

# NMR studies of host-pathogen interactions

**Katja Petzold**



**Department of Medical Biochemistry and Biophysics**  
Umeå University  
Umeå 2009

Copyright©Katja Petzold  
ISBN: 978-91-7264-816-6  
ISSN: 0346-6612  
Cover picture: Vanessa Kunkel  
Printed by: VMC, KBC, Umeå University,  
901 87 Umeå, Sverige, 2009.

I know that I know nothing  
Socrates

Coffee, Tea, vodka?  
Rima Sulniute

*To my family, past and present*



# Table of Contents

<b>1</b>	<b>List of Papers</b> .....	<b>6</b>
<b>2</b>	<b>Abbreviations</b> .....	<b>7</b>
<b>3</b>	<b>Abstract</b> .....	<b>8</b>
<b>4</b>	<b>Introduction &amp; Aim of the Thesis</b> .....	<b>9</b>
<b>5</b>	<b>Regulatory “epsilon” RNA of the <i>Hepatitis B virus</i></b> .....	<b>11</b>
5.1	The <i>Hepatitis B virus</i> and its lifecycle.....	11
5.1.1	<i>Epsilon interacts with the viral reverse transcriptase</i> .....	12
5.2	Structural biology of RNA .....	14
5.3	NMR of RNA .....	15
5.4	NMR studies of the apical loop of the epsilon RNA.....	17
<b>6</b>	<b>Lipids of <i>Helicobacter pylori</i></b> .....	<b>18</b>
6.1	<i>Helicobacter pylori</i> and its infection .....	18
6.2	Biological role of lipids .....	20
6.3	Background: <i>H. pylori</i> lipids.....	24
<b>7</b>	<b>NMR background</b> .....	<b>25</b>
7.1	Spins and Spectrometer .....	25
7.2	Manipulation of spins .....	27
7.3	Spin interaction.....	28
7.3.1	<i>Chemical Shift</i> .....	28
7.3.2	<i>J coupling</i> .....	30
7.3.3	<i>Dipolar interactions</i> .....	30
7.3.4	<i>Relaxation</i> .....	31
7.4	Design of NMR experiments .....	33
7.4.1	<i>COSY (correlation spectroscopy)</i> .....	34
7.4.2	<i>ROESY: (Rotating frame Overhauser Effect spectroscopy)</i> .....	35
7.4.3	<i>Heteronuclear relaxation reveals fast intramolecular motion</i> .....	37
<b>8</b>	<b>Conclusion and Outlook</b> .....	<b>39</b>
	Paper I: Relaxation measurements of the <i>HBV</i> apical loop .....	39
	Paper II: EASY ROESY – distance measurement for medium-sized molecules.....	40
	Paper III & IV: Design of the SCT P,H-COSY experiment and characterization of <i>H. pylori</i> phospholipids .....	40
<b>9</b>	<b>Acknowledgment</b> .....	<b>42</b>
<b>10</b>	<b>References</b> .....	<b>44</b>

# 1 List of Papers

This thesis is based on the following papers, which are referred to in the text by their Roman numerals I-IV.

**I:** **Petzold K.**, Duchardt E., Flodell S., Larsson G., Kidd-Ljunggren K., Wijmenga S. and Schleucher J.; Conserved nucleotides in an RNA essential for *hepatitis B virus* replication show distinct mobility patterns  
Nucleic Acids Research 2007; 35(20):6854-61.

**II:** Thiele C.M., **Petzold K.**, Schleucher J.; EASY ROESY: Reliable cross-peak integration in adiabatic symmetrized ROESY  
Chemistry – A European Journal 2009; 15(3):585-8

**III:** **Petzold K.**, Olofsson A., Arnqvist A., Gröbner G., Schleucher J.; Semi-constant-time P,H-COSY NMR provides direct insight into the phospholipid pattern associated with the virulence of *Helicobacter pylori*.  
under revision

**IV:** Olofsson A., Vallström A., **Petzold K.**, Schleucher J., Carlsson S., Haas R., Backert S., Nyunt Wai S., Gröbner G., Arnqvist A.; Characterization of *Helicobacter pylori* vesicles and their cognate properties for intimate host interactions  
submitted

Paper I and II were reprinted with kind permission from the publishers

Papers not included in this Thesis:

a: Robertsson J., **Petzold K.**, Löfvenberg L., Backman L.; Folding of spectrin's SH3 domain in the presence of spectrin repeats.  
Cellular & Molecular Biology Letters 2005; 10(4):595-612.

b: **Petzold K.**, Öhman A., Backman L.; Folding of the  $\alpha$ II-spectrin SH3 domain under physiological salt conditions.  
Archives of Biochemistry and Biophysics 2008; 474(1):39-47.

## 2 Abbreviations

CL	cardiolipin
COSY	correlation spectroscopy
CPG	cholesteryl phosphoglycoside
DD	dipole-dipole (e.g. in dipole-dipole coupling)
DNA	deoxyribonucleic acid
$\epsilon$	epsilon
EASY	efficient adiabatic symmetrized
<i>HBV</i>	<i>Hepatitis B Virus</i>
<i>H. pylori</i>	<i>Helicobacter pylori</i>
INEPT	insensitive nuclei enhanced by polarization transfer
LPS	lipopolysaccherides
lysoPL	lysophospholipid
NMR	Nuclear Magnetic Resonance
NOESY	nuclear Overhauser effect spectroscopy
P,H-COSY	$^{31}\text{P}$ , $^1\text{H}$ correlation spectroscopy
PA	phosphatidic acid
PC	phosphatidylcholine
PE	phosphatidylethanolamine
pgRNA	pregenomic RNA
PL	phospholipid
ppm	parts per million
PS	phosphatidylserine
RDC	residual dipolar coupling
RNA	ribonucleic acid
ROESY	rotating frame nuclear Overhauser effect spectroscopy
RT	reverse transcriptase
TOCSY	total correlation spectroscopy

### 3 Abstract

This thesis describes the use of Nuclear Magnetic Resonance (NMR) for characterizing two host-pathogen interactions: The behavior of a regulatory RNA of the *Hepatitis B virus (HBV)* and the attachment of *Helicobacter pylori (H. pylori)* to the gastric mucosa. NMR is a powerful tool in biomedicine, because molecules ranging from small ligands to biomacromolecules can be studied with atomic resolution. Different NMR experiments are designed to determine structures, or to monitor interactions, folding, stability or motion.

