Further exploration of the imidazole pK_{aH} motif in DNAzymes through NMR studies of sequence permutations

Thesis submitted to obtain

the degree of Master of Science in Chemistry by

Anne-Mare DE VRIES

Academic year 2015 - 2016

DANKWOORD

De route van bachelorstudente naar het indienen van een Master thesis loopt niet altijd via gebaande wegen. Op de weg die ik gegaan ben, zijn er veel mensen geweest die een goede richting aangaven of me een stuk hebben vergezeld op mijn reis. Allereerst wil ik graag mijn promotoren José Martins en Annemieke Madder bedanken. Zij maakten het mogelijk te werken aan een zeer uitdagend project en hebben me altijd bijgestaan met raad en daad gedurende mijn laatste jaar. Mijn begeleider, die direct betrokken was bij mijn werk en mij dagelijks bijstond met het beantwoorden van vragen en van wie ik ontzettend veel heb geleerd, ben ik zeer veel dank verschuldigd: bedankt Dieter Buyst. Lars Verdonck, bedankt voor alle praktische begeleiding en de altijd goed gezinde tips en antwoorden.

In een jaar waarin je aan je thesis werkt, is ook de werkomgeving zeer bepalend. Maar als je zulke leuke collega's hebt als ik heb gehad, dan ga je iedere ochtend fluitend naar je 'werk'. Bij deze bedankt aan Tim Courtin, Niels Geudens, Matthias De Vleeschouwer, Kim De Nolf, en de andere collega's van de NMRSTR-, de OBCR- en de PCN-groep. Daarnaast ook veel dank aan Roos Naves.

Buiten de mensen die direct verbonden waren aan mijn thesis zijn er ook vrienden die me altijd steunen en zo nu en dan zorgden voor de nodige ontspanning. Als eerste wil ik mijn kotgenote Eline Van den Dungen bedanken, lief en leef hebben we al gedeeld en ik ben nog steeds erg dankbaar dat je zo een toffe huisgenote bent. Verder mijn vrienden Esther De Rycke, Yamin Abdouni, Jelle Caestecker en Karel Zegers: fijn dat jullie er altijd voor me zijn.

De boog kan niet altijd gespannen zijn en afgelopen jaar is mijn handbalteam van handbalclub Don Bosco Gent van onschatbare waarde geweest, merci aan iedereen.

Al zit ik dan vaak op een grote afstand, mensen van wie je houdt zijn altijd in de buurt. Mijn papa en mama zijn het beste voorbeeld. Mijn zussen Maartje en Nelleke en hun mannen en Jasmijn en Madelief zijn erg belangrijk voor me en wil ik hier graag noemen. Ook John Meertens en Marc Meertens mogen niet ontbreken. Carla Pera en Ria de Jong zijn een klasse apart, bedankt voor jullie onvoorwaardelijke steun de afgelopen 24 jaar.

Tot slot zijn er nog enkelen die genoemd moeten worden. Vragen stellen is soms belangrijker dan antwoorden geven: bedankt Eddy Blaas. Ann De Muyt, ontzettend bedankt voor de warme opname binnen je gezin, toen ik zelf getroffen was door een blessure. En tenslotte veel dank aan meneer Smits, mijn interesse voor de scheikunde begon bij u in de les.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

A	Adenine
Å	Angstrom
ACN	Acetonitrile
AMBER	Assisted model building energy refinement
C	Cytosine
DCM	Dichloromethane
DIPEA	N,N-diisopropylethylamine
DMF	Dimethylformamide
DMTr	4,4-dimethoxytrityl
DNA	Deoxyribonucleic acid
DSS	4,4-dimethyl-4-silapentane-1-sulfonic acid
E	Enzyme
E_a	Activation energy
EDTA	Ethylenediaminetetraacetic acid
G	Guanine
HPLC	High-performance liquid chromatography
Im	Imidazole
IDT	Integrated DNA technologies
MALDI TOF	Matrix assisted laser desorption/ionization time of flight
MD	Molecular dynamics
MHz	Mega hertz
NMR	Nuclear magnetic resonance
NOESY	Nuclear Overhauser effect spectroscopy
OBCR	Organic and biomimetic chemistry research
P	Reaction product

PCN	Physics and chemistry of nanostructures
ppm	Parts per million
RMSD	Root mean square deviation
RNA	Ribonucleic acid
S	Substrate
T	Thymine
TCA	Trichloroacetic acid
TEAA	Triethylammonium acetate
TFA	Trifluoroacetic acid
TOCSY	Total correlations spectroscopy
TS	Transition state
UV-Vis	Ultraviolet-visible light spectroscopy
WT	Wild type

LIST OF FIGURES

HOOFDSTUK 8

LIST OF TABLES

CHAPTER 1 INTRODUCTION

1.0 INTRODUCTION

This work entitled: ***Further exploration of the imidazole pK_{aH} motif in DNAzymes through NMR studies of sequence permutation*** is part of a collective research project within Ghent University, in particular with scientists from the NMRSTR-group (NMR and Structure Analysis Research) and the OBCR-group (Organic and Biomimetic Chemistry Research) both from the Organic and Macromolecular Chemistry Department. The aim of this work was formulated based on findings from earlier research by Vicky Gheerardijn, Bjorn Van Gasse and in particular by Dieter Buyst[1-3]. The research questions being answered are more of a fundamental nature than directly linked to applications on the short term. Nevertheless it can be of great importance for future enzymatic studies.

The ultimate goal pursued in the common research project is to develop a histidine based hydrolyse-type DNAzyme founded on a predictable B-DNA framework. This is a continuation of the development of a serine protease mimic, using histidine, serine and aspartic acid-like nucleotide building blocks[1,2]. In a short DNA duplex unnatural nucleotides, adapted with functional groups similar to histidine, serine and aspartic acid, are introduced to mimic a so-called catalytic triad to obtain enzymatic activity. Systematic studies were executed to determine the ideal length of the DNA-duplex as well as the type, number and position of the modified nucleotides. During the PhD of Dieter Buyst a pK_{aH} regulating motif induced by a histidine-like modified nucleotide within the DNA scaffold was described based on UV Vis thermal degradation data, NMR measurements and MD simulations. As a result of a specific interaction, caused by hydrogen bond formation, they observed that the protonated imidazole within the local DNA environment (G_9, G_{19}, O_6 carbonyl groups, see Figure 1.1) experiences a significantly increased imidazolium pK_{aH}. Furthermore, they discovered via *in silico* MD simulations that by adaptation of the local DNA environment (base pair context), this specific interaction and the correlated up regulation of the imidazolium pK_{aH} could either be present or absent.
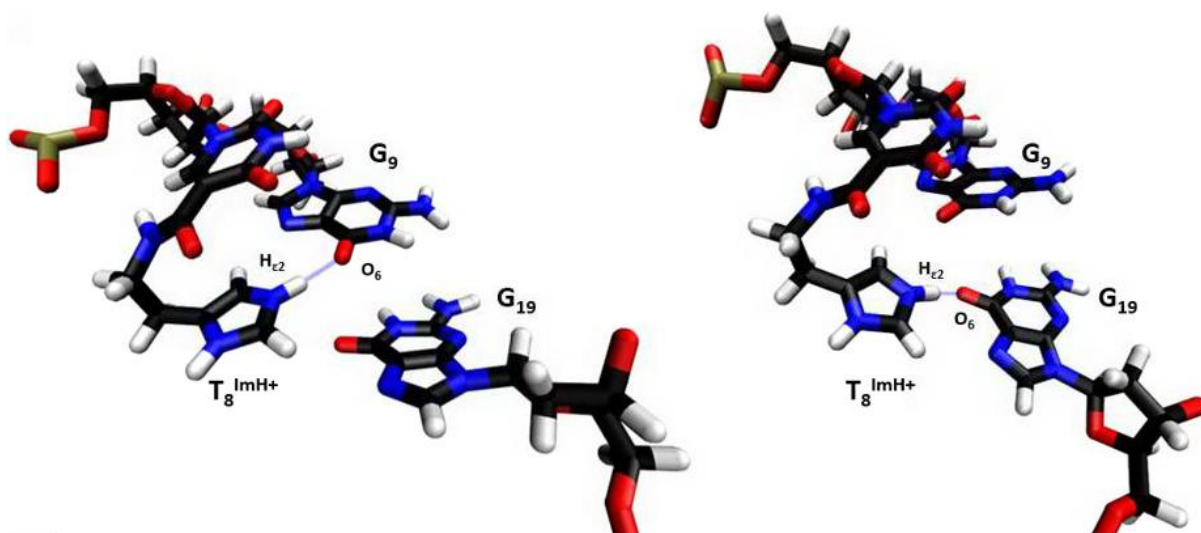
Figure 1.1. Two snapshots of an MD simulation where two specific interaction patters of the pK_{aH} regulating motif are highlighted. In each case the imidazole functionality (T_8^{ImH+}) is the hydrogen bond donor, while G_9 and G_{19} act as hydrogen bond acceptor. The hydrogen bonds are represented by blue dotted lines. (Adapted from [3]).

The aim of this thesis is to investigate the robustness of this motif and to examine its resistance to changes in the nucleotide sequence.

The OBCR and NMRSTR group based this thesis on their research showing that MD simulations can predict the behaviour of the duplexes and the formation of the pK_{aH} motif in particular. Based on these predictions we decided which sequences should be synthesized and analyzed or not. UV-Vis thermal melting experiments and NMR data serve as our main source of information.

Before presenting the results of the research a few concepts need clarification to contextualize and show the relevance of the research question. Section 1.1 and 1.2 therefore introduce enzymes and DNA found in nature, where 1.3 in contrast presents artificial enzymes. In the last part of this first chapter a project outline is presented.

1.1 ENZYMES

Chemical reactions in all living organisms are possible because of the existence and properties of catalysts. In biochemistry a crucial set of catalysts are protein based and are known as enzymes. These enzymes are of vital importance, because without them all the reactions necessary for a living organism become impossible at physiological temperature due to the large activation energy (E_a) associated with these reactions. The role of the enzymes is to decrease the activation energy by stabilizing the transition state ($TS^{\ddagger}$) of the substrates. The Michaelis-Menten model gives an accurate description of most enzymatic reactions involving a single substrate[4]:

$$E + S \rightleftharpoons E \cdot S \longrightarrow E \cdot P \rightleftharpoons E + products \qquad \text{Eq. 1.1}$$
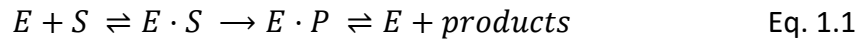
Where E, S and P refer to enzyme, substrate and product respectively. At the end of the reaction the enzyme is released in its original form and is capable to participate in a new catalytic cycle. Different models of enzyme-substrate complex (E·S) formation exists. Figure 1.2 illustrates the main differences between the lock and key (a), conformational isomerism (b), induced fit (c) and conformational selection (d) model.
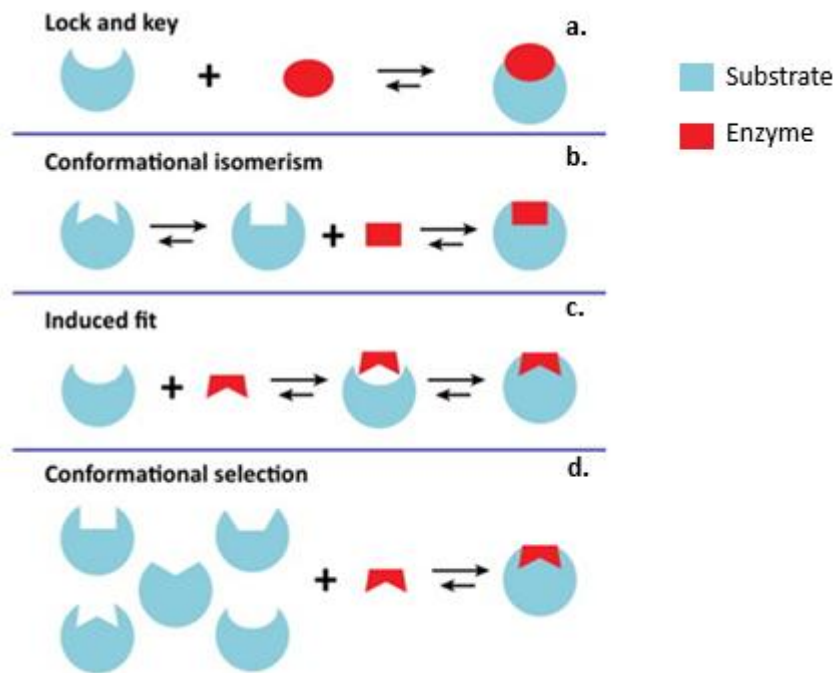
Figure 1.2. Enzyme - substrate binding models. (Adapted from [5]).

The most simple representation of formation of an enzyme-substrate complex (a) is proven to be too simplistic to occur *in vivo* and was replaced by the induced fit model (c). Both (b) and (d) can exist next to the induced fit model and are an extension of the possible binding models[6].

1.2 DNA

The extensive knowledge we have to date on the structure and function of DNA is based on decades of research and breakthroughs. Here I want to highlight two important moments in the history of DNA. In 1869 J.F. Miescher first isolated and described DNA[7]. An iconic breakthrough took place in 1954 when Watson and Crick[8] elucidated the helical structure of DNA based on high quality X-ray spectra obtained by Rosalind Franklin.

Deoxyribo**n**ucleic **a**cid (DNA) constitutes the genetic information of all living organisms which is present in each cell. Eukaryotes have, in contrast to prokaryotes and viruses, a

separate compartment which contains the DNA, the nucleus. The main task of this macromolecule is its involvement in the inheritance of characteristics from parents to their offspring. Zooming in on the structural aspects of DNA we can distinguish nucleobases attached to a backbone (Figure 1.3B). The backbone is a negatively charged chain of alternating sugar-phosphate groups. The bases can be subdivided into two classes: the purines (adenine and guanine) and the pyrimidines (cytosine and thymine), as illustrated in Figure 1.3D.
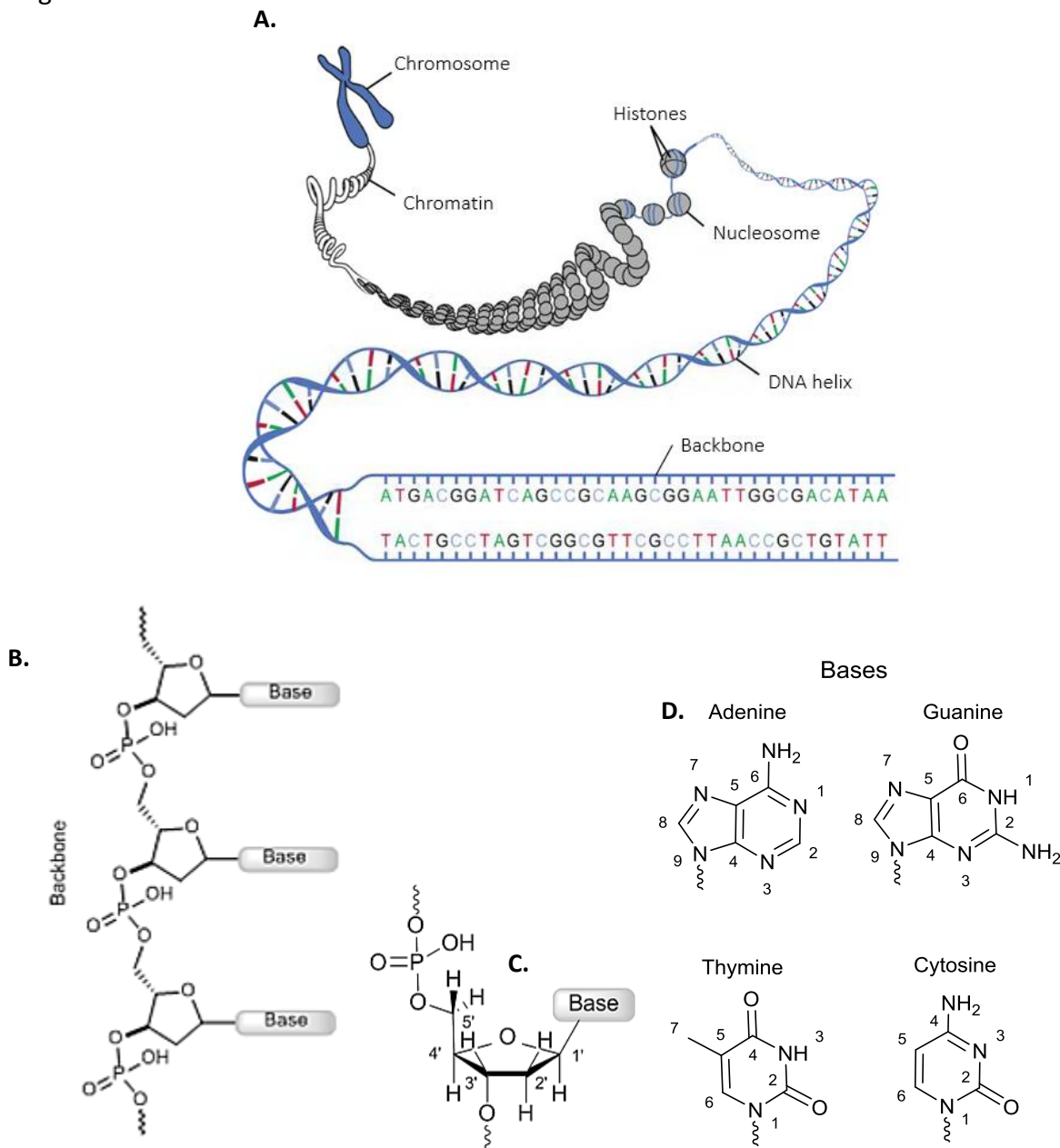
Figure 1.3. (A). Natural occurring DNA in the nucleus. (B). Detail of the backbone. (C). The sugar ring and the numbering and the numbering scheme used for the hydrogen atoms during NMR assignment is indicated. (D). The nucleobases with the purines adenine and guanine and the pyrimidines thymine and cytosine.

Base pairing, through mutual hydrogen bonds across two complementary strands, occurs both between adenine and thymine, and guanine and cytosine with the formation of the so-called Watson-Crick base pairs (Figure 1.4). If two complementary strands are present they are capable to adopt a folded form. The stabilisation of a DNA helix depends on a fine balance of interactions including hydrogen bonds between bases, hydrogen bonds between bases and surrounding water molecules, and base-stacking interactions between adjacent bases. As can be expected from the number of hydrogen bonds, the stability of a G•C pair is higher than that of an A•T pair. The base-stacking interaction is more prevalent in duplexes and depends on the aromaticity of the bases and their dipole moments and has shown to be sequence dependent[9]. Besides internal interactions, external factors also affect the degree of stabilisation. With higher salt concentration the stability increases because the high cation concentration masks the destabilising charge repulsion between the two negatively charged phosphodiester backbones. *In vivo,* proteins named histones are involved in both stabilizing and condensation of the DNA helix by ordering the DNA helix into structural units called nucleosomes (Figure 1.3A). The total length of a chain of DNA differs between species and range from $110 \cdot 10^3$ (smallest DNA strand of a bacteria) to $250 \cdot 10^6$ base pairs (longest human DNA strand).
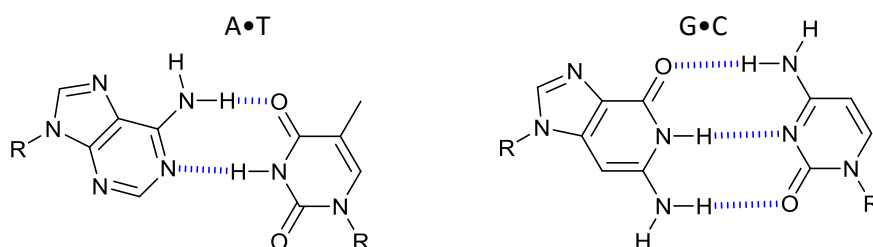
Figure 1.4. The hydrogen bonds involved in the inter-strand Watson-Crick base pairing (blue), the sugar ring and backbone of the DNA molecule are not shown but represented by R.

Several stable double helix conformations are known from X-ray analyses[10]. Some examples of DNA conformations are A-, B- and Z-DNA (Figure 1.5), all of which are believed to have a biological function. A- and B-helices are both right handed, but an A-helix is more compact and shorter in contrast to B-DNA which is more elongated and slimmer. Z-DNA is the complete opposite of A-DNA: it is left-handed, long and even more elongated than B-DNA[11-15]. In addition Z-DNA consists of only G and C base pairs, which means that the occurrence of this type is sequence dependent. In contrast, the presence of A- or B-DNA depends on the surrounding conditions. More specifically, conversion from the B- to the A-form occurs

under dehydrating conditions. The different folded forms are characterized by different sizes for the grooves, i.e. the accessible space between the intertwined strands (Figure 1.5). They are referred to as the major and minor groove. In this work we focus on the biological most common DNA conformation, namely B-DNA being part of the secondary structure of the systems of interest.
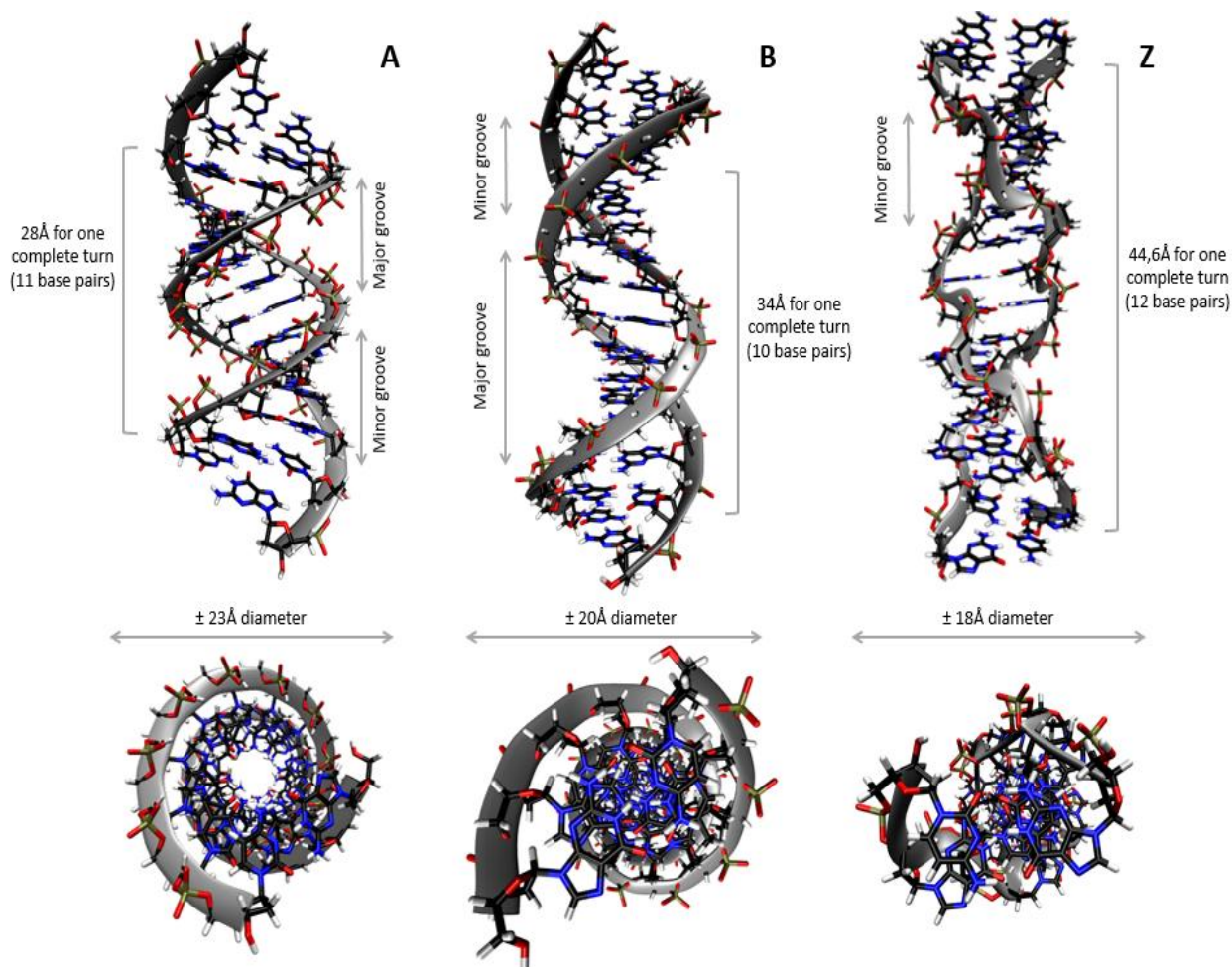
Figure 1.5. General overview of three types of DNA secondary structure, A-DNA (left), B-DNA (middle) and Z-DNA (right). All sequences are built using the crystallographic parameters provided by Discovery Studio 4.0 and visualised using VMD 1.9.1. (Adapted from [3]).

1.3 ARTIFICIAL ENZYMES

With increasing knowledge about the behaviour and structure of several enzymes, scientists began to design and develop artificial enzymes based on artificial scaffolds or inspired by motifs found in nature[16]. Changing the primary structure of a protein-based enzyme by introduction of non-natural occurring amino acids or by deleting some of them can cause major changes in the secondary or even tertiary structure of the protein. The change is caused by the protein's folding and stability which is regulated by a whole range of processes and interactions to which each individual amino acid may contribute. This has led to efforts

where simpler structures are used to organize the functionalities, required for catalytic activity in a desired way. One of the simpler structures which is being investigated to support catalysis, is the class of nucleic acids[17]. Nucleic acids in general have a significant drawback in comparison with proteins, because of the lack of functional groups. Proteins are superior to catalyze chemical reactions: 20 different natural amino acids exists with a large variety of functional groups. The combination of these 20 amino acids leads to an extended number of possible molecules. In contrast, nucleic acids only have 4 very similar building blocks and the functional group diversity is more limited and characterized by pK_{aH} values outside the interesting range (see Figure 1.6). Therefore we must use nucleotides modified with functional groups suitable for catalysis in this thesis[18,19].
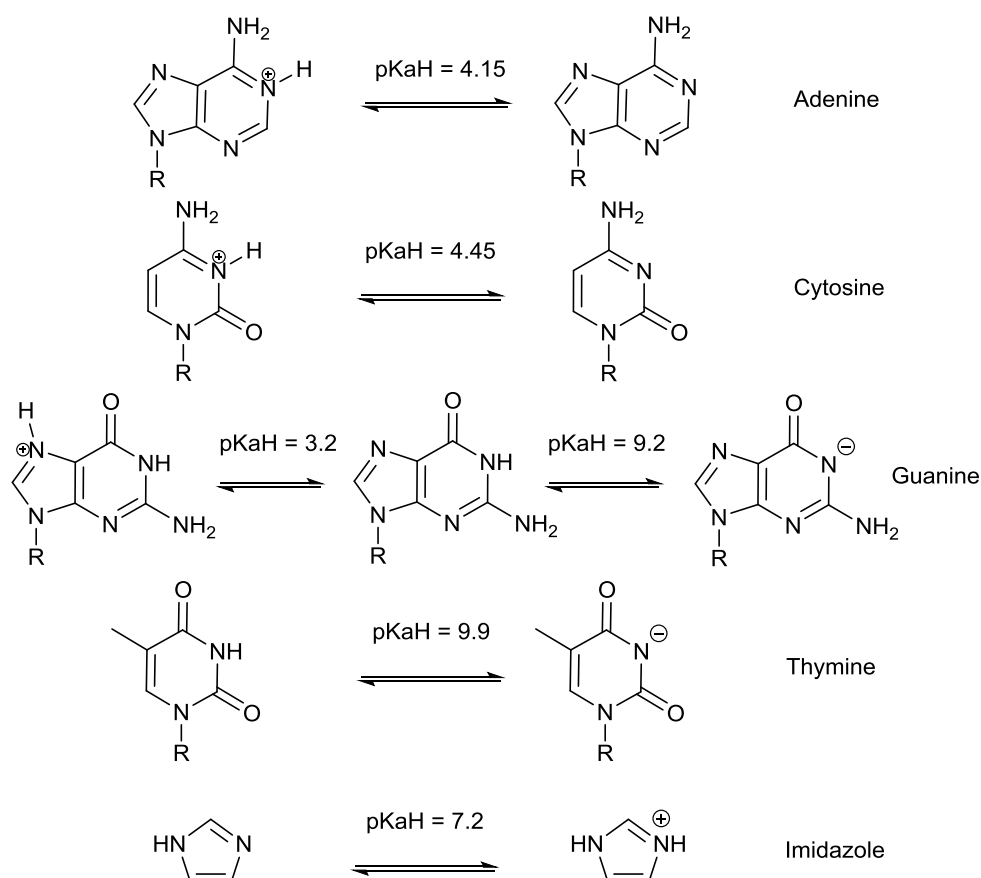
Figure 1.6. Protonation equilibria of the different DNA bases and the imidazole functionality. (Adapted from [11]).

1.3.0 DNA*ZYMES*

Processes catalyzed by functionalities within a natural DNA moiety have not (yet) been discovered in nature[20,21], but since the detection of catalytic RNA[22,23] by T. Cech in the early 1980s, scientists became aware of the ability of nucleic acids to function as an enzyme as well. This important discovery was awarded with a Nobel prize in 1989[24]. The discovery of the RNAzymes came not as a total surprise, because Watson and Crick already predicted

the ability of RNA to have a catalytic function[25,26]. Enzymes based on RNA are called ribozymes. An example of a natural occurring ribozyme is the ribosome[27, 28]. Artificial DNAzymes can be used to assist different reactions including inactivation of a target RNA, probing of structured RNA or assisting in the manipulation of recombinant DNA[29, 30].

The DNAzyme rate enhancements of artificial DNAzymes have been shown to be in the same range as ribozymes[31] and there is no other proof that ribozymes are more capable of catalyzing reactions. The choice to use DNA as a scaffold for an enzyme instead of RNA is mainly based on practical considerations. First, DNA is more stable and has a more rigid structure compared to RNA. The increased stability of DNA is caused by the absence of a reactive hydroxyl group at the 2' position of the sugar ring and since DNA mostly occurs as a double helix, the conformation is more predictable compared to the less rigid conformation adopted by single strands of RNA. As a result of this, DNA enables us to work with reliable positioning of synthetic functionalities. Secondly, the synthesis of DNA on solid-phase is more cost-efficient. This is related to the expense and ease during the monomer synthesis, the efficiency of the introduction of new nucleotides and the simplicity of deprotection after the reaction cycle[29]. Finally the method in this thesis (NMR) fits better to DNA.

