FACULTY
OF MATHEMATICS
AND PHYSICS
Charles University

DOCTORAL THESIS

Václav Římal

NMR Study of Oligonucleotide Structures

Department of Low-Temperature Physics

Supervisor of the doctoral thesis:
Prof. RNDr. Helena Štěpánková, CSc.

Study programme:
Physics

Specialisation:
Biophysics, Chemical and Macromolecular Physics

Prague 2018

I declare that I carried out this doctoral thesis independently, and only with the cited sources, literature, and other professional sources.

I understand that my work relates to the rights and obligations under the Act No. 121/2000 Coll., the Copyright Act, as amended, in particular the fact that Charles University has the right to conclude a licence agreement on the use of this work as a school work pursuant to Section 60 paragraph 1 of the Copyright Act.

In Prague, 26 March 2018 Václav Římal

I would like to thank my supervisor, Helena Štěpánková, for her infinite support during all the years and for her understanding of my peculiar wishes and needs. I am grateful to Pepa Štěpánek for helping me understand what is the aim of the research and for numerous consultations of my results and manuscripts.

My thanks belong to Ondra Socha for a pleasant collaboration and inspiring new ideas. I thank Honza Lang for valuable discussions, critical remarks, and his generous assistance of any kind. I thank all the members of the lab, former and present, for broadening my views and skills and for reminding me that there are always people behind science.

I appreciate the help of the co-workers involved in various parts of this thesis: Pavla Pečínková who guided me through the UV and Raman measurements; Larisa Starovoytová who assisted at the NMR experiments on the 300 MHz spectrometer; Peter Mojzeš and Katka Lindnerová Mudroňová who introduced me the DNA quadruplexes and invited me to join their team; the colleagues from the Faculty of Pharmacy in Hradec Králové, above all Zdeněk Novák, who asked for my contribution to their project on thiobenzanilides, and Vašek Profant for enriching this fruitful collaboration by his quantum-chemical calculations.

I pay tribute to the memory of my mother, who always wished me to finish the doctoral studies, for her endless care since the very beginning until the very end. I thank my father and my brothers Kuba and Vojta for holding together under all circumstances. I thank my wife Katka – she knows me better than I do – and my daughters Máňa and Helenka for showing me what is really important in life.

My ways to work and back and elsewhere have become much more enjoyable thanks to my bike Meriďák.

I thank LaTeX that I used to typeset this thesis in the Palatino font.

> Copak je něco, o čem se dá říci: Pojď se podívat na něco nového? Vždyť to tu bylo už celé věky, bylo to na světě dávno před námi!
>
> *Kazatel 1:10*

> Is there any thing whereof it may be said, See, this is new? It hath been already of old time, which was before us.
>
> *Ecclesiastes 1:10*

During the years of my postgraduate study, I have come across several distinct scientific projects. None of them has ever had any potential to become of general and instant importance to the public. Some have not even led to any publication yet. Where can one find motivation then?

But this is just how the basic research looks like. Science wanders around vainly trying to describe the complexity of the world in simple words. It has all been here before, created – not important whether by Nature or God. And only a small portion of our tasks is successfully completed with satisfactory results.

From the rest

we

learn

Title: NMR Study of Oligonucleotide Structures

Author: Václav Římal

Department: Department of Low-Temperature Physics

Supervisor: Prof. RNDr. Helena Štěpánková, CSc., Department of Low-Temperature Physics

Abstract: The dynamics of nucleic acids plays a fundamental role in the interactions with proteins. Some processes are governed by changes in DNA stability and flexibility caused by sequence alterations or chemical modifications without substantial structural impact.

We employed line-shape analysis of variable-temperature ^{1}H NMR spectra to deeply investigate the melting transition of DNA structures. We observed a significant influence of the sequence beyond the nearest neighbours on chemical shifts and thermodynamics of the double helix. The CpG motif is especially sensitive to its distant surroundings and for a particular oligonucleotide it was even found to undergo an unexpected local conformational transition. A strong cooperativity in the duplex disruption was witnessed by both the equilibrium and kinetic aspects as well as by distant effects of cytosine methylation. We propose a two-state melting scheme also for a fragment of a transcription element proved here to fold as a hairpin with a six-membered loop. In addition, we describe the properties of self-assemblies of riboguanosine and deoxyriboguanosine 5′-monophosphates.

This thesis presents a novel strategy for a detailed description of the thermodynamics of various structural motives of nucleic acids. It can be readily extended to other suitable chemical processes. We open new questions concerning the cooperativity inside nucleic-acid complexes, a promising phenomenon in DNA recognition.

Keywords: DNA, CpG motif, chemical exchange, nuclear magnetic resonance

CONTENTS

1 INTRODUCTION

1.1 Why and how?

Since the original proposal of the base-paired double-helical nature of DNA by Watson and Crick [1], many aspects of NUCLEIC ACIDS (NA) have been studied. Apart from the double helix, various structural motifs, such as hairpin loops, triple helices, and quadruplexes, have been recognised for both RIBONUCLEIC ACID (RNA) and DEOXYRIBONUCLEIC ACID (DNA) [2]. The spatial arrangement plays a fundamental role in NA–protein and NA–NA interactions in living cells. It influences storage and protection of the genetic information, regulation of the gene expression, or the immune response.

The living organism does not behave like a static molecular system, though. The molecules are in a permanent motion and biological processes depend on the stability of intermolecular complexes and flexibility of the chemical species involved. Therefore, the kinetic and dynamic aspects should form integral parts of the NA research. As the experimental and computational methods are evolving, deeper and deeper details about the molecular features are becoming accessible. However, their potential is far from being fully exploited to find answers to the ever-growing range of issues concerning molecular mechanisms of NA function. While the number of three-dimensional NA structures solved continuously grows, works dealing e.g. with the hybridisation kinetics remain relatively scarce. Although the methodology for detailed investigation of molecular processes involving association and dissociation has been developed and widely applied for small molecules, its implementation on biological macromolecules appears rather rarely.

One of the experimental techniques suitable for aqueous solutions is NUCLEAR MAGNETIC RESONANCE (NMR). Its advantage lies not only in the well-known ability to provide structural data of biomolecules, but

also in the high sensitivity of the frequency and shape of the NMR line to the arrangement and interactions of the molecules under study and, moreover, to the fate of these characteristics within a certain time window. Taking into account that the NMR-determined quantities can be obtained separately for individual atomic positions of resonating nuclei, NMR can be agreed as a potentially powerful tool for detailed studies of biomolecular processes.

Prior to the work on the PhD thesis, I devoted myself to an accurate description of the NMR line during the melting of a NA duplex. The dynamic equilibrium between duplexes and single-stranded NA usually corresponds to a fast or intermediate asymmetric chemical exchange. It was shown that a simultaneous least-square fit of a due analytical formula to the experimental NMR spectrum can provide the temperature dependence not only of the equilibrium ratio between duplexes and single strands, but also of the kinetics of duplex formation and dissociation. These previous achievements have been further developed in the course of the doctoral study and used for various molecular systems in solution.

The aim of this dissertation is to apply this approach to determining detailed differences in the temperature-induced changes in short oligonucleotides. Emphasis was placed on comparing the response obtained from distinct positions in the NA chain and on the interpretation of possible differences. For this purpose, a huge amount of measurements of temperature-dependent NMR spectra were performed on a large set of model oligonucleotides. Above all, these were self-complementary DNA octamers with the same proportion of all major bases. The study also involved DNA containing 5-methylcytosine base in the CpG motif, sequences potentially forming hairpin loops, and nucleotides constituting guanine tetrads. In order to illustrate the general robustness of the thermodynamic analysis of chemical processes by the line-shape fitting of variable-temperature NMR spectra, the isomerisation equilibrium and kinetics of a non-nucleic-acid substance – a secondary thiobenzanilide – were also studied in this way.

1.2 Structure of the thesis

The introductory parts relate to NA chemical structure and basic structural motifs (Chapter 2), relevant models of the thermodynamics of chemical processes (Chapter 3), and NMR spectroscopy (Chapter 4). The next two chapters are devoted to the methods used for acquisition (Chapter 5) and analysis (Chapter 6) of the experimental NMR data.

These tools are applied to the duplex disruption of short, eight base-pairs containing palindromic fragments of DNA in Chapter 7. Apart from the evaluation of the melting thermodynamics, the chemical shifts are also subjected to a thorough analysis of their dependence on the nucleotide sequence, temperature, and concentration. Based on the measured values, current approximations are challenged and anomalous behaviour is revealed in a few cases.

The effect of methylated cytosine in the CpG motif, which serves as a regulatory epigenetic mark in the genome, on the spectral and thermodynamic properties of the DNA double helix is discussed in Chapter 8. The NMR measurements are supplemented by the absorption of ULTRAVIOLET (UV) radiation and the RAMAN SCATTERING (RS) spectroscopy.

Chapter 9 is devoted to the application of the methodology established originally for the double-helical DNA to different structural motifs. The question whether an oligodeoxynucleotide based on a transcription element folds as an unimolecular hairpin loop rather than as a duplex is addressed by employing several synergical approaches. In addition, multimolecular complexes of guanosine mononucleotides are investigated.

The line-shape analysis of variable-temperature NMR spectra and determination of equilibrium and kinetic parameters of secondary thiobenzanilide isomerisation are described in Chapter 10.

2 NUCLEIC ACIDS

2.1 Biology

Nucleic acids are the carriers of the genetic information: DNA contains all the details about the primary sequence of all the proteins of an organism and is replicated during cell division; RNA – the so-called MESSENGER RNA (mRNA) – transfers this information to ribosomes where the proteins are created. This is the basic scheme, but much more functions can be ascribed to NA. Many RNA molecules are designated to have different specific functions, such as RIBOSOMAL RNA (rRNA), TRANSFER RNA (tRNA), or various enzymatically active RNA, RIBOZYMES and VIROIDS. Other NON-CODING RNAS (ncRNA) are currently under intensive investigation and numerous purposes have been discovered. These include SMALL NUCLEAR RNA (snRNA) targeting splicing, SMALL NUCLEOLAR RNA (snoRNA) targeting rRNA modifications [3], and MICRO RNA (miRNA) and SMALL INTERFERING RNA (siRNA) post-transcriptionally regulating gene expression [4, 5].

Protein-coding sequences in the DNA – genes *sensu stricto* – constitute only a small part of the genome: 1 %–2 % in humans. This fraction strongly varies between taxons and does not depend on the number of genes [6]. Thus, many regions of DNA are not transcribed into RNA but some of them have regulatory functions instead: promotor and other regulatory sequences interact with transcription factors or other protein molecules which enhance or suppress gene expression. TRANSPOSONS used to be transcribed and translated in the past during evolution, but are – usually – suppressed now. Repetitive SATELLITE DNA in CENTROMERES is bound by proteins forming the central part of chromosome that is essential during cell division. TELOMERES, located at the ends of chromosomes, are shortened after each replication. Intriguingly, there are

even long fragments of DNA whose function is not yet understood, and these are called JUNK DNA. [7]

All the genetic information is stored in DNA, but not every gene is ready to use. How is the cell differentiation established during ontogenesis? This remains a puzzle. However, many aspects are already known. EPIGENETIC phenomena [8, 9] are involved, currently classified into two main categories: METHYLATION OF DNA, which is going to be described in § 2.6 on page 18, and methylation and acetylation of histone proteins [10].

Hence, nature provides various examples of NA–protein and NA–NA interactions, which can be affected by chemical modifications and sequence alterations. Their physical basis often lies in subtle changes in DNA conformation or dynamics – this topic still deserves much interest in the NA research.

2.2 Chemical composition

The basic building block of NA, the NUCLEOTIDE, is formed from a NUCLEOBASE (briefly, a BASE), a saccharide, and a phosphate group [11].

DNA and RNA differ in the choice of the sugar: D-ribose is found in RNA and 2-deoxy-D-ribose in DNA, both locked in the β-furanose heterocyclic form (Fig. 2.1). The base is attached by an N-glycosidic bond to the anomeric carbon of the pentose to form a NUCLEOSIDE. Five major bases are prevalent in natural NA: three of them derived from pyrimidine – CYTOSINE, THYMINE, and URACIL – and two derived from purine – ADENINE and GUANINE (Fig. 2.2). Uracil rarely occurs in DNA and thymine only scarcely resides in RNA. Accordingly, there are four possible bases which can usually be placed in a particular DNA or RNA nucleoside.

The nucleobases are heterocyclic aromatic compounds with planar geometry of the rings. They contain substituents many of which are polar and can give rise to HYDROGEN BONDS as acceptors (lone electron pairs localised on nitrogen and oxygen) or donors (hydrogen atoms bound to nitrogen). These hydrogen bonds mainly serve for base recognition and formation of BASE PAIRS. It should also be mentioned that the amino groups occupy a planar conformation due to the delocalisation of the

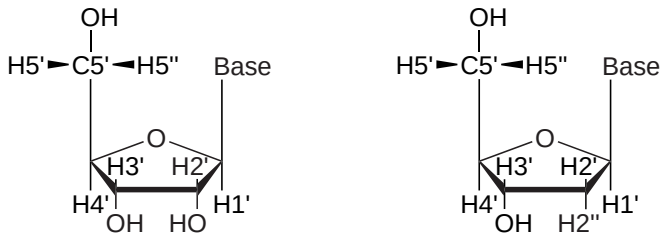


Figure 2.1 Schemes of D-ribose (left) and 2-deoxy-D-ribose (right) in their β-furanose forms and atom numbering in nucleosides. 'Base' denotes the position where a base is bound

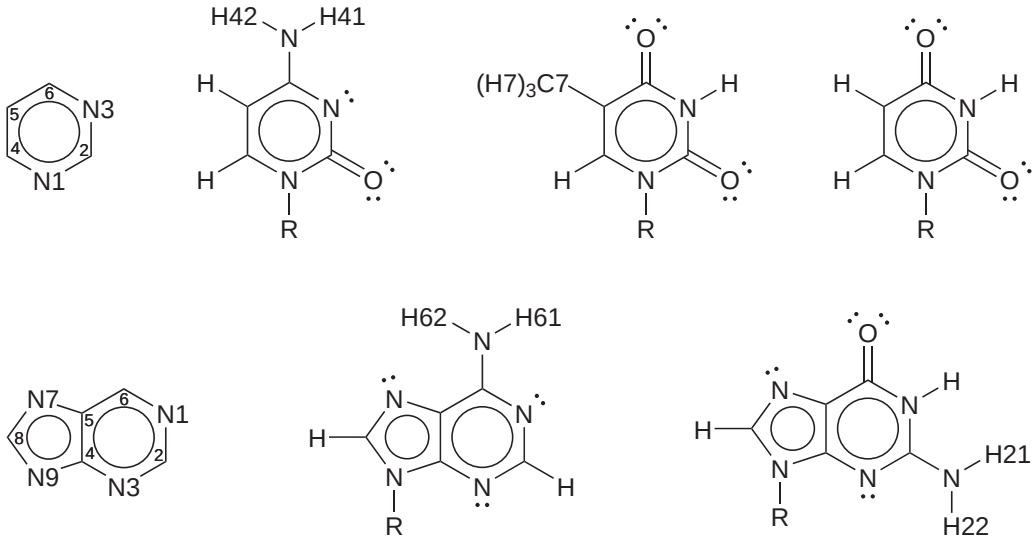


Figure 2.2 Schemes of the major bases in NA. Top row: pyrimidine and pyrimidine-derived bases – uracil, thymine, and cytosine (from left to right). Bottom row: purine and purine-derived bases – adenine and guanine (from left to right). The numbering of the atoms [12] is indicated in the pyrimidine and purine molecules and in the outer moieties of the bases where different from the ring atom bound to it. 'R' denotes the position where a saccharide is bound in nucleosides. Circles represent the delocalized π electrons. The lone electron pairs in sp^2 orbitals of the bases are indicated as double dots

amino-nitrogen lone electron pair into the aromatic system that causes the sp^2 hybridisation of the other valence orbitals.

Phosphate can form an ester bond next to the C5′ carbon of a nucleoside, thus creating a nucleotide. Longer NA molecules are composed of nucleotides covalently bound by linkages between the phosphate and C3′ into a non-branching chain. This results in an asymmetric NA strand having distinguished 5′ and 3′ ends (Fig. 2.3). A short, usually synthetic molecule comprising a small number of nucleotides is called an OLIGONUCLEOTIDE or an OLIGODEOXYNUCLEOTIDE (ODN).

Apart from the above mentioned components, alternatives exist in the native nucleic acids. DNA contains some fractions of minor bases, such as 5-METHYLCYTOSINE (m^5C) and 5-HYDROXYMETHYLCYTOSINE (hm^5C) that are going to be discussed later (§ 2.6 on page 18). RNA reaches much greater variability: XANTHINE, HYPOXANTHINE, 7-METHYLGUANINE, DIHYDROURACIL and other nucleobases as well as PSEUDOURIDINE, an example of a modified nucleoside, are commonly present, especially in tRNA. Different configuration of the phosphate linkages are known in living cells, too, e.g. (2′ → 5′)-RNA oligonucleotides or cyclic nucleoside 3′, 5′-monophosphates. Other substituents can also appear in NA after an action of some foreign chemical compound, chronic inflammation, or ultraviolet or X-ray irradiation, which are most often cytotoxic or mutagenic and not at all desired by the cell[1] [13].

Various modifications may be introduced artificially to any part of NA: for instance, fluorescent or paramagnetic bases, deaza modifications of bases, non-polar or aliphatic bases, abasic nucleosides, 2′-methyl or O2′, C4′-methylene ribose (the latter forms the so-called LOCKED NUCLEIC ACIDS), XENO NUCLEIC ACIDS containing different sugar residues or even non-saccharide moieties instead, phosphonate and phosphorothioate linkers, or PEPTIDE NUCLEIC ACIDS are widely used in research, chemical industry and pharmacology for numerous purposes.

[1] evolution of species may gain in this way, though, as well as the therapy against cancer

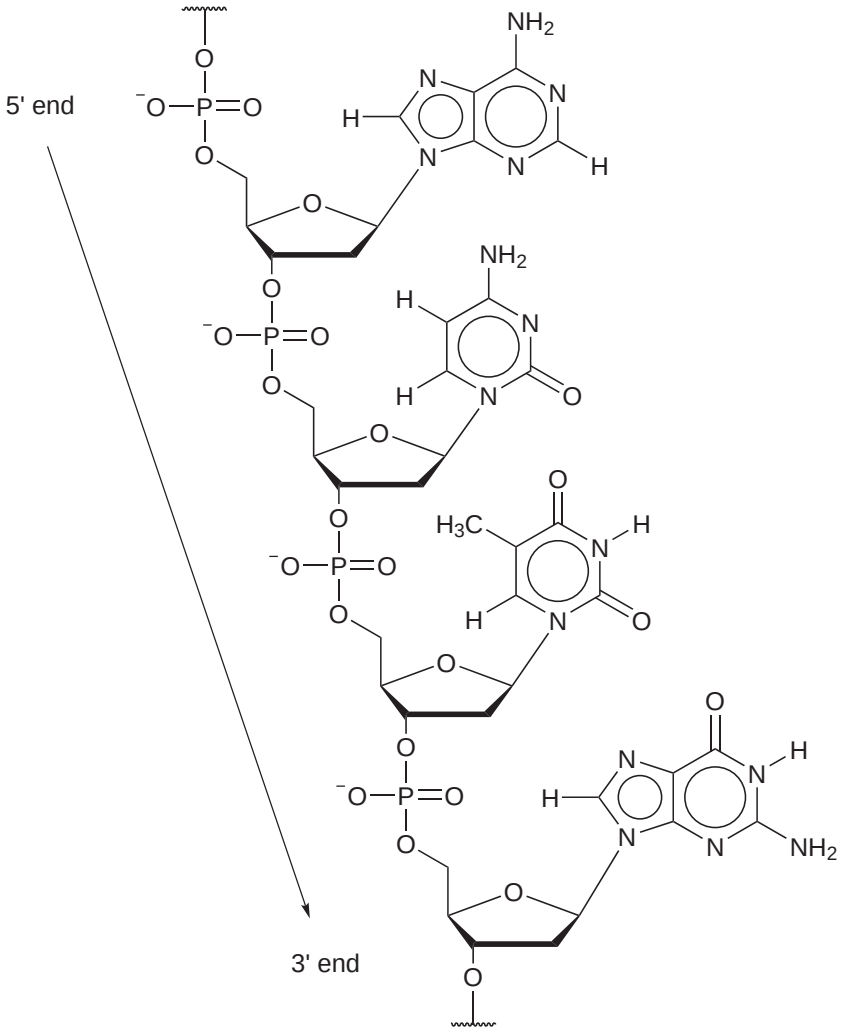


Figure 2.3 Scheme of a DNA fragment composed of four nucleotides in a chain with the ACTG sequence. The hydrogen atoms of deoxyribose are omitted for clarity

2.3 Used nomenclature

Throughout this work, conventions of naming NA and their constituents are based on the recommendations by the International Union of Pure and Applied Chemistry (IUPAC) and the International Union of Biochemistry (IUB) [12, 14, 15].

The names of nucleobases are abbreviated to single characters (A, C, G, T, and U), which may also mean the whole nucleoside or nucleotide. To stress that a free nucleotide, i.e., a nucleoside 5′-phosphate, rather than

a nucleotide as a part of a larger molecule is concerned, the abbreviations AMP, GMP, and so on are to be used.

A NA molecule is described by the sequence of base abbreviations, read from the 5′ to the 3′ end. A 'p' can be added to represent the 3′ → 5′ phosphate linkage explicitly. It is assumed that phosphates are not present at the ends of synthetic oligonucleotides. When not clear from the context (most of the experiments in this thesis were conducted on DNA), a prefix 'd' or 'r' should be used to specify DNA or RNA, respectively. As an example, a DNA octanucleotide can be depicted as d(CpApTpCpGpApTpG) or, more conveniently, CATCGATG.

A non-covalent interaction between nucleobases or whole molecules is indicated by a central dot as in a C·G base pair.

Incomplete specification of a base in a sequence can be expressed by a single letter, too [15]: N represents any nucleotide (or nucleoside or base), R any purine, Y any pyrimidine, W adenine or thymine, S cytosine or guanine, etc. These abbreviations will be reminded whenever encountered. Their use is suitable when a set of similar NA chains are described together or when talking about general properties of short pieces of NA having some bases in common and some bases different.

A particular atom in a NA molecule is addressed by the one-character symbol of the base in the referred nucleotide, its number (starting from 1 on the 5′ end), the element symbol, and the number of atom according to Fig. 2.1 or Fig. 2.2[2]. Thus, the methyl carbon of the third thymine from CATCGATG is T3C7 and the two hydrogens bound to carbon C2′ of the sixth adenosine are A6H2′ and A6H2″. Omitting any part of this pattern means that it covers all possible options, e.g. A2 represents the second nucleotide, adenosine phosphate, and CH5 stands for hydrogens H5 of all cytosine bases from the molecule.

2.4 Structural properties

Although the chemical composition of NA and their PRIMARY STRUCTURE given by the nucleotide sequence are well defined, spatial arrange-

[2] no punctuation is inserted since this would be superfluous

ments of NA molecules are diverse due to a broad variability of combination of dihedral angles and topologies of non-covalent interactions.

2.4.1 Conformation of nucleotides

The torsion angle around the glycosidic bond, χ (Fig. 2.4), occupies two main ranges: in the *anti* orientation, the amino and keto groups and the C2 carbon of the base are found apart from the sugar, whereas in the *syn* orientation, they lie above it. In both cases, the plane of the base is roughly perpendicular to the plane of the pentose and it approximately bisects the O4′—C1′—C2′ angle. In nucleotides, the *anti* conformation is much more preferred, while the *syn* orientation can mainly be acquired by guanosine in certain cases, including GMP in solution. [16]

The five dihedral angles in the pentose ring are kept tightly interconnected due to the limitations of the cycle. The SUGAR PUCKER usually falls into two distinct types: C2′-*endo* and C3′-*endo*, called also SOUTH (S) and NORTH (N), respectively (Fig. 2.4) [12]. The torsions along the phosphodiester backbone (Fig. 2.4) are, most usually, *trans* at the dihedral angles β and ε and $+$*gauche* at γ [12]. This leads to a preference for the atoms P—O5′—C5′—C4′—H4′ to lie in one plane arranged in a shape of a 'W' and the atoms C4′—C3′—O3′—P to lie in one plane, too. [16]

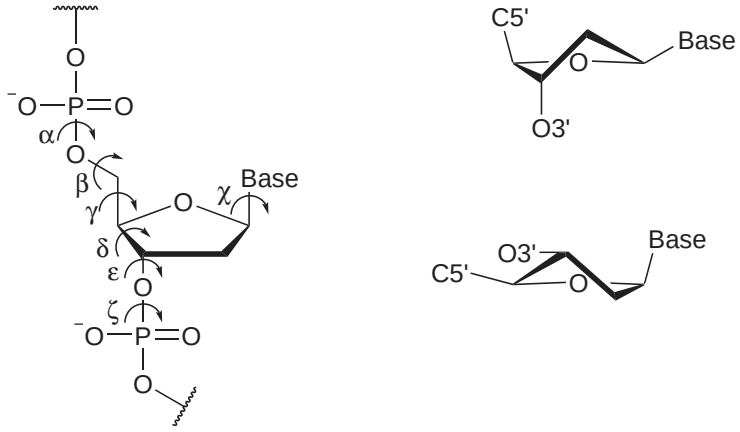


Figure 2.4 Left: a DNA fragment showing the dihedral angles in the sugar–phosphate backbone (α–ζ) and the orientation of the base (χ). Right: the two major sugar puckers, C2′-*endo* (top) and C3′-*endo* (bottom)

The most common configurations described above are subjected to exceptions that can be found in some unusual motifs. One shall also keep in mind that in solution, more conformations may be present in a dynamic equilibrium.

2.4.2 Secondary structure

NA chains can fold into various mono- and multimolecular structural motifs, giving the complexes a SECONDARY STRUCTURE. Only non-covalent phenomena are responsible for these arrangements:

a. ELECTROSTATIC FORCES between ions: electrostatics repulses the negatively charged phosphate groups from each other and cations are needed to compensate their charge;
b. STACKING INTERACTION between aromatic moieties arises from the combination of electrostatic interactions between charge distributions associated with the electron density and VAN DER WAALS FORCES between static or instantaneous electrical dipoles;
c. hydrogen bonds between partially positively charged hydrogen atoms bound to electronegative elements and lone electron pairs of oxygen or nitrogen;
d. HYDROPHOBIC EFFECT: a consequence of the different dipole–dipole forces between water molecules and polar and non-polar parts of NA.

Several basic types of secondary structure are recognised and these are defined by their stoichiometry, topology, and ordinary ranges of conformational parameters. The main monomolecular motif is a HAIRPIN when the NA folds onto itself. Bimolecular complexes are called DUPLEXES and they occur in a double-helical overall conformation (Fig. 2.5). Helical structures incorporating more NA strands, TRIPLEXES and QUADRUPLEXES, have also been described. More complicated, branching complexes can be represented by CRUCIFORMS and three-, four-, and even five-way JUNCTIONS [17, 18].

The three-dimensional structures of real systems are usually determined by combinations of experimental work and computer simulations. X-RAY DIFFRACTION (XRD) is the most widely employed experimental method for this purpose, followed by NMR spectroscopy. Both of them

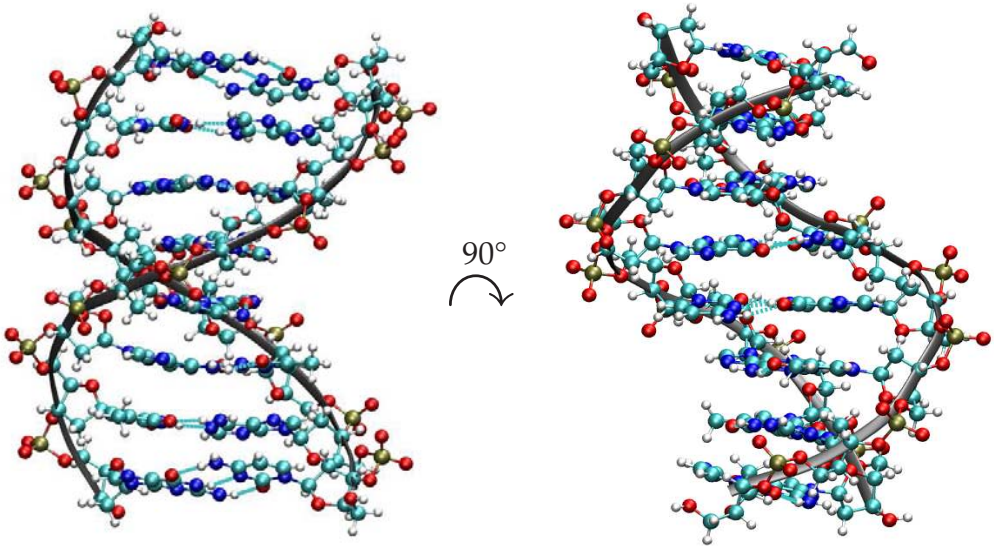


Figure 2.5 NMR-determined structure of a B-DNA double helix made by self-complementary sequence CATGCATG. The all-atoms ball-and-stick model, colour-coded by element type, with the backbones highlighted by grey ribbons, was created in VMD [19] from the Protein Data Bank (PDB) entry 1D18 [20]. Left: a view perpendicular to the helical axis. Right: a view rotated by 90° around the vertical axis, clockwise when viewed from the top

have their advantages and their drawbacks. XRD provides a straightforward connection between the structure and the diffraction pattern, but requires preparation of a sufficiently large monocrystal, which can turn out to be a very difficult task. NMR works fine in solution, which is a more natural environment than a solid crystal, and the sample is prepared more easily. However, the elaborate pulse sequences used for structural studies are quite time-consuming as is also their analysis requiring deep understanding of the subject. Moreover, the heavier the molecule is, the broader lines and hence the lower the signal-to-noise ratio the acquired spectra have. Both XRD and NMR are followed by iterative MOLECULAR DYNAMICS (MD) refining the spatial structure.

2.4.3 Duplexes

A duplex can be formed by two DNA molecules, two RNA molecules, or even by one DNA and one RNA molecule, forming a hybrid duplex. The polar sugar–phosphate backbone is exposed to the solvent, whereas

the more hydrophobic nucleobases are hidden in the middle of the DOUBLE HELIX (Fig. 2.5). The bases from the opposing strands are paired to each other: WATSON–CRICK BASE PAIRING is the major scheme in which COMPLEMENTARY BASES form hydrogen bonds, two in A·T and three in G·C (Fig. 2.6). A pair of grooves, called MAJOR GROOVE and MINOR GROOVE, propagates along the helix.

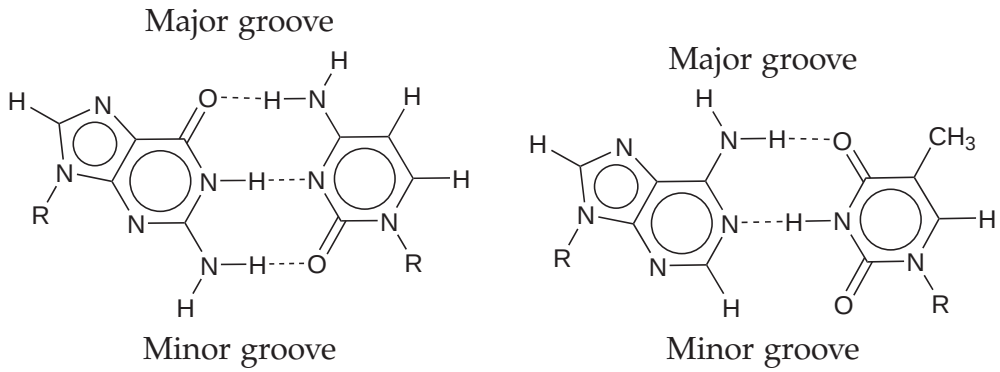


Figure 2.6 Watson–Crick pairing between complementary bases. Left: G·C. Right: A·T. Dashed lines indicate hydrogen bonds, 'R' denotes attachment of the saccharide. Positions of the grooves in a double helix are depicted

Three basic structural families are distinguished: A-FORM, B-FORM (Fig. 2.5), and Z-FORM, the first two being right-handed helices, the last one left-handed. These duplexes have antiparallel orientation of the two strands, so that the 3′ end of one molecule is aligned with the 5′ end of the other. What kind of the double helix a duplex adopts depends on the primary structure and outer conditions, such as pH, ionic composition, or – in crystal – the level of humidity [16].

Local differences from the global averages are commonly found in real duplex structures. Intermediate structures, as well as junctions between the standard conformations can appear. Moreover, deviations from the Watson–Crick pairing, such as MISMATCHES, BULGES, WOBBLE or HOOGSTEEN pairs, can be introduced.

2.4.4 Hairpins

Hairpins are monomolecular motifs consisting of a double-helical STEM and a LOOP. There is a large structural variability mainly in the loop

which usually has three [21] or four [22] nucleotides in DNA but longer turns are not scarcely present in RNA molecules. The stem–loop junction usually consists of one or more non-canonical base pairs.

2.4.5 G-tetrads

One important example of stable Hoogsteen pairs deserves our attention: the G-TETRAD (Fig. 2.7), often called also G-QUARTET. These motifs are formed by self-association of GMP in highly concentrated solutions [23, 24] or in solid phase and are also found in G-QUADRUPLEXES, structures made of four NA chains rich in guanine [25, 26].

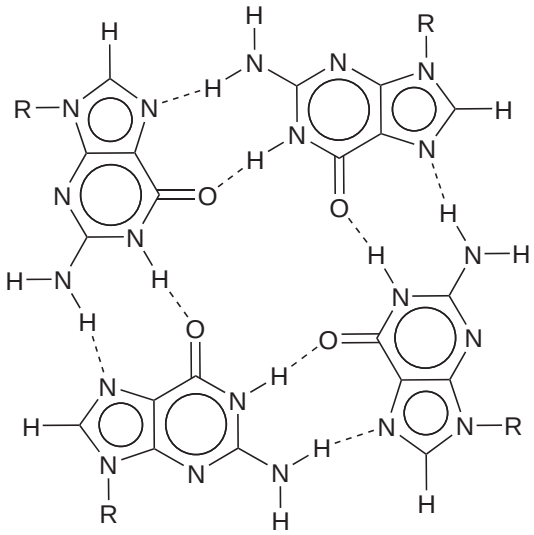


Figure 2.7 A G-tetrad. Dashed lines indicate hydrogen bonds, R denotes attachment of the saccharide

G-tetrad is a planar ring of four guanines joined together by Hoogsteen hydrogen bonds. A cation residing in the tetrad centre can stabilise the motif and its type often directs the resulting structure and stability of the whole complex [27]. In both GMP self-associates and G-quadruplexes, the sugar pucker can vary and there is no general preference for the *anti* or *syn* conformation either, although each group of parallel strands has the same base orientation within a single G-tetrad of a quadruplex [28, 29].

2.5 Stability of secondary structures

The choice of a particular folding of a NA complex depends on many conditions. Among them, the amount, valence, as well as type of cations present in the solution have profound impact on the stability of all structural motifs. Therefore, during *in vitro* studies of NA properties, buffers with a stable pH and salt concentrations in excess of that of nucleotides are commonly employed. Determination of the thermodynamic parameters describing the structural stability is an important cue to the understanding of NA functions.

2.5.1 Melting temperature

Temperature can be easily handled during experiment and can induce dramatic structural changes. Since NA structures are generally well ordered and therefore entropically unfavourable, increasing temperature destabilises them. Accordingly, there is a transition midpoint at which half of the molecules is folded and the other half is in the RANDOM-COIL state[3]; the corresponding temperature is called the MELTING TEMPERATURE, T_m. It is quite an easy-to-use quantity and possible differences in T_m monitor how the stability is affected. Most often, the TWO-STATE MODEL neglecting any partially folded species is selected to describe the melting of short oligonucleotides, but this assumption may be violated in higher salt concentrations [30] and traces of unfolded molecules can persist to low temperature [31]. More details on the melting thermodynamics are going to be discussed in Chapter 3.

Many experimental methods which can measure T_m have been developed. Some of them can be considered as more or less global since they provide information on the overall stability:

a. UV absorption spectroscopy mostly relies on the HYPOCHROMIC EFFECT – base stacking reduces the extinction coefficients in the UV range of the electromagnetic radiation [32, 33];
b. CIRCULAR DICHROISM (CD) is the chirality-sensitive modification of UV absorption;

[3] things become more complicated when more than two types of structure are involved

c. DIFFERENTIAL SCANNING CALORIMETRY (DSC) directly measures the heat transferred during a temperature change;
d. FÖRSTER RESONANCE ENERGY TRANSFER (FRET) senses the distance between two fluorescent probes that changes after melting;
e. ELECTRON PARAMAGNETIC RESONANCE (EPR) can measure the distance between two paramagnetic labels [34].

Apart from these techniques, there are others that can also reveal some local structural changes caused by the temperature. It is worth to mention in particular:

a. NMR, especially ^{1}H and ^{31}P spectroscopy well suitable for NA, profits from the dependence of chemical shifts on local environment;
b. Raman scattering spectroscopy (RS) and its extensions, such as RESONANCE RS (RRS), exploit the differences in frequencies and amplitudes of vibrational bands of the folded state and of the random coil.

Further information about the stability of a complex can be obtained by some other approaches measuring different quantities instead of T_m, including ISOTHERMAL TITRATION CALORIMETRY (ITC). The experimental methods can be supplemented with MD or other simulations which can bring more insight into the thermodynamic properties of NA structures.

2.5.2 Stability predictions

Lots of efforts have been spent to predict the stability of NA structures. Ideally, one would obtain a complete thermodynamic description from some calculation whose input would just be the nucleotide sequence. Of course, *ab initio* quantum-mechanical computations are far too much complicated. Thus, empirical approaches based on analyses of large sets of various experimental data have been developed. The predominant simplification occurring in these calculations is the so-called NEAREST-NEIGHBOUR (NN) model[4]: the global extensive quantities of a transition are obtained as sums of the values valid for individual base pairs and base-pair steps. Some modifications or penalties are added in some ap-

[4] approximations used in estimating other NA properties, such as structural or spectral parameters, can carry the same name, too

proaches to properly reflect the terminal base pairs, helix initiation, symmetry of self-complementary strands, and other miscellanies.

The methods were pioneered on RNA duplexes [35, 36]. The NN parameters for B-DNA obtained in 1980s from DSC and UV measurements [37] were improved a decade later using a more extensive set of sequences by UV melting experiments [38]. The existing data from different laboratories were then gathered and the model was unified, together with a prediction of salt-dependence of the duplex thermodynamics [39]. The NN model has also been refined for RNA duplexes [40] and extended for DNA complexes containing mismatches [41, 42, 43, 44, 45, 46], dangling ends [47], hypoxanthine bases [48], or hairpin loops [49] and even for DNA–RNA hybrid duplexes [50]. More folding possibilities can be combined, too [51]. Several web interfaces or stand-alone applications for predicting the stabilities of NA complexes have been built upon these thermodynamic data, namely DINAMelt [52], mFold [53], dnaMATE [54], and OligoCalc [55].

2.6 Cytosine methylation in DNA

Not only the four major bases (A, T, C, and G) are present in the native DNA. Above all, cytosine is very often the subject for a naturally occurring chemical modification: its methylation on carbon C5. The methyl group of 5-methylcytosine (m^5C, Fig. 2.8) doesn't alter the Watson–Crick base pairing scheme with guanine: having the same position as the methyl group of thymine, it points toward the major groove of a double helix.

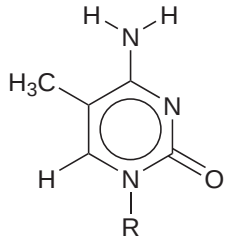


Figure 2.8 5-methylcytosine. 'R' denotes the position where a saccharide is bound in nucleosides. Atom numbering follows Fig. 2.2 (page 7) and the atoms of the methyl group all carry the number 7 in analogy with thymine

2.6.1 CpG motif

The hotspot for the DNA methylation is the CG dinucleotide sequence, the so-called CPG MOTIF. In vertebrates[5], its global frequency is surprisingly low (only 1 %–3 % [57, 58] compared to the unbiased probability $1 : 16 = 6.25\,\%$) and m^5C is found in 70 %–90 % of CpG sites [59], despite the total degree of cytosine methylation reaches only $\sim 1\,\%$ [60, 61]. CHG and CHH contexts[6] are commonly methylated in plants, too [62].

The CpG methylation is an EPIGENETIC mark that has a key role in controlling gene expression as it usually represses it [63, 64, 65]. It is a dominant mechanism of cell differentiation [10, 66, 67], X-chromosome inactivation [68], and genomic imprinting [69]. Defective changes in DNA methylation occur in cancer cells [70, 71, 72], as well as in other serious diseases [73, 74, 75].

The methylation pattern of the genome undergoes several reprogramming periods during ontogenesis [63, 76, 77, 78, 79]. It may even be modified during prenatal or even postnatal period and transferred to the mature germ cells [80, 81] – acquired traits can be propagated to descendants, leading to a non-Mendelian, Lamarckian-like contribution to the evolutions of species [82, 83]. The molecular mechanisms of the methylation management are not precisely known, although some functions of ncRNA have been discovered [84, 85, 86, 87] and active demethylation in animals involves 5-hydroxymethylcytosine (hm^5C) as an intermediate [88, 89]. Plants are already proven to utilise elaborate RNA machineries [62, 90].

While highly methylated in vertebrates, the CpG motif in bacteria is mostly found unmodified [56]. This difference is used by the immune system to distinguish the intruder's DNA [91, 92, 93]. Various synthetic ODN with unmethylated CpG have immunostimulatory effects which promises their use in prevention or treatment of numerous diseases [94, 95, 96].

[5] 'Our understanding of the molecular biology of animals is increasingly biased towards those with backbones. In fact the vertebrate branch accounts for a relatively small fraction (less than 5 %) of the animal kingdom, though it does include scientists (and grant givers).' [56]

[6] H = not G

Apart from its notable biological importance, there are remarkable features concerning the structure and dynamics of the CpG motif. Different geometry of CpG in otherwise B-DNA duplexes was observed both in crystal [97] and in solution [98]. A greater flexibility of the duplex $(CATCGATG)_2$ relative to $(GTACGTAC)_2$ was described [98]. The sugar pucker, helicoidal and backbone parameters of CpG [58, 99, 100] as well as its conformational dynamics [101, 102] can also significantly differ in DNA duplexes sharing the same NCGN tetranucleotides.

The above mentioned works lead to a conclusion that the world of NA structures and dynamics is not as simple as it might seem. Generally, the NN model is not accurate enough to describe detailed properties of a DNA double-helix. The extraordinarily variable and flexible CpG motif continually attracts the curiosity of many researchers. Its peculiarities might be very important in targeting DNA by proteins and in intercalating or surface-binding of small molecules.

2.6.2 A weak impact on structure

The B-DNA changes its overall geometry only very slightly after the CpG sequence is methylated (or hydroxymethylated). Such are conclusions from various works employing XRD [103, 104, 105], NMR [106, 107], and MD [107, 108]: structural parameters are weakly altered [103, 104, 107], in particular the sugar–phosphate backbone torsion angles and the helical twist [101, 106]. Even outside the CpG motif, conformational changes induced by cytosine methylation remain small [108]. Exceptions are found near A-tracts where methylation makes the minor groove narrower [109] and changes the overall bend of the duplex [110].

2.6.3 Increased duplex stability and rigidity

Although the impact of cytosine methylation on the structure is marginal, duplex stability increases after CpG methylation. Despite the stabilisation can be also very weak, it has been experimentally observed for ODN [103, 106, 107, 109, 111], as well as for longer DNA strands counting less than a hundred [112, 113] or even several hundreds to thousands base pairs [108]. Taking the intrinsic polymorphism of the unmethylated CpG motif into account, the structural changes induced by cytosine methy-

lation seem to be closely tied to the conformation and rigidity of the unmethylated duplex. Complete or partial reversal of the m^5C effect has been reported after oxidation to hm^5C [108, 111, 113]. Some data concerning self-complementary ODN duplexes collected from the literature are shown in Table 2.1. Increase of T_m was also measured for DNA forming a hairpin [114], a triplex [115], and a G-quadruplex [116].

A very similar effect on the duplex stability was observed by comparing DNA duplexes with thymine and its demethylated counterpart, uracil [118, 119]. These results are in accordance with the idea of duplex stabilisation by adding a methyl group to C5 of any pyrimidine nucleobase: it affects the electronic distribution in the aromatic ring resulting in decrease of its electric dipole moment by 1.2 D [120] that makes the base more hydrophobic. DENSITY FUNCTIONAL THEORY (DFT) *ab initio* calculations revealed that the stacking interaction is enhanced [121]. This was already suggested in the NMR study of short ODN [106].

The duplex flexibility in the terms of local motions is also lowered by cytosine methylation [122]. MD has revealed that the steric hindrance combined with changes in polarity of the substituent [107, 108] and the modified stacking [123] are responsible for the increased rigidity.

2.6.4 CpG methylation: an open question

The biological importance of the CpG methylation described above implicates that many protein–DNA interactions are methylation-dependent. Since it is generally accepted that the spatial structure of CpG-containing DNA is not strongly altered after the methylation of cytosine, it seems that it is the reduced flexibility of the methylated CpG motif that plays a substantial role in the discrimination between unmodified and methylated DNA. The everlasting confrontation of the INDUCED FIT and CONFORMATIONAL SELECTION paradigms is probably an indispensable issue, too [124].

Although there has been written a lot about the malleable CpG motif and its methylation, conclusive explanations of its extraordinary behaviour on a detailed level are still missing. How does the sequence context around CpG influence the duplex stability? And *vice versa*, how far along the DNA double helix does the CpG dinucleotide influence its sur-

Table 2.1 Review of experimentally determined melting temperatures, T_m, and enthalpies, ΔH, of duplex formation of self-complementary ODN together with the total concentration of single strands, c, and methods used

DNA sequence[a]	$\frac{c}{\mathrm{mM}}$	$\frac{T_m}{\mathrm{K}}$	$\frac{\Delta H}{\mathrm{kJ \cdot mol^{-1}}}$	Method[b]	Solvent composition[c]	Source
CATCGATG	0.1	320.5	−236	UV (var. conc.)	10 mM NaCacB, 1 M NaCl	[38]
	0.006	305		UV (inflex)	PB, $I = 100$ mM	[117]
	6	328		NMR ^{31}P (inflex)	PB, $I = 100$ mM	[117]
	7.0	326.6	−341.8	NMR AH2, TH7[e]	NaKPB, $I = 100$ mM	[106]
CAT**M**GATG	7.0	329.8	−316.2	NMR AH2, TH7[e]	NaKPB, $I = 100$ mM	[106]
CTTCGAAG	0.006	292		UV (inflex)	PB, $I = 100$ mM	[117]
	6	318		NMR ^{31}P (inflex)	PB, $I = 100$ mM	[117]
	7.0	323.6	−266.5	NMR AH2[e]	NaKPB, $I = 100$ mM	[106]
CTT**M**GAAG	7.0	322.6	−272.7	NMR A6H2[e]	NaKPB, $I = 100$ mM	[106]
CGCGAATTCGCG	[d]	344.8		NMR H2, H5, H6, H7[e]	100 mM PB	[30]
	0.65	344.4	−214	DSC (integral)	100 mM NaCl	[30]
	0.65	344.4	−155	DSC (van 't Hoff)	100 mM NaCl	[30]
	0.055	332.0	−240	UV (van 't Hoff)	10 mM NaPB, 50 mM NaCl	[103]
CG**M**GAATTCGCG	0.055	332.0	−219	UV (van 't Hoff)	10 mM NaPB, 50 mM NaCl	[103]
CGCGAATT**M**GCG	0.055	333.3	−206	UV (van 't Hoff)	10 mM NaPB, 50 mM NaCl	[103]
CG**H**GAATTCGCG	0.055	332.6	−244	UV (van 't Hoff)	10 mM NaPB, 50 mM NaCl	[103]
CGCGAATT**H**GCG	0.055	329.5	−184	UV (van 't Hoff)	10 mM NaPB, 50 mM NaCl	[103]

[a]Modified bases (M = m^{5}C, H = hm^{5}C) in bold. [b]NMR: analysis of $\delta(T)$, mean values of indicated nuclei; integral: area under the curve; van 't Hoff: fit of the melting curve; inflex: maximum of the melting-curve first derivative at T_m; var. conc.: fit of $T_m(c)$. [c]CacB: cacodylate buffer; PB: phosphate buffer; alkali ions prefixed if specified; I: ionic strength. [d]Not provided but we expect the same as for DSC. [e]T_m as transition midpoints; where provided, ΔH from the slope of a linearised plot

roundings? Additional experiments are needed in order to gain better insight into the molecular mechanisms governing the methylation-induced changes in DNA.

2.7 Serum response element

The SERUM RESPONSE FACTOR (SRF), containing the DNA-binding MADS domain, controls gene transcription by binding its target DNA sequence, the SERUM RESPONSE ELEMENT (SRE) [125]. SRF binds the DNA duplex as a dimer and promotes a 70° DNA bending [126, 127]. The SRF dimer attached to SRE is then bound by a TERNARY COMPLEX FACTOR [128, 129] to activate the transcription.

2.7.1 Sequence and structure

SRE carries the CArG-box sequence CCWWWWWWGG[7]. The promoter of *c-fos* proto-oncogene contains SRE with mutually complementary extensions in both directions [125, 129]. Such pattern containing complementary fragments separated by a non-complementary part has been found in various other regulatory sequences in the genome. These can potentially create cruciforms since its two strands are partially palindromic, in principle allowing formation of hairpins in both strands [130, 131]. Such an unusual structure would strongly facilitate its recognition by SRF.

The key question that needs to be answered is: do the individual DNA strands of *c-fos* SRE really fold as hairpins? UV-melting studies on solutions of SRE single strands revealed a clear transition between a folded and unfolded state [132]. It has been suggested that this could be the hairpin topology; however, an irrefutable proof is still missing. There remains a second option that a duplex with imperfect base pairing could be formed from two identical molecules. Both folding possibilities (Fig. 2.9) share the five terminal Watson–Crick base pairs but differ in the topology of the six-nucleotide central part of the ODN sequences (A6–A11).

2.7.2 DNA hairpins with long loops

No other structural or thermodynamic studies of the naked SRE DNA are described in the literature. In general, DNA hairpins containing more

[7] W stands for A or T

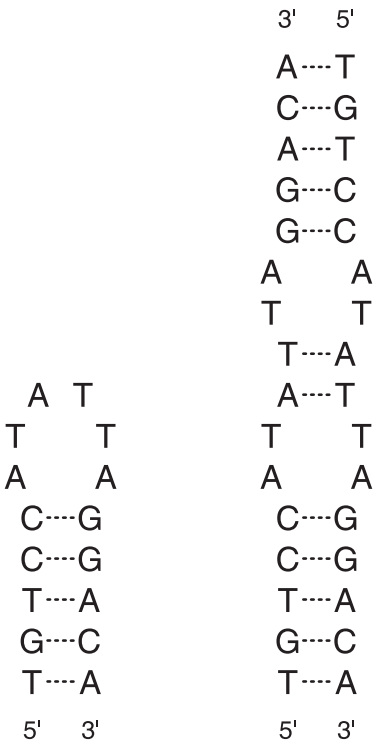


Figure 2.9 Schemes of base pairing in SRE folded into hairpin (left) and duplex (right) based on predictions by the DINAMelt web server [52]

than three loop nucleotides are only scarcely studied [21, 22, 133, 134]. Therefore, confirmation of the hypothetical hairpin fold of the SRE fragment with six-membered loop would bring relatively new type of a secondary motif.

3 CHEMICAL PROCESSES IN SOLUTION

This chapter is dedicated to the general thermodynamics beyond the processes happening in dilute solutions. We describe the equilibrium and the kinetics of an arbitrary chemical transition – a chemical reaction, a complex formation, or a conformational change. The theory works with CONCENTRATIONS equal to the MOLARITIES, the amount of a substance per volume.

3.1 Two-state equilibrium

3.1.1 First-order transition

Suppose a transition between state A and state B with the forward and backward RATE CONSTANTS k_f and k_b, respectively,

$$\mathrm{A} \underset{k_\mathrm{b}}{\overset{k_\mathrm{f}}{\rightleftharpoons}} \mathrm{B}, \tag{3.1}$$

in a thermodynamic equilibrium in which the concentrations of the two species are not evolving in time:

$$k_\mathrm{f}[\mathrm{A}] = k_\mathrm{b}[\mathrm{B}]. \tag{3.2}$$

The ratio between the concentrations of the product and the reactant is the forward EQUILIBRIUM CONSTANT:

$$K_\mathrm{f} = \frac{[\mathrm{B}]}{[\mathrm{A}]}, \tag{3.3}$$

which can be immediately connected to the rate constants according to Eq. (3.2):

$$K_\mathrm{f} = \frac{k_\mathrm{f}}{k_\mathrm{b}}. \tag{3.4}$$

Under a constant pressure, the distribution of individual populations is governed by the GIBBS FREE ENERGY, G. When the difference between

G of the two states from Eq. (3.1) at a temperature T, $\Delta G_\mathrm{f}(T)$, is expressed in the terms of DIFFERENCES IN ENTHALPY, ΔH_f, and ENTROPY, ΔS_f, as

$$\Delta G_\mathrm{f}(T) = G_\mathrm{B}(T) - G_\mathrm{A}(T) = \Delta H_\mathrm{f} - T\Delta S_\mathrm{f}, \tag{3.5}$$

the equilibrium constant follows the VAN 'T HOFF EQUATION that can be written in the form

$$K_\mathrm{f}(T) = \mathrm{e}^{-\Delta G_\mathrm{f}/RT} = \mathrm{e}^{-\Delta H_\mathrm{f}/RT + \Delta S_\mathrm{f}/R}, \tag{3.6}$$

where R is the molar gas constant.

The rate constants depend on temperature, too. The original phenomenological ARRHENIUS EQUATION,

$$k_\mathrm{f}(T) = A\mathrm{e}^{-E_\mathrm{a}/RT}, \tag{3.7}$$

introducing an ACTIVATION ENERGY, E_a, and a FREQUENCY FACTOR, A, was later revised by the transition-state theory that produced the EYRING EQUATION:

$$k_\mathrm{f}(T) = \frac{kT}{h}\,\mathrm{e}^{-\Delta H^\dagger/RT + \Delta S^\dagger/R}. \tag{3.8}$$

Here, the universal Boltzmann (k) and Planck (h) constants are used and the transition state is described by $\Delta H^\dagger$ and $\Delta S^\dagger$, the forward ACTIVATION ENTHALPY and ENTROPY, respectively, that may be combined to give the forward GIBBS FREE ENERGY OF ACTIVATION,

$$\Delta G^\dagger(T) = \Delta H^\dagger - T\Delta S^\dagger. \tag{3.9}$$

The forward direction is taken by default when dealing with the kinetics, i.e. $\Delta G^\dagger = \Delta G^\dagger_\mathrm{f}$. Nevertheless, the above equations hold also for the backward process and the thermodynamic quantities are simply interconnected, e.g. the backward activation Gibbs free energy is:

$$\Delta G^\dagger_\mathrm{b} = \Delta G^\dagger_\mathrm{f} - \Delta G. \tag{3.10}$$

3.1.2 Second-order transition

Let us now consider a slightly more complicated transition that is applicable also to the formation and dissociation of a bimolecular complex:

$$\mathrm{A} \underset{k_\mathrm{b}}{\overset{k_\mathrm{f}}{\rightleftharpoons}} \mathrm{B} + \mathrm{C}. \tag{3.11}$$

It is still first-order in the forward direction but it is second-order in the backward sense. Therefore, the equilibrium definition (3.2) changes to

$$k_\mathrm{f}[\mathrm{A}] = k_\mathrm{b}[\mathrm{B}][\mathrm{C}] \tag{3.12}$$

and the forward equilibrium constant of this process is

$$K_\mathrm{f} = \frac{[\mathrm{B}][\mathrm{C}]}{[\mathrm{A}]}\,. \tag{3.13}$$

Eq. (3.4) tying K_f with the rate constants is still valid, indeed.

The unit of K_f has altered so that it is no longer a dimensionless quantity. Therefore, the van 't Hoff equation (3.6) must be modified accordingly to

$$K_\mathrm{f}(T) = c_\mathrm{ref}\,\mathrm{e}^{-\Delta H_\mathrm{f}/RT + \Delta S_\mathrm{f}/R}, \tag{3.14}$$

where a standard reference concentration, c_ref, is introduced. As a rule, $c_\mathrm{ref} = 1$ M is selected for its simplicity even though there is no physical reason [135]. This parameter appears in the Eyring equation for the backward rate constant, too, but it is not needed for the forward direction in Eq. (3.8) that has been set as default.

Analogously to the way applied here, the thermodynamic description can be extended to any two-state transition. However, elementary chemical transitions usually stay at most second-order because processes involving higher number of interacting species commonly proceed *via* some intermediate states.

3.1.3 Melting of NA secondary structures

The general theory demonstrated above finds applications in the thermodynamic analysis of the stability of various NA motifs [136]. In this thesis, the folded structure is going to represent the state A in both Eq. (3.1) and (3.11). This state is, naturally, of the greatest interest, so the changes in quantities referring to the backward direction, identical with the folding transition, are going to be written without any subscript:

$$\Delta H = \Delta H_\mathrm{b} = -\Delta H_\mathrm{f} \tag{3.15}$$

and so on.

The nature of NA structures held together by non-covalent interactions implies that the stable folds should have $\Delta H < 0$. Additionally, the

folded state has usually a higher symmetry than the random coil, therefore the folding is accompanied with $\Delta S < 0$. These two parameters are opposing each other in the resulting ΔG (3.5) that is the cause of the instability of the NA structural motifs at higher temperatures.

Unimolecular structures, such as hairpins, are well described by the general approach for the first-order transitions. Their melting temperature, given by

$$K_\mathrm{f}(T_\mathrm{m}) = 1, \tag{3.16}$$

$$\Delta G(T_\mathrm{m}) = 0, \tag{3.17}$$

$$T_\mathrm{m} = \frac{\Delta H}{\Delta S}, \tag{3.18}$$

is independent on the total NA concentration,

$$c = [\mathrm{A}] + [\mathrm{B}]. \tag{3.19}$$

On the other hand, formation of NA **duplexes from two identical strands** can be treated as a special case of the general second-order scheme (3.11) with B = C:

$$\mathrm{A} \underset{k_\mathrm{b}}{\overset{k_\mathrm{f}}{\rightleftharpoons}} \mathrm{B} + \mathrm{B}. \tag{3.20}$$

The total concentration of the NA molecules is then

$$c = 2[\mathrm{A}] + [\mathrm{B}]. \tag{3.21}$$

At T_m, exactly half of the NA molecules are paired to duplexes, therefore

$$K_\mathrm{f}(T_\mathrm{m}) = c, \tag{3.22}$$

and the van 't Hoff equation (3.14) yields generally non-zero

$$\Delta G(T_\mathrm{m}) = RT_\mathrm{m} \ln \frac{c}{c_\mathrm{ref}}. \tag{3.23}$$

The melting temperature is then depending on c:

$$T_\mathrm{m} = \frac{\Delta H}{\Delta S + R \ln c/c_\mathrm{ref}}. \tag{3.24}$$

3.1.4 Melting curve

The melting thermodynamics is experimentally assessed by measurements of structure-dependent quantities (chemical shift, absorbance at a

given wavelength or other) at variable temperatures [33]. Very often their observed values, x, are the population-weighted averages of their values of the folded, x_A, and unfolded, x_B, molecules (which are generally temperature-dependent):

$$x(T) = p_A x_A + p_B x_B. \tag{3.25}$$

The populations of the folded and unfolded species, p_A and p_B, respectively, are relative counts of single molecules normalised to one:

$$p_A + p_B = 1. \tag{3.26}$$

This makes them substantially different from the concentrations [A] and [B] of the complexes, which depend on the stoichiometry of the process.

Determination of the thermodynamic quantities from the experimental MELTING CURVES (3.25) requires a connection between x and K_f. This depends on the molecularity of the transition under study. Substituting the concentrations in the definitions of the equilibrium constants (3.3) or (3.13) by the relative populations and extracting p_A give:

a. for the melting of a **monomolecular hairpin**,

$$p_A(T) = \frac{1}{1 + K_f}; \tag{3.27}$$

b. for the melting of a **self-complementary duplex**,

$$p_A(T) = \frac{4c + K_f - \sqrt{K_f^2 + 8cK_f}}{4c}. \tag{3.28}$$

Fitting these relations – using the appropriate form of the van 't Hoff equation, either Eq. (3.6) or (3.14), and the normalisation condition (3.26) – to the melting data described by Eq. (3.25) leads to the knowledge of ΔH and ΔS of the folding.

A strong correlation between ΔH and ΔS and their activation counterparts obtained from fitting of the melting curves impedes their precise estimation [137]. Additional systematic error arises from neglecting their temperature dependence caused by different heat capacities of A and B [138]. However, calculation of $\Delta G_{310\,K} = \Delta G(310\ K)$ and T_m is much more accurate [139].

3.2 Three-state equilibrium

3.2.1 Application to the duplex melting

Certain temperature dependences of ^{1}H NMR spectra of self-complementary DNA duplexes were found not to match the two-state model. In these cases, a more general scheme considering also an intermediate state besides the low-temperature duplex and the high-temperature state of separate single strands had to be employed.

Two possible types of the intermediate state were assumed. Either the intermediate state is a duplex of different geometry and the temperature-induced changes correspond to a duplex–duplex transition as the first step and a duplex–single strand transition as the second one:

$$\mathrm{D}_1 \rightleftharpoons \mathrm{D}_2 \rightleftharpoons \mathrm{S} + \mathrm{S}, \tag{3.29}$$

or there are two single-strand forms. In the latter case, the duplex melting is followed by a temperature-induced transition between the two single-strand forms:

$$\mathrm{D} \rightleftharpoons \mathrm{S}_1 + \mathrm{S}_1 \rightleftharpoons \mathrm{S}_2 + \mathrm{S}_2. \tag{3.30}$$

In both models, one step is concentration dependent (the melting temperature T_m should increase with increased c), while the other does not depend on it.

Regardless of the particular scheme, we now label the three species as A, B, and C. The relative number of molecules present in these states, p_A, p_B, and p_C, fulfil the normalisation condition

$$p_\mathrm{A} + p_\mathrm{B} + p_\mathrm{C} = 1. \tag{3.31}$$

The melting curve is then, in analogy with Eq. (3.25),

$$x(T) = p_\mathrm{A} x_\mathrm{A} + p_\mathrm{B} x_\mathrm{B} + p_\mathrm{C} x_\mathrm{C}. \tag{3.32}$$

3.2.2 Duplex intermediate

The first possibility, Eq. (3.29), can be described as

$$\mathrm{A} \underset{K_\mathrm{A}}{\rightleftharpoons} \mathrm{B} \underset{K_\mathrm{B}}{\rightleftharpoons} \mathrm{C} + \mathrm{C}, \tag{3.33}$$

where A and B are the two duplex states, C denotes the single strand, and K_A and K_B are the backward equilibrium constants of the two transitions.

Given that

$$p_{\mathrm{A}} = \frac{2[\mathrm{A}]}{c}, \tag{3.34}$$

$$p_{\mathrm{B}} = \frac{2[\mathrm{B}]}{c}, \tag{3.35}$$

$$p_{\mathrm{C}} = \frac{[\mathrm{C}]}{c}, \tag{3.36}$$

the equilibrium constants can be written as

$$K_{\mathrm{A}} = \frac{p_{\mathrm{A}}}{p_{\mathrm{B}}}, \tag{3.37}$$

$$K_{\mathrm{B}} = \frac{p_{\mathrm{B}}}{2cp_{\mathrm{C}}^2}. \tag{3.38}$$

We now have enough equalities to be able to find the formulae for p_{A}, p_{B}, and p_{C} in the terms of K_{A}, K_{B}, and c. First, inserting p_{A} derived from the normalisation condition (3.31) into Eq. (3.37) and substituting p_{B} by

$$p_{\mathrm{B}} = 2cK_{\mathrm{B}}p_{\mathrm{C}}^2 \tag{3.39}$$

coming from Eq. (3.38) yield a quadratic equation for p_{C}:

$$2c(K_{\mathrm{A}}+1)K_{\mathrm{B}}p_{\mathrm{C}}^2 + p_{\mathrm{C}} - 1 = 0. \tag{3.40}$$

There is only one positive solution:

$$p_{\mathrm{C}} = \frac{\sqrt{8c(K_{\mathrm{A}}+1)K_{\mathrm{B}}+1}-1}{4c(K_{\mathrm{A}}+1)K_{\mathrm{B}}}. \tag{3.41}$$

The other populations, p_{B} and p_{A}, can be subsequently calculated by Eq. (3.39) and the normalisation condition (3.31), respectively[1].

3.2.3 Single-strand intermediate

We are now about to describe a scheme equivalent to Eq. (3.30):

$$\begin{aligned} \mathrm{A} &\underset{K_{\mathrm{A}}}{\rightleftharpoons} \mathrm{B}+\mathrm{B}, \\ \mathrm{B} &\underset{K_{\mathrm{B}}}{\rightleftharpoons} \mathrm{C}. \end{aligned} \tag{3.42}$$

[1] it is not very nice when expanded

In analogy with the previous case, the equilibrium constants are

$$K_A = \frac{p_A}{2cp_B^2}, \tag{3.43}$$

$$K_B = \frac{p_B}{p_C}. \tag{3.44}$$

Substitution of p_C in Eq. (3.44) using the normalisation (3.31) followed by replacement of p_A according to Eq. (3.43) gives a quadratic equation for p_B:

$$2cK_AK_Bp_B^2 + (K_B + 1)p_B - K_B = 0. \tag{3.45}$$

Its only positive root is

$$p_B = \frac{\sqrt{(K_B + 1)^2 + 8cK_AK_B^2} - K_B - 1}{4cK_AK_B}. \tag{3.46}$$

The other populations, p_A and p_C, can then be directly calculated by Eq. (3.43) and Eq. (3.31), respectively[2].

[2] the previous footnote applies

4 NUCLEAR MAGNETIC RESONANCE

This chapter deals with nuclear magnetic resonance in isotropic liquids. Only the phenomena which are later used to interprete the experimental results are briefly explained here. The methods applied to studies on nucleic acids are also described.

4.1 Interactions of nuclear spins

4.1.1 Zeeman effect

The atomic nuclei of many isotopes have a non-zero spin in their ground state. A nucleus with a spin quantum number I possesses a dipolar magnetic moment $\boldsymbol{\mu}$, whose projection to an axis z is quantised to the values

$$\mu_z = m\hbar\gamma, \tag{4.1}$$

where the quantum number

$$m \in \{-I,\ -I+1,\ \ldots,\ I-1,\ I\}, \tag{4.2}$$

$\hbar$ is the reduced Planck constant and γ is the isotope-specific GYROMAGNETIC RATIO.

In a static magnetic field $\boldsymbol{B_0}$ collinear with z ($|\boldsymbol{B_0}| = B_0 = B_z$), the stationary magnetic energy levels of a nucleus,

$$E_m = -\boldsymbol{B_0} \cdot \boldsymbol{\mu} = -B_0\mu_z = -m\hbar\gamma B_0 \tag{4.3}$$

belonging to particular values of m, are not equal, which is known as ZEEMAN EFFECT. Therefore, the individual levels are differently populated in a sample containing N nuclei in a thermodynamic equilibrium according to the Boltzmann distribution, leading to a macroscopic NUCLEAR MAGNETISATION

$$\boldsymbol{M} = \frac{NI(I+1)\hbar^2\gamma^2}{3kT}\,\boldsymbol{B_0}, \tag{4.4}$$

given that

$$\hbar\gamma B_0 \ll kT, \tag{4.5}$$

satisfied in the achievable laboratory conditions [140].

We shall now turn our attention to the time evolution of $\boldsymbol{M}$ and the properties of the resulting NMR spectra in isotropic liquids.

4.1.2 Bloch equations

The interaction with magnetic field $\boldsymbol{B}$ produces a torque experienced by the nuclear magnetisation. Introducing the LONGITUDINAL RELAXATION TIME T_1 and the TRANSVERSE RELAXATION TIME T_2, the classical phenomenological equations of motion of $\boldsymbol{M}$, the BLOCH EQUATIONS, are

$$\frac{\mathrm{d}M_x}{\mathrm{d}t} = \gamma(\boldsymbol{M}\times\boldsymbol{B})_x - \frac{M_x}{T_2}, \tag{4.6}$$

$$\frac{\mathrm{d}M_y}{\mathrm{d}t} = \gamma(\boldsymbol{M}\times\boldsymbol{B})_y - \frac{M_y}{T_2}, \tag{4.7}$$

$$\frac{\mathrm{d}M_z}{\mathrm{d}t} = \gamma(\boldsymbol{M}\times\boldsymbol{B})_z - \frac{M_z - M_0}{T_1}, \tag{4.8}$$

where M_0 is the magnitude of the equilibrium nuclear magnetisation.

4.1.3 Free precession

The Bloch equations imply that when only the z-oriented static field $\boldsymbol{B_0}$ is present, the complex quantity

$$M_\perp = M_x + \mathrm{i}M_y \tag{4.9}$$

describing the component of $\boldsymbol{M}$ perpendicular to $\boldsymbol{z}$ follows

$$M_\perp(t) = M_0\mathrm{e}^{-\mathrm{i}\gamma B_0 t - t/T_2}, \tag{4.10}$$

assuming that $M_\perp(0) = M_0$. Eq. (4.10) shows that a precession of $\boldsymbol{M}$ around the direction of $\boldsymbol{B_0}$ occurs with angular velocity

$$\boldsymbol{\omega_0} = -\gamma\boldsymbol{B_0}, \tag{4.11}$$

called LARMOR FREQUENCY, and it is recorded as FREE INDUCTION DECAY (FID). Selected isotopes with their γ are listed in Table 4.1, showing the amplitudes of their RESONANT FREQUENCIES

$$\nu_0 = \frac{\omega_0}{2\pi}. \tag{4.12}$$

Table 4.1 Spin numbers, gyromagnetic ratios, resonant frequencies at 11.7 T, and natural abundances, β, of nuclei with non-zero spin used in this thesis [141]

	I	$\frac{\gamma \cdot 10^{-6}}{\mathrm{T^{-1}s^{-1}}}$	$\frac{\nu_0}{\mathrm{MHz}}$	$\frac{\beta}{\%}$
^{1}H	$\frac{1}{2}$	267.5	500.0	99.985
^{2}H	1	41.1	76.8	0.015
^{13}C	$\frac{1}{2}$	67.3	125.8	1.108
^{31}P	$\frac{1}{2}$	108.4	202.6	100.0

Fourier transform of Eq. (4.10) yields the spectral line that is proportional to the LORENTZIAN CURVE:

$$\mathcal{L}(\omega) = \frac{T_2}{1 + T_2^2(\omega - \omega_0)^2} + \mathrm{i}\,\frac{T_2^2(\omega - \omega_0)}{1 + T_2^2(\omega - \omega_0)^2}\,. \tag{4.13}$$

The real part of Eq. (4.13) represents the ABSORPTION and the imaginary part the DISPERSION curve. The FULL WIDTH AT HALF MAXIMUM (FWHM) of the absorption line[1] is twice the inverse of T_2.

4.1.4 Chemical shift

The externally applied field $\boldsymbol{B_0}$ is modified by the electrons in the matter. The measurable time-average magnitude of the magnetic field at the nucleus location in isotropic liquids can be described as

$$B_{\mathrm{loc}} = (1 - \sigma)B_0, \tag{4.14}$$

where σ stands for CHEMICAL SHIELDING. The Larmor frequency changes accordingly and it is usually presented in the terms of CHEMICAL SHIFT,

$$\delta = \frac{\omega - \omega_0}{\omega_0}\,, \tag{4.15}$$

relative to the frequency ω_0 of a CHEMICAL SHIFT STANDARD; δ is commonly expressed in the units of parts per million, ppm.

For historical reasons, spectra are plotted with reverse direction of the chemical-shift axis, i.e. δ increases from right to left. DOWNFIELD refers to lower σ and higher δ and conversely for UPFIELD.

[1] in the units of angular velocity

4.1.5 Indirect dipole–dipole interaction

The INDIRECT DIPOLE–DIPOLE INTERACTION between chemically non-equivalent nuclei connected by one or more chemical bonds to each other results in splitting of the spectral lines of the nuclei. In the case of an isolated pair of nuclei with $I = \frac{1}{2}$, the resonance frequencies in the angular-velocity scale are changed by

$$\Delta\omega = \pm\pi J. \tag{4.16}$$

Here, J is called the SCALAR COUPLING CONSTANT. Eq. (4.16) is an approximation for nuclei with large mutual difference of chemical shifts – the WEAK COUPLING limit:

$$\Delta\delta \gg \frac{J}{\omega_0}\,. \tag{4.17}$$

Apart from the peak splitting, the phases of the components of the multiplets are modified when a pulse sequence containing a SPIN ECHO is used for the acquisition. Due to the anti-phase term arising from the coupling evolution [142, 143], the phase differences of the two peaks in any doublet are[2]

$$\Delta\varphi = \pm\pi J\tau_{\text{echo}} = \Delta\omega\,\tau_{\text{echo}}. \tag{4.18}$$

This leads to mixing of the absorption and dispersion components of the Lorentzian curve (4.13).

The scalar coupling is useful in discovering connections between atoms by chemical bonds. Particularly, TWO-DIMENSIONAL (2D) NMR spectra exploiting the coherent MAGNETISATION TRANSFER are commonly employed: homonuclear CORRELATION SPECTROSCOPY (COSY) and TOTAL CORRELATION SPECTROSCOPY (TOCSY) or various heteronuclear experiments, such as HETERONUCLEAR CORRELATION (HETCOR), HETERONUCLEAR SINGLE-QUANTUM COHERENCE (HSQC), and HETERONUCLEAR MULTIPLE-BOND COHERENCE (HMBC) suppressing the strongest one-bond couplings represent the basic techniques that can be extensively modified [144, 145, 146].

[2] relative to the signal of nuclei without the indirect interaction

4.1.6 Direct dipole–dipole interaction

Although the DIRECT DIPOLE–DIPOLE INTERACTION between nuclear magnetic moments does not manifest itself in the line shapes of NMR spectra of isotropic liquids, it has a strong effect on nuclear relaxation. Importantly, the interaction also leads to a CROSS-RELAXATION that causes a non-coherent magnetisation transfer due to NUCLEAR OVERHAUSER EFFECT (NOE) [147].

The cross-relaxation rate depends on the isotopes involved, the kinetics of molecular tumbling, and the distance between the nuclei. The latter is used to determine (both qualitatively and quantitatively) the inter-atomic distances in molecules. Common experiments devoted to this application, the two-dimensional NUCLEAR OVERHAUSER EFFECT SPECTROSCOPY[3] (NOESY) and ROTATING-FRAME OVERHAUSER EFFECT SPECTROSCOPY (ROESY) [146], show off-diagonal cross peaks for pairs of nuclei close in space to each other. The cross-peak intensity (volume) in NOESY decreases with the sixth power of the inter-nuclear distance and is hardly measurable for distances longer than 5 Å.

4.2 Chemical exchange

Whenever a nucleus changes its chemical environment, we speak about CHEMICAL EXCHANGE. No matter what is the physical–chemical mechanism and its molecularity, the net effect on the nuclei is always first-order – the total number of particles is preserved[4].

