethyl acetate (3x50mL). The organic layer was concentrated to get the crude compound. The crude compound was purified by column chromatography (15% EtOAc in Pet ether) to get colorless liquid (1.2g, 81%); $[\alpha]_D^{20}$ -11.91 (c 1.7, CHCl₃); ¹H NMR (200 MHz, CDCl₃): δ (ppm) 1.24-1.31 (m, 3H, rotameric mixture), 1.41-1.45 (9H, rotameric mixture), 2.95-3.03 (m, 1H), 3.10-3.38 (m, 6H, rotameric mixture), 3.39-3.77(m, 6H), 4.02-4.20(m, 5H) ; ¹³C NMR (50 MHz, CDCl₃): δ (ppm) 14.2, 28.41, 49.97, 57.91, 60.75, 70.09, 71.04, 73.23, 73.33, 79.20, 127.7, 127.74, 128.38, 138.12, 155.81, 172.56 ; HRMS calcd. for $C_{17}H_{31}O_7N_2ClNa$: 433.1712, Observed mass: 433.1727

(4*R*, 5*R*)-ethyl 2-(*N*-(3-((*tert*-butoxycarbonyl)amino)-1,4-dimethoxybutan-2-yl)-2-chloroacetamido)acetate (9a₁)
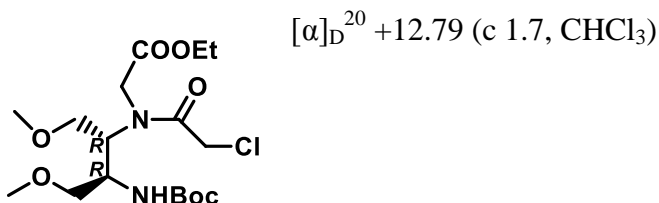
$[\alpha]_D^{20}$ +12.79 (c 1.7, CHCl₃)

(4*S*, 5*S*)-ethyl 2-(*N*-(1, 4-bis(benzyloxy)-3-((*tert*-butoxycarbonyl)amino)butan-2-yl)-2-chloroacetamido) acetate (9b$_2$)
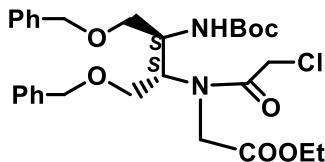
Chloroacetyl chloride (2.1mL, 26.7mmol) was added in 2-3 portions to the solution of compound **8b$_2$** (2.6g, 5.3mmol) and NaHCO$_3$ (4.5g, 53.4mmol) in 25mL dioxane, water (1:1) at 0°C. The pH of the reaction mixture was maintained between 8-9. The mixture was allowed to stir for 1/2h. After completion of the reaction the dioxane was removed under vacuum and the water was extracted with ethyl acetate (3x50mL). The organic layer was concentrated to acquire the crude compound. The crude compound was purified by column chromatography (15% EtOAc in Pet ether) to get colorless liquid (2.6g, 84%).[α]$_D^{20}$ -6.7 (c 3.7, CHCl$_3$); ^{1}H NMR (200 MHz, CDCl$_3$): δ (ppm) 1.40-1.51 (m, 3H, 9H,rotameric mixture), 3.23-3.29(m, 1H), 3.48-3.60(m, 5H), 3.84-4.19(m, 3H), 4.35-4.37(m, 4H), 4.48-4.56(m, 5H),5.27-5.29(br s, 1H), 7.33(s, 10H); ^{13}C NMR (50 MHz, CDCl$_3$): δ (ppm) 14.19, 28.33, 42.50, 57.69, 60.39,69.00, 69.41, 71.04, 73.36, 73.60, 81.04, 127.80, 128.33, 128.57, 137.09, 152.69, 166.92, 171.05; HRMS calcd for C$_{29}$H$_{40}$O$_7$N$_2$Cl: 563.2519, Observed mass: 563.2519

(4*R*, 5*R*)-ethyl 2-(*N*-(1,4-bis(benzyloxy)-3-((*tert*-butoxycarbonyl)amino)butan-2-yl)-2-chloroacetamido) acetate (9a$_2$)
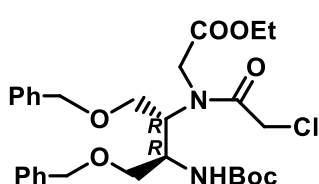
[α]$_D^{20}$ +6.3 (c 3.5, CHCl$_3$)

(4*S*, 5*S*)-ethyl2-(*N*-(3-((*tert*-butoxycarbonyl)amino)-1,4-dimethoxybutan-2-yl)-2-thymi nyl acetate (10b$_1$)
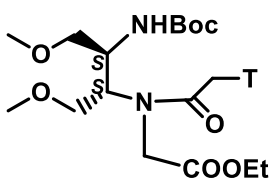
Compound **9b$_1$** (2.1g, 5.1mmol), activated K$_2$CO$_3$ (0.85g, 6.1mmol) and Thymine (0.7g, 5.6mmol) was suspended in 20mL of dry DMF and the reaction mixture was allowed to stir for 6h at rt. The solvent was removed under vacuum and the residue was extracted with ethyl acetate (3x100mL). The organic layer was washed with brine (3x50mL) and was dried over anhydrous Na$_2$SO$_4$. Ethyl acetate was removed under vacuum to get the crude compound. The crude compound was purified by column chromatography (30% EtOAc in Pet ether) to

get white solid (1.9g, 80%); $[\alpha]_D^{20}$ -6.9(c 0.6, CHCl$_3$); ^{1}H NMR (200 MHz, CDCl$_3$): δ (ppm) 1.23-1.32(3H, rotameric mixture), 1.4, 1.44(9H, rotameric mixture), 1.91(s, 3H), 3.18-3.40(6H, rotameric mixture), 3.52-3.76(m,4H), 4.01-4.49(m, 6H), 4.69-5.16(m,2H), 7.00,7.05(1H, rotameric mixture), 8.55,8.61(1H, rotameric mixture); ^{13}C NMR (50 MHz, CDCl$_3$): δ (ppm) 12.28, 14.16, 28.31, 45.93, 47.88, 56.64, 59.31, 61.24, 61.78, 71.36, 72.23, 79.99, 110.57, 141.64, 151.24, 164.70, 167.93, 168.75, 170.14;HRMS calcd for C$_{22}$H$_{36}$N$_4$O$_9$Na: 523.2374, Observed mass: 523.2387.

(4*R*, 5*R*)-ethyl2-(*N*-(3-(((*tert*-butoxycarbonyl)amino)-1,4-dimethoxybutan-2-yl)-2-thym inyl acetate (10a$_1$)
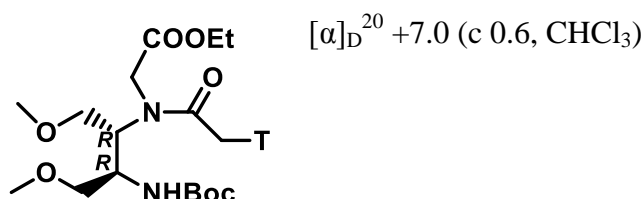
$[\alpha]_D^{20}$ +7.0 (c 0.6, CHCl$_3$)

(4*S*, 5*S*)-ethyl 2-(*N*-(1,4-bis(benzyloxy)-3-(((*tert*-butoxycarbonyl)amino)butan-2-yl)-2-thyminyl acetate (10b$_2$)
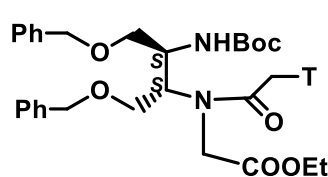
Compound **9b$_2$** (1g, 1.8mmol), activated K$_2$CO$_3$ (0.3g, 2.1mmol) and Thymine (0.25g, 1.9mmol) was suspended in 10mL of dry DMF and the reaction mixture was allowed to stir for 6h at rt. The solvent was removed under vacuum and the residue was extracted with ethyl acetate (3x50mL). The organic layer was washed with brine (3x25mL) and was dried over anhydrous Na$_2$SO$_4$. Ethyl acetate was removed under vacuum to get the crude compound. The crude compound was purified by column chromatography (30% EtOAc in Pet ether) to get white solid (0.99g, 83%); $[\alpha]_D^{20}$ -23.8(c 5, CHCl$_3$); ^{1}H NMR (200 MHz, CDCl$_3$): δ (ppm) 1.10-1.26(m, 3H, rotameric mixture), 1.38, 1.44(9H, rotameric mixture), 3.47-3.93(m, 4H), 4.01-4.24(m, 6H), 4.34-4.82(m, 6H), 7.22-7.45(m, 11H), 7.96-8.08(br s, 1H), 9.09-9.15(br s, 1H); ^{13}C NMR (50 MHz, CDCl$_3$): δ (ppm) 12.35, 14.19, 28.30, 45.97, 48.01, 56.83, 60.41, 61.07, 61.68, 69.14, 73.25, 79.63, 110.52, 127.18, 127.81, 128.39, 137.35, 141.01, 151.05, 155.88, 164.48, 168.57, 170.01; HRMS calcd for C$_{34}$H$_{44}$O$_9$N$_4$Na: 675.3001, Observed mass: 675.3010.

(4*R*, 5*R*)-ethyl 2-(*N*-(1,4-bis(benzyloxy)-3-((*tert*-butoxycarbonyl)amino)butan-2-yl)-2-thyminyl acetate (10a₂)
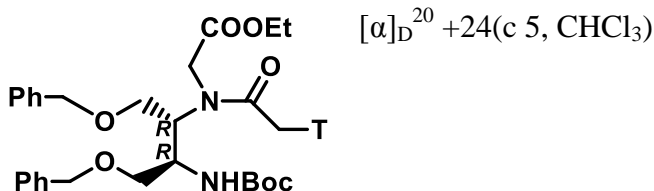
$[\alpha]_D^{20}$ +24(c 5, CHCl₃)

**(4*S*, 5*S*)-ethyl 2-(*N*-(3-((*tert*-butoxycarbonyl)amino)-1,4-dimethoxybutan-2-yl)-2-(6 benzyloxy carbonyl adeninyl) acetate (11b₁)**
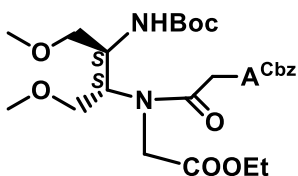
Compound **9b₁**(1.2g, 2.9mmol), activated K₂CO₃ (0.48g, 3.5 mmol) and N^6- benzyloxycarbonyladenine (0.87g, 3.2mmol) was suspended in 15mL of dry DMF and the reaction mixture was allowed to stir for 6h at rt. The solvent was removed under vacuum and the residue was extracted with ethyl acetate (3x50mL). The organic layer was washed with brine (3x25mL) and was dried over anhydrous Na₂SO₄. Ethyl acetate was removed under vacuum to get the crude compound. The crude compound was purified by column chromatography (30% EtOAc in Pet ether) to get white solid (1.3g, 72%); $[\alpha]_D^{20}$ -19.6(c 5, CHCl₃); ^{1}H NMR (200 MHz, CDCl₃): δ (ppm) 1.21-1.34(3H, rotameric mixture), 1.39,1.41(9H, rotameric mixture), 3.19-3.44(m, 7H, rotameric mixture), 3.53-3.82(m, 3H), 4.10-4.58(m, 5H), 4.86-5.46(m, 5H), 7.34-7.42(m, 5H), 8.10-8.11(s,1H, rotameric mixture), 8.72,8.75 (1H, rotameric mixture); ^{13}C NMR (50 MHz, CDCl₃): δ (ppm) 14.16, 28.23, 43.85, 56.91, 58.64, 59.41,61.12, 67.63, 71.85,79.56, 121.07, 128.40, 128.59, 135.58, 144.13, 149.43, 151.09, 152.75, 155.80, 166.73, 167.62, 169.85; HRMS calcd for C₃₀H₄₂O₉N₇: 644.3039, Observed mass: 644.3034.

(4*R*, 5*R*)-ethyl 2-(*N*-(3-((*tert*-butoxycarbonyl)amino)-1,4-dimethoxybutan-2-yl)-2-(6 benzyloxy carbonyl adeninyl) acetate (11a₁)
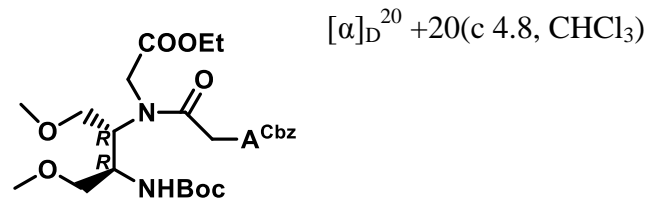
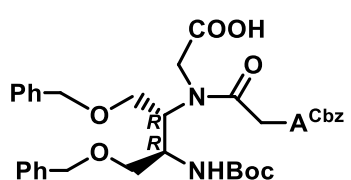$[\alpha]_D^{20}$ +20(c 4.8, CHCl₃)

(4*S*, 5*S*)-ethyl 2-(*N*-(1,4-bis(benzyloxy)-3-((*tert*-butoxycarbonyl)amino)butan-2-yl)-2-(6 benzyloxy carbonyl adeninyl) acetate (11b₂)
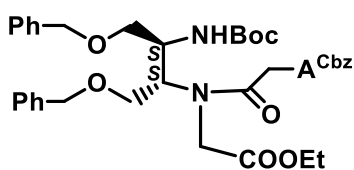
Compound **9b₂** (1g, 1.8mmol), activated K_2CO_3 (0.3g, 2.1 mmol) and N^6- benzyloxycarbonyladenine (0.5g, 1.9mmol) was suspended in 10mL of dry DMF and the reaction mixture was allowed to stir for 6h at r.t. The solvent was removed under vacuum and the residue was extracted with ethyl acetate (3x50mL). The organic layer was washed with brine (3x25mL) and was dried over anhydrous Na_2SO_4. Ethyl acetate was removed under vacuum to get the crude compound. The crude compound was purified by column chromatography (20% EtOAc in Pet ether) to get white solid (1.1g, 76%); $[\alpha]_D^{20}$ -26.5(c 5, CHCl₃); ^1H NMR (200 MHz, CDCl₃): δ (ppm) 1.08-1.17 (3H, rotameric mixture), 1.29, 1.38 (9H, rotameric mixture), 3.43-3.75(m, 4H, rotameric mixture), 3.9-4.31(m, 7H), 4.40-4.92 (m, 5H),5.27,5.28 (2H,rotameric mixture), 7.27(m, 15H), 7.68 (1H, rotameric mixture), 7.99 (1H, rotameric mixture); ^{13}C NMR (50 MHz, CDCl₃): δ (ppm) 13.92, 28.32, 47.39, 60.39, 60.84, 61.62, 68.02, 69.21, 69.28, 73.02, 73.27, 79.62, 127.41,127.59, 1258.35, 128.60, 129.10, 135.21, 135.47, 137.64, 139.44, 147.97, 148.66, 155.87, 162.03, 167.83, 167.94; HRMS calcd for $C_{42}H_{50}O_9N_7$: 796.3665, Observed mass: 796.3657

(4*R*, 5*R*)-ethyl 2-(*N*-(1,4-bis(benzyloxy)-3-((*tert*-butoxycarbonyl)amino)butan-2-yl)-2-(6 benzyloxy carbonyl adeninyl) acetate(11a₂)
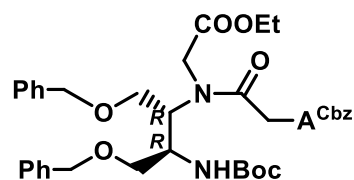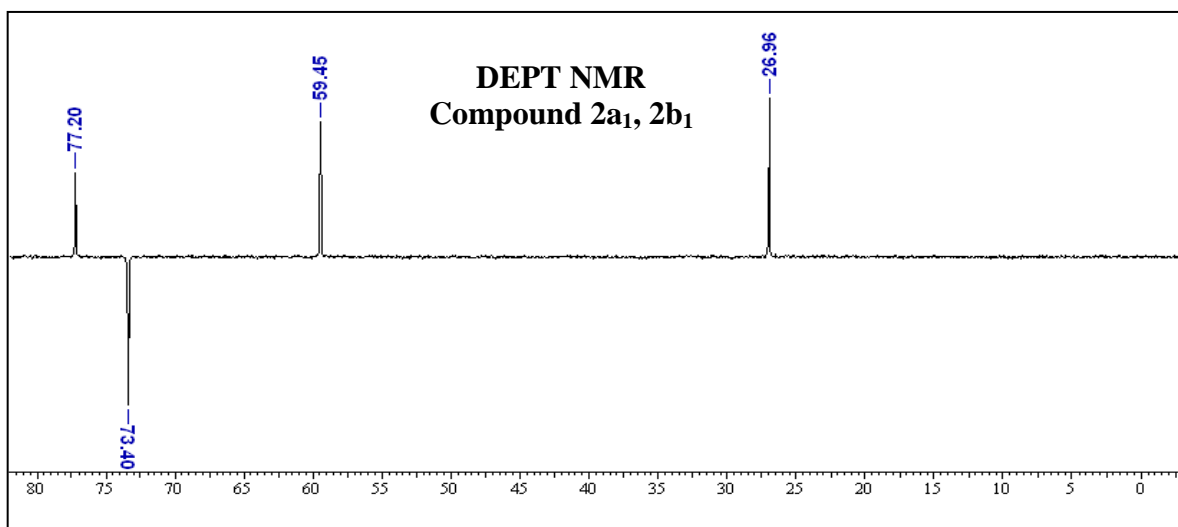
$[\alpha]_D^{20}$ +25.3(c 4.8, CHCl₃)

(4*S*, 5*S*)-2-(*N*-(3-((*tert*-butoxycarbonyl) amino)-1, 4-dimethoxybutan-2-yl)-2- thyminyl acetic acid (12b₁)
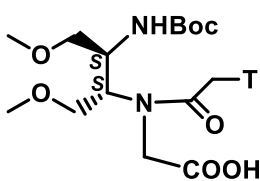
The compound **10b₁** (0.4g, 0.8mmol) was dissolved in 1mL methanol and to it 2mL of 1N LiOH solution was added. The completion of the reaction was monitored by TLC. After 30mins, methanol was removed under vacuum and the aqueous layer was neutralized with Dowex H⁺ resin. The resin was separated by filtration. The aqueous layer was washed with ethyl acetate (3x50mL) and the combined organic layer was lyophilized to get the free acid

(0.28g, 74%). $[\alpha]_D^{20}$ +10.8(c 1.3, MeOH); ^{1}H NMR (200 MHz, D$_2$O): δ (ppm) 1.32, 1.36 (9H, rotameric mixture), 1.78,1.82 (3H, rotameric mixture), 3.21-3.34 (6H, rotameric mixture), 3.48-3.65(m, 4H), 3.91-4.51(m, 6H), 7.28, 7.34 (1H, rotameric mixture); ^{13}C NMR (50 MHz, DMSO-d6): δ (ppm) 12.16, 48.17, 49.90, 58.28, 58.47, 70.63, 77.72, 79.41, 107.81, 138.09, 151.21, 151.77, 155.75, 164.64, 165.18; HRMS calcd for C$_{20}$H$_{32}$O$_9$N$_4$Na: 495.2061, Observed mass: 495.2060.

(4R, 5R)-2-(N-(3-((tert-butoxycarbonyl) amino)-1, 4-dimethoxybutan-2-yl)-2- thyminyl acetic acid (12a₁)

$[\alpha]_D^{20}$ -10.8(c 1.2, MeOH)
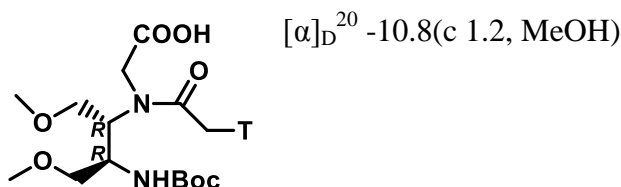
(4S, 5S)-2-(N-(1, 4-bis (benzyloxy)-3-((tert-butoxycarbonyl) amino) butan-2-yl)-2-thyminyl acetic acid (12b₂)
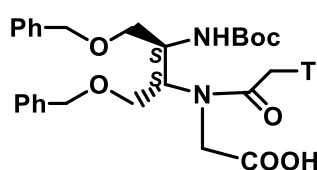
The compound **10b₂** (1g, 1.5mmol) was dissolved in 2mL methanol and to it 4mL of 1N LiOH solution was added**.** The completion of the reaction was monitored by TLC. After 30mins, methanol was removed under vacuum and the aqueous layer was neutralized with Dowex H$^+$ resin. The resin was separated by filtration. The aqueous layer was washed with ethyl acetate (3x50mL) and the combined organic layer was lyophilized to get the free acid (0.75g, 79%). $[\alpha]_D^{20}$ +28.3(c 1.5, MeOH); ^{1}H NMR (200 MHz, CD$_3$OD): δ (ppm) 1.29,1.32 (9H, rotameric mixture), 1.79 (s, 3H), 3.41-3.70(m, 4H), 3.74-4.13 (m, 4H), 4.17-4.65(m, 6H), 7.18-7.20(m, 1H); ^{13}C NMR (50 MHz, DMSO-d6): δ (ppm) 12.36, 28.59, 62.31, 62.55, 62.95, 68.51, 70.56, 72.28, 72.60, 78.14, 108.45, 127.87, 128.00, 128.57, 138.85, 142.72, 151.65, 156.02, 165.26, 168.03, 169.10; HRMS calcd for C$_{32}$H$_{40}$O$_9$N$_4$Na: 647.2687, Observed mass: 647.2687.

(4R, 5R)-2-(N-(1, 4-bis (benzyloxy)-3-((tert-butoxycarbonyl) amino) butan-2-yl)-2-thyminyl acetic acid (12a₂)
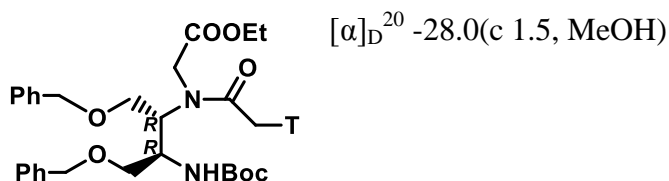
$[\alpha]_D^{20}$ -28.0(c 1.5, MeOH)

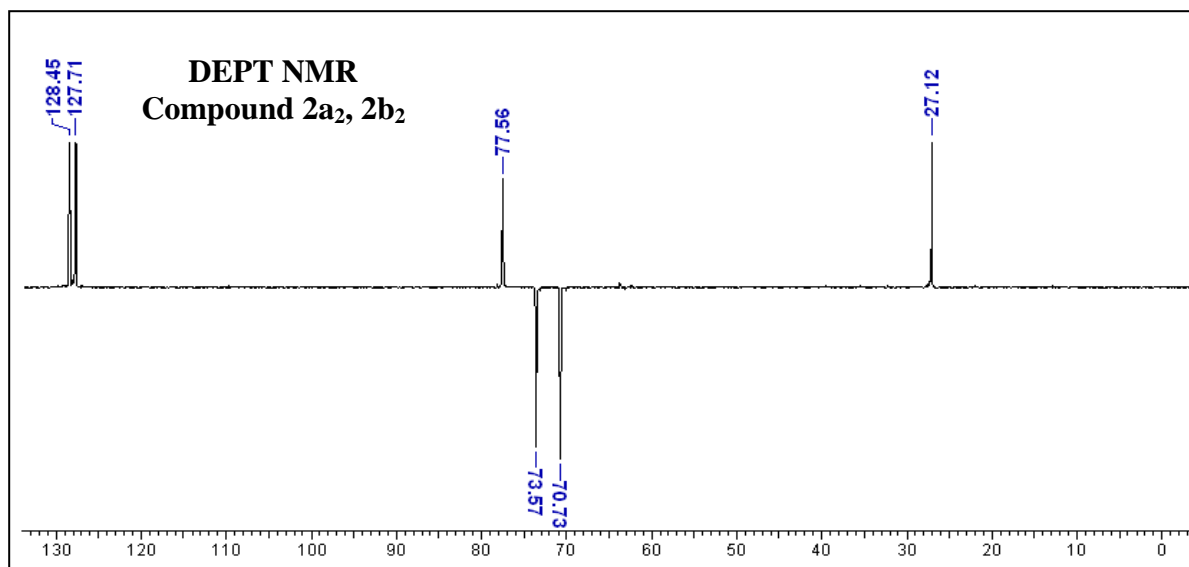**(4*S*, 5*S*)-2-(*N*-(3-((*tert*-butoxycarbonyl) amino)-1, 4-dimethoxybutan-2-yl)-2-(6-benzyl-oxy carbonyl adeninyl) acetic acid (13b₁)**
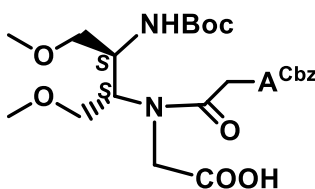
The compound **11b₁** (1g, 1.6mmol) was dissolved in 2mL methanol and to it 4mL of 1N LiOH solution was added**.** The completion of the reaction was monitored by TLC. After 30mins, methanol was removed under vacuum and the aqueous layer was neutralized with Dowex H⁺ resin. The resin was separated by filtration. The aqueous layer was washed with ethyl acetate (3x50mL) and the combined organic layer was lyophilized to get the free acid (0.68g, 71%). $[\alpha]_D^{20}$ +6.8(c 2.5, MeOH); ^{1}H NMR (400 MHz, CDCl₃): δ (ppm) 1.30, 1.42(9H, rotameric mixture), 3.14-3.30(6H, rotameric mixture), 3.42-3.68(m, 4H), 4.11-4.27(m,4H), 4.98-5.27(m, 4H), 7.33-7.35(m, 5H), 8.14-8.18(1H,rotameric mixture), 8.70, 8.72(1H, rotameric mixture); ^{13}C NMR (10 MHz, CDCl₃): δ (ppm) 28.36, 44.29, 46.34, 48.43, 49.31, 58.51, 58.95,59.15,67.68,70.98, 72.08, 79.57, 120.46, 128.46, 128.59, 135.48, 144.46, 149.15, 151.47, 152.74, 155.62, 155.83, 167.26, 172.67. HRMS calcd for $C_{28}H_{37}O_9N_7$: 616.2726, Observed mass: 616.2727

(4*R*, 5*R*)-2-(*N*-(3-((*tert*-butoxycarbonyl) amino)-1, 4-dimethoxybutan-2-yl)-2-(6 benzyl-oxy carbonyl adeninyl) acetic acid (13a₁)
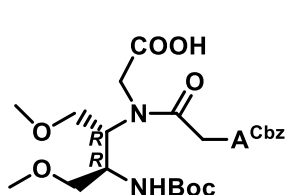
$[\alpha]_D^{20}$ -6.4(c 2.5, MeOH)

(4*S*, 5*S*)-2-(*N*-(1, 4-bis (benzyloxy)-3-((*tert*-butoxycarbonyl) amino) butan-2-yl)-2-(6 benzyloxy carbonyl adeninyl) acetic acid (13b₂)
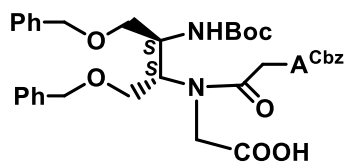
The compound **11b₂** (0.5g, 0.6mmol) was dissolved in 1mL methanol and to it 2mL of 1N LiOH solution was added**.** The completion of the reaction was monitored by TLC. After 30mins, methanol was removed under vacuum and the aqueous layer was neutralized with Dowex H⁺ resin. The resin was separated by filtration. The aqueous layer was washed with ethyl acetate (3x50mL) and the combined organic layer was lyophilized to procure the free acid (0.38g, 73%). $[\alpha]_D^{20}$ +26.7 (c 2.0, MeOH); ^{1}H NMR (200 MHz, CD₃OD): δ (ppm)

1.22,1.35(9H, rotameric mixture), 3.42-3.71(m,6H), 4.09-4.27(m, 4H), 4.33-4.55(m, 6H), 7.21-7.24(m, 15H), 8.03(s, 1H), 8.10(s, 1H); ^{13}C NMR (100 MHz, CDCl$_3$): δ (ppm) 28.33, 47.18, 57.82, 58.38, 65.08, 69.57, 72.94, 73.32, 73.81, 81.16, 119.14, 126.97, 127.80, 128.36, 129.02, 135.55, 137.68, 137.75, 140.53, 149.99, 153.13, 155.56, 165.90, 168.99; HRMS calcd for C$_{40}$H$_{44}$O$_9$N$_7$: 766.3195, Observed mass: 766.3198

(4*R*, 5*R*)-2-(*N*-(1, 4-bis (benzyloxy)-3-((*tert*-butoxycarbonyl) amino) butan-2-yl)-2-(6 benzyloxy carbonyl adeninyl) acetic acid (13a$_2$)

$[\alpha]_D^{20}$ -26.4(c 1.9, MeOH)

3A.7 Appendix

....

¹H NMR

Compound 2a₁, 2b₁

**¹³C NMR
Compound 2a₁, 2b₁**

**DEPT NMR
Compound 2a₁, 2b₁**

Chloroform-d

¹H NMR

Ph—O
Ph—O

Compound 2a₂, 2b₂

7.30
7.21
4.55
4.05
4.03
4.02
3.60
3.59
3.58
1.43
TMS
0.00

10.04 4.06 1.99 4.08 6.00

7.5 7.0 6.5 6.0 5.5 5.0 4.5 4.0 3.5 3.0 2.5 2.0 1.5 1.0 0.5 0.0

Chloroform-d

¹³C NMR
Compound 2a₂, 2b₂

137.96
128.34
127.60
109.62
77.47
77.76
77.12
73.47
70.64
27.01
TMS
0.00

140 130 120 110 100 90 80 70 60 50 40 30 20 10 0

DEPT NMR
Compound 2a₂, 2b₂

128.45
127.71
77.56
27.12
73.57
70.73

130 120 110 100 90 80 70 60 50 40 30 20 10 0

^{1}H NMR
Compound 3a$_1$, 3b$_1$

^{13}C NMR
Compound 3a$_1$, 3b$_1$

DEPT NMR
Compound 3a$_1$, 3b$_1$

¹H NMR

Ph—O—OH
Ph—O—OH

Compound 3a₂, 3b₂

¹³C NMR
Compound 3a₂, 3b₂

DEPT NMR
Compound 3a₂, 3b₂

¹H NMR

Compound 4a₁, 4b₁

¹³C NMR
Compound 4a₁, 4b₁

**DEPT NMR
Compound 4a₁, 4b₁**

¹H NMR
Compound 4a₂, 4b₂

¹³C NMR
Compound 4a₂, 4b₂

DEPT NMR
Compound 4a₂, 4b₂

^{1}H NMR

Compound 5a$_1$, 5b$_1$

^{13}C NMR
Compound 5a$_1$, 5b$_1$

DEPT NMR
Compound 5a$_1$, 5b$_1$

¹H NMR

Compound 5a₂, 5b₂

¹³C NMR
Compound 5a₂, 5b₂

DEPT NMR
Compound 5a₂, 5b₂

^{1}H NMR

NHBoc

Compound 7a$_1$, 7b$_1$

Chloroform-d
7.28
5.35
5.31
3.70
3.68-3.47
3.35
3.28
3.25
1.45
TMS
0.00

0.83   1.05 10.97   9.00

7.0 6.5 6.0 5.5 5.0 4.5 4.0 3.5 3.0 2.5 2.0 1.5 1.0 0.5 0.0

^{13}C NMR
Compound 7a$_1$, 7b$_1$

Chloroform-d
155.83
79.16
77.67
77.03
76.40
73.48
58.85
51.27
50.79
28.31
TMS
0.00

160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10 0

DEPT NMR
Compound 7a$_1$, 7b$_1$

58.91
51.33
50.83
28.36
75.07
73.53

75 70 65 60 55 50 45 40 35 30 25 20 15 10 5 0

¹H NMR
Compound 7a₂, 7b₂

¹³C NMR
Compound 7a₂, 7b₂

DEPT NMR
Compound 7a₂, 7b₂

¹H NMR

NHBoc

NH

COOEt

Compound 8a₁, 8b₁

Chloroform-d

TMS

¹³C NMR
Compound 8a₁, 8b₁

Chloroform-d

TMS

DEPT NMR
Compound 8a₁, 8b₁

TMS

^{1}H NMR

Compound 8a$_2$, 8b$_2$

^{13}C NMR
Compound 8a$_2$, 8b$_2$

DEPT NMR
Compound 8a$_2$, 8b$_2$

¹H NMR

Compound 9a₁, 9b₁

¹³C NMR
Compound 9a₁, 9b₁

DEPT NMR
Compound 9a₁, 9b₁

^{1}H NMR
Compound 9a$_2$, 9b$_2$

^{13}C NMR
Compound 9a$_2$, 9b$_2$

DEPT NMR
Compound 9a$_2$, 9b$_2$

¹H NMR

Compound 10a₁, 10b₁

Chloroform-d

8.61
8.55
7.27
7.05
7.00
5.16
5.12
5.03
4.77
4.69
4.49
4.34
3.75
3.70
3.55
3.54
3.40
3.33
3.27
3.18
1.92
1.91
1.44
1.40
1.32
1.26
1.23
TMS
0.00

0.79
0.85
2.01
6.07
4.00
6.03
3.09
9.00

¹³C NMR
Compound 10a₁, 10b₁

Chloroform-d
TMS

170.14
168.75
167.93
164.70
151.24
141.64
110.57
79.99
77.81
77.17
76.53
72.23
71.36
61.78
61.24
59.31
56.64
47.88
45.93
28.31
14.16
12.28
0.00

DEPT NMR
Compound 10a₁, 10b₁

TMS

141.64
59.31
28.31
14.22
12.29
0.00
71.97
71.36
61.79
61.25
47.75
45.94

¹H NMR
Compound 10a₂, 10b₂

¹³C NMR
Compound 10a₂, 10b₂

DEPT NMR
Compound 10a₂, 10b₂

^{1}H NMR

Compound 11a$_1$, 11b$_1$

^{13}C NMR
Compound 11a$_1$, 11b$_1$

DEPT NMR
Compound 11a$_1$, 11b$_1$

^{1}H NMR

Compound 11a$_2$, 11b$_2$

^{13}C NMR
Compound 11a$_2$,
11b$_2$

DEPT NMR
Compound 11a$_2$, 11b$_2$

¹H NMR
Compound 12a₁, 12b₁

¹³C NMR
Compound 12a₁, 12b₁

DEPT NMR
Compound 12a₁, 12b₁

¹H NMR

Compound
12a₂, 12b₂

¹³C NMR
Compound 12a₂, 12b₂

DEPT NMR
Compound 12a₂, 12b₂

^{1}H NMR
Compound 13a$_1$, 13b$_1$

^{13}C NMR
Compound 13a$_1$,
13b$_1$

DEPT NMR
Compound 13a$_1$, 13b$_1$

¹H NMR
Compound 13a₂, 13b₂

¹³C NMR
Compound
13a₂, 13b₂

DEPT NMR
Compound 13a₂, 13b₂

D:\Data\T-1   10/28/2014 4:04:48 PM

T-1 #979 RT: 4.36 AV: 1 NL: 1.62E8
T: FTMS + p ESI Full ms [66.70-1000.00]

213.1101
R=77607
$C_9 H_{18} O_4 Na = 213.1097$
1.7699 ppm

Compound 2a$_1$
Expected mass: 213.1097
Observed mass: 213.1101

187.0581
R=82400

191.0231
R=80602

205.0388
R=76502

209.0338
R=65702

217.0386
R=75002

220.9349
R=75702

226.9519
R=75702

232.9285
R=74502

239.0895
R=71807

245.0788
R=71107

m/z

D:\Data\T-2_141028161558   10/28/2014 4:15:58 PM

T-2_141028161558 #2103 RT: 9.37 AV: 1 NL: 3.34E5
T: FTMS + p ESI Full ms [66.70-1000.00]

365.1731
R=54402
$C_{21} H_{26} O_4 Na = 365.1723$
2.0955 ppm

Compound 2a$_2$
Expected mass: 365.1723
Observed mass: 365.1731

365.1066
R=52702

364.9641
R=38000

364.9200
R=39700

m/z

D:\Data\T-3   10/28/2014 4:27:06 PM

T-3 #867 RT: 3.86 AV: 1 NL: 1.66E9
T: FTMS + p ESI Full ms [66.70-1000.00]

Compound 3a$_1$
Expected mass: 173.0784
Observed mass: 173.0789

173.0789
R=86607
C$_6$ H$_{14}$ O$_4$ Na = 173.0784
2.7333 ppm

154.0573
R=90407
C$_6$ H$_{11}$ O$_3$ Na = 154.0600
-17.8095 ppm

134.9969
R=94207

143.0020
R=87907

179.0753
R=75006

195.0609
R=76807
C$_8$ H$_{12}$ O$_4$ Na = 195.0628
-9.4930 ppm

215.0896
R=77002
C$_{17}$ H$_{11}$ = 215.0855
18.8639 ppm

m/z

D:\Data\T-4   10/28/2014 4:38:16 PM

T-4 #985 RT: 4.39 AV: 1 NL: 1.54E8
T: FTMS + p ESI Full ms [66.70-1000.00]

Compound 3a$_2$
Expected mass: 325.1410
Observed mass: 325.1418

325.1418
R=62807
C$_{18}$ H$_{22}$ O$_4$ Na = 325.1410
2.3199 ppm

315.1572
R=59107
C$_{17}$ H$_{24}$ O$_4$ Na = 315.1567
1.6361 ppm

304.8960
R=61100

329.0057
R=58402

334.9157
R=45100

339.1211
R=60507

344.8762
R=58702

350.8933
R=59802

m/z

D:\Data\T-5   10/28/2014 4:49:26 PM

T-5 #878 RT: 3.92 AV: 1 NL: 4.00E8
T: FTMS + p ESI Full ms [66.70-1000.00]

329.0347
R=62807
$C_8 H_{18} O_8 Na S_2$ = 329.0335
3.4584 ppm

Compound 4a₁
Expected mass: 329.0335
Observed mass: 329.0347

293.0676   301.1421   308.0865   315.1583   324.0797   338.0948   345.0086   352.1106   362.9278   381.1210 387.2885
R=64807   R=64802   R=59502   R=54802   R=62407   R=59507   R=60007   R=58402   R=55402   R=44900 R=44100

290   300   310   320   330   340   350   360   370   380   390
m/z

D:\Data\T-6   10/28/2014 5:00:37 PM

T-6 #913 RT: 4.07 AV: 1 NL: 7.64E8
T: FTMS + p ESI Full ms [66.70-1000.00]

481.0977
R=51307
$C_{20} H_{26} O_8 Na S_2$ = 481.0961
3.2353 ppm

Compound 4a₂
Expected mass: 481.0961
Observed mass: 481.0977

459.1104
R=49204
$C_{20} H_{27} O_8 S_2$ = 459.1142
-8.3510 ppm

476.1427   490.1586   497.0712   504.1737   518.1885   532.2048
R=51807   R=44807   R=50107   R=48907   R=45607   R=45102

430   440   450   460   470   480   490   500   510   520   530   540
m/z

Compound 5a$_2$
Expected mass: 375.1540
Observed mass: 375.1537

Compound 7a$_1$
Expected mass: 249.1809
Observed mass: 249.1815

D:\Data\TB-3 12/2/2014 3:06:46 PM

TB-3 #462 RT: 2.06 AV: 1 NL: 1.00E9
T: FTMS + p ESI Full ms [66.70-1000.00]

401.2435
R=56307
$C_{23}H_{33}O_4N_2$ = 401.2435
0.1153 ppm

Compound 7a$_2$
Expected mass: 401.2435
Observed mass: 401.2435

D:\Data\T-8 10/28/2014 5:22:59 PM

T-8 #510 RT: 2.27 AV: 1 NL: 7.62E9
T: FTMS + p ESI Full ms [66.70-1000.00]

335.2183
R=61207
$C_{15}H_{31}O_6N_2$ = 335.2177
1.9677 ppm

Compound 8a$_1$
Expected mass: 335.2177
Observed mass: 335.2183

321.2024
R=60007
$C_{14}H_{29}O_6N_2$ = 321.2020
1.0865 ppm

357.2001
R=59907
$C_{15}H_{30}O_6N_2Na$ = 357.1996
1.4739 ppm

350.8932
R=26500

362.9254
R=42000

D:\Data\TB-4   12/2/2014 3:26:16 PM

TB-4 #540 RT: 2.41 AV: 1 NL: 2.47E9
T: FTMS + p ESI Full ms [66.70-1000.00]

487.2806
R=50907
$C_{27}H_{39}O_6N_2$ = 487.2803
0.7722 ppm

Compound 8a$_2$
Expected mass: 487.2803
Observed mass: 487.2806

473.2652
R=51707

509.2620
R=50007
$C_{27}H_{38}O_6N_2Na$ = 509.2622
-0.3615 ppm

525.2369
R=47507

D:\Data\T-10   10/28/2014 5:45:17 PM

T-10 #926 RT: 4.13 AV: 1 NL: 1.30E9
T: FTMS + p ESI Full ms [66.70-1000.00]

433.1727
R=54506
$C_{17}H_{31}O_7N_2ClNa$ = 433.1712
3.3897 ppm

Compound 9a$_1$
Expected mass: 433.1712
Observed mass: 433.1727

393.2244
R=56007

413.1907
R=55907

427.2068
R=54207

442.1861
R=51507

457.2170
R=52607

475.1617
R=48706

485.2125
R=42800

495.1434
R=50407

D:\Data\TB-5 12/2/2014 3:37:25 PM

TB-5 #701 RT: 3.13 AV: 1 NL: 1.98E4
T: FTMS + p ESI Full ms [66.70-1000.00]

Compound 9a$_2$
Expected mass: 563.2159
Observed mass: 563.2156

D:\Data\T-11 10/28/2014 5:56:25 PM

T-11 #862 RT: 3.84 AV: 1 NL: 8.01E8
T: FTMS + p ESI Full ms [66.70-1000.00]

Compound 10a$_1$
Expected mass: 523.2374
Observed mass: 523.2387

D:\Data\T-12   10/28/2014 6:07:33 PM

T-12 #1017   RT: 4.53   AV: 1   NL: 3.91E8
T: FTMS + p ESI Full ms [66.70-1000.00]

675.3010
R=43407
$C_{34}H_{44}O_9N_4Na = 675.3001$
1.3540 ppm

Compound 10a$_2$
Expected mass: 675.3001
Observed mass: 675.3010

653.3191
R=44002
$C_{34}H_{45}O_9N_4 = 653.3181$
1.5099 ppm

672.2933
R=42907

664.2661
R=38606

683.2839
R=41807

689.2798
R=42607

697.2826
R=42306

704.2706
R=33704

710.7850
R=32200

D:\Data\TB-6   12/2/2014 3:48:34 PM

TB-6 #912   RT: 4.07   AV: 1   NL: 9.31E8
T: FTMS + p ESI Full ms [66.70-1000.00]

644.3034
R=44307
$C_{30}H_{42}O_9N_7 = 644.3039$
-0.6930 ppm

Compound 11a$_1$
Expected mass: 644.3039
Observed mass: 644.3034

599.2670
R=42500
$C_{26}H_{38}O_8N_7Na = 599.2674$
-0.6315 ppm

626.2924
R=40307

658.2830
R=37806

666.2845
R=43507

682.2585
R=41307

692.2573
R=28000

702.3078
R=41307

718.2282
R=37100

D:\Data\TANAYA   12/8/2014 3:20:55 PM

TANAYA #1085 RT: 4.83 AV: 1 SB: 52 5.79 , 4.73-4.95 NL: 6.91E7
T: FTMS + p ESI Full ms [66.70-1000.00]