Paper I describes the analysis of the motions of a regulatory RNA of *HBV*. The NMR structure of the RNA had revealed before that several well-conserved nucleotides adopt multiple conformations. Therefore an analysis of possible underlying motions was undertaken using two different NMR techniques, one of which (off-resonance ROESY) was applied to nucleic acids for the first time. The observed motions suggest an explanation why the structurally poorly defined nucleotides are highly conserved.

In paper II we improved the ROESY NMR experiment, which is used to measure internuclear distances for structure determination of medium-sized molecules. Using a small protein and an organometallic complex as examples, we demonstrated that the new EASY ROESY experiment yields clean spectra that can directly be integrated to derive interatomic distances.

*H. pylori*, the bacterium involved in peptic ulcer disease and gastric cancer, survives in the harsh acidic environment of the stomach. It possesses many membrane proteins which mediate adherence, raising the question, if their activity is related to membrane composition. In paper III & IV we analyzed therefore the phospholipid composition of *H. pylori* membranes.

In paper III, an advanced method for the analysis of the phospholipid composition of biological membranes was developed. The two-dimensional semi-constant-time  $^{31}\text{P},^1\text{H}$ -COSY experiment combines information from phosphorus and hydrogen atoms of phospholipids for their unambiguous identification. Furthermore, the high resolution of the two-dimensional experiment allows the quantification of phospholipids where conventional methods fail.

In paper IV we applied the new experiment to analyze the lipid composition of whole *H. pylori* cells, their inner and outer membranes, and of vesicles shed by the bacterium. The goal of this study was to characterize the vesicles which are suggested to play a role in the inflammation process. We established that the outer membrane and the vesicles have similar phospholipid compositions, suggesting that the vesicles are largely derived from the outer membrane.

The NMR results presented here elucidate details of molecular systems engaged in pathogenicity, as basis for therapeutic strategies against these pathogens.

Keywords: NMR, regulatory RNA, *HBV*, *H. pylori*, phospholipids, ROESY

## 4 Introduction & Aim of the Thesis

The structure of a molecule is intimately linked to its function – this is a paradigm of biology and chemistry, leading to the field of structural biology. Different biophysical techniques can be used to observe molecular structure on a variety of resolutions – e.g. cryo-Electron Microscopy (structure of the ribosome with a resolution of 8Å (Mitra et al. 2006)), or Circular Dichroism spectroscopy (showing overall secondary structure of chiral molecules (Kelly et al. 2000)). But only two methods can deliver atom-size resolution: X-ray Crystallography and Nuclear Magnetic Resonance (NMR).

A wide range of NMR observables delivers information on the chemical environment of each atom in a molecule. Using internuclear distances, local conformation and orientation of domains, molecular structure is determined at atomic resolution. Other advantages of NMR are the possibility to analyze complex structures under native like conditions (even weak interactions of highly complex systems can be retained), the complete recovery of the intact sample for possible further analysis, and the possibility to simultaneously quantify components of complex mixtures and derive structural information on the components. But structural information is not enough for explaining the function of a molecule, because the structure of a pair of scissors does not explain how they cut. Therefore information in addition to the structure-function paradigm became important – motion. Only motion can explain how bonds are broken and formed, molecules fold or interact with each other accompanied by the deformation of the interaction surface. NMR is the only method which can report on molecular motion over the whole range of time scales from bond vibrations to macromolecular folding (Al-Hashimi 2007; Shajani et al. 2007).

High resolution – liquid state NMR is a relatively young spectroscopy and has developed tremendously over the last decades. Key steps on the way to become a modern tool in chemistry and biology were the discovery of the absorption of radio frequency by the nuclear spin in a matching magnetic field (Bloch et al. 1946), the discovery of the chemical shift effect (Dickinson 1950), Fourier-Transform NMR (Ernst et al. 1966; Aue et al. 1976) and two-dimensional (2D) NMR (Nagayama et al. 1977). Modern biomolecular NMR became feasible through technical progress leading to extremely stable and strong (over 20 Tesla) helium-cooled superconductive magnets, in concert with stable isotope labeling of biomolecules (Freeman 1995).

The aim of my PhD, and therefore this thesis, was to apply NMR to different biological systems. I wanted to analyze a range of molecules and extract a variety of information. For this purpose, NMR experiments were designed or modified, to be able to address questions on disease-related systems.

**Specific Aims:**

- Monitor the phospholipid composition of the bacterium *H. pylori* and identify possible relationships between membrane protein activity and membrane composition during different bacterial growth phases
- Simplify the ROESY experiment to extract more reliable distance information of medium-sized molecules for the structure determination process
- Identify patterns of motion of a viral RNA using relaxation NMR experiments, which could be linked to the RNAs protein-binding mechanism

## 5 Regulatory “epsilon” RNA of the *Hepatitis B virus*

### 5.1 The *Hepatitis B virus* and its lifecycle

The *Hepatitis B virus* (*HBV*) belongs to the family of the hepadnaviruses and has a 3,2 kilobase-big genome consisting of partially double-stranded DNA, so called relaxed circular DNA (rcDNA). During replication, *HBV* passes a stage of RNA and is therefore classified as a pararetrovirus (Nassal 2008). The genome of the virus codes for only 7 proteins in 4 open reading frames: 3 versions of the envelope protein, the nucleocapsid protein core, the secretory hepatitis ‘B e’ antigen, the X-protein and the most important protein ‘P-protein’, containing the reverse transcriptase (RT) domain (Ghany et al. 2007).