1.3.1 MODIFIED NUCLEOSIDE

To overcome the limited presence of functional groups within the DNA scaffold a modified thymine (T^{Im}) is introduced into the DNA scaffold (Figure 1.7). The reason to choose a thymine as a basis for the modification is based on synthetic reasons. Thymine is the only base without an exocyclic amino function which simplifies the synthetic steps to incorporate the desired modification. Besides that, the modification can be introduced at the place of the methyl-group of thymine, which is always positioned in the major groove. Since the interaction of substrates with our DNA-scaffold takes place at the minor and/or major groove, the presence of our modification is required at this position. In this specific case, the methyl-group of the thymine is replaced by a short linker bearing an imidazole moiety at its end. Imidazole functionalities are capable of general metal coordination, acid-base catalysis and nucleophilic attack in systems as proteins, ribozymes and DNAzymes[32]. In protein-based enzymes the imidazole function can be found in histidine, an amino acid frequently occurring in the active sites of hydrolytic enzymes[33-35]. Therefore, reconsidering the wide

8

variety of possible interactions or catalysis modes, imidazole seems to be the most logical choice.
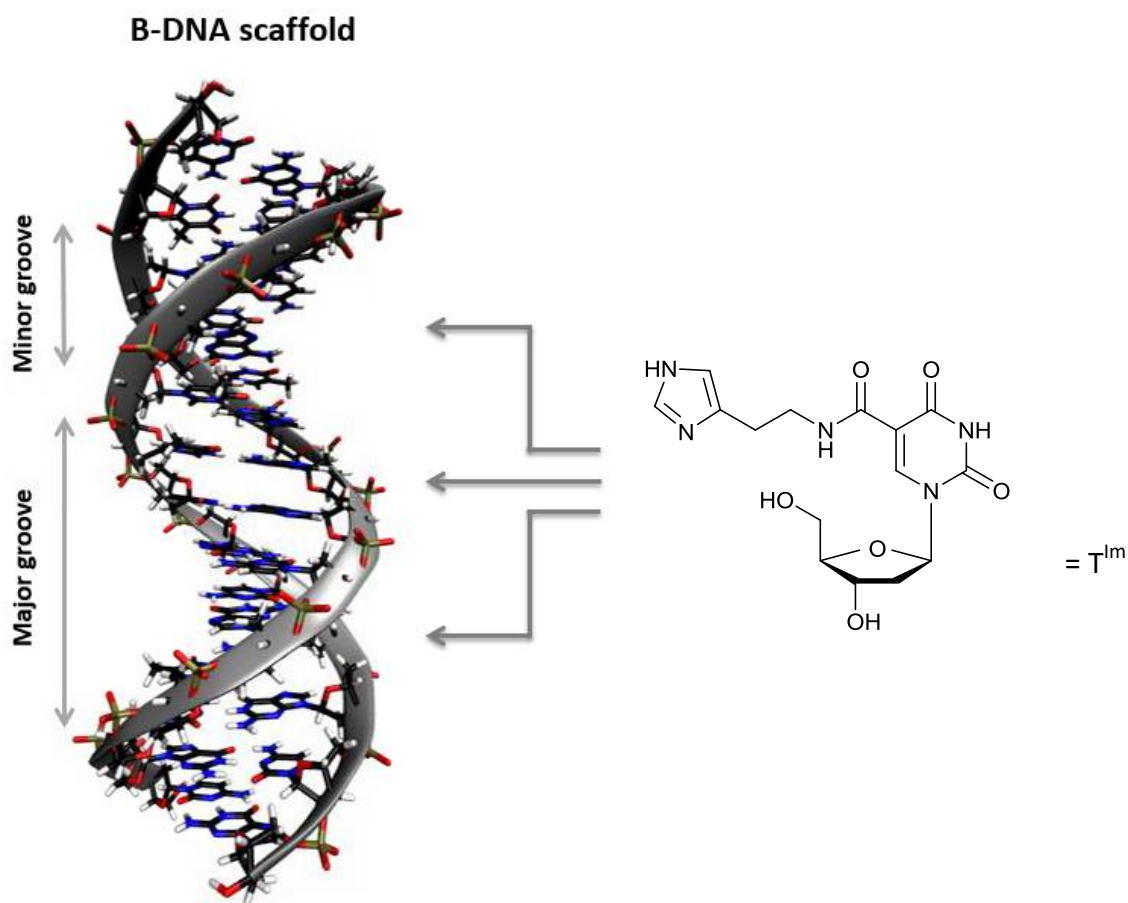
B-DNA scaffold

Figure 1.7. Introduction of the modified thymine building block (T[Im]) in the DNA scaffold of choice. Represented by a model obtained from MD simulations. (Adapted from [3]).

1.3.2 STATE OF THE ART

In earlier systematic studies a DNA scaffold was used[1-3]. This scaffold had to meet the following conditions:

- The sequence must have enough A•T base pairs, since the modification is synthetically limited to thymine building blocks.

- The sequences should have an appropriate length and/or enough G•C base pairs because A•T base pairs are less stable, and also the introduction of the modification can lead to destabilisation.

- The sequence should have sufficient spectral dispersion in terms of ^{1}H signals and should be within reasonable length. This results in reliable UV-Vis melting data and assignable NMR spectra.

A systematic study to identify systems that meet the above conditions resulted in a 14mer sequence. Originally there was a T•T mismatch present at the position of the modified

nucleotide, but this was not requisite to obtain an interaction pattern between the imidazole moiety and the DNA scaffold. An interesting position for T^{ImH} in this wild type sequence proved to be position 8, because only T_8^{ImH+} systems show a persistent and specific hydrogen bond pattern between its imidazole functionality and the DNA scaffold[3]. Figure 1.8 shows this sequence, originally defined by Buyst *et al*[36] with its three base pair cassette of T_8^{ImH+}, G_9 and G_{19}, from now on referred to as the original motif, $T_8^{ImH+}(G_9C_{20})(C_{10}G_{19})$.
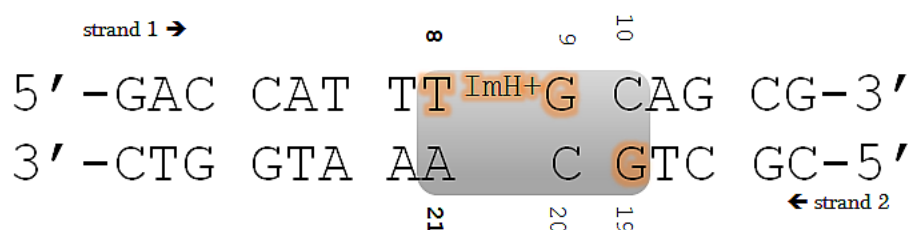
Figure 1.8. The original motif, with the base pair cassette indicated in the grey box.

The original motif is characterized by a pK_{aH} of 8.7 ± 0.02, an increase of 1.5 units compared to the isolated nucleoside building block, and an increase of melting temperature of $5.2°C \pm 0.1°C$ relative to the unmodified duplex sequence. A set of specific nOe-contacts can be found between the imidazole moiety and the guanines at position 9 and 19, which endorses the MD simulations as shown before (Figure 1.1). In general a sequence with a T^{Im} at position n, followed by two guanines at position n+1 (same strand as the modification) and n+2 (opposite strand), will lead to a pK_{aH} regulating motif.

1.4 PROJECT OUTLINE

So far, the position and nature of all nucleotides within the shaded box (Figure 1.8) were considered to contribute to the properties of the pK_{aH} interaction motif. To investigate whether stabilizing interactions of imidazole building blocks are allowed in other combinations within the DNA scaffold, a few permutations of the above mentioned sequence were proposed for *in silico* investigation. This allowed us to examine the robustness and toleration of different base pairs within the stabilizing motif and test whether a similar interaction motif can be obtained without G as the main acceptor. The next sequences (Figure 1.9) are of interest in this Master Thesis:

strand 1 →

$T_8^{ImH+}(A_9T_{20})(C_{10}G_{19})$

```
5'-GAC CAT TT ImH+A CAG CG-3'
3'-CTG GTA AA      T GTC GC-5'  ← strand 2
```

strand 1 →

$T_8^{ImH+}(T_9A_{20})(C_{10}G_{19})$

```
5'-GAC CAT TT ImH+T CAG CG-3'
3'-CTG GTA AA      A GTC GC-5'  ← strand 2
```

strand 1 →

$T_8^{ImH+}(G_9C_{20})(A_{10}T_{19})$

```
5'-GAC CAT TT ImH+G AAG CG-3'
3'-CTG GTA AA      C TTC GC-5'  ← strand 2
```

strand 1 →

$T_8^{ImH+}(G_9C_{20})(T_{10}A_{19})(T_{11}A_{18})$

```
5'-GAC CAT TT ImH+G TTG CG-3'
3'-CTG GTA AA      C AAC GC-5'  ← strand 2
```
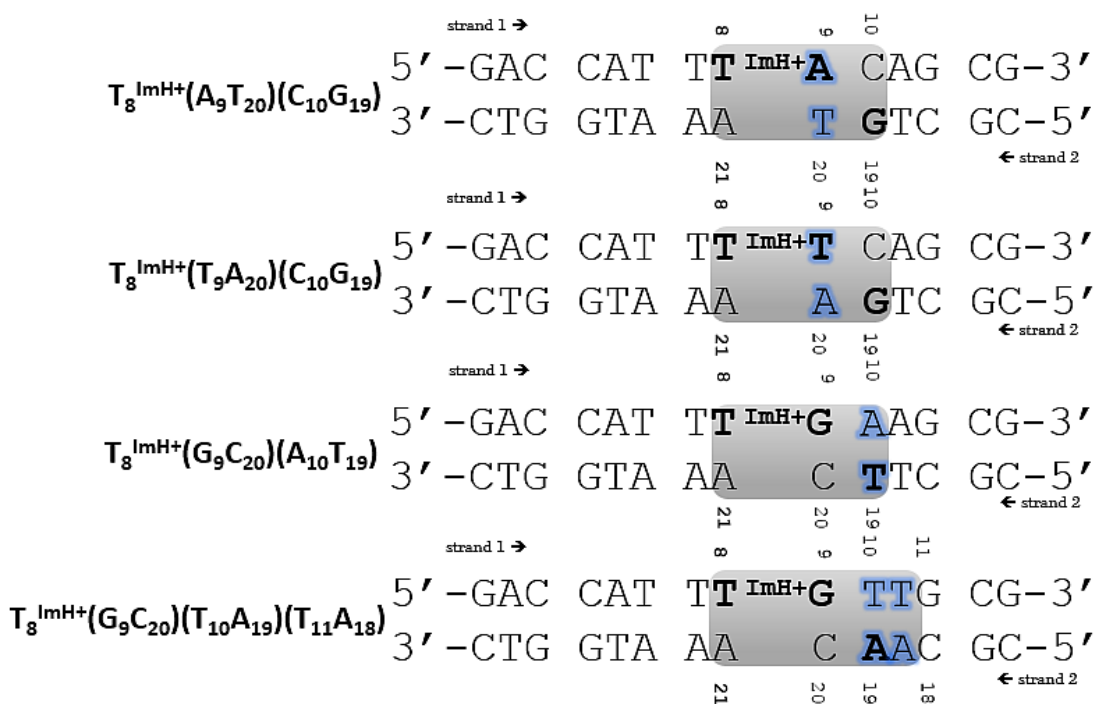
Figure 1.9. The sequences of interest during this thesis and their shortened name.

Investigation of sequences $T_8^{ImH+}(G_9C_{20})(A_{10}T_{19})$ and $T_8^{ImH+}(G_9C_{20})(T_{10}A_{19})(T_{11}A_{18})$ can tell whether a thymine or adenine can act as a hydrogen bond acceptor for the positively charged imidazole group (see 3.2). Sequences $T_8^{ImH+}(T_9A_{20})(C_{10}G_{19})$ and $T_8^{ImH+}(A_9T_{20})(C_{10}G_{19})$ give information about the toleration of an A•T/T•A base pair at position 9 (see 3.3). The sequences involving permutations with G•C base pairs are not considered in this thesis as *in silico* studies have shown these to be of little value, i.e. only a disruption of the original motif could be observed (Buyst *et al*, unpublished results).

The interaction behaviour of the four A•T/T•A sequences was already investigated by making use of MD simulations, but definitive conclusions can only be made when the predictions are backed by sound experimental proof, and this is the main achievement aimed for in this thesis.

The experimental work included the following steps/activities:

- Synthesis of the sequences/duplexes
- 2D NMR measurements and full assignment
- pH-titrations and pK_{aH} determination of the imidazole functionality
- Determination of melting temperatures

In the remainder of this work, the following chapters will be presented:

Chapter 2 **Methods**: this chapter contains a short description of the methods of choice, synthesis details and data analysis techniques.

Chapter 3 **Results and Discussion**: the results of the studied systems based on UV-Vis thermal melting data, NMR assignment and pK_{aH} titrations are represented and discussed.

Chapter 4 **General conclusion and Perspectives**

Chapter 5 **Materials and Chemicals**

Chapter 6 **Protocols**

Chapter 7 **Appendix**: the spectroscopic and chromatographic data is included in this chapter, additionally safety issues are highlighted.

Chapter 8 **Dutch summary**

Chapter 9 **References**

CHAPTER 2 METHODS

2.0 INTRODUCTION

This chapter gives more information on the used techniques and methodologies, in addition to the subsequent data processing steps. More specifically, the synthesis (2.1), thermal melting experiments (2.2) and NMR analyses (2.3) are discussed in the corresponding paragraphs.

2.1 SYNTHESIS

2.1.0 SYNTHESIS OF THE IMIDAZOLIUM BUILDING BLOCK

The synthesis from 5-iodo-2'-deoxyuridine to the modified thymine is a three step process described by Holmes and Gait (Figure 2.1)[37]. The last step is a substitution on the phosphoramidite reagent using DIPEA as a base. One should be careful with the resulting phosphoramidite derivative, because it is sensitive to air, moisture and light.
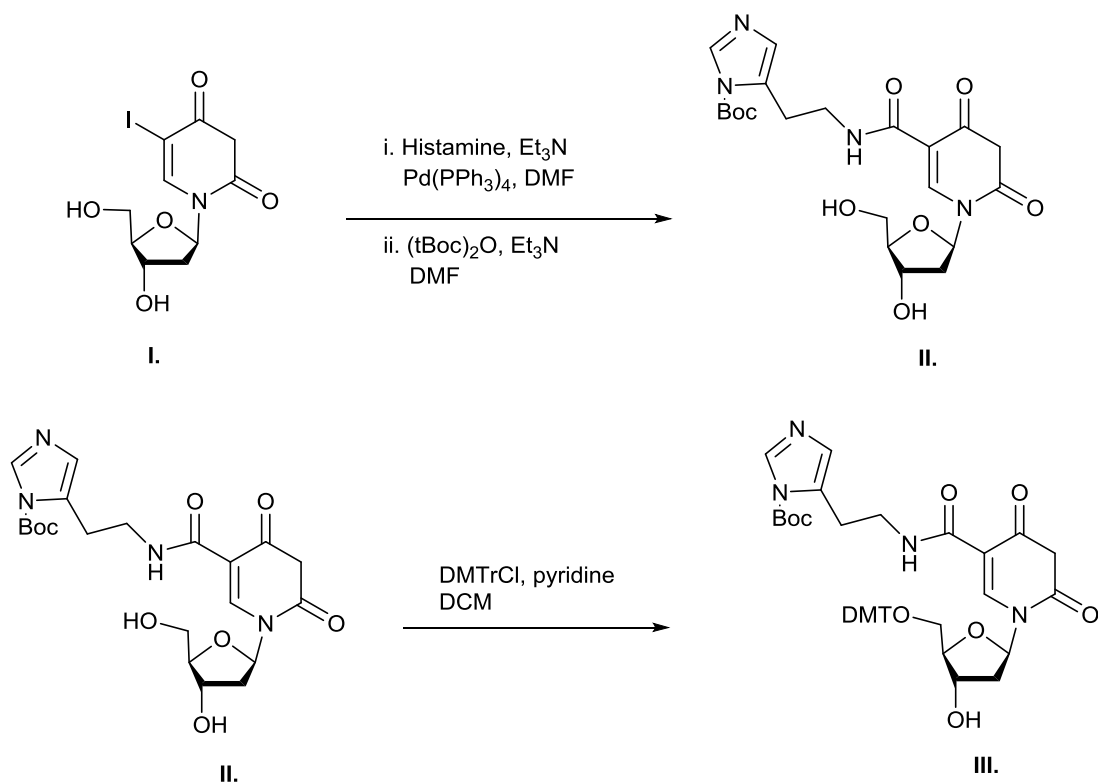
Figure 2.1. Three step synthesis of the modified thymine.

2.1.1 DNA SYNTHESIS

The synthesis of the modified strands is based on an automated solid phase phosphoramidite protocol[38]. Alkyl phosporamidites react as stable 3'-derivatives of nucleosides efficiently and rapidly with nucleoside 5' hydroxyl groups with an azole catalyst present. This results in unstable phosphite triesters which must be oxidized immediately to the stable phosphate triester. Up to 100 nucleosides can be introduced using this iterative synthesis cycle. Figure 2.2 shows the different steps which include detritylation (removal of the 5'-DMTr group), this is carried out with trichloroacetic acid in dichloromethane. Activation of the phosphoramidite (Step 1) occurs when it is mixed with 4,5-dicyanoimidazole in an acetonitrile solution as coupling agent. Step 2 is the addition of the activated phosphoramidite to the growing chain. A safety step is included, called capping, to block the chains that have not reacted during the coupling reaction and this limits the number of deletion sequences. Oxidation of the intermediate phosphite to the phosphate triester is achieved with iodine and water in THF to obtain the stable phosphate triester[39].
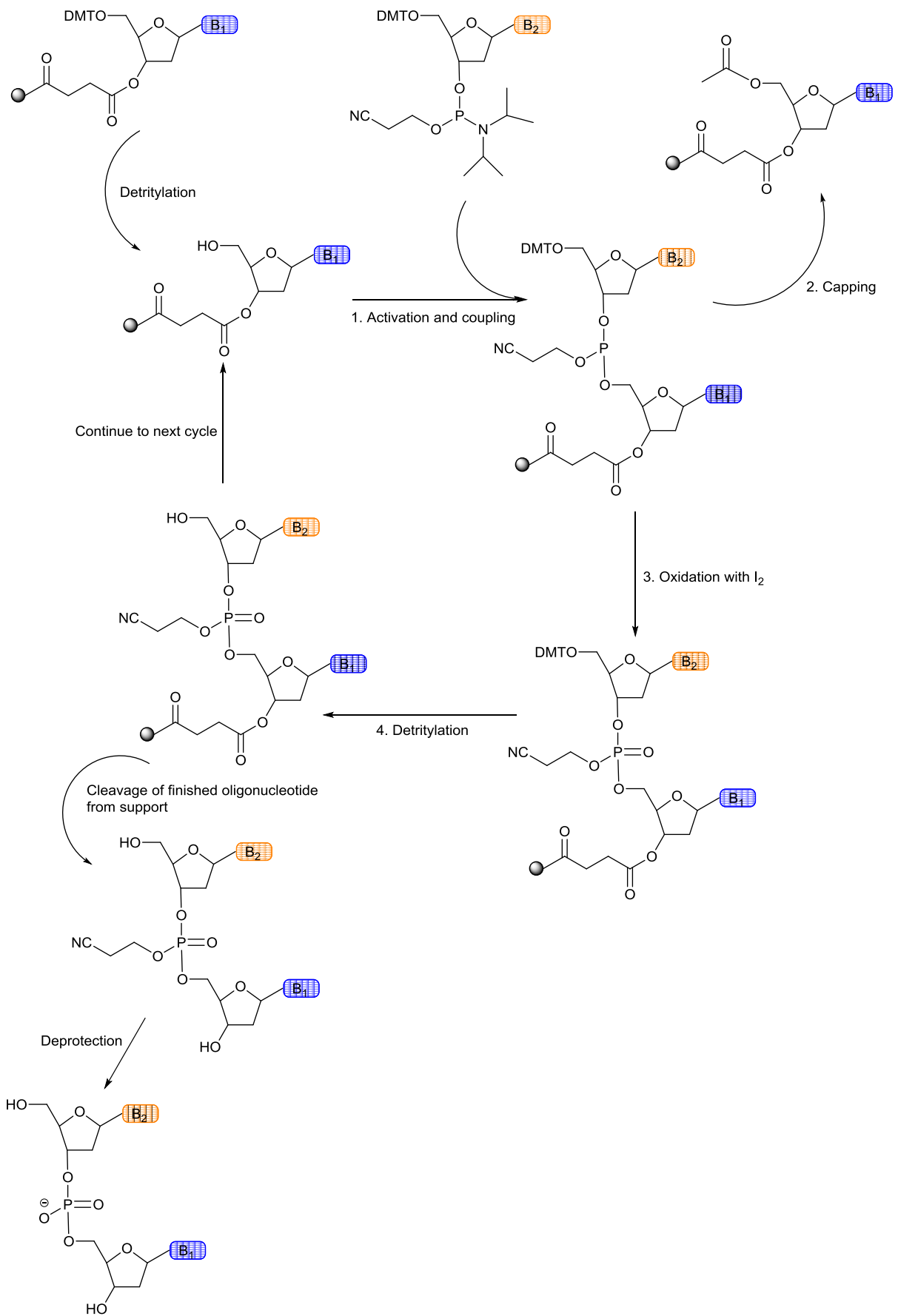
Figure 2.2. Phosphoramidite DNA synthesis. (Adapted from [40]).

The coupling of the nucleotides can be followed through visual control of the trityl cation (Figure 2.3) released after the detritylation step, because of its orange colour.
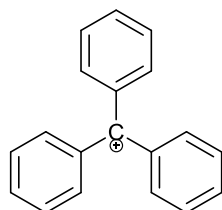
Figure 2.3. Trityl cation.

2.2 THERMAL MELTING CURVES

The stability of the duplexes is determined by considering the thermal melting temperature T_m. The sequences, in duplex form, are exposed to several heating and cooling cycles to obtain a reliable T_m (see Figure 2.4 for three typical heating curves). At this temperature the duplex is 'melted' for ± 50%. More precisely an equilibrium comes between the folded- and unfolded state, because with increasing temperature a transition is induced to the unbound state.

$$Folded\ state\ (duplex) \underset{T\searrow}{\overset{T\nearrow}{\rightleftarrows}} Unfolded\ state \qquad\qquad Eq.\ 2.1$$
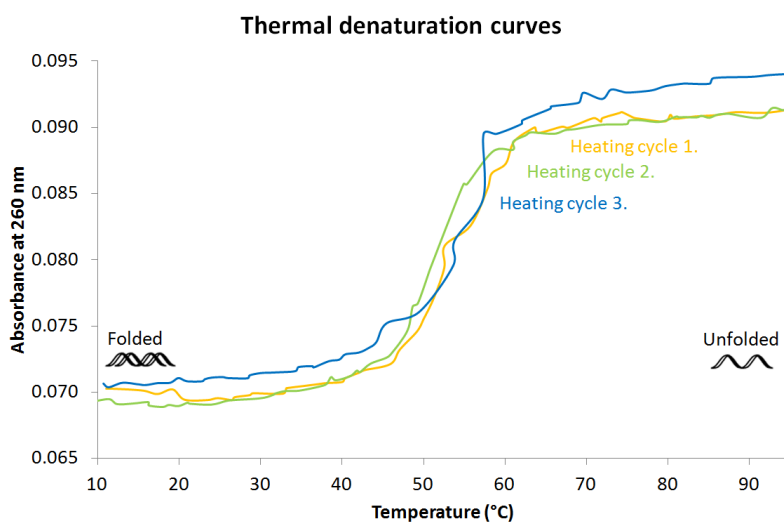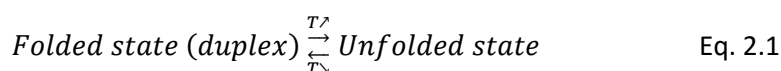
Figure 2.4. Thermal melting curves which are used to determine the melting temperature of the DNA duplexes.

A series of practical issues have to be considered during this melting process, for instance the pH, the concentration of salts, the heating rate and the concentration of the duplexes have to be controlled. This is of importance because T_m depends on all of these parameters. All of the measurements must be performed under identical conditions to make mutual comparison possible.

16

Two methods exist to determine the melting temperature based on these absorbance spectra. On the one hand there is a simple and user-independent method which derives the T_m from the maximum of the first derivative of the denaturation curves. The only drawback of this method is that the value obtained is not always the T_m[42]. On the other hand one can use the baseline method where an upper and lower baseline is fitted to the absorbance curve. The next step is to draw a median line between these two baselines. The point where this median intersects the absorbance curve, corresponds to the T_m. The second method comes with an increased uncertainty, because baseline determination is in most cases subjective. Figure 2.5 shows the different methods.
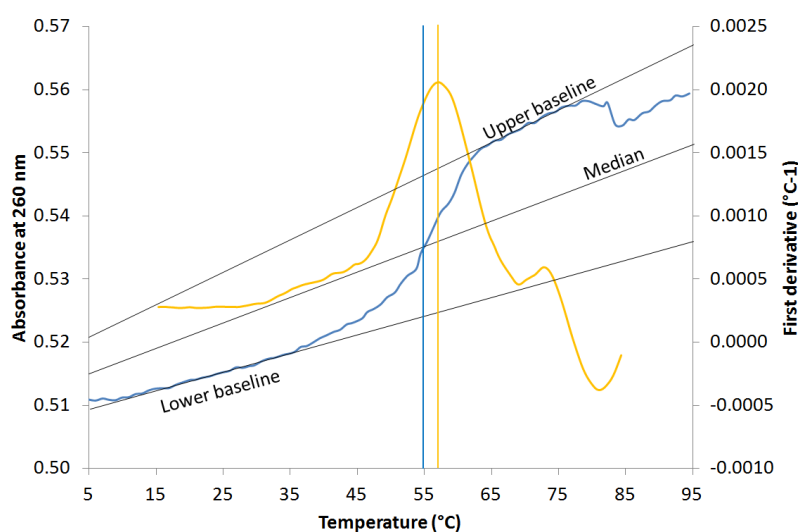
Figure 2.5. On an average thermal denaturation curve both the baseline method and the determination of the T_m via the first derivative are shown.

The T_m determined by the two methods is slightly different, sometimes up to 1.5°C. We here chose to work with the first derivative method, because it is more reliable and less subjective.

2.3 NMR

In this work both 1D and 2D NMR techniques will be used to investigate the systems of interest. The 2D techniques include NOESY and TOCSY spectra to elucidate the internucleotide contacts and deoxyribose protons respectively. Beside these assignment methodologies of the DNA sequence, ^{1}H spectra are used to determine the pK_{aH} value of the imidazole moiety.

2.3.0 FULL ASSIGNMENT

After recording the 2D spectra, the actual assignment is carried out[43]. The specific hydrogen atoms of the DNA units are characterized by unique chemical shifts (Table 2.1).

Table 2.1. Overview of the chemical shift distribution (adapted from [44]).

		Δ (ppm)	Comments
Sugar	2', 2''	1.8 - 3.0	
	4', 5', 5''	3.7 - 4.5	All hydrogen atoms of the sugar ring
	3'	4.4 - 5.2	
	1'	5.3 - 6.3	
Base	CH_3	1.2 - 1.6	Thymine
	5	5.3 - 6.0	Cytosine
	6	7.1 - 7.6	Thymine or cytosine
	2, 8	7.3 - 8.4	Guanine or adenine
	NH_2	6.6 - 9.0	Adenine, cytosine or guanine
	NH	10.0 - 15.0	Guanine or thymine

In Figure 2.6 the nOe-contacts induced by through space interaction between the specific hydrogen atoms is shown.

18
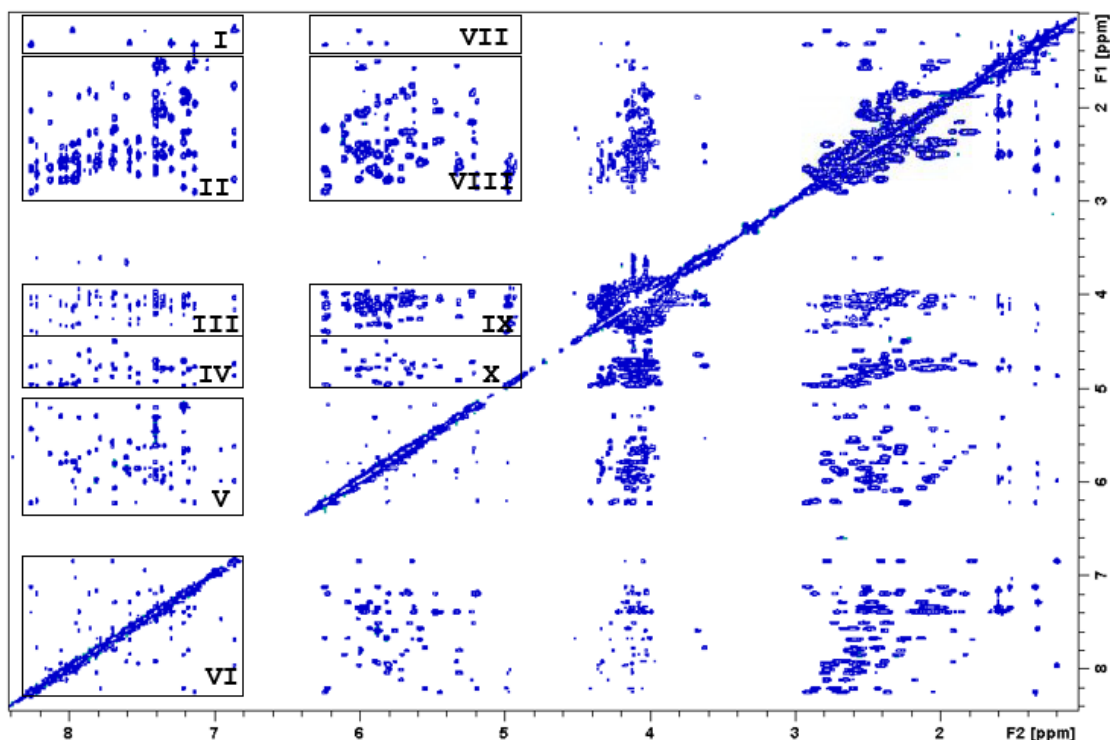
Figure 2.6. A 2D-NOESY spectrum of the WT sequence $(G_9C_{20})(A_{10}T_{19})$ with the characteristic checkerboard motif. Interaction regions are specified, (I). H7*-H6/H8 (II). H2'/H2''-H6/H8 (III). H5'/H4'-H6/H8 (IV). H3'-H6/H8 (V). H1'-H6/H8 (VI). H6/H8-H6/H8 (VII). H7*-H1' (VIII). H2'/H2''-H1' (IX). H4'/H5'-H1' (X). H3'-H1'. (D_2O, 25°C, 700 MHz).

A 2D-NOESY experiment is of great interest, because it allows for the sequential assignment of all resonances to the individual nucleotides in each strand. This is obtained by exploiting through space interactions and resulting intranucleotide nOe contacts between the H1'of the sugar ring and the H6 (in case of thymine and cytosine) or H8 (in case of adenine or guanine) of the associated heteronuclear bases. Because of the base-stacking in a duplex, additional through space contacts also exist between the H6/H8 and the H1' of the subsequent nucleotide in the sequence. Therefore, except for the nucleotides at the ends, every H1' is connected to two successive nucleotide bases. In this way, starting with H1'of G1, a so-called 'sequential walk' to G14 is possible and the same applies for the opposing strand, C15 to C28[45], as shown in Figures 2.6 and 2.7 respectively.
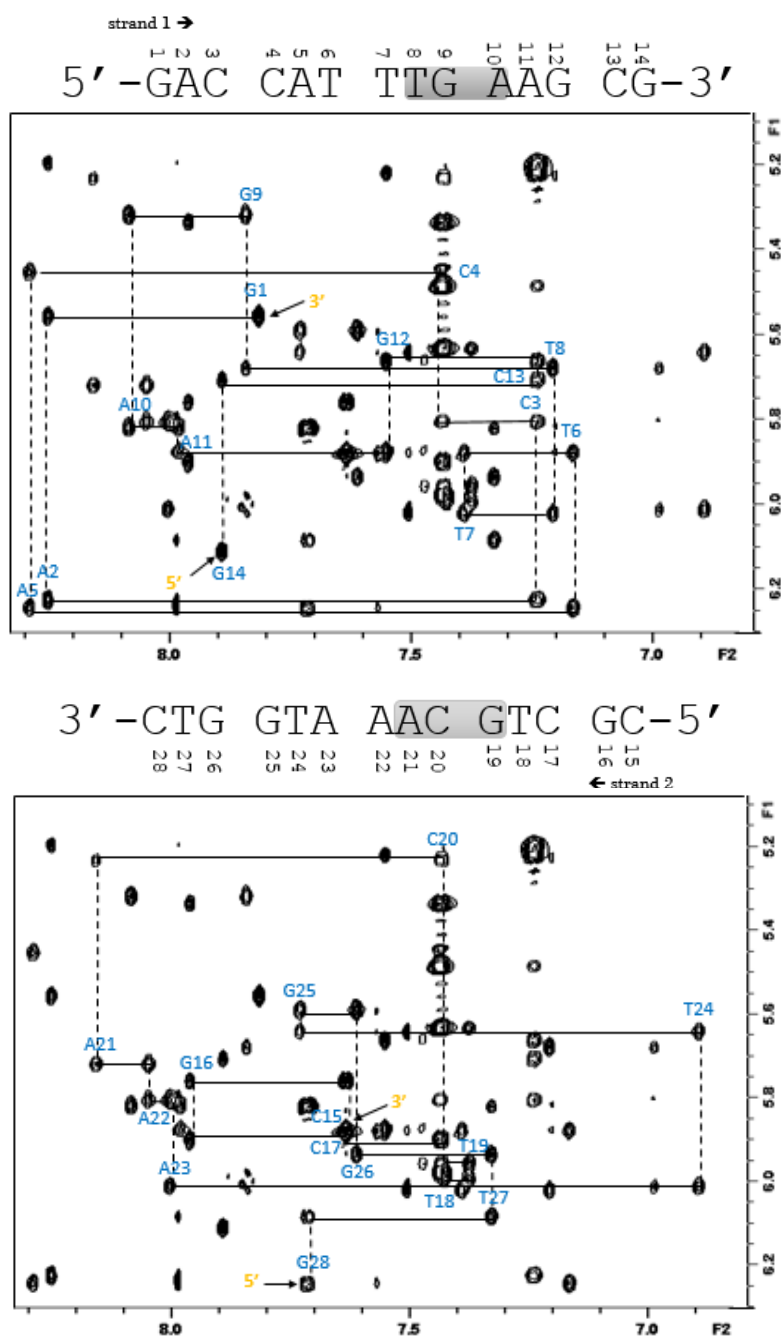
Figure 2.7. Detail of the 2D NOESY spectrum with the assignment of the H1' and H6/H8 region of WT sequence $(G_9C_{20})(A_{10}T_{19})$ by a sequential walk through the strands. (Mixing time 200 ms, D_2O, 25°C, 700 MHz).

