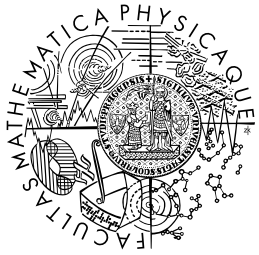


FACULTY
OF MATHEMATICS
AND PHYSICS
Charles University

BACHELOR THESIS

Hana Bušková

Melting of a methylated CpG-containing DNA double helix

Department of Low-Temperature Physics

Supervisor of the bachelor thesis: Mgr. Václav Římal, Ph.D.

Study programme: Physics

Study branch: Applied Physics

Prague 2018

I declare that I carried out this bachelor 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 Sb., the Copyright Act, as amended, in particular the fact that the Charles University has the right to conclude a license agreement on the use of this work as a school work pursuant to Section 60 subsection 1 of the Copyright Act.

In Prague, 19th of July, 2018 signature of the author

I wish to express my sincere thanks to my thesis supervisor, Mgr. Václav Římal, Ph.D., for valuable help and immense patience. I would also like to thank my family, for pretending to listen when I needed to explain a topic in order to understand it, and for always having chocolate around.

Title: Melting of a methylated CpG-containing DNA double helix

Author: Hana Bušková

Department: Department of Low-Temperature Physics

Supervisor: Mgr. Václav Římal, Ph.D., Department of Low-Temperature Physics

Abstract: The melting of a DNA double helix into two single strands is a crucial process for life on Earth. Previous experimental data on the self-complementary sequence CTTm^{5}CGAAG containing 5-methylcytosine suggests that the melting of an oligonucleotide might be a three-state, rather than two-state process. To assess this, ^{1}H NMR spectra were measured at varied concentrations of the oligonucleotide in temperatures between 274 and 366 K. Through analysis of the chemical shifts of aromatic hydrogens undergoing fast chemical exchange, we have determined that the most likely intermediate state between a double helix and a single stranded random coil is a structure consisting of two oligonucleotide molecules rather than a structured single strand.

Keywords: NMR spectroscopy, DNA, duplex melting

Contents

Introduction

The deoxyribonucleic acid (DNA) is the carrier of the genetic information of all eucaryotic and procaryotic life on Earth. Its four bases (as per Watson–Crick canonical pairing) form a source code of life. Research shows that our genetic information isn't just read-only, but it's a molecule subject to change as much as all the matter around it. It's subject to laws of physics and chemistry like everything else. It's subject to an increase of entropy and degradation, like everything else is.

Nuclear magnetic resonance (or NMR) can lend itself to a variety of experiments and experimental subjects. Its principle is vastly different from other spectral methods, and it's exceptionally useful when researching short and simple oligonucleotides, while also being indispensable in assessing their thermal parameters. In studies of melting DNA duplexes, presence of a two-step process has been revealed, with a transition state between the duplex and single strand secondary structures of the deoxynucleotide CTTCGAAG [1]. A decisive conclusion about the nature of the intermediate state appears to still be missing. Methylation of the central cytosine seems to increase the population of the intermediate, perhaps because it has a greater thermal stability than its non-methylated counterpart [2]. The main goal of this thesis is to distinguish between two possible melting schemes based on an analysis of concentration and temperature dependent ^{1}H NMR spectra of CTTm^{5}CGAAG.

1. Deoxyribonucleic acid

1.1 Primary structure of DNA

The DNA consists of a hierarchical structure. First, there are the bases, heterocyclic compounds shown in Fig. 1.1. For the sake of clarity, all members of the rings are numbered. Thymine and cytosine, together with uracil (an RNA base) are derivates of pyrimidine, so they're collectively referred to as pyrimidine bases. For similar reasons, adenine and guanine are dubbed purine bases.

Next, we attach our deoxyribose via a glycosidic bond between its first carbon and a nitrogen of a base on the first (for pyrimidines) or ninth position (for purines), as shown in Fig. 1.2 for an adenine. If we substitute the -OH group on the 5' carbon of the deoxyribose by a phosphate group, we finally get a nucleotide. An esterification reaction between 3' carbon of this nucleotide, and a phosphate group of another, is a mechanism sufficient for polymerisation, and synthesis, of an oligonucleotide of any length. This also establishes an unambiguous direction of the strand, from the 5' end to the 3' end.

1.2 Secondary structure of DNA

The spatial structure of DNA is determined by non-covalent interactions, one of them being hydrogen bonds. Watson–Crick base pairs, shown in Fig. 1.3, are the most well-known form of such bonding, however, less traditional bondings of canonical base pairs, such as Hoogsteen pairings [3], non-canonical base pairs [4], as well as structures of three [5] or four [6] bases bonded by hydrogen bonds at once, have been observed both *in vivo* and *in vitro*.

The aromatic rings of the bases have the properties of electrical quadrupoles, thanks to the π electrons above and under the plane of the ring. This is one of the effects behind another interaction, which is stacking. As the name implies, stacking leads to a parallel position of the aromatic rings. Other important interactions

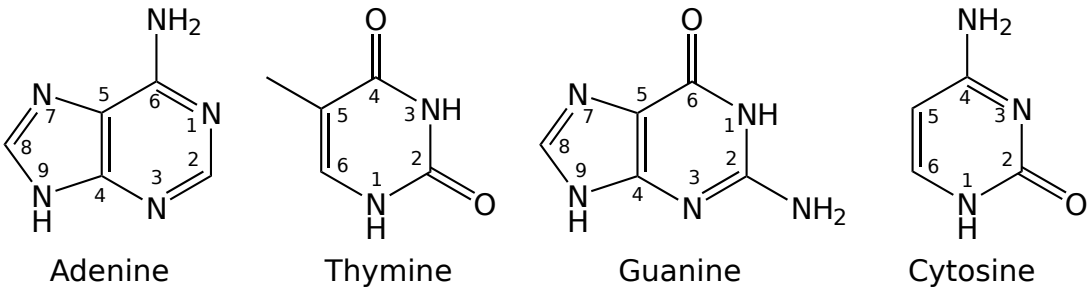


Figure 1.1: The four bases of DNA

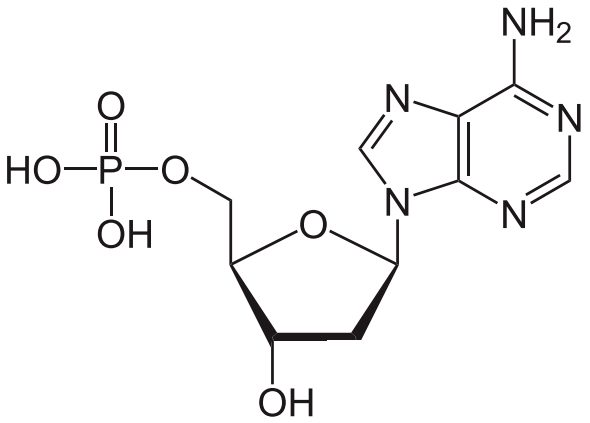


Figure 1.2: Attachment of adenine to a deoxyribose, with a phosphate group on its 5' end

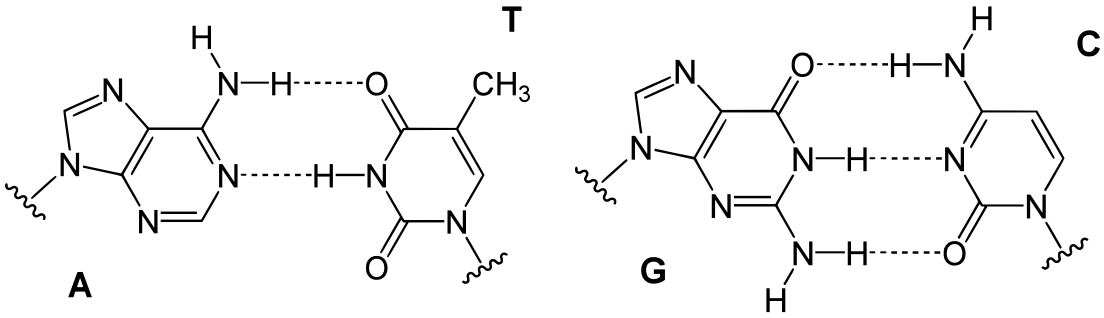


Figure 1.3: Watson–Crick canonical base pairing

responsible for the secondary structure of DNA include hydrophobic interaction and the electrostatic forces. They are present thanks to the charges and partial charges of the atoms in the molecule. Bases are hydrophobic, deoxyribose and phosphate are both hydrophilic. Since living cells consist of mostly water, the two strands wrap themselves around each other in a way that exposes the least of the bases to the surroundings. In crystalline form, this leads to the right-handed A-DNA structure observed by Franklin [7]. B-DNA, the form predicted by Watson and Crick [8] differs from it by the base-pair orientation being perpendicular relative to the fibre axis. The Watson–Crick hydrogen bonding can be seen on Figure 1.3. Alternative structures, such as protonated DNA [9] or the left handed Z-DNA form [10] have also been observed and play an important role in DNA transcription *in vivo* [11].

1.3 Melting of DNA duplex

At low enough temperatures, two complementary oligonucleotide strands that are close enough to each other usually form hydrogen bonds together. However, once the temperature is sufficiently high, the Brownian motion of the duplex and of the water around it makes it impossible to keep the hydrogen bonds intact. At

temperatures close to the boiling point of water, most molecules exist as single-strand only, likely forming transient hydrogen bonds with the surrounding water instead. To accurately judge the energetic characteristics of each of these two arrangements, it is advisable to use the Gibbs free energy

$$\Delta G = \Delta H - T\Delta S. \tag{1.1}$$

ΔH and ΔS represent the change of enthalpy and enthropy, while T is the temperature in kelvin.

1.3.1 Two-state system

Let us assume for now that there are two states only, single strand (S) and a duplex (D), between which an equilibrium is established for any given temperature. For self-complementary oligonucleotide, the reaction

$$\mathrm{D} \rightleftharpoons 2\mathrm{S} \tag{1.2}$$

has an equilibrium constant K for the backwards process, given as

$$K = \frac{c_\mathrm{D}}{c_\mathrm{S}^2} c_\mathrm{ref}, \tag{1.3}$$

where c_D and c_S mean the concentrations of double strands and single strands respectively. c_ref is the reference concentration, set to 1M. This way, we obtain a dimensionless equilibrium constant. The overall concentration of the oligonucleotide, c_0, remains constant during the experiment. It only follows that

$$c_0 = 2c_\mathrm{D} + c_\mathrm{S}. \tag{1.4}$$

We will also need the van't Hoff equation

$$\Delta G = -RT \ln K \tag{1.5}$$

and its linear form

$$\ln K = -\frac{\Delta H}{RT} + \frac{\Delta S}{R}. \tag{1.6}$$

It is more useful in our case to determine results in populations instead of concentrations. Population references the amount of molecules in a given state, so the sum of all self-complementary oligonucleotides in all states should add up to one. The conversion of populations and concentrations is as follows:

$$p_\mathrm{S} = \frac{c_\mathrm{S}}{c_0} \tag{1.7}$$

$$p_\mathrm{D} = 2\frac{c_\mathrm{D}}{c_0} \tag{1.8}$$

The population p_D of duplexes in the specimen is then calculated from equations (1.3), (1.4), and (1.8) as

$$p_{\mathrm{D}} = \frac{1 + \kappa - \sqrt{1 + 2\kappa}}{\kappa}, \tag{1.9}$$

where κ is determined by

$$\kappa = \frac{4c_0}{c_{\mathrm{ref}}} \exp\Big(- \frac{\Delta G}{RT}\Big). \tag{1.10}$$

If we further define the melting point of the duplex as the temperature at which half of the present oligonucleotides are single stranded, it is determined by

$$T_{\mathrm{m}} = \frac{\Delta H}{\Delta S - R \ln \frac{c_0}{c_{\mathrm{ref}}}}. \tag{1.11}$$

The equation that describes the average chemical shift of a hydrogen on an aromatic ring under fast chemical exchange conditions can then be described as

$$\delta = p_{\mathrm{D}}(a + bT) + (1 - p_{\mathrm{D}})(c + dT). \tag{1.12}$$

We assume (and have observed) that δ changes linearly with temperature for a one-state system. Therefore, it is necessary to fit the chemical shift as the weighted mean of two linear functions for a two-state system. The letters a, b, c and d represent the parameters of such functions.