4.2.1 Two-site chemical exchange

A two-site chemical exchange can be depicted as

$$\mathrm{A} \underset{k_\mathrm{B}}{\overset{k_\mathrm{A}}{\rightleftharpoons}} \mathrm{B}. \tag{4.19}$$

The EXCHANGE RATES k_A and k_B are generally not the same as the rate constants k_f and k_b defined in Chapter 3, because their mutual relations depend on the particular type of the process. Selecting two important cases, these are:

[3] alternatively, NUCLEAR OVERHAUSER ENHANCEMENT SPECTROSCOPY

[4] because the stable isotopes are not easy to split or to merge

a. for the **first-order transition** (3.1) on page 25:

$$k_\mathrm{A} = k_\mathrm{f}, \tag{4.20}$$
$$k_\mathrm{B} = k_\mathrm{b}; \tag{4.21}$$

b. for the **self-association second-order transition** (3.20) on page 28:

$$k_\mathrm{A} = k_\mathrm{f}, \tag{4.22}$$
$$k_\mathrm{B} = 2k_\mathrm{b}[\mathrm{B}]. \tag{4.23}$$

We now label the relative counts of the nuclei in state A and B as p_A and p_B, respectively. The symbols are selected intentionally the same as in § 3.1.4 on page 28 in order to stress that they are equal, indeed. Since the normalisation (3.26) stays valid and provided that the system is in a detailed equilibrium, fulfilling

$$p_\mathrm{A} k_\mathrm{A} = p_\mathrm{B} k_\mathrm{B}, \tag{4.24}$$

only two independent parameters characterise the two-site exchange.

4.2.2 Bloch–McConnell equations

In order to describe the time evolution of nuclear magnetisations of site A, $\boldsymbol{M}^\mathrm{A}$, and site B, $\boldsymbol{M}^\mathrm{B}$, the Bloch equations (4.6)–(4.8) are extended in a straightforward fashion to the set of Bloch–McConnell equations containing the two-site exchange [148, 149, 150]:

$$\frac{\mathrm{d}M_x^\mathrm{A}}{\mathrm{d}t} = \gamma(\boldsymbol{M}^\mathrm{A} \times \boldsymbol{B}_\mathrm{loc}^\mathrm{A})_x - \frac{M_x^\mathrm{A}}{T_2^\mathrm{A}} - k_\mathrm{A} M_x^\mathrm{A} + k_\mathrm{B} M_x^\mathrm{B}, \tag{4.25}$$
$$\frac{\mathrm{d}M_y^\mathrm{A}}{\mathrm{d}t} = \gamma(\boldsymbol{M}^\mathrm{A} \times \boldsymbol{B}_\mathrm{loc}^\mathrm{A})_y - \frac{M_y^\mathrm{A}}{T_2^\mathrm{A}} - k_\mathrm{A} M_y^\mathrm{A} + k_\mathrm{B} M_y^\mathrm{B}, \tag{4.26}$$
$$\frac{\mathrm{d}M_z^\mathrm{A}}{\mathrm{d}t} = \gamma(\boldsymbol{M}^\mathrm{A} \times \boldsymbol{B}_\mathrm{loc}^\mathrm{A})_z - \frac{M_z^\mathrm{A} - M_0^\mathrm{A}}{T_1^\mathrm{A}} - k_\mathrm{A} M_z^\mathrm{A} + k_\mathrm{B} M_z^\mathrm{B}, \tag{4.27}$$
$$\frac{\mathrm{d}M_x^\mathrm{B}}{\mathrm{d}t} = \gamma(\boldsymbol{M}^\mathrm{B} \times \boldsymbol{B}_\mathrm{loc}^\mathrm{B})_x - \frac{M_x^\mathrm{B}}{T_2^\mathrm{B}} + k_\mathrm{A} M_x^\mathrm{A} - k_\mathrm{B} M_x^\mathrm{B}, \tag{4.28}$$
$$\frac{\mathrm{d}M_y^\mathrm{B}}{\mathrm{d}t} = \gamma(\boldsymbol{M}^\mathrm{B} \times \boldsymbol{B}_\mathrm{loc}^\mathrm{B})_y - \frac{M_y^\mathrm{B}}{T_2^\mathrm{B}} + k_\mathrm{A} M_y^\mathrm{A} - k_\mathrm{B} M_y^\mathrm{B}, \tag{4.29}$$
$$\frac{\mathrm{d}M_z^\mathrm{B}}{\mathrm{d}t} = \gamma(\boldsymbol{M}^\mathrm{B} \times \boldsymbol{B}_\mathrm{loc}^\mathrm{B})_z - \frac{M_z^\mathrm{B} - M_0^\mathrm{B}}{T_1^\mathrm{B}} + k_\mathrm{A} M_z^\mathrm{A} - k_\mathrm{B} M_z^\mathrm{B}, \tag{4.30}$$

where the relaxation times of the nuclei at the two sites are labelled by superscripts as well as the local magnetic fields, already respecting the two chemical shifts, δ_A and δ_B, analogously to Eq. (4.15).

Under free precession, the time-derivatives of the transverse magnetisations in the complex notation as in Eq. (4.9) become

$$\frac{\mathrm{d}M_\perp^\mathrm{A}}{\mathrm{d}t} = \left(-\mathrm{i}\omega_\mathrm{A} - \frac{1}{T_2^\mathrm{A}} - k_\mathrm{A}\right) M_\perp^\mathrm{A} + k_\mathrm{B} M_\perp^\mathrm{B}, \tag{4.31}$$

$$\frac{\mathrm{d}M_\perp^\mathrm{B}}{\mathrm{d}t} = \left(-\mathrm{i}\omega_\mathrm{B} - \frac{1}{T_2^\mathrm{B}} - k_\mathrm{B}\right) M_\perp^\mathrm{B} + k_\mathrm{A} M_\perp^\mathrm{A}, \tag{4.32}$$

where ω_A and ω_B are the Larmor frequencies of the two sites. Treatment of this set of two linear ordinary differential equations describing the general case of two-site chemical exchange basically consists of computing a matrix exponential, Fourier transform to the frequency domain, and a tedious algebraical editing. The available literature [151, 152, 150, 153] only approaches the complete solution. The fully analytical formula was derived in my diploma thesis [154].

The exchange spectral line is proportional to

$$\mathcal{X}(\omega) = \frac{k_\mathrm{A} + k_\mathrm{B} + p_\mathrm{A}\alpha_\mathrm{B} + p_\mathrm{B}\alpha_\mathrm{A}}{\alpha_\mathrm{A}\alpha_\mathrm{B} + k_\mathrm{A}\alpha_\mathrm{B} + k_\mathrm{B}\alpha_\mathrm{A}}, \tag{4.33}$$

whilst the characteristics of the two sites are contained in

$$\alpha_\mathrm{A} = \frac{1}{T_2^\mathrm{A}} + \mathrm{i}(\omega - \omega_\mathrm{A}), \tag{4.34}$$

$$\alpha_\mathrm{B} = \frac{1}{T_2^\mathrm{B}} + \mathrm{i}(\omega - \omega_\mathrm{B}). \tag{4.35}$$

Depending on the $p_\mathrm{A}/p_\mathrm{B}$ ratio, the intrinsic line widths, the exchange rates, and the difference between ω_A and ω_B, three distinct regimes may be observed in the spectrum:

a. SLOW EXCHANGE with two separated peaks;
b. INTERMEDIATE EXCHANGE with the two signals coalesced into one broad peak;
c. FAST EXCHANGE with only one moderately broadened peak.

Whereas the line shapes during the slow and fast exchange can be approximated by Lorentzian curves, the spectrum is severely influenced in

the intermediate case and the signal can even be virtually vanishing in the spectral noise and background.

4.2.3 Thermodynamic analysis of variable-temperature spectra

The theoretical relations described above can be used to estimate the thermodynamic parameters of chemically exchanging systems. Acquisition of spectra at various temperatures and analysis of their line shapes according to (4.33) yield the relative populations p_A and p_B as well as the exchange rates, k_A and k_B.

The temperature dependence of p_A can be treated as a melting curve provided in Eq. (3.25), letting $x_A = 1$ and $x_B = 0$, therefore directly fitting its appropriate relation to the equilibrium constant, such as Eq. (3.27) or (3.28).

The exchange rates k_A are first converted to the chemical rate constants according to the trivial Eq. (4.20) and (4.22). Next, the Arrhenius equation (3.7) or Eyring equation (3.8) can be fitted to their temperature dependence.

4.3 NMR of nucleic acids

4.3.1 ^{1}H resonances in nucleic acids and their assignment

Assignment of the observed NMR resonances to particular hydrogen nuclei of individual nucleotides is essential for studies of local properties of NA structures. The discrimination between chemically distinct groups is performed based on known ranges of chemical shifts: aromatic hydrogens (adenine H2, pyrimidine H6, and purine H8), cytosine H5, thymine methyl groups $(H7)_3$, amino and imino protons of the bases (the so-called EXCHANGEABLE PROTONS), and different sets of locations at the deoxyribose ring are well separated from each other [155, 156]. In the aromatic part of the spectrum, the doublets of cytosine H6 are readily identified; some other rules can be more or less reliably applied to discern between other base-proton types as well [157].

Apart from straightforward but relatively expensive treatments such as selective isotopical labelling [158, 159], the position in the ODN sequence is most commonly revealed from two-dimensional ^{1}H NOESY.

Intra-nucleotide together with inter-nucleotide distances between hydrogen atoms in a B-form double-helix [160, 161, 162] allow a SEQUENTIAL WALK: the most useful cross-peaks appear between H6/H8 protons and H1′, H2′, and H2″ protons of the same nucleoside and of its neighbour in the direction towards the 5′ end. Once the pyrimidine H6 resonances are identified, their strong cross-peaks with the CH5 or TH7 of the same base are easily found. In DNA, the H1′ protons are closer to H2″ protons than to H2′, thus the corresponding cross-peaks have different intensities that makes the diastereotopic protons identifiable – especially at short mixing times.

Provided that non-deuterated water is used as the solvent, exchangeable protons are assigned from NOESY, too. A sequential walk is possible along the amino and imino groups forming the hydrogen bonds between the base pairs or lying close to them. Resonances from both DNA strands of the double-helix play part of this scheme. AH2 and CH5 are also connected to this spin system which leads to completion of the interpretation of the aromatic lines. [157, 109]

Accidental overlaps can often be overcome by analysis of another spectral region. The symmetric portions across the diagonal of the NOESY spectrum shall not be forgotten. More information is obtained from various other parts, such as those including thymine methyl groups or inter-aromatic cross-peaks close to the diagonal. In this way, all the exchangeable and non-exchangeable protons visible in the ^{1}H spectrum can be, in principle, assigned. However, the deoxyribose protons H3′, H4′, H5′, and H5″ are difficult to assign in $^{1}H_2O$ because their resonances are generally very close to that of water protons and have low dispersion of chemical shifts. [163]

Even after a careful acquisition, processing and analysis of ^{1}H NOESY, some resonances may remain unassigned or assigned ambiguously. Supplementary experiments are then to be performed. Variation of conditions (temperature, concentration) may help. Pulse sequences tailored for particular goals are also available of which we select a handful:

a. ^{1}H COSY and TOCSY (preferably in D_2O) useful for the deoxyribose protons [164];
b. various ^{1}H–^{31}P correlation experiments [165, 166, 167] (preferably

in D_2O), optionally combined with ^1H TOCSY or NOESY [168, 169], for assigning ^{31}P resonances together with the deoxyribose protons, utilising heteronuclear coupling constants $^3J_{H3'-P}$ and $^4J_{H4'-P}$ of few hertz [170];

c. ^1H–^{13}C HMBC[5] for assigning AH2 protons by connecting them to H8 signals of the same base *via* the couplings to C4 [171] with the magnitudes around $^3J_{H2-C4} = 12$ Hz and $^3J_{H8-C4} = 5$ Hz [161, 164];
d. similarly, the ^1H–^{13}C HMBC can be utilised even for connecting GH8 and TH6 to the imino protons GH1 and TH3, respectively [172].

The above-mentioned assignment techniques have been mainly designed for nucleic acids folded in double helices or other structural motifs which are generally of the greatest interest. Although the literature on the NMR properties of unfolded NA chains is very scarce [173, 174, 175], the experiments described can be used for assigning random-coil resonances as well. Naturally, some limitations must be considered: e.g., the observation of NOE contacts is problematic at high temperatures due to changed inter-atomic distances, enhanced flexibility and different (cross-)relaxation rates. Generally, the indirect dipole–dipole interaction serves better than the direct through-space interaction in the random-coil DNA.

4.3.2 Prediction of chemical shifts by the nearest-neighbour model

Although chemical shifts are well known to reproducibly report on the secondary structural motifs in proteins [176], nucleic acids with their small number of residue types and their inherent dependence of shifts on adjacent bases due to the ring-current effect [177] bring a considerable challenge. Efforts have been spent for developing methods to assign the NMR resonances without the need of NOESY and other experiments [178, 179]. Ideally, the chemical shifts would be calculated based on the nucleotide sequence and folding motifs only. This would in turn facilitate structure determination based on chemical shifts [180].

The *ab initio* quantum-chemical calculations [181] are still too complicated to be routinely applied for macromolecules, although they are already emerging [182, 183]. Instead, the common estimates of the NMR

[5] the carbon nuclei ^{13}C can be utilised even at natural isotopic composition of ODN in the inverse-detected experiment

chemical shifts of ^{1}H [177, 163, 173, 184], ^{31}P [175], or ^{13}C [174] in DNA rely on statistical analyses of large sets of experimental data. These methods are gathered and the calculations are automated in DSHIFT web server [184] freely accessible over the Internet. The database of ^{1}H chemical shifts of the more structurally variable RNA molecules can be analysed with respect to their secondary structure [185, 186].

The simplest approximation which takes the surrounding nucleotides into account is the nearest-neighbour model widely applied in the literature to various properties of individual nucleotides or Watson–Crick pairs. Some empirical prediction protocols supplement the NN approach by corrections for oligonucleotide ends or next-nearest neighbours. The double-helical shifts of penultimate positions are modified by adding a value derived for each type of proton; shielding induced by pyrimidine as a next-nearest neighbour relative to purine are pointed out but not incorporated in the prediction protocol [163]. Additive constants reflecting different next-nearest-neighbouring bases but ignoring the type of the base under question are introduced in the random-coil prediction [173, 184].

The effect of the third nearest neighbour was analysed and, in single strands, it was found to be always lower than 0.02 ppm and 0.007 ppm in average [173]. The authors ignore it in their predictions, but it should be considered as a possible source of errors in the order of 0.01 ppm. Temperature dependence of shifts is also generally neglected both in duplex and single strands with the exception of imino protons of G and T, having the slopes $(-2.2 \pm 0.6) \cdot 10^{-3}$ ppm/K and $(-5.4 \pm 0.5) \cdot 10^{-3}$ ppm/K, respectively [163].

Corrections for double helices containing mismatch base pairs are available, too [187, 188, 189]. We have found no predictions for ODN containing modified backbone or minor bases, such as 5-methylcytosine, in the literature yet.

5 EXPERIMENTAL DETAILS

5.1 Sample preparation

The oligodeoxynucleotides were purchased as dry powders purified by ion-exchange high-performance liquid chromatography and repeated lyophilisation from the Department of Experimental Biology, Faculty of Science of Masaryk University, Brno (the laboratory was later made part of the Core Facility Proteomics of CEITEC – Central European Institute of Technology), or from ATDBio Ltd, Southampton, United Kingdom. The rGMP and dGMP mononucleotides and their salts were purchased from Sigma–Aldrich spol. s.r.o., Prague, or Carbosynth Limited, Compton (Berkshire), United Kingdom.

The ODN samples were dissolved in sodium phosphate buffer in H_2O prepared from Na_2HPO_4 and NaH_2PO_4 and NaCl in appropriate amounts to give pH 7.0 and 25 mM phosphate and total 200 mM Na^+ concentrations. The solvent used for NMR spectroscopy contained 10 % volume fraction of D_2O for lock purposes and sodium 4,4-dimethyl-4-silapentane-1-sulfonate (DSS) as an internal chemical-shift standard in a concentration 10- to 100-fold lower than that of the ODN.

5.2 NMR spectrometer

The NMR experiments on octanucleotide DNA solutions were performed on Bruker Avance 500 spectrometer connected to a Bruker UltraShield 11.7 T superconducting magnet with the proton frequency 500.13 MHz. The experiments were carried on a 5 mm triple-channel probehead (^{1}H, ^{31}P, and broad band; the outer coil connected in the ^{1}H resonance circuit) with deuterium lock and single-axis gradient in the direction of the static magnetic field $\boldsymbol{B_0}$. The ^{31}P spectra were acquired utilising the broad-band channel due to its higher sensitivity than the ^{31}P circuit devoted primarily

to decoupling. Alternatively, a dual-channel probehead (^{1}H and broad band) with analogous setup was used.

In the course of the work on this thesis, the spectrometer console was replaced by a Bruker Avance III HD machine, which was employed for the measurements of the SRE-based DNA and of the GMP samples and for some subsidiary experiments, such as ^{31}P spectra for the determination of ODN concentration and ^{1}H–^{13}C HMBC. Differently from the former spectrometer, automatic shimming of the magnetic-field homogeneity was available. A dual-channel probehead similar to the one described above but supplied with automatic tuning and matching system was used.

A flow of dry nitrogen gas, pre-cooled by an external unit and warmed by a heater inside the probehead, served for achieving the target temperature in both spectrometers.

Volumes between 500 µl and 750 µl of the solutions were placed into 5 mm Norell Ultra Precision NMR tubes. For the variable-concentration series of CTTCGAAG, only 250 µl were inserted into a 5 mm Shigemi microtube, magnetic-susceptibility matched to D_2O, in order to reach higher ODN concentration.

5.3 Variable-temperature ^{1}H NMR experiments

The melting of DNA secondary structures was monitored by VARIABLE-TEMPERATURE (VT) ONE-DIMENSIONAL (1D) ^{1}H spectra acquired with water-signal suppression by excitation sculpting using pulsed field gradients (Fig. 5.1) [190]. The water-selective pulses were 4 ms long. The total relaxation period composed of the acquisition time and the recovery delay, D_1, was at least 2.3 s. Number of scans varied between 64 and 1024 depending on the ODN concentration and signal-to-noise requirements. Besides the ^{1}H spectra, proton-decoupled ^{31}P spectra were also acquired, but these are not analysed in this thesis.

Temperature was decreased[1] from a point above 350 K by steps of 2 K. The lowest temperature was limited by the freezing point. The acquisition started not sooner than 15 min after the temperature had been set – this

[1] tests with increasing temperature have shown full reproducibility of the spectra

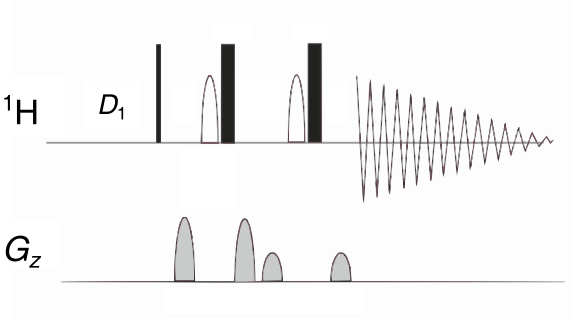


Figure 5.1 One-dimensional pulse sequence with suppression of solvent signal, `zgesgp`. ^{1}H radiofrequency and the field gradient G_z channels are shown. Thin and thick black bars represent hard 90° and 180° pulses, respectively, and the rounded pulses mean shaped selective 180° pulses in the ^{1}H channel. Adapted from [191]

waiting time was previously proved to provide a fully equilibrated homogeneous temperature in the sample volume. The precise temperature of the sample (with the error about 0.5 K) was determined from calibration experiments with methanol and ethylene glycol [192]. Field homogeneity and probehead tuning were being manually adjusted when necessary, usually after a change of about 10 K.

Fourier transforms with no apodisation were followed by zero- and first-order phase corrections and linear-baseline subtractions by the Bruker software Topspin. Chemical shifts were referenced to the methyl resonance of the internal DSS standard ($\delta_{\mathrm{DSS}} = 0$ ppm).

5.4 Assignment strategies

With the aim of assigning the ^{1}H resonances in the **folded state**, two-dimensional ^{1}H NOESY spectra were acquired by a pulse sequence employing water suppression by excitation sculpting [190] shown in Fig. 5.2. A temperature between 280 K and 286 K was selected to minimise overlaps of spectral lines. Two mixing times D_8, 100 ms and 250 ms, were used for each ODN sample. Sweep width was set to 20 ppm in both dimensions. 256 or 512 FIDs (incrementing D_0) of 4096 data points were

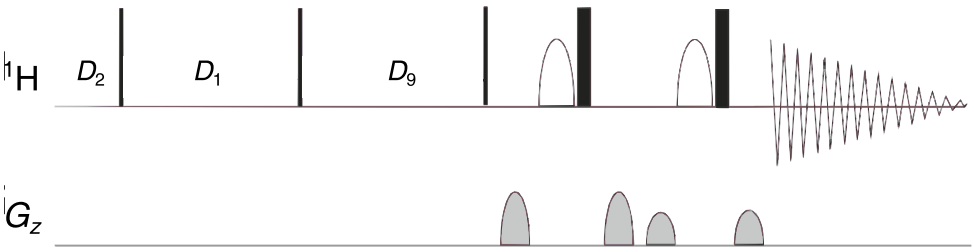


Figure 5.2 NOESY pulse sequence with suppression of solvent signal, `noesyesgpph`. Objects in the scheme have the same meaning as in Fig. 5.1. Adapted from [191]

collected in the States–TPPI[2] mode. The number of scans in each FID varied between 32, 64, and 128. The recovery delay D_1 was at least 1 s. Zero-filling, apodisations by squared cosines, Fourier transforms, first-order phase corrections and 4- or 5-order baseline corrections in both dimensions were performed by NMRPipe [193]. The cross-peaks were assigned in Sparky [194].

Having the ^{1}H spectrum in the folded state fully assigned, the temperature series of one-dimensional spectra was used to track the resonances to the **unfolded state**. Fitting of the spectral lines has turned out to be very helpful in this task: the temperature dependence of chemical shifts can be followed much more reliably from their values than from visual comparisons of raw spectra.

When the tracking did not yield trustful assignment in single strands, **auxiliary experiments** were employed. The ambiguities were dealt with by acquiring temperature-dependent spectra of more concentrated solutions not only in order to gain better signal-to-noise ratio, but mainly to speed up the chemical exchange by elevating the melting temperature[3] which results in higher rate constants at T_m. When insufficient, measurements on a 300 MHz spectrometer at the Institute of Macromolecular Chemistry of the Academy of Sciences were even conducted in order to

[2] a TIME-PROPORTIONAL PHASE INCREMENTATION method that leads to phase-sensitive complex spectrum in the indirect dimension

[3] valid for duplexes but not for unimolecular complexes

further increase the ratio between the exchange rate and the difference between the resonant frequencies of the exchanging sites.

This approach can still fail in some cases, depending on the differences of chemical shifts between folded and unfolded state and spectral overlaps. NMR experiments based on magnetisation transfer by indirect dipole–dipole interactions within the nucleobases were then used for they are independent on the spatial structure. The ambiguously assigned thymine methyl protons, TH7, were identified using their interactions with TH6; similarly for the m^5C base. One-dimensional selective COSY has been found very reliable, in spite of the very weak four-bond coupling between TH6 and TH7 of about 1 Hz and the phase-distorted peaks in the resulting multiplets. The homonuclear decoupling was employed, too, but perfect shimming must be achieved in order to observe the collapse of the splitting in the spectra of TH6 and TH7. In all these cases mentioned, presaturation of water signal was applied.

Since there are no means to correlate the highly isolated AH2 protons to any other hydrogen nucleus by any homonuclear NMR experiment in a single-stranded DNA molecule, a two-dimensional ^1H–^{13}C HMBC was measured when needed [171]. We applied the pulse sequence in Fig. 5.3 which employs two-fold low-pass J-filter to suppress one-bond ^1H–^{13}C

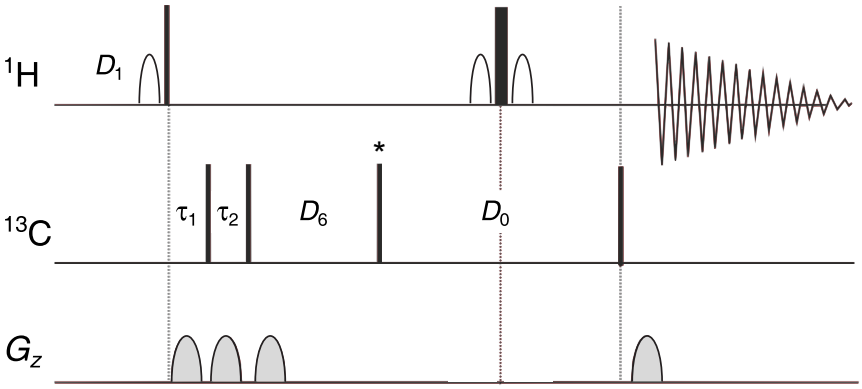


Figure 5.3 HMBC pulse sequence with WATERGATE, `hmbcgpl2ndwg`. Objects in the scheme have the same meaning as in Fig. 5.1 except that the ^1H selective pulses make 90° flip angle. Adapted from [191]

correlations and WATERGATE[4] suppression of water signal [195, 196]. A suitable temperature was selected based on 1D spectra. 96 FIDs with incrementing D_0 in the States–TPPI mode after 600 scans with recovery delay $D_1 = 1.0$ s were acquired into 2048 real data points. The 60 ppm wide ^{13}C spectral window was centred at 135 ppm. We set the delay D_6 for magnetisation transfer to 50 ms and the delays τ_1 and τ_2 for suppression of the one-bond coherences to 4.2 ms and 3.1 ms, respectively.

In order to improve the resolution in the indirect dimension and to avoid spectral folding at the same time, the pulse on ^{13}C which converts the single-quantum coherence into a double-quantum one (labelled by an asterisk in Fig. 5.3) was made selective. We refer to this experiment as a SELECTIVE HMBC. The particular setting used for different samples evolved in time; we achieved the best results with a 4-ms sinc-shaped pulse and the ^{13}C dimension centred at 149.3 ppm with 3.0 ppm width which contains the C4 resonances only. 32 FIDs were acquired with variable number of scans in order to obtain the best signal-to-noise ratio in the time available (several days).

The processing of both types of the ^{1}H–^{13}C HMBC, performed in NMRPipe [193], included zero-filling to 4096 and 512 data points in the direct and indirect dimensions, respectively, followed by squared-cosine apodisations and Fourier transforms of both dimensions. First-order phase correction was applied in the carbon dimension, while magnitude processing was necessary in hydrogens. The baselines were corrected using a fifth-order polynomial in both dimensions. Sparky software [194] was used to visualise the spectra and assign the cross-peaks.

5.5 Ultraviolet absorption

Absorption spectra in the wavelength range from 230 nm to 340 nm were acquired by a Lambda 12 (PerkinElmer) double-beam spectrophotometer, equipped with a deuterium lamp. Buffer with no ODN dissolved served as the reference. The sample chamber was continuously flushed with dry air to avoid water condensation on the outer surface of the cuvettes.

1200 µl of the solutions were placed in standard 10 mm cuvettes.

[4] water suppression by gradient-tailored excitation

A 200-µl drop of mineral oil was placed on the solution level to prevent evaporation at high temperatures. The spectra were measured at temperatures increasing from 273 K to 347 K in 3 K steps. After reaching the target temperature, we waited 10 min to equilibrate the temperature before the measurement of the spectrum.

Spectral background was corrected by subtracting the absorbance at 330 nm where no absorbance of DNA is expected.

5.6 Raman scattering

The 532.2 nm excitation Nd:YAG laser with frequency doubling (Verdi V2, Coherent) produced a power of approximately 0.5 W at the sample. Scattered light collected in a right-angle geometry was analysed by a Spex 270 M spectrograph (Jobin Yvon, single grating with 1800 grooves per millimetre) with a CCD detector cooled by liquid nitrogen. An edge filter in front of the spectrograph was used to suppress the elastic scattering. Raman spectra were recorded in the region of Stokes shifts between 500 cm^{-1} and 1824 cm^{-1} with the resolution of 1 cm^{-1}. Total acquisition time for every spectrum was 1000 s.

Cylindrical quartz microcuvette with 12 µl of the solution was placed into a thermo-stabilised chamber. Raman spectra were measured at temperatures from 275 K to 353 K in 2 K steps, a 10-min delay was used after reaching the target temperature for its equilibration. After each measurement, spectrum of a neon lamp was recorded for precise spectral calibration.

The background of Raman spectra was corrected by subtracting optimal fifth-degree polynomial function and properly scaled Raman spectra of the buffer, pure water, and the quartz cuvette wall.

5.7 Determining the oligodeoxynucleotide concentrations

5.7.1 Phosphorus quantitative NMR

For the thermodynamic analysis of the stability of DNA structures, a reliably accurate method for the determination of the ODN concentration in solution, c, is needed. The common approach based on UV absorbance

measured at 260 nm uses extinction coefficients estimated by some approximation. Considering the base composition only is the simplest one; alternatively, a nearest-neighbour model can be applied [197]. During the course of our NMR studies, these methods, suffering also from imperfections during the sample manipulation, were found to provide insufficiently precise results.

For this reason, ^{31}P QUANTITATIVE NMR SPECTROSCOPY (qNMR) was employed. Fully relaxed spectra were acquired by a single-pulse sequence with a 50-s recovery time and no decoupling. A weak exponential window function (line broadening 0.5 Hz) was applied prior to the Fourier transform. Automatic second-order polynomial baseline subtraction was performed. Extra care was taken to the manual first-order phase correction. The molar ratio between the oligonucleotide and the inorganic phosphate of the buffer, which thus served as an internal concentration reference, was determined with high precision from the integral intensities of the ^{31}P spectral lines. All this processing and analysis was done in Topspin. We estimate the relative error of qNMR concentrations to be 5 %, considering the inaccuracy in buffer concentration and the processing-related uncertainty (especially from phasing).

The more concentrated samples of the dilution series of CTTCGAAG were not available anymore at the time the qNMR was deployed. We used the ^{31}P spectra with ^{1}H decoupling already measured during the VT experiments for determining the concentrations. As the repetition time had not been long enough to achieve the fully relaxed state before each scan and the decoupling influences the spectral intensities, too, we introduced a correction factor. Its value was obtained by a subsidiary experiment comparing a fully relaxed spectrum and that with the decoupling and shorter relaxation delay. The relative error is estimated to increase up to 10 % in this case.

5.7.2 Concentrations of samples for ultraviolet and Raman spectra

Additional melting experiments employing UV and RS also require the knowledge of the ODN concentrations. We determined them as the average values calculated from the measured absorbances (of the proper solutions used for the UV melting and of highly diluted RS samples) at

260 nm both in the duplex and in the melted state and the extinction coefficients predicted by the NN method [197].

Serious estimates of UV extinction coefficients are unavailable for m^5C-containing sequences. We performed auxiliary experiments comparing ^{31}P qNMR with UV spectra of a 50-fold diluted solution [198] that yielded extinction coefficients of unmethylated DNA the same as the NN prediction [197] (within the experimental error). No significant effect of methylation was found, therefore, we used the same NN predictions for determining the ODN concentrations for the unmodified as well as for the methylated sequences.

We estimate the accuracy of the concentrations as 2 % for the samples used in UV and 5 % for those used in RS.

6 METHODS OF DATA ANALYSIS

6.1 Asymexfit

As a part of my work on this thesis, ASYMEXFIT (Asymmetric exchange fitting) toolbox for MATLAB environment has been developed. This in-house software was needed in order to analyse the VT series of relatively complicated ^{1}H spectra of ODN because the existing packages (WinDNMR [199], gNMR [200], the Dynamic NMR module of Bruker Topspin [201], matNMR [202], or SIMPSON [203]) are not satisfactory: our data need not only the equations for asymmetric chemical exchange for many peaks in the spectrum with explicit independent line widths, but also temperature-dependent estimates of the spectral parameters of the exchanging sites, robust determination of errors, and automatic extraction of the results. Easily made custom extensions and batch jobs are also desirable. Apart from the least-square fitting of the NMR spectra, Asymexfit handles the thermodynamic models underlying the melting curves.

Asymexfit has been continuously updated and extended. It has been made freely available to download over the Internet at the department website `http://nmr.mff.cuni.cz` together with the manual. Its latest stable version is 2.3.

Asymexfit exploits the possibility of OBJECT-ORIENTED PROGRAMMING (OOP) in MATLAB: its main class, `NMRspectra`, includes – among other – the methods for loading the spectra from the Bruker spectrometer format (`loadBruker`), manipulations with the many-peak parameters and the spectral series (`addpeaks`, `removepeaks`, `next`, etc.), and, finally, simulating and fitting of the Lorentzian as well as two-site-exchange resonances (`fit`) and error estimation (`estimateerrors`).

Other classes (abstract `TDdata` and its descendants `populationData` and `EyringData`) are also implemented in order to fit the equilibrium and

kinetic data. Straightforward conversions of the results from `NMRspectra` class to the thermodynamic analysis form integral parts of the package.

6.2 NMR line-shape fitting

The peaks in the VT ^{1}H NMR spectra were fitted by their model line shapes by Asymexfit. The ^{1}H spectra were divided into several frequency domains. All the peaks in each domain were fitted together with independent parameters, comparing the experimental spectrum with the sum of the individual resonance lines and a linear baseline. A common phase correction was also optimised.

Doublets and higher multiplets caused by the indirect spin–spin interactions were treated by the weak-coupling limit as sums of individual transitions with their relative intensities fixed according to the coupling scheme. For pulse sequences with spin echoes, their phases according to Eq. (4.18) were introduced as fixed values $-\tau_{\text{echo}}$ was determined from the acquisition parameters.

6.2.1 Lorentzian curves

Lorentzian curves $\mathcal{L}$ (4.13) were fitted to the spectral lines that have their widths unmodified by chemical exchange. This can happen if:

a. no exchange is present;
b. the exchange is too fast resulting in one peak;
c. the exchange is too slow resulting in two (or more) peaks.

The chemical shifts δ of the resonances, their apparent transverse relaxation times[1], T_2^*, and their intensities were obtained by this approach. Errors in δ are reported as one fifth of FWHM of the peak,

$$\Delta\delta = \frac{1}{5}\,\frac{2}{T_2^*\omega_0}\,, \tag{6.1}$$

but at least 0.002 ppm. Where the spectral fit failed to give correct estimate of T_2^* and fell into a boundary, it was treated as if $T_2^* = 10$ ms in order to respect the wrong convergence of the fit[2].

[1] we use the apparent T_2^* that include the inhomogeneous line broadening instead of the real T_2 that are inaccessible by the Lorentzian line-shape fitting

[2] for the DNA octamers, the usual values of T_2^* are between 30 ms and 100 ms in duplex and they are roughly doubled in single strands

6.2.2 Extrapolation of chemical shifts and line widths

The analysis of intermediate two-site chemical exchange requires some independent estimates of the chemical shifts δ_A and δ_B of the exchanging sites as well as T_{2A}^* and T_{2B}^*.

The ^{1}H chemical shifts obtained from the Lorentzian fitting at low and high temperatures, where the effect of the chemical exchange was negligible, were subjected to linear regressions with respect to T. The parameters of the straight lines, together with their prediction limits on 99 % level, then served for extrapolations to the temperatures where the chemical exchange modifies the resonances.

The times T_{2A}^* and T_{2B}^* were treated differently: they were taken as constants with respect to T. These were obtained as maximal and average values from the Lorentzian fits performed on the low- and high-temperature spectra, respectively. Their confidence limits were selected to include the average values of all T_{2A}^* and as three-times the standard deviations (3σ) of all T_{2B}^* of the particular nucleus.

6.2.3 Exchange line shapes

Peaks modified by a two-site chemical exchange were fitted by the corresponding line shape $\mathcal{X}$ (4.33). The extrapolated values of δ_A, δ_B, T_{2A}^*, and T_{2B}^* were used as fixed parameters. The exchange-related quantities, p_A and k_A were optimised by the fit, together with the line intensities. In this way, an adequate number of degrees of freedom was associated with the whole procedure.

In order to estimate the errors of the resulting p_A and k_A, spectral fits were repeated with varying the fixed values of δ_A and δ_B in the first round and of T_{2A}^* and T_{2B}^* in the second round within their confidence intervals. As the third round, the fits were performed on a series of 1000 artificial spectra calculated for the optimal parameters with added Gaussian noise on the same level as the experimental spectra [204]. The final errors were estimated as square roots of the sums of squares of 3σ of the distributions obtained by the three rounds.

6.2.4 Application to the DNA melting

The low density of overlaps in the aromatic and methyl region allowed the fit of the exchange line shapes around the melting point. On the other hand, the crowded region of H1′ and CH5 peaks with relatively complicated J-coupling patterns permitted only the Lorentzian fits. The exchangeable protons give no signal in single strands, therefore no exchange fit is possible. This general scheme is presented in Table 6.1. Consequently, the non-labile hydrogens are naturally divided into two main classes[3]: the aromatic (H2, H6, H7, and H8, collectively referred to as Ha, fitted by exchange) and anomeric protons (H1′).

Table 6.1 Functions selected for the fits of DNA-melting spectra at low, medium and high temperatures. The types of hydrogens are approximately sorted by their chemical shifts. $\mathcal{L}$: Lorentzian curve (4.13); $\mathcal{X}$: exchange curve (4.33). Missing values for the labile protons indicate unobservable resonances

	H1, H3	H41, H42	H6, H8	H2	H5	H1′	H7
Low T	$\mathcal{L}$	$\mathcal{L}$	$\mathcal{L}$	$\mathcal{L}$	$\mathcal{L}$	$\mathcal{L}$	$\mathcal{L}$
Medium T			$\mathcal{X}$	$\mathcal{X}$	$\mathcal{L}$	$\mathcal{L}$	$\mathcal{X}$
High T			$\mathcal{L}$	$\mathcal{L}$	$\mathcal{L}$	$\mathcal{L}$	$\mathcal{L}$

In all the ODN studied, the chemical exchange was fast or intermediate, therefore a single, usually moderately broadened peak was present for each proton. The exchange fit was not feasible for several particular Ha resonances which were too narrow (owing to small differences between δ_A and δ_B) or which overlap with some other signal in such a way that spoils the results. Lorentzian curves replaced the exchange line shapes in these cases.

Spectral regions containing all the deoxyribose protons other than H1′ were completely omitted due to multiplet patterns leading to too frequent overlaps.

[3] all the cytosine H5 signals, though aromatic, have been modelled by Lorentzian curves only, so we leave them intentionally off these groups

6.3 Employment of thermodynamic models

6.3.1 Individual fits

The temperature-dependent parameters obtained from the spectral analyses (δ or p_A and k_A) were fitted by Asymexfit in the sense of the presumed two-state equilibrium using:

a. Eq. (3.25) for δ as x with $x_A = \delta_A$ and $x_B = \delta_B$ as linear functions of T with their parameters optimised by the fitting procedure;
b. Eq. (3.27) or (3.28) for p_A;
c. Eq. (3.8) for $k_A = k_f$.

The errors were used for weighting in the χ^2 calculations. By these fits, ΔH and ΔS were determined individually from each proton resonance, allowing the calculation of $\Delta G_{310\,K}$ and T_m. For the spectral lines fitted by the exchange shape, activation parameters, $\Delta H^\dagger$, $\Delta S^\dagger$, and $\Delta G^\dagger_{310\,K}$ calculated from them, were obtained, too.

6.3.2 Global fits

The melting profiles – either $p_A(T)$ or $\delta(T)$ – were also fitted simultaneously for all values obtained from a class of proton resonances of each ODN sample. The optimised global enthalpy, ΔH^{glob}, and entropy, ΔS^{glob}, were kept common for all the data in each set. In this way, folding parameters were achieved as global quantities describing the overall properties of the molecule under study: ΔH^{glob}, ΔS^{glob}, T_m^{glob}, and $\Delta G^{glob}_{310\,K}$.

The kinetics has been analysed in the same manner: simultaneous fits of $k_A(T)$ by the Eyring equation (3.8) yielded global thermodynamic quantities of the activation of unfolding: $\Delta H^{\dagger,glob}$, $\Delta S^{\dagger,glob}$, and $\Delta G^{\dagger,glob}_{310\,K}$.

7 SELF-COMPLEMENTARY DNA DUPLEXES

In this chapter, various properties and spectral characteristics derived from ^{1}H NMR spectra of self-complementary (palindromic) oligodeoxynucleotides are presented and discussed. The chemical shifts are confronted with the nearest-neighbour model. Their dependence on base sequence, temperature, and ODN concentration is investigated. Then, the thermodynamics of the disruption (melting) of the DNA double-helix is described at the level of individual nucleotides, which brings new insights into the cooperativity and local deviations in the denaturation process. Exploiting the possibilities of NMR spectroscopy, values of both the equilibrium and kinetic quantities are estimated. The two-state approximation of the duplex melting is challenged.

Some of the key methods have been tested during my diploma thesis on two ODN samples [154] and one sample was measured during the bachelor thesis by Ondřej Socha under my supervision [205]. In this dissertation, the set of samples has been significantly expanded and the approaches used have been further developed into a consistent workflow that is even directly transferable to solve any similar problem.

7.1 Samples

Our study employed a set of eight ODN each containing two occurrences of all of the four major nucleobases. Their compositions and concentrations in solutions used in the NMR experiments are listed in Table 7.1.

7.2 Assignment of ^{1}H NMR spectra

Chemical shifts of ^{1}H nuclei were assigned by the standard procedures described in § 4.3.1 on page 40. The 2D ^{1}H NOESY spectra were acquired in duplex and the temperature dependence of 1D spectra served for tracking the resonances to single strands (§ 5.4 on page 47).

Table 7.1 Self-complementary ODN samples and their concentrations, c, measured by ^{31}P qNMR and estimated from the provider UV data using the extinction coefficients calculated by the provider from the base composition (Bases) and calculated from the NN model [197] (NN)

	c / mM		
	^{31}P qNMR	Bases	NN
Basic series			
CAACGTTG	1.10 ± 0.05	1.34	1.36
CATCGATG	1.17 ± 0.06	1.22	1.22
CTACGTAG	0.90 ± 0.05	0.97	0.95
CTTCGAAG	1.06 ± 0.05	1.10	1.12
CTTGCAAG	0.90 ± 0.04	0.94	0.97
GAACGTTC	0.94 ± 0.05	1.13	1.13
GATGCATC	0.87 ± 0.04	0.89	0.89
GTAGCTAC	0.95 ± 0.05	1.04	1.03
Additional concentrations			
CTACGTAG	4.98 ± 0.25	4.75	4.67
CTTCGAAG	3.92 ± 0.39^b	5.00	5.07
	1.80 ± 0.18^b	2.50	2.54
	0.95 ± 0.10^b	1.25	1.27
	0.53 ± 0.03	0.63	0.63
GTAGCTAC	[a]	4.05	4.00

[a]Spectra of ^{31}P not acquired. [b]Concentrations from decoupled ^{31}P spectra

7.2.1 Assignment in duplex

NOESY spectra at a single temperature were always sufficient to assign the proton spectra in duplex. The sequential walks over the H1′–H6/H8 cross-peaks, shown for all ODN sequences in the figures of § A.1 of Appendix (page A-1), were supplemented and confirmed by the region with the H2′–H6/H8 and the H2″–H6/H8 connectivities as well as by other spectral areas. Example of this strategy is shown in Fig. 7.1. All the ^{1}H chemical shifts of all the eight self-complementary duplexes under study are collected in Table A.1 in Appendix.

The signals of several diastereotopic pairs H2′ and H2″, mostly in the 3′-terminal nucleotides, remain unseparated due to their mutual overlap, which is already known to happen [109]. The too little difference between

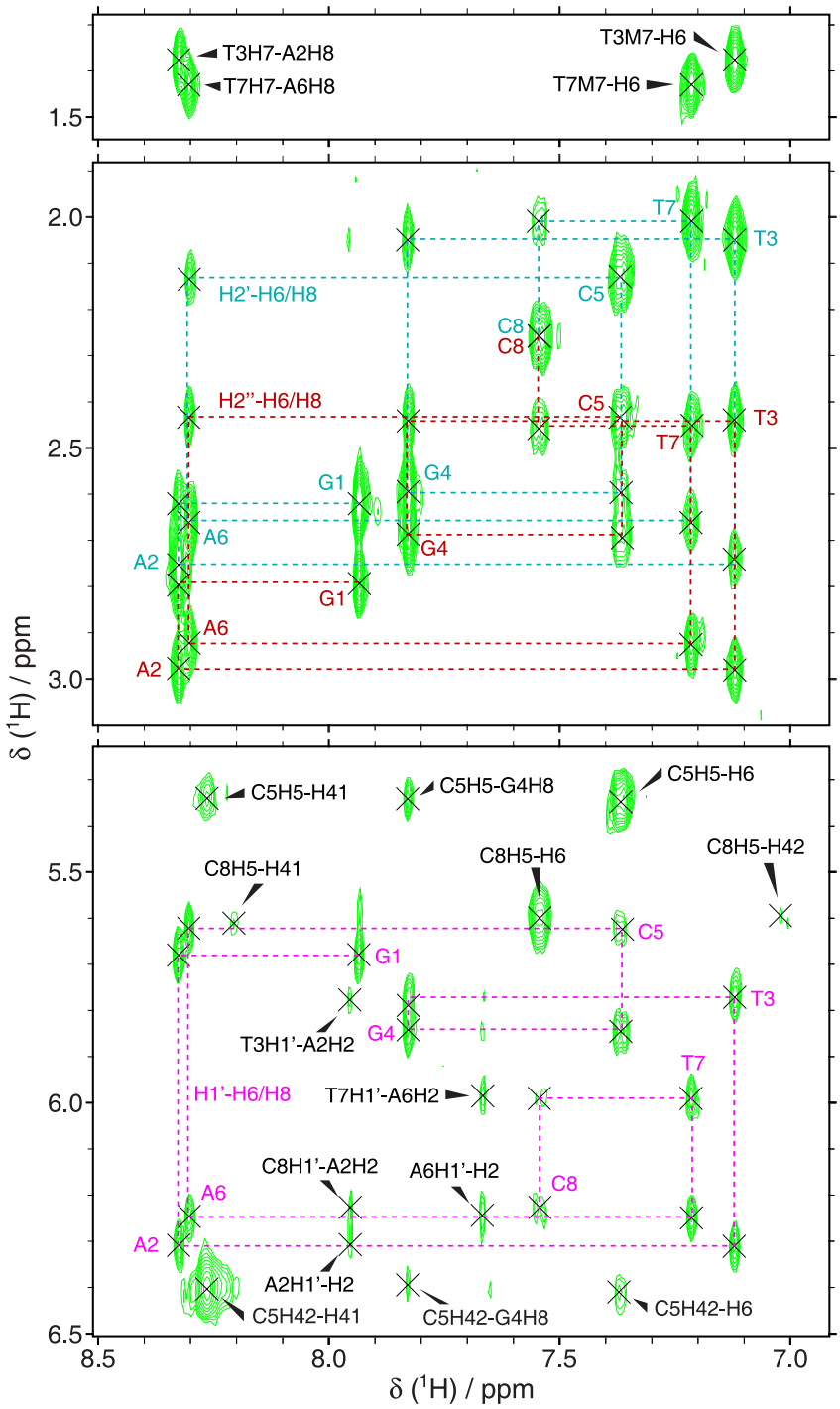


Figure 7.1 Portions of ^{1}H NOESY spectrum of GATGCATC (284 K). Sequential walks depicted by dashed lines (H1′–H6/H8 in magenta, H2′–H6/H8 in cyan, H2″–H6/H8 in dark red) with the intra-nucleotide cross peaks labelled. Extra cross-peaks annotated

their chemical shifts is stressed by the complicated multiplet patterns of the H2′ and H2″ lines experiencing strong geminal coupling.

Only some types of exchangeable hydrogen nuclei are observable in the spectra: guanine and thymine imino (GH1 and TH3) and cytosine amino protons (CH41 and CH42). The other types of hydrogens bound to nitrogen atoms exchange with water hydrogens so quickly that they remain invisible under our experimental conditions [157]. Therefore, we cannot measure the resonances of adenine (AH61 and AH62) and guanine (GH21 and GH22) amino protons, even though they form parts of the hydrogen bonds in the Watson–Crick base pairing.

7.2.2 Assignment in single strands

Tracking of proton resonances assigned in duplex along the VT 1D spectra to the unfolded state resulted in determining the single-strand ^{1}H chemical shifts of almost all hydrogen nuclei (Table A.3 in Appendix).

Some spectral lines experience strong broadening, overlap with other signals, and virtually vanish during the duplex melting due to the chemical exchange. For assigning such resonances in single strands, additional experiments have been performed: homonuclear decoupling (Fig. 7.2), se-

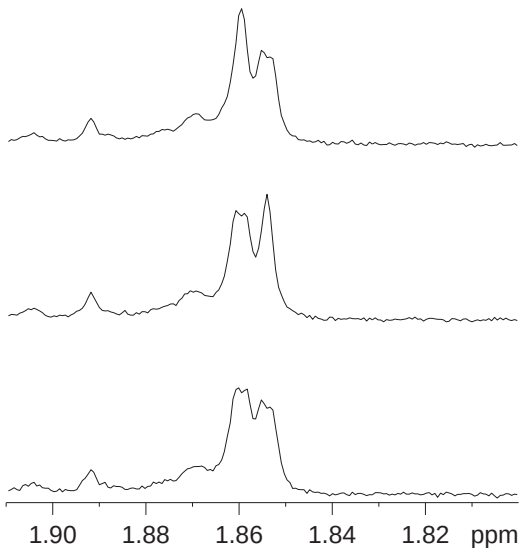


Figure 7.2 Assignment of TH7 in CTTCGAAG (0.22 mM, 350 K) by homonuclear decoupling. From top to bottom: irradiating T2H6, T3H6, and an empty spectral region at 7 ppm as a control

lective 1D COSY, and selective ^{1}H–^{13}C HMBC (Fig. 7.3) at elevated temperatures and measurements on more concentrated solutions, optionally at a lower magnetic field. C4H1′ and T6H1′ of CTACGTAG were distinguished only by prediction by DSHIFT [184] which we consider safe in this case as the chemical shift difference between the two peaks is sufficiently large, 0.15 ppm. The only unresolved resonances remaining are three peaks of GATGCTAC, T2H1′, G4H1′, and T6H1′, which lie so close to each other that any attempt to their separation fails. All this treatment is gathered in Table 7.2.

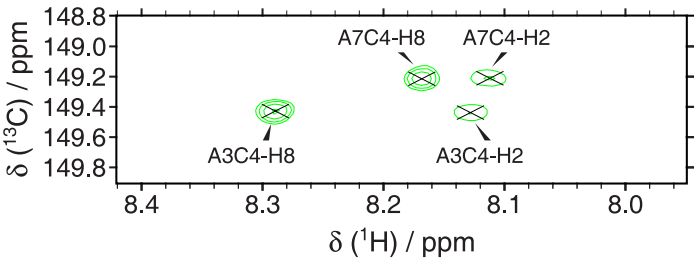


Figure 7.3 Selective ^{1}H–^{13}C HMBC spectrum of CTACGTAG (346 K)

Table 7.2 Resonances tracked ambiguously from duplex to single strands and methods of resolving them with marks indicating whether the approach has been successful

Sample	Ambiguities	Method applied, conditions	
CATCGATG	TH7	selective COSY, 338 K	✓
CTACGTAG	GH8	higher concentration, 300 MHz	✓
	AH2	higher concentration, 300 MHz	×
		selective ^{1}H–^{13}C HMBC, 346 K	✓
	C4, T6H1′	prediction by DSHIFT	✓
CTTCGAAG	TH7	selective COSY, 350 K	✓
		homonuclear decoupling, 350 K	✓
	AH2	higher concentration	✓
CTTGCAAG	TH7	selective COSY, 340 K	✓
GTAGCTAC	CH6	higher concentration	✓
	AH2	higher concentration	×
		selective ^{1}H–^{13}C HMBC, 334 K	✓
	T2, G4, T6H1′	no means available	×

7.2.3 A complete picture

The successful assignment of the ^{1}H peaks led to full interpretation of the NMR spectra both in duplex and in single strands; an example is shown in Fig. 7.4. The chemical shifts determined from the fitting of 1D spectra by Lorentzian curves (only H2′ and H2″ originate directly from 2D NOESY) are going to be analysed in the following sections. Next, the VT series of all the ODN sequences, as in Fig. 7.5, will serve to the detailed investigation of the melting process.

7.3 Sequence dependence of ^{1}H chemical shifts

The rich set of experimentally determined chemical shifts for eight octamer sequences was analysed with the aim to answer the following questions: Does the NN model described in § 4.3.2 on page 42 satisfy the real data to a reasonable precision? Can the prediction protocols be reliably used to assign resonances instead of or in conjunction with experimental strategies? Can we predict anomalous local structure comparing the estimated chemical shifts with the measured values?

7.3.1 Next-nearest-neighbour effects

The experimental values of ^{1}H chemical shifts were first compared among themselves. Supposing that the identities of adjacent bases represent the strongest contribution, proton shifts of the central nucleotides of various base triplets inside the sequence were compared. Terminal-nucleotide shifts in doublets on either ODN end were also considered. The differences between δ of protons in the same NN environment found in different remote surroundings are shown in Fig. 7.6 (duplex) and Fig. 7.7 (single strands) for the internal nucleotides and in Fig. 7.8 (duplex) and Fig. 7.9 (single strands) for the terminal nucleotides. Correlation plots of δ are located in § A.3 of Appendix (page A-17).

The comparison of the values present at least twice in our data shows that while in many cases the agreement is good, there are not seldom cases where the differences are reaching $\Delta\delta \approx 0.1$ ppm, which is well above the proclaimed error of the prediction method, 0.03 ppm [163].

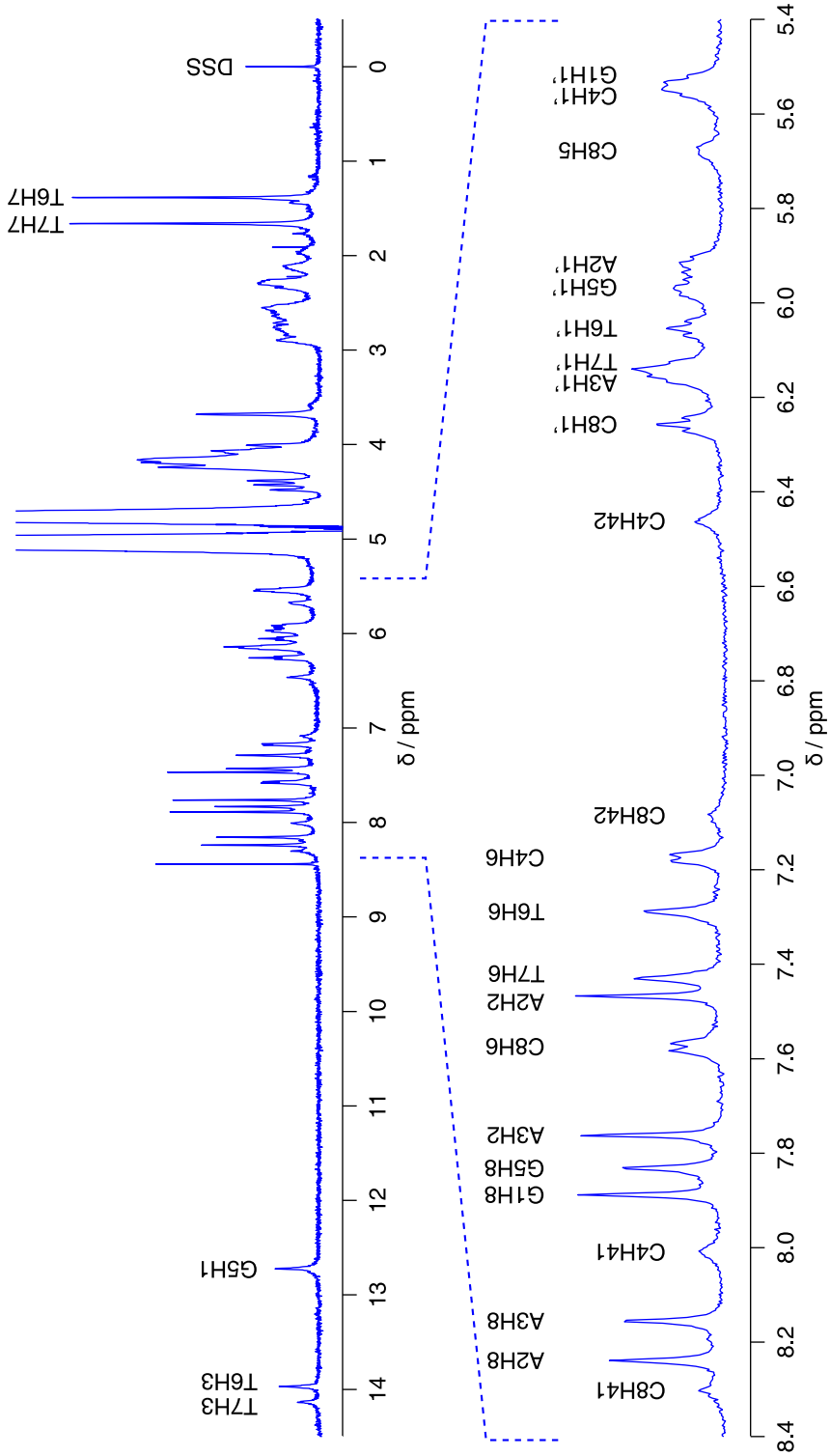


Figure 7.4 Spectrum of ^{1}H NMR of GAACGTTC duplex at 285.7 K. Top: full spectrum; bottom: expansion of the part with aromatic and anomeric protons. Assignments of selected resonances are indicated in the figure

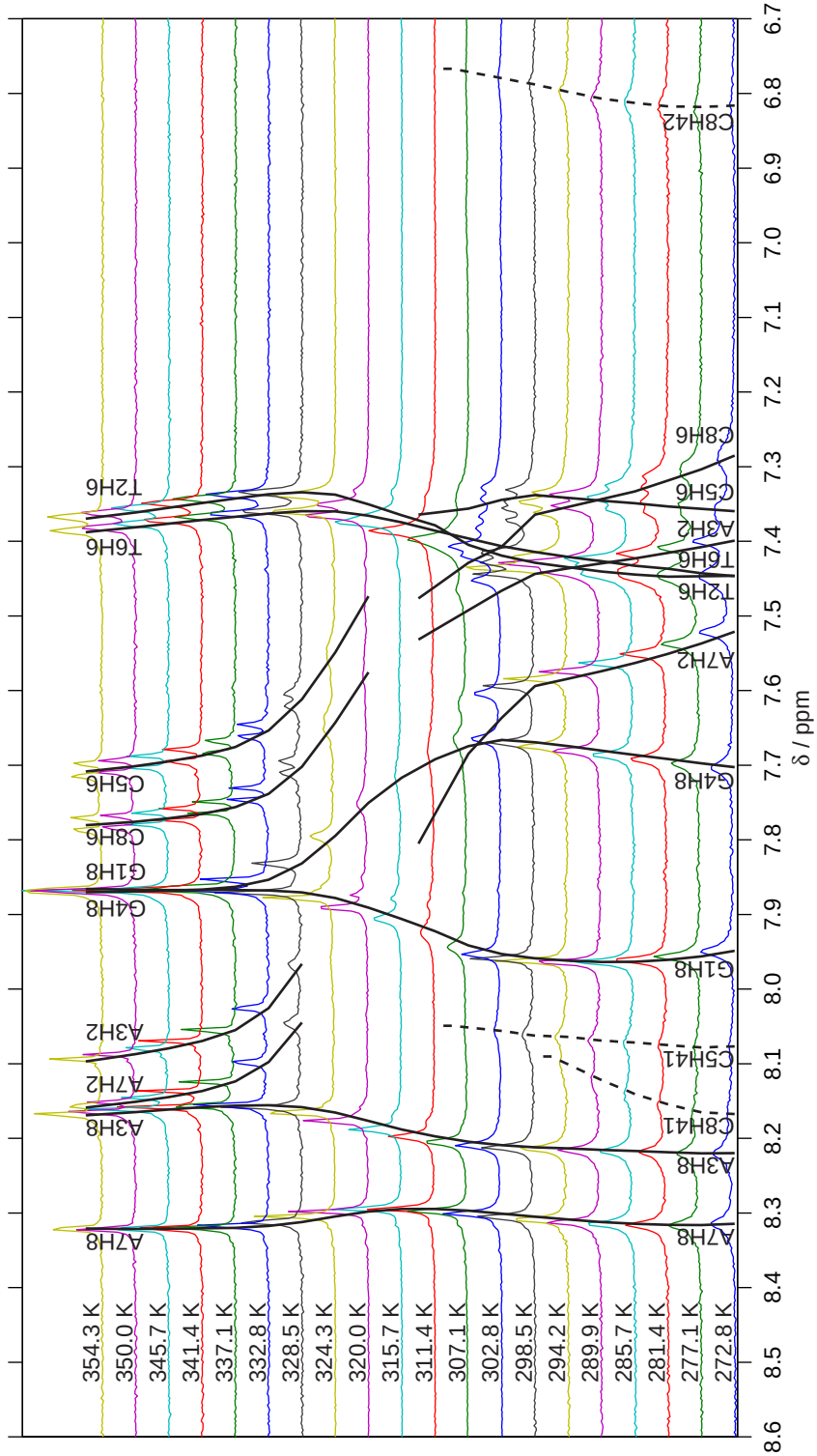


Figure 7.5 Typical ^{1}H VT NMR spectra: aromatic part of GTAGCTAC at every second temperature measured (indicated on the left). Peak positions, coming from the exchange line-shape fits, are marked by annotated solid (non-exchangeable hydrogens) and dashed (amino protons) black curves; their parts are missing where the increased line widths did not permit the fits

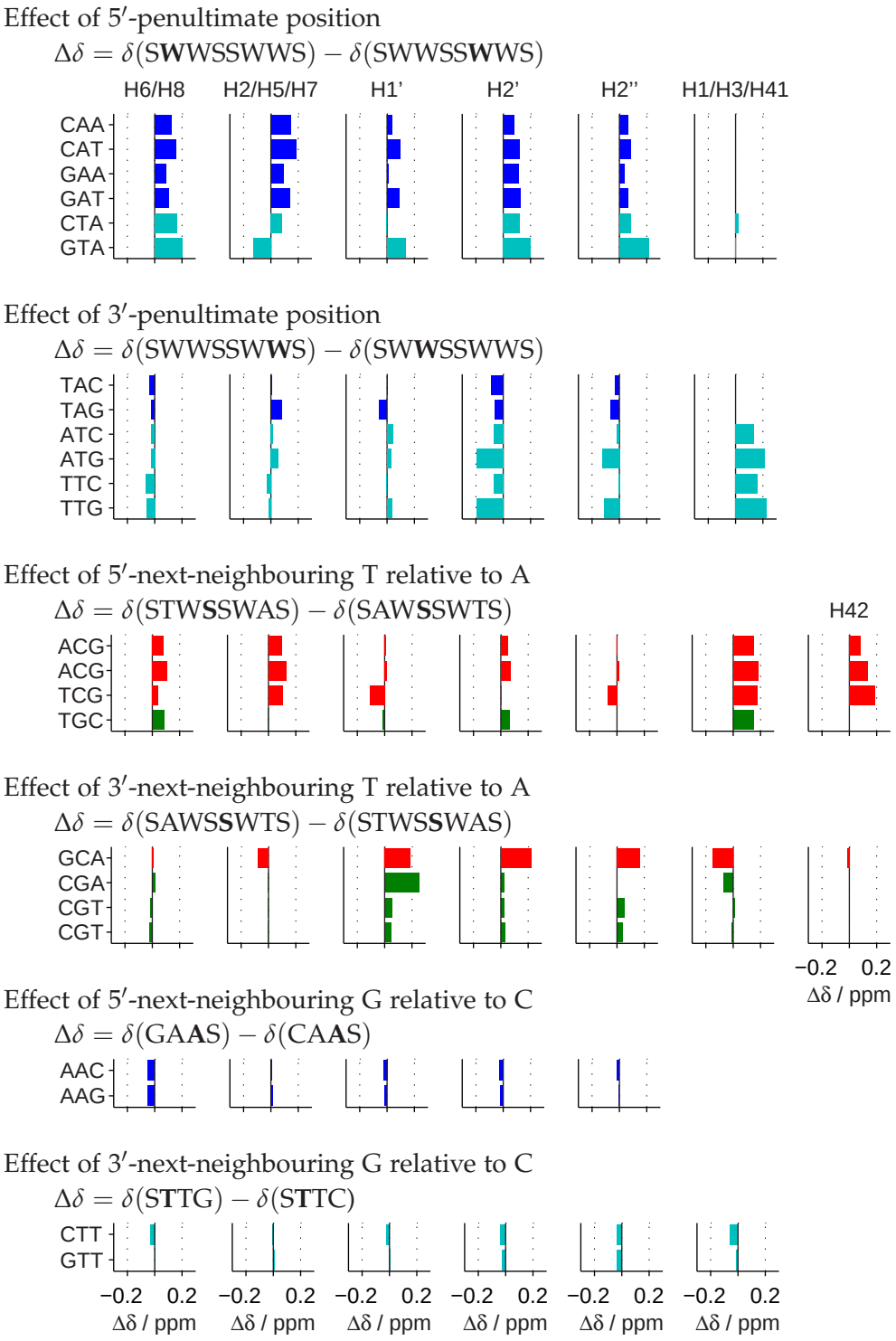
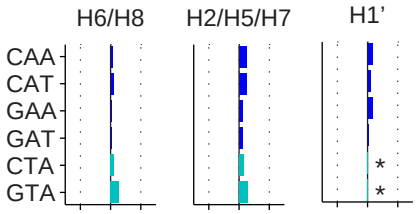


Figure 7.6 Next-nearest-neighbour effects on δ in nucleosides in bold (duplex, 284 K). ACG and CGT found three times in our ODN set. Bars colour-coded by the base type. W = A or T; S = C or G

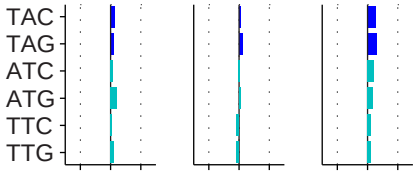
Effect of 5′-penultimate position

$\Delta\delta = \delta(\mathrm{S\mathbf{W}WSSWWS}) - \delta(\mathrm{SWWSS\mathbf{W}WS})$



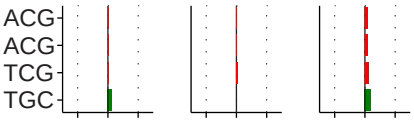
Effect of 3′-penultimate position

$\Delta\delta = \delta(\mathrm{SWWSSW\mathbf{W}S}) - \delta(\mathrm{SW\mathbf{W}SSWWS})$



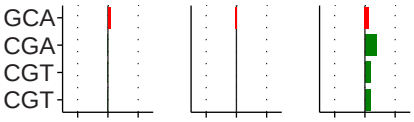
Effect of 5′-next-neighbouring T relative to A

$\Delta\delta = \delta(\mathrm{STW\mathbf{S}SWAS}) - \delta(\mathrm{SAW\mathbf{S}SWTS})$



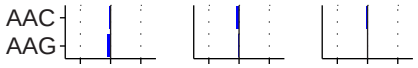
Effect of 3′-next-neighbouring T relative to A

$\Delta\delta = \delta(\mathrm{SAWS\mathbf{S}WTS}) - \delta(\mathrm{STWS\mathbf{S}WAS})$



Effect of 5′-next-neighbouring G relative to C

$\Delta\delta = \delta(\mathrm{GA\mathbf{A}S}) - \delta(\mathrm{CA\mathbf{A}S})$



Effect of 3′-next-neighbouring G relative to C

$\Delta\delta = \delta(\mathrm{S\mathbf{T}TG}) - \delta(\mathrm{S\mathbf{T}TC})$

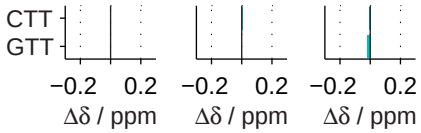


Figure 7.7 Next-nearest-neighbour effects on δ in nucleosides in bold (single strands, 354 K). ACG and CGT found three times in our ODN set. Bars colour-coded by the base type. W = A or T; S = C or G. Asterisks: missing values due to ambiguous assignment of resonances

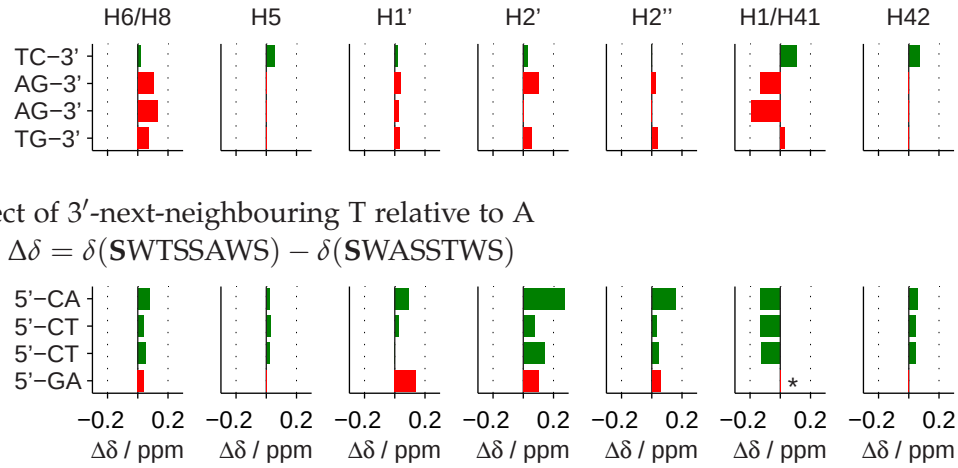


Figure 7.8 Next-nearest-neighbour effects on δ in terminal nucleosides in bold (duplex, 284 K). AG-3′ and 5′-CT found three times in our ODN set. Bars colour-coded by the base type. W = A or T; S = C or G. Asterisk: missing value due to an unobserved peak

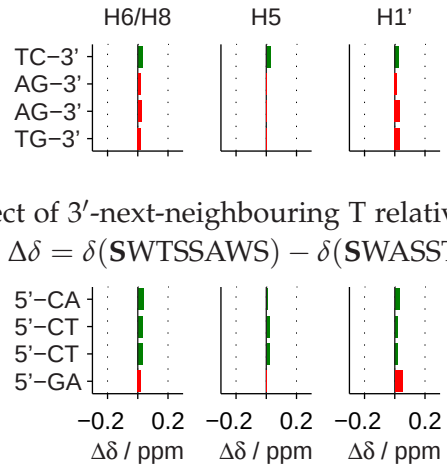


Figure 7.9 Next-nearest-neighbour effects on δ in terminal nucleosides in bold (single strands, 354 K). AG-3′ and 5′-CT found three times in our ODN set. Bars colour-coded by the base type. W = A or T; S = C or G

Such demonstrations of the sequence dependence are clearly violating the assumptions of the NN models.

Inspection of the influences of the next-nearest neighbours in **duplex** reveals several main features:

a. proximity of the 5′-end moves all non-exchangeable protons except TH7 downfield;
b. proximity of the 3′-end moves H2′ and H2″ upfield and TH1 downfield;
c. the base (including exchangeable) protons of internal C and G are sensitive to the 5′-next-nearest neighbour whereas their sugar protons respond rather to the 3′-next-nearest neighbour: in both cases, T causes downfield shifts relative to A;
d. the previous point applies for the non-exchangeable protons in the terminal nucleotides as well.

The end effects mentioned above correspond to the penultimate corrections in the prediction protocol for double-helical shifts [163]. Conversely, the effect of 5′-penultimate position on TH3 shifts published as 0.10 ppm is not seen in our data. The cited work also notes decreased shielding of H5, H6, H7, and H8 after replacing a 5′-next-neighbouring purine by pyrimidine and a similar shift of H1′, H2′, and H2″ upon the same change but in the 3′-direction. These changes are reproduced in our collection (the substitution of A by T); in addition, our data indicate strong uniform changes in AH2 and exchangeable protons upon the same replacement, which is not recorded in the literature and should probably deserve more attention.

Contrary to these effects which seem to be universal, the T-for-A substitution at the 5′-side causes opposite shift changes for two out of three GH1 of terminal guanines than what was observed for the internal residues. Moreover, TH7 moves in opposite directions when facing the 5′-penultimate position in CTA and GTA contexts. This proves that the problem with next-nearest neighbours cannot be solved now by a simple set of additive corrections. Generalisation of the observed long-range effects would require large data sets containing experimentally obtained shifts at least in all possible pentanucleotides.

In **single strands** (Fig. 7.7 and Fig. 7.9), the effect is not that strong

as in duplex and cannot be looked at as persuasive considering our limited data set, although several cases indicate small, yet systematic drifts. These are in agreement with the corrections for the next-nearest neighbours implemented in the method for predicting single-strand chemical shifts [173, 184]. On the other hand, such little shielding and deshielding differences in the order of 0.01 ppm might be overridden by the effect of temperature or even third neighbours, both neglected in the original paper.

We can **conclude** that taking only NN into account creates differences frequently exceeding 0.1 ppm in duplex but only in the order of 0.01 ppm in single strands. In contrast, the published works neglect the next-nearest neighbour effects in duplex with the exception of ODN ends [163] but treat it in single strands [173, 184]. In the view of our experimental chemical shifts, the former approach seems oversimplified whilst the latter one complicated too much regarding other sources of comparable imperfections.

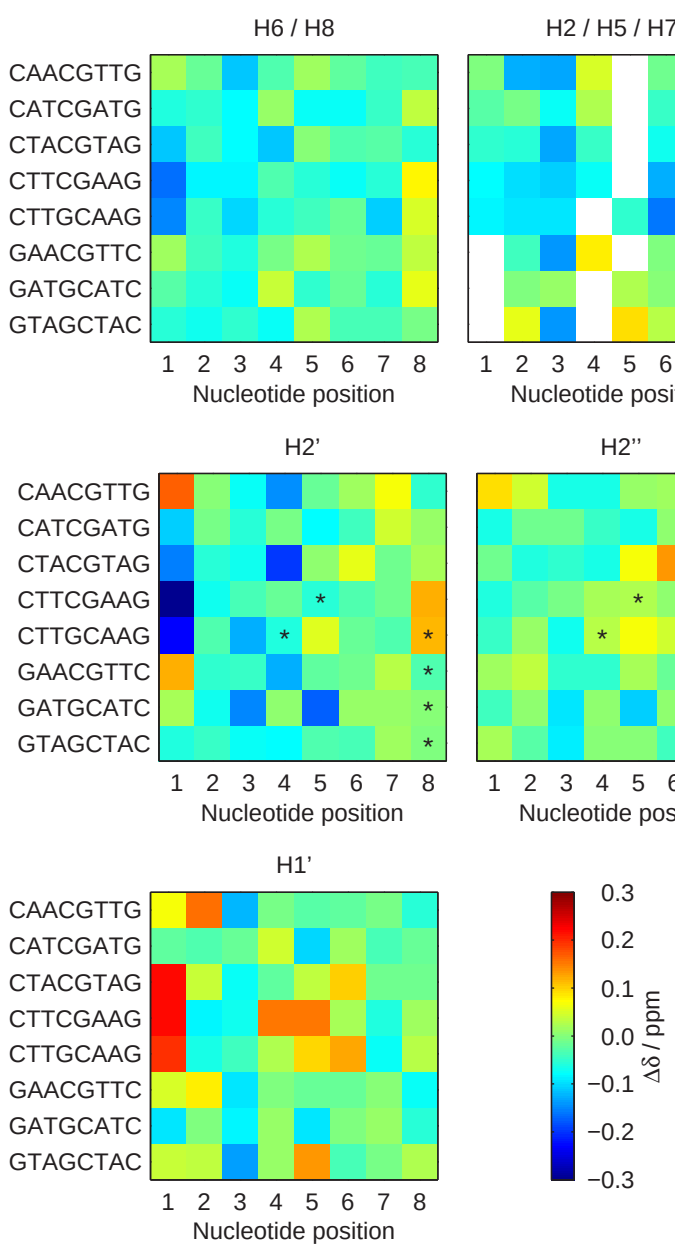
7.3.2 Differences between prediction and experiment

The next step of our analysis was a comparison of our experimental values of ^{1}H chemical shifts, δ_{Exp}, with the predictions, δ_{Pred}, calculated by the DSHIFT web server [184]. For duplex, the method of Altona [163] was chosen. Unfortunately, the selection of the ODN sequences in the original works of Lam [173, 184] does not allow to estimate chemical shifts of the terminal nucleosides.

