

CHARLES UNIVERSITY

Faculty of Pharmacy

---

Course of Study: Pharmacy

Subject: Pharmacy



Jakub Radek Štoček

INTERACTIONS OF MODIFIED NUCLEOBASES STUDIED  
BY NMR SPECTROSCOPY

Studium interakcí mezi modifikovanými nukleobazemi pomocí  
NMR spektroskopie

Master Thesis

Supervisors: doc. PharmDr. Jiří Kuneš, CSc.

doc. RNDr. Martin Dračínský, Ph.D.

Prague 2020



# Prohlášení

Prohlašuji, že jsem tuto závěrečnou práci zpracoval samostatně a že jsem uvedl všechny použité informační zdroje a literaturu. Tato práce ani její podstatná část nebyla předložena k získání jiného nebo stejného akademického titulu.

V Praze 20. dubna 2020

# Abstract (EN)

Nucleic acids are among the largest and most complex compounds in nature and without a doubt the most important molecules for the existence of life. Understanding the chemical behaviour and thermodynamic parameters of their building components enables us to make ever so increasingly accurate predictions of nucleic acid structures, reactions and functions. These predictions in turn provide great help in research fields such as DNA nanotechnology, nucleic acid targeted drug development and DNA material sciences. My research work targeting nucleobases and their analogues is focused on exploring effects of their conformational isomers and tautomers on the thermodynamics of intermolecular complexes that these compounds can theoretically form with each other through hydrogen bonds.

Doc. RNDr. Martin Dračinský, Ph.D. and I have conducted two separate studies concerning this topic at the NMR department of the Institute of Organic Chemistry and Biochemistry of the Czech Academy of Sciences. The first research project concentrated on developing a method for the determination of nucleobase-pairing free energies from rotamer equilibria of various 2-(methylamino)pyrimidines [1], while the second project aimed to shed light on the tautomerism of guanine analogues [2].

## Keywords

Nuclear Magnetic Resonance (NMR) Spectroscopy, Rotational Isomers, Tautomerism, Complex Formation

# Abstrakt (CZ)

Nukleové kyseliny jsou jedny z největších a nejsložitějších sloučenin vyskytujících se v přírodě a nepochybně představují nejdůležitější molekuly pro existenci života. Pochopení chemického chování a termodynamických parametrů stavebních kamenů nukleových kyselin nám umožňuje činit stále přesnější predikce jejich struktury, reakcí a funkcí. Tyto predikce usnadňují výzkum v oborech jako je DNA nanotechnologie, vývoj léčiv zaměřujících nukleové kyseliny a materiálové vědy studující DNA. Moje práce týkající se nukleobází a jejich analog je zaměřena na studium efektů jejich konformačních izomerů a tautomerů na termodynamiku mezimolekulových komplexů, které tyto sloučeniny mohou teoreticky navzájem tvořit pomocí vodíkových vazeb.

Doc. RNDr. Martin Dračínský, Ph.D. a já jsme uskutečnili dva oddělené výzkumy týkající se těchto témat na Ústavu Organické Chemie a Biochemie Akademie Věd České Republiky. První z těchto projektů se zaměřil na vývoj metody pro určení volné Gibbsovy energie párování nukleobází na základě rotamerních rovnováh různých derivátů 2-(methylamino)pyrimidinu [1] a druhý projekt se věnuje tautomerii analog guaninu [2].

## Klíčová Slova

Spektroskopie Nukleární Magnetické Rezonance (NMR), Rotamery, Tautomery, Tvorba Komplexů

# Acknowledgements

My deepest thanks go to doc. RNDr. Martin Dračínský, Ph.D. and doc. PharmDr. Jiří Kuneš, CSc. for their guidance, support, patience, helpfulness, constructive criticism and educational support. I also owe a great deal of gratitude to Ing. Ján Tarábek, Ph.D. for showing me the beauty of writing in L<sup>A</sup>T<sub>E</sub>X and RNDr. Vladimír Vít for initially igniting my enthusiasm for science. Finally, I would like to thank Bc. Zuzana Osifová and Mgr. Jana Maříková for being wonderful and supportive colleagues.

# Contents

<b>Abstract (EN)</b>	<b>ii</b>
<b>Abstrakt (CZ)</b>	<b>iii</b>
<b>Acknowledgements</b>	<b>iv</b>
<b>1 Nuclear Magnetic Resonance</b>	<b>1</b>
1.1 NMR Spectroscopy as a Method . . . . .	1
1.2 The Principles of NMR . . . . .	1
1.2.1 Particles . . . . .	1
1.2.2 Spin . . . . .	2
1.2.3 Magnetogyric Ratio . . . . .	2
1.2.4 Basics of NMR Spectroscopy . . . . .	3
1.2.5 Isotopes and Nuclear Spin Quantum Number . . . . .	4
1.2.6 Shielding . . . . .	4
1.2.7 Acquisition of Spectra . . . . .	5
1.2.8 Chemical Shift . . . . .	7
1.3 Dynamic NMR . . . . .	7
1.3.1 Basics of dNMR . . . . .	7
1.3.2 Variable Temperature dNMR . . . . .	8
1.3.3 Chemical Exchange . . . . .	9
<b>2 Computational Chemistry</b>	<b>11</b>
2.1 The Importance of Quantum Chemistry . . . . .	11
2.2 Practical Aspects of Computational Chemistry . . . . .	11
2.3 Theory Behind Used Experiments . . . . .	12
2.3.1 Mathematical Model . . . . .	13
2.3.2 Basis Set . . . . .	13

---

2.3.3	Geometry Optimization . . . . .	13
2.3.4	Reaction Field . . . . .	14
2.3.5	Dispersion . . . . .	14
<b>3</b>	<b>The Hydrogen Bond</b>	<b>15</b>
3.1	Definition of the Hydrogen Bond . . . . .	15
3.2	The Importance of Hydrogen Bonds in Biology . . . . .	15
3.3	NMR Spectroscopy of the Hydrogen Bond . . . . .	16
3.4	Hydrogen Bonding and Quantum Chemistry . . . . .	17
<b>4</b>	<b>Conformational Isomerism and Tautomerism</b>	<b>18</b>
4.1	Conformers . . . . .	18
4.2	Tautomers . . . . .	19
<b>5</b>	<b>Nucleobase-Pairing Free Energies from Rotameric Equilibria</b>	<b>21</b>
5.1	Introduction . . . . .	21
5.2	Mathematical Aparatus . . . . .	23
5.3	DFT Calculations . . . . .	24
5.4	Experimental Results and Discussion . . . . .	24
<b>6</b>	<b>Tautomerism of Guanine Analogues</b>	<b>31</b>
6.1	Introduction . . . . .	31
6.2	DFT Calculations . . . . .	32
6.3	Experimental Results and Discussion . . . . .	36
<b>7</b>	<b>Technical and Experimental Data</b>	<b>40</b>
7.1	Hardware . . . . .	40
7.2	Software . . . . .	40
7.3	Chemicals . . . . .	41
7.4	Sample Preparation . . . . .	41
7.5	Calculations . . . . .	41
	<b>Bibliography</b>	<b>42</b>



# Chapter 1

## Nuclear Magnetic Resonance

### 1.1 NMR Spectroscopy as a Method

Nuclear Magnetic Resonance (NMR) spectroscopy is a powerful analytical method with a wide span of possible applications. The main areas of interest in modern science range anywhere from total or partial structural elucidation of unknown substances, structure confirmation, monitoring of reaction mixtures, studying physical processes, to uses such as 3D studies of materials through NMR imaging, which also established an irreplaceable position in medical diagnostics. To understand and use the potential that lies within this method, we first need to discuss the phenomenon of nuclear magnetic resonance itself, although before we delve any deeper, it is necessary to address the basics it builds on. NMR was first discovered by I. I. Rabi in 1938 on isolated nuclei, followed by E. M. Purcell in 1946 on a solid sample and F. Bloch et al. on a liquid the same year [3–5]. This section relies on texts [6–9].

### 1.2 The Principles of NMR

#### 1.2.1 Particles

Particles that embody the universe we exist within possess a set of intrinsic properties that define them among others and are of great influence on both their macroscopic and microscopic behaviour, the latter being of utmost importance concerning NMR. The composite particles that interest us, as the name of the method implies, are the atomic nuclei. Each atomic nucleus has four important physical properties, namely mass, electric charge, magnetism and spin [6]. The nucleus is always composed of at least one proton

and zero or more neutrons. The property of concern for NMR is the nuclear spin, which arises from and depends on total proton and neutron count and spin orientation in the ground energy state of the nucleus [6].

### 1.2.2 Spin

Spin is a difficult concept to grasp, with any remotely tangible description being inaccurate at best, although it might be rather vaguely stated that spin is a property of a particle that makes it behave as if it was spinning around, hence the name, even though in reality it is not. In other words, spin is a form of angular momentum that does not arise from actual rotation. Luckily for us, quantum mechanics enable us to understand the behaviour of spin and all sorts of possible operations with it. NMR spectroscopy is all about manipulating spins in a tightly controlled and planned manner, resulting in acquiring valuable data about the atomic nuclei and their surroundings. Each particle possessing spin has a spin value given as a multiple of  $\frac{1}{2}$ , which implies that spin for a given particle can either be 0, a half integer or an integer. The angular momentum provided by spin in combination with positive electric charge coming from the up quarks in protons give rise to magnetic moment ( $\mu$ ) [6]. This, along with nuclear spin number ( $I$ ) is used to describe a very important property of a given nucleus, the magnetogyric ratio ( $\gamma$ ), as displayed by eq. (1.1) [8].

$$\gamma = \frac{2\pi\mu}{hI} \quad (1.1)$$

where  $h$  is the Planck constant.

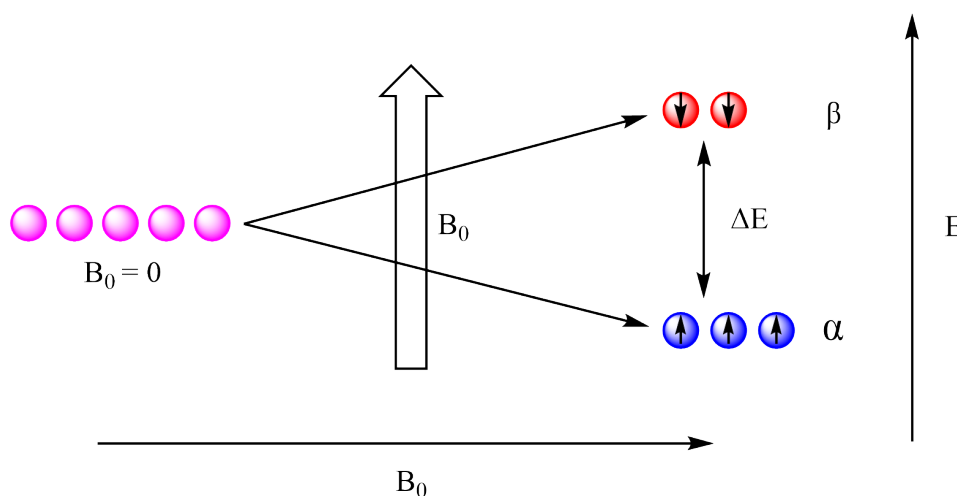
### 1.2.3 Magnetogyric Ratio

The magnetogyric ratio  $\gamma$  is a constant that has a unique value for every distinct isotope, and in practice, describes how strongly magnetic a particular nuclide is. When a nucleus with  $\mu$  is put in an external static field ( $\mathbf{B}_0$ ), it precesses around the field axis with a frequency ( $\nu$ ) depending on the field strength. This motion is called the Larmor precession ( $\omega_0$ ) and is described by eq. (1.2) [6].

$$\omega_0 = \frac{-\gamma\mathbf{B}_0}{2\pi} \text{ [Hz]} \quad (1.2)$$

## 1.2.4 Basics of NMR Spectroscopy

For a nucleus to be observable by NMR spectroscopy, it is required that it possesses a non-zero nuclear spin value  $I$ , arising from the mutual proton and neutron spin configurations in the ground energy state. Nuclei with spin greater than zero possess  $(2I + 1)$  energy sublevels, all of which are degenerate in the absence of an external electric or magnetic field. The splitting between these levels is called the nuclear Zeeman splitting, shown in fig. 1.1.



**Figure 1.1:** A simple illustration of nuclear Zeeman splitting. Note the slight spin population excess in the lower energy state ( $\alpha$ ) which is aligned with the direction of applied external field  $B_0$  (denoted by the large arrow). The energy difference ( $\Delta E$ ) between spin states  $\alpha$  and  $\beta$  is directly proportional to magnetic field strength.

In the case of an external field  $B_0$  being applied, the degeneracy breaks, giving rise to distinct energy sublevels [6]. The spin population of these states at equilibrium is not equal, with a very slight majority of spins being in the lower energy state ( $N_\alpha$ ), having the same  $\mu$  orientation as  $B_0$ . This equilibrium is described by the Boltzmann distribution in eq. (1.3).

$$\frac{N_\alpha}{N_\beta} = e^{\frac{-\Delta E}{k_B T}} \quad (1.3)$$

Where  $k_B$  is the Boltzmann constant and  $T$  is temperature.