1.3.2 Three-state system

In the previous subsection, we dealt with a simple two-state reaction. However, it should be possible for the oligonucleotide to assume a third state between the fully disconnected single strand state, and the fully bonded double strand. There are two possible routes for that kind of reaction, case A:

$$\mathrm{D} \rightleftharpoons 2\varphi \rightleftharpoons 2\mathrm{S}; \tag{1.13}$$

or case B:

$$\mathrm{D} \rightleftharpoons \chi \rightleftharpoons 2\mathrm{S}. \tag{1.14}$$

The equation (1.13) refers to the case where, as the temperature goes up, a double strand melts into two single strands. These assume a non-trivial secondary structure, denoted by φ. The secondary structure might be enforced by hydrogen bonding among the bases of the strand. With a sufficiently high temperature, the bonds break and the strand assumes the random coil structure. The other case, represented by equation (1.14), shows that the two oligonucleotides in the transition state stay connected, but in a different configuration than the original double strand. This is represented by the letter χ. However, the result for sufficiently high temperatures is two single strands again, same as in the previous

reaction. To avoid confusion, all variables and constants that assume the first reaction is happening will be denoted by the A subscript, while the ones that assume the second will have a B subscript henceforth.

The equilibrium constants of the reactions from the intermediate state to the double strand are given as

$$K_{1\mathrm{A}} = \frac{c_{\mathrm{DA}}}{c_{\varphi}^2} c_{\mathrm{ref}} \tag{1.15}$$

and

$$K_{1\mathrm{B}} = \frac{c_{\mathrm{DB}}}{c_{\chi}}. \tag{1.16}$$

The reactions between the intermediate states and the fully melted states then have the equilibrium constants

$$K_{2\mathrm{A}} = \frac{c_{\varphi}}{c_{\mathrm{SA}}} \tag{1.17}$$

and

$$K_{2\mathrm{B}} = \frac{c_{\chi}}{c_{\mathrm{SB}}^2} c_{\mathrm{ref}}. \tag{1.18}$$

The amount of the sample remains the same, so it follows that

$$c_0 = 2c_{\mathrm{DA}} + c_{\varphi} + c_{\mathrm{SA}} \tag{1.19}$$

or, in case of reaction (1.14) taking place

$$c_0 = 2c_{\mathrm{DB}} + 2c_{\chi} + c_{\mathrm{SB}}. \tag{1.20}$$

Alternatively, we can use populations instead of concentrations. Population of nuclei for both systems is then normalised as

$$p_{\mathrm{D}} + p_{\mathrm{i}} + p_{\mathrm{S}} = 1 \tag{1.21}$$

where p_{i} denotes the intermediate stages of both possible reaction routes. We obtain the populations by dividing the whole Eq. (1.19) or (1.20) by their c_0. By solving the systems of equations (1.15), (1.17), and (1.19) for reaction A, and (1.16), (1.18), and (1.20) for reaction B, we obtain

$$p_{\varphi} = \frac{\sqrt{(K_{2\mathrm{A}}+1)^2 + 8c_0 K_{1\mathrm{A}} K_{2\mathrm{A}}^2 / c_{\mathrm{ref}}} - K_{2\mathrm{A}} - 1}{4K_{1\mathrm{A}} K_{2\mathrm{A}} c_0 / c_{\mathrm{ref}}} \tag{1.22}$$

for the first reaction, and

$$p_{\mathrm{SB}} = \frac{\sqrt{1 + 8c_0 K_{2\mathrm{B}}(1+K_{1\mathrm{B}})/c_{\mathrm{ref}}} - 1}{4c_0(K_{1\mathrm{B}}+1)K_{2\mathrm{B}}/c_{\mathrm{ref}}} \tag{1.23}$$

for the second. The remaining populations can be easily calculated from these two as

$$p_{\mathrm{DA}} = 2K_{1\mathrm{A}}c_0 p_\varphi^2/c_{\mathrm{ref}} \tag{1.24}$$

$$p_{\mathrm{SA}} = \frac{1}{K_{2\mathrm{A}}} p_\varphi \tag{1.25}$$

$$p_{\mathrm{DB}} = 2K_{1\mathrm{B}}K_{2\mathrm{B}}c_0 p_{\mathrm{SB}}^2/c_{\mathrm{ref}} \tag{1.26}$$

$$p_\chi = 2K_{2\mathrm{B}}c_0 p_{\mathrm{SB}}^2/c_{\mathrm{ref}}. \tag{1.27}$$

The resulting average chemical shift under fast chemical exchange is then determined by an expression similar to equation (1.12):

$$\delta = p_{\mathrm{D}}(a + bT) + p_{\mathrm{i}}(d + cT) + p_{\mathrm{S}}(e + fT). \tag{1.28}$$

2. Nuclear Magnetic Resonance

A spinning electric charge generates a magnetic field around it, and possesses a magnetic moment, $\boldsymbol{\mu}$. Analogically to classical physics, the magnetic moment of a nucleus is

$$\boldsymbol{\mu} = \gamma \boldsymbol{I}, \tag{2.1}$$

where $\boldsymbol{I}$ is the spin of the nucleus, and γ is the gyromagnetic ratio of the nucleus, unique for every nuclide and its state [12]. Nuclei with an odd number of nucleons have a spin $k/2$, and nuclei with an even number of nucleons have a spin k ($k \in \mathbb{Z}_0$). If a magnetic field $\boldsymbol{B_0}$ is present, the degeneration of energy levels corresponding to different projections of the spin is removed and the energies split. For example, in the case of a spin-1/2 nucleus, like ^{1}H, two are possible. The difference between adjacent energies (determined by Zeeman effect) is given as

$$\Delta E = \gamma B_0 \hbar, \tag{2.2}$$

and the frequency of a radiofrequency field (the Larmor frequency) that can induce a transition between energy levels is given by

$$\omega_0 = \gamma B_0. \tag{2.3}$$

2.1 Bloch Equations

The Bloch equations relate the magnetic field $\boldsymbol{B}$ to the nuclear magnetisation $\boldsymbol{M}$, which is a sum of all magnetic moments $\boldsymbol{\mu}$.

$$\frac{\mathrm{d}M_\mathrm{x}}{\mathrm{d}t} = \gamma(\boldsymbol{M} \times \boldsymbol{B})_\mathrm{x} - \frac{M_\mathrm{x}}{T_2} \tag{2.4}$$

$$\frac{\mathrm{d}M_\mathrm{y}}{\mathrm{d}t} = \gamma(\boldsymbol{M} \times \boldsymbol{B})_\mathrm{y} - \frac{M_\mathrm{y}}{T_2} \tag{2.5}$$

$$\frac{\mathrm{d}M_\mathrm{z}}{\mathrm{d}t} = \gamma(\boldsymbol{M} \times \boldsymbol{B})_\mathrm{z} - \frac{M_\mathrm{z} - M_0}{T_1} \tag{2.6}$$

T_1 is called longitudal relaxation time. Other terms for longitudal relaxation are thermal, or spin–lattice relaxation. Similarly, T_2 is the transverse relaxation time. Another term for it is spin–spin relaxation.

2.1.1 Free precession

Let us simplify (2.4), (2.5), and (2.6). By assuming $T_1 \to \infty, T_2 \to \infty$, we obtain the system

$$\frac{\mathrm{d}M_\mathrm{x}}{\mathrm{d}t} = \gamma(\boldsymbol{M} \times \boldsymbol{B})_\mathrm{x} \tag{2.7}$$
$$\frac{\mathrm{d}M_\mathrm{y}}{\mathrm{d}t} = \gamma(\boldsymbol{M} \times \boldsymbol{B})_\mathrm{y} \tag{2.8}$$
$$\frac{\mathrm{d}M_\mathrm{z}}{\mathrm{d}t} = \gamma(\boldsymbol{M} \times \boldsymbol{B})_\mathrm{z}. \tag{2.9}$$

We will further specify the problem by assuming the constant magnetic field $\boldsymbol{B}(t) = (0, 0, B_0)$, so the problem reduces to

$$\frac{\mathrm{d}M_\mathrm{x}}{\mathrm{d}t} = \gamma B_0 M_y \tag{2.10}$$
$$\frac{\mathrm{d}M_\mathrm{y}}{\mathrm{d}t} = -\gamma B_0 M_x \tag{2.11}$$
$$\frac{\mathrm{d}M_\mathrm{z}}{\mathrm{d}t} = 0 \tag{2.12}$$

with the non-trivial solution being, without loss of generality,

$$M_\mathrm{x} = C\cos(\gamma B_0 t) \tag{2.13}$$
$$M_\mathrm{y} = -C\sin(\gamma B_0 t) \tag{2.14}$$
$$M_\mathrm{z} = const. \tag{2.15}$$

Therefore, the transversal component of magnetisation vector in complex notation becomes

$$M_\perp(t) = Ce^{-i\omega t}. \tag{2.16}$$

Hence, if there was no damping due to relaxation, the magnetization vector would precess with the angular frequency ω. This applies to all nuclei with a non-zero gyromagnetic ratio that are present in the field. The real spectrum, however, undergoes both transverse and longitudal relaxation. Applying Fourier transform to the resulting signal results in a spectrum of frequencies corresponding to the nuclei present.