I have reported a discrepancy between the chemical shift calculated by DSHIFT and presented in the original paper [184], concerning H2 of the central adenine in GAA triplet, to the DSHIFT authors. The difference revealed, 0.011 ppm, is very subtle and the authors updated the program immediately. Nevertheless, one should check the values in the articles from time to time, even though using DSHIFT is much more comfortable.

The differences $\Delta\delta = \delta_{\mathrm{Pred}} - \delta_{\mathrm{Exp}}$ are plotted in Fig. 7.10 (non-exchangeable protons) and Fig. 7.11 (exchangeable protons) for duplex and in Fig. 7.12 for single strands.

The values of $\Delta\delta$ in **duplex** are in most cases exceeding the published root-mean-square error (0.01–0.03) ppm of δ_{Pred} [163], but no globally sys-



$\sqrt{\langle(\Delta\delta)^2\rangle}$ / ppm

H6/8	H2/5/7
0.05	0.08
0.06	0.04
0.07	0.11
0.08	0.11
0.08	0.11
0.03	0.08
0.05	0.06
0.05	0.09

H2′	H2″
0.09	0.05
0.05	0.04
0.09	0.06
0.12	0.03
0.11	0.05
0.07	0.03
0.08	0.05
0.05	0.04

H1′
0.08
0.04
0.09
0.12
0.10
0.05
0.06
0.07

Figure 7.10 Differences between predicted and experimental chemical shifts of indicated non-exchangeable protons in duplex. $\Delta\delta = \delta_{\mathrm{Pred}} - \delta_{\mathrm{Exp}}$. The overlapping resonances of H2′ and H2″ are labelled by asterisks. Root-mean-square deviations are tabulated

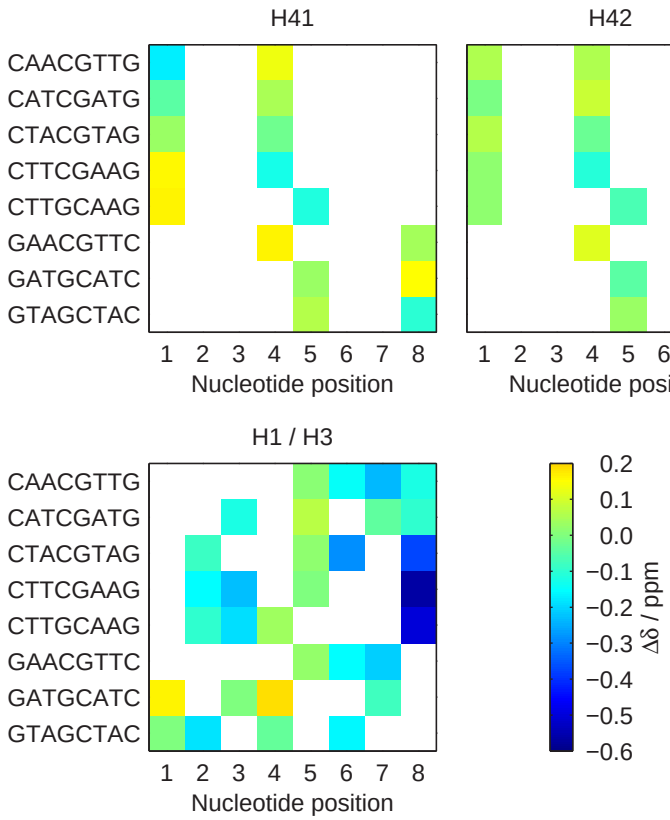
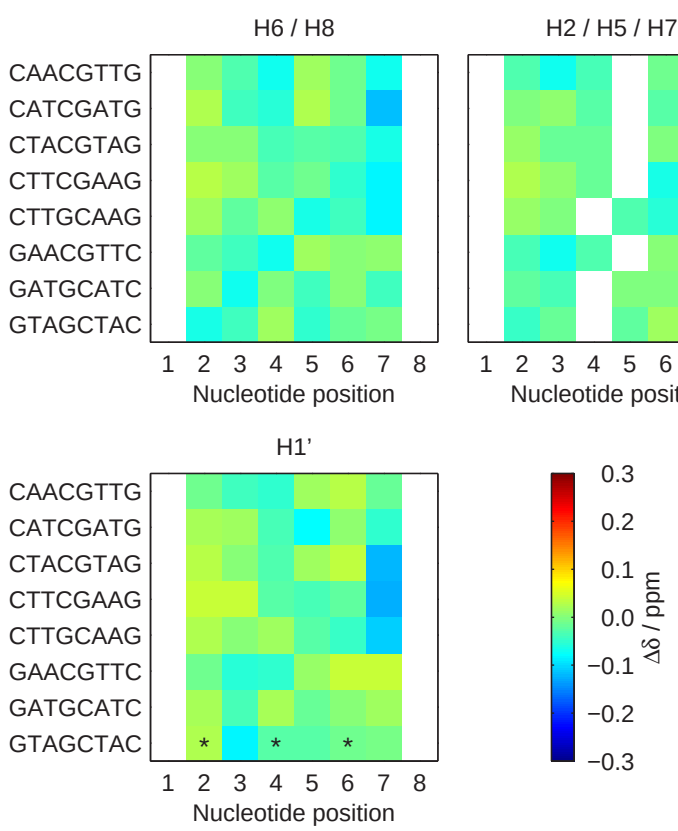


Figure 7.11 Differences between predicted and experimental chemical shifts of exchangeable protons in duplex. $\Delta\delta = \delta_{\mathrm{Pred}} - \delta_{\mathrm{Exp}}$

tematic drift can be observed. The largest discrepancies are found in the 5′-terminal and then in the 3′-terminal nucleosides. This could be explained by the large sensitivity of the ODN ends to the solvent which makes the prediction method more fragile than for internal residues. Alternatively, formation of low-populated higher-order structures at low temperatures could be responsible, which is going to be discussed based on other observations as well in § 7.5.4 on page 95.

Non-exchangeable ^{1}H δ_{Exp} in the internal nucleotides differ from δ_{Pred} usually by less than 0.1 ppm. Nuclei exceeding this limit are listed in Table 7.3. For the majority of them, δ_{Pred} lie upfield from δ_{Exp} except the anomeric protons. The relatively large deviations seen for eleven out of sixteen AH2 protons may seem to be tied to their strong sensitivity to temperature, which is ignored in the prediction: they possess upfield shift



$\sqrt{\langle(\Delta\delta)^2\rangle}$ ppm	
H6/8	H2/5/7
0.04	0.04
0.06	0.02
0.03	0.01
0.04	0.04
0.05	0.03
0.03	0.04
0.04	0.02
0.04	0.03

H1′
0.03
0.04
0.05
0.06
0.05
0.04
0.02
0.04

Figure 7.12 Differences between predicted and experimental chemical shifts of indicated protons in single strands. $\Delta\delta = \delta_{\mathrm{Pred}} - \delta_{\mathrm{Exp}}$. The ambiguously assigned resonances of H1′ protons are labelled by asterisks. Root-mean-square deviations are tabulated

with decreasing T (see below in § 7.4.1 on page 82). Therefore, lowering T at which δ_{Exp} are collected would move them closer towards δ_{Pred}. On the other hand, the slopes ($1.6 \cdot 10^{-3}$ ppm/K in average) are not strong enough to achieve the exact match – it would require cooling far below the freezing point of water.

The discrepancies between δ_{Pred} and δ_{Exp} of other than H1′ are most probably caused by the approximations taken in the prediction method since the signs of $\Delta\delta$ are conserved among the ODN sequences. Some but not all of them can be explained by the next-nearest-neighbour effects de-

Table 7.3 Non-exchangeable protons of non-terminal nucleotides with chemical shifts differing from the predicted values by more than 0.1 ppm in duplex. $\Delta\delta = \delta_{\mathrm{Pred}} - \delta_{\mathrm{Exp}}$

Sample	Nucleus	$\frac{\Delta\delta}{\mathrm{ppm}}$
CAACGTTG	A3H8	−0.11
CTACGTAG	C4H6	−0.11
CAACGTTG	A2H2	−0.12
	A3H2	−0.13
CTACGTAG	A3H2	−0.13
	A7H2	−0.22
CTTCGAAG	A6H2	−0.13
	A7H2	−0.17
CTTGCAAG	A6H2	−0.16
	A7H2	−0.16
GAACGTTC	A3H2	−0.14
GTAGCTAC	A3H2	−0.14
	A7H2	−0.14
CTTCGAAG	T3H7	−0.11

Sample	Nucleus	$\frac{\Delta\delta}{\mathrm{ppm}}$
CAACGTTG	A2H1′	0.16
	A3H1′	−0.12
CTTCGAAG	C4H1′	0.16
	G5H1′	0.16
CTTGCAAG	A6H1′	0.13
GTAGCTAC	A3H1′	−0.13
	C5H1′	0.13
CAACGTTG	C4H2′	−0.14
CTACGTAG	C4H2′	−0.19
CTTGCAAG	T3H2′	−0.12
GAACGTTC	C4H2′	−0.12
GATGCATC	T3H2′	−0.15
	C5H2′	−0.17
CTACGTAG	T6H2″	0.14

scribed above (§ 7.3.1 on page 72). The H1′ resonances present a different issue: there are tandem deviations with variable signs located at the same or adjacent pairs in three samples (Table 7.3). These outliers might indicate some unusual structure of the nucleic acid. On the other hand, they are not reproduced by other protons of the corresponding nucleotides but of A3 in CAACGTTG.

The protons of cytosine amino groups (Fig. 7.11) have δ_{Pred} always within 0.2 ppm deviation[1] from δ_{Exp}. Not counting the terminal residues, most of the imino protons differ also by less than 0.2 ppm with the exceptions listed in Table 7.4; all GH1 have the δ_{Pred} downfield from or very close to δ_{Exp} while TH3 upfield. Interestingly, T7H3 of CAACGTTG belongs to the same base pair as A2H1′ pointed out above in Table 7.3. Opposed to AH2, the discrepancies in TH3 resonances cannot be even

[1] we select larger limit than in the case of non-exchangeable protons since the shifts of exchangeable ones are generally much more sensitive to the outer conditions and the duplex stability and structure

Table 7.4 Exchangeable protons of non-terminal nucleotides with chemical shifts differing from the predicted values by more than 0.2 ppm in duplex. $\Delta\delta = \delta_{\mathrm{Pred}} - \delta_{\mathrm{Exp}}$

Sample	Nucleus	$\frac{\Delta\delta}{\mathrm{ppm}}$
CAACGTTG	T7H3	−0.23
CTACGTAG	T6H3	−0.28
CTTCGAAG	T3H3	−0.22

partially corrected by lowering the temperature at which δ_{Exp} is obtained because of their negative slope of $\delta(T)$ seen in our data as well as in [163].

In **single strands**, the observed differences between δ_{Pred} and δ_{Exp} (Fig. 7.12) mostly exceed the root-mean-square deviation 0.02 ppm published in the Lam's works [173, 184]. However, only a handful of the differences[2] reaches over 0.1 ppm (Table 7.5), which brings a significant improvement compared to duplex. Note that the three outlying anomeric protons all belong to the penultimate A7 in the 3′-terminal AG. Altogether, we consider the chemical shift predictions in single strands for our samples generally in accordance with the experimental values even though the temperature dependence of chemical shifts is not coped with in the predictions: our data are collected at 81 °C, whereas the predictions are estimated for 25 °C. The different solvent (mainly, the addition of 8 M urea in [173]) is an indispensable issue, too.

The importance of the predictions of chemical shifts in single strands grows for the resonances the assignment of which failed using the basic tracking procedure. What happens if the predictions are applied to the resonances of AH2 in single strands that were first assigned ambiguously? Since we have already resolved these conflicts by increased ODN concentration or ^{1}H–^{13}C HMBC (Table 7.2 on page 65), we can now answer this question: for all the three sequences, CTACGTAG, CTTCGAAG, and GTAGCTAC, the order of AH2 resonances is estimated correctly by the prediction, even though there are non-negligible differences between the predicted and experimental values.

[2] note that the predictions for terminal nucleotides are not available

Table 7.5 Protons of non-terminal nucleotides with chemical shifts differing from the predicted values by more than 0.1 ppm in single strands. $\Delta\delta = \delta_{\mathrm{Pred}} - \delta_{\mathrm{Exp}}$

Sample	Nucleus	$\frac{\Delta\delta}{\mathrm{ppm}}$
CATCGATG	T7H6	−0.11
CTACGTAG	A7H1′	−0.12
CTTCGAAG	A7H1′	−0.13
CTTGCAAG	A7H1′	−0.11

7.3.3 Chemical shift changes induced by duplex melting

The differences between chemical shifts in single strands (δ_{SS}, 354 K) and in duplex (δ_{Duplex}, 284 K) are gathered in Fig. 7.13 as $\Delta\delta = \delta_{\mathrm{SS}} - \delta_{\mathrm{Duplex}}$. It is clearly visible that there is a general downfield shift after melting. Relatively high $\Delta\delta$ are observed for CH1′ in both central dinucleotides, CpG and GpC. These trends are in a qualitative agreement with the predictions [156].

The increased shielding of the aromatic protons in the vicinity of the 5′-end has not been reported yet. It is connected to the large deviation of the predictions from our experimental data in duplex discussed above. Additionally, G5H8 of two of the CpG steps display upfield shifts, too. One of the sequences in question, CTTCGAAG, has already shown anomalous chemical shifts of H1′ protons of the CpG motif (Fig. 7.10, Table 7.3).

7.3.4 Advantages and drawbacks of the predictions

Taking the previous discussion into account, we are persuaded that the current predictions can estimate the chemical shifts in short ODN with about 0.1 ppm precision both in duplex and single strand. Worse predictions are made for terminal residues (available only for the double helix so far), exchangeable protons, and other resonances with strong temperature dependence, mainly AH2.

Large deviations between δ_{Exp} and δ_{Pred} mainly in the adenine repeat of CAACGTTG and in the cental CpG motif of CTTCGAAG probably mark local conformational anomalies. Otherwise, our data agree with the

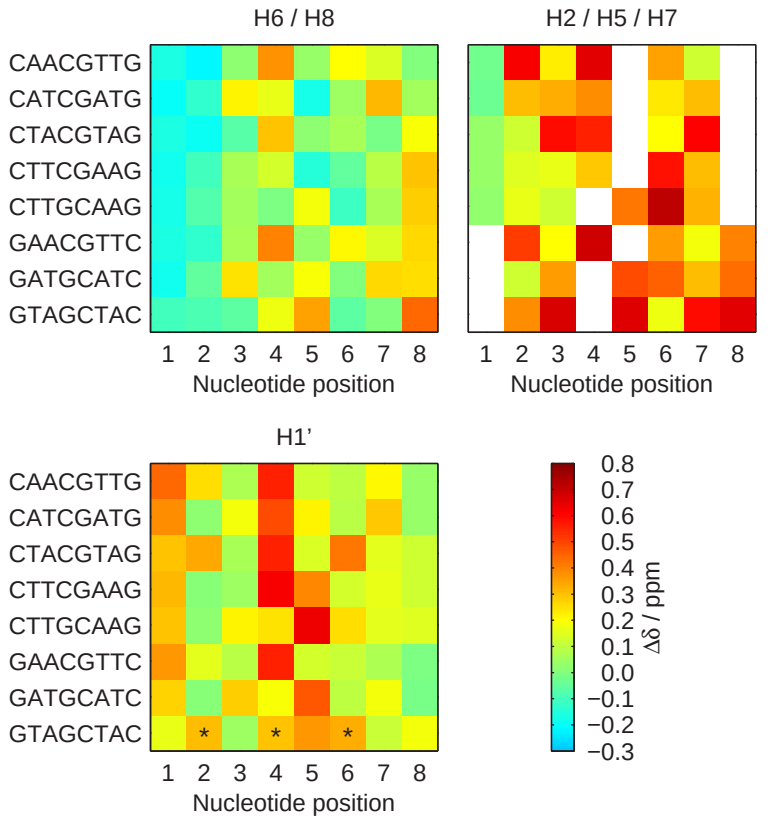


Figure 7.13 Melting-induced changes in chemical shifts of indicated protons, $\Delta\delta = \delta_{SS} - \delta_{Duplex}$. The ambiguously assigned resonances of H1′ protons in single strands are labelled by asterisks

NN models to a reasonable level. Together with the standard patterns found in NOESY spectra, this allows us to presume that the spatial structures of our ODN duplexes are close to the canonical B-form.

Following conclusions, regarding short ODN, can be drawn from our study:

a. the influence of base composition more distant than the nearest neighbours cannot be neglected in many cases, some of which show general tendencies but some show more complicated sequence dependence;
b. the next-nearest neighbours can induce changes in the order of 0.1 ppm in duplex while they stay under 0.1 ppm in single strands, which is not properly reflected in the prediction methods [163, 173, 184];

c. many δ_{Exp} in terminal nucleosides strongly differ from δ_{Pred} which indicates increased malleability of the ODN ends;
d. although the differences between δ_{Exp} and δ_{Pred} often exceed the published deviations, the predictions can still serve as rough estimates, e.g. as a helpful tool in resolving assignment ambiguities, especially for non-exchangeable protons in internal nucleosides;
e. temperature dependence of chemical shifts, which is generally neglected in the available prediction methods, can be partially responsible for the discrepancies observed;
f. large differences in duplex ($|\Delta\delta| > 0.2$ ppm) can indicate local deviations from the B-from structure of the double-helix.

7.4 Temperature dependence of ^{1}H chemical shifts

Because the literature usually ignores the sensitivity of ^{1}H chemical shifts on temperature (outside the region of the temperature-induced unfolding), we have opened the following questions: Can we really justify neglecting the temperature dependence of chemical shifts? Or does it bring any new piece of information which could be helpful in the research on DNA structure and dynamics? Similarly as for the values of chemical shifts at a given temperature, our data provide insight into the sequence effects on their temperature dependence.

7.4.1 Linear fits of chemical shifts

Linear fits of the chemical shifts, cleaned from the deviating points, with respect to temperature were performed. The magnitudes of the slopes usually reach multiples of 0.001 ppm·K^{-1} for aromatic protons and are generally similar in duplex and in single strands. Steeper slopes than for other nuclei are observed for exchangeable protons (mainly CH42 and TH3 and in terminal nucleotides) but also for 3′-terminal CH5 in duplex.

Our slopes of exchangeable resonances roughly agree with the values mentioned alongside the NN prediction of chemical shifts [163]. However, all the other aromatic and deoxyribose shifts are generally regarded as constant in the cited article, although we found significant temperature

dependence. The work establishing the prediction protocols for random-coil chemical shifts [173] claims the slopes are less than 0.001 ppm·K^{-1}.

The slopes, a, display a broadly scattered distribution and the correlation among different occurrences of triplets is low – Fig. 7.14 and Fig. A.16–A.19 of § A.3 of Appendix (page A-17). The situation improves much in single strands compared to duplex but still remains unsatisfactory from the NN-model point of view. Nevertheless, some simple rules for shielding changes with increased temperature can be suggested:

a. all H6, H8, and exchangeable protons but 3′-terminal CH6 move upfield in duplex;
b. AH2, CH5, and TH7 move downfield in duplex;
c. different surrounding cause different directions of H1′, but CH1′ generally move downfield in duplex;
d. apart from GH8 and terminal CH6 showing incoherent directions, aromatic and anomeric protons move downfield in single strands.

The most striking exceptions to these points are presented by A2H8 and T7H6 of the $(GAACGTTC)_2$ duplex, both from the second base pair, going downfield with higher T ($a > 0$ in Fig. 7.14), and C5H5 of CTTGCAAG with $a < 0$. These cases might indicate some anomaly in the duplex properties. On the other hand, considering that there is no objective reason that $a = 0$ should be the relevant threshold and these singular cases are supported by no other spectral characteristics, accepting any conclusion would be premature.

The experimental chemical shifts of most ^{1}H in duplex are not strictly linear at the whole temperature range. At about 284 K and below[3], the shifts are declining to either direction from the trend followed above this point. This behaviour can be observed for all classes of protons analysed and it is more evident close to the ODN ends, especially for CH5. It might indicate some low-temperature structural transition of the double-helix, condensation of the duplexes into multi-molecular complexes, or changes in the interactions between water molecules or ions and the DNA.

[3]precise value is difficult to estimate and it varies between particular samples and nuclei

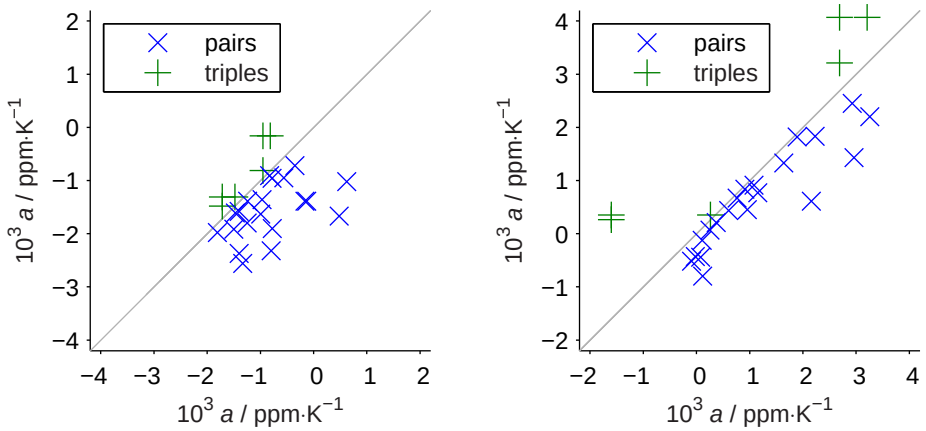


Figure 7.14 Correlation plots of chemical-shift slopes in the same trinucleotides for H6 and H8 protons. Left: duplex; right: single strands. The diagonals are indicated by grey lines. Two occurrences of a trinucleotide are denoted as pairs, three occurrences as triples

7.4.2 Sequence effect on the chemical-shifts slopes

In order to test the applicability of the NN model, the differences between the slopes, Δa, caused by the next-nearest neighbours are shown in Fig. 7.15 (duplex) and Fig. 7.16 (single strands) for internal and in Fig. 7.17 (duplex) and Fig. 7.18 (single strands) for terminal nucleotides in the same way as for the values of chemical shifts in § 7.3.1 on page 66.

The agreement between slopes obtained for the same protons in the same **internal triplets** from different occurrences in the data set is much worse than in the case of shifts at a given temperature (Fig. 7.6–7.9). Differences between the aromatic protons of the central nucleotide of a particular triplet commonly exceed the differences between hydrogen nuclei found in different triplets. In other words: the effect of bases further than one base away is comparable in magnitude with the influences of the nearest neighbours. The proximal termini affect all ^{1}H in duplex especially strongly.

Although some tendencies common for a set of triplets can be found, mainly in single strands, there are no simple rules which could be obtained from our limited data showing the effects of next-nearest neighbours on a. The nucleotide sequence determines the temperature dependence in much more complicated manner than what has been deduced

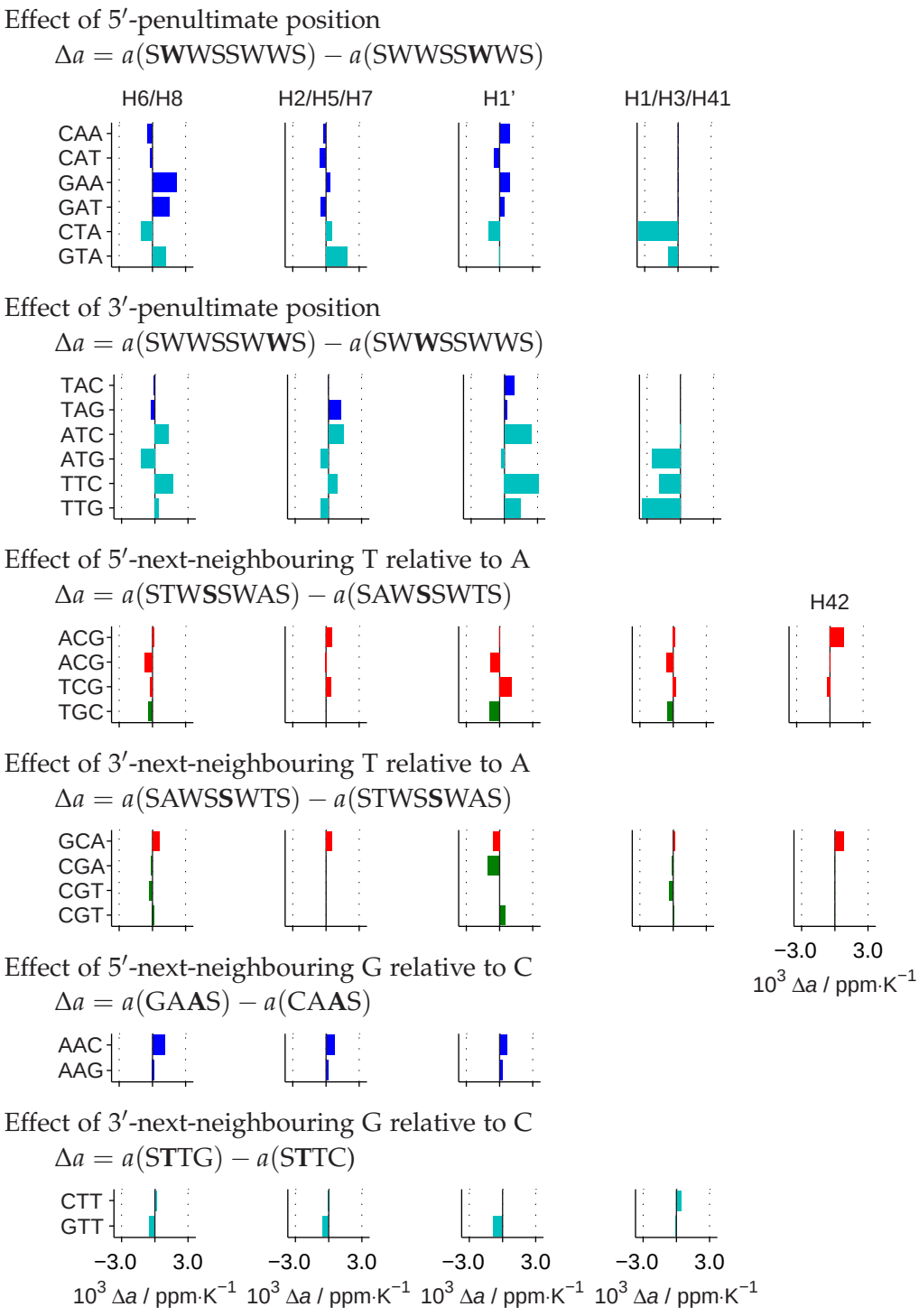


Figure 7.15 Next-nearest-neighbour effects on chemical-shift slopes in nucleotides in bold (duplex). ACG and CGT found three times in our ODN set. Bars colour-coded by the base type. W = A or T; S = C or G

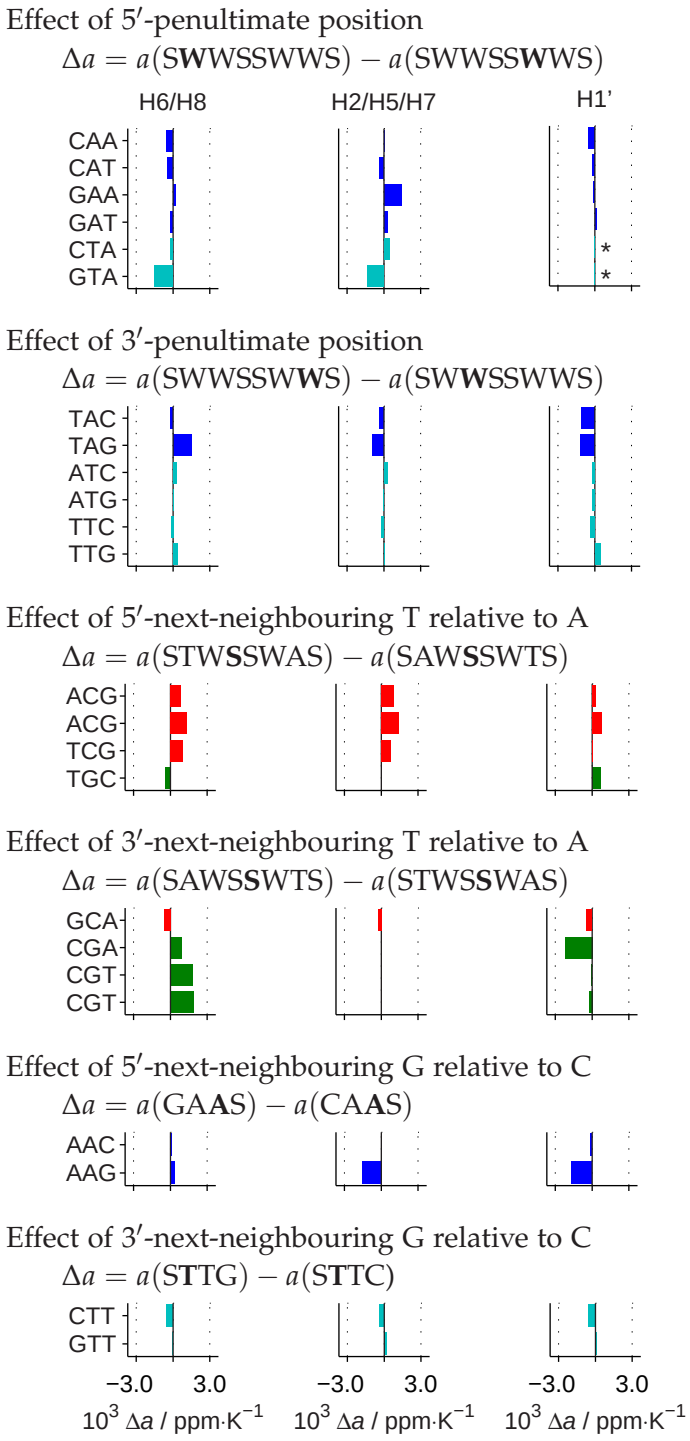


Figure 7.16 Next-nearest-neighbour effects on chemical-shift slopes in nucleotides in bold (single strands). ACG and CGT found three times in our ODN set. Bars colour-coded by the base type. W = A or T; S = C or G. Asterisks: missing values due to ambiguous assignment

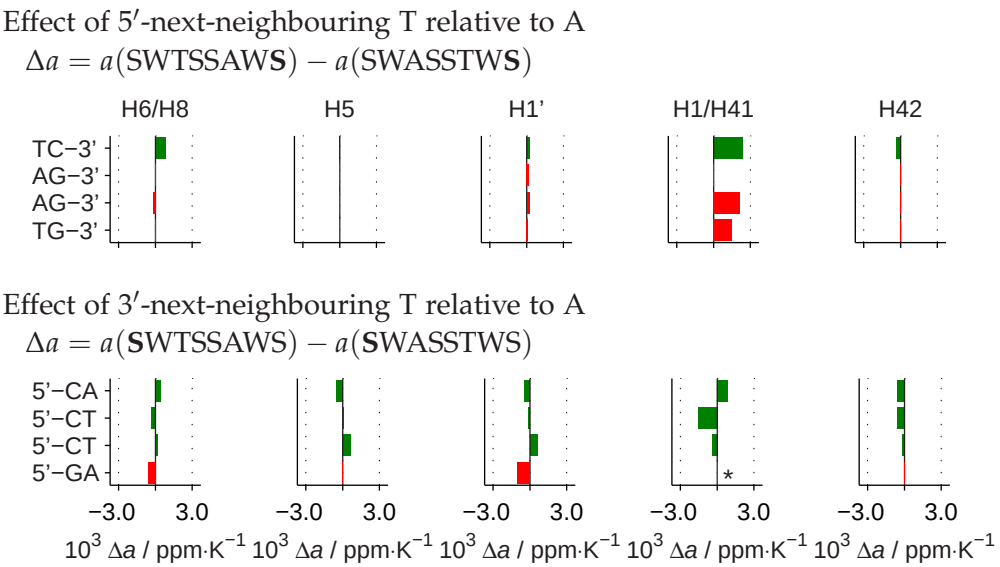


Figure 7.17 Next-nearest-neighbour effects on chemical-shift slopes in terminal nucleotides in bold (duplex). AG-3′ and 5′-CT found three times in our ODN set. Bars colour-coded by the base type. W = A or T; S = C or G. Asterisk: missing value due to an unobserved peak

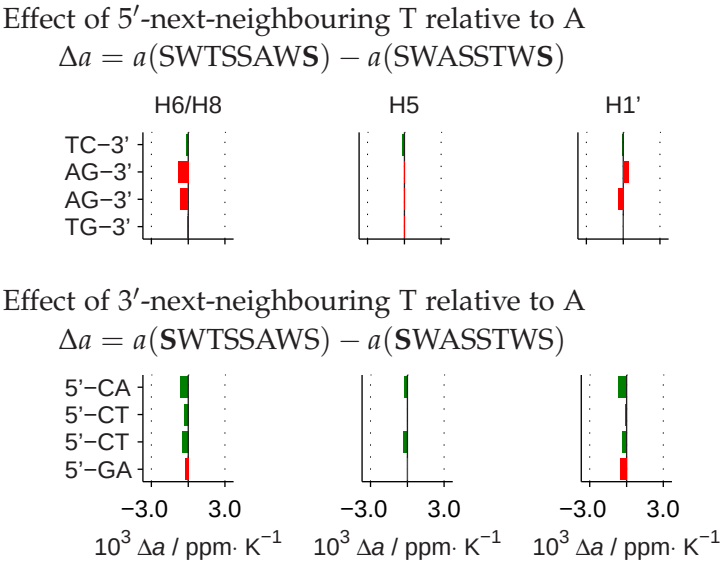


Figure 7.18 Next-nearest-neighbour effects on chemical-shift slopes in terminal nucleotides in bold (single strands). AG-3′ and 5′-CT found three times in our ODN set. Bars colour-coded by the base type. W = A or T; S = C or G

for values of δ at a single temperature. Based on our data we suggest that at least five nucleotides in a row should be used in any potential study attempting to reveal the base-context effects on the temperature dependence of δ.

The **ODN ends** provide a bit different picture: the slopes span quite large a range of values and extremely steep ones are found in the aromatics of 3′-terminal cytosine in duplex, even exceeding 0.008 ppm·K^{-1} for CH5 in the 3′-terminal TC. Although the absolute values of a in the terminal residues are larger compared to the internal ones, the differences between the same terminal doublets stay low, under 0.001 ppm·K^{-1} for all the non-exchangeable protons and CH42 (Fig. 7.17 and Fig. 7.18). Therefore, the absence of an adjacent nucleotide does much stronger an effect than a sole variation of the neighbouring base.

7.4.3 Sensitivity to subtle conformational variations

The sensitivity of δ to T is caused by several phenomena, which commit any attempt for quantitative analysis to become gravely non-elementary: the average geometry and strength of hydrogen bonds varies with respect to temperature [206, 207]; the stacking between bases can be affected, too. As a consequence, the base-pair parameters and also the positions of hydration water molecules change and that influences the electronic distribution in the DNA molecule.

We have seen a complicated dependence of the temperature-induced shielding changes on the DNA sequence above. It means that even a very subtle conformational variation can cause observable differences in a. They are found more sensitive to the base environment than the values of δ at a single temperature. We now summarise what have been revealed:

a. temperature change about 10 K causes differences in ^{1}H chemical shifts in the order of 0.01 ppm which is well observable and should not be neglected – a reference temperature at which the chemical-shift values are presented should be part of any publication;
b. non-linearities of chemical shifts with respect to temperature indicate some – yet unspecified – structural transitions occurring in duplex;
c. the NN model seriously fails to describe the sequence effects on

temperature dependence of chemical shifts in the internal nucleotides, although the chemical-shift slopes are surprisingly well conserved in equal terminal dinucleotides.

7.5 Concentration dependence of ^{1}H chemical shifts

Some of the ODN sequences have been measured by NMR spectroscopy at two or more different concentrations, c, retaining the same buffer composition (Table 7.1 on page 62). Various goals have been challenged by this approach: in two cases (CTACGTAG and GTAGCTAC), uncertain assignments of resonances in single strands have been resolved (Table 7.2 on page 65); in one case (CTTCGAAG), the possibility of a three-state process during duplex melting has been addressed.

In this section, the effect of ODN concentration on ^{1}H chemical shifts is analysed, which has not been thoroughly done before and the literature is silent on this subject.

7.5.1 CTACGTAG in duplex

Duplex ^{1}H spectra for two concentrations of CTACGTAG sequence are shown in Fig. 7.19. Clearly enough, all the peaks moved upfield after c

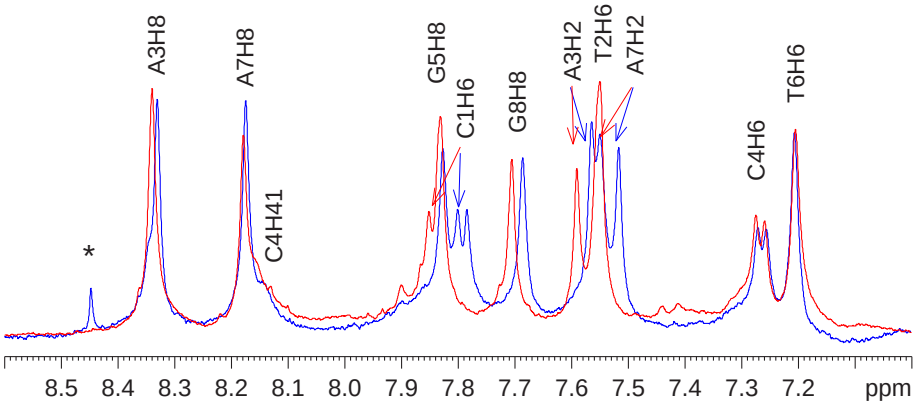


Figure 7.19 Aromatic parts of ^{1}H NMR spectra of CTACGTAG measured at two different concentrations in duplex (300.7 K). Red: 0.90 mM; blue: 4.98 mM. The asterisk, *, marks an impurity. The spectral intensities are scaled to approximately match each other

had been increased[4]. The uniform direction of shift is observed for all the aromatic, anomeric, and exchangeable protons. The resonances from the terminal residues and AH2 protons appear to be the most sensitive.

Only the data in duplex are available for the more concentrated sample – the experiments were performed only up to 330 K where the duplex is still not completely melted and the spectra cannot be interpreted as of pure single strands.

7.5.2 GTAGCTAC in duplex and single strands

Comparison of ^{1}H NMR spectra of GATGCTAC acquired at two different ODN concentrations (Fig. 7.20) reveals that the resonances move upfield in duplex and the differences grow toward the ODN termini, which agrees perfectly with the CTACGTAG sequence just discussed. On the other hand, downfield shifts of all peaks are observed after the concentration increase in single strands[5]. The amplitudes of the differences in δ are much smaller in random coil than in the double helix.

7.5.3 CTTCGAAG at four concentrations

Additionally to the basic sample, four different concentrations of CTTCGAAG in decreasing order were prepared (Table 7.1 on page 62) and their ^{1}H VT NMR spectra were acquired in the whole usual range.

This sample series was prepared from another ODN batch and another buffer solution, although all nominal parameters stayed the same. The NMR spectra were measured from a special Shigemi tube for small sample volumes, hence the temperature calibration may be slightly different from the basic sample. We tried to compare the chemical shifts of the basic sample with this four-membered series, but it was found to decline slightly from the trend observed in the dilution series, although it follows the same qualitative rules which are common also to the two ODN sequences described above. This illustrates how sensitive the NMR spectroscopy is: subtle changes in the solvent composition caused just by ordinary imprecisions in handling with chemicals or placement in an-

[4] with the only exception presented by T6H6; however, the difference is only 1.0 Hz

[5] only two tiny exceptions were found in this case, namely C5H42 which moves a little downfield in duplex (by 1.2 Hz) and A7H8 which moves a little upfield in single strand (by 0.4 Hz)

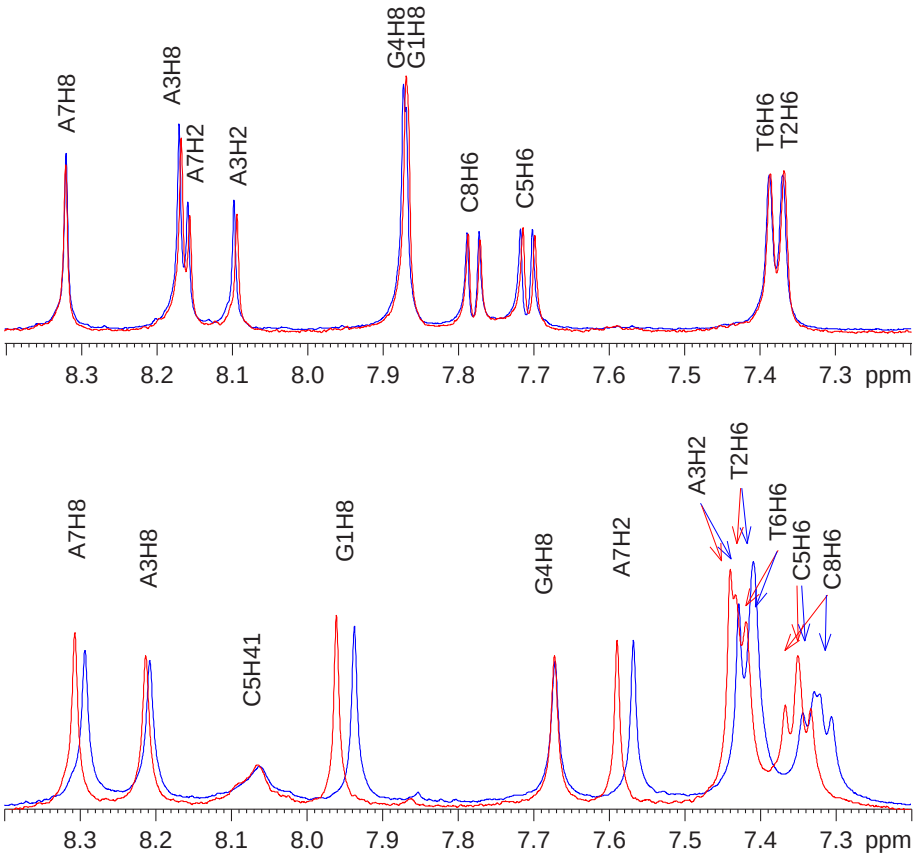


Figure 7.20 Aromatic parts of ^{1}H NMR spectra of GTAGCTAC measured at two different concentrations. Red: 0.95 mM; blue: 4.05 mM. Top: single strands (354.3 K); bottom: duplex (296.4 K). The spectral intensities are scaled to approximately match each other

other tube can measurably influence the chemical shifts. Only the four-membered series is thus discussed below.

The chemical shifts at all temperatures obtained by Lorentzian curve fitting are graphically presented in Fig. 7.21 for a selected nucleus and in § A.4 of Appendix (page A-26) for all ^{1}H. The values of δ at one T in duplex and at another one in single strands are extracted as functions of c in Fig. 7.22, Fig. 7.23, and Fig. 7.24.

We can directly see that completely all the protons[6] move upfield with increasing c in duplex. On the other hand, downfield change of all the

[6]we excluded the deoxyribose protons H2′ and H2″ from this analysis because the NOESY spectra were not acquired for all the concentrations without which the assign-

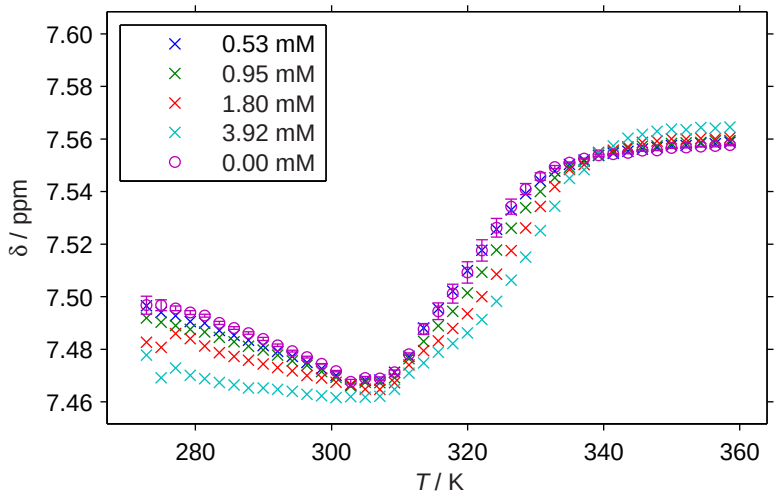


Figure 7.21 Chemical shifts of T3H6 of CTTCGAAG in the whole temperature range measured (crosses) and extrapolations to $c = 0$ (circles with errorbars)

single-strand shifts is observed, although the effect is not that strong. The apparent linearity of $\delta(c)$ both in duplex and in single strands allows extrapolation to zero concentration which could serve as a hypothetical reference[7]. These values from linear regressions of $\delta(c)$ are included in Fig. 7.21 and § A.4 of Appendix (page A-26).

One could argue that all the observed effects are quite weak, but the chemical-shift differences well exceed the accuracy of their determination because they are at least comparable to the line widths. Moreover, the uniformity of change direction is striking.

7.5.4 Higher-order structures and salt effect

The previous paragraphs have shown that increased ODN concentration causes upfield shift of all aromatic, H1′, and exchangeable protons in duplex and downfield shift of all resonances in single strands. Among the various effects on chemical shifts discussed (temperature in § 7.4 on page 81, melting in § 7.3.3 on page 79, and methylation in § 8.3 on

ment would not be reliable; nevertheless, visual comparison of the spectra reveals that the upfield shift of all the saccharide resonances is very likely

[7] in reality, there would be no duplexes in an infinitely diluted sample, but this reference could serve as a model for non-interacting duplexes – an ideal-gas approximation

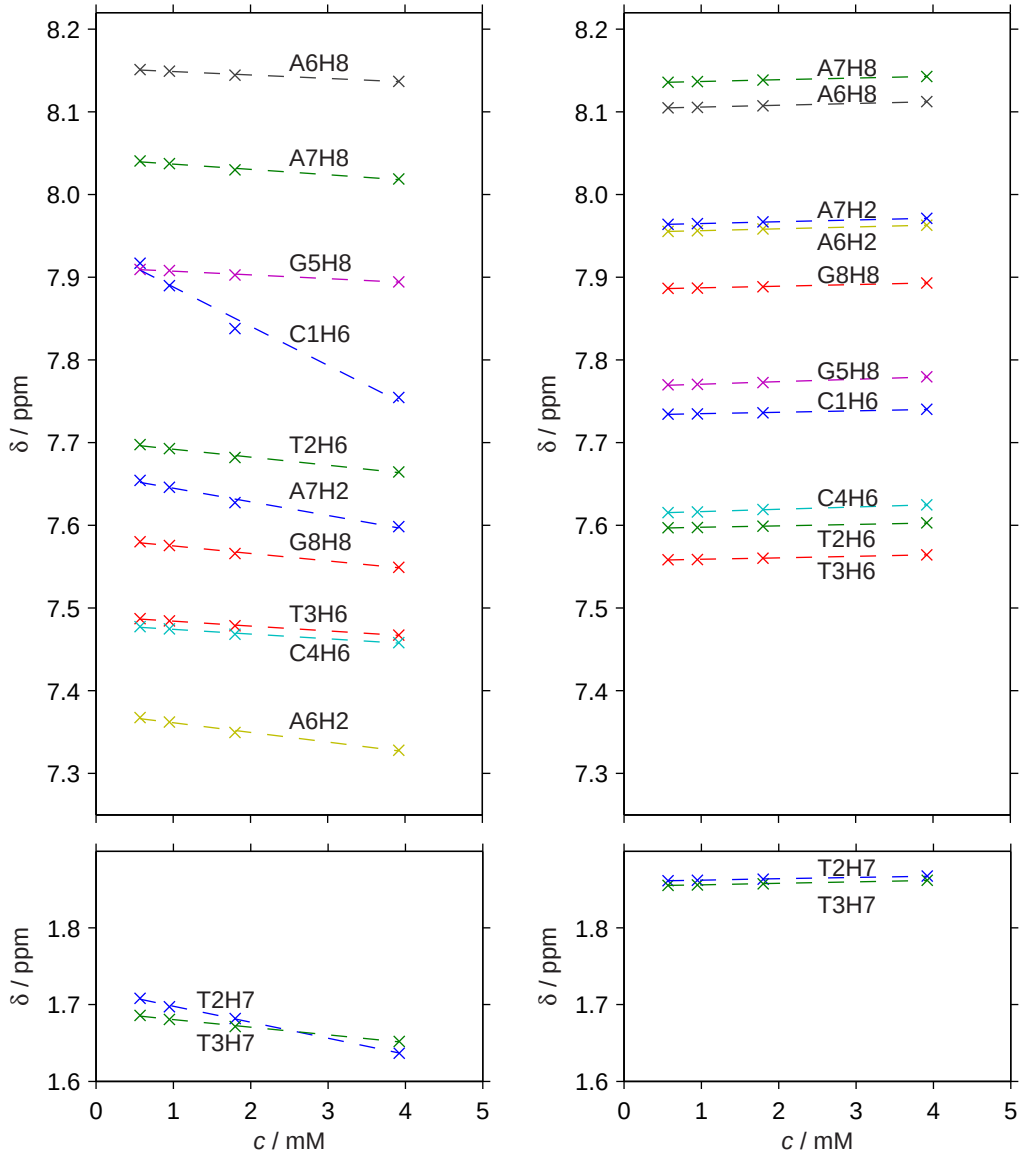


Figure 7.22 Concentration dependence of ^{1}H chemical shifts of H6, H8, and H2 (top) and TH7 (bottom) of CTTCGAAG. Left: duplex, 283.5 K; right: single strands, 354.3 K. Dashed lines are linear fits

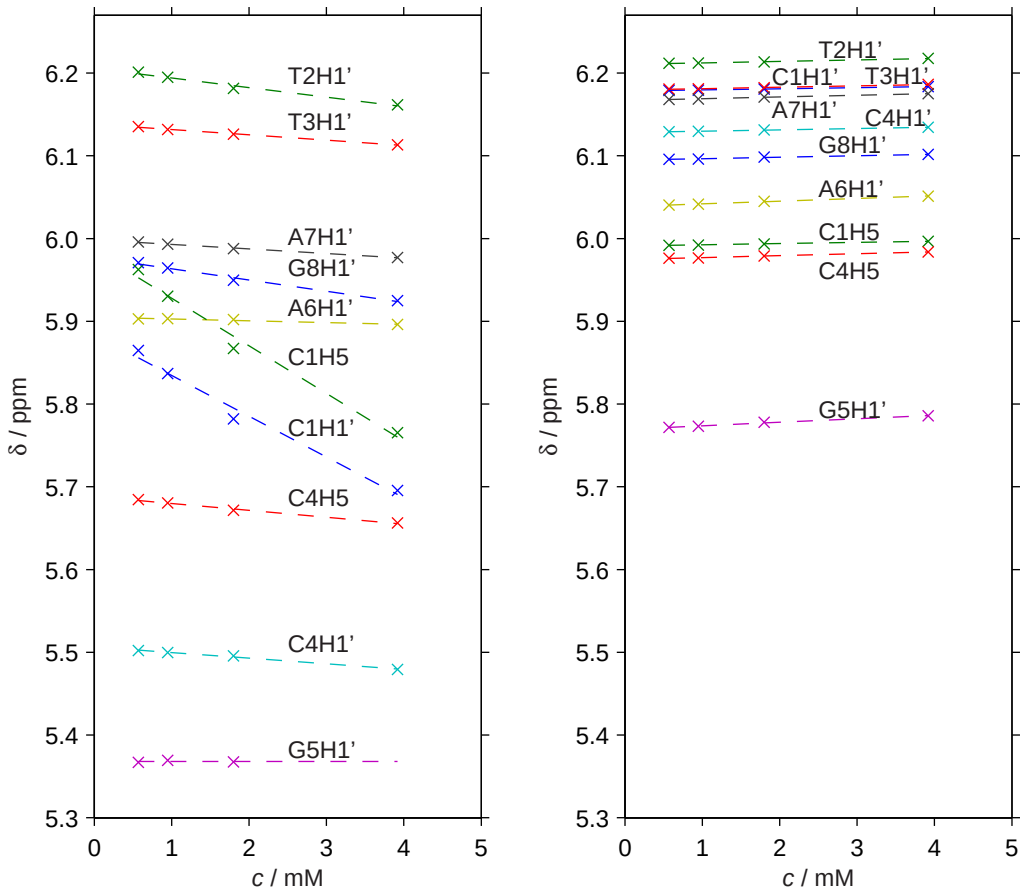


Figure 7.23 Concentration dependence of ^{1}H chemical shifts of CH5 and H1′ of CTTCGAAG. Left: duplex, 283.5 K; right: single strands, 354.3 K. Dashed lines are linear fits. Value of G5H1′ missing at the highest concentration in duplex due to overlap with residual water signal. The scales are the same as in Fig. 7.22

page 125), this is the only case of a concerted influence on all hydrogen nuclei in distinct chemical moieties.

How can this behaviour be explained? One possible answer arises immediately: the increased c increases the duplex stability, therefore δ would be more 'duplex-like' at all temperatures. However, this option cannot hold because there are both directions of the chemical-shift differences induced by melting present for different kind of protons (Fig. 7.13 on page 80), whereas the concentration affects all the hydrogen nuclei uniformly. Any change in macroscopic properties of the solution, mainly its

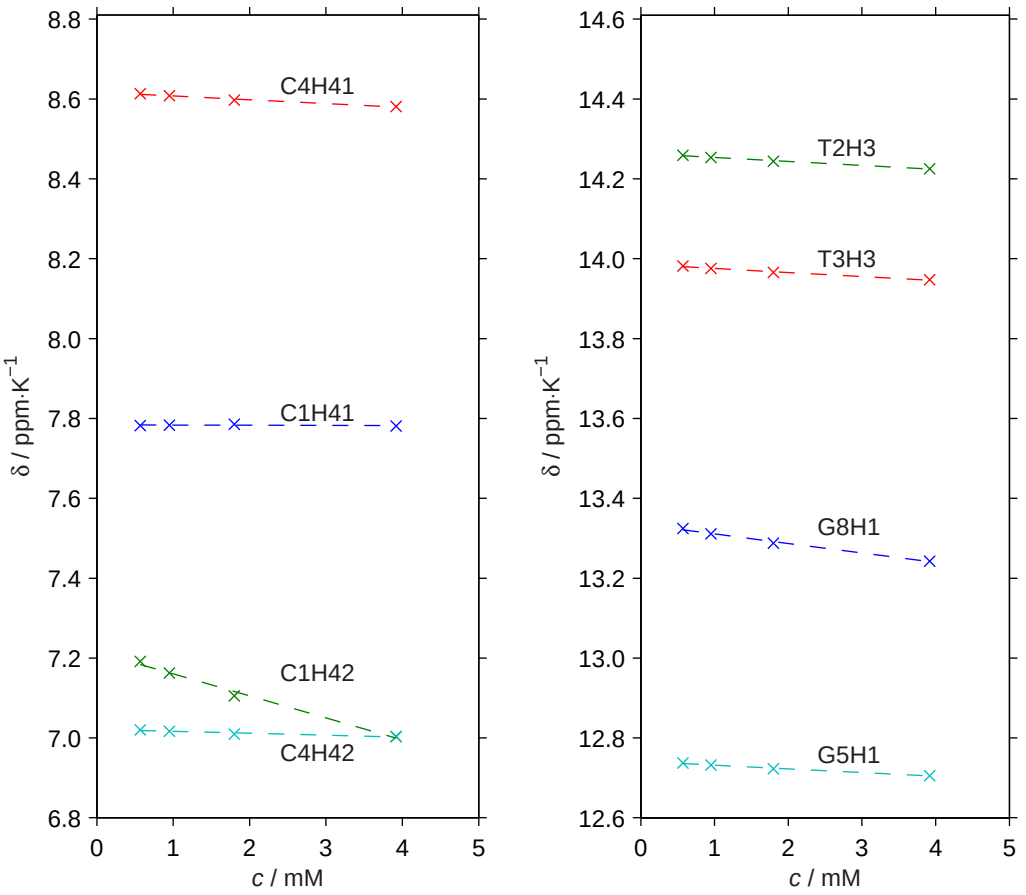


Figure 7.24 Concentration dependence of ^{1}H chemical shifts of exchangeable protons of CTTCGAAG in duplex (283.5 K). Left: amino, right: imino protons. Dashed lines are linear fits. Different vertical scales than in figures Fig. 7.22 and Fig. 7.23

magnetic susceptibility, can be discarded, too, because it would affect the resonances of DSS and water in the same way as those of the DNA which we do not observe.

The change in c modifies the molar ratio between ODN and ions. The acidity slightly changed in the CTTCGAAG concentration series: estimated from ^{31}P chemical shift of the inorganic phosphate [208, 209, 210], pH gradually increased with decreasing c from 6.6 to 7.0. Therefore, in the search for the reasons of concentration-induced changes of chemical shifts, additional experiments with varying the solvent composition were also conducted.

The following points have been revealed:

a. the sensitivity of δ of H6, H8, and H1′ on c in duplex increases towards the ODN ends (Fig. 7.25);
b. the chemical shifts in duplex are the most sensitive to c at lower temperatures – Fig. 7.21 and § A.4 of Appendix (page A-26);
c. increase in NaCl level from 80 mM to 250 mM shifts all ^{1}H peaks in the same direction and with magnitudes comparable to the changes induced by increased c in single strands but not in duplex;
d. acidity of the solution has a negligible effect (difference 0.004 ppm between pH 6.0 and 9.5 measured during titrations by NaOH and HCl without any buffer).

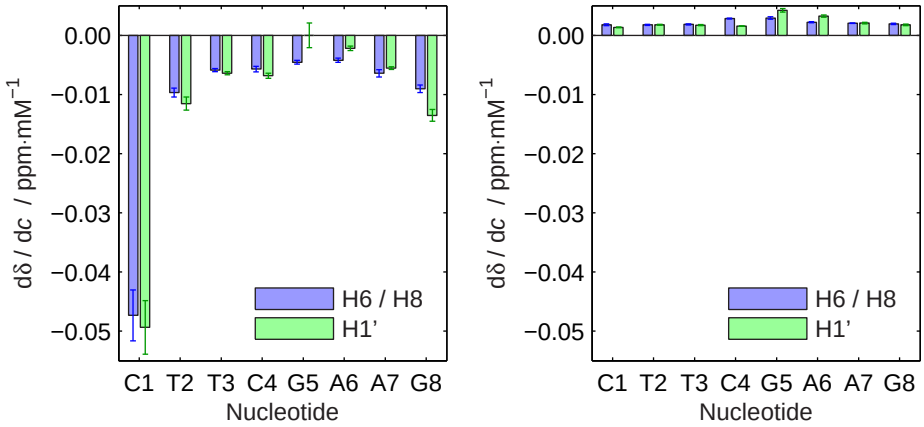


Figure 7.25 Slopes of chemical shifts of H6, H8, and H1′ in CTTC-GAAG with respect to c. Left: duplex, 283.5 K; right: single strands, 354.3 K

These pieces of evidence indicate that some multi-molecular complexes, possibly stacked duplexes, are formed at lower temperatures which agrees with the heavy discrepancies between experimental and predicted chemical shifts of terminal nucleosides (§ 7.3.2 on page 75) and with the non-linearities of $\delta(T)$ in duplex (§ 7.4.1 on page 82). In the literature, this phenomenon is described only for much higher DNA concentrations [211, 212] and in the presence of divalent cations [213]. On the other hand, a favourable free energy of the end-to-end stacking was estimated by MD [214] which would allow the presence of such complexes even at our conditions.

Well above T_m, where no complexes are expected, the main reason of the changes of chemical shifts with c lies probably in the ionic strength of the solution relative to the ODN concentration.

7.6 Duplex melting observed by ^{1}H NMR

In this section, the results obtained from the line-shape analysis of ^{1}H VT NMR experiments are presented and discussed.

The least-square fitting of exchange or Lorentzian line shapes according to § 6.2 on page 56 to the ^{1}H spectra in the aromatic, methyl, and anomeric regions was performed at all temperatures measured (examples are shown in Fig. 7.26). For every resonance, the melting curves and, if applicable, the kinetic data were extracted. Examples of the temperature dependence of the duplex fraction, p_A, and of the dissociation rate constant, k_A, are shown in Fig. 7.27 and Fig. 7.28, respectively. Exceptions from the general approach were encountered; these cases are not considered in the following treatments.

The results of the thermodynamic analyses by the van 't Hoff and Eyring models are presented in the tables of § A.5 of Appendix (page A-32). The global fits were conducted according to § 6.3.2 on page 59 – their results are in Table 7.6 and Table 7.7.

The NMR-determined T_m^{glob} and also ΔH^{glob} are in a good agreement with literature (Table 2.1 on page 22). Also, our data concerning two sequences studied earlier [117, 106], CATCGATG and CTTCGAAG, reproduce the lower stability of the latter.

7.6.1 Sequence effects on global characteristics

The eight ODN sequences can be grouped according to their stabilities obtained from Ha (Table 7.6):

a. CAACGTTG and CATCGATG with $T_m^{glob} \approx 324$ K;
b. CTTGCAAG, GAACGTTC, and GATGCATC with $T_m^{glob} \approx 321$ K;
c. CTTCGAAG and GTAGCTAC with T_m^{glob} around 319 K and 320 K;
d. CTACGTAG with $T_m^{glob} \approx 314$ K.

This classification is valid in the terms of both T_m^{glob} and $\Delta G_{310\,K}^{glob}$. On the other hand, it is not clear from the underlying ΔH^{glob} and ΔS^{glob} except

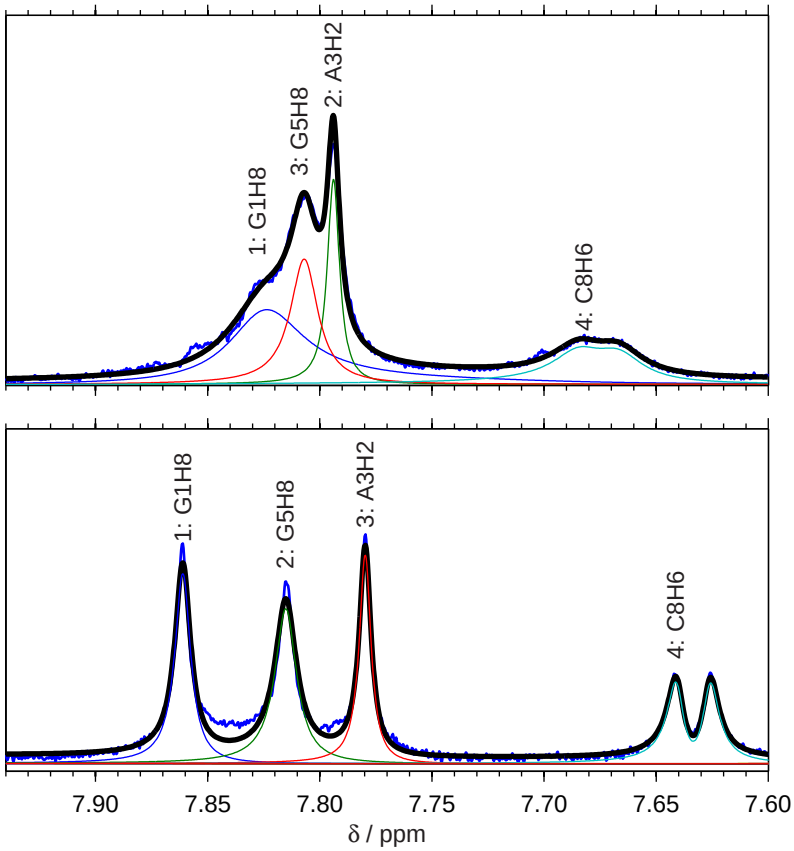


Figure 7.26 A part of ^{1}H NMR spectrum of GAACGTTC (edited outputs of the plotting method of Asymexfit class `NMRspectra` with arbitrary vertical scales). Top: 309.3 K (melting, exchange line shapes); bottom: 298.5 K (duplex, Lorentzian curves). Blue: experimental spectrum; black: fit; other colours: fit decomposition into resonances of individual protons, labelled accordingly

for the least stable CTACGTAG having their lowest absolute values. However, the errors in ΔH^{glob} and ΔS^{glob} are underestimated (for instance, we do not handle the error in temperature determination and neglect the heat-capacity change, ΔC_{p}).

Among these four classes, the CA dinucleotide at the 5′ end appears to have the most stabilising effect. Interestingly, there is a very big difference in the impact of the reversal of an A·T base pair in CAACGTTG: while the A–T substitution at position 3 hardly has any influence, the same change in the second nucleotide decreases $T_{\mathrm{m}}^{\mathrm{glob}}$ by astonishing 10 K.

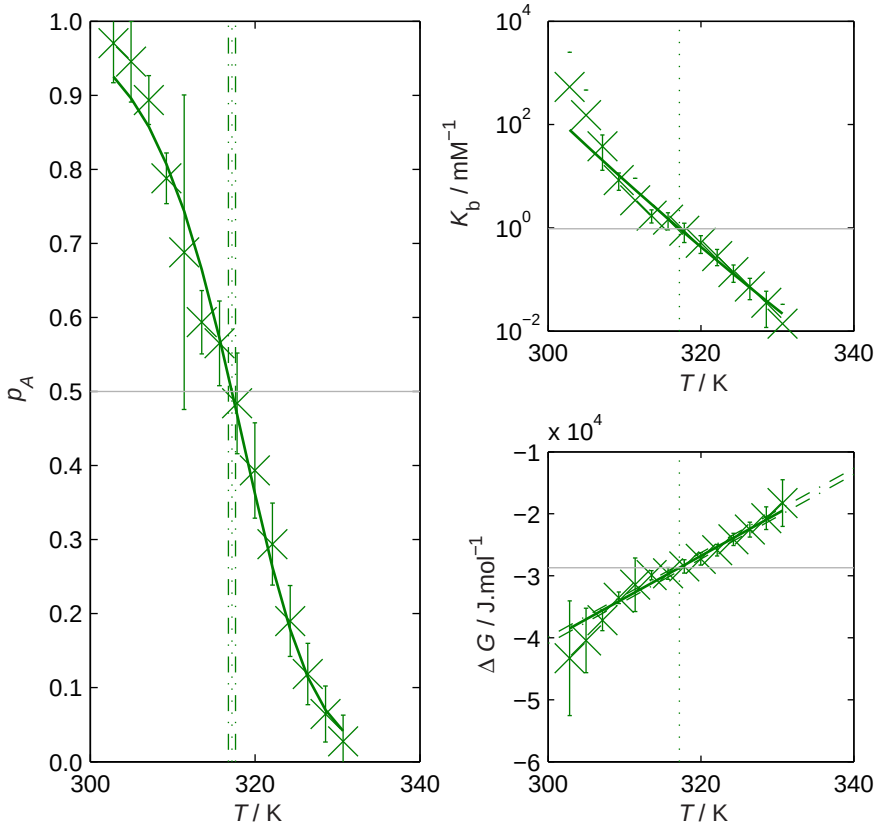


Figure 7.27 Melting curve of T2H6 in GATGCTAC (modified output of the plotting method of Asymexfit class `populationData`). Left: duplex fraction, p_A; top right: logarithmic plot of the association equilibrium constant, K_b; bottom right: Gibbs free energy of association, ΔG. Crosses and error bars: results from spectral fits; in the logarithmic scale, only the upper edges of the error bars are shown for relative errors exceeding 100 %. Solid curves: van't Hoff fit; dash–dot lines in the bottom-right panel show the confidence limits of $\Delta G_{310\,\mathrm{K}}$. Vertical dotted lines mark T_m and the dash–dot vertical lines in the left panel correspond to its confidence limits. Grey horizontal lines indicate the values at T_m

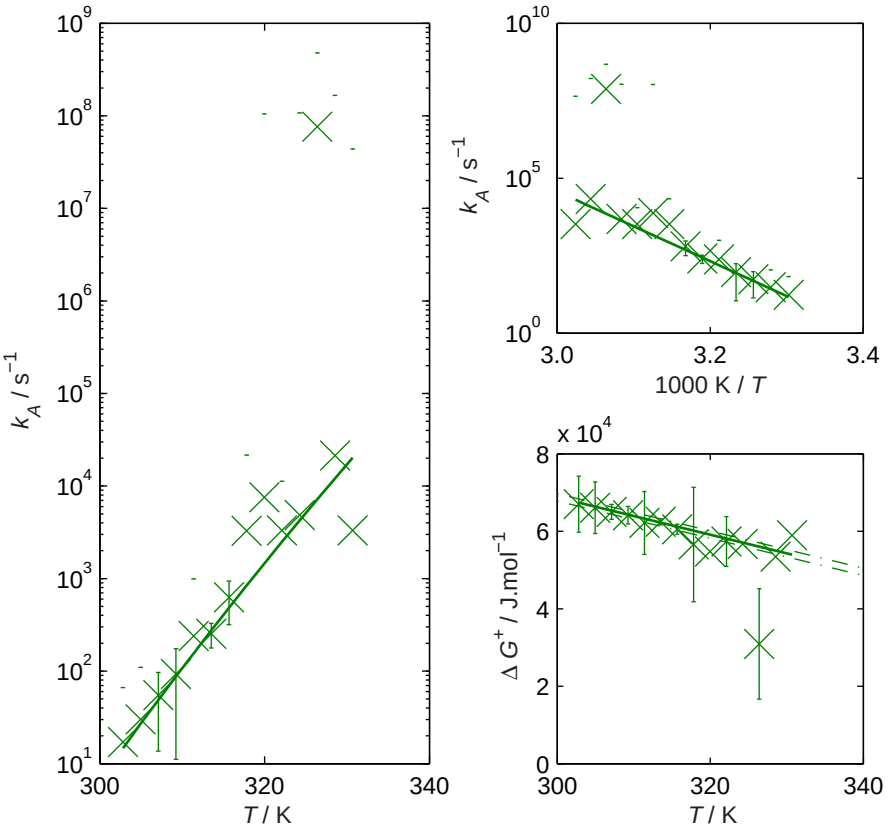


Figure 7.28 Exchange rates of T2H6 in GATGCTAC (modified output of the plotting method of Asymexfit class `EyringData`). Left: logarithmic plot of the rate constant of duplex dissociation, k_A; top right: Eyring plot of k_A; bottom right: Gibbs free energy of activation, $\Delta G^\dagger$. Crosses and error bars: results from spectral fits; in the logarithmic scales, only the upper edges of the error bars are shown for relative errors exceeding 100 %. Solid curves: Eyring fit; dash–dot lines in the bottom-right panel show the confidence limits of $\Delta G^\dagger_{310\,\mathrm{K}}$

No other simple rule can be extracted. There is no general trend seen upon the sequence inversion encountered in the couples CATCGATG–GTAGCTAC, CTACGTAG–GATGCATC, and CTTGCAAG–GAACGTTC, but this is not surprising as it only stresses the asymmetry of the NA chain. The members of ODN pairs having their first and second halves swapped, CATCGATG–GATGCATC and CTACGTAG–GTAGCTAC, also fall into different stability categories.

Table 7.6 Results of global fitting of $p_A(T)$ for Ha (H2, H6, H7, and H8) and $\delta(T)$ for H1′ protons. T_m^{glob} are calculated for 1 mM ODN concentration. The errors presented are coming purely from the fits

	$\frac{T_m^{glob}}{K}$	$\frac{\Delta G_{310\,K}^{glob}}{kJ \cdot mol^{-1}}$	$\frac{\Delta H^{glob}}{kJ \cdot mol^{-1}}$	$\frac{\Delta S^{glob}}{J \cdot mol^{-1} \cdot K^{-1}}$
CAACGTTG				
Ha	324.0 ± 0.1	-28.7 ± 0.2	-252 ± 4	-722 ± 12
H1′	324.7 ± 0.1	-29.7 ± 0.3	-264 ± 5	-755 ± 16
CATCGATG				
Ha	323.5 ± 0.1	-28.7 ± 0.1	-261 ± 2	-748 ± 7
H1′	324.1 ± 0.1	-29.4 ± 0.2	-268 ± 4	-770 ± 13
CTACGTAG				
Ha	314.5 ± 0.1	-20.6 ± 0.1	-195 ± 2	-564 ± 8
H1′	318.6 ± 0.2	-24.8 ± 0.2	-257 ± 5	-749 ± 15
CTTCGAAG				
Ha	318.7 ± 0.1	-24.6 ± 0.1	-252 ± 4	-732 ± 12
H1′	320.7 ± 0.2	-27.6 ± 0.4	-294 ± 7	-858 ± 22
CTTGCAAG				
Ha	321.3 ± 0.2	-26.1 ± 0.2	-236 ± 5	-677 ± 16
H1′	322.6 ± 0.1	-27.6 ± 0.2	-252 ± 4	-724 ± 13
GAACGTTC				
Ha	321.6 ± 0.1	-27.1 ± 0.1	-259 ± 3	-748 ± 9
H1′	321.6 ± 0.2	-27.4 ± 0.3	-266 ± 5	-769 ± 16
GATGCATC				
Ha	321.9 ± 0.1	-26.7 ± 0.2	-242 ± 5	-693 ± 16
H1′	321.7 ± 0.1	-26.8 ± 0.2	-248 ± 4	-714 ± 13
GTAGCTAC				
Ha	320.1 ± 0.1	-25.5 ± 0.1	-244 ± 3	-706 ± 10
H1′	321.7 ± 0.2	-27.9 ± 0.4	-277 ± 8	-804 ± 26

There are significant and uniform differences between the parameters obtained from Ha and H1′: T_m^{glob} and $\left|\Delta G_{310\,K}^{glob}\right|$ are always higher in the case of H1′. Based on tests comparing the results from exchange fitting of Ha with their Lorentzian fits (see also § 7.7.1 on page 112), it cannot be ascribed to a systematic error connected to the different methods of analysis of the spectra of Ha and H1′. Therefore, some real process is responsible; perhaps some local structural transitions accompanying the duplex melting.

Table 7.7 Results of global fitting of $k_A(T)$ of all relevant H2, H6, H7, and H8 protons. The errors presented are coming purely from the fits

	$\Delta G^{\dagger,\mathrm{glob}}_{310\,\mathrm{K}}$ / kJ · mol^{-1}	$\Delta H^{\dagger,\mathrm{glob}}$ / kJ · mol^{-1}	$\Delta S^{\dagger,\mathrm{glob}}$ / J · mol^{-1} · K^{-1}
CAACGTTG	66.3 ± 0.1	210 ± 2	465 ± 8
CATCGATG	66.5 ± 0.1	242 ± 2	566 ± 7
CTACGTAG	63.2 ± 0.1	193 ± 3	420 ± 10
CTTCGAAG	66.4 ± 0.1	210 ± 5	465 ± 18
CTTGCAAG	66.4 ± 0.1	247 ± 5	583 ± 14
GAACGTTC	65.4 ± 0.1	250 ± 6	596 ± 18
GATGCATC	65.6 ± 0.1	243 ± 4	572 ± 14
GTAGCTAC	64.7 ± 0.1	233 ± 2	544 ± 8

Looking at the dissociation activation energy of the least stable ODN, CTACGTAG (Table 7.7) reveals that its $\Delta G^{\dagger,\mathrm{glob}}_{310\,\mathrm{K}}$ is also significantly smaller but the difference in $\Delta G^{\dagger,\mathrm{glob}}_{310\,\mathrm{K}}$ of this ODN compared to other sequences is less than in $\Delta G^{\mathrm{glob}}_{310\,\mathrm{K}}$. Therefore, relative to single strands, the free energy at 310 K of the $(\mathrm{CTACGTAG})_2$ duplex as well as of the transition state is increased. No such trend can be observed for the other seven duplexes – the correlation between $\Delta G^{\mathrm{glob}}_{310\,\mathrm{K}}$ and $\Delta G^{\dagger,\mathrm{glob}}_{310\,\mathrm{K}}$ is very poor. This contrasts with the strong correlation found in a set of octamer DNA, RNA, and hybrid duplexes [215]. In addition, our $\Delta G^{\dagger,\mathrm{glob}}_{310\,\mathrm{K}}$ are lower than determined by stopped-flow spectroscopy although $\Delta G^{\mathrm{glob}}_{310\,\mathrm{K}}$ are comparable [215], but it might be connected to the different buffer compositions.