The sequential walk through the strands is not restricted to the H1'-H6/H8 region, but is also present in the H2'/H2''-H6/H8 region. The similarity between both regions gives a first internal control for the executed assignment. Secondly a 2D-TOCSY spectrum is used to check the assignment. For instance, the intra nucleotide nOe between H1'and H3' should also show a through chain coupling in a TOCSY spectrum. In this way the deoxyribose protons belonging to one spin system (nucleotide) can be grouped. Based on the information of both the NOESY and TOCSY spectra, a full assignment can be obtained.

For both the modified and non-modified, wild-type (WT) sequences of all the proposed systems 2D-NMR spectra were recorded and assigned, giving the opportunity to graphically represent the influence of the modification on the chemical shift. The absolute value of the chemical shifts of the modified systems is subtracted from those of the non-modified (wild type) sequences. As an example, Figure 2.8 shows the perturbation of both strands caused by the introduced imidazole at position 8. Only perturbations above 0.1 ppm are considered to be significant (dotted lines in figure below). Remarkable is the large change in chemical shift of the H6 of the modified thymine and the methyl group of the thymine, at position 18 (located at the opposing strand compared to the modification). This methyl group of T_{18} can be used as chemical reporter for the interaction between the imidazole modified thymine and the guanine at position 19.
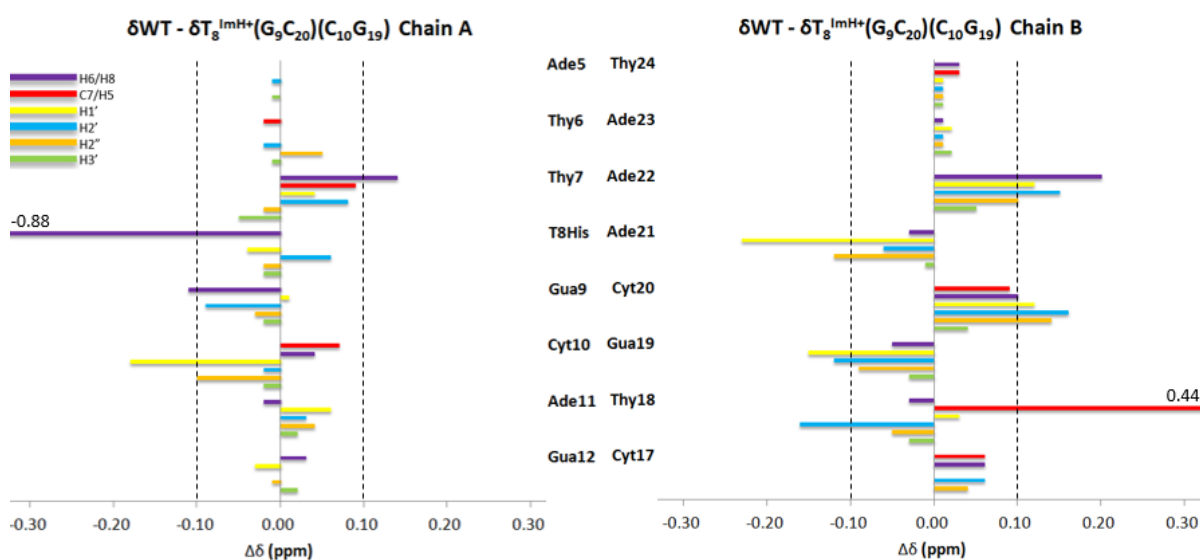
Figure 2.8. Chemical shift perturbation mapping of the original motif with respect to the non-modified WT sequence. Only perturbations with absolute value > 0.1 ppm (dashed lines) are considered to be significant. Where necessary the corresponding perturbation value has been added.

2.3.2 NOE CONTACTS

With respect to the nature of NOESY experiments, the position and behaviour of the imidazole can be followed. If the imidazole protons ε_1 and/or δ_2 stay long enough in close proximity (< 0.5 Å) to the protons of the DNA scaffold, this should result in a nOe-contact (Figure 2.9). After the assignment of the spectrum and the specific contacts of the imidazole protons with the DNA strand, it then becomes clear that our imidazole functionality shows a persistent conformation within the sequence. The proximity can, for instance, be explained by the existence of stabilizing interactions of the imidazole with the DNA scaffold, in contrast

to exposure to the solvent. The reason why only ε_1 is present on Figure 2.9 is the inherent difficulty to detect the δ_2 peak, due to overlap with other signals between 7.0 and 8.0 ppm.
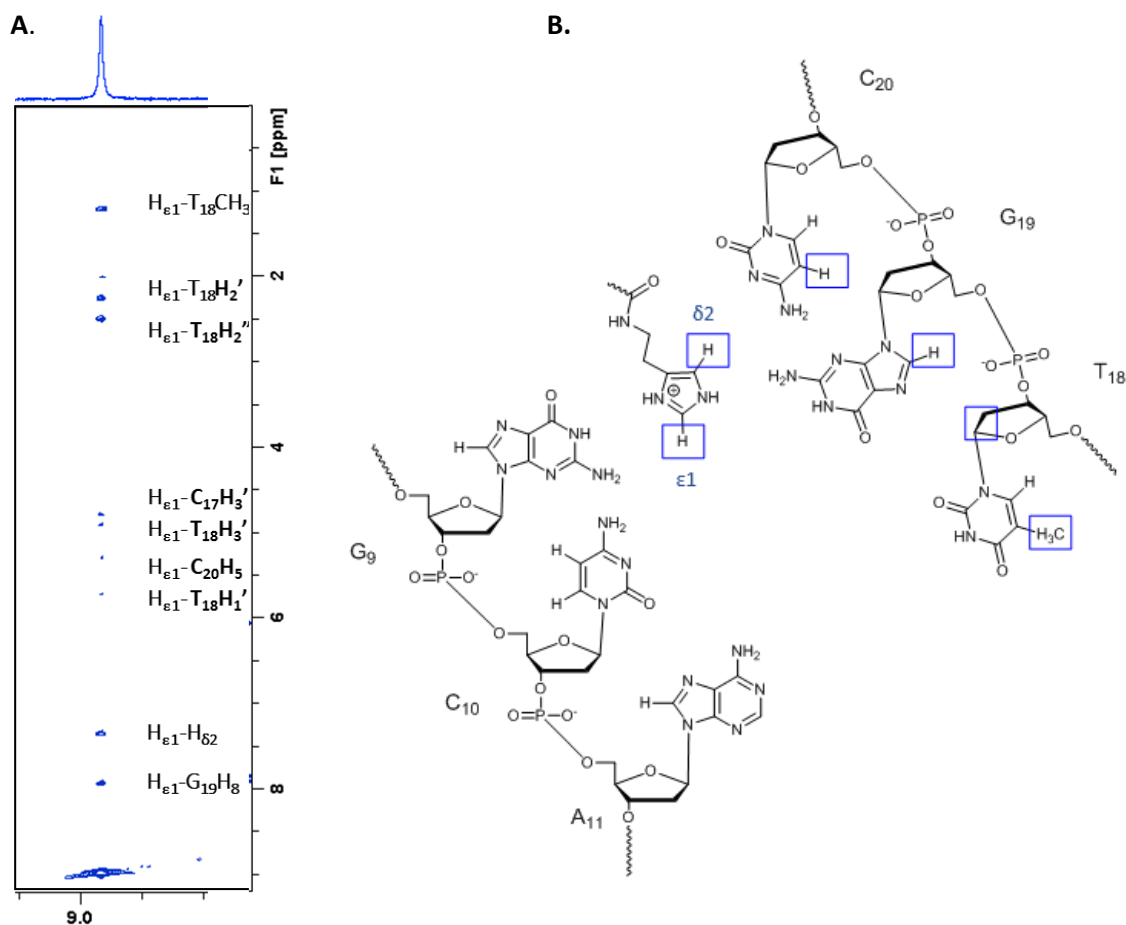
Figure 2.9. (A) Zoom of a 2D-NOESY spectrum of the original motif $T_8^{ImH+}(G_9C_{20})(C_{10}G_{19})$ with specific nOe-contacts of ε_1. (Mixing time of 200 ms, D_2O, 25°C, 700 MHz). (B) A structural representation of the interaction partners of the ε_1 peak (indicated in blue boxes). (Based on data of Buyst *et al*[3]).

2.3.3 PK$_{AH}$ TITRATIONS

Both the ε_1 and δ_2 proton are sensitive to the protonation state of the δ_1 atom of the imidazole moiety and can be used to detect the pK$_{aH}$. As can be seen on Figure 2.10A the ε_1 proton is nearer to the δ_1 atom, but the chemical shift of both protons can be used to fit the pK$_{aH}$ based on an S-shaped curve[46-48]. In both acidic and basic pH region the chemical shift of the ε_1 and δ_2 proton is followed, after which the S-shaped curve can be drawn. To enable identification of the ε_1 and δ_2 peak in the crowded ^1H spectra, a weak $^4J_{HH}$-coupling between the ε_1 and δ_2 peak should be detected in a 2D-TOCSY experiment (Figure 2.10B). Starting with a TOCSY experiment at low pH, a titration in the acidic region is executed following the shift in ppm of the peak of both the ε_1 and δ_2 peak. The same approach is executed in the basic region.

22
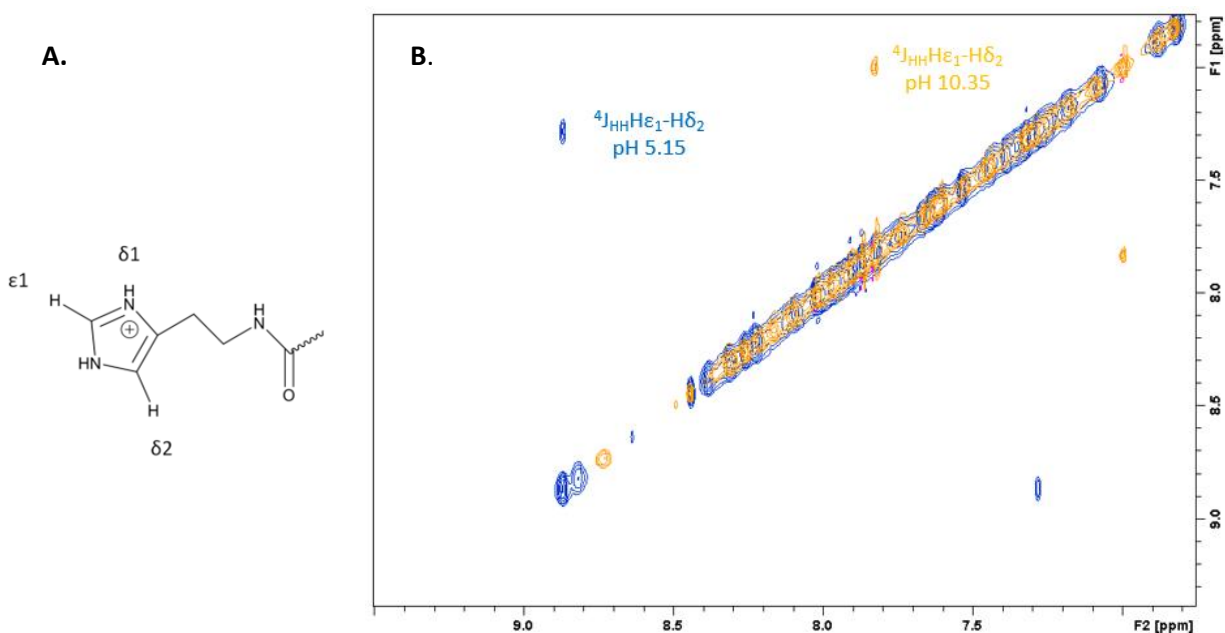
Figure 2.10. (A) Structure of the protonated imidazole moiety . (B) Overlay of two homonuclear 2D-TOCSY experiments (mixing time 60 ms) to detect the ε1 and δ2 peak in both acidic (blue) and basic (orange) region. Only the aromatic region of the spectra is shown. (H_2O/D_2O, 25°C, 700 MHz).

A figure detailing the shift of the ε_1 peak in acidic region with increasing pH is shown in Figure 2.11 (boxed area). This figure not only illustrates the shift but also the broadening of the imidazole peak with increasing pH. Above 7.42 the peak will disappear completely, only to reappear at higher pH within the DNA signal background.
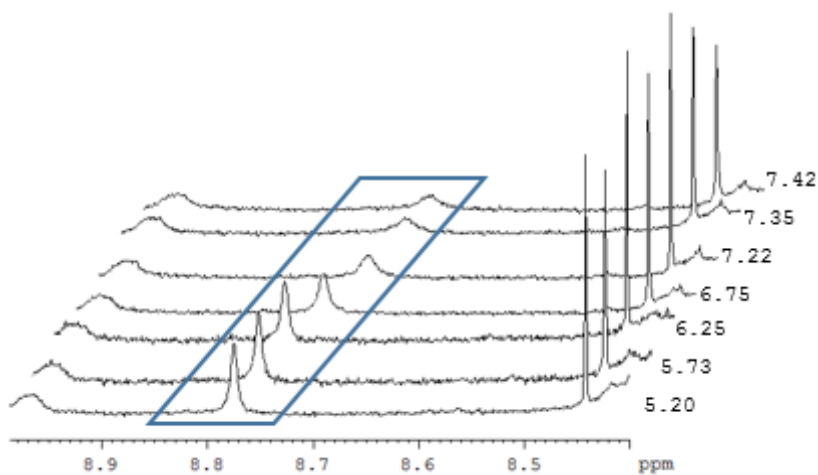
Figure 2.11. Overlay of multiple ^{1}H spectra at different pH values in the acidic region. (H_2O/D_2O, 25°C, 700 MHz).

Plotting the chemical shift versus the pH value of both acidic and basic regions, an S-shaped curve is obtained (Figure 2.12). By fitting the Henderson-Hasselbalch equation (see equation 2)[49] to this data the pK$_{aH}$ is found.

$$\delta_{obs} = \delta_{pr} - \frac{(\delta_{pr}-\delta_{depr})}{1+10^{(pK_{aH}-pH)}}$$ Eq. 2.2

Where δ_{obs} represents the observed chemical shift, δ_{pr} the chemical shift of the protonated state (at low pH) and δ_{depr} the chemical shift of the deprotonated state (at high pH).
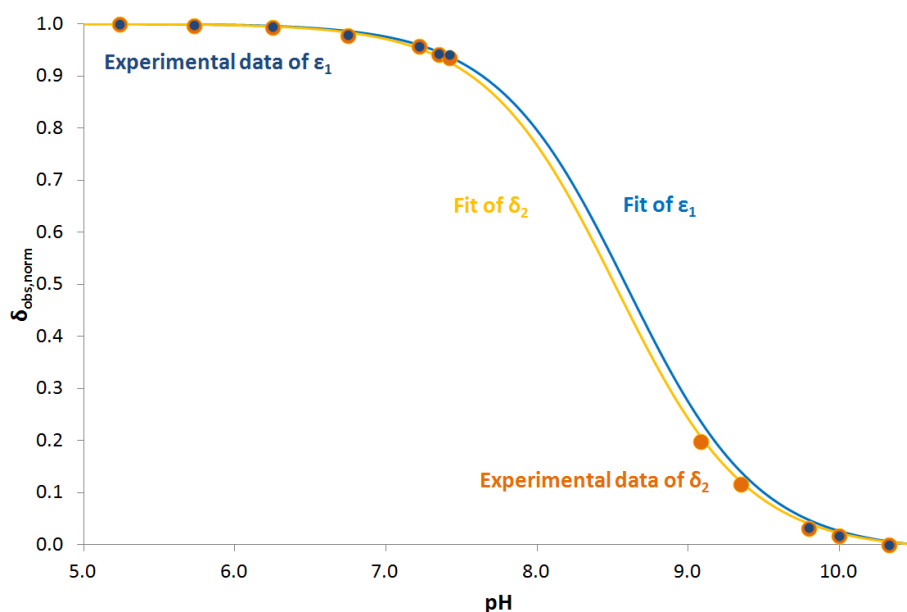
Figure 2.12. Normalized observed chemical shifts as a function of pH for the sequence $T_8^{ImH+}(G_9C_{20})(T_{10}A_{19})(T_{11}A_{18})$. Both the experimental data points and fitted curves are shown, the corresponding calculated pK_{aH} value is 8.5 ± 0.03.

More experimental points for the δ_2 protons were determined giving more data and a decreased fitting error. For this reason I chose to represent in this thesis only pK_{aH} values based on δ_2 curves.

Chapter 3 Results and Discussion

3.0 Introduction

In this section the results are presented which allow to answer the research question: "How robust is the original motif $T_8^{ImH+}(G_9C_{20})(C_{10}G_{19})$ towards permutations of the base pairs present at position 9•20 and 10•19?". Figure 3.1 summarizes the sequences of interest and their shortened name that will be used throughout the text. The sequences involving permutations with G•C base pairs are not considered in this thesis as *in silico* studies have shown these to be of little value, i.e. only a disruption of the original motif could be observed.
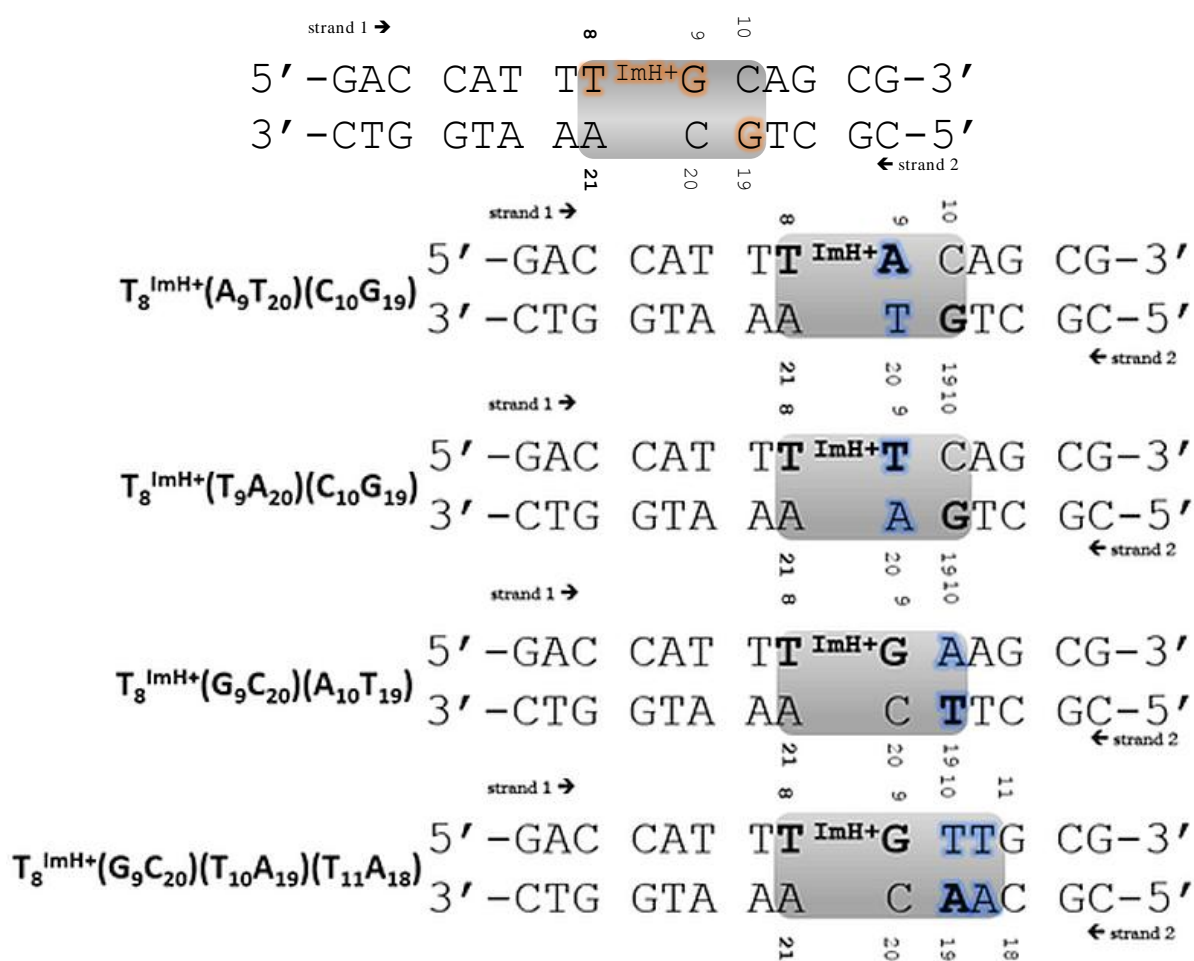
Figure 3.1 The original motif (top) and the four permutated sequences of interest during this work. The permutations relative to the original motif (top) are indicated in blue.

3.1 Is an A•T/T•A base pair at position 9 tolerated by the specific interaction pattern?

The original motif and the base pair modifications to the motif considered in this section are shown in Figure 3.2.
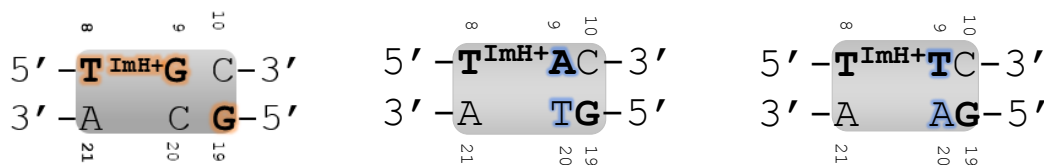
Figure 3.2. The original motif and the sequences with an A•T/T•A base pair at position 9. The base pairs that are altered with respect to the original interaction motif are indicated in blue.

The pH titration curves to characterize the pK_{aH} for both studied sequences were recorded and are shown in Figure 3.3 together with the titration curves for the T^{ImH+} building block and the original motif $T_8^{ImH+}(G_9C_{20})(C_{10}G_{19})$ reported previously[1,3,36].
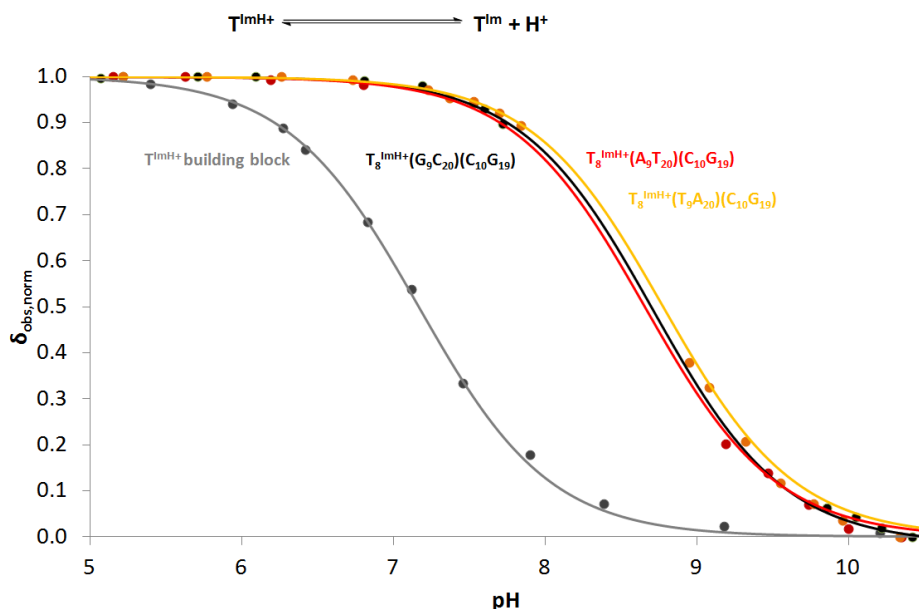
Figure 3.3. Normalized observed δ_2 chemical shifts as a function of pH for the isolated nucleoside building block (grey), sequences $T_8^{ImH+}(A_9T_{20})(C_{10}G_{19})$ (red), $T_8^{ImH+}(T_9A_{20})(C_{10}G_{19})$ (orange) and the original motif $T_8^{ImH+}(G_9C_{20})(C_{10}G_{19})$ (black). Both the experimental data points and fitted curves are shown. T^{Im} and T^{ImH+} refer to the neutral and positively charged state of the imidazole functionality respectively.

Table 3.1 gives an overview of the melting temperatures (T_m) and the pK_{aH} values. To assist in the interpretation, ΔT_m values are additionally included. ΔT_m values are calculated for the unmodified systems relative to the values for the unmodified sequence of the original motif (reported in shaded cells) while ΔT_m values for the imidazole modified sequences are all relative to those for the corresponding unmodified counterparts. Finally the ΔpK_{aH} values relative to the T_8^{ImH+} nucleoside building block are included.

First, a decrease in T_m ($\Delta T_m < 0$, shaded cells) of the unmodified sequences $(A_9T_{20})(C_{10}G_{19})$ and $(T_9A_{20})(C_{10}G_{19})$ is observed compared to $WT(G_9C_{20})(C_{10}G_{19})$. This can be related to the introduction of the less stable A•T or T•A base pair at position 9 compared to the original G•C base pair.

Table 3.1. Melting temperature and pK_{aH} overview of the sequences with a T•A and A•T pair at position 9. [a] T_m values were determined in 100 mM NaCl, 10 mM phosphate buffer at pH = 7. [b] Not relevant since T^{ImH+} represents the imidazole nucleoside building block and not a duplex based system, hence there is no melting process. [c] The wild type sequences has no imidazole functionality.

System	Building Block T^{ImH+}	Original motif		A•T base pair		T•A base pair	
		$WT(G_9C_{20})$ $(C_{10}G_{19})$	$T_8^{ImH+}(G_9C_{20})$ $(C_{10}G_{19})$	(A_9T_{20}) $(C_{10}G_{19})$	$T_8^{ImH+}(A_9T_{20})$ $(C_{10}G_{19})$	(T_9A_{20}) $(C_{10}G_{19})$	$T_8^{ImH+}(T_9A_{20})$ $(C_{10}G_{19})$
T_m (°C)[a]	n.r.[b]	58.9 ± 0.2	64.1 ± 0.06	50.8 ± 0.03	60.5 ± 0.13	54.5 ± 0.6	56.4 ± 0.07
ΔT_m (°C)	n.r.[b]	ref	5.2 ± 0.6	-8.1 ± 0.2	9.7 ± 0.1	-4.4 ± 0.6	1.9 ± 0.6
pK_{aH}	7.2 ± 0.02	/[c]	8.7 ± 0.02	/[c]	8.7 ± 0.07	/[c]	8.8 ± 0.02
ΔpK_{aH}	Ref		1.5 ± 0.03		1.5 ± 0.07		1.6 ± 0.03

In general, introducing an imidazole can cause a stabilizing effect due to the fact that at pH 7 the imidazole functionality is at least partially existing in the protonated imidazolium form (Figure 3.4) and can interact with the negatively charged phosphodiester backbone thus causing an electrostatic stabilisation in the order of 1-2°C[3].
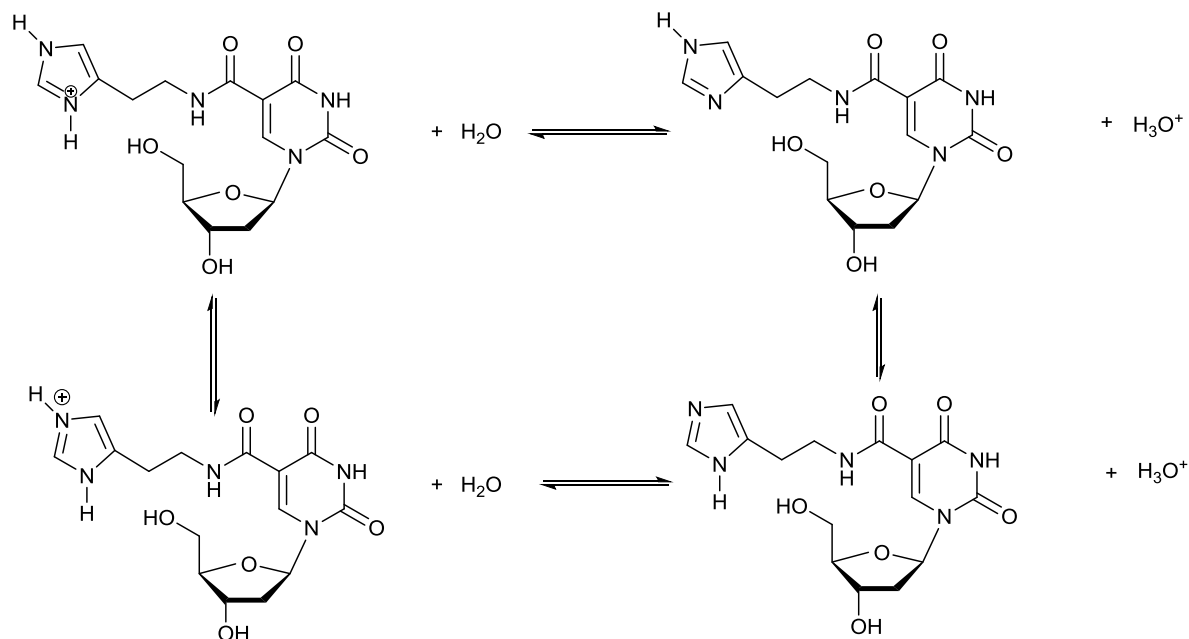
Figure 3.4. The tautomeric equilibrium is present from top to bottom (both left and right sight) and the protonation equilibria are present from left to right (both top and bottom).

In addition, in earlier work[36] it was shown that the original motif $T_8^{ImH+}(G_9C_{20})(C_{10}G_{19})$ increased the T_m by 5.2°C compared to the unmodified sequence $(G_9C_{20})(C_{10}G_{19})$.