Compound 11a$_2$
Expected mass: 796.3665
Observed mass: 796.3657

796.3657
R=38791
$C_{42} H_{50} O_9 N_7$ = 796.3665
-0.8883 ppm

818.3471
R=37565
$C_{42} H_{49} O_9 N_7 Na$ = 818.3484
-1.5768 ppm

D:\Data\TB-8   12/2/2014 4:10:54 PM

TB-8 #990 RT: 4.42 AV: 1 NL: 3.73E7
T: FTMS + p ESI Full ms [66.70-1000.00]

495.2060
R=50407
$C_{20} H_{32} O_9 N_4 Na$ = 495.2061
-0.3767 ppm

Compound 12a$_1$
Expected mass: 495.2061
Observed mass: 495.2060

536.1653
R=48007
$C_{26} H_{23} O_5 N_7 Na$ = 536.1653
-0.0089 ppm

464.1661
R=49506
$C_{21} H_{26} O_9 N_3$ = 464.1664
-0.6013 ppm

484.2090
R=47300
$C_{26} H_{26} O_3 N_7$ = 484.2092
-0.3083 ppm

517.1880
R=49307

D:\Data\TB-9   12/2/2014 4:22:04 PM

TB-9 #1092   RT: 4.87   AV: 1   NL: 5.85E6
T: FTMS + p ESI Full ms [66.70-1000.00]

647.2687
R=43707
$C_{32} H_{40} O_9 N_4 Na = 647.2687$
-0.0187 ppm

Compound 12a$_2$
Expected mass: 647.2687
Observed mass: 647.2687

610.1840
R=44606

615.1394
R=43607

663.2340
R=39107

669.2508
R=42107

605.4753
R=27900

632.1140
R=28800

644.2589
R=38300

655.2512
R=43400

679.2051
R=31700

685.4348
R=35400

Relative Abundance

m/z

D:\Data\TB-10   12/2/2014 4:33:12 PM

TB-10 #960   RT: 4.28   AV: 1   NL: 5.62E7
T: FTMS + p ESI Full ms [66.70-1000.00]

616.2727
R=45307
$C_{28} H_{38} O_9 N_7 = 616.2726$
0.2482 ppm

Compound 13a$_1$
Expected mass: 616.2726
Observed mass: 616.2727

605.4756
R=33600

614.2582
R=42202

626.2932
R=43907

632.2664
R=31000

Relative Abundance

m/z

Compound 13a₂
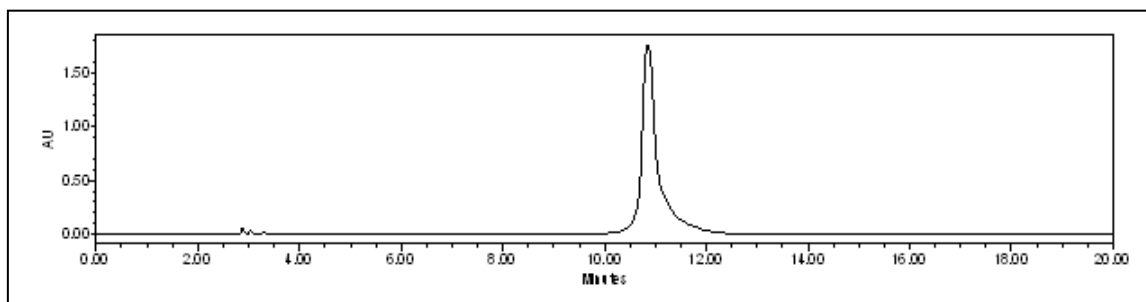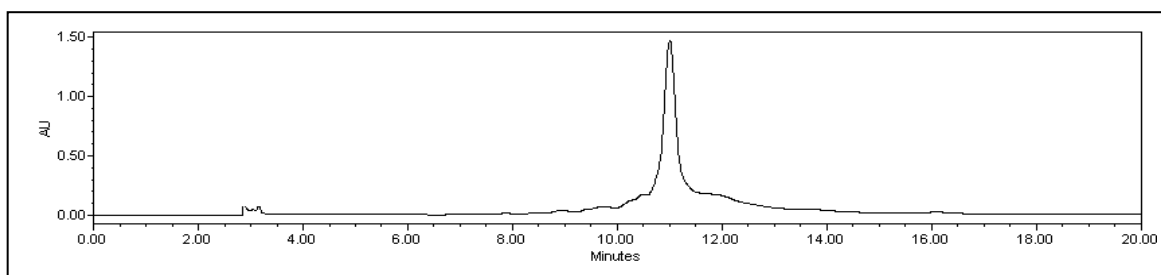**Expected mass**: 766.3195
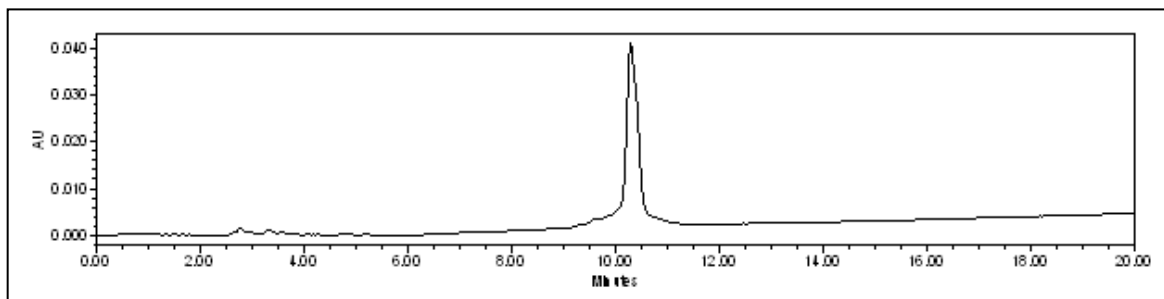**Observed mass**: 766.3198

Chiral HPLC of 10a₁/10b₁

Chiral HPLC was done on Kromasil 5-Amycoat (4.6x250mm) column in mobile phase isopropyl alcohol: Petroleum ether (30:70)
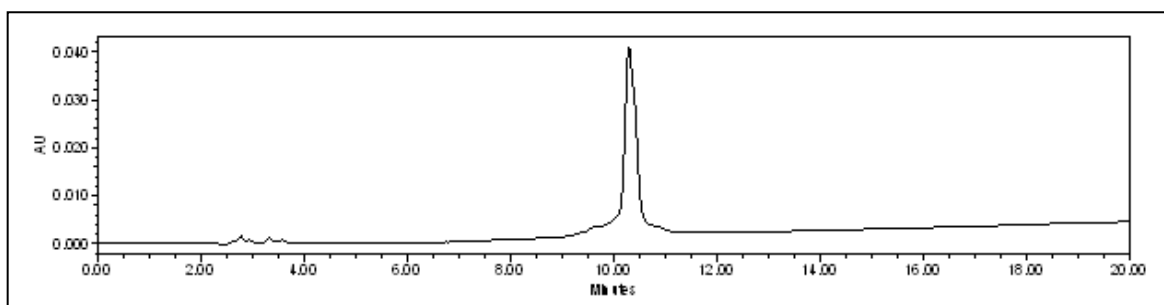
a) Chiral HPLC of **10a₁ + 10b₁**
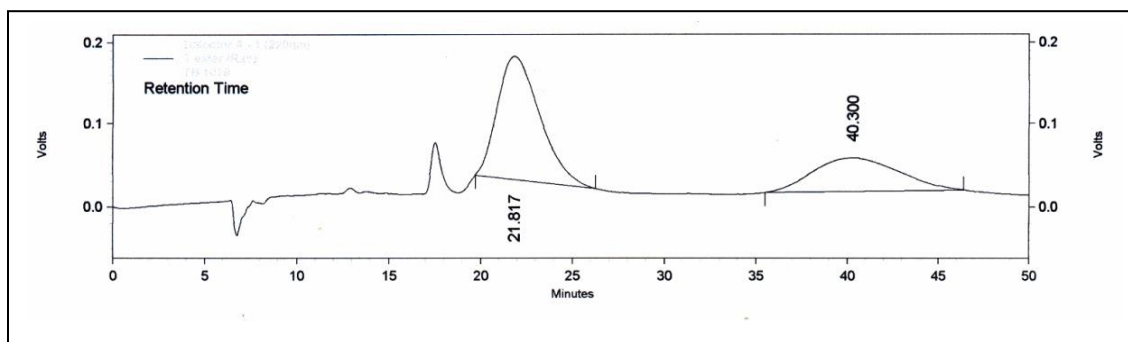
b) Chiral HPLC of **10a₁**
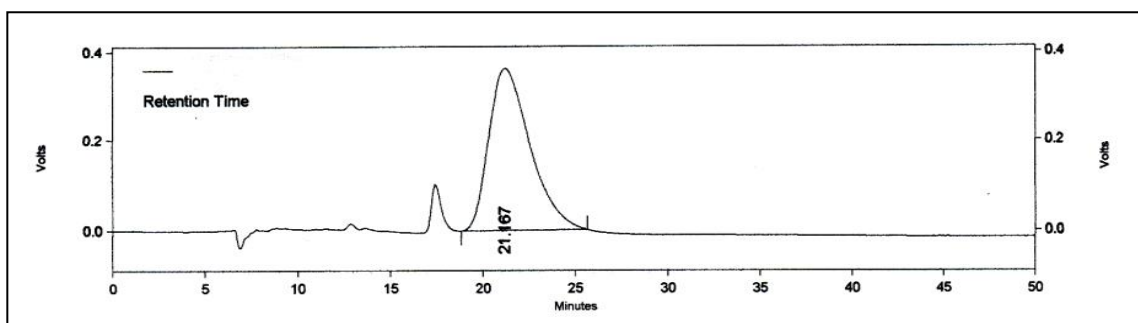
c) Chiral HPLC of **10b₁**
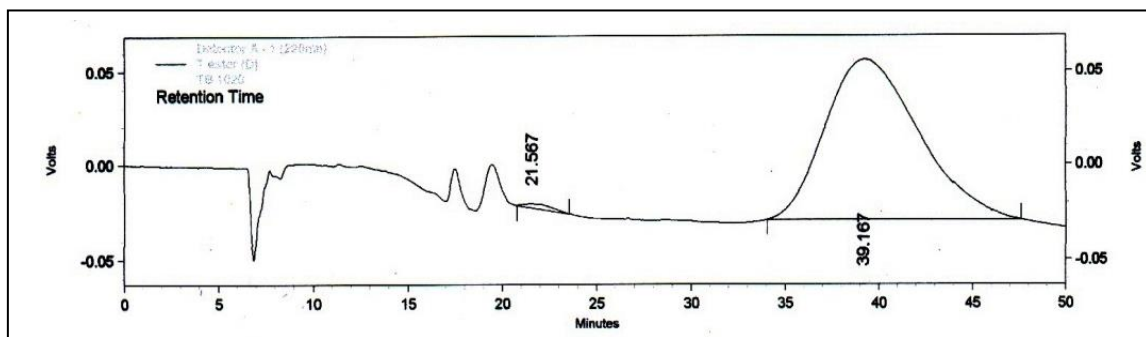
HPLC chromatogram of purified **PNA 1**
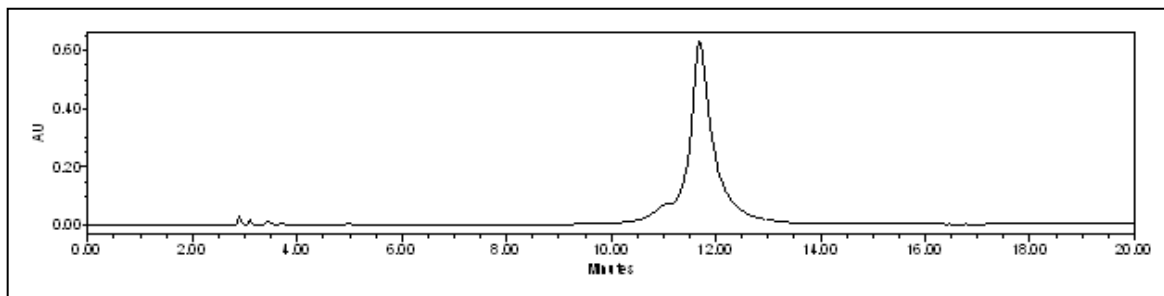

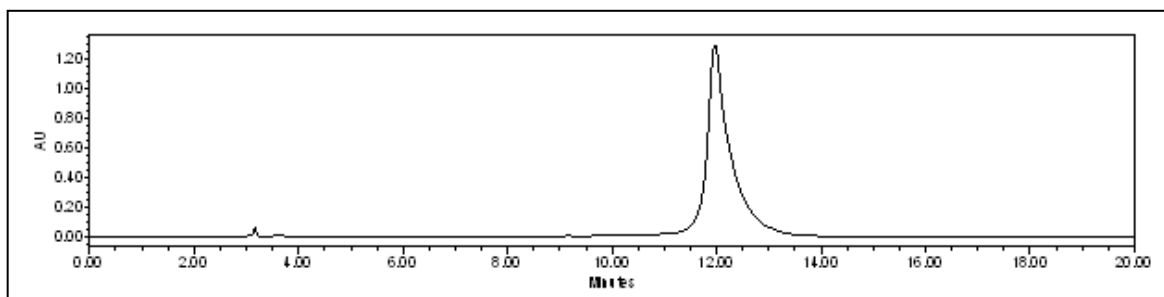HPLC chromatogram of purified **PNA 2**

HPLC chromatogram of purified **PNA 3a**

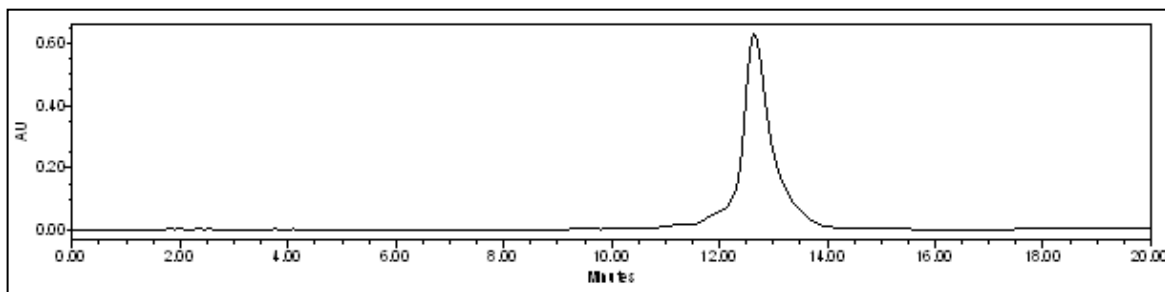HPLC chromatogram of purified **PNA 3b**

HPLC chromatogram of purified **PNA 4a**

HPLC chromatogram of purified **PNA 4b**

HPLC chromatogram of purified **PNA 5a**

HPLC chromatogram of purified **PNA 5b**

HPLC chromatogram of purified **PNA 6a**

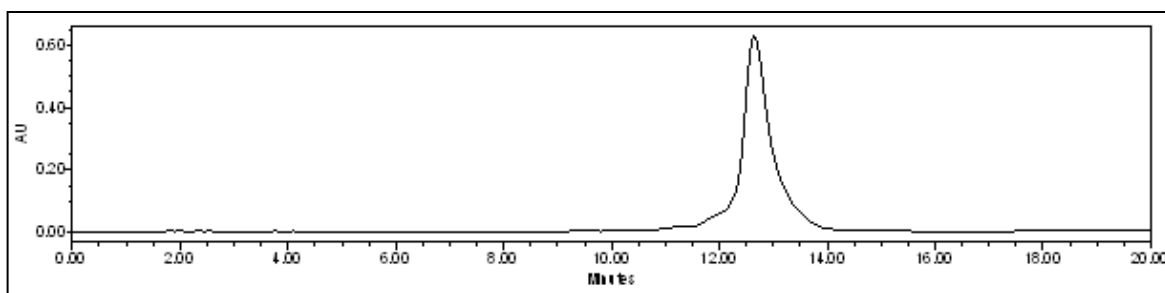**Fig S17:** HPLC chromatogram of purified **PNA 6b**

HPLC chromatogram of purified **PNA 7a**

HPLC chromatogram of purified **PNA 7b**

HPLC chromatogram of purified **PNA 8a**
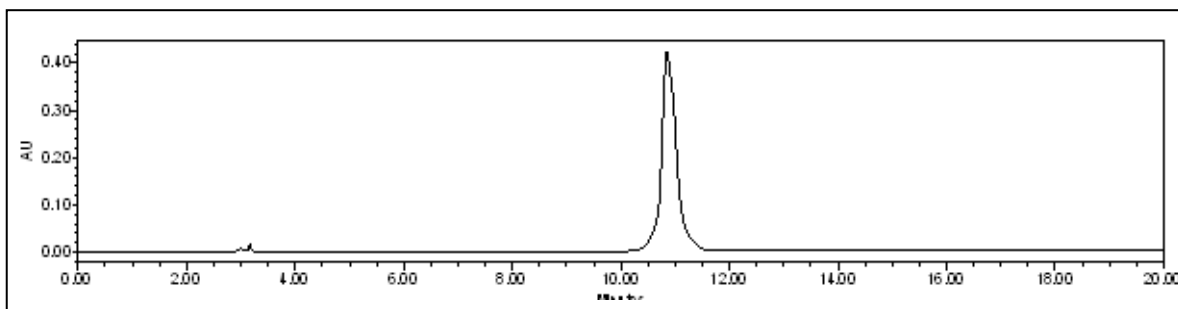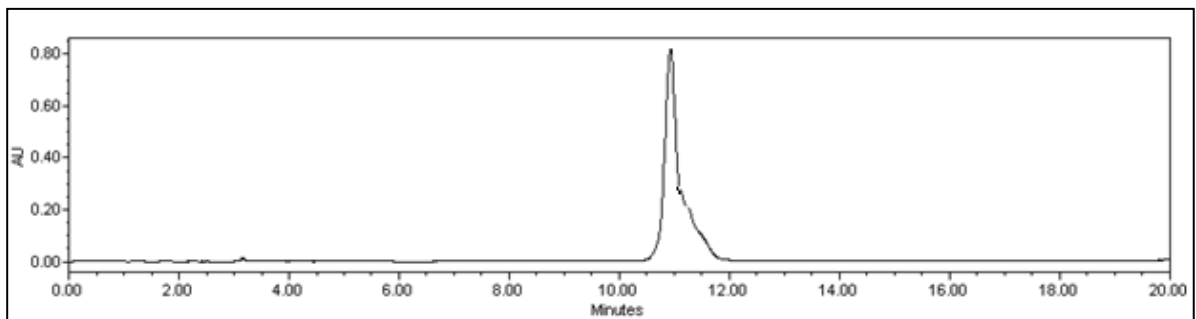

HPLC chromatogram of purified **PNA 8b**
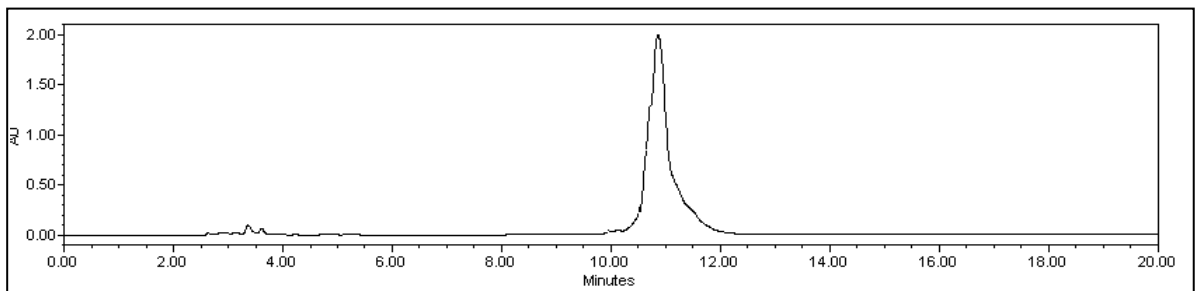
HPLC chromatogram of purified **PNA 9a**
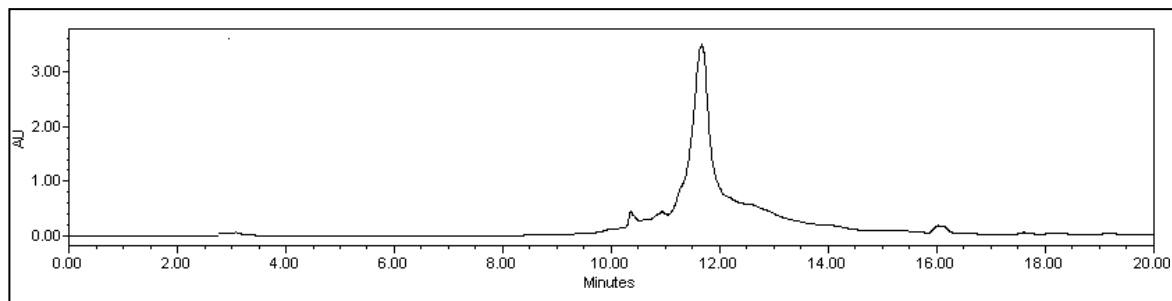
HPLC chromatogram of purified **PNA 9b**

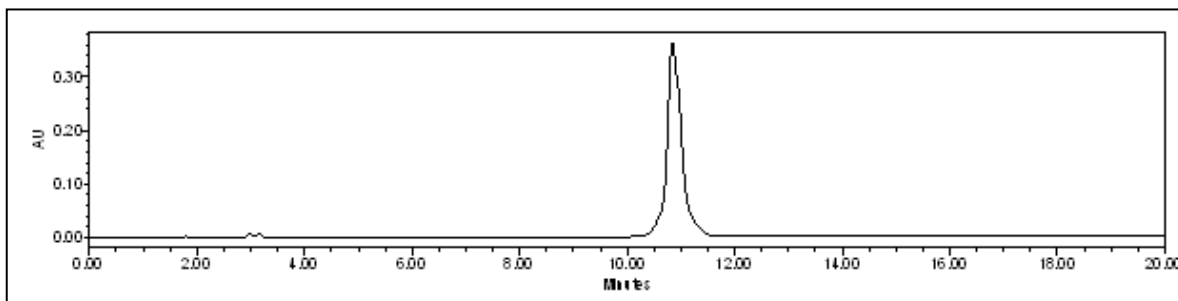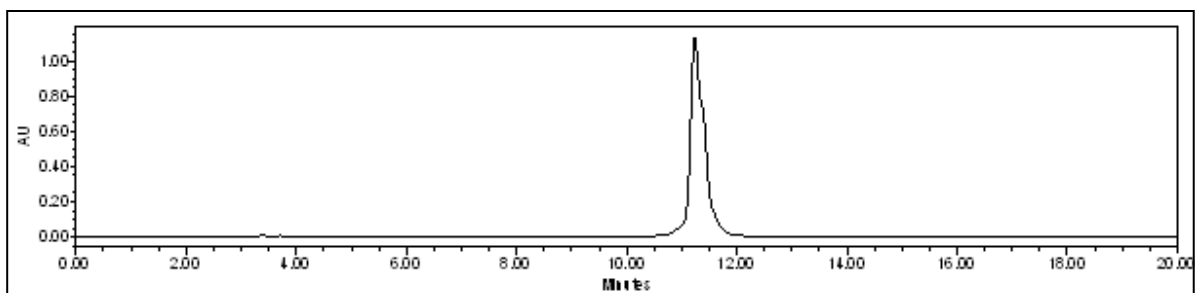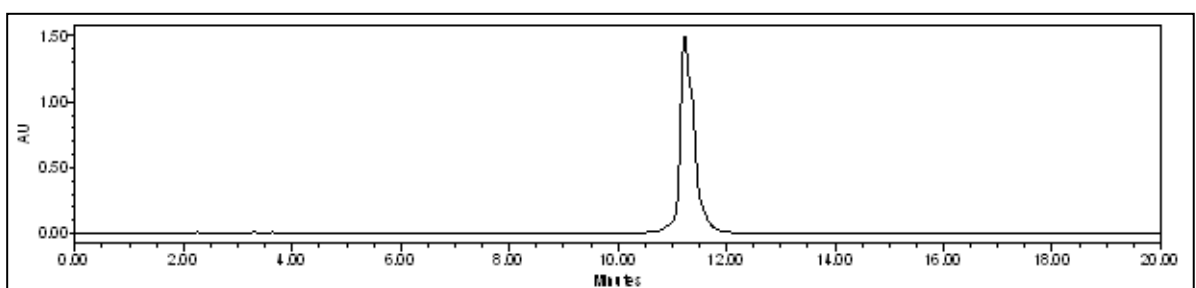HPLC chromatogram of purified **PNA 10a**

PNA 3a
Expected MW: 3439.4
Observed MW: 3478.6 (M+K)

PNA 3b
Expected MW: 3439.4
Observed MW: 3477.9(M+K)

PNA 5a
Expected MW: 3555.5
Observed MW: 3558.2

PNA 5b
Expected MW: 3555.5
Observed MW: 3597.0 (M+K)

PNA 6a
Expected MW: 3439.42
Observed MW: 3464.9 (M+Na)

PNA 6b
Expected MW: 3439.42
Observed MW: 3478.2 (M+K)

PNA 7a
Expected MW: 3467.5
Observed MW: 3491.8(M+Na)

PNA 7b
Expected MW: 3467.5
Observed MW: 3490.7 (M+K)

PNA 8a
Expected MW: 3499.4
Observed MW: 3505.0

PNA 8b
Expected MW: 3499.4
Observed MW: 3497.5

AB Sciex TOF/TOF™ Series Explorer™ 72085
TOF/TOF™ Linear Spec #1=>AdvBC(32,0.5,0.1)=>AdvBC(10,0.5,0.1)=>SM5[BP = 3597.0, 766]

PNA 9a
Expected MW: 3555.5
Observed MW: 3598.0(M+K)

AB Sciex TOF/TOF™ Series Explorer™ 72085
TOF/TOF™ Linear Spec #1=>AdvBC(32,0.5,0.1)[BP = 3593.2, 34947]

PNA 9b
Expected MW: 3555.5
Observed MW: 3593.3(M+K)

TOF/TOF™ Linear Spec #1=>BC=>AdvBC(32,0.5,0.1)=>SM5[BP = 4101.8, 1737]

PNA 10a
Expected MW: 4067.88
Observed MW: 4104.7(M+K)

Section B

Biophysical evaluation of (*R, R*)/(*S, S*) β,γ-Bis Diol/Dimethoxy *aeg*PNA incorporated oligomers

3B.1 Synthesis of complementary Oligonucleotides

The complementary and parallel complementary DNA oligomers were synthesized on Bioautomation Mer-Made 4 synthesizer using standard β-cyanoethyl phosphoramidite chemistry. The complementary RNA and mismatch oligonucleotides were obtained commercially.

List of DNA sequences

cDNA: 5'-CTGAAATCGGTT;

Parallel DNA: 5'-TTGGCTAAAGTC

List of RNA sequences

cRNA: 5'-CUGAAAUCGGUU

Mismatch RNA: CUGAA*U*UCGGUU

3B.2 UV-Melting Study

The T_m values of the modified PNAs hybridized with cDNA in antiparallel (*ap*) and parallel (*p*) orientations, antiparallel complementary RNA (cRNA) and mismatched RNA (mmRNA) in 1: 1 stoichiometry were determined by temperature-dependent UV absorbance plots. The T_m values were determined at 10mM and 100mM salt concentrations. The sample preparation and instrumental procedure is same as described in *chapter 2*.

Table 2: UV melting at 10mM NaCl concentration of aaccgaT^{OH}ttcag-K

Seq Code	Sequences	*ap*-cDNA	*p*-cDNA	cRNA	mmRNA
PNA 1	aaccgatttcag-K	58.3	46.4	63.1	46.3
PNA 3a	aaccga$^{(R, R)}$**T^{OH}**ttcag-K	n.t.	n.t.	n.t.	n.t.
PNA 3b	aaccga$^{(S, S)}$**T^{OH}**ttcag-K	n.t.	n.t.	n.t.	n.t.
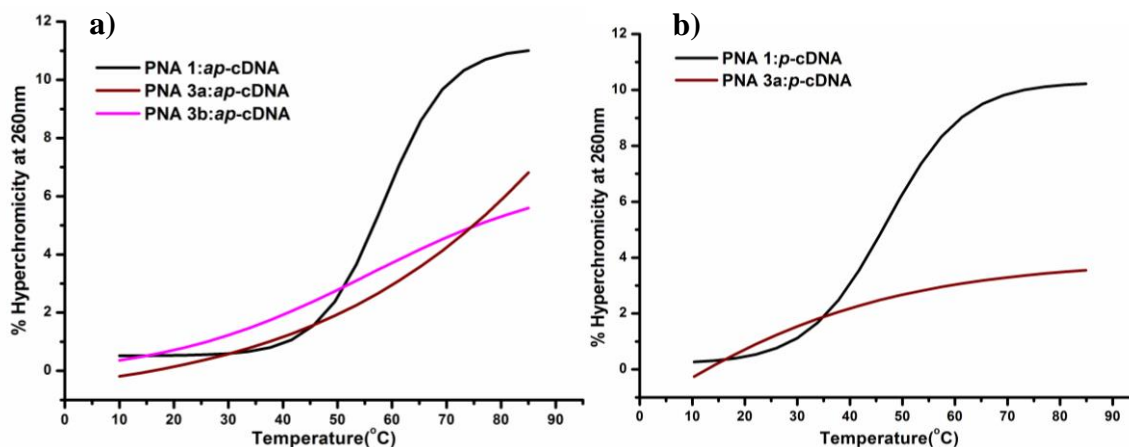
Figure 3.3: UV melting plots of PNA 1, 3a, 3b with a) *ap*-cDNA b) *p*-cDNA at 10mM NaCl concentration

Figure 3.4: UV melting plots of PNA 1, 3a, 3b with a) cRNA b) mmRNA at 10mM NaCl concentration

Table 3: UV meltings at 100mM NaCl salt concentration of aaccgaT^OHttcag-K

Seq Code	Sequences	*ap*-cDNA	cRNA
PNA 1	aaccgatttcag-K	53.7	60.9
PNA 3a	aaccga$^{(R, R)}$T^{OH}ttcag-K	n.t.	n.t.
PNA 3b	aaccga$^{(S, S)}$T^{OH}ttcag-K	n.t.	n.t.

Figure 3.5: UV melting plots of PNA 1, 3a, 3b with a) *ap*-cDNA b) cRNA at 100mM NaCl concentration

Table 4: UV meltings at 10 mM NaCl salt concentration of aaccga**T**^{OMe}ttcag-K

Seq Code	Sequences	*ap*-cDNA	*p*-cDNA	cRNA	mmRNA
PNA 1	aaccgatttcag-K	58.3	46.4	63.1	46.3
PNA 4a	aaccga$^{(R, R)}$**T**^{OMe}ttcag-K	59.2	38.5	64.7	51.8
PNA 4b	aaccga$^{(S, S)}$**T**^{OMe}ttcag-K	43.6	n.d.	51.7	42.8

Figure 3.6: UV melting plots of PNA 1, 4a, 4b with a) *ap*-cDNA **b)** *p*-cDNA **c) cRNA d) mmRNA at 10mM NaCl concentration**

Table 5: UV meltings at 100mM NaCl salt concentration of aaccgaT^{OMe}ttcag-K

Seq Code	Sequences	*ap*-cDNA	cRNA
PNA 1	aaccgatttcag-K	53.7	60.9
PNA 4a	aaccga$^{(R, R)}$T^{OMe}ttcag-K	53.7	59.9
PNA 4b	aaccga$^{(S, S)}$T^{OMe}ttcag-K	36.1	44.3

Figure 3.7: UV melting plots of PNA 1, 4a, 4b with a) *ap*-cDNA b) cRNA at 100mM NaCl concentration

Table 6: **UV melting at 10mM NaCl concentration of** aaccgaT^{OMe}tT^{OMe}cag-K

Seq Code	Sequences	*ap*-cDNA	*p*-cDNA	cRNA	mmRNA
PNA 1	aaccgatttcag-K	58.3	46.4	63.1	46.3
PNA 5a	aaccga$^{(R, R)}$T^{OMe}t$^{(R, R)}$T^{OMe}cag-K	59.2	38.5	64.7	51.8
PNA 5b	aaccga$^{(S, S)}$T^{OMe}t$^{(S, S)}$T^{OMe}cag-K	43.6	n.d.	51.7	42.8

Figure 3.8: UV melting plots of PNA 1, 5a, 5b with a) *ap*-cDNA b) *p*-RNA c) cRNA d) mmRNA at 10mM NaCl concentration

Table 7: UV melting at 100mM NaCl concentration of of aaccga $T^{OMe}tT^{OMe}$ cag-K

Seq Code	Sequences	*ap*-cDNA	cRNA
PNA 1	aaccgatttcag-K	53.7	60.9
PNA 5a	aaccga $^{(R, R)}T^{OMe}t^{(R, R)}T^{OMe}$cag-K	54.5	63.9
PNA 5b	aaccga $^{(S, S)}T^{OMe}t^{(S, S)}T^{OMe}$cag-K	n.t.	n.t.

Figure 3.9: UV melting plots of PNA 1, 5a, 5b with a) *ap*-cDNA b) cRNA at 100mM NaCl concentration

Table 8: UV melting at 10mM salt concentration of aaccg**A**^{OH}tttcag-K

Seq Code	Sequences	*ap*-cDNA	*p*-cDNA	cRNA	mmRNA
PNA 1	aaccgatttcag-K	58.3	46.4	63.1	46.3
PNA 6a	aaccg$^{(R, R)}$**A**^{OH}tttcag-K	n.t.	n.t.	n.t.	n.t.
PNA 6b	aaccg$^{(S, S)}$**A**^{OH}tttcag-K	n.t.	n.t.	n.t.	n.t.
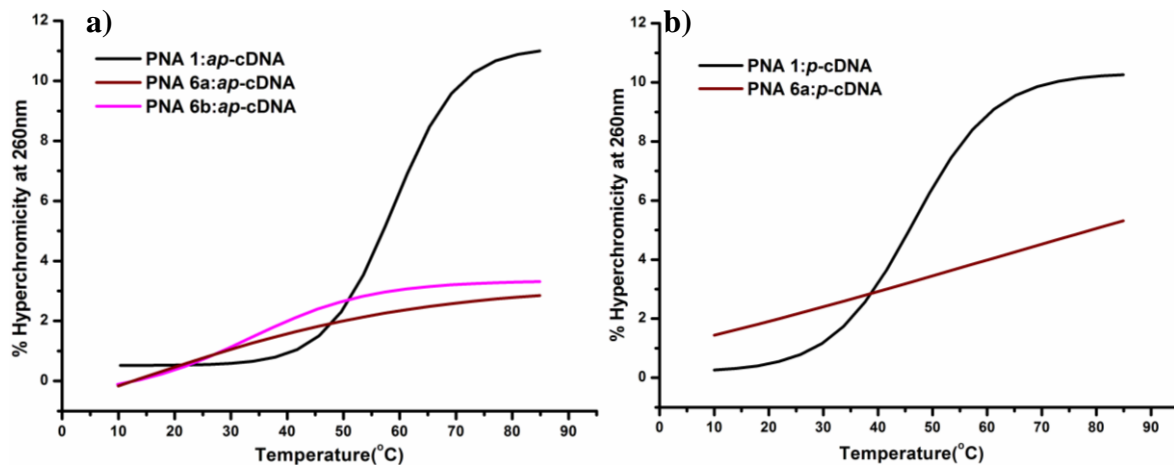
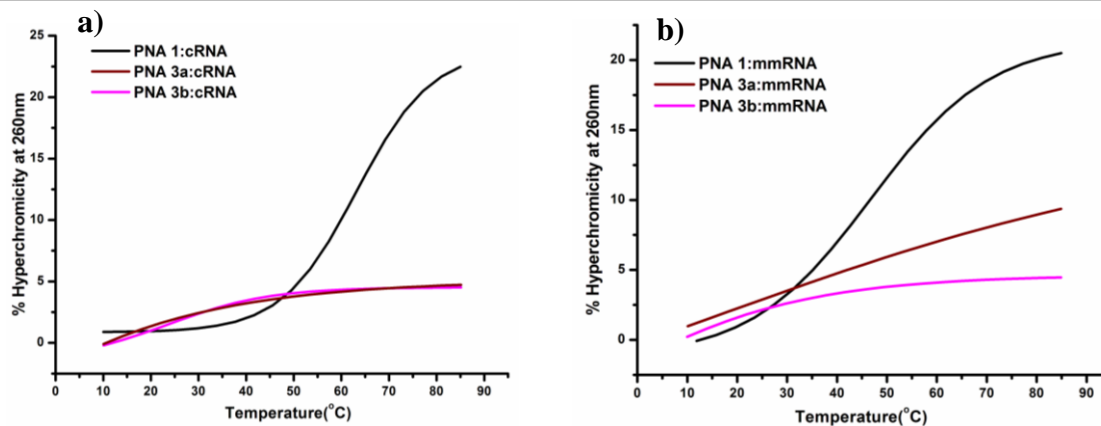**Figure 3.10: UV melting plots of PNA 1, 6a, 6b with a) *ap*-cDNA b) *p*-cDNA at 10mM NaCl concentration**

Figure 3.11: UV melting plots of PNA 1, 6a, 6b with a) cRNA b) mmRNA at 10mM NaCl concentration

Table 9: UV melting at 100mM NaCl concentration of aaccg$\mathbf{A^{OH}}$tttcag-K

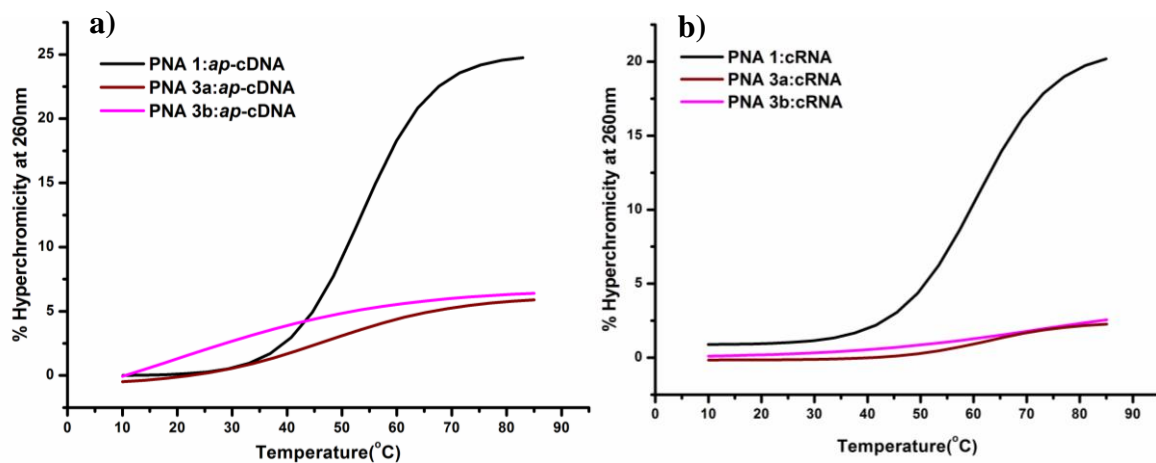| Seq Code	Sequences	*ap*-cDNA	cRNA
PNA 1	aaccgatttcag-K	53.7	60.9
PNA 6a	aaccg$^{(R, R)}$$\mathbf{A^{OH}}$tttcag-K	n.t.	n.t.
PNA 6b	aaccg$^{(S, S)}$$\mathbf{A^{OH}}$tttcag-K	n.t.	n.t.