Simply speaking, by applying electromagnetic radiation in the radiofrequency region ( $R_f$ ) with a frequency matching  $\omega$ , this equilibrium can be broken by energy absorption and some of the spins can be raised to a higher energy state ( $N_\beta$ ). The energy difference

( $\Delta E$ ) is described by eq. (1.4) [9]. The return back to equilibrium after absorption is associated with rise of the NMR signal, which can be recorded as a decaying interferogram, also known as the free induction decay (FID) which becomes a conventional spectrum after transformation from time domain to frequency domain by the Fourier transformation [8].

$$\Delta E = h\nu = \frac{h\gamma\mathbf{B}_0}{2\pi} \quad (1.4)$$

### 1.2.5 Isotopes and Nuclear Spin Quantum Number

Nuclei with  $I = \frac{1}{2}$  are best suited for NMR spectroscopy since they are spherical in shape and possess just two Zeeman sublevels in accordance to the  $(2I+1)$  rule. Nuclei with  $I > \frac{1}{2}$  are quadrupolar nuclei with a non-spherical shape, which leads to many technical complications and their measurement is not nearly as common as the  $I = \frac{1}{2}$  nuclei. The by far most important quadrupolar nucleus is  $^2\text{H}$ , alternatively D for deuterium with  $I = 1$ , which replaces  $^1\text{H}$  in liquid state NMR solvents to reduce the intensity of solvent signals and serve as the field stability lock anchor. Nuclei with  $I = 0$  are unfortunately not detectable by NMR spectroscopy and are thus often dubbed as NMR silent. While not many isotopes actually have  $I = 0$ , fate had it that almost all the important and abundant isotopes in organic chemistry except  $^1\text{H}$ ,  $^{19}\text{F}$  and  $^{31}\text{P}$  are NMR silent. This is not as much of a setback as it would initially seem, as isotopes  $^{13}\text{C}$ ,  $^{15}\text{N}$ ,  $^{17}\text{O}$  and others are stable and albeit not naturally very abundant, can be readily measured, just with much lower sensitivity [8].

### 1.2.6 Shielding

The reason of the massive success of NMR spectroscopy as an analytical tool lies in the concept that while all distinct NMR active isotopes share the same  $\gamma$  between their own kind, the precise  $\omega$  of a nucleus strongly depends on its molecular environment. The magnetic field observed for each of the nuclei in a molecule can be altered in several different ways, the most intense effect being caused by the immediate electron cloud. Electrons themselves have spin  $\frac{1}{2}$  and a negative charge, and as such they also exhibit magnetic moment  $\mu$ , but due to the opposite nature of their charge compared to the atomic nuclei, they are aligned opposite to  $\mathbf{B}_0$  and thus exhibit a phenomenon we call shielding. The effect of shielding on the Larmor precession frequency  $\omega$  is explained with eq. (1.5). Other effects include direct dipole-dipole interactions, which can be exploited

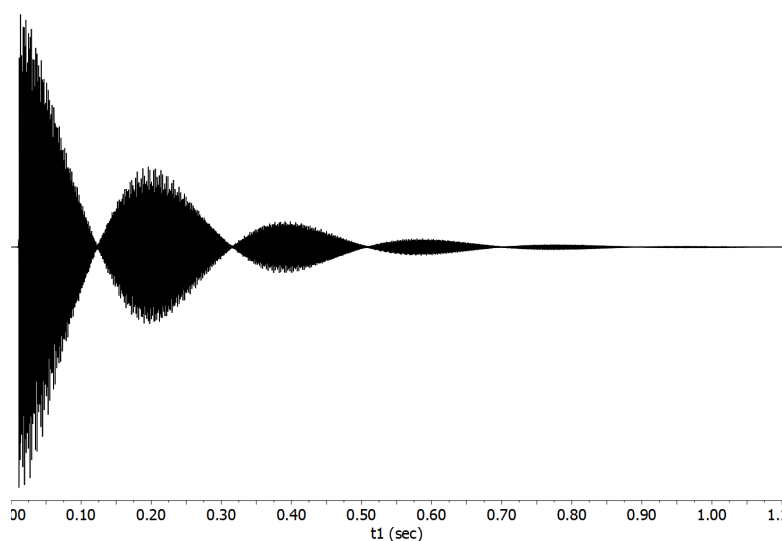
for gaining 3D spatial information about the molecule and of course the indirect spin coupling, which can be described as a through-bond interaction of nuclear spins caused by hyperfine interactions of nuclei and their surrounding electrons. This phenomenon, known as the J coupling, is tremendously helpful when assigning a spectrum, as it splits the signals in accordance to their molecular surroundings [6, 8].

$$\omega = \frac{\gamma}{2\pi} \mathbf{B}_0 (1 - \sigma) \quad (1.5)$$

Where  $\sigma$  is the shielding constant.

### 1.2.7 Acquisition of Spectra

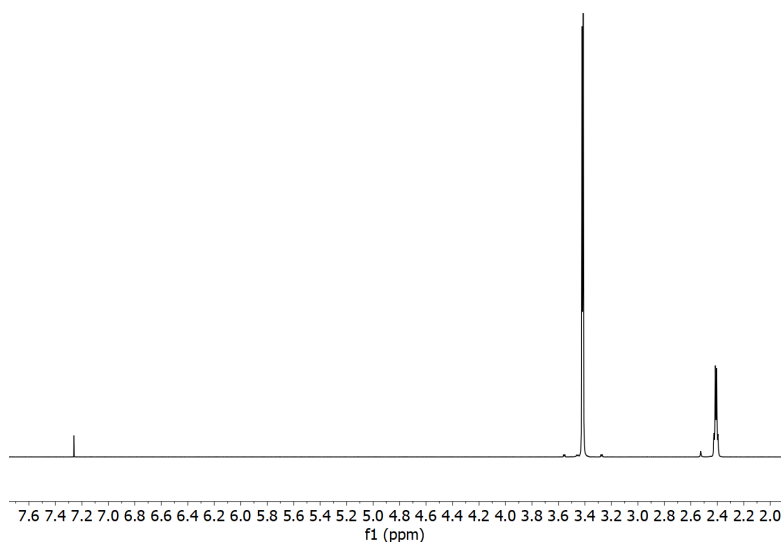
Before the transformed spectrum is addressed, I find it necessary to roughly explain the details of FID acquisition.



**Figure 1.2:** The  $^1\text{H}$  free induction decay (FID) of methanol in  $\text{CDCl}_3$  is a recording of all  $^1\text{H}$  spins precessing at their specific Larmor frequencies after a pulse. It provides good visual indication of spin relaxation and is useful for understanding data apodization, although the structural information is not readily available.

The normal sample for standard liquid state NMR usually consists of a few milligrams of a compound dissolved in 500-600  $\mu\text{l}$  of deuterated solvent, enclosed by a polyethylene cap within a 5mm outer diameter high purity borosilicate glass sample tube. It is clear that there is a large amount of spins present in the total sample volume. When the sample tube is placed within a probe mounted inside a NMR spectrometer, the total net magnetization vector ( $\mathbf{M}_0$ ) aligns with the external magnetic field  $\mathbf{B}_0$  in the Cartesian frame, parallel

with the  $z$ -axis by convention. The  $\mathbf{M}_0$  is composed of all the individual populations of equivalent spin magnetic moments  $\boldsymbol{\mu}$  precessing with a Larmor frequency  $\omega$  in compliance with eq. (1.5) [8]. It is immediately apparent that the role of  $\mathbf{B}_0$  within eq. (1.2) places an extremely high demand on overall field homogeneity throughout the probe's acquisition zone in the sample. Modern superconducting high field instruments can handle this issue with little to no trouble. Once the sample is set in place, the  $R_f$  transmitter coil located perpendicular to the sample can apply a short lasting ( $t_p$ ) electromagnetic pulse with a frequency  $\nu$  matching  $\omega_0$  of a phase ( $\phi$ ) generating magnetic field ( $\mathbf{B}_1$ ) along an axis in the  $x$ - $y$  plane, which tips the  $\mathbf{M}_0$  by nutation angle ( $\theta$ ) described by eq. (1.6) in the Cartesian frame. Right after the pulse, the magnetization vector ( $\mathbf{M}$ ) of each equivalent precessing spin population is now undergoing a process called relaxation, which is essentially a return back to equilibrium. The oscillating  $\mathbf{M}$  induces an electric current which can be detected along the  $y$ -axis by the receiver coil located in the same plane as the  $R_f$  coil. The relaxation process is responsible for signal decay and can be described as a return of net magnetization to equilibrium by energy transfer ( $T_1$ ) and net magnetization dephasing ( $T_2$ ). The overall FID is a recording of all the contributing  $\mathbf{M}$  precessing at  $\omega$  (see fig. 1.2). The FID itself is impossible to read in practical situations, so the data are made far easier to interpret by converting it to a frequency domain by fourier transformation (FT) where all the contributing Larmor frequencies  $\omega$  can be found as more or less sharp signals along a scale, as seen in fig. 1.3 [9].



**Figure 1.3:** The  $^1\text{H}$  spectrum of methanol in  $\text{CDCl}_3$  obtained by fourier transformation of the FID in fig. 1.2

---

A great advantage provided by the pulsed FT approach is the ability to add up FIDs from each acquisition to greatly improve the signal-to-noise ratio by eliminating inconsistent artifacts [8].

$$\theta = \frac{\gamma \mathbf{B}_1 t_p}{2\pi} [^\circ] \quad (1.6)$$

## 1.2.8 Chemical Shift

Now that it is underlined how a spectrum is created, we can look upon some practical aspects. As eq. (1.2) suggests, the precise Larmor frequency  $\omega$  is proportionally depending on the used field strength  $\mathbf{B}_0$ . This causes interpretation problems in practice as many different spectrometers of different field strengths are used. The concept of chemical shift ( $\delta$ ) was created to make the spectra independent of this effect by making them dimensionless. This is achieved by relating the signal Larmor precession frequency  $\omega$  to  $\omega_{ref}$  of a reference compound as portrayed by eq. (1.7). This reference compound is usually tetramethylsilane (TMS). The units of  $\delta$  are dimensionless and written as parts per million (ppm) [8].

$$\delta = \frac{\omega - \omega_{ref}}{\omega_{ref}} \quad (1.7)$$

The chemical shift scale is traditionally increasing from right to left, with 0 defined by  $\omega_{ref}$ . The area under signals is quantifiable by integration with good accuracy, given that adequate relaxation delay was implemented between pulses [9].

## 1.3 Dynamic NMR

### 1.3.1 Basics of dNMR

NMR spectroscopy stands as an excellent tool for dynamic process studies and often serves as the best available method for observation of molecular motion, chemical exchange and non-covalent interactions. The main limitation of dynamic NMR (dNMR) comes from an intrinsic attribute of conventional NMR spectroscopy itself, the so called NMR timescale. Please note that there are other NMR methods to measure dynamic systems with rate constants outside of further discussed region, usually involving relaxation or the use of other nuclei, but they are omitted for the purposes of this thesis. If the frequency of a process, called the rate constant ( $k$ ), has a value  $k > 10^4 \text{ s}^{-1}$ , it is considered too

fast for dNMR as the lifetime ( $\tau = \frac{1}{k}$ ) of each state is too short ( $\tau < 0.1$  ms) to make an impact on the FID. On the other hand, should the process have  $k < 1$  s<sup>-1</sup>, the lifetime of each state ( $\tau > 1$  s) is so long that the FID acquisition happens faster than a single iteration and the system is thus seen as static. Any process between these extremes ( $1 > k > 10^{-4}$  s<sup>-1</sup>) can be subjected to dNMR analysis. Note that the term frequency is used rather than speed as the exchange itself is extremely fast regardless of  $k$ , which actually describes how often it happens. The rate constant, among other valuable data, can be extracted from a dynamic system through computational simulation of spectra, often achieved by a method called lineshape fitting, followed by complete spectral lineshape analysis. My personal choice for this task is the Bruker TopSpin dnmr utility. There are many dynamic processes suitable for dNMR, including but not limited to: restricted bond rotation, tautomerism, intermolecular exchange, hydrogen bonding, non-covalent interactions, structural rearrangement, ring inversions, chiral centre inversion and fluxionality. Dynamic parameters exceeding <sup>1</sup>H NMR timescale can be studied through more challenging magnetization transfer experiments or <sup>13</sup>C dNMR. This section relies on text [9].

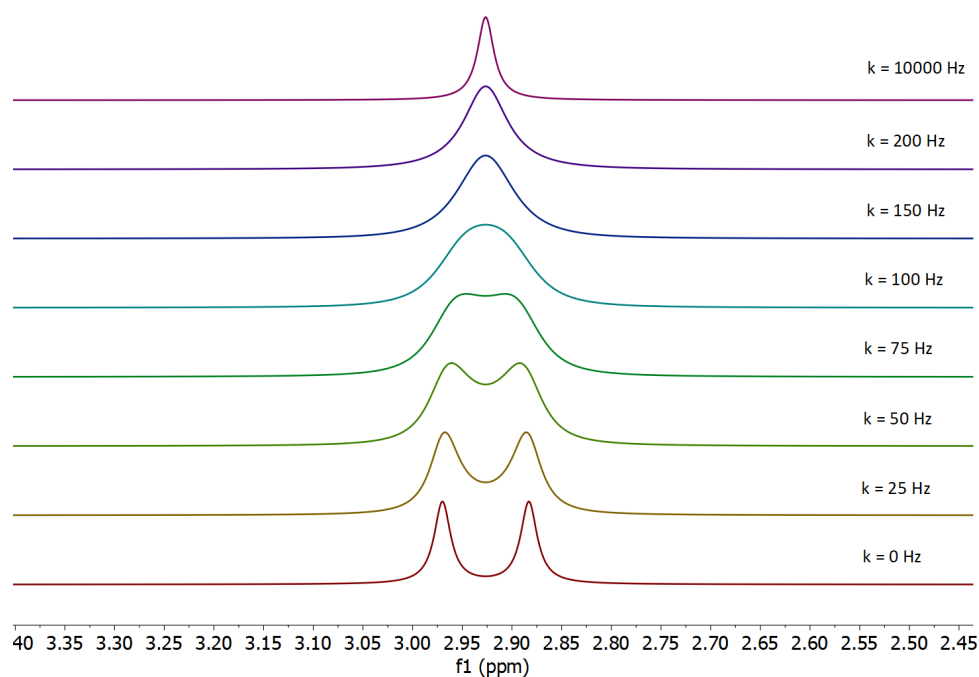
### 1.3.2 Variable Temperature dNMR

A great example of dynamic process observation is the restricted tertiary amide bond rotation in *N,N*-dimethylformamide.