Here, the change in T_m when introducing T^{ImH+} at position 8 in duplexes featuring an A•T or T•A at position 9 clearly diverges from this. Either a less pronounced increase (+1.9°C for $T_8^{ImH+}(T_9A_{20})(C_{10}G_{19})$) or a more pronounced increased one (+9.7°C for $T_8^{ImH+}(A_9T_{20})(C_{10}G_{19})$) is evident from the data. Based on melting temperature alone we do not expect the presence of the pK_{aH} specific interaction pattern for $T_8^{ImH+}(T_9A_{20})(C_{10}G_{19})$, while its presence appears for the $T_8^{ImH+}(A_9T_{20})(C_{10}G_{19})$ system, but with stronger stabilisation than observed before.

Second, the value of the pK_{aH} in the A•T and T•A sequences is almost identical to that of the original motif $T_8^{ImH+}(G_9C_{20})(C_{10}G_{19})$ (see Figure 3.3) with an increase of 1.5-1.6 units, showing that no real impact on the pK_{aH} motif is caused by the introduction of an adenine or thymine at position 9, i.e. it appears present in both. Thus, there appears to be no clear correlation between ΔT_m and the pK_{aH} value, since complete opposite changes in T_m does not result in a clear differentiation of pK_{aH}.

Table 3.2. H bond persistence of the original motif $T_8^{ImH+}(G_9C_{20})(C_{10}G_{19})$ and the three other sequences with different base pairs at position 9•20. (Adapted from[3]).

Sequence	Donor	Acceptor	Persistence (%)
$T_8^{ImH+}(G_9C_{20})(C_{10}G_{19})$	T_8^{ImH+} Hε_2	G_{19} O6	27.1
	T_8^{ImH+} Hε_2	G_9 O6	1.5
$T_8^{ImH+}(A_9T_{20})(C_{10}G_{19})$	T_8^{ImH+} Hε_2	G_{19} O6	3.0
$T_8^{ImH+}(T_9A_{20})(C_{10}G_{19})$	T_8^{ImH+} Hε_2	G_{19} O6	27.7
	T_8^{ImH+} Hε_2	G_{19} N7	0.21
	T_8^{ImH+} Hδ_1	T_9 O4	0.35
$T_8^{ImH+}(C_9G_{20})(C_{10}G_{19})$	T_8^{ImH+} Hε_2	G_{19} O6	1.04
	T_8^{ImH+} Hε_2	G_{19} N7	0.56

To assist in interpretation of these results, we turned to MD simulations of these sequences performed in previous work[3]. Here, the persistence of short distances required for successful interaction of T^{ImH+} with the Hoogsteen side of G_{19} is extracted from a 50 ns explicit solvent simulation of the particular sequence (Table 3.2). These are confronted with the presence (or absence) of nOe data connecting the imidazole moiety to the DNA scaffold. Using these we are capable to compare the experimental data with the analyses of the *in silico* experiments. From the MD trajectories, the behaviour of $T_8^{ImH+}(T_9A_{20})(C_{10}G_{19})$ was expected to be similar to the original motif $T_8^{ImH+}(G_9C_{20})(C_{10}G_{19})$ and the experimental NMR

data confirms this. The specific nOe contact interaction pattern (Figure 3.5) and the chemical shift perturbation figures (see Appendix) appear to be very similar to those observed in the original motif, and results in almost identical pK_{aH} value. Nevertheless, the increase in T_m is significantly smaller than what was observed for the original motif $T_8^{ImH+}(G_9C_{20})(C_{10}G_{19})$.

In contrast, based on a first simulation run, the interaction pattern of $T_8^{ImH+}(A_9T_{20})(C_{10}G_{19})$ was expected to be strongly decreased (almost absent) in persistence (3% see Table 3.2). From the simulation, the lack of interaction cannot be allocated to steric hindrance, but is interpreted to be a result of change in the local electrostatic surface. It appears that the exocyclic amino groups of A_9 and C_{10} taken together effectively form a predominantly positively charged local electrostatic field, that could cause repulsion of the positively charged imidazole functionality[3]. The experimental data however, proves different. The pK_{aH} value is identical to the one of the original motif and the interaction pattern, as can be deduced from the nOe-contacts (Figure 3.5) and chemical shift perturbation mapping figures (Appendix) show also the similarity with the original motif. In addition, there is a high increase in melting temperature. Therefore, the change in electrostatic surface caused by the introduction of the adenine appears insufficient to perturb the interaction. We conclude from these observations that, in short, the base pair at position 9 need not to be a guanine, a thymine and even an adenine equally allowing for the interaction leading to the pK_{aH} motif.
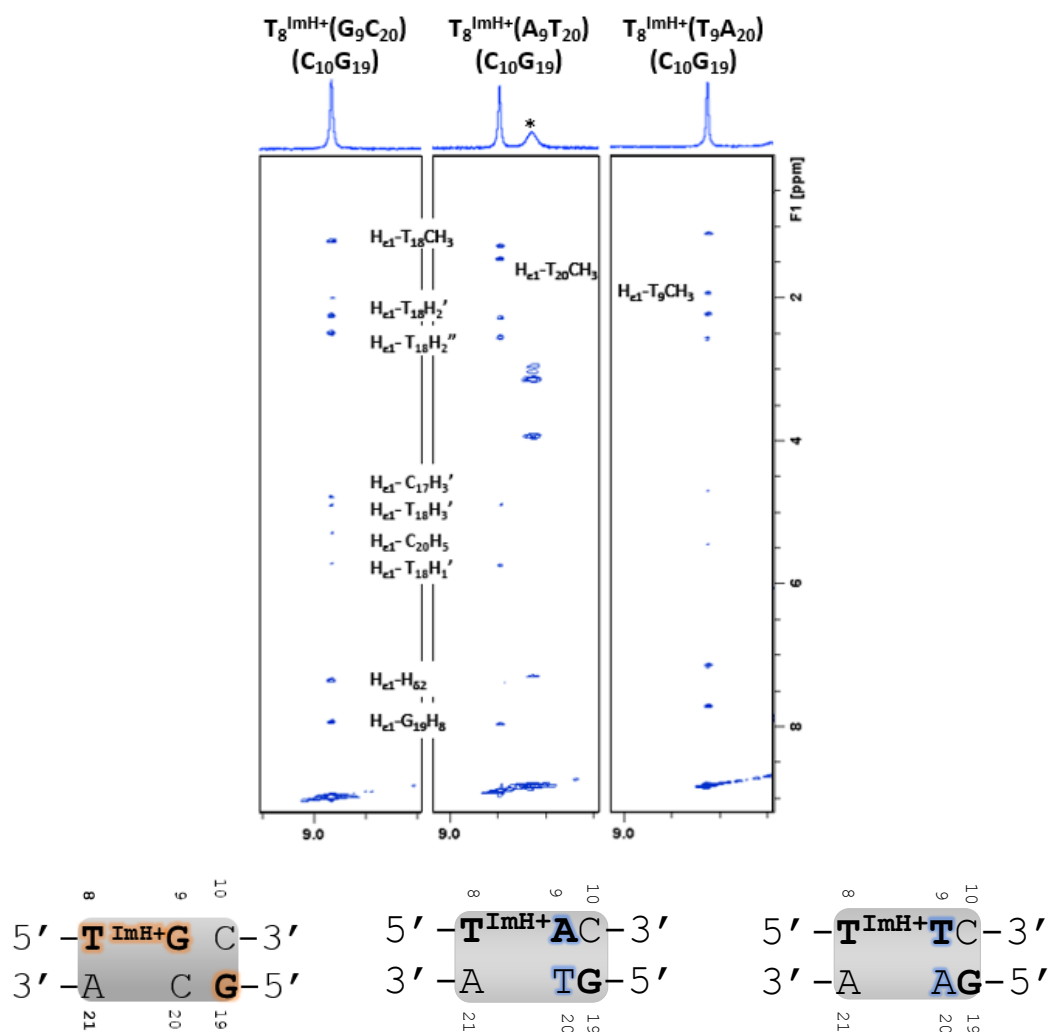
Figure 3.5. At the top an overview of the specific nOe contacts starting from the Hε$_1$ proton for the $T_8^{ImH+}(G_9C_{20})(C_{10}G_{19})$, $T_8^{ImH+}(A_9T_{20})(C_{10}G_{19})$ and $T_8^{ImH+}(T_9A_{20})(C_{10}G_{19})$ sequences to the DNA non-exchangeable protons is present. *corresponds to the exchangeable amide linker proton of the T^{ImH+} functionality. (mixing time 200 ms, D_2O, pD 6, 25°C, 700 MHz). The original motif and the permutated sequences (bottom).

So far, three of the four possible base pair combinations at position 9•20 have been considered. A final fourth system is $T_8^{ImH+}(C_9G_{20})(C_{10}G_{19})$ where a (C_9G_{20}) base pair occurs instead of the (G_9C_{20}) in the original WT sequence. This sequence was only considered *in silico* and the hydrogen bond persistence was below 1% (see Table 3.2). The expected interruption of the interaction motif due to the (C_9G_{20}) base pair, could be ascribed to changes in the electrostatic surface. The presence of the repulsing exocyclic amine group of the cytosine base pair leads to shrinkage of the accessible electronegative hot spot surrounding G_{19} (see Figure 3.6), resulting in a disruption of the interaction motif. The cytosine amino groups are also likely to result in steric hindrance for an effective interaction in the major groove[3]. As a result, the opinion was that this sequence should not be synthesized and explored experimentally.
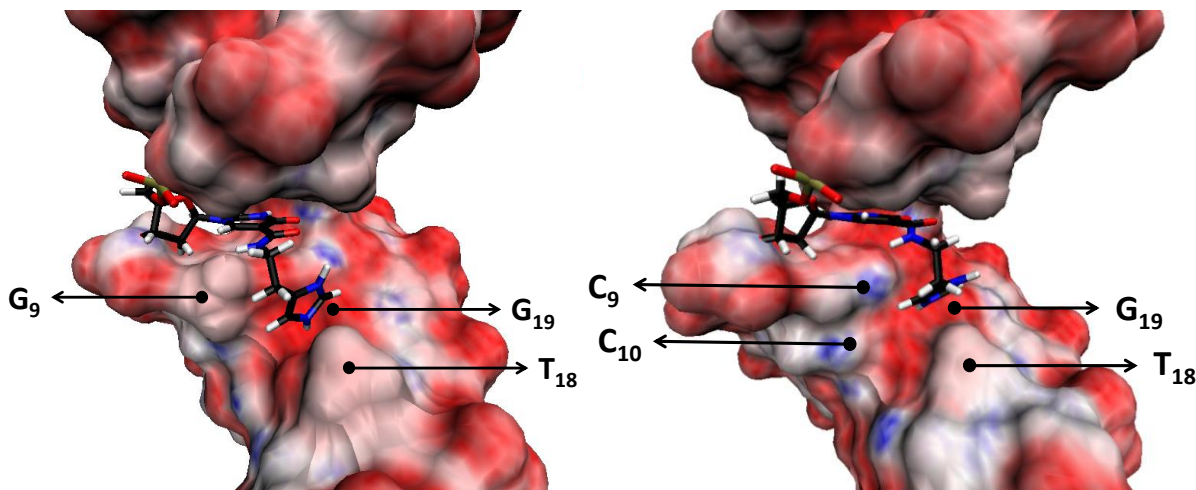
Figure 3.6. Electrostatic surfaces of the T_8^{ImH+}-G_{19} O6 interaction of sequences $T_8^{ImH+}(G_9C_{20})(C_{10}G_{19})$ (left) and $T_8^{ImH+}(C_9G_{20})(C_{10}G_{19})$ (right). The electronegative (electropositive) hot spots in the structure are colour coded in different shades of red (blue) and depend on the intensity. Visualized using VMD 1.9.1. (Adapted from [3]).

While predictions via MD simulations of the behaviour of the sequences have mostly agreed with experimental data in the past, a strong discrepancy is established here in the case of sequence $T_8^{ImH+}(A_9T_{20})(C_{10}G_{19})$ for the first time. As was shown for this sequence, even without a high hydrogen bond persistence in the simulation (3%), the specific interaction pattern is nevertheless present. This puts the interpretation of the even lower H-bond persistence of $T_8^{ImH+}(C_9G_{20})(C_{10}G_{19})$ as excluding formation of the motif in doubt. Therefore, to come to a complete conclusion, the latter sequence should also be investigated during future studies.

3.2 Can a thymine or adenine act as a hydrogen bond acceptor for the positively charged imidazole group?

The original motif and the base pair modifications considered in this section, together with all experimental data are represented in Figure 3.7, Figure 3.9 and Table 3.3.
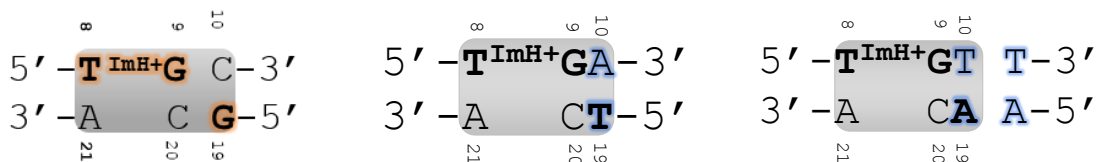
Figure 3.7. The original motif $T_8^{ImH+}(G_9C_{20})(C_{10}G_{19})$ (left) and sequences $T_8^{ImH+}(G_9C_{20})(A_{10}T_{19})$ (middle) and $T_8^{ImH+}(G_9C_{20})(T_{10}A_{19})(T_{11}A_{18})$ (right) to test the A•T/T•A base pair as interaction partner. The additional flipping of $(A_{11}T_{18})$ into a $(T_{11}A_{18})$ base pair was introduced for chemical shift mapping reasons (see text). The base pairs that are altered with respect to the original motif are indicated in blue.

As can be seen in Figure 3.7, the $T_8^{ImH+}(G_9C_{20})(T_{10}A_{19})(T_{11}A_{18})$ sequence features an additional modification with respect to the sequence of the original motif $T_8^{ImH+}(G_9C_{20})(C_{10}G_{19})$. Indeed,

the base pair at position 11 is flipped between the strands in this sequence. The reason for this is that in this fashion, the T_{11} methyl can be used as a chemical shift reporter for the presence of a specific interaction with the neighbouring base via perturbation of the attached methyl resonance as explained in Chapter 2, section 2.3.1 (Figure 3.8).
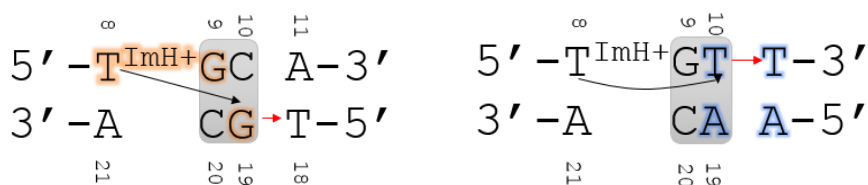
Figure 3.8 Interaction of the T^{ImH+} with G in the original motif $T_8{}^{ImH+}(G_9C_{20})(C_{10}G_{19})$ (left) and with T in sequence $T_8{}^{ImH+}(G_9C_{20})(T_{10}A_{19})(T_{11}A_{18})$ (right). The chemical shift reporter is indicated with a red arrow.
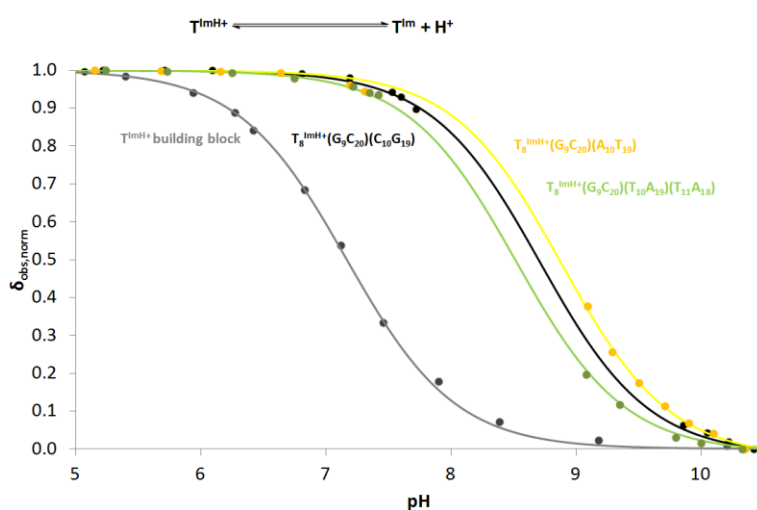
Figure 3.9. Normalized observed chemical shifts as a function of pH for the isolated nucleoside building block (grey), sequences $T_8{}^{ImH+}(G_9C_{20})(A_{10}T_{19})$ (orange), $T_8{}^{ImH+}(G_9C_{20})(T_{10}A_{19})(T_{11}A_{18})$ (green) and the original motif $T_8{}^{ImH+}(G_9C_{20})(C_{10}G_{19})$ (black). Both the experimental data points and fitted curves are shown. T^{Im} and T^{ImH+} refer to the neutral and positively charged state of the imidazole functionality respectively.

Taking Table 3.3 into account, a decrease in T_m of the unmodified sequences is observed, relative to the original motif induced when the $(C_{10}G_{19})$ is replaced by an A•T or T•A base pair, as expected. The introduction of the imidazole modified thymine in both sequences stabilizes the systems with 4-5°C, a value similar to the 5.2°C in original motif $T_8{}^{ImH+}(G_9C_{20})(C_{10}G_{19})$. The pK_{aH} value are similar to those of the original motif, 'sandwiching' the original value. Indeed, the increase of pK_{aH} value is somewhat more pronounced for $T_8{}^{ImH+}(G_9C_{20})(A_{10}T_{19})$ than for $T_8{}^{ImH+}(G_9C_{20})(T_{10}A_{19})(T_{11}A_{18})$ as can be seen in Figure 3.9 and Table 3.3, but remains close in absolute value (8.9 and 8.5) to that of the original motif (8.7). This together with the almost uniform increase in T_m shows that in both cases is interpreted to indicate an interaction is present.

Table 3.3. Melting temperature and pK$_{aH}$ overview of the sequences with an A•T and T•A pair at position 10. [a] T$_m$ values were determined in 100 mM NaCl, 10 mM phosphate buffer at pH = 7. [b] Not relevant since T^{ImH+} represents the imidazole nucleoside building block and not a duplex based system, hence there is no melting process. [c] The wild type sequences have no imidazole functionality.

	Building Block	Original motif		A•T base pair		T•A base pair	
System	T^{ImH+}	(G$_9$C$_{20}$)(C$_{10}$G$_{19}$)	T$_8$$^{ImH+}$(G$_9C_{20}$)(C$_{10}G_{19}$)	(G$_9$C$_{20}$)(A$_{10}$T$_{19}$)	T$_8$$^{ImH+}$(G$_9C_{20}$)(A$_{10}T_{19}$)	(G$_9$C$_{20}$)(T$_{10}$A$_{19}$)(T$_{11}$A$_{18}$)	T$_8$$^{ImH+}$(G$_9C_{20}$)(T$_{10}A_{19}$)(T$_{11}A_{18}$)
T$_m$ (°C)[a]	n.r.[b]	58.9 ± 0.2	64.1 ± 0.06	52.3 ± 0.4	57.3 ± 0.1	56.2 ± 0.2	60.2 ± 0.3
ΔT$_m$ (°C)	n.r.[b]	ref	5.2 ± 0.6	-6.6 ± 0.4	4.9 ± 0.4	-2.7 ± 0.3	4.0 ± 0.3
pK$_{aH}$	7.2 ± 0.02	/[c]	8.7 ± 0.02	/[c]	8.9 ± 0.02	/[c]	8.5 ± 0.03
Δ pK$_{aH}$	ref		1.5 ± 0.03		1.7 ± 0.03		1.3 ± 0.03

The existence of the pK$_{aH}$ motif was already predicted based on *in silico* experiments where both sequences displayed significant persistence for the interaction between T$_8$$^{ImH+}$ Hε$_2$ and T$_{19}$ O4 (for sequence T$_8$$^{ImH+}$(G$_9C_{20}$)(A$_{10}T_{19}$)) and T$_8$$^{ImH+}$ Hε$_2$ to T$_{10}$ O4 (for sequence T$_8$$^{ImH+}$(G$_9C_{20}$)(T$_{10}A_{19}$)(T$_{11}A_{18}$)) albeit less than that in the original motif T$_8$$^{ImH+}$(G$_9C_{20}$)(C$_{10}G_{19}$) (see shaded boxes in Table 3.4).

Table 3.4. H bond persistence of the original motif T$_8$$^{ImH+}$(G$_9C_{20}$)(C$_{10}G_{19}$) and the three other sequences with different base pairs at position 10•19. Adapted from [3].

Sequence	Donor	Acceptor	Persistence
T$_8$$^{ImH+}$(G$_9C_{20}$)(C$_{10}G_{19}$)	T$_8$$^{ImH+}$ Hε$_2$	G$_{19}$ O6	27.1
	T$_8$$^{ImH+}$ Hε$_2$	G$_9$ O6	1.5
T$_8$$^{ImH+}$(G$_9C_{20}$)(A$_{10}T_{19}$)	T$_8$$^{ImH+}$ Hε$_2$	T$_{19}$ O4	9.26
	T$_8$$^{ImH+}$ Hε$_2$	G$_9$ O6	0.60
	T$_8$$^{ImH+}$ Hδ$_1$	G$_9$ N7	1.44
T$_8$$^{ImH+}$(G$_9C_{20}$)(T$_{10}A_{19}$)(T$_{11}A_{18}$)	T$_8$$^{ImH+}$ Hε$_2$	T$_{10}$ O4	16.43
	T$_8$$^{ImH+}$ Hε$_2$	G$_9$ O6	0.54
	T$_8$$^{ImH+}$ Hδ$_1$	G$_9$ N7	0.53

First we consider the T$_8$$^{ImH+}$(G$_9C_{20}$)(A$_{10}T_{19}$) sequence with an A•T base pair at position 10•19. In previous *in silico* work it was established that the duration of the interaction of this sequence is significantly lower (in total 11.3%) than what would be expected in comparison with the original motif (28.6% Table 3.4)[3]. Nevertheless, the values of both T$_m$ and pK$_{aH}$, as well as the chemical shift perturbation figure (Figure 3.9) are pointing at the presence of the specific interaction pattern, indicating again that the MD simulations do not tell the whole story. The presence of the nOe contacts (Figure 3.10) are partially supporting the presence of the specific interaction motif for the sequence with an adenine on position 9 since,

despite their presence, the intensity and number of nOe's is lower compared to what is observed in the original motif $T_8^{ImH+}(G_9C_{20})(C_{10}G_{19})$. This could result from structural features. For instance the distances between T_{19} and the interacting protons during interaction could on average be longer, which results in nOe contacts with lower intensity.

For the $T_8^{ImH+}(G_9C_{20})(T_{10}A_{19})(T_{11}A_{18})$ sequence, a major difference was observed *in silico*. Here, it is found that the protonated imidazole T_8^{ImH+} interacts with the Hoogsteen side of T_{10}, thus leading to an intrastrand interaction. The total *in silico* persistence amounts to 17.5%, which is more than observed for the previous sequence where the interaction, as usual, crosses the duplex from one strand to the other, here from T_8^{ImH+} to T_{19}.

In this case, only one nOe peak is found present in the 2D NOESY spectrum. This finding apparently contradicts with the notable increase in pK_{aH} and T_m value that points to the presence of the specific interaction pattern. While this might seem contradictory, the absence of nOe contacts could be linked to the presence of an intrastrand interaction rather than an interstrand one, leading to a different disposition of the imidazolium group with respect to the backbone. In this specific case, sample issues may also play a role. Indeed, the synthesis yield of the synthesized strands was markedly lower than for the systems studied before. This is illustrated by the poor S/N ratio of the corresponding 1D ^{1}H spectrum (Figure 3.10, top). More nOe contacts could be expected if the 2D NOESY experiment is repeated with a higher sample concentration since in general the specific nOe contacts are already fairly weak, even in the original interaction motif.

Figure 3.10. At the top an overview of the specific nOe contacts starting from the $H\varepsilon_1$ proton for the $T_8^{ImH+}(G_9A_{20})(C_{10}G_{19})$, $T_8^{ImH+}(G_9C_{20})(A_{10}T_{19})$ and $T_8^{ImH+}(G_9C_{20})(T_{10}A_{19})(T_{11}A_{18})$ sequences to the DNA non-exchangeable protons is presented. *corresponds to the exchangeable amide linker proton of the T^{ImH+} functionality. (mixing time 200ms, D_2O, pD 6, 25°C, 700 MHz). Also the sequence of the original motif and the permutated systems are indicated (bottom).

The structural impact of the change from the $(C_{10}G_{19})$ to the $(A_{10}T_{19})$ base-pair on the formation of the specific interstrand interaction motif is highlighted in Figure 3.11. Besides the slightly different position of the carbonyl of both base pairs in the major groove, it remains capable of acting as hydrogen bond acceptor for the imidazolium group introduced at position 8, and thus show the specific interaction leading to the pK_{aH} increase, nOe contacts and T_m increase observed experimentally.
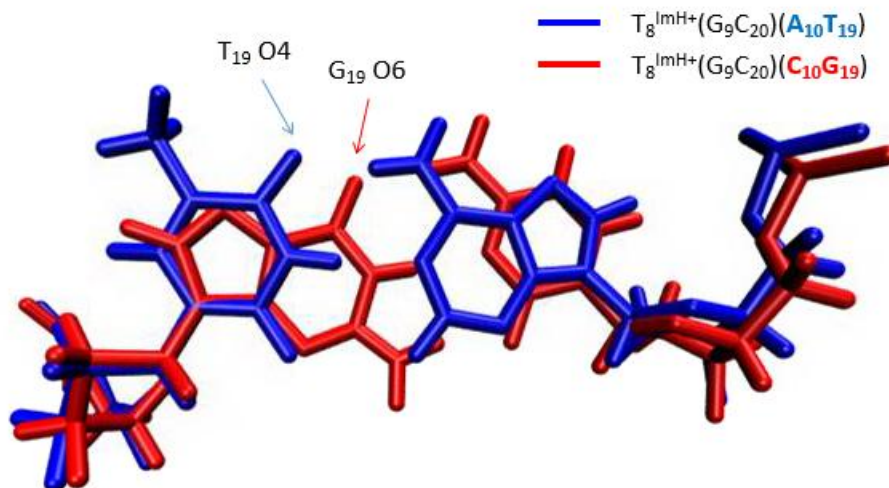
Figure 3.11. Overlay of the $T_8^{ImH+}(G_9C_{20})(A_{10}T_{19})$ (blue) and the $T_8^{ImH+}(G_9C_{20})(C_{10}G_{19})$ (red) base pair after minimizing the RMSD of the upper and lower base pairs of the corresponding sequences. The relevant hydrogen bond acceptor carbonyl groups are indicates where applicable. Visualized using VMD 1.9.1. (Adapted from [3]).

A final source of information with regards to the position of the imidazole are the chemical shift perturbation mapping figures (Figure 3.12) of the sequences with an adenine or thymine introduced at position 10•19. For $T_8^{ImH+}(G_9C_{20})(A_{10}T_{19})$, the observed perturbation pattern is similar to the original motif, indicating the presence of a similar specific interaction pattern. Noteworthy is that the perturbation of the CH_3 group of the thymine adjacent to the interaction T_{19} (T_{18}) features a value of 0.24 ppm. Such effect was also seen in the original motif, where the base at position 19 is a guanine, and a characteristic perturbation of 0.44 ppm is observed (see Chapter 2, Figure 2.9).
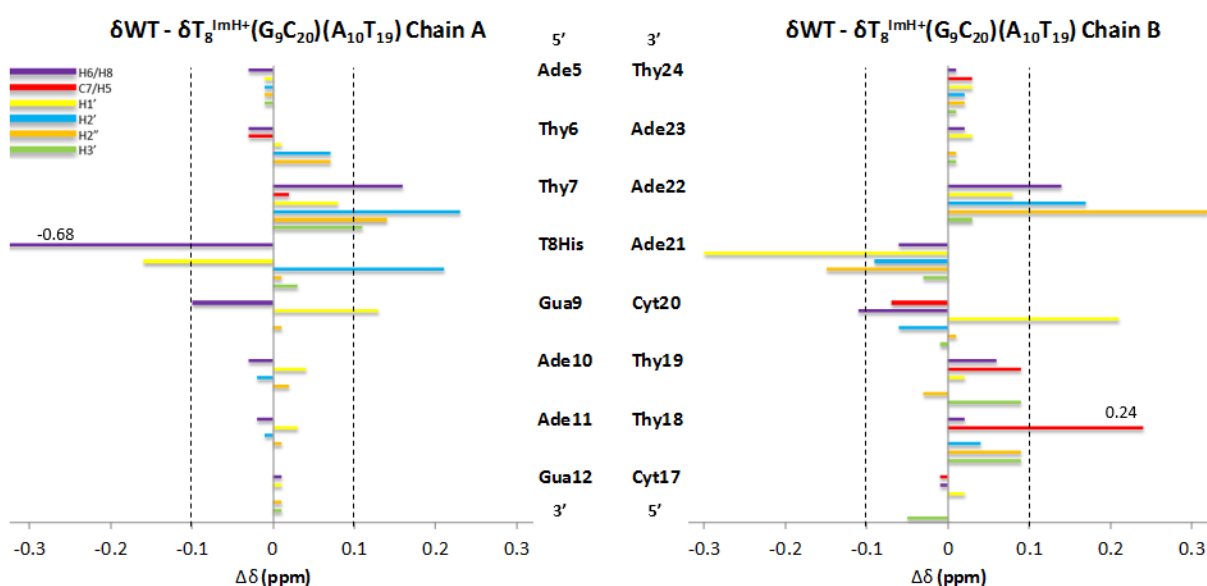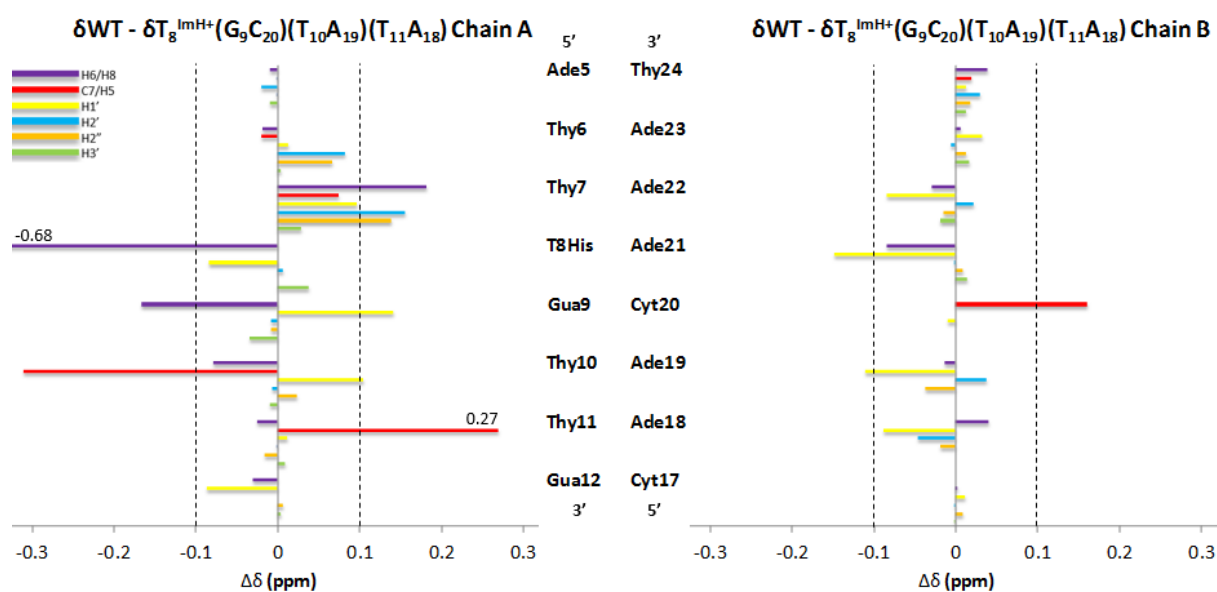
Figure 3.12. Chemical shift perturbation mapping of $T_8^{ImH+}(G_9C_{20})(A_{10}T_{19})$ (top) and $T_8^{ImH+}(G_9C_{20})(T_{10}A_{19})(T_{11}A_{18})$ (bottom) with respect to the non-modified sequences. Only perturbations >0.1 ppm (dashed lines) are considered to be significant. Where necessary the corresponding perturbation value has been added.