Next, we compare T_m obtained from the p_A melting curves with the nearest-neighbour predictions (Fig. 7.29). Although some overall similarities are present, the NN approaches do not reproduce the experimentally determined order of stabilities, neither the categorisation into the four classes. Therefore, there must be some longer-range effects underlying the DNA duplex thermodynamics.

7.6.2 Sequence dependence of the labile-protons exchange

The integral intensities of amino and imino resonances, calculated as the ratios between fitted values of the peak maxima and T_2^*, decrease at gradually lower temperatures coming closer to the ODN ends (Fig. 7.30).

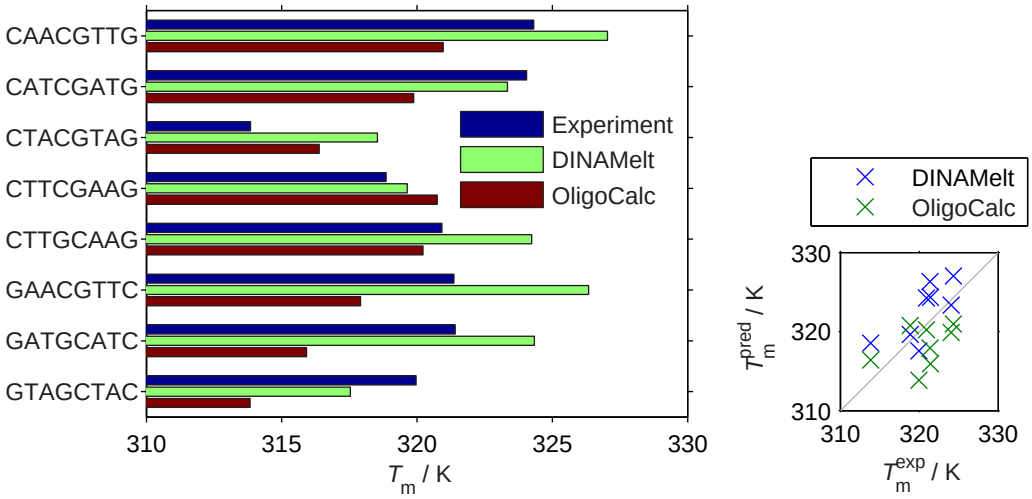


Figure 7.29 Experimental T_m coming from the global NMR analysis of Ha and predictions based on the NN model by DINAMelt [52] and OligoCalc [55]. Right: correlation plot between experimental, T_m^{exp}, and predicted values, T_m^{pred}, with the diagonal drawn in grey. The correlation coefficients are 0.72 and 0.41 for DINAMelt and OligoCalc, respectively

Therefore, their exchange with water is faster. There are two typical explanations that are not mutually exclusive:

a. the solvent accessibility increases;
b. the base-pair opening rate increases.

Based only on the peak intensities, it cannot be discriminated between them. There is no simple connection between the intensities of the labile protons and the duplex melting since many aspects influence the exchange rate. Chiefly, there are more possible ways how the exchange happens: the water molecule can attain the base protons from the ODN-end face or from one of the grooves; a flipped-out base can facilitate the reaction. Detailed investigation of the exchange mechanisms involving many steps requires specially devoted experiments involving titrations by a catalyst [216, 217].

The data in Fig. 7.30 are not corrected for the temperature dependence of the transverse relaxation times. This can be seen from the initial increases of the integral intensities at low temperatures, caused by faster relaxations at lower T and by the spin echo used for acquisition.

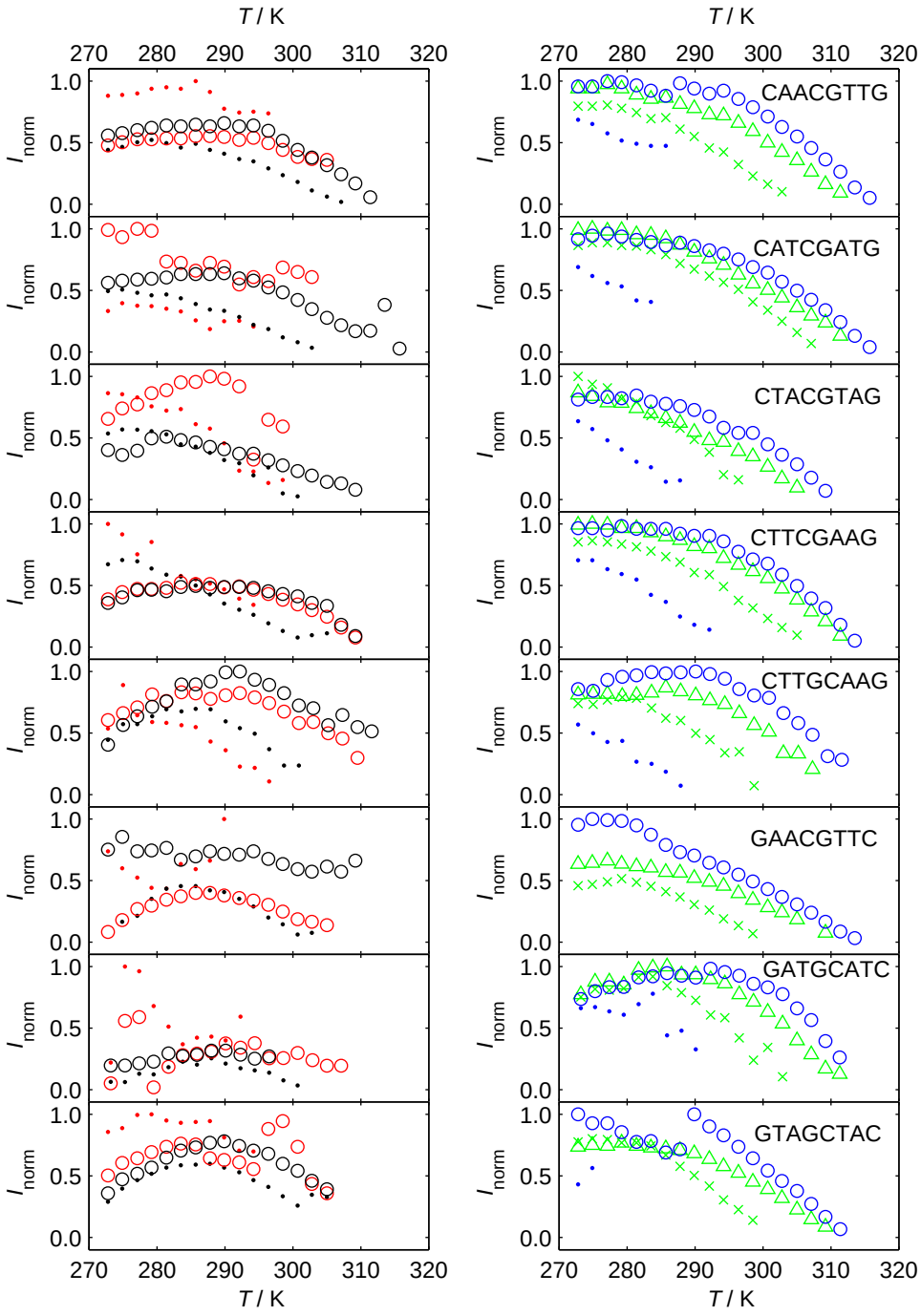


Figure 7.30 Integral intensities of amino (left) and imino (right) protons normalised to the maximal value in each plot, I_{norm}. ODN sequences in the right column. Dots, crosses, triangles, and circles used for protons in the first, second, third, and fourth base pair, respectively. CH41 in red, CH42 in black, GH1 in blue, and TH3 in green

7.6.3 Local differences from global properties

One of the great capabilities of high-resolution NMR is that it can study individual resonances belonging to different groups of equivalent nuclei. Application of this experimental technique to the DNA-duplex melting promises detection of local differences of its stability and flexibility. Once the analysis of all the acquired spectra is done, what can we deduce from the results?

The differences between T_m, $\Delta G_{310\,\mathrm{K}}$, and $\Delta G^{\dagger}_{310\,\mathrm{K}}$ obtained from individual spectral lines and the results from the global analysis of NMR data, $T_\mathrm{m}^{\mathrm{glob}}$, $\Delta G^{\mathrm{glob}}_{310\,\mathrm{K}}$, and $\Delta G^{\dagger,\mathrm{glob}}_{310\,\mathrm{K}}$, are shown in Fig. 7.31, Fig. 7.32, and Fig. 7.33. There are no general trends that would be common for all the ODN sequences. Above all, the central regions of the duplexes do not appear more stable than the outer parts. The activation energies stay rel-

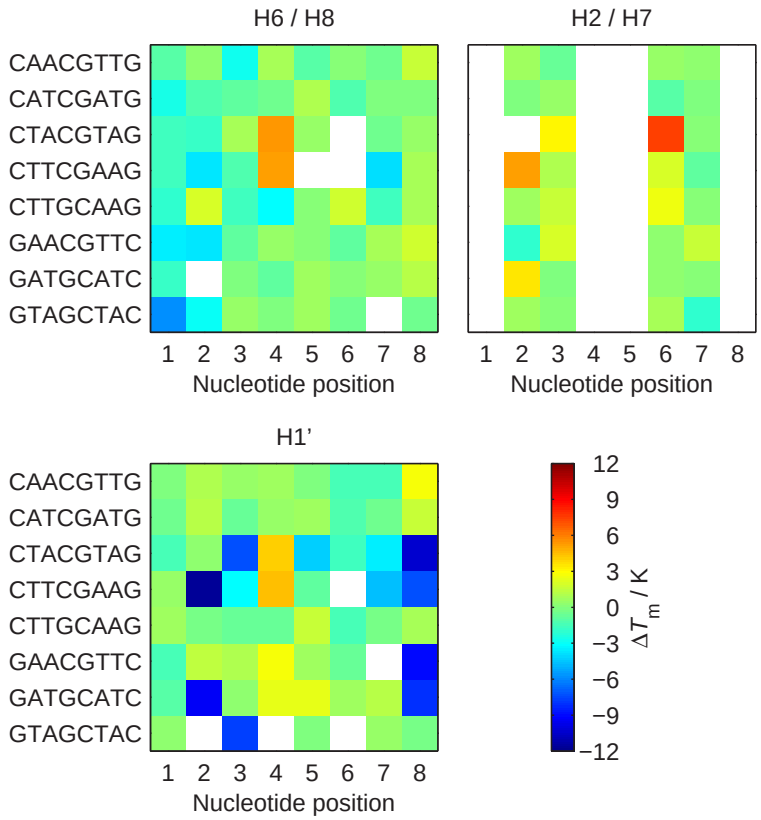


Figure 7.31 Differences between T_m obtained from independent and global fits, $\Delta T_\mathrm{m} = T_\mathrm{m} - T_\mathrm{m}^{\mathrm{glob}}$

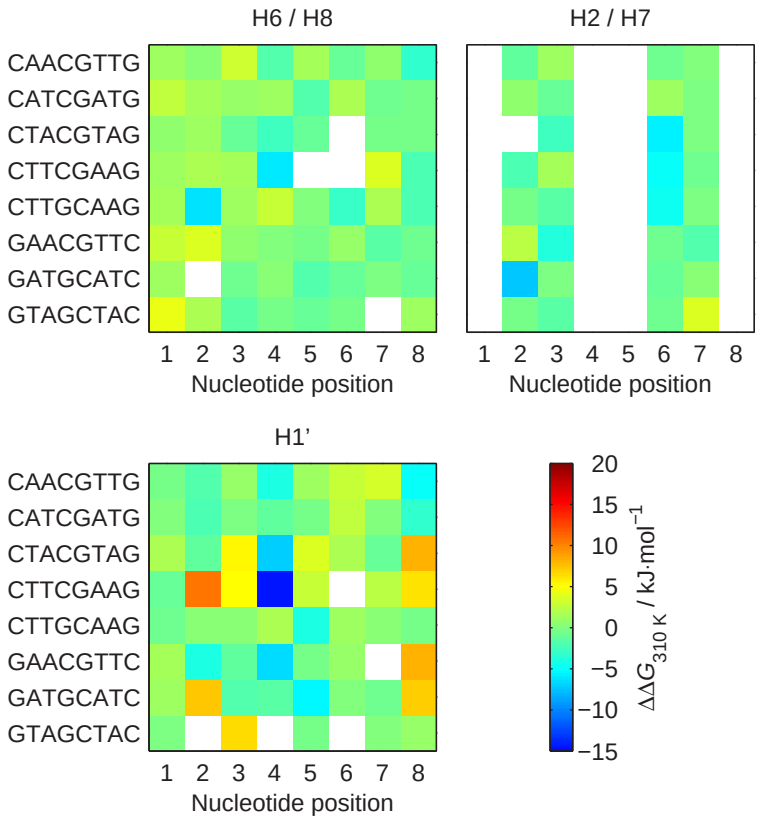


Figure 7.32 Differences between $\Delta G_{310\,\mathrm{K}}$ obtained from independent and global fits, $\Delta\Delta G_{310\,\mathrm{K}} = \Delta G_{310\,\mathrm{K}} - \Delta G^{\mathrm{glob}}_{310\,\mathrm{K}}$

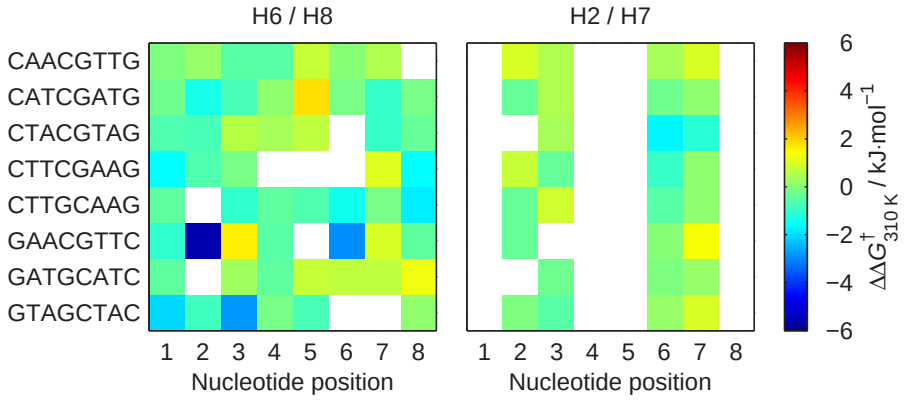


Figure 7.33 Differences between $\Delta G^{\dagger}_{310\,\mathrm{K}}$ obtained from independent and global fits, $\Delta\Delta G^{\dagger}_{310\,\mathrm{K}} = \Delta G^{\dagger}_{310\,\mathrm{K}} - \Delta G^{\dagger,\mathrm{glob}}_{310\,\mathrm{K}}$

atively well conserved within every duplex, too. Therefore, we conclude that the melting of the eight-base-pair double helices occurs in a very cooperative way.

On the other hand, there are several related deviations that should not be overlooked:

a. lower T_m and higher $\Delta G_{310\,K}$ for the 5′-terminal base protons H6 and H8;
b. lower T_m and higher $\Delta G_{310\,K}$ for the 3′-terminal protons H1′;
c. higher T_m and lower $\Delta G_{310\,K}$ for C4H6 and C4H1′ in CTACGTAG and CTTCGAAG compared to the other nucleotides in these two ODN sequences;
d. several singular deviations, in particular, T_m and $\Delta G_{310\,K}$ of T6H7 in CTACGTAG and T2H1′ in CTTCGAAG and $\Delta G^{\ddagger}_{310\,K}$ of A2H8 in GAACGTTC.

7.6.4 Fraying?

The apparently lower stabilities of the duplex ends are asymmetric: it is not seen for both nucleotides forming the terminal pair. On that account we cannot simply attribute it to a lower stability of the closing base pair.

How can be this difference between the two opposing nucleotides at a duplex end explained? Looking at the spatial structure (Fig. 7.34), the aromatic protons H6 and H8 in the major groove located at the 5′-terminal base are more exposed to the environment than the hydrogens of the 3′-terminal base, shielded by the sugar–phosphate backbone to a similar extent as in the inner nucleosides. The situation seems reversed for the H1′ hydrogens: at the 5′-end they point towards the neighbouring phosphate while they are more freely accessible by the solvent molecules at the 3′-end. Hence, the differentially lower T_m and $\Delta G_{310\,K}$ determined for the protons in terminal nucleotides can be ascribed to some changes in the solvation shell which are not necessarily tied to the duplex melting. It can also be connected to the formation of higher-order structures proposed in § 7.5.4 on page 95 based on the chemical shifts.

Contrary to this, the NMR signal intensities of exchangeable hydrogens in both DNA strands clearly decrease at lower temperatures for those located near the ends of the duplex (Fig. 7.30). But, except for the

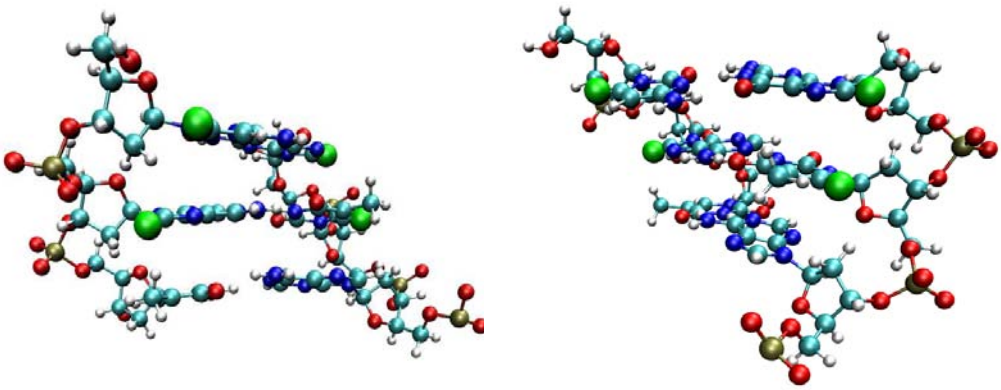


Figure 7.34 A view from the major groove of three nucleotide pairs of a B-DNA duplex close to its end (top side) with H6 and H8 of the ultimate and penultimate bases shown as larger green spheres. Situation at the 5′-end (left) and at the 3′-end (right). Created in VMD [19] from the solution structure of $(CATGCATG)_2$, PDB 1D18 [20]

terminal nucleotides, the melting parameters obtained from line-shape fitting of ^{1}H NMR spectra do not show any observable dependence on the position of a base pair in the sequence.

How this discrepancy should be imagined?

Is it FRAYING?

Very often, the duplex ends are imagined as more unstable than its core [30, 218, 219, 220]. We understand our results from the ^{1}H NMR spectroscopy in another way: the picture of free motions of the terminal nucleotides is in contradiction with the position-independence of the melting parameters. The faster exchange of the terminal amino and imino hydrogens with water can be explained by their easier accessibility from the solvent: although the base-pair disruption frequency and amplitude is very similar along the whole molecule, the middle is more buried in the structure whilst the ends are much more exposed to the outer space. As a consequence, the probability of a hydrogen exchange during a single transient duplex-dissociation event grows towards the ODN ends.

7.6.5 Anomalies in the CpG motif

The H6 and H1′ protons in C4 of the central CpG steps in CTACGTAG and CTTCGAAG show abnormally large stabilities in terms of both T_m

(Fig. 7.31) and $\Delta G_{310\,\mathrm{K}}$ (Fig. 7.32). In contrast, this is not accompanied by any significant deviation in $\Delta G^{\dagger}_{310\,\mathrm{K}}$ (Fig. 7.33).[8]

The outlying values are located in distinct sequence context, ACGT and TCGA, and, *vice versa*, this behaviour is not conserved in different ODN samples sharing the same central tetrads. Two of the four cases under discussion here have also been convicted of abnormally large deviations of experimental duplex chemical shifts from the NN model (Table 7.3 on page 77). These two anomalies together bring new pieces of evidence for the extraordinary malleability of the CpG dinucleotide and the substantial importance of its surroundings in the DNA sequence.

7.6.6 Melting cooperativity and local deviations

From the above analysis of the results on the thermodynamics of duplexes obtained from ^{1}H NMR it follows that:

a. the NN model fails to predict the stabilities of the short octamer ODN duplexes better than 3 K precision in T_m and cannot be reliably used to compare the properties of closely related sequences;
b. generally, the stability parameters as well as the activation energies do not significantly vary along the double-helix which suggests that a high level of long-range cooperativity is associated with the melting process;
c. we observe no signs of fraying of the duplex ends;
d. anomalous melting is locally demonstrated by the duplex termini and at the CpG motif.

Although the duplex melting itself has been found to proceed in a two-state manner with negligibly populated intermediates, some modifications of this general scheme have been revealed. Since the deviating values of T_m and $\Delta G_{310\,\mathrm{K}}$ located at the ODN ends and in some central CpG steps contrast with non-significant variations of $\Delta G^{\dagger}_{310\,\mathrm{K}}$, it is a fast chemical exchange that influences the apparent stabilities observed for the affected nucleotides. Most probably, some rapid temperature-sensitive equilibrium is present in the system, additionally to the duplex–single strand transition. Condensation of the duplexes into higher-order complexes, conformational changes in the DNA backbone, locally different

[8] although only one value, namely for C4H6 in CTACGTAG, is available

degree of solvent accessibility, or some rearrangements in the water–DNA interactions or changes in their dynamics may underlie our observations.

7.7 A three-state process in CTTCGAAG

The VT NMR spectra of G5H8 in CTTCGAAG cannot be reliably modelled by the usual model assuming a two-state process with linear chemical-shift asymptotes: there appears a hump in the peak position in the vicinity of 310 K (Fig. 7.35). This phenomenon has not been observed for any other resonance of any other ODN. In addition, the melting-induced upfield change of its chemical shift is unusual among GH8 (Fig. 7.13 on page 80).

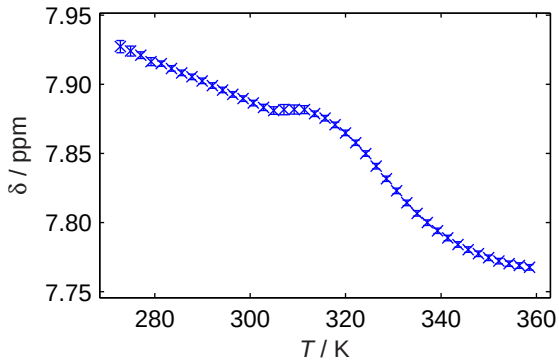


Figure 7.35 Chemical shifts of G5H8 of CTTCGAAG ($c = 1.06$ mM)

The character of the melting curve may indicate that the unfolding is in fact a two-state process but suffers from a non-linearity of the chemical shifts in single strands. A model with parabolic asymptotes has been tested but it did not explain the data correctly. We assume that a three-state equilibrium accompanies the duplex melting.

Properties of the two transitions, even their stoichiometries, can hardly be estimated from a single melting curve. With the objective of a more profound explanation of the processes happening around the central CpG step, the ^{1}H VT NMR spectra have been acquired for four ODN concentrations, spanning almost an order of magnitude (Table 7.1 on page 62).

7.7.1 Global two-state characteristics at variable ODN concentration

Before any three-state model was employed, the usual two-state analysis of the experimental data was performed. Both the aromatic and the thymine methyl regions were fitted by the Lorentzian curves even at medium temperatures, additionally to the exchange line shapes.

The results of the global fits of $\delta(T)$ for each of the concentrations and for all concentrations jointly are presented in Table 7.8. Results of the thermodynamic fits of $p_A(T)$ and $k_A(T)$ obtained from exchange fits of Ha resonances are collected in Table 7.9 and Table 7.10. The values of T_m^{glob} were also fitted by their estimated dependence on c, Eq. (3.24).

As expected, the overall double-helix **stability** in the terms of T_m^{glob} increases with concentration. On the other hand, the values of both ΔH^{glob} and ΔS^{glob} obtained from the fits of $\delta(T)$ clearly increase with increasing c which is also true for $\Delta G^{glob}_{310\,K}$ calculated from them (Table 7.8). This

Table 7.8 Results of global fitting of $\delta(T)$ assuming the two-state process for various c of CTTCGAAG. All: joint fit of all profiles at all concentrations; vH: van 't Hoff fit of $T_m(c)$ by Eq. (3.24)

c / mM	T_m^{glob} / K	$\Delta G^{glob}_{310\,K}$ / kJ · mol^{-1}	ΔH^{glob} / kJ · mol^{-1}	ΔS^{glob} / J · mol^{-1} · K^{-1}
H2, H6, H7, H8				
0.53	319.7 ± 0.1	-29.1 ± 0.3	-317 ± 7	-927 ± 21
0.95	320.7 ± 0.1	-27.0 ± 0.3	-273 ± 5	-793 ± 16
1.80	322.6 ± 0.1	-25.9 ± 0.3	-245 ± 5	-708 ± 14
3.92	326.0 ± 0.1	-25.0 ± 0.2	-219 ± 4	-625 ± 11
All	321.1 ± 0.1^a	-26.6 ± 0.1	-255 ± 2	-737 ± 7
vH	321.2 ± 0.1^a	-26.8 ± 0.3	-260 ± 7	-751 ± 23
H1′				
0.53	320.9 ± 0.2	-31.0 ± 0.4	-339 ± 9	-993 ± 28
0.95	322.0 ± 0.2	-28.1 ± 0.3	-275 ± 6	-796 ± 19
1.80	324.5 ± 0.2	-27.4 ± 0.4	-248 ± 6	-713 ± 18
3.92	327.5 ± 0.2	-26.9 ± 0.4	-236 ± 5	-673 ± 17
All	322.2 ± 0.1^a	-27.4 ± 0.1	-253 ± 3	-726 ± 8
vH	322.5 ± 0.1^a	-27.4 ± 0.3	-248 ± 8	-713 ± 25

[a] T_m calculated for $c = 1$ mM

Table 7.9 Results of global fitting of $p_A(T)$ assuming the two-state process for various c of CTTCGAAG. All: joint fit of all profiles at all concentrations; vH: van 't Hoff fit of $T_m(c)$ by Eq. (3.24)

c / mM	T_m^{glob} / K	$\Delta G^{glob}_{310\,K}$ / kJ · mol^{-1}	ΔH^{glob} / kJ · mol^{-1}	ΔS^{glob} / J · mol^{-1} · K^{-1}
H2, H6, H7, H8				
0.53	318.0 ± 0.1	-25.6 ± 0.1	-245 ± 4	-707 ± 11
0.95	320.8 ± 0.1	-26.4 ± 0.2	-251 ± 4	-723 ± 13
1.80	323.0 ± 0.1	-26.5 ± 0.3	-254 ± 5	-733 ± 16
3.92	326.1 ± 0.1	-26.4 ± 0.3	-246 ± 4	-707 ± 14
All	320.7 ± 0.1^a	-25.8 ± 0.1	-242 ± 2	-697 ± 6
vH	320.6 ± 0.1^a	-24.7 ± 0.1	-210 ± 4	-597 ± 12

[a] T_m calculated for $c = 1$ mM

Table 7.10 Results of global fitting of $k_A(T)$ assuming the two-state process for various c of CTTCGAAG. All: joint fit of all profiles at all concentrations; Ave: error-weighted averages and standard deviations

c / mM	$\Delta G^{\dagger,glob}_{310\,K}$ / kJ · mol^{-1}	$\Delta H^{\dagger,glob}$ / kJ · mol^{-1}	$\Delta S^{\dagger,glob}$ / J · mol^{-1} · K^{-1}
H2, H6, H7, H8			
0.53	66.6 ± 0.1	194 ± 4	411 ± 13
0.95	66.4 ± 0.2	188 ± 5	391 ± 14
1.80	66.4 ± 0.2	193 ± 5	409 ± 16
3.92	67.3 ± 0.3	234 ± 7	539 ± 21
All	66.6 ± 0.1	196 ± 2	419 ± 7
Ave	66.6 ± 0.3	200 ± 17	429 ± 54

effect cannot be explained by the simple thermodynamic model applied. It is not observed for the results of $p_A(T)$ fitting where these quantities stay – surprisingly well, almost perfectly – intact (Table 7.9). This discrepancy might be connected to a duplex condensation witnessed already by chemical shifts (§ 7.5.4 on page 95) to which the exchange fits would be immune due to extrapolation of δ from the linear regions only.

The fits of $T_m^{glob}(c)$ by Eq. (3.24) give very similar ΔH and ΔS to those obtained from joint global fits of individual melting profiles (Fig. 7.36).

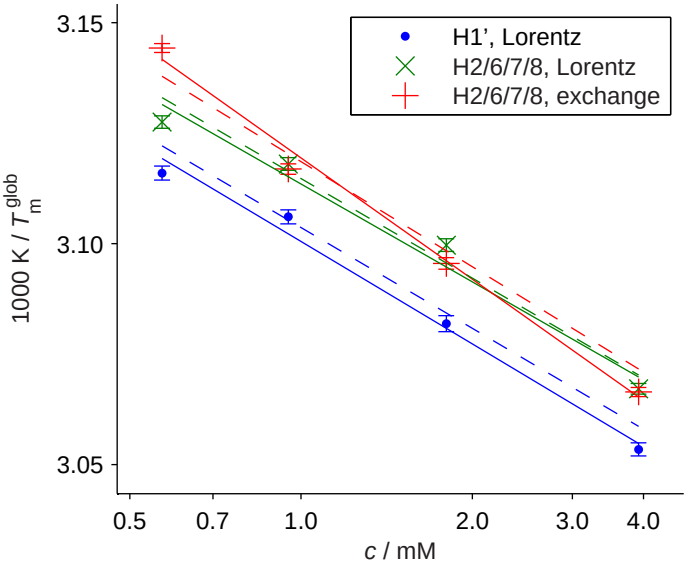


Figure 7.36 Van 't Hoff plot of $T_m^{glob}(c)$ for indicated classes of nuclei and fitting methods. Solid lines: fits by Eq. (3.24); dashed lines: results of joint fits of all $\delta(T)$ of the corresponding proton classes at all concentrations

The largest differences are for the exchange-determined aromatic values. We can ascribe it to the fact that the van 't Hoff fit handles only the few T_m^{glob} values, which are already products of previous fits of the melting curves. The joint fits of $p_A(T)$, on the other hand, work with all the original data together which leads to suppression of imprecisely obtained numbers by the error weighting. This approach appears more reliable and, indeed, the enthalpy and entropy changes lie closer to those from the $\delta(T)$ dependence.

We can see that T_m^{glob} obtained from the H1′ resonances are more than 1 K higher than from the aromatic hydrogens, similarly to the findings for other ODN sequences described in § 7.6.1 on page 99. Since the comparison of T_m^{glob} determined by the fits of Lorentzian and exchange line shapes of Ha shows no big impact of the selection of a particular spectral-fitting method, the higher apparent stabilities of H1′ relative to Ha are not caused by the different approaches to these two classes of hydrogen nuclei. It indicates that there is an intrinsic structural transition to which the Ha and H1′ protons are differently sensitive.

The **kinetic parameters** do not change significantly with c (Table 7.10) which is in agreement with the presumptions. There is only a deviation for the highest c but removing several values of k_A that apparently have accidentally underestimated errors much improves the match with the other concentrations.

7.7.2 Deviations from the global two-state fits

Despite there are no obvious discrepancies in the global thermodynamic quantities, a violation of the two-state model can be seen in the residuals of the CTTCGAAG chemical shifts from their global fits[9] (Fig. 7.37 and Fig. 7.38). The deviations were not that high for any other ODN sequence in our set.

The residuals from global fitting of p_A are also larger than for other ODN sequences (Fig. 7.39). The absence of any oscillations similar to what was seen in the residuals of δ is probably caused by the exchange line-shape analysis that presumes the two-state behaviour in its definition.

7.7.3 Equilibrium witnessed by G5H8

The most deviating chemical shifts are observed for the G5H8 nucleus. Fits by the three-state processes described by Eq. (3.29) and Eq. (3.30) have been performed for the G5H8 shifts. In accordance with the previous analyses of NMR melting data, we suppose that δ of the states prevalent at the lowest and highest temperatures are linear with respect to T. On the other hand, δ of the third state is very poorly and indirectly determined because there is no temperature at which this site dominates. In order to improve the stability of the fits we model it as temperature independent.

First, we assumed the **D_1–D_2–S** scheme (3.29) with a duplex–duplex transition in the first step. During the fit of $\delta(T)$ optimising all the thermodynamic parameters together with the chemical-shift asymptotes, the characteristics of the first transition, D_1–D_2, were common for all c but the melting transition, D_2–S, was independent at each c. This analysis yielded T_m of the D_2–S process increasing with c, while ΔH and ΔS are relatively unchanged (Fig. 7.40 and Table 7.11).

[9] performed independently for each concentration

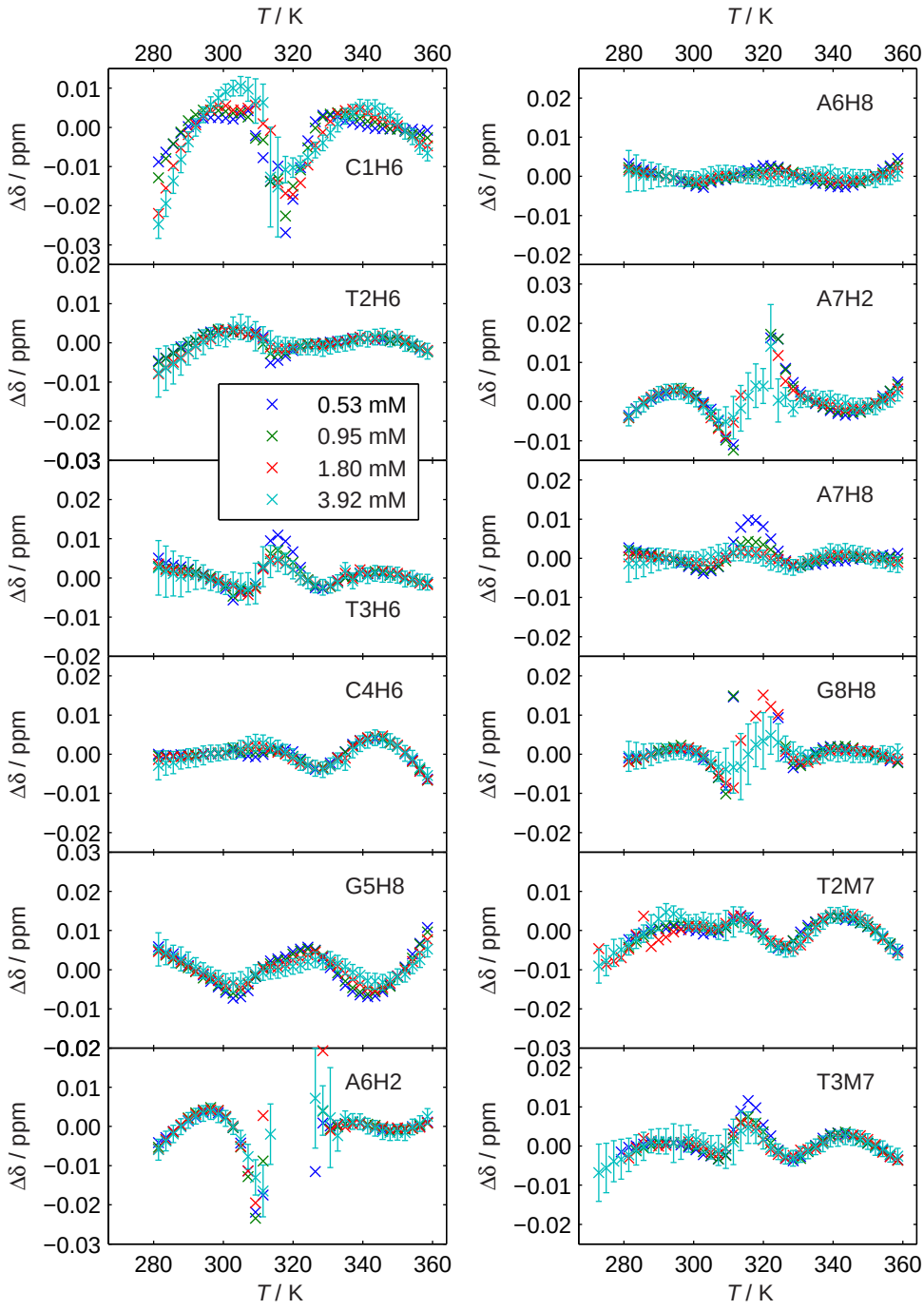


Figure 7.37 Residuals of δ of aromatic protons from their global fits by the two-state model. All plots with the same scales. Error bars shown only for the highest c for better clarity

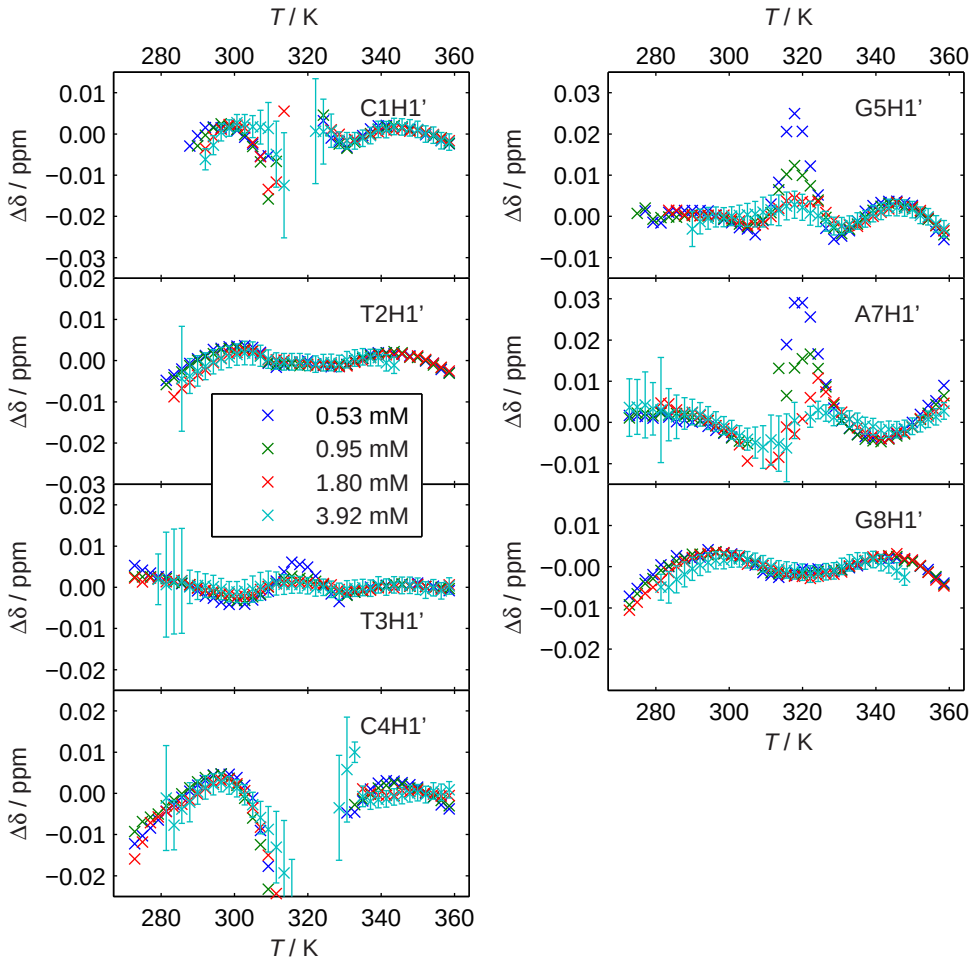


Figure 7.38 Residuals of δ of H1′ from their global fits by the two-state model. All plots with the same scales (matching those in Fig. 7.37). Error bars shown only for the highest c for better clarity

Next, the second type of the three-state model, **D–S$_1$–S$_2$** (3.30), was applied, with the parameters of the second step kept independent on c. Supposing that the fully unfolded single strands are identified as S$_2$ resulted in non-monotonic dependence of the duplex–single strand T_m. However, the model in Eq. (3.30) can also be interpreted in another way: the fully melted state can be S$_1$ whereas S$_2$ can be some unimolecular fold, e.g. a hairpin, which is unstable at high temperatures. This gave more realistic results but the temperature dependence of chemical shifts is not reliably reproduced by the fit, especially around the local maxima near 310 K, and

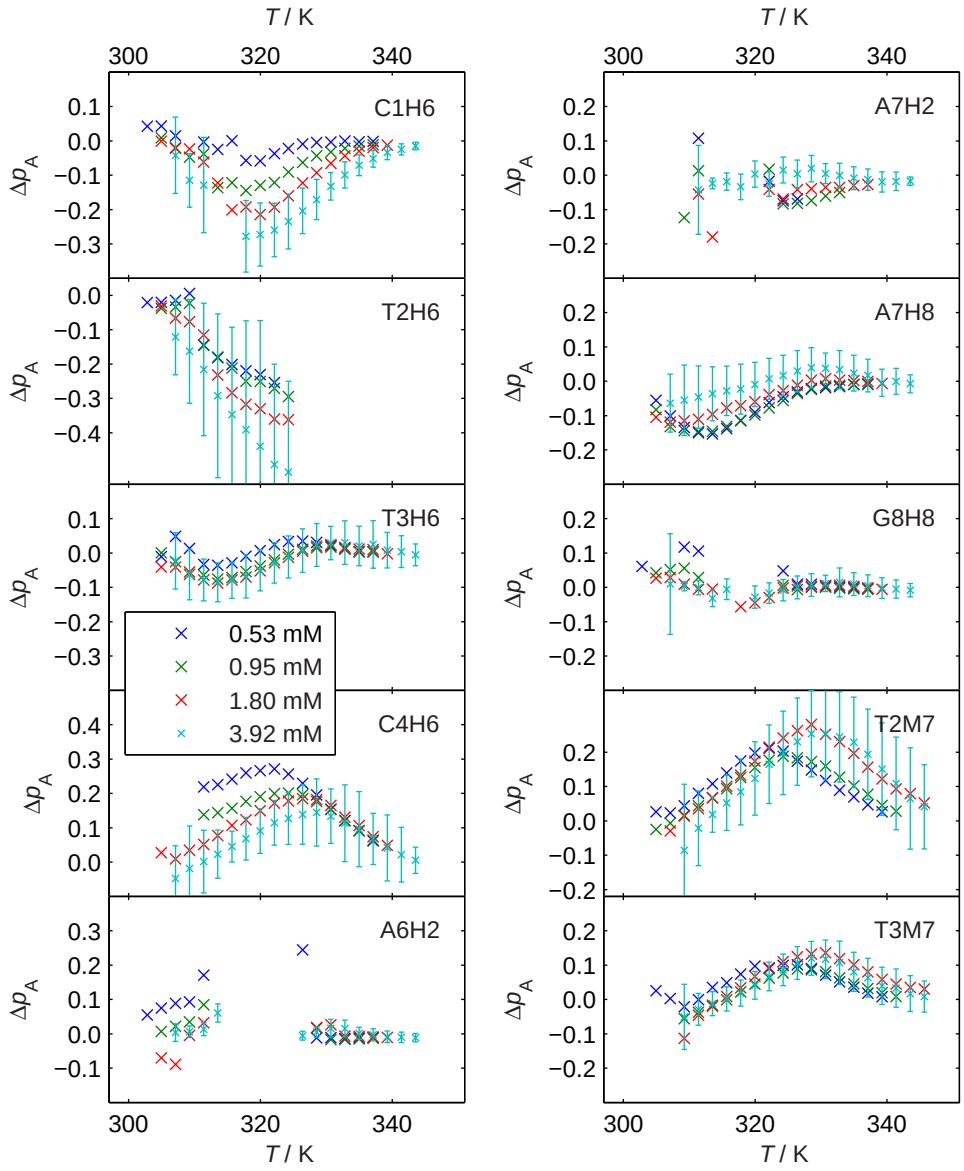


Figure 7.39 Residuals of p_A from their global fits. All plots with the same scales. Error bars shown only for the highest concentration for better clarity

Table 7.11 Results of fitting of the D_1–D_2–S three-state process (3.29) to $\delta(T)$ of G5H8 in CTTCGAAG for various c

c / mM	T_m / K	$\Delta G_{310\,K}$ / kJ · mol^{-1}	ΔH / kJ · mol^{-1}	ΔS / J · mol^{-1} · K^{-1}
$D_1 \rightleftarrows D_2$				
All	308 ± 3	13 ± 3	-350 ± 160	-1190 ± 520
$D_2 \rightleftarrows S + S$				
0.53	325 ± 5	-27 ± 4	-168 ± 47	-450 ± 140
0.95	326 ± 4	-27 ± 4	-174 ± 44	-470 ± 130
1.80	327 ± 3	-26 ± 4	-178 ± 43	-490 ± 130
3.92	330 ± 3	-26 ± 4	-187 ± 47	-520 ± 140

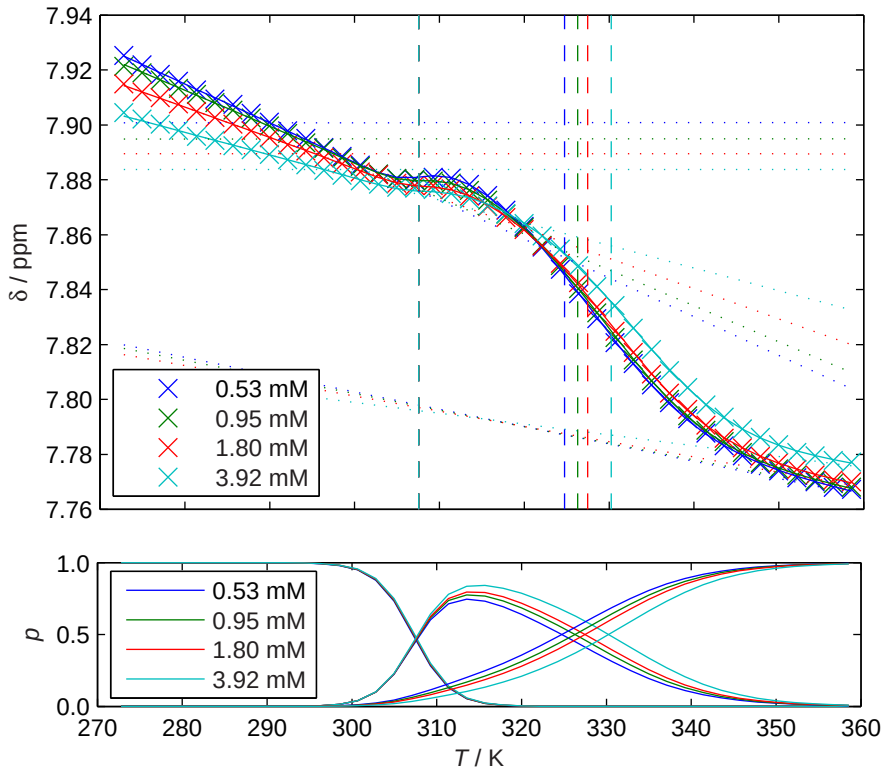


Figure 7.40 Results of the fits of $\delta(T)$ of G5H8 in CTTCGAAG at various c by the D_1–D_2–S three-state model (3.29). Top: chemical shifts; dotted lines: chemical shifts of individual sites; dashed vertical lines: transition temperatures of the two processes; error bars omitted for clarity. Bottom: relative populations of D_1, D_2, and S

the chemical shifts of the two single-strand species appear to be too much away from the measured values (Fig. 7.41).

Based on the above analysis of G5H8 chemical shifts we could **conclude** that the D_1–D_2–S model involving a duplex–duplex transition followed by its denaturation to single strands is more likely. On the other hand, there are large errors in the fitted parameters that make the results far from satisfactory without any additional experiments. Also, the highest deviations from the two-state model located close to T_m unexpectedly decrease with increasing c (Fig. 7.37) which would be more consistent with the D–S_1–S_2 model.

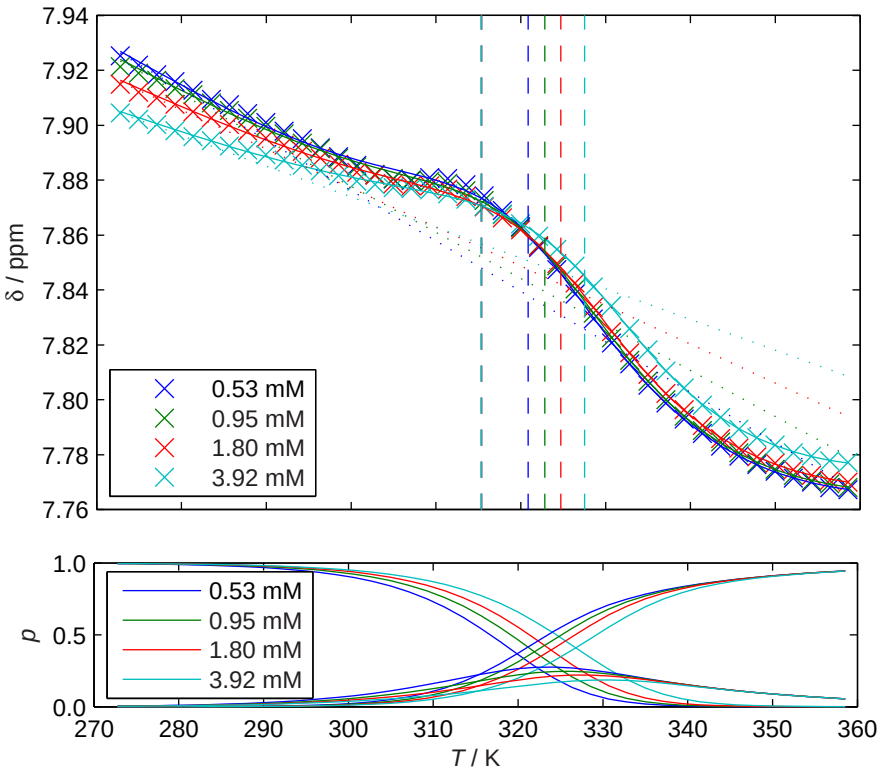


Figure 7.41 Results of the fits of $\delta(T)$ of G5H8 in CTTCGAAG at various c by the D–S_1–S_2 three-state model (3.30). Top: chemical shifts; dotted lines: chemical shifts of duplex (δ of S_1 below 7.76 ppm and δ of S_2 around 8.3 ppm are outside the plot limits); dashed vertical lines: transition temperatures of the two processes; error bars omitted for clarity. Bottom: relative populations of D, S_2, and S_1 (their maxima from left to right)

Analyses by the three-state model have also been tried for other ^{1}H nuclei of CTTCGAAG. Nevertheless, no reliable results can be obtained because the melting curves can be relatively well explained by the ordinary two-state melting. The two transitions which are observed separately in the case of G5H8 resonance could still be present for the other spectral lines but merged into apparently one process.

7.7.4 Comparison with CATCGATG

The central tetranucleotide, TCGA, is shared with another ODN in our series, namely CATCGATG, that differs from CTTCGAAG by the reversed second base pair. This modification removes the series of four pyrimidines followed by four purines and breaks the T_2 and A_2 repetitions.

In the melting profiles of CATCGATG, no severe violations of the two-state model have been discovered. It explains the melting data much more neatly for all the ^{1}H resonances, except for those in the C1 terminal residue (Fig. 7.42).

7.7.5 Melting curves from UV and Raman

During the research of the methylation impact in Chapter 8, the CTTCGAAG sample was also subjected to the measurements of UV and RS.

In UV, the hypothetical duplex–duplex transition around 308 K from Table 7.11 would not be detectable because it is already above $T_\mathrm{m} = 300$ K due to low $c = 12.0$ μM. Any unimolecular folding process in the D–S_1–S_2 model (3.30) is not detected either although this would be possible if the transition temperatures were as in Fig. 7.41.

The two-state model suits also well the Raman spectra measured at $c = 2.8$ mM with $T_\mathrm{m} = 323$ K, which sufficiently separates the proper duplex melting from the other transition.

It can be summarised that although at least one of these experimental techniques in principle allows observation of the extra transition revealed in NMR spectrum mainly for G5H8, it has not been detected. The spectral changes in UV and RS are probably too small to be observed in the VT measurements.

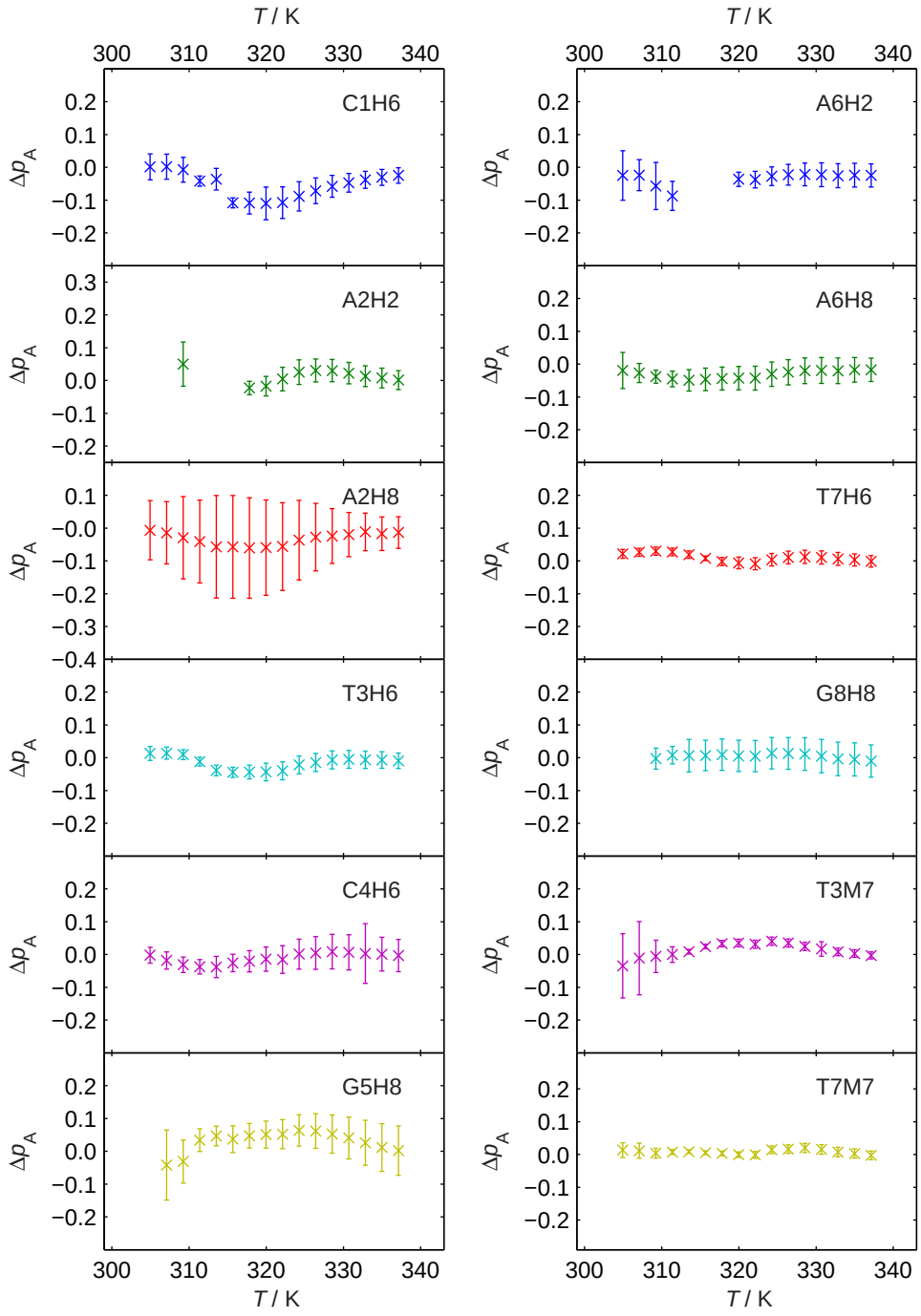


Figure 7.42 Residuals of p_A of CATCGATG from their global fits. All plots with the same scales (matching those in Fig. 7.39)

7.7.6 Extraordinary melting thermodynamics

The series of ^{1}H VT NMR spectra with variable concentration of CTTCGAAG has revealed a deviation from the otherwise two-state DNA duplex melting. We observed that:

a. the overall concentration-dependent two-state melting follows well the van 't Hoff equation and the kinetics is conserved;
b. exchange and Lorentzian fits of the NMR spectra yield the same thermodynamic results but the exchange approach profits from the assessment of the kinetics;
c. different sensitivity of the aromatic and saccharide hydrogens to the melting process has been confirmed;
d. there is clearly a non-two-state behaviour of the G5H8 chemical shifts;
e. fitting of three-state processes to the G5H8 $\delta(T)$ favours the D_1–D_2–S model (3.29) over the D–S_1–S_2 scheme (3.30) but a contradictory conclusion comes from larger deviations at lower concentrations;
f. UV and RS conform to the two-state melting;
g. the anomalous behaviour is not detected in a closely related ODN sequence CATCGATG.

Although we succeeded in modelling the data by the three-state equilibria, no evidence is strong enough to rule out any of the two possible melting pathways. An unusual chemical exchange, ascribed to two different duplex conformations, was also reported earlier for the ^{31}P resonance of the CpG phosphorus [221]. The interesting properties of this particular DNA double-helix are yet to be investigated in more detail, employing additional experiments with a wider range of ODN concentration.

8 DUPLEXES WITH 5-METHYLCYTOSINE

8.1 Samples

To find out the effect of the methylation of cytosine in the CpG motif of self-complementary DNA fragments on the spectral and thermodynamic properties, variable-temperature NMR experiments were carried on solutions of three ODN sequences carrying the symmetrically methylated (m^{5}CpG)·(m^{5}CpG) central segment. They are listed in Table 8.1.

Table 8.1 Concentrations, c, of ODN samples with m^{5}C measured by ^{31}P qNMR and calculated from the provider data (Provider)

	c / mM	
Sequence	^{31}P qNMR	Provider
CAAm^{5}CGTTG	1.01 ± 0.05	1.40
CATm^{5}CGATG	1.12 ± 0.06	1.36
CTTm^{5}CGAAG	0.92 ± 0.05	1.09

All of these DNA fragments have their unmodified counterparts in the previous chapter dealing with the self-complementary duplexes. ^{1}H NMR spectra of one sample, CAAm^{5}CGTTG, were measured by a bachelor student under my supervision [205] and the results are significantly extended in this dissertation.

8.2 Assignment of ^{1}H NMR spectra

The standard assignment protocols (§ 4.3.1 on page 40) are not affected by the presence of m^{5}C because the main scheme is based on H6 and H8 resonances. The methyl hydrogens H7 of m^{5}C experience the same chemical shift range and the same direct dipole–dipole interactions as the thymine methyl.

All the aromatic, exchangeable, and anomeric protons have been successfully identified from the NOESY spectra (Fig. A.9–A.11 in Appendix). Even in the complicated region of H2′ and H2″ resonances, all signals were distinguished stereospecifically in the NOESY spectra except for the G5 deoxyribose in CTTm^{5}CGAAG, which is a feature conserved from the unmethylated duplex. The resonances have been tracked to single strands along the temperature series of 1D spectra.

Special treatments were needed in a few unclear cases (Table 8.2). An example of the assignment of equivocal T and m^{5}C methyl protons in single strands is shown in Fig. 8.1.

The duplex shifts are presented in Table A.2 and the single-strand shifts in Table A.4 in Appendix.

Figure 8.1 Assignment of methyl groups in CTTm^{5}CGAAG at 350 K. From top to bottom: selective ^{1}H COSY with selective irradiation of T2H6, T3H6, and m^{5}C4H6 protons; normal 1D ^{1}H spectrum

Table 8.2 Resonances assigned ambiguously and methods of resolving them with marks indicating whether the approach has been successful. Duplex is indicated by 'd' and single strands by 's'

Sample	Ambiguities	Method applied, conditions	
CATm^{5}CGATG	AH2 (d)	NOESY, 294 K	✓
	TH3 (d)	NOESY, 294 K	✓
	TH7 (s)	selective COSY, 338 K	✓
CTTm^{5}CGAAG	TH7 (s)	selective COSY, 326 K, 350 K	✓
	AH2 (s)	analogy with CTTCGAAG	✓

8.3 Effect of cytosine methylation on ^{1}H chemical shifts

This section evaluates the influence of the methyl group added to cytosine on δ. The experimental chemical shifts for unmodified and methylated ODN, obtained from Lorentzian-line fits, and their dependence on temperature are analysed. In addition, new parameters for m^{5}C-containing ODN are modestly attached to the state-of-the-art prediction protocols.

8.3.1 Changes of chemical shifts

Differences between chemical shifts after methylation, δ_{Me}, and in unmodified ODN, δ_{Norm}, are shown in Fig. 8.2 and Fig. 8.3 for duplex and in Fig. 8.4 for single strands.

The increased shielding ($\Delta\delta \doteq -0.17$ ppm) of the base protons H6 of the modified cytosine itself is commonly explained by the positive inductive effect caused by the electron-donating methyl group [105]. The shifts of nucleotides adjacent to the chemical modification are also affected. In duplex, this effect is propagated further along the double-helix compared to the situation in the random-coil state.

A closer look on the data in **duplex** reveals that larger changes are observed for the base and sugar protons of the 5′-adjacent nucleotide W3 (i.e., A3 or T3) than for the 3′-neighbouring G5. This finding agrees well with a previous work [105] where changes very similar to our values were measured for thymine in a (TCG)·(CGA) triplet. We explain such differentially larger sensitivity of the 5′-neighbour by the spatial proximity of the affected hydrogens to CH5/m^{5}CH7.

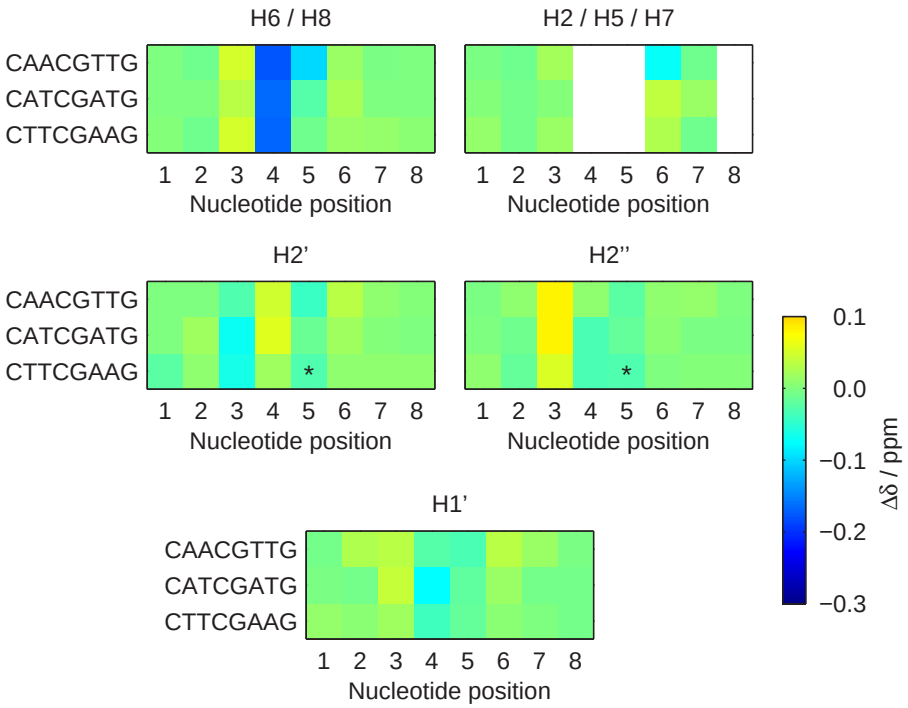


Figure 8.2 Changes in chemical shifts of non-exchangeable protons in duplex (284 K) induced by methylation of C4. $\Delta\delta = \delta_{\mathrm{Me}} - \delta_{\mathrm{Norm}}$. The asterisks mark overlapping H2′ and H2″ resonances

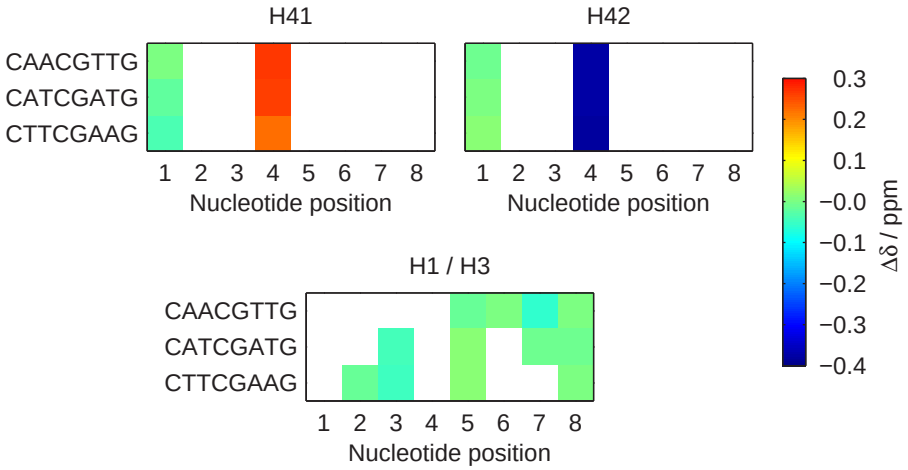


Figure 8.3 Changes in chemical shifts of exchangeable protons in duplex (284 K) induced by methylation of C4. $\Delta\delta = \delta_{\mathrm{Me}} - \delta_{\mathrm{Norm}}$

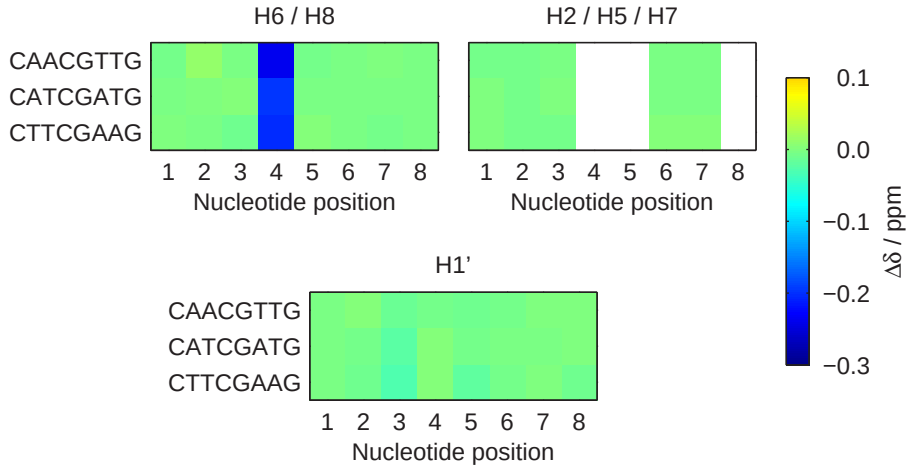


Figure 8.4 Changes in chemical shifts of protons in single strands (354 K) induced by methylation of C4. $\Delta\delta = \delta_{\mathrm{Me}} - \delta_{\mathrm{Norm}}$

Remarkably large $\Delta\delta$ are presented by G5H8, T6H7, and T7H3 protons of the $(\mathrm{CAACGTTG})_2$ duplex (Table 8.3). The values for T6H7 are close to those reported previously in the RACGTY[1] fragments [109, 107] (Table 8.3). While only a subtle conformational variation is seen after the methylation in the GACGTC context [107], the central part of the double helix containing an A-tract is significantly affected [109]. Changes in the solvation shell were suggested as the cause of the shielding differences [107]. We also observed the differential sensitivity of TH7 to the 5′-surrounding one base apart (Fig. 7.6 on page 69). Therefore, because of similar effects on δ regardless of the structural change, a net next-nearest-neighbour influence or an inter-strand interaction is responsible (Fig. 8.5). Experiments with asymmetrically methylated duplexes would help to confirm or reject this hypothesis.

The above observations prove that the effect of methylation depends on wider sequence context and not only on the nature of the neighbouring bases. It should be regarded as a part of evidence that the DNA double-helix possesses a long-range sensitivity to its modifications.

In **single strands**, the changes in neighbouring nucleotides are largely suppressed. It seems that the much more flexible structure with weaker stacking diminishes the through-space interactions between adjacent bases.

[1] R = A or G (purine bases); Y = C or T (pyrimidine bases)

Table 8.3 Chemical-shift differences of selected ^{1}H induced by symmetric methylation of cytosines in the central (CpG)·(CpG) in duplex. $\Delta\delta = \delta_{\mathrm{Me}} - \delta_{\mathrm{Norm}}$. The numbering of nucleotides refers to the octamer

	$\Delta\delta$ / ppm			
	G5H8	T6H7	T7H3	Source
CAACGTTG	−0.097	−0.073	−0.055	This work
GAAAACGTTTTC	−0.10	−0.07	−0.01	[109]
GAGATGACGTCATCTC	−0.08	−0.05		[107]

Figure 8.5 Scheme of the possible interactions of m^{5}C methyl groups (cyan dots) with T6H7 and T7H3 in CAACGTTG. The two-fold symmetry of the self-complementary duplex is stressed by the rotation of the bottom strand. Purines are shown larger than pyrimidines

8.3.2 New predictions for m^{5}C-containing ODN

As described above, the effect of cytosine methylation on ^{1}H chemical shifts in adjacent nucleotides represents a complex problem. Despite this, we propose new values for predictions of δ in DNA sequences containing the m^{5}C base according to the protocols developed by other researchers for the major bases [163, 173]. We are aware of the lack of sufficiently large set of experimental results to reliably establish statistically robust predictions. Additionally, we don't have enough data to predict every possible triplet containing m^{5}C due to the limited number of sequences used. However, this could be a first approximation when dealing with methylated oligonucleotides, which is not available in the literature yet.

The ^{1}H chemical shifts of the m^{5}C nucleotide in **duplex** are shown in Table 8.4. The effects of cytosine methylation on nucleotides adjacent to m^{5}C are gathered in Table 8.5: these values should just be added to the shifts predicted for duplexes bearing regular C [163]. The presented values originate from our work joint with chemical shifts published else-

Table 8.4 Average chemical shifts, δ, of m^{5}C in various sequence contexts in duplex together with their standard deviations, σ, and counts of experimental data used for the averaging, N

	Non-exchangeable, (282–300) K					Amino, 283 K	
	H6	H7	H1′	H2′	H2″	H41	H42
	δ / ppm						
Am^{5}CG	6.96	1.49	5.52	1.98	2.34	8.31	6.16
ATm^{5}CG	7.25	1.69	5.54	2.04	2.35	8.70	6.45
TTm^{5}CG	7.28	1.78	5.52	1.99	2.28	8.83	6.64
	σ / ppm						
Am^{5}CG	0.07	0.10	0.01	0.08	0.01		
ATm^{5}CG	0.02	0.03	0.02	0.02	0.02	0.01	0.02
TTm^{5}CG	0.02	0.04	0.09	0.04	0.02	0.02	0.01
	N						
Am^{5}CG	2	2	2	2	2	1	1
ATm^{5}CG	2	2	2	2	2	2	2
TTm^{5}CG	3	2	3	2	2	2	2

where [105, 106, 109] from which all the values were used except for the amino hydrogens from [109] which lie more than 0.3 ppm upfield from ours. Due to the relatively strong influence of the 5′-next-nearest neighbour on cytosine shifts in the TCG triplet and on methylation-induced changes of guanine shifts in ACGT and TCGA, we treat these cases separately.

Another approach is taken for prediction of the **single-strand** chemical shifts [173]: the trinucleotides are considered independently and there are supplementary numbers accounting for the effects of next-nearest neighbours. We present only the average δ of the central residues in m^{5}C-containing triplets found in our samples (Table 8.6) which we suggest to be used directly for eventual predictions of shifts in these triplets. We ignore the next-nearest-neighbour corrections as they were found negligible, less than 0.01 ppm. As opposed to the duplex state, we cannot join our data with literature values because, unfortunately, only the double-helical chemical shifts have been published [105, 106, 109].

Table 8.5 Increments of chemical shifts in duplex, $\Delta\delta = \delta_{\mathrm{Me}} - \delta_{\mathrm{Norm}}$, of nucleotides adjacent to cytosine caused by its methylation, together with their standard deviations, σ, and counts of experimental data used for the averaging, N

	Non-exchangeable, (282–300) K					Imino, 283 K
	H6/H8	H2/H7	H1′	H2′	H2″	H1/H3
	$\Delta\delta$ / ppm					
5′-A	0.05	0.02	0.03	−0.02	0.08	
5′-T	0.04	0.02	0.03	−0.07	0.07	−0.03
3′-G in ACGT	−0.09		−0.04	−0.03	−0.03	−0.01
3′-G in TCGA	−0.01		−0.01	−0.01	−0.02	0.02
	σ / ppm					
5′-A	0.00	0.00	0.01	0.00	0.00	
5′-T	0.01	0.01	0.02	0.01	0.01	0.03
3′-G in ACGT	0.01		0.01	0.01	0.01	0.01
3′-G in TCGA	0.01		0.01	0.02	0.01	0.01
	N					
5′-A	2	2	2	2	2	
5′-T	5	5	5	4	4	4
3′-G in ACGT	2		2	2	2	2
3′-G in TCGA	5		5	4	4	5

8.3.3 Sensitivity of chemical-shift slopes to methylation

The differences between the slopes of $\delta(T)$, obtained by linear fitting, in methylated, a_{Me}, and in unmodified ODN, a_{Norm}, are shown in Fig. 8.6 and Fig. 8.7 for duplex and in Fig. 8.8 for single strands.