2.2 Chemical Shift

There is a slight difference between the outside field applied, B_0, and the actual field around a nucleus, B_{local}. The reason for this difference is the electron density surrounding the nucleus itself. We will denote the shielding effect strength as σ. With this correction applied, a nucleus in an isotropic liquid in the outside field B_0 with σ will have an angular resonance frequency

$$\omega = \gamma(1 - \sigma)B_0. \tag{2.17}$$

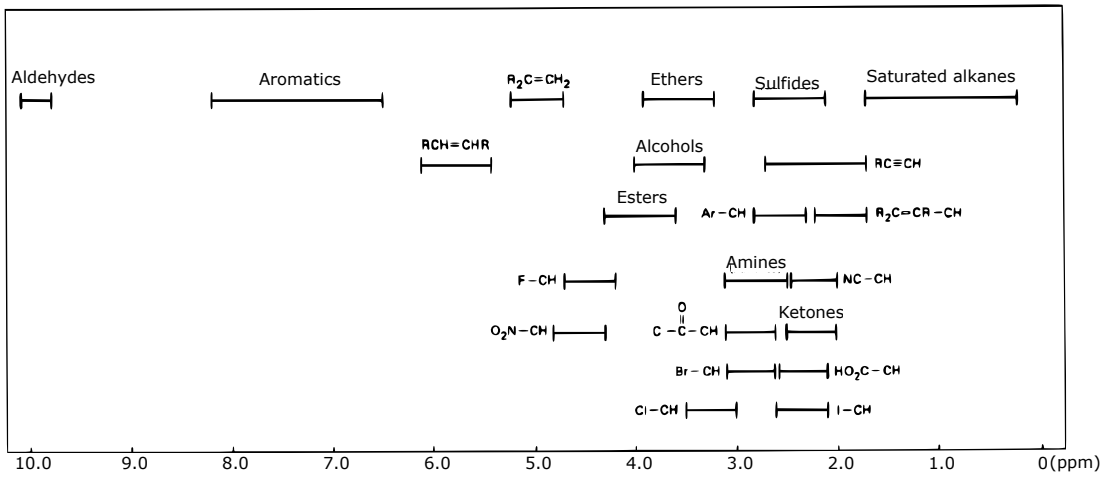


Figure 2.1: Chemical shifts in ^{1}H spectra for common organic functional groups, reproduction from [12]

In liquid samples, suspended in homogenous B_0, signals from nuclei in different places in a molecule are commonly discernible. Customarily, chemical shift δ is used instead of angular frequency. We define it as

$$\delta = \frac{\omega - \omega_{\mathrm{s}}}{\omega_{\mathrm{s}}}, \tag{2.18}$$

usually in the units of parts per million (ppm); ω_s denotes a standard for the experiment, for example, tetramethylsilane (TMS) for experiments in non-polar solvent, or 4,4-dimethyl-4-silapentane-1-sulfonic acid (DSS) in water.

2.3 Chemical exchange

NMR methods can be used to observe the kinetics of chemical reactions or conformational changes. This chemical exchange means that the nucleus's surroundings, and therefore its signal on the spectrum, change. The reaction rate constant k can be either greater than, equal, or smaller than the difference in resonance frequencies corresponding to the two states occurung, $\Delta\nu$. In slow chemical exchange ($k \ll \Delta\nu$), the result is two distinct peaks corresponding to the two states discernible in NMR. For $k \approx \Delta\nu$, the spectral shape is severely affected and coalescence of the two peaks into one line occurs at a certain rate. Fast chemical exchange, which occurs for $k \gg \Delta\nu$, is characterized by one peak at the weighted average position that corresponds to the two states. The change in conformation of short oligonucleotide duplexes, such as in reactions (1.2), (1.13), and (1.14), is an example of fast chemical exchange [1].

3. Materials and methods

The spectra were measured on a Bruker Avance 500 III HD spectrometer, in a magnetic field of 11.7 T, which corresponds to hydrogen (^{1}H) resonance frequency 500 MHz. We used a BBFO probe, which allows for a liquid sample in a tube with a diameter of 5 mm. The probe has got three in/out ports: ^{1}H, ^{2}H, and a port for broadband channel. We used this port for ^{31}P.

In order to measure the chemical shift as a function of temperature, we warmed the samples to 366 K, adjusted the shimming coils in the spectrometer and the wobble, let them rest for at least half an hour, and measured the spectra. The computer was then programmed to decrease the temperature by 2 K, wait 15 minutes, adjust the shimming coils and wobble, measure the spectra, and repeat until the temperature 274 K was reached.[1] To ensure similar conditions for all spectra, the flow of nitrogen gas through the probe was set at 600 l/h for the whole temperature range.

3.1 Preparation of the samples

To allow for the lock on the signal of deuterium, we have used water with D_2O as our solvent. DSS sodium salt purchased as 1% solution in D_2O from Sigma–Aldritch was used as an internal standard. The exact amounts used are in Table 3.1. As we used a sample that was more concentrated than DNA in living cells, we chose for the buffer solution to match the increased concentration. Both of the solutions in Table 3.2 were mixed as 75 mM PO_4^{3-} salt[2] with 200 mM Na^{1+} in the buffer. We have mixed these together until the pH of the resulting solution was 7.0, which was measured with a pH meter.

The oligonucleotide with the sequence CTTm^{5}CGTTG was obtained from

[1] Although the described process cools the sample down, we assume the opposite direction to be true throughout this thesis, due to the somehow more cumbersome numerics and language of the former problem. With slow enough change of temperature, the melting and binding can be considered equivalent for our purposes and the system in a thermodynamic equilibrium.

[2] As the salts used were not stored in airtight containers, they absorbed humidity and contained an unknown fraction of water.

	V / ml
H_2O	89.98
D_2O	9.84
$99,9\,\%$ $D_2O + 1\,\%$ DSS	0.18

Table 3.1: The composition of the solvent used for buffer preparation

	V / ml	m / mg	m_{NaCl} / mg
$Na_2HPO_4 \cdot 2H_2O$	25	333.59	73.13
$NaH_2HPO_4 \cdot 2H_2O$	25	234.07	146.28

Table 3.2: Saline solutions used for buffer preparation

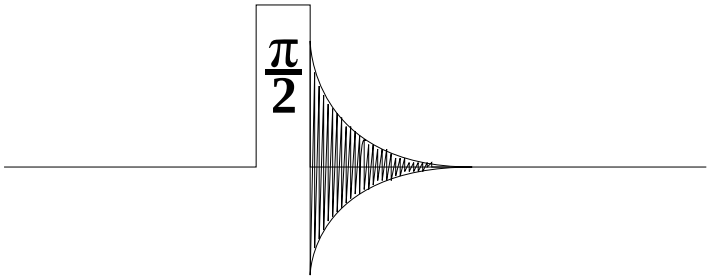


Figure 3.1: zg sequence

ATDBio, Ltd. Each sample had approximately quarter of the concentration of the previous one. The same buffer was used for all samples, resulting in increasing relative amount of inorganic phosphate ions to the oligonucleotide molecules.

3.2 Pulse Experiments

The simplest pulse sequence consists of a $\pi/2$ pulse, followed directly by the signal acquisition (Figure 3.1). It is an impractical sequence for biochemical applications when done on hydrogen atoms, due to the most common solvent being water in those cases. However, we have used the sequence on phosphorus (^{31}P) at the temperature 298 K.

For hydrogen, we use sequences that target water directly, and dampen its signals. It should be noted that the nuclei in the immediate proximity of water on the frequency scale are still impossible to analyse, but this approach grants a greater range of the spectrum for evaluation. The zgesgp sequence (Fig. 3.2) [13] uses a combination of square $\pi/2$ and π pulses, together with 4 ms long selective Gauss pulses, which enable us to choose the resonance frequency of water hydrogens, without affecting most of the others.

Table 3.3 lists the acquisition parameters of the three samples. In all cases, 4 dummy scans were run before the acquisition itself on ^{1}H. The data was then Fourier transformed in Topspin environment with no apodization. Zero order phase correction and a second order baseline correction were performed. The chemical shifts were referenced to the DSS standard 0 ppm.

	^{1}H	^{31}P
Pulse program	zgesgp	zg
NS		
$c = 9.5$ mM	64	32
$c = 2.4$ mM	512	128
$c = 0.54$ mM	2400	1150
D1 / s	1.00	60.0
AQ / s	1.64	2.03
TD	32768	16384

Table 3.3: Concentrations (c, from ^{31}P) and acquisition parameters. NS – number of scans; D1 – relaxation delay; AQ – acquisition time; TD – number of data points

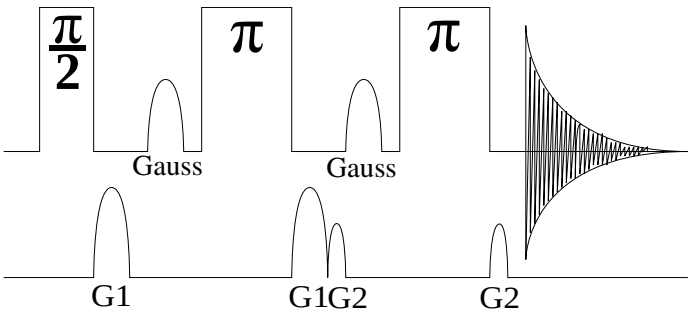


Figure 3.2: zgesgp sequence. Top row: ^{1}H channel; bottom row: z-gradients

3.3 Methods of analysis

Because the integral of a peak in an NMR spectrum is determined by its molar concentration, the results of the zg experiment were used to determine the concentrations of the oligonucleotide in the samples. Cross-multiplication of the peaks of the oligonucleotides with a peak of a known concentration (in our case, the inorganic phosphate nuclei) gives us a concentration of the unknown component with an error of about 5%.

The hydrogen spectra were processed partly through MATLAB toolbox Asymexfit [14], and partly through Octave with standard libraries. We used the toolbox to fit the chemical shifts in the aromatic region of the ^{1}H spectra.

This made it possible to track the peaks as they evolved over the temperature range, forming smooth curves. Asymexfit already contained a class populationData to fit curves without an intermediate state. The fitting of peaks with a discernible intermediate was done in the Octave environment with functions based on the Asymexfit package. The fitting function in Octave used Levenberg–Marquardt (dampened least squares) algorithm.

4. Results and Discussion

4.1 Concentrations

We have used the phosphorus zg spectrum of each sample to determine the concentration of oligonucleotides accurately. This was done in the Topspin environment, by integration of the peak belonging to the inorganic pshosphate, and comparison against the integral of the phosphate groups in the oligonucleotide. The ^{31}P zg spectrum of 9.5 mM sample is in the attached Fig. 4.1, and the sequence used is in Fig. 3.1. The resulting concentrations are included in Tab. 3.3.

4.2 Fits of 1D ^{1}H spectra

The zgesgp spectra of ^{1}H were aligned with the DSS standard at zero. The oligonucleotide spectrum consists of two major regions, the aromatics and the deoxyribose. Because of its well-resolved peaks, the aromatic region was loaded in MATLAB with Asymexfit package. Best fit of the measured zgesgp spectra in the aromatic region was done as a sum of Lorentz curves in Asymexfit.