*HBV* has infected the liver of ca. 2 billion people and about 150 million of them are chronic carriers and are at risk for chronic hepatitis, cirrhosis and primary liver cancer. *HBV* causes approximately 800 000 deaths each year, with liver cancer being the 5<sup>th</sup> most frequent cancer in the world (Lavanchy 2008). Vaccines against the envelope proteins of the virus are existing, but up to 10% of the population are non-responders and because these envelope proteins are non-essential, high escape mutation rates impair the vaccines further (Glebe et al. 2007). Treatment is possible by using either cytokines, which are antiviral and immunomodulatory, or nucleos(t)ide analogs, which block the reverse transcriptase step (see Fig. 1). Both are not perfect drugs: Cytokines have severe side-effects and are hard to apply and application of nucleotide analogs leads to drug resistance. Further complications arise from the extraordinary high mutation rate, meaning that every single-base can be changed in one day, considering that the length of treatment is several years. Unfortunately, not even the active center of the RT is extremely sensitive to mutation, often double mutations occur, where one mutation escapes the drug and the second reconstitutes the function (Ghany et al. 2007).

A *HBV* infection starts by docking onto the cell to yet unknown receptors and thereafter the internalization of the virus, possibly via endocytosis. Once in the cell, the virus gets stripped of its envelope proteins and the relaxed circular (rc)DNA containing capsid gets transported to the nucleus (Fig. 1). Here the rcDNA gets repaired to covalently closed circular (ccc)DNA, while the P-protein, which is covalently linked to the rcDNA, is released. This cccDNA is an extremely stable plasmid, present in many copies and it can even be integrated into the host genome. It often resists the treatment, hence the need for long-term treatment. In a next step the cccDNA is transcribed by the host polymerase into pregenomic (pg)RNA and viral messenger (m)RNA, which then gets exported to the cytosol and translation of viral

proteins occurs. In a next step the P-protein and the nucleocapsid core interact with a stem structure of the 5' pgRNA, called epsilon ( $\epsilon$ ), triggering encapsidation and reverse transcription (see next chapter), which also requires additional host-factors. After complete reverse transcription the new capsid containing again rcDNA linked to the P-protein, which then either gets coated by the envelope proteins and excreted, or it enters into the nucleus to enrich the pool of cccDNA (Glebe et al. 2007; Nassal 2008).

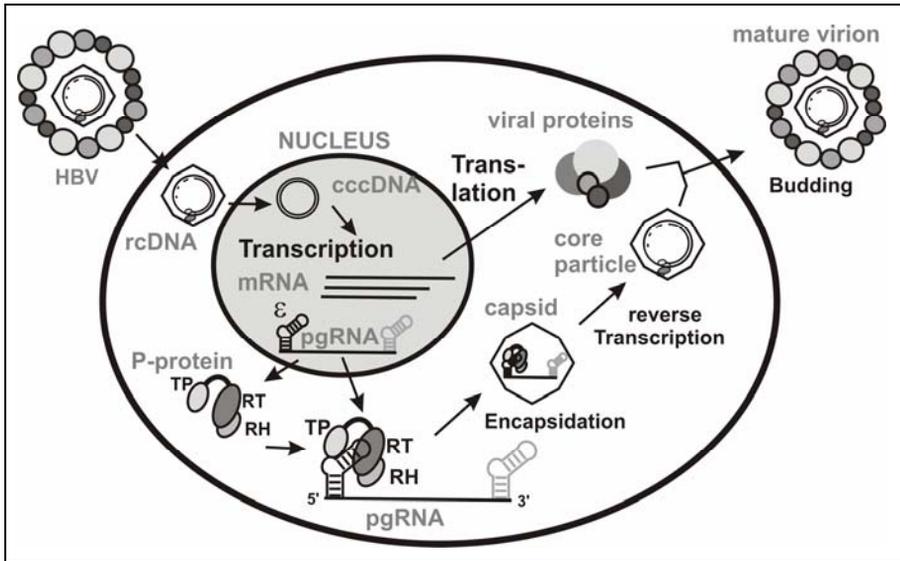


Figure 1: Schematic lifecycle of the *Hepatitis B Virus*. The pregenomic RNA (pgRNA) plays a central role in the replication of new viruses by interacting with the reverse transcriptase (RT) complex and so allowing transcription of new viral DNA and at the same time triggering encapsidation (description of the viral life cycle see text).

### 5.1.1 Epsilon interacts with the viral reverse transcriptase

One of the two main components in the RNA-protein interaction is the 90 kDa P-protein. The *HBV* P-protein is quite unique, because it contains three functional subunits: First, the reverse transcriptase (RT), a protein all retroviruses code in their genome for. Second, a RNase H domain that degrades the pgRNA upon DNA synthesis. Third, the terminal protein (TP) domain, connected via a flexible linker. Via a highly conserved tyrosine, the TP domain carries the covalently bound primer, leading to a unique protein priming RT reaction (Beck et al. 2007). The RT and the RNase H domain are well conserved among Hepadnaviridae and other retroviruses such as HIV, but there is absolutely no conservation of the TP domain, which is also exceptionally hydrophobic and therefore hard to express (Ghany et al. 2007).

Until now, it was not possible to over-express functional human RT, which has hampered studies of its interaction with RNA.

The second main component in the interaction leading to reverse-transcription is the 5' structure of the pgRNA, the stem-loop epsilon ( $\epsilon$ ).  $\epsilon$  is a 60-nucleotide (nt) regulatory RNA present at the 5' and 3' ends of the pgRNA. The  $\epsilon$  stem loop has a six-nucleotide bulge, which is the template for the primer synthesis, a lower stem which is not conserved and therefore thought to be a structural spacer requirement, and the 27-nucleotide apical loop (Fig. 2). The whole  $\epsilon$  is necessary for priming competent binding, but for simple binding of the P-protein, the apical loop is not required in human *HBV* (Beck et al. 2007). Therefore the hypothesis was recently developed, that other host factors might interact with the apical loop and initiate a structural change necessary for encapsidation and completion of reverse transcription. However, the postulated interaction partner has not been identified yet, because nobody has yet succeeded to produce a fully functional in vitro model (Nassal 2008; Hu et al. 2009).