Figure 3.12: UV melting plots of PNA 1, 6a, 6b with a) *ap*-cDNA b) cRNA at 100mM NaCl concentration

Table 10: UV melting at 10mM salt concentration of aaccgA^{OMe}tttcag-K

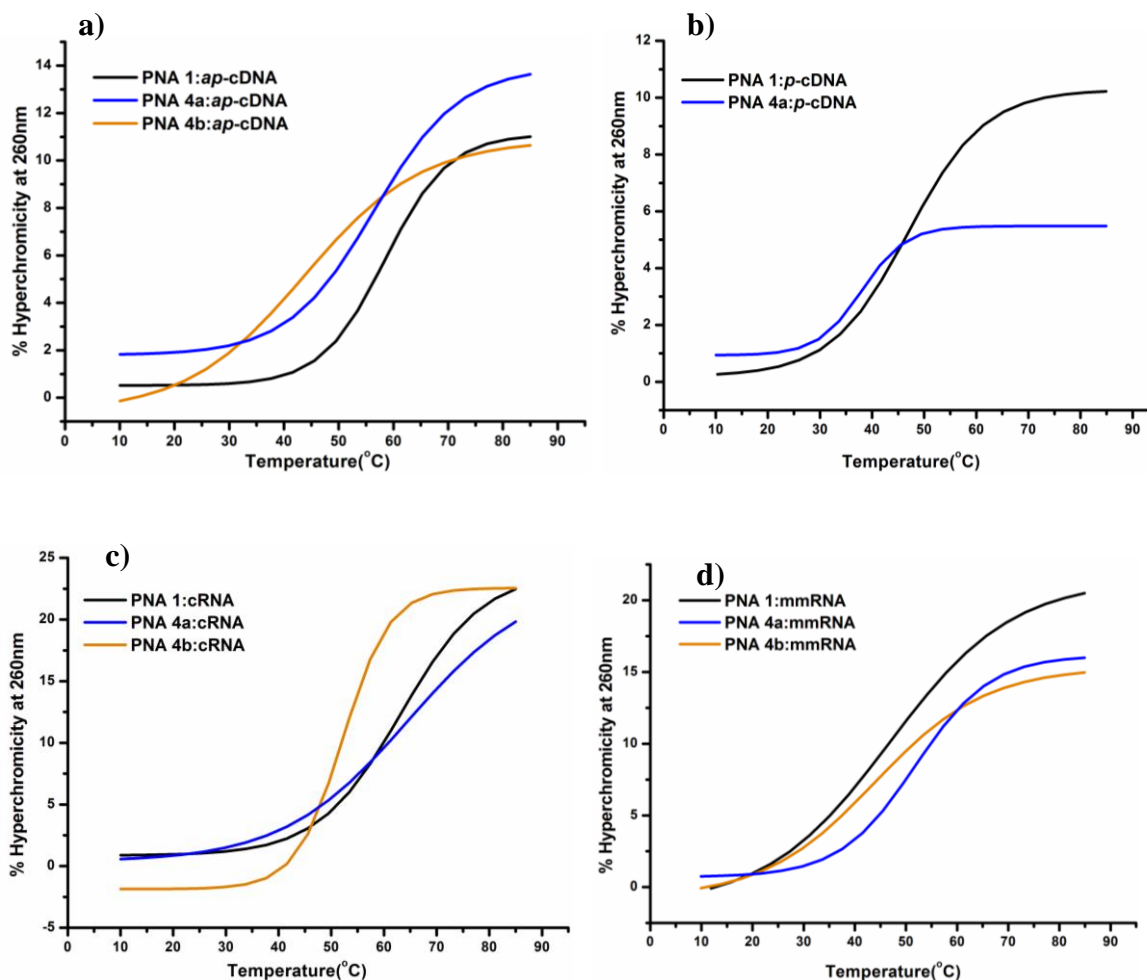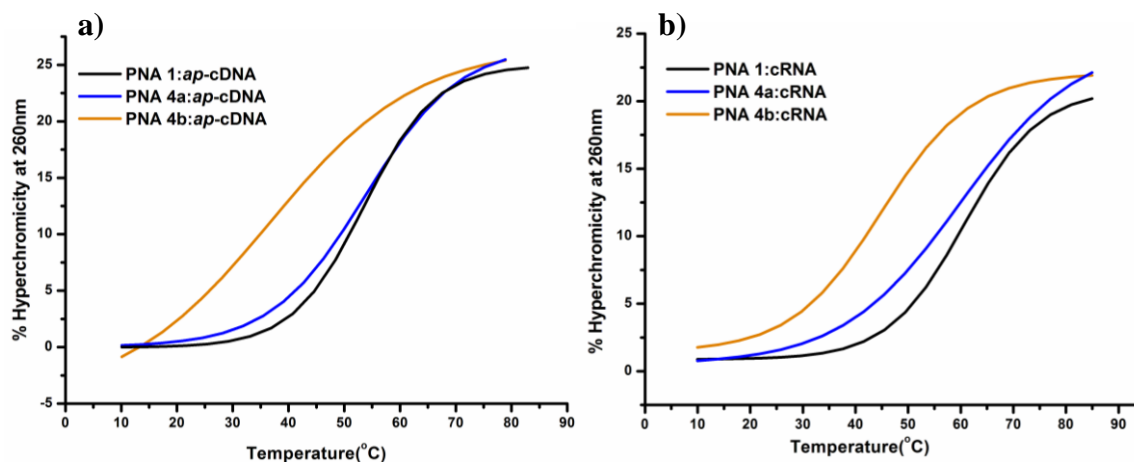| Seq Code	Sequences	*ap*-cDNA	*p*-cDNA	cRNA	mmRNA
PNA 1	aaccgatttcag-K	58.3	46.4	63.1	46.3
PNA 7a	aaccg$^{(R, R)}$A^{OMe}tttcag-K	58.2	38.8	61.7	52.0
PNA 7b	aaccg$^{(S, S)}$A^{OMe}tttcag-K	38.1	n.d.	49.2	40.4

Figure 3.13: UV melting plots of PNA 1, 7a, 7b with a) *ap*-cDNA b) *p*-cDNA c) cRNA d) mmRNA at 10mM NaCl concentration

Table 11: UV melting at 100mM NaCl concentration of aaccg**A**^{OMe}tttcag-K

Seq Code	Sequences	*ap*-cDNA	cRNA
PNA 1	aaccgatttcag-K	53.7	60.9
PNA 7a	aaccg$^{(R,R)}$**A**^{OMe}tttcag-K	50.7	57.3
PNA 7b	aaccg$^{(S,S)}$**A**^{OMe}tttcag-K	28.7	41.9
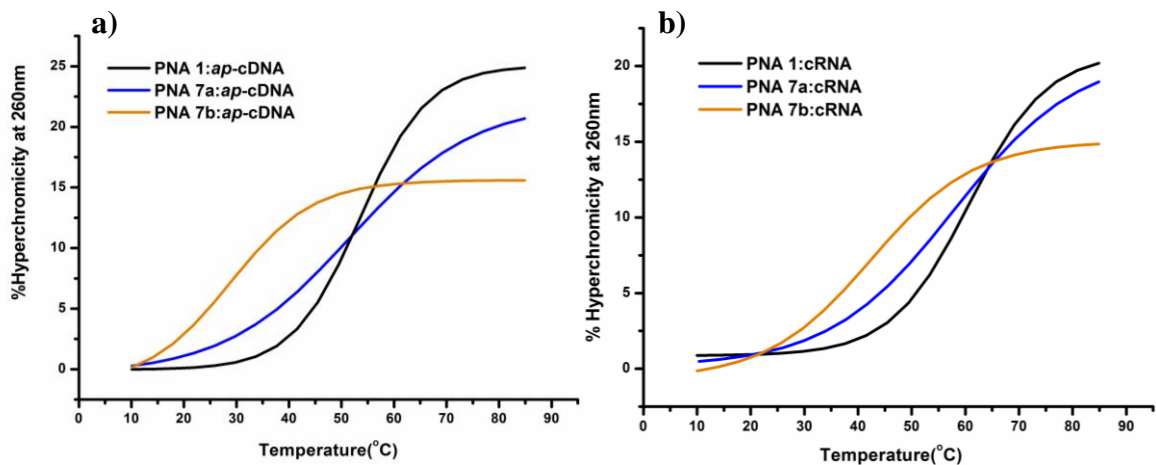
Figure 3.14: UV melting plots of PNA 1, 7a, 7b with a) *ap*-cDNA **b) cRNA at 100mM NaCl concentration**

Table 12: UVmelting at 10mM NaCl concentration of A^{OH}**A**^{OH}**CCGATTTCAG-K**

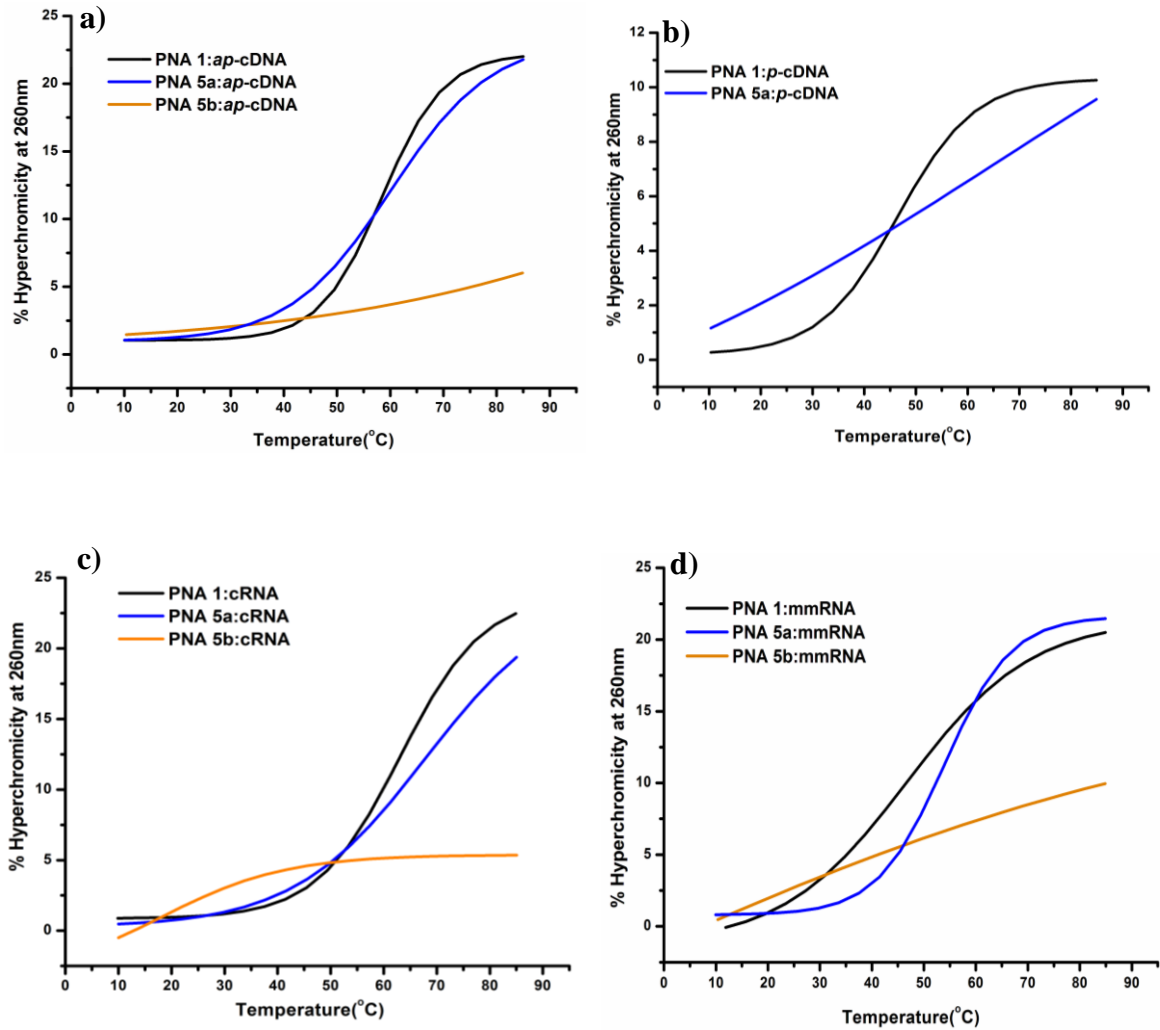| Seq Code	Sequences	*ap*-cDNA	*p*-cDNA	cRNA	mmRNA
PNA 1	aaccgatttcag-K	58.3	46.4	63.1	46.3
PNA 8a	$^{(R,R)}$**A**$^{OH(R,R)}$**A**^{OH}CCGATTTCAG-K	58.2	38.8	61.7	52.0
PNA 8b	$^{(S,S)}$**A**$^{OH(S,S)}$**A**^{OH}CCGATTTCAG-K	38.1	n.d.	49.2	40.4
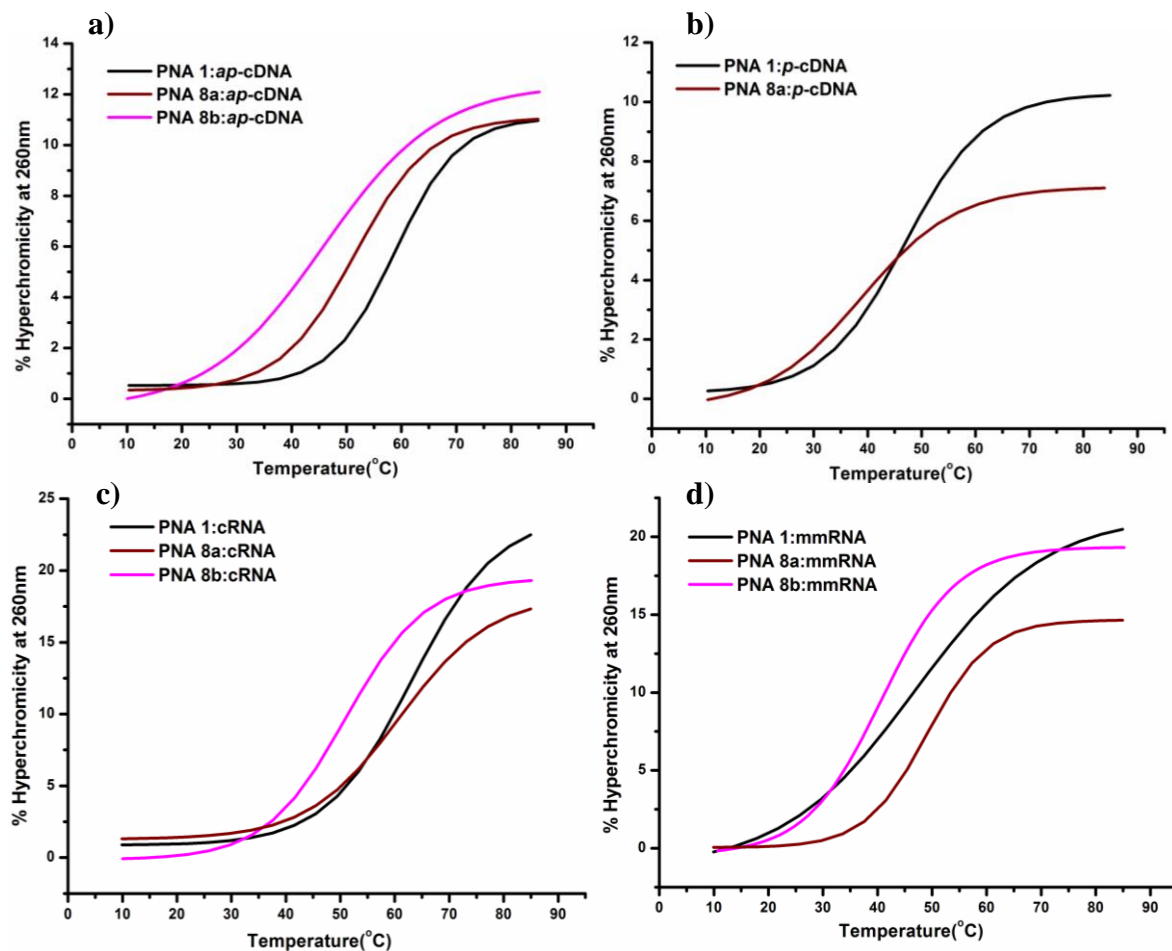
Figure 3.15: UV melting plots of PNA 1, 8a, 8b with a) *ap*-cDNA b) *p*-cDNA c) cRNA d) mmRNA at 10mM NaCl concentration

Table 13: UVmelting at 100mM NaCl concentration of $A^{OH}A^{OH}$CCGATTTCAG-K

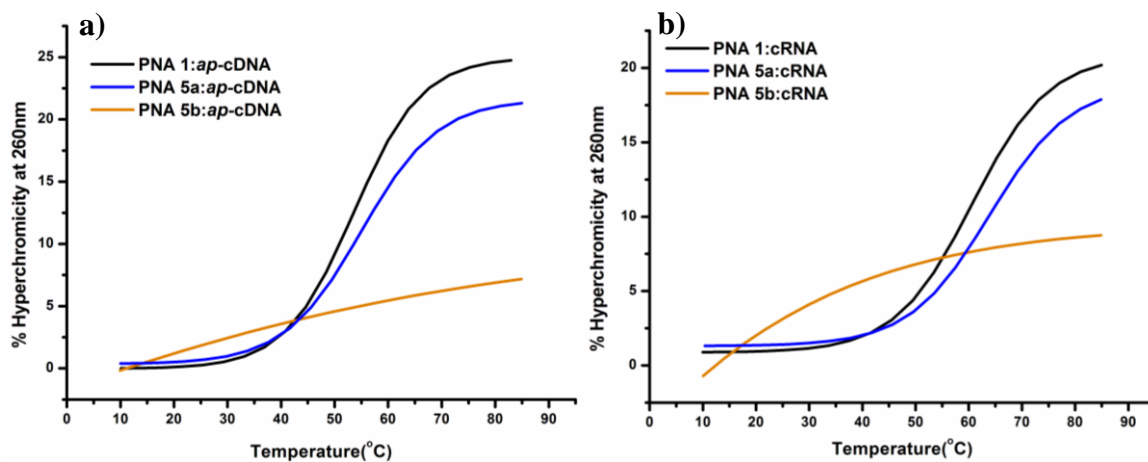| Seq Code	Sequences	*ap*-cDNA	cRNA
PNA 1	aaccgatttcag-K	53.7	60.9
PNA 8a	$^{(R, R)}A^{OH\,(R, R)}A^{OH}$CCGATTTCAG-K	49.7	58.5
PNA 8b	$^{(S, S)}A^{OH\,(S, S)}A^{OH}$CCGATTTCAG-K	42.1	46.6

Figure 3.16: UV melting plots of PNA 1, 8a, 8b with a) *ap*-cDNA **b)** cRNA at 100mM NaCl concentration

Table 14: UV meltings at 10mM NaCl concentration of A^{OMe}A^{OMe}ccgatttcag-K

| Seq Code	Sequences	*ap*-cDNA	*p*-cDNA	cRNA	mmRNA
PNA 1	aaccgatttcag-K	58.3	46.4	63.1	46.3
PNA 9a	$^{(R, R)}$A^{OMe}$^{(R, R)}$A^{OMe}ccgatttcag-K	57.6	38.5	63.2	50.3
PNA 9b	$^{(S, S)}$A^{OMe}$^{(S, S)}$A^{OMe}ccgatttcag-K	50.6	n.d.	57.3	37.8
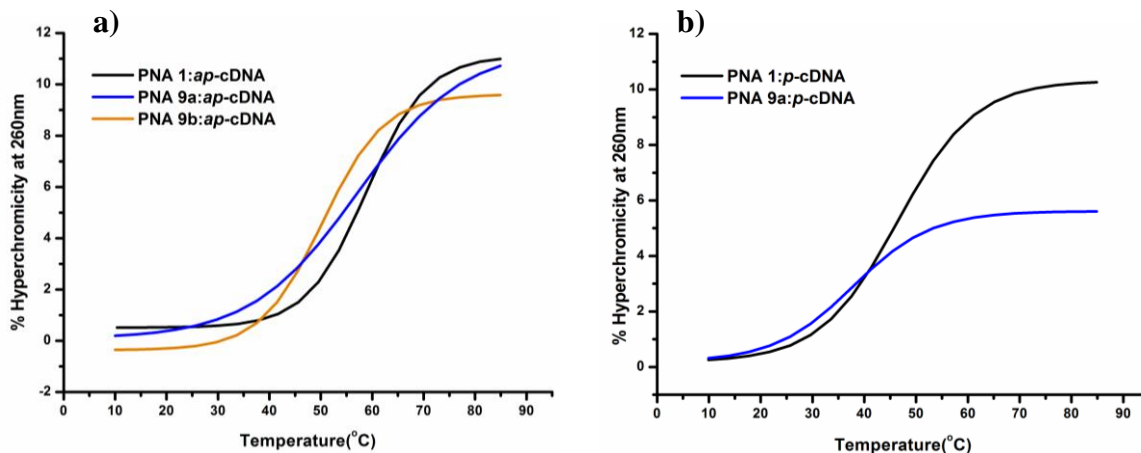
Figure 3.17: UV melting plots of PNA 1, 8a, 8b with a) *ap*-cDNA **b)** *p*-cDNA at 10mM NaCl concentration
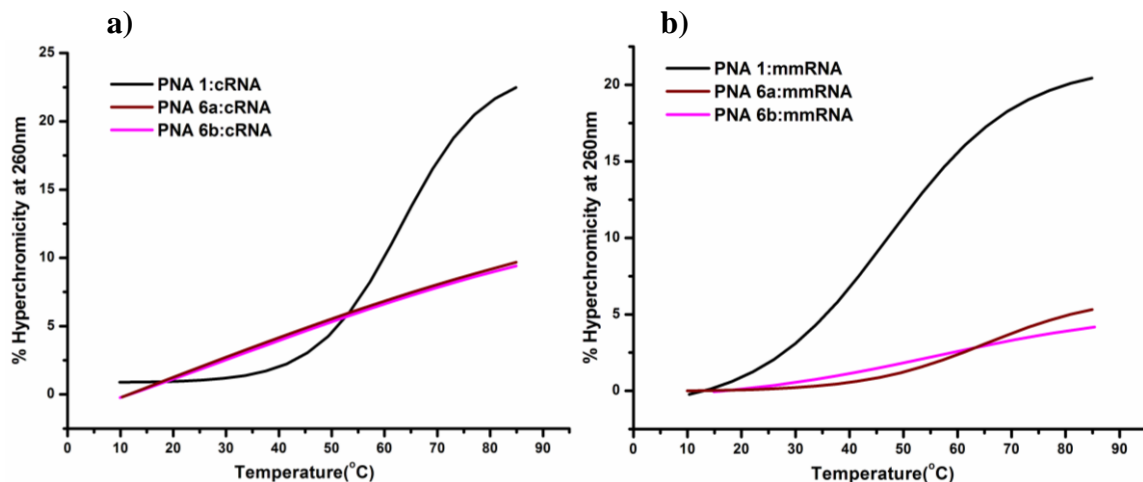
Figure 3.18: UV melting plots of PNA 1, 9a, 9b with a) cRNA b) mmRNA at 10mM NaCl concentration

Table 15: UV meltings at 100mM NaCl concentration of $A^{OMe}A^{OMe}$ccgatttcag-K

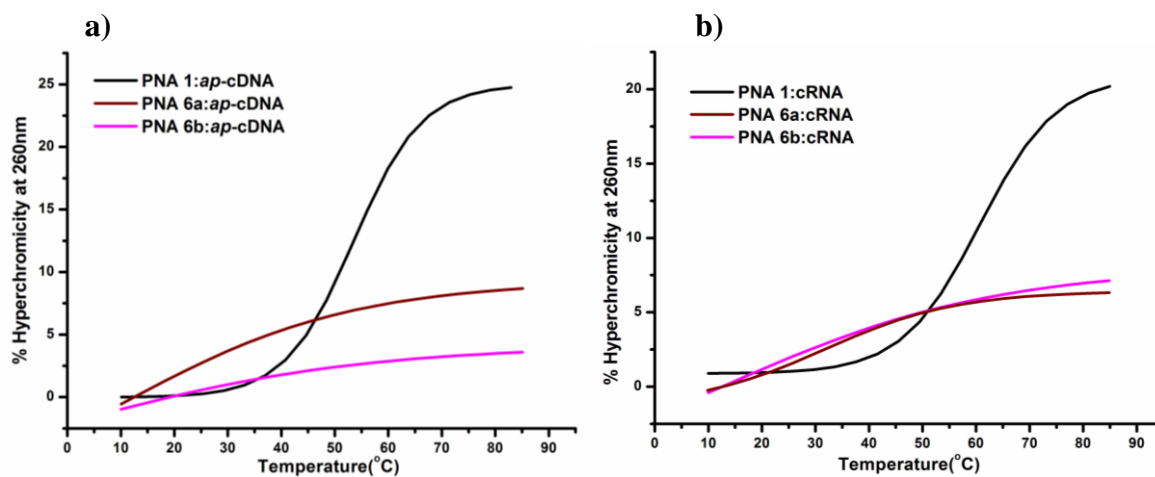| Seq Code	Sequences	*ap*-cDNA	cRNA
PNA 1	aaccgatttcag-K	53.7	60.9
PNA 9a	$^{(R, R)}A^{OMe}{}^{(R, R)}A^{OMe}$ccgatttcag-K	52.1	61
PNA 9b	$^{(S, S)}A^{OMe}{}^{(S, S)}A^{OMe}$ccgatttcag-K	47	50.4
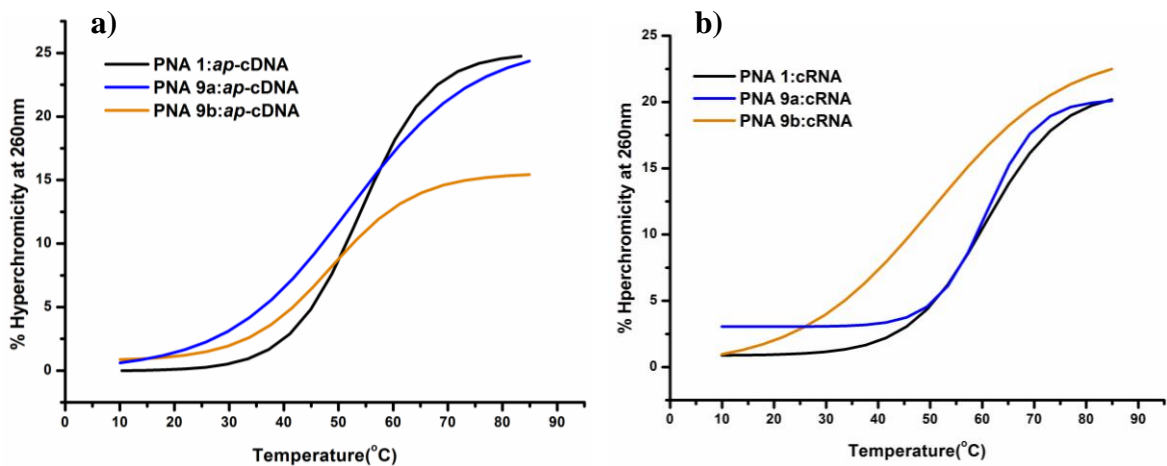
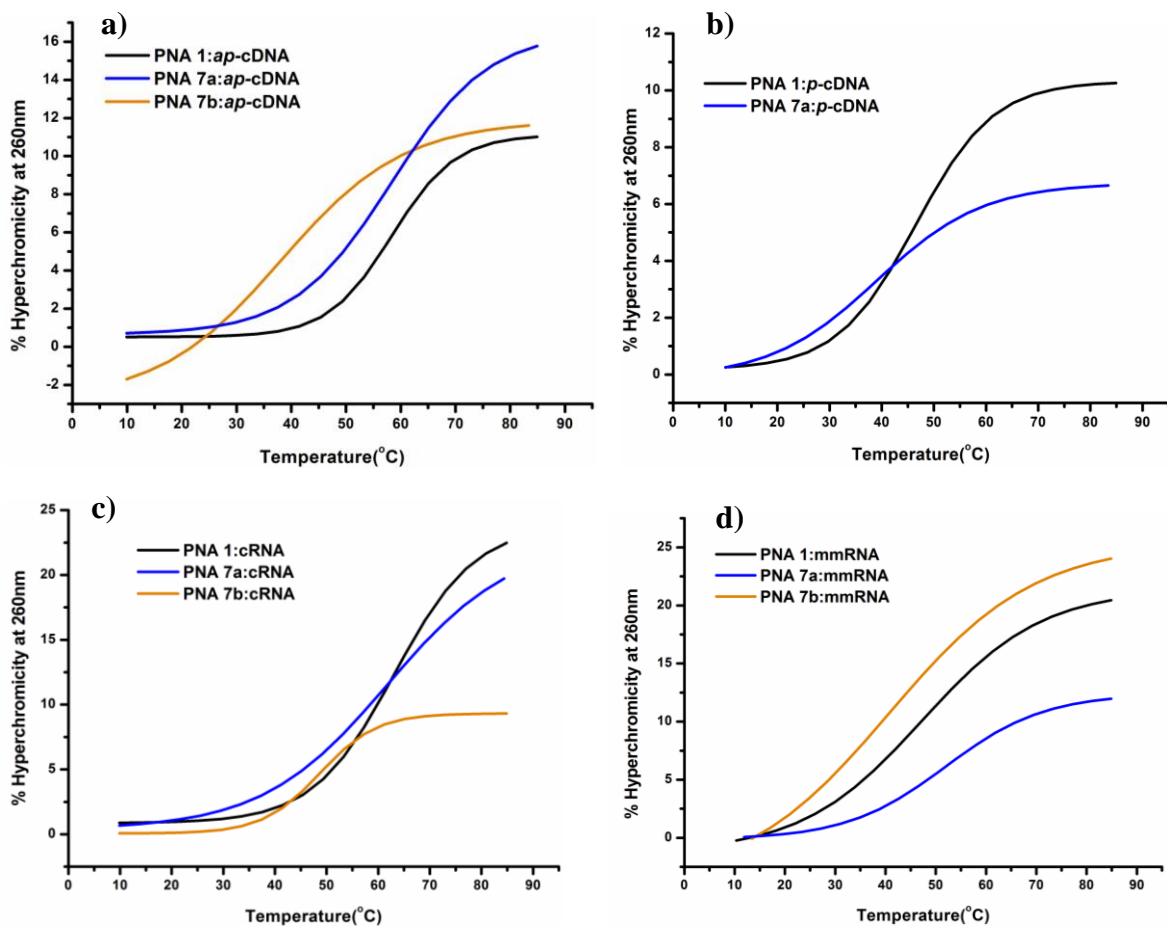**Figure 3.19: UV melting plots of PNA 1, 9a, 9b with a) *ap*-cDNA b) cRNA at 100mM NaCl concentration**

Table 16: UV meltings at 10mM NaCl concentration of lysine attached PNA oligomers

Seq Code	Sequences	*ap*-cDNA	*p*-cDNA	cRNA	mmRNA
PNA 2	K_4-aaccgatttcag-K	61.5	49.5	66.6	53.6
PNA 10a	K_4-accga$^{(R, R)}T^{OMe}$t$^{(R, R)}T^{OMe}$cag-K	61.5	n.t.	68.5	53.7
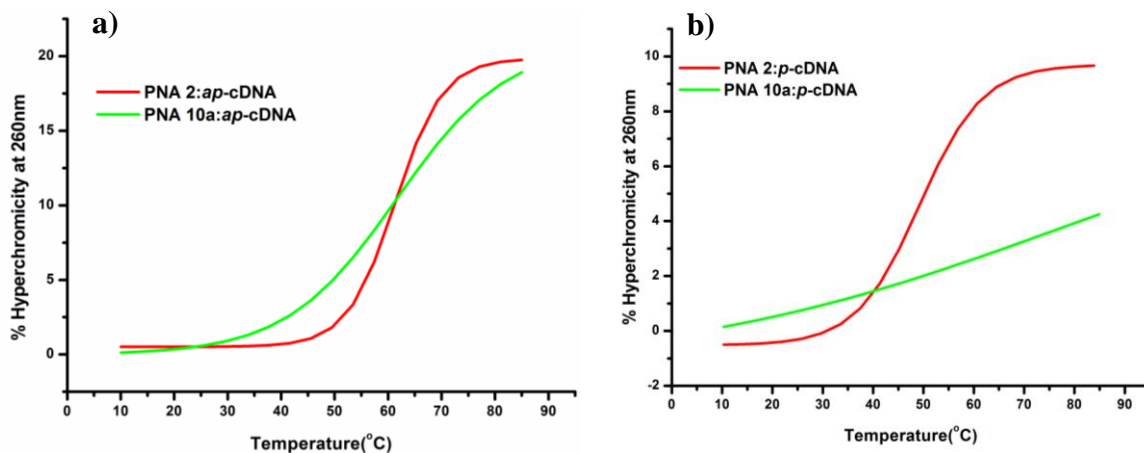
Figure 3.20: UV melting plots of PNA 1, 10a with a) *ap*-cDNA b) *p*-cDNA at 10mM NaCl concentration
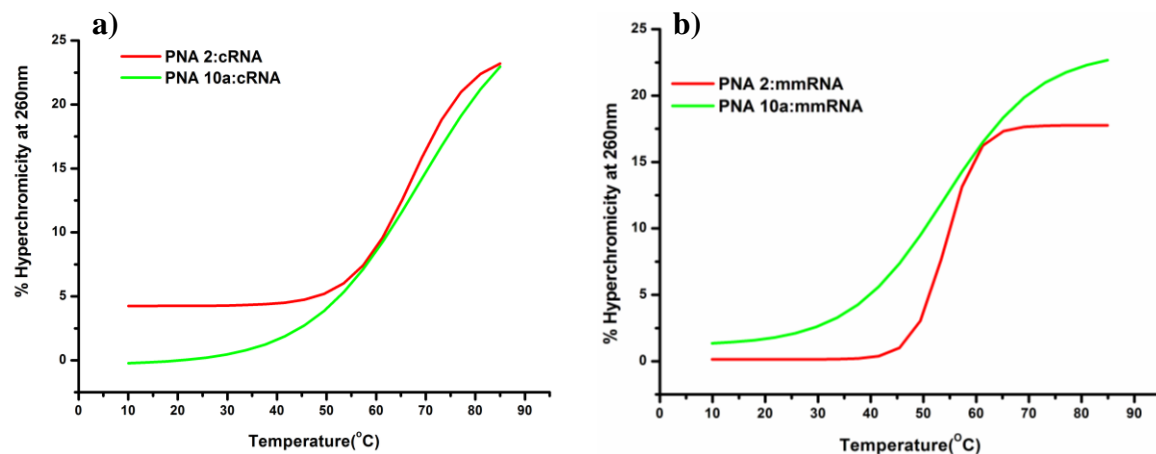
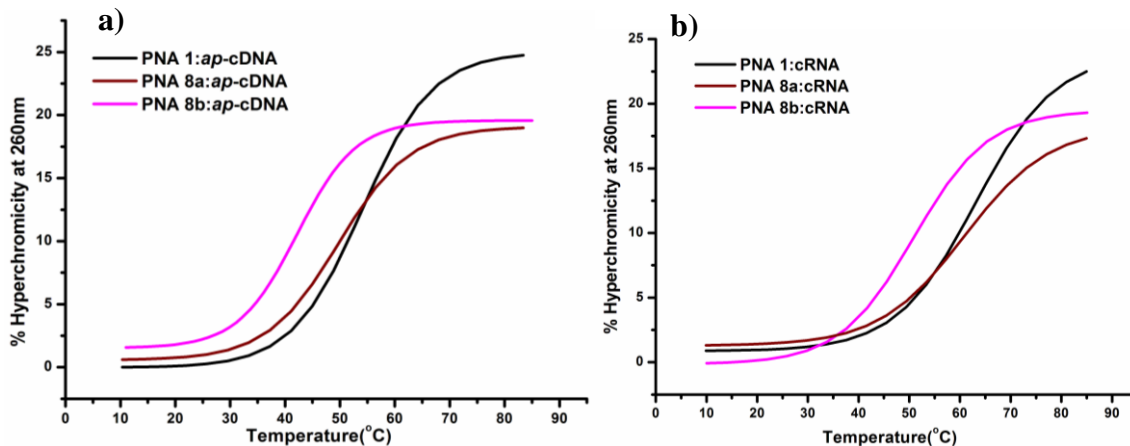**Figure 3.21: UV melting plots of PNA 1, 10a with a) cRNA b) mmRNA at 10mM NaCl concentration**

Table 17: UV meltings at 100mM NaCl concentration of lysine attached PNA oligomers

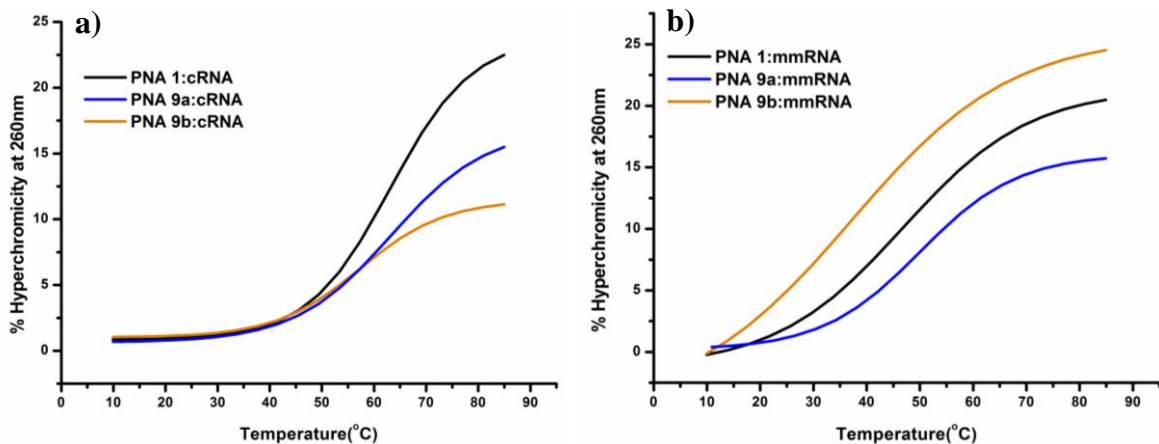| Seq Code	Sequences	*ap*-cDNA	cRNA
PNA 2	K$_4$-aaccgatttcag-K	58.5	62.7
PNA 10a	K$_4$-accga$^{(R, R)}$T^{OMe}t$^{(R, R)}$T^{OMe}cag-K	55.1	63.1
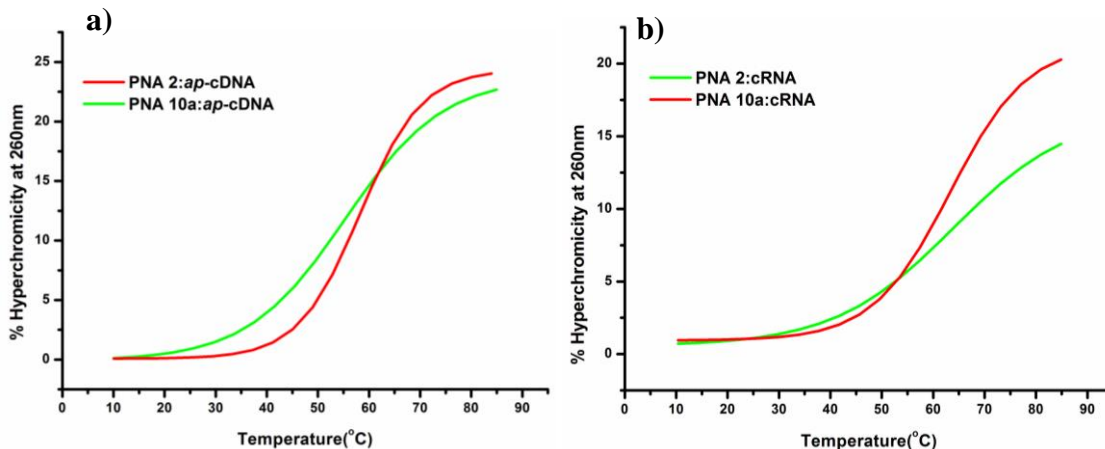
Figure 3.22: UV melting plots of PNA 1, 10a with a) *ap*-cDNA b) cRNA at 100mM NaCl concentration

(β, γ) bis-hydroxymethyl modified oligomers did not form stable duplex with complementary DNA and RNA while (β, γ) bis-methoxymethyl modified oligomers hybridized with complementary DNA and RNA. Complexes of (*R, R*) (β, γ) bis-methoxymethyl modified oligomers with cDNA as well as RNA were found to be more stable in comparison to those with (*S, S*) (β, γ) bis-methoxymethyl modified oligomers. The change in salt concentration showed a slight destabilizing effect of -3 to 4°C uniformly for the modified and unmodified PNA oligomers, when four lysine units were attached at *N*-terminus. The unmodified acyclic flexible *aeg*PNA binds to the *ap*-cDNA and *ap*-cRNA sequences to form stable PNA: DNA and PNA: RNA duplexes. The PNA1 and PNA2 also bind to the *p*-cDNA but with comparatively lower T_m than with the *ap*-cDNA. The replacement of *aeg*PNA-thymine units in the centre by one or two (*R, R*)-T^{OMe} units stabilized the complexes with cRNA by 1.6 °C/modification while (*S, S*)-T^{OMe} units completely destabilized the sequences when present in the internal positions. The (*R, R*)-A^{OMe} units in the sequences formed stable complexes in the centre or at the end of the sequences. The stereochemistry dependent destabilization of the duplexes was less at the *C*-terminal end of the sequence where two adenine *aeg*PNA units were replaced by two (*S, S*) - or (*R, R*)-A^{OMe} units. The PNA sequence with two modified units (*R, R*)-T^{OMe} in internal positions was organized such that unlike *aeg*PNA they did not show any binding with *p*-cDNA. Introduction of four lysine units at the *N*-terminus further enhanced the stability of duplexes with both *ap*-cDNA and cRNA for both PNA 1 and PNA 10a.

It was observed that the modified PNAs with a single $(S, S)/(R, R)$-T^{OH}/A^{OH} modification in the centre of the sequence completely destabilized the duplexes. The sequence in which the modified (R, R)-A^{OH} units were at *N*-terminus could form duplexes with *ap*-cDNA and RNA with slight destabilization. As expected, the (S, S)-A^{OH} units at *N*-terminus of the sequence led to further destabilization. This functional group driven destabilizing effect was further confirmed by electrophoretic gel mobility shift assay.

3B.3 Electrophoretic gel mobility shift assay

Electrophoretic gel experiment was employed to establish that β, γ-bis-hydroxy oligomers do not bind to complementary *ap*-cDNA while β, γ-bis-methoxy substituted oligomers do. **PNA 1, 3a, 4a** were mixed individually with cDNA (350μM) in water.

Lane 1: *ap*-cDNA: **PNA 3a** (aaccga$^{(R, R)}$**T**^{OH}ttcag-K)

Lane 2: *ap*-cDNA: **PNA 4a (**aaccga$^{(R, R)}$**T**^{OMe}ttcag-K)

Lane 3: *ap*-cDNA: **PNA 1** (aaccgatttcag-K)

Lane 4: Single strand *ap*-**cDNA**

Lane 5: Bromophenol Blue

Figure 3.23: Electrophoretic gel mobility shift assay of PNA 1, 3a, 4a

3B.4 Rationalizing the observations

We considered to rationalize this change in PNA-DNA/RNA duplex stability on changing the substitution from β,γ-bis-CH$_2$OMe to β,γ-bis-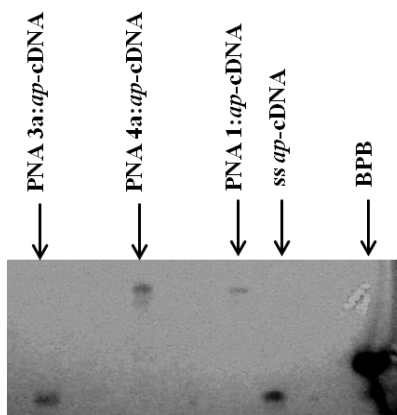CH$_2$OH on the basis of recent work on conformational analysis carried out in 1,4-butanediol system as a prototype.[35]

Theoretical calculations suggested that the preferred lowest energy conformation in the 1,4 butanediol system is largely confined to a lowest energy conformation which can be attributed to the possible linear H--O--H hydrogen bonding in a 7-membered ring.[36] This conformation also holds good in the presence of added water molecules. Such hydrogen bonded conformations containing linear H--O--H hydrogen bonding in a 7-membered ring are shown to be preferred conformations even in aqueous medium for glycerol[37]. Our results in fact give a chemical evidence for the preference of the hydrogen bonded structure for 1, 4-butanediol segment of our modified β,γ-bis-CH$_2$OH-substituted PNA, in solution, considering the known facts about the geometries in PNA: RNA structure. The previous structural analysis, based on NMR and X-ray crystallography studies, pointed out that the preferred dihedral angle *β*, in the ethylenediamine segment (N –CH$_2$ –CH$_2$ –N , **Figure 3.24**) would be close to 60-70° in PNA:RNA duplex and the same was ~140° in PNA:DNA duplex. The designed (*R, S*)-*cis*-cyclohexyl[27,28,29] based PNA, in which, the dihedral angle *β* - was locked at 60-70°, was highly successful in stabilizing the PNA:RNA duplexes. The *trans/cis* cyclopentyl PNA[25, 26, 27] however could bind to both DNA/RNA successfully due to the relatively flexible cyclopentyl ring system in which the dihedral angle in ethylenediamine segment could span the required range more easily. In *trans*-cyclohexyl system due to closed rigid ring, the dihedral angle *β* could be in the range to form the duplex in at least one of the chair conformations. The (*S, S*)-*trans*-cyclohexyl[24] unit caused marginal duplex destabilization effect whereas the (*R, R*)-*trans*-cyclohexyl based PNAs showed large destabilization of complexes with DNA and RNA. The rigid nature of cyclohexyl ring was very effective in binding with RNA in stereoselective manner when the geometry of substitution was *cis* and the dihedral angle *β* was restricted to the preferred value (~ 60°) in either of the chair conformations of cyclohexane in RNA: PNA duplexes. In the present modification, the possible hydrogen bonding interactions in vicinal hydroxymethyl substitution could restrict the conformation of modified bis- hydroxymethyl PNA in which this angle would be far away (180°) than is suitable for PNA: RNA/DNA duplex formation due to free rotation around *C-C* single bond and thus formation of such duplexes is precluded. In the bis-methoxymethyl PNA, in the absence of hydrogen bonding interactions, and free *C-C* bond rotation, a conformation suitable for RNA/DNA binding is maintained (**Figure 3.24**) and the formation of stable duplexes with RNA as well as DNA were observed. We therefore pro-

pose that the difference in binding properties of bis-hydroxymethyl and bis-methoxymethyl substitution could arise from the possible intramolecular hydrogen bonding in adjacent free hydroxyl groups in bis-hydroxymethyl monomers and thus freezing the conformations of bis-hydroxymethyl units that is not suitable for binding with either RNA or DNA (**Figure 3.24**). It is also noteworthy to find that the (*R, R*) stereoisomer is preferred over (*S, S*) isomer while binding to the natural nucleic acids.
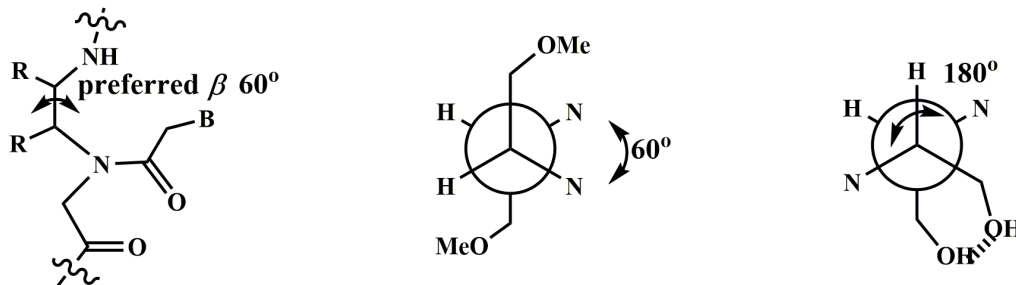
N-CHRCHR-N dihedral angle (*β*) in ethylene diamine segment of modified PNA

Figure 3.24: PNA and Newman projection of β,γ-substituted PNA monomers

3B.5 Summary

- PNA substituted with (*R, R*)-β,γ-bis- methoxymethyl thymine and adenine residues stabilize the duplexes while the (*S, S*)-stereochemistry completely destabilizes the duplexes.