In contrast to the shifts at a single temperature, the methylation-induced changes in the slopes have no general tendency to diminish further from the methylated base. In single strands, the amplitudes of Δa are not significantly lowered compared to duplex. In addition to these global statements, two main phenomena are found:

a. the slopes of C4H6 and G5H1′ shifts in duplex are affected homogenously in the three sequences studied;

Table 8.6 Chemical shifts δ of the central nucleotides N in various triplets in single strands (354 K). The triplets Tm^{5}CG and m^{5}CGA occur twice in our sequences, hence their averages are shown and standard deviations σ appended

N	Triplet	H6/H8	H2/H7	H1′
		δ / ppm		
m^{5}C	Am^{5}CG	7.34	1.76	6.10
	Tm^{5}CG	7.41	1.90	6.12
A	AAm^{5}C	8.23	7.98	6.24
T	ATm^{5}C	7.47	1.75	6.11
	TTm^{5}C	7.55	1.85	6.15
G	m^{5}CGA	7.77		5.80
	m^{5}CGT	7.88		6.09
		σ / ppm		
m^{5}C	Tm^{5}CG	< 0.01	0.01	0.02
G	m^{5}CGA	< 0.01		0.06

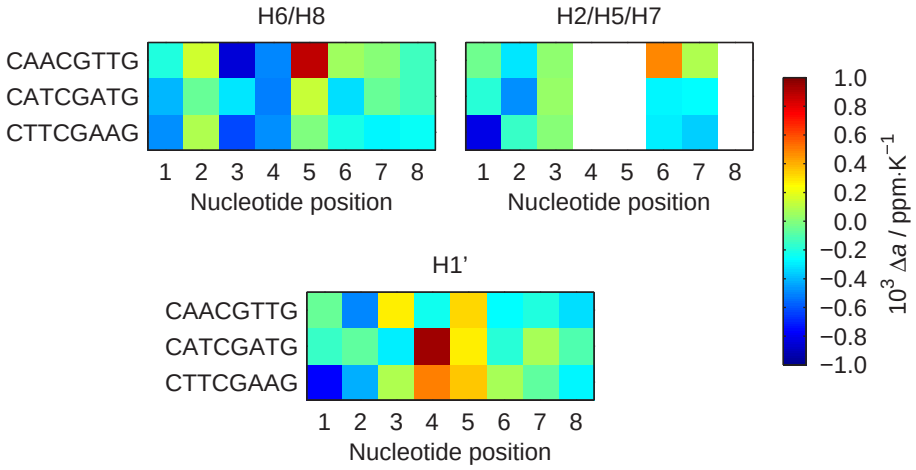


Figure 8.6 Changes in chemical-shift slopes of non-exchangeable protons in duplex induced by methylation of C4. $\Delta a = a_{\mathrm{Me}} - a_{\mathrm{Norm}}$

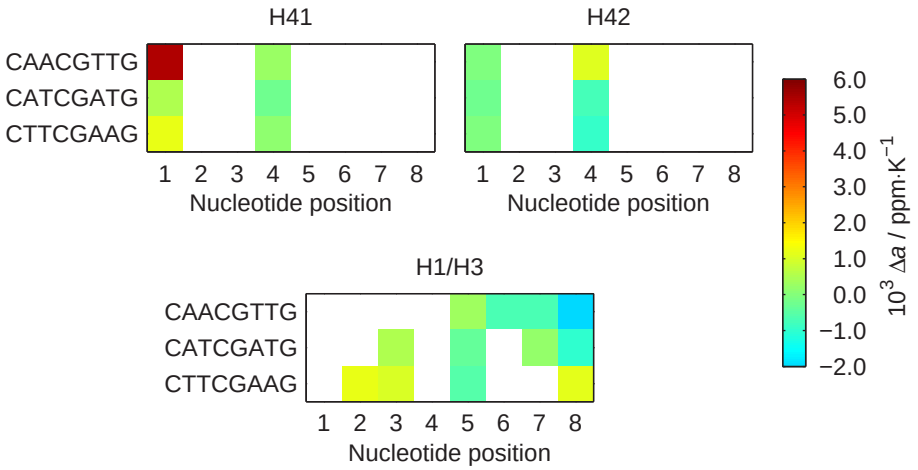


Figure 8.7 Changes in chemical-shift slopes of exchangeable protons in duplex induced by methylation of C4. $\Delta a = a_{\mathrm{Me}} - a_{\mathrm{Norm}}$

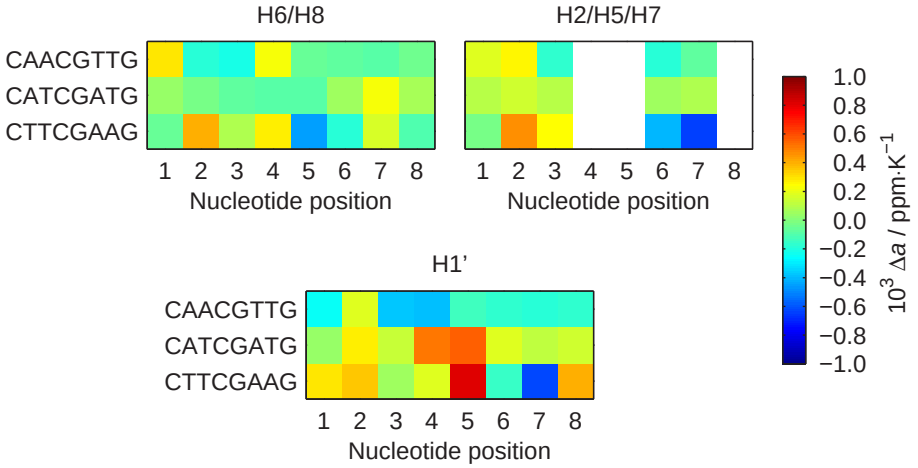


Figure 8.8 Changes in chemical-shift slopes in single strands induced by methylation of C4. $\Delta a = a_{\mathrm{Me}} - a_{\mathrm{Norm}}$

b. a strong long-range influence is found in the double-helical CAACGTTG: the highly positive Δa of G5H8 and T6H7, correlating with the anomalies observed for duplex chemical shifts at a single temperature (Table 8.3), are supported by large slope differences of the exchangeable protons throughout the entire molecule.

8.3.4 Delocalised shielding changes

We have shown that the cytosine methylation strongly affects not only the chemical shifts of the most proximal hydrogen nuclei but also those several base pairs away. This is also underlying the parameters for prediction of chemical shifts proposed in this section. The importance of the sequence context makes the nearest-neighbour approach untrustworthy when accurate description of the duplex properties is needed. The DNA double helix cannot be imagined as a series of isolated base-pair couples. Instead, much longer segments must be considered in future models, reflecting the propagation of the stacking interaction along the double-stranded DNA.

Our results reveal that the temperature dependence of chemical shifts can also serve as a sensitive probe of the double-helix properties. As an example, a globally anomalous response to the CpG methylation is observed for the CAACGTTG sequence.

8.4 Methylated-duplex melting observed by ^{1}H NMR

The line-shape fitting of ^{1}H spectra was performed in the same way as for unmodified ODN (§ 7.6 on page 96). The results of the thermodynamic analysis by the van 't Hoff and Eyring models are tabulated in § A.5 of Appendix (page A-32) and compared to the unmethylated analogues in Fig. 8.9–8.11. The results of the global fits are in Table 8.7 and Table 8.8.

8.4.1 Methylation effect on duplex stability

Comparison of the results from the global fits of unmodified (Table 7.6 on page 100 and Table 7.7 on page 101) and methylated ODN (Table 8.7 and Table 8.8) reveals that the methylated analogues are a bit more stable under the same conditions and that the activation energies for duplex disruption are higher when m^{5}C is present.

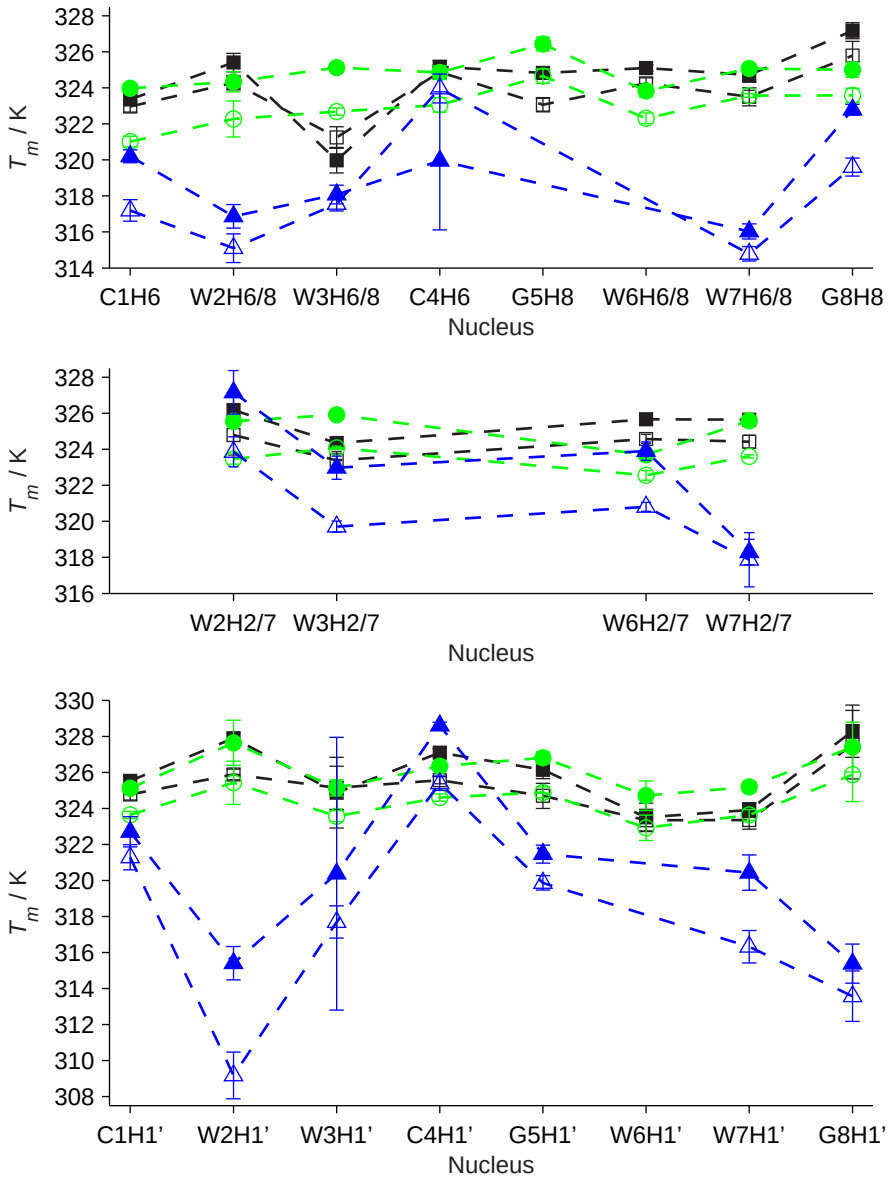


Figure 8.9 Individual T_m of ^{1}H resonances in methylated and unmodified ODN recalculated for $c = 1$ mM. Open symbols: unmethylated; filled symbols: methylated. Black squares: CAACGTTG; green circles: CATCGATG; blue triangles: CTTCGAAG. W = A or T. Dashed lines only connecting points for better clarity

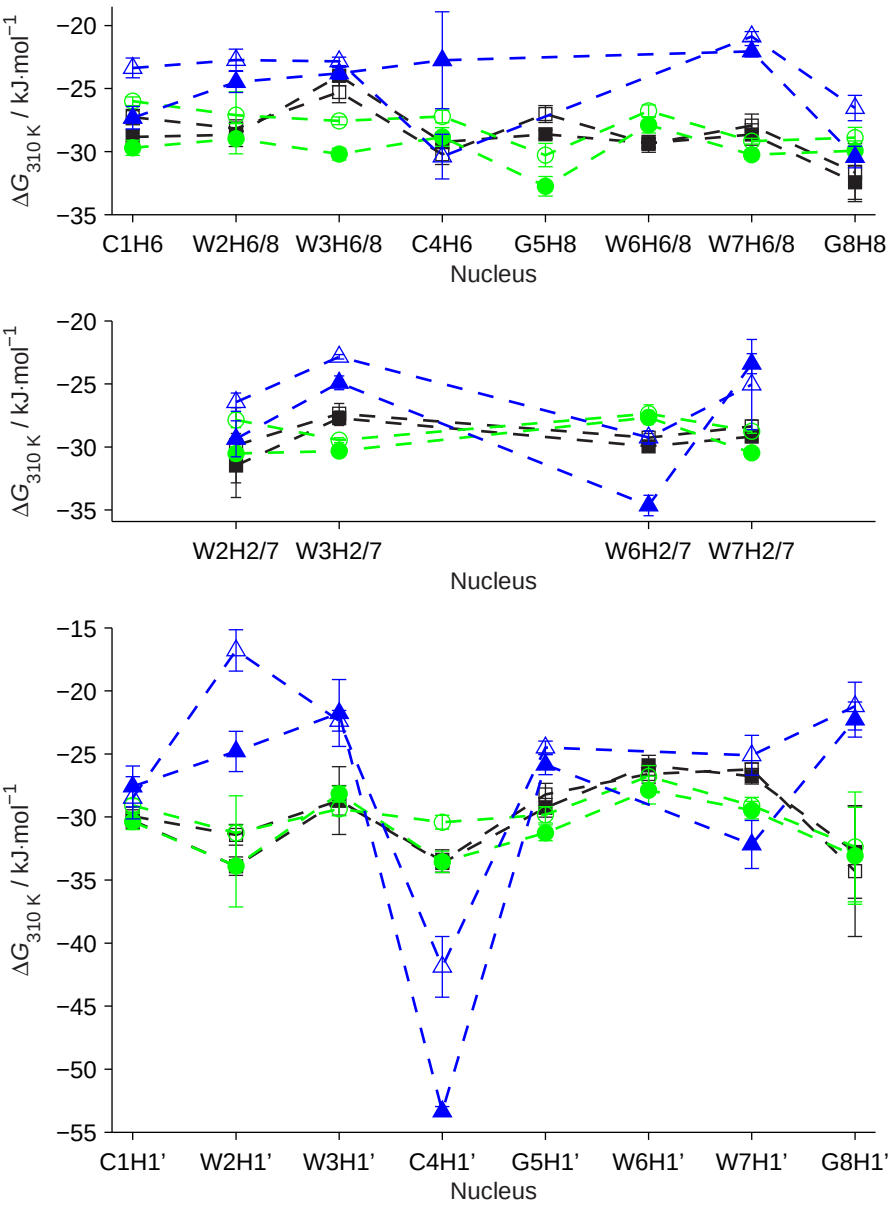


Figure 8.10 Individual $\Delta G_{310\,K}$ of ^{1}H resonances in methylated and unmodified ODN. Open symbols: unmethylated; filled symbols: methylated. Black squares: CAACGTTG; green circles: CATCGATG; blue triangles: CTTCGAAG. W = A or T. Dashed lines only connecting points for better clarity

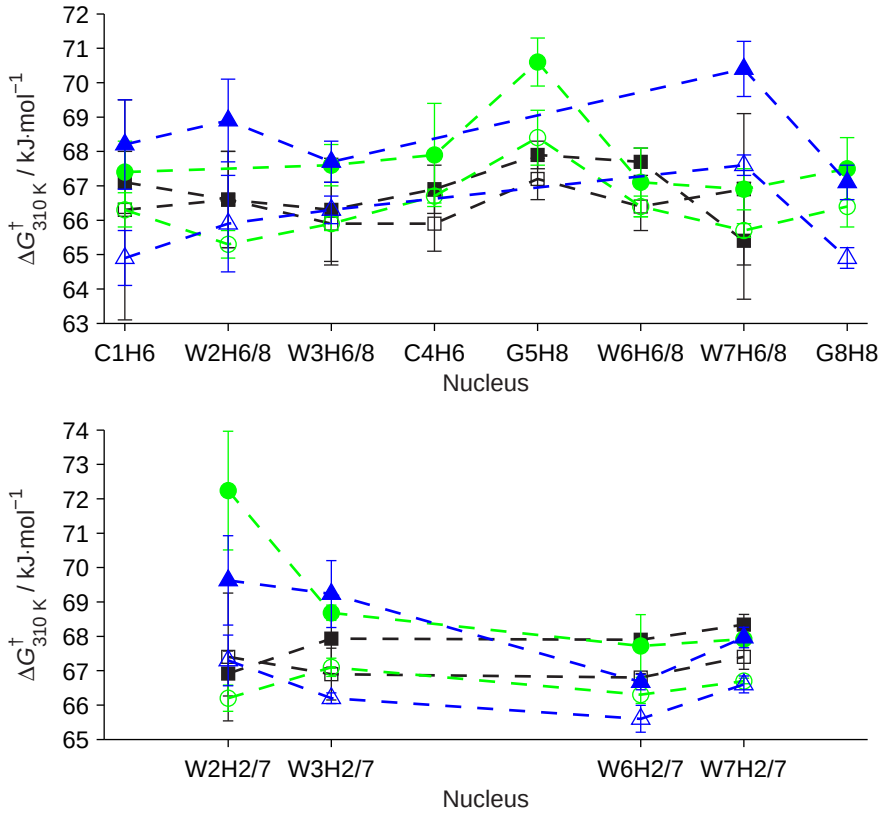


Figure 8.11 Individual $\Delta G^{\dagger}_{310\,K}$ of ^{1}H resonances in methylated and unmodified ODN. Open symbols: unmethylated; filled symbols: methylated. Black squares: CAACGTTG; green circles: CATCGATG; blue triangles: CTTCGAAG. W = A or T. Dashed lines only connecting points for better clarity

The differences in duplex stability and in the activation energies are not restricted to the central parts of the double helix (Fig. 8.9–8.11). Instead, the methylation-induced changes are distributed evenly among the whole DNA strand. The duplex structure is highly cooperative and the chemical modification in its middle base pair has a similar influence on all the nucleotides. No particular anomaly has been observed even for CAACGTTG although there are unexpectedly high and delocalised changes in chemical shifts after methylation. Similar extent of the melting cooperativity has been seen also after adenine substitution by 2-aminopurine [222].

Table 8.7 Results of global fitting of $p_A(T)$ for Ha (H2, H6, H7, and H8) and $\delta(T)$ for H1′ protons. T_m^{glob} are calculated for 1 mM ODN concentration. The errors presented are coming purely from the fits

	$\frac{T_m^{glob}}{K}$	$\frac{\Delta G_{310\,K}^{glob}}{kJ \cdot mol^{-1}}$	$\frac{\Delta H^{glob}}{kJ \cdot mol^{-1}}$	$\frac{\Delta S^{glob}}{J \cdot mol^{-1} \cdot K^{-1}}$
CAAm^{5}CGTTG				
Ha	325.1 ± 0.1	-29.7 ± 0.1	-257 ± 2	-733 ± 7
H1′	326.0 ± 0.1	-30.7 ± 0.3	-263 ± 4	-750 ± 13
CATm^{5}CGATG				
Ha	325.4 ± 0.1	-30.4 ± 0.1	-266 ± 2	-761 ± 7
H1′	325.5 ± 0.1	-30.7 ± 0.3	-270 ± 5	-772 ± 14
CTTm^{5}CGAAG				
Ha	321.1 ± 0.1	-28.5 ± 0.2	-310 ± 4	-907 ± 13
H1′	324.1 ± 0.1	-40.6 ± 0.2	-525 ± 3	-1561 ± 10

Table 8.8 Results of global fitting of $k_A(T)$ of all relevant H2, H6, H7, and H8 protons. The errors presented are coming purely from the fits

	$\frac{\Delta G_{310\,K}^{\dagger,glob}}{kJ \cdot mol^{-1}}$	$\frac{\Delta H^{\dagger,glob}}{kJ \cdot mol^{-1}}$	$\frac{\Delta S^{\dagger,glob}}{J \cdot mol^{-1} \cdot K^{-1}}$
CAAm^{5}CGTTG	68.0 ± 0.1	256 ± 2	606 ± 7
CATm^{5}CGATG	68.2 ± 0.1	239 ± 2	551 ± 8
CTTm^{5}CGAAG	67.8 ± 0.1	186 ± 4	380 ± 13

Based on our results, it can be concluded that the symmetric cytosine methylation of CpG:

a. increases the stability of the DNA double-helix;
b. slows down the DNA double-helix disruption;
c. affects at least three base pairs in both directions away from the methylated site.

8.4.2 A three-state process in CTTm^{5}CGAAG

Similarly to the non-methylated analogue (§ 7.7 on page 109), $\delta(T)$ of G5H8 in CTTm^{5}CGAAG was also subjected to the fit by the three-state models according to the schemes in Eq. (3.29) and (3.30), although only one ODN concentration is available (Fig. 8.12). In the case of the hypothet-

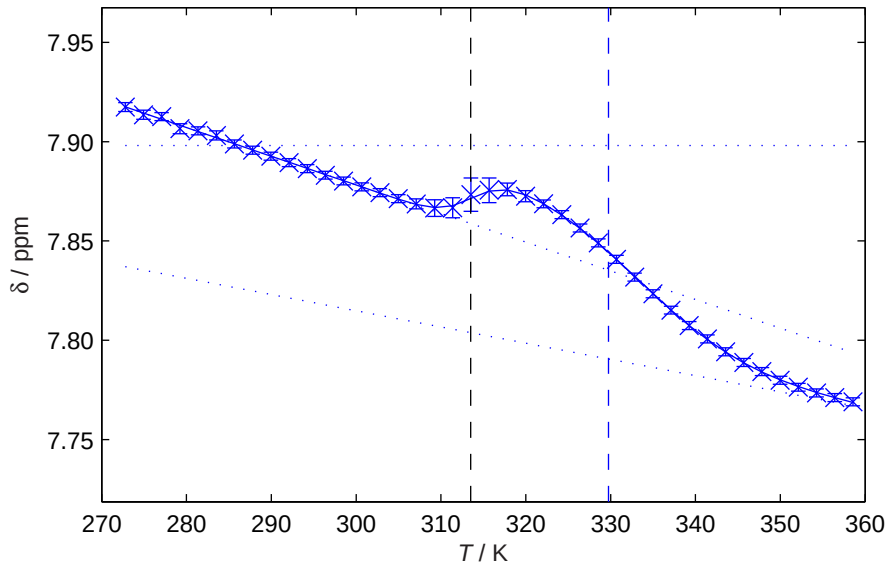


Figure 8.12 Chemical shifts of G5H8 in CTTm^{5}CGAAG together with their fit by the D_1–D_2–S three-state model (3.29). Dotted lines: chemical shifts of individual sites. Dashed vertical lines: transition temperatures of the two processes

ical D_1–D_2–S process (3.29), both the first and second transition move to higher temperatures by (4–5) K relative to the unmethylated ODN. This is consistent with the generally observed higher rigidity of the methylated double helix. Detailed description of the underlying thermodynamics requires additional experiments, at least the variable-concentration measurements, which are currently in progress.

8.5 Joint melting analysis by NMR, UV, and RS

In order to evaluate the melting parameters with lower mutual correlation than from a single melting curve, we measured the temperature-dependent UV and RS spectra of the methylated sequences as well as of their unmodified counterparts in a large concentration range.

The absorption at 260 nm was used as the melting curve and fitted by Eq. (3.25) with linear base lines x_A and x_B.

The Raman spectra show remarkable changes mainly in the intensities of certain bands. We extracted the main melting signatures by a SINGULAR-VALUE DECOMPOSITION (SVD, Fig. 8.13) of each VT data set

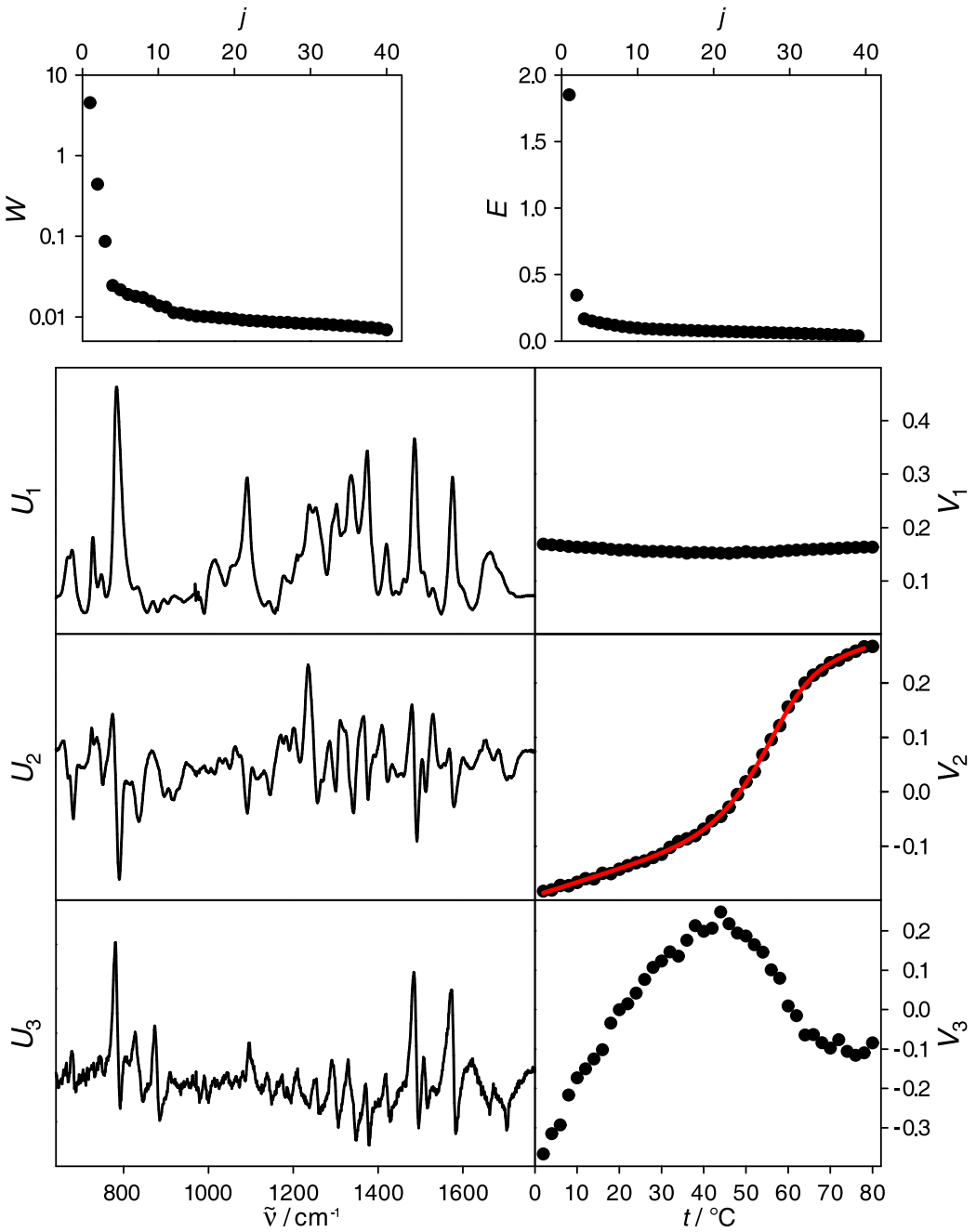


Figure 8.13 SVD of VT RS of CATCGATG. Top: statistical weights of the subspectrum j, W, and residual errors caused by considering j subspectra, E. Bottom: the first three subspectra, U_1, U_2, and U_3 (with the wavenumber $\tilde{\nu}$ on the horizontal axis), and their relative contributions, V_1, V_2, and V_3, to the experimental spectra acquired at the temperature t indicated. Solid red curve: fit of V_2

of the six samples. We performed the fits of the melting curve (3.25) with linear base lines to the scores of the second spectral component, V_2.

The results of UV and RS analyses, together with the global ^{1}H NMR values from Ha (Table 7.6 on page 100 and Table 8.7 on page 137), yielded T_m at a wide concentration window spanning almost three orders of magnitude of c. Their errors were estimated as sums of squares of the T_m errors coming from the underlying fits and the uncertainties in the temperature settings during the experiments. It allows us to quite reliably estimate the underlying changes in enthalpy and entropy and their methylation-caused differences by using Eq. (3.24) – the fit results are presented in Fig. 8.14 and Table 8.9. Interestingly, we found that upon methylation, the values of ΔH and ΔS of the duplex formation have risen in the case of TT preceding and AA following the CpG motif, they have remained unchanged in the AA–TT surrounding and have decreased for AT dinucleotide before and after CpG. Hence, the detailed thermodynamic reasons for the methylation-induced stabilisation of DNA double-helix vary in different sequence contexts.

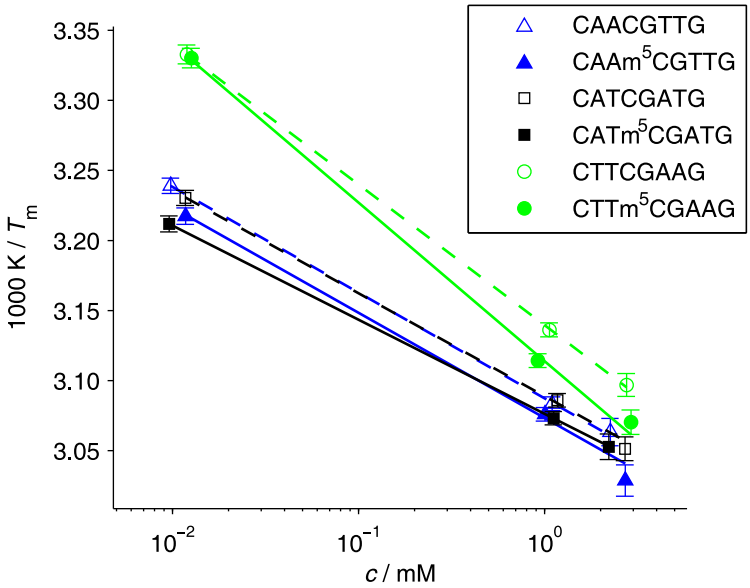


Figure 8.14 Van 't Hoff plot of $T_m(c)$. Lines: fits by Eq. (3.24). Empty symbols and dashed lines:unmethylated; filled symbols and solid lines: methylated samples

Table 8.9 Enthalpies and entropies of duplex formation obtained from joint analysis of NMR, UV, and RS. The methylation-induced changes in ΔH and ΔS are shown as $\Delta\Delta H$ and $\Delta\Delta S$, respectively

	ΔH / kJ mol^{-1}	$\Delta\Delta H$ / kJ mol^{-1}	ΔS / J mol^{-1} K^{-1}	$\Delta\Delta S$ / J mol^{-1} K^{-1}
CAACGTTG	-254 ± 11		-725 ± 35	
CAAm^{5}CGTTG	-255 ± 13	-2 ± 17	-727 ± 39	-1 ± 52
CATCGATG	-260 ± 11		-747 ± 36	
CATm^{5}CGATG	-285 ± 14	-24 ± 18	-819 ± 44	-72 ± 57
CTTCGAAG	-191 ± 7		-541 ± 23	
CTTm^{5}CGAAG	-169 ± 6	21 ± 9	-469 ± 19	71 ± 30

9 NON-CANONICAL DNA MOTIFS

9.1 Sequences based on serum response element

We employed various strategies to determine whether ODN molecules based on the serum response element (SRE) fold into a duplex or a hairpin (Fig. 2.9 on page 24). Substantial part of this study is formed by NOESY analysis and thermodynamic characterisation by variable-temperature ^{1}H NMR. Earlier, UV melting experiments were performed for selected ODN sequences [132].

In our laboratory, the NMR work on ODN sequences coming from one SRE strand was started during a master's thesis under my supervision [223] and it was further developed in this dissertation. Three ODN were studied (Table 9.1) at concentrations approximately 1 mM.

Table 9.1 ODN sequences based on SRE. Individual nucleotides are numbered from 5′- to 3′-end of the longest sequence, s1fos16. Sequences s1fos14 and s1fos12 are numbered starting from 2 and 3, respectively, so that the nucleotide numbers match across the three sequences

Label	Sequence
s1fos16	TGTCCATATTAGGACA
s1fos14	GTCCATATTAGGAC
s1fos12	TCCATATTAGGA
numbering	1 ... 16

9.1.1 Assignment of ^{1}H chemical shifts

The standard assignment procedures (§ 4.3.1 on page 40) were applied. For this purpose, it is not important whether the real structure adopted by the ODN is duplex or hairpin: although the NOE connectivities would be distinct in the two cases, the stem (the double-helical part formed by

the complementary bases) would be approximately shared (Fig. 2.9 on page 24).

Despite the basic sequential walks in NOESY were possible to follow unambiguously (Fig. 9.1), there were modulations in the cross-peak intensities in the central part of the ODN sequences compared to B-DNA. No thymine imino protons from the central segment have been observed. In addition, we failed to assign the H2 signals from the A6, A8, and A11 bases from NOESY. Therefore, ^{1}H–^{13}C HMBC experiments were run and successfully interpreted in the folded state (Fig. 9.2). Some adenine amino protons, most likely from A6, A8, or A11, also appear in the 1D ^{1}H spectrum but are not assigned with certainty due to the small number of their cross peaks and exchange with water.

The connectivities observed in the central part of s1fos16 are schematically depicted in Fig. 9.3. Several inter-residue cross peaks in NOESY are rather unexpected (especially A6H2–A8H1′ and A8H2–A11H3′) and favour the hairpin conformation. There is also a relative lack of inter-nucleotide cross peaks involving T9 and T10 nucleotides.

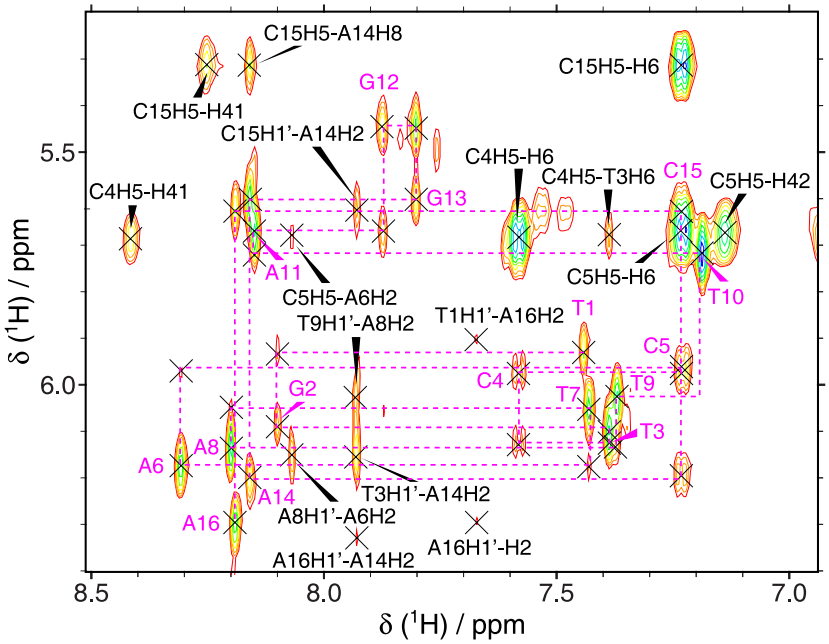


Figure 9.1 A portion of ^{1}H NOESY spectrum of s1fos16 (286 K). Sequential walk H1′–H6/H8 depicted by dashed magenta line with the intra-nucleotide cross peaks labelled. Other peaks fully annotated

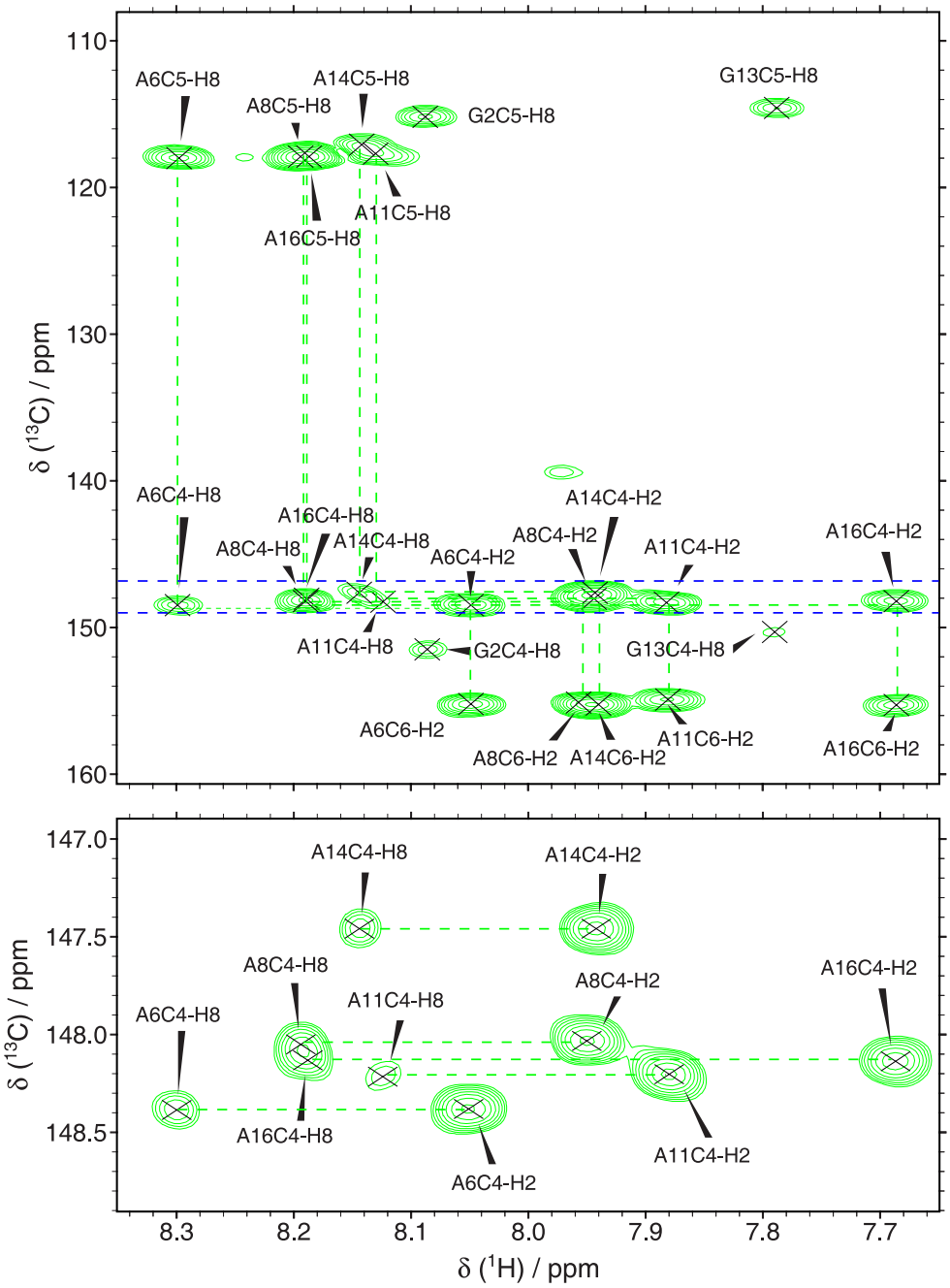


Figure 9.2 Parts of ^{1}H–^{13}C HMBC (top) and selective ^{1}H–^{13}C HMBC (bottom) spectra of s1fos16 at 298 K. The green dashed lines connect cross peaks from the same adenine bases. The blue dashed lines in the upper plot show the limits of the lower one

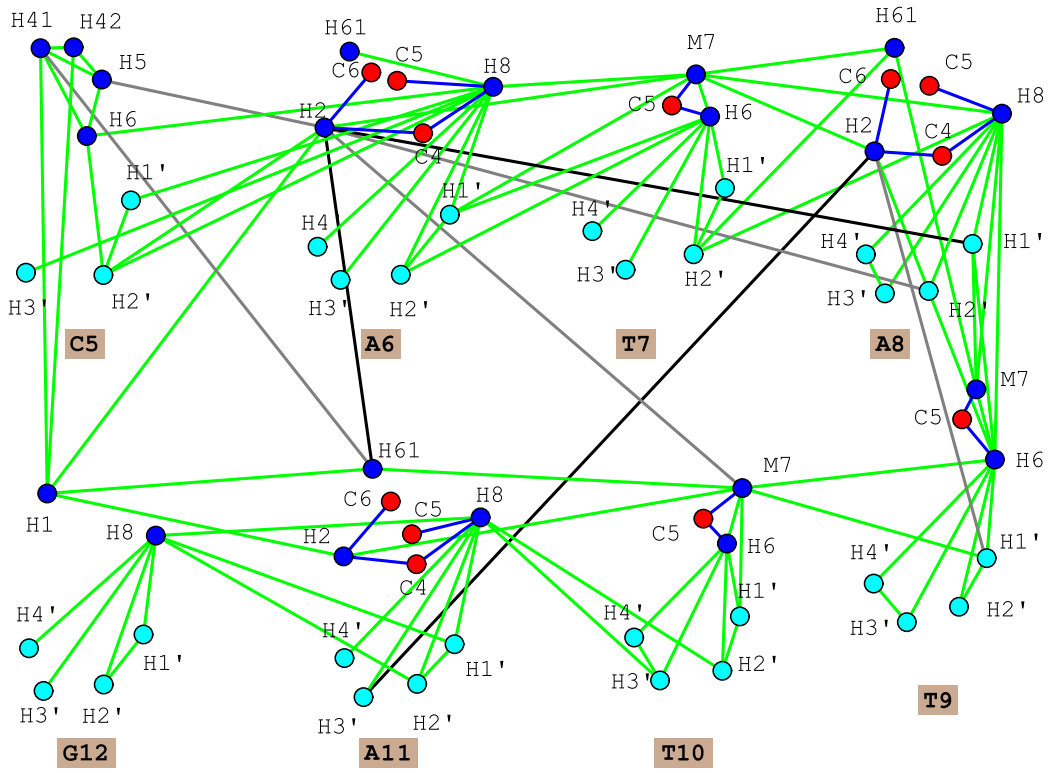


Figure 9.3 Diagram of connectivities in the central part of s1fos16 observed in ^{1}H NOESY (green) and in ^{1}H–^{13}C HMBC (blue) spectra. Nuclei are designated by blue (base protons), cyan (deoxyribose protons; H2″ omitted and not all H3′ and H4′ are shown), and red circles (carbons). Selected long-range NOESY peaks coloured by grey and black as weak and strong, respectively. AH61 not assigned with certainty. Created by Sparky [194]

In the VT ^{1}H spectra, no strong line broadening is observed. Therefore, tracking the signals to single strands brought no problems.

All ^{1}H chemical shifts of s1fos16 and the aromatic resonances of s1fos12 and s1fos14 are collected in § A.2.3 of Appendix (page A-13) for the folded and in § A.2.4 of Appendix (page A-15) for the unfolded state.

9.1.2 Comparison of chemical shifts with predictions

Differences between δ_{Pred} calculated by the NN model for double helix [163] and single strands [173, 184] and the experimental values δ_{Exp} are shown in Fig. 9.4.

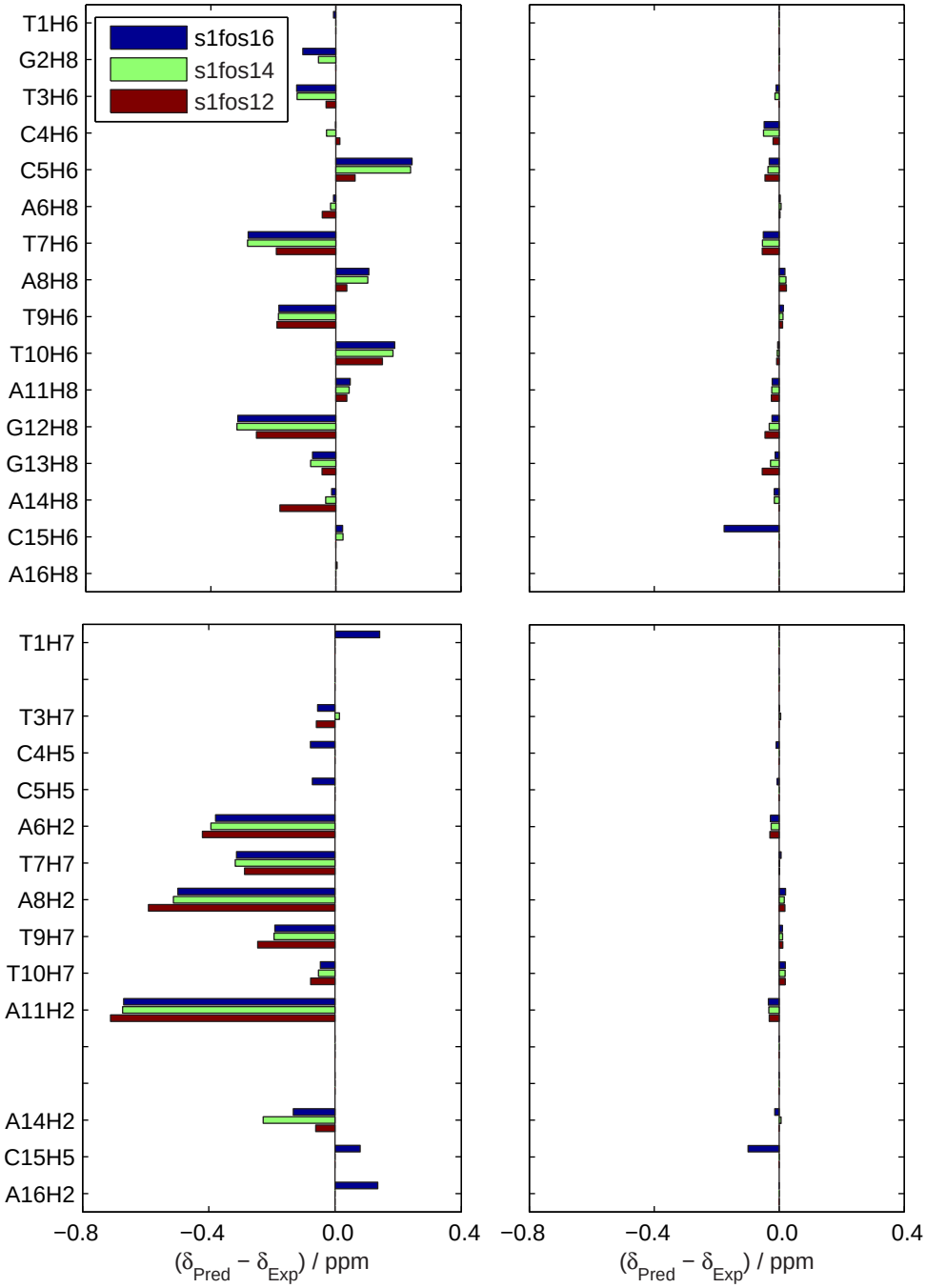


Figure 9.4 Differences between predicted, δ_{Pred}, and experimental chemical shifts, δ_{Exp}, of aromatic hydrogens in folded (298 K, left) and unfolded state (360 K, right). Top: H6 and H8, bottom: H2, H5, and H7

Since AH2 from the central part of the ODN sequences were not assigned successfully from NOESY and we cannot rely on the duplex prediction because of an unknown structure yet to be investigated, their chemical shifts in random coil are of a special importance. The prediction suggests that the order (from high to low chemical shift) would be A8H2, A6H2, A11H2. However, ^{1}H–^{13}C HMBC revealed that A8H2 and A6H2 are transposed. The difference between these shifts is 0.01 ppm at 360 K which is clearly resolvable so the predictions should always be taken with doubts in similar cases. Therefore, the experimental assignment by a method alternative to NOESY is invaluable in resolving ambiguities.

The predictions in the folded state are closer to the experiment in the self-complementary part of the ODN than in the central segment. This was expected to happen in both the hairpin and the mismatch-containing duplex. Corrections for the A6·A11 and T7·T10 mismatches, implemented from a combination of [187] and [188] (the uncertainties published around 0.07 ppm), were applied to δ_{Pred} of the central part of the ODN. The comparison with uncorrected predictions (Fig. 9.5) shows that better agreement with our experimental data was achieved at the boundary between the complementary region and the central segment than in the central part itself. This favours the possibility of formation of a hairpin over a mismatched duplex.

9.1.3 Line widths

At low temperatures, the s1fos16 line widths are comparable to the octamer DNA duplexes. This indicates that the molecular weights of the complexes are also similar. If 16-nucleotide duplexes were dominating in the s1fos16 solution, the line widths would be approximately doubled, using the simplest assumption that T_2^{-1} is proportional to the rotational correlation time τ_c given by the Einstein–Stokes relation for isotropic motions as

$$\tau_c = \frac{\eta V}{kT}, \tag{9.1}$$

where η is the solvent viscosity and V is the effective volume of the molecule [224]. Therefore, this leads to a preference for the hairpin hypothesis.

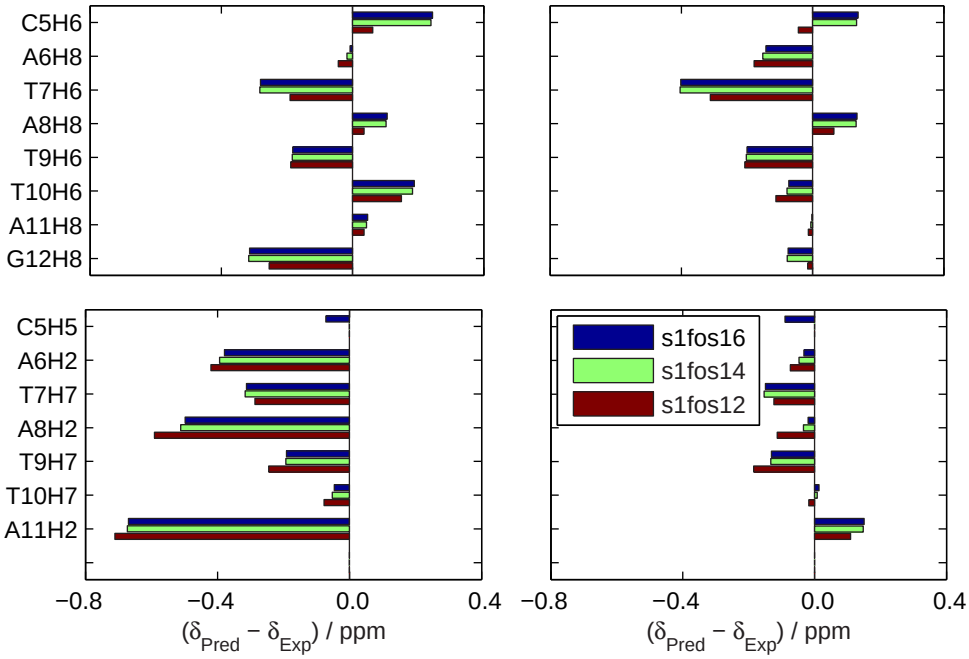


Figure 9.5 Differences between predicted, δ_{Pred}, and experimental chemical shifts, δ_{Exp}, in folded state (298 K) without (left) and with (right) corrections for the A6·A11 and T7·T10 mismatch base pairs

9.1.4 Temperature dependence of chemical shifts

Linear fits of $\delta(T)$ in s1fos14 (because of its mediocre T_{m} among the three ODN) in the folded and unfolded states have been performed in the temperature regions outside melting where a good linearity is observed. The slopes, shown in Fig. 9.6, mark unusual trends around the central part of the ODN: as opposed to B-DNA double helix, where all H6 and H8 in non-terminal nucleotides have negative slopes (§ 7.4.1 on page 82), we find positive slopes in C5H6, T9H6, and T10H6 in the folded state. Additionally, the H2, H5, and H7 protons having positive slopes in B-DNA show negative signs in A6H2, T7M7, and T10H7. In single strands, the shifts move downfield with increasing T which agrees with the findings for self-complementary DNA octamers.

We can conclude that there is some non-B-DNA structure in the middle part of the ODN s1fos14 (from C5 to A11). We cannot specify it more precisely because both the mismatched base pairs and a hairpin loop could be responsible, but the temperature sensitivities of δ clearly serve

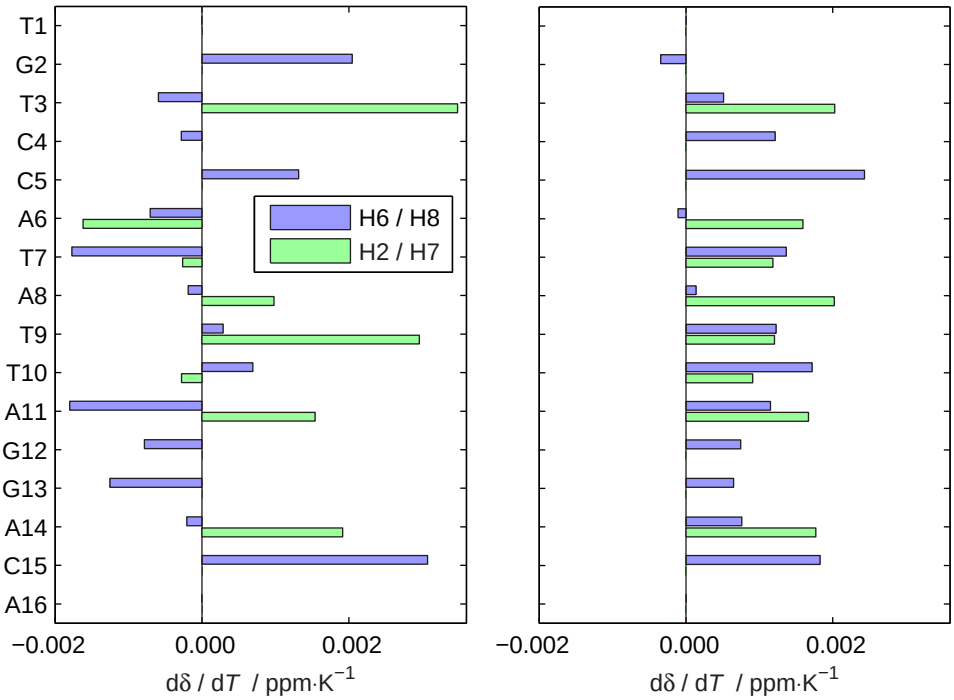


Figure 9.6 Slopes of chemical shifts with respect to T in folded (left) and unfolded state (right) of s1fos14

as a marker of an unusual motif. Interestingly, this feature is reflected also by the Watson–Crick C5·G12 base pair.

9.1.5 NMR melting curves

The thermodynamic stability is now evaluated by the analysis of ^{1}H NMR resonances. The peaks of H2, H6, H7, and H8 of all the three ODN sequences at all temperatures were fitted by Lorentzian curves because of the fast exchange. Additionally, the variable-temperature H1′ and CH5 signals were fitted for the longest ODN, s1fos16. Owing to relatively frequent overlaps in this spectral region, the fitting was performed with common amplitudes[1] of all the spectral lines at each temperature.

Global fits of $\delta(T)$ have been performed assuming two distinct folded states, hairpin and duplex, using Eq. (3.27) and Eq. (3.27), respectively. The fitted curves suit the experimental points quite neatly in both cases

[1]meaning the equilibrium magnetisations; due to dephasing by the spin echo this does not imply the integral intensities are the same

with comparable residuals. However, confrontation of the folding enthalpies and entropies obtained from the two models with the results from self-complementary octamer duplexes (Table 9.2) shows that they are too low in absolute values to be associated with melting of a duplex, which would have at least ten regular base pairs. Instead, ΔH and ΔS lie between four and five base pairs[2] which suits perfectly the idea of a hairpin formation.

Table 9.2 Folding enthalpies and entropies obtained from global fits of $\delta(T)$ or $p_{\mathrm{A}}(T)$ by hairpin and duplex melting models. Ha: H2, H6, H7, and H8. For SRE, $c = 1$ mM was assumed for the duplex model

	Hairpin model		Duplex model	
	ΔH^{glob} / kJ mol^{-1}	ΔS^{glob} / J mol^{-1} K^{-1}	ΔH^{glob} / kJ mol^{-1}	ΔS^{glob} / J mol^{-1} K^{-1}
SRE fragments, fits of $\delta(T)$				
s1fos16 – Ha	-137 ± 1	-409 ± 3	-211 ± 1	-576 ± 4
s1fos16 – H1′	-146 ± 2	-435 ± 6	-220 ± 3	-604 ± 8
s1fos14 – Ha	-129 ± 1	-397 ± 4	-196 ± 2	-549 ± 6
s1fos12 – Ha	-62 ± 2	-204 ± 7	-83 ± 3	-220 ± 10
Self-complementary DNA, fits of $p_{\mathrm{A}}(T)$				
CAACGTTG	-169 ± 3	-524 ± 8	-252 ± 4	-722 ± 12
CATCGATG	-170 ± 1	-525 ± 4	-261 ± 2	-748 ± 7
CTTCGAAG	-170 ± 3	-534 ± 8	-252 ± 4	-732 ± 12

We present the thermodynamic parameters obtained from the fits of individual $\delta(T)$ by the unimolecular model in § A.5.5 of Appendix (page A-49).

$T_{\mathrm{m}}^{\mathrm{glob}}$ determined from Ha of s1fos12, s1fos14, and s1fos16 are (305 ± 2) K, (325.8 ± 0.1) K, and (333.9 ± 0.1) K, respectively. Except for the shortest ODN which was found much less stable than predicted, these values agree within a 4 K precision with the NN estimates for hairpin melting and are at least 20 K higher than expected for duplex [52]. In addition, $T_{\mathrm{m}} = 330$ K of s1fos16 from UV absorption at $c = 3$ µM [132]

[2] supposing the average values $\Delta H = -32$ kJ $\cdot$ mol^{-1} and $\Delta S = -92$ J $\cdot$ mol^{-1} $\cdot$ K^{-1} per base pair from our data or $\Delta H = -35$ kJ $\cdot$ mol^{-1} and $\Delta S = -89$ J $\cdot$ mol^{-1} $\cdot$ K^{-1} per base pair from [39]

agrees well with the 1 mM T_m from NMR which marks a concentration-independent, therefore unimolecular process.

The residuals between the experimental data and the optimal global fits are shown in Fig. 9.7, Fig. 9.8, and Fig. 9.9. Overall, the deviations stay very small, lower than 0.01 ppm. Apart from the terminal nucleotides, the largest differences – similar to those in CTTCGAAG (Fig. 7.37 and 7.38 on pages 114 and 115) – are found for T7H6, T9H7, and A6H1′ located in the central segment. It seems that some minor unexpected transition occurs at lower temperatures rather than during the melting itself; removing the data points below 300 K remarkably reduces the residuals at the whole temperature range (Fig. 9.10).

What can we conclude from the above thoughts? First, the melting curves are more likely to represent a hairpin than a duplex disruption. Second, there is no clear violation of the two-state approximation. The folded state melts in a strongly cooperative way and no local deviations from the global behaviour can be seen during the temperature-induced unfolding. However, declinations from linear trends at low temperatures indicate that some structural rearrangements or formation of other complexes with lower relative populations occur at low temperatures.

9.1.6 Duplex or hairpin?

We demonstrate that the folded state is a hairpin rather than a duplex. Several reasons that have been described above lead to this conclusion:

a. in the central parts of the ODN molecules, the predictions of chemical shifts match our experimental values much worse than in the outer regions, even after the corrections for mismatches (Fig. 9.4);
b. we observe no resonances of the thymine imino protons from the central parts of the ODN sequences;
c. several long-range cross peaks in NOESY are incompatible with the double-helical arrangement (Fig. 9.3);
d. molecular weight of the folded s1fos16 is similar to the octamer duplexes as witnessed by the NMR line widths;
e. the extensive thermodynamic quantities describing the melting process correspond to disruption of 4 or 5 base pairs only;

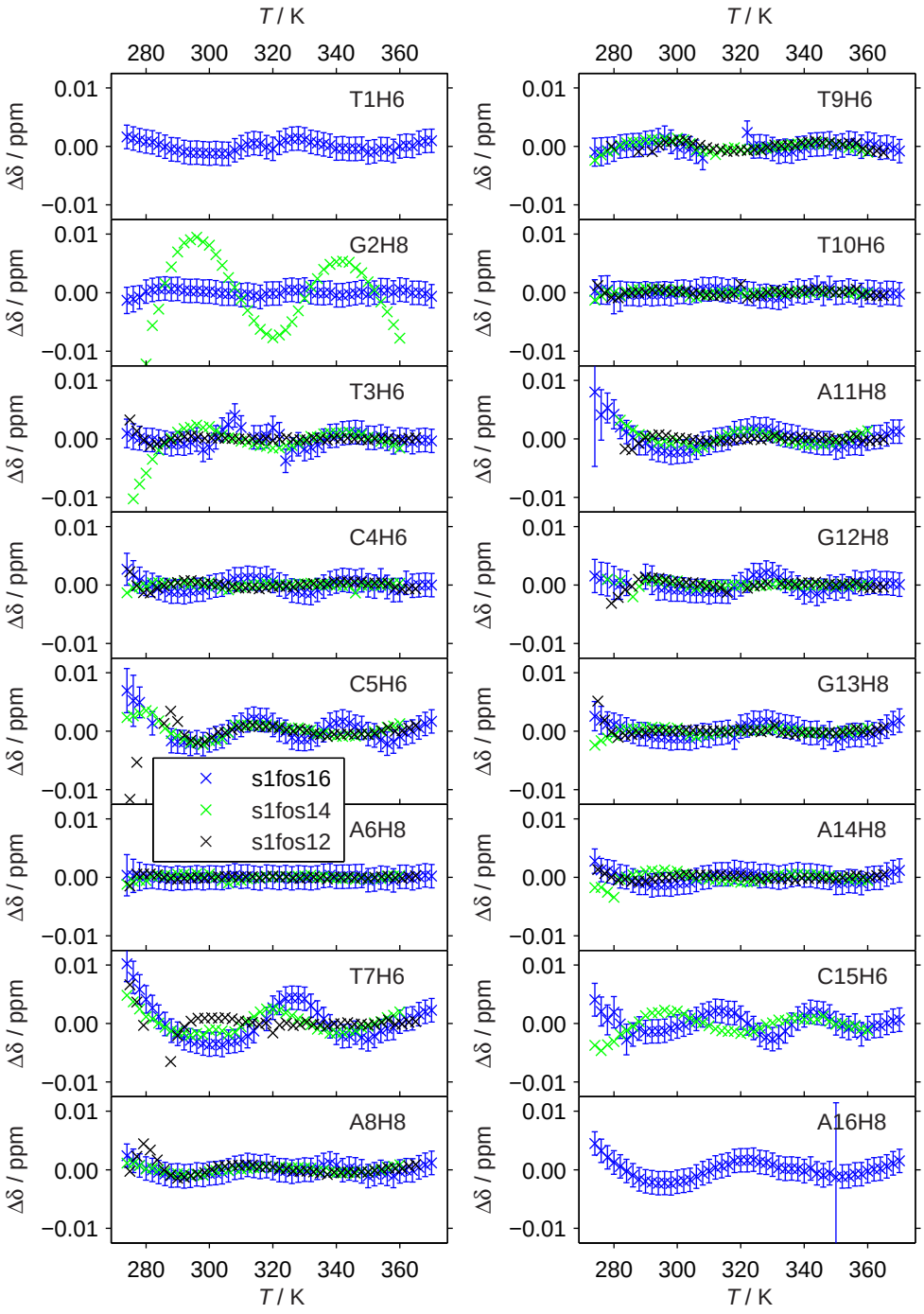


Figure 9.7 Residuals of chemical shifts of H6 and H8 from the global fits by an unimolecular process. Error bars shown only for s1fos16 for better clarity. All scales are the same

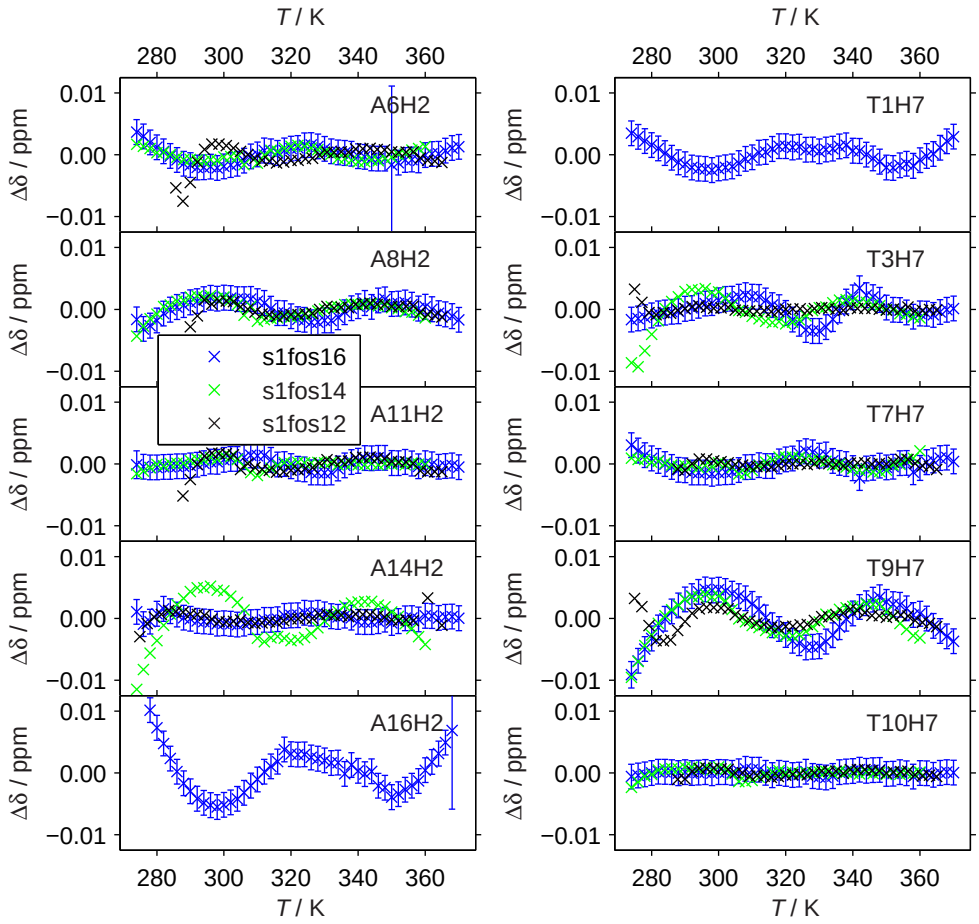


Figure 9.8 Residuals of chemical shifts of H2 (left) and H7 (right) from the global fits by an unimolecular process. Error bars shown only for s1fos16 for better clarity. All scales are the same

f. the stabilities predicted by DINAMelt web server [52] for hairpins agree with our results and are much higher than for the double helices;

g. very similar melting parameters of s1fos16 have been obtained by NMR ($c = 1$ mM) and UV absorption ($c = 3$ µM) [132].

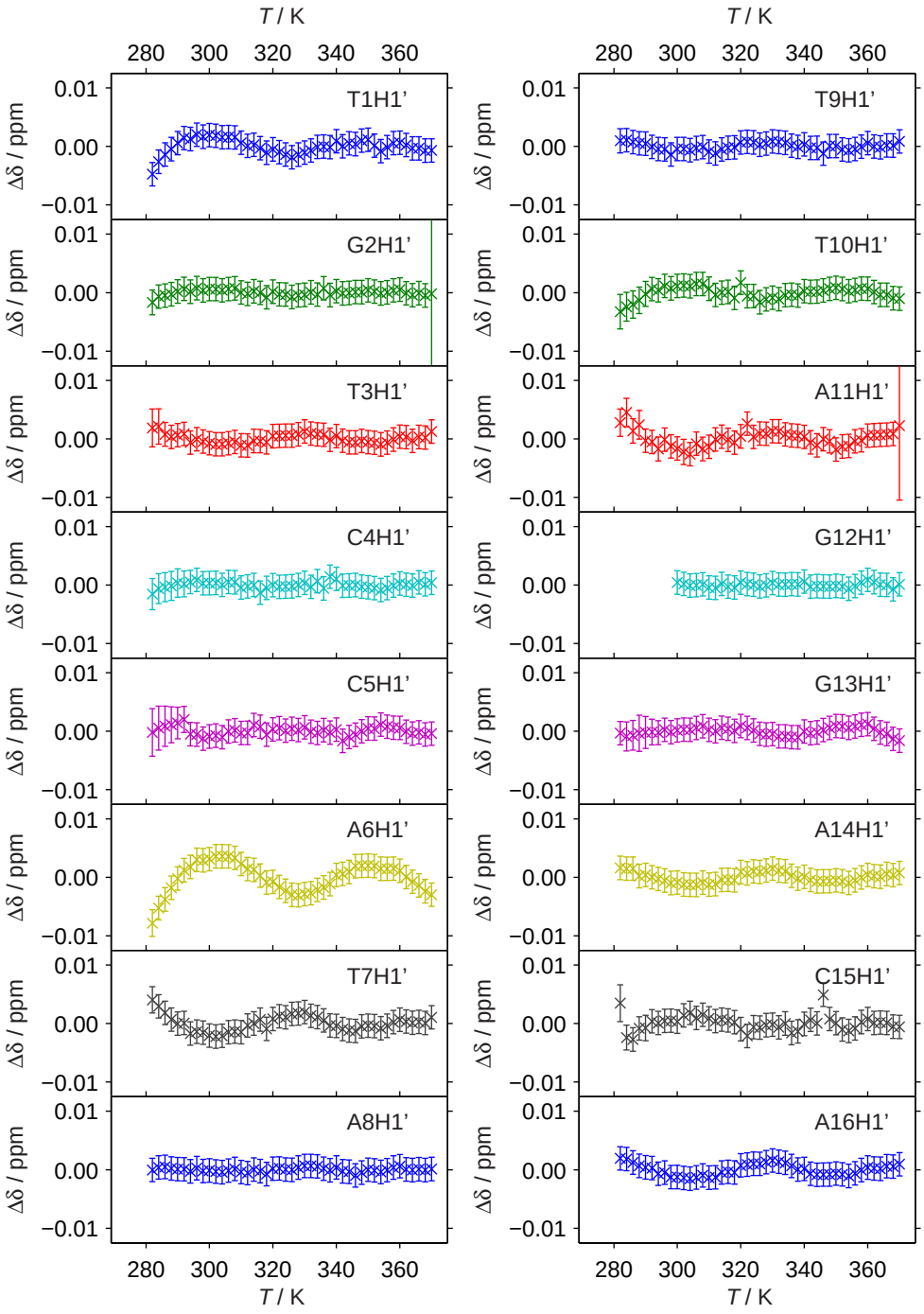


Figure 9.9 Residuals of H1′ chemical shifts of s1fos16 from the global fit by an unimolecular process. All scales are the same

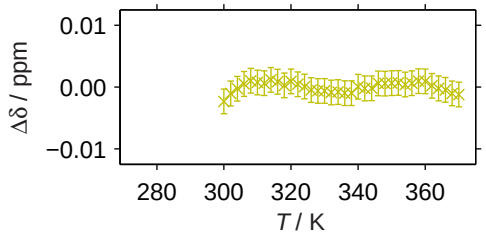


Figure 9.10 Residuals of δ of A6H1′ in s1fos16 from the global fit by an unimolecular process after deleting the points below 300 K. The vertical axis and the colour are the same as in Fig. 9.9

9.2 Self-assemblies of guanosine monophosphate

The abilities of riboguanosine 5′-monophosphate (rGMP) and of 2′-deoxyriboguanosine 5′-monophosphate (dGMP) to form multimolecular complexes in various conditions were studied by ^{1}H and ^{31}P NMR spectroscopy, conducted by the author of this thesis. These experiments were inspired by earlier results from Raman scattering spectroscopy, measured by the co-workers at the Institute of Physics of Charles University. Since the available literature on dGMP complexes is much less numerous than on the ribonucleotide counterpart, our work is focused more on the deoxynucleotide.

It was shown by RS that the sodium salt of rGMP (Na_2rGMP) in water solution with concentration above 250 mM forms supramolecular complexes at ambient temperature. Vibrational band around 1730 cm^{-1} [225] marks the presence of G-tetrads. Contrary to rGMP, only traces of Na_2rGMP assemblies have been revealed. The complexes are greatly stabilised when K^+ is added and the potassium-induced changes in RS are very similar in rGMP and dGMP.

9.2.1 Sodium solutions

Even at concentration of Na_2dGMP as high as 800 mM, ^{1}H and ^{31}P NMR spectra show small changes (Fig. 9.11). Based on spectral integration, we estimated that only around 20 % of the molecules are involved in the supramolecular complexes after lowering the temperature down to 278 K. This confirms the finding from RS that dGMP in sodium solutions forms only weak assemblies.

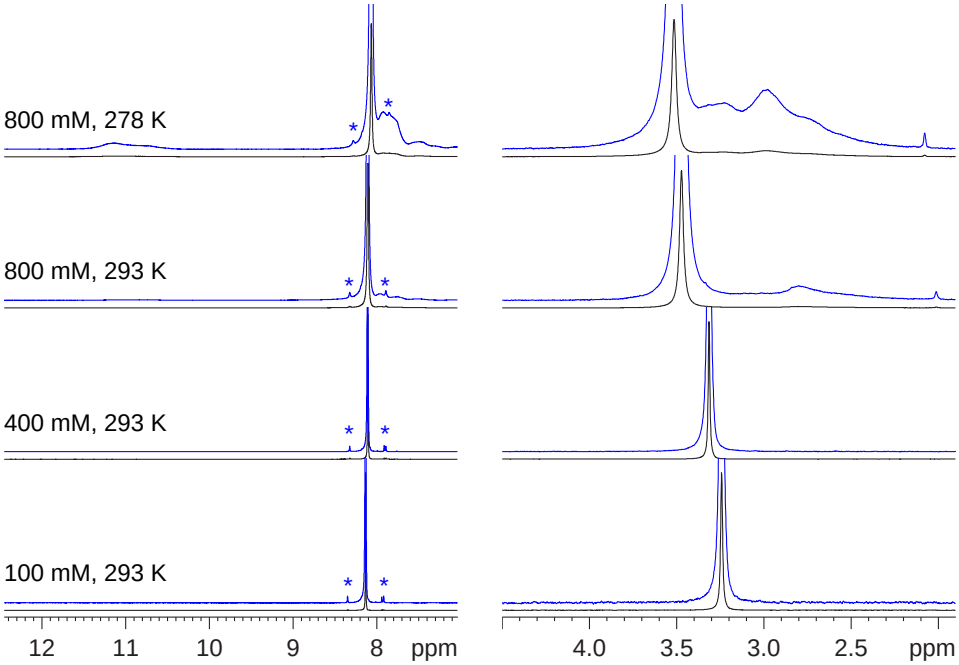


Figure 9.11 NMR spectra of Na_2dGMP. Concentrations and temperatures are indicated in the plot. Intensities normalised to the maxima. Spectra in blue have ten-fold magnified intensity scale relative to black. Left: ^{1}H spectra showing imino and H8 resonances; asterisks, *, mark ^{13}C-coupling satellites. Right: ^{31}P spectra with ^{1}H decoupling

9.2.2 Potassium effect

Addition of 400 mM KCl to the 400 mM Na_2dGMP solution induces strong complex formation clearly visible in NMR (Fig. 9.12). Less than a half of the dGMP molecules remains free. Several overlapping peaks in NMR show that there are more nonequivalent positions in the structure, comprising comparable fractions of the individual molecules. This is in accordance with the expected presence of multiple stacked G-tetrads in the assembly whose number is limited because the outer segments still constitute a significant portion. Imino signals appear in the ^{1}H spectrum, proving the formation of stable hydrogen bonds within the G-tetrads.

While such dramatic changes are invoked by KCl, control experiments with addition of equimolar NaCl instead had a negligible impact on the spectra (Fig. 9.12).

We obtained qualitatively different NMR spectra of the potassium-

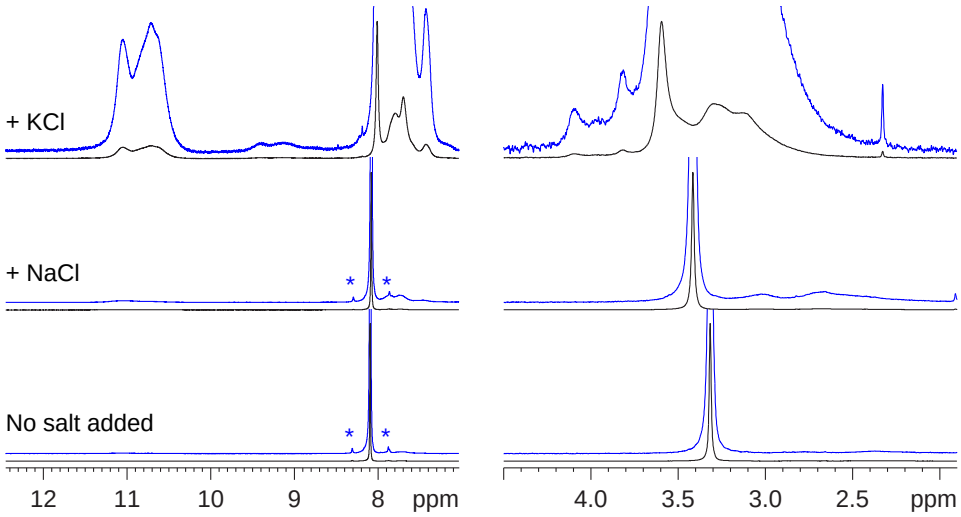


Figure 9.12 NMR spectra of 400 mM Na_2dGMP at 278 K without and with addition of 400 mM salt as indicated in the plot. Intensities are normalised to the maxima. Spectra in blue have ten-fold magnified intensity scale relative to black. Left: ^{1}H spectra showing imino and H8 resonances; asterisks, *, mark ^{13}C-coupling satellites. Right: ^{31}P spectra with ^{1}H decoupling

containing dGMP and rGMP solutions[3] (Fig. 9.13). On the other hand, the Raman spectra were very similar in the two cases. This could mean that while the local conformational parameters observed by the vibrational spectra are pertained, the mutual arrangements of the individual nucleotides and stacking configuration of the G-tetrads, to which NMR lines are sensitive, are changed. Furthermore, alternative folds with lower populations are distinguished in the ^{1}H and ^{31}P NMR spectra of rGMP.