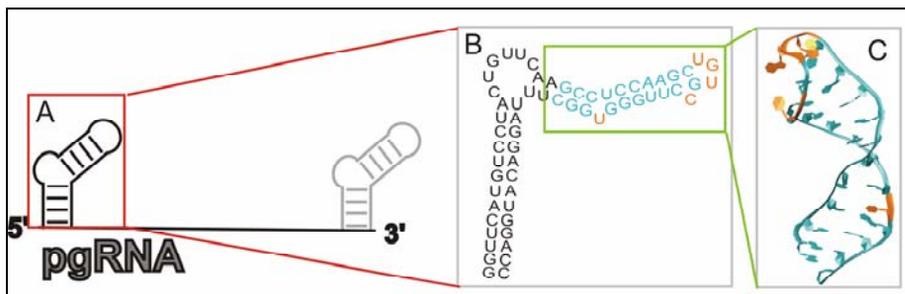


Figure 2: A: Sketch of the structure of the pregenomic RNA (pgRNA) with the 5' epsilon loop marked by a red box. B: Secondary structure of the epsilon loop with the apical loop highlighted by the green box. C: The tertiary structure of the apical loop as published previous in our group (Flodell et al. 2006), stem nucleotides are shown in turquoise, and nucleotides in non-helical structures (loop nucleotides, C<sub>16</sub> and the bulged U<sub>23</sub>) in orange.

The whole  $\epsilon$  stem-loop is highly conserved (Flodell et al. 2002; Hu et al. 2004; Beck et al. 2007), however, several nucleotides of the apical loop are poorly defined in the structure. The related duck *HBV* forms an analogous P-protein- $\epsilon$  complex, which contains a partially melted apical loop and is priming competent without further factors (Girard et al. 2007). This leads to the question if a defined RNA structure is recognized by the P-protein, or a deformable structure with characteristic motional and thermodynamic parameters. The latter binding mechanism has been observed for the HIV-TAR RNA (Al-Hashimi 2005). The binding mechanism in the p- $\epsilon$  interaction will be revealed by studies of intramolecular motion of the apical

loop. By comparison with a structure of the complex further detailed information can be added, when becoming available.

## 5.2 Structural biology of RNA

Ribonucleic acid (RNA) is a long, linear polymer of only four different nucleotides and is assembled by the RNA polymerase. Nucleotides consist of a phosphate group, a ribose and a base (Fig. 3). The phosphate group is esterified to the nucleotide's own ribose at the C<sub>5'</sub> OH and to the next ribose at the C<sub>3'</sub> OH, to form the phosphodiester backbone. Four different bases are connected via a N-glycosidic bond to the C<sub>1'</sub> and give the identity/specificity to each nucleotide: Adenine (A) base pairs via two hydrogen bonds with Uracil (U) and Guanine (G) base pairs via three hydrogen bonds with Cytosine (C), giving rise to the classical Watson-Crick base pairs. The base pairs stack on top of each other and form the canonical, rigid structure of the oligomer – the double helix – two RNA strands linked via base pairs (Stryer et al. 2002; Hall 2008) (Fig. 2B & C). RNA compared to DNA poses an OH group on the C<sub>2'</sub> of the ribose and is therefore more sensitive to cleavage of the oligomer by an ester exchange from the C<sub>5'</sub> to the C<sub>2'</sub>, building a cyclic-phosphorous ester. On the other hand, the C<sub>2'</sub> OH group also adds structural and catalytic features which allow RNA to adopt structures and perform functions that are unknown for DNA.

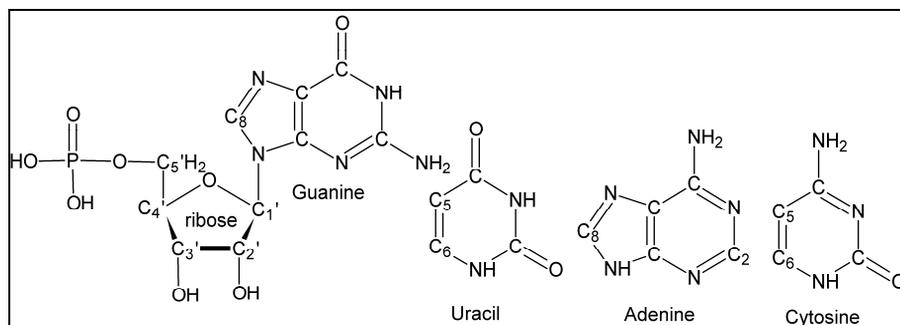


Figure 3: Drawing of a nucleotide and of the three other bases occurring in RNA. Selected C atoms are numbered according to NMR standards, besides for C<sub>5'</sub> the protons in C-H groups were omitted for clarity. The nucleotides are the building blocks for the RNA structure shown in Fig. 2. Formulas were drawn with ChemSketch.

While RNA can form a double helix, it does not normally exist as two complementary strands, but is transcribed as a single strand. Single-stranded RNA can adopt different flexible structures besides a helix, such as stem-loops, hairpins, bulges or complex three dimensional structures (see Fig. 2). These structures play fundamental biological roles, as e.g. in the ribosome or in the spliceosome. Non-Watson-Crick base pairs are central building blocks

of these non-helical structures (Leontis et al. 2003), which are of importance for interactions with other molecules.

The functions of RNA can be divided in 3 classes: First, transport of genetic information, as performed by messenger (m)RNA. Second, catalytic activity performed by structures called ribozymes, with ribosomal (r)RNA, RNase P or self-splicing RNA, as prominent examples. Third, regulatory RNAs, such as small interfering RNAs (siRNA), riboswitches (Reichow et al. 2006; Schwalbe et al. 2007) and RNA thermometers (Stryer et al. 2002; Narberhaus et al. 2006; Waters et al. 2009). Often two functions are combined in one RNA polymer, for example when a regulatory structure is present in the 5' untranslated region (UTR) of an mRNA regulating the translation of the mRNA. This regulation in turn can be achieved e.g. by ligand binding or by a thermal transition. RNA functions are often accompanied by drastic conformational changes, therefore when thinking of RNA one must not consider it as a single structure, but must see it as an ensemble of highly dynamic conformations (Al-Hashimi 2007). One such regulatory RNA with a complex function is the  $\epsilon$  RNA of *HBV* (see chapter 5.4).