- The chirality of the monomers, substitution on the backbone as well as position of modification affected the binding properties while cross pairing. A simple change in backbone from bis-hydroxmethyl to methoxymethyl in *(R, R)*-dimethoxy PNA drastically changed the binding to DNA and RNA.

3B.6 Experimental

3B.6.1 Preparation of buffer for UV-melting experiment

UV-melting experiments were performed in 10mM sodium phosphate buffer with 10mM or 100mM sodium chloride concentration as mentioned. All buffers were prepared in milliQ-water and the pH of the buffer maintained to 7.2.

3B.6.2 Preparation of sample for polyacrylamide gel electrophoresis

The samples were lyophilized and resuspended in 2μL sodium phosphate buffer (10mM, pH 7.2). The samples were annealed by heating to 90°C followed by slow cooling to room temperature and refrigeration at 4°C for 6h. To this 2μL of 40% sucrose solution in TBE buffer (pH 8.0) was added and loaded on the gel. Bromophenol Blue (BPB) was used as the tracer dye separately in adjacent well. Gel electrophoresis was performed on a 15% non-denaturing polyacrylamide gel (acrylamide: bis-acrylamide, 29:1) at a constant power supply of 150V, until the BPB migrated to three-fourth of the gel length. During electro-phoresis the temperature was maintained at 10°C. The spots were visualized through UV shadowing by illuminating the gel placed on a pre-coated silica gel plate (F_{254}), using UV-light.

Section C

Cellular uptake studies

3C.1 Carboxyfluorescein tagging of PNA oligomers

Carboxyfluorescein was attached to the synthesized PNA oligomers **PNA 1, 2, 5a, 10a** for studying the internalization of the oligomers into cells. To synthesize carboxyfluo-rescein attached PNA oligomers, couplings were carried out using **PNA 1, 2, 5a, 10a** on MBHA resin in presence of ten equivalents of 5(6)-carboxyfluorescein, HOBt, DIPCDI (Diisopropyl carbodiimide) in DMF overnight. The oligomers were cleaved from solid phase in presence of TFA, TFMSA employing the regular protocol. The crude peptide was purified by semipreparative C18 column.

Table 18: HPLC purification of Carboxyfluorescein tagged PNA 1, 2, 5a, 10a and their MALDI-TOF analysis

Seq code	Sequences	R.t.	Calc. mass	Obsvd. mass
PNA 1-CF	CF-aaccgatttcag-K	14.6	3737.4	3736.0
PNA 2-CF	CF-K$_4$-aaccgatttcag-K	14.9	4249.8	4248.4
PNA 5a-CF	CF-aaccga $\mathbf{T}^{\mathbf{OMe}(R,R)}$t $\mathbf{T}^{\mathbf{OMe}(R,R)}$ cag-K	15.2	3915.7	3918.3
PNA 10a-CF	CF-K$_4$-aaccga $\mathbf{T}^{\mathbf{OMe}(R,R)}$t $\mathbf{T}^{\mathbf{OMe}(R,R)}$cag-K	14.7	4428.4	4431.95

3C.2 Cell culture and cellular uptake using Flow Cytometry

HCT-116 cells were treated with CF-labeled PNAs at 1μM concentration and incubated for 10h at 37°C. In order to remove the cell surface bound PNAs, the cells were washed with heparin (1mg/ml) before flow cytometric analysis. Thus, the fluorescent positive cells that were obtained after FACS analysis were indicative of internalized PNAs and not associated with the cell membrane.

To see the internalisation of carboxyfluorescein-labeled PNA conjugates, intracellular fluorescence was determined by flow cytometry analysis using a FACSAria flow cytometer (Becton Dickinson). Experiments were performed in triplicate with acquisition of total of 1000 events in each sample was done. FITC (530/30nm) band pass filter was used for fluorescence analysis of the cells.
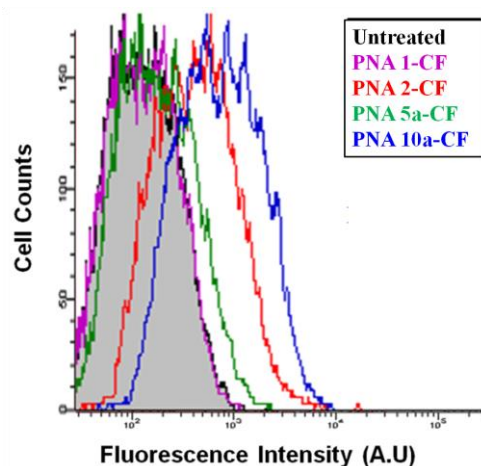
Figure 3.25: FACS analysis of the PNA 1, 2, 5a, 10a oligomers with carboxyfluorescein (CF) at 37°C
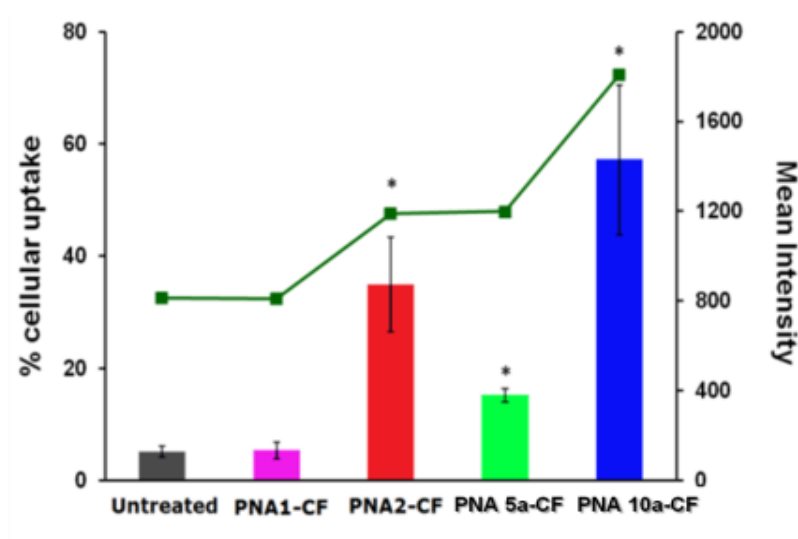
Figure 3.26: Quantitative cellular uptake of PNAs in absence of transfection agent. Flow cytometric analysis of CF-conjugated PNAs in HCT-116 cell line. The graphs depicts % cellular uptake (left axis) and mean fluorescence intensity (right axis) of each PNA after internalization. PNA10a-CF having four lysine residues and modified monomers showed most efficient cellular uptake as compared to other PNAs. Experiments were conducted in triplicate (n= 3). Error bars represent ± SD with * p-value < 0.01 (Student's *t* test).

In order to examine the ability of PNA [PNA (1, 2, 5a, 10a)-CF] to enter the cells without the help of transfection agent, we performed flow cytometry. The untreated cells (~5.5%) and unmodified PNA (PNA1-CF) showed ~5.4% fluorescent positive cells. The presence of bis-hydroxymethyl units in the sequence slightly improved the uptake to ~15%. Lysine residues are known to improve the cellular uptake of PNA. The unmodified PNA 2-CF with four lysine residues improved internalization upto ~35 %.The bis-hydroxymethyl modified PNA containing 4-lysine units (PNA10a-CF) was most efficiently internalized without the aid of transfection agent and ~57% fluorescent positive cells were observed.

3C.3 Conclusion

⸶ The results presented here reveal for the first time an acyclic β, γ substituted PNA and the specificity of stereochemical requirements of the modified PNA while binding to target DNA and RNA. PNA substituted with (*R, R*) β, γ-bis-methoxymethyl thymine and adenine residues stabilizes the duplexes while the (*S, S*) - stereochemistry completely destabilizes the duplexes.

⸹ The change from bis-methoxymethyl to bis-hydroxymethyl substitution is unsuitable for duplex formation in the present stereochemistry. Furthermore, the PNAs with (*R, R*)- β, γ-bis-methoxymethyl substituted thymine significantly improved the access for PNA to intracellular space.

3C.4 Experimental

3C.4.1 Appendix

HPLC chromatogram of purified **PNA 1-CF**

HPLC chromatogram of purified **PNA 2-CF**
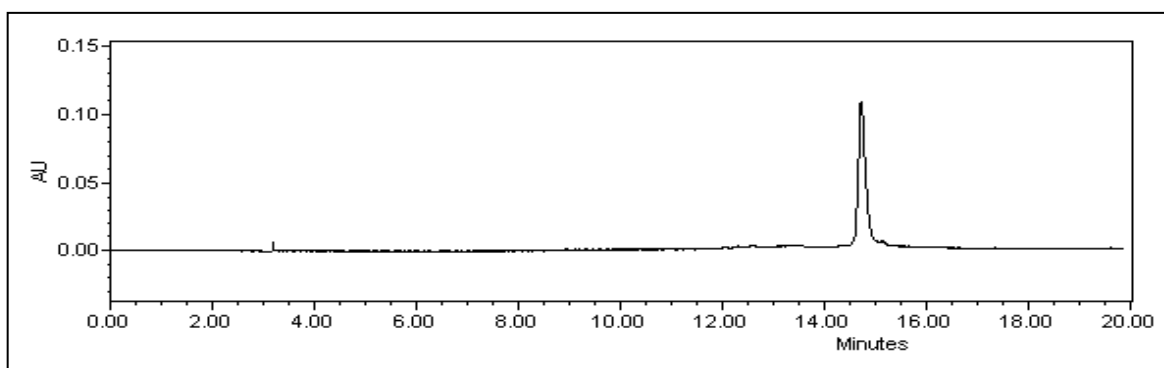
HPLC chromatogram of purified **PNA 5a-CF**

HPLC chromatogram of purified **PNA 10a-CF**

PNA 5a-CF
Expected MW: 3915.7
Observed MW: 3918.3
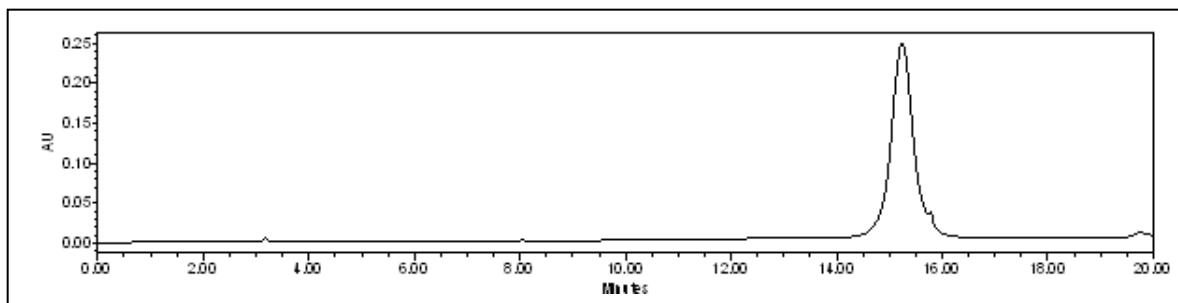
PNA 10a-CF
Expected MW: 4428.4
Observed MW: 4431.95
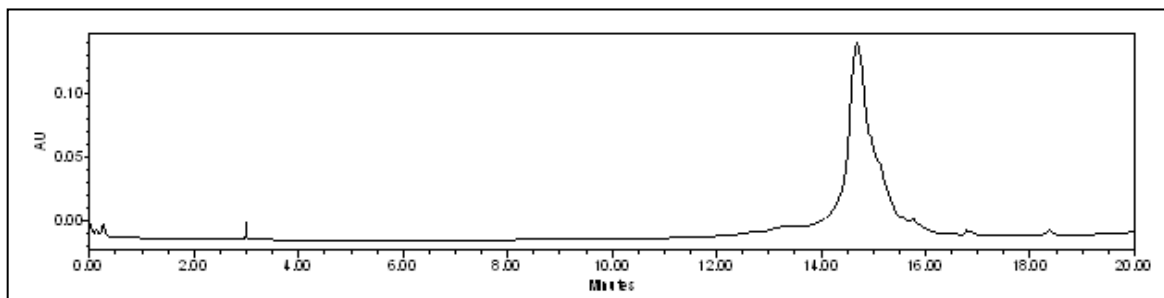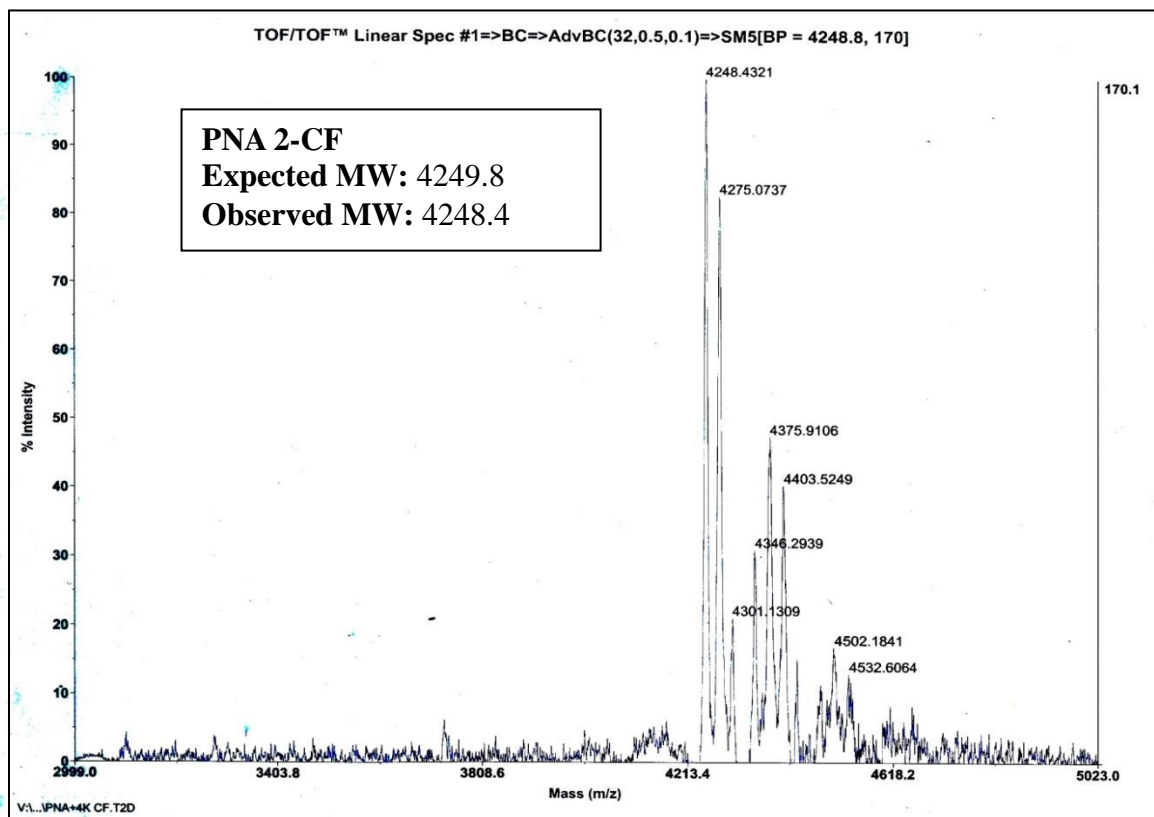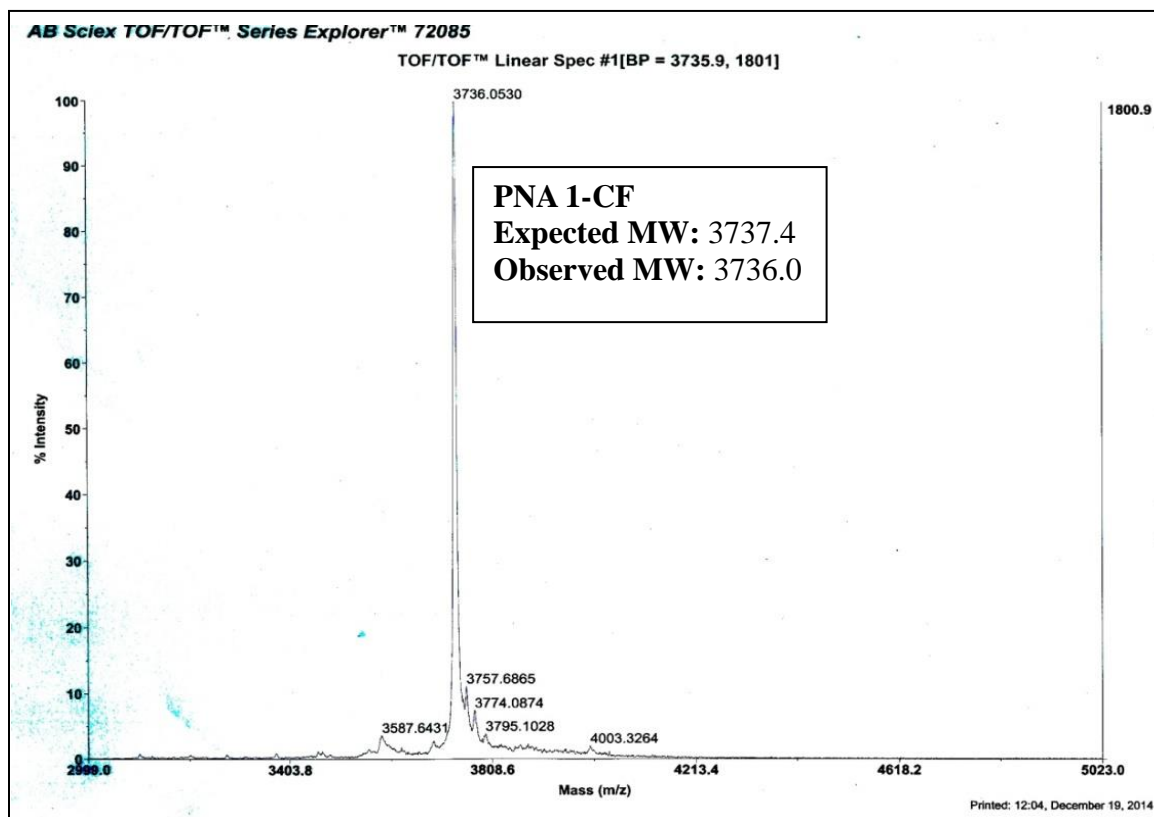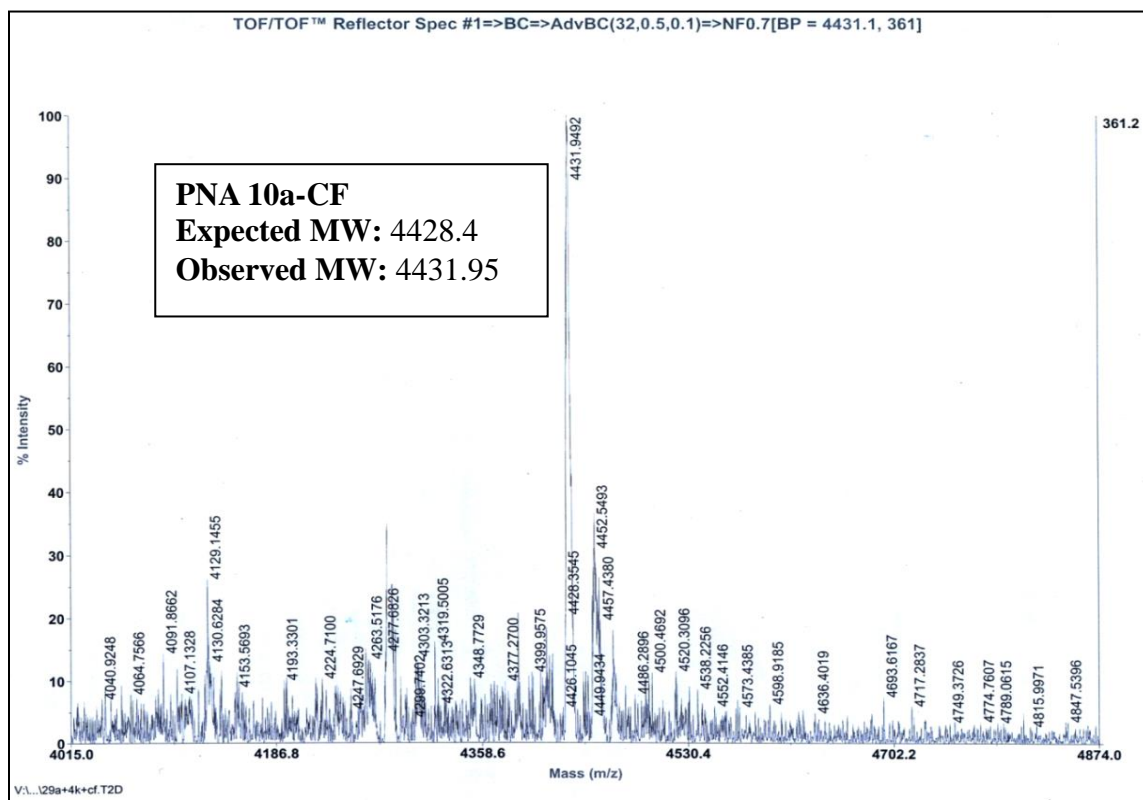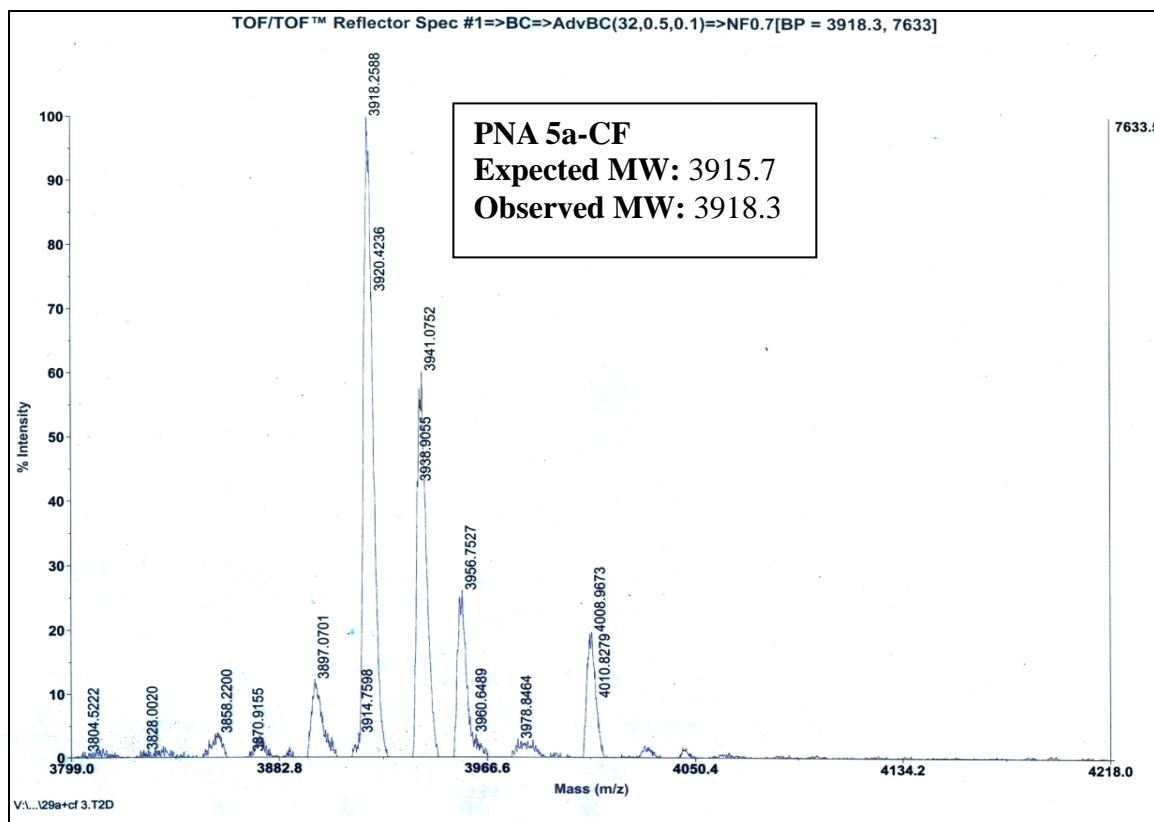
3C.4.2 Experimental procedure and sample preparation for flow cytometry

HCT-116 (human colorectal carcinoma) were cultured in Dulbecco's modified Eagle's medium (DMEM, GlutaMAX, 4.5 g/L D-glucose, pyruvate ; Life technologies) supplemented with 10% Fetal Bovine serum (FBS) at 37°C in 5% CO_2 atmosphere. HCT-116 cells were seeded in 12- well plates at 6×10^4 cells per well and grown overnight until 70% confluency. Next day, media was removed and cells were washed with 1X PBS. To each well, PNA (PNA1, 2, 5a, 10a) was added in serum free media (OptiMEM, Invitrogen) at a final concentration of 1μM and incubated for 10h at 37°C/5% CO_2. Untreated cells were used as a negative control. After incubation, cells were washed with 1xPBS supplemented with Heparin (1mg/ml) to remove the cell surface bound PNA. Next, the cells were washed with 1:1 mixture of Trypan Blue and 1xPBS to further remove the extracellular fluorescent artifacts. Trypsin (0.25%) was then added, and the cells were harvested in complete media (DMEM, Life Technologies) after 5 min. The cells were centrifuged down, washed with 1xPBS twice and resuspended in 1xPBS.

3C.5 References

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

2.   B. Hyrup and P. E. Nielsen, *Bioorg. Med. Chem.*, 1996, **4**, 5.

3.   M. Egholm, O. Buchardt, L. Christensen, C. Behrens, S. M. Freier, D.A. Driver, R. H. Berg, S. K. Kim, B. Norden and P. E. Nielsen, *Nature*, **1993**, *365*, 566.

4.   R. Kole, A. R. Krainer and S. Altman, *Nat. Rev. Drug Disc.*, 2012, **11**, 125

5.   V. A. Kumar and K. N. Ganesh, *Acc. Chem. Res*. 2005, 38, 404.

6.   J. K. Watts, G. F. Deleavey and M. J. Damha, *Drug Discovery Today*, 2008, **13**, 842.

7.   J. C. Hanvey, N. J. Peffer, J. E. Bisi, S. A. Thomson, R. Cadilla, J. A. Josey, D. J. Ricca, C. F. Hassman, M. A. Bonham, K. G. Au, S. G. Carter, D. A. Bruckenstein, A. L. Boyd, S. A. Noble and L. E. Babiss, *Science*, 1992, **258**, 1481.

8.   P. E. Nielsen, *Q. Rev. Biophys.*, 2005, **38**, 345.

9.   M. Masuko, *Nucleic Acid Res. Suppl.,* 2003, **1**, 145.

10. A.Cattani-Scholz, D. Pedone, F. Blobner, G. Abstreiter, J. Schwartz, M. Tornow, L. Andruzzi, *Biomacromolecules,* 2009, **10**, 489.

11. G. Haaima, A. Lohse, O. Buchardt and P. E. Nielsen, *Angew. Chem. Int. Ed.,* 1996, **35**, 1939

12. A. Puschl, S. Sforza, G. Haaima, O. Dahl and P. E. Nielsen, *Tetrahedron Lett.,* 1998, **39**, 4707

13. P. Gupta, O. Muse and E. Rozners, *Biochemistry*, 2012, **51**, 63.

14. P. Zhou, M. Wang, L. Du, G.W. Fisher, A. Waggoner and D. H. Ly, *J. Am. Chem. Soc.*, 2003, **125**, 6878.

15. A. Dragulescu-Andrasi, P. Zhou, G. He and D. H. Ly, *Chem. Commun.*, 2005, **41**, 244.

16. M. C. De Koning, L. Petersen, J. J. Weterings, M. Overhand, G. A. van der Marel and D.V. Filippov, *Tetrahedron*, 2006, **62**, 3248.

17. B. Sahu, V. Chenna, K. L. Lathrop, S. M. Thomas, G. Zon, K. J. Livak and D. H. Ly, *J. Org. Chem.*, 2009, **74**, 1509.

18. A. Manizardi, A. Calabretta, M. Bencivenni, T. Tedeschi, S. Sforza, R. Corradini and R. Marchelli, *Chirality*, 2010, **22**, E161.

19. T. Tedeschi, S. Sforza, R. Corradini, R. Marchelli, *Tetrahedron Lett.*, 2005, **46**, 8395.

20. S. Sforza, T. Tedeschi, R. Corradini, R. Marchelli, Eur.J.Org.Chem., 2007, **2007**, 5879.

21. R. Mitra and K. N. Ganesh, *Chem. Commun.*, 2011, **47**, 1198.

22. R. Mitra and K.N. Ganesh, *J. Org.Chem.*, 2012, **77**, 5696.

23. T. Sugiyama, Y. Imamura, Y. Demizu, M. Kurihara, M. Takano and A. Kittaka, *Bioorg. Med.Chem. Lett.*, 2011, **21**, 7317.

24. P. Lagriffoule, P. Wittung, M. Ericksson, D. K. Jensen, B. Norden, O. Buchardt and P. E. Nielsen, *Chem. Eur. J.*, 1997, **3**, 912.

25. M. C. Myers, M. A. Witschi, N. V. Larionova, J. M. Frank, R. D. Haynes, T. Hara, A. Grajkowski and D. H. Appella, *Org. Lett.*, 2003, **5**, 2695.

26. J. K. Pokorski, M. A. Witschi, B. L. Purnell and D. H. Appella, *J. Am. Chem. Soc.* 2004, **126**, 15067.

27. T. Govindaraju, V. A. Kumar and K. N. Ganesh, *Chem. Commun.*, 2004,860.

28. T. Govindaraju, V. A. Kumar and K. N. Ganesh, *J. Am. Chem. Soc.*, 2005, **127**, 4144.

29. T. Govindaraju, R. G. Gonnade, M. M. Bhadbhade, V. A. Kumar and K. N. Ganesh, *Organic Letters*, 2003, **17**, 3013.

30. T. Govindaraju, V. A. Kumar and K. N. Ganesh, *J. Org. Chem.,* 2004, **69**, 1858.

31. A. Scheurer, P. Mosset and R. W. Saalfrank, *Tetrahedron: Asymmetry*, 1999, **10**, 3559.

32. A. Scheurer, P. Mosset and R. W. Saalfrank, *Tetrahedron: Asymmetry*, 1997, **8**, 1243, 3161.

33. A. Dragulescu-Andrasi , S. Rapireddy , B. M. Frezza , C. Gayathri , R. R. Gil and D. H. Ly, *J. Am. Chem. Soc.*, 2006, **128**, 10258.

34. B. Sahu, I. Sacui, S. Rapireddy, K. J. Zanotti, R. Bahal, B. A. Armitage and Danith H. Ly, *J. Org. Chem.*, 2011, **76**, 5614.

35. L. Evangelisti, Q. Gou, L. Spada, G. Feng and W. Caminati, *Chem. Phys. Lett.*, 2013, **556**, 55.

36. S. Kozuch, S. M. Bachrach, and J. M.L. Martin, *J. Phys. Chem. A*, 2014, **118**, 293.

37. S. M. Bachrach, *J. Phys. Chem. A*, 2014, **118**, 1123.

38. J. Amato, M. I. Stellato, E. Pizzo, L. Petraccone, G. Oliviero, N. Borbone, G. Piccialli, A. Orecchia, B. Bellei, D. Castiglia, C. Giancola, *Mol. BioSyst.*, 2013, **9**, 3166.

39. S. Abes, J. J. Turner, G. D. Ivanova, D.Owen, D. Williams, A. Arzumanov, P. Clair, M. J. Gait and B. Lebleu, *Nucleic Acids Res.,* 2007, **35**, 4495.

40. M. Lundberg, S. Wikström and M. Johansson, *Molecular Therapy,* 2003 **8**, 143.

Chapter 4

Structural and functional evaluation of PNA and modified PNA as G-quadruplex

4.1 Introduction to G-quadruplex

G-quadruplexes are a distinct class of highly ordered nucleic acid structures comprising tandem repeats of guanine. They have become a common feature in naturally occurring nucleic acids (DNA and RNA) and perform regulatory functions in many biological processes.[1] G-quadruplex structures are stable and detectable in human genomic DNA.[2] This regular occurrence of quadruplex structures in the genome stimulates the designing of drug molecules based on DNA/RNA G-quartets (**Figure 4.1**).[3] Aptamers are chemically synthesized oligonucleotides with abilities to bind to small molecules, peptides, proteins with selectivity, specificity and affinity equal or superior to those of antibodies.[4] Thrombin binding aptamer (TBA), a 15-mer DNA sequence, 5'-GGTTGGTGTGGTTGG-3' capable of forming G-quadruplex, was discovered in 1992 by *in vitro* selection.[5] This aptamer was found to inhibit fibrin-clot formation by binding to the thrombin protein with high selectivity and affinity.
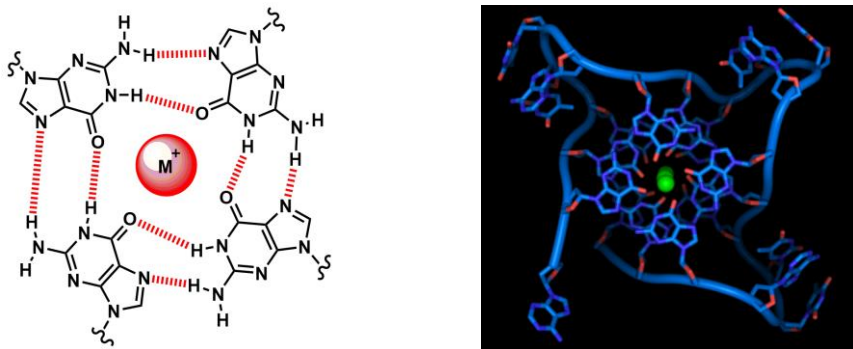
Figure 4.1: The G-tetrad motif

4.1.1 Structure and topology of G-quadruplexes

G- quadruplexes can be formed from one, two or four different strands of DNA or RNA. Their topologies depend on strand polarity, loop length and sequence of bases. A quadruplex forming sequence can exhibit equilibrium between several different G-tetrad conformations. However, the biological function and molecular recognition properties may well be associated with one predominant conformation. G-quadruplexes can be designated parallel, antiparallel and mixed type according to strand polarity. They are termed antiparal-

lel when at least two of the four strands are antiparallel to each other; and parallel only when all the strands are parallel to each other.

Different types of G-quadruplexes are also formed depending on loop sequence and on loop length (**Figure 4.2**). Both loop length and sequence play a role in determining the folded conformation and thermodynamic stability of a quadruplex-forming sequence.[6] Loop length determines the topology of unimolecular as well as multimolecular quadruplex. Few of such commonly found G-quadruplex structures are depicted below.
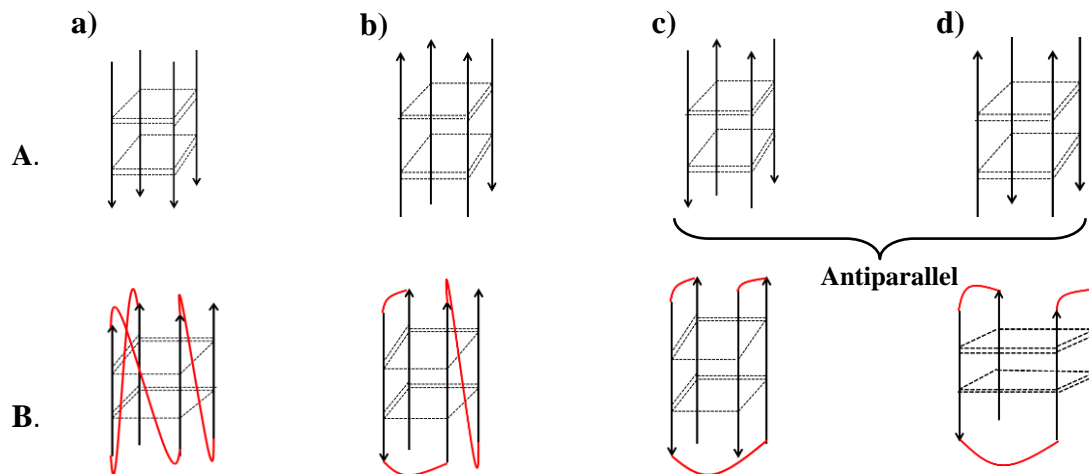
Figure 4.2: (A) Different arrangements of strand polarity, arrows show 5'→3' polarity: (a) all strands parallel, (b) three parallel and one antiparallel, (c) two pairs of adjacent parallel strands, and (d) alternating antiparallel strands. (B) Various loop topologies leading to different arrangements of strand polarity.

4.1.2 Dependence of quadruplex stability on metal ions

The central core of the G-quartet produces a specific geometric arrangement of lone pairs of electrons from the four G-$O6$, which can coordinate a monovalent cation of the correct size.[7] Potassium and sodium were found to be optimal in increasing the G-quadruplex stability. The smaller Na^+ ion can occupy the plane formed by these atoms, whereas the larger K^+ requires a non-planar component, which may in fact lie between two such G-quartets. This allows additional coordination of the metal ions, i.e. to satisfy the usual hexacoordinate stereochemistry of the alkali metal ions. In order to accommodate this stereochemistry, the individual nucleobases may protrude out of the plane,[8] such that they are balanced by the stacking energies. Other monovalent cations such as Li^+, Cs^+, etc., and also the divalent cations such as Ba^{++}, Ca^{++}, Sr^{++} [9] etc are also known to stabilize G-quartet

structures. In addition, non-metal cations such as NH_4^+ [10] have also been shown to stablize G-quadruplexes. Changes in the stabilizing cations have also been exploited as conformation[11] and redox switches[12] in literature.

4.1.3 Molecular crowding conditions

Interior of all the cells contain high concentration of macromolecules. Such environment are therefore, termed 'crowded,' because, in general, single macromolecular species do not occur at high concentration. Collectively, the macromolecules occupy a significant fraction (typically 20-30%) of the total volume. This fraction of volume is thus, physically unavailable to other molecules (**Figure 4.3**).
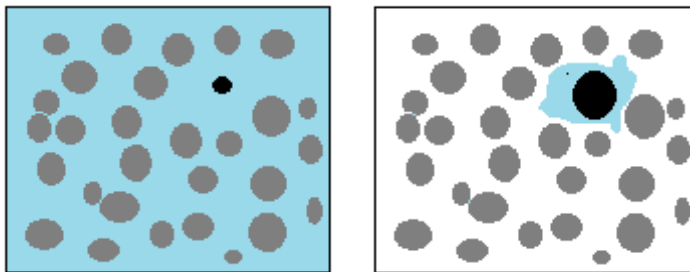
Figure 4.3: Illustration of the volume of solvent available (blue) for two molecules of different sizes (black spheres) in a compartment containing a high concentration of macromolecules (grey spheres). This reduction in available volume increases the effective concentration of macromolecules in concentrated solutions - an effect known as macromolecular crowding

The numbers, sizes and shapes of all the molecules present in each compartment regulate the intracellular volume unavailable to other macromolecules.[13] Crowding is not confined to cellular interiors, but also occurs in the extracellular matrix of tissues such as cartilage and blood plasma.

Crowding explains the existence of molecular chaperones despite the ability of many proteins to assemble. Crowding should favor aggregation because of its effect of increasing the thermodynamic activity of partly folded polypeptide chains and also increase the functional activity of chaperones by stimulating their association with partly folded chains.

In the context of G-quadruplexes, molecular crowding was found to induce human telomere G-quadruplex formation under cation-deficient conditions.[14] Similarly, TBA was

shown to fold into a quadruplex in the absence of stabilizing cations, but in the presence of molecular crowding agents or thrombin.[15] The activity of most G-quadruplex stabilizing ligands has been shown to be noticeably reduced under molecular crowding conditions that mimic the physiological milieu.[16] Recently, however, a G-quadruplex stabilizing ligand was shown to enhance the stability of G-quadruplexes, resulting in an increase in anti-telomerase activity in the presence of molecular crowding conditions.[17]

Commonly used synthetic crowding agents include Ficolls, dextrans, polyethylene glycol and polyvinyl alcohol.

