Winter 2009

Molecular Dynamics Study of Single Stranded Peptide Nucleic Acids

Anna K. Manukyan
Old Dominion University

MOLECULAR DYNAMICS STUDY OF SINGLE STRANDED PEPTIDE NUCLEIC ACIDS

by

Anna K. Manukyan
B.S. May 2002, Armenian State University

A Dissertation Submitted to the Faculty of
Old Dominion University in Partial Fulfillment of the
Requirement for the Degree of

DOCTOR OF PHILOSOPHY

BIOMEDICAL SCIENCES

OLD DOMINION UNIVERSITY
December 2009

Approved by:

Jennifer Poutsma (Director)

Patricia Pleban (Member)

Kenneth Brown (Member)

Christopher Osgood (Member)

Craig Bayse (Member)

ABSTRACT

MOLECULAR DYNAMICS STUDY OF SINGLE STRANDED PEPTIDE NUCLEIC ACIDS

Anna K. Manukyan
Old Dominion University, 2009
Director: Dr. Jennifer Poutsma

A PNA molecule is a DNA strand where the sugar-phosphate backbone has been replaced by a structurally homomorphous pseudopeptide chain consisting of *N* (2-amino-ethyl)-glycine units. PNA binds strongly to both DNA and RNA. However, an analysis of the X-ray and NMR data show that the dihedral angles of PNA/DNA or PNA/RNA complexes are very different from those of DNA:DNA or RNA:RNA complexes. In addition, the PNA strand is very flexible. One way to improve the binding affinity of PNA for DNA/RNA is to design a more pre-organized PNA structure. An effective way to rigidify the PNA strand is to introduce ring structures into the backbone. In several experimental studies, the ethylenediamine portion of aminoethylglycine peptide nucleic acids (*aeg*PNA) has been replaced with one or more (S,S)-*trans* cyclopentyl (*cp*PNA) units. This substitution has met with varied success in terms of DNA/RNA recognition.

In the present work, molecular modeling studies were performed to investigate PNA and modified PNA analogs. Molecular dynamics (MD) simulations is a principal tool in the theoretical study of biological molecules. This computational method calculates the time dependent behavior of a molecular system and provides detailed information on the fluctuations and conformational changes. The MD simulation uses an empirical parameterized energy functions. These parameters play an important role in the quality of the simulations. Therefore, novel empirical force field parameters were developed for cyclopentane modified PNA analogs. We demonstrate that our parameterization can accurately reproduce high level quantum mechanical calculations.

Detailed investigations on the conformational and dynamical properties of single stranded *aeg*PNA and *cp*PNA were undertaken to determine how the cyclopentane ring will improve binding and to determine the contributions of both entropy and dihedral

angle preference to the observed stronger binding. The effects of single and multiple modifications of the PNA backbone were also analyzed to understand changes in conformational and dynamical properties.

This dissertation is dedicated to my family and to my husband, Hovhannes Grigoryan

ACKNOWLEDGMENTS

I would like to address special thanks to my advisor, Dr. Jennifer Poutsma, for her mentoring and support. I am very grateful to Professors Brown and Osgood for their encouragement and help in the achievement of this work. I also thank my committee members, Professors Pleban and Bayse, for their insightful comments.

I thank Dr. J.C Poutsma for motivation and helpful advice. Additionally, I would like to thank the members of my group, Chuanyin Shi and Jeff Tibbitt, for many conversations and helpful comments. I would also like to thank my colleagues and friends Sonia Antony and Prakash Nallathamby for moral support. I thank Prof. Roux and his group for offering to let me visit his lab for the last year. Finally, I thank Dr. Lennart Nilsson for providing the PNA parameters.

TABLE OF CONTENTS

Page

LIST OF TABLES vii

LIST OF FIGURES xii

Chapter

I. INTRODUCTION 1
NUCLEIC ACID ANALOGS 3
PEPTIDE NUCLEIC ACIDS 8
PHYSICAL PROPERTIES AND HYBRIDIZATION OF PNA 9
APPLICATIONS OF PNA 13
LIMITING FACTORS FOR PNA APPLICATIONS 17
STRATEGIES OF PNA MODIFICATIONS 19

II. THEORETICAL BACKGROUND 33
EMPIRICAL FORCE FIELDS 33
QUANTUM CHEMICAL METHODS 37
PARAMETERIZATION STRATEGY 45

III. FORCEFIELD PARAMETERS FOR CYCLOPENTYL PNA. 49
COMPUTATIONAL METHODS 49
RESULTS AND DISCUSSION 53
PARTIAL CHARGES 55
GEOMETRIC FACTORS 61
TORSIONAL PROFILES 68
PSEUDOROTATION AND STRUCTURAL NOMENCLATURE 78
TESTING OF THE PARAMETERIZATION 83

IV. STUDIES WITH SINGLE STRANDED PNA 90
INTRODUCTION AND MOTIVATION 90
THEORY AND METHODOLOGY 94
EFFECT OF A SINGLE CYCLOPENTANE MODIFICATION 104
THE EFFECT OF THE TERMINAL LYSINE 141
THE EFFECT OF THE POSITION 155
THE EFFECT OF MULTIPLE MODIFICATIONS 177

V. SUMMARY 231

REFERENCES 234

VITA 249

LIST OF TABLES

Table		Page
1.	Comparison of the melting temperatures of the duplexes	10
2.	Backbone dihedral angles for the modified PNA structures	30
3.	T_m data for PNA:DNA duplexes	31
4.	Model compound-water interaction energies and distances	58
5.	Average interaction energies (kcal/mol) between water and the model compound	59
6.	Atom types and assigned partial charges for the *trans-cp*PNA model	60
7.	Average bond length and angle parameters for the model compound	62
8.	Vibrational frequency assignments for the model molecule	65
9.	Dipole moments for the seven conformers of the model	68
10.	Torsional parameter set for the *cp*PNA model molecule	77
11.	Puckering phase angles and relative energies from *ab initio* and CHARMM calculations	81
12.	Geometric data (°) at the energy minimum	82
13.	The barriers to interconversion obtained by QM and MM methods	83
14.	Conformer types found by *ab initio* and MM calculations	84
15.	Dihedral angles of the central conformation of each cluster	88
16.	Selected dihedrals and relative energies of the model compound conformations obtained by *ab initio* calculations at the HF/6-31G(d) level of theory	89
17.	The numbering scheme and the modified oligonucleotides used	93
18.	Average torsion angles for each cluster from the *aeg*PNA**1** simulation	106

Table Page

19. Similar clusters identified from the different clustering Schemes for *aeg*PNA**1** 107

20. Average dihedral values (°) and RMSF (bold, °) of the simulated *aeg*PNA**1** 108

21. Average torsion angles for each cluster from the *cp*PNA**2** simulation 111

22. Similar clusters identified from the different clustering Schemes for *cp*PNA**2** 112

23. Average dihedrals (°) and RMSF (bold, °) from cluster analysis of *cp*PNA**2** 113

24. Average torsion angles for each cluster from the *aeg*PNA**3** simulation 116

25. Similar clusters identified from the different clustering Schemes for *aeg*PNA**3** 117

26. Average torsion angles for each cluster from the *cp*PNA**4** simulation 120

27. Similar clusters identified from the different clustering Schemes for *cp*PNA**4** 121

28. Average dihedrals (°) and RMSF (bold, °) of the simulated *aeg*PNA**3** 121

29. Average dihedrals (°) and RMSF (bold, °) of the simulated *cp*PNA**4** 122

30. Average torsion angles for each cluster from the *aeg*PNA**12** simulation 126

31. Similar clusters identified from the different clustering Schemes for *aeg*PNA**12** 127

32. Average dihedrals (°) and RMSF (bold, °) of the simulated *aeg*PNA**12** 128

33. Average torsion angles for each cluster from the *cp*PNA**13** simulation 129

34. Similar clusters identified from the different clustering Schemes for *cp*PNA**13** 130

35. Average dihedrals (°) and RMSF (bold, °) of the simulated *cp*PNA**13** 132

36. Average torsion angles for each cluster from the *aeg*PNA**14** simulation 135

Table Page

37. Similar clusters identified from the different clustering Schemes for *aeg*PNA**14** 136

38. Average torsion angles for each cluster from the *cp*PNA**15** simulation 138

39. Similar clusters identified from the different clustering Schemes for *cp*PNA**15** 139

40. Average dihedrals (°) and RMSF (bold, °) of the simulated *aeg*PNA**14** 139

41. Average dihedrals (°) and RMSF (bold, °) of the simulated *cp*PNA**15** 139

42. Average torsion angles for each cluster from the *cp*PNA**4a** simulation 144

43. Similar clusters identified from the different clustering Schemes for *cp*PNA**4a** 145

44. Average dihedrals (°) and RMSF (bold, °) of the simulated *cp*PNA**4a** 147

45. Average torsion angles for each cluster from the *cp*PNA**15a** simulation 152

46. Similar clusters identified from the different clustering Schemes for *cp*PNA**15a** 152

47. Average dihedrals (°) and RMSF (bold, °) of the simulated *cp*PNA**15a** 154

48. Average torsion angles for each cluster from the *cp*PNA**5** simulation 157

49. Similar clusters identified from the different clustering Schemes for *cp*PNA**5** 158

50. Average dihedrals (°) and RMSF (bold, °) of the simulated *cp*PNA**5** 159

51. Average torsion angles for each cluster from the *cp*PNA**6** simulation 162

52. Similar clusters identified from the different clustering Schemes for *cp*PNA**6** 163

53. Average dihedrals (°) and RMSF (bold, °) of the simulated *cp*PNA**6** 164

54. Average torsion angles for each cluster from the *cp*PNA**16** simulation 167

55. Similar clusters identified from the different clustering Schemes for *cp*PNA**16** 168

Table Page

56. Average dihedrals (°) and RMSF (bold, °) of the simulated *cp*PNA**16**.....169

57. Average torsion angles for each cluster from the *cp*PNA**17** simulation.....172

58. Similar clusters identified from the different clustering Schemes for *cp*PNA**17**.....173

59. Average dihedrals (°) and RMSF (bold, °) of the simulated *cp*PNA**17**.....174

60. Average torsion angles for each cluster from the *cp*PNA**18** simulation.....177

61. Similar clusters identified from the different clustering Schemes for *cp*PNA**18**.....178

62. Average dihedrals (°) and RMSF (bold, °) of the simulated *cp*PNA**18**.....179

63. Average torsion angles for each cluster from the *cp*PNA**7** simulation.....183

64. Similar clusters identified from the different clustering Schemes for *cp*PNA**7**.....184

65. Average dihedrals (°) and RMSF (bold, °) of the simulated *cp*PNA**7**.....185

66. Average torsion angles for each cluster from the *cp*PNA**8** simulation.....188

67. Similar clusters identified from the different clustering Schemes for *cp*PNA**8**.....189

68. Average dihedrals (°) and RMSF (bold, °) of the simulated *cp*PNA**8**.....190

69. Average torsion angles for each cluster from the *cp*PNA**9** simulation.....193

70. Similar clusters identified from the different clustering Schemes for *cp*PNA**9**.....194

71. Average dihedrals (°) and RMSF (bold, °) of the simulated *cp*PNA**9**.....195

72. Average torsion angles for each cluster from the *cp*PNA**10** simulation.....198

73. Similar clusters identified from the different clustering Schemes for *cp*PNA**10**.....199

74. Average dihedrals (°) and RMSF (bold, °) of the simulated *cp*PNA**10**.....200

Table Page

75. Torsion angle data for each cluster from simulation of *cp*PNA**11** 202

76. Similar clusters identified from the different clustering Schemes for *cp*PNA**11** 202

77. Average dihedrals (°) and RMSF (bold, °) of the simulated *cp*PNA**11**.....204

78. Average torsion angles for each cluster from the *cp*PNA**19** simulation.....208

79. Similar clusters identified from the different clustering Schemes for *cp*PNA**19** 209

80. Average dihedrals (°) and RMSF (bold, °) of the simulated *cp*PNA**19**.....210

81. Average torsion angles for each cluster from the *cp*PNA**20** simulation.....213

82. Similar clusters identified from the different clustering Schemes for *cp*PNA**20** 214

83. Average dihedrals (°) and RMSF (bold, °) of the simulated *cp*PNA**20**.....215

84. Average torsion angles for each cluster from the *cp*PNA**21** simulation.....217

85. Similar clusters identified from the different clustering Schemes for *cp*PNA**21** 218

86. Average dihedrals (°) and RMSF (bold, °) of the simulated *cp*PNA**21**.....219

87. Average torsion angles for each cluster from the *cp*PNA**22** simulation.....222

88. Similar clusters identified from the different clustering Schemes for *cp*PNA**22** 222

89. Average dihedrals (°) and RMSF (bold, °) of the simulated *cp*PNA**22**.....225

90. Dihedral angles in DNA and RNA duplexes 227

LIST OF FIGURES

Figure		Page
1.	Structural representation of the DNA molecule	2
2.	Base triplets that can form with PNA (Watson-Crick and Hoogsteen)	2
3.	Sites for chemical modifications of ribonucleotides	4
4.	Structures of various nucleic acid analogs	4
5.	Examples of nucleobase and sugar modifications in oligonucleosides	6
6.	Alternatives to the phosphodiester backbone	7
7.	Comparison of the DNA and PNA structures	8
8.	Schematic model of strand displacement and triplex formation	11
9.	Potential mechanisms of antisense and antigene oligonucleotides	14
10.	Strategies for detecting gene expression: fluorescently labeled PNA probes (top) and molecular beacons (bottom)	17
11.	Potential sites for the modification of the PNA backbone	19
12.	Examples of modified PNA nucleobase moieties	20
13.	Double duplex invasion	21
14.	Olefinic Peptide nucleic acids (OPA)	23
15.	Examples of modified PNA backbone moieties	24
16.	Examples of PNA backbone modifications with cyclic moieties	26
17.	Definition of the PNA backbone dihedral angles	29
18.	Schematic representation of force field energy terms	35
19.	Atom orientation in the torsion angle	36
20.	Parameter optimization cycle	47

Figure Page

21. CHARMM atom types assigned to the model compound 54

22. Chemical structure and optimized geometries of the model compound at the HF/6-31G(d) level of theory 55

23. Locations of the model-water interaction sites 57

24. The lowest energy QM structures for the model-water complexes 59

25. Vibrational frequencies of the model molecule 66

26. Backbone dihedral angles used for parameterization 69

27. Torsional energy plots for the model compound 73

28. Structure (top) and five torsional energy plots for the model compound 76

29. The position of the torsional angles on the cyclopentane ring 78

30. A common representation of the pseudorotation of five-membered rings 80

31. Energy profile along the pseudorotation angle, *P* 81

32. QM and MM relative energies of the model compound 85

33. The correlation of the relative energies calculated by both methods: QM and MM 86

34. RMSD plot for the model compound relative to its initial structure 87

35. Notations for the backbone torsion 94

36. RMSD of the backbone atoms as a function of time with respect to the starting structure for the simulations of PNA-T_8 sequences 99

37. RMSD time course of the PNA backbone atoms with reference to initial structure of PNA-MB_{10} simulations 100

38. Radius of gyration of octamer (PNA-T_8) structures plotted against time 102

39. Radius of gyration of decamer structures (PNA-MB_{10}) plotted against time 103

40. Some representative backbone structures observed 104

Figure Page

41. Representative structures of the most populated clusters from the MD simulation of *aeg*PNA**1** using the different clustering Schemes 107

42. Representative structures of the most populated clusters from the MD simulation of *cp*PNA**2** using the different clustering Schemes 110

43. Visualization of the most populated molecular structures corresponding to *aeg*PNA**3** cluster averages 115

44. Conformation of the central members of the most populated clusters in the simulations of *cp*PNA**4** 119

45. The most populated conformers observed in MD simulations of *aeg*PNA**12** 125

46. Central members of the most populated clusters of structures from *cp*PNA**13** simulations 131

47. Structures of central cluster members of the simulation of *aeg*PNA**14** 134

48. Structures of central cluster members of the simulation of *cp*PNA**15** 137

49. Representative structures of the most populated clusters from the MD simulation of *cp*PNA**4a** 146

50. Representative structures of the most populated clusters from the MD simulation of *cp*PNA**15a** 151

51. Structures of cluster centers of most populated clusters from *cp*PNA**5** simulations 156

52. Representative structures of the most populated clusters from the MD simulation of *cp*PNA**6** 161

53. Representative structures of the most populated clusters from the MD simulation of *cp*PNA**16** 166

54. Cluster centroid structures obtained from simulations of *cp*PNA**17** 171

55. Representative structures of the most populated clusters from the MD simulation of *cp*PNA**18** 176

56. Representative structures of the most populated clusters from the MD simulation of *cp*PNA**7** 182

Figure Page

57. Representative structures of the most populated clusters from the MD simulation of *cp*PNA**8**.....187

58. Representative structures of the most populated clusters from the MD simulation of *cp*PNA**9**.....192

59. Representative structures of the most populated clusters from the MD simulation of *cp*PNA**10**.....197

60. Representative structures of the most populated clusters from the MD simulation of *cp*PNA**11**.....201

61. Representative structures of the most populated clusters from the MD simulation of *cp*PNA**19**.....207

62. Representative structures of the most populated clusters from the MD simulation of *cp*PNA**20**.....212

63. Representative structures of the most populated clusters from the MD simulation of *cp*PNA**21**.....216

64. Representative structures of the most populated clusters from the MD simulation of *cp*PNA**22**.....221

CHAPTER I

INTRODUCTION

The importance of nucleic acids in such processes as growth, reproduction, heredity and viral infection led to the establishment of oligonucleotide therapeutics. The two types of nucleic acid molecules known as ribonucleic acid, or RNA, and deoxyribonucleic acid, or DNA, perform different functions in these vital processes. DNA contains genetic code necessary for the development and function of living organisms and is considered as the most important biomolecule since its structure was determined by Watson and Crick.[1] The storage and transmission of the genetic information is facilitated via replication, transcription and translation, mechanisms that involve nucleic acid-protein regulatory interactions. This observation suggests that in all domains of life these two polymers are strongly dependent on each other; DNA sequence encodes the amino acid content of proteins, while proteins are required for the maintenance and replication of the genome. The active communication between DNA and proteins is carried out by RNA. The DNA/RNA/protein *central dogma*[2] of molecular biology is very complicated as it requires both nucleic acids and proteins to be formed from simple organic molecules.

Early investigations into the structure of oligonucleotides led to the proposal of the double stranded model of DNA (Figure 1), a helical structure composed of a sugar-phosphate backbone and four different nitrogenous bases stacked in the center of the chains. Both helical polynucleotide strands of the complex are right-handed, but run in opposite directions. The DNA bases consist of planar, aromatic, heterocyclic molecules that are divided into two groups: purine bases, adenine (A) and guanine (G), and pyrimidine bases, cytosine (C) and thymine (T). Within the DNA double helix, one base from each strand forms a coplanar base pair. Adenine pairs with thymine via two hydrogen bonds, and guanine forms three hydrogen bonds with cytosine (Figure 2) to form the standard complementary Watson-Crick base pairs.

The citations on the following pages follow the style of the *Journal of Chemical Theory and Computation.*

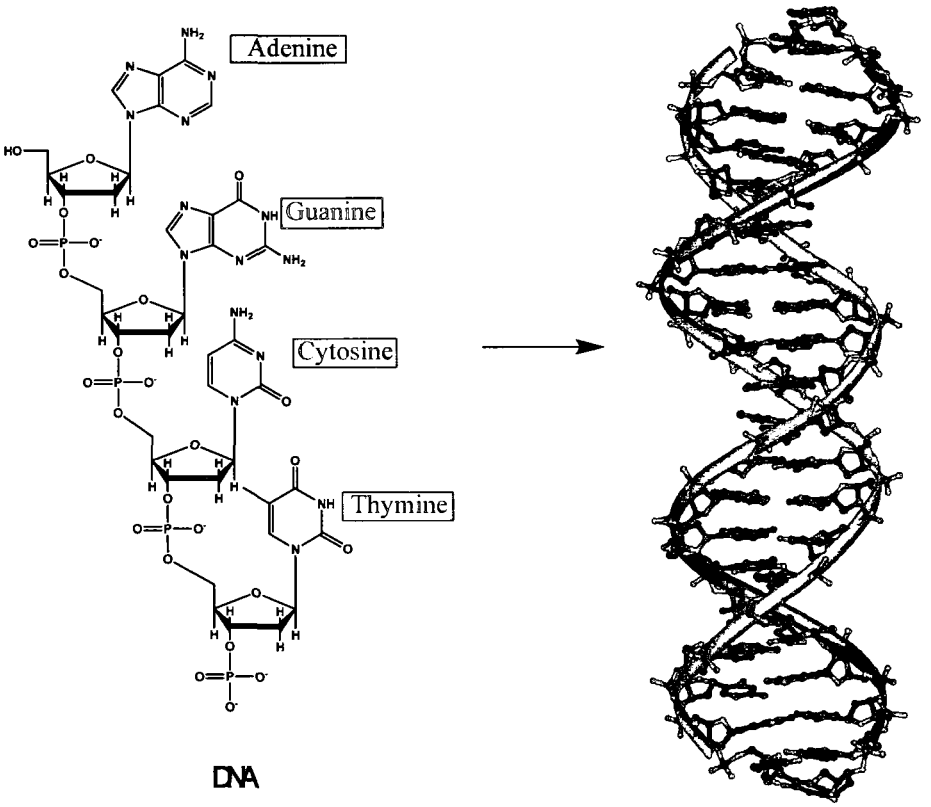

Figure 1. Structural representation of the DNA molecule

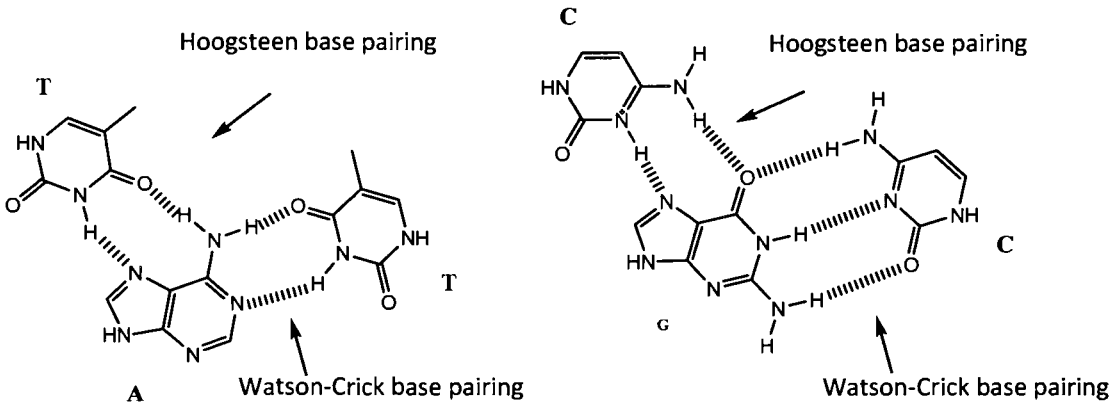

Figure 2. Base triplets that can form with PNA (Watson-Crick and Hoogsteen)

Hydrophobic interactions between nucleotide bases and the aqueous environment of the cell induce base stacking that minimizes their contact with water. The stability of the DNA molecule is governed by various non-covalent interactions that maintain its secondary structure in aqueous solution.[3] Among these, Watson-Crick hydrogen bonding and base stacking are always stabilizing. On the other hand, electrostatic repulsion between adjacent phosphate groups is always destabilizing. Steric effects vary in their contribution depending on the nature of the interacting bases. The inherent plasticity of the DNA molecule allows formation of right-handed B-DNA and A-DNA helices and left-handed Z-DNA,[4,5] as well as parallel and anti-parallel orientations of binding and multi-stranded complexes.

NUCLEIC ACID ANALOGS

The ability of certain regulatory proteins and nucleic acids to recognize and bind specific nucleotide sequences has led to their widespread applications in research. Although natural molecules have a high affinity to complementary RNA and DNA strands, their phosphodiester backbone is extremely susceptible to degradation by cellular nucleases. Furthermore, the products of degradation are cytotoxic and inhibit cell growth in living cells.[6,7] To improve upon these shortcomings, which prevent their widespread application, chemists developed the ability to synthesize nucleic acid analogues. Synthetic oligonucleotides have been of considerable interest not only from the biological viewpoint, but also as simplified models for the study of nucleic acid chemistry. For a successful application of such polynucleotides in DNA diagnostics and genetic therapy, high sequence-specificity and binding affinity are crucial. To date approaches to the design of oligomers have included modifications of the sugar-phosphate backbone or the nucleobase (Figure 3).

The first generation of analogues involved alteration of the internucleotide linkages with the purpose of enhancing enzymatic resistance. The earliest used and most extensively tested analogues of natural phosphates are phosphorothioate oligodeoxynucleotides developed by Eckstein and coworkers.[8] In these molecules, one of

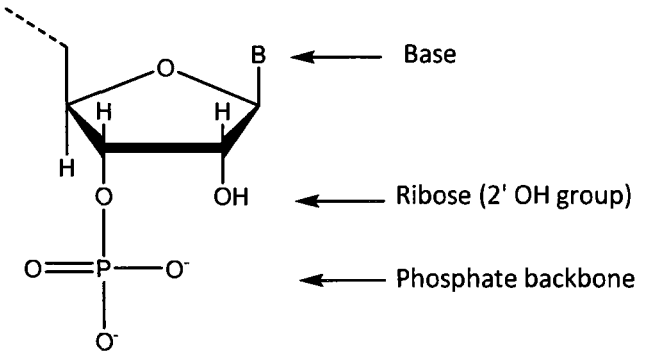

Figure 3. Sites for chemical modifications of ribonucleotides

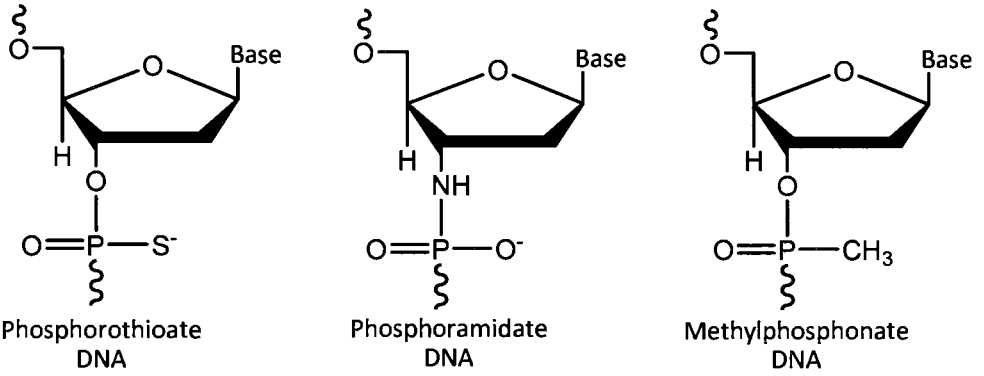

Figure 4. Structures of various nucleic acid analogs

the non-bridge phosphate oxygen atoms is replaced by sulfur. Phoshorothioates are more resistant to nucleases[9] and show improvements in their pharmacokinetic properties with increased *in vivo* half-life.[10] However, anticancer and antiviral applications revealed major drawbacks, namely low affinity of binding and non-sequence specific interactions under certain circumstances.[11] Since charged molecules have a relatively poor ability to penetrate cell membranes, much effort has been concentrated on introducing charge-neutral phosphate backbones.[12] Methylphosphonates[13] and phosphoramidates[14, 15] are examples of nonionic analogues of olionucleotides (Figure 4). Both are highly resistant to exonucleases, form very stable duplexes with complementary sequences, and

demonstrate effective sequence-specific inhibition of gene expression, but lack the ability to recruit RNase H activity.[16] The RNase H mode of action involves the recognition of the heteroduplex (RNA:oligonucleotide) by this intracellular nuclease which cleaves the target RNA strand thereby preventing translation of mRNA into protein and leaving the antisense oligonucleotide strand intact.

A variety of heterocyclic base modifications were designed as novel purine and pyrimidine mimics with the expectation that their properties would be improved. Molecules containing base analogues were shown to be modified in thermal stability, double-helical conformation or base-pairing arrangements. The resulting oligonucleotides displayed different hybridization strength and protein recognition patterns. Examples of such modifications include thiothymine,[17] 2,4-dithiopyrimidine,[18] N^4-methyloxycytosine,[19] xanthine,[20] O^6-methylguanine[21] and 8-hydroxyguanine[22] (Figure 5). These nucleosides containing non-natural nucleobases demonstrated significant antiviral, antibacterial and anticancer activity.

The incorporation of functional groups has been further extended to include the sugar ring of DNA. Most experiments are carried out with DNA, because RNA molecules are easily degraded by RNases due to the free C2'-OH group.[23, 24] Interestingly, modification at the C2'- position leads to steric hindrance, which increases the nuclease resistance of the RNA.[25] Moreover, the C2' position of a nucleoside can be easily chemically modified and is more tolerant towards changes in size and structure because of its location on the exterior of the helix.[26] A significant number of such sugar modifications are reported in the literature.[27] The most prominent ones, shown in Figure 5, are 2'-*O*-methyl and 2'-*O*-methoxyethyl RNA. These oligonucleotides are significantly more stable against nucleases than unmodified ones, have slightly enhanced affinity for their complementary RNA target, show potent antisense effects and are less toxic.

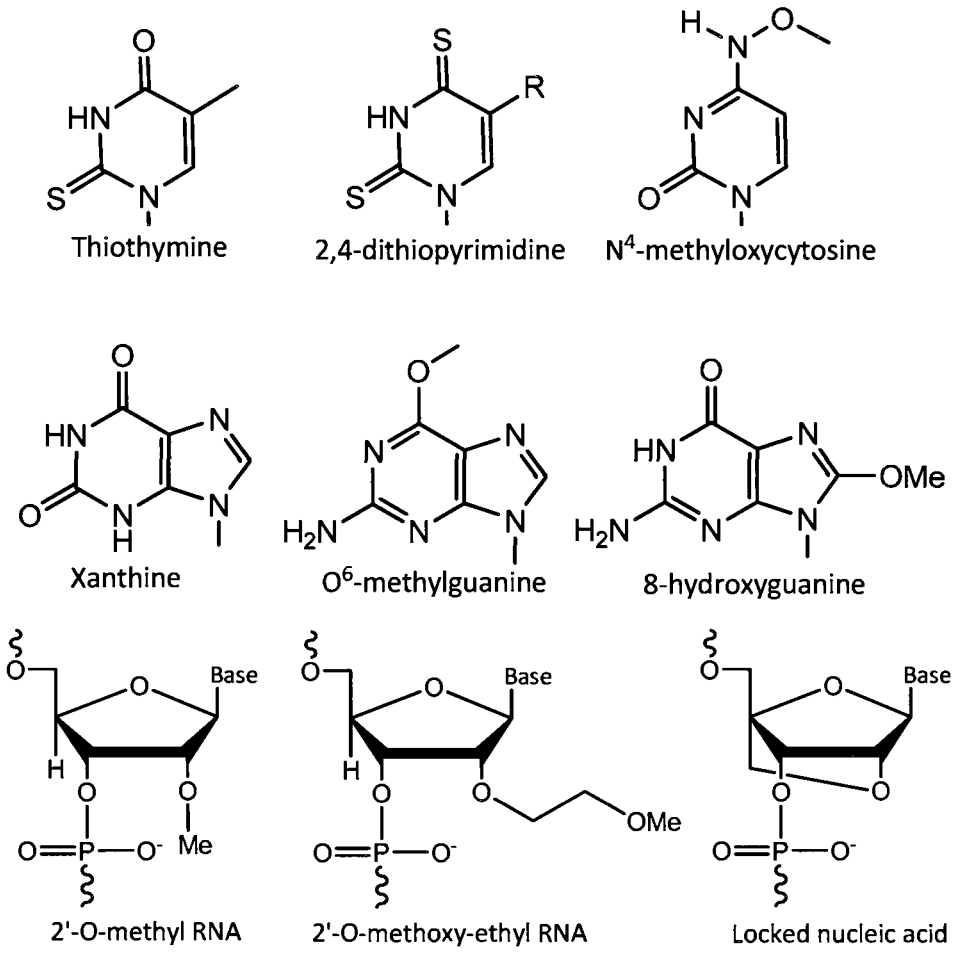

Figure 5. Examples of nucleobase and sugar modifications in oligonucleosides

An alternative strategy for sugar modification has been introduced with the aim of inducing high binding affinity via a rigid structural pre-organization. Locked nucleic acids (LNA),[28-30] also known as bridged nucleic acids (BNA), are RNA analogues that represent a very elegant and simple modification of the pentose sugar. They contain a methylene linkage that connects the 2'-hydroxyl and 4' -carbon of the ribose ring. The conformational flexibility is reduced by the presence of the bridge that confers a C3'-*endo* (β-D-LNA) or C2'-*endo* (α-D-LNA) conformation to the sugar.[31] This design yields oligomers that exhibit remarkably increased nucleic acid recognition and outstanding binding affinity to complementary DNA with melting temperature increases of up to 10°C per substitution.[29] The constrained conformation (or conformational

restriction) of LNA nucleotides in LNA/DNA complexes organizes the phosphate backbone in a way that leads to a higher population of the C3'-*endo* conformation. These changes extensively reduce the loss of entropy upon hybridization and present a favorable enthalpy associated with the improved stacking of bases. Because of their unprecedented hybridization properties and resistance to nucleolytic degradation, LNA proved to be suitable for a wide variety of biological applications.[32] However, recent studies of their effects as antisense agents in liver tissue suggest that the observed hepatoxicity can be a limiting factor in the future medicinal use of LNA.[33]

Because of their potential in many research areas, a diversity of modified oligomers has been developed. Such alterations not only affect the affinity and specificity of binding, but also possess acceptable pharmacokinetic properties. Both DNA and RNA were further modified dramatically by substituting the entire phosphate and furanose ring with completely different components while still preserving the nucleobases for hybridization properties. Replacement of the negatively charged phosphates with neutral groups can reduce the repulsion between DNA strands and enhance duplex stability. The resulting novel oligonucleotide analogues (Figure 6) were carefully tested for their biological activity. Although these molecules are not degraded by nucleases, they suffer from serious problems such as low solubility in aqueous solution and poor membrane penetration.[34]

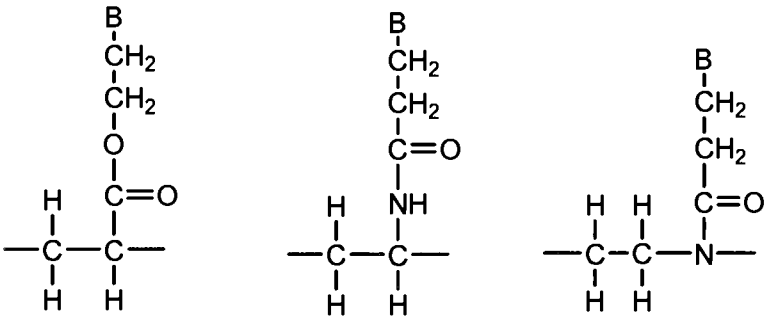

Figure 6. Alternatives to the phosphodiester backbone

PEPTIDE NUCLEIC ACIDS

Nucleic acid analogues were designed with the aim of improving upon molecular recognition and biophysical properties and/or developing novel properties that are unique from those of their natural counterparts. Peptide nucleic acids (PNAs), developed by Nielsen and coworkers,[35] were the first successful example of oligomers in which the entire sugar-phosphate scaffold of DNA/RNA has been replaced with an acyclic polyamide backbone (Figure 7). Its original design has been developed by means of computer-aided atom by atom replacement of DNA. These oligomers were created as a reagent to sequence-specifically target double stranded DNA to form a triplex via Hoogsteen base pairing in the major groove.[36]

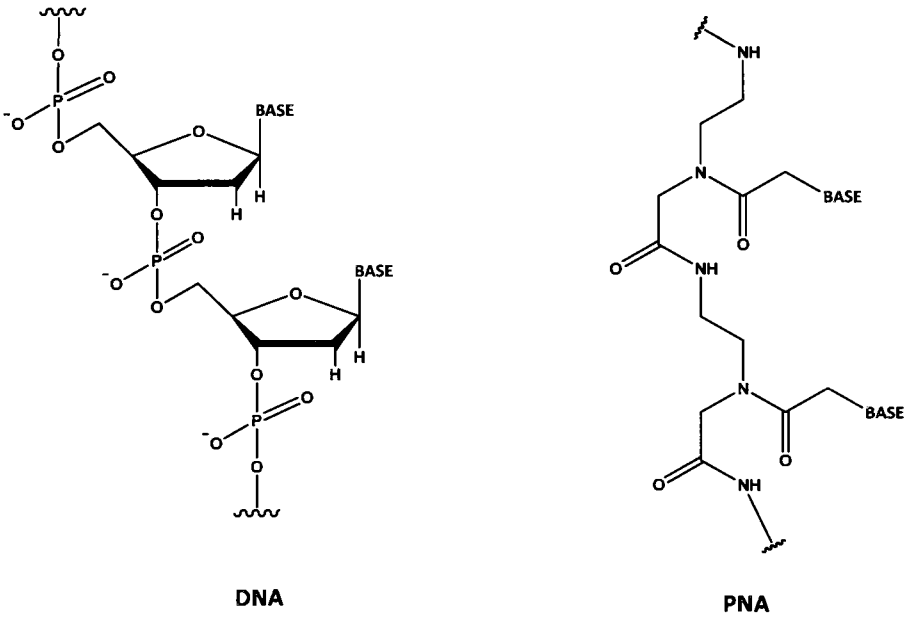

Figure 7. Comparison of the DNA and PNA structures

PHYSICAL PROPERTIES AND HYBRIDIZATION OF PNA

PNA is a synthetic DNA/RNA mimic consisting of *N*-(2-aminoethyl) glycine units (*aeg*PNA). The four natural nucleobases are attached to the glycine amino group of the backbone via methylene carbonyl linkages.[37] Geometric analysis of PNA and DNA backbones reveal that these molecules share a common "6+3" bond spacing motif. Each backbone unit is separated from the next by six bonds while the distance between the nucleobase and the backbone is three bonds. The conservation of this bond spacing motif has been suggested to contribute to the DNA mimicking properties of PNA. In analogy with regular peptides, PNA molecules are written in the *N*- to *C*- terminal direction. Egholm *et al.*[38] suggested terminology where the amino terminus of PNA corresponds to the 5'-terminus of DNA and the carboxyl terminus of PNA corresponds to the 3'-end of DNA.

The lack of the phosphate groups creates a neutral charge, and due to this neutrality PNA molecules are less soluble in water. To confer added solubility, a lysine moiety is often conjugated at the end. The non-standard, pseudo-peptide backbone of PNA exhibits increased resistance to nucleases that damage DNA samples[39] thereby, conferring advantages for its *in vivo* applications.[40] Also of significant importance is its stability to strong acids and weak bases.[41]

In general, single-stranded DNA/RNA can form complexes through Watson-Crick hydrogen bonding with any oligonucleotide bearing the correct sequence of complementary bases. Thus, duplex formation, or hybridization, is not limited to DNA:DNA or RNA:RNA complexes, but can involve RNA and artificial oligomers such as PNA. The synthetic backbone provides PNA with unique hybridization characteristics. It binds to complementary DNA, RNA and PNA in a sequence-dependent manner to form duplexes. Interestingly, PNA:PNA duplexes are even more stable than PNA:DNA duplexes of the same sequence. The higher stability of the complexes is reflected in the higher T_m values (Table 1).[26] The melting temperature (T_m) of the DNA molecule is defined as the temperature at which half of the DNA strands are in the double-helical state and half are in the "random-coil" state. The melting temperature is assayed by measuring changes in absorption in the UV range. The T_m values of PNA containing

duplexes follow the order PNA:PNA>PNA:RNA>PNA:DNA>DNA:DNA. Melting temperatures were shown to be dependent on G-C content and strand length,[42] and provided useful information about structural transitions.

Table 1. Comparison of the melting temperatures of the duplexes

Complex	T_m (°C)		T_m (°C)
DNA:DNA antiparallel	53.3	PNA:PNA antiparallel	78.6
PNA:DNA parallel	56.1	PNA:RNA parallel	51.2
PNA:DNA antiparallel	69.5	PNA:RNA antiparallel	72.3

PNA oligomers containing only thymine and cytosine (i.e. polypyrimidines) and polyadenine DNA can form very stable $(PNA)_2$:DNA triplexes by means of Watson-Crick-Hoogsteen base pairing. The stability of the PNA_2:DNA triplexes is enhanced further when the cytosine residues are replaced with a protonated cytosine mimic, such as pseudo-isocytosine.[43] The triplex formation can be used by PNA to invade double-stranded homopurine/homopyrimidine DNA targets by a P-loop strand displacement mechanism (Figure 8).[44] PNAs were also observed to form PNA_3 triplex, PNA_4[45, 46] and PNA_2:DNA_2[47] quadruplex structures.

DNA complex formation is highly dependent on the concentration of the salt in the solution, which shields the charges on the sugar-phosphate backbone, and allows access to all available binding sites on the sequence of interest.[49] Contrary to that of double-stranded DNA, the structure of PNA lacks the destabilizing electrostatic repulsions resulting in a greater binding constant at low to medium ionic strengths.[38] However, $(PNA)_2$:DNA triplex formation proved to be dependent on the ionic strength of the solution.[50]

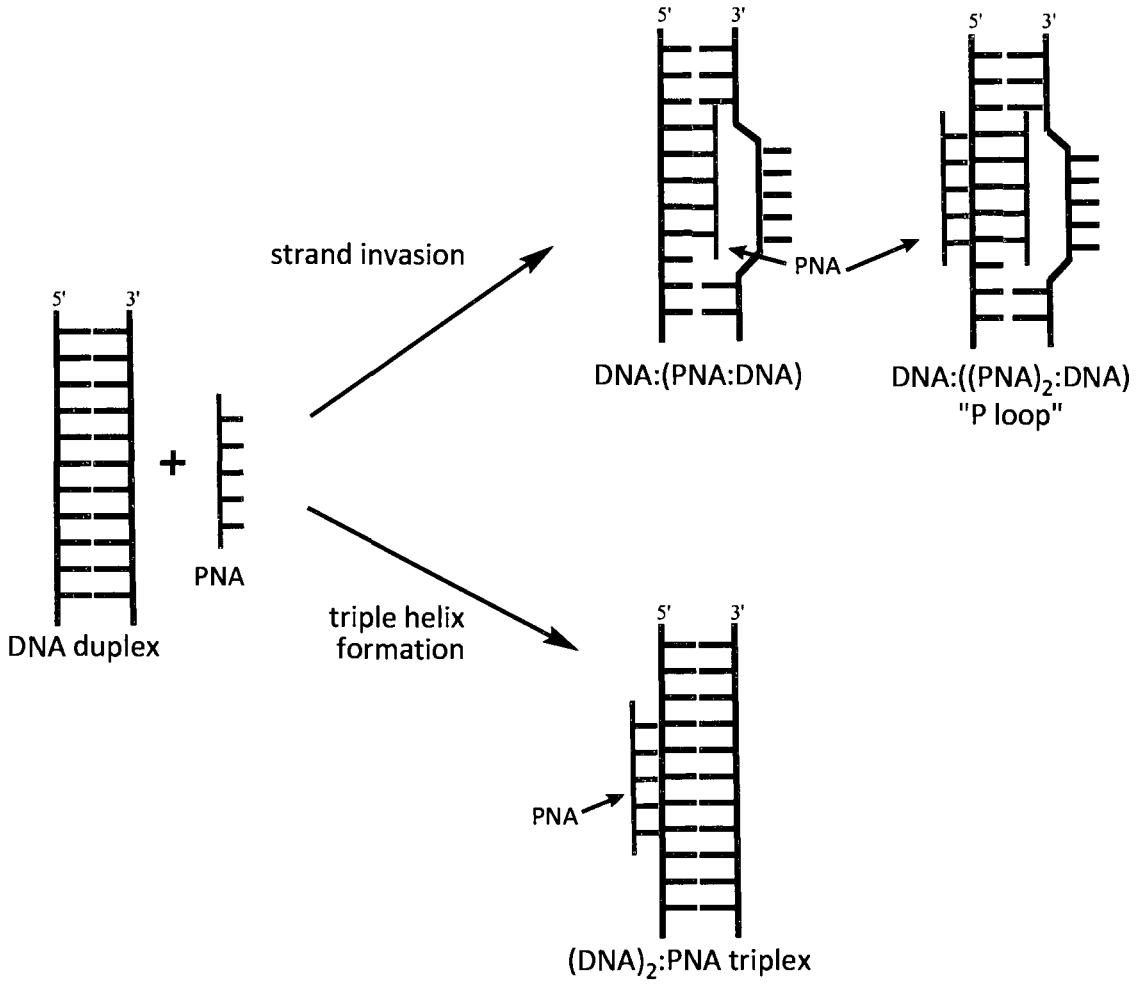

Figure 8. Schematic model of strand displacement and triplex formation

Structure of the PNA complexes

Due to the potential of PNA as a drug or a diagnostic tool, the structures of PNA-containing complexes are of great interest. Data is available on the NMR structures of the PNA:RNA,[51] PNA:DNA[48, 52] and PNA:PNA duplexes,[53] and the crystallographic structures of the $(PNA)_2$:DNA triplex[54] and PNA:PNA duplex.[55] Complexes involving PNA are helical in nature, with the PNA strand able to adapt to its nucleic acid partner. The PNA:RNA duplex was found to be a right-handed helix of "A"-form with structural features reminiscent of RNA:RNA duplexes. The primary amide bonds were in the *trans*

conformation with the carbonyl groups of the PNA directed towards the C-terminus. From the spectroscopic data collected on the RNA strand, the ribose ring adopted a C3'-*endo* conformation as in RNA and had puckered glycosidic bonds. NMR studies of the PNA:DNA duplexes revealed that they have elements of both A- and B-form DNA:DNA duplexes, demonstrating the conformational flexibility of PNA. Comparison of the PNA:DNA duplex structure with standard A- and B-DNA by superposition of the helical axes showed that while the lateral positioning of the base pairs was A-like, the backbone curvature, base-pair inclination, and helical rise were all B-like.

The X-ray structures of the $(PNA)_2$:DNA triplex and PNA:PNA duplex revealed an unusual helix, named a P-helix after PNA, characterized by a small twist angle, a large base-displacement, and a wide, deep major groove. In contrast to natural nucleic acids, PNA contains no asymmetric centers. It can form either right- or left-handed helices, depending on the chirality of the amino acid at the carboxylic end of the molecule.[56] The finding that PNA prefers a helical structure that is very different from those of the PNA:RNA and PNA:DNA helices indicates that PNA, despite its favorable hybridization properties, is not an optimal DNA mimic, leaving ample room for improvements in its design.

With the availability of NMR and X-ray crystallography data, a few computational studies were performed for further characterization of the peptide nucleic acid structures.[57-61] Nilsson *et al.* investigated the structural and dynamical properties of duplexes involving PNA molecules, as well as single stranded PNA, RNA and DNA oligomers by using molecular dynamics techniques. The obtained results were in excellent agreement with the available experimental data. Using energetic analyses, the molecular dynamics simulation studies of a PNA:DNA duplex and a $(PNA)_2$:DNA triplex in aqueous solution concluded that the hybrid PNA:DNA complexes have a characteristic helicity that is quite different from the actual DNA counterpart. The reason is not simply the replacement of the sugar phosphate backbone with a PNA backbone, but some definite conformational preference of the PNA strand that affects the conformational flexibility. These computational (modeling) studies were based on predicted structures and intended to explain why PNA never forms very B-like structures.

APPLICATIONS OF PNA

By virtue of its unique properties, PNA has found wide application in molecular biology, gene therapeutics, biosensors and nucleic acid purification. Recent progress on the studies of PNA properties and applications has been reviewed.[62-66,67] Applications of particular interest and of considerable promise include antigene and antisense biotechnologies, biosensor detection via nucleic acid hybridization and plasmid purification for pharmaceutical development.

Inhibition of gene expression (antisense and antigene)

Protein biosynthesis consists of transcription and translation processes. In the transcription step of protein production, DNA is used as a template to produce an RNA molecule, called precursor messenger RNA (pre-mRNA). During the maturation phase, the non-coding (introns) portion of the pre-mRNA is removed (splicing), leaving only the coding sequences (exons). To initiate translation, mRNA travels to the cytoplasm and forms a complex with a ribosome that assembles proteins based on the instructions contained in the mRNA. In principle, the disruption of gene expression is possible at any of these stages via antigene and antisense agents (Figure 9).

Most natural diseases such as cancer are a result of inappropriate protein production due to over-expression, silencing or absence of a particular gene. While traditional drug therapies consist of compounds that block or inhibit deficient proteins, antisense therapies focus on preventing their production. Antisense drugs are based on small oligonucleotide molecules that bind to the protein-coding portion of mRNA, preventing the translation of the disease-causing protein. By binding to a specific genetic segment of mRNA, antisense drugs prevent the genetic code from being read by the ribosome (translational arrest via steric blocking). Additionally, the bound antisense/mRNA complex is enzymatically degraded so the protein cannot be synthesized. This mechanism involves activation of RNase H that specifically cleaves the mRNA strand of the complex by means of its exo- and endonuclease activities.

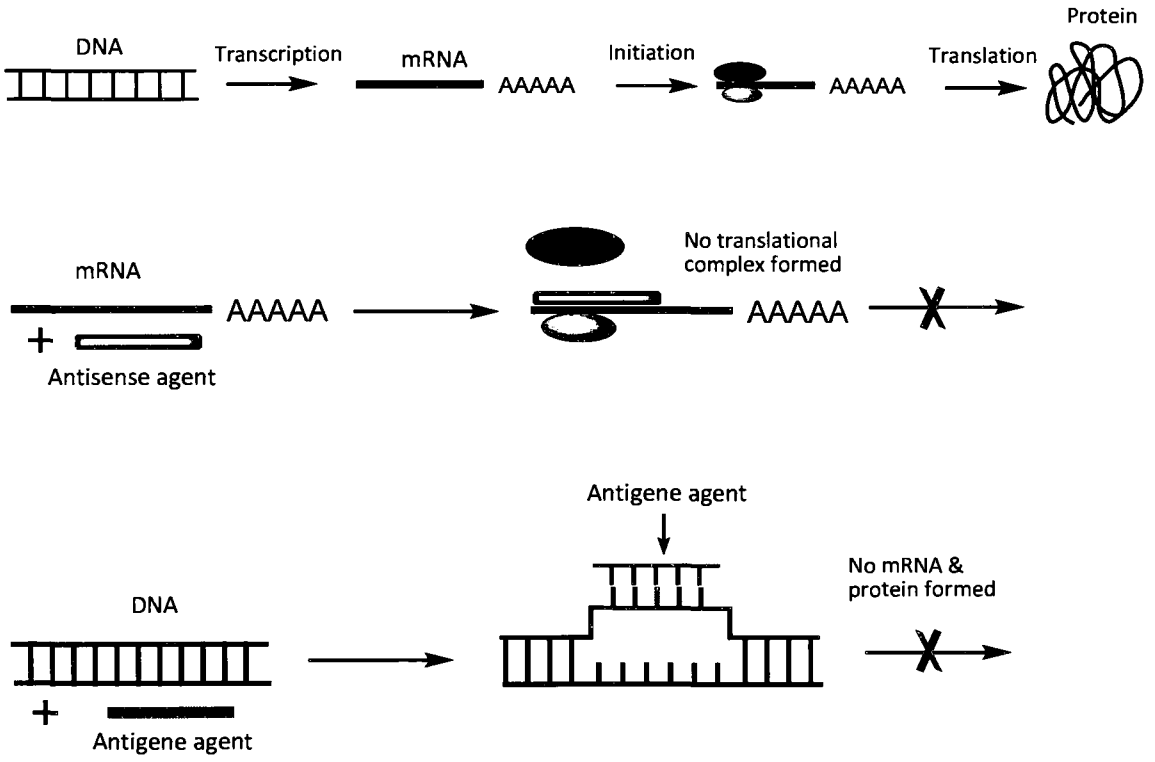

Figure 9. Potential mechanisms of antisense and antigene oligonucleotides

Due to its RNA and DNA binding capabilities, PNA has obvious potential for use in a variety of genetic procedures. Antisense PNA selectively inhibits the expression of brain proteins.[68] The design of anticancer[69-71] and antiviral[72, 73] drugs based on PNA seems to be a very promising approach. Some PNA derivatives manifest antibacterial and antisense activities towards eukaryotic cells and animal organisms.[74-78]

PNA effectively inhibits gene expression primarily through steric blockage of translation, but fails to invoke RNase H activity.[79] There are many examples of the use of duplex and triplex forming PNA to block translation in mammalian cells. Most of the research in this area is in cell-free systems, and *in vitro* translation experiments indicate that regions around the translation initiation codon (AUG) site of mRNA are very sensitive to inhibition by triplex forming PNA.[80, 81] The potent antisense effects of PNA are due to the high specificities and stabilities of the triplex $(PNA)_2$:RNA. The translational arrest of mRNA occurs even upon addition of short (6-mer) homopyrimidine

PNA strands. The antisense efficiencies of duplex-forming PNA are lower than those of triplex-forming ones because longer PNA chains are required for the inhibition. Duplex-forming PNA can inhibit translation, being specifically directed against the binding sites of ribosomes, whereas triplex forming PNA are more specific against polypurine sites located below the translation initiation point.

In order to develop a strategy to inhibit gene transcription, antigene PNAs were synthesized and characterized.[82] One of the most important studies in this regard has been reported by Corey and coworkers. PNA-mediated inhibition of gene transcription probably occurs via a process that involves $(PNA)_2$:DNA triplex formation, strand invasion of the target DNA, and formation of a PNA:DNA hybrid or strand displacement with generation of a $(DNA)_2$:PNA complex. The manner in which PNA interacts with double stranded DNA depends on the sequence composition.[83]

Several *in vitro* studies demonstrated that PNA:DNA strand displacement complexes are efficient inhibitors of transcription and hinder the binding of proteins such as transcription factors that trigger transcription initiation. Although PNA gene targeting forms stable complexes, the invasion process under physiological conditions is slow, therefore, PNA demonstrates a weak cellular antigene action. The presence of various cations that have a stabilizing effect on the DNA helix reduce the rate of duplex invasion *in vivo*.[84,85] The strand invasion process of polypyrimidine PNAs can be facilitated by attaching short cationic peptides, introducing pH-independent pseudoisocytosines, or using *bis*-PNA in which the two hybridizing chains are connected by a flexible linker.[43] It should be noted, however, that helix invasion does not necessarily lead to inhibition of transcription. As already stated above, the initiation of transcription involves formation of an open complex, equivalent to a P-loop structure of the PNA triple helix. During strand displacement, RNA polymerase can recognize this complex and initiate transcription *in vitro*.[86] Therefore, in addition to inhibiting transcription, PNAs can be considered as possible "artificial promoters".

PNA in diagnostics

The diagnosis of human diseases relies on the examination of a DNA sample for the presence of known genetic anomalies. Therefore, detection of DNA sequences for aberrant genes via nucleic acid hybridization forms the basis of research, medical diagnosis and treatment of genetic diseases. Consequently, much effort has been devoted to the development of DNA sensing methods.[87] The most popular methods to determine specific genetic sequences through hybridization are based on biosensor type platforms (DNA microarray). A typical DNA microarray is comprised of single-stranded DNA probes covalently attached to a transducer surface such as glass indium tin oxide, gold, carbon or silver. Target strands are modified with fluorescent or radio labels and those with complementary sequences to the immobilized DNA form double stranded complexes on the surface. The capture of a complementary strand on the array is coupled with a change in the properties of the duplex and is transduced into a useful electrical signal, usually in the form of light, mass (frequency) or an electrochemical change.[88,89] Due to their high binding affinity and specificity, PNAs can be used in hybridization applications as an alternative to natural nucleotide probes. PNA-based biosensors were developed for improved mismatch-sensitive hybridization detection of nucleic acids, with applications ranging from single nucleotide polymorphism detection to nucleic acid sequencing.[39,90-97] Moreover, hairpin forming PNA[98] with fluorescent and quencher moieties at opposite termini of the PNA oligomer, termed "molecular beacons", found applications in biomolecular assays by binding to designated DNA and RNA targets.[99-104] Upon binding the target sequence, the molecular beacon opens and forms a complex with RNA/DNA as a single stranded PNA probe, thereby separating the fluorescent and quenching moieties in the process (Figure 10).

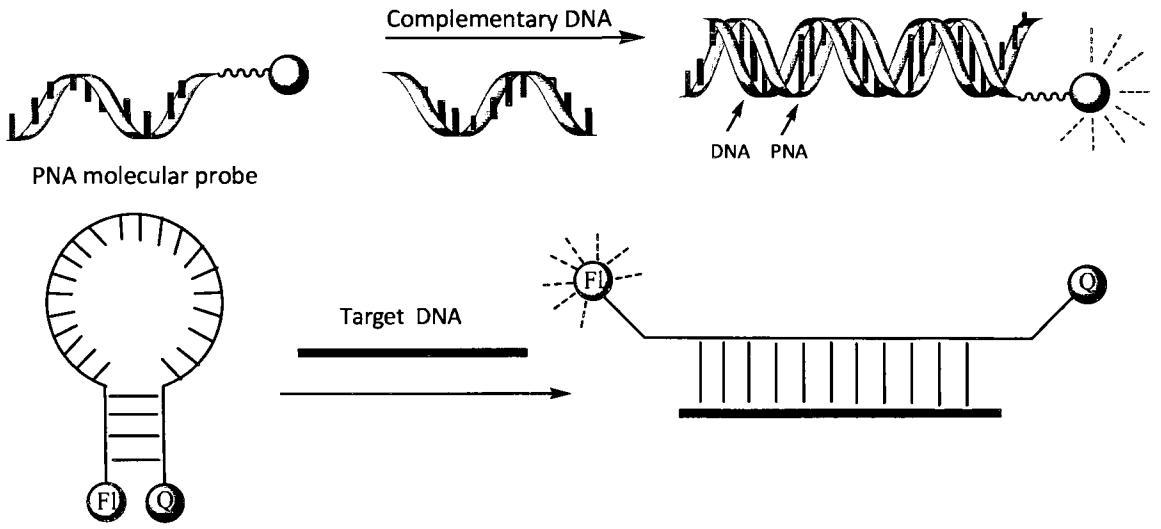

Figure 10. Strategies for detecting gene expression: fluorescently labeled PNA probes (top) and molecular beacons (bottom)

PNA oligomers can also be used to silence polymerase chain reaction (PCR) amplifications in single mutation analyses.[105] The PCR clamping method for the detection of point mutations is based on the ability of PNAs to bind more strongly to complementary nucleic acids combined with its inability to act as a primer for DNA polymerases. Targeting of the PNA oligomer against the primer binding site can block the formation of the PCR product. The procedure is so powerful that the expert choice of the primer length can allow discrimination of alleles which differ only in one base pair (single point mutation).[106-108]

LIMITING FACTORS FOR PNA APPLICATIONS

Despite the many appealing properties of PNA, there are limitations for the use of PNA as sequence specific binding agents that have not been resolved. The unique helix invasion process is highly dependent on salt concentration.[49,80,86,109,110] The pronounced inhibitory effect at moderate salt levels (>50mM NaCl) is due to the neutralization of the DNA backbone by monovalent cations that decreases DNA

breathing dynamics (transient openings of base pairs i.e. local unzipping).[84,111] Hence, even though PNA exhibits a high degree of sequence specificity, the rate of PNA strand invasion is too slow in high salt concentrations. Kurakin and coworkers, however, reported that the presence of a preformed PNA strand-displacement complex near the target facilitates the binding of the second PNA at physiologically relevant ionic strengths. Subsequent experimental results illustrated and support the hypothesis that creating distortions or openings in the DNA duplex will accelerate PNA binding dramatically in the presence of salts.

The efficacy of oligonucleotides in various biological applications as *in vitro* and *in vivo* reagents relies on the ability to traverse the cellular membrane, and for PNA, this requirement represents a major challenge.[112] Ideally, the intracellular transport should be accomplished by simple exogenous delivery. PNA oligomers, however, are poorly taken up this way and require additional means of transportation for efficient cellular uptake. Consequently, a great deal of effort has been invested into developing alternate methods for improved intracellular delivery. The most frequently used approaches include covalent attachment of cell-penetrating peptides, conjugation with targeting ligands, incorporation into liposomes and backbone modifications.[113-117] Another simple strategy for cellular delivery involves hybridization of PNA to a short DNA molecule followed by complexation with cationic lipid-based agents.[118] These agents deliver DNA inside the cell where PNA is released to carry out its task such as hybridization to mRNA. This method has been successfully used to deliver PNA to endothelial and different types of mammalian cells.[119] The ability of the aforementioned conjugates to move through the membrane presumably relies merely on their physical properties. Nevertheless, an instance of increased receptor-dependent delivery of PNA has also been reported, in which a D-peptide derived from the insulin-like growth factor has been attached to PNA.[120] However, this type of uptake only occurs in cells that express the specific receptor for insulin-like growth factor making this type of delivery specific to particular cells or tissue.[121]

Over the last decade, considerable effort has been devoted towards developing and improving different means to deliver PNA into cells, and of the many approaches described, the method of peptide conjugation has been extensively used. However, most

cell-penetrating peptides have amphipathic character or lead to formation of amphiphilic moieties via conjugation to PNA that destabilize or disrupt cellular membranes and are thus toxic to cells.[122] Recognizing the challenges, many strategies have been developed to design PNA analogues with improved solubility, increased ability for nonspecific binding to natural oligomers and increased transport across cellular membranes. The relatively simple structure of PNA allows for various modifications of the basic architecture as one way to improve its physical properties. The following sections summarize some of the recently reported PNA derivatives.

STRATEGIES FOR PNA MODIFICATIONS

From a chemical standpoint, PNA represents a hybrid of an oligonucleotide and a peptide, and this structural-and-functional duality of PNA determines its unique properties. Indeed, these molecules combine the inherent recognizing ability of nucleic acids with the flexibility and stability of proteins. Despite its desirable characteristics, the lack of charge and polar groups in the backbone reduces their performance in *in vivo* applications. Therefore, efforts have been made to synthesize new PNA analogues comprised of modifications at the nucleobase, nucleobase-backbone linker and backbone (Figure 11).

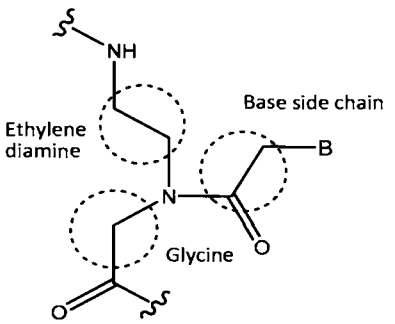

Figure 11. Potential sites for the modification of the PNA backbone

Nucleobase modifications

Inhibition of gene expression by antisense and antigene approaches relies on the stability of the oligonucleotide complex. In this context, the use of modified nucleobases represents an efficient way to enhance the biomolecular recognition and hybridization through increased hydrogen bonding and/or base stacking interactions. Several PNA analogues bearing nonstandard nucleobases have been studied, and some representative examples are shown in Figure 12. Haaima *et al.* reported that substitution of adenine with 2,6-aminopurine (D) increases the thermal stability of the PNA:DNA duplex.[123] This improvement is probably due to an increased affinity for thymine, since diaminopurine-thymine base pairing involves an additional hydrogen bond and extends the inter-base stacking.

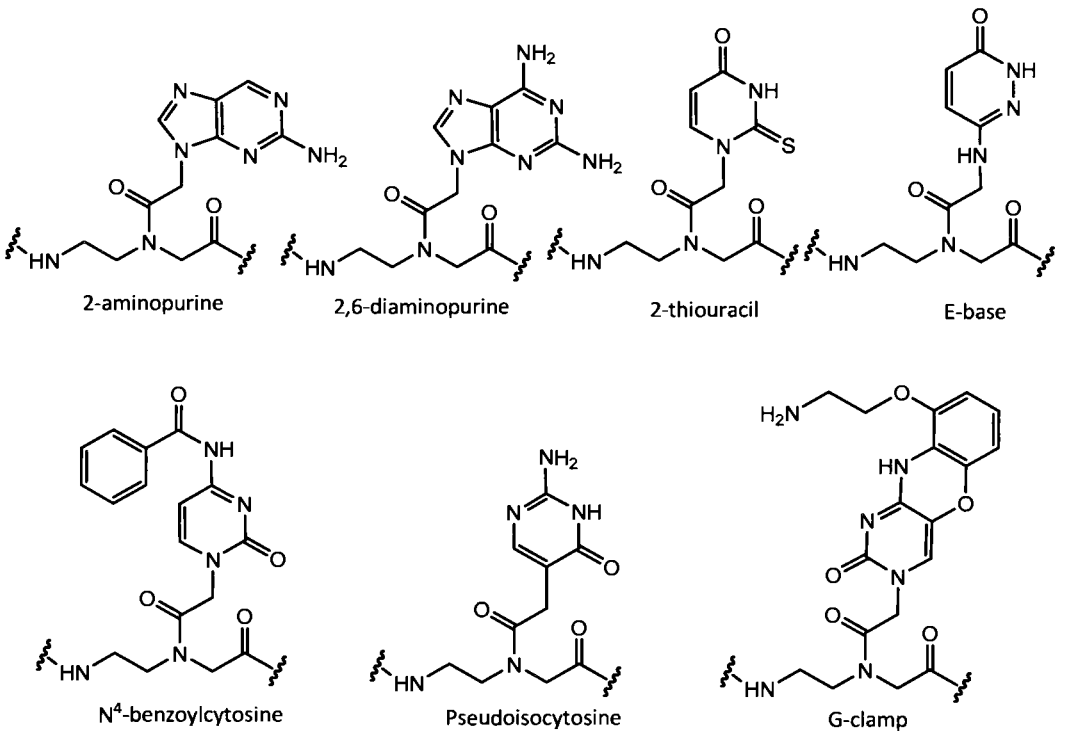

Figure 12. Examples of modified PNA nucleobase moieties

Moreover, it displayed greater mismatch discrimination compared to adenine. 2-Thiouracil along with 2,6-diaminopurine have been used in PNA-DNA recognition and have been shown to form stable complexes with DNA by so-called "double duplex invasion" (Figure 13).[124] Pseudoisocytosine is another example of a modified nucleobase that forms an extraordinarily stable triplex by mimicking a protonated cytosine at neutral or alkaline conditions.[43]

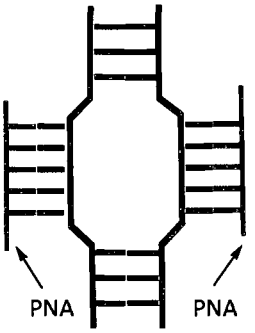

Figure 13. Double duplex invasion

Incorporation of N^4-benzoyl cytosine causes inhibition of triple helix formation due to steric interference,[125] whereas E-base proved to be a more useful substitution when forming triplexes.[126] The unique design of E-base allows for specific A-T base pair recognition in the major groove and has been found to bind more strongly to thymine than guanine. In fact, PNA_2:DNA triplex formation at neutral pH is restricted to purine rich sequences. Homopyrimidine strands rich in cytosine require an acidic environment to combine into a DNA_2:PNA complex that is of limited stability. In order to increase the stability of hybridized triplexes, pyrimidine bases that possess aromatic moieties with extended π-surface area for greater hydrophobic/stacking interactions were synthesized. Two groups worked separately on one of the most pronounced nucleobase analogues termed G-clamp.[127,128] This modification led to increased stabilization through the formation of extra hydrogen bonds as well as enhanced stacking due to a larger surface

area. Moreover, it maintained an excellent sequence discrimination ability. Modified nucleobases that are capable of selective binding and possess intrinsic fluorescence were used to study interactions between different classes of biopolymers. 2-Aminopurine is the first analogue to be prepared for this purpose, whereas thiazole orange exhibits a remarkable base stacking ability combined with a sensitivity for single base-pair mismatches.[129-132]

Nucleobase-backbone linker modifications

Free PNA oligomers are shown to be composed of an equilibrium mixture of *E* and *Z* rotamers about the linker amide. Upon hybridization to a complementary sequence of PNA, RNA or DNA, the amide bonds of PNA are mostly in the *Z* conformation. With the aim of understanding the role of the nucleobase-backbone linker in the binding process, new PNA analogues were synthesized in which a tertiary amide bond between the base and the backbone is replaced by an olefin (Figure 14).[133] It has been speculated that freezing the nucleobase-backbone bond in the correct conformation (*Z* or *E*) would result in improved affinity and specificity of binding. Experimental results using olefinic peptide nucleic acids (OPA) showed that both *E* and *Z* isomers can bind to complementary DNA, but with decreased thermal stability when compared to *aeg*PNA. *E*-OPA oligomers prefer the parallel binding mode over the antiparallel one, however, unlike *aeg*PNA, homopyrimidine OPAs are not able to form triplexes. Incorporation of a fluorine atom on the linker double bond (F-OPA) appears to stabilize the resulting complex with complementary DNA compared to unmodified *aeg*PNA.[134] F-OPA can also hybridize with DNA in a parallel fashion although with lower affinity. Like with OPAs, there appears to be a sequence dependence on the thermal stability of F-OPAs. Despite the observed destabilization of duplexes, locking the base in aconformation suitable for hybridization results in increased binding selectivity compared to *aeg*PNA.

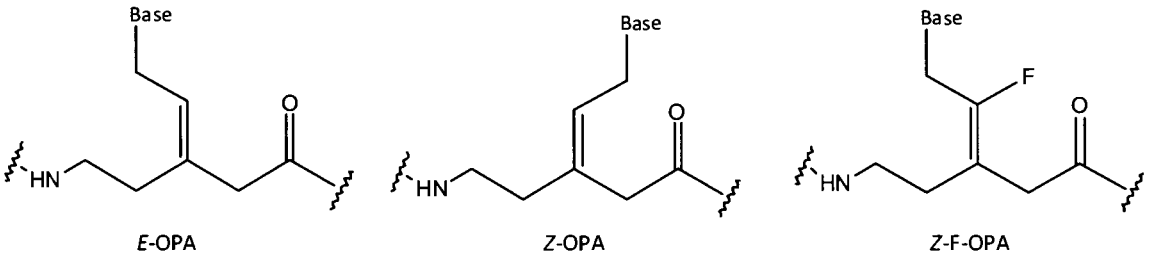

Figure 14. Olefinic peptide nucleic acids (OPA)

Backbone modifications

In addition to the aforementioned modifications, a large number of backbone derivatives were investigated in order to improve some of the limitations of the PNA structure and its use in therapeutic and diagnostic applications. The PNA backbone can be modified in several ways: (i) extension of the backbone by one or more atoms, (ii) rigidification of the backbone via conformational constraint in order to pre-organize the PNA and (iii) introduction of cationic functional groups.

In an attempt to understand if retaining the same number of bonds as in DNA influences hybridization properties, oligomers with an increased number of methylenes in the backbone were synthesized (Figure 15).[135-137] It was reasoned that in addition to imposing unfavorable steric or structural constraints, these changes would also increase the flexibility of the molecules. The binding properties of different analogues were studied by experiment using UV melting techniques. Extended backbone PNA oligomers resulted in decreased melting temperatures of PNA:DNA duplexes, demonstrating that correct inter-base distance is crucial for stable duplex formation. Modified oligomers with the correct number of bonds in the backbone, such as oxy-PNA, C-PNA and retro-inverse PNA, have decreased melting temperatures, but high sequence selectivity and improved solubility profiles over original PNA.[138-140] α-Helical PNA is another example of a PNA analogue, which has a true peptide backbone composed of Ser-Ala_2-Lys tetrapeptide units organized into an α-helix.[141] The formation by α-PNA of very stable

duplexes with DNA in a sequence-specific manner suggests that structural organization of the backbone is essential for hybridization.

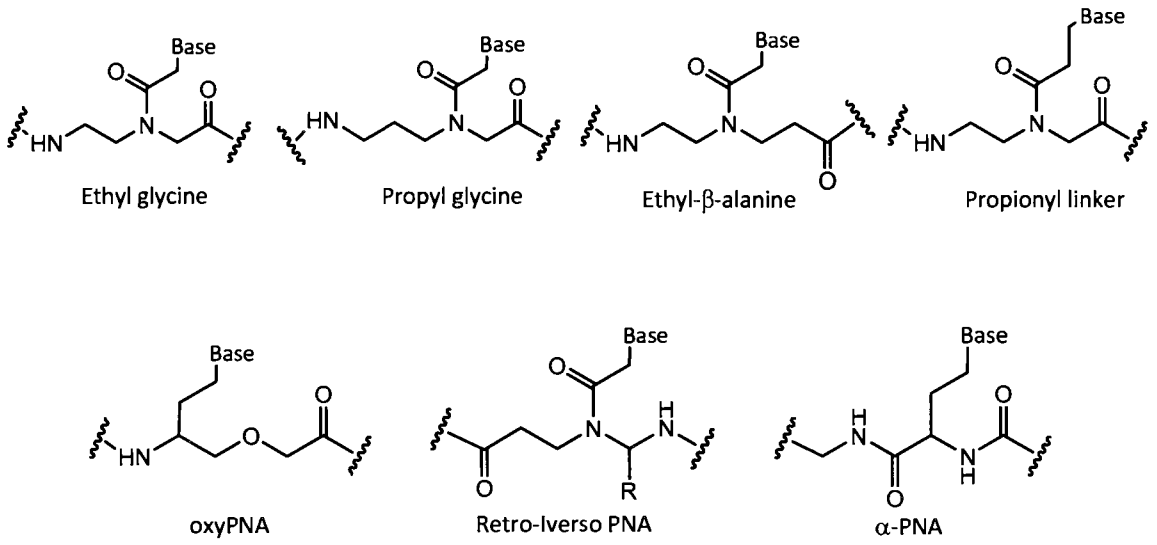

Figure 15. Examples of modified PNA backbone moieties

Introduction of charged moieties into the PNA backbone presents an important strategy that has led to PNA oligomers with improved physicochemical characteristics. Cationic groups, such as lysines, are usually used to increase solubility, decrease self-aggregation, improve cellular uptake and increase binding affinity with DNA/RNA. However, these groups can only be introduced at the terminal ends of the PNA chain. Several classes of cationic PNA analogues that bind to DNA/RNA strands with high affinity have been reported.[142-145] This stabilization is most likely due to the electrostatic interactions between the positively charged backbone of PNA and the negatively charged DNA backbone. However, detailed information on the structure of these analogues is needed to allow the design of new analogues with favorable features.