### 5.3 NMR of RNA

The classical application of NMR is the determination of molecular structures. Structure determination is based on a complete assignment of all signals to their corresponding atoms in the molecule. This approach is fully applicable to RNA, but several properties of RNA make it a challenge: RNA is only built up of four different nucleotides, which in fact only differ in their bases. It has a low  $^1\text{H}$  density compared to proteins, leading to a low number of  $^1\text{H}$ - $^1\text{H}$  distances that can be measured by NOESY and used as distance constraints in structure determination. Finally, large parts of RNA structures exist as stable A-form double helix, with poor resolution of NMR signals. All this leads to a massive overlap in resonances, especially in the ribose moiety and the phosphate backbone, which have small chemical shift dispersion to begin with.

Therefore the assignment and the extraction of structural constraints are challenging. On the other hand, stretches of A-form helix provide starting points for signal assignment: The signals of imino hydrogens in the bases are visible because the hydrogen bonding of the base pairs protects them from solvent exchange. Furthermore, imino hydrogens of consecutive base pairs are close in space, so that in a NOESY experiment, cross-peaks connect the imino signal of each base pair to the imino signals of both adjacent base pairs. Therefore one can follow the sequence of imino-imino cross-peaks, to assign the signals along the nucleotide sequence, which is called a sequential walk. The imino signals also allow examination of folding and stability of

the RNA molecule with a simple one-dimensional  $^1\text{H}$  spectrum, because the number of imino signals equals the number of base pairs. The non-helical secondary structure elements (loops, bulges, kinks, etc – Fig 2) often interact with ligands and are therefore biological relevant. These secondary structures have more highly resolved chemical shifts and are consequently easier to assign. As they usually do not contain base-pairs, their imino signals cannot normally be observed. Due to the more modular built of RNA as compared to proteins, certain secondary structure elements have the same chemical shift properties in different molecules, and are therefore easy to discover. For complete assignment of RNA molecules, isotope labeling with  $^{13}\text{C}$  and  $^{15}\text{N}$  is required, so that advanced multidimensional heteronuclear NMR experiments can be applied (Furtig et al. 2003). Structure determination requires further experiments to define local conformations, such as the sugar pucker of the ribose, the conformation of the phosphate backbone and the glycoside torsion angle between the base and ribose. A long-standing problem for NMR structure determination originates from the fact that all classical NMR structure parameters, such as distances and dihedral angles, report only on local structures. Therefore relative orientation of secondary structure elements is hard to determine, especially in non-globular molecules such as RNA. To constrain global molecular shape, such as the relative orientation of two helices, residual dipolar couplings (RDCs) have recently been introduced (Furtig et al. 2003). Using the whole range of isotope labeling and NMR techniques, structure determination of RNA molecules up to about 100 nucleotides can be tackled today.

But there is much more to NMR of RNA besides structure determination. A structure alone does not explain RNA function such as mechanisms of binding or catalysis; the timescale and amplitudes of motion must also be described to explain biological activity (Getz et al. 2007). A wide range of NMR experiments can be employed to monitor motion ranging from picoseconds to slower than seconds (paper I). Motions of different entities, such as a bond, a nucleotide or a whole domain, can give rise to weakly populated conformations that can nevertheless be crucial for biological function (Al-Hashimi 2007; Shajani et al. 2007). Using other dedicated NMR experiments, it is possible to define ligand binding surfaces (e.g. chemical shift perturbation (Zuiderweg 2002)). Bound metal ions can be localized or structural water molecules found, even folding pathways of RNA structures can be investigated (Furtig et al. 2003). However, few comprehensive studies of folding or motions of RNAs have been performed to date (Hall 2008).

## 5.4 NMR studies of the apical loop of the epsilon RNA

The 27-nucleotide apical loop of *HBV*  $\epsilon$  RNA forms a stable structure (melting point around 65°C) which was recently solved in our group (Fig. 2C). It consists of two A-form helices, which are bent by approximately 20° relative to each other, induced by the bulged U<sub>23</sub> (Flodell et al. 2002; Flodell et al. 2006). It was discovered to possess a pseudo-triloop, consisting of the loop nucleotides U<sub>12</sub>, G<sub>13</sub> and U<sub>14</sub>, a closing base pair (C<sub>11</sub>G<sub>15</sub>) and the bulged C<sub>16</sub>. This pseudo-triloop might be an important protein binding motif. Unfortunately, the exact molecular role of the pseudo-triloop is not yet understood, mainly because the P-protein cannot be handled in vitro and other probably required interaction partners are not yet identified. Although the apical loop is not needed for binding of the RT, it is necessary for priming (see chapter 5.1.1). In the human apical loop, several loop nucleotides were found to be structurally ill-defined (Flodell et al. 2006), yet these nucleotides are highly conserved and therefore essential for function. The apical loop of duck *HBV*, a model system for the human apical loop, may give a hint for an explanation for this puzzle. The duck apical loop is unstable, so that at body temperature a part of the upper helix has melted unstructured (Girard et al. 2007). Similarly, the structurally ill-defined nucleotides in the human apical loop might be mobile and consequently prone to sample minor conformations, which then might be trapped upon binding of the reverse transcriptase in a conformational capture mechanism. We therefore analyzed motions of this RNA and found that nucleotides U<sub>12</sub>, G<sub>13</sub>, C<sub>16</sub> and U<sub>23</sub> are highly mobile (paper I).

## **6 Lipids of *Helicobacter pylori***

### **6.1 *Helicobacter pylori* and its infection**

The Gram-negative bacterium *Helicobacter pylori* (*H. pylori*) is the oldest known and most successful human pathogen (Chitcholtan et al. 2008). It colonizes the human stomach in a lifelong fashion, by adhering to the gastric mucosa. Infection often occurs in childhood and more than 50% of the population is a carrier. The bacterium causes gastric inflammation which leads in about 15% of the cases to peptic ulcer disease and in some of those cases to gastric cancer (Atherton 2006). It is responsible for more than 60% of all gastric cancers, in total 8% of all cancers, and was therefore declared as a class I carcinogen by the World Health Organization (Parkin 2004).

Currently, treatment of *H. pylori* infection consists of a triple combination of antibiotics. First resistances have already emerged. It was found that the outcome of the infection depends on the bacterial strain and the host genome. Interestingly, it is unclear if *H. pylori* infection is only negative. Because only a fraction of the infected individuals experiences symptoms, therefore it is speculated that *H. pylori* can also be beneficial to the host, by preventing other bacterial infections or diseases (Atherton 2006).