4.1.4 Aptamers or Decoy oligonucleotides: An emerging class of therapeutics

Numerous nucleic acid ligands also termed decoys or aptamers have been developed during the past 25 years that can inhibit the activity of many pathogenic proteins. Several properties of aptamers make them an attractive class of therapeutic compounds. Their affinity and specificity for a given protein makes it possible to isolate a ligand to virtually any target and further adjusting their bioavailability expands their clinical utility. Development of aptamers that retain activity in multiple organisms facilitates preclinical development. Thus they may prove useful in the treatment of a variety of human maladies, including infectious diseases, cancer, and cardiovascular disease.[18]

Aptamers[19] are usually synthetic oligonucleotides that are specifically selected for binding to a certain target, which may range from small molecules to peptides, proteins, or even whole cells. However, natural aptamers also exist in riboswitches. Their binding affinity and specificity may be equal or even superior to that of antibodies (The reported dissociation constants are in the picomolar to micromolar range), and aptamers are known to be able to differentiate and discriminate between even small structural differences that may exist as a result of the presence or absence of a small functional group such as methyl or hydroxyl or even as a result of differing chirality. The word 'aptamer' was first coined by Ellington and Szostak[20] in 1990. Aptamers were first developed by an *in vitro* selection technique termed SELEX (*S*ystematic *E*volution of *L*igands by *EX*ponential enrichment) independently in the laboratories of Joyce,[21] Szostak[20] and Gold[22] in 1990 (**Figure 4.4**).
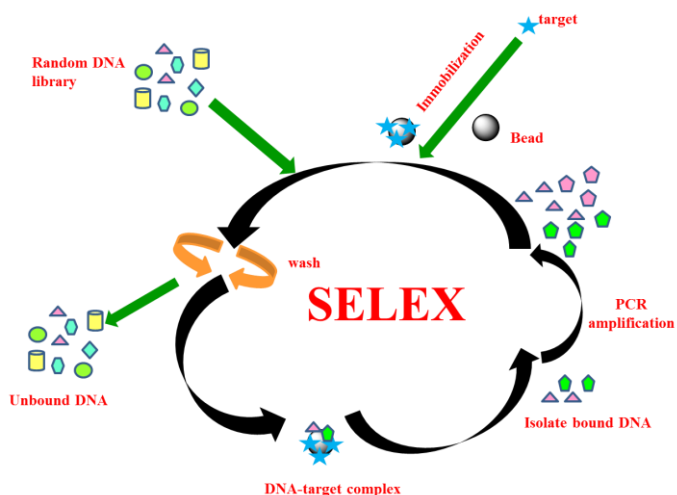
Figure 4.4: Isolation of aptamers using SELEX

4.1.5 Spiegelmers

The susceptibility of the DNA/RNA backbone to nuclease attack spurred a search for modifications that could resist nuclease degradation, but would still be amenable to a suitable selection procedure. 'Spiegelmers', first developed by Fuerste and co-workers[23] are based on the enantiomeric L-ribose sugars in nucleotides, the name being derived from the German 'spiegel' meaning 'mirror'. They are highly resistant to nuclease degradation and form a relatively new class of potential drugs, some being currently tested in clinical trials. The first step involves the synthesis of the enantiomeric form of the target. For example, in peptides, this would involve the synthesis using D-aminoacids. This is followed by the selection of a natural D-ribose-based RNA against the unnatural enantiomeric form of the target by the SELEX procedure. In the next step, the mirror-image RNA (based on the enantiomeric L-ribose sugars), termed the Spiegelmer, is chemically synthesized, which binds specifically, the natural target. Very recently, the discovery of DNA- and DNA/RNA-mixed Spiegelmers that bind glucagon with a K_d of 3nM were reported,[24] as potential therapeutic candidates to attenuate hyperglycemia in type 1 and type 2 diabetes. However the drawdack of spiegelmers is that the production of a speigelmer involves selection of a preliminary oligonucleotide against the synthetic enantiomer of the chosen target, which is not always feasible, and thus the route of chemical modification of oligonucleotide library used in SELEX is more universally applicable.

4.1.6 Thrombin-binding aptamer-TBA: Discovery, structural features, function, modifications with effect on structure and function

4.1.6.1 Anticoagulation Aptamer Target: Thrombin

Thrombin is a key regulatory enzyme in the coagulation cascade. It is a serine protease produced from prothrombin by the action of factor Xa. Thrombin, in turn, converts fibrinogen into fibrin, which is the building block of the fibrin matrix of blood clots.[25]

The thrombin binding aptamer (TBA) unravelled by SELEX method, inhibited fibrin- clot formation with high selectivity and affinity. This single-stranded DNA aptamer prolonged clotting time from 25s to 169s in purified fibrinogen and from 25s to 43s in human plasma.[5] In cynomolgus monkeys, the prothrombin time (PT) increased by 1.7-fold 10min after infusion of the aptamer and returned to baseline 10min after the infusion was terminated.[26] In a canine cardiopulmonary bypass (CPB) model, the aptamer-treated group exhibited increases in PT, activated partial thromboplastin time (aPTT), and activated clotting time that subsequently returned to baseline after aptamer infusion ceased.[27] Currently the only approved anticoagulant for coronary artery bypass graft is heparin. TBA exhibits a K_d of 2nM for thrombin, 50nM for prothrombin and binding to other serum proteins or proteolytic enzymes is essentially undetectable.[15] It could be useful as a heparin alternative in cardiovascular surgery. It has the key advantages that it avoids the risk of thrombocytopenia, associated with heparin, and is a specific inhibitor, effective at inhibiting clot-bound thrombin. No significant toxicities or excessive bleeding intraoperatively has been observed.

The 3D solution structure of TBA was solved by NMR studies[28,29] and showed this sequence [d(GGTTGGTGTGGTTGG)] to be a chair-like intramolecular antiparallel G-quadruplex, with two G-quartets stacked on each other and linked by two TT loops and one TGT loop (**Figure 4.5**).
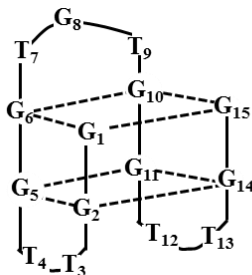
Figure 4.5: Thrombin binding aptamer[31]

X-ray crystallographic studies[30] have shown the importance of the loops in protein recognition and binding. This has prompted a number of studies aimed at improving the anti-thrombin effect of TBA.

4.1.6.2 Modifications in TBA

The era following the discovery of TBA has surfaced various modifications of TBA. Modifications involving replacement of sugar units, phosphate group as well as nucleobase has been reported. This section describes some of the significant modifications of TBA. Single-base substitution studies with unlocked nucleic acids, for example, identified a single position out of 15 investigated, that resulted in enhanced activity.[31] Most substitutions with LNA, on the contrary, led to reduction of anticoagulant activity.[32] The effect on quadruplex stability of North-nucleoside in the loops of TBA was reported by Eritja and coworkers.[33] The replacement of thymidines in the TGT loop of the TBA quadruplex by uridine (U) and 2'-fluorouridine (FU) induced greater stability to the antiparallel quadruplex structure determined by UV-T_m experiments and CD spectroscopy. However the presence of North-methanocarbathymidine (NT) in the same positions destabilized the quadruplex structure. Also, the substitution of thymidines in the TT loops by U, FU and NT destabilized the antiparallel quadruplex structure. Thus, sugar conformations of the nucleotides in the loop region of TBA are important in determining the stability of TBA quadruplex structure. (**Figure 4.6**)

Phosphorothioates (**Figure 4.7a**) in the internucleotide linkage of the guanines resulted in reduced stability, but in the TT loop increased nuclease resistance, at the same time, retaining the stability and anticoagulant activity.[34] A series of 15-mer oligonucleotides with
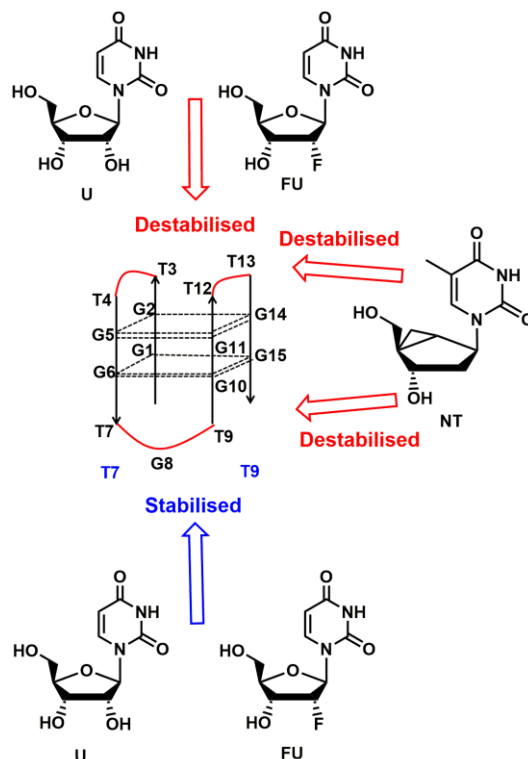
Figure 4.6: Effect of sugar modification in loop region of TBA.

one or more negatively charged phosphodiester groups replaced by a neutral formacetal group (**Figure 4.7b**) were also synthesized and evaluated for their thrombin inhibitory activity *in vitro* and *in vivo*. Replacement of the negatively charged phosphodiester linkage with a neutral formacetal linkage[35] was studied with the goal to understand the role of the phosphodiester group in binding to thrombin, as well as the position of the group or groups critical for the interactions. Another target was to derive analogues with increased *in vivo* half-life through structural modifications.
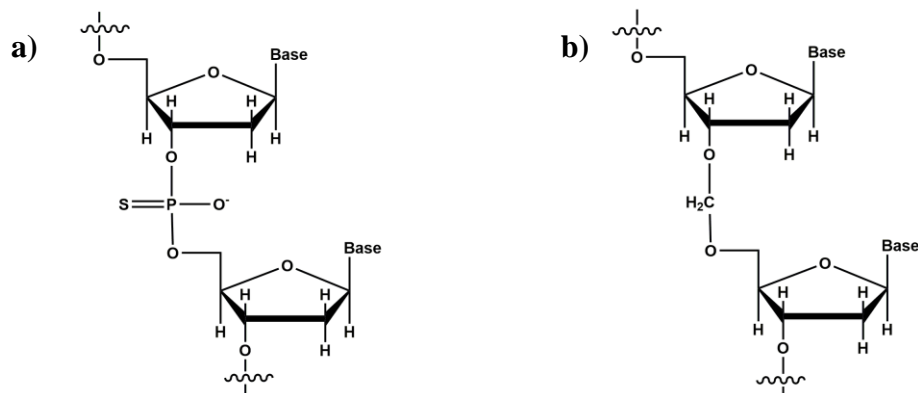
Figure 4.7: Dinucleotides showing phosphorothioate and formacetal internucleoside linkages

TBA has a short half- life which makes it an ideal anticoagulant for cardiopulmonary bypass surgeries, where a quick onset of anticoagulation and rapid return to normal clotting value is desirable. However analogues of TBA with an increased *in vivo* half-life could be useful in clinical cases such as deep vein thrombosis where extended anticoagulant effect is necessary. The short half-life of TBA is mainly due to rapid tissue uptake.[36] Cellular uptake of oligonucleotide (ODN) occurs mainly through receptor-mediated endocytosis[37] and binding of an ODN to the cell surface is mainly based on charge-charge interactions.[38] It was found that more than one of the phosphodiester groups are simultaneously involved in the binding with thrombin. For the ODNs containing two non-contiguous formacetal groups, additive results were obtained for the effect of the substitution on the thrombin inhibitory activity. The *in vivo* anticoagulation study in monkeys showed that the ODN containing four formacetal groups shows an increased *in vivo* PT and extended *in vivo* half-life compared to the unmodified ODN, suggesting that replacement of negatively charged phosphodiester groups with neutral formacetal groups may decrease the tissue uptake of the ODN.

Modified nucleobases in the form of 4-thio-dU in a position-dependent manner resulted in significant improvements in anticoagulant activity.[39] Similarly, *iso*guanine nucleobases were used to increase the stability of modified TBA.[40] Introduction of 5-nitroindole as nucleobase modification at a particular position enhanced the clotting time to nearly 5-fold.[41]Inversion of strand polarity[42] was also shown to have a favorable effect on thermal stability as well as thrombin-affinity. However, decreased thrombin inhibition was observed in this case. It has also been demonstrated that the quadruplex stability of TBA is not the only factor influencing anticoagulant activity. Specific TBA-thrombin interactions are equally important, particularly those involving the loop regions.

Backbone modifications of TBA are reported to have profound effects on the structural topology of the resulting quadruplex. The *iso*-sequential RNA (rTBA) was shown to form a multimolecular parallel G-quadruplex,[43] in contrast to TBA, which is a unimolecular antiparallel G-quadruplex.[27,5] A mixed DNA/RNA backbone TBA sequence folded in either DNA-like or RNA-like quadruplex structure, depending on the position of the ribo- or deoxyribo-nucleotides in the sequence.[44] An RNA-based SELEX process for thrombin-binding yielded a nucleobase sequence completely different from TBA.[45,46] The nucleobase

sequence of the recently reported 3',2'-TNA aptamer (**Figure 4.8a**) for thrombin-binding, although a G-rich sequence,[47] was also different from TBA. The loop configuration of the quadruplex can also be a deciding factor for various structural topologies with varying loop configurations.[1b] Replacement of 3'5' backbone with 2'5' backbone (**Figure 4.8b**) by our group earlier not only formed G- quadruplex but also efficiently enhanced the clotting time.[48]
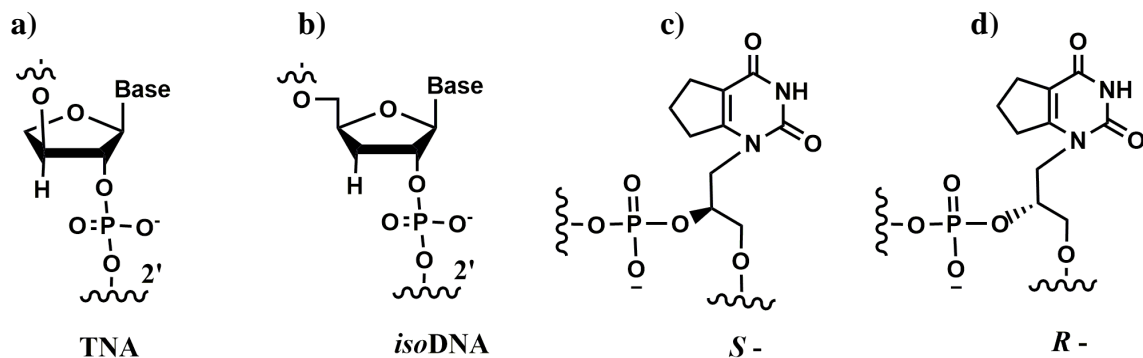

Figure 4.8: Reported modifications of thrombin binding aptamer

Other backbone modifications include replacement of the sugar moiety with optically active acyclic nucleoside as well as modifying the nucleobase. While introduction of both *R*, *S* modification stabilized the quadruplex structure, the stabilization was rather dependent on the position of modification and not on the stereochemistry of the centre (**Figure 4.8 c, d**).[49]

4.2 Introduction to present work

Rationally designed biomolecules and their mimics leading to self-assembled supramolecular structures are currently being studied as a highly exciting research topic.[50] The last decade has seen the advent of DNA-[51] and protein[52]- based supramolecular nanostructures with various shapes and properties, defined by the intra/intermolecular interactions of their primary sequences comprising nucleotides and amino acids, respectively. Watson-Crick nucleobase-complementarity[53] and base-stacking interactions have been mainly used to dictate the highly predictable DNA/PNA nanostructures.[54] In the case of engineered self-assembling polypeptide structures, complex co-operative interactions be-

tween the native/modified α or β- amino acids allowed the formation supramolecular contact surfaces. The long range interaction of polypeptide chains leading to three dimensional arrays remains still unpredictable and therefore highly challenging.[55] Peptides form different three dimensional structures through hydrogen bonding. However, the commonly formed secondary structures of peptides are i) random coil ii) β-sheet iii) α-helix
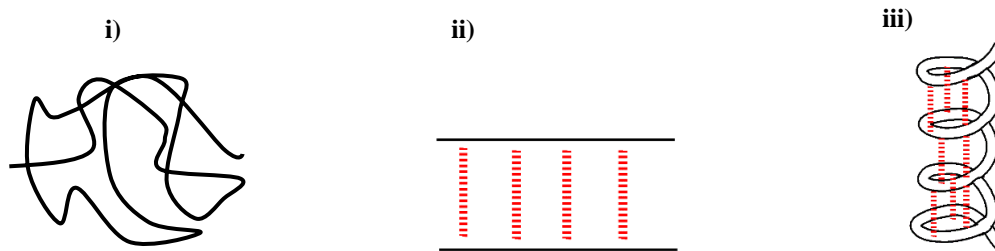
i)   ii)   iii)

Figure 4.9: Secondary structures of peptides

Nucleic acids are also known to fold into three dimensional space by using Hoogsteen hydrogen bonding as in G-quadruplex structure.[56] The quadruplex structures with PNA backbone have been previously reported in the literature for short PNA sequences that form either four stranded multimolecular structures or mixed bimolecular and multimolecular structures.[57] Balsubramanian and co-workers reported short PNA sequence G_4T-Lys, to form a four stranded parallel quadruplex.[58] Tetramolecular PNA quadruplex motif was also studied by the same group to explore dynamic covalent chemistry on self-templating PNA oligomers.[59] PNA quadruplexes were reported earlier by Datta et al[60] for a $G_4T_4G_4$-Lys sequence which could form either two and/or four stranded quadruplexes. Lusvarghi et al also reported G-quadruplexes with PNA backbone in which the thymine units in the loop region were replaced by miniPEG[61] units as abasic sites to differentiate the formation of heteroduplex and heteroquadruplex structures with DNA and RNA. Unimolecular folded G-quadruplex structures with PNA backbone are not yet reported in the literature. Our first objective in these studies was to find if the homologous PNA sequence could form folded unimolecular tetraplex.

4.3 Rationale

TBA forms a chair-like unimolecular G-quadruplex structure that specifically binds thrombin via specific interactions with the adjacently placed TT loop regions.[62] In this case the TT loops are placed proximally due to intramolecular folded G-quadruplex formation. The specific binding of TBA with thrombin probably is due to the array of specifically oriented phosphate groups in the TT-loop regions of TBA.[61b] The stability of the secondary and tertiary structures of the unimolecular folded quadruplexes is governed by the backbone modifications as well as the loop length and the loop sequences.[64] Inspired by the three dimensional structure of TBA, as a starting point, we chose to study if the PNA backbone of TBA. Further, we envisaged to systematically replace the loop thymine monomers by a suitable length of peptide chain. This would allow the confluence of the G-quadruplex structure and spatially organized three dimensional peptide arrays in the adjacent loop regions of the probable G-quadruplex structures. Here, we propose to study the TBA sequence comprising PNA backbone and the designed hybrid PNA-peptide oligomers, having guanine-PNA monomers in the G-quadruplex forming region, and different amino acid sequences with probable ability of forming α-helices and β-turns in the loop regions. We would thus be addressing the question if G-quadruplex formation would dictate the α-helical loops to turn and form unimolecular G-tetraplexes or would the stable β-turn motifs stabilize the unimolecular chair-like structural features of G-quadruplex. In such a case, the two long distanced peptide chains could be predictably proximally arranged. Our choice of PNA sequence was that of a well-studied intramolecular folded TBA sequence having DNA backbone. This sequence contains two TT loops and a single TGT loop. *i.e.* $G_2T_2G_2TGTG_2T_2G_2$.

We have thus created a possibility of bringing together the helix bundles with the help of stable G-quadruplexes or bring the long range amino acids in the loop region in close proximity with the help of defined unimolecular quadruplex structure.

We also planned to incorporate the modPNA-TBA (discussed in Chapter 3)[63] in the loop region keeping in mind the –OH appendages which can be further derivatized according to necessity.

4.4 Synthesis of PNA-TBA and modPNA-TBA oligomers

In the homogeneous PNA sequence we used L-lys as the C-terminus amino acid to enhance the solubility of PNA oligomer. However, we used β-alanine at the C-termini of modPNA-TBA sequence. PNA-TBA and modPNA-TBA oligomers were synthesised on solid phase utilizing Boc chemistry on Lysine or β-Alanine derivatized MBHA resin. The oligomer was synthesized with the aid of CEM peptide synthesizer under microwave conditions. The crude PNA was purified and characterized following the same procedure as drafted in *chapter 2*.

Table 1: PNA-TBA sequence synthesized and characterized

Seq code	Sequences	HPLC t_R (min)	Calc. mass	Obsvd. mass
PNA-TBA	ggttggtgtggttgg-Lys	11.8	4364.1	4369.2(M+Na)
modPNA-TBA	ggttggtgtggttgg-β-Ala	11.2	4547.66	4550.74

4.5 Synthesis of DNA Oligonucleotides

The DNA-TBA G-quadruplex and complementary DNA sequences were synthesized on Bioautomation Mer-Made 4 synthesizer using standard β-cyanoethyl phosphoramidite chemistry. The sequences were synthesized on polystyrene solid support, followed by ammonia treatment. The purity of the oligomers was ascertained by RP-HPLC on a C18 column to be more than 98% and was used without further purification for biophysical studies.

List of DNA sequences

DNA-TBA: 5'-GGTTGGTGTGGTTGG-3'   **cDNA**: 5'-CCAACCACACCAACC-3'

4.6 UV melting experiment

A plot of absorbance scan recorded between 220nm to 320nm *versus* wavelength at 10ºC and 90ºC exhibits an isobestic point around 280 nm, whereas a net hyperchromism is observed upon G-quartet formation at 285 nm or higher wavelength. This difference is maximal at around 295 nm (**Figure 4.10**). The wavelength of 260 nm chosen for most nucleic

acid UV-melting studies is trivial for analysis of quadruplex structures. This is because absorbance recorded at 260nm might show increase, decrease or no co-operative change at 260nm with increasing temperature.[64] Each structure has a characteristic differential absorbance signature, meaning that a simple absorbance analysis at high and low temperatures will help to determine the proper wavelengths for melting analysis and will also give indications on the nature of the folded conformation. In the case of quadruplexes, it is possible to follow co-operative decrease of absorbance at 295 nm. On the other hand, the absence of any cooperative transition at 295 nm provides evidence that the structure formed is not a G-quadruplex.
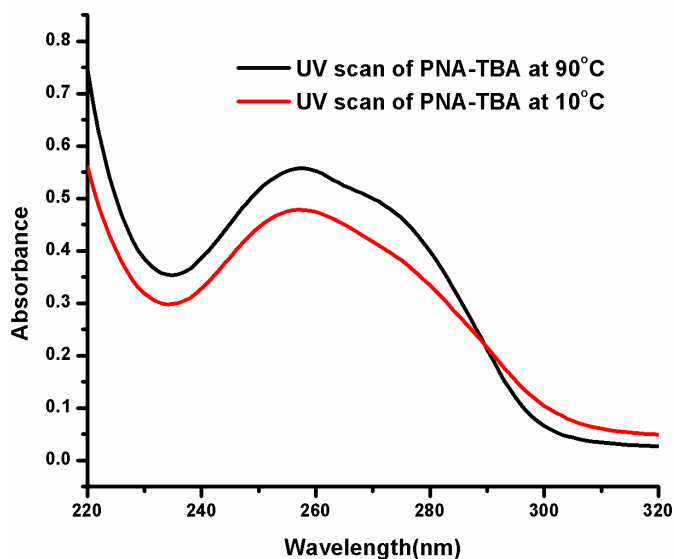

Figure 4.10: Representative absorbance scan of PNA-TBA at 10ºC and 90ºC

Thus, stability of the DNA-TBA, PNA-TBA and modPNA-TBA sequences was followed by the change in the UV absorbance at 295 nm with temperature.

Table 2: UV-melting experiments of DNA-TBA, PNA-TBA in water, NaCl, KCl

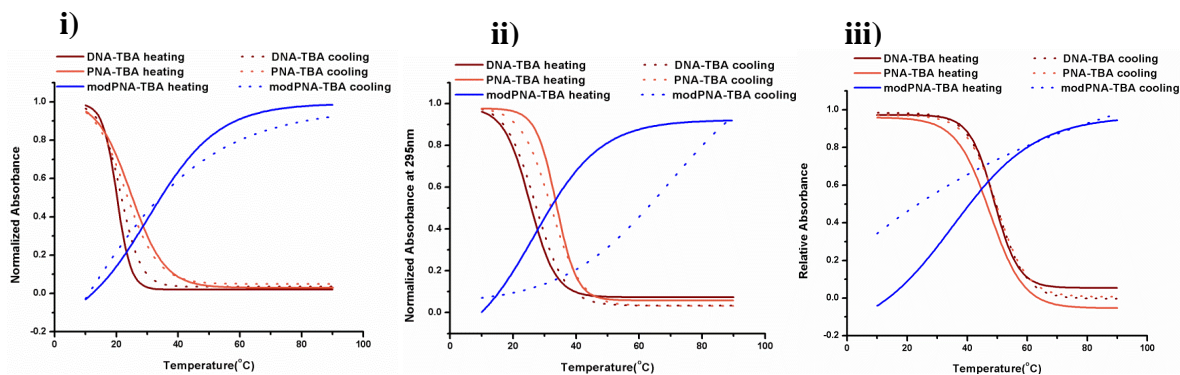| Sequences	T_m water at 295nm		T_m NaCl at 295nm		T_m KCl at 295nm	
	Heat	**Cool**	**Heat**	**Cool**	**Heat**	**Cool**
DNA-TBA	20.45	20.2	22.6	22.6	50.0	49.7
PNA-TBA	21.2	21.7	31.9	32.3	48.4	48.7
modPNA-TBA	n.t.	n.t.	n.t.	n.t.	n.t.	n.t.

Figure 4.11: UV melting and cooling curves of DNA-TBA, PNA-TBA and modPNA-TBA in i) water ii) NaCl iii) KCl

The monovalent cations Na^+ and K^+ were necessary for the stability of the quadruplex structure, K^+ being most favored. PNA-TBA like DNA-TBA was found to form unimolecular quadruplex as evident from the absence of hysteresis between the heating and cooling curves. The modified PNA used in this study where both the 'tt' loops in **modPNA-TBA** sequence are modified using β, γ-bis-(hydroxymethyl)- substituted PNA, did not exhibit quadruplex melting, as increase in absorbance with temperature was observed at 295nm (**Figure 4.11**). The modified thyminyl units in 'tt' loop regions are highly substituted and could have caused steric destabilization of either inter- or intra-molecular quadruplex structure. Another reason for this could be the extended structure of the substituted PNA monomers due to possible internal hydrogen bonding as suggested earlier while duplexation with DNA/RNA.[63]

PNA-TBA sequence, when mixed with complementary DNA, displayed linear increase in absorbance at 295nm. A melting was observed only when recorded at 260nm, the T_m as tabulated below.

Table 3: UV-melting experiments of PNA-TBA with cDNA in KCl

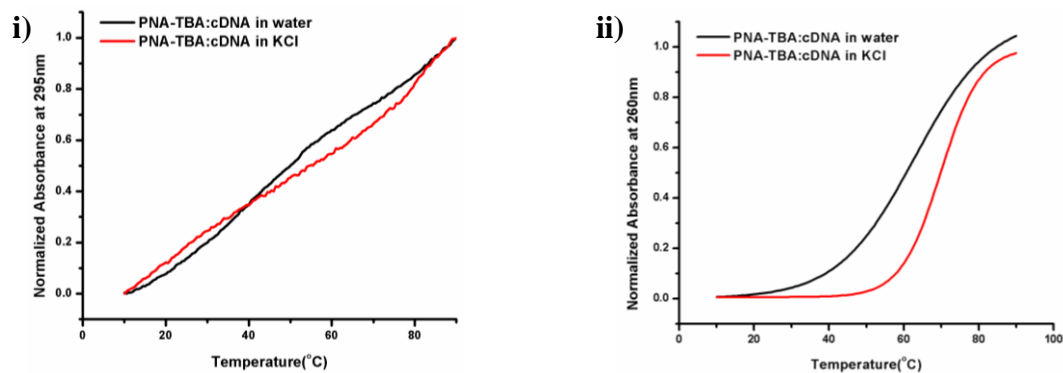| Sequence	At 260nm		At 295nm	
	water	100mM KCl	water	100mM KCl
PNA-TBA	69.8	63.2	n.t.	n.t.

Figure 4.12: UV melting and cooling curves of PNA-TBA with cDNA at i) 295nm ii) 260nm

UV melting of 1:1 mixture of PNA-TBA: cDNA at 260nm denotes that PNA-TBA forms primarily duplex structure in presence of complementary DNA. Absence of co-operative melting at 295nm further confirms inability of PNA-TBA to form quadruplex in presence of cDNA.

4.7 Thermal Difference Spectra (TDS)

TDS is a distinct spectroscopic signature or a 'finger print' which is characteristic of each kind of nucleic acid structure viz single strand, duplex, triplex, i-motif and quadruplexes. [65] The spectra obtained from arithmetic difference between the spectral scan recorded above and below the melting temperature is known as thermal difference spectra. A characteristic thermal difference spectrum of quadruplex shows a positive peak at 243nm, 273nm and a negative peak at 295nm. Thus, thermal difference spectra of the sequences were noted to find out if PNA-TBA and modPNA-TBA sequences formed quadruplex.
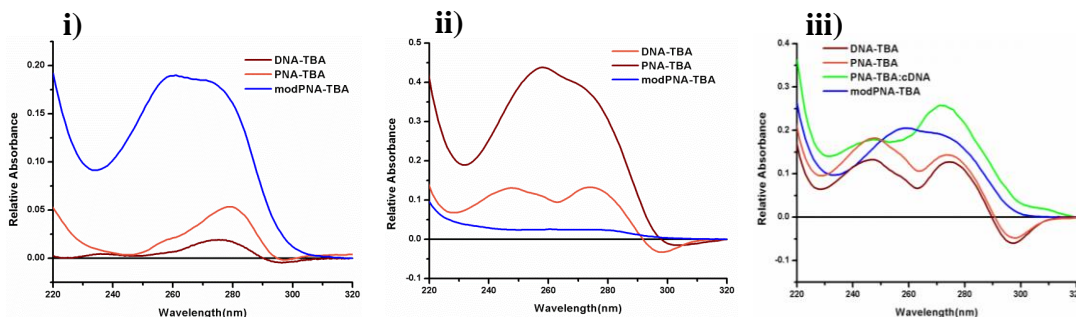

Figure 4.13: Thermal difference spectra of i) DNA-TBA, PNA-TBA and modPNA-TBA in water, ii) DNA-TBA, PNA-TBA and modPNA-TBA in NaCl iii) DNA-TBA, PNA-TBA, modPNA-TBA and PNA-TBA: cDNA in KCl

Thermal difference spectra of PNA-TBA confirmed its ability to form quadruplex. The TDS spectrum of the quadruplex was distinctly different from that of duplex formed from PNA-TBA and cDNA (**Figure 4.13 iii**). The 'tt' substituted **modPNA-TBA** did not show characteristic TDS of quadruplex structure which was in accordance to the UV-melting experiments.

4.8 Peptide modification in PNA-TBA sequences

The loop region of thrombin binding aptamer was then modified with amino acids keeping in mind the ability of amino acids to form three dimensional structures and the ease with which the side chains can be modified according to necessity. The simplest, achiral amino acid glycine was chosen to study the effect of incorporation of amino acids on quadruplex structure. The 'tt' loop regions were then modified with chiral amino acid L-alanine and D-Alanine to study the effect of induced chirality in the quadruplex. A chain of four amino acids was incorporated in the loop region to meet the requirements of replacing the dinucleotide loop stretch (**Figure 4.14**). The methyl groups in the side chains in alanine are known to be promoting helical structures as compared to the randomness of gly-oligomers.[66]
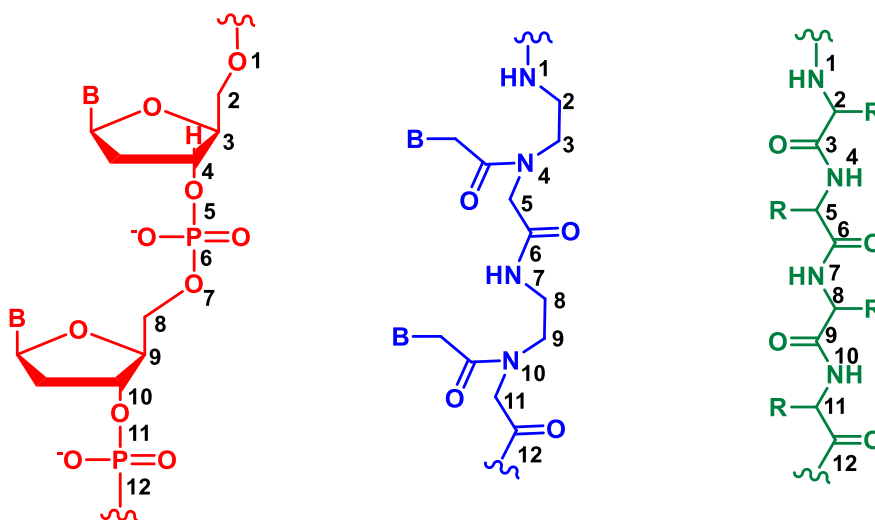

Figure 4.14: Designing the peptide modified loop region

4.8.1 Peptide modified PNA-TBA sequences synthesized and characterized

The monomers employed for solid phase peptide synthesis are commercially available Boc-Glycine, Boc-L-Alanine, Boc-D-Alanine. The peptide modified PNA-TBA sequences were synthesized, purified and characterized following the protocol described earlier.

Table 4: PNA-Gly, PNA-L-Ala and PNA-D-Ala sequence synthesized and characterized

Seq code	Sequences	HPLC t_R (min)	Calc. mass	Obsvd. mass
PNA-Gly	gg(Gly)$_4$ggtgtgg(Gly)$_4$gg-β-Ala	10.8	3696.41	3693.8
PNA-L-Ala	gg(Gly-L-Ala)$_2$ggtgtgg(Gly-L-Ala)$_2$gg -β-Ala	11.3	3752.47	3754.5
PNA-D-Ala	gg(Gly-D-Ala)$_2$ ggtgtgg(Gly-D-Ala)$_2$ gg-β-Ala	11.5	3752.47	3754.6

4.8.2 UV melting experiment and thermal difference spectra of PNA-Gly, PNA-L-Ala and PNA-D-Ala

UV-melting and cooling experiments were recorded for PNA-Gly, PNA-L-Ala and PNA-D-Ala sequences.

Table 5: UV-melting experiments of PNA-Gly, PNA-L-Ala and PNA-D-Ala

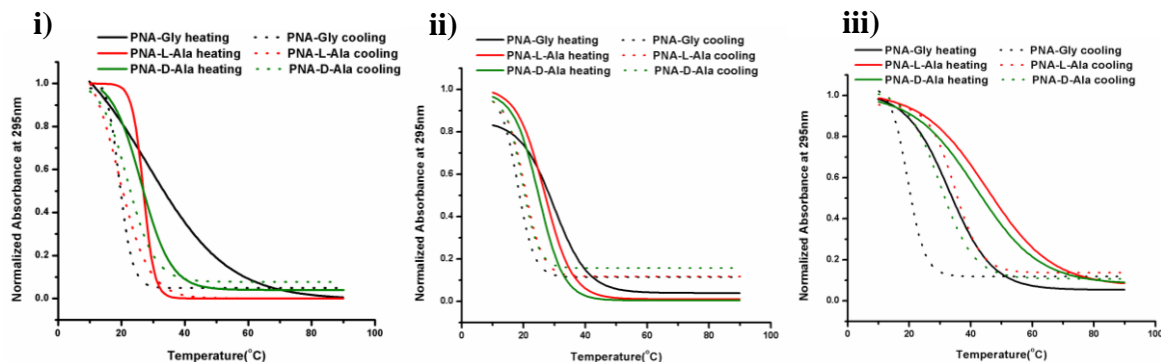| Sequences	T_m water at 295nm		T_m NaCl at 295nm		T_m KCl at 295nm	
	Heat	Cool	Heat	Cool	Heat	Cool
PNA-Gly	28.8	19.6	30.8	21.8	34.1	20.7
PNA-L-Ala	27	19.1	26.9	20.7	45.3	35.5
PNA-D-Ala	26.8	20.2	24.8	19.2	41.8	30.2

Figure 4.15: UV melting and cooling curve of PNA-Gly, PNA-L-Ala and PNA-D-Ala in i) water, ii) NaCl, iii) KCl

Modifying the loop region with glycine and alanine changed the structure from unimolecular to multimolecular quadruplex which was evident from the hysteresis observed in melting and annealing curves in cases of modified PNA oligomers when monitored at 295 nm wavelength.
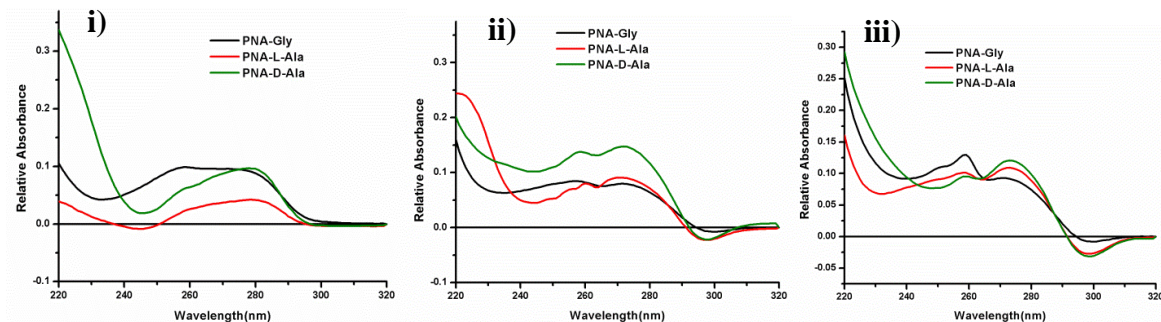

Figure 4.16: Thermal difference spectra of PNA-Gly, PNA-L-Ala and PNA-D-Ala in i) water, ii) NaCl and iii) KCl

Glycine is the most flexible prochiral amino acid whereas introduction of alanine is expected to increase the propensity of formation of α-helical peptide geometry. The studied quadruplexes with peptides containing glycine and alanine may not have induced the turns in the oligomers required for the unimolecular quadruplex and thus might be responsible for the formation of multimolecular quadruplex by association of G-tetrads using two or four different strands.

4.9 Hydroxyproline modifications of PNA-TBA

Proline and its derivatives play a salient role in protein folding and refolding. They influence a reversal in backbone conformation, triggering nucleation of turns as well as breaking of helices in proteins.[67] Literature report proposes that presence of alternate L-proline and D-proline results in stronger and explicit β-turn in a peptide.[68] Role of 4-hydroxy proline[69] has been studied extensively among other derivatives of proline. As increasing the hydrophobicity using proline would diminish the solubility of PNA in water, we used hydroxyprolines instead of prolines in our design. Therefore, we replaced the alanine amino acids in the loop region with *trans*-4-hydroxy-L-proline (L-Hyp) and *trans*-4-hydroxy-D-proline (D-Hyp) to get a unimolecular quadruplex of TBA sequence. They were also used together to get stronger β-turn. We further attempted to introduce valine and thre-

onine adjacent to alternating D-Hyp and L-Hyp respectively in the loop region to study their effect on the quadruplex stability.[70]
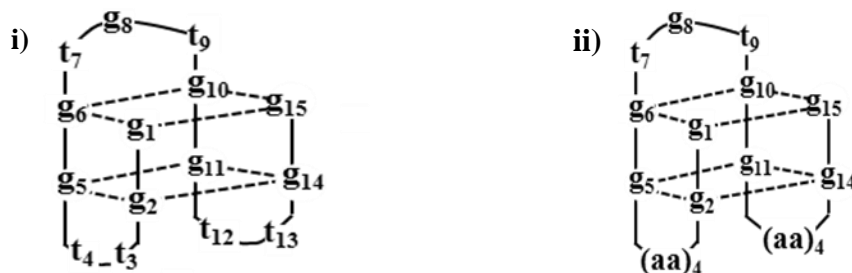

Figure 4.17: i) structure of PNA-TBA ii) projected structure of *trans*-hydroxyproline modified PNA-TBA

4.9.1 Synthesis and characterization of *trans*-hydroxyproline modifications of PNA-TBA

The *trans*-Boc-L-Hyp-OBn monomers utilized for synthesis of oligomer were synthesized following the route reported earlier (**Scheme 4.1**).[71,72] The *trans*-Boc-D-Hyp-OBn [73] monomer was also synthesized from *trans*-4-hydroxy-L-proline as described below (**Scheme 4.2**). Epimerisation of *trans*-4-hydroxy-L-proline to *cis*-4-hydroxy-L-proline was performed as one pot reaction according to Vanek *et al.*[74] Successive ester and Boc protection of free acid and amine of compound **4** produced compound **6** in quantitative yield**.** *p*-nitrobenzyl ester has been extensively used to invert the stereochemistry of chiral alcohol via Mitsunobu reaction.[75] Compound **6** was esterified with *p*-nitrobenzyl alcohol and was orthogonally deprotected in presence of methyl ester using NaN_3 in methanol to achieve compound **8**.[76] Compound **8** was converted to **3b** following the same procedure of **3a.**
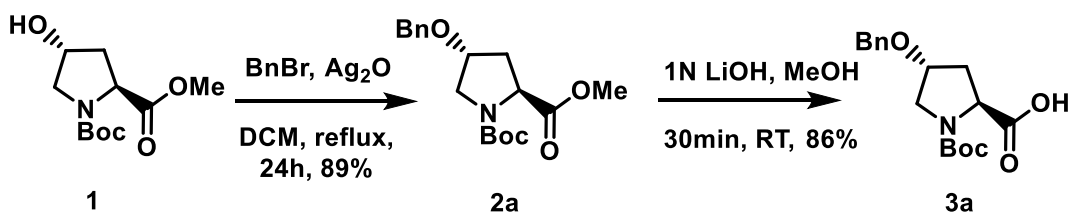
Scheme 4.1: Synthesis of *trans*-Boc-L-Hyp-OBn monomer
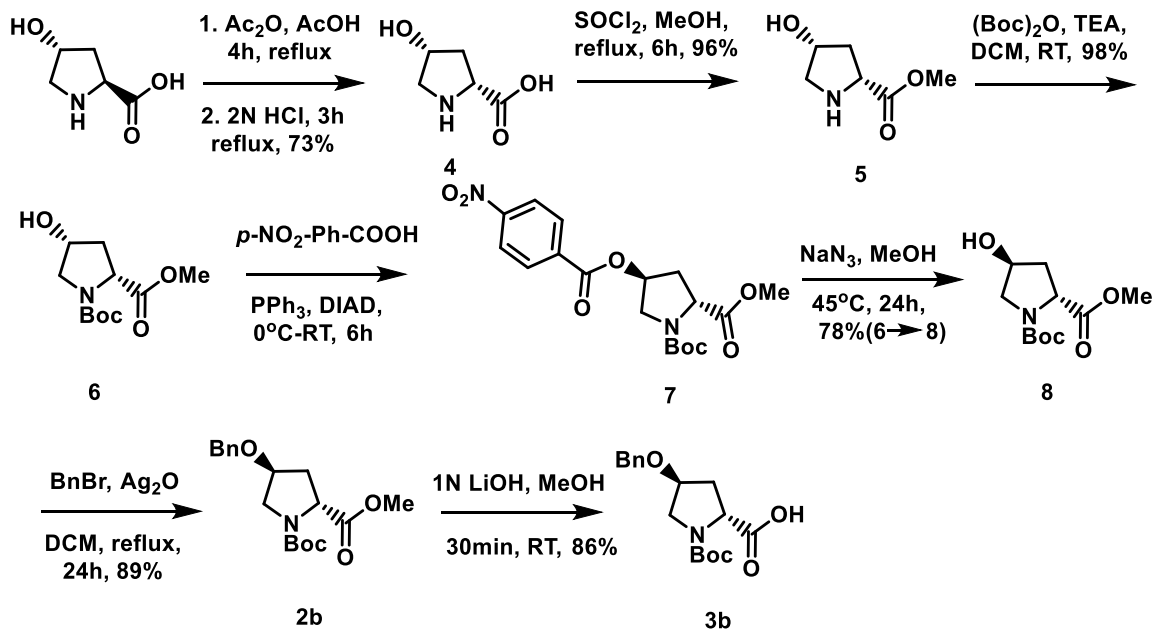
Scheme 4.2: Synthesis of *trans*-Boc-D-Hyp-OBn monomer

Synthesis of all the oligomers was carried out under microwave conditions. Boc-Val and Boc-Thr-OBn monomers applied for synthesis of PNA-Val were commercially available.