While both *N*-methyl groups are virtually indistinguishable, they experience different magnetic environment in respect to their alignment being either *cis* or *trans* to the carbonyl oxygen. The fact that we observe a sharp signal for each state at room temperature in real sample indicates that the rate constant for the rotation around the amide bond is sufficiently slow, given by its high rotational energy barrier. As the temperature rises, the signals are progressively broadening and shifting towards each other (see fig. 1.4), until they finally merge in a coalescence point at  $k = 75$  Hz, from which the signal only gets narrower until only one sharp averaged peak is observed at values of  $k > 1000$  Hz. The broadening effect occurs due to relaxation (especially  $T_2$ ) being accelerated by random frequency skips of the involved spins, which lead to faster net dephasing of  $\mathbf{M}_0$ . This effect is the strongest at coalescence point and fades in intensity as the rate constant increases. The observed spectral broadening and signal averaging is strongly dependent on used magnetic field strength in accordance to eq. (1.5), lower field strength means smaller frequency difference, which leads to signal coalescence being observed at a lower temperature. One should also be aware of the fact that some signals, especially those of exchangeable protons,



exhibit thermal shift in addition to the shift induced by a dynamic process, which causes inaccurate readings and calculations. An ability to precisely control sample temperature during experiments is a necessity and requires possession of a calibrated probe able to withstand and operate under a wide thermal range. Cooling to sub-zero temperatures is usually achieved by feeding evaporated nitrogen from a Dewar flask directly to the probe through a specialised setup.

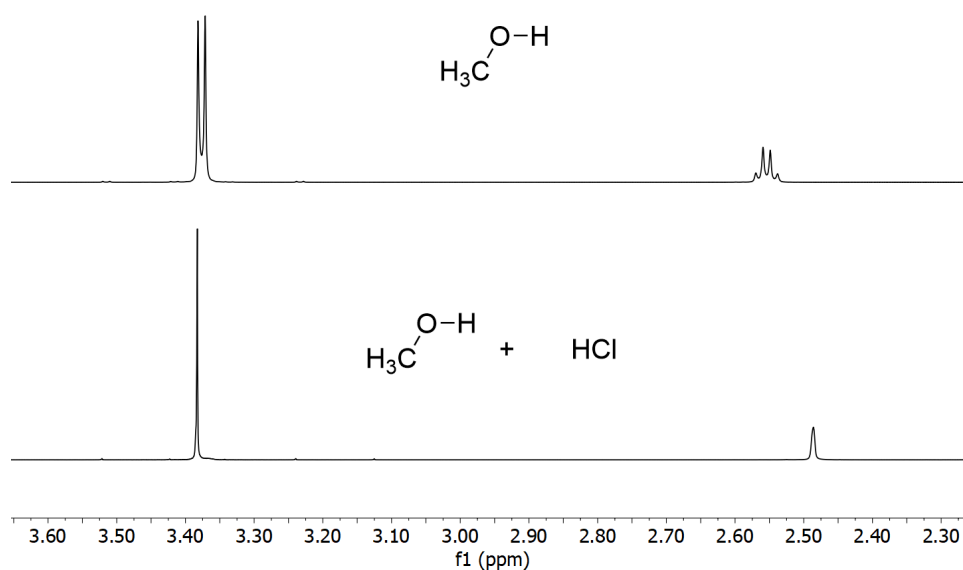


**Figure 1.4:** Simulated  $^1\text{H}$  spectra of aliphatic region of *N,N*-dimethylformamide with various exchange rate constants depicting an array of variable temperature experiments where the sample would be progressively heated.

### 1.3.3 Chemical Exchange

Factors other than temperature can also significantly alter the dynamic parameters of a system. This is especially true for intermolecular phenomena such as hydrogen bonding, hydrogen exchange, complex formations and other non-covalent interactions. Hydrogen exchange between highly polarised functional groups is a great model process for intermolecular dNMR. Basic instinct tells us that exchangeable hydrogens present on functional groups are normal parts of a molecule and should as such exhibit standard J coupling behaviour with neighbouring nuclei. In practical situations however, they are often either not seen at all or simply as broad signals bearing no coupling structure,

complemented by the fact that the supposed coupling partners rarely show anything beyond almost invisible broadening caused by residual interaction. A practical example of this concept can be seen in fig. 1.5.



**Figure 1.5:** Trace amounts of  $\text{HCl}$  cause exchange decoupling of otherwise clearly coupled  $^1\text{H}$  signals of methanol. Sample measured in  $\text{CDCl}_3$

Signal disappearance is usually an aftermath of choosing a protic solvent that, if deuterated, readily displaces a vast majority of exchangeable hydrogens in the sample with deuterium, causing total signal loss. The exchange process is often further catalyzed by acids, including water residue, a notable broadening factor in aprotic environments. Contrary to what could previously be seen in fig. 1.4, increasing rate constant of intermolecular processes to the upper limits of NMR timescale does not lead to signal averaging, but rather to a phenomenon called exchange decoupling, where  $\tau$  of each potentially coupled spin state is too short to influence each other (see fig. 1.5). The rate constant  $k$  and other data for exchanging hydrogens can again be extracted through lineshape analysis of signals coupled to the exchanging site, provided that the coupling constant ( $^3J_{HH}$ ) can be obtained.

# Chapter 2

## Computational Chemistry

### 2.1 The Importance of Quantum Chemistry

Nowadays, quantum chemical calculations are an immensely helpful tool for researchers all over the world [10]. A number of user-friendly programs has been developed for routine use of computational chemistry in order to gain valuable insight before running actual experiments, saving precious time and money on resources. Possible jobs range anywhere from predicting spectral parameters to complex conformational studies of protein-ligand complexes and much more [11]. Calculating ground state energies for many isomers of a molecule is a matter of hours, provided that a capable workstation is available. Everything is done by methods developed to provide the best possible accuracy to computational cost ratio when approximating solution of the Schrödinger equation, presented in the time-independent three dimensional multi-particle general form in eq. (2.1) [10, 12, 13].

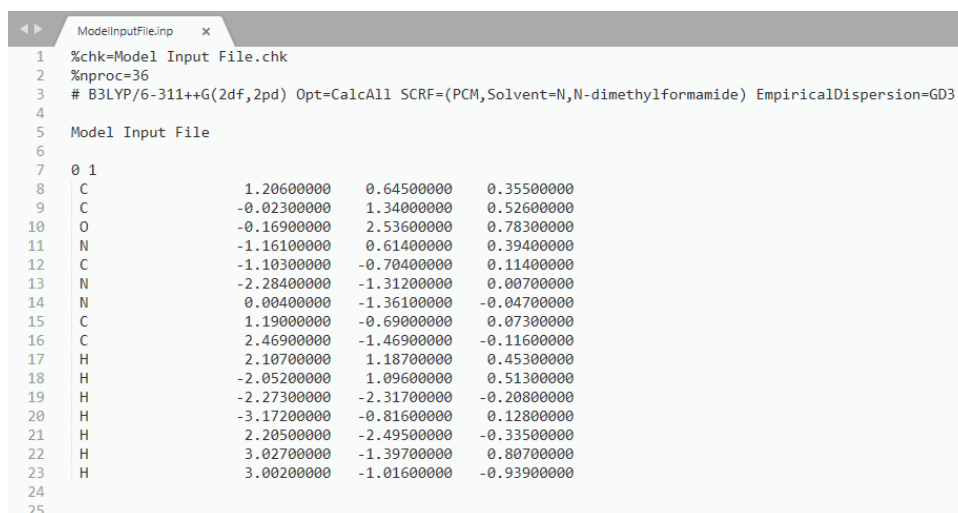
$$\hat{H}|\Psi\rangle = E|\Psi\rangle \quad (2.1)$$

where  $\hat{H}$  is the Hamiltonian operator,  $\Psi$  is a time-independent wave function and  $E$  is total energy of the system. The energy of a system is often measured in Hartrees ( $E_h$ ), where  $1E_h = 627.509$  kcal/mol.

### 2.2 Practical Aspects of Computational Chemistry

In order to perform calculations on a system, one must first put together an input containing all necessary commands and information for the software. In case of Gaussian, this input is composed of two major sections, as seen in fig. 2.1. All computational chemistry

within this thesis was done using exclusively HyperChem 6 for molecular modelling and Gaussian 03 [14] and Gaussian 16 [15] for the quantum chemical calculations themselves.



```
ModelInputFile.inp x
1 %chk=Model Input File.chk
2 %nproc=36
3 # B3LYP/6-311++G(2df,2pd) Opt=CalcAll SCRF=(PCM,Solvent=N,N-dimethylformamide) EmpiricalDispersion=GD3
4
5 Model Input File
6
7 0 1
8 C 1.20600000 0.64500000 0.35500000
9 C -0.02300000 1.34000000 0.52600000
10 O -0.16900000 2.53600000 0.78300000
11 N -1.16100000 0.61400000 0.39400000
12 C -1.10300000 -0.70400000 0.11400000
13 N -2.28400000 -1.31200000 0.00700000
14 N 0.00400000 -1.36100000 -0.04700000
15 C 1.19000000 -0.69000000 0.07300000
16 C 2.46900000 -1.46900000 -0.11600000
17 H 2.10700000 1.18700000 0.45300000
18 H -2.05200000 1.09600000 0.51300000
19 H -2.27300000 -2.31700000 -0.20800000
20 H -3.17200000 -0.81600000 0.12800000
21 H 2.20500000 -2.49500000 -0.33500000
22 H 3.02700000 -1.39700000 0.80700000
23 H 3.00200000 -1.01600000 -0.93900000
24
25
```

**Figure 2.1:** An example of a complete input file for calculations in Gaussian.

The first section specifies available system resources, applied method and basis set, optimization procedure and any additional optional command the user wants to utilize, such as applied reaction field and various correction factors. The second section specifies the system on which the first section commands will be applied. This specification begins by stating the charge and spin multiplicity for each molecule and complex contained, followed by atom description with Cartesian coordinates of their relative location. To get the coordinates, I first model the system manually and then use a fast nonlinear conjugate gradient method geometry optimization, in my case the Polak-Ribière conjugate gradient, to roughly pre-optimize the structure/s [16, 17]. This is then saved as a .sjf, .mol or .pdb file, converted to a native gaussian format and stripped of all redundant data. Once the commands are entered and checked, the calculation can be run.

## 2.3 Theory Behind Used Experiments

To limit the size of this section to acceptable levels, I shall only address the theory behind models seen in fig. 2.1, as they represent the settings which were used for all performed calculations. This section relies on texts [18, 19].

---

### 2.3.1 Mathematical Model

The most important part of the calculation is the used mathematical model. B3LYP is a member of the density functional theory (DFT) method group. DFT uses functionals to calculate electron density through computing its components: the kinetic energy, electron-nuclear interaction, Coulomb repulsion and electron-electron interaction, which boasts separate exchange and correlation terms [20]. This approach provides better accuracy while only slightly increasing the computational cost compared to classical Hartree-Fock *ab-initio* methods [21]. B3LYP itself is a hybrid functional, consisting of Becke-style 3 parameter exchange functional and the Lee-Yang-Parr gradient corrected correlation functional [22, 23].

### 2.3.2 Basis Set

The basis function set, such as 6-31G\*\*, expresses the atomic and molecular orbitals through functions  $\chi_r$ . The choice of appropriate basis set is vital for obtaining satisfactory results. The 6-31G\*\* basis set uses Gaussian-type polarization functions [24] to express orbital layers ( $s, p, d, f$ ) that can be applied to certain types of atoms according to their character. The set 6-31G\*\* (syn. 6-31G( $d, p$ )) is suitable for medium sized systems containing atoms up to chlorine, which adds up to type  $d$  Gaussian for heavy atoms and up to type  $p$  Gaussian for hydrogens. Larger basis sets mean less spatial constraints on electrons and more accurate approximations, although with increased computational cost. The 6-311++G(2df,2pd) basis set is further discussed in section 3.4.

### 2.3.3 Geometry Optimization

The geometry optimization protocol is set by the command Opt. The energy of a system can be plotted as a multidimensional surface with the number of degrees of freedom specifying the dimension count. Every point of this surface is described by a unique set of geometric variables of the system. The usual goal of geometry optimizations is to find the surface energy minima with global minimum being the ground energy state. Gaussian uses an algorithm utilizing local surface gradients to evaluate whether convergence criteria have been met for the calculation to terminate. Force constants play a very important role within these criteria and in order to ensure that the true energy minimum is found, we use the CalcAll command, which forces Gaussian to compute force constants at every step of the calculation. Coordinate format can be chosen freely by a command, such as Cartesian, otherwise Gaussian uses its internal coordinates system [18].

### 2.3.4 Reaction Field

Any calculation is performed on a system in gas phase by default. As we are modelling systems in solution, solvent effects must be taken into consideration. We simulate the solvent environment by applying a so called self-consistent reaction field (SCRF) in the polarizable-continuum model (PCM) [25, 26]. This method places the system in a cavity formed by spheres surrounding every atom, having radius of 1.2 times the van der Waals radius of their corresponding atom [27]. The cavity represents the solvent as a continuum with relative permittivity ( $\epsilon$ ), with its polarization represented numerically. The input values are specified by listing the solvent, in our case *N,N*-dimethylformamide.

### 2.3.5 Dispersion

In order to increase the accuracy of DFT calculations, empirical dispersion corrections are applied to the methods. We applied the GD3 correction in all cases [28].

# Chapter 3

## The Hydrogen Bond

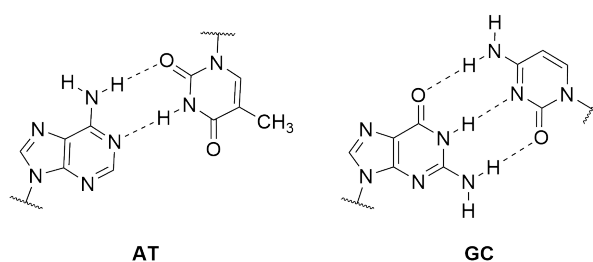
### 3.1 Definition of the Hydrogen Bond

The official IUPAC definition, postulated in 2011 by Arunan et al. says: “The hydrogen bond is an attractive interaction between a hydrogen atom from a molecule or a molecular fragment X–H in which X is more electronegative than H, and an atom or a group of atoms in the same or a different molecule, in which there is evidence of bond formation.” [29]

The Hydrogen bonding is a weak to strong type of interaction depending on the molecular environment, which is suggested to have an upper energy limit of 167 kJ/mol and no lower limit boundaries. Stronger hydrogen bonds usually occur between an electron deprived hydrogen and an electron rich interacting molecular fragment. The group possessing a covalent bond with hydrogen is commonly dubbed as the hydrogen bond donor (D) while the interacting group is called the hydrogen bond acceptor (A). The phenomenon itself is mainly of electrostatic nature but also consists of a covalent contribution [29].