Bacterial adherence to the gastric mucosa plays a key role in the persistent infection. It provides protection from the acidic environment of the stomach, avoids clearance by peristaltic movements and enables easier access to nutrients for the metabolically impaired *H. pylori* (Atherton 2006). The bacterium has over time evolved to avoid clearance by shedding of stomach epithelial cells or changes in the inflamed gastric mucosa (Atherton 2006). Adherence, which has to cope with these phenomena, is mediated by two major adherence proteins. These proteins belong to the family of *H. pylori* outer membrane proteins (HOP) (Alm et al. 2000; Shao et al. 2005) and bind to oligosaccharides present on the gastric mucosa. The blood group antigen binding adhesin (BabA) binds to the fucosylated ABO blood group antigen, expressed on gastric epithelial cells, at first contact with host cells (Fig. 4) (Ilver et al. 1998). Upon infection the activation or release of many other virulence factors is triggered (Ilver et al. 1998; Atherton 2006).

The bacterium possesses many virulence factors, like the cytotoxin-associated gene A (CagA), the vacuolating cytotoxin A (VacA), the urease system and outer membrane phospholipase A (OMPLA) (Fig. 4). These factors are thought to induce an immune response of the host with a subsequent inflammation, provide resistance to the harsh environment of the stomach and play a role in the delivery of nutrients to the metabolically impaired *H. pylori*.

The cytotoxin CagA is transported directly to the epithelial cells via the type IV secretion system and causes the hummingbird phenotype (cell form changes) in the host cells. It interferes with cell-signaling pathways and has both pro- and anti-apoptotic influences (Atherton 2006). The protein VacA is secreted and upon uptake by the host cells it can create acidifying vacuoles by forming selective pores in endosomal compartments, where small anions and neutral charged molecules can enter (Atherton 2006). This process causes oxidative stress and is therefore related to genomic instability, which may lead to cancer (Chitcholtan et al. 2008). It is also believed to play a role in delivering nutrients to the bacterial cells, by making the host-membranes more permeable (Atherton 2006). The urease system buffers the pH in the cell by creating ammonia from urea and therefore allows the cell to survive in acidic conditions. Furthermore, by applying chemotaxis and flagellae, *H. pylori* “swims” to the right environment, the epithelial cell lining (Atherton 2006). The secreted OMPLA protein degrades phospholipids to lysophospholipids, which have detergent properties and thus destroy the integrity of the host membrane, both leading to release of nutrients and an immune response (Tannaes et al. 2005). The wealth of virulence factors produced by *H. pylori* creates a sophisticated machinery to survive in and profit from the human host without damaging it excessively in the short term, thus enabling life-long colonization. The Inflammation induced by *H. pylori* is thought to be a key factor in disease development but the complete infection mechanism is not yet completely revealed (Atherton 2006).

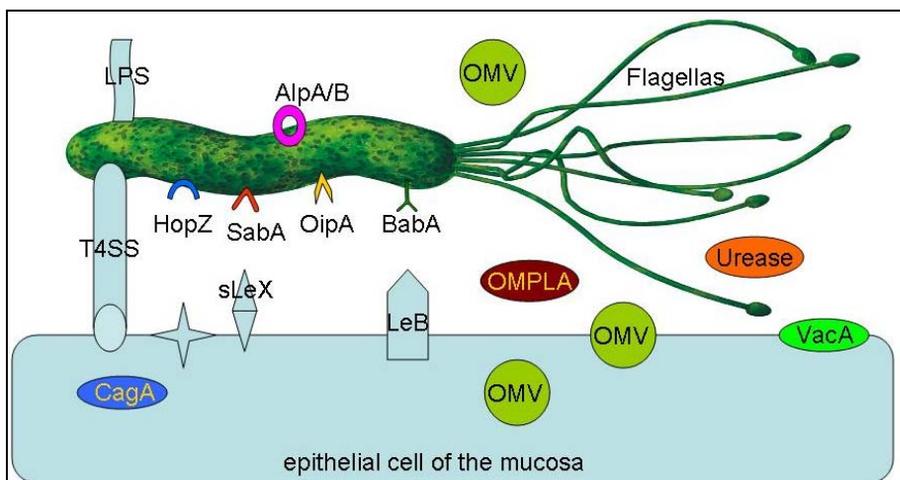


Figure 4: Sketch of the virulence factors produced by the bacterium *H. pylori* (dark green, see text for abbreviations). The interaction of different virulence factors with the host is depicted (the epithelial lining in the human stomach, light blue). The bacterium adheres via BabA and SabA, membrane proteins HopZ and AlpA/B are

thought to promote adhesion or adherence associated proteins. Thereafter it exports virulence factors CagA to the host by using the Type IV secretion system and VacA by secretion. Vesicles (omv) also deliver the virulence factors, such as CagA and VacA, to the epithelial cells. *H. pylori* furthermore secrete OMPLA which destroys the host membrane, and it posses urease which changes the pH of the bacterium. With this battery of virulence factors, the bacterium is very efficient and can adapt to various scenarios and so guarantee a persistent infection. (*H. pylori* bacterium adapted from Luke Marshall)

Upon inflammation, the polysaccharides sialyl Lewis x and a (sLex/a) are upregulated on epithelial cells and the second adhesin from the HOP family of membrane proteins, the sialic acid-binding adhesin (SabA), supports adhesion (Mahdavi et al. 2002). By having two independent adhesion possibilities, *H. pylori* can regulate binding activity to avoid being flushed out of the stomach. Several regulation mechanism are known for BabA, which can for example fine tune its binding to the pH value, where lower pH leads to the release of the blood group antigen (Bugaytsova et al. 2009), or *H. pylori* can turn off BabA activity by phase variation, caused by recombination and slipped-strand mispairing (Bäckstrom et al. 2004). Being able to adapt to changes in the host is one of the reasons why *H. pylori* is so successful in its constant colonization of the human gastric mucosa.