The relatively high binding affinity of PNAs toward natural oligonucleotides is attributed to the lack of electrostatic repulsion between the uncharged PNA backbone and the negatively charged sugar-phosphate backbone of DNA/RNA. The stability of duplex

structures in solution can be interpreted in terms of enthalpy-entropy compensation. The hybridization of the PNA strand with DNA/RNA increases enthalpy while decreasing entropy. The effect of introducing a modification can be explained in terms of a change in the free energy of the backbone torsions (ΔG_t) and the free energy related to base stacking (ΔG_s). The change in the base stacking free energy is dominated by an enthalpy change ($\Delta G_s \sim \Delta H_s$), whereas the change in the backbone torsion free energy is affected by both entropy and enthalpy. The single-stranded PNA, being acyclic, is conformationally more flexible in its different structural segments. Consequently, formation of the PNA:DNA/RNA complex is accompanied by conformational changes in the PNA in order to gain enthalpic advantage of hydrogen bonds and base-stacking interactions. This gain is accompanied by an undesirable entropy loss, and possible undesirable enthalpy loss due to increased torsional strain. The decrease in entropy upon hybrid formation is due to the formation of a highly ordered and fairly rigid duplex structure from two flexible and less ordered single strands. Therefore, constraining the single stranded PNA in a conformation identical to or close to that found in the hybrid should greatly reduce the entropy and enthalpy losses and increase the free energy of binding.

Any favorable structural pre-organization of PNA that facilitates binding should allow the strand to assume the preferred range of dihedral angles observed in PNA:DNA/RNA complexes. With this in mind, numerous attempts have been made to modify the aminoethylglycine PNA backbone through introducing alkyl and cyclic substituents to improve oligonucleotide selectivity and binding affinity.[146,147] The presence of rotamers around the tertiary amide bond interferes with the hybridization process. Thus, in order to hinder the rotation, the nucleobase was connected to a cyclic system. Examples of conformationally blocked PNA are represented by aminoprolyl (*ap*PNA),[148,149] aminoethylprolyl (*aep*PNA),[150-152] pyrrolidinone (*pyr*PNA),[153,154] piperidinone (*pip*PNA),[155] pyrrolidine,[156,157] cyclopropane (*cpr*PNA),[158] cyclopentyl (*cp*PNA)[159-164] and cyclohexyl PNA (*ch*PNA)[165-167](Figure 16). These monomers are obtained by bridging atoms in the backbone or atoms of the backbone and base-backbone linker. Interestingly, *ap*PNA, *aep*PNA and *pyr*PNA are all synthesized utilizing a hydroxyproline derivative as a precursor. In the 4-aminoproline derived PNA, the two

chiral centers at C2 and C3 result in two diastereoisomers. PNA molecules having a single chiral L-*trans*-4-aminoprolyl monomer bind to DNA with high affinity, whereas PNAs with the other three diastereoisomers do not stabilize PNA:DNA duplexes.[149] Additionally, the stereochemistry of the monomer affects the preference for parallel or antiparallel binding. Homochiral thymine *ap*PNA does not bind to DNA strands due to the high rigidity of the backbone, which results in structural incompatibility.[168] However, incorporation of L-*trans* aminoprolyl units alternating with aminoethyl glycine units has been found to improve binding to DNA.

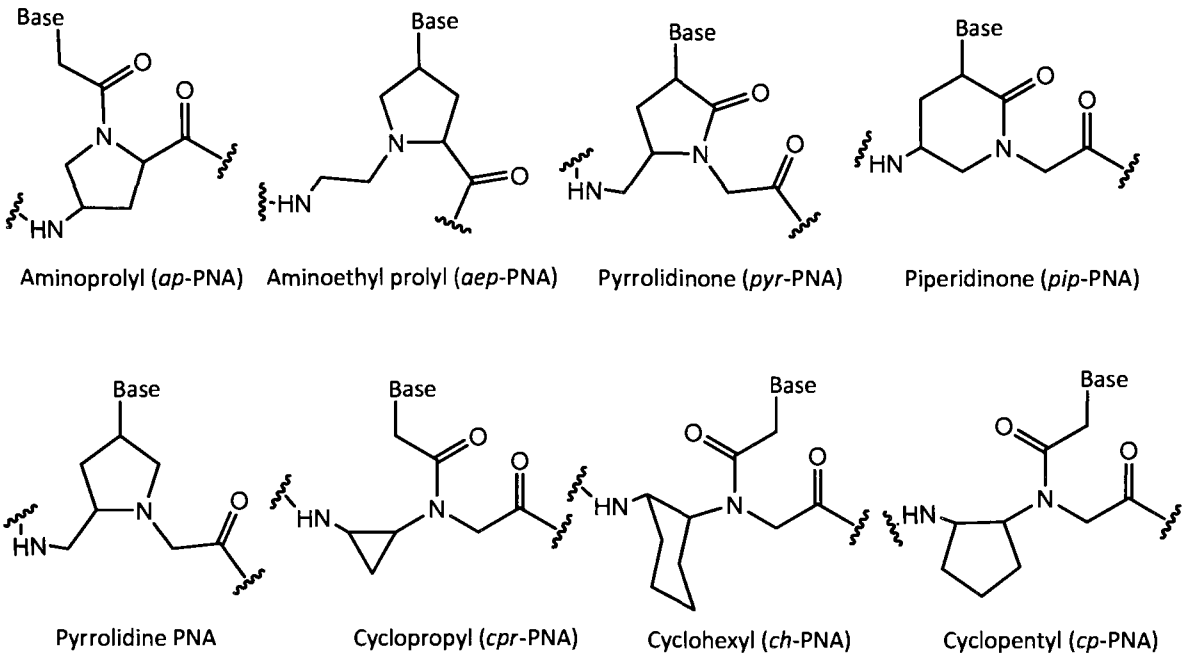

Figure 16. Examples of PNA backbone modifications with cyclic moieties

The replacement of the tertiary amide carbonyl on the backbone by a methylene group yields *aep*PNA that retains the flexibility of the aminoethyl segment, but restricts the rotamers through a methylene bridge between the glycine and the side chain.[150-152] In these derivatives, the nucleobase is attached directly to the positively charged pyrrolidine

ring. When hybridized to DNA, *aep*PNA has a higher affinity than *aeg*PNA, owing to the presence of the protonated ring nitrogen. In general, the antiparallel binding mode is preferred over the parallel one. Furthermore, *aep*PNA exhibits excellent sequence discrimination and forms *aep*PNA_2:DNA triplexes. However, under physiological conditions (pKa 6.8) the protonation can influence ring puckering and alter the conformation of certain nucleobases, thereby reducing duplex stability.

An alternative to pre-organizing the methylcarbonyl linker is to introduce a ring between the amide carbonyl and the backbone restoring the amide character of the pyrrolidine ring nitrogen (*pyr*PNA).[153] Of all the four stereoisomers of *pyr*PNA, the (3*S*,5R) isomer has been found to be the most efficient in binding to complementary RNA. It is also better able to discriminate between RNA and DNA. A slightly lowered thermal stability is observed when comparing the binding of (3*S*,5R) *pyr*PNA with RNA to that of unmodified PNA with RNA. Taken altogether, these results suggest that, of the four isomers, the (3*S*,5*R*) stereoisomer best approximates the conformation of PNA in a PNA:RNA double helix. In addition, the five-membered ring may not be the optimal conformational restraint modification, because it causes a reduced thermal stability when compared to *aeg*PNA.

A reduced version of *pyr*PNA called pyrrolidine PNA has also been synthesized by Nielsen and Vilaivan via two different methods.[155,169,170] Incorporation of a single (2*R*,4*S*) pyrrolidine monomer into an *aeg*PNA mixed sequence resulted in an increase in binding affinity to both DNA and RNA, and protonation of the nitrogen did not seem to affect the stability of the hybrids. On the contrary, the (2*S*,4*R*) isomers do not stabilize the formation of hybrids with DNA/RNA. However, a homoadenylate pyrrolidine decamer exhibited an increase in binding affinity compared to an isosequential unmodified PNA.

PNA variants containing six-membered structures received much attention due to their unique conformational preferences and sequence recognition abilities. Molecular modeling studies suggested that conformations of the rigid six-membered piperidinone PNA would allow formation of hybridization-competent oligomers. However, significant destabilization of the piperidinone PNA:DNA duplexes was observed compared to *aeg*PNA.[155] The C1 of the glycyl unit and the C6 of the nucleobase linker are bridged by

a two-carbon ethylene to form the aminoethylpipecolyl PNA (*aepip*PNA).[171,172] This chiral six-membered analogue has been found to stabilize PNA_2:DNA triplexes.

Although there are many different PNAs employing cyclic constraints, monomers with carbocyclic backbone modifications of the C2-C3 bond represent a particularly promising group of cyclic PNAs. The ethylene diamine portion of the PNA backbone is the most flexible part; therefore, restricting the bond rotation will reduce the flexibility and pre-organize PNA oligomers for binding. Moreover, when nucleobases are directly attached to the ring of the cyclic PNA, they constrain the ring in specific conformations leading to distinct differences in binding properties. The earliest example of backbone modification was reported by Nielsen and co-workers who replaced the flexible aminoethyl segment at the Cβ and Cγ positions of the PNA backbone with a 1,2-diaminocyclohexyl moiety. Both enantiomers of *trans* cyclohexyl PNA (*ch*PNA) were inserted into mixed base sequences and their hybridization to DNA/RNA complementary strands was analyzed. Unfortunately, neither the (*S,S*) nor the (*R,R*) isomer of *ch*PNA improved the stability of the complexes formed, however, introduction of (*R,R*) isomers resulted in a dramatic destabilization of the PNA:DNA or RNA complexes. It should be noted that the (*S,S*) enantiomer decreased entropy as expected, but this was compensated for by a reduced enthalpy gain.[165] Following Nielsen's study, Ganesh and Kumar reported synthesis of *cis*-(*1R, 2S*) and (*1S, 2R*)-cyclohexylthyminyl PNA monomers.[173,174] The *cis* cyclohexane modification reduced the stability of the triplexes and duplexes of poly-T PNA.[175] PNA containing the (*R,S*) isomer exhibited higher hybrid stability in the PNA:DNA complexes than the (*S,R*) isomer. In the case of RNA complexes, *cis*-(*R,S*)-*ch*PNA monomers seem to be better tolerated in the duplex structure than (*S,R*).[176] However, when incorporated into mixed base sequences, *cis*-*ch*PNAs showed high selectivity for RNA over DNA. While this structural discrimination emphasizes the potential of PNA for antisense applications, the *cis* cyclohexane modification does not induce optimal conformations for DNA hybridization.

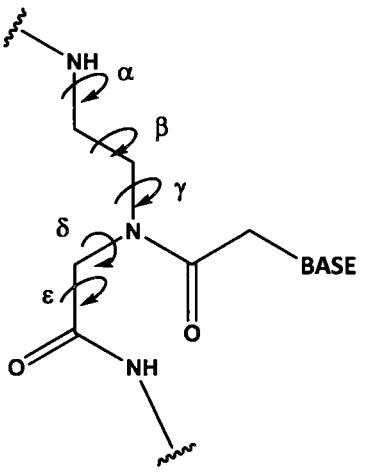

Figure 17. Definition of the PNA backbone dihedral angles

Analysis of the hybrid structures revealed that the dihedral angle values were restricted to a specific range in the PNA:DNA/RNA duplexes (Figure 17 and Table 2).[51,177] In an attempt to adjust the dihedral angles for improved hybridization, Ganesh and coworkers replaced the cyclohexyl unit with cyclopentane and synthesized *cis-(1R,2S)* and *(1S,2R)*-cyclopentyl PNA monomers (*cp*PNA).[159,160] *Cp*PNAs carry a five-membered ring bridging the Cβ and Cγ of the aminoethyl moiety. The substituted cyclopentane ring system is an interesting modification due to its inherent flexibility. Cyclopentane has two preferred conformations: half-chair and envelope. These conformations have several energy minima and the barrier for interconversion is low, which implies that cyclopentane modified PNA should be flexible enough to adopt the necessary dihedral angles for stable complex formation. On the contrary, *trans*-cyclohexane PNAs have two chair conformations: diaxial and diequatorial. The diaxial conformation is unfavored and makes the cyclohexane a rigid modification that forbids structural readjustments during hybridization.

Table 2. Backbone dihedral angles for the modified PNA structures

Compound	α	β	γ	δ
PNA/PNA[a]	-107	62	74	96
DNA/DNA[b]	71	-165	49	110
RNA/RNA[a]	-81	174	66	95
PNA/DNA[b]	105	141	78	139
PNA/RNA[b]	170	67	79	84
*(S,R)-ch*PNA[a,c]	128	-63	76	119
*(R,S)-ch*PNA[a,c]	-129	66	-78	-119
*(S,R)-cp*PNA[a,c]	84	-24	86	90
*(R-S)-cp*PNA[a,c]	-84	25	-86	-90

Data obtained from [a]X-ray or [b]NMR. [c]Monomer crystal structures

Cis-cyclopentyl modified PNAs exhibit higher affinity towards RNA than DNA: the *(R,S)-cp*PNA enantiomer formed higher stability duplexes with DNA than the (S,R)-*cp*PNA monomer. For polyribonucleotides (rA) the reverse trend has been observed. In all cases the PNA:RNA hybrids are more stable than the corresponding PNA:DNA duplexes. In comparison to cis-*ch*PNA, cis-*cp*PNA oligomers had much higher T_m values; however, stereochemical discrimination of DNA and RNA was expressed better by the *ch*PNA oligomers.[176]

Based on molecular modeling studies and published NMR data, the *(S,S)*-cyclopentadiamine ring was used for conformational restraint of the C2-C3 dihedral angle of the PNA backbone.[164] Appella *et al.* first reported the incorporation of *trans-(S,S)-cp*PNA monomers into polythymine heptamers through standard solid-phase synthesis and tested the ability of the strands to hybridize with DNA utilizing UV thermal denaturation techniques.[178,179] The *(S,S)* enantiomer was chosen because this stereochemistry can promote right-handed helix formation. On the contrary, introduction of the *(R,R)* monomer had a serious adverse effect on duplex stability. The *cp*PNA monomers were also incorporated into a number of PNA strands containing a mixed-base sequence that was extensively studied by Nielsen.[165] The effect of the addition of the *trans-cp*PNA modification on the binding affinity to complementary DNA is summarized in Table 3. The addition of a single *trans-cp*PNA monomer increased the T_m by

approximately 5°C. The effect was independent of the position in the PNA strand and the nucleobase attached to the modified monomer. The influence of the cyclopentane modification was also additive, in that the T_m continued to increase as multiple residues were incorporated.[180] Moreover, incorporation of a single cyclopentane monomer showed an improvement for discriminating against mismatches. The results of thermal denaturation studies also provided an excellent basis for the development of PNA microarrays.[181]

Table 3. T_m data for PNA:DNA duplexes

PNA sequence[a]	T_m(°C)
GTAGATCACT-Lys	48.9
GTAGAT*CACT-Lys	54.9
GTAGATCA*CT-Lys	54.5
GTAGATC*ACT-Lys	54.2
GTAGAT*C*ACT-Lys	60.2
GTAGA*T*C*ACT-Lys	64.4

*= *trans-cp* residue, [a] Cyclopentane stereochemistry is (S,S)

Summary

The design of *cp*PNA is an outcome of optimization of the dihedral angles that constrain the PNA backbone for differential DNA/RNA binding and discrimination via pre-organization. The flexible *trans*-cyclopentane modification, which changes conformation easily, allows reorganization of the ring puckering, which leads to high affinity binding to both DNA and RNA, but does not improve binding selectivity. The goal of our research is to explore and analyze the structure and dynamics of rigidified PNAs containing one or multiple *trans* cyclopentyl groups in the backbone in order to determine distinctive structural features and preferences. We advance the hypothesis that the restriction of the *aeg*PNA backbone dihedral angles will lead to increased binding affinity and oligonucleotide selectivity based on correct PNA pre-organization. Due to their properties *cp*PNAs can provide deep insight into the interplay between the various

interactions which stabilize oligonucleotide systems. To exploit this advantage, however, more detailed information about their geometrical arrangements and stability are needed in combination with a critical test of the torsion angles. It is the goal of the present work to provide such information with the help of theoretical approaches.

CHAPTER II

THEORETICAL BACKGROUND

The basis of computational chemistry relies on the assumption that chemical behavior can be predicted by solving mathematical equations that describe the chemical system. The main goal of computational methods is the consistent and accurate investigation of the chemical properties and the dynamical behavior of molecular systems using theoretical approaches. Theoretical investigations can be carried out at two different levels, quantum mechanical[182-185] and empirical (molecular mechanics).[186] Quantum mechanical methods deal with the electronic structure of molecules, and such computational results can reach the accuracy of experimental data. However, these calculations are time-consuming and limited to only small systems. On the other hand, for complex biomolecules such as nucleic acids and proteins, molecular mechanical or force field methods are used as an effective alternative. This approach, however, suffers from the limitation that experimental or *ab initio* data is required for parameterization; information that is not generally available for novel types of compounds. As a consequence, parameterization of empirical force fields to accurately describe these systems becomes an important concern, and is the focus of the work presented in this chapter.

EMPIRICAL FORCE FIELDS

Molecular mechanics (MM) methods are developed by empirically fitting experimental parameters to classical mechanical models of molecules. The simplest representation of an atom is a sphere of fixed radius, with a mass and charge located at the center of the sphere. MM calculations model molecules with a set of classical potential energy functions that define the force field.[187] These equations describe the dependence of a molecule's energy on the coordinates of its constituent atoms.[188,189] The nature of these functions is chosen to provide a good approximation to the real systems.

At the same time the calculations require little computational effort. The molecules are modeled as a collection of atoms of different sizes and softness and bonds of different stiffness.[190] The expression for the total potential energy (E) of a molecule has contributions from each of the major internal valence coordinates (terms for distortion of bond lengths E_b, bond angles E_a, and torsion angles E_t), and intra- and intermolecular non-bonded interactions (van der Waals forces E_{vdW}, and electrostatic interactions E_{el}). The total potential energy is then given by the contribution of the different terms:

$$E = E_b + E_a + E_t + E_{vdW} + E_{el} \quad (1)$$

Force fields employ various functional forms for the terms in equation (1) as well as sets of parameters for bonding and non-bonding interactions. Each force field was developed to solve specific problems, thus there is no single "best" force field for the accurate description of all molecular systems.[191] Methods based on empirical force fields can be used to calculate both the static and dynamical properties of molecules. The first group includes minimum energy structures, conformational energies, rotational barriers, vibrational frequencies, etc. The evolution of a molecular structure in time (trajectories) and time averages of structural and thermodynamic properties can be calculated using molecular dynamics simulations. Force field methods, however, are unable to describe systems and processes where an extensive rearrangement of the electron density takes place (bond breaking events, excited states, etc.) Also, most currently available force fields are pairwise-additive, thus limiting their accuracy for systems where polarization of the molecular charge distribution is important.[192,193]

There are several MM force fields available for performing simulations of biomolecules including AMBER,[194] GROMOS[195] and others.[196,197] CHARMM[198,199] (acronym for Chemistry at HARvard Macromolecular Mechanics) is a molecular mechanics and dynamics program developed in the laboratory of Martin Karplus, and is one of the most widely used force fields in the simulation of large biomolecules. It possesses the necessary parameters for the common nucleosides and amino acids. The CHARMM force field includes the following terms (Figure 18):

$$E_{total} = \sum_{bonds} K_r(r - r_0)^2 + \sum_{angles} K_\theta(\theta - \theta_0)^2 +$$

$$\sum_{dihedrals} K_\phi[1 + \cos(n\omega - \gamma)] + \sum_{i<j}\left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}}\right) + \sum_{i<j}\frac{q_i q_j}{\varepsilon r_{ij}} \quad (2)$$

The first MM energy component in the form of a simple harmonic potential, is the stretching term that describes the bonds. The stretching energy increases when the bond is deformed (compressed or elongated) from its equilibrium position. It obeys Hooke's law, which considers both the magnitude of deformation and the stretch constant (stiffness) of the bond. In order to do this in MM, an equilibrium bond length, r_0, needs to be determined from experiment or *ab initio* calculations. A second parameter, the quadratic stretch constant, K_r, must be assigned to each bond to model the stiffness (or strength) of the bond. This factor is important for primarily determining the vibrational frequency of a bond, although it will indirectly affect the bond length. Besides bond stretching, angle bending is an important feature of polyatomic systems and can be modeled classically in an analogous manner. It is also described by a simple harmonic function, where K_θ is the force constant for bending the angle formed by three atoms bonded to each other and θ_0 is the equilibrium angle for the arrangement.

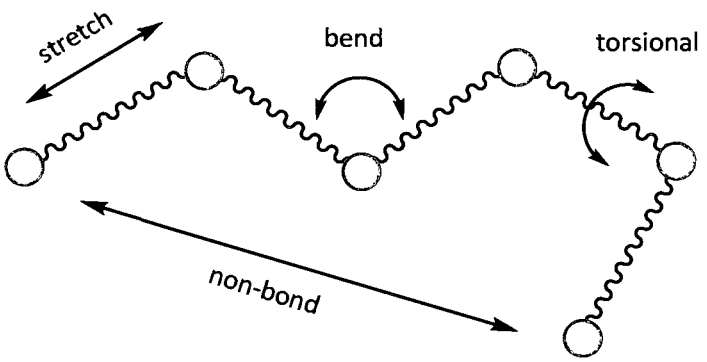

Figure 18. Schematic representation of force field energy terms. Figure adapted from Jensen, F. *Introduction to Computational Chemistry*, John Wiley, 2007

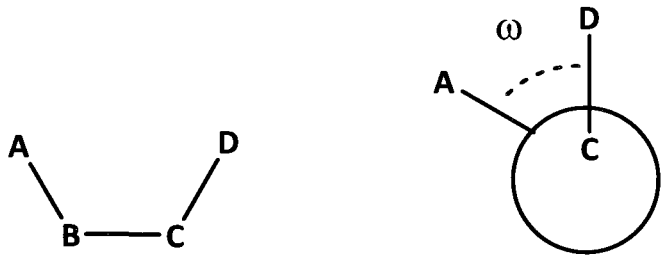

Figure 19. Atom orientation in the torsion angle

The next term in Equation 2 is the torsion angle potential, which represents the change in energy due to internal bond rotation. For four atoms bound in sequence *ABCD*, the clockwise angle from *A* to *D* while looking down the *CB* bond represents the torsion angle, ω, between atoms *A* and *D* (Figure 19). Although this is also commonly referred to as a dihedral angle, it should be noted that a dihedral is actually the complement of a torsion angle with a value of 180°- ω. The torsion potential is assumed to be periodic and is modeled with a truncated Fourier series. In the torsion angle term, K_ϕ is the constant associated with the height of the torsional barrier, n is the periodicity, which dictates the number of minima a bond of a given chemical type has as it rotates through 360°, ω is the torsion angle, and γ is an angle that determines the phase of the function. The last term in Equation 2 contains two potentials describing the non-bonded interactions. The Lennard-Jones potential accounts for both the attractive (negative term which is proportional to r^{-6}) and repulsive interactions (positive term proportional to r^{-12}). A_{ij} and B_{ij} are the associated parameters which depend on the pair of atoms *i* and *j*. In addition to potential energy steric interactions, the presence of charges and polar groups requires incorporation of electrostatic interaction terms. Since the electrostatic charges are relatively constant for a particular functional group, each atom is assigned a partial charge. The potential energy of interaction between the two terms is then written in the form of a standard Coulomb interaction, where q_i and q_j are partial atomic charges, r_{ij} is the distance between the charges, and ε is the dielectric constant of the medium.

As mentioned above, these potential functions define the force field and the total potential energy of the system is defined as the sum of all the contributing energy terms. However, the quality of the MM force field depends strongly on the adjustable parameters. In order to obtain meaningful and reliable results, the parameters must reproduce the desired structural and energetic properties. This goal is achieved by fitting them to experimentally determined data or high level *ab initio* calculations on small molecules with similar chemical properties. For example, equilibrium bond and angle values can be taken from X-ray structural data and the corresponding force constants determined from vibrational spectral data. This kind of approach allows for parameter interpolation, i.e. force field parameters for a particular class of compounds or functional groups can be used to calculate the structures of other similar classes of compounds. However, the CHARMM force field has been parameterized only for proteins, carbohydrates and naturally occurring nucleic acids. To the best of our knowledge, only two sets of parameters for regular *aeg*PNA have been reported.[58,59] Therefore, all parameters related to the cyclopentane ring modification needed to be assigned in the present work. Since high level quantum calculations have an accuracy comparable with experiments, the parameters in this study were determined solely from first principle calculations of molecular geometries and their energies.

QUANTUM CHEMICAL METHODS

Electronic structure methods avoid many limitations of the empirical force fields, but their major restriction is the maximum size of the molecular system that can be treated (about 100 atoms). These techniques are based on the concept that electrons are not localized at particular points in space but behave like waves. The probability of finding an electron within a region of space, or probability density, is given by Equation 3, where Ψ is the electronic wave function. With Ψ normalized, the probability of finding the particle in all space must be 1. Although the wave function itself does not have a physical meaning, it is postulated to contain all the information about the system. The wavefunction depends on both the position of the particle and its spin states.

$$P = \int_a^b \Psi^*(x)\,\Psi(x)\,dx \tag{3}$$

In quantum mechanical or *ab initio* calculations, the Schrodinger differential wave equation (Equation 4) for an electronic system is formulated and solved by approximations.[200,201]

$$\mathrm{H}\Psi = E\Psi \tag{4}$$

Central to this equation is the linear Hamiltonian operator $\hat{H}$, which acts upon the wave function to give the total energy of the system. The Hamiltonian is made up of the sum of the kinetic and potential energy contributions of the nuclei and electrons in the system. The E's are the discrete eigenvalues of the different stationary states. The solutions (the wavefunctions, Ψ) of the time-independent Schrodinger equation correspond to various stationary states of the system (the solutions are called stationary-state wave functions because they are independent of time); the lowest energy solution is called the ground state. An accurate solution to this equation can be obtained for the hydrogen atom only; therefore, for molecular systems, various mathematical approximations are introduced. The Born-Oppenheimer approximation assumes that the electronic and nuclear wave functions can be separated and treated independently due to the fact that the nuclei are much heavier than electrons and, therefore, move more slowly. Specifically, one is assuming that the motion of an electron is instantaneous relative to the fixed nuclei. Quantitatively, the Born-Oppenheimer approximation may be formulated by writing down the Schrodinger equation for electrons in the field of fixed nuclei, i.e. $\hat{T}^{nuclei} = 0$,

$$\hat{H} = \hat{T}^{elec} + \hat{V} \tag{5}$$

where $\hat{T}^{elec}$ is the electronic kinetic energy, and $\hat{V}$ is the coulomb potential energy.

The resulting molecular Hamiltonian is a sum of the electronic and nuclear kinetic terms, plus the potential applied to the electrons and nuclei. Then, the wave function

becomes a product of electronic and nuclear components. The final goal of most quantum chemistry calculations is to determine the molecular wave function, which is accurate enough to calculate the desired properties to an acceptable degree of uncertainty. The multi-electron Hamiltonian is a difficult entity with which to work because of the mathematical complexity involved in the electron-electron repulsion term. Using the orbital approximation, in which each electron occupies an orbital, the wave function is separated into a product of several one-electron functions:

$$\Psi = \phi_1\phi_2\phi_3 \ldots \phi_n \tag{6}$$

The wave function must remain antisymmetric with respect to exchange of electrons due to the Pauli Exclusion Principle. This condition is a consequence of the fact that electrons are *fermions* having ½ spins. In order to give a correct description of the state of a many-electron atom, the correct wave function describing the state of the electron must be the product of spatial orbital functions and spin functions. Then, the many-electron molecular wave function can be written as a Slater determinant,[202] with each column representing a one spin orbital and each row representing one electron. Since exchange of any two rows in a determinant changes its sign (multiplication by -1), the Slater determinant satisfies the antisymmetry requirements of the wavefunction. Moreover, the determinant, which mixes all electrons and orbitals, can be varied easily to determine the wave function that provides the lowest energy eigenvalues.

By introducing a set of known spatial functions, the Schrodinger equation for the many-electron problem can be transformed into a set of equations and solved by matrix methods. Each of the molecular orbital functions in Equation 6 is expressed as a linear combination of atomic orbital functions:

$$\phi_i = \sum_{k=1}^{n} C_{ik}X_k \tag{7}$$

Here ϕ_i is a molecular orbital and n is the number of the atomic orbital (basis set) function X_k, which can be selected to be a Slater-type, a Gaussian or any other function. The expansion coefficients C_{ik} are variational parameters. In theory, an accurate orbital

description is attained using an infinite number of basis functions. However, this type of representation is impossible for computers. Therefore, practical concerns limit the description of a molecular system to a finite number of basis functions. Two types of atomic basis functions received widespread use. Slater-type orbitals (STO) provide a very good description of the orbital space, however, since they are computationally too expensive, their use in practical calculations is limited. As an alternative, Gaussian-type atomic functions have been commonly used for orbital approximation. The advantage to using these functions is that all of the integrals in the computations can be evaluated explicitly without the use of numerical integration. However, it should be noted that they decay faster at large $\boldsymbol{r}$ (distance from the nucleus) and, therefore, give poor representation of atomic orbitals. As a consequence, several Gaussian functions (or primitives) are needed to adequately describe an orbital. Primitives are multiplied by an angular function to give orbitals the proper symmetry and orientation.[203]

Two popular basis sets that are used in *ab initio* calculations are the minimal basis set and the split-valence basis set. Minimal basis sets, such as STO-nG, are computationally inexpensive, but may not give reliable results. The most commonly used basis sets are the split-valence basis sets developed by John Pople's group.[204] These basis sets have a common notation of *KMN*-G, where *K* denotes the number of primitive gaussians for the core orbital basis function; *M* and *N* indicate the linear combination of primitive gaussians for the valence orbitals. For biological applications, which deal predominantly with first-row elements, satisfactory results can be obtained using a Pople-style split-valence basis set of the form 6-31G*. This basis set uses two basis functions with different orbital exponents (composed of three and one primitive, respectively) to represent each valence atomic orbital, and one contracted function (composed of six primitives) for inner-shell atomic orbitals. A single "*" indicates the use of one set of polarization functions on non-hydrogen atoms. A polarization function is an orbital with an angular momentum quantum number (l) greater than necessary (specifically three p-type functions on hydrogen and five or six d-type functions on first row atoms such as carbon and nitrogen). Use of polarization functions is important for the accurate representation of molecular electron density distribution, that is, the orbital polarization. Additionally, diffuse functions (s or p type) indicated by a "+" symbol may

be employed to allow electron density at a significant distance from the nuclear center. These functions are most useful for anions, molecules with lone pairs or systems where long-range electronic interactions are of interest.[205]

Ab initio methods solve the Schrodinger equation and solely depend on the fundamentals of quantum mechanics. In principle, they can be applied to any geometry, but for systems with more than two electrons it leads to the *N*-body problem, and the computational cost scales as a factor of *N* factorially. Therefore, these methods are only realistic for smaller molecular systems with less than 100 atoms.[206] There are two major methods for dealing with chemical bonding: Molecular Orbital (MO) theory and Valence Bond (VB) theory. The current computational approaches in quantum chemistry are based almost entirely on MO theory. The simplest type of *ab initio* calculation is the Hartree-Fock method (HF), in which an electron is considered to move in the field of the electron cloud caused by other electrons. This reduces the many electron equation to a set of one-electron HF equations (Equation 8),

$$F_i \phi_i = \varepsilon_i \phi_i \tag{8}$$

where ϕ_i is the wavefunction, ε_i is the energy of the orbital, and F_i is the Fock operator. When an HF calculation is started, neither the F_i nor ϕ_i are known. An initial guess is made for the ϕ_i, which in turn allows calculation of the F_i. Hartree utilized the notion that the wave function may be optimized by an iterative process in response to the influence of the total electric field potential. Since the solution of each HF equation affects the overall probability distribution, the equations are solved until convergence in electron density and energy is achieved. Therefore, the HF method is called a self-consistent field (SCF) calculation. This method uses the variational principle; therefore, the calculated energy will be equal to or greater than the true energy of the system.

HF calculations are perhaps the most common *ab initio* method used and it is important to understand their limitations. There are two approximations made in the method that lead to an error in the calculated energies. First, the Schrodinger equation does not take into account relativistic effects, and these play an important role as the number of core electrons increase. An assumption is made that the electrons have the

same mass regardless of how fast they travel. For heavy atoms, the core electrons begin traveling at speeds approaching the speed of light and their masses become non-negligible. However, since most biological applications only involve first and second row elements, the relativistic effects are of minor significance.

Despite its utility, the Hartree-Fock theory has the limitation that the interelectronic potential is an averaged field approximation; it calculates the electron-electron repulsion of electron **'a'** with regards to the average field of electron **'b'**. As a consequence, it does not take into account the fact that in reality the motions of electrons **'a'** and **'b'** are correlated and the probability of finding electrons **'a'** and **'b'** on the same side of the nuclei is less than that of finding them on opposite sides. When considering conformational energies, errors due to dynamical correlation effects become of considerable importance since they can vary strongly with local changes in angles and torsions. There exist a number of techniques that improve upon this lack and include the calculation of electron correlation, or explicit instantaneous interactions between electrons (start from HF wave function and then correct for the correlation). The most common routine is based on many-body perturbation theory. The *ab initio* calculations done in this study utilize second-order Møller-Plesset perturbation (MP2) in which two electrons are excited from occupied orbitals to virtual ones.[207] If relativistic contributions are not included, the correlation energy is the difference between the exact energy and the HF energy. In MP2 calculations, electron correlation is added as a perturbation to the calculated HF ground-state wave function. In MP theory, the Hamiltonian operator of a molecule is the sum of the HF Hamiltonian (zero-order operator) and a "perturbed" Hamiltonian that is modified by an arbitrary λ parameter. Hence, the unknown eigenvalues and eigenfunctions of the perturbed system are related to the known eigenvalues and eigenfunctions of the unperturbed system. The perturbation $\hat{H}^{(1)}$ is the difference between the true electronic Hamiltonian and the zero-order operator.

$$\hat{H} = \hat{H}^{(0)} + \lambda\hat{H}^{(1)} \tag{9}$$

Since the perturbed and unperturbed systems are similar and the perturbed wave function and energy are continuous functions of the variable parameter λ, it is assumed

that both can be expanded in a Taylor series to include first and second order corrections. The series are assumed to converge for $\lambda = 1$. The sum of the unperturbed and first order energies is equal to the HF energy, and the second-order correction introduces electron correlation. The second-order term has the form:

$$E^{(2)} = \sum_{t<u}^{unocc} \sum_{s<r}^{occ} \frac{|\langle tu||rs\rangle|^2}{(\varepsilon_r+\varepsilon_s-\varepsilon_t-\varepsilon_u)} \quad (10)$$

where $\langle tu||rs\rangle$ is the two-electron integral over the particular MOs, $\varepsilon_{s,r}$ and $\varepsilon_{t,u}$ are the orbital energies of the occupied (r,s) and unoccupied (t,u) MOs, respectively. It has been shown that the main contribution (>90 %) to the correlation energy of an isolated system comes from the doubly excited states, third and higher order corrections are typically an order of magnitude smaller than those of the second order. This result suggests that the expansion can be truncated. The higher orders of Møller-Plesset theory perturbation are denoted as MP*n*, where *n* is the order of perturbation. The MP1 energy is the same as the Hartree-Fock energy; MP2 includes the effects of double excitation and is the most practical treatment for electron correlation.[208,209] Third, fourth, and higher orders of perturbation are derived similarly to the second-order. Although the use of energy corrections of the third or fourth-order from MP3 and MP4 calculations may improve the quality of the calculations, they require more CPU time and become extremely computationally expensive.

Geometry optimizations and frequency calculations

The potential energy function, derived either from empirical methods or from solving the electronic Schrödinger equation, can be used to calculate relative energies, equilibrium geometries and vibrational frequencies. In principle, the function can be applied to any geometry in the ground or excited state on the potential energy surface. Geometry optimization is aimed at finding stationary points of the function where the first derivatives are zero.[210] In the majority of cases, it is of interest to find a configuration that corresponds to a minimum energy or ground-state structure, i.e. all the second derivatives are positive. In some cases, the desired structures are first-order saddle

points (transition structures), i.e. the second derivative is negative in one direction and positive in all other directions. Most optimization methods determine the nearest stationary point, but a multi-dimensional function may contain a large number of stationary points of the same kind (local minima). The task of finding the global energy minimum in an energy landscape with many minima separated by high energy barriers is not trivial.

Most approaches to the multiple minima problem that have been developed so far are based on gradient techniques.[211] All commonly used methods assume that the first derivative of the function with respect to all variables, the gradient **g**, can be calculated analytically. The gradient vector points in the direction of the largest increase in the function, therefore, the function value can be lowered by following in the direction opposite the gradient (steepest descent). It is also possible to arrive at a multi-dimensional local energy minimum by performing function evaluations along the line that is conjugate to the previous search direction, i.e. choosing gradients conjugate to each other. In addition to the first derivative techniques, some methods also use the second derivative matrix, the Hessian $\mathbf{H}_i$. Analytical gradients are computed at each point, and the matrix is updated based on the gradients from previous points. At the energy minimum, the diagonal eigenvalues of the Hessian matrix are all positive, and the step direction is opposite to the gradient direction. However, if one of the Hessian eigenvalues is negative, the step direction is along the gradient component (function increases), and thus the optimization may end up at a stationary point with one negative Hessian eigenvalue, i.e. a first-order saddle point. For successful optimization, the quality of the initial Hessian is very important. Therefore, calculation of an exact Hessian (rather than an approximate Hessian) at the first point often gives better convergence. Usually, gradient optimizations allow the intra- and intermolecular degrees of freedom to relax simultaneously so that a fully optimized structure can be obtained. However, subsequent evaluation of the vibrational frequencies is needed to verify the nature of the optimized structure. Stationary points are identified as local minima by the absence of imaginary frequencies, whereas transition structures are recognized by the presence of one imaginary frequency.

Harmonic vibrational frequencies are calculated from the analytical second derivatives of the energy with respect to the nuclear coordinates, and to be meaningful,

they must be evaluated at an equilibrium geometry. It is important to remember that both the optimized structure and the shape of the energy surface depend on the theoretical method and basis set used. Changes in frequencies are partly due to changes in the geometry, since the force constant decreases with increasing bond length. Moreover, *ab initio* calculations determine bond lengths at the theoretical bottom of the atomic interaction well (absolute zero) and do not include the effect of zero-point vibrational motion on bond length. As a result, the calculated bond lengths are generally shorter than experimentally determined values. One source of error in the vibrational frequencies is from the harmonic approximation used in their determination.[212-214] In the harmonic potential, energy levels are evenly spaced. Due to anharmonicity in the potential, the vibrational levels are actually condensing, which leads to an overestimation of the vibrational frequencies. Calculated HF frequencies are usually higher by about 7-10% relative to the experimental values. This overestimation is also due to the incorrect dissociation and the bond lengths being too short. The inclusion of electron correlation normally lowers the force constants (i.e. frequencies decrease) since the correlation energy increases as a function of bond length. In order to partly compensate for such systematic errors, calculated frequencies are often scaled by factors derived from benchmarking calculations.

PARAMETERIZATION STRATEGY

The success of empirical molecular modeling is critically dependent on both the potential functions and the accuracy and continued refinement of the parameters used. To obtain reliable results, it is crucial to develop force field parameters that correctly describe molecular systems and adequately reproduce desired energetic and structural properties. Developing such parameters can be a formidable task for various reasons where the inadequacy of the experimental data and the contradictory nature of experimental observations (more than one value exists for some experimentally determined properties) are just two examples. Therefore, theoretical results from *ab initio* calculations are increasingly used in parameter development. The quality of a molecular

mechanics calculation performed with a particular force field is based on the methods and target data used to optimize the parameters. Biomolecules, such as nucleic acids and proteins, are made up of smaller organic subunits that are used as the initial targets for data collection in force field development.

The mathematical terms in the CHARMM force field that describe the intra- and intermolecular forces in biomolecules are relatively simple; the intermolecular component consists of electrostatic and vdW terms and the intramolecular component includes terms for bond stretching, angle bending and torsional energies. The parameters needed to extend the force field for a particular class of compounds are generally optimized using different sources. The geometry parameters such as standard bond lengths and bond angles are often optimized to reproduce gas-phase geometries of simple model compounds obtained from quantum mechanical calculations, electron diffraction, or microwave experiments. The internal force constants used for stretch and bend deformations are usually optimized to fit vibrational spectra (calculated or experimental), which contain individual frequencies and their assignments. Torsional energy parameters are optimized to fit the rotational barriers about single bonds in simple molecules. The parameters used for nonbonded interactions are undoubtedly the most difficult part of the parameter optimization. The optimization of the electrostatic parameters is based on the reproduction of target data from QM calculations, in particular, the partial atomic charges should be able to reproduce the electrostatic field of the molecule. The vdW parameters are optimized by reproducing experimental heats of vaporization, molecular volumes and free energies of hydration.[215]

Parameter optimization approaches differ for various force fields such as CHARMM, AMBER and OPLS. If the force field must be extended to treat a new class of compounds, the optimization method used must be consistent with the original method used to develop the force field. The procedure used to parameterize the force field in this work was developed by MacKerell *et al.* and is summarized in Figure 20.[216] This multi-step iterative optimization scheme allows for a balanced optimization between the inter- and intramolecular parts of the force field.

The parameterization of the CHARMM force field presented in this work is based on *ab initio* data for small molecules, as well as macromolecular simulation data, which

capture the condensed phase properties. The methods of *ab initio* quantum mechanics are not yet applicable to large biomolecules, therefore, in parameter optimization the macromolecule is divided into smaller model compounds, and target data is collected on the different geometries of these compounds.

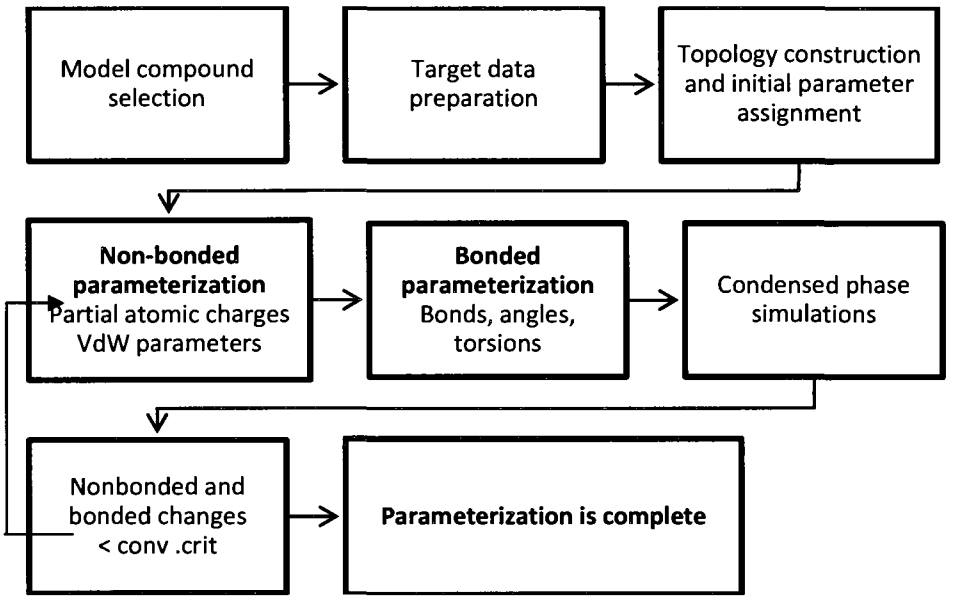

Figure 20. Parameter optimization cycle

A model compound should include all relevant functional groups required to properly describe the local environment (e.g., dihedrals) but be small enough to be studied computationally. In the present study, the initial parameters assigned to the model compound are extracted directly from the CHARMM parameter set. Empirical force field calculations are performed on the model compound and the computed properties are compared with the target data. Then, the parameters are manually adjusted to better reproduce the target data. The entire parameterization process involves iterative steps since changing certain parameters may affect previously optimized parameters and require additional iterations to bring the parameters closer to the target data. The optimization process is considered complete when all sets of parameters satisfy the

convergence criteria associated with the target data for the model compounds. Then, the resultant parameters can be used to perform simulations of macromolecules.

When parameterizing the force field for a new set of compounds, it should be noted that the force field is "empirical". Even if the same form of the potential energy function is used, the quality and results of the force field can differ if different target data are utilized. A great amount of correlation exists between the different parameters of the force field, and therefore, different combinations of parameters can reproduce the same set of target data. Thus, the iterative optimization process is extremely important to the quality and results of the force field. Even though automatic parameterization procedures have been attempted, a significant amount of manual work is still required.

CHAPTER III

FORCE FIELD PARAMETERS FOR CYCLOPENTYL PNA

COMPUTATIONAL METHODS

The cyclopentane structure is not implemented in the standard version of the CHARMM27 force field. Thus, it is essential to parameterize this structure in order to investigate all experimentally tested *cp*PNA oligomers. This goal will be achieved by adding the missing parameters to the CHARMM force field. Parameterization occurs in two stages. In the first phase, we optimized the non-bonded partial charge parameters, and in the second, we optimized the bond, angle and dihedral parameters. All parameters developed in this work were based on *ab initio* results from a training set of small molecules that represented the desired target systems. The data necessary for these parameterizations included bond lengths, bond angles, torsion angles, dipole moments, and vibrational frequencies.

Quantum mechanical calculations were carried out with the Gaussian'03[217] program and are described below. The investigation started with the construction of a model structure followed by location of all of its possible conformations. Optimized geometries for the model conformers and their complexes with water were obtained by HF-6-31G(d) and MP2/6-31G(d) calculations. All optimized geometries were identified as local minima by the absence of imaginary normal mode frequencies. To correct for the harmonic nature of the *ab initio* calculations, all QM frequencies where adjusted using a scaling factor of 0.94.[218] In order to obtain the potential energy surfaces (PES) for selected backbone torsion angles, torsion energies were calculated by holding the selected torsion angle fixed at different increments (30° increments from –180° to 180°) at the MP2/6-31G(d) level of theory. For the cyclopentane ring, the torsion angle range was limited to -40° to 40°.

Empirical force field calculations on model geometries were performed using the CHARMM program and the CHARMM27 force field. This newer force field was chosen

over CHARMM22, because it provides an improved description of the conformational properties of nucleic acids. Unfortunately, the topology and parameters for PNA residues were not directly available in the standard molecular dynamics package. To the best of our knowledge, only two sets of parameters for PNA have been published in the literature.[59,222] We obtained one set from Dr. Lennart Nilsson (Karolinska Institutet, Sweden). The initial parameters for the *X*PC residues in *cp*PNA were obtained from existing similar molecules, such as proline, already parameterized in the force field. The atom types were assigned based on CHARMM atom type definitions. This effort makes most of the initial bonded and non-bonded parameters directly available from the CHARMM parameter set. Any missing parameters were assigned by employing values for chemically similar bonds and angles.

Full geometry minimizations of all model conformers were performed with the CHARMM force field in vacuum using the conjugate gradient minimization algorithm (CG). For the model compound calculations, no truncation of the van der Waals or Coulomb interactions was performed. Starting geometries for the minimizations were taken directly from the QM results and read into CHARMM. MM vibrational frequencies were calculated using the VIBRAN module in CHARMM. Constrained minimizations were employed for the calculation of the torsional profiles. A torsional constraint with a large (10,000 kcal/mol/rad) force constant was set for the dihedral angle of interest, whereas the other degrees of freedom were set to their minimum energy values. A fifteen degree increment and 2000 steps of CG minimization were used for all PES calculations.

The parameterization procedure was begun by optimizing the partial charges. Partial atomic charges were based on the CHarges from the Electrostatic Potentials using a Grid method (CHelpG) algorithm.[219,220] which fits atomic centered point charges to the molecular electrostatic potential. This method, which uses a grid of point charges to represent the electrostatic potential near the van der Waals surface, is widely used and regarded as a superior method for determining atomic charges. It is often used to obtain input charges for molecular mechanics calculations. However, a shortcoming inherent to this method is that the assignment of grid points does not reach buried (sp^3-hybridized) atoms well enough,[51] possibly causing inaccuracies in sterically crowded environments. Moreover, the CHelpG method does not sample points far enough away from the vdW

surface. Despite these limitations, electrostatically derived partial charges are significantly better in quality than those calculated by the default Mulliken method. A particular advantage of CHelpG compared to other electrostatic methods is its rotational invariance. Because a uniform grid of probe points is used to sample the electrostatic potential, the resulting charges are not dependent upon the molecule's orientation in the coordinate system.

The HF/6-31G(d) CHelpG charges were subsequently modified to be consistent with the original force field and further adjustments were made to the charges for polar and nonpolar hydrogens in accordance with the original parameterization procedure: the charges of all nonpolar hydrogens were set to a value of 0.09 e and the excessive positive charge was added into the adjacent carbon atoms. As in the original CHARMM27 force field, transferring charges from the small model compounds to the PNA fragments was accomplished by adding the charge of the removed hydrogen atom to the heavy atom from which it was deleted. Finally the charges on the polar groups were adjusted to better reproduce scaled HF/6-31G(d) water interaction energies.

The CHelpG charges used in the present work are different from the Mulliken charges that were used as a starting point for optimization of the original CHARMM27 force field. In general, the CHelpG charges are larger in magnitude and are better able to preserve the molecular dipole moments. The Mulliken charges on the other hand are considerably basis-set dependent, particularly when diffuse functions are used, and generally smaller in magnitude. It should be emphasized that the Mulliken charges used as a starting point for parameter optimization of CHARMM27 were often significantly altered upon optimization to obtain appropriate energies for intermolecular interactions with TIP3P[221] water molecules. The final charges in many cases were actually closer to the CHelpG charges.

Once the partial charge optimization was complete, equilibrium geometry parameters and force constants were optimized iteratively until satisfactory fitting to the target data was achieved. After this, iterative adjustment of the charges was coupled with adjustment of the internal parameters, until overall convergence was reached.

MD simulations for testing of the *cp*PNA parameters

In general, a molecular dynamics simulation comprises four steps: minimization, heating, equilibration, and production. In the first phase, the initial configuration of the system (which may be obtained from crystallographic data or a graphically built model) is subjected to energy minimization to relieve any major stresses. Velocities are then assigned to each atom and increased slowly until the target temperature is achieved. Equilibration follows, in which velocities corresponding to the target temperature are reassigned constantly, and finally one enters the production stage of the simulation.

The initial coordinates of the atoms were taken from the quantum chemical calculations. Next, each molecule to be studied was embedded into a box of 3108 TIP3P water molecules. A minimization that allowed the positions of the water molecules to vary, but kept the model structure fixed was performed.

The investigated systems consisted of 36 heavy atoms plus 486 water. Molecular dynamics was performed for 20 ps while the temperature was raised from 60 K to 298 K. The SHAKE[225] procedure was applied to all hydrogen atoms so that a time step of 2 fs was possible. A cutoff of 13 Å was applied to the non-bonding Lennard-Jones interactions and the Particle Mesh Ewald method[226] was used to account for long-range interactions. The simulations were performed at constant NVE.

After heating, the equilibration (80 ps) and production steps were carried out. For a statistically meaningful representation, long simulation times on the order of nanoseconds are essential. Our simulations were carried out for 10 ns. We used VMD, version 1.8.6[227] for visualization.

RESULTS AND DISCUSSION

In order to increase the PNA binding affinity for complementary oligonucleotides, numerous PNA backbone modifications have been investigated during the last decade. To enhance the oligonucleotide binding affinity, *trans*-1,2-cyclopentane units were introduced into the PNA backbone. It was demonstrated that the sensitivity of these new PNA structures to binding DNA/RNA was improved by three orders of magnitude.[181] Because of their biological importance and potential therapeutic uses, it is of interest to understand the conformational (particularly the dihedral angles) and dynamical properties of these novel *cp*PNAs at the molecular level. Molecular dynamics simulations are a powerful tool for elucidating the atomistic details of biomolecular systems; however, the empirical force field parameters used must be optimized so that they accurately treat the class of compounds that one is attempting to model. In this work we focused on developing new CHARMM force field parameters for cyclopentane, namely partial atomic charges, equilibrium bond lengths, angles and dihedrals. The determination of force field parameters for modeling a cyclic ring structure is rather complicated. Numerous aspects have to be taken into consideration. Development of an appropriate force field will enable us to run MD simulations on *cp*PNA oligonucleotides and to better understand the conformational dynamics of these systems.

Optimization of force field parameters

In CHARMM, every atom must be assigned an atom type. Each atom type is given a number that is used, along with its connectivity to other atoms, to assign it molecular parameters from the parameter table. Some elements have multiple possible atom types to account for a variety of chemical environments. For example, carbon is described by 27 different definitions ranging from alkane carbons to carbons in heme prosthetic groups. In our study, we needed to add four new atom definitions to CHARMM for the ring moieties and amide nitrogens in the backbone. The respective CHARMM atom types of the atoms in our model system, which are used in the tables and the text, are depicted in Figure 21.

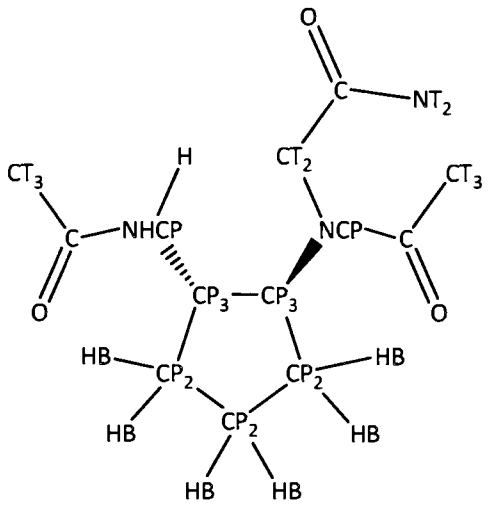

Figure 21. CHARMM atom types assigned to the model compound. The parameters in Table 6 and 4 are given in terms of the atom types defined here

Parameter development for empirical force fields such as CHARMM is accelerated by the modular character of these force fields. Following standardized protocols based on the original development of these force fields ensures the transferability of parameters for new chemical compounds. A detailed description of the parameterization procedures and its motivations were previously described. As in the original force field parameterization, a self-consistent step-wise optimization approach was taken that involved the iterative adjustments of internal and external parameters until convergence was obtained.

To begin this study, a representative model compound was designed and subjected to *ab initio* geometry optimization at the HF/6-31G(d) and MP2/6-31G(d) levels of theory (Figure 22). In addition, various configurations of the model structure were generated in order to study the local conformational behavior of the *trans*-cyclopentyl moiety. These *ab intio* structures were then fully optimized in CHARMM with the initial parameter set described above and vibrational frequencies were calculated.

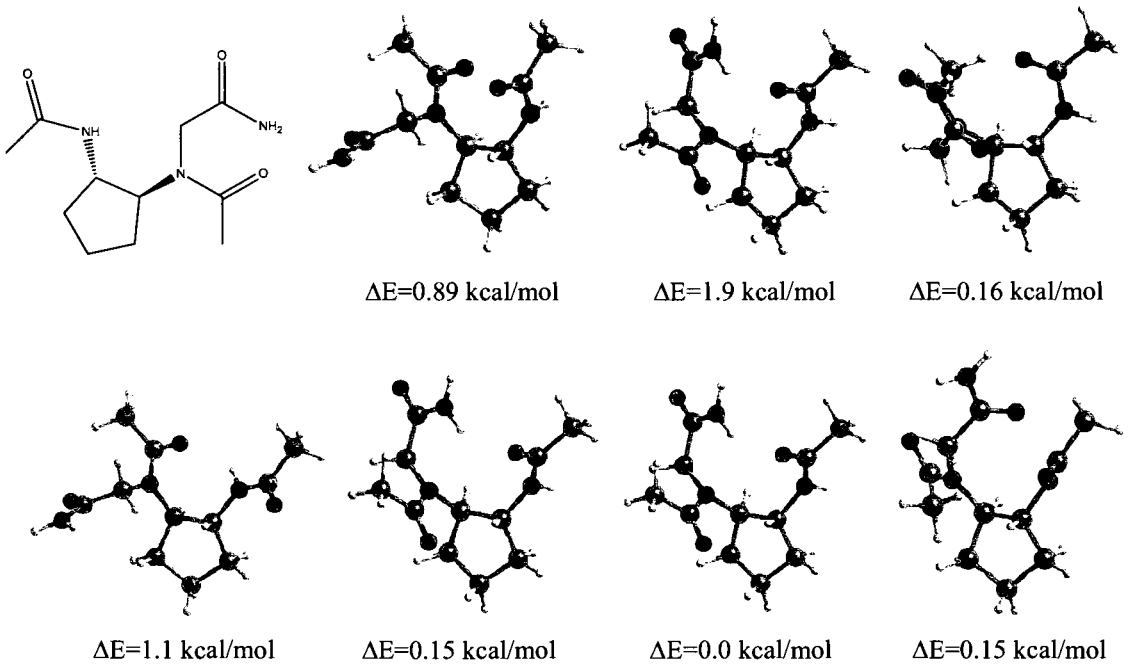

Figure 22. Chemical structure and optimized geometries of the model compound at the HF/6-31G(d) level of theory

PARTIAL CHARGES

The accurate representation of the partial charges involved in electrostatic interactions is an important element in any force field. The more accurate the charge model, the better the quality of the force field. It is necessary to obtain a partial charge set from a number of different conformations to avoid, as much as possible, any conformational dependency. To derive new atomic partial charges for the *cp*PNA molecule, seven different conformations of the model compound were obtained and optimized using the HF level of theory. CHelpG charges were then calculated for the atoms of the optimized structures. Since it was important to determine high quality charges, the CHelpG method was chosen.[220]

Adjustments were made to create charges more consistent with the CHARMM force field parameters. For instance, in the CHARMM force field, all non-polar hydrogens are constrained to have a charge of 0.09 *e*, while polar hydroxyl hydrogens are

assigned a charge of 0.43 *e*. Thus, the hydrogens were given the appropriate CHARMM partial charges and the difference created with the CHelpG charges was added to the adjoining heavy atoms. Charges on equivalent atom types were averaged.

Water-solute interactions are of major importance, and it is necessary to scale the charges of the solute to reproduce the quantum mechanically obtained interaction energies. For this purpose, four independent water molecules were placed within hydrogen bonding distance of the polar groups (Figure 23). HF/6-31G(d) geometry optimizations were performed on the initial water positions while fixing the position of the model molecule. The original CHARMM van der Waals parameters have been kept for all atoms. Thus, since the vdW parameters were not adjusted, it was not necessary to move the individual water molecules to various positions. Interaction energies between an individual water molecule and the system were calculated as the difference between the energy of the model–water complex and the sum of the individual monomers, water and the model compound. For neutral hydrogen bonded complexes, HF/6-31G(d) calculations generally give geometries and interaction energies in good accord with very high level *ab initio* calculations, owing to cancellations of errors from the basis-set extension and correlation energy.[229-231] In addition, HF/6-31G(d) interaction energies were used to adjust some of the partial charges already in CHARMM in order to maintain consistency among the partial charges of the force field.

Usually, the best correspondence between MM and *ab initio* interaction energies is achieved if the former is scaled by a factor of 1.16. This adjustment corrects for the absence of polarization and attractive Lennard-Jones contributions in the HF calculations. Thus, the scaled interaction energies were employed when comparing the CHARMM results to the *ab initio* results. Hydrogen bond lengths are uniformly 0.2-0.3 Å shorter from the force field based calculations than from the HF/6-31G(d) optimizations.[232] Without this contraction in the CHARMM bond length parameters, the computed liquid, density of water would be too low.

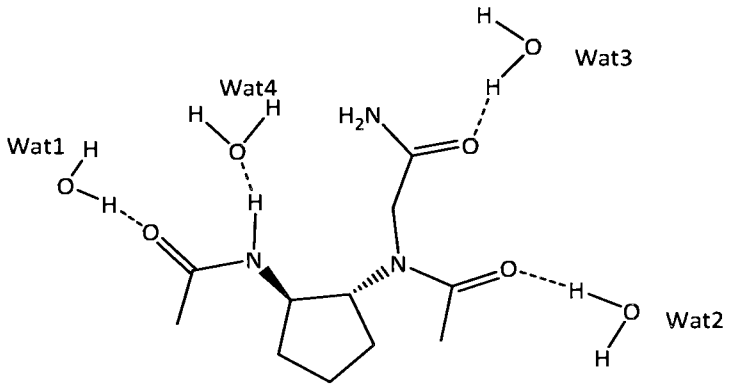

Figure 23. Locations of the model-water interaction sites

A comparison of the force field dimer interaction energies and the *ab initio* target data is presented in Table 4, while Figure 24 contains the optimized geometries of the lowest energy complex for each interaction site. To estimate how well the optimized CHARMM parameters reproduce the total interaction energy of the four waters, we compared the HF/6-31G (d) and the CHARMM values. The sum of the individual water-model interaction energies shown in Table 5 is -31.19 kcal/mol for the optimized CHARMM calculations and -31.66 kcal/mol for the HF/6-31G(d) calculations, yielding a difference in overall interaction energies of 0.47 kcal/mol. Averaging over only one water position gives differences below 0.2 kcal/mol for all four water locations.

The hydrogen bond distance between the model and the water molecules is given in Table 4. The average QM and MM values for (OH) distances are 1.88 Å and 1.80 Å, respectively. Interactions 1 and 3 are unsymmetrical with CO...H distances of 2.04 and 1.87 Å; the hydrogen bonding OH bond is in the plane of the ring, while the other water hydrogen is perpendicular to this plane. Interaction 4, where the oxygen of water acts as a hydrogen bond acceptor, has the N-C=O plane coplanar with the plane of the water molecule (Figure 24).

Table 4. Model compound-water interaction energies and distances. The numbers indicate the water location for which data is being presented

	Energy (kcal/mol)					Distance (Å)		
1)O2..H OH	**HF/6-31G(d)[a]**	**MM Initial**	**Difference**	**MM Optimized**	**Difference**	**HF/6-31G(d)**	**Optimized**	**Difference**
Conf 1	-10.2	-10.7	-0.5	-10.8	-0.6	1.8	1.73	-0.07
Conf 2	-10	-10.2	-0.2	-10.1	-0.1	2.03	1.76	-0.27
Conf 3	-7.8	-7.8	0	-7.3	0.5	1.93	1.71	-0.22
Conf 4	-9.4	-9.6	-0.2	-9.9	-0.5	1.83	1.7	-0.13
Conf 5	-7.8	-7.7	0.1	-7.3	0.5	1.94	1.77	-0.17
Conf 6	-7.8	-11	-3.2	-7.3	0.5	1.79	1.69	-0.1
Conf 7	-7.7	-7.8	-0.1	-7.3	0.4	1.94	1.71	-0.23
Average			-0.59		0.1			-0.17
2)O2..H OH								
Conf 1	-12.2	-13.9	-1.7	-11.1	1.1	1.72	1.75	0.03
Conf 2	-5.5	-7.5	-2	-5.8	-0.3	1.84	1.75	-0.09
Conf 3	-6.6	-8.9	-2.3	-6.5	0.1	1.82	1.82	0
Conf 4	-6.6	-8	-1.4	-6.1	0.5	1.78	1.75	-0.03
Conf 5	-7.7	-8.5	-0.8	-7.8	-0.1	2.01	1.79	-0.22
Conf 6	-7.9	-9	-1.1	-7.6	0.3	1.81	1.77	-0.04
Conf 7	-7.2	-8.5	-1.3	-7.8	-0.6	2.04	1.82	-0.22
Average			-1.51		0.14			-0.08
3)O2..H OH								
Conf 1	-9.3	-10.9	-1.6	-9.3	0	1.85	1.75	-0.1
Conf 2	-8.6	-12.5	-3.9	-10.4	-1.8	1.75	1.69	-0.06
Conf 3	-8	-9.5	-1.5	-8	0	1.81	1.73	-0.08
Conf 4	-9.2	-10.3	-1.1	-8.8	0.4	1.87	1.75	-0.12
Conf 5	-8	-9.5	-1.5	-8	0	1.81	1.69	-0.12
Conf 6	-8.3	-12.9	-4.6	-7.3	1	1.75	1.73	-0.02
Conf 7	-9.5	-10.2	-0.7	-8.5	1	1.8	1.74	-0.06
Average			-2.13		0.09			-0.08
4)NH..O HH								
Conf 1	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Conf 2	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Conf 3	-6	-6.8	-0.8	-6.2	-0.2	1.95	2	0.05
Conf 4	-6.6	-8	-1.4	-6.1	0.5	1.88	1.98	0.01
Conf 5	-6.2	-11	-4.8	-6.5	-0.3	1.93	1.96	0.03
Conf 6	-8.3	-12.9	-4.6	-7.4	0.9	1.95	1.91	-0.04
Conf 7	-6	-6.8	-0.8	-6.2	-0.2	1.95	1.9	-0.05
Average			-2.48		0.14			0.06

[a]HF target energies were scaled by 1.16 and distances were shortened by 0.20 Å.

Table 5. Average interaction energies (kcal/mol) between water and the model compound

Position	CHARMM Optimized	HF/6-31G(d)	Difference
Wat 1	-8.57	-8.67	0.1
Wat 2	-7.53	-7.67	0.14
Wat 3	-8.61	-8.7	0.09
Wat 4	-6.48	-6.62	0.14

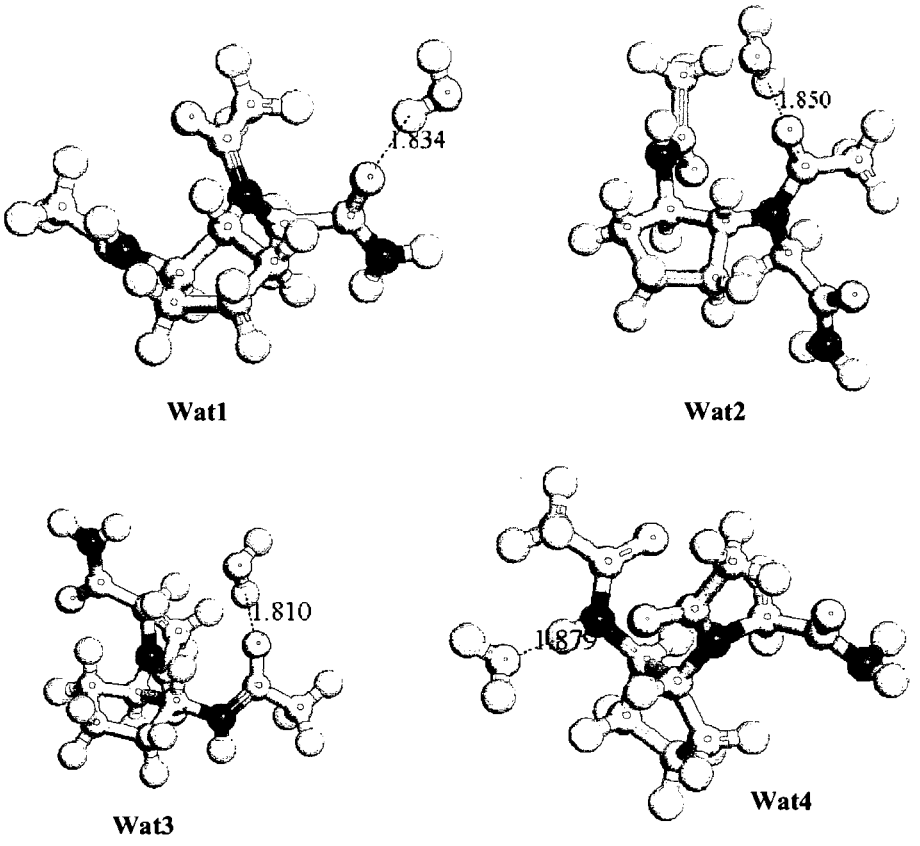

Figure 24. The lowest energy QM structures for the model-water complexes. The dashed lines represent the hydrogen bonding distances

Despite an excellent agreement between the *ab initio* values and CHARMM values for the interaction energies and target distances, the partial charges for the model compound were modified to achieve an even better agreement between the two sets of data. The convergence criterion for charges was a <5% difference between the CHARMM and QM energies. The newly modified CHARMM parameters exhibited distance and energy deviations of 0.038 Å and 0.6 kcal/mol, respectively.

The final partial atomic charge parameters are shown in Table 6. The adjacent electronegative N–H and amide groups induce a large positive charge in the CP3 atom of the ring. On the other hand, the electron donation of the adjacent C=O group creates a negative charge on the CT2 atom of the methylene group. It should be noted that constraints were used on the charge fitting to create groups within the model compound that had a total charge of zero.

Table 6. Atom types and assigned partial charges for the *trans-cp*PNA model

Name	Type	Initial	Optimized
C	C	0.95	0.95
O1'	O	-0.7	-0.64
C2'	CT2	-0.3	-0.2
H2'/H2"	HB	0.09	0.09
N2'	NCP	-0.52	-0.45
C5'	CP3	0.3	0.32
H5'	HB	0.09	0.09
C6'	CP3	0.64	0.49
H6'	HB	0.09	0.09
N	NHCP	-1.04	-1.01
H1'	H	0.31	0.4
C3'	C	0.82	0.58
O3'	O	-0.65	-0.47
C4'	CT3	-0.35	-0.38
H3'/H3"	HB	0.09	0.09
C7'	CP2	-0.18	-0.26
H7'/H7"	HB	0.09	0.09
C8'	CP2	-0.18	-0.24
H8'/H8"	HB	0.09	0.09
C9'	CP2	-0.18	-0.08
H9'/H9"	HB	0.09	0.09

GEOMETRIC FACTORS

Optimization of the bond length and valence angle parameters was based on the geometric target data from the QM calculations. The conformations of the model compound were optimized at the MP2/6-31G(d) level of theory in order to obtain more accurate geometries. The new parameter set was initially built from existing parameters, since these values have been tested over a wide range of molecules. Missing bond stretching and angle bending parameters were extrapolated from related types in the CHARMM library. Using the CHARMM force field, each model compound conformation was energy minimized with steepest descent followed by Newton-Raphson. Then, the CHARMM parameters were modified to yield a good agreement between the quantum mechanical results and CHARMM. In cases where the bond and angle deviations between the original CHARMM force field and the QM structure were larger than 0.05 Å or 60°, the minimized bond lengths and angles obtained from the HF optimized structures were used as the equilibrium bond lengths and equilibrium angles for the new CHARMM parameters. When a choice had to be made between matching a skeletal bond angle formed by heavy atoms versus another bond angle, preference was given to the skeletal bond angle.

Table 7 compares the geometries of the model compound determined by CHARMM and QM. The reported values are averages over the different conformations. The MM geometries for the model are in excellent agreement with the MP2 results. The average error between CHARMM and *ab initio* is 0.02 Å for the bond distances and 0.28° for the angles. The MM C-N and ring C-C bond lengths and the N-C-C and ring C-C-C bond angles are within 0.001 Å and 0.15° of the *ab initio* values, respectively. Overall, the RMS deviations between the CHARMM and MP2 geometries are very good with values of 0.0018 Å and 0.2° for the 12 bond distances and 27 angles, respectively.

The most significant deviations in bond lengths are observed for the NHCP-CP3, CP3-CP2 and NHCP-C bonds. The QM bond length of the NCP-CP3 bond in conformation 2 is about 0.02 Å longer than the average calculated length in all other conformations. This elongation, which may be due to steric effects, is not entirely reproduced by the molecular mechanics calculations. In CHARMM, the deviation of this

Table 7. Average bond length and angle parameters for the model compound. The bonds and angles are listed using atom types from Figure 21

Bonds (Å)	**QM**	**CHARMM Initial**	**CHARMM Optimized**	**Force Constant Initial**	(kcal mol $^{-1}$ Å^{-2}) **Optimized**
NHCP -CP3	1.45	1.43	1.437	320	320
CP3-CP2	1.54	1.53	1.526	222.5	222.5
CP2-CP2	1.54	1.53	1.538	222.5	222.5
CP3-CP3	1.54	1.5	1.487	222.5	410
CP3-NCP	1.46	1.48	1.408	220	320
NHCP-C	1.37	1.34	1.402	370	320
O-C	1.22	1.23	1.22	620	620
NCP-C	1.38	1.37	1.36	463	463
NCP-CT2	1.45	1.49	1.371	261	320
NHCP-H	1	1	1	440	440
CP3-HB	1.09	1.08	1.09	330	330
CP2-HB	1.09	1.08	1.09	330	330
Ave. Difference		0.0267	0.0019		
Std. Deviation		0.036	0.0018		
Angles (°)					
NHCP-CP3-CP2	114.6	110.5	120	70	-
NHCP-CP3-CP3	112.9	114	109	70	-
NCP-CP3-CP3	114.8	108	106.8	67.7	-
CP2-CP2-CP2	105.6	114	113.1	58.35	-
NCP-CP3-CP2	115.9	108	113.2	67.7	-
C-NHCP-CP3	121.2	120	123	50	-
CP3-NCP-C	120.2	117	125.5	50	-
CP3-NCP-CT2	118.2	120	124.4	35	-
CP3-CP3-CP2	104	111	105.1	53.35	-
CP2-CP2-CP3	104	111	105.9	53.35	-
H-NHCP-CP3	117.5	117	114.5	35	-
NHCP-CP3-HB	107	108	102.8	48	-
HB-CP3-CP3	108.3	111	112	35	-
HB-CP3-CP2	109.2	111	113.1	35	-
HB-CP2-CP3	110.6	110	110.1	35	-
HB-CP2-CP2	111.5	110	110.1	35	-
HB-CP2-HB	108.4	115	108.9	36	-
H-NHCP-C	118.1	123	120.2	34	-
O-C-NHCP	122.9	123	127.5	80	-
NHCP-C-CT3	114.2	117	116.4	80	-
C-NCP-CT2	112.4	110	97	50	-
NCP-CT2-HB	109.6	108	108.3	51.5	-
O-C-NCP	121.8	123	122.8	80	-
NCP-C-CT3	117	108	105.4	50	-
NCP-CP3-HB	105.5	108	104	51.5	-
HB-CT3-HB	109.2	115	108.5	36	-
HC-NH2-C	119.5	111	125.6	50	-
Ave. Difference		2.876	0.2785		
Std. Deviation		3.729	0.1972		

bond length from the average of the other NCP-CP3 bonds is only 0.014 Å. For the NHCP-H bond, the agreement between *ab initio* and CHARMM calculated bond lengths is quite good, with an absolute average difference of 0.004 Å and an RMSD of 0.0078 Å. The only conformation which exhibits a moderate error is number 7, where the N-H bond lengths are 0.008 Å too long in CHARMM. However, because the r_o values are assigned to provide reasonable results over all conformations, it is occasionally necessary to accept error in one member of a series of conformations to maintain agreement with the other conformers.

The key geometric features for this parameterization, the CP3-CP3, CP2-CP2 and CP3-CP2 bond lengths, are well reproduced by molecular mechanics; deviations are primarily observed in conformations 1 and 7. If we look directly at the RMSD for just the CP2-CP2 bond, the result is quite small (0.064 Å), for a bond length that varies by 0.14 Å. Overall, CHARMM gives slightly longer carbon-carbon single bonds (~0.02 Å) than MP2 does.

The absolute average bond length deviation for C-H bonds over all conformations is 0.001 Å. This value includes hydrogens which are on carbons adjacent to nitrogens. Therefore, as might be expected, the most deviant C-H bond lengths are those far from the amide. In particular, the bonds on the terminal methyl groups show the largest deviations of 0.005 Å.

