Incorporating modified bases and base-backbone linkers into PNAs for targeting RNA secondary structures

Ong, Alan Ann Lerk

2019

Ong, A. A. L. (2019). Incorporating modified bases and base-backbone linkers into PNAs for targeting RNA secondary structures. Doctoral thesis, Nanyang Technological University, Singapore.

<https://hdl.handle.net/10356/136984>

<https://doi.org/10.32657/10356/136984>

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0).

Downloaded on 26 Apr 2023 19:29:12 SGT

**NANYANG
TECHNOLOGICAL
UNIVERSITY**

SINGAPORE

**Incorporating modified bases and base-backbone
linkers into PNAs for targeting RNA secondary
structures**

ALAN ONG ANN LERK
Interdisciplinary Graduate School
NTU Institute of Health Technologies (HealthTech NTU)

**Incorporating modified bases and base-backbone
linkers into PNAs for targeting RNA secondary
structures**

ALAN ONG ANN LERK

**Interdisciplinary Graduate School
NTU Institute of Health Technologies (HealthTech NTU)**

A thesis submitted to the Nanyang Technological University in
partial fulfillment of the requirement for the degree of
Doctor of Philosophy

2019

Statement of Originality

I hereby certify that the work embodied in this thesis is the result of original research, is free of plagiarised materials, and has not been submitted for a higher degree to any other University or Institution.

12 July 2019

.....
Date

.....
Alan Ong Ann Lerk

Supervisor Declaration Statement

I have reviewed the content and presentation style of this thesis and declare it is free of plagiarism and of sufficient grammatical clarity to be examined. To the best of my knowledge, the research and writing are those of the candidate except as acknowledged in the Author Attribution Statement. I confirm that the investigations were conducted in accord with the ethics policies and integrity standards of Nanyang Technological University and that the research data are presented honestly and without prejudice.

12 July 2019

.....

Date

.....

Chen Gang

Authorship Attribution Statement

This thesis contains material from one paper published in the following peer-reviewed journal(s) where I was the first and/or corresponding author.

Chapter 2 is published as Ong, A. A. L., Toh, D.F.K., Patil, K.M., Meng, Z., Yuan, Z., Krishna, M.S., Devi, G. Haruehanroengra, P., Lu, Y., Xia, K., Okamura, K., Sheng, J., and Chen, G. . Recognition of U-G, U-A, and C-G pairs by double-stranded RNA-binding PNAs incorporated with an artificial nucleobase. *Biochemistry* **58**, 1319-1331 (2019). doi: 10.1021/acs.biochem.1028b01313.

The contributions of the co-authors are as follows:

- Dr Chen directed the project, analyzed the data, and prepared the manuscript.
- I performed laboratory work including synthesis of monomers and oligomers, characterization of monomers and oligomers and binding studies. I also analyzed the data and prepared the manuscript.
- Dr Patil assisted in the synthesis of the intermediates of the final monomers.
- Ms. Toh assisted in obtaining the binding data, analyzing the binding data, and preparing the manuscript.
- Dr. Meng, Dr Lu, and Dr Xia did the modeling studies.
- Dr. Yuan and Dr. Devi assisted in the early stage of the work.
- Mr. Krishna helped synthesize some of materials.
- Dr Okamura helped with the binding studies of the miRNA hairpin precursor.
- Ms Haruehanroengra and Dr Sheng provided an RNA hairpin sample

12 July 2019

.....
Date

.....
Alan Ong Ann Lerk

Abstract

The structure diversity of RNA allows RNA to be involved in various biological processes. RNA secondary structures include the single-stranded (ss) loops and double-stranded (ds) stems and are involved in protein binding and gene regulation. Peptide nucleic acid (PNA) is a synthetic analog of DNA and RNA with a chemically stable and neutral peptide-like backbone. Targeting RNA secondary structures shows great promise in RNA-targeting therapeutics. Traditional antisense PNAs (asPNAs) can be used to target ssRNA and weakly formed dsRNA regions. Short chemically-modified PNAs have been developed to recognize dsRNAs at near physiological conditions by major-groove PNA·RNA-RNA triplex formation. However, the development of dsRNA-binding PNAs (dbPNAs) as a class of dsRNA-targeting drug is hindered by complicated RNA secondary structures and canonical or non-canonical base pairings. Here, we utilize PNAs incorporating modified bases and base-backbone linkers to improve the recognition of model RNA secondary structures and disease-related RNA structures.

Project 1. Chemically-modified short PNAs recognize RNA duplexes at near physiological conditions by major-groove PNA·RNA-RNA triplex formation and show a great promise in developing RNA-targeting probes and therapeutics. Thymine (T) and uracil (U) are often incorporated into PNAs to recognize A-U pairs through major-groove T·A-U and U·A-U triplex formation. Incorporation of a modified nucleobase, 2-thiouracil (s^2U), into triplex-forming oligonucleotides (TFOs) stabilizes both DNA and RNA triplexes. Thiolation of uracil causes a decrease of the pKa of N3 nitrogen atom which may result in improved hydrogen bonding in addition to

enhanced base stacking interactions, similar to the previously reported thiolation effect of pseudoisocytosine (J to L substitution). Here, we incorporated s^2U into short PNAs, followed by binding studies of a series of s^2U -modified PNAs. We demonstrated by nondenaturing polyacrylamide gel electrophoresis (PAGE) and thermal melting experiments that s^2U and L incorporated into dbPNAs enhance the recognition of A-U and G-C pairs, respectively, in RNA duplexes in a position-independent manner, with no appreciable binding to DNA duplex. Combining s^2U and L modifications in dbPNAs facilitates enhanced and selective recognition of dsRNAs over ssRNAs. We further demonstrated through a cell-free assay the application of the s^2U - and L-modified dbPNAs in the inhibition of the pre-microRNA-198 maturation in a substrate specific manner. Thus, s^2U -modified dbPNAs may be generally useful for the enhanced and selective recognition of RNA duplexes and for the regulation of RNA functions.

Project 2. Triplex formation is favored for RNA duplexes with a purine tract within one of the RNA duplex strands, and is severely destabilized if the purine tract is interrupted by pyrimidine residues. Here, we report the synthesis of a PNA monomer incorporated with an artificial nucleobase S, followed by the binding studies of a series of S-modified PNAs. Our data suggest that an S residue incorporated into short 8-mer dbPNAs can recognize internal Watson–Crick C-G and U-A, and wobble U-G base pairs (but not G-C, A-U, and G-U pairs) in RNA duplexes. The short S-modified PNAs show no appreciable binding to DNA duplexes or single-stranded RNAs. Interestingly, replacement of the C residue in an S·C-G triple with a 5-methyl C results in the disruption of the triplex, probably due to a steric clash between S and 5-methyl C. Previously reported PNA E base shows recognition of U-A and A-U pairs,

but not a U-G pair. Thus, S-modified dbPNAs may be uniquely useful for the general recognition of RNA U-G, U-A, and C-G pairs. Shortening the succinyl linker of our PNA S monomer by one carbon atom to have a malonyl linker causes a severe destabilization of triplex formation. Our experimental and modeling data indicate that part of the succinyl moiety in a PNA S monomer may serve to expand the S base forming stacking interactions with adjacent PNA bases.

Project 3. Developing a general platform for the molecular recognition of both dsRNAs and ssRNAs is important for probing, detecting, and targeting RNA structures. Traditional asPNAs can be used to target ssRNAs and weakly formed dsRNA regions. Relatively stable dsRNA regions may be targeted by nucleobase modified dbPNAs. Since natural dsRNAs are often interrupted by ssRNA loop regions, it is of general interest to biochemists and biologists to be able to target dsRNA-ssRNA junctions in a sequence- and structure-specific manner. Here, we describe our design of combining dbPNAs and asPNAs (designated as daPNAs) for the targeting of a model dsRNA-ssRNA junction, as well as dsRNA-ssRNA regions present in a miRNA hairpin precursor and tau pre-mRNA splice site hairpin. Our data suggest that it is promising to combine traditional asPNA (with a 4-letter code: T, C, A, and G) and dbPNA (with a 4-letter code: T, L, Q, and E or S) scaffolds to achieve RNA secondary structure-specific targeting of RNA with micromolar to nanomolar affinities. Thus, daPNAs may be a useful 7-letter platform for developing novel chemical probes and therapeutic ligands targeting RNA structures.

Acknowledgements

This thesis finally becomes reality after spending 4 years in pursuing a PhD degree at Nanyang Technological University in Singapore. Throughout the journey, I gained a lot of knowledge as well as personal skillsets. This journey may not be easy without the support, guidance, encouragement from different individuals.

Foremost, I would like to express my deepest appreciation to my main supervisor, Professor Chen Gang, who provides guidance and persistent advice which is useful in my study and research. He is very supportive and gives me freedom to make my own decision in my research projects. Being a kind and friendly supervisor, he is willing to listen to my suggestions and feedbacks.

Besides my main supervisor, I would also like to thank the rest of my committee members: Professor Xavier Roca Castella, Prof Zhao Yanli and Dr. Chong Mei Sian who are willing to spend their time giving me valuable advice and solutions whenever I need without any hesitation.

I would like to thank my fellow labmates in Professor Chen Gang's group, especially Dr. Kiran Patil, Ms Desiree-Faye Toh and Dr. Tan Jiazi for not only sharing their experience and skills in research to me, but also becoming important and helpful friends in my life. I will never forget the fun time we had inside and outside the laboratory in this 4 years of my phd study.

Most importantly, I would like to express my greatest gratitude to my family members who always give me sufficient freedom, support, encouragement, financial assistance and love. Last but not least, I would like to thanks my best friend as well as my beloved companion, Ms Koo Ling Ya for her love, understanding and encouragement.

List of Publications

1. **Ong, A. A. L.**,= Toh, D.F.K.,= Patil, K.M.,= Meng, Z., Yuan, Z., Krishna, M.S., Devi, G. Haruehanroengra, P., Lu, Y., Xia, K., Okamura, K., Sheng, J., and Chen, G. Recognition of U-G, U-A, and C-G pairs by double-stranded RNA-binding PNAs incorporated with an artificial nucleobase. *Biochemistry* 58, 1319-1331 (2019). (= These authors contributed equally to this work).
2. **Ong, A. A. L.**,= Toh, D. K.,= Krishna, M. S., Patil, K.M., Okamura.K, Chen, G. Incorporating 2-thiouracil into short double-stranded RNA binding PNAs (dbPNAs) for enhanced recognition of A-U pairs and for targeting a microRNA hairpin precursor. *Biochemistry* 58, 3444-3453 (2019). (= These authors contributed equally to this work).
3. **Ong, A. A. L.**,= Tan, J.,= Bhadra, M., Dezanet.C., Patil, K.M., Kierzek, R., Decout, J.L., Roca, X., Chen, G. RNA secondary structure-based design of antisense peptide nucleic acids for modulating disease-associated aberrant tau pre-mRNA alternative splicing. *Molecules* 24, 3020 (2019). (= These authors contributed equally to this work).
4. Krishna, M. S., Toh, D. K., Meng, Z., **Ong, A. A. L.**, Wang, Z., Lu, Y., Xia, K., Prabakaran, M., Chen, G. Sequence- And Structure-Specific Probing of RNAs by Short Nucleobase-Modified dsRNA-Binding PNAs Incorporating a Fluorescent Light-up Uracil Analog. *Anal Chem* 91, 5331-5338 (2019).
5. Tan, J.,= Yang, L.,= **Ong, A. A. L.**, Shi, J., Zhong, Z., Lye, M. L., Liu, S., Lisowiec-Wachnicka, J., Kierzek, R., Roca, X., Chen, G. A disease-causing intronic point mutation C19G alters tau exon 10 splicing via RNA secondary structure rearrangement. *Biochemistry* 58, 1565-1578 (2019). (= These authors contributed equally to this work).
6. Keszy, J., Patil, K.M., Kumar, S.M. Shu, Z., Yee, Y.H., Zimmermann, L., **Ong, A.A.L.**, Toh, D.F.K., Krishna, M.S., Yang, L., Decout, J.L., Luo, D., Prabakaran, M., Chen, G., and Kierzek, E. A short chemically modified dsRNA-binding PNA (dbPNA) inhibits influenza viral replication by targeting viral RNA panhandle structure. *Bioconjugate Chem* 30, 931-943 (2019).
7. Patil, K. M., Toh, D. K., Yuan, Z., Meng, Z., Shu, Z., Zhang, H., **Ong, A. A. L.**, Krishna, M. S., Lu, L., Lu, Y., Chen, G. Incorporating uracil and 5-

halouracils into short peptide nucleic acids for enhanced recognition of A-U pairs in dsRNAs. *Nucleic Acids Res* 46, 7506-7521 **(2018)**.

8. Puah, R. Y., Jia, H., Maraswami, M., Toh, D. K., Ero, R., Yang, L., Patil, K. M., **Ong, A.A.L.**, Krishna, M. S.; Sun, R.; Tong, C.; Huang, M.; Chen, X.; Loh, T. P.; Gao, Y. G.; Liu, D. X.; Chen, G. Selective Binding to mRNA Duplex Regions by Chemically Modified Peptide Nucleic Acids Stimulates Ribosomal Frameshifting. *Biochemistry* 57, 149-159 **(2018)**.

Table of Contents

Abstract	1
Acknowledgements	4
List of Publications	6
Table of Contents	8
List of Abbreviations	12
Chapter 1 Introduction	15
1.1 RNAs and their and Biological Functions	15
1.2 RNA-targeting antisense oligonucleotide drugs and their mechanisms ..	18
1.3 Peptide Nucleic Acids (PNAs) and their properties	20
1.4 Double-stranded RNAs (dsRNAs)-binding Peptide Nucleic Acids (dbPNAs) and their application	20
References	25
Chapter 2 General Recognition of U-G, U-A, and C-G Pairs by Double-Stranded RNA-Binding PNAs Incorporated with an Artificial Nucleobase	30
2.1 Introduction	30
2.2 Materials and Methods	35
2.2.1 Molecular modelling	35
2.2.2 General methods	35
2.2.3 Synthesis of PNA monomers	36
2.2.3.1 Synthesis of PNA monomer S	37

2.2.3.2	Synthesis of PNA monomer S1.	39
2.2.4	Synthesis of PNA oligomers.	41
2.2.5	Nondenaturing polyacrylamide gel electrophoresis.	42
2.2.6	Fluorescence titration binding study.	42
2.2.7	Thermal melting.	43
2.3	Results and Discussion	45
2.3.1	Recognition of an internal C-G, U-A, and U-G pair in an RNA duplex (but not in DNA duplex) by S-modified PNAs	45
2.3.2	Effects of linker length on PNA·RNA ₂ triplex formation	59
2.3.3	Effects of sequence context and cytosine-5 methylation on PNA·RNA ₂ triplex formation	64
2.3.4	Targeting miR-26b precursor structure.	67
2.3.5	S-modified PNAs show no binding to ssRNAs	67
2.4	Conclusions	72
	References	74

Chapter 3	Incorporating 2-thiouracil into short dsRNA-binding PNAs (dbPNAs) for enhanced recognition of A-U pairs and for targeting a microRNA hairpin precursor	80
3.1	Introduction	80
3.2	Materials and Methods	83
3.2.1	General methods and synthesis of PNA oligomers	83
3.2.2	Thermal melting.	84
3.2.3	Nondenaturing Polyacrylamide Gel Electrophoresis	85
3.2.4	Data analysis	85
3.2.5	<i>In vitro</i> dicing assay	85
3.3	Results and Discussion	86

3.3.1	Replacement of T with s ² U in an modified dbPNA improves the recognition of dsRNA in a position-independent manner	86
3.3.2	s ² U-modified PNAs show sequence specific recognition of dsRNA but no appreciable recognition to dsDNA	100
3.3.3	Combining s ² U and L modifications in PNAs facilitates selective binding to dsRNAs over ssRNAs	101
3.3.4	Binding of dbPNAs to miR-198 hairpin precursor structure results in a substrate-specific inhibition of dicing <i>in vitro</i>	104
3.4	Conclusions	110
	References	111

Chapter 4	Recognition of RNA secondary structures with a programmable PNA-based 7-letter platform	115
4.1	Introduction	115
4.2	Materials and Methods	119
4.2.1	Synthesis of PNA oligomers	119
4.2.2	Molecular modelling	119
4.2.3	Nondenaturing polyacrylamide gel electrophoresis	120
4.2.4	Data analysis	120
4.3	Results and Discussion	121
4.3.1	Recognition of dsRNA containing a 3'-overhang poly (U) tail by combining dsRNA-binding PNAs (dbPNAs) and antisense peptide nucleic acids (asPNAs)	121
4.3.2	Targeting dsRNA-ssRNA junctions presents in Tau pre-mRNA hairpin structure by combining dsRNA-binding PNAs (dbPNAs) and antisense peptide nucleic acids (asPNAs).....	123
4.3.3	Targeting dsRNA-ssRNA junction presents in miRNA-26b hairpin structure by combining dsRNA-binding PNAs (dbPNAs) and antisense peptide nucleic acids (asPNAs).....	125
4.4	Conclusions	127

References	129
 Chapter 5 Conclusion and Future Work	 132
References	135
 Appendixes	 136
Appendix 1. ^1H , ^{13}C NMR Spectra for the synthesized compounds	136
Appendix 2. MALDI-TOF characterization of PNA oligomers	142

List of Abbreviations

δ	Chemical shift
$^{\circ}\text{C}$	degrees Celsius
2'-AE	2'-O-aminoethoxy
2AP	2-Aminopurine
5'ss	5' splice-site
^{Br} U	5-bromouracil
A	Adenine
AEG	N-(2-aminoethyl)-glycine
ASO	Antisense oligonucleotide
BNA	Bridged nucleic acid
Boc	tert-Butyloxycarbonyl
C	Cytosine
CHCA	α -cyano-4-hydroxycinnamic acid
Cy3	Cyanine-3
daPNAs	combining dbPNAs and asPNAs
dbPNA	dsRNA-binding PNA
DCM	Dichloromethane
DIPEA	N,N-Diisopropylethylamine
DMD	Duchenne Muscular Dystrophy
DMF	Dimethylformamide
DMSO-d ₆	Deuterated DMSO
DNA	Deoxyribonucleic acid
ds	double-stranded
dsDNA	double-stranded DNA
dsRNA	double-stranded RNA
E	3-oxo-2,3-dihydropyridazine
EDTA	Ethylenediaminetetraacetic acid

FTDP-17	Frontotemporal dementia with Parkinsonism linked to chromosome 17
G	Guanine
GROMACS	GRoningen MAchine for Chemical Simulations
HBTU	3-[Bis(dimethylamino)methylumyl]-3 <i>H</i> -benzotriazol-1-oxide hexafluorophosphate
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HRMS-EI	High-resolution mass spectrometry with electron ionization
J	pseudoisocytosine
L	thio-pseudoisocytosine
LCMS-ESI	Liquid chromatography-mass spectroscopy with electrospray ionization source
LNA	Locked Nucleic Acid
M	Molar
m ⁵ C	5-methylcytosine
<i>m/z</i>	mass-to-charge ratio
MALDI-TOF	Matrix-assisted laser desorption/ionization-time of flight
MAPT	Microtubule-associated protein tau gene
MBHA·HCl	4-Methylbenzhydrylamine hydrochloride
MD	Molecular Dynamic
MES	2-(N-morpholino)ethanesulfonic acid
mHz	MegaHertz
miRISC	miRNA-induced silencing complex
miRNA	microRNA
MM	Molecular Mechanic
mRNAs	messenger RNAs
N ₂	Nitrogen gas
ncRNA	non-coding RNA
nm	nanometer
NMM	N-Methylmorpholine
NMR	Nuclear Magnetic Resonance
Nt	nucleotide

PAGE	nondenaturing polyacrylamide gel electrophoresis
Pd/C	Palladium on carbon
PNA	Peptide Nucleic Acid
ppm	part per million
pre-miRNA	precursor miRNA
pre-mRNAs	precursor mRNAs
pri-miRNA	primary miRNA
PS	Phosphorothioate
PyBOP	Benzotriazol-1-yl-oxytripyrrolidinophosphonium hexafluorophosphate
Q	guanidine-modified 5-methylcytosine
RBP	RNA-binding protein

Chapter 1

Introduction

1.1 RNAs and their biological functions

Ribonucleic acid (RNA) is an essential biopolymer, which consists of a ribose sugar, a phosphate group, and a nitrogenous base or nucleobase. Each natural RNA nucleotide contains a nucleobase. Interestingly, these bases can be modified in numerous ways during the maturation of RNA. Currently, there are hundreds of naturally occurring modified nucleotides with modifications on the base and/or sugar-phosphate back that have been reported¹. Unlike DNA, the presence of 2' hydroxyl group of the ribose sugar provides a more stable A-form conformation than B-form conformation. RNA is a single-stranded molecule which folds itself into more complex secondary and three-dimensional structures through intramolecular base pairings and stacking interactions. Different kinds of structural elements can be found in RNA secondary and tertiary structure motifs including hairpins, internal loops, bulges, junctions, and pseudoknots (**Figure 1.1**)^{2,3}. The structural diversity of RNA allows it to be vital in myriad of biological functions such as transcription, translation, gene expression and regulation.

RNA secondary structure motifs, including duplex, stem-loop, and pseudoknot, are important in gene regulation and expression. Messenger RNAs (mRNAs) are a class of RNA molecules that transfer the genetic information from DNA to protein. RNA

secondary structures in the 5'- and 3'-untranslated regions of mRNAs are involved in the regulation of mRNAs' function⁴. In addition, RNA secondary structures have been found in the untranslated regions of many viral RNAs to regulate the viral life cycle⁵. Eukaryotic mRNAs undergo several processing steps which allow precursor mRNAs (pre-mRNAs) to be matured into mRNAs. Alternative splicing is one of the pre-mRNA processing steps which generates proteomic diversity⁶. Approximately 95% of the human genes undergo alternative splicing⁷. RNA secondary structure plays an vital role in splicing regulation. The presence or absence of an RNA secondary structure may promote exon skipping or exon inclusion, which then produce different splicing variants⁸. For example, a self-complementary RNA hairpin structure at the 3'-end of exon 10 and 5'-end of intron 10 in microtubule-associated protein tau gene (*MAPT*) pre-mRNA regulates the alternative splicing of exon 10 to maintain the ratio of 4R:3R isoforms⁸⁻¹³. The formation of this hairpin inhibits the recognition of U1 small nuclear ribonucleoprotein (U1 snRNP) by masking the 5' splice-site (5'ss) which is a key initial step in splicing¹⁴⁻¹⁶.

In general, RNAs can be categorized into different types based on their biological functions: messenger RNAs (mRNAs) and non-coding RNAs (ncRNAs) including transfer RNAs (tRNAs), ribosomal RNAs (rRNAs), small nuclear RNAs (snRNAs) . mRNAs, tRNAs, and rRNAs are involved in the protein synthesis. While 85% of the human genome is transcribed into RNAs, the majority of the RNAs are not directly involved in protein coding¹⁷. In the recent years, more ncRNAs have been discovered with important roles in processes such as DNA replication and gene regulation¹⁸⁻²⁰. MicroRNAs (miRNAs) are a class of ncRNAs containing around 22 nucleotides and are involved in gene silencing and regulation. Similar as the maturation of mRNAs,

the maturation of miRNAs requires several processing steps starting from primary miRNA (pri-miRNA) transcripts which are then processed to precursor miRNAs (pre-miRNAs) in the nucleus. Then, pre-miRNAs are processed into mature miRNAs by dsRNA-binding RNase Dicer or Argonaute in the cytoplasm. Both pri-miRNA and pre-miRNA form stem-loop structures (hairpins) while mature miRNAs are double-stranded RNAs. Numerous studies have shown that dysregulation of miRNAs are associated with many diseases such as cancers, neurodegenerative diseases, heart disease, kidney disease, and obesity²¹⁻²⁵.

RNA and RNA-binding proteins (RBPs) form ribonucleoprotein complexes. RBPs are involved in post-transcriptional processes of RNAs, gene regulation, and cellular processes including cellular function, transport, and localization²⁶⁻²⁸. Mutations and dysregulation of the expression of RNAs cause dysregulation and leads to diseases such as cancers and Alzheimer's disease²⁹. Hence, RNA has emerged as a class of potential therapeutic target for both small molecules and antisense oligonucleotides^{30,31}.

Figure 1.1: (a) Illustration of an RNA structure and a major-groove RNA triple helix consists of consecutive U·A·U base triples. (b) 3D structure of the RNA structure shown in panel a (PDB 3P22)³².

1.2 RNA-targeting antisense oligonucleotide drugs and their mechanisms

Recognition of RNA by antisense oligonucleotides (ASOs) can be achieved by complementary base pairing. Unmodified DNA and RNA oligonucleotides are not ideal candidates as RNA-targeting drugs due to several factors: degradation by nucleases, poor cellular uptake, unfavourable pharmacokinetics properties, and sub-optimal binding affinity to the complementary target sequence^{31,33,34}. Over the years, several types of synthetic oligonucleotides have been developed to aid RNA recognition including phosphorothioate (PS), 2'-ribose, bridged nucleic acid (BNA) and many others³¹ (**Figure 1.2**). These synthetic oligonucleotides can target different types of disease-associated RNAs through different mechanisms to achieve therapeutic applications. In general, the mechanisms of action of ASOs can be either

inducing steric hindrance to inhibit recognition of proteins or causing cleavage of RNA by the recruitment of RNase H upon base pairing with the target RNAs which can be mRNA, pre-mRNA, miRNA and others³¹. For example, the activity of miRNA can be inhibited by an anti-miR oligonucleotide which is complementary to the mature miRNA guide strand³⁵. The designed anti-miR will bind to the target of mature miRNA in the miRNA-induced silencing complex (miRISC)³⁶. The recent FDA-approved ASO drug, nusinersen (Spinraza) was developed by targeting the spinal muscular atrophy (SMN) pre-mRNA transcript and modulates the splicing of SMN transcript which increases the expression of alternatively-spliced variant³⁷. Hence, splice-switching antisense oligonucleotides (SSOs) are emerging as promising therapeutic agents for neurodegenerative and other diseases. However, toxicity and off-target effects of ASOs are the concerns in the development of RNA-targeting oligonucleotide drug.

Figure 1.2: Chemical structures of (a) RNA (b) PS (c) 2'-OMe RNA (d) LNA and (e) PNA

1.3 Peptide Nucleic Acids (PNAs) and their properties

Peptide Nucleic Acid (PNA), which is a polyamide analog of naturally occurring DNA and RNA, was first reported by biochemist Peter E. Nielsen and his coworkers in 1991³⁸. PNA was studied to target double-stranded DNA through Hoogsteen base pairing. The chemical structure of a canonical PNA consists of a neutral N-(2-aminoethyl)-glycine (AEG) backbone, a methylenecarbonyl linker, and a nucleobase (**Figure 1.2e**). Similar to traditional oligonucleotides, the binding properties of PNA is governed by base pairing³⁹. A PNA can form Watson-Crick base pairing to the complementary strand of DNA, RNA or even PNA itself to form a stable PNA-DNA, PNA-RNA, and PNA-PNA duplex, respectively^{40,41}. Strand invasion occurs when PNA disrupts the pre-formed DNA or RNA hairpins to form a new PNA-DNA or PNA-RNA duplex with an unpaired DNA or RNA single strand, respectively. The unique polyamide structure of PNA provides enhanced chemical stability and resistance towards nucleases and proteases as compared to that of DNA and RNA or peptides. In addition, the neutral PNA backbone gives rise to a better binding affinity to the complementary oligonucleotides by reducing the electrostatic repulsion between the complementary strands. The strong hybridization properties of PNA enable PNA to be utilized in anti-gene (transcription) and antisense (translation) application^{42,43}. However, moderate water solubility, poor cellular uptake and biodistribution hinder the development of PNA-based antisense therapeutics^{44,45}.

1.4 Double-stranded RNAs (dsRNAs)-binding Peptide Nucleic Acids (dbPNAs) and their application

RNA duplexes or double-stranded RNAs (dsRNAs) can be targeted by a third strand, which is termed triplex-forming oligonucleotides (TFOs) though Hoogsteen base

pairing. Locked Nucleic Acids (LNAs)-modified TFOs have been reported to recognize dsRNAs as well as dsDNAs through the formation of TFO·RNA₂ and TFO·DNA₂ triplex⁴⁶. However, the drawback of utilizing TFO to form TFO·RNA₂ triplex is that the binding affinity of TFO is weak at physiological condition and TFO often binds more strongly to DNA duplexes⁴⁶⁻⁵¹.

Similar to TFOs, PNAs can bind to the RNA duplex without strand invasion to form major groove PNA·RNA₂ triplex through the formation of consecutive PNA-RNA-RNA base triples formed by Hoogsteen base pairings⁵²⁻⁵⁸ (**Figure 1.3**). Remarkably, unlike TFO, PNA often binds more strongly to dsRNA than dsDNA, presumably due to the geometry compatibility of the major groove of RNA and favourable backbone-backbone interaction between RNA and PNA³. Similar to the naturally existing RNA major groove triplex, a pyrimidine motif (Thymine, Uracil and Cytosine) is grafted on PNA for the recognition of RNA Watson-Crick base pairs to form T·A-U, U·A-U and C⁺·G-C base triples, respectively (**Figure 1.3a,b**). The formation of C⁺·G-C base triples is favoured at relatively low pH (<6.0) due to the fact that cytosine N³ position (pK_a=4.5 for a C monomer) in triplex-forming oligonucleotides (TFOs) needs to be protonated to form stable C⁺·G Hoogsteen pairs. Triplex formation is severely destabilized if the purine tract is intervened by pyrimidine bases which form Watson-Crick C-G and U-A inverted base pairs. Unmodified nucleobases (A, G, C, T) are not able to recognize the inverted Watson-Crick base pairs. Therefore, modifications on nucleobase have been studied in order to overcome these problems.