*H. pylori* constantly sheds vesicles (characterization see paper IV), which display the adhesins BabA and SabA. These vesicles can play a role in infection as well as in delivering nutrients to the bacterial cell, e.g. by delivering virulence factors, such as CagA and VacA, to epithelial cells.

Other outer membrane proteins from the HOP family, like the outer inflammatory protein A (OipA), adherence-associated lipoprotein (AlpA&B) and HopZ are thought to be involved in adherence and immune response, though their exact function remains to be discovered (Odenbreit 2005; Shao et al. 2005; Atherton 2006; Doerrler 2006). All these proteins are embedded into the outer membrane of the bacterium and therefore activity-regulating interactions may occur between the proteins and the phospholipids, the main component of membranes. These interactions can be of interest for future drug development (Engelman 2005; Sachs et al. 2006).

## 6.2 Biological role of lipids

The word lipid describes a broad range of molecules, which are of medium size, soluble in organic solvents and mostly amphipathic (Yeagle et al. 2005). This text focuses on the subgroup of phospholipids (specifically glycerophospholipids, abbreviated PLs). PLs are found, among others, in Gram-negative bacterial membranes. They consist of a glycerol body, where two hydrophobic fatty acids with the alkyl chains R' and R'' are esterified at the C atoms sn1 and sn2 and a hydrophilic phosphate, connected to the head-

group (R), is bound at position sn3 (see Fig. 5). The biologically most relevant PLs carry a saturated fatty acid chain at position sn1 and a medium to highly unsaturated fatty acid chain at position sn2 (Fuchs et al. 2009) and are described in Table 1.

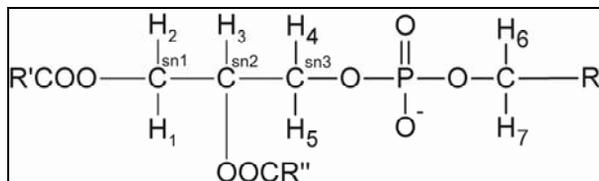


Figure 5: General scheme of a glycerophospholipid. R stands for the head-group (for example choline in PC), R' and R'' are the alkyl chains of the fatty acids esterified to the glycerol moiety. The Glycerol moiety with the esterified fatty acids is called diacylglycerol (DAG). The formula was drawn with ChemSketch.

Lipids are essential to life, having the unique feature of spontaneous assembly into bilayers forming membranes, which allow cells to create compartments (Yeagle et al. 2005; van Meer et al. 2008). Major components of prokaryotic membranes are phospholipids, proteins, glycolipids and lipopolysaccharides (LPS – only in Gram-negative bacteria). When many PL molecules come together in an aqueous solution, the hydrophobic fatty acid chains orient themselves towards each other so that only the hydrophilic head-groups interact with water, so lipids create a barrier non-permeable to water, an ordered bilayer. This barrier, the membrane, allows the cell to protect the cytoplasm from the extracellular environment, critically controls transport and signaling between cell and environment (van Meer et al. 2008). Gram-negative bacteria have an inner membrane (IM) and an outer membrane (OM), which is made permeable for small hydrophilic molecules by porin proteins (Epanand et al. 2009). Both of these membranes consist of an inner- and an outer leaflet; where the outer leaflet of the OM is mainly composed of LPS, while the other leaflets contain mainly PLs (Doerrler 2006). Each of the membranes consists of more than 100 individual lipid species; this huge variety is based on the fatty acid chain composition combined with a range of different head-groups (Yeagle et al. 2005). While cholesterol and its derivatives play important roles in eukaryotic membranes, they are almost completely absent in bacterial membranes (Yeagle et al. 2005). The most common PLs are phosphatidylethanolamine (PE), cardiolipin (CL), phosphatidylcholine (PC), phosphatidylserine (PS), phosphatidylinositol (PI) and phosphatidic acid (PA), their structure, properties and function are given in table 1 (Yeagle et al. 2005). Variation in the PL composition can change the bulk biophysical properties of a membrane, or specific PLs can engage in interactions with membrane proteins. Hence membranes can adapt to environmental changes and are

therefore not just passive barriers. For example, it has been observed for *Staphylococcus aureus*, that the composition of phospholipids is changing upon pH stress encounter. It was hypothesized that the change of concentration of PG and lysoPG leads to a decrease of the membrane permeability for protons and therefore protect the cytoplasm (Haest et al. 1972).

Table 1 presents the head-groups (=R in Fig. 5) of the most common bacterial phospholipids. DAG represents diacylglycerol, and molecules are presented in the right column. The formulae were drawn with ChemSketch according to: (Yeagle et al. 2005; Matsumoto et al. 2006; van Meer et al. 2008; Fuchs et al. 2009).

Headgroup	Properties/functions	structure
Phosphatidyl-ethanolamine (PE)	Zwitterionic, main component of bacterial membranes, conical geometry (small head-group); imposes stress curvature on eukaryotic cells (points for budding etc); not bilayer forming; often used to accommodate proteins and regulate function, can have two hydrogen bonds, therefore very compact and rigid network, PE micro domains possible	$\text{DAG-O-P(=O)(O}^-\text{)-O-CH}_2\text{-CH}_2\text{-NH}_3^+$
Cardiolipin (Diphosphatidyl glycerol, CL, DPG)	Negatively charged under physiological conditions, because it complexes a proton between the phosphates, conical geometry, imposes stress curvature on eukaryotic cells, not bilayer forming; often used to accommodate positively charged proteins and regulate their function; major component in mitochondria, chloroplasts and some bacteria	$\begin{array}{c} \text{O}^- \\   \\ \text{O}=\text{P}-\text{O}-\text{CH}_2-\text{C}-\text{CH}_2-\text{O}-\text{P}=\text{O} \\   \quad   \quad   \\ \text{O} \quad \text{H} \quad \text{O}^- \\ \text{DAG} \quad \text{OH} \quad \text{DAG} \end{array}$
Phosphatidyl-choline (PC, Lecithin)	Zwitterionic; most common in eukaryotic membranes; self-organizes in planar bilayers, nearly cylindrical shape	$\text{DAG-O-P(=O)(O}^-\text{)-O-CH}_2\text{-CH}_2\text{-N}^+\text{(CH}_3\text{)}_2$
Phosphatidyl-glycerol (PG)	Negatively charged, builds extensive hydrogen bond network, can form ionic bonds	$\text{DAG-O-P(=O)(O}^-\text{)-O-CH}_2\text{-C(OH)(H)-CH}_2\text{OH}$