### 3.2 The Importance of Hydrogen Bonds in Biology

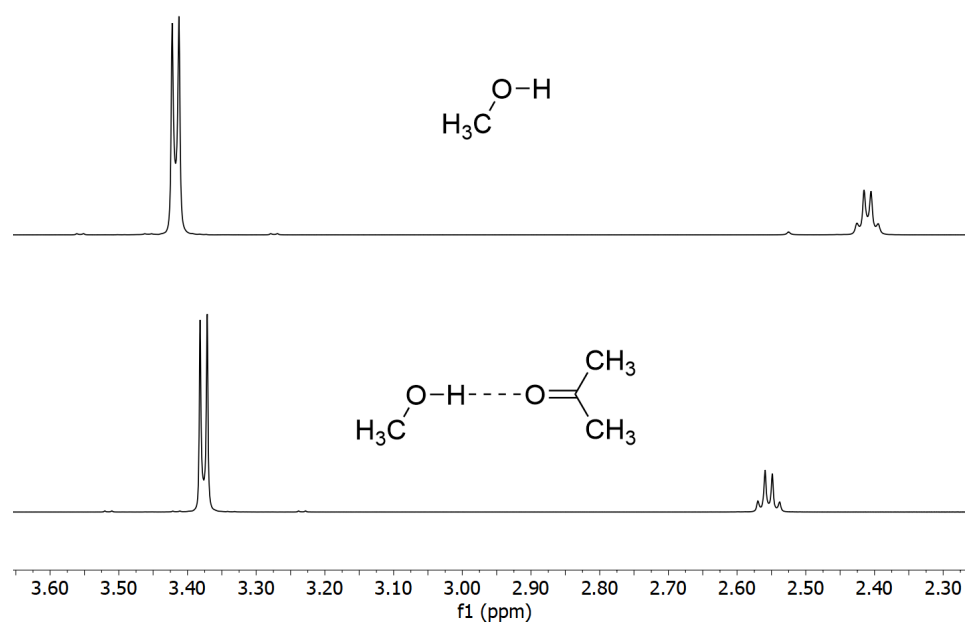
Hydrogen bonds are ubiquitous in all biological systems. Without them, the existence of life as we know it would be impossible, as they represent a key component of substrate recognition, ligand interaction and nucleobase pairing (see fig. 3.1) in nucleic acids [30–32]. Moreover, they are also responsible for the high melting and boiling points of water and greatly contribute to solvation effects, providing a suitable reaction environment in cells [33, 34].



**Figure 3.1:** The two canonical DNA base pairs. Adenine (A) pairs with Thymine (T) through 2 hydrogen bonds and Guanine (G) pairs with Cytosine (C) through 3 hydrogen bonds.

### 3.3 NMR Spectroscopy of the Hydrogen Bond

Nowadays, NMR spectroscopy is widely regarded as one of the best and most practical analytical methods for hydrogen bond observation. Even in the early days of NMR, when field strengths were low, instrumentation rather primitive and spectra measured by the continuous wave (CW) approach, it was possible to obtain indirect evidence of hydrogen bonding with good sensitivity by measuring chemical shift change ( $\Delta\delta$ ) of the involved hydrogens caused by their deshielding, as seen in fig. 3.2 [8].



**Figure 3.2:**  $^1\text{H}$  spectrum of methanol in  $\text{CDCl}_3$  before (top) and after adding a small amount of acetone (bottom). The formed hydrogen bond between methanol and acetone changes the O-H bond length and charge distribution, causing deshielding of the hydroxyl hydrogen.



This principle is also applicable to NMR of other nuclei involved in hydrogen bonding, most notably  $^{19}\text{F}$ ,  $^{17}\text{O}$  and  $^{15}\text{N}$ , which get more shielded. Exploitation of the partly covalent nature of hydrogen bonds can yield direct evidence (J-coupling) with the use of methods such as transverse relaxation-optimized spectroscopy (TROSY), which is often used in bioNMR, and much less practical INADEQUATE [35–37].

### 3.4 Hydrogen Bonding and Quantum Chemistry

Special care must be taken when using *ab initio* calculations on systems with prominent hydrogen bonding activity and dispersion interactions, especially intermolecular complexes. All calculations performed with finite basis sets are susceptible to the so called basis set superposition error (BSSE), which causes an artificial alteration of complex energy compared to separated monomers. BSSE produced by calculations scales inversely with the used basis set size, thus the use of large sets is highly advised [19]. A large set such as 6-311++G(2df,2pd) puts two *d* and one *f* functions on heavy atoms and two *p* and one *d* functions on hydrogens, while the plus signs indicate the use of diffusion functions on both heavy atoms and hydrogens [18].

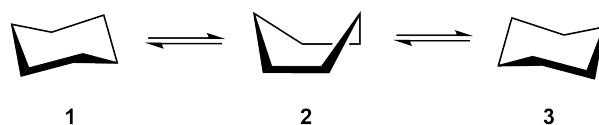
In account to compensate for BSSE, one may apply the counterpoise correction (CP) in gas phase calculations which attempts to compute the error and subtracts it from calculated energies, although the improvement in correlation to experimental results is often not as notable as expected [38–40]. A significant problem arising from the CP approach is the inability to use SCRF simultaneously. Since the implicit solvent can be a crucial parameter during calculations, the logical alternative is to omit CP entirely and instead mitigate the problem by using the largest acceptable basis set [18].

# Chapter 4

## Conformational Isomerism and Tautomerism

### 4.1 Conformers

According to IUPAC, a conformer is described as “One of a set of stereoisomers, each of which is characterised by a conformation corresponding to a distinct potential energy minimum.” [41] A conformation is “The spatial arrangement of the atoms affording distinction between stereoisomers which can be interconverted by rotations about formally single bonds.” [41] A typical example of conformational isomerism is the set of cyclohexane conformers.



**Figure 4.1:** The chair conformations of cyclohexane (1; 3) represent the global energy minimum, being the most abundant form at any given time, while the boat (2) and far less stable half-chair and twist-boat conformations correspond to local minima. The forms can freely transform between each other and specific substitution may even lead to the boat form being the global minimum due to reasons discussed below.

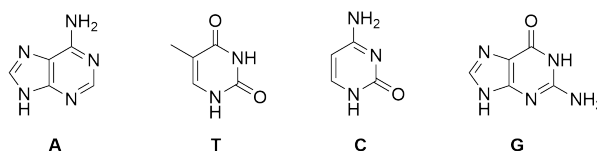
Should the rotational potential energy barrier of such bond in acyclic systems (best defined by dihedral torsion angles) be high enough to make observation of separate conformations possible as distinct species on the experiment timescale due to restricted rotation, then such conformations are called rotamers [41].

The potential energy barrier between conformers is most commonly increased through phenomena such as steric hinderance, bond order higher than 1, ring strain (see fig. 4.1) and dipole-dipole interactions. [1, 41]. Different conformations of molecules usually produce different magnetic environment observed by the nuclei and are suitable to analysis by dNMR.

## 4.2 Tautomers

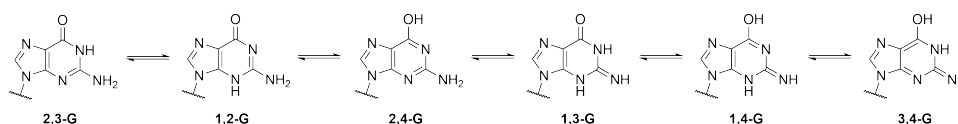
“Tautomers are structural isomers that differ from one another based on the position of proton(s) and double bonds.” [42] Tautomeric forms behave similarly to conformational isomers in sense of being able to freely interchange between each other while their relative abundance is highly dependent on the potential energy difference [43]. The major influencing parameters for tautomeric equilibria are aromatic resonance, lone-pair and dipolar repulsion, hydrogen bonding, local electron density and solvent effects [43].

Just like hydrogen bonding, tautomerism plays a vital role in biological systems. The major tautomeric forms of nucleobases (see fig. 4.2) ensure the correct pairing of nucleotides (see fig. 3.1) in nucleic acids, providing an effective, universal and selective way of storing information, shaping life as we know it [34, 44].



**Figure 4.2:** The four canonical DNA nucleobases adenine (A), thymine (T), cytosine (C) and guanine (G).

The rare tautomeric forms are suspected of contributing to replication errors in DNA, but also at the same time suggested to enhance structural and functional diversity of RNA enzymes and aptamers [42].



**Figure 4.3:** Possible tautomeric forms of 9-substituted guanine. The 2,3-G form is the only canonical tautomer present in DNA.

As tautomerism is a fast dynamic process defined by an equilibrium, isolation of separate

tautomeric forms is impossible and in cases of strong thermodynamical preference of the major tautomer, observation in polar solvents at room temperature is nigh impossible with conventional and widespread methods except for UV-VIS, IR and Raman spectroscopies, which on the other hand often have problems with structural assignment [42, 43].

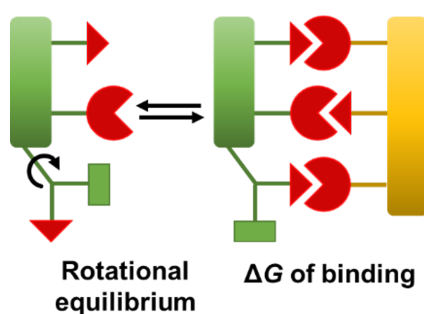
Dynamic NMR stands as a well established method for detection of tautomers in solution and solids. Each tautomeric form can be seen either as a distinct entity or an averaged signal in the spectra and concentrations at equilibrium can be obtained by signal integration or calculated by lineshape fitting [43].

# Chapter 5

## Nucleobase-Pairing Free Energies from Rotameric Equilibria

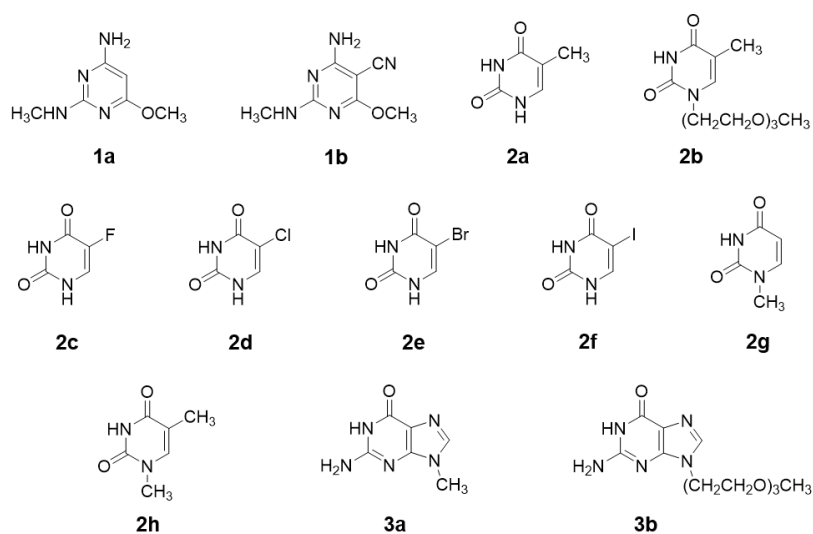
### 5.1 Introduction

Our research of rotameric equilibria is based on the fact that each rotamer of 2-(methylamino)pyrimidine (2-MAP) has a different hydrogen bonding pattern, one of which is complementary to a selected nucleobase, as illustrated by fig. 5.1. A modest library of 2-MAP derivatives and binding partners was carefully selected to explore the electronic effects of 2-MAP substitution on complex formation free energies (see fig. 5.2).



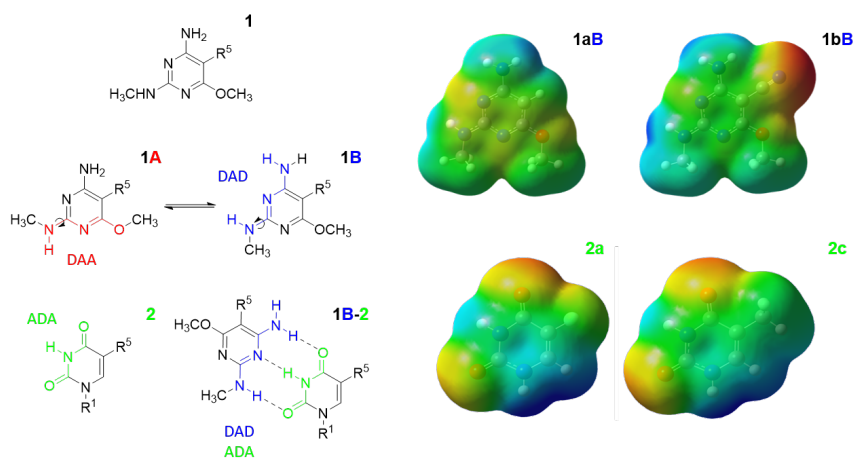
**Figure 5.1:** One of the rotamers has a complementary binding pattern to the ligand, allowing the formation of a hydrogen bonded complex [1].

The native ratio of rotamers at equilibrium (which will be further referenced to as A and B, depending on their orientation) is changed upon the addition of a suitable binding partner. Monitoring and quantifying these ratios by variable temperature NMR enables us to determine the complex formation free energies by a simple calculation [1].



**Figure 5.2:** Library of compounds used for the research of rotameric equilibria.

The rotamer ratio is also strongly solvent-dependent. Our endeavours began with compound **1a**, which was first observed in *N,N*-dimethylformamide- $d_7$  (DMF- $d_7$ ), where rotamer A (see fig. 5.3) is the major component. In methanol- $d_4$  however, rotamer B is naturally more abundant [1]. Rotamer B of compounds **1a** and **1b** with the donor-acceptor-donor (DAD) binding pattern is capable of forming a hydrogen bonded complex with substituted uracils **2a-2h**, as shown in fig. 5.3. There are two possible orientations of the substituted uracil in regards to the 2-(methylamino)pyrimidines, analogous to Watson-Crick and reverse Watson-Crick pairing.