The oligonucleotide and the buffer were both the same as the ones used in [1],

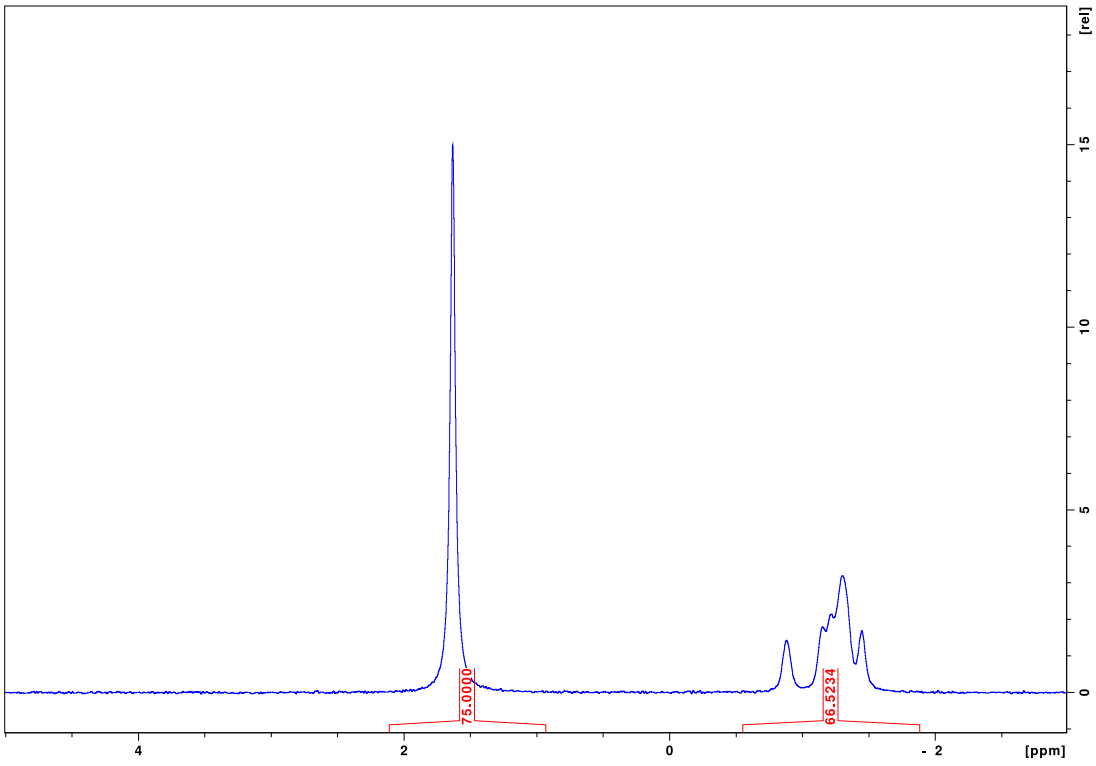


Figure 4.1: ^{31}P zg spectrum of 9.5 mM sample and its integrals

with the main difference being the concentration of samples. We therefore used peak assignment from there in this thesis. The result were smooth curves across the temperature axis, depicted on Figures A.1, A.2, and A.3.

4.3 Two-state model

The chemical shift curves from the precious step were used for the analysis of thermodynamic parameters of the nuclei. The Asymexfit package provided the class populationData for fitting of sigmoid curves from Eq. (1.12). The fitting function needed a good initial estimate in order to converge to a solution. The data, together with their fitting functions, are on Fig. 4.3 and 4.2. The calculated thermodynamic parameters are shown in Table A.1. Most peaks displayed the 2-state behavior clearly, with exceptions, which could be however still approximated by the same (see Equation (1.2)). With a few exceptions, the melting temperature T_m increases with concentration.

4.4 Three-state melting

For some peak curves, however, a sigmoid curve was not a good enough model, which suggested a third, intermediate, state. This effect was particularly strong in the G5H8 or T2H6 nuclei (Fig. 4.3 (c) and (e)), where it was visible even for the highest concentration. We have deduced that there were two possible mechanisms for this to occur, reactions (1.13) and (1.14). As the partial reactions $2\varphi \rightleftharpoons 2S$ in process (1.13) and $D \rightleftharpoons \chi$ in process (1.14) are of first order, they had to be the ones that did not change their melting temperature, T_m, with concentration, which wasn't the case for the other partial reactions, $D \rightleftharpoons 2\varphi$ and $\chi \rightleftharpoons 2S$.

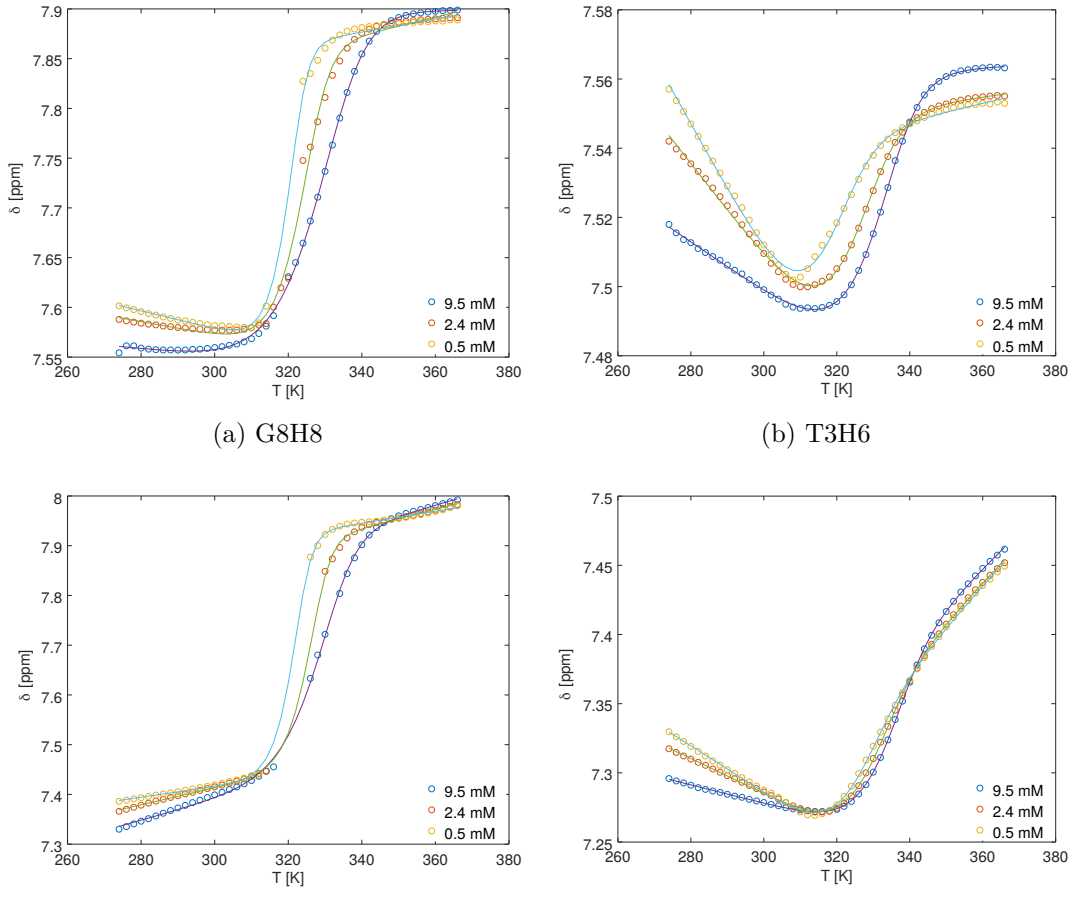
The first curves fitted for both models A and B and for both peaks G5H8 and T2H6 were for the 0.54 mM concentration, as it was the one where the effect of the intermediate was most pronounced.

As the formulas had 10 parameters, and because the remaining concentrations would inherit two of those parameters from the 0.54 mM samples, we had to ensure the parameters in the model would be as close to the global minimum of χ^2 as possible. This was achieved through multiple iterations of the fitting process for the first spectrum, with the parameters $a, b, c, d, e,$ and f from equation (1.28) and any of the two pairs of thermodynamic constants being kept rigid while the other parameters were destabilized on purpose to check if the fit would return to the same values. In higher concentrations, the same process was applied, but without allowing variance for the appropriate pair of thermodynamic constants for each process, so the ΔH and ΔS corresponding to the first-order reaction in their respective models were the same for all concentrations. The result can be

seen on Figures 4.4, 4.6, 4.8 and 4.10, with thermodynamic data in Table A.2.

4.5 Discussion

A closer look at Figures 4.4 and 4.8 reveals that the model A (φ intermediate) deviates from the data in 9.5 mM significantly, due to the fact that we forced it to assume a melting temperature from the lower concentrations. Model B with the χ on Figures 4.6 and 4.7 appears to be more accurate in this sense. A somewhat more robust way to look at the data is through the populations of the respective states, as on Figs. 4.5, 4.7, 4.9, and 4.11. The populations corresponding to the model A show the left melting temperature to either decrease with concentration (Fig. 4.5), or they do not appear to have a monotonous relation to the concentration. On the other hand, Figs. 4.7 and 4.11 seem to describe the data better. Figure 4.12 shows the plots of melting point temperature in Kelvin against concentration



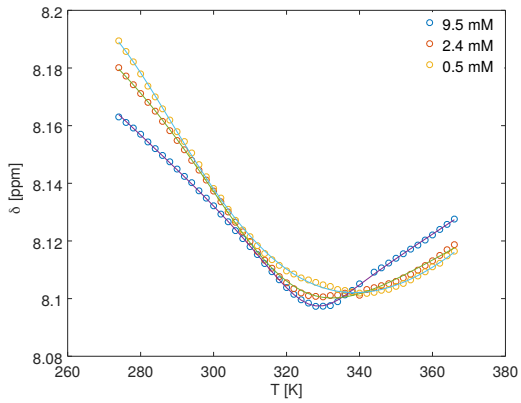
(a) G8H8

(b) T3H6

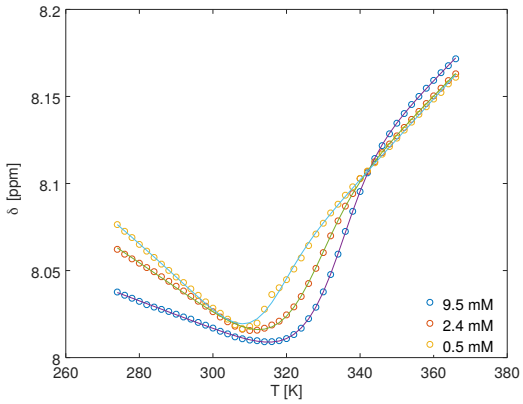
(c) A6H2

(d) mC4H6

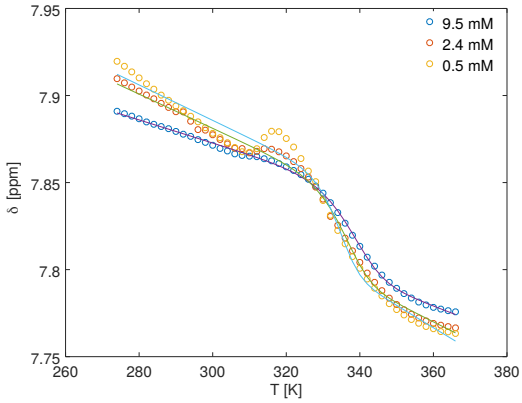
Figure 4.2: Chemical shift of aromatic ^{1}H nuclei. Circles represent the values from Lorentzian fits of the NMR spectra; the lines are the two-state melting fits described by Eq. (1.12).