The chemical shift perturbation mapping figure of $T_8^{ImH+}(G_9C_{20})(T_{10}A_{19})(T_{11}A_{18})$ looks quite different when compared to that of the original motif $T_8^{ImH+}(G_9C_{20})(C_{10}G_{19})$. Since the motif only involves the strand carrying the T^{ImH+} modification this comes as no surprise. Even so, a chemical shift perturbation of 0.27 ppm at the level of the methyl group of T_{11}, following the motif is observed here, similar to the value of 0.24 ppm in the $T_8^{ImH+}(G_9C_{20})(A_{10}T_{19})$, and is best explained by the intrastrand motif. The impact of the specific positioning of the imidazole unit thus remains an independent and characteristic trait of the presence of the interaction.

The *in silico* observation that T^{ImH+} finds an interaction site with a partner on the same strand in $T_8^{ImH+}(G_9C_{20})(T_{10}A_{19})(T_{11}A_{18})$ – a feature so far never observed before – is thus supported by the experimental data. However, being an intrastrand interaction, it cannot be excluded that this interaction could also arise, at least partially, in the single oligonucleotide strand. Therefore, it is important to investigate whether this behaviour is already present in the non-duplexed single strand.

3.2.0 SINGLE STRAND $T_8^{ImH+}(G_9)(T_{10})(T_{11})$

Since this sequence is being investigated in single strand configuration, our source of information is limited compared to duplex sequences. Full assignment of the 2D NOESY spectrum is not possible, due to the lack of base-stacking and a rigid predictable structure. The latter also allows for the sequential walk, which is absent in the case of a single stranded

system. Thermal melting data cannot be obtained either, because no melting process takes place with single strands. Conclusions about the interaction pattern of the single strand sequence are therefore only possible based on specific nOe contacts involving the imidazole (Figure 3.13) and a pH titration of $T_8^{ImH+}(G_9)(T_{10})(T_{11})$.

Table 3.5. pK_{aH} overview of the single strand and duplex form of the $T_8^{ImH+}(G_9C_{20})(T_{10}A_{19})(T_{11}A_{18})$ sequence.

	Building Block	Duplex	Single strand
System	T^{ImH+}	$T_8^{ImH+}(G_9C_{20})$ $(T_{10}A_{19})(T_{11}A_{18})$	$T_8^{ImH+}(G_9)$ $(T_{10})(T_{11})$
pK_{aH}	7.2 ± 0.02	8.5 ± 0.03	7.6 ± 0.07
$\Delta\ pK_{aH}$	ref	1.3 ± 0.03	0.4 ± 0.03

The data of Table 3.5 shows a small but notable shift in pK_{aH} value of the single strand sequence relative to the T^{ImH+} building block that remains distinctly lower than those of all the studied double helix systems by Buyst *et al*[3] where a specific interaction was experimentally found to be absent (pK_{aH} values from 7.92 ± 0.04 to 8.13 ± 0.04 [1,3]). Also, the presence of a specific interaction pattern should be evident from nOe's involving the imidazole group. But, as can be seen in Figure 3.13, no nOe contacts are visible.

Two conclusions can be drawn based on these observations. Firstly, this conclusively shows that no real interaction is present in the unfolded single strand and that the shift in pK_{aH} should be understood as resulting from the change in chemical background experienced by the imidazole in the nucleoside building block, when the latter is incorporated into a single strand.

Secondly, the double helix plays an important role in the possibility to form a specific interaction pattern. Besides the possible interaction partners present on the opposite strand of the modification, the double helix provides the rigid structure and associated chemical environment to allow the imidazole moiety to interact.
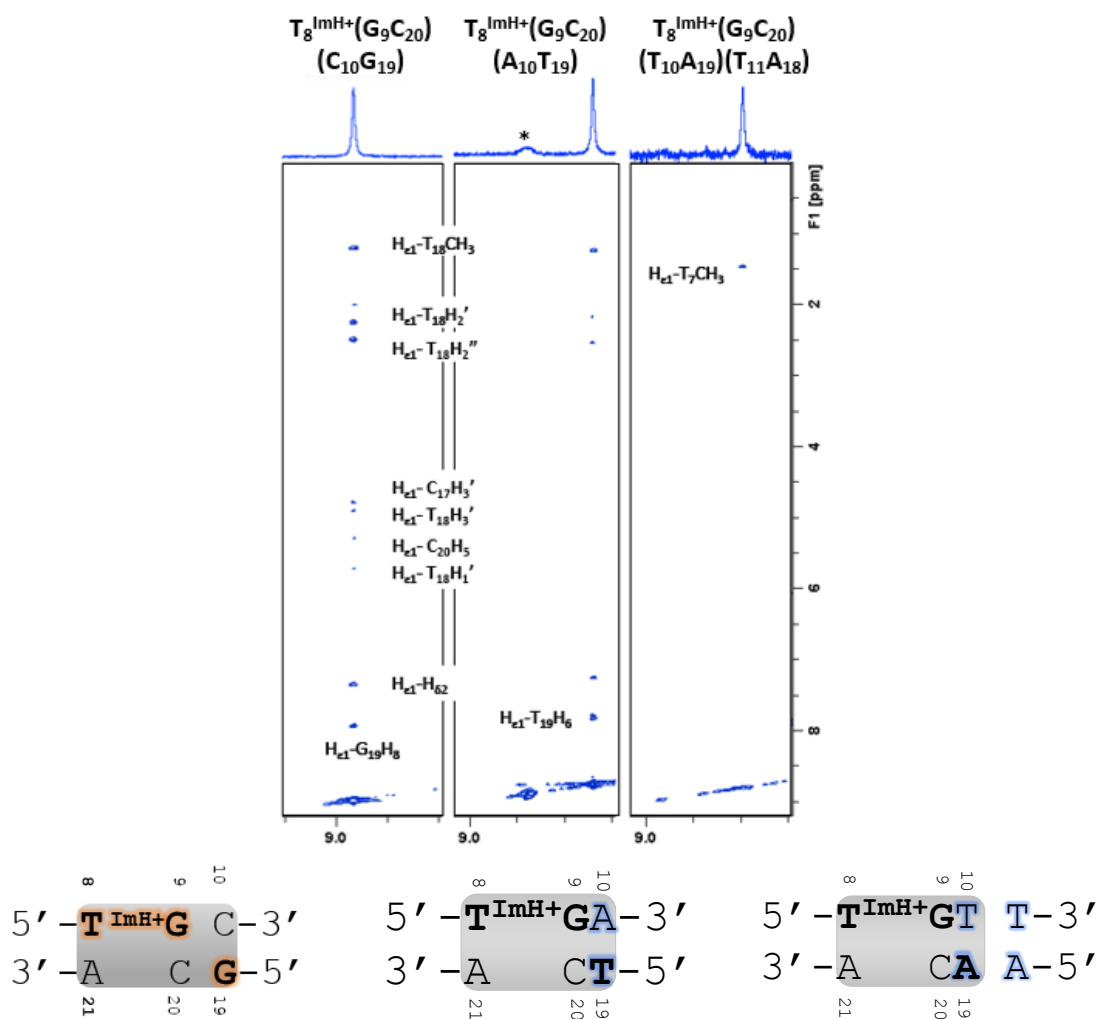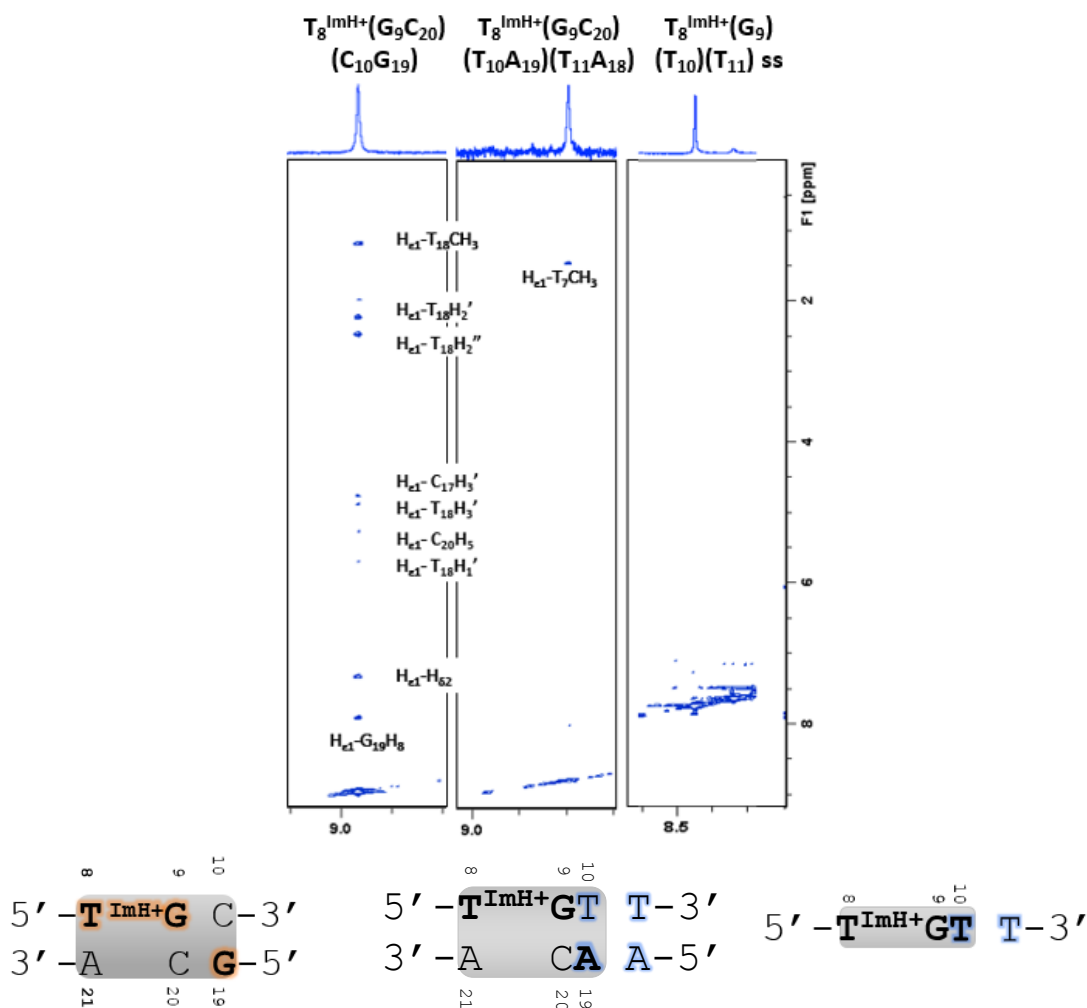
Figure 3.13. At the top an overview of the specific nOe contacts starting from the $H\epsilon_1$ proton for the $T_8^{ImH+}(G_9C_{20})(C_{10}G_{19})$, $T_8^{ImH+}(G_9C_{20})(T_{10}A_{19})(T_{11}A_{18})$ and $T_8^{ImH+}(G_9)(T_{10})(T_{11})$ss sequences to the DNA non-exchangeable protons is represented. *corresponds to the exchangeable amide linker proton of the T^{ImH+} functionality. (mixing time 200ms, D_2O, pD 6, 25°C, 700 MHz).Also the sequence of the original motif and the duplex and single strand form of sequence $T_8^{ImH+}(G_9C_{20})(T_{10}A_{19})(T_{11}A_{18})$ are present (bottom).

3.3 IS IT POSSIBLE TO REPLACE BOTH OF THE GUANINES FOR THYMINES IN THE HYDROGEN BOND ACCEPTOR REGION?

So far, the experimental data for the 4 sequences investigated (Figure 3.1) revealed that the $T_8^{ImH+}(T_9A_{20})(C_{10}G_{19})$ and $T_8^{ImH+}(G_9C_{20})(A_{10}T_{19})$ systems show a specific interaction pattern and a correlated increase of at least 1.5 pK_{aH} units for the imidazole. Substitution of a single A•T/T•A base pair at position 9•20 or 10•19 in the original sequence is thus tolerated and does not perturb the formation of the pK_{aH} interaction motif involving G_{19}. Hence this raised the question whether it is possible to combine *both* permutations, leading to the $T_8^{ImH+}(T_9A_{20})(A_{10}T_{19})$ duplex and maintain the specific pK_{aH} interaction across the duplex.

To examine this possibility, the classical approach is executed. First of all MD simulations are carried out to test whether any specific interaction can be observed. In Figure 3.14 the original motif and the two sequences of interest here are presented.
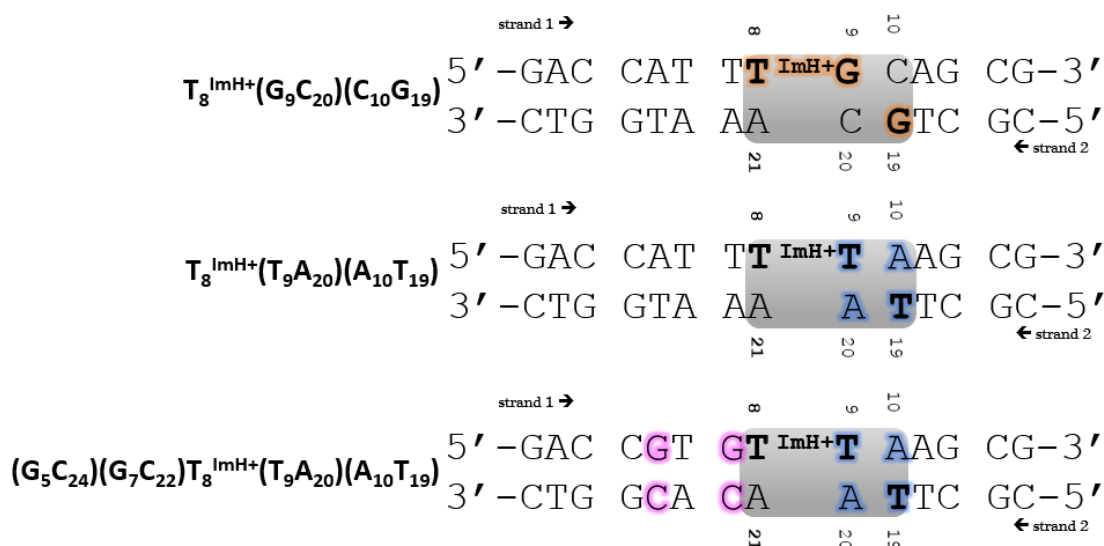
Figure 3.14. The original motif and the sequences where two A•T/T•A base pairs are introduced, the permutations within the motif are indicated in blue, the changes upstream relative to the modification are indicated in pink.

Simulation of the $T_8^{ImH+}(T_9A_{20})(A_{10}T_{19})$ sequence (Fig. 3.14, middle) reveals that the carbonyl oxygen of T_{19} can be involved in hydrogen bond formation with the imidazole, but with a persistence of 2.85% only (Table 3.6). This is less than a third of the persistence observed in the $T_8^{ImH+}(G_9C_{20})(A_{10}T_{19})$ sequence. An additional hydrogen bond formation with the phosphate backbone of T_7 is also apparent. No immediate explanation for this significant drop could be proposed. However, it may be clear from the sequence that it now contains 7 subsequent A•T pairs in its centre. In literature it has been observed that such regions, so-called A-tracts, are characterised by their own unique conformational features such as major groove narrowing[50]. While it remains unclear whether this is also the case for this sequence, it may be possible that a similar conformational feature inhibits a persistent interaction in this sequence. For this reason, a second sequence was proposed (Figure 3.14 bottom) where two G•C base pairs are introduced in order to interrupt the central T•A/A•T base pair sequence, reducing it to 4 subsequent A•T pairs. In spite of this change, the MD simulations showed identical hydrogen bond persistence involving the T_{19} carbonyl oxygen (Table 3.6). Based on experience prior to this work, these *in silico* results suggest no pK_{aH} motif is expected to form in either sequence, a rather unexpected find given the results obtained when introducing a single A•T base pair.

Given the earlier discrepancies between *in silico* and experimental observations, it was decided to synthesize and characterize both sequences to experimentally investigate

whether the interaction motif is maintained in $T_8^{ImH+}(T_9A_{20})(A_{10}T_{19})$ and whether disrupting the A-tract in the latter sequence has any influence.

Table 3.6. H bond persistence of the original motif $T_8^{ImH+}(G_9C_{20})(C_{10}G_{19})$ and the two other sequences with introduction of a thymine at position 9 and 19. ([a] Adapted from [3]).

Sequence	Donor	Acceptor	Persistence
$T_8^{ImH+}(G_9C_{20})(C_{10}G_{19})$	$T_8^{ImH+} H\varepsilon_2$	$G_{19} O6$	27.1^a
	$T_8^{ImH+} H\varepsilon_2$	$G_9 O6$	1.5^a
$T_8^{ImH+}(T_9A_{20})(A_{10}T_{19})$	$T_8^{ImH+} H\delta_1$	$T_7 OP_2$	3.32 ± 0.32^1
	$T_8^{ImH+} H\varepsilon_2$	$T_{19} O4$	2.85 ± 0.40
$(G_5C_{24})(G_7C_{22})T_8^{ImH+}(T_9A_{20})(A_{10}T_{19})$	$T_8^{ImH+} H\delta_1$	$G_7 OP_1$	13.53 ± 0.67
	$T_8^{ImH+} H\delta_1$	$G_7 OP_2$	0.79 ± 0.00
	$T_8^{ImH+} H\varepsilon_2$	$T_{19} O4$	3.32 ± 1.11

The pK_{aH} curves for both sequences are shown in Figure 3.15, as well as those of the original motif and the T^{ImH+} building block. In Table 3.7 the values for T_m and pK_{aH} can be found.
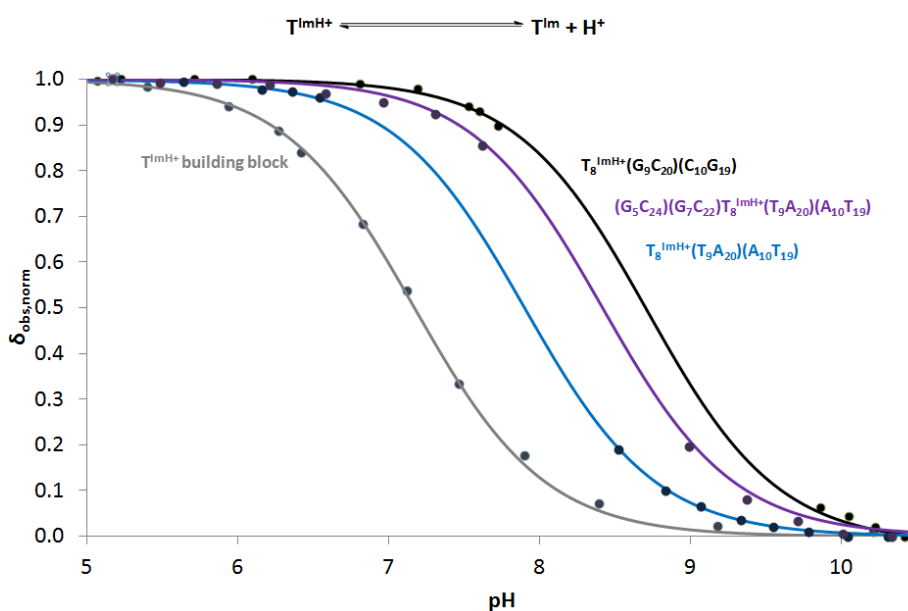
Figure 3.15. Normalized observed chemical shifts as a function of pH for the isolated T nucleoside building block (grey), sequences $T_8^{ImH+}(T_9A_{20})(A_{10}T_{19})$ (blue), $(G_5C_{24})(G_7C_{22})T_8^{ImH+}(T_9A_{20})(A_{10}T_{19})$ (purple) and the original motif $T_8^{ImH+}(G_9C_{20})(C_{10}G_{19})$ (black). Both the experimental data points and fitted curves are shown. T^{Im} and T^{ImH+} refer to the neutral and positively charged state of the imidazole functionality respectively.

---

[1] Standard deviations are included here, to indicate variations observed when comparing two independent simulation trajectories of the same sequence, an approach performed for the first time in the context of this MSc project.

Table 3.7. Melting temperature and pK_{aH} overview of the sequences with two T•A/A•T base pairs introduced within the motif. [a] T_m values were determined in 100 mM NaCl, 10 mM phosphate buffer at pH = 7. [b] Not relevant since T^{ImH+} represents the imidazole nucleoside building block and not a duplex based system, hence there is no melting process. [c] The wild type sequences have no imidazole functionality.

	Building Block	Original motif		$(T_9A_{20})(A_{10}T_{19})$		$(G_5C_{24})(G_7C_{22})(T_9A_{20})(A_{10}T_{19})$	
System	T^{ImH+}	(G_9C_{20}) $(C_{10}G_{19})$	$T_8^{ImH+}(G_9C_{20})$ $(C_{10}G_{19})$	(T_9A_{20}) $(A_{10}T_{19})$	$T_8^{ImH+}(T_9A_{20})$ $(A_{10}T_{19})$	$(G_5C_{24})(G_7C_{22})$ $(T_9A_{20})(A_{10}T_{19})$	$(G_5C_{24})(G_7C_{22})T_8^{ImH+}$ $(T_9A_{20})(A_{10}T_{19})$
T_m (°C)[a]	n.r.[b]	58.9 ± 0.2	64.1 ± 0.06	51.7 ± 1.0	50.6 ± 0.1	56.1 ± 0.1	57.9 ± 0.07
ΔT_m (°C)		ref	5.2 ± 0.6	-7.2 ± 1.0	-1.1 ± 1.0	-2.8 ± 0.2	1.7 ± 0.1
pK_{aH}	7.2 ± 0.02	/[c]	8.7 ± 0.02	/[c]	7.9 ± 0.02	/[c]	8.4 ± 0.03
ΔpK_{aH}	ref		1.5 ± 0.03		0.7 ± 0.03		1.2 ± 0.03

As usual, the general trend of decreasing T_m by introduction of thymine or adenine base pairs can be observed in Table 3.7 (shaded cells), and conversely, substitution of the T•A base pairs for G•C ones at position 5 and 7 causes the T_m to increase again, though remaining somewhat under that of the original motif. Noteworthy, the melting temperature of $T_8^{ImH+}(T_9A_{20})(A_{10}T_{19})$ is lower than that of the matching unmodified sequence. This is the first time ever that introduction of a T^{ImH+} nucleoside causes a destabilisation. When T^{ImH+} is introduced in a sequence where the A-rich tract has been reduced, the T_m value indicates stabilisation of the DNA strand with 1.7°C. Such a value is typical for the stabilisation by 1-2°C caused by the nonspecific interaction between the protonated imidazole and the negatively charged backbone, *vide supra.*

The pK_{aH} value for $T_8^{ImH+}(T_9A_{20})(A_{10}T_{19})$ is 0.8 units lower than the value for the original motif (see Figure 3.15) and fits the value observed before in T_6^{ImH+} and T_7^{ImH+} sequences where it was shown that the imidazole group does not interact[3]. The T_6^{ImH+} and T_7^{ImH+} sequences are similar to the original motif, only the position of the imidazole modification differs. Only two nOe contacts each time involving the methyl group can be seen in the NOESY spectrum: one with the neighbouring T_9 and a clearly weaker one to T_{19} methyl. This indicates some level of proximity, but puts the persistence of the motif in doubt. The latter appears to be confirmed from the absence of a characteristic shift for the adjacent T_{18} methyl, which is limited to 0.11 ppm.

The impact of reducing the length of the A-rich tract by investigating the $(G_5C_{24})(G_7C_{22})T_8^{ImH+}(T_9A_{20})(A_{10}T_{19})$ sequence, is significant. The pK_{aH} value is 0.5 units higher compared to $T_8^{ImH+}(T_9A_{20})(A_{10}T_{19})$ and comes within range of the values for sequences where

the presence of the motif has been established. Also, the specific nOe contact figure (Figure 3.16) shows more nOe's peaks including one to the T_{18} methyl, which also shows the characteristic chemical shift perturbation of 0.24 ppm. Therefore, the interaction motif appears to be largely intact. Taken together, this shows that not only the type of acceptor base pairs is a determinant for the pK_{aH} value and the persistent interaction behaviour, but the overall sequence context and its impact on the structure of the duplex scaffold plays an important role as well.
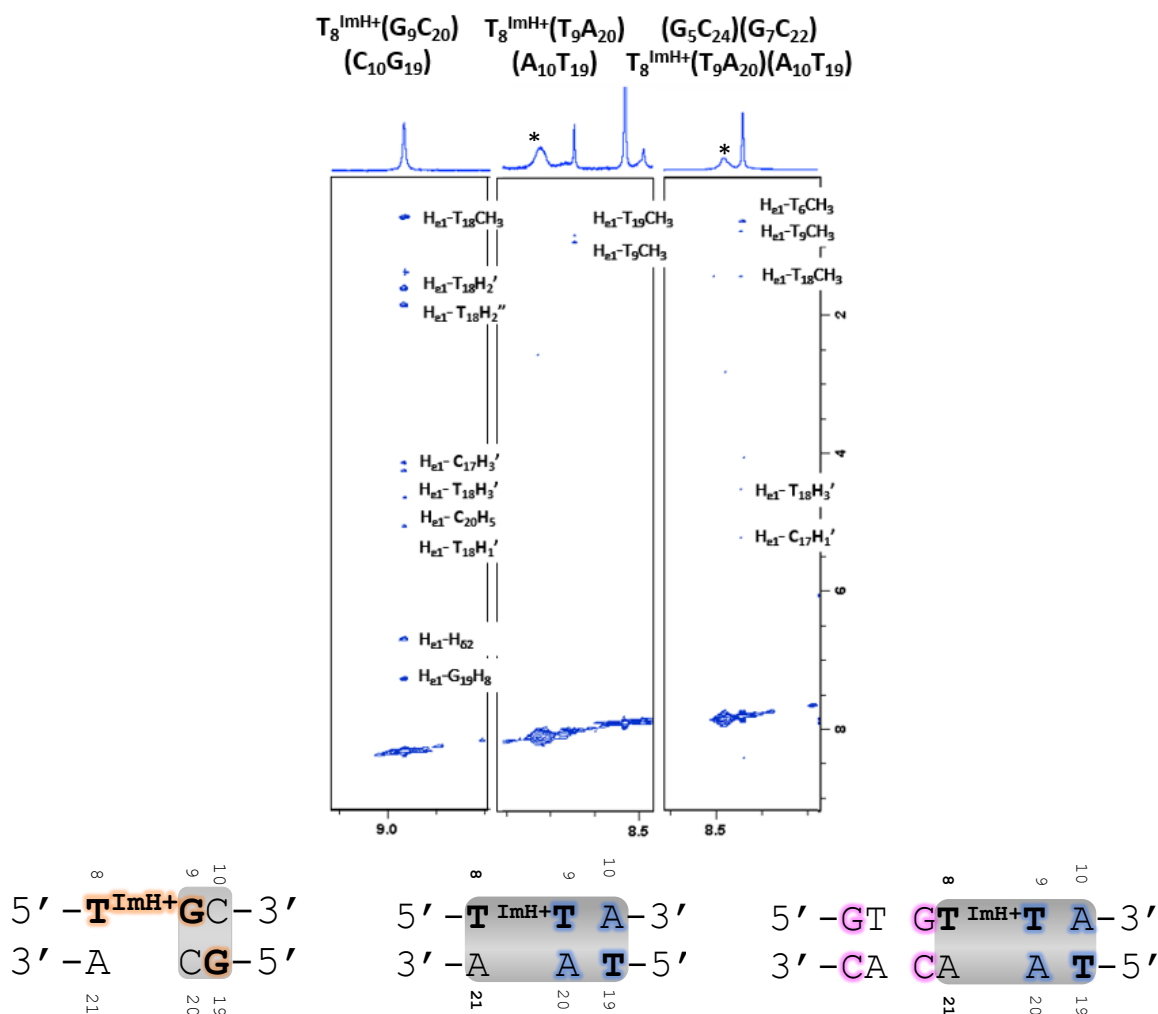
Figure 3.16. At the top an overview of the specific nOe contacts starting from the $H\varepsilon_1$ proton for the $T_8^{ImH+}(G_9A_{20})(C_{10}G_{19})$, $T_8^{ImH+}(T_9A_{20})(A_{10}T_{19})$ and $(G_5C_{24})(G_7C_{22})T_8^{ImH+}(T_9A_{20})(A_{10}T_{19})$ sequences to the DNA non-exchangeable protons is present. *corresponds to the exchangeable amide linker proton of the T^{ImH+} functionality. (mixing time 200ms, D_2O, pD 6, 25°C, 700 MHz).The corresponding sequences are also shown (bottom).

CHAPTER 4 GENERAL CONCLUSION AND PERSPECTIVES

Proteins are characterized by a wide variety of functional groups but lack predictable folding. On the contrary, nucleic acids only have 4 similar building blocks but the secondary structure in duplex form is very predictable. To be able to design and synthesize an active enzyme via a bottom up approach, it was proposed to try and combine the best of both worlds and design a DNAzyme bearing a protein-like imidazole modification. The imidazole moiety was chosen because of its capability to act as hydrogen bond donor and acceptor, but also assist in general acid/base catalysis and nucleophilic catalysis, pointing at the involvement in both binding and catalysis events.

Systematic studies were carried out previously to determine the ideal length and sequence of the DNA duplex[1-3]. Considering the sequence: enough A•T base pairs should be present to experimentally test several positions for introducing the modified building block – which is synthetically limited to thymine – to be able to find the ideal position. Because of stabilisation and structural reasons, also G•C base pairs should be present within the DNA scaffold. Our sources of information to investigate these systems are thermal melting analysis and NMR spectroscopy and, to ensure experimental data analysis, the duplex should stay within reasonable length.

Buyst et al[3] discovered the existence of a pK_{aH} regulating motif consisting of the imidazole modified thymine, T^{Im}, on position 8, followed by a guanine on positions 9 and 19 (Figure 4.1).

Figure 4.1 Original motif with the pK_{aH} regulating motif indicated in the grey box.

This disposition of bases, in combination with a stable duplex structure, leads to formation of hydrogen bonds involving the carbonyl function of the guanines and the protonated imidazole functionality. It results in a significant increase in pK_{aH} (+1.5 compared to the isolated T^{Im} building block). Also, a stabilisation against thermal denaturation of 5-6°C is observed, compared to the same sequence without modification. A marked increase compared to the 1–2°C increase observed when only nonspecific stabilisation occurs. When

the three-base pair motif highlighted in the grey box (Figure 4.1) is introduced in another duplex sequence context, it could be established that the motif is preserved.

In this work, we aimed to contribute in establishing which base pair modifications can be tolerated within the three-base pair box corresponding to the original motif. To this end, permutations of the base pairs within the pK_{aH} regulating motif were proposed and investigated *in silico* and experimentally, to establish the robustness of the motif.

Based on a comparison of all permutations studied so far[3,36, this work] we can now fine-tune the rules of thumb for obtaining a specific interaction motif.

First, at position n+1 (base pairs 9•20) there are no restrictions in terms of nucleobases, with G•C, A•T and T•A all being tolerated. In general terms, the interaction with the G_{19} at position n+2 on the opposing strand is preserved with a carbonyl acceptor present in the major groove at this level.

Second, when an A•T base pair is introduced at position n+2 (base pairs 10•19) the interaction is maintained with thymine capable of acting both as an acceptor at position 19 (opposite strand vs. T^{ImH+} modification) or position 10 (same strand vs. T^{ImH+} modification).