**Figure 5.3:** Rotamers of 2-(methylamino)pyrimidines and the pairing of rotamer B to substituted uracils. The calculated electrostatic potential plots of structures **1aB** and **1bB** in contrast to **2a** and **2c** nicely show how the binding regions are complementary to each other.

## 5.2 Mathematical Aparatus

In order to calculate the complex formation free energy ( $\Delta_r G$ ), one must first obtain the association constant  $K_2$ . We begin with outlining the experimentally available concentration values. Only the rotamer B is capable of forming a hydrogen bonded complex with the ligand.

$$\begin{aligned}
 [L_{\text{TOT}}] &= [L] + [BL] \\
 [S_{\text{TOT}}] &= [A] + [B] + [BL] \\
 [A_{\text{TOT}}] &= [A] \\
 [B_{\text{TOT}}] &= [B] + [BL]
 \end{aligned}
 \tag{5.1}$$

Where L is the ligand, S stands for substrate, A and B represent each of the rotamers and BL indicates a complex between rotamer B and the ligand. TOT indicates the total amount present in solution. The concentrations themselves are calculated from known sample mass and relative integral intensities.

The equilibrium constant  $K_1$  of reaction  $B \rightarrow A$  can easily be calculated from relative concentrations of rotamers A and B.

$$K_1 = \frac{[A]}{[B]} \tag{5.2}$$

The association constant  $K_2$  of reaction  $B+L \rightarrow BL$  is defined by eq. (5.3).

$$K_2 = \frac{[BL]}{[B][L]} \tag{5.3}$$

Note that while eq. (5.3) is fairly simple, neither of the concentrations is readily available from the spectra. They can, however, be obtained from eq. (5.1) by utilizing eq. (5.2).

$$\begin{aligned}
 [B] &= \frac{[A_{\text{TOT}}]}{K_1} \\
 [L] &= [L_{\text{TOT}}] - [B_{\text{TOT}}] + \frac{[A_{\text{TOT}}]}{K_1} \\
 [BL] &= [B_{\text{TOT}}] - \frac{[A_{\text{TOT}}]}{K_1}
 \end{aligned}
 \tag{5.4}$$

Solving eq. (5.3) with eq. (5.4) yields the final eq. (5.5).

$$K_2 = \frac{K_1[B_{\text{TOT}}] - [A_{\text{TOT}}]}{[A_{\text{TOT}}]([L_{\text{TOT}}] - [B_{\text{TOT}}])} \tag{5.5}$$

### 5.3 DFT Calculations

Density functional theory calculations in the Gaussian 03 program package were implemented to obtain optimized geometries, electronic energies and free energies of all used compounds in fig. 5.2 (including rotamers A and B for compounds **1a** and **1b**) with simulated DMF solvation at B3LYP/6-31G(d,p) level except **2f** (used method is unsuitable for iodine) and **2b/3b**, which were substituted with **2h** and **3a** for simplicity of the calculation. Furthermore, complexes of compound **1a** with **2a**, **2c**, **2d**, **2e**, **2h** and compound **1b** with **2h** (both Watson-Crick like and reverse Watson-Crick like mutual orientations in all cases), plus a dimer of **3a** were also calculated to obtain their geometries and electronic/free energies. All detailed data charts are available in the ESI of [1].

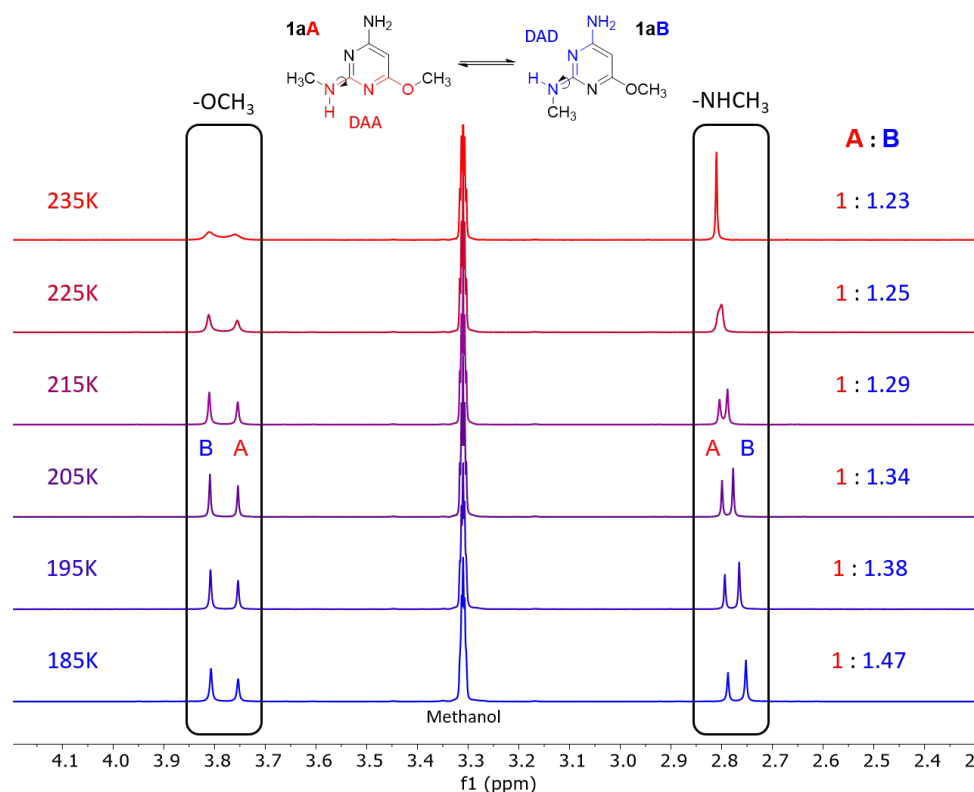
### 5.4 Experimental Results and Discussion

Experiments began with measuring native rotamer ratios of compound **1a** in DMF- $d_7$ , methanol- $d_4$  and their various mixtures (see ESI of [1] for a full chart). Two most important results are that the A:B ratio is 6:4 for DMF- $d_7$  and 4:6 for methanol- $d_4$ .

Since there is no need for an aprotic environment as the rotamer ratios can be obtained by integration of either the methoxy or methylamino signals, pure methanol- $d_4$  was chosen because it has a significantly lower melting point than DMF (by about 45K depending on purity and water contamination) and is also noticeably cheaper. This choice was also very convenient as working with solvent mixtures (which can advantageously have a lower melting point), would come with various theoretical problems concerning unstable rotamer equilibrium due to solvent evaporation changing the mixture ratios, and also some practical issues such as difficulty shimming at lower temperatures caused by weak solvent signals.

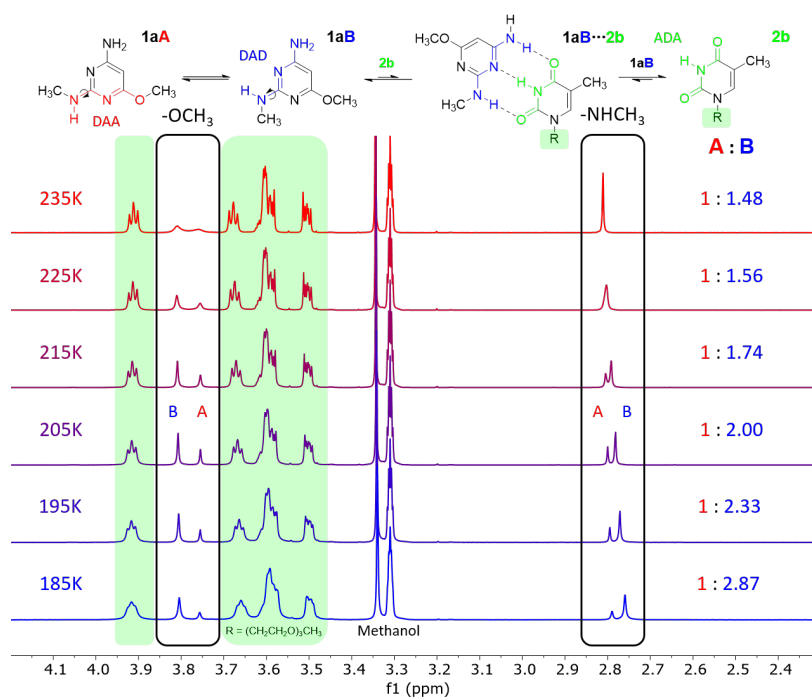
A 10 mmol solution of compound **1a** in methanol- $d_4$  was measured from 185 K up to the methylamino group signal coalescence at 235 K with 10 K steps. (see fig. 5.4). Applying eq. (5.2) on rotamer ratios obtained by manual integration gives temperature dependent  $K_1$  values.



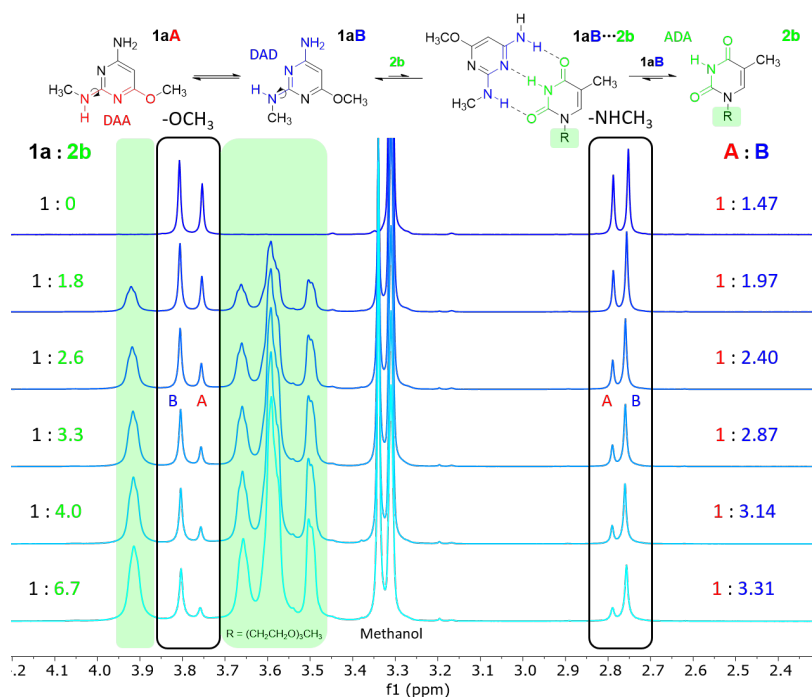


**Figure 5.4:**  $^1\text{H}$  NMR variable temperature spectral array of a 10 mmol/l solution of compound **1a** in methanol- $d_4$ .

To quantify the influence of intermolecular interactions on the rotameric equilibria, we proceeded with a 1-N-oligoethyleneglycol substituted thymine, which serves as an ideal example for it has a close structural relation to thymine found in DNA and the substitution guarantees satisfying solubility. Substituted uracils always bind with an acceptor-donor-acceptor (ADA) pattern regardless of their relative orientation to the rotamer B, which has a DAD pattern. We expected to see a change in the rotamer ratios in a direction further favoring B in methanol- $d_4$ . Variable temperature experiments identical to those shown in fig. 5.4 were executed for mixtures with increasing increments of **2b** from 1.8 up to 6.7 equivalents. These measurements yielded a large data set enabling analysis of the effect of both temperature and concentration of binding partner, as visualised by fig. 5.5 and fig. 5.6. Increasing concentration of **2b** and lowering the temperature does indeed favor the complex formation and inherently increases rotamer B concentration.



**Figure 5.5:**  $^1\text{H}$  NMR variable temperature spectral array of a 1:3.3 mixture of compounds **1a** (10 mmol/l) and **2b** in methanol- $d_4$ .

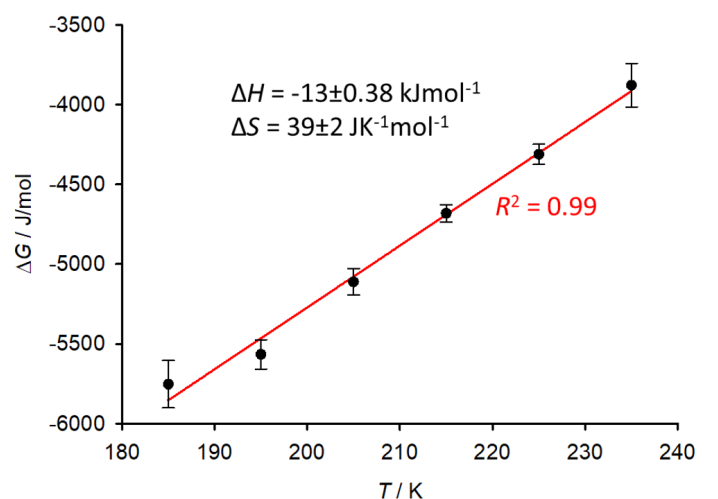


**Figure 5.6:**  $^1\text{H}$  NMR spectral array of compound **1a** (10 mmol/l) and **2b** mixtures with variable concentration ratios in methanol- $d_4$  at 175 K.