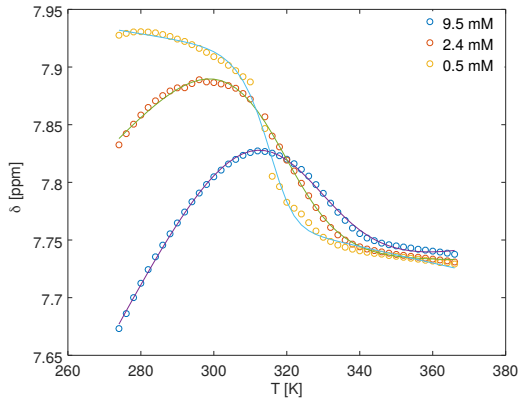
(a) A6H8



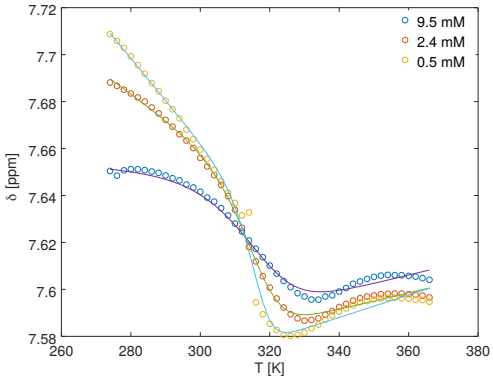
(b) A7H8



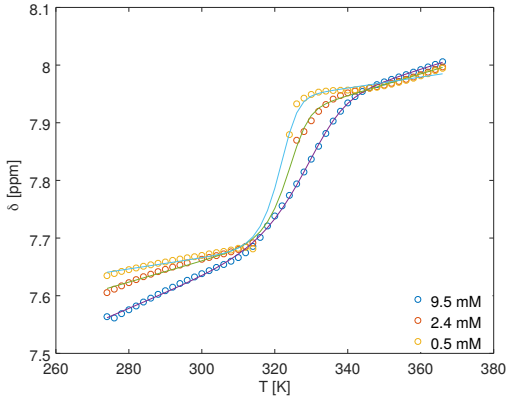
(c) G5H8



(d) C1H6



(e) T2H6



(f) A7H2

Figure 4.3: Chemical shift of aromatic ^{1}H nuclei. Circles represent the values from Lorentzian fits of the NMR spectra; the lines are the two-state melting fits described by Eq. (1.12).

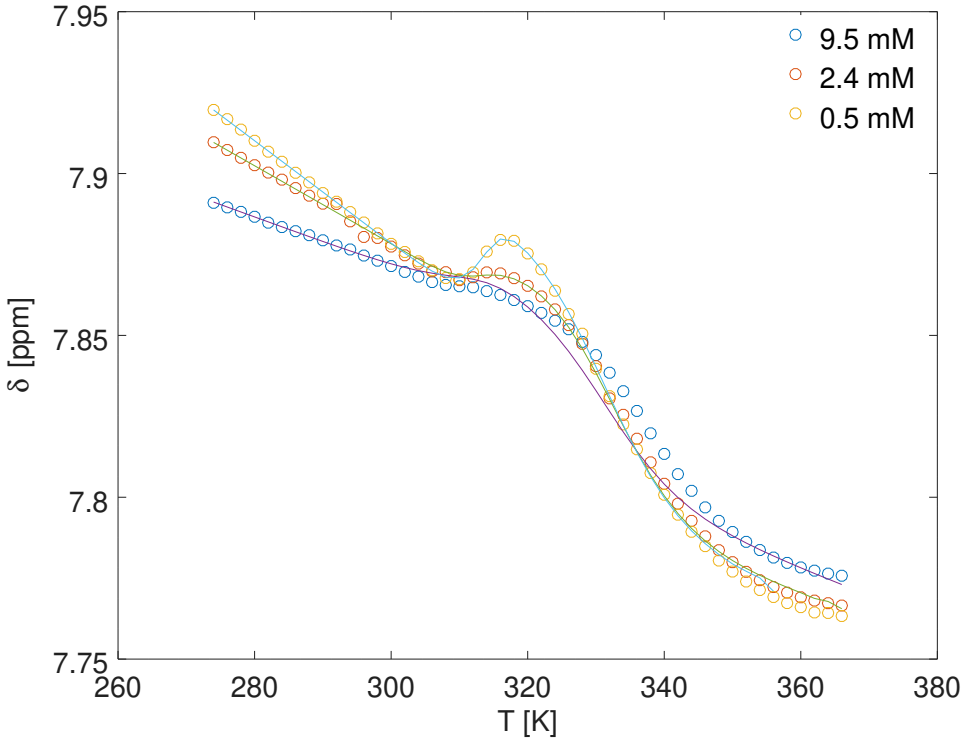


Figure 4.4: Chemical shift of G5H8 nuclei. Circles represent the values from Lorentzian fits of the NMR spectra; the lines are the three-state fits described by Eq. (1.28), fitting model A.

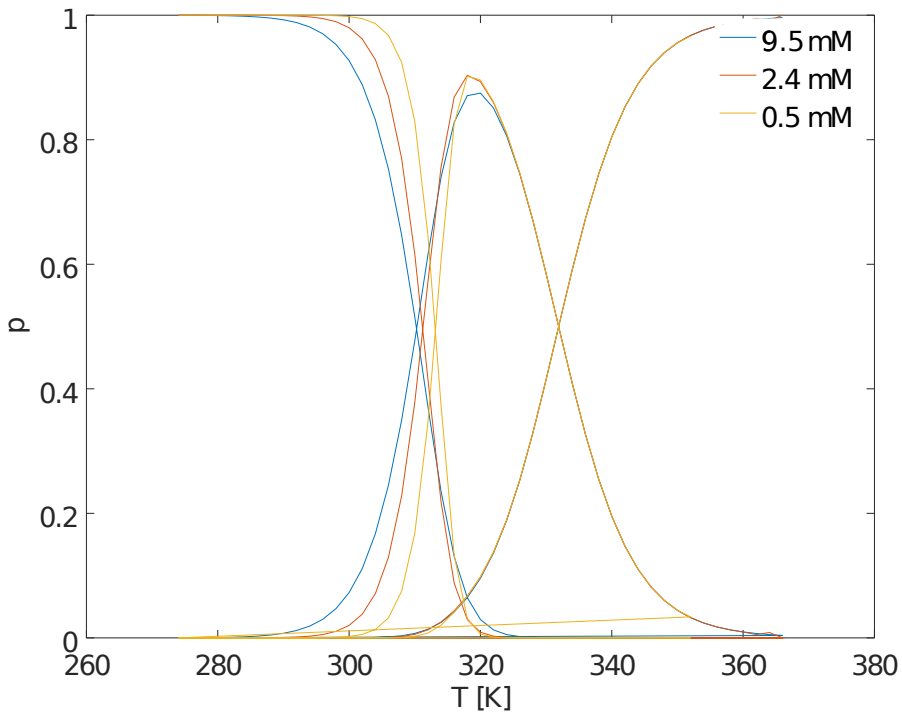


Figure 4.5: Populations of states by concentration for G5H8, results of fits by model A. Colours are the same for all states in a sample for clarity.

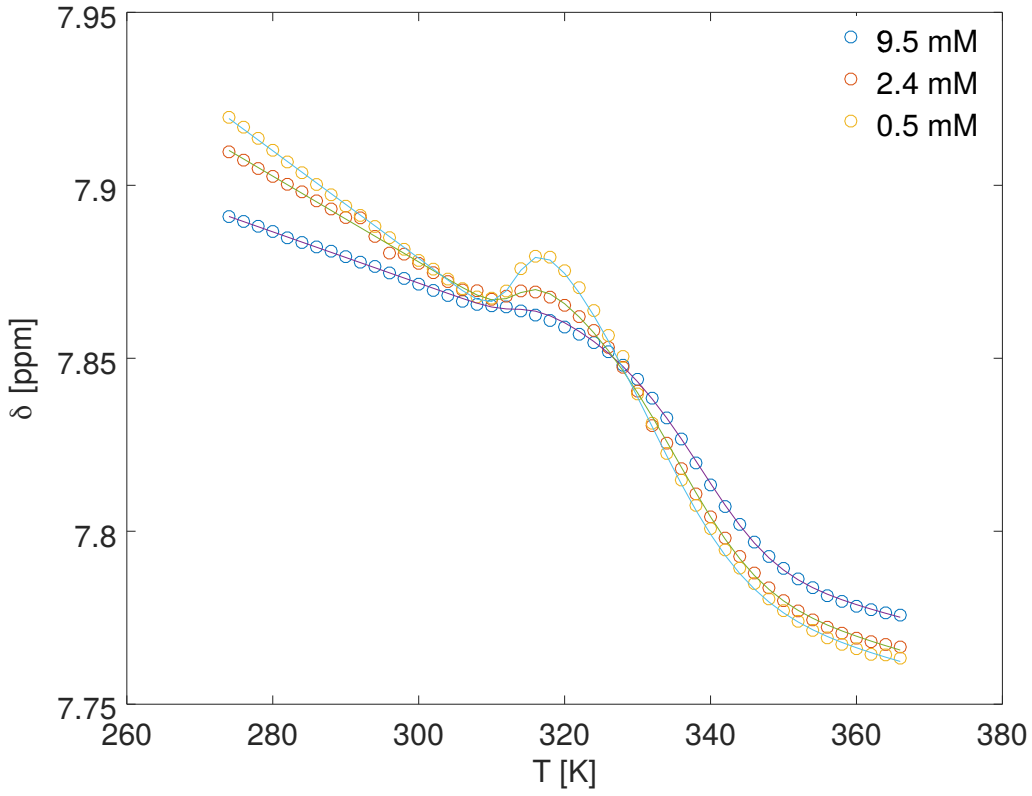


Figure 4.6: Chemical shift of G5H8 nuclei. Circles represent the values from Lorentzian fits of the NMR spectra; the lines are the three-state fits described by Eq. (1.28), fitting model B.

of the samples. From our results, the more likely scenario seems to be B with the intermediate state χ and equation (1.14).

It is highly unlikely that only the two nuclei do enter an intermediate state. The effect is probably present in all nuclei, and might be hard to discern because of a similar chemical shift to either a single strand or a duplex.

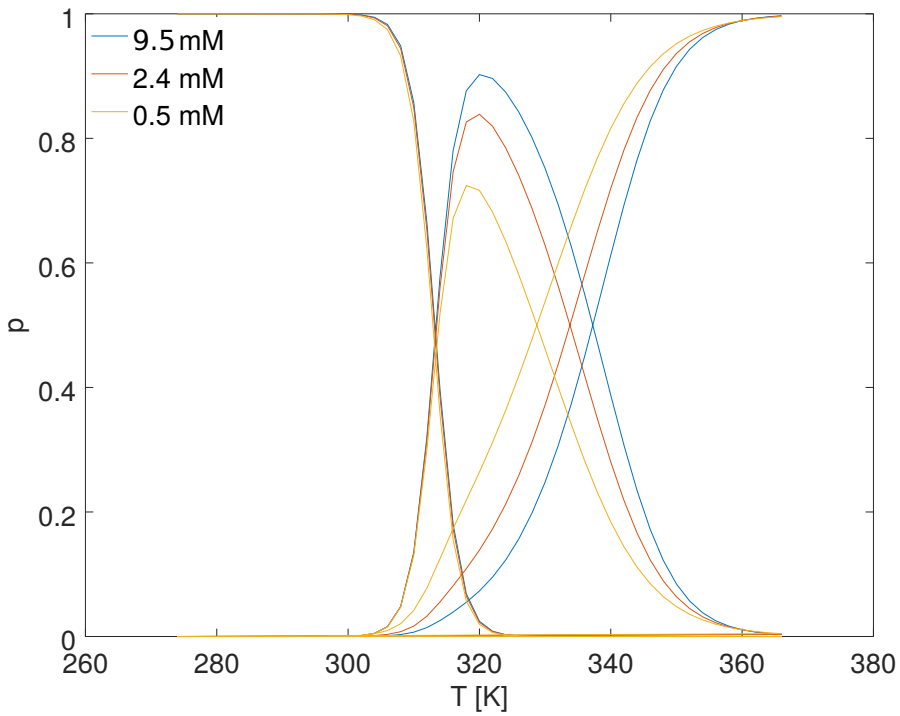


Figure 4.7: Populations of states by concentration for G5H8, results of fits by model B. Colours are the same for all states in a sample for clarity.