Recognition of dsRNAs by chemically-modified PNAs has been widely studied. Recognition of dsRNAs containing A-U base pair is important because of the

comparatively high occurrence in biologically relevant RNA duplexes. Unmodified nucleobases, thymine (T) and uracil (U) are grafted on PNAs to target Watson-Crick A-U base pairs via the formation of T·A-U and U·A-U base triples^{3,52,58-61} (**Figure 1.3a**). Short PNAs incorporating uracils and halouracils facilitate the recognition of dsRNA containing an A-U base pair but not dsDNAs at near physiological conditions with minimal pH dependence⁵⁸ (**Figure 1.3e**). Short PNAs incorporating modified base thio-pseudoisocytosine (L) have been reported to facilitate recognition of Watson-Crick G-C pair in dsRNAs at near physiological conditions with weakened pH dependence (**Figure 1.3c**). Furthermore, L-modified PNAs show selective recognition of dsRNAs over ssRNAs because of the steric hindrance arising from the formation of an unstable Watson-Crick-like L-G pair⁵³ (**Figure 1.3g**).

On the other hand, targeting the Watson-Crick C-G and U-A pairs is challenging due to a decreased number of hydrogen bond donors and acceptors on the Hoogsteen edge of C and U bases^{3,62-64}. PNAs incorporated with guanidine-modified 5-methylcytosine (Q)⁶⁵ show sequence-specific and selective binding to C-G pair-containing dsRNAs over dsDNAs⁵⁴ (**Figure 1.3d**). Although Q is a C derivative, Q does not form a Watson-Crick-like Q-G pair, mainly because alkylation of the amino group in the Q base results in a steric clash in a Watson-Crick-like Q-G pair⁵⁴ (**Figure 1.3h**). Thus, Q can recognize C-G pair in dsRNAs but not a G in ssRNAs. PNAs incorporated with 3-oxo-2,3-dihydropyridazine (E) monomer⁶⁶ have been utilized for targeting dsRNAs containing Watson-Crick U-A pairs (**Figure 1.3f**)⁶³. However, an E·U-A base triple may be stabilized by only one hydrogen bond formed between E base and U base in a U-A pair (**Figure 1.3f**). Thus, modified nucleobases capable of recognizing RNA

Watson-Crick and inverted Watson-Crick base pairs with improved binding affinity as well as specificity are required in order to achieve superior recognition of dsRNAs.

In recent years, double-stranded RNAs-binding PNAs (dbPNAs) have been utilized in targeting several biologically-relevant RNA duplexes in vitro and in cell cultures⁶⁷⁻⁶⁹. Rozners and his coworkers reported the suppress of translation by triplex formation between a PNA and mRNA in vitro and in cell culture⁶⁷. Chen and his coworkers reported the sequence-selective recognition of mRNA duplex regions by a 8-mer dbPNA stimulates ribosomal frameshifting in vitro⁶⁸. The same group then reported the recognition of viral RNA panhandle structure by a 10-mer dbPNA which inhibited influenza viral replication⁶⁹. Cellular uptake of dbPNAs has been improved by incorporating artificial nucleobases carrying positive charges^{54,59}. The use of dbPNAs with relatively shorter length (10-mer or lesser) as compared to the antisense PNAs with at least 13-mer in length may improve the water solubility and bioavailability of PNA. Hence, dbPNAs has emerged as a potential dsRNA-targeting drug.

Figure 1.3: Chemical structures of (a) T/U·A·U, (b) C⁺·G·C, (c) L·G·C (d) Q·C·G (e) X·U·A·U and (f) E·U·A base triples, and (g) unstable Watson–Crick-like L-G and (h) unstable Watson–Crick-like Q-G pairs. The sugar-phosphate backbone of RNA is represented by letter R. Black dashed lines indicate hydrogen bonds.

References

- (1) Boccaletto, P.; Machnicka, M. A.; Purta, E.; Piatkowski, P.; Baginski, B.; Wirecki, T. K.; de Crecy-Lagard, V.; Ross, R.; Limbach, P. A.; Kotter, A.; Helm, M.; Bujnicki, J. M. MODOMICS: a database of RNA modification pathways. 2017 update. *Nucleic Acids Res* **2018**, *46*, D303-D307.
- (2) Wyatt, J. R.; Puglisi, J. D.; Tinoco, I., Jr. RNA folding: pseudoknots, loops and bulges. *Bioessays* **1989**, *11*, 100-106.
- (3) Devi, G.; Zhou, Y.; Zhong, Z.; Toh, D. F. K.; Chen, G. RNA Triplexes: from structural principles to biological and biotech applications. *Wiley Interdiscip Rev RNA* **2015**, *6*, 111-128.
- (4) Wilkie, G. S.; Dickson, K. S.; Gray, N. K. Regulation of mRNA translation by 5'- and 3'-UTR-binding factors. *Trends Biochem Sci* **2003**, *28*, 182-188.
- (5) Liu, Y.; Wimmer, E.; Paul, A. V. Cis-acting RNA elements in human and animal plus-strand RNA viruses. *Biochim Biophys Acta* **2009**, *1789*, 495-517.
- (6) Nilsen, T. W.; Graveley, B. R. Expansion of the eukaryotic proteome by alternative splicing. *Nature* **2010**, *463*, 457-463.
- (7) Sultan, M.; Schulz, M. H.; Richard, H.; Magen, A.; Klingenhoff, A.; Scherf, M.; Seifert, M.; Borodina, T.; Soldatov, A.; Parkhomchuk, D.; Schmidt, D.; O'Keefe, S.; Haas, S.; Vingron, M.; Lehrach, H.; Yaspo, M. L. A global view of gene activity and alternative splicing by deep sequencing of the human transcriptome. *Science* **2008**, *321*, 956-960.
- (8) Buratti, E.; Baralle, F. E. Influence of RNA secondary structure on the pre-mRNA splicing process. *Mol Cell Biol* **2004**, *24*, 10505-10514.
- (9) Hutton, M.; Lendon, C. L.; Rizzu, P.; Baker, M.; Froelich, S.; Houlden, H.; Pickering-Brown, S.; Chakraverty, S.; Isaacs, A.; Grover, A.; Hackett, J.; Adamson, J.; Lincoln, S.; Dickson, D.; Davies, P.; Petersen, R. C.; Stevens, M.; de Graaff, E.; Wauters, E.; van Baren, J.; Hillebrand, M.; Joosse, M.; Kwon, J. M.; Nowotny, P.; Che, L. K.; Norton, J.; Morris, J. C.; Reed, L. A.; Trojanowski, J.; Basun, H.; Lannfelt, L.; Neystat, M.; Fahn, S.; Dark, F.; Tannenberg, T.; Dodd, P. R.; Hayward, N.; Kwok, J. B.; Schofield, P. R.; Andreadis, A.; Snowden, J.; Craufurd, D.; Neary, D.; Owen, F.; Oostra, B. A.; Hardy, J.; Goate, A.; van Swieten, J.; Mann, D.; Lynch, T.; Heutink, P. Association of missense and 5'-splice-site mutations in tau with the inherited dementia FTDP-17. *Nature* **1998**, *393*, 702-705.
- (10) McCarthy, A.; Loneragan, R.; Olszewska, D. A.; O'Dowd, S.; Cummins, G.; Magennis, B.; Fallon, E. M.; Pender, N.; Huey, E. D.; Cosentino, S.; O'Rourke, K.; Kelly, B. D.; O'Connell, M.; Delon, I.; Farrell, M.; Spillantini, M. G.; Rowland, L. P.; Fahn, S.; Craig, P.; Hutton, M.; Lynch, T. Closing the tau loop: the missing tau mutation. *Brain* **2015**, *138*, 3100-3109.
- (11) Donahue, C. P.; Muratore, C.; Wu, J. Y.; Kosik, K. S.; Wolfe, M. S. Stabilization of the tau exon 10 stem loop alters pre-mRNA splicing. *J Biol Chem* **2006**, *281*, 23302-23306.
- (12) Varani, L.; Hasegawa, M.; Spillantini, M. G.; Smith, M. J.; Murrell, J. R.; Ghetti, B.; Klug, A.; Goedert, M.; Varani, G. Structure of tau exon 10 splicing regulatory element RNA and destabilization by mutations of frontotemporal dementia and parkinsonism linked to chromosome 17. *Proc Natl Acad Sci U S A* **1999**, *96*, 8229-8234.
- (13) Grover, A.; Houlden, H.; Baker, M.; Adamson, J.; Lewis, J.; Prihar, G.; Pickering-Brown, S.; Duff, K.; Hutton, M. 5' splice site mutations in tau associated with the inherited dementia FTDP-17 affect a stem-loop structure that regulates alternative splicing of exon 10. *J Biol Chem* **1999**, *274*, 15134-15143.
- (14) Roca, X.; Krainer, A. R.; Eperon, I. C. Pick one, but be quick: 5' splice sites and the problems of too many choices. *Genes Dev* **2013**, *27*, 129-144.
- (15) Roca, X.; Akerman, M.; Gaus, H.; Berdeja, A.; Bennett, C. F.; Krainer, A. R. Widespread recognition of 5' splice sites by noncanonical base-pairing to U1 snRNA involving bulged nucleotides. *Genes Dev* **2012**, *26*, 1098-1109.

- (16) Tan, J.; Ho, Jia Xin J.; Zhong, Z.; Luo, S.; Chen, G.; Roca, X. Noncanonical registers and base pairs in human 5' splice-site selection. *Nucleic Acids Research* **2016**, *44*, 3908-3921.
- (17) Hangauer, M. J.; Vaughn, I. W.; McManus, M. T. Pervasive transcription of the human genome produces thousands of previously unidentified long intergenic noncoding RNAs. *PLoS Genet* **2013**, *9*, e1003569.
- (18) Zhang, A. T.; Langley, A. R.; Christov, C. P.; Kheir, E.; Shafee, T.; Gardiner, T. J.; Krude, T. Dynamic interaction of Y RNAs with chromatin and initiation proteins during human DNA replication. *J Cell Sci* **2011**, *124*, 2058-2069.
- (19) Espinoza, C. A.; Allen, T. A.; Hieb, A. R.; Kugel, J. F.; Goodrich, J. A. B2 RNA binds directly to RNA polymerase II to repress transcript synthesis. *Nat Struct Mol Biol* **2004**, *11*, 822-829.
- (20) Hirota, K.; Miyoshi, T.; Kugou, K.; Hoffman, C. S.; Shibata, T.; Ohta, K. Stepwise chromatin remodelling by a cascade of transcription initiation of non-coding RNAs. *Nature* **2008**, *456*, 130-134.
- (21) Penna, E.; Orso, F.; Taverna, D. miR-214 as a key hub that controls cancer networks: small player, multiple functions. *J Invest Dermatol* **2015**, *135*, 960-969.
- (22) Maes, O. C.; Chertkow, H. M.; Wang, E.; Schipper, H. M. MicroRNA: Implications for Alzheimer Disease and other Human CNS Disorders. *Curr Genomics* **2009**, *10*, 154-168.
- (23) Thum, T.; Galuppo, P.; Wolf, C.; Fiedler, J.; Kneitz, S.; van Laake, L. W.; Doevendans, P. A.; Mummery, C. L.; Borlak, J.; Haverich, A.; Gross, C.; Engelhardt, S.; Ertl, G.; Bauersachs, J. MicroRNAs in the human heart: a clue to fetal gene reprogramming in heart failure. *Circulation* **2007**, *116*, 258-267.
- (24) Phua, Y. L.; Chu, J. Y.; Marrone, A. K.; Bodnar, A. J.; Sims-Lucas, S.; Ho, J. Renal stromal miRNAs are required for normal nephrogenesis and glomerular mesangial survival. *Physiol Rep* **2015**, *3*.
- (25) Zhu, H.; Shyh-Chang, N.; Segre, A. V.; Shinoda, G.; Shah, S. P.; Einhorn, W. S.; Takeuchi, A.; Engreitz, J. M.; Hagan, J. P.; Kharas, M. G.; Urbach, A.; Thornton, J. E.; Triboulet, R.; Gregory, R. I.; Consortium, D.; Investigators, M.; Altshuler, D.; Daley, G. Q. The Lin28/let-7 axis regulates glucose metabolism. *Cell* **2011**, *147*, 81-94.
- (26) Hentze, M. W.; Castello, A.; Schwarzl, T.; Preiss, T. A brave new world of RNA-binding proteins. *Nat Rev Mol Cell Biol* **2018**, *19*, 327-341.
- (27) Glisovic, T.; Bachorik, J. L.; Yong, J.; Dreyfuss, G. RNA-binding proteins and post-transcriptional gene regulation. *FEBS Lett* **2008**, *582*, 1977-1986.
- (28) Hogan, D. J.; Riordan, D. P.; Gerber, A. P.; Herschlag, D.; Brown, P. O. Diverse RNA-binding proteins interact with functionally related sets of RNAs, suggesting an extensive regulatory system. *PLoS Biol* **2008**, *6*, e255.
- (29) Cooper, T. A.; Wan, L.; Dreyfuss, G. RNA and disease. *Cell* **2009**, *136*, 777-793.
- (30) Connelly, C. M.; Moon, M. H.; Schneekloth, J. S., Jr. The Emerging Role of RNA as a Therapeutic Target for Small Molecules.
- (31) Shen, X.; Corey, D. R. Chemistry, mechanism and clinical status of antisense oligonucleotides and duplex RNAs. *Nucleic Acids Res* **2018**, *46*, 1584-1600.
- (32) Mitton-Fry, R. M.; DeGregorio, S. J.; Wang, J.; Steitz, T. A.; Steitz, J. A. Poly(A) tail recognition by a viral RNA element through assembly of a triple helix. *Science* **2010**, *330*, 1244-1247.
- (33) Juliano, R. L. The delivery of therapeutic oligonucleotides. *Nucleic Acids Res* **2016**, *44*, 6518-6548.
- (34) Crooke, S. T.; Wang, S.; Vickers, T. A.; Shen, W.; Liang, X. H. Cellular uptake and trafficking of antisense oligonucleotides. *Nat Biotechnol* **2017**, *35*, 230-237.
- (35) Stenvang, J.; Petri, A.; Lindow, M.; Obad, S.; Kauppinen, S. Inhibition of microRNA function by antimiR oligonucleotides. *Silence* **2012**, *3*, 1.

- (36) Fabani, M. M.; Gait, M. J. miR-122 targeting with LNA/2'-O-methyl oligonucleotide mixmers, peptide nucleic acids (PNA), and PNA-peptide conjugates. *RNA* **2008**, *14*, 336-346.
- (37) Hua, Y.; Vickers, T. A.; Baker, B. F.; Bennett, C. F.; Krainer, A. R. Enhancement of SMN2 exon 7 inclusion by antisense oligonucleotides targeting the exon. *PLoS Biol* **2007**, *5*, e73.
- (38) Nielsen, P. E.; Egholm, M.; Berg, R. H.; Buchardt, O. Sequence-selective recognition of DNA by strand displacement with a thymine-substituted polyamide. *Science* **1991**, *254*, 1497-1500.
- (39) Egholm, M.; Buchardt, O.; Christensen, L.; Behrens, C.; Freier, S. M.; Driver, D. A.; Berg, R. H.; Kim, S. K.; Norden, B.; Nielsen, P. E. PNA hybridizes to complementary oligonucleotides obeying the Watson-Crick hydrogen-bonding rules. *Nature* **1993**, *365*, 566-568.
- (40) Egholm, M.; Nielsen, P. E.; Buchardt, O.; Berg, R. H. Recognition of guanine and adenine in DNA by cytosine and thymine containing peptide nucleic acids (PNA). *Journal of the American Chemical Society* **1992**, *114*, 9677-9678.
- (41) Dueholm, K. L.; Egholm, M.; Behrens, C.; Christensen, L.; Hansen, H. F.; Vulpius, T.; Petersen, K. H.; Berg, R. H.; Nielsen, P. E.; Buchardt, O. Synthesis of Peptide Nucleic Acid Monomers Containing the Four Natural Nucleobases: Thymine, Cytosine, Adenine, and Guanine and Their Oligomerization. *The Journal of Organic Chemistry* **1994**, *59*, 5767-5773.
- (42) Hanvey, J. C.; Peffer, N. J.; Bisi, J. E.; Thomson, S. A.; Cadilla, R.; Josey, J. A.; Ricca, D. J.; Hassman, C. F.; Bonham, M. A.; Au, K. G.; et al. Antisense and antigene properties of peptide nucleic acids. *Science* **1992**, *258*, 1481-1485.
- (43) Hyrup, B.; Nielsen, P. E. Peptide nucleic acids (PNA): synthesis, properties and potential applications. *Bioorg Med Chem* **1996**, *4*, 5-23.
- (44) Mardirossian, G.; Lei, K.; Rusckowski, M.; Chang, F.; Qu, T.; Egholm, M.; Hnatowich, D. J. In vivo hybridization of technetium-99m-labeled peptide nucleic acid (PNA). *J Nucl Med* **1997**, *38*, 907-913.
- (45) McMahon, B. M.; Mays, D.; Lipsky, J.; Stewart, J. A.; Fauq, A.; Richelson, E. Pharmacokinetics and tissue distribution of a peptide nucleic acid after intravenous administration. *Antisense Nucleic Acid Drug Dev* **2002**, *12*, 65-70.
- (46) Zhou, Y.; Kierzek, E.; Loo, Z. P.; Antonio, M.; Yau, Y. H.; Chuah, Y. W.; Geifman-Shochat, S.; Kierzek, R.; Chen, G. Recognition of RNA duplexes by chemically modified triplex-forming oligonucleotides. *Nucleic Acids Res* **2013**, *41*, 6664-6673.
- (47) Roberts, R. W.; Crothers, D. M. Stability and properties of double and triple helices: dramatic effects of RNA or DNA backbone composition. *Science* **1992**, *258*, 1463-1466.
- (48) Escude, C.; Francois, J. C.; Sun, J. S.; Ott, G.; Sprinzl, M.; Garestier, T.; Helene, C. Stability of triple helices containing RNA and DNA strands: experimental and molecular modeling studies. *Nucleic Acids Res* **1993**, *21*, 5547-5553.
- (49) Wilds, C. J.; Damha, M. J. Duplex recognition by oligonucleotides containing 2'-deoxy-2'-fluoro-D-arabinose and 2'-deoxy-2'-fluoro-D-ribose. Intermolecular 2'-OH-phosphate contacts versus sugar puckering in the stabilization of triple-helical complexes. *Bioconjug Chem* **1999**, *10*, 299-305.
- (50) Han, H.; Dervan, P. B. Sequence-specific recognition of double helical RNA and RNA-DNA by triple helix formation. *Proceedings of the National Academy of Sciences of the United States of America* **1993**, *90*, 3806-3810.
- (51) Wang, S.; Kool, E. T. Relative stabilities of triple helices composed of combinations of DNA, RNA and 2'-O-methyl-RNA backbones: chimeric circular oligonucleotides as probes. *Nucleic Acids Res* **1995**, *23*, 1157-1164.
- (52) Li, M.; Zengeya, T.; Rozners, E. Short peptide nucleic acids bind strongly to homopurine tract of double helical RNA at pH 5.5. *J Am Chem Soc* **2010**, *132*, 8676-8681.

- (53) Devi, G.; Yuan, Z.; Lu, Y.; Zhao, Y.; Chen, G. Incorporation of thio-pseudoisocytosine into triplex-forming peptide nucleic acids for enhanced recognition of RNA duplexes. *Nucleic Acids Res* **2014**, *42*, 4008-4018.
- (54) Toh, D.-F.; Devi, G.; Patil, K. M.; Qu, Q.; Maraswami, M.; Xiao, Y.; Loh, T.-P.; Zhao, Y.; Chen, G. Incorporating a guanidine-modified cytosine base into triplex-forming PNAs for the recognition of a C-G pyrimidine-purine inversion site of an RNA duplex. *Nucleic Acids Res* **2016**, *44*, 9071-9082.
- (55) Zengeya, T.; Gupta, P.; Rozners, E. Sequence selective recognition of double-stranded RNA using triple helix-forming peptide nucleic acids. *Methods Mol Biol* **2014**, *1050*, 83-94.
- (56) Ong, A. A. L.; Toh, D. K.; Patil, K. M.; Meng, Z.; Yuan, Z.; Krishna, M. S.; Devi, G.; Haruehanroengra, P.; Lu, Y.; Xia, K.; Okamura, K.; Sheng, J.; Chen, G. General recognition of U-G, U-A, and C-G pairs by double-stranded RNA-binding PNAs (dbPNAs) incorporated with an artificial nucleobase. *Biochemistry* **2019**.
- (57) Krishna, M. S.; Toh, D. K.; Meng, Z.; Ong, A. A. L.; Wang, Z.; Lu, Y.; Xia, K.; Prabakaran, M.; Chen, G. Sequence- And Structure-Specific Probing of RNAs by Short Nucleobase-Modified dsRNA-Binding PNAs Incorporating a Fluorescent Light-up Uracil Analog. *Anal Chem* **2019**, *91*, 5331-5338.
- (58) Patil, K. M.; Toh, D. K.; Yuan, Z.; Meng, Z.; Shu, Z.; Zhang, H.; Ong, A. A. L.; Krishna, M. S.; Lu, L.; Lu, Y.; Chen, G. Incorporating uracil and 5-halouracils into short peptide nucleic acids for enhanced recognition of A-U pairs in dsRNAs. *Nucleic Acids Res* **2018**, *46*, 7506-7521.
- (59) Hnedzko, D.; McGee, D. W.; Karamitas, Y. A.; Rozners, E. Sequence-selective recognition of double-stranded RNA and enhanced cellular uptake of cationic nucleobase and backbone-modified peptide nucleic acids. *RNA* **2017**, *23*, 58-69.
- (60) Virta, P. M.; Tahtinen, V.; Granqvist, L.; Murtola, M.; Stromberg, R. 19F NMR spectroscopic analysis of the binding modes in triple helical PNA/microRNA-complexes. *Chem Eur J* **2017**, *23*, 7113-7124.
- (61) Sato, T.; Sato, Y.; Nishizawa, S. Triplex-Forming Peptide Nucleic Acid Probe Having Thiazole Orange as a Base Surrogate for Fluorescence Sensing of Double-stranded RNA. *J Am Chem Soc* **2016**, *138*, 9397-9400.
- (62) Zengeya, T.; Li, M.; Rozners, E. PNA containing isocytidine nucleobase: synthesis and recognition of double helical RNA. *Bioorg Med Chem Lett* **2011**, *21*, 2121-2124.
- (63) Gupta, P.; Zengeya, T.; Rozners, E. Triple helical recognition of pyrimidine inversions in polypurine tracts of RNA by nucleobase-modified PNA. *Chem Commun* **2011**, *47*, 11125-11127.
- (64) Gowers, D. M.; Fox, K. R. Towards mixed sequence recognition by triple helix formation. *Nucleic Acids Res* **1999**, *27*, 1569-1577.
- (65) Semenyuk, A.; Darian, E.; Liu, J.; Majumdar, A.; Cuenoud, B.; Miller, P. S.; Mackerell, A. D., Jr.; Seidman, M. M. Targeting of an interrupted polypurine:polypyrimidine sequence in mammalian cells by a triplex-forming oligonucleotide containing a novel base analogue. *Biochemistry* **2010**, *49*, 7867-7878.
- (66) Eldrup, A. B.; Dahl, O.; Nielsen, P. E. A novel peptide nucleic acid monomer for recognition of thymine in triple-helix structures. *J Am Chem Soc* **1997**, *119*, 11116-11117.
- (67) Endoh, T.; Hnedzko, D.; Rozners, E.; Sugimoto, N. Nucleobase-Modified PNA Suppresses Translation by Forming a Triple Helix with a Hairpin Structure in mRNA In Vitro and in Cells. *Angew Chem Int Ed Engl* **2016**, *55*, 899-903.
- (68) Puah, R. Y.; Jia, H.; Maraswami, M.; Kaixin Toh, D. F.; Ero, R.; Yang, L.; Patil, K. M.; Lerk Ong, A. A.; Krishna, M. S.; Sun, R.; Tong, C.; Huang, M.; Chen, X.; Loh, T. P.; Gao, Y. G.; Liu, D. X.; Chen, G. Selective Binding to mRNA Duplex Regions by Chemically Modified Peptide Nucleic Acids Stimulates Ribosomal Frameshifting. *Biochemistry* **2018**, *57*, 149-159.
- (69) Keszy, J.; Patil, K. M.; Kumar, S. R.; Shu, Z.; Yong, H. Y.; Zimmermann, L.; Ong, A. A. L.; Toh, D. K.; Krishna, M. S.; Yang, L.; Decout, J. L.; Luo, D.; Prabakaran, M.;

Chen, G.; Kierzek, E. A Short Chemically Modified dsRNA-Binding PNA (dbPNA) Inhibits Influenza Viral Replication by Targeting Viral RNA Panhandle Structure. *Bioconjug Chem* **2019**, *30*, 931-943.

Chapter 2

General Recognition of U-G, U-A, and C-G Pairs by Double-Stranded RNA-Binding PNAs Incorporated with an Artificial Nucleobase

2.1 Introduction

Targeting single-stranded RNAs (ssRNAs) or partially structured RNAs by antisense oligonucleotides through duplex formation has shown great potential with clinical applications in fighting diseases such as Duchenne Muscular Dystrophy (DMD) and Spinal Muscular Atrophy (SMA)¹⁻³. Triplex-Forming Oligonucleotides (TFOs) can bind to double-stranded RNAs (dsRNAs) through TFO·RNA₂ triplex formation⁴. In a parallel major-groove RNA triplex, a TFO strand binds to the major groove of an RNA duplex through Hoogsteen and Hoogsteen-like base pairing interactions, such as T/U·A·U and C⁺·G·C base triples (**Figure 2.1a,b**)⁴. However, traditional TFO·RNA₂ triplexes are intrinsically not stable due to the charge-charge repulsion among the phosphate backbone of three strands and the relatively low pH required for the C⁺·G·C base triple formation⁵⁻¹¹. In addition, C-G and U-A base pairs are not easily recognized by natural bases^{8-9,12-16}. Furthermore, many of the TFOs containing unmodified and modified sugar-phosphate backbones bind more strongly to dsDNA than dsRNAs^{4-5,12,16-19}.

Peptide Nucleic Acids (PNAs) have a peptide-like backbone and a carbonyl-methylene linker between the PNA backbone and bases (**Figure 2**)²⁰⁻²¹, which have been recently shown to form stable PNA·RNA₂ triplexes (**Figure 2.2**)²²⁻²⁵,

presumably mainly due to the absence of charge-charge repulsion between the backbones of dsRNA-binding PNA (dbPNA) and dsRNA. In addition, the backbone of PNAs is chemically stable resulting in PNAs to be resistant against proteases and nucleases²⁶⁻²⁹. Furthermore, dbPNAs show sequence-specific binding to targeted RNA duplexes with significantly reduced binding to DNA duplexes^{22-25,30-32}.

We have shown that incorporating a modified base thio-pseudoisocytosine (L) (**Figure 2.1c**) into PNAs facilitates enhanced recognition of Watson-Crick G-C base pairs in dsRNAs at physiological pH with reduced pH dependence²³⁻²⁴. A steric clash between the sulphur atom in the L base and the amino group of guanine destabilizes a potential Watson-Crick L-G base pair, making the L-modified dbPNAs selective in binding to dsRNAs over ssRNAs.

Targeting the C-G, U-A, and U-G pairs is challenging due to a reduced number of hydrogen bond donors and acceptors on the Hoogsteen edge of C and U bases^{4,32-34}. PNAs incorporating guanidine-modified 5-methylcytosine (Q)¹⁴ show sequence-specific and selective binding to C-G pair-containing dsRNAs over dsDNAs (**Figure 2.1d**)²⁴. Although Q is a C derivative, Q does not form a Watson-Crick-like Q-G pair, mainly because alkylation of the amino group in the Q base results in a steric clash in a Watson-Crick-like Q-G pair^{24, 35-37}. Thus, Q can recognize the major groove of a C-G pair in dsRNAs but not a G in ssRNAs.

Figure 2.1. Chemical structures of base triples (a) T·A-U, (b) C⁺·G-C, (c) L·G-C, (d)

Q·C-G, (e) E·U-A, (f) proposed shifted E·U-A, (g) S·U-A, (h) S·C-G, and (i) S·U·G.

The letter R represents the sugar-phosphate backbone of RNA. Hydrogen bonding interactions are indicated by black dashed lines. Based on the modelling data (see **Figure 2.3**), a C-H···O hydrogen is proposed for S·U-A and S·U·G. The green dashed line indicates enhanced van der Waals interaction. The atoms shown in red may be involved in van der Waals interactions. Relacement of the hydrogen atom at 5 position of C in a C-G pair (see panels **d** and **h**) with a methyl group, results in the disruption of Q·C-G triple and S·C-G triple, probably due to steric clashes of the methyl group with Q and S bases, respectively.