Table 6: PNA-L-Hyp, PNA-D-Hyp, PNA-(L+D)-Hyp and PNA-Val sequence synthesized and characterized

Seq code	Sequences	HPLC t_R(min)	Calc. mass	Obsvd. mass
PNA-L-Hyp	gg(Gly-L-Hyp-L-Hyp-Gly)ggtgtgg (Gly-L-Hyp-L-Hyp-Gly)gg-β-Ala	11.6	3920.5	3922.2
PNA-D-Hyp	gg(Gly-D-Hyp-D-Hyp-Gly)ggtgtgg (Gly-D-Hyp-D-Hyp-Gly)gg-β-Ala	11.3	3920.5	3922.0
PNA-(L+D)-Hyp	gg(Gly-L-Hyp-D-Hyp-Gly)ggtgtgg (Gly-L-Hyp-D-Hyp-Gly)gg-β- Ala	10.8	3920.5	3919.5
PNA-Val	gg(Val-D-Hyp-L-Hyp-Thr)-ggtgtgg (Val-D-Hyp-L-Hyp-Thr)-gg-β- Ala	11.2	4092.6	4093.8

4.9.2 UV melting and thermal difference spectra of *trans*-hydroxyproline modified PNA oligomers

UV melting experiments were performed to test the ability of the *trans*-hydroxyproline modified PNA oligomers to form unimolecular quadruplex. Their ability to form quadruplex was further confirmed by thermal difference spectra.

Table 7: UV melting experiment of PNA-{L, D, (L+D)–Hyp} and PNA-Val oligomers

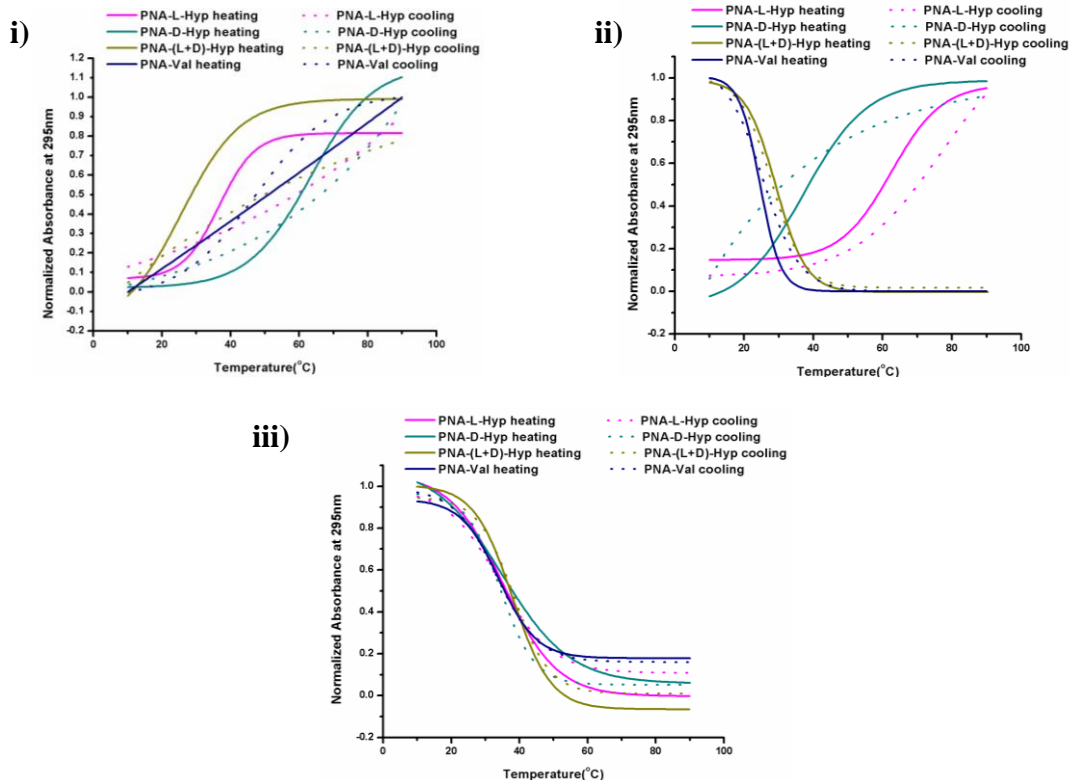| Sequences	T_m water at 295nm		T_m NaCl at 295nm		T_m KCl at 295nm	
	Heat	**Cool**	**Heat**	**Cool**	**Heat**	**Cool**
PNA-L-Hyp	n.t	n.t	n.t	n.t	35.3	34.9
PNA-D-Hyp	n.t	n.t	n.t	n.t	34.9	34.1
PNA-(L+D)-Hyp	n.t	n.t	29.5	28.9	38.2	38.8
PNA-Val	n.t.	n.t.	24.8	25.4	33.2	33.8

Figure 4.18: UV melting and annealing curves of PNA-L-Hyp, PNA-D-Hyp, PNA-(L+D)-Hyp, PNA-Val in i) water ii) NaCl iii) KCl
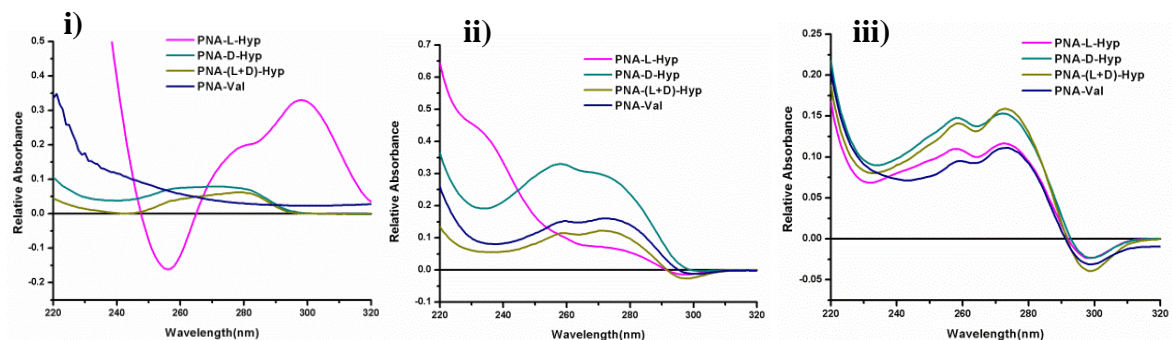
Figure 4.19: Thermal difference spectra of i) PNA-L-Hyp ii) PNA-D-Hyp iii) PNA-(L+D)-Hyp, iv) PNA-Val

PNA-L-Hyp, PNA-D-Hyp, PNA-(L+D)-Hyp and PNA-Val were unable to form quadruplex structure in water. Na^+ was able to fit in the central void of PNA-(L+D)-Hyp and PNA-Val only to form G-quartet. However, all hydroxyproline modified PNA-TBA sequences viz PNA-L-Hyp, PNA-D-Hyp, PNA-(L+D)-Hyp and PNA-Val formed stable quadruplex in presence of KCl. The expected topology would be intramolecular folding and quadruplex formation would be unimolecular as no hysteresis was observed in heating and cooling profiles. It was observed that PNA-peptide mixmers containing hydroxyproline residues form unimolecular quadruplexes but these are thermally less stable as compared to DNA-TBA or PNA-TBA.

The thermal difference spectra further confirmed the formation of quadruplex of the hydroxyproline modified oligomers. Furthermore, replacement of glycines with valine and threonine resulted in destabilizing the quadruplex.

4.10 CD spectra of PNA-TBA and modified PNA-TBA oligomers

DNA quadruplexes are often followed by CD spectroscopy.[77] Study of circular dichroism gives an insight of the orientation and folding of strands i.e. parallel or antiparallel. The spectra of "parallel" quadruplexes, in which four strands run parallel to each other have a dominant positive band at 260 nm, and a negative peak at 240 nm.[78, 79] Conversely, the spectra of "antiparallel" quadruplexes where the strands are opposite to each other have a negative band at 260 nm and positive band at around 290 nm.[77, 80] Both quadruplex types display an additional characteristic positive peak at 210 nm. The empirical simple relationship

§ Positive CD at 260 nm and negative CD at 240 nm → parallel G-quadruplex

§ Positive CD at 290 nm and negative CD at 270 nm → antiparallel G-quadruplex

is a confirmatory outcome of CD experiment. Thus, CD spectroscopy gives a vivid idea of the strand polarity of G-quartets and has become an important tool for researchers working on G-quadruplexes.

However, for PNA-TBA and modified PNA-TBA quadruplexes no significant CD signal was observed. This does not necessarily rule out the possibility of PNA-TBA or modified PNA-TBA oligomers forming quadruplexes. The absence of CD signal is most likely due to the achiral backbone of PNA. Introduction of chiral amino acids like L-Ala, D-Ala, L-Hyp and D-Hyp also could not induce any CD signals.
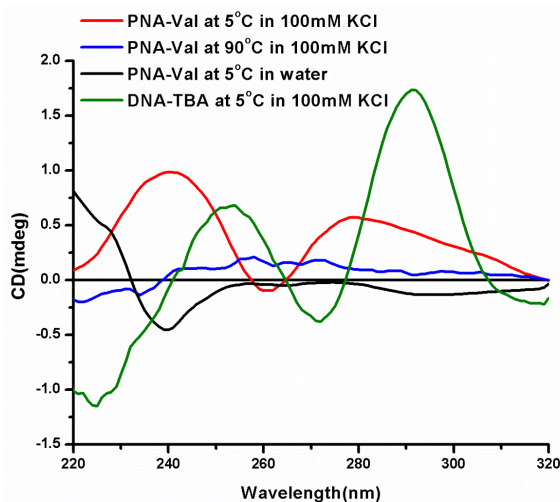

Figure 4.20: CD spectra of PNA-Val (30μM oligomer concentration) in water at 5ºC, in 100mM KCl at 5 ºC and 90ºC and DNA-TBA (1μM oligomer concentration) in 100mM KCl at 5 ºC

CD signals were successfully observed only in case of PNA-Val at an oligomer concentration of 30μM at 5°C. The CD signal disappeared when the sample was heated to 90°C. No CD signal was observed in water which was in accordance with the UV-melting experiment done in water. This illustrated that the CD signal was observed due to the formation of quadruplex. A weak but distinct positive CD signal around 280nm and a weak negative band at 260nm indicate the formation of antiparallel quadruplex for PNA-Val.

4.11 UV melting experiment at higher oligomer concentration

Increasing the oligomer concentration is known to show increase in the T_m for multimolecular quadruplexes but remains unchanged for the unimolecular quadruplex.[81] Thus melting experiments were performed at different oligomer concentrations of 10 μM and 20 μM to further confirm the molecularity of PNA-TBA and modified PNA quadruplexes.

Table 8: UV melting experiment of modified PNA oligomers at higher oligomer concentrations

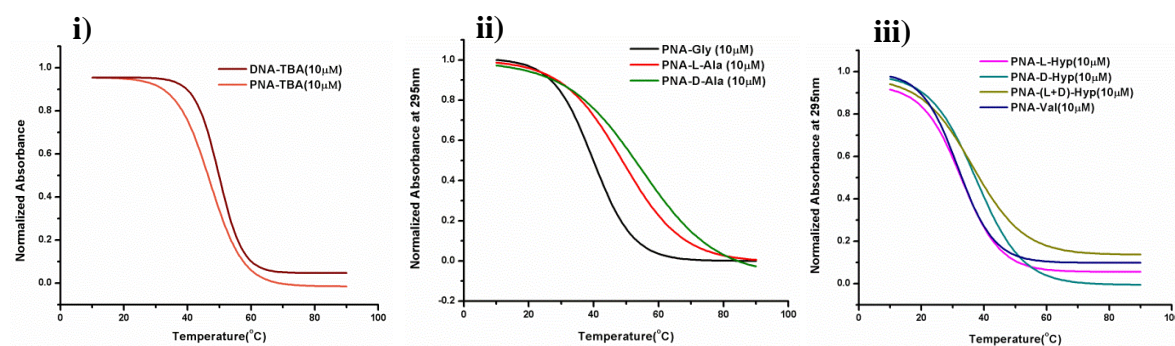| Sequence Code	5μM	10μM	20μM	Composition
DNA-TBA	50.0	49.6	48.2	unimolecular
PNA-TBA	48.4	46.8	47.8	unimolecular
PNA-Gly	34.1	39.8	65.3	multimolecular
PNA-L-Ala	45.3	48.2	61.6	multimolecular
PNA-D-Ala	41.8	53.8	63.5	multimolecular
PNA-L-Hyp	35.3	32.8	36.9	unimolecular
PNA-D-Hyp	34.9	36.8	38.1	unimolecular
PNA-(L+D)-Hyp	38.2	36.5	34.6	unimolecular
PNA-Val	33.2	32.8	35.6	unimolecular

Figure 4.21: UV melting experiments of i) DNA-TBA, PNA-TBA ii) PNA-Gly, PNA-L-Ala, PNA-D-Ala iii) PNA-L-Hyp, PNA-D-Hyp, PNA-(L+D)-Hyp, PNA-Val at 10 μM oligomer concentration
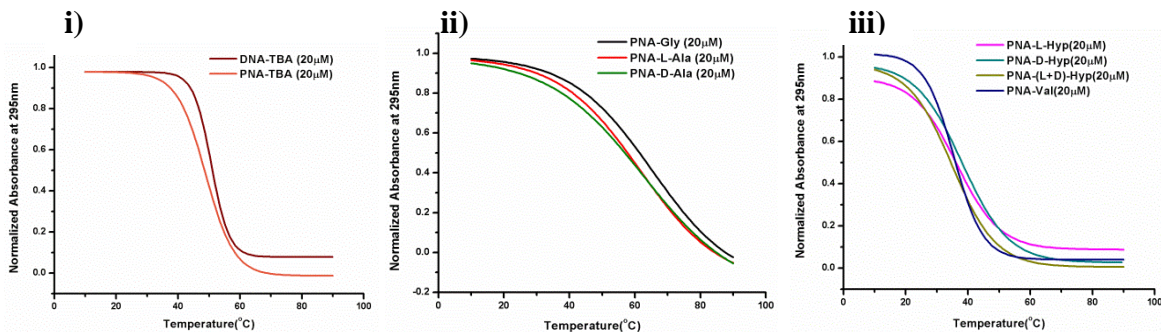
Figure 4.22: UV melting experiments of i) DNA-TBA, PNA-TBA ii) PNA-Gly, PNA-L-Ala, PNA-D-Ala iii) PNA-L-Hyp, PNA-D-Hyp, PNA-(L+D)-Hyp, PNA-Val at 20 µM oligomer concentration

The UV-melting of unimolecular DNA-TBA and PNA-TBA quadruplexes did not show much variation and within experimental error ($\pm$ 1-2 $^{\circ}$C) at higher strand concentrations. The effect of strand concentrations on the melting temperatures however, is clearly evident on the PNA peptide conjugates PNA-Gly, PNA-L-Ala and PNA-D-Ala sequences as T_m increased steadily with increased strand concentrations, supporting intermolecular quadruplex formation in the case of peptide-PNA mixmer sequences comprising glycine and L/D-alanine. The quadruplex formation thus would bring together the peptide strands in a very specific way considering the composition and formation of bimolecular or tetramolecular complexes. In the case of PNA-pepide mixmers comprising gly-L/D-Hyp-gly and val-(L+D)-Hyp-thr peptide sequence, the UV- T_m of the conjugates were found to be unaffected by changing individual strand concentrations. This further substantiates the formation of unimolecular complexes when the turn inducing peptides are utilized in PNA. A unimolecular quadruplex in this case could be similar to antiparallel chair-like quadruplex of DNA-TBA (**Figure 4.23**).
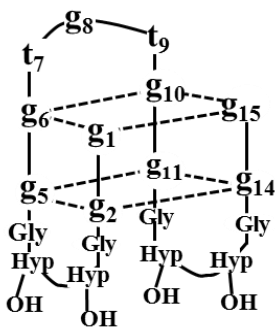

Figure 4.23: Proposed structure of hydroxyproline modified PNA-TBA

The strand orientation in multimolecular complexes could be variable because of sterically unrestricted attachment of the nucleobase to the PNA backbone.

4.12 UV melting experiment at higher salt concentration

Increase of specific alkali metal ion concentration (K^+, Na^+) is known to enhance the stability of G-quadruplex.[82] Hence, next we investigated the role of changed NaCl concentration in G-quadruplex formation for synthesized PNA G-quadruplexes.[47] On increasing the salt concentration there was indeed a rise in the melting temperature and reduction of salt led to decrease of Tm. This is indicative of electrostatic contribution to G-quadruplex stability due to the specifically bound Na+ or K+ ions.

Table 9: UV melting experiment of modified PNA-TBA oligomers at different salt concentration

Sequence Code	T_m at 295 nm (50 mM NaCl)		T_m at 295 nm (100 mM NaCl)		T_m at 295 nm (150 mM NaCl)	
	Heat	Cool	Heat	Cool	Heat	Cool
DNA-TBA	19.2	19.5	22.6	22.6	22.7	23.3
PNA-TBA	19.1	18.2	31.9	32.3	32.2	32.6
PNA-L-Ala	22.6	19.7	26.9	20.7	28.2	22.9
PNA-(L+D)-Hyp	25.2	24.8	29.5	28.9	33.08	32.2

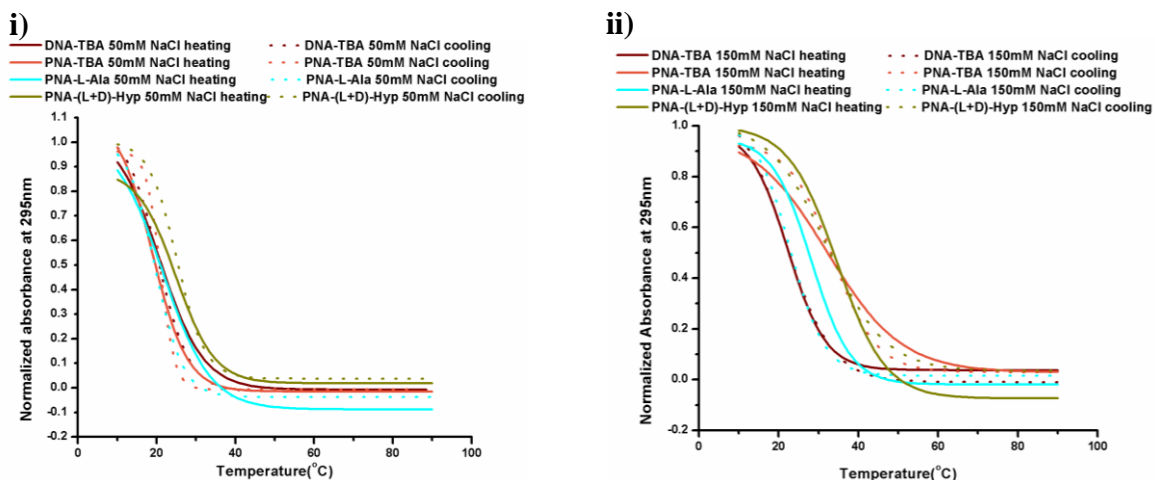**Figure 4.24: UV melting and annealing curves of peptide modified PNA oligomers in i) 50mM NaCl salt concentration ii) 150mM NaCl concentration**

4.13 Thermal Difference Spectra Factor (TDS-Factor)

A careful observation of a representative TDS spectrum shows that in the spectral region of interest (220–320 nm) there are four distinct positive and negative bands around 243, 255, 273, and 295 nm, respectively. Thus, TDS-Factor was defined as the absolute values of $\Delta A240nm/\Delta A295nm$, $\Delta A255nm/\Delta A295nm$, and $\Delta A275nm/\Delta A295nm$, where ΔA is the difference between the absorbance above and below the melting temperature at a particular wavelength.[83] The values obtained are plotted against temperature on X- axis. For parallel quadruplexes, the magnitudes for this TDS-Factor appear above 4, and for antiparallel quadruplexes below 2. This effect can be attributed to the substantially different stacking of electronic transition dipole moments (hypochromicity)
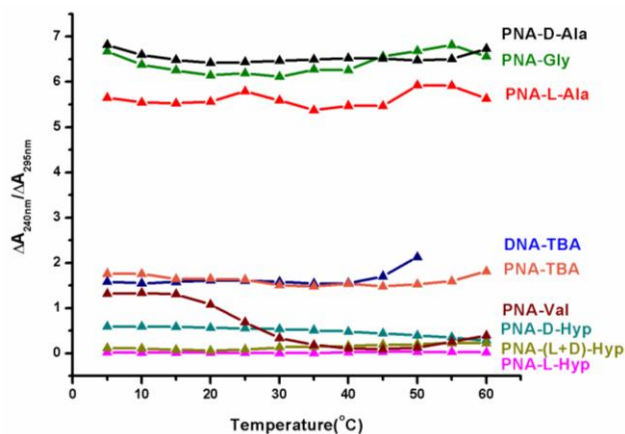

Figure 4.25: Thermal difference spectra factor of PNA and modified PNA oligomers

Although, achiral, acyclic PNA is not restricted by glycosidic bond angles like in DNA, TDS, very similar to that observed for DNA quadruplexes was observed. It was indeed very interesting to see the results obtained for the different PNA oligomers. Although in PNA, nucleobase geometry is not restricted in syn/anti conformations, the quadruplex structures of PNA probably follow the trend that is followed by DNA quadruplex structures. All the PNA oligomers e.g. PNA-DNA, PNA-L-Hyp, PNA D-Hyp, PNA-(L+D)-Hyp and PNA-Val that were found to be unimolecular quadruplexes by UV-T_m measurements exhibited TDS-factor below 2, indicating the antiparallel orientation of the strands and syn/anti conformations for the nucleobases. In the multimolecular PNA quadruplexes, with gly and

glycyl-alanyl loops however, the TDS factor was found to be above 4, suggesting parallel orientation of the multiple strands.

4.14 Electron spray ionization mass spectrometry

Electron spray ionization mass spectrometry (ESI-MS) has been reported earlier to analyse non-covalent complexes of nucleic acids such as G-quadruplexes[84] or their interaction with metal ions in gas phase and antitumor drugs.[85,86] UV melting experiments indicate that PNA-TBA forms unimolecular quadruplex while PNA-Gly, PNA-L-Ala, PNA-D-Ala forms multimolecular quadruplex. Multimolecular quadruplexes are primarily of two kinds, bimolecular and tetramolecular, each of which may display similar mass/charge (m/z) ratios. Nevertheless, when odd numbers of metal ions are associated with four stranded G-quadruplex it indicates formation of tetrameric quadruplex and not dimeric quadruplex. For example, a dimeric quadruplex with x protons and one metal cation(Y) would exhibit m/z of $(2M+Y^{+}+ xH^{+})^{n+}$. However, a tetrameric quadruplex can also display same m/z value for $(4M+2Y^{+}+ 2xH^{+})^{2n+}$. But, quadruplexes showing m/z of $(4M+Y^{+}+7H^{+})^{2n+}$ (for example) is possible only in case of tetrameric quadruplex. This is because the metal ion associated with the quadruplex must be an integral value. This principle has been used earlier to determine the molecularity of quadruplexes[87] as well as complexes arising due to polypeptide chains.[88]

Table 10: Observed and expected ESI-MS of G-quadruplex

S. No.	Species	m/z Obsvd.	m/z Calcd.	Possible molecularity
				PNA-TBA (Molecular weight: 4364.1)
1	$(M+2H^{+})^{2+}$	N.O.	2183.1	unimolecular
2	$(M+K+H)^{2+}$	2202.9	2202.1	unimolecular/bimolecular/tetramolecular
3	$(2M+K+3H)^{4+}$	N.O.	2192.6	bimolecular/tetramolecular
4	$(2M+3K+H)^{4+}$	N.O.	2211.6	bimolecular/tetramolecular
5	$(4M+K+7H)^{8+}$	N.O.	2187.8	tetramolecular
6	$(4M+3K+5H)^{8+}$	N.O.	2197.3	tetramolecular
7	$(4M+5K+3H)^{8+}$	N.O.	2207.2	tetramolecular
8	$(4M+7K+H)^{8+}$	N.O.	2216.3	Tetramolecular

N.O. = Not Observed

PNA-L-Ala (Molecular weight: 3752.47)				
1	$(M+2H^+)^{2+}$	N.O.	1877.2	unimolecular
2	$(M+K+H)^{2+}$	1897.2	1896.2	unimolecular/bimolecular/tetramolecular
3	$(2M+K+3H)^{4+}$	N.O.	1886.7	bimolecular/tetramolecular
4	$(2M+3K+H)^{4+}$	N.O.	1905.7	bimolecular/tetramolecular
5	$(4M+K+7H)^{8+}$	N.O.	1882	tetramolecular
6	$(4M+3K+5H)^{8+}$	N.O.	1891.5	tetramolecular
7	$(4M+5K^++3H^+)^{8+}$	1901.3	1901	Tetramolecular
8	$(4M+7K+H)^{8+}$	1911.2	1910.4	Tetramolecular
PNA-D-Ala (Molecular weight: 3752.47)				
1	$(M+2H^+)^{2+}$	1877.6	1877.2	unimolecular
2	$(M+K+H)^{2+}$	N.O.	1896.2	unimolecular/bimolecular/tetramolecular
3	$(2M+K+3H)^{4+}$	N.O.	1886.7	bimolecular/tetramolecular
4	$(2M+3K+H)^{4+}$	1905.7	1905.7	bimolecular/tetramolecular
5	$(4M+K+7H)^{8+}$	N.O.	1882	tetramolecular
6	$(4M+3K+5H)^{8+}$	N.O.	1891.5	tetramolecular
7	$(4M+5K^++3H^+)^{8+}$	1901.2	1901	Tetramolecular
8	$(4M+7K+H)^{8+}$	N.O.	1910.4	Tetramolecular
PNA-D-Hyp (Molecular weight: 3920.5)				
1	$(M+2H^+)^{2+}$	N.O.	1961.3	unimolecular
2	$(M+K+H)^{2+}$	1981.4	1980.3	unimolecular/bimolecular/tetramolecular
3	$(2M+K+3H)^{4+}$	N.O.	1970.8	bimolecular/tetramolecular
4	$(2M+3K+H)^{4+}$	N.O.	1989.8	bimolecular/tetramolecular
5	$(4M+K+7H)^{8+}$	N.O.	1966	tetramolecular
6	$(4M+3K+5H)^{8+}$	N.O.	1975.5	tetramolecular
7	$(4M+5K^++3H^+)^{8+}$	N.O.	1985	Tetramolecular
8	$(4M+7K+H)^{8+}$	N.O.	1994.5	Tetramolecular
PNA-(L+D)-Hyp (Molecular weight: 3920.5)				
1	$(M+2H^+)^{2+}$	N.O.	1961.3	unimolecular
2	$(M+K+H)^{2+}$	1981.4	1980.3	unimolecular/bimolecular/tetramolecular
3	$(2M+K+3H)^{4+}$	N.O.	1970.8	bimolecular/tetramolecular
4	$(2M+3K+H)^{4+}$	N.O.	1989.8	bimolecular/tetramolecular
5	$(4M+K+7H)^{8+}$	N.O.	1966	tetramolecular
6	$(4M+3K+5H)^{8+}$	N.O.	1975.5	tetramolecular
7	$(4M+5K^++3H^+)^{8+}$	N.O.	1985	Tetramolecular
8	$(4M+7K+H)^{8+}$	N.O.	1994.5	Tetramolecular
PNA-Val (Molecular weight: 4092.66)				
1	$(M+2H^+)^{2+}$	N.O.	2047.3	unimolecular
2	$(M+K+H)^{2+}$	2065.5	2066.3	unimolecular/bimolecular/tetramolecular
3	$(2M+K+3H)^{4+}$	N.O.	2056.8	bimolecular/tetramolecular
4	$(2M+3K+H)^{4+}$	N.O.	2075.83	bimolecular/tetramolecular
5	$(4M+K+7H)^{8+}$	N.O.	2052.1	tetramolecular
6	$(4M+3K+5H)^{8+}$	N.O.	2061.6	tetramolecular
7	$(4M+5K^++3H^+)^{8+}$	N.O.	2071.1	Tetramolecular
8	$(4M+7K+H)^{8+}$	N.O.	2080.6	Tetramolecular

N.O. = Not Observed

In presence of KCl, ions arising from K^+ adducts of PNA-TBA (entry 2) was observed. Presence of odd number of potassium ions bound to the tetrameric complexes of PNA-L-Ala (entry 7, 8) was also observed. Such species are only possible in the case of terameric quadruplex structure. K^+ and H^+ adducts were observed in case of PNA-D-Ala (entry 1, 2) Presence of $(2M+3K+H)^{4+}$ species (entry 4) indicates formation of either bimolecular or teramolecular quadruplex while $(4M+5K^++3H^+)^{8+}$ (entry 7) and $(4M+7K+H)^{8+}$ (entry 8) confirms formation of tetrameric quadruplex. However, absence of such species in case of PNA-TBA, PNA-D-Hyp, PNA-(L+D)-Hyp and PNA-Val further confirms the formation of unimolecular quadruplex. Relative abundance in ESI-MS is known to reflect the availability of different species in the sample.[89] Thus PNA-L-Ala and PNA-D-Ala might favourably be existing as tetrameric quadruplex.
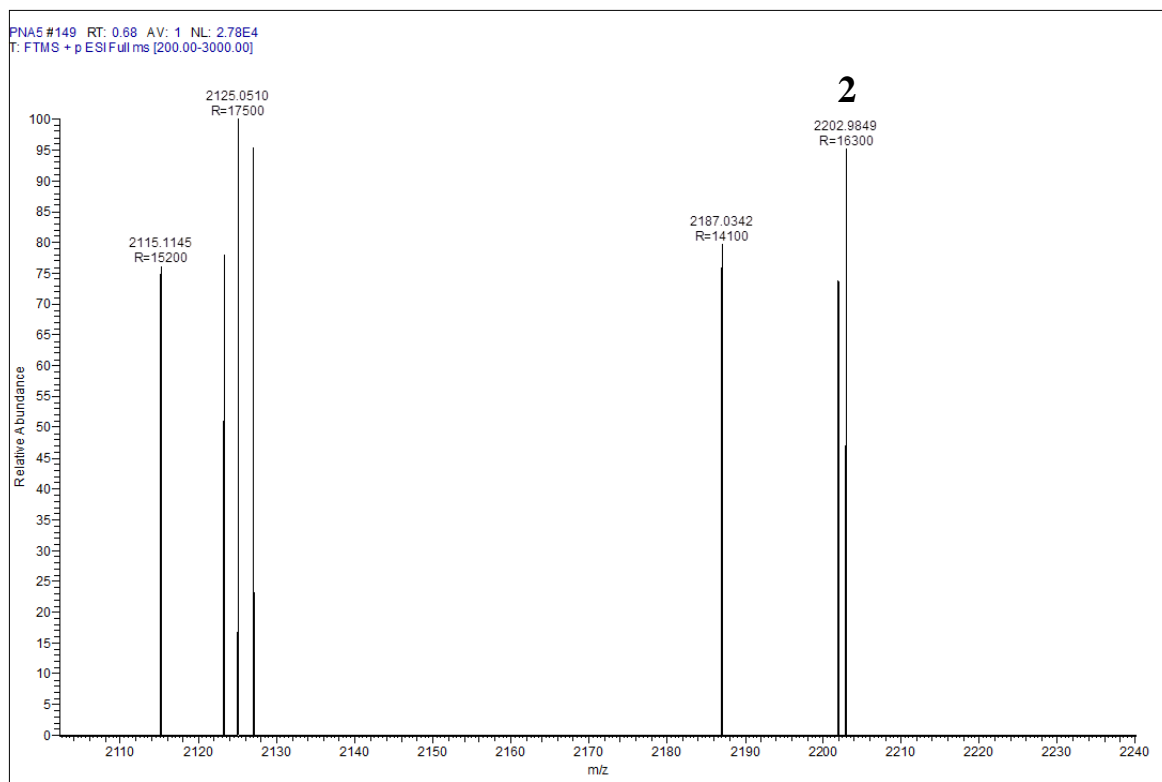
Figure 4.26: ESI-MS of PNA-TBA
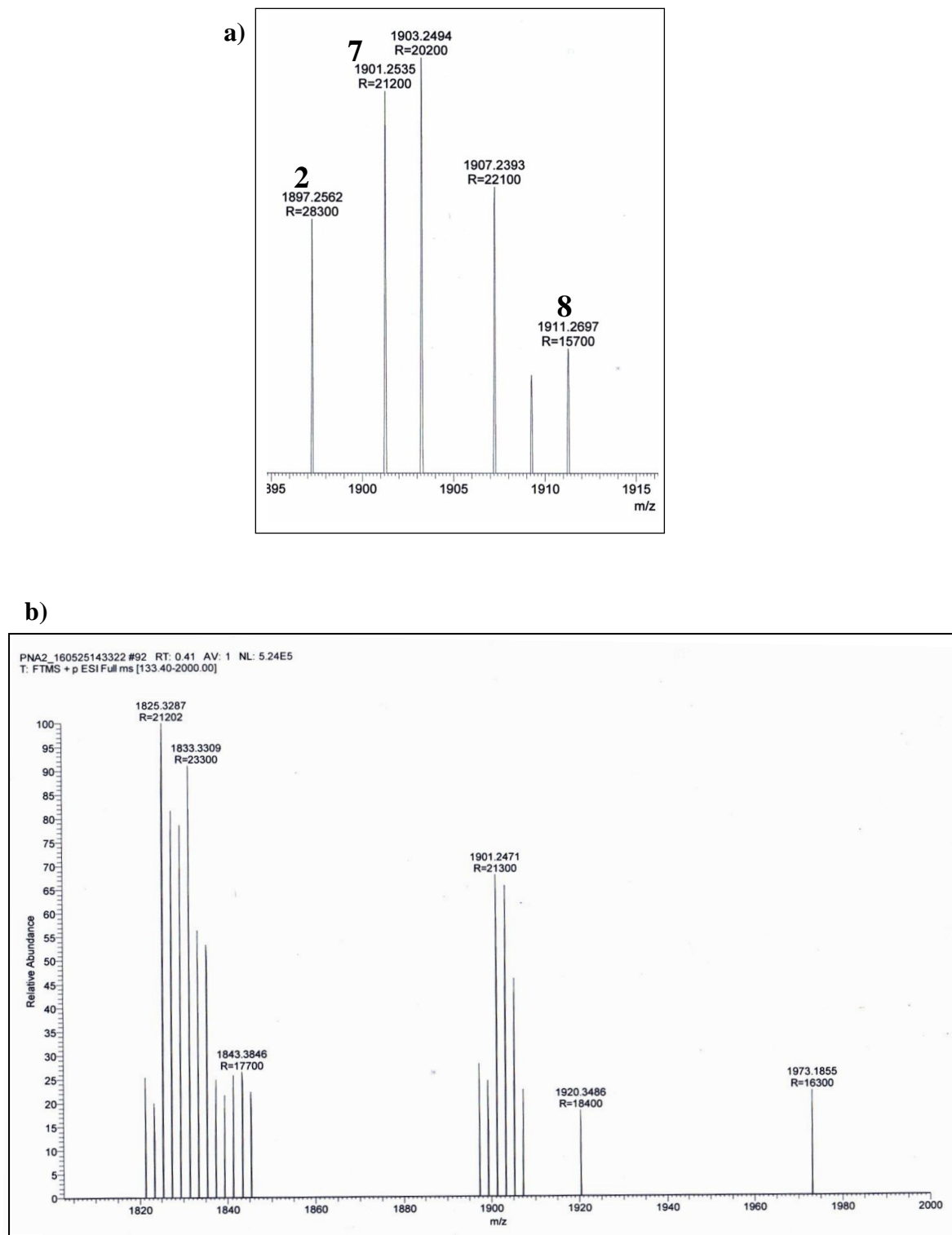
Figure 4.27: ESI-MS chromatogram of PNA-L-Ala a) zoomed region of the required peaks b) complete mass spectrum
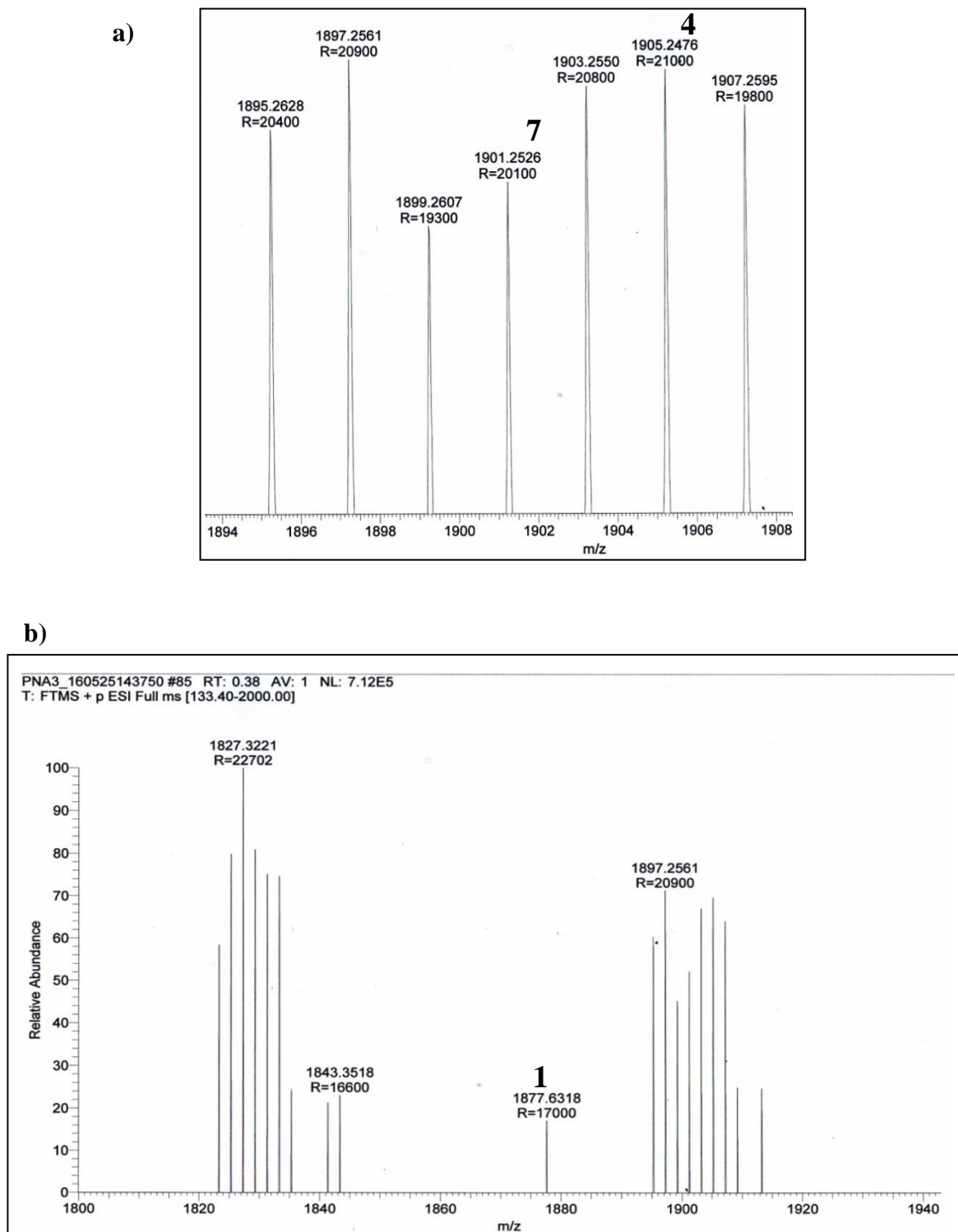
Figure 4.28: ESI-MS of PNA-D-Ala a) zoomed region of the required peaks b) complete mass spectrum
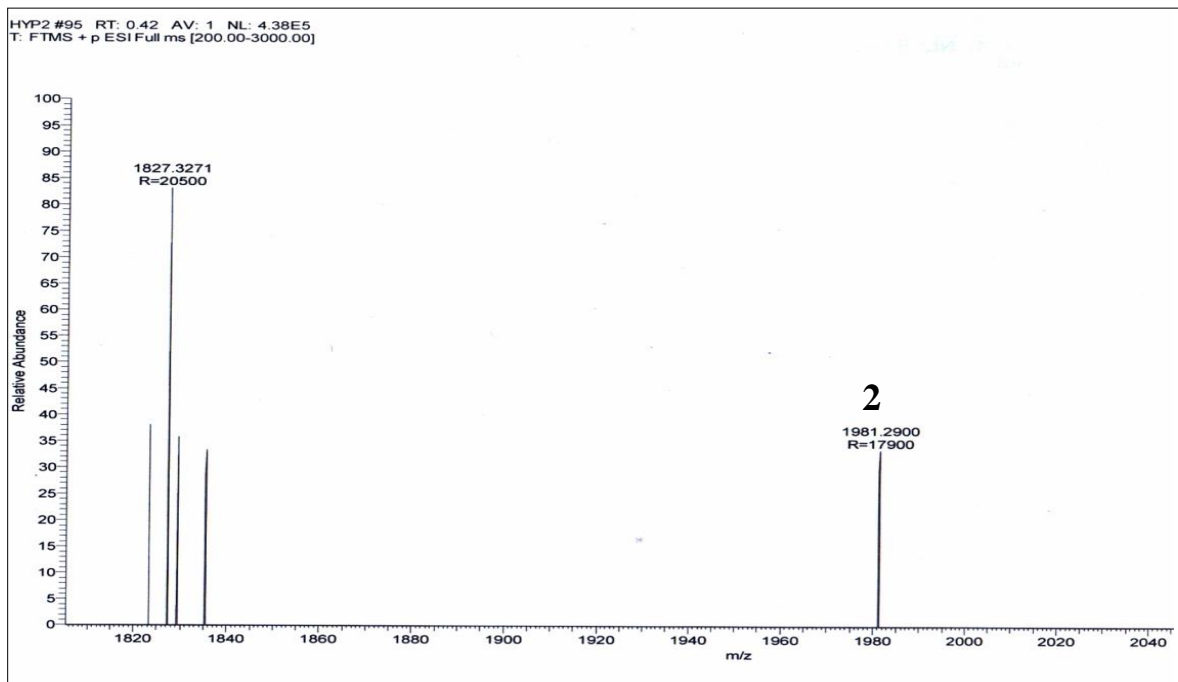
Figure 4.29: ESI-MS of PNA-D-Hyp
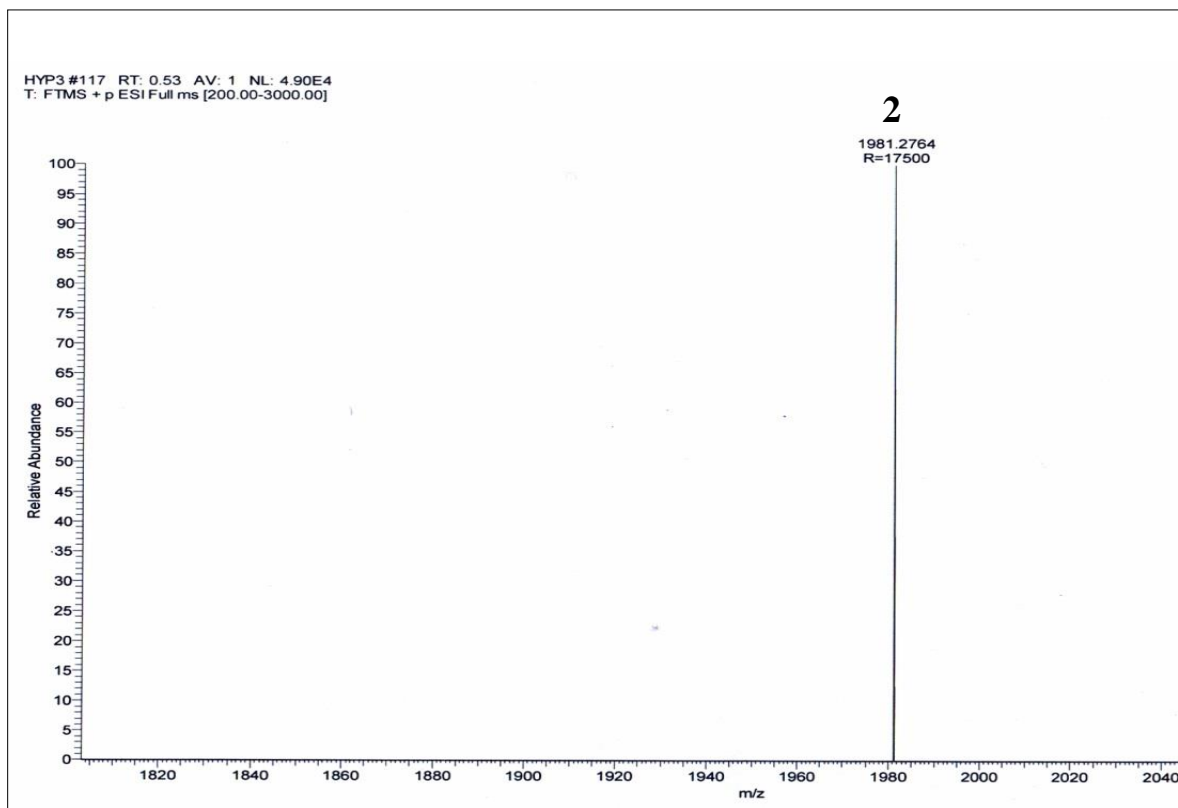
Figure 4.30: ESI-MS of PNA-(L+D)-Hyp
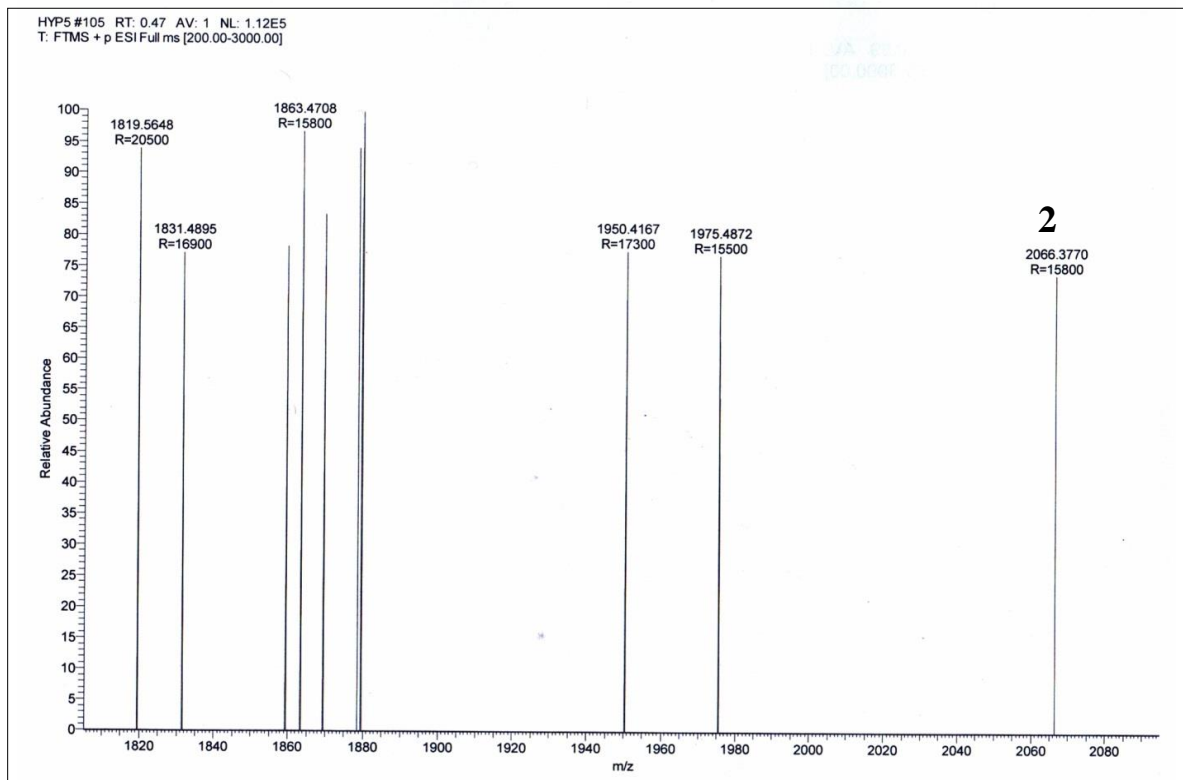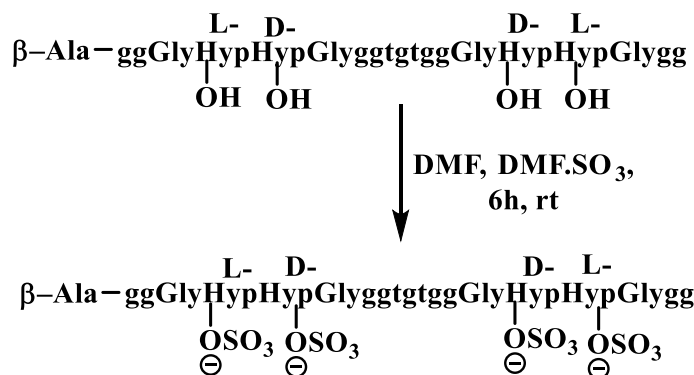
Figure 4.31: ESI-MS of PNA-Val

4.15 Sulphation of hydroxy groups

Thrombin is a serine protease which is proteolytically cleaved from prothrombin in co-agulation cascade.[90] Thrombin molecules possess positively charged residues in the potential binding sites (exosite I and II) with negatively charged sulfated glycosylamine ligands.[91] Exosite I or anion binding exosite is known to bind to thrombin receptor,[92] heparin,[93] fibrinogen,[94] hirudin[95] and thrombin binding aptamer.[96] Anticoagulant and antithrombic activities are among the most widely studied properties of sulphated macromolecules. Hence sulphation of the hydroxy groups might enable the binding of modified oligomers to thrombin.