9.2.3 Differences between rGMP and dGMP

Our results arising from the combination of two spectroscopic techniques reveal that there are substantial differences in the sensitivity of rGMP and dGMP to the cation type. Previous observations made by RS have been supported and extended by ^{1}H and ^{31}P NMR:

a. although there is a strong affinity to complex formation in sodium solutions of rGMP, the deoxynucleotide hardly makes any supra-

[3] the conditions vary slightly

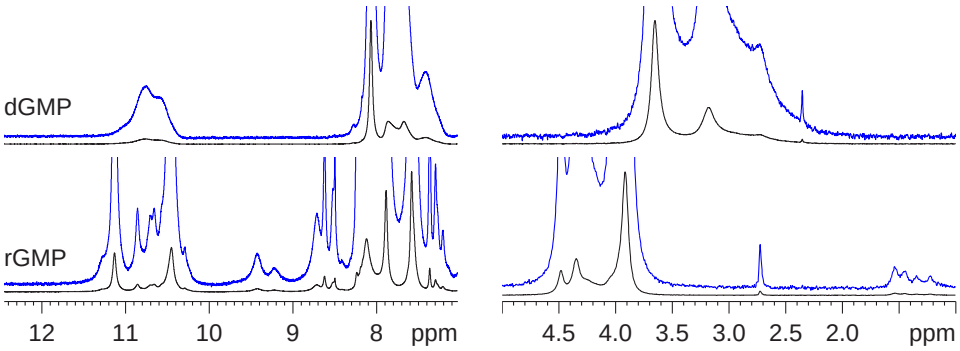


Figure 9.13 NMR spectra of 400 mM Na_2dGMP with 400 mM KCl at 293 K (top) and of 400 mM K_2rGMP at 298 K (bottom). Intensities are normalised to the maxima. Spectra in blue have ten-fold magnified intensity scale relative to black. Left: ^{1}H spectra showing imino and H8 resonances. Right: ^{31}P spectra with ^{1}H decoupling

molecular assemblies in Na^+, unless its high concentration and low temperature are set;

b. both rGMP and dGMP readily create stable G-tetrad-containing complexes in K^+ solutions, but their structures are different.

The astonishing variability of the supramolecular assemblies of the simple nucleotides may be connected to the general polymorphism of G-quadruplexes formed by RNA and DNA oligonucleotides.

10 THIOBENZANILIDE

This chapter deals with an *E*–*Z* conversion observed in a class of secondary thiobenzanilides. All the samples were prepared and their NMR spectra as well as other experiments were taken by colleagues of the Department of Inorganic and Organic Chemistry, Faculty of Pharmacy in Hradec Králové of Charles University. My task was to analyse the acquired NMR spectra with respect to the undergoing chemical exchange, using the strategy originally developed for the melting of nucleic acids. The thermodynamic characterisation of the equilibrium and kinetics of the transition between *E* and *Z* conformations of a selected compound was then confronted with the results of quantum-chemical calculations performed by Václav Profant from the Institute of Physics of Charles University, Faculty of Mathematics and Physics.

10.1 Conformers of secondary thiobenzanilides

Various derivatives of 2-methoxy-2′-hydroxybenzanilides were tested for their antimycobacterial activity. ^{1}H NMR spectra of thiobenzanilides present among these substances contained two sets of resonances which were unambiguously attributed to two distinct orientations around the partially double CS—NH bond [226]. A detailed study of 5-chloro-N-(5-chloro-2-hydroxyphenyl)-2-methoxybenzothioamide (compound **1**) followed. Assignment of the resonances, based on a ^{1}H NOESY spectrum, proved that there is an undergoing *E*–*Z* transition (Fig. 10.1), which is accompanied with chemical exchange observable in ^{1}H and ^{13}C NMR spectra at 11.7 T. The *Z* conformer is more populated than *E*.

It was speculated that intramolecular hydrogen bonds could stabilise the *Z* conformer. A set of derivatives with non-polar moieties instead of the methoxy and hydroxy groups were prepared with the aim of verifying or rejecting this possibility. The conformational equilibrium hardly

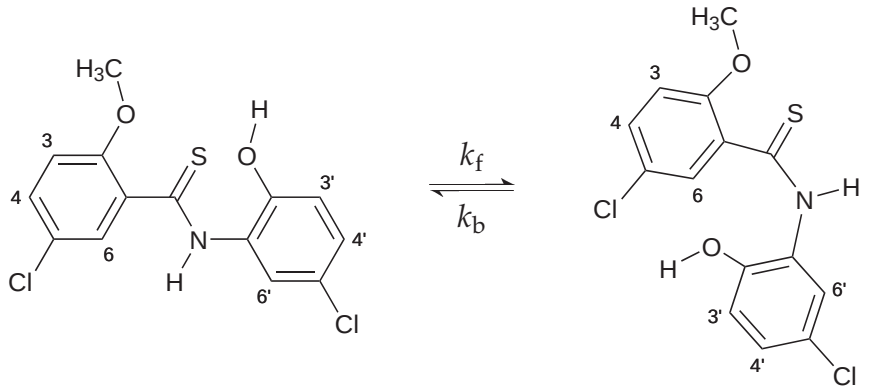


Figure 10.1 Equilibrium between *Z* (left) and *E* (right) conformation around the CS—NH bond of **1**. The rate constants of the forward and backward process and the atom numbering are indicated

changed. Therefore, it is the steric hindrance of the *ortho* substituents rather than a hydrogen bond which is the most important source of the observed properties of the rotation around the CS—NH bond.

10.2 Characterisation of the *E*–*Z* transition by ^{1}H NMR

In order to address the thermodynamics of the process under study, the ^{1}H spectra of **1** in deuterated dimethyl sulfoxide, DMSO-d_6, were acquired in a broad temperature range (Fig. 10.2). The spectral series brings qualitative changes: a slow exchange is present and two sets of resonances are distinguished at lower temperatures, whilst the lines broaden when increasing T and the minor species is no longer resolved above some point. The general procedure of line-shape analysis during the two-site chemical exchange (§ 6.2 on page 56) was followed, but several modifications were needed.

For each temperature, the fitting was done simultaneously for H3, H4, H6, H3′, H4′, H6′, and OMe resonances with common equilibrium constant of the forward process from Fig. 10.1, K_f, and exchange rate, k_f. Resonances of NH and OH protons were excluded from the analysis because of their exchange with water.

The line widths in the terms of T^*_{2A} and T^*_{2B} of the two exchanging sites could not be estimated directly because the exchange modifies the lines

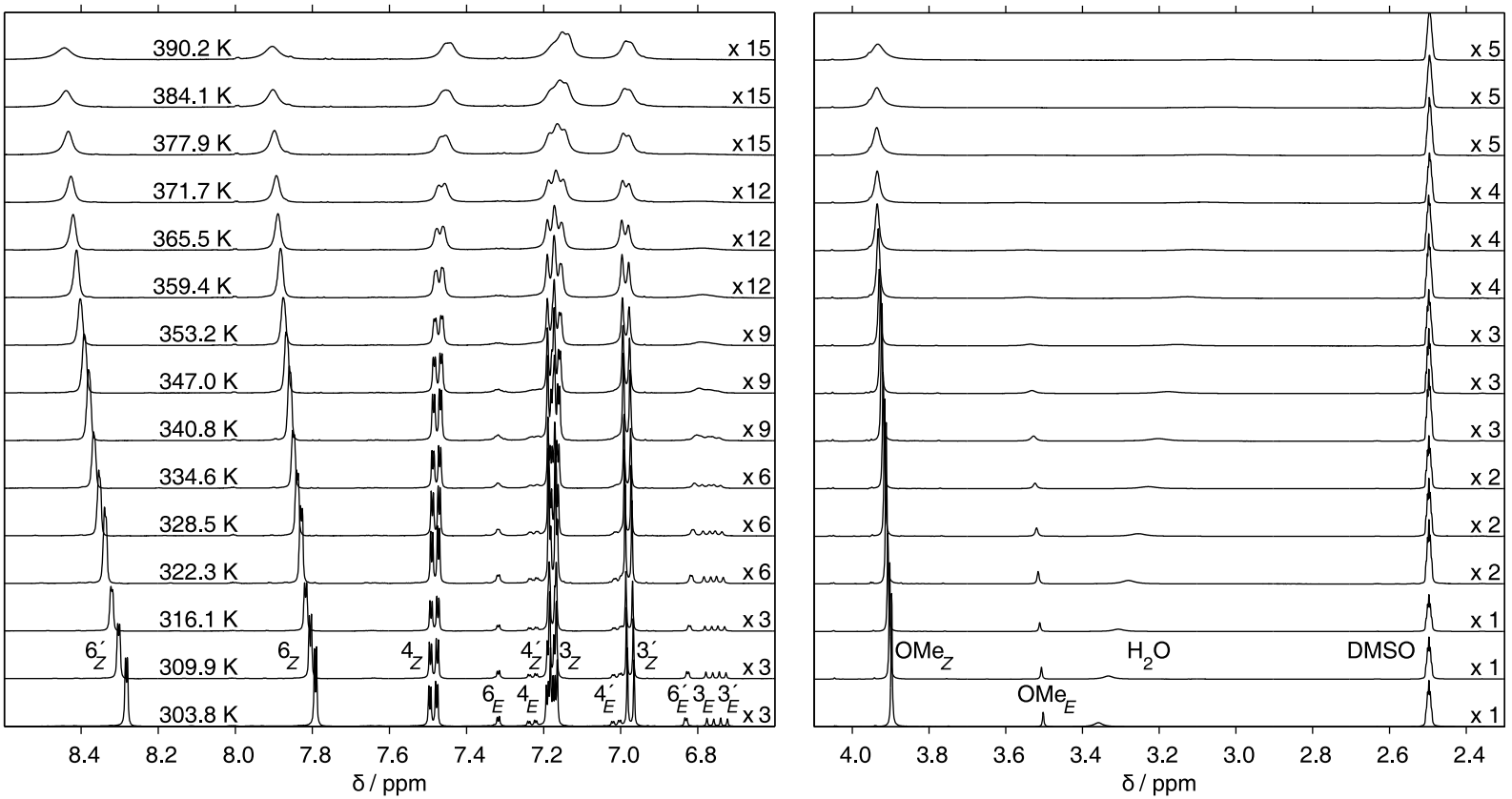


Figure 10.2 Aromatic (left) and aliphatic (right) parts of variable-temperature 500 MHz ^{1}H NMR spectra of **1** in DMSO-d_6. The vertical scales are changing with increasing temperature to improve the clarity of the broadening resonance lines as indicated in the plots. The scales differ three times in the two spectral regions because of the methyl groups. Assignment of the resonances is shown and subscripts denote the two conformers

even for the lowest temperature measured. We followed a protocol based on [224]: possible rates k_f at 316.1 K were estimated by repeated fitting with different fixed values of k_f during which all the other parameters (δ_A, δ_B, T^*_{2A}, and T^*_{2B} of all peaks and the common K_f) were optimised. Visual inspection of the results, comparison of chi-square, and rejection of relaxation times that were too high yielded minimal and maximal possible rates $k_{min} = 0.001\ s^{-1}$ and $k_{max} = 0.5\ s^{-1}$. Optimal $k_{opt} = 0.2\ s^{-1}$ served for determining the corresponding values of T^*_{2A} and T^*_{2B} of all the peaks. These were used in all further analysis as fixed parameters; values obtained for k_{min} and k_{max} were taken as their confidence limits.

In the spectra up to 328.5 K, where the exchange is in a slow regime and two relatively narrow peaks can be distinguished for each pair of exchanging nuclei, δ_A and δ_B were optimised by the fitting procedure, along with the exchange parameters, K_f and k_f. For higher temperatures, severe broadening of the resonance lines of E occurs. This, together with the overlap of $H3_E$, $H3'_E$, and $H6'_E$, disables independent fitting of the chemical shifts of this conformer. Therefore, their values were fixed to linear extrapolations from the lower temperatures, while the chemical shifts of the hydrogen nuclei belonging to Z were still taken as free parameters.

The functions $K_f(T)$ and $k_f(T)$ obtained from the line-shape analysis were fitted by van 't Hoff (3.6) and Eyring (3.8) equations for a first-order process, respectively (Fig. 10.3). The following values of the thermodynamic quantities describing the transition of **1** from Z to E were determined:

$$\begin{aligned}
\Delta H &= (-3.1 \pm 1.3)\ \mathrm{kJ \cdot mol^{-1}},\\
\Delta S &= (-26.9 \pm 3.6)\ \mathrm{J \cdot mol^{-1} \cdot K^{-1}},\\
\Delta G_{310\,K} &= (5.2 \pm 0.3)\ \mathrm{kJ \cdot mol^{-1}},\\
\Delta H^{\dagger} &= (77.7 \pm 1.7)\ \mathrm{kJ \cdot mol^{-1}},\\
\Delta S^{\dagger} &= (-12.5 \pm 4.6)\ \mathrm{J \cdot mol^{-1} \cdot K^{-1}},\\
\Delta G^{\dagger}_{310\,K} &= (81.6 \pm 0.4)\ \mathrm{kJ \cdot mol^{-1}}.
\end{aligned}$$

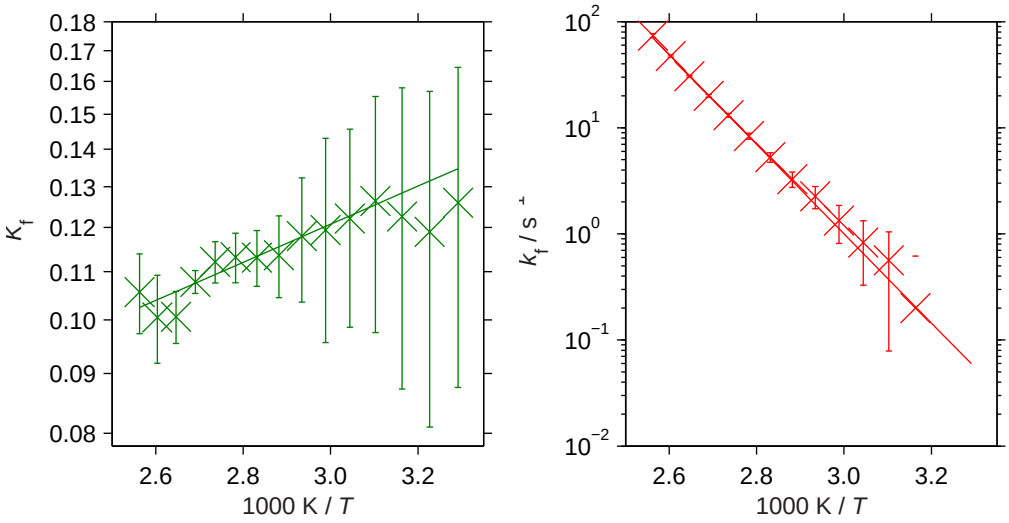


Figure 10.3 Linearised van 't Hoff (left) and Eyring (right) plots of the temperature dependence of the experimental values and errors of equilibrium and rate constants, respectively, and their fits

10.3 Comparison with quantum-chemical calculations

The DFT calculations with one explicit DMSO molecule and implicit solvent around showed that there is no intramolecular hydrogen bond in **1** which supports the experimental finding. Instead, there is a hydrogen bridge between the thioamide hydrogen and the DMSO oxygen.

Free rotation of the two aryl moieties was revealed in the *Z* conformer. This freedom is much limited in *E* due to steric hindrance between the substituents on the rings but the barriers still correspond to a relatively high reorientation rate above 10^5 s^{-1}. Such results are in agreement with the entropy loss observed experimentally. The enthalpy decrease can be explained by the π–π stacking interaction which is, however, too weak to overweight the entropic penalty at the ambient temperatures.

An energy change 4.1 kJ · mol^{-1} for the *Z* to *E* transition, obtained as the difference between population-weighted mean DFT energies of the two conformers, is near the free-energy change, $\Delta G_{310\,\mathrm{K}}$, coming from VT NMR. Two transition states of the *Z*–*E* equilibrium were found to pose similar activation barriers of 80.8 kJ · mol^{-1} in average – again, a nice match with the NMR-determined value was achieved.

10.4 Synergy of the chemical treatment, NMR, and calculations

By combination of chemical modifications, variable-temperature ^{1}H NMR spectroscopy, and detailed quantum-chemical DFT calculations, we revealed that 5-chloro-N-(5-chloro-2-hydroxyphenyl)-2-methoxybenzothioamide (**1**) undergoes a *Z*–*E* transition and that:

a. no intramolecular hydrogen bond stabilises either conformer;
b. the *Z* conformer is more populated at a wide temperature range due to its higher entropy arising from free rotations of the aryl groups;
c. moderately high activation barrier permits the *Z*–*E* interconversion around the partially double thioamide bond;
d. consensus was reached in quantitative thermodynamic analysis of the transition by NMR and DFT.

11 CONCLUSIONS

This dissertation is devoted to detailed NMR studies of short nucleic-acid segments. The structural variability of their intra- and intermolecular complexes in aqueous solutions was investigated under various environmental conditions such as temperature, oligonucleotide concentration, or type and concentration of metal counter-ions. The obtained results clearly show that the NMR technique allied with detailed analysis of the positions and shapes of the resonance lines provides – even in the case of the commonly used magnetic field (500 MHz spectrometer) – valuable information that is often not achievable by other methods. It comprises not only the basic structural data but also the equilibrium and kinetic parameters of temperature-induced conformational transitions, above all unfolding of the complexes. These thermodynamic quantities can be obtained independently for numerous sites along the oligonucleotide chain, thus providing insight into the physical processes on the level of individual nucleotides. The NMR results represent then a priceless complement to the outcomes of other methods such as electron or vibrational spectroscopy.

The temperature dependence of ^{1}H chemical shifts outside the temperature region of the folding–unfolding transition was shown to be an important NMR characteristic. At first, the magnitude of the chemical-shift changes (about 0.01 ppm per 10 K) indicates the necessity to specify the temperature whenever the values of oligonucleotide ^{1}H chemical shifts are presented. Next, the slope of the ^{1}H chemical-shift change was found to depend remarkably on the oligonucleotide sequence, being sensitive not only to the nearest neighbours of the nucleoside in question but also to the next-nearest and even to more distant nucleosides. The only exceptions are terminal nucleosides where the chemical-shift slopes are surprisingly well conserved for equal terminal doublets.

A particular attention was paid to analyse the NMR spectra in the area of the temperature-induced transition. The relation between the transition properties and the spectral shape had been verified in my previous work [154]. This doctoral thesis demonstrated the applicability of the approach also in the case of a *Z–E* conformational transition in a secondary thiobenzanilide, where the effect of the transition on the ^{1}H NMR corresponded to a slow exchange regime within a large temperature range. NMR results revealed higher entropy of the *Z* conformer as the reason for its preferential population and a moderately high activation barrier.

In the case of oligodeoxynucleotides, duplex formation and dissociation occurs in a medium to fast exchange regime when investigated using the 500 MHz spectrometer. The melting transition is manifested by gradual changes of positions of resonance lines. Moreover, the ^{1}H NMR line widths allowed determining the activation parameters from individual aromatic hydrogens of the studied series of self-complementary octamers. It can be assumed that the importance of the line-shape analysis during melting would be even greater if the nucleic acids were studied in higher magnetic fields, which relatively slow down the chemical-exchange regime.

Besides the findings mentioned above, this thesis has brought about further results, mainly:

a. Systematic study of a set of eight self-complementary DNA octamers allowed considering the applicability of the ^{1}H chemical-shift predictions based on the nearest-neighbour model. Although the predictions offered rough estimates, numerous differences exceeding 0.1 ppm were found for duplexes. This concerns mainly the terminal nucleosides, but the differences often exceeded the published deviations in the internal nucleotides, too, showing rather complicated sequence effects.
b. Several duplexes possess a surprisingly strong dependence of ^{1}H chemical shifts on concentration, probably suggesting an end-to-end duplex stacking *via* terminal C·G base pairs.
c. For all the studied octamers, a long-range cooperativity in the DNA duplex melting and no fraying of the ends were observed. Although the spatial structures of all the double helices were close to

the canonical B-form according to the standard patterns found in NOESY spectra, several anomalies in the local structure were found: above all, the octamers with a central CpG motif are remarkably sensitive to the remote sequence context. In the case of CTTCGAAG, even a three-state duplex melting was revealed, most probably caused by a local conformational transition in the double-helical state.

d. Symmetric methylation of the central CpG motif was found to increase the duplex stability, the effect being equally transmitted at least three base pairs away. The joint treatment of the melting parameters obtained by NMR, UV absorption, and Raman scattering in a wide concentration range revealed that the bases of the nearest and next-nearest neighbours of the CpG motif strongly influence the thermodynamic parameters of the duplex formation, particularly the entropic penalty.

e. It was possible to distinguish unambiguously the duplex dissociation from a single-molecular conformational transition and it was confirmed that the single-strand transcription element SRE forms a stable hairpin loop.

f. The results obtained from ^{1}H and ^{31}P NMR spectra of rGMP and dGMP in aqueous solutions supported the conclusions from Raman experiments that dGMP can form supramolecular tetrads only in the presence of potassium ions, whereas rGMP forms stable tetrads also in the case of sodium cations. The NMR experiments indicated structural differences between the tetrads formed by the two kinds of mononucleotides, which may explain the different stabilisation effect of particular counter-ions. This experience has been later employed in an NMR study of concentration-dependent polymorphism of human telomeric G-quadruplexes d[$G_3(TTAG_3)_3$] and d[$AG_3(TTAG_3)_3$], which is currently in progress.

To conclude, numerous findings indicate that the ability of individual nucleotides to create various types of differently strong non-covalent bonds to other nucleotides or other molecules in the environment is a key aspect of nucleic-acid polymorphism. These bonds are realised according to the nucleotide position in a particular sequence depending not only on

the nearest neighbours but also on the long-range sequence composition, which takes place even in the case of relatively simple segments of nucleic acids.

BIBLIOGRAPHY

[1] Watson, J. D. and Crick, F. H. C., *Molecular structure of nucleic acids*, Nature **171** (1953), no. 4356, 737–738, DOI 10.1038/171737a0.

[2] Rich, A., *DNA comes in many forms*, Gene **135** (1993), no. 1-2, 99–109, DOI 10.1016/0378-1119(93)90054-7.

[3] Allen, S., Blackburn, G. M., Egli, M., Fisher, J., Flavell, A. J., Gait, M. J., Haq, I., Laughton, C., Loakes, D., Luisi, B., Pyle, A. M., Stollar, E., Williams, D. M., and Williams, N. H., *RNA Structure and Function*, In: Nucleic Acids in Chemistry and Biology (Blackburn, G. M., Gait, M. J., Loakes, D., and Williams, D. M., eds.), The Royal Society of Chemistry, Cambridge, 3rd ed., 2006, ISBN 978-0-85404-654-6, ch. 7, pp. 253–293, DOI 10.1039/9781847555380-00253.

[4] Lee, R. C., Feinbaum, R. L., and Ambros, V., *The C. elegans heterochronic gene* lin-4 *encodes small RNAs with antisense complementarity to* lin-14, Cell **75** (1993), no. 5, 843–854, DOI 10.1016/0092-8674(93)90529-Y.

[5] Fire, A., Xu, S., Montgomery, M. K., Kostas, S. A., Driver, S. E., and Mello, C. C., *Potent and specific genetic interference by double-stranded RNA in* Caenorhabditis elegans, Nature **391** (1998), no. 6669, 806–811, DOI 10.1038/35888.

[6] Gregory, T. R., *The C-value enigma in plants and animals: a review of parallels and an appeal for partnership*, Annals Bot. **95** (2005), no. 1, 133–146, DOI 10.1093/aob/mci009.

[7] Allen, S., Blackburn, G. M., Egli, M., Fisher, J., Flavell, A. J., Gait, M. J., Haq, I., Laughton, C., Loakes, D., Luisi, B., Pyle, A. M., Stollar, E., Williams, D. M., and Williams, N. H., *Genes and Genomes*, In: Nucleic Acids in Chemistry and Biology (Blackburn, G. M., Gait, M. J., Loakes, D., and Williams, D. M., eds.), The Royal Society of Chemistry, Cambridge, 3rd ed., 2006, ISBN 978-0-85404-654-6, ch. 6, pp. 209–250, DOI 10.1039/9781847555380-00209.

[8] Berger, S. L., Kouzarides, T., Shiekhattar, R., and Shilatifard, A., *An operational definition of epigenetics*, Genes Dev. **23** (2009), no. 7, 781–783, DOI 10.1101/gad.1787609.

[9] Bird, A., *Perceptions of epigenetics*, Nature **447** (2007), no. 7143, 396–398, DOI 10.1038/nature05913.

[10] Bernstein, B. E., Meissner, A., and Lander, E. S., *The Mammalian Epigenome*, Cell **128** (2007), no. 4, 669–681, DOI 10.1016/j.cell.2007.01.033.

[11] Neidle, S., *The building-blocks of DNA and RNA*, In: Principles of Nucleic Acid Structure, Academic Press, Amsterdam, 2008, ISBN 978-0-12-369507-9, ch. 2, pp. 20–37.

[12] Markley, J. L., Bax, A., Arata, Y., Hilbers, C. W., Kaptein, R., Sykes, B. D., Wright, P. E., and Wüthrich, K., *Recommendations for the presentation of NMR structures of proteins and nucleic acids*, J. Mol. Biol. **280** (1998), no. 5, 933–952, DOI 10.1006/jmbi.1998.1852.

[13] Lonkar, P. and Dedon, P. C., *Reactive species and DNA damage in chronic inflammation: Reconciling chemical mechanisms and biological fates*, Int. J. Cancer **128** (2011), no. 9, 1999–2009, DOI 10.1002/ijc.25815.

[14] International Union of Pure and Applied Chemistry and Internation Union of Biochemistry, *Abbreviations and Symbols for Nucleic Acids, Polynucleotides and their Constituents*, Pure Appl. Chem. **40** (1974), no. 3, 278–290, DOI 10.1351/pac197440030277.

[15] Internation Union of Biochemistry, Nomenclature Committee, *Nomenclature for incompletely specified bases in nucleic acid sequences. Recommendations 1984*, Eur. J. Biochem. **150** (1985), no. 1, 1–5, DOI 10.1111/j.1432-1033.1985.tb08977.x.

[16] Allen, S., Blackburn, G. M., Egli, M., Fisher, J., Flavell, A. J., Gait, M. J., Haq, I., Laughton, C., Loakes, D., Luisi, B., Pyle, A. M., Stollar, E., Williams, D. M., and Williams, N. H., *DNA and RNA Structure*, In: Nucleic Acids in Chemistry and Biology (Blackburn, G. M., Gait, M. J., Loakes, D., and Williams, D. M., eds.), The Royal Society of Chemistry, Cambridge, 3rd ed., 2006, ISBN 978-0-85404-654-6, ch. 2, pp. 13–76, DOI 10.1039/9781847555380-00013.

[17] Kaushik, M., Kaushik, S., Roy, K., Singh, A., Mahendru, S., Kumar,

M., Chaudhary, S., Ahmed, S., and Kukreti, S., *A bouquet of DNA structures: Emerging diversity*, Biochem. Biophys. Reports **5** (2016), 388–395, DOI 10.1016/j.bbrep.2016.01.013.

[18] Neidle, S., *Nonstandard and Higher-Order DNA Structures: DNA–DNA Recognition*, In: Principles of Nucleic Acid Structure, Academic Press, Amsterdam, 2008, ISBN 978-0-12-369507-9, ch. 4, pp. 81–131.

[19] Humphrey, W., Dalke, A., and Schulten, K., *VMD: Visual molecular dynamics*, J. Mol. Graph. **14** (1996), no. 1, 33–38, DOI 10.1016/0263-7855(96)00018-5.

[20] Baleja, J. D., Germann, M. W., van de Sande, J. H., and Sykes, B. D., *Solution Conformation of Purine-pyrimidine DNA Octamers using Nuclear Magnetic Resonance, Restrained Molecular Dynamics and NOE-based Refinement*, J. Mol. Biol. **215** (1990), no. 3, 411–428, DOI 10.1016/S0022-2836(05)80361-4.

[21] Blommers, M. J. J., van de Ven, F. J. M., van der Marel, G. A., van Boom, J. H., and Hilbers, C. W., *The three-dimensional structure of a DNA hairpin in solution. Two-dimensional NMR studies and structural analysis of d(ATCCTATTTATAGGAT)*, Eur. J. Biochem. **201** (1991), 33–51, DOI 10.1111/j.1432-1033.1991.tb16253.x.

[22] Gupta, G., Sarma, M. H., Sarma, R. H., Bald, R., Engelke, U., Oei, S. L., Gessner, R., and Erdmann, V. A., *DNA Hairpin Structures in Solution: 500-MHz Two-Dimensional ^{1}H NMR Studies on d(CGCCGCAGC) and d(CGCCGTAGC)*, Biochemistry **26** (1987), no. 24, 7715–7723, DOI 10.1021/bi00398a027.

[23] Led, J. J. and Gesmar, H., *^{31}P Magnetization-Transfer NMR Studies of the Interchange between Structures of Self-Assembled Disodium Guanosine 5′-Monophosphate in Solution*, J. Phys. Chem. **89** (1985), no. 4, 583–588, DOI 10.1021/j100250a007.

[24] Spindler, L., Drevenšek Olenik, I., Čopič, M., Romih, R., Cerar, J., Škerjanc, J., and Mariani, P., *Dynamic light scattering and ^{31}P NMR spectroscopy study of the self-assembly of deoxyguanosine 5′-monophosphate*, Eur. Phys. J. E **7** (2002), no. 1, 95–102, DOI 10.1140/epje/i200101099.

[25] Dolinnaya, N. G., Ogloblina, A. M., and Yakubovskaya, M. G.,

Structure, properties, and biological relevance of the DNA and RNA G-quadruplexes: Overview 50 years after their discovery, Biochemistry (Moscow) **81** (2016), no. 13, 1602–1649, DOI 10.1134/S0006297916130034.

[26] Davis, J. T., *G-Quartets 40 Years Later: From 5′-GMP to Molecular Biology and Supramolecular Chemistry*, Angew. Chem. Int. Ed. **43** (2004), 668–698, DOI 10.1002/anie.200300589.

[27] Spindler, L., Drevenšek-Olenik, I., Čopič, M., Cerar, J., Škerjanc, J., and Mariani, P., *Dynamic light scattering and ^{31}P NMR study of the self-assembly of deoxyguanosine 5′-monophosphate: The effect of added salt*, Eur. Phys. J. E **13** (2004), no. 1, 27–33, DOI 10.1140/epje/e2004-00037-0.

[28] Williamson, J. R., *G-Quartet Structures in Telomeric DNA*, Annu. Rev. Biophys. Biomol. Struct. **23** (1994), 703–730, DOI 10.1146/annurev.bb.23.060194.003415.

[29] Feigon, J., Koshlap, K. M., and Smith, F. W., *^{1}H NMR Spectroscopy of DNA Triplexes and Quadruplexes*, Methods Enzymol. **261** (1995), 225–255, DOI 10.1016/S0076-6879(95)61012-X.

[30] Patel, D. J., Kozlowski, S. A., Marky, L. A., Broka, C., Rice, J. A., Itakura, K., and Breslauer, K. J., *Premelting and Melting Transitions in the d(CGCGAATTCGCG) Self-Complementary Duplex in Solution*, Biochemistry **21** (1982), no. 3, 428–436, DOI 10.1021/bi00532a002.

[31] Wu, P. and Sugimoto, N., *Transition characteristics and thermodynamic analysis of DNA duplex formation: a quantitative consideration for the extent of duplex association*, Nucleic Acids Res. **28** (2000), no. 23, 4762–4768, DOI 10.1093/nar/28.23.4762.

[32] Marky, L. A. and Breslauer, K. J., *Calculating thermodynamic data for transitions of any molecularity from equilibrium melting curves*, Biopolymers **26** (1987), no. 9, 1601–1620, DOI 10.1002/bip.360260911.

[33] Mergny, J.-L. and Lacroix, L., *Analysis of Thermal Melting Curves*, Oligonucleotides **13** (2003), no. 6, 515–537, DOI 10.1089/154545703322860825.

[34] Sahu, I. D., McCarrick, R. M., and Lorigan, G. A., *Use of Electron Paramagnetic Resonance To Solve Biochemical Problems*, Biochemistry **52** (2013), no. 35, 5967–5984, DOI 10.1021/bi400834a.

[35] Borer, P. N., Kan, L. S., and Ts'o, P. O. P., *Conformation and interaction of short nucleic acid double-stranded helices. I. Proton magnetic resonance studies on the nonexchangeable protons of ribosyl ApApGpCpUpU*, Biochemistry **14** (1975), no. 22, 4847–4863, DOI 10.1021/bi00693a012.

[36] Freier, S. M., Kierzek, R., Jaeger, J. A., Sugimoto, N., Caruthers, M. H., Neilson, T., and Turner, D. H., *Improved free-energy parameters for predictions of RNA duplex stability*, Proc. Natl. Acad. Sci. USA **83** (1986), no. 24, 9373–9377, DOI 10.1073/pnas.83.24.9373.

[37] Breslauer, K. J., Frank, R., Blöcker, H., and Marky, L. A., *Predicting DNA duplex stability from the base sequence*, Proc. Natl. Acad. Sci. USA **83** (1986), no. 11, 3746–50.

[38] SantaLucia, Jr., J., Allawi, H. T., and Seneviratne, P. A., *Improved nearest-neighbor parameters for predicting DNA duplex stability*, Biochemistry **35** (1996), no. 11, 3555–62, DOI 10.1021/bi951907q.

[39] SantaLucia, Jr., J., *A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics*, Proc. Natl. Acad. Sci. USA **95** (1998), no. 4, 1460–1465.

[40] Xia, T., SantaLucia, Jr., J., Burkard, M. E., Kierzek, R., Schroeder, S. J., Jiao, X., Cox, C., and Turner, D. H., *Thermodynamic Parameters for an Expanded Nearest-Neighbor Model for Formation of RNA Duplexes with Watson–Crick Base Pairs*, Biochemistry **37** (1998), no. 42, 14719–14735, DOI 10.1021/bi9809425.

[41] Allawi, H. T. and SantaLucia, Jr., J., *Thermodynamics and NMR of internal G·T mismatches in DNA*, Biochemistry **36** (1997), no. 34, 10581–10594, DOI 10.1021/bi962590c.

[42] Allawi, H. T. and SantaLucia, Jr., J., *Thermodynamics of internal C·T mismatches in DNA*, Nucleic Acids Res. **26** (1998), no. 11, 2694–2701, DOI 10.1093/nar/26.11.2694.

[43] Allawi, H. T. and SantaLucia, Jr., J., *Nearest-neighbor thermodynamics of internal A·C mismatches in DNA: Sequence dependence and pH effects*, Biochemistry **37** (1998), no. 26, 9435–9444, DOI 10.1021/bi9803729.

[44] Allawi, H. T. and SantaLucia, Jr., J., *Nearest Neighbor Thermodynamic Pparameters for Internal G·A Mismatches in DNA*, Biochemistry **37** (1998), no. 8, 2170–2179, DOI 10.1021/bi9724873.

[45] Peyret, N., Seneviratne, P. A., Allawi, H. T., and SantaLucia, Jr., J.,

Nearest-Neighbor Thermodynamics and NMR of DNA Sequences with Internal A·A, C·C, G·G, and T·T Mismatches, Biochemistry **38** (1999), no. 12, 3468–3477, DOI 10.1021/bi9825091.

[46] Tikhomirova, A., Beletskaya, I. V., and Chalikian, T. V., *Stability of DNA Duplexes Containing GG, CC, AA, and TT Mismatches.*, Biochemistry **45** (2006), no. 35, 10563–10571, DOI 10.1021/bi060304j.

[47] Bommarito, S., Peyret, N., and SantaLucia, Jr., J., *Thermodynamic parameters for DNA sequences with dangling ends*, Nucleic Acids Res. **28** (2000), no. 9, 1929–1934, DOI 10.1093/nar/28.9.1929.

[48] Watkins, N. E. and SantaLucia, Jr., J., *Nearest-neighbor thermodynamics of deoxyinosine pairs in DNA duplexes*, Nucleic Acids Res. **33** (2005), no. 19, 6258–6267, DOI 10.1093/nar/gki918.

[49] SantaLucia, Jr., J. and Hicks, D., *The Thermodynamics of DNA Structural Motifs*, Annu. Rev. Biophys. Biomol. Struct. **33** (2004), 415–440, DOI 10.1146/annurev.biophys.32.110601.141800.

[50] Sugimoto, N., Nakano, S., Katoh, M., Matsumura, A., Nakamuta, H., Ohmichi, T., Yoneyama, M., and Sasaki, M., *Thermodynamic Parameters To Predict Stability of RNA/DNA Hybrid Duplexes*, Biochemistry **34** (1995), no. 35, 11211–11216, DOI 10.1021/bi00035a029.

[51] Dimitrov, R. A. and Zuker, M., *Prediction of Hybridization and Melting for Double-Stranded Nucleic Acids.*, Biophys. J. **87** (2004), no. 1, 215–226, DOI 10.1529/biophysj.103.020743.

[52] Markham, N. R. and Zuker, M., *DINAMelt web server for nucleic acid melting prediction*, Nucleic Acids Res. **33** (2005), no. Web Server issue, W577–W581, DOI 10.1093/nar/gki591.

[53] Zuker, M., *Mfold web server for nucleic acid folding and hybridization prediction*, Nucleic Acids Res. **31** (2003), no. 13, 3406–3415, DOI 10.1093/nar/gkg595.

[54] Panjkovich, A., Norambuena, T., and Melo, F., *dnaMATE: a consensus melting temperature prediction server for short DNA sequences*, Nucleic Acids Res. **33** (2005), no. Web Server issue, W570–W572, DOI 10.1093/nar/gki379.

[55] Kibbe, W. A., *OligoCalc: an online oligonucleotide properties calculator*, Nucleic Acids Res. **35** (2007), no. Web Server issue, W43–W46, DOI 10.1093/nar/gkm234.

[56] Bird, A. P., *CpG islands as gene markers in the vertebrate nucleus*, Trends Genet. **3** (1987), no. 12, 342–347, DOI 10.1016/0168-9525(87)90294-0.

[57] Adams, R. L., Davis, T., Rinaldi, A., and Eason, R., *CpG deficiency, dinucleotide distributions and nucleosome positioning*, Eur. J. Biochem. **165** (1987), no. 1, 107–115, DOI 10.1111/j.1432-1033.1987.tb11200.x.

[58] Dans, P. D., Faustino, I., Battistini, F., Zakrzewska, K., Lavery, R., and Orozco, M., *Unraveling the sequence-dependent polymorphic behavior of d(CpG) steps in B-DNA*, Nucleic Acids Res. **42** (2014), no. 18, 11304–11320, DOI 10.1093/nar/gku809.

[59] Adams, R. L., *DNA methylation. The effect of minor bases on DNA–protein interactions*, Biochem. J. **265** (1990), no. 2, 309–320, DOI 10.1042/bj2650309.

[60] Ehrlich, M., Gama-Sosa, M. A., Huang, L. H., Midgett, R. M., Kuo, K. C., McCune, R. A., and Gehrke, C., *Amount and distribution of 5-methylcytosine in human DNA from different types of tissues or cells*, Nucleic Acids Res. **10** (1982), no. 8, 2709–2721, DOI 10.1093/nar/10.8.2709.

[61] Gama-Sosa, M. A., Midgett, R. M., Slagel, V. A., Githens, S., Kuo, K. C., Gehrke, C. W., and Ehrlich, M., *Tissue-specific differences in DNA methylation in various mammals*, Biochim. Biophys. Acta **740** (1983), no. 2, 212–219, DOI 10.1016/0167-4781(83)90079-9.

[62] Henderson, I. R. and Jacobsen, S. E., *Epigenetic inheritance in plants*, Nature **447** (2007), no. 7143, 418–424, DOI 10.1038/nature05917.

[63] Reik, W., *Stability and flexibility of epigenetic gene regulation in mammalian development*, Nature **447** (2007), no. 7143, 425–432, DOI 10.1038/nature05918.

[64] Auclair, C., *Structural and Functional Regulation of DNA: Geometry, Topology and Methylation*, In: Nanoscience (Boisseau, P., Houdy, P., and Lahmani, M., eds.), Springer Berlin Heidelberg, Berlin, Heidelberg, 2009, ISBN 978-3-540-88632-7, ch. 1, pp. 3–27, DOI 10.1007/978-3-540-88633-4.

[65] Berger, S. L., *The complex language of chromatin regulation during transcription*, Nature **447** (2007), no. 7143, 407–412, DOI 10.1038/nature05915.

[66] Pasini, A., Bonafè, F., Fiumana, E., Guarnieri, C., Morselli, P. G., Or-

anges, C. M., Caldarera, C. M., Muscari, C., and Giordano, E., *Impact of CpG methylation in addressing adipose-derived stem cell differentiation towards the cardiac phenotype*, In: Epigenomics: From Chromatin Biology to Therapeutics (Appasani, K., ed.), Cambridge University Press, Cambridge, 2012, ISBN 978-1-107-00382-8, ch. 10, pp. 134–145.

[67] Ehrlich, M. and Lacey, M., *DNA methylation and differentiation: silencing, upregulation and modulation of gene expression*, Epigenomics **5** (2013), no. 5, 553–568, DOI 10.2217/epi.13.43.

[68] Csankovszki, G., Nagy, A., and Jaenisch, R., *Synergism of Xist RNA, DNA Methylation, and Histone Hypoacetylation in Maintaining X Chromosome Inactivation*, J. Cell Biol. **153** (2001), no. 4, 773–783, DOI 10.1083/jcb.153.4.773.

[69] Weaver, J. R., Sarkisian, G., Krapp, C., Mager, J., Mann, M. R. W., and Bartolomei, M. S., *Domain-Specific Response of Imprinted Genes to Reduced DNMT1*, Mol. Cell. Biol. **30** (2010), no. 16, 3916–3928, DOI 10.1128/MCB.01278-09.

[70] Baylin, S. B. and Herman, J. G., *DNA hypermethylation in tumorigenesis: Epigenetics joins genetics*, Trends Genet. **16** (2000), no. 4, 168–174, DOI 10.1016/S0168-9525(99)01971-X.

[71] Esteller, M., Silva, J. M., Dominguez, G., Bonilla, F., Matias-Guiu, X., Lerma, E., Bussaglia, E., Prat, J., Harkes, I. C., Repasky, E. A., Gabrielson, E., Schutte, M., Baylin, S. B., and Herman, J. G., *Promoter Hypermethylation and BRCA1 Inactivation in Sporadic Breast and Ovarian Tumors*, J. Natl. Cancer Inst. **92** (2000), no. 7, 564–569, DOI 10.1093/jnci/92.7.564.

[72] Kim, J., Bretz, C. L., and Lee, S., *Epigenetic instability of imprinted genes in human cancers*, Nucleic Acids Res. **43** (2015), no. 22, 10689–10699, DOI 10.1093/nar/gkv867.

[73] Robertson, K. D. and Wolffe, A. P., *DNA methylation in health and disease*, Nat. Rev. Genet. **1** (2000), no. 1, 11–19, DOI 10.1038/35049533.

[74] Herzing, L. B. K., *Epigenetic regulation in human neurodevelopmental disorders including autism, Rett syndrome, and epilepsy*, In: Epigenomics: From Chromatin Biology to Therapeutics (Appasani, K., ed.), Cambridge University Press, Cambridge, 2012, ISBN 978-1-

107-00382-8, ch. 28, pp. 404–419.

[75] Feinberg, A. P., *Phenotypic plasticity and the epigenetics of human disease*, Nature **447** (2007), no. 7143, 433–440, DOI 10.1038/nature05919.

[76] Cedar, H. and Bergman, Y., *Programming of DNA Methylation Patterns*, Annu. Rev. Biochem. **81** (2012), no. 1, 97–117, DOI 10.1146/annurev-biochem-052610-091920.

[77] Arand, J., Lepikhov, K., Wossidlo, M., and Walter, J., *Active DNA demethylation: the enigma starts in the zygote*, In: Epigenomics: From Chromatin Biology to Therapeutics (Appasani, K., ed.), Cambridge University Press, Cambridge, 2012, ISBN 978-1-107-00382-8, ch. 7, pp. 91–103.

[78] Law, J. A. and Jacobsen, S. E., *Establishing, maintaining and modifying DNA methylation patterns in plants and animals*, Nat. Rev. Genet. **11** (2010), no. 3, 204–220, DOI 10.1038/nrg2719.

[79] Hajkova, P., Erhardt, S., Lane, N., Haaf, T., El-Maarri, O., Reik, W., Walter, J., and Surani, M., *Epigenetic reprogramming in mouse primordial germ cells*, Mech. Dev. **117** (2002), no. 1-2, 15–23, DOI 10.1016/S0925-4773(02)00181-8.

[80] Burton, T. and Metcalfe, N. B., *Can environmental conditions experienced in early life influence future generations?*, Proc. Royal Soc. B **281** (2014), no. 1785, 20140311, DOI 10.1098/rspb.2014.0311.

[81] Van Soom, A., Peelman, L., Holt, W. V., and Fazeli, A., *An Introduction to Epigenetics as the Link Between Genotype and Environment: A Personal View*, Reproduction Domest. Animals **49** (2014), no. Suppl. 3, 2–10, DOI 10.1111/rda.12341.

[82] Ng, S.-F., Lin, R. C. Y., Laybutt, D. R., Barres, R., Owens, J. A., and Morris, M. J., *Chronic high-fat diet in fathers programs β-cell dysfunction in female rat offspring*, Nature **467** (2010), no. 7318, 963–966, DOI 10.1038/nature09491.

[83] Dias, B. G. and Ressler, K. J., *Parental olfactory experience influences behavior and neural structure in subsequent generations*, Nat. Neurosci. **17** (2014), no. 1, 89–96, DOI 10.1038/nn.3594.

[84] Nagano, T., *Long non-coding RNA in epigenetic gene silencing*, In: Epigenomics: From Chromatin Biology to Therapeutics (Appasani,

K., ed.), Cambridge University Press, Cambridge, 2012, ISBN 978-1-107-00382-8, ch. 6, pp. 73–87.

[85] Mondal, T. and Kanduri, C., *Maintenance of epigenetic information: a noncoding RNA perspective*, Chromosom. Res. **21** (2013), no. 6–7, 615–625, DOI 10.1007/s10577-013-9385-5.

[86] Matzke, M. A. and Mosher, R. A., *RNA-directed DNA methylation: an epigenetic pathway of increasing complexity*, Nat. Rev. Genet. **15** (2014), no. 6, 394–408, DOI 10.1038/nrg3683.

[87] Weinberg, M. S. and Morris, K. V., *Transcriptional gene silencing in humans*, Nucleic Acids Res. **46** (2016), no. 5, 511–520, DOI 10.1093/nar/gkw139.

[88] Tahiliani, M., Koh, K. P., Shen, Y., Pastor, W. a., Bandukwala, H., Brudno, Y., Agarwal, S., Iyer, L. M., Liu, D. R., Aravind, L., and Rao, A., *Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1*, Science **324** (2009), no. 5929, 930–935, DOI 10.1126/science.1170116.

[89] Hackett, J. A., Sengupta, R., Zylicz, J. J., Murakami, K., Lee, C., Down, T. A., and Surani, M. A., *Germline DNA demethylation dynamics and imprint erasure through 5-hydroxymethylcytosine*, Science **339** (2013), no. 6118, 448–452, DOI 10.1126/science.1229277.

[90] Wassenegger, M., Heimes, S., Riedel, L., and Sänger, H. L., *RNA-directed de novo methylation of genomic sequences in plants*, Cell **76** (1994), no. 3, 567–576, DOI 10.1016/0092-8674(94)90119-8.

[91] Dalpke, A. H. and Heeg, K., *CpG-DNA as immune response modifier*, Int. J. Med. Microbiol. **294** (2004), no. 5, 345–354, DOI 10.1016/j.ijmm.2004.07.005.

[92] Hemmi, H., Takeuchi, O., Kawai, T., Kaisho, T., Sato, S., Sanjo, H., Matsumoto, M., Hoshino, K., Wagner, H., Takeda, K., and Akira, S., *A Toll-like receptor recognizes bacterial DNA*, Nature **408** (2000), no. 6813, 740–745, DOI 10.1038/35047123.

[93] Krieg, A. M., *CpG Motifs in Bacterial DNA and Their Immune Effects*, Annu. Rev. Immunol. **20** (2002), 709–760, DOI 10.1146/annurev.immunol.20.100301.064842.

[94] Klinman, D. M., *Immunotherapeutic uses of CpG oligodeoxynucleotides*, Nat. Rev. Immunol. **4** (2004), no. 4, 249–259, DOI 10.1038/nri1329.

[95] Gangloff, M., Ludidi, P. L., and Gay, N. J., *Structures and motifs involved in Toll signaling*, In: Toll and Toll-Like Receptors: An Immunologic Perspective (Rich, T., ed.), Molecular Biology Intelligence Unit, Springer, Boston, 2005, ISBN 978-0-387-27445-4, ch. 2, pp. 56–93, DOI 10.1007/0-387-27445-6_3.

[96] Morecki, S. and Slavin, S., *Immunoregulation of GVHD by triggering the innate immune system with CpG*, Expert. Rev. Hematol. **2** (2009), no. 4, 443–453, DOI 10.1586/ehm.09.29.

[97] Dickerson, R. E. and Drew, H. R., *Structure of a B-DNA dodecamer. II. Influence of base sequence on helix structure*, J. Mol. Biol. **149** (1981), no. 4, 761–786, DOI 10.1016/0022-2836(81)90357-0.

[98] Lefebvre, A., Mauffret, O., Hartmann, B., Lescot, E., and Fermandjian, S., *Structural behavior of the CpG step in two related oligonucleotides reflects its malleability in solution*, Biochemistry **34** (1995), no. 37, 12019–12028.

[99] Cordier, C., Marcourt, L., Petitjean, M., and Dodin, G., *Conformational variation of the central CG site in d(ATGACGTCAT)$_2$ and d(GAAAACGTTTTC)$_2$. An NMR, molecular modelling and 3D-homology investigation*, Eur. J. Biochem. **261** (1999), no. 3, 722–733, DOI 10.1046/j.1432-1327.1999.00314.x.

[100] Dans, P. D., Pérez, A., Faustino, I., Lavery, R., and Orozco, M., *Exploring polymorphisms in B-DNA helical conformations*, Nucleic Acids Res. **40** (2012), no. 21, 10668–10678, DOI 10.1093/nar/gks884.

[101] Lefebvre, A., Mauffret, O., Lescot, E., Hartmann, B., and Fermandjian, S., *Solution structure of the CpG containing d(CTTCGAAG)$_2$ oligonucleotide: NMR data and energy calculations are compatible with a BI/BII equilibrium at CpG*, Biochemistry **35** (1996), no. 38, 12560–12569, DOI 10.1021/bi9606298.

[102] Pasi, M., Maddocks, J. H., Beveridge, D., Bishop, T. C., Case, D. A., Cheatham, T., Dans, P. D., Jayaram, B., Lankas, F., Laughton, C., Mitchell, J., Osman, R., Orozco, M., Pérez, A., Petkevičiūtė, D., Spackova, N., Sponer, J., Zakrzewska, K., and Lavery, R., *μABC: a systematic microsecond molecular dynamics study of tetranucleotide sequence effects in B-DNA*, Nucleic Acids Res. **42** (2014), no. 19, 12272–12283, DOI 10.1093/nar/gku855.

[103] Renčiuk, D., Blacque, O., Vorlíčková, M., and Spingler, B., *Crystal structures of B-DNA dodecamer containing the epigenetic modifications 5-hydroxymethylcytosine or 5-methylcytosine*, Nucleic Acids Res. **41** (2013), no. 21, 9891–9900, DOI 10.1093/nar/gkt738.

[104] Heinemann, U. and Hahn, M., *C-C-A-G-G-C-m^{5}C-T-G-G. Helical fine structure, hydration, and comparison with C-C-A-G-G-C-C-T-G-G*, J. Biol. Chem. **267** (1992), no. 11, 7332–7341.

[105] Theruvathu, J. A., Yin, Y. W., Pettitt, B. M., and Sowers, L. C., *Comparison of the Structural and Dynamic Effects of 5-Methylcytosine and 5-Chlorocytosine in a CpG Dinucleotide Sequence*, Biochemistry **52** (2013), no. 47, 8590–8598, DOI 10.1021/bi400980c.

[106] Lefebvre, A., Mauffret, O., El Antri, S., Monnot, M., Lescot, E., and Fermandjian, S., *Sequence dependent effects of CpG cytosine methylation. A joint ^{1}H-NMR and ^{31}P-NMR study*, Eur. J. Biochem. **229** (1995), no. 2, 445–454, DOI 10.1111/j.1432-1033.1995.0445k.x.

[107] Derreumaux, S., Chaoui, M., Tevanian, G., and Fermandjian, S., *Impact of CpG methylation on structure, dynamics and solvation of cAMP DNA responsive element*, Nucleic Acids Res. **29** (2001), no. 11, 2314–2326, DOI 10.1093/nar/29.11.2314.

[108] Wanunu, M., Cohen-Karni, D., Johnson, R. R., Fields, L., Benner, J., Peterman, N., Zheng, Y., Klein, M. L., and Drndic, M., *Discrimination of Methylcytosine from Hydroxymethylcytosine in DNA Molecules*, J. Am. Chem. Soc. **133** (2011), no. 3, 486–492, DOI 10.1021/ja107836t.

[109] Marcourt, L., Cordier, C., Couesnon, T., and Dodin, G., *Impact of C5-cytosine methylation on the solution structure of d(GAAAACGTTTTC)$_2$. An NMR and molecular modelling investigation*, Eur. J. Biochem. **265** (1999), no. 3, 1032–1042, DOI 10.1046/j.1432-1327.1999.00819.x.

[110] Hodges-Garcia, Y. and Hagerman, P. J., *Cytosine methylation can induce local distortions in the structure of duplex DNA*, Biochemistry **31** (1992), no. 33, 7595–7599, DOI 10.1021/bi00148a022.

[111] Thalhammer, A., Hansen, A. S., El-Sagheer, A. H., Brown, T., and Schofield, C. J., *Hydroxylation of methylated CpG dinucleotides reverses stabilisation of DNA duplexes by cytosine 5-methylation*, Chem. Commun. **47** (2011), no. 18, 5325–5327, DOI 10.1039/c0cc05671e.

[112] Rodríguez López, C. M., Guzmán Asenjo, B., Lloyd, A. J., and

Wilkinson, M. J., *Direct detection and quantification of methylation in nucleic acid sequences using high-resolution melting analysis*, Anal. Chem. **82** (2010), no. 21, 9100–9108, DOI 10.1021/ac1024057.

[113] Rodríguez López, C. M., Lloyd, A. J., Leonard, K., and Wilkinson, M. J., *Differential Effect of Three Base Modifications on DNA Thermostability Revealed by High Resolution Melting*, Anal. Chem. **84** (2012), no. 17, 7336–7342, DOI 10.1021/ac301459x.

[114] Xodo, L. E., Manzini, G., Quadrifoglio, F., van der Marel, G., and van Boom, J., *DNA hairpin loops in solution. Correlation between primary structure, thermostability and reactivity with single-strand-specific nuclease from mung bean*, Nucleic Acids Res. **19** (1991), no. 7, 1505–1511, DOI 10.1093/nar/19.7.1505.

[115] Xodo, L. E., Manzini, G., Quadrifoglio, F., van der Marel, G. A., and van Boom, J. H., *Effect of 5-methylcytosine on the stability of triple-stranded DNA – a thermodynamic study*, Nucleic Acids Res. **19** (1991), no. 20, 5625–5631, DOI 10.1093/nar/19.20.5625.

[116] Zamiri, B., Mirceta, M., Bomsztyk, K., Macgregor, R. B. J., and Pearson, C. E., *Quadruplex formation by both G-rich and C-rich DNA strands of the* C9orf72 *(GGGGCC)8·(GGCCCC)8 repeat: effect of CpG methylation*, Nucleic Acids Res. **43** (2015), no. 20, 10055–10064, DOI 10.1093/nar/gkv1008.

[117] El Antri, S., Bittoun, P., Mauffret, O., Monnot, M., Convert, O., Lescot, E., and Fermandjian, S., *Effect of distortions in the phosphate backbone conformation of six related octanucleotide duplexes on CD and* ^{31}P *NMR spectra*, Biochemistry **32** (1993), no. 28, 7079–7088, DOI 10.1021/bi00079a003.

[118] Delort, A.-M., Neumann, J. M., Molko, D., Hervé, M., Téoule, R., and Tran Dinh, S., *Influence of uracil defect on DNA structure:* ^{1}H *NMR investigation at 500 MHz*, Nucleic Acids Res. **13** (1985), no. 9, 3343–3355, DOI 10.1093/nar/13.9.3343.

[119] Wärmländer, S., Sponer, J. E., Sponer, J., and Leijon, M., *The influence of the thymine C5 methyl group on spontaneous base pair breathing in DNA*, J. Biol. Chem. **277** (2002), no. 32, 28491–28497, DOI 10.1074/jbc.M202989200.

[120] Bianchi, C. and Zangi, R., *How to Distinguish Methyl-Cytosine from*

Cytosine with High Fidelity, J. Mol. Biol. **424** (2012), no. 3-4, 215–224, DOI 10.1016/j.jmb.2012.09.024.

[121] Acosta-Silva, C., Branchadell, V., Bertran, J., and Oliva, A., *Mutual Relationship between Stacking and Hydrogen Bonding in DNA. Theoretical Study of Guanine–Cytosine, Guanine–5-methylcytosine, and Their Dimers*, J. Phys. Chem. B **114** (2010), no. 31, 10217–10227, DOI 10.1021/jp103850h.

[122] Geahigan, K. B., Meints, G. a., Hatcher, M. E., Orban, J., and Drobny, G. P., *The dynamic impact of CpG methylation in DNA*, Biochemistry **39** (2000), no. 16, 4939–4946, DOI 10.1021/bi9917636.

[123] Yusufaly, T. I., Li, Y., and Olson, W. K., *5-Methylation of Cytosine in CG:CG Base-Pair Steps: A Physicochemical Mechanism for the Epigenetic Control of DNA Nanomechanics*, J. Phys. Chem. B **117** (2013), no. 51, 16436–16442, DOI 10.1021/jp409887t.

[124] Vogt, A. D. and Di Cera, E., *Conformational Selection Is a Dominant Mechanism of Ligand Binding*, Biochemistry **52** (2013), no. 34, 5723–5729, DOI 10.1021/bi400929b.

[125] Gilman, M. Z., *The c-*fos *serum response element responds to protein kinase C–dependent and –independent signals but not to cyclic AMP*, Genes Dev. **2** (1988), no. 4, 394–402, DOI 10.1101/gad.2.4.394.

[126] Pellegrini, L., Tan, S., and Richmond, T. J., *Structure of serum response factor core bound to DNA*, Nature **376** (1995), no. 6540, 490–498, DOI 10.1038/376490a0.

[127] Huet, A., Parlakian, A., Arnaud, M.-C., Glandières, J.-M., Valat, P., Fermandjian, S., Paulin, D., Alpert, B., and Zentz, C., *Mechanism of binding of serum response factor to serum response element*, FEBS J. **272** (2005), no. 12, 3105–3119, DOI 10.1111/j.1742-4658.2005.04724.x.

[128] Hassler, M. and Richmond, T. J., *The B-box dominates SAP-1-SRF interactions in the structure of the ternary complex*, EMBO J. **20** (2001), no. 12, 3018–3028, DOI 10.1093/emboj/20.12.3018.

[129] Mo, Y., Ho, W., Johnston, K., and Marmorstein, R., *Crystal structure of a ternary SAP-1/SRF/c-fos SRE DNA complex.*, J. Mol. Biol. **314** (2001), no. 3, 495–506, DOI 10.1006/jmbi.2001.5138.

[130] Bikard, D., Loot, C., Baharoglu, Z., and Mazel, D., *Folded DNA in Action: Hairpin Formation and Biological Functions in Prokary-*

otes, Microbiol. Mol. Biol. Rev. **74** (2010), no. 4, 570–588, DOI 10.1128/MMBR.00026-10.

[131] Brázda, V., Laister, R. C., Jagelská, E. B., and Arrowsmith, C., *Cruciform structures are a common DNA feature important for regulating biological processes*, BMC Mol. Biol. **12** (2011), DOI 10.1186/1471-2199-12-33.

[132] Řezáčová, B., *Strukturní flexibilita regulačního segmentu DNA*, Master thesis, Univerzita Karlova v Praze, 2009.

[133] Bonnet, G., Krichevsky, O., and Libchaber, A., *Kinetics of conformational fluctuations in DNA hairpin-loops*, Proc. Natl. Acad. Sci. USA **95** (1998), no. 15, 8602–8606, DOI 10.1088/1367-2630/15/11/113010.

[134] Lin, C. H. and Patel, D. J., *Structural basis of DNA folding and recognition in an AMP-DNA aptamer complex: distinct architectures but common recognition motifs for DNA and RNA aptamers complexed to AMP*, Chem. Biol. **4** (1997), no. 11, 817–832, DOI 10.1016/S1074-5521(97)90115-0.

[135] Atkins, P. and de Paula, J., *Chemical equilibrium*, In: Atkins' Physical Chemistry, Oxford University Press, Oxford, 8th ed., 2006, ISBN 9780198700722, ch. 7, pp. 200–239.

[136] Owczarzy, R., *Melting temperatures of nucleic acids: discrepancies in analysis*, Biophys. Chem. **117** (2005), no. 3, 207–215, DOI 10.1016/j.bpc.2005.05.006.

[137] Chaires, J. B., *Possible origin of differences between van't Hoff and calorimetric enthalpy estimates*, Biophys. Chem. **64** (1997), no. 1-3, 15–23, DOI 10.1016/S0301-4622(96)02205-3.

[138] Wu, P., Nakano, S., and Sugimoto, N., *Temperature dependence of thermodynamic properties for DNA/DNA and RNA/DNA duplex formation*, Eur. J. Biochem. **269** (2002), no. 12, 2821–2830, DOI 10.1046/j.1432-1033.2002.02970.x.

[139] McPhail, D. and Cooper, A., *Thermodynamics and kinetics of dissociation of ligand-induced dimers of vancomycin antibiotics*, J. Chem. Soc. Faraday Trans. **93** (1997), no. 13, 2283–2289, DOI 10.1039/a701327b.

[140] Abragam, A., *General introduction*, In: Principles of Nuclear Magnetism, International Series of Monographs on Physics, vol. 32, Oxford University Press, Oxford, 1983, ISBN 978-0-19-852014-6, ch. 1,

pp. 1–18.

[141] Hore, P. J., *Introduction*, In: Nuclear Magnetic Resonance, Oxford University Press, Oxford, 1995, ISBN 978-0-198-55682-4, ch. 1, pp. 1–7.

[142] Keeler, J., *Product operators*, In: Understanding NMR spectroscopy, John Wiley & Sons, Ltd, Chichester, 2nd ed., 2010, ISBN 978-0-470-74608-0, ch. 7, pp. 139–182.

[143] Levitt, M. H., *Homonuclear AX System*, In: Spin dynamics: basics of nuclear magnetic resonance, John Wiley & Sons, Ltd, Chichester, 2nd ed., 2008, ISBN 978-0-470-51117-6, ch. 15, pp. 369–408.

[144] Griesinger, C., Schwalbe, H., Schleucher, J., and Sattler, M., *Proton-Detected Heteronuclear and Multidimensional NMR*, In: Two-Dimensional NMR Spectroscopy (Croasmun, W. R. and Carlson, R. M. K., eds.), Methods in Stereochemical Analysis, VHC Publishers, Inc., New York, 2nd ed., 1994, ISBN 1-56081-664-3, ch. 3, pp. 457–580.

[145] Friebolin, H., *Two-Dimensional NMR Spectroscopy*, In: Basic One- and Two-Dimensional NMR Spectroscopy, WILEY-VHC Verlag GmbH & Co. KGaA, Weinheim, 5th ed., 2011, ISBN 978-3-527-32782-9, ch. 9, pp. 239–296.

[146] Berger, S. and Braun, S., *The Second Dimension*, In: 200 and More NMR Experiments – A Practical Course, WILEY-VHC Verlag GmbH & Co. KGaA, Weinheim, 2004, ISBN 3-527-31067-3, ch. 10, pp. 362–452.

[147] Friebolin, H., *The Nuclear Overhauser Effect*, In: Basic One- and Two-Dimensional NMR Spectroscopy, WILEY-VHC Verlag GmbH & Co. KGaA, Weinheim, 5th ed., 2011, ISBN 978-3-527-32782-9, ch. 10, pp. 297–312.

[148] McConnell, H. M., *Reaction rates by nuclear magnetic resonance*, J. Chem. Phys. **28** (1958), no. 3, 430–431, DOI 10.1063/1.1744152.

[149] Bain, A. D., *Chemical exchange in NMR*, Prog. Nucl. Magn. Reson. Spectrosc. **43** (2003), no. 3-4, 63–103, DOI 10.1016/j.pnmrs.2003.08.001.

[150] Hansen, D. F. and Led, J. J., *Implications of using approximate Bloch–McConnell equations in NMR analyses of chemically exchanging systems:*

application to the electron self-exchange of plastocyanin, J. Magn. Reson. **163** (2003), no. 2, 215–227, DOI 10.1016/S1090-7807(03)00062-4.

[151] Palmer, III, A. G., *Chemical Exchange Effects in Biological Macromolecules*, In: Encyclopedia of Nuclear Magnetic Resonance (Grant, D. M. and Harris, R. K., eds.), John Wiley, 2002.

[152] Günther, H., *The Influence of Dynamic Effects on ^1H Nuclear Magnetic Resonance Spectra*, In: NMR Spectroscopy: Basic Principles, Concepts, and Applications in Chemistry, John Wiley & Sons Ltd, Chichester, 2nd ed., 1995, ISBN 978-0-471-95201-X, ch. 9, pp. 335–390.

[153] Kaplan, J. I. and Fraenkel, G., *NMR Lineshapes for Exchanging Systems under Conditions of Low rf Power*, In: NMR of Chemically Exchanging Systems, Academic Press, New York, 1980, ISBN 0-12-397550-6, ch. VI, pp. 71–129.

[154] Římal, V., *NMR studium lokální strukturní stability v molekule DNA*, Master thesis, Charles University in Prague, 2009.

[155] Wemmer, D. E., *Nucleic Acid Structure and Dynamics from ^1H NMR*, In: NMR Spectroscopy and its Application to Biomedical Research (Sarkar, S. K., ed.), Elsevier Science B. V., Amsterdam, 1996, ISBN 978-0-444-89410-6, ch. 6, pp. 281–312, DOI 10.1016/B978-044489410-6/50008-8.

[156] Lam, S. L. and Chi, L. M., *Use of chemical shifts for structural studies of nucleic acids*, Prog. Nucl. Magn. Reson. Spectrosc. **56** (2010), no. 3, 289–310, DOI 10.1016/j.pnmrs.2010.01.002.

[157] Feigon, J., Sklenář, V., Wang, E., Gilbert, D. E., Macaya, R. F., and Schultze, P., *^1H NMR Spectroscopy of DNA*, Methods Enzymol. **211** (1992), no. 1977, 235–253, DOI 10.1016/0076-6879(92)11015-B.

[158] Do, N. Q. and Phan, A. T., *Monomer–Dimer Equilibrium for the 5′–5′ Stacking of Propeller-Type Parallel-Stranded G-Quadruplexes: NMR Structural Study*, Chem. Eur. J. **18** (2012), no. 46, 14752–14759, DOI 10.1002/chem.201103295.

[159] Phan, A. T. and Patel, D. J., *A Site-Specific Low-Enrichment ^{15}N, ^{13}C Isotope-Labeling Approach to Unambiguous NMR Spectral Assignments in Nucleic Acids*, J. Am. Chem. Soc. **124** (2002), no. 7, 1160–1161, DOI 10.1021/ja011977m.

[160] Wüthrich, K., *NOE-Observable ^{1}H–^{1}H Distances in Nucleic Acids*, In: NMR of proteins and nucleic acids, John Wiley, Chichester, 1986, ISBN 978-0-471-82893-8, ch. 11, pp. 203–219.

[161] Wijmenga, S. S. and van Buuren, B. N. M., *The use of NMR methods for conformational studies of nucleic acids*, Prog. Nucl. Magn. Reson. Spectrosc. **32** (1998), no. 4, 287–387, DOI 10.1016/S0079-6565(97)00023-X.

[162] Goljer, I. and Bolton, P. H., *Studies of Nucleic Acid Structures Based on NMR Results*, In: Two-Dimensional NMR Spectroscopy (Croasmun, W. R. and Carlson, R. M. K., eds.), Methods in Stereochemical Analysis, VHC Publishers, Inc., New York, 2nd ed., 1994, ISBN 1-56081-664-3, ch. 7, pp. 699–740.

[163] Altona, C., Faber, D. H., and Hoekzema, A. J. A. W., *Double-helical DNA ^{1}H chemical shifts: an accurate and balanced predictive empirical scheme*, Magn. Reson. Chem. **38** (2000), no. 2, 95–107, DOI 10.1002/(SICI)1097-458X(200002)38:2<95::AID-MRC592>3.0.CO;2-M.

[164] Hilbers, C. W. and Wijmenga, S. S., *Nucleic Acids: Spectra, Structures, and Dynamics*, In: Encyclopedia of Nuclear Magnetic Resonance (Grant, D. M. and Harris, R. K., eds.), John Wiley, 2002.

[165] Sklenář, V., Miyashiro, H., Zon, G., Miles, H. T., and Bax, A., *Assignment of the ^{31}P and ^{1}H resonances in oligonucleotides by two-dimensional NMR spectroscopy*, FEBS Lett. **208** (1986), no. 1, 94–98, DOI 10.1016/0014-5793(86)81539-3.

[166] Ono, A., Makita, T., Tate, S., Kawashima, E., Ishido, Y., and Kainosho, M., *C5′ Methylene Proton Signal Assignment of DNA/RNA Oligomers Labeled with C5′-Monodeuterated Nucleosides by ^{1}H–^{31}P HSQC Spectroscopy*, Magn. Reson. Chem. **34** (1996), no. 13, S40–S46, DOI 10.1002/(SICI)1097-458X(199612)34:13<S40::AID-OMR43>3.0.CO;2-Y.

[167] Tisne, C., Simenel, C., Hantz, E., Schaeffer, F., and Delepierre, M., *Backbone Conformational Study of a Non-Palindromic 16 Base Pair DNA Duplex Exploring 2D ^{31}P–^{1}H Heteronuclear Inverse Spectroscopy: Assignment of all NMR Phosphorus Resonances and Measurement of ^{3}J ^{31}P–$^{1}H3′$ Coupling Constants*, Magn. Reson. Chem. **34** (1996), no. 13,

S115–S124, DOI 10.1002/(SICI)1097-458X(199612)34:13<S115::AID-OMR66>3.0.CO;2-C.

[168] Kellogg, G. W. and Schweitzer, B. I., *Two- and three-dimensional ^{31}P-driven NMR procedures for complete assignment of backbone resonances in oligodeoxyribonucleotides*, J. Biomol. NMR **3** (1993), no. 5, 577–595, DOI 10.1007/BF00174611.

[169] Castagné, C., Murphy, E. C., Gronenborn, A. M., and Delepierre, M., *^{31}P NMR analysis of the DNA conformation induced by protein binding. SRY/DNA complexes*, Eur. J. Biochem. **267** (2000), no. 4, 1223–1229, DOI 10.1046/j.1432-1327.2000.01124.x.

[170] Gorenstein, D. G., *Conformation and Dynamics of DNA and Protein-DNA Complexes by ^{31}P NMR*, Chem. Rev. **94** (1994), no. 5, 1315–1338, DOI 10.1021/cr00029a007.

[171] van Dongen, M. J. P., Wijmenga, S. S., Eritja, R., Azorín, F., and Hilbers, C. W., *Through-bond correlation of adenine H2 and H8 protons in unlabeled DNA fragments by HMBC spectroscopy*, J. Biomol. NMR **8** (1996), no. 2, 207–212, DOI 10.1007/BF00211166.

[172] Phan, A. T., *Long-range imino proton-^{13}C J-couplings and the through-bond correlation of imino and non-exchangeable protons in unlabeled DNA*, J. Biomol. NMR **16** (2000), no. 2, 175–178, DOI 10.1023/A:1008355231085.

[173] Lam, S. L., Ip, L. N., Cui, X., and Ho, C. N., *Random coil proton chemical shifts of deoxyribonucleic acids*, J. Biomol. NMR **24** (2002), no. 4, 329–337, DOI 10.1023/A:1021671531438.

[174] Kwok, C. W., Ho, C. N., Chi, L. M., and Lam, S. L., *Random coil carbon chemical shifts of deoxyribonucleic acids*, J. Magn. Reson. **166** (2004), no. 1, 11–18, DOI 10.1016/j.jmr.2003.10.001.

[175] Ho, C. N. and Lam, S. L., *Random coil phosphorus chemical shift of deoxyribonucleic acids*, J. Magn. Reson. **171** (2004), no. 2, 193–200, DOI 10.1016/j.jmr.2004.08.024.

[176] Cavalli, A., Salvatella, X., Dobson, C. M., and Vendruscolo, M., *Protein structure determination from NMR chemical shifts*, Proc. Natl. Acad. Sci. USA **104** (2007), no. 23, 9615–9620, DOI 10.1073/pnas.0610313104.

[177] Wijmenga, S. S., Kruithof, M., and Hilbers, C. W., *Analysis of ^{1}H*

chemical shifts in DNA: Assessment of the reliability of ^1H chemical shift calculations for use in structure refinement, J. Biomol. NMR **10** (1997), no. 4, 337–350, DOI 10.1023/A:1018348123074.

[178] Hader, P. A., Alkema, D., Bell, R. A., and Neilson, T., *Parameters for proton chemical shift calculation in oligoribonucleotides*, J. Chem. Soc. Chem. Commun. (1982), no. 1, 10–12, DOI 10.1039/c39820000010.

[179] Bell, R., Alkema, D., Coddington, J., Hader, P., Hughes, D., and Neilson, T., *Prediction of ^1H NMR chemical shifts of DNA oligomers*, Nucleic Acids Res. **11** (1983), no. 4, 1143–1149, DOI 10.1093/nar/11.4.1143.

[180] Abi-Ghanem, J., Heddi, B., Foloppe, N., and Hartmann, B., *DNA structures from phosphate chemical shifts*, Nucleic Acids Res. **38** (2010), no. 3, 1119–1128, DOI 10.1093/nar/gkp1061.

[181] Giessner-Prettre, C. and Pullman, B., *Quantum mechanical calculations of NMR chemical shifts in nucleic acids*, Q. Rev. Biophys. **20** (1987), no. 3-4, 113–172, DOI 10.1017/S0033583500004169.

[182] Pauwels, E., Claeys, D., Martins, J. C., Waroquier, M., Bifulco, G., Speybroeck, V. V., and Madder, A., *Accurate prediction of ^1H chemical shifts in interstrand cross-linked DNA*, RSC Adv. **3** (2013), no. 12, 3925–3938, DOI 10.1039/c3ra22408b.

[183] Victora, A., Moller, H. M., and Exner, T. E., *Accurate* ab initio *prediction of NMR chemical shifts of nucleic acids and nucleic acids/protein complexes*, Nucleic Acids Res. **42** (2014), no. 22, e173, DOI 10.1093/nar/gku1006.

[184] Lam, S. L., *DSHIFT: a web server for predicting DNA chemical shifts*, Nucleic Acids Res. **35** (2007), no. Web Server issue, W713–W717, DOI 10.1093/nar/gkm320.

[185] Barton, S., Heng, X., Johnson, B. A., and Summers, M. F., *Database proton NMR chemical shifts for RNA signal assignment and validation*, J. Biomol. NMR **55** (2013), no. 1, 33–46, DOI 10.1007/s10858-012-9683-9.

[186] Brown, J. D., Summers, M. F., and Johnson, B. A., *Prediction of hydrogen and carbon chemical shifts from RNA using database mining and support vector regression*, J. Biomol. NMR **63** (2015), no. 1, 39–52, DOI 10.1007/s10858-015-9961-4.

[187] Lam, S. L., Lai, K. F., and Chi, L. M., *Proton chemical shift prediction of A·A mismatches in B-DNA duplexes*, J. Magn. Reson. **187** (2007), no. 1, 105–111, DOI 10.1016/j.jmr.2007.04.005.

[188] Kwok, C. K. and Lam, S. L., *NMR proton chemical shift prediction of T·T mismatches in B-DNA duplexes*, J. Magn. Reson. **234** (2013), 184–189, DOI 10.1016/j.jmr.2013.06.022.

[189] Ng, K. S. and Lam, S. L., *NMR proton chemical shift prediction of C·C mismatches in B-DNA*, J. Magn. Reson. **252** (2015), 87–93, DOI 10.1016/j.jmr.2015.01.005.

[190] Hwang, T.-L. and Shaka, A., *Water Suppression That Works. Excitation Sculpting Using Arbitrary Waveforms and Pulsed Field Gradients*, J. Magn. Reson. Ser. A **112** (1995), no. 2, 275–279, DOI 10.1006/jmra.1995.1047.