PNAs incorporating 3-oxo-2,3-dihydropyridazine (E) monomer³⁸ have been utilized for targeting dsRNAs containing Watson-Crick U-A pairs (**Figure 2.1e**)³³. However, an E·U-A base triple may be stabilized by only one hydrogen bond formed between E base and U base in a U-A pair (**Figure 2.1e,f**). A modified base S has been incorporated into relatively long (e.g., 18-mer) DNAs and 2'-O-aminoethoxy (2'-AE) modified RNAs for the recognition of T-A base pairs in dsDNAs through TFO·DNA₂ triplex formation (**Figure 2.3**)³⁹⁻⁴¹. Here, we developed a synthetic route to the synthesis of a PNA monomer S containing a novel succinyl linker connecting S base and PNA backbone (**Figure 2.1g**), and studied the binding of S-modified short PNAs (8-mer) to dsRNAs, dsDNAs, and ssRNAs by non-denaturing polyacrylamide gel electrophoresis (PAGE), fluorescence titration, and thermal melting experiments. In addition, we tested how the linker connecting the PNA backbone and the T and S bases affects PNA·RNA₂ triplex formation.

Figure 2.3. Chemical structures of (a) DNA S·T-A (1,2) and (b) PNA S·U-A. The letters R and D represent the sugar-phosphate backbones of RNA and DNA, respectively. Hydrogen bonding interactions are indicated by black dashed lines. The methyl group of T base is shown in red.

2.2 Materials and methods

2.2.1 Molecular modelling

The S-containing triplex structure was modelled based on the previously reported method⁴². The triplex was put into a periodic box and solvated using TIP3P water model. After minimization and NVT (constant temperature and volume) /NPT (constant temperature and pressure) equilibrium stage, the system underwent a 20 ns product Molecular Dynamics (MD). The product MD trajectory was then clustered to form the representative structure. We used a docking simulation to modify the local structure involving S residue. The S residue in MD representative structure was removed, with its centre regarded as the docking centre. We set the docking pocket the space between the two neighbouring bases to force the plane of the expanded S base (including the carbonyl methylene attached to aniline) to be parallel to the neighboring bases. Thereafter, we docked the expanded S base into the docking pocket, with the expanded S base maintained as a planar structure (by fixing the rotational single bonds). The docked expanded S base and PNA backbone was then covalently connected by the carbonyl methylene linker, followed by optimization without bond angle constraints by energy minimization using Molecular Mechanics (MM)⁴³. The MD simulation, docking simulation, and the remaining procedure were conducted using GROMACS⁴⁴, Autodock Vina⁴⁵, and Discovery studio (Dassault Systèmes BIOVIA, San Diego CA, USA), respectively.

2.2.2 General methods

Reagents and solvents used were obtained from commercial sources and used without further purification. All organic reactions were monitored with the use of thin layer chromatography (TLC) using aluminum sheets silica gel 60 F254 (Merck). Compounds were purified by flash column chromatography using silica gel with ethyl

acetate/petroleum ether mixture as the eluting solvent. All ^1H and ^{13}C NMR spectra were obtained at room temperature on 300 MHz (75 MHz, ^{13}C), 400 MHz (100 MHz, ^{13}C) or 500 MHz Bruker spectrometers. The chemical shifts (δ) are shown in parts per million (ppm). The residual solvent peaks were used as references for the ^1H (chloroform- d : 7.26; dimethyl sulfoxide- d_6 : 2.50; methanol- d_4 : 3.31) and ^{13}C (chloroform- d : 77.0; dimethyl sulfoxide- d_6 : 39.5; methanol- d_4 : 49.1) NMR spectra. Mass spectra of the compounds were obtained *via* liquid chromatography-mass spectroscopy with electrospray ionization source (LCMS-ESI) and high-resolution mass spectrometry (electron ionization) (HRMS-EI). Reverse-phase high performance liquid chromatography (RP-HPLC) purified RNA and DNA oligonucleotides were purchased from Sigma-Aldrich Singapore. The RNA containing 5-methyl C was synthesized by a DNA/RNA synthesizer and purified by HPLC.

2.2.3 Synthesis of PNA monomers

The detailed synthesis procedures for the PNA monomers are shown in **Appendix 1** and **Appendix 2**. The PNA monomers were synthesized in few steps of reactions shown in **Scheme 2.1**. The artificial S base (compound **1**) was synthesized based on the previously reported methods^{39, 46}. Commercially available methyl 4-chloro-4-oxobutanoate was attached to compound **1** as a linker (compound **2**). Hydrolysis was carried out on compound **2** using aqueous lithium hydroxide and hydrochloric acid to give compound **3**. In the presence of HBTU and NMM in DMF, compound **3** was coupled with PNA backbone benzyl (2-((*tert*-butoxycarbonyl)amino)ethyl)glycinate to give the benzyl ester compound **4**. Catalytic hydrogenation of compound **4** yielded PNA monomer **S** (compound **5**).

To synthesize PNA monomer **S**₁, malonyl dichloride was attached to compound **1**, after which PNA backbone benzyl (2-((*tert*-butoxycarbonyl)amino)ethyl)glycinate was added in the presence of HBTU and NMM in DMF to give compound **6**. The desired PNA monomer **S**₁ (compound **7**) was obtained through the catalytic hydrogenation of compound **6**.

2.2.3.1 Synthesis of PNA monomer **S**

Methyl 4-((3-(2-acetamidothiazol-4-yl)phenyl)amino)-4-oxobutanoate (2). Under N₂ atmosphere, methyl 4-chloro-4-oxobutanoate (5.246 mL, 42.9 mmol) was added dropwise into an ice-cold stirring solution of compound **1** (2 g, 8.58 mmol) in anhydrous DCM (70 mL) and the mixture was cooled in an ice bath. Upon the completion of the addition, the reaction mixture was stirred for 2 h under room temperature. DCM was removed under reduced pressure followed by extraction with ethyl acetate (100mL, 2×). The combined organic phase was washed with brine (50 mL) and dried over Na₂SO₄. Product was purified by flash column chromatography to obtain 2.68 g of product with an isolated yield of 90%. ¹H NMR (300 MHz, DMSO-d₆): δH 12.28 (s, 1H), 10.09 (s, 1H), 8.30 (s, 1H), 7.55-7.49 (m, 2H), 7.38-7.30 (m, 2H), 3.60 (s, 3H), 2.62 (s, 4H), 2.16 (s, 3H); ¹³C NMR (75 MHz, DMSO-d₆): δC 172.94, 169.91, 168.71, 157.93, 148.75, 139.59, 134.82, 129.00, 120.50, 118.45, 116.66, 107.96, 51.38, 30.84, 28.52, 22.50; ESI-MS: *m/z*: calculated for C₁₆H₁₈N₃O₄S [M+H]⁺ 348.10, found: 348.10.

4-((3-(2-acetamidothiazol-4-yl)phenyl)amino)-4-oxobutanoic acid (3). Aqueous 1 M LiOH was added dropwise to an ice-cold solution of **2** (2 g, 5.76 mmol) dissolved in THF (30 mL). The reaction was left to stir for 2 h at room temperature and the completion of the reaction was checked by TLC. The reaction mixture was then

neutralized by 2 M aqueous HCl. Product was obtained as precipitate after filtration by using a Büchner funnel. The product was washed with water and dried in a vacuum desiccator to get 1.5 g of product with an isolated yield of 80%. ¹H NMR (400 MHz, MeOH-d₄): δH 8.14 (s, 1H), 7.63-7.61 (d, 1H), 7.45-7.43 (d, 1H), 7.34-7.30 (m, 2H), 2.69 (s, 4H), 2.22 (s, 3H); ¹³C NMR (100 MHz, MeOH-d₄): δC 176.86, 172.93, 170.91, 159.73, 150.92, 140.26, 136.64, 130.01, 122.79, 120.57, 118.86, 108.84, 32.42, 30.07, 22.58; ESI-MS: *m/z*: calculated for C₁₅H₁₆N₃O₄S [M+H]⁺ 334.09, found: 334.09.

Benzyl N-(4-((3-(2-acetamidothiazol-4-yl)phenyl)amino)-4-oxobutanoyl)-N-(2-((tert-butoxycarbonyl)amino)ethyl)glycinate (4). NMM (1.2 mL, 9.7 mmol) and HBTU (1.6 g, 4.2 mmol) were added to a solution of **3** (1 g, 3.2 mmol) dissolved in DMF (20 mL) under N₂. The acid was activated by stirring at room temperature for 15 mins. *Benzyl (2-((tert-butoxycarbonyl)amino)ethyl)glycinate* (1.1 g, 3.6 mmol) was added to the above reaction mixture. The reaction was allowed to stir for another 12 h at room temperature, after which the reaction mixture was extracted by ethyl acetate (100 mL × 3). The combined organic phase was washed with saturated aqueous KHSO₄ (50 mL), brine (50 mL), and dried over Na₂SO₄. Product was purified by flash chromatography to afford 2.1 g of product with an isolated yield of 60 %. ¹H NMR (500 MHz, CDCl₃). This compound exists as two rotamers. The chemical shifts for the well separated minor rotamer are given in brackets): δH 10.75 (10.85) (s, 1H), 8.83 (8.92) (s, 1H), 7.94 (s, 1H), 7.53-7.27 (m, 7H), 7.25-7.21 (m, 1H), 7.04 (7.05) (s, 1H), 5.09 (5.17) (s, 2H), 4.05 (4.15) (s, 2H), 3.51-3.08 (m, 4H), 2.78 (2.84, 2.65) (s, 4H), 2.14 (s, 3H), 1.39 (1.45, 1.37) (s, 9H); ¹³C NMR (100 MHz, CDCl₃): δC 173.63, 171.30, 169.96, 169.43, 168.86, 156.25, 149.31, 138.80, 135.22, 134.96, 129.21,

128.86, 128.77, 128.64, 128.57, 128.38, 121.57, 119.74, 118.11, 108.04, 79.91, 67.32, 49.44, 49.26, 38.98, 32.59, 29.83, 28.49, 23.13; ESI-MS: m/z : calculated for $C_{31}H_{38}N_5O_7S$ $[M+H]^+$ 624.25, found: 624.25.

N-(4-((3-(2-acetamidothiazol-4-yl)phenyl)amino)-4-oxobutanoyl)-*N*-(2-((tert-butoxycarbonyl)amino)ethyl)glycine (**5**). Benzyl ester compound **4** (1g, 1.6 mmol) was dissolved in MeOH:EtOH (100 mL, 1:1) and purged with N_2 . 10% Pd/C (100 mg) was added and the solution was saturated with H_2 over 6 h at room temperature. The reaction mixture was filtered through a short pad of Celite and the pad was washed with MeOH (50 mL). The solvent was removed under reduced pressure to give 770 mg of product with an isolated yield 70%. 1H NMR (500 MHz, MeOH- d_4). This compound exists as two rotamers. The chemical shifts for the well separated minor rotamer are given in brackets.): δH 8.13 (8.55) (s, 1H), 7.62-7.60 (d, 1H), 7.48-7.46 (d, 1H), 7.33-7.29 (m, 2H), 3.97 (3.94) (s, 2H), 3.52-3.48 (m, 2H), 3.28-3.15 (m, 2H), 2.70 (2.84) (s, 4H), 2.22 (2.42) (s, 3H), 1.41 (1.43) (s, 9H); ^{13}C NMR (100 MHz, MEOH- d_4): δC 180.31, 176.23, 175.27, 173.47, 170.92, 159.36, 150.97, 140.35, 136.58, 129.95, 122.70, 120.65, 118.91, 108.80, 80.01, 54.35, 51.40, 39.46, 33.17, 30.80, 28.76, 24.15; ESI-MS: m/z : calculated for $C_{24}H_{32}N_5O_7S$ $[M+H]^+$ 534.20, found: 534.20.

2.2.3.2 Synthesis of PNA monomer **S₁**

Benzyl N-(3-((3-(2-acetamidothiazol-4-yl)phenyl)amino)-3-oxopropanoyl)-*N*-(2-((tert-butoxycarbonyl)amino)ethyl)glycinate (**6**). Under N_2 atmosphere, compound **1** (1.5 g, 6.4 mmol) was added portion wise into an ice-cold stirring solution of malonyl dichloride (3.1 mL, 32 mmol) in anhydrous DCM (70 mL) and the mixture was stirred in ice bath for 2 h. The reaction progress was monitored by TLC. Upon the reaction

completion, DCM was evaporated under reduced pressure and the reaction mixture was co-evaporated by DCM several times. The reaction mixture was then dissolved in anhydrous DCM (70 mL) and stirred in ice bath. Under N₂ atmosphere, Benzyl (2-((*tert*-butoxycarbonyl)amino)ethyl)glycinate (2.2 g, 7.1 mmol) and TEA (1.8 mL, 13 mmol) were added into the above ice-stirring solution and the mixture was stirred in ice bath for 2 h. The reaction mixture was washed with water, brine and dried over Na₂SO₄. Product was purified by flash chromatography to afford 1.2 g of product with an isolated yield of 30%. ¹H NMR (500 MHz, MeOH-d₄). This compound exists as two rotamers. The chemical shifts for the well separated minor rotamer are given in brackets.): δH 8.15 (8.14) (s, 1H), 7.65-7.63 (d, 1H), 7.49-7.46 (m, 1H), 7.37-7.29 (m, 7H), 5.19 (5.20) (s, 2H), 4.21 (4.42) (s, 2H), 3.67-3.51 (m, 4H), 3.30-3.27 (3.24-3.21) (m, 2H), 2.22 (s, 3H), 1.43 (1.41) (s, 9H); ¹³C NMR (100 MHz, Acetone-d₆): δC 171.35, 170.40, 169.74, 169.04, 165.33, 158.93, 150.21, 140.31, 137.09, 136.33, 129.83, 129.40, 129.37, 129.12, 128.98, 128.88, 121.90, 119.56, 117.99, 108.34, 79.27, 67.22, 49.83, 49.09, 43.19, 39.56, 28.62, 22.84; ESI-MS: *m/z*: calculated for C₃₀H₃₆N₅O₇S [M+H]⁺ 610.23, found: 610.23.

N-(3-((3-(2-acetamidothiazol-4-yl)phenyl)amino)-3-oxopropanoyl)-*N*-(2-((*tert*-butoxycarbonyl)amino)ethyl)glycine (**7**). Benzyl ester compound **6** (1g, 1.6 mmol) was dissolved in MeOH:EtOH (100 mL, 1:1) and purged with N₂. 10% Pd/C (100 mg) was added and the solution was saturated with H₂ over 6 h at room temperature. The reaction mixture was filtered through a short pad of Celite and the pad was washed with MeOH (50 mL). The solvent was removed under reduced pressure to give 767 mg of product with an isolated yield of 70%. ¹H NMR (500 MHz, MeOH-d₄). This compound exists as two rotamers. The chemical shifts for the well separated minor

rotamer are given in brackets.): δ H 8.15 (8.09) (s, 1H), 7.64-7.61 (d, 1H), 7.53-7.51 (d, 1H), 7.35-7.30 (m, 2H), 4.01 (s, 2H), 3.64-3.48 (m, 3H), 3.28-3.26 (m, 3H), 2.23 (s, 3H), 1.39 (1.42) (s, 9H); ^{13}C NMR (100 MHz, MeOH- d_4): δ C 175.99, 171.00, 170.11, 167.89, 163.38, 158.45, 150.81, 140.06, 136.56, 130.02, 122.93, 120.60, 118.82, 108.94, 80.16, two methylene carbons are merged into MeOH- d_4 residual peak, 42.91, 39.55, 28.73, 22.61; ESI-MS: m/z : calculated for $\text{C}_{23}\text{H}_{30}\text{N}_5\text{O}_7\text{S}$ $[\text{M}+\text{H}]^+$ 520.19, found: 520.19.

2.2.4 Synthesis of PNA oligomers

The PNA thymine (T) and cytosine (C) monomers were purchased from ASM Research Chemicals. PNA monomer L was synthesized following the reported method²³. PNA oligomers were synthesized manually using Boc chemistry via a Solid-Phase Peptide Synthesis (SPPS) protocol. 4-Methylbenzhydrylamine hydrochloride (MBHA·HCl) polystyrene resins were used. The loading value used for the synthesis of the oligomers was 0.3 mmol/g and acetic anhydride was used as the capping reagent. Benzotriazol-1-yl-oxytripyrrolidinophosphonium hexafluorophosphate (PyBOP) and *N,N*-diisopropylethylamine (DIPEA) were used as the coupling reagent. The oligomerization of PNA was monitored by Kaiser test. Cleavage of the PNA oligomers was done using trifluoroacetic acid (TFA) and trifluoromethanesulfonic acid (TFMSA) method, after which the oligomers were precipitated with diethyl ether, dissolved in deionized water and purified by reverse-phase high performance liquid chromatography (RP-HPLC) using water–ACN–0.1% TFA as the mobile phase. Matrix-assisted laser desorption/ionization-time of flight (MALDI-TOF) analysis was used to characterize the oligomers (**Table A2.1, Figure A2.1**), with the use of α -cyano-4-hydroxycinnamic acid (CHCA) as the sample crystallization matrix.

2.2.5 Non-denaturing polyacrylamide gel electrophoresis

Non-denaturing (12wt%) polyacrylamide gel electrophoresis (PAGE) experiments were conducted with incubation buffers containing 200 mM NaCl, 0.5 mM EDTA, 20 mM MES at pH 6.0, or 200 mM NaCl, 0.5 mM EDTA, 20 mM HEPES at pH 7.5 or pH 8.0. The loading volume for samples containing RNA and DNA hairpins were 20 μ L and 10 μ L, respectively. The concentration of RNA and DNA are 1 or 0.25 μ M. The samples were prepared by snap cooling of the hairpin, followed by annealing with PNA oligomers by slow cooling from 65°C to room temperature and incubation at 4°C overnight. Prior to loading the samples into the wells, 35% glycerol (20% of the total volume) was added to the sample mixtures. 1 $\times$ Tris–borate–EDTA (TBE) buffer, pH 8.3 was used as the running buffer for all gel experiments. The gel was run at 4°C at 250 V for 5 h. The gels were then stained with ethidium bromide and imaged by the Typhoon Trio Variable Mode Imager.

2.2.6 Fluorescence titration binding study

Fluorescence emission spectra were recorded on a Varian Cary Eclipse fluorescence spectrophotometer at room temperature using a 1 cm square cuvette. All samples contain 1 μ M of 2-aminopurine-labeled double-stranded RNA or DNA (dsRNA or dsDNA) in 70 μ L of incubation buffer. The PNA concentration ranges from 0 to 50 μ M. The incubation buffers used is 200 mM NaCl, 0.5 mM EDTA, 20 mM HEPES at pH 7.5. The samples containing the RNA duplex and PNA were prepared by slow cooling of the RNA duplex from 95°C to room temperature, followed by annealing with PNA oligomers at room temperature for 1–2 h and incubation at 4°C overnight. The emission spectra of 2-aminopurine were measured at room temperature over a wavelength range of 330–550 nm with an excitation wavelength of 303 nm.

2.2.7 Thermal melting

UV-absorbance-detected thermal melting experiments were conducted using the Shimadzu UV-2550 UV-Vis spectrophotometer with the use of an 8-microcell cuvette. The absorbance at 260 nm was recorded with the temperature increasing from 15 to 95°C followed by the temperature decreasing from 95 to 15°C. The temperature ramp rate is 0.5°C/min. The optical path length of the 8-microcell cuvette is 1 cm. All samples contain 5 μ M RNA and 5 μ M PNA in 130 μ L buffer. The incubation buffer is 200 mM NaCl, 0.5 mM EDTA, 20 mM NaH₂PO₄, pH 7.5. The samples containing the single-stranded RNA and PNA were annealed by slow cooling from 95°C to room temperature, followed by incubation at 4°C overnight. Data were normalized at high temperature and the melting temperatures were determined based on the Gaussian fit of the first derivative of the curves.

Scheme 2.1. Synthesis schemes of PNA S (compound 5) and S₁ (compound 7) monomers. (i) Methyl 4-chloro-4-oxobutanoate, DCM, 0°C- room temperature (rt), 2 h, 90%. (ii) 1 M aq. LiOH, tetrahydrofuran (THF), rt, 1 h, 1 M HCl, 0°C, 80%. (iii) Benzyl (2-((*tert*-butoxycarbonyl)amino)ethyl)glycinate, Hexafluorophosphate

Benzotriazole Tetramethyl Uronium (HBTU), *N*-methylmorpholine (NMM), anhydrous dimethylformamide (DMF), rt, 12 h, 60%. (iv) H₂, 10% Pd/C, MeOH:EtOH (1:1), rt, 6h, 70%. (v) Malonyl dichloride, DCM, 0°C, rt, 2h, Benzyl (2-((*tert*-butoxycarbonyl)amino)ethyl)glycinate, Triethylamine (TEA), DCM, 0°C, 2 h, 30%. (vi) H₂, 10% Pd/C, MeOH:EtOH (1:1), rt, 6h, 70%.

2.3 Results and Discussion

2.3.1 Recognition of an internal C-G, U-A, and U-G pair in an RNA duplex (but not in DNA duplex) by S-modified PNAs

To enhance the PNA·RNA₂ triplex formation at near-physiological conditions (e.g., 200 mM NaCl, pH 7.5), we incorporated into a series of short PNAs (**Figure 2.2l, o-r, t**) a previously developed PNA monomer L for enhanced recognition of G-C pairs (**Figure 2.1c**)^{23, 47-48}. The non-denaturing PAGE result suggests that PNA P11 (**Figure 2.2o**), which contains both S and L modifications, binds to rHP2 ($K_d = 10.6 \pm 1.9 \mu\text{M}$), rHP4 ($K_d = 9.3 \pm 2.0 \mu\text{M}$) and rHP8 ($K_d = 7.5 \pm 1.5 \mu\text{M}$) to form a PNA·RNA₂ triplexes in a near physiological buffer (200 mM NaCl, pH 7.5) (**Figures 2.5 and 2.7b, Table 2.1, Figure 2.6a-c**). Thus, the S residue in a PNA can facilitate the binding to RNA duplexes containing a Watson-Crick C-G pair (rHP2), Watson-Crick U-A pair (rHP4), or Wobble U-G pair (rHP8). We propose that an S base may form three hydrogen bonds with a C-G base pair and four hydrogen bonds with both U-A and U-G base pairs, respectively (**Figure 2.1g-i**). Our modelling result reveals that an S·U-A triple can form and is compatible within a PNA·RNA-RNA triplex (**Figure 2.4**). The 2-aminothiazole moiety in the S residue may be protonated⁸, induced by the recognition of a G base in a C-G and a U-G pair (**Figure 2.2h,i**). Our non-denaturing PAGE data reveal that lowering the pH from 8.0 to 6.0 moderately (within 6 fold) enhances the binding of P11 (NH₂-Lys-TLTSTTTT-CONH₂) to the studied dsRNAs (**Table 2.1, Figures 2.8—2.11**). The L residue has minimal pH dependence as compared to the unmodified C base²³⁻²⁴. The data may thus indicate that the protonation and tautomerization can occur near physiological pH for the formation of S·C-G and S·U·G triples in addition to an S·U-A triple, with the main driving force resulting from stabilizing base triple formation (**Figure 2.1g-i**) as has been observed

for other nucleic acid structure-induced protonation^{7, 49-60} and tautomerization⁶¹⁻⁶⁴. It is possible that the RNA base pairs may form tautomers induced by triple formation as well⁶⁵⁻⁶⁶.

The favourable non-covalent interaction between carbonyl oxygen and ring sulfur atom⁶⁷⁻⁷¹ may allow the acetylated 2-aminothiazole group to be pre-oriented for hydrogen bonding with purine bases in RNA (**Figures 2.1 and 2.4**). In addition, we propose that the amide bond linking succinyl linker and aniline moiety in S base may form a tautomer induced upon triple formation with a C-G pair (**Figure 2.2h**). The proposed base triple structure formation results in minimal distortion in the linker connecting the PNA base and backbone (**Figure 2.4**, see below on the critical importance of the linker). Thus, our non-denaturing PAGE data suggest that it is possible to incorporate S residues into dsRNA-binding PNAs for the recognition of dsRNAs containing canonical pairs C-G, U-A, and U-G pairs (**Figure 2.2b,d,e,f,h**).

Figure 2.4. Modelled three-dimensional structure and potential stacking pattern of a PNA•RNA-RNA triples involving S base. (a) PNA•RNA-RNA triplex construct used for modelling. We used the PNA sequence of AcNH-TLTSTTTL-CONH₂ with the N-terminus capped by an acetyl group based on a previously modelled structure⁵⁸. (b,c) Potential stacking patterns of PNA•RNA-RNA triples involving S base based on the modelled triplex structure (panel a). The stacking patterns of the base triples involving PNA residues from positions 3-5 are shown. The base triple with the carbon atoms shown in green is closer to the viewer. In a TS step (see panel b), T base is stacked with S through the amide bond⁷²⁻⁷⁴ attached to aniline moiety. In a ST step (see panel c), however, there is essentially no stacking interactions formed between S and T.

Figure 2.5. Non-denaturing PAGE with an incubation buffer of 200 mM NaCl, 0.5 mM EDTA, 20 mM HEPES, pH 7.5. The running buffer is 1× TBE, pH 8.3. The loaded RNA hairpins are at 1 μM in 20 μL . The PNA concentrations in lanes from left to right are 0, 0.2, 0.4, 1, 1.6, 2, 4, 10, 16, 20, 28 and 50 μM , respectively. (a-e) PNA P11 binds to rHP2, rHP4 and rHP8 with K_d values of (10.6 ± 1.9), (9.3 ± 2.0), and (7.5 ± 1.5) μM , respectively. (f) PNA P3 shows no binding to rHP8. (g-j) PNA P12 shows binding to rHP5 (C-G) and rHP7 (U-A) with K_d values of (11.9 ± 2.2) and (6.0 ± 1.5) μM , respectively. The gels shown are representative gels. The errors given are standard errors.

Figure 2.6. K_d determination by non-denaturing PAGE for (a-c) P11 and (d-e) P12 binding to various RNA hairpins at pH 7.5. The fraction of triplex formation was calculated according to $Y = I_{\text{triplex}} / (I_{\text{triplex}} + aI_{\text{duplex}})$. The band intensities were normalized according to $a = I_{\text{triplex max}} / I_{\text{duplex max}}$. $I_{\text{duplex max}}$ is the band intensity for hairpin alone without the addition of PNA. $I_{\text{triplex max}}$ is the triplex band intensity with the highest concentration of PNA added. The data were fit to the equation: $Y = Y_0 +$

$(B/2/R_0)(R_0 + X + K_d - ((R_0 + X + K_d)^2 - 4R_0X)^{1/2})$ where R_0 is the RNA hairpin concentration (1 μ M). Y_0 and B are the minimum and maximum fraction of triplex formation respectively. X is the total PNA concentration and K_d is the dissociation constant.

Table 2.1. K_d (μ M) values for PNA binding to RNA or DNA duplexes obtained by non-denaturing PAGE.^a

	rHP1 (G-C)	rHP2 (C-G)	rHP3 (A-U)	rHP4 (U-A)	rHP8 (U-G)	rHP5 (C-G)	rHP6 (G-C)	rHP7 (U-A)	dHP1	dHP2	dHP4
P10 (S)	NB	NB	NB	NB	-	-	-	-	-	-	-
P10-S ₁	NB	NB	NB	NB	-	-	-	-	-	-	-
P11 (S)	NB	(9.4 $\pm$ 3.0) 10.6 $\pm$ 1.9 (26.7 $\pm$ 5.2)	(>28) >50	(5.8 $\pm$ 1.5) 9.3 $\pm$ 2.0 (32.8 $\pm$ 11.9)	(5.9 $\pm$ 1.6) 7.5 $\pm$ 1.5 (19.3 $\pm$ 4.5)	-	-	-	NB	NB	NB
P11-S ₁	NB	>50	NB	>20	>50	-	-	-	-	-	NB
P11-E	NB	NB	2.6 $\pm$ 0.5	3.4 $\pm$ 0.8	>50	-	-	-	-	-	NB
P12 (S)	>50	-	-	-	-	(4.7 $\pm$ 1.5) 11.9 $\pm$ 2.2 (25.6 $\pm$ 7.5)	NB	(3.0 $\pm$ 0.8) 6.0 $\pm$ 1.5 (9.8 $\pm$ 3.5)	NB	-	-
P13	NB	NB	0.2 $\pm$ 0.1 ^c	NB	-	-	-	-	-	-	-
P13-xT	NB	NB	NB	NB	-	-	-	-	-	-	-
P13-xxT	NB	NB	NB	NB	-	-	-	-	-	-	-
P3 (Q) ^b	NB	(1.2 $\pm$ 0.2) 4.4 $\pm$ 0.5 (8.8 $\pm$ 1.4)	NB	NB	NB	-	-	-	NB	NB	-
P6 (Q) ^b	NB	-	-	-	-	1.6 $\pm$ 0.2	NB	>50	-	-	-

^aThree incubation buffers are: 200 mM NaCl, 0.5 mM EDTA, 20 mM MES, pH 6.0, or 200 mM NaCl, 0.5 mM EDTA, 20 mM HEPES, pH 7.5 or pH 8.0. The data shown in parentheses were taken at pH 6.0 (top) and pH 8.0 (bottom). “-” represents that the data were not obtained. “NB” indicates that no binding was observed. The RNAs and DNAs were at 1 μ M (20 μ L) and 0.25 μ M (10 μ L), respectively, unless otherwise noted. P13-xT and P13-xxT have the fourth T residue in P13 replaced with the

original carbonyl methylene linker lengthened with one (^xT) and two (^{xx}T) carbon atoms, respectively (**Figure S17a-c**). ^bData obtained from previous study ²⁴ except P3 binding to rHP8. ^crHP3 was kept at 0.25 μ M in 25 μ L.