The most important bond angles in the new parameter set involve the new atom types: CP2-CP3-CP3, CP2-CP2-CP3, CP2-CP2-CP2, NCP-CP3-CP2 and NHCP-CP3-CP2. Using the final optimized parameters, MM is in close agreement with the *ab initio* results for these angles with an average deviation of 0.35°. It is noteworthy that the angles vary significantly over the conformations of the model molecule.

Other important skeletal bond angles are CP3-NCP-CT2, NCP-CP3-CP3 and NHCP-CP3-CP3. The absolute average angle deviations were 1.33° for CP3-NCP-CT2, 1.93° for CP3-NCP-CT2 and 0.77° for NHCP-CP3-CP3. Conformation 6 had a deviation of 4.09° from the QM result for the CP3-NCP-CT2 angle. However, the error for this deviation was less than 3.5%, which is quite adequate for a parameterization scheme. Based on the RMSD analysis, in general, most deviations from the *ab inito* results were around 1°. As with the methyl C-H bond length, the CT2-C-O bond angle was already

parameterized in CHARMM and we did not make any adjustments to the established parameters. These results for bond angles should be more than adequate for producing reasonable structures for larger *cp*PNA molecules.

The vibrational frequencies for normal modes produced by CHARMM and *ab initio* values were compared and analyzed for accuracy. For the QM results, the qualitative assignment of the modes was done visually using the GaussView program. The different spectral regions correspond to different types of vibrations. Vibrational frequencies and their assignments for the model molecule in its lowest-energy conformation are listed in Table 8. These modes are distributed in a wide spectral range-from 30 to 3700 cm^{-1}. The region below 700 cm^{-1} involves torsions and angular motions of the heavy atoms (e.g. C-C-C, C-C-O). Vibrational modes in this region are highly coupled and a large number of internal coordinates have small contributions to the potential energy distribution. The vibrational modes associated with C-C torsions are located between 70 and200 cm^{-1}, whereas the modes associated with C-NH torsions are higher than 200 cm^{-1}. The region between 800 and 1200 cm^{-1} is characteristic of the heavy atom stretching modes. An exception is found for 1057 cm^{-1} and 1093 cm^{-1}, which correspond to the N-C-H bending modes. The region between 1200 and 1500 cm^{-1} is characteristic of angular modes involving hydrogen atoms (C-C-H, H-C-H). The high frequency region (>2500 cm^{-1}) is characterized by stretching modes. CH stretching vibrations were found between 2900 and 3058 cm^{-1}, with the CT2-HB vibrational frequency lower than the CT3-HB frequency, which is lower than the CP2-HB frequency.

To derive the force constants for the model molecule, 19 of the most prominent normal modes were chosen for the fitting procedure. The *ab initio* vibrational frequencies were scaled by 0.94 to account for the typical computational overestimation of frequency values.[218] The accuracy of the results was assessed by performing a root-mean-square deviation (RMSD) statistical analysis on the differences between the MM and QM values. Then, an iterative fitting procedure was applied, where the parameters (force constants and distances) were varied until the difference between the vibrational frequencies calculated using the CHARMM force field and QM was minimal. For example, with a force constant of 320 kcal mol^{-1} $Å^{-2}$ for the CP3-NCP bond, the QM value of 1.46 Å could be reproduced within 0.01 Å by CHARMM. The equilibrium

values for CP3-NHCP and C-NCP were changed from 1.43 to 1.437 Å and from 1.37 to 1.36 Å, respectively, but the CHARMM force constants were left unchanged.

In the fitting procedure, the following requirements for the target force constants were considered: correspondence between the CHARMM and *ab initio* results, including the similarity of the structures, the similarity of the frequencies, the resemblance of the potential energy curves, and consistency with the other CHARMM force constants so that interactions with other parts of the system defined by original CHARMM parameters are appropriate.

Table 8. Vibrational frequency assignments for the model molecule

Assignment	MP2[a], cm^{-1}	CHARMM Optimized, cm^{-1}	Diff.
Asym. N—H str	3511.2	3472.5	-38.7
Sym. N—H str	3416.8	3380.5	-36.3
Sym. N—H str	3362.9	3328.8	-34.1
Asym. C—H str (Me)	3058.6	3020.4	-38.2
Asym. C—H str (ring)	3018.6	2987.2	-31.4
Asym. C—H str (CH2)	2960.5	2923	-37.5
Sym. C—H str (Me)	2935.1	2901.4	-33.7
Sym. C—H str (ring)	2905.8	2872.1	-33.7
C=O str + H—N—C bend	1711.2	1747.2	36
C=O str + H—N—C bend	1678.7	1704.8	26.1
H—N—H bend	1563.7	1602.3	38.6
Sym. C—H str (ring) + C—N—H bend	1481.6	1509.7	28.1
H—C—H bend (Me)	1455.9	1466.4	10.5
H—C—H bend (CH2)	1441.3	1450.8	9.5
Asym. N—C str	1418.1	1439.6	21.5
N—C—H+C—N—C bend	1093.4	1068.6	-24.8
Asym. C—C stretch (ring)	1079.5	1051.5	-28
N—C—H bend	1057.1	1042.6	-14.5
Sym. N—C str	847.1	809.2	-37.9
Std. Deviation			29.4
RMSD (all)			39.04
Corr. Coefficient			0.9982

[a]Weighted by the factor of 0.9434

The plot of ν_{QM} versus the optimized ν_{MM} for the model compound is shown in Figure 25. The closer the points are to the diagonal, the better the fit. There were modes that

could not be matched properly by only adjusting bond and angle terms. The MM frequencies of those modes were ~100 cm^{-1} higher than the corresponding QM vibration.

There are several types of bending modes: twist, wag, umbrella and rocking modes as well as normal A-B-C bends. To further complicate matters, some bending modes are heavily coupled with other bending, stretching and torsional modes. For example, in Table 8, the frequency at 1093 cm^{-1} has at least two bending modes coupled together. As a reminder, the CHARMM bending energy does not account for any vibrational coupling as each angle is treated separately. Because only two adjustable parameters are used in CHARMM for each bending mode, the differences between the QM and MM results have a wide range of values. However, these deviations when taken as a whole are quite modest considering the number of internal coordinates involved in each QM vibration.

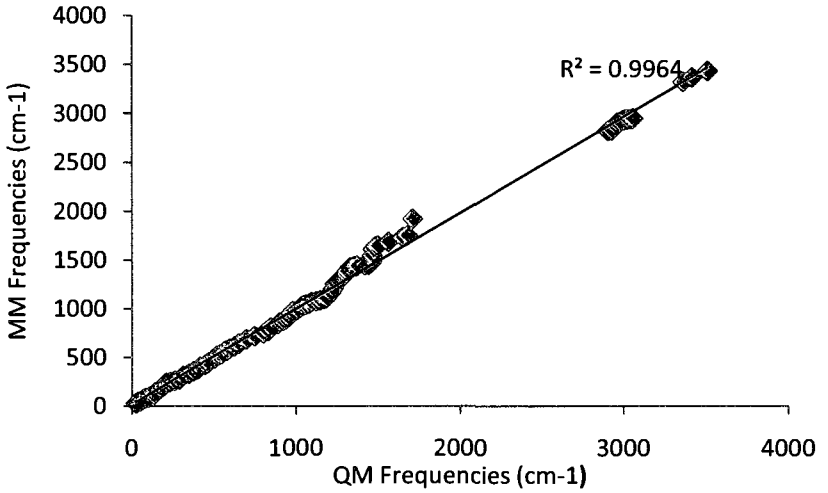

Figure 25. Vibrational frequencies of the model molecule. The plotted line shows the ideal fit between QM and MM data.

The average difference (ν_{QM}-ν_{MM}) for the 15 vibrational frequencies located between 1200 and 1400 cm^{-1} is about 17 cm^{-1}. The standard deviation is about 9.6 cm^{-1}

for the 58 frequencies between 800 and 3000 cm^{-1}. The comparison of the results shows that the NHCP-H, CP3-NCP and CP2-CP2 stretching modes match quite well, with differences of 34, 38 and 28 cm^{-1}, respectively. The QM vibrational frequency of the C-N stretch is identified to be the peak at 847 cm^{-1}, while CHARMM had a value of 809 cm^{-1}. The CT2-NCP and CT2-HB stretching modes have differences in frequencies of 21.5 and 37.5 cm^{-1}. The relatively small deviations observed for the vibrational frequencies are sufficient to validate the accuracy of our parameterization.

The assignment of the frequencies between QM and MM was difficult due to vibrational mixing, different theoretical treatments, and slightly different final geometries. As a consequence, the resulting QM assignments for a few stretching modes are not in perfect agreement with CHARMM. However, the final RMSD and correlation coefficient between CHARMM and QM are 39 cm^{-1} and 0.9982, respectively. Therefore, the developed force field gives a good performance when calculating the equilibrium geometry and vibrational properties of the molecule.

Dipole moments

To mimic the effect of electronic polarizability, which is not explicitly taken into account in fixed-charge additive force fields, atomic charges are overestimated purposely (15-30 %),[233] which leads to an enhanced molecular dipole moment. In molecular mechanics, the dipole moments are calculated by summing any bond moment vectors present in the molecule. In contrast, the *ab initio* derived dipole moment is a one-electron operator property with both nuclear and electronic terms.[205] In neutral molecules, where the number of protons and electrons is equal, μ does not depend upon the position of the molecule within the coordinate system.

The dipole moments from *ab initio* and molecular mechanics calculations are shown in Table 9. For conformation 3, the difference in the dipole moments is unusually large. The observed difference may be due to the limitation of not including electron density shifts in molecular mechanics calculations. Comparing only the deviation in magnitudes, close agreement to the QM calculated molecular dipoles was obtained with an average error of 18.1%. The average absolute difference between *ab initio* and

CHARMM dipole moments is 0.98 D, with the average QM gas-phase dipole moment being 5.42 D and the MM value 6.4 D.

Table 9. Dipole moments for the seven conformers of the model

Molecule	MP2[a]	CHARMM[a]	Deviation
Conf1	5.27	5.94	0.67
Conf2	7.06	7.27	0.21
Conf3	4.26	6.64	2.38
Conf4	5.08	6.04	0.96
Conf5	5.36	6.34	0.98
Conf6	5.09	6.00	0.91
Conf7	5.87	6.63	0.76
Ave.			0.98

[a]All values are in Debye

TORSIONAL PROFILES

In addition to bonds and angles, the torsional terms were also parameterized. The torsion parameters, in conjunction with the atomic charges and vdW parameters are the primary determinants of the relative conformational energies of the molecule. The most flexible areas of the polypeptide are along the backbone where the heterocyclic bases are connected to the amide. The polypeptide will have to rotate around these bonds in order to optimize interactions with DNA/RNA. The bases themselves are aromatic, so the rings will remain planar.

The CHARMM force field uses the rotational barrier height, the dihedral phase angle and the periodicity to define the torsional potential energy function. Therefore, each bonded series of atoms A-B-C-D must have at least one set of these dihedral parameters in the force field. The dihedral parameters are derived by the "fitting philosophy", i.e., the quantum potential was compared to the molecular mechanics potential and the parameters were adjusted until the MM potential reproduced the quantum potential.

The QM potential was generated by rotating the side-chain dihedral angles by 30° increments and optimizing the other degrees of freedom at the MP2/6-31G(d) level of

theory. This scheme provides an energy profile as a function of the dihedral angle and is recommended by the developers of CHARMM for the derivation of the torsional parameters. The lowest energy conformer is set to zero and the energies of the other conformers are given relative to it, since we are interested only in the conformational component of the total energy. The CHARMM potential was generated by using a 15° increment to generate structures and minimizing each of these conformers in CHARMM with the selected torsion angle held fixed. Here, the dihedral term was deliberately set to zero, and we obtained the second curve. If the CHARMM potential did not match the MP2 results, the multiplicity, phase shift, or force constant were adjusted as necessary.

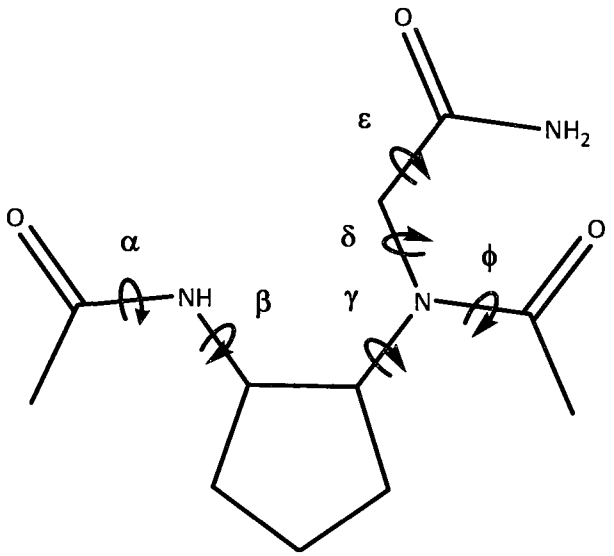

Figure 26. Backbone dihedral angles used for parameterization

Many different torsional profiles were calculated to evaluate the model structure. Initially, six backbone/side-chain dihedrals were selected for parameterization: the rotation of the methyl carbonyl group and the rotation of the two amide groups on the ring (Figure 26). When a specific dihedral angle is altered, many MM energy components may also change (the vdW energy, other torsional energies, and the stretching and

bending energy). All of these terms contribute to the total energy and the overall shape of the torsional energy profile. In initial calculations only the dihedral under examination was kept constant, while the rest of the molecule was allowed to relax. This method, however, produces very complex energy profiles as variation of the dihedrals examined induces large distortions in the rest of the molecule. Therefore, to accurately model a specific dihedral potential, it was necessary to constrain the remaining five dihedral angles to their values in the global energy minimum structure.

The initial CHARMM parameters for the dihedrals needing parameterization did not agree with the MP2/6-31G(d) potential energy surfaces (Figure 7). For example, the C-NHCP (α) dihedral has a minimum at 150° instead of 190°, and the energy barrier is ~3.5 kcal/mol too high. The large barrier at 120° illustrates that rotation about the C-N bond is restricted. For the first amide group (β), the CHARMM curve does not have the correct phase; the MP2 plot has a barrier at 150°, whereas CHARMM has a minimum at that location. The second amide rotation (γ) does not give the lowest energy at 240° and has energy barriers that are too high (~3.6 kcal/mol). The (ϕ) torsion has energy barriers that are ~4.1 kcal/mol too high, but the minimum at 180° does match the MP2 curve. The CHARMM results are in reasonable agreement with the MP2 results for the δ dihedral; however, for the CHARMM curve, the location of the maximum is 30° out of phase. Rotation around C-CT2 (ε) should have a maximum at 270° and 60°; instead there is one maximum at 60°.

To accomplish the necessary parameterization of the torsions, an automated dihedral matching method[234] or Monte Carlo simulated annealing (MCSA) protocol was employed, which optimizes the MM parameter set until the best fit with the QM reference set is obtained. In this procedure, the quantum mechanical energies derived from the rotational profile analysis are used as target data which provide the dihedral parameters by minimizing the squared difference between the MM and QM energies. Initially, all 12 K_ϕ parameters were set to zero and then randomly sampled and selected based on the Metropolis criteria[235] where the target function was the difference between the empirical and MP2/6-31G* energy surfaces. Dihedral multiplicities (n) of 1, 2, and 3 were included for each dihedral, and the corresponding K values, constrained in the range

of -3 to 3 kcal/mol, were optimized to minimize the root-mean-square error (RMSE) between the empirical and QM energetics. The merit function is given by

$$RMSE = \sqrt{\frac{\sum_i (E_i^{QM} - E_i^{MM} + c)^2}{\sum_i 1}}, \tag{11}$$

where E_i^{QM} and E_i^{MM} are the QM and MM energies of conformation i and c is a constant that aligns the QM and MM data to minimize the *RMSE* (Equation 12). Although allowing the phase angles to vary during MCSA can yield a better fit, phase angle values were constrained to be either 0° or 180° in order to preserve the symmetry of the cosine function about 0°. Failure to do so will lead to different energies for a molecule and its mirror image, which is nonphysical.

$$\frac{\partial RMSE}{\partial c} = 0 \tag{12}$$

In the MCSA calculation, the Boltzmann distribution temperature factor (RT) was lowered at the end of each MC cycle, and an exponential cooling schedule was used to decrease the value of RT

$$T_m = T_0 \exp(-\frac{4m}{m_{max}}) \tag{13}$$

where T_0 is the starting temperature, m is the current Monte Carlo step number, m_{max} is the maximum number of Monte Carlo steps, and T_m is the temperature at step m. E_i^{MM} and c are recomputed and these values are then used to calculate $RMSE_m$ (*RMSE* at Monte Carlo step m). Acceptance or rejection of parameters is in accordance with the Metropolis criteria with $\Delta E = RMSE_m - RMSE_{m-1}$. Refinement calculations consisted of multiple (5-10) independent 5000-step MCSA runs with random initial values for the dihedral force constants and exponential cooling. Convergence was achieved when the values of the RMSE in all successive iterations of MCSA runs changed by less than 0.1%. Clearly, agreement between the empirical model and the QM data was improved via enhanced

sampling of the parameters via the MCSA approach. However, in some instances, empirical adjustment of these automatically generated parameters was undertaken to further refine the CHARMM potential energies.

The final CHARMM parameters give correct rotation barriers and have the right multiplicity. Although the agreement between the QM and MM data is not perfect, the rather complex shapes are reproduced satisfactorily. During parameterization, special care was taken to reproduce as accurately as possible the low energy ($\leq$3 kcal/mol) regions. When these new parameters are fully incorporated and used to study large macromolecules, it is important to make sure we have the lowest energy conformer, which is paramount in studying fluctuations that could affect binding ability. Consequently, this feature is far more critical in parameterization than the barrier heights. Thus, the most important criteria for determining an acceptable torsional fit is whether or not the MM dihedral angle of the minimum energy structure matches that of the QM profile.

For the dihedrals α and γ, the reproduction of the high-energy regions was achieved at the expense of reproducing perfectly the low-energy regions (Figure 27). The CT3-C-NHCP-CP3 torsion curve (α) is a symmetrical curve with an energy maximum at 300° and an energy minimum at 180°. CHARMM is able to fit this energy surface with a RMSD of 0.86 kcal/mol over the entire curve. In the case of γ, the rotational barriers are overestimated. This discrepancy was deemed necessary since smaller dihedral constants which lower the barriers also shift the position of the MM minimum. The C-NHCP-CP3-CP2 (β) torsion curve has a maximum at 120° and two minima at 60° and 270°. MM tends to overestimate the energy of this rotation by about 2 kcal/mol, but has the maximum and minima in approximately the correct locations. Concerning the CP3-NCP-C-CT3 (ϕ) torsion, the minimum energy wells for both *ab initio* and CHARMM occur at 0° and 180°, and therefore are in very solid agreement. The barriers to rotation around the NCP-C bond are somewhat smaller compared to the *ab initio* ones: 18.6 and 23.5 kcal/mol versus 19.1 and 28.2 kcal/mol, respectively. The δ torsion shows a maximum at 180° and a minimum at 270°. CHARMM shows correct phase and multiplicity but the barrier at 180° is 1.3 kcal/mol higher than with MP2.

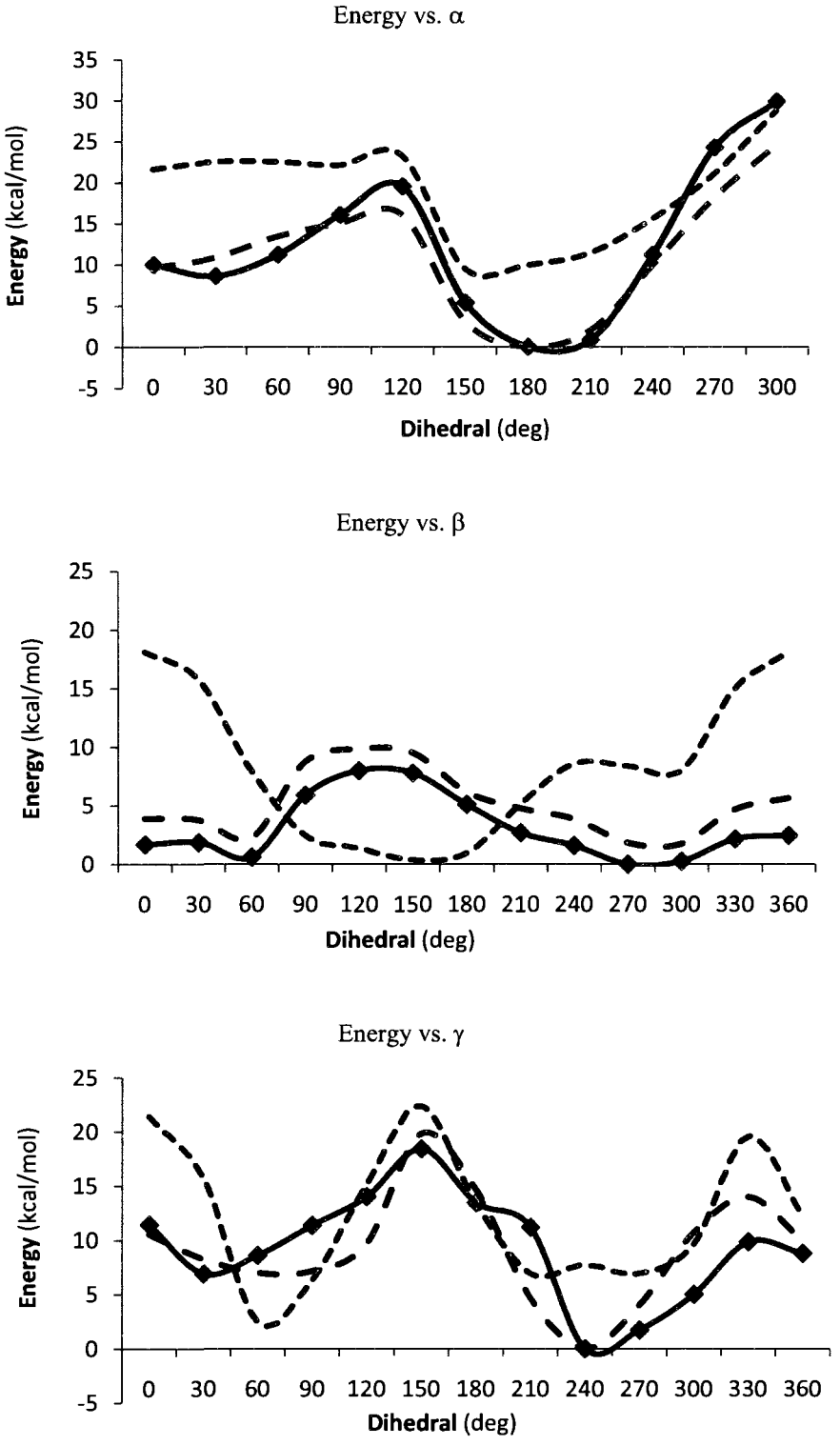

Figure 27. Torsional energy plots for the model compound. Blue diamond (♦), MP2/6-31G(d); green short dash (--), initial CHARMM; red long dash (--), final CHARMM

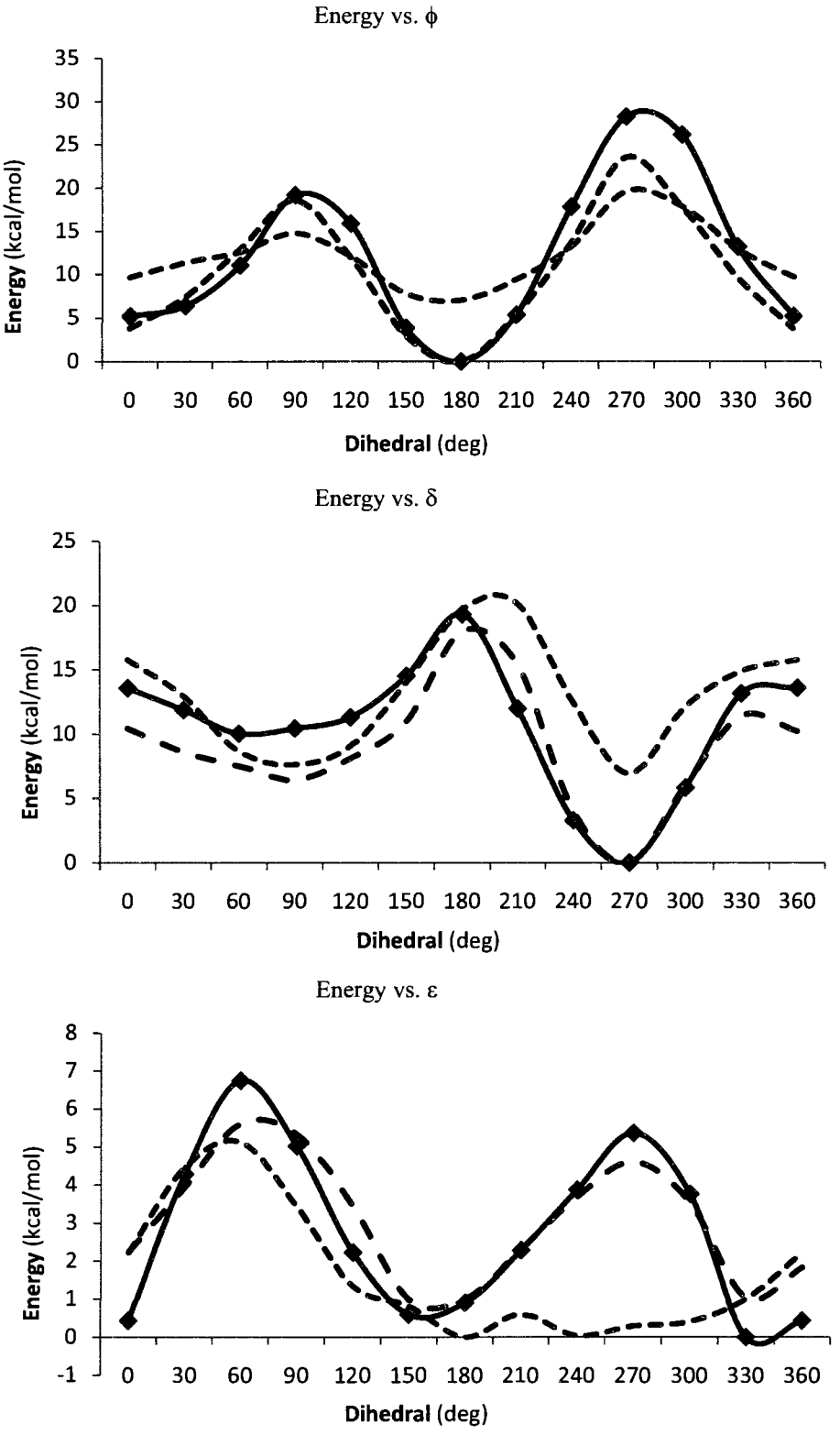

Figure 27. Continued. Blue diamond (♦), MP2/6-31G(d); green short dash (--), initial CHARMM; red long dash (--), final CHARMM

Ab initio calculations of the model molecule reveal two energy minima for the torsion angle ε: 0° and 150°, with the first minimum lower by 0.2 kcal/mol. In addition, the peaks have different widths, and the energy barriers separating the stationary states are 6.7 and 5.3 kcal/mol. Out of all six torsion curves, the torsional profile of the CP3-NCP-C-CT3 dihedral (ϕ) has the largest RMSD. CHARMM calculated the -90° energy to be lower than the QM energy by over 4.5 kcal/mol and this difference causes the large error. However, better agreement could not be obtained since alteration of this dihedral reduced the agreements found for the other dihedrals.

Knowledge of the shape and the height of the energy barriers for the ring torsions of cyclopentane play a significant role in determining the configuration of the *cp*PNA macromolecule, and therefore, special care was taken to obtain excellent parameters for the equivalent dihedrals of the model system: ω2 and 5: CP2-CP2-CP3-CP3 and ω3 and 4: CP3-CP2-CP2-CP2, although the parameters for all the dihedrals of the cyclopentane ring were optimized. Increments of 10° over the range from -40° to 40° were used to generate the conformational energy profiles of the ring dihedrals for both QM and MM. All of the ring torsions were incremented simultaneously and held fixed at each value, while the remaining geometric variables were optimized. In the CHARMM calculations, the force constants of the ring dihedrals were set to zero, so that they did not contribute to the energy. Only the torsional energy parameters for dihedral angles involving heavy atoms were modified in order to have the smallest possible impact on the original CHARMM force field. The lowest energy structures from CHARMM were compared to the *ab initio* calculations. The conformational energies versus dihedral angle are plotted in Figure 28. The data shows that even though the torsional parameters that are associated with the heavy atom ring dihedral angles were set to zero, the energy barriers are still greatly overestimated. The energy differences between the CHARMM potential energy curves and the corresponding *ab initio* energy curves was minimized via the MCSA procedure previously mentioned (Equation 11).

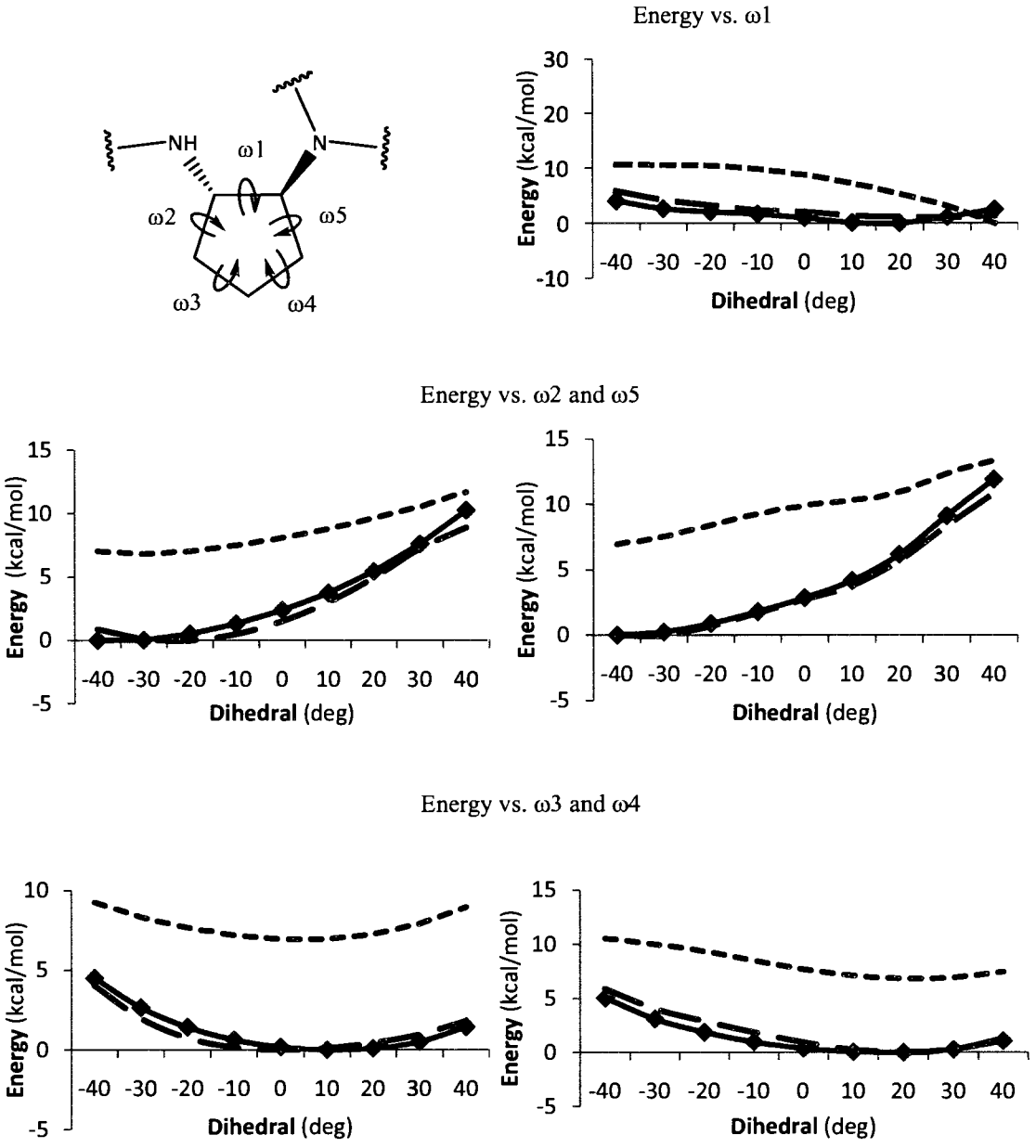

Figure 28. Structure (top) and five torsional energy plots for the model compound. Blue diamond (♦),MP2/6-31G(d); green short dash (--), initial CHARMM; red long dash (--), final CHARMM

New CHARMM ring torsional energies were calculated using the new set of parameters obtained from the MCSA optimization and compared to the MP2 energy

curves. It can be seen from Figure 28 that a significant improvement was made, particularly for the height of the barrier. However, the shapes of the energy curves were not always reproduced very well since the CHARMM torsional expression does not provide enough flexibility to account for the shape of the *ab initio* profiles. In spite of these differences, the maximum deviation of the fitted MM profile from the QM one does not exceed 0.5 kcal/mol and thus, the parameters should not lead to significantly different dynamical behaviors. The torsional parameters are summarized in Table 10.

Table 10. Torsional parameter set for the *cp*PNA model molecule

	CHARMM Initial			CHARMM Optimized		
Dihedral	K_ϕ	n	δ (deg)	K_ϕ	n	δ (deg)
CT3 - C - NHCP - CP3	1.6	1	0	2.5	2	180
CP3 - CP3 - NHCP - C	1.6	1	0	1.8	2	180
CP3 - NCP - CT2 - C	0	6	180	0.07	2	180
NCP - CT2 - C - NH2	0	6	180	1.83	1	180
C - NCP - CP3 - CP2	0.14	3	0	2.56	3	180
NHCP - CP3 - CP3 - NCP	0.2	3	0	1.63	3	0
CP3 - CP2 - CP2 - CP2	0.16	3	0	2.74	2	180
CP3 - CP3 - CP2 - CP2	0.14	3	0	1.4	2	180
CP2 - CP3 - CP3 - CP2	0.2	3	0	2.5	2	180

Comparison of the three surfaces (ω1, ω3 and ω4) shows that they are all similar, with minor differences, as expected. All of the potential energy curves have a long narrow minimum energy region in the vicinity of 10-20°, since the five-membered ring places restraints upon the torsions. For example, the CP3-CP2-CP2-CP2 (ω3 and 4) torsion curve has an energy maximum at -40° and an energy minimum at 10°. CHARMM does a good job at reproducing the entire curves of the ω3 and ω4 torsions, with RMSDs of 0.29 kcal/mol for both. For ω4, there is some disagreement in the energies at the maximum, with a deviation of 0.8 kcal/mol for that point. For the CP2-CP2-CP3-CP3 (ω2 and 5) torsion, CHARMM matches the QM curve with a RMSD of 0.3 kcal/mol. The largest errors are at the -40° and 0° points with deviations of ~0.8 kcal/mol for both. However, the error falls within the accepted error range for MP2 calculations. The final

curve, CP2-CP3-CP3-CP2 (ω1), has an energy maximum at -40° and a minimum at 20°. The RMSD between CHARMM and MP2 is 0.47 kcal/mol. Here again, the largest error associated with the CHARMM curve is 1.8 kcal/mol at -40°.

PSEUDOROTATION AND STRUCTURAL NOMENCLATURE

To assess the validity of the ring torsional parameters, a comparison was made between the QM conformational energies versus the pseudorotational angle, *P* and the CHARMM conformational energies. "Pseudorotation"[236] was developed as a way to describe interconversion between different conformers of cyclopentane rings. The formula for calculating the pseudorotational coordinate *P* from values φ_0, φ_1, φ_2, φ_3 and φ_4 (Figure 29) is given in Equation 14.

$$P = \arctan\left[\frac{(\varphi_2+\varphi_4)-(\varphi_1+\varphi_3)}{2\varphi_0\,(\sin 144^0+\sin 72^0)}\right] \qquad (14)$$

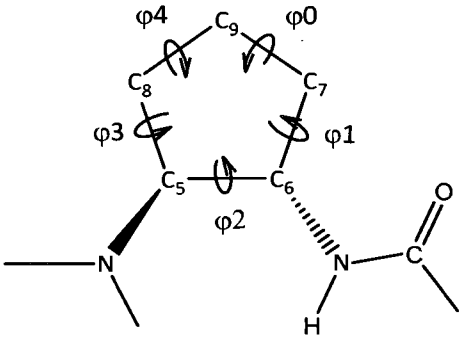

Figure 29. The position of the torsional angles on the cyclopentane ring. All φ torsional angles are used to calculate the pseudorotation angle and amplitude

Puckering is a common phenomenon observed in all cycloalkanes. For example, if cyclopentane were forced to be completely planar then ten fully eclipsed hydrogens

create 10 kcal/mol of torsional strain[237] which can only be alleviated by assuming one of the low energy conformers: envelope (*E*) or half-chair / twist (*T*). The envelope has one atom positioned out of the plane defined by the other four atoms. In the half-chair, three atoms form a plane and the two remaining atoms lie above or below this plane. The cyclopentane ring has five atoms, each of which can be either above or below the plane at different stages of the pseudorotation cycle, so there will be ten possible envelope conformations, with *P* values of 0,36, 72°, etc., alternating with ten twist conformations, with *P* values of 18, 54, 90°, etc.. It is important to note that every 36° ± 4° change in the phase angle alters one envelope conformation of the cyclopentane ring to another, i.e. a different atom emerges from the plane of the other four atoms. At half way between two adjacent envelope conformations along *P*, there exists a twisted conformation of the cyclopentane ring. Given a starting conformation, an increase of 360° in *P* takes us completely around the pseudorotation cycle, returning to the starting configuration. Figure 30 visualizes the geometries of the possible preferred puckered states for cyclopentane.

Through pseudorotation, ring conformations having different out-of-plane atoms may interconvert without traversing a high energy planar structure. For cyclopentane, this process of pseudorotation has little or no energy barrier resulting in essentially "free" ring pucker motion. However, in nucleic acids pseudorotation is hindered because of the substituents on the ring. Five-membered rings are actually quite sensitive to the electronic effect of the substituents attached to it. Thus, the nucleobases cause a strong puckering effect on the sugar; hence, some of the conformers become more energetically favored than others.

Correct representation of the cyclopentane ring pucker is essential for an accurate description of the dynamics of cyclopentane rings in the PNA molecules. To ensure that the torsional parameters correctly represent the relative energies of the pucker conformations, *P* was calculated for each QM structure and the corresponding CHARMM structure. The conformations of the rings were further analyzed in terms of whether they are envelopes or half-chair (Table 11). Given the limitation of the force field, the emphasis was on whether or not the trends in the ring conformation energetics

were reproduced overall. The MP2/6-31G(d) and MM pseudorotational potential energies are plotted in Figure 31.

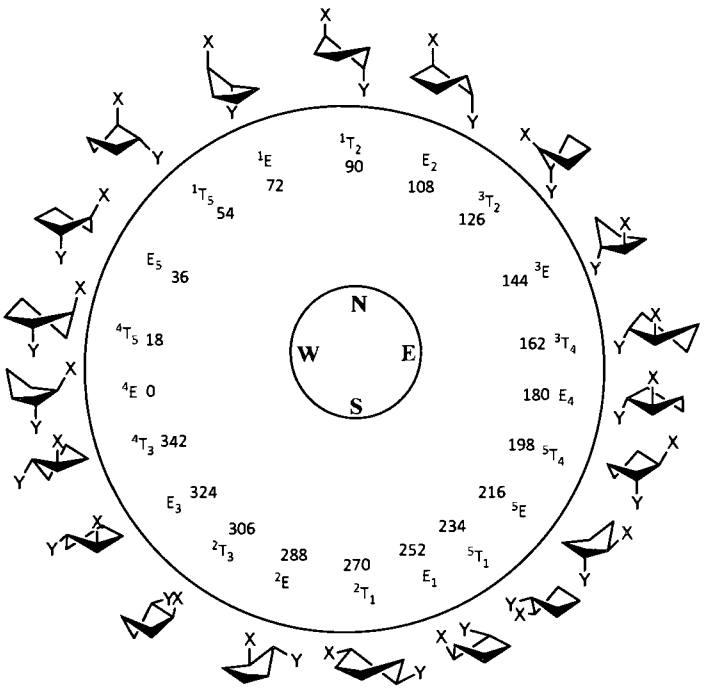

Figure 30. A common representation of the pseudorotation of five-membered rings

CHARMM reproduces the QM surface and both curves have the same general trend with a global minimum twist conformation. The two low energy regions on the QM plot span 340° to180° and are commonly referred to as North and East on the pseudorotation potential. The North family of conformations comprises three pseudorotamers ($^7T_6$, E_6, $^5T_6$). The subscript and superscript refer to which atoms (atom numbers are given in Figure 29) are above or below the plane of the other atoms. The lowest energy North conformer adopts a $^5hT_6$ shape with a puckering phase angle of P= 40.7°. This conformation is the global minimum with an energy slightly lower than the East minimum at P=126.9°, $^5T_8$ (ΔE=0.107 kcal/mol). The East energy region contains four pseudorotamers, ranging from ^5E to E_8.

Table 11. Puckering phase angles and relative energies from *ab initio* and CHARMM calculations

MP2/6-31G(d)			CHARMM		
P (deg)	Conf.[a]	ΔE (kcal/mol)	***P*** (deg)	Conf.	ΔE (kcal/mol)
3	E_6	0.057	4	^5E	0.121
20	$^5T_6$	0.023	20	$^5T_6$	0.010
57	$^5hT_6$	0.002	94	E_6	0.073
103	^5E	0.052	160	$^7T_6$	0.196
145	$^5T_8$	0.253	183	^7E	0.592
162	$^5hT_8$	0.419	199	$^7T_9$	0.833
177	E_8	0.763	207	E_9	1.407
193	$^9T_8$	1.354	220	$^8T_9$	1.970
199	$^9hT_8$	1.525	235	^8E	2.828
211	^9E	1.851	244	$^8T_5$	2.725
223	$^7T_9$	2.285	253	E_5	2.806
285	^8E	4.242	262	$^6T_5$	2.670
304	$^8T_9$	2.750	272	^6E	2.334
315	E_9	1.642	287	E_7	1.750
323	$^7T_9$	1.320	296	$^9T_7$	1.593
333	^7E	0.662	310	^9E	1.886
344	$^7hT_6$	0.297	320	$^9T_8$	1.711
353	$^7T_6$	0.117	335	E_8	1.249
358	E_6	0.085	351	$^5T_8$	0.451

[a]The numbers in the upper right and lower left corners correspond to the atom numbers located above or below the plane, respectively.(E), envelope; (T), twist; (hT), half-twist (Figure 30)

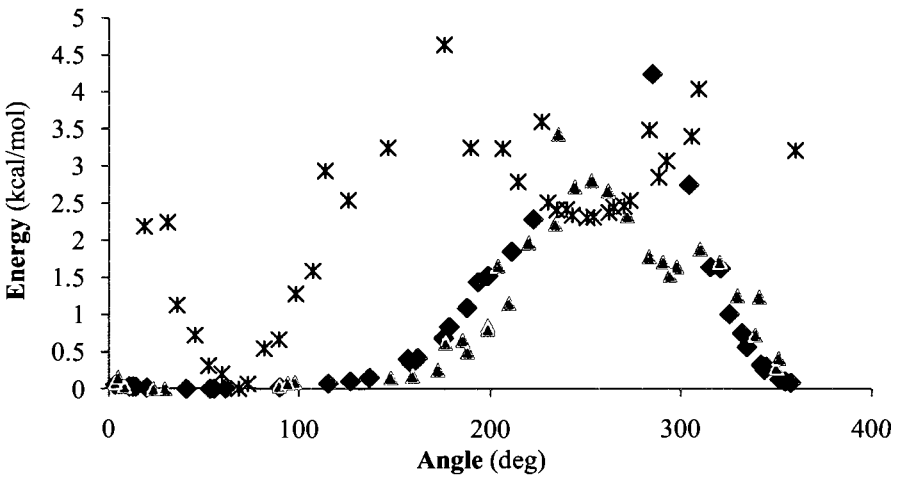

Figure 31. Energy profile along the pseudorotation angle, ***P***. Blue diamond (♦), MP2/6-31G(d); green star (*), initial CHARMM; magenta triangle (Δ), final CHARMM

The conformers calculated with the CHARMM force field also have a strong energy preference towards the North and East regions with similar puckering phases. The MM North and East conformations include three (^5E, $^5T_6$ and E_6) and four ($^7T_6$, ^7E, $^7T_9$, E_9) ring shapes, respectively. The force field calculations also indicate a preference for the North conformations. The most stable North conformation adopts a $^5T_6$ form with P=23.5° and an energy difference of 0.152 kcal/mol when compared with the East minimum. The structural parameters of the $^5T_6$ conformation optimized with both methods are shown in Table 12 and are similar for both methods.

The energy profiles of CHARMM closely parallel their *ab initio* counterparts for similar conformations, with an absolute average difference of 0.33 kcal/mol. If the conformational change of the ring compound occurs on the one-dimensional potential surface along the pseudorotation path, the pseudorotation barrier can be directly defined by the energy difference between maximum and minimum. For the CHARMM curve, the maximum occurs at a point that has two conformations with the same P value. Thus, the barrier is calculated as the average energy of these two structures (3.43 and 2.22 kcal/mol). The barriers obtained from the MP2 and CHARMM calculations are 4.24 and 2.83 kcal/mol, respectively.

Table 12. Geometric data (°) at the energy minimum

Method	P	φ_0	φ_1	φ_2	φ_3	φ_4
Twist conformation, $^5T_6$						
MP2	19.6	10.0	-33.4	-62.2	-38.1	17.2
CHARMM	23.5	10.0	-31.4	-66.5	-33.8	14.7

The fit can be further analyzed by looking at the pathways for conformational changes. For each energy minimum on the potential energy surface, the conformational energies along the pseudorotation pathway were calculated using QM and MM methods. The interconversion between two different twist conformations can proceed along the pseudorotation pathway passing through the envelope form. The barrier for this path is defined as the difference between the averaged energies of the twist conformers and the

averaged energies of the envelope forms. The calculated barriers are summarized in Table 13. Some of the barrier heights could not be reproduced exactly by CHARMM. In general, the CHARMM energies are lower compared to the *ab initio* results. The CHARMM energy barrier for the half-chair to envelope interconversion was 0.35 kcal/mol. Conformational exchange in the opposite direction gave a barrier of 0.27 kcal/mol. The MP2 value for the envelope-half-chair exchange was 0.52 kcal/mol, which is approximately twice as large as the value obtained by CHARMM. The energy barrier for the reverse process was determined to be 0.68 kcal/mol. The MP2 barriers are estimated to be 1.05 kcal/mol for the *E*-*E* path and 0.78 kcal/mol for the *T*-*T* path.

Table 13. The barriers to interconversion obtained by QM and MM methods

	Energy (kcal/mol)		
Path	MP2	CHARMM	Diff.
E to *E*	1.05	0.58	-0.47
T to *T*	0.78	0.63	-0.15
E to *T*	0.52	0.27	-0.25
T to *E*	0.68	0.35	-0.33

In conclusion, it is important to note that all parameterization was done in a self-consistent fashion. Whenever any bonded or nonbonded parameter was changed, all the targeted properties were recalculated so as to determine whether or not they were within the desired tolerance. If not, other parameters were adjusted as required, and all the targeted properties were calculated again, and so on, to converge to a set of final parameters. The presented data reflect this final self-consistently optimized parameter set.

TESTING OF THE PARAMETERIZATION

It is obvious that the validity of the simulation depends on the accuracy of the force field parameters of the potential energy function. These parameters were refined

based on calculations on the electronic structure of different conformations of the model compound. However, the question arises how well these parameters will perform when applied to a larger macromolecular system. To inspect the validity of the force field for *cp*PNA, CHARMM was used to perform energy minimization and MD simulations in explicit water and the results were compared with the *ab initio* data.

The quality of the generated parameters was checked by comparison of force-field-derived geometries with *ab initio* minimized structures. In addition to our initial model compound, we also used a smaller model compound with N-methyl amine and N,N-dimethyl amine ring substituents. All seventeen conformations of both models were minimized once with the original CHARMM using the initial parameters, and a second time using the modified CHARMM parameters. Regarding the geometry, the structure minimized with the modified force field fit the *ab initio* minimized structure significantly better than the molecule optimized with the original version. The cyclopentane rings are in different conformations for three structures (out of 17 studied): the QM optimized structures show one envelope and two different half-chair conformations for the ring, while CHARMM gives two different envelope and one half-chair conformations, respectively (Table 14).

Table 14. Conformer types found by *ab initio* and MM calculations

	Number of confs. identified		
Conformer	QM	MM Initial	MM Final
C6 up, C7 down half-chair	3	1	2
C6 up, C5 down half-chair	4	1	5
C5 up, C6 down half-chair	0	2	1
C9 up, C7 down half-chair	1	1	1
C5 down envelope	1	2	1
C6 down envelope	4	2	3
C7 down envelope	3	3	2
C5 up envelope	0	0	1
C6 up envelope	1	4	1
C8 up envelope	0	1	0

In principle the quality of the parameters should be tested by comparing relative energies of different conformations of a molecule against *ab initio* results. In total, 57 conformations were studied and the results are presented in Figure 32. These conformations were generated by changing various side-chain dihedrals of the original model. The new parameter set turned out to reproduce the *ab initio* relative energies to a reasonable precision, in contrast to the previous parameter set. The relative energies of the various conformers are reproduced within 2.2 kcal/mol. The RMS relative energies to QM are 4.1 1.32 and kcal/mol for the initial and modified CHARMM parameter sets, respectively. The correlation between the relative QM and CHARMM energies is an acceptable 0.88 (Figure 33).

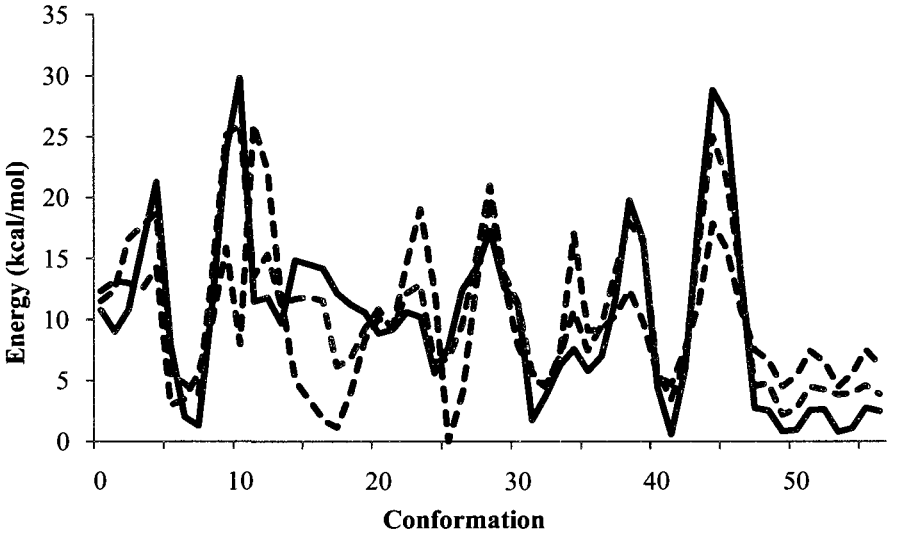

Figure 32. QM and MM relative energies of the model compound. Blue line (—), MP2/6-31G(d); green short dash (--), initial CHARMM; purple long dash (--), final CHARMM

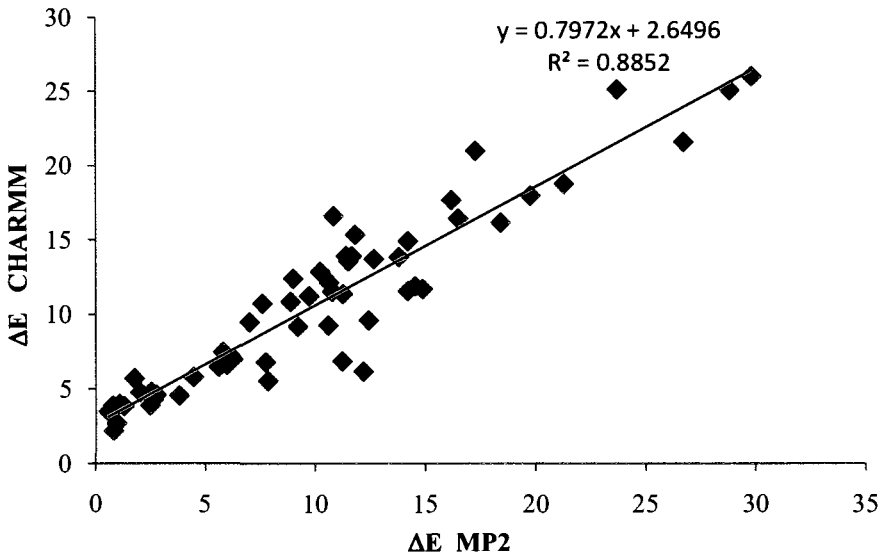

Figure 33. The correlation of the relative energies calculated by both methods: QM and MM

Solvation effects play a significant role in determining the spatial structures and dynamics of molecules.[242] Therefore, it is interesting to check the force field in simulations with explicit solvent. With this aim in view, a 10 ns MD simulation of the model molecule in a box of explicit water was calculated. The average structure was calculated from the 5,000,000 trajectory coordinate sets and overlapped with the initial structure. The mass-weighted root mean square difference between the average and starting structures was calculated and is an important indicator of the conformational flexibility of the molecule.

Analysis of the resulting MD trajectory shows stable behavior of the *cp*PNA since no significant distortions of the structure were observed: all bond lengths and bond angles fluctuated near their equilibrium values. The average RMSD of the simulation was 0.86 Å, as shown in Figure 34. A good parameter set should yield a small RMS difference from the X-ray structure; however a good value is not sufficient to conclude that a parameter set is acceptable. Exploration of the differences between the simulated and starting structures was done by examining the RMSDs of the cyclopentane ring and

peptide backbone (side-chain). It was found that the backbone region moves more than any other region during the simulation. The backbone has the largest contribution to the RMSD with an average value of 0.76 Å. As expected, the ring contributes the least to the RMSD with an average value of 0.39 Å.

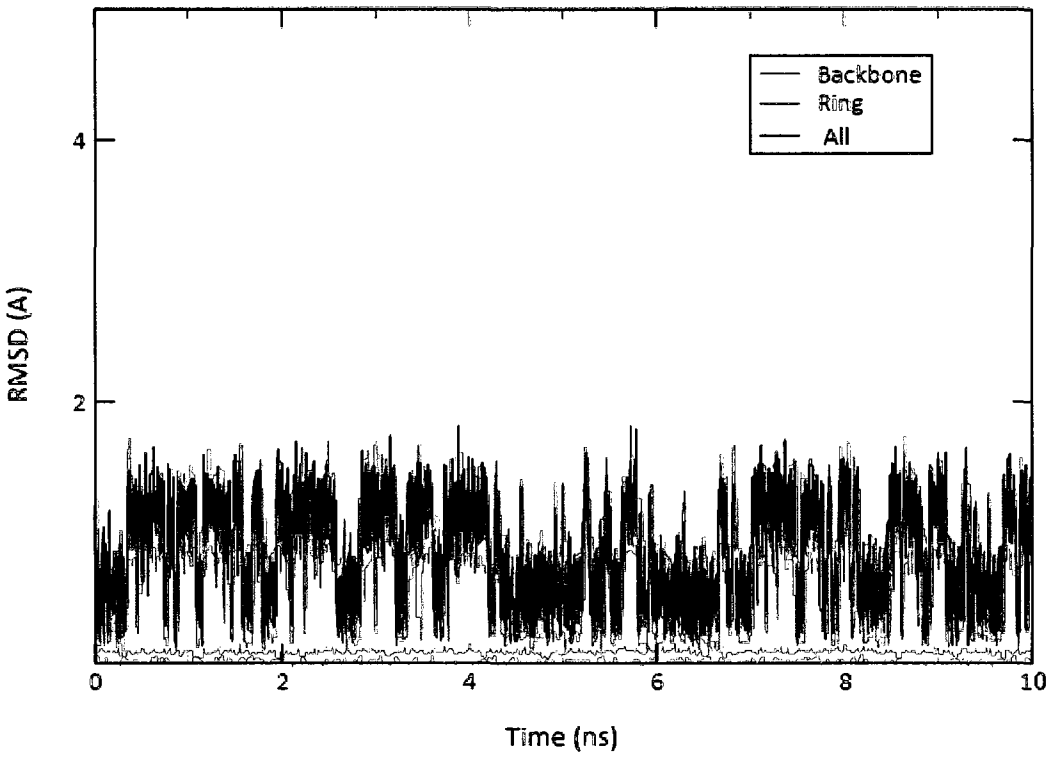

Figure 34. RMSD plot for the model compound relative to its initial structure. The ring region (red) has a stable RMSD while the backbone (black) shows greater mobility

To further characterize the structures adopted by the model molecule in the simulations, we performed conformational cluster analysis based on the backbone dihedral angles. Several conformations of the model molecule were identified and compared to the structures and corresponding relative energies obtained by QM optimizations. The more populated conformations are most likely separated from each

other by barriers and the population ratios of the clusters are proportional to the free-energy differences. The resultant clusters could give intuitive information about the underlying free-energy surface. Using a radius of 35° for clustering resulted in the location of 9 clusters. The mean dihedral angles of the different clusters along with their populations are given in Table 15. Clusters T1-T6 had higher memberships, which suggests that they are energy basins with lower energies. The molecule is in a T5 type conformer nearly 20% of the trajectory, indicating that it represents a deep basin. T7, T8 and T9 have the lowest populations, and they may be separated from other conformations by a higher barrier. The lowest-energy conformations generated by the QM procedure were compared to representatives of the most populated clusters. Table 16 shows the similarity between the computed torsion angles in the low energy conformations and the torsion angles for the clusters with occupancy ≥13%. The α dihedral angle is significantly different between the two sets. Consequently, in the simulation the model molecule was able to explore similar but not identical conformational regions as obtained by the *ab initio* calculations.

Table 15. Dihedral angles of the central conformation of each cluster

Cluster	No. of members	α	β	γ	δ
T1	7604	-49.80	-61.32	-59.40	-97.27
T2	7757	-61.42	-67.87	-66.50	-97.09
T3	6701	-59.08	-62.76	-53.95	-109.08
T4	8553	-61.26	-55.44	-67.24	-92.13
T5	10227	-70.11	-58.53	-57.24	-101.42
T6	6462	-74.88	-57.55	-69.56	-92.69
T7	2497	-87.80	-57.01	-61.66	-99.73
T8	141	-162.34	-60.85	-63.37	-96.90
T9	58	150.83	-63.06	-63.80	-97.14

Taken together, the minimization and MD results demonstrate quite a reasonable behavior of the model molecule, and, therefore, provide strong grounds to believe in the validity of the developed force field. How well will the results obtained for a simple

model system apply to biologically relevant compounds? To answer this question, we converted the force field parameters for the model compound into parameters for *X*PC residues (where *X* represents one of the four bases) and then ran explicit solvent MD simulations on *cp*PNA molecules.

Table 16. Selected dihedrals and relative energies of the model compound conformations obtained by *ab initio* calculations at the HF/6-31G(d) level of theory

Conf.	ΔE(kcal/mol)	α	β	γ	δ
1	0.899	85.25	-59.14	-75.31	-90.43
2	1.904	89.11	-96.55	-71.16	-85.14
3	0.155	139.7	-76.68	9.615	-76.81
4	1.102	162.7	-65.12	-86.25	-84.32
5	0.155	100.36	-89.83	-61.89	-100.82
6	0.000	112.2	-61.93	116.77	-81.31
7	0.156	100.3	-89.86	-61.84	-100.86

Summary

The increased role of PNA in various biological applications provides an important demand for the development of force field parameters for cyclopentyl modified PNA molecules suitable for use with the well established CHARMM force field. In order to use molecular dynamics simulations to determine the effects that govern the flexibility of *cp*PNA molecules, this class of compounds has been explicitly parameterized. An iterative process of refinement was used to develop the parameters for cyclopentyl PNAs, while remaining compatible with the regular PNA parameters. A complete set of the CHARMM force field parameters was developed for *cp*PNA molecules based on *ab initio* calculations, which properly account for the five-membered ring modification present in the backbone of these compounds. The reliability of this new force field was proved by molecular mechanics calculations and MD simulations.

CHAPTER IV

STUDIES WITH SINGLE STRANDED PNAs

INTRODUCTION AND MOTIVATION

Rationally designed PNAs that show complementary binding exclusively to RNA or DNA will aid in the discovery of new molecular biology applications, and drug candidates. The rules governing PNA:DNA/RNA complexes are not well understood because the interplay between the hydrogen bonding, stacking and conformational mobility has yet to be defined. However, limited progress in creating PNAs with improved binding and better selectivity for RNA and DNA was achieved by correctly pre-organizing the PNA structure using synthetically accessible modifications.[244] In order to increase the PNA binding affinity for complementary oligonucleotides, numerous PNA backbone modifications have been investigated during the last decade. Recently, Appella's group reported a novel modified PNA for detection purposes.[164] To enhance the oligonucleotide binding affinity, the authors introduced a *trans*-1,2-cyclopentane diamine into the PNA backbone. The thermal denaturation studies revealed that the sensitivity of these new PNA structures to binding DNA/RNA was improved by three orders of magnitude.[181] The *cp*PNA strands showed improved binding affinity regardless of nucleobase or position modified. In contrast to singly modified strands, PNAs with numerously modified residues showed better binding affinity toward DNA, suggesting that the effect of increased T_m was additive from a single favorable modification to quadruply modified oligomer sequences. The *trans*-cyclopentane modification was the first and, to date, only modification that provided a consistent additive increase in thermal stability when incorporated into PNA oligomers. The thermodynamic properties of *aeg*PNA:DNA and *cp*PNA:DNA duplexes were also investigated. Van't Hoff enthalpies and extrapolated ΔG values determined from the UV melting studies on the duplexes indicated that the increased thermodynamic stability of the *cp*PNA:DNA hybrid is due to an entropic contribution.[158,161,163]

Our calculations extend these studies, and the results have provided a picture of the flexibility of single strands in explicit water on the nanosecond time scale and a physical interpretation for the origins of the relative stabilities of the PNA:DNA/RNA hybrids. They also provide insight into the changes in PNA structures, induced by the cyclopentane backbone modifications, which might be important for understanding the improved binding of this chemically modified oligonucleotide to complementary DNA/RNA.

*Cp*PNA was designed to optimize the dihedral angles of the PNA backbone for duplex formation. When a modification is made to PNA to make the backbone more rigid, helix formation should now be more entropically favorable (ΔS_t). However, the modification could also change the conformational preferences of the strand (ΔH_t). The modifications to PNA that improve its binding obviously have decreased the unfavorability of the entropy. As long as the modification doesn't make the torsional strain worse, the increased rigidity will make the binding better. The question is: have they also improved the torsional potential energy surface (ΔH_t)?

The torsion angles around the bonds of the sugar-phosphate DNA/RNA backbone greatly influence the secondary structure as well as base-base recognition during the double helix formation. From the inspection of the backbone dihedrals, one can ascertain if the modified PNAs should be sufficiently malleable for efficient stacking and Watson-Crick base pairing. The more the nucleotide is pre-organized in a helical conformer closely resembling that of a bound natural oligomer (DNA/DNA or RNA/RNA), the higher should be the stability of the duplex formed between the modified oligomer and its DNA/RNA complement. The stability of partly modified PNA:DNA/RNA duplexes may be explained as follows. The constrained conformation of the PNA nucleotide locally organizes the peptide backbone (decreased loss of entropy upon duplex formation) in the direction of a favorable conformation for efficient stacking of the nucleobases in the duplex (increase in enthalpy upon formation). Thus, the formation of a PNA:DNA/RNA duplex should be favored by both enthalpy and entropy compared to the corresponding DNA/DNA or RNA/RNA duplex. A straightforward consequence of this binding pattern is that the increase in the stability of modified PNA:DNA/RNA duplexes should not saturate as the number of modifications increases.

In the study presented here, we performed detailed investigations on the torsional preferences of *aeg*PNA and *trans-cp*PNA monomers using molecular dynamic techniques which gave insights into the effect of the rings on the PNA's behavior. This work will reveal if these modifications induce pre-organization of the PNA oligomers into a helix for optimal DNA/RNA binding, and whether these modifications cause changes in conformational entropy. This dissertation focuses on several aspects of PNA structure and dynamics: (i) the stability of the MD simulations, (ii) the conformational sampling of the modified regions and their flexibility, (iii) the characterization of monomer dynamics and the changes due to the presence of a charged terminal Lys residue, (iv) single and (v) multiple cyclopentane modifications at various positions in the strand.

Since in RNA/DNA chemistry, stabilization is greatly influenced by the sequence composition, it is important to take into account the effects of nucleobase composition. To assess this both the energetics and flexibility of the classical and modified PNA backbones, were studied for two sequences: polythymine (T_1-T_2-T_3-T_4-T_5-T_6-T_7-T_8) and mixed base (G_1-T_2-A_3-G_4-A_5-T_6-C_7-A_8-C_9-T_{10}). These are the two most common sequences used to study PNA modifications experimentally.[102,129,144-152,156-163,173-181]

Chiral PNAs containing positively charged residues, such as *D*- or *L*-Lys-based monomers, can be used for modulating the directionality of the binding, the preferential handedness of the helices and the stability of the PNA:DNA duplexes.[91,92] Based on circular dichroism studies and melting experiments, it was proposed that *D*-Lys containing PNAs promote the formation of right-handed structures when bound in the antiparallel mode, whereas *L*-Lys containing PNAs give rise to left-handed structures.[245] In parallel PNA:PNA duplexes, the reversed helicities were observed.[246] In agreement with this preference, the *D*-Lys containing chiral PNAs were found to form more stable complexes than their *L* counterparts when bound to antiparallel DNA,[247] which is always right-handed. Therefore, we tested the effect of the lysine residue on *aeg*PNA and *cp*PNA conformations, and its ability to govern the trends in preferential binding to complementary DNA strands. Last, the effect of single and multiple ring modifications at various positions was also analyzed. The complete list of structures studied is presented in Table 17.

Molecular dynamics simulations are a special discipline of molecular modeling. It is used as an important tool for structural determination and refinement in crystallography and NMR. However, more significantly, just as its name implies, MD can be used to provide dynamic information, in other words, time-dependent properties of molecular systems. This type of information is not available from experimental structures. Conformational transitions are the usual subject of molecular dynamics studies. Therefore, in this work, we utilized molecular dynamics simulations to facilitate our study on classical and modified PNAs.

Table 17. The numbering scheme and the modified oligonucleotides used

Name	Sequence (PNA-T_8)	Name	Sequence (PNA-MB_{10})
*aeg*PNA**1**	T-T-T-T-T-T-T-T	*aeg*PNA**12**	G-T-A-G-A-T-C-A-C-T
*cp*PNA**2**	T-T-T-T*-T-T-T-T	*cp*PNA**13**	G-T-A-G-A-T*-C-A-C-T
*aeg*PNA**3**	T-T-T-T-T-T-T-T-Lys	*aeg*PNA**14**	G-T-A-G-A-T-C-A-C-T-Lys
*cp*PNA**4**	T-T-T-T*-T-T-T-T-Lys	*cp*PNA**15**	G-T-A-G-A-T*-C-A-C-T-Lys
*cp*PNA**4a**	T-T-T-T*-T-T-T-T-(D)-Lys	*cp*PNA**15a**	G-T-A-G-A-T*-C-A-C-T-(D)-Lys
*cp*PNA**5**	T*-T-T-T-T-T-T-T-Lys	*cp*PNA**16**	G-T-A-G*-A-T-C-A-C-T-Lys
*cp*PNA**6**	T-T-T-T-T-T-T-T*-Lys	*cp*PNA**17**	G-T-A-G-A-T-C*-A-C-T-Lys
*cp*PNA**7**	T*-T-T-T*-T-T-T-T*-Lys	*cp*PNA**18**	G-T-A-G-A-T-C-A*-C-T-Lys
*cp*PNA**8**	T*-T-T-T-T-T-T-T*-Lys	*cp*PNA**19**	G-T-A-G-A-T*-C*-A-C-T-Lys
*cp*PNA**9**	T-T-T-T*-T-T-T-T*-Lys	*cp*PNA**20**	G-T-A-G-A-T*-C-A*-C-T-Lys
*cp*PNA**10**	T*-T-T-T*-T-T-T-T-Lys	*cp*PNA**21**	G-T-A-G-A-T-C*-A*-C-T-Lys
*cp*PNA**11**	T*-T*-T*-T*-T*-T*-T*-T*-Lys	*cp*PNA**22**	G-T-A-G-A-T*-C*-A*-C-T-Lys

* indicates residue with *trans*-(*S,S*)-cyclopentane modification

One of the objectives of this study is to understand the differences in conformations and conformational dynamics between single stranded *aeg*PNA and cyclopentyl PNAs in various sequence contexts. To investigate these, the backbone dihedral angles were examined (Figure 35). As already briefly described in Chapter 2, statistical clustering based on dihedral angles was performed. For each sequence and modified variant, the conformational space is explored during the simulation. Analysis of the percentage of cluster occupancies provides information on the stable conformation for

each sequence/modification. Conformational cluster analysis will show if the modifications to the backbone cause the single-stranded PNA to pre-organize into a helical structure that closely resembles the PNA conformation in PNA/DNA or PNA/RNA double helix. Additionally, whether the pre-organization correlates with high binding affinity to DNA/RNA can be determined.

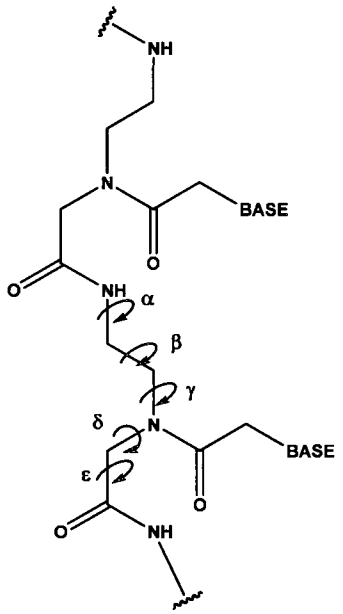

Figure 35. Notations for the backbone torsion angles

THEORY AND METHODOLOGY

Conformational diversity concept

Conformational analysis is the characterization of the structures that a molecule can adopt and how these influence its properties. A key component of a conformational analysis is a conformational search, the object of which is to identify the preferred conformations of a molecule, which usually are minima on the potential energy surface.

Due to PNA's high conformational flexibility, it has such a large number of minima that it is impractical to characterize them all. This conundrum is known as the multiple minima problem[248] and is the main difficulty in structurally characterizing a peptide. Specifically, most of the peptides under physiological conditions exist as a mixture of interchangeable conformations with similar energies, populated according to the Boltzmann distribution. The conformation with the lowest potential energy is referred to as the global minimum. Although the global minimum exhibits the lowest energy value, it may not be highly populated because of the contribution of the vibrational entropy to the statistical weight of each structure. Moreover, the global minimum may not be the active (i.e. the functional) structure. Indeed, in some cases, the active conformation may not correspond to any minimum on the energy surface of the molecule.

Experimentally the crystal structure represents only a snapshot of the average structure of the most populated ensemble in the solid state, which may or may not correspond to the lowest energy conformer. Computational methods constitute an alternative approach to determine the active conformation. The more populated this conformation, the lower the configurational entropy loss upon binding, since the loss in flexibility upon binding will be smaller. This conclusion is supported by the fact that conformationally constrained analogs exhibit higher affinity than more flexible compounds.

Computational conformational search methods can be divided into the following categories: systematic search algorithms, model-building methods, random approaches, distance geometry and MD. Independent from the strategy selected, four key elements are needed to carry out the exploration of the conformational space of an oligonucleotide. The first consists of employing a peptide/nucleic acid model description based on classic mechanics, i.e. a force field that permits the calculation of the energy of a conformation. The second is to find a method capable of generating different conformations in order to explore all the low energy regions of the conformational space. The third key element consists of minimizing the different conformations, whereas the fourth and last element is to find a convergence criterion to assess if the conformational space has been sufficiently explored. However, in many biopolymer applications, conformational space cannot be

exhaustively searched and a representative subset of conformational space is sampled instead.

Sampling conformations of biomolecules through computational techniques is a key step toward understanding molecular function. Computational methods based on molecular dynamics can now routinely identify energetically predominant polymer structures evident over nanosecond (ns) to microsecond (μs) trajectories. Unfortunately representative sets of structural conformations for large biomolecules are not sampled completely by these trajectories due to the complexity of the configuration space and energetic barriers that localize sampling. Elevated temperature simulations provide a higher probability that these energy barriers will be surpassed; however the resulting energy surface is not necessarily representative of the biologically relevant lower temperature regime.

Cluster analysis

Computational techniques produce large numbers of conformations of a given chemical structure and are used to find a particular quantity of interest (e.g., the lowest-energy conformation or the free energy of the system). After performing a conformational search, it is of interest to know whether the conformations found form groupings of structurally related clusters or if they are distributed randomly throughout conformational space. Following a dynamics simulation, one can investigate how many structurally distinct classes were visited during the simulation, as well as the time sequence. The conformations sampled are usually sorted according to two properties: the potential energy associated with the conformation and its distance from other molecular conformations. The clustering method is presented in detail in Ref. 243 and in Chapter 2.

System setup and computational procedures

All the calculations and analyses were carried out within the CHARMM package and the newly developed force field parameters for PNAs which are summarized in

Chapter 2. To model the water molecules the TIP3P model was used, one of the most widely used water models in biomolecular simulations. A summary of simulated systems and sequences is reported in Table 17. The crystal structure of PDBID:1PUP was used for the starting structure of the poly-T_8 PNA simulations. The initial configurations for the mixed base decamer simulations were taken from the PDBID:1NR8 structure. The initial solute conformations of all the systems containing cyclopentane fragments were built *in silico* using the SYBYL program. The system setup procedure was started by adding hydrogen atoms using the HBUILD algorithm. The structure was minimized using 500 cycles of steepest descent followed by minimization with the conjugate gradient method. For each simulation the solute was placed at the center of a periodic truncated octahedral box large enough to avoid any interactions between mirror images of the PNA strand. The solvent box was built from a cubic periodic configuration of a smaller box of pre-equilibrated water molecules. Then, the system was minimized fixing the coordinates of the PNA and allowing the water molecules to move.

MD simulations were initialized from the energy-minimized configurations with atomic velocities taken from the Maxwell-Boltzmann distribution at 60 K, while the PNA atoms were restrained using a harmonic potential. Each system was then gradually brought to the desired temperature of 298 K during 100 ps of MD simulations. The electrostatic interactions were calculated by the Particle Mesh Ewald (PME) algorithm, with an interpolation order of 4 and a grid spacing of 1.0 Å. All simulations were run under periodic boundary conditions in the NVE ensemble. The SHAKE algorithm with a tolerance of 10^{-5}Å was applied to constrain the stretching of all bonds containing hydrogen atoms. The time step for the simulations was 2 fs and the trajectory coordinates were stored every 0.2 ps for analysis. The van der Waals (vdW) forces were treated by using a cutoff of 11 Å and a switching function. The neighborlist was generated up to a cutoff of 13 Å and updated whenever any atom had moved more than 1 Å since the last update.