The rotamer ratios of all **1a-2b** mixtures were obtained simultaneously by both manual integration of spectra and line shape fitting of the spectra. Applying eq. (5.5) gives us everything we need to calculate Gibbs free energy by eq. (5.6).

$$\Delta G = RT \ln K_2 \quad (5.6)$$

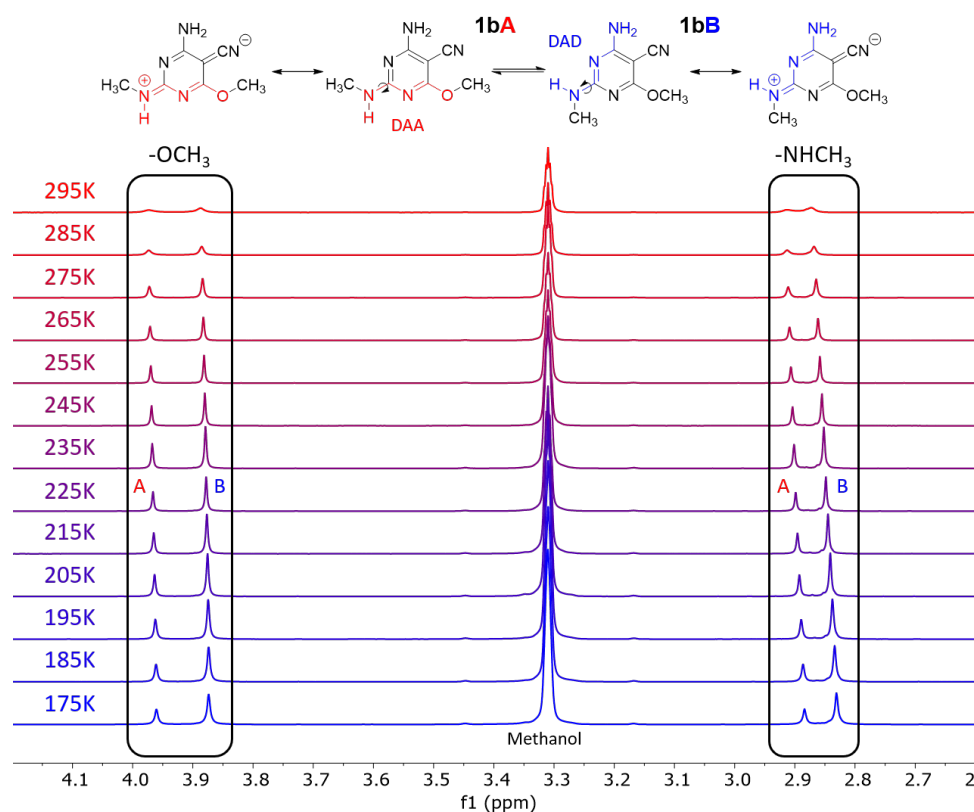
Where R is the universal gas constant ( $8.314 \text{ JK}^{-1} \text{ mol}^{-1}$ ) and T is temperature. Solving eq. (5.6) for every temperature gives us the linear dependence graph in fig. 5.7.



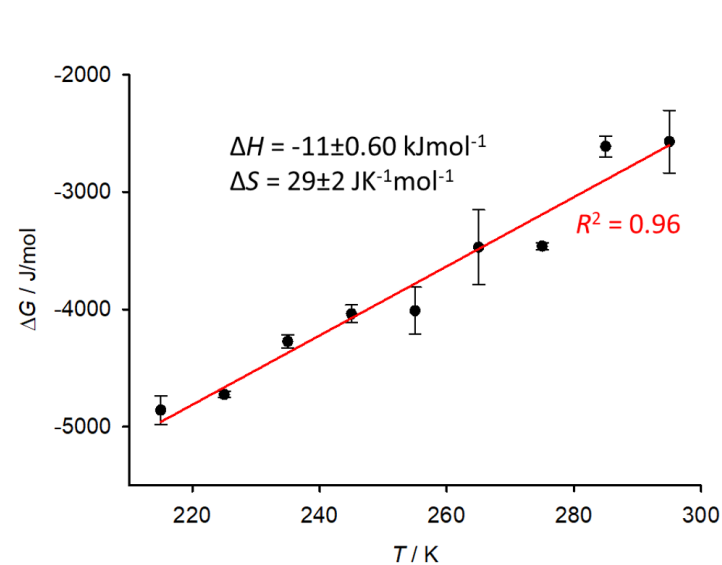
**Figure 5.7:** Temperature dependence graph of complex formation Gibbs free energy ( $\Delta G$ ) between compounds **1a** and **2b** in methanol- $d_4$

There is a possibility of greatly expanding the suitable temperature range by introducing an electron-withdrawing group (EWG) in position 5 of the 2-(methylamino)pyrimidine structure. Our experiments with compound **1b** possessing a nitrile group in position 5 prove that this theoretical concept works flawlessly in practice and both rotamers can be observed even at room temperature, as displayed in fig. 5.8. The coalescence temperature itself is increased by a higher rotational energy barrier caused by push-pull interactions of the substituents.

An unexpected downside regarded to compound **1b** in particular, that has not been apparent at first sight, is probably a very slight insolubility of either the compound itself or the formed complex at very low temperatures, which has caused the  $\Delta G/T$  graph in fig. 5.9 to be less accurate than the one in fig. 5.7.



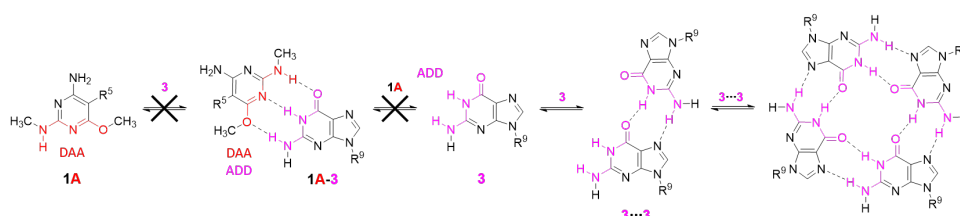
**Figure 5.8:**  $^1\text{H}$  NMR variable temperature spectral array of a 10 mmol/l solution of compound **1b** in methanol- $d_4$ . Note the wide range of available temperatures in comparison to fig. 5.4.



**Figure 5.9:** Temperature dependence graph of complex formation free energy between compounds **1b** and **2b** in methanol- $d_4$

Rotamer A has a donor-acceptor-acceptor (DAA) pattern, suitable to form hydrogen bonded complexes with compounds having an accessible complementary pattern. We decided to attempt some experiments with 9-substituted guanine **3b**, again keeping in touch with the DNA substrate theme, utilizing a short oligoethylene glycol substitution for increased solubility.

No matter the solvent and mixture ratio though, no effects on the rotameric equilibria of either compound **1a** or **1b** was observed. However, especially the H1 and H8 signals of compound **3b** were shifting significantly as the concentration of **3b** in the solution increased. This suggests that homodimers of **3b** or higher supramolecular structures such as a tetramers are forming (see fig. 5.10). This type of guanine complexes was previously found in telomeres. [45]



**Figure 5.10:** Hypothetical scheme of complex formation between the rotamer A of compounds **1a** and **1b** and substituted guanines **3a** and **3b**. The scheme also contains proposed dimer and tetramer structures of these guanine analogues.

We have experimentally determined the complex formation  $\Delta G$  values in  $\text{DMF}-d_7$  at 240 K for complexes of compound **1a** with various position 5 substituted uracils (see table 5.1).

There is a clearly visible trend of increasing complex stability with higher electron-donating character of the substituents, although the differences are not as large as one would anticipate. The small differences are probably originating from counterbalancing of the opposing effects the substituents have on the hydrogen-bonding strength in regards to how well the involved groups serve as a hydrogen bond donors or acceptors [1].

**Table 5.1:** Experimental  $\Delta G$  values ( $\text{kJ mol}^{-1}$ ) of hydrogen-bonded complex formation between compound **1a** and **2a-2g** in  $\text{DMF}-d_7$  at 240 K.

	$R^5$	$\Delta G_{\text{exp}}$
<b>2a</b>	$\text{CH}_3$	$-3.0 \pm 0.3$
<b>2b</b>	$\text{CH}_3$	$-2.6 \pm 0.2$
<b>2c</b>	F	$-0.6 \pm 0.1$
<b>2d</b>	Cl	$-0.7 \pm 0.1$
<b>2e</b>	Br	$-0.9 \pm 0.1$
<b>2f</b>	I	$-1.5 \pm 0.3$
<b>2g</b>	H	$-2.3 \pm 0.1$

The complex formation free energy of **1a-2b** in  $\text{DMF}-d_7$  is significantly lower in comparison to values determined in methanol- $d_4$ , which serves as an evidence of good solvation of the compounds by  $\text{DMF}-d_7$ . This provides a good test of the applicability of the method, for both the resulting magnitudes of  $K_2$  and complex formation  $\Delta G$  are very small. Other methods for determination of these values, such as microcalorimetry, usually require concentrations of interacting partners in the order of  $K_2^{-1}$ , which would be impossible to achieve with the presented compounds [1].

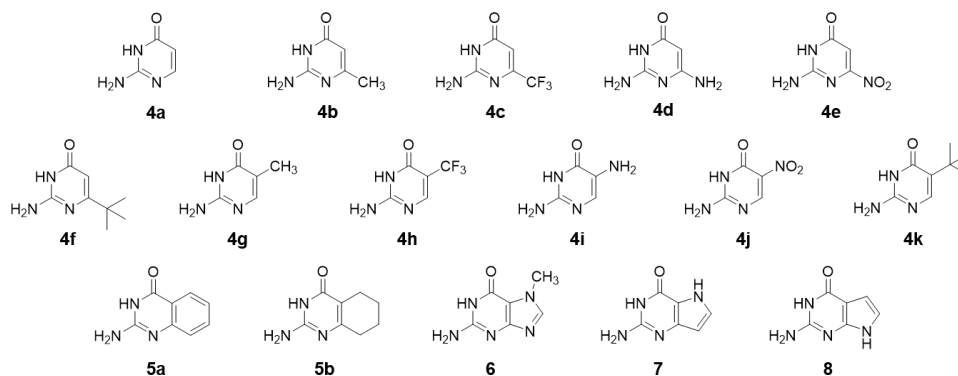
In conclusion, we have developed a fast, straightforward and widely applicable method for determination of complex formation free energy in solutions of modified nucleobases by NMR.

# Chapter 6

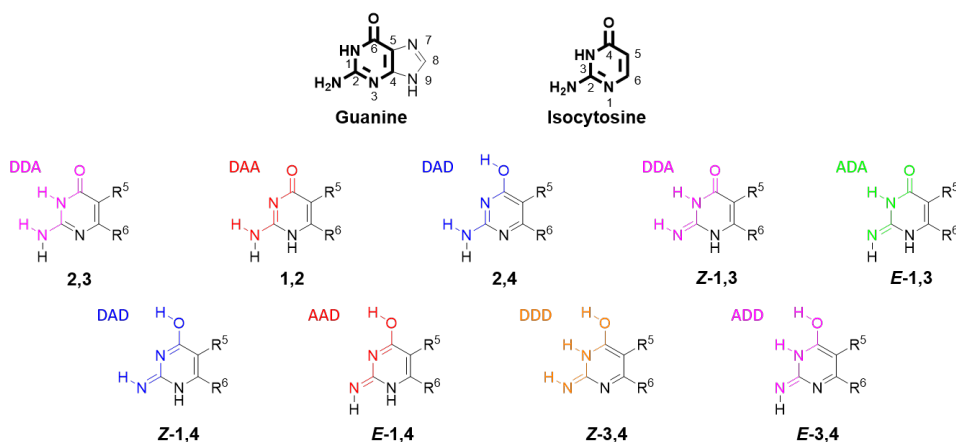
## Tautomerism of Guanine Analogues

### 6.1 Introduction

Tautomerism of nucleobases plays an important role in ensuring that the replication and transcription of genetic information within cells happens correctly. Uncommon tautomers of nucleobases are suggested to be participating components of several processes catalysed by nucleic acid enzymes [42] and responsible for some nucleic acid mutations [46]. Guanine in particular has a large number of possible tautomers with several different hydrogen-bonding patterns. Based on previous research of rotameric equilibria [1] and isocytosine [47, 48], which can be thought of as a structural fragment of guanine (see fig. 6.2), we became interested in studying guanine and its structural analogues (see fig. 6.1) through variable temperature NMR spectroscopy and DFT calculations to explore the effects responsible for the remarkable stability of the canonical guanine tautomer [2].



**Figure 6.1:** Library of compounds used for the research of tautomerism. Compounds **2b**, **2h**, **3a** and **3b** from fig. 5.2 were also used.



**Figure 6.2:** All possible tautomers of guanine and isocytosine analogues with their respective binding patterns (hydroxyl group rotamers excluded).

The employed tautomer labeling represents exchangeable hydrogen positions according to isocytosine numbering (different to that of guanine), which was also used for substituent numbering. The stereochemical labels are valid for all compounds used for the purposes of this research, see fig. 6.1. Should the atomic number of connecting atom of  $R^6$  exceed 16, the stereochemical notation ( $E/Z$ ) would be inverted.

## 6.2 DFT Calculations

Quantum chemical calculations had a vital role in the initial screening of various substituent effects on the relative energies of distinct tautomeric forms of selected isocytosine and guanine analogues. We ran geometry optimizations at B3LYP/6-311++G(2df,2pd) level with implicit DMF solvation in Gaussian 16 program package for tautomers **2,3**; **1,2**; **2,4**; **Z-1,3** and **E-1,3** of all structures in fig. 6.1, including structure **3a** from fig. 5.2 (see table 6.1 and table 6.2). Tautomers **1,4** and **3,4** (iminoenol) were omitted from the calculations as previous calculations of isocytosine tautomers [49] had shown that their energy is significantly higher than that of the most stable tautomer. Initial calculations were performed at B3LYP/6-31G(d,p) level and their results can be seen in the ESI of [2].

The tautomer **2,3** (3H-ketoamino, corresponding to the canonical guanine tautomer) has the lowest relative energy in all cases, followed by either tautomer **1,2** (1H-ketoamino, most cases) or **2,4** (enol). Substituents in position 5 have a strong influence on the tautomer **1,2** relative energy. Compared to isocytosine (**4a**), electron donating groups (**4i**) have a stabilizing effect while electron withdrawing groups (**4j/4h**) destabilize the



tautomer **1,2**. This is in contrast with substitution in position 6, where electron donating groups (**4d**) just as electron withdrawing groups (**4c/4e**) destabilize the tautomer **1,2** with the exception of alkyl substitution, which very slightly stabilises tautomer **1,2** regardless of both position (**4b/4g**) and sterical demands (**4f/4k**). Tautomers **Z-1,3** and **E-1,3** (imino) are always very similar in relative energies, destabilized by electron withdrawing groups (especially in position 6) and like tautomers **2,4** (enol), never fall below 20 kJ/mol.