Figure 2.7. Summary of binding properties obtained by PAGE. The incubation buffer is 200 mM NaCl, 0.5 mM EDTA, 20 mM HEPES, pH 7.5. Those with measured $K_d > 50 \mu$ M are indicated at a value at 50 μ M. **(a)** PNAs P5 and P13 binding to various RNA hairpins. RNA hairpins were kept at 0.25 μ M for the PAGE assay. P5 has the sequence of $\text{NH}_2\text{-Lys-TLTLTTTL-CONH}_2$ ²³⁻²⁴. **(b)** PNAs P3, P11, and P11-E binding to various RNA hairpins. RNA hairpins were kept at 1 μ M for the PAGE assay.

Remarkably, PNA P11 shows no or weak binding to other RNA hairpins with the C-G, U-A, and U-G pairs replaced with a G-C pair (rHP1) or A-U pair (rHP3) (**Figure 2.5a,c and 2.7b, Table 2.1**). The data suggest that S residue is not simply an intercalator non-specifically intercalating into RNA duplexes⁷⁵⁻⁷⁹. In addition, PNA P11 does not show binding to all the studied DNA hairpins with the sequence homologous to the studied RNA hairpins (**Figures 2.2i-k, Table 2.1, Figure 2.8**), presumably because the major groove of a DNA duplex is not structurally compatible

for short PNA binding. The selective binding of PNA P11 to dsRNA over dsDNA observed here is consistent with previous non-denaturing PAGE and isothermal titration calorimetry (ITC) studies^{22-24, 33}.

Figure 2.8. Non-denaturing PAGE for (a-c) P11, (d) P12 and (e) P11-S₁ binding to various DNA hairpins. The incubation buffer is 200 mM NaCl, 0.5 mM EDTA, 20 mM HEPES, pH 7.5. The loaded samples contain 1 μM of DNA hairpins in 10 μL. The PNA concentrations in lanes from left to right are 0, 0.2, 0.4, 1, 1.6, 2, 4, 10, 16, 20, 28, and 50 μM, respectively.

Our non-denaturing PAGE data reveal that substitution of the single S residue in P11 with a Q residue (P3, NH₂-LysTLTQT²⁴TTTL-CONH₂, **Figure 2.2l**) and an E residue (P11-E, NH₂-Lys-TLTETTTTL-CONH₂, **Figure 2.2q**) causes no or weak binding to RNA hairpin with U-G pair (rHP8) (**Figure 2.7, Table 2.1, Figures 2.9 and 2.10**). The results show that, compared to the previously reported PNA Q and E monomers which are able to recognize Watson-Crick C-G and U-A base pair, respectively^{24, 33}, our PNA S monomer is advantageous for the recognition of a dsRNA region containing a Wobble U-G base pair (**Figure 2.7**).

Figure 2.9. Proposed structures of (a) E·A-U and (b) shifted E·U-A. The letter R represents the sugar-phosphate backbone of RNA. Hydrogen bonding interactions are indicated by black dashed lines. (c-h) Non-denaturing PAGE data for P11-E (NH₂-Lys-TLTETTTTL-CONH₂) binding to various RNA hairpins and a DNA hairpin. The incubation buffer is 200 mM NaCl, 0.5 mM EDTA, 20 mM HEPES, pH 7.5. The loaded samples contain 1 μM of RNA hairpins and DNA hairpin in 20 and 10 μL , respectively. The PNA concentrations in lanes from left to right are 0, 0.2, 0.4, 1, 1.6, 2, 4, 10, 16, 20, 28, and 50 μM , respectively.

Figure 2.10. K_d determination by non-denaturing PAGE for P11-E binding to (a) rHP3, (b) rHP4, and (c) P13 binding to rHP3 at pH 7.5. The fraction of triplex formation was calculated according to $Y = I_{\text{triplex}} / (I_{\text{triplex}} + aI_{\text{duplex}})$. The band intensities were normalized according to $a = I_{\text{triplex max}} / I_{\text{duplex max}}$. $I_{\text{duplex max}}$ is the band intensity for hairpin alone without the addition of PNA. $I_{\text{triplex max}}$ is the triplex band intensity with the highest concentration of PNA added. The data were fit to the equation: $Y = Y_0 + (B/2/R_0)(R_0 + X + K_d - ((R_0 + X + K_d)^2 - 4R_0X)^{1/2})$ where R_0 is the RNA hairpin concentration. Y_0 and B are the minimum and maximum fraction of triplex formation respectively. X is the total PNA concentration and K_d is the dissociation constant.

We carried fluorescence titration studies to further confirm the binding affinities for the PNA·RNA₂ formation in solution. A fluorescent 2-aminopurine residue was incorporated into dsRNAs (**Figure 2.11**) for the fluorescence measurements. It is

known that the fluorescence signal of 2-aminopurine is sensitive to its local structural environment^{24, 35, 80-81}. We observed quenching of the 2-aminopurine fluorescence intensity upon the binding of PNA P11 (**Figure 2.12a-d, Figure 2.13**), which is consistent with our previous studies^{24, 35}. Our fluorescence titration data reveal that PNA P11 binds relatively weakly to dsRNA1-2AP ($K_d > 20 \mu\text{M}$) and dsRNA3-2AP ($K_d = 7.7 \pm 1.1 \mu\text{M}$), suggesting S base does not show strong recognition of G-C and A-U pairs (**Figure 2.12a-b, Figure 2.13**). However, PNA P11 (NH₂-Lys-TLTSTTTL-CONH₂) shows relatively stronger binding to dsRNA2-2AP (forming an S·C-G triple, $K_d = 3.9 \pm 1.1 \mu\text{M}$), dsRNA4-2AP (forming an S·U-A triple, $K_d = 2.8 \pm 0.4 \mu\text{M}$), and dsRNA8-2AP (forming an S·U-G triple, $K_d = 2.4 \pm 0.5 \mu\text{M}$) in a near physiological buffer (200 mM NaCl, pH 7.5) (**Figure 2.12c-d, Figure 2.13**). Substitution of the single S residue in P11 with a Q residue (P3, NH₂-LysTLTQTTTL-CONH₂, **Figure 2.21**)²⁴ results in weak binding to dsRNA8-2AP, suggesting no stable formation of a Q·U-G triple, $K_d = 13.0 \pm 1.3 \mu\text{M}$, **Figure 2.12e,f**). Thus, consistent with our PAGE data, a U-G pair in dsRNAs is recognized by S base but not Q base. The fluorescence titration data suggest no binding between PNA P11 and dsDNAs (dsDNA1-2AP and dsDNA2-2AP) (**Figure 2.13**), consistent with our PAGE results (**Table 2.1**).

Figure 2.11. Constructs for fluorescence titration study. The 2-aminopurine residue is designated as “2” in the dsRNAs and dsDNAs. **(a-e)** Model 2-aminopurine-labelled dsRNAs. **(f-g)** Model 2-aminopurine-labelled dsDNAs. **(h)** PNA P11. **(i)** A PNA·RNA₂ triplex formed between PNA P11 and dsRNA4-2AP.

Figure 2.12. Fluorescence titration study of PNAs P11 and P3 binding to 2-aminopurine-labelled dsRNAs. The 2-aminopurine-labelled dsRNA were kept at 1 μM. The buffer is 200 mM NaCl, 0.5 mM EDTA, 20 mM HEPES, pH 7.5. The K_d values were fit based on the plots of 2-aminopurine fluorescence intensity at 370 nm. The peak at 475 nm is due to the weak fluorescence emission of the L base in the PNA. **(a,b)** P11 shows weak binding to dsRNA1-2AP. **(c,d)** P11 shows binding to dsRNA8-2AP. **(e,f)** P3 shows relatively weakened binding to dsRNA8-2AP.

Figure 2.13. Fluorescence titration study of PNA P11 binding to various 2-aminopurine-labelled dsRNAs and dsDNAs. The 2-aminopurine residue is designated as “2” in the dsRNA and dsDNA sequences. **(a,b)** P11 binds to dsRNA2-2AP. **(c,d)** P11 shows relatively weakened binding to dsRNA3-2AP. **(e,f)** P11 shows binding to dsRNA4-2AP. **(g-j)** P11 shows no binding to dsDNA1-2AP and dsDNA2-2AP.

2.3.2 Effects of linker length on PNA·RNA₂ triplex formation

The S base is connected to the PNA backbone via a succinyl linker, which has the same length as a deoxyribose-containing linker (**Figure 2.3**)³⁹. We synthesized a derivative of PNA S monomer (S₁) by replacing the succinyl linker with a malonyl linker (**Scheme 2.1**). The substitution of the single S residue in P11 with an S₁ residue (PNA P11-S₁, see **Figure 2.2p**) results in significantly weakened binding to all the tested RNA hairpins including rHP1, rHP2, rHP3, rHP4, and rHP8 as compared to P11 (**Figure 2.14i-m**). In addition, PNA P11-S₁ does not show binding to the DNA hairpin with a T-A pair opposite to S₁ (dHP4) (**Figures 2.2k, Table 2.1, Figure 2.8e**). The data demonstrate that the number of carbon atoms in the linker connecting the artificial nucleobase S and PNA backbone is crucial in forming a stable PNA·RNA₂ triplex via Hoogsteen base pairing. Our modelling result suggests that having a succinyl linker may allow the S base to adopt an optimized orientation and backbone-base distance for the formation of S·U-A, S·C-G and S·U-G base triples, respectively (**Figure 2.1g-i**). Critically, the carbonyl methylene connected to aniline seems to expand the planar structure of S base and shows amide stacking interactions⁷²⁻⁷⁴ in a TS step within the PNA of a PNA·RNA-RNA triplex. The importance of the length of the linker observed here may also support our proposed base triple geometries, because formation of the three triples causes no distortion on the linker (**Figures 2.1g-i and 2.4**).

In order to further investigate how the length of the linker affects the recognition of RNA duplexes by PNAs, we synthesized two modified PNA thymine monomers by replacing the original carbonyl methylene linker with lengthened linkers with one (^xT) and two (^{xx}T) carbon atoms, respectively (**Figure 2.15a-c**)⁸². The linker length in T monomer may affect the recognition of a Watson-Crick A-U base pair through T·A-U base triple formation (**Figure 2.1A**). We carried non-denaturing PAGE binding studies of 8-mer PNAs incorporating T (P13, **Figure 2.2s**), ^xT (P13-^xT), and ^{xx}T (P13-^{xx}T), respectively, to an RNA duplex containing (rHP3). PNA P13 binds to rHP3 ($K_d = 0.2 \pm 0.1 \mu\text{M}$) to form a PNA·RNA₂ triplex in a near physiological buffer (200 mM NaCl, pH 7.5) (**Figure 2.7a, Table 2.1, Figures 2.15d-g and 2.16c**). Significantly, the replacement of the single thymine at position 4 in P13 by ^xT (P13-^xT) and ^{xx}T (P13-^{xx}T) results in no binding toward rHP3 (**Table 2.1, Figure 2.15h-o, 2.16**). It is possible that lengthening of the linker with one and two carbons may shift the T base away from the A base in an A-U base pair resulting in the disruption of a Hoogsteen T·A pair (**Figure 2.1a**). In addition, a lengthened linker may increase the flexibility in the ^xT and ^{xx}T residues, resulting in an increased entropy penalty^{75, 82} for base triple formation.

It was previously reported that, replacing the a standard carbonyl methylene linker in a P monomer (2-pyrimidinone base with carbonyl methylene linker) with a carbonyl ethylene linker (P_{ex}) results in enhanced recognition of a C-G pair, with reduced specificity though³³. It is possible that the P_{ex} residue may adopt different base triple structures with an elongated linker. Similarly, E base may also form alternative base triple structures in recognizing a U-A pair (**Figure 2.1e,f**). Consistently, we observed

that an E residue can recognize both U-A and A-U pairs (**Figures 2.2q and 2.7, Table 2.1 and Figure 2.9**). The aniline moiety in S base, however, may not form alternative hydrogen bonds in the recognition of RNA base pairs, except the possible tautomeric amide bond formation in the recognition a C-G pair (**Figure 2.1h**). Our results also imply that the carbonyl methylene part of the succinyl linker adjacent to the aniline moiety may be considered as part of an expanded S base (**Figures 2.1g-i and 2.4**). The remaining carbonyl methylene of the succinyl linker can serve the role of a standard carbonyl methylene linker in PNAs (see **Figure 2.1a-d**). Taken together, our results show that the linker length of a PNA residue in a short dbPNA plays a critical role for the recognition of RNA base pair.

Figure 2.14. Non-denaturing PAGE for **(a-d)** P10, **(e-h)** P10-S₁, and **(i-m)** P11-S₁ binding to various RNA hairpins. The incubation buffer is 200 mM NaCl, 0.5 mM EDTA, 20 mM HEPES, pH 7.5. The loaded samples contain 1 μ M of rHP in 20 μ L. The PNA concentrations in lanes from left to right are 0, 0.2, 0.4, 1, 1.6, 2, 4, 10, 16, 20, 28, and 50 μ M, respectively.

Figure 2.15. Chemical structures of (a) T·A-U, (b) ^xT and (c) ^{xx}T. The letter R represents the sugar-phosphate backbone of RNA. Hydrogen bonding interactions are indicated by black dashed lines. (d-o) Non-denaturing PAGE for P13 (NH₂-Lys-TCTTTTTC-CONH₂), P13-xT (NH₂-Lys-TCT^xTTTTC-CONH₂), and P13-xxT (NH₂-Lys-TCT^{xx}TTTTC-CONH₂) binding to various RNA hairpins. The incubation buffer is 200 mM NaCl, 0.5 mM EDTA, 20 mM HEPES, pH 7.5. All loaded samples contain 1 μM of RNA hairpins in 20 μL , except for that shown in panel (f), where the loaded samples contain 0.25 μM of rHP3 in 25 μL . The PNA concentrations in lanes from left to right are 0, 0.2, 0.4, 1, 1.6, 2, 4, 10, 16, 20, 28, and 50 μM , respectively. The PNA concentrations in (f) from left to right are 0, 0.02, 0.05, 0.1, 0.15, 0.2, 0.4, 1, 2, 4, 10, and 20 μM , respectively.

Figure 2.16. K_d determination by non-denaturing PAGE for P11-E binding to (a) rHP3, (b) rHP4, and (c) P13 binding to rHP3 at pH 7.5. The fraction of triplex formation was calculated according to $Y = I_{\text{triplex}} / (I_{\text{triplex}} + aI_{\text{duplex}})$. The band intensities were normalized according to $a = I_{\text{triplex max}} / I_{\text{duplex max}}$. $I_{\text{duplex max}}$ is the band intensity for hairpin alone without the addition of PNA. $I_{\text{triplex max}}$ is the triplex band intensity with the highest concentration of PNA added. The data were fit to the equation: $Y = Y_0 + (B/2/R_0)(R_0 + X + K_d - ((R_0 + X + K_d)^2 - 4R_0X)^{1/2})$ where R_0 is the RNA hairpin concentration. Y_0 and B are the minimum and maximum fraction of triplex formation respectively. X is the total PNA concentration and K_d is the dissociation constant.

2.3.3 Effects of sequence context and cytosine-5 methylation on PNA·RNA₂ triplex formation

In PNA P11 (NH₂-Lys-TLTSTTTL-CONH₂), the S residue is at position 4 and flanked by two T residues. We made PNA P12 (NH₂-Lys-TLTLSTTL-CONH₂), with the S residue at position 5 and flanked by L and T residues, to study the effect of the sequence context of S modification on PNA·RNA₂ triplex formation. We studied the binding of P12 to RNA hairpins with A-U pair (rHP1), C-G pair (rHP5), G-C pair (rHP6), and U-A pair (rHP7) opposite to the S residue, respectively (**Figure 2.2a,f-h**). Our non-denaturing PAGE data reveal that P12 binds to rHP5 ($K_d = 11.9 \pm 2.2 \mu\text{M}$) and rHP7 ($K_d = 6.0 \pm 1.5 \mu\text{M}$) at 200 mM NaCl, pH 7.5 (**Figure 2.5h-j**, **Table 2.1**, **Figure 2.6d-e**), with weak binding to rHP1 ($K_d > 50 \mu\text{M}$) and no observable binding to (rHP6) (**Figure 2.5g**). Thus, PNAs containing S residues are useful for targeting dsRNAs containing canonical U-A, C-G, and U-G pairs with varied sequence contexts.

Next, we studied the effect of cytosine-5 methylation on PNA·RNA₂ triple formation, as the modelling suggests that attaching a methyl group may cause steric clash within an S·C-G triple (**Figure 2.5b,c**). Our non-denaturing PAGE data reveal that P3, P11 and P11-E do not show any observable binding to rHP2-m⁵C, which contains a m⁵C-G pair (**Figure 2.17**). It is not surprising that E base in P11-E does not recognize a m⁵C-G because E base does not recognize an unmodified C-G pair. The fact that a m⁵C-G pair is not recognized by either Q in P3 or S in P11 may further suggest that that Q and S recognize a C-G pair through relatively compact base triple formation (**Figure 2.5b,c**). The presence of a methyl group at 5-position of cytosine in a C-G pair may indeed generate a steric clash with the corresponding Q and S, in P3 and P11,

Figure 2.18. Testing the binding of PNAs targeting miR-26b hairpin precursor. (a,b) PNAs P14 and P14-E. (c) miR-26b hairpin precursor construct. The top part of the stem region is close to the Dicer cleavage site. The residues shown in the box were modified for the binding study. (d) A PNA·RNA₂ triplex formed between PNA P14 and miR-26b hairpin precursor. (e-f) Non-denaturing PAGE study of PNAs binding to miR-26b hairpin precursor. The gels were run with a running buffer of 1× TBE, pH 8.3 for 5 h at 250 V. The incubation buffer contains 200 mM NaCl, 0.5 mM EDTA, 20 mM HEPES, pH 7.5, or 200 mM NaCl, 0.5 mM EDTA, 20 mM MES, pH 6.0. The loaded miR-26b hairpin precursor is at 1 μM in 20 μL . The PNA concentration in lanes from left to right are (e) 0, 0.2, 0.4, 1, 1.6, 2, 4, 10, 16, 20, 28, and 50 μM , or (f-g) 0, 0.2, 0.4, 1, 1.6, 2, 4, 10, 16, 20, and 50 μM , respectively. (e) PNA P14 binds to miR-26b hairpin precursor with K_d values of approximately 50 and 10 μM , respectively, at pH 7.5 and 6.0. Due to the relatively weak binding and the weak band

intensity of the triplex band (probably due to the weakened staining upon PNA binding), we didn't quantify the K_d values. PNA P14-E shows no binding to miR-26b hairpin precursor at pH 6.0.

2.3.4 Targeting miR-26b precursor structure

We next incorporated S into a PNA (P14, NH₂-LysTLLTSTLL-CONH₂, **Figure 2.18a**) to target miRNA-26b hairpin precursor structure (**Figure 2.18**)⁸³⁻⁸⁴. P14 shows pH dependent binding to the miRNA-26b hairpin precursor (K_d values of around 10 and 50 μ M, respectively, at pH 6.0 and 7.5, **Figure 2.18**). The hairpin construct shows two distinct bands probably due to a relatively unstable stem region near the terminal ends resulting in the formation of an alternative secondary structure or the presence of a small fraction of a shortened sequence. A relatively unstable stem (containing multiple A-U/U-A and G-U/U-G pairs) may cause weakened binding by dsRNA-binding PNAs. Significantly, substitution of the single S residue in P14 with an E residue (P14-E, NH₂-LysTLLTETLL-CONH₂, **Figure 2.18b**) results in no binding to miRNA-26b hairpin precursor even at pH 6.0 (**Figure 2.18**). Taken together, the results show that, compared to the previously reported PNA E monomer which is able to recognize a Watson-Crick U-A base pair (**Figures 2.1e**), the PNA S monomer is advantageous for the recognition of a dsRNA region containing a Wobble U-G base pair probably through the formation of a unique S·U-G triple (**Figure 2.1i**). Targeting C-G, U-A, and U-G base pairs has been a significant challenge. Our data provide a starting point for further modifying S base to achieve improved binding affinity and specificity.

2.3.5 S-modified PNAs show no binding to ssRNAs

We next tested whether the S-modified PNAs bind to ssRNAs. The UV-absorbance-detected thermal melting results reveal that the short PNAs incorporating an S residue

(P10, P11 and P12) show no appreciable binding to the ssRNAs (**Figure 2.19, Figure 2.20**). Clearly, the S residue is useful for incorporating into dsRNA-binding PNAs for the sequence-specific and selective recognition dsRNAs over ssRNAs. Our previous studies have shown that PNAs incorporating L and Q residues show significantly weakened binding to ssRNAs^{23-24, 35-37, 58}. Thus, an S residue may be combined with L, Q, T, and other bases for designing dsRNA-binding PNAs for targeting biologically-important RNA structures.

Figure 2.19. Thermal melting results for potential PNA-RNA duplexes. The buffer is 200 mM NaCl, 0.5 mM EDTA, 20 mM NaH₂PO₄, pH 7.5. (a-d) ssRNAs studied. (e-h) PNAs studied. (i) PNA-RNA duplex formed between PNA P1 and ssRNA1. Melting curves for samples containing (j) ssRNA1, (k) ssRNA2, (l) ssRNA3, and (m) ssRNA4. For the curves with transitions, hysteresis between heating and cooling was observed (see **Figure 2.20**). Only the heating curves are shown. The melting temperatures are shown for the curves with melting transitions. S modification causes destabilization of PNA-RNA duplexes. The potential destabilizing base pairs formed

between PNA and RNA are indicated. We did not obtain the thermal melting curves for the triplexes due to the relatively low hyperchromicity of triplex to duplex transitions.

Figure 2.20. Thermal melting results for RNA-PNA duplexes. The incubation buffer is 200 mM NaCl, 0.5 mM EDTA, 20 mM NaH₂PO₄, pH 7.5. The melting temperatures are shown for the curves with melting transitions. For the curves with transitions, hysteresis between heating and cooling was observed, presumably due to the slow binding and/or dissociation rates compared to the temperature ramp rate (0.5°C/min).

2.4 Conclusion

In summary, we have developed a synthesis method for a PNA monomer incorporating the artificial S base. We have shown that the length of linker connecting the artificial S base and PNA backbone is critical for the recognition of internal C-G, U-A and U-G pairs in dsRNA regions at near-physiological conditions. Importantly, the relatively short 8-mer PNAs show no appreciable binding to dsDNAs and ssRNAs. Our modelling data suggest that S base can form the proposed hydrogen bonds and van der Waals contacts. The carbonyl methylene fragment of the succinyl linker

directly attached to the aniline ring of S may help expand S base by amide stacking interactions⁷²⁻⁷⁴ with an upstream base within the dbPNA (**Figure 2.4b**). Enhancing the stacking interactions within the triplex-forming strands⁸⁵ would certainly facilitate the development of further improved sequence-specific dsRNA-binding ligands. The recognition of a m⁵C-G pair by S base and previously developed Q base is severely weakened, which may suggest that S and Q recognize a C-G pair by relatively compact S·C-G and Q·C-G base triple formation, respectively, but not by intercalation. Our work provides new insights into the development of dbPNAs incorporating further improved nucleobases for targeting Watson-Crick C-G and U-A pairs as well as a wobble U-G pair, which is one of the most common non-Watson-Crick pairs in RNAs⁸⁶⁻⁸⁹.

References

1. Wilton, S. D.; Fall, A. M.; Harding, P. L.; McClorey, G.; Coleman, C.; Fletcher, S.. Antisense oligonucleotide-induced exon skipping across the human dystrophin gene. *Mol Ther* **2007**, *15* (7), 1288-96.
2. Crooke, S. T.; Witztum, J. L.; Bennett, C. F.; Baker, B. F., RNA-Targeted Therapeutics. *Cell Metab* **2018**, *27* (4), 714-739.
3. Shen, X.; Corey, D. R., Chemistry, mechanism and clinical status of antisense oligonucleotides and duplex RNAs. *Nucleic Acids Res* **2018**, *46* (4), 1584-1600.
4. Devi, G.; Zhou, Y.; Zhong, Z.; Toh, D. F. K.; Chen, G., RNA Triplexes: from structural principles to biological and biotech applications. *Wiley Interdiscip Rev RNA* **2015**, *6*, 111-28.
5. Zhou, Y.; Kierzek, E.; Loo, Z. P.; Antonio, M.; Yau, Y. H.; Chuah, Y. W.; Geifman-Shochat, S.; Kierzek, R.; Chen, G., Recognition of RNA duplexes by chemically modified triplex-forming oligonucleotides. *Nucleic Acids Res* **2013**, *41* (13), 6664-73.
6. Sugimoto, N.; Wu, P.; Hara, H.; Kawamoto, Y., pH and cation effects on the properties of parallel pyrimidine motif DNA triplexes. *Biochemistry* **2001**, *40* (31), 9396-405.
7. Bloese, J. M.; Silverman, S. K.; Bevilacqua, P. C., A simple molecular model for thermophilic adaptation of functional nucleic acids. *Biochemistry* **2007**, *46* (14), 4232-4240.
8. Rusling, D. A.; Powers, V. E.; Ranasinghe, R. T.; Wang, Y.; Osborne, S. D.; Brown, T.; Fox, K. R., Four base recognition by triplex-forming oligonucleotides at physiological pH. *Nucleic Acids Res* **2005**, *33* (9), 3025-32.
9. Ohkubo, A.; Yamada, K.; Ito, Y.; Yoshimura, K.; Miyauchi, K.; Kanamori, T.; Masaki, Y.; Seio, K.; Yuasa, H.; Sekine, M., Synthesis and triplex-forming properties of oligonucleotides capable of recognizing corresponding DNA duplexes containing four base pairs. *Nucleic Acids Res* **2015**, *43* (12), 5675-86.
10. Vuyisich, M.; Beal, P. A., Regulation of the RNA-dependent protein kinase by triple helix formation. *Nucleic Acids Res* **2000**, *28* (12), 2369-74.
11. Moses, A. C.; Huang, S. W.; Schepartz, A., Inhibition of Rev-RRE complexation by triplex tethered oligonucleotide probes. *Bioorg Med Chem* **1997**, *5* (6), 1123-9.
12. Fox, K. R.; Brown, T., Formation of stable DNA triplexes. *Biochemical Society transactions* **2011**, *39* (2), 629-34.
13. Griffin, L. C.; Dervan, P. B., Recognition of thymine adenine base pairs by guanine in a pyrimidine triple helix motif. *Science* **1989**, *245* (4921), 967-71.
14. Semenyuk, A.; Darian, E.; Liu, J.; Majumdar, A.; Cuenoud, B.; Miller, P. S.; Mackerell, A. D., Jr.; Seidman, M. M., Targeting of an interrupted polypurine:polypyrimidine sequence in mammalian cells by a triplex-forming oligonucleotide containing a novel base analogue. *Biochemistry* **2010**, *49* (36), 7867-78.
15. Abu Almakarem, A. S.; Petrov, A. I.; Stombaugh, J.; Zirbel, C. L.; Leontis, N. B., Comprehensive survey and geometric classification of base triples in RNA structures. *Nucleic Acids Res* **2012**, *40* (4), 1407-23.
16. Wang, S.; Kool, E. T., Relative stabilities of triple helices composed of combinations of DNA, RNA and 2'-O-methyl-RNA backbones: chimeric circular oligonucleotides as probes. *Nucleic Acids Res* **1995**, *23* (7), 1157-64.

