# Synthesis and Properties of Amino- and Guanidino-Peptoid Peptide Nucleic Acids

**Lu Xiao Wei**

**Supervisor: Professor Liu Chuan-Fa**

School of Biological Sciences

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

**2010**

# Acknowledgements

Upon completion of this thesis, there are innumerable people who have affected, inspired and supported me along the way and they deserve gratitude.

First of all, I would like to express my greatest appreciation to my supervisor A/Prof. Liu Chuan-Fa for his insightful ideas, instructive guidance and encouragement. Obviously I would not have been able to accomplish this degree without his support. His profound knowledge and rich experience in chemistry and biology, his rigorous attitude towards science as well as his kindness and patience will have a great impact on my entire research career and life. It is just like that I have conducted my PhD project at home with my family during the past 5 years.

I would deeply express my sincere thanks to all the previous and present members of our lab, Dr Zeng Yun, Dr Yang Yi, Dr S. Thennarasu, Dr Zhang Xiaohong, Dr Tan Xiaohong, Ms Qin Wei, Mr. Andre Wirjo, Ms Hou Wen, Mr. Yang Renliang, Mr. Li Fupeng and Ms Ding Yingjie for their kindness and friendship. A special appreciation goes to Dr Zeng Yun, who has given me fundamental technical supports in completing this thesis.

I am very grateful to Ms Wong Minghwa from Assi/Prof. Thanabalu Thirumaran's lab for her help in performing the cellular uptake assay. I would like to thank Mr. Kalyan Kumar Pasunooti from Assi/Prof. Liu Xuewei's lab in SPMS for his help in HRMS analysis data. Thanks are also extended to my friends in SBS,

especially Ms Miao Huang, who has helped me a lot during my study.

I would like to acknowledge members of General Office in the School of Biological Sciences, especially Ms Chong Chye Hong May, Ms Chong Pui Ling and Ms Tan Soh Tin Eileen for their assistance during my graduate study. I would also thank the School of Biological Sciences, Nanyang Technological University for giving me this opportunity and much assistance to pursue my PhD study.

I am deeply and forever indebted to my parents, my sister and brother for their endless love and encouragement during my study. Finally, I would like to express my gratitude to my beloved husband – Mr. Yan Chaoyi. You have given me so much of yourself and have enhanced far beyond what I can express in words. I will be forever grateful for your strength, understanding and love.

# Table of Contents

# List of Illustrations

**Chapter 1**



**Chapter 2**

**Chapter 3**

# List of Tables

## Chapter 1



## Chapter 2



## Chapter 3

# Abstract

Peptide nucleic acid (PNA) is a synthetic mimic of DNA and RNA that can bind to complementary DNA and RNA with high affinity and sequence specificity. The structural simplicity, high hybridization affinity and *in vivo* stability of PNA have made it an attractive agent for antigene and antisense applications in basic biology and medicine. However, the inherent limitations of PNA, such as low solubility and poor cell permeability, have greatly limited its use. Extensive modification work has been conducted on PNA in an effort to further improve its desirable properties and overcome its limitations.

This thesis presents the results of a new type of PNA analogs with a peptoid-like side chain appending from the γ-N of the PNA backbone. The side chain of varying lengths is further functionalized with a positively charged amino or guanidino group. An interesting relationship between the hybridization affinity of the resultant amino-peptoid PNA (AP-PNA) or guanidino-peptoid PNA (GP-PNA) and the length of the side chain is observed. This makes it possible to modulate the hybridization and pharmacological properties of PNA in a more predictable way.

Chapter 1 of this thesis contains an introduction to peptide nucleic acid, their chemo-physical and biological properties as well as their applications. Information on PNA analogs with a chiral backbone is also provided.

Chapter 2 pertains to the synthesis of AP-PNA monomers, AP-PNA oligomers

and GP-PNA oligomers. The AP-PNA monomers of different side chain length (2, 3, 4, 5 or 6 carbon) were synthesized with the exocyclic amines of their four natural nucleobases protected by an acyl or Bhoc protecting group. The AP-PNA oligomers were synthesized on Rink amide PEGA resin using standard solid-phase Fmoc peptide synthesis protocols. The GP-PNA oligomers were achieved by converting the peptoidic amino group on AP-PNA to the corresponding guanidino-peptoid via a guanylation reaction. Other AP-PNA derivatives were obtained by acylating the amino group with interesting functional groups.

Chapter 3 demonstrates the hybridization and cellular uptake properties of AP-PNA, GP-PNA and the side-chain amine-acylated AP-PNA derivatives. A systematic structure–activity relationship study reveals that the length of side chain plays an important role in the hybridization affinity of the peptoid PNAs. The hybridization affinity increases with increasing length of the peptoid side chain with a cap reached at the 5 or 6-carbon length. The cellular uptake property of GP-PNAs is better than that of AP-PNAs and normal PNAs but not as efficient as TAT peptide.

In the fourth and final chapter, I briefly summarize the dissertation and discuss future work for the research in this area.

# Chapter 1 Introduction

## 1.1 Introduction to peptide nucleic acid

Recent advances in genomics have revealed tens of thousands of human and microbial genes. Modulating gene expression with specific molecules holds the key to mechanistic investigation on the functions of these genes and to the identification and validation of potential drug targets. A simple but attractive strategy for gene expression modulation involves the use of the antigene or antisense approach.[1-3] It concerns the use of an oligonucleotide or oligonucleotide analog for the targeting of a complementary nucleic acid sequence on the basis of Watson-Crick base pairing rules. Using this approach, a gene's expression can be modulated on different levels. One can target its genome DNA to modulate replication or transcription - an antigene approach, or mRNA to block translation - an antisense approach. Antigene oligonucleotides bind to DNA with Hoogsteen bonds to form a triplex helix and prevent gene expression by interfering with the binding of transcription factors or physically blocking RNA synthesis. Antisense oligonuclotides bind to complementary RNA and prevent gene expression by invoking RNase H-mediated degradation of the targeted RNA or physically blocking mRNA translation.

Applying this seemingly straightforward approach is not without obstacles.

Since the first demonstration of antisense activity in cultured cells in 1971, only one antisense oligonucleotide, Vitravene$^{R}$, has achieved FDA approval for treating cytomegalovirus retinitis. One obstacle is that natural oligonucleotides are rapidly degraded by nucleases and hence are not suitable as antigene or antisense agents *in vivo*. A number of modifications on the deoxyribose phosphate backbone of DNA have been tried to increase nuclease stability and cell membrane permeability. Consequently, derivatives such as mono- or dithiophosphates, methyl phosphonates and borano phosphates, as well as formacetal, carbamate, siloxane or dimethylenethio-, sulfoxido- and sulfono-linked species were prepared. The use of these agents, which are based on modified phosphodiester oligos, have been met with limited success.[4] Another big obstacle is ineffective hybridization due to the presence of secondary and tertiary structures in a DNA or RNA target.[5-8] For genomic DNA, the sequence being targeted is already engaged in hybridization with its complementary strand in the double helix. For mRNA, it is well known that the single stranded RNA is prone to fold into a rather stable 3-D structure. In both cases, the introduced oligo probe must invade the existing structure and seize the opportunity to hybridize with the transiently exposed target sequence during natural nucleic acid "breathing" events. Thus, hybridization of antigene or antisense oligos with structured DNA or RNA is unfavorable both thermodynamically and kinetically, relative to an unstructured target.[8] An attractive strategy to overcome this problem is to augment the displacing power, i.e., the affinity, of the hybridizing oligomer, so that the energetic barrier imposed

by a folded nucleic acid structure can be compensated by the increased stability of the newly formed duplex. Peptide nucleic acid or PNA is an example of such a high affinity oligomer.[8, 9]

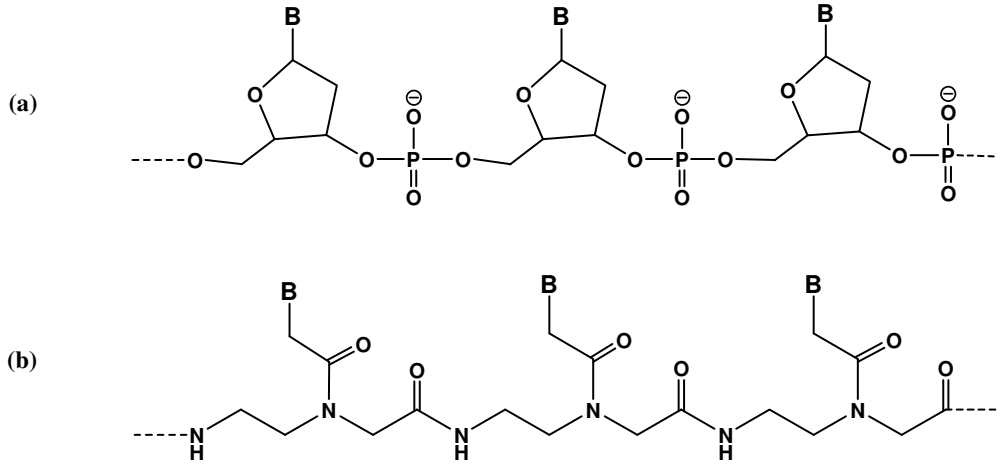



**Figure 1-1.** Comparison of structures of a single-stranded DNA (a) and PNA (b). B stands for nucleobase.

Peptide Nucleic Acid (PNA) is a DNA/RNA mimic in which the nucleobases are attached via methylenecarbonyl linkers to a poly-*N*-(2-aminoethyl)glycine (aeg) backbone (Figure 1-1).[9] PNA binds to complementary DNA/RNA through Watson-Crick base pairing with high affinity and sequence specificity as compared to most synthetic DNA mimics.[10] The high binding affinity of PNA to DNA/RNA owns partly to the lack of negative charge on its backbone and presumably to the proper interbase distances, the high flexibility of the aminoethyl linkers, and eventually intramolecular hydrogen bonding.[11] The simplicity of the achiral PNA structure makes it easily synthetically accessible by well-established solid phase peptide synthesis protocols and subsequent modifications with reporter groups or peptides with relative ease. Its resistance to both nucleases and peptidases confers

high *in vivo* stability. The inert amide linkages that join the different components together also give it excellent chemical stability and long shelf-life. They show few interactions with proteins or other biomolecules. These remarkable properties have stimulated considerable interest in using PNA for antisense and antigene applications in basic biology and medicine.

## 1.2 Chemical and physical properties of PNA

### 1.2.1 Chemical stability

PNA is completely acid stable even under HF conditions. PNA is also stable in the presence of weak bases, such as the ammonia and piperidine. Therefore, PNA can be assembled either by standard Boc or Fmoc solid phase peptide synthesis protocols. The only chemical instability of PNA is the free amino group at the N-terminus where a slow N-acyl transfer[12] of the nucleobase can occur under alkaline conditions. This type of side reaction can be eliminated by capping the final N-terminal amino group.

### 1.2.2 Solubility

Pure PNAs are neutral compounds with limited water solubility and a tendency for self-aggregation. The poor water solubility can be improved by introduction of charged groups, such as a C-terminal lysine amide. For example,

H-$T_{10}$-Lys-$NH_2$ is water soluble at concentrations in excess of 1.5 mM.[13] Alternatively, the introduction of negative charges, such as those in PNA/DNA chimeras, also enhances water solubility.

## 1.2.3 Binding affinity

PNA oligomers bind superbly well and with high sequence discrimination to complementary DNA, RNA or another PNA. The stability of PNA-nucleic acid complexes has been inferred by the thermal melting temperature (Tm) which is the temperature that allows dissociation of 50% of the complex, and is usually obtained from hyperchromicity measurements at 260 nm. It has been shown to correlate well with their thermodynamic dissociation constants, $K_d$. The Tm for identical sequences follows the order: PNA/PNA > PNA/RNA > PNA/DNA > RNA/DNA > DNA/DNA.[10, 14]

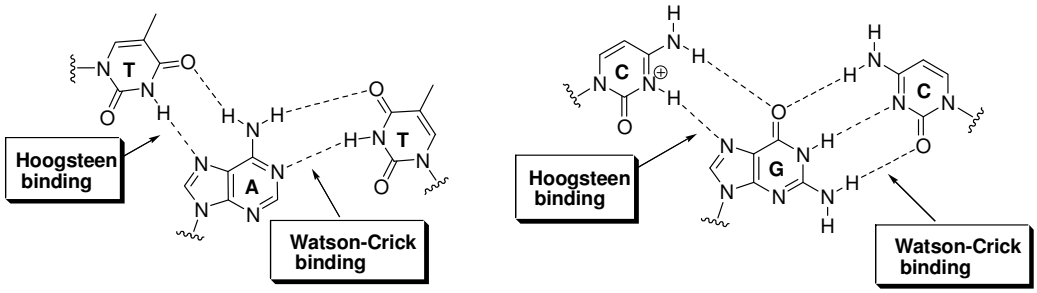



**Figure 1-2. Watson-Crick and Hoogsteen base pairs.**

### 1.2.3.1 Triplex helix formation

The first PNA oligomers were reported in 1991 by the Nielsen group.[9]

Oligomers containing different numbers of thymine residues were prepared because the thymine nucleobase has no exocyclic amino group and is thus easy to prepare.[13] The Tm values of PNA/DNA complexes are much higher than those of the corresponding DNA/DNA complex and the Tm values are dependent on the number of T units showing a ΔTm of ~10 $^{o}$C per unit. Furthermore, the stoichiometry determined by UV-titration curves is 2:1, indicating $(PNA)_2$/DNA triplexes. The Tm value of the complexes containing one cytosine monomer in the middle of poly$T_{10}$ is pH dependent (the triplex is the most stable at pH 5), and the stoichiomerty is 2:1 indicating that Hoogsteen base paring is involved in the complex.[15] The extremely high thermal stability of $(PNA)_2$/DNA triplexes may partially be due to the neutral backbone. However, there is always a very strong hysteresis (up to 30 $^{o}$C )[16, 17] in the UV-melting curves of the $(PNA)_2$/DNA triplex helix, indicating that the triple helix formation is rather slow.[18] The hysteresis can be largely suppressed with the so-called bis-PNAs, where the Watson-Crick strand (binds antiparallel to DNA) and the Hoogsteen strand (binds parallel to DNA) are joined together by a proper linker.

### 1.2.3.2 Duplex helix formation with complementary DNA and RNA

A pentadecamer PNA sequence containing all four natural nucleobases reported in 1993 hybridized to the complementary oligonucleotides obeying the Watson-Crick rules.[10] In contrast to DNA or RNA, the pentadecamer PNA binds

DNA and RNA at both antiparallel orientation and parallel orientation, whereby the PNA C-terminus corresponds to the 3' end and the N-terminus to the 5' end of the normal oligonucleotide as a convention. The antiparallel orientation of the two strands is more stable than the parallel orientation. The difference in Tm for the two orientations is more than 15 $^{o}$C for a PNA/DNA duplex and more than 20 $^{o}$C for PNA/RNA duplex. The antiparallel PNA/DNA and PNA/RNA duplexes are more stable than the corresponding DNA/DNA and DNA/RNA duplexes. The thermal stability of PNA/RNA duplexes is consistently higher than that of PNA/DNA duplexes.[19] The parallel PNA/DNA and PNA/RNA duplexes have almost the same stabilities as the corresponding DNA/DNA and DNA/RNA duplexes. The kinetic study of PNA complex formation using capillary gel electrophoresis showed that the antiparallel complex formed much faster than the parallel complex.[20]

In contrast to the hybrids between two anionic oligos like DNA/DNA, or DNA/RNA, the stability of PNA-nucleic acid complexes is independent of ionic strength because of the neutral PNA backbone.[21] The Tm of DNA/DNA duplexes increases considerably with increasing salt concentration, while the Tm of PNA/DNA duplexes decreases. The uncharged PNA/PNA duplex shows no significant destabilization with increasing salt concentration.

The sequence discrimination is generally more efficient for PNA recognizing DNA than for DNA recognizing DNA. One base pair mismatch results in a reduction of the Tm value of 8-20 $^{o}$C,[10] about double that observed for a

DNA/DNA duplex.

### 1.2.3.3 Duplex helix formation with complementary PNA

Bengt Norden et al[14] found that two complementary PNA strands can hybridize to one another to form a Watson-Crick base-paired helical duplex. For a PNA decamer H-GTAGATCACT-Lys-$NH_2$ and the complementary antiparallel sequence H-AGTGATCTAC-Lys-$NH_2$, the thermal stability of the PNA-PNA duplex (melting temperature) is 67 $^{o}$C, which is considerably higher than that of the antiparallel PNA/DNA duplex (Tm = 51 $^{o}$C) or the DNA/DNA duplex (Tm = 33.5 $^{o}$C). Analogously to what was observed in the PNA/DNA duplex, the thermal stability of the antiparallel PNA duplex is much more stable (around 20 $^{o}$C) than that of the corresponding parallel PNA duplex. The kinetics investigated by CD spectrum show that the PNA/PNA duplex formation is relatively slower than that of the DNA/DNA duplex. This may be due to the inversion of the duplex to the preferred chirality. A PNA-PNA-PNA triplex was also observed with Watson-Crick-Hoogsteen base paring.[22]

## 1.2.4 PNA structure

Several high resolution structures of PNA complexes are available at present. The detailed structural information of the PNA/DNA duplex,[23, 24] the PNA/RNA duplex,[25] the $PNA_2$/DNA triplex[26, 27] and the PNA/PNA duplex[28] have been

obtained by CD spectroscopy, NMR and X-ray crystallography.

### 1.2.4.1 PNA/DNA duplexes

The solution structures of two antiparallel PNA/DNA duplexes, the octamer sequence H-(GCTATGTC)-$NH_2$ and decamer sequence H-(GTAGATCACT)-$NH_2$ with their corresponding antiparallel DNA strands were studied by NMR spectra.[23] PNA/DNA hybrids are base paired by Watson-Crick hydrogen bonds and the DNA strand in the PNA/DNA complex adopts a B form with the deoxyribose sugars in the C2'-*endo* conformation. The solution structure of the octamer PNA/DNA duplex was studied more in detail.[24] The structure is a right-handed double helix. The pitch is 42 Å, with close to 13 base pairs per turn. The helical diameter is approximately 23 Å and the average helical twist is 28º. The major groove is wide and deep and the minor groove is shallow and narrow. This makes the base pair displaced towards the minor groove, allowing the major groove to extend to near the center of the helix. The primary amide bonds are always in the *trans* conformation (or *Z* configuration) and the ethylene regions exhibit substantial heterogeneity. The carbonyl groups that link the nucleobases to the PNA backbone-nucleobase are oriented along the backbone, pointing in the C-terminal direction. The sugar rings of the DNA strand are in a B-like form with *anti* glycosidic torsion angles and are predominantly in the $C_2$'-*endo* conformation. However, the DNA backbone torsion angles ($\alpha$, $\beta$, $\gamma$, $\varepsilon$ and $\xi$ ) are close to those in the A-form. Thus the PNA/DNA duplex exhibits base pair stacking similar to the

A-form and the DNA sugar is more like the B-form.

### 1.2.4.2 PNA/RNA duplex

The solution structures of antiparallel PNA/RNA duplex, GAACTC-$NH_2$·r(GAGUUC) was studied by NMR spectra.[25] The PNA/RNA duplex is a right-handed helix with Watson-Crick base paring. Similar to the PNA/DNA duplex, the carbonyl groups on the PNA backbone-nucleobase linkers are pointing towards the C-terminus. The primary amide bonds are in the *trans* conformation and the torsion angle in the ethylenediamine region of the backbone is predominantly *gauche* (NCCN ≈ +60º). The RNA strand is similar to the A-form, with $C_3$'-*endo* sugar puckers and glycosidic torsion angles near -160º.

### 1.2.4.3 $PNA_2$/DNA triplex

The crystal structure of $PNA_2$/DNA triplex was obtained by X-ray crystallography.[27] The triplex containing one polypurine DNA strand d(GAAGAAGAG) hybridizes to a bisPNA CTCTTCTTC-peptide-CTTCTTCTC with Watson-Crick and Hoogsteen base paring. Thus the regular T·A-T and $C^+$·G-C triplexes are formed. The triplex structure revealed a helix different from either the A-form or the B-form DNA and forms a "P-form" with a very large pitch (18 bp). The P-form helix has a large cavity along the helical axis, and is wide (the diameter is ~26 Å). There are hydrogen bonds between the DNA phosphate oxygens and the Hoogsteen PNA amide, and extensive van der Waals forces

between the two strands. The DNA stand is similar to the A-form with $C_3$'-*endo* sugar conformation with an average interphosphate distance of 6.0 Å. There are 35 clearly defined water molecules mostly binding in the minor groove, specifically to the amide hydrogen of the Watson-Crick PNA backbone. These water molecules, as well as the hydrogen bonds and the van der Waals forces, are the main factors responsible for the enormous stability of the triplex helix. The CD spectra of a $PNA_2$/DNA triplex formed between poly(dA) and PNA-$T_8$ is similar to the crystal structure.[26]

### 1.2.4.4 PNA/PNA duplex

The crystal structure of a self-complementary PNA/PNA duplex (H-CGTACG-$NH_2$) was obtained by X-ray crystallography.[28] The duplex exists as both right- and left- handed helices. All bases in the duplex are engaged in Watson-Crick hydrogen bonding and the base pairs are perpendicular to the helical axis. The helices are wide (28 Å) with a very large pitch of ~18 base pairs (compared to the 11 and 10 base pairs per turn in A-DNA and B-DNA). The two PNA strands are close together to produce a very wide and deep major groove and a very narrow and shallow minor groove. The amide groups of the backbone are in *trans* configuration and the carbonyl oxygens of the backbone-nucleobase linker point towards the carboxyl terminal. Thus the structure has a strong similarity to the P-form of the $PNA_2$/DNA triplexes described above.

#### 1.2.4.5 Comparison of the hybrid structures

There are several common features in the hybrids described above. [29, 30] Firstly, the carbonyl groups of the backbone-nucleobase linker all point towards the C-terminus. Secondly, the Watson-Crick paired bases stack in a manner similar to an A-form RNA helix. Thirdly, the secondary amide bonds are in the *trans* conformation. Lastly, the base pairs are perpendicular to the helical axis, similar to B-DNA helices with only minor variations in slide, tilt and propeller twist angles between individual base pairs.

## 1.3 Biological properties of PNA

### 1.3.1 Stability in biological systems

The use of oligonucleotides or derivatives as antisense or antigene therapeutics must show high biostability in serum and in cells. In contrast to the unmodified oligonucleotides which are rapidly digested by nucleases in serum, PNAs, due to their non peptide and non nucleic acid structure, have a remarkably high biostability in both human serum and cell extracts.[31] For example, H-$T_{10}$-Lys-$NH_2$ cultured in human blood serum, *Eschericia coli*, and *Micrococcus luteus* extracts, as well as nuclear and cytoplasmic extracts from mouse Ehrlich ascite tumor cells shows no significant degradation by HPLC, while the oligonucleotide controls were degraded rapidly under the same conditions. PNA/DNA chimeras are much more stable (25-50 times) in serum than their

corresponding unmodified oligonucleotides.[32, 33]

## 1.3.2 Cellular uptake of PNAs

The cellular uptake of PNAs is extremely low.[34-36] The poor cellular uptake property of PNA has greatly hampered its use as an antigene or antisense agent. Many efforts have been devoted to enhance PNA cell permeability. The first used method was direct intracellular microinjection.[37-39] Although feasible, microinjection is limited to small-scale experimental set-ups, due to the intensive labor required.[40] Electroporation[41-43] was also used to transfer PNAs to cells. This method relies on the use of electrical pulses to transiently perturb the lipid bilayers of cell membranes, allowing PNA to diffuse freely into cells. Conjugating PNAs to DNA oligomers in the presence of cationic lipids,[44-46] receptor ligands,[47-49] lipophilic moieties[49-51] or cell permeable peptides (CPPs)[52-70] has been reported to improve PNA cellular uptake. Among these, the PNA-CPP conjugate shows the most promising uptake properties. CPPs are short amphiphilic cationic/hydrophobic peptides, which are reported to transport biomolecules (oligonucleotides, proteins, and peptides) and even large molecular weight material (liposomes, virus) across biological membranes. The natural CPPs are derived from proteins such as the Drosophila Antennapedia transcription factor[71] or the HIV-1 Tat transactivating protein.[72, 73] A series of these chimeric peptide analogs, like MTS-NLS, transportan and so on have been designed and studied in terms of

their cellular uptake mechanism and applications. Some of the CPPs are listed in Table 1-1.

**Table 1-1.** Examples of Cell Penetrating Peptides

| CPP | Sequence | Ref. |
|---|---|---|
| HIV-1 Tat | GRKKRRQRRRPPQ | [74] |
| Penetratin | RQIKIWFQNRRMKWKK | [71] |
| pAntp | RQIKIWFQNRRMKWKK | [75] |
| Retro-invserso penetratin | (D)-KKWKMRRNQFWVKVQR | [76] |
| Transportan | GWTLNSAGYLLGKINLKALAALAKKIL | [77] |
| NLS | PKKKKRKV | [78] |
| Cell membrane active peptide | KFFKFFKFFK | [79] |
| Polyarginine | RRRRRRRRRR | [80] |

## 1.4 Applications of PNA

The stable and sequence specific binding to the complementary DNA or RNA, as well as its high biological and chemical stability makes PNA a promising antisense and antigene drug candidate.

### 1.4.1 Antisense agents

PNA has obvious potential in antisense strategies as PNA hybridizes mRNA at high binding affinity and selectivity. However, unlike other antisense agents, PNA/RNA duplexes are not subjected to degradation by RNase-H. The antisense effect acts by steric interference of RNA processing, cytoplasmic transport, translation, caused by binding to the mRNA.[81] The first antisense PNA was demonstrated in 1992.[37] Nuclear microinjection of cells constitutively expressing SV40 large T antigen (TAg) with either a 15-mer or 20-mer PNA targeted to the

TAg mRNA suppressed TAg expression. This inhibition was both sequence-specific and dose-dependent. To test this potential of PNA in bacteria, Nielsen and Good[82] designed antisense PNAs targeting the start codon regions of the *Escherichia coli* β-galactosidase and β-lactamase genes. Dose-dependent and specific gene inhibition was observed in vitro using low nanomolar PNA concentrations and in vivo using low micromolar concentrations. Rabie et al[76] showed that PNA coupled to a retro-inverso peptide was rapidly internalized in cultured neurons and had enough antisense activity to depress the targeted mRNA. Nielsen and Good[79] showed that 9- to 12-mer PNAs, especially when attached to the cell membrane-active peptide KFFKFFKFFK, provided improvements in antisense potency in *E. coli* amounting to two orders of magnitude while retaining target specificity. Subsequent uptake kinetic studies[83] of the peptide-PNAs showed that membrane permeability was more rapid in the presence of peptide-PNA conjugates relative to the free components used alone or in combination. The enhanced cell-permeability properties of peptide-PNAs can explain their potent antisense activity, and such peptide–PNA conjugates open exciting possibilities for anti-infective drug development and provide new tools for microbial genetics.

## 1.4.2 Antigene agents

PNAs can inhibit gene transcription by triplex helix formation, strand invasion

or strand displacement.[9, 37] Strand invasion, in which PNA invades DNA to form stable PNA/DNA hybrids, has profound implications for both positive and negative control of gene activity. Several *in vitro* studies have shown that binding of PNA or bis-PNA to double-stranded DNA can effectively block transcriptional elongation and inhibit the binding of transcriptional factors to DNA helices.[17] Allfrey *et al*[84, 85] have demonstrated that PNA invasion of the tandem CAG repeat of the human androgen receptors and the TATA binding protein inhibits the transcription of these genes. Freier *et al*[86] demonstrated PNA mediated inhibition of NF-κB specific transcriptional activation. Andrea Pession *et al*[87] developed an anti-gene PNA which was conjugated with a nuclear localization signal peptide at the N-terminus and selectively inhibited *MYCN* transcription in neuroblastoma cells by targeting a unique sequence in the antisense DNA strand of exon 2 of *MYCN*. Corey *et al*[88, 89] demonstrated PNA as an antigene agent to inhibit transcription by targeting the transcriptional start sites for the human progesterone receptor B and A isomers at sequences predicted to be single-stranded within the open complex of chromosomal DNA.

Elliott Richelson *et al*[90] demonstrated an *in vivo* study of antigene PNA targeted to Angiotensinogen, the precursor protein that leads to Angotensin I and II. The study demonstrates on the molecular, protein, and physiological levels that antigene PNAs are effective *in vivo* upon systemic administration.

Although these results are encouraging, the binding of PNAs to double strand DNA is influenced by the presence of salts and the process of strand invasion is

very slow. Because the binding process requires replacing one of the complementary DNA strands in the DNA/DNA duplex, the presence of positively charged ions, especially at high concentrations stabilizes the DNA/DNA helix resulting in inhibition of the strand invasion.[91] However, incorporation of positively charged groups such as lysine or arginine into the PNA backbone[92-94] or conjugating a positively charged peptide to PNA[95] leads to an increase in PNA strand invasion in the presence of high concentrations of salts. This may be due to the presence of positive charges contained in PNA competing with other positive ions to interact with the negatively charged DNA targets. Furthermore, incorporating DNA intercalators (acridine),[96, 97] the G-clamp[98] or single-stranded DNA binding protein (SSB)[99] to single stranded PNAs or modified PNAs have also sufficiently promoted strand invasion. Komiyama *et al*[100] demonstrated the "double-duplex invasion" model where two complementary PNAs bearing positively charges (D-lys) at the α-C of the PNA backbone targeting the double stranded DNA resulted in a highly efficient strand invasion.

## 1.4.3 Anti-cancer agents

PNAs have also been used in anti-cancer applications. Corey *et al*[101] reported that PNAs targeting the RNA component of human telomerase (hTR) inhibited telomerase activity with high binding affinity and selectivity ($IC_{50}$ values in the picomolar to nanomolar range). The activity of PNAs as potential telomerase

inhibitors have been scoped by conjugating to cell penetrating peptides (CPPs).[44, 102-105] Since telomerase is almost ubiquitously expressed in human tumors, this data pointed out the potential use of PNAs as anticancer drugs.[106] PNAs with various CPP conjugations were also reported to inhibit various sequences of bcl-2[107-112]. PNAs targeting other cancer cells were also reported.[113-119]

## 1.4.4 PCR clamping

The PCR clamping[120] method involves using a PNA oligomer to inhibit the amplification of a specific target by direct competition with a PCR primer (because PNA recognize and bind to their complementary nucleic acid sequences with higher thermal stability and specificity than the corresponding deoxyribooligonucleotides and they cannot function as primers for DNA polymerases). This technique is so powerful that it can be used to detect single base gene variants for mutation screening and gene isolation.[121-130]

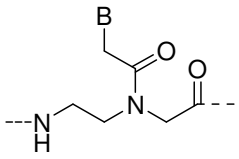
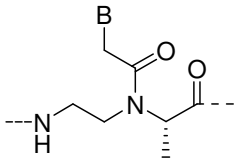
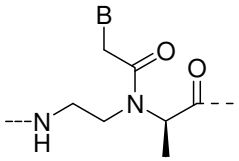
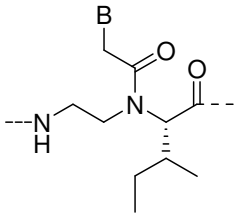
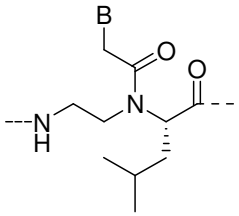
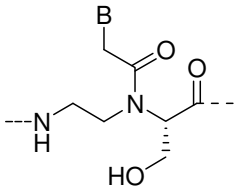
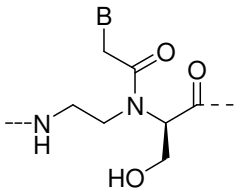
## 1.4.5 Other applications

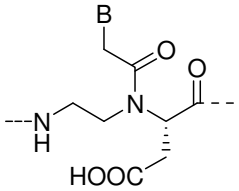
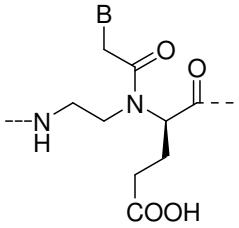
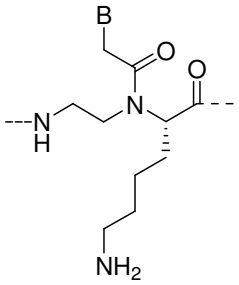
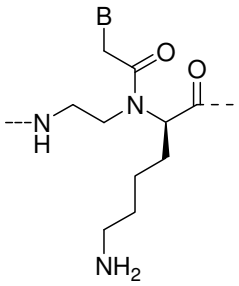
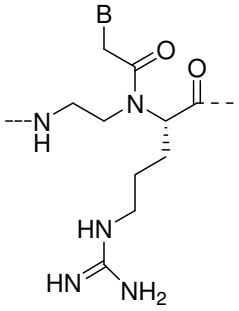
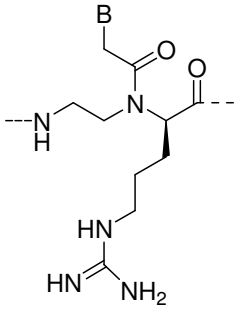
Due to their high binding affinity and specificity, PNA probes have proven extremely useful, especially in FISH (fluorescence *in situ* hybridization) assays,[131-134] solid phase hybridization techniques,[19, 135] and nucleic acid capture.[136]

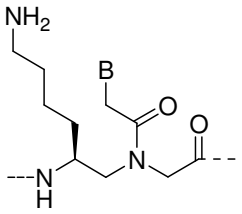
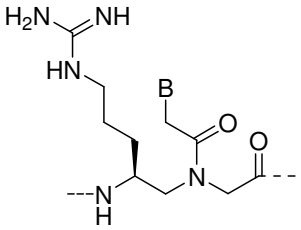
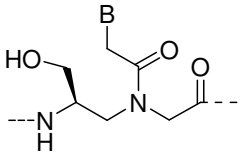
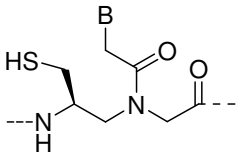
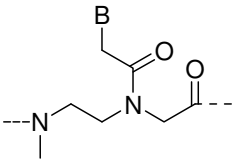
## 1.5 PNA backbone modifications

Since the first report on PNA,[9] a large number of PNA backbone derivatives have been described and investigated in order to further improve the properties of PNA. Such properties include binding affinity, solubility and cellular uptake. These PNA backbone modifications include introducing chiral centers with amino acids side chains on the α-C or γ-C of PNA's aeg backbone, or constraining the PNA backbone with five- or six-member cyclic structures.[137-139] The PNA derivatives based on different chiral amino acids introduced on the α-C or γ-C aegPNA backbone is listed in Table 1-2. The results show that the stability of modified PNA/DNA is decided by the nature of the side chain, the position and the configuration of monomers. As far as the configuration of the α-C chiral monomer is concerned, PNAs containing D-monomers bind to the complementary antiparallel DNA with higher binding affinity than that of the L-monomer.[140] For the configuration of the γ-C chiral monomers, the L-monomer binds to the antiparallel DNA better than the D-monomer. The observed enantioselectivity maybe due to the preferential PNA helicity induced by the configuration of the stereogenic centre.[141, 142] For example, the right-handed nature of the α-D-monomers binds the right-handed DNA better than the corresponding left-handed nature of α-L-monomers binding to the right-handed DNA. Moreover, introduction of negatively charged side chain, such as aspartic or glutamic acid, decreases hybridization potency, while a positively charged side chain, such as arginine or lysine has a beneficial effect on the stability of PNA/DNA duplex.

**Table 1-2.** Effect on thermal stability (ΔTm, °C) per modified unit for structural modified PNA as compared to unmodified PNA.

| Entry | Structure | Chirality | ΔTm DNA per unit | ΔTm RNA per unit | Ref |
|---|---|---|---|---|---|
| 1 |  | Gly | 0 | 0 | |
| 2 |  | α-L-Ala | -1.5 | ND | [143] |
| 3 |  | α-D-Ala | -0.3 | ND | [143] |
| 4 |  | α-L-Ile | -2.6 | -3.0 | [140] |
| 5 |  | α-L-Leu | -1.33 ~ -2 | ND | [141] |
| 6 |  | α-L-Ser | -1.0 | -1.0 | [140] |
| 7 |  | α-D-Ser | -0.6 | -1.0 | [140] |

| | | | | | |
|---|---|---|---|---|---|
| 8 |  | α-L-Asp | -3.3 | ND | [140, 141] |
| 9 |  | α-D-Glu | -2.3 | ND | [140, 141] |
| 10 |  | α-L-Lys | -2 ~ 0 | ND | [140, 141] |
| 11 |  | α-D-Lys | 1 ~ 2 | ND | [140, 141, 144] |
| 12 |  | α-L-Arg | -7.3 ~ -8 | ND | [145, 146] |
| 13 |  | α-D-Arg | -2.6 ~ 0 | ND | [145, 146] |

| | | | | | |
|---|---|---|---|---|---|
| 14 |  | γ-L-Lys | 1.6 | ND | [147] |
| 15 |  | γ-L-Arg | 3.2 | 2 | [148] |
| 16 |  | γ-L-Ser | 2 ~ 3 | ND | [149] |
| 17 |  | γ-L-Cys | -0.1 | ND | [150] |
| 18 |  | γN-methyl | -2.8 | -1.8 | [151] |

## 1.6 Thesis overview

This thesis describes a new type of PNA analogs in which the γ-N of the PNA backbone is modified with an alkyl side chain carrying a functional headgroup. Results on the synthesis of these PNA analogs containing such peptoid-like side chains of various lengths and characterization of their hybridization and uptake properties are presented in the following chapters.

Chapter 1 includes an introduction to peptide nucleic acid and their chemical, physical and biological properties, as well as their applications. Some information

on modifications based on introducing amino acids into the PNA backbone is also provided. Chapter 2 covers the synthesis of AP-PNA monomers, and the assembling of these monomers to AP-PNA oligomers as well as the conversion of the AP-PNA oligomers to the corresponding GP-PNA oligomers and AP-PNA derivatives. In chapter 3, we investigate the effects of side chain length and the head functional group on the hybridization potency of peptoid-PNA. The CD spectra of selected sequences are also studied as well as the cellular uptake property for the selected sequences.

## 1.7 References


[1] D. A. Braasch, D. R. Corey, *Biochemistry* **2002**, *41*, 4503.

[2] K. S. Lavery, T. H. King, *Current Opinion in Drug Discovery & Development* **2003**, *6*, 561.

[3] K. F. Pirollo, A. Rait, L. S. Sleer, E. H. Chang, *Pharmacology & Therapeutics* **2003**, *99*, 55.

[4] J. Kurreck, *European Journal of Biochemistry* **2003**, *270*, 1628.

[5] P. Nielsen, *Drug Discovery Today* **2003**, *8*, 440.

[6] V. V. Demidov, *Drug Discovery Today* **2003**, *8*, 390.

[7] S. A. Kushon, J. P. Jordan, J. L. Seifert, H. Nielsen, P. E. Nielsen, B. A. Armitage, *Journal of the American Chemical Society* **2001**, *123*, 10805.

[8] B. A. Armitage, *Drug Discovery Today* **2003**, *8*, 222.

[9] P. E. Nielsen, M. Egholm, R. H. Berg, O. Buchardt, *Science* **1991**, *254*, 1497.

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

[11] B. Hyrup, M. Egholm, P. E. Nielsen, P. Wittung, B. Norden, O. Buchardt, *Journal of the American Chemical Society* **1994**, *116*, 7964.

[12] M. Eriksson, L. Christensen, J. Schmidt, G. Haaima, L. Orgel, P. E. Nielsen, *New Journal Of Chemistry* **1998**, *22*, 1055.

[13] M. Egholm, O. Buchardt, P. E. Nielsen, R. H. Berg, *Journal of the American Chemical Society* **1992**, *114*, 1895.

[14] P. Wittung, P. E. Nielsen, O. Buchardt, M. Egholm, B. Norden, *Nature* **1994**, *368*, 561.

[15] M. Egholm, P. E. Nielsen, O. Buchardt, R. H. Berg, *Journal of the American Chemical Society* **1992**, *114*, 9677.

[16] M. Egholm, L. Christensen, K. L. Dueholm, O. Buchardt, J. Coull, P. E. Nielsen, *Nucleic Acids Research* **1995**, *23*, 217.

[17] M. C. Griffith, L. M. Risen, M. J. Greig, E. A. Lesnik, K. G. Sprankle, R. H. Griffey, J. S. Kiely, S. M. Freier, *Journal of the American Chemical Society* **1995**, *117*, 831.

[18] E. A. Lesnik, L. M. Risen, D. A. Driver, M. C. Griffith, K. Sprankle, S. M. Freier, *Nucleic Acids Research* **1997**, *25*, 568.

[19] K. K. Jensen, H. Orum, P. E. Nielsen, B. Norden, *Biochemistry* **1997**, *36*, 5072.

[20] D. J. Rose, *Analytical Chemistry* **1993**, *65*, 3545.

[21] S. Tomac, M. Sarkar, T. Ratilainen, P. Wittung, P. E. Nielsen, B. Norden, A. Graslund, *Journal of the American Chemical Society* **1996**, *118*, 5544.

[22] P. Wittung, P. Nielsen, B. Norden, *Journal of the American Chemical Society* **1997**, *119*, 3189.

[23] M. Leijon, A. Graslund, P. E. Nielsen, O. Buchardt, B. Norden, S. M. Kristensen, M. Eriksson, *Biochemistry* **1994**, *33*, 9820.

[24] M. Eriksson, P. E. Nielsen, *Nature Structural Biology* **1996**, *3*, 410.

[25] S. C. Brown, S. A. Thomson, J. M. Veal, D. G. Davis, *Science* **1994**, *265*, 777.

[26] S. K. Kim, P. E. Nielsen, M. Egholm, O. Buchardt, R. H. Berg, B. Norden, *Journal of the American Chemical Society* **1993**, *115*, 6477.

[27] L. Betts, J. A. Josey, J. M. Veal, S. R. Jordan, *Science* **1995**, *270*, 1838.

[28] H. Rasmussen, J. Sandholm, *Nature Structural Biology* **1997**, *4*, 98.

[29] M. Eriksson, P. E. Nielsen, *Quarterly Reviews of Biophysics* **1996**, *29*, 369.

[30] P. E. Nielsen, G. Haaima, *Chemical Society Reviews* **1997**, *26*, 73.

[31] V. V. Demidov, V. N. Potaman, M. D. Frankkamenetskii, M. Egholm, O. Buchard, S. H. Sonnichsen, P. E. Nielsen, *Biochemical Pharmacology* **1994**, *48*, 1310.

[32] A. Peyman, E. Uhlmann, K. Wagner, S. Augustin, G. Breipohl, D. W. Will, A. Schafer, H. Wallmeier, *Angewandte Chemie-International Edition in English* **1996**, *35*, 2636.

[33] E. Uhlmann, D. W. Will, G. Breipohl, D. Langner, A. Ryte, *Angewandte Chemie-International Edition in English* **1996**, *35*, 2632.

[34] O. Buchardt, M. Egholm, R. H. Berg, P. E. Nielsen, *Trends in Biotechnology* **1993**, *11*, 384.

[35] B. Hyrup, P. E. Nielsen, *Bioorganic & Medicinal Chemistry* **1996**, *4*, 5.

[36] P. Wittung, M. Eriksson, R. Lyng, P. E. Nielsen, B. Norden, *Journal of the American Chemical Society* **1995**, *117*, 10167.

[37] J. C. Hanvey, N. J. Peffer, J. E. Bisi, S. A. Thomson, R. Cadilla, J. A. Josey, D. J. Ricca, C. F. Hassman, M. A. Bonham, K. G. Au, S. G. Carter, D. A. Bruckenstein, A. L. Boyd, S. A. Noble, L. E. Babiss, *Science* **1992**, *258*, 1481.

[38] M. A. Bonham, S. Brown, A. L. Boyd, P. H. Brown, D. A. Bruckenstein, J. C. Hanvey, S. A. Thomson, A. Pipe, F. Hassman, J. E. Bisi, B. C. Froehler, M. D. Matteucci, R. W. Wagner, S. A. Noble, L. E. Babiss, *Nucleic Acids Research* **1995**, *23*, 1197.

[39] S. A. Noble, M. A. Bonham, J. E. Bisi, D. A. Bruckenstein, P. H. Brown, S. C. Brown, R. Cadilla, M. D. Gaul, J. C. Hanvey, C. F. Hassman, J. A. Josey, M. J. Luzzio, P. M. Myers, A. J. Pipe, D. J. Ricca, C. W. Su, C. L. Stevenson, S. A. Thomson, R. W. Wiethe, L. E. Babiss, *Drug Development Research* **1995**, *34*, 184.

[40] U. Koppelhus, P. E. Nielsen, *Advanced Drug Delivery Reviews* **2003**, *55*, 267.

[41] M. A. Shammas, C. G. Simmons, D. R. Corey, R. J. S. Reis, *Oncogene* **1999**, *18*, 6191.

[42] G. Wang, X. X. Xu, P. Pace, D. A. Dean, P. M. Glazer, P. Chan, S. R. Goodman, I. Shokolenko, *Nucleic Acids Research* **1999**, *27*, 2806.

[43] J. G. Karras, M. A. Maier, T. Lu, A. Watt, M. Manoharan, *Biochemistry* **2001**, *40*, 7853.

[44] S. E. Hamilton, C. G. Simmons, I. S. Kathiriya, D. R. Corey, *Chemistry & Biology* **1999**, *6*, 343.

[45] D. F. Doyle, D. A. Braasch, C. G. Simmons, B. A. Janowski, D. R. Corey, *Biochemistry* **2001**, *40*, 53.

[46] B. S. Herbert, A. E. Pitts, S. I. Baker, S. E. Hamilton, W. E. Wright, J. W. Shay, D. R. Corey, *Proceedings of the National Academy of Sciences of the United States of America* **1999**, *96*, 14276.

[47] S. Basu, E. Wickstrom, *Bioconjugate Chemistry* **1997**, *8*, 481.

[48] X. Zhang, C. G. Simmons, D. R. Corey, *Bioorganic & Medicinal Chemistry Letters* **2001**, *11*, 1269.

[49] L. C. Boffa, S. Scarfi, M. R. Mariani, G. Damonte, V. G. Allfrey, U. Benatti, P. L. Morris, *Cancer Research* **2000**, *60*, 2258.

[50] L. Mologni, E. Marchesi, P. E. Nielsen, C. Gambacorti-Passerini, *Cancer Research* **2001**, *61*, 5468.

[51] A. Muratovska, R. N. Lightowlers, R. W. Taylor, D. M. Turnbull, R. A. J. Smith, J. A. Wilce, S. W. Martin, M. P. Murphy, *Nucleic Acids Research* **2001**, *29*, 1852.

[52] M. Pooga, U. Soomets, M. Hallbrink, A. Valkna, K. Saar, K. Rezaei, U. Kahl, J. X. Hao, X. J. Xu, Z. Wiesenfeld-Hallin, T. Hokfelt, A. Bartfai, U. Langel, *Nature Biotechnology* **1998**, *16*, 857.

[53] L. Fisher, U. Soomets, V. C. Toro, L. Chilton, Y. Jiang, U. Langel, K. Iverfeldt, *Gene Therapy* **2004**, *11*, 1264.

[54] J. Oehlke, G. Wallukat, Y. Wolf, A. Ehrlich, B. Wiesner, H. Berger, M. Bienert, *European Journal of Biochemistry* **2004**, *271*, 3043.

[55] L. Petersen, M. C. de Koning, P. van Kuik-Romeijn, J. Weterings, C. J. Pol, G. Platenburg, M. Overhand, G. A. van der Marel, J. H. van Boom, *Bioconjugate Chemistry* **2004**, *15*, 576.

[56] J. J. Diaz-Mochon, L. Bialy, J. Watson, R. M. Sanchez-Martin, M. Bradley, *Chemical Communications* **2005**, 3316.

[57] J. J. Turner, G. D. Ivanova, B. Verbeure, D. Williams, A. A. Arzumanov, S. Abes, B. Lebleu, M. J. Gait, *Nucleic Acids Research* **2005**, *33*, 6837.

[58] S. Abes, D. Williams, P. Prevot, A. Thierry, M. J. Gait, B. Lebleu, *Journal of Controlled Release* **2006**, *110*, 595.

[59] N. Bendifallah, F. W. Rasmussen, V. Zachar, P. Ebbesen, P. E. Nielsen, U. Koppelhus, *Bioconjugate Chemistry* **2006**, *17*, 750.

[60] T. Shiraishi, P. E. Nielsen, *Febs Letters* **2006**, *580*, 1451.

[61] S. Abes, J. J. Turner, G. D. Ivanova, D. Owen, D. Williams, A. Arzumanov, P. Clair, M. J. Gait, B. Lebleu, *Nucleic Acids Research* **2007**, *35*, 4495.

[62] J. X. Hu, D. R. Corey, *Biochemistry* **2007**, *46*, 7581.

[63] J. J. Turner, S. Jones, M. M. Fabani, G. Ivanova, A. A. Arzumanov, M. J. Gait, *Blood Cells Molecules and Diseases* **2007**, *38*, 1.

[64] G. D. Ivanova, A. Arzumanov, R. Abes, H. Yin, M. J. A. Wood, B. Lebleu, M. J. Gait, *Nucleic Acids Research* **2008**, *36*, 6418.

[65] G. D. Ivanova, A. A. Arzumanov, J. J. Turner, M. M. Fabani, R. Abes, B. Lebleu, M. J. Gait, in *14th Symposium on Chemistry of Nucleic Acid Components* (Ed.: M. Hocek), Cesky Krumlov, CZECH REPUBLIC, **2008**, pp. 103.

[66] M. Kosuge, T. Takeuchi, I. Nakase, A. T. Jones, S. Futaki, *Bioconjugate Chemistry* **2008**, *19*, 656.

[67] I. Nakase, T. Takeuchi, G. Tanaka, S. Futaki, *Advanced Drug Delivery Reviews* **2008**, *60*, 598.

[68] K. M. Stewart, K. L. Horton, S. O. Kelley, *Organic & Biomolecular Chemistry* **2008**, *6*, 2242.

[69] P. A. Wender, W. C. Galliher, E. A. Goun, L. R. Jones, T. H. Pillow, *Advanced Drug Delivery Reviews* **2008**, *60*, 452.

[70] A. Ziegler, *Advanced Drug Delivery Reviews* **2008**, *60*, 580.

[71] D. Derossi, A. H. Joliot, G. Chassaing, A. Prochiantz, *Journal of Biological Chemistry* **1994**, *269*, 10444.

[72] A. D. Frankel, C. O. Pabo, *Cell* **1988**, *55*, 1189.

[73] M. Green, P. M. Loewenstein, *Cell* **1988**, *55*, 1179.

[74] E. Vives, P. Brodin, B. Lebleu, *Journal of Biological Chemistry* **1997**, *272*, 16010.

[75] M. Pooga, U. Soomets, M. Hallbrink, A. Valkna, K. Saar, K. Rezaei, U. Kahl, J. X. Hao, X. J. Xu, Z. Wiesenfeld-Hallin, T. Hokfelt, T. Bartfai, U. Langel, *Nature Biotechnology* **1998**, *16*, 857.

[76] G. Aldrian-Herrada, M. G. Desarmenien, H. Orcel, L. Boissin-Agasse, J. Mery, J. Brugidou, A. Rabie, *Nucleic Acids Research* **1998**, *26*, 4910.

[77] M. Pooga, M. Hallbrink, M. Zorko, U. Langel, *Faseb Journal* **1998**, *12*, 67.

[78] G. Cutrona, E. M. Carpaneto, M. Ulivi, S. Roncella, O. Landt, M. Ferrarini, L. C. Boffa, *Nature Biotechnology* **2000**, *18*, 300.

[79] L. Good, S. K. Awasthi, R. Dryselius, O. Larsson, P. E. Nielsen, *Nature Biotechnology* **2001**, *19*, 360.

[80] J. B. Rothbard, S. Garlington, Q. Lin, T. Kirschberg, E. Kreider, P. L. McGrane, P. A. Wender, P. A. Khavari, *Nature Medicine* **2000**, *6*, 1253.

[81] H. Knudsen, P. E. Nielsen, *Nucleic Acids Research* **1996**, *24*, 494.

[82] L. Good, P. E. Nielsen, *Nature Biotechnology* **1998**, *16*, 355.

[83] M. Eriksson, P. E. Nielsen, L. Good, *Journal of Biological Chemistry* **2002**, *277*, 7144.

[84] L. C. Boffa, E. M. Carpaneto, V. G. Allfrey, *Proceedings of the National Academy of Sciences of the United States of America* **1995**, *92*, 1901.

[85] L. C. Boffa, P. L. Morris, E. M. Carpaneto, M. Louissaint, V. G. Allfrey, *Journal of Biological Chemistry* **1996**, *271*, 13228.

[86] T. A. Vickers, M. C. Griffith, K. Ramasamy, L. M. Risen, S. M. Freier, *Nucleic Acids Research* **1995**, *23*, 3003.

[87] R. Tonelli, S. Purgato, C. Camerin, R. Fronza, F. Bologna, S. Alboresi, M. Franzoni, R. Corradini, S. Sforza, A. Faccini, J. M. Shohet, R. Marchelli, A. Pession, *Molecular Cancer Therapeutics* **2005**, *4*, 779.

[88] B. A. Janowski, K. E. Huffman, J. C. Schwartz, R. Ram, D. Hardy, D. S. Shames, J. D.

Minna, D. R. Corey, *Nature Chemical Biology* **2005**, *1*, 216.

[89] B. A. Janowski, K. Kaihatsu, K. E. Huffman, J. C. Schwartz, R. Ram, D. Hardy, C. R. Mendelson, D. R. Corey, *Nature Chemical Biology* **2005**, *1*, 210.

[90] B. M. McMahon, J. A. Stewart, M. D. Bitner, A. Fauq, D. J. McCormick, E. Richelson, *Life Sciences* **2002**, *71*, 325.

[91] V. V. Demidov, M. D. Frank-Kamenetskii, *Methods* **2001**, *23*, 108.

[92] S. V. Smulevitch, C. G. Simmons, J. C. Norton, T. W. Wise, D. R. Corey, *Nature Biotechnology* **1996**, *14*, 1700.

[93] T. Bentin, H. J. Larsen, P. E. Nielsen, *Biochemistry* **2003**, *42*, 13987.

[94] K. Kaihatsu, R. H. Shah, X. Zhao, D. R. Corey, *Biochemistry* **2003**, *42*, 13996.

[95] K. Kaihatsu, D. A. Braasch, A. Cansizoglu, D. R. Corey, *Biochemistry* **2002**, *41*, 11118.

[96] T. Bentin, P. E. Nielsen, *Journal of the American Chemical Society* **2003**, *125*, 6378.

[97] S. Rapireddy, G. He, S. Roy, B. A. Armitage, D. H. Ly, *Journal of the American Chemical Society* **2007**, *129*, 15596.

[98] V. Chenna, S. Rapireddy, B. Sahu, C. Ausin, E. Pedroso, D. H. Ly, *Chembiochem* **2008**, *9*, 2388.

[99] T. Ishizuka, K. Otani, J. Sumaoka, M. Komiyama, *Chemical Communications* **2009**, 1225.

[100] T. Ishizuka, J. Yoshida, Y. Yamamoto, J. Sumaoka, T. Tedeschi, R. Corradini, S. Sforza, M. Komiyama, *Nucleic Acids Research* **2008**, *36*, 1464.

[101] J. C. Norton, M. A. Piatyszek, W. E. Wright, J. W. Shay, D. R. Corey, *Nature Biotechnology* **1996**, *14*, 615.

[102] J. G. Harrison, C. Frier, R. Laurant, R. Dennis, K. D. Raney, S. Balasubramanian, *Bioorganic & Medicinal Chemistry Letters* **1999**, *9*, 1273.

[103] R. Villa, M. Folini, S. Lualdi, S. Veronese, M. G. Daidone, N. Zaffaroni, *Febs Letters* **2000**, *473*, 241.

[104] A. Whitney, G. Gavory, S. Balasubramanian, *Chemical Communications* **2003**, 36.

[105] M. Folini, R. Bandiera, E. Millo, P. Gandellini, G. Sozzi, P. Gasparini, N. Longoni, M. Binda, M. G. Daidone, K. Berg, N. Zaffaroni, *Cell Proliferation* **2007**, *40*, 905.

[106] F. Pellestor, P. Paulasova, *Int J Mol Med* **2004**, *13*, 521.

[107] L. Mologni, P. E. Nielsen, C. Gambacorti-Passerini, *Biochemical and Biophysical Research Communications* **1999**, *264*, 537.

[108] M. R. Lewis, F. Jia, F. Gallazzi, Y. Wang, J. L. Zhang, N. Shenoy, S. Z. Lever, M. Hannink, *Bioconjugate Chemistry* **2002**, *13*, 1176.

[109] F. Gallazzi, F. Jia, L. A. Landon, N. Shenoy, G. Sivaguru, M. Hannink, S. Z. Lever, M. R. Lewis, in *18th American Peptide Symposium*, Boston, Massachusetts, **2003**, p. P475.

[110] F. Gallazzi, Y. Wang, F. Jia, N. Shenoy, L. A. Landon, M. Hannink, S. Z. Lever, M. R. Lewis, *Bioconjugate Chemistry* **2003**, *14*, 1083.

[111] F. Jia, F. Gallazzi, N. Shenoy, J. Zhang, S. Z. Lever, M. Hannink, M. R. Lewis, in *50th Annual Meeting of the Society-of-Nuclear-Medicine*, New Orleans, Louisiana, **2003**, p. 1318.

[112] F. Jia, S. D. Figueroa, F. GallazZi, B. S. Balaji, M. Hannink, S. Z. Lever, T. J. Hoffman, M. R. Lewis, *Journal of Nuclear Medicine* **2008**, *49*, 430.

[113] H. Knudsen, P. E. Nielsen, *Anti-Cancer Drugs* **1997**, *8*, 113.

[114] L. Mologni, C. Gambacorti-Passerini, *Letters in Peptide Science* **2003**, *10*, 297.

[115] N. V. Amirkhanov, E. Wickstrom, in *16th International Roundtable of the International-Society-for-Nucleosides-Nucleotides-and-Nucleic-Acids (IS3NA)*, Minneapolis, MN, **2004**, pp. 423.

[116] A. Chakrabarti, M. R. Aruva, S. P. Sajankila, M. L. Thakur, E. Wickstrom, in *16th International Roundtable of the International-Society-for-Nucleosides-Nucleotides-and-Nucleic-Acids (IS3NA)*, Minneapolis, MN, **2004**, pp. 409.

[117] T. Shiraishi, P. E. Nielsen, *Nucleic Acids Research* **2004**, *32*, 4893.

[118] X. B. Tian, R. Winter, M. R. Aruva, K. J. Zhang, C. A. Cardi, M. L. Thakur, E. Wickstrom, in *AACR/NCI/EORTC International Conference on Molecular Targets and Cancer Therapeutics*, Philadelphia, PA, **2005**, pp. 9130S.

[119] A. Zannetti, S. Del Vecchio, A. Romanelli, S. Scala, M. Saviano, G. Cali, M. P. Stoppelli, C. Pedone, M. Salvatore, *Biochemical Pharmacology* **2005**, *70*, 1277.

[120] H. Orum, P. E. Nielsen, M. Egholm, R. H. Berg, O. Buchardt, C. Stanley, *Nucleic Acids Research* **1993**, *21*, 5332.

[121] C. Thiede, E. Bayerdorffer, R. Blasczyk, B. Wittig, A. Neubauer, *Nucleic Acids Research* **1996**, *24*, 983.

[122] Y. Myal, A. Blanchard, P. Watson, M. Corrin, A. Huang, S. Troup, B. Iwasiow, *American Journal of Pathology* **1998**, *153*, ST16.

[123] D. G. Murdock, N. C. Christacos, D. C. Wallace, *Nucleic Acids Research* **2000**, *28*, 4350.

[124] K. A. Kreuzer, P. le Coutre, O. Landt, L. K. Na, M. Schwarz, K. Schultheis, A. Hochhaus, B. Dorken, *Annals of Hematology* **2003**, *82*, 284.

[125] T. Takiya, Y. Horie, S. Futo, Y. Matsumoto, K. Kawai, T. Suzuki, *Journal of Bioscience and Bioengineering* **2003**, *96*, 588.

[126] T. Takiya, S. Futo, M. Tsuna, T. Namimatsu, T. Sakano, K. Kawai, T. Suzuki, *Bioscience Biotechnology and Biochemistry* **2004**, *68*, 360.

[127] M. Urata, Y. Wada, S. H. Kim, W. Chumpia, Y. Kayamori, N. Hamasaki, D. C. Kang, *Clinical Chemistry* **2004**, *50*, 2045.

[128] J. Dabritz, J. Hanfler, R. Preston, J. Stieler, H. Oettle, *British Journal of Cancer* **2005**, *92*, 405.

[129] M. Miyake, K. Sugano, K. Kawashima, H. Ichikawa, K. Hirabayashi, T. Kodama, H. Fujimoto, T. Kakizoe, Y. Kanai, K. Fujimoto, Y. Hirao, *Biochemical and Biophysical Research Communications* **2007**, *362*, 865.

[130] M. Beau-Faller, M. Legrain, A. C. Voegeli, E. Guerin, T. Lavaux, A. M. Ruppert, A. Neuville, G. Massard, J. M. Wihlm, E. Quoix, P. Oudet, M. P. Gaub, *British Journal of Cancer* **2009**, *100*, 985.

[131] H. Stender, *Expert Review of Molecular Diagnostics* **2003**, *3*, 649.

[132] V. Gonzalez, E. Padilla, M. Gimenez, C. Vilaplana, A. Perez, G. Fernandez, M. D. Quesada, M. A. Pallares, V. Ausina, *European Journal of Clinical Microbiology & Infectious Diseases* **2004**, *23*, 396.

[133] G. N. Forrest, *Expert Review of Molecular Diagnostics* **2007**, *7*, 231.

[134] S. B. Selvaraju, R. Kapoor, J. S. Yadav, *Molecular and Cellular Probes* **2008**, *22*, 273.

[135] H. PerryOkeefe, X. W. Yao, J. M. Coull, M. Fuchs, M. Egholm, *Proceedings of the National Academy of Sciences of the United States of America* **1996**, *93*, 14670.

[136] C. Seeger, H. G. Batz, H. Orum, *Biotechniques* **1997**, *23*, 512.

[137] A. Puschl, T. Boesen, G. Zuccarello, O. Dahl, S. Pitsch, P. E. Nielsen, *Journal of Organic Chemistry* **2001**, *66*, 707.

[138] L. D. Fader, Y. S. Tsantrizos, *Organic Letters* **2002**, *4*, 63.

[139] A. Porcheddu, G. Giacomelli, *Current Medicinal Chemistry* **2005**, *12*, 2561.

[140] G. Haaima, A. Lohse, O. Buchardt, P. E. Nielsen, *Angewandte Chemie-International Edition in English* **1996**, *35*, 1939.

[141] S. Sforza, G. Haaima, R. Marchelli, P. E. Nielsen, *European Journal of Organic Chemistry* **1999**, 197.

[142] J. O. Smith, D. A. Olson, B. A. Armitage, *Journal of the American Chemical Society* **1999**, *121*, 2686.

[143] K. L. Dueholm, K. H. Petersen, D. K. Jensen, M. Egholm, P. E. Nielsen, O. Buchardt, *Bioorganic & Medicinal Chemistry Letters* **1994**, *4*, 1077.

[144] S. Sforza, R. Corradini, S. Ghirardi, A. Dossena, R. Marchelli, *European Journal of Organic Chemistry* **2000**, 2905.

[145] P. Zhou, M. M. Wang, L. Du, G. W. Fisher, A. Waggoner, D. H. Ly, *Journal of the American Chemical Society* **2003**, *125*, 6878.

[146] A. Dragulescu-Andrasi, P. Zhou, G. F. He, D. H. Ly, *Chemical Communications* **2005**, 244.

[147] E. A. Englund, D. H. Appella, *Angewandte Chemie International Edition in English* **2007**, *46*, 1414.

[148] B. Sahu, V. Chenna, K. L. Lathrop, S. M. Thomas, G. Zon, K. J. Livak, D. H. Ly, *Journal of Organic Chemistry* **2009**, *74*, 1509.

[149] A. Dragulescu-Andrasi, S. Rapireddy, B. M. Frezza, C. Gayathri, R. R. Gil, D. H. Ly, *J Am Chem Soc* **2006**, *128*, 10258.

[150] M. C. de Koning, L. Petersen, J. J. Weterings, M. Overhand, G. A. van der Marel, D. V. Filippov, *Tetrahedron* **2006**, *62*, 3248.