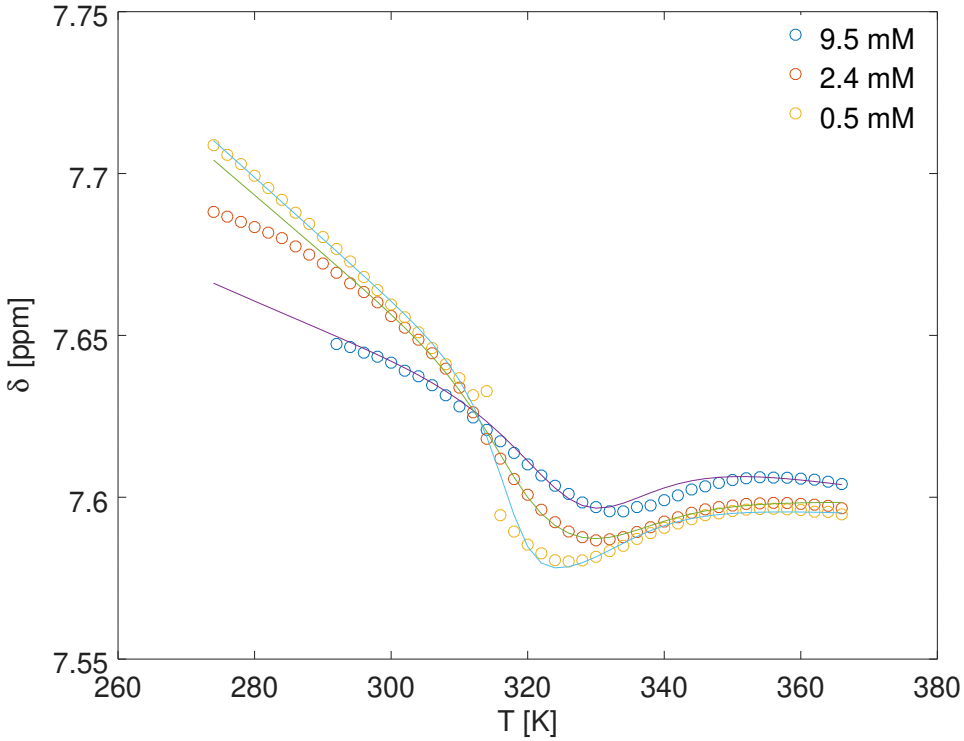


Figure 4.8: Chemical shift of T2H6 nuclei. Circles represent the values from Lorentzian fits of the NMR spectra; the lines are the three-state fits described by Eq. (1.28), fitting model A.

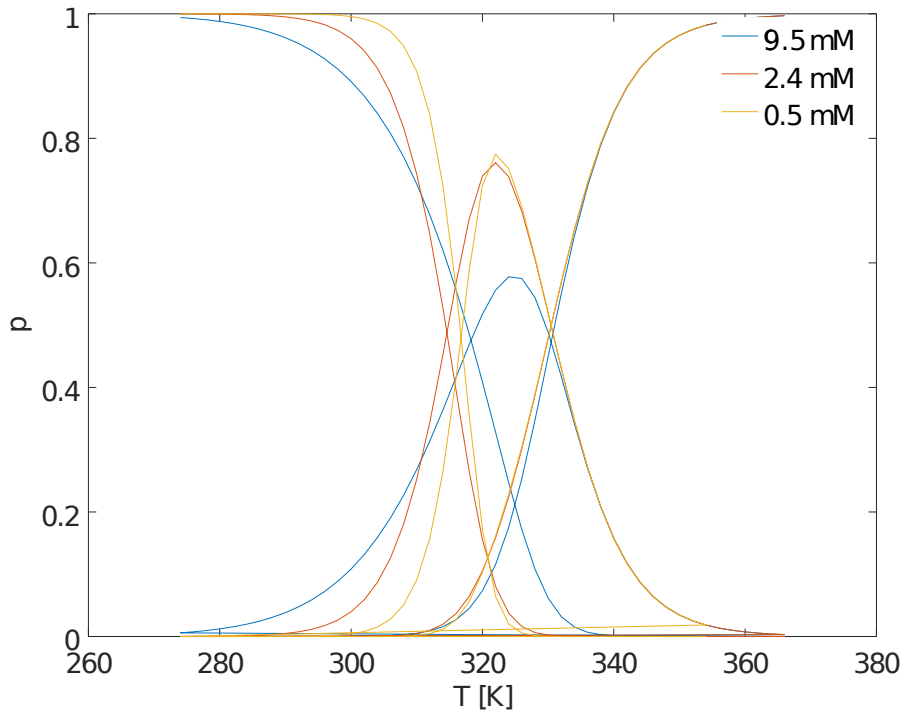


Figure 4.9: Populations of states by concentration for T2H6, results of fits by models A. Colours are the same for all states in a sample for clarity.

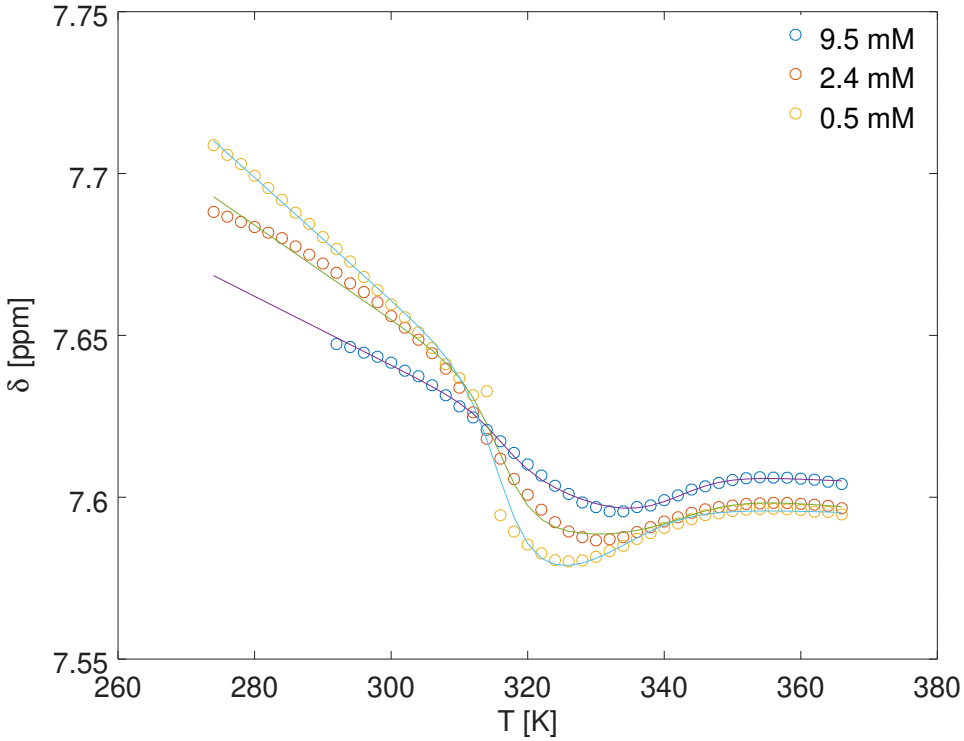


Figure 4.10: Chemical shift of T2H6 nuclei. Circles represent the values from Lorentzian fits of the NMR spectra; the lines are the three-state fits described by Eq. (1.28), fitting model B.

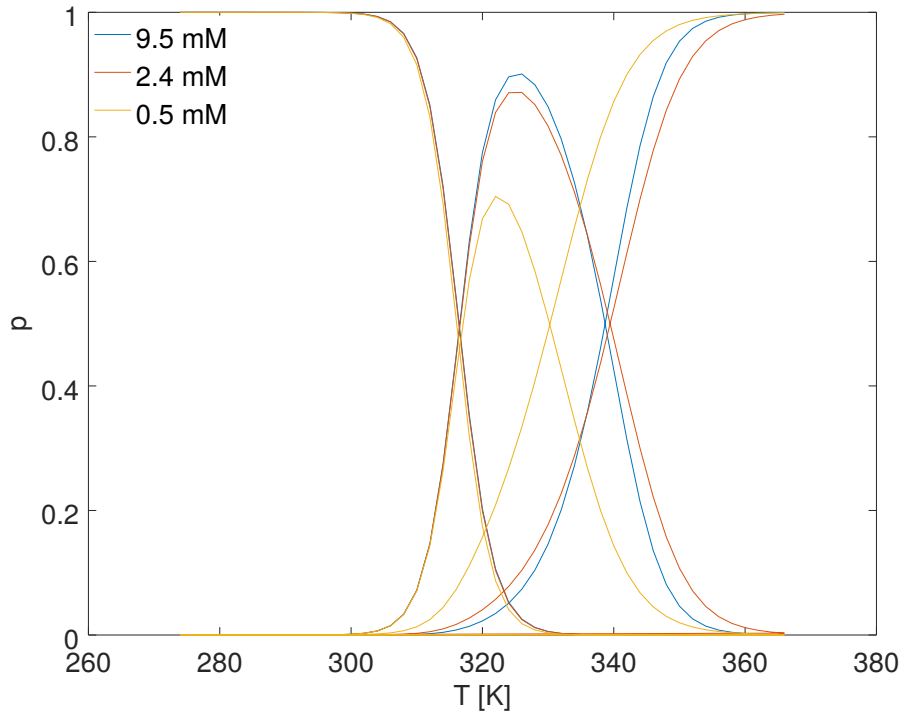


Figure 4.11: Populations of states by concentration for T2H6, results of fits by model B. Colours are the same for all states in a sample for clarity.

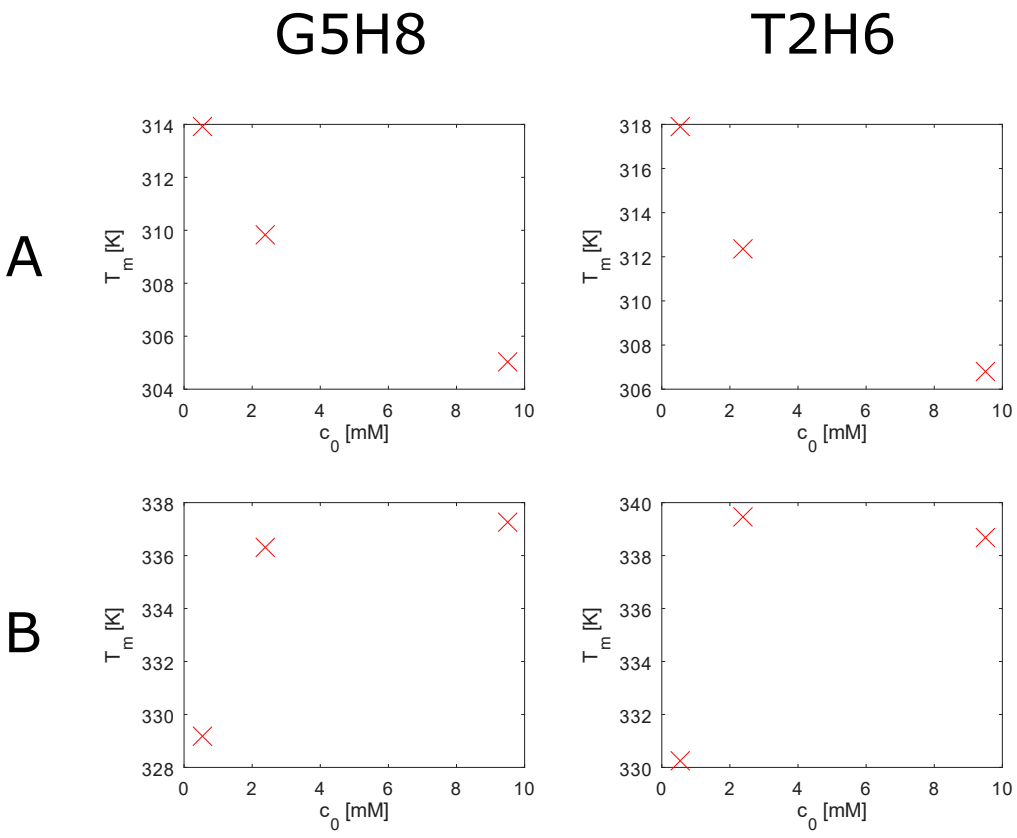


Figure 4.12: T_m against sample concentration for G5H8 and T2H6.