Third, simultaneous introduction of a thymine at position 9 and 19 in an otherwise T•A/A•T rich sequence completely disrupts the motif, an unexpected finding in view of the previous results. Sequence $T_8^{ImH+}(T_9A_{20})(A_{10}T_{19})$ shows also the unique feature that introduction of an imidazole modified thymine does not increase the T_m. The motif can be recovered however, by interrupting the A-rich tract with G•C base pairs. This observation in turn highlights the importance of the sequence context where the motif is introduced.

Last but not least, the interaction motif requires a duplex environment to completely come to fruition as illustrated by the single strand sequence $T_8^{ImH+}(G_9)(T_{10})(T_{11})$.

A final conclusion is that whenever the interaction motif is formed, its presence can be detected by the perturbation of the methyl chemical shift of a thymine, positioned immediately following the interacting base when this is located on the same strand as the T^{Im} modification, or preceding it when located on the opposite strand. The perturbation is fairly constant over all sequences studies, but differs when the interacting base at position 19 is a G or T, being 0.37 ppm (A•T or T•A base pair present on position 9•20) and 0.44 ppm (in case of the original motif) or 0.24 ppm to 0.27 ppm respectively.

This work shows the possibilities and relevance for the executed bottom up approach of nucleic acid based compounds where both experimental and *in silico* experiments were of great value. In future studies it would be worth executing the experimental analysis of the last four permutations (Figure 4.2) within the three-base pair box corresponding to the original motif. Introduction of a C•G base pair on position 9•20 resulting in sequence $T_8^{ImH+}(C_9G_{20})(C_{10}G_{19})$ would clarify if all combinations of base pairs are tolerated by the specific interaction pattern at this position, while investigation of sequence $T_8^{ImH+}(G_9C_{20})(G_{10}C_{19})$ will show if also an intrastrand interaction is possible with a G as interaction partner on the strand carrying the modification. At last, two sequences can be considered where the base pairs on position n+1 and n+2 would be flipped. In the system with G•C and C•G base pairs this results in sequence $T_8^{ImH+}(C_9G_{20})(G_{10}C_{19})$ (an inversion of the base pairs compared to the original motif), while the similar sequence with T•A and A•T base pairs results in $T_8^{ImH+}(A_9T_{20})(T_{10}A_{19})$. *In silico* analysis of the $T_8^{ImH+}(C_9G_{20})(G_{10}C_{19})$ sequence showed no interaction motif[3], but the predictions based on MD simulations of the imidazole behaviour does not always agree with the findings based on experimental work, as seen in several cases in this work. Sequence $T_8^{ImH+}(A_9T_{20})(T_{10}A_{19})$ could be investigated by both *in silico* and experimental work. This sequence will result in an A•T rich tract which will not show an interaction motif, as shown for sequence $T_8^{ImH+}(T_9A_{20})(A_{10}T_{19})$ (Section 3.3). Interrupting this tract while introducing G•C base should be necessary, *vide supra.*
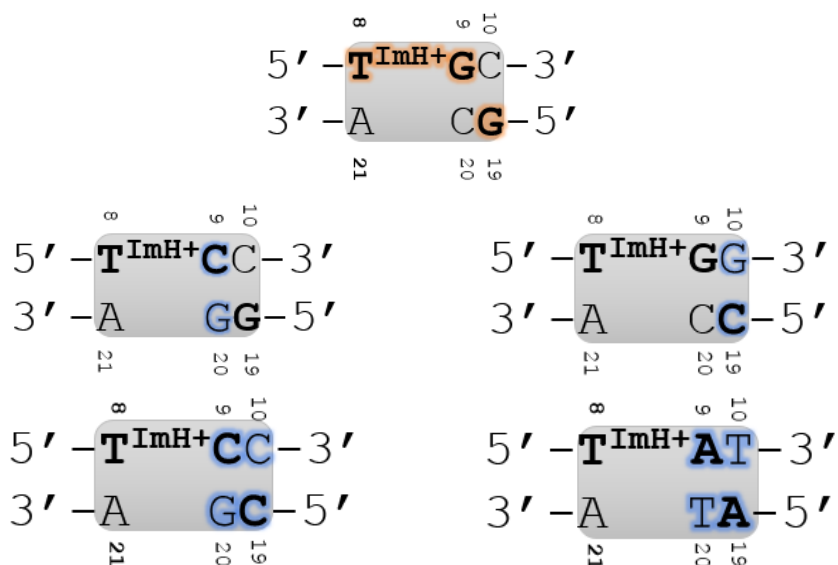
Figure 4.2 The original motif (top) and the last 4 possibilities of permutations within the grey box.

46

Besides the experimental investigation of the above mentioned sequences, sequence $T_8^{ImH+}(G_9C_{20})(T_{10}A_{19})(T_{11}A_{18})$ should also be reanalyzed. Recording a new 2D NOESY spectrum would allow to investigate the presence of specific nOe contacts, which are absent in Figure 3.10 due to low concentration issues. The presence of these expected nOe contacts would support the conclusions already made and allow to complete the investigation of permutated systems of the interaction motif.

CHAPTER 5 MATERIALS AND CHEMICALS

5.0 SYNTHESIS OF THE MODIFIED DNA STRANDS

The synthesis from 5-iodo-2'-deoxyuridine (purchased from Sigma-Aldrich) to the modified thymine is a three step process as described by Gait and Holmes[37]. The first step includes a palladium-catalyzed carboxamidation with histidine (Sigma-Aldrich). The next two steps are protection of both the imidazole NH group with a Boc protecting group and DMTr-protection of the OH group present at the 5'-position of the sugar ring. Afterwards the DNA synthesis is executed based on a phosphoramidite protocol[38] on an Applied Biosystems 394 DNA/RNA synthesizer (1 µM scale) with a 20 minute manual coupling of the modified thymine. The synthesis starts with the first nucleotide bound to a solid phase (Sigma-Aldrich) and is followed by introduction of the nucleotides via a cyclic reaction.

5.0.1 WORK UP

Overnight deprotection and cleavage from the solid phase is carried out in concentrated ammonia at 55°C. During this step deprotection of both the cyano-ethyl groups from the phosphotriester and the nucleobases also occurs. Afterwards a purification step on a Sep-pak C_{18}-cartridge takes place (Protocols)[51]. Before and after the purification step an HPLC-chromatogram is recorded (Appendix). The RP-HPLC analyses were carried out on an Agilent 1100 Series Instrument equipped with an X-bridge Oligonucleotide BEH C18 RP column (2.1 x 50 mm, 2.5 µM) with a gradient of 0-30% ACN in TEAA buffer in 15 minutes at 50°C (0.8 mL/min). The synthesized modified single strands are characterized by use of MALDI-TOF analyses, recorded on an Applied Biosystems Voyager-DE STR Biospectrometry Workstation (Appendix). The deprotection step leads to a high concentration of salts in the solution, especially ammonium salts. These salts could lead to interferences during NMR experiments. Therefore an additional desalting step (Protocols) is necessary to prevent these unintentional interferences. In order to be able to exchange all the triethylammonium cations for Na^+ cations, 1.2 equivalents of NaOAc per equivalent DNA should be used. Therefore, the concentration of the DNA strands has to be determined. This is done by using a Trinean Dropsense96 (Drop plate-D/S) spectrophotometer for UV/Vis spectral analysis. To obtain a DNA duplex an annealing step of the two complementary strands is executed. All unmodified strands were obtained from IDT, Belgium. The last step before any NMR measurements can be started is dialysis in H_2O. A membrane with a 3.5 kDa molecular

weight cut-off is used to wash away all the deletion sequences, unfolded strands and any excess salts during 48 hours. Finally the duplex is lyophilised with a Speedvac SPD 111V Savant.

5.0.2 ^{31}P-NMR OF THE MODIFIED THYMINE

The building block was dissolved in 700 µL of D-chloroform. The spectra were recorded on a Bruker Avance II spectrometer operating at a ^{1}H frequency of 300.13 MHz equipped with a BBO-Z probe.

5.1 MELTING TEMPERATURE MEASUREMENTS

The T_m values were obtained in 10 mM phosphate buffer at pH 7, 100 mM NaCl, 1 µM duplex using a Varian UV-Vis spectrophotometer at a wavelength of 260 nm at regular temperature intervals. The samples were heated from 10°C to 95°C with a rate of 0.3°C per minute. Three heating and cooling cycles were determined for each strand and taking into account the obtained values of T_m (calculated using the Cary 300 Bio Software) an average value ($T_{m,av}$) was calculated, with the standard deviation serving as an error margin. Baseline corrections were executed using reference samples containing 10 mM phosphate buffer at pH 7 and 100 mM NaCl without any DNA.

5.2 NMR STUDIES OF THE MODIFIED OLIGONUCLEOTIDES

For the measurement of the 2D spectra, the duplexes were dissolved in 550 µL of D_2O solution at pD 6.0. The pK_{aH} determination required that the duplexes were dissolved in 550 µL of stock solution containing 90/10 H_2O/D_2O, 100 mM NaCl, 0.1 mM EDTA, 0.05 mM NaN_3 and 0.05 mM DSS (4,4-dimethyl-4-silapentane-1-sulfonic acid). All spectra were recorded on a Bruker Avance II spectrometer operating at a ^{1}H frequency of 700.13 MHz operating under Topspin 3.1pl6 and using a standard 5mm inverse TXI-Z ATMA probe head. All measurements were performed at 25°C. Standard pulse sequences from the Bruker library were used throughout. In D_2O solution, the residual HDO signal was suppressed. All spectra extended over 25.0 ppm (90/10 H_2O/D_2O) or 12.0 ppm (D_2O) along the ^{1}H dimension. The CCPN data model was used for the complete assignment of all duplexes and provides chemical shift data for the assigned non-exchangeable protons[52]. This afforded the collection of ^{1}H chemical shift perturbations caused by introduction of the imidazolium group and the identification of nOe contacts involving the T^{Im} residue in the duplexes.

CHAPTER 6 PROTOCOLS

6.0 PURIFICATION OF THE MODIFIED DNA STRANDS AND CLEAVAGE FROM THE SOLID SUPPORT

The ammonia solutions containing the cleaved oligonucleotide can be purified by different methods such as HPLC, gel-electrophoresis and solid phase extraction (SPE). The currently used procedure is to do a solid phase extraction purification. The SPE cartridge contains C18 reversed phase silica. The ammonia solution can be applied to this cartridge in the following protocol:

Mark the top of the cartridge and put the number of your sample on it.

Make more solution then needed due to the dead-volume of the needle.

- 10 mL acetonitrile (HPLC-grade).
- 10 mL 5mM triethylammonium acetate (TEAA) buffer (fridge).
- 3x solution with the DMTr-protected oligonucleotide in concentrated ammonia solution.
- 15 mL 2,5% aqueous ammonia solution.
- 10 mL milliQ water.
- 10 mL 1,5% trifluoroacetic acid in milliQ (DMTr-deprotection = orange colour).
- 10 mL milliQ.
- 5x1 mL 20% acetonitrile (HPLC-grade).

The first addition of acetonitrile is for activation of the C18-phase. Next the TEAA buffer removes all acetonitrile and prepares the solid support for absorbing the DMTr-protected oligonucleotide. Due to the low polarity of the DMTr group, all oligonucleotides bearing a DMTr will be retained by the solid support but not the acetyl capped sequences. The aqueous ammonia solution in the fourth step removes protecting groups and capped sequences from the solid support. This is followed by a rinsing step before the DMTr protecting group can be cleaved with a 1.5% TFA acid solution in milliQ. This is again followed by a washing step to neutralize the solid support and to remove the DMTr cations before collecting the oligonucleotide with a 20% ACN solution.

6.1 DESALTING OF OLIGONUCLEOTIDES[53]

- Dissolve the single stranded DNA in 20µL milliQ and add
    - 5 µL 3M NaOAc solution per OD unit (1.23 g in 5mL)
        - 1 OD unit = 33 µg/mL
    - 100 µL isopropanol per OD unit
- Vortex the solution (if the solution is too cloudy, add approximately 50 µL milliQ).
- Cool the solution in the freezer during 30min (-20°C) (put on the centrifuge: 10min at 5°C and 7000 rpm).
- Centrifuge the mixture for 10min at 7000 rpm and remove the supernatant carefully, without breaking the pellet.
- Add again the calculated amount of isopropanol and shake the eppendorf (=washing step, pellet does not need to dissolve).
- Centrifuge the mixture for 10min at 7000 rpm and remove the supernatant carefully by making use of a pipette.
- Dry the pellet on the rotary evaporator or in air. Dissolve the pellet in milliQ to determine the concentration using a Trinean protocol.

This protocol was executed on the mixture obtained after Sep-pak purification. The solution still contained triethylammonium ions as counter ions. During this protocol the counter ions are exchanged for Na ions and remaining ammonium salts are removed.

6.1.1 PROTOCOL OPTIMIZATION

Because the above mentioned protocol caused major losses of material, the protocol was optimized. The old protocol was executed in duplicate simultaneous with a slightly adapted protocol, to make sure a comparison between the two methods after optimization was possible. The adaptations included addition of 150 µL isopropanol per OD unit instead of 100 µL isopropanol per OD unit. The loss of DNA in the aqueous phase should be decreased since the percentage of organic phase increased with the larger amount of added isopropanol.

The amount of desalted DNA after the adapted protocol was 20% higher compared to the original protocol, so it was suggested that in future desalting steps 150 µL isopropanol per OD unit should always be added. Another cause for the loss of material could be linked to the drying step. In order to speed up the drying process, an argon flow was used to increase the speed of evaporation of the isopropanol. But, even though the flow used was low, the risk to lose DNA is however too high compared to air-dried DNA.

CHAPTER 7 APPENDIX

7.0 P-NMR
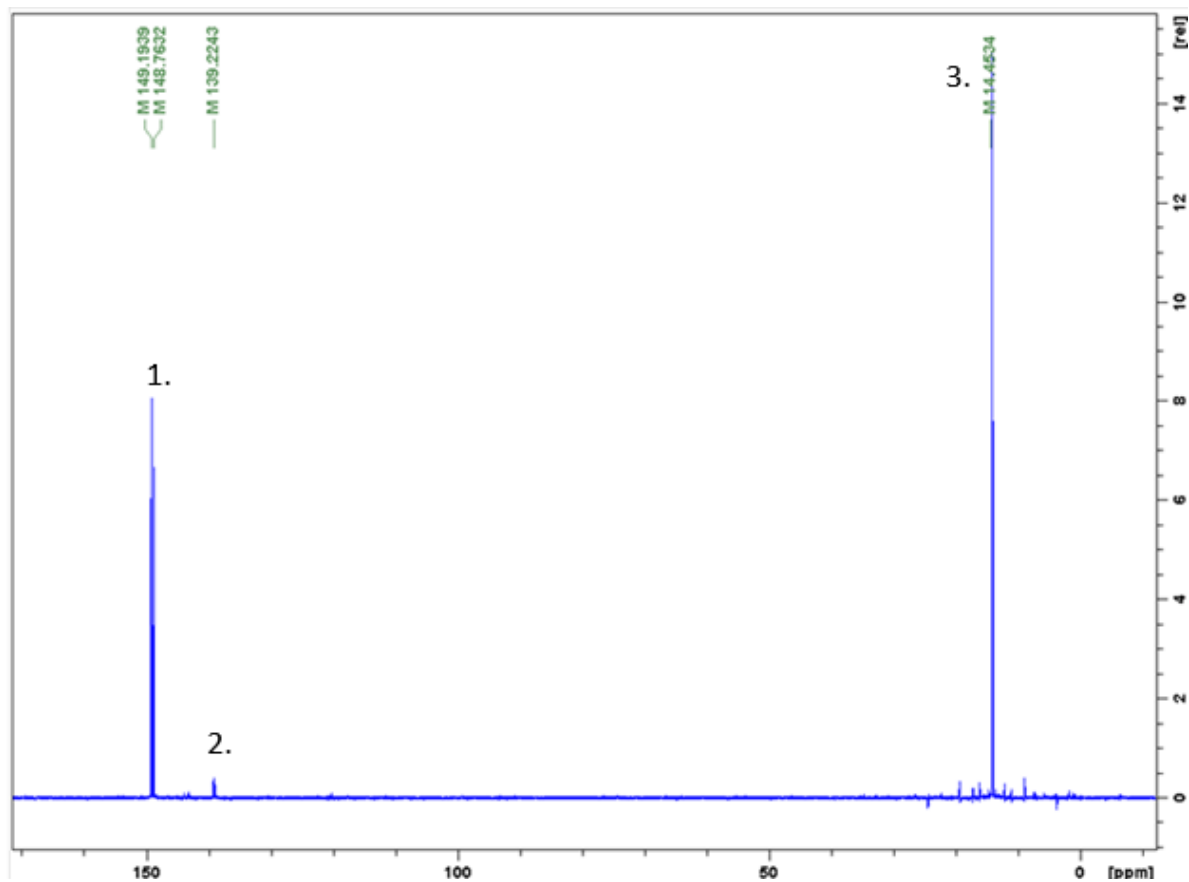
Figure 7.1. ^{1}H decoupled ^{31}P-NMR spectrum of the modified building block (CDCl$_3$, 25°C, 300 MHz)

In the ^{31}P-NMR spectra (Figure 7.1) the following peaks can be observed. Around 14 ppm H-phosphonate impurities derived from the starting material are present and around 150 ppm the end product of the reaction can be found. ^{31}P-NMR experiments cannot be interpreted quantitatively. The peaks which are annotated represent:
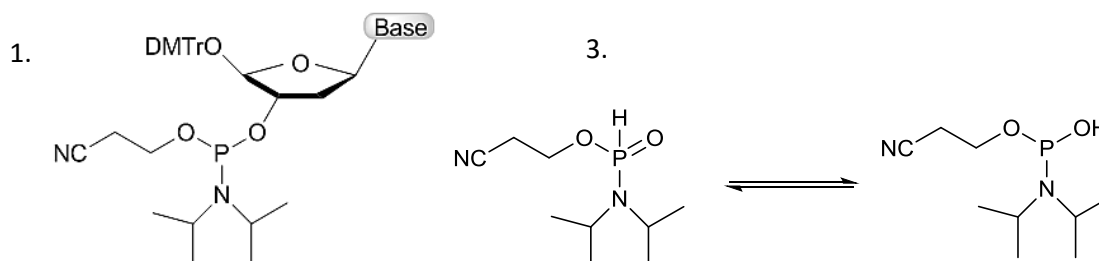
Figure 7.2. Chemical structures present in Figure 7.1.

Peak 1 *appears* as a doublet that results from a contribution of the two diastereomers created by introducing the stereocentre at the P atom. The minor peaks (2) are caused by phosphate triester impurities.

All the ^{31}P-NMR spectra recorded during this thesis are included in Figure 7.3.
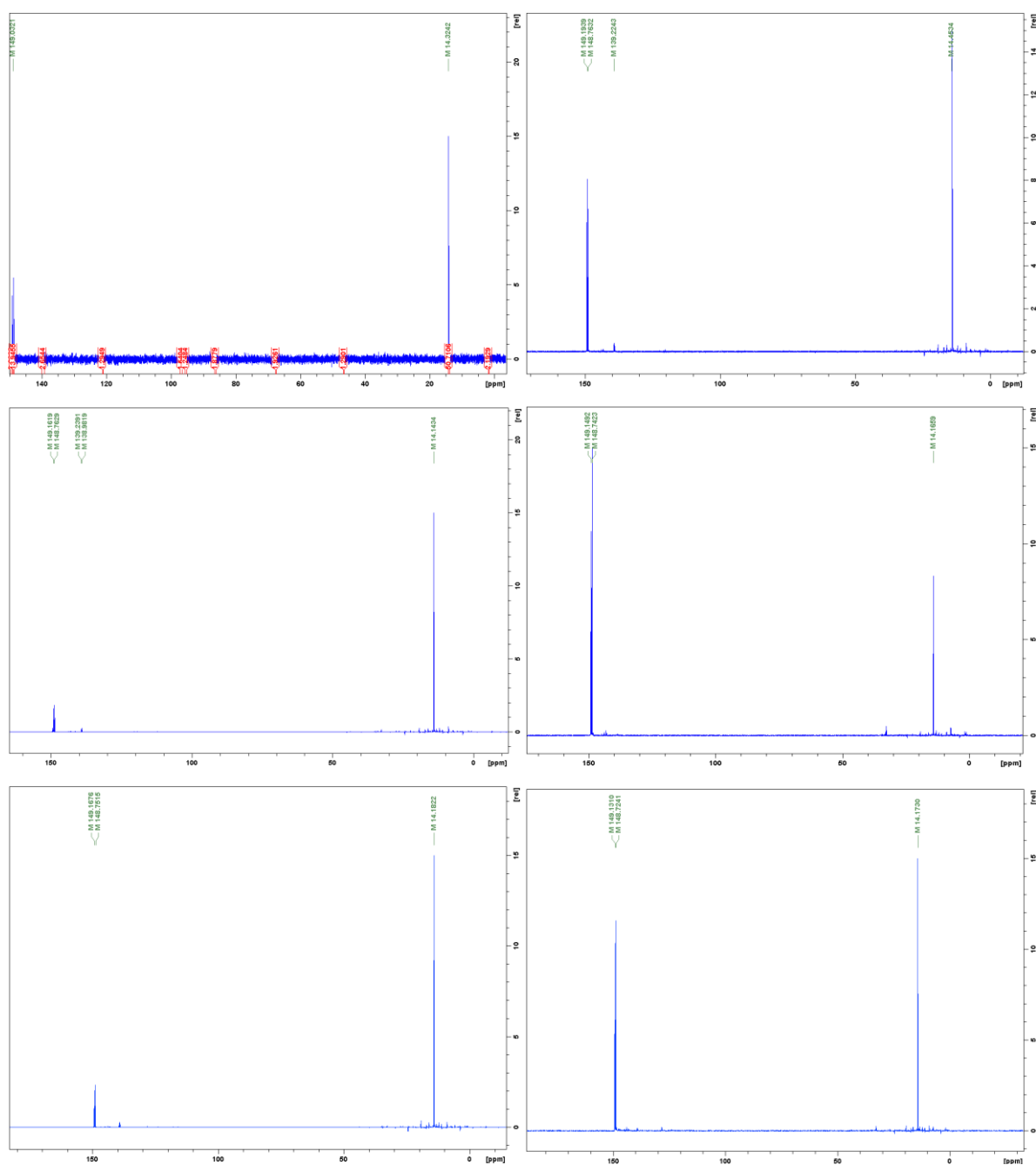
Figure 7.3 ^{31}P-NMR spectra recorded (from top left to bottom right) at the following dates: 9/30/15, 10/13/15, 12/7/15, 2/15/16, 3/7/16 and 3/16/16 (CDCl$_3$, 25°C, 300 MHz).

7.1 HPLC
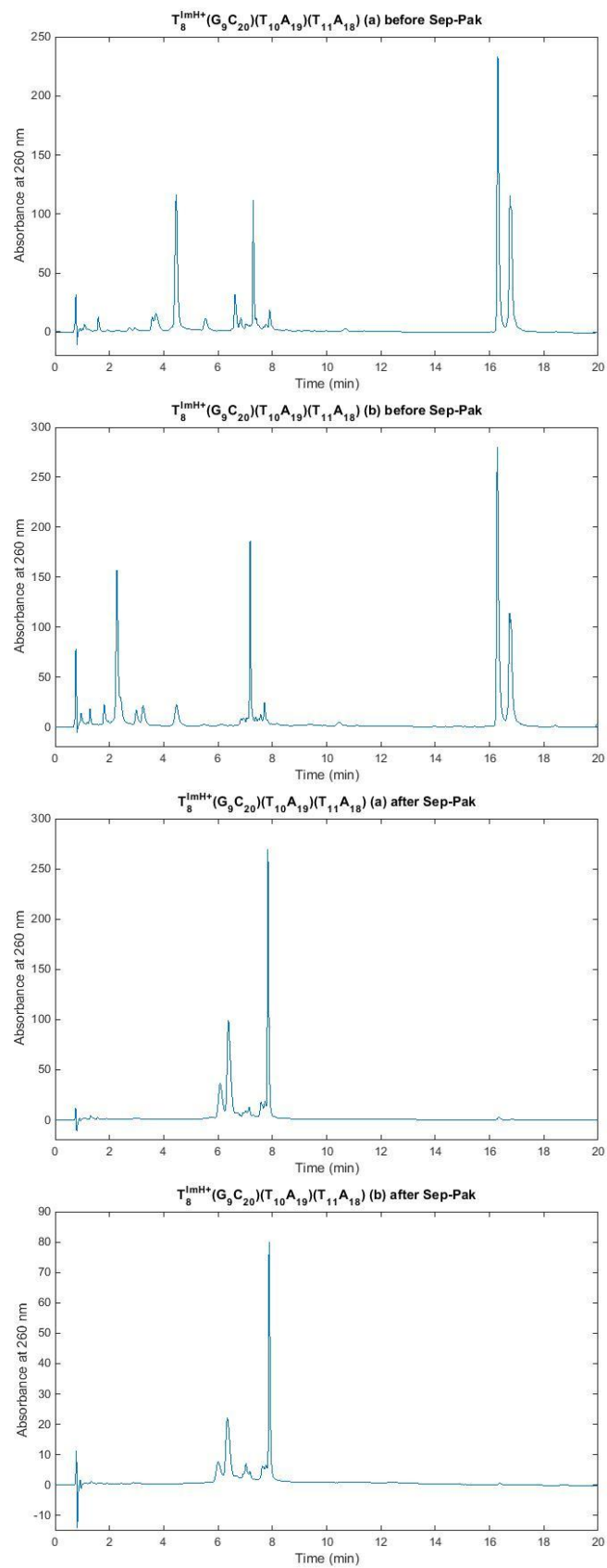
Figure 7.4 HPLC chromatograms of sequence $T_8^{ImH+}(G_9C_{20})(T_{10}A_{19})(T_{11}A_{18})$ in duplicate (a, b) both before and after Sep-Pak purification.
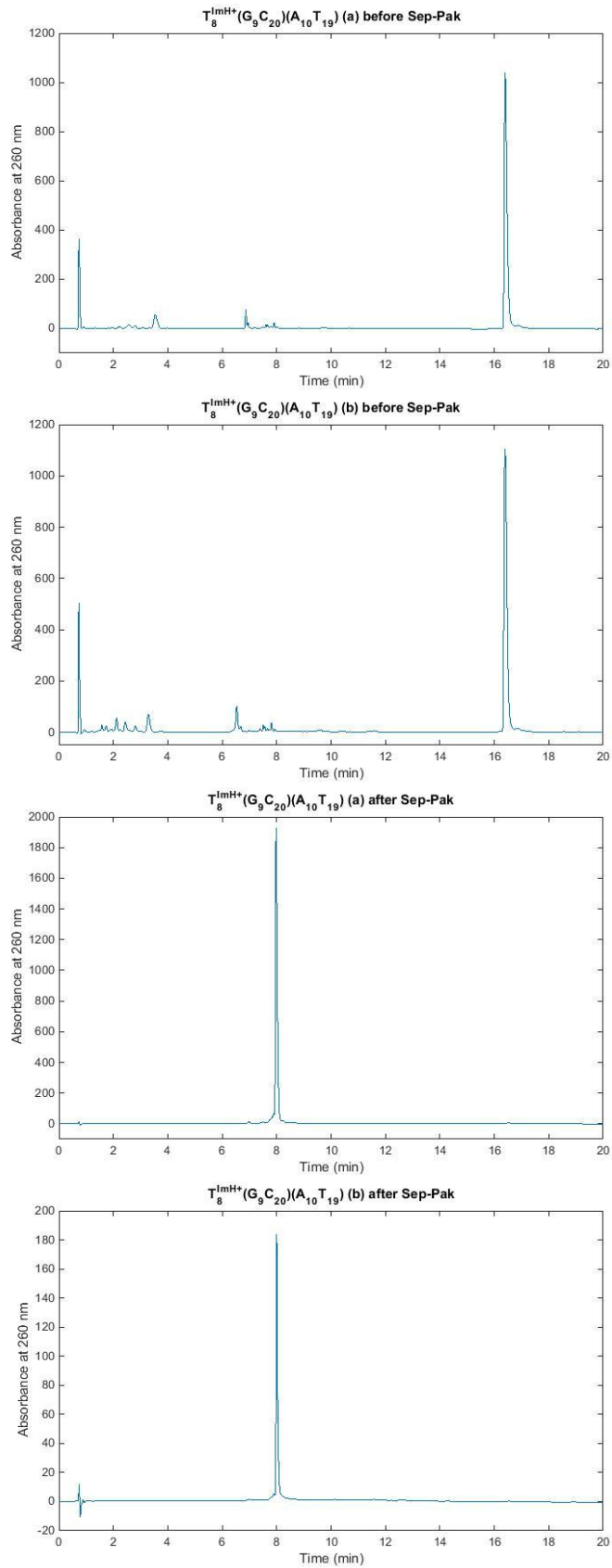
Figure 7.5 HPLC chromatograms of sequence $T_8^{ImH+}(G_9C_{20})(A_{10}T_{19})$ in duplicate (a, b) both before and after Sep-Pak purification.

The chromatograms of sequence $T_8^{ImH+}(G_9C_{20})(A_{10}T_{19})$ (Figure 7.5) show a higher purity compared with the sequence $T_8^{ImH+}(G_9C_{20})(T_{10}A_{19})(T_{11}A_{18})$ (Figure 7.4) due to working under stringent dry conditions during the synthesis. Because this approach resulted in better chromatograms, the subsequent syntheses were carried out under these conditions. Because of the similarity between the HPLC data, not all the chromatograms are included.

7.2 MALDI-TOF

MALDI-TOF (matrix-assisted laser desorption/ionization time of flight) experiments were executed on the single strand modified sequences after the purification step. HPA (3-hydroxypinolinic acid) in combination with citric acid was used as matrix. The samples were mixed with Dowex 50WX 8-100 exchange resin before spotting on the MALDI-plate
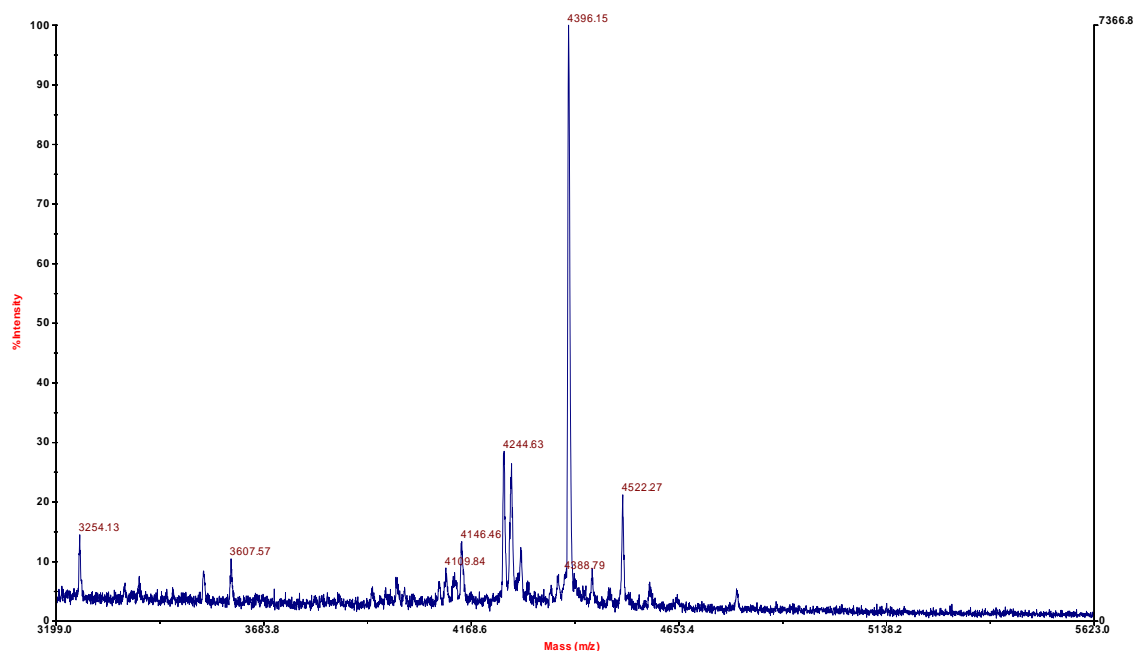
Figure 7.3. MALDI-TOF spectrum of sequence $T_8^{ImH+}(G_9C_{20})(T_{10}A_{19})(T_{11}A_{18})$.