[151] G. Haaima, H. Rasmussen, G. Schmidt, D. K. Jensen, J. S. Kastrup, P. W. Stafshede, B. Norden, O. Buchardt, P. E. Nielsen, *New Journal of Chemistry* **1999**, *23*, 833.

# Chapter 2 Design and synthesis of AP-PNA and GP-PNA oligomers

## 2.1 Design rationale

PNA is a DNA/RNA mimic with a poly-N-(2-aminoethyl)glycine (aeg) backbone (Figure 2-1a) and binds to complementary oligonucleotides through Watson-Crick base pairing with high sequence specificity. The structural simplicity, high hybridization affinity and *in vivo* stability of PNA have made it an attractive agent for antisense and antigene applications in basic biology and medicine.

However, like oligonucleotides and many large oligomeric compounds of similar nature, PNA has poor cell membrane permeability which has severely limited its usage in biomedical research. As discussed in chapter 1, many methods have been developed to improve PNA cellular uptake during the past two decades. Among these, conjugating PNA to cell penetrating peptides (CPPs) is a popular approach. When attached to PNA, those peptides were found to be good vehicles to deliver PNA into living cells. However, problems may arise when using such conjugate systems *in vivo*. Peptides usually have very short half lives *in vivo* due to their susceptibility to proteases. This is especially true in view of the Arg-rich nature of most cell-permeable peptides. So, the peptide component may be degraded and lose its delivering function before the conjugate reaches target cells. Including a peptide moiety may also bring about issues of immunogenicity.

Many structure-activity relationship studies on CPPs [1-3] have revealed some

key elements for cell penetration. Firstly, all CPPs share the universal property of net positive charge. Arg residues contribute more to CPP internalization than Lys residues.[4, 5] This can be explained with the donation of two hydrogen bonds by guanidino group of Arg, compared to one hydrogen bond by amino group of Lys at physiological pH.[6] Secondly, hydrophobicity play a key role in the uptake of short cationic peptides.[7] However, over hydrophobicity will make the peptide stuck in the membrane rather than internalization.

Therefore, it is highly desirable to synthesize a PNA-based antigene or antisense agent with built-in cell permeability through incorporating the membrane translocation features of cell-permeable peptides onto the PNA backbone. Many reports have been published describing backbone-modified PNA analogs for this and other purposes.[8-14] Among these, introducing an amino acid side chain bearing a positive charge, such as lysine, ornithine or arginine at the α- or γ-C of aegPNA backbone (Figure 2-1a), results in improvement of water solubility and cellular uptake.[8-14] Notably, PNA with either an L-Arg or D-Arg side chain at the α-carbon of the aeg backbone, i.e., *D*-Arg$_\alpha$-PNA and *L*-Arg$_\alpha$-PNA, have been reported to have good cell permeability.[10, 11] It should be noted that almost all the modifications have been focused on the aeg backbone carbons which inevitably generate a chiral centre. Often, the two stereoisomers exhibit distinct helical structures and binding behaviors in hybridizing with DNA and RNA sequences, possibly as a result of differential interstrand or intrastrand steric interactions caused by the two configurations.[15] It is therefore of utmost importance to

preserve the chiral integrity of such PNA analogs during the synthetic process.

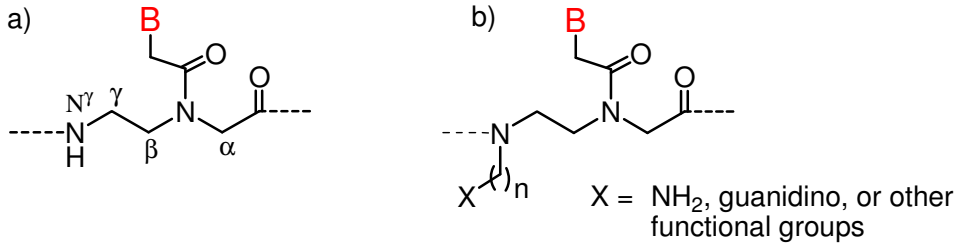



**Figure 2-1.** Structures of (a) an aegPNA residue, (b) a peptoid PNA residue (an amino-peptoid PNA residue, when X = $NH_2$). B denotes nucleobase.

Modifying the γ-nitrogen of the aeg backbone has also been considered. However, in an early study by the Nielsen group,[16] $N^{\gamma}$ methylation was found to have a negative impact on the hybridizing affinity of the modified PNAs, which probably has deterred further attempts to modify this position.

In our efforts to develop PNA-based agents, we decided to reinvestigate the modifications at the γ-nitrogen by appending to it an alkyl side chain carrying a functional headgroup (Figure 2-1b). We name the side chain a peptoid-like side chain, as it is analogous to the $N^{\alpha}$-side chains in N-substituted oligoglycines or peptoids. Thus two types of new PNA analogs, amino-peptoid PNA (AP-PNA) and guanidino-peptoid PNA (GP-PNA) were designed. To be determined is the optimal length of the spacer that links the headgroup to the backbone. AP-PNA and GP-PNA preserves the achiral nature of PNA and therefore causes no stereochemistry complications synthetically. Introducing such a side chain may also bring some of the beneficial effects observed for a similar side chain extended from the α- or γ-C. In addition, the functional headgroup could also serve as a

suitable anchor point to attach various structural moieties of biophysical and biochemical interest. Furthermore, given the ease in choosing the length of the peptoid side chain and the nature of the functional headgroup, the electrostatic effects of such a side chain can be examined systematically.

## 2.2 Synthesis of AP-PNA monomers

PNA oligomer assembling was first accomplished by Boc/Cbz strategy where the Boc group was used for the protection of the amino group of the backbone and the Cbz was used for protection of the exocyclic amino groups of the nucleobases. However, the Boc/Cbz strategy requires harsh conditions (HF or TFMSA) to remove the Cbz group. Thus other strategies with milder deprotection conditions, such as Mmt/acyl, Fmoc/acyl, Fmoc/Bhoc, Fmoc/Boc and Fmoc/Cbz, were investigated. Other alternative amino protecting groups have also been prepared, such as Dde/Mmt, NVOC/acyl, and azide/Bhoc. In this study, Fmoc/acyl and Fmoc/Bhoc strategies are investigated.

For the synthesis of GP-PNA oligomers, the guanidino group can be introduced to the amine of the peptoid side-chain at the monomer stage or the oligomer stage. If introduced at the monomer stage, specialized and expensive guanidine protecting groups, such as Pmc or Pbf, are needed. If introduced at the oligomer stage, one can convert the AP-PNA oligomers to polyguanidine GP-PNA in good yields by treatment with excess amount of a guanylation reagent like

N,N'-bis-Boc-S-methylisothiourea. In this study, we used the latter, i.e., the global guanylation method. In addition, the amino group could be a suitable anchor point to attach various structural moieties of biophysical and biomedical interest.

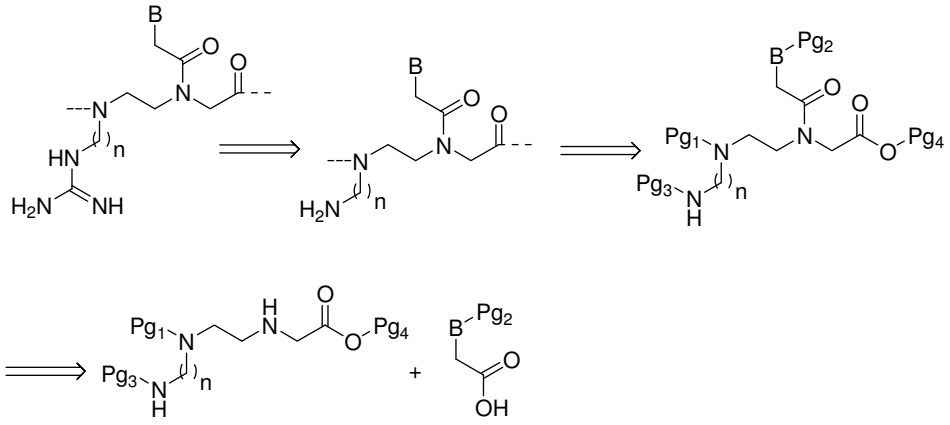



**Figure 2-2.** Retrosynthetic scheme of GP-PNA and AP-PNA oligomers.

Synthesis of AP-PNA monomers can be divided into two parts, the protected nucleobase acetic acids and AP-PNA backbones. As shown in Figure 2-2, synthesizing AP-PNA monomers involves the use of 4 protecting groups, i.e., the γ–N protecting group (Pg1), nucleobase protecting group (Pg2), peptoid side chain amino protecting group (Pg3) and the carboxyl protecting group (Pg4). We use the Fmoc group to protect the γ–N where the chain elongates, Boc group to protect the peptoid side chain amino group, methyl to protect the carboxylic acid on the backbone, and Bhoc or acyl to protect the nucleobases.

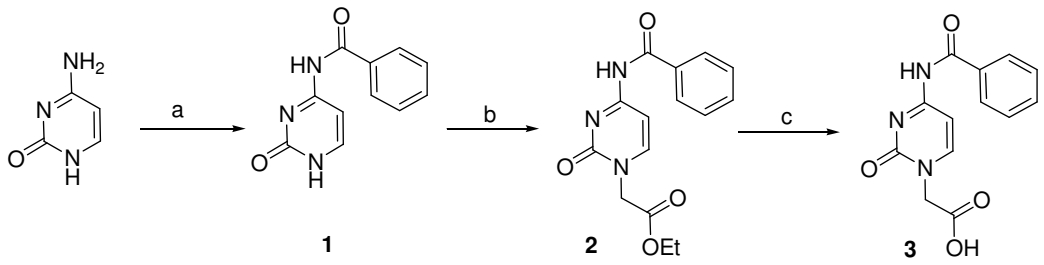
### 2.2.1 Synthesis of nucleobase acetic acid derivatives

Orthogonally protected nucleobases and their derivatives are required for the chemical synthesis of PNA as the reactive groups present on exocyclic nucleobases can interfere with the synthesis and give rise to side reactions, such as acetylation during the capping procedure and chain extension from these groups during assembling process. Therefore choosing proper protecting groups is necessary to prevent undesired acetylation or chain extension. Carboxymethylated thymine is commercially available and does not require protection during synthesis. Cytosine, adenine and guanine have an exocyclic amino group and so need to be protected properly. Both Bhoc and acyl protecting groups are chosen to protect the exocyclic nucleobase amino groups. Fmoc/Bhoc PNA monomers are commercially available and are used widely for routine PNA synthesis. The Bhoc protecting group can be removed simultaneously with Boc under 95% TFA treatment. However, introducing Bhoc group onto the nucleobases requires expensive reagents such as carbonyldiimidazole or triphosgene and the Bhoc group is unstable under milder acidic conditions. On the other hand, acyl protecting group which is removed by 35% ammonia is quite stable under basic or acidic conditions and is suitable for the synthesis of PNA-DNA chimeras. In addition, acyl protecting group is stable under Boc protecting group removal condition and therefore is suitable for anchoring other functional groups on the amino groups of the peptoid side chain, while keeping the exocyclic nucleobase amino group intact. Therefore, both Bhoc and acyl protected nucleobase acetic acid derivatives were prepared according to

literature procedures,[17-20] which were then used to synthesize the corresponding PNA monomers. Bhoc protected guanine acetic acid was obtained commercially.

## 2.2.1.1 Synthesis of cytosine acetic acids

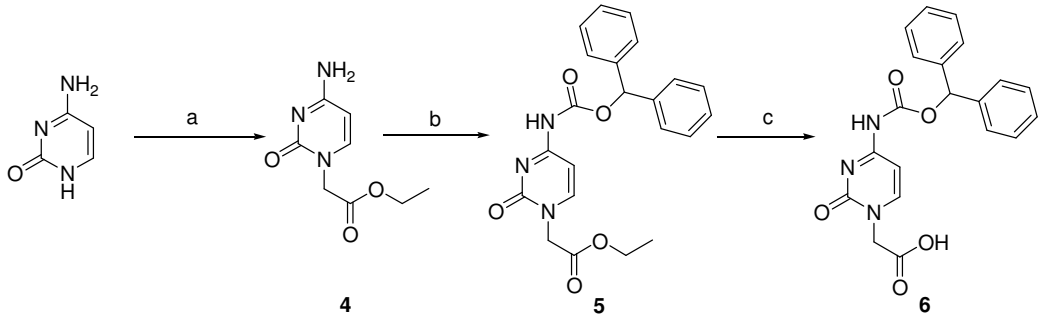
### 2.2.1.1.1 Synthesis of 4-*N*-(benzoyl)cytosine-1-acetic acid




**Scheme 2-1. Synthesis of 4-*N*-(benzoyl)cytosine-1-acetic acid: (a) Benzoyl chloride, Pyridine; (b) Ethyl bromoacetate, NaH/DMF; (c)NaOH, $H_2O$.**

Synthesis of acyl protected cytosine acetic acid was relatively easy as no other mono- or double-alkylation by-products were produced in the process. Cytosine reacted with benzoyl chloride in dry pyridine to form $N^4$-benzoyl cytosine **1.** The poor solubility of compound **1** (even in DMSO) impeded chemical manipulation and structure determination using NMR spectroscopy. However, alkylation of **1** enhanced the solubility in organic solvents. Compound **1** was converted to its sodium salt using NaH in dry DMF and alkylated by ethyl bromoacetate. The resulting ethyl ester **2** was hydrolyzed using two equivalents of NaOH in water. After adjusting the pH to 3, the desired product **3** was collected as a white precipitate.

### 2.2.1.1.2 Synthesis of 4-*N*-(benzhydryloxycarbonyl)cytosine-1-acetic acid

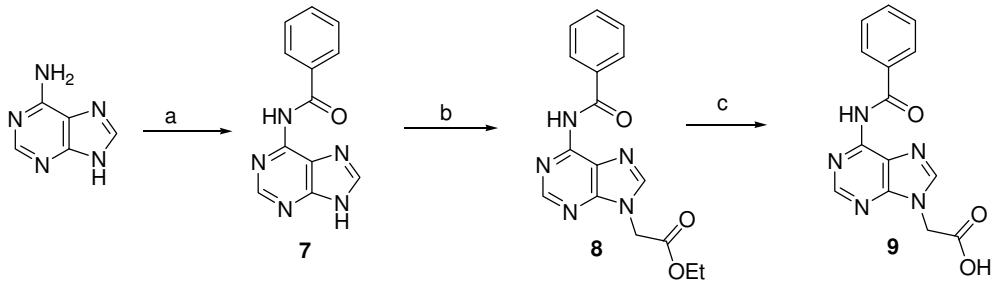



**Scheme 2-2. Synthesis of 4-*N*-(benzhydryloxycarbonyl)cytosine-1-acetic acid: (a) Ethyl bromoacetate, NaH, DMF; (b) N,N'-carbonyldiimidazole, benzhydrol, DMF; (c) NaOH, HCl, $H_2O$.**

To synthesize Bhoc protected cytosine acetic acid, cytosine was alkylated by ethyl bromoacetate in the presence of NaH in dry DMF. The alkylated product **4** was reacted with carbonyldiimidazole and benzhydrol at 60 $^{o}$C to get Bhoc protected Cytosine ester **5** which was further hydrolyzed using NaOH to afford the corresponding 4-*N*-(benzhydryloxycarbonyl)cytosine-1-acetic acid **6** in good yield.

## 2.2.1.2 Synthesis of adenine acetic acid

### 2.2.1.2.1 Synthesis of 6-*N*-(benzoyl)adenine-9-acetic acid




**Scheme 2-3. Synthesis of 6-*N*-(benzoyl)adenine-9-acetic acid: (a) Benzoyl chloride, Pyridine; (b) Ethyl bromoacetate, NaH/DMF; (c)NaOH, $H_2O$.**

Alkylation of the purine bases is more difficult compared to that of the pyrimidine bases, because purine has an imidazole ring fused to a pyrimidine ring. The two nitrogen atoms on the imidazole ring are chemically equivalent as a result the purine bases exist as a pair of tautomers. Therefore alkylation of purine bases will give a mixture of the monoalkylated $N^9$- and $N^7$-regioisomers. In the case of adenine, both isomers (**8** and **8'**) were obtained after alkylation. HPLC data showed that the required $N^9$- derivative **8** (more polar) prevailed in the ratio of 2:1 to **8'** (less polar) as shown in Figure 2-3A.

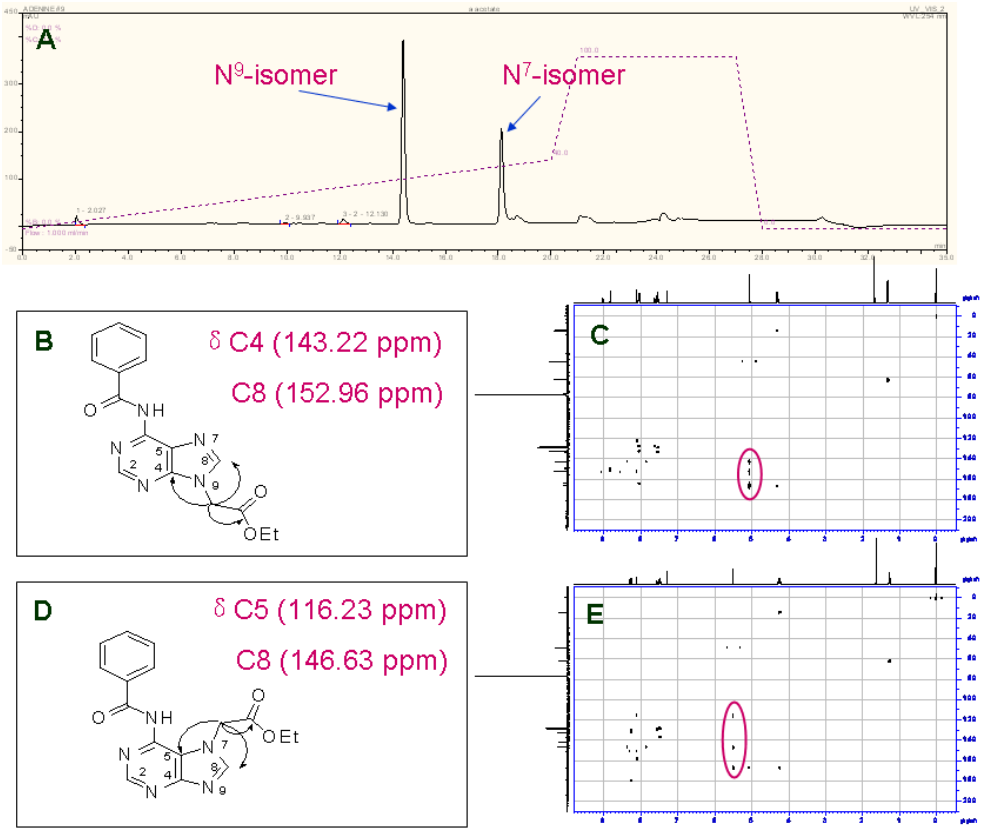
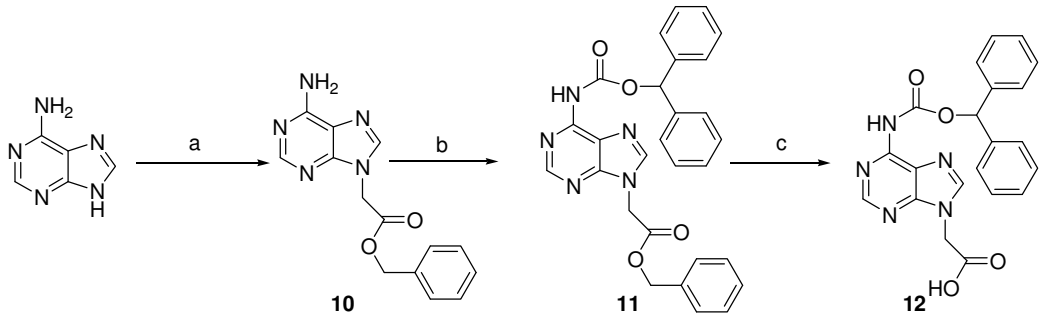



**Figure 2-3.** The reversed-phase HPLC trace and the HMBC results of the $N^7$- and $N^9$-regioisomers of benzoyl protected ethoxycarbonylmethyladenine. (A) HPLC trace of the two isomers; (B) The cross coupling region of $N^9$-regioisomers and their chemical shift of the coupling carbon; (C) HMBC spectrum of $N^9$-regioisomers; (D) The cross coupling region of $N^7$-regioisomers and their chemical shift of the coupling carbon; (E) HMBC spectrum of $N^7$-regioisomers.

To identify $N^9$- and $N^7$-regioisomers, 2D NMR HMBC experiment was carried out. The regiochemical assignment of $^{13}C$ signals in substituted purine is known from INEPT experiments.[21, 22] The $N^9$- and $N^7$-isomers can be distinguished from the cross-peaks arising from the coupling of methylene protons of the $\underline{CH_2}COOEt$ fragment with C4 and C5 carbons. In $N^9$-isomer **8**, the methylene protons (singlet at $\delta$ = 5.07 ppm) are coupled to both C4 and C8 carbons but not C5, while the analogous methylene protons of $N^7$-isomer **8'** (singlet at $\delta$ = 5.50 ppm) shows coupling to C5 and C8 but not C4. The isomer **8**, purified by column

chromatography, was subjected to hydrolytic conditions to yield $N^6$-benzoyl-9-carboxymethyladenine **9**. As it will be mentioned later, the undesired $N^7$ alkylation reaction can be avoided by using a bulky alkylation reagent.

### 2.2.1.2.2 Synthesis of 6-*N*-(benzhydryloxycarbonyl)adenine-9-acetic acid

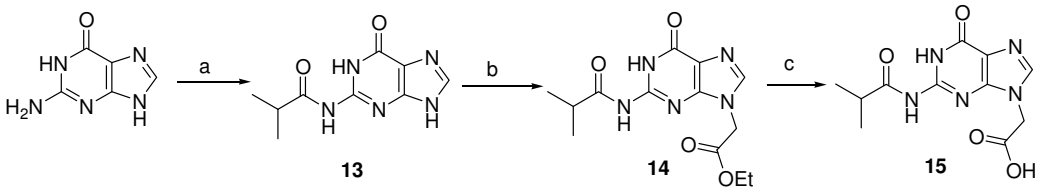



**Scheme 2-4. Synthesis of 6-*N*-(benzhydryloxycarbonyl)adenine-9-acetic acid: (a) Benzyl bromoacetate, NaH, DMF; (b) N,N'-carbonyldiimidazole, benzhydrol, DMF; (c) LiOH, HCl, $H_2O$.**

To synthesize Bhoc protected adenine acetic acid, the same method as for the synthesis of Bhoc protected cytosine-acetic acid was used. The adenine was alkylated using benzyl bromoacetate in the presence of NaH in DMF. It was surprising to find that only one product (benzyl adenine-9-acetate **10**) was formed. **10** was then reacted with carbonyldiimidazole and benzhydrol in DMF at 60 $^{o}$C to get the Bhoc protected adenine ester **11**, which was hydrolyzed using LiOH to give 6-*N*-(benzhydryloxycarbonyl)adenine-9-acetic acid **12**.

## 2.2.1.3 Synthesis of guanine acetic acid

### 2.2.1.3.1 Synthesis of 2-*N*-isobutyrylguanine-9-acetic acid




**Scheme 2-5. Synthesis of 2-*N*-isobutyrylguanine-9-acetic acid: (a) Isobutyryl chloride, TEA/DMF; (b) Ethyl bromoacetate, $Na_2CO_3$/DMF; (c)NaOH, $H_2O$.**

Although the guanine 2-amino group is not nucleophilic enough to interfere with many transformations, the poor solubility of unprotected guanine is a real problem in subsequent reactions. Therefore, a lipophilic group is needed to increase its solubility in organic solvents. Hence, the isobutyryl protecting group was introduced at the $N^2$-position by reacting it with isobutyryl chloride in the presence of one equivalent of TEA in dry DMF. Alkylation of the crude product at 0 ℃ and in the presence of NaH yielded not only the monoalkylated $N^9$- and $N^7$-regioisomers, but also the double alkylated guanine derivatives. This made the separation of desired products difficult. Therefore, alkylation was performed at 60 ℃ in the presence of two equivalents of $Na_2CO_3$. This reaction resulted in the formation of monoalkylated $N^9$- and $N^7$-regioisomers with essentially no double alkylation products. The required $N^9$- derivative **14** (more polar) prevailed in the ratio 55:45 to **14'** (less polar) as shown in Figure 2-4A. The two isomers were then separated by column chromatography.

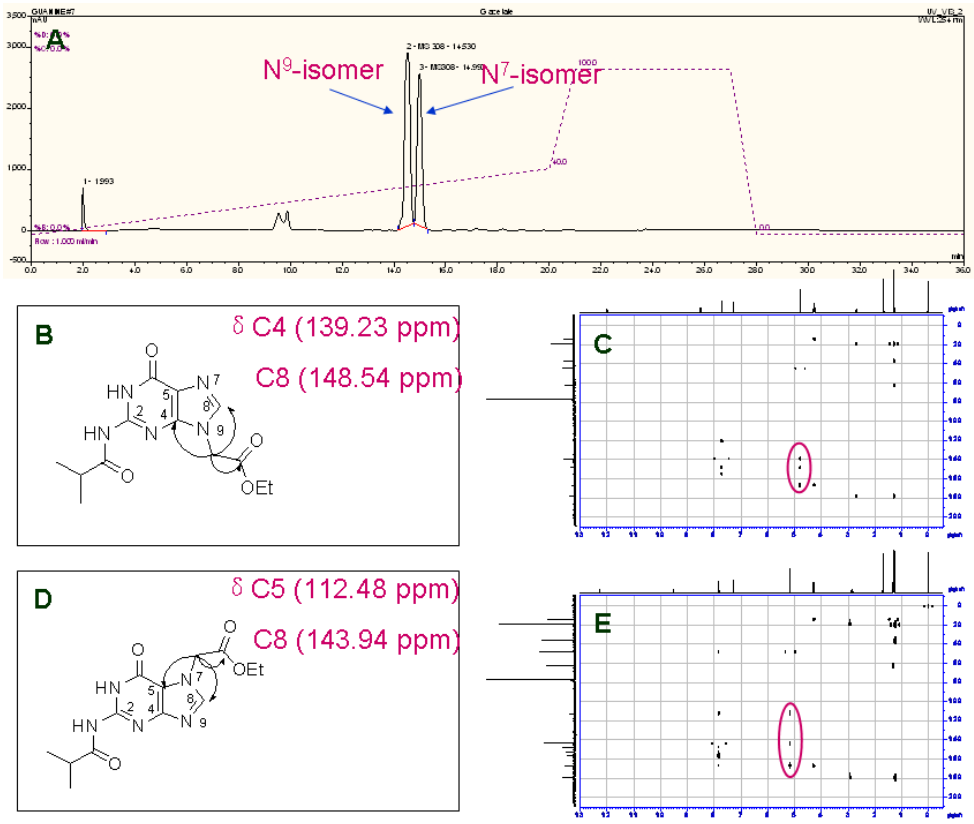
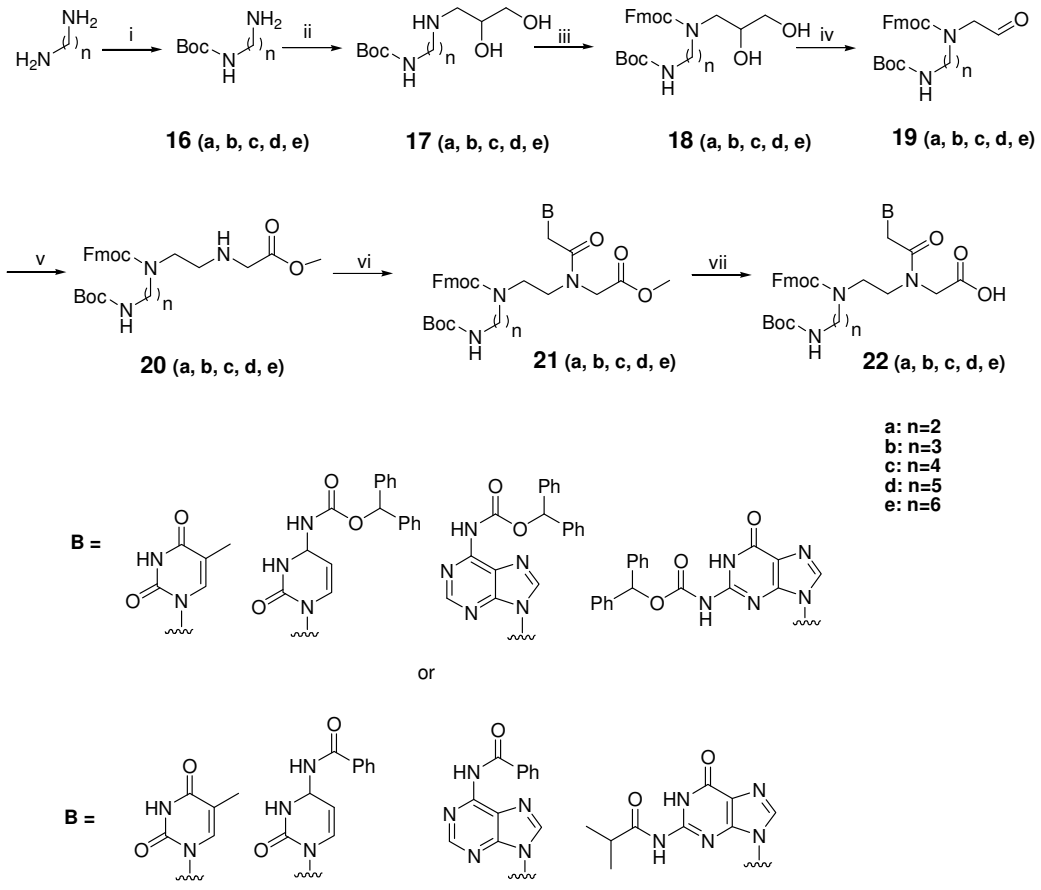



**Figure 2-4.** The reversed-phase HPLC trace and HMBC spectra of the $N^7$- and $N^9$-regioisomers of $N^2$-isobutyryl-ethoxycarboxymethylguanine A) HPLC trace of the two isomers; B) The cross coupling region of $N^9$-regioisomers and their chemical shift of the coupling carbon; C) HMBC spectrum of $N^9$-regioisomers; D) The cross coupling region of $N^7$-regioisomers and their chemical shift of the coupling carbon; E) HMBC spectrum of $N^7$-regioisomers.

Identification of $N^9$- and $N^7$-regioisomers **14** and **14'** was performed using HMBC spectra. As shown in the spectra, the methylene protons of $N^9$-derivative **14** (singlet at $\delta$ = 4.80 ppm) are coupled to both C4 and C8 carbons but not C5. The analogous methylene protons of $N^7$-derivative **14'** (singlet at $\delta$ = 5.17 ppm) show coupling to C5 and C8 but not C4. Base promoted hydrolysis and neutralization of the required isomer **14** resulted in **15** in quantitative yields.

## 2.2.2 Synthesis of AP-PNA monomers




**Scheme 2-6. Synthesis of AP-PNA monomers. (i) $(Boc)_2O$, DCM; (ii) Glycidol, THF; (iii) Fmoc-OSu, DCM; (iv) $NaIO_4$, THF/$H_2O$; (v) H-Gly-OMe, $NaCNBH_3$, THF/MeOH; (vi)$BCH_2COOH$, HATU, DIEA, DMF; (vii) NaOH.**

The method used by Nielsen *et al.*[16] for the synthesis of N-methyl monomers was adapted for the synthesis of AP-PNA monomers. Scheme 2-6 shows the reaction steps for the preparation of the AP-PNA monomers with different peptoid side chain length. All starting materials are very cheap and easy to obtain except 1, 5-diaminopentane. Most of the reactions are conducted at room temperature and

the reaction can be easily monitored either by HPLC and/or ESI-MS. Di-*tert*-butyl dicarbonate ($Boc_2O$) was reacted with excess amounts of alkylenediamine to form monoprotected alkylenediamine **16**. Compound **16** was then reacted with equi-molar glycidol to produce **17** which, after Fmoc protection of the backbone amine, was oxidatively cleaved with $NaIO_4$ to produce **19**. Compound **19** was subsequently reacted with glycine methyl ester hydrochloride at neutral pH to yield the corresponding imine *in situ*, which was reduced with $NaBH_3CN$ to give the desired AP-PNA backbone **20**. After purification by flash chromatography, the pure oil-like **20** was coupled to the acyl- or Bhoc-protected nucleobase-acetic acids in the presence of HATU and DIEA to give the corresponding AP-PNA monomer esters **21** which, after alkaline hydrolysis, yielded the corresponding AP-PNA monomers **22**. All the structures of the so-synthesized AP-PNA monomers were shown in figure 2-5. The name of the monomers were based on this form: Bn-P-OH, where B stands for the four nucleobases (A, T, C, G), n is the side chain length (n = 2, 3, 4, 5, 6), P is protecting group for the nucleobases (Bhoc, Bz or iBu), OH means it is an acid form.

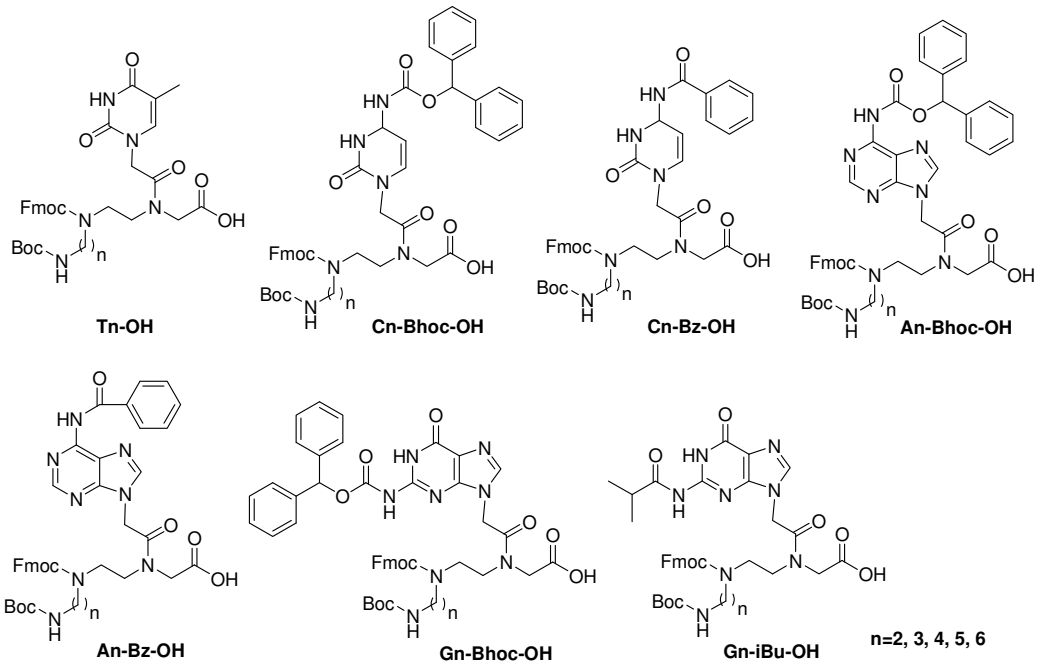
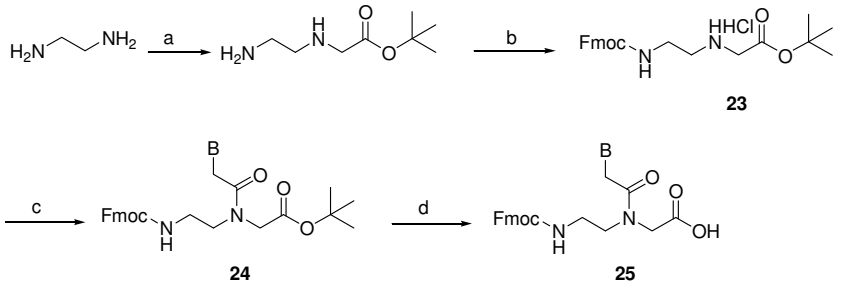



**Figure 2-5.** Structure of AP-PNA monomers containing T, A, C, G nucleobases.

## 2.3 Synthesis of aegPNA monomers




**Scheme 2-7. Synthesis of aegPNA monomer: (a) t-butyl bromoacetate, DCM; (b) Fmoc-Cl, TEA, DCM; (c) Nucleobases, HATU, NMM, DMF; (d) 95% TFA, DCM. B = T, thymine; C, $N^4$-Benzoylcytosine; A, $N^6$-Benzoyladenine; G, $N^2$-isoButyrylguanine**

Synthesis of aegPNA monomers also includes two parts: the properly protected nucleobases and the aegPNA backbone. The Fmoc/Bhoc aegPNA monomers are commercially available, while the Fmoc/acyl aegPNA monomers

are not. Synthesis of Fmoc/acyl aegPNA monomers is relatively easy. The Fmoc-protected aegPNA linker was prepared according to the method described previously.[23] Ethylenediamine was alkylated with *t*-butyl bromoacetate and the Fmoc group was installed at the primary amine selectively. The pure HCl salt **23** can be obtained by washing the crude reaction mixture with diluted aqueous HCl and concentrated to get an oil-like product which was storing at -20 $^{o}$C overnight to get the white precipitate. The installation of nucleobases to aegPNA monomer was achieved using the same method as AP-PNA monomers.

## 2.4 Manual solid phase synthesis of AP-PNA oligomers

AP-PNA oligomers were assembled on Rink amide PEGA resin using the above obtained Fmoc-protected AP-PNA and aegPNA monomers. Rink amide PEGA resin (50 – 100 mg) was first swelled in DMF/DCM (5/1) for 2 h. The Fmoc group was then deprotected by 20% piperidine in DMF (5 min × 2). After washing the resin with DMF and DCM, Fmoc-Gly-OH was coupled onto the resin using PyBOP/DIEA. Kaiser Test was used throughout the synthesis to monitor the coupling and deprotection steps. The coupling of the monomers (aegPNA or AP-PNA) was conducted with one of the following methods. Method A: monomer (2 eq), PyBOP (2 eq) and DIEA (10 eq); method B: monomer (2 eq), HATU (2 eq) and DIEA (10 eq). When the monomer was coupled to a primary amine, i.e., to an

aegPNA monomer, method A was used; when coupled to a secondary amine, i.e., to an AP-PNA monomer, method B was used. The coupling mixture was added directly to the Fmoc-deprotected resin. The coupling reaction required 4 – 6 h at room temperature as checked by Kaiser Test, and the resin was then washed with DMF and DCM thoroughly. The Fmoc deprotection and monomer coupling cycles were continued until the last residue. After the final Fmoc deprotection, the terminal amine was capped by adding 5% acetic anhydride in DMF in the presence of DIEA for 15 min. The resin was washed thoroughly with DMF and DCM and dried *in vacuo*. For acyl protected nucleobases, the dried resin was treated with 35% ammonium overnight at room temperature to remove the acyl protecting groups and washed with DMF/DCM and dried *in vacuo*. For the Bhoc protected nucleobases, this step is not required. Final deprotection of Boc and cleavage from the resin was carried out by adding 2 mL 95% TFA (2.5% TIS, 2.5% $H_2O$) to the dry resin and the suspension was shaken for 1 h at room temperature. The cleavage cocktail was filtered and the resin was washed with TFA (1 mL $\times$ 2). The combined filtration was precipitated by adding 15 mL cold diethyl ether, and centrifuged to get the crude white solid. The crude white solid was redissolved in $H_2O$ and purified by semi preparative reverse-phase HPLC. The product was then lyophilized to get a white powder. The yield usually depends on the PNA sequence and the number of modifications in the sequence. After cleavage and HPLC purification, the overall yield of a typical synthesis was around 5%. The expected product was detected by MALDI-TOF mass spectroscopy. All the expected and

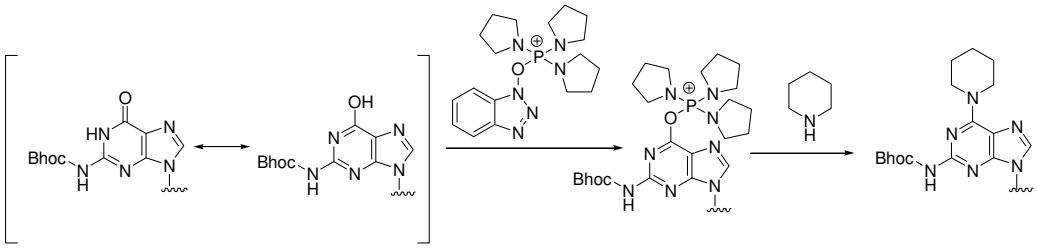
observed masses for the AP-PNA oligomers are summarized in Table 2-1.

**Table 2-1. MALDI-TOF mass spectral analysis of synthesized AP-PNA oligomers**

| | Sequence | $M_w$ calcd (isotopic) | $m/z$ $[M+H]^+$ found |
|---|---|---|---|
| aegPNA | Ac-GTAGATCACT-Gly-$NH_2$ | 2824.12 | 2826.14 |
| AP-PNA 2-1 | Ac-GTAGA**T**$^{2a}$CACT-Gly-$NH_2$ | 2867.13 | 2867.78 |
| AP-PNA 2-2 | Ac-GTAG**A**$^{2a}$ T**C**$^{2a}$ ACT-Gly-$NH_2$ | 2910.21 | 2911.05 |
| AP-PNA 2-3 | Ac-**G**$^{2a}$ TAG**A**$^{2a}$ T**C**$^{2a}$ ACT-Gly-$NH_2$ | 2953.25 | 2953.74 |
| AP-PNA 2-4 | Ac-G**T**$^{2a}$A**G**$^{2a}$A**T**$^{2a}$C**A**$^{2a}$C T-Gly-$NH_2$ | 2996.29 | 2996.91 |
| AP-PNA 3-1 | Ac-GTAGA**T**$^{3a}$CACT-Gly-$NH_2$ | 2881.18 | 2882.45 |
| AP-PNA 3-2 | Ac-GTAG**A**$^{3a}$ T**C**$^{3a}$ ACT-Gly-$NH_2$ | 2938.24 | 2939.24 |
| AP-PNA 3-3 | Ac-**G**$^{3a}$ TAG**A**$^{3a}$ T**C**$^{3a}$ ACT-Gly-$NH_2$ | 2995.3 | 2997.37 |
| AP-PNA 3-4 | Ac-G**T**$^{3a}$A**G**$^{3a}$A**T**$^{3a}$C**A**$^{3a}$C T-Gly-$NH_2$ | 3052.35 | 3053.22 |
| AP-PNA 4-1 | Ac-GTAGA**T**$^{4a}$CACT-Gly-$NH_2$ | 2895.2 | 2897.53 |
| AP-PNA 4-2 | Ac-GTAG**A**$^{4a}$ T**C**$^{4a}$ ACT-Gly-$NH_2$ | 2966.27 | 2967.76 |
| AP-PNA 4-3 | Ac-**G**$^{4a}$ TAG**A**$^{4a}$ T**C**$^{4a}$ ACT-Gly-$NH_2$ | 3037.34 | 3037.96 |
| AP-PNA 4-4 | Ac-G**T**$^{4a}$A**G**$^{4a}$A**T**$^{4a}$C**A**$^{4a}$C T-Gly-$NH_2$ | 3108.42 | 3109.16 |
| AP-PNA 5-1 | Ac-GTAGA**T**$^{5a}$CACT-Gly-$NH_2$ | 2909.23 | 2909.81 |
| AP-PNA 6-1 | Ac-GTAGA**T**$^{6a}$CACT-Gly-$NH_2$ | 2923.21 | 2924.25 |
| AP-PNA 6-2 | Ac-GTAG**A**$^{6a}$ T**C**$^{6a}$ ACT-Gly-$NH_2$ | 3022.33 | 3022.94 |
| AP-PNA 6-3 | Ac-**G**$^{6a}$ TAG**A**$^{6a}$ T**C**$^{6a}$ ACT-Gly-$NH_2$ | 3121.44 | 3121.89 |
| AP-PNA 6-4 | Ac-G**T**$^{6a}$A**G**$^{6a}$A**T**$^{6a}$C**A**$^{6a}$C T-Gly-$NH_2$ | 3220.54 | 3221.34 |
| AP-PNA 6-5 | Ac-G**T**$^{6a}$A**G**$^{6a}$A**T**$^{6a}$C**A**$^{6a}$C **T**$^{6a}$ -Gly-$NH_2$ | 3319.65 | 3321.71 |

| | | | |
|---|---|---|---|
| AegPNA-$T_{10}$ | Ac- TTTTTTTTTT-Gly-$NH_2$ | 2777.07 | 2778.70 |
| AP-PNA-$T_{10}$ 2-1 | Ac- TTTTT$T^{2a}$TTTT-Gly-$NH_2$ | 2820.12 | 2821.97 |
| AP-PNA-$T_{10}$ 2-3 | Ac- T$T^{2a}$TTT$T^{2a}$TTT$T^{2a}$-Gly-$NH_2$ | 2906.20 | 2907.48 |
| AP-PNA-$T_{10}$ 2-5 | Ac- T$T^{2a}$T$T^{2a}$T$T^{2a}$T$T^{2a}$T$T^{2a}$-Gly-$NH_2$ | 2992.28 | 2993.32 |
| AP-PNA-$T_{10}$ 6-1 | Ac- TTTTT$T^{6a}$TTTT-Gly-$NH_2$ | 2876.18 | 2877.68 |
| AP-PNA-$T_{10}$ 6-3 | Ac- T$T^{6a}$TTT$T^{6a}$TTT$T^{6a}$-Gly-$NH_2$ | 3074.39 | 3074.94 |
| AP-PNA-$T_{10}$ 6-5 | Ac- T$T^{6a}$T$T^{6a}$T$T^{6a}$T$T^{6a}$T$T^{6a}$-Gly-$NH_2$ | 3272.60 | 3273.09 |

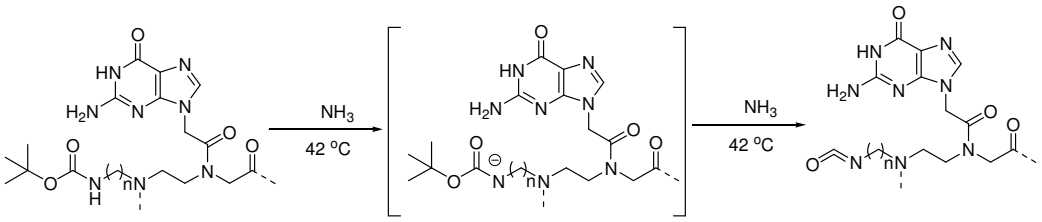
Stephan and co-workers[24] have found that in the PyBOP mediated synthesis of PNA oligomers, the obtained products showed M + 67n peaks of molecular masses in the mass spectra. And this was also observed in our case, especially when the Bhoc protected guanine was present. The proposed mechanism is that the protected guanine undergoes a keto-enol conversion under alkaline condition (Scheme 2-8). The enol form is activated by the excess of PyBOP on the reaction solution which is further substituted by the good nucleophile piperidine during Fmoc deprotection to get the 6-piperidine-guanine by-product. The presence of these by-products makes purification very difficult. Thus, a guanine derivative, such as $N^2$-isobutyryl-7-ethoxycarbonylmethylguanine, was added as a scavenger to the reaction mixture and meanwhile using less amount of PyBOP was effective to suppress this side reaction. It is most likely that the added guanine derivative competes with guanine residues to react with PyBOP.



**Scheme 2-8. Proposed rationale for the PyBOP induced side reaction.** [24]

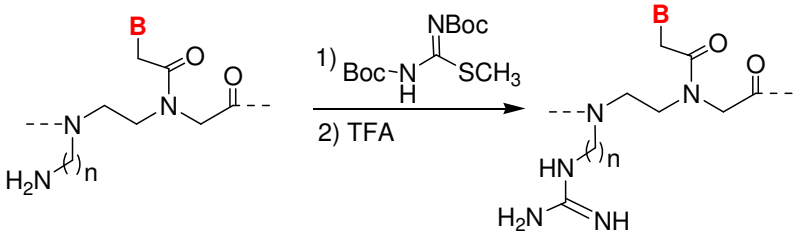
There is another side reaction during removal of the acyl protecting group. After capping the N-terminal amine, the dried resin was transferred to a 1.5 mL eppendorf tube containing 0.5 mL 35% ammonia, which was sealed and kept at 42 $^{o}$C overnight. TFA cleavage of the ammonia-treated resin revealed side products which showed M + 26n mass peaks besides the expected mass in the mass spectra. Usually, the Boc protecting group is strongly resistant toward many nucleophilic reagents. However, many groups[25, 26] have reported on the conversion of Boc-protected amines into ureas via the formation of isocyanate intermediate in the presence of strong base. Therefore we proposed the following mechanism for this side reaction (Scheme 2-9). The N-H protons of the Boc amino are acidic. The N-H protons undertook deprotonation and attacked the carbonyl carbon followed by the expulsion of a $tBu^{-}$ group from the Boc group under alkaline at high temperature. The so-formed isocyanate by-product was confirmed by MALDI-TOF mass which showed M+26n peaks. This was also confirmed indirectly by the experiment that deprotecting the acyl protecting group by 35% ammonia at 42 $^{o}$C overnight after TFA cleavage, i.e. in the absence of Boc

protecting group on the peptoid side chain amine, or treating the resin by 35% ammonia at room temperature overnight followed by TFA cleavage, no M + 26n peaks appeared on the mass spectra of the crude products in either case.




**Scheme 2-9. Proposed mechanism for ammonium induced side reaction.**

## 2.5 Converting AP-PNA to GP-PNA in solution.




**Scheme 2-10. Converting the peptoid amine to guanidine to give a guanidino-peptoid PNA (GP-PNA)**

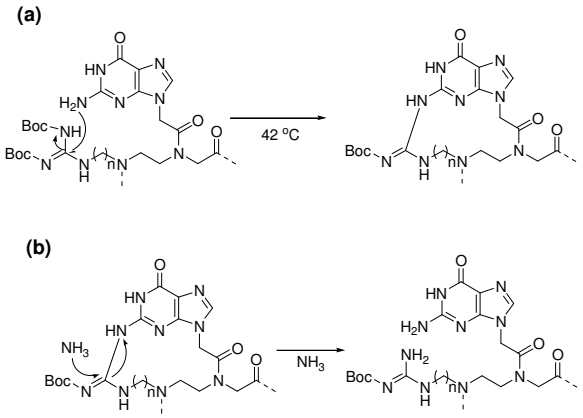
With AP-PNA in hand, one can convert the amino groups on the peptoid side chains to guanidino groups using guanylation reagent to yield the corresponding GP-PNAs. Therefore the pure and dried AP-PNA oligomer (1-2 mg) and N,N'-bis-Boc-S-methylisothiourea (10 eq per amino group) were dissolved in 50 µL DMF in a 1.5 mL Eppondorf tube. Triethylamine (2 eq per amino group) was added to the clear solution. Some precipitate came out as the TEA·TFA salt. Then the mixture was kept at 42 °C overnight. DMF was removed under high

vacuum and the residue was then washed with hexane to remove the excess N,N'-bis-Boc-S-methylisothiourea. If necessary, the product was purified by HPLC at this stage. The obtained N,N'-bis-Boc-protected guanidine product was then treated with 95% TFA (2.5%$H_2O$ and 2.5% TIS) at room temperature for 1 h. The deprotected mixture was diluted with $H_2O$ and purified directly by semi preparative reverse-phase HPLC. The product was then lyophilized to give a white solid. The expected product was detected by MALDI-TOF mass spectroscopy.

**Table 2-2**.MALDI-TOF mass spectral analysis of synthesized GP-PNA oligomers.

| | Sequence | $M_w$ calcd (isotopic) | $m/z$ $[M+H]^+$ found |
|---|---|---|---|
| GP-PNA 2-1 | Ac-GTAGA**T$^{2g}$**CACT-Gly-$NH_2$ | 2909.19 | 2909.65 |
| GP-PNA 2-2 | Ac-GTAG**A$^{2g}$** T**C$^{2g}$**ACT-Gly-$NH_2$ | 2994.25 | 2995.21 |
| GP-PNA 2-3 | Ac-**G$^{2g}$**TAG**A$^{2g}$**T**C$^{2g}$**ACT-Gly-$NH_2$ | 3079.31 | 3079.96 |
| GP-PNA 2-4 | Ac-G**T$^{2g}$**A**G$^{2g}$**A**T$^{2g}$**C**A$^{2g}$**C T-Gly-$NH_2$ | 3164.38 | 3165.61 |
| GP-PNA 3-1 | Ac-GTAGA**T$^{3g}$**CACT-Gly-$NH_2$ | 2923.2 | 2924.14 |
| GP-PNA 3-2 | Ac-GTAG**A$^{3g}$**T**C$^{3g}$**ACT-Gly-$NH_2$ | 3022.28 | 3023.45 |
| GP-PNA 3-3 | Ac-**G$^{3g}$**TAG**A$^{3g}$**T**C$^{3g}$**ACT-Gly-$NH_2$ | 3121.36 | 3122.46 |
| GP-PNA 3-4 | Ac-G**T$^{3g}$**A**G$^{3g}$**A**T$^{3g}$**C**A$^{3g}$**C T-Gly-$NH_2$ | 3220.44 | 3221.35 |
| GP-PNA 4-1 | Ac-GTAGA**T$^{4g}$**CACT-Gly-$NH_2$ | 2937.22 | 2937.68 |
| GP-PNA 4-2 | Ac-GTAG**A$^{4g}$**T**C$^{4g}$**ACT-Gly-$NH_2$ | 3050.31 | 3051.30 |
| GP-PNA 4-3 | Ac-**G$^{4g}$**TAG**A$^{4g}$**T**C$^{4g}$**ACT-Gly-$NH_2$ | 3163.41 | 3164.10 |
| GP-PNA 4-4 | Ac-G**T$^{4g}$**A**G$^{4g}$**A**T$^{4g}$**C**A$^{4g}$**C T-Gly-$NH_2$ | 3276.5 | 3277.11 |

| | | | |
|---|---|---|---|
| GP-PNA 5-1 | Ac-GTAGA$T^{5g}$CACT-Gly-$NH_2$ | 2951.23 | 2951.87 |
| GP-PNA 6-1 | Ac-GTAGA$T^{6g}$CACT-Gly-$NH_2$ | 2965.25 | 2966.06 |
| GP-PNA 6-2 | Ac-GTAG$A^{6g}$T$C^{6g}$ACT-Gly-$NH_2$ | 3106.38 | 3106.97 |
| GP-PNA 6-3 | Ac-$G^{6g}$TAG$A^{6g}$T$C^{6g}$ ACT-Gly-$NH_2$ | 3247.15 | 3248.75 |
| GP-PNA 6-4 | Ac-G$T^{6g}$A$G^{6g}$A$T^{6g}$C$A^{6g}$C T-Gly-$NH_2$ | 3388.63 | 3389.23 |
| GP-PNA 6-5 | Ac-G$T^{6g}$A$G^{6g}$A$T^{6g}$C$A^{6g}$C $T^{6g}$-Gly-$NH_2$ | 3529.75 | 3530.71 |
| GP-PNA-$T_{10}$ 2-1 | Ac- TTTTT$T^{2g}$TTTT-Gly-$NH_2$ | 2862.14 | 2863.29 |
| GP-PNA-$T_{10}$ 2-3 | Ac- T$T^{2g}$TTT$T^{2g}$TTT$T^{2g}$-Gly-$NH_2$ | 3032.27 | 3033.27 |
| GP-PNA-$T_{10}$ 2-5 | Ac- T$T^{2g}$T$T^{2g}$T$T^{2g}$T$T^{2g}$T$T^{2g}$-Gly-$NH_2$ | 3203.39 | 3203.04 |
| GP-PNA-$T_{10}$ 6-1 | Ac- TTTTT$T^{6g}$TTTT-Gly-$NH_2$ | 2918.2 | 2919.27 |
| GP-PNA-$T_{10}$ 6-3 | Ac- T$T^{6g}$TTT$T^{6g}$TTT$T^{6g}$-Gly-$NH_2$ | 3200.45 | 3201.08 |
| GP-PNA-$T_{10}$ 6-5 | Ac- T$T^{6g}$T$T^{6g}$T$T^{6g}$T$T^{6g}$T$T^{6g}$-Gly-$NH_2$ | 3482.71 | 3483.05 |




**Scheme 2-11. (a) Proposed mechanism for the Base nucleic acid induced side reactions; (b) Conversion of the by-product back to the guanidino product by treatment of 35 % ammonia to confirm the proposed mechanism.**

Although converting AP-PNA to GP-PNA in solution is convenient, the yields are very low especially when there are more than three amino groups. The major

problem is incomplete conversion. When reacting for longer time, such as 2 days at 42 $^{o}$C, some by-products were observed. For example, the free amines were converted to their Boc protected forms rather than the bis-Boc protected guanidines. And even if all were converted to the corresponding GP-PNAs, some M-16n by-products were also observed on the MALDI-TOF mass spectra. We proposed a mechanism for such side reactions (shown in Scheme 2-11). Under high temperature at alkaline condition (pH ≈ 9), the free exocyclic amino groups of the base nucleic acid were nucelophile enough to attack the imino carbon displacing the -NH(Boc) group. The observed mass peaks were M-16n (the Boc group is sensitive to the MALDI condition, therefore can not be seen at the MALDI-TOF mass spectrum). When these by-products were treated with 35% ammonia at room temperature, some of the M-16n peak disappeared and the M peak showed up. This observation indirectly confirmed our hypothesis.

## 2.6 Modifying the peptoid amine of AP-PNA oligomers by acylation on solid phase.

The amino groups on the peptoid side chains can be modified in various ways using an acylation reaction. After the terminal amine was capped by acetylation, the resin was washed with DMF and DCM, and dried thoroughly *in vacuo*. The dried resin was treated with a mixture of $BF_3$·Ether 2.5%, $CH_3COOH$ 5% in dry DCM (200 μL × 3) for a total time of 2 min to remove the side-chain Boc

protecting group.[27] The deprotection reaction was checked by the Kaiser Test. After washing with DCM and DMF, the newly exposed free amine was then reacted with either acetic anhydride/DIEA for 20 min or with β-alanine or biotin for 4-5 h. The β-alanine (Boc-β-Ala-OH) or biotin (D-biotin) was activated by PyBOP in the presence of DIEA as described previously in the coupling Method A. The reaction was monitored by Kaiser test. The subsequent acyl deprotection, cleavage, purification and characterization experiments were performed using the same method as in AP-PNA oligomer synthesis.

**Table 2-3. MALDI-TOF mass spectral analysis of synthesized AP-PNA derivatives.**

| | Sequence | $M_w$ calcd | $m/z$ $[M+H]^+$ found |
|---|---|---|---|
| AP-PNA 2-1(βAla) | Ac-GTAGA$T^{2a(\beta Ala)}$CACT-Gly-$NH_2$ | 2938.2 | 2938.44 |
| AP-PNA 3-1(βAla) | Ac-GTAGA$T^{3a(\beta Ala)}$CACT-Gly-$NH_2$ | 2952.22 | 2952.9 |
| AP-PNA 4-1(βAla) | Ac-GTAGA$T^{4a(\beta Ala)}$CACT-Gly-$NH_2$ | 2966.23 | 2967.42 |
| AP-PNA 6-1(Ac) | Ac-GTAGA$T^{6a(Ac)}$CACT-Gly-$NH_2$ | 2965.24 | 2966.10 |
| AP-PNA 6-1(Biotin) | Ac-GTAGA$T^{6a(Biotin)}$CACT-Gly-$NH_2$ | 3149.30 | 3151.25 |
| AP-PNA 6-1(βAla) | Ac-GTAGA$T^{6a(\beta Ala)}$CACT-Gly-$NH_2$ | 2994.26 | 2995.05 |

# 2.7 Experiments

## 2.7.1 Materials and Methods

### 2.7.1.1 Materials

All solvents used were of analytical grade. Pyridine, DMF and DCM were

dried using anhydrous $Na_2SO_4$ for one to two days, distilled and stored over molecular sieves prior to usage. NaH was washed with dry hexane and dried *in vacuo* prior to use. Guanine was purified by dissolving it in 0.5 M NaOH, and subsequently adjusting the pH of the solution to 10 using 1 M HCl. The white precipitate formed was dried and grounded to fine powder prior to usage. All other chemicals from Aldrich, Merk and Lancaster were used without further purification.

### 2.7.1.2 Chromatography

Thin layer chromatography (TLC) was carried out on aluminum sheets coated with silica [silica 60 F254; layer thickness 0.2 mm (Merk)]. Analytes were visualized under UV light ($\lambda$ = 254 or 365 nm). Flash chromatography was carried out using commercial silica column (Redisep) and preparative column chromatography was carried out using silica gel 60 (Lancaster, 0.040-0.063 mm).

### 2.7.1.3 HPLC

Analytical HPLC was performed on a DIONEX system equipped with a PDA-100 photodiode array detector and a C18 reversed-phase column (Agilent, 4.6×150 mm). Semi-preparative HPLC was carried out on a SHIMADZU system with a C18 semi-preparative column (GRACE). The analytes were eluted using a gradient mixture of two solvents: Solvent A was distilled deionized water containing 0.045% TFA and solvent B was 90% ACN in distilled deionized water

containing 0.045% TFA. The mobile phase flow rate was 1 mL/min for analytical HPLC and 2.5 mL/min for semi-preparative HPLC, respectively. The separation temperature was 23 ℃.

### 2.7.1.4 Mass Spectrometry

ESI mass spectra were recorded using a bench-top ion trap mass spectrometer (FINNIGAN LCQ Deca XP MAX) equipped with standard ESI sources. Samples were introduced by direct injection using a syringe. The carrier buffer was composed of two solvents in 1:1 v/v ratio: solvent A (dd$H_2O$ containing 0.045%TFA) and solvent B (90% ACN in dd$H_2O$ containing 0.04% TFA), and delivered by a LC pump. Nitrogen served both as the nebulizer gas and the dry gas. Nitrogen was generated by a PEAK SCIENTIFIC lab gas generator. Helium served as the damping gas for the ion trap and collision gas for $MS^n$ experiments. Data analysis software Xcalibur 1.4 was used for processing the spectra.

MALDI-TOF mass spectra of PNA oligomers were recorded on a 4800 MALDI TOF/TOF Analyzer operating in MS reflector positive ion mode and using α-cyano-4-hydroxycinnamic acid as the matrix.

### 2.7.1.5 NMR Spectroscopy

The $^1H$ and $^{13}C$ NMR spectra were recorded on a Bruker DRX-400 (400 MHz) NMR spectrophotometer. $^{13}C$ spectra were recorded in broad band decoupled mode. $^1H$ and $^{13}C$ chemical shifts are quoted in ppm relative to the internal standard

tetramethylsilane (TMS) or the residual solvent proton signals. 2D NMR HMBC spectra were recorded with a spectral width of 15 ppm in $F_2$ and with 4096 data points (acquisition time 228 ms). The HMBC experiments were run with standard pulse programs built-in NMR spectrophotometer (without low-pas $J$ filter, relaxation delay D1 = 1.6 s, evolution delay $^{n}J_{CH}$ $D_6$ = 80 ms, sweep width $F_1$ 200 ppm, acquisitions 4, and increments 512). The spectra were apodizated with a squared π/2-shift sine bell in $F_1$ and π/4-shift sine bell in $F_2$.

## 2.7.2 Abbreviations

ACN – Acetonitrile

Bhoc – Benzhydryloxycarbonyl

Boc – tert-Butyloxycarbonyl

Bz – Benzoyl

Cbz – Carbobenzyloxy

DCM – Dichloromethane

Dde – N-(1-(4,4-dimethyl-2,6-dioxocyclohexylidene)ethyl)

DIEA – N,N-Diisopropylethylamine

DMF – N,N-Dimethylformamide

DMSO – Dimethyl sulphoxide

ESI MS – Electro Spray Ionization Mass Spectrometry

EtOH – Ethanol

Fmoc – Fluorenylmethoxycarbonyloxy

Fmoc-OSu – N-(9-Fluorenylmethoxycarbonyloxy) succinimide

HATU – O-(Benzotriazol-1-yl)-N,N,N',N'-tetramethyluronium

hexafluorophosphate

HMBC – Heteronuclear Multiple Bond Correlation

iBu – isoButyryl

INEPT – Insensitive Nuclei Enhanced by Polarization Transfer

Mmt – Monomethoxytrityl

$NaBH_3CN$ – Sodium cyanoborohydride

NMR – Nuclear Magnetic Resonance

NVOC – Nitroveratryloxycarbonyl

PyBOP – Benzotriazole-1-yl-oxy-tris-pyrrolidino-phosphonium hexafluorophosphate

TEA – Triethylamine

TFA – Trifluoroacetic acid

TFMSA –Trifluoromethanesulfonic acid

THF – Tetrahydrofuran

TIS – Triisopropylsilane

## 2.7.3 Experimental section

**4-*N*-Benzoylcytosine 1.** Benzoyl chloride (2.3 mL, 20 mmol) was added dropwise over 30 min to a stirred suspension of cytosine (1.1 g, 10 mmol) in dry pyridine, and stirring was continued at room temperature for another 1 h at room temperature, after which the pH was adjusted to 5 using 1 M HCl. The precipitate formed was washed with EtOH and dried *in vacuo* to give **1** (1.92 g, yield 90%).