The trajectories were first analyzed by evaluating the root-mean-square fluctuations (RMSF) and deviations from the crystal structure (RMSD) for the backbone. The neural network based clustering algorithm ART-2 included with the CHARMM distribution was used for cluster analysis of the trajectory. A cluster analysis allows one

to classify the frames of the MD trajectory formally into a number of groups which physically correspond to local minima on the potential energy surface. The MD trajectory could be represented therefore as travelling a pathway through potential energy 'basins'. In the cluster analysis, a set of descriptors is chosen to describe the clusters, in this case, the backbone torsion angles. Then, the MD trajectory is used to calculate a time series of the backbone torsion angles. Finally, the trajectory structures are clustered into batches of similar configurations using the backbone RMSD of the dihedral angles as similarity criterion.

Our clustering plan began with a pair of parameters: the set of torsion angles and a cluster radius. For the results presented in this section, three clustering schemes were chosen after evaluation of the dependence of the cluster populations and the total number of clusters on these two criteria. Due to complexity of the dihedral space sampled, convergence among the clustering schemes was not observed. Thus, results from three different schemes are presented. The following three clustering schemes were employed: (1) a cutoff radius of 55°, use of every other backbone dihedral excluding the terminal residues, and use of every 500-th frame of the trajectory (10000 frames total, Scheme I), (2) a cutoff radius of 50°, every other backbone dihedral excluding the terminal residues, and every 1000-th frame of the trajectory (5000 frames total, Scheme II); and (3) a cutoff radius of 55°, all the backbone dihedrals excluding the terminal residues, and every 500-th frame of the trajectory (10000 frames total, Scheme III). The clusters from each scheme are checked for uniqueness. Those for which all backbone dihedral angles are different by less than 55° with respect to a cluster from another scheme are not considered unique. The clusters that were located by more than one scheme are listed in a separate table along with the population percentages and Scheme I cluster numbers. A structural snapshot that is closest to the final cluster center was selected as a representative of each cluster. For each PNA strand, the dihedral angle values from the most populated clusters were compared with the average dihedral values of native DNA and RNA duplexes.

Results and discussion: stability of the simulations

The MD of single helical *aeg*PNA and modified *cp*PNA monomers with and without a terminal lysine residue were simulated for 10 ns each. To assess the stability of all the MD simulations, a collection of properties was monitored as a function of time. The atom-positional root mean square deviations with respect to the initial structure as a function of simulation time are shown in Figures 36 and 37.

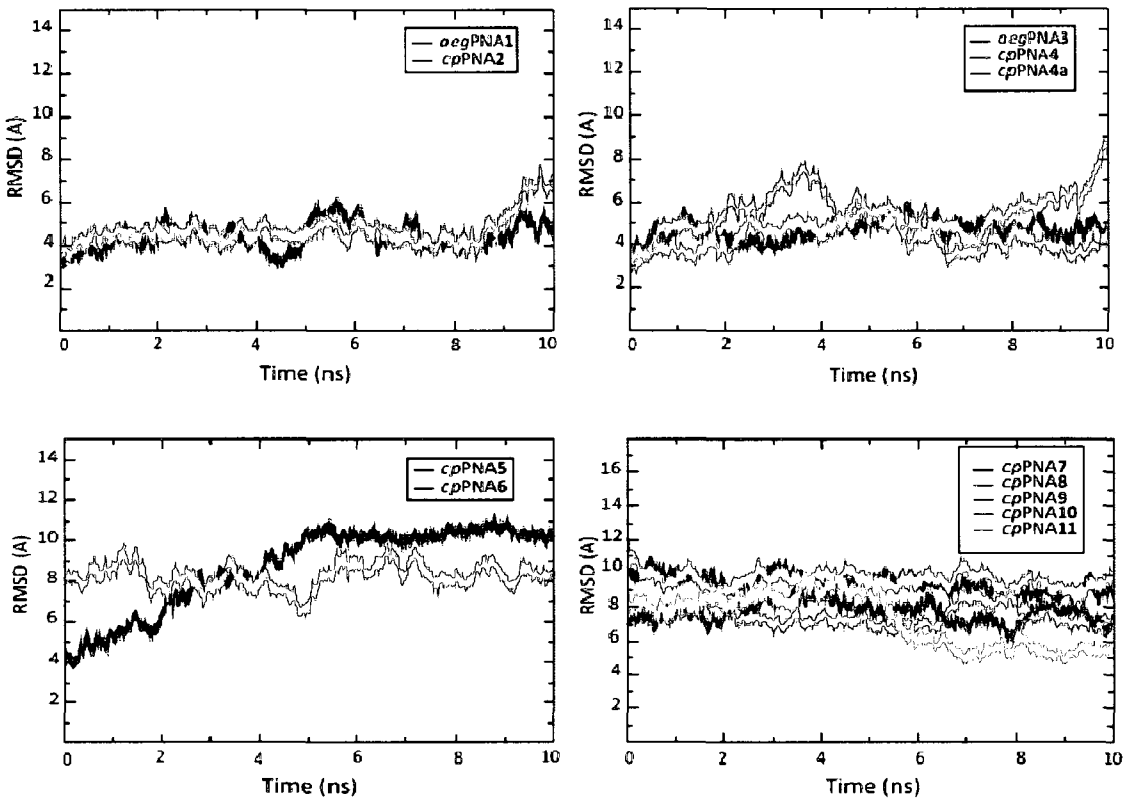

Figure 36. RMSD of the backbone atoms with respect to the starting structure as a function of time for the PNA-T_8 simulations

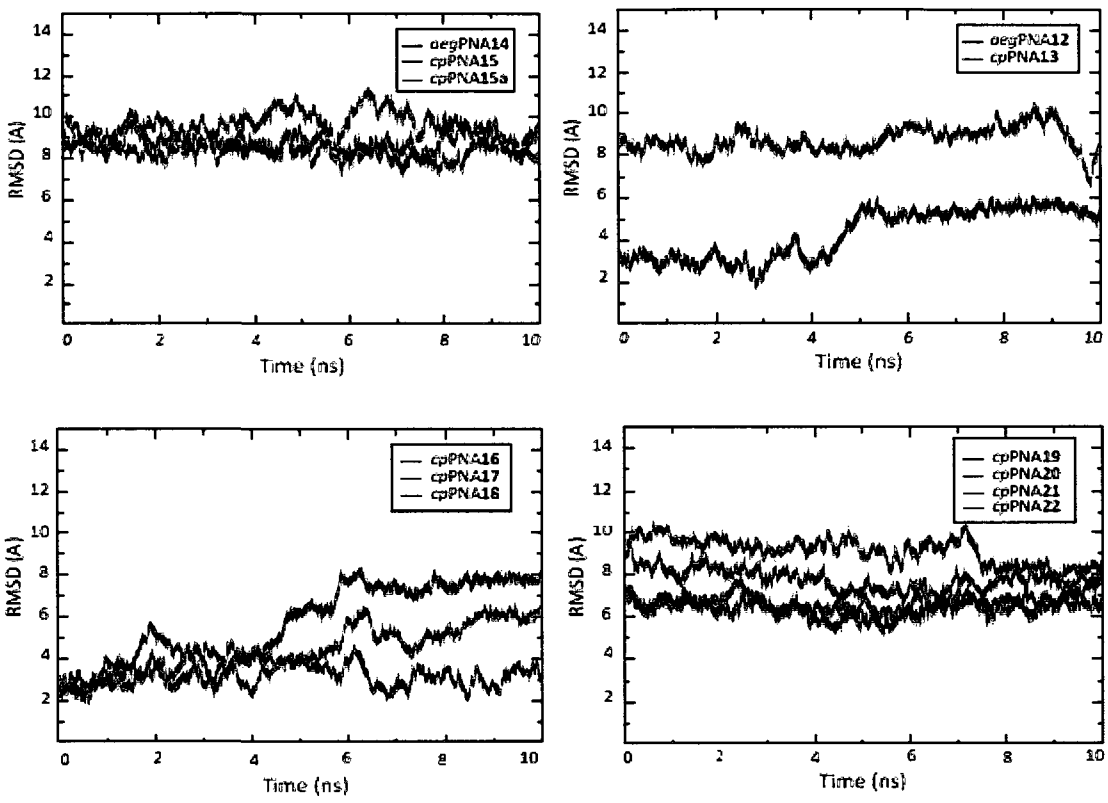

Figure 37. RMSD of the PNA backbone atoms with respect to the initial structure as a function of time for the PNA-MB_{10} simulations

The RMSD of the atomic positions was calculated according to the expression:

$$RMSD = \left(\frac{1}{N}\sum_{i=1}^{N}\langle \Delta r_i^2 \rangle\right)^{1/2} \tag{15}$$

where N is the number of atoms, Δr_i is the difference between the instantaneous and starting position for the ith atom, and the angular brackets indicate a time average over the simulation time. Because PNA is a flexible molecule, it can adopt several important conformations or conformational families. This is confirmed by the high RMSD between the starting X-ray and NMR configurations, which was calculated to be ~10 Å.

The radius of gyration (RGYR) was also calculated throughout the MD trajectories. The variation in the RGYR reports on molecular compactness. The RGYR of a group of atoms is defined as the root-mean-square distance from each atom of the molecule to their centroid

$$RGYR = \frac{1}{N} \langle \sum_{j=1}^{N} (r_j - r_{mean})^2 \rangle^{1/2} \quad (16)$$

where r_j is the position coordinate of atoms j, r_{mean} is the mean position of the atoms and N is the number of atoms.

Figures 38 and 39 show RGYR plotted against time. For most of the systems these values are stable along the 10 ns periods and fluctuate in the range of 8-15 Å. The largest RGYR, 14.85, was observed for the *cp*PNA**19** simulation at time point 5 ns. The obtained results and analysis of the trajectories show that the RGYR values for some extended states are similar to that of the "folded" molecule (~8 Å). This indicates that the folded and some extended states are similarly compact; therefore, RGYR per se is a necessary, but not sufficient requirement, for a folded structure.

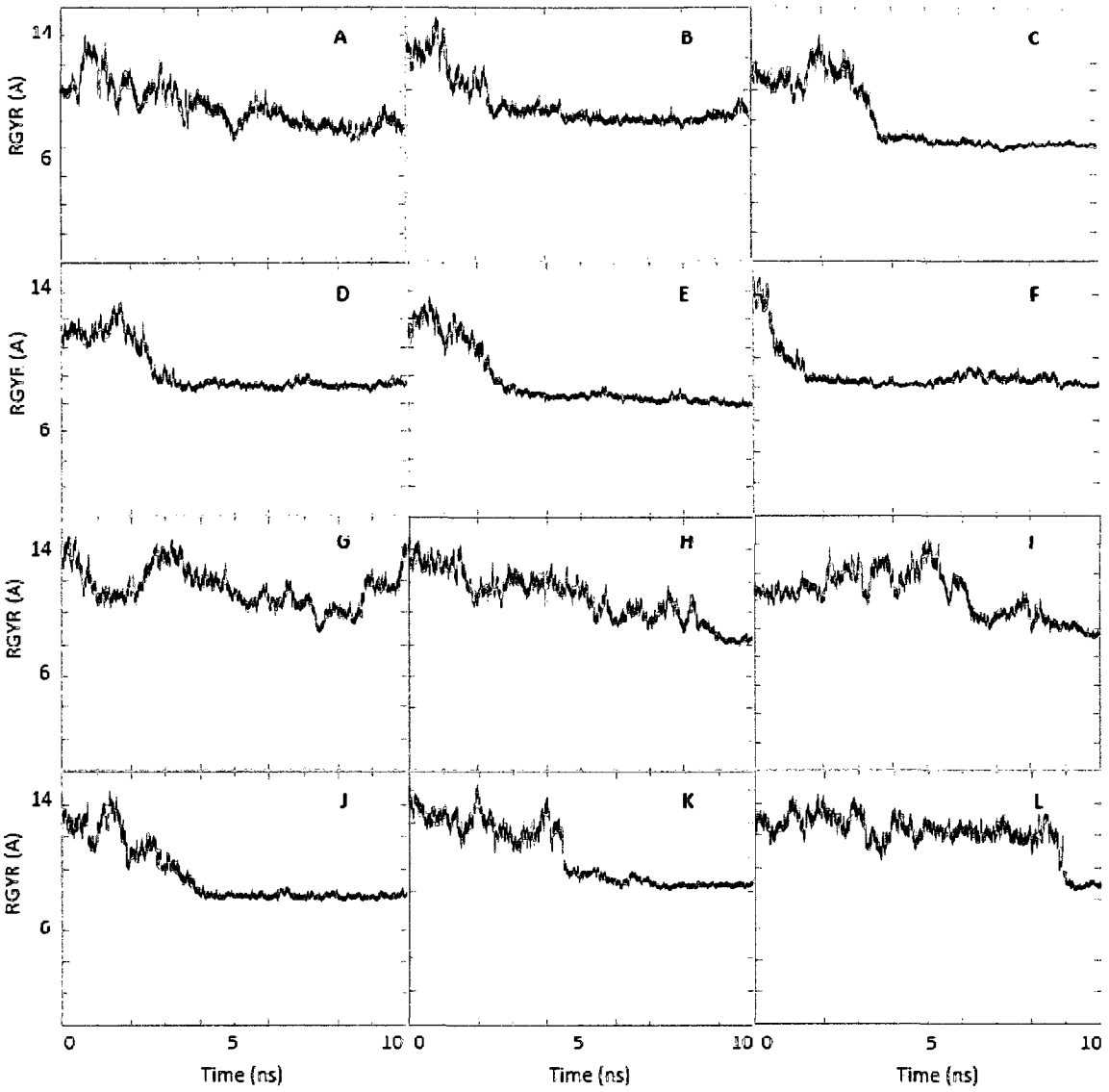

Figure 38. Radius of gyration of the octamer (PNA-T_8) structures plotted against time: *aeg*PNA**1** (A), *cp*PNA**2**(B), *aeg*PNA**3** (C), *cp*PNA**4** (D), *cp*PNA**4a** (E), *cp*PNA**5** (F), *cp*PNA**6** (G), *cp*PNA**7** (H), *cp*PNA**8** (I), *cp*PNA**9** (J), *cp*PNA**10** (K), *cp*PNA**11** (L)

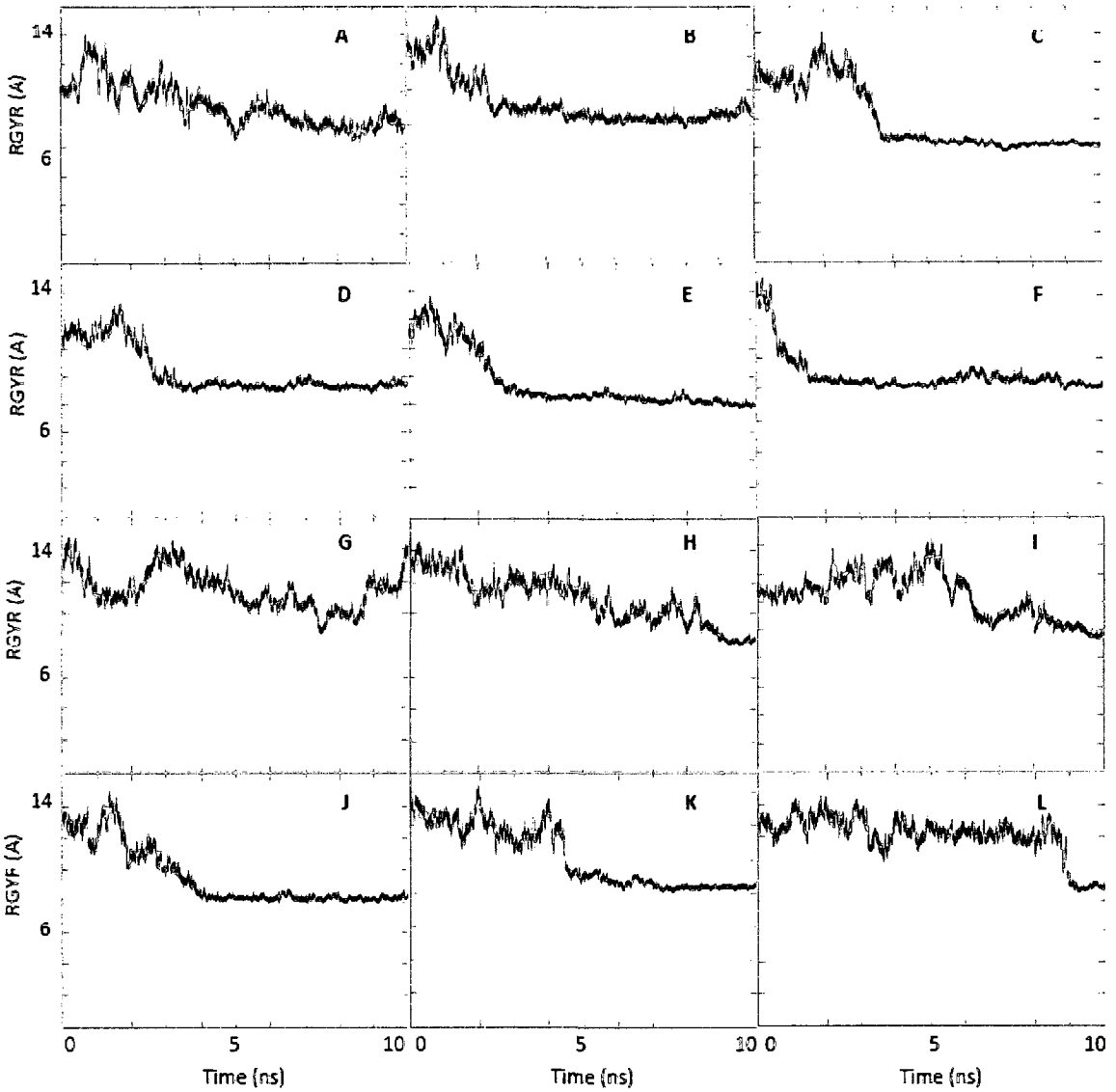

Figure 39. Radius of gyration of the decamer structures (PNA-MB_{10}) plotted against time: *aeg*PNA**12** (A), *cp*PNA**13** (B), *aeg*PNA**14** (C), *cp*PNA**15** (D), *cp*PNA**15a** (E), *cp*PNA**16** (F), *cp*PNA**17** (G), *cp*PNA**18** (H), *cp*PNA**19** (I), *cp*PNA**20** (J), *cp*PNA**21** (K), *cp*PNA**22** (L)

EFFECT OF A SINGLE CYCLOPENTANE MODIFICATION

PNA-T_8 sequence: *aeg*PNA**1** and *cp*PNA**2**

The ability to adopt a particular secondary structure is important in the binding process of biopolymers. The possibility that the free, single-stranded structure pre-organizes into the bound conformation before binding to its complement, which affects the affinity and rate of oligonucleotide hybridization, must also be considered. Thus, we examined the simulation trajectories in the context of structural features (e.g. turn, helix, loop, knot, coil, hairpin, etc.) and backbone conformations (Figure 40). Such properties were assigned to each cluster. Helical structures represent one of the most common structural motifs and recognition sites in nucleic acids. Therefore, we focused on identifying helical elements because the monomer needs to adopt a helix to bind DNA/RNA.

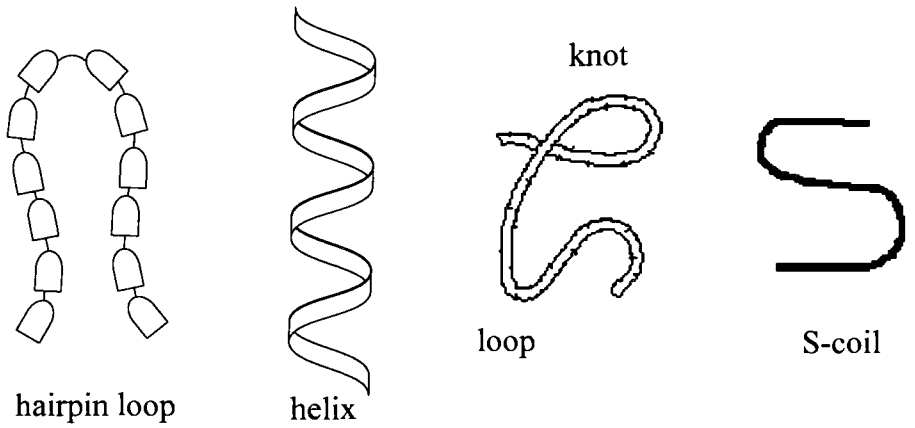

Figure 40. Some representative backbone structures observed

The 10 ns MD trajectory of *aeg*PNA**1** was analyzed. We computed the RMSD deviation over the course of the MD simulation, relative to the first set of coordinates from the production MD simulation. During the first 3 ns, the RMSD values increased to

~5 Å (Figure 36). During this time, the molecule underwent conformational changes, but remained relatively stable for the rest of the simulation. These RMSD fluctuations indicate the significant flexibility of the structure. The process of conformational change of the PNA can be analyzed in view of the evolution of patterns sampled along time. Since the simulation started from the helical conformation from the PNA/DNA duplex, the first patterns sampled correspond to structures with no conformational motifs or adopt the dihedral angles of a helix.

To characterize the conformational states of the molecule, cluster analysis was conducted using the three criteria mentioned previously and the data is summarized in Tables 18 and 19. We first present properties of the entire ensemble of structures sampled, followed by a more detailed discussion of specific preferred conformations. The process yielded sixteen distinct clusters. There are no significant differences in the predominant structures obtained using all three criteria. Four clusters were identical between Schemes I and II, and the remaining three clusters from Scheme I were replaced by six new clusters in Scheme II. Cluster **3**, the most populous cluster of the three was similar to cluster **10**, the most populous cluster of the six. Five of the clusters from Scheme I were also located using the parameters of Scheme III, though the population percentage of cluster **3** was much lower using Scheme III. The remaining clusters of Scheme I were replaced by three new clusters in Scheme III. None of the new clusters from Scheme II matched those of Scheme III.

Cluster **1** (11.6 % of the total number of structures), which was also located with Schemes II and III, contains a single pattern that exhibits no conformational motifs, though remnants of the starting helical structure remain. The remaining clusters (**2**, **3**, **4**, **5**, **6**, **7**) from Scheme I, correspond to a conformation representing an S-type coil. The matching clusters in Schemes II and III as well as clusters **11**, **12** and **15** also have an S-type coil structure. Cluster **10** (18.7%) exhibits a hairpin-like motif running from residues 3 to 8. The remaining clusters have low populations and most likely do not contribute significantly to the conformational preference of the molecule. Figure 41 shows some characteristic conformations of the 8-mer.

Table 18. Average torsion angles for each cluster from the *aeg*PNA**1** simulation

Cluster	No. of members	%	T2		T3			T4		T5			T6		T7		
			β	δ	α	γ	ε	β	δ	α	γ	ε	β	δ	α	γ	ε
Scheme I																	
1	1163	**11.6**	176	86	173	96	-100	-179	81	-178	84	73	64	84	178	81	-50
2	643	**6.4**	-160	86	167	108	49	173	81	170	75	-1	112	93	-170	96	-140
3	2446	**24.5**	-169	97	166	99	68	179	108	-173	93	-49	171	-99	-177	97	-113
4	1851	**18.5**	-162	88	167	103	81	-174	98	-163	92	50	169	-90	-178	94	-56
5	1385	**13.9**	-167	87	159	106	74	-173	88	-159	91	66	178	-93	-178	98	155
6	877	**8.8**	-159	86	160	104	-108	177	134	-170	92	21	175	-99	-176	98	48
7	1635	**16.4**	-165	88	172	106	80	180	112	-167	93	-40	171	-105	175	87	33
Scheme II																	
8	206	**4.1**	-162	85	166	107	27	170	80	173	76	66	64	95	-168	92	-111
9	160	**3.2**	-160	86	168	109	53	178	82	165	76	-71	163	90	-173	98	-173
10	937	**18.7**	-175	86	160	105	71	-180	114	-164	93	-12	170	-99	172	100	-154
11	661	**13.2**	-161	100	173	98	79	179	104	-175	92	-55	171	-99	-173	93	-65
12	411	**8.2**	-162	87	158	108	78	-170	72	-159	92	82	-175	-91	-164	98	117
13	124	**2.5**	-164	127	170	81	-49	172	100	175	86	-71	170	-99	-169	86	-63
Scheme III																	
14	709	**7.1**	-164	126	173	85	44	172	89	166	82	-82	168	-92	-164	87	-86
15	4738	**47.4**	-164	87	169	105	78	-179	111	-164	93	-9	169	-101	175	95	-66
16	406	**4.1**	165	84	156	105	81	179	125	-166	95	-33	172	-99	174	92	163

Table 19. Similar clusters identified from the different clustering Schemes for *aeg*PNA**1**

Cluster	No. of mem	%	Scheme I Cluster #[a]	Cluster	No. of mem	%	Scheme I Cluster #[a]
Scheme II				*Scheme III*			
A	541	**10.8**	1	A	1063	**10.6**	1
B	875	**17.5**	4	B	761	**7.6**	2
C	391	**7.8**	6	C	593	**5.9**	4
D	694	**13.9**	7	D	855	**8.6**	5
				E	875	**8.8**	6
Total number of clusters			10				8

[a] numbers correspond to identical cluster from Table 18

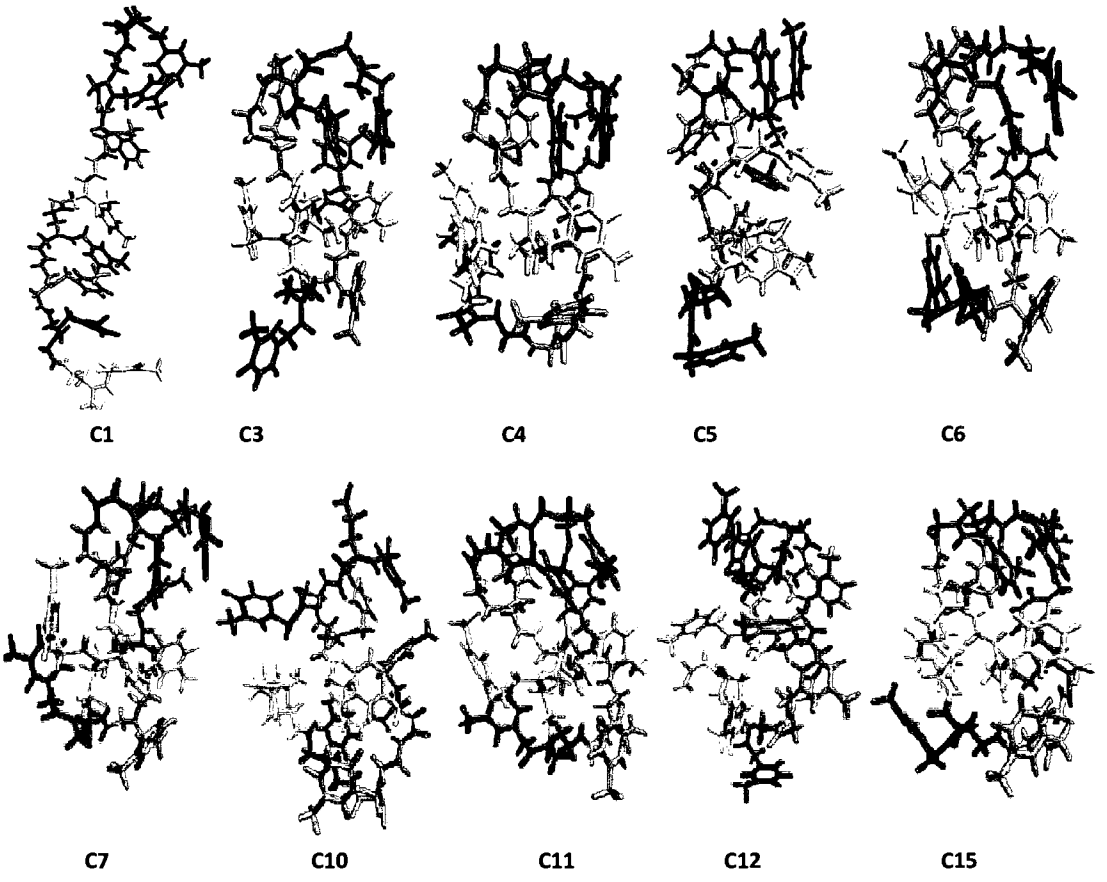

Figure 41. Representative structures of the most populated clusters from the MD simulation of *aeg*PNA**1** using the different clustering Schemes. Residues are colored by position: **T1**-red, **T2**-dark grey, **T3**-orange, **T4**-yellow, **T5**-dark yellow, **T6**-grey, **T7**-green, and **T8**-white. The structures are positioned with **T1** on top and **T9** on bottom

Therefore, the conformational state of the *aeg*PNA**1** is not entirely random and is represented by a few stable conformational ensembles. The results also show that the *aeg*PNA conformation is very flexible and not pre-formed for oligonucleotide binding, since none of the observed conformers are remotely helical in nature.

A more detailed description of the flexibility of the *aeg*PNA**1** residues was investigated by calculating the root mean square fluctuation (RMSF) of each backbone dihedral (Table 20). RMSF is a good measure of the flexibility of the system during a given time period, and can provide valuable information about the deviations of the torsions involved in binding and about variations in the secondary structure with respect to the starting conditions. We also averaged the RMSFs over all residues including the terminal residues. However, this method is a very qualitative measure of PNA flexibility, but it can give us a view of the conformational dynamics.

Table 20. Average dihedral values (°) and RMSF (bold, °) of the simulated *aeg*PNA**1**

Residue	α	β	γ	δ	ε	Total RMSF	Average RMSF
T1	—	—	9.4	109.7	86.6		
	—	—	**91**	**21**	**48**	**160**	**53**
T2	166.1	-166.9	-95.7	89.6	57.0		
	21	**18**	**83**	**18**	**33**	**173**	**35**
T3	166.6	-48.8	102.7	95.3	68.5		
	21	**34**	**13**	**16**	**84**	**168**	**34**
T4	-178.0	-178.7	71.5	101.3	-37.0		
	22	**13**	**54**	**24**	**64**	**177**	**35**
T5	-169.3	87.2	90.2	-151.3	10.0		
	23	**40**	**11**	**80**	**62**	**216**	**43**
T6	-175.8	155.9	-45.7	-64.0	133.7		
	27	**41**	**79**	**73**	**143**	**363**	**73**
T7	-178.7	176.5	93.1	136.7	18.9		
	29	**52**	**17**	**79**	**118**	**295**	**59**
T8	178.6	177.6	40.3	—	—		
	28	**13**	**83**	—	—	**123**	**41**

In *aeg*PNA**1**, T_6 and T_7 which are the least involved in base stacking interactions, were the most flexible of all the residues. The stacking interactions between T_1, T_2, T_4 and T_5 are present for most of the simulation time and are also present in the central structures of the most populated clusters. The simulation yielded mean values in the range of 155 to 177° for β (excluding T_3 and T_5), 70 to100° for γ (excluding T_1, T_2, T_6 and T_8), 90 to 130° for δ (excluding T_5 and T_6), and 166 to -178° for α. The ε torsion varied greatly during the simulation and no specific range could be identified.

Starting from the conformation found in the PNA:DNA duplex, a 10 ns MD simulation of *cp*PNA**2** was conducted to address the effect of a single cyclopentane modification on the conformational dynamics (structural sampling). It appears that the PNA fluctuated around an average until it underwent a significant conformational change at ~8 ns. The RMSD remained at ~4 Å during the first 9 ns, before increasing to 7 Å (Figure 36). Apparently, modification in the backbone restricts the conformational dynamics of the entire oligonucleotide. At the beginning of the trajectory, there is a sampling of extended conformation structures. Then (at 1 ns), the molecule adopts a bend conformation with a loop at the N-terminus (**C5**, Figure 42). At 1.3 ns, the oligonucleotide falls into a more coiled conformation (**C12**). The central turn motif appears again at 5.7 ns and becomes wider (**C15**). As a result, the molecule adopts a bend conformation that lasts for 3.5 ns. Finally, the last 0.8 ns correspond to the sampling of coiled motifs (residues 2 to 9) together with the two central turns in the molecule (**C3**). In summary, the simulation points to the tendency of *cp*PNA**2** to attain a coiled structure; however, this pattern is not very stable and the molecule fluctuates around an ensemble of two conformations (coil and bend).

As in the simulation with *aeg*PNA**1**, the simulation trajectory was clustered using the backbone dihedral angles. To compare the results between two simulations, the same clustering plan was used (Tables 21 and 22). Six clusters were identical between Schemes I and II. The remaining four clusters from Scheme I were replaced by five new clusters in Scheme II. Five of the clusters from Scheme I were also located using the criteria of Scheme III, though the population percentages of all clusters, except cluster **8** were higher using Scheme III. The remaining Scheme I clusters were replaced by one cluster in Scheme III. Only one cluster from Scheme III was identical to a cluster in Scheme II.

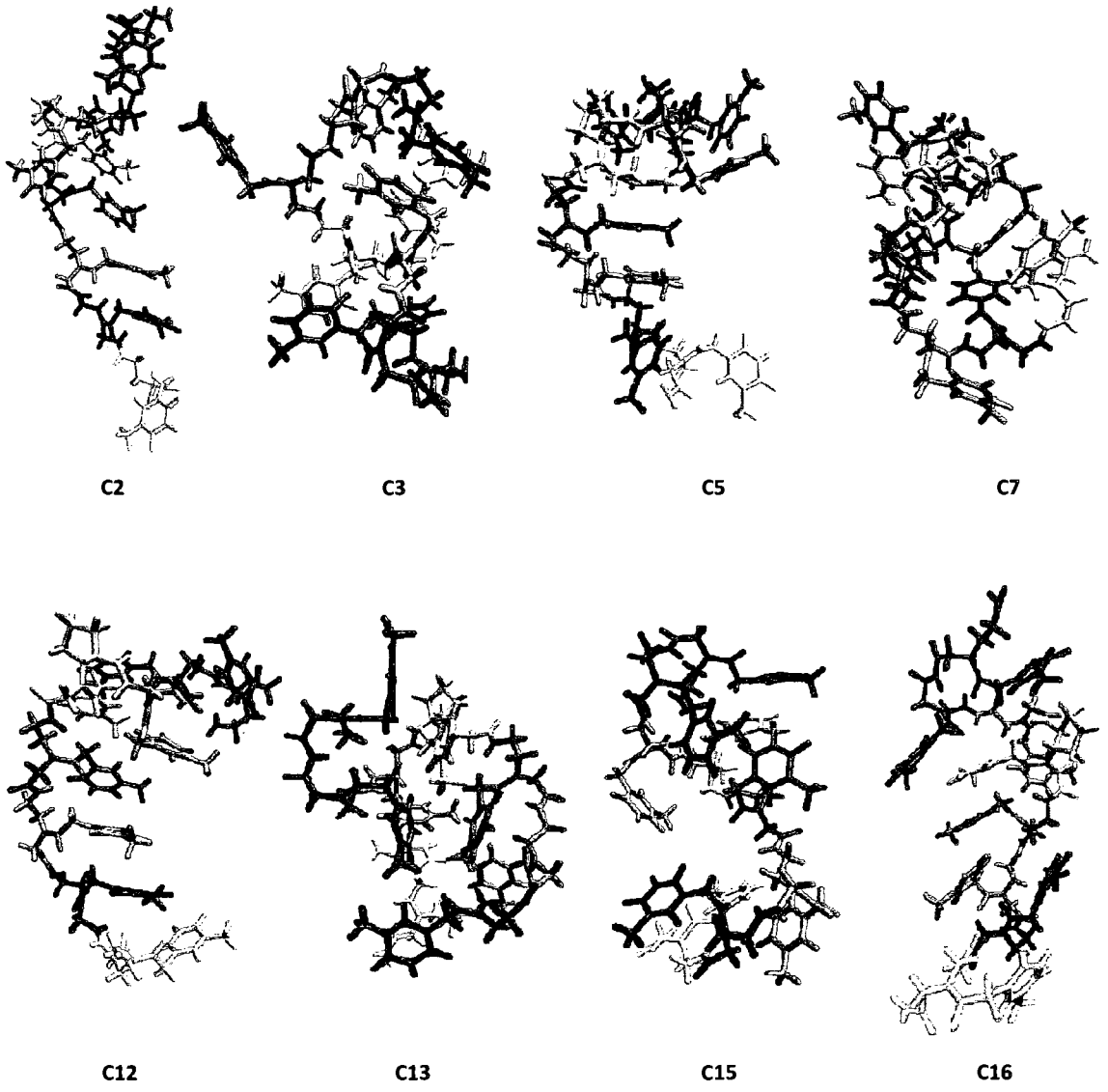

Figure 42. Representative structures of the most populated clusters from the MD simulation of *cp*PNA**2** using the different clustering Schemes. Residues are colored by position: **T1**-red, **T2**- dark grey, **T3**-orange, **T4**-yellow, **T5**-dark yellow, **T6**-grey, **T7**-green, and **T8**-white. The structures are positioned with **T1** on top and **T8** on bottom

Table 21. Average torsion angles for each cluster from the *cp*PNA**2** simulation

Cluster	No. of members	%	T2		T3			T4		T5			T6		T7		
			β	δ	α	γ	ε	β	δ	α	γ	ε	β	δ	α	γ	ε
Scheme I																	
1	359	**3.6**	68	103	170	96	-123	67	115	178	97	60	-164	84	170	101	73
2	1361	**13.6**	-178	97	-179	94	-100	66	101	-176	78	-55	172	88	180	91	81
3	991	**9.9**	-58	107	179	103	-136	67	104	179	85	-105	172	96	175	97	84
4	571	**5.7**	57	97	169	97	-122	69	106	166	131	-94	-81	100	-179	101	75
5	652	**6.5**	-63	93	172	93	-112	68	103	-163	85	67	75	86	178	86	80
6	318	**3.2**	-52	79	168	83	80	148	95	-131	86	69	63	87	174	71	65
7	4475	**44.8**	-63	109	175	101	-138	68	102	-168	80	97	174	95	177	90	78
8	415	**4.2**	-57	80	-179	80	-170	143	92	-166	80	73	60	88	171	82	72
9	389	**3.9**	-180	84	-167	89	-96	66	108	-169	81	65	67	84	179	84	82
10	469	**4.7**	-93	110	-180	95	-119	68	100	-167	79	-62	175	89	175	100	-119
Scheme II																	
11	309	**6.2**	-58	81	172	81	137	147	94	-155	82	66	66	81	-179	80	69
12	459	**9.2**	-111	85	-180	89	-102	78	105	-165	84	65	68	83	179	84	80
13	904	**18.1**	-59	109	174	101	-134	69	102	-168	82	75	168	82	157	101	80
14	79	**1.6**	54	94	-177	79	-138	146	94	-150	78	70	60	70	173	79	72
15	720	**14.4**	-64	107	174	102	-147	68	102	-155	80	99	174	104	174	86	90
Scheme III																	
16	885	**9.0**	-102	102	-178	101	-125	66	101	-172	78	-71	171	91	175	108	84

Table 22. Similar clusters identified from the different clustering Schemes for *cp*PNA**2**

Cluster	No. of mem	%	Scheme I Cluster #[a]	Cluster	No. of mem	%	Scheme I Cluster #[a]
Scheme II				*Scheme III*			
A	678	**13.6**	2	A	1588	**15.9**	8
B	229	**4.6**	10	B	1100	**11.0**	2
C	287	**5.7**	4	C	2094	**20.9**	13
D	180	**3.6**	1	D	2967	**29.7**	7
E	418	**8.4**	3	E	934	**9.3**	4
F	737	**14.7**	7	F	432	**4.3**	10
Total number of clusters			11				7

[a] numbers correspond to identical cluster from Table 21

Using the three criteria, sixteen distinct clusters remained (Figure 42). The most populated clusters obtained from all three Schemes have very similar dihedral angle values. The results show that the most populated conformation, cluster **7**, contains an N-terminal loop and bend motifs and the structure resembles that of cluster **11** from Scheme II. The second most abundant pattern represented by clusters **2**, **3** and **5** from Scheme I, cluster **15** from Scheme II and cluster **16** from Scheme III, groups structures having a coil with a helical motif from residues 4 to 6. The major difference between the clusters is the orientation of the terminal residues. The rest of the clusters have low populations and exhibit less frequent structural features or do not present any conformational motifs.

Conformations obtained from clustering of *aeg*PNA**1** trajectories are predominantly coil-like in comparison to those from *cp*PNA**2**, which exist in a bent conformation most of the simulation time. One of the clusters in *cp*PNA**2** that accounts for only 9% of the total number of structures, has a coil conformation, however upon comparison of the dihedral angles between *aeg*PNA**1** and *cp*PNA**2**, differences are observed for the residues at and around the modification. Thus, these two structures are significantly different from each other.

Compared to *aeg*PNA**1**, the single modification at residue T_4 has a large effect on the flexibility of the structure (i.e. less flexible). The number of clusters from these two trajectories is approximately the same, which seems unexpected at first, since the modification should induce rigidity and thereby, reduce the number of conformations.

*Aeg*PNA**1** has many clusters with approximately the same energy, which is proportional to the population number; while *cp*PNA**2** has one to two very low energy conformers, a couple of low energy conformers and the rest are high energy (except in Scheme II). Thus, *cp*PNA**2** is less flexible as it will spend most of its time in the very low and low energy conformers while *aeg*PNA**1** will be moving among many similar energy conformers.

According to the RMSFs of the backbone dihedrals, residues T_3 and T_4 in *cp*PNA**2** (over which the cyclopentane is introduced) are less flexible than T_3 and T_4 in *aeg*PNA, while the non-terminal residues T_5 and T_7, have approximately the same flexibility. T_1 and T_8 show higher flexibility when compared to the unmodified *aeg*PNA**1** simulation. In summary, the modification resulted in decreased flexibility of the two residues at the modification site but increased RMSF values for T_1 and T_8. The data in Table 23 show that the dihedral flexibility distribution is different for the *cp*PNA**2** molecule compared to *aeg*PNA**1**. These results are in agreement with the clustering data.

Table 23.Average dihedrals (°) and RMSF (bold, °) from cluster analysis of *cp*PNA**2**

Residue	α	β	γ	δ	ε	Total RMSF	Average RMSF
T1	—	—	-51.4	105.1	49.0		
	—	—	**84**	**53**	**72**	**209**	**70**
T2	176.8	-70.3	-58.6	102.3	85.1		
	23	**65**	**76**	**17**	**28**	**209**	**42**
T3	176.3	163.6	97.0	92.1	-132.2		
	21	**32**	**15**	**14**	**37**	**119**	**24**
T4	-107.0	74.6	68.0	102.5	81.3		
	17	**24**	**20**	**12**	**48**	**121**	**24**
T5	-170.7	38.2	84.0	101.5	-140.9		
	31	**59**	**18**	**23**	**135**	**266**	**53**
T6	-179.1	161.6	72.6	91.3	25.1		
	27	**52**	**54**	**20**	**65**	**218**	**44**
T7	177.3	7.4	90.9	90.9	-40.5		
	26	**62**	**17**	**17**	**169**	**291**	**58**
T8	-177.3	-94.5	-30.5	—	—		
	25	**71**	**84**	—	—	**180**	**60**

The consequences on the mean dihedral angle distributions are also shown in Table 23. The most significant change is that the backbone β angles are more broadly distributed than those for the original *aeg*PNA**1**. The γ and δ torsions over the central residues (3 to 7) become restricted to a narrower range: 70-90° and 90-110°, respectively. The range of variations for the terminal γ torsions (residues 1, 2 and 8) is 30-70°. Similar to the *aeg*PNA**1**, the ε torsion still does not have a preferred range of values. However, the ε dihedrals do differ significantly between *aeg*PNA**1** and *cp*PNA**2**. Thus, the cyclopentane modification has had a significant effect on the dihedral profile of the backbone and on its flexibility. These alterations could have significant effects on the ability of *cp*PNA**2** to bind to RNA or DNA.

PNA-T_8 sequence with Lys: *aeg*PNA**3** *and cp*PNA**4**

In the simulation of *aeg*PNA**3**, the RMSD plot with respect to the starting structure shows that the system is relatively stable, with the RMSD in the range of 4-5 Å for most of the simulation trajectory. As in the case of the *aeg*PNA**1** MD trajectory, the first patterns sampled correspond to structures that exhibit no secondary structure or exhibit some residues with dihedral angles corresponding to a helix, since the simulation was started in a helical conformation. A bend motif appears at residue 1 at 400 ps, and disappears and reappears during the rest of the simulation. The central region alternatively exhibits turns and loops extending from residues 2 to 7, and sporadically unfolds. Subsequently, PNA adopts a hairpin-like conformation (~4 ns) that folds later into a knot (~7 ns).

Ten clusters were identical between the Schemes I and II, and the remaining six clusters from Scheme I were replaced by four new clusters in Scheme II. Nine of the clusters from Scheme I were also identified in Scheme III with similar cluster memberships. The remaining five clusters of Scheme I were replaced by only one new cluster in Scheme III. One of the new clusters identified in Scheme II matched that of Scheme III. Thus, twenty one distinct clusters were identified and most of them are representatives of different folded conformations (Figure 43, Tables 24 and 25).

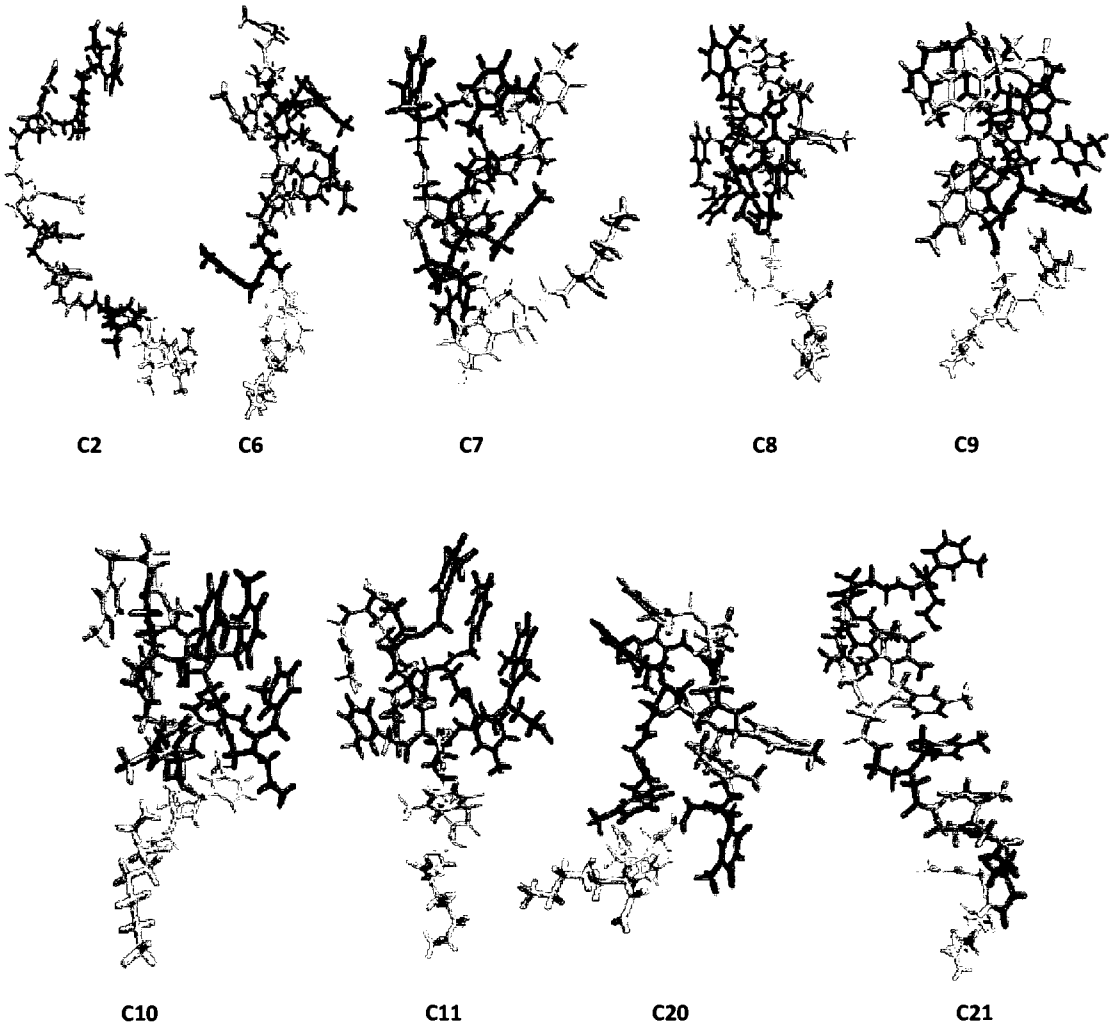

Figure 43. Visualization of the most populated molecular structures corresponding to *aeg*PNA**3** cluster averages. Residues are colored by position: **T1**-red, **T2**- dark grey, **T3**-orange, **T4**-yellow, **T5**-dark yellow, **T6**-grey, **T7**-green, **T8**-white, and **K9**-pink. The structures are positioned with **T1** residue on top and **K9** on bottom

Table 24. Average torsion angles for each cluster from the *aeg*PNA3 simulation

Cluster	No. of Members	%	T2		T3			T4		T5			T6		T7		
			β	δ	α	γ	ε	β	δ	α	γ	ε	β	δ	α	γ	ε
Scheme I																	
1	296	**3.0**	56	116	-174	74	-94	64	105	-178	95	68	62	78	175	74	-42
2	887	**8.9**	161	-103	-160	91	-87	140	88	176	82	44	66	99	-175	87	66
3	152	**1.5**	171	-100	-150	82	-117	163	71	174	81	62	70	-135	-169	90	68
4	485	**4.9**	175	-94	170	101	-44	-141	76	179	85	-133	174	91	178	88	-84
5	449	**4.5**	-152	-94	172	100	-67	-72	66	-175	79	-110	66	93	163	91	-110
6	1250	**12.5**	179	-94	172	81	-58	-61	94	-177	97	-141	179	98	178	88	90
7	1513	**15.1**	-179	-90	-171	85	1	177	83	169	91	-166	179	97	178	98	90
8	1265	**12.7**	-178	-104	175	82	91	168	-91	169	91	75	-176	94	171	85	-115
9	704	**7.0**	-176	-88	175	117	-156	176	-87	-134	99	67	180	98	178	80	-86
10	773	**7.7**	179	-87	166	109	-169	170	-106	-151	92	72	158	104	-177	105	96
11	1161	**11.6**	174	-85	160	109	-147	170	-108	-177	90	88	177	-127	175	96	99
12	207	**2.1**	180	-97	166	77	-73	23	85	-179	98	-171	-179	91	-179	94	42
13	108	**1.1**	177	-97	174	82	84	165	-88	-176	107	-148	175	99	-171	108	139
14	61	**0.6**	179	-96	-172	84	72	163	-69	-177	104	-150	174	104	-173	110	-156
15	349	**3.5**	180	-84	172	116	-164	173	-95	-143	95	55	-76	79	169	88	-102
16	340	**3.4**	-178	-93	170	108	-133	170	-109	152	89	76	168	104	-178	99	98
Scheme II																	
17	131	**2.6**	128	-100	-155	89	-103	63	113	-173	98	65	68	80	-178	82	70
18	284	**5.7**	175	-99	-170	96	-56	174	81	172	85	-112	150	99	-179	95	90
19	204	**4.1**	-179	-82	175	111	-179	170	-108	-148	92	83	117	96	175	104	151
20	335	**6.7**	-176	-88	-173	80	48	178	90	171	93	-159	-179	96	179	99	94
Scheme III																	
21	822	**8.2**	173	-103	-159	90	-85	170	77	172	76	32	71	130	-174	89	66

Table 25. Similar clusters identified from the different clustering Schemes for *aeg*PNA**3**

Cluster	No. of mem	%	Scheme I Cluster #[a]	Cluster	No. of mem	%	Scheme I Cluster #[a]
Scheme II				*Scheme III*			
A	146	**2.9**	1	A	280	**3**	1
B	760	**15.2**	7	B	288	**3**	17
C	601	**12.0**	6	C	605	**6**	4
D	240	**4.8**	4	D	282	**3**	5
E	268	**5.4**	5	E	1513	**15**	6
F	79	**1.6**	12	F	1478	**15**	7
G	577	**11.5**	8	G	1285	**13**	8
H	320	**6.4**	9	H	469	**5**	15
I	474	**9.5**	10	I	727	**7**	9
J	581	**11.6**	11	J	655	**7**	10
				K	1596	**16**	11
Total number of clusters			14				12

[a] numbers correspond to an identical cluster from Table 24

There are no significant differences in the predominant structures obtained using the parameters of Schemes I and II. However, the most populous structure from Scheme III differs from the predominant clusters of Schemes I and II in the values of the $T_3\delta$, $T_4\gamma$, $T_5\delta$ and $T_6\gamma$ dihedrals. The structures in the three most abundant clusters for all three schemes (**6**, **7** and **8**) contain a hairpin loop running from residues 1 to 8. The second group of most populated clusters, again across all three schemes, (**9**, **10** and **11**) is characterized by a knot, expanding from residues 2 to 7, and a C-terminal bend. The remaining clusters from Schemes I, II and III, have low populations and thus, these structural motifs are less important. Cluster **4** from Scheme I (4.85 %) exhibits an N-terminal bend and a turn from residues 5 to 7. Clusters **13** (1.1%) and **15** (3.5%) from Scheme I, and cluster **19** (4.1%) from Scheme II, exhibit a combination of turns and loops for the region covering residues 2 to 7. These structures represent rare events and occur when the PNA is fluctuating between the two most probable conformations. Finally, cluster **1** from Scheme I (2.9%), has no conformational motifs, while cluster **2** from Scheme I (8.8%) and **20** (6.7%) from Scheme II have an overall bend in the

structure. The clustering shows that *aeg*PNA**3** is a very flexible molecule with many shallow minima.

Like *cp*PNA**2**, *cp*PNA**4** has a cyclopentane modification at residue 4. The RMSD of *cp*PNA**4** is higher than that of *cp*PNA**2**, but it follows a similar pattern of fluctuation. A loop appears at the modification site at 400 ps. This loop continues to widen and fold until 4.8 ns at which point the PNA reaches a folded state which exists for the remainder of the simulation.

The most populated clusters from the three schemes have very similar dihedral angle values. Eight clusters from Scheme II and three clusters from Scheme III matched those of Scheme I. The remaining one cluster from Scheme I was replaced by three new clusters in Scheme II and 6 clusters of Scheme I were replaced by thirteen new clusters in Scheme III. Thus, the structures of *cp*PNA**4** were classified into twenty five distinct clusters (Figure 44, Tables 26 and 27). Scheme II had a lower population for cluster **5** but a higher one for cluster **9**. None of the new clusters from Scheme II matched those of Scheme III. The most populated cluster for all three Schemes, **7**, accounts for 42% of the structures, and has a G-type fold over residues 2 to 8. A similar structure was observed for cluster **23** with a population of 13% and cluster **16** from Scheme III. Cluster **5** from Scheme I and cluster **24** from Scheme III, have a wide loop motif from residues 2 to 7. The rest of the clusters have lower memberships. Cluster **2** (Scheme I), cluster **11** (Scheme II) and cluster **13** (Scheme III), show an extended conformation with a small bulge formed at the modification site.

For Schemes I and II, *cp*PNA**4** had fewer clusters than *aeg*PNA**3**. For Scheme III, *cp*PNA**4** had five more clusters than *aeg*PNA**3**, but six of the clusters had a population of 1% or less (none of the *aeg*PNA**3** Scheme III clusters were lower than 3%). In addition, there is a cluster with a very large population in all Schemes for *cp*PNA**4**, but *aeg*PNA**3** has several clusters with similar populations. These results all suggest that the addition of the cyclopentane ring decreased the flexibility of the oligonucleotide. The most populated clusters of *aeg*PNA**3** and *cp*PNA**4** also differ in their conformations and structural motifs, since the modification produces a kink at the site of modification that gives rise to a turn. The clusters exhibit different types of folded structures (hairpin vs. G-type), and the

terminal regions of the unmodified strand form loops, which was not observed for *cp*PNA**4**.

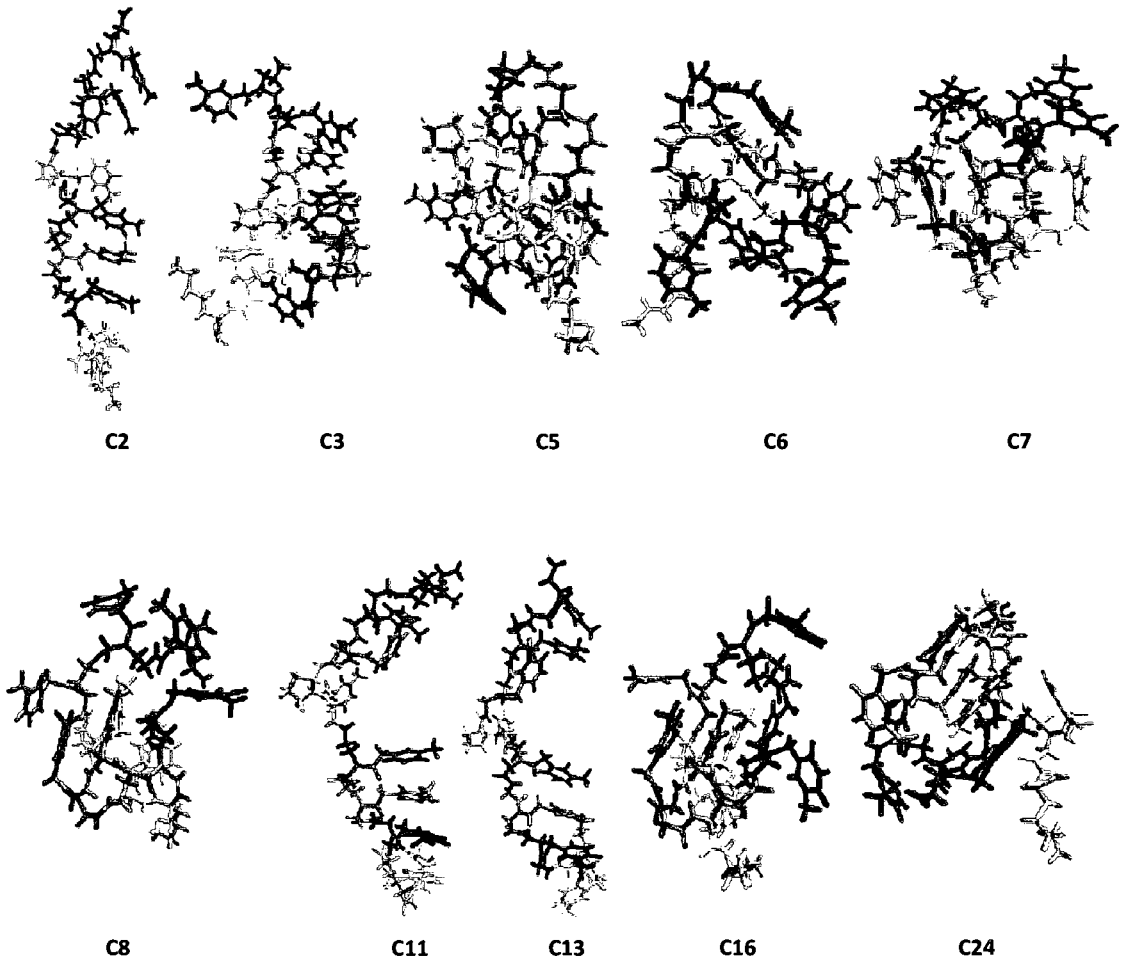

Figure 44. Conformation of the central members of the most populated clusters in the simulations of *cp*PNA**4**. Residues are colored by position: **T1**-red, **T2**- dark grey, **T3**-orange, **T4**-yellow, **T5**-dark yellow, **T6**-grey, **T7**-green, **T8**-white, and **K9**-pink. The structures are positioned with **T1** residue on top and **K9** on bottom

Table 26. Average torsion angles for each cluster from the *cp*PNA**4** simulation

Cluster	No. of members	%	T2		T3			T4		T5			T6		T7		
			β	δ	α	γ	ε	β	δ	α	γ	ε	β	δ	α	γ	ε
Scheme I																	
1	290	**2.9**	59	113	-173	76	-102	124	95	-173	85	65	68	79	-173	78	74
2	1314	**13.1**	67	104	-172	85	-123	67	102	-175	-99	61	65	82	179	86	6
3	895	**9.0**	-72	94	-172	91	-154	66	115	-167	-96	64	66	94	175	91	89
4	402	**4.0**	-51	95	-171	94	-156	65	121	-163	-93	-67	153	111	-179	85	43
5	1201	**12.0**	-66	-89	-180	86	159	67	107	-169	-94	62	77	85	167	105	-98
6	713	**7.1**	63	-99	-166	80	112	66	108	-164	-88	-61	144	85	172	107	-76
7	4200	**42.0**	63	-102	179	82	122	67	107	-169	-91	61	76	91	174	106	-120
8	831	**8.3**	63	-95	172	81	121	66	107	-156	-90	-35	134	82	160	101	117
9	154	**1.5**	-72	-94	-178	85	-141	70	106	-163	-89	-54	146	91	148	85	-99
Scheme II																	
10	265	**5.3**	61	97	-171	89	-145	67	100	-175	-100	58	65	78	174	83	-90
11	414	**8.3**	67	109	-172	83	-111	67	104	-175	-99	63	65	83	-179	87	57
12	36	**0.7**	-80	-88	-175	92	164	66	112	-157	-94	-52	150	95	153	88	-102
Scheme III																	
13	941	**9.4**	61	114	-173	79	-99	85	99	-175	-38	64	67	77	179	84	61
14	672	**6.7**	-8	89	-178	91	-157	66	115	-169	-97	65	63	108	-170	81	61
15	341	**3.4**	-46	93	-171	94	-156	65	121	-165	-94	-66	151	112	-180	81	32
16	133	**1.3**	-79	-114	-172	88	-170	67	101	-171	-93	56	75	75	146	80	-108
17	95	**1.0**	-62	104	-171	91	-156	66	119	-153	-90	-71	156	110	-178	104	91
18	84	**0.8**	-152	-70	164	82	-176	65	109	-163	-97	59	78	72	136	80	-117
19	98	**1.0**	-70	131	-158	86	-156	65	107	-169	-96	63	74	70	144	81	-126
20	371	**3.7**	-59	88	-168	89	-152	65	117	-166	-98	64	62	70	153	109	93
21	78	**0.8**	-103	-91	164	86	179	67	100	180	-96	55	75	78	165	84	-90
22	400	**4.0**	-48	-71	-165	94	145	68	107	-175	-95	60	77	79	172	108	-77
23	1264	**12.6**	61	-106	-169	83	112	67	107	-174	-90	59	76	73	161	106	-68
24	93	**0.9**	-46	-98	-173	82	-110	72	101	-164	-86	-61	155	91	150	85	-98
25	288	**2.9**	53	-99	164	81	130	67	108	-166	-90	-32	112	106	178	101	-151

Table 27. Similar clusters identified from the different clustering Schemes for *cp*PNA**4**

Cluster	No. of mem	%	Scheme I Cluster #[a]	Cluster	No. of mem	%	Scheme I Cluster #[a]
Scheme II				*Scheme III*			
A	145	**2.9**	1	A	3630	**36.4**	7
B	425	**8.5**	3	B	753	**7.5**	8
C	200	**4.0**	4	C	759	**7.6**	6
D	339	**6.8**	5				
E	353	**7.1**	6				
F	2100	**42.0**	7				
G	412	**8.2**	8				
H	311	**6.3**	9				
Total number of clusters			11				16

[a] numbers correspond to identical cluster from Table 26

Table 28. Average dihedrals (°) and RMSF (bold, °) of the simulated *aeg*PNA**3**

Residue	α	β	γ	δ	ε	Total RMSF	Average RMSF
T1	—	—	-63.9	-88.0	-75.1		
	—	—	**69**	**36**	**47**	**152**	**51**
T2	176.2	175.4	-31.9	-97.6	-112.8		
	23	**30**	**87**	**31**	**68**	**239**	**48**
T3	176.3	-26.0	94.2	-125.1	55.3		
	26	**95**	**19**	**51**	**125**	**316**	**63**
T4	-169.1	-166.4	92.7	-2.1	16.4		
	29	**65**	**20**	**94**	**75**	**283**	**57**
T5	-176.7	171.5	91.1	84.8	175.7		
	34	**63**	**15**	**17**	**118**	**247**	**49**
T6	-176.9	160.9	91.0	113.9	33.0		
	28	**52**	**15**	**50**	**87**	**232**	**46**
T7	177.2	-158.0	91.5	123.7	-82.5		
	27	**128**	**16**	**71**	**148**	**390**	**78**
T8	-178.0	170.3	48.7	—	—		
	29	**59**	**79**	—	—	**167**	**56**

The RMSF values show that residues T_3 and T_7 are the most flexible in *aeg*PNA**3** (Tables 28 and 29). During the MD simulation, the T_5 and T_6 residues tend to have the most stacking interactions and have moderate flexibility. The *cp*PNA**4** simulation shows that T_4, which has the modification, is the least flexible, while residues T_1, T_2, T_7 are the

most flexible. In fact, T_4 is very inflexible in the *cp*PNA**4** simulation. Residues T_5 and T_6 undergo minimal changes in their secondary structure during the simulation and, therefore, are able to establish stacking interactions and have low flexibility. Comparison between *aeg*PNA**3** and *cp*PNA**4** shows that the fluctuations of residues 3, 4, 5, 6 and 8 are reduced with the addition of the cyclopentane modification. As with *cp*PNA**2**, the modification leads to decreased flexibility, particularly at the modification site, together with increased mobility of other residues, though not the same ones (T_1 and T_8, *cp*PNA**2**; T_1 and T_2, *cp*PNA**4**).

Table 29. Average dihedrals (°) and RMSF (bold, °) of the simulated *cp*PNA**4**

Residue	α	β	γ	δ	ε	Total RMSF	Average RMSF
T1	—	—	13.9	90.0	-89.4		
	—	—	**91**	**28**	**124**	**243**	**81**
T2	179.4	31.1	92.2	-145.0	-122.3		
	21	**59**	**18**	**76**	**136**	**310**	**62**
T3	-177.8	162.5	83.8	140.4	157.1		
	26	**36**	**13**	**58**	**58**	**191**	**38**
T4	-104.9	68.5	57.9	107.5	165.7		
	19	**13**	**16**	**11**	**29**	**88**	**18**
T5	-167.6	-100.6	87.7	95.9	38.0		
	20	**48**	**32**	**25**	**53**	**178**	**36**
T6	168.8	87.4	90.9	88.8	13.7		
	28	**33**	**12**	**21**	**55**	**149**	**30**
T7	172.6	-74.9	99.6	-26.5	-53.9		
	28	**92**	**19**	**86**	**109**	**334**	**67**
T8	-176.1	-146.2	96.1	—	—		
	22	**66**	**16**	—	—	**104**	**35**

Comparing *aeg*PNA**3** and *cp*PNA**4**, the cyclic constraint caused a decrease in the fluctuations of the β, γ and δ torsions, while the fluctuations of the α and ε dihedrals remained about the same. The flexibility of the α torsions is small and approximately equivalent for all investigated molecules and thus is not mentioned further. The ε torsion has the highest fluctuations for all four simulations, which means that the modification has no impact on the flexibility of this dihedral. The modification produces significant changes in the preferred conformations of the molecule as well as in the overall

flexibility. The most significant changes in the *cp*PNA**4** strand are that the β dihedrals vary significantly along the chain, rather than having values around 170°, which is what was observed in *aeg*PNA**3**. On the contrary, the δ torsion, which has a large range of values in *aeg*PNA**3**, covers a range of 90-140° in *cp*PNA**4**. Thus, introduction of the constraint considerably changed the behavior of these dihedral angles. The γ torsion remains primarily at its original value of 90° for both simulations.

Overall, these results indicate that *aeg*PNA has various conformational ensembles and therefore higher conformational entropy, while *cp*PNA is more conformationally restricted and thus, has lower conformational entropy. The conformational constraint of the strand caused by *cp*PNA was evidenced by the differences in the dihedral fluctuations and the reduced numbers of highly populated clusters for *cp*PNA. In addition, the modification affected the distributions of average dihedrals, restricting them into a specific range of preferred values. The β angles were adjusted to be in the range of those for typical DNA/DNA and RNA/RNA duplexes. Consequently, *cp*PNA should have a stronger duplex forming ability than unmodified PNA.

PNA-MB$_{10}$ sequence: *aeg*PNA**12** *and cp*PNA**13**

In the simulations of *aeg*PNA**12**, the RMSD stays below 4 Å, but there is a great amount of fluctuation at ~5 ns (Figure 37). Analysis of the trajectory structures shows that this movement corresponds to formation of a bend in the molecule. After 6 ns, the RMS deviations start to converge to around 6 Å. The evolution of new conformations during the simulation was examined. At the beginning of the trajectory, the oligonucleotide is still in its starting conformation. At 2.5 ns, many conformations appear which have different structural motifs. At 3 ns, a wide loop, from residues 4 to 8 is observed. This structure lasts for a nanosecond and becomes an ω-loop. At 5ns, the conformation changes to a wide bend in the middle portion of the strand that eventually becomes a coil. At 9.2 ns, the strand folds again into an ω-loop. In summary, the simulation points to the tendency of the PNA molecule to attain different conformations, however the structures seem to be similar in energy and the molecule fluctuates around an ensemble of conformations characterized by the presence of various motifs.

Conformations obtained from clustering were classified into twenty two unique patterns (Tables 30 and 31). Six of the clusters from Scheme I were also located using the parameters of Schemes II and III. Four additional clusters were identical between Schemes I and II. The remaining cluster from Scheme I was substituted by seven new clusters in Scheme II. Four clusters of Scheme I were replaced by four new clusters in Scheme III. Only cluster **17** of Scheme II was found to be identical to a cluster in Scheme III. There are no significant differences in the predominant structures between Schemes I and III. Most of the conformations are coil-like, but there are also clusters with an extended structure, as shown in Figure 45. The backbone structures are very diverse; hence, a stable structure could not be detected.

The most abundant cluster for Schemes I and III, **6**, exhibits a wide bend with a slight helical turn. The second most abundant pattern in Scheme I, but also present in Schemes II and III, **10**, exhibits an ω-loop. The third most populated cluster in Scheme I, **7**, has a C-terminal loop and is a coiled structure. Cluster **5** (Scheme I) contains structures with a wide loop, amounting for 8% of the total number of structures. Finally, cluster **1**, observed for all three Schemes, resembles the initial helical conformation. The clusters located only by Schemes II and III have low population numbers, and, therefore, are not considered to be significant. As in the *aeg*PNA**1** simulations, the predominant structures from different clustering schemes are very similar. As demonstrated by the RMSD plot, the molecule remains in the helical-like extended conformation of the starting structure for a smaller period of time.

