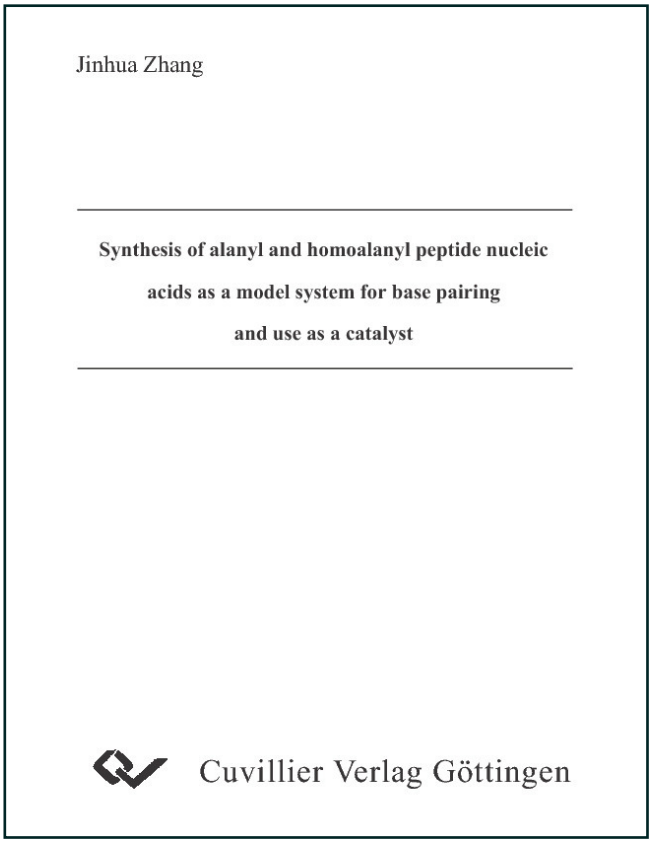
# Synthesis of alanyl and homoalanyl peptide nucleic acids as a model system for base pairing and use as a catalyst



## 1. Introduction

Every living organism on our planet stores its genetic information in the same chemical form as double-stranded deoxyribonucleic acid (DNA), the carrier of genetic information in all cellular life as well as in many viruses. The structure of DNA consists of two strands of linked nucleotides, each of which is composed of a deoxyribose sugar residue, a phosphoryl group and one of the four nucleobases: adenine (A), thymine (T), guanine (G), or cytosine (C). It generally forms an anti-parallel right-handed double helix based on the Watson-Crick hydrogen bonding (H-bonding) exclusively of adenine-thymine (two H-bonds) and guanine-cytosine (three H-bonds) base pairs. Furthermore stacking with the ideal distance of basepairs of 3.4 Å is provided by helicalization. Each nucleobase is hydrogen bonded to a nucleobase of the opposite strand, the two strands are complementary, with the sequence of one strand implys the sequence of the other. The genetic information is given by the sequence of the nucleobases.

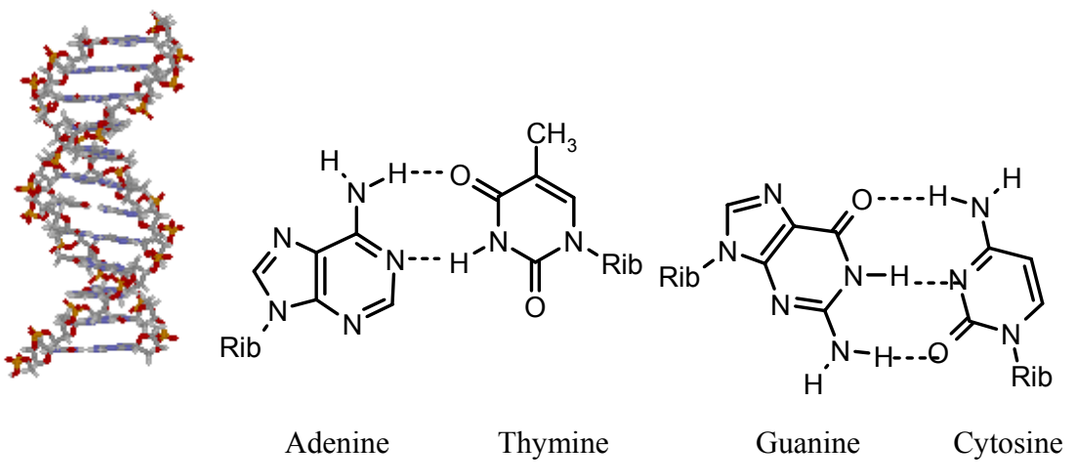



Figure 1.1: B-DNA double strands and A-T G-C base pairing.