**Ethyl (4-*N*-benzoylcytosine)-1-acetate 2.** NaH (0.43 g, 18 mmol) was added

to a stirred suspension of **1** (1.92 g, 9 mmol) in dry DMF (100 mL) at 0 °C and the mixture was stirred at room temperature for 1 h. Ethyl bromoacetate (1.5 g, 9 mmol) was added dropwise at room temperature and stirred for 1 h. Methanol (2 mL) was added to quench the reaction, and the solvents were removed under reduced pressure. The residue obtained was dissolved in DCM, washed with water (3 × 50 mL), dried over $Na_2SO_4$, and evaporated to give **2** as a white solid: 1.38 g, yield 54%.

**4-*N*-Benzoylcytosine-1-acetic acid 3.** Compound **2** (1.38 g, 4.6 mmol) and NaOH (0.36 g, 9.2 mmol) were dissolved in water (10 mL) and the solution was stirred at room temperature for 2 h. Then the pH was adjusted to 3 using 1 M HCl. The precipitate formed was separated by filtration, washed with 5 mL water and dried *in vacuo* to afford **3** as a white solid. Yield: 1.16 g (92%); $^1$H NMR (400 MHz, $D_2O$, sodium salt): δ 7.78 (d, 1H), 7.75 (d, 2H), 7.55 (t, 1H), 7.39 (t, 2H), 7.19 (d, 1H), 4.31 (s, 2H); ESI-MS m/z $[M+H]^+$: 274.

**Ethyl cytosine-1-acetate 4.** To a suspension of cytosine (4.44 g, 40 mmol) in dry DMF (100 mL) was added NaH (1.008 g, 42 mmol) in two portions for a period of 30 min at 0 °C and the mixture was stirred at 0 °C for 2 h. Ethyl bromoacetate (4.65 mL, 42 mmol) was added to the rapidly stirring solution for a period of 1 h at 0 °C. After addition, the reaction was stirred at room temperature overnight. The reaction was quenched by adding methanol and the mixture was neutralized by pouring into ice containing HCl (45 mmol). The mixture was concentrated and the precipitated solid was collected by filtration, which was

washed with water and dried under vacuum to get 6.4 g of ethyl cytosine-1-acetate, yield: 81.2%. $^{1}$H NMR (400 MHz, $d_6$-DMSO): δ 8.31 (bs, 1H), 7.73 (d, 1H), 5.87 (d, 1H), 4.52 (s, 2H), 4.12 (q, 2H), 1.18 (t, 3H). $^{13}$C NMR (400 MHz, $d_6$-DMSO): δ 168.91, 163.51, 152.43, 128.89, 94.05, 61.62, 50.22, 14.49; ESI-MS m/z $[M+H]^+$: 274.

**Ethyl [4-*N*-(benzhydryloxycarbonyl)cytosine]-1-acetate 5.** To **4** (4.3 g, 21.8 mmol) dissolved in dry DMF (80 mL) was added N,N'-carbonyldiimidazole (5.7 g, 35 mmol). The mixture was stirred at room temperature for 2 h and benzhydrol (5.2 g, 28 mmol) was added at 60 °C and the temperature was maintained at 60 °C for about 6 h. Then two more batches of benzhydrol (520 mg) were added at 1 h intervals. The reaction mixture was then allowed to stir overnight while cooling to room temperature. Methanol was added to quench the reaction and the solvents were removed *in vacuo*. The resulting residue was separated with water and DCM. Water layer was re-extracted with DCM (50 mL × 3), the combined organic layer was washed with water (100 mL × 2) and brine (100 mL), respectively, dried with anhydrous $Na_2SO_4$ and removed *in vacuo*. The desired product was separated by flash chromatography (Hexane/ Ethyl acetate: 1/1) to get 6.8 g of **5** as a white solid, yield 78%. $^{1}$H NMR (400 MHz, $d_6$-DMSO): δ 7.83 (br, 1H), 7.50 (d, 1H), 7.21-7.34 (m, 10H), 7.10 (d, 1H), 6.83 (s, 1H), 4.58 (s, 2H), 4.23 (q, 2H), 1.27 (t, 3H). $^{13}$C NMR (400 MHz, DMSO): δ 167.26, 162.80, 155.37, 151.54, 148.78, 139.14, 128.68, 128.34, 127.03, 95.24, 79.27, 62.19, 50.79, 14.07; ESI-MS m/z $[M+Na]^+$: 430.0.

**4-*N*-(Benzhydryloxycarbonyl)cytosine-1-acetic acid 6.** To a solution of **5** (6.8 g, 18 mmol ) dissolved in THF (20 mL) was added NaOH (1.4 g, 36 mmol) in water (20 mL). The progression of the reaction was monitored by TLC and when the starting material was all consummated, concentrated HCl was carefully added to adjust the pH to 4~5 at 0 $^{o}$C. THF was removed *in vacuo* and the solid precipitated was filtered, washed with water and dried to get 5.2 g white solid, yield (76%). $^{1}$H NMR (400 MHz, $d_6$-DMSO): $\delta$ 10.96 (s, 1H), 8.00 (d, 1H), 7.25-7.44 (m, 10H), 6.92 (d, 1H), 6.77 (s, 1H), 4.49 (s, 2H). $^{13}$C NMR (400 MHz, DMSO): $\delta$ 167.26, 162.80, 155.37, 151.54, 148.78, 139.14, 128.68, 128.34, 127.03, 95.24, 79.27, 50.79; ESI-MS m/z $[M+Na]^+$: 402.06.

**6-*N*-Benzoyladenine 7.** Benzoyl chloride (1.3 mL, 11 mmol) was added dropwise over 30 min to a stirred suspension of adenine (1.35 g, 10 mmol) in dry pyridine, and stirring was continued at 100 $^{o}$C for a further 3 h, and allowed to stand overnight at room temperature. The reaction was quenched with methanol and the solvents were removed under reduced pressure. The residue was triturated in hot isopropanol and dried *in vacuo* to give **7** as a white solid: yield 2.15 g (90%); $^{1}$H NMR (400 MHz, $d_6$-DMSO): $\delta$ 11.50 (s, 1H), 8.74 (s, 1H), 8.52 (s, 1H), 8.11 (d, 2H), 7.66 (t, 1H), 7.58 (t, 2H); ESI-MS m/z $[M+H]^+$: 240.

**Ethyl (6-*N*-benzoyladenine)-9-acetate 8.** Compound **7** (2.39 g, 10 mmol) was dissolved in dry DMF (100 mL), cooled to 0 $^{o}$C, sodium hydride (0.48 g, 20 mmol) was added, and the mixture stirred at room temperature for 30 min. Ethyl bromoacetate (1.22 mL, 11 mmol) was added dropwise through a syringe. Stirring

was continued for 2 h at room temperature and the reaction was quenched with methanol (2 mL). The solvents were evaporated under reduced pressure. The residue was subjected to column chromatography (DCM/MeOH; 9.5/0.5) to obtain **8** as a white solid: yield 1.3 g (40%); $^{1}$H NMR (400 MHz, $CDCl_3$): δ 9.02 (s, 1H), 8.83 (s, 1H), 8.11 (s, 1H), 8.05 (d, 2H), 7.62 (t, 1H), 7.54 (t, 2H), 5.06 (s, 2H), 4.31 (q, 2H), 1.38 (t, 3H); $^{13}$C NMR (400 MHz, $CDCl_3$): δ 166.76, 164.53, 152.96, 152,50, 149.83, 143.22, 133.68, 132.80, 128.89, 127.85, 122.56, 62.53, 44.33, 14.09; ESI-MS m/z $[M+H]^+$ : 326.

**6-*N*-benzoyladenine-9-acetic acid 9.** Compound **8** (1.63 g, 5 mmol) and NaOH (0.4 g, 10 mmol) were dissolved in water (20 mL) and the solution was stirred for 2 h at room temperature. The pH was adjusted to 3 using 1 M HCl, causing precipitation of the product. The precipitate was separated by filtration, washed with water, and dried *in vacuo* to give **9** as a white solid: yield 1.4 g (95%); $^{1}$H NMR (400 MHz, $d_6$-DMSO): δ 11.682 (s, 1H), 9.11 (s, 1H), 8.81 (s, 1H), 8.52 (s, 1H), 8.09 (d, 2H), 7.67 (t, 1H), 7.60 (t, 2H), 5.16 (s, 2H); $^{13}$C NMR (400 MHz, $CDCl_3$): δ 169.43, 166.34, 152.89, 152,09, 150.29, 145.63, 133.74, 132.98, 128.45, 122.56, 44.88; ESI-MS m/z $[M+H]^+$: 298.

**Benzyl adenine-9-acetate 10.** To a suspension of adenine (4 g, 29.6 mmol) in dry DMF (70 mL) was added NaH (1.2 g, 50 mmol) in two portions for a period of 30 min at 0 $^{o}$C and the mixture stirred at 0 $^{o}$C for 2 h. To the rapidly stirring solution was added benzyl bromoacetate (5.1 mL, 32.6 mmol) dropwise for a period of 1 h at 0 $^{o}$C. After addition the reaction was allowed to stir overnight

while warming to room temperature. The reaction was quenched by adding methanol. The mixture was filtered and the solid was washed with water. The combined filtration was concentrated *in vacuo*. Water was added to the residue and the solid filtered. The combined solid was dried under vacuum to get 7.36 g of light yellow solid, yield (76%). $^{1}$H NMR (400 MHz, $d_6$-DMSO): $\delta$ 8.14 (s, 1H), 7.33-7.41 (m, 5H), 7.26 (s, 1H), 5.20 (s, 2H), 5.15 (s, 2H). $^{13}$C NMR (400 MHz, $d_6$-DMSO): $\delta$ 168.38, 156.44, 153,13, 150.19, 141.692, 135.90, 128.93, 128.72, 128.49, 119.87, 67.10, 44.43; ESI-MS m/z $[M+H]^+$: 284.28.

**Benzyl [6-*N*-(benzhydryloxycarbonyl)adenine]-9-acetate 11.** To a suspension of **10** (2.27 g, 8 mmol) in dry DMF (20 mL) was added N,N'-carbonyldiimidazole (2.08 g, 12.8 mmol). The mixture was stirred at 60 $^{o}$C for 4 h during which time the solution became clear and benzhydrol (1.92 g. 10.4 mmol) was added at 60 $^{o}$C and the temperature was maintained at 60 $^{o}$C for about 6 h. Then two more batches of benzhydrol (192 mg) were added at 1 h intervals. The reaction mixture was then allowed to stir overnight while cooling to room temperature. Methanol was added to quench the reaction and the solvents were removed *in vacuo*. The resulting residue was separated with water and DCM. Water layer was re-extracted with DCM (50 mL $\times$ 3), the combined organic layer was washed with water (100 mL $\times$ 2) and brine (100 mL), respectively, dried with anhydrous $Na_2SO_4$ and removed *in vacuo*. The desired product was separated by flash chromatography (Hexane/ Ethyl acetate: 1/1) to get 1.7 g as an oil, yield 45%. $^{1}$H NMR (400 MHz, $CDCl_3$): $\delta$ 9.75 (s, 1H), 8.72 (s, 1H), 7.94 (s, 1H),

7.23-7.39 (m, 15H), 7.07 (s, 1H), 5.15 (s, 2H), 4.90 (s, 2H). $^{13}$C NMR (400 MHz, $CDCl_3$): δ 166.69, 152.97, 151.60, 150.61, 149.67, 143.46, 139.66, 135.90, 128.84, 128.79, 128.70, 128.52, 128.06, 127.34, 121.43, 78.57, 67.97, 44,23; ESI-MS m/z $[M+Na]^+$: 516.19.

**6-*N*-(Benzhydryloxycarbonyl)adenine-9-acetic acid 12.** To a solution of **11** (1.7 g, 3.4 mmol) dissolved in THF (10 mL) was added LiOH (826 mg, 34 mmol) in water (10 mL). The progression of the reaction was monitored by TLC and when the start material was all consummated, HCl was carefully added to adjust the pH to 4~5 at ice-bath. THF was removed *in vacuo* and the solid precipitated out was filtered, washed with water and dried to get 986 mg white solid, yield 72%. $^1$H NMR (400 MHz, $d_6$-DMSO): δ 10.89 (s, 1H), 8.59 (s, 1H), 8.41 (s, 1H), 7.28-7.51 (m, 10H), 6.79 (s, 1H), 5.06 (s, 2H). $^{13}$C NMR (400 MHz, $d_6$-DMSO): δ 169.50, 152.97, 152.15, 151.57, 149.81, 145.30, 141.34, 128.94, 128.16, 126.96, 122.84, 77.74, 44,73; ESI-MS m/z $[M+H]^+$: 403.94.

**2-*N*-isobutyrylguanine 13.** Isobutyryl chloride (2.02 mL, 19.9 mmol) was added dropwise using a syringe to a stirred suspension of guanine (3.02 g, 20 mmol) and TEA (1.45 mL, 10.4 mmol) in anhydrous pyridine (40 mL). The mixture was stirred for 3 h at 100 ℃. The reaction was quenched using methanol, and the solvents were evaporated to dryness. The residue obtained was triturated in hot isopropanol, and dried *in vacuo* to give **13** (2.67 g, yield 60%); ESI-MS m/z $[M+H]^+$: 222.

**Ethyl (2-*N*-isobutyrylguanine)-9-acetate 14.** Compound **13** (2.21 g, 10

mmol) and $Na_2CO_3$ (2.12 g, 20 mmol) were suspended in dry DMF, and the mixture was warmed to 60 $^oC$ with stirring. After 30 min the heating was terminated and ethyl bromoacetate (1.22 mL, 11 mmol) was added dropwise. The stirring was continued for an additional 2 h. The reaction mixture was filtered and the solid was washed with DCM. The organic solvents were removed under reduced pressure. The residue was subjected to column chromatography ((DCM/MeOH; 9.5/0.5) to obtain **14** as a white solid: yield 0.77 g (25%); MS(+ESI) 308 $[M+H]^+$; $^1H$ NMR (400 MHz, $CDCl_3$): δ 11.99 (s, 1H), 8.50(s, 1H), 7.71(s, 1H), 4.80 (s, 2H), 4.26 (q, 2H), 2.70 (septet, 1H), 1.31 (t, 3H), 1.28 (d, 6H); $^{13}C$ NMR (400 MHz, $CDCl_3$): δ 178.3, 166.9, 155.4, 148.5, 147.6, 139.2, 120.8, 62.3, 44.5, 36.5, 18.9, 14.1.

**2-*N*-isobutyrylguanine-9-acetic acid 15.** Compound **14** (3.07 g, 10 mmol) and NaOH (0.8 g, 20 mmol) were dissolved in water (20 mL). The mixture was stirred at room temperature for 2 h. The pH was adjusted to 3 using 1 M HCl, to obtain a white precipitate. The solution was filtered off, and the precipitate was washed with a small amount of cold water and dried *in vacuo* to give 1**5** as a white solid 2.65 g, yield 95%; $^1H$ NMR (400 MHz, $d_6$-DMSO): δ 13.32 (s, 1H), 12.06 (s, 1H), 11.64 (s, 1H), 7.93 (s, 1H), 4.88 (s, 2H), 2.74 (septet, 1H), 1.10 (d, 6H); $^{13}C$ NMR (400 MHz, $CDCl_3$): δ 180.63, 169.52, 155.29, 149.47, 148.52, 140.78, 120.17, 44.88, 39.38,19.33; ESI-MS m/z $[M+H]^+$: 280.

**Boc-alkylenediamines 16.** Di-*tert*-butyl dicarbonate ($Boc_2O$) (12.3 g, 0.056 mol) dissolved in 100 mL DCM was added dropwisely to a solution of

alkylenediamine (0.28 mol) in 150 mL DCM during 4 hours with vigorous stirring at 0 $^{o}$C. The reaction was stirred for additional 20 hours at room temperature and was then washed with water (4 × 100 mL). The organic phase was dried over $Na_2SO_4$, evaporated to dryness under reduced pressure to give as colorless oil. The crude product **1** was used for the next step without purification.

**Synthesis of products 18 a-e.** Glycidol (4.85 mL, 0.073 mol) in THF was added to the solution of **16** (0.066 mol) in THF dropwisely at room temperature. The reaction was stirred at room temperature for 20 h. THF was removed under reduced pressure and the solid precipitated out was filtered and washed with DCM. The filtration was concentrated and used at the next step without any purification. Fmoc-OSu (16.8 g, 0.05 mol) was added portionwise to a solution of **17** in DCM at 0 $^{o}$C for a period of 3 h. TEA (7 mL, 0.05mol) was added during Fmoc-OSu addition. The mixture was stirred at 0 $^{o}$C for 2 h and then at room temperature overnight. The reaction mixture was concentrated under reduced pressure. The residue was extracted by DCM (100 mL) / $H_2O$ (100 mL). The combined organic phase was washed with brine and dried over sodium sulfate. It was purified by flash column chromatography. The product was obtained as colorless oil. The products were characterized by ESI-MS and NMR data.

**18a** (n = 2): yield 41%; $^{1}$H NMR ($CDCl_3$): δ 7.68 (d, 2H); 7.48 (d, 2H), 7.32 (t, 2H), 7.25 (t, 2H), 5.21 (s, 1H), 4.52 (d, 2H), 4.14 (m, 1H), 4.04 (m, 1H), 3.66 (m, 1H), 3.4 (m, 1H), 3.31-2.78 (m, 6H), 1.96 (d, 2H), 1.33 (s, 9H); $^{13}$C NMR ($CDCl_3$): δ 143.9, 141.4, 127.8, 127.3, 124.6, 120.0, 79.8, 70.6, 67.0, 64.1, 51.5,

49.3, 47.4, 39.0, 28.4; ESI-MS m/z $[M+Na]^+$: 479.43.

**18b** (n = 3): yield 46% ; $^1H$ NMR ($CDCl_3$): δ 7.67 (d, 2H), 7.47 (d, 2H), 7.32 (t, 2H), 7.24 (t, 2H), 5.19 (s, 1H), 4.52 (d, 2H), 4.1 (m, 1H), 4.0 (m, 1H), 3.6 (m, 1H), 3.37 (m, 1H), 3.29-2.67 (m, 6 H), 1.94 (s, 1H), 1.87 (s, 1H), 1.34 (s, 9H), 1.2 (m, 2H); $^{13}C$ NMR ($CDCl_3$): δ 157.5, 155.4, 143.8, 141.4, 127.7, 127.2, 124.6, 119.9, 79.8, 70.7, 66.5, 64.2, 53.5, 50.4, 47.4, 38.0, 30.0, 28; ESI-MS m/z $[M+Na]^+$:493.12.

**18c** (n = 4): yield 47% ; $^1H$ NMR ($CDCl_3$): δ 7.67 (d, 2H), 7.48 (d, 2H), 7.30 (t, 2H), 7.24 (t, 2H), 4.54 (d, 2H), 4.45 (t, 1H), 4.12 (t, 1H), 3.62 (m, 1H), 3.4 (m, 1H), 3.33-2.87 (m, 6H), 1.37 (s, 9H), 1.17-1.07 (m, 4H); $^{13}C$ NMR ($CDCl_3$): δ 157.5, 155.4, 143.9, 141.5, 127.7, 127.2, 124.6, 120.0, 79.4, 70.8, 66.8, 63.4, 49.9, 48.3, 47.4, 39.9, 28.4, 27.1, 25.3; ESI-MS m/z $[M+H]^+$: 485.08.

**18d** (n = 5): yield 40%; $^1H$ NMR ($CDCl_3$): δ 7.68 (d, 2H), 7.46 (d, 2H), 7.29 (t, 2H), 7.23 (t, 2H), 4.53 (m, 2H), 4.12 (t, 1H), 4.02 (m, 1H), 3.61 (m, 1H), 3.41 (m, 1H), 3.32-2.83 (m, 6H), 1.96 (d, 1H), 1.37 (s, 9H), 1.25-0.92 (m, 6H); $^{13}C$ NMR ($CDCl_3$): δ 157.7, 156.2, 143.8, 141.5, 127.7, 127.1, 124.6, 120.0, 79.2, 70.7, 66.5, 63.4, 50.1, 48.6, 47.4, 40.3, 29.6, 28.4, 27.8, 23.7; ESI-MS m/z $[M+H]^+$: 499.09.

**18e** (n = 6): yield 34.6%. $^1H$ NMR ($CDCl_3$): δ 7.68 (d, 2H), 7.48 (d, 2H), 7.32 (t, 2H), 7.23 (t, 2H), 5.21 (s, 1H), 4.51 (d, 2H), 4.14 (t, 1H), 4.06 (m, 1H), 3.64 (m, 1H), 3.47 (m, 1H), 3.44-2.87 (m, 6H), 1.97 (s, 1H), 1.37 (s, 9H), 1.36-0.97 (m, 6H); $^{13}C$ NMR ($CDCl_3$): δ 143.9, 141.4, 127.7, 127.1, 124.6,

120.0, 79.3, 70.8, 67.1, 63.4, 49.9, 48.8, 47.4, 40.5, 29.9, 28.4, 28.1, 26.3, 25.4; ESI-MS m/z $[M+H]^+$: 512.94.

**Synthesis of products 20 a-e.** Compound **18** (4.87 mmol) was dissolved in THF/ $H_2O$ (10 mL, 5/1) followed by the addition of sodium periodate (1.15 g, 5.36 mmol) in one portion. The mixture was sonicated in normal ultra sonic for 15 min and stirred for another 30 min at room temperature. The solid was then removed by filtration, and the solid was washed with DCM. The combined filtrate was concentrated and was extracted using DCM/$H_2O$. The organic phase was washed with brine, dried over sodium sulfate, and concentrated under reduced pressure. The residue was then dissolved in THF (4 mL) and added portionwise to the solution of glycine methyl ester hydrochloride (611 mg, 5.36 mmol) in MeOH which was neutralized to pH 7 by $Et_3N$ (678 µL) at room temperature. It was stirred at room temperature for 1 h. Sodium cyanoborohydride (918 mg, 14.61 mmol) was added to the above solution portionwise at 0 $^oC$, followed by the addition of acetic acid (279 µL, 4.87 mmol). It was stirred for 6 h at room temperature. $H_2O$ (10 mL) was added to quench the reaction. The organic solvent was removed under reduced pressure, the residue was extracted by EtOAc/$H_2O$, and the organic phase was washed with brine and dried over sodium sulfate. After removing the solvent, the product was purified by flash chromatography as pale yellow sticky oil.

**20a** (n = 2): yield 41%; $^1H$ NMR ($CDCl_3$): $\delta$ 7.66 (d, 2H), 7.47 (d, 2H), 7.30 (t, 2H), 7.22 (t, 2H), 5.5 (br, 1H), 5.2 (br, 1H), 4.46 (d, 2H), 4.12 (t, 1H), 3.61 (s,

3H), 3.29-2.63 (m, 10H), 1.32 (s, 9H); $^{13}$C NMR ($CDCl_3$): δ 172.7, 156.5, 156.0, 144.0, 141.4, 127.7, 127.1, 124.7, 119.9, 79.1, 66.8, 51.8, 50.3, 48.5, 47.9, 47.6, 47.4, 39.2, 29.7 ESI-MS m/z $[M+H]^+$: 498.07.

**20b** (n = 3): yield 46%; $^1$H NMR ($CDCl_3$): δ 7.67 (d, 2H), 7.49 (d, 2H), 7.31 (t, 2H), 7.23 (t, 2H), 5.2 (s, 1H), 4.50 (d, 2H), 4.14 (t, 1H), 3.65 (s, 3H), 3.30-2.61 (m, 10H), 1.89 (s, 1H), 1.35 (s, 9H), 1.2 (m, 2H); $^{13}$C NMR ($CDCl_3$): δ 156.1, 144.0, 141.4, 127.7, 127.1, 124.7, 119.9, 78.9, 66.6, 51.8, 50.5, 47.9, 47.5, 45.3, 44.9, 37.1, 29.7, 28.4; ESI-MS m/z $[M+H]^+$: 512.16.

**20c** (n = 4): yield 52%; $^1$H NMR ($CDCl_3$): δ 7.66 (d, 2H), 7.47(d, 2H), 7.29 (t, 2H), 7.21 (t, 2H), 5.18 (s, 1H), 4.45 (d, 2H), 4.11 (t, 1H), 3.61 (s, 3H), 3.42-2.83 (m, 10H), 1.34 (s, 9H), 1.16-1.09 (m, 4H); $^{13}$C NMR ($CDCl_3$): δ 172.9, 155.9, 144.1, 141.4, 127.6, 127.1, 124.7, 119.9, 79.0, 66.5, 53.4, 51.7, 50.5, 48.0, 47.4, 47.2, 40.1, 29.7, 28.4, 27.2; ESI-MS m/z $[M+H]^+$: 526.39..

**20d** (n = 5): yield 48%; $^1$H NMR ($CDCl_3$): δ 7.68 (d, 2H), 7.47(d, 2H), 7.33-7.18 (m, 4H), 4.8 (br, 1H), 4.3 (d, 2H), 4.0 (t, 1H), 3.44 (s, 3H), 3.42-2.23 (m, 10H), 1.74 (d, 1H), 1.34 (s, 9H), 1.2-0.81 (m, 6H); $^{13}$C NMR ($CDCl_3$): δ 172.4, 155.9, 144.0, 141.2, 127.5, 127.0, 124.6, 119.8, 78.5, 66.4, 54.2, 51.6, 50.2, 47.3, 46.7, 42.5, 40.3, 29.5, 28.4, 27.2, 23.7;ESI-MS m/z $[M+H]^+$: 540.19.

**20e** (n = 6): yield 49%; $^1$H NMR ($CDCl_3$): δ 7.67 (d, 2H), 7.48 (d, 2H), 7.31 (t, 2H), 7.23 (t, 2H), 4.53 (d, 2H), 4.14 (t, 1H), 3.69 (s, 3H), 3.16-2.98 (m, 10H), 1.92 (s, 1H), 1.36 (s, 9H), 1.35-1.13 (m, 8H); $^{13}$C NMR ($CDCl_3$): δ 166.8, 143.8, 141.4, 127.7, 127.1, 124.7, 120.0, 79.1, 67.4, 53.1, 48.4, 47.9, 47.3, 45.3, 44.5,

40.4, 29.1, 28.4, 28.3, 26; ESI-MS m/z $[M+H]^+$: 554.49.

**Synthesis of AP-PNA monomer acids 22.** The appropriately protected nucleobase acetic acid (1 mmol) was added to a clear solution of **20** (1 mmol) in DMF (3 mL), followed by the addition of HATU (570 mg, 1.5 mmol) and DIEA (679 μL, 4 mmol) at 0 $^oC$. The reaction was stirred at room temperature overnight. The reaction mixture was then poured into cold water (40 mL) to form a white precipitation. The precipitation was filtered out and washed with water and hexane to give a white solid which was purified by column chromatography to yield the pure methyl ester product **21**. **21** (1 mmol) was dissolved in a mixture of THF (3 mL) and $H_2O$ (1 mL) at 0 $^oC$. NaOH (2 M, 2 mmol) was added dropwise and stirred at that temperature for 20 min. When the reaction was completed, the mixture was adjusted to pH 4 by the careful addition of potassium hydrogen sulfate solution (1 M). THF was removed under reduced pressure; the residue was extracted with ethyl acetate (30 mL × 3). The combined organic phase was dried over sodium sulfate and concentrated under reduced pressure. The product was lyophilized to obtain a white or pale yellow powder.

**T2-OH**: yield 96%; $^1H$ NMR ($d_6$-DMSO): δ 9.86 (d, 1H), 7.68 (d, 2H), 7.48 (d, 2H), 7.31 (t, 2H), 7.23 (t, 2H), 6.94 (d, 1H), 4.54 (d, 2H), 4.32 (t, 1H), 4.13 (s, 2H), 3.95 (s, 2H), 3.69-2.94 (m, 8H), 1.79 (d, 3H), 1.31 (m, 9H); $^{13}C$ NMR ($d_6$-DMSO): δ 172.8, 168.1, 164.9, 156.3, 151.3, 143.8, 141.6, 141.4, 127.8, 127.3, 124.7, 119.9, 110.6, 79.8, 67.7, 50.2, 48.3, 47.8, 47.3, 46.5, 45.3, 38.9, 28.4, 12.2; ESI-MS m/z $[M+Na]^+$: 672.33; ESI-HRMS calcd for m/z $[M+H]^+$

($C_{33}H_{40}N_5O_9$): 650.2826, found 650.2856.

**A2-Bz-OH**: yield 89%; $^1$H NMR ($d_6$-DMSO): δ 11.8 (s, 1H), 11.2 (s, 1H), 8.67 (d, 1H), 8.07 (d, 1H), 7.89 (d, 2H), 7.65 (d, 2H), 7.57-7.32 (m, 9H), 6.7 (m, 1H), 4.6-2.4 (m, 15H), 1.46 (s, 9H); $^{13}$C NMR ($d_6$-DMSO): δ 169.1, 166.6, 164.9, 156.3, 152.6, 151.9, 150.4, 144.2, 143.8, 141.4, 133.8, 132.8, 128.9, 127.6, 127.2, 125.6, 120.6, 79.8, 67.1, 50.6, 49.7, 48.7, 47.3, 46.6, 43.7, 43.2, 28.7; ESI-MS m/z $[M+H]^+$: 763.71; ESI-HRMS calcd for m/z $[M+H]^+$ ($C_{40}H_{43}N_8O_8$): 763.3204, found 763.3198.

**C2-Bz-OH**: yield 87%; $^1$H NMR ($CDCl_3$): δ 8.56 (br, 1H), 8.0-7.28 (m, 15H), 4.94-3.7 (m, 7H), 3.6-2.85 (m, 8H), 1.4 (s, 9H); $^{13}$C NMR ($CDCl_3$): δ 171.1, 168.1, 166.8, 161.4, 156.5, 154.6, 150.2, 143.8, 141.3, 134, 131.3, 128.8, 127.8, 127.3, 124.7, 119.9, 96.4, 79.9, 67.1, 50.8, 50.3, 49.1, 47.8, 47.2, 46.5, 38.9, 28.3; ESI-MS m/z $[M+H]^+$: 739.17; ESI-HRMS calcd for m/z $[M+H]^+$ ($C_{39}H_{43}N_6O_9$): 739.3092, found 739.3095.

**G2-iBu-OH**: yield 91%; $^1$H NMR ($d_6$-DMSO): δ 12.09 (m, 1H), 11.64 (m, 1H), 7.9 (d, 1H), 7.87 (d, 2H), 7.62 (d, 2H), 7.41 (d, 2H), 7.35 (d, 2H), 6.89 (s, 1H), 5.19-3.74 (m, 7H), 3.21-2.67 (m, 9H), 1.34 (s, 9H), 1.11 (s, 6H); $^{13}$C NMR ($d_6$-DMSO): δ 180.1, 170.2, 166.8, 157.3, 156.2, 149.2, 147.3, 144.4, 141.4, 140.9, 128.1, 127.6, 125.1, 120.6, 77.9, 67.4, 51.3, 48.8, 47.3, 46.4, 44.9, 40.4, 39.7, 35.1, 28.6, 19.3; ESI-MS m/z $[M+H]^+$: 767.36; ESI-HRMS calcd for m/z $[M+H]^+$ ($C_{37}H_{45}N_8O_9$): 745.3310, found 745.3293.

**A2-Bhoc-OH:** yield 92%; $^1$H NMR ($CDCl_3$): δ 10.02 (br, 1H), 8.43-7.28 (m,

20H), 7.01 (s, 1H), 5.87 (s, 1H), 5.28-4.04 (m, 7H), 3.78-2.82 (m, 8H), 1.39 (d, 9H); $^{13}$C NMR ($CDCl_3$): δ 171.2, 166.0, 162.1, 156.4, 151.6, 150.7, 149.7, 145.0, 143.7, 141.4, 141.2, 139.2, 128.8, 127.9, 127.6, 127.3, 127.0, 124.6, 121.3, 120.0, 80.9, 79.9, 67.0, 51.1, 48.1, 47.3, 47.1, 45.1, 44.7, 39.8, 28.3; ESI-MS m/z $[M+H]^+$: 869.11; ESI-HRMS calcd for m/z $[M+H]^+$ ($C_{47}H_{49}N_8O_9$): 869.3623, found 869.3604.

**C2-Bhoc-OH:** yield 91%; $^1$H NMR ($CDCl_3$): δ 10.19 (d, 1H), 8.81 (br, 1H), 7.77-7.01 (m, 20H), 6.78 (s, 1H), 6.78 (s, 1H), 5.88 (s, 1H), 4.69-2.94 (m, 15H), 1.38 (s, 9H); $^{13}$C NMR ($CDCl_3$): δ 171.6, 167.1, 162.5, 156.8, 156.5, 151.8, 149.9, 143.8, 141.3, 139.3, 128.7, 128.3, 127.6, 127.1, 126.9, 124.7, 120.0, 95.5, 80.9, 79.9, 66.9, 50.2, 49.8, 47.9, 47.3, 46.6, 44.5, 40.1, 28.4; ESI-MS m/z $[M+H]^+$: 844.88; ESI-HRMS calcd for m/z $[M+H]^+$ ($C_{46}H_{49}N_6O_{10}$): 845.3510, found 845.3508.

**G2-Bhoc-OH:** yield 93%; $^1$H NMR ($CDCl_3$): δ 11.4 (br, 1H), 10.68 (d, 1H), 8.53-7.18 (m, 19H), 6.89 (d, 1H), 6.1 (s, 1H), 4.82-2.94 (m, 15H), 1.39 (s, 9H); $^{13}$C NMR ($CDCl_3$): δ 170.6, 166.0, 160.7, 157.4, 156.5, 153.6, 149.2, 148.2, 143.8, 141.4, 140.7, 139.1, 128.7, 128.5, 127.8, 127.3, 127.1, 124.4, 121.0, 120.0, 79.8, 66.9, 66.5, 51.1, 47.7, 46.9, 45.8, 44.8, 44.3, 38.7, 28.4; ESI-MS m/z $[M+H]^+$: 885.25; ESI-HRMS calcd for m/z $[M+H]^+$ ($C_{47}H_{49}N_8O_{10}$): 885.3572, found 885.3560.

**T3-OH**: yield 92%; $^1$H NMR ($d_6$-DMSO): δ 11.29 (s, 1H), 7.88 (d, 2H), 7.52 (d, 2H), 7.4-7.25 (m, 4H), 6.7 (s, 1H), 4.66-3.97 (m, 5H), 3.73-2.73 (m, 10H),

1.73 (s, 3H), 1.41 (s, 9H), 1.40-0.9 (m, 2H); $^{13}$C NMR ($d_6$-DMSO): δ 170.9, 167.8, 164.8, 156.4, 156, 151.4, 144.4, 143.8, 141.3, 128.1, 127.5, 125.4, 120.5, 108.5, 78.0, 66.9, 49.1, 48.9, 48.3, 47.3, 47, 45.8, 38.0, 29.4, 28.7, 12.4; ESI-MS m/z $[M+H]^+$: 664.25; ESI-HRMS calcd for m/z $[M+H]^+$ ($C_{34}H_{42}N_5O_9$): 664.2983, found 664.2971.

**A3-Bz-OH**: yield 86%; $^1$H NMR ($d_6$-DMSO): δ 11.17 (s, 1H), 8.63 (d, 1H), 8.32 (s, 1H), 8.06 (d, 2H), 7.90 (d, 2H), 7.67 (d, 2H), 7.58 (d, 2H), 7.41-7.32 (m, 5H), 6.8 (br, 1H), 5.18-4.01 (m, 7H), 3.76-2.33 (m, 8H), 1.49 (m, 2H), 1.35 (s, 9H); $^{13}$C NMR ($d_6$-DMSO): δ 171.8, 169.4, 166.3, 156.2, 153.4, 151.5, 150.9, 144.4, 143.8, 141.3, 133.3, 132.8, 128.9, 128.1, 127.6, 125.4, 120.5, 78.0, 66.9, 51.1, 48.3, 47.3, 46.5, 44.8, 42.3, 38.0, 29.2, 28.7; ESI-MS m/z $[M+H]^+$: 777.83; ESI-HRMS calcd for m/z $[M+H]^+$ ($C_{41}H_{45}N_8O_8$): 777.3360, found 777.3356.

**C3-Bz-OH**: yield 84%; $^1$H NMR ($d_6$-DMSO): δ 11.2 (s. 1H), 7.96-7.16 (m, 15H), 4.89-2.57 (m, 15H), 1.32 (s, 9H); $^{13}$C NMR ($d_6$-DMSO): δ 170.8, 167.9, 167.5, 163.8, 156.1, 155.6, 151.6, 144.4, 141.4, 133.7, 133.2, 128.9, 128.0, 127.6, 125.4, 120.5, 96.2, 77.9, 66.8, 50.6, 50.1, 48.4, 47.9, 47.3, 45.9, 39.6, 29.7, 28.7; ESI-MS m/z $[M+H]^+$: 753.67; ESI-HRMS calcd for m/z $[M+H]^+$ ($C_{40}H_{45}N_6O_9$): 753.3248, found 753.3254.

**G3-iBu-OH**: yield 89%; $^1$H NMR ($d_6$-DMSO): δ 12.1 (s, 1H), 11.6 (s, 1H), 7.72-7.64 (m, 3H), 7.57-7.28 (m, 6H), 4.93-3.99 (m, 7H), 3.75-2.59 (m, 9H), 1.35 (s, 9H), 1.34 (m, 2H), 1.10 (m, 6H); $^{13}$C NMR ($d_6$-DMSO): δ 180.5, 173.2, 155.9, 155.7, 149.8, 148.4, 144.4, 141.3, 135.9, 128.4, 128.1, 127.6, 125.4, 121.7,

120.5, 78, 66.9, 50.2, 47.3, 46.4, 44.9, 40.4, 37.9, 36.8, 34.2, 28.7, 25.7, 19.3; ESI-MS m/z $[M+H]^+$: 759.98; ESI-HRMS calcd for m/z $[M+H]^+$ ($C_{38}H_{47}N_8O_9$): 759.3466, found 759.3458.

**A3-Bhoc-OH:** yield 90%; $^1$H NMR ($CDCl_3$): δ 10.05 (s, 1H), 9.3 (br, 1H), 8.58-7.23 (m, 20 H), 6.86 (s, 1H), 6.09 (s, 1H), 5.78 (s, 1H), 5.32-5.15 (m, 2H), 4.82-3.98 (m, 5 H), 3.78-2.83 (m, 8H), 1.39 (d, 9H), 1.38-1.21 (m, 2H); $^{13}$C NMR ($CDCl_3$): δ 171.2, 166.0, 162.1, 156.5, 151.6, 151.2, 150.7, 149.7, 144.9, 143.7, 141.4, 139.2, 128.5, 128.3, 127.6, 127.2, 127.0, 124.6, 121.3, 120.0, 81.0, 79.9, 67.0, 51.1, 48.1, 47.3, 47.1, 45.1, 44.7, 39.8, 28.4, 28.3; ESI-MS m/z $[M+H]^+$: 883.04; ESI-HRMS calcd for m/z $[M+H]^+$ ($C_{48}H_{51}N_8O_9$): 883.3779, found 883.3790.

**C3-Bhoc-OH:** yield 93%; $^1$H NMR ($CDCl_3$): δ 9.82 (s, 1H), 8.81 (s, 1H), 7.72-7.01 (m, 20H), 6.78 (d, 1H), 6.1 (s, 1H), 5.88 (s, 1H), 4.77-2.83 (m, 15H), 1.44 (s, 9H), 1.43-1.41 (m, 2H); $^{13}$C NMR ($CDCl_3$): δ 171.6, 167.1, 160.5, 156.5, 151.8, 149.9, 148.8, 143.8, 141.3, 139.3, 128.7, 128.3, 127.8, 127.3, 126.9, 124.6, 120.0, 95.5, 80.9, 79.9, 66.9, 50.2, 49.8, 47.9, 47.3, 46.6, 44.5, 38.5, 28.4, 28.1; ESI-MS m/z $[M+H]^+$: 859.19; ESI-HRMS calcd for m/z $[M+H]^+$ ($C_{47}H_{51}N_6O_{10}$): 859.3667, found 859.3654.

**G3-Bhoc-OH:** yield 90%; $^1$H NMR ($CDCl_3$): δ 11.4 (br, 1H), 10.6 (d, 1H), 9.43 (br, 1H), 8.6 (br, 1H), 7.74-7.28 (m, 19H), 6.89 (s, 1H), 6.10 (s, 1H), 5.16-3.96 (m, 7H), 3.52-2.88 (m, 8H), 1.59-1.53 (m, 2H), 1.39 (s, 9H); $^{13}$C NMR ($CDCl_3$): δ 169.8, 165.9, 160.6, 157.4, 156.5, 153.5, 149.4, 148.2, 143.7, 141.4,

140.5, 139.0, 128.7, 128.5, 127.8, 127.3, 126.9, 124.4, 120.2, 120.0, 79.9, 66.9, 66.4, 51.1, 47.8, 46.8, 45.6, 44.3, 42.9, 37.9, 28.4, 27.6; ESI-MS m/z $[M+H]^+$: 899.19; ESI-HRMS calcd for m/z $[M+H]^+$ ($C_{48}H_{51}N_8O_{10}$): 899.3728, found 899.3738.

**T4-OH**: yield 95%; $^1$H NMR ($d_6$-DMSO): δ 11.28 (s, 1H), 7.89 (d, 2H), 7.63 (d, 2H), 7.41 (d, 2H), 7.33 (d, 2H), 6.78 (s, 1H), 4.64-3.95 (m, 5H), 3.16-2.33 (m, 10H), 1.74 (d, 3H), 1.37 (s, 9H), 1.23-1.08 (m, 4H); $^{13}$C NMR ($d_6$-DMSO): δ 170.8, 167.8, 165.1, 156.1, 155.8, 151.2, 144.5, 142.3, 141.4, 128.0, 127.5, 125.2, 120.5, 109.8, 77.8, 66.7, 49.1, 48.1, 47.7, 47.3, 45.8, 43.3, 39.2, 29.4, 28.7, 27.3, 12.3; ESI-MS m/z $[M+Na]^+$: 700.42; ESI-HRMS calcd for m/z $[M+H]^+$ ($C_{35}H_{44}N_5O_9$): 678.3139, found 678.3142.

**A4-Bz-OH**: yield 93%; $^1$H NMR ($d_6$-DMSO): δ 11.2 (s, 1H), 8.65 (s, 1H), 8.32 (d, 1H), 8.02 (d, 2H), 7.89 (d, 2H), 7.62 (m, 4H), 7.45-7.19 (m, 5H), 6.8 (br, 1H), 5.38-5.12 (m, 2H), 4.56-4.07 (m, 5H), 3.67-2.88 (m, 8H), 1.37 (s, 9H), 1.35-1.06 (m, 4H); $^{13}$C NMR ($d_6$-DMSO): δ 171.8, 169.3, 166.5, 156.6, 152.7, 150.9, 150.1, 146.3, 144.7, 140.8, 133.3, 132.8, 128.7, 128.2, 127.7, 124.3, 119.8, 79.8, 66.9, 51.1, 48.3, 47.3, 46.5, 44.9, 40.4, 39.7, 28.4, 26.3, 25.2; ESI-MS m/z $[M+H]^+$: 791.74; ESI-HRMS calcd for m/z $[M+H]^+$ ($C_{42}H_{47}N_8O_8$): 791.3517, found 791.3522.

**C4-Bz-OH**: yield 88%; $^1$H NMR ($d_6$-DMSO): δ 11.2 (s, 1H), 8.01-6.92 (m, 15H), 4.81-4.09 (m, 7H), 3.8-2.88 (m, 8H), 1.37 (s, 9H), 1.38-1.19 (m, 4H); $^{13}$C NMR ($d_6$-DMSO): δ 170.8, 167.9, 167.5, 163.8, 156.1, 155.6, 151.6, 144.5,

141.4, 133.7, 133.2, 129.9, 128.0, 127.5, 125.2, 120.5, 96.2, 77.8, 66.6, 50.6, 50.1, 48.4, 47.9, 47.3, 45.9, 42.0, 28.4, 27.3, 26.1; ESI-MS m/z $[M+H]^+$: 767.78; ESI-HRMS calcd for m/z $[M+H]^+$ ($C_{41}H_{47}N_6O_9$): 767.3405, found 767.3419.

**G4-iBu-OH**: yield 89%; $^1$H NMR ($d_6$-DMSO): δ 12.1 (s, 1H), 11.6 (s, 1H), 7.89-7.19 (m, 9H), 6.79 (br, 1H), 5.15-3.98 (m, 7H), 3.75-2.50 (m, 9H), 1.37 (s, 9H), 1.34 (m, 4H), 1.10 (m, 6H); $^{13}$C NMR ($d_6$-DMSO): δ 180.1, 170.3, 166.8, 156, 155.3, 149.7, 146.8, 144.4, 141.4, 128.0, 127.5, 125.1, 120.5, 77.8, 67.3, 51.3, 48.8, 47.3, 46.4, 44.9, 40.4, 39.7, 35.1, 28.7, 27.3, 26.5, 19.3; ESI-MS m/z $[M+H]^+$: 773.85; ESI-HRMS calcd for m/z $[M+H]^+$ ($C_{39}H_{49}N_8O_9$): 773.3623, found 773.3621.

**A4-Bhoc-OH:** yield 90%; $^1$H NMR ($CDCl_3$): δ 9.89 (s, 1H), 8.60-7.19 (m, 20 H), 6.92 (s, 1H), 6.0 (s, 1H), 5.77 (s, 1H), 5.32-5.15 (m, 2H), 4.98-4.27 (m, 5 H), 4.10-2.72 (m, 8H), 1.39 (d, 9H), 1.38-1.02 (m, 4H); $^{13}$C NMR ($CDCl_3$): δ 171.3, 167.4, 162.1, 156.6, 155.3, 150.7, 149.0, 144.8, 143.7, 141.4, 137.8, 128.5, 128.4, 127.3, 127.2, 127.1, 124.5, 121.0, 120.0, 81.0, 79.8, 67.0, 49.8, 48.1, 47.3, 47.1, 44.7, 40.1, 39.8, 29.7, 27.1, 25.7; ESI-MS m/z $[M+H]^+$: 897.10; ESI-HRMS calcd for m/z $[M+H]^+$ ($C_{49}H_{53}N_8O_9$): 897.3936, found 897.3912.

**C4-Bhoc-OH:** yield 93%; $^1$H NMR ($CDCl_3$): δ 9.83 (s, 1H), 8.93 (s, 1H), 7.87-7.28 (m, 20H), 6.80 (d, 1H), 6.1 (s, 1H), 5.88 (s, 1H), 4.81-2.80 (m, 15H), 1.44 (s, 9H), 1.43-1.41 (m, 4H); $^{13}$C NMR ($CDCl_3$): δ 171.6, 167.1, 163.5, 156.5, 155.4, 151.8, 148.8, 143.8, 141.3, 139.3, 128.7, 128.3, 127.8, 127.2, 126.9, 124.6, 120.0, 95.5, 80.1, 79.5, 66.9, 50.2, 49.8, 47.9, 47.3, 46.6, 44.5, 40.1, 28.4, 26.9,

25.6; ESI-MS m/z $[M+Na]^+$: 895.58; ESI-HRMS calcd for m/z $[M+H]^+$ ($C_{48}H_{53}N_6O_{10}$): 873.3823, found 873.3829.

**G4-Bhoc-OH:** yield 91%; $^1H$ NMR ($CDCl_3$): δ 11.4 (br, 1H), 10.6 (d, 1H), 8.58-7.15 (m, 19H), 6.89 (d, 1H), 4.83-2.86 (m, 15H), 1.43 (s, 9H), 1.42-1.27 (m, 4H); $^{13}C$ NMR ($CDCl_3$): δ 171.8, 166.0, 160.6, 157.2, 156.5, 153.3, 149.3, 148.2, 143.4, 141.4, 140.6, 138.9, 128.7, 128.5, 127.8, 127.3, 127.0, 124.4, 120.3, 120.0, 79.8, 67.1, 66.3, 51.1, 48.5, 47.7, 46.8, 45.7, 44.3, 40.1, 29.2, 28.4, 26.9; ESI-MS m/z $[M+H]^+$: 913.22; ESI-HRMS calcd for m/z $[M+H]^+$ ($C_{49}H_{53}N_8O_{10}$): 913.3885, found 913.3893.

**T5-OH**: yield 89%; $^1H$ NMR ($CDCl_3$): δ 9.87 (br, 1H), 7.8 (d, 2H), 7.6 (d, 2H), 7.4 (t, 2H), 7.3 (t, 2H), 7.0 (s, 1H), 5.17 (br, 1H), 4.76-4.08 (m, 7H), 3.68-2.96 (m, 8H), 1.83 (d, 3H), 1.45 (s, 9H), 1.34-1.12 (m, 6H); $^{13}C$ NMR ($CDCl_3$): δ 171.9, 167.7, 164.7, 156.2, 151.5, 143.9, 143.4, 141.4, 127.8, 127.1, 124.6, 120.0, 110.5, 79.2, 66.8, 49.1, 48.9, 48.3, 47.7, 47.3, 44.8, 40.4, 29.6, 28.4, 28.2, 23.8, 12.2; ESI-MS m/z $[M+H]^+$: 692.01; ESI-HRMS calcd for m/z $[M+H]^+$ ($C_{36}H_{46}N_5O_9$): 692.3296, found 692.3285.

**A5-Bhoc-OH:** yield 90%; $^1H$ NMR ($CDCl_3$): δ 9.3 (br, 1H), 8.56 (m, 1H), 8.06 (m, 1H), 7.76 (m, 2H), 7.54-7.02 (m, 16H), 6.93 (s, 1H), 6.0 (s, 1H), 5.19-4.07 (m, 7H), 3.92-2.85 (m, 8H), 1.47 (s, 9H), 1.32-1.02 (m, 6H); $^{13}C$ NMR ($CDCl_3$): δ 172.3, 167.1, 156.1, 152.3, 151.0, 149.1, 144.7, 143.9, 141.4, 139.8, 128.5, 128.0, 127.7, 127.3, 127.2, 124.6, 121.0, 120.0, 79.0, 78.6, 66.7, 51.3, 50.2, 48.4, 47.3, 46.5, 44.0, 40.4, 29.6, 28.4, 28.2, 23.8; ESI-MS m/z $[M+H]^+$: 911.65;

ESI-HRMS calcd for m/z $[M+Na]^+$ ($C_{50}H_{54}N_8O_9Na$): 933.3911, found 933.3899.

**C5-Bhoc-OH:** yield 93%; $^1H$ NMR ($CDCl_3$): δ 7.78-7.17 (m, 20H), 6.84 (d, 1H), 4.73-4.02 (m, 7H), 3.81-2.98 (m, 8H), 1.34 (s, 9H), 1.31-0.92 (m, 6H); $^{13}C$ NMR ($CDCl_3$): δ 173.2, 167.6, 163.1, 156.3, 155.3, 151.8, 150.1, 144.0, 141.4, 139.6, 128.7, 128.2, 127.7, 127.1, 126.9, 124.7, 120.0, 95.4, 79.0, 78.4, 66.7, 51.2, 50.5, 49.6, 48.3, 47.4, 42.7, 40.4, 29.6, 28.4, 28.2, 23.7; ESI-MS m/z $[M+H]^+$: 887.38; ESI-HRMS calcd for m/z $[M+H]^+$ ($C_{49}H_{55}N_6O_{10}$): 887.3980, found 887.3976.

**G5-Bhoc-OH:** yield 91%; $^1H$ NMR ($CDCl_3$): δ 11.2 (br, 1H), 10.4 (br, 1H), 7.73-7.15 (m, 19H), 6.78 (m, 1H), 4.83-4.12 (m, 7H), 3.71-2.66 (m, 8H), 1.43 (s, 9H), 1.38-0.99 (m, 6H); $^{13}C$ NMR ($CDCl_3$): δ 169.2, 167.1, 157.0, 156.1, 155.8, 153.6, 149.3, 147.2, 143.4, 141.3, 139.3, 128.7, 128.2, 127.8, 127.2, 127.0, 124.4, 120.3, 119.1, 79.2, 79.0, 66.7, 51.1, 50.2, 48.6, 47.5, 46.4, 44.3, 40.4, 29.6, 28.4, 26.7, 23.7; ESI-MS m/z $[M+H]^+$: 927.77; ESI-HRMS calcd for m/z $[M+H]^+$ ($C_{50}H_{55}N_8O_{10}$): 927.4041, found 927.4019.

**T6-OH**: yield 90%; $^1H$ NMR ($d_6$-DMSO): δ 7.77 (d, 2H), 7.58 (d, 2H), 7.4 (t, 2H), 7.34 (t, 2H), 7.01 (s, 1H), 4.87-4.14 (m, 7H), 3.74-2.95 (m, 8H), 1.97 (d, 3 H), 1.37 (s, 9 H), 1.38-0.81 (m, 8 H); $^{13}C$ NMR ($d_6$-DMSO): δ 167.7, 165.1, 156.5, 153.4, 143.8, 141.4, 140.9, 128.0, 127.5, 125.1, 120.5, 110.7, 71.2, 66.8, 63.4, 51.2, 50.8, 48.9, 47.4, 42.1, 40.3, 30.0, 28.4, 26.2, 12.3. ESI-MS m/z $[M+H]^+$:706.27; ESI-HRMS calcd for m/z $[M+H]^+$ ($C_{37}H_{48}N_5O_9$): 706.3452, found 706.3417.

**A6-Bz-OH**: yield 88%; $^{1}$H NMR ($CDCl_3$): δ 8.6 (m, 2H), 8.04 (m, 2H), 7.71-7.26 (m, 11H), 5.85 (br, 1H), 5.42-4.44 (m, 5H), 4.22-3.05 (m, 10H), 1.45 (s, 9H), 1.44-1.09 (m, 8H); $^{13}$C NMR ($CDCl_3$): δ 166.2, 161.2, 161.0, 156.2, 151.2, 144.8, 143.7, 141.4, 133.5, 128.8, 128.5, 127.8, 127.1, 124.6, 119.9, 77.8, 66.4, 50.2, 48.8, 47.8, 47.3, 45.9, 42.1, 40.3, 30.0, 28.4, 26.3. ESI-MS m/z $[M+H]^+$: 819.58; ESI-HRMS calcd for m/z $[M+H]^+$ ($C_{44}H_{51}N_8O_8$): 819.3830, found 819.3823.

**C6-Bz-OH**: yield 85%; $^{1}$H NMR ($CDCl_3$): δ 8.04-7.29 (m, 15H), 5.82 (br, 1H), 4.85-4.46 (m, 3H), 4.22-3.72 (m, 4H), 3.32-2.96 (m, 8H), 1.45 (s, 9H), 1.42-1.09 (m, 8H); $^{13}$C NMR ($CDCl_3$): δ 167.7, 165.3, 156.1, 151.7, 149.8, 143.8, 141.3, 134.2, 128.9, 128.8, 127.8, 127.1, 124.6, 119.9, 96.2, 79.2, 66.4, 50.3, 50.1, 47.8, 47.3, 45.9, 42.1, 40.3, 30.0, 28.4, 26.6; ESI-MS m/z $[M+H]^+$: 795.26; ESI-HRMS calcd for m/z $[M+H]^+$ ($C_{43}H_{51}N_6O_9$): 795.3718, found 795.3726.

**G6-iBu-OH**: yield 84%; $^{1}$H NMR ($d_6$-DMSO): δ 12.1 (m, 1H), 10.74 (m, 1H), 8.45 (m, 1H), 7.81-7.16 (m, 8H), 5.62 (br, 1H), 4.95-4.65 (m, 3H), 4.42-3.74 (m, 4H), 3.52-2.81 (m, 8H), 2.71 (m, 1H), 1.45 (s, 9H), 1.43-1.18 (m, 14H); $^{13}$C NMR ($d_6$-DMSO): δ 166.2, 161.3, 157.2, 148.5, 147.9, 143.6, 141.4, 140.4, 127.9, 127.2, 125.6, 124.2, 120.5, 77.8, 66.4, 47.4, 40.4, 36.6, 36.2, 29.9, 28.7, 26.5, 26.4, 18.8; ESI-MS m/z $[M+H]^+$: 802.20; ESI-HRMS calcd for m/z $[M+H]^+$ ($C_{41}H_{53}N_8O_9$): 801.3936, found 801.3920.

**A6-Bhoc-OH:** yield 90%; $^{1}$H NMR ($d_6$-DMSO): δ 10.92 (s, 1H), 8.50 (s, 1H), 8.31 (d, 1H), 7.87 (d, 2H), 7.63 (d, 2H), 7.5 (m, 4H), 7.32-7.19 (m, 10H),

6.80 (s, 1H), 5.32-5.15 (m, 2H), 4.41-4.35 (m, 3 H), 4.05 (s, 2H), 3.87-2.72 (m, 8H), 1.39 (d, 9H), 1.38-0.72 (m, 8H); $^{13}$C NMR ($d_6$-DMSO): δ 171.3, 167.4, 156.6, 152.8, 151.3, 150.2, 149.1, 143.8, 141.4, 141.3, 139.8, 128.8, 128.1, 127.7, 127.3, 127.0, 124.7, 121.0, 120.5, 79.6, 68.2, 66.9, 53.4, 48.5, 47.3, 45.5, 44.0, 41.7, 38.7, 29.9, 29.3, 28.9, 28.4, 26.3. ESI-MS m/z $[M+H]^+$: 925.16; ESI-HRMS calcd for m/z $[M+H]^+$ ($C_{51}H_{57}N_8O_9$): 925.4249, found 925.4255.

**C6-Bhoc-OH:** yield 91%; $^1$H NMR ($d_6$-DMSO): δ 10.98 (br, 1H), 7.87-6.92 (m, 20H), 6.77 (s, 1H), 4.81-3.96 (m, 7H), 3.57-2.80 (m, 8H), 1.37 (s, 9H), 1.36-1.09 (m, 8H); $^{13}$C NMR ($d_6$-DMSO): δ 170.8, 167.5, 163.5, 156.1, 155.4, 152.8, 151.5, 144.5, 141.5, 140.9, 129.0, 128.3, 128.0, 127.5, 126.9, 125.1, 120.5, 94.2, 77.9, 77.8, 66.4, 50.1, 50.0, 48.4, 47.9, 47.4, 45.9, 40.2, 30.2, 28.7, 27.9, 26.6, 26.4; ESI-MS m/z $[M+H]^+$: 901.11; ESI-HRMS calcd for m/z $[M+H]^+$ ($C_{50}H_{57}N_6O_{10}$): 901.4136, found 901.4131.

**G6-Bhoc-OH:** yield 91%; $^1$H NMR ($CDCl_3$): δ 11.26 (br, 1H), 10.61 (d, 1H), 8.5 (d, 1H), 7.72-7.11 (m, 19H), 6.88 (d, 1H), 5.58 (br, 1H), 4.84-2.92 (m, 15H), 1.43 (s, 9H), 1.41-1.28 (m, 8H); $^{13}$C NMR ($CDCl_3$): δ 169.8, 164.7, 160.6, 157.2, 153.3, 149.3, 148.2, 143.4, 141.4, 140.6, 139.1, 128.7, 128.5, 127.8, 127.3, 126.9, 124.4, 120.0, 119.2, 79.9, 67.1, 66.3, 51.1, 48.5, 47.7, 46.8, 45.6, 44.3, 40.1, 29.9, 28.4, 27.3, 26.1, 26.0; ESI-MS m/z $[M+H]^+$: 941.16; ESI-HRMS calcd for m/z $[M+H]^+$ ($C_{51}H_{57}N_8O_{10}$): 941.4198, found 941.4195.

**t-Butyl N-[2-(N-9-fluorenylmethoxycarbonyl)aminoethyl] glycinate hydrochloride 23.** t-Butylbromoacetate (7.5 ml, 31 mmol) in dry DCM (10 mL)

was added dropwise to a solution of ethylenediamine (30 mL, 450mmol) at 0 $^{o}$C over 1.5 h. The resulting mixture was stirred at 0 $^{o}$C for 3h and at room temperature for 18 h. The reaction mixture was diluted with DCM (20 mL), washed with brine (20 mL) [caution: the product is also partially soluble in water] which was re-extracted with DCM (3 × 20 mL). The combined organic layer was washed with brine (20 mL). TLC (colored by ninhydrin) showed ethylenediamine was gone. The organic layer was dried and concentrated to get 7.551 g (85%) as yellow oil. The yellow oil was dissolved in DCM (25 mL) and TEA (6 mL) was added. To this solution was added dropwise a solution of Fmoc-Cl (11.2 g) in DCM (50 mL) at 0 $^{o}$C over a period of 2 h. After addition, the reaction was allowed to stir at room temperature for a further 12 h. The resulting solution was washed with 1 M HCl ( 2 × 50 mL) and brine (50 mL). The organic layer was dried and concentrated to give light yellow oil. Cooling at - 20 $^{o}$C overnight resulted in a precipitate which was collected by filtration and washed with DCM until the solid was colorless. The filtrate was concentrated and cooled to yield more solid. All the combined solid was dried under vacuum to give **23** (5.53 g) as a white solid, yield 45%. $^{1}$H NMR (400 MHz, $d_6$-DMSO): δ 9.20 (s, 2H), 7.94 (d, 2H), 7.73 (d, 2H), 7.57 (t, 1H), 7.46 (t, 2H), 7.37 (t, 2H), 4.38 (d, 2H), 4.27(t, 1H), 3.91 (s, 2H), 3.37 (t, 2H), 3.03 (t, 2H). $^{13}$C NMR (400 MHz, $d_6$-DMSO): δ 166.27, 156.77, 144.27, 141.22, 128.32, 127.55, 125.59, 120.13, 83.50, 66.06, 47.79, 47.12, 46.89, 37.11, 28.09; ESI-MS m/z $[M+H]^+$ : 396.92.

**General procedure for the synthesis of the aegPNA monomer acid.**

Nucleobase-acetic acid (1 eq) was dissolved in dry DMF, HATU (1.5 eq) and NMM (5 eq) was added. The yellow green mixture was stirred at 0 $^{o}$C for 10 min. **23** (1 eq) was added in one portion and the stirring continued at 0 $^{o}$C for 30 min. The reaction was allowed to stir at room temperature for 22 h. The progression of the reaction was monitored by HPLC and ESI MS. After consumption of starting material, the reaction mixture was concentrated *in vacuo* and the residue was diluted with DCM, which was washed with 1M HCl (3 × 30 mL) to adjust the pH to 6 ~ 7 and brine (1 × 30 mL). The organic layer was dried and removed to get a yellow oil which was subject to flash column chromatography (DCM/ MeOH = 9.5/0.5). The ester was isolated as a white solid. The ester was treated with 5 mL 95% TFA at 0 $^{o}$C for 30 min. The mixture was continued to stir for 2 h while warming to room temperature. The progression of the reaction was monitored by HPLC and ESI MS. After consumption of the start material, TFA was removed and the residue was separated with DCM/$H_2O$, and the water layer was re-extracted with DCM (3 × 30 mL), the combined DCM layer was washed with water (2 × 50 mL) and brine (1 × 50 mL). The organic solvent was removed and the residue was dried under vacuum to give the acid as a white solid.

**T-OH:** yield 84%; $^{1}$H NMR (400 MHz, $d_6$-DMSO): δ 11.23 (s, 1H), 7.81 (d, 2H, *J* = 7.2 Hz), 7.61 (d, 2H, *J* = 7.2 Hz), 7.34 (m, 3H), 7.25 (m, 2H), 7.24 (s, 1H), 4.57 (s, 1.2H), 4.40 (s, 0.8 H), 4.25 (d, 2H, *J* = 6.4 Hz), 4.13(t, 1H, *J* = 6.4 Hz), 2.66-3.27 (m, 4H), 1.66 (s, 3H). $^{13}$C NMR (400 MHz, $d_6$-DMSO): δ 171.24, 170.88, 168.08, 167.68, 164.86, 156.79, 156.54, 151.45, 144.33, 142.58, 141.20,

128.09, 127.54, 125.60, 120.59, 108.53, 65.96, 49.47, 48.14, 47.19, 12.37; ESI-MS m/z $[M+H]^+$: 506.89.