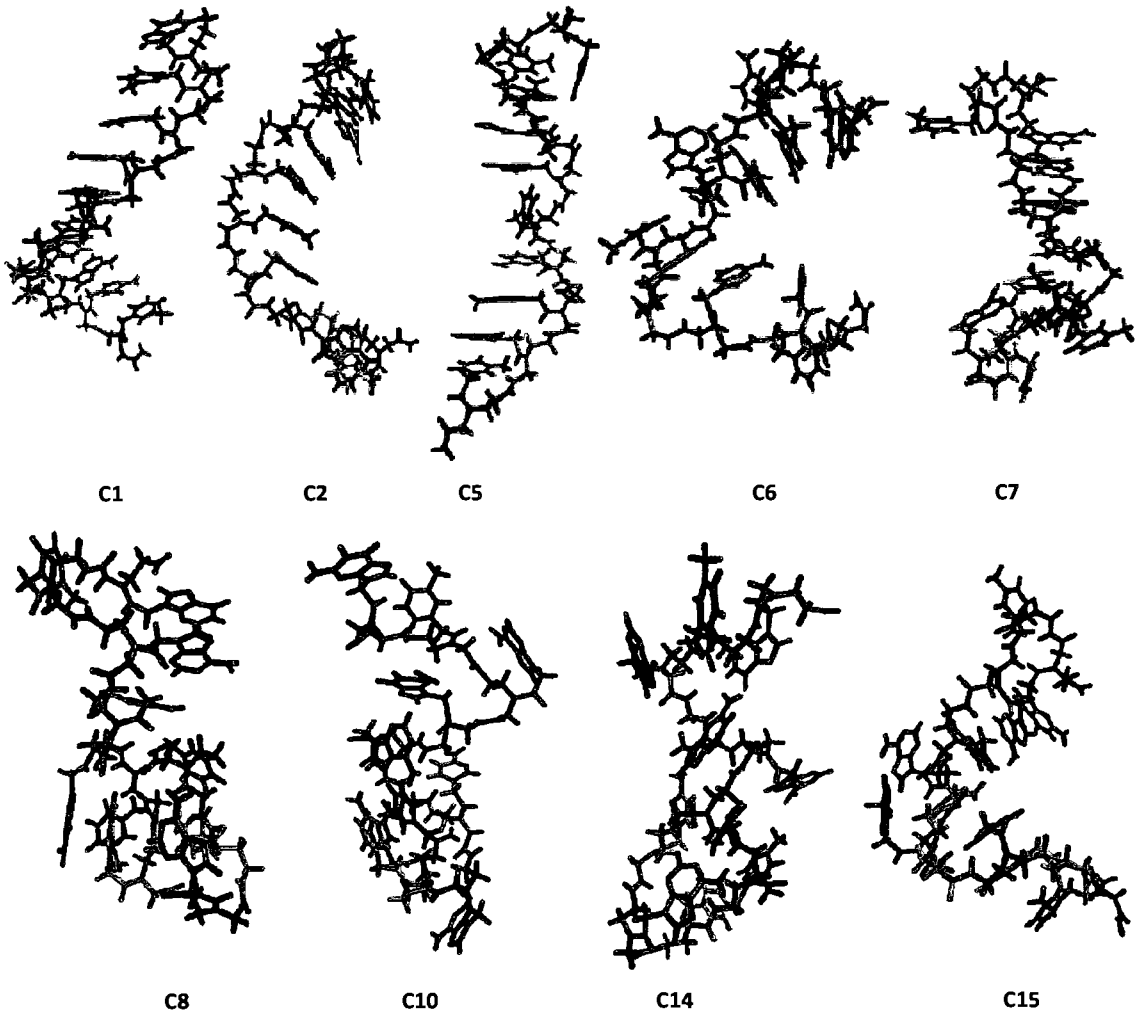

Figure 45. The most populated conformers observed in MD simulations of *aeg*PNA**12**. Residues are colored by name: **APN**-blue, **CPN**-orange, **TPN**-green, and **GPN**-red. The structures are positioned with **GPN1** on top and **TPN10** on bottom (see Table 17 for the MB_{10} sequence)

Table 30. Average torsion angles for each cluster from the *aeg*PNA**12** simulation

Cluster	No. of mem	%	T2		A3			G4		G5			A6		C7			A8		C9		
			β	δ	α	γ	ε	β	δ	α	γ	ε	β	δ	α	γ	ε	β	δ	α	γ	ε
Scheme I																						
1	1025	**10**	69	91	179	82	61	68	80	175	80	64	65	85	175	89	71	61	86	175	82	88
2	1092	**11**	165	-90	176	76	68	65	103	179	84	39	67	95	179	87	-55	173	83	170	86	31
3	84	**1**	175	-108	159	78	-73	144	116	176	66	70	66	70	170	88	-148	175	84	163	86	31
4	471	**5**	169	-72	179	77	-55	83	101	168	93	45	-175	74	175	84	66	65	82	177	82	31
5	840	**8**	177	-76	176	77	64	63	101	165	94	76	-172	90	175	87	68	68	82	178	85	97
6	2113	**21**	17	-109	178	94	59	68	91	176	-164	74	-167	101	172	-100	55	64	88	175	96	60
7	1380	**14**	60	-112	176	93	65	64	92	169	-96	51	176	85	151	-101	-37	180	94	179	88	46
8	699	7	57	-103	174	-105	64	66	105	170	-104	54	172	85	152	-102	70	-172	89	161	106	75
9	377	**4**	39	103	163	-103	30	93	107	180	-101	-66	177	102	160	-89	111	-92	109	164	101	80
10	1736	**17**	130	89	179	-84	118	71	94	171	-98	-50	179	88	153	-90	-82	-178	103	162	103	97
11	183	**2**	64	83	178	84	69	60	89	171	75	-66	148	100	170	98	67	63	98	180	93	60
Scheme II																						
12	31	**1**	57	97	177	75	-63	125	81	169	74	46	71	97	172	94	73	56	84	175	76	59
13	90	**2**	170	-99	180	73	59	63	109	175	99	47	-177	77	169	85	68	67	82	176	83	142
14	401	**8**	-16	-111	175	88	52	70	95	178	95	75	-168	106	168	-101	61	66	66	164	92	51
15	87	**2**	52	93	174	-93	-104	168	107	174	-101	-71	174	94	157	-89	-106	-174	108	163	114	85
16	212	**4**	-180	97	174	-83	110	63	91	170	-101	62	-176	96	159	-90	-83	-178	100	157	98	101
17	41	**1**	-17	104	170	-92	-69	173	112	177	-102	-72	167	107	156	-87	114	-112	105	167	98	77
18	80	**2**	165	-61	175	79	-18	72	90	179	96	-57	164	68	173	85	62	67	76	176	80	-22
Scheme III																						
19	256	**3**	62	110	179	80	40	82	72	166	78	48	73	89	163	94	71	57	83	170	83	22
20	526	**5**	42	101	167	-100	-13	117	107	177	-101	-70	176	100	158	-88	156	-115	110	164	105	82
21	214	**2**	164	-68	177	79	-3	69	95	172	94	-53	162	72	173	84	54	77	77	179	82	31
22	156	**2**	177	-89	165	77	-9	106	119	171	73	70	63	67	157	86	-132	176	83	166	85	26

Table 31. Similar clusters identified from the different clustering Schemes for *aeg*PNA**12**

Cluster	No. of mem	%	Scheme I Cluster #[a]	Cluster	No. of mem	%	Scheme I, II Cluster #[a]
Scheme II				*Scheme III*			
A	485	**10**	1	A	953	**9**	1
B	88	**2**	11	B	1012	**10**	2
C	404	**8**	2	C	1122	**11**	5
D	41	**1**	3	D	2188	**22**	6
E	308	**6**	5	E	1261	**13**	7
F	1345	**27**	7	F	739	**7**	8
G	352	**7**	8	G	981	**10**	10
H	140	**3**	9	H	592	**6**	17
I	574	**11**	10				
J	321	**6**	4				
Total number of clusters			17				12

[a] numbers correspond to identical cluster from Table 30

The calculated RMSF profile provides a more detailed view of residue flexibility (Table 32). The N-terminal residues, G_1 and T_2, have large RMSF values, indicating that the N-terminal region in classical PNA is very flexible. Of the non-terminal residues, 6 and 7, which are turn residues, showed the most flexibility. The central residue G_4 had the lowest RMSF. The β and ε torsion angles had larger fluctuations than the other angles. The γ dihedral also exhibited significant fluctuations, except for the residues involved in base stacking, G_4 and C_9. The α dihedrals have the lowest fluctuations followed by the δ dihedrals for residues 4-10. Conformational analysis indicates that the β torsions lie in the range 130-170° (except for G_4 and C_9). As seen in Table 27, nearly all values seem to be possible for the ε torsion. The δ dihedrals in the central region remain constant at 60-90°. In contrast, three different ranges of γ values were observed: 70-90° (residues 1, 2, 4, 9), 20-40° (residues 3, 6, 8) and -15 to -50° (residues 5, 7, 10).

Table 32. Average dihedrals (°) and RMSF (bold, °) of the simulated *aeg*PNA**12**

Residue	α	β	γ	δ	ε	Total RMSF	Average RMSF
G1	—	—	70.1	169.7	161.9		
	—	—	**63**	**83**	**77**	**223**	**74**
T2	177.8	-41.9	77.2	-79.4	-83.9		
	26	**151**	**47**	**107**	**108**	**439**	**88**
A3	-177.4	142.1	36.2	56.0	58.5		
	24	**53**	**81**	**62**	**49**	**269**	**54**
G4	169.0	69.3	85.7	95.0	7.1		
	26	**24**	**12**	**20**	**66**	**148**	**30**
A5	176.4	149.4	-15.0	89.5	34.1		
	26	**53**	**93**	**20**	**67**	**259**	**52**
T6	178.4	156.5	21.9	90.7	11.4		
	34	**51**	**87**	**19**	**117**	**308**	**62**
C7	166.9	159.1	-28.7	51.5	148.3		
	27	**51**	**90**	**69**	**153**	**390**	**78**
A8	179.6	129.9	27.8	91.1	11.2		
	31	**65**	**87**	**21**	**74**	**278**	**56**
C9	173.8	40.9	92.7	96.4	74.8		
	29	**85**	**15**	**19**	**61**	**209**	**42**
T10	172.5	173.7	-49.5	—	—		
	25	**31**	**79**	—	—	**135**	**45**

The *cp*PNA**13** has the cyclopentane modification at the T_6 residue. The RMSDs with respect to the starting conformation of *cp*PNA**13** are characterized by fluctuations of up to 6 Å (Figure 37). Since the simulation started from a helical conformation, the first conformations sampled correspond to similar, helical structures. Subsequently, PNA adopts an S-type coil (1.8 ns) that unfolds (6.8 ns) and then reaches the final, coil conformation (7.5 ns).

The clustering procedure identified twelve distinct clusters (Tables 33 and 34). Four clusters were identical between Schemes I and II, and only two clusters from Scheme I matched those of Scheme III, though the population percentage of the cluster similar to cluster **4** was much lower. Two clusters from Scheme I were substituted by four new clusters in Scheme II, while four clusters from Scheme I were replaced by two clusters in Scheme III. In Figure 46, the typical structures from each cluster are shown.

Table 33. Average torsion angles for each cluster from the *cp*PNA**13** simulation

Cluster	No. of Members	%	T2		A3			G4		G5			A6		C7			A8		C9		
			β	δ	α	γ	ε	β	δ	α	γ	ε	β	δ	α	γ	ε	β	δ	α	γ	ε
Scheme I																						
1	1182	**11.8**	78	78	177	82	68	61	81	172	82	-20	67	103	-174	83	71	62	91	177	85	78
2	988	**9.9**	58	170	179	84	71	62	80	172	90	-92	68	87	-180	90	-102	178	99	175	84	66
3	161	**1.6**	54	-110	178	91	65	63	88	176	95	-102	69	94	-179	88	-88	-175	81	143	-79	64
4	6794	**67.9**	58	-109	162	89	-120	69	105	-151	105	-113	65	111	-171	85	-41	179	79	170	73	-112
5	540	**5.4**	-57	-5	174	89	-114	60	104	-146	106	-109	64	109	-168	84	-31	174	82	165	66	-121
6	335	**3.4**	57	-102	180	89	50	27	86	-174	99	-99	66	98	-176	88	-78	-179	76	179	78	-74
Scheme II																						
7	9	**0.2**	68	76	172	84	75	55	109	177	80	-97	66	107	-127	-70	45	70	122	162	79	76
8	279	**5.6**	60	99	179	80	71	60	68	163	83	-90	69	80	176	91	-109	175	99	160	84	67
9	325	**6.5**	55	-103	179	89	74	63	92	-178	98	-98	67	96	-175	87	-92	-178	92	180	47	62
10	1051	**21.0**	59	-107	160	88	-121	64	103	-149	106	-113	65	111	-172	84	18	175	77	169	61	-118
Scheme III																						
11	1752	**17.5**	72	86	179	82	69	61	77	169	82	-43	68	96	-177	86	82	99	93	172	84	73
12	5602	**56.0**	47	-98	163	89	-118	62	103	-149	106	-113	65	111	-171	84	-25	179	77	166	66	-116

Table 34. Similar clusters identified from the different clustering Schemes for *cp*PNA**13**

Cluster	No. of mem	%	Scheme I Cluster #[a]	Cluster	No. of mem	%	Scheme I Cluster #[a]
Scheme II				*Scheme III*			
A	581	**11.6**	1	A	711	**7.1**	6
B	2348	**47.0**	4	B	1935	**19.4**	4
C	269	**5.4**	5				
D	138	**2.8**	6				
Total number of clusters			8				4

[a] numbers correspond to identical cluster from Table 33

The most populous and presumably lowest energy clusters are clearly enriched in conformations containing turns and loops. Cluster **12** from Scheme III has the highest population for that Scheme and differs from cluster **4**, which has the highest population for Schemes I and II, only in the value of the $C_7\delta$ dihedral. The most abundant pattern from Scheme I and II, cluster **4**, contains a series of loops involving residues 1 to 4, 5 to 6 and 7 to 9. Clusters **2** (Scheme I) and **12** (Scheme III) exhibit an S-type coil conformation. PNA seems to prefer this pattern and it can be combined with a C-terminal loop or turn motifs. Cluster **10** from Scheme II has a bent backbone with N- and C-terminal loops. Cluster **1**, which is observed for Scheme II as well, contains helical motifs which resemble the initial conformation. A comparison of the clustered conformations of *aeg*PNA**12** and those of *cp*PNA**13** shows major differences. The unmodified PNA strand does not have a preferred structure and numerous conformations with various structural motifs were observed. On the contrary, the most probable conformations of the modified *cp*PNA**13** have a single coil pattern. In addition, the number of clusters is greatly reduced upon addition of the cyclopentane ring and one cluster with a very large population is observed for *cp*PNA**13**, unlike the numerous clusters with similar populations observed for *aeg*PNA**12**. Therefore, the cyclopentane-modified strand has reduced flexibility and should bind stronger to DNA due to a reduction of the entropy penalty.

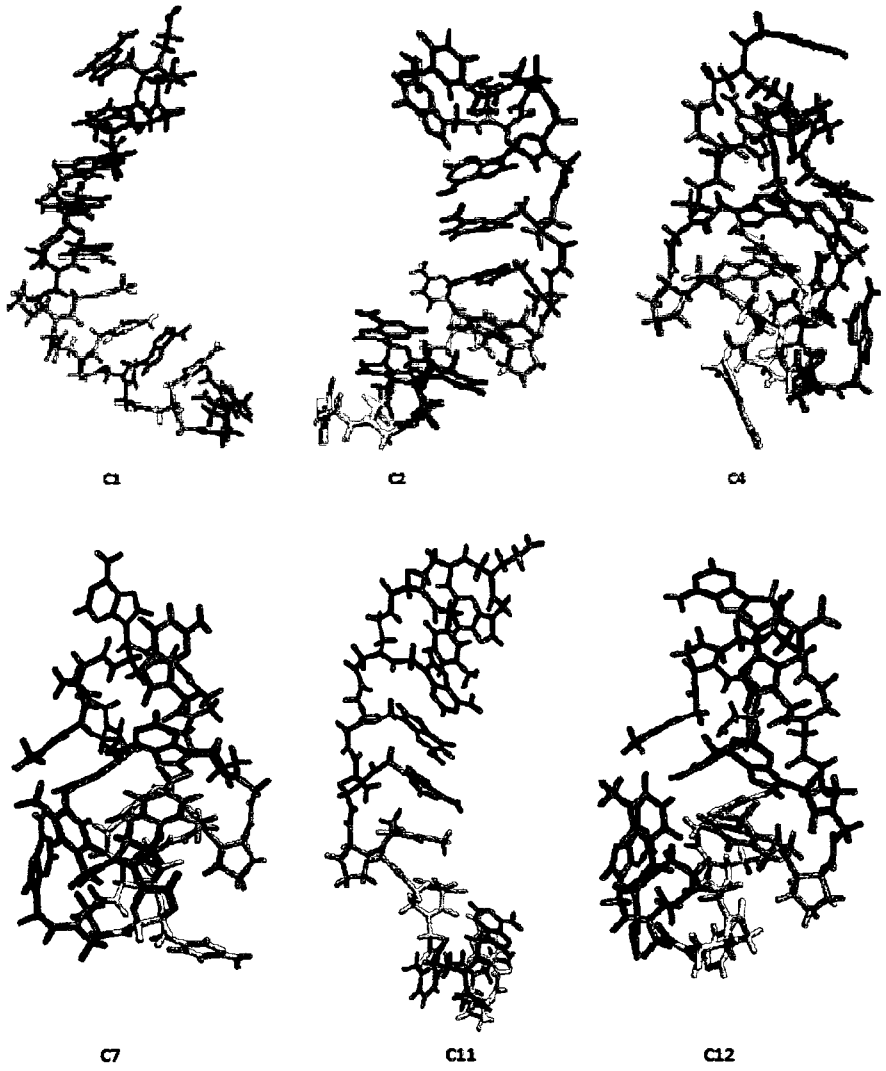

Figure 46. Central members of the most populated clusters of structures from *cp*PNA**13** simulations. Residues are colored by name: **APN**-blue, **CPN**-orange, **TPN**-green, **TPC**-light blue and **GPN**-red. The structures are positioned with **GPN1** on top and **TPN10** on bottom (see Table 17 for the MB_{10} sequence)

The analysis of the *cp*PNA**13** RMSFs show that the C_9 residue has the largest fluctuations, which can be attributed to the mobility of the turn over residues C_7 to C_9. In addition to G_4, which had the lowest fluctuations in the unmodified PNA, the T_6 and C_7 residues, where the modification is located, have low fluctuations in *cp*PNA**13**. It is

immediately apparent that the RMSF values for *aeg*PNA**12** are much larger in magnitude than for *cp*PNA**13**. This reduction in flexibility is assumed to be largely a consequence of the modification. Similar to *aeg*PNA**12**, in *cp*PNA**13** the largest dynamical motion was observed for the torsion angle ε. However, the RMSF values for ε are much lower for the strand with the modification, particularly for residues 1, 2, 6 and 7. Even though the modification is at residue 6, the flexibility of the whole strand is reduced. The lower dihedral flexibility agrees with the clustering results. Modification of the backbone of the decamer seems to have a greater effect on stability than was observed for the polythymine monomers.

Table 35. Average dihedrals (°) and RMSF (bold, °) of the simulated *cp*PNA**13**

Residue	α	β	γ	δ	ε	Total RMSF	Average RMSF
G1	—	—	24.5	83.6	61.3		
	—	—	**93**	**16**	**26**	**135**	**45**
T2	172.3	54.1	85.0	-131.4	3.2		
	21	**31**	**12**	**73**	**59**	**196**	**39**
A3	-167.5	156.4	87.8	87.1	-72.3		
	25	**44**	**12**	**19**	**86**	**186**	**37**
G4	178.6	65.1	80.4	99.0	-48.1		
	25	**22**	**14**	**20**	**60**	**141**	**28**
A5	-160.3	134.8	100.4	-123.0	-98.9		
	26	**39**	**16**	**71**	**41**	**193**	**39**
T6	-111.8	65.5	61.8	106.8	13.8		
	19	**7**	**18**	**14**	**35**	**93**	**19**
C7	-172.1	165.8	85.5	95.3	-35.1		
	22	**27**	**13**	**17**	**63**	**142**	**28**
A8	-175.1	165.1	90.9	82.2	-56.8		
	31	**40**	**21**	**17**	**71**	**180**	**36**
C9	-171.1	50.9	72.1	61.9	-68.7		
	27	**129**	**30**	**45**	**84**	**315**	**63**
T10	175.1	32.2	93.2	—	—		
	22	**66**	**17**	—	—	**105**	**35**

To further analyze the character of the fluctuations caused by the presence of the cyclopentane ring, the average dihedral angles were examined (Table 35). Although the

backbone β angles still have values in the 140-160° region, many residues now have values of ~60° (2, 4, 6 and 10). The γ and δ torsions become restricted to a narrower range and with the exception of certain residues (1, 6 and 2, 5, respectively), adopt values around 90°. The range of variations for the ε torsions (except T_2 and T_6) is from -40 to -70°. The largest differences in the average angle ranges are observed for the γ and ε dihedrals. The modification also had a small influence on the β and δ torsions in residues 2, 6, 10 and 1, 2, 5, respectively. Overall, the modification has reduced the flexibility and altered the conformational preference of the strand.

PNA-MB_{10} sequence with Lys: *aeg*PNA**14** and *cp*PNA**15**

For the *aeg*PNA**14** simulation, the RMSDs over time compared to the starting X-ray structure are under 5 Å (Figure 36). At the beginning of the simulation, the *aeg*PNA**14** is still in an extended conformation. Subsequently, the molecule forms a stable, closed loop structure at 1.7 ns and undergoes a very limited number of structural fluctuations for the next 4 ns. Then, the molecule stays in an alternate conformation for the rest of the MD trajectory (6-10 ns), and no new conformations are observed.

*Aeg*PNA**14** was classified into seventeen clusters (Figure 47, Tables 36 and 37). The predominant structures obtained using all three criteria have similar dihedral angles. Eight clusters from Scheme II were similar to the clusters obtained from Scheme I. The other three clusters from Scheme I were replaced by four new clusters in Scheme II. Four of the clusters from Scheme I were also identified in Scheme III, though with higher population percentages. The remaining seven clusters were substituted with three clusters, one of which matched a cluster in Scheme II. Only three clusters matched between all three Schemes.

Cluster **6** from Schemes I and II, cluster **17** from Scheme III and cluster **8** from all three Schemes represent the most abundant pattern and suggest that most of the structures are grouped in only three clusters. These patterns are characterized by a G-type loop from residues 1 to 7. Cluster **4**, which exists in all three Schemes, has a closed loop structure (residues 1 to 10). Cluster **1** has helical dihedral angles from residues 3 to 7. In summary,

the simulation shows that the PNA folds in the first 2 ns, to adopt a fold that is fairly well maintained for the rest of the MD trajectory.

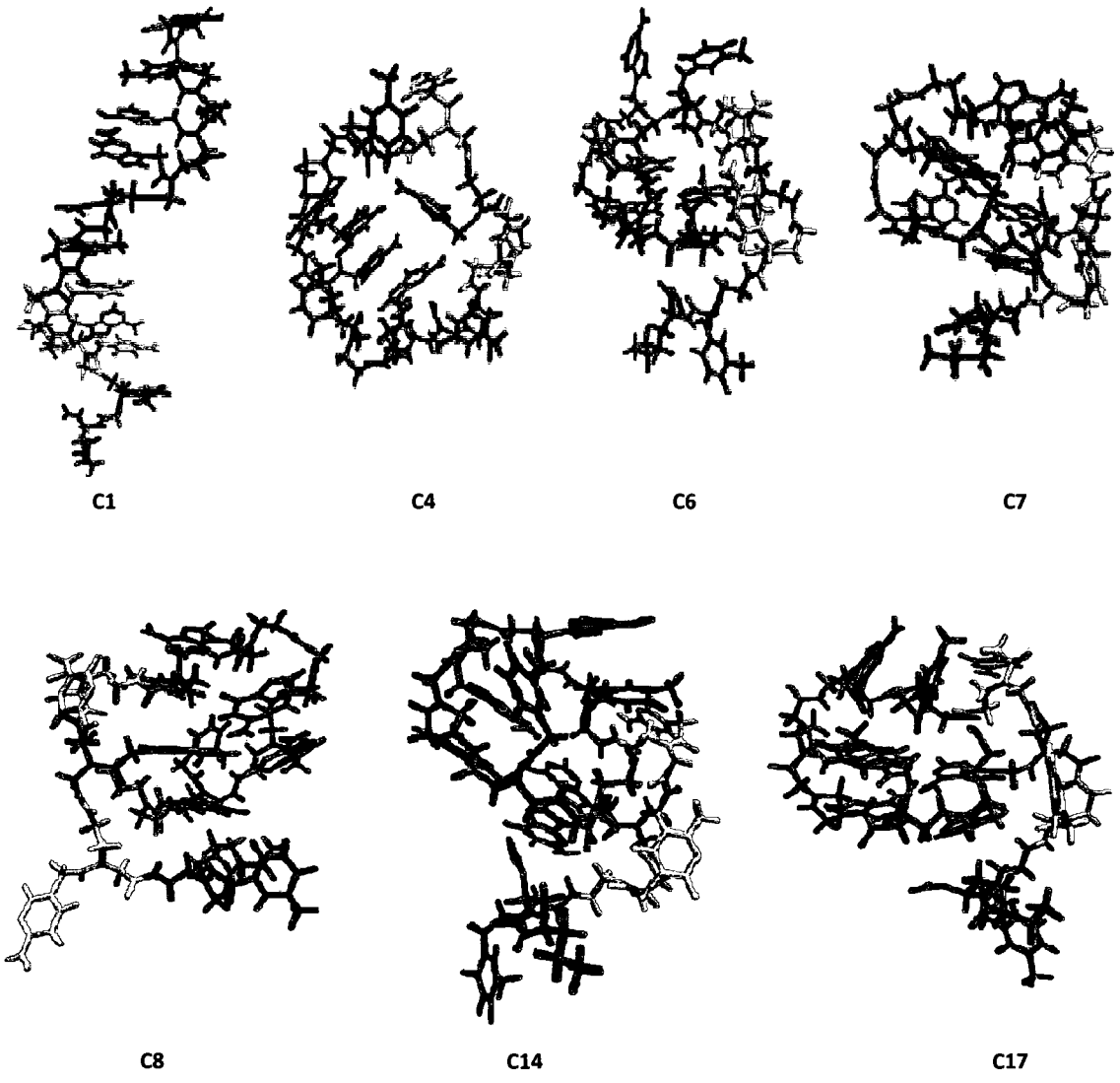

Figure 47. Structures of central cluster members of the simulation of *aeg*PNA**14**. Residues are colored by name: **APN**-blue, **CPN**-orange, **TPN**-green, **GPN**-red and **K**-purple. The structures are positioned with **GPN1** on top and **K11** on bottom (see Table 17 for the sequence)

Table 36. Average torsion angles for each cluster from the *aeg*PNA**14** simulation

Cluster	No. of Members	%	T2		A3			G4		G5			A6		C7			A8		C9		
			β	δ	α	γ	ε	β	δ	α	γ	ε	β	δ	α	γ	ε	β	δ	α	γ	ε
Scheme I																						
1	1327	**13.3**	68	81	-179	82	62	68	79	176	83	65	65	81	178	84	68	62	83	175	79	72
2	287	**2.9**	61	89	179	81	64	67	76	179	82	63	68	71	176	73	-85	177	91	177	107	77
3	203	**2.0**	65	96	-180	86	63	59	81	174	76	-64	117	114	178	84	119	69	84	170	100	70
4	1652	**16.5**	60	121	-167	89	62	64	99	171	87	86	-49	85	164	-92	175	162	97	173	95	-77
5	165	**1.7**	57	107	-146	-95	63	79	82	157	103	71	66	106	150	-105	99	179	81	171	89	91
6	2031	**20.3**	60	93	-175	82	38	69	103	180	85	82	-17	82	164	-91	93	-175	95	178	94	114
7	1004	**10.0**	46	87	173	88	-72	157	87	165	95	75	62	84	158	-101	55	-173	98	180	102	127
8	2399	**24.0**	59	94	172	94	78	175	88	165	101	71	66	92	167	-99	45	-162	116	176	96	-156
9	220	**2.2**	63	104	-177	89	66	65	75	174	82	69	61	102	177	86	-163	-58	85	171	103	80
10	255	**2.6**	58	113	172	84	56	65	96	172	88	70	66	101	176	85	106	69	94	-75	94	-135
11	457	**4.6**	55	109	173	87	67	59	91	180	85	63	64	87	170	84	136	68	89	172	105	95
Scheme II																						
12	231	**4.6**	61	95	179	84	64	66	75	177	83	65	65	83	177	80	-110	-137	88	177	107	78
13	245	**4.9**	57	104	154	84	46	71	100	179	79	77	62	103	167	-102	92	161	93	176	91	-106
14	360	**7.2**	60	86	-175	83	65	73	95	180	83	78	57	83	155	-97	52	-165	94	178	106	110
15	109	**2.2**	61	93	174	96	76	-174	84	170	105	70	65	83	170	-93	-92	-77	128	165	97	-137
Scheme III																						
16	273	**2.7**	55	105	-178	-13	61	67	70	152	106	72	64	106	159	-104	117	179	83	168	90	99
17	2005	**20.1**	53	86	-179	85	-4	110	96	174	89	78	30	81	155	-96	52	-169	98	179	105	120

Table 37. Similar clusters identified from the different clustering Schemes for *eg*PNA**14**

Cluster	No. of mem	%	Scheme I Cluster #[a]	Cluster	No. of mem	%	Scheme I Cluster #[a]
Scheme II				*Scheme III*			
A	891	**17.8**	1	A	1669	**16.7**	1
B	101	**2.0**	3	B	1059	**10.6**	11
C	144	**2.9**	10	C	544	**5.4**	13
D	101	**2.0**	5	D	2290	**22.9**	4
E	730	**14.6**	4	E	2160	**21.6**	8
F	475	**9.5**	7				
G	941	**18.8**	8				
H	670	**13.4**	6				
Total number of clusters			12				7

[a] numbers correspond to identical cluster from Table 36

The *cp*PNA**15** strand has a cyclopentane modification at residue T_6. The RMSD over time is under 5 Å. The conformation of the strand evolves from the starting helical structure to a hairpin loop (1.6 ns). This structure exists interchangeably with a closed loop form for the remainder of the simulation (2-10 ns). The simulation of *cp*PNA**15** yielded thirteen clusters (Tables 38 and 39). Scheme I only provided three clusters. Two of them were also located with Scheme II and III. The remaining cluster from Scheme I was replaced by five new clusters in Scheme II and nine new clusters in Scheme III. Three of the new clusters from Scheme III matched those of Scheme II. Cluster **2** from Scheme I and cluster **3** from all three Schemes have a closed loop conformation (residues 1to 10) (Figure 48). Cluster **5** from Scheme II has a similar conformation. Clusters **6** and **8** from Scheme II along with cluster **10** from Scheme III contain structures with a hairpin loop involving residues 1 to 8. Cluster **1** (Scheme I), **4** and **7** (Scheme II) feature helical turns from residues 3 to 8.

For Schemes I and II, the number of clusters was greatly reduced indicating the decreased flexibility of the *cp*PNA**15** strand. In addition, clusters with an increased population were observed for all three Schemes of *cp*PNA**15**. Taken together, the simulations of *aeg*PNA**14** and *cp*PNA**15** show that the conformation of the decamer is significantly changed by modification. The *aeg*PNA**14** simulations sample a set of conformers, none of which are very dominant and the conformers exhibit various secondary structural elements. On the other hand, the most common backbone

conformations identified for *cp*PNA**15** are restricted to hairpin-like and closed loop conformations.

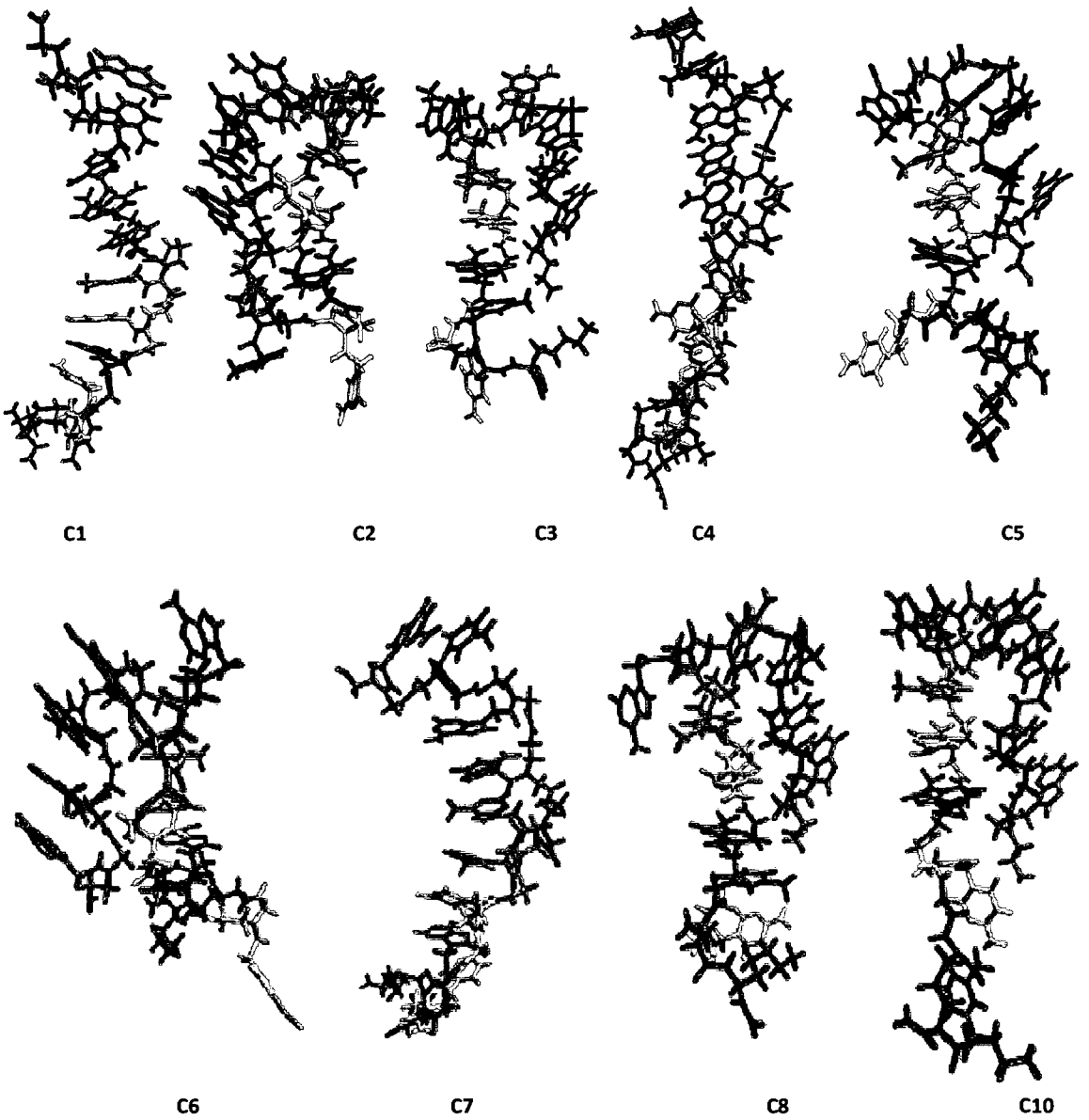

Figure 48. Structures of central cluster members of the simulation of *cp*PNA**15**. Residues are colored by name: **APN**-blue, **CPN**-orange, **TPN**-green, **TPC**- light blue, **GPN**-red and **K**-purple. The structures are positioned with **GPN1** on top and **K11** on bottom (see Table 17 for the sequence)

Table 38. Average torsion angles for each cluster from the *cp*PNA**15** simulation

Cluster	No. of Members	%	T2		A3			G4		G5			A6		C7			A8		C9		
			β	δ	α	γ	ε	β	δ	α	γ	ε	β	δ	α	γ	ε	β	δ	α	γ	ε
Scheme I																						
1	2133	**21.3**	91	96	-173	80	-86	168	77	173	90	28	67	100	164	80	72	61	82	172	81	0
2	3627	**36.3**	61	114	-157	95	-20	88	90	-173	92	-113	110	100	169	76	73	58	102	172	99	-112
3	4240	**42.4**	60	107	-168	105	-58	66	92	153	-112	-138	141	96	172	74	71	59	93	173	100	53
Scheme II																						
4	507	**10.1**	75	96	-169	80	-82	171	81	173	91	32	66	101	168	81	74	62	82	172	81	-83
5	844	**16.9**	61	116	-160	96	-64	106	90	-168	92	-113	102	102	170	78	74	57	98	178	98	-89
6	656	**13.1**	61	108	-171	103	-58	64	92	151	-102	-137	143	96	167	74	71	58	96	160	105	-42
7	547	**10.9**	105	95	-175	80	-92	167	74	172	89	26	67	100	161	78	70	59	82	172	81	68
8	652	**13.0**	62	113	-156	97	-19	78	90	-177	95	-115	120	99	169	75	72	57	105	167	98	107
Scheme III																						
9	282	**2.8**	62	119	-164	84	-68	164	85	-154	92	-115	71	112	173	89	70	61	69	162	83	-127
10	876	**8.8**	58	112	-159	102	-65	82	87	-166	91	-112	119	97	174	72	76	60	111	171	80	-66
11	434	**4.3**	62	115	-156	91	59	62	89	-180	93	-111	114	98	165	74	75	56	96	167	107	-38
12	267	**2.7**	61	113	-157	99	-61	88	91	-169	92	-118	125	100	170	75	70	56	132	168	122	-146
13	176	**1.8**	62	113	-155	93	53	62	89	-170	90	-115	127	100	169	75	71	57	127	162	114	-137

Table 39. Similar clusters identified from the different clustering Schemes for *cp*PNA**15**

Cluster	No. of mem	%	Scheme I Cluster #[a]	Cluster	No. of mem	%	Scheme I, II Cluster #[a]
Scheme II				*Scheme III*			
A	1362	**27.2**	3	A	1258	**12.6**	7
B	426	**8.5**	1	B	4026	**40.3**	3
				C	423	**4.2**	5
				D	283	**2.8**	1
				E	847	**8.5**	4
				F	1119	**11.2**	8
Total number of clusters			7				12

[a] numbers correspond to identical cluster from Table 38

Table 40. Average dihedrals (°) and RMSF (bold, °) of the simulated *aeg*PNA**14**

Residue	α	β	γ	δ	ε	Total RMSF	Average RMSF
G1	—	—	70.2	103.4	-38.2		
	—	—	**60**	**21**	**57**	**138**	**46**
T2	-171.4	59.4	83.8	97.5	63.3		
	27	**20**	**12**	**22**	**42**	**123**	**25**
A3	-179.4	145.5	84.4	83.9	48.0		
	29	**44**	**27**	**21**	**54**	**175**	**35**
G4	171.8	101.6	43.7	91.3	5.7		
	27	**55**	**78**	**20**	**66**	**246**	**49**
A5	175.5	28.5	89.9	86.3	72.0		
	31	**66**	**15**	**16**	**28**	**156**	**31**
T6	176.8	30.7	91.2	87.0	45.8		
	23	**58**	**15**	**18**	**86**	**200**	**40**
C7	169.3	165.8	-65.1	91.9	145.4		
	27	**46**	**120**	**23**	**131**	**347**	**69**
A8	171.5	93.7	159.4	98.0	61.1		
	34	**173**	**86**	**21**	**56**	**370**	**74**
C9	178.1	63.6	94.7	94.9	42.4		
	23	**95**	**18**	**21**	**113**	**270**	**54**
T10	-179.4	154.6	100.8	—	—		
	24	**53**	**41**	—	—	**118**	**39**

To further address the effect of the single cyclopentane modification, the RMSF profiles of *aeg*PNA**14**, and *cp*PNA**15** were compared. According to Tables 40 and 41, the dynamics of many regions were affected by the modification. Most notably, residues G_4

to A_8, except for A_5 are more rigid in *cp*PNA**15**. The reduced flexibility of the modified T_6 residue is not surprising but the increased flexibility of the immediate neighboring residue A_5 is unexpected. The change in overall flexibility is not as pronounced as was observed for the strands without lysine. Most likely, the already reduced flexibility of the lysine containing *aeg*PNA accounts for this discrepancy.

The two simulations were also compared by analyzing the average dihedral angles. Initially, it seems that the presence of the ring did not affect the average values since all the dihedrals, except for ε, stayed close to the values observed for the *aeg*PNA**14**. In particular, the δ torsion remains fixed in its preferred value of ~100°. However, upon detailed examination, a significant contrast was found when comparing the β, γ and ε dihedrals of residues 5-9. The values of the δ dihedrals are more consistent in *cp*PNA**15** and new values are observed for the β and ε angles.

Table 41. Average dihedrals (°) and RMSF (bold, °) of the simulated *cp*PNA**15**

Residue	α	β	γ	δ	ε	Total RMSF	Average RMSF
G1	—	—	78.6	87.7	58.6		
	—	—	**55**	**15**	**32**	**102**	**34**
T2	177.3	66.6	78.3	107.2	61.9		
	17	**26**	**10**	**17**	**62**	**132**	**26**
A3	-164.8	152.1	96.0	93.9	50.2		
	21	**34**	**15**	**29**	**48**	**147**	**29**
G4	-169.5	95.3	90.6	87.9	-45.7		
	29	**45**	**12**	**17**	**38**	**141**	**28**
A5	169.2	160.7	157.7	87.9	-93.5		
	29	**43**	**82**	**16**	**69**	**239**	**48**
T6	-124.4	114.1	89.0	98.5	48.7		
	23	**32**	**14**	**9**	**31**	**109**	**22**
C7	-169.1	66.3	76.0	80.9	-72.0		
	27	**24**	**10**	**15**	**16**	**92**	**18**
A8	172.0	58.6	79.8	93.7	-66.3		
	18	**11**	**11**	**20**	**56**	**116**	**23**
C9	-175.7	-95.8	95.6	-152.3	9.4		
	25	**111**	**17**	**86**	**93**	**332**	**66**
T10	-179.7	-179.2	39.5	—	—		
	32	**31**	**86**	—	—	**149**	**50**

The overall results show that as far as the structural properties are concerned, *aeg*PNA and *cp*PNA are very different. Reduction in the number of conformations indicates a clear modification in the dynamical properties of the PNA strand. Evidence of the greater stiffness with respect to the unmodified form is revealed by analysis of the backbone dihedral RMSF, and is also confirmed by the clustering data. Such findings suggest that the cyclopentane ring plays an important role in determining the flexibility of PNA. The present simulation studies indicate that the cyclopentane modification caused changes in the average values of the backbone torsion angles (β, γ, ε) around the residue with the modification. We suggest that a single cyclopentane ring in PNA functions as an entropic regulator, which reduces the conformational entropy of PNA and increases the binding to DNA. The modification also affects the torsional potential energy surface which could also have consequences on the DNA binding energy.

THE EFFECT OF THE TERMINAL LYSINE

PNA-T_8 sequence

The RMSF results were used to estimate the effect of the terminal lysine on the strand flexibility; therefore, comparisons of the flexibility profiles of *aeg*PNA**1** vs. *aeg*PNA**3** and *cp*PNA**2** vs. *cp*PNA**4**/**4a** were performed. *Aeg*PNA**3** and *cp*PNA**4** contain a Lys residue attached at the C-terminus. The backbone dihedral fluctuations of the residues in *aeg*PNA**3** were significantly larger than those in *aeg*PNA**1**. Generally, the N-terminal exhibited larger fluctuations during the *aeg*PNA**3** simulation. Two other regions, residues 3 and 4, and the C-terminal residues 7 and 8 had considerably larger fluctuations in *aeg*PNA**3** than in *aeg*PNA**1**. At the same time, central residue 6 showed moderate but decreased flexibility in *aeg*PNA**3**. In general, the distributions of the average values tend to be somewhat narrower for *aeg*PNA**3** than for *aeg*PNA**1**. This difference is particularly noticeable for the γ torsion (90-95°). Some major changes are observed for the N-terminal dihedrals γ1 and γ2 which change from 9 to -64° and from -32 to -96°, respectively. The β torsion primarily remains in its original range of 160 - 178° with the

exception of residue 3 which has a value of ~35° in both simulations. The δ dihedrals of all residues but T_7 change significantly when the Lys is attached to the C-terminus.

A comparative analysis of *cp*PNA**2** and *cp*PNA**4** was also performed. The N-terminal residues, 1 to 3, in *cp*PNA**4** have greater RMSF values than in *cp*PNA**2**, indicating that the N-terminal region is more flexible. Just as in *aeg*PNA**3**, this change in residue flexibility is assumed to be largely a consequence of the lysine, even though the lysine is attached at the C-terminus. Residue 7 also had higher RMSF values. A similar pattern of enhanced flexibility was also observed for *aeg*PNA**1** and *aeg*PNA**3**. Residues 4 to 6 and 8 were less flexible than the corresponding residues in *aeg*PNA**1**. While the flexibility of some residues was slightly altered; overall, the flexibility of the strands remained the same. Further examination revealed marked differences in the γ and δ dihedral fluctuations, with γ being less flexible and δ more flexible in *cp*PNA**4**. These results provide additional evidence that the lysine has an impact on flexibility. In particular, the γ6 and δ3 dihedrals are commonly affected by Lys.

The Lys attachment produces significant changes in the average values of the β and ε torsion angles. Even though a particular range could not be identified for either of the torsions, large differences exist in the respective angle values for the two strands. However, the four angles β3, β4, ε6 and ε7 remain essentially unchanged despite the changes in the other dihedrals. The most significant changes occur at the termini where γ1, γ2 and γ8 adopt values of -51°, -59° and -30°, respectively, rather than values of 13°, 92° and 84°, which are the corresponding angles in *cp*PNA**4**. The two strands have similar average values for the γ torsion in the central region of the strand.

The results show that the terminal lysine, although only slightly altering the structure, significantly affects the thermodynamic stability of the molecule. The number of clusters is increased when comparing *aeg*PNA**1** to *aeg*PNA**3**, but approximately the same for *cp*PNA**2** and *cp*PNA**4**. However, the RMSFs are significantly increased for both strands containing the lysine. This increase in total conformers and RMSFs shows the higher conformational variability of the strands with the lysine. In spite of the presence of the cyclopentane ring, the *cp*PNA**4** strand has greater flexibility than the *cp*PNA**2** strand, except at the modified T_4 residue. This result is surprising, because modified PNA strands usually have smaller fluctuations, and could be an indication that lysine is

involved in determining the flexibility and dynamical features of PNA. Also, we note comparatively higher fluctuations at the N-terminal region and differences in the γ and δ average values. Overall, the results indicate that the introduction of the *L*-Lys at the C-terminus of PNA dramatically increases its inherent flexibility and impacts the entropy of binding.

The effect of the stereochemistry of the lysine

D-lysine is thought to add stereochemistry to the PNA strand, which causes an increase in the strength of binding to DNA and also creates a preference for right-handed helices. Therefore, incorporation of a *D*-lysine vs. *L*-lysine should alter the interaction with DNA and its helical preference. In order to evaluate the impact of *D*-Lys on the conformational behavior, an MD simulation of 10 ns was run. RMSDs from the starting structure were calculated for *cp*PNA**4a**.The trajectory stabilized after 6 ns with an average RMSD of 4.0 Å (Figure 2). During the simulations, the secondary structure elements are stable during the 10 ns period, albeit the number of helical elements decreases and the formation of bend and loop elements increases.

Conformational cluster analysis shows that the trajectories of the simulations yield twenty nine distinct clusters (Tables 42 and 43). Fourteen clusters were identical between Schemes I and II and eight clusters were found to be similar between Schemes I and III. The other five clusters from Scheme I were replaced by four new clusters in Schemes II, while the remaining thirteen clusters of Scheme I were replaced with five clusters in Scheme III. One of the clusters from Scheme II was also located using the parameters of Scheme III. A detailed analysis was performed on the most representative structures. The central members of the most populated, clusters are shown in Figure 49. The predominant clusters from Scheme I match those of Scheme II. On the contrary, the highly populated cluster from Scheme III (45%) is similar to cluster **3** from Scheme I, which contains only 2% of the structures.

Table 42. Average torsion angles for each cluster from the *cp*PNA**4a** simulation

Cluster	No. of mem	%	T2		T3			T4		T5			T6		T7		
			β	δ	α	γ	ε	β	δ	α	γ	ε	β	δ	α	γ	ε
Scheme I																	
1	326	**3.3**	173	71	162	87	77	68	109	-179	91	70	57	77	167	83	-63
2	778	**7.8**	-73	74	-145	92	-102	68	108	-163	97	59	177	85	175	99	81
3	227	**2.3**	-72	103	-159	93	113	68	107	160	94	81	172	102	176	93	-75
4	965	**9.7**	-74	-91	-168	102	-150	68	102	-172	85	14	170	92	172	100	77
5	189	**1.9**	-96	-92	-169	109	-163	66	100	-163	83	20	167	88	178	93	-57
6	410	**4.1**	-64	114	178	101	23	67	104	-173	92	69	58	79	180	81	96
7	329	**3.3**	-72	105	150	97	51	65	104	-178	94	-35	176	82	-179	85	95
8	464	**4.6**	-65	110	-167	90	-155	67	107	152	99	50	169	87	179	107	60
9	208	**2.1**	-67	109	173	97	-7	69	105	-172	100	73	177	80	177	79	80
10	285	**2.9**	-62	105	-175	83	-139	68	111	146	101	16	170	89	178	104	-46
11	152	**1.5**	-66	118	-178	99	19	67	104	171	95	10	180	77	176	80	89
12	775	**7.8**	-70	102	-179	98	-146	67	110	-156	95	57	176	83	171	100	78
13	856	**8.6**	-175	96	180	88	27	66	108	176	92	67	61	87	176	79	90
14	667	**6.7**	-73	96	-176	98	65	67	105	-172	97	43	169	87	167	102	-114
15	883	**8.8**	-72	90	-180	92	-108	67	102	154	99	61	176	84	177	101	78
16	319	**3.2**	-66	115	-168	88	-157	66	106	-179	92	-16	173	88	-180	107	73
17	118	**1.2**	-64	118	180	100	-3	69	106	-176	89	-52	-180	80	174	78	84
18	1034	**10.3**	-71	93	170	88	-102	69	108	-154	98	60	176	84	-178	101	73
19	200	**2.0**	-70	93	176	87	-112	76	109	-176	95	-20	178	85	176	99	78
20	598	**6.0**	-73	93	-175	88	-91	72	104	-165	101	55	175	84	172	104	-89
21	200	**2.0**	-72	106	173	97	45	63	104	-172	94	59	175	78	174	79	76
Scheme II																	
22	85	**1.7**	-62	118	-177	103	24	66	106	-169	91	68	58	104	-136	83	99
23	87	**1.7**	-73	95	-180	97	58	66	106	-167	96	-2	172	88	178	111	-112
24	244	**4.9**	-66	117	-168	88	-157	66	104	-177	92	-17	172	87	179	107	78
25	187	**3.7**	-64	107	-172	89	-78	68	108	135	101	44	169	87	-179	107	53
Scheme III																	
26	219	**2.2**	-165	129	-174	97	14	67	107	-171	94	64	62	71	175	77	97
27	180	**1.8**	-168	92	-178	87	25	65	107	-175	92	76	57	120	-175	87	87
28	21	**0.2**	-169	88	172	88	-14	72	108	123	84	67	61	117	-174	88	94
29	303	**3.0**	-180	84	177	83	23	66	109	166	93	67	62	80	174	75	95

Table 43. Similar clusters identified from the different clustering Schemes for *cp*PNA**4a**

Cluster	No. of mem	%	Scheme I Cluster #[a]	Cluster	No. of mem	%	Scheme I, II Cluster #[a]
Scheme II				*Scheme III*			
A	162	**3.2**	1	A	409	**4.1**	1
B	428	**8.6**	13	B	4533	**45.3**	3
C	474	**9.5**	18	C	1460	**14.6**	14
D	652	**13.0**	2	D	1148	**11.5**	4
E	104	**2.1**	3	E	644	**6.4**	7
F	482	**9.6**	4	F	296	**3.0**	6
G	94	**1.9**	5	G	594	**5.9**	9
H	124	**2.5**	6	H	44	**0.4**	13
I	337	**6.7**	8	I	144	**1.4**	22
J	138	**2.8**	10				
K	465	**9.3**	14				
L	439	**8.8**	15				
M	113	**2.3**	21				
N	44	**0.9**	9				
Total number of clusters			18				13

[a] numbers correspond to identical cluster from Table 42

In the three most representative clusters from Scheme I (cluster **12**, **15** and **18**), which are equally populated, residues 2 to 4 form an inner loop. Cluster **4** from Scheme I and cluster **24** from Scheme II group structures that exhibit a wide hairpin loop. Cluster **2** contains structures that adopt an Ω-loop conformation spanning residues 2 to 7. Cluster **13** that was also identified in Schemes II and III very closely resembles the starting conformation. The number of conformations sampled is larger in the simulation of the *cp*PNA**4a** single strand than in the simulation of the *cp*PNA**4** with *L*-Lys for Scheme I and II and is similar for Scheme III. In addition, the largest population of *cp*PNA**4a** is ~10%, except for Scheme III (45%), while *cp*PNA**4** has a highly populated cluster (~40%) for all three Schemes. Comparison of the conformations of the most populated clusters revealed that the stereochemistry of lysine also has an effect on the shape of the PNA molecule. In the presence of *D*-Lys, PNA assumes a hairpin-like conformation while with *L*-Lys the main structure has a G-type fold.

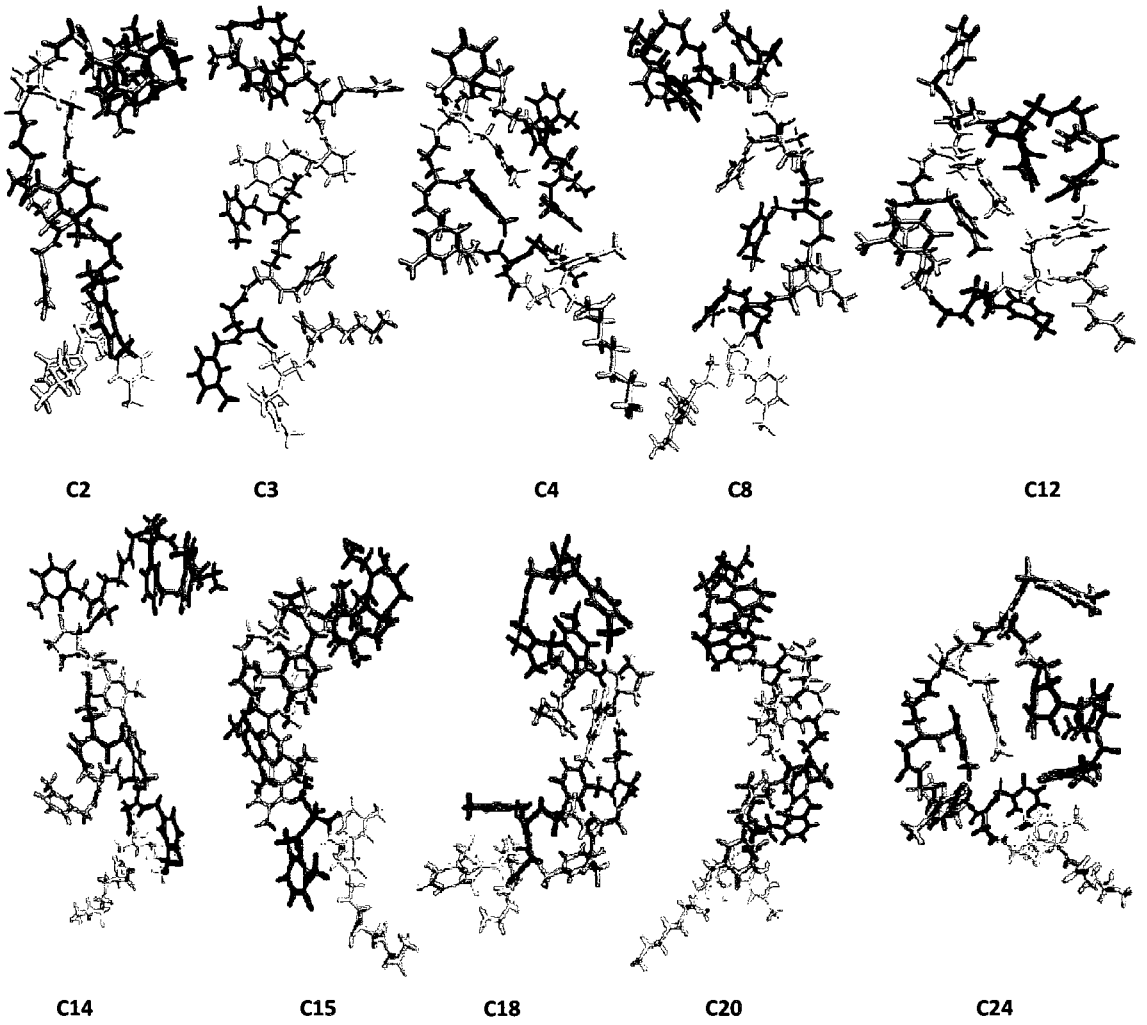

Figure 49. Representative structures of the most populated clusters from the MD simulation of *cp*PNA**4a.** Residues are colored by position: **T1**-red, **T2**-dark grey, **T3**-orange, **T4**-yellow, **T5**-dark yellow, **T6**-grey, **T7**-green, **T8**-white, and **K9**-pink. The structures are positioned with **T1** on top and **K9** on bottom

As noted previously for *cp*PNA**4**, *cp*PNA**4a** also has a rigid central region that is surrounded by flexible residues (Table 44). In particular, residues 1 to 3 along with the C-terminal residue 8 exhibit significant flexibility. However, unlike *cp*PNA**4**, the flexibility of residue 4 is slightly increased due to large fluctuations in the ε dihedral. On the whole, the flexibility of the *cp*PNA**4** strand is about the same as that of *cp*PNA**4a**.

Table 44. Average dihedrals (°) and RMSF (bold, °) of the simulated *cp*PNA**4a**

Residue	α	β	γ	δ	ε	Total RMSF	Average RMSF
T1	—	—	10.3	118.5	30.6		
	—	—	**93**	**56**	**76**	**225**	**75**
T2	178.7	-83.2	90.4	117.1	74.8		
	21	**38**	**41**	**59**	**86**	**245**	**49**
T3	-175.9	175.1	93.5	89.4	21.0		
	26	**50**	**15**	**17**	**130**	**238**	**48**
T4	-108.4	68.1	71.9	105.9	65.5		
	21	**8**	**11**	**11**	**118**	**169**	**34**
T5	-176.1	159.9	94.8	88.9	42.0		
	30	**41**	**14**	**17**	**42**	**144**	**29**
T6	172.1	155.5	87.0	85.4	63.0		
	28	**44**	**12**	**16**	**36**	**136**	**27**
T7	175.1	-29.4	95.7	92.6	41.7		
	23	**67**	**15**	**20**	**73**	**198**	**40**
T8	-179.7	-93.5	67.1	—	—		
	25	**66**	**71**	—	—	**162**	**54**

An initial analysis of the average angles yields distributions of torsions very similar to those found for the *cp*PNA**2** molecule. Thus, γ varies from 70 to 95°, δ assumes values in the range of 90-120°, ε is restricted to values in the range of 20-70°, and for the β torsion, various values are observed. There are, however, differences in the averages of several angles: β5, β7, γ1, γ2, γ8, ε3 and ε5. Thus, the *D*-Lys strand has different preferences for the β, δ and ε dihedrals when compared to the *L*-Lys strand. In summary, the *D*-Lys has a milder influence on the average angle distributions in comparison to the *L* configuration. Nevertheless, both cause a similar change in the pattern of the γ torsions.

For the modified strands, it can be concluded that the stereochemistry of the terminal Lys has a small impact on the flexibility and a large impact on the conformational dynamics of PNA. *Cp*PNA**4** (*L*-Lys) single strand conformations sampled a larger conformational space, i.e. broader dihedral angle distribution and more conformers than *cp*PNA**4a**. Thus, the torsional PES is different between the two strands. On the other hand, the presented results suggest that the entropic contribution to the DNA binding energy will be approximately the same for *L*- and *D*-Lys.

PNA-MB$_{10}$ sequence: *aeg*PNA**12** and *aeg*PNA**14**

To address the role of the terminal lysine in the mixed base strand, the RMSF profiles of *aeg*PNA**12** and *aeg*PNA**14** were compared. According to the results, there are many regions where residue mobility appears to be affected by the presence of the Lys. The most noticeable one is the increased rigidity of the T_1 to T_3 residues, and to a lesser extent the residues from T_5 to T_7. More surprising is that the neighboring residues, 4, 8 and 9, become more mobile. The change in overall flexibility due to the presence of the lysine is very pronounced and the lysine causes a reduction in the flexibility of the strand. While the average values of the β torsion were shifted and this dihedral experienced increased fluctuations at certain residues (4, 8 and 9) in *aeg*PNA**14**, the γ, δ and ε dihedrals (residues 2, 3, 5, 6 and 7) exhibited lower RMSF values in comparison to *aeg*PNA**12**. Comparison of the clustering shows that Schemes II and III gave fewer clusters for *aeg*PNA**14**, while Scheme I gave the same number of clusters. However, the main *aeg*PNA**14** clusters had slightly higher populations. This data is in agreement with the RMSF results.

Variations in the average torsion angles were also examined. The significant changes in the average values of the β and γ dihedrals were most likely caused by the Lys. The two strands have similar average values for the β torsion in residues 1, 3, 7, 9 and 10, but different values for the remainder of the residues. The largest difference between the two strands is the alteration in the γ average values. In the strand with Lys, all of the γ torsions are around 90°, whereas a range of values was seen for *aeg*PNA**12**.

The average δ and ε backbone dihedral angles had similar values to the corresponding ones from *aeg*PNA**12**, showing changes only for a few residues.

PNA-MB$_{10}$ sequence: *cp*PNA**13** and *cp*PNA**15**

The average residue fluctuations of the *cp*PNA**13** and *cp*PNA**15** were examined and compared to further assess the effect of the lysine. When the lysine is added, the region at and around the modification site (residues 6, 7 and 8), remained fairly rigid, with even lower RMSFs for residues 7 and 8 in *cp*PNA**15**. Except for the C-terminal region (residues 9 and 10) and T_4-T_6, all residues showed lower flexibility in *cp*PNA**15** compared to *cp*PNA**13**. As a consequence, the *cp*PNA**15** is somewhat more constrained than *cp*PNA**13**, though a small increase in flexibility was observed for residues 9 and 10. These changes in chain mobility reflect the influence of the lysine and suggest that residues 1 to 3 of the N-terminus experience lower fluctuations whenever a Lys is attached at the C-terminus. Thus, the lysine affects the dynamics of the entire strand and not just the residue to which it is attached. Evaluation of the dihedral RMSFs shows that the reduced fluctuations of specific dihedrals (δ3, δ7, ε2 and ε5) are largely a consequence of the Lys attached at the terminus. The clustering seems to show that the two strands have equivalent flexibility since fewer clusters were observed for Scheme I for *cp*PNA**15** but more clusters were observed for Scheme III. This data reinforces the RMSF results which show only a slight decrease in flexibility for *cp*PNA**15**.

Analysis of the average values reveals a wider distribution of the β dihedral as opposed to γ, δ or ε, which have similar ranges to the dihedrals in *cp*PNA**13**. Major changes for the β dihedrals are only observed in the central or C-terminal regions (residues 6-10). The γ torsion exists in its original range of 80-90°. Likewise, δ stays close to its preferred values of ~100°, while ε is either around -60° or 60°, except in residue 9.

Computation of the conformational properties of *aeg*PNA**12**, *aeg*PNA**14**, *cp*PNA**13** and *cp*PNA**15** gives a picture on the effect of the lysine residue at the C-terminus. The attachment of *L*-Lys had a pronounced influence on the average values of the β torsions. The conformational constraint of the PNA strand by the *L*-Lys was

evidenced by the differences in RMSFs as well as cluster numbers, though the effect was much more pronounced for the unmodified strand. In addition, dihedral angle fluctuation changes were distinct for specific δ and ε torsions. The lower number of clusters and RMSF values found for *aeg*PNA**14** and *cp*PNA**15** indicate that in the presence of the *L*-Lys the flexibility of the strand is reduced which could counterbalance entropy loss upon binding.

The effect of the stereochemistry of the lysine

The RMSD plot from the simulation of the mixed base strand containing *D*-Lys, *cp*PNA**15a**, provides insight into the dynamical behavior of the system. The graph clearly shows that the major structural deviations occur in the 3-5 ns range, and unlike in *cp*PNA**4a**, the RMSD does not go above 7 Å. After 3 ns the helical geometry is lost in favor of a more turn-like geometry. This conformational change is paralleled by an increase in the RMSD value with respect to the NMR starting geometry. A hairpin loop over the residues near the N-terminal region is formed at 3.5 ns of this simulation. At 7 ns another coiled structure is observed that is stable for the rest of the simulation time. Clustering was applied to define the most populated structures during the long MD simulation (Tables 45 and 46, Figure 50).

Cluster analysis yielded fourteen distinct clusters. There are no significant differences in the most abundant structures obtained from the three Schemes. Four clusters were identical between Schemes I and II and three Scheme I clusters were located with Scheme III. The remaining clusters from Scheme 1 were replaced by six new clusters in Scheme II and two new clusters in Scheme III. The population of the top two clusters from Scheme I (**3** and **6**) represents 63% of the total number of conformations.

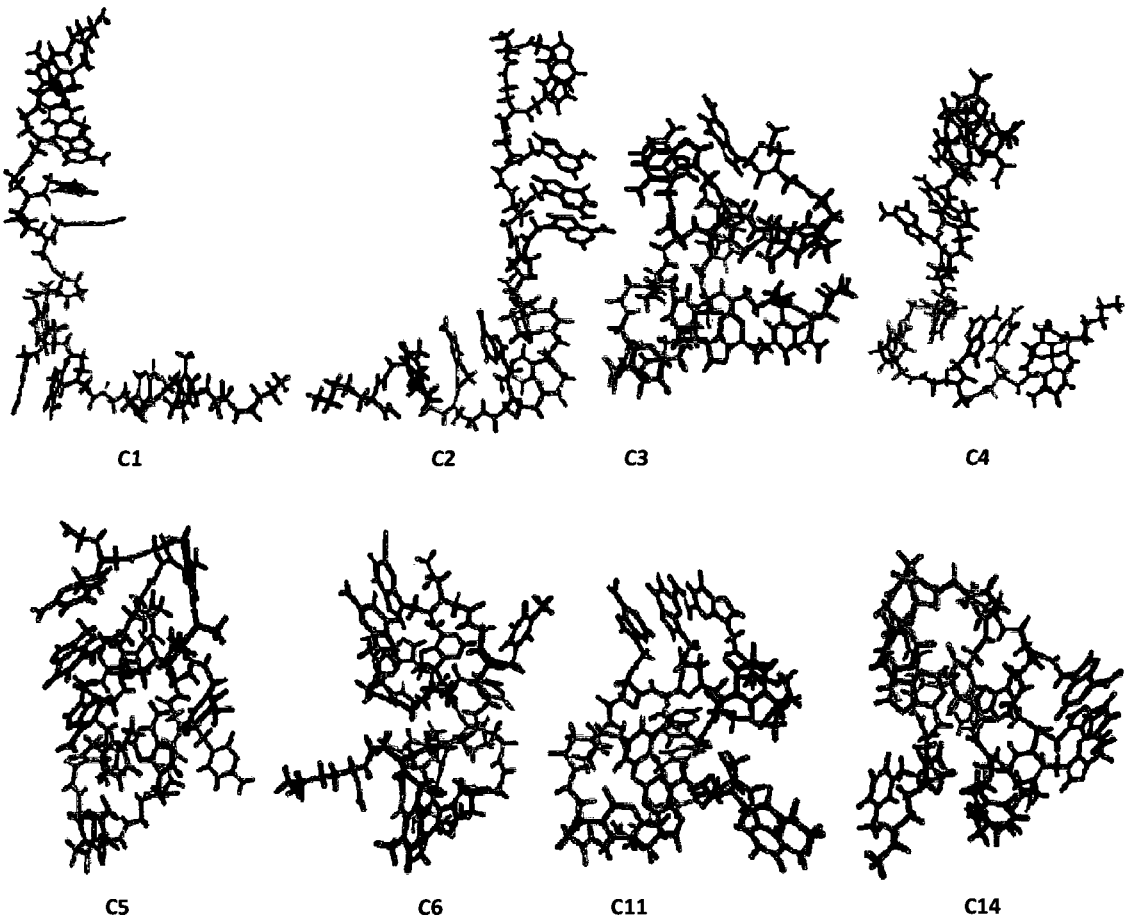

Figure 50. Representative structures of the most populated clusters from the MD simulation of *cp*PNA**15a**. Residues are colored by name: **APN**-blue, **CPN**-orange, **TPN**-green, **TPC**- light blue, **GPN**-red and **K**-purple. The structures are positioned with **GPN1** on top and **K11** on bottom (see Table 17 for the sequence)

Table 45. Average torsion angles for each cluster from the *cp*PNA**15a** simulation

Cluster	No. of Members	%	T2		A3			G4		G5			A6		C7			A8		C9		
			β	δ	α	γ	ε	β	δ	α	γ	ε	β	δ	α	γ	ε	β	δ	α	γ	ε
Scheme I																						
1	1193	**11.9**	66	83	178	83	69	63	81	176	80	132	68	107	161	86	30	63	94	171	91	69
2	713	**7.1**	158	-102	176	90	65	63	82	172	83	-161	67	104	171	90	-69	96	87	173	85	88
3	1944	**19.4**	-115	-83	158	95	73	-75	70	172	71	-139	71	103	177	100	-101	175	90	177	101	-141
4	754	**7.5**	-127	-86	175	83	8	93	102	179	98	-132	65	112	170	91	-91	177	96	174	86	-62
5	983	**9.8**	-118	-100	156	99	53	-68	60	157	79	-120	73	100	179	101	-101	173	96	168	94	79
6	4413	**44.1**	-154	87	167	91	73	-53	-88	171	62	-167	66	102	177	84	-99	173	90	162	86	158
Scheme II																						
7	168	**3.4**	64	92	177	87	68	65	84	176	78	-171	66	104	169	88	-68	65	119	173	91	90
8	207	**4.1**	177	-94	158	95	40	-65	62	148	87	-111	70	104	173	101	-107	168	93	168	91	87
9	238	**4.8**	-89	-95	173	86	61	58	98	169	98	-145	67	106	170	94	-98	178	85	175	88	-43
10	161	**3.2**	-174	-75	173	81	-62	128	101	161	96	-114	64	119	171	90	-77	171	111	171	84	-81
11	811	**16.2**	-74	-108	154	100	68	-73	68	173	70	-143	74	99	177	99	-100	175	90	178	101	142
12	265	**5.3**	-177	89	162	90	72	-58	-141	-140	-46	167	69	100	171	84	-99	174	92	159	83	137
Scheme III																						
13	673	**6.7**	170	-72	175	77	12	24	85	155	91	-112	65	116	169	97	-91	177	115	177	88	-78
14	2567	**25.7**	-105	-92	155	99	66	-73	66	169	72	-135	72	101	180	100	-100	174	89	177	100	114

Table 46. Similar clusters identified from the different clustering Schemes for *cp*PNA**15a**

Cluster	No. of members	%	Scheme I Cluster #[a]	Cluster	No. of members	%	Scheme I Cluster #[a]
Scheme II				*Scheme III*			
A	434	**8.7**	1	A	1196	**12.0**	1
B	451	**9.0**	3	B	1137	**11.4**	2
C	1916	**38.3**	6	C	4427	**44.3**	6
D	349	**7.0**	2				
Total number of clusters			10				5

[a] numbers correspond to identical cluster from Table 45

Comparison of the cluster analysis obtained from *cp*PNA**15** and *cp*PNA**15a**, suggests that the *cp*PNA**15a** explores a greater diversity of structures. More clusters were characterized for *cp*PNA**15a** than for *cp*PNA**15** using Schemes I and II. The reverse is observed for Scheme III though six of the *cp*PNA**15** clusters have populations less than 5%. On the other hand, comparison of the *cp*PNA**15** conformations with the *cp*PNA**15a** conformations from the dominant clusters shows that similar hairpin-like structures were adopted in the simulations of both strands. Therefore, the configuration of lysine does not appear to have an effect on the overall shape of the *cp*PNA molecule.

A residue-based description of the strand flexibility was obtained by calculating the RMSF values for *cp*PNA**15a** (Table 47). The peptide has a rigid central core that is surrounded by flexible residues. In particular, residues 2, 4, 9 and 10 exhibit significant flexibility. The data in Table 47 indicates that when the *D*-Lys is present the N- and C-terminal regions (residues 1 to 4) are more flexible than in the case when the amino acid is absent. We also see a significant decrease in the flexibility at the modification site, residues 6 to 8. The switch from *L*-Lys to *D*-Lys also has an effect on the flexibility of the strand. For example, when Lys is in the *D* configuration (*cp*PNA**15a**), the resultant RMSF values for the central residues are lower or approximately the same. On the other hand, the flexibility of the *cp*PNA**15a** C-terminus region has significantly increased. The overall flexibility of the strand is enhanced considerably by *D*-Lys compared to *L*-Lys and a strand with no lysine (*cp*PNA**13**).

Comparison of the average angles shows that the γ and δ torsions have values very similar to those found in *cp*PNA**15**. Thus, γ varies from 70 to 100° and δ from 80° to100°, except for δ2, δ4, γ1 and γ10, which have differences in their average values. As for the β and ε torsions, they are found to have broadly distributed values with ε having an increased distribution over *cp*PNA**15**. To conclude, the presence of the *D*-Lys has a stronger effect on the average angle values than the *L* isomer since the *L* isomer had almost no effect. Nevertheless, both show a similar pattern of changes in the δ torsion values.

Table 47. Average dihedrals (°) and RMSF (bold, °) of the simulated *cp*PNA**15a**

Residue	α	β	γ	δ	ε	Total RMSF	Average RMSF
G1	—	—	82.6	94.2	103.2		
	—	—	**44**	**18**	**81**	**143**	**48**
T2	-176.1	-160.9	96.9	-34.6	-54.2		
	22	**73**	**17**	**125**	**64**	**301**	**60**
A3	-167.5	-86.6	90.6	78.4	64.8		
	27	**141**	**13**	**20**	**32**	**233**	**47**
G4	174.1	-25.6	105.9	3.5	122.2		
	19	**62**	**26**	**86**	**94**	**287**	**57**
A5	179.1	75.5	71.7	80.6	-161.3		
	30	**37**	**32**	**22**	**44**	**165**	**33**
T6	-109.9	67.9	74.6	103.4	8.1		
	17	**8**	**13**	**11**	**23**	**72**	**14**
C7	177.3	-178.8	90.1	78.1	-81.3		
	25	**17**	**15**	**16**	**51**	**124**	**25**
A8	-155.3	155.3	94.5	91.4	37.1		
	28	**43**	**14**	**20**	**56**	**161**	**32**
C9	172.5	-36.2	90.2	88.9	-104.7		
	26	**99**	**13**	**18**	**125**	**281**	**56**
T10	-175.9	173.5	26.3	—	—		
	31	**32**	**89**	—	—	**152**	**51**

In summary, the simulations on the mixed base strands show that in the presence of the C-terminal *L*-Lys, the mobility of the entire strand is reduced. This is correlated with the clustering data where the decreased number of clusters also indicates a lower conformational freedom of PNA. On the contrary, attachment of *D*-Lys has the opposite effect and enhanced fluctuations are observed along with a higher variability in the structures from clustering. This indicates that the terminal lysine can also play a role in stabilizing (or destabilizing) the duplexes. These results differ significantly from the polythymine studies where the Lys either increased or had no effect on the flexibility. The Lys also had an effect on the sampled conformational space for both MB_{10} and T_8 sequences, though no consistent effect could be detected.

THE EFFECT OF THE POSITION

To investigate the effect of the location of the modified cyclopentane on the structural properties, a comparative analysis of oligomers having a single modification at various positions was performed.

PNA-T_8 sequence: *cp*PNA**5** and *cp*PNA**6**

The RMSD for the *cp*PNA**5**, which has the modification at T_1, showed that large fluctuations were occurring during the simulation and the RMSD continually increased to 10 Å until 5 ns. Then, the RMSD value stabilized and ranged from 9.0 to 10.5 Å for the remainder of the simulation. Twenty four different clusters were characterized, but there were only two distinct patterns (Tables 48 and 49). The structures are provided in Figure 51. The most populated clusters obtained using all three criteria have very similar dihedral angle values. Fourteen clusters from Scheme II were identical to those from Scheme I and the remaining three clusters from Scheme I were replaced by three new clusters in Scheme II. Five of the clusters from Scheme I were also identified in Scheme III. The twelve clusters of Scheme I were replaced by only four new clusters in Scheme III.

The most abundant pattern from Scheme I, clusters **3, 4, 6** and **7** that were also located with Schemes II and III, exhibit a hairpin motif with two loops, one in the N-terminal and one in the C-terminal region. Cluster **20** from Scheme II and cluster **22** from Scheme III also have a hairpin like structure. The second most abundant clusters from Scheme I (**6, 7**) represent a knot expanding from residues 4 to 6. Similar conformations, although with lower cluster memberships, were located with Schemes II and III and are represented by clusters **19** and **23**, respectively.