17. Han, H.; Dervan, P. B., Sequence-specific recognition of double helical RNA and RNA.DNA by triple helix formation. *Proceedings of the National Academy of Sciences of the United States of America* **1993**, 90 (9), 3806-10.
18. Roberts, R. W.; Crothers, D. M., Stability and properties of double and triple helices: dramatic effects of RNA or DNA backbone composition. *Science* **1992**, 258 (5087), 1463-6.
19. Escude, C.; Francois, J. C.; Sun, J. S.; Ott, G.; Sprinzl, M.; Garestier, T.; Helene, C., Stability of triple helices containing RNA and DNA strands: experimental and molecular modeling studies. *Nucleic Acids Res* **1993**, 21 (24), 5547-53.
20. Nielsen, P. E.; Egholm, M.; Berg, R. H.; Buchardt, O., Sequence-selective recognition of DNA by strand displacement with a thymine-substituted polyamide. *Science* **1991**, 254 (5037), 1497-500.
21. Egholm, M.; Buchardt, O.; Christensen, L.; Behrens, C.; Freier, S. M.; Driver, D. A.; Berg, R. H.; Kim, S. K.; Norden, B.; Nielsen, P. E., PNA hybridizes to complementary oligonucleotides obeying the Watson-Crick hydrogen-bonding rules. *Nature* **1993**, 365 (6446), 566-8.
22. Li, M.; Zengeya, T.; Rozners, E., Short peptide nucleic acids bind strongly to homopurine tract of double helical RNA at pH 5.5. *J Am Chem Soc* **2010**, 132 (25), 8676-81.
23. Devi, G.; Yuan, Z.; Lu, Y.; Zhao, Y.; Chen, G., Incorporation of thio-pseudoisocytosine into triplex-forming peptide nucleic acids for enhanced recognition of RNA duplexes. *Nucleic Acids Res* **2014**, 42, 4008-18.
24. Toh, D.-F.; Devi, G.; Patil, K. M.; Qu, Q.; Maraswami, M.; Xiao, Y.; Loh, T.-P.; Zhao, Y.; Chen, G., Incorporating a guanidine-modified cytosine base into triplex-forming PNAs for the recognition of a C-G pyrimidine-purine inversion site of an RNA duplex *Nucleic Acids Res* **2016**, 44, 9071-82.
25. Zengeya, T.; Gupta, P.; Rozners, E., Sequence selective recognition of double-stranded RNA using triple helix-forming peptide nucleic acids. *Methods Mol Biol* **2014**, 1050, 83-94.
26. Hyrup, B.; Nielsen, P. E., Peptide nucleic acids (PNA): synthesis, properties and potential applications. *Bioorg Med Chem* **1996**, 4 (1), 5-23.
27. Demidov, V.; Frank-Kamenetskii, M. D.; Egholm, M.; Buchardt, O.; Nielsen, P. E., Sequence selective double strand DNA cleavage by peptide nucleic acid (PNA) targeting using nuclease S1. *Nucleic Acids Res* **1993**, 21 (9), 2103-7.
28. Winssinger, N.; Damoiseaux, R.; Tully, D. C.; Geierstanger, B. H.; Burdick, K.; Harris, J. L., PNA-encoded protease substrate microarrays. *Chem Biol* **2004**, 11 (10), 1351-60.
29. Komiyama, M.; Ye, S.; Liang, X.; Yamamoto, Y.; Tomita, T.; Zhou, J. M.; Aburatani, H., PNA for one-base differentiating protection of DNA from nuclease and its use for SNPs detection. *J Am Chem Soc* **2003**, 125 (13), 3758-62.
30. Gupta, P.; Muse, O.; Rozners, E., Recognition of double-stranded RNA by guanidine-modified peptide nucleic acids. *Biochemistry* **2012**, 51 (1), 63-73.
31. Zengeya, T.; Gupta, P.; Rozners, E., Triple-helical recognition of RNA using 2-aminopyridine-modified PNA at physiologically relevant conditions. *Angew Chem Int Ed Engl* **2012**, 51 (50), 12593-6.
32. Zengeya, T.; Li, M.; Rozners, E., PNA containing isocytidine nucleobase: synthesis and recognition of double helical RNA. *Bioorg Med Chem Lett* **2011**, 21 (7), 2121-4.

33. Gupta, P.; Zengeya, T.; Rozners, E., Triple helical recognition of pyrimidine inversions in polypurine tracts of RNA by nucleobase-modified PNA. *Chem Commun* **2011**, 47 (39), 11125-7.
34. Gowers, D. M.; Fox, K. R., Towards mixed sequence recognition by triple helix formation. *Nucleic Acids Res* **1999**, 27 (7), 1569-77.
35. Toh, D.-F. K.; Patil, K. M.; Chen, G., Sequence-specific and Selective Recognition of Double-stranded RNAs over Single-stranded RNAs by Chemically Modified Peptide Nucleic Acids. *J Vis Exp* **2017**, 127, e56221.
36. Puah, R. Y.; Jia, H.; Maraswami, M.; Kaixin Toh, D. F.; Ero, R.; Yang, L.; Patil, K. M.; Lerk Ong, A. A.; Krishna, M. S.; Sun, R.; Tong, C.; Huang, M.; Chen, X.; Loh, T. P.; Gao, Y. G.; Liu, D. X.; Chen, G., Selective Binding to mRNA Duplex Regions by Chemically Modified Peptide Nucleic Acids Stimulates Ribosomal Frameshifting. *Biochemistry* **2018**, 57, 149-159.
37. Patil, K. M.; Chen, G., Recognition of RNA sequence and structure by duplex and triplex formation: targeting miRNA and pre-miRNA. In *Modified Nucleic Acids in Biology and Medicine*, Erdmann, V. A.; Jurga, S.; Barciszewski, J., Eds. Springer: 2016; pp 299-317.
38. Eldrup, A. B.; Dahl, O.; Nielsen, P. E., A novel peptide nucleic acid monomer for recognition of thymine in triple-helix structures. *J Am Chem Soc* **1997**, 119 (45), 11116-11117.
39. Guianvarc'h, D.; Benhida, R.; Fourrey, J. L.; Maurisse, R.; Sun, J. S., Incorporation of a novel nucleobase allows stable oligonucleotide-directed triple helix formation at the target sequence containing a purine.pyrimidine interruption. *Chem Commun (Camb)* **2001**, (18), 1814-5.
40. Guianvarc'h, D.; Fourrey, J. L.; Maurisse, R.; Sun, J. S.; Benhida, R., Design of artificial nucleobases for the recognition of the AT inversion by triple-helix forming oligonucleotides: a structure-stability relationship study and neighbour bases effect. *Bioorg Med Chem* **2003**, 11 (13), 2751-9.
41. Wang, Y.; Rusling, D. A.; Powers, V. E.; Lack, O.; Osborne, S. D.; Fox, K. R.; Brown, T., Stable recognition of TA interruptions by triplex forming oligonucleotides containing a novel nucleoside. *Biochemistry* **2005**, 44 (15), 5884-92.
42. Patil, K. M.; Toh, D. F. K.; Yuan, Z.; Meng, Z. Y.; Shu, Z. Y.; Zhang, H. P.; Ong, A. A. L.; Krishna, M. S.; Lu, L. Y.; Lu, Y. P.; Chen, G., Incorporating uracil and 5-halouracils into short peptide nucleic acids for enhanced recognition of A-U pairs in dsRNAs. *Nucleic Acids Res* **2018**, 46 (15), 7506-7521.
43. Hahn, M., Receptor Surface Models .1. Definition and Construction. *J Med Chem* **1995**, 38 (12), 2080-2090.
44. Van der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C., GROMACS: Fast, flexible, and free. *J Comput Chem* **2005**, 26 (16), 1701-1718.
45. Trott, O.; Olson, A. J., AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem* **2010**, 31 (2), 455-61.
46. Hurd, C. D.; Kharasch, N., The Dipolar Ion Structures of the 2-Thiazolylsulfamic Acids. *Journal of the American Chemical Society* **1946**, 68 (4), 653-658.
47. Egholm, M.; Christensen, L.; Dueholm, K. L.; Buchardt, O.; Coull, J.; Nielsen, P. E., Efficient pH-independent sequence-specific DNA binding by pseudoisocytosine-containing bis-PNA. *Nucleic Acids Res* **1995**, 23 (2), 217-22.

48. Ono, A.; Tso, P. O. P.; Kan, L. S., Triplex formation of oligonucleotides containing 2'-O-methylpseudoisocytidine in substitution for 2'-deoxycytidine. *J Am Chem Soc* **1991**, *113* (10), 4032-4033.
49. Bevilacqua, P. C.; Brown, T. S.; Chadalavada, D.; Lecomte, J.; Moody, E.; Nakano, S. I., Linkage between proton binding and folding in RNA: implications for RNA catalysis. *Biochem Soc Trans* **2005**, *33* (Pt 3), 466-70.
50. Siegfried, N. A.; O'Hare, B.; Bevilacqua, P. C., Driving forces for nucleic acid pKa shifting in an A(+)·C wobble: effects of helix position, temperature, and ionic strength. *Biochemistry* **2010**, *49* (15), 3225-36.
51. Wilcox, J. L.; Bevilacqua, P. C., A Simple Fluorescence Method for pKa Determination in RNA and DNA Reveals Highly Shifted pKa's. *Journal of the American Chemical Society* **2013**, *135* (20), 7390-7393.
52. Wilcox, J. L.; Bevilacqua, P. C., pKa shifting in double-stranded RNA is highly dependent upon nearest neighbors and bulge positioning. *Biochemistry* **2013**, *52* (42), 7470-6.
53. Legault, P.; Pardi, A., Unusual dynamics and pKa shift at the active site of a lead-dependent ribozyme. *Journal of the American Chemical Society* **1997**, *119* (28), 6621-6628.
54. Acharya, P.; Cheruku, P.; Chatterjee, S.; Acharya, S.; Chattopadhyaya, J., Measurement of nucleobase pKa values in model mononucleotides shows RNA-RNA duplexes to be more stable than DNA-DNA duplexes. *J Am Chem Soc* **2004**, *126* (9), 2862-9.
55. Tang, C. L.; Alexov, E.; Pyle, A. M.; Honig, B., Calculation of pKas in RNA: on the structural origins and functional roles of protonated nucleotides. *J Mol Biol* **2007**, *366* (5), 1475-96.
56. Chen, G.; Kennedy, S. D.; Turner, D. H., A CA(+) pair adjacent to a sheared GA or AA pair stabilizes size-symmetric RNA internal loops. *Biochemistry* **2009**, *48* (24), 5738-52.
57. Wolter, A. C.; Weickhmann, A. K.; Nasiri, A. H.; Hantke, K.; Ohlenschlager, O.; Wunderlich, C. H.; Kreutz, C.; Duchardt-Ferner, E.; Wohnert, J., A Stably Protonated Adenine Nucleotide with a Highly Shifted pKa Value Stabilizes the Tertiary Structure of a GTP-Binding RNA Aptamer. *Angew Chem Int Ed Engl* **2017**, *56* (1), 401-404.
58. Patil, K. M.; Toh, D. K.; Yuan, Z.; Meng, Z.; Shu, Z.; Zhang, H.; Ong, A. A. L.; Krishna, M. S.; Lu, L.; Lu, Y.; Chen, G., Incorporating uracil and 5-halouracils into short peptide nucleic acids for enhanced recognition of A-U pairs in dsRNAs. *Nucleic Acids Res* **2018**, *46* (15), 7506-7521.
59. Allred, B. E.; Gebala, M.; Herschlag, D., Determination of Ion Atmosphere Effects on the Nucleic Acid Electrostatic Potential and Ligand Association Using AH⁺·C Wobble Formation in Double-Stranded DNA. *Journal of the American Chemical Society* **2017**, *139* (22), 7540-7548.
60. Yang, L.; Zhong, Z.; Tong, C.; Jia, H.; Liu, Y.; Chen, G., Single-Molecule Mechanical Folding and Unfolding of RNA Hairpins: Effects of Single A-U to A·C Pair Substitutions and Single Proton Binding and Implications for mRNA Structure-Induced -1 Ribosomal Frameshifting. *J Am Chem Soc* **2018**, *140* (26), 8172-8184.
61. Kan, L. S.; Lin, W. C.; Yadav, R. D.; Shih, J. H.; Chao, I., NMR studies of the tautomerism in pseudoisocytidine. *Nucleosides & Nucleotides* **1999**, *18* (4-5), 1091-1093.
62. Mir, A.; Chen, J.; Robinson, K.; Lendy, E.; Goodman, J.; Neau, D.; Golden, B. L., Two Divalent Metal Ions and Conformational Changes Play Roles in the Hammerhead Ribozyme Cleavage Reaction. *Biochemistry* **2015**, *54* (41), 6369-81.

63. Singh, V.; Fedeles, B. I.; Essigmann, J. M., Role of tautomerism in RNA biochemistry. *RNA* **2015**, *21* (1), 1-13.
64. Ai, H.; Chen, J.; Zhang, C., Amino-imino adenine tautomerism induced by the cooperative effect between metal ion and H₂O/NH₃. *J Phys Chem B* **2012**, *116* (46), 13624-36.
65. Rozov, A.; Westhof, E.; Yusupov, M.; Yusupova, G., The ribosome prohibits the G*U wobble geometry at the first position of the codon-anticodon helix. *Nucleic Acids Res* **2016**, *44* (13), 6434-41.
66. Kimsey, I. J.; Szymanski, E. S.; Zahurancik, W. J.; Shakya, A.; Xue, Y.; Chu, C. C.; Sathyamoorthy, B.; Suo, Z.; Al-Hashimi, H. M., Dynamic basis for dG*dT misincorporation via tautomerization and ionization. *Nature* **2018**, *554* (7691), 195-201.
67. Hansell, D. P.; Robinson, K. J.; Wallis, J. D.; Povey, D. C., N-[3-methyl-2(3H)-thiazolylidene]benzamide containing a short intramolecular S center dot center dot center dot O contact. *Acta Crystallogr C* **1996**, *52*, 136-139.
68. Peacock, H.; Luo, J. H.; Yamashita, T.; Luccarelli, J.; Thompson, S.; Hamilton, A. D., Non-covalent S--O interactions control conformation in a scaffold that disrupts islet amyloid polypeptide fibrillation. *Chem Sci* **2016**, *7* (10), 6435-6439.
69. Creuven, I.; Norberg, B.; Olivier, A.; Evrard, C.; Evrard, G.; Wigerinck, P.; Herdewijn, P.; Durant, F., 5-(thien-2-yl) uracil analogs: 5-(5-methylthien-2-yl)-2'-deoxyuridine, 5-(5-thien-2-yl)-2'-deoxyuridine, and 5-(5-bromothien-2-yl)-2'-deoxyuridine. *Journal of Chemical Crystallography* **1996**, *26* (11), 777-789.
70. Collins, I.; Moyes, C.; Davey, W. B.; Rowley, M.; Bromidge, F. A.; Quirk, K.; Atack, J. R.; McKernan, R. M.; Thompson, S. A.; Wafford, K.; Dawson, G. R.; Pike, A.; Sohal, B.; Tsou, N. N.; Ball, R. G.; Castro, J. L., 3-heteroaryl-2-pyridones: Benzodiazepine site ligands with functional selectivity for alpha 2/alpha 3-subtypes of human GABA(A) receptor-ion channels. *Journal of Medicinal Chemistry* **2002**, *45* (9), 1887-1900.
71. Beno, B. R.; Yeung, K. S.; Bartberger, M. D.; Pennington, L. D.; Meanwell, N. A., A Survey of the Role of Noncovalent Sulfur Interactions in Drug Design. *Journal of Medicinal Chemistry* **2015**, *58* (11), 4383-4438.
72. Wang, Y.; Wang, J.; Yao, L., Computational study of peptide plane stacking with polar and ionizable amino acid side chains. *J Phys Chem A* **2015**, *119* (14), 3471-8.
73. James, W. H., 3rd; Muller, C. W.; Buchanan, E. G.; Nix, M. G.; Guo, L.; Roskop, L.; Gordon, M. S.; Slipchenko, L. V.; Gellman, S. H.; Zwier, T. S., Intramolecular amide stacking and its competition with hydrogen bonding in a small foldamer. *J Am Chem Soc* **2009**, *131* (40), 14243-5.
74. Priyakumar, U. D.; Hyeon, C.; Thirumalai, D.; MacKerell, A. D., Urea Destabilizes RNA by Forming Stacking Interactions and Multiple Hydrogen Bonds with Nucleic Acid Bases. *Journal of the American Chemical Society* **2009**, *131* (49), 17759-17761.
75. Sato, T.; Sato, Y.; Nishizawa, S., Optimization of the Alkyl Linker of TO Base Surrogate in Triplex-Forming PNA for Enhanced Binding to Double-Stranded RNA. *Chemistry* **2016**, *23*, 4079-4088.
76. Sato, T.; Sato, Y.; Nishizawa, S., Triplex-Forming Peptide Nucleic Acid Probe Having Thiazole Orange as a Base Surrogate for Fluorescence Sensing of Double-stranded RNA. *J Am Chem Soc* **2016**, *138* (30), 9397-400.
77. Filichev, V. V.; Pedersen, E. B., Stable and selective formation of Hoogsteen-type triplexes and duplexes using twisted intercalating nucleic acids (TINA) prepared via

postsynthetic sonogashira solid-phase coupling reactions. *Journal of the American Chemical Society* **2005**, *127* (42), 14849-14858.

78. Doluca, O.; Boutorine, A. S.; Filichev, V. V., Triplex-Forming Twisted Intercalating Nucleic Acids (TINAs): Design Rules, Stabilization of Antiparallel DNA Triplexes and Inhibition of G-Quartet-Dependent Self-Association. *Chembiochem* **2011**, *12* (15), 2365-2374.

79. Koshlap, K. M.; Gillespie, P.; Dervan, P. B.; Feigon, J., Nonnatural Deoxyribonucleoside-D(3) Incorporated in an Intramolecular DNA Triplex Binds Sequence-Specifically by Intercalation. *Journal of the American Chemical Society* **1993**, *115* (17), 7908-7909.

80. Hochstrasser, R. A.; Carver, T. E.; Sowers, L. C.; Millar, D. P., Melting of a DNA helix terminus within the active site of a DNA polymerase. *Biochemistry* **1994**, *33* (39), 11971-11979.

81. Sakakibara, Y.; Abeysirigunawardena, S. C.; Duc, A. C.; Dremann, D. N.; Chow, C. S., Ligand- and pH-induced conformational changes of RNA domain helix 69 revealed by 2-aminopurine fluorescence. *Angew Chem Int Ed Engl* **2012**, *51* (48), 12095-8.

82. Hyrup, B.; Egholm, M.; Nielsen, P. E.; Wittung, P.; Norden, B.; Buchardt, O., Structure-Activity Studies of the Binding of Modified Peptide Nucleic-Acids (PNAs) to DNA. *Journal of the American Chemical Society* **1994**, *116* (18), 7964-7970.

83. Kozomara, A.; Griffiths-Jones, S., miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res* **2014**, *42* (Database issue), D68-73.

84. Absalon, S.; Kochanek, D. M.; Raghavan, V.; Krichevsky, A. M., MiR-26b, upregulated in Alzheimer's disease, activates cell cycle entry, tau-phosphorylation, and apoptosis in postmitotic neurons. *J Neurosci* **2013**, *33* (37), 14645-59.

85. Brown, J. A.; Bulkley, D.; Wang, J.; Valenstein, M. L.; Yario, T. A.; Steitz, T. A.; Steitz, J. A., Structural insights into the stabilization of MALAT1 noncoding RNA by a bipartite triple helix. *Nat Struct Mol Biol* **2014**, *21* (7), 633-40.

86. Varani, G.; McClain, W. H., The G x U wobble base pair. A fundamental building block of RNA structure crucial to RNA function in diverse biological systems. *EMBO Rep* **2000**, *1* (1), 18-23.

87. Xu, D.; Landon, T.; Greenbaum, N. L.; Fenley, M. O., The electrostatic characteristics of G.U wobble base pairs. *Nucleic Acids Res* **2007**, *35* (11), 3836-47.

88. Leontis, N. B.; Westhof, E., Geometric nomenclature and classification of RNA base pairs. *RNA* **2001**, *7* (4), 499-512.

89. Chen, J. L.; Dishler, A. L.; Kennedy, S. D.; Yildirim, I.; Liu, B.; Turner, D. H.; Serra, M. J., Testing the Nearest Neighbor Model for Canonical RNA Base Pairs: Revision of GU Parameters. *Biochemistry* **2012**, *51* (16), 3508-3522.

Chapter 3

Incorporating 2-thiouracil into short dsRNA-binding PNAs (dbPNAs) for enhanced recognition of A-U pairs and for targeting a microRNA hairpin precursor

3.1 Introduction

Triplex-forming oligonucleotides (TFOs) are capable of recognizing RNA duplexes through the formation of parallel major-groove TFO·RNA₂ triplexes stabilized by consecutive base triples such as T/U·A·U and C⁺·G·C (**Figure 3.1a,b**)¹⁻⁶. TFOs show better binding affinity to DNA duplexes than RNA duplexes^{1-2, 7-11}, resulting in a significant challenge in the development of TFO-based ligands for selectively targeting RNA duplexes. Peptide Nucleic Acids (PNAs) (**Figure 3.1**) are a synthetic mimic of natural nucleic acids (DNA and RNA), consisting of neutral, peptide-like backbone with the nucleobases typically attached by a methylene carbonyl linker¹²⁻¹³. The chemically stable backbone allows PNA to be resistant against nucleases and proteases¹⁴⁻¹⁷. PNAs form stable Watson-Crick duplexes and parallel major-groove triplexes with DNA and RNA^{2, 13-14, 18-24}. Interestingly, short PNAs are capable of selectively target double-stranded RNA (dsRNA) but not dsDNAs, showing a great promise in developing dsRNA-binding PNAs (dbPNAs) as novel RNA-targeting ligands^{2, 18-37}.

We have previously discovered that incorporation into short dbPNAs with thio-pseudoisucytosine (L, **Figure 3.1c**), a derivative of pseudoisocytosine (J)³⁸⁻³⁹, facilitates recognition of Watson-Crick G-C pair in dsRNAs at physiological

condition with minimized pH dependence²². Furthermore, L-modified PNAs show selective recognition of dsRNAs with significantly reduced binding to single-stranded RNAs (ssRNAs) because of the steric hindrance arisen from the formation of a unstable Watson-Crick-like L-G pair (**Figure 3. 1d**)²².

PNAs incorporating T residue have been utilized to target A-U base pairs via the formation of T·A-U base triples (**Figure 3.1a**). We have showed that halouracils-modified PNAs may improve the recognition of dsRNAs containing an A-U pair at physiological condition with minimized pH dependence²⁴. However, halouracils show sequence dependent (de)stabilization upon incorporating into L-containing PNAs for the PNA·RNA₂ triplex formation. Incorporating a relatively bulky benzothiophene-modified uracil also results in a destabilization of the formation of PNA·RNA₂ triplex²⁹. Thus, a modified A-U pair-recognizing base is needed for the enhanced formation of PNA·RNA₂ triplex in a generally applicable manner.

Incorporation of 2-thiouracil (s²U) into oligonucleotides stabilizes the formation of duplexes and triplexes through the formation of Watson-Crick s²U-A base pair and s²U·A-U base triple, respectively (**Figure 3.1e,f**)^{11, 40-47}. An s²U modification in oligonucleotides favors C3'-endo sugar pucker induced by a steric effect between thiocarbonyl group and hydroxyl group⁴⁸. In addition, the pK_a of N3 nitrogen atom is decreased upon the thiolation of Uracil, which may strengthen the hydrogen bonding involving imino proton H3 in s²U^{43, 49}. The enhanced stacking interaction of 2-thiocarbonyl group with adjacent bases may also contribute to the formation of stable TFO·RNA₂ triplexes^{40, 45}.

Incorporation of s^2U into PNAs favors double-duplex invasion of dsDNAs⁵⁰. However, the effect of s^2U incorporation into chemically-modified dbPNAs on the recognition of RNA duplexes is not understood. Here, we studied the binding of s^2U - and L-modified short PNAs (8-mer) to dsRNAs, dsDNA, and ssRNAs by PAGE and thermal melting experiments. Furthermore, we tested the application of an s^2U - and L-modified PNA in targeting a dsRNA region of pre-microRNA-198 for the inhibition of the binding and cleavage activity of Dicer.

Figure 3.1. Chemical structures of PNA·RNA₂ base triples, PNA-RNA base pairs, and PNA L- s^2U -5. The sugar-phosphate backbone of RNA is shown as "R". The black dashed lines indicate Hydrogen bonding. (a-c,e) PNA·RNA₂ base triples base

synthesized manually using Boc chemistry via a Solid-Phase Peptide Synthesis (SPPS) protocol. 4-Methylbenzhydrylamine hydrochloride (MBHA·HCl) polystyrene resins were used. The loading value used for the synthesis of the oligomers was 0.3 mmol/g and acetic anhydride was used as the capping reagent. Benzotriazol-1-yl-oxytripyrrolidinophosphonium hexafluorophosphate (PyBOP) and *N,N*-diisopropylethylamine (DIPEA) were used as the coupling reagent. The coupling of each PNA monomer was monitored by Kaiser test. The cleavage of the PNA oligomers was done using trifluoroacetic acid (TFA) and trifluoromethanesulfonic acid (TFMSA) method, after which the oligomers were precipitated with diethyl ether, dissolved in deionized water and purified by RP-HPLC using H₂O–CH₃CN–0.1% TFA as the mobile phase. Matrix-assisted laser desorption/ionization-time of flight (MALDI-TOF) analysis was used to characterize the oligomers (**Table A2.2, Figure A2.2**), with the use of α -cyano-4-hydroxycinnamic acid (CHCA) as the sample crystallization matrix.

3.2.2 Thermal melting

UV-absorbance-detected thermal melting experiments were conducted using a Shimadzu UV-2550 UV-Vis spectrophotometer with the use of an 8-microcell cuvette. The absorbance at 260 nm was recorded with the temperature increasing from 15 to 95°C followed by the temperature decreasing from 95 to 15°C. The temperature ramp rate is 0.5°C/min. The optical path length is 1 cm. All samples contain 5 μ M RNA and 5 μ M PNA in 130 μ L buffer of 200 mM NaCl, 0.5 mM EDTA, 20 mM NaH₂PO₄, pH 7.5. The samples containing the single-stranded RNA and PNA were annealed by slow cooling from 95 °C to room temperature, followed by incubation at 4°C overnight. Data were normalized at high temperature and the melting temperatures were determined based on the Gaussian fit of the first derivative of the curves.

3.2.3 Nondenaturing polyacrylamide gel electrophoresis

Nondenaturing polyacrylamide gel electrophoresis (PAGE) (12wt%) experiments were conducted with incubation buffers containing 200 mM NaCl, 0.5 mM EDTA, 20 mM HEPES at pH 7.5 or pH 8.0. The loading volume for samples containing RNA and DNA hairpins were 20 μ L or 25 μ L and 10 μ L, respectively. The concentration of RNA and DNA are 50 nM, 0.1 μ M, or 0.25 μ M. The samples were prepared by snap cooling of the hairpins, followed by annealing with PNA oligomers by slow cooling from 65°C to room temperature and incubation at 4°C overnight. Prior to loading the samples into the wells, 35% glycerol (20% of the total volume) was added to the sample mixtures. 1 $\times$ Tris–borate–EDTA (TBE) buffer, pH 8.3 was used as the running buffer for all gel experiments. The gel was run at 4 °C at 250 V for 5 h. The gels were then stained with ethidium bromide and imaged by a Typhoon Trio Variable Mode Imager.

3.2.4 Data analysis

The fraction of triplex formation was calculated based on equation $Y = I_{\text{triplex}} / (I_{\text{triplex}} + aI_{\text{duplex}})$. The band intensities were normalized according to $a = I_{\text{triplex max}} / I_{\text{duplex max}}$. $I_{\text{duplex max}}$ is the band intensity for hairpin alone without the addition of PNA. $I_{\text{triplex max}}$ is the triplex band intensity with the highest concentration of PNA added. The data were fit to the equation: $Y = Y_0 + (B/2/R_0)(R_0 + X + K_d - ((R_0 + X + K_d)^2 - 4R_0X)^{1/2})$ where R_0 is the RNA hairpin concentration (50 nM). Y_0 and B are the minimum and maximum fraction of triplex formation respectively. X is the total PNA concentration and K_d is the dissociation constant.

3.2.5 In vitro dicing assay

In vitro dicing assay was performed using purified Dicer enzyme (Genlantis) and RNA oligonucleotides (IDT) containing the sequences of human *pre-mir-124-2*

(5'CGUGUUCACAGCGGACCUUGAUUUAAUGUCAUACAAUUAAGGCACGC
GGUGAAUGCCAA3') and *pre-mir-198*
(5'GGUCCAGAGGGGAGAUAGGUUCCUGUGAUUUUCCUUCUUCUCUAUA
GAAUAA3'). 5-10 nM 5'-³²P-labeled pre-miRNA was incubated in Dicing buffer (14
mM Tris-HCl pH 7.6, 2 mM MgCl₂, 1 mM DTT) containing 0.017 unit/μL Dicer and
0.017 Unit/μL RNaseOut (Invitrogen). The PNA (with final concentrations of 0.2-0.6
μM) was added to the reaction. The reactions were incubated at 37 °C for two hours.
Dicing reaction was stopped by adding the same volume of Proteinase K solution
(200 μg/mL Proteinase K, 8 mM NaCl, 10 mM Tris (pH 8.5), 0.05% SDS, 5 mM
EDTA) and incubating at 37 °C for 30 minutes. RNA was precipitated by ethanol
precipitation and the pellet was dissolved in 2× RNA gel loading dye (Invitrogen).
RNA was resolved on a denaturing polyacrylamide gel (12% Sequagel, National
Diagnostics) and transferred onto Hybond N+ (GE Healthcare). Radioactivity was
detected by the Typhoon Phosphorimager (GE Healthcare) and signals quantified by
Image Quant (GE Healthcare).