**C-Bz-OH:** yield 80%; $^1$H NMR (400 MHz, $d_6$-DMSO): δ 11.26 (s, 1H), 7.37-8.08 (m, 17H), 4.93 (s, 1.2H), 4.75 (s, 0.8 H), 4.35-4.41 (m, 3H), 4.07 (s, 2H), 2.72-3.10 (m, 4H). $^{13}$C NMR (400 MHz, $d_6$-DMSO): δ 171.23, 170.89, 167.51, 167.12, 163.75, 156.83, 156.54, 151.65, 144.36, 141.21, 133.66, 133.16, 128.90, 128.08, 127.54, 125.61, 120.59, 96.19, 65.98, 50.10, 48.20, 47.37, 47.18; ESI-MS m/z $[M+H]^+$:595.99.

**A-Bz-OH:** yield 88%; $^1$H NMR (400 MHz, $d_6$-DMSO): δ 11.26 (s, 1H), 7.37-8.08 (m, 17H), 4.93 (s, 1.2H), 4.75 (s, 0.8 H), 4.35-4.41 (m, 3H), 4.07 (s, 2H), 2.72-3.10 (m, 4H). $^{13}$C NMR (400 MHz, $d_6$-DMSO): δ 171.28, 170.77, 167.38, 166.90, 156.90, 156.61, 151.96, 145.97, 144.32, 141.21, 133.76, 132.98, 128.96, 128.08, 127.52, 125.57, 120.60, 65.98, 49.64, 48.18, 47.42, 47.20, 44.76; ESI-MS m/z $[M+H]^+$: 620.18.

**G-iBu-OH:** yield 83%; $^1$H NMR (400 MHz, $d_6$-DMSO): δ 11.98 (s, 1H), 11.56 (s, 0.5 H), 11.49 (s, 0.5H), 7.24-7.82 (m, 10H), 4.04 (s, 1.2H), 4.88 (s, 0.8 H), 4.16-4.29 (m, 3H), 3.94 (s, 2H), 2.26-3.42 (m, 5H), 1.03 (d, 6H). $^{13}$C NMR (400 MHz, $d_6$-DMSO): δ 180.59, 171.26, 170.85, 167.46, 166.92, 156.84, 155.334, 148.37, 144.32, 141.23, 128.07, 127.51, 125.57, 125.50, 120.59, 65.97, 48.30, 47.38, 47.20, 44.55, 35.12, 19.31; ESI-MS m/z $[M+H]^+$: 602.15.

**Manual solid phase synthesis of AP-PNA oligomers.** The experimental details for manual solid phase synthesis of AP-PNA oligomers are as follows. Rink

amide PEGA resin (50 mg, 0.012 mmol) was swelled in DMF/DCM (5/1) mixture for 2 h. The Fmoc group was deprotected by 20 % piperidine in DMF (5 min $\times$ 2). After washing, the resin was coupled with a mixture of Fmoc-Gly-OH (18 mg, 0.06 mmol), PyBOP (31 mg, 0.06 mmol), DIEA (20 μL, 0.12 mmol) for 1 h. After complete coupling which was monitored with Kaiser Test, the following synthetic cycles consisted of three steps: (1) removal of the Fmoc protecting group from the terminal amine using 20 % piperidine in DMF, (2) coupling of aegPNA monomers or AP-PNA monomers on the N-terminus of the growing PNA chain. The monomer was activated by PyBOP when coupled to a primary amine, i.e., to an aegPNA; Or it was activated by HATU when coupled to a secondary amine, i.e., to an AP-PNA. (3) capping the unreacted amines with acetic anhydride. The couplings were allowed for 4 to 6 hours at room temperature. After the final Fmoc deprotection, the terminal amine was capped by adding 5% acetic anhydride in DMF in the presence of DIEA for 15 min. The resin was washed thoroughly with DMF and DCM and dried *in vacuo*. For acyl protected nucleobases, the dried resin was treated with 35% ammonium overnight at room temperature to remove the acyl protecting groups and washed with DMF/DCM and dried *in vacuo*. For the Bhoc protected nucleobases, this step is not required. Final deprotection of Boc and cleavage from the resin was carried out by adding 2 mL 95% TFA (2.5% TIS, 2.5% $H_2O$) to the dry resin and the suspension was shaken for 1 h at room temperature. The cleavage cocktail was filtered and the resin was washed with TFA (1 mL $\times$ 2). The combined filtration was precipitated by adding 15 mL cold

diethyl ether and centrifuged to get the crude white solid. The crude white solid was redissolved in $H_2O$ and purified by semi preparative reverse-phase HPLC. The product was then lyophilized to get a white powder. After cleavage and HPLC purification, the overall yield of a typical synthesis was usually around 5%. The expected product was detected by MALDI-TOF mass spectroscopy.

**Modifying the peptoid amine of AP-PNA oligomers by acylation on solid phase.** After the terminal amine was capped by acetylation, the resin was washed with DMF and DCM, and dried thoroughly *in vacuo*. The dried resin was treated with a mixture of $BF_3$·Ether 2.5%, $CH_3COOH$ 5% in dry DCM (200 μL × 3) for a total time of 2 min to remove the side-chain Boc protecting group. The deprotection reaction was checked by the Kaiser test and the resin was washed with DCM and DMF. Then Boc-β-Ala-OH (2 eq) or D-Biotin (2 eq) which was activated by PyBOP (2 eq) in the presence of DIEA (10 eq) was added to the resin. The coupling time was usually 3-4 h to ensure complete coupling. After coupling, the subsequent acyl deprotection, cleavage, purification and characterization experiments were performed using the same method as in AP-PNA oligomer synthesis.

## 2.8 References

[1] P. A. Wender, D. J. Mitchell, K. Pattabiraman, E. T. Pelkey, L. Steinman, J. B. Rothbard, *Proceedings of the National Academy of Sciences of the United States of America* **2000**, *97*, 13003.

[2] J. B. Rothbard, E. Kreider, C. L. Vandeusen, L. Wright, B. L. Wylie, P. A. Wender, *Journal of Medicinal Chemistry* **2002**, *45*, 3612.

[3] K. Albertshofer, A. M. Siwkowski, E. V. Wancewicz, C. C. Esau, T. Watanabe, K. C. Nishihara, G. A. Kinberger, L. Malik, A. B. Eldrup, M. Manoharan, R. S. Geary, B. P. Monia, E. E. Swayze, R. H. Griffey, C. F. Bennett, M. A. Maier, *Journal of Medicinal Chemistry* **2005**, *48*, 6741.

[4] D. J. Mitchell, D. T. Kim, L. Steinman, C. G. Fathman, J. B. Rothbard, *Journal of Peptide Research* **2000**, *56*, 318.

[5] J. B. Rothbard, T. C. Jessop, R. S. Lewis, B. A. Murray, P. A. Wender, *Journal of the American Chemical Society* **2004**, *126*, 9506.

[6] M. Hansen, K. Kilk, U. Langel, *Advanced Drug Delivery Reviews* **2008**, *60*, 572.

[7] C. N. Carrigan, B. Imperiali, *Analytical Biochemistry* **2005**, *341*, 290.

[8] B. Hyrup, M. Egholm, P. E. Nielsen, P. Wittung, B. Norden, O. Buchardt, *Journal of the American Chemical Society* **1994**, *116*, 7964.

[9] P. E. Nielsen, G. Haaima, A. Lohse, O. Buchardt, *Angewandte Chemie International Edition in English* **1996**, *35*, 1939.

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

[11] A. Dragulescu-Andrasi, P. Zhou, G. He, D. H. Ly, *ChemComm* **2005**, 244.

[12] A. Dragulescu-Andrasi, S. Rapireddy, B. M. Frezza, C. Gayathri, R. R. Gil, D. H. Ly, *J Am Chem Soc* **2006**, *128*, 10258.

[13] Y. Yamamoto, J. Yoshida, T. Tedeschi, R. Corradini, S. Sforza, M. Komiyama, *Nucleic Acids Symposium Series* **2006**, *50*, 109.

[14] E. A. Englund, D. H. Appella, *Angewandte Chemie International Edition in English* **2007**, *46*, 1414.

[15] S. Sforza, G. Galaverna, A. Dossena, R. Corradini, R. Marchelli, *Chirality* **2002**, *14*, 591.

[16] G. Haaima, H. Rasmussen, G. Schmidt, D. K. Jensen, J. S. Kastrup, P. W. Stafshede, B. Norden, O. Buchardt, P. E. Nielsen, *New Journal of Chemistry* **1999**, *23*, 833.

[17] F. Debaene, N. Winssinger, *Organic Letters* **2003**, *5*, 4445.

[18] Z. Timar, L. Kovacs, G. Kovacs, Z. Schmel, *J. Chem. Soc., Perkin Trans. 1* **2000**, 19.

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

[20] D. W. Will, G. Breipohl, D. Langner, J. Knolle, E. Uhlmann, *Tetrahedron* **1995**, *51*, 12069.

[21] R. M. Osterman, B. A. McKittrick, T.-M. Chan, *Tetrahedron Letters* **1992**, *33*, 4867.

[22] R. H. E. Hudson, M. Goncharenko, A. P. Wallman, F. Wojciechowski, *Synlett* **2005**, 1442.

[23] S. A. Thomson, J. A. Josey, R. Cadilla, M. D. Gaul, C. F. Hassman, M. J. Luzzio, A. J. Pipe, K. L. Reed, D. J. Ricca, R. W. Wiethe, S. A. Noble, *Tetrahedron* **1995**, *51*, 6179.

[24] S. Pritz, Y. Wolf, C. Klemm, M. Bienert, *Tetrahedron Letters* **2006**, *47*, 5893.

[25] M. Lamothe, M. Perez, V. ColovrayGotteland, S. Halazy, *Synlett* **1996**, 507.

[26] H. Miel, S. Rault, *Tetrahedron Letters* **1998**, *39*, 1565.

[27] E. F. Evans, N. J. Lewis, I. Kapfer, G. Macdonald, R. J. K. Taylor, *Synthetic Communications* **1997**, *27*, 1819.

# Chapter 3 Hybridization and cellular uptake properties of AP-PNA, GP-PNA and AP-PNA derivatives

## 3.1 Hybridization properties of AP-PNA and GP-PNA

### 3.1.1 Triplex formation and stability

The sequences used here are PNA-$T_{10}$ oligomers which are known to form $(PNA)_2$/DNA triplexes with the complementary deoxyoligonucleotide $dA_{10}$. The C-terminus has a glycine amide residue and the N-terminus is acetylated. There are 1, 3 and 5 separate or alternately placed peptoid modifications in the $T_{10}$ sequence to study the effect of peptoid modification on hybridization property. For example, the sequence Ac-TTTTTT*TTTT-Gly-$NH_2$ means that one modification is introduced in the middle of the sequence; Ac-TT*TTTT*TTTT*-Gly-$NH_2$ means three modifications placed separately; Ac-TT*TT*TT*TT*TT*-Gly-$NH_2$ means five modifications that are placed alternately. The nomenclature AP-PNA-$T_{10}$ x-y or GP-PNA-$T_{10}$ x-y is used for any PNA-$T_{10}$ containing at least one amino-peptoid or guanidino-peptoid side chain. The first number “x” in the name specifies the length (the number of methylene carbons) of the side chain and the second “y” denotes the number of peptoid modifications in the PNA. For instance, AP-PNA-$T_{10}$ 2-1 means an AP-PNA-$T_{10}$ oligomer containing one 2-carbon amino-peptoid side chain in the sequence. Similarly, the superscript number on a residue specifies the length of a peptoid side chain on that residue and the

superscript letter "a" denotes the amino headgroup on the peptoid side chain, while "g" denotes the guanidino headgroup on the peptoid side chain. For example, the sequence Ac- TTTTTT$^{6g}$TTTT-Gly-$NH_2$ indicates the presence of one guanidinohexylpeptoid side chain on the thymine residue (in red) in the $T_{10}$ sequence.

### 3.1.1 Hybridization properties of (AP-PNA-$T_{10}$)$_2$/DNA triplexes

The AP-PNA-$T_{10}$ sequences with 1, 3 and 5 aminopeptoid side chains prepared from the 2- and 6-carbon series which are called aminoethyl-PNAs (entry **2**, **3** and **4**) and aminohexyl-PNAs (entry **5**, **6** and **7**) were prepared. All UV melting studies were performed with $dA_{10}$ and $T_{10}$ in 1:2 stoichiometries. The Tm values are presented in Table 3-1 and Figure 3-1.

**Table 3-1.** Melting temperature (Tm, $^{o}$C) of (AP-PNA)$_2$/DNA triplexes.

| Entry | Name of PNA | Sequence | DNA[a] Heating | DNA Cooling | ΔT [b] |
|---|---|---|---|---|---|
| 1 | aegPNA-$T_{10}$ | Ac-TTTTTTTTTT-$NH_2$ | 70.7 (0)[c] | 57 | 13.7 |
| 2 | AP-PNA-$T_{10}$ 2-1 | Ac- TTTTT$^{2a}$TTTT-Gly-$NH_2$ | 62.6 (4.05) [c] | 52.5 | 10.1 |
| 3 | AP-PNA-$T_{10}$ 2-3 | Ac- T$^{2a}$TTT$^{2a}$TTT$^{2a}$-Gly-$NH_2$ | 52 (3.1) [c] | 50 | 2 |
| 4 | AP-PNA-$T_{10}$ 2-5 | Ac- T$^{2a}$T$^{2a}$TT$^{2a}$TT$^{2a}$T$^{2a}$-Gly-$NH_2$ | 39.6 (3.1) [c] | 40.9 | -1.3 |
| 5 | AP-PNA-$T_{10}$ 6-1 | Ac- TTTTT$^{6a}$TTTT-Gly-$NH_2$ | 68.9 (0.9) [c] | 56.6 | 12.3 |
| 6 | AP-PNA-$T_{10}$ 6-3 | Ac- T$^{6a}$TTT$^{6a}$TTT$^{6a}$-Gly-$NH_2$ | 66.4 (0.75) [c] | 60.3 | 6.1 |
| 7 | AP-PNA-$T_{10}$ 6-5 | Ac- T$^{6a}$T$^{6a}$TT$^{6a}$T$^{6a}$T$^{6a}$-Gly-$NH_2$ | 59.3 (1.1) [c] | 58.1 | 1.2 |

[a] 5'd(AAAAAAAAAA); [b] Difference of Tm values between heating and cooling; [c] ΔTm per modified unit given in parentheses.

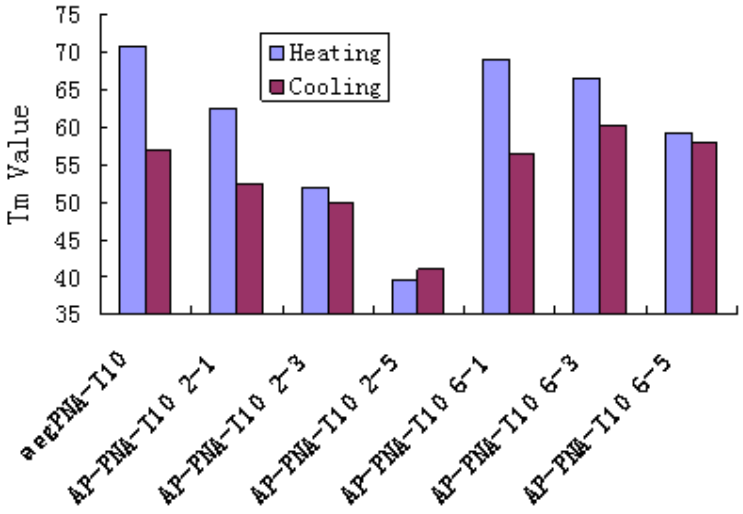



**Figure 3-1. Tm values of $(AP\text{-}PNA)_2$/DNA triplexes.**

As shown in Table 3-1 and Figure 3-1, the presence of a single aminoethyl-PNA residue in the middle of the $T_{10}$ sequence, i.e., AP-PNA-$T_{10}$ 2-1, has a detrimental effect on the stability of the $(AP\text{-}PNA\text{-}T_{10})_2$/DNA triplex (Table 3-1, entry 2). The Tm of the complex decreases by 8.1 $^{o}$C, or 4.05 $^{o}$C per aminoethylPNA unit when taking into account the two AP-PNA chains in the triplex, as compared to $(aegPNA)_2$/DNA triplex. The detrimental effect becomes severer when increasing the number of aminoethyl-PNA units in the sequence, as the Tm of $(AP\text{-}PNA\text{-}T_{10}\ 2\text{-}3)_2$/DNA triplex (Table 3-1, entry 3) and for $(AP\text{-}PNA\text{-}T_{10}\ 2\text{-}5)_2$/DNA triplex (Table 3-1, entry 4) is 18.7 $^{o}$C and 31.1 $^{o}$C, respectively, lower than that of $(aegPNA)_2$/DNA triplex. Interestingly, the cost per aminoethyl-PNA unit is the same (3.1 $^{o}$C) for both AP-PNA-$T_{10}$ 2-3 and AP-PNA-$T_{10}$ 2-5, which is lower than that for AP-PNA-$T_{10}$ 2-1. So the detrimental effect of incorporating aminoethyl-PNA units seems to be “additive”.

The presence of aminohexyl side chains in the $T_{10}$ sequence, on the other hand, generally shows a less detrimental effect than that of aminoethyl side chains. As seen from Table 3-1, introducing a single aminohexyl-PNA unit in the middle of $T_{10}$ sequence, i.e., AP-PNA-$T_{10}$ 6-1 (Table 3-1, entry 5), reduced the stability of its (AP-PNA-$T_{10}$)$_2$/DNA triplex by only 0.9 $^{o}$C per unit with respective to (aegPNA-$T_{10}$)$_2$/DNA. Introducing 3 aminohexyl-PNA units results in a cost of 0.75 $^{o}$C per unit (Table 3-1, entry 6), while having 5 such residues gives a higher cost of 1.1 $^{o}$C per unit (Table 3-1, entry 7). This may be due to the larger steric crowding effects exerted by the 10 aminohexyl side chains from the two AP-PNA-$T_{10}$ 6-5 copies in the triplex.

For a given number of modifications in the sequences, a linear increase in thermal stability of (AP-PNA-$T_{10}$)$_2$/DNA triplexes is observed from 2-carbon to 6-carbon amino-peptoid side-chain length. For example, the Tm value of (AP-PNA-$T_{10}$ 6-1)$_2$/DNA triplex is 6.3 $^{o}$C higher than that of (AP-PNA-$T_{10}$ 2-1)$_2$/DNA triplex; the Tm value of (AP-PNA-$T_{10}$ 6-3)$_2$/DNA triplex is 14.4 $^{o}$C higher than that of (AP-PNA-$T_{10}$ 2-3)$_2$/DNA triplex; and the Tm value of (AP-PNA-$T_{10}$ 6-5)$_2$/DNA triplex is 19.7 $^{o}$C higher than that of (AP-PNA-$T_{10}$ 2-5)$_2$/DNA triplex. This linear increase in $\Delta$Tm values with the side-chain length for a given number of positive charges (note that the amino group is positively charged at neutral pH) indicates that the length of side chain plays an important role in the stability of (AP-PNA-$T_{10}$)$_2$/DNA triplexes and that a longer side chain is better tolerated than a shorter side chain. The positive charges on the aminoalkyl

side chains do not seem to be involved in electrostatic attractions with the negatively charged DNA backbone in the (AP-PNA-$T_{10}$)$_2$/DNA triplexes, which would otherwise increase the thermal stability of the triplexes when increasing the number of positive charges.

Despite the negative effects of the aminoalkyl modifications on the thermal stability of (AP-PNA-$T_{10}$)$_2$/DNA triplex, one interesting result is that the strong hysteresis seen in (aegPNA-$T_{10}$)$_2$/DNA triplex is well suppressed in the (AP-PNA-$T_{10}$)$_2$/DNA triplexes. The hysteresis herein is represented by the difference between the melting temperature measured in the heating process and the annealing temperature measured in the cooling process. For a given side-chain length, the hysteresis becomes smaller with an increasing number of modifications in the sequence. For example, the $\Delta$T value of (AP-PNA-$T_{10}$ 2-1)$_2$/DNA triplexes is 10.1 $^{o}$C (Table 3-1, entry 2) compared to $\Delta$T of 13.7 $^{o}$C for (aegPNA-$T_{10}$)$_2$/DNA triplexes (Table 3-1, entry 1). The $\Delta$T values of (AP-PNA-$T_{10}$ 2-3)$_2$/DNA and (AP-PNA-$T_{10}$ 2-5)$_2$/DNA triplexes are 2 $^{o}$C and -1.3 $^{o}$C (Table 3-1, entry 3 and 4), respectively. For the aminohexyl series, the $\Delta$T values of (AP-PNA-$T_{10}$ 6-1)$_2$/DNA, (AP-PNA-$T_{10}$ 6-3)$_2$/DNA and (AP-PNA-$T_{10}$ 6-5)$_2$/DNA triplexes are 12.3, 6.1 and 1.2 $^{o}$C (Table 1, entry 5, 6 and 7), respectively. Although the 6-carbon series do not display a smaller hystersis than the 2-carbon series, they exhibit significantly higher annealing temperatures. In fact, their annealing temperature is either comparable to or even higher than that of the (aegPNA-$T_{10}$)$_2$/DNA triplexes. These results indicate that the hysteresis is suppressed as result of increasing positive

charges on AP-PNA which provide electrostatic attractions between PNA and DNA and help bring them together during the triplex formation process. The strong hysteresis in the hybridization studies of $(aegPNA)_2$/DNA triplexes is due to a slow rate of forming the $(aegPNA)_2$/DNA triplex, i.e., slow annealing, under physiological conditions, which is one of the major obstacles in using PNA as an antigene agent.

The above data indicate that, although the electrostatic attractions between the AP-PNA and DNA do not seem to play a role in contributing to the stability of $(AP\text{-}PNA\text{-}T_{10})_2$/DNA triplexes, they are favorable for the formation of these $(AP\text{-}PNA\text{-}T_{10})_2$/DNA triplexes and hence can facilitate their formation rate.

### 3.1.2 Hybridization properties of $(GP\text{-}PNA\text{-}T_{10})_2$/DNA triplexes

To demonstrate the effect of guanidino groups on the stability of $(GP\text{-}PNA)_2$/DNA triplexes, all the amino groups in AP-PNA-$T_{10}$ oligomers shown in Table 3-1 were converted to guanidino groups to give the corresponding GP-PNA-$T_{10}$ oligomers. Their hybridization properties with the complementary $dA_{10}$ are presented in Table 3-2 and Figure 3-2.

**Table 3-2.** Melting temperature (Tm, **°C**) of $(GP\text{-}PNA)_2$-DNA triplexes[a].

| Entry | Name of PNA | Sequence | DNA[a] Heating | DNA Cooling | ΔTm [b] |
|---|---|---|---|---|---|
| 1 | aegPNA-$T_{10}$ | Ac-TTTTTTTTTT-$NH_2$ | 70.7 (0) [c] | 57 | 13.7 |
| 8 | GP-PNA-$T_{10}$ 2-1 | Ac- TTTTT$T^{2g}$TTTT-Gly-$NH_2$ | 62.8 (3.95) [c] | 54 | 8.8 |
| 9 | GP-PNA-$T_{10}$ 2-3 | Ac- T$T^{2g}$TTT$T^{2g}$TTT$T^{2g}$-Gly-$NH_2$ | 55.6 (2.5) [c] | 51.2 | 4.4 |
| 10 | GP-PNA-$T_{10}$ 2-5 | Ac- T$T^{2g}$T$T^{2g}$T$T^{2g}$T$T^{2g}$T$T^{2g}$-Gly-$NH_2$ | 49.5 (2.1) [c] | 49.3 | 0.2 |
| 11 | GP-PNA-$T_{10}$ 6-1 | Ac- TTTTT$T^{6g}$TTTT-Gly-$NH_2$ | 69.7 (0.5) [c] | 55.1 | 14.6 |
| 12 | GP-PNA-$T_{10}$ 6-3 | Ac- T$T^{6g}$TTT$T^{6g}$TTT$T^{6g}$-Gly-$NH_2$ | 68.7 (0.33) [c] | 62.8 | 5.9 |
| 13 | GP-PNA-$T_{10}$ 6-5 | Ac- T$T^{6g}$T$T^{6g}$T$T^{6g}$T$T^{6g}$T$T^{6g}$-Gly-$NH_2$ | 62.5 (0.8) [c] | 61.3 | 1.2 |

[a] 5'd(AAAAAAAAAA); [b] the difference of Tm value between heating and cooling; [c] ΔTm per modified unit given in parentheses.

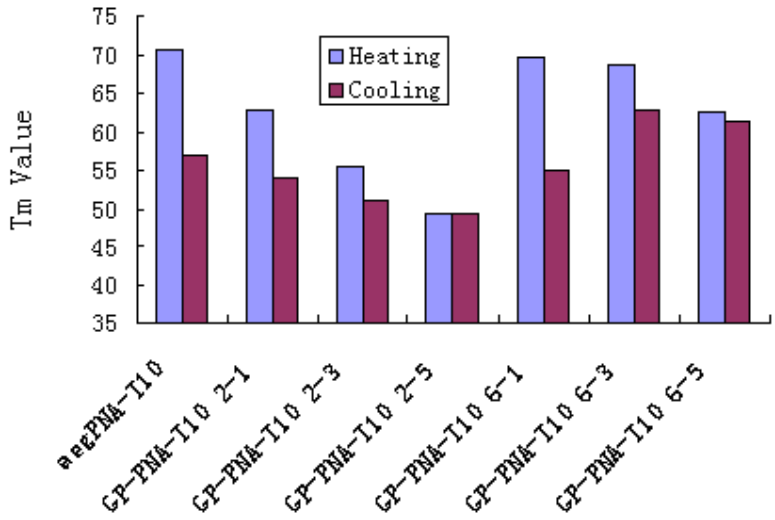



**Figure 3-2. Tm values of $(GP\text{-}PNA)_2$-DNA triplexes.**

As shown in Table 3-2 and Figure 3-2, similar trends as in the $(AP\text{-}PNA\text{-}T_{10})_2$/DNA triplexes are obtained in the $(GP\text{-}PNA\text{-}T_{10})_2$/DNA triplexes. First of all, the presence of a GP-PNA unit in the middle of $T_{10}$ sequence results in a decrease in the thermal stability of $(GP\text{-}PNA\text{-}T_{10})_2$/DNA triplex, and the negative effect becomes severer when increasing the number of modifications. Second, for a given number of modifications, the longer the side chain, the higher the Tm value.

For example, the Tm value of (GP-PNA-$T_{10}$ 6-1)$_2$/DNA triplex is 6.9 $^{o}$C higher than that of (GP-PNA-$T_{10}$ 2-1)$_2$/DNA triplex (Table 3-2, entry 11 and 8), which is similar to that between (AP-PNA-$T_{10}$ 6-1)$_2$/DNA triplex and (AP-PNA-$T_{10}$ 2-1)$_2$/DNA triplex (Table 3-1, entry 5 and 2). Furthermore, for a given side-chain length, increasing the number of cationic groups favors the formation of (GP-PNA-$T_{10}$)$_2$/DNA triplexes and hence increases the formation rate. For example, the $\Delta$T values (differences between the heating and cooling) of (GP-PNA-$T_{10}$ 2-1)$_2$/DNA triplexes, (GP-PNA-$T_{10}$ 2-3)$_2$/DNA and (GP-PNA-$T_{10}$ 2-5)$_2$/DNA triplexes are 8.8, 4.4 and 0.2 $^{o}$C (Table 3-2, entry 8, 9 and 10), respectively. The $\Delta$T values of (GP-PNA-$T_{10}$ 6-1)$_2$/DNA triplexes, (GP-PNA-$T_{10}$ 6-3)$_2$/DNA and (GP-PNA-$T_{10}$ 6-5)$_2$/DNA triplexes are 14.6, 5.9 and 1.2 $^{o}$C (Table 3-2, entry 11, 12 and 13), respectively. Remarkably, the annealing temperatures of (GP-PNA-$T_{10}$ 6-3)$_2$/DNA triplex (62.8 $^{o}$C) and (GP-PNA-$T_{10}$ 6-5)$_2$/DNA triplex (61.3 $^{o}$C) are even higher than that of (aegPNA)$_2$/DNA triplex (57 $^{o}$C).

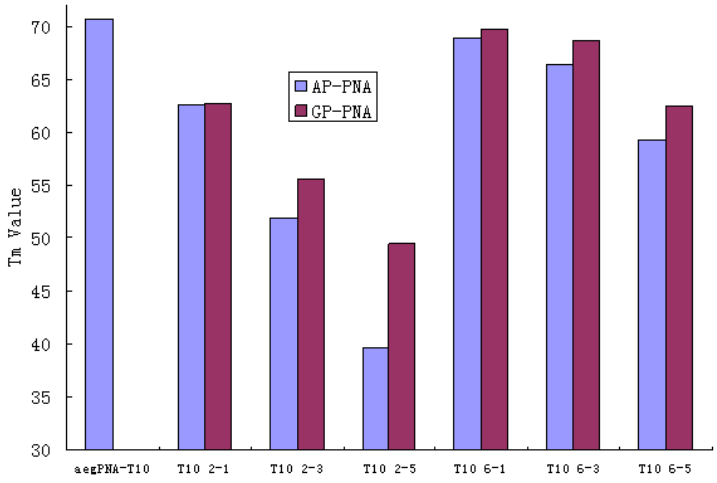



**Figure 3-3. The comparison of Tm values (heating) between (AP-PNA)$_2$/DNA and (GP-PNA)$_2$/DNA triplexes.**

Comparing the data in Table 3-1 and Table 3-2, and as shown in Figure 3-3, for a given side-chain length, converting the amino group to guanidino group results in an increase of the $(PNA\text{-}T_{10})_2$/DNA triplex's stability. For example, the Tm value of $(GP\text{-}PNA\text{-}T_{10}$ 2-5$)_2$/DNA triplex is about 10 $^oC$ higher than that of $(AP\text{-}PNA\text{-}T_{10}$ 2-5$)_2$/DNA triplex (Table 3-2, entry 10 and Table 3-1, entry 4) and the Tm value of $(GP\text{-}PNA\text{-}T_{10}$ 6-5$)_2$/DNA triplex is about 3.2 $^oC$ higher than that of $(AP\text{-}PNA\text{-}T_{10}$ 6-5$)_2$/DNA triplex (Table 3-2, entry 13 and Table 3-1, entry 7). Moreover, the cost per guanidinoethylPNA unit and guanidinohexylPNA unit is less than that of the corresponding aminoethylPNA unit and aminohexylPNA unit. For example, the cost per guanidinoethylPNA unit for the $GP\text{-}PNA\text{-}T_{10}$ 2-3 oligomer is 2.5 $^oC$ whereas the cost per aminoethylPNA unit for the $AP\text{-}PNA\text{-}T_{10}$ 2-3 oligomer is 3.1 $^oC$ (Table 3-2, entry 9 and Table 3-1, entry 3). The cost per guanidinohexylPNA unit for the $GP\text{-}PNA\text{-}T_{10}$ 6-3 oligomer is 0.33 $^oC$ whereas the cost per aminohexylPNA unit for the $AP\text{-}PNA\text{-}T_{10}$ 6-3 oligomer is 0.75 $^oC$ (Table 3-2, entry 12 and Table 3-1, entry 6). Therefore it is safe to conclude that in these PNA series, converting the amino to guanidino headgroup improves the thermal stability of the corresponding $PNA_2$/DNA triplexes. Converting the amino to guanidino headgroup increases the size of the side chain, and a larger side chain might play a favorable role in stabilizing the peptoid-$PNA_2$/DNA triplexes. It is also possible that, although their net charge is equal, a guanidino group provides more hydrogen bond donors than does an amino group.

### 3.1.3 Conclusions on the triplex formation and stability

In summary, we have demonstrated that for a given positively charged group, the Tm values of $(PNA)_2$/DNA triplexes increased with a longer side chain length. And for a side chain with a given number of methylene groups, the guanidino headgroup has a beneficial effect on the stability of $(PNA)_2$/DNA triplexes as compared with the amino headgroup. Therefore, although the $N^{\gamma}$ modifications are detrimental to the triplexes stability, which is consistent with what is observed with $N^{\gamma}$-methyl modification,[1] this can be compensated by introducing a longer side chain with a guanidino headgroup. Furthermore, we have demonstrated that the introduction of positively charged groups on the peptoid side chain would promote the formation of $(PNA\text{-}T_{10})_2$/DNA triplexes.

## 3.2 Duplex formation and stability

A decamer sequence Ac-GTAGATCACT-Gly-$NH_2$ was used in the duplex formation and stability study. Various peptoid modifications have been done on the decamer sequence. For instance, the sequence Ac-GTAGAT*CACT-Gly-$NH_2$ means one modification at the marked thymine residue; Ac-GTAGA*TC*ACT-Gly-$NH_2$ means two modifications at the marked adenine residue and cytosine residue; Ac-G*TAGA*TC*ACT-Gly-$NH_2$ means that three modifications were introduced separately; Ac-GT*AG*AT*CA*CT-Gly-$NH_2$ means that four modifications were introduced alternatively and Ac-GT*AG*AT*CA*CT*-Gly-$NH_2$ sequence means that five modifications were

introduced alternatively. The peptoid modifications in these PNAs are different in side-chain length and headgroups. The nomenclature AP-PNA x-y or GP-PNA x-y is used for the PNA sequence containing at least one amino-peptoid or guanidino-peptoid side chain modification. "x" denotes the number of methylene carbons of side chain and "y" is the number of modifications in the PNA sequence. These peptoid-PNA oligomers as well as the unmodified aegPNA were hybridized with fully matched antiparallel DNA or RNA, fully matched parallel DNA, and antiparallel DNA or RNA with a single mismatched base pair ($T_{PNA}$-$G_{DNA/RNA}$) in the middle of sequence. Their hybridization data were used to systematically examine the effect of side-chain length, the number of modifications and the nature of the headgroup on the duplex stability, orientation preferences and sequence selectivity. The sequences for the antiparallel complementary DNA and RNA are 5'd(AGTGATCTAC) and 5'(AGUGAUCUAC), for parallel complementary DNA is 5'd(CATCTAGTGA) and for mismatched DNA and RNA are 5'd(AGTGGTCTAC) and 5'(AGUGGUCUAC), respectively.

### 3.2.1 Hybridization properties of AP-PNA containing one peptoid side chain of varying length in the middle of the decamer sequence

In the previous section, we have found that the side chain length seems to play an important role in the thermal stability of $(PNA\text{-}T_{10})_2$/DNA triplexes. To determine whether the length of the peptoid side chain also has an influence on the

thermal stability of the PNA/DNA or PNA/RNA duplexes, a set of five AP-PNAs containing a single aminopeptoid thymine residue at the middle of the sequence were first synthesized. The spacer length between the peptoid amino headgroup and the backbone in the five AP-PNAs is 2, 3, 4, 5 and 6 carbons, respectively. Table 3-3 and Figure 3-4 present the hybridization data of these AP-PNAs in comparison with the unmodified aegPNA.

**Table 3-3.** The thermal stability data (Tm, ºC) of AP-PNA/DNA and AP-PNA/RNA duplexes. The decamer AP-PNAs contain a single amino-peptoid side chain at the thymine residue at position 6.

| Entry | Name of PNA | Sequence | DNA (ap) | DNA(p) | DNA (m) | RNA (ap) | RNA (m) |
|---|---|---|---|---|---|---|---|
| 1 | aegPNA | Ac-GTAGATCACT-Gly-$NH_2$ | 48.6 (0)[a] | 34.9 | 36.4 (12.2) [b] | 51.9 (0)[a] | 42.7 (9.2) [c] |
| 14 | AP-PNA 2-1 | Ac-GTAGA**T**$^{2a}$CACT-Gly-$NH_2$ | 40.2 (-8.4)[a] | 34.4 | 29.0 (11.2) [b] | 48.9 (-3)[a] | 41.0 (7.9) [c] |
| 15 | AP-PNA 3-1 | Ac-GTAGA**T**$^{3a}$CACT-Gly-$NH_2$ | 43.6 (-5)[a] | 32.7 | 30.0 (10.9) [b] | 50.1 (-0.8)[a] | 40.3 (9.8) [c] |
| 16 | AP-PNA 4-1 | Ac-GTAGA**T**$^{4a}$CACT-Gly-$NH_2$ | 46.1 (-2.5)[a] | 31.5 | 31.1 (15) [b] | 51.5 (-0.4)[a] | 42.2 (9.3) [c] |
| 17 | AP-PNA 5-1 | Ac-GTAGA**T**$^{5a}$CACT-Gly-$NH_2$ | 49.1 (+0.5)[a] | 34.9 | 34.4 (14.7) [b] | 52.0 (+0.1)[a] | 43.3 (8.7) [c] |
| 18 | AP-PNA 6-1 | Ac-GTAGA**T**$^{6a}$CACT-Gly-$NH_2$ | 48.9 (+0.3)[a] | 33.9 | 33.8 (14.8) [b] | 51.8 (-0.1)[a] | 39.3 (12.5) [c] |

[a] ΔTm between aegPNA and modified PNA given in parentheses. [b] ΔTm between the perfect match PNA/DNA(ap) duplex and the corresponding PNA/DNA duplex with a single base mismatch given in parentheses. [c] ΔTm between the perfect match PNA/RNA(ap) duplex and the corresponding PNA/RNA duplex with a single base mismatch given in parentheses.

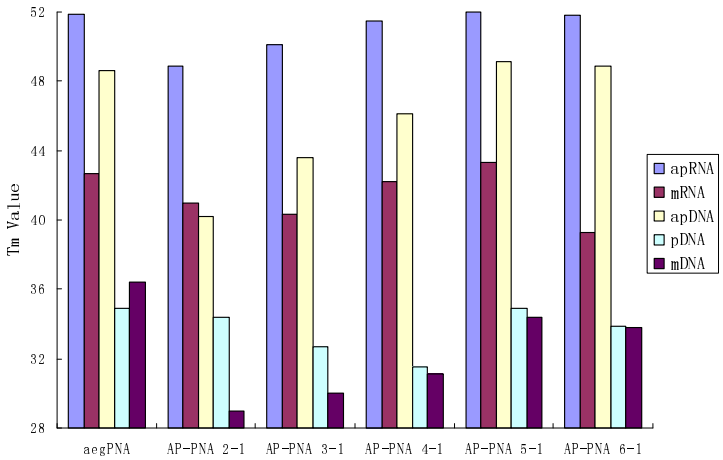



**Figure 3-4 Tm values of AP-PNA/DNA and AP-PNA/RNA duplexes.**

The data in Table 3-3 and Figure 3-4 clearly show that the length of the aminopeptoid side chain plays a critical role in determining the thermal stability of the AP-PNA/DNA(ap) duplex. Inclusion of a shorter amino-peptoid side chain (n = 2, 3 or 4) has a negative effect on the thermal stability of the AP-PNA/DNA(ap) duplex. For example, compared with the Tm value of aegPNA/DNA(ap), AP-PNA 2-1 with a 2-carbon spacer causes 8.4 ºC drop (Table 3-3, entry 14), AP-PNA 3-1 with a 3-carbon spacer causes 5 $^{o}$C drop (Table 3-3, entry 15), and AP-PNA 4-1 with a 4-carbon spacer causes 2.5 $^{o}$C drop (Table 3, entry16). The negative effect becomes less obvious with the increase of the side chain length. When the aminopeptoid side chain is increased to 5 or 6 carbons, the introduced $N^{\gamma}$-aminopentyl or $N^{\gamma}$-aminohexyl group maintains the high thermal stability of the AP-PNA/DNA(ap) duplexes. As seen in Table 3 (entry 17 and 18), the resultant AP-PNA/DNA(ap) duplex has a Tm that is 0.3 and 0.5 $^{o}$C higher than that of aegPNA/DNA(ap), respectively. As shown in Figure 3-4, a trend can be observed that the hybridization affinity increases with an increasing length of the peptoid side chain with a cap reached at the 5 or 6-carbon length.

A similar spacer length-dependent effect on the thermal stability of AP-PNA/RNA(ap) duplexes is observed. The Tm values of AP-PNA/RNA(ap) duplexes increase with the side-chain length until 5 to 6 carbons side chain length. It is worth noticing that the Tm values of AP-PNA 5-1/DNA(ap) and AP-PNA 5-1/RNA(ap) duplexes as well as AP-PNA 6-1/DNA(ap) and AP-PNA 6-1/RNA(ap) duplexes are similar to that of aegPNA. Interestingly, like aegPNA,

the same AP-PNA displays a stronger binding affinity towards RNA than towards DNA.

The antiparallel binding preferences were examined by hybridizing these AP-PNAs to the complementary DNA in both antiparallel and parallel orientations. The data show that AP-PNAs preserve the antiparallel binding preference, as the affinities of AP-PNAs to the parallel DNA sequence are significantly reduced (Table 3-3, column 5).

The discrimination ability of AP-PNAs to the mismatched sequences was also evaluated. The thermal stabilities of AP-PNA/DNA(m) and AP-PNA/RNA(m) are largely reduced as compared with the perfect match AP-PNA/DNA(ap) and AP-PNA/RNA(ap). For example, the Tm values decrease by 10.9 ~ 15 $^{o}$C with a single base pair mismatch for DNA binding (Table 3-3, column 5) and 7.9 ~ 12.5 $^{o}$C with a single base pair mismatch for RNA binding (Table 3-3, column 7). Therefore, introducing an amino-peptoid side chain onto aegPNA does not compromise the binding specificity of the modified PNAs and these AP-PNAs maintain the ability to discriminate closely related sequences. Among these, AP-PNA with a 6-carbon spacer shows the best discrimination ability.

### 3.2.2 Hybridization properties of GP-PNA containing one peptoid side chain of varying length in the middle of the decamer sequence

All the AP-PNAs shown in Table 3-3 were converted to the five corresponding GP-PNAs. Thus a set of GP-PNAs containing a single guanidinopeptoid thymine residue at the middle of the sequence were obtained. The spacer length between the peptoid guanidino headgroup and the backbone in the five GP-PNAs is of 2, 3, 4, 5 or 6 carbons, respectively. The hybridization data of these GP-PNAs in comparison with the unmodified aegPNA were presented in Table 3-4 and Figure 3-5.

**Table 3-4.** Thermal stability data (Tm, ℃) of GP-PNA/DNA and GP-PNA/RNA duplexes. The decamer GP-PNAs contain a single guanidino-peptoid side chain at the thymine residue at position 6.

| Entry | Name of PNA | Sequence | DNA (ap) | DNA (p) | DNA (m) | RNA (ap) | RNA (m) |
|---|---|---|---|---|---|---|---|
| 1 | aegPNA | Ac-GTAGATCACT-Gly-$NH_2$ | 48.6 (0)[a] | 34.9 | 36.4 (12.2)[b] | 51.9(0)[a] | 42.7 (9.2)[c] |
| 19 | GP-PNA 2-1 | Ac-GTAGA$T^{2g}$CACT-Gly-$NH_2$ | 41.5 (-7.1)[a] | 31.5 | 27.3 (14.2)[b] | 50.3 (-1.6)[a] | 39.3 (11)[c] |
| 20 | GP-PNA 3-1 | Ac-GTAGA$T^{3g}$CACT-Gly-$NH_2$ | 45.4 (-3.2)[a] | 31.5 | 31.9 (13.5)[b] | 50.4 (-1.5)[a] | 40.3 (10.1)[c] |
| 21 | GP-PNA 4-1 | Ac-GTAGA$T^{4g}$CACT-Gly-$NH_2$ | 48 (-0.6)[a] | 33.3 | 33.9 (14.1)[b] | 52.5 (+0.6)[a] | 39.5 (13)[c] |
| 22 | GP-PNA 5-1 | Ac-GTAGA$T^{5g}$CACT-Gly-$NH_2$ | 49.4 (+0.8)[a] | 35 | 34.3 (15.1)[b] | 53.2 (+1.3)[a] | 42.7 (10.5)[c] |
| 23 | GP-PNA 6-1 | Ac-GTAGA$T^{6g}$CACT-Gly-$NH_2$ | 48.5 (-0.1)[a] | 32.8 | 27.9 (20.6)[b] | 52.4 (+0.5)[a] | 42.7 (9.7)[c] |

[a] ΔTm between aegPNA and modified PNA given in parentheses. [b] ΔTm between the perfect match PNA/DNA(ap) duplex and the corresponding PNA/DNA duplex with a single base mismatch given in parentheses. [c] ΔTm between the PNA/RNA(ap) duplex and the corresponding PNA/RNA duplex with a single base mismatch given in parentheses.

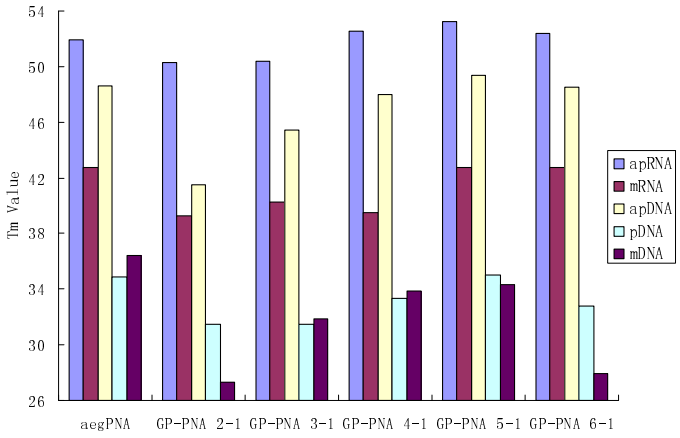



**Figure 3-5. Tm values of GP-PNA/DNA and GP-PNA/RNA duplexes.**

As shown in Table 3-4 and Figure 3-5, the length of the guanidinopeptoid side chain again plays a critical role in determining the thermal stability of the GP-PNA/DNA(ap) and GP-PNA/RNA(ap) duplexes. For example, inclusion of a shorter guanidino-peptoid side chain (n = 2, 3, and 4) results in a decrease in the Tm values of GP-PNA/DNA(ap) duplexes of 7.1 $^{o}$C, 3.2 $^{o}$C and 0.6 $^{o}$C, respectively. However, inclusion of a 5 or 6-carbon spacer gives the GP-PNA with a DNA-binding affinity that is slightly higher than or similar to that of aegPNA (Table 3-4, entry 22 and 23). The trend of the Tm values of GP-PNA/RNA(ap) duplexes are similar with that of GP-PNA/DNA(ap). It increases with the side-chain length till 5 methylene carbons and slightly decreases for 6 methylene carbons. Thus, the general trend that the hybridization affinity increases with an increasing length of the peptoid side chain with a cap reached at the 5 or 6-carbon length is also observed here for the GP-PNAs.

Their antiparallel binding preference was also preserved. As seen in Table 3-4 (column 5), the Tm values of GP-PNA/DNA(p) duplexes decrease significantly (Table 3-4, column 5) as compared with those of GP-PNA/DNA(ap) duplexes.

Also, GP-PNAs exhibit significantly lower binding affinity towards single base pair mismatched DNA (by 13.5 ~ 20.6 $^{o}$C, Table 3-4, column 6) as well as toward single base pair mismatched RNA (by 9.7 ~ 13 $^{o}$C, Table 3-4, column 8). Therefore introducing a guanidino-peptoid side chain onto aegPNA does not compromise the binding specificity of the modified PNA. As shown in Table 3-4 (column 6), the sequence selectivity increases with the side chain length and a similar increasing trend was also found in the case of AP-PNAs (Table 3-3, column 6). Since the headgroup is the same (amino or guanidino), the increasing discrimination ability is also a consequence of the increased in the length of the peptoid side chain.

Comparing the hybridization data of AP-PNAs and the corresponding GP-PNAs (Table 3-3 and 3-4), one can also find that for a given side chain length (with the exception of the 6-carbon side chain), converting the amino to guanidino headgroup resulted in an increase in Tm value. For example, the Tm value towards DNA(ap) increases by 1.3 $^{o}$C for the 2-carbon spacer (Table 3-3, entry 14 and Table 3-4, entry 19), 1.8 $^{o}$C for the 3-carbon spacer (Table 3-3, entry 15 and Table 3-4, entry 20), 1.9 $^{o}$C for the 4-carbon spacer (Table 3-3, entry 16 and Table 3-4, entry 21) and 0.3 $^{o}$C for the 5-carbon spacer (Table 3-3, entry 17 and Table 3-4, entry 22). These data indicate that introducing a guanidinopeptoid side chain has a

beneficial effect on the thermal stabilities of the GP-PNA/DNA(ap) duplexes, which is also consistent with what was found with the $(GP\text{-}PNA\text{-}T_{10})_2$/DNA triplexes. By and large, GP-PNAs also discriminate better against base-pair mismatches than do AP-PNAs.

Therefore, it is safe to draw the following conclusions: (1) Although $N^{\gamma}$ modification tends to destabilize the PNA/DNA duplexes, the negative effects can be compensated by introducing a longer peptoid side chain, such as those with a 5 or 6-carbon spacer. (2) The hybridization affinity of PNA/DNA(ap) and PNA/RNA(ap) duplexes increases with an increasing length of the peptoid side chain with a cap at the 5 or 6-carbons. (3) A guanidino headgroup will increase the stability of the PNA/DNA(ap) and PNA/RNA(ap) duplexes as compared to an amino headgroup. (4) The increase in side-chain length benefits not only the binding affinity of peptoid PNAs, but also the sequence selectivity.

### 3.2.3 Hybridization properties of AP-PNA containing multiple peptoid side chains of varying length in the decamer sequence

To examine the cumulative effect of amino-peptoid or guanidino-peptoid side chains on the stability of PNA/DNA and PNA/RNA duplexes, PNA sequences containing multiple amino-peptoid side chains with the 2, 3, 4 or 6 carbon

side-chain spacer were prepared. Their detailed hybridization properties were evaluated and shown in Table 3-5 and Table 3-6.

**Table 3-5.** Effects of multiple amino-peptoid modifications on the thermal stability (Tm, ºC) of AP-PNA/DNA and AP-PNA/RNA duplexes.

| Entry | Name of PNA | Sequence | DNA (ap) | DNA(p) | DNA (m) | RNA (ap) | RNA (m) |
|---|---|---|---|---|---|---|---|
| 1 | aegPNA | Ac-GTAGATCACT-Gly-$NH_2$ | 48.6 (0)[a] | 34.9 | 36.4 (12.2)[b] | 51.9 (0)[a] | 42.7 (9.2)[c] |
| 14 | AP-PNA 2-1 | Ac-GTAGA**T$^{2a}$**CACT-Gly-$NH_2$ | 40.2 (-8.4)[a] | 34.4 | 29.0 (11.2)[b] | 48.9 (-3)[a] | 41.0 (7.9)[c] |
| 24 | AP-PNA 2-2 | Ac-GTAGA**$^{2a}$** T**C$^{2a}$** ACT-Gly-$NH_2$ | 39 (-9.6)[a] | 31.9 | 31.6 (7.4)[b] | 44.5 (-7.4)[a] | 37.6 (6.9)[c] |
| 25 | AP-PNA 2-3 | Ac-**G$^{2a}$** TAGA**$^{2a}$** T**C$^{2a}$** ACT-Gly-$NH_2$ | 36.2 (-12.4)[a] | 32.9 | 31.4 (4.8)[b] | 42.3 (-9.6)[a] | 35.6 (6.7)[c] |
| 26 | AP-PNA 2-4 | Ac-G**T$^{2a}$**A**G$^{2a}$**A**T$^{2a}$**C**A$^{2a}$**C T-Gly-$NH_2$ | 31.1 (-17.5)[a] | 31.4 | 29 (2.1)[b] | 36.9 (-15)[a] | 33.9 (3)[c] |
| 15 | AP-PNA 3-1 | Ac-GTAGAT**$^{3a}$**CACT-Gly-$NH_2$ | 43.6 (-5)[a] | 32.7 | 30.0 (10.9)[b] | 50.1 (-0.8)[a] | 40.3 (9.8)[c] |
| 27 | AP-PNA 3-2 | Ac-GTAGA**$^{3a}$** T**C$^{3a}$** ACT-Gly-$NH_2$ | 40.6 (-8.2)[a] | 33.8 | 25.9 (14.7)[b] | 50.4 (-1.5)[a] | 40.9 (9.5)[c] |
| 28 | AP-PNA 3-3 | Ac-**G$^{3a}$** TAGA**$^{3a}$** T**C$^{3a}$** ACT-Gly-$NH_2$ | 37.3 (-11.3)[a] | 34 | 23.1 (14.2)[b] | 50.2 (-1.7)[a] | 39.2 (11)[c] |
| 29 | AP-PNA 3-4 | Ac-G**T$^{3a}$**A**G$^{3a}$**A**T$^{3a}$**C**A$^{3a}$**C T-Gly-$NH_2$ | 34.3 (-14.3)[a] | 29.5 | 32.4 (1.9)[b] | 44.6 (-7.3)[a] | 34.3 (10.3)[c] |
| 16 | AP-PNA 4-1 | Ac-GTAGAT**$^{4a}$**CACT-Gly-$NH_2$ | 46.1 (-2.5)[a] | 31.5 | 31.1 (15)[b] | 51.5 (-0.4)[a] | 42.2 (9.3)[c] |
| 30 | AP-PNA 4-2 | Ac-GTAGA**$^{4a}$** T**C$^{4a}$** ACT-Gly-$NH_2$ | 46.6 (2)[a] | 34.3 | 31.9 (14.7)[b] | 49.7 (2.2)[a] | 39.2 (10.5)[c] |
| 31 | AP-PNA 4-3 | Ac-**G$^{4a}$** TAGA**$^{4a}$** T**C$^{4a}$** ACT-Gly-$NH_2$ | 46 (2.6)[a] | 32.8 | 33.3 (12.7)[b] | 50.1 (1.8)[a] | 41.1 (9)[c] |
| 32 | AP-PNA 4-4 | Ac-G**T$^{4a}$**A**G$^{4a}$**A**T$^{4a}$**C**A$^{4a}$**C T-Gly-$NH_2$ | 45.5 (3.1)[a] | 33.5 | 34.6 (10.9)[b] | 48.7 (3.2)[a] | 37.6 (11.1)[c] |
| 18 | AP-PNA 6-1 | Ac-GTAGAT**$^{6a}$**CACT-Gly-$NH_2$ | 48.9 (+0.3)[a] | 33.9 | 33.8 (14.8)[b] | 51.8 (-0.1)[a] | 39.3 (12.5)[c] |
| 33 | AP-PNA 6-2 | Ac-GTAGA**$^{6a}$** T**C$^{6a}$** ACT-Gly-$NH_2$ | 47.2 (-1.4)[a] | 35.5 | 31.7 (15.5)[b] | 52.9 (+1)[a] | 43.9 (9)[c] |
| 34 | AP-PNA 6-3 | Ac-**G$^{6a}$** TAGA**$^{6a}$** T**C$^{6a}$** ACT-Gly-$NH_2$ | 46.6 (-2)[a] | 35 | 31 (15.6)[b] | 52.4 (+0.5)[a] | 43.4 (9)[c] |
| 35 | AP-PNA 6-4 | Ac-G**T$^{6a}$**A**G$^{6a}$**A**T$^{6a}$**C**A$^{6a}$**C T-Gly-$NH_2$ | 45.8 (-2.8)[a] | 33.5 | 34.3 (11.5)[b] | 53.6 (+1.7)[a] | 44 (9.6)[c] |
| 36 | AP-PNA 6-5 | Ac-G**T$^{6a}$**A**G$^{6a}$**A**T$^{6a}$**C**A$^{6a}$**C **T$^{6a}$** –Gly-$NH_2$ | 46.4 (-2.2)[a] | 36 | 33.5 (12.9)[b] | 54.3 (+2.4)[a] | 44 (10.3)[c] |

[a] ΔTm between aegPNA and modified PNA given in parentheses. [b]ΔTm for the deference in Tm between the PNA/DNA(ap) duplex to the corresponding PNA/DNA duplex with a base mismatch given in parentheses.. [c] ΔTm for the deference in Tm between the PNA/RNA(ap) duplex to the corresponding PNA/RNA duplex with a base mismatch given in parentheses..

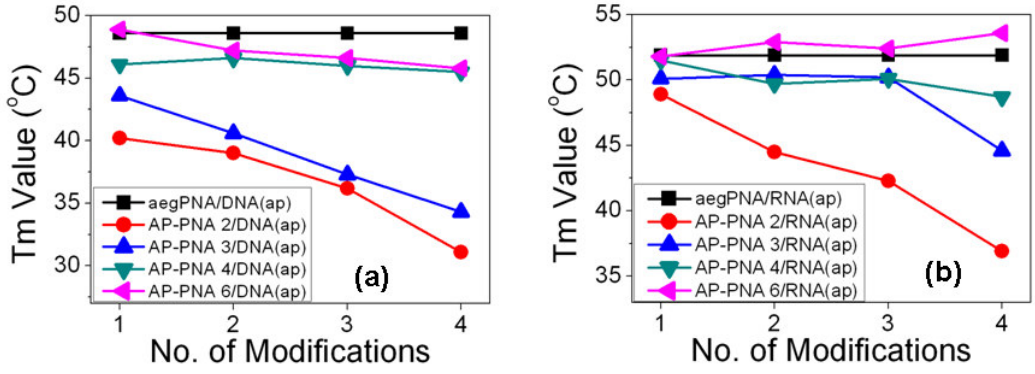



**Figure 3-6. Tm values of (a) AP-PNA/DNA(ap) duplexes and (b) AP-PNA/RNA (ap) duplexes.**

Figure 3-6 shows the Tm values of all the AP-PNA/DNA(ap) duplexes and AP-PNA/RNA(ap) duplexes containing 1, 2, 3 and 4 modifications with the 2, 3, 4 or 6 carbon amino-peptoid side chain. From the figure, one can see that the negative effect caused by a short 2 or 3-carbon amino-peptoid side chain on the stability of AP-PNA/DNA(ap) duplex increases with the increasing number of modifications in the sequence; however, the negative effect is not additive. For example, for the Tm values of AP-PNA/DNA(ap) duplex with the 2-carbon side chain length, introducing one modification causes a drop of 8.4 $^{o}$C drop (Table 3-5, entry 14), and introducing 2, 3 and 4 modifications in the sequences causes a drop of 9.6 $^{o}$C, 12.4 $^{o}$C and 17.5 $^{o}$C, respectively (Table 3-5, entry 24, 25 and 26). On the other hand, for the 4 and 6-carbon series, only a very small or no decrease in thermal stability of the AP-PNA/DNA(ap) duplex is seen when the number of peptoid modification is increased from 1 to 4 or even 5. Taking the 6-carbon side chain length as an example, there are no observable relationships between the Tm values of AP-PNA/DNA(ap) duplexes and the modifications in the sequences. The Tm values of AP-PNA/DNA(ap) duplexes increases by 0.3 $^{o}$C upon one modification (Table 3-3, entry 18), whereas having 2, 3, 4 and 5 such modifications causes a drop of 1.4 $^{o}$C, 2 $^{o}$C, 2.8 $^{o}$C and 2.2 $^{o}$C, respectively (Table 3-5, entry 33-36).

A similar trend is also obtained with that of AP-PNA/RNA(ap) duplexes, where a short peptoid side chain gives lower Tm value of AP-PNA/RNA(ap) duplexes as compared with longer peptoid side chain (Figure 3-6b). A small

difference here is that for the 6-carbon amino-peptoid side chain, increasing the number of modifications in the sequence causes a slight increase in Tm values.

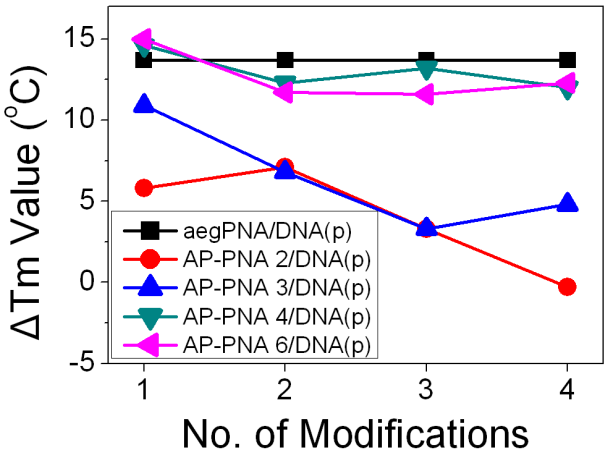



**Figure 3-7. ΔTm values between AP-PNA/DNA(ap) and AP-PNA/DNA(p) duplexes.**

Their antiparallel binding preferences were also examined. Figure 3-7 shows the difference of binding affinity in Tm values between the antiparallel DNA binding and parallel DNA binding. Except the 2 and 3-carbon AP-PNA series, all other AP-PNAs with a longer side chain preserve the antiparallel binding preference, as the difference in binding affinity of AP-PNAs toward antiparallel and parallel DNA is similar to that of aegPNA. And the antiparallel binding preference is not affected with the increasing number of modifications in the sequence.

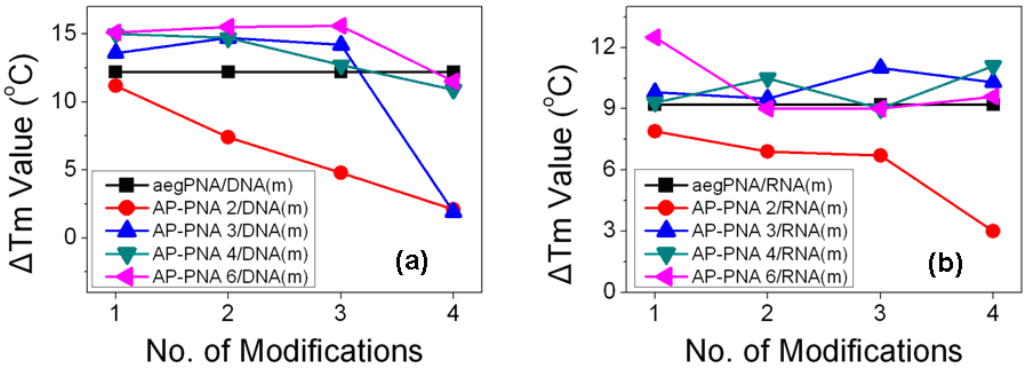



**Figure 3-8. ΔTm values between (a) AP-PNA/DNA(ap) and AP-PNA/DNA(m) duplexes and (b) AP-PNA/RNA(ap) and AP-PNA/RNA(m) duplexes.**

The discrimination abilities of AP-PNAs against mismatched DNA or RNA were also evaluated. Figure 3-8 shows the difference in Tm value between AP-PNA binding to the perfect matched antiparallel DNA or RNA binding and to a single base-pair mismatched DNA or RNA. Except for the 2-carbon peptoid side chain, introducing a longer peptoid side chain in PNAs increases the sequence selectivity. AP-PNAs with a 4 and 6-carbon peptoid side chain show even higher sequence discrimination abilities than aegPNAs. And the selectivity is not affected by increasing the number of modifications in the sequences.

### 3.2.4 Hybridization properties of GP-PNA containing multiple peptoid side chains of varying length in the decamer sequence

All the AP-PNAs shown in Table 3-5 were converted to the corresponding GP-PNAs. Their hybridization data in comparison with the unmodified aegPNA were presented in Table 3-6.