**Table 6.1:** Relative energies (kJ/mol) of tautomers **2,3**; **1,2**; **2,4**; **Z-1,3** and **E-1,3** of monocyclic isocytosine analogues **4a-4k** calculated with implicit DMF solvation at B3LYP/6-311++G(2df,2pd) level.

	R <sup>5</sup>	R <sup>6</sup>	<b>2,3</b>	<b>1,2</b>	<b>2,4</b>	<b>Z-1,3</b>	<b>E-1,3</b>
<b>4a</b>	H	H	0.0	13.9	20.7	23.3	24.4
<b>4b</b>	H	CH <sub>3</sub>	0.0	12.6	28.2	21.3	22.4
<b>4c</b>	H	CF <sub>3</sub>	0.0	32.3	29.0	45.3	46.6
<b>4d</b>	H	NH <sub>2</sub>	0.0	34.8	28.0	36.7	38.8
<b>4e</b>	H	NO <sub>2</sub>	0.0	32.5	20.0	49.1	49.8
<b>4f</b>	H	C(CH <sub>3</sub> ) <sub>3</sub>	0.0	12.8	27.8	22.3	23.6
<b>4g</b>	CH <sub>3</sub>	H	0.0	12.0	30.5	23.8	24.5
<b>4h</b>	CF <sub>3</sub>	H	0.0	16.3	24.3	28.7	30.3
<b>4i</b>	NH <sub>2</sub>	H	0.0	6.5	25.9	29.8	29.6
<b>4j</b>	NO <sub>2</sub>	H	0.0	24.9	24.6	34.8	37.5
<b>4k</b>	C(CH <sub>3</sub> ) <sub>3</sub>	H	0.0	12.8	22.0	22.3	23.1

Guanine analogues **3a**, **6**, **7**, **8** and isocytosine derivatives **5a** and **5b** were subjected to the same calculations as structures **4a-4k**. The trends of tautomer **2,3** (3H-ketoamino) always being the most stable and tautomers **2,4** (enol) and **1,3** (imino) having relatively high relative energies persist. The results (table 6.2) show a remarkable instability of all tautomers other than **2,3** for guanine analogues **3a** and **8**, both of which have a heterocyclic nitrogen in position 9. These two compounds are the only ones with the lowest relative energy difference between tautomer **2,3** and the second most stable tautomer higher than 30 kJ/mol.

**Table 6.2:** Relative energies (kJ/mol) of tautomers **2,3**; **1,2**; **2,4**; **Z-1,3** and **E-1,3** of bicyclic guanine analogues **3a**, **6**, **7** and **8** and isocytosine analogues **5a** and **5b** calculated with implicit DMF solvation at B3LYP/6-311++G(2df,2pd) level.

	<b>2,3</b>	<b>1,2</b>	<b>2,4</b>	<b>Z-1,3</b>	<b>E-1,3</b>
<b>3a</b>	0.0	41.0	33.8	49.1	50.1
<b>5a</b>	0.0	8.5	46.8	20.4	20.1
<b>5b</b>	0.0	10.8	30.6	21.0	21.7
<b>6</b>	0.0	17.2	34.3	32.8	32.6
<b>7</b>	0.0	11.5	48.7	27.5	27.3
<b>8</b>	0.0	38.5	38.3	45.0	45.9

Based on the data shown above, we chose a handful of compounds to be tested experimentally to best illustrate our proposed concepts. In order to underline our observations with calculations, geometry optimizations were performed for dimers of complementary tautomers **2,3** (3H-ketoamino, DDA) and **1,2** (1H-ketoamino, AAD) (experimental example in fig. 6.3) of 9-methylguanine **3a** and isocytosine derivatives **4a**, **4b**, **4c** and **5a** (see table 6.3). Furthermore, we calculated intermolecular complexes of 9-methylguanine **3a** with isocytosine analogues **4b** and **4c**. Since isocytosine (**4a**) had shown promising relative energy values of tautomers **2,4** (enol) and **E-1,3** (imino) in table 6.1, we calculated complexes of these tautomers with suitable complementary hydrogen-bonding partners, 1-methylthymine **2h** and 2,6-diaminopyridine (**DAP**).

**Table 6.3:** Complexation and stabilization energies of selected dimers and complexes of isocytosine and guanine analogues calculated at a B3LYP/6-311++G(2df,2pd) level with implicit DMF solvation.

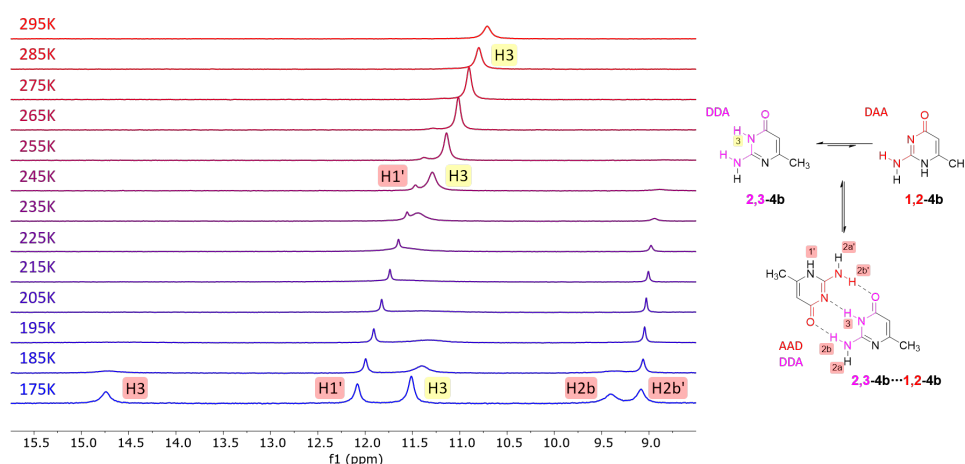
	$E_{\text{complexation}}$	$E_{\text{stabilization}}$
<b>2,3-3a + 1,2-3a</b>	-78.1	-37.2
<b>2,3-4a + 1,2-4a</b>	-79.9	-66.0
<b>2,3-4b + 1,2-4b</b>	-80.1	-67.5
<b>2,3-4c + 1,2-4c</b>	-79.5	-47.2
<b>2,3-5a + 1,2-5a</b>	-77.3	-68.8
<b>2,3-3a + 1,2-4b</b>	-79.2	-66.6
<b>2,3-3a + 1,2-5a</b>	-78.1	-69.7
<b>2,4-4a + 2h</b>	-68.7	-47.9
<b>1,3-4a + DAP</b>	-55.0	-32.6

The values of  $E_{\text{complexation}}$  represent the complexation energy calculated by subtracting the sum of participating tautomers from the total complex energy and  $E_{\text{stabilization}}$  is the stabilization energy calculated as a difference of the total complex energy and the most stable tautomers of participating molecules. The  $E_{\text{complexation}}$  thus displays the strength of intermolecular interactions forming the intermolecular complex and the  $E_{\text{stabilization}}$  expands on it by adding the energy penalty needed to first form the participating tautomers, should they not be the most stable form. Since the compounds are well solvated by DMF in solution and we only use a polarizable continuum model solvation in our calculations, it is reasonable to expect that the complex formation will be disfavored experimentally because of the need to overcome a higher energy barrier of solvation. The results of these calculations will be further discussed below in correlation to experimental observations.

### 6.3 Experimental Results and Discussion

$^1\text{H}$  NMR variable temperature arrays of studied compounds were measured in a 3:1 DMF- $d_7$  : dichloromethane- $d_2$  (DCM- $d_2$ ) mixture, utilizing a temperature range of 295 K-175 K with 10 K steps at 10 mmol/l sample concentrations. The spectral range was set sufficiently wide to enable the observation of eventually forming hydrogen-bonded complexes between different complementary tautomers of the compounds.

Based on the calculations above and solubility testing, our main interest focused on thymine analogue **2b**, 9-substituted guanine **3b** (**2h** and **3a** in calculations for simplicity), isocytosine derivatives **4a**, **4b**, **4c**, **5a** and 2,6-diaminopyridine (**DAP**), for which we ran complex geometry optimizations seen in table 6.3 to back up our observations.

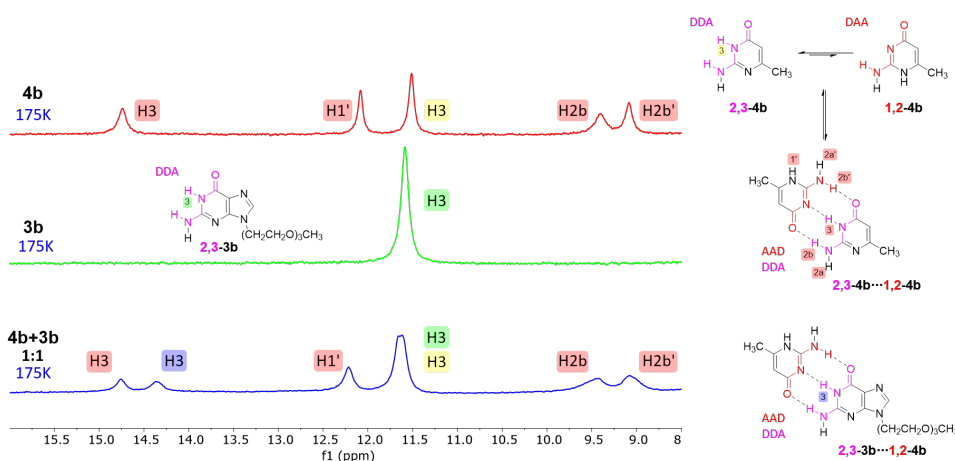


**Figure 6.3:**  $^1\text{H}$  NMR variable temperature spectral array of a 10 mmol/l solution of 6-methylisocytosine **4b** in a 3:1 N,N-dimethylformamide- $d_7$  and DCM- $d_2$  mixture.

As seen in table 6.1, alkyl substitution in either position 5 or 6 decreases the relative energy difference between the tautomers **2,3** and **1,2** (3H- and 1H-ketoamino). The fact we are able to clearly detect distinct signals of the forming **2,3-4b-1,2-4b** complex as high as 255 K in fig. 6.3 confirm the calculations and our initial expectations of behaviour in solution similar to isocytosine. A very distinct signal around 15 ppm appears below 195 K and serves as a very good identifier of the complex.

The next step was introducing our 9-substituted guanine **3b** to 6-methylisocytosine **4b** in a 1:1 ratio. Guanine analogues with a heterocyclic nitrogen in position 9 exist as an exceptionally stable **2,3** (DDA) tautomer (see table 6.2) and thus strongly promote the formation of tautomers **1,2** (AAD) of isocytosine derivatives which have a relatively stable **1,2** tautomer by nature in order to form a hydrogen bonded complex in solution

(see table 6.3). Three examples of isocytosine analogues are **4a**, **4b** (see fig. 6.4) and **5a**, all of which have relative energy of tautomer **1,2** below 15 kJ/mol in comparison to tautomer **2,3** (table 6.1).

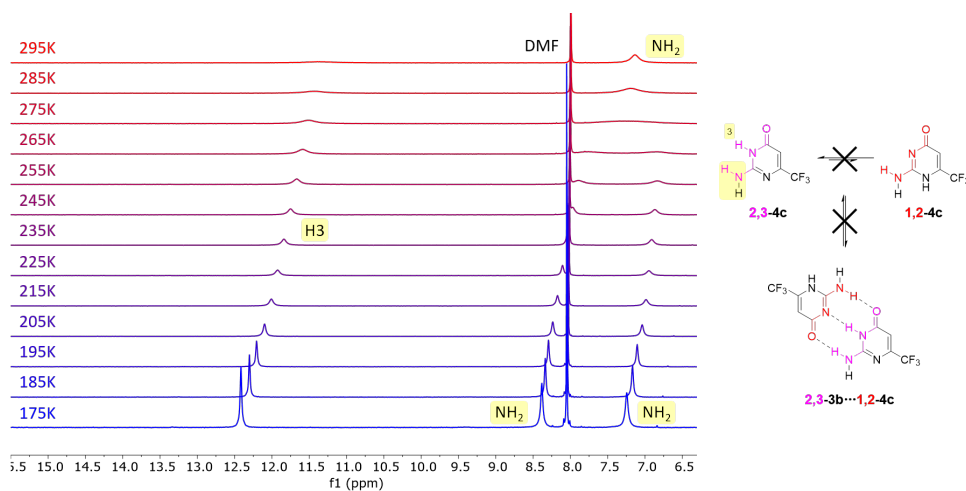


**Figure 6.4:**  $^1\text{H}$  NMR spectra of 6-methylisocytosine **4b**, 9-substituted guanine **3b** and their 1:1 mixture (10 mmol/l) in a 3:1 DMF- $d_7$  and DCM- $d_2$  solution at 175 K

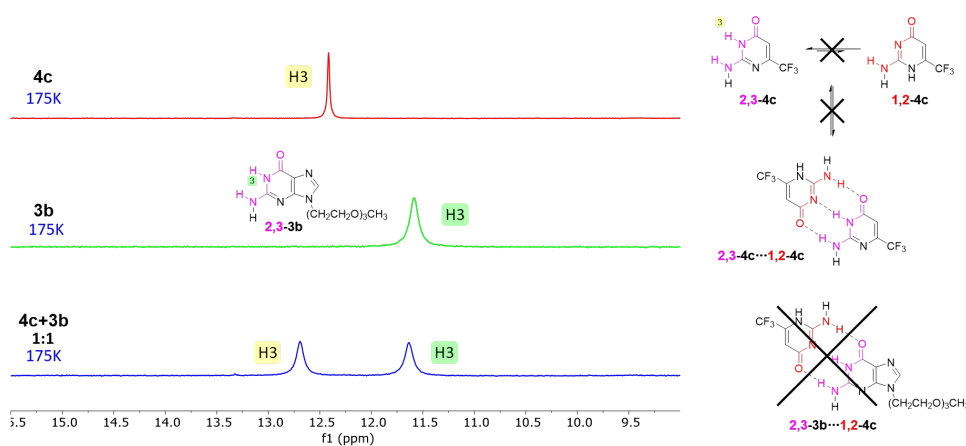
The experimental spectra of isocytosine derivatives **4a**, **4b** and **5a** yield practically identical results as in fig. 6.4, congruent with the positive prediction of complex formation, having  $E_{\text{stabilization}}$  well below  $-60$  kJ/mol (table 6.3). A second signal above 14 ppm undoubtedly confirms the existence of a hydrogen-bonded complex with 9-substituted guanine **3b**.