[191] Parella, T., *Pulse Program Catalogue: I. 1D & 2D NMR Experiments*, Bruker BioSpin GmbH, 2010.

[192] Günther, H., *Experimental aspects of nuclear magnetic resonance spectroscopy*, In: NMR Spectroscopy: Basic Principles, Concepts, and Applications in Chemistry, John Wiley & Sons Ltd, Chichester, 2nd ed., 1995, ISBN 978-0-471-95201-X, ch. 3, pp. 53–67.

[193] Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A., *NMRPipe: A multidimensional spectral processing system based on UNIX pipes*, J. Biomol. NMR **6** (1995), no. 3, 277–293, DOI 10.1007/BF00197809.

[194] Goddard, T. D. and Kneller, D. G., *Sparky 3*.

[195] Piotto, M., Saudek, V., and Sklenář, V., *Gradient-tailored excitation for single quantum NMR spectroscopy of aqueous solutions*, J. Biomol. NMR **2** (1992), no. 6, 661–665, DOI 10.1007/BF02192855.

[196] Sklenář, V., Piotto, M., Leppik, R., and Saudek, V., *Gradient-Tailored Water Suppression for ^{1}H–^{15}N HSQC Experiments Optimized to Retain Full Sensitivity*, J. Magn. Reson. Ser. A **102** (1993), no. 2, 241–245, DOI 10.1006/jmra.1993.1098.

[197] Tataurov, A. V., You, Y., and Owczarzy, R., *Predicting ultraviolet spectrum of single stranded and double stranded deoxyribonucleic acids*, Biophys. Chem. **133** (2008), no. 1-3, 66–70, DOI 10.1016/j.bpc.2007.12.004.

[198] Cavaluzzi, M. J. and Borer, P. N., *Revised UV extinction coefficients for nucleoside-5′-monophosphates and unpaired DNA and RNA*, Nucleic Acids Res. **32** (2004), no. 1, e13, DOI 10.1093/nar/gnh015.

[199] Reich, H. J., *WinDNMR: Dynamic NMR Spectra for Windows*, J. Chem. Educ. **72** (1995), no. 12, 1086, DOI 10.1021/ed072p1086.1.

[200] Budzelaar, P. H., *gNMR*, 2006.

[201] Rohonczy, J., *TEDDY – Dynamic NMR Module*.

[202] van Beek, J. D., *matNMR: A flexible toolbox for processing, analyzing and visualizing magnetic resonance data in Matlab®*, J. Magn. Reson. **187** (2007), no. 1, 19–26, DOI 10.1016/j.jmr.2007.03.017.

[203] Bak, M., Rasmussen, J. T., and Nielsen, N. C., *SIMPSON: A General Simulation Program for Solid-State NMR Spectroscopy*, J. Magn. Reson. **147** (2000), no. 2, 296–330, DOI 10.1006/jmre.2000.2179.

[204] Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P., *Modeling of Data*, In: Numerical Recipes in C: The Art of Scientific Computing, Cambridge University Press, Cambridge, 2nd ed., 1992, ISBN 978-0521431088, ch. 15, pp. 656–706.

[205] Socha, O., *Studium lokálních změn ve stabilitě dvoušroubovice DNA po methylaci cytosinu pomocí jaderné magneticé rezonance*, Bachelor thesis, Charles University in Prague, 2014.

[206] Dougherty, R. C., *Temperature and pressure dependence of hydrogen bond strength: A perturbation molecular orbital approach*, J. Chem. Phys. **109** (1998), no. 17, 7372, DOI 10.1063/1.477343.

[207] Dračínský, M., Bouř, P., and Hodgkinson, P., *Temperature Dependence of NMR Parameters Calculated from Path Integral Molecular Dynamics Simulations*, J. Chem. Theory Comput. **12** (2016), 968–973, DOI 10.1021/acs.jctc.5b01131.

[208] Ackerman, J. J., Soto, G. E., Spees, W. M., Zhu, Z., and Evelhoch, J. L., *The NMR Chemical Shift pH Measurement Revisited: Analysis of Error and Modeling of a pH Dependent Reference*, Magn. Reson. Med. **36** (1996), no. 5, 674–683, DOI 10.1002/mrm.1910360505.

[209] Pietri, S., Miollan, M., Martel, S., Le Moigne, F., Blaive, B., and Culcasi, M., *α- and β-Phosphorylated Amines and Pyrrolidines, a New Class of Low Toxic Highly Sensitive ^{31}P NMR pH Indicators*, J. Biol. Chem. **275** (2000), no. 26, 19505–19512, DOI 10.1074/jbc.M001784200.

[210] Lutz, N. W., Fur, Y. L., Chiche, J., Pouysse, J., and Cozzone, P. J., *Quantitative in vivo characterization of intracellular and extracellular pH profiles in heterogeneous tumors: A novel method enabling multiparametric pH analysis*, Cancer Res. **73** (2013), no. 15, 4616–4628, DOI 10.1158/0008-5472.CAN-13-0767.

[211] Nakata, M., Zanchetta, G., Chapman, B. D., Jones, C. D., Cross, J. O., Pindak, R., Bellini, T., and Clark, N. A., *End-to-End Stacking and Liquid Crystal Condensation of 6- to 20-Base Pair DNA Duplexes*, Science **318** (2007), no. 5854, 1276–1279, DOI 10.1126/science.1143826.

[212] Alam, T. M. and Drobny, G., *Magnetic ordering in synthetic oligonucleotides. A deuterium nuclear magnetic resonance investigation*, J. Chem. Phys. **92** (1990), no. 11, 6840–6846, DOI 10.1063/1.458270.

[213] Li, L., Pabit, S. A., Lamb, J. S., Park, H. Y., and Pollack, L., *Closing the lid on DNA end-to-end stacking interactions*, Appl. Phys. Lett. **92** (2008), no. 22, DOI 10.1063/1.2937402.

[214] Maffeo, C., Luan, B., and Aksimentiev, A., *End-to-end attraction of duplex DNA*, Nucleic Acids Res. **40** (2012), no. 9, 3812–3821, DOI 10.1093/nar/gkr1220.

[215] Rauzan, B., Mcmichael, E., Cave, R., Sevcik, L. R., Ostrosky, K., Whitman, E., Stegemann, R., Sinclair, A. L., Serra, M. J., and Deckert, A. A., *Kinetics and Thermodynamics of DNA, RNA, and Hybrid Duplex Formation*, Biochemistry **52** (2013), no. 5, 765–772, DOI 10.1021/bi3013005.

[216] Nonin, S., Leroy, J. L., and Guéron, M., *Terminal Base Pairs of Oligodeoxynucleotides: Imino Proton Exchange and Fraying*, Biochemistry **34** (1995), no. 33, 10652–10659, DOI 10.1021/bi00033a041.

[217] Wärmländer, S., Sen, A., and Leijon, M., *Imino Proton Exchange in DNA Catalyzed by Ammonia and Trimethylamine: Evidence for a Secondary Long-Lived Open State of the Base Pair*, Biochemistry **39** (2000), no. 3, 607–615, DOI 10.1021/bi991863b.

[218] Lane, A. N., *N.m.r. assignments and temperature-dependent conformational transitions of a mutant trp operator-promoter in solution*, Biochem. J. **259** (1989), no. 3, 715–724, DOI 10.1042/bj2590715.

[219] Wenke, B. B., Huiting, L. N., Frankel, E. B., Lane, B. F., and Núñez, M. E., *Base Pair Opening in a Deoxynucleotide Duplex Containing a*

cis-syn *Thymine Cyclobutane Dimer Lesion*, Biochemistry **52** (2013), no. 51, 9275–9285, DOI 10.1021/bi401312r.

[220] Vaitiekunas, P., Crane-Robinson, C., and Privalov, P. L., *The energetic basis of the DNA double helix: a combined microcalorimetric approach*, Nucleic Acids Res. **43** (2015), no. 17, 8577–8589, DOI 10.1093/nar/gkv812.

[221] El Antri, S., Mauffret, O., Monnot, M., Lescot, E., Convert, O., and Fermandjian, S., *Structural deviations at CpG provide a plausible explanation for the high frequency of mutation at this site. Phosphorus Nuclear Magnetic Resonance and Circular Dichroism Studies*, J. Mol. Biol. **230** (1993), no. 2, 373–378, DOI 10.1006/jmbi.1993.1153.

[222] Dallmann, A., Dehmel, L., Peters, T., Mügge, C., Griesinger, C., Tuma, J., and Ernsting, N. P., *2-Aminopurine incorporation perturbs the dynamics and structure of DNA*, Angew. Chem. Int. Ed. **49** (2010), no. 34, 5989–5992, DOI 10.1002/anie.201001312.

[223] Socha, O., *Charakterizace strukturních vlastností a stability DNA vlásenek pomocí NMR spektroskopie*, Master thesis, Charles University, 2016.

[224] Geyer, M., Schweins, T., Herrmann, C., Prisner, T., Wittinghofer, A., and Kalbitzer, H. R., *Conformational transitions in p21ras and in its complexes with the effector protein Raf-RBD and the GTPase activating protein GAP*, Biochemistry **35** (1996), no. 32, 10308–10320, DOI 10.1021/bi952858k.

[225] Audet, P., Simard, C., and Savoie, R., *A vibrational spectroscopic study of the self-association of* 5′*-GMP in aqueous solution*, Biopolymers **31** (1991), no. 2, 243–251, DOI 10.1002/bip.360310211.

[226] Kozic, J., Novotná, E., Volková, M., Stolaříková, J., Trejtnar, F., Wsól, V., and Vinšová, J., *Synthesis and in vitro antimycobacterial and isocitrate lyase inhibition properties of novel 2-methoxy-*2′*-hydroxybenzanilides, their thioxo analogues and benzoxazoles*, Eur. J. Med. Chem. **56** (2012), 108–119, DOI 10.1016/j.ejmech.2012.08.016.

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

1D	one-dimensional
2D	two-dimensional
A	adenine
C	cytosine
COSY	correlation spectroscopy
DFT	density functional theory
dGMP	2′-deoxyriboguanosine 5′-monophosphate
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
DSC	differential scanning calorimetry
DSS	4,4-dimethyl-4-silapentane-1-sulfonate
FID	free induction decay
FWHM	full width at half maximum
G	guanine
GMP	guanosine 5′-monophosphate
HMBC	heteronuclear multiple-bond coherence
hm^5C	5-hydroxymethylcytosine
m^5C	5-methylcytosine
MD	molecular dynamics
NA	nucleic acid
ncRNA	non-coding ribonucleic acid
NMR	nuclear magnetic resonance
NN	nearest neighbour
NOE	nuclear Overhauser effect
NOESY	nuclear Overhauser enhancement spectroscopy
ODN	oligodeoxynucleotide
PDB	Protein Data Bank
qNMR	quantitative nuclear magnetic resonance spectroscopy
rGMP	riboguanosine 5′-monophosphate
RNA	ribonucleic acid

RS	Raman scattering
SRE	serum response element
SRF	serum response factor
SVD	singular-value decomposition
T	thymine
TOCSY	total correlation spectroscopy
TPPI	time-proportional phase incrementation
UV	ultraviolet
VT	variable temperature
XRD	X-ray diffraction

LIST OF CONFERENCE CONTRIBUTIONS

I. Římal, V., Štěpánková, H., and Štěpánek, J., *DNA Duplex Melting Monitored via Lineshape Analysis of ^1H NMR Spectra*, poster, 8th Discussions in Structural Molecular Biology, Nové Hrady, Nové Hrady, Czech Republic (2010).

II. Římal, V., Štěpánková, H., and Štěpánek, J., *DNA Duplex Melting Monitored via Lineshape Analysis of ^1H NMR Spectra*, talk, NMR Valtice 2010, 25th Central European NMR Meeting, Valtice, Czech Republic (2010).

III. Římal, V., Štěpánková, H., and Štěpánek, J., *DNA Duplex Melting Monitored via Lineshape Analysis of ^1H NMR Spectra*, talk, 19th Annual Student Conference, Week of Doctoral Students 2010, Prague, Czech Republic (2010).

IV. Římal, V., Štěpánková, H., and Štěpánek, J., *DNA Duplex Melting Monitored via Lineshape Analysis of ^1H NMR Spectra*, poster, Ampere NMR school, Poznań and Wierzba, Poland (2010).

V. Římal, V., Štěpánková, H., and Štěpánek, J., *Asymmetrical Chemical Exchange During the Melting of DNA Double-Helix*, poster, EMBO Multidimensional NMR in Structural Biology, Il Ciocco (LU), Italy (2010).

VI. Římal, V., Štěpánková, H., and Štěpánek, J., poster, *Sequence-Specific Local Stability and Dynamics of DNA Double-Helix Studied by Hydrogen NMR*, EUROMAR, Frankfurt am Main, Germany (2011).

VII. Římal, V., Štěpánková, H., and Štěpánek, J., *Sequence-Specific Local Stability and Dynamics of DNA Double-Helix Studied by Hydrogen NMR*, poster, Sixth Cambridge Symposium on Nucleic Acids Chemistry and Biology, Cambridge, United Kingdom (2011).

VIII. Římal, V., Štěpánková, H., and Štěpánek, J., talk, *Influence of Cytosine Methylation on the Dynamics of DNA Double-Helix*, NMR Valtice

2012, 27th Central European NMR Meeting, Valtice, Czech Republic (2012).

IX. Římal, V., Štěpánková, H., and Štěpánek, J., *Effects of Cytosine Methylation and Mismatch Base Pairs on Local Melting Properties of DNA Double-Helix*, poster, 12th Chianti/INSTRUCT Workshop on BioNMR 2012, Montecatini Terme (PT), Italy (2012).

X. Římal, V., Štěpánková, H., and Štěpánek, J., *Stabilization of DNA Double-Helix by Cytosine Methylation Observed by Nuclear Magnetic Resonance Spectroscopy*, poster, The 5th EMBO Meeting, Amsterdam, Netherlands (2013).

XI. Římal, V., Socha, O., Štěpánková, H., and Štěpánek, J., *Cytosine Methylation Stabilizes All Base Pairs in a Short DNA Duplex*, talk, NMR Valtice 2015, 30th Central European NMR Meeting, Valtice, Czech Republic (2015).

XII. Římal, V., Lindnerová Mudroňová, K., and Mojzeš, P., *Thermal Changes and Concentration-Dependent Folding of GGGN Repeat*, poster, G4thering, 6th International Meeting on Quadruplex Nucleic Acids, Prague, Czech Republic (2017).

LIST OF PAPERS

Publications connected to this thesis

I. **Římal, V.**, Štěpánková, H., and Štěpánek, J., *Analysis of NMR Spectra in Case of Temperature-Dependent Chemical Exchange Between Two Unequally Populated Sites*, Concepts Magn. Reson. Part A **38A** (2011), no. 3, 117–127, DOI 10.1002/cmr.a.20214.

II. **Římal, V.**, Socha, O., Štěpánek, J., and Štěpánková, H., *Spectroscopic Study of Cytosine Methylation Effect on Thermodynamics of DNA Duplex Containing CpG Motif*, J. Spectrosc. **2015** (2015), Article ID 842810, DOI 10.1155/2015/842810.

III. Mudroňová, K., **Římal, V.**, and Mojzeš, P., *Effect of ribose versus 2′-deoxyribose residue in guanosine 5′-monophosphates on formation of G-quartets stabilized by potassium and sodium cations*, Vib. Spectrosc. **82** (2016), 60–65, DOI 10.1016/j.vibspec.2015.12.002

IV. Kozic, J., Novák, Z., **Římal, V.**, Profant, V., Kuneš, J., and Vinšová, J., *Conformations, equilibrium thermodynamics and rotational barriers of secondary thiobenzanilides*, Tetrahedron **72** (2016), no. 17, 2072–2083, DOI 10.1016/j.tet.2016.02.035

Other publications by the author

I. Sedláčková, E., **Římal, V.**, Štěpánková, H., and Štěpánek, J., *NMR Investigation of Magnesium Specific Binding to DNA Duplex*, AIP Conf. Series, vol. 1204, 2010, ISBN 978-0-73540-741-1, pp. 236–238, DOI 10.1063/1.3295660

II. Uličná, K., Štěpánková, H., **Římal, V.**, Hanzlíček, T., and Perná, I., *NMR Study of Aluminium Coordination in Clays*, WDS'13 Proceedings of Contributed Papers, Part III, 2013, ISBN 978-8-07378-252-8, pp. 104–109.

III. Tesař, T., Mušálek, R., Medřický, J., Kotlan, J., Lukáč, F., Pala, Z., Ctibor, P., Chráska, T., Houdková, Š., **Římal, V.**, and Curry, N., *Development of suspension plasma sprayed alumina coatings with high enthalpy plasma torch*, Surf. Coat. Technol. **325** (2017), 277–288, DOI 10.1016/j.surfcoat.2017.06.039.

A APPENDIX

A.1 Assignment of ^{1}H resonances from NOESY

This section contains figures with H1′–H6/H8 sequential walks in ^{1}H NOESY spectra of all octamer duplexes. These figures form the basis of the resonance assignments described in § 7.2.1 on page 62 for unmethylated and in § 8.2 on page 123 for methylated ODN.

Portions of ^{1}H NOESY spectra that contain the H1′–H6/H8 are shown. Sequential walks are depicted by dashed magenta lines and the intra-nucleotide cross peaks are labelled. Extra cross-peaks not involved in these sequential walks are annotated, if the assignments are known.

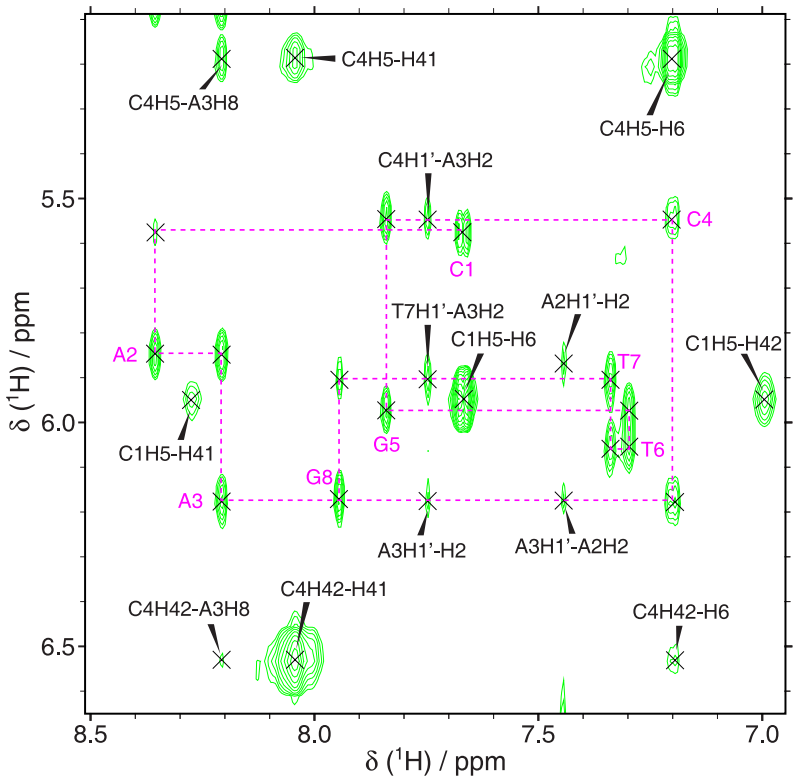


Figure A.1 CAACGTTG: the H1′–H6/H8 sequential walk

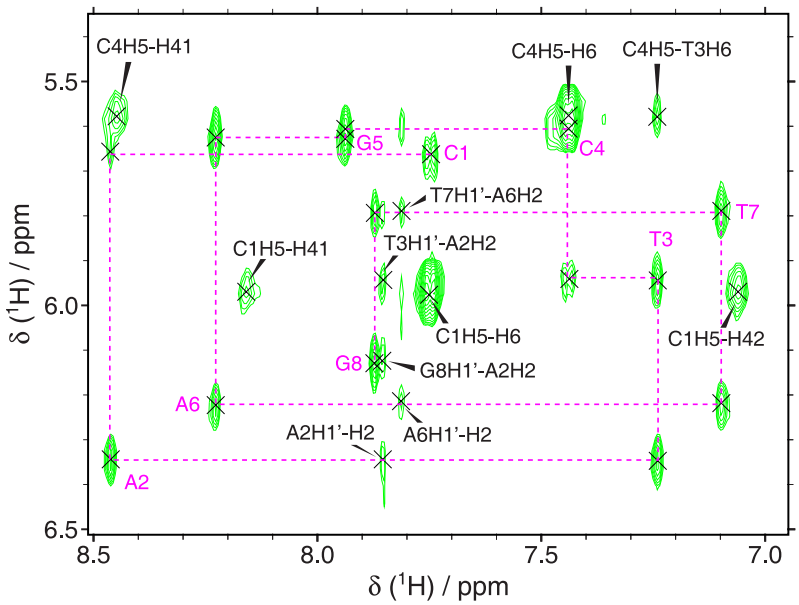


Figure A.2 CATCGATG: the H1′–H6/H8 sequential walk

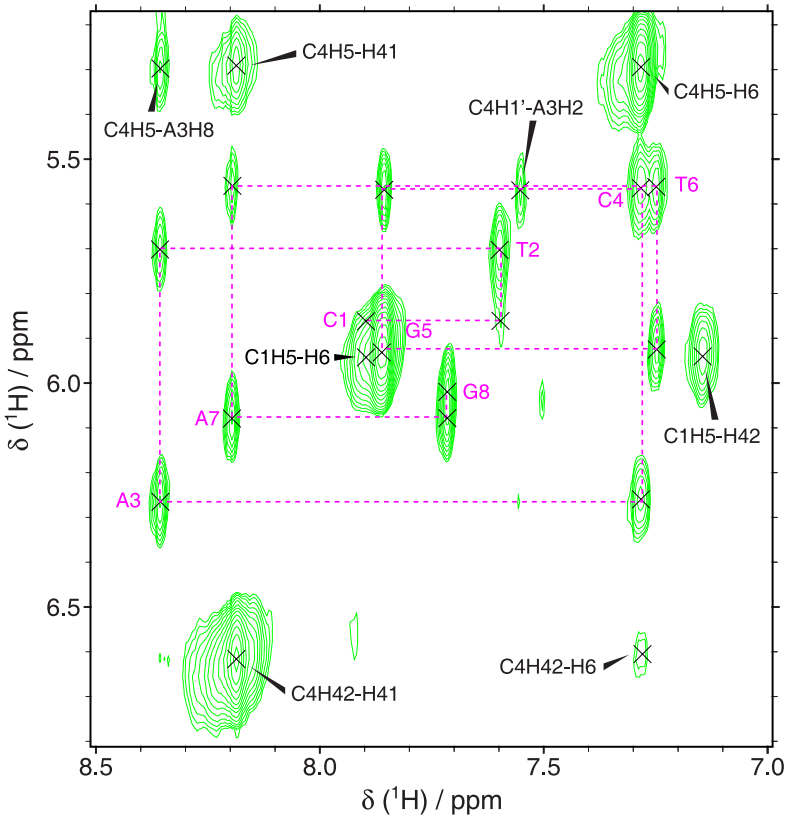


Figure A.3 CTACGTAG: the H1′–H6/H8 sequential walk

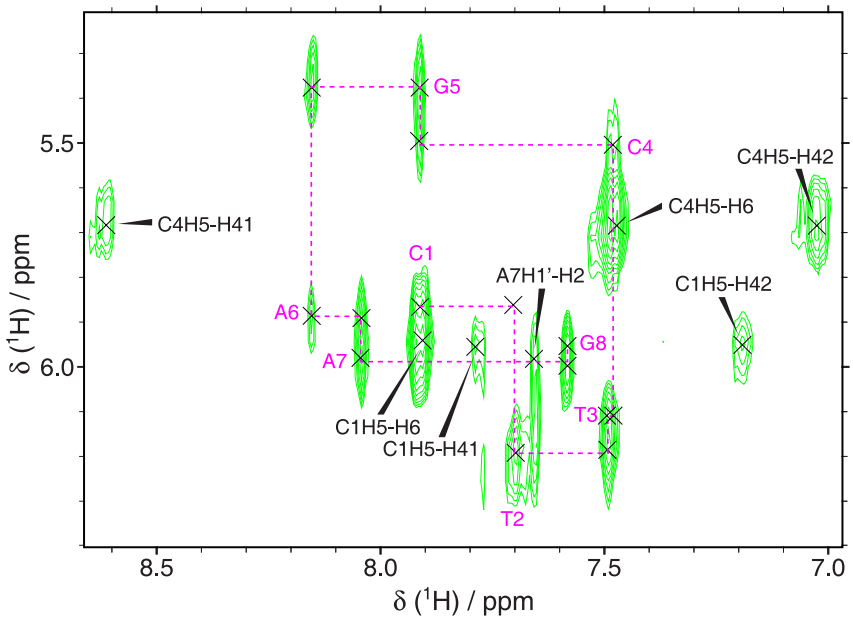


Figure A.4 CTTCGAAG: the H1′–H6/H8 sequential walk

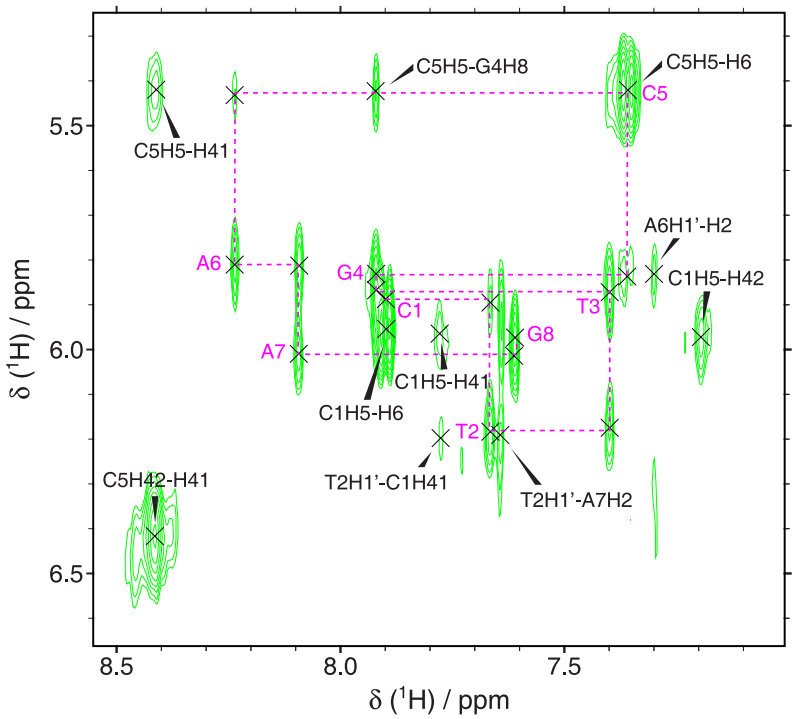


Figure A.5 CTTGCAAG: the H1′–H6/H8 sequential walk

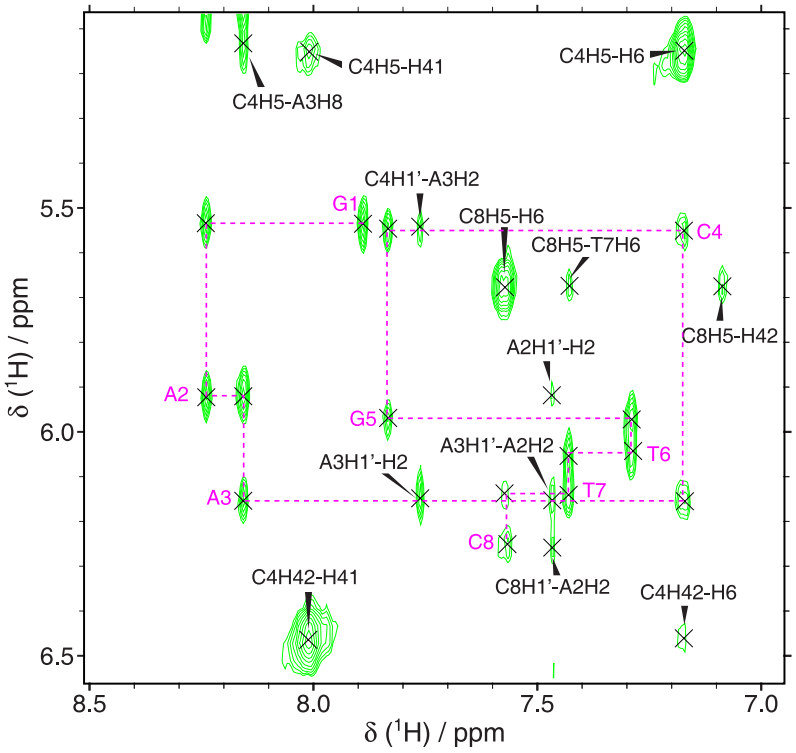


Figure A.6 GAACGTTC: the H1′–H6/H8 sequential walk

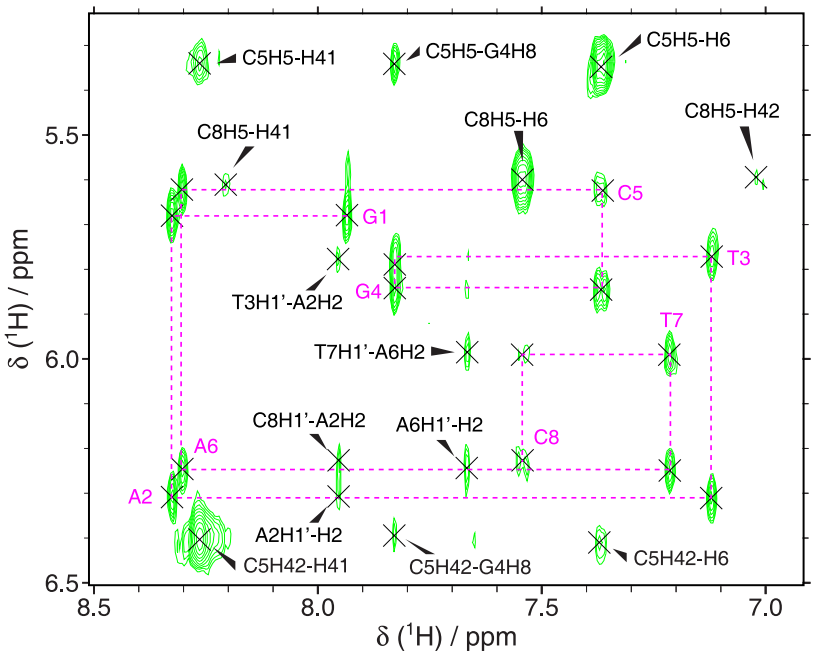


Figure A.7 GATGCATC: the H1′–H6/H8 sequential walk

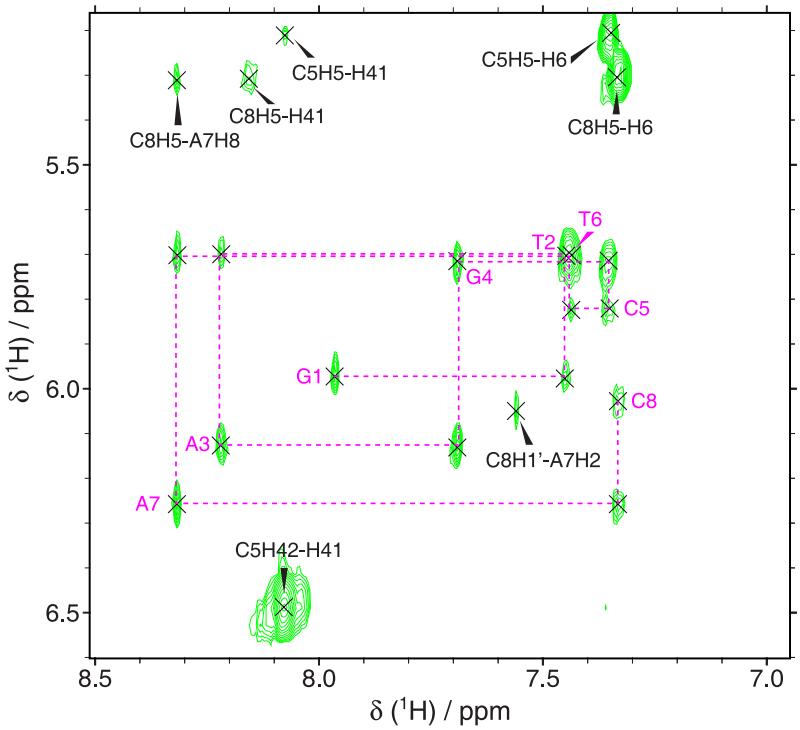


Figure A.8 GTAGCTAC: the H1′–H6/H8 sequential walk

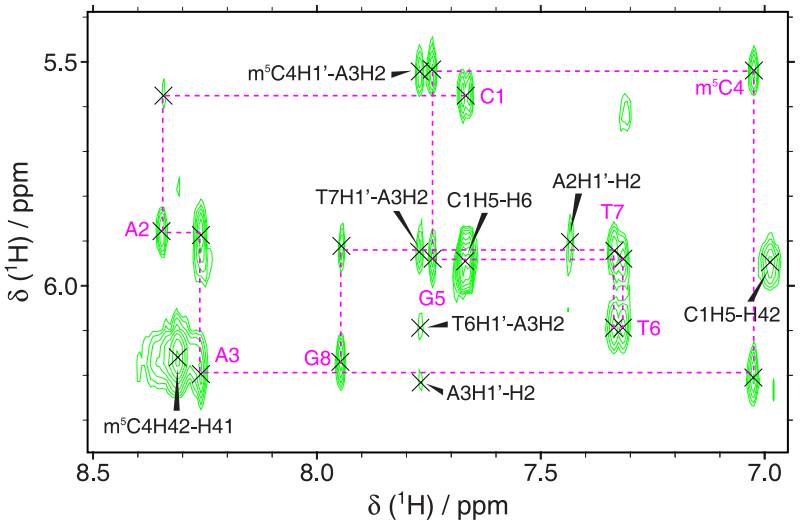


Figure A.9 CAAm^{5}CGTTG: the H1′–H6/H8 sequential walk

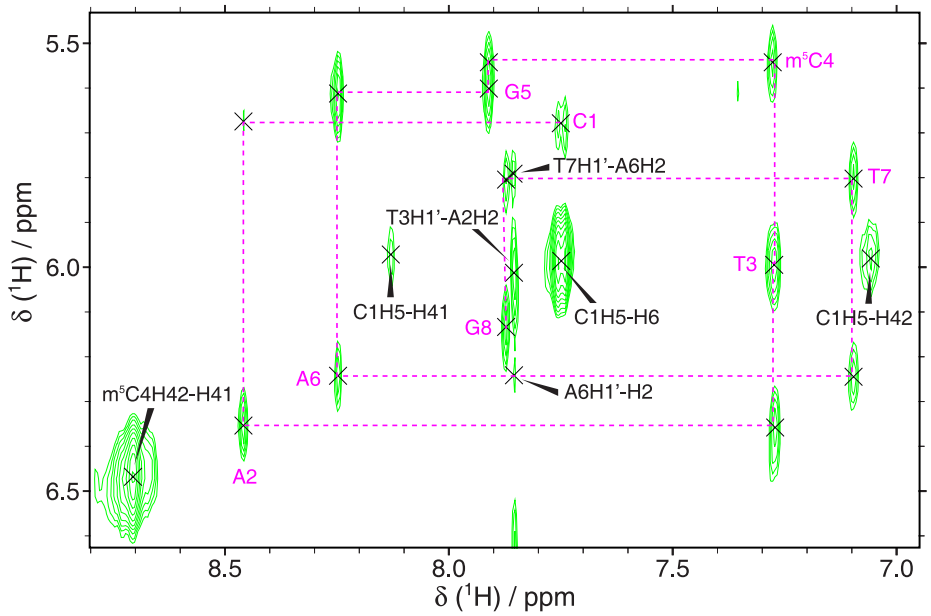


Figure A.10 CATm^{5}CGATG: the H1′–H6/H8 sequential walk

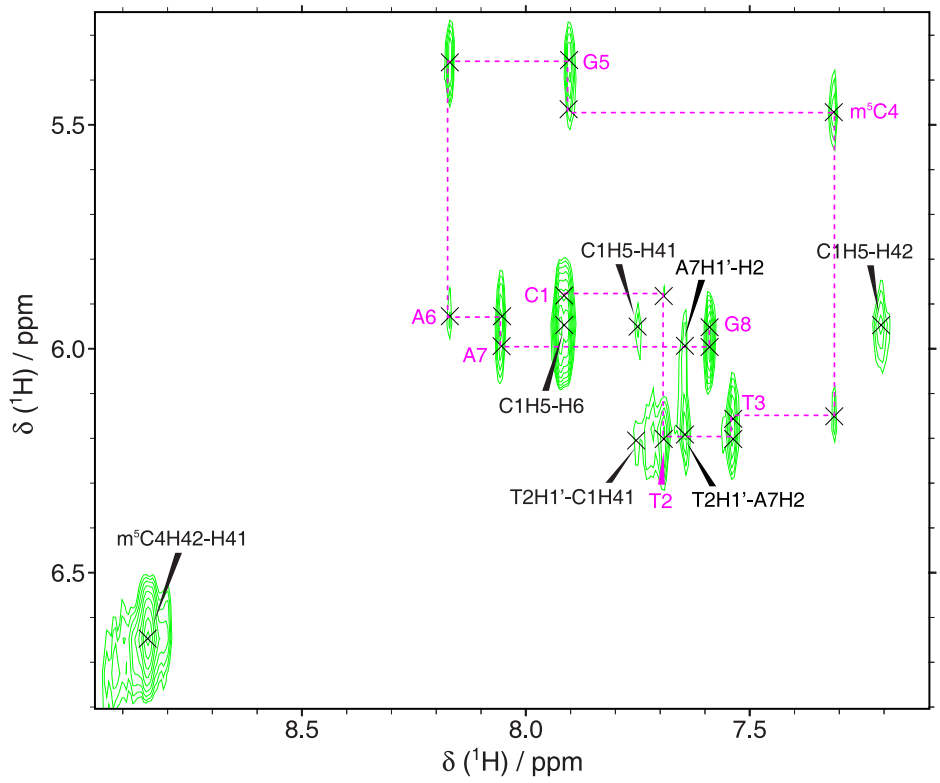


Figure A.11 CTTm^{5}CGAAG: the H1′–H6/H8 sequential walk

A.2 Chemical shifts of ^{1}H resonances

This section accompanies the texts of § 7.2 on page 61, § 8.2 on page 123, and § 9.1.1 on page 143.

The ^{1}H chemical shifts of all ODN samples (self-complementary octamers without and with m^{5}C, as well as sequences based on SRE) are collected in the following tables.

A.2.1 Duplex

Table A.1 ^{1}H chemical shifts of self-complementary DNA strands in duplex. All values obtained from Lorentzian-line fitting of 1D spectra at temperature T_{1D} except for H2′ and H2″ protons which come from 2D NOESY acquired at T_{2D}

	δ/ppm							
	Non-exchangeable					Exchangeable		
	Base		Deoxyribose			Amino		Imino
	H6/8	H2/5/7	H1′	H2′	H2″	H41	H42	H1/3
CAACGTTG			$T_{1D} = 283.5$ K			$T_{2D} = 282.2$ K		
C1	7.668	5.951	5.577	1.818	2.325	8.271	6.995	
A2	8.356	7.443	5.851	2.835	2.904			
A3	8.208	7.747	6.177	2.642	2.871			
C4	7.200	5.189[a]	5.547	1.982	2.342	8.044	6.533	
G5	7.840		5.972	2.633	2.778			12.738
T6	7.298	1.389	6.057	2.083	2.549			13.971
T7	7.339	1.704	5.904	1.979	2.394			14.160
G8	7.945		6.170	2.636	2.372			12.850
CATCGATG			$T_{1D} = 283.5$ K			$T_{2D} = 281.4$ K		
C1	7.747	5.973	5.669	2.095	2.481	8.142	7.053	
A2	8.458	7.856	6.349	2.787	3.001			
T3	7.236	1.412	5.946	2.074	2.471			13.627
C4	7.436	5.580	5.611	1.996	2.394	8.439	6.832	
G5	7.933		5.628	2.736	2.800			12.666
A6	8.221	7.813	6.222	2.616	2.911			
T7	7.093	1.429	5.795	1.855	2.314			13.713
G8	7.870		6.134	2.577	2.335			12.820

Table A.1 (continued)

	δ/ppm							
	H6/8	H2/5/7	H1′	H2′	H2″	H41	H42	H1/3
CTACGTAG			$T_{1D} = 283.5$ K			$T_{2D} = 284$ K[b]		
C1	7.858	5.929	5.856	2.190	2.559	7.905	7.146	
T2	7.598	1.722	5.698	2.263	2.556			13.823
A3	8.355	7.546	6.253	2.768	2.907			
C4	7.280	5.284	5.558	2.032	2.342	8.195	6.614	
G5	7.855		5.914	2.609	2.721			12.729
T6	7.247	1.505	5.555	2.029	2.303			13.743
A7	8.194	7.499	6.069	2.700	2.873			
G8	7.713		6.000	2.428	2.271			13.137
CTTCGAAG			$T_{1D} = 283.5$ K			$T_{2D} = 283.5$ K		
C1	7.911	5.954	5.858	2.333	2.607	7.782	7.190	
T2	7.698	1.706	6.201	2.284	2.654			14.260
T3	7.488	1.685	6.135	2.205	2.556			13.982
C4	7.478	5.683	5.504	2.002	2.326	8.615	7.024	
G5	7.911		5.372	2.713[c]				12.736
A6	8.152	7.365	5.906	2.670	2.870			
A7	8.043	7.653	5.997	2.569	2.826			
G8	7.581		5.972	2.327	2.243		13.329	
CTTGCAAG			$T_{1D} = 283.6$ K			$T_{2D} = 283.6$ K		
C1	7.896	5.958	5.880	2.269	2.591	7.773	7.191	
T2	7.664	1.699	6.180	2.247	2.617			14.200
T3	7.399	1.721	5.868	2.173	2.506			13.935
G4	7.921		5.832	2.665[c]				12.688
C5	7.359	5.419	5.429	1.905	2.261	8.412	6.420	
A6	8.235	7.296	5.811	2.753	2.842			
A7	8.092	7.637	6.011	2.589	2.830			
G8	7.609		5.959	2.334[c]				13.268

Table A.1 (continued)

	δ/ppm							
	H6/8	H2/5/7	H1′	H2′	H2″	H41	H42	H1/3
GAACGTTC			$T_{1D} = 283.5$ K			$T_{2D} = 285.7$ K		
G1	7.892		5.540	2.526	2.731			[d]
A2	8.238	7.460	5.916	2.786	2.904			
A3	8.157	7.758	6.151	2.611	2.857			
C4	7.174	5.154[a]	5.541	1.961	2.325	8.012	6.476	
G5	7.834		5.966	2.639	2.766			12.722
T6	7.291	1.381	6.053	2.108	2.586			13.981
T7	7.425	1.655	6.134	2.137	2.550			14.144
C8	7.560	5.646	6.251	2.286[c]		8.314	7.080	
GATGCATC			$T_{1D} = 283.8$ K			$T_{2D} = 283.8$ K		
G1	7.934		5.682	2.627	2.787			12.615
A2	8.324	7.952	6.309	2.747	2.972			
T3	7.119	1.375	5.768	2.048	2.440			13.500
G4	7.828		5.848	2.599	2.689			12.538
C5	7.367	5.343	5.622	2.129	2.433	8.264	6.402	
A6	8.303	7.665	6.249	2.666	2.919			
T7	7.212	1.426	5.988	2.007	2.452			13.763
C8	7.537	5.588	6.233	2.256[c]		8.206	7.007	
GTAGCTAC			$T_{1D} = 283.5$ K			$T_{2D} = 284.2$ K		
G1	7.962		5.968	2.669	2.777			12.787
T2	7.446	1.380	5.692	2.233	2.522			13.739
A3	8.218	7.419	6.122	2.759	2.936			
G4	7.689		5.714	2.544	2.593			12.803
C5	7.352	5.207[a]	5.825	2.037	2.457	8.074	6.485	
T6	7.435	1.639	5.702	2.141	2.476			13.798
A7	8.315	7.557	6.256	2.680	2.875			
C8	7.326	5.290	6.034	2.108[c]		8.149	6.816	

[a]Peak position in the region of water signal doesn't permit direct analysis of 1D spectrum, value from NOESY presented instead. [b]Temperature was not calibrated at the corresponding conditions. [c]Separate signals of the diastereotopic protons H2′ and H2″ have not been distinguished. [d]The terminal imino proton not observed

Table A.2 ^{1}H chemical shifts of self-complementary DNA strands with 5-methylcytosine in duplex. All values obtained from Lorentzian-line fitting of 1D spectra at temperature T_{1D} except for H2′ and H2″ protons which come from 2D NOESY acquired at T_{2D}

	δ/ppm							
	Non-exchangeable					Exchangeable		
	Base		Deoxyribose			Amino		Imino
	H6/8	H2/5/7	H1′	H2′	H2″	H41	H42	H1/3
CAAm^{5}CGTTG			$T_{1D} = 283.5$ K			$T_{2D} = 282.2$ K		
C1	7.668	5.949	5.573	1.818	2.323	8.277	6.989	
A2	8.347	7.433	5.881	2.837	2.915			
A3	8.260	7.770	6.211	2.616	2.953			
m^{5}C4	7.025	1.556	5.522	2.030	2.351	8.313	6.163	
G5	7.743		5.940	2.593	2.758			12.729
T6	7.316	1.316	6.093	2.119	2.558			13.978
T7	7.337	1.695	5.919	1.988	2.407			14.106
G8	7.947		6.169	2.641	2.374			12.854
CATm^{5}CGATG			$T_{1D} = 283.5$ K			$T_{2D} = 283.5$ K		
C1	7.747	5.978	5.667	2.097	2.483	8.126	7.056	
A2	8.458	7.852	6.344	2.808	2.991			
T3	7.273	1.415	5.987	2.006	2.554			13.589
m^{5}C4	7.273	1.706	5.537	2.054	2.362	8.704	6.461	
G5	7.910		5.611	2.725	2.784			12.680
A6	8.246	7.852	6.237	2.635	2.919			
T7	7.095	1.446	5.789	1.860	2.312			13.706
G8	7.872		6.130	2.579	2.334			12.816
CTTm^{5}CGAAG			$T_{1D} = 283.5$ K			$T_{2D} = 283.5$ K		
C1	7.914	5.966	5.872	2.312	2.617	7.750	7.206	
T2	7.689	1.702	6.207	2.293	2.638			14.247
T3	7.538	1.702	6.154	2.143	2.611			13.937
m^{5}C4	7.312	1.799	5.466	2.021	2.295	8.842	6.649	
G5	7.903		5.358	2.687[a]				12.751
A6	8.170	7.395	5.912	2.680	2.870			
A7	8.056	7.645	5.998	2.578	2.830			
G8	7.589		5.965	2.336	2.246			13.336

[a]Separate signals of the diastereotopic protons H2′ and H2″ have not been distinguished

A.2.2 Single strands

Table A.3 ^{1}H chemical shifts of self-complementary DNA strands in single strands. All values obtained from Lorentzian line-shape fitting of 1D spectra at temperatures T indicated

	δ/ppm				δ/ppm		
	H6/8	H2/5/7	H1′		H6/8	H2/5/7	H1′
CAACGTTG	$T = 354.3$ K			CTTCGAAG	$T = 354.3$ K		
C1	7.515	5.932	6.026	C1	7.735	5.992	6.180
A2	8.150	8.065	6.109	T2	7.598	1.862	6.213
A3	8.234	7.982	6.253	T3	7.559	1.856	6.182
C4	7.578	5.850	6.106	C4	7.617	5.976	6.130
G5	7.880		6.097	G5	7.770		5.773
T6	7.501	1.746	6.156	A6	8.106	7.956	6.042
T7	7.488	1.833	6.118	A7	8.138	7.965	6.170
G8	7.950		6.211	G8	7.887		6.097
CATCGATG	$T = 354.3$ K			CTTGCAAG	$T = 354.8$ K		
C1	7.553	5.945	6.059	C1	7.733	5.990	6.179
A2	8.332	8.170	6.376	T2	7.595	1.868	6.206
T3	7.464	1.746	6.133	T3	7.461	1.847	6.096
C4	7.610	5.966	6.106	G4	7.915		6.085
G5	7.769		5.850	C5	7.547	5.840	6.070
A6	8.273	8.053	6.315	A6	8.132	8.016	6.073
T7	7.417	1.743	6.090	A7	8.159	7.968	6.174
G8	7.927		6.175	G8	7.893		6.116
CTACGTAG	$T = 354.3$ K			GAACGTTC	$T = 354.3$ K		
C1	7.700	5.969	6.159	G1	7.735		5.907
T2	7.411	1.847	6.039	A2	8.120	7.978	6.081
A3	8.292	8.144	6.322	A3	8.225	7.966	6.244
C4	7.580	5.848	6.122	C4	7.578	5.845	6.105
G5	7.883		6.061	G5	7.880		6.100
T6	7.311	1.710	5.974	T6	7.505	1.745	6.172
A7	8.188	8.117	6.236	T7	7.568	1.843	6.207
G8	7.912		6.131	C8	7.829	6.053	6.248

Table A.3 (continued)

	δ/ppm				δ/ppm		
	H6/8	H2/5/7	H1′		H6/8	H2/5/7	H1′
GATGCATC	$T = 353.7$ K			GTAGCTAC	$T = 354.3$ K		
G1	7.759		5.959	G1	7.866		6.144
A2	8.283	8.078	6.326	T2	7.367	1.767	6.005[a]
T3	7.373	1.735	6.052	A3	8.168	8.094	6.174
G4	7.888		6.045	G4	7.870		6.019[a]
C5	7.565	5.832	6.098	C5	7.706	5.870	6.193
A6	8.312	8.121	6.351	T6	7.386	1.816	6.036[a]
T7	7.483	1.739	6.175	A7	8.321	8.156	6.378
C8	7.797	6.025	6.220	C8	7.779	5.947	6.225

[a] Assignments not possible due to mutual overlaps; values shown in increasing order

Table A.4 ^{1}H chemical shifts of self-complementary DNA strands with 5-methylcytosine in single strands. All values obtained from Lorentzian line-shape fitting of 1D spectra at temperatures T indicated

	δ/ppm				δ/ppm		
	H6/8	H2/5/7	H1′		H6/8	H2/5/7	H1′
CAAm^{5}CGTTG	$T = 354.3$ K			CTTm^{5}CGAAG	$T = 354.3$ K		
C1	7.510	5.927	6.025	C1	7.735	5.992	6.177
A2	8.162	8.061	6.113	T2	7.595	1.857	6.205
A3	8.232	7.980	6.240	T3	7.551	1.849	6.154
m^{5}C4	7.344	1.764	6.100	m^{5}C4	7.416	1.910	6.134
G5	7.875		6.089	G5	7.774		5.755
T6	7.500	1.745	6.152	A6	8.105	7.959	6.038
T7	7.488	1.832	6.118	A7	8.132	7.969	6.171
G8	7.949		6.212	G8	7.886		6.089
CATm^{5}CGATG	$T = 354.3$ K						
C1	7.551	5.945	6.057				
A2	8.333	8.166	6.372				
T3	7.468	1.747	6.111				
m^{5}C4	7.414	1.888	6.111				
G5	7.767		5.844				
A6	8.271	8.052	6.314				
T7	7.414	1.740	6.089				
G8	7.926		6.175				

A.2.3 SRE – folded state

Table A.5 ^{1}H chemical shifts of folded s1fos16, TGTCCATATTAG-GACA. All values were obtained from Lorentzian-line fitting of 1D spectra at 284 K except for H2′ and H2″ protons which come from 2D NOESY acquired at 286 K

	δ/ppm							
	Non-exchangeable					Exchangeable		
	Base		Deoxyribose			Amino		Imino
	H6/8	H2/5/7	H1′	H2′	H2″	H41/61	H42	H1/3
T1	7.446	1.622	5.928	2.016	2.351			[c]
G2	8.102		6.091	2.796	2.900			12.665
T3	7.389	1.417	6.131	2.262	2.579			13.731
C4	7.583	5.686	5.973	2.111	2.427	8.418	6.934	
C5	7.228	5.675	5.970	1.500	1.919	8.558	7.152	
A6	8.309	8.072	6.175	2.749[b]		6.845		
T7	7.434	1.727	6.055	2.028	2.312			[c]
A8	8.201	7.931	6.136	2.581[b]				
T9	7.370	1.502	6.030	2.116	2.293			[c]
T10	7.186	1.703	5.718	1.915	2.187			[c]
A11	8.153	7.869	5.668	2.586[b]		6.748		
G12	7.876		5.448[a]	2.699[b]				12.797
G13	7.805		5.606	2.653	2.744			12.853
A14	8.163	7.927	6.198	2.606	2.831			
C15	7.234	5.313[a]	5.617	1.879	2.251	8.260	6.764	
A16	8.193	7.672	6.302	2.443	2.609			

[a]Spectral overlap doesn't permit direct analysis of 1D spectrum, value from NOESY presented instead. [b]Separate signals of the diastereotopic protons H2′ and H2″ have not been distinguished. [c]Imino proton not observed

Table A.6 Aromatic ^{1}H chemical shifts of folded s1fos12 and s1fos14 obtained from Lorentzian-line fitting of 1D spectra at 284 K

	δ/ppm			δ/ppm	
	H6/8	H2/5/7		H6/8	H2/5/7
GTCCATATTAGGAC			TCCATATTAGGA		
s1fos14			s1fos12		
G2	7.934				
T3	7.542	1.308	T3	7.565	1.733
C4	7.605	[b]	C4	7.714	[b]
C5	7.224	[b]	C5	7.355	[b]
A6	8.317	8.087	A6	8.346	7.968[c]
T7	7.438	1.730	T7	7.347[c]	1.701[c]
A8	8.201	7.931	A8	8.252	8.095[c]
T9	7.370	1.502	T9	7.370[c]	1.530
T10	7.188	1.707	T10	7.210	1.717[c]
A11	8.153	7.879	A11	8.153	7.950[c]
G12	7.879[a]		G12	7.849	
G13	7.808		G13	7.769	
A14	8.165	8.010	A14	8.214	7.978
C15	7.255	[b]			

[a]Spectral overlap with A11H2. [b]Resonance not analysed. [c]Signal too broad at 284 K, value from 292 K instead

A.2.4 SRE – unfolded state

Table A.7 ^{1}H chemical shifts of unfolded s1fos16, TGTCCATATTAG-GACA. All values presented were obtained from Lorentzian line-shape fitting of 1D spectra at 360 K

	δ/ppm		
	H6/8	H2/5/7	H1′
T1	7.439	1.844	6.094
G2	7.958		6.173
T3	7.515	1.782	6.190
C4	7.750	6.017	6.216
C5	7.621	5.996	6.097
A6	8.303	8.137	6.335
T7	7.320	1.720	5.996
A8	8.294	8.120	6.328
T9	7.459	1.723	6.114
T10	7.342	1.803	5.959
A11	8.155	8.094	6.142
G12	7.739		5.855
G13	7.746		5.793
A14	8.230	8.026	6.288
C15	7.625	5.882	6.123
A16	8.293	8.149	6.380

Table A.8 Aromatic ^{1}H chemical shifts of unfolded s1fos12 and s1fos14 obtained from Lorentzian-line fitting of 1D spectra at 360 K

	δ/ppm			δ/ppm	
	H6/8	H2/5/7		H6/8	H2/5/7
GTCCATATTAGGAC			TCCATATTAGGA		
s1fos14			s1fos12		
G2	7.889				
T3	7.536	1.790	T3	7.530	1.845
C4	7.752	[a]	C4	7.763	[a]
C5	7.625	[a]	C5	7.635	[a]
A6	8.300	8.135	A6	8.303	8.139
T7	7.323	1.725	T7	7.324	1.725
A8	8.290	8.118	A8	8.289	8.117
T9	7.460	1.722	T9	7.462	1.722
T10	7.344	1.803	T10	7.345	1.803
A11	8.157	8.098	A11	8.158	8.096
G12	7.748		G12	7.761	
G13	7.761		G13	7.785	
A14	8.262	8.069	A14	8.274	8.100
C15	7.761	[a]			

[a] Resonance not analysed

A.3 Chemical shifts and their temperature dependence in the same neighbourhood

The chemical shifts obtained for the same triplets or terminal doublets are shown in Fig. A.12 through Fig. A.15 as correlation plots between the experimental values, accompanying the text of § 7.3.1 on page 66. These graphs are constructed as follows: where exactly two occurrences of a nucleotide triplet are found within our ODN (these case are labelled as 'pairs' in the figure legends), each of the two shifts of the middle nucleotide is put on its own axis of the graph in a way that the created point lies under the diagonal. Where three occurrences of a nucleotide triplet are found ('triples'), the three middle-nucleotide shifts are coupled into three pairs and treated as above but placed over the diagonal.

The same approach is used for the slopes of chemical shifts with respect to temperature in the graphs from Fig. A.16 to Fig. A.19 that accompany the text of § 7.4.1 on page 81.

A.3.1 Chemical shifts

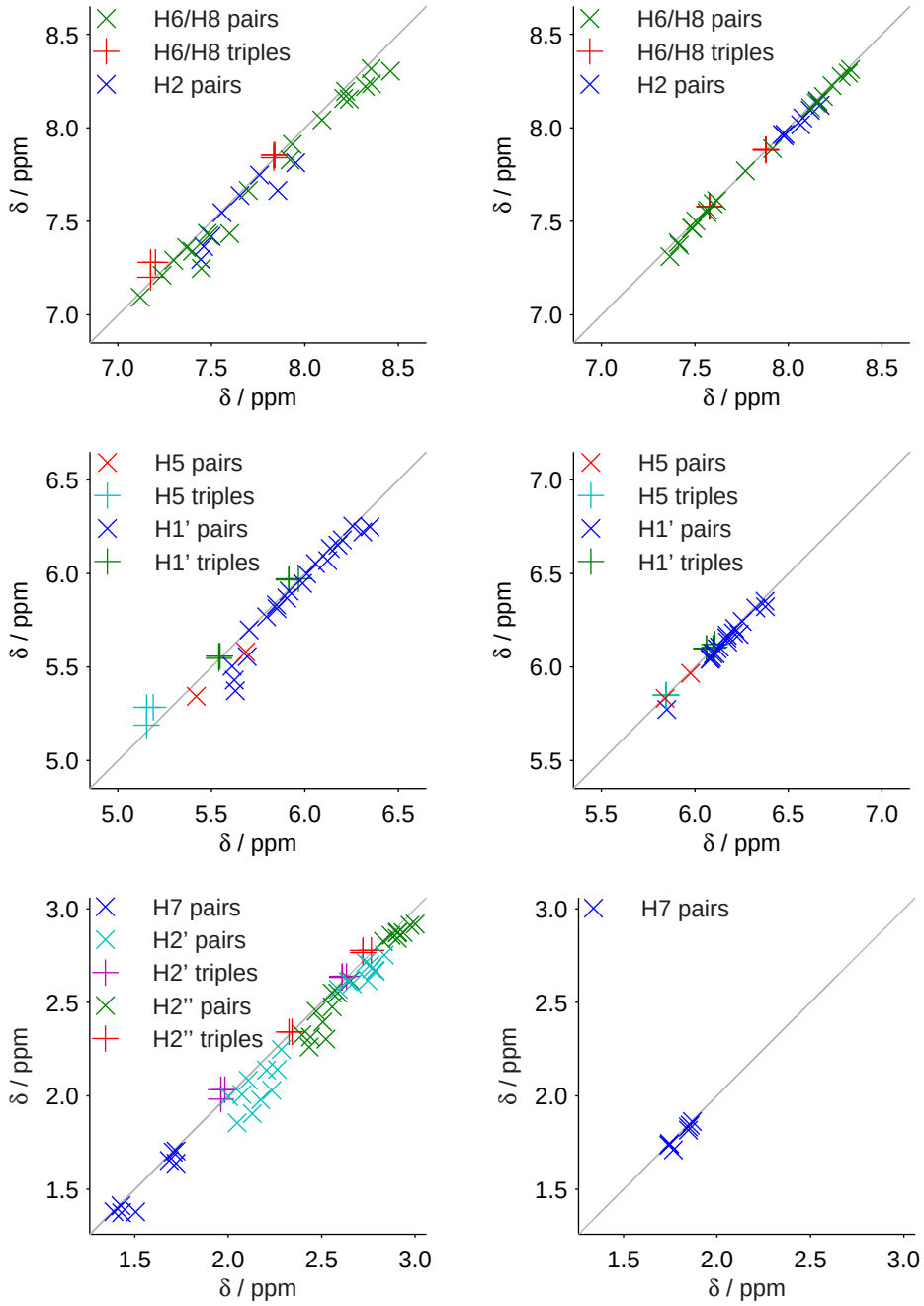


Figure A.12 Correlation plots of non-exchangeable chemical shifts in the same trinucleotides. Left: **duplex**; right: **single strands**

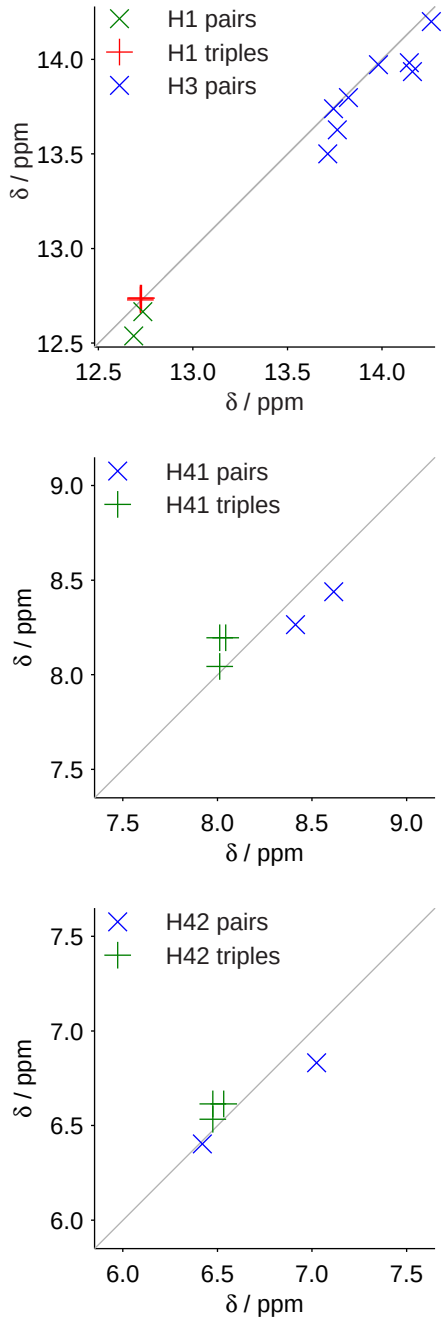


Figure A.13 Correlation plots of chemical shifts of exchangeable protons in the same trinucleotides in **duplex**

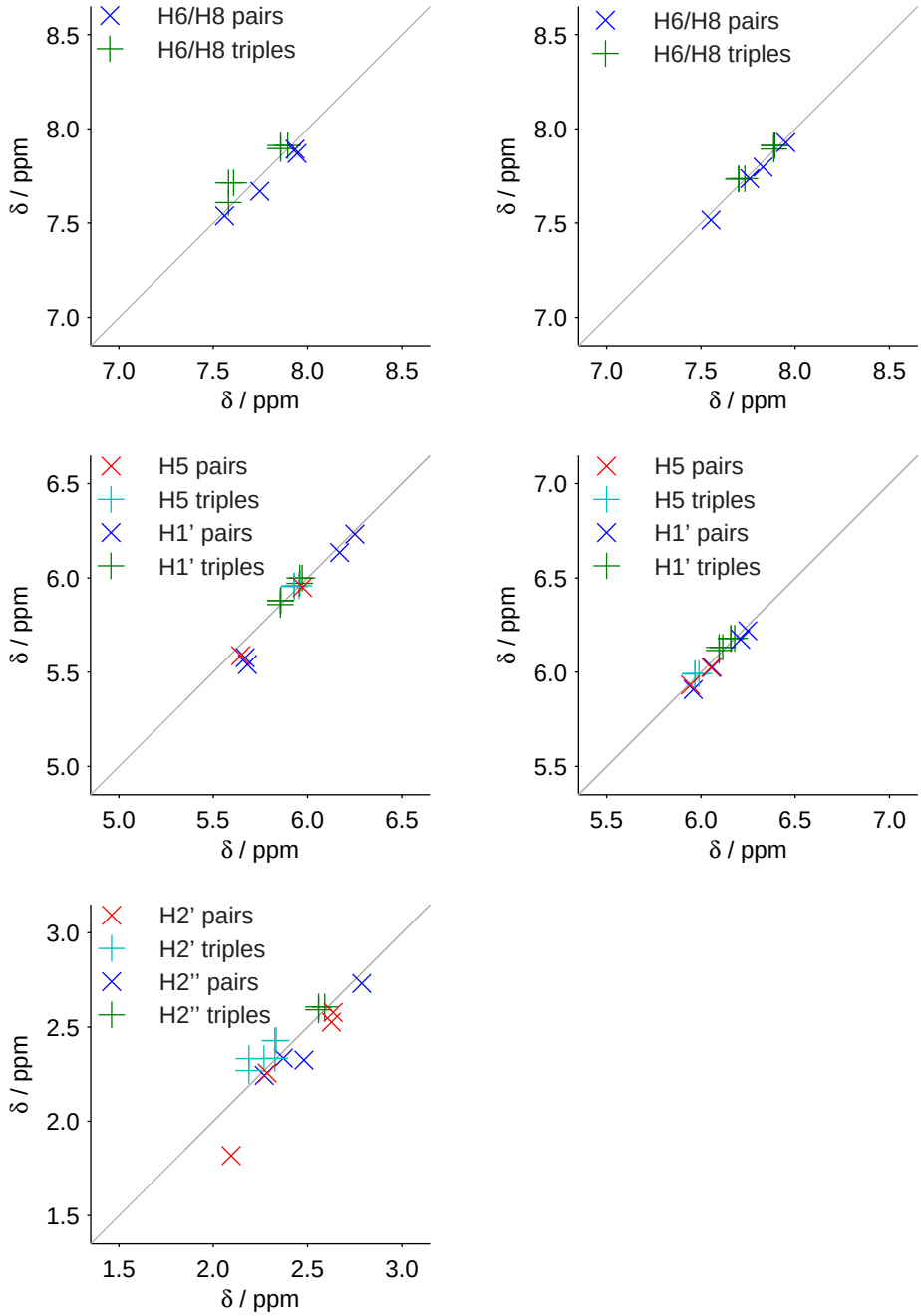


Figure A.14 Correlation plots of chemical shifts of non-exchangeable protons of terminal nucleotides with the same neighbours. Left: **duplex**; right: **single strands**

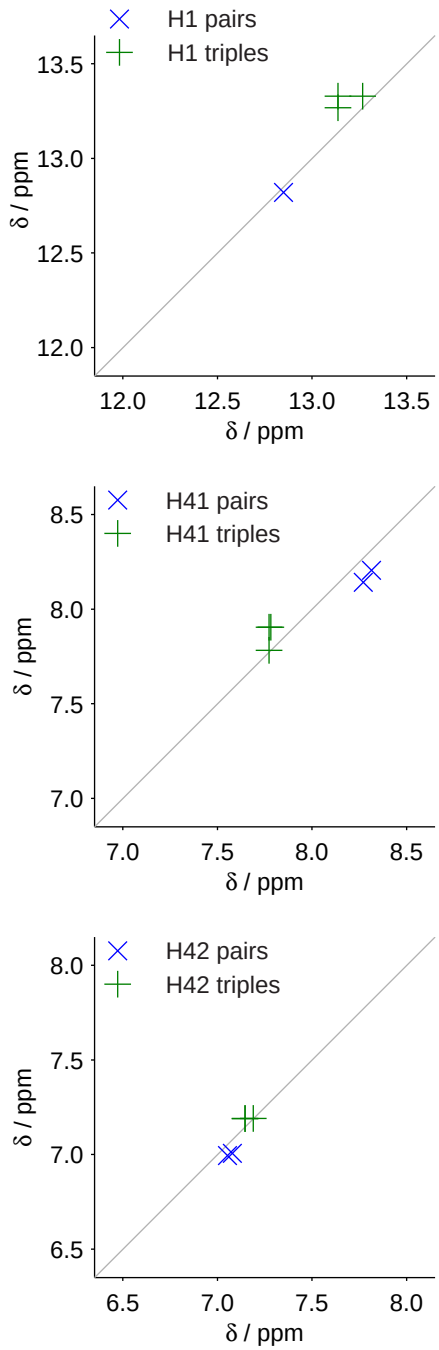


Figure A.15 Correlation plots of chemical shifts of exchangeable protons of terminal nucleotides with the same neighbours in **duplex**

A.3.2 Slopes of chemical shifts

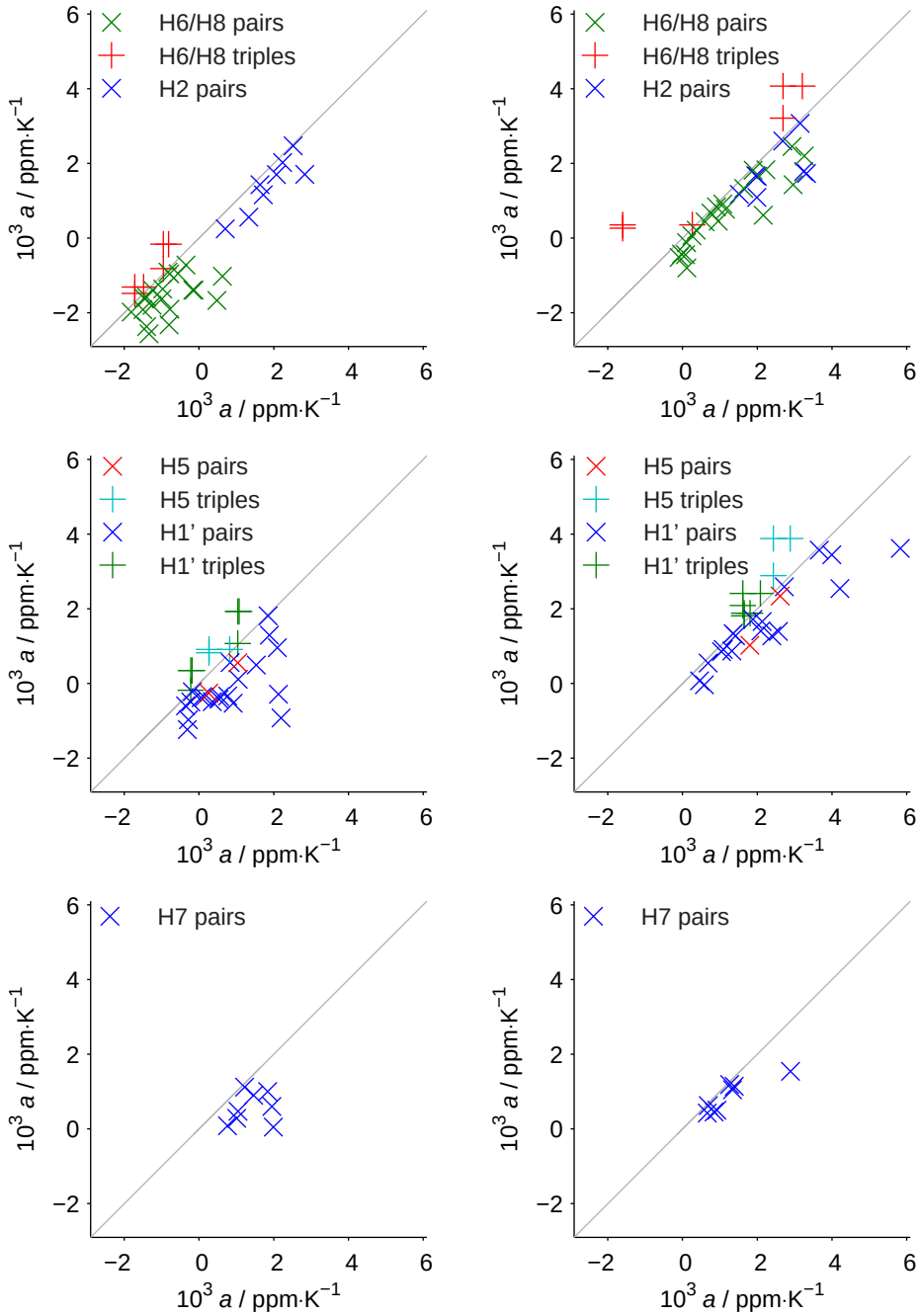


Figure A.16 Correlation plots of chemical-shift slopes in the same trinucleotides. Left: **duplex**; right: **single strands**

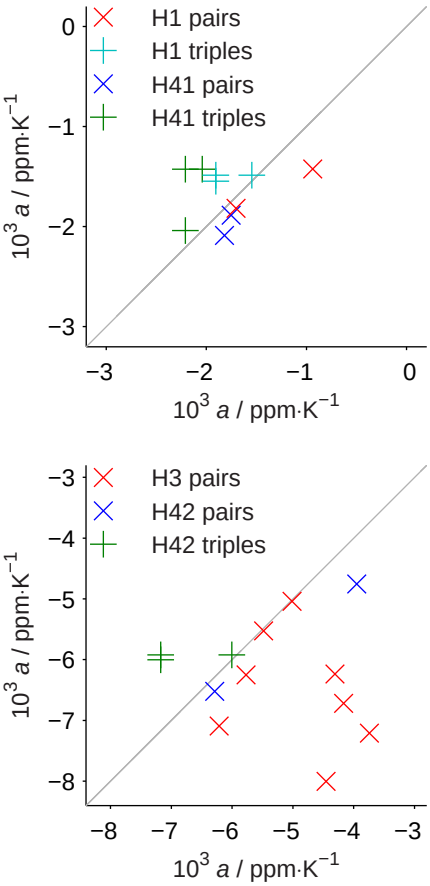


Figure A.17 Correlation plots of chemical-shift slopes of exchangeable protons in the same trinucleotides in **duplex**

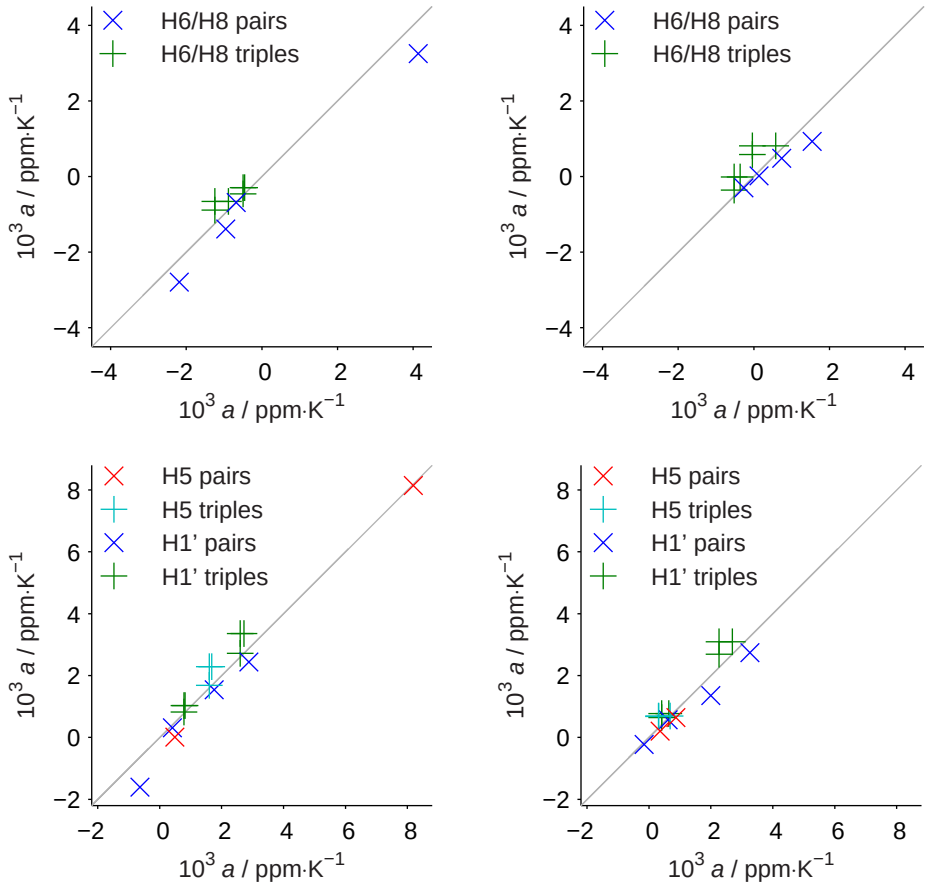


Figure A.18 Correlation plots of chemical-shift slopes of non-exchangeable protons in terminal nucleotides with the same neighbours. Left: **duplex**; right: **single strands**

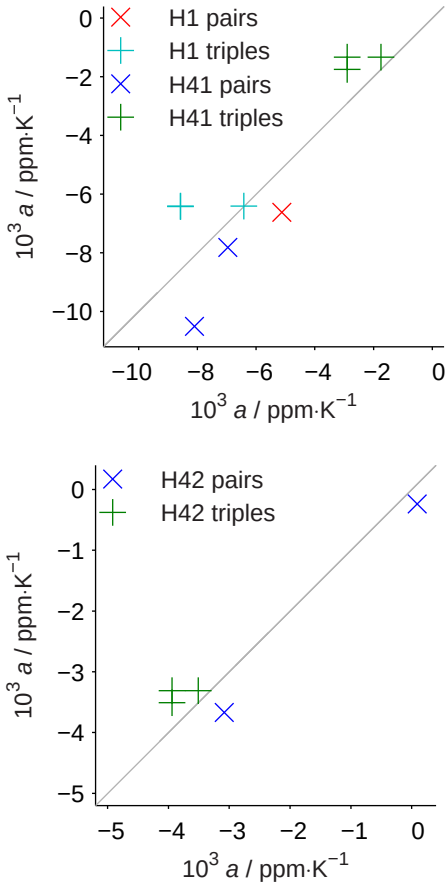


Figure A.19 Correlation plots of chemical-shift slopes of exchangeable protons in terminal nucleotides with the same neighbours in **duplex**

A.4 CTTCGAAG chemical shifts at various concentrations

The following plots of ^{1}H $\delta(T)$ accompany the text of § 7.5.3 on page 89.