**Table 3-6. Effects of multiple guanidino-peptoid modifications on the thermal stability (Tm, ºC) of GP-PNA/DNA and GP-PNA/RNA duplexes.**

| Entry | Name of PNA | Sequence | DNA (ap) | DNA (p) | DNA (m) | RNA (ap) | RNA (m) |
|---|---|---|---|---|---|---|---|
| 1 | aegPNA | Ac-GTAGATCACT-Gly-$NH_2$ | 48.6 (0)[a] | 34.9 | 36.4 (12.2)[b] | 51.9(0)[a] | 42.7 (9.2)[c] |
| 19 | GP-PNA 2-1 | Ac-GTAGA$T^{2g}$CACT-Gly-$NH_2$ | 41.5 (-7.1)[a] | 31.5 | 27.3 (14.2)[b] | 50.3 (-1.6)[a] | 39.3 (11)[c] |
| 37 | GP-PNA 2-2 | Ac-GTAG$A^{2g}$ $TC^{2g}$ ACT-Gly-$NH_2$ | 39.3(-9.3)[a] | 30.6 | 25 (14.3)[b] | 46 (-5.9)[a] | 36.3 (9.7)[c] |
| 38 | GP-PNA 2-3 | Ac-$G^{2g}$ TG$A^{2g}$ $TC^{2g}$ ACT-Gly-$NH_2$ | 36.8(-11.8)[a] | 35.8 | 27.3 (9.5)[b] | 45.9 (-6)[a] | 37.9 (8)[c] |
| 39 | GP-PNA 2-4 | Ac-$GT^{2g}AG^{2g}AT^{2g}CA^{2g}C$ T-Gly-$NH_2$ | 33.3(-15.3)[a] | 32.7 | 28.7 (4.6)[b] | 38.9 (-13)[a] | 32.9 (6)[c] |
| 20 | GP-PNA 3-1 | Ac-GTAGA$T^{3g}$CACT-Gly-$NH_2$ | 45.4 (-3.2)[a] | 31.5 | 31.9 (13.5)[b] | 50.4 (-1.5)[a] | 40.3 (10.1)[c] |
| 40 | GP-PNA 3-2 | Ac-GTAG$A^{3g}TC^{3g}$ ACT-Gly-$NH_2$ | 45.8 (-2.8)[a] | 31 | 27 (12.8) | 52 (+0.1)[a] | 44.3 (7.7) |
| 41 | GP-PNA 3-3 | Ac-$G^{3g}$ TG$A^{3g}$ $TC^{3g}$ ACT-Gly-$NH_2$ | 41 (-7.6)[a] | 35.5 | 26.3 (14.7) | 52 (+0.1)[a] | 39.9 (12.1) |
| 42 | GP-PNA 3-4 | Ac-$GT^{3g}AG^{3g}AT^{3g}CA^{3g}C$ T-Gly-$NH_2$ | 39.1 (-9.5)[a] | 35 | 27 (12.1) | 46.9 (-5)[a] | 38 (8.9) |
| 21 | GP-PNA 4-1 | Ac-GTAGA$T^{4g}$CACT-Gly-$NH_2$ | 48 (-0.6)[a] | 33.3 | 33.9 (14.1)[b] | 52.5 (+0.6)[a] | 39.5 (13)[c] |
| 43 | GP-PNA 4-2 | Ac-GTAG$A^{4g}$ $TC^{4g}$ ACT-Gly-$NH_2$ | 47.9 (-0.7)[a] | 33 | 29.6 (18.3)[b] | 52.5 (+0.6)[a] | 43.3 (9.2)[c] |
| 44 | GP-PNA 4-3 | Ac-$G^{4g}$ TG$A^{4g}$ $TC^{4g}$ACT-Gly-$NH_2$ | 49 (+0.4)[a] | 33.9 | 35.8 (13.2)[b] | 50.8 (-1.1)[a] | 41.8 (9)[c] |
| 45 | GP-PNA 4-4 | Ac-$GT^{4g}AG^{4g}AT^{4g}CA^{4g}C$ T-Gly-$NH_2$ | 48.1 (-0.5)[a] | 33 | 33.1 (15)[b] | 50.6 (-1.3)[a] | 38.6 (12)[c] |
| 23 | GP-PNA 6-1 | Ac-GTAGA$T^{6g}$CACT-Gly-$NH_2$ | 48.5 (-0.1)[a] | 32.8 | 27.9 (20.6)[b] | 52.4 (+0.5)[a] | 42.7 (9.7)[c] |
| 46 | GP-PNA 6-2 | Ac-GTAG$A^{6g}$ $TC^{6g}$ ACT-Gly-$NH_2$ | 46.7 (-1.9)[a] | 32 | 21.6 (25.1)[b] | 52 (+0.1)[a] | 44.5 (7.5)[c] |
| 47 | GP-PNA 6-3 | Ac-$G^{6g}$TG$A^{6g}$ $TC^{6g}$ACT-Gly-$NH_2$ | 48.9 (+0.3)[a] | 32.2 | 33.4 (15.5)[b] | 52.7 (+0.8)[a] | 43.2 (9.5)[c] |
| 48 | GP-PNA 6-4 | Ac-$GT^{6g}AG^{6g}AT^{6g}CA^{6g}C$ T-Gly-$NH_2$ | 47 (-1.6)[a] | 33.6 | 27.9 (19.1)[b] | 55.4 (+3.5)[a] | 44.2 (11.2)[c] |
| 49 | GP-PNA 6-5 | Ac-$GT^{6g}AG^{6g}AT^{6g}CA^{6g}C$ $T^{6g}$ –Gly-$NH_2$ | 46.5 (-2.1)[a] | 31.9 | 29 (17.5)[b] | 54 (+2.1)[a] | 43.7 (10.3)[c] |

[a] ΔTm between aegPNA and modified PNA given in parentheses. [b]ΔTm between the perfect match PNA/DNA(ap) duplex and the corresponding PNA/DNA duplex with a single base mismatch given in parentheses. [c] ΔTm between the perfect match PNA/RNA(ap) duplex and the corresponding PNA/RNA duplex with a single base mismatch given in parentheses.

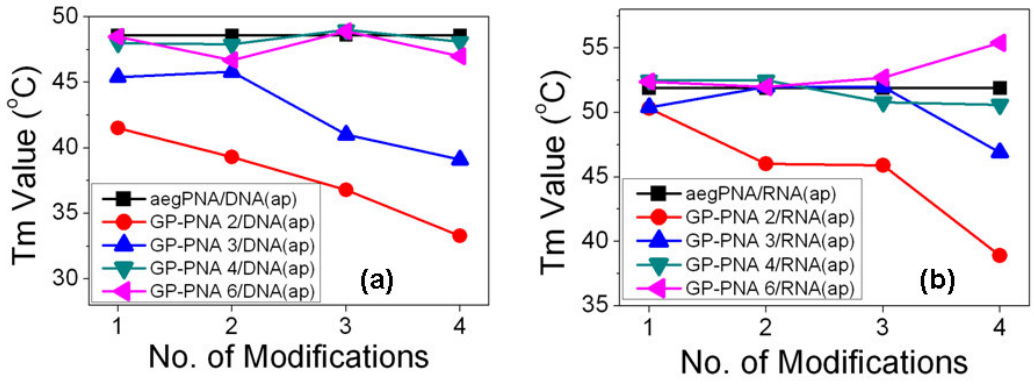



**Figure 3-9. Tm values of (a) GP-PNA/DNA(ap) duplexes and (b) GP-PNA/RNA(ap) duplexes.**

Figure 3-9 shows the Tm values of all the GP-PNA/DNA(ap) duplexes and GP-PNA/RNA(ap) duplexes containing 1, 2, 3 and 4 modifications with the 2, 3, 4 or 6-carbon guanidino-peptoid side chain. A similar length-dependent trend was

obtained. The hybridization affinity increases with an increasing length of the peptoid side chain. Those with short peptoid side chains (2 or 3-carbon) have a negative effect on the stability of the GP-PNA/DNA(ap) and GP-PNA/RNA(ap) duplexes. The detrimental effects become severer when increasing the number of modifications in the sequence. Those with longer peptoid side chains (4 or 6-carbon), on the other hand show no detrimental effects on the thermal stability of the GP-PNA/DNA(ap) and GP-PNA/RNA(ap) duplexes, even with an increasing number of modifications in the sequences. For example GP-PNA 6-5, which has 5 guanidinohexyl peptoid side chains, behaves similarly with aegPNA in binding to DNA or RNA – they exhibit similar binding affinities, antiparallel binding preferences and sequence selectivity.

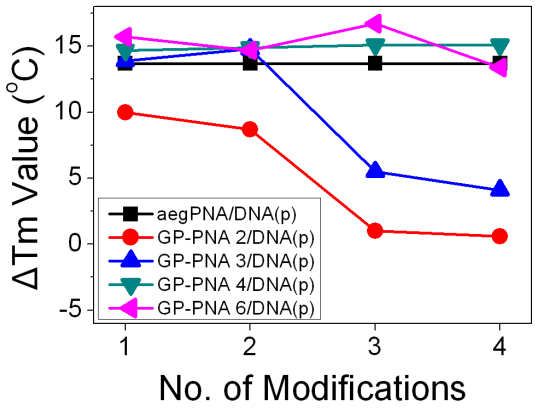



**Figure 3-10. ΔTm values between GP-PNA/DNA(ap) and GP-PNA/DNA(p) duplexes.**

The antiparallel binding preferences were also examined. Figure 3-10 shows the difference of binding affinity in Tm value between the antiparallel DNA binding and parallel DNA binding. Except for some with the 2 or 3-carbon side chains, most of GP-PNAs preserve the antiparallel binding preference, as the

difference in binding affinity of GP-PNAs toward antiparallel and parallel DNA is similar with that of aegPNA. And the antiparallel binding preference is not affected when increasing the number of modifications in the sequence.

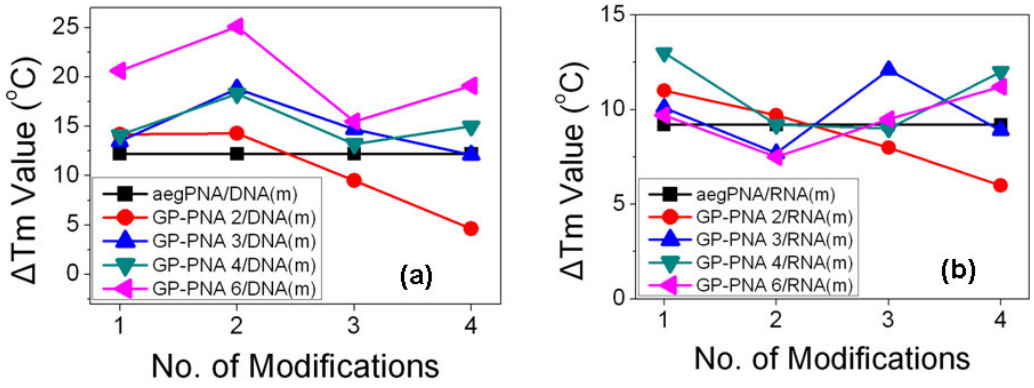



**Figure 3-11. ΔTm values between (a) GP-PNA/DNA(ap) and GP-PNA/DNA(m) duplexes and (b) GP-PNA/RNA(ap) and GP-PNA/RNA(m) duplexes.**

The discrimination abilities of GP-PNAs to the mismatched sequences were also evaluated. Figure 3-11 shows the difference in Tm value between the antiparallel DNA or RNA binding and mismatched DNA or RNA binding. Interestingly, introducing even the 2-carbon guanidino-peptoid side chain in PNAs does not compromise much of the binding specificity. GP-PNAs with 3, 4 or 6-carbon guanidino-peptoid side chain show even higher sequence discrimination ability than aegPNAs. And this ability is not affected when increasing the number of modifications in the sequences.

## 3.2.11 Discussion on duplex formation and stability

Until now, a systematic work has been done to evaluate the effects of side chain length, the headgroup and cumulative modifications on the duplex thermal stability. Limited by the high cost of the starting material 1, 5-diaminopentane, the cumulative effect of the 5-carbon side chain in AP-PNAs and GP-PNAs is not examined in this study. The hybridization properties of the others were performed and discussed in detail.

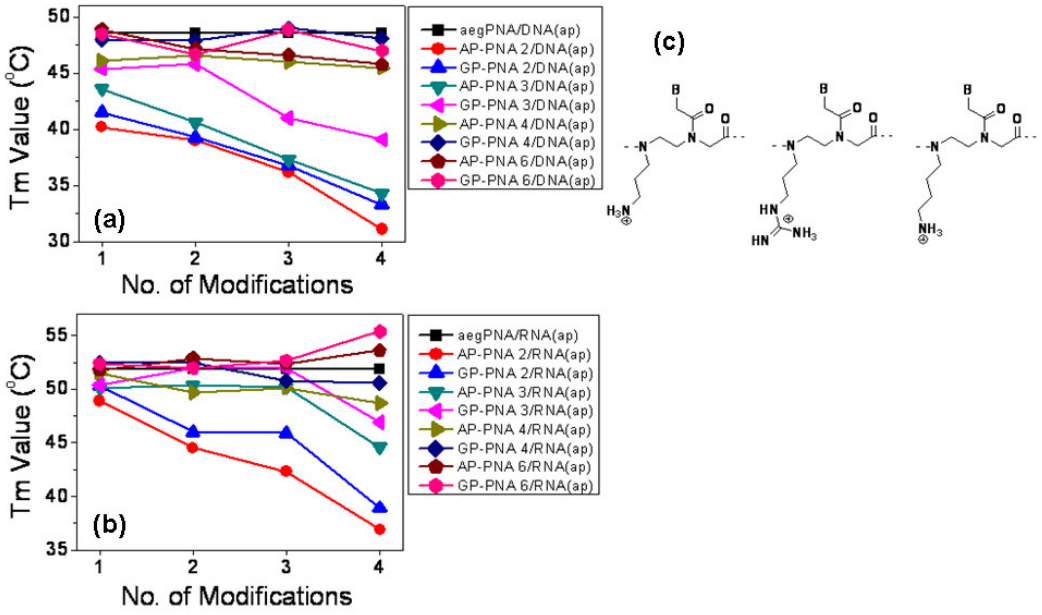



**Figure 3-12. Tm Values of all the AP-PNAs and GP-PNAs binding toward (a) DNA(ap) and (b) RNA(ap); (c) Schematic illustration of the distances of the cationic centre to the PNA backbones.**

As shown in Figure 3-12a and Figure 3-12b, an interesting phenomenon that the 'crossed-increasing' of the Tm values between AP-PNA and GP-PNA toward antiparallel DNA and RNA binding is observed. Generally, for a given number of modifications in the sequence, the Tm values of AP-PNA and GP-PNA toward

DNA(ap) and RNA(ap) binding reveal such a tendency: aegPNA ≥ GP-PNA 4 > AP-PNA 4 > GP-PNA 3 > AP-PNA 3 > GP-PNA 2 > AP-PNA 2. As discussed previously, this may due to the donation of bihydrogen bond from guanidino group as compared one hydrogen bond from amino group. However, our data have shown that the electrostatic attraction between PNA and DNA does not impact on the binding affinity toward DNA(ap) or RNA(ap). Another possible explanation is the different distance between the cation centre and the PNA backbone. As shown in Figure 3-12c, the distance between the cationic centre at the guanidino headgroup and the backbone is around one carbon longer than that of the amino headgroup. This can be consolidated by the following two findings. One is that for a 6-carbon side chain, the affinities of GP-PNA are not higher than that of AP-PNA toward DNA(ap) or RNA(ap) binding. The other one is that the hybridization behaviours of AP-PNA with 6-carbon side chain are similar with that of GP-PNA with 4-carbon side chain.

Also, as shown in Figure 3-12 (a and b), the cumulative effects are length-dependent. No matter an amino or guanidino headgroup it bears, the negative impact of a 2 or 3-carbon side chain on duplex stability becomes severer when increasing the number of modifications in the sequences; on the other hand, those with 4 or 6-carbon side chain do not show a cumulative negative effect with the increasing number of modifications in the sequences. This is quite interesting, since increasing the number of modifications in the sequence, may increase the non-specific electrostatic attractions between the cationic PNA and anionic DNA.

However, the data obtained so far and in the following section show that these non-specific electrostatic attractions do not contribute much to the thermal stability of the duplexes nor affect the sequence binding selectivity.

It is also worth noticing that, despite the negative impact of a short peptoid side chain on hybridizing affinity, almost all the AP-PNAs and GP-PNAs still bind to complementary anti-parallel DNA or RNA with a higher affinity than does the corresponding DNA oligonucleotide (Tm of a comparable DNA/DNA or DNA/RNA duplex is about 33-34 $^{o}$C [2]). Considering the extremely high binding affinity of an unmodified aegPNA of 15 to 20 residues – the optimal length required of an antisense agent for unique targeting,[2] the moderate hybridizing affinity of some of the peptoid PNAs, such as the 3-C series, may have some practical value for antisense or antigene applications. It has been determined that a 15-mer PNA binds to the perfect-matched complementary RNA at a Tm of over 70 $^{o}$C and to the RNAs with a single-base mismatch at a Tm of 54-63 $^{o}$C,[2] which raises the concern of non-specific binding to other unintended targets at physiological temperature. Our data suggest that, by choosing an appropriate side-chain length and/or degree of peptoid modification, it is possible to design PNA-based agents with tailored hybridization strength. One can even envisage using several mixed peptoid side chains of different spacer lengths to fine-tune the hybridization behaviours of PNA. Our data also illustrate again the excellent versatility of PNA's aeg backbone which allows it to be easily modified to suit a variety of applications.

## 3.2.12 The hybridization properties of AP-PNA derivatives

To further examine the effect of side-chain length and charge state on hybridization properties, the amino headgroup of four mono-$N^{\gamma}$-aminoalkylated PNAs (AP-PNA **14**, **15**, **16** and **18** from Table 3-3) were modified by an aminopropionyl or β-alanyl group, which extended the side chain by three atoms, to give PNA **50**, **51**, **52** and **53**, respectively. Also the amino group of the mono-$N^{\gamma}$-aminohexyl PNA (AP-PNA 6-1) was acetylated with acetic anhydride or biotinylated to give PNA **54** and **55**, respectively. Their hybridization properties with DNA and RNA were examined and the data were shown in Table 3-7 and Figure 3-13.

**Table 3-7. Effects of modifying the peptoid side-chain amine on hybridization property.**

| Entry | Name of PNA | Sequence | DNA (ap) | DNA (p) | DNA (m) | RNA (ap) | RNA (m) |
|---|---|---|---|---|---|---|---|
| 1 | aegPNA | Ac-GTAGATCACT-Gly-$NH_2$ | 48.6 | 34.9 | 36.4 (12.2) [b] | 51.9 | 42.7 (9.2) [c] |
| 50 | AP-PNA 2-1(βAla) | Ac-GTAGA**T**$^{2a(\beta Ala)}$CACT-Gly-$NH_2$ | 46 (-2.6)[a] | 36.2 | 32.5 (13.5) [b] | 52.6 (+0.7)[a] | 42.6 (10) [c] |
| 51 | AP-PNA 3-1(βAla) | Ac-GTAGA**T**$^{3a(\beta Ala)}$CACT-Gly-$NH_2$ | 47.7 (-0.9)[a] | 36.2 | 33.8 (13.9) [b] | 52.4 (+0.5)[a] | 43.2 (9.2) [c] |
| 52 | AP-PNA 4-1(βAla) | Ac-GTAGA**T**$^{4a(\beta Ala)}$CACT-Gly-$NH_2$ | 48.7 (+0.1)[a] | 36.6 | 32.0 (16.7) [b] | 53.4 (+1.5)[a] | 41.2 (12.2) [c] |
| 53 | AP-PNA 6-1(βAla) | Ac-GTAGA**T**$^{6a(\beta Ala)}$CACT-Gly-$NH_2$ | 48.8 (+0.2)[a] | 35.2 | 31.8 (17) [b] | 53.9 (+2)[a] | 44 (9.9) [c] |
| 54 | AP-PNA 6-1(Ac) | Ac-GTAGA**T**$^{6a(Ac)}$CACT-Gly-$NH_2$ | 48.9 (+0.3)[a] | 40.5 | 40.8 (8.1) [b] | 53.9 (+2)[a] | 44.5 (9.4) [c] |
| 55 | AP-PNA 6-1(biotin) | Ac- GTAGA**T**$^{6a(Biotin)}$CACT-Gly-$NH_2$ | 48.6 (0)[a] | 42.9 | 40 (8.6) [b] | 53.9 (+2)[a] | 44.4 (9.5) [c] |

[a] ΔTm between aegPNA and modified PNA given in parentheses. [b]ΔTm between the perfect match PNA/DNA(ap) duplex and the corresponding PNA/DNA duplex with a single base mismatch. [c] ΔTm between the perfect match PNA/RNA(ap) duplex and the corresponding PNA/RNA duplex with a single base mismatch.

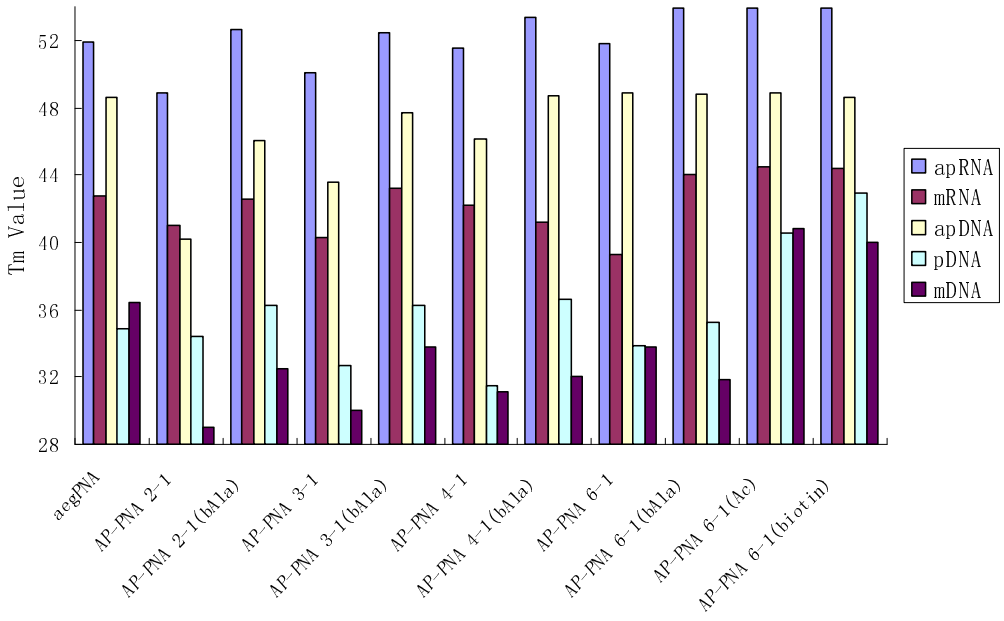



**Figure 3-13. Tm values of modified AP-PNA/DNA and AP-PNA/RNA duplexes.**

As shown in Figure 3-13 and Table 3-7, aminopropionylation on the amino-peptoid side chain results in an increases of the thermal stability of most AP-PNA/DNA(ap) and AP-PNA/RNA(ap) duplexes. The effect of this modification is notably significant for AP-PNAs with a short side chain (such as AP-PNA 2-1, AP-PNA 3-1 and AP-PNA 4-1), but not for AP-PNA 6-1 which already has a long side-chain spacer. For example, AP-PNA 2-1(βAla), whose six-atom side chain spacer is extended from the two atoms in AP-PNA 2-1, gains 5.8 ºC in Tm for binding with DNA(ap) and 3.7 ºC in Tm with RNA(ap) as compared to its parent AP-PNA 2-1. AP-PNA 3-1(βAla), whose side-chain spacer is extended from three atoms in AP-PNA 3-1 to seven atoms, gains 4.1 ºC in Tm

for binding with DNA(ap) and 2.3 ºC in Tm with RNA(ap) as compared to its parent AP-PNA 3-1. AP-PNA 4-1(βAla), whose side-chain spacer is extended from four atoms in AP-PNA 4-1 to eight atoms, gains 2.6 ºC for binding with DNA(ap) and 1.9 ºC in Tm with RNA(ap) as compared to its parent AP-PNA 4-1. AP-PNA 6-1(βAla) which has its side-chain spacer extended from six atoms in AP-PNA 6-1 to 10 atoms, on the other hand, exhibits similar binding affinity in Tm with its parent AP-PNA 6-1. The fact that the Tm values of AP-PNA 2-1(βAla), AP-PNA 3-1(βAla) and AP-PNA 4-1(βAla) are still not as good as AP-PNA 6-1 despite having spacers with the same number of atoms or longer, suggest that in addition to the spacer length, factors such as intrachain hydrogen-bonding ability and/or conformational rigidity of the peptoid side chain may also affect the binding affinity of AP-PNA toward complementary oligonucleotides.

The antiparallel binding preference is retained since the binding affinity toward parallel DNA was reduced largely as compared with antiparallel DNA (Table 3-7). Compared with their parent PNAs (Table 3-3), aminopropionylation on AP-PNAs leads to higher sequence specificity against base pair mismatches as compared to the parent AP-PNAs.

To examine the effect of side-chain charge state on the thermal stability of duplexes, the amino group in AP-PNA 6-1 was acetylated to get the resultant AP-PNA 6-1(Ac) and acylated with biotin to get the AP-PNA 6-1(biotin). Their hybridization properties were also studied (Table 3-7, entry 54 and 55). AP-PNA

6-1(Ac) and AP-PNA 6-1(biotin) show no much change in Tm for binding with DNA(ap) and RNA(ap) as compared to its parent AP-PNA 6-1. However, their antiparallel binding preferences and sequence selectivity are significantly reduced. The results further indicates that the positive charges on the amino headgroup do not contribute much to the hybridizing affinity of AP-PNAs, but do to maintaining good antiparallel binding preference and sequence discrimination abilities.

The above data illustrate again the beneficial effects of a long side chain in AP-PNA, although the effects seem capped at the 5 or 6-carbon spacer length. This observed spacer length-dependent effect on the hybridizing affinity of AP-PNA is intriguing. From the data obtained in this study and those of Nielsen et al[1] showing the negative impact of $N^{\gamma}$-methylation, one might be enticed to postulate a favourable role of the positive charge of the peptoid amine for charge-charge interactions with the phosphodiester backbone which would compensate for the negative effect of an $N^{\gamma}$-alkyl group, and a longer 5- or 6-carbon spacer would provide the optimal distance for such interactions. However, acetylation of the positive charge-bearing peptoid amine of AP-PNA 6-1 causes no reduction in the hybridization affinity of the resultant AP-PNA 6-1(Ac). Also, when the side-chain amine of AP-PNA 6-1 was acylated with biotin, a useful biochemical probe, the resultant AP-PNA 6-1(biotin) exhibits the same hybridizing affinity as does AP-PNA 6-1. Therefore, the positive charge on the amino headgroup of the peptoid side chain does not seem to play a role in contributing to the hybridizing affinity of AP-PNA. This observation is also similar to what was found in a

previous study on PNAs with a γ-C side chain.[3] The present data do not allow us to pinpoint the cause of this length-dependent effect. One possible reason is that a short amino-peptoid side chain might induce a higher content of E configuration about the tertiary amide due to hydrogen bonding potential of the peptoid ammonium with the amide carbonyl, whereas PNA's backbone amide is known to adopt a Z configuration for oligonucleotide binding.[4] It is also possible that a longer side chain may help stabilize better the helical structure of the AP-PNA/DNA duplex, as previous studies on peptoids have shown that increasing the size of the peptoid side chain imposes restrictions to the conformational flexibility of the oligoglycine backbone.[5-7] However, structural investigations by NMR or X-ray crystallography are needed to find out whether these or any other factors are involved. The data on AP-PNA 6-1(Ac) and particularly AP-PNA 6-1(biotin) also confirm that one can indeed use the peptoid side-chain amine in an AP-PNA as the attachment site to introduce a relatively large molecular moiety for preparing PNA conjugates with potential diagnostic and therapeutic interest.

## 3.3 CD spectroscopy

We have demonstrated that GP-PNAs and AP-PNAs with a 6-carbon side-chain spacer show similar hybridization properties as aegPNA. Therefore they represent an interesting candidate as antigene or antisense agents. In this and the following part, we will further evaluate the circular dichroic properties and cellular uptake properties of GP-PNA or AP-PNA oligomers containing the 6-carbon

peptoid side chains.

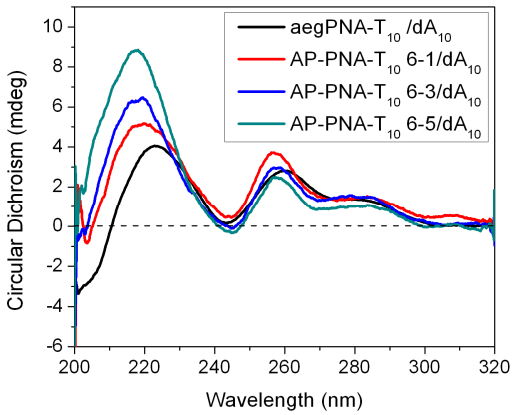



**Figure 3-14. CD spectra of $(aegPNA\text{-}T_{10})_2/dA_{10}$ and $(AP\text{-}PNA\text{-}T_{10})_2/dA_{10}$ triplexes.**

Figure 3-14 presents a comparison between the CD spectrum of $(aegPNA\text{-}T_{10})_2/dA_{10}$ and those of all three $(AP\text{-}PNA\text{-}T_{10})_2/dA_{10}$ triplexes with the 6-carbon side chains. The CD spectra of the three $(AP\text{-}PNA\text{-}T_{10})_2/dA_{10}$ triplexes with different number of modifications in the sequence are quite similar with each other as well as that of $(aegPNA\text{-}T_{10})_2/dA_{10}$.

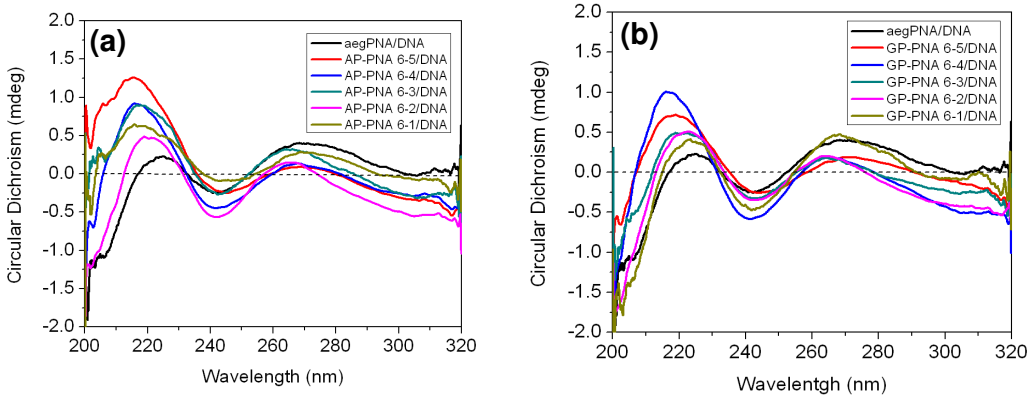



**Figure 3-15. CD spectra of (a) the aegPNA/DNA(ap) and AP-PNA/DNA(ap) duplexes and the aegPNA/DNA(ap) and GP-PNA/DNA(ap) duplexes.**

We also investigated circular dichroic properties of aegPNA/DNA(ap), AP-PNA/DNA(ap) and GP-PNA/DNA(ap) duplexes to address whether the effects of modifications include structural changes of the duplexes. Figure 3-15a and Figure 3-15b represent the comparisons between the CD spectra of aegPNA/DNA(ap) duplexes and AP-PNA/DNA(ap) duplexes as well as GP-PNA/DNA(ap) duplexes. Interestingly, like in the structure of triplexes, no major structural change occurs in the AP-PNA/DNA(ap) and GP-PNA/DNA(ap) duplexes, even upon introduction of 5 modifications in the PNA strands. However, the different intensities of CD signals also point to some minor differences in the conformations adopted by these duplexes.

## 3.4 Cellular uptake properties

Many requirements needed to be fulfilled for the practical application of an antisense or antigene agent.[8] The first is the high binding affinity and sequence specificity towards the target sequences. Secondly, it must be stable under physiological conditions and have a sufficiently long half-life under *in vivo* conditions for it to be able to display its desired action in the cell. It must therefore be resistant to enzymes (e.g. nucleases) that degrade nucleic acids. The most important is that it must be able to pass through the cell membranes to reach its site of action. Uniform distribution will normally be desirable; i.e., it should not accumulate either in particular organs such as the liver or in particular cell compartments such as the lysosomes.

The extremely high stability *in vivo* as well as the high binding affinity and the sequence selectivity toward the target sequences are the main advantages of PNA. However, the poor cellular uptake property has hampered its use as potential antisense and antigene agents. Therefore, PNAs that are cell-permeable are highly desirable as antisense and antigene candidates.

The dodecamer of GP-PNA (Fl-TA$^{6g}$GT$^{6g}$AG$^{6g}$AT$^{6g}$CA$^{6g}$CT$^{6g}$-Gly-$NH_2$), and AP-PNAs (Fl-TA$^{6a}$GT$^{6a}$AG$^{6a}$AT$^{6a}$CA$^{6a}$CT$^{6a}$-Gly-$NH_2$) were synthesized and labeled with fluorescein at the N-termini. The backbone of the AP-PNA oligomer was alternatively spaced between an $N^{\gamma}$-aminohexylpeptoid PNA monomer and aegPNA monomer. The GP-PNA oligomers are prepared from the corresponding AP-PNAs by converting the amino groups to the guanidine groups. The aegPNA (Fl-TAGTAGATCACT-Gly-$NH_2$) and TAT domain (Fl-YGRKKRRQRRR-$NH_2$) were used as negative and positive controls, respectively. These oligomers were incubated with Hela cells for 24 h at 37 $^{o}$C at a concentration of 0.1μM. Following the incubations, the cells were thoroughly washed with PBS (1x) buffer. In order to avoid the artifactual intracellular distribution caused by cell-fixation,[9] the experiments were performed using unfixed cells. The cells were grown on the glass-bottom cultured plates and viewed directly under the fluorescent microscope.

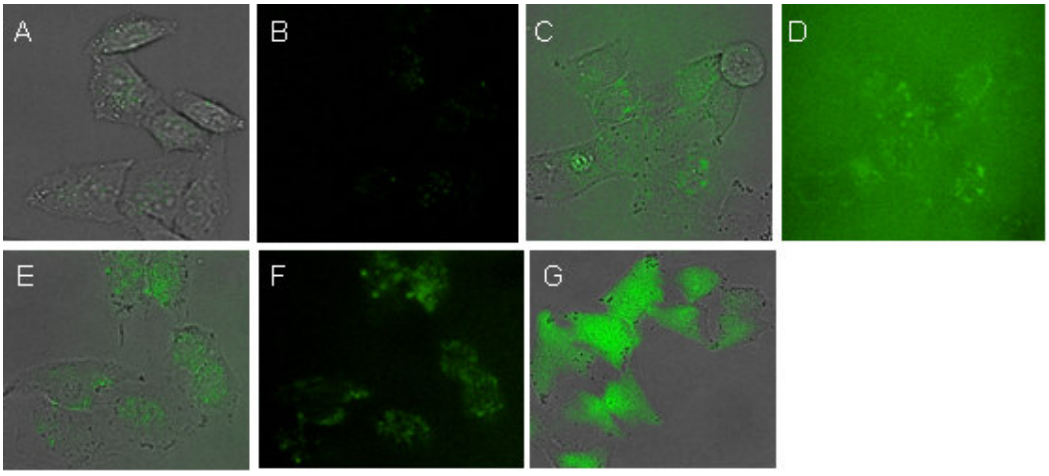



**Figure 3-16. Fluorescent confocal microscopy images of the uptake of fluorescein(FAM)-labelled aegPNA(A, B), AP-PNA(C, D), GP-PNA(E, F) and TAT domain (G) incubated for 24 h with Hela cells at a concentration of 0.1 μM. A, C, E and G are the overlay of DIC and fluorescent images; B, D and F are the fluorescent images.**

As shown in Figure 3-16, the efficiency of GP-PNA taken up by cells is better than that of AP-PNA, which is better than that of aegPNA. For example, no obvious uptake was observed with aegPNA (Figure 3-16A and B). The fluorescent images in Hela cells are too faint to be detected. The AP-PNA oligomer (Figure 3-16C and D) showed a slightly better cell uptake property than that of aegPNA. The fluorescent signal was faint and most of the cells had no fluorescent absorbance, which means that the AP-PNA did not traverse all of the cells. For the GP-PNA oligomer (Figure 3-16E and F), the majority of the cells showed clear and strong fluorescent signals, suggesting that the GP-PNA oligomer was taken up by the cells. However, the cell-uptake property of GP-PNA is not as efficient as the TAT peptide. The TAT traversed the whole cells and mostly in the cell nucleus (Figure 3-16G).

In conclusion, we have demonstrated that the GP-PNA shows better cellular uptake properties than that of aegPNA, but still not as good as the TAT peptide under identical conditions. One possible explanation is that the cell permeability is

affected by the number of guanidino headgroups, the molecular weight of the oligomers as well as the type of cell-line. Therefore, more structure-function relationship studies will be performed in the future on fluorescein-tagged probes of GP-PNA and GP-PNA/PNA chimeras and other control transporters to determine the minimum and optimal number of guanidino groups required for cell permeability and the effect of the size of GP-PNA on cellular uptake. The kinetics and mechanism of cellular entry will also be studied. Intracellular distribution and localization of the probe will be analyzed concurrently.

## 3.5 Experiments

### 3.5.1 Melting curves and melting temperature (Tm)

DNA/RNA oligomers were purchased from SIGMA PROLIGO. Thermal denaturation studies (Tm) were run on a Cary 300 Bio UV-Visible spectrophotometer. The oligomers were hybridized in a buffer solution containing 100 mM NaCl, 10 mM sodium phosphate and 0.1 mM EDTA, pH 7.0. The concentrations of PNAs were quantified by measuring the absorbance ($A_{260}$) of the PNA solution at 260 nm. The value for the molar extinction coefficients (ε260) of the single base are: $\varepsilon_{260}$ (A) = 13.7 mL/(μmole × cm), $\varepsilon_{260}$ (C)= 6.6 mL/(μmole × cm), $\varepsilon_{260}$ (G) = 11.7 mL/(μmole × cm), $\varepsilon_{260}$ (T) = 8.6 mL/(μmole × cm). Molar extinction coefficient of PNA is represented as the sum of those of individual bases that comprise the oligomer. $\varepsilon_{260} = \Sigma\ \varepsilon_i \times n_i$ ($\varepsilon_i$ = molar extinction coefficient of base, $n_i$ = number of base). For duplex formation, the concentrations of DNA, aegPNA,

AP-PNA and GP-PNA oligomers were 2 μM each; for triplex formation, the concentrations of DNA was 2 μM and concentration of aegPNA, AP-PNA or GP-PNA oligomers were 4 μM each. The samples were first heated to 90 $^{o}$C for 5 min, followed by gradually cooling to room temperature. UV-absorption was monitored at 260 nm from 85 to 15 $^{o}$C and 15 to 85 $^{o}$C at the rate of 0.5 $^{o}$C per minute. All values are accurate to ± 1 ºC. Experiments are repeated at least three times and all values are the average of three valid measurements.

## 3.5.2 CD spectroscopy

CD spectra were recorded on an Applied Photophysics Chirascan spectrophotometer. Each spectrum is the average of five scans, recorded at 25 $^{o}$C at a 1 mm optical path length. The samples were kept in a buffer solution containing 100 mM NaCl, 10 mM sodium phosphate, and 0.1 mM EDTA, pH 7.0. The samples were heated at 90 $^{o}$C for 5 min, followed by gradually cooling to room temperature prior to recording CD spectrum. Except the sequence of aegPNA-$T_{10}$, all the terminal amines in the other sequences are capped with acetic anhydride.

## 3.5.3 Cellular uptake properties

Culture and transfection of Hela cells with fluorescein-labeled aegPNA, AP-PNA, GP-PNA oligomers and Tat peptide were carried out according to the published protocols.[10, 11] Hela cells were grown in 75 $cm^2$ culture flasks at 37 $^{o}$C in an atmosphere of 5% $CO_2$ in Hyclone DMEM/High Glucose medium without

antibiotics and were plated at 100,000 cells per well on 6-well Lab-Tek chambered coverglass slides in DMEM medium and cultured two days before transfection. Oligomer solutions were prepared by diluting a 200 μM stock solution at 1×PBS buffer to 0.1 μM with warmed (37 $^{o}$C) medium. The cell medium was discarded and cells were washed with 1×PBS buffer followed by incubation with the above prepared oligomers at a concentration of 0.1 μM for 24 h at 37 $^{o}$C in an atmosphere of 5% $CO_2$. The medium was removed and cells were washed with 1×PBS buffer (2 mL × 3). The cells were imaged with Olympus IX51 fitted with Cool SNAP$^{HQ}$ camera and analyzed using Metamorph software (Molecular Devices).

## 3.6 References


[1] G. Haaima, H. Rasmussen, G. Schmidt, D. K. Jensen, J. S. Kastrup, P. W. Stafshede, B. Norden, O. Buchardt, P. E. Nielsen, *New Journal of Chemistry* **1999**, *23*, 833.

[2] S. Sforza, G. Galaverna, A. Dossena, R. Corradini, R. Marchelli, *Chirality* **2002**, *14*, 591.

[3] E. A. Englund, D. H. Appella, *Angewandte Chemie International Edition in English* **2007**, *46*, 1414.

[4] M. Eriksson, P. E. Nielsen, *Nature Structural Biology* **1996**, *3*, 410.

[5] P. Armand, K. Kirshenbaum, A. Falicov, R. L. Dunbrack, K. A. Dill, R. N. Zuckermann, F. E. Cohen, *Folding & Design* **1997**, *2*, 369.

[6] P. Armand, K. Kirshenbaum, R. A. Goldsmith, S. Farr-Jones, A. E. Barron, K. T. V. Truong, K. A. Dill, D. F. Mierke, F. E. Cohen, R. N. Zuckermann, E. K. Bradley, *Proceedings of the National Academy of Sciences of the United States of America* **1998**, *95*, 4309.

[7] K. Kirshenbaum, A. E. Barron, R. A. Goldsmith, P. Armand, E. K. Bradley, K. T. V. Truong, K. A. Dill, F. E. Cohen, R. N. Zuckermann, *Proceedings of the National Academy of Sciences of the United States of America* **1998**, *95*, 4303.

[8] E. Uhlmann, A. Peyman, *Chemical Reviews* **1990**, *90*, 543.

[9] J. P. Richard, K. Melikov, E. Vives, C. Ramos, B. Verbeure, M. J. Gait, L. V. Chernomordik, B. Lebleu, *Journal of Biological Chemistry* **2003**, *278*, 585.

[10] B. A. Janowski, J. Hu, D. R. Corey, *Nature Protocols* **2006**, *1*, 436.

[11] T. Shiraishi, P. E. Nielsen, *Nature Protocols* **2006**, *1*, 633.

# Chapter 4 Conclusions

The overall goal of this dissertation was to (A) design and synthesize peptoid-like PNA analogs with positively charged amino or guanidino groups attached to the γ-N PNA backbone with various side chain lengths; (B) study the hybridization properties of synthesized PNA analogs and (C) study their uptake activities in cultured cells.

## 4.1 Design and synthesis of AP-PNA and GP-PNA oligomers and AP-PNA derivatives

We have synthesized a series of AP-PNA monomers with various side chain lengths containing four natural nucleobases. The exocyclic nucleobases amino group was protected either by acyl or Bhoc protecting group. Boc was used to protect the peptoid side chain amino group and Fmoc for the backbone secondary amino group where the PNA oligomer would elongate. The AP-PNA oligomers were synthesized on Rink amide PEGA resin using standard solid-phase Fmoc peptide synthesis protocols. While coupling onto an aegPNA residue was easily affected using PyBOP as the coupling reagent, coupling onto an AP-PNA residue was more difficult owing to the steric hindrance of its backbone secondary amine and was achieved through the action of HATU for ~ 6 h. Meanwhile, a guanine derivative compound was added to the coupling reaction mixture to suppress any modification on guanine residues during PNA oligomer assembly using the PyBOP

as coupling reagent. The AP-PNA oligomers were readily soluble in water, which facilitated their purification and hybridization studies.

GP-PNA oligomers were obtained by converting the corresponding AP-PNA oligomers in solution using N,N'-bis-Boc-S-methylisothiourea in the presence of DIEA. Albeit convenient, the yields are very low especially when the amino groups in sequence are more than three. In the future, the guanidino headgroup can be introduced at the monomer stages using sulfonamide as a protecting group, or through optimizing the guanylation reaction promoted by $HgCl_2$.

The amino headgroup in AP-PNAs can be modified by an acylation reaction. Several functional moieties, such as β-Ala and biotin, have been introduced onto the amino headgroup. Hybridization studies have shown that the modifications would not affect their binding affinity, suggesting that the peptoid side chain can bear a variety of interesting moieties to modulate the hybridization and pharmacological properties of the PNA analogues.

## 4.2 Hybridization properties of synthesized PNA analogs

The effects of the side chain length, headgroup and the cumulative modifications in the sequences have been examined systematically. The hybridization properties of AP-PNA oligomers, GP-PNA oligomers as well as AP-PNA derivatives were studied and compared with that of aegPNA.

One intriguing finding is that the length of side chain plays an important role in the thermal stability of the PNA/DNA duplexes. For AP-PNA and GP-PNA oligomers, their hybridization affinities toward antiparallel DNA and RNA increase with increasing length of the peptoid side chain with a cap reached at the 5 or 6-carbon length. For instance, AP-PNAs and GP-PNAs with peptoid short side chains (2, 3 or 4 carbons) show an inferior duplex stability, and the detrimental effects of the short side chain become severer when increasing the number of modifications in sequences. However, those with longer peptoid side chain (5 or 6 carbons) exhibit a similar ability with aegPNA in binding affinities, antiparallel binding preferences and sequence selectivity. β-Alanylation on the peptoid side chain amine significantly increase the DNA- and RNA-binding affinity of those AP-PNAs with a short peptoid side chain (2, 3 and 4 carbons, notably), although these β-alanylated derivatives are still not as good as AP-PNA 6-1. The AP-PNA 6-1(βAla) exhibits similar binding affinity in Tm with its parent AP-PNA 6-1. Therefore, in addition to the peptoid side chain length, factors such as intrachain hydrogen-bonding ability and/or conformational rigidity of the peptoid side chain may also affect the binding affinity of AP-PNA toward complementary oligonucleotides. Future structural investigations by NMR or X-ray crystallography are needed to find out whether these or any other factors are involved.

The cumulative effects are also length-dependent. The negative impact caused by AP-PNA oligomers and GP-PNA oligomers with short peptoid side chains (2, 3

or 4 carbon) become severer when increasing the number of modifications in the sequences. On the other hand, the DNA- and RNA-binding affinities of AP-PNA and GP-PNA with 6-C side chain length are not affected with the number of modifications in the sequences. Therefore the non-specific electrostatic attractions between the negatively charged DNA or RNA and the positively charged AP-PNA or GP-PNA do not contribute much to the thermal stability of duplexes.

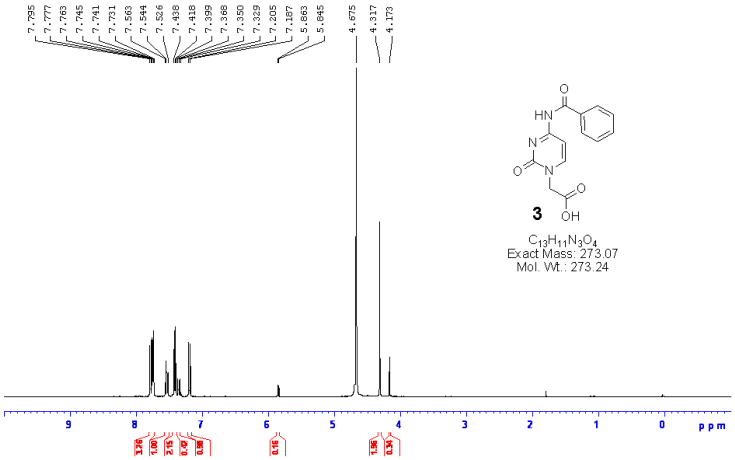
The 'crossed-increasing' tendency: aegPNA ≥ GP-PNA 4-carbon > AP-PNA 4-carbon > GP-PNA 3-carbon > AP-PNA 3-carbon > GP-PNA 2-carbon > AP-PNA 2-carbon, as well as the hybridization properties of AP-PNA 6-1(Ac) and AP-PNA 6-1(Biotin), indicate that it is the side chain length other than the nature of the headgroup that affects the hybridization affinity. However, the well suppressed hysteresis indicates that the presence of positive charges (the nature of headgroups) plays an important role in facilitate the triplex formation.

## 4.3 Cellular uptake activity in cultured cells

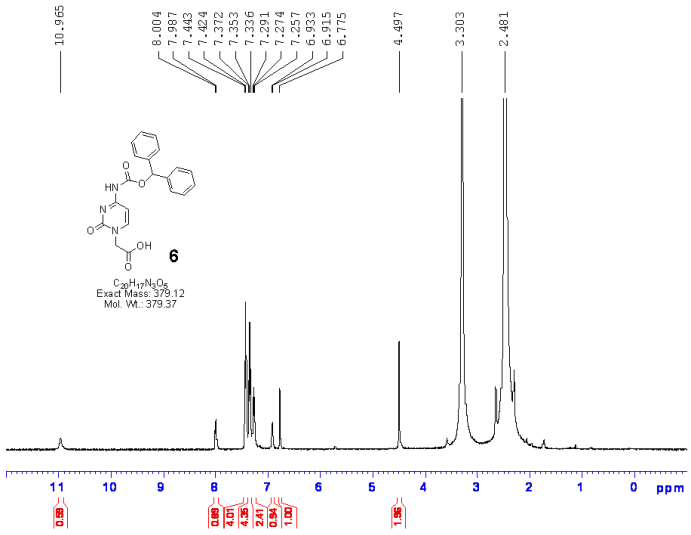
Fluorescein-tagged dodecamer AP-PNA and GP-PNA oligomers containing alternated 6-carbon peptoid side chains were used to study the cellular uptake activity in cultured Hela cells. Their uptake properties were compared with aegPNA and TAT peptide. We have demonstrated that GP-PNA traversed the Hela cells more efficiently than AP-PNA and aegPNA, but still not as good as TAT peptide. One possible explanation is that the cell permeability is affected by the

number of guanidino headgroups, the molecular weight of the oligomers as well as the type of cell-line. Therefore, structure-function relationship studies will be performed on fluorescein-tagged probes of GP-PNA and other control transporters to determine the minimum and optimal number of guanidino groups required for cell permeability. The kinetics and mechanism of cellular entry will also be studied. Intracellular distribution and localization of the probe will be analyzed concurrently.

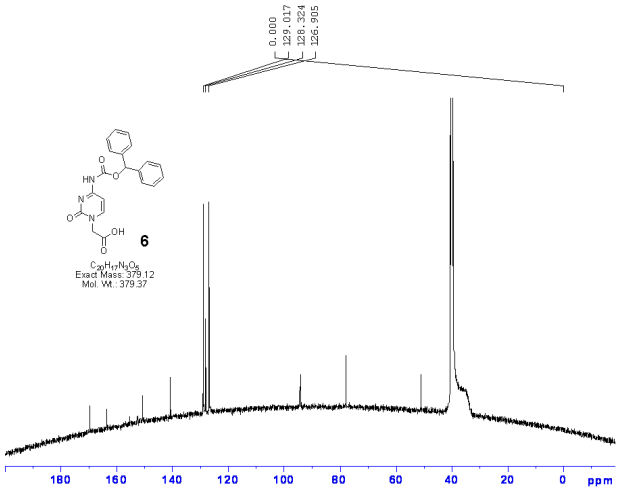
# Appendix



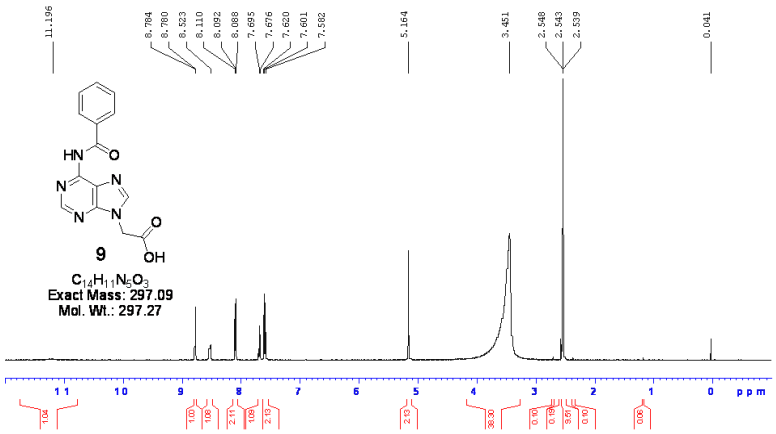

$^1$H NMR (d-DMSO) spectrum of **3**



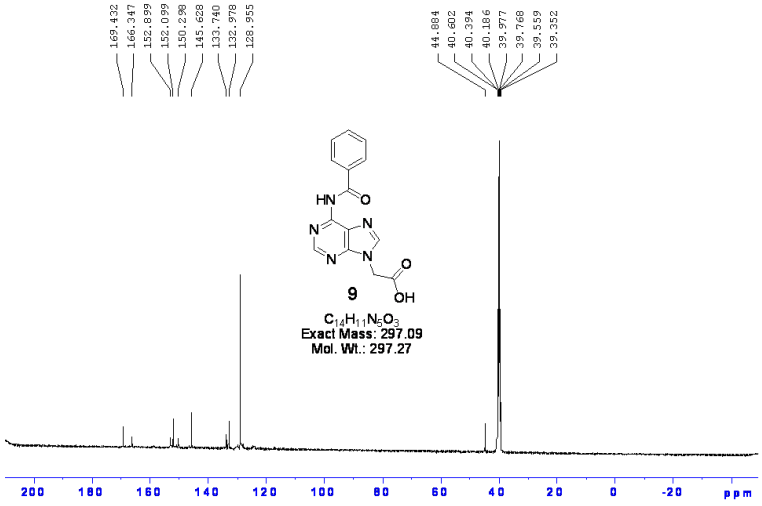

$^1$H NMR (d-DMSO) spectrum of **6**



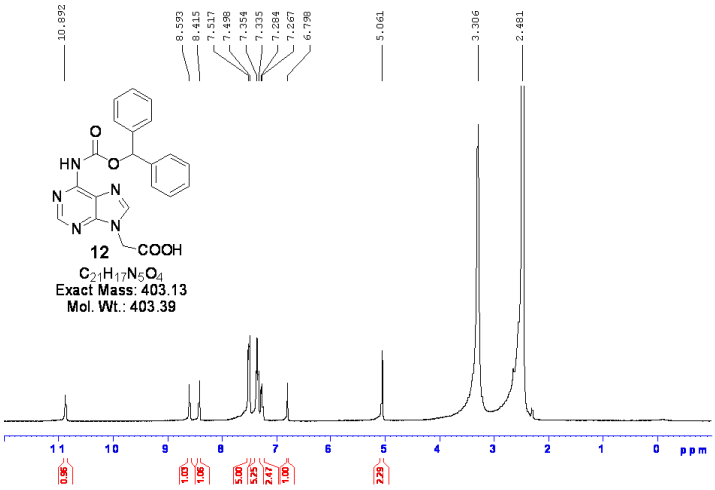

$^{13}$C NMR (d-DMSO) spectrum of **6**



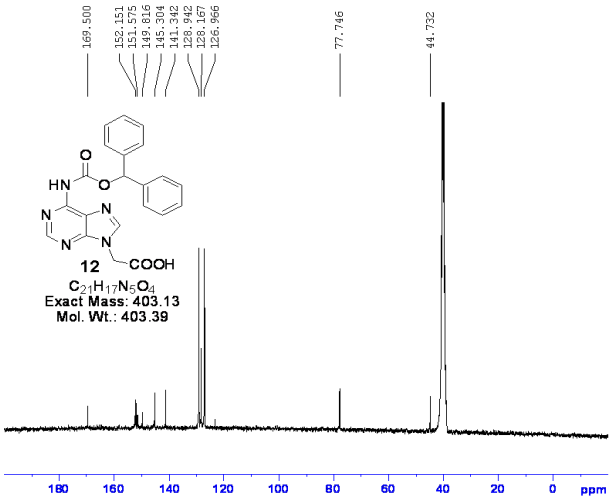

$^{1}$H NMR (d-DMSO) spectrum of **9**



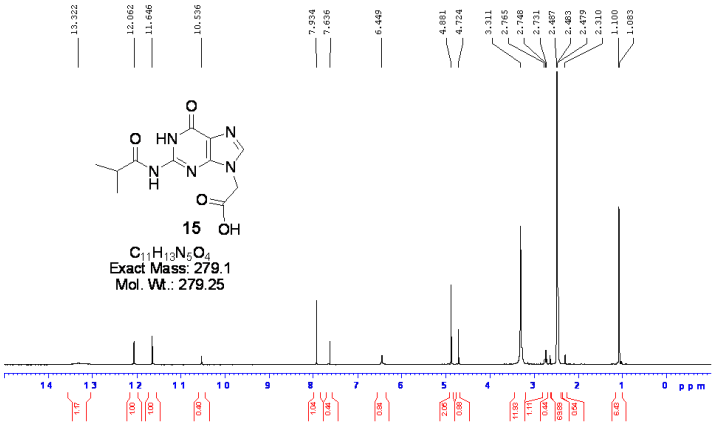

$^{13}$C NMR (d-DMSO) spectrum of **9**



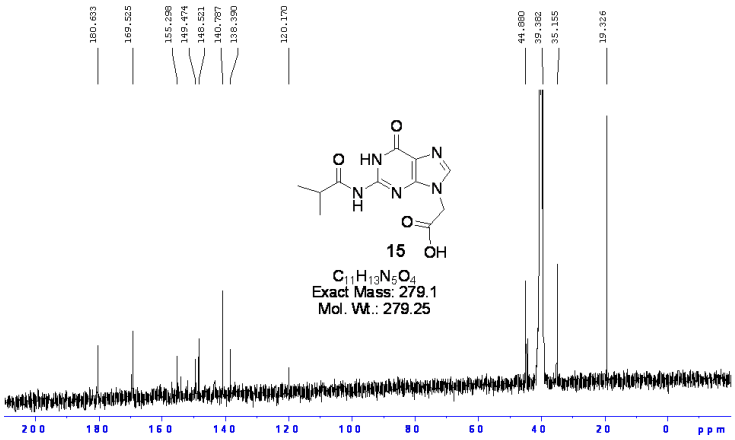

$^{1}$H NMR (d-DMSO) spectrum of **12**



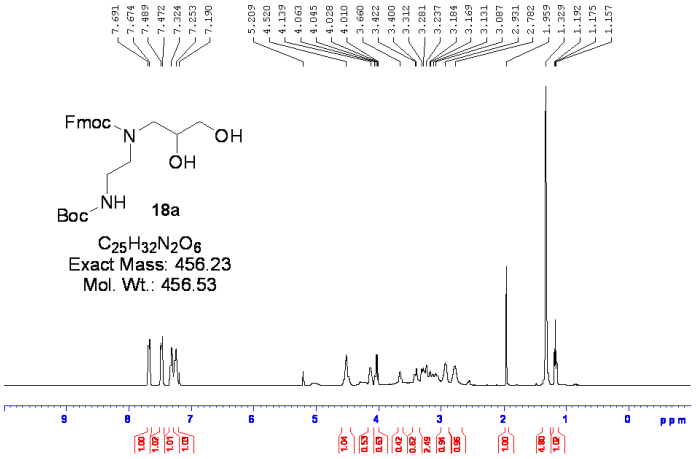

$^{13}$C NMR (d-DMSO) spectrum of **12**



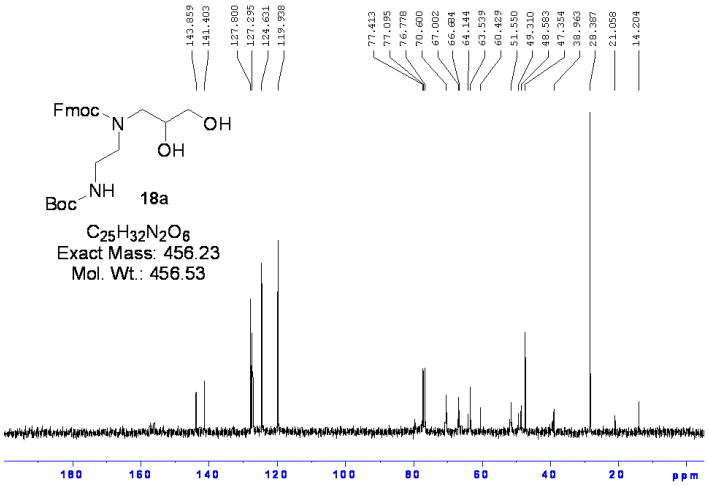

$^{1}$H NMR (d-DMSO) spectrum of **15**



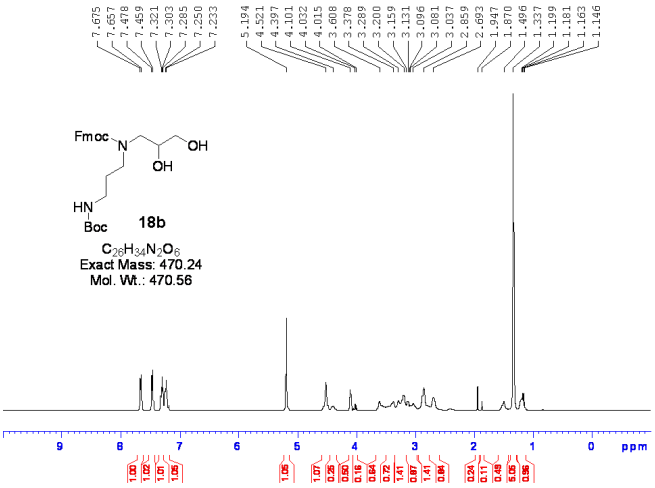

$^{13}$C NMR (d-DMSO) spectrum of **15**



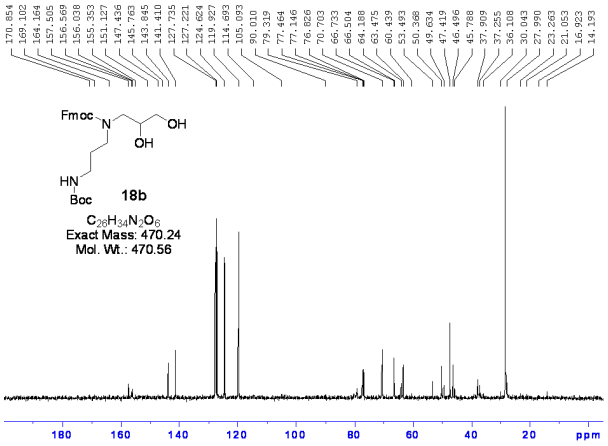

$^{1}$H NMR ($CDCl_3$) spectrum of **18a**



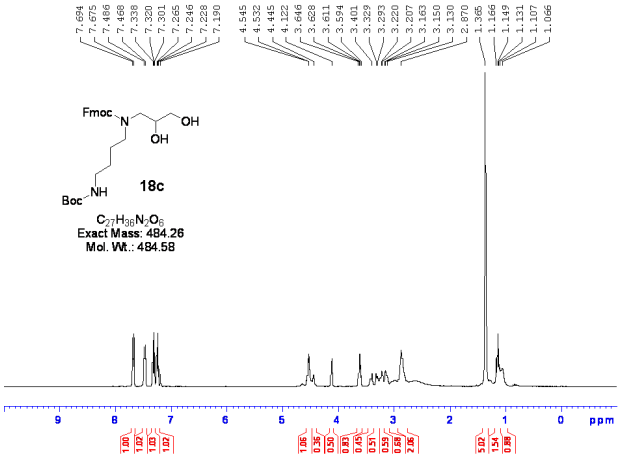

$^{13}$C NMR ($CDCl_3$) spectrum of **18a**



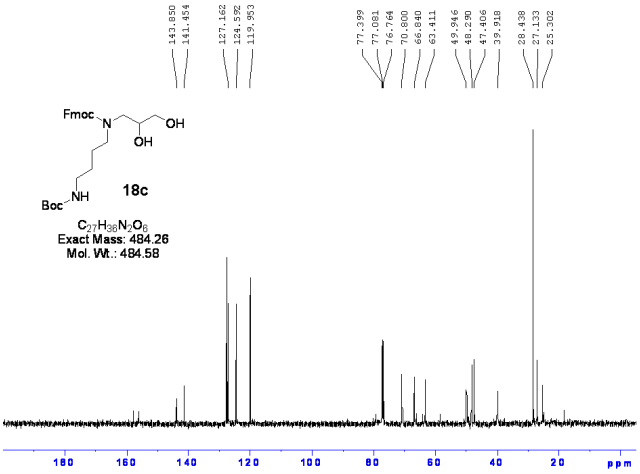

$^{1}$H NMR ($CDCl_3$) spectrum of **18b**



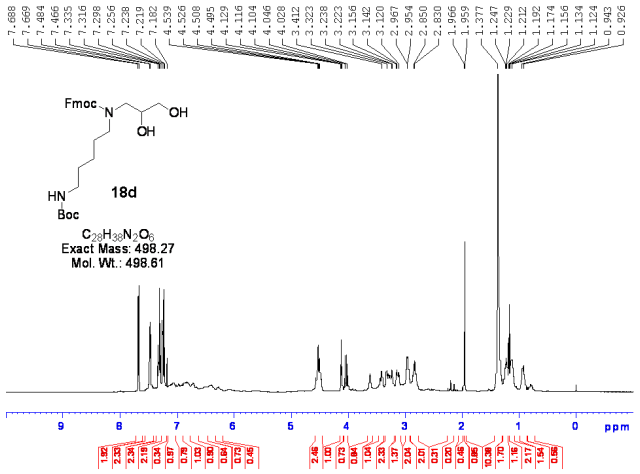

$^{13}$C NMR ($CDCl_3$) spectrum of **18b**



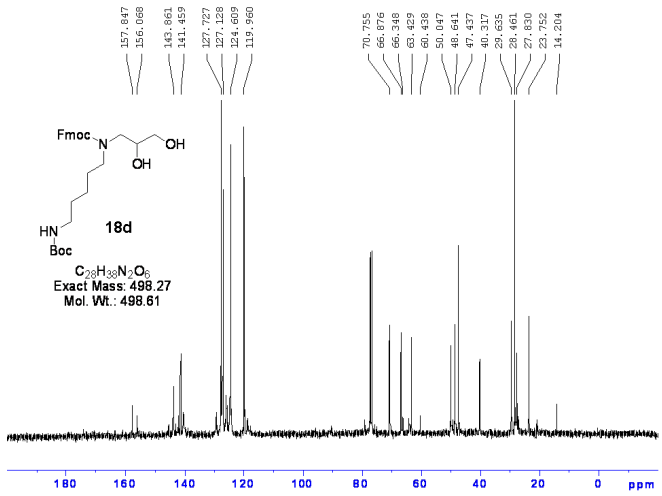

$^{1}$H NMR ($CDCl_3$) spectrum of **18c**



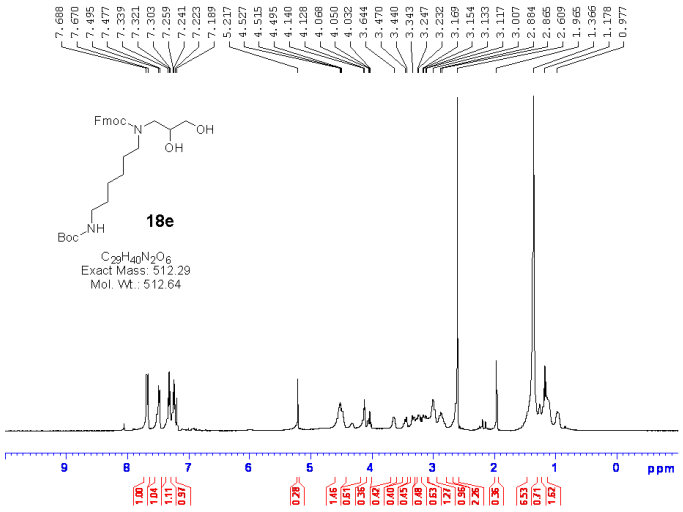

$^{13}$C NMR ($CDCl_3$) spectrum of **18c**



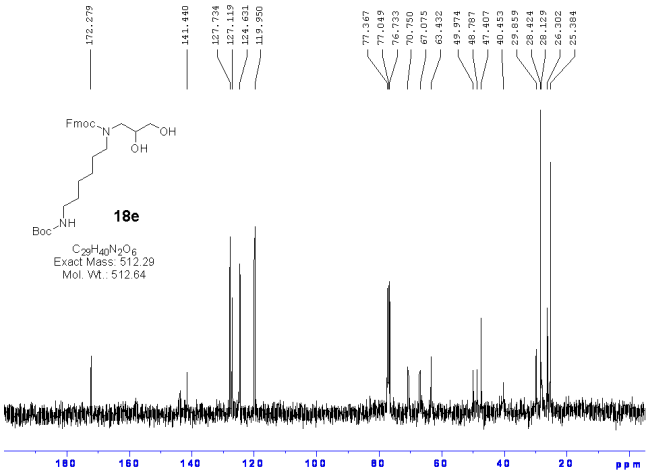

$^{1}$H NMR ($CDCl_3$) spectrum of **18d**



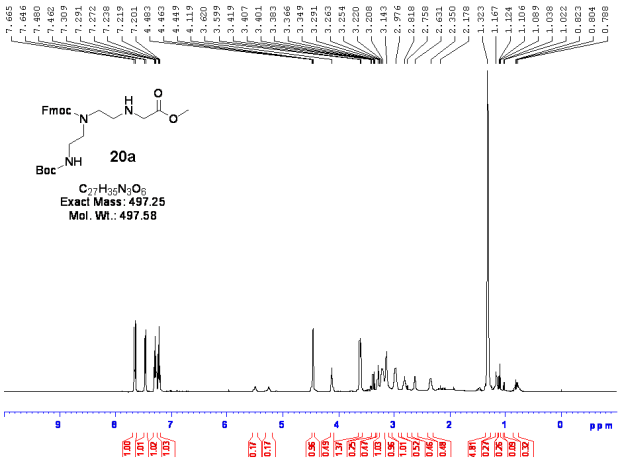

$^{13}$C NMR ($CDCl_3$) spectrum of **18d**



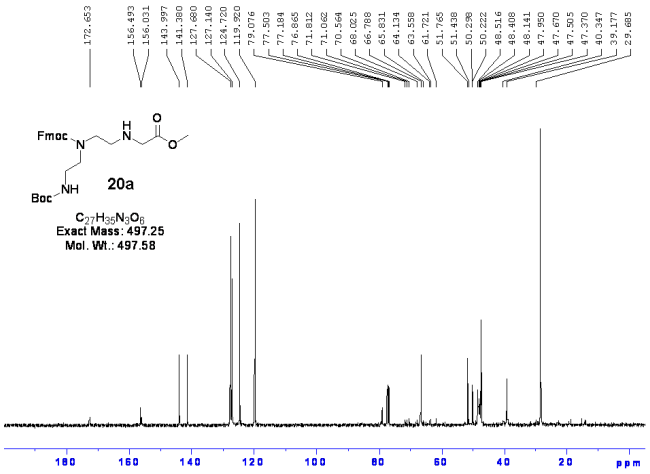

$^{1}$H NMR ($CDCl_3$) spectrum of **18e**



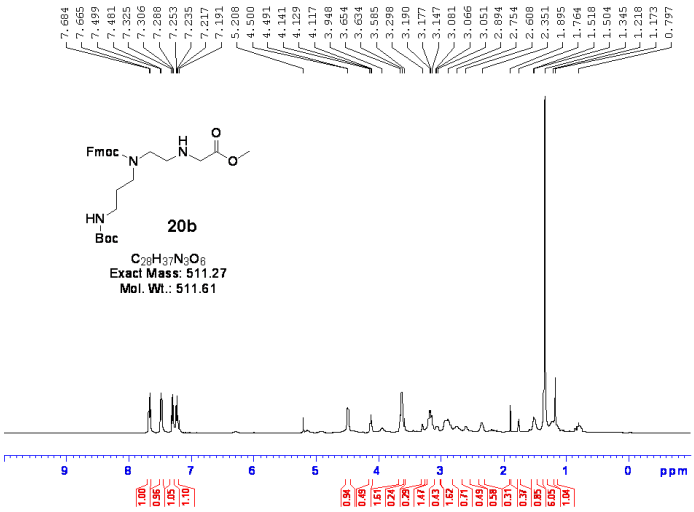

$^{13}$C NMR ($CDCl_3$) spectrum of **18e**



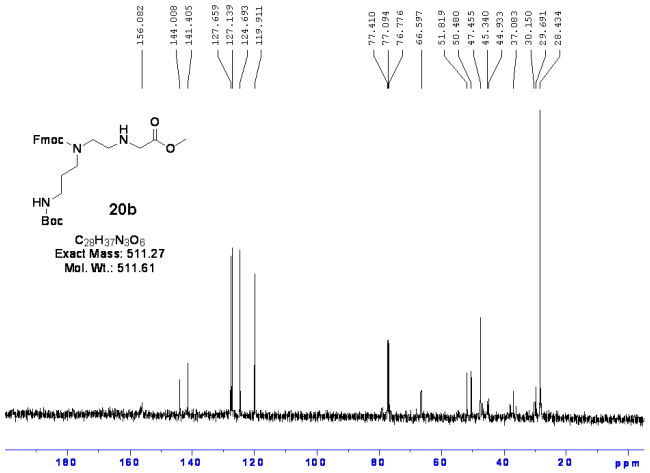

$^{1}$H NMR ($CDCl_3$) spectrum of **20a**



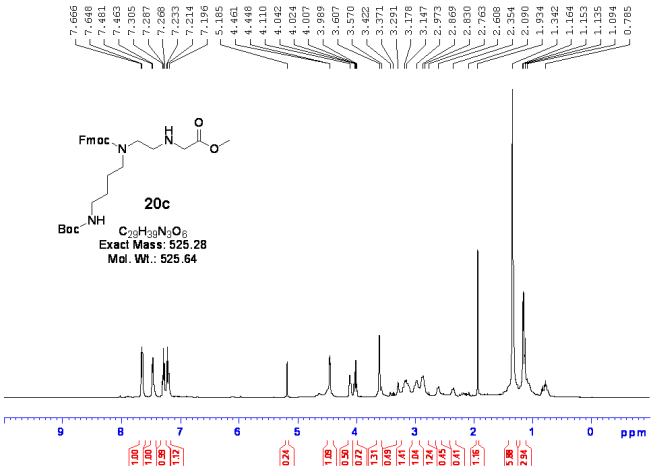

$^{13}$C NMR ($CDCl_3$) spectrum of **20**[a]