The expression of genetic information is based on transcription, in which a DNA strand serves as a template to transcribe portions of its genomic information into molecules of the closely related polymer messenger ribonucleic acids (mRNAs), which at the ribosome composed of RNA and protein guide the synthesis of protein molecules by the more complex machinery known as translation. The order in which the amino acids linked together is prescribed by the base sequences of mRNAs, and

the function of each protein depends on its amino acid sequence. Consequently, the genetic information encoded by DNA is translated through the intermediacy of RNA to the structures and functions of proteins (Figure 1.2).[1]

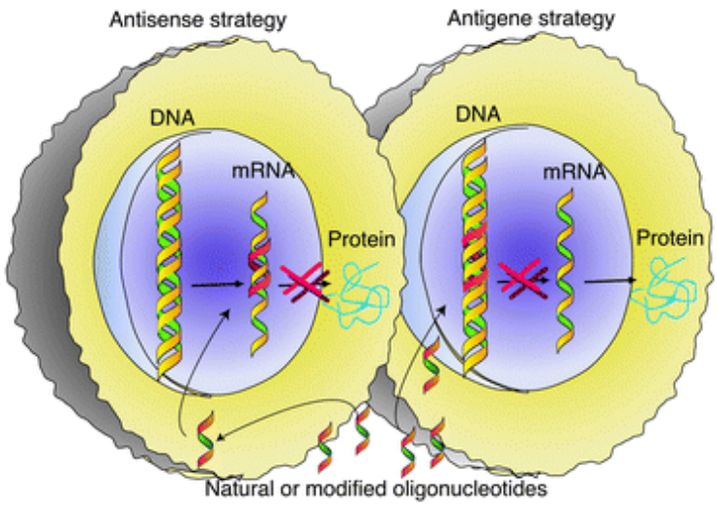



Figure 1.2: The cell process of from genome to RNA to protein and representation of antisense and antigene strategies.[2]

The quest to develop new drug therapies based on sequence-specific interactions between complementary nucleic acids is an exciting and rapidly growing field of chemical research.[3] One way to gene therapy consists in introducing a molecule into a cell, that activates or represses gene expression. This is the concept of antisense or antigene therapy. The antigene and antisense therapy avails itself looking for oligomers, which are qualified for the recognition of nucleic acids and can bind selectively with DNA or RNA. Thereby therapeutic hope is based on the blockade of defect genetic sequences preventing the biosynthesis of defected proteins. For this purpose in 1991 the group of P. E. Nielsen developed the oligonucleotide surrogates that are capable of maintaining Watson-Crick (Hoogsteen) base-pairing to native RNA (DNA) targets but do not incorporate the usual phosphodiester linkages which are susceptible to nucleases and proteases and incompatable with passive membrane transport.[4] They developed aminoethylglycine peptide nucleic acids by molecular modelling studies, in which the complete sugar-phosphate backbone of DNA is replaced by a polyamide backbone containing N-(2-aminoethyl) glycine units. The nucleobase is attached to the nitrogen of the glycine through an acetic acid linker. The number of atoms between the ends of two monomeric units is the same as in DNA.

Due to their structure they are not substrates for enzymatic degradation. They hybridize selectively to their target mRNAs with specific interactions. Antisense oligonucleotides are of great interest as potential molecular biological tools and therapeutic agents.[5] They have the ability to interact with a complementary sequence of mRNA and specifically to inhibit gene expression and protein biosynthesis. In contrast to conventional protein targeting drugs, less or no undesired side effects are anticipated.

In order to examine the contributing factors such as stericical interactions, stacking, hydrogen bonding, solvation etc. affecting the thermodynamic stability of DNA , there was a need to reduce the complicated systems of a DNA double strand to a geometrically simpler system. In this background alanyl peptide nucleic acids (alanyl-PNAs) were introduced.[6] They were built of alternating configured alanyl nucleo amino acid units, whereby the nucleobases are linked covalently over the β-position with the amino acid alanine. In the pairing complex alanyl PNA has a linear peptide backbone and subjects therefore to no selectivity-determining topological obligations by helicalization. Colinear oriented backbone strands permit all sizes of basepairs as well as all orientations of the nucleobases. Alanyl-PNA can carry out different pairing modes. Molecular recognition results from Watson-Crick and Hoogsteen site basepairing. Homoalanyl peptide nucleic acid (homoalanyl PNA) can be obtained from homologization of the side chain by a methylene group.[7] From the investigation of the pairing characteristics of alanyl-PNAs and their homologous oligomers results the fundamental knowledge of pairing affinities of the nucleobases and their stacking contribution in dependence of the nucleobase orientation.[6]

The objective of this work is to take the alanyl-PNA as a model system for investigating the pairing characteristics of nucleobases. Alanyl-PNAs are well suited for the study of base pairing and stacking since the linear topology of this artificial pairing system allows the formation of all base pair combinations and various pairing modes.[8] This contains the structure of the peptide nucleic acid with the nucleobases thymine, adenine, cytosine, and guanine. Xanthine-xanthine pairing alanyl-PNA hexamers turned out to be of special interest since they were found to be far more

stable than expected for two dentate base pairing.[9] Such systems also exhibit unexpected new effects which arise due to the high number of possible pairing modes. The pairing behaviour of xanthines and N-methylated xanthines are of interest, the structural pairing properties of xanthines and N1 and N3 methylated xanthines were investigated. The xanthine oligomer and N1 or N3 methylated xanthine oligomers were synthesized and the stabilities of their pairing complexes were examined by means of temperature-dependent UV spectroscopy. Comparison of the stabilities of the pairing complexes by the structural pairing possibilities of the series of xanthine alanyl-PNA oligomers in cooperation with theoretical work, finally lead to the pairing mode of the xanthine oligomer.