3.3 Results and Discussion

3.3.1 Replacement of T with s²U in an modified dbPNA improves the recognition of dsRNA in a position-independent manner

We conducted nondenaturing PAGE to study the PNA·RNA₂ triplex formation. Our
previous nondenaturing PAGE data show that an unmodified PNA (P1, NH₂-Lys-
TCTCTTTC-CONH₂, **Figure 3.2**) binds to rHP1 without strand invasion ($K_d = 2.1 \pm$
0.4 μM at pH 7.5)²⁴. We have previously shown that a single T-to-U substitution at
position 3 in P1 (C-U-3) moderately improved binding toward rHP1²⁴. The T-to-U
substitution stabilized the formation of PNA·RNA₂ triplex by a strengthened
hydrogen bond caused by the decrease of pK_a value of monomer N3 from ~9.8 for T

toward neutrality for U^{24, 51-52}. In this study, we used a Cyanine 3 (Cy3)-labelled rHP1 (50 nM instead of 0.25 μ M) and obtained comparable K_d values (P1, $K_d = 3.07 \pm 0.67$ μ M; C-U-3, $K_d = 1.22 \pm 0.17$ μ M at pH 7.5, **Figure 3.3a, 3.4b, Table 3.2**).

Compared to T-to-U substitutions, T-to-s²U substitutions in PNAs may result in more significant strengthening of triplex formation by strengthened base stacking and hydrogen bonding interaction^{11, 40-47}. Stacking interactions are often sequence context dependent, which may result in sequence dependent stabilization of T-to-s²U substitutions. We studied the effects of single T-to-s²U substitution at varied positions (positions 1, 3, 5, 6, and 7, **Figure 3.2f, Table 3.1**) of PNA P1 on the formation of triplex. The T residue of PNA P1 at position 1 is adjacent to a C residue (s²UC, C-s²U-1), while the T residues of PNA P1 at positions 3, 5, 6, and 7 are surrounded by two C residues (Cs²UC, C-s²U-3), C and T residues (Cs²UT, C-s²U-5), two T residues (Ts²UT, C-s²U-6), and T and C residues (Ts²UC, C-s²U-7), respectively. We observed that, compared to the control PNAs (P1 and C-U-3), all s²U-modified PNAs show enhanced binding to rHP1-Cy3 (**Figure 3.3a,c, 3.4, 3.5, Table 3.2**). Among all the s²U-modified PNAs at varied positions, the PNA containing a s²U modification at position 6 (C-s²U-6, with a local sequence context of Ts²UT, $K_d = 0.12 \pm 0.01$ μ M) shows the best improvement in binding to rHP1-Cy3 at pH 7.5, followed by s²U modification at position 5 (C-s²U-5, Cs²UT, $K_d = 0.21 \pm 0.07$ μ M), position 7 (C-s²U-7, Ts²UC, $K_d = 0.32 \pm 0.04$ μ M), position 1 (C-s²U-1, s²UC, $K_d = 0.44 \pm 0.06$ μ M), and position 3 (C-s²U-3, Ts²UC, $K_d = 0.72 \pm 0.13$ μ M) (**Figure 3.2f,i, 3.3, 3.4, Table 3.2**).

We further tested the effect of pH on triplex formation. Upon increasing the pH from 7.5 to 8.0, the binding of P1 to rHP1-Cy3 is weakened (by about 4 fold, with K_d values of 3.07 and 13.17 μM , respectively) (**Table 3.2, 3.7 and 3.8**), mainly due to the relatively low pH requirement for the formation of $\text{C}^+\cdot\text{G}\cdot\text{C}$ base triple (**Figure 3.1b**). Replacing T with s^2U at positions 3, 5, and 6 in P1 shows a similar pH dependence, with the K_d value increased by about 4 fold upon increasing the pH from 7.5 to 8.0. However, the two PNAs with s^2U substitutions at positions 1 and 7, respectively, show a relatively large triplex destabilization (about a 20-fold increase in K_d values) upon pH increase from 7.5 to 8.0. It is probable that a C residue at and near the terminal ends of a PNA is relatively more sensitive to an adjacent T-to- s^2U substitution, due to the fact that, compared to internal C residues, C residues at terminal ends have a relatively lower $\text{p}K_a$ to form a $\text{C}^+\cdot\text{G}\cdot\text{C}$ triple⁵³. More studies are needed to understand if the stacking of an s^2UC step (present in PNAs C- s^2U -1 and C- s^2U -7) may affect the $\text{p}K_a$ of C, resulting in an increased pH dependence.

We have previously demonstrated that replacement of C by L (**Figure 3.1c**) stabilizes the $\text{PNA}\cdot\text{RNA}_2$ triplex formation²². Thus, it is important to study the effect of simultaneous incorporation of s^2U and L modifications on $\text{PNA}\cdot\text{RNA}_2$ triplex formation. We studied the effects of single s^2U substitution at different positions (positions 1, 3, 5, 6, and 7, **Figure 3.2g, Table 3.1**) of the L-modified PNA P5 (NH_2 -Lys-TLTLTTTL- CONH_2) on $\text{PNA}\cdot\text{RNA}_2$ triplex formation (**Figure 3.2j**). PNA P5 binds to rHP1-Cy3 with a K_d value of $0.18 \pm 0.07 \mu\text{M}$ (**Figure 3.3b, Table 3.2**). All the T-to- s^2U substitutions in PNA P5 result in enhanced binding, with the K_d values lies between a small range of 0.06-0.16 μM (**Table 3.2, Figures 3.3b,d, 3.4, 3.5 and**

3.12a). Note that PNA L-s²U-3 may bind to rHP1-Cy3 with a 2:1 binding ratio as revealed by the presence of two complex bands in PAGE (**Figures 3.4f and 3.6**).

Upon increasing the pH from 7.5 to 8.0, the binding of L- and s²U-containing PNAs is weakened slightly (with the K_d values increased by about 2 fold) (**Figures 3.7, 3.8, Table 3.2**), presumably due to the relatively pH-independent formation of L·G-C base triples. It is of note that s²U and L are compatible in stabilizing the formation of PNA·RNA₂ triplex, without complex sequence and pH dependence as compared to that for the PNAs without L modifications. We have previously observed 5-halouracils may show sequence dependent (de)stabilization upon incorporating into L-containing PNAs for the PNA·RNA-RNA triplex formation²⁴. Thus, compared to halouracils, s²U is advantageous and one may incorporate s²U into PNAs at varied positions for enhanced binding to dsRNAs.

Figure 3.3. Nondenaturing PAGE study of different PNAs binding to rHP1-Cy3. rHP1-Cy3 is loaded at 50 nM in 20 μ L. The PNA concentrations in lanes from left to right are (a) 0, 0.05, 0.1, 0.2, 0.4, 1, 1.6, 2, 4, 10, 16, 20, 28, and 50 μ M and (b-d) 0, 0.005, 0.01, 0.02, 0.05, 0.1, 0.2, 0.4, 0.7, 1, 1.5, 2, 5, and 10 μ M, respectively.

(a) P13 + rHP3-Cy3 ($K_d = 0.38 \pm 0.05 \mu\text{M}$)

(b) C-U-3 + rHP1-Cy3 ($K_d = 1.22 \pm 0.17 \mu\text{M}$)

(c) C-s²U-1 + rHP1-Cy3 ($K_d = 0.44 \pm 0.06 \mu\text{M}$)

(d) L-s²U-1 + rHP1-Cy3 ($K_d = 0.10 \pm 0.05 \mu\text{M}$)

(e) C-s²U-3 + rHP1-Cy3 ($K_d = 0.72 \pm 0.13 \mu\text{M}$)

(f) L-s²U-3 + rHP1-Cy3

(g) C-s²U-4 + rHP3-Cy3 ($K_d = 0.04 \pm 0.01 \mu\text{M}$)

(h) L-s²U-4 + rHP3-Cy3 ($K_d = 0.08 \pm 0.03 \mu\text{M}$)

(i) C-s²U-5 + rHP1-Cy3 ($K_d = 0.21 \pm 0.07 \mu\text{M}$)

(j) L-s²U-5 + rHP1-Cy3 ($K_d = 0.11 \pm 0.05 \mu\text{M}$)

(k) C-s²U-6 + rHP1-Cy3 ($K_d = 0.12 \pm 0.01 \mu\text{M}$)

(l) L-s²U-6 + rHP1-Cy3 ($K_d = 0.06 \pm 0.02 \mu\text{M}$)

Figure 3.4. Nondenaturing PAGE for dPNAs binding to rHP1-Cy3 and rHP3-Cy3 at pH 7.5. The loaded samples contain 50 nM (panel a-n) and 0.25 μM (panel o-q) of RNA hairpin in 20 μL and 25 μL , respectively. The PNA concentrations in lanes from left to right are: (panel a-n) 0, 0.005, 0.01, 0.02, 0.05, 0.1, 0.2, 0.4, 0.7, 1, 1.5, 2, 5, and 10 μM respectively, or (panel o-q) 0, 0.05, 0.1, 0.2, 0.4, 1, 1.6, 2, 4, 10, 16, 20, 28, and 50 μM , respectively.

Figure 3.5. K_d determination by non-denaturing PAGE for various PNAs binding to rHP1-Cy3 and rHP3-Cy3 at pH 7.5.

Figure 3.6. Proposed 2:1 complex structure with the formation of two coaxially-stacked parallel triplexes with the PNA T residues at position 1 and/or 3 replaced with s^2 U (see **Figures 3.4f** and **3.7d,f**).

Figure 3.7. Nondenaturing PAGE for various PNAs binding to rHP1-Cy3 and rHP3-Cy3 at pH 8.0. The loaded samples contain 50 nM of RNA hairpin in 20 μ L. The PNA concentrations in lanes from left to right are: (**left column**) 0, 0.05, 0.1, 0.2, 0.4,

1, 1.6, 2, 4, 10, 16, 20, 28, and 50 μM respectively, or (**right column**) 0, 0.005, 0.01, 0.02, 0.05, 0.1, 0.2, 0.4, 0.7, 1, 1.5, 2, 5, and 10 μM , respectively.

Figure 3.8. K_d determination by nondenaturing PAGE for various PNAs binding to rHP1-Cy3 at pH 8.0.

Table 3.1. PNAs studied in this paper

PNA	Sequence (N-terminus to C-terminus)
P1	NH ₂ -Lys-TCTCTTTC-CONH ₂
P5	NH ₂ -Lys-TLTLTTTL-CONH ₂
P13	NH ₂ -Lys-TCT T TTTC-CONH ₂
C-U-3	NH ₂ -Lys-TC U CTTTC-CONH ₂
C-s ² U-1	NH ₂ -Lys- s ² U CTCTTTC-CONH ₂
C-s ² U-3	NH ₂ -Lys-TC s ² U CTTTC-CONH ₂
C-s ² U-4	NH ₂ -Lys-TCT s ² U TTTC-CONH ₂
C-s ² U-5	NH ₂ -Lys-TCTC s ² U TTTC-CONH ₂
C-s ² U-6	NH ₂ -Lys-TCTCT s ² U TC-CONH ₂
C-s ² U-7	NH ₂ -Lys-TCTCTT s ² U C-CONH ₂
L-s ² U-1	NH ₂ -Lys- s ² U LTTLTTTL-CONH ₂
L-s ² U-3	NH ₂ -Lys-TL s ² U LTTL-CONH ₂
L-s ² U-4	NH ₂ -Lys-TLT s ² U TTTL-CONH ₂
L-s ² U-5	NH ₂ -Lys-TLTL s ² U TTL-CONH ₂
L-s ² U-6	NH ₂ -Lys-TLTLT s ² U TL-CONH ₂
L-s ² U-7	NH ₂ -Lys-TLTLTT s ² U L-CONH ₂
dbPNA-198	NH ₂ -Lys-LLTLTTLL-CONH ₂
dbPNA-198-s ² U	NH ₂ -Lys-LLTL s ² U TLL-CONH ₂
dbPNA-NC ^a	NH ₂ -Lys-LLQTLEL-CONH ₂

^aThe PNA E residue has been used for the recognition of DNA T-A and RNA U-A pairs^{25, 33, 35-36, 54}. The guanidine-modified 5-methylcytosine (Q)⁵⁵ has been developed to selectively recognize dsRNAs containing an Watson-Crick C-G base pair²³.

Table 3.2. K_d (μM) values for PNA binding to RNA or DNA duplexes obtained by non-denaturing PAGE.^a

	rHP1 (G-C)	rHP1-Cy3 (G-C)	rHP2 (C-G)	rHP3 (A-U)	rHP3-Cy3 (A-U)	rHP4 (U-A)	dHP1 (A-U)
P1	2.1 ± 0.4^b (27.0 ± 5.3) ^b	3.07 ± 0.67 (13.17 ± 6.72)	NB ^b	NB ^b	-	NB ^b	NB ^b
P5	0.2 ± 0.1^b (0.3 ± 0.10) ^b	0.18 ± 0.07 (0.36 ± 0.13)	NB ^b	0.4 ± 0.1^b	-	1.9 ± 0.4^b	-
P13	NB ^b	-	NB ^b	0.2 ± 0.1^b	0.38 ± 0.05	NB ^b	-
C-U-3	1.2 ± 0.3^b	1.22 ± 0.17 (14.74 ± 3.59)	-	-	-	-	NB ^b
C-s ² U-1	-	0.44 ± 0.06 (8.77 ± 1.33)	-	-	-	-	-
L-s ² U-1	-	0.10 ± 0.05 (0.22 ± 0.06)	-	-	-	-	-
C-s ² U-3	-	0.72 ± 0.13 (3.98 ± 0.62)	-	-	-	-	-
L-s ² U-3	-	2:1 binding ^c	-	-	-	-	-
C-s ² U-4	NB	-	NB	< 0.25	0.04 ± 0.01	NB	-
L-s ² U-4	-		-	-	0.08 ± 0.03	-	-
C-s ² U-5	-	0.21 ± 0.07 (0.82 ± 0.22)	-	-	-	-	-
L-s ² U-5	-	0.11 ± 0.05 (0.20 ± 0.08)	-	-	-	-	-
C-s ² U-6	-	0.12 ± 0.01 (0.57 ± 0.12)	-	-	-	-	NB
L-s ² U-6	-	0.06 ± 0.02 (0.16 ± 0.06)	-	-	-	-	-
C-s ² U-7	-	0.32 ± 0.04 (6.30 ± 1.35)	-	-	-	-	-
L-s ² U-7	-	0.16 ± 0.07 (0.25 ± 0.10)	-	-	-	-	-

^aTwo incubation buffers were used 200 mM NaCl, 0.5 mM EDTA, 20 mM HEPES, pH 7.5 or pH 8.0. The data shown in parentheses were taken at pH 8.0. “-” represents that the data were not obtained. “NB” indicates that no binding was observed. ^bData obtained from a previous study²⁴. The concentration of Cy3-labelled RNA hairpins is 50 nM, while the concentration of unlabelled RNA hairpins is 0.25 μM . ^cPAGE data reveal a 2:1 binding (**Figures 3.4f and 3.7f**).

3.3.2 s^2 U-modified PNAs show sequence specific recognition of dsRNA but no appreciable recognition of dsDNA

We studied the binding specificity of the s^2 U-modified PNA (C- s^2 U-4). We have previously reported that PNA P13 (NH₂-Lys-TCTTTTTC-CONH₂) binds to rHP3 with a K_d value of $0.2 \pm 0.1 \mu\text{M}$ ²⁵. Here, we used a Cy3-labelled rHP3 (rHP3-Cy3, 50 nM instead of 0.25 μM) and obtained a comparable K_d value ($0.38 \pm 0.05 \mu\text{M}$ at pH 7.5, **Figure 3.4a, Table 3.2**). We replaced the T residue at 4th position in PNA P13 with an s^2 U residue (C- s^2 U-4) and our nondenaturing PAGE result suggests that PNA C- s^2 U-4 binds to rHP3-Cy3 with a K_d value of $0.04 \pm 0.01 \mu\text{M}$ (**Figures 3.4g, Table 3.2**), which again suggests that incorporation of an s^2 U residue enhances the binding toward dsRNA containing an A-U pair. No triplex formation was observed between C- s^2 U-4 and mutated RNA hairpins rHP2, rHP1 and rHP4 (**Figure 3.2, 3.4o-q**). The results indicate that s^2 U-modified PNAs possesses an excellent binding specificity via s^2 U·A-U triple formation.

We next studied the binding of an s^2 U-incorporated PNA to a DNA hairpin (dHP1). We observed that PNA C- s^2 U-6 shows a strong binding to rHP1 but no binding to dHP1 at pH 7.5 (**Figure 3.9, Table 3.2**), which is consistent with previously reported results²²⁻²⁵. It is probable that a PNA is more compatible with the relatively deep and narrow major groove of an RNA duplex with a relatively small twist (compared to a DNA duplex), resulting in the selective recognition of RNA duplex over a DNA duplex by PNAs. Overall, it is convincing to utilize s^2 U-modified PNAs to selectively target dsRNAs containing an A-U pair.

Figure 3.9. Nondenaturing PAGE for PNA C-s²U-6 binding to DNA hairpin dHP1. The loaded samples contain 1 μ M of dHP1 in 20 μ L. The PNA concentrations in lanes from left to right are: 0, 0.2, 0.4, 1, 1.6, 2, 4, 10, 16, 20, 28, and 50 μ M, respectively.

3.3.3 Combining s²U and L modifications in PNAs facilitates selective binding to dsRNAs over ssRNAs

Incorporation of s²U residues into a PNA is expected to stabilize a PNA-RNA duplex and a PNA·RNA₂ triplex through the formation of stabilized Watson-Crick s²U-A pairs and s²U·A-U triples, respectively (**Figure 3.1e,f**)⁴⁰⁻⁴⁵. However, L-modified PNAs facilitates the formation of PNA·RNA-RNA triplex but destabilize PNA-RNA duplex, respectively, with the formation of stable L·G-C triples and unstable L-G Watson-Crick pairs (**Figure 3.1c,d**)²². Thus, it is important to test if PNAs containing s²U and L residues (**Figure 3.1c,e**) still show the selective recognition of dsRNAs over ssRNAs.

We studied the binding of the s²U-modified PNAs to ssRNAs (**Figure 3.10, 3.12b**) by thermal melting. Compared to P1, the s²U-incorporated PNA sequences have relatively high melting temperatures (with the melting temperature increased by 5-10 °C, based on the heating curves) (**Figure 3.10a, 3.12b**). Importantly, incorporating L residues into the s²U-modified PNAs results in significantly decreased melting temperatures in the formation of PNA-RNA duplexes (with the melting temperatures

decreased by 5-15 °C, **Figure 3.10b, 3.12b**). Thus, incorporating s^2U into dbPNAs containing L and other dsRNA-specific residues can facilitate enhanced and selective recognition of dsRNAs over ssRNAs.

Figure 3.10. Thermal melting results for RNA-PNA duplexes containing s^2U - modified PNAs. All samples contain 5 μM RNA and 5 μM PNA. The heating curves for P1, P5, and s^2U -modified PNAs with ssRNA1 (5'-AGAGAAAG-3') suggest that incorporation of L and s^2U is useful for enhancing the binding to a dsRNA with minimal binding to an ssRNA.

Figure 3.11. Thermal melting results for potential (miR-198-ssRNA)-PNA duplexes. All samples contain 5 μ M RNA and 5 μ M PNA. The data suggest no strong binding between the PNAs and the miR-198-ssRNA (5'-GGGAGAUAGG-3').

Figure 3.12. Summary of binding properties. (a) Comparison of K_d values (obtained by PAGE) of rHP1-Cy3 binding to PNAs P1, P5, s²U-modified PNAs, and L- and s²U-modified PNAs. (b) Comparison of thermal melting temperature (T_m) values (heating) of PNA P1, P5, s²U-modified, and L- and s²U-modified PNAs binding to ssRNA1.

3.3.4 Binding of dbPNAs to miR-198 hairpin precursor structure results in a substrate-specific inhibition of dicing *in vitro*

Our previous work shows that a 6-mer dbPNA (NH₂-Lys-LLLTLT-CONH₂) binds to microRNA-198 (miR-198) hairpin⁵⁶⁻⁵⁷, through formation of base triples with the RNA base pairs from position 7 to 12 ($K_d = 2.1 \mu\text{M}$ at pH 7.5, **Figure 3.12a**)²⁴. Unfortunately, upon replacing the two T residues with ^{Br}U (^{Br}U = 5-bromouracil) to obtain NH₂-Lys-LLL^{Br}UL^{Br}U-CONH₂, we observed a destabilization effect for the PNA·RNA₂ triplex formation ($K_d = 5.6 \mu\text{M}$ at pH 7.5), probably due to an unfavorable stacking interactions in a L^{Br}U step²⁴. It is important to test if s²U modification can be applied to enhance the targeting of biological RNAs such as pre-miR-198.

To enhance the binding, we designed two lengthened 8-mer dbPNAs with and without s²U modification (**Figure 3.13c,d**) for the targeting of pre-miR-198. The bulge U residue in pre-miR-198 is expected to be fully looped out upon dbPNA binding to allow for the triplex formation. Our PAGE results suggest that dbPNA-198 (NH₂-Lys-LLTLTTLL-CONH₂, without the incorporation of s²U residues) indeed binds to miR-198 hairpin precursor at pH 7.5 (with a K_d value of $0.16 \pm 0.07 \mu\text{M}$, **Figure 3.13f, 3.14**). Replacing the T residue at 5th position in dbPNA-198 with an s²U (dbPNA-198-s²U, NH₂-Lys-LLTLs²UTLL-CONH₂,) results in an enhanced binding at pH 7.5 (with a K_d value of $0.08 \pm 0.05 \mu\text{M}$, **Figure 3.13g, 3.14**). Thus, our data further suggest that, compared to a ^{Br}U modification²⁴, an s²U modification is advantageous in that incorporation of an Ls²U step stabilizes a PNA·RNA₂ triplex formation. Significantly, increasing the pH from 7.5 to 8.0 causes no weakening of the binding (**Figure 3.13f-i**). Importantly, both dbPNA-198-s²U and dbPNA-198 show relatively low melting

temperatures (< 31 °C) in binding to single-stranded miR-198-ssRNA (**Figure 3.13b, 3.11**). Thus, our data clearly show that dbPNAs incorporated with s²U and L have enhanced and selective binding to dsRNAs over ssRNAs.

Encouraged by the observation that binding of dbPNA-198 could be enhanced by the introduction of an s²U residue, we tested whether dicing of full-length pre-miR-198 (**Figure 3.15**) could be inhibited *in vitro*. We prepared 5'-³²P-labeled pre-miR-198 and incubated the substrate with purified Dicer enzyme. After incubation, a ~20-nt mature miRNA product was observed in the absence of PNA (**Figure 3.16a**). Similar dicing activity was observed in the presence of a control PNA (dbPNA-NC) at the concentrations of 0.2-0.6 μM. Addition of dbPNA-198 at 0.4 and 0.6 μM noticeably decreased dicing efficiency, and the inhibition effect was stronger for dbPNA-198-s²U with a clear reduction of the product even at 0.2 μM. To test if the effect of the PNAs against pre-miRNA dicing is specific to pre-miR-198, we prepared a control pre-miRNA hairpin with an unrelated stem sequence (pre-miR-124, **Figure 3.15**). As expected, no strong effect against pre-miR-124 dicing was observed for the PNAs (**Figure 3.16b**), although sequence non-specific effects were seen at higher concentrations (> 0.8 μM, data not shown). The results demonstrate that specific binding of a dbPNA to a pre-miRNA inhibits dicing *in vitro* and s²U modification can enhance inhibition, consistent with the lower *K_d* for dbPNA-198-s²U compared with dbPNA-198.

Figure 3.14. K_d determination by nondenaturing PAGE for PNA dbPNA-198 and dbPNA-198-s²U binding to miR-198 hairpin precursor at pH 7.5 and 8.0. For panels b and d, the points at 1, 2, and 4 μM were removed to facilitate fitting.

Figure 3.15. Secondary structure models of full-length pre-miR-198 and pre-miR-124-2 used for dicing assay. The mature miRNA sequences are shown in red. The sequence of dbPNA-198 designed for binding to pre-miR-198 is shown in blue. Binding of dbPNA-198 and dbPNA-198-s²U inhibits the binding of Dicer and thus Dicer cleavage activity at the junctions of red and black sequences. Pre-miR-124-2 is used as a negative control.

Figure 3.16. Sequence-specific inhibition of pre-miR-198 dicing by dbPNAs. Dose response analysis was performed using ~8 nM full-length pre-miR substrates, 0.017 unit/ul Dicer and the PNA at the indicated concentrations. ³²P-labeled full-length pre-miR-198 **(a)** and pre-miR-124-2 **(b)** were used (see Materials and Methods and **Figure 3.15**). The substrate and product signals on the autoradiography were quantified and plotted. The data points on the charts are color coded according to the PNA used (dbPNA-NC: Green, dbPNA-198: Dark blue, dbPNA-198-s²U: Light blue). Pre-miR-198 dicing was inhibited most strongly by dbPNA-198-s²U, followed by dbPNA-198, while no clear effects were seen for pre-miR-124 dicing, demonstrating sequence-specific inhibition of dbPNA-198-s²U and dbPNA-198.

3.4 Conclusions

In summary, we have demonstrated that s^2U can be combined with L for recognition of dsRNAs containing an A-U and G-C pairs, respectively, with minimal binding to dsDNAs and ssRNAs. A T-to- s^2U substitution involves the replacements of the methyl and carbonyl groups with a hydrogen and a thiocarbonyl group, respectively, both of which can help facilitate improved binding to dsRNAs. A U-to- s^2U substitution in PNAs results in further enhanced base stacking and hydrogen bonding interactions^{11, 40-47}. The binding enhancement of s^2U incorporation is generally applicable with s^2U immediately adjacent to T, C, and L with no significant sequence context dependence. Our previous work shows that halouracils adjacent to L may show sequence dependence in the stabilization of PNA·RNA-RNA triplexes²⁴, suggesting that s^2U is advantageous for general stabilization of PNA·RNA-RNA triplexes. PNAs with s^2U and other modified residues such as L may have broad applications in targeting various RNAs including pre-miRNAs as demonstrated in this paper. Our work provides the foundation for developing dbPNAs containing s^2U , L, and other modified residues for enhanced dsRNA recognition in a further broadened sequence context.

References

1. Fox, K. R.; Brown, T., Formation of stable DNA triplexes. *Biochemical Society transactions* **2011**, *39* (2), 629-34.
2. Devi, G.; Zhou, Y.; Zhong, Z.; Toh, D. F. K.; Chen, G., RNA Triplexes: from structural principles to biological and biotech applications. *Wiley Interdiscip Rev RNA* **2015**, *6*, 111-28.
3. Bacolla, A.; Wang, G.; Vasquez, K. M., New Perspectives on DNA and RNA Triplexes As Effectors of Biological Activity. *PLoS Genet* **2015**, *11* (12), e1005696.
4. Buske, F. A.; Mattick, J. S.; Bailey, T. L., Potential in vivo roles of nucleic acid triple-helices. *RNA Biol* **2011**, *8* (3), 427-39.
5. Conrad, N. K., The emerging role of triple helices in RNA biology. *Wiley Interdiscip Rev RNA* **2014**, *5* (1), 15-29.
6. Li, Y.; Syed, J.; Sugiyama, H., RNA-DNA Triplex Formation by Long Noncoding RNAs. *Cell Chem Biol* **2016**, *23* (11), 1325-1333.
7. Roberts, R. W.; Crothers, D. M., Stability and properties of double and triple helices: dramatic effects of RNA or DNA backbone composition. *Science* **1992**, *258* (5087), 1463-6.
8. Escude, C.; Francois, J. C.; Sun, J. S.; Ott, G.; Sprinzl, M.; Garestier, T.; Helene, C., Stability of triple helices containing RNA and DNA strands: experimental and molecular modeling studies. *Nucleic Acids Res* **1993**, *21* (24), 5547-53.
9. Han, H.; Dervan, P. B., Sequence-specific recognition of double helical RNA and RNA-DNA by triple helix formation. *Proceedings of the National Academy of Sciences of the United States of America* **1993**, *90* (9), 3806-10.
10. Wang, S.; Kool, E. T., Relative stabilities of triple helices composed of combinations of DNA, RNA and 2'-O-methyl-RNA backbones: chimeric circular oligonucleotides as probes. *Nucleic Acids Res* **1995**, *23* (7), 1157-64.
11. Ohkubo, A.; Yamada, K.; Ito, Y.; Yoshimura, K.; Miyauchi, K.; Kanamori, T.; Masaki, Y.; Seio, K.; Yuasa, H.; Sekine, M., Synthesis and triplex-forming properties of oligonucleotides capable of recognizing corresponding DNA duplexes containing four base pairs. *Nucleic Acids Res* **2015**, *43* (12), 5675-86.
12. Nielsen, P. E.; Egholm, M.; Berg, R. H.; Buchardt, O., Sequence-selective recognition of DNA by strand displacement with a thymine-substituted polyamide. *Science* **1991**, *254* (5037), 1497-500.
13. Egholm, M.; Buchardt, O.; Christensen, L.; Behrens, C.; Freier, S. M.; Driver, D. A.; Berg, R. H.; Kim, S. K.; Norden, B.; Nielsen, P. E., PNA hybridizes to complementary oligonucleotides obeying the Watson-Crick hydrogen-bonding rules. *Nature* **1993**, *365* (6446), 566-8.
14. Hyrup, B.; Nielsen, P. E., Peptide nucleic acids (PNA): synthesis, properties and potential applications. *Bioorg Med Chem* **1996**, *4* (1), 5-23.
15. Demidov, V.; Frank-Kamenetskii, M. D.; Egholm, M.; Buchardt, O.; Nielsen, P. E., Sequence selective double strand DNA cleavage by peptide nucleic acid (PNA) targeting using nuclease S1. *Nucleic Acids Res* **1993**, *21* (9), 2103-7.
16. Winssinger, N.; Damoiseaux, R.; Tully, D. C.; Geierstanger, B. H.; Burdick, K.; Harris, J. L., PNA-encoded protease substrate microarrays. *Chem Biol* **2004**, *11* (10), 1351-60.
17. Komiyama, M.; Ye, S.; Liang, X.; Yamamoto, Y.; Tomita, T.; Zhou, J. M.; Aburatani, H., PNA for one-base differentiating protection of DNA from nuclease and its use for SNPs detection. *J Am Chem Soc* **2003**, *125* (13), 3758-62.
18. Li, M.; Zengeya, T.; Rozners, E., Short peptide nucleic acids bind strongly to homopurine tract of double helical RNA at pH 5.5. *J Am Chem Soc* **2010**, *132* (25), 8676-81.
19. Rozners, E., Recent advances in chemical modification of Peptide nucleic acids. *J Nucleic Acids* **2012**, *2012*, 518162.
20. Zengeya, T.; Gupta, P.; Rozners, E., Sequence selective recognition of double-stranded RNA using triple helix-forming peptide nucleic acids. *Methods Mol Biol* **2014**, *1050*, 83-94.