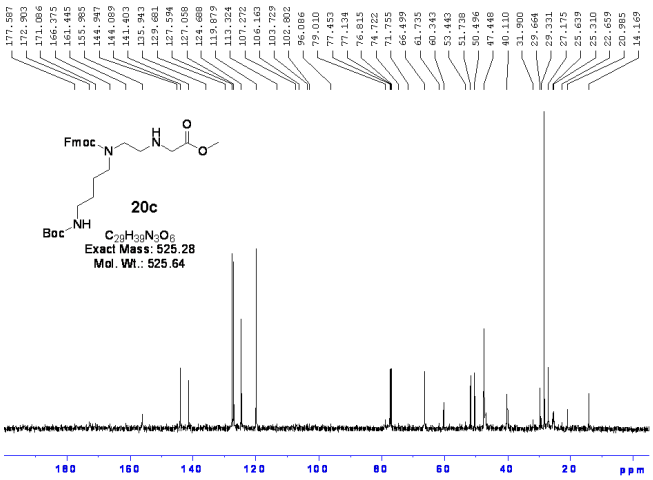

$^{1}$H NMR ($CDCl_3$) spectrum of **20b**



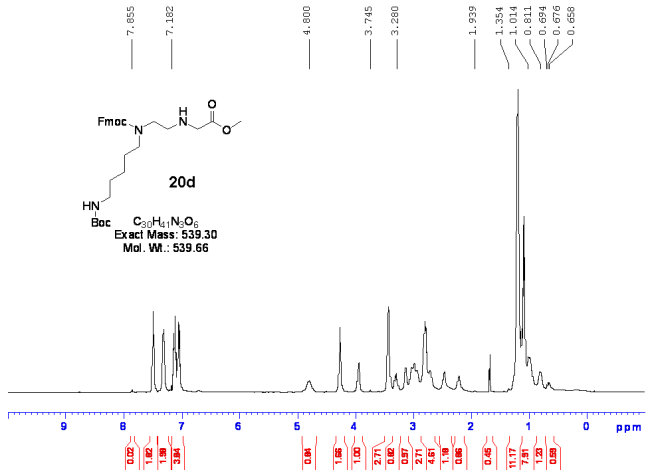

$^{13}$C NMR ($CDCl_3$) spectrum of **20b**



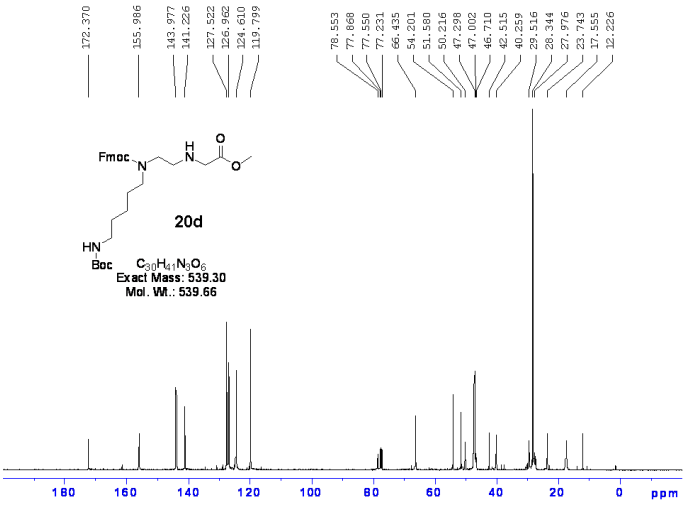

$^{1}$H NMR ($CDCl_3$) spectrum of **20c**



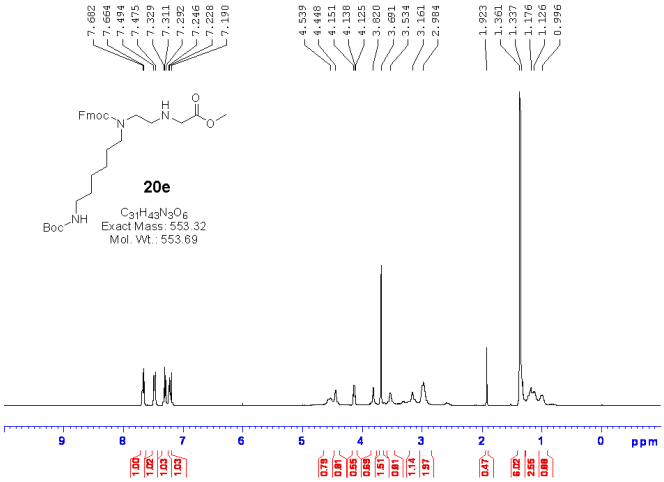

$^{13}$C NMR ($CDCl_3$) spectrum of **20c**



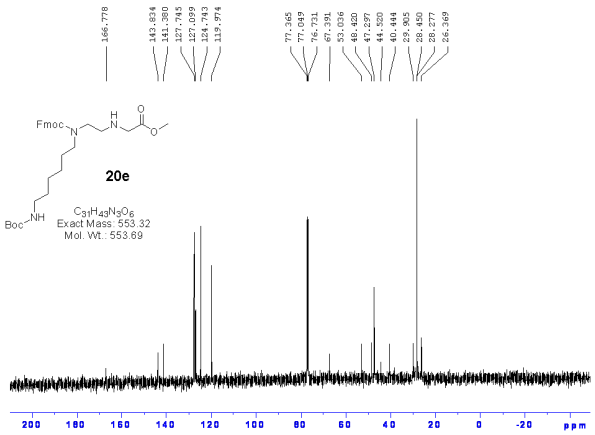

$^{1}$H NMR ($CDCl_3$) spectrum of **20d**



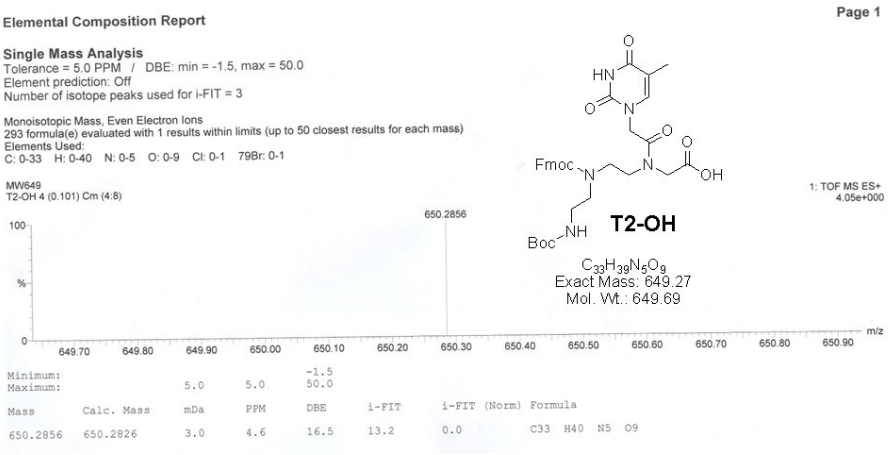

$^{13}$C NMR ($CDCl_3$) spectrum of **20d**



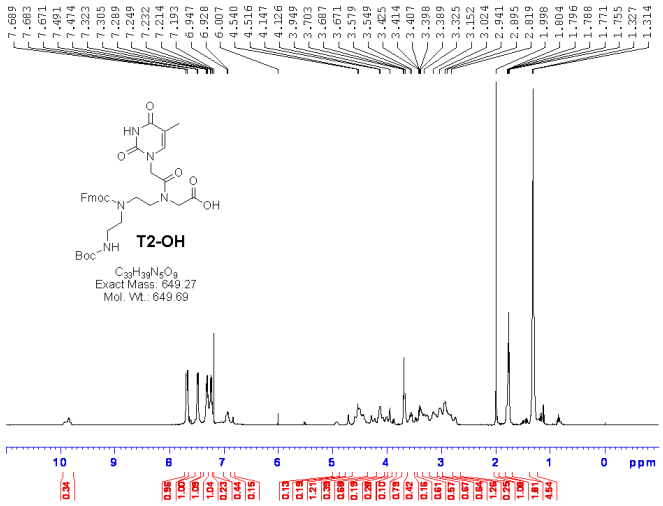

$^{1}$H NMR ($CDCl_3$) spectrum of **20e**



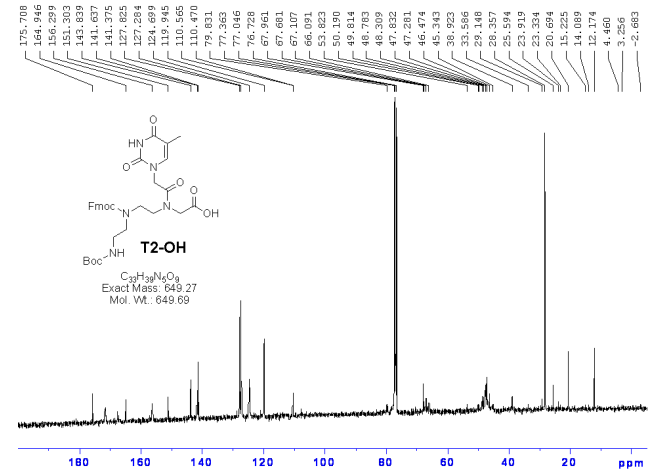

$^{13}$C NMR ($CDCl_3$) spectrum of **20e**



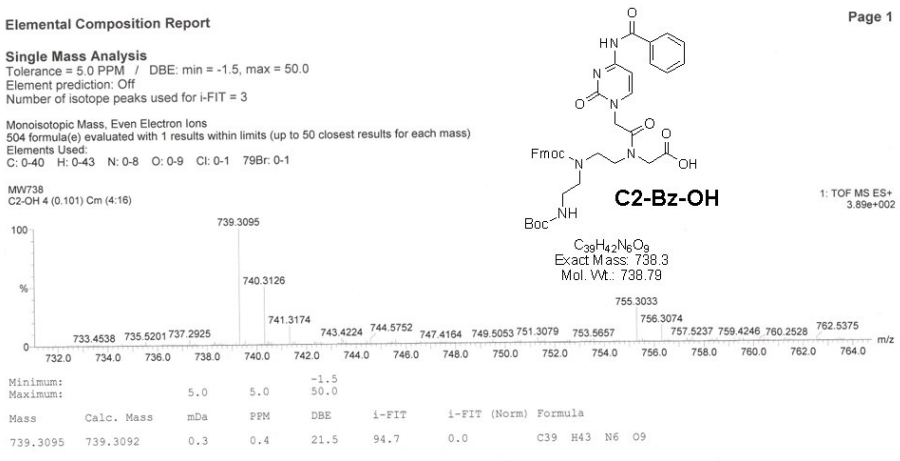

HRMS spectrum of **T2-OH**



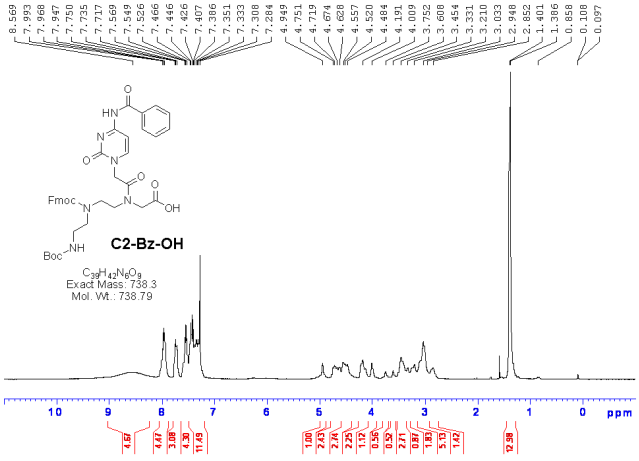

$^{1}$H NMR ($CDCl_3$) spectrum of **T2-OH**



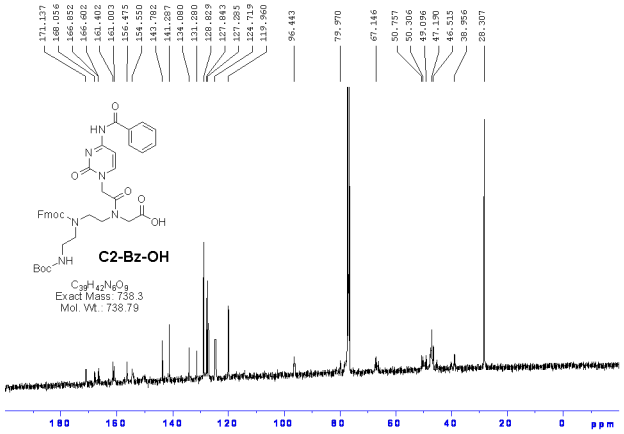

$^{13}$C NMR ($CDCl_3$) spectrum of **T2-OH**



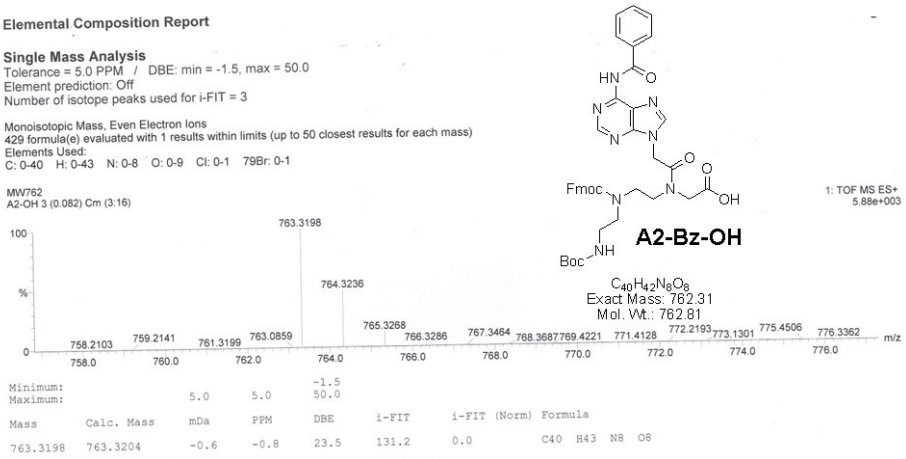

HRMS spectrum of **C2-Bz-OH**



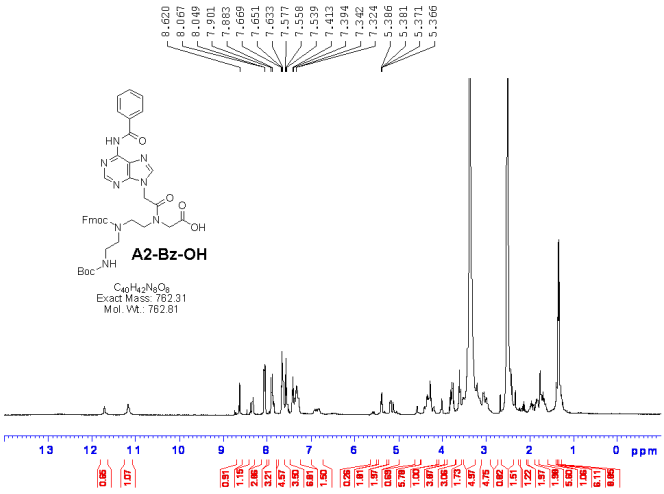

$^{1}$H NMR ($CDCl_3$) spectrum of **C2-Bz-OH**



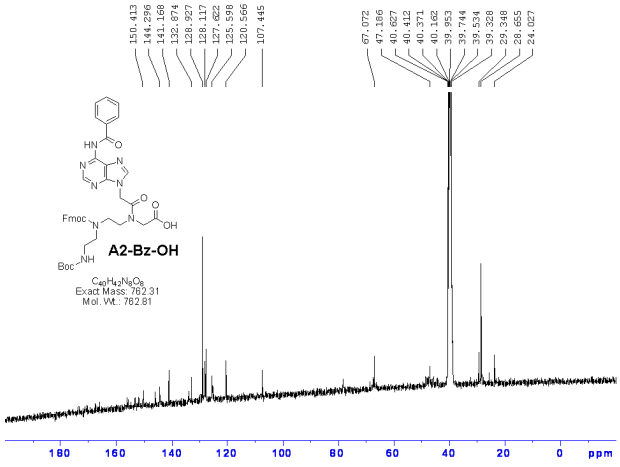

$^{13}$C NMR ($CDCl_3$) spectrum of **C2-Bz-OH**



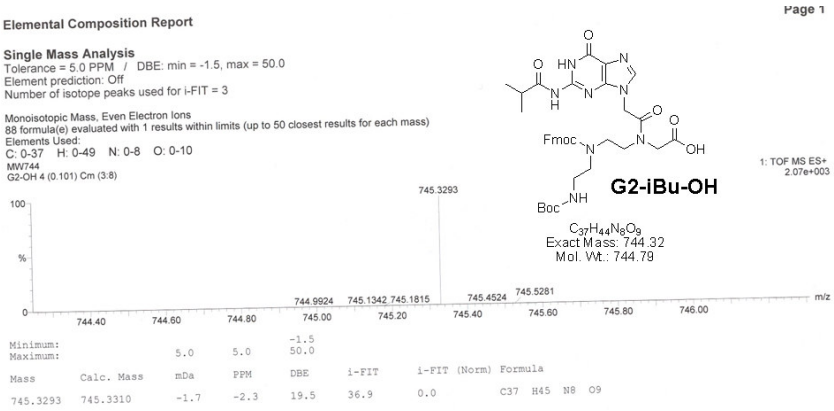

HRMS spectrum of **A2-Bz-OH**



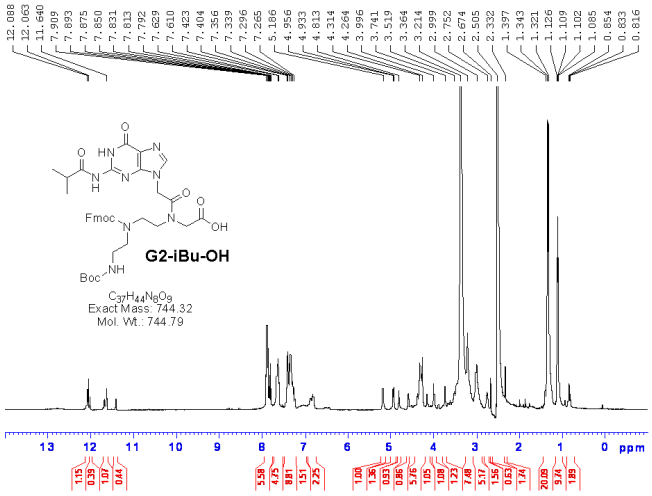

$^{1}$H NMR (d-DMSO) spectrum of **A2-Bz-OH**



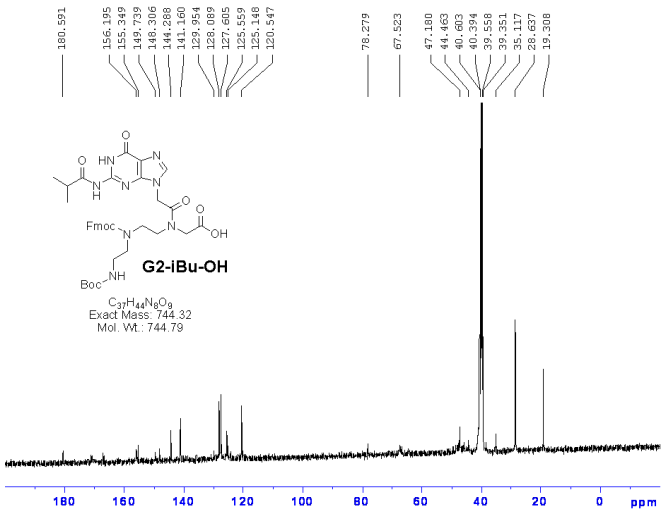

$^{13}$C NMR (d-DMSO) spectrum of **A2-Bz-OH**



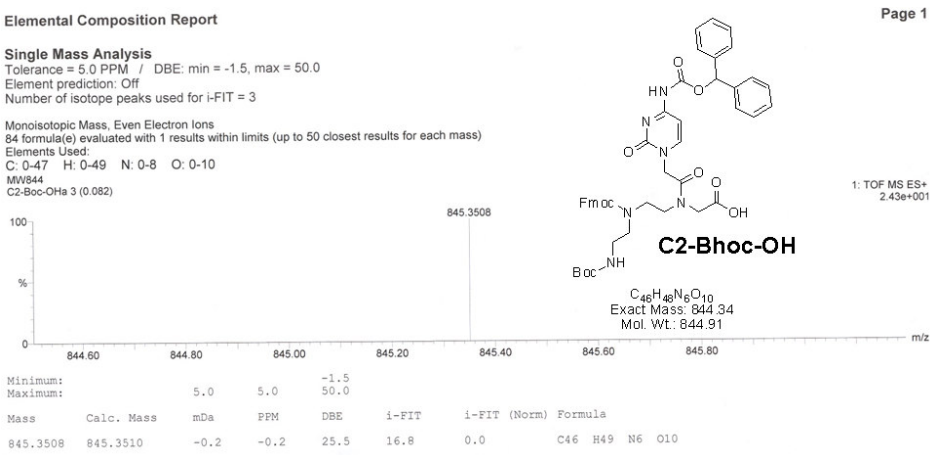

HRMS spectrum of **G2-iBu-OH**



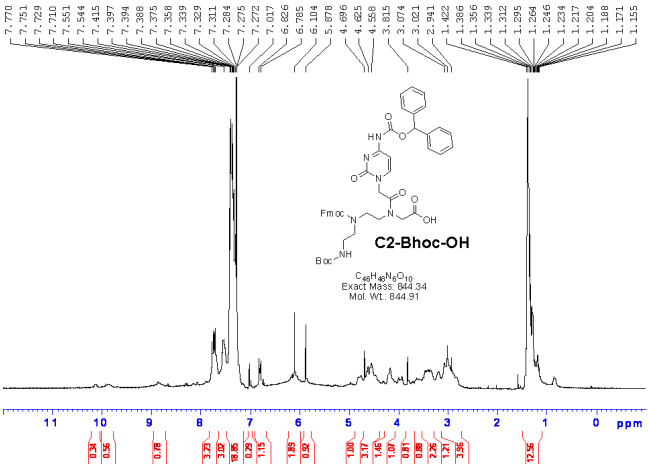

$^{1}$H NMR (d-DMSO) spectrum of **G2-iBu-OH**



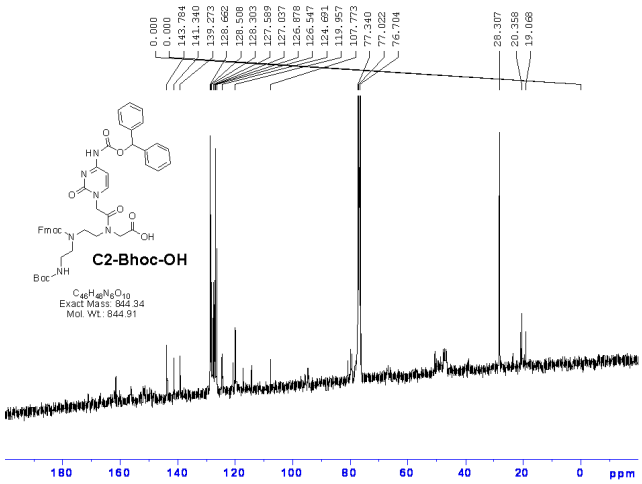

$^{13}$C NMR (d-DMSO) spectrum of **G2-iBu-OH**



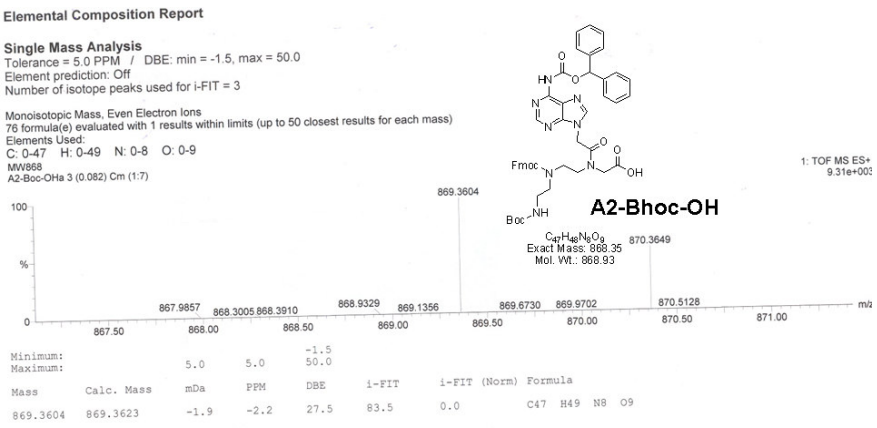

HRMS spectrum of **C2-Bhoc-OH**



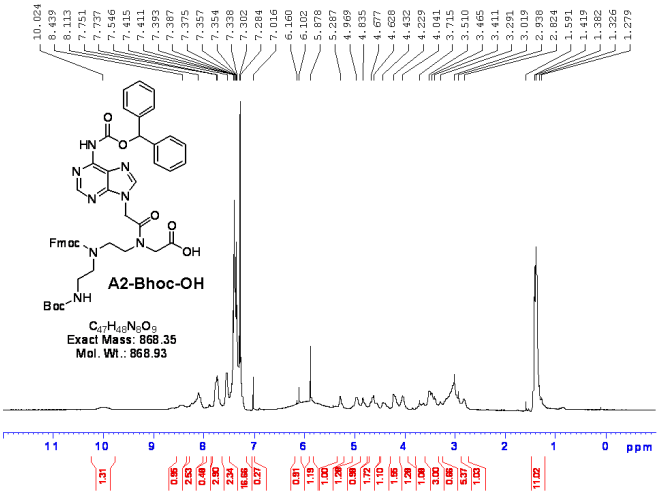

$^{1}$H NMR ($CDCl_3$) spectrum of **C2-Bhoc-OH**



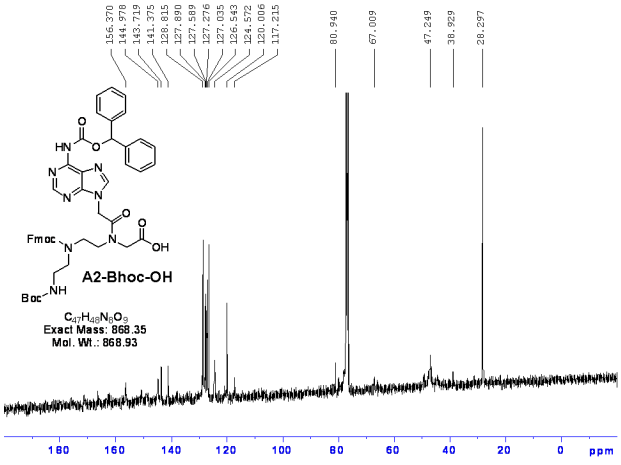

$^{13}$C NMR ($CDCl_3$) spectrum of **C2-Bhoc-OH**



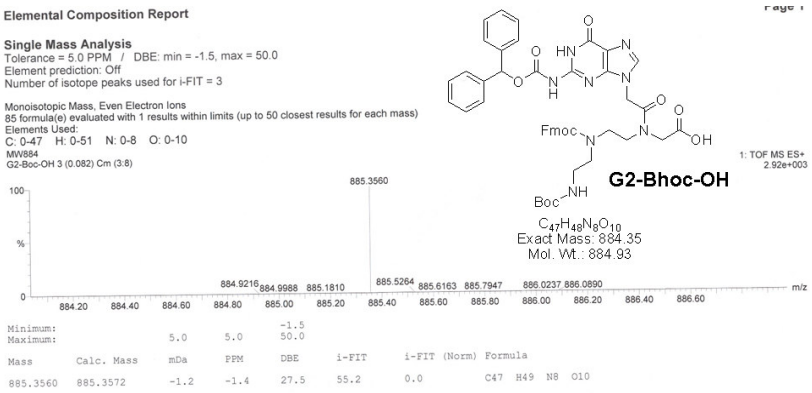

HRMS spectrum of **A2-Bhoc-OH**



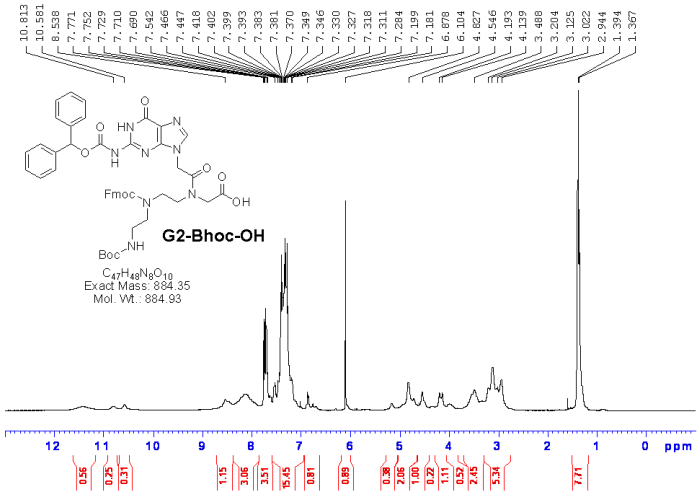

$^{1}$H NMR ($CDCl_3$) spectrum of **A2-Bhoc-OH**



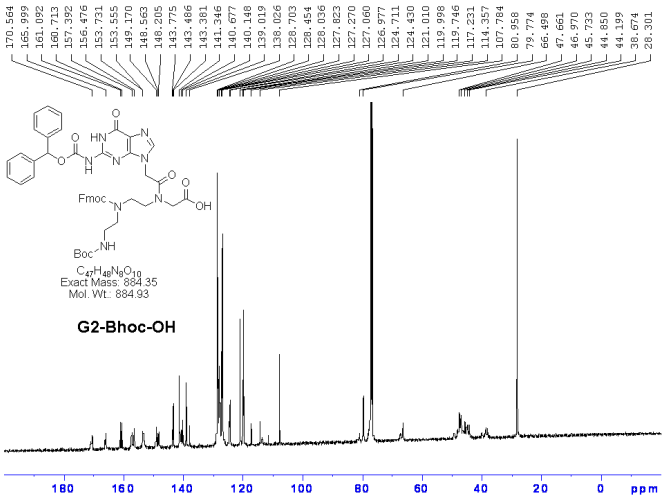

$^{13}$C NMR ($CDCl_3$) spectrum of **A2-Bhoc-OH**



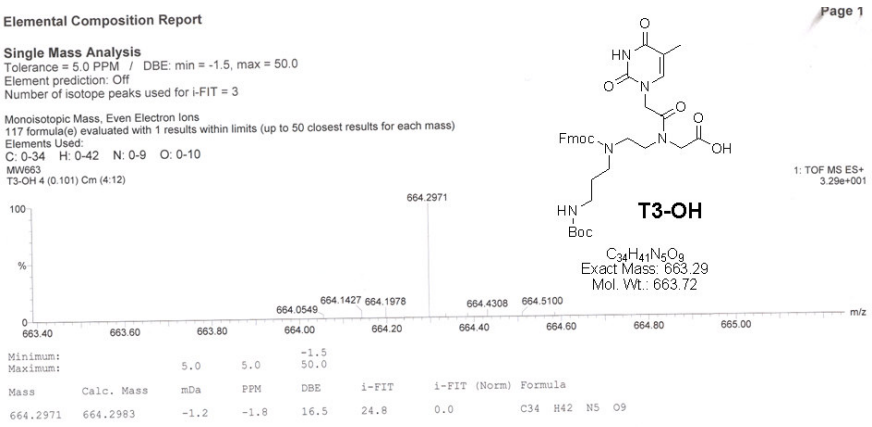

HRMS spectrum of **G2-Bhoc-OH**



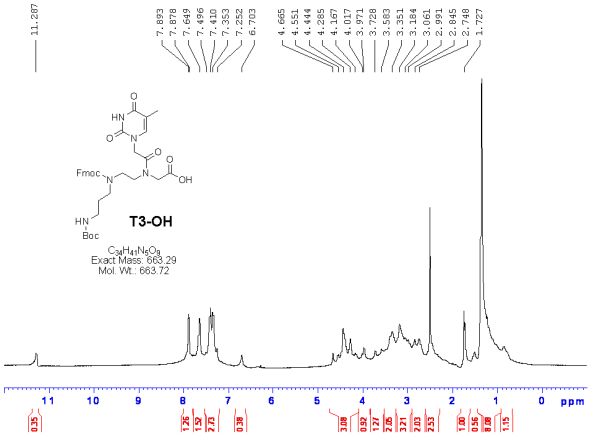

$^{1}$H NMR ($CDCl_3$) spectrum of **G2-Bhoc-OH**



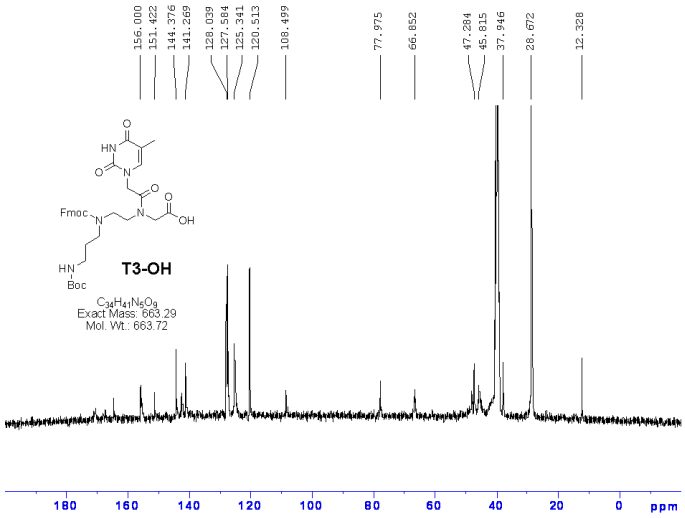

$^{13}$C NMR ($CDCl_3$) spectrum of **G2-Bhoc-OH**



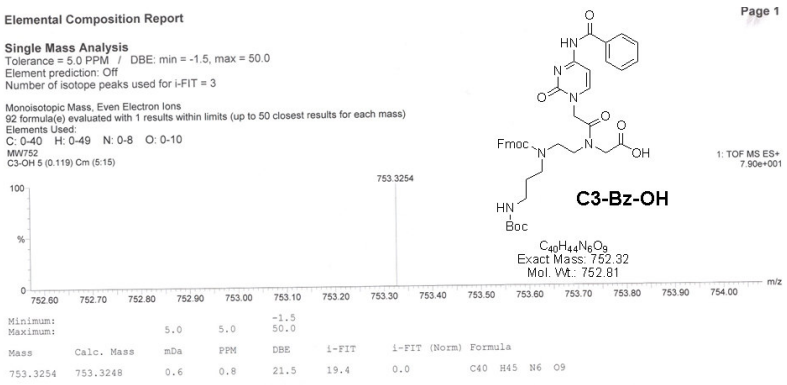

HRMS spectrum of **T3-OH**



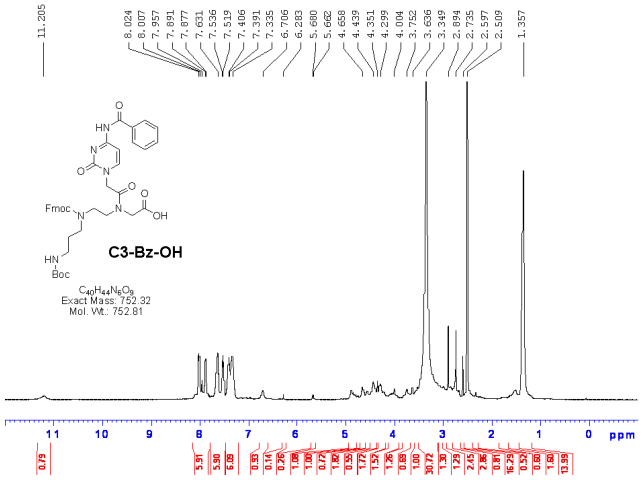

$^{1}$H NMR (d-DMSO) spectrum of **T3-OH**



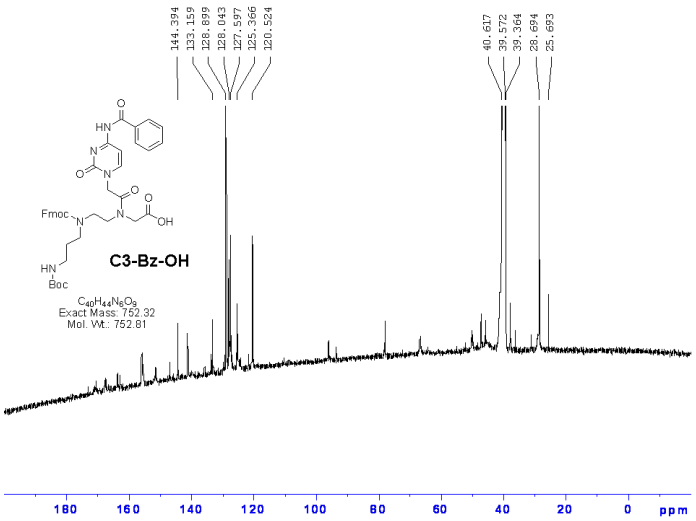

$^{13}$C NMR (d-DMSO) spectrum of **T3-OH**



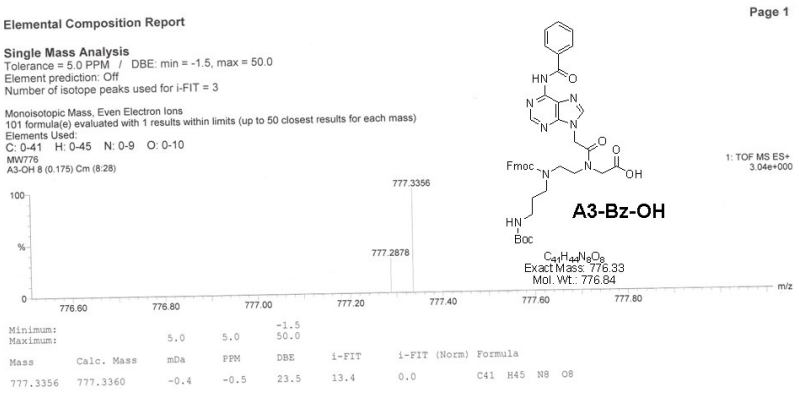

HRMS spectrum of **C3-Bz-OH**



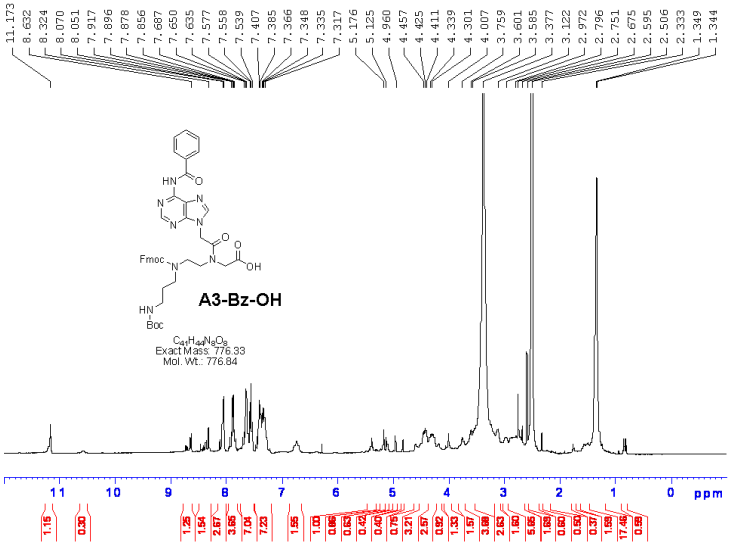

$^{1}$H NMR (d-DMSO) spectrum of **C3-Bz-OH**



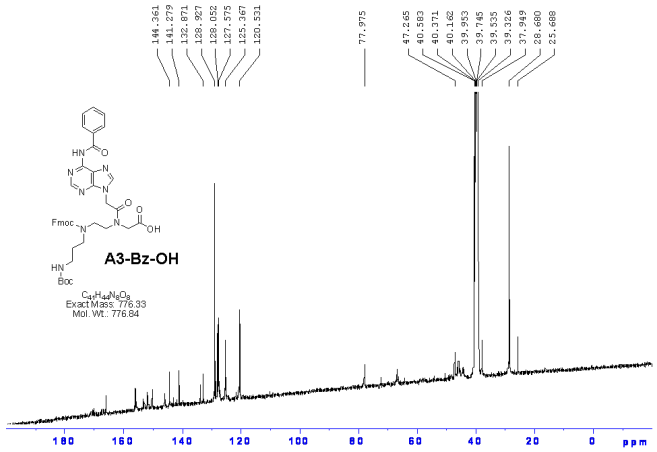

$^{13}$C NMR (d-DMSO) spectrum of **C3-Bz-OH**



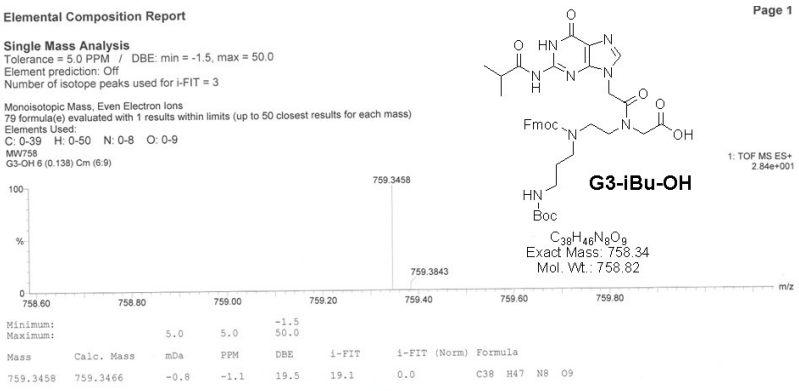

HRMS spectrum of **A3-Bz-OH**



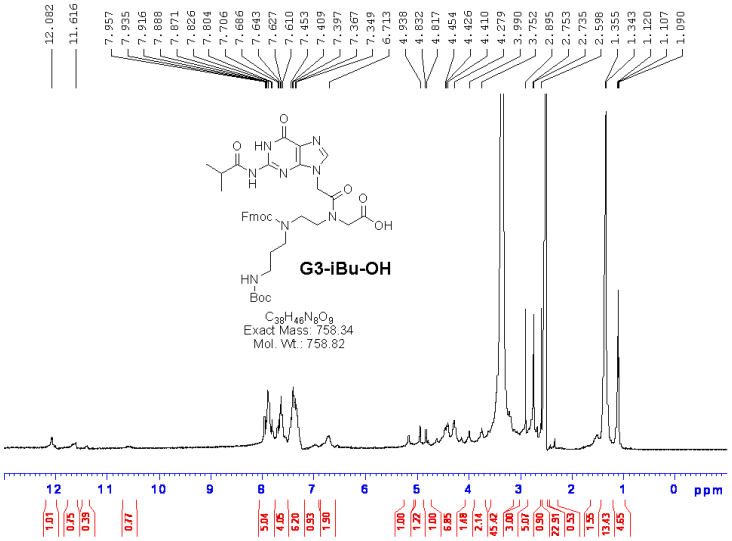

$^{1}$H NMR (d-DMSO) spectrum of **A3-Bz-OH**



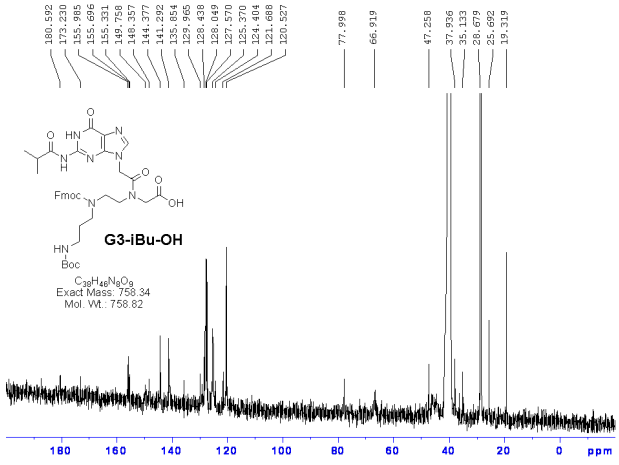

$^{13}$C NMR (d-DMSO) spectrum of **A3-Bz-OH**



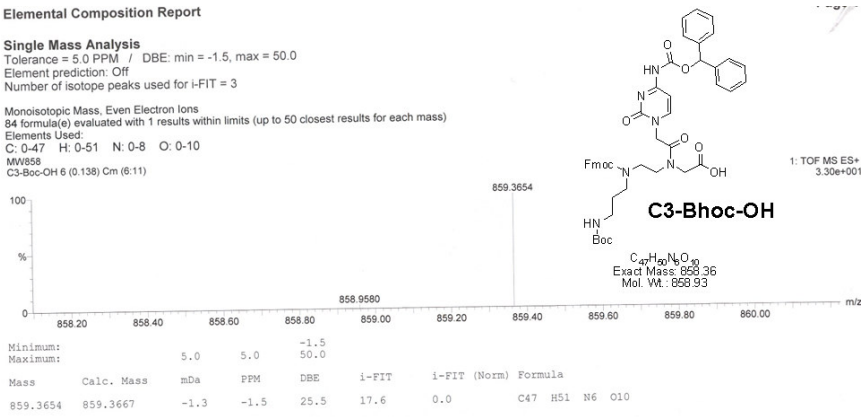

HRMS spectrum of **G3-iBu-OH**



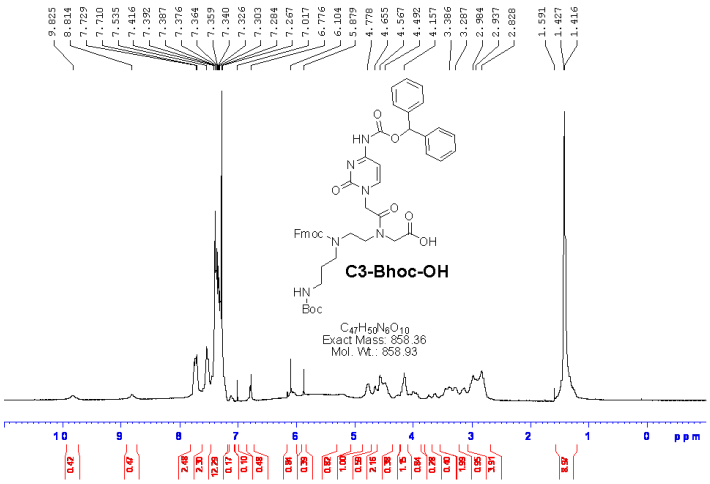

$^{1}$H NMR (d-DMSO) spectrum of **G3-iBu-OH**



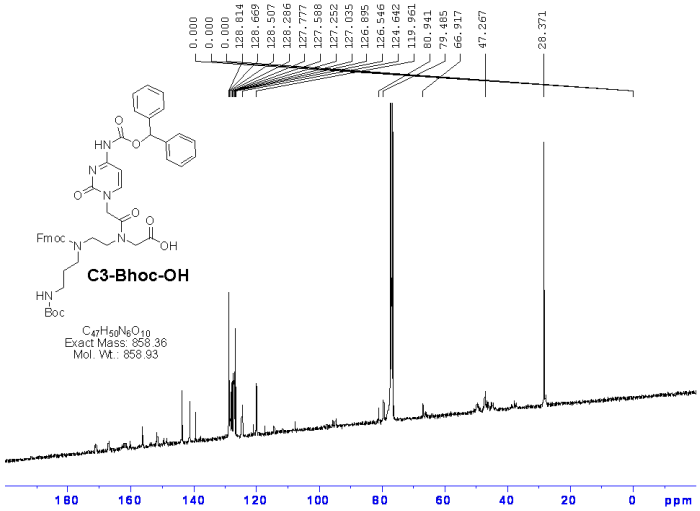

$^{13}$C NMR (d-DMSO) spectrum of **G3-iBu-OH**



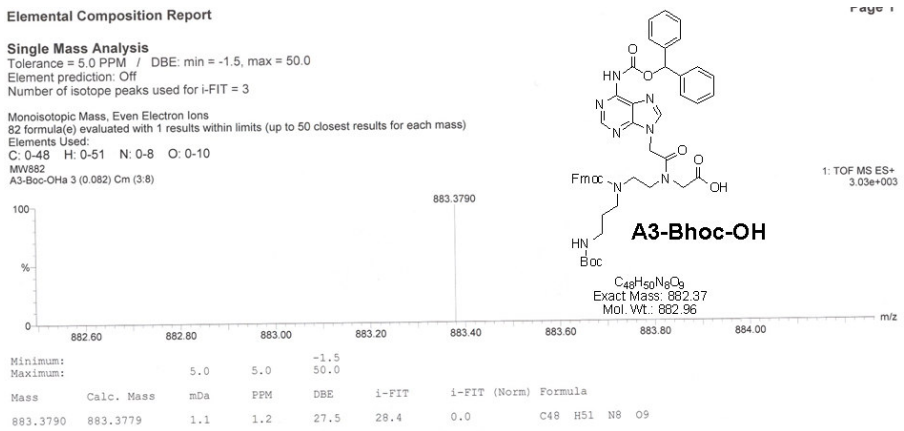

HRMS spectrum of **C3-Bhoc-OH**



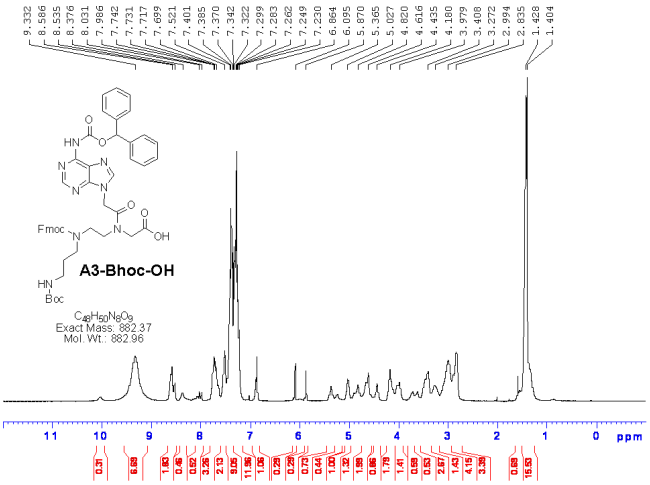

$^{1}$H NMR (CDCl3) spectrum of **C3-Bhoc-OH**



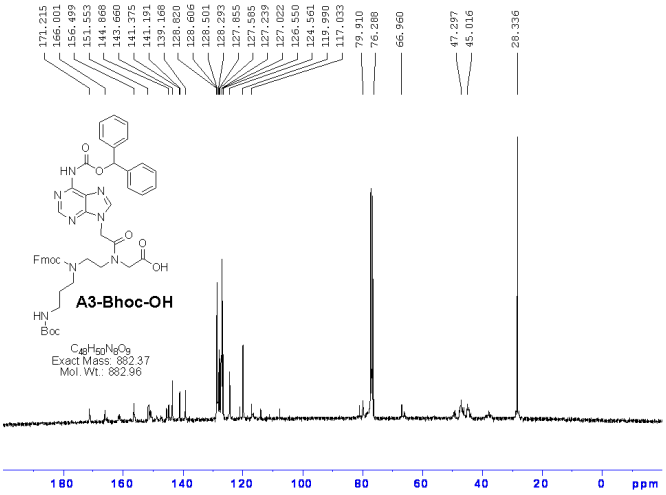

$^{13}$C NMR (CDCl3) spectrum of **C3-Bhoc-OH**



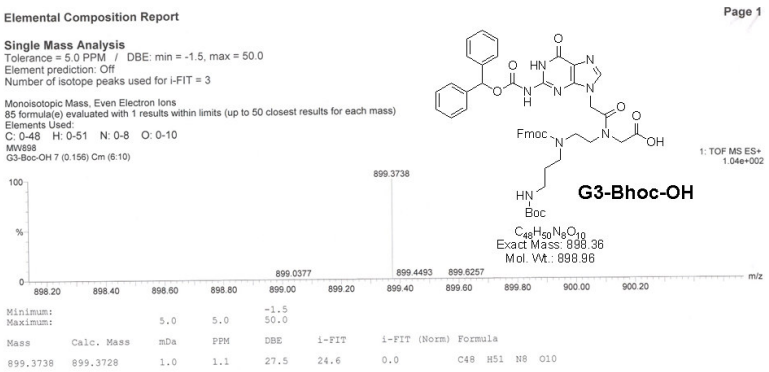

HRMS spectrum of **A3-Bhoc-OH**



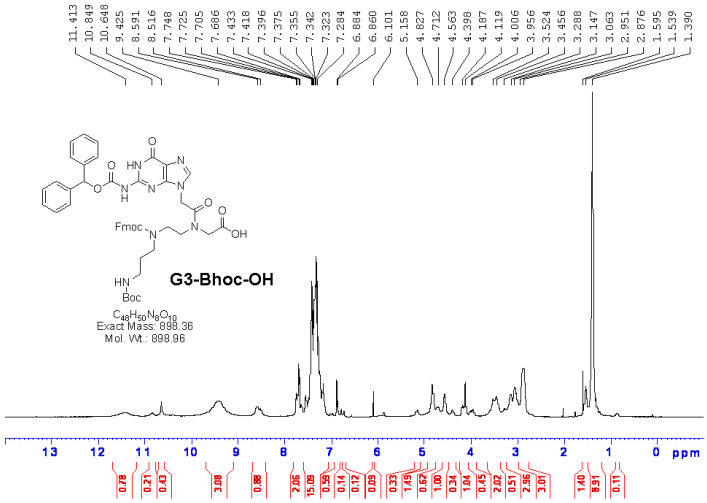

$^{1}$H NMR (CDCl3) spectrum of **A3-Bhoc-OH**



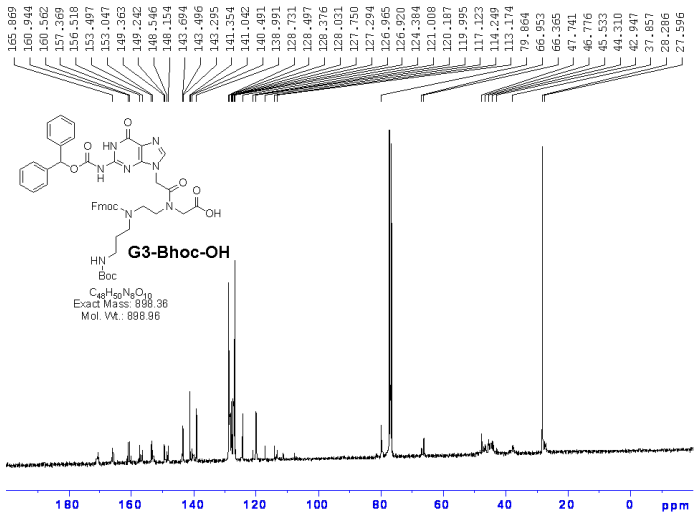

$^{13}$C NMR (CDCl3) spectrum of **A3-Bhoc-OH**



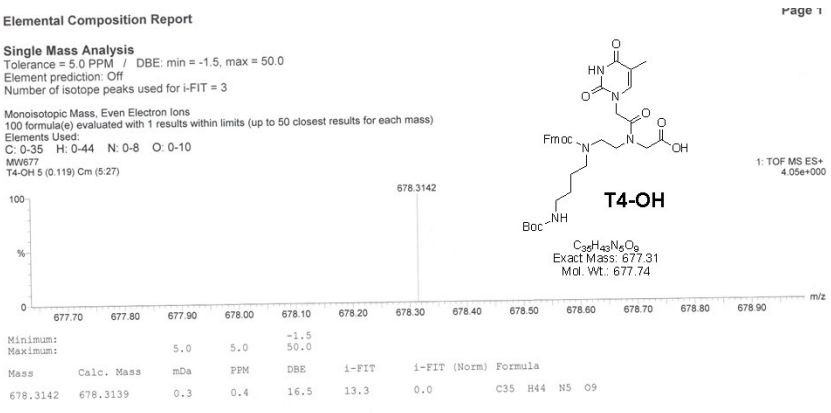

HRMS spectrum of **G3-Bhoc-OH**



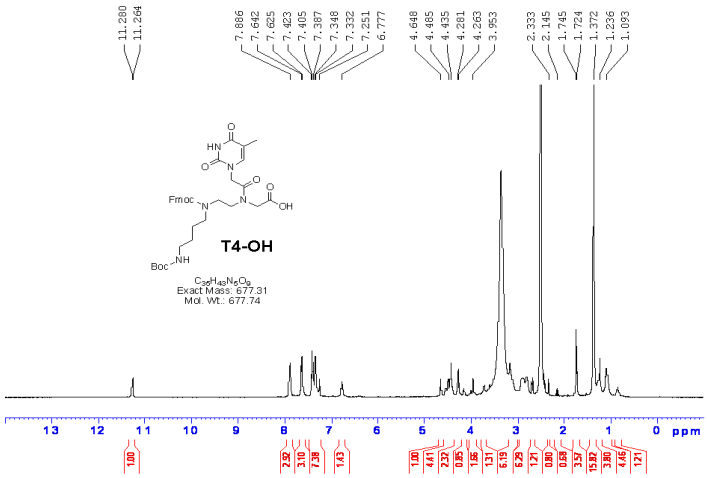

$^{1}$H NMR (CDCl3) spectrum of **G3-Bhoc-OH**



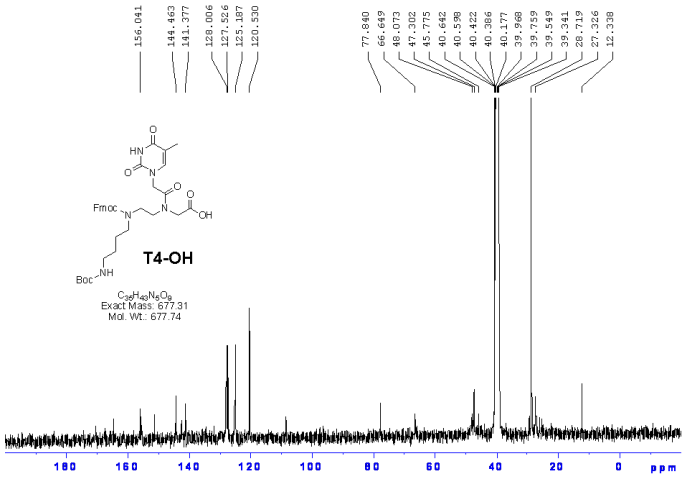

$^{1}$H NMR (CDCl3) spectrum of **G3-Bhoc-OH**



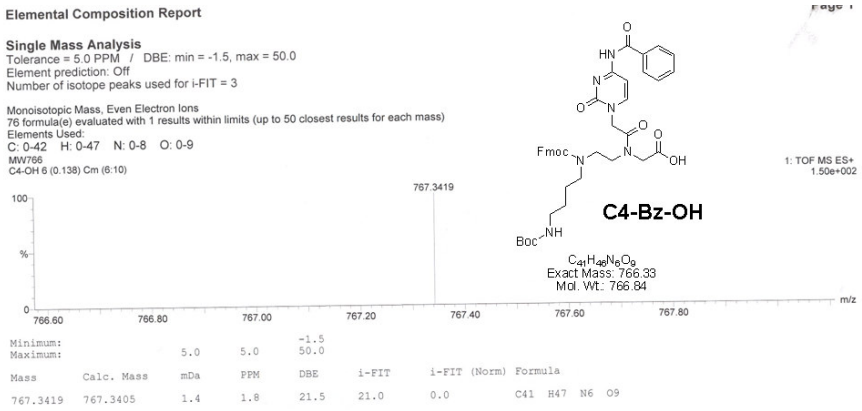