4.15.1 Synthesis of sulphated oligomers

The hydroxyl groups of hydroxyproline in PNA-peptide conjugates were sulphated. The oligomers PNA-L-Hyp, PNA-D-Hyp and PNA-(L+D)-Hyp was dissolved in dry DMF. To it DMF.SO$_3$ was added and the reaction mixture was stirred for 6h.[97] The pH of the so-

lution was adjusted to 5 with NaHCO$_3$ and the solvent was removed. The sulphated oligo-
mers were purified by HPLC and characterized by MALDI.

$$\text{L- D-} \qquad\qquad\qquad \text{D- L-}$$
$$\beta-\text{Ala}-\text{ggGlyHypHypGlyggtgtggGlyHypHypGlygg}$$
$$\text{OH OH} \qquad\qquad\quad \text{OH OH}$$

DMF, DMF.SO$_3$,
6h, rt

$$\text{L- D-} \qquad\qquad\qquad \text{D- L-}$$
$$\beta-\text{Ala}-\text{ggGlyHypHypGlyggtgtggGlyHypHypGlygg}$$
$$\underset{\ominus}{\text{OSO}_3} \; \underset{\ominus}{\text{OSO}_3} \qquad\qquad \underset{\ominus}{\text{OSO}_3} \; \underset{\ominus}{\text{OSO}_3}$$

Scheme 4.3: Synthesis of sulphated oligomers

Table 11: Synthesis and characterization of PNA-{L, D, (L+D)}-Hyp and PNA-Val sulphate oligomers

Seq code	Sequences	HPLC t$_R$(min)	Calc. mass	Obsvd. mass
PNA-L-Hyp sulphate	ggGlyL-Hyp(OSO$_3^-$)L-Hyp(OSO$_3^-$)Gly ggtgtggGlyL-Hyp(OSO$_3^-$) L-Hyp (OSO$_3^-$)Gly gg-β-Ala	10.8	4238.9	4264.82(M+Na)
PNA-D-Hyp sulphate	GGGlyD-Hyp(OSO$_3^-$)D-Hyp(OSO$_3^-$) GlyGGTGTGGGlyD-Hyp(OSO$_3^-$) D-Hyp (OSO$_3^-$) GlyGG-β-Ala	10.4	4238.9	4280.5(M+K)
PNA-(L+D)-Hyp sulphate	GGGlyL-Hyp(OSO$_3^-$)D-Hyp(OSO$_3^-$) GlyGGTGTGGGlyD-Hyp(OSO$_3^-$) L-Hyp(OSO$_3^-$)Gly GG-β- Ala	10.4	4238.9	4269.3(M+Na)
PNA-Val sulphate	GGValD-HypL-HypThreoGGTGTGG ThreoL-HypD-HypValGG-β- Ala	11.7	4565.4	4628.8(M+Na+K)

4.15.2 UV melting experiment of sulphated hydroxyproline modified oligomers

UV melting experiments were performed with the sulphated hydroxyproline modi-
fied PNA oligomers in 100mM KCl concentration. Absence of cooperative transition at
295nm indicated that they were unable to form quadruplex.

Table 12: UV melting experiment of PNA-{L, D, (L+D), Val} sulphate

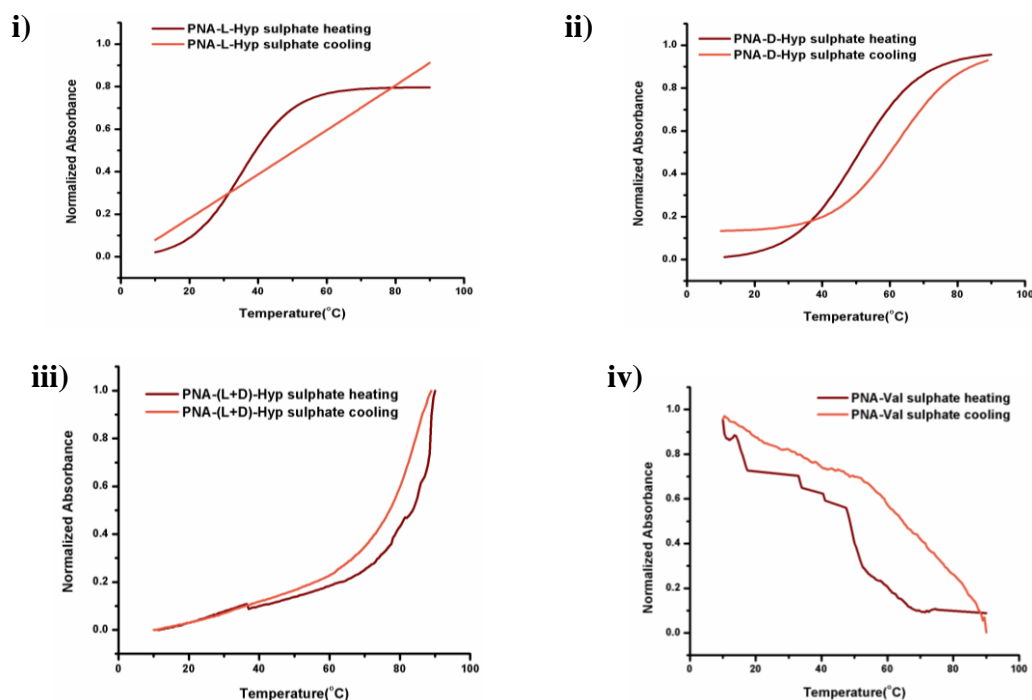| Sequences	T_m KCl at 295nm	
	Heat	**Cool**
PNA-L-Hyp sulphate	n.t.	n.t.
PNA-D-Hyp sulphate	n.t.	n.t.
PNA-(L+D)-Hyp sulphate	n.t.	n.t.
PNA-Val sulphate	n.t.	n.t.

Figure 4.32: UV melting and annealing curves of i) PNA-L-Hyp sulphate ii) PNA-D-Hyp sulphate iii) PNA-(L+D)-Hyp sulphate iv) PNA-Val sulphate

In each case of PNA-L-Hyp, PNA-D-Hyp, PNA-(L+D)-Hyp and PNA-Val, hyperchromicity was observed at 295nm. This indicates presence of other secondary structures in these sequences.

4.16 Conclusion

- In conclusion, 15mer PNA-TBA and 19mer modified PNA-TBA sequences were synthesized under microwave condition.

- The PNA-TBA as well as modified oligomers formed quadruplexes. However, while PNA-Gly, PNA-L-Ala and PNA-D-Ala formed multimolecular parallel quadruplex, PNA-L-Hyp, PNA-D-Hyp, PNA-(L+D)-Hyp, PNA-Val successfully formed unimolecular antiparallel quadruplex.

- The sulphated oligomers of PNA-L-Hyp, PNA-D-Hyp, PNA-(L+D)-Hyp, PNA-Val were unable to form quadruplex.

4.17 Experimental

1-(*tert*-butyl) 2-methyl (2*S*, 4*R*)-4-(benzyloxy) pyrrolidine-1, 2-dicarboxylate (2a)
1-(*tert*-butyl) 2-methyl (2*R*, 4*S*)-4-(benzyloxy) pyrrolidine-1, 2-dicarboxylate (2b)
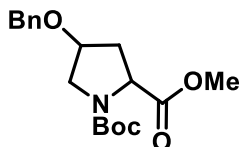
To a solution of **1** (0.5g, 2.04mmol) in CH_2Cl_2 (7mL), benzyl bromide (0.30 mL, 2.4mmol), silver oxide (1.4g, 6.1mmol) was added and the mixture was refluxed for 24h. The contents were filtered over celite, concentrated and chromatographed over silica gel (10% ethyl acetate in pet ether) to obtain **2a/2b** (1.2g, 89%) as colorless viscous liquid; $[\alpha]_D^{20}$ -18.8(c 1.0, $CHCl_3$) for **2a** and +18.0 (c 1.2, $CHCl_3$) for **2b**, ^{1}HNMR (200 MHz, CD_3OD): δ (ppm) 1.41(s, 9H), 2.00-2.13(m, 1H), 2.35-2.47(m, 1H),3.58-3.69 (m,2H), 3.734 (s, 3H), 4.16-4.19 (m, 1H),4.33-4.44(s, 1H), 4.52(s, 2H), 7.33 (s, 5H); ^{13}C NMR (50 MHz, CD_3OD): δ (ppm) 28.6, 37.54,50.35, 52.77, 59.62, 72.00, 77.62, 81.85, 128.84, 128.89, 129.50, 139.46, 156.05, 176.81; HRMS calcd for $C_{18}H_{25}O_5NNa$: 358.1625, Observed mass: 358.1618.

(2*S*, 4*R*)-4-(benzyloxy)-1-(*tert*-butoxycarbonyl) pyrrolidine-2-carboxylic acid (3a)
(2*R*, 4*S*)-4-(benzyloxy)-1-(*tert*-butoxycarbonyl) pyrrolidine-2-carboxylic acid (3b)
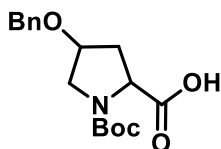
The compound **2a/2b** (0.4g, 1.2mmol) was dissolved in 1mL methanol and to it 2mL of 1N LiOH solution was added**.** The completion of the reaction was monitored by TLC. After 30mins, methanol was removed under vacuum and the aqueous layer was neutralized with Dowex H[+] resin. The resin was separated by filtration. The aqueous layer was washed with ethyl acetate (3 x 50mL) and the combined organic layer was lyophilized to get the free acid (0.31g, 81%). $[\alpha]_D^{20}$ -35.2(c 1.3, MeOH) for **3a**, +34.8(c 1.1, MeOH) for **3b**;^{1}HNMR(200 MHz, CD_3OD): δ (ppm) 1.43(s, 9H),1.99-2.12(m, 1H), 2.40-2.49(m, 1H),3.49-3.67 (m,2H), 4.20-

4.32 (m, 2H), 4.62(s, 2H), 7.32 (s, 5H); ^{13}C NMR (50 MHz, CD$_3$OD): δ (ppm) 28.6, 37.55, 52.77, 59.62, 72.01, 77.62, 81.85, 128.84, 128.89, 129.50, 139.46, 156.05, 176.81; HRMS calcd for C$_{17}$H$_{23}$O$_5$NNa: 344.1468, Observed mass: 344.1466.

(2*R*, 4*R*)-4-hydroxypyrrolidine-2-carboxylic acid (4)
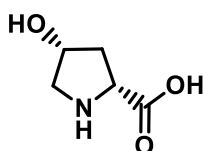
5g (20.4mmol) of *trans*- hydroxy L-proline was suspended in 7mL gla-cial acetic acid and treated with acetic anhydride (17.7mL, 102mmol). After 4h of reflux, solvent was removed under vacuo and dark brown syrup was obtained. The residue was taken up in water (2 x 100mL) and evaporated each time to syrup. The syrup was then dissolved in 7.5mL of 2N HCl and re-fluxed for 3h. The solvent was removed under vacuum and dissolved in water (100mL). The solution was treated with 90% absolute alcohol and a crystalline precipitate formed immediately. The procedure was repeated further twice to get pure product (3.8g, 76%). $[\alpha]_D^{20}$ +54.3 (c 1.1, H$_2$O); ^{1}HNMR (200 MHz, D$_2$O): δ (ppm) 2.16-2.28 (m, 1H), 2.31-2.51(m, 1H), 3.38-3.46 (m, 2H), 4.52-4.65 (m, 2H); ^{13}C NMR (50 MHz, D$_2$O): δ (ppm) 36.83, 53.33, 58.27, 69.56, 171.70; HRMS calcd for C$_5$H$_9$O$_3$NNa: 154.0475, Observed mass: 154.0475.

1-(*tert*-butyl) 2-methyl (2*R*, 4*R*)-4-hydroxypyrrolidine-1, 2-dicarboxylate (6)
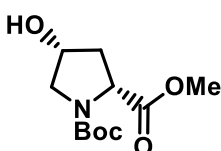
To a stirred slurry of *cis*-4-hydroxy-D-proline (**4**) (1g, 4.36mmol) in methanol (10mL) was added thionyl chloride (0.4mL, 4.8 mmol) at 0 $^\circ$C. After the completion of the addition, the resulting slurry was heated to 60°C and stirred for 6 h to give a clear light brown solution. The mix-ture was evaporated under vacuum to yield methyl ester as a light-brown solid (1.3g, 96%), which was used in the next step without further purification. Boc$_2$O (1.4mL, 7.58mmol) was added to a stirred suspension of crude methyl ester in CH$_2$Cl$_2$ (15mL). To the mixture triethylamine (2.1mL, 14.8mmol) was added dropwise at 0ºC. After completion of the start-ing material as observed by TLC analysis, solvent was removed under reduced pressure to get the crude compound. Purification of the crude over silica gel (30% EtOAc in pet ether) provided **6** as white solid (1.8 g, 99%); $[\alpha]_D^{20}$ +63.3 (c 1.5, CHCl$_3$); ^{1}HNMR (200 MHz, CDCl$_3$): δ (ppm) 1.41, 1.46 (9H, rotameric mixture), 2.02-2.13 (m, 1H), 2.22–2.36 (m, 1H),

3.43-3.62 (m, 2H), 3.74 (s, 3H), 4.36–4.50 (m, 2H); ^{13}C NMR (50 MHz, CDCl$_3$): δ (ppm) 28.22, 39.03, 52.04, 54.68, 57.87, 69.45, 80.37, 153.95, 173.63. HRMS calcd for C$_{11}$H$_{19}$O$_5$NNa: 268.1155, Observed mass: 268.1151.

1-(*tert*-butyl) 2-methyl (2*R*, 4*S*)-4-((4-nitrobenzoyl) oxy) pyrrolidine-1, 2-dicarboxylate (7)
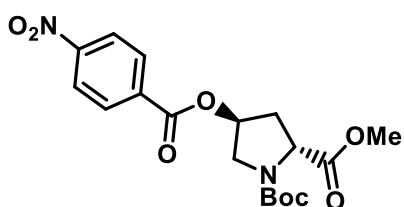
Ph$_3$P (0.19g, 0.73mmol) was dissolved in dry THF (20mL) at 0°C in a round bottomed flask. To it DIAD (0.16µL, 0.73mmol) was added dropwise resulting in formation of white solid. *p*-nitrobenzoic acid (0.1mg, 0.59mmol) was dissolved in dry THF (5mL) was added dropwise to the mixture maintaining the temperature to 0°C. Gradually clear solution was formed after addition. To this clear solution **6** (0.12g, 0.49mmol) dissolved in THF (5mL) was added dropwise. The reaction mixture was allowed to reach room temperature and stirred for another 6h. The solvent was removed under vacuum and the residue was extracted with water (3 x 25mL) and brine (2 x 20mL). The organic layer was concentrated and used for the next step without further purification.

1-(*tert*-butyl) 2-methyl (2*R*, 4*S*)-4-hydroxypyrrolidine-1, 2-dicarboxylate (8)
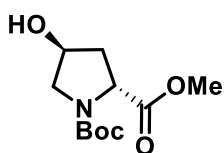
Compound **7** (0.3g, 0.81mmol) was dissolved in 1mL methanol. To this sodium azide (0.27g, 3.28mmol) was added and heated at 45°C for 24h. The solvent was removed under vacuum and the residue was dissolved in DCM and washed with water (3 x 10mL), brine (2 x 10mL). DCM was then removed under reduced pressure to get the crude compound. Purification of the crude compound over silica gel provided **8** as white solid (0.15g, 78% from 6 to 8). [α]$_D^{20}$ +54.3 (c 1.5, CHCl$_3$); ^{1}HNMR (200 MHz, CDCl$_3$): δ (ppm) 1.37, 1.42 (9H, rotameric mixture), 2.01(m, 1H), 2.25 (m, 1H), 3.52-3.56 (m, 2H), 3.70 (s, 3H), 4.36–4.43 (m, 2H); ^{13}C NMR (50 MHz, CDCl$_3$): δ (ppm) 28.22, 39.01, 52.05, 54.62, 57.87, 69.18, 80.30, 154.09, 173.75. HRMS calcd for C$_{11}$H$_{19}$O$_5$NNa: 268.1155, Observed mass: 268.1158.

4.17.1 Preparation of sample for UV-melting, thermal difference spectra and CD experiments

All samples were prepared at 5μM oligomer concentration in 10mM sodium or potassium phosphate buffer. The salt concentrations used are 100mM NaCl or KCl. The oligomer and salt concentration remains same for all the experiments unless otherwise mentioned.
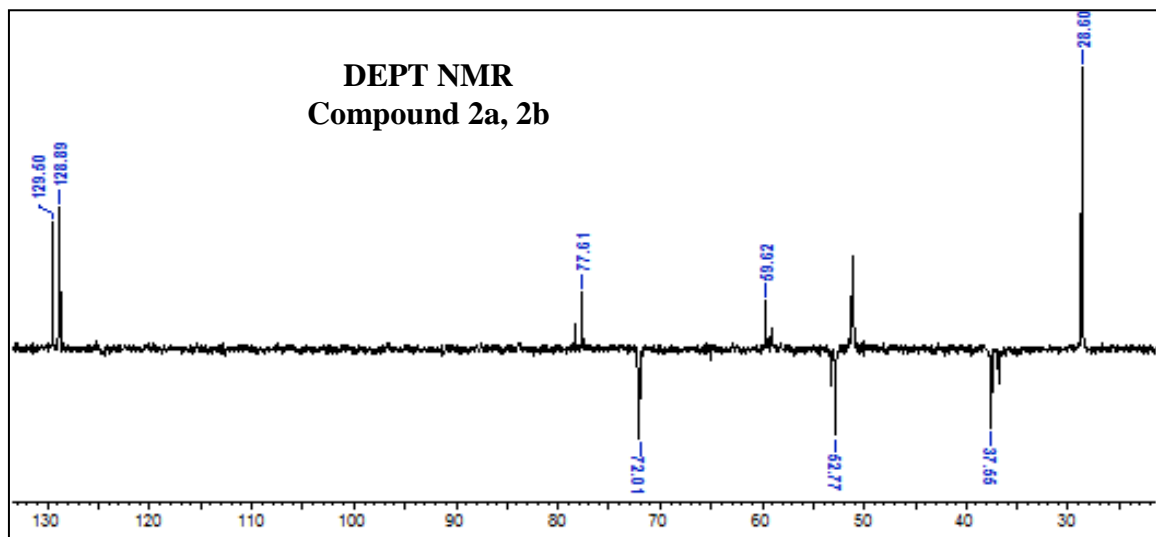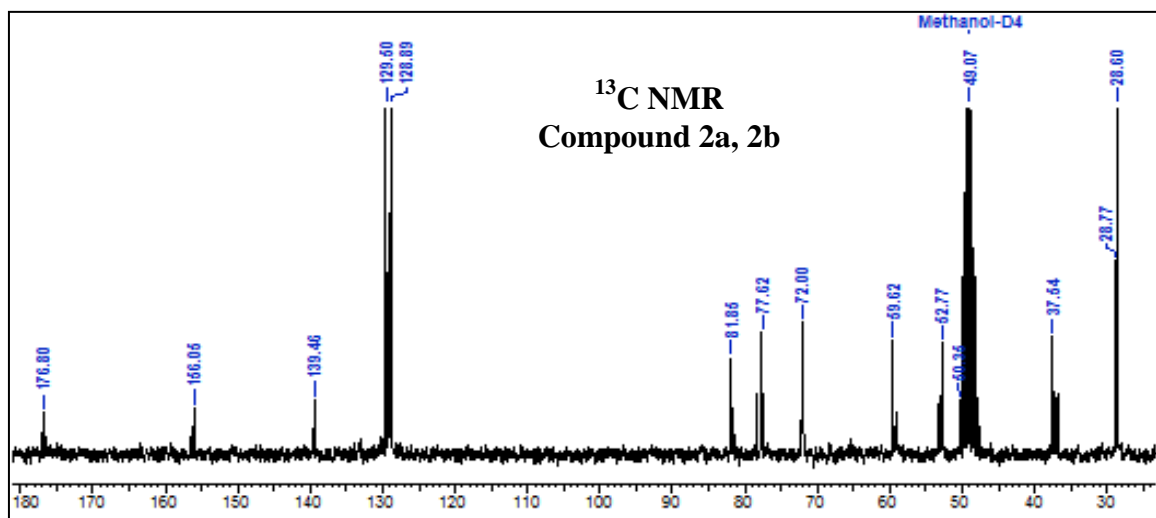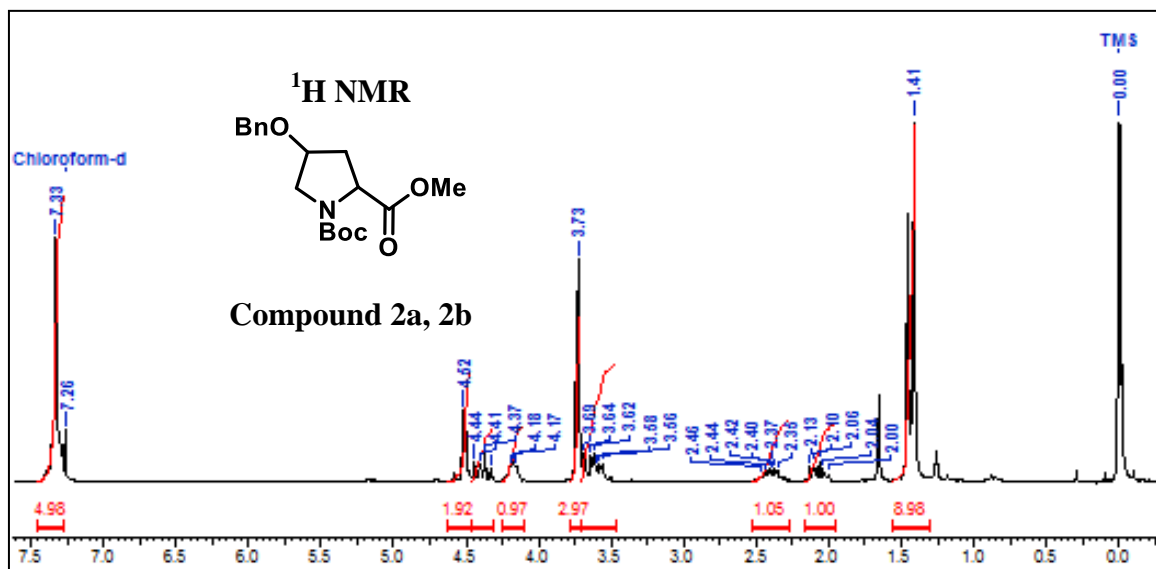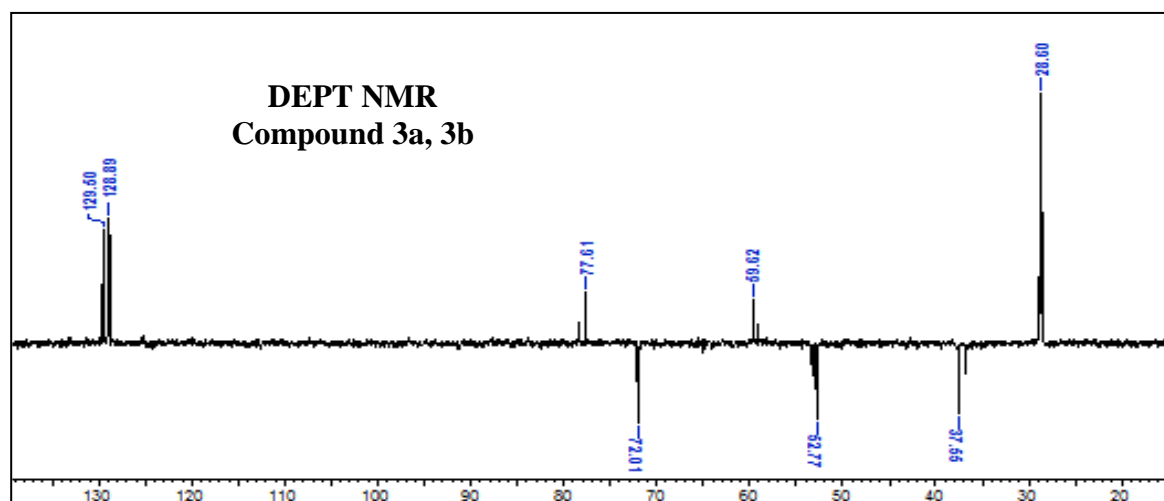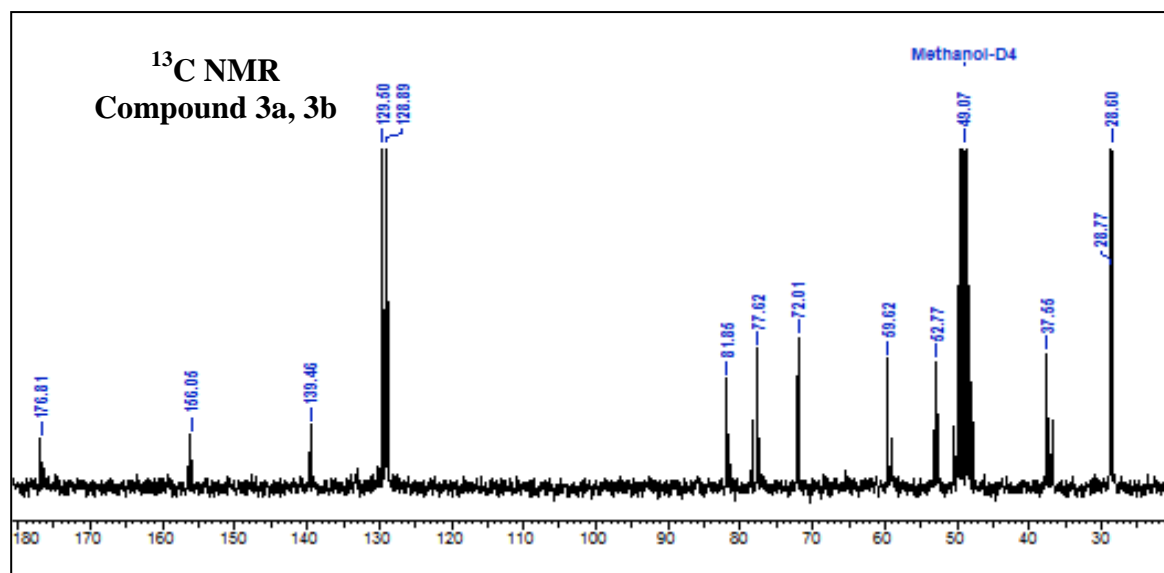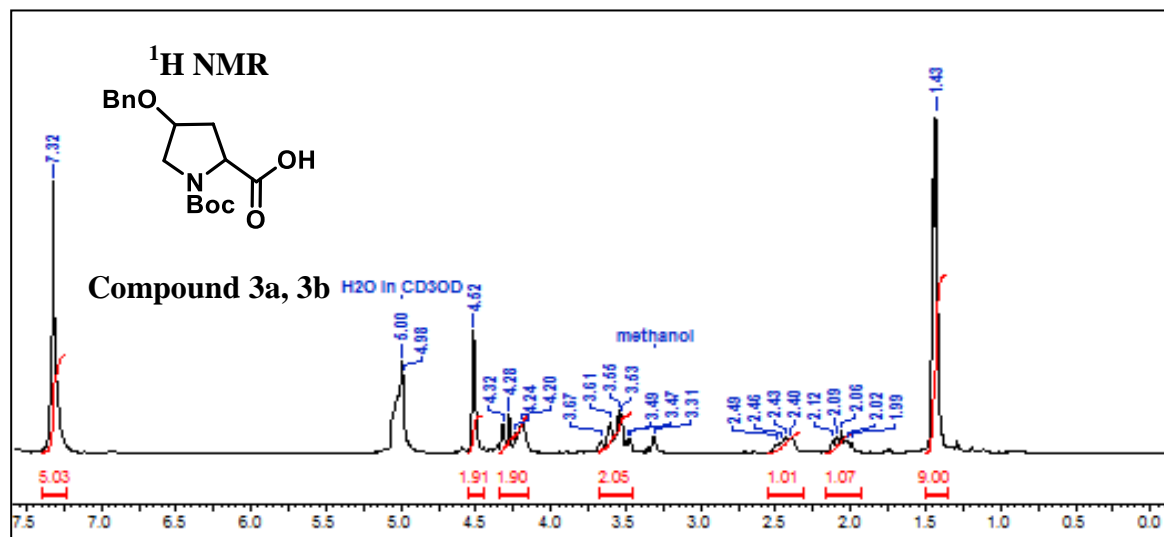
4.17.2 Preparation of sample for ESI-MS experiments

Samples containing 10μM PNA-TBA and modified PNA-TBA were prepared in water/methanol (1:1) with 500μM of KCl and 0.1% acetic acid. All experiments were performed under positive-ion ESI conditions. Typical conditions utilized a source voltage 4.2kV, capillary temperature of 320 °C, nitrogen sheath gas set at 35 arbitrary units, and 2-4 scans acquired (each consisting of 1 microscan). PNA-TBA and modified PNA-TBA samples were admitted through Hypersil Gold C18 column (2.1*50mm, 19μm particle size).
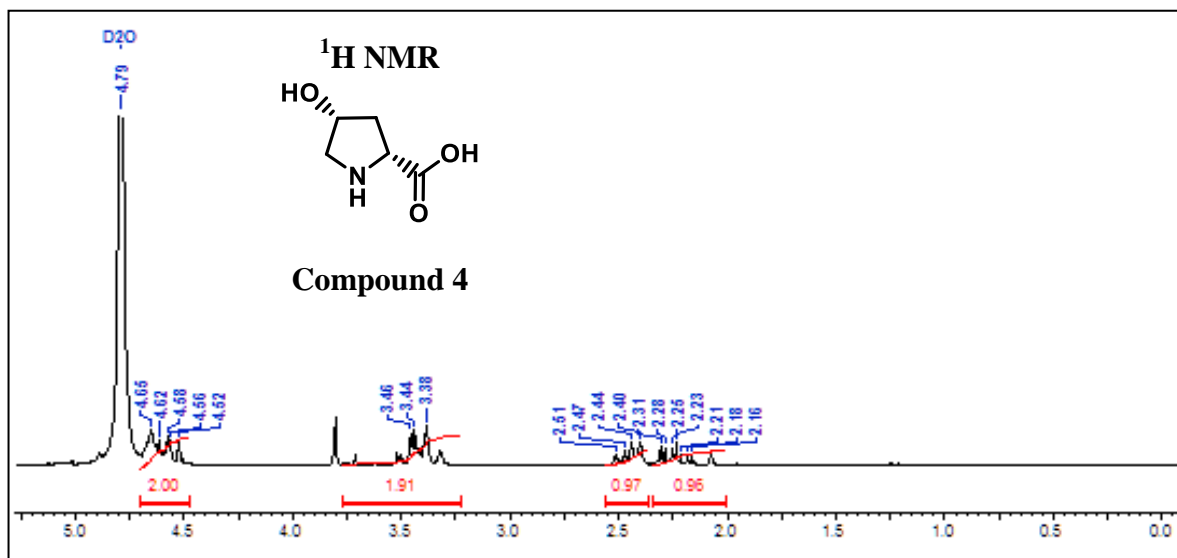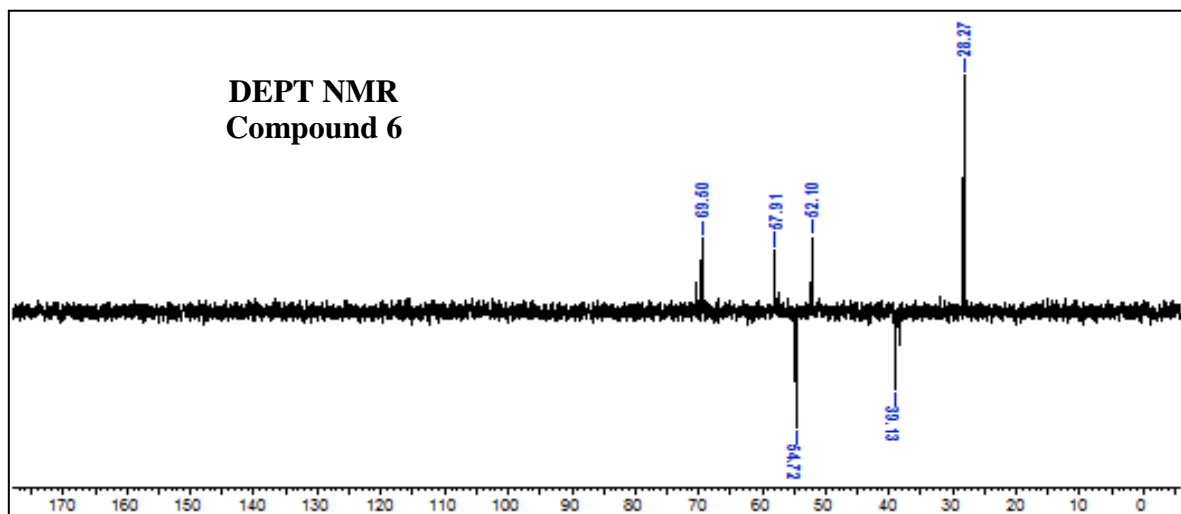
4.17.3 Appendix

...