Conclusion

In this thesis, the transition state between double helix and single stranded forms of DNA was studied on a model self-complementary strand with the sequence CTTm^{5}CGAAG through NMR spectroscopy.

In most nuclei, only two structures, without an intermediate, were visible. The melting temperatures (T_m) for the two state spectra were between 310 and 335 K. The transition state was clearly observed in two peaks, G5H8 and T2H6. Their thermodynamic parameters were calculated for two possible processes, $D \rightleftharpoons 2\varphi \rightleftharpoons 2S$ and $D \rightleftharpoons \chi \rightleftharpoons 2S$. The second reaction described seems to be the more likely suspect, i.e. the double strand with increasing temperature first changes conformation, and then falls apart. The difference between the two melting temperatures for three-state systems was found to be 5-10 K.

Bibliography

[1] V. Římal. *NMR study of oligonucleotide structures*. Phd thesis, Charles University, 2018.

[2] L. Marcourt, C. Cordier, T. Couesnon, and G. Dodin. Impact of C5-cytosine methylation on the solution structure of d(GAAAACGTTTTC)2. an NMR and molecular modelling investigation. *European Journal of Biochemistry*, 265(3):1032–1042, 1999.

[3] K. Hoogsteen. The crystal and molecular structure of a hydrogen-bonded complex between 1-methylthymine and 9-methyladenine. *Acta Crystallographica*, 16(9):907–916, 9 1963.

[4] S. Lemieux and F. Major. RNA canonical and non-canonocal base pairing types: a cecognition method and complete repertoire. pages 4250–4263, 10 2002.

[5] V. Kumar, V. Kesavan, and K. V. Gothelf. Highly stable triple helix formation by homopyrimidine (l)-acyclic threoninol nucleic acids with single stranded DNA and RNA". *Organic & Biomolecular Chemistry*, 13(8):2366–2374, 2015.

[6] D. Sen and W. Gilbert. Formatiom of parallel four-stranded complexes by guanine-rich motifs in DNA and its implications for meiosis. *Nature*, 339:364–366, 1988.

[7] R. E. Franklin and R. G. Gosling. The structure of sodium thymonucleate fibres. I. The influence of water content. *Acta Crystallographica*, 6(8–9):673–677, 1953.

[8] J.D. Watson and F. H. C. Crick. Molecular structure of nucleic acids. *Nature*, 171:737–738, 1953.

[9] Bikash P. A., L. Haque, S. Bhuiya, and S. Das. Exploring the mode of binding of the bioflavonoid kaempferol with B and protonated forms of DNA by spectroscopic and molecular docking study. *RSC Advances*, 5(14):10219–10230, 01 2015.

[10] T. Mitsui, R. Langridge, B. E. Shortle, C. R. Cantor, R. C. Grant, K. Masahiko, and R. D. Wells. Physical and enzymatic studies on Poly d(I–C).Poly d(I–C), an unusual double-helical DNA. *Nature*, 228:1166–1169, 12 1970.

[11] A. Rich and S. Zhang. Z-DNA: the long road to biological function. *Nature Reviews Genetics*, 4:566–572, 2003.

[12] J. B. Lambert and E. P. Mazzola. *Nuclear Magnetic Resonance Spectroscopy: an introduction to principles, applications, and experimental methods*. Pearson Education, Inc., New Jersey, 2003.

[13] T. Hwang and A. Shaka. Water suppression that works. Excitation sculpting using arbitrary wave-forms and pulsed-field gradients. *Journal of Magnetic Resonance, Series A*, 2(112), 2 1995.

[14] V. Římal, H. Štěpánková, and J. Štěpánek. Analysis of NMR spectra in case of temperature-dependent chemical exchange between two unequally populated sites. *Concepts in magnetic resonance*, 38A(3):117–127, 2011.

List of Figures

List of Tables

List of Abbreviations

A adenine
C cytosine
DNA deoxyribonucleic acid
DSS 4,4-dimethyl-4-silapentane-1-sulfonic acid
G guanine
NMR nuclear magnetic resonance
T thymine
TMS tetramethylsilane

A. Attachments

	$\frac{\Delta H}{\mathrm{kJ \cdot mol^{-1}}}$	$\frac{\Delta S}{\mathrm{J \cdot mol^{-1} \cdot K^{-1}}}$	$\frac{\Delta G_{310\mathrm{K}}}{\mathrm{kJ \cdot mol^{-1}}}$	$\frac{T_\mathrm{m}}{\mathrm{K}}$
$c = 9.5$ mM				
C1H6	-109.5 ± 3.9	-362 ± 16	$+2.8 \pm 2.5$	318.4 ± 0.6
T2H6	-168.2 ± 10	-514 ± 20	-8.8 ± 0.1	314.9 ± 0.1
T3H6	-231.5 ± 2.6	-718 ± 9	-9 ± 2	330.9 ± 0.1
mC4H6	-238.6 ± 2.7	-735 ± 9	-10.7 ± 2	332.8 ± 0.1
G5H8	-306.2 ± 15.8	-926 ± 52	-19.2 ± 4.8	337.4 ± 0.4
A6H2	-252.1 ± 9.8	-786 ± 34	-8.4 ± 3.8	328.3 ± 0.2
A6H8	-231.8 ± 8	-731 ± 28	-5.3 ± 3.4	325.4 ± 0.4
A7H2	-229.6 ± 9.4	-718 ± 33	-7.1 ± 3.7	328.2 ± 0.3
A7H8	-238.8 ± 1.8	-739 ± 6	-9.6 ± 1.6	331.2 ± 0.1
G8H8	-205.7 ± 4.4	-645 ± 16	-5.8 ± 2.6	328.4 ± 0.2
$c = 2.4$ mM				
C1H6	-131.3 ± 5.2	-426 ± 21	$+0.7 \pm 2.9$	313.7 ± 0.5
T2H6	-253 ± 18.9	-771 ± 64	-14 ± 5.2	319.8 ± 0.1
T3H6	-209.2 ± 5.5	-652 ± 20	-7 ± 2.9	324.3 ± 0.2
mC4H6	-220.1 ± 4.1	-678 ± 14	-9.8 ± 2.5	327.9 ± 0.2
G5H8	-362.5 ± 53.6	-1087 ± 175	-25.5 ± 8.7	335.6 ± 0.8
A6H2	-418.4 ± 21.6	-1292 ± 70	-17.7 ± 5.5	325.5 ± 0.3
A6H8	-92.6 ± 6.4	-305 ± 26	$+2 \pm 3.2$	310.9 ± 2.6
A7H2	-428 ± 14.2	-1330 ± 47	-15.6 ± 4.5	323.5 ± 0.3
A7H8	-197.8 ± 4	-619 ± 14	-5.8 ± 2.4	323.2 ± 0.2
G8H8	-328.8 ± 20.7	-1023 ± 70	-11.6 ± 5.5	323.6 ± 0.3
$c = 0.54$ mM				
C1H6	-319.7 ± 25.1	-1012 ± 87	-6.1 ± 6.1	314.4 ± 0.4
T2H6	-433.9 ± 0.8	-1330 ± 3	-21.6 ± 1.1	321.5 ± 0.1
T3H6	-196.9 ± 8.3	-617 ± 30	-5.6 ± 3.5	316.4 ± 0.3
mC4H6	-200.9 ± 6.9	-615 ± 25	-10.1 ± 3.2	323.8 ± 0.4
G5H8	-453.1 ± 126.6	-1351 ± 407	-34.1 ± 13.4	334 ± 1.1
A6H2	-475.2 ± 0.4	-1474 ± 1	-18.3 ± 0.7	321.3 ± 0.1
A6H8	-24.6 ± 6.6	-97 ± 225	$+5.4 \pm 8.4$	241.5 ± 32.3
A7H2	-469.9 ± 8.3	-1458 ± 28	-18 ± 3.5	321.2 ± 0.3
A7H8	-227.5 ± 13.6	-723 ± 50	-3.4 ± 4.5	312.4 ± 0.5
G8H8	-449.6 ± 0.5	-1400 ± 2	-15.6 ± 0.8	319.9 ± 0.2

Table A.1: Thermodynamic parameters for two-state model

		G5H8			T2H6		
		9.5 mM	2.4 mM	0.5 mM	9.5 mM	2.4 mM	0.5 mM
A	$\Delta H_1/\mathrm{kJ \cdot mol^{-1}}$	−336	−490	−681	−156	−392	−479
	$\Delta S_1/\mathrm{J \cdot mol^{-1} \cdot K^{-1}}$	−1044	−1524	−2112	−450	−922	−1450
	$\Delta G_1^{310\mathrm{K}}/\mathrm{kJ \cdot mol^{-1}}$	−12.3	−17.5	−26.3	−16.3	−106.2	−29.7
	T_{m1}/K	305	310	314	307	312	318
	$\Delta H_2/\mathrm{kJ \cdot mol^{-1}}$	−164			−164		
	$\Delta S_2/\mathrm{J \cdot mol^{-1} \cdot K^{-1}}$	−437			−439		
	$\Delta G_2^{310\mathrm{K}}/\mathrm{kJ \cdot mol^{-1}}$	−28.6			−28.0		
	T_{m2}/K	330			332		
B	$\Delta H_1/\mathrm{kJ \cdot mol^{-1}}$	−450			−322		
	$\Delta S_1/\mathrm{J \cdot mol^{-1} \cdot K^{-1}}$	−1379			−959		
	$\Delta G_1^{310\mathrm{K}}/\mathrm{kJ \cdot mol^{-1}}$	−22.6			−24.4		
	T_{m1}/K	313			317		
	$\Delta H_2/\mathrm{kJ \cdot mol^{-1}}$	−230	−197	−163	−320	−253	−223
	$\Delta S_2/\mathrm{J \cdot mol^{-1} \cdot K^{-1}}$	−644	−537	−434	−905	−694	−613
	$\Delta G_2^{310\mathrm{K}}/\mathrm{kJ \cdot mol^{-1}}$	−30.6	−31.0	−28.9	−39.1	−37.5	−33.1
	T_{m2}/K	337	334	329	339	339	330