HRMS spectrum of **T4-OH**



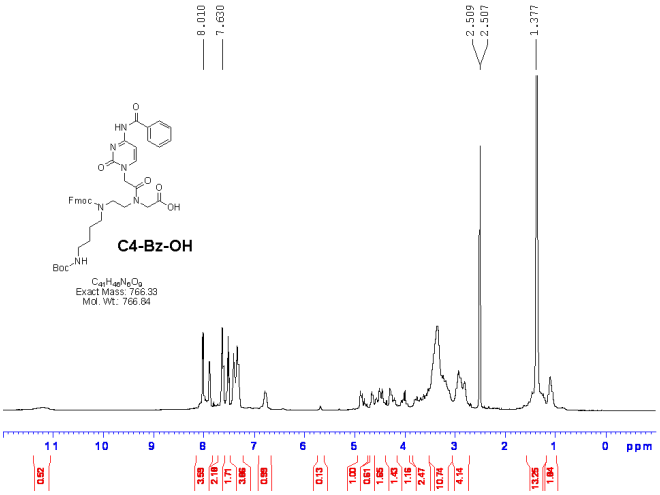

$^{1}$H NMR (DMSO) spectrum of **T4-OH**



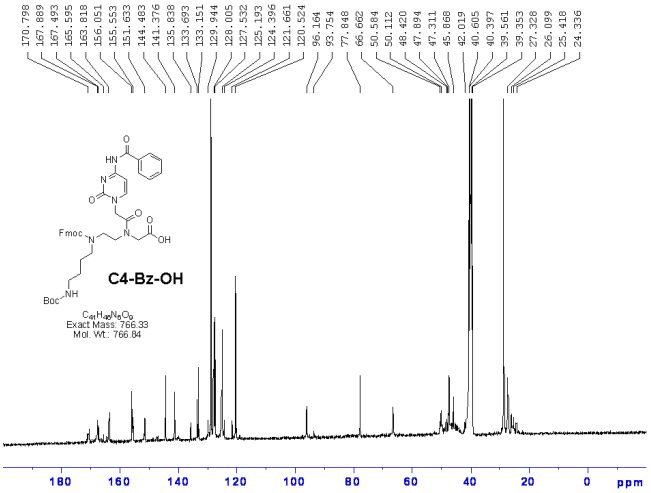

$^{13}$C NMR (DMSO) spectrum of **T4-OH**



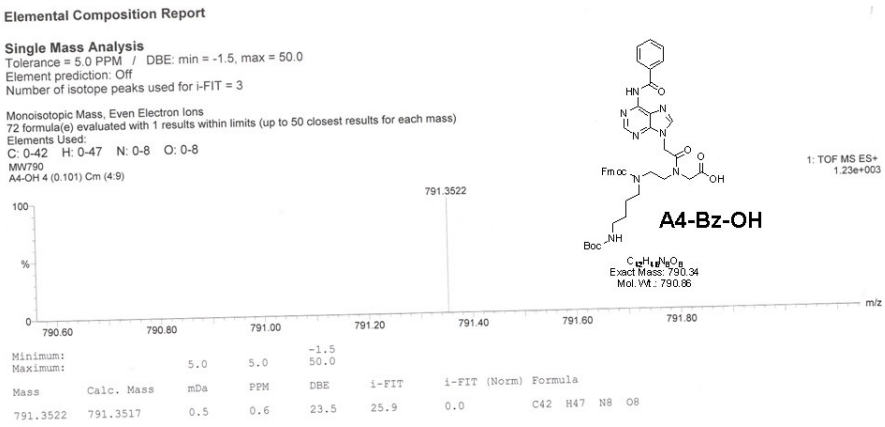

HRMS spectrum of **C4-Bz-OH**



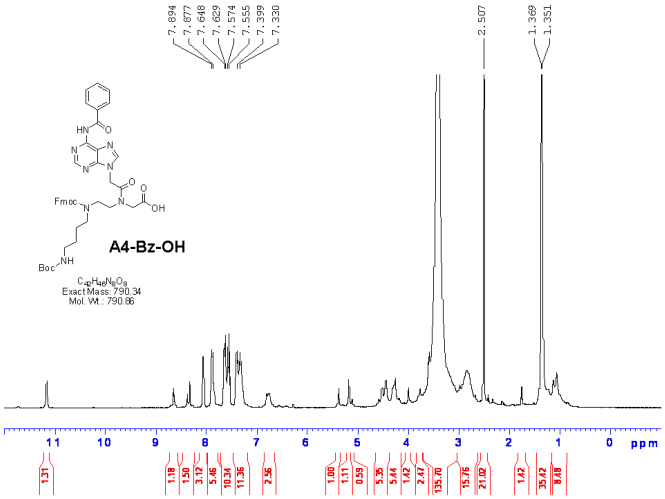

$^{1}$H NMR (DMSO) spectrum of **C4-Bz-OH**



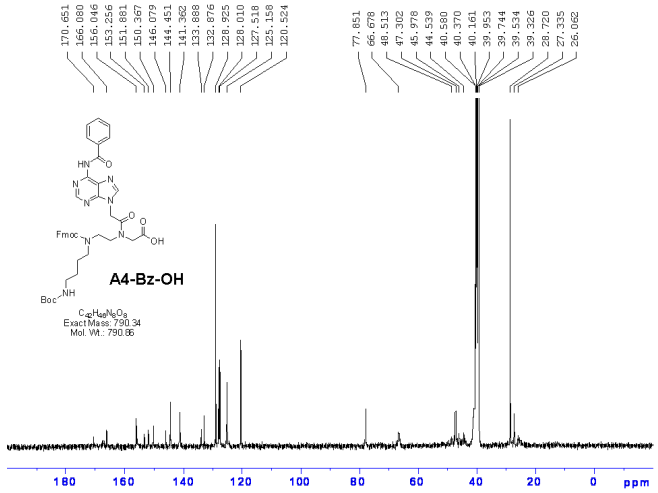

$^{13}$C NMR (DMSO) spectrum of **C4-Bz-OH**



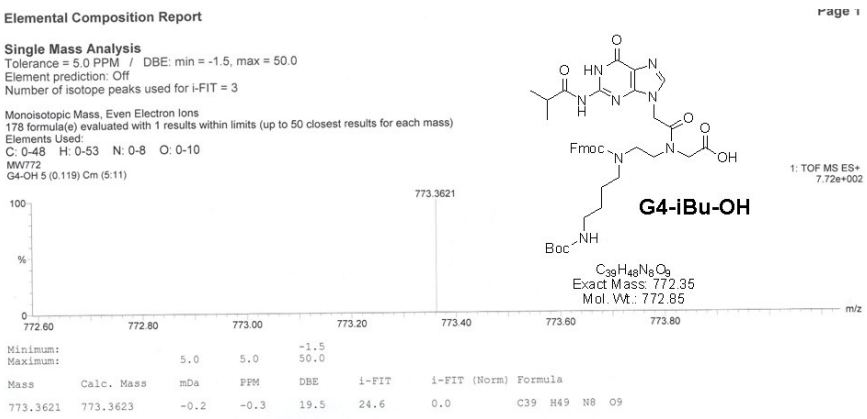

HRMS spectrum of **A4-Bz-OH**



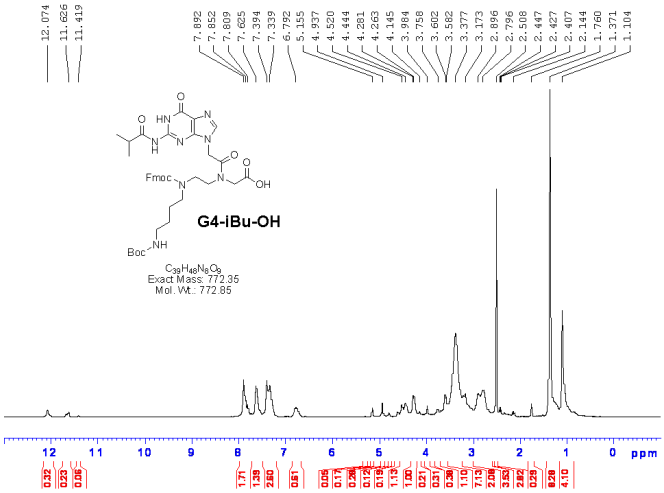

$^{1}$H NMR (DMSO) spectrum of **A4-Bz-OH**



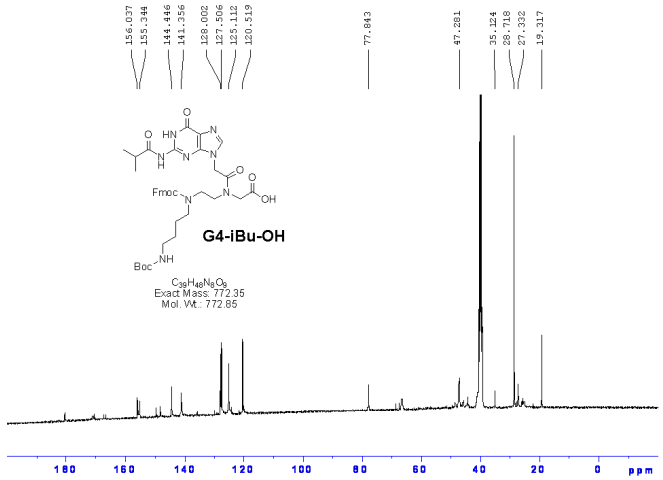

$^{13}$C NMR (DMSO) spectrum of **A4-Bz-OH**



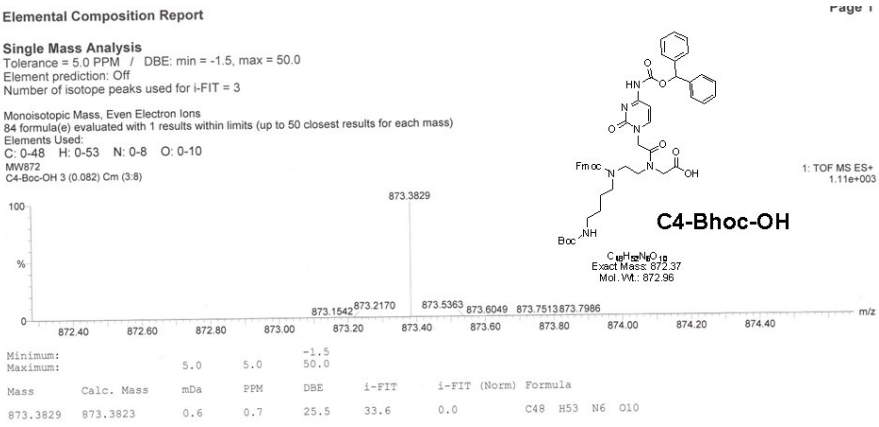

HRMS spectrum of **G4-iBu-OH**



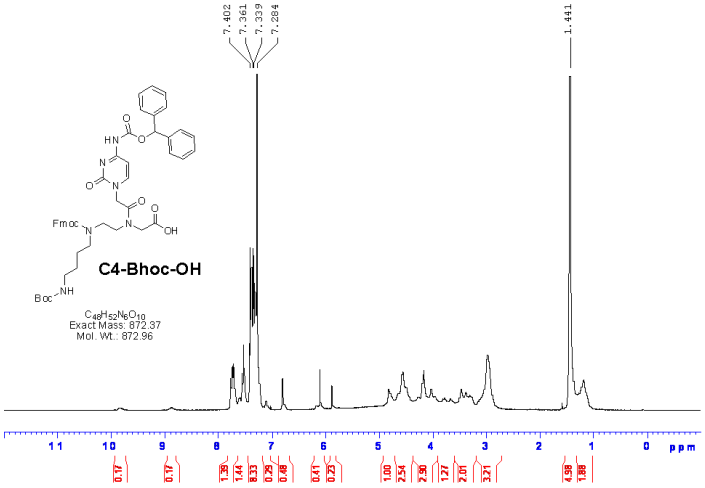

$^{1}$H NMR (DMSO) spectrum of **G4-iBu-OH**



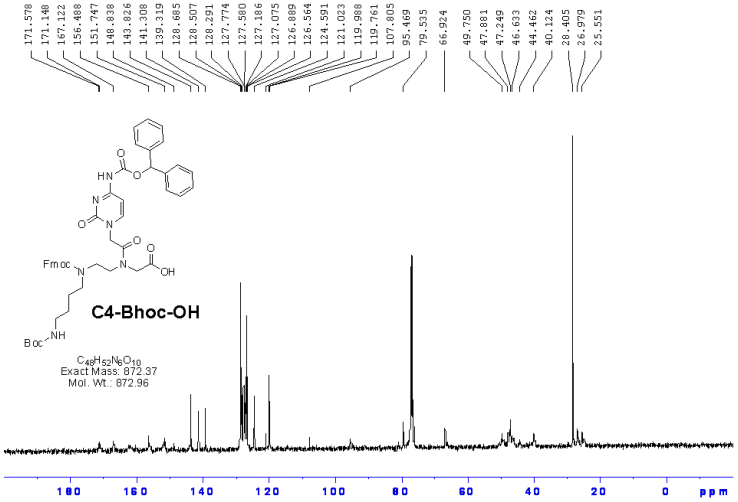

$^{13}$C NMR (DMSO) spectrum of **G4-iBu-OH**



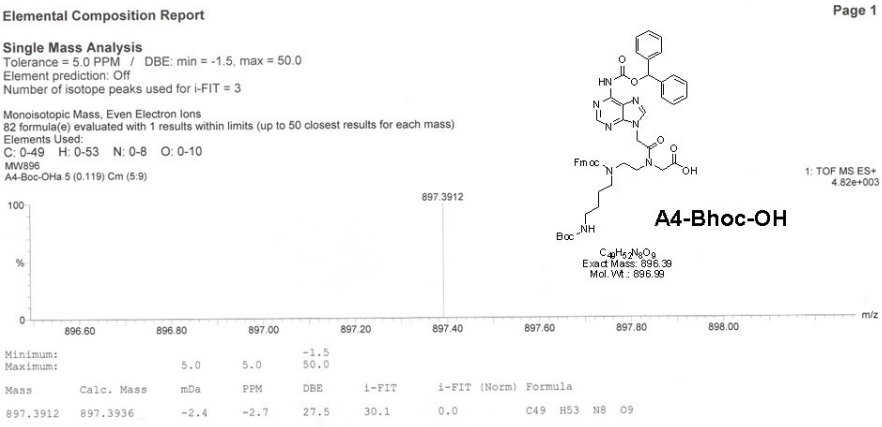

HRMS spectrum of **C4-Bhoc-OH**



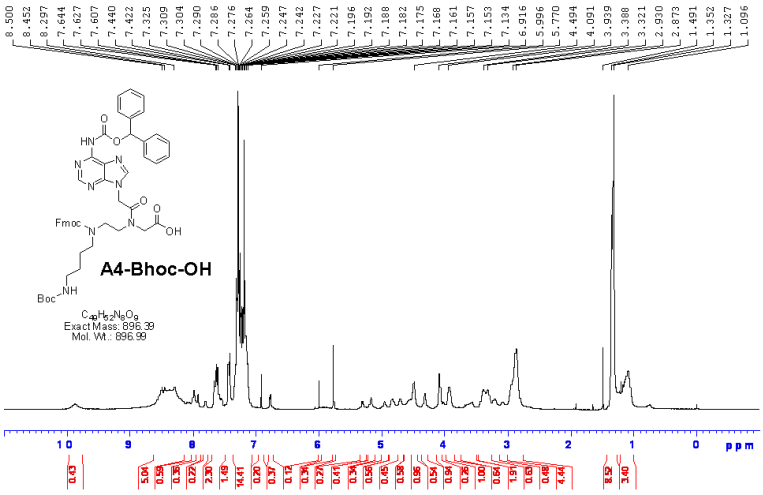

$^{1}$H NMR ($CDCl_3$) spectrum of **C4-Bhoc-OH**



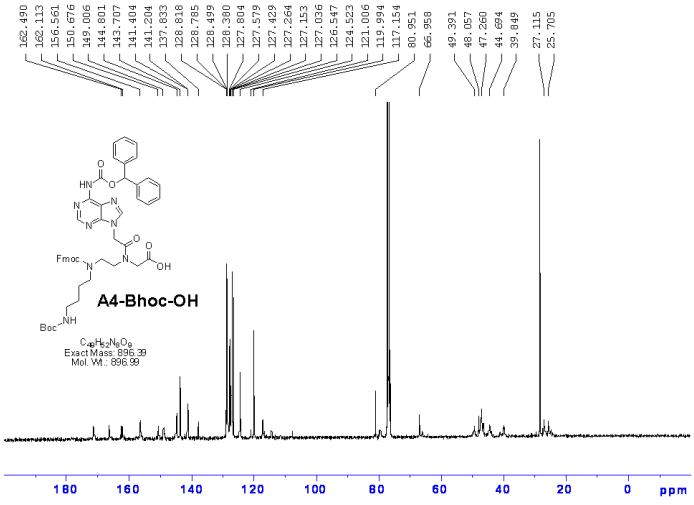

$^{13}$C NMR ($CDCl_3$) spectrum of **C4-Bhoc-OH**



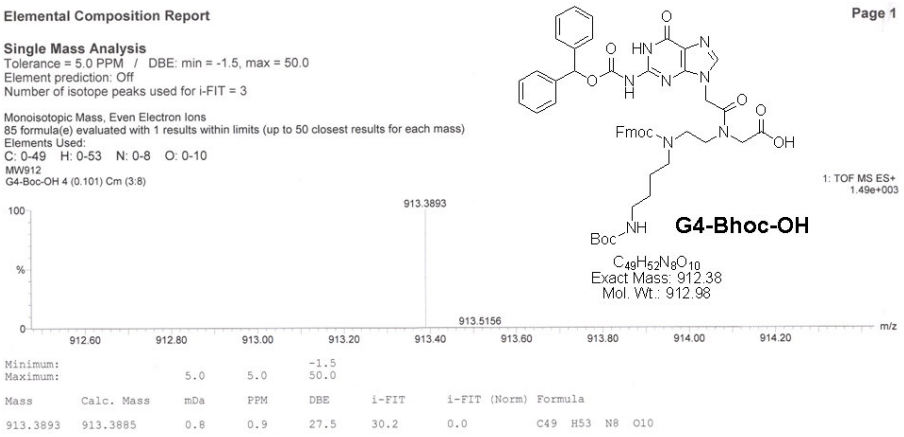

HRMS spectrum of **A4-Bhoc-OH**



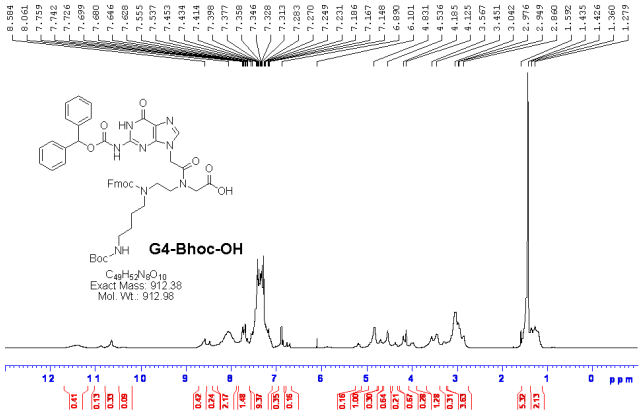

$^{1}$H NMR ($CDCl_3$) spectrum of **A4-Bhoc-OH**



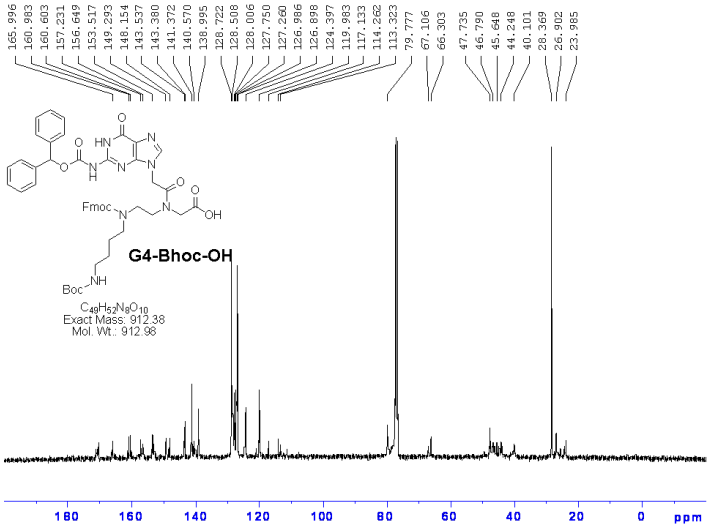

$^{13}$C NMR ($CDCl_3$) spectrum of **A4-Bhoc-OH**



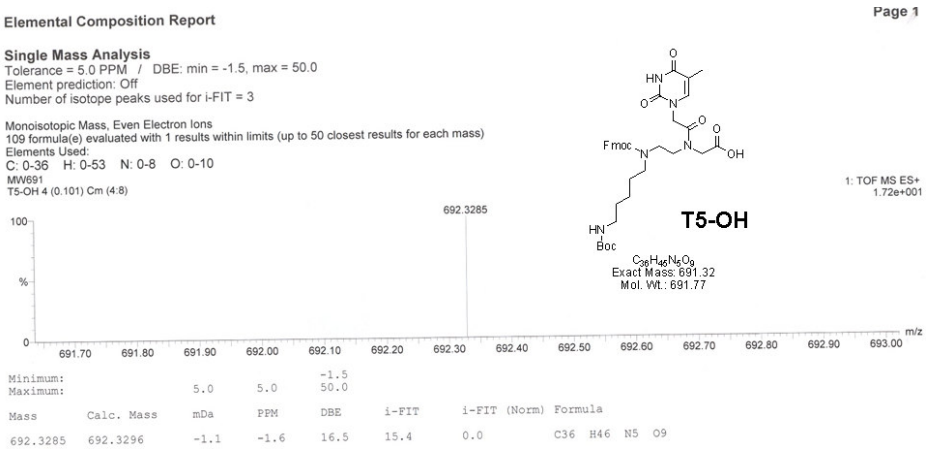

HRMS spectrum of **G4-Bhoc-OH**



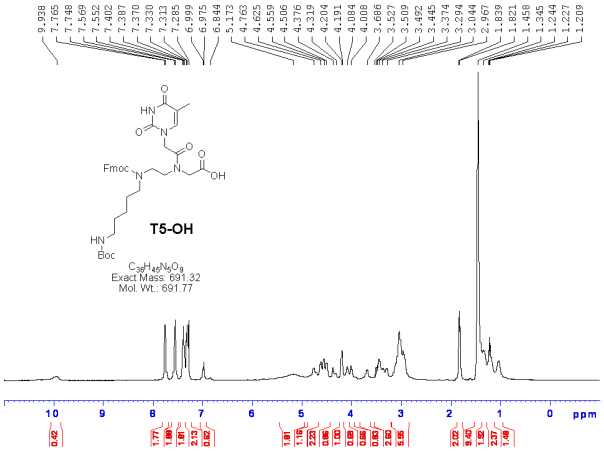

$^{1}$H NMR ($CDCl_3$) spectrum of **G4-Bhoc-OH**



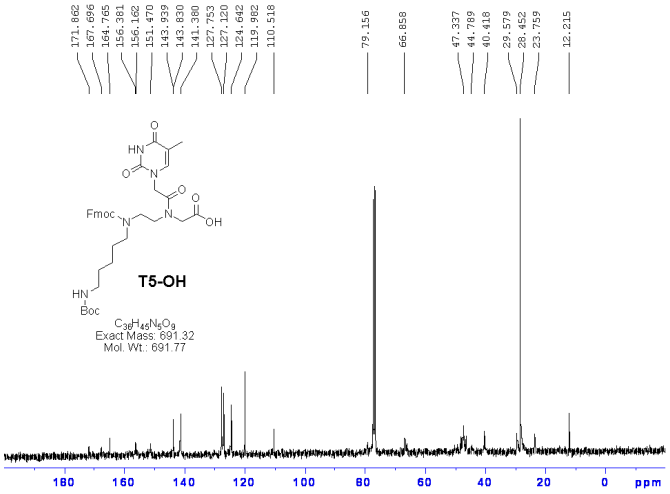

$^{13}$C NMR ($CDCl_3$) spectrum of **G4-Bhoc-OH**



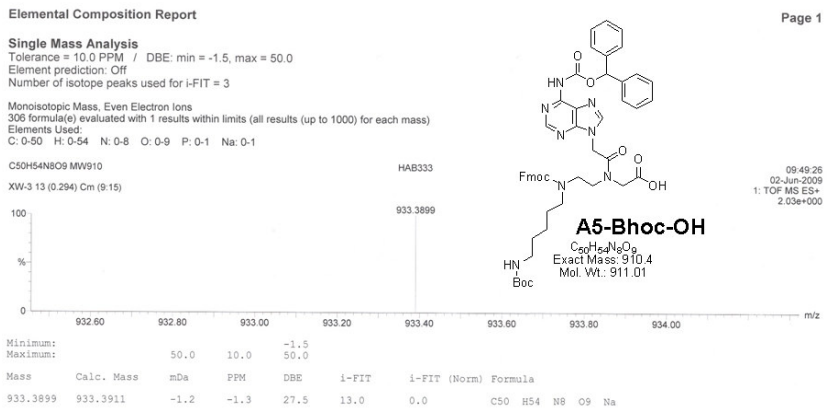

HRMS spectrum of **T5-OH**



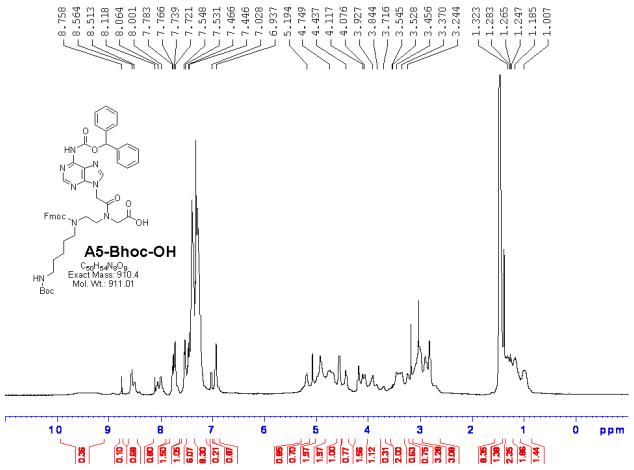

$^{1}$H NMR (DMSO) spectrum of **T5-OH**



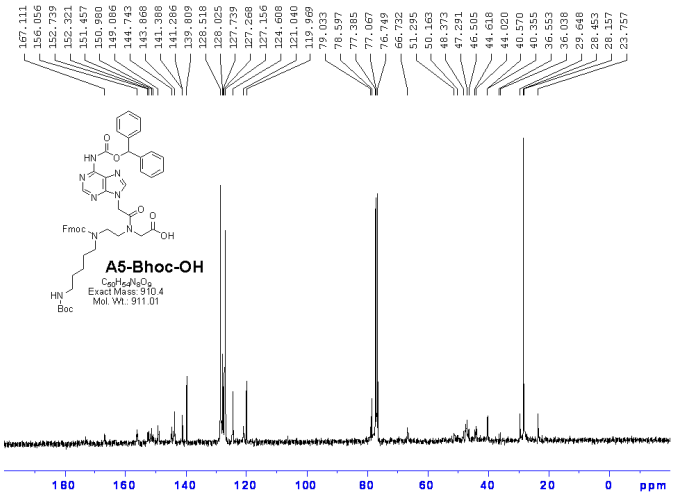

$^{13}$C NMR (DMSO) spectrum of **T5-OH**



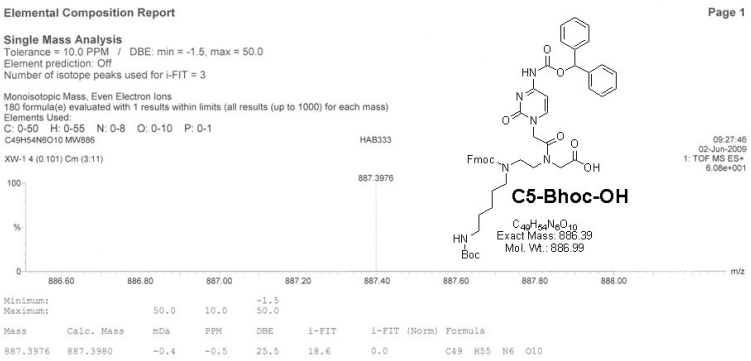

HRMS spectrum of **A5-Bhoc-OH**



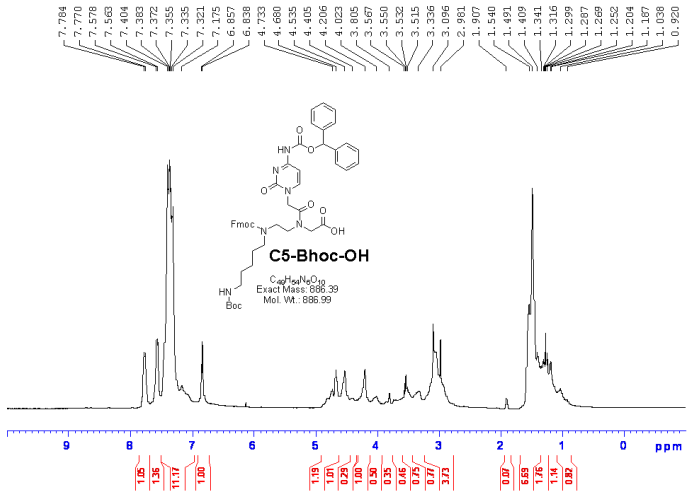

$^{1}$H NMR ($CDCl_3$) spectrum of **A5-Bhoc-OH**



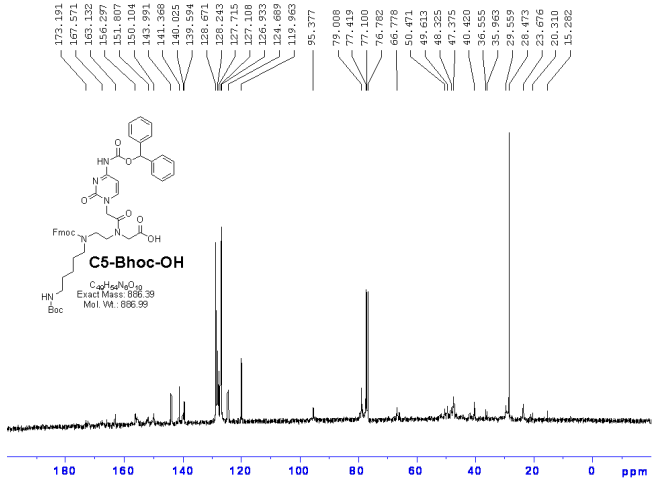

$^{13}$C NMR ($CDCl_3$) spectrum of **A5-Bhoc-OH**



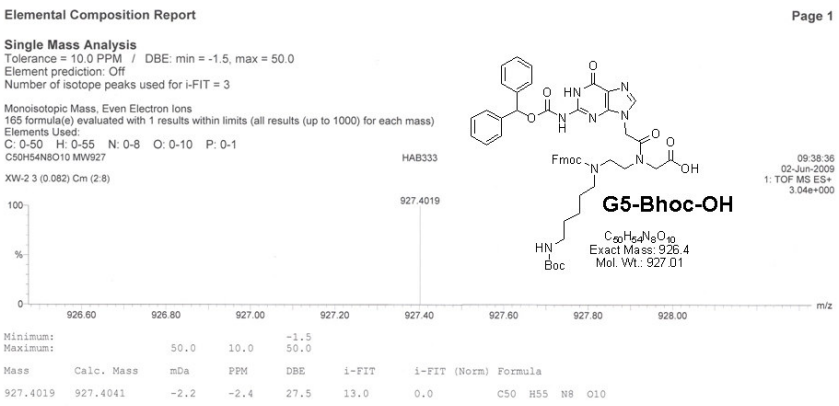

HRMS spectrum of **C5-Bhoc-OH**



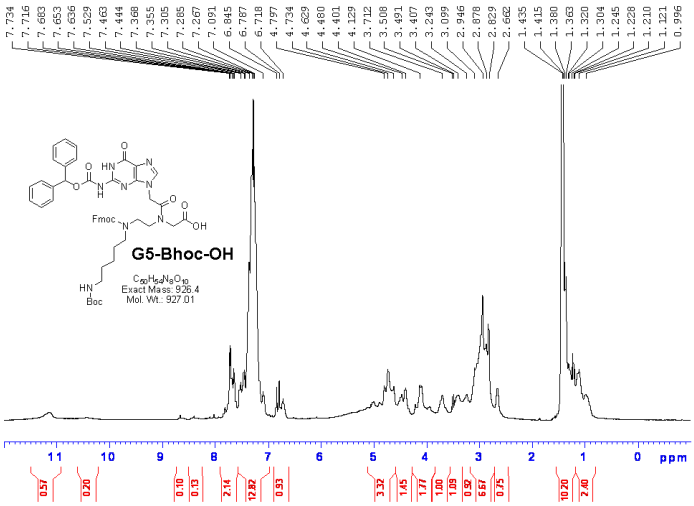

$^{1}$H NMR ($CDCl_3$) spectrum of **C5-Bhoc-OH**



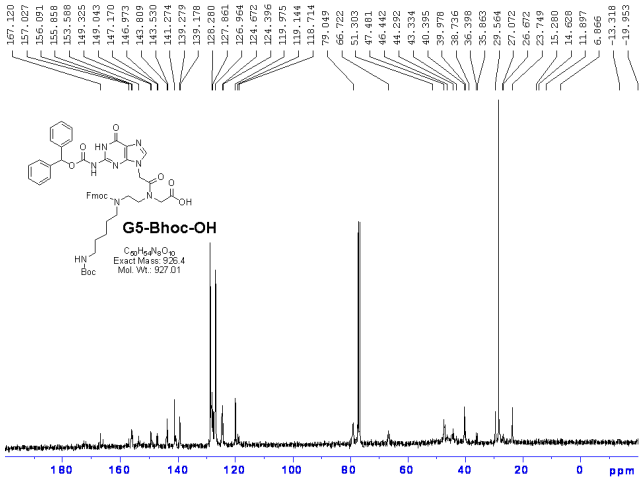

$^{13}$C NMR ($CDCl_3$) spectrum of **C5-Bhoc-OH**



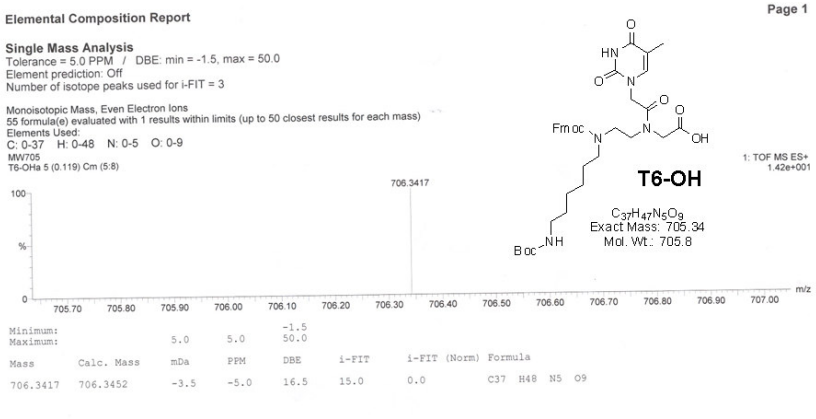

HRMS spectrum of **G5-Bhoc-OH**



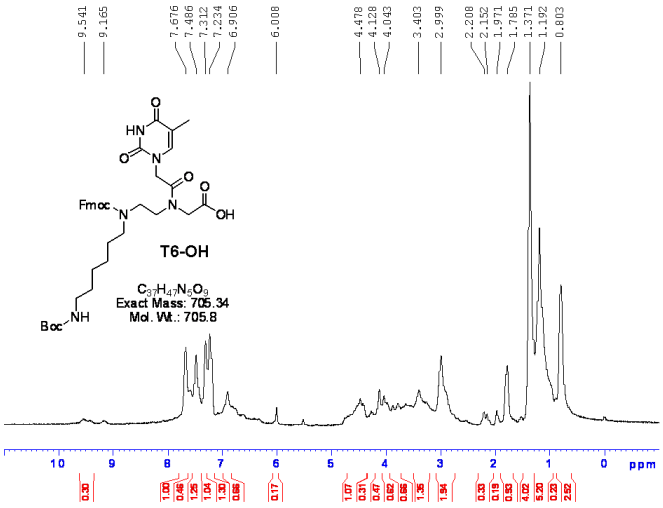

$^{1}$H NMR ($CDCl_3$) spectrum of **G5-Bhoc-OH**



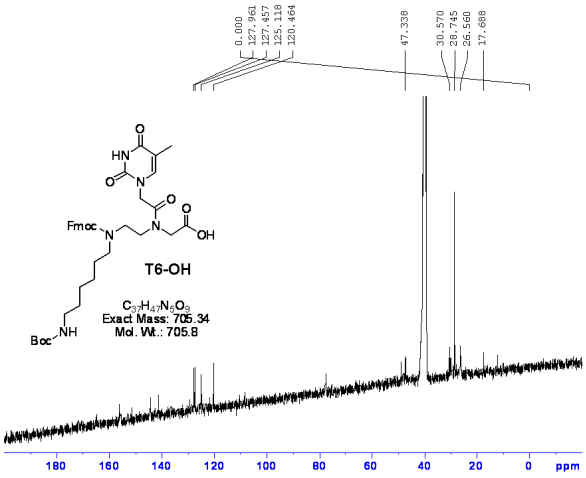

$^{13}$C NMR ($CDCl_3$) spectrum of **G5-Bhoc-OH**



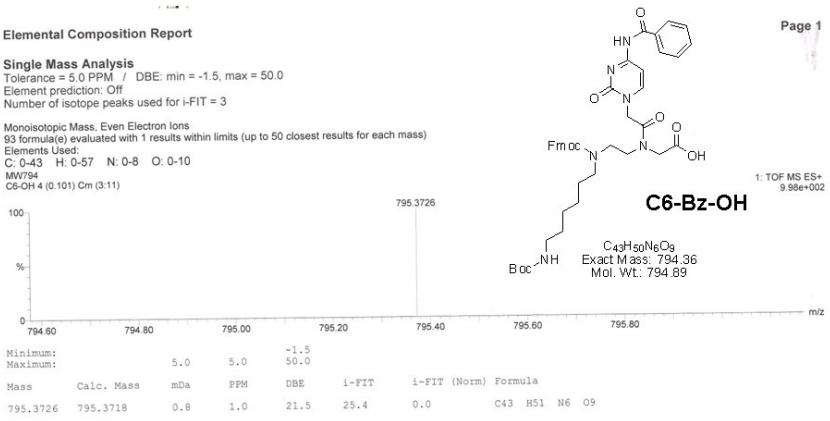

HRMS spectrum of **T6-OH**



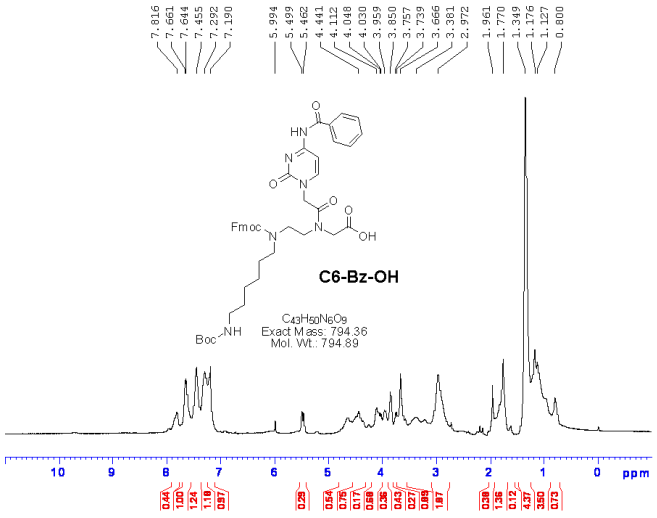

$^{1}$H NMR (DMSO) spectrum of **T6-OH**



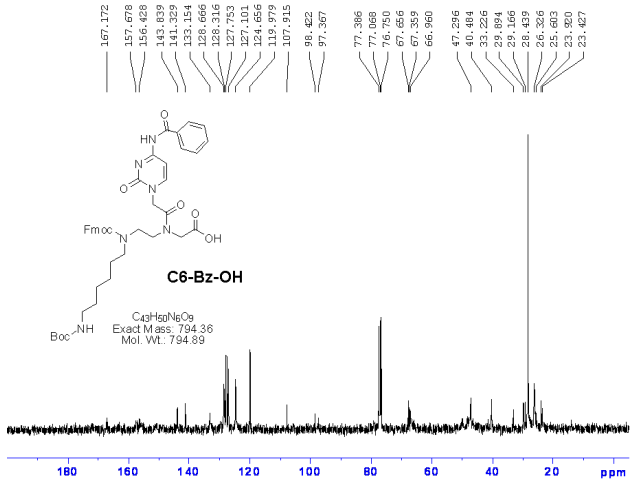

$^{13}$C NMR (DMSO) spectrum of **T6-OH**



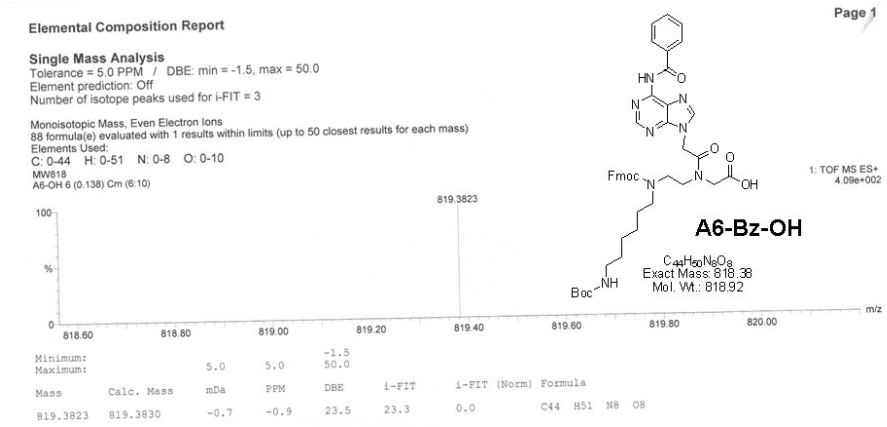

HRMS spectrum of **C6-Bz-OH**



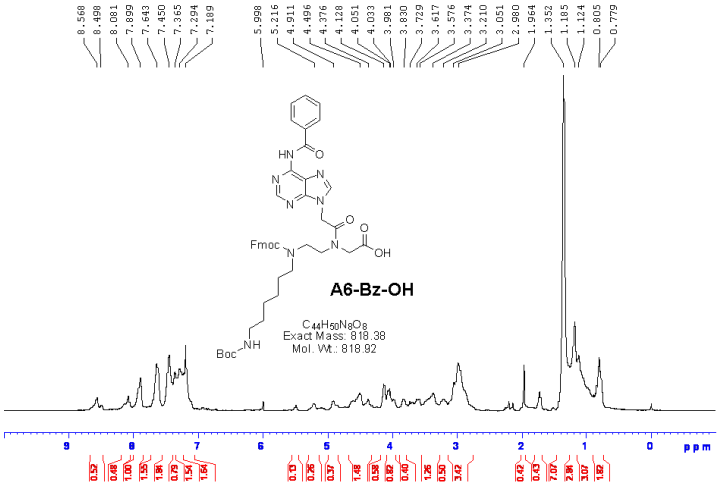

$^{1}$H NMR (DMSO) spectrum of **C6-Bz-OH**



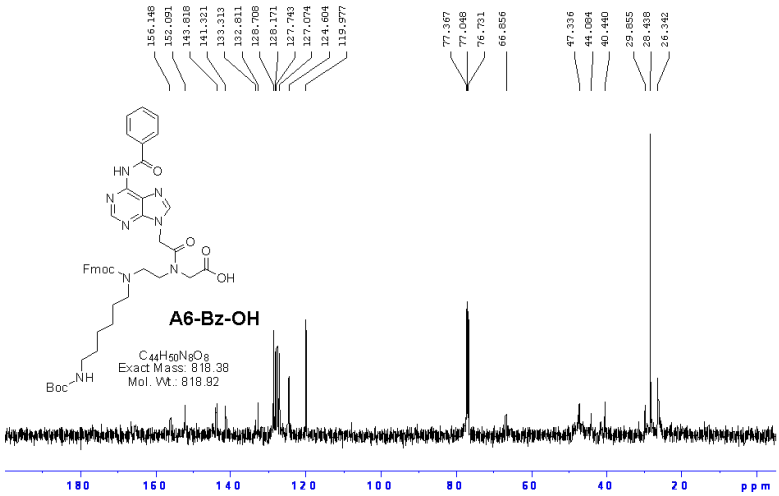

$^{13}$C NMR (DMSO) spectrum of **C6-Bz-OH**



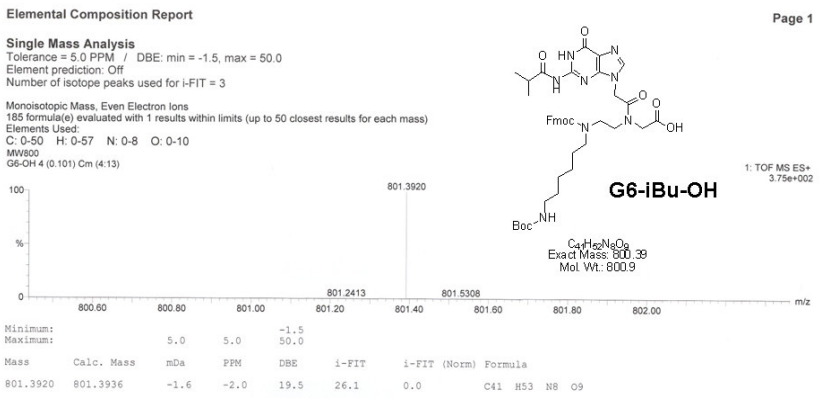

HRMS spectrum of **A6-Bz-OH**



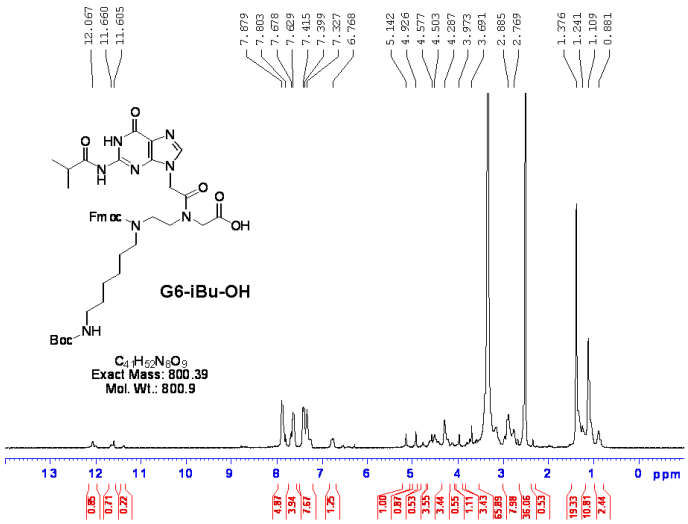

$^{1}$H NMR (DMSO) spectrum of **A6-Bz-OH**



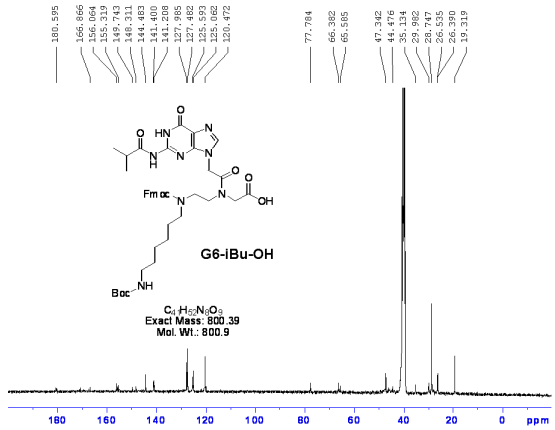

$^{13}$C NMR (DMSO) spectrum of **A6-Bz-OH**



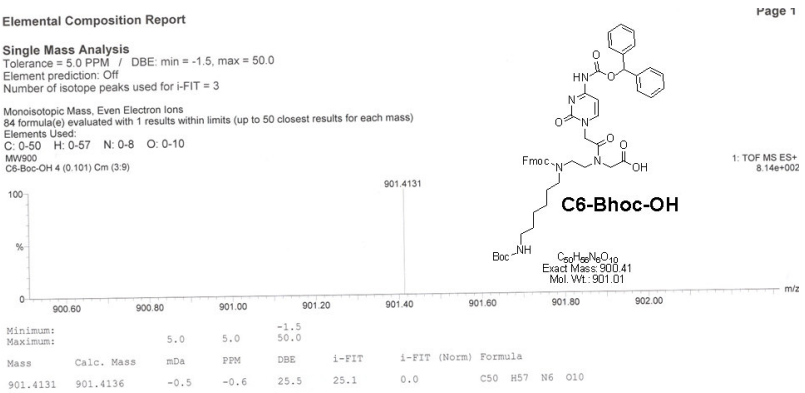

HRMS spectrum of **G6-iBu-OH**



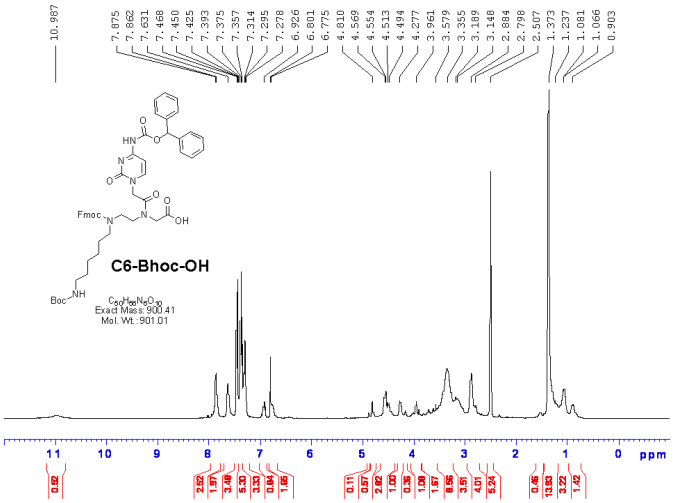

$^{1}$H NMR (DMSO) spectrum of **G6-iBu-OH**



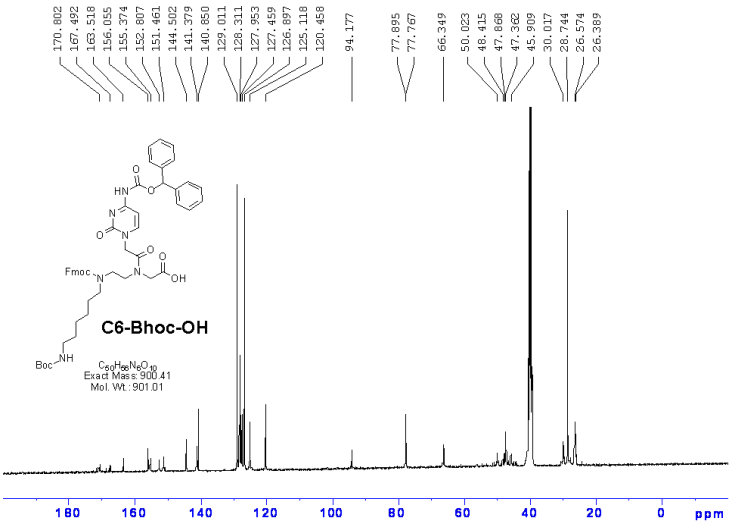

$^{13}$C NMR (DMSO) spectrum of **G6-iBu-OH**



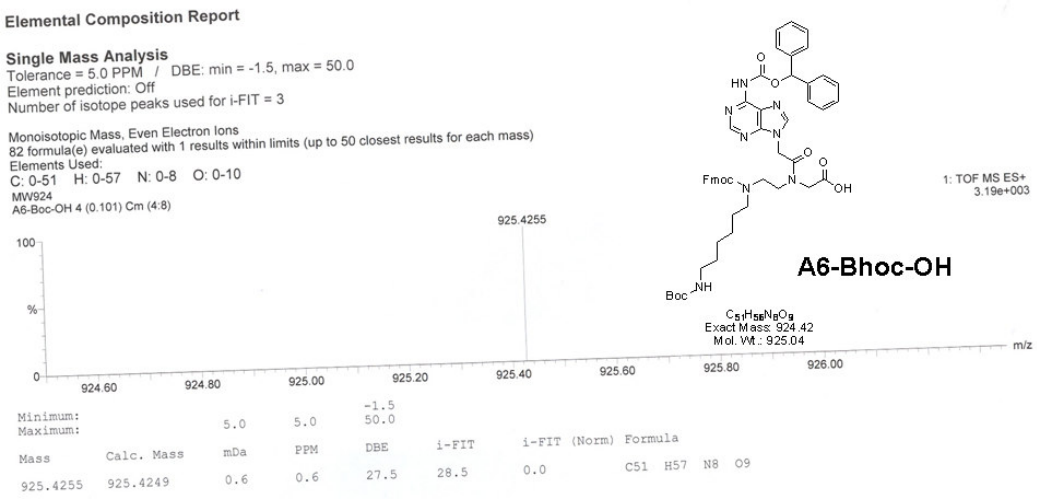

HRMS spectrum of **C6-Bhoc-OH**



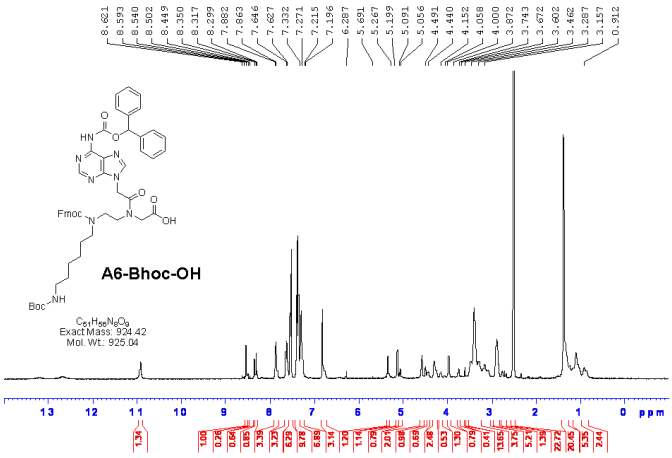

$^{1}$H NMR ($CDCl_3$) spectrum of **C6-Bhoc-OH**



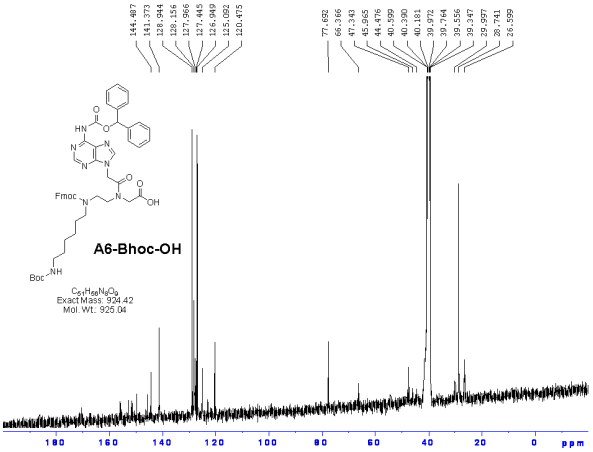

$^{1}$H NMR ($CDCl_3$) spectrum of **C6-Bhoc-OH**



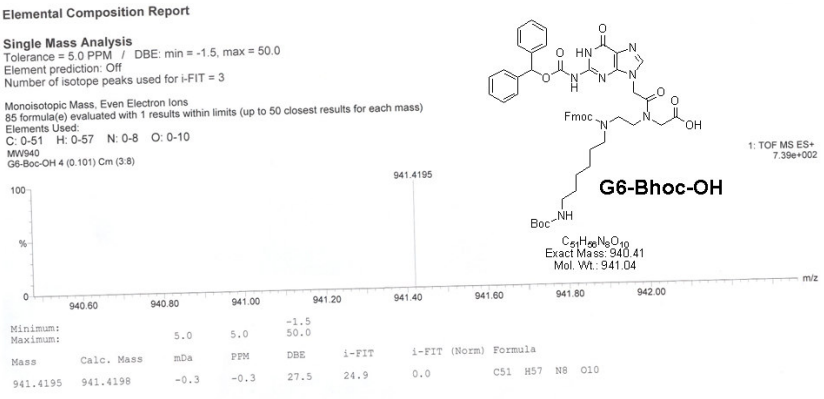

HRMS spectrum of **A6-Bhoc-OH**



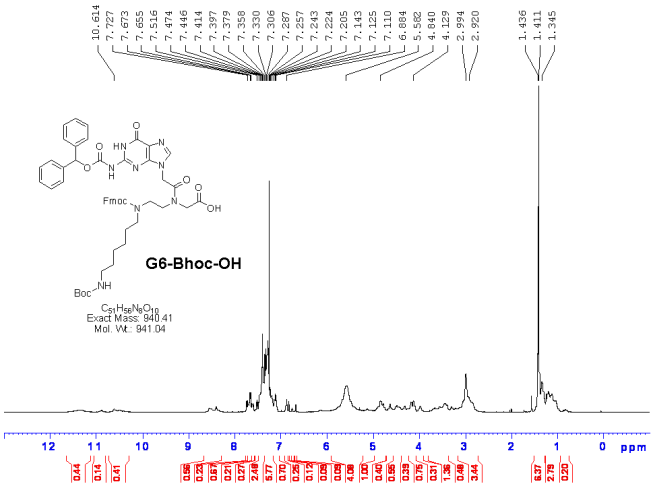

$^{1}$H NMR ($CDCl_3$) spectrum of **A6-Bhoc-OH**



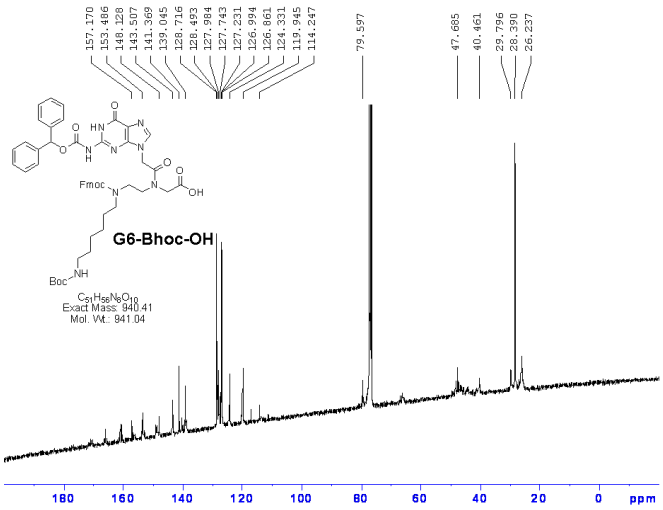

$^{13}$C NMR ($CDCl_3$) spectrum of **A6-Bhoc-OH**



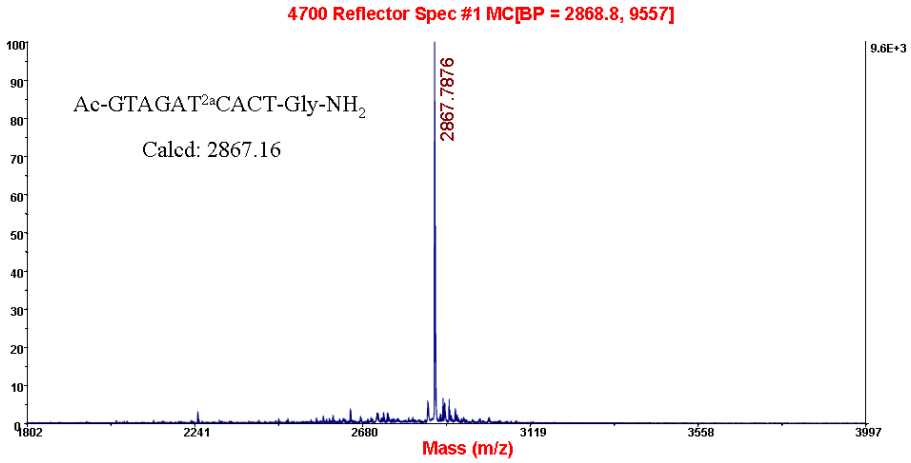