As can be derived from Figure A.6 the most intense peak corresponds to the mass of the ssDNA strand.

7.3 CHEMICAL SHIFT PERTURBATION MAPPING

The following figures are all chemical shift mapping of T_8^{ImH+} sequences with respect to the non-modified WT sequences. Only perturbations with absolute value >0.1 (dashed lines) are considered to be significant. Where necessary the corresponding perturbation value has been added. The order of showing the figures corresponds to the order of discussions in the thesis.
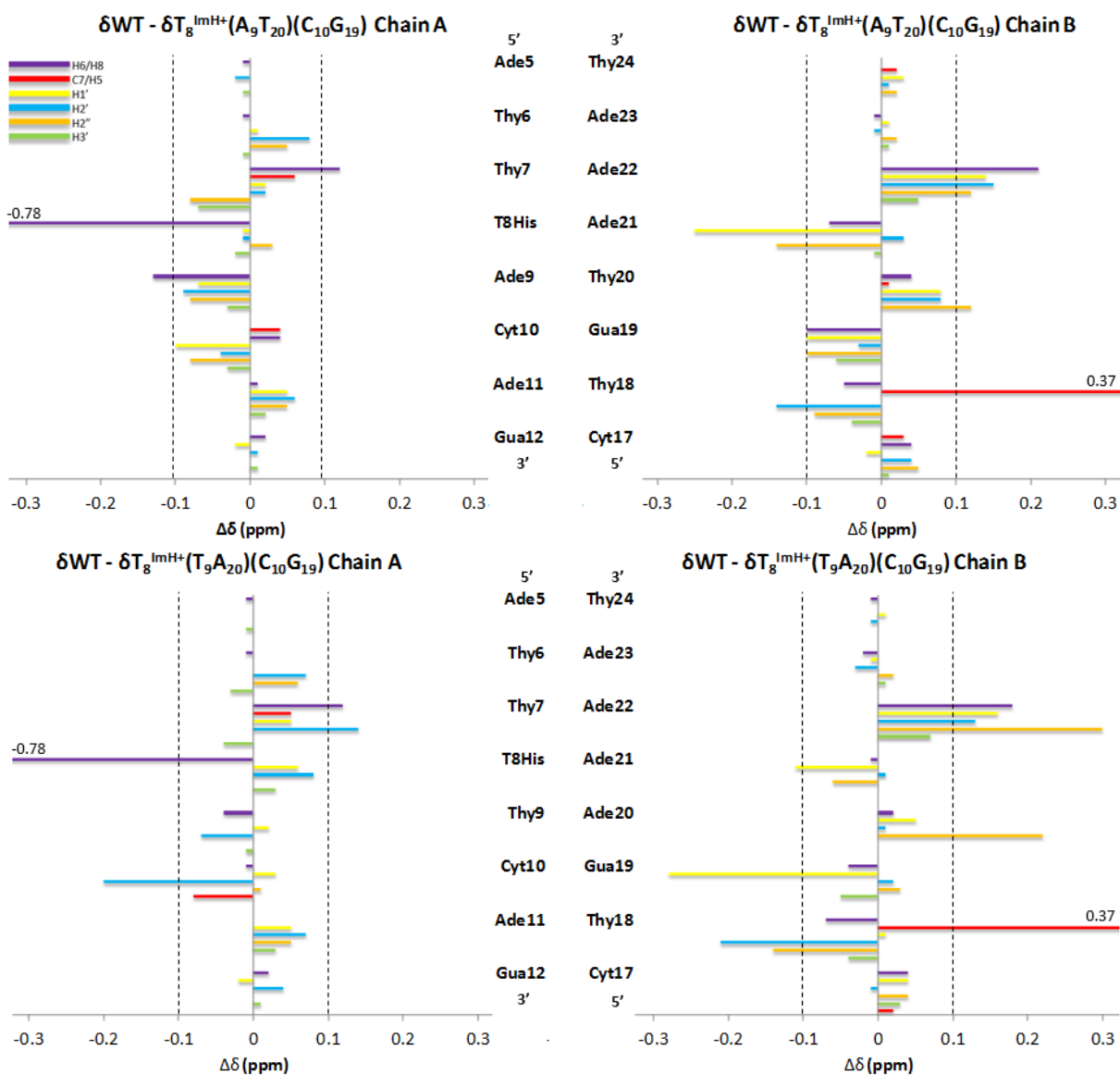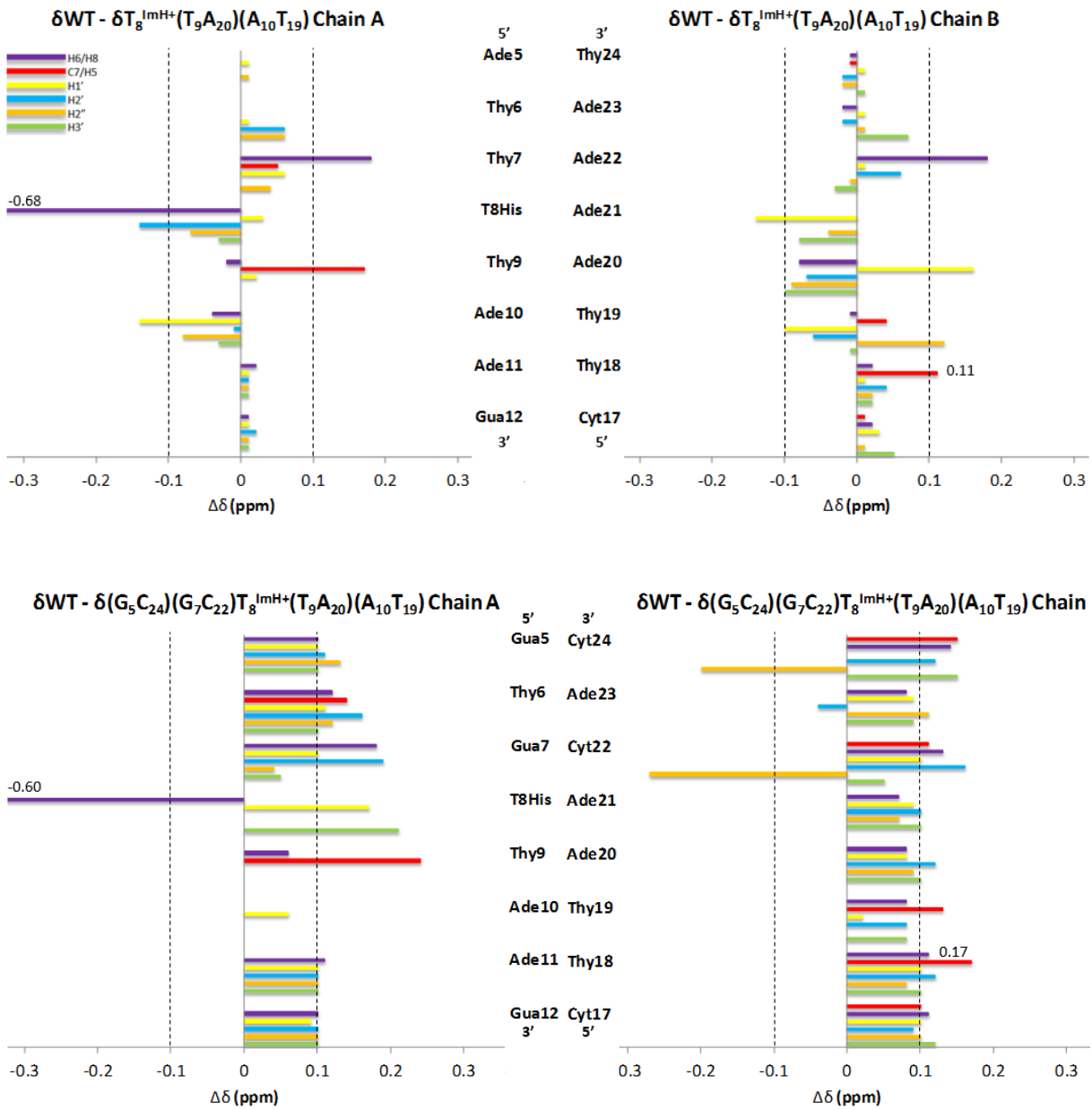
Figure 7.7 Chemical shift perturbation mapping of all synthesized sequences.

7.4 SAFETY

Table 7.1. Hazard symbols of the used chemicals.

Product	⚠	🔥	☣	🌿	🧪	☠
ACN	x	x				
Ammonia	x		x	x		
Brine						
CaH_2		x				
Cap B		x			x	
Chloroform	x		x			
$CO_{2 (s)}$						
D_2O						
DCl	x				x	
DCM	x		x			
DIPEA		x			x	x
EDTA	x					
I_2	x		x			
Isopropanol	x	x				
Na_2SO_4						
NaCl	x					
$NaHCO_3$						
NaN_3			x	x		x
NaOAc						
NaOD					x	
Phosphoramidite						
TCA deblock	x		x	x	x	
TEAA						
TFA	x				x	

Figure 7.8. Safety issues in an NMR-room. (Adapted from [54]).

HOOFDSTUK 8 NEDERLANDSE SAMENVATTING

Veel reacties worden *in vivo* mogelijk gemaakt door het bestaan van katalysatoren. Een belangrijke klasse van katalysatoren is gebaseerd op eiwitstructuren en worden ook wel enzymen genoemd. Door hun grote chemische verscheidenheid en voorkomen in verschillende vormen zijn deze enzymen uitermate geschikt voor het katalyseren van reacties. Tijdens een katalytische reactie vindt interactie plaats met het substraat en wordt de transitietoestand verlaagt om zo een reactie mogelijk te maken. In tegenstelling tot deze enzymen en proteïnen in het algemeen, bezitten nucleïnezuren slechts 4 zeer gelijkaardige bouwstenen die in staat zijn onderlinge waterstofbruggen te vormen. Katalyse door nucleïnezuren werd niet mogelijk geacht tot de ontdekking in het begin van de jaren 80 van enzymatisch RNA[24], ook wel ribozymes genoemd. Deze ontdekking was de start van decennia lang onderzoek naar artificiële enzymen gebaseerd op nucleïnezuren. Niet enkel RNA werd gebruikt als basismolecule, ook DNA kan dienen als scaffold. Om de katalytische performantie van DNA te verhogen, kunnen gemodificeerde bouwstenen geïntroduceerd worden met bijkomende functionele groepen. Op basis van voorafgaande systematische studies[1-3] werd een 14meer duplex DNA uitgerust met een imidazool dragende thymine. Studies naar de beste plaats voor de introductie van de gemodificeerde bouwsteen leidden tot de ontdekking van een pK_{aH} regulerend motief[36], verder het originele motief genoemd (zie Figuur N.1). Dit motief wordt gekarakteriseerd door de aanwezigheid van de imidazool gemodificeerde thymine (T^{ImH}) op positie 8, gevolgd door guanines op positie 9 en 19. Kenmerken voor dit motief zijn een toename in smeltpunt van 5,2°C ten opzichte van de ongemodificeerde sequentie en een toename van 1,5 pK_{aH} eenheden voor T_8^{ImH+} in de duplex ten opzichte van de vrije bouwsteen. De toename in smeltpunt is significant hoger dan de toename van 1-2°C die toegeschreven kan worden aan een aspecifieke interactie van de imidazool met de backbone.
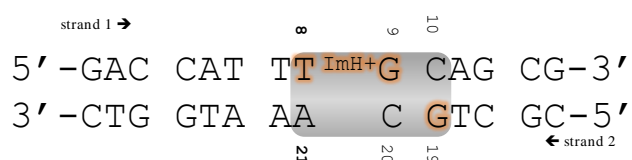
Figuur 8.1. Het originele motief.

Een relevante vraag na deze ontdekking is in welke mate de precieze volgorde van basen binnen de grijze box in Figuur 8.1 vereist is voor het verkrijgen van een pK_{aH} regulerend motief. Teneinde een antwoord te formuleren op deze vraag werden een aantal permutaties

van het originele motief voorgesteld. Het conformationele gedrag van deze gepermuteerde sequenties werd allereerst onderzocht via MD simulaties en op basis hiervan werden sequenties geselecteerd om te synthetiseren en experimenteel te analyseren. Figuur 8.2 toont de sequenties die toelaten de volgende twee vragen te beantwoorden: "Wordt een adenine of thymine getolereerd op positie 9?" en "Voldoet een adenine of thymine als waterstofbrug acceptor op positie 19?"
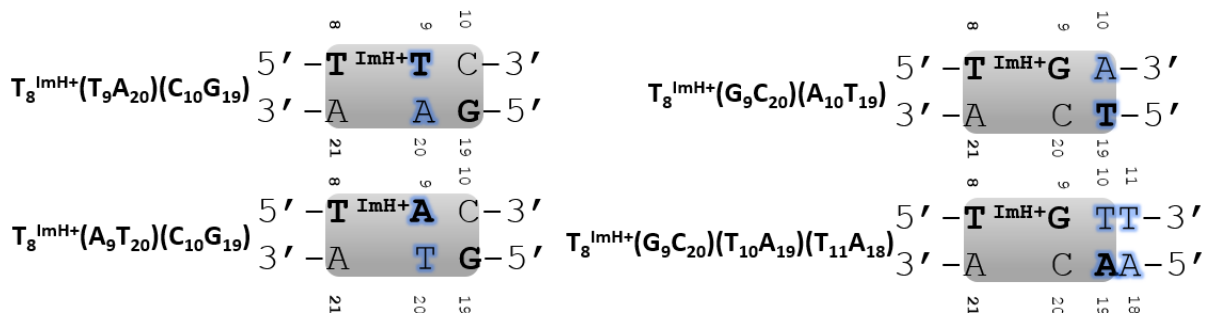
Figuur 8.2. De sequenties onderzocht in deze thesis en hun verkorte naam, de permutaties ten opzichte van het originele motief zijn aangeduid in blauw.

Om een antwoord te formuleren op deze twee vragen werden de gemodificeerde sequenties gesynthetiseerd via de vaste fase fosforamidit methode en onderzocht via UV-Vis thermische analyse en NMR spectroscopie.

Op basis van smelttemperatuuranalyse, pH titraties, bepalen van specifieke nOe-contacten en chemische verschuiving kunnen de volgende conclusies getrokken worden:

- Bij introductie van zowel een adenine als een thymine op positie 9 blijft het motief behouden.

- Een adenine op positie 10 zorgt voor het behoud van het interactiemotief. Een thymine op positie 10 leidt tot een specifieke interactie van T_8^{ImH+} naar T_{10} op dezelfde streng, de toename in T_m en pK_{aH} is minder uitgesproken.

- Tenslotte kan gesteld worden dat de algemene DNA sequentie ook bepalend is voor het al dan niet voorkomen van het interactie motief. Enkel bij een DNA duplex is het motief aanwezig en er werd ook aangetoond dat teveel opeenvolgende A•T base paren het motief verbreken.

Om het permutatie verhaal te vervolledigen zou het interessant zijn om de laatste twee ontbrekende sequenties $T_8^{ImH+}(C_9G_{20})(C_{10}G_{19})$ en $T_8^{ImH+}(G_9C_{20})(G_{10}C_{19})$ ook experimenteel nog te onderzoeken. MD simulaties tonen aan dat deze sequenties het motief niet zouden vertonen, maar zoals bleek uit de analyses van de overige sequenties komen MD simulaties niet altijd overeen met experimentele data en is verdere experimentele verificatie nodig.

CHAPTER 9 REFERENCES

1. Van Gasse, B. (2012). Investigating the structure and stability of modified DNA duplexes with NMR spectroscopy. PhD. Ghent University.

2. Gheerardijn, V. (2014). Functional nucleic acids as tools in catalysis. PhD. Ghent University.

3. Buyst, D. (2015). Design and systematic study of imidazole based DNAzymes. PhD. Ghent University.

4. Kirby, A. (1996). Enzyme mechanisms, models, and mimics. Angewandte Chemie-International Edition in English, 35(7), 707-724.

5. Trends in pharmacological sciences, (2014). Beware of docking. [online] Available at: Cell.com [Accessed 21 Mar. 2016].

6. Boehr, D., Nussinov, P. and Wright, P. (2009). The role of dynamic conformational ensembles in biomolecular recognition. Nature Chemical Biology, 5, 789-796.

7. Dahm, R. (2005). Friedrich Miescher and the discovery of DNA. Developmental Biology, 278(2), 74-288.

8. Crick, F. (1968). The origin of the genetic code. Journal of Molecular Biology, 38(3), 367-379.

9. Nucleic Acids Book, (2005-2016). DNA duplex stability. [online] Available at: Atdbio.com/nucleic-acids-book. [Accessed 21 Mar. 2016].

10. Neidle, S. (2008). Principles of Nucleic Acid Structure. Academic Press, Elsevier.

11. Dickerson, R., H. Drew, et al. (1982). The anatomy of A-DNA, B-DNA, and Z-DNA. Science 216(4545), 475-485.

12. Bochman, M., Paeschke K., et al. (2012). DNA secondary structures: stability and function of G-quadruplex structures. *Nature Reviews. Genetics*, 13(11), 770-780.

13. Rich, A., Nordheim A., et al. (1984). The chemistry and biology of left handed Z-DNA. Annual Review of Biochemistry, 53, 791-846.

14. Arnott, S. and D. Hukins (1972). Optimized parameters for A-DNA and B-DNA. Biochemical and Biophysical Research Communications, 47(6), 1504-&.

15. Wing, R., H. Drew, et al. (1980). "Crystal-structure analysis of a complete turn of B-DNA. Nature, 287(5784), 755-758.

16. Nanda, V. and R. Koder (2010). Designing artificial enzymes by intuition and computation. Nat Chem, 2(1), 15-24.

17. Joyce, G. (1998). Nucleic acid enzymes: Playing with a fuller deck. Proceedings of the National Academy of Sciences of the United States of America, 95(11), 5845-5847.

18. Perrin, D., Garestier T., et al. (2001). Bridging the gap between proteins and nucleic acids: A metal-independent RNAseA mimic with two protein-like functionalities. Journal of the American Chemical Society, 123(8), 1556-1563.

19. Perrin, D., Garestier T., et al. (1999). Expanding the catalytic repertoire of nucleic acid catalysts: Simultaneous incorporation of two modified deoxyribonucleoside triphosphates bearing ammonium and imidazolyl functionalities. Nucleosides Nucleotides & Nucleic Acids 18(3), 377-391.

20. Breaker, R. (1997). In vitro selection of catalytic polynucleotides. Chemical Reviews, 97(2), 371-390.

21. Close, D. (2013). Calculated pK_a's of the DNA base radicals. The journal of physical chemistry, 117, 473-480.

22. Fedor, M. and Williamson J. (2005). The catalytic diversity of RNAs. Nature Reviews. Molecular Cell Biology, 6(5), 399-412.

23. Strobel, S. and Cochrane J. (2007). RNA Catalysis: Ribozymes, Ribosomes and Riboswitches. Current opinion in chemical biology, 11(6), 636-643.

24. Nobel prizes, (2014). The Nobel Prize in Chemistry 1989 – Press Release. [online] Available at: Nobelprize.org. [Accessed. 11 Apr. 2016].

25. Robertson, M. and Joyce G. (2012). The Origins of the RNA World. Cold Spring Harbor Perspectives in Biology, 4(5).

26. Bernhardt, H. (2012). The RNA world hypothesis: the worst theory of the early evolution of life (except for all the others). Biology Direct, **7**.

27. Cech, T.(2000). Structural biology - The ribosome is a ribozyme. Science, 289(5481), 878-879.

28. Thompson, J. and Raines T. (1994). Value of general acid-base catalysis to ribonuclease-A. Journal of the American Chemical Society, 116(12), 5467-5468.

29. Silverman, S. (2008). Catalytic DNA (deoxyribozymes) for synthetic applications - current abilities and future prospects. Chemical Communications, (30), 3467-3485.

30. Lermer, L., Roupioz, et al. (2002). Toward an RNaseA mimic: a DNAzyme with imidazoles and cationic amines. Journal of the American Chemical Society, 124(34), 9960-9961.

31. Breaker, R. and Joyce G. (1994). A DNA enzyme that cleaves RNA. Chemistry & Biology 1(4), 223-229.

32. Breslow, R. and Chung S. (1989). A novel synthesis of substituted imidazoles, and a reexamination of a purported chymotrypsin model. Tetrahedron Letters, 30(33), 4353-4356.

33. Santoro, S.and Joyce G. (1997). A general purpose RNA-cleaving DNA enzyme. Proceedings of the National Academy of Sciences of the United States of America, 94(9), 4262-4266.

34. Raines, R. (1998). Ribonuclease A. Chemical Reviews, 98(3), 1045-1065.

35. Razkin, J., Nilsson H., et al. (2007). Catalysis of the cleavage of uridine 3'-2,2,2-trichloroethylphosphate by a designed Helix-Loop-Helix motif peptide. *Journal of the American Chemical Society,* 129(47), 14752-14758.

36. Buyst, D., Gheerardijn, V. et al. (2015). Identification of a pKa-regulating motif stabilizing imidazole-modified double-stranded DNA. Nucleic Acids Research, 43(1), 51-62.

37. Holmes, S. C and Gait, M. (2005). Syntheses and oligonucleotide incorporation of nucleoside analogues containing pendant imidazolyl or amino functionalities - The search for sequence-specific artificial ribonucleases. European Journal of Organic Chemistry, 2005(24), 5171-5183.

38. Beaucage, S. and Caruthers, M. (1981). Deoxynucleoside phosphoramidites – a new class of key intermediates for deoxypolynucleotide synthesis. Tetrahedron Letters, 22(20), 1859-1862.

39. Blackburn, G., Gait, M., Loakes, D. and Williams, D. (2006). Nucleic acids in chemistry and biology. 3rd ed. Cambridge: The Royal Society of Chemistry.

40. Nucleic Acids Book, (2005-2016). The phosphoramidite method. [online] Available at: Atdbio.com/nucleic-acids-book. [Accessed 28 Sep. 2016].

41. Vargeese, C., Carter J., et al. (1998). Efficient activation of nucleoside phosphoramidites with 4,5-dicyanoimidazole during oligonucleotide synthesis. Nucleic Acids Research, 26(4), 1046-1050.

42. Mergny, J. and Lacroix L. (2003). Analysis of thermal melting curves. Oligonucleotides, 13(6), 515-537.

43. Boelens, R., Scheek, R. et al. (1985). Sequential assignment of imono-proton and amino-proton resonances in H-1-NMR spectra of oligonucleotides by two-dimensional NMR-spectroscopy – Application to a LAC operator fragment. Journal of Magnetic Resonance, 62(3), 378-386.

44. Wütrich, K. (1986). NMR of Proteins and Nucleic Acids. New York, Wiley Interscience.

45. Weiss, M., Patel, D. et al. (1984). Two dimentional H-1-NMR of the lambda-operator site OL1- A sequential assignment strategy and its application. Proceedings of the National Academy of Sciences of the United States of America-Biological Sciences, 81(1), 130-134.

46. Ash, E., Sudmeier, J. et al. (2000). Unusual H-1 NMR chemical shifts support (His) C-epsilon 1-H center dot center dot center dot O = C H-bond, Proposal for reaction-driven ring flip mechanism in serine protease catalysis. Proceedings of the National Academy of Sciences of the United States of America, 97(19), 10371-10376.

47. Robillar.G and Shulman, R. (1972). High-resolution nuclear magnetic-resonance study of histidine-aspartate hydrogen-bond in chemotrypsin and chymotrypsinogen. Journal of Molecular Biology, 71(2), 507-&.

48. Cruz-Gallardo, I., Del Conte, R. et al. (2015). A Non-Invasive NMR Method Based on Histidine Imidazoles to Analyze the pH-Modulation of Protein-Nucleic Acid Interfaces. Chemistry-a European Journal, 21(20), 7588-7595.

49. Silverstein, T. (2012). Fitting Imidazole 1H NMR Titration Data to the Henderson–Hasselbalch Equation. Journal of Chemical Education, 89(11), 1474-1475.

50. Calladine, C. and Drew, H. (1986). Principles of sequence dependent flexure of DNA. Journal of Molecular Biology, 192(4), 907-918.

51. Andrus, A. and Kuimelis, G. (2001). Overview of Purification and Analysis of Synthetic Nucleic Acids in Current Protocols in Nucleic Acid Chemistry. John Wiley & Sons, Inc.

52. Vranken, W., et al. (2005).The CCPN data model for NMR spectroscopy: Development of a software pipeline. *Proteins-Structure Function and Bioinformatics*, 59(4), 687-696.

53. James, T. (2000), Current Protocols in Nucleic Acid Chemistry. John Wiley & Sons, Inc.

54. Bruker AVANCE, (2003). Beginners Guide. [online] Available at: https://www.auburn.edu. [Accessed. 14 May 2016].

Further Exploration of the Imidazole pK$_{aH}$ Motif in DNAzymes through NMR Studies of Sequence Permutations

A. S. de Vries[a,b], D. Buyst[a,b], L. Verdonck[b], A. Madder[b], and J. C. Martins[a]

[a] NMR and Structural Analysis Unit, Department of Organic and Macromolecular Chemistry, Ghent University, Ghent, 9000, Belgium
[b] Organic and Biomimetic Chemistry Research Group, Department of Organic and Macromolecular Chemistry, Ghent University, Ghent, 9000, Belgium

Buyst et al[1] reported the existence of a pK$_{aH}$ regulating motif within a 14mer DNA duplex. This motif is characterized by an imidazole modified thymine at position 8 followed by two guanines at position 9 and 19 (opposite strand). A•T/T•A base pairs are introduced at position 9 and 10 to investigate the tolerance of this motif to other bases. It is been demonstrated that at position 9 both adenine and thymine are tolerated by the motif. Also at position 10 the introduction of A•T base pairs is possible. Nevertheless, the presence of a thymine as hydrogen bond acceptor at position 10 resulted in a less pronounced interaction pattern, caused by an intrastrand regulating motif. The conclusion made after the investigation of a sequence where both the guanines are replaced for thymines is that also the overall DNA sequence is of importance.

Introduction

Processes catalyzed by functionalities within a natural DNA moiety have not (yet) been discovered in nature[2.3], but since the detection of catalytic RNA[4,5] by T. Cech in the early 1980s, scientists became aware of the ability of nucleic acids to function as an enzyme as well. This important discovery was awarded with a Nobel prize in 1989[6]. As enzymes based on RNA are called ribozymes, DNA based enzymes are named DNAzymes. Artificial DNAzymes can be used to assist in different reactions including targeted hydrolysis of RNA, probing of structured RNA or assisting in the manipulation of recombinant DNA[7,8].

The choice to use DNA as a scaffold for an enzyme instead of RNA is mainly based on practical considerations. First, DNA is more stable and has a more rigid structure compared to RNA. The increased stability of DNA is caused by the absence of a reactive hydroxyl group at the 2' position of the sugar ring and since DNA mostly occurs as a double helix, the conformation is more predictable compared to the less rigid conformations adopted by single stranded RNA. As a result of this DNA enables us to work with reliable positioning of synthetic functionalities. Secondly, the synthesis of DNA on solid-phase is more cost-efficient. This is related to the expense and ease during the monomer synthesis, the efficiency of the introduction of new nucleotides and the simplicity of deprotection after the reaction cycle[9]. To overcome the limited functional group diversity in DNA, an imidazole modified thymine is introduced in a DNA duplex. The imidazole moiety was chosen because of its capability to act as hydrogen bond donor and acceptor, but also assist in general acid/base catalysis and nucleophilic catalysis, pointing at the involvement in both binding and catalysis events. In earlier work systematic studies were carried out[10-12] to determine the ideal position of introducing the imidazole modified thymine and length of the DNA duplex. To be able to record thermal melting data and NMR spectra and interpret the data properly, the duplex

should be of reasonable length. Buyst *et al*[1] reported the existence of a pK_{aH} regulating motif within this 14mer DNA duplex. Figure 1. shows hydrogen bonds present between the imidazole modified thymine and the DNA scaffold.
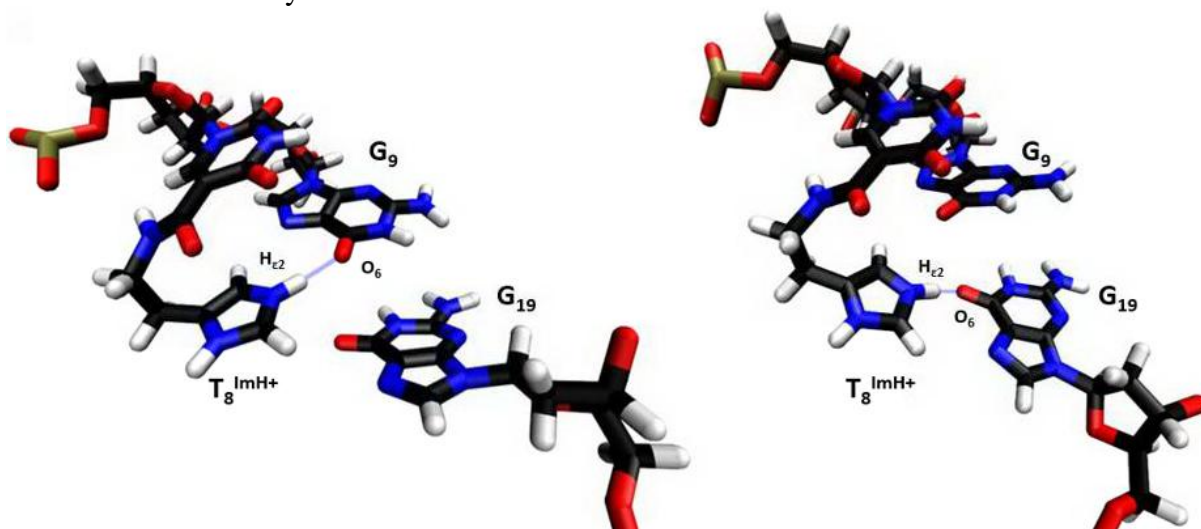
Figure 1. Two specific interaction patters of the pK_{aH} regulating motif. In each case the imidazole functionality, T_8^{ImH+}, is the hydrogen bond donor, while G_9 and G_{19} act as hydrogen bond acceptor. The hydrogen bonds are represented by blue dotted lines. (Adapted from [12]).

As graphically represented in Figure 2, the imidazole modified thymine, T^{Im}, is introduced at position 8, followed by a guanine at position 9 (n + 1 compared to the modified thymine) and a guanine at position 19 (n +2 compared to the modified thymine and on the opposite strand). This motif gives rise to an increase in thermal melting point of 5.2°C relative to the wild type (WT) sequence and an increase in pK_{aH} of 1.5 units relative to the T^{Im} building block.
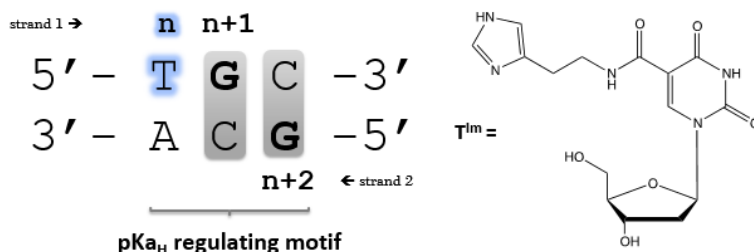
Figure 2. Specific sequence which results in a pK_{aH} regulating motif (left), the modified thymine (right) which is introduced at position n.