21. Patil, K. M.; Chen, G., Recognition of RNA sequence and structure by duplex and triplex formation: targeting miRNA and pre-miRNA. In *Modified Nucleic Acids in Biology and Medicine*, Erdmann, V. A.; Jurga, S.; Barciszewski, J., Eds. Springer: 2016; pp 299-317.
22. Devi, G.; Yuan, Z.; Lu, Y.; Zhao, Y.; Chen, G., Incorporation of thio-pseudoisocytosine into triplex-forming peptide nucleic acids for enhanced recognition of RNA duplexes. *Nucleic Acids Res* **2014**, *42*, 4008-18.
23. Toh, D.-F.; Devi, G.; Patil, K. M.; Qu, Q.; Maraswami, M.; Xiao, Y.; Loh, T.-P.; Zhao, Y.; Chen, G., Incorporating a guanidine-modified cytosine base into triplex-forming PNAs for the recognition of a C-G pyrimidine-purine inversion site of an RNA duplex. *Nucleic Acids Res* **2016**, *44*, 9071-82.
24. Patil, K. M.; Toh, D. K.; Yuan, Z.; Meng, Z.; Shu, Z.; Zhang, H.; Ong, A. A. L.; Krishna, M. S.; Lu, L.; Lu, Y.; Chen, G., Incorporating uracil and 5-halouracils into short peptide nucleic acids for enhanced recognition of A-U pairs in dsRNAs. *Nucleic Acids Res* **2018**, *46* (15), 7506-7521.
25. Ong, A. A. L., Toh, D.F.K., Patil, K.M., Meng, Z., Yuan, Z., Krishna, M.S., Devi, G. Haruehanroengra, P., Lu, Y., Xia, K., Okamura, K., Sheng, J., and Chen, G. , Recognition of U-G, U-A, and C-G pairs by double-stranded RNA-binding PNAs incorporated with an artificial nucleobase. *Biochemistry* **2019**, *58*, 1319-1331.
26. Toh, D.-F. K.; Patil, K. M.; Chen, G., Sequence-specific and Selective Recognition of Double-stranded RNAs over Single-stranded RNAs by Chemically Modified Peptide Nucleic Acids. *J Vis Exp* **2017**, *127*, e56221.
27. Puah, R. Y.; Jia, H.; Maraswami, M.; Toh, D.-F. K.; Ero, R.; Yang, L.; Patil, K. M.; Ong, A. A. L.; Krishna, M. S.; Sun, R.; Tong, C.; Huang, M.; Chen, X.; Loh, T. P.; Gao, Y. G.; Liu, D. X.; Chen, G., Selective Binding to mRNA Duplex Regions by Chemically Modified Peptide Nucleic Acids Stimulates Ribosomal Frameshifting. *Biochemistry* **2018**, *57* (1), 149-159.
28. Keszy, J., Patil, K.M., Kumar, S.M. Shu, Z., Yee, Y.H., Zimmermann, L., Ong, A.A.L., Toh, D.-F. K., Krishna, M.S., Yang, L., Decout, J.L., Luo, D., Prabakaran, M., Chen, G., and Kierzek, E. , A short chemically modified dsRNA-binding PNA (dbPNA) inhibits influenza viral replication by targeting viral RNA panhandle structure. *Bioconjugate Chem* **2019**, *30* (3), 931-943.
29. Krishna, M. S., Toh, D.F.K., Meng, Z., Ong, A.A.L., Wang, Z., Lu, Y., Xia, K., Prabakaran, M., and Chen, G. , Sequence- and structure-specific probing of RNAs by short nucleobase-modified dsRNA-binding PNAs (dbPNAs) incorporating a fluorescent light-up uracil analog. *Anal Chem* **2019**, *91*, 5331-5338.
30. Zengeya, T.; Gupta, P.; Rozners, E., Triple-helical recognition of RNA using 2-aminopyridine-modified PNA at physiologically relevant conditions. *Angew Chem Int Ed Engl* **2012**, *51* (50), 12593-6.
31. Hnedzko, D.; McGee, D. W.; Karamitas, Y. A.; Rozners, E., Sequence-selective recognition of double-stranded RNA and enhanced cellular uptake of cationic nucleobase and backbone-modified peptide nucleic acids. *RNA* **2017**, *23* (1), 58-69.
32. Muse, O.; Zengeya, T.; Mwaura, J.; Hnedzko, D.; McGee, D. W.; Grewer, C. T.; Rozners, E., Sequence selective recognition of double-stranded RNA at physiologically relevant conditions using PNA-peptide conjugates. *ACS Chem Biol* **2013**, *8* (8), 1683-6.
33. Gupta, P.; Zengeya, T.; Rozners, E., Triple helical recognition of pyrimidine inversions in polypurine tracts of RNA by nucleobase-modified PNA. *Chem Commun* **2011**, *47* (39), 11125-7.
34. Kotikam, V.; Kennedy, S. D.; MacKay, J. A.; Rozners, E., Synthetic, Structural, and RNA Binding Studies on 2-Aminopyridine-Modified Triplex-Forming Peptide Nucleic Acids. *Chem Eur J* **2019**, *25* (17), 4367-4372.
35. Kim, K. T.; Chang, D. L.; Winssinger, N., Double-Stranded RNA-Specific Templated Reaction with Triplex Forming PNA. *Helvetica Chimica Acta* **2018**, *101* (3), e1700295.
36. Virta, P. M.; Tahtinen, V.; Granqvist, L.; Murtola, M.; Stromberg, R., 19F NMR spectroscopic analysis of the binding modes in triple helical PNA/microRNA-complexes. *Chem Eur J* **2017**, *23*, 7113-7124.

37. Sato, T.; Sato, Y.; Nishizawa, S., Triplex-Forming Peptide Nucleic Acid Probe Having Thiazole Orange as a Base Surrogate for Fluorescence Sensing of Double-stranded RNA. *J Am Chem Soc* **2016**, *138* (30), 9397-400.
38. Egholm, M.; Christensen, L.; Dueholm, K. L.; Buchardt, O.; Coull, J.; Nielsen, P. E., Efficient pH-independent sequence-specific DNA binding by pseudoisocytosine-containing bis-PNA. *Nucleic Acids Res* **1995**, *23* (2), 217-22.
39. Ono, A.; Tso, P. O. P.; Kan, L. S., Triplex formation of oligonucleotides containing 2'-O-methylpseudoisocytidine in substitution for 2'-deoxycytidine. *J Am Chem Soc* **1991**, *113* (10), 4032-4033.
40. Okamoto, I.; Seio, K.; Sekine, M., Triplex forming ability of oligonucleotides containing 2'-O-methyl-2-thiouridine or 2-thiothymidine. *Bioorg Med Chem Lett* **2006**, *16* (12), 3334-6.
41. Testa, S. M.; Disney, M. D.; Turner, D. H.; Kierzek, R., Thermodynamics of RNA-RNA duplexes with 2-or 4-thiouridines: Implications for antisense design and targeting a group I intron. *Biochemistry* **1999**, *38* (50), 16655-16662.
42. Carlucci, M.; Kierzek, E.; Olejnik, A.; Turner, D. H.; Kierzek, R., Chemical synthesis of LNA-2-thiouridine and its influence on stability and selectivity of oligonucleotide binding to RNA. *Biochemistry* **2009**, *48* (46), 10882-93.
43. Kumar, R. K.; Davis, D. R., Synthesis and studies on the effect of 2-thiouridine and 4-thiouridine on sugar conformation and RNA duplex stability. *Nucleic acids research* **1997**, *25* (6), 1272-1280.
44. Szabat, M.; Kierzek, E.; Kierzek, R., Modified RNA triplexes: Thermodynamics, structure and biological potential. *Sci Rep* **2018**, *8* (1), 13023.
45. Zhou, Y.; Kierzek, E.; Loo, Z. P.; Antonio, M.; Yau, Y. H.; Chuah, Y. W.; Geifman-Shochat, S.; Kierzek, R.; Chen, G., Recognition of RNA duplexes by chemically modified triplex-forming oligonucleotides. *Nucleic Acids Res* **2013**, *41* (13), 6664-73.
46. Sismour, A. M.; Benner, S. A., The use of thymidine analogs to improve the replication of an extra DNA base pair: a synthetic biological system. *Nucleic Acids Res* **2005**, *33* (17), 5640-6.
47. Sheng, J.; Larsen, A.; Heuberger, B. D.; Blain, J. C.; Szostak, J. W., Crystal structure studies of RNA duplexes containing s(2)U:A and s(2)U:U base pairs. *J Am Chem Soc* **2014**, *136* (39), 13916-24.
48. Yamamoto, Y.; Yokoyama, S.; Miyazawa, T.; Watanabe, K.; Higuchi, S., NMR Analyses on the Molecular Mechanism of the Conformational Rigidity of 2-Thioribothymidine, a Modified Nucleoside in Extreme Thermophile Transfer-RNAs. *Febs Letters* **1983**, *157* (1), 95-99.
49. Gilli, P.; Pretto, L.; Bertolasi, V.; Gilli, G., Predicting hydrogen-bond strengths from acid-base molecular properties. The pK(a) slide rule: toward the solution of a long-lasting problem. *Accounts of chemical research* **2009**, *42* (1), 33-44.
50. Lohse, J.; Dahl, O.; Nielsen, P. E., Double duplex invasion by peptide nucleic acid: A general principle for sequence-specific targeting of double-stranded DNA. *Proc Natl Acad Sci USA* **1999**, *96* (21), 11804-11808.
51. Privat, E. J.; Sowers, L. C., A proposed mechanism for the mutagenicity of 5-formyluracil. *Mutat Res* **1996**, *354* (2), 151-6.
52. Povsic, T. J.; Dervan, P. B., Triple helix formation by oligonucleotides on DNA extended to the physiological pH range. *J Am Chem Soc* **1989**, *111* (8), 3059-3061.
53. Asensio, J. L.; Lane, A. N.; Dhesi, J.; Bergqvist, S.; Brown, T., The contribution of cytosine protonation to the stability of parallel DNA triple helices. *J Mol Biol* **1998**, *275* (5), 811-22.
54. Eldrup, A. B.; Dahl, O.; Nielsen, P. E., A novel peptide nucleic acid monomer for recognition of thymine in triple-helix structures. *J Am Chem Soc* **1997**, *119* (45), 11116-11117.
55. Semenyuk, A.; Darian, E.; Liu, J.; Majumdar, A.; Cuenoud, B.; Miller, P. S.; Mackerell, A. D., Jr.; Seidman, M. M., Targeting of an interrupted polypurine:polypyrimidine

sequence in mammalian cells by a triplex-forming oligonucleotide containing a novel base analogue. *Biochemistry* **2010**, *49* (36), 7867-78.

56. Kozomara, A.; Griffiths-Jones, S., miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res* **2014**, *42* (Database issue), D68-73.

57. Sundaram, G. M.; Common, J. E.; Gopal, F. E.; Srikanta, S.; Lakshman, K.; Lunny, D. P.; Lim, T. C.; Tanavde, V.; Lane, E. B.; Sampath, P., 'See-saw' expression of microRNA-198 and FSTL1 from a single transcript in wound healing. *Nature* **2013**, *495* (7439), 103-6.

Chapter 4

Recognition of RNA secondary structures with a programmable

PNA-based 7-letter platform

4.1 Introduction

RNAs perform a diverse array of catalytic and regulatory functions in viruses and cells. RNA structures are mainly stabilized by base paired double-stranded (ds) stem regions¹. Together with single-stranded (ss) loop regions, RNAs can fold into complex secondary and tertiary structures², facilitating the molecular recognition of RNA structures by small molecules³⁻⁵ and peptides/proteins⁶. Antisense oligonucleotides (ASOs) are a successful class of RNA-targeting drugs which bind to target RNAs by Watson-Crick base pairing. A few ASOs have clinical applications for treatment of certain diseases such as Duchenne Muscular Dystrophy (DMD) and Spinal Muscular Atrophy (SMA)⁷. ASOs are often 13-25 nucleotides in length⁷.

Currently, programmable RNA structure-specific and tight-binding ligands are relatively unexplored. Peptide nucleic acid (PNA) is characterized by a neutral, peptide-like backbone, which imparts chemical stability and nuclease resistance (**Figure 4.1**)⁸. PNA possesses stronger hybridization towards complementary DNA and RNA to form PNA-DNA and PNA-RNA duplexes which are thermally more stable as compared to DNA-DNA and RNA-RNA duplexes⁹. Strong binding of a PNA to a complementary sequence allows the strand invasion of a dsRNA or a dsDNA, with one of the strands of the dsRNA/dsDNA displaced by the PNA. Importantly, PNAs show promising bioactivities in animal models¹⁰⁻¹⁵. However,

traditional antisense PNAs (asPNAs) do not bind to RNAs through the recognition of RNA structures, which may give rise off-target binding.

Short (e.g., 10 mer) dsRNA-binding PNAs (dbPNAs) incorporated with modified nucleobases can form sequence-specific triplexes with dsRNAs through the formation of consecutive PNA·RNA-RNA base triples (**Figure 4.1a-f**), with significantly weakened binding to dsDNAs¹⁶⁻²⁴ and ssRNAs¹⁶⁻²⁰. Naturally occurring dsRNAs in functional RNAs are often interrupted by ssRNA loop regions, which may limit the application of dbPNAs. Thus, it is critical to develop a scaffold for the simultaneous recognition of dsRNA and ssRNA regions. We herein designed novel PNAs combining dbPNAs (containing bases of T, L, Q, and E or S, **Figure 4.1a-f**)^{16,17,19-24} and traditional antisense PNAs (asPNAs, containing unmodified bases T, C, G, and A) (designated as daPNAs, see **Figure 4.1g,h**) for the simultaneous recognition of dsRNA and immediately adjacent ssRNA regions. We show that novel daPNAs can be utilized to target dsRNA-ssRNA regions with high sequence/structure specificity and strong binding affinity, potentially overcoming the challenges facing other approaches such as small molecules and traditional ASOs including asPNAs.

We designed a series of short daPNAs with the asPNA segment antiparallel to the ssRNA region and the dbPNA segment forming a parallel triplex with the dsRNA region (see **Figure 4.1g for example**, see all the PNAs studied in **Table A2.3 and Figure A2.3**). Our structural modelling study^{18,25} suggests that such a novel fold is structurally compatible without significant distortion in the backbones and bases (**Figures 4.1h and 4.2**).

Figure 4.1. Chemical structures of PNA-RNA-RNA base triples and PNA daPNA-tau and modelled structure of a daPNA-bound RNA. **(a-f)** Base triples of T·A-U, C⁺·G-C, L·G-C, Q·C-G, E·U-A and S·U-G¹⁶⁻²⁴. Compared to unmodified C base, modified L base shows enhanced binding affinity towards G-C pair with reduced pH dependence¹⁶. L base can also recognize a G-U pair. The sugar-phosphate backbone of RNA is represent by "R". Orange dashed lines indicate hydrogen bonding interactions. **(g)** daPNA-tau (NH₂-Lys-TQTQTGCCG-CONH₂). The dbPNA and asPNA segments are shown in blue and green, respectively. **(h)** A modelled daPNA·(dsRNA-ssRNA) structure (bottom) with the sequence and secondary structure shown on the top. The Hoogsteen and Watson-Crick pairs formed between PNA and RNA are shown as orange dots and lines, respectively.

Figure 4.2. Structural modelling procedure for a daPNA·(dsRNA-ssRNA) structure.

4.2 Materials and methods

4.2.1 Synthesis of PNA oligomers

Reverse-phase high performance liquid chromatography (RP-HPLC) purified RNA oligonucleotides were purchased from Sigma-Aldrich Singapore. The PNA monomers thymine (T), cytosine (C), Guanine (G) and Adenine (A) were purchased from ASM Research Chemicals. PNA monomers L, Q, E, and S were synthesized following the reported methods^{16,17,20,26}. PNA oligomers were synthesized manually using Boc chemistry via a Solid-Phase Peptide Synthesis (SPPS) protocol. 4-Methylbenzhydrylamine hydrochloride (MBHA·HCl) polystyrene resins were used. The loading value used for the synthesis of the oligomers was 0.3 mmol/g and acetic anhydride was used as the capping reagent. Benzotriazol-1-yl-oxytripyrrolidinophosphonium hexafluorophosphate (PyBOP) and *N,N*-diisopropylethylamine (DIPEA) were used as the coupling reagent. The oligomerization of PNA was monitored by Kaiser test. Cleavage of the PNA oligomers was done using trifluoroacetic acid (TFA) and trifluoromethanesulfonic acid (TFMSA) method, after which the oligomers were precipitated with diethyl ether, dissolved in deionized water and purified by reverse-phase high performance liquid chromatography (RP-HPLC) using water–ACN–0.1% TFA as the mobile phase. Matrix-assisted laser desorption/ionization-time of flight (MALDI-TOF) analysis was used to characterize the oligomers (**Table A2.3**, **Figure A2.3**), with the use of α -cyano-4-hydroxycinnamic acid (CHCA) as the sample crystallization matrix. The extinction coefficients of T, C, L, Q, E, and S were estimated to be 8.8, 7.3, 7.3, 7.3, 6.0, and 16.4 mM⁻¹cm⁻¹, respectively, at 260 nm.

4.2.2 Molecular modelling

We modelled the structure shown in Figure 1h, based on the major-groove PNA·RNA-RNA triplex structure we previously modelled¹⁸. We removed the bottom

triple and base pairs (shown in red, **Figure 4.2**, left panel). A three-base-pair PNA-RNA duplex segment (shown in green, **Figure 4.2**, middle panel) was extracted from a reported structure (PDB ID: 1PNN), followed by connecting the PNA-RNA duplex segment to the remaining PNA·RNA-RNA triplex structure (**Figure 4.2**, middle panel). The obtained structure was then solvated by TIP3P water molecules, followed by optimization, and NVT and NPT equilibrium. Then a 100-ns product MD was conducted, with the trajectory used for clustering. The center structure of the most populated cluster was regarded as the representative structure. The MD simulations were conducted using GROMACS 5 package²⁵.

4.2.3 Nondenaturing polyacrylamide gel electrophoresis

Nondenaturing (12wt%) polyacrylamide gel electrophoresis (PAGE) experiments were conducted with incubation buffer containing 200 mM NaCl, 0.5 mM EDTA, 20 mM HEPES at pH 7.5. The loading volume for samples containing Cyanine 3 (Cy3)-labelled RNA hairpins were 20 μ L while the loading volume for samples containing non-labelled RNA hairpins were 25 μ L. The concentration of Cy3-labelled RNA hairpins is 20 nM while the concentration of non-labelled RNAs is 0.25 μ M. The samples were prepared by snap cooling of the hairpin, followed by annealing with PNA oligomers by slow cooling from 65°C to room temperature and incubation at 4°C overnight. Prior to loading the samples into the wells, 35% glycerol (20% of the total volume) was added to the sample mixtures. 1 $\times$ Tris–borate–EDTA (TBE) buffer, pH 8.3 was used as the running buffer for all gel experiments. The gel was run at 4 °C at 250 V for 5 h. The gels for the Cy3-labelled samples were imaged Typhoon Trio Variable Mode Imager. The gels for the non-labelled samples were stained with ethidium bromide and imaged by the Typhoon Trio Variable Mode Imager.

4.2.4 Data analysis

The fraction of triplex formation was calculated based on equation $Y = I_{\text{triplex}} / (I_{\text{triplex}} + aI_{\text{duplex}})$. The band intensities were normalized according to $a = I_{\text{triplex max}} / I_{\text{duplex max}}$. $I_{\text{duplex max}}$ is the band intensity for hairpin alone without the addition of PNA. $I_{\text{triplex max}}$ is the triplex band intensity with the highest concentration of PNA added. The data were fit to the equation: $Y = Y_0 + (B/2/R_0)(R_0 + X + K_d - ((R_0 + X + K_d)^2 - 4R_0X)^{1/2})$ where R_0 is the RNA hairpin concentration (50 nM). Y_0 and B are the minimum and maximum fraction of triplex formation respectively. X is the total PNA concentration and K_d is the dissociation constant.

4.3 Results and Discussion

4.3.1 Recognition of dsRNA containing a 3'-overhang poly (U) tail by combining dsRNA-binding PNAs (dbPNAs) and antisense peptide nucleic acids (asPNAs)

We constructed a hairpin containing a 6-nucleotide (nt) 3' dangling end as a model of dsRNA-ssRNA motif (**Figure 4.3**). We carried out nondenaturing polyacrylamide gel electrophoresis (PAGE) to characterize the recognition of the model dsRNA-ssRNA junction by PNAs (**Figure 4.3a-e**). The 10-mer daPNA-1 (NH₂-Lys-TAAAAALTLT-CONH₂) (**Figure 4.3a**) contains a 6-mer segment (underlined) for the formation of an antiparallel PNA-RNA duplex with the 6-nt 3' dangling end and a 4-mer dbPNA segment for the formation of a major-groove PNA·dsRNA triplex. The nondenaturing PAGE result indicates that daPNA-1 indeed binds tightly to the model dsRNA-ssRNA junction (with a $K_d < 0.2 \mu\text{M}$) at pH 7.5, **Figure 4.3f**). Two control daPNAs daPNA-NC1 (NH₂-Lys-TCACACLTLT-CONH₂) and daPNA-NC2 (NH₂-Lys-TAAAAAETET-CONH₂) (**Figure 4.3b,c**), however, show no binding to the model dsRNA-ssRNA junction (**Figure 4.3g,h**). The asPNA segment of daPNA-NC2 can only form five A-U and one T-A PNA-RNA pairs without a more stabilizing PNA-

PNA daPNA-NC1 and daPNA-NC2 show no binding to model dsRNA-ssRNA junction. The gels shown are representative gels.

4.3.2 Targeting dsRNA-ssRNA junctions presents in Tau pre-mRNA hairpin structure by combining dsRNA-binding PNAs (dbPNAs) and antisense peptide nucleic acids (asPNAs)

We next tested the generality of the daPNAs platform by designing a daPNA targeting a biomedically important regulatory tau pre-mRNA hairpin structure (**Figure 4.4**) located at the junction site between exon 10 and intron 10²⁷⁻³⁰. Point mutations in this splice site hairpin structure results in (de)stabilizing of the hairpin and thus aberrant alternative splicing leading to neurodegenerative diseases such as frontotemporal dementia with Parkinsonism linked to chromosome 17 (FTDP-17)²⁹⁻³³. Significantly, our nondenaturing PAGE result suggests that 9-mer daPNA-tau (NH₂-Lys-TQTQTGCCG-CONH₂) (**Figure 4.4a**) binds strongly to the tau pre-mRNA splice site hairpin (tau-Cy3) at pH 7.5 ($K_d = 0.03 \pm 0.01 \mu\text{M}$, **Figure 4.4g**). Importantly, a control daPNA-tau-NC (NH₂-Lys-TETETGCCG-CONH₂) (**Figure 4.4b**) shows no observable binding to tau pre-mRNA hairpin (**Figures 4.4h**), which again suggests that the 4-mer dbPNA segment in daPNA-tau is critical in enhancing the binding, even though the 5-mer asPNA segment is G/C rich. dbPNA-NC (NH₂-Lys-TQTQ-CONH₂) (**Figure 4.4c**) does not affect the gel shift of tau-Cy3 hairpin (**Figure 4.4i**), consistent with the fact that an asPNA segment is needed to cooperatively stabilize a 4-base-triple triplex formed by dbPNA-NC.

We also tested the binding of a 10-mer asPNA-tau (NH₂-Lys-ACACTGCCGC-CONH₂) designed to be complementary to the 5' arm of tau pre-mRNA hairpin (**Figure 4.4d,e**). Interestingly, 10-mer asPNA-tau binds to the tau hairpin (tau-Cy3) (with a K_d of $\sim 0.03 \mu\text{M}$) (**Figures 4.4j**), with the binding affinity comparable to 9-mer daPNA-tau. The advantage of a daPNA compared to a traditional asPNA is that daPNA is sequence and structure specific and has a relatively lower energy barrier for melting the preformed RNA structures. In this case, the relatively weak bottom stem of tau pre-mRNA splice hairpin structure^{28,30} is targeted by strand invasion by a 5-mer asPNA segment of daPNA-tau to form a PNA-RNA duplex, which is cooperatively stabilized by a 4-base-triple PNA·dsRNA triplex (**Figure 4.4f**).

Figure 4.4. Targeting tau pre-mRNA splice site hairpin structure by daPNA. (a-d) PNAs studied. (e) Cy3 labelled tau pre-mRNA hairpin construct (tau-Cy3). (f) A complex structure formed between daPNA-tau and tau-Cy3. (g-j) Nondenaturing

PAGE data conducted at pH 7.5. The loaded RNA hairpins are at 20 nM in 20 μ L. The PNA concentrations in lanes from left to right are 0, 0.005, 0.007, 0.01, 0.015, 0.02, 0.03, 0.05, 0.07, 0.1, 0.15, and 0.2 μ M, respectively. **(g)** daPNA-tau (panel a) binds to tau-Cy3 with a K_d of 0.04 ± 0.01 μ M. **(h,i)** Control PNAs daPNA-tau-NC and dbPNA-NC (see the sequence in panels b and c) show weak or no binding to tau-Cy3. **(j)** asPNA-tau (see the sequence in panel d) binds to tau-Cy3 with a K_d of ~ 0.03 μ M. The gels shown are representative gels. The errors given are standard errors.

4.3.3 Targeting dsRNA-ssRNA junction presents in miRNA-26b hairpin structure by combining dsRNA-binding PNAs (dbPNAs) and antisense peptide nucleic acids (asPNAs)

We further tested the application of a daPNA in targeting miRNA-26b (miR-26b) hairpin precursor (**Figure 4.5**). Over production of mature miR-26b is related to taupathy and Alzheimer's disease^{34,35}. We have previously shown that 8-mer dbPNA-14 (NH₂-Lys-TLLTSTLL-CONH₂, **Figure 4.5a,d**) binds to miR-26b hairpin structure³⁶ relatively weakly at pH 7.5 ($K_d \sim 50$ μ M)²⁰. There are two distinct bands for the miR-26b hairpin construct²⁰, probably due to the presence of an alternatively folded secondary structure at the bottom part of the stem. We designed an 11-mer daPNA by adding a 3-mer asPNA segment at the N-terminal side of 8-mer dbPNA-14 (daPNA-26b, NH₂-Lys-TGGTLLTSTLL-CONH₂, **Figure 4.5b,e**). daPNA-26b indeed binds to miR-26b hairpin precursor structure ($K_d = 1.7 \pm 0.8$ μ M) at pH 7.5 (**Figures 4.5f**). Interestingly, we observe only one band for the daPNA-26b-bound structure (**Figure 4.5f**). Both bands for the free miR-26b hairpin disappear upon the addition of daPNA-26b, indicating that the two bands contain the same RNA sequence. Titration of a control PNA daPNA-NC1 (**Figure 4.3b**) does not show any observable effect on the free miR-26b hairpin, with the two distinct bands observed throughout the lanes

(Figure 4.5g). The results further suggest that the free miR-26b hairpin is flexible forming two conformations. The relatively unstable bottom part of the stem region near the terminal ends is expected to favor the binding by the 3-mer asPNA segment of daPNA-26b through strand invasion. It is remarkable that by simply adding a 3-mer asPNA segment to the 8-mer dbPNA-14 (Figure 4.5d,e) can enhance the binding.