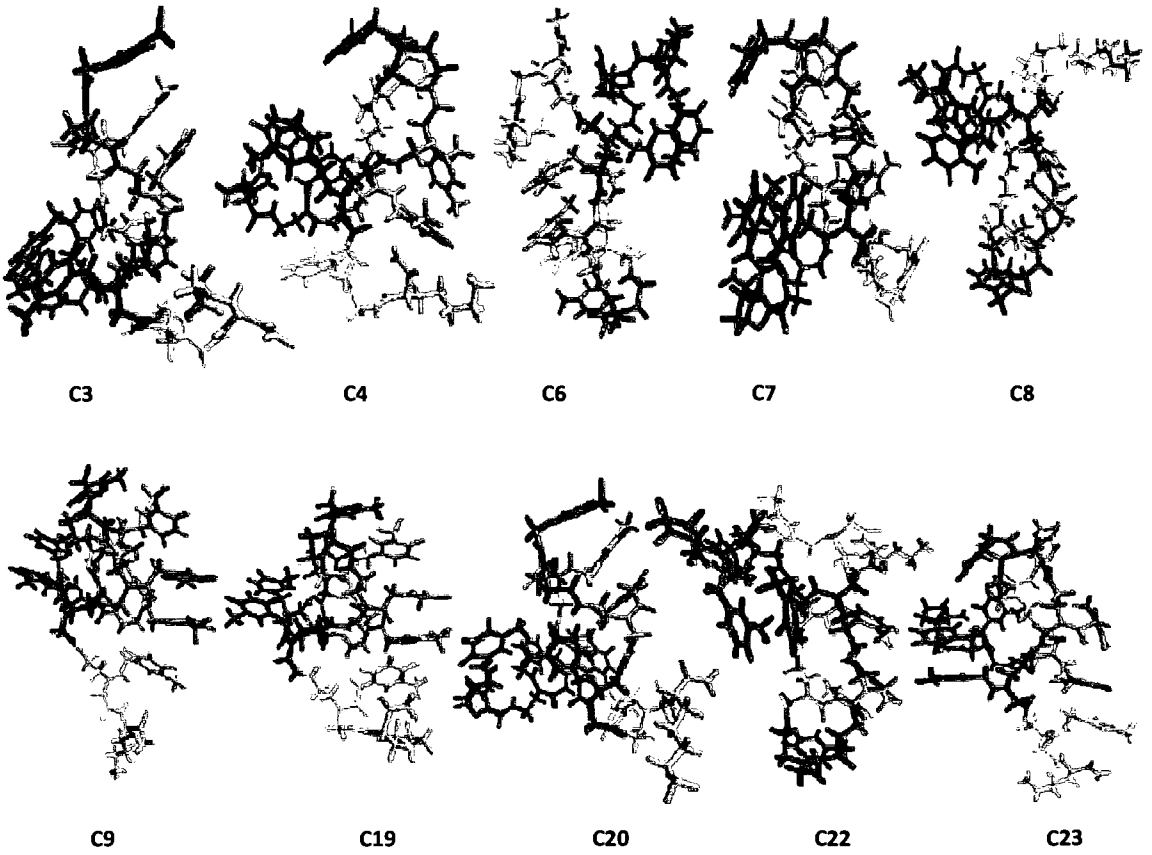

Figure 51. Structures of cluster centers of most populated clusters from *cp*PNA**5** simulations. Residues are colored by position: **T1**-red, **T2**- dark grey, **T3**-orange, **T4**-yellow, **T5**-dark yellow, **T6**-grey, **T7**-green, **T8**-white, and **K9**-pink. The structures are positioned with **T1** on top and **K9** on bottom

Table 48. Average torsion angles for each cluster from the *cp*PNA**5** simulation

Cluster	No. of members	%	T2		T3			T4		T5			T6		T7		
			β	δ	α	γ	ε	β	δ	α	γ	ε	β	δ	α	γ	ε
Scheme I																	
1	113	**1.1**	55	96	-167	86	71	60	118	-176	86	42	64	90	179	84	68
2	215	**2.2**	63	63	167	81	-76	166	115	-180	83	33	60	84	178	79	73
3	1976	**19.8**	57	98	163	95	91	177	97	163	93	-108	175	76	180	102	-92
4	626	**6.3**	55	106	177	90	62	-179	93	176	-94	63	-173	98	-178	103	-106
5	446	**4.5**	50	103	172	94	-146	174	97	-175	-82	-71	-157	80	-179	101	-174
6	652	**6.5**	-53	96	175	92	124	-179	113	176	-89	3	-176	120	177	97	-119
7	655	**6.6**	57	100	171	87	44	-178	82	178	-95	66	-172	97	-179	108	100
8	539	**5.4**	49	104	169	94	103	180	99	-159	-83	-78	-157	96	178	98	-118
9	3868	**38.7**	56	36	157	-114	60	172	99	176	-81	-41	-170	78	177	102	106
10	82	**0.8**	57	62	166	80	-104	-176	98	-165	87	-57	70	50	156	93	-70
11	127	**1.3**	168	96	-172	106	-86	179	112	154	100	-87	104	112	175	99	-75
12	50	**0.5**	58	70	143	77	-64	-169	97	-169	86	40	57	48	147	89	-67
13	208	**2.1**	65	91	-172	105	-67	172	98	-165	95	-78	171	99	-165	104	-68
14	38	**0.4**	158	89	164	101	-85	173	100	-164	94	58	80	106	-176	96	-85
15	74	**0.7**	61	84	-171	109	-72	174	100	-171	93	-95	175	100	134	102	-71
16	261	**2.6**	53	94	-175	106	-77	158	81	155	90	-134	-175	99	-172	95	-61
17	70	**0.7**	51	92	-168	90	42	-168	77	-169	-81	-88	-86	85	171	102	91
Scheme II																	
18	90	**1.8**	61	107	167	94	-156	175	93	-177	-81	-73	-170	77	-179	106	115
19	582	**11.6**	57	37	160	-112	62	172	92	177	-92	34	-176	81	174	103	109
20	426	**8.5**	57	98	178	95	88	179	101	163	95	-116	176	84	-174	109	-55
Scheme III																	
21	327	**3.3**	60	75	176	83	-23	130	116	-179	84	36	61	86	178	81	71
22	1717	**17.2**	10	100	172	93	141	179	105	-174	-85	-40	-167	103	179	99	-133
23	1300	**13.0**	57	102	176	83	48	-178	85	178	-94	56	-166	95	-180	105	173
24	375	**3.8**	108	92	-176	104	-73	174	101	-168	95	-65	145	105	-175	102	-73

Table 49. Similar clusters identified from the different clustering Schemes for *cp*PNA**5**

Cluster	No. of members	%	Scheme I Cluster #[a]	Cluster	No. of members	%	Scheme I Cluster #[a]
Scheme II				*Scheme III*			
A	44	**0.9**	1	A	305	**3.1**	16
B	107	**2.1**	2	B	1979	**19.8**	3
C	126	**2.5**	13	C	3829	**38.3**	*9*
D	558	**11.2**	3	D	102	**1.0**	10
E	263	**5.3**	8	E	66	**0.7**	12
F	109	**2.2**	17				
G	313	**6.3**	4				
H	324	**6.5**	7				
I	1353	**27.1**	9				
J	40	**0.8**	10				
K	62	**1.2**	11				
L	133	**2.7**	16				
M	152	**3.0**	5				
N	314	**6.3**	6				
Total number of clusters			17				9

[a] numbers correspond to identical cluster from Table 48

For *cp*PNA**5**, the number of clusters obtained is higher than for *cp*PNA**4** for Schemes I and II, but both have a similar number of highly populated clusters; though, the most populated cluster for *cp*PNA**4** has a slightly larger population. This finding suggests that the flexibility of the oligonucleotide is increased or approximately the same for *cp*PNA**5**. Comparing to *aeg*PNA**3**, the number of clusters are approximately the same, but *cp*PNA**5** has a highly populated cluster, which *aeg*PNA**3** lacks. Thus, the T_1 modification seems to have slightly reduced the flexibility of the strand. The comparison between cluster conformations of *cp*PNA**4** and *cp*PNA**5** reveals that modification at the N-terminal region affects the backbone structure as well, since hairpin-like conformations are sampled instead of wide G-type loops.

By analyzing the RMSF values of the modified T_1 residue and other nearby bases, important changes in the flexibility are observed (Table 50). For example, the RMSF values of T_1 and T_2 are clearly lower than those observed for the *aeg*PNA**3** and *cp*PNA**4** simulations. Several distant residues also have decreased fluctuations (4, 6 and 7). On the

other hand, T_3, T_5 and T_8 have a similar flexibility as in the unmodified strand (*aeg*PNA**3**) (see Table 12). The residues with the larger RMSF values in the *cp*PNA**5** strand correspond to the larger fluctuations in the RMSD relative to the crystal structure as it is alterations in this section of the strand that leads to the high RMSD. Thus, both modification sites caused a decrease in dynamics around the modification site and at distant locations from the modification, so that the overall strand flexibility was greatly reduced from *aeg*PNA**3**.

Table 50. Average dihedrals (°) and RMSF (bold, °) of the simulated *cp*PNA**5**

Residue	α	β	γ	δ	ε	Total RMSF	Average RMSF
T1	—	—	-89.3	90.2	39.0		
	—	—	**21**	**12**	**29**	**62**	**21**
T2	173.3	50.7	75.9	73.4	86.1		
	22	**34**	**15**	**35**	**46**	**152**	**30**
T3	165.9	-39.6	13.0	90.1	64.2		
	28	**95**	**102**	**21**	**65**	**311**	**62**
T4	176.4	174.2	19.4	97.7	60.0		
	22	**17**	**89**	**16**	**48**	**192**	**38**
T5	176.2	-99.5	-29.2	92.5	-43.2		
	22	**61**	**83**	**28**	**71**	**265**	**53**
T6	-170.7	-179.6	94.9	85.4	-126.7		
	37	**34**	**16**	**20**	**60**	**167**	**33**
T7	178.6	174.6	100.9	-104.4	-172.5		
	22	**28**	**14**	**38**	**97**	**199**	**40**
T8	179.6	156.2	13.2	—	—		
	23	**47**	**93**	—	—	**163**	**54**

As in *aeg*PNA**3** and *cp*PNA**4**, the largest fluctuations were recorded for the ε dihedrals, but they were also much smaller in *cp*PNA**5**. Some of the flexible residues (3 to 5) that occasionally participate in base stacking interactions had γ torsions with increased fluctuations when compared to the unmodified *aeg*PNA**3** and *cp*PNA**4**. Unlike *aeg*PNA**3**, but similar to *cp*PNA**4**, the lowest fluctuations are observed for the β and δ torsions, though the δ fluctuations are significantly lower in *cp*PNA**5**. The angles that

were affected by the modification in a similar way to *cp*PNA**4** include β4, γ2 and δ4. These changes in the conformational mobility reflect the influence of the ring modification at this specific position. Thus, the location of the ring does not appear to matter for some changes in dynamics but does matter for others.

Evaluation of the average angles of *aeg*PNA**3** and *cp*PNA**5** shows considerable change in the β2 and β5 values, 175 to 51° and 171 to -99°, respectively. Most of the residues have β values in the 155-175° range (except for 2, 3 and 5). For the γ dihedrals, two ranges of average values were observed: 80-100° (residues 2, 6, 7) and 13-20° (residues 3, 4, 8). The δ and ε torsions did not stay within a specific range. Consequently, the chief differences between *aeg*PNA**3** and *cp*PNA**5** are in the δ and ε angles and the previously mentioned β angles. In addition, the γ angles in residues 3 to 5 undergo particularly large changes in their average values. When the modification is at T_4 (*cp*PNA**4**), the β and δ torsions show changes in average values compared to the unmodified PNA. On the other hand, β angles remain intact in *cp*PNA**5** and changes are observed mainly for δ and ε torsions.

Analysis of the *cp*PNA**6** trajectory, with the modification at residue 8, showed that at 4.5 ns the RMSD value jumped from 3.5 to 6.2 Å. After 5 ns, the RMSD value ranged from 3.0 to 5 Å. Formation of a hairpin structure in the backbone together with the formation of a turn in the C-terminal region was observed during the first 5.1 ns of the simulation, analogous to the observations for the *cp*PNA**5** simulations. Using cluster analysis, the structures were classified into twelve distinct clusters (Figure 52, Tables 51 and 52). The most populated clusters from Schemes I and II have similar structures and dihedrals. However, the second predominant cluster from Scheme III has very different dihedral angle values. Five clusters were similar between Schemes I and II and five clusters from Scheme III were identical to those of Scheme I. The remaining two clusters from Scheme I were replaced by three new clusters in Scheme II. Analogously, two of the clusters from Scheme I were substituted by four new clusters in Scheme III. One of the new clusters from Scheme III was similar to a new cluster in Scheme II. The most abundant clusters from Scheme I (**4** and **5**) that were also located with Schemes II and III group structures with a hairpin motif. The second most populated conformation adopts an extended, helical form and is represented by clusters **2** and **3** from Scheme I. Cluster **11**

from Scheme III also exhibits a similar conformation, though with a higher population percentage.

A different number of clusters were found for the different simulations. *Aeg*PNA**3** and *cp*PNA**5** have a similar number of clusters, while *cp*PNA**4** has less clusters and *cp*PNA**6** has even fewer clusters. However, the *cp*PNA**6** most populous cluster has a higher membership than the clusters of the other singly modified PNA strands. This result indicates that the addition of the cyclopentane ring at the C-terminus decreased the flexibility of the strand more than modification at the two other locations tested. The PNA conformations sampled in both simulations of *cp*PNA**5** and *cp*PNA**6** resemble hairpin-like structures. Consequently, the modification at either of the termini has a similar effect on the overall shape of the structure.

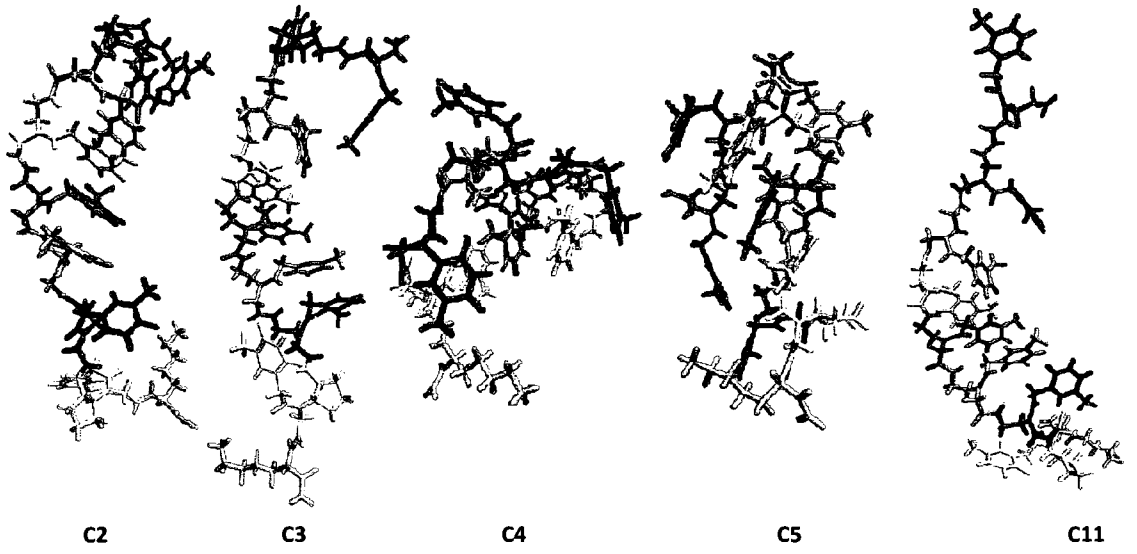

Figure 52. Representative structures of the most populated clusters from the MD simulation of *cp*PNA**6**. Residues are colored by position: **T1**-red, **T2**- dark grey, **T3**-orange, **T4**-yellow, **T5**-dark yellow, **T6**-grey, **T7**-green, **T8**-white, and **K9**-pink. The structures are positioned with **T1** on top and **K9** on bottom

Table 51. Average torsion angles for each cluster from the *cp*PNA**6** simulation

Cluster	No. of Members	%	T2		T3			T4		T5			T6		T7		
			β	δ	α	γ	ε	β	δ	α	γ	ε	β	δ	α	γ	ε
Scheme I																	
1	599	**6.0**	160	82	171	81	-55	176	79	-179	82	69	60	75	174	82	-105
2	889	**8.9**	101	114	-173	96	74	63	79	178	81	68	61	78	171	84	152
3	692	**6.9**	176	-98	178	92	79	-66	73	-178	86	62	67	77	-178	84	155
4	5403	**54.0**	173	-102	167	-104	53	-64	77	176	80	81	115	97	179	93	-109
5	1989	**19.9**	-174	101	-172	-97	68	-77	83	173	74	24	161	94	172	97	-131
6	26	**0.3**	-169	89	-171	-120	80	-79	82	174	75	-34	175	69	-97	-81	-179
7	402	**4.0**	178	-108	160	-91	-79	53	61	165	82	49	148	94	-172	90	-103
Scheme II																	
8	255	**5.1**	131	114	-168	97	73	65	80	178	80	70	61	81	175	84	-153
9	227	**4.5**	63	115	-177	93	78	62	78	177	81	67	61	73	165	85	79
10	322	**6.4**	-167	98	-172	-107	68	-78	82	172	76	-50	173	93	170	96	-135
Scheme III																	
11	1984	**19.8**	142	-165	-179	118	73	-6	76	-180	84	65	64	78	177	84	172
12	280	**2.8**	179	102	-171	-97	66	-78	82	172	76	-52	175	87	166	104	-135
13	161	**1.6**	-125	90	-175	-125	71	-77	79	174	77	-5	143	77	125	87	-153

Table 52. Similar clusters identified from the different clustering Schemes for *cp*PNA**6**

Cluster	No. of members	%	Scheme I Cluster #[a]	Cluster I	No. of members	%	Scheme I Cluster #[a]
Scheme II				*Scheme III*			
A	274	**5.5**	1	A	598	**6.0**	1
B	346	**6.9**	3	B	4994	**49.9**	4
C	2699	**54.0**	4	C	748	**7.5**	5
D	673	**13.5**	5	D	405	**4.1**	7
E	201	**4.0**	7	E	265	**2.7**	10
				F	563	**5.6**	5
Total number of clusters			8				9

[a] numbers correspond to identical cluster from Table 51

A residue-based description of the local flexibility was obtained by calculating the RMSF values (Table 53). Examination of the RMSFs showed that residues 1, 2 and 3 are the most flexible. The residues with the smallest dynamical motions are 4 and 5, which are closely associated with stacking, and 8. The RMSF values in Table 53 indicate that when T_8 is modified, residue T_1 becomes more flexible compared to its counterpart in unmodified *aeg*PNA**3**. Moreover, the *cp*PNA**6** simulation yields a larger region (residues 4 to 8) of less flexible residues than is observed for *aeg*PNA**3**. As with *aeg*PNA**3** and *cp*PNA**5**, the T_3 residue exhibits significant flexibility. This observation was found to be relatively constant over all calculations, unless T_4 was modified. The strand flexibility was essentially the same for *cp*PNA**4**, *cp*PNA**5** and *cp*PNA**6**, with *cp*PNA**4** having slightly higher fluctuations; though, the individual residue dynamics are dependent on the location of the modification. The reduced flexibilities of *cp*PNA**4**, *cp*PNA**5** and *cp*PNA**6** indicate that modification at residues 1, 4 or 8 should lead to tighter binding to DNA/RNA, provided the modifications do not restrict the strand to dihedrals that would prevent binding.

Table 53. Average dihedrals (°) and RMSF (bold, °) of the simulated *cp*PNA**6**

Residue	α	β	γ	δ	ε	Total RMSF	Average RMSF
T1	—	—	-140.9	93.7	-149.7		
	—	—	**161**	**23**	**95**	**279**	**93**
T2	179.8	169.1	87.6	29.5	-110.8		
	22	**31**	**14**	**87**	**79**	**233**	**47**
T3	173.6	24.1	-59.5	89.7	47.9		
	28	**156**	**81**	**19**	**45**	**329**	**66**
T4	168.5	-36.4	99.5	77.7	71.1		
	22	**70**	**14**	**14**	**18**	**138**	**28**
T5	175.6	60.3	79.4	98.8	65.0		
	19	**12**	**11**	**27**	**41**	**110**	**22**
T6	-179.5	114.4	81.4	91.5	-54.0		
	24	**51**	**12**	**22**	**82**	**191**	**38**
T7	177.3	165.3	91.1	84.6	-128.4		
	27	**70**	**17**	**19**	**49**	**182**	**36**
T8	-108.4	73.1	58.5	—	—		
	18	**18**	**21**	—	—	**57**	**19**

The ε torsion shows the greatest variability in *cp*PNA**6**, while the β torsions of the central region exhibit larger flexibilities in comparison to the terminal β torsions. This is partly consistent with the results for PNA**3-5**. The torsion angles γ (except T_1 and T_3) and δ (except T_2) exhibited very low RMSF values. When compared to the unmodified strand, reduced fluctuations are observed for the β, δ and ε dihedrals of the central residues (except T_6). The β and δ dihedrals have even lower RMSFs than in *cp*PNA**4** or *cp*PNA**5**, whereas the γ torsions have higher fluctuations. The specific angles that were influenced similarly as in *cp*PNA**4** and *cp*PNA**6** are γ2, γ8, δ2, δ4 and β7, γ2, γ3, δ3, δ4, respectively. In summary, the two angles whose dynamics were most affected by a constraint, irrespective of the modification position, were γ2 and δ4.

The torsional angle data for *cp*PNA**6** shows that aside from residue 8, the α dihedral adopts various values between 168° and -179°. Restriction of the torsion angle δ (60-90°) can be noticed when compared to *aeg*PNA**3**. The ε dihedral angle also changes

somewhat; it now lies at ~60° in the central region and assumes values from -110 to -150° for the terminal residues. In addition, the preferred dihedral values for β differ between the *aeg*PNA**3** and *cp*PNA**6** strands. The average values of the β dihedrals were affected in both the *cp*PNA**4** and the *cp*PNA**6** strands. Changes in averages were observed for ε in *cp*PNA**5** and *cp*PNA**6**. Thus, the addition of a cyclopentane ring does cause changes in the torsional PES, but the types of alterations are location dependent. For productive binding to DNA/RNA, the modified strands should assume backbone dihedrals that allow for standard Watson-Crick base pairing without major strain. A more detailed comparison between these strands and DNA/RNA is performed in a later section.

PNA-MB$_{10}$ sequence: *cp*PNA**16**, *cp*PNA**17** and *cp*PNA**18**

In the case of *cp*PNA**16**, which has the modification at residue G_4, the RMSD was around 3 Å for the initial 3 ns, then continually increased to over 8 Å and finally stabilized at ~7 ns. In comparison to *cp*PNA**15**, the secondary structure content is relatively stable: a coil structure involving residues 2 to 8 appears several times (1.1 ns and 6.7 ns) during the simulation, while the N-and C-terminal regions remain mostly disordered. Clustering resulted in three highly populated clusters for Schemes I and II and two highly populated clusters for Scheme III (Figure 53, Tables 54 and 55). There are no major differences in the predominant structures located with Schemes I and III. On the contrary, most populated clusters from Schemes I and II show very different dihedral angle values. Only two clusters were found to be similar between Schemes I and II. The remaining nine clusters from Scheme I were substituted by three new clusters in Scheme II. Two clusters from Scheme III matched those of Scheme I and only one cluster was located using the parameters of Scheme II. Even though limited sampling of the potential energy surface occurred during this simulation, many backbone states appear to be possible. The most abundant pattern from Schemes I and II, clusters **1**, **10**, **12** and **13**, has a bent hairpin loop that appears at 3.3 ns. The second (clusters **2** and **4**) and third (cluster **9**) most populated patterns of Scheme I represent two different coil structures expanding from residues 2 to 8 and 4 to 10, respectively. The geometry of cluster **9** is very similar to that observed in the NMR structure of the PNA:DNA complex.

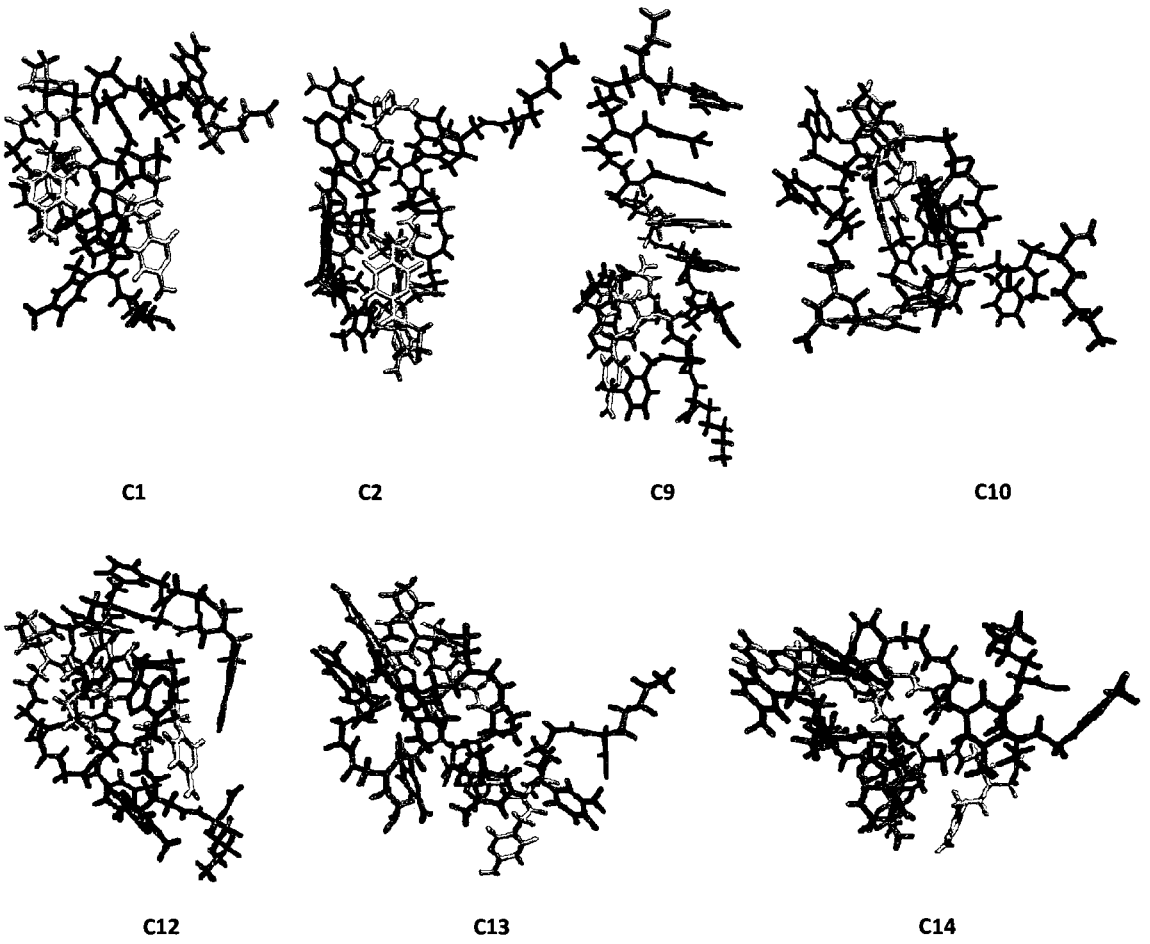

Figure 53. Representative structures of the most populated clusters from the MD simulation of *cp*PNA**16**. Residues are colored by name: **APN**-blue, **CPN**-orange, **TPN**-green, **GPC**- light green, **GPN**-red and **K**-purple

Table 54. Average torsion angles for each cluster from the *cp*PNA**16** simulation

Cluster.	No. of Members	%	T2		A3			G4		G5			A6		C7			A8		C9		
			β	δ	α	γ	ε	β	δ	α	γ	ε	β	δ	α	γ	ε	β	δ	α	γ	ε
Scheme I																						
1	3339	**33.4**	69	89	178	75	-29	69	108	178	76	96	-46	-57	180	106	-90	157	-69	-173	87	60
2	2872	**28.7**	20	101	172	93	-165	68	102	-166	75	105	-53	-69	175	117	-34	178	-87	-171	-94	173
3	250	**2.5**	63	79	176	87	39	66	99	-125	80	63	67	108	174	86	-66	162	-98	171	91	69
4	55	**0.6**	71	93	179	74	44	63	108	176	84	-98	67	-101	166	97	-75	161	-70	-173	87	53
5	31	**0.3**	65	75	165	81	-20	72	101	-136	76	66	64	66	178	91	16	156	-113	-130	86	63
6	111	**1.1**	65	82	177	84	-9	68	102	-179	71	63	63	88	174	91	-49	164	-84	-163	81	75
7	34	**0.3**	70	65	156	78	-3	68	96	-129	62	74	65	71	174	86	49	164	-99	159	88	62
8	264	**2.6**	68	80	178	83	21	65	103	-167	72	63	53	-90	161	104	-80	158	-75	-175	83	79
9	1834	**18.3**	95	91	178	90	57	62	107	-175	77	95	-45	-64	178	106	-80	162	-78	-179	80	72
10	962	**9.6**	109	97	168	106	76	65	107	-178	75	100	-47	-74	174	112	-62	171	-92	179	70	-53
11	127	**1.3**	69	94	178	73	-29	69	106	173	89	-101	63	-75	176	105	-84	158	-68	-172	86	54
Scheme II																						
12	3137	**62.7**	82	91	177	84	13	66	107	-179	76	95	-42	-63	178	107	-83	160	-75	-176	82	49
13	832	**16.6**	42	100	173	93	-174	68	102	-163	74	104	-53	-69	177	118	-36	178	-90	-163	-91	-112
14	627	**12.5**	-7	103	170	94	-155	66	103	-170	77	107	-54	-70	172	117	-32	177	-84	179	-91	59

Table 55. Similar clusters identified from the different clustering Schemes for *cp*PNA**16**

Cluster	No. of members	%	Scheme I Cluster #[a]	Cluster	No. of members	%	Scheme I, II Cluster #[a]
Scheme II				*Scheme III*			
A	307	**6.1**	7	A	607	**6.1**	3
B	97	**1.9**	11	B	6522	**65.2**	12
				C	2871	**28.7**	2
Total number of clusters			5				3

[a] numbers correspond to identical cluster from Table 54

The distribution of conformations is different for cpPNA**16** not only in comparison with the unmodified PNA but also in comparison with *cp*PNA**15**. The number of conformations sampled is larger for Scheme I in the simulation of the *cp*PNA**16** single strand than in the simulation of the modified *cp*PNA**15**, but *cp*PNA**16** has fewer clusters for Schemes II and III. A comparison of conformations between *cp*PNA**15** and *cp*PNA**16** shows that both adopt similar hairpin-like structures, but for *cp*PNA**16** coil conformations were also observed. *Cp*PNA**16** had fewer clusters than *aeg*PNA**14** for all three Schemes. Thus, this modification has also reduced the flexibility of the strand.

To assess the importance of the modification position, the residue mobility was examined by comparing the average RMSF values per residue (Table 56). The terminal residue T_1 has the highest fluctuations. Some other residues, including residues 3 and 9, are also flexible. On the whole, however, the lowest deviations are found in the region covering residues 4 to 8. This marked decrease in flexibility is most likely due to the specific location of the cyclic constraint at residue G_4. Differences between the RMSF values of the unmodified strand and *cp*PNA**16** are evident over the entire chain except for residues T_2, T_3, T_9 and T_{10}, and the overall flexibility of the strand is reduced. In *cp*PNA**15**, the modification also caused a decrease in the overall flexibility, but the decrease is larger for *cp*PNA**15**. Therefore, a strand containing a modification at T_6 should bind slightly tighter to DNA than a strand with a modification at G_4.

The RMSF values of most torsion angles are widely distributed with γ (G_1, A_8, C_9), δ (G_1, T_6, C_7) and ε (G_1, A_3, C_9) dihedrals having the highest values. In comparison to the reference structure *aeg*PNA**14**, the fluctuations observed for the β and ε dihedrals

are significantly reduced in the central residues located around the cyclopentane modification at residue G_4. Additionally, the δ dihedrals showed enhanced fluctuations not only for the residue of the N-terminus (G_1) but for the central region as well (T_6). Interestingly, enhanced fluctuations were observed for both the δ and the γ dihedrals compared to *cp*PNA**15**. Finally, both strands (*cp*PNA**15** and *cp*PNA**16**) share a common (reduced) effect on the fluctuations of several angles, namely β8, β9, γ4, γ7, ε6, ε7 and ε9. Both modifications cause a decrease in flexibility at the modification site but also cause reduced fluctuations at similar sites on the strand.

Table 56. Average dihedrals (°) and RMSF (bold, °) of the simulated *cp*PNA**16**

Residue	α	β	γ	δ	ε	Total RMSF	Average RMSF
G1	—	—	-170.1	150.7	-167.1		
	—	—	**156**	**88**	**106**	**350**	**117**
T2	-178.5	63.1	87.0	93.1	76.8		
	31	**52**	**12**	**15**	**25**	**135**	**27**
A3	-175.5	-27.0	86.2	89.2	65.5		
	22	**104**	**16**	**18**	**94**	**254**	**51**
G4	-161.3	66.7	67.7	105.4	62.7		
	26	**7**	**10**	**9**	**32**	**84**	**17**
A5	-173.1	56.6	75.7	110.7	92.7		
	29	**16**	**11**	**21**	**32**	**109**	**22**
T6	177.8	-36.8	119.8	-77.8	74.0		
	17	**35**	**15**	**52**	**36**	**155**	**31**
C7	-177.3	-80.5	109.1	-118.8	64.6		
	16	**39**	**12**	**40**	**37**	**144**	**29**
A8	-171.7	164.8	45.6	-80.1	60.9		
	25	**17**	**84**	**18**	**23**	**167**	**33**
C9	-174.2	112.6	32.2	91.6	33.9		
	22	**84**	**81**	**18**	**71**	**276**	**55**
T10	174.0	161.6	82.4	—	—		
	28	**41**	**40**	—	—	**109**	**36**

Table 56 contains the average dihedral angles as well. Let us consider the various angles in turn: the β angle has different values along the chain and undergoes transitions

at a number of residues when the modification is introduced. Both γ and δ remain in the 70-120° and ±90-100° regions, respectively, except for γ1, δ1, γ8, γ9 which are centered around 35° or 180°. For the ε angle, the range of 60-70°, which was observed in the unmodified PNA, is maintained except for residues 1 and 9. Consequently, the major differences between *aeg*PNA**14** and *cp*PNA**16** are in the β angles. Additionally, γ (residues 7 to 9) and δ (residues 6 to 8) angles show particularly large changes in their average values. In comparison to *cp*PNA**16**, differences were observed in the average values of β (residues 3, 5, 6 and 8) and γ (residues 1, 5, 6, 9 and 10) dihedrals.

There is a large jump in RMSD for the *cp*PNA**17**, which has a cyclopentane ring at C_7, at ~ 5.8 ns. This jump is due to a large structural distortion. The trajectory data was classified into nineteen clusters of similar members (Tables 57 and 58, Figure 54). First, the molecule persists in the potential energy basin corresponding to cluster **19** from Scheme 3 (1.3-3 ns). Then, a transition to another state occurs and the strand fluctuates between clusters **2** , **15** and **17** for approximately 4 ns, and the last 2.7 ns of the MD simulation are spent within cluster **3**, with a final jump to the state represented by cluster **4**. Clusters **1** and **16** turn out to be more or less short-lived transient states, with the second cluster corresponding (as usual) to the starting structure. Thus, the main low-energy states of *cp*PNA**17** differ from each other solely in orientation of the N-terminal residues.

The most populated cluster from Scheme I was also located with both Schemes II and III; however, in Scheme II the population percentage was much lower. Four clusters were identical between Schemes I and II and the remaining eleven clusters from Scheme I were replaced by only four new clusters in Scheme II. Two of the clusters from Scheme I and two clusters from Scheme II were located using the criteria of Scheme III, though the population percentage for clusters **19** and **3** were significantly higher when using Scheme III. The most populated pattern from the three schemes (**3, 19**) has an inner loop motif with an N-terminal bend and 4 little turns. The second most populated pattern from Scheme II (**18**) adopts a coiled conformation. The rest of the clusters are equally populated and exhibit an Ω-loop expanding from residues 4 to 9, and an N-terminal loop.

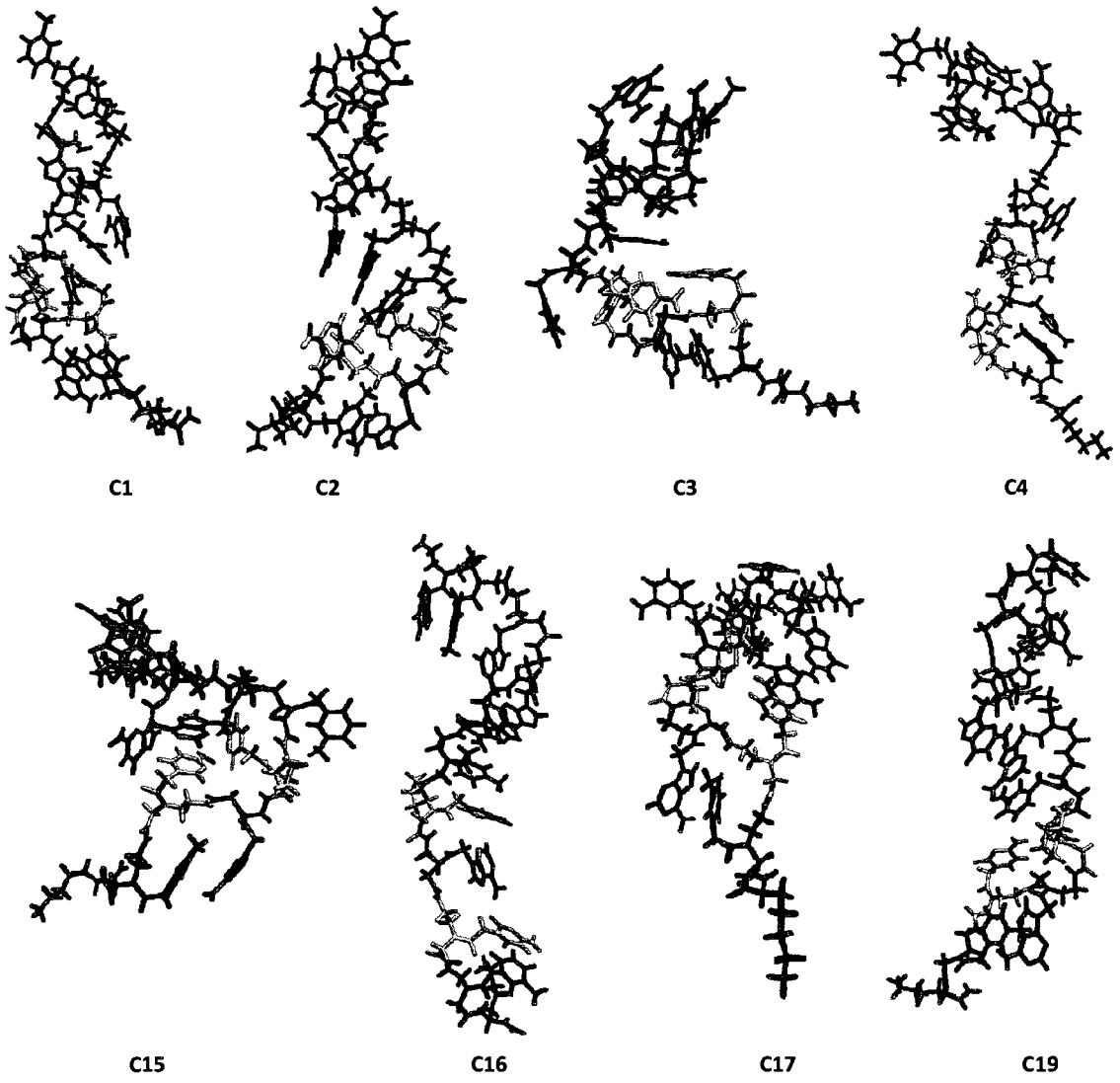

Figure 54. Cluster centroid structures obtained from simulations of *cp*PNA**17**. Residues are colored by name: **APN**-blue, **CPN**-orange, **TPN**-green, **CPC**- pink, **GPN**-red and **K**-purple. The structures are positioned with **GPN1** on top and **K11** on bottom (see Table 17 for the sequence)

Table 57. Average torsion angles for each cluster from the *cp*PNA**17** simulation

Cluster	No. of Members	%	T2		A3			G4		G5			A6		C7			A8		C9		
			β	δ	α	γ	ε	β	δ	α	γ	ε	β	δ	α	γ	ε	β	δ	α	γ	ε
Scheme I																						
1	1073	**10.7**	170	86	-143	92	74	179	79	172	80	58	63	86	-119	68	26	-59	82	173	83	-161
2	1064	**10.6**	171	84	-144	91	77	-179	81	177	77	66	63	85	-130	75	-5	-82	43	166	97	-156
3	2625	**26.3**	178	67	-146	100	109	-60	-95	175	95	32	-116	84	-109	63	15	-55	78	180	87	-157
4	1149	**11.5**	171	83	-135	104	-107	-141	-88	171	101	70	-172	108	-120	23	56	-55	79	173	87	-162
5	227	**2.3**	90	127	180	103	18	179	69	176	90	59	66	75	-147	79	12	-61	92	156	85	-101
6	305	**3.1**	91	125	-179	101	-14	-176	70	178	87	63	66	86	-106	77	4	-58	94	174	85	-103
7	217	**2.2**	93	125	177	104	10	-179	69	180	90	67	64	87	-109	77	13	-59	92	159	85	-62
8	333	**3.3**	103	121	178	87	53	-177	76	177	84	66	63	89	-104	69	0	-52	91	170	87	-153
9	208	**2.1**	92	125	179	100	-21	-175	70	179	87	64	65	77	-160	77	2	-58	91	178	83	-83
10	144	**1.4**	97	122	179	97	-13	-177	75	179	85	70	63	76	-151	73	5	-58	97	168	85	-149
11	103	**1.0**	-179	78	166	98	47	-180	70	167	85	61	66	76	-126	73	36	-59	80	172	81	-68
12	141	**1.4**	174	89	-122	102	0	-180	75	166	88	57	65	80	-131	74	8	-60	87	171	85	-117
13	107	**1.1**	175	71	150	99	-19	-173	70	172	81	59	71	79	-115	71	11	-55	79	176	80	-59
14	190	**1.9**	169	90	-152	90	46	174	101	179	86	45	-60	87	-105	78	6	-62	80	170	88	-159
15	1084	**10.8**	162	89	-153	90	67	174	108	177	91	-62	-55	84	-104	74	6	-54	91	178	86	-155
Scheme II																						
16	433	**8.7**	94	84	179	91	1	168	74	178	86	62	66	81	-129	70	27	26	96	177	84	66
17	11	**0.2**	178	82	-135	90	82	-178	78	174	92	67	67	73	-128	76	3	-48	94	-85	-72	-169
18	684	**13.7**	178	82	-142	101	108	-60	-92	170	89	-7	-63	79	-108	68	5	-55	78	179	85	-157
19	837	**16.7**	106	119	-179	100	4	-178	71	179	86	62	66	81	-131	75	8	-57	92	173	84	-98

Table 58. Similar clusters identified from the different clustering Schemes for *cp*PNA**17**

Cluster	No. of members	%	Scheme I Cluster #[a]	Cluster	No. of members	%	Scheme I, II Cluster #[a]
Scheme II				*Scheme III*			
A	1203	**24.1**	1	A	792	**7.9**	16
B	627	**12.5**	15	B	3763	**37.6**	3
C	639	**12.8**	3	C	1279	**12.8**	15
D	566	**11.3**	4	D	4166	**41.7**	19
Total number of clusters			8				4

[a] numbers correspond to identical cluster from Table 57

For Schemes II and III, *cp*PNA**17** had fewer clusters than *aeg*PNA**14**. For Schemes I and II, *cp*PNA**17** had five more clusters than *cp*PNA**15**. On the contrary, the number of clusters is increased when comparing *cp*PNA**17** to *cp*PNA**16**. This increase in total number of conformers shows that the strands with the modification at G_4 exhibit higher flexibility. Consequently, *cp*PNA**17** will form more stable duplexes with DNA than *aeg*PNA**14**, *cp*PNA**15**, but the binding affinity will be lower compared to *cp*PNA**16**. The position of the cyclopentane also imparts preferences for the backbone conformation, since Ω-coils were sampled instead of hairpin-like structures.

The RMSF data is given in Table 59. The largest values in the RMSF profile correspond to residues 1 and 4. Residues 7 to 10, (modification is at C_7) have the lowest RMSF values and are extremely stable whereas the $G_1T_2A_3G_4A_5T_6$ region shows significant flexibility. Residues 2 and 4-6 have much higher RMSF values in *cp*PNA**17** than in *cp*PNA**15**, *cp*PNA**16** and *cp*PNA**18** (see below), though the overall flexibility is still lower than in *aeg*PNA**14**. As in *cp*PNA**16**, the highest fluctuations were observed for the β and ε dihedrals, whereas γ and δ showed the lowest deviations. In general, the *cp*PNA**17** exhibits patterns of flexibility that are very similar to *cp*PNA**16**, although less pronounced. When compared to the unmodified PNA strand, reduced fluctuations are registered for the β and γ dihedrals of the C-terminal residues (7 to 10). The β, γ and ε dihedrals have higher RMSF values than in *cp*PNA**15** or *cp*PNA**18**, whereas the γ torsions have lower fluctuations in comparison to *cp*PNA**16** and *cp*PNA**18**. The specific

angles that were affected in *cp*PNA**15** and *cp*PNA**16** are β8, β9, γ4, γ7, γ8, ε7, ε9 and β8, β9, γ1, γ4, γ7, δ1, ε7, ε9, respectively.

Table 59. Average dihedrals (°) and RMSF (bold, °) of the simulated *cp*PNA**17**

	α	β	γ	δ	ε	Total RMSF	Average RMSF
G1	—	—	-29.3	-100.4	-59.9		
	—	—	**89**	**51**	**44**	**184**	**61**
T2	179.8	153.7	-88.2	86.4	-69.0		
	22	**39**	**52**	**28**	**55**	**196**	**39**
A3	-154.6	169.0	96.3	20.7	84.6		
	33	**35**	**14**	**85**	**77**	**244**	**49**
G4	177.4	-145.5	98.9	16.2	-126.9		
	24	**58**	**19**	**87**	**96**	**284**	**57**
A5	178.3	-32.0	89.0	84.6	41.5		
	23	**118**	**16**	**18**	**54**	**229**	**46**
T6	-176.1	-25.0	93.3	86.2	128.0		
	26	**99**	**14**	**17**	**107**	**263**	**53**
C7	-168.6	69.2	64.4	101.2	17.4		
	24	**8**	**20**	**13**	**34**	**99**	**20**
A8	173.9	-51.5	104.9	79.9	80.0		
	26	**34**	**15**	**23**	**35**	**133**	**27**
C9	-179.1	173.5	86.6	82.7	-128.1		
	23	**21**	**14**	**16**	**69**	**143**	**29**
T10	-177.9	165.0	86.5	—	—		
	20	**30**	**11**	—	—	**61**	**20**

Examination of the torsion angle data for *cp*PNA**17** shows that the β and ε dihedrals adopt various values in different residues. The torsion angles γ and δ remain almost exclusively in the region centered around 90° with an occasional tendency to have a negative value. The *aeg*PNA**14** and *cp*PNA**17** strands differ from one another in their respective preferences for the β and ε dihedrals. Changes in average values at the modification site are seen for most β, ε and even γ (A_8) torsions. The δ dihedral undergoes a transition only for the central A_3 and G_4 residues. When comparing to

*cp*PNA**15**, the major changes were observed for the β (residues 2 to 6 and 9) and ε (residues 4 to 10) dihedrals. Differences between *cp*PNA**17** and *cp*PNA**16** were recorded for the β (residues 2 to 5) and γ (residues 7 to 9) torsions.

The RMSD of the *cp*PNA**18**, which has a modification at residue A_8, shows initial fluctuations, but appears to stabilize at ~7 ns. Using conformational cluster analysis, eight distinct clusters were obtained (Figure 55, Tables 60 and 61). The most populated cluster from Scheme I was also located with Schemes II and III. Two clusters were found to be similar between the Schemes I and II and only one cluster from Scheme III matched that of Scheme I. The remaining cluster from Scheme I was substituted by three new clusters in Scheme II and two new clusters from Scheme III replaced two other clusters from Scheme I. Here again, none of the new clusters from Scheme II were identical with those of Scheme III.

One of the most abundant patterns (6.2-10 ns) found in Scheme I (cluster **3**) contains a wide bend with a helical motif running from residues 5 to 7. The matching clusters in Schemes II and III along with **4** and **8** also exhibit a similar structure. Cluster **6** from Scheme II and cluster **7** from Scheme 3 represent another abundant pattern that corresponds to a coil conformation with a helical element involving residues 4 to 7. Finally, clusters **2** (Scheme I) and **5** (Scheme II) group structures that have a hairpin loop extending from residues 1 to 6 together with a helical motif. At the beginning of the simulations, the molecule exists in cluster **1** which resembles the starting helical geometry (~1 ns). Then, it interchanges between the hairpin and coiled structures for 4.5 ns and finally falls into the cluster **8** conformation.

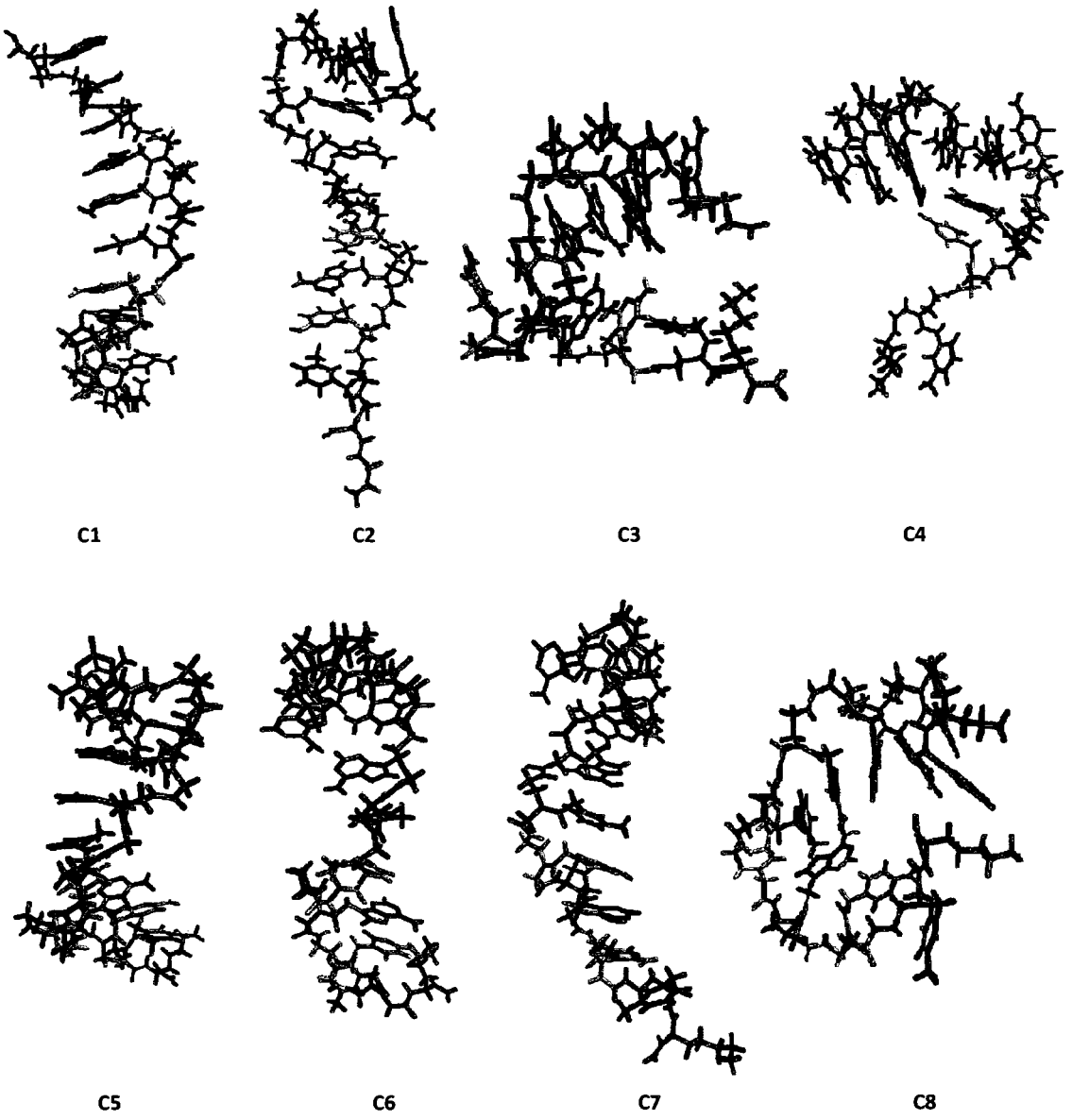

Figure 55. Representative structures of the most populated clusters from the MD simulation of *cp*PNA**18**. Residues are colored by name: **APN**-blue, **CPN**-orange, **TPN**-green, **APC**- light brown, **GPN**-red and **K**-purple. The structures are positioned with **GPN1** on top and **K11** on bottom (see Table 17 for the sequence)

Table 60. Average torsion angles for each cluster from the *cp*PNA**18** simulation

Cluster	No. of Members	%	T2		A3			G4		G5			A6		C7			A8		C9		
			β	δ	α	γ	ε	β	δ	α	γ	ε	β	δ	α	γ	ε	β	δ	α	γ	ε
Scheme I																						
1	3134	**31.3**	67	90	-174	80	22	99	81	180	83	65	63	79	177	87	12	68	101	-162	80	56
2	2504	**25.0**	66	86	-173	73	-67	170	90	176	87	13	100	87	177	93	41	67	102	-174	79	-83
3	4362	**43.6**	59	91	-174	72	-41	176	97	178	88	-65	160	100	172	101	73	68	102	-172	80	73
Scheme II																						
4	663	**13.3**	62	85	-175	71	-62	175	98	177	90	-70	159	100	175	103	76	67	101	-176	78	-91
5	715	**14.3**	68	88	-172	74	-68	164	82	176	83	65	64	79	178	87	18	67	102	-172	79	-74
6	594	**11.9**	69	89	-172	71	-63	172	94	178	84	60	64	78	177	88	14	69	101	-169	81	77
Scheme III																						
7	4649	**46.5**	67	89	-173	78	-8	121	82	179	83	64	63	79	178	87	13	68	101	-166	80	20
8	910	**9.1**	47	93	176	74	40	-171	84	161	86	-65	165	96	174	106	72	66	102	-174	82	52

Table 61. Similar clusters identified from the different clustering Schemes for *cp*PNA**18**

Cluster	No. of members	%	Scheme I Cluster #[a]
	Scheme II		
A	1041	**20.8**	1
B	1984	**39.7**	3
	Scheme III		
A	4441	**44.4**	3
Total number of clusters:		Schemes II – **5**;	Scheme III - **3**

[a] numbers correspond to identical cluster from Table 60

For Schemes II and III, *cp*PNA**18** had fewer clusters than *aeg*PNA**14**, *cp*PNA**15** (T_6) and *cp*PNA**17** (C_7), but a similar number of clusters to *cp*PNA**16** (G_4). On the other hand, for Scheme I, *cp*PNA**18** and *cp*PNA**15** had the same number of clusters, but the number of clusters for *cp*PNA**18** was still lower than *aeg*PNA**14**, *cp*PNA**16** and *cp*PNA**17**. These results all suggest that the addition of the cyclopentane ring at position T_8 significantly decreased the flexibility of the strand when compared to the unmodified PNA and systems with single modifications at T_6 and C_7. However, it exhibits a similar level of flexibility as *cp*PNA with the cyclopentane at G_4. The highly populated clusters of *cp*PNA**18** differ from the unmodified and other singly modified *cp*PNAs in the backbone conformations and structural motifs. The clusters of *aeg*PNA**14** (G-type), *cp*PNA**15**-*cp*PNA**16** (hairpin), *cp*PNA**17** (Ω-loop) exhibit different types of coils and folded conformations, while the modification at T_8 leads to structures that assume helical conformations and maintain the helical element during the entire simulation time.

Examining the RMSF profile for the *cp*PNA**18** simulation, we see that the terminal residue torsions have relatively large flexibilities in comparison to the rest of the chain (Table 62). Stacking interactions are observed for all residues except T_{10}. The residues G_1 and T_{10} have increased flexibility compared to the flexibility of these residues in the *aeg*PNA**14** simulation, while the flexibility of residues T_6 to C_9 is considerably reduced. A slight increase in flexibility of residue A_3 is also observed. The RMSF values of the three systems (*cp*PNA**16-18**) showed similar patterns for residues 7 to 9. The

fluctuations of modified residues were more reduced in *cp*PNA**16** (G_4) and *cp*PNA**18** (A_8) than in *cp*PNA**15** (T_6) or *cp*PNA**17** (C_7). Addition of the ring at T_6 had distinct local effects and small effects on the residues, very distant from the modification site. In the case of *cp*PNA**16** and *cp*PNA**17**, decreased RMSF values were seen for all residues located downstream (toward the C-terminus) from the altered residue. On the other hand, in *cp*PNA**18** most residues showed reduced fluctuations, although residues closer to the site of the modification had the largest changes in RMSFs. The biggest effect on the overall chain flexibility was observed for *cp*PNA**15** and *cp*PNA**18**; thus, the modifications at T_6 or A_8 significantly impaired the mobility of the strand. This finding partially agrees with the clustering results, where *cp*PNA**18** had the lowest conformational flexibility. Finally, the *cp*PNA**16** and *cp*PNA**17** strands have comparable flexibility.

Table 62. Average dihedrals (°) and RMSF (bold, °) of the simulated *cp*PNA**18**

Residue	α	β	γ	δ	ε	Total RMSF	Average RMSF
G1	—	—	-153.2	81.9	50.1		
	—	—	**155**	**18**	**35**	**208**	**69**
T2	-177.3	63.0	84.6	89.8	61.0		
	23	**22**	**12**	**17**	**43**	**117**	**23**
A3	-173.7	63.6	74.4	-135.5	-27.9		
	20	**31**	**13**	**65**	**60**	**189**	**38**
G4	178.7	150.5	64.2	90.2	13.4		
	21	**48**	**56**	**25**	**65**	**215**	**43**
A5	178.1	79.3	86.2	90.2	-5.1		
	32	**36**	**13**	**22**	**69**	**172**	**34**
T6	-175.8	114.5	85.8	90.2	74.7		
	25	**50**	**12**	**20**	**35**	**142**	**28**
C7	174.5	11.7	94.5	93.0	46.0		
	21	**60**	**14**	**17**	**46**	**158**	**32**
A8	-109.0	67.7	74.1	101.5	41.6		
	20	**9**	**9**	**9**	**30**	**77**	**15**
C9	-169.2	65.8	79.9	93.4	21.1		
	29	**12**	**10**	**18**	**83**	**152**	**30**
T10	170.1	162.4	-28.6	—	—		
	27	**53**	**85**	—	—	**165**	**55**

When compared to the reference *aeg*PNA strand, the RMSF values of β, γ and ε indicate a reduction in flexibility while fluctuations observed for δ are quite similar with the exception of residue 3 where an increase in fluctuation is recorded. As observed in *cp*PNA**16** and *cp*PNA**17** simulations, the γ torsion has the lowest RMSF values whereas in *cp*PNA**15** it is the highest. Furthermore, in all mixed base sequences examined (*aeg*PNA**14** and *cp*PNA**15-18**) the δ torsion shows the lowest fluctuations. Similar to *cp*PNA**17** but unlike *cp*PNA**16**, the highest fluctuations were seen for the ε dihedral. As for the β dihedral, it demonstrates higher fluctuations in all systems except for *cp*PNA**15** where it has the lowest RMSFs. Lastly, the dihedrals that were identified as the most influenced by the constraint, regardless of its position, are β8, β9, γ4, γ7, ε7 and ε9.

Table 62 also contains the average dihedral angles of the decamer backbone. The average β values remain predominantly in the 60-80° range for most of the nucleotides, though a few *trans* (~160°) values are also observed (residues 4, 6 and 10). Most residues have average values in the 60-90° range for γ. The average δ torsion stays in its original range of 80-100°. Interestingly, the ε dihedral assumes a range of values from -30° to 75°. This observation is in contrast to the corresponding results on the unmodified oligomer where values in the 50-70° range were seen. Along with β and ε, the γ dihedral showed differences in average values compared with the regular *aeg*PNA strand. The preferred dihedral values for β and δ differ between the singly modified *cp*PNA strands as well.

The following conclusions emerge on the dynamics of the 10-mer *cp*PNAs bearing cyclopentane at various positions. First, the conformational space sampled shows a sizable reduction in the presence of the cyclopentane. The lowest number of clusters was obtained for *cp*PNA**18** and the highest for *cp*PNA**17** and *aeg*PNA**14**. Second, the conformational sampling of the *cp*PNA**18** is sizably reduced when quantified by analyzing backbone flexibility and the distributions of the dihedral spaces accessed by the residues. Unlike *cp*PNA**16** and *cp*PNA**17**, but similar to *cp*PNA**15**, the effect on strand flexibility was not local and covered more residues located far from the modification site. The tendency of the strands to adopt a helical structure was seen only for *cp*PNA**18**. For *cp*PNA**15** and *cp*PNA**16**, the most prevalent structures observed were hairpin

conformations. In all *cp*PNAs, the average dihedrals in the modified residues were limited to certain ranges of values: α=-170°, β=60-70°, γ=70-85°, δ=90-100°, ε=45-60°. These effects are determinants of PNA binding thermodynamics and should result in significantly different conformational enthalpies for the modified strands. Owing to the flexibility, the stability of the *cp*PNA:DNA complexes is strongly influenced by entropic contributions. More specifically, the alterations at T_6 or A_8 restricted the flexibility of PNA more than those at C_7 or G_4 and, as a consequence, these strands should have a lower entropy penalty when binding to DNA. A comparison of results between polythymine and mixed base sequences revealed that in both systems, the position of the alteration changes the flexibility of the residue considerably. However, the effect of the modification position was more pronounced in poly-T_8 sequences, since the reduced fluctuations were not localized to the site of the modification but also included the adjacent residues. This higher restriction of the *cp*PNA-T_8 strands will result in higher binding affinity to target DNA/RNA.

THE EFFECT OF MULTIPLE MODIFICATIONS

PNA-T_8 sequence: *cp*PNA**7**-*cp*PNA**11**

In Figure 2, the RMSD between the trajectory structures of *cp*PNA**7**, which has the modification at residues 1, 4 and 8, and the initial structure is shown as a function of time. For the initial 2 ns, the RMSD was at 4 Å, and then continually increased to 7 Å. After 4 ns, the RMSD values ranged from 5.2 to 6.8 Å. Starting from a helical initial structure, the helix is lost after 2.5 ns and is not formed again during the remaining 7.5 ns. The conformational space that is sampled in the MD simulation was analyzed using conformational cluster analysis and sixteen clusters were obtained (Tables 63 and 64). In Figure 56, the central member structure of each of the three most populated clusters is shown.

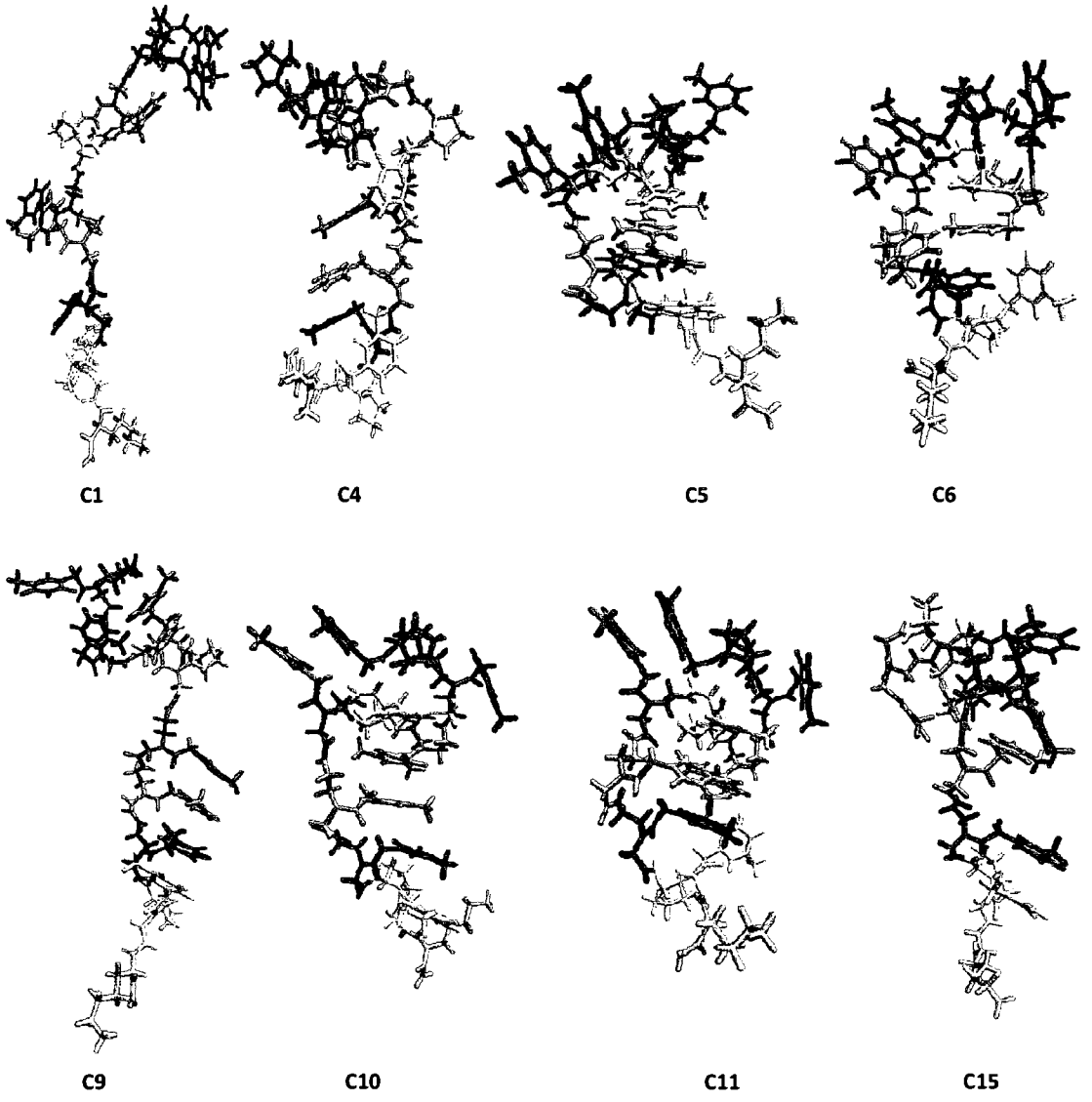

Figure 56. Representative structures of the most populated clusters from the MD simulation of *cp*PNA7. Residues are colored by position: **T1**-red, **T2**- dark grey, **T3**-orange, **T4**-yellow, **T5**-dark yellow, **T6**-grey, **T7**-green, **T8**-white, and **K9**-pink. The structures are positioned with **T1** on top and **K9** on bottom

Table 63. Average torsion angles for each cluster from the *cp*PNA7 simulation

Cluster	No. of members	%	T2		T3			T4		T5			T6		T7		
			β	δ	α	γ	ε	β	δ	α	γ	ε	β	δ	α	γ	ε
Scheme I																	
1	1562	**15.6**	148	88	-165	99	-130	69	107	-179	95	65	64	86	-179	84	38
2	694	**6.9**	178	75	171	-87	-49	67	107	165	92	63	65	76	178	87	24
3	779	**7.8**	176	78	157	-89	97	65	104	-178	96	66	63	75	178	83	57
4	858	**8.6**	177	86	178	95	-125	66	108	-180	90	65	60	73	166	85	-102
5	3106	**31.3**	-178	-101	-168	109	-79	117	97	176	93	62	138	92	-177	94	90
6	2716	**27.2**	-177	-96	-166	108	-80	137	92	177	97	137	180	83	176	84	-40
7	285	**2.9**	174	72	143	-95	-158	66	112	-160	92	69	58	93	175	85	-149
Scheme II																	
8	223	**4.5**	66	94	-167	91	-177	68	106	-174	97	63	65	87	-178	87	22
9	397	**7.9**	-176	80	-167	104	-81	68	108	-178	96	65	63	84	177	81	52
10	664	**13.3**	-176	-97	-162	108	-82	132	96	-171	99	-132	175	79	174	85	7
11	837	**16.7**	-179	-95	-170	109	-80	140	89	166	96	74	-178	87	178	85	-79
12	180	**3.6**	173	98	-157	96	-178	69	106	175	92	64	63	91	-174	85	23
Scheme III																	
13	731	**7.3**	175	75	151	-92	147	66	108	-168	94	68	60	80	175	82	131
14	443	**4.4**	173	-171	-150	110	-78	69	107	-174	99	65	67	78	179	82	87
15	2157	**21.6**	-174	-101	-164	108	-82	113	101	-171	94	-35	131	84	176	83	31
16	265	**2.7**	179	41	-171	107	-78	70	110	-163	99	70	57	96	167	79	63

Table 64. Similar clusters identified from the different clustering Schemes for *cp*PNA7

Cluster	No. of members	%	Scheme I Cluster #[a]	Cluster	No. of members	%	Scheme I, II Cluster #[a]
Scheme II				*Scheme III*			
A	340	**6.8**	2	A	734	**7.3**	8
B	393	**7.9**	3	B	1022	**10.2**	2
C	145	**2.9**	7	C	3380	**33.8**	5
D	412	**8.2**	4	D	774	**7.7**	4
E	1409	**28.2**	5	E	494	**4.9**	9
Total number of clusters			10				9

[a] numbers correspond to identical cluster from Table 63

The predominant structures identified in all three Schemes have similar dihedral angle values. Between Schemes I and II, five clusters had similar dihedrals. The other two clusters from Scheme I were replaced by five different clusters in Scheme II. Three of the clusters from Scheme I were also located with the Scheme III criteria. Four new clusters were identified in Scheme III, while two of the clusters from Scheme III matched those of Scheme II. Many of the most populated conformers from Schemes I and II (**5**, **6**, **10** and **11**) have a coil structure with an N-terminal loop. Cluster **15** is another highly populated cluster located with Scheme III and contains extended structures with an N-terminal loop. However, the lowly populated structures are of irregular character, showing a variety of structural motifs. They have no helical character.

For *cp*PNA**7**, the number of clusters obtained is higher than for *cp*PNA**6** for Schemes II and III but both have a similar number of clusters according to Scheme I. In comparison to the unmodified *aeg*PNA**3** and singly modified *cp*PNA**4** (T_4) and *cp*PNA**5** (T_1), the increased number of cyclopentane rings in *cp*PNA**7** results in reduction of the clusters over all Schemes. Thus, the flexibility of the *cp*PNA**7** oligonucleotide is higher or approximately the same as for *cp*PNA**6**, but is significantly lower than for *aeg*PNA**3**, *cp*PNA**4** or *cp*PNA**5**. The comparison between cluster conformations reveals that modifications at the C- and N-terminal residues along with the T_4, affects the backbone conformations, since instead of hairpin-like or G-type folded structures, more extended coiled structures are sampled. Consequently, the *cp*PNA**7**:DNA complex stability will be

larger than the PNA:DNA duplex stability of the unmodified or *cp*PNA**4** and *cp*PNA**5** strands.

Table 65. Average dihedrals (°) and RMSF (bold, °) of the simulated *cp*PNA**7**

Residue	α	β	γ	δ	ε	Total RMSF	Average RMSF
T1	—	—	-87.6	93.7	25.0		
	—	—	**19**	**14**	**37**	**70**	**23**
T2	167.2	175.4	-25.5	-164.5	-131.9		
	40	**26**	**92**	**94**	**92**	**344**	**69**
T3	-173.9	-116.3	70.8	-100.5	-174.8		
	26	**57**	**75**	**15**	**156**	**329**	**66**
T4	-108.9	101.7	61.6	99.7	53.3		
	27	**36**	**27**	**13**	**57**	**160**	**32**
T5	177.4	-179.8	94.5	77.3	40.7		
	26	**14**	**14**	**16**	**72**	**142**	**28**
T6	179.4	117.8	23.2	84.7	65.9		
	22	**60**	**87**	**18**	**31**	**218**	**44**
T7	178.5	37.8	87.3	92.3	-10.1		
	22	**51**	**14**	**20**	**145**	**252**	**50**
T8	-118.4	68.3	59.1	—	—		
	26	**8**	**24**	—	—	**58**	**19**

By analyzing the RMSF profile of *cp*PNA**7**, we can see that the terminal residues 1 and 8, as well as the central residue 5 have much lower RMSF values than in the unmodified strand, indicating that these regions are constrained by the modifications (Table 65). Residue 2 showed an increased RMSF in comparison to *aeg*PNA**3**,which indicates increased flexibility. Residue 3 also showed greater flexibility than the other residues in the *cp*PNA**7** strand; however, it remains similar to the unmodified strand. In *cp*PNA**7**, residues 5 and 8 are more rigid than in *cp*PNA**4** and *cp*PNA**5**. In comparison to *cp*PNA**5** and *cp*PNA**6**, higher RMSF values were seen for residues 2, 6 and 7. Based on RMSF data, the *cp*PNA**7** and *cp*PNA**5** strands had lower flexibility compared to *cp*PNA**4** but were more flexible than *cp*PNA**6**. This is in line with the observations from the clustering analysis. The addition of multiple rings diminished the flexibility of *cp*PNA**7**

and decreased the conformational entropy of the strand. Thus, *cp*PNA**7** will favor association to DNA/RNA more than *aeg*PNA**3**, *cp*PNA**4** and *cp*PNA**5**, but less than *cp*PNA**6**.

The greatest deviations are found for the ε dihedral. The β torsions show moderate fluctuations that are slightly greater than the RMSFs of the γ and δ angles. The analysis of the average values shows that the β and ε dihedrals are widely distributed. Consequently, significant changes in the average values are observed for the β and ε dihedrals (residues 2, 3, 5, 6, 7) when compared to the unmodified and most cyclopentyl PNAs. In contrast, as in all multiple cyclopentane systems, flexibility of the γ and δ angles is affected by the addition of the cyclopentane since values in the 60-90° and 80-100° range are observed for these torsions, respectively. Exceptions include the γ2 and δ3 dihedrals that show different averages and upon comparison to other sequences show similarity only with the unmodified *aeg*PNA**3** strand.

When compared to *aeg*PNA**3**, fluctuations of the β and δ dihedrals were reduced the most. Moreover, in comparison to all molecules carrying modifications, the β torsion (central residues) is more rigid than in *cp*PNA**4-6**. The δ dihedral showed a similar level of flexibility in all strands. Similarities between the RMSF values of *cp*PNA**7** and the strands carrying a single cyclopentane are evident for certain dihedrals: β3, γ8 and δ4.

The RMSD of the *cp*PNA**8** sequence with the modifications at residues 1 and 8, shows initial fluctuation, but appears to stabilize at ~2 ns (Figure 36). The initial helix is lost at 0.9 ns and no helix is formed within the remaining 9 ns. Cluster analysis of the simulation data resulted in the generation of fourteen different clusters (Figure 47, Tables 66 and 67). The majority of the structures exhibit a hairpin motif, suggesting that this conformation is highly stable. All three highly populated clusters identified in Scheme I were also located with Schemes II and III. Six clusters were found to be identical in Schemes I and II and the remaining cluster from Scheme I was replaced by six new clusters in Scheme II. Also, five of the clusters from Scheme I were similar to those of Scheme III and only one of the new clusters from Scheme II was located with Scheme III. The remaining two clusters of Scheme I were substituted in Scheme III by a single

new cluster with a high number of members and by the aforementioned cluster identical to **12** from Scheme II.