HRMS spectrum of **G6-Bhoc-OH**



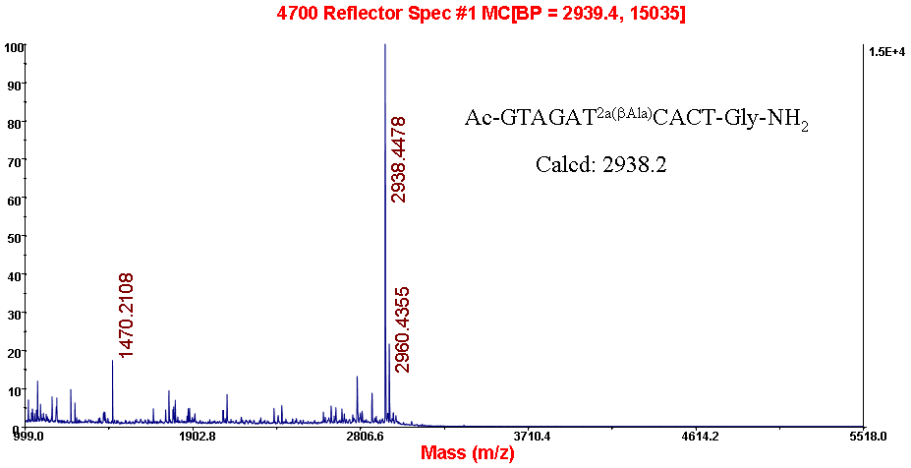

$^{1}$H NMR ($CDCl_3$) spectrum of **G6-Bhoc-OH**



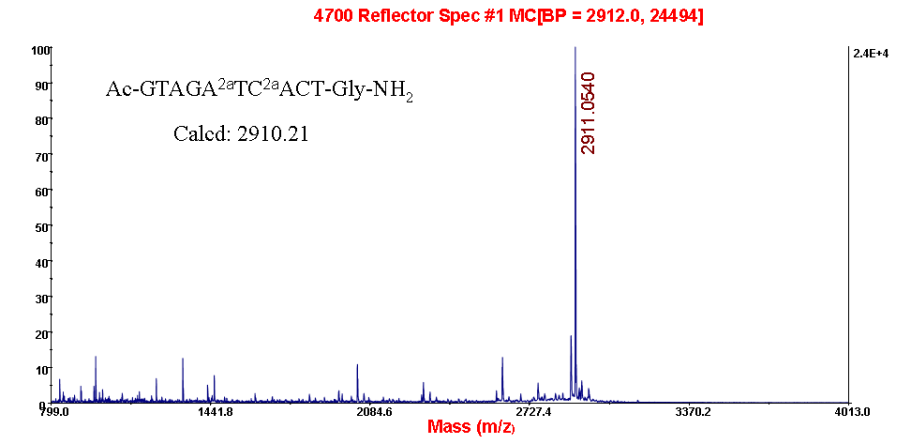

$^{13}$C NMR ($CDCl_3$) spectrum of **G6-Bhoc-OH**



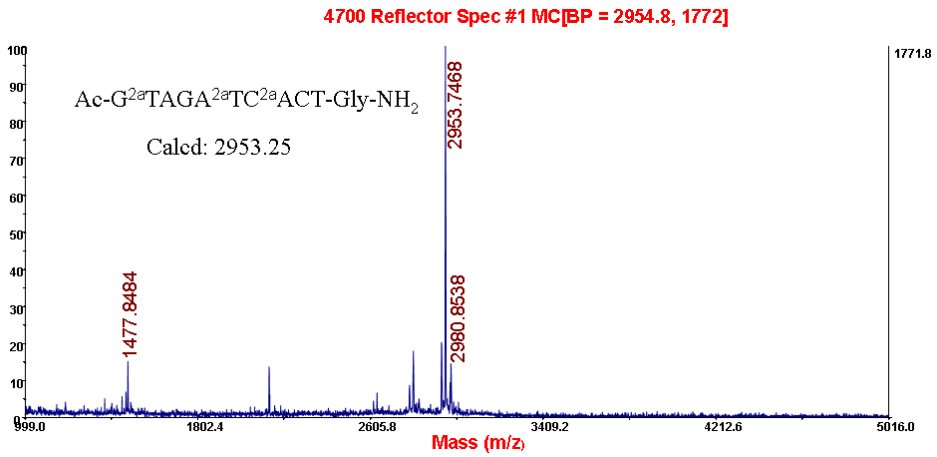

MALDI-TOF MASS spectrum of **AP-PNA 2-1**



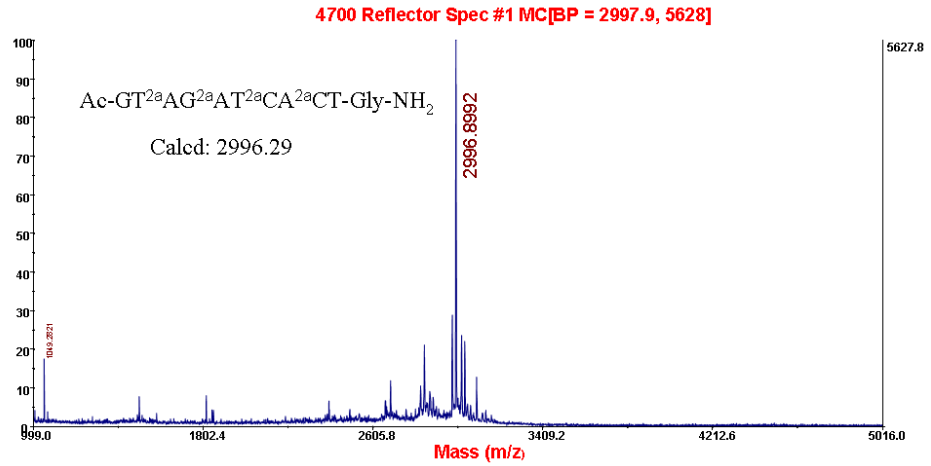

MALDI-TOF MASS spectrum of **AP-PNA 2-1(βAla)**



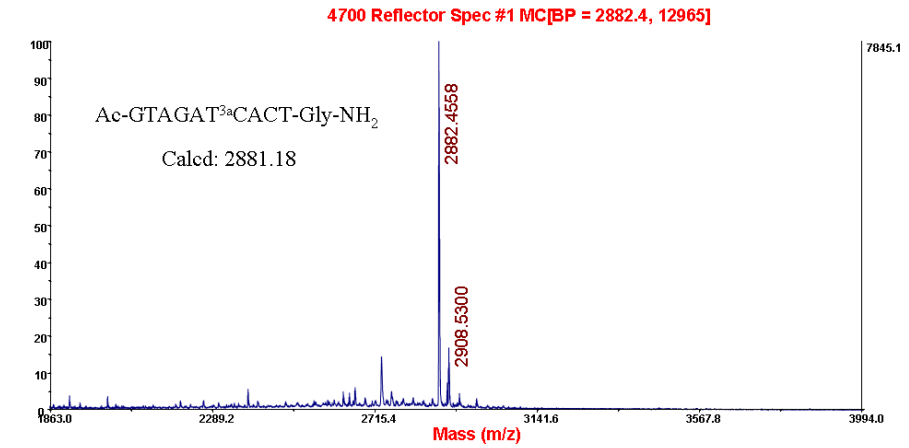

MALDI-TOF MASS spectrum of **AP-PNA 2-2**



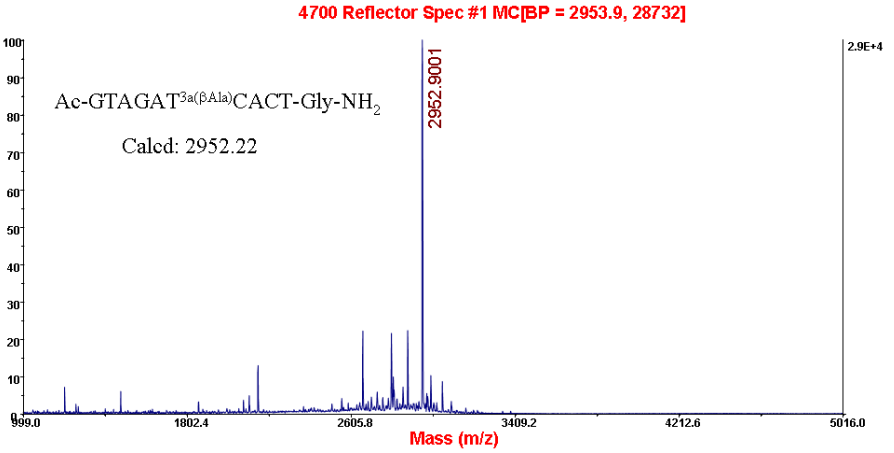

MALDI-TOF MASS spectrum of **AP-PNA 2-3**



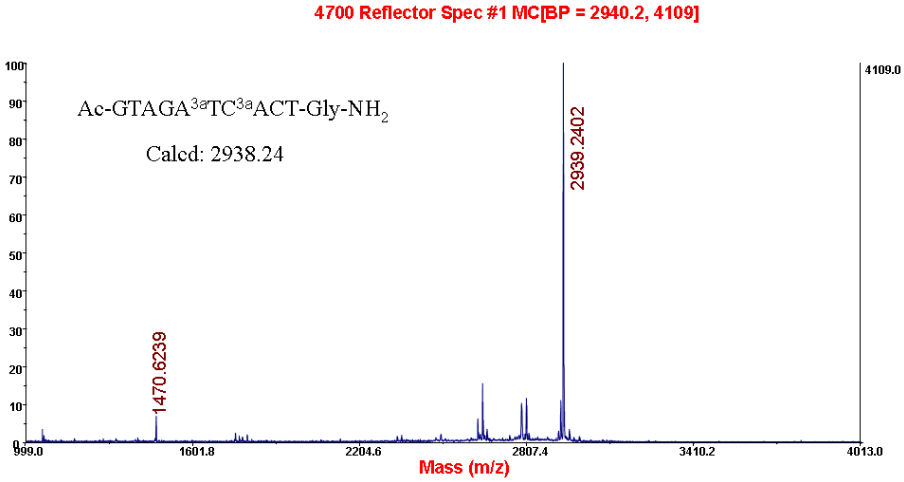

MALDI-TOF MASS spectrum of **AP-PNA 2-4**



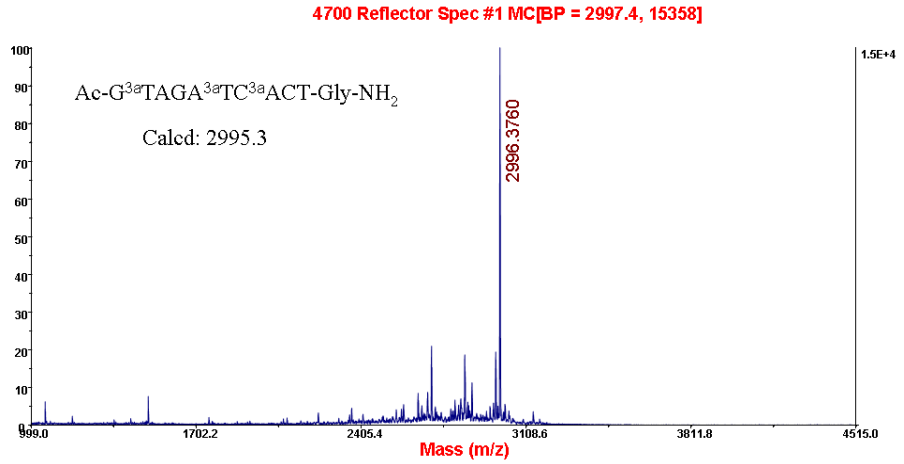

MALDI-TOF MASS spectrum of **AP-PNA 3-1**



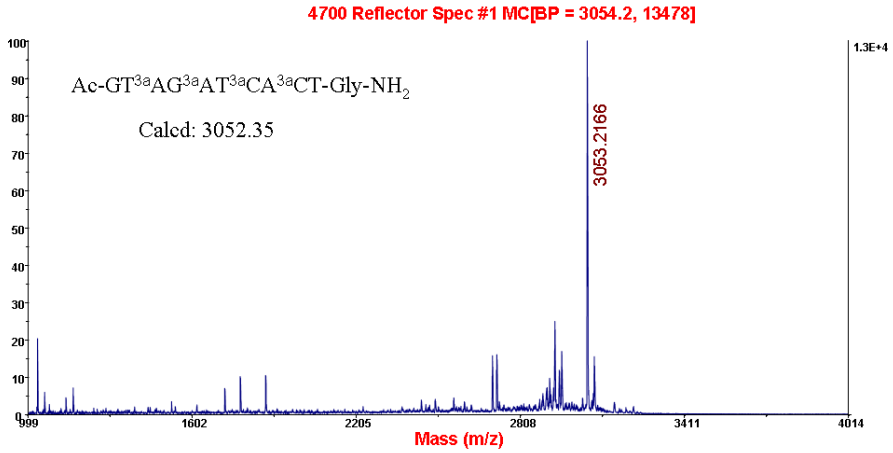

MALDI-TOF MASS spectrum of **AP-PNA 3-1(βAla)**



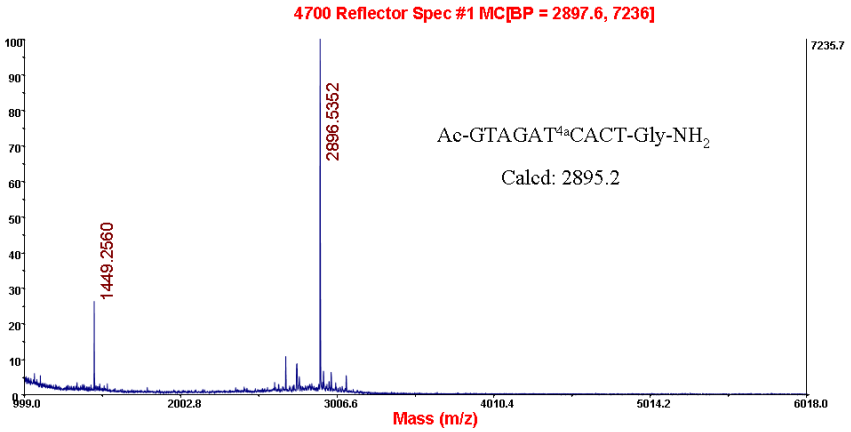

MALDI-TOF MASS spectrum of **AP-PNA 3-2**



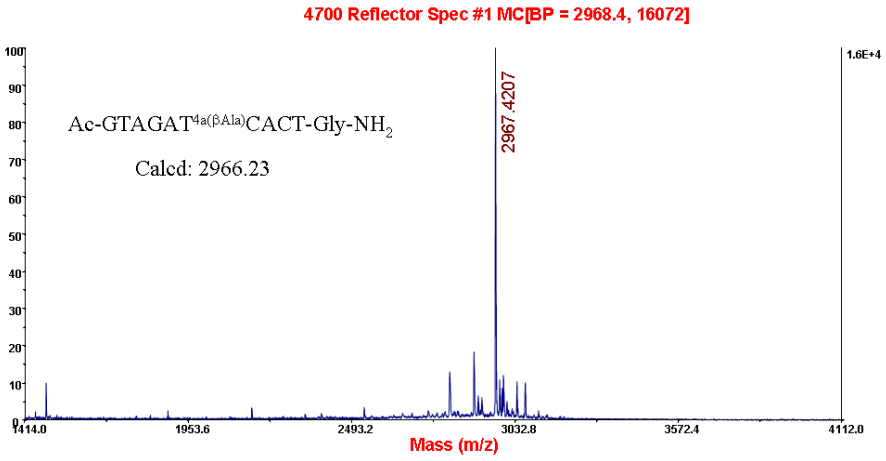

MALDI-TOF MASS spectrum of **AP-PNA 3-3**



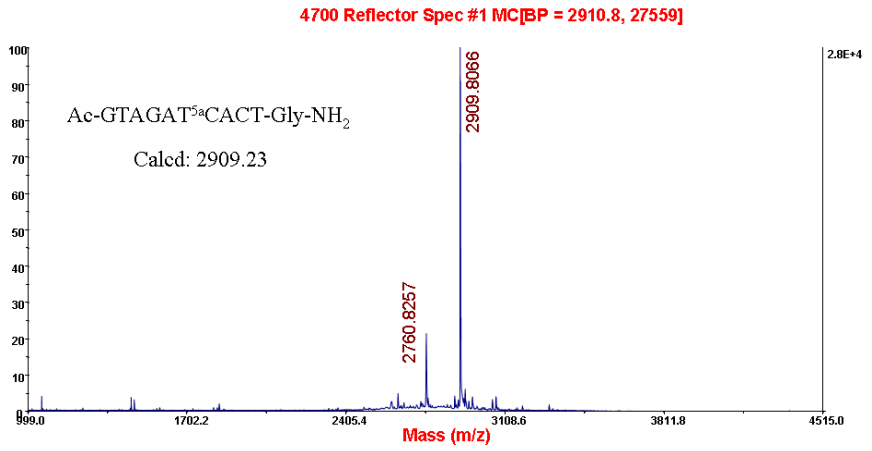

MALDI-TOF MASS spectrum of **AP-PNA 3-4**



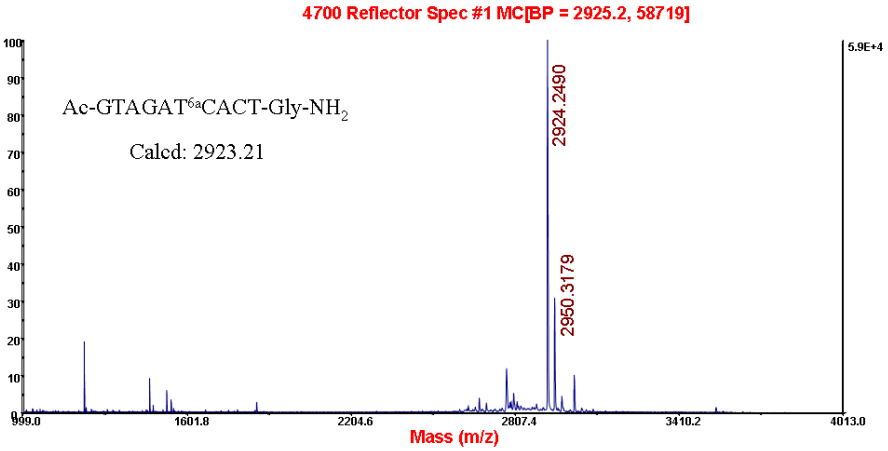

MALDI-TOF MASS spectrum of **AP-PNA 4-1**



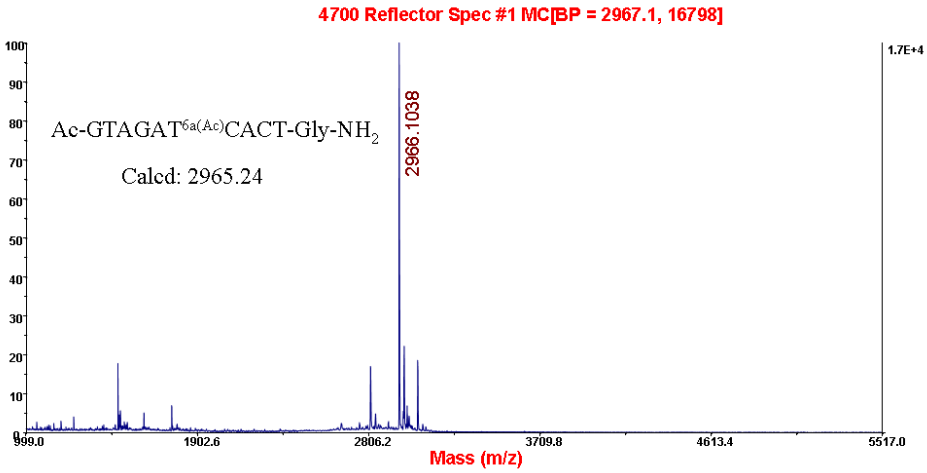

MALDI-TOF MASS spectrum of **AP-PNA 4-1(βAla)**



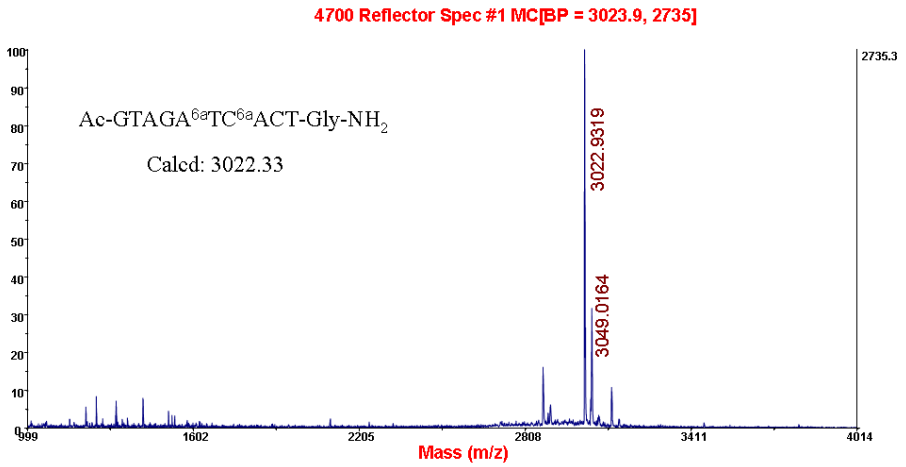

MALDI-TOF MASS spectrum of **AP-PNA 5-1**



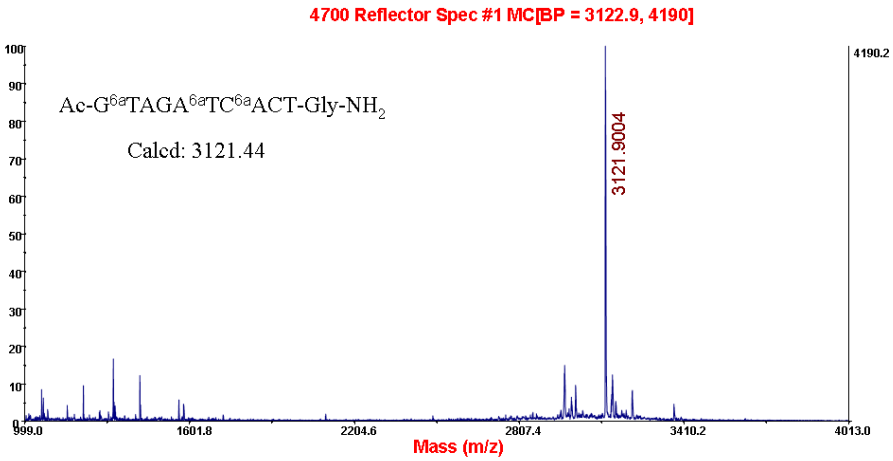

MALDI-TOF MASS spectrum of **AP-PNA 6-1**



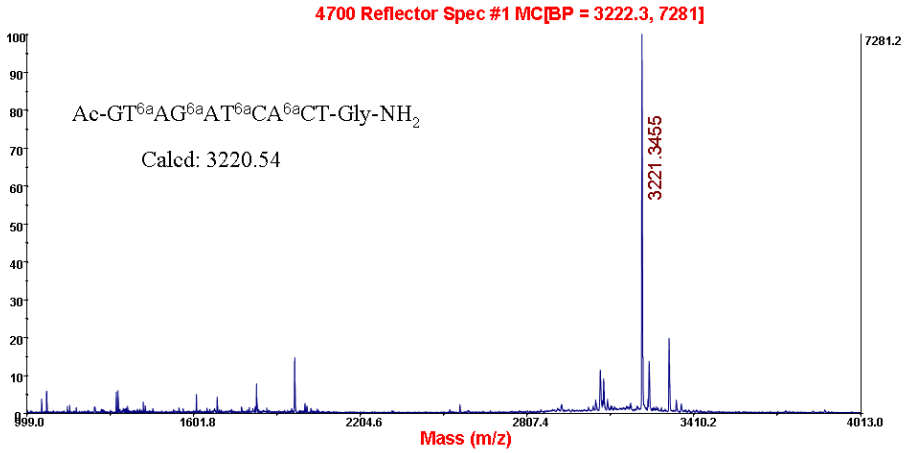

MALDI-TOF MASS spectrum of **AP-PNA 6-1(Ac)**



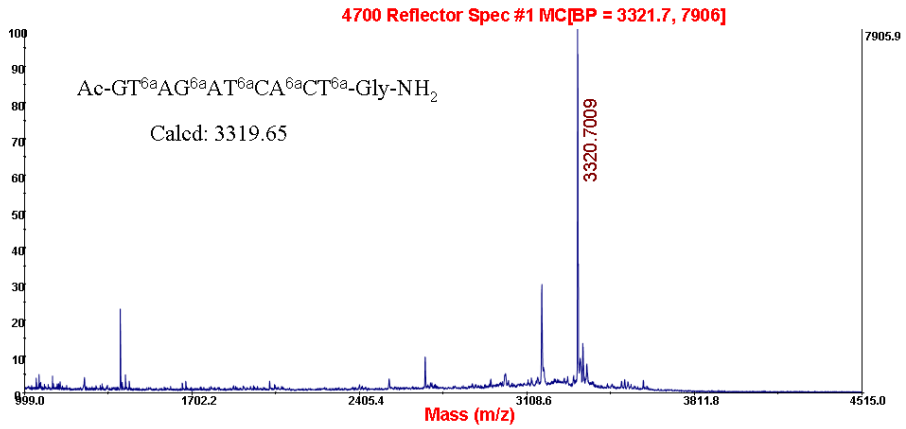

MALDI-TOF MASS spectrum of **AP-PNA 6-2**



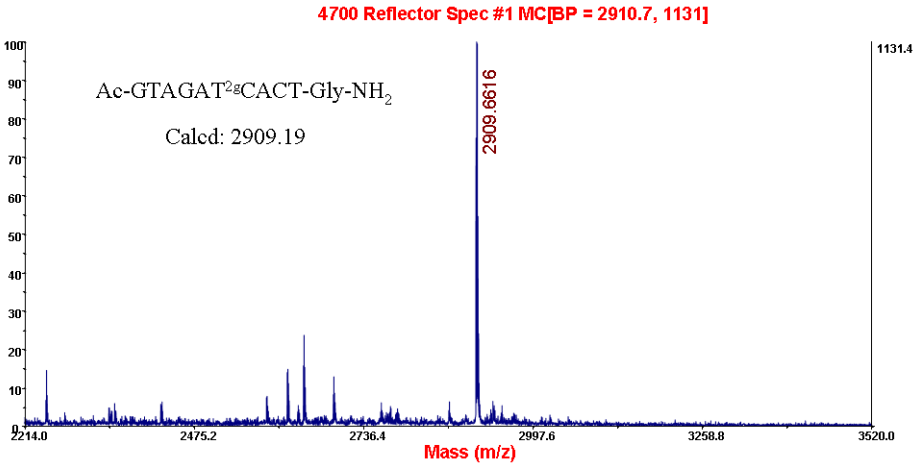

MALDI-TOF MASS spectrum of **AP-PNA 6-3**



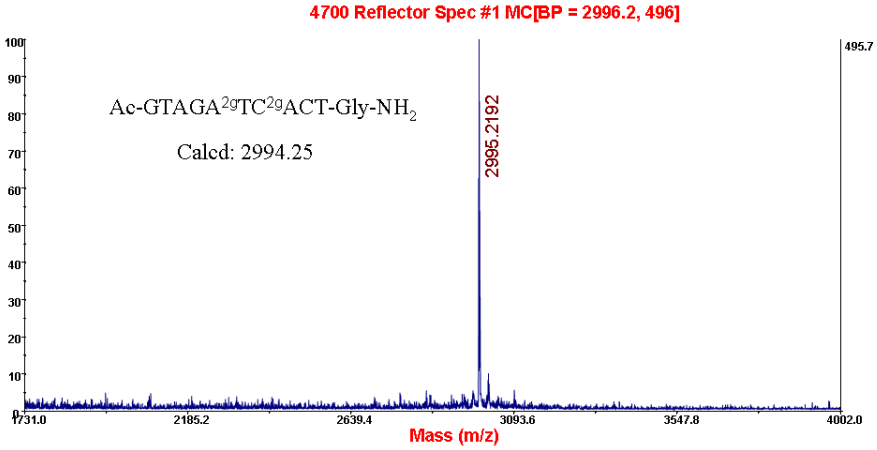

MALDI-TOF MASS spectrum of **AP-PNA 6-4**



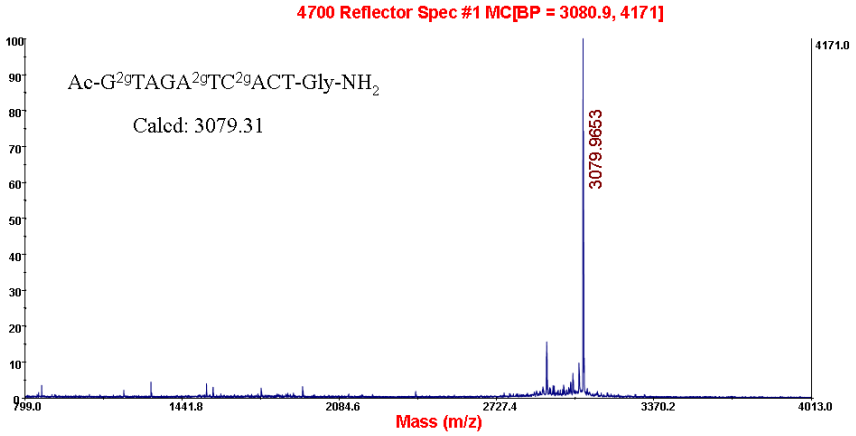

MALDI-TOF MASS spectrum of **AP-PNA 6-5**



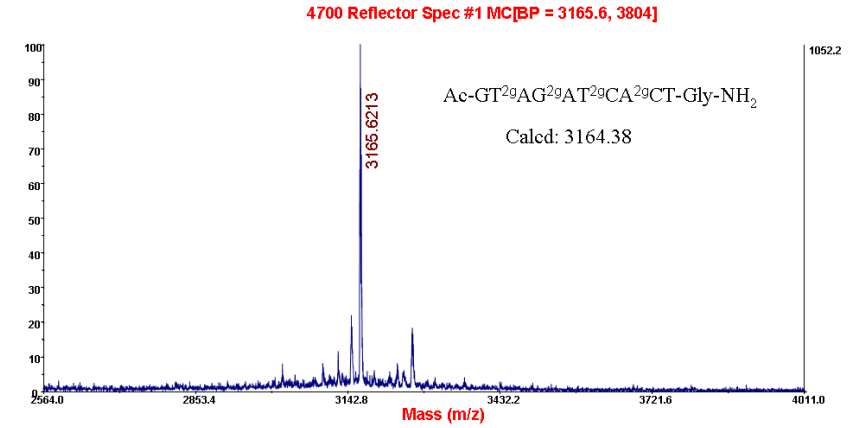

MALDI-TOF MASS spectrum of **GP-PNA 2-1**



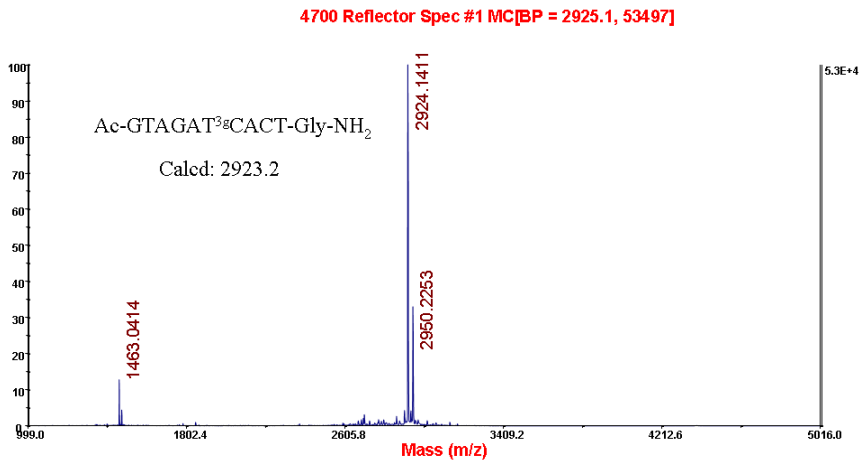

MALDI-TOF MASS spectrum of **GP-PNA 2-2**



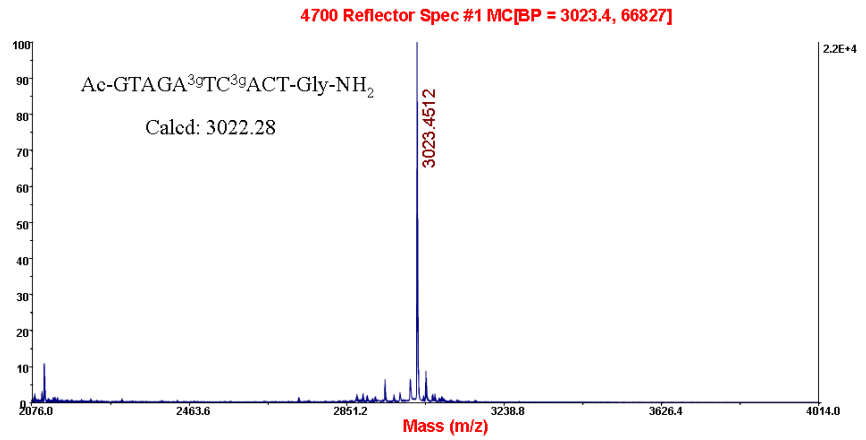

MALDI-TOF MASS spectrum of **GP-PNA 2-3**



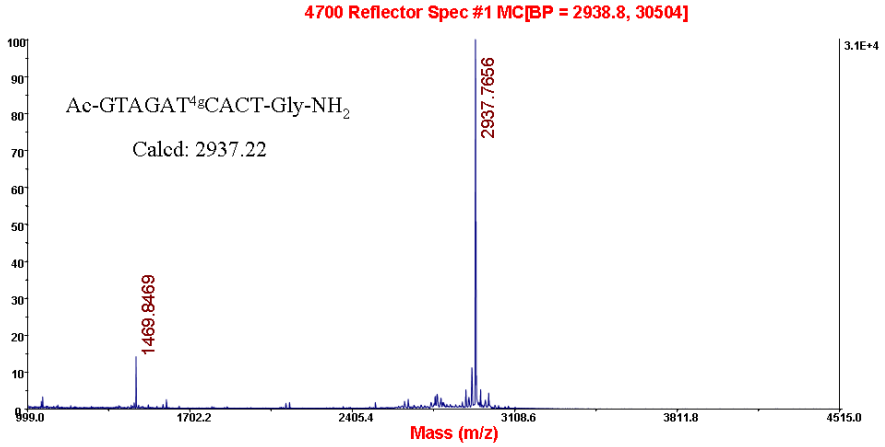

MALDI-TOF MASS spectrum of **GP-PNA 2-4**



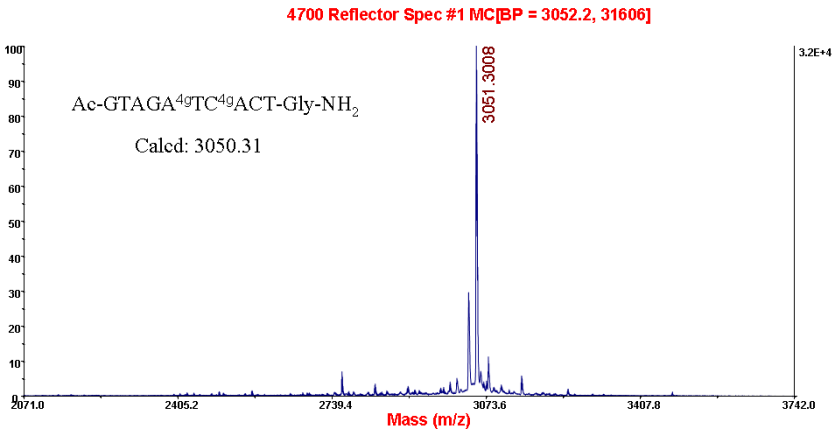

MALDI-TOF MASS spectrum of **GP-PNA 3-1**



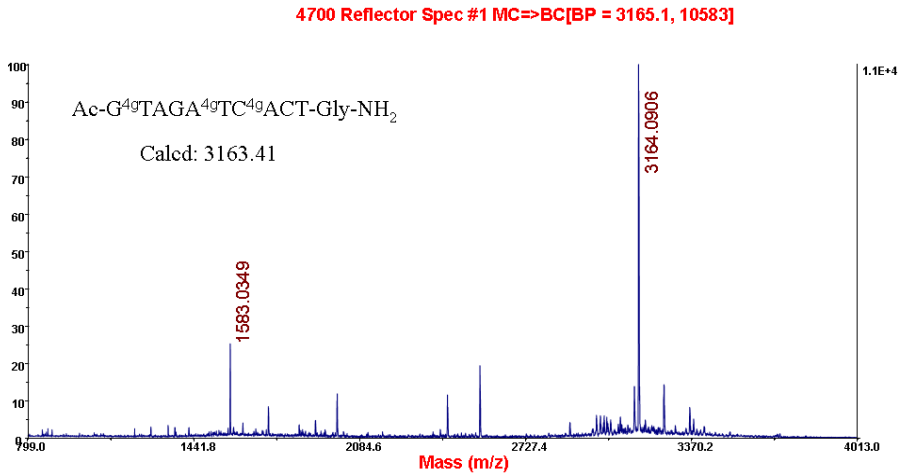

MALDI-TOF MASS spectrum of **GP-PNA 3-2**



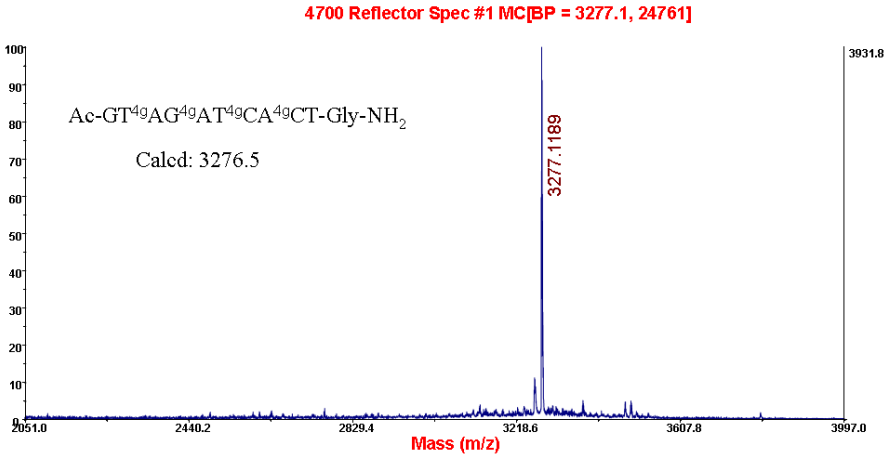

MALDI-TOF MASS spectrum of **GP-PNA 4-1**



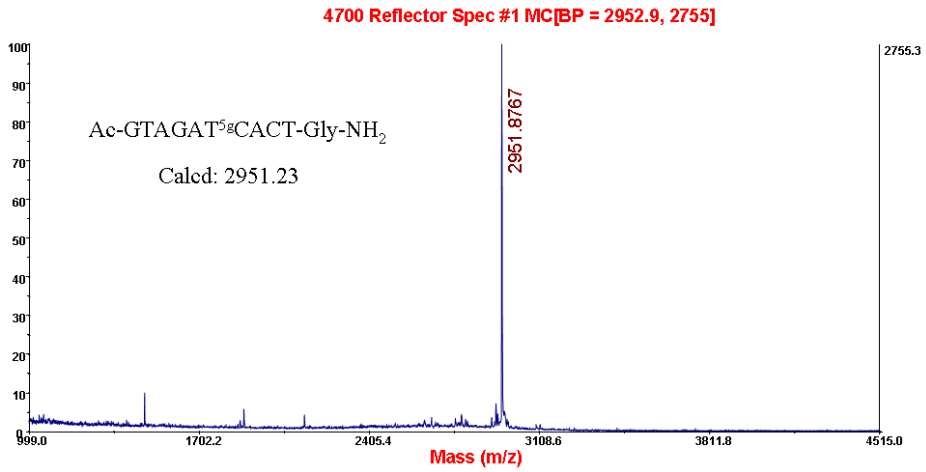

MALDI-TOF MASS spectrum of **GP-PNA 4-2**



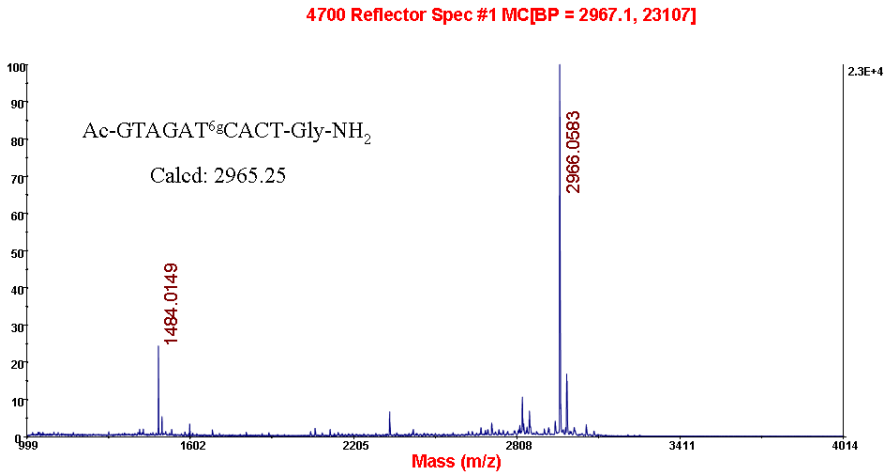

MALDI-TOF MASS spectrum of **GP-PNA 4-3**



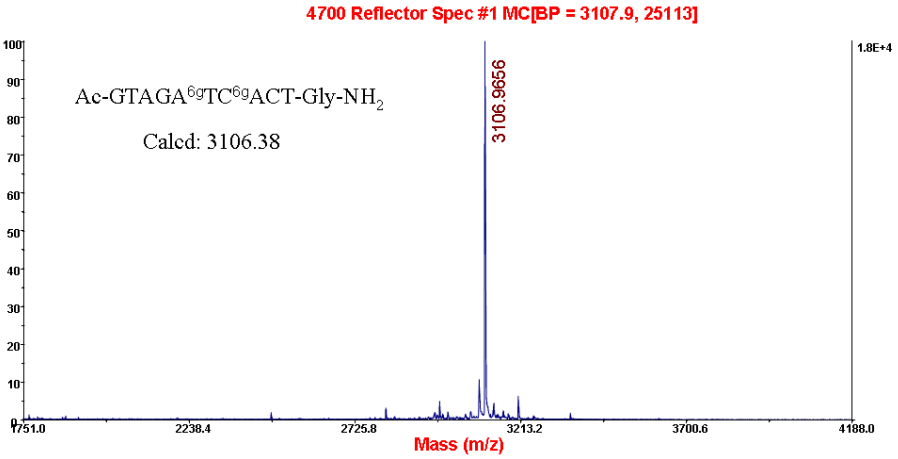

MALDI-TOF MASS spectrum of **GP-PNA 4-4**



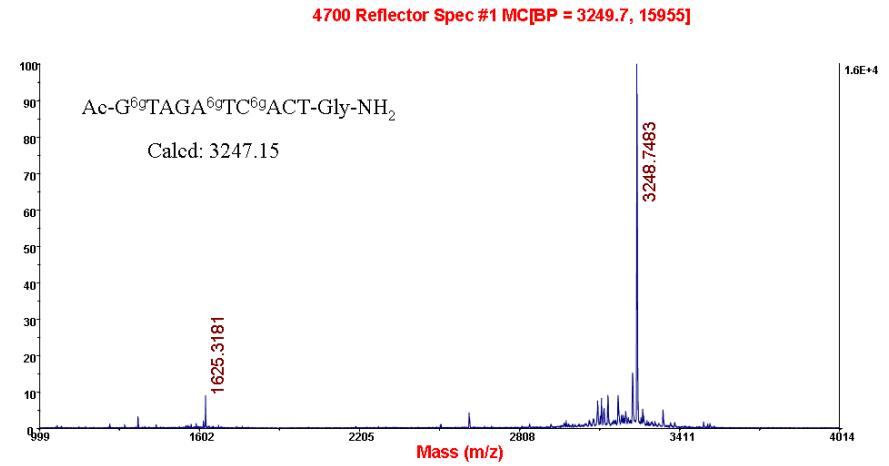

MALDI-TOF MASS spectrum of **GP-PNA 5-1**



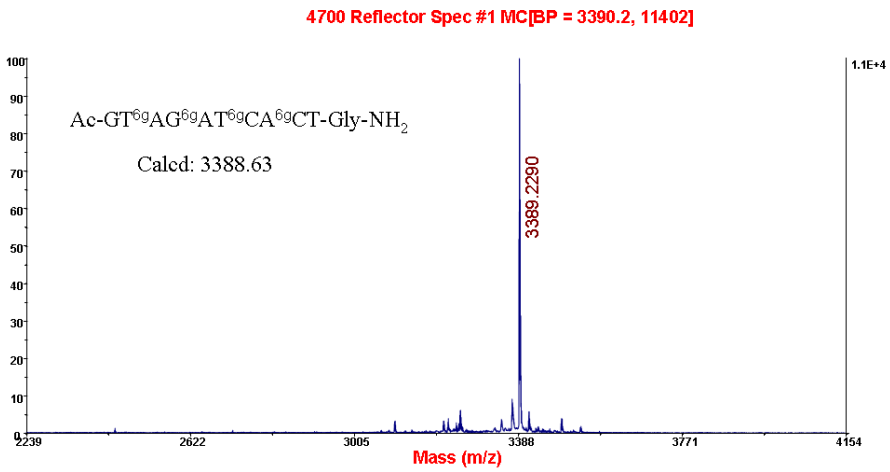

MALDI-TOF MASS spectrum of **GP-PNA 6-1**



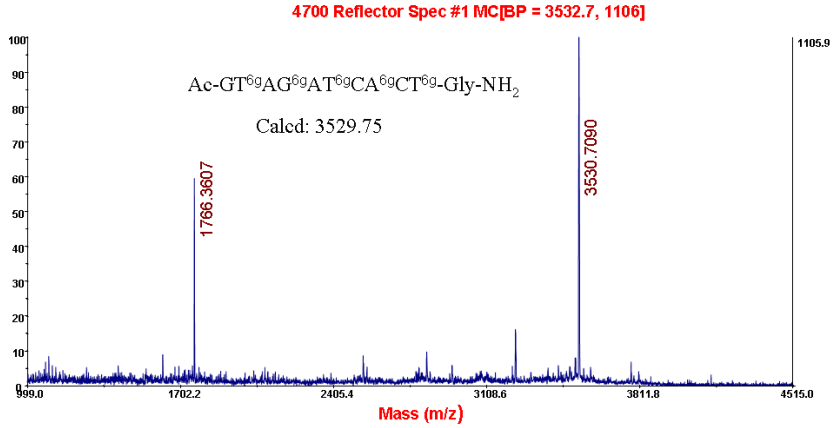

MALDI-TOF MASS spectrum of **GP-PNA 6-2**



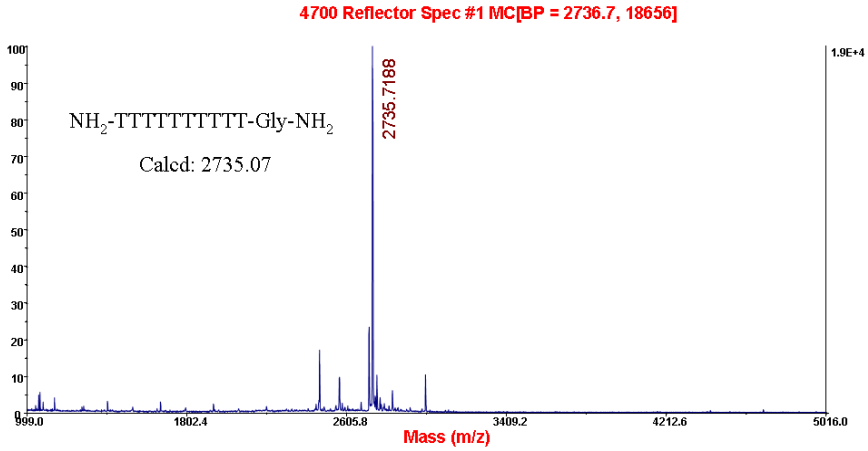

MALDI-TOF MASS spectrum of **GP-PNA 6-3**



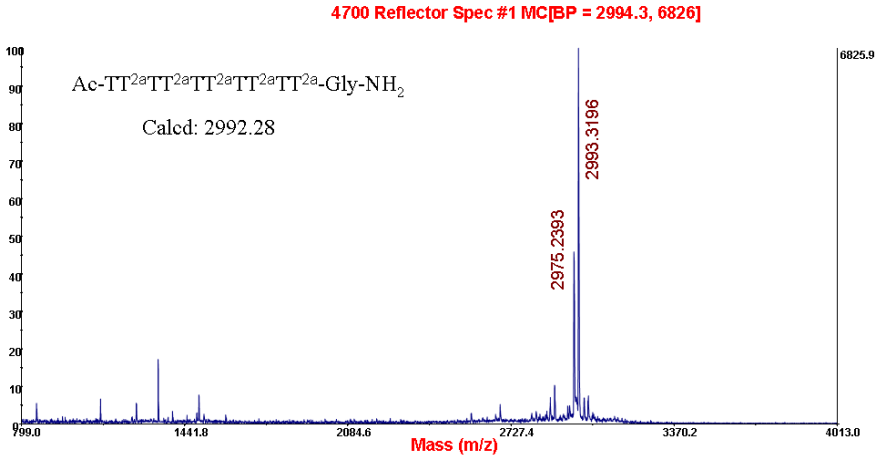

MALDI-TOF MASS spectrum of **GP-PNA 6-4**



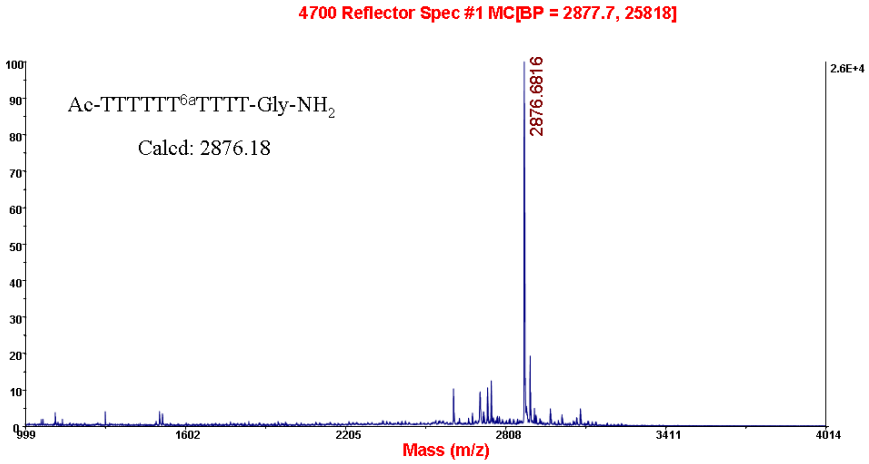

MALDI-TOF MASS spectrum of **GP-PNA 6-5**



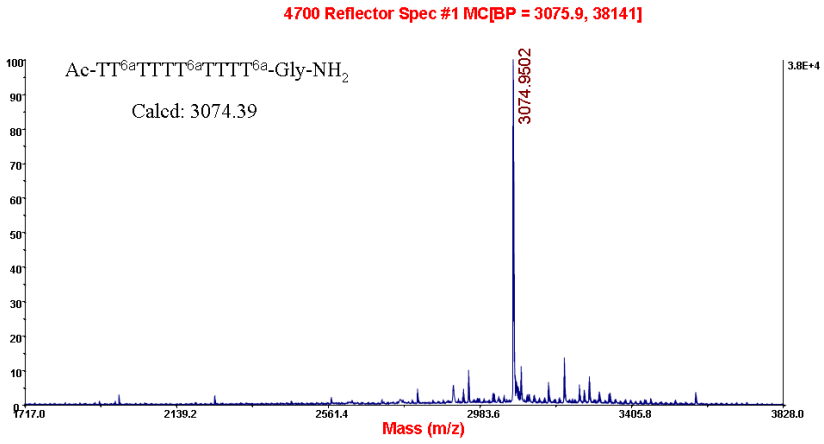

MALDI-TOF MASS spectrum of **aegPNA-$T_{10}$**



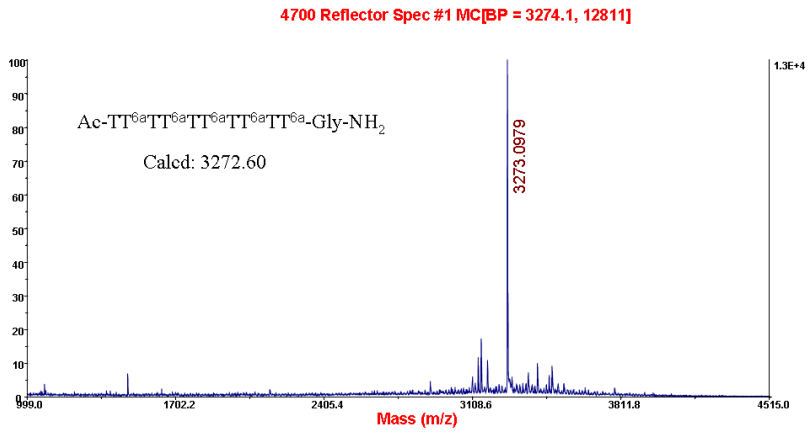

MALDI-TOF MASS spectrum of **AP-PNA-$T_{10}$ 2-5**



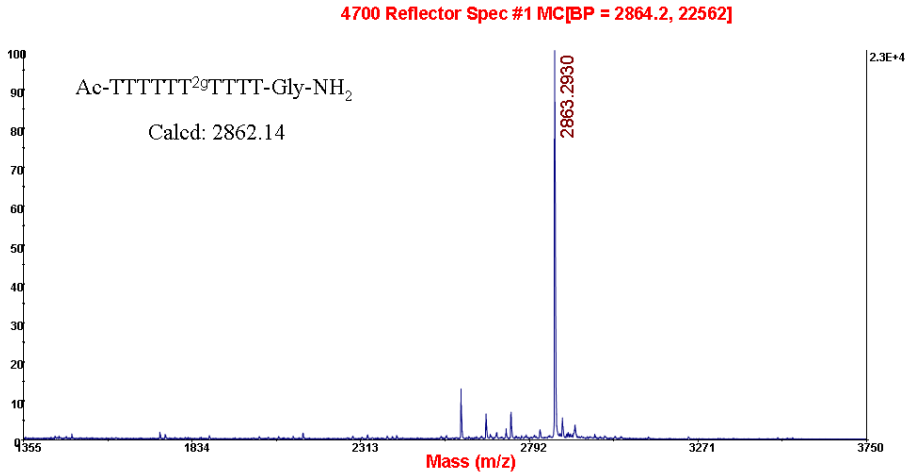

MALDI-TOF MASS spectrum of **AP-PNA-$T_{10}$ 6-1**



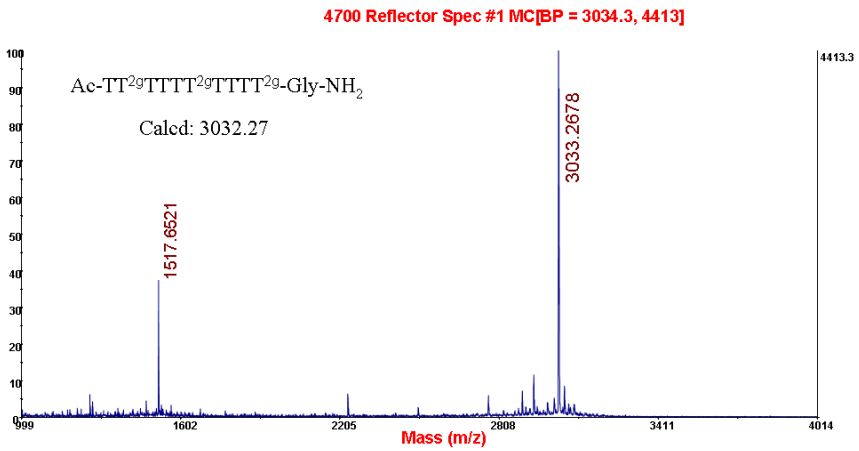

MALDI-TOF MASS spectrum of **AP-PNA-$T_{10}$ 6-2**



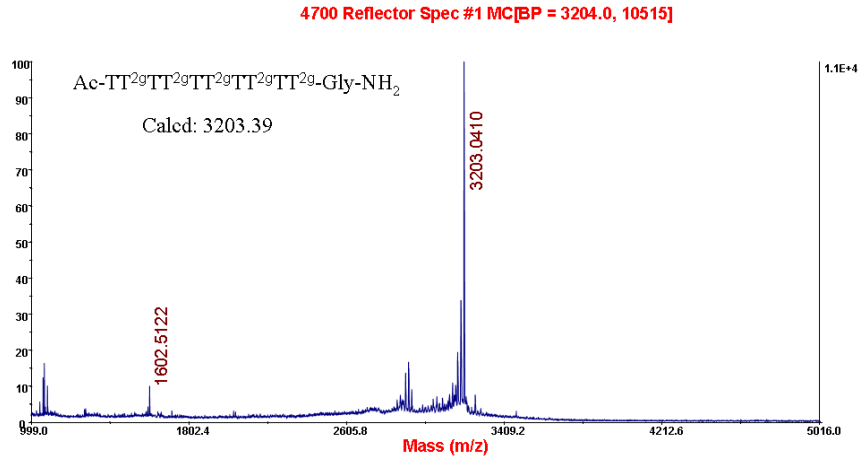

MALDI-TOF MASS spectrum of **AP-PNA-$T_{10}$ 6-5**



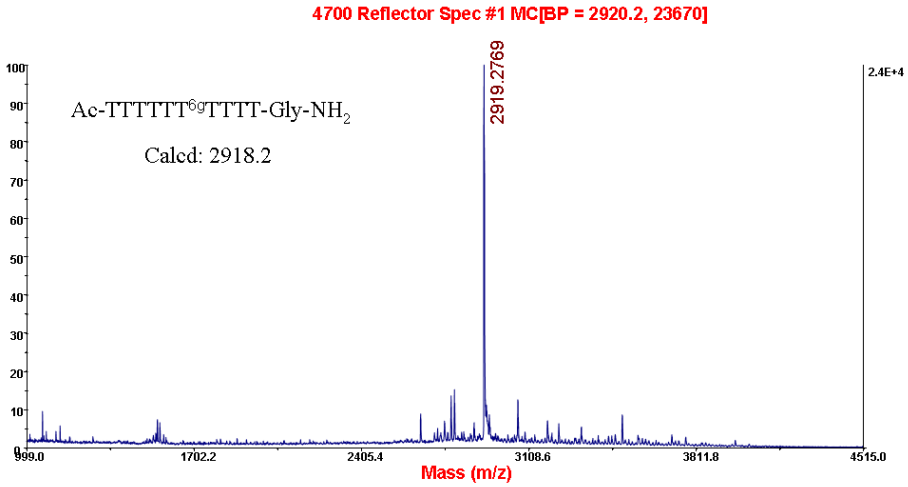

MALDI-TOF MASS spectrum of **GP-PNA-$T_{10}$ 2-1**



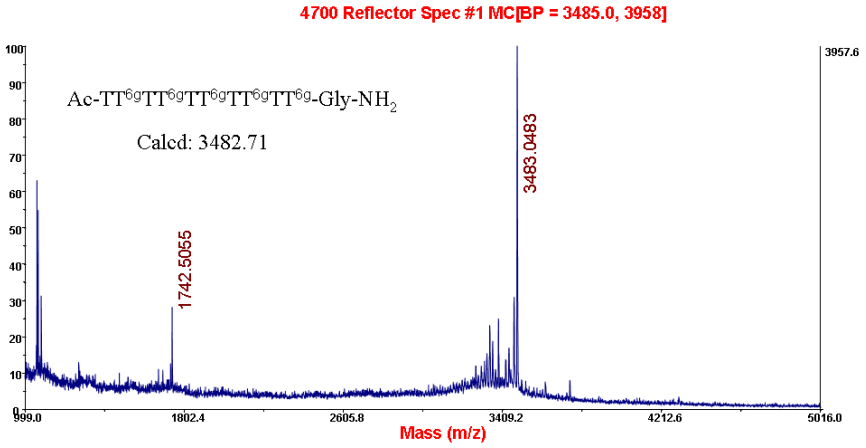

MALDI-TOF MASS spectrum of **GP-PNA-$T_{10}$ 2-3**



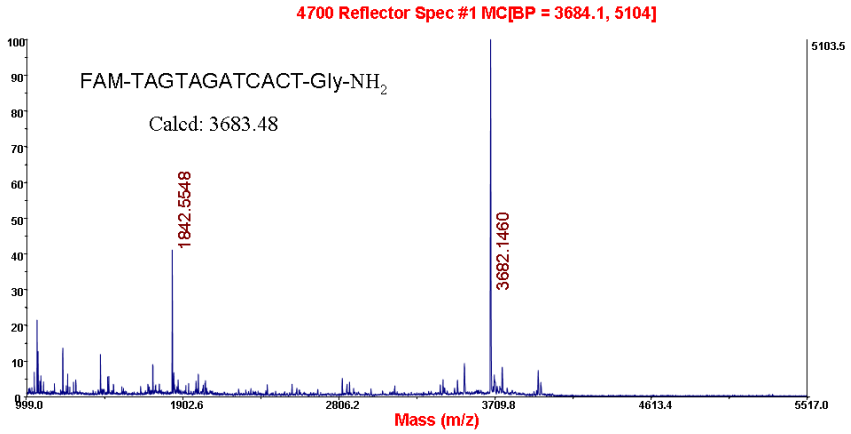

MALDI-TOF MASS spectrum of **GP-PNA-$T_{10}$ 2-5**



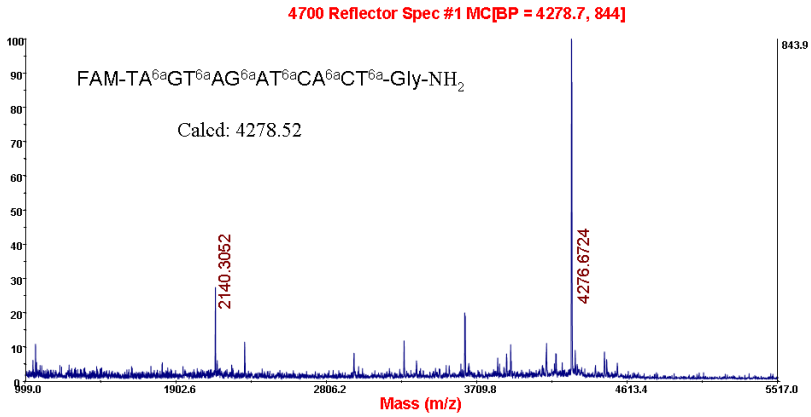

MALDI-TOF MASS spectrum of **GP-PNA-$T_{10}$ 6-1**



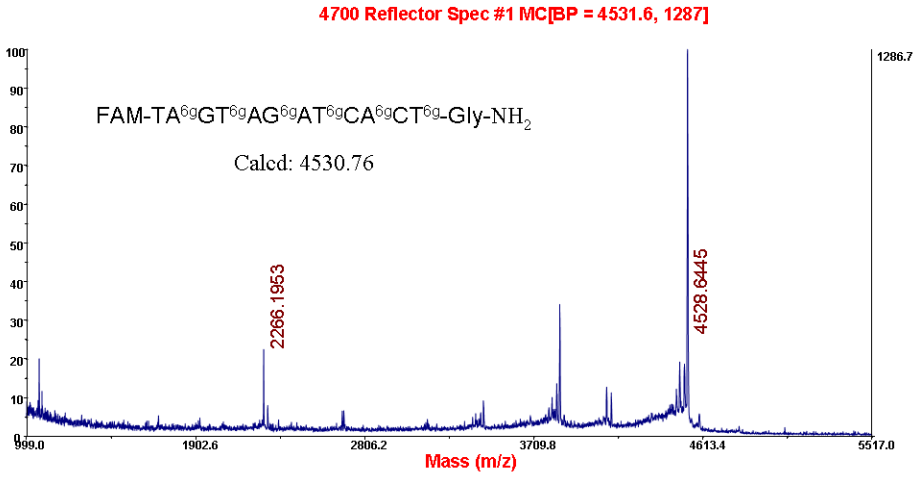

MALDI-TOF MASS spectrum of **GP-PNA-$T_{10}$ 6-5**




MALDI-TOF MASS spectrum of **Fam-aegPNA**




MALDI-TOF MASS spectrum of **Fam-AP-PNA 6-6**




MALDI-TOF MASS spectrum of **Fam-GP-PNA 6-6**