It is already established that alanyl and homoalanyl-PNA as a pairing system have their own selectivity rules.[6] For example, an alanyl PNA duplex possesses with a suitable antiparallel strand, a linear, rigid β-sheet structure. In order to deeply understand the molecular structure of mixed alanyl and homoalanyl-PNA with G, C nucleobase pairing, a mixed alanyl and homoalanyl G-C pairing hexamer was synthesized. In cooperation with Professor Sheldrick (Göttingen), the analysis of its structure by crystallization followed by X-ray analysis is attempted.

The apolar environment in the inside pocket of an alanyl-PNA base stack [10]shows that this might be a suitable environment for catalytic reactions. By incorporation of histidine into the sequence, metal ions like $Zn^{2+}$ might be coordinated with its aromatic side chain. The $Zn^{2+}$-PNA binding system might be a preorganized model system for metallohydrolase. Therefore, a series of alanyl and homoalanyl-PNA hexameric oligomers were synthesized. Their UV-dependent melting curves with and without the presence of $Zn^{2+}$are analyzed and hopefully lead to metallohydrolases.

## 2. Peptide nucleic acids

Up to now the highly specific recognition of DNA/RNA via artificial oligonucleotides in the use of antisense and antigene therapy is in the focus of medicinal research. A new attempt towards improved cell uptake, resistance to degradation by proteases and nucleases as well as sequence specificity is represented by the development of peptide nucleic acids (PNAs). In 1991, Nielsen, Egholm, Berg, and Buchardt reported the synthesis of these biooligomers as a completely artificial DNA/RNA analog.[3]

The PNA monomer is 2-aminoethyl glycine linked by a methylenecarbonyl linkage to one of the four DNA-bases (adenine, guanine, thymine, or cytosine). That means that the nucleobases of DNA were retained, but the deoxyribose phosphodiester backbone of DNA was replaced by a pseudo-peptide backbone homomorphous with the DNA (Figure 2.1). Due to this structural benefit it was originally designed as a ligand for the recognition of double stranded DNA. The concept was to mimic an oligonucleotide binding to double stranded DNA via Hoogsteen site base pairing in the major groove to form stable triple strands.[3]

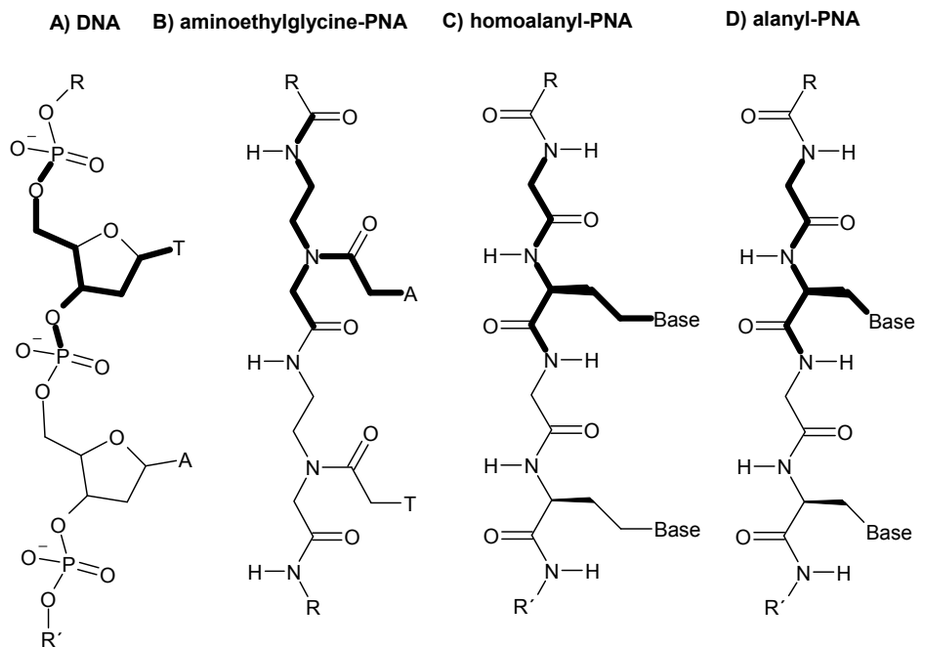



Figure 2.1: Comparison of A) DNA, B) aminoethylglycine-PNA (*Nielsen-PNA*), C) chiral homoalanyl-PNA and D) chiral alanyl-PNA, the last two are described in the following. The isosteric units are highlighted. “Base” represents one of the natural nucleobases found in DNA.