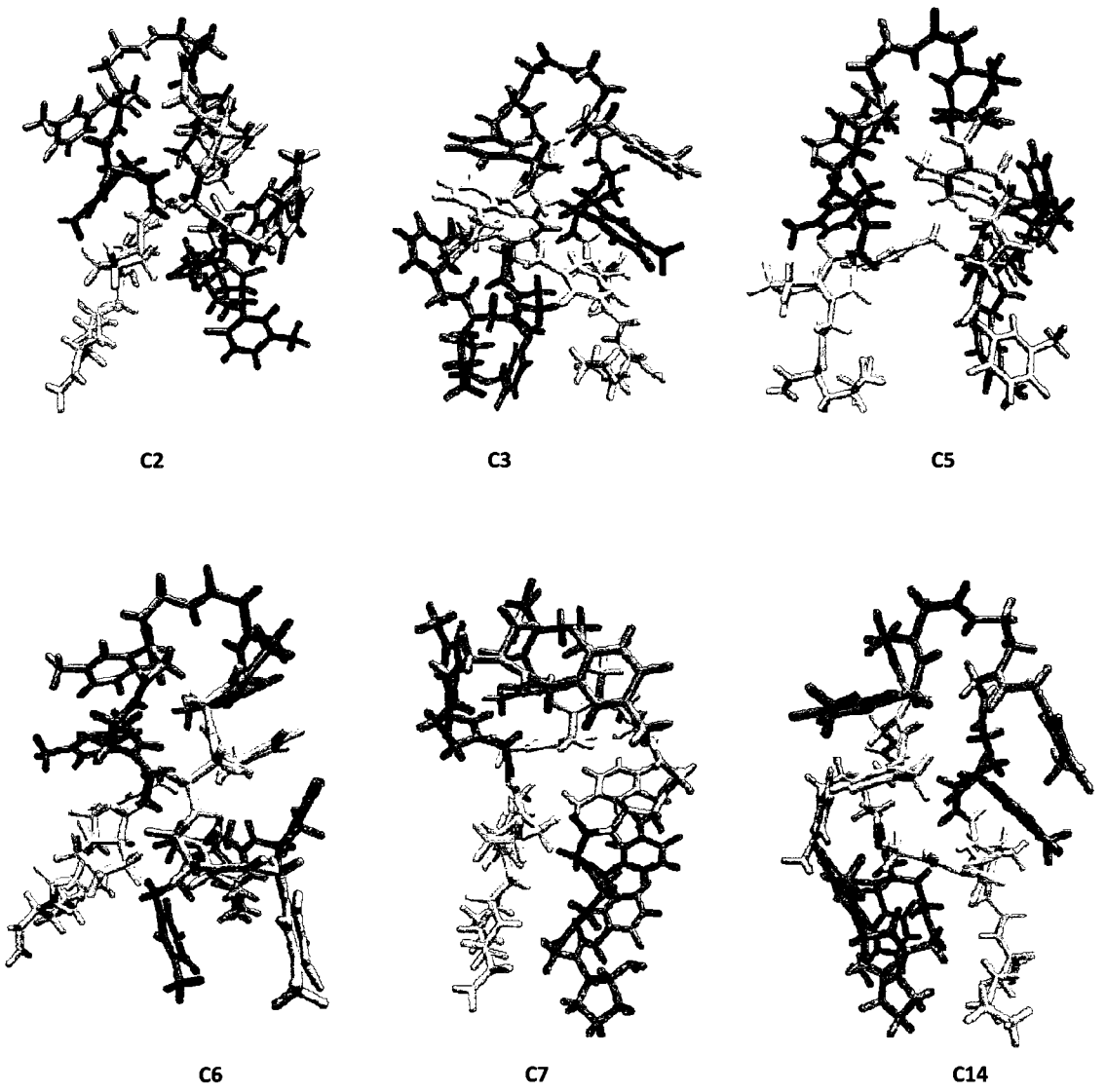

Figure 57. Representative structures of the most populated clusters from the MD simulation of *cp*PNA**8**. Residues are colored by position: **T1**-red, **T2**- dark grey, **T3**-orange, **T4**-yellow, **T5**-dark yellow, **T6**-grey, **T7**-green, **T8**-white, and **K9**-pink. The structures are positioned with **T1** on top and **K9** on bottom

Table 66. Average torsion angles for each cluster from the *cp*PNA**8** simulation

Cluster	No. of members	%	T2		T3			T4		T5			T6		T7		
			β	δ	α	γ	ε	β	δ	α	γ	ε	β	δ	α	γ	ε
Scheme I																	
1	305	**3.1**	19	92	175	96	-60	-180	78	-178	84	65	52	81	-174	100	88
2	2829	**28.3**	-59	93	-179	94	92	80	84	175	82	66	-70	84	169	82	-108
3	2270	**22.7**	-60	79	180	94	-47	168	89	171	87	64	-102	84	169	79	-93
4	639	**6.4**	-58	101	179	97	98	84	79	178	85	70	-111	80	172	81	109
5	705	**7.1**	-60	101	-159	98	105	61	68	170	87	-62	29	84	168	83	12
6	898	**9.0**	-61	78	-179	91	-47	164	99	168	85	-59	-11	89	169	77	-78
7	2354	**23.5**	70	108	-171	93	-60	162	92	169	75	-49	179	67	160	103	-102
Scheme II																	
8	166	**3.3**	-57	101	172	99	104	62	86	178	84	102	-173	83	174	82	148
9	189	**3.8**	-61	105	-173	96	103	67	67	175	89	0	-64	80	164	84	67
10	62	**1.2**	-60	80	174	85	-37	173	98	177	86	26	-63	84	164	90	69
11	52	**1.0**	62	115	-156	102	69	68	76	172	77	-51	-179	58	151	110	-98
12	269	**5.4**	-59	87	172	93	67	177	85	-180	83	77	-63	77	170	84	-152
13	106	**2.1**	-61	80	174	94	89	-179	85	-177	89	73	-174	80	169	83	-96
Scheme III																	
14	2661	**26.6**	-61	77	180	91	-33	148	93	168	87	13	-50	90	168	77	-83

Table 67. Similar clusters identified from the different clustering Schemes for *cp*PNA**8**

Cluster	No. of members	%	Scheme I Cluster #[a]	Cluster	No. of members	%	Scheme I, II Cluster #[a]
Scheme II				*Scheme III*			
A	134	**2.7**	1	A	267	**2.7**	1
B	1165	**23.3**	2	B	2186	**21.9**	2
C	1015	**20.3**	3	C	999	**10.0**	5
D	278	**5.6**	5	D	2365	**23.7**	7
E	439	**8.8**	6	E	837	**8.4**	3
F	1125	**22.5**	7	F	685	**6.9**	12
Total number of clusters			12				7

[a] numbers correspond to identical cluster from Table 66

The clusters obtained from all Schemes for *cp*PNA**8** were compared with the clustering results of the unmodified *aeg*PNA**3** and cyclopentyl PNAs having single modifications at similar locations: *cp*PNA**5** (T_1) and *cp*PNA**6** (T_8). For *cp*PNA**8** and *cp*PNA**6**, a similar number of clusters was obtained by Scheme I, while Scheme II and III identified fewer clusters for *cp*PNA**6** than for *cp*PNA**8**. This result indicates that the addition of a single cyclopentane ring at the C-terminus decreased the flexibility of the strand more than two modifications at both the N- and C-terminals. However, *cp*PNA**8** has lower flexibility than *aeg*PNA**3** and *cp*PNA**5**, since clustering resulted in a higher number of conformations for these strands than for the doubly modified strand. The conformations of *cp*PNA**5** and *cp*PNA**6** resemble hairpin-like structures. Consequently, modifications at both locations, although not altering the overall structure of the PNA, affect the thermodynamic stability of the molecule. The *cp*PNA**8** strand with T_1 and T_8 modifications will show much enhanced duplex stability compared to the non-modified PNA and *cp*PNA**5**; however, the strand with only one modification at T_8 (*cp*PNA**6**) will exhibit better DNA binding properties.

The RMSFs of *cp*PNA**8** and *aeg*PNA**3** were compared in order to assess the effect of multiple modifications on the strand flexibility (Table 68). The most evident differences are found at the modification sites (at position 1 and 8) and at central residue 4. Both the terminal residues and the T_4 are no longer flexible, and thus, have low

deviation values. Further differences, directly ascribable to the modification sites, are noted for residue 6, for which the RMSF value is higher compared to the unmodified one.

Table 68. Average dihedrals (°) and RMSF (bold, °) of the simulated *cp*PNA**8**

Residue	α	β	γ	δ	ε	Total RMSF	Average RMSF
T1	—	—	-89.6	87.2	43.5		
	—	—	**16**	**11**	**33**	**60**	**20**
T2	-177.2	-26.5	99.4	93.1	4.6		
	21	**60**	**17**	**23**	**65**	**186**	**37**
T3	-176.2	166.7	94.2	-74.8	-26.2		
	24	**40**	**13**	**58**	**127**	**262**	**52**
T4	179.2	129.0	86.1	86.5	68.1		
	25	**53**	**13**	**17**	**23**	**131**	**26**
T5	172.1	62.9	82.2	-177.1	18.5		
	20	**15**	**12**	**77**	**64**	**188**	**38**
T6	-177.3	-89.9	96.8	80.1	153.9		
	30	**77**	**15**	**17**	**135**	**274**	**55**
T7	167.7	34.4	86.3	92.7	40.1		
	20	**50**	**16**	**22**	**156**	**264**	**53**
T8	-113.4	101.7	68.6	—	—		
	25	**37**	**32**	—	—	**94**	**31**

Modification at the end of the strand caused a decrease in the RMSF for that residue and the neighboring residue regardless of whether the modification was at the N-terminus or C-terminus. Reductions in RMSFs were seen for other residues in the strand with the addition of the modification, but the location of these residues was dependent on the location of the modification. When the modification is added to both the N-terminus and C-terminus, the RMSFs of the modified residues were comparable between *cp*PNA**8** and *cp*PNA**6** or *cp*PNA**5**. However, the RMSF of residue 2 in *cp*PNA**8** was comparable to that of *cp*PNA**5**, but the reduction in RMSF was much smaller for *cp*PNA**8** than for *cp*PNA**6**. For the other residues, the RMSFs for residues 4 and 5 can be interpreted as additive effects from *cp*PNA**5** and *cp*PNA**6**, but not for residues 6 and 7. For residue 4, both *cp*PNA**5** and *cp*PNA**6** had low RMSFs and this reduction could have caused the even further reduction in RMSF for this residue in *cp*PNA**8**. For residue 5, the RMSF in *cp*PNA**8** is in between the RMSF of *cp*PNA**5** and *cp*PNA**6**. *Cp*PNA**8** has higher RMSFs

for residues 6 and 7 than both *cp*PNA**5** and *cp*PNA**6**. Thus, while the modification had similar effects locally on the RMSF in the singly- and doubly-modified strands, the effects on the rest of the strand could not be predicted for *cp*PNA**8** from *cp*PNA**6** and *cp*PNA**5**.

RMSF results indicate that the flexibility of *cp*PNA**8** is lower compared to *aeg*PNA**3**, *cp*PNA**5** or *cp*PNA**6**, which only partially correlates with clustering data, where *cp*PNA**5** was found to be more flexible than *cp*PNA**8**. These results suggest that restrictions in the flexibility of *cp*PNA**8** (compared to *cp*PNA**6** and possibly *cp*PNA**5**), caused by the introduction of a second cyclopentane modification, will result in larger contributions to the DNA/RNA binding free energy.

A more detailed picture of the similarities and differences was obtained from the direct comparison of the dihedral RMSFs between the *cp*PNA**8** and all other strands. The β, γ and δ torsions exhibit lower variability as compared to *aeg*PNA**3** and *cp*PNA**4**. In residues 1 to 5, the γ dihedrals have lower RMSF values than in *cp*PNA**5**, **6** and **7**. Significant changes in the ε dihedral are only seen for the T_4-T_7 residues, for which increased fluctuations are observed compared to *cp*PNA**5** and *cp*PNA**7**. The strongest similarities between *cp*PNA**8** and singly modified strands are observed only for three (β7, β2 and δ4) dihedrals.

Analysis of the *cp*PNA**8** structure shows that the β and ε dihedrals spanned a significant portion of torsional space. The ranges for the γ and δ torsions were 70-100° and 80-95°, respectively. The major differences between the *cp*PNA**8** and *aeg*PNA**3** angles are seen for the β, δ and ε dihedrals. However, compared to other singly modified cyclopentyl PNAs, most changes in values are observed for the β and ε dihedrals only (i.e. the modification affects the values of the β and ε dihedrals).

The RMSD of the *cp*PNA**9**, which has modifications at T_4 and T_8, shows a great amount of fluctuation during the simulation, with the most fluctuation at 5-7 ns. The initial helix was only maintained for 1.1 ns and was not formed again during the remaining 8.9 ns. Cluster selection criteria resulted in twelve distinct clusters. Relevant information on the clusters is given in Figure 58 and Tables 69 and 70. Five clusters from Scheme I were identified in Schemes II and III and one cluster from Scheme II was

located with Scheme III. The remaining three clusters from Scheme I were substituted with four new clusters in Scheme II. New patterns were not located with the Scheme III parameters.

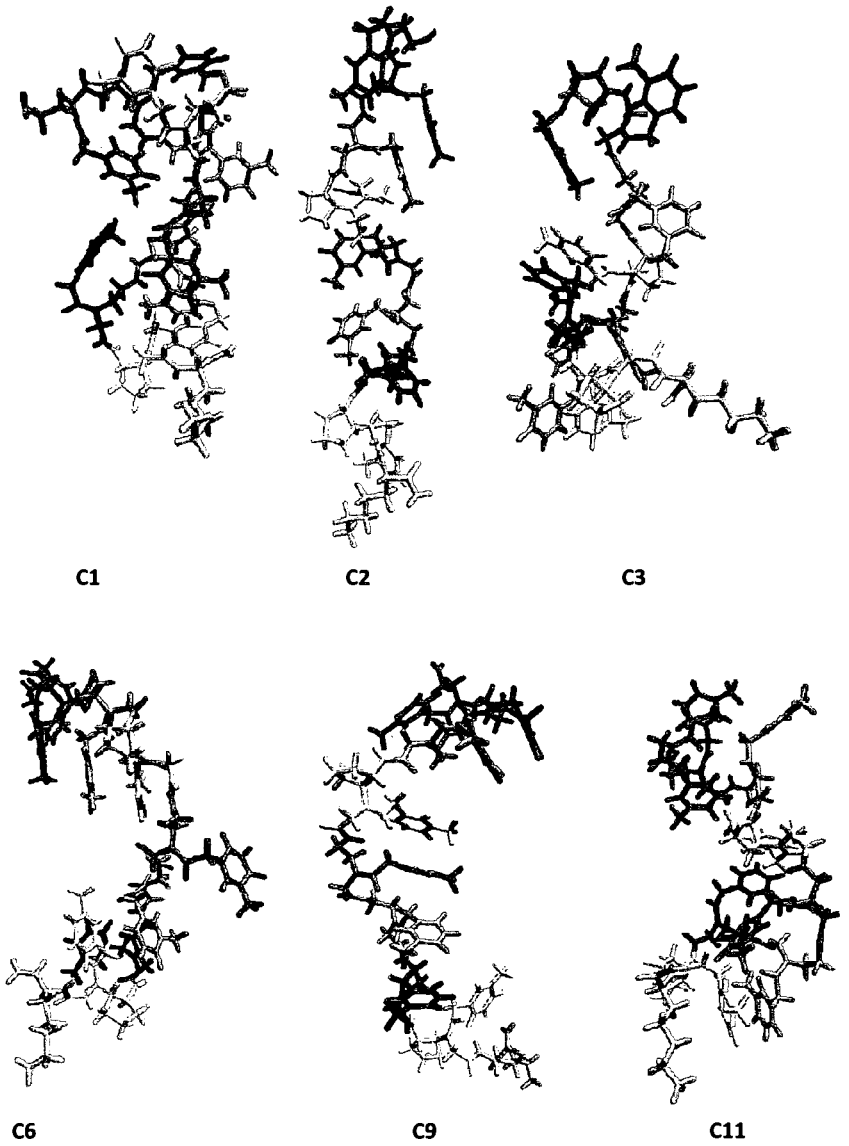

Figure 58. Representative structures of the most populated clusters from the MD simulation of *cp*PNA**9**. Residues are colored by position: **T1**-red, **T2**- dark grey, **T3**-orange, **T4**-yellow, **T5**-dark yellow, **T6**-grey, **T7**-green, **T8**-white, and **K9**-pink. The structures are positioned with **T1** on top and **K9** on bottom

Table 69. Average torsion angles for each cluster from the *cp*PNA**9** simulation

Cluster	No. of members	%	T2		T3			T4		T5			T6		T7		
			β	δ	α	γ	ε	β	δ	α	γ	ε	β	δ	α	γ	ε
Scheme I																	
1	3612	**36.1**	-81	79	176	97	-131	63	99	175	99	85	-65	98	-170	86	-140
2	1183	**11.8**	-63	71	168	81	-100	68	107	-176	-94	64	66	95	179	96	-135
3	3054	**30.5**	-33	94	-179	84	72	68	107	176	85	85	-66	101	-177	87	-146
4	499	**5.0**	-66	77	175	83	-162	90	105	-176	90	63	67	80	-178	85	-134
5	336	**3.4**	-179	77	169	81	-158	149	93	-156	77	71	56	75	170	83	-125
6	703	**7.0**	-60	79	180	88	23	68	105	-171	99	71	66	91	178	99	-126
7	540	**5.4**	-67	75	172	82	-163	68	102	-157	-97	49	62	94	-176	89	-123
8	73	**0.7**	-54	75	178	86	41	67	105	175	-90	70	65	92	160	91	-109
Scheme II																	
9	430	**8.6**	-110	77	173	82	-156	113	101	-168	86	66	63	78	177	85	-130
10	43	**0.9**	-67	71	170	85	-128	69	102	-166	-96	-43	68	83	178	94	-119
11	348	**7.0**	50	113	177	79	102	68	111	178	81	85	-67	104	-175	85	-144
12	263	**5.3**	-66	70	166	84	-113	68	107	-149	-97	63	66	97	174	99	-128

Table 70. Similar clusters identified from the different clustering Schemes for *cp*PNA**9**

Cluster	No. of members	%	Scheme I Cluster #[a]	Cluster	No. of members	%	Scheme I, II Cluster #[a]
Scheme II				*Scheme III*			
A	1513	**30.3**	3	A	1773	**17.7**	2
B	1810	**36.2**	1	B	3133	**31.3**	3
C	189	**3.8**	7	C	3529	**35.3**	1
D	370	**7.4**	2	D	50	**0.5**	5
E	34	**0.7**	8	E	784	**7.8**	9
				F	731	**7.3**	6
Total number of clusters			9				6

[a] numbers correspond to identical cluster from Table 69

A comparison of the clustered conformations of *cp*PNA**9** and those of *aeg*PNA**3**, *cp*PNA**4** and *cp*PNA**6** shows major differences. The preferred structures of *cp*PNA**4** include wide loop and G-type conformations, while for *cp*PNA6 hairpin-like structures are predominant. On the contrary, the most probable conformations of the doubly modified *cp*PNA**9** have a Ω-loop structure. The number of clusters located by all schemes is reduced when comparing *cp*PNA**9** to *aeg*PNA**3** and *cp*PNA**4** (T_4). This decrease in the total number of conformers shows that the two cyclic constraints at T_4 and T_8 greatly affected the conformational flexibility. In spite of the presence of the multiple rings, the *cp*PNA**9** strand has greater flexibility than the *cp*PNA**6** strand, which is only restrained at the T_8 residue. Thus, the introduction of the cyclopentane at both positions increased its inherent flexibility with respect to *cp*PNA**6** and may substantially decrease the *cp*PNA:DNA complex stability.

The dynamics of the *cp*PNA**9** residues along the polypeptide chain were examined by means of the root mean square fluctuations (Table 71). Residues 2, 4, 7 and 8 remained fairly rigid. Residues 5 and 6 showed higher flexibility, although the average values of the RMSFs were not as high as the most flexible regions of the strand (residues 1 and 3). The N-terminus was found to be mostly unstructured and very flexible. In comparison to the unmodified PNA, *cp*PNA**9** possessed regions with relatively decreased RMSF values (residues 2, 3, 4, 7, 8). Comparing to singly modified *cp*PNA**4** (T_4) and *cp*PNA**6** (T_8), the N- and C-terminal residues (1, 2, 7 and 8) showed lower fluctuations, whereas central residues 3 to 6 had enhanced flexibility. RMSF data suggests that the

overall chain flexibility of *cp*PNA**9** is lower than those of *cp*PNA**4** or *cp*PNA**6**, which partially contradicts the clustering data. Thus, the entropic penalty incurred by restriction of *cp*PNA**9** may result in higher affinity of binding.

Table 71. Average dihedrals (°) and RMSF (bold, °) of the simulated *cp*PNA**9**

Residue	α	β	γ	δ	ε	Total RMSF	Average RMSF
T1	—	—	175.5	91.6	52.8		
	—	—	**135**	**16**	**44**	**195**	**65**
T2	170.6	-64.4	98.7	82.0	69.6		
	23	**41**	**14**	**18**	**25**	**121**	**24**
T3	176.7	13.2	88.2	85.0	71.0		
	21	**65**	**13**	**37**	**145**	**281**	**56**
T4	-127.8	69.9	52.7	102.8	69.1		
	28	**19**	**28**	**13**	**68**	**156**	**31**
T5	-179.5	-76.8	58.8	93.0	78.1		
	31	**111**	**74**	**15**	**25**	**256**	**51**
T6	178.0	-22.2	96.7	95.9	156.2		
	24	**62**	**15**	**18**	**119**	**238**	**48**
T7	-175.8	155.7	88.2	81.4	-138.0		
	22	**40**	**14**	**16**	**21**	**113**	**23**
T8	-107.7	67.0	68.8	—	—		
	16	**7**	**13**	—	—	**36**	**12**

Here again, the lowest variations are observed for the γ and δ dihedrals as well as for some of the β and ε (residues 2, 4 and 7) angles. Residues 3 and 6 exhibit the highest fluctuations for the β and ε torsions. Examination of the differences between unmodified *aeg*PNA**3** and *cp*PNA**9** revealed that the β, δ and ε torsions experienced the most changes and showed significant reductions in their RMSF values. For all systems considered, the least differences are observed in the fluctuations of the δ dihedral as opposed to the β and γ torsions, which exhibited the highest degree of differences. The flexibility of the γ2 and δ4 dihedrals were similarly affected in all the singly modified strands. However, *cp*PNA**9** and *cp*PNA**8**, the strands with several modifications have more dihedrals in common when it comes to flexibility. In this case, the list of angles that showed similar changes included the β3, β7, γ2, γ8, δ4, δ7 and ε5 dihedrals. Based on these observations, the

*cp*PNA**9** strand shares the most similarity with *cp*PNA**5**, which is unusual since the modification in this strand is at C1.

Further investigation of the average backbone dihedral values was conducted and it was found that the β dihedral has a wide permissible range. Although free to change, the ε torsions of residues 1 to 5 become restricted to the range of 50-80°. The γ and δ torsions exhibit essentially the same behavior as in all cyclopentyl PNAs and stay within their ranges of 60-90° and 80-100°. In comparison to the unmodified PNA molecule, the fewest changes in average values were registered for the γ torsions. However, when compared to the modified PNA strands, the averages differ mainly for the β and ε torsions. On the contrary, γ and δ remain close to the values seen for the other strands, aside from the dihedrals γ1-2, δ3 and δ5.

Although the *cp*PNA**10** simulation with modifications at T_1 and T_4 is very stable during the 10 ns simulation with an RMSD around 5.5 Å, the conformational spaces sampled are quite different. Twelve clusters were obtained from cluster analysis (Tables 72 and 73). Three highly populated clusters from Scheme I were also located with Scheme II. Cluster **11** from Scheme III is another abundant cluster that differs from the other three in the value of the $T_5\delta$ dihedral. All nine clusters in Scheme I were located with Scheme II, though an extra cluster was located with Scheme II. On the other hand, only three clusters from Scheme I were found to be similar with Scheme III and the remaining clusters were replaced by two new clusters in Scheme III. In Figure 59, we show representative structures for the top five clusters, each of which makes up 10-30% of the structures of *cp*PNA**10**. This is unlike *cp*PNA**8**, in which more than half of the entire ensemble is comprised of a single cluster. All of the *cp*PNA**10** clusters contain S-coil structures, except for clusters **4** and **5** from Schemes I and II, which have structures that are helical in the center (residues 4 to 6). Cluster **11** also represents an S-coil conformation. The smaller clusters appear to sample random coil and turn-like structures.

For *cp*PNA**10**, the number of conformations located with all three schemes is lower in comparison to the unmodified and the singly modified *cp*PNA**4** (T_4) and *cp*PNA**5** (T_1) strands. *Cp*PNA**10** mainly adopts S-coil conformations with a helical motif, while for *cp*PNA**4** and *cp*PNA**5** G-type and hairpin conformations were observed. Thus, the modifications at T_1 and T_4 not only reduced the flexibility of the strand, but also

induced helical structure. The thermodynamic stability of the *cp*PNA**10**:DNA complexes will increase due to larger decrease in unfavorable entropy since this oligonucleotide has an organized structure in its single-stranded state (therefore, low initial entropy).

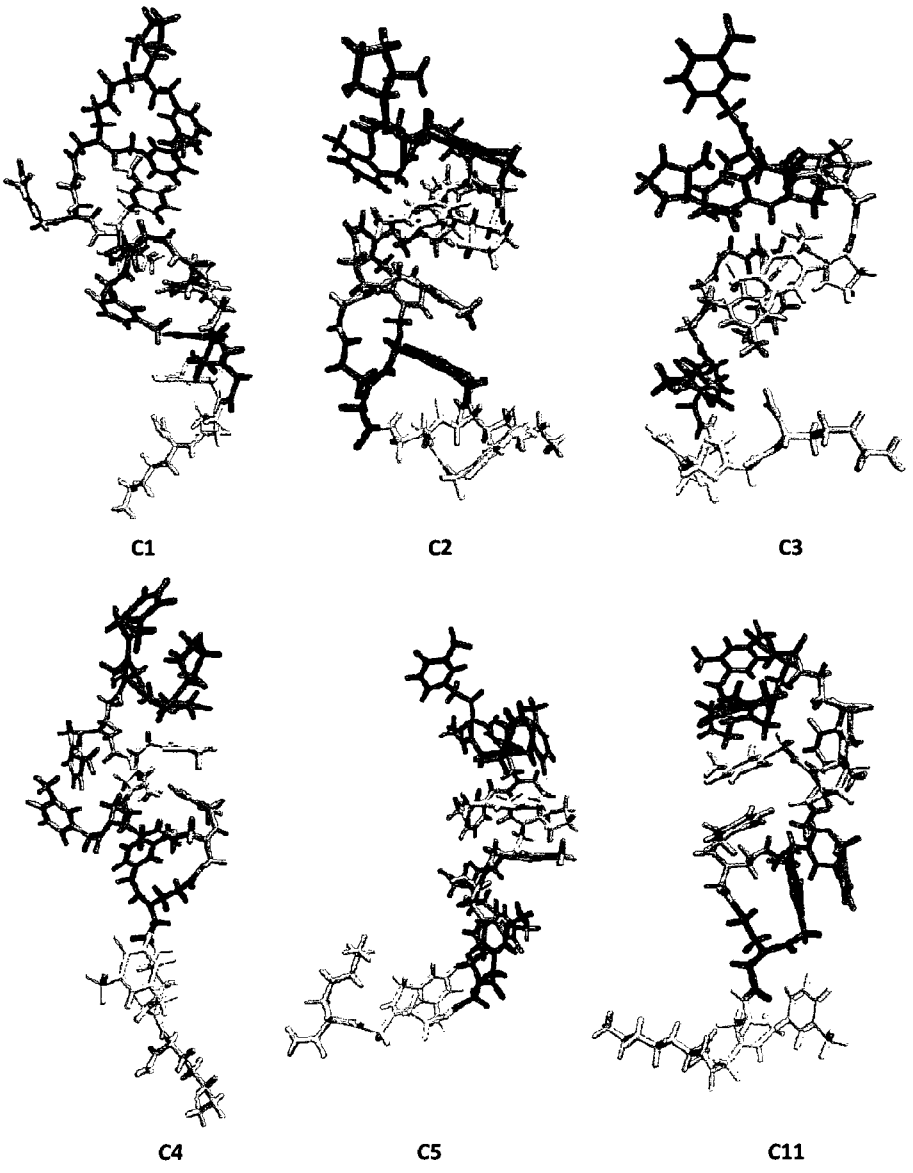

Figure 59. Representative structures of the most populated clusters from the MD simulation of *cp*PNA**10**. Residues are colored by position: **T1**-red, **T2**- dark grey, **T3**-orange, **T4**-yellow, **T5**-dark yellow, **T6**-grey, **T7**-green, **T8**-white, and **K9**-pink. The structures are positioned with **T1** on top and **K9** on bottom

Table 72. Average torsion angles for each cluster from the *cp*PNA**10** simulation

Cluster	No. of members	%	T2		T3			T4		T5			T6		T7		
			β	δ	α	γ	ε	β	δ	α	γ	ε	β	δ	α	γ	ε
Scheme I																	
1	1110	**11.1**	72	86	-166	91	-111	76	109	-166	89	-49	176	81	172	79	52
2	1119	**11.2**	-179	88	177	106	-126	65	108	-155	94	-38	175	79	173	93	47
3	1315	**13.2**	-59	85	-178	99	-133	67	105	-158	93	-50	-177	81	179	80	50
4	2342	**23.4**	-61	63	-175	99	-116	67	105	-178	91	101	-176	92	-169	84	-110
5	1719	**17.2**	-65	66	-175	95	-118	67	106	-168	105	-48	-176	80	-177	85	-109
6	333	**3.3**	96	92	-171	86	-96	71	104	-164	84	-147	-178	78	172	85	30
7	687	**6.9**	173	85	-168	93	-134	67	108	-163	93	-54	175	82	178	86	-47
8	548	**5.5**	72	87	-168	88	-100	82	105	-162	88	-75	-180	78	173	82	-50
9	827	**8.3**	-62	87	-179	99	-135	67	105	-173	89	99	179	97	-176	84	61
Scheme II																	
10	163	**3.3**	154	86	-174	104	-119	64	111	-158	89	44	168	80	175	91	44
Scheme III																	
11	4761	**47.6**	-63	68	-176	97	-120	67	105	-174	96	51	-177	89	-173	85	-86
12	146	**1.5**	71	81	-173	82	-113	148	95	-167	76	-65	165	90	170	73	69

Table 73. Similar clusters identified from the different clustering Schemes for *cp*PNA**10**

Cluster	No. of members	%	Scheme I Cluster #[a]	Cluster	No. of members	%	Scheme I Cluster #[a]
Scheme II				*Scheme III*			
A	523	**10.5**	1	A	3070	**30.7**	2
B	330	**6.6**	7	B	1462	**14.6**	3
C	465	**9.3**	2	C	561	**5.6**	6
D	654	**13.1**	3				
E	1155	**23.1**	4				
F	410	**8.2**	9				
G	875	**17.5**	5				
H	151	**3.0**	6				
I	274	**5.5**	8				
Total number of clusters			10				5

[a] numbers correspond to identical cluster from Table 72

RMSF values showed that the dihedral angles of the terminal residues 1 and 8 are the least flexible along with residues 3 and 4 (Table 74). Residues 2, 6 and 7 are characterized by higher RMSF values and are seen to engage in base stacking interactions. Differences between the RMSF values of the unmodified *aeg*PNA**3** and *cp*PNA**10** indicate an overall decrease in strand flexibility. In comparison to *cp*PNA**4**, the N- and C-terminal residues (1, 2, 7 and 8) of *cp*PNA**10** show decreased fluctuations, while residues 4 and 6 have enhanced flexibility. In the case of *cp*PNA**5**, a decrease in the RMSF values was observed for residues in the central region (3 to 5). The overall strand flexibility of *cp*PNA**10** is significantly lower than those of *cp*PNA**4** and *cp*PNA**5**. Therefore, this strand will have increased affinity toward DNA/RNA as a consequence of a lower entropic penalty.

An interesting finding is that in all cases except for *cp*PNA**9**, the modification at position 4 results in increased flexibility of residue 2 compared to the unmodified PNA. When compared to classical PNA, strands with more than one modification exhibit an overall decrease in flexibility, however, in systems like *cp*PNA**7** and *cp*PNA**9** wide portions of the chain remain very flexible, meaning the effect of the modifications is local.

Table 74. Average dihedrals (°) and RMSF (bold, °) of the simulated *cp*PNA**10**

Residue	α	β	γ	δ	ε	Total RMSF	Average RMSF
T1	—	—	-85.0	89.0	46.6		
	—	—	**18**	**13**	**28**	**59**	**20**
T2	-173.6	-168.3	96.9	77.7	-51.1		
	22	**91**	**14**	**28**	**86**	**241**	**48**
T3	-174.6	174.5	96.8	88.2	-120.6		
	23	**21**	**16**	**17**	**27**	**104**	**21**
T4	-109.5	68.5	69.8	106.1	116.0		
	17	**14**	**13**	**10**	**66**	**120**	**24**
T5	-166.9	165.8	93.4	-102.8	-5.6		
	23	**29**	**14**	**23**	**82**	**171**	**34**
T6	-165.5	-179.6	58.6	84.4	149.8		
	39	**12**	**73**	**15**	**83**	**222**	**44**
T7	-179.1	152.2	84.2	89.9	-29.5		
	22	**64**	**13**	**20**	**82**	**201**	**40**
T8	-177.4	168.1	88.7	—	—		
	27	**29**	**13**	—	—	**69**	**23**

As with *aeg*PNA**3**, *cp*PNA**4** and *cp*PNA**7-8**, the highest fluctuations are recorded for the ε dihedral whereas the least flexibility is seen for the γ and δ torsions. The β dihedral shows low fluctuations when compared to the above mentioned sequences. Nevertheless, the β, γ and ε angles show decreased fluctuations for the central residues in comparison to most systems studied. For example, in *cp*PNA**10**, the β dihedral has a lower RMSF than in *aeg*PNA**3** and *cp*PNA**6-9**. The RMSF values of the γ dihedral are lower than in *aeg*PNA**3**, *cp*PNA**5-7** and *cp*PNA**9**. The ε torsion in *cp*PNA**10** has the lowest fluctuations when compared to all other strands. The strands with multiple rings share more similarities since seven dihedrals were influenced analogously: β3, β7, γ1, γ8, δ4, δ7 and ε5.

Examination of the torsion values indicates that the modifications influenced the β dihedral substantially when compared to the unmodified strand. The constraint produced a clearer range of average values for the β dihedral with a shift to ~170°. On the

other hand, the ε dihedrals still vary significantly along the chain and differ from those of the unmodified and other *cp*PNA strands. It appears that the cyclopentane carrying residues help to pre-organize PNA by freezing both the γ and δ angles. The values of these angles range from 60° to 96° and from 80° to 100° for most residues. Departures of γ and δ from their respective preferred values are noted only in residues 1 and 5 for *cp*PNA**10**.

The structure of the *cp*PNA**11** octamer with modifications at all residues is remarkably stable, and no "folding" was observed in the simulations as one can see in Figures 36 and 38, where neither the RMSD nor the RGYR indicate a complete loss of the extended helix. Clustering of the structures shows that at a cutoff radius of 55° all the structures exist in only one cluster (Schemes I and III), while an RMS threshold of 50° (Scheme II) produced three new clusters (Tables 75 and 76, Figure 60). This indicates that the conformation of *cp*PNA**11** is very stable.

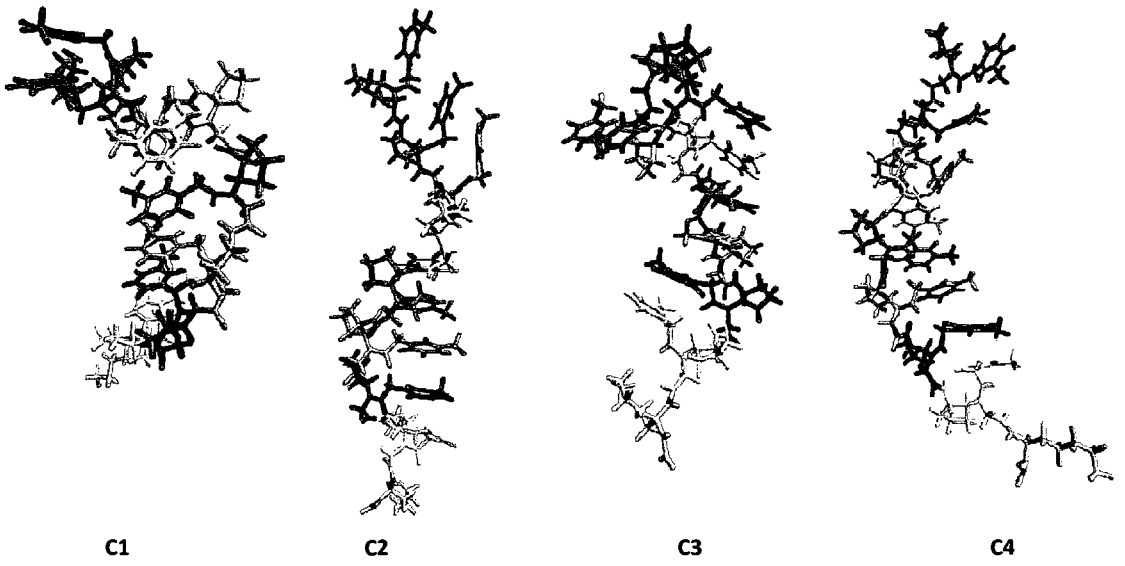

Figure 60. Representative structures of the most populated clusters from the MD simulation of *cp*PNA**11**. Residues are colored by sequence: **T1**-red, **T2**- dark grey, **T3**-orange, **T4**-yellow, **T5**-dark yellow, **T6**-grey, **T7**-green, **T8**-white, and **K9**-pink. The structures are positioned with **T1** on top and **K9** on bottom

Table 75.Torsion angle data for each cluster from simulation of *cp*PNA**11**

Clus	No. of Mem	%	T2		T3			T4		T5			T6		T7		
			β	δ	α	γ	ε	β	δ	α	γ	ε	β	δ	α	γ	ε
Scheme I																	
1	10000	**100**	67	89	-130	40	-169	70	96	-119	82	6	68	97	-118	79	4
Scheme II																	
2	176	**3.5**	66	85	-140	49	-154	142	90	-118	80	7	70	97	-118	75	6
3	1903	**38.1**	66	94	-126	21	-144	66	99	-120	81	6	68	97	-118	79	3
4	88	**1.8**	65	112	-146	33	-83	127	92	-119	85	10	66	100	-118	69	-16

Table 76. Similar clusters identified from the different clustering Schemes for *cp*PNA**11**

Cluster	No. of members	%	Scheme I Cluster #[a]
Scheme II			
A	2833	**56.7**	1
Scheme III			
A	10000	**100**	1
Total number of clusters : Scheme II – **4**; Scheme III - **1**			

[a] numbers correspond to identical cluster from Table 75

It appears that residues 1 to 3 are more flexible and this region of the molecule makes a loop motif. Cluster **1**, which was also identified by Schemes II and III, together with clusters **2** and **3** contain a wide helical structure from residues 4 to 7. On the other hand, cluster **4** appears to be similar to the initial conformation, sampling an extended helix, but cluster **4** accounts for <2% of the structures. Therefore, it does not make a significant contribution to the ensemble average.

From clustering analysis of the PNAs with multiple modifications, the highest number of clusters was located with Scheme II. For *cp*PNA**7**, the number of clusters obtained with Scheme I was lower than in Scheme III. On the other hand, Scheme III identified the lowest number of clusters for *cp*PNA**9** and *cp*PNA**10**. The lowest number

of clusters over the three Schemes was reported for *cp*PNA**11**. In comparison to all doubly modified strands, *cp*PNA**8**, *cp*PNA**9** and *cp*PNA**10**, *cp*PNA**7** with three cyclopentane rings shows a higher degree of flexibility based on the clustering results. While the strands with rings at the N or C terminals and position 4 (*cp*PNA**9** and *cp*PNA**10**) show similar levels of flexibility, the *cp*PNA**8** with constraints at both terminals has a slightly enhanced strand flexibility. Clearly, the clustering results indicate that *cp*PNA**11** is the least flexible strand. The greater restricted conformational freedom for *cp*PNA**11** enforces a decrease in the conformational entropy.

The MD simulations of *cp*PNA**7**, *cp*PNA**8** and *cp*PNA**9**, which start from a helical structure, predict that the *cp*PNA strand will neither maintain nor adopt a helical conformation. These strands have coil and hairpin-like conformations. However, the backbone conformations identified for *cp*PNA**10** are S-type coils with some helical character and the PNA conformation corresponding to the helical strand is the most populous. If PNA is pre-organized for binding to DNA/RNA, then its conformations in solution should overlap with the conformation adopted in its complex with either DNA or RNA. Examination of structures from the predominant clusters of *cp*PNA**11** reveals that the lowest energy conformation does strongly resemble a helical structure. The calculations indicate that the fully modified *cp*PNA**11** does have an intrinsic tendency to assume the conformation found in native complexes even when it is free in solution; that is, the strand is significantly pre-organized.

In *cp*PNA**11**, the overall stiffness of the chain results in low RMSF values (Table 77). In comparison to all modified PNAs, the flexibility of the *cp*PNA**11** appears to be dramatically reduced, suggesting that the increase in stiffness of the structure is uniformly distributed along the backbone. Thus, structural pre-organization and the reduced flexibility of *cp*PNA**11** will contribute to the hybridization stability of PNA. Low RMSF values are found for *cp*PNA**5**, *cp*PNA**8** and *cp*PNA**10** around residues 1 and 2, indicating that the modification at position 1 alters the dynamics of the N-terminal region. As expected, all dihedrals, except for ε1 and ε2, exhibit low RMSF values for *cp*PNA**11**. In comparison to all systems, the β and ε dihedrals have the lowest RMSF values for *cp*PNA**11**. Fluctuations of the δ dihedral are lower relative to the ones seen in *cp*PNA**4**

and *cp*PNA**9** while the γ torsion has reduced fluctuations when compared to *cp*PNA**5-7** and *cp*PNA**9**.

Table 77. Average dihedrals (°) and RMSF (bold, °) of the simulated *cp*PNA**11**

Residue	α	β	γ	δ	ε	Total RMSF	Average RMSF
T1	—	—	-75.6	104.8	-37.1		
	—	—	**18**	**14**	**65**	**97**	**32**
T2	-142.8	66.7	27.7	89.0	158.9		
	24	**7**	**22**	**17**	**75**	**145**	**29**
T3	-130.1	68.7	39.8	91.5	-169.4		
	21	**7**	**25**	**16**	**32**	**101**	**20**
T4	-122.6	70.1	72.8	96.1	10.1		
	20	**18**	**21**	**11**	**21**	**91**	**18**
T5	-118.9	68.3	82.2	96.1	6.6		
	19	**7**	**11**	**9**	**19**	**65**	**13**
T6	-119.4	68.3	81.9	96.4	6.7		
	18	**7**	**11**	**9**	**19**	**64**	**13**
T7	-117.9	68.3	79.3	98.3	3.6		
	17	**7**	**11**	**10**	**20**	**65**	**13**
T8	-117.6	68.3	70.9	—	—		
	18	**7**	**11**	—	—	**36**	**12**

Evaluation of the torsion angle data yields γ values in the ranges of 70-80° for the central residues 4 to 8, and 30-40° for residues 2 and 3. Likewise, the ε dihedrals adopt a value of 10° in residues 4 to 7, but remain primarily at 180° for residues 2 and 3. Furthermore, for the β and δ dihedrals the distribution is somewhat narrow; only values around 70° are observed for the β torsion and values around 90° for the δ torsion. The chief differences between the unmodified strand and *cp*PNA**11** are in the preferences of the β, δ and ε dihedrals. In the unmodified *aeg*PNA**3** the β torsions occupy the *trans* region, whereas in *cp*PNA**11**, the average β values are restricted to ~70°. The δ dihedral assumes various values in the unmodified strand, but has a fixed value of ~100° in *cp*PNA**11**. In the case of the γ torsions, changes in the averages are evident only for residues 2 and 3. When compared to the *cp*PNAs with multiple modifications, different sampling of the α and δ dihedrals is noted. The pre-organized structure of the fully

modified *cp*PNA**11** adopts dihedral values characteristic of RNA duplexes; therefore, is more compatible with RNA-type complexes. These results suggest that the increased stability of the duplexes with *cp*PNA**11** will be a result of large increases in favorable enthalpy of hybridization that is not compensated for by increases in unfavorable entropy (due to reduced flexibility).

More generally, the overall results show that fluctuations in the central part (residues 4 and 5) of any polythymine strand containing cyclopentane at positions 1 or 8, regardless of other alterations, are reduced. Modifications at both the T_4 and T_8 positions affect the RMSFs of residue 3, resulting in a higher flexibility compared to the rest of the chain residues. For *cp*PNA**7** (modified residues: T_1, T_4 and T_8) and *cp*PNA**9** (modified residues: T_4 and T_8), the reduced conformational fluctuations are localized to the site of the modifications with only a small influence on the fluctuations of adjacent residues. According to clustering and RMSF data, among all *cp*PNAs with multiple modifications, the fully modified *cp*PNA**11** strand had the lowest flexibility, while *cp*PNA**7** (T_1, T_4 and T_8) exhibited the highest mobility. *Cp*PNA**10** (T_1 and T_4) showed higher fluctuations in comparison to *cp*PNA**11**, but it had lower flexibility than *cp*PNA**7**-**9**. Based on RMSF results, *cp*PNA**8** and *cp*PNA**9** had comparable flexibility. Analysis of average dihedral angles revealed that in strands with multiple modifications changes were observed the in β and ε values. In *cp*PNA**7** and *cp*PNA**10**, the β angle is mostly in the *trans* conformation, while in *cp*PNA**9** and *cp*PNA**11** it is in the *gauche*± conformation. *Cp*PNA**18** adopts conformations where this torsion angle populates the *gauche* and *trans* conformations. The δ dihedral values are centered around 90° in all *cp*PNAs. These results indicate that the number of modifications and the position have an effect on the torsional potential energy surface.

Evidence of greater stiffness of the cyclopentyl modified PNAs with respect to the classical *aeg*PNA strand suggests that the constraint plays an important role in determining the flexibility of the strand as well as the dynamics of certain regions of the molecular structure. The structures for single as well as multiple modifications suggest that the overall geometry attained by the molecule restricts the conformational of the strand. Again, this result can be attributed to the presence of the modification in the backbone of PNA, which in turn imparts some rigidity to the molecular geometry. Thus,

the constrained *cp*PNAs (*cp*PNA**10**, *cp*PNA**11**) will bind with more favorable binding entropies than the flexible ones (*cp*PNA**7** or *cp*PNA**8-9**).

PNA-MB$_{10}$ sequence: *cp*PNA**19**-*cp*PNA**22**

The *cp*PNA**19** (modifications at T_6 and C_7) and *cp*PNA**20** (modifications at T_6 and A_8) simulations at 298 K are stable along the 10 ns periods. The RMSD suggests that the structures remain close to the helical structure. Overall, the RMSD and RGYR indicate that at least half of the structures populated by *cp*PNA**19** have residual structural motifs from the initial conformation (Figures 37 and 39). The conformations of *cp*PNA**19** were grouped into sixteen distinct clusters (Figure 61, Tables 78 and 79). The highly populated clusters from Scheme I were also identified in Schemes II and III. Five clusters were found to be similar in Schemes I and II while only three were identical between Schemes I and III. The remaining clusters in Scheme I were replaced by seven and two new clusters in Schemes II and III, respectively.

A stable helical motif is observed in the central region of the molecule and is present for 4 ns. Cluster analysis indicates that *cp*PNA**19** samples significant populations of coil structures with helical content (cluster **7**). Many structures in the ensemble of *cp*PNA**19** have N- and C-terminal loops, which are represented by clusters **2** and **3** from Schemes I and III. However, in cluster **3** the loops become more bent and resemble two knots at the termini. Cluster **16** has a limited population of coil structures with two inner loops running from residues 3 to 6 and 8 to 10.

Based on the Scheme I and III results, *aeg*PNA**14** has more clusters than *cp*PNA**19**. In comparison to *cp*PNA**15**, the *cp*PNA**19** strand had a larger number of clusters in Schemes II and III, but it had fewer clusters with Scheme I. Similarly, when compared to *cp*PNA**19**, *cp*PNA**17** had more conformations with Schemes I and II but a lower number was observed for Scheme III. The simulations also show that the conformation of the *cp*PNA**19** decamer is significantly changed by multiple modifications at these specific positions. *Cp*PNA**15** and *cp*PNA**17** exhibit hairpin-like and Ω–loop conformations, while the predominant structures identified for *cp*PNA**19** include coil conformations with helical content. These results suggest that the flexibility

of the *cp*PNA**19** (T_6, C_7) strand with two modifications is higher or comparable to that of *cp*PNA**15** and *cp*PNA**17**, which contain a single cyclopentane at either T_6 or C_7 positions. Thus, the observed pre-organized structure for *cp*PNA**19** indicates no significant entropic advantage to forming a duplex, but it may have an enthalpic advantage. However, the thermodynamic stability of the complexes with DNA will be increased compared to the non-modified *aeg*PNA**14**.

Figure 61. Representative structures of the most populated clusters from the MD simulation of *cp*PNA**19**. Residues are colored by name: **APN**-blue, **CPN**-orange, **TPN**-green, **TPC**-light blue, **CPC**–pink, **GPN**-red and **K**-purple (see Table 17 for sequence)

Table 78. Average torsion angles for each cluster from the *cp*PNA**19** simulation

Cluster	No. of Members	%	T2		A3			G4		G5			A6		C7			A8		C9		
			β	δ	α	γ	ε	β	δ	α	γ	ε	β	δ	α	γ	ε	β	δ	α	γ	ε
Scheme I																						
1	1947	**19.5**	68	81	178	81	68	63	79	176	85	33	66	96	-116	75	39	65	81	174	79	69
2	2324	**23.2**	-171	108	169	110	-62	64	106	178	98	30	66	97	-115	77	12	-180	95	-178	97	12
3	1692	**16.9**	-177	102	178	101	-88	55	111	-178	96	13	68	99	-115	72	40	76	125	-175	100	112
4	385	**3.9**	-169	87	175	105	-65	92	96	-178	102	28	67	96	-115	74	153	65	113	174	100	66
5	298	**3.0**	-168	122	176	110	-65	88	107	-166	104	28	67	97	-118	75	41	72	110	-176	98	53
6	232	**2.3**	-156	104	170	90	-79	165	72	174	88	30	66	96	-114	70	56	60	60	-161	92	88
7	3122	**31.2**	-175	92	171	96	68	64	85	179	91	30	67	96	-116	74	39	67	83	176	85	55
Scheme II																						
8	237	**4.7**	-177	111	164	110	-60	61	108	177	95	33	66	97	-115	76	1	180	96	155	92	-90
9	617	**12.3**	-171	104	170	109	-60	66	102	-173	101	27	67	98	-116	75	5	-179	92	167	96	72
10	99	**2.0**	-172	112	178	104	-66	144	103	173	100	53	64	98	-114	72	36	65	92	177	88	53
11	316	**6.3**	-168	113	169	112	-63	61	112	165	95	34	66	97	-114	80	33	-180	101	-127	106	-25
12	35	**0.7**	-158	84	156	86	-85	158	76	160	90	28	66	98	-115	71	159	67	106	-176	94	46
13	74	**1.5**	-176	92	178	103	-52	79	103	-165	102	30	67	95	-121	73	40	73	93	-170	97	61
14	74	**1.5**	-175	88	179	109	-59	73	106	-163	99	-7	70	95	-113	74	148	67	117	160	100	82
Scheme III																						
15	3263	**32.6**	113	86	179	86	66	64	77	174	86	34	66	95	-117	75	38	66	81	175	81	67
16	1032	**10.3**	-162	103	168	94	5	95	83	178	90	31	66	97	-115	71	63	64	94	-171	97	54

Table 79. Similar clusters identified from the different clustering Schemes for *cp*PNA**19**

Cluster	No. of members	%	Scheme I Cluster #[a]	Cluster I	No. of members	%	Scheme I Cluster #[a]
Scheme II				*Scheme III*			
A	1547	**30.9**	7	A	1281	**12.8**	7
B	973	**19.5**	1	B	2679	**26.8**	2
C	849	**17.0**	3	C	1745	**17.5**	3
D	84	**1.7**	4				
E	51	**1.0**	5				
Total number of clusters			12				5

[a] numbers correspond to identical cluster from Table 78

The average RMSF per residue was calculated and the results are presented in Table 80. The most mobile residues are located on the termini, (residues 1 and 10), and at the two loops (residues 3, 8 and 9). A similar but less pronounced effect was observed for *cp*PNA**17** with modification at C_7. The modification at two positions shows lower RMSFs for residues 6 to 8 but enhances the mobility of the N- and C-terminal residues (1 to 3 and 10) when compared to the unmodified strand. This is in contrast to *cp*PNA**15**, where the flexibility of the N-terminal residues is reduced upon single modification at T_6 and only the C-terminal residues have increased motions. Moreover, the overall flexibility of the *cp*PNA**19** is comparable to *aeg*PNA and this strand is the most flexible strand for all sequences containing a cyclopentane modification. Consequently, the modified *cp*PNA**19** will exhibit less favorable entropy of binding and will bind to DNA with lower affinity than the rigid cyclopentyl analogs.

A range of RMSFs is seen for the four dihedrals: β, γ, ε and ε. Nevertheless, most γ (residues 3 to 7, 9) and δ (residues 2, 5 to 8) torsions have low RMSFs. Similar observations were made for all sequences with modifications, irrespective of position or number. The fluctuations of the β and ε torsions in residues 6 to 9 are smaller than the ones in *aeg*PNA**14**. However, several β torsions in the central as well as terminal regions exhibit higher variability when compared to *cp*PNA**15** or *cp*PNA**17**. The terminal residue γ dihedrals have higher RMSF values than those in *aeg*PNA**14**, *cp*PNA**15** and *cp*PNA**17**. As for δ, it showed enhanced fluctuations in comparison to the rest of the PNA strands.

The *cp*PNA**19** strand shares the strongest similarity with the *cp*PNA**16** and *cp*PNA**17**. Consequently, five commonly affected dihedrals were identified: β8, β9, γ7, ε7 and ε9.

Table 80. Average dihedrals (°) and RMSF (bold, °) of the simulated *cp*PNA**19**

Residue	α	β	γ	δ	ε	Total RMSF	Average RMSF
G1	—	—	51.0	-18.0	77.1		
	—	—	**102**	**157**	**81**	**340**	**113**
T2	172.5	163.5	-19.3	96.0	62.9		
	22	**50**	**88**	**20**	**36**	**216**	**43**
A3	-179.6	174.9	98.1	-175.4	-1.3		
	25	**69**	**17**	**111**	**73**	**295**	**59**
G4	179.2	66.2	82.6	117.4	-37.3		
	23	**26**	**13**	**63**	**102**	**227**	**45**
A5	179.1	131.6	93.1	77.7	27.4		
	26	**56**	**15**	**16**	**34**	**147**	**29**
T6	-135.1	66.7	83.5	96.7	4.5		
	29	**8**	**11**	**10**	**17**	**75**	**15**
C7	-116.0	67.7	74.2	101.2	37.8		
	16	**7**	**11**	**10**	**42**	**86**	**17**
A8	-162.4	94.3	1.8	94.1	27.7		
	33	**49**	**91**	**26**	**67**	**266**	**53**
C9	179.3	138.3	90.4	15.1	57.7		
	30	**88**	**15**	**124**	**54**	**311**	**62**
T10	172.4	139.0	2.4	—	—		
	26	**61**	**86**	—	—	**173**	**58**

Closer inspection of the average dihedral values revealed two ranges for the β torsions: 140-170° (residues 2, 3, 5, 9, 10) and 70-95° (residues 4, 6 to 8). As mentioned earlier, the γ torsion assumes values in the 70-100° range, except for T_2 and A_8, where the averages are 2° and -20°, respectively. Analogously, the δ torsion varies from 80 to 100° throughout the simulations, except for A_3 and C_9. Finally, the average values for the ε torsion vary from 30-60° in all nucleotides but A_3, G_4 and T_6. The major differences between *cp*PNA**19** and *aeg*PNA**14** angles are seen for the β and ε dihedrals. A similar trend is observed when cyclopentyl PNAs are compared. One exception to this

exists: in the case of *cp*PNA**16**, with a modification at G_4, the average values for all dihedrals have wider ranges.

After conformational sampling, the resulting ensemble for *cp*PNA**20** contained structures mostly in the closed loop conformation (Tables 81 and 82, Figure 62). Two of the highly populated clusters from Scheme I were located with Scheme III, but only one was found with Scheme II. All seven clusters of Scheme I were identified in Scheme III and six clusters matched those from Scheme II. Scheme II located two new clusters that were not similar to the additional clusters located with Scheme III. The termini are more flexible with approximately 40% of the structures from Scheme I containing an N-terminal loop (cluster **3**). The identical clusters in Schemes II and III as well as cluster **8** also have similar conformations. Other types of secondary structural elements are less prevalent. Cluster **1** contained exclusively *cp*PNAs adopting conformations similar to the starting NMR structure.

*Cp*PNA**20** had fewer clusters than *aeg*PNA**14** for Schemes I and II. The number of conformations sampled is larger for Schemes I and II in the simulation of the *cp*PNA**20** single strand than in the simulation the modified *cp*PNA**15**, but *cp*PNA**20** has fewer clusters for Scheme III. The number of clusters across all three schemes was higher when comparing *cp*PNA**20** to *cp*PNA**18**. Both the *cp*PNA**20** (T_6 and A_8) and *cp*PNA**15** (T_6) strands adopt similar closed-loop structures, but for *cp*PNA**18** (A_8) helical coil conformations were observed. Thus, multiple modifications reduce the flexibility of the PNA strand in comparison to the *aeg*PNA**14**; however, the strands with single modification at either T_6 or A_8 show greater decrease in flexibility than that with both residues substituted. This finding suggests that the thermodynamic stability of the *cp*PNA**20** complexes with DNA/RNA will not be greatly increased in comparison to the unmodified PNA. The higher flexibility of this strand will lead to larger entropic losses upon duplex formation and decreased binding affinity in comparison to singly modified *cp*PNA**15** or *cp*PNA**18**.

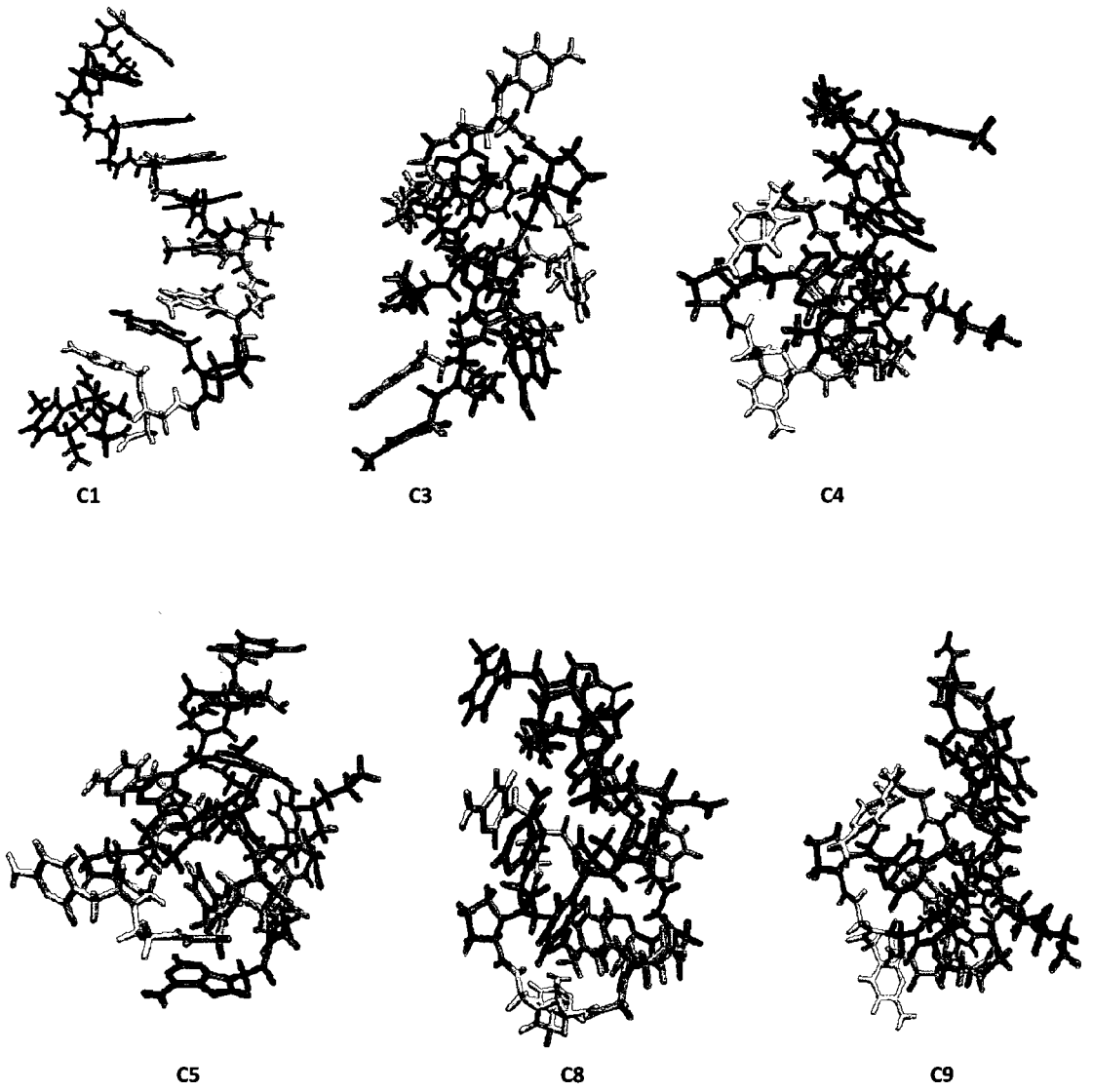

Figure 62. Representative structures of the most populated clusters from the MD simulation of *cp*PNA**20**. Residues are colored by name: **APN**-blue, **CPN**-orange, **TPN**-green, **TPC**-light blue, **APC**-light brown, **GPN**-red and **K**-purple. The structures are positioned with **GPN1** on top and **K11** on bottom (see Table 17 for the sequence)

Table 81. Average torsion angles for each cluster from the *cp*PNA**20** simulation

Cluster	No. of Members	%	T2		A3			G4		G5			A6		C7			A8		C9		
			β	δ	α	γ	ε	β	δ	α	γ	ε	β	δ	α	γ	ε	β	δ	α	γ	ε
Scheme I																						
1	1192	**11.9**	67	94	178	89	63	65	78	177	81	26	66	98	-169	81	11	67	102	167	76	75
2	579	**5.8**	-176	84	176	92	-132	175	-116	-166	103	23	68	100	-156	95	123	69	98	143	105	127
3	3729	**37.3**	-58	110	177	94	-88	178	-92	179	95	-105	80	105	172	89	177	70	95	177	74	-84
4	2200	**22.0**	-4	102	176	83	60	54	-100	170	94	-115	91	105	-178	93	179	67	100	177	75	-96
5	1408	**14.1**	-24	112	172	84	81	178	-91	-175	86	-120	76	105	174	102	106	67	107	175	92	-109
6	637	**6.4**	177	102	176	93	63	65	126	179	94	34	67	101	-179	94	62	67	105	176	102	81
7	255	**2.6**	-76	125	175	102	121	180	-113	-166	95	6	66	105	-177	88	170	71	92	164	103	169
Scheme II																						
8	748	**15.0**	-57	108	179	82	34	82	-92	168	96	-111	108	105	171	87	175	70	91	176	82	-94
9	691	**13.8**	57	97	171	86	84	72	-105	178	91	-118	72	104	-165	99	171	64	111	178	66	-105
Scheme III																						
10	541	**5.4**	-66	119	174	101	-105	174	-96	179	89	-87	73	107	-170	88	143	66	121	155	105	-29

Table 82. Similar clusters identified from the different clustering Schemes for *cp*PNA**20**

Cluster	No. of members	%	Scheme I Cluster #[a]	Cluster I	No. of members	%	Scheme I Cluster #[a]
Scheme II				*Scheme III*			
A	595	**11.9**	1	A	1246	**12.5**	1
B	320	**6.4**	6	B	583	**5.8**	6
C	292	**5.8**	2	C	3223	**32.2**	3
D	1792	**35.8**	3	D	2467	**24.7**	4
E	439	**8.8**	5	E	1089	**10.9**	5
F	123	**2.5**	7	F	453	**4.5**	2
				G	398	**4.0**	7
Total number of clusters			8				8

[a] numbers correspond to identical cluster from Table 81

The RMSFs of *cp*PNA**20** compared with the respective values from the unmodified strand show that modifications at positions 6 and 8, cause an increase in flexibility for residues 1 to 4 (Table 83). The mobility of the remaining part of the strand is significantly reduced. Overall, the flexibility of the strand is comparable to *aeg*PNA**14**. A single modification at C_7 (*cp*PNA**17**) produced a similar effect, although to a higher extent. The analysis of the RMSFs between *cp*PNA**15** (modification at T_6), *cp*PNA**18** (modification at A_8) and *cp*PNA**20** (modifications at T_6 and A_8) show differences in the mobility of the central residues. A single modification at either the T_6 or A_8 position has a profound effect on flexibility, spanning more residues in the central part than was observed for *cp*PNA**20**. Modifying both residues rigidifies the backbone only locally. Thus, weaker binding to DNA would be predicted for this strand.

Here again, the β and ε torsional fluctuations dominate; whereas, the γ and δ dihedrals only vary slightly. If compared to the unmodified sequence, the β, γ and ε dihedrals demonstrate decreased fluctuations for residues 6 to 9. Yet, as opposed to most *cp*PNA strands, the β, δ and ε fluctuations are significantly higher in residues 2 to 4.

The extent of the variations in the PNA average values suggests that possible conformational changes are caused by the modifications. The key conformational parameters, the γ and δ torsions, do not exhibit wide variations in the oligonucleotide structure. Regarding the distribution of the β and ε angles, there exist differences between

*aeg*PNA**14** and *cp*PNA**20**. Modification at T_6 and A_8 favors a broader distribution of these angles. In comparison to *cp*PNA**15** and *cp*PNA**18**, the main differences are in the β, γ and ε dihedrals.

Table 83. Average dihedrals (°) and RMSF (bold, °) of the simulated *cp*PNA**20**

Residue	α	β	γ	δ	ε	Total RMSF	Average RMSF
G1	—	—	55.5	110.3	43.7		
	—	—	**73**	**43**	**72**	**188**	**63**
T2	177.8	-84.6	-49.5	105.2	-110.4		
	25	**99**	**76**	**23**	**99**	**322**	**64**
A3	177.7	-64.8	89.6	1.2	-10.3		
	27	**123**	**15**	**149**	**94**	**408**	**82**
G4	179.1	129.3	-45.1	-127.1	-109.3		
	28	**63**	**81**	**70**	**102**	**344**	**69**
A5	179.0	156.4	92.2	89.3	-74.9		
	24	**41**	**14**	**19**	**66**	**164**	**33**
T6	-111.3	78.5	71.8	103.6	43.3		
	27	**22**	**15**	**10**	**32**	**106**	**21**
C7	179.6	172.6	91.1	92.1	137.5		
	28	**50**	**14**	**14**	**67**	**173**	**35**
A8	-116.7	68.3	51.0	99.2	54.8		
	19	**7**	**24**	**15**	**46**	**111**	**22**
C9	177.6	20.0	81.5	-161.1	-42.2		
	29	**54**	**26**	**71**	**90**	**270**	**54**
T10	176.5	139.3	90.6	—	—		
	22	**72**	**14**	—	—	**108**	**36**

There is a large jump in RMSD for *cp*PNA**21** (modifications at C_7 and A_8) at ~ 4 ns, which corresponds to a further deviation from the NMR reference structure. The conformational behavior of *cp*PNA**21** was studied by cluster analysis and thirteen distinct clusters were obtained (Tables 84 and 85, Figure 63). There are no differences between the highly populated clusters obtained by the three criteria. Nine identical clusters were located in Schemes I and II. The remaining cluster of Scheme I was substituted by a different cluster in Scheme II. On the contrary, only three clusters from Scheme I were found to be similar in Scheme III. The seven remaining clusters were replaced by only

two clusters with low percentages. The clusters were also analyzed from a conformational flexibility point of view. The conformations fit only two large families. During the first 4.2 ns, the molecule exists in a somewhat helical conformation (clusters **1** and **10**). The rest of the clusters represent a pattern with a more compact, S-coil conformation and accounts for 65% of the structures obtained with the parameters of Scheme I.

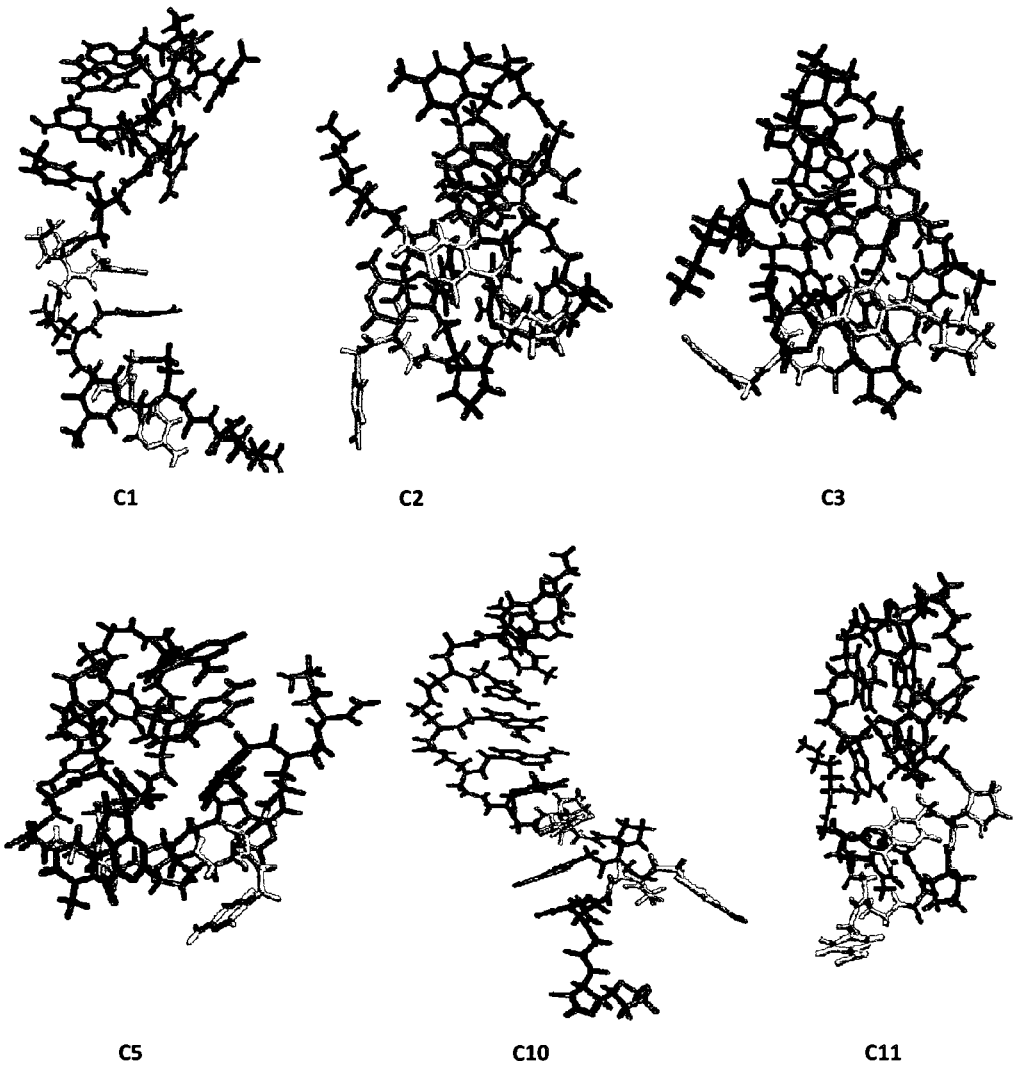

Figure 63. Representative structures of the most populated clusters from the MD simulation of *cp*PNA**21**. Residues are colored by name: **APN**-blue, **CPN**-orange, **TPN**-green, **CPC**- pink, **APC**-light brown, **GPN**-red and **K**-purple. The structures are positioned with **GPN1** on top and **K11** on bottom (see Table 17 for the sequence)

Table 84. Average torsion angles for each cluster from the *cp*PNA**21** simulation

Cluster	No. of Members	%	T2		A3			G4		G5			A6		C7			A8		C9		
			β	δ	α	γ	ε	β	δ	α	γ	ε	β	δ	α	γ	ε	β	δ	α	γ	ε
Scheme I																						
1	2791	**27.9**	175	87	155	106	74	60	83	179	88	67	68	85	-112	73	2	66	107	-176	102	-131
2	2146	**21.5**	174	89	158	106	77	56	102	173	115	60	168	82	-140	22	53	64	102	-164	102	-177
3	1467	**14.7**	175	89	159	106	77	56	101	170	113	-24	176	78	-135	25	55	65	100	-144	103	-167
4	644	**6.4**	72	79	180	84	69	63	77	177	81	65	65	81	-127	75	-1	67	103	-168	84	75
5	1029	**10.3**	175	85	158	106	77	56	98	161	107	-48	-177	81	-116	63	2	65	91	175	97	159
6	106	**1.1**	66	82	171	74	-57	155	96	174	82	67	68	73	-141	72	-5	66	109	168	99	-156
7	129	**1.3**	65	105	176	89	70	60	81	174	76	-59	159	96	-103	72	-7	67	110	172	96	-131
8	515	**5.2**	179	97	162	101	73	60	85	174	99	-38	170	91	-107	68	-7	66	107	169	104	-122
9	379	**3.8**	66	80	178	83	68	65	78	180	80	66	64	74	-129	72	-5	66	108	-174	104	173
10	767	**7.7**	67	80	178	84	65	66	78	180	81	66	65	78	-129	76	-3	67	107	174	107	-115
Scheme II																						
11	506	**10.1**	175	87	156	107	74	60	89	180	94	68	71	80	-116	69	4	67	103	-164	102	173
Scheme III																						
12	310	**3.1**	102	115	177	85	60	70	75	171	79	-46	148	87	-106	71	-5	67	111	174	97	-132
13	230	**2.3**	167	88	170	97	64	67	71	172	82	21	103	86	-108	70	-5	67	110	174	98	-139

Table 85. Similar clusters identified from the different clustering Schemes for *cp*PNA**21**

Cluster	No. of members	%	Scheme I Cluster #[a]	Cluster	No. of members	%	Scheme I Cluster #[a]
Scheme II				*Scheme III*			
A	922	**18.4**	1	A	2103	**21.0**	1
B	1049	**21.0**	2	B	5520	**55.2**	2
C	521	**10.4**	10	C	1837	**18.4**	9
D	362	**7.2**	4				
E	715	**14.3**	3				
F	529	**10.6**	5				
G	50	**1.0**	6				
H	64	**1.3**	7				
I	266	**5.3**	8				
Total number of clusters			10				5

[a]numbers correspond to identical cluster from Table 84

For Schemes II and III, *cp*PNA**21** (C_7 and A_8) had more clusters than *cp*PNA**17** (C_7) but the number of clusters from Scheme I was higher for *cp*PNA**17**. For all three schemes, *cp*PNA**21** had fewer clusters than *aeg*PNA**14** and *cp*PNA**18** (A_8). These results all suggest that the addition of the cyclopentane ring decreased the flexibility of the strand in comparison to the unmodified PNA; however, the two modifications at C_7 and A_8 did not have as large of an effect on strand flexibility as single alterations at either position alone. A comparison of conformations between *cp*PNA**21** and *cp*PNA**17** or *cp*PNA**18** shows that they adopt various coil conformations, but for *cp*PNA**18** helical structure was also observed. Consequently, multiple rings at both locations only slightly altered the overall shape of the structure.