On the other hand, no interesting activity whatsoever can be seen in the spectra of isocytosine **4c** substituted with a trifluoromethyl group (EWG) in position 6 in fig. 6.5. This is in agreement with the calculated relative energy of tautomer **1,2-4c** (1H-ketoamino) being more than 30 kJ/mol higher than tautomer **2,3-4c** (3H-ketoamino) (table 6.1) and further backed by the calculated  $E_{\text{stabilization}}$  for complex between tautomers **2,3-4c** and **1,2-4c** above  $-50$  kJ/mol.

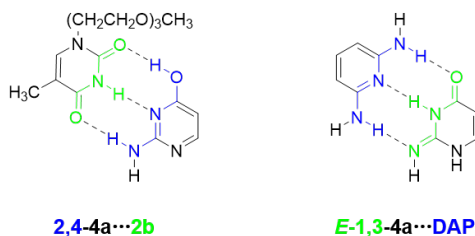
The concept holds true even upon the addition of 9-substituted guanine **3b** and no complex formation can be observed in the spectra. The calculated  $E_{\text{stabilization}}$  in DMF for complex of tautomers **3,2-3a** and **1,2-3a** is  $-37.2$  kJ/mol and 9-methylguanine (**3a**) is thus a lot less likely to form a **1,2-3a** tautomer in solution than compound **4c**, which shows no activity at all. This is evidenced in the spectra of compound **3b** seen in fig. 6.4 and fig. 6.6.



**Figure 6.5:**  $^1\text{H}$  NMR variable temperature spectral array of a 10 mmol/l solution of compound **4c** in a 3:1 DMF- $d_7$  and DCM- $d_2$  mixture.



**Figure 6.6:**  $^1\text{H}$  NMR spectra of compounds **4c**, **3b** and their 1:1 mixture (10 mmol/l) in a 3:1 DMF- $d_7$  and DCM- $d_2$  solution at 175 K



**Figure 6.7:** Structures of hypothetical complexes of isocytosine **2,4-4a** with 1-substituted thymine **2b** and **E-1,3-4a** with 2,6-diaminopyridine (**DAP**). Neither of these complexes was observed experimentally.

---

Naturally, we were also interested in the formation of hydrogen-bonded complexes by less stable tautomers **2,4** (enol, DAD) and **E-1,3** (imino, ADA) with suitable partners (see fig. 6.7). Isocytosine (**4a**) was selected as the best candidate with the lowest, albeit still not very promising calculated relative energies of mentioned tautomers (see table 6.1).

Experiments were carried out with mixtures of isocytosine with modified thymine **2b** in hopes to see complex formation employing tautomer **2,4-4a** and with 2,6-diaminopyridine (**DAP**) to promote a complex with **E-1,3**, but to no avail. The failure to observe any kind of complex is supported by highly unsatisfactory calculated  $E_{\text{stabilization}}$ , being  $-47.9$  kJ/mol for complex of tautomer **2,4-4a** with substituted thymine **2b** and  $-32.6$  kJ/mol for complex of tautomer **E-1,3** with **DAP**.

In conclusion, we have been successful at theoretical determination of the effects responsible for the relative stability of guanine and isocytosine analogues/derivatives and delivered an experimental proof thereof. No other tautomers of guanine aside from the canonical **3,2** (3H-ketoamino) form were observed under any condition.

# Chapter 7

## Technical and Experimental Data

### 7.1 Hardware

All experiments were performed on an Oxford Instruments 500MHz (11.7T) superconducting magnet equipped with a Bruker triple resonance broad-band probe with ATM (5 mm PATBO BB- $^1\text{H}/^{19}\text{F}/\text{D}$  Z-GRD) connected to B-VT 3200 (for low temperatures) and BCU-05 (for normal and high temperatures) temperature control units, accompanied by a Bruker Avance II 500 (two channels, BOSS II (34 magnetic field homogeneity corrections), BSMS 2, amplifiers BLAXH 300/100, BLAXH 300 and BLAX 500) console.

### 7.2 Software

- Gaussian 03 revision C.02 (calculations)
- Gaussian 16 revision C.01 (calculations)
- GaussView 6 (molecular modelling)
- HyperChem 6 (molecular modelling)
- MestReNova 12.0.3 (data processing)
- TopSpin 3.2 (data acquisition)
- TopSpin 4.1 (dNMR simulations)



---

## 7.3 Chemicals

All experiments were performed in solutions of commercially available deuterated solvents. N,N-dimethylformamide- $d_7$  and methanol- $d_4$  were purchased from Eurisotop in bottles and dichloromethane- $d_2$  was obtained from Merck in ampoules. Compounds **1a** and **1b** were prepared by Ing. Lucie Čechová and compounds **2b** and **3b** were prepared by RNDr. Michal Šála, Ph.D. (see ESI and references in [1] for synthesis and characterization). All other physically used compounds were purchased directly from Sigma-Aldrich.

## 7.4 Sample Preparation

Weighted samples were dissolved in deuterated solvents to obtain solutions with concentrations of 10, 20, 30, 40 or 50mmol/l. Increased temperatures and ultrasonic bath were often used to speed up the dissolution. Should a mixture of two compounds be prepared, the compound with worse solubility would be dissolved first and the solution would then be used to dissolve the second compound. The mixtures were prepared in stoichiometric ratios with increasing multiples of one compound with the other one having constant concentration. In cases where solvent mixtures were used to obtain low melting point aprotic environments, N,N-dimethylformamide- $d_7$  and dichloromethane- $d_2$  were mixed during sample preparation in volumetric ratios of 1:1 at first [1], later testing then revealed that 1:3 ratio also provides sufficiently low melting point while being easier to work with [2]. In these mixtures, the trace signal of N,N-dimethylformamide with chemical shift of 2.75 ppm was used as an internal reference. Samples in methanol- $d_4$  solutions were referenced to the trace signal of methanol at 3.31 ppm.

## 7.5 Calculations

All DFT calculations related to [1] were ran at B3LYP/6-31G(d,p) level using the polarizable continuum model SCRF with GD3 empirical dispersion correction in Gaussian 03 [14] program package. Computations related to [2] were ran at B3LYP/6-311++G(2df,2pd) level using the SCRF PCM implicit solvent model (N,N-dimethylformamide) with GD3 empirical dispersion correction in Gaussian 16 [15] program package. The oligoethyleneglycol groups in structures **2b** and **3b** were substituted by a methyl group to simplify the calculations, effectively using structures **2h** and **3a**.

# Bibliography

- [1] J. R. Štoček, K. Bártová, L. Čechová, M. Šála, O. Socha, Z. Janeba, M. Dračínský, *Chemical Communications* **2019**, *55*, 11075–11078.
- [2] J. R. Štoček, M. Dračínský, *Biomolecules* **2020**, *10*, DOI 10.3390/biom10020170.
- [3] I. I. Rabi, J. R. Zacharias, S. Millman, P. Kusch, *Physical Review* **1938**, *53*, 318–318.
- [4] E. M. Purcell, H. C. Torrey, R. V. Pound, *Physical Review* **1946**, *69*, 37–38.
- [5] F. Bloch, W. W. Hansen, M. Packard, *Physical Review* **1946**, *70*, 474–485.
- [6] M. H. Levitt, *Spin Dynamics*, 2nd, Wiley, UK, **2008**.
- [7] N. E. Jacobsen, *NMR Spectroscopy Explained: Simplified Theory, Applications and Examples for Organic Chemistry and Structural Biology*, 1st, Wiley, USA, **2007**.
- [8] R. M. Silverstein, F. X. Webster, D. J. Kiemle, *Spectrometric Identification of Organic Compounds*, 7th, Wiley, USA, **2005**.
- [9] T. D. W. Claridge, *High Resolution NMR Techniques in Organic Chemistry*, 3rd, Elsevier, UK, **2016**.
- [10] R. A. Friesner, *Proceedings of the National Academy of Sciences* **2005**, *102*, 6648–6653.
- [11] P. Deglmann, A. Schäfer, C. Lennartz, *International Journal of Quantum Chemistry* **2015**, *115*, 107–136.
- [12] R. A. Mata, M. A. Suhm, *Angewandte Chemie International Edition* **2017**, *56*, 11011–11018.
- [13] E. Schrödinger, *Physical Review* **1926**, *28*, 1049–1070.

- 
- [14] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. A. Montgomery, Jr., T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, C. Gonzalez, J. A. Pople, Gaussian~03 Revision C.02, Gaussian Inc. Wallingford CT, **2004**.
- [15] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman, D. J. Fox, Gaussian~16 Revision C.01, Gaussian Inc. Wallingford CT, **2016**.
- [16] R. Fletcher, *The Computer Journal* **1964**, *7*, 149–154.
- [17] E. Polak, G. Ribiere, *Revue française d'informatique et de recherche opérationnelle. Série rouge* **1969**, *3*, 35–43.
- [18] J. B. F. A. Foresman, *Exploring Chemistry with Electronic Structure Methods*, 3rd, Gaussian Inc., USA, **2015**.
- [19] I. N. Levine, *Quantum Chemistry*, 7th, Pearson, USA, **2014**.

- [20] R. O. Jones, *Reviews of Modern Physics* **2015**, *87*, 897–923.
- [21] D. R. Hartree, *Mathematical Proceedings of the Cambridge Philosophical Society* **1928**, *24*, 89–110.
- [22] A. D. Becke, *The Journal of Chemical Physics* **1993**, *98*, 1372–1377.
- [23] C. Lee, W. Yang, R. G. Parr, *Physical Review B* **1988**, *37*, 785–789.
- [24] S. F. Boys, *Proceedings of the Royal Society of London Series A Mathematical and Physical Sciences* **1950**, *200*, 542–554.
- [25] L. Onsager, *Journal of the American Chemical Society* **1936**, *58*, 1486–1493.
- [26] J. Tomasi, B. Mennucci, R. Cammi, *Chemical Reviews* **2005**, *105*, 2999–3094.
- [27] A. Bondi, *The Journal of Physical Chemistry* **1964**, *68*, 441–451.
- [28] S. Grimme, J. Antony, S. Ehrlich, H. Krieg, *The Journal of Chemical Physics* **2010**, *132*, 154104.
- [29] E. Arunan, G. R. Desiraju, R. A. Klein, J. Sadlej, S. Scheiner, I. Alkorta, D. C. Clary, R. H. Crabtree, J. J. Dannenberg, P. Hobza, H. G. Kjaergaard, A. C. Legon, B. Mennucci, D. J. Nesbitt, *Pure and Applied Chemistry* **2011**, *83*, 1637–1641.
- [30] D. Chen, N. Oezguen, P. Urvil, C. Ferguson, S. M. Dann, T. C. Savidge, *Science Advances* **2016**, *2*, e1501240.
- [31] J. Nowakowski, I. Tinoco, *Seminars in Virology* **1997**, *8*, 153–165.
- [32] J. D. Watson, F. H. C. Crick, *Nature* **1953**, *171*, 737–738.
- [33] P. Atkins, J. de Paula, *Fyzikální Chemie*, VŠCHT Praha, Czech Republic, **2013**.
- [34] B. Alberts, D. Bray, A. Johnson, J. Lewis, M. Raff, K. Roberts, P. Walter, *Základy buněčné biologie*, Espero, Czech Republic, **2005**.
- [35] C. Fernandez, *Current Opinion in Structural Biology* **2003**, *13*, 570–580.
- [36] K. Pervushin, A. Ono, C. Fernandez, T. Szyperski, M. Kainosho, K. Wuthrich, *Proceedings of the National Academy of Sciences* **1998**, *95*, 14147–14151.
- [37] S. P. Brown, M. Pérez-Torrallba, D. Sanz, R. M. Claramunt, L. Emsley, *Journal of the American Chemical Society* **2002**, *124*, 1152–1153.
- [38] S. F. Boys, F. Bernardi, *Molecular Physics* **1970**, *19*, 553–566.
- [39] M. Gutowski, J. G. C. M. Van Duijneveldt-Van De Rijdt, J. H. Van Lenthe, F. B. Van Duijneveldt, *The Journal of Chemical Physics* **1993**, *98*, 4728–4737.

- 
- [40] B. Brauer, M. K. Kesharwani, J. M. L. Martin, *Journal of Chemical Theory and Computation* **2014**, *10*, 3791–3799.
- [41] G. P. Moss, *Pure and Applied Chemistry* **1996**, *68*, DOI 10.1351/pac199668122193.
- [42] V. Singh, B. I. Fedeles, J. M. Essigmann, *RNA* **2015**, *21*, 1–13.
- [43] P. Taylor, G. Van der Zwan, L. Antonov, *Tautomerism: Methods and Theories*, Wiley, **2014**.
- [44] F. Crick, *Nature* **1970**, *227*, 561–563.
- [45] C. Fonseca Guerra, H. Zijlstra, G. Paragi, F. M. Bickelhaupt, *Chemistry - A European Journal* **2011**, *17*, 12612–12622.
- [46] J. Florián, J. Leszczyński, *Journal of the American Chemical Society* **1996**, *118*, 3010–3017.
- [47] R. Pohl, O. Socha, P. Slavíček, M. Šála, P. Hodgkinson, M. Dračínský, *Faraday Discussions* **2018**, *212*, 331–344.
- [48] R. Pohl, O. Socha, M. Šála, D. Rejman, M. Dračínský, *European Journal of Organic Chemistry* **2018**, *2018*, 5128–5135.
- [49] M. Dračínský, P. Jansa, K. Ahonen, M. Buděšínský, *European Journal of Organic Chemistry* **2011**, *2011*, 1544–1551.