Figure 4.5. Targeting miR-26b hairpin precursor by daPNA. (a,b) PNAs studied. (c) miR-26b hairpin precursor construct. A hairpin loop (boxed) is used to stabilize the construct for the binding studies. (d) A PNA·RNA₂ triplex formed between PNA dbPNA-14 (panel a) and miR-26b hairpin. Our previous work shows that dbPNA-14 binds to miR-26b hairpin weakly with a K_d of $\sim 50 \mu\text{M}$ ²⁰. (e) A complex formed between daPNA-26b (panel b) and miR-26b hairpin. (f,g) Nondenaturing PAGE data conducted at pH 7.5. The loaded miR-26b hairpin precursor is at 1 μM in 20 μL . The

PNA concentrations in lanes from left to right are 0, 0.2, 0.4, 1, 1.6, 2, 4, 10, 16, 20, 28, and 50 μM , respectively. (f) PNA daPNA-26b binds to miR-26b hairpin with a K_d of $1.7 \pm 0.8 \mu\text{M}$, which is in contrast to the weak binding by dbPNA-14 with a K_d of $\sim 50 \mu\text{M}$ ²⁰. (g) Control daPNA-NC1 (**Figure 4.3b**) shows no binding to miR-26b hairpin. The gels shown are representative gels. The errors given are standard errors.

Figure 4.6. K_d determination by nondenaturing PAGE conducted at pH 7.5. (a) daPNA-tau binding to Cy3 labelled tau pre-mRNA hairpin construct (tau-Cy3). (b) daPNA-26b binding to miR-26b hairpin precursor.

4.4 Conclusions

In summary, we have developed a novel and generally applicable binding mode by combining dbPNA and asPNA for the recognition of a dsRNA region and an immediately adjacent ssRNA region (dsRNA-ssRNA junction) through the simultaneous formation of an asPNA-RNA duplex and a major-groove dbPNA·RNA₂ triplex. We demonstrated the application of the new molecular recognition strategy in targeting tau pre-mRNA splice site hairpin and miR-26b hairpin precursor. A dsRNA-ssRNA junction structure can be pre-formed or induced by the binding of daPNAs (through the strand invasion of weakly formed dsRNA regions). Thus, the daPNAs

platform would serve as useful junction-specific molecular glues for the probing and targeting of many biologically and medically important RNA structures in transcriptomes.

References

- (1) Turner, D. H. Fundamental interactions in RNA: Questions answered and remaining. *Biopolymers* **2013**, *99*, 1097-1104.
- (2) Dethoff, E. A.; Boerneke, M. A.; Gokhale, N. S.; Muhire, B. M.; Martin, D. P.; Sacco, M. T.; McFadden, M. J.; Weinstein, J. B.; Messer, W. B.; Horner, S. M.; Weeks, K. M. Pervasive tertiary structure in the dengue virus RNA genome. *Proc Natl Acad Sci USA* **2018**, *115*, 11513-11518.
- (3) Thomas, J. R.; Hergenrother, P. J. Targeting RNA with small molecules. *Chem Rev* **2008**, *108*, 1171-1224.
- (4) Childs-Disney, J. L.; Disney, M. D. Approaches to Validate and Manipulate RNA Targets with Small Molecules in Cells. *Annu Rev Pharmacol Toxicol* **2016**, *56*, 123-140.
- (5) Warner, K. D.; Hajdin, C. E.; Weeks, K. M. Principles for targeting RNA with drug-like small molecules. *Nat Rev Drug Discov* **2018**, *17*, 547-558.
- (6) Chen, Y.; Yang, F.; Zubovic, L.; Pavelitz, T.; Yang, W.; Godin, K.; Walker, M.; Zheng, S.; Macchi, P.; Varani, G. Targeted inhibition of oncogenic miR-21 maturation with designed RNA-binding proteins. *Nat Chem Biol* **2016**, *12*, 717-723.
- (7) Shen, X.; Corey, D. R. Chemistry, mechanism and clinical status of antisense oligonucleotides and duplex RNAs. *Nucleic Acids Res* **2018**, *46*, 1584-1600.
- (8) Nielsen, P. E.; Egholm, M.; Berg, R. H.; Buchardt, O. Sequence-selective recognition of DNA by strand displacement with a thymine-substituted polyamide. *Science* **1991**, *254*, 1497-1500.
- (9) Egholm, M.; Buchardt, O.; Christensen, L.; Behrens, C.; Freier, S. M.; Driver, D. A.; Berg, R. H.; Kim, S. K.; Norden, B.; Nielsen, P. E. PNA hybridizes to complementary oligonucleotides obeying the Watson-Crick hydrogen-bonding rules. *Nature* **1993**, *365*, 566-568.
- (10) Fabani, M. M.; Abreu-Goodger, C.; Williams, D.; Lyons, P. A.; Torres, A. G.; Smith, K. G.; Enright, A. J.; Gait, M. J.; Vigorito, E. Efficient inhibition of miR-155 function in vivo by peptide nucleic acids. *Nucleic Acids Res* **2010**, *38*, 4466-4475.
- (11) Cheng, C. J.; Bahal, R.; Babar, I. A.; Pincus, Z.; Barrera, F.; Liu, C.; Svoronos, A.; Braddock, D. T.; Glazer, P. M.; Engelman, D. M.; Saltzman, W. M.; Slack, F. J. MicroRNA silencing for cancer therapy targeted to the tumour microenvironment. *Nature* **2015**, *518*, 107-110.
- (12) Quijano, E.; Bahal, R.; Ricciardi, A.; Saltzman, W. M.; Glazer, P. M. Therapeutic Peptide Nucleic Acids: Principles, Limitations, and Opportunities. *Yale J Biol Med* **2017**, *90*, 583-598.
- (13) Gupta, A.; Quijano, E.; Liu, Y.; Bahal, R.; Scanlon, S. E.; Song, E.; Hsieh, W. C.; Braddock, D. E.; Ly, D. H.; Saltzman, W. M.; Glazer, P. M. Anti-tumor Activity of miniPEG-gamma-Modified PNAs to Inhibit MicroRNA-210 for Cancer Therapy. *Mol Ther Nucleic Acids* **2017**, *9*, 111-119.
- (14) Ricciardi, A. S.; Bahal, R.; Farrelly, J. S.; Quijano, E.; Bianchi, A. H.; Luks, V. L.; Putman, R.; Lopez-Giraldez, F.; Coskun, S.; Song, E.; Liu, Y.; Hsieh, W. C.; Ly, D. H.; Stitelman, D. H.; Glazer, P. M.; Saltzman, W. M. In utero nanoparticle delivery for site-specific genome editing. *Nat Commun* **2018**, *9*, 2481.
- (15) Ozes, A. R.; Wang, Y.; Zong, X.; Fang, F.; Pilrose, J.; Nephew, K. P. Therapeutic targeting using tumor specific peptides inhibits long non-coding RNA HOTAIR activity in ovarian and breast cancer. *Sci Rep* **2017**, *7*, 894.
- (16) Devi, G.; Yuan, Z.; Lu, Y.; Zhao, Y.; Chen, G. Incorporation of thio-pseudoisocytosine into triplex-forming peptide nucleic acids for enhanced recognition of RNA duplexes. *Nucleic Acids Res* **2014**, *42*, 4008-4018.
- (17) Toh, D.-F.; Devi, G.; Patil, K. M.; Qu, Q.; Maraswami, M.; Xiao, Y.; Loh, T.-P.; Zhao, Y.; Chen, G. Incorporating a guanidine-modified cytosine base into triplex-forming PNAs for the recognition of a C-G pyrimidine-purine inversion site of an RNA duplex. *Nucleic Acids Res* **2016**, *44*, 9071-9082.

- (18) Patil, K. M.; Toh, D. K.; Yuan, Z.; Meng, Z.; Shu, Z.; Zhang, H.; Ong, A. A. L.; Krishna, M. S.; Lu, L.; Lu, Y.; Chen, G. Incorporating uracil and 5-halouracils into short peptide nucleic acids for enhanced recognition of A-U pairs in dsRNAs. *Nucleic Acids Res* **2018**, *46*, 7506-7521.
- (19) Keszy, J.; Patil, K.M., Kumar, S.M. Shu, Z., Yee, Y.H., Zimmermann, L., Ong, A.A.L., Toh, D.F.K., Krishna, M.S., Yang, L., Decout, J.L., Luo, D., Prabakaran, M., Chen, G., and Kierzek, E. . A short chemically modified dsRNA-binding PNA (dbPNA) inhibits influenza viral replication by targeting viral RNA panhandle structure. *Bioconjugate Chem* **2019**, *In Press*, doi: 10.1021/acs.bioconjchem.1029b00039.
- (20) Ong, A. A. L., Toh, D.F.K., Patil, K.M., Meng, Z., Yuan, Z., Krishna, M.S., Devi, G. Haruehanroengra, P., Lu, Y., Xia, K., Okamura, K., Sheng, J., and Chen, G. . Recognition of U-G, U-A, and C-G pairs by double-stranded RNA-binding PNAs incorporated with an artificial nucleobase. *Biochemistry* **2019**, *In press*, doi: 10.1021/acs.biochem.1028b01313.
- (21) Li, M.; Zengeya, T.; Rozners, E. Short peptide nucleic acids bind strongly to homopurine tract of double helical RNA at pH 5.5. *J Am Chem Soc* **2010**, *132*, 8676-8681.
- (22) Gupta, P.; Zengeya, T.; Rozners, E. Triple helical recognition of pyrimidine inversions in polypurine tracts of RNA by nucleobase-modified PNA. *Chem Commun* **2011**, *47*, 11125-11127.
- (23) Zengeya, T.; Gupta, P.; Rozners, E. Triple-helical recognition of RNA using 2-aminopyridine-modified PNA at physiologically relevant conditions. *Angew Chem Int Ed Engl* **2012**, *51*, 12593-12596.
- (24) Zengeya, T.; Gupta, P.; Rozners, E. Sequence selective recognition of double-stranded RNA using triple helix-forming peptide nucleic acids. *Methods Mol Biol* **2014**, *1050*, 83-94.
- (25) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015**, *1-2*, 19-25.
- (26) Eldrup, A. B.; Dahl, O.; Nielsen, P. E. A novel peptide nucleic acid monomer for recognition of thymine in triple-helix structures. *J Am Chem Soc* **1997**, *119*, 11116-11117.
- (27) Grover, A.; Houlden, H.; Baker, M.; Adamson, J.; Lewis, J.; Prihar, G.; Pickering-Brown, S.; Duff, K.; Hutton, M. 5' splice site mutations in tau associated with the inherited dementia FTDP-17 affect a stem-loop structure that regulates alternative splicing of exon 10. *J Biol Chem* **1999**, *274*, 15134-15143.
- (28) Varani, L.; Hasegawa, M.; Spillantini, M. G.; Smith, M. J.; Murrell, J. R.; Ghetti, B.; Klug, A.; Goedert, M.; Varani, G. Structure of tau exon 10 splicing regulatory element RNA and destabilization by mutations of frontotemporal dementia and parkinsonism linked to chromosome 17. *Proc Natl Acad Sci U S A* **1999**, *96*, 8229-8234.
- (29) Donahue, C. P.; Muratore, C.; Wu, J. Y.; Kosik, K. S.; Wolfe, M. S. Stabilization of the tau exon 10 stem loop alters pre-mRNA splicing. *J Biol Chem* **2006**, *281*, 23302-23306.
- (30) Tan, J., Yang, L., Ong, A.A.L., Shi, J., Zhong, Z., Lye, M.L., Liu, S., Lisowiec-Wachnicka, J., Kierzek, R., Roca, X., Chen, G. . A disease-causing intronic point mutation C19G alters tau exon 10 splicing via RNA secondary structure rearrangement. *Biochemistry* **2019**, *In press*, doi: 10.1021/acs.biochem.1029b00001.
- (31) Hutton, M.; Lendon, C. L.; Rizzu, P.; Baker, M.; Froelich, S.; Houlden, H.; Pickering-Brown, S.; Chakraverty, S.; Isaacs, A.; Grover, A.; Hackett, J.; Adamson, J.; Lincoln, S.; Dickson, D.; Davies, P.; Petersen, R. C.; Stevens, M.; de Graaff, E.; Wauters, E.; van Baren, J.; Hillebrand, M.; Joosse, M.; Kwon, J. M.; Nowotny, P.; Che, L. K.; Norton, J.; Morris, J. C.; Reed, L. A.; Trojanowski, J.; Basun, H.; Lannfelt, L.; Neystat, M.; Fahn, S.; Dark, F.; Tannenberg, T.; Dodd, P. R.; Hayward, N.; Kwok, J. B.; Schofield, P. R.; Andreadis, A.; Snowden, J.; Craufurd, D.; Neary, D.; Owen, F.; Oostra, B. A.; Hardy, J.; Goate, A.; van Swieten, J.; Mann, D.; Lynch, T.; Heutink, P. Association of missense and 5'-splice-site mutations in tau with the inherited dementia FTDP-17. *Nature* **1998**, *393*, 702-705.

- (32) Schoch, K. M.; DeVos, S. L.; Miller, R. L.; Chun, S. J.; Norrbom, M.; Wozniak, D. F.; Dawson, H. N.; Bennett, C. F.; Rigo, F.; Miller, T. M. Increased 4R-Tau Induces Pathological Changes in a Human-Tau Mouse Model. *Neuron* **2016**, *90*, 941-947.
- (33) McCarthy, A.; Lonergan, R.; Olszewska, D. A.; O'Dowd, S.; Cummins, G.; Magennis, B.; Fallon, E. M.; Pender, N.; Huey, E. D.; Cosentino, S.; O'Rourke, K.; Kelly, B. D.; O'Connell, M.; Delon, I.; Farrell, M.; Spillantini, M. G.; Rowland, L. P.; Fahn, S.; Craig, P.; Hutton, M.; Lynch, T. Closing the tau loop: the missing tau mutation. *Brain* **2015**, *138*, 3100-3109.
- (34) Absalon, S.; Kochanek, D. M.; Raghavan, V.; Krichevsky, A. M. MiR-26b, upregulated in Alzheimer's disease, activates cell cycle entry, tau-phosphorylation, and apoptosis in postmitotic neurons. *J Neurosci* **2013**, *33*, 14645-14659.
- (35) Millan, M. J. Linking deregulation of non-coding RNA to the core pathophysiology of Alzheimer's disease: An integrative review. *Prog Neurobiol* **2017**, *156*, 1-68.
- (36) Kozomara, A.; Griffiths-Jones, S. miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res* **2014**, *42*, D68-73.

Chapter 5

Conclusion and Future Work

RNA secondary structure motifs are involved in gene regulation and expression. Mutations and dysregulation of the expression of RNAs cause dysregulation and lead to diseases¹. Hence, disease-associated RNA secondary structure becomes a potential therapeutic target^{2,3}. RNA duplexes can be recognized by ASOs through Watson-Crick base pairing. In addition, RNA duplexes or dsRNAs can be targeted TFOs through Hoogsteen base pairing⁴. PNA is a polyamide analog of naturally occurring DNA and RNA which possesses several advantages over the conventional oligonucleotides⁵. RNA duplexes or dsRNAs can be recognized by asPNAs through Watson-Crick base pairing to form a stable PNA-RNA duplex⁶⁻⁸. Besides, the recognition of dsRNAs can be achieved by dbPNAs through Hoogsteen base pairing to form a PNA·RNA-RNA triplex⁹⁻¹³. However, the development of dsRNA-binding PNAs (dbPNAs) as a class of dsRNA-targeting drug is hindered by complicated RNA secondary structures and canonical or non-canonical base pairings. Hence, modifications on nucleobases and the development of novel binding mode may improve the recognition of RNA secondary structures and disease-related RNA structures.

In summary, in this thesis, we utilized PNAs incorporating modified bases and base-backbone linkers and performed binding studies using non-denaturing PAGE studies, thermal melting experiments, fluorescence titration experiments and in vitro dicing assays to study the recognition of RNA secondary structures.

Our studies in this thesis revealed that PNAs incorporating the artificial S base with the optimized linker is critical for the recognition of C-G, U-A and U-G pairs in dsRNA regions at physiological conditions. Short 8-mer PNAs incorporating S base show no appreciable recognition of dsDNAs and ssRNAs. We demonstrated the recognition of miRNA-26b hairpin precursor by short 8-mer PNAs incorporating S monomer. Hence, PNA S monomer is advantageous for the recognition of a dsRNA region containing a Wobble U-G base pair probably through the formation of a unique S·U-G triple.

We also demonstrated that PNAs incorporating s^2U can be combined with L to recognize dsRNAs containing an A-U and G-C pairs, respectively, with minimal binding to dsDNAs and ssRNAs¹⁰. The binding enhancement of s^2U incorporation is generally applicable with s^2U immediately adjacent to T, C, and L with no significant sequence context dependence. We demonstrated through a cell-free assay the application of the s^2U - and L-modified dbPNAs in the inhibition of the pre-microRNA-198 maturation in a substrate specific manner. Hence, PNAs with s^2U and other modified residues such as L may have broad applications in targeting various RNAs including pre-miRNAs as demonstrated in our studies in this thesis.

In addition to the incorporation of modified nucleobases, we also developed a novel and generally applicable binding mode by combining dbPNA and asPNA, which is termed daPNA, for the recognition of a dsRNA region and an immediately adjacent ssRNA region (dsRNA-ssRNA junction) through the simultaneous formation of an asPNA-RNA duplex and a major-groove dbPNA·RNA₂ triplex. We demonstrated the recognition of tau pre-mRNA splice site hairpin and miR-26b hairpin precursor by

daPNAs. Thus, the daPNAs would be a potentially useful platform for the probing and targeting of many biologically and medically important RNA secondary structures or dsRNAs.

In future, further modifications and optimizations can be done in order to further improve the recognition of RNA secondary structures or dsRNAs by utilizing PNA. We can further improve the binding properties of S base by further modifying the chemical structure so as to achieve better affinity and selectivity to the recognition of Wobble U-G base pair. We can demonstrate the binding properties of dbPNAs incorporating S base or modified S base by targeting dsRNAs in cells. Finally, we can achieve recognition of biologically-relevant dsRNAs by dbPNA with a 4-letter code: s^2U , L, Q, and S^{10,11}. Similarly, we can combine traditional asPNA (with a 4-letter code: T, C, A, and G) and dbPNA (with a 4-letter code: s^2U , L, Q, and S) scaffolds to achieve RNA secondary structure-specific targeting of dsRNA. Our work provides the foundation for developing dsRNAs-targeting drugs by utilizing dbPNAs and daPNAs containing modified residues which are capable of recognizing dsRNA selectively in a further broadened sequence context.

References

- (1) Cooper, T. A.; Wan, L.; Dreyfuss, G. RNA and disease. *Cell* **2009**, *136*, 777-793.
- (2) Connelly, C. M.; Moon, M. H.; Schneekloth, J. S., Jr. The Emerging Role of RNA as a Therapeutic Target for Small Molecules. *Cell Chem Biol* **2016**, *23*, 1077-1090.
- (3) Shen, X.; Corey, D. R. Chemistry, mechanism and clinical status of antisense oligonucleotides and duplex RNAs. *Nucleic Acids Res* **2018**, *46*, 1584-1600.
- (4) Zhou, Y.; Kierzek, E.; Loo, Z. P.; Antonio, M.; Yau, Y. H.; Chuah, Y. W.; Geifman-Shochat, S.; Kierzek, R.; Chen, G. Recognition of RNA duplexes by chemically modified triplex-forming oligonucleotides. *Nucleic Acids Res* **2013**, *41*, 6664-6673.
- (5) Nielsen, P. E.; Egholm, M.; Berg, R. H.; Buchardt, O. Sequence-selective recognition of DNA by strand displacement with a thymine-substituted polyamide. *Science* **1991**, *254*, 1497-1500.
- (6) Egholm, M.; Buchardt, O.; Christensen, L.; Behrens, C.; Freier, S. M.; Driver, D. A.; Berg, R. H.; Kim, S. K.; Norden, B.; Nielsen, P. E. PNA hybridizes to complementary oligonucleotides obeying the Watson-Crick hydrogen-bonding rules. *Nature* **1993**, *365*, 566-568.
- (7) Egholm, M.; Nielsen, P. E.; Buchardt, O.; Berg, R. H. Recognition of guanine and adenine in DNA by cytosine and thymine containing peptide nucleic acids (PNA). *Journal of the American Chemical Society* **1992**, *114*, 9677-9678.
- (8) Dueholm, K. L.; Egholm, M.; Behrens, C.; Christensen, L.; Hansen, H. F.; Vulpius, T.; Petersen, K. H.; Berg, R. H.; Nielsen, P. E.; Buchardt, O. Synthesis of Peptide Nucleic Acid Monomers Containing the Four Natural Nucleobases: Thymine, Cytosine, Adenine, and Guanine and Their Oligomerization. *The Journal of Organic Chemistry* **1994**, *59*, 5767-5773.
- (9) Li, M.; Zengeya, T.; Rozners, E. Short peptide nucleic acids bind strongly to homopurine tract of double helical RNA at pH 5.5. *J Am Chem Soc* **2010**, *132*, 8676-8681.
- (10) Devi, G.; Yuan, Z.; Lu, Y.; Zhao, Y.; Chen, G. Incorporation of thio-pseudoisocytosine into triplex-forming peptide nucleic acids for enhanced recognition of RNA duplexes. *Nucleic Acids Res* **2014**, *42*, 4008-4018.
- (11) Toh, D.-F.; Devi, G.; Patil, K. M.; Qu, Q.; Maraswami, M.; Xiao, Y.; Loh, T.-P.; Zhao, Y.; Chen, G. Incorporating a guanidine-modified cytosine base into triplex-forming PNAs for the recognition of a C-G pyrimidine-purine inversion site of an RNA duplex. *Nucleic Acids Res* **2016**, *44*, 9071-9082.
- (12) Zengeya, T.; Gupta, P.; Rozners, E. Sequence selective recognition of double-stranded RNA using triple helix-forming peptide nucleic acids. *Methods Mol Biol* **2014**, *1050*, 83-94.
- (13) Patil, K. M.; Toh, D. K.; Yuan, Z.; Meng, Z.; Shu, Z.; Zhang, H.; Ong, A. A. L.; Krishna, M. S.; Lu, L.; Lu, Y.; Chen, G. Incorporating uracil and 5-halouracils into short peptide nucleic acids for enhanced recognition of A-U pairs in dsRNAs. *Nucleic Acids Res* **2018**, *46*, 7506-7521.

Appendixes

Appendix 1. ^1H , ^{13}C NMR Spectra for the synthesized compounds

Appendix 2. MALDI-TOF characterization of PNA oligomers

Table A2.1. MALDI-TOF data for the synthesized PNAs studied in chapter 2.

PNA	Sequence	Molecular Formula	Calculated MW	Observed MW
P10 (S)	NH ₂ -Lys-TCTSTTTC-CONH ₂	C ₁₀₀ H ₁₃₂ N ₃₈ O ₃₁ S	2392.96	2393.29
P10-S ₁	NH ₂ -Lys-TCTS ₁ TTTC-CONH ₂	C ₉₉ H ₁₃₀ N ₃₈ O ₃₁ S	2378.95	2380.64
P11 (S)	NH ₂ -Lys-TLTSTTTL-CONH ₂	C ₁₀₀ H ₁₃₂ N ₃₈ O ₂₉ S ₃	2424.92	2427.64
P11-S ₁	NH ₂ -Lys-TLTS ₁ TTTL-CONH ₂	C ₉₉ H ₁₃₀ N ₃₈ O ₂₉ S ₃	2410.90	2411.85
P11-E	NH ₂ -Lys-TLTETTTL-CONH ₂	C ₉₂ H ₁₂₆ N ₃₈ O ₂₈ S ₂	2274.90	2276.29
P12	NH ₂ -Lys-TLTLSTTL-CONH ₂	C ₉₉ H ₁₃₁ N ₃₉ O ₂₇ S ₄	2425.90	2428.12
P13	NH ₂ -Lys-TCTTTTTC-CONH ₂	C ₉₂ H ₁₂₅ N ₃₇ O ₃₁	2243.93	2245.24
P13-xT	NH ₂ -Lys-TCT ^x TTTTTC-CONH ₂	C ₉₃ H ₁₂₇ N ₃₇ O ₃₁	2259.27	2259.72
P13-xxT	NH ₂ -Lys-TCT ^{xx} TTTTTC-CONH ₂	C ₉₄ H ₁₂₉ N ₃₇ O ₃₁	2273.29	2273.44
P14	NH ₂ -Lys-TLLTSTLL-CONH ₂	C ₉₈ H ₁₃₀ N ₄₀ O ₂₅ S ₅	2428.67	2428.97
P14-E	NH ₂ -Lys-TLLTETLL-CONH ₂	C ₉₀ H ₁₂₄ N ₄₀ O ₂₄ S ₄	2278.48	2278.22

Table A2.2. MALDI-TOF data for the synthesized PNAs studied in chapter 3.

PNA	Sequence	Molecular Formula	Calculated MW	Observed MW
C-s ² U-1	NH ₂ -Lys-s ² UCTCTTTC-CONH ₂	C ₉₀ H ₁₂₂ N ₃₈ O ₂₉ S	2230.90	2231.90
L-s ² U-1	NH ₂ -Lys-s ² ULTLT ² TTL-CONH ₂	C ₉₀ H ₁₂₂ N ₃₈ O ₂₆ S ₄	2278.83	2280.29
C-s ² U-3	NH ₂ -Lys-TCs ² UCTTTC-CONH ₂	C ₉₀ H ₁₂₂ N ₃₈ O ₂₉ S	2230.90	2231.90
L-s ² U-3	NH ₂ -Lys-TLs ² ULTT ² TTL-CONH ₂	C ₉₀ H ₁₂₂ N ₃₈ O ₂₆ S ₄	2278.83	2280.75
C-s ² U-4	NH ₂ -Lys-TCTs ² UTTTC-CONH ₂	C ₉₁ H ₁₂₃ N ₃₇ O ₃₀ S	2245.90	2246.90
L-s ² U-4	NH ₂ -Lys-TLTs ² UTT ² TTL-CONH ₂	C ₉₁ H ₁₂₃ N ₃₇ O ₂₈ S ₃	2277.85	2279.16
C-s ² U-5	NH ₂ -Lys-TCTCs ² UTTC-CONH ₂	C ₉₀ H ₁₂₂ N ₃₈ O ₂₉ S	2230.90	2231.61
L-s ² U-5	NH ₂ -Lys-TLT ² LTs ² UT ² TTL-CONH ₂	C ₉₀ H ₁₂₂ N ₃₈ O ₂₆ S ₄	2278.83	2280.14
C-s ² U-6	NH ₂ -Lys-TCTCTs ² UTC-CONH ₂	C ₉₀ H ₁₂₂ N ₃₈ O ₂₉ S	2230.90	2232.61
L-s ² U-6	NH ₂ -Lys-TLT ² LTs ² UTL-CONH ₂	C ₉₀ H ₁₂₂ N ₃₈ O ₂₆ S ₄	2278.83	2280.59
C-s ² U-7	NH ₂ -Lys-TCTCTTs ² UC-CONH ₂	C ₉₀ H ₁₂₂ N ₃₈ O ₂₉ S	2230.90	2231.92
L-s ² U-7	NH ₂ -Lys-TLT ² LT ² Ts ² UL-CONH ₂	C ₉₀ H ₁₂₂ N ₃₈ O ₂₆ S ₄	2278.83	2279.82
PNAs targeting miR-198 precursor				
dbPNA-198	NH ₂ -Lys-LLTLT ² TLL-CONH ₂	C ₈₉ H ₁₂₂ N ₄₀ O ₂₃ S ₅	2278.82	2280.22
dbPNA-198-s ² U	NH ₂ -Lys-LLT ² LTs ² UTLL-CONH ₂	C ₈₈ H ₁₂₀ N ₄₀ O ₂₂ S ₆	2280.78	2283.31
dbPNA-NC	NH ₂ -Lys-LLQTLEL-CONH ₂	C ₈₂ H ₁₂₀ N ₄₀ O ₁₉ S ₄	2096.85	2097.9

Table A2.3. MALDI-TOF data for the synthesized PNAs studied in chapter 4.

PNA	Sequence	Molecular formula	Calculated MW	Observed MW
asPNA-tau	NH ₂ -Lys-ACACTGCCGC-CONH ₂	C ₁₁₁ H ₁₄₆ N ₆₀ O ₃₀	2799.17	2801.31
daPNA-1	NH ₂ -Lys-TAAAAALTLT-CONH ₂	C ₁₁₄ H ₁₄₈ N ₆₀ O ₂₇ S ₂	2853.15	2854.25
daPNA-26b	NH ₂ -Lys-TGGTLLTSTLL-CONH ₂	C ₁₃₁ H ₁₇₀ N ₅₈ O ₃₅ S ₅	3275.19	3279.02
daPNA-NC1	NH ₂ -Lys-TCACACLTLT-CONH ₂	C ₁₁₁ H ₁₄₈ N ₅₄ O ₃₀ S ₂	2781.12	2783.56
daPNA-NC2	NH ₂ -Lys-TAAAAETET-CONH ₂	C ₁₁₆ H ₁₅₂ N ₆₀ O ₂₉	2849.23	2852.75
daPNA-tau	NH ₂ -Lys-TQTQTGCCG-CONH ₂	C ₁₀₉ H ₁₅₇ N ₅₅ O ₃₁	2732.24	2733.34
daPNA-tau-NC	NH ₂ -Lys-TETETGCCG-CONH ₂	C ₁₀₃ H ₁₃₉ N ₄₉ O ₃₁	2558.08	2559.10
dbPNA-NC	NH ₂ -Lys-TQTQ-CONH ₂	C ₅₆ H ₉₁ N ₂₇ O ₁₅	1381.72	1382.90