The analysis of the RMSFs for *cp*PNA**21** shows a maximum for the N-terminal residue, which can be interpreted as higher mobility for the turn (Table 86). Residues 3, 4, 7 and 8 have low RMSF values and are most often seen in base stacking interactions. Furthermore, high fluctuations are also found for residues 5 and 6, which belong to the central part of the strand and are located just above the modification. In contrast, the entire *cp*PNA**18** strand is rigidified as a consequence of the single modification at position 8. A similar increase in the flexibility of the N-terminal residues is observed for the strand carrying a single cyclopentane at residue C_7 (*cp*PNA**17**). In comparison to

*aeg*PNA**14**, the mobility of the chain was constrained in the regions of residues 3, 4 and 7 to 10. Differences between the RMSF values of the C-terminal residues for *cp*PNA**17**, **18** and **21** indicate that the limited mobility of this region is induced by modification at A_8. The overall strand flexibility of *cp*PNA**21** is lower than those of *cp*PNA**17**, *cp*PNA**19** and *cp*PNA**20**. However, it has higher flexibility in comparison to *cp*PNA**15** and *cp*PNA**18**, which suggests that *cp*PNA**21** will lose more entropy upon binding to DNA.

Table 86. Average dihedrals (°) and RMSF (bold, °) of the simulated *cp*PNA**21**

Residue	α	β	γ	δ	ε	Total RMSF	Average RMSF
G1	—	—	68.06	92.44	86.20		
	—	—	**106**	**19**	**44**	**169**	**56**
T2	-170.58	160.77	-52.21	87.13	70.48		
	27	**49**	**79**	**15**	**34**	**204**	**41**
A3	162.34	-29.79	101.06	85.77	72.22		
	26	**27**	**13**	**14**	**21**	**101**	**20**
G4	172.60	60.29	80.82	90.60	74.32		
	18	**15**	**10**	**17**	**18**	**78**	**16**
A5	-175.43	-5.83	98.38	86.49	32.62		
	20	**62**	**19**	**17**	**52**	**170**	**34**
T6	173.85	123.13	92.79	81.99	154.90		
	34	**55**	**17**	**15**	**92**	**213**	**43**
C7	-124.23	67.27	54.00	93.71	19.25		
	22	**7**	**27**	**15**	**32**	**103**	**21**
A8	-131.71	65.66	67.89	102.82	60.23		
	28	**7**	**17**	**10**	**35**	**97**	**19**
C9	-170.54	68.12	100.73	88.00	64.04		
	37	**42**	**12**	**18**	**55**	**164**	**33**
T10	-173.88	107.98	80.45	—	—		
	22	**49**	**12**	—	—	**83**	**28**

The lowest fluctuations are noticed for the γ, δ and some β and ε dihedrals (residues 3, 4, 7 and 8). Upon comparison to the *aeg*PNA and *cp*PNA strands, significant reductions in the RMSFs are noted for the β, γ and ε torsions (residues 3, 4 and 7 to 9). In addition, the δ dihedrals have a similar level of flexibility as in the unmodified reference PNA, but demonstrate considerable rigidity when compared to other *cp*PNAs. It is worth

noting that the modification-induced changes in the β and ε fluctuations are registered mainly in residues carrying the cyclopentane. This finding suggests that the impact of the constraint on the flexibility of these dihedrals is limited to the site of the modification. Based on our observations, the *cp*PNA**21** strand shares the most similarity with *cp*PNA**16**. As mentioned previously, the flexibility of the β8, β9, γ7, ε7 and ε9 dihedrals was affected (reduced) in comparison to all PNAs with a single cyclopentane. It should be noted, however, that the strands with multiple rings have more similarities to each other since a larger number of dihedrals become restricted as a consequence of the modifications (β8, β9, γ2, γ7, γ8, ε7 and ε9).

The average backbone dihedral values are reported in Table 86. A number of residues have average β values in the 60-70° range. Exceptions are noted for T_2, A_3, A_5 and T_6. The γ torsions mainly range from 70 to 100° except in T_2 where the average is -52°. Similar average values observed for all *cp*PNAs were also seen for the δ torsions (90-100°). The torsion values for ε vary from 70 to 90° in most nucleotides, except for A_5, T_6 and C_7. In comparison to the unmodified PNA molecule, the most changes in the average values were registered for the β and ε torsions, whereas no significant differences are observed for δ. However, when compared to *cp*PNA strands, the averages vary mainly for the β and ε torsions. By contrast, γ along with δ remain close to their predominant values.

The structures sampled for *cp*PNA**22** (modifications T_6, C_7 and A_8) are not very different from the initial helical reference, resulting in an RMSD of 4.3 Å. Evaluation of the generated ensembles indicates that fast conformational changes are occurring within small domains. The cluster analysis shows that the two largest clusters from Scheme I account for ~60% of the ensemble (Figure 64, Tables 87 and 88). One of the most populated clusters from Scheme I was identified in both Schemes II and III. The other highly populated cluster from Scheme I was only located with Scheme III. In Scheme II, the cluster with the highest membership differs from the aforementioned cluster in the value of the C_9δ torsion. Three clusters were identical between Schemes I, II and III. The remaining three clusters from Scheme I were replaced by two new clusters in Scheme II and only one new cluster in Scheme III.

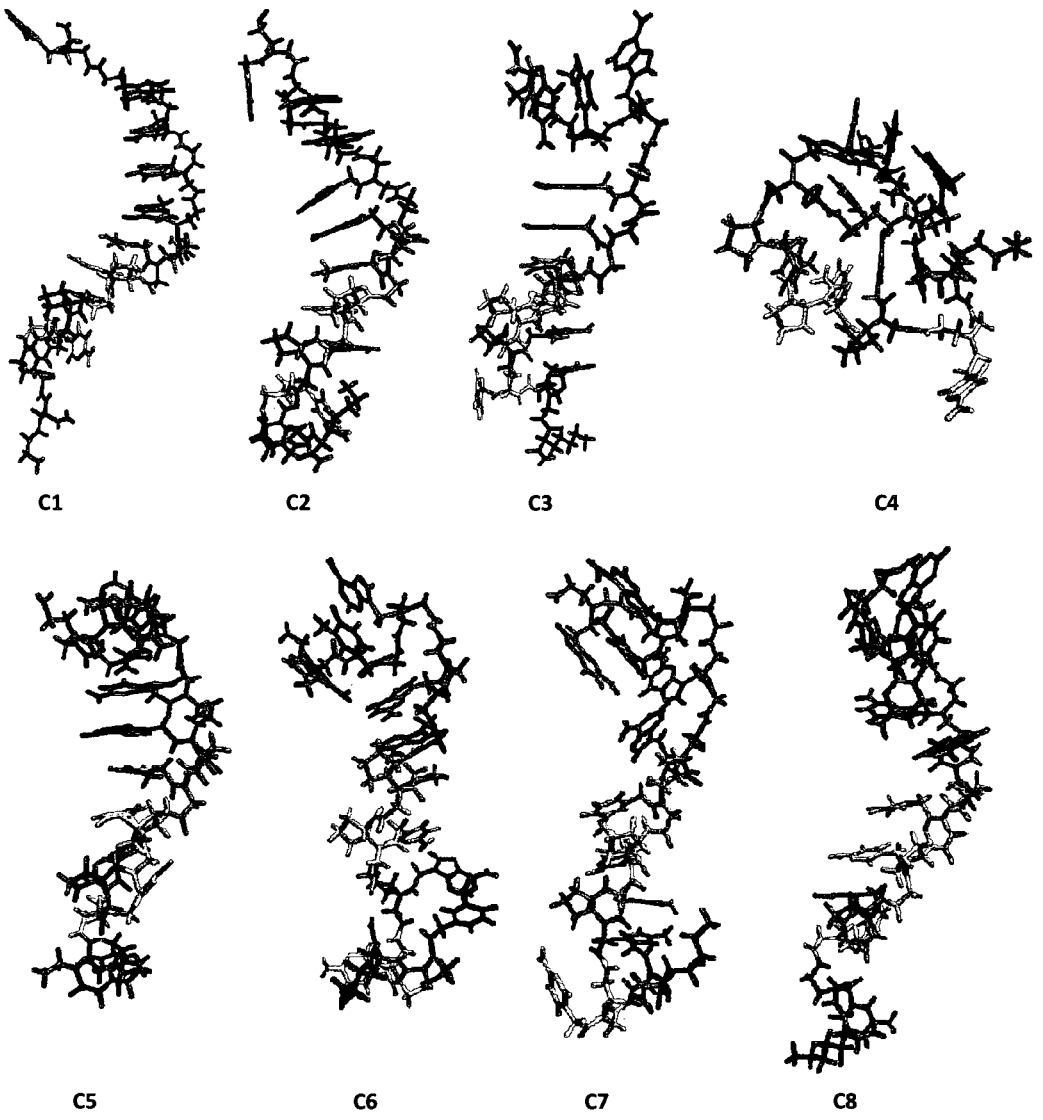

Figure 64. Representative structures of the most populated clusters from the MD simulation of *cp*PNA**22**. Residues are colored by name: **APN**-blue, **CPN**-orange, **TPN**-green, **TPC**- light blue, **CPC**–pink, **APC**-light brown, **GPN**-red and **K**-purple. The structures are positioned with **GPN1** on top and **K11** on bottom (see Table 17 for the sequence)

Table 87. Average torsion angles for each cluster from the *cp*PNA**22** simulation

Cluster	No. of members	%	T2		A3			G4		G5			A6		C7			A8		C9		
			β	δ	α	γ	ε	β	δ	α	γ	ε	β	δ	α	γ	ε	β	δ	α	γ	ε
Scheme I																						
1	3539	**35.4**	166	89	179	88	82	66	79	173	85	31	66	97	-118	79	1	68	107	175	92	-83
2	750	**7.5**	172	87	175	89	85	67	79	173	86	31	66	98	-120	80	3	69	107	169	93	53
3	2846	**28.5**	-162	97	169	88	-65	178	87	167	65	29	67	99	-119	78	3	68	102	178	100	107
4	1667	**16.7**	-159	-127	176	91	-22	180	98	168	-97	165	64	97	-121	75	-2	69	100	180	100	110
5	1198	**12.0**	-160	108	173	121	-95	172	85	166	81	15	76	98	-119	76	-1	69	107	174	100	-61
Scheme II																						
6	1781	**35.6**	-161	102	171	100	-75	176	86	164	81	24	70	99	-120	77	1	69	103	177	100	52
7	402	**8.0**	-163	152	178	85	-30	178	95	179	-91	61	66	97	-114	78	2	68	103	179	99	112
Scheme III																						
8	3018	**30.2**	-166	97	175	102	144	98	82	170	84	25	69	98	-119	79	1	69	108	177	100	-41

Table 88. Similar clusters identified from the different clustering Schemes for *cp*PNA**22**

Cluster	No. of members	%	Scheme I Cluster #[a]	Cluster I	No. of members	%	Scheme I Cluster #[a]
Scheme II				*Scheme III*			
A	1793	**35.9**	1	A	2209	**22.1**	1
B	380	**7.6**	2	B	3090	**30.9**	3
C	644	**12.9**	4	C	1683	**16.8**	4
Total number of clusters			5				4

[a]numbers correspond to identical cluster from Table 87

Clusters **1** and **2** from Scheme I contain structures with helices from residues 2 to 7. The matching clusters in Scheme II together with cluster **8** exhibit similar conformations. On examination of clusters **3** and **5**, we found that the structures are more conformationally disordered, but retain helical motifs. Scheme II located clusters (**6** and **7)** that contain helical conformations together with N- and C-terminal loops. The last significant cluster **4** from Scheme I is comprised of hairpin-like loop conformations that emerged at ~9 ns. This data seems to suggest that exhaustive sampling of the configurational space was not accomplished because the helical element did not remain for the entire simulation and disappeared around 9 ns. Thus, the simulations only captured a snapshot of the energy landscape.

Due to conformational constraints imposed by the three cyclopentanes, the number of clusters obtained (all three schemes) from the simulations of *cp*PNA**22** (T_6, C_7 and A_8) was lower than for *cp*PNA**17** (C_7), *cp*PNA**19** (T_6 and C_7), *cp*PNA**20** (T_6 and A_8) and *cp*PNA**21** (C_7 and A_8). For Schemes II and III *cp*PNA**22** had fewer clusters than *cp*PNA**15** while the number of conformers obtained by Scheme I was higher for *cp*PNA**22**. These results suggest that among these strands, *cp*PNA**22** exhibits the lowest flexibility; therefore, the changes in overall binding free energy will be due to a smaller entropic penalty. Clustering of *cp*PNA**18** (A_8) resulted in a lower number of groups in comparison to *cp*PNA**22**, which demonstrates that the dynamics of *cp*PNA**18** are more restricted. The simulations sampled both helical and hairpin conformations for *cp*PNA**22**. Similar conformations were observed for *cp*PNA**15** and *cp*PNA**18** that have modifications at T_6 and A_8, respectively. Thus, the *cp*PNA:DNA association process will be more entropically favorable for *cp*PNA**18** than for *cp*PNA**22**.

The effect of multiple cyclopentane modifications can be studied via analysis of the conformational properties of the 10-mer *cp*PNAs. Over the different *cp*PNA strands, a great variety of structural elements is present in different regions of the PNA. Indeed, *cp*PNA**19** has a significant number of structures with loops and turns. A coil conformation with a helical motif was also present. The *cp*PNA**22** strand with three rings retained an overall helical structure during the simulation. Thus, modifications at both T_6 and C_7 or only at A_8 induce a helical structure since all predominant structures obtained from clustering of *cp*PNA**18** and *cp*PNA**19** had a helical element. In addition, since the

number of strands having a helical character was higher for the PNA-T_8 sequence, it can also be inferred that the polythymine sequences show a larger tendency toward helical elements than the mixed base sequence.

To further investigate the influence of multiple modifications on the chain flexibility, the RMSF profile was analyzed (Table 89). The N-terminal residues 1, 2 and 3 showed greater flexibility than the other residues. This is partly consistent with the results received for *cp*PNA**17** (C_7), *cp*PNA**18** (A_8), *cp*PNA**19** (T_6 and C_7), *cp*PNA**20** (T_6 and A_8) and *cp*PNA**21** (C_7 and A_8), which also had higher N-terminal flexibility. The region comprising the modifications (residues 6, 7 and 8) showed considerably lower fluctuations. By comparing the RMSF values between *aeg*PNA**14** and *cp*PNA**22**, important changes in the overall strand flexibility upon modification are onserved. The resultant RMSF values of residues 6 to 10 are clearly lower than those observed in the unmodified PNA simulation. On the other hand, residues 1 to 5 show enhanced flexibility. Similar to *cp*PNA**15** (T_6), *cp*PNA**16** (G_4), *cp*PNA**17** (C_7) and *cp*PNA**21** (C_7 and A_8), but unlike *cp*PNA**18** (A_8), *cp*PNA**19** (T_6 and C_7) and *cp*PNA**20** (T_6 and A_8), the residues located on the N-terminal side of the modifications were among the most mobile residues. Yet, the C-terminal side residue almost always showed reduced fluctuations. As noted previously for *cp*PNA**16** and *cp*PNA**20**, the *cp*PNA**22** also has a moderately flexible C-terminal region (residues 9 and 10). In comparison to all the *cp*PNA strands, *cp*PNA**22** had the lowest flexibility, and since the extent of immobilization is a decisive determinant for DNA binding, the greater restriction of mobility will contribute favorably to the free energy of binding.

The reduced flexibilities of β, γ and ε in the modified residues 6, 7 and 8 are assumed to translate into their reduced mobility. On the other hand, the RMSF values for ε dihedral in residues 2 to 5 are larger in magnitude and demonstrate increased flexibility in comparison to the reference PNA. The γ torsion underwent increased fluctuations at G_1 and T_2 upon introduction of the constraints. For the β and δ dihedrals, a range of RMSF values was obtained. Normally, δ is restricted upon modification, but in this case β was restricted and δ experienced enhanced fluctuations in the terminal regions. Residues bearing the ring constraint had lower RMSFs for those dihedrals. A similar pattern of

dihedral flexibility was found in *cp*PNA**17** (C_7), *cp*PNA**19** (T_6 and C_7) and *cp*PNA**20** (T_6 and A_8), although less pronounced. In comparison to all other systems, the β dihedrals of *cp*PNA**22** have the lowest RMSF values. Fluctuations of the δ and ε dihedrals are greater relative to the ones seen for other *cp*PNAs, with the exception of *cp*PNA**20**, which had larger RMSFs.

Table 89. Average dihedrals (°) and RMSF (bold, °) of the simulated *cp*PNA**22**

Residue	α	β	γ	δ	ε	Total RMSF	Average RMSF
G1	—	—	30.7	50.4	56.9		
	—	—	**85**	**145**	**78**	**308**	**62**
T2	171.8	-174.5	50.8	97.3	43.0		
	22	**34**	**78**	**55**	**72**	**261**	**52**
A3	175.9	31.3	92.6	-12.3	-155.3		
	23	**50**	**19**	**93**	**119**	**304**	**61**
G4	-178.7	129.5	78.2	85.4	72.1		
	26	**57**	**26**	**15**	**23**	**147**	**29**
A5	169.7	88.8	45.5	80.6	55.8		
	21	**50**	**74**	**18**	**75**	**238**	**48**
T6	-138.3	67.1	76.6	97.8	6.1		
	30	**12**	**17**	**10**	**21**	**90**	**18**
C7	-119.4	67.6	77.5	99.5	0.9		
	18	**7**	**11**	**9**	**17**	**62**	**12**
A8	-114.0	68.5	75.0	104.4	-80.5		
	15	**7**	**10**	**11**	**11**	**54**	**11**
C9	-176.1	159.9	96.4	-56.1	11.6		
	26	**39**	**15**	**66**	**97**	**243**	**49**
T10	-177.2	-0.6	99.3	—	—		
	21	**58**	**22**	—	—	**101**	**34**

The average β values vary from 70 to 80° in the central residues (5 to 8), while the rest of the residues assume various values. Likewise, the γ dihedrals stay within the 80-90° range in most residues, but vary around 40° for residues 1, 2 and 5. For the δ dihedrals, values around 90° are seen, while the ε torsions are populated over a broad range. The major differences between the unmodified strand and *cp*PNA**22** are in the preferences of the β and ε dihedral values. In the case of the γ and δ torsions, changes in

the averages are apparent only for residues 2, 7, 8 and 3, 9, respectively. When comparing the two strands, the average values for the α and δ dihedrals differ between *cp*PNA**19**, *cp*PNA**20** and *cp*PNA**22**. The *cp*PNA strands with multiple modifications had the following approximate backbone torsion angles: α=170°, β=60° or 160°, γ=60-90°, δ=90-100° and ε=30-60°. The position of the modification had a significant effect on the average values of the β and ε torsions. The γ values were also shifted, though to a lesser extent. The effect these changes will have on the relative magnitudes of the enthalpies of binding cannot be evaluated. Nevertheless, the variations in the preferred torsion values indicate that structural factors will lead to differential binding enthalpies of the constrained PNAs.

In summary, the modification effect can be local to the site of the alteration (*cp*PNA**19** and *cp*PNA**20**) or very distant from the modification site. These varying results clearly indicate that the effect of the modification is position- and residue-specific. In addition, our calculations appear to be consistent with the experimental melting temperatures, in that there is a decrease in flexibility as the number of constraints is increased. In comparison to the polythymine sequences, the effects on the overall strand flexibility were less pronounced with the increased number of cyclopentanes for the MB_{10} systems; however, in these systems, more distinct changes in fluctuations (i.e. rigidification) were observed depending on the location of the modified residues. The two sequences differ in the torsional preferences for the β dihedral, as more strands in T_8 adopt *gauche* values than in MB_{10}. These observations suggest that the *cp*PNA-T_8 sequence, being more constrained, will bind to DNA/RNA with a binding entropy that is more favorable than the mixed base decamer. In addition, the restriction of the β dihedrals in the desirable range, will lead to stronger preferences for binding to RNA or DNA.

Comparison to DNA and RNA duplexes

We used the dihedral angles to evaluate the conformational similarity between the *cp*PNA single strands and DNA/RNA duplex structures. If we can find a cluster having

similar values to the native duplexes, its backbone conformation will not change appreciably during binding. In addition, this conformer is presumably a minimum on the PES of the strand. The lower the energy of this conformer, the more stable a duplex the strand should form with DNA or RNA. Essentially, the torsional strain of the duplex will have been reduced. This section describes a more direct comparison between calculation and experiment, in which computed backbone dihedral angles of the PNA single strands are compared with the corresponding values from NMR/X-ray studies of native DNA and RNA duplexes.[249,250]

Table 90 lists the backbone dihedral angles of PNA:DNA, DNA and RNA duplexes averaged over residues. The experimental structures of DNA and RNA and the theoretically obtained conformations of PNA share similar α values (~180°), though the modified residue has an average value of ~130°. A comparison of RNA dihedrals with *aeg*PNA shows two major differences. First, the β torsion is fixed in a gauche (60°) region for the RNA conformations; but, in the calculations, this angle can be gauche+ (+ 60°), gauche- (-60°), or trans (180°), prompting early speculation that PNA analogues that stabilize this torsion might exhibit an increased affinity towards RNA. Second, δ adopts values around ±90° in the calculations; but, prefers the *trans* (~180°) domain in the NMR and crystal structures. The similarity of the RNA conformation with *aeg*PNA is also evident; the values of the α, γ and ε dihedral angles for *aeg*PNA**1**, *aeg*PNA**12** and *aeg*PNA**14** are within 20° of the dihedrals of RNA. However, in the *aeg*PNA**3** strand with lysine, the majority of the ε torsions adopt values in the 100-160° range. In addition, no clusters with a dihedral angle pattern similar to the mean dihedral angles of the DNA duplex could be located.

Table 90. Dihedral angles in DNA and RNA duplexes[249,250]

Compound	α	β	γ	δ	ε
A-RNA[249]	174	77	80	-159	69
B-DNA[250]	-172	59	134	173	-99
PNA:DNA[54,177]	105	141	78	148	-

The *cp*PNA molecules were also compared to the DNA/RNA dihedrals. First, the polythymine sequence is analyzed, followed by the mixed base strands. Only the torsion angles of the most populated clusters were used for comparison. In all cases, the γ dihedral angles were similar to the corresponding values for the natural RNA duplexes and the PNA:DNA duplexes. The largest difference between the modified PNAs and DNA/RNA is in the value of δ. The calculations yield values of roughly 95°, while the NMR refinement of duplex structures yielded a value of 160-170° for this angle. Thus, the PNA and the DNA/RNA appear to differ significantly in this regard. However, in the PNA:DNA duplex, the δ value is closer to the calculated values. With respect to the dihedral angle value differences, modification(s) has the biggest effect on the β and ε torsions. Modification at T_4 (*cp*PNA**2**, *cp*PNA**4** and *cp*PNA**4a**) favors values near ~70° for the β dihedral, whereas values near 170° were observed in the *aeg*PNA structures. Conversely, our calculations on the *cp*PNA**5** (T_1) and *cp*PNA**6** (T_8) conformations indicate that the β dihedrals prefer values in the 160-180° range, which agrees with the PNA:DNA duplex. However, gauche+ and gauche- β dihedral angles are also observed for these strands. The dihedral angle ε displays a comparatively broader distribution and additionally accesses the dihedral space around 0°. The computed values for the *cp*PNA**2**, *cp*PNA**4a** and *cp*PNA**6** strands lie between the DNA and RNA values for ε, but are slightly closer to the RNA values. On the other hand, *cp*PNA**4** and *cp*PNA**5** display ε dihedral angle distributions that are similar to the corresponding values found in the DNA duplex structure. Interestingly, an evident shift of about 65° is observed when the α dihedral values for the modified residues of all *cp*PNA strands are compared. With the exception of *cp*PNA**5** and *cp*PNA**6** structures, which bear a modification at either of the termini, the α angles in the modified residues adopt values of ~100° and significantly deviate from the values in the RNA/DNA complex, but are in excellent agreement with values of the PNA:DNA complex.

The detailed analyses of the monosubstituted *cp*PNA decamers shows that the vast majority of the residues in the data set have their torsion angle β in the gauche orientation, which matches with the RNA duplex dihedrals. The ranges of the α, γ and δ dihedrals are similar to those observed for the polythymine strands. The configurational

spaces accessed upon modification undergo a pronounced reduction. However, the distributions of the ε angles are different between the various *cp*PNA strands. The calculated dihedral values for *cp*PNA**15** (T_6), *cp*PNA**16** (G_4) and *cp*PNA**18** (A_8) match most of the trends seen in the RNA duplexes with the exception of a few residues. The *cp*PNA**13** (T_6, no Lys) and *cp*PNA**15a** (T_6, *D*-Lys) display significantly broader distributions and overlap with the DNA and RNA values in several residues. Remarkably, *cp*PNA**17** (C_7) can access the ε dihedral region (~-100°), characteristic of the DNA configuration in the duplex.

Considering multiple-site modifications, all nucleotides in the *cp*PNA chains had the same combination of the three torsion angles α, γ and δ. However, restrictions of the conformational space upon modification appeared for both the β and ε torsions in *cp*PNA**8** (T_1 and T_8), *cp*PNA**11** (T_1–T_8) and *cp*PNA**21** (C_7 and A_8). In these *cp*PNA chains, the residues had β and ε torsions with values near those observed in the RNA duplex. In the remaining six strands with multiple modifications (polyT_8 - *cp*PNA**7**, *cp*PNA**9**, *cp*PNA**10** and MB_{10} - *cp*PNA**19**, *cp*PNA**20**, *cp*PNA**22**), both torsions were not able to assume these values at the same time. The β torsion shows two ranges of values and additionally samples the dihedral region around 170° in all cases. The ε backbone dihedral angles in *cp*PNA**7** (T_1, T_4 and T_8) and *cp*PNA**19** (T_6 and C_7) were similar to the corresponding values for the RNA duplexes, while in *cp*PNA**9** (T_4 and T_8), *cp*PNA**10** (T_1 and T_4), *cp*PNA**20** (T_6 and A_8) and *cp*PNA**22** (C_7 and A_8) the corresponding values of certain clusters are close to the values found in the RNA or DNA complex. Furthermore, clusters with the correct values for RNA duplex formation, for the torsion angles β and ε automatically have α and γ torsion angles in the correct range for binding. However, the δ dihedrals still had a wide range of values and differed from the required ones. Although a strand can have β or ε values similar to those of RNA, the combination of both being near the RNA values seems to be the decisive factor in assuming a correct conformation.

A PNA strand, which is able to adopt the dihedral values of native duplexes, will be able to bind stronger to DNA or DNA. As shown by MD simulations, this combination of angle values can be realized in a single-stranded *cp*PNA chain to some extent. Several *cp*PNA structures had restricted flexibility of the β and γ torsion angles.

With the exception of *cp*PNA**5**, *cp*PNA**6** and *cp*PNA**17**, the *cp*PNA strands favored β dihedral angles around 65° (between 60° and 70°), which seems to be an important criteria for RNA and DNA specificity. The present calculations suggest that *cp*PNA can pre-organize the strand into conformations that are better able to bind RNA or DNA. In particular, polythymine *cp*PNA**4a**, *cp*PNA**7**, *cp*PNA**8** and mixed base *cp*PNA**15**, *cp*PNA**16**, *cp*PNA**18**, *cp*PNA**19** and *cp*PNA**21** have β and γ dihedrals in the preferred range for RNA binding. On the other hand, *cp*PNA**4**, *cp*PNA**5** and *cp*PNA**17** have the corresponding values suitable for DNA binding. Thus, the rigidification of the PNA backbone by a cyclopentyl constraint may be suitable for imparting DNA/RNA discrimination properties. In addition, the cyclopentane modification should strengthen binding to RNA and experimental studies will need to be conducted to test this hypothesis. The cyclopentane modification also gave α and γ values similar to the PNA:DNA duplex. This outcome is not surprising since the cyclopentane modification was designed to limit the strand to the dihedrals observed in the PNA:DNA duplex and probably helps contribute to the strong binding to DNA that is seen for this modification. The attachment of lysine (*L* and *D*) had an effect only on the ε dihedral in the poly-T strands. In the future, computational results for the single strand could guide the design of a variety of modifications, which could improve the binding properties of PNA. In particular, designing modifications that reproduce the dihedrals of the native DNA and RNA duplexes instead of the PNA:DNA duplex could lead to improvements in binding strength. Simulations on the PNA:DNA/RNA duplex systems will also be needed to help further our understanding of the affinity and stability of these systems.

CHAPTER V

SUMMARY

Molecular dynamics simulations are a powerful theoretical tool for investigation of biological molecular systems. Simulations allow for the observation of structure at the level of individual molecules instead of the ensemble averages typically provided by experiments. Computational studies can also provide atomic level detail concerning specific interactions that may not be readily available from experimental studies of rapidly interconverting ensembles. This insight enhances our understanding of the mechanistic details of conformational behavior and structure.

The first problem of study in this thesis concerned the development of force field parameters for PNA. The MD simulation uses parameterized empirical energy functions in its calculations. Development of good parameters sets is essential for the success of the MD simulation. Therefore, in this work we have also focused on developing new CHARMM force field parameters for cyclopentane modified PNA, namely partial atomic charges and equilibrium bond-lengths, angles and dihedrals.

In the second problem, MD simulations were performed to study the conformation of *aeg*PNA and a number of cyclopentyl modified structures and to reveal to what extent PNA structure, flexibility and dynamics are affected by the modification. It is generally assumed that pre-organizing a flexible molecule in a way that corresponds to the conformations they adopt upon binding to DNA/RNA, will lead to increased binding because the rigidified molecule will benefit from lower entropic penalty. This common assumption, however, ignores the important fact that pre-organization has both enthalpic and entropic components and obtaining a more favorable enthalpy of binding can be achieved by adopting conformations with similar torsional angles as in the target DNA/RNA. Therefore, we undertook a systematic analysis of the conformational properties of the single stranded *cp*PNA molecules bearing single and multiple constraints in the backbone. In particular, we set out to identify and compare the preferred conformations of these molecules as well as to map out the internal degrees of

freedom that govern their conformational flexibility. Clustering analysis was used for identifying the structural similarity between different conformers of the molecules and examining the cluster conformations were examined in detail.

We have shown that the dominant configurations in solution can be successfully captured using clustering. Introduction of a single modification significantly affects the flexibility of the strand as shown by both RMSF values and cluster analysis. Results also indicate that the position of the modification plays an important role in determining the flexibility. Strong effects (rigidification) are also observed as the number of cyclopentyl groups increases. Overall, the results indicate the cyclopentane restricted the conformational freedom of PNA. Thus, the rigidified molecule will exhibit a higher affinity toward DNA or RNA due to lower entropic penalty.

The *cp*PNA was designed to maximize pre-organization into a helical conformation. Simulations of singly modified PNAs predict that modifications are at specific positions in the strand can lead to a helical-like conformer. In addition, certain *cp*PNA strands with multiple modifications were more prone toward helical elements. Thus, the additional moiety in PNA is essential for helix stability. Our simulations support the general picture of a cyclopentane modification as a determinant of the pre-organization and are in line with the experimental data that constrained PNA analogues form stable duplexes with DNA/RNA. When comparing PNA dihedral angle values calculated using the MD to the experimentally determined values from the RNA duplex, better agreement was found for the strands carrying multiple rings than for the singly modified strands, further supporting the fact that the modification alters the dihedral potential energy surface and does not just reduce flexibility.

These results suggest that conformationally constrained *cp*PNAs have a high potential for selective binding to DNA since they exhibit profound effects on the backbone geometry. Future directions include the extension of this study to increase sampling of the PES and to perform double-stranded simulations. An additional development would be to relate the torsion angle values to conformational features, and seek correlations between the values of different torsion angles. The information about residue and dihedral flexibility can then be employed to study the effect of nucleobase

composition on the stability and propose improved modifications which will hopefully bind even stronger to DNA or RNA.

REFERENCES

(1) Watson, J. D.; Crick, F. H. *Nature* **1953**, 737-738.

(2) Crick, F. H. *Nature* **1970**, 561-563.

(3) Kool, E. T. *Chem. Rev.* **1997**, *5,* 1473-1488.

(4) Wang, A.; Quigley, G.; Kolpak, F.; Crawford, J.; van Boom, J.; van der Marel, G.; Rich, A. *Nature* **1979**, *5740,* 680.

(5) Arnott, S. ***Polynucleotide secondary structures****: a historical perspective;* Oxford University Press: 1999; pp 1-38.

(6) Koziolkiewicz, M.; Gendaszewska, E.; Maszewska, M.; Stein, C. A.; Stec, W. J. *Blood* **2001**, *4,* 995-1002.

(7) Vaerman, J. L.; Moureau, P.; Deldime, F.; Lewalle, P.; Lammineur, C.; Morschhauser, F.; Martiat, P. *Blood* **1997**, *1,* 331-339.

(8) Eckstein, F. *Acc. Chem. Res.* **1979**, *6,* 204-210.

(9) Stein, C. A.; Subasinghe, C.; Shinozuka, K.; Cohen, J. S. *Nucleic Acids Res.* **1988**, *8,* 3209-3221.

(10) Campbell, J. M.; Bacon, T. A.; Wickstrom, E. *J. Biochem. Biophys. Methods* **1990**, *3,* 259-267.

(11) Crooke, R. M. *Anticancer Drug Des.* **1991**, *6,* 609-646.

(12) Ts'o, P. O.; Aurelian, L.; Chang, E.; Miller, P. S. *Ann. N. Y. Acad. Sci.* **1992**, 159-177.

(13) Marugg, J. E.; de Vroom, E.; Dreef, C. E.; Tromp, M.; van der Marel, G. A.;vanBoom, J. H. *Nucleic Acids Res.* **1986**, *5,* 2171-2185.

(14) *Jager, A.; Levy, M. J.; Hecht, S. M. Biochemistry* ***1988****, 19, 7237-7246.*

(15) Letsinger, R. L.; Alul, R. A.; Farooqui, F.; Gryaznov, S. M.; Kinstler, O. *Nucleic Acids Symp. Ser.* **1991**, *24,* 75-78.

(16) Heidenreich, O.; Gryaznov, S.; Nerenberg, M. *Nucleic Acids Res.* **1997**, *4,* 776-780.

(17) Ishikawa, T.; Yoneda, F.; Tanaka, K.; Fuji, K. *Bioorg. Med. Chem. Lett.* **1991**, *10,* 523-526.

(18) Coleman, R. S.; Siedlecki, J. M. *Tetrahedron Letters* **1991**, *26*, 3033-3034.

(19) Anand, N. N.; Brown, D. M.; Salisbury, S. A. *Nucleic Acids Res.* **1987**, *20*, 8167-8176.

(20) Eritja, R.; Horowitz, D. M.; Walker, P. A.; Ziehler-Martin, J. P.; Boosalis, M. S.; Goodman, M. F.; Itakura, K.; Kaplan, B. E. *Nucleic Acids Res.* **1986**, *20*, 8135-8153.

(21) Smith, C. A.; Xu, Y. Z.; Swann, P. F. *Carcinogenesis* **1990**, *5*, 811-816.

(22) Komatsu, H.; Kim, S.; Sakabe, I.; Ichikawa, T.; Nakai, M.; Takaku, H. *Bioorg. Med. Chem. Lett.* **1992**, *6*, 565-570.

(23) Alderfer, J. L.; Tazawa, I.; Tazawa, S.; Ts'o, P. *Biochemistry* **1974**, *8*, 1615-1622.

(24) Perreault, J. P.; Wu, T. F.; Cousineau, B.; Ogilvie, K. K.; Cedergren, R. *Nature* **1990**, *6266*, 565-567.

(25) Iribarren, A. M.; Sproat, B. S.; Neuner, P.; Sulston, I.; Ryder, U.; Lamond, A. I. *Proc. Natl. Acad. Sci. U. S. A.* **1990**, *19*, 7747-7751.

(26) Freier, S. M.; Altmann, K. H. *Nucleic Acids Res.* **1997**, *22*, 4429-4443.

(27) Cook, P. D. Chapter 31. Second Generation Antisense Oligonucleotides: 2'-Modifications. In *Annual Reports in Medicinal Chemistry;* James A. Bristol, Ed.; Academic Press: 1998; Vol. Volume 33, pp 313-325.

(28) Kurreck, J.; Wyszko, E.; Gillen, C.; Erdmann, V. A. *Nucleic Acids Res.* **2002**, *9*, 1911-1918.

(29) Braasch, D. A.; Corey, D. R. *Chem. Biol.* **2001**, *1*, 1-7.

(30) Koshkin, A. A.; Singh, S. K.; Nielsen, P.; Rajwanshi, V. K.; Kumar, R.; Meldgaard, M.; Olsen, C. E.; Wengel, J. *Tetrahedron* **1998**, *14*, 3607-3630.

(31) Petersen, M.; Nielsen, C. B.; Nielsen, K. E.; Jensen, G. A.; Bondensgaard, K.; Singh, S. K.; Rajwanshi, V. K.; Koshkin, A. A.; Dahl, B. M.; Wengel, J.; Jacobsen, J. P. *J. Mol. Recognit.* **2000**, *1*, 44-53.

(32) Petersen, M.; Wengel, J. *Trends Biotechnol.* **2003**, *2*, 74-81.

(33) Swayze, E. E.; Siwkowski, A. M.; Wancewicz, E. V.; Migawa, M. T.; Wyrzykiewicz, T. K.; Hung, G.; Monia, B. P.; Bennett, C. F. *Nucleic Acids Res.* **2007**, *2*, 687-700.

(34) Pitha, J. *Unusual Properties of New Polymers* **1983**, 1-16.

(35) Nielsen, P. E.; Egholm, M.; Berg, R. H.; Buchardt, O. *Science* **1991**, *5037*, 1497-1500.

(36) Egholm, M.; Buchardt, O.; Nielsen, P. E.; Berg, R. H. *J. Am. Chem. Soc.* **1992**, *5*, 1895-1897.

(37) Nielsen, P. E. In *Peptide Nucleic Acids: Protocols and Applications, Second Edition.* Section Title: Biochemical Genetics; 2004; pp 318.

(38) Egholm, M.; Buchardt, O.; Christensen, L.; Behrens, C.; Freier, S. M.; Driver, D. A.; Berg, R. H.; Kim, S. K.; Norden, B.; Nielsen, P. E. *Nature* **1993**, *6446*, 566-568.

(39) Gaylord, B. S.; Heeger, A. J.; Bazan, G. C. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *17*, 10954-10957.

(40) Demidov, V. V.; Potaman, V. N.; Frank-Kamenetskii, M. D.; Egholm, M.; Buchard, O.; Sonnichsen, S. H.; Nielsen, P. E. *Biochem. Pharmacol.* **1994**, *6*, 1310-1313.

(41) Breipohl, G.; Knolle, J.; Langner, D.; O'Malley, G.; Uhlmann, E. *Bioorg. Med. Chem. Lett.* **1996**, *6*, 665-670.

(42) Giesen, U.; Kleider, W.; Berding, C.; Geiger, A.; Orum, H.; Nielsen, P. E. *Nucleic Acids Res.* **1998**, *21*, 5004-5006.

(43) Egholm, M.; Christensen, L.; Dueholm, K. L.; Buchardt, O.; Coull, J.; Nielsen, P. E. *Nucleic Acids Res.* **1995**, *2*, 217-222.

(44) Wittung, P.; Nielsen, P.; Norden, B. *J. Am. Chem. Soc.* **1997**, *13*, 3189-3190.

(45) Datta, B.; Bier, M. E.; Roy, S.; Armitage, B. A. *J. Am. Chem. Soc.* **2005**, *12*, 4199-4207.

(46) Amato, J.; Gabelica, V.; Borbone, N.; Rosu, F.; De Pauw, E.; Oliviero, G.; Piccialli, G.; Mayol, L. *Nucleic Acids Symp. Ser.* **2008**, *52*, 167-168.

(47) Datta, B.; Schmitt, C.; Armitage, B. A. *J. Am. Chem. Soc.* **2003**, *14*, 4111-4118.

(48) Eriksson, M.; Nielsen, P. E. *Q. Rev. Biophys.* **1996**, *4*, 369-394, 3 plates.

(49) Tomac, S.; Sarkar, M.; Ratilainen, T.; Wittung, P.; Nielsen, P. E.; Norden, B.; Graeslund, A. *J. Am. Chem. Soc.* **1996**, *24*, 5544-5552.

(50) Wittung, P.; Nielsen, P.; Norden, B. *J. Am. Chem. Soc.* **1996**, *30*, 7049-7054.

(51) Brown, S. C.; Thomson, S. A.; Veal, J. M.; Davis, D. G. *Science* **1994**, *5173*, 777-777.

(52) Leijon, M.; Graeslund, A.; Nielsen, P. E.; Buchardt, O.; Norden, B.; Kristensen, S. M.; Eriksson, M. *Biochemistry* **1994**, *33*, 9820-9825.

(53) He, W.; Hatcher, E.; Balaeff, A.; Beratan, D. N.; Gil, R. R.; Madrid, M.; Achim, C. *J. Am. Chem. Soc.* **2008**, *40*, 13264-13273.

(54) Betts, L.; Josey, J. A.; Veal, J. M.; Jordan, S. R. *Science* **1995**, *5243*, 1838-1841.

(55) Rasmussen, H.; Kastrup, J. S.; Nielsen, J. N.; Nielsen, J. M.; Nielsen, P. E. *Nat. Struct. Biol.* **1997**, *2*, 98-101.

(56) Wittung, P.; Eriksson, M.; Lyng, R.; Nielsen, P. E.; Norden, B. *J. Am. Chem. Soc.* **1995**, *41*, 10167-10173.

(57) Sen, S.; Nilsson, L. *J. Am. Chem. Soc.* **2001**, *30*, 7414-7422.

(58) Shields, G. C.; Laughton, C. A.; Orozco, M. *J. Am. Chem. Soc.* **1998**, *24*, 5895-5904.

(59) Sen, S.; Nilsson, L. *J. Am. Chem. Soc.* **1998**, *4*, 619-631.

(60) Soliva, R.; Sherer, E.; Luque, F. J.; Laughton, C. A.; Orozco, M. *J. Am. Chem. Soc.* **2000**, *25*, 5997-6008.

(61) Chen, S. M.; Mohan, V.; Kiely, J. S.; Griffith, M. C.; Griffey, R. H. *Tetrahedron Lett.* **1994**, *29*, 5105-5108.

(62) Hyrup, B.; Nielsen, P. E. *Bioorg. Med. Chem.* **1996**, *1*, 5-23.

(63) Buchardt, O.; Egholm, M.; Berg, R. H.; Nielsen, P. E. *Trends Biotechnol.* **1993**, *9*, 384-386.

(64) Ray, A.; Norden, B. *FASEB J.* **2000**, *9*, 1041-1060.

(65) Nielsen, P. E. *Pharmacol.Toxicol.* **2000**, *1*, 3-7.

(66) Eldrup, A. B.; Nielsen, P. E. *Adv. Amino Acid Mimetics Peptidomimetics* **1999**, 221-245.

(67) Lundin, K. E.; Good, L.; Stroemberg, R.; Graeslund, A.; Smith, C. I. E. *Adv. Genet.* **2006**, 1-51.

(68) Tyler-McMahon, B. M.; Stewart, J. A.; Jackson, J.; Bitner, M. D.; Fauq, A.; McCormick, D. J.; Richelson, E. *Biochem. Pharmacol.* **2001**, *7*, 929-932.

(69) Petraccone, L.; Barone, G.; Giancola, C. *Curr. Med. Chem. Anticancer Agents* **2005**, *5*, 463-475.

(70) Lech-Maranda, E.; Korycka, A.; Robak, T. *Mini Rev. Med. Chem.* **2006**, *5*, 575-581.

(71) Ishibashi, M.; Harada, K.; Horiya, S.; Matsufuji, S. Patent 2007004341, 2007.

(72) Ivanova, G.; Arzumanov, A. A.; Turner, J. J.; Reigadas, S.; Toulme, J. J.; Brown, D. E.; Lever, A. M.; Gait, M. J. *Ann. N. Y. Acad. Sci.* **2006**, 103-115.

(73) Chen, Z. Y.; Cheng, A. C.; Wang, M. S.; Xu, D. W.; Zeng, W.; Li, Z. *Acta Pharmacol. Sin.* **2007**, *10*, 1652-1658.

(74) Nielsen, P. E. *Expert Opin. Investig. Drugs* **2001**, *2*, 331-341.

(75) Eriksson, M.; Nielsen, P. E.; Good, L. *J. Biol. Chem.* **2002**, *9*, 7144-7147.

(76) Tan, X. X.; Actor, J. K.; Chen, Y. *Antimicrob. Agents Chemother.* **2005**, *8*, 3203-3207.

(77) Stock, R. P.; Olvera, A.; Sanchez, R.; Saralegui, A.; Scarfi, S.; Sanchez-Lopez, R.; Ramos, M. A.; Boffa, L. C.; Benatti, U.; Alagon, A. *Nat. Biotechnol.* **2001**, *3*, 231-234.

(78) Huang, X.; Pan, J.; An, X.; Zhuge, H. *Zhonghua Weishengwuxue He Mianyixue Zazh* **2007**, *4*, 358-362.

(79) Knudsen, H.; Nielsen, P. E. *Nucleic Acids Res.* **1996**, *3*, 494-500.

(80) Hanvey, J. C.; Peffer, N. J.; Bisi, J. E.; Thomson, S. A.; Cadilla, R.; Josey, J. A.; Ricca, D. J.; Hassman, C. F.; Bonham, M. A.; Au, K. G. *Science* **1992**, *5087*, 1481-1485.

(81) Nielsen, P. E. *Antisense Drug Technol. 2nd Ed.* **2008**, 507-518.

(82) Corey, D. R. *Trends Biotechnol.* **1997**, *6*, 224-229.

(83) Uhlmann, E.; Peyman, A.; Breipohl, G.; Will, D. W. *Angew.Chem., Int.Ed.* **1998**, *20*, 2796-2823.

(84) Demidov, V. V.; Yavnilovich, M. V.; Belotserkovskii, B. P.; Frank-Kamenetskii, M. D.; Nielsen, P. E. *Proc. Natl. Acad. Sci. U. S. A.* **1995**, *7*, 2637-2641.

(85) Bentin, T.; Nielsen, P. E. *Biochemistry* **1996**, *27*, 8863-8869.

(86) Mollegaard, N. E.; Buchardt, O.; Egholm, M.; Nielsen, P. E. *Proc. Natl. Acad. Sci. U. S. A.* **1994**, *9*, 3892-3895.

(87) Kristensen, V. N.; Kelefiotis, D.; Kristensen, T.; Borresen-Dale, A. L. *BioTechniques* **2001**, *2*, 318-22, 324, 326 passim.

(88) Ferguson, J. A.; Boles, T. C.; Adams, C. P.; Walt, D. R. *Nat. Biotechnol.* **1996**, *13*, 1681-1684.

(89) Patolsky, F.; Lichtenstein, A.; Willner, I. *J. Am. Chem. Soc.* **2001**, *22*, 5194-5205.

(90) Arlinghaus, H. F.; Kwoka, M. N.; Jacobson, K. B. *Anal. Chem.* **1997**, *18*, 3747-3753.

(91) Feriotto, G.; Corradini, R.; Sforza, S.; Bianchi, N.; Mischiati, C.; Marchelli, R.; Gambari, R. *Lab. Invest.* **2001**, *10*, 1415-1427.

(92) Corradini, R.; Feriotto, G.; Sforza, S.; Marchelli, R.; Gambari, R. *J. Mol. Recognit.* **2004**, *1*, 76-84.

(93) Chun, K. H.; Hwang, H. J. Patent 2004033476, 2004.

(94) Al Attar, H. A.; Norden, J.; O'Brien, S.; Monkman, A. P. *Biosens. Bioelectron.* **2008**, *10*, 1466-1472.

(95) Joung, H. A.; Lee, N. R.; Lee, S. K.; Ahn, J.; Shin, Y. B.; Choi, H. S.; Lee, C. S.; Kim, S.; Kim, M. G. *Anal. Chim. Acta* **2008**, *2*, 168-173.

(96) Pellestor, F.; Paulasova, P.; Hamamah, S. *Curr. Pharm. Des.* **2008**, *24*, 2439-2444.

(97) Mateo-Marti, E.; Briones, C.; Pradier, C. M.; Martin-Gago, J. A. *Biosens. Bioelectron.* **2007**, *9-10*, 1926-1932.

(98) Armitage, B.; Ly, D.; Koch, T.; Frydenlund, H.; Orum, H.; Schuster, G. B. *Biochemistry* **1998**, *26*, 9417-9425.

(99) Kuhn, H.; Demidov, V. V.; Gildea, B. D.; Fiandaca, M. J.; Coull, J. C.; Frank-Kamenetskii, M. D. *Antisense Nucleic Acid Drug Dev.* **2001**, *4*, 265-270.

(100) Kuhn, H.; Demidov, V. V.; Coull, J. M.; Fiandaca, M. J.; Gildea, B. D.; Frank-Kamenetskii, M. D. *J. Am. Chem. Soc.* **2002**, *6*, 1097-1103.

(101) Petersen, K.; Vogel, U.; Rockenbauer, E.; Nielsen, K. V.; Kolvraa, S.; Bolund, L.; Nexo, B. *Mol. Cell. Probes* **2004**, *2*, 117-122.

(102) Englund, E. A.; Appella, D. H. *Org. Lett.* **2005**, *16*, 3465-3467.

(103) Thurley, S.; Roglin, L.; Seitz, O. *J. Am. Chem. Soc.* **2007**, *42*, 12693-12695.

(104) Li, X.; Morgenroth, E.; Raskin, L. *Appl. Environ. Microbiol.* **2008**, *23*, 7297-7305.

(105) Orum, H.; Nielsen, P. E.; Egholm, M.; Berg, R. H.; Buchardt, O.; Stanley, C. *Nucleic Acids Res.* **1993**, *23*, 5332-5336.

(106) Schuermann, M.; Behn, M. *Methods Mol. Biol.* **2002**, 165-179.

(107) Galbiati, S.; Restagno, G.; Foglieni, B.; Bonalumi, S.; Travi, M.; Piga, A.; Sbaiz, L.; Chiari, M.; Damin, F.; Smid, M.; Valsecchi, L.; Pasi, F.; Ferrari, A.; Ferrari, M.; Cremonesi, L. *Ann. N. Y. Acad. Sci.* **2006**, 137-143.

(108) Miyake, M.; Sugano, K.; Kawashima, K.; Ichikawa, H.; Hirabayashi, K.; Kodama, T.; Fujimoto, H.; Kakizoe, T.; Kanai, Y.; Fujimoto, K.; Hirao, Y. *Biochem. Biophys. Res. Commun.* **2007**, *4*, 865-871.

(109) Nielsen, P. E.; Egholm, M.; Berg, R. H.; Buchardt, O. *Nucleic Acids Res.* **1993**, *2*, 197-200.

(110) Peffer, N. J.; Hanvey, J. C.; Bisi, J. E.; Thomson, S. A.; Hassman, C. F.; Noble, S. A.; Babiss, L. E. *Proc. Natl. Acad. Sci. U. S. A.* **1993**, *22*, 10648-10652.

(111) Kurakin, A.; Larsen, H. J.; Nielsen, P. E. *Chem. Biol.* **1998**, *2*, 81-89.

(112) Wittung, P.; Kajanus, J.; Edwards, K.; Haaima, G.; Nielsen, P. E.; Norden, B.; Malmstrom, B. G. *FEBS Lett.* **1995**, *3*, 27-29.

(113) Rebuffat, A. G.; Nawrocki, A. R.; Nielsen, P. E.; Bernasconi, A. G.; Bernal-Mendez, E.; Frey, B. M.; Frey, F. J. *FASEB J.* **2002**, *11*, 1426-1428.

(114) Muratovska, A.; Lightowlers, R. N.; Taylor, R. W.; Turnbull, D. M.; Smith, R. A.; Wilce, J. A.; Martin, S. W.; Murphy, M. P. *Nucleic Acids Res.* **2001**, *9*, 1852-1863.

(115) Koppelhus, U.; Awasthi, S. K.; Zachar, V.; Holst, H. U.; Ebbesen, P.; Nielsen, P. E. *Antisense Nucleic Acid Drug Dev.* **2002**, *2*, 51-63.

(116) Ljungstrom, T.; Knudsen, H.; Nielsen, P. E. *Bioconjug. Chem.* **1999**, *6*, 965-972.

(117) Abes, S.; Ivanova, G. D.; Abes, R.; Arzumanov, A. A.; Williams, D.; Owen, D.; Lebleu, B.; Gait, M. J. *Methods Mol. Biol.* **2009**, 85-99.

(118) Braasch, D. A.; Corey, D. R. *Methods Mol. Biol.* **2002**, 211-223.

(119) Yuan, X.; Ma, Z.; Zhou, W.; Niidome, T.; Alber, S.; Huang, L.; Watkins, S.; Li, S. *Biochem. Biophys. Res. Commun.* **2003**, *1*, 6-11.

(120) Basu, S.; Wickstrom, E. *Bioconjug. Chem.* **1997**, *4*, 481-488.

(121) Liang, K. W.; Hoffman, E. P.; Huang, L. *Mol. Ther.* **2000**, *3*, 236-243.

(122) Scheller, A.; Oehlke, J.; Wiesner, B.; Dathe, M.; Krause, E.; Beyermann, M.; Melzig, M.; Bienert, M. *J. Pept. Sci.* **1999**, *4*, 185-194.

(123) Haaima, G.; Hansen, H. F.; Christensen, L.; Dahl, O.; Nielsen, P. E. *Nucleic Acids Res.* **1997**, *22*, 4639-4643.

(124) Lohse, J.; Dahl, O.; Nielsen, P. E. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *21*, 11804-11808.

(125) Christensen, L.; Hansen, H. F.; Koch, T.; Nielsen, P. E. *Nucleic Acids Res.* **1998**, *11*, 2735-2739.

(126) Eldrup, A. B.; Dahl, O.; Nielsen, P. E. *J. Am. Chem. Soc.* **1997**, *45*, 11116-11117.

(127) Rajeev, K. G.; Maier, M. A.; Lesnik, E. A.; Manoharan, M. *Org. Lett.* **2002**, *25*, 4395-4398.

(128) Ausin, C.; Ortega, J. A.; Robles, J.; Grandas, A.; Pedroso, E. *Org. Lett.* **2002**, *23*, 4073-4075.

(129) Gangamani, B. P.; Kumar, V. A.; Ganesh, K. N. *Chem.Commun.* **1997**, *19*, 1913-1914.

(130) Robertson, K. L.; Yu, L.; Armitage, B. A.; Lopez, A. J.; Peteanu, L. A. *Biochemistry* **2006**, *19*, 6066-6074.

(131) Socher, E.; Jarikote, D. V.; Knoll, A.; Roglin, L.; Burmeister, J.; Seitz, O. *Anal. Biochem.* **2008**, *2*, 318-330.

(132) Kohler, O.; Jarikote, D. V.; Seitz, O. *Chembiochem* **2005**, *1*, 69-77.

(133) Schutz, R.; Cantin, M.; Roberts, C.; Greiner, B.; Uhlmann, E.; Leumann, C. *Angew.Chem., Int.Ed.* **2000**, *7*, 1250-1253.

(134) Hollenstein, M.; Leumann, C. J. *Org. Lett.* **2003**, *11*, 1987-1990.

(135) Ferrer, E.; Shevchenko, A.; Eritja, R. *Bioorg. Med. Chem.* **2000**, *2*, 291-297.

(136) Petersen, K. H.; Buchardt, O.; Nielsen, P. E. *Bioorg. Med. Chem. Lett.* **1996**, *7*, 793-796.

(137) Nielsen, P. E.; Haaima, G. *Chem. Soc. Rev.* **1997**, 73-78.

(138) Ciapetti, P.; Mann, A.; Schoenfelder, A.; Taddei, M.; Trifilieff, E.; Canet, I.; Canet, J. *Int. J. Pept. Res. Ther.* **1997**, *4*, 341-349.

(139) Kuwahara, M.; Arimitsu, M.; Sisido, M. *J. Am. Chem. Soc.* **1999**, *1*, 256-257.

(140) Krotz, A. H.; Larsen, S.; Buchardt, O.; Eriksson, M.; Nielsen, P. E. *Bioorg. Med. Chem.* **1998**, *11*, 1983-1992.

(141) Garner, P.; Dey, S.; Huang, Y. *J. Am. Chem. Soc.* **2000**, *10*, 2405-2406.

(142) Barawkar, D. A.; Bruice, T. C. *J. Am. Chem. Soc.* **1999**, *44*, 10418-10419.

(143) Zhou, P.; Wang, M.; Du, L.; Fisher, G. W.; Waggoner, A.; Ly, D. H. *J. Am. Chem. Soc.* **2003**, *23*, 6878-6879.

(144) Govindaraju, T.; Kumar, V. A. *Chem.Commun.* **2005**, *4*, 495-497.

(145) Govindaraju, T.; Kumar, V. A. *Tetrahedron* **2006**, *10*, 2321-2330.

(146) Kumar, V. A.; Ganesh, K. N. *Acc. Chem. Res.* **2005**, *5*, 404-412.

(147) Kumar, V. A. *Eur.J.Org.Chem.* **2002**, *13*, 2021-2032.

(148) Gangamani, B. P.; Kumar, V. A.; Ganesh, K. N. *Tetrahedron* **1996**, *47*, 15017-15030.

(149) Gangamani, B. P.; Kumar, V. A.; Ganesh, K. N. *Tetrahedron* **1999**, *1*, 177-192.

(150) Vilaivan, T.; Khongdeesameor, C.; Harnyuttanakorn, P.; Westwell, M. S.; Lowe, G. *Bioorg. Med. Chem. Lett.* **2000**, *22*, 2541-2545.

(151) D'Costa, M.; Kumar, V. A.; Ganesh, K. N. *Org. Lett.* **1999**, *10*, 1513-1516.

(152) D'Costa, M.; Kumar, V.; Ganesh, K. N. *Org. Lett.* **2001**, *9*, 1281-1284.

(153) Puschl, A.; Boesen, T.; Zuccarello, G.; Dahl, O.; Pitsch, S.; Nielsen, P. E. *J. Org. Chem.* **2001**, *3*, 707-712.

(154) Sharma, N. K.; Ganesh, K. N. *Chem.Commun.* **2003**, *19*, 2484-2485.

(155) Pueschl, A.; Tedeschi, T.; Nielsen, P. E. *Org. Lett.* **2000**, *26*, 4161-4163.

(156) Kumar, V.; Pallan, P. S.; Meena; Ganesh, K. N. *Org. Lett.* **2001**, *9*, 1269-1272.

(157) Lonkar, P. S.; Kumar, V. A. *Bioorg. Med. Chem. Lett.* **2004**, *9*, 2147-2149.

(158) Pokorski, J. K.; Myers, M. C.; Appella, D. H. *Tetrahedron Lett.* **2005**, *6*, 915-917.

(159) Govindaraju, T.; Kumar, V. A.; Ganesh, K. N. *J. Org. Chem.* **2004**, *17,* 5725-5734.

(160) Govindaraju, T.; Kumar, V. A.; Ganesh, K. N. *Chem.Commun.* **2004**, *7,* 860-861.

(161) Pokorski, J. K.; Appella, D. H. *In Cyclopentane PNA: Binding properties and applications.* 2004; pp 15-444.

(162) Pokorski, J. K.; Witschi, M. A.; Purnell, B. L.; Appella, D. H. *J. Am. Chem. Soc.* **2004**, *46,* 15067-15073.

(163) Pokorski, J. K.; Witschi, M. A.; Purnell, B. L.; Appella, D. H. *J. Am. Chem. Soc.* **2004**, *46,* 15067-15073.

(164) Myers, M. C.; Witschi, M. A.; Larionova, N. V.; Franck, J. M.; Haynes, R. D.; Hara, T.; Grajkowski, A.; Appella, D. H. *Org. Lett.* **2003**, *15,* 2695-2698.

(165) Lagriffoule, P.; Wittung, P.; Eriksson, M.; Jensen, K. K.; Norden, B.; Buchardt, O.; Nielsen, P. E. *Chem.--Eur.J.* **1997**, *6,* 912-919.

(166) Govindaraju, T.; Gonnade, R. G.; Bhadbhade, M. M.; Kumar, V. A.; Ganesh, K. N. *Org. Lett.* **2003**, *17,* 3013-3016.

(167) Govindaraju, T.; Kumar, V. A.; Ganesh, K. N. *J. Am. Chem. Soc.* **2005**, *12,* 4144-4145.

(168) Meena; Kumar, V. A. *Bioorg. Med. Chem.* **2003**, *16,* 3393-3399.

(169) Lowe, G.; Vilaivan, T. *J. Chem. Soc., Perkin Trans.1* **1997**, *4,* 555-560.

(170) Vilaivan, T.; Suparpprom, C.; Harnyuttanakorn, P.; Lowe, G. *Tetrahedron Lett.* **2001**, *32,* 5533-5536.

(171) Shirude, P. S.; Kumar, V. A.; Ganesh, K. N. *Tetrahedron Lett.* **2004**, *15,* 3085-3088.

(172) Puschl, A.; Boesen, T.; Tedeschi, T.; Dahl, O.; Nielsen, P. *J. Chem. Soc., Perkin Trans. 1* **2001**, 2757-2763.

(173) Govindaraju, T.; Gonnade, R. G.; Bhadbhade, M. M.; Kumar, V. A.; Ganesh, K. N. *Org. Lett.* **2003**, *17,* 3013-3016.

(174) Govindaraju, T.; Kumar, V. A.; Ganesh, K. N. *J. Org. Chem.* **2004**, *6,* 1858-1865.

(175) Govindaraju, T.; Kumar, V. A.; Ganesh, K. N. *J. Am. Chem. Soc.* **2005**, *12*, 4144-4145.

(176) Govindaraju, T.; Madhuri, V.; Kumar, V. A.; Ganesh, K. N. *J. Org. Chem.* **2006**, *1*, 14-21.

(177) Eriksson, M.; Nielsen, P. E. *Nat. Struct. Biol.* **1996**, *5*, 410-413.

(178) Pokorski, J. K.; Witschi, M. A.; Purnell, B. L.; Appella, D. H. *J. Am. Chem. Soc.* **2004**, *46*, 15067-15073.

(179) Witschi, M. A.; Appella, D. H. *In Asymmetric synthesis of trans-cyclopentane PNA monomers.* 2004; pp 23-446.

(180) Englund, E. A.; Xu, Q.; Witschi, M. A.; Appella, D. H. *J. Am. Chem. Soc.* **2006**, *51*, 16456-16457.

(181) Pokorski, J. K.; Nam, J.; Vega, R. A.; Mirkin, C. A.; Appella, D. H. *Chem.Commun.* **2005**, *16*, 2101-2103.

(182) Levine, I. N. *Quantum chemistry;* Pearson Prentice Hall: Upper Saddle River, N.J., 2009; pp 751.

(183) Mortimer, R. G. *Physical chemistry;* Harcourt/Academic Press: San Diego, 2000; pp 1116.

(184) Yamaguchi, Y. *A New dimension to quantum chemistry: analytic derivative methods in ab initio molecular electronic structure theory;* Oxford University Press: New York, 1994; Vol. 29, pp 471.

(185) Richards, W. G.; Cooper, D. L. *Ab initio molecular orbital calculations for chemists;* Clarendon Press: Oxford, 1983; pp 116.

(186) Burkert, U.; Allinger, N. L. *Molecular mechanics;* American Chemical Society: Washington, D.C., 1982; Vol. 177, pp 339.

(187) Clark, T. *A handbook of computational chemistry: a practical guide to chemical structure and energy calculations;* Wiley: New York, 1985; pp 332.

(188) Jensen, F. *Introduction to computational chemistry;* John Wiley & Sons: Chichester, England; Hoboken, NJ, 2007; pp 599.

(189) Leach, A. R. *Molecular modelling: principles and applications;* Prentice Hall: Harlow, England; Reading, Mass., 2001; pp 744.

(190) Murrell, J. N. *Molecular potential energy functions;* J. Wiley: Chichester West Sussex; New York, 1984; pp 197.

(191) Schleyer, P. v. R. *Encyclopedia of computational chemistry;* John Wiley: Chichester; New York, 1998; pp 3429.

(192) Dixon, R.; Kollman, P. *J. Comput. Chem.* **1997**, *13,* 1632-1646.

(193) Halgren, T. A.; Damm, W. *Curr. Opin. Struct. Biol.* **2001**, *2,* 236-242.

(194) Case, D. A.; Cheatham, T. E.,3rd; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M.,Jr; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. *J. Comput. Chem.* **2005**, *16,* 1668-1688.

(195) Christen, M.; Hunenberger, P. H.; Bakowies, D.; Baron, R.; Burgi, R.; Geerke, D. P.; Heinz, T. N.; Kastenholz, M. A.; Krautler, V.; Oostenbrink, C.; Peter, C.; Trzesniak, D.; van Gunsteren, W. F. *J. Comput. Chem.* **2005**, *16,* 1719-1751.

(196) Jorgensen, W. L.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1988**, *6,* 1657.

(197) Halgren, T. *J. Comput. Chem.* **1996**, *5-6,* 490-519.

(198) Brooks, B. R.; Bruccoleri, R. E.; Olafson, D. J.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, 187-217.

(199) MacKerel Jr., A. D.; Brooks III, C. L.; Nilsson, L.; Roux, B.; Won, Y.; Karplus, M. *CHARMM: The Energy Function and Its Parameterization with an Overview of the Program;*John Wiley & Sons: Chichester: 1998; Vol. 1, pp 271-277.

(200) Szabo, A.; Ostlund, N. S. *Modern quantum chemistry: introduction to advanced electronic structure theory;* Dover Publications: Mineola, N.Y., 1989; pp 466.

(201) Atkins, P. W.; Friedman, R. S. *Molecular quantum mechanics;* Oxford University Press: Oxford; New York, 2005; pp 573.

(202) Slater, J. C. *Phys. Rev.* **1929**, *10,* 1293.

(203) Young, D. C. *Computational chemistry: a practical guide for applying techniques to real world problems;* Wiley: New York, 2001; pp 381.

(204) Frisch, M. J.; Pople, J. A.; Binkley, J. S. *J. Chem. Phys.* **1984**, *7,* 3265-3269.

(205) Hehre, W. J. *Ab initio molecular orbital theory;* Wiley: New York, 1986; pp 548.

(206) Friesner, R. A.; Beachy, M. D. *Curr. Opin. Struct. Biol.* **1998**, *2*, 257-262.

(207) Møller, C.; Plesset, M. S. *Phys. Rev.* **1934**, *7*, 618.

(208) Ratner, M. A.; Schatz, G. C. *An introduction to quantum mechanics in chemistry;* Prentice Hall: Upper Saddle River, NJ, 2000; pp 305.

(209) Hinchliffe, A. *Molecular modelling for beginners;* Wiley: Hoboken, NJ, 2008; pp 411.

(210) Foresman, J. B.; Frisch, A.; Gaussian, I. *Exploring chemistry with electronic structure methods;* Gaussian, Inc.: Pittsburgh, PA, 1996.

(211) Kincaid, D.; Cheney, E. W. *Numerical analysis: mathematics of scientific computing;* American Mathematical Society: Providence, R.I., 2002; Vol. 2, pp 788.

(212) Liebman, J. F.; Greenberg, A. *Molecular structure and energetics;* VCH Publishers: Deerfield Beach, Fla., 1989.

(213) Pople, J. A.; Krishnan, R.; Schlegel, H. B.; Binkley J. S. *Int. J. Quantum Chem.* **1978**, *5*, 545-560.

(214) Koch, W.; Frenking, G.; Gauss, J.; Cremer, D.; Collins, J. R. *J. Am. Chem. Soc.* **1987**, *20*, 5917.

(215) Chen, I. J.; Yin, D.; MacKerell, A. D.,Jr *J. Comput. Chem.* **2002**, *2*, 199-213.

(216) MacKerell, A. D.; Wiorkiewicz-Kuczera, J.; Karplus, M. *J. Am. Chem. Soc.* **1995**, *48*, 11946-11975.

(217) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul,

A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Laham, A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. **2003**.

(218) Scott, A. P.; Radom, L. *J. Phys. Chem.* **1996**, *41*, 16502-16513.

(219) Chirlian, L.; Francl, M. *J. Comput. Chem.* **1987**, *6*, 894-905.

(220) Breneman, C. M.; Wiberg, K. B. *J. Comput. Chem.* **1990**, *3*, 361-373.

(221) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *2*, 926-935.

(222) Shields, G. C.; Laughton, C. A.; Orozco, M. *J. Am. Chem. Soc.* **1999**, *7*, 1625.

(223) Brunger, A. T.; Karplus, M. *Proteins* **1988**, *2*, 148-156.

(224) Hindle, S. A.; Rarey, M.; Buning, C.; Lengaue, T. *J. Comput. Aided Mol. Des.* **2002**, *2*, 129-149.

(225) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. *J. Comput. Phys.* **1977**, 327.

(226) Essmann, U.; Perera, L.; Berkowitz, M.; Darden, T.; Lee, H.; Pedersen, L. *J. Chem. Phys.* **1995**, *19*, 8577-8593.

(227) Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graph.* **1996**, *1*, 33-8, 27-8.

(228) Sigfridsson, E.; Ryde, U. *J. Comput. Chem.* **1998**, *4*, 377-395.

(229) Dill, J. D.; Allen, L. C.; Topp, W. C.; Pople, J. A. *J. Am. Chem. Soc.* **1975**, *25*, 7220-7226.

(230) Frisch, M. J.; Pople, J. A.; Del Bene, J. E. *J. Phys. Chem.* **1985**, *17*, 3664-3669.

(231) Zheng, J.; Merz, K. Jr., *J. Comput. Chem.* **1992**, *9*, 1151-1169.

(232) Pranata, J.; Wierschke, S. G.; Jorgensen, W. L. *J. Am. Chem. Soc.* **1991**, *8*, 2810-2819.

(233) Mackerell, A. D.,Jr *J. Comput. Chem.* **2004**, *13*, 1584-1604.

(234) Guvench, O.; MacKerell, A. D.,Jr *J. Mol. Model.* **2008**, *8*, 667-679.

(235) Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A. H.; Teller, E. *J. Chem. Phys.* **1953**, *6*, 1087-1092.

(236) Kilpatrick, J. E.; Pitzer, K. S.; Spitzer, R. *J. Am. Chem. Soc.* **1947**, *10*, 2483-2488.

(237) Solomons, T. W. G.; Fryhle, C. B. *Organic chemistry;* John Wiley: Hoboken, NJ, 2008; pp 1191.

(238) Tomimoto, M.; Go, N. *J. Phys. Chem.* **1995**, *2,* 563-577.

(239) Altona, C.; Sundaralingam, M. *J. Am. Chem. Soc.* **1972**, *23,* 8205-8212.

(240) Cremer, D.; Pople, J. A. *J. Am. Chem. Soc.* **1975**, *6,* 1358-1367.

(241) Cremer, D.; Pople, J. A. *J. Am. Chem. Soc.* **1975**, *6,* 1354-1358.

(242) Tironi, I. G.; Sperb, R.; Smith, P.; van Gunsteren, W. F. *J. Chem. Phys.* **1995**, *13,* 5451-5459.

(243) Carpenter, G.; Grossberg, S. *Appl. Opt.* **1987**, *23,* 4919-4930.

(244) Misumi, A.; Iwanga, K.; Furuta, K.; Yamamoto, H. *J. Am. Chem. Soc.* **1985**, *107*, 3343-3345.

(245) Wittung, P.; Nielsen, P.; Buchardt, O.; Egholm, M.; Norden, B. *Nature.***1994**, *368*, 561-563.

(246) Wittung, P.; Eriksson, M.; Reidar, L.; Nielsen, P.; Norden, B. *J. Am. Chem. Soc.* **1995**, *117*, 10167-10173.

(247) Jain, V.; Green, M.; Faccini, A.; Tedeschi, T.; Corradini, R.; Armitage, B. *Polym. Prepr.* **2006**, *47*, 26–27.

(248) Gibson, K.; Scheraga, H. *J. Comput. Chem.* **1987**, *8*, 826–834.

(249) Leporc, S.; Mauffret, O.; Tevanian, G.; Lescot, E.; Monnot, M.; Fermandjian, S. *Nucleic Acids Res.***1999**, *24*, 4759-67.

(250) Perbandt, M.; Vallazza, M.; Lippmann, C.; Betzel, C. *Acta Crystallogr. D Biol. Crystallogr.* **2001**, *57*, 219-24.

VITA

Anna K. Manukyan

EDUCATION

Ph.D in Biomedical Sciences, Biochemistry track **2003-2009**

Department of Chemistry & Biochemistry
Old Dominion University, Norfolk, VA

Dissertation: "Molecular dynamics study of single stranded peptide nucleic acids"
Advisor: Prof. Jennifer Radkiewicz-Poutsma

B.Sc Honors in Biochemistry and Environmental Science **1997-2002**

Department of Biochemistry
Armenian State University, Yerevan, Armenia

Thesis: "The role of noradrenargic system in the mechanisms of emotional stress"
Advisor: Prof. Mesrop Khanbabyan

PUBLICATIONS

A. Manukyan, J. Poutsma, "Development of Force Field Parameters for Cyclopentyl PNA," *in preparation*

T. Ribelin, C. Katz, D. Withrow, S. Smith, A. Manukyan, V. Day, B. Neuenswander, J. Poutsma and J. Aubé, "Highly Stereoselective Ring Expansion Reactions Mediated by Attractive Cation-p Interactions." *Angewandte Chemie*, International Edition, **2008**, 47(33), 6233

C. Katz, T. Ribelin, D. Withrow, Y. Basseri, A. Manukyan, A. Bermudez, C. Nuera, V. Day, D. Powell, J. Poutsma and J. Aubé, "Non-Bonded, Attractive Cation–Interactions in Azide-Mediated Asymmetric Ring Expansion Reactions." *Journal of Organic Chemistry*, **2008**, 73, 3318

A. Manukyan, J. Poutsma, "The Effect of Solvent on the Stereoselectivity of Reactions between Hydroxyalkyl Azides and Ketones", Journal of Molecular Structure: *THEOCHEM* **2006**, 766, 105