To investigate whether stabilizing interactions of imidazole building blocks are allowed in other combinations within the DNA scaffold, a few permutations of the above mentioned sequence were investigated with *in silico* MD simulations. Afterwards the most promising sequences were experimentally tested. This allowed us to examine the robustness and tolerance for different base pairs within the stabilizing motif and test whether a similar interaction motif could be obtained without G as the main acceptor. The following sequences are of interest:
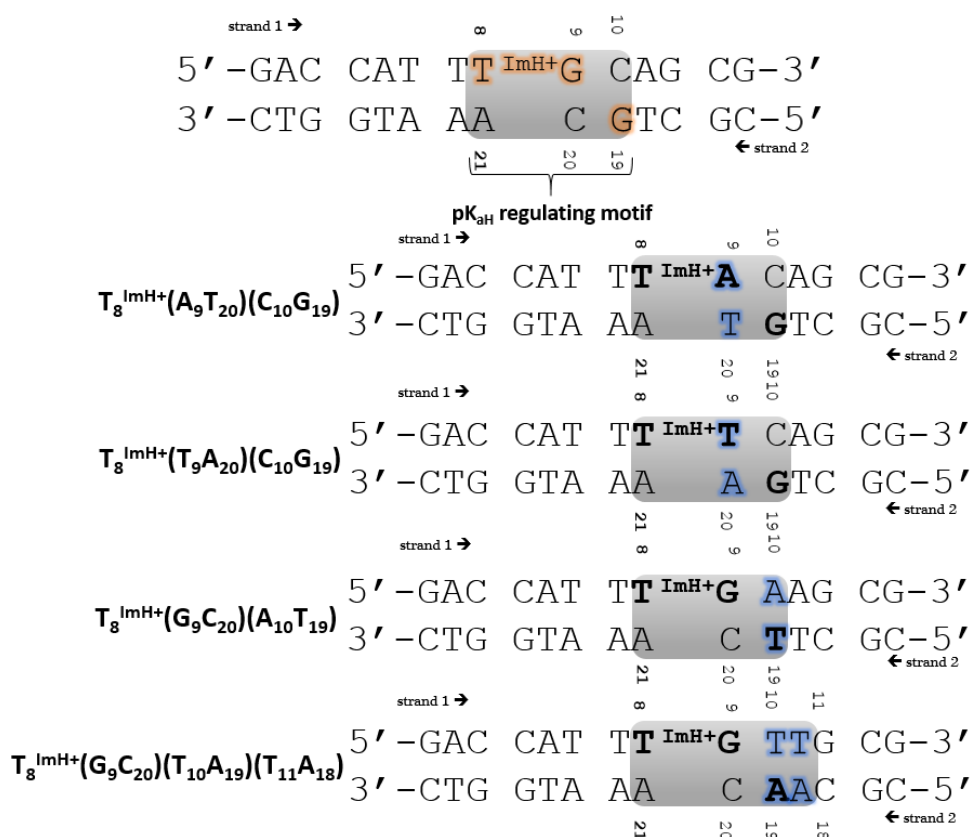
Figure 3. The original motif (top) and the investigated sequences together with their shortened name that will be used throughout the text.

The experimental analysis of these sequences allow to answer the following questions: "Is an A•T or T•A base pair tolerated at position 9?" and "Can an adenine or thymine act as a hydrogen bond acceptor for the positively charged imidazolium group?".

Materials and methods

Synthesis of T^{Im} modified DNA

Starting from 5-iodo-2'-deoxyuridine (Sigma Aldrich) the modified nucleotide was obtained via a three step synthesis described by Holmes and Gait[13]. The first step included a palladium-catalyzed one-pot carboxamidation followed by protection of the imidazole NH group and the 5' OH group of the sugar ring with a tBoc and DMTr group respectively. Afterwards, the modified building block could be introduced in a DNA sequence via phosphoramidite chemistry using standard protocols[14]. Subsequent deprotection and cleavage from the solid support was executed in concentrated ammonia followed by a Sep-pak purification. The complementary strands, which are non-modified, were purchased from Integrated DNA technologies (IDT, Belgium). Desalting using an ion exchange method[15] was executed on both the modified and purchased strands. The oligonucleotides were dissolved in a 97/3 mixture of isopropanol and a 3M NaOAc solution. The mixture was centrifuged (10 minutes, 7000 rpm) and the oligonucleotide precipitate was air-dried. The molar quantities of single strands were determined spectrophotometrically by measuring the absorbance at room temperature at 260 nm (Trinean DropSense 96). To obtain the desired double stranded DNA, equimolar amounts of single strand were annealed by heating to 95°C for 30 minutes, followed by gradual cooling down to room temperature. The annealed duplex sample was dialyzed using a dialysis membrane with a 3.5 kDa molecular weight cut-off.

UV thermal denaturation experiments

The T_m values were obtained in 10 mM phosphate buffer at pH 7, 100 mM NaCl, 1 µM duplex using a Varian UV-Vis spectrophotometer at a wavelength of 260 nm at regular intervals. The samples were heated from 10°C to 95°C using a rate of 0.3°C per minute. Three heating and cooling cycles were determined for each strand and taking into account the obtained values of T_m (calculated using the Cary 300 Bio Software) an average value ($T_{m,av}$) was calculated, with the standard deviation serving as an error margin. Baseline corrections were executed using reference samples containing 10 mM phosphate buffer at pH 7 and 100 mM NaCl without any DNA.

NMR studies of the modified oligonucleotides

2D spectra for assignment used the duplexes dissolved in 550 µL of D_2O solution at pD 6.0. The pK_{aH} determination required that the duplexes were dissolved in 550 µL of stock solution containing 90/10 H_2O/D_2O, 100 mM NaCl, 0.1 mM EDTA, 0.05 mM NaN_3 and 0.05 mM DSS (4,4-dimethyl-4-silapentane-1-sulfonic acid). All spectra were recorded on a Bruker Avance II spectrometer operating at a ^1H frequency of 700.13 MHz operating under Topspin 3.1pl6 and using a standard 5mm inverse TXI-Z ATMA probe head. All measurements were performed at 25°C. Standard pulse sequences from the Bruker library were used throughout. All spectra extended over 25.0 ppm (90/10 H_2O/D_2O) or 12.0 ppm (D_2O) along the ^1H dimension. Mixing times for the TOCSY and NOESY spectra were 75 ms and 200 ms, respectively.

The CCPN data model was used for the complete assignment of all duplexes and provides chemical shift data for the assigned non-exchangeable protons[16]. This afforded the collection of ^1H chemical shift perturbations caused by introduction of the imidazolium group and the identification of nOe contacts involving the T^{Im} residue in the duplexes.

T^{Im} pK_{aH} determination

To determine the pK_{aH} of the imidazole the ε1 and δ2 protons are followed in 1D ^1H spectra recorded at different pH values. To be able to identify these protons, a 2D TOCSY experiment is recorded. The weak $^4J_{HH}$ coupling between ε1 and δ2 in a spectral region otherwise devoid of cross peaks, makes it possible to detect the chemical shift of both these protons[17,18]. A complete assignment of all the peaks present in the ^1H spectra is not required, only the chemical shift values of our ε1 and δ2 are of interest. Extraction of the T_m is based on fitting the observed chemical shift values to the Henderson-Hasselbach equation[19]. Error values were calculated by using an in-house written Monte-Carlo based algorithm[20]. All reported pK_{aH} values were derived from the Hδ2 chemical shift data since more experimental points for the $δ_2$ protons were determined giving a decreased fitting error.

Introduction of an adenine or thymine at position 9.

The original motif and the base pair modifications to the motif considered in this section are shown in Figure 4.
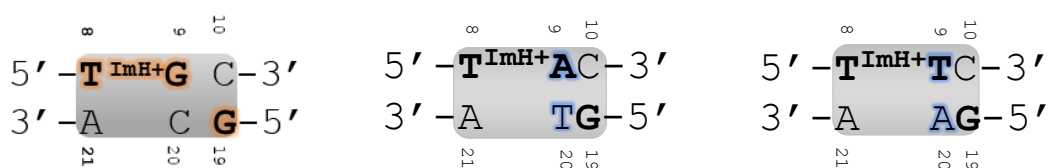

Figure 4. The original motif and the sequences with an A•T/T•A base pair at position 9. The base pairs that are altered with respect to the original interaction motif are indicated in blue.

Table I gives an overview of the melting temperatures (T_m) and the pK_{aH} values. To assist in the interpretation, ΔT_m values are additionally included. ΔT_m values are calculated for the unmodified systems relative to the values for the unmodified sequence of the original motif (reported in shaded cells) while ΔT_m values for the imidazole modified sequences are all relative to those for the corresponding unmodified counterparts. Finally the ΔpK_{aH} values relative to the T_8^{ImH+} nucleoside building block are included. The introduction of A•T base pairs leads to a destabilisation of the unmodified sequences relative to the $WT(G_9C_{20})(C_{10}G_{19})$ (Table 1, shaded cells).

TABLE I. Melting temperature and pK_{aH} overview of the sequences with a T•A and A•T pair at position 9. [a] T_m values were determined in 100 mM NaCl, 10 mM phosphate buffer at pH = 7. [b] Not relevant since T^{ImH+} represents the imidazole nucleoside building block and not a duplex based system, hence there is no melting process. [c] The wild type sequences has no imidazole functionality.

System	Building Block T^{ImH+}	Original motif		A•T base pair		T•A base pair	
		$WT(G_9C_{20})$ $(C_{10}G_{19})$	$T_8^{ImH+}(G_9C_{20})$ $(C_{10}G_{19})$	(A_9T_{20}) $(C_{10}G_{19})$	$T_8^{ImH+}(A_9T_{20})$ $(C_{10}G_{19})$	(T_9A_{20}) $(C_{10}G_{19})$	$T_8^{ImH+}(T_9A_{20})$ $(C_{10}G_{19})$
T_m (°C)[a]	n.r.[b]	58.9 ± 0.2	64.1 ± 0.06	50.8 ± 0.03	60.5 ± 0.13	54.5 ± 0.6	56.4 ± 0.07
ΔT_m (°C)	n.r.[b]	ref	5.2 ± 0.6	-8.1 ± 0.2	9.7 ± 0.1	-4.4 ± 0.6	1.9 ± 0.6
pK_{aH}	7.2 ± 0.02	/[c]	8.7 ± 0.02	/[c]	8.7 ± 0.07	/[c]	8.8 ± 0.02
ΔpK_{aH}	ref		1.5 ± 0.03		1.5 ± 0.07		1.6 ± 0.03

Based on T_m values only, a pK_{aH} regulating motif is only expected to be present in $T_8^{ImH+}(A_9T_{20})(C_{10}G_{19})$ (+ 9.7°C), since the increase for $T_8^{ImH+}(T_9A_{20})(C_{10}G_{19})$ is not pointing at specific imidazole interaction behaviour. Stabilisation in T_m of 1-2°C can be ascribed to nonspecific imidazole backbone interaction. In contrast, the pK_{aH} values (+1.5-1.6 units) indicates the presence of a regulating motif for both the sequences. Specific nOe-contacts and chemical shift perturbation mapping for both sequences however, support the presence of the specific interaction pattern in both cases. Introduction of an A•T or T•A base pair at position 9•20 is therefore tolerated.

<u>Introduction of an adenine or thymine at position 10.</u>

Figure 5 represents the sequences of interest in this section and Table II shows the experimental values of the two sequences with an adenine or thymine at position 10.
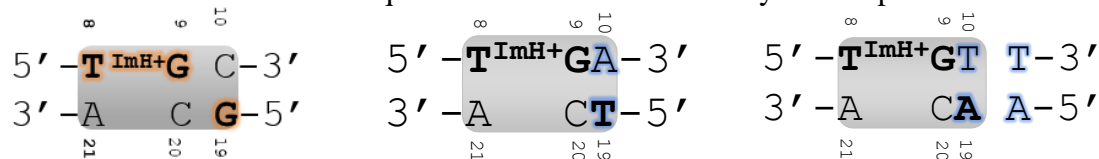

Figure 5. The original motif and the sequences with an A•T/T•A base pair at position 10. The base pairs that are altered with respect to the original interaction motif are indicated in blue.

TABLE II. Melting temperature and pK_{aH} overview of the sequences with a T•A and A•T pair at position 10. [a] T_m values were determined in 100 mM NaCl, 10 mM phosphate buffer at pH = 7. [b] Not relevant since T^{ImH+} represents the imidazole nucleoside building block and not a duplex based system, hence there is no melting process. [c] The wild type sequences have no imidazole functionality.

System	Building Block	Original motif		A•T base pair		T•A base pair	
	T^{ImH+}	(G_9C_{20}) $(C_{10}G_{19})$	$T_8^{ImH+}(G_9C_{20})$ $(C_{10}G_{19})$	(G_9C_{20}) $(A_{10}T_{19})$	$T_8^{ImH+}(G_9C_{20})$ $(A_{10}T_{19})$	$(G_9C_{20})(T_{10}A_{19})$ $(T_{11}A_{18})$	$T_8^{ImH+}(G_9C_{20})$ $(T_{10}A_{19})(T_{11}A_{18})$
T_m (°C)[a]	n.r.[b]	58.9 ± 0.2	64.1 ± 0.06	52.3 ± 0.4	57.3 ± 0.1	56.2 ± 0.2	60.2 ± 0.3
ΔT_m (°C)	n.r.[b]	ref	5.2 ± 0.6	-6.6 ± 0.4	4.9 ± 0.4	-2.7 ± 0.3	4.0 ± 0.3
pK_{aH}	7.2 ± 0.02	/[c]	8.7 ± 0.02	/[c]	8.9 ± 0.02	/[c]	8.5 ± 0.03
ΔpK_{aH}	ref		1.5 ± 0.03		1.7 ± 0.03		1.3 ± 0.03

The stabilisation caused by the introduction of the modified thymine is less pronounced (4-5°C) in comparison with the original motif, but together with the increase in pK_{aH} also these sequences show a stabilizing interaction motif. A final source of information for the imidazole behaviour is the chemical shift perturbation mapping. For $T_8^{ImH+}(G_9C_{20})(A_{10}T_{19})$ (Figure 6), the observed perturbation pattern is similar to the original motif, indicating the presence of a similar specific interaction pattern. Noteworthy is that the perturbation of the CH_3 group of the thymine adjacent to the interaction T_{19} (T_{18}) features a value of 0.24 ppm. Such effect was also seen in the original motif, where the base at position 19 is a guanine, and a characteristic perturbation of 0.44 ppm is observed. Based on the chemical shift perturbation mapping of $T_8^{ImH+}(G_9C_{20})(T_{10}A_{19})(T_{11}A_{18})$ (Figure 6) it becomes clear that the interaction of this sequence is predominantly present on the same strand as the modification, as find not seen before.
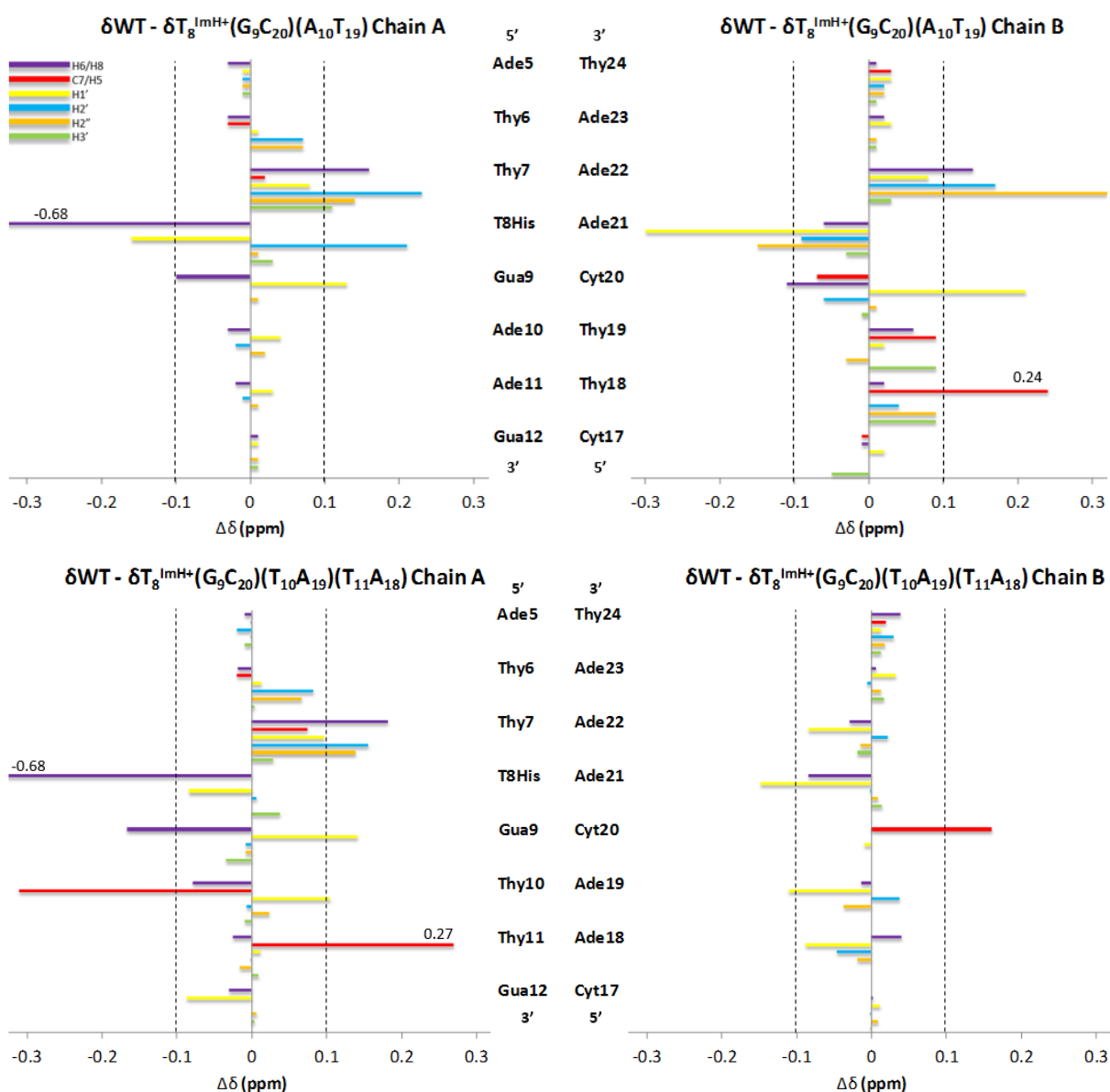
Figure 6. Chemical shift perturbation mapping of $T_8^{ImH+}(G_9C_{20})(T_{10}A_{19})(T_{11}A_{18})$ with respect to the non-modified WT sequences. Only perturbations >0.1 ppm (dashed lines) are considered to be significant. Where necessary the corresponding perturbation value has been added.

The interaction between the imidazole modified thymine on position 8 and a carbonyl donor at position 10 is firstly observed here. Hence, the interaction is less pronounced compared to the original motif.

Could both guanines be replaced by thymines?
From the experimental data it appears that both $T_8^{ImH+}(T_9A_{20})(C_{10}G_{19})$ and $T_8^{ImH+}(G_9C_{20})(A_{10}T_{19})$ systems show a specific interaction pattern and a correlated increase of at least 1.5 pK_{aH} units. Substitution of a single A•T/T•A base pair at position 9•20 or 10•19 in the original motif is tolerated, hence this raised the question whether it is possible to combine these two permutations and obtain a sequence where both G•C/C•G base pairs are substituted by T•A/A•T base pairs as hydrogen bond acceptors in the specific interaction. The classical approach was followed to investigate the systems. First *in silico* MD simulations were carried

out followed by experimental analysis. The introduction of two additional T/A base pairs results in a tract of 7 subsequent less stable A•T base pairs, so a second $T_8^{ImH+}(T_9A_{20})(A_{10}T_{19})$ system is proposed where 2 A•T base pairs are replaced by G•C base pairs in order to interrupt the central TA/AT base pair sequence as can be seen in Figure 7.
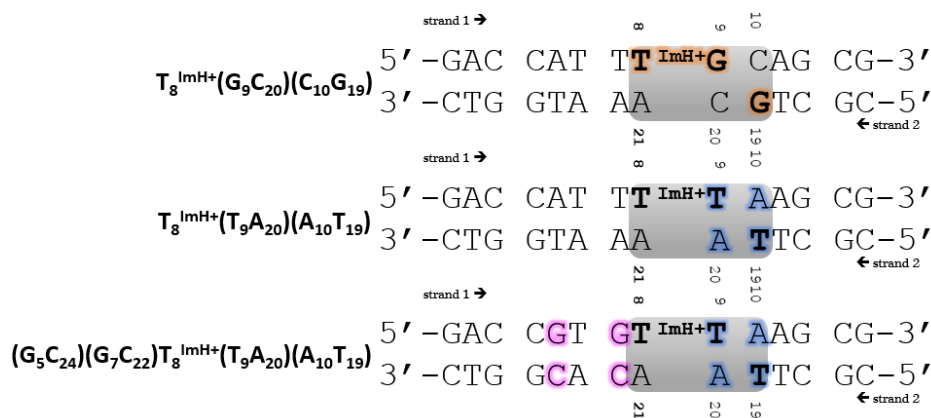

Figure 7. The original motif and the sequences where two A•T/T•A base pairs are introduced, the permutations within the motif are indicated in blue, the changes upstream relative to the modification are indicated in pink.

Although the hydrogen bond persistence (Table III) was low for both the sequences, it was decided to synthesize and characterize them both.

TABLE III. H bond persistence of the original motif $T_8^{ImH+}(G_9C_{20})(C_{10}G_{19})$ and the two other sequences with introduction of a thymine at position 9 and 19. [a] Adapted from [12].

Sequence	Donor	Acceptor	Persistence
$T_8^{ImH+}(G_9C_{20})(C_{10}G_{19})$	$T_8^{ImH+} H\varepsilon_2$	$G_{19} O6$	27.1[a]
	$T_8^{ImH+} H\varepsilon_2$	$G_9 O6$	1.5[a]
$T_8^{ImH+}(T_9A_{20})(A_{10}T_{19})$	$T_8^{ImH+} H\delta_1$	$T_7 OP_2$	3.32 ± 0.32[2]
	$T_8^{ImH+} H\varepsilon_2$	$T_{19} O4$	2.85 ± 0.40
$(G_5C_{24})(G_7C_{22})T_8^{ImH+}(T_9A_{20})(A_{10}T_{19})$	$T_8^{ImH+} H\delta_1$	$G_7 OP_1$	13.53 ± 0.67
	$T_8^{ImH+} H\delta_1$	$G_7 OP_2$	0.79 ± 0.00
	$T_8^{ImH+} H\varepsilon_2$	$T_{19} O4$	3.32 ± 1.11

As usual, the general trend of decreasing T_m by introduction of thymine or adenine base pairs can be observed in Table IV (shaded cells), and conversely, substitution of the T•A base pairs for G•C ones at position 5 and 7 causes the T_m to increase again, though remaining somewhat under that of the original motif. Noteworthy, the melting temperature of $T_8^{ImH+}(T_9A_{20})(A_{10}T_{19})$ is lower than that of the matching unmodified sequence. This is the first time ever that introduction of a T^{ImH+} nucleoside causes a destabilisation. When T^{ImH+} is introduced in a sequence where the A-rich tract has been reduced, the T_m value indicates stabilisation of the DNA strand with 1.7°C. Such a value is typical for the stabilisation by 1-2°C caused by the nonspecific interaction between the protonated imidazole and the negatively charged backbone, *vide supra.*

---

[2] Standard deviations are included here, to indicate variations observed when comparing two independent simulation trajectories of the same sequence, an approach performed for the first time in the context of this research project.

TABLE IV. Melting temperature and pK_{aH} overview of the sequences with two T•A and A•T base pairs introduced. [a] T_m values were determined in 100 mM NaCl, 10 mM phosphate buffer at pH = 7. [b] Not relevant since T^{ImH+} represents the imidazole nucleoside building block and not a duplex based system, hence there is no melting process. [c] The wild type sequences have no imidazole functionality.

System	Building Block T^{ImH+}	Original motif		$(T_9A_{20})(A_{10}T_{19})$		$(G_5C_{24})(G_7C_{22})(T_9A_{20})(A_{10}T_{19})$	
		$(G_9C_{20})(C_{10}G_{19})$	$T_8^{ImH+}(G_9C_{20})(C_{10}G_{19})$	$(T_9A_{20})(A_{10}T_{19})$	$T_8^{ImH+}(T_9A_{20})(A_{10}T_{19})$	$(G_5C_{24})(G_7C_{22})(T_9A_{20})(A_{10}T_{19})$	$(G_5C_{24})(G_7C_{22})T_8^{ImH+}(T_9A_{20})(A_{10}T_{19})$
T_m (°C)[a]	n.r.[b]	58.9 ± 0.2	64.1 ± 0.06	51.7 ± 1.0	50.6 ± 0.1	56.1 ± 0.1	57.9 ± 0.07
ΔT_m (°C)		ref	5.2 ± 0.6	-7.2 ± 1.0	-1.1 ± 1.0	-2.8 ± 0.2	1.7 ± 0.1
pK_{aH}	7.2 ± 0.02	/[c]	8.7 ± 0.02	/[c]	7.9 ± 0.02	/[c]	8.4 ± 0.03
ΔpK_{aH}	ref		1.5 ± 0.03		0.7 0.03		1.2 0.03

The impact of reducing the length of the A-rich tract by investigating the $(G_5C_{24})(G_7C_{22})T_8^{ImH+}(T_9A_{20})(A_{10}T_{19})$ sequence, is significant. The pK_{aH} value is 0.5 units higher compared to $T_8^{ImH+}(T_9A_{20})(A_{10}T_{19})$ and comes within range of the values for sequences where the presence of the motif has been established. Also, the specific nOe contact figure shows more nOe contacts, therefore, the interaction motif appears to be largely intact. Taken together, this shows that not only the type of acceptor base pairs is determinative for the pK_{aH} value and the persistent interaction behaviour, but the overall sequence context and its impact on the structure of the duplex scaffold plays an important role as well.

Conclusions

When comparing all permutations studied so far[1,12,21] we can fine-tune the rules of thump for obtaining a specific interaction motif.

- First, at position 9•20 there are no restrictions in terms of nucleobases and in general terms, the interaction is preserved with a carbonyl acceptor present in the major groove at this level.
- Second, this observation remains true for position 10•19 where it has been shown that a thymine can function both as an acceptor at position 19 (opposite strand vs. T^{ImH+} modification) or position 10 (same strand vs. T^{ImH+} modification).
- Third, simultaneous introduction of a thymine at position 9 and 19 in an otherwise T•A/A•T rich sequence almost completely disrupts the motif. This can be (partially) recovered by interrupting the T•A/A•T-like tract by introducing G•C base pairs. This observation in turn highlights the importance of the sequence context where the motif is introduced.
- A final conclusion is that whenever the interaction motif is formed, its presence can be signified by the perturbation of the methyl chemical shift of a thymine, positioned immediately following the interacting base when this is located on the same strand as the T^{Im} modification, or preceding it when located on the opposite strand. The perturbation is fairly constant over all sequences studies, but differs when the interacting base at position 19 is a G or T, being 0.37 ppm (A•T or T•A base pair present on position 9•20) and 0.44 ppm (in case of the original motif) or 0.24 ppm to 0.27 ppm respectively.

This work shows the possibilities and relevance for the executed bottom up approach of nucleic acid based compounds where both experimental and *in silico* experiments were of great value. In future studies it would be worth executing the experimental analysis of the last four permutations within the three-base pair box corresponding to the original motif. Investigation of sequences $T_8^{ImH+}(C_9G_{20})(C_{10}G_{19})$, $T_8^{ImH+}(G_9C_{20})(G_{10}C_{19})$,

$T_8^{ImH+}(C_9G_{20})(G_{10}C_{19})$ and $T_8^{ImH+}(A_9T_{20})(T_{10}A_{19})$ should complete the story of permutations within the original interaction motif.

Acknowledgments

References

1. Buyst, D., Gheerardijn, V. et al. (2015). Identification of a pKa-regulating motif stabilizing imidazole-modified double-stranded DNA. *Nucleic Acids Research,* 43(1), 51-62
2. Breaker, R. (1997). In vitro selection of catalytic polynucleotides. *Chemical Reviews,* 97(2), 371-390.
3. Close, D. (2013). Calculated pK$_a$'s of the DNA base radicals. *The journal of physical chemistry*, 117, 473-480.
4. Fedor, M. and Williamson J. (2005). The catalytic diversity of RNAs. *Nature Reviews. Molecular Cell Biology,* 6(5), 399-412.
5. Strobel, S. and Cochrane J. (2007). RNA Catalysis: Ribozymes, Ribosomes and Riboswitches. *Current opinion in chemical biology,* 11(6), 636-643.
6. Nobel prizes, (2014). The Nobel Prize in Chemistry 1989 – Press Release. [online] Available at: Nobelprize.org. [Accessed. 11 Apr. 2016].
7. Silverman, S. (2008). Catalytic DNA (deoxyribozymes) for synthetic applications - current abilities and future prospects. *Chemical Communications*, (30), 3467-3485.
8. Lermer, L., Roupioz, et al. (2002). Toward an RNaseA mimic: a DNAzyme with imidazoles and cationic amines. *Journal of the American Chemical Society,* 124(34), 9960-9961.
9. Silverman, S. (2008). Catalytic DNA (deoxyribozymes) for synthetic applications - current abilities and future prospects. Chemical Communications, (30), 3467-3485.
10. Van Gasse, B. (2012). Investigating the structure and stability of modified DNA duplexes with NMR spectroscopy. PhD. Ghent University.
11. Gheerardijn, V. (2014). Functional nucleic acids as tools in catalysis. PhD. Ghent University.
12. Holmes, S.C. and M.J. Gait, *Syntheses and oligonucleotide incorporation of nucleoside analogues containing pendant imidazolyl or amino functionalities - The search for sequence-specific artificial ribonucleases.* European Journal of Organic Chemistry, 2005. 2005(24): p. 5171-5183.
13. Beaucage, S.L. and M.H. Caruthers, *Deoxynucleoside phosphoramidites - a new class of key intermediates for deoxypolynucleotide synthesis.* Tetrahedron Letters, 1981. 22(20): p. 1859-1862.
14. Andrus, A. and R.G. Kuimelis, *Overview of Purification and Analysis of Synthetic Nucleic Acids*, in *Current Protocols in Nucleic Acid Chemistry*2001, John Wiley & Sons, Inc.
15. Vranken, W.F., et al., *The CCPN data model for NMR spectroscopy: Development of a software pipeline.* Proteins-Structure Function and Bioinformatics, 2005. 59(4): p. 687-696.
16. Markley, J.L. and M.A. Porubcan, *Charge-relay system of serine proteinases - Proton magnetic-resonance titration studies of 4 histidines of porcine trypsin.* Journal of Molecular Biology, 1976. 102(3): p. 487-509.

17. Roberts, G.C.K., ed., *NMR of macromolecules: a practical approach*. The Practical Approach Series, ed. B.D.H. D. Rickwood. Vol. 134. 1993: Oxford University Press: Oxford.

18. Silverstein, T. (2012). Fitting Imidazole 1H NMR Titration Data to the Henderson–Hasselbalch Equation. Journal of Chemical Education, 89(11), 1474-1475.

19. Alper, J.S. and R.I. Gelb, *Standard errors and confidence-intervals in nonlinear-regression - Comparison of Monte-Carlo and parametric statistics.* Journal of Physical Chemistry, 1990. 94(11): p. 4747-4751.

20. Buyst, D. (2015). *Design and systematic study of imidazole based DNAzymes.* PhD. Ghent University.

21. De Vries, A. (2016). *Further exploration of the imidazole pK_{aH} motif in DNAzymes through NMR studies of sequence permutations.* Master thesis. Ghent University.