^{1}H NMR

Compound 2a, 2b

^{13}C NMR
Compound 2a, 2b

DEPT NMR
Compound 2a, 2b

¹H NMR

Compound 3a, 3b

¹³C NMR
Compound 3a, 3b

DEPT NMR
Compound 3a, 3b

¹H NMR

Compound 4

D2O
4.79

4.65
4.62
4.58
4.56
4.52

3.46
3.44
3.38

2.51
2.47
2.44
2.40
2.31
2.28
2.25
2.23
2.21
2.18
2.16

2.00   1.91   0.97   0.96

5.0   4.5   4.0   3.5   3.0   2.5   2.0   1.5   1.0   0.5   0.0

---

¹³C NMR
Compound 4

171.70

69.56

58.27

53.33

36.83

170 160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10 0

DEPT NMR
Compound 4

53.30

36.80

58.23

69.64

90 85 80 75 70 65 60 55 50 45 40 35 30 25 20 15 10 5 0

¹H NMR

Compound 6

¹³C NMR
Compound 6

DEPT NMR
Compound 6

¹H NMR

Compound 8

**¹³C NMR
Compound 8**

**DEPT NMR
Compound 8**

Compound 2a, 2b
Expected Mass: 358.1625
Observed Mass: 358.1618

Compound 3a, 3b
Expected Mass: 344.1468
Observed Mass: 344.1466

T: FTMS + p ESI Full ms [100.00-1500.00]

154.0475
R=89102
$C_5 H_9 O_3 N Na = 154.0475$
-0.0619 ppm

Compound 4
Expected Mass: 154.0475
Observed Mass: 154.0475

153.9485
R=83500

154.1306
R=64000

Relative Abundance

m/z

GB-1 #97 RT: 0.43 AV: 1 NL: 6.51E8
T: FTMS + p ESI Full ms [100.00-1500.00]

268.1151
R=69007
$C_{11} H_{19} O_5 N Na = 268.1155$
-1.4958 ppm

Compound 6
Expected Mass = 268.1155
Observed Mass = 268.1151

Relative Abundance

230.0333
R=75007

248.0438
R=69807

284.0889
R=66707

301.1407
R=62402

330.0854
R=62102

359.6382
R=58307

387.6695
R=56407

m/z

T: FTMS + p ESI Full lock ms [100.00-700.00]

Compound 8
Expected Mass = 268.1155
Observed Mass = 268.1158

268.1158
R=66702
$C_{11}H_{19}O_5N\,Na = 268.1155$
0.2093 mmu

256.1512
R=66502

275.1641
R=63306

234.1339
R=69307

265.6615
R=66006

242.8902
R=53300

260.6511
R=61906

270.1608
R=60306

245.0513
R=63702

251.0004
R=59600

Relative Abundance

m/z

HPLC chromatogram of purified **PNA-TBA** oligomer
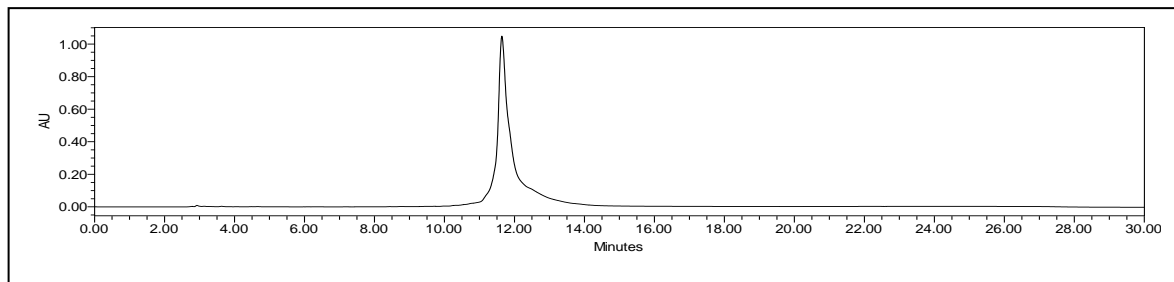
HPLC chromatogram of purified **modPNA-TBA** oligomer
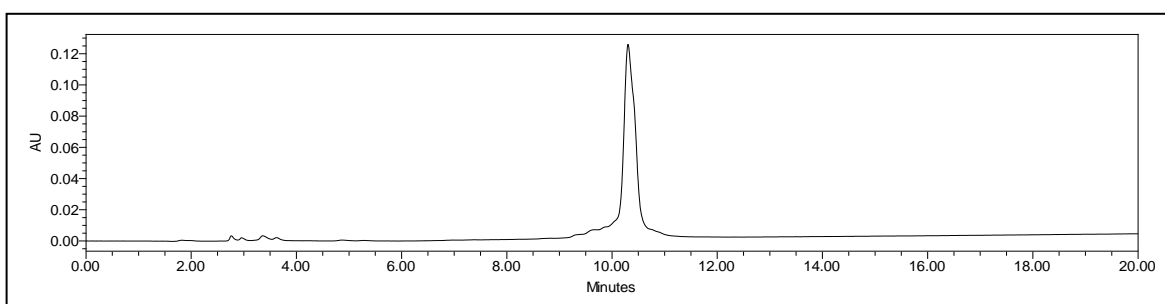
HPLC chromatogram of purified **PNA-Gly** oligomer
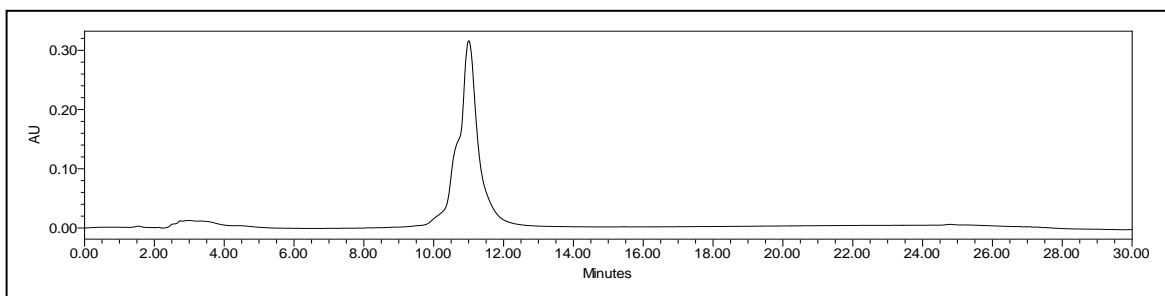
HPLC chromatogram of purified **PNA-L-Ala** oligomer
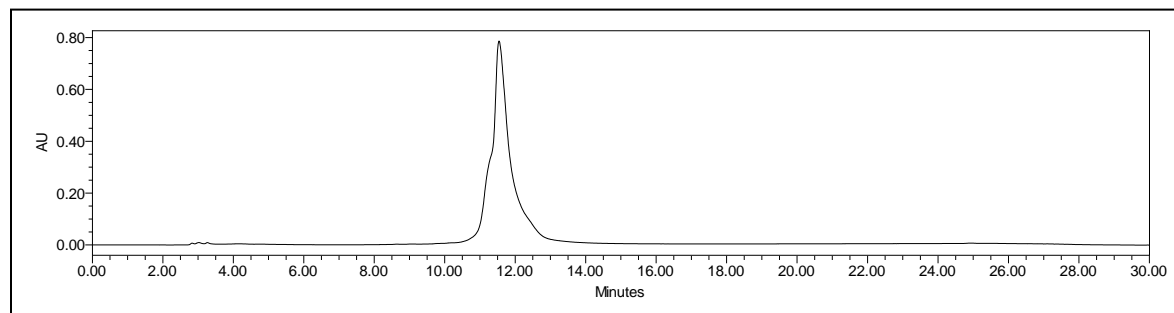
HPLC chromatogram of purified **PNA-D-Ala** oligomer
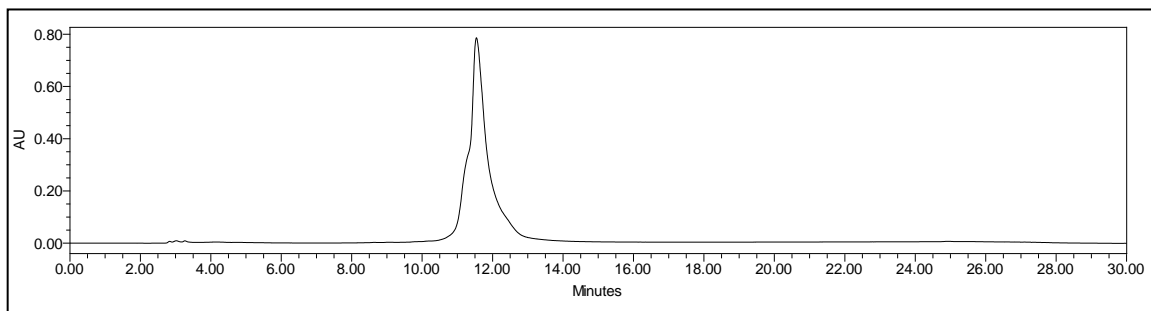
HPLC chromatogram of purified **PNA-L-Hyp** oligomer
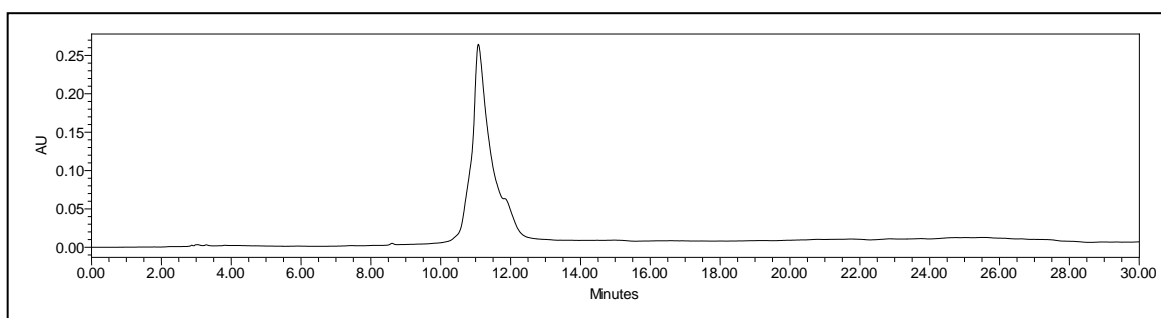
HPLC chromatogram of purified **PNA-D-Hyp** oligomer
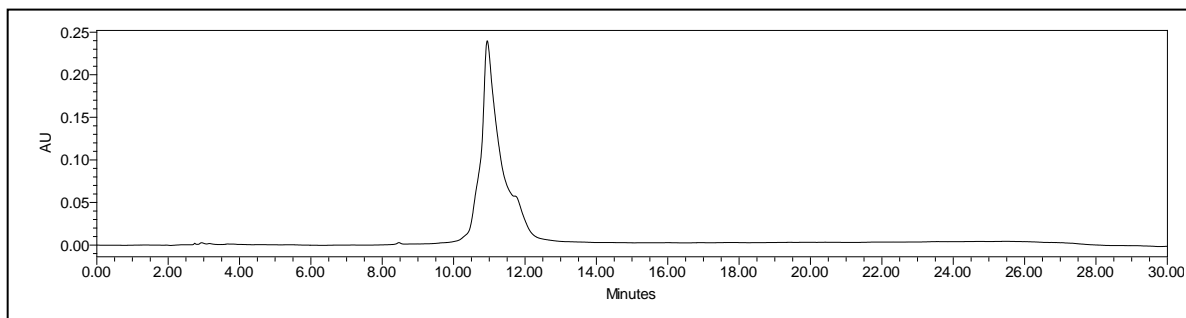
HPLC chromatogram of purified **PNA-(L+D)-Hyp** oligomer
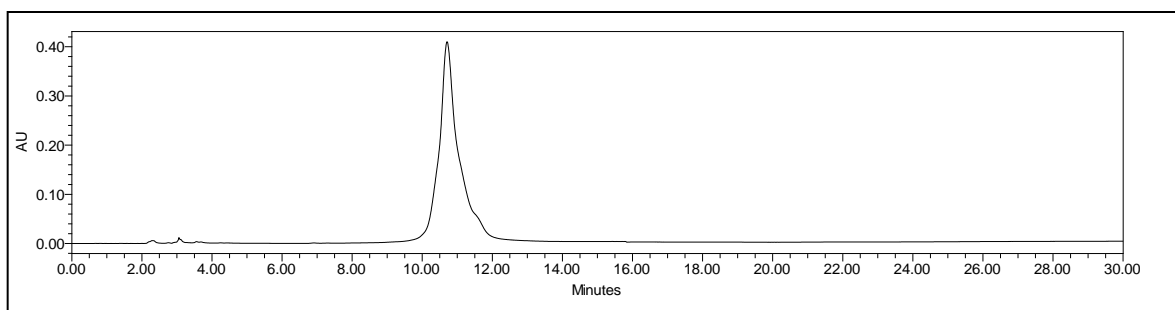
HPLC chromatogram of purified **PNA-Val** oligomer
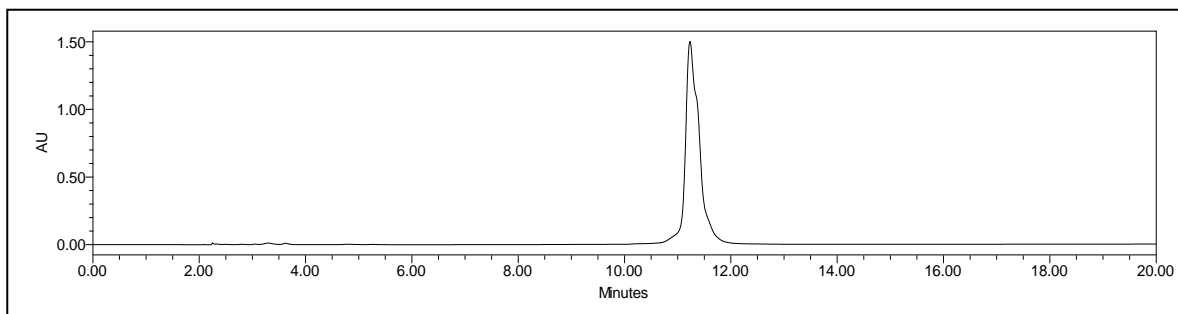
HPLC chromatogram of purified **PNA-L-Hyp sulphate** oligomer
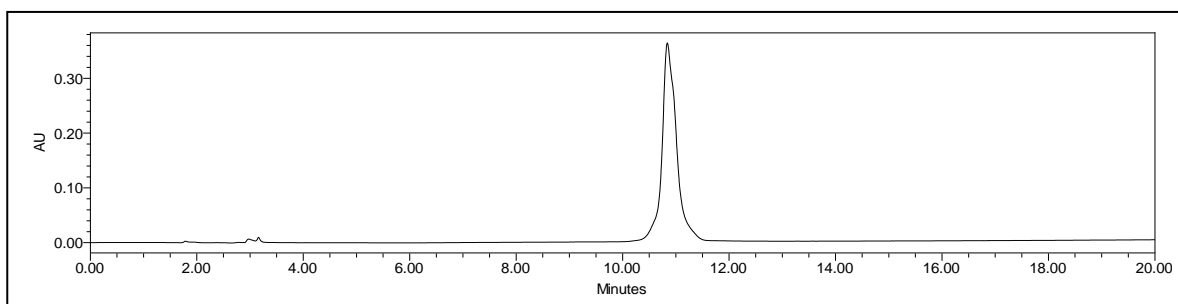
HPLC chromatogram of purified **PNA-D-Hyp sulphate** oligomer
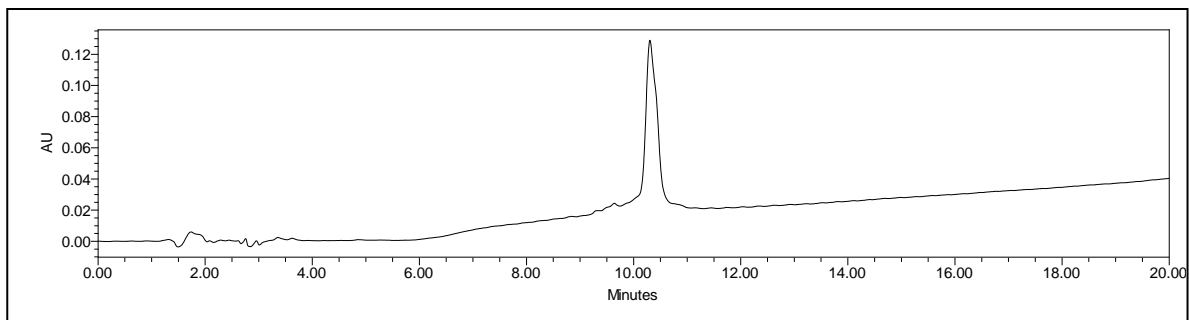
HPLC chromatogram of purified **PNA-(L+D)-Hyp sulphate** oligomer
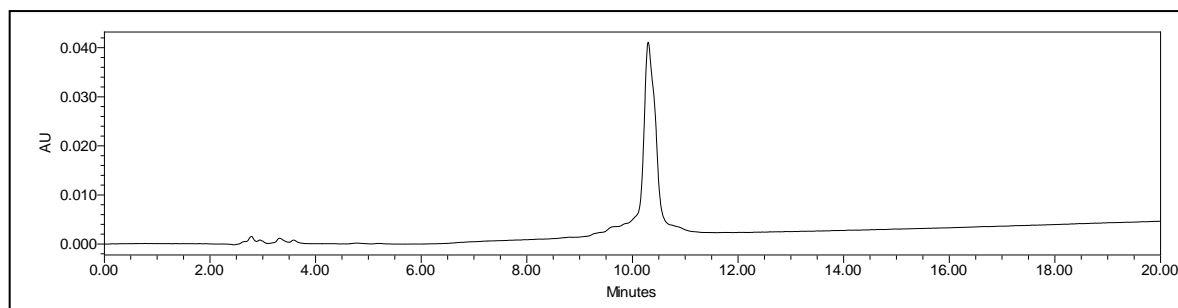
HPLC chromatogram of purified **PNA-Val sulphate** oligomer

TOF/TOF™ Linear Spec #1[BP = 4369.2, 4274]

PNA-TBA
Expected MW: 4364.17
Observed MW: 4369.2(M+Na)

TOF/TOF™ Linear Spec #1=>BC=>AdvBC(32,0.5,0.1)=>SM51=>BC=>AdvBC(32,0.5,0.1)[BP = 4548.8, 12]

modPNA-TBA
Expected MW: 4547.66
Observed MW: 4550.74

4550.7432

11.8

%Intensity

Mass (m/z)

TOF/TOF™ Linear Spec #1=>BC=>SM5[BP = 3693.3, 93]

PNA-Gly
Expected MW: 3696.4
Observed MW: 3693.8

3693.8389

92.7

3709.0361
3750.3040

%Intensity

Mass (m/z)

TOF/TOF™ Linear Spec #1=>BC=>SM5[BP = 3754.7, 43]

PNA-L-Ala
Expected MW: 3752.47
Observed MW: 3752.5

3754.5093

3811.6177

42.6

TOF/TOF™ Linear Spec #1=>BC=>AdvBC(32,0.5,0.1)=>SM5[BP = 3754.6, 25]

PNA-D-Ala
Expected MW: 3752.47
Observed MW: 3752.6

3754.6238

3812.3105

24.6

TOF/TOF™ Linear Spec #1=>BC=>AdvBC(32,0.5,0.1)=>SM5[BP = 3919.1, 40]

PNA-(L+D)-Hyp
Expected MW: 3920.5
Observed MW: 3919.5

TOF/TOF™ Linear Spec #1=>BC=>AdvBC(32,0.5,0.1)=>SM5=>SM5=>AdvBC(10,0.5,0.1)=>AdvBC(10,0.5,0.1)=>SM5[BP = 4092.6, 143]

PNA-Val
Expected MW: 4092.7
Observed MW: 4093.8

TOF/TOF™ Reflector Spec #1=>BC=>AdvBC(32,0.5,0.1)=>NF0.7[BP = 4264.8, 255]

PNA-L-Hyp sulphate
Expected MW: 4238.9
Observed MW: 4264.82(M+Na)

TOF/TOF™ Reflector Spec #1=>BC=>BC=>NF0.7[BP = 4064.0, 254]

PNA-D-Hyp sulphate
Expected MW: 4238.9
Observed MW: 4280.5(M+K)

TOF/TOF™ Reflector Spec #1=>BC=>AdvBC(32,0.5,0.1)=>NF0.7[BP = 4269.3, 80]

PNA-(L+D)-Hyp sulphate
Expected MW: 4238.9
Observed MW: 4269.3(M+Na)

4269.3

79.5

Mass (m/z)

TOF/TOF™ Linear Spec #1=>BC=>AdvBC(32,0.5,0.1)=>SM5[BP = 2376.1, 649]

4628.8687

PNA-Val sulphate
Expected MW: 4565.4
Observed MW: 4628.8(M+Na+K)

434.1

4553.0742

4571.7354

3087.5891

3619.7195

4444.7344

Mass (m/z)

4.18 References

1. a) J. E. Johnson, J. S. Smith, M. L. Kozak and F. B. Johnson, *Biochimie*, 2008, **90**, 1250. b) S. Burge, G. N. Parkinson, P. Hazel, A. K. Todd and S. Neidle, *Nucleic Acid Res.*, 2006, **34**, 5402. c) A. Siddiqui-Jain, C. L. Grand, D. J. Bearss and L. H. Hurley, *Proc. Natl. Acad. Sci. U S A*, 2002, **99**, 11593. d) J. L. Huppert and S. Balasubramanian, *Nucleic Acids Res*, 2005, **33**, 2908. e) Y. Xu, K. Kaminaga and M. Komiyama, *J. Am. Chem. Soc.*, 2008, **130**, 11179. f) Huppert, J.L., *Chem. Soc. Rev.,* 2009, **37**, 1375.

2. a) G. Biffi, D. Tannahill, J. Mc. Cafferty and S. Balasubramanian, *Nat. Chem.*, 2013, **5**, 182. b) E. Y. Lam, D. Beraldi, D. Tannahill and S. Balasubramanian, *Nat. Commun.*, 2013, **4**, 1796.

3. a) P. J. Bates, J. B. Kahlon, S. D. Thomas, J. O. Trent and D. M. Miller, *J Biol. Chem.*, 1999, **274**, 26369. b) T. Mashima, A. Matsugami, F. Nishikawa, S. Nishikawa and M. Katahira, *Nucleic Acids Res.*, 2009, **37**, 6249. c) J. L. Huppert, *Philos. Trans. A Math Phys. Eng. Sci.*, 2007, **365**, 2969-. d) S. Balasubramanian and S. Neidle, *Curr. Opin. Chem. Biol.*, 2009, **13**, 345. e) T. M. Ou, Y. J. Lu, J. H. Tan, Z. S. Huang, K. Y. Wong and L. Q. Gu, *Chem. Med. Chem.*, 2008, **3**, 690.
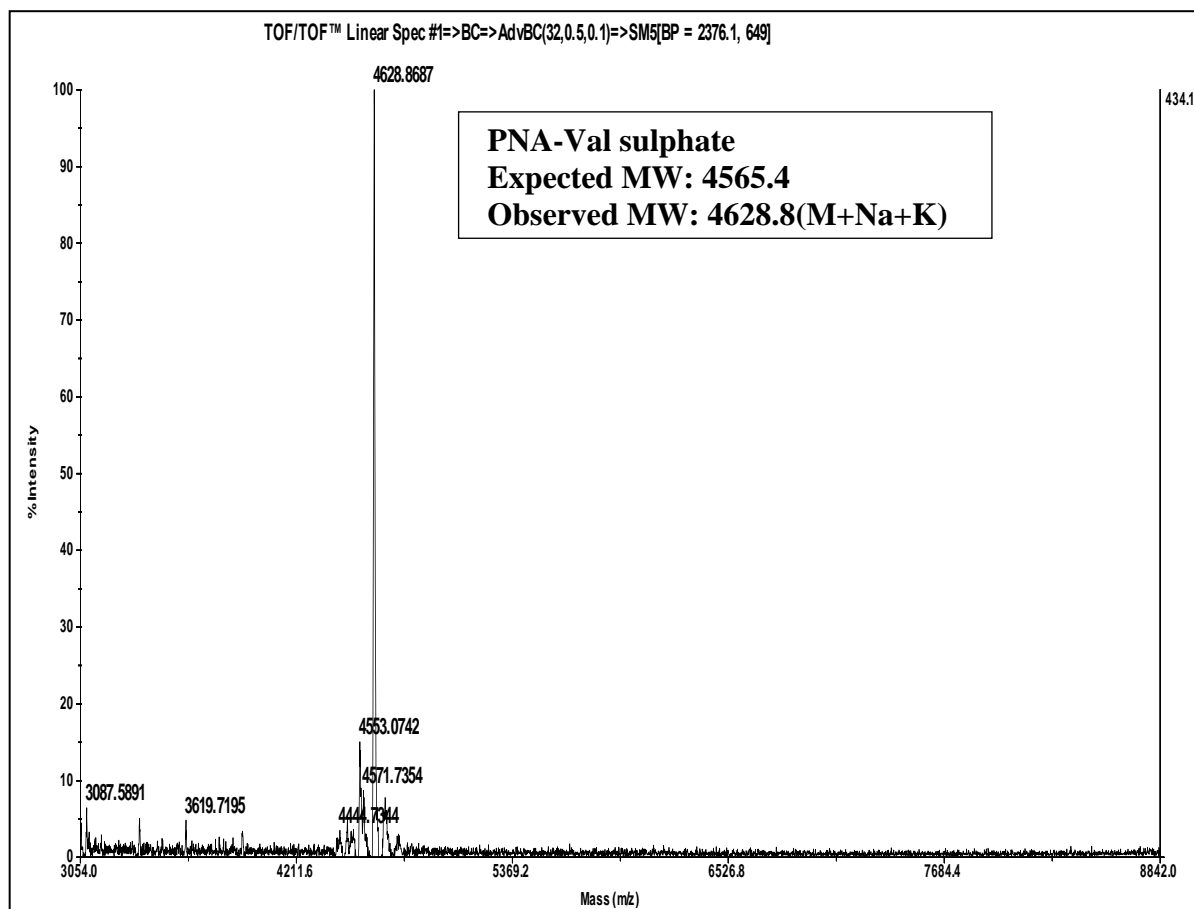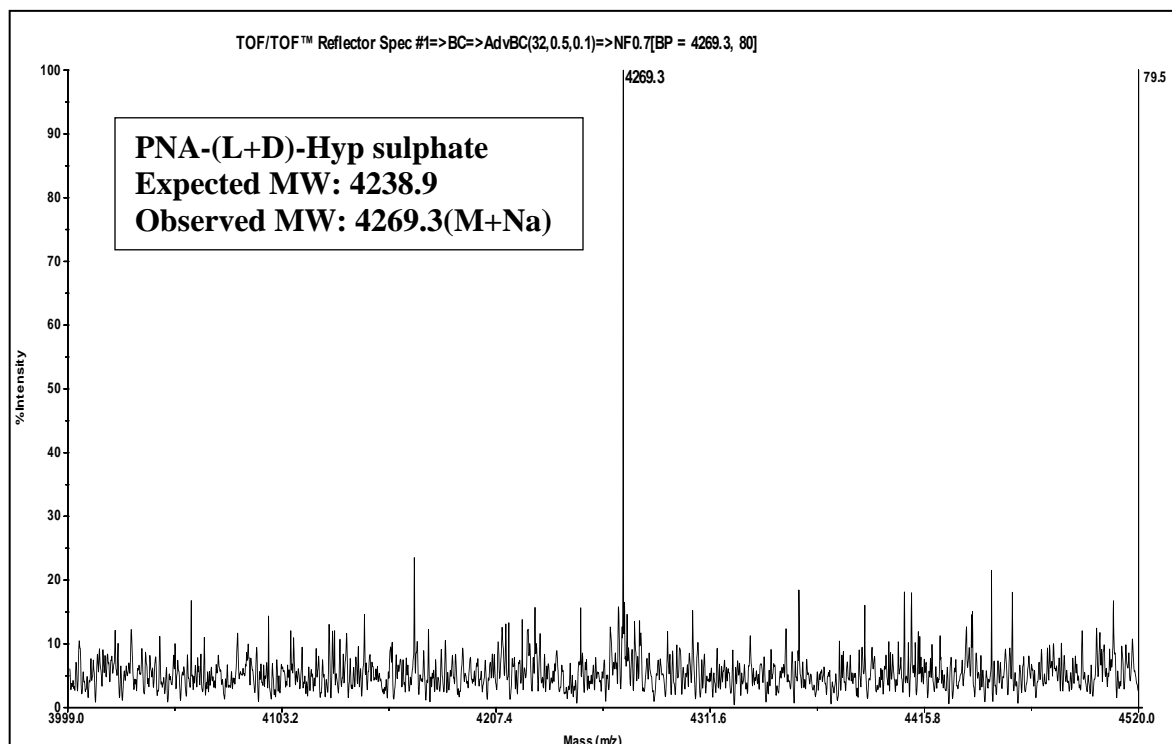
4. a) J. Hesselberth, M. P. Robertson, S. Jhaveri and A. D. Ellington, *J. Biotechnol.*, 2000, **74**, 15. b) A. A. Haller and P. Sarnow, *Proc Natl. Acad. Sci. U S A*, 1997, **94**, 8521. c) K. Gebhardt, A. Shokraei, E. Babaie and B. H. Lindqvist, *Biochemistry*, 2000, **39**, 7255. d) D. S. Wilson, A. D. Keefe and J. W. Szostak, *Proc. Natl. Acad. Sci. U S A*, 2001, **98**, 3750.e) J. Kawakami, H. Imanaka, Y. Yokota and N. Sugimoto, *J Inorg. Biochem.*, 2000, **82**, 197. f) E. N. Brody and L. Gold, *J. Biotechnol.*, 2000, **74**, 5.

5. L. C. Bock, L. C. Griffin, J. A. Latham, E. H. Vermaas and J. J. Toole, *Nature*, 1992, **355**, 564.

6. I. Smirnov and R. H. Shafer, *Biochemistry*, 2000, **39**, 1462.

7. A. N. Lane, J. B. Chaires, R. D. Gray and J. O. Trent, *Nucleic Acids Res.*, 2008, **36**, 5482.

8. A. E. Engelhart, J. Plavec, O. Persil and N.V. Hud, 2008, Nucleic Acid-Metal Ion Interactions, Royal Society of Chemistry, London.

9. a) B. I. Kankia and L. A. Marky, *J. Am. Chem. Soc.*, 2001, **123**, 10799. b) X. Mao, L. A. Marky and W. H. Gmeiner, *J. Biomol. Struct. Dyn.*, 2004, **22**, 25.c) X. A. Mao and W. H. Gmeiner, *Biophys. Chem.*, 2005, **113**, 155. d) I. Smirnov and R. H. Shafer, *J. Mol. Biol.*, 2000, **296**, 1. e) I. V. Smirnov, F. W. Kotch, I. J. Pickering, J. T. Davis and R. H. Shafer, *Biochemistry*, 2002, **41**, 12133. f) M. Vairamani and M. L. Gross, *J. Am. Chem. Soc.*, 2003, **125**, 42. g) V. M. Marathias, K. Y. Wang, S. Kumar, T. Q. Pham, S. Swaminathan and P. H. Bolton, *J. Mol. Biol.*, 1996, **260**, 378.

10. M. Trajkovski, P. Sket and J. Plavec, *Org. Biomol. Chem.*, 2009, **7**, 4677.

11. A. De Rache, I. Kejnovska, M. Vorlickova and C. Buess-Herman, *Chem. Eur. J.*, 2012, **18**, 4392.

12. V. C. Diculescu, A. M. Chiorcea-Paquim, R. Eritja and A. M. Oliveira-Brett, *J. Nucleic Acids*, 2010**, 84**, 1932.

13. R. J. Ellis, *Trends Biochem. Sci.*, 2001, **26**, 597.

14. Z. Y. Kan, Y. Yao, P. Wang, X. H. Li, Y. H. Hao and Z. Tan, *Angew. Chem. Int. Ed. Engl.*, 2006, **45**, 1629.

15. S. Nagatoishi, Y. Tanaka and K. Tsumoto, *Biochem. Biophys. Res. Commun.*, 2007, **352**, 812.

16. a) A. Arora, C. Balasubramanian, N. Kumar, S. Agrawal, R. P. Ojha and S. Maiti, *FEBS Lett.*, 2008, **275**, 3971. b) Z. Chen, K. W. Zheng, Y. H. Hao and Z. Tan, *J. Am. Chem. Soc.*, 2009, **131**, 10430.

17. T. Agarwal, D. Pradhan, I. Geci, A. M. El-Madani, M. Petersen, E. B. Pedersen and S. Maiti, *Nucleic Acid. Ther.*, 2012, **22**, 399.

18. S. M. Nimjee, C. P. Rusconi and B. A. Sullenger, *Annu. Rev. Med.*, 2005, **56**, 555.

19. a) W. James in *Aptamers* in *Encyclopedia of Analytical Chemistry* (Ed.: R. A. Meyers), Wiley, Chichester, UK, 2000, 4848.b) The Aptamer Handbook. Edited by S. Klussmann. (2006) WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim.

20. A. D. Ellington and J. W. Szostak, *Nature*, 1990, **346**, 818.

21. D. L. Robertson and J. F. Joyce, *Nature*, 1990, **344**, 467.

22. C. Tuerk and L. Gold, *Science*, 1990, **249**, 505.

23. a) S. Klussmann, A. Nolte, R. Bald, V. A. Erdmann and J. P. Furste, *Nat. Biotechnol.*, 1996, **14**, 1112-1115. b) A. Nolte, S. Klussmann, R. Bald, V. A. Erdmann and J. P. Furste, *Nat. Biotechnol.*, 1996, **14**, 1116.

24. A. Vater, S. Sell, P. Kaczmarek, C. Maasch, K. Buchner, E. Pruszynska-Oszmalek, P. Kolodziejski, W. G. Purschke, K. W. Nowak, M. Z. Strowski and S. Klussmann, *J. Biol. Chem.*, 2013, **288**, 21136.

25. S. R. Coughlin, *Nature,* 2000*,* **407**, 258.

26. L. C. Griffin, G. F. Tidmarsh, L. C. Bock, J. J. Toole and L. L. Leung, *Blood*, 1993, **81**, 3271.

27. A. De Anda, Jr., S. E. Coutre, M. R. Moon, C. M. Vial, L. C. Griffin, V. S. Law, M. Komeda, L. L. Leung and D. C. Miller, *Ann. Thorac. Surg.*, 1994, **58**, 344.

28. K. Y. Wang, S. McCurdy, R. G. Shea, S. Swaminathan and P. H. Bolton, *Biochemistry*, 1993, **32**, 1899.

29. P. Schultze, R. F. Macaya and J. Feigon, *J Mol Biol*, 1994, **235**, 1532.

30. R. F. Macaya, P. Schultze, F. W. Smith, J. A. Roe and J. Feigon, *Proc. Natl. Acad. Sci. U S A*, 1993, **90**, 3745.

31. A. Pasternak, F. J. Hernandez, L. M. Rasmussen, B. Vester and J. Wengel, *Nucleic Acids Res.*, 2010, **39**, 1155.

32. L. Bonifacio, F. C. Church and M. B. Jarstfer, *Int. J. Mol. Sci.*, 2008, **9**, 422.

33. A. Avino, S. Mazzini, R. Ferreira, R. Gargallo, V. E. Marquez and R. Eritja, *Bioorg. Med. Chem.*, 2012, **20**, 4186.

34. M. Zaitseva, D. Kaluzhny, A. Shchyolkina, O. Borisova, I. Smirnov and G. Pozmogova, *Biophys Chem*, 2010, **146**, 1.

35. G. X. He, J. P. Williams, M. J. Postich, S. Swaminathan, R. G. Shea, T. Terhorst, V. S. Law, C. T. Mao, C. Sueoka, S. Coutre and N. Bischofberger, *J. Med. Chem.*, 1998, **41**, 4224.

36. a) J. P. Shaw, J. A. Fishback, K. C. Cundy and W. A. Lee, *Pharm. Res.*, 1995, **12**, 1937. b) Lee, W. A., Fishback, J. A., Shaw, J.-P., Bock, L. C., Griffin, L.C., Cundy, K. C. , *Pharm. Res.,* **1995**, *12*, 1943.

37. a) S. L. Loke, C. A. Stein, X. H. Zhang, K. Mori, M. Nakanishi, C. Subasinghe, J. S. Cohen and L. M. Neckers, *Proc. Natl. Acad. Sci. U S A*, 1989, **86**, 3474. b) L. A. Yakubov, E. A. Deeva, V. F. Zarytova, E. M. Ivanova, A. S. Ryte, L. V. Yurchenko and V. V. Vlassov, *Proc. Natl. Acad. Sci. U.S.A.*, 1989, **86**, 6454.

38. G. F. Beck, W. J. Irwin, P. L. Nicklin and S. Akhtar, *Pharm. Res.*, 1996, **13**, 1028.

39. S. Mendelboum Raviv, A. Horvath, J. Aradi, Z. Bagoly, F. Fazakas, Z. Batta, L. Muszbek and J. Harsfalvi, *J. Thromb. Haemost.*, 2008, **6**, 1764.

40. S. R. Nallagatla, B. Heuberger, A. Haque and C. Switzer, *J. Comb. Chem.*, 2009, **11**, 364.

41. V. B. Tsvetkov, A. M. Varizhuk, G. E. Pozmogova, I. P. Smirnov,N. A. Kolganova and E. N. Timofeev, *Sci. Rep.*, 2015, **5**, 16337.

42. a) L. Martino, A. Virno, A. Randazzo, A. Virgilio, V. Esposito, C. Giancola, M. Bucci, G. Cirino and L. Mayol, *Nucleic. Acids. Res.*, 2006, **34**, 6653.b) B. Pagano, L. Martino, A. Randazzo and C. Giancola, *Biophys. J.*, 2008, **94**, 562.

43. A. Joachimi, A. Benz and J. S. Hartig, *Bioorg. Med. Chem.*, 2009, **17**, 6811.

44. C. F. Tang and R. H. Shafer, *J. Am. Chem. Soc.*, 2006, **128**, 5966.

45. M. L. Jeter, L. V. Ly, Y. M. Fortenberry, H. C. Whinna, R. R. White, C. P. Rusconi, B. A. Sullenger and F. C. Church, *FEBS Lett.*, 2004, **568**, 10.

46. R. White, C. Rusconi, E. Scardino, A. Wolberg, J. Lawson, M. Hoffman and B. Sullenger, *Mol. Ther.*, 2001, **4**, 567.

47. H. Yu, S. Zhang and J. C. Chaput, *Nat. Chem.*, 2012, **4**, 183.

48. A. D. Gunjal, M. Fernandes, N. Erande, P. R. Rajamohanan and V. A. Kumar, *Chem. Commun.*, **50**, 605.

49. [49] M. Scuotto, M. Persico, M. Bucci, V. Vellecco, N. Borbone, E. Morelli, G. Oliviero, E. Novellino, G. Piccialli, G. Cirino, M. Varra, C. Fattorusso and L. Mayol, *Org. Biomol. Chem.*, 2014, **12**, 5235.

50. K. Petkau-Milroy and L. Brunsveld, *Org. Biomol. Chem.*, 2013, **11**, 219.

51. C. K. McLaughlin, G. D. Hamblin and H. F. Sleiman, *Chem. Soc. Rev.*, 2011, **40**, 5647.

52. Q. Luo, Z. Dong, C. Hou and J. Liu, *Chem. Commun.*, 2014, **50**, 9997.

53. J. D. Watson and F. H. C. Crick, *Nature*, 1953, **171**, 737.

54. F. Xuan, T. W. Fan and I. M. Hsing, *ACS Nano*, 2015, **9**, 5027.

55. R. Volkmer, V. Tapia and C. Landgraf, *FEBS Lett*, 2012, **586**, 2780.

56. T. Simonsson, *Biol. Chem.*, 2001, **382**, 621.

57. B. Datta, M. E. Bier, S. Roy and B. A. Armitage, *J Am Chem Soc*, 2005, **127**, 4199.

58. Y. Krishnan-Ghosh, E. Stephens and S. Balasubramanian, *J. Am. Chem. Soc.*, 2004, **126**, 5944.

59. Y. Krishnan-Ghosh, A. M. Whitney and S. Balasubramanian, *Chem. Commun.*, 2005, 3068.

60. B. Datta, C. Schmitt and B. A. Armitage, *J. Am. Chem. Soc.*, 2003, **125**, 4111.

61. S. Lusvarghi, C. T. Murphy, S. Roy, F. A. Tanious, I. Sacui, W. D. Wilson, D. H. Ly and B. A. Armitage, *J. Am. Chem. Soc.*, 2009, **131**, 18415.

62. K. Padmanabhan, K. P. Padmanabhan, J. D. Ferrara, J. E. Sadler and A. Tulinsky, *J Biol. Chem.*, 1993, **268**, 17651. b) J. A. Kelly, J. Feigon and T. O. Yeates, *J. Mol. Biol.*, 1996, **256**, 417. c) K. Padmanabhan and A. Tulinsky, *Acta. Crystallogr. D. Biol. Crystallogr.*, 1996, **52**, 272. d) I. Russo Krauss, A. Merlino, A. Randazzo, E. Novellino, L. Mazzarella and F. Sica, *Nucleic Acids Res.*, 2012, **40**, 8119.

63. T. Bose, A. Banerjee, S. Nahar, S. Maiti and V. A. Kumar, *Chem. Commun.*, 2015, **51**, 7693.

64. J. L. Mergny, A. T. Phan and L. Lacroix, *FEBS Lett.*, 1998, **435**, 74.

65. J. L. Mergny, A. De Cian, A. Ghelab, B. Sacca and L. Lacroix, *Nucleic Acids Res.*, 2005, **33**, 81.

66. A. Chakrabartty, J. A. Schellman and R. L. Baldwin, *Nature*, 1991, **351**, 586.

67. a) B. L. Sibanda, T. L. Blundell and J. M. Thornton, *J. Mol. Biol.*, 1989, **206**, 759. b) J. E. Bell and T. E. Bell, *Proteins and Enzymes*; Prentice Hall: Englewood Cliffs, NJ, 1988. c) D. J. Barlow and J. M. Thornton, *J. Mol. Biol*, 1988, **201**, 601. d) G. D. Rose, L. M. Gierasch and J. A. Smith, *Adv. Protein Chem*, 1985, **37**, 1.

68. G. Srinivasulu, M.H.V. Ramana Rao, S. K. Kumar and A. C. Kunwar, *ARKIVOC*, 2004, **8,** 69.

69. a) R. Improta, F. Mele, O. Crescenzi, C. Benzi and V. Barone, *J. Am. Chem. Soc.*, 2002, **124**, 7857. b) M. L. DeRider, S. J. Wilkens, M. J. Waddell, L. E. Bretscher, F. Weinhold, R. T. Raines and J. L. Markley, *J. Am. Chem. Soc.*, 2002, **124**, 2497. c) R. Improta, C. Benzi and V. Barone, *J. Am. Chem. Soc.*, 2001, **123**, 12568.

70. J. P. Schneider, D. J. Pochan, B. Ozbas, K. Rajagopal, L. Pakstis and J. Kretsinger, *J. Am. Chem. Soc.*, 2002, **124**, 15030.

71. R. Singh and G. Panda, *RSC Advances*, 2013, **3**, 19533.

72. PCT Int. Appl., 2016014982, 28 Jan 2016.

73. B. A. Ellsworth, E. A. Jurica, J. Shi, W. R. Ewing, X. Ye, X. Wu, Y. Zhu, C. Sun, *PCT Int. Appl.*, 2014078609, 22 May 2014.

74. V. Vanek, M. Budesınsky, M. Rinnova and I. Rosenberg, *Tetrahedron*, 2009, **65**, 862.

75. J. A. Gomez-Vidal, M. T. Forrester and R. B. Silverman, *Org. Lett.*, 2001, **3**, 2477.

76. A. K. Pandey, D. Naduthambi, K. M. Thomas and N. J. Zondlo, *J. Am. Chem. Soc.*, 2013, **135**, 4333.

77. S. Paramasivan, I. Rujan and P. H. Bolton, *Methods*, 2007, **43**, 324.

78. P. Balagurumoorthy, S. K. Brahmachari, D. Mohanty, M. Bansal and V. Sasisekharan, *Nucleic Acids Res.*, 1992, **20**, 4061.

79. D. M. Gray, *Biopolymers*, 1974, **13**, 2087.

80. M. Lu, Q. Guo and N. R. Kallenbach, *Biochemistry*, 1993, **32**, 598.

81. L. C. Bock, L. C. Griffin, J. A. Latham, E. H. Vermaas and J. J. Toole, *Nature*, 1992, **355**, 564.

82. J. R. Williamson, *Annu. Rev. Biophys. Biomol. Struct.*, 1994, **23**, 703.

83. A. I. Karsisiotis, N. M. Hessari, E. Novellino, G. P. Spada, A. Randazzo and M. Webba da Silva, *Angew. Chem. Int. Ed. Engl.*, 2011, **50**, 10645.

84. M. Vairamani and M. L. Gross, *J. Am. Chem. Soc.*, 2003, **125**, 42.

85. W. M. David, J. Brodbelt, S. M. Kerwin and P. W. Thomas, *Anal. Chem.*, 2002, **74**, 2029.

86. F. Rosu, V. Gabelica, C. Houssier, P. Colson and E. D. Pauw, *Rapid Commun. Mass Spectrom.*, 2002, **16**, 1729.

87. B. Datta, M. E. Bier, S. Roy and B. A. Armitage, *J. Am. Chem. Soc.*, 2005, **127**, 4199.

88. H. Wendt, E. Durr, R. M. Thomas, M. Przybylski and H. R. Bosshard, *Protein Sci.*, 1995, **4**, 1563.

89. M. Przybylski and M. O. Glocker, *Angew. Chem. Int. Ed. Engl.*, 1996, **35**, 806.

90. N. J. Royle, D. M. Irwin, M. L. Koschinsky, R. T. MacGillivray and J. L. Hamerton, *Somat. Cell. Mol. Genet.*, 1987, **13**, 285.

91. W. Bode, D. Turk and A. Karshikov, *Protein Sci.*, 1992, **1**, 426.

92. L. W. Liu, T. K. Vu, C. T. Esmon and S. R. Coughlin, *J. Biol. Chem.*, 1991, **266**, 16977.

93. J. P. Sheehan, Q. Wu, D. M. Tollefsen and J. E. Sadler, *J. Biol. Chem.*, 1993, **268**, 3639.

94. J. W. Fenton, 2nd, T. A. Olson, M. P. Zabinski and G. D. Wilner, *Biochemistry*, 1988, **27**, 7106.

95. T. J. Rydel, K. G. Ravichandran, A. Tulinsky, W. Bode, R. Huber, C. Roitsch and J. W. Fenton, *Science*, 1990, **249**, 277.

96. Q. Wu, M. Tsiang and J. E. Sadler, *J. Biol. Chem.*, 1992, **267**, 24408.

97. C. Avitabile, L. Moggio, G. Malgieri, D. Capasso, S. Di Gaetano, M. Saviano, C. Pedone and A. Romanelli, *PLoS One*, 2012, **7**, e35774.