Table A.2: Thermodynamic parameters for 3-state systems

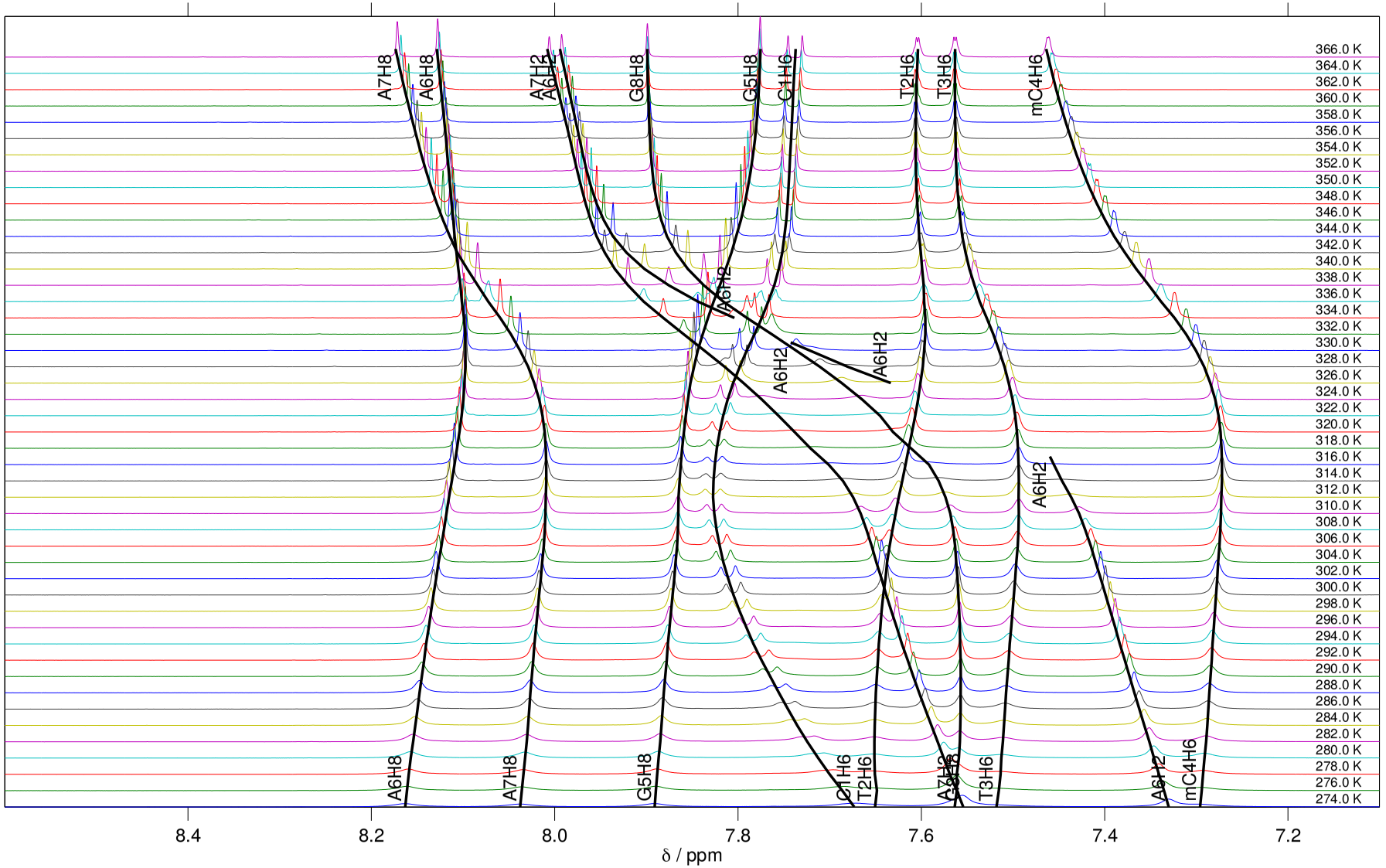


Figure A.1: zgesgp spectra for 9.5 mM sample

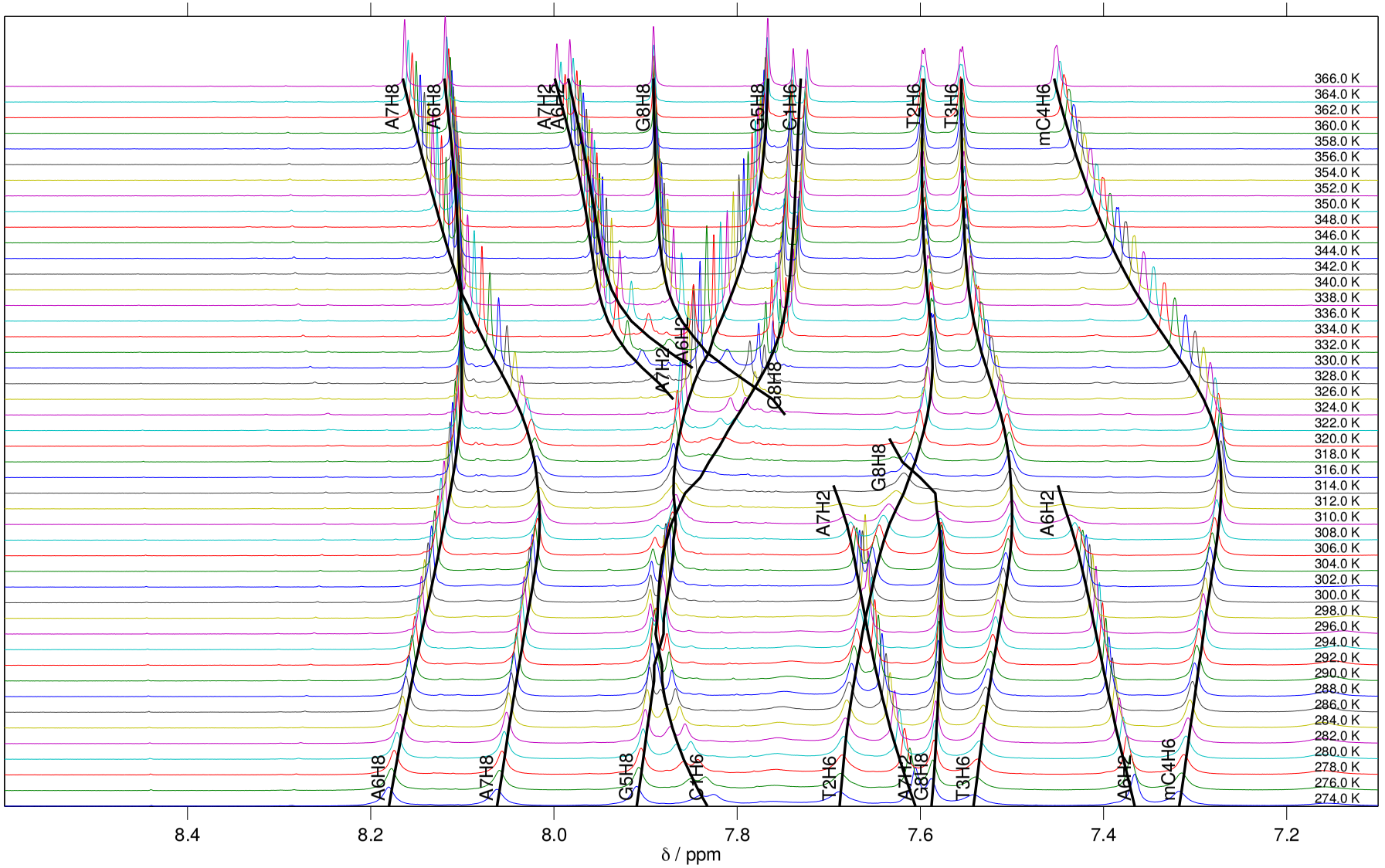


Figure A.2: zgesgp spectra for 2.4 mM sample

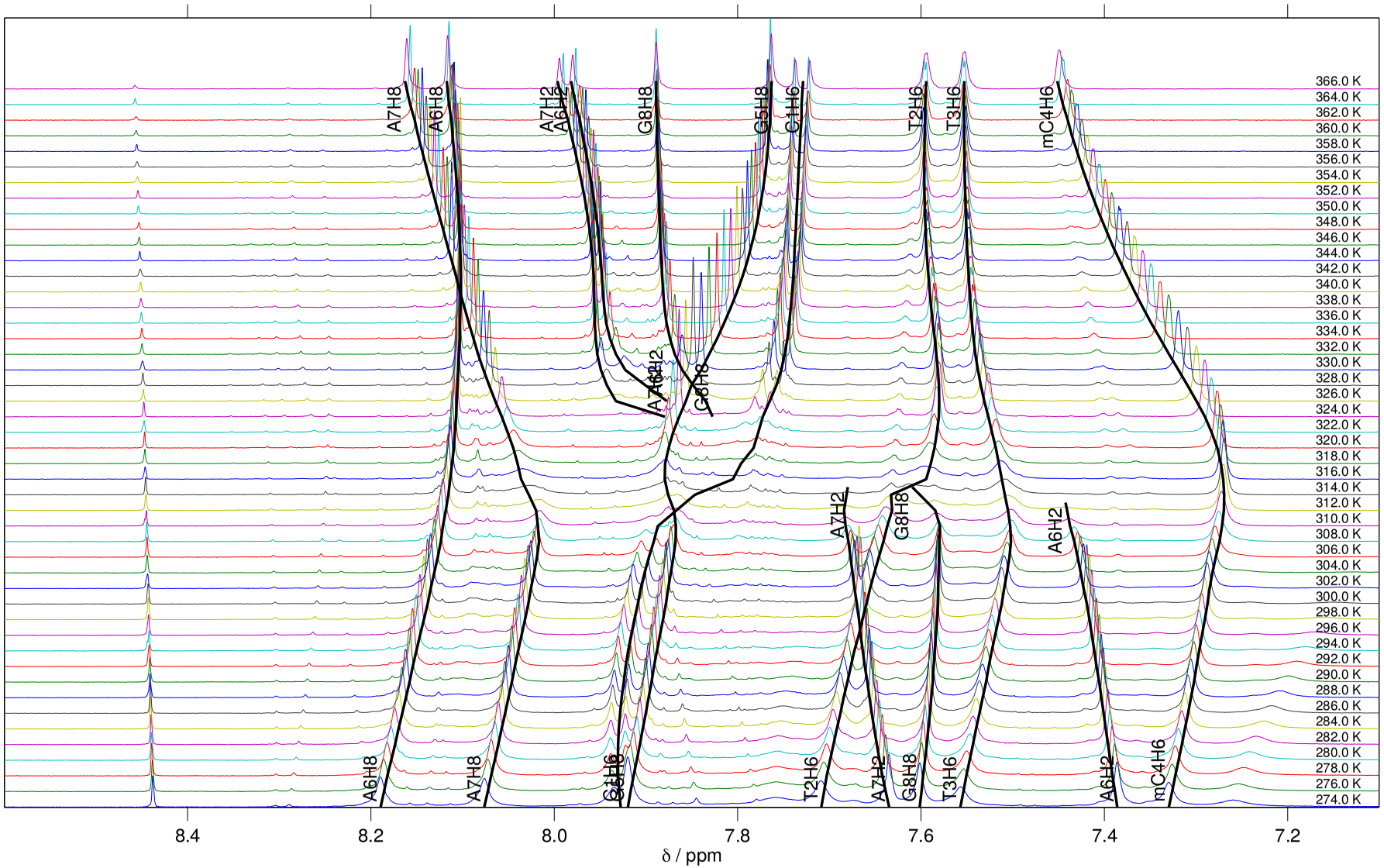


Figure A.3: zgesgp spectra for 0.54 mM sample