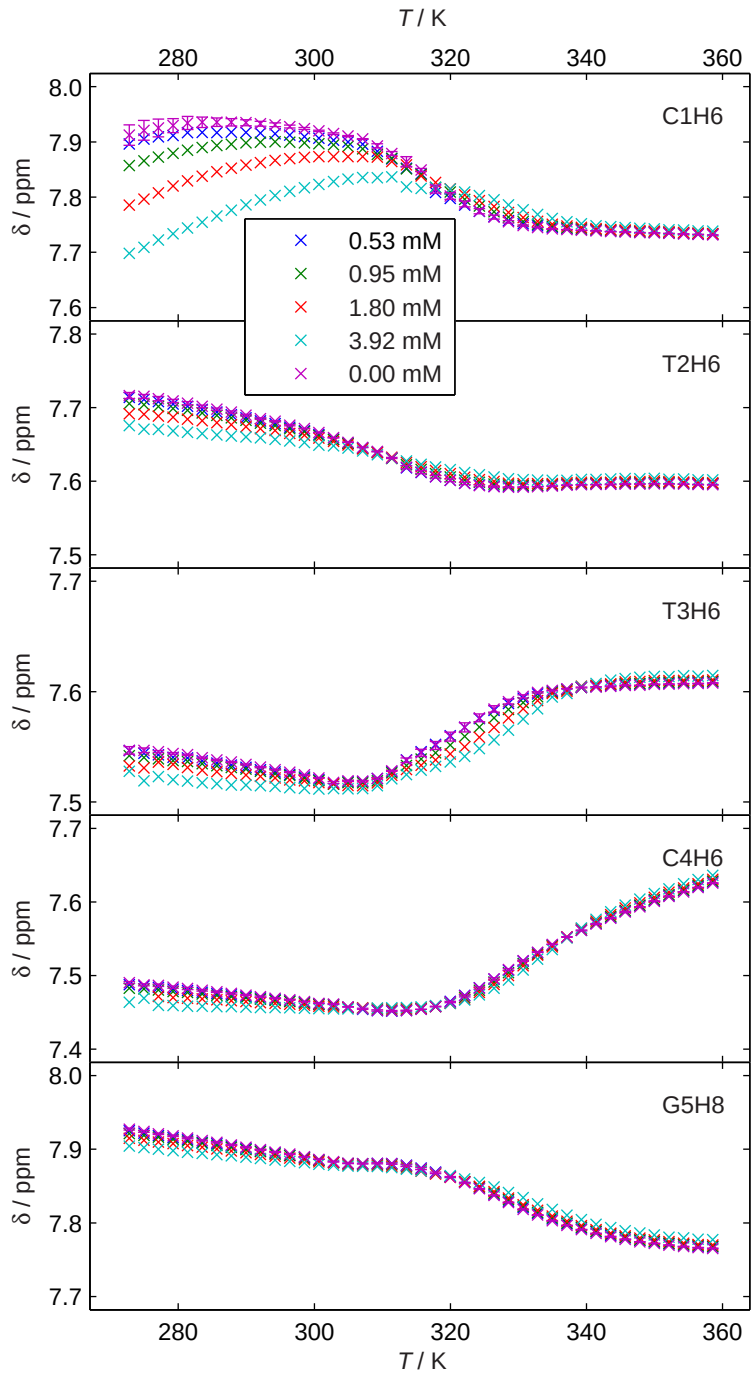


Figure A.20 (to be continued)

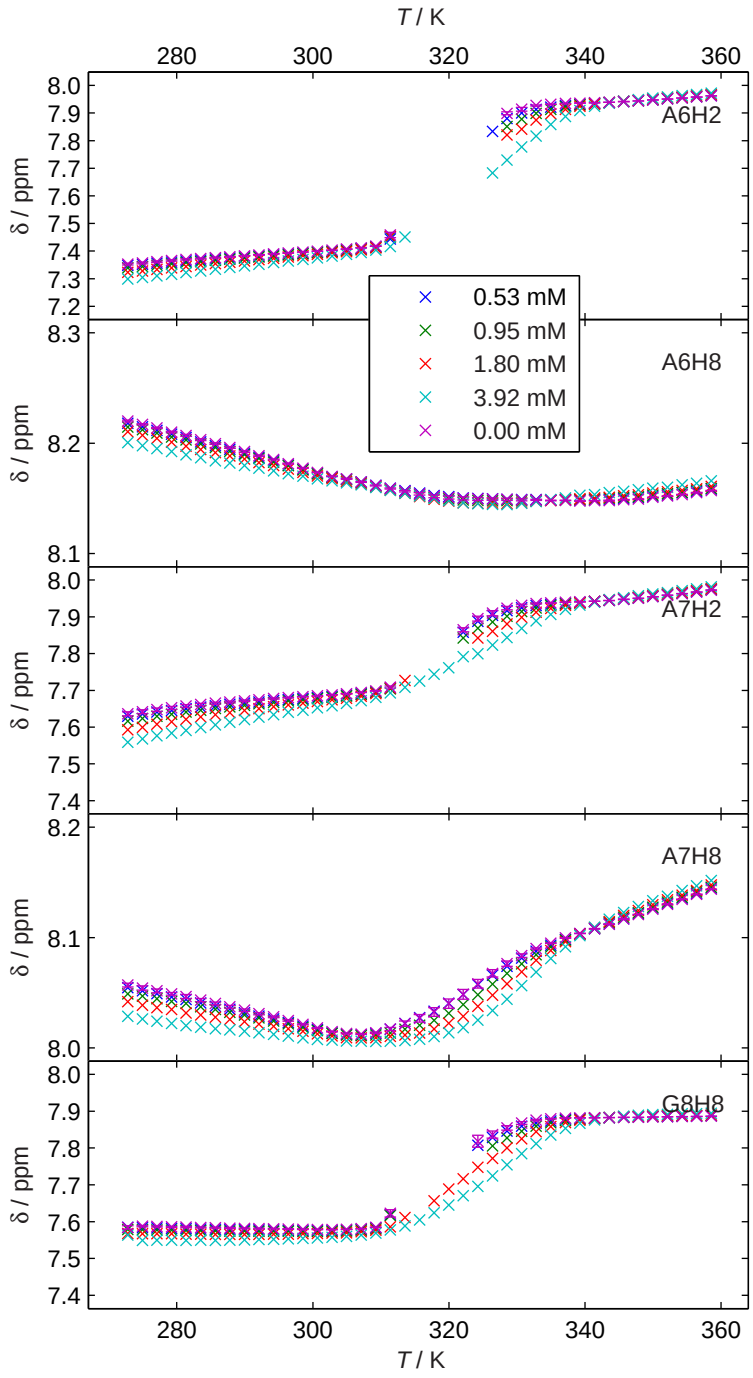


Figure A.20 Chemical shifts of **H6, H8, and AH2** of CTTCGAAG in the whole temperature range measured (crosses) and extrapolations to $c = 0$ (circles with errorbars). Vertical axes are not in the same scales

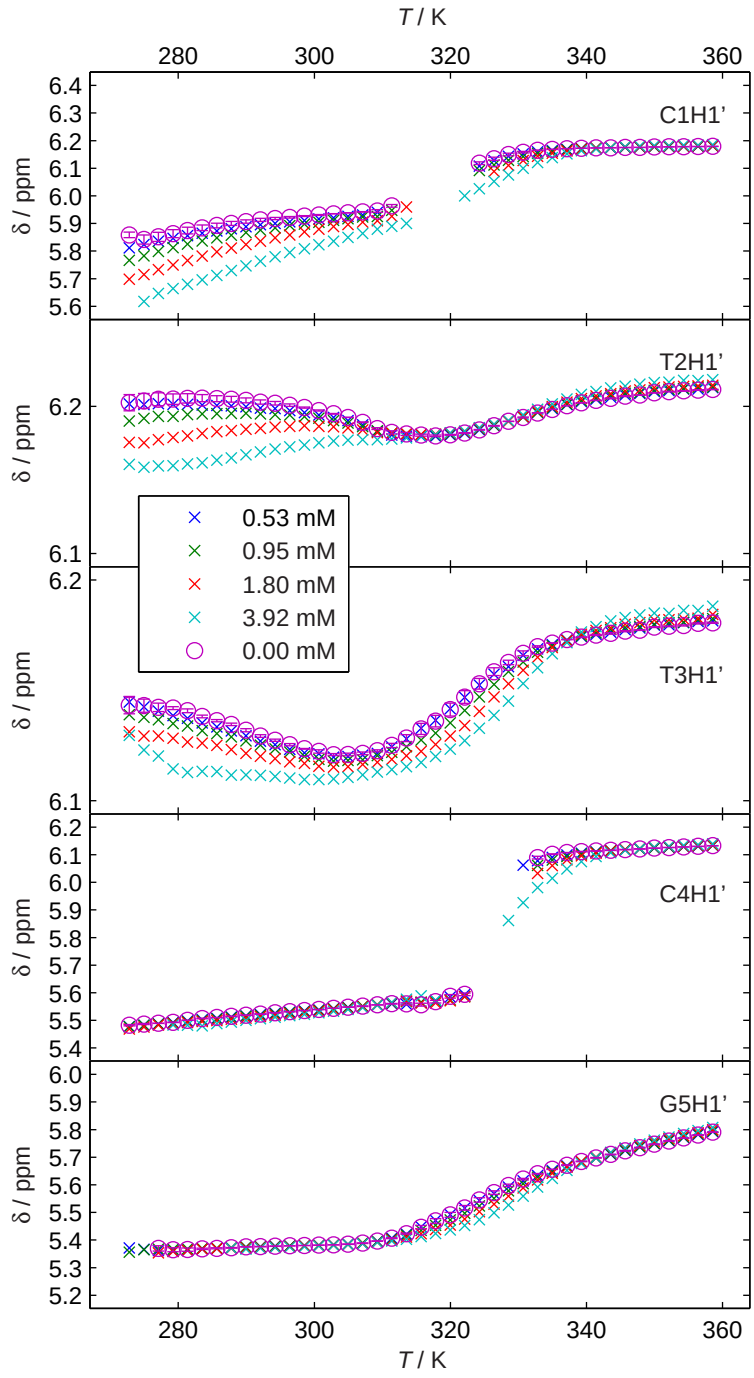


Figure A.21 (to be continued)

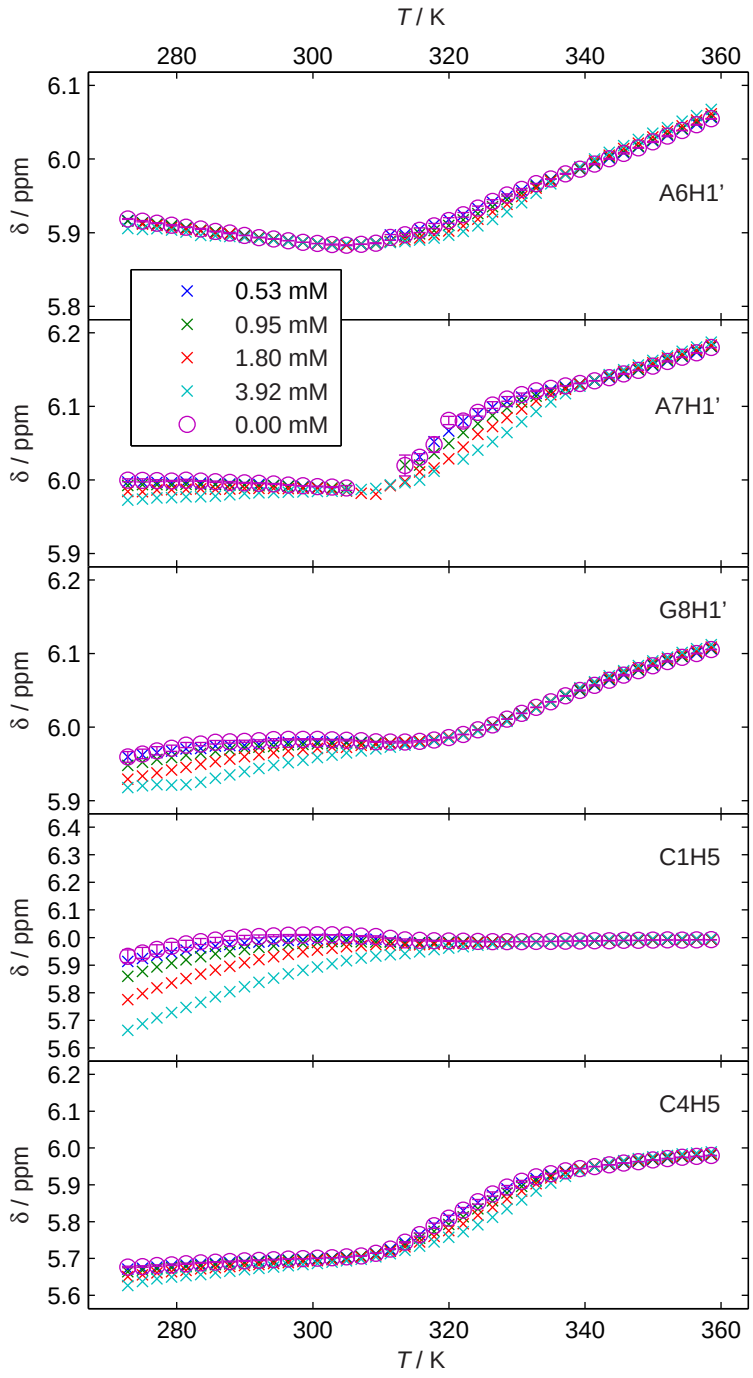


Figure A.21 Chemical shifts of **H1′ and CH5** of CTTCGAAG in the whole temperature range measured (crosses) and extrapolations to $c = 0$ (circles with errorbars). Vertical axes are not in the same scales

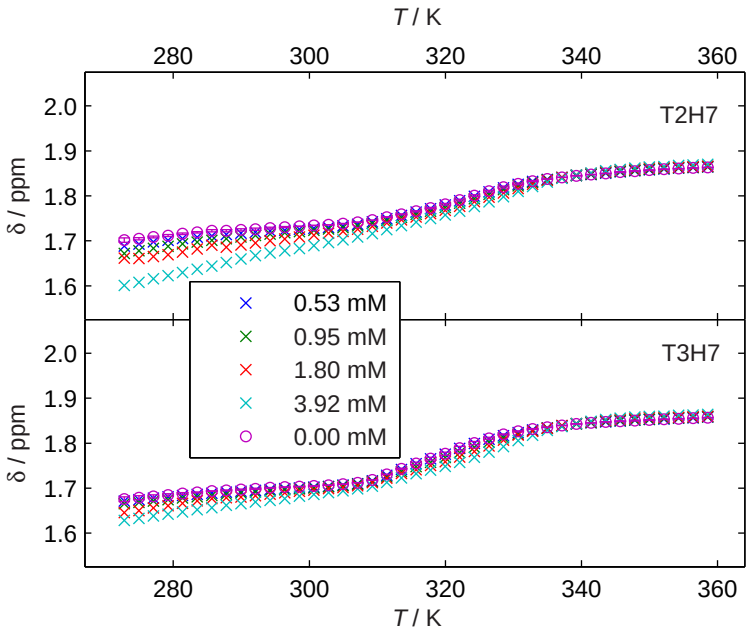


Figure A.22 Chemical shifts of **TH7** of CTTCGAAG in the whole temperature range measured (crosses) and extrapolations to $c = 0$ (circles with errorbars). Vertical axes are in the same scales

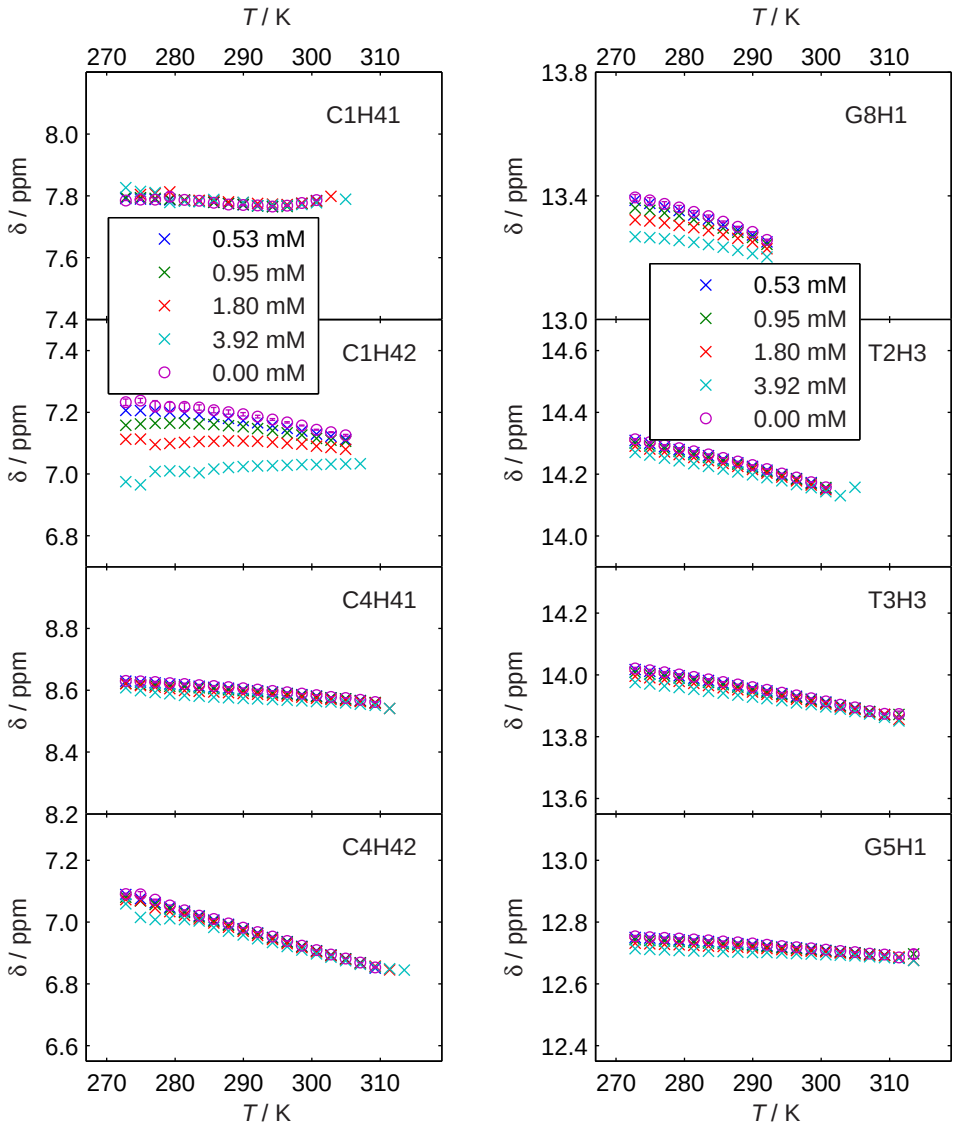


Figure A.23 Chemical shifts of **exchangeable protons** of CTTCGAAG in the whole temperature range measured (crosses) and extrapolations to $c = 0$ (circles with errorbars). Vertical axes are in the same scales. Left: amino, right: imino protons

A.5 Thermodynamic parameters obtained from ^{1}H NMR

This section contains tabulated results of analysis of melting profiles obtained from ^{1}H NMR spectroscopy of self-complementary ODN – without (§ 7.6 on page 96) and with methylation (§ 8.4 on page 133) of cytosine in the central CpG dinucleotide – and SRE-based ODN (§ 9.1.5 on page 150).

A.5.1 Equilibrium – non-methylated ODN

Table A.9 Association parameters of self-complementary duplex DNA obtained from ^{1}H NMR of **H6 and H8**

	$\frac{T_m}{K}$	$\frac{\Delta G_{310\,K}}{kJ \cdot mol^{-1}}$	$\frac{\Delta H}{kJ \cdot mol^{-1}}$	$\frac{\Delta S}{J \cdot mol^{-1} \cdot K^{-1}}$
CAACGTTG				
C1	323.3 ± 0.3	-27.3 ± 0.6	-236 ± 13	-673 ± 39
A2	324.6 ± 0.4	-28.2 ± 0.7	-235 ± 14	-669 ± 42
A3	321.6 ± 0.6	-25.3 ± 0.8	-214 ± 16	-608 ± 49
C4	325.2 ± 0.4	-30.2 ± 0.8	-271 ± 16	-777 ± 49
G5	323.4 ± 0.4	-27.0 ± 0.7	-228 ± 13	-648 ± 41
T6	324.6 ± 0.3	-29.4 ± 0.7	-264 ± 14	-755 ± 44
T7	323.8 ± 0.5	-27.9 ± 0.9	-243 ± 18	-694 ± 55
G8	326.1 ± 0.8	-31.7 ± 2.3	-286 ± 42	-820 ± 130
CATCGATG				
C1	321.6 ± 0.3	-26.0 ± 0.3	-239 ± 9	-688 ± 28
A2	322.8 ± 1.0	-27.1 ± 1.8	-245 ± 34	-700 ± 100
T3	323.2 ± 0.2	-27.6 ± 0.3	-249 ± 7	-713 ± 23
C4	323.6 ± 0.4	-27.2 ± 0.5	-233 ± 12	-664 ± 38
G5	325.1 ± 0.4	-30.3 ± 0.9	-277 ± 20	-795 ± 63
A6	322.9 ± 0.3	-26.8 ± 0.5	-235 ± 11	-672 ± 34
T7	324.1 ± 0.1	-29.2 ± 0.2	-271 ± 6	-781 ± 18
G8	324.1 ± 0.4	-28.9 ± 0.8	-264 ± 18	-760 ± 54

Table A.9 (continued)

	T_m / K	$\Delta G_{310\,K}$ / kJ · mol^{-1}	ΔH / kJ · mol^{-1}	ΔS / J · mol^{-1} · K^{-1}
CTACGTAG				
C1	312.4 ± 0.6	−19.9 ± 0.5	−236 ± 18	−698 ± 56
T2	312.2 ± 0.2	−19.3 ± 0.1	−173 ± 5	−497 ± 15
A3	314.8 ± 0.2	−21.4 ± 0.2	−217 ± 10	−630 ± 32
C4	319.2 ± 0.7	−23.1 ± 0.7	−174 ± 18	−488 ± 57
G5[c]	314.4 ± 0.3	−21.4 ± 0.4	−239 ± 14	−703 ± 44
T6[a]	318.3 ± 0.8	−23.1 ± 0.6	−191 ± 13	−541 ± 42
A7	313.5 ± 0.4	−20.8 ± 0.3	−249 ± 24	−736 ± 78
G8	314.4 ± 0.4	−21.0 ± 0.4	−207 ± 10	−600 ± 31
CTTCGAAG				
C1	317.4 ± 0.6	−23.4 ± 0.8	−246 ± 20	−717 ± 61
T2	315.3 ± 0.8	−22.7 ± 0.8	−305 ± 53	−910 ± 170
T3	317.8 ± 0.4	−22.8 ± 0.3	−212 ± 14	−611 ± 46
C4	324.1 ± 0.8	−30.4 ± 1.8	−292 ± 40	−840 ± 120
G5[cd]	317.1 ± 1.9	−22.0 ± 0.5	−195 ± 48	−560 ± 150
A6[ab]	280 ± 56	−13.0 ± 6.0	−43 ± 36	−98 ± 97
A7	315.0 ± 0.4	−20.9 ± 0.4	−202 ± 14	−586 ± 43
G8	319.8 ± 0.5	−26.5 ± 1.0	−291 ± 21	−854 ± 64
CTTGCAAG				
C1	319.0 ± 0.7	−24.5 ± 1.0	−229 ± 25	−659 ± 76
T2	323.1 ± 2.4	−32.0 ± 7.6	−340 ± 140	−1010 ± 430
T3	319.5 ± 0.4	−24.7 ± 0.5	−222 ± 14	−637 ± 44
G4	317.9 ± 0.5	−23.2 ± 0.4	−206 ± 16	−591 ± 52
C5	321.2 ± 0.5	−25.9 ± 0.8	−225 ± 21	−641 ± 64
A6	322.7 ± 0.5	−28.7 ± 0.7	−270 ± 21	−778 ± 64
A7	319.4 ± 0.7	−24.1 ± 1.0	−206 ± 23	−585 ± 70
G8	321.9 ± 0.5	−28.1 ± 1.4	−272 ± 30	−787 ± 92

Table A.9 (continued)

	$\frac{T_m}{K}$	$\frac{\Delta G_{310\,K}}{kJ \cdot mol^{-1}}$	$\frac{\Delta H}{kJ \cdot mol^{-1}}$	$\frac{\Delta S}{J \cdot mol^{-1} \cdot K^{-1}}$
GAACGTTC				
G1	318.0 ± 0.5	-24.2 ± 0.8	-247 ± 20	-720 ± 63
A2	317.8 ± 0.6	-23.1 ± 0.8	-210 ± 19	-602 ± 60
A3	320.6 ± 0.5	-26.4 ± 1.1	-255 ± 28	-738 ± 85
C4	322.0 ± 0.3	-26.8 ± 0.5	-237 ± 15	-678 ± 47
G5	321.5 ± 0.5	-27.4 ± 1.3	-264 ± 29	-764 ± 90
T6	320.6 ± 0.3	-26.0 ± 0.6	-243 ± 17	-700 ± 52
T7	322.3 ± 0.6	-28.5 ± 1.3	-276 ± 35	-800 ± 110
C8	323.3 ± 1.4	-27.6 ± 0.4	-236 ± 15	-673 ± 50
GATGCATC				
G1	319.7 ± 0.5	-25.4 ± 0.7	-238 ± 18	-687 ± 57
A2[a]	313.2 ± 5.1	-20.2 ± 4.0	-198 ± 86	-570 ± 270
T3	321.4 ± 0.3	-27.1 ± 0.5	-252 ± 13	-726 ± 39
G4	320.7 ± 0.4	-26.2 ± 0.5	-241 ± 13	-692 ± 42
C5	322.1 ± 0.4	-28.3 ± 0.7	-268 ± 17	-774 ± 54
A6	321.6 ± 0.6	-27.5 ± 1.1	-257 ± 23	-741 ± 71
T7	321.9 ± 0.4	-26.6 ± 0.8	-232 ± 18	-663 ± 57
C8	322.8 ± 0.5	-27.7 ± 1.1	-240 ± 23	-685 ± 72
GTAGCTAC				
G1	314.2 ± 0.4	-21.1 ± 0.2	-236 ± 15	-694 ± 50
T2	317.2 ± 0.4	-23.5 ± 0.5	-245 ± 16	-714 ± 51
A3	320.4 ± 0.7	-26.9 ± 0.8	-274 ± 27	-796 ± 86
G4	319.9 ± 0.1	-25.7 ± 0.2	-251 ± 6	-728 ± 19
C5	320.8 ± 0.4	-26.5 ± 0.9	-254 ± 21	-733 ± 64
T6	319.6 ± 1.6	-25.8 ± 1.7	-264 ± 57	-770 ± 180
A7[a]	322.4 ± 1.0	-29.5 ± 2.3	-299 ± 51	-870 ± 160
C8	319.7 ± 0.4	-24.2 ± 0.7	-207 ± 17	-589 ± 52

[a]Fitting of the exchange line shape not possible, the results are from chemical shifts only. [b]The two-state model fails, peak positions in the melting region outside the interval defined by extrapolations from duplex and single strands. [c]Parabolic extrapolation of single-strand chemical shifts. [d]Anomalous behaviour violating the two-state model, discussed in § 7.7 on page 109

Table A.10 Association parameters of self-complementary duplex DNA obtained from ^{1}H NMR of **AH2, CH5 and TH7**

	$\frac{T_m}{K}$	$\frac{\Delta G_{310\,K}}{kJ \cdot mol^{-1}}$	$\frac{\Delta H}{kJ \cdot mol^{-1}}$	$\frac{\Delta S}{J \cdot mol^{-1} \cdot K^{-1}}$
CAACGTTG				
C1[ab]	314.5 ± 1.3	-20.5 ± 1.1	-204 ± 27	-594 ± 84
A2	325.1 ± 1.1	-29.9 ± 3.0	-265 ± 47	-760 ± 140
A3	323.7 ± 0.5	-27.4 ± 0.8	-232 ± 16	-660 ± 50
C4[c]				
T6	324.9 ± 0.2	-29.2 ± 0.3	-255 ± 9	-727 ± 27
T7	324.8 ± 0.4	-28.3 ± 0.5	-237 ± 12	-672 ± 36
CATCGATG				
C1[ab]	315.9 ± 0.9	-22.2 ± 1.2	-257 ± 30	-756 ± 92
A2	324.0 ± 0.3	-27.8 ± 0.6	-241 ± 13	-689 ± 39
T3	324.6 ± 0.1	-29.5 ± 0.2	-270 ± 4	-774 ± 13
C4[a]	324.7 ± 0.2	-30.2 ± 0.6	-282 ± 10	-814 ± 30
A6	323.1 ± 0.3	-27.4 ± 0.7	-247 ± 15	-708 ± 47
T7	324.1 ± 0.1	-28.7 ± 0.2	-260 ± 4	-747 ± 12
CTACGTAG				
C1[ab]	305.6 ± 1.1	-13.6 ± 1.0	-311 ± 20	-959 ± 64
T2[a]	321.9 ± 1.2	-24.9 ± 1.4	-186 ± 26	-520 ± 79
A3	317.1 ± 1.8	-23.3 ± 2.5	-231 ± 58	-670 ± 180
C4[a]	318.8 ± 0.5	-23.8 ± 0.6	-208 ± 12	-595 ± 35
T6	321.4 ± 1.0	-26.3 ± 1.0	-231 ± 28	-660 ± 86
A7	314.0 ± 1.0	-20.5 ± 0.8	-194 ± 15	-560 ± 45
CTTCGAAG				
C1[ab]	307.8 ± 1.0	-15.1 ± 1.1	-346 ± 8	-1069 ± 25
T2	324.1 ± 0.8	-26.5 ± 0.7	-202 ± 19	-568 ± 61
T3	320.0 ± 0.3	-22.8 ± 0.2	-166 ± 7	-463 ± 21
C4[a]	321.0 ± 0.4	-25.9 ± 0.7	-240 ± 12	-690 ± 38
A6	321.0 ± 0.2	-29.2 ± 0.5	-339 ± 12	-999 ± 36
A7	318.0 ± 1.5	-25.0 ± 3.6	-292 ± 90	-860 ± 280

Table A.10 (continued)

	T_m / K	$\Delta G_{310\,K}$ / kJ · mol^{-1}	ΔH / kJ · mol^{-1}	ΔS / J · mol^{-1} · K^{-1}
CTTGCAAG				
C1[ab]	311.6 ± 1.0	−19.8 ± 1.2	−330 ± 51	−1000 ± 160
T2	321.8 ± 0.5	−26.4 ± 0.5	−227 ± 13	−647 ± 42
T3	322.6 ± 0.6	−27.5 ± 1.0	−240 ± 24	−687 ± 73
C5[a]	323.3 ± 0.2	−29.3 ± 0.6	−271 ± 11	−780 ± 33
A6	323.5 ± 0.9	−30.4 ± 2.3	−295 ± 37	−850 ± 110
A7	321.1 ± 0.6	−26.0 ± 1.1	−229 ± 27	−655 ± 83
GAACGTTC				
A2	319.5 ± 1.9	−24.5 ± 3.6	−220 ± 61	−630 ± 180
A3	323.4 ± 1.2	−30.7 ± 4.0	−308 ± 86	−890 ± 260
C4[c]				
T6	321.8 ± 0.1	−27.6 ± 0.2	−255 ± 5	−734 ± 17
T7	323.1 ± 0.5	−28.7 ± 0.9	−258 ± 18	−738 ± 56
C8[ab]	325.9 ± 1.3	−30.8 ± 2.1	−263 ± 28	−750 ± 84
GATGCATC				
A2	325.0 ± 0.9	−34.0 ± 3.4	−344 ± 64	−1000 ± 200
T3	321.4 ± 0.2	−26.7 ± 0.4	−248 ± 11	−714 ± 34
C5[a]	322.8 ± 0.3	−29.4 ± 0.7	−284 ± 12	−820 ± 36
A6	321.9 ± 0.4	−27.7 ± 0.9	−257 ± 18	−741 ± 55
T7	321.5 ± 0.3	−26.2 ± 0.4	−233 ± 10	−666 ± 32
C8[ab]	324.4 ± 0.7	−30.9 ± 1.6	−287 ± 27	−825 ± 83
GTAGCTAC				
T2	320.7 ± 0.2	−25.9 ± 0.2	−240 ± 8	−691 ± 25
A3	320.1 ± 1.1	−26.9 ± 2.1	−283 ± 40	−830 ± 120
C5[a]	322.6 ± 0.4	−30.2 ± 1.0	−314 ± 17	−916 ± 53
T6	320.8 ± 0.2	−26.0 ± 0.2	−238 ± 7	−683 ± 22
A7	318.1 ± 0.9	−21.7 ± 0.6	−147 ± 13	−405 ± 41
C8[ab]	322.5 ± 0.4	−30.3 ± 1.0	−320 ± 18	−933 ± 55

[a] Fitting of the exchange line shape not possible, the results are from chemical shifts only.
[b] Serious deviation from chemical-shift linearity in duplex makes the analysis unreliable.
[c] Values in duplex missing due to spectral overlap with residual water signal

Table A.11 Association parameters of self-complementary duplex DNA obtained from ^{1}H NMR of **H1′**. All values obtained from fitting of chemical shifts (no exchange line shape)

	$\frac{T_m}{K}$	$\frac{\Delta G_{310\,K}}{kJ \cdot mol^{-1}}$	$\frac{\Delta H}{kJ \cdot mol^{-1}}$	$\frac{\Delta S}{J \cdot mol^{-1} \cdot K^{-1}}$
CAACGTTG				
C1	325.1 ± 0.3	-29.9 ± 0.7	-267 ± 13	-765 ± 39
A2	326.2 ± 0.3	-31.4 ± 0.8	-279 ± 15	-800 ± 46
A3	325.5 ± 1.2	-28.7 ± 1.2	-234 ± 21	-661 ± 65
C4	325.8 ± 0.2	-33.5 ± 0.9	-328 ± 15	-952 ± 45
G5	325.1 ± 0.7	-28.2 ± 0.9	-230 ± 18	-652 ± 55
T6	323.7 ± 0.6	-26.6 ± 0.7	-213 ± 15	-602 ± 47
T7	323.8 ± 0.5	-26.2 ± 0.5	-204 ± 10	-574 ± 30
G8	327.8 ± 1.9	-34.3 ± 5.2	-308 ± 91	-880 ± 280
CATCGATG				
C1	324.2 ± 0.2	-29.1 ± 0.5	-268 ± 9	-770 ± 29
A2	325.9 ± 1.2	-31.3 ± 2.9	-284 ± 54	-820 ± 170
T3	324.1 ± 0.3	-29.4 ± 0.5	-276 ± 11	-795 ± 32
C4	325.1 ± 0.2	-30.4 ± 0.6	-281 ± 9	-808 ± 29
G5	325.4 ± 0.3	-29.8 ± 0.6	-262 ± 12	-751 ± 37
A6	323.5 ± 0.7	-26.8 ± 0.9	-225 ± 18	-638 ± 54
T7	324.1 ± 0.3	-29.1 ± 0.7	-268 ± 12	-772 ± 37
G8	326.3 ± 1.5	-32.4 ± 4.4	-299 ± 78	-860 ± 240
CTACGTAG				
C1	316.9 ± 0.5	-22.9 ± 0.6	-224 ± 12	-648 ± 36
T2	318.7 ± 0.4	-25.9 ± 0.6	-287 ± 13	-843 ± 40
A3	311.1 ± 0.9	-19.4 ± 1.0	-380 ± 4	-1164 ± 11
C4	322.5 ± 0.2	-31.7 ± 0.8	-350 ± 16	-1028 ± 50
G5	314.2 ± 0.8	-20.8 ± 0.7	-206 ± 20	-597 ± 62
T6	316.7 ± 0.4	-22.8 ± 0.5	-225 ± 10	-653 ± 31
A7	314.9 ± 0.4	-25.5 ± 0.7	-482 ± 2	-1473 ± 5
G8	308.1 ± 1.6	-16.5 ± 1.1	-253 ± 49	-760 ± 160

Table A.11 (continued)

	$\frac{T_m}{K}$	$\frac{\Delta G_{310\,K}}{kJ \cdot mol^{-1}}$	$\frac{\Delta H}{kJ \cdot mol^{-1}}$	$\frac{\Delta S}{J \cdot mol^{-1} \cdot K^{-1}}$
CTTCGAAG				
C1	321.4 ± 0.7	-28.5 ± 1.7	-304 ± 31	-889 ± 94
T2	309.3 ± 1.3	-16.8 ± 1.6	-377 ± 6	-1162 ± 18
T3	318.0 ± 0.9	-22.4 ± 0.8	-189 ± 20	-537 ± 61
C4	325.5 ± 0.4	-41.9 ± 2.4	-509 ± 39	-1510 ± 120
G5	320.1 ± 0.4	-24.5 ± 0.5	-217 ± 11	-622 ± 34
A6[a]	305.2 ± 3.2	-15.4 ± 1.6	-141 ± 17	-404 ± 57
A7	316.5 ± 0.9	-25.1 ± 1.6	-366 ± 35	-1100 ± 110
G8	313.7 ± 1.4	-21.2 ± 1.9	-300 ± 61	-900 ± 190
CTTGCAAG				
C1	322.9 ± 0.3	-28.3 ± 0.6	-256 ± 11	-734 ± 34
T2	322.0 ± 1.1	-27.1 ± 1.9	-241 ± 40	-690 ± 120
T3	321.6 ± 0.2	-27.1 ± 0.5	-248 ± 10	-714 ± 30
G4	321.6 ± 0.3	-25.8 ± 0.6	-214 ± 11	-607 ± 35
C5	324.0 ± 0.3	-31.4 ± 0.8	-308 ± 13	-891 ± 40
A6	320.9 ± 0.3	-26.3 ± 0.6	-242 ± 12	-694 ± 38
A7	322.1 ± 0.7	-27.1 ± 0.8	-241 ± 18	-689 ± 55
G8	323.1 ± 0.7	-28.0 ± 1.2	-245 ± 25	-700 ± 78
GAACGTTC				
G1	320.2 ± 0.3	-25.7 ± 0.4	-245 ± 8	-707 ± 24
A2	322.9 ± 0.4	-31.3 ± 1.3	-334 ± 26	-975 ± 80
A3	322.5 ± 1.0	-28.6 ± 1.9	-274 ± 40	-790 ± 120
C4	324.3 ± 0.4	-33.6 ± 1.5	-356 ± 23	-1039 ± 71
G5	322.2 ± 0.6	-27.7 ± 1.1	-257 ± 22	-741 ± 69
T6	320.8 ± 0.5	-26.3 ± 0.9	-247 ± 18	-713 ± 57
T7[a]	317.1 ± 7.0	-20.8 ± 2.6	-125 ± 27	-336 ± 88
C8	312.5 ± 1.6	-19.1 ± 0.9	-147 ± 22	-414 ± 69

Table A.11 (continued)

	$\frac{T_m}{K}$	$\frac{\Delta G_{310\,K}}{kJ \cdot mol^{-1}}$	$\frac{\Delta H}{kJ \cdot mol^{-1}}$	$\frac{\Delta S}{J \cdot mol^{-1} \cdot K^{-1}}$
GATGCATC				
G1	320.2 ± 0.2	-25.5 ± 0.3	-230 ± 7	-660 ± 21
A2	312.0 ± 1.5	-19.5 ± 1.1	-200 ± 28	-583 ± 88
T3	321.7 ± 0.3	-28.6 ± 0.7	-286 ± 14	-830 ± 44
G4	323.7 ± 0.4	-28.1 ± 0.7	-234 ± 14	-664 ± 45
C5	323.8 ± 0.4	-32.0 ± 1.1	-325 ± 18	-946 ± 56
A6	321.9 ± 0.8	-26.4 ± 1.2	-223 ± 24	-635 ± 75
T7	322.6 ± 0.6	-27.5 ± 1.1	-239 ± 21	-681 ± 66
C8	313.3 ± 1.7	-19.6 ± 0.9	-138 ± 21	-381 ± 65
GTAGCTAC				
G1	322.0 ± 0.4	-28.0 ± 0.9	-269 ± 19	-779 ± 60
T2[b]	322.2 ± 1.2	-30.7 ± 3.5	-338 ± 63	-990 ± 190
A3	313.9 ± 1.6	-21.1 ± 1.7	-258 ± 45	-760 ± 140
G4[b]	320.6 ± 0.4	-25.6 ± 0.6	-231 ± 13	-663 ± 39
C5	321.5 ± 0.3	-28.1 ± 0.6	-284 ± 12	-826 ± 37
T6[b]	320.9 ± 0.3	-27.1 ± 0.6	-269 ± 14	-780 ± 44
A7	322.0 ± 0.7	-27.6 ± 1.2	-259 ± 27	-748 ± 84
C8	321.3 ± 0.5	-26.9 ± 0.9	-254 ± 20	-732 ± 61

[a]The two-state model fails, peak positions in the melting region outside the interval defined by extrapolations from duplex and single strands. [b]Ambiguous assignment of resonances in single strands (see Table A.3)

A.5.2 Equilibrium – methylated ODN

Table A.12 Association parameters of self-complementary duplex DNA with methylated cytosine obtained from ^{1}H NMR of **H6 and H8**

	$\frac{T_m}{K}$	$\frac{\Delta G_{310\,K}}{kJ \cdot mol^{-1}}$	$\frac{\Delta H}{kJ \cdot mol^{-1}}$	$\frac{\Delta S}{J \cdot mol^{-1} \cdot K^{-1}}$
CAAm^{5}CGTTG				
C1	323.4 ± 0.3	−28.8 ± 0.6	−267 ± 10	−769 ± 31
A2	325.4 ± 0.5	−28.7 ± 0.9	−230 ± 17	−648 ± 53
A3	320.0 ± 0.7	−24.0 ± 0.8	−199 ± 12	−563 ± 37
m^{5}C4	325.2 ± 0.3	−29.2 ± 0.5	−245 ± 11	−697 ± 33
G5	324.9 ± 0.2	−28.6 ± 0.3	−237 ± 7	−674 ± 21
T6	325.1 ± 0.2	−29.3 ± 0.4	−248 ± 8	−707 ± 23
T7	324.7 ± 0.4	−28.7 ± 0.8	−240 ± 15	−680 ± 45
G8	327.2 ± 0.4	−32.4 ± 1.3	−279 ± 23	−795 ± 70
CATm^{5}CGATG				
C1	324.3 ± 0.3	−29.7 ± 0.6	−276 ± 12	−795 ± 36
A2	324.7 ± 0.6	−29.0 ± 1.2	−253 ± 22	−722 ± 66
T3	325.5 ± 0.2	−30.2 ± 0.5	−267 ± 9	−763 ± 28
m^{5}C4	325.3 ± 0.4	−28.8 ± 0.7	−241 ± 13	−686 ± 41
G5	326.8 ± 0.4	−32.8 ± 0.8	−297 ± 17	−853 ± 52
A6	324.3 ± 0.3	−27.9 ± 0.6	−236 ± 11	−673 ± 34
T7	325.4 ± 0.2	−30.2 ± 0.4	−268 ± 7	−768 ± 21
G8	325.4 ± 0.4	−29.9 ± 0.8	−263 ± 16	−751 ± 50
CTTm^{5}CGAAG				
C1	320.0 ± 0.4	−27.3 ± 0.9	−299 ± 23	−876 ± 72
T2	316.6 ± 0.6	−24.5 ± 0.8	−308 ± 48	−920 ± 150
T3	317.8 ± 0.5	−23.8 ± 0.5	−237 ± 18	−686 ± 58
m^{5}C4^e	319.5 ± 3.8	−22.8 ± 3.8	−160 ± 70	−440 ± 210
G5^{cd}	320.0 ± 0.5	−25.3 ± 0.5	−231 ± 21	−663 ± 65
A6^{ab}	294 ± 42	−15.3 ± 5.6	−50 ± 39	−110 ± 110
A7	315.7 ± 0.4	−22.0 ± 0.5	−223 ± 17	−647 ± 54
G8	322.6 ± 0.3	−30.4 ± 0.9	−319 ± 16	−931 ± 49

[a]Fitting of the exchange line shape not possible, the results are from chemical shifts only. [b]The two-state model fails, peak positions in the melting region outside the interval defined by extrapolations from duplex and single strands. [c]Parabolic extrapolation of single-strand chemical shifts. [d]Anomalous behaviour violating the two-state model, discussed in § 8.4.2 on page 137. [e]Fitting of the exchange line shape not possible for several temperatures near duplex which affects the precision

Table A.13 Association parameters of self-complementary duplex DNA with methylated cytosine obtained from ^{1}H NMR of **AH2, CH5, m^{5}CH7, and TH7**

	$\frac{T_m}{K}$	$\frac{\Delta G_{310\,K}}{kJ \cdot mol^{-1}}$	$\frac{\Delta H}{kJ \cdot mol^{-1}}$	$\frac{\Delta S}{J \cdot mol^{-1} \cdot K^{-1}}$
CAAm^{5}CGTTG				
C1^{ab}	315.7 ± 1.1	-21.9 ± 1.2	-225 ± 26	-655 ± 82
A2	326.2 ± 1.0	-31.5 ± 2.5	-276 ± 37	-790 ± 110
A3	324.4 ± 0.3	-27.7 ± 0.6	-224 ± 11	-634 ± 34
m^{5}C4	325.4 ± 0.1	-29.0 ± 0.2	-238 ± 5	-675 ± 15
T6	325.7 ± 0.1	-29.9 ± 0.2	-253 ± 5	-718 ± 14
T7	325.7 ± 0.2	-29.2 ± 0.3	-237 ± 6	-669 ± 18
CATm^{5}CGATG				
C1^{ab}	312.8 ± 1.2	-19.0 ± 0.8	-172 ± 19	-495 ± 61
A2	325.9 ± 0.2	-30.5 ± 0.6	-266 ± 10	-761 ± 29
T3	326.3 ± 0.1	-30.3 ± 0.4	-257 ± 7	-730 ± 22
m^{5}C4	325.8 ± 0.2	-30.3 ± 0.3	-264 ± 7	-754 ± 21
A6	324.1 ± 0.4	-27.7 ± 0.6	-233 ± 12	-663 ± 38
T7	326.0 ± 0.1	-30.5 ± 0.3	-265 ± 6	-755 ± 18
CTTm^{5}CGAAG				
C1^{ab}	308.7 ± 1.1	-16.6 ± 1.2	-344 ± 15	-1056 ± 46
T2	326.8 ± 1.2	-29.4 ± 1.4	-220 ± 34	-620 ± 100
T3	322.6 ± 0.6	-24.9 ± 0.5	-177 ± 17	-490 ± 54
m^{5}C4	321.9 ± 0.4	-24.7 ± 0.2	-180 ± 10	-500 ± 32
A6	323.7 ± 0.5	-34.6 ± 0.8	-393 ± 17	-1154 ± 52
A7	318.0 ± 0.7	-23.4 ± 0.8	-215 ± 15	-617 ± 47

[a] Fitting of the exchange line shape not possible, the results are from chemical shifts only.
[b] Serious deviation from chemical-shift linearity in duplex makes the analysis unreliable.

Table A.14 Association parameters of self-complementary duplex DNA with methylated cytosine obtained from ^{1}H NMR of **H1′**. All values obtained from fitting of chemical shifts (no exchange line shape)

	$\frac{T_m}{K}$	$\frac{\Delta G_{310\,K}}{kJ \cdot mol^{-1}}$	$\frac{\Delta H}{kJ \cdot mol^{-1}}$	$\frac{\Delta S}{J \cdot mol^{-1} \cdot K^{-1}}$
CAAm^{5}CGTTG				
C1	325.6 ± 0.2	-30.3 ± 0.6	-261 ± 10	-746 ± 31
A2	327.9 ± 0.3	-33.9 ± 0.7	-295 ± 13	-843 ± 38
A3	324.9 ± 2.0	-28.7 ± 2.7	-238 ± 39	-680 ± 120
m^{5}C4	327.1 ± 0.2	-33.6 ± 0.7	-303 ± 11	-869 ± 33
G5	326.2 ± 0.5	-29.3 ± 0.7	-232 ± 14	-653 ± 42
T6	323.5 ± 0.8	-25.9 ± 0.8	-194 ± 16	-543 ± 49
T7	323.9 ± 0.4	-26.8 ± 0.6	-209 ± 12	-588 ± 37
G8	328.3 ± 1.4	-32.8 ± 3.7	-269 ± 59	-760 ± 180
CATm^{5}CGATG				
C1	325.5 ± 0.2	-30.4 ± 0.6	-270 ± 11	-774 ± 35
A2	328.0 ± 1.2	-33.9 ± 3.2	-299 ± 55	-850 ± 170
T3	325.6 ± 0.5	-28.2 ± 0.6	-223 ± 11	-627 ± 35
m^{5}C4	326.7 ± 0.2	-33.6 ± 0.8	-315 ± 13	-907 ± 39
G5	327.2 ± 0.3	-31.3 ± 0.6	-262 ± 11	-743 ± 34
A6	325.2 ± 0.8	-27.9 ± 1.1	-222 ± 21	-628 ± 64
T7	325.6 ± 0.3	-29.5 ± 0.7	-250 ± 12	-711 ± 36
G8	327.8 ± 1.4	-33.1 ± 3.8	-288 ± 64	-820 ± 200
CTTm^{5}CGAAG				
C1	322.4 ± 0.8	-27.6 ± 1.6	-249 ± 27	-713 ± 83
T2	315.2 ± 0.9	-24.8 ± 1.6	-409 ± 55	-1240 ± 170
T3	319.8 ± 7.6	-21.8 ± 2.7	-122 ± 30	-324 ± 97
m^{5}C4	328.5 ± 0.2	-53.3 ± 0.4	-628 ± 4	-1855 ± 11
G5	321.1 ± 0.5	-25.9 ± 0.8	-226 ± 15	-646 ± 47
A6[a]	308.5 ± 2.9	-17.1 ± 1.8	-193 ± 34	-570 ± 110
A7	320.3 ± 1.0	-32.2 ± 1.9	-442 ± 26	-1321 ± 80
G8	315.1 ± 1.1	-22.3 ± 1.4	-262 ± 40	-770 ± 130

[a]The two-state model fails, peak positions in the melting region outside the interval defined by extrapolations from duplex and single strands

A.5.3 Kinetics – non-methylated ODN

Table A.15 Activation parameters of dissociation of self-complementary duplex DNA obtained from ^{1}H NMR of **H6 and H8**

	$\Delta G^\dagger_{310\,\mathrm{K}}$ / kJ · mol^{-1}	$\Delta H^\dagger$ / kJ · mol^{-1}	$\Delta S^\dagger$ / J · mol^{-1} · K^{-1}
CAACGTTG			
C1	66.3 ± 3.2	214 ± 94	478 ± 293
A2	66.6 ± 1.4	224 ± 46	506 ± 144
A3	65.9 ± 1.2	138 ± 36	233 ± 114
C4	65.9 ± 0.8	200 ± 18	432 ± 55
G5	67.2 ± 0.6	263 ± 26	633 ± 81
T6	66.4 ± 0.7	242 ± 33	568 ± 106
T7	66.9 ± 2.2	395 ± 70	1060 ± 221
G8[b]	58 ± 32	920 ± 860	2800 ± 2700
CATCGATG			
C1	66.3 ± 0.5	249 ± 23	591 ± 73
A2	65.3 ± 0.4	280 ± 19	692 ± 60
T3	65.9 ± 0.2	235 ± 10	547 ± 32
C4	66.7 ± 0.2	233 ± 13	537 ± 43
G5	68.4 ± 0.8	300 ± 40	750 ± 130
A6	66.4 ± 0.3	240 ± 15	559 ± 46
T7	65.7 ± 0.2	237 ± 7	554 ± 23
G8	66.4 ± 0.6	245 ± 28	576 ± 89
CTACGTAG			
C1	62.7 ± 1.2	148 ± 60	280 ± 190
T2	62.6 ± 0.1	191 ± 7	415 ± 24
A3	63.9 ± 0.3	259 ± 28	628 ± 89
C4	63.7 ± 0.2	268 ± 28	657 ± 89
G5[c]	64.0 ± 0.3	180 ± 14	374 ± 44
T6[a]			
A7	62.4 ± 0.3	212 ± 32	480 ± 100
G8	62.9 ± 0.1	217 ± 4	497 ± 13

Table A.15 (continued)

	$\Delta G^{\dagger}_{310\,\mathrm{K}}$ / kJ · mol^{-1}	$\Delta H^{\dagger}$ / kJ · mol^{-1}	$\Delta S^{\dagger}$ / J · mol^{-1} · K^{-1}
CTTCGAAG			
C1	64.9 ± 0.8	230 ± 36	530 ± 110
T2	65.9 ± 1.4	220 ± 140	490 ± 440
T3	66.3 ± 0.4	289 ± 28	719 ± 88
C4[b]	66 ± 36	400 ± 1500	1000 ± 5000
G5[cd]	65.6 ± 0.4	213 ± 40	470 ± 130
A6[a]			
A7	67.6 ± 0.3	260 ± 20	622 ± 64
G8	64.9 ± 0.3	194 ± 10	418 ± 32
CTTGCAAG			
C1	66.0 ± 1.6	276 ± 41	680 ± 130
T2[b]	66 ± 48	300 ± 1000	700 ± 3100
T3	65.5 ± 0.5	269 ± 20	656 ± 63
G4	66.0 ± 0.5	239 ± 21	559 ± 67
C5	65.8 ± 0.7	248 ± 21	587 ± 67
A6	65.1 ± 1.3	270 ± 53	660 ± 170
A7	66.3 ± 0.8	247 ± 24	584 ± 77
G8	64.7 ± 1.9	191 ± 42	410 ± 130
GAACGTTC			
G1	64.5 ± 0.7	192 ± 45	410 ± 140
A2	60.0 ± 0.9	105 ± 39	150 ± 120
A3	67.1 ± 1.0	274 ± 51	670 ± 160
C4	65.0 ± 0.2	255 ± 11	614 ± 35
G5[b]	68.1 ± 2.6	190 ± 220	380 ± 720
T6	62.6 ± 0.7	169 ± 23	344 ± 73
T7	66.5 ± 1.4	286 ± 72	710 ± 230
C8	65.1 ± 0.4	166 ± 35	320 ± 110
GATGCATC			
G1	65.2 ± 1.2	270 ± 58	660 ± 180
A2[a]			
T3	66.0 ± 0.3	262 ± 10	633 ± 30
G4	65.2 ± 0.7	274 ± 45	670 ± 140
C5	66.4 ± 0.7	253 ± 24	603 ± 75
A6	66.3 ± 2.1	290 ± 150	730 ± 470
T7	66.3 ± 0.5	262 ± 15	633 ± 46
C8	66.9 ± 0.7	257 ± 26	614 ± 84

Table A.15 (continued)

	$\Delta G^\dagger_{310\,\mathrm{K}}$ / kJ · mol^{-1}	$\Delta H^\dagger$ / kJ · mol^{-1}	$\Delta S^\dagger$ / J · mol^{-1} · K^{-1}
GTAGCTAC			
G1	62.7 ± 1.0	188 ± 79	400 ± 250
T2	63.9 ± 0.9	213 ± 77	480 ± 250
A3	61.9 ± 1.1	290 ± 180	730 ± 580
G4	64.6 ± 0.1	219 ± 6	499 ± 18
C5	64.0 ± 0.3	229 ± 9	532 ± 27
T6[b]	52 ± 17	570 ± 320	1670 ± 980
A7[a]			
C8	64.9 ± 0.2	214 ± 9	479 ± 27

[a]Fitting of the exchange line shape not possible. [b]Errors of exchange rates too high to obtain a reliable Eyring fit. [c]Parabolic extrapolation of single-strand chemical shift. [d]Anomalous behaviour violating the two-state model, discussed in § 7.7 on page 109

Table A.16 Activation parameters of dissociation of self-complementary duplex DNA obtained from ^{1}H NMR of **AH2 and TH7**

	$\Delta G^\dagger_{310\,\mathrm{K}}$ / kJ · mol^{-1}	$\Delta H^\dagger$ / kJ · mol^{-1}	$\Delta S^\dagger$ / J · mol^{-1} · K^{-1}
CAACGTTG			
A2	67.4 ± 1.9	205 ± 32	445 ± 96
A3	66.9 ± 0.8	239 ± 23	555 ± 71
T6	66.8 ± 0.1	240 ± 4	560 ± 12
T7	67.4 ± 0.4	261 ± 13	624 ± 41
CATCGATG			
A2	66.2 ± 0.4	178 ± 19	361 ± 60
T3	67.1 ± 0.3	238 ± 7	552 ± 23
A6	66.3 ± 0.2	229 ± 11	524 ± 35
T7	66.7 ± 0.1	261 ± 3	627 ± 10
CTACGTAG			
T2[a]			
A3	63.7 ± 1.2	231 ± 29	540 ± 89
T6	61.6 ± 1.3	313 ± 62	810 ± 200
A7	62.2 ± 0.5	174 ± 14	362 ± 44

Table A.16 (continued)

	$\Delta G^{\dagger}_{310\,K}$ / kJ · mol^{-1}	$\Delta H^{\dagger}$ / kJ · mol^{-1}	$\Delta S^{\dagger}$ / J · mol^{-1} · K^{-1}
CTTCGAAG			
T2	67.3 ± 0.7	311 ± 63	785 ± 203
T3	66.2 ± 0.2	226 ± 12	515 ± 39
A6	65.6 ± 0.4	180 ± 13	368 ± 41
A7	66.6 ± 0.2	186 ± 17	387 ± 56
CTTGCAAG			
T2	66.1 ± 0.2	239 ± 8	557 ± 26
T3	67.3 ± 0.3	302 ± 15	756 ± 48
A6	65.9 ± 2.3	240 ± 37	560 ± 110
A7	66.6 ± 0.9	242 ± 23	567 ± 71
GAACGTTC			
A2	65.2 ± 3.2	226 ± 79	520 ± 250
A3[b]	80 ± 22	910 ± 970	2700 ± 3100
T6	65.5 ± 0.3	248 ± 9	588 ± 29
T7	66.9 ± 1.3	280 ± 61	690 ± 190
GATGCATC			
A2[b]	79 ± 32	510 ± 690	1400 ± 2100
T3	65.4 ± 0.1	240 ± 5	562 ± 17
A6	65.6 ± 0.4	240 ± 14	562 ± 44
T7	65.9 ± 0.3	247 ± 10	584 ± 32
GTAGCTAC			
T2	64.7 ± 0.1	250 ± 4	599 ± 13
A3	64.2 ± 0.2	231 ± 7	537 ± 22
T6	65.0 ± 0.1	240 ± 6	566 ± 20
A7	65.7 ± 0.4	228 ± 15	525 ± 47

[a]Fitting of the exchange line shape not possible. [b]Errors of exchange rates too high to obtain a reliable Eyring fit

A.5.4 Kinetics – methylated ODN

Table A.17 Activation parameters of dissociation of self-complementary duplex DNA with methylated cytosine obtained from ^{1}H NMR of **H6 and H8**

	$\Delta G^\dagger_{310\,\mathrm{K}}$ / kJ · mol^{-1}	$\Delta H^\dagger$ / kJ · mol^{-1}	$\Delta S^\dagger$ / J · mol^{-1} · K^{-1}
CAAm^{5}CGTTG			
C1	67.1 ± 0.9	229 ± 24	524 ± 74
A2	66.6 ± 1.4	233 ± 47	540 ± 150
A3	66.3 ± 1.5	214 ± 42	480 ± 130
m^{5}C4	66.9 ± 0.7	247 ± 15	582 ± 47
G5	67.9 ± 0.4	245 ± 10	570 ± 32
T6	67.7 ± 0.4	245 ± 15	571 ± 48
T7	65.4 ± 1.7	232 ± 54	540 ± 170
G8[b]	60 ± 640	200 ± 10000	500 ± 32000
CATm^{5}CGATG			
C1	67.4 ± 0.9	234 ± 22	538 ± 69
A2[b]	65.5 ± 7.8	530 ± 760	1500 ± 2400
T3	67.6 ± 0.6	238 ± 12	551 ± 38
m^{5}C4	67.9 ± 1.5	283 ± 57	700 ± 180
G5	70.6 ± 0.7	372 ± 37	970 ± 120
A6	67.1 ± 1.0	232 ± 26	531 ± 79
T7	66.9 ± 0.6	232 ± 15	534 ± 45
G8	67.5 ± 0.9	253 ± 28	599 ± 88
CTTm^{5}CGAAG			
C1	68.2 ± 1.3	183 ± 38	370 ± 120
T2	68.9 ± 1.2	229 ± 60	520 ± 190
T3	67.7 ± 0.6	228 ± 38	520 ± 120
m^{5}C4[b]	73 ± 50	600 ± 1500	1800 ± 4700
G5[cd]	67.4 ± 0.2	196 ± 16	414 ± 51
A6[a]			
A7	70.4 ± 0.8	255 ± 30	597 ± 93
G8	67.1 ± 0.5	179 ± 10	363 ± 30

[a]Fitting of the exchange line shape not possible. [b]Errors of exchange rates too high to obtain a reliable Eyring fit [c]Parabolic extrapolation of single-strand chemical shift. [d]Anomalous behaviour violating the two-state model, discussed in § 8.4.2 on page 137

Table A.18 Activation parameters of dissociation of self-complementary duplex DNA with methylated cytosine obtained from ^{1}H NMR of **AH2, m^{5}CH7, and TH7**

	$\Delta G^\dagger_{310\,K}$ / kJ · mol^{-1}	$\Delta H^\dagger$ / kJ · mol^{-1}	$\Delta S^\dagger$ / J · mol^{-1} · K^{-1}
CAAm^{5}CGTTG			
A2	66.9 ± 0.7	246 ± 12	579 ± 38
A3	67.9 ± 0.8	265 ± 27	635 ± 84
m^{5}C4	68.0 ± 0.2	249 ± 5	585 ± 17
T6	67.9 ± 0.1	258 ± 3	613 ± 9
T7	68.3 ± 0.3	264 ± 8	631 ± 24
CATm^{5}CGATG			
A2	72.2 ± 1.7	302 ± 30	741 ± 93
T3	68.7 ± 0.2	253 ± 7	594 ± 21
m^{5}C4	68.8 ± 0.3	262 ± 9	624 ± 28
A6	67.7 ± 0.9	237 ± 25	546 ± 76
T7	67.9 ± 0.1	228 ± 5	518 ± 15
CTTm^{5}CGAAG			
T2	69.6 ± 1.3	321 ± 76	810 ± 240
T3	69.2 ± 1.0	257 ± 45	600 ± 140
m^{5}C4	68.0 ± 0.2	239 ± 16	552 ± 50
A6	66.7 ± 0.2	164 ± 12	315 ± 39
A7	68.0 ± 0.3	163 ± 13	305 ± 43

[a] Errors of exchange rates too high to obtain a reliable Eyring fit

A.5.5 Melting of SRE ODN

Table A.19 Folding parameters of SRE-based DNA obtained from ^{1}H NMR of **H6 and H8**. All values obtained from fitting of chemical shifts (no exchange line shape)

	$\frac{T_m}{K}$	$\frac{\Delta G_{310\,K}}{kJ \cdot mol^{-1}}$	$\frac{\Delta H}{kJ \cdot mol^{-1}}$	$\frac{\Delta S}{J \cdot mol^{-1} \cdot K^{-1}}$
TGTCCATATTAGGACA		s1fos16		
T1	333.5 ± 1.2	-8.6 ± 0.4	-122 ± 8	-367 ± 26
G2	333.8 ± 0.8	-9.9 ± 0.6	-138 ± 11	-414 ± 32
T3	334.2 ± 0.5	-10.5 ± 0.3	-145 ± 6	-434 ± 18
C4	334.1 ± 0.4	-9.9 ± 0.3	-138 ± 4	-413 ± 14
C5	333.8 ± 0.2	-9.8 ± 0.2	-137 ± 3	-412 ± 9
A6	333.3 ± 1.8	-9.9 ± 1.2	-141 ± 22	-424 ± 67
T7	336.5 ± 0.5	-14.9 ± 1.1	-190 ± 14	-563 ± 43
A8	333.4 ± 0.6	-9.6 ± 0.4	-137 ± 7	-411 ± 22
T9	335.2 ± 1.6	-10.0 ± 0.8	-133 ± 14	-395 ± 44
T10	334.2 ± 0.8	-10.1 ± 0.5	-140 ± 9	-419 ± 27
A11	334.8 ± 1.8	-7.8 ± 0.4	-105 ± 7	-314 ± 23
G12	334.4 ± 0.5	-10.9 ± 0.5	-149 ± 7	-445 ± 22
G13	344.2 ± 7.9	-9.7 ± 3.0	-97 ± 16	-282 ± 44
A14	333.5 ± 0.7	-9.3 ± 0.4	-132 ± 7	-395 ± 20
C15	334.0 ± 0.2	-10.0 ± 0.1	-140 ± 2	-419 ± 7
A16	333.6 ± 0.7	-8.4 ± 0.3	-119 ± 6	-355 ± 18

Table A.19 (continued)

	$\frac{T_m}{K}$	$\frac{\Delta G_{310\,K}}{kJ \cdot mol^{-1}}$	$\frac{\Delta H}{kJ \cdot mol^{-1}}$	$\frac{\Delta S}{J \cdot mol^{-1} \cdot K^{-1}}$
GTCCATATTAGGAC		s1fos14		
G2	317.0 ± 1.3	−1.5 ± 0.2	−69 ± 6	−219 ± 18
T3	311.3 ± 2.2	−0.6 ± 1.0	−130 ± 25	−417 ± 79
C4	325.5 ± 0.6	−6.3 ± 0.3	−132 ± 7	−406 ± 23
C5	325.5 ± 0.3	−5.9 ± 0.1	−124 ± 3	−380 ± 10
A6	323.6 ± 3.8	−6.3 ± 2.0	−149 ± 54	−460 ± 170
T7	327.9 ± 0.6	−7.9 ± 0.5	−145 ± 10	−442 ± 32
A8	325.2 ± 0.8	−5.8 ± 0.3	−125 ± 9	−384 ± 27
T9	326.5 ± 1.8	−7.6 ± 1.0	−150 ± 24	−461 ± 74
T10	325.7 ± 1.1	−6.4 ± 0.5	−133 ± 12	−409 ± 38
A11	324.2 ± 3.5	−4.5 ± 0.8	−102 ± 14	−314 ± 45
G12	325.8 ± 0.5	−6.2 ± 0.4	−127 ± 8	−391 ± 25
G13	324.5 ± 2.7	−4.5 ± 1.4	−102 ± 21	−313 ± 64
A14	326.1 ± 0.8	−7.1 ± 0.5	−144 ± 12	−441 ± 36
C15	326.1 ± 0.2	−6.7 ± 0.1	−136 ± 3	−416 ± 9
TCCATATTAGGA		s1fos12		
T3	302 ± 10	1.1 ± 1.4	−42 ± 28	−138 ± 92
C4	315.5 ± 9.8	−1.0 ± 1.6	−55 ± 35	−170 ± 110
C5	308.4 ± 3.2	0.4 ± 0.8	−75 ± 7	−244 ± 22
A6	303 ± 47	1.3 ± 8.3	−55 ± 31	−181 ± 88
T7	314.6 ± 6.6	−0.8 ± 1.1	−54 ± 5	−172 ± 17
A8	300.0 ± 7.2	2.6 ± 1.5	−79 ± 18	−262 ± 56
T9	329 ± 17	−3.3 ± 5.0	−57 ± 68	−170 ± 210
T10	307 ± 10	0.4 ± 1.5	−51 ± 27	−165 ± 85
A11	302 ± 13	2.2 ± 3.6	−79 ± 15	−263 ± 54
G12	303.4 ± 7.2	1.5 ± 1.6	−68 ± 7	−225 ± 22
G13	308 ± 19	0.4 ± 3.2	−51 ± 16	−166 ± 59
A14	298.8 ± 8.5	2.2 ± 1.3	−59 ± 16	−198 ± 50

Table A.20 Folding parameters of SRE-based DNA obtained from ^{1}H NMR of **AH2, CH5, and TH7**. All values obtained from fitting of chemical shifts (no exchange line shape)

	$\frac{T_m}{K}$	$\frac{\Delta G_{310\,K}}{kJ \cdot mol^{-1}}$	$\frac{\Delta H}{kJ \cdot mol^{-1}}$	$\frac{\Delta S}{J \cdot mol^{-1} \cdot K^{-1}}$
TGTCCATATTAGGACA		s1fos16		
T1	333.3 ± 0.3	-9.3 ± 0.2	-133 ± 3	-398 ± 10
T3	334.0 ± 0.2	-10.6 ± 0.2	-148 ± 3	-443 ± 9
C4	334.3 ± 0.2	-10.9 ± 0.2	-150 ± 3	-450 ± 10
C5	334.1 ± 0.2	-10.4 ± 0.2	-143 ± 3	-429 ± 10
A6	333.9 ± 0.9	-8.6 ± 0.3	-120 ± 6	-358 ± 19
T7	337.5 ± 1.9	-18.1 ± 5.5	-222 ± 72	-660 ± 210
A8	334.8 ± 0.6	-12.0 ± 0.7	-161 ± 11	-481 ± 33
T9	337.3 ± 0.7	-17.2 ± 1.6	-212 ± 22	-630 ± 66
T10	333.9 ± 0.7	-10.1 ± 0.5	-142 ± 8	-424 ± 25
A11	334.1 ± 0.4	-10.5 ± 0.3	-145 ± 6	-435 ± 18
A14	337.7 ± 3.8	-16.2 ± 8.6	-200 ± 110	-590 ± 330
C15	334.2 ± 0.1	-10.8 ± 0.2	-150 ± 3	-447 ± 8
A16	333.4 ± 0.2	-8.7 ± 0.1	-125 ± 2	-374 ± 7
GTCCATATTAGGAC		s1fos14		
T3	326.2 ± 0.2	-7.0 ± 0.2	-142 ± 3	-435 ± 11
C4[a]				
C5[a]				
A6	324.5 ± 1.7	-5.1 ± 0.5	-115 ± 9	-353 ± 30
T7	330.4 ± 2.2	-7.2 ± 1.3	-116 ± 24	-350 ± 73
A8	326.5 ± 0.8	-8.0 ± 0.6	-158 ± 15	-483 ± 46
T9	330.0 ± 0.9	-10.4 ± 1.1	-171 ± 21	-520 ± 64
T10	325.4 ± 0.9	-6.5 ± 0.4	-138 ± 11	-425 ± 35
A11	325.5 ± 0.6	-6.5 ± 0.3	-137 ± 8	-422 ± 25
A14	320.2 ± 1.2	-2.7 ± 0.3	-84 ± 8	-261 ± 24
C15[a]				

Table A.20 (continued)

	T_m / K	$\Delta G_{310\,K}$ / kJ · mol^{-1}	ΔH / kJ · mol^{-1}	ΔS / J · mol^{-1} · K^{-1}
TCCATATTAGGA		s1fos12		
T3	302.7 ± 7.1	1.3 ± 1.1	−56 ± 12	−185 ± 35
C4[a]				
C5[a]				
A6	310.2 ± 6.3	−0.1 ± 2.4	−121 ± 30	−391 ± 99
T7	299 ± 20	2.7 ± 4.7	−73 ± 15	−243 ± 43
A8[b]				
T9	275 ± 19	6.7 ± 3.7	−53 ± 6	−192 ± 14
T10	325 ± 25	−2.9 ± 6.4	−65 ± 42	−200 ± 120
A11[b]				
A14	308.6 ± 3.9	0.4 ± 1.1	−83 ± 19	−267 ± 63

[a]Resonance not analysed. [b]Fit unfeasible due to absence of the sigmoidal character of $\delta(T)$

Table A.21 Folding parameters of SRE-based s1fos16, TGTCCATATTAGGACA, obtained from ^{1}H NMR of **H1′**. All values obtained from fitting of chemical shifts (no exchange line shape)

	T_m / K	$\Delta G_{310\,K}$ / kJ · mol^{-1}	ΔH / kJ · mol^{-1}	ΔS / J · mol^{-1} · K^{-1}
T1	335.2 ± 0.6	−12.8 ± 0.9	−171 ± 13	−509 ± 38
G2	334.9 ± 1.0	−11.4 ± 0.8	−154 ± 13	−459 ± 40
T3	333.9 ± 1.0	−9.6 ± 0.7	−134 ± 12	−400 ± 37
C4	334.6 ± 0.4	−11.0 ± 0.4	−150 ± 6	−449 ± 18
C5	336.8 ± 2.6	−8.9 ± 1.4	−112 ± 22	−334 ± 67
A6	342.0 ± 1.9	−24.2 ± 8.1	−259 ± 92	−760 ± 270
T7	337.3 ± 1.1	−15.8 ± 2.5	−195 ± 32	−579 ± 95
A8	334.6 ± 0.6	−10.5 ± 0.5	−142 ± 8	−425 ± 25
T9	334.5 ± 1.1	−9.8 ± 0.7	−133 ± 11	−399 ± 35
T10	335.6 ± 0.6	−12.0 ± 0.8	−158 ± 11	−470 ± 34
A11	334.5 ± 0.3	−10.1 ± 0.2	−138 ± 4	−413 ± 12
G12	334.8 ± 0.3	−10.7 ± 0.3	−144 ± 5	−431 ± 14
G13	336.2 ± 1.0	−11.7 ± 0.6	−150 ± 10	−445 ± 29
A14	334.8 ± 2.5	−8.3 ± 0.8	−112 ± 14	−335 ± 45
C15	335.1 ± 0.2	−11.3 ± 0.3	−151 ± 4	−450 ± 12
A16	333.4 ± 1.7	−8.2 ± 0.8	−117 ± 14	−352 ± 44