	and coordination bonds; PG in clusters might exclude CL	
Phosphatidic-acid (PA)	Negative charged, important as an intermediate in biosynthesis, negatively charged, can have detergent properties	$\text{DAG}-\text{O}-\overset{\text{O}}{\parallel}{\text{P}}-\text{OH}$ $\quad \quad \quad  $ $\quad \quad \quad \text{O}^-$
Phosphatidyl-serine (PS)	Negative charged, mostly present in inner leaflet, can be specifically bound by proteins	$\text{DAG}-\text{O}-\overset{\text{O}}{\parallel}{\text{P}}-\text{O}-\text{CH}_2-\overset{\text{H}_3\text{N}^+}{\text{C}}-\overset{\text{O}}{\parallel}{\text{C}}-\text{O}^-$ $\quad \quad \quad   \quad \quad \quad  $ $\quad \quad \quad \text{H} \quad \quad \quad \text{H}$
Phosphatidyl-inositol (PI)	Negative charged, phosphorylated on the inositol moiety it is an important hormone signaling factor	$\text{DAG}-\text{O}-\overset{\text{O}}{\parallel}{\text{P}}-\text{O}-\text{C}$

Building a membrane barrier might be the most important function of lipids, but it is only one function. Membranes also change structure as a function of lipid composition and are involved in transport of nutrients and in signaling. Therefore membranes host proteins, imbedded or peripheral, which can build channels or are receptors for signaling molecules or relay signals (Cronan 2003). Recently it was discovered that certain PLs can also regulate the function of membrane proteins, either by specific binding, or by forcing proteins to aggregate, or by stabilizing a particular protein structure by providing a specific environment in microscopic “rafts” (Engelman 2005). These discoveries change the picture of a membrane from a uniform fluid distribution to a raft-like inhomogeneous state. This inhomogeneous distribution suggests another function of PLs: To change the local structure of membranes as needed for fundamental biological functions such as cell division or shedding of vesicles. The most important local structural parameter is its curvature, which can be influenced by lipids’ size in head-group and fatty acid chains. For example, PE and CL have been shown to preferably localize at a place where the membrane curvature is increased, which is energetically favorable for the geometry of their larger head-groups. Thus, enriching the membrane locally in specific PLs allows the cells to lower the energy barriers locally for processes that require membrane curvature, such as cell division and budding of vesicles.

Specific lipids, such as PI and its phosphorylated forms, also play important roles as messenger molecules (Yeagle et al. 2005). Furthermore, upon processing with phospholipases, several messenger molecules are formed from the original phospholipid (Fuchs et al. 2009). This processing creates signaling molecules with strongly different properties, making them suitable for different signaling roles. In lysophospholipids (lysoPLs) most commonly the sn2 fatty acid chain is hydrolyzed and in PA the head-group at sn3 is

hydrolyzed. Diacylglycerol (DAG) is formed by hydrolysis of the phosphate moiety, and lysophosphatidic acid (lysoPA) by hydrolysis of both the head-group and one fatty acid (van Meer et al. 2008). These molecules can either stay in the membrane and change its properties thereby recruiting proteins, or leave the membrane and act as cytoplasmic signals, such as LPA and inositols from PI. LysoPLs have detergent properties and therefore destabilize membranes. Under pathological conditions increased lysoPL concentrations are found which in turn decreases membrane integrity and thus increases pathogenicity. Therefore high lysoPL concentrations are commonly observed in inflammation-related diseases as asthma, rheumatoid arthritis and sepsis, but even play a role in some cancers, diabetes and reproductive failure (Fuchs et al. 2009).

### **6.3 Background: *H. pylori* lipids**

Hirai and coworkers (Hirai et al. 1995) analyzed the PL composition of *H. pylori* and found, besides the common phospholipids PE, CL, PG, PS and PC, also three cholesteryl glycosides (CG), of which one of them was phosphorylated (CPG) and not discovered in bacteria before. As described before, bacteria cannot synthesize cholesterol, it is therefore believed that *H. pylori* internalizes cholesterol and then synthesizes the cholesteryl glycosides, which can make up to 25% of its membrane. Two years later, a change of PL composition related to acidic pH stress was described, where an increase of stress induces an increase in lysoPE, caused by the bacterial outer membrane phospholipase A (OMPLA) (Bukholm et al. 1997).

In this context, lysoPE was thought to be lytic, either by making the bacterial membrane permeable (at moderate concentration of lysoPE) or destroying it (at high concentration of lysoPE). It was hypothesized that higher membrane permeability aids the release of virulence factors (VacA, CagA, OMPLA and Urease) to the host cells. It was also suggested that released fatty acids and LPE molecules can destroy the host membranes (Tannaes et al. 2001). Furthermore, the release of CPG was suggested to cause hemolysis (Hirai et al. 1995). The Bukholm group could later proof a relationship of the OMPLA gene activity and the concentration of lysoPE, and therefore a correlation to peptic ulcer disease (Tannaes et al. 2001; Tannaes et al. 2005). Hirai and coworkers showed that a change of PL composition over time causes a change of morphology. Over a course of 8 days, bacteria passed from exponential growth (spiral), to stationary phase and finally to coccoid stage. This transition was accompanied by increases in CPG and PC, a decrease in PE, and an increase in CL at the expense of PG. Similar to the Bukholm group's results, causal relationships were postulated between the increase of CPG and hemolysis, and between the decrease of PE and destabilization of the membrane (Shimomura et al. 2004).

## 7 NMR background

Questions related to the function of biological systems are usually quite complex. As a particular strength of NMR, specific experiments can be designed to address each of these questions. Using these experiments, NMR yields information on an atomic resolution level on molecular structure, dynamics or intermolecular interactions. For example, relaxation measurements can analyze motions of functional groups in complex biological molecules. Hydrogen exchange experiments monitor biomolecular folding, COSY-type experiments define which chemical groups are connected via bonds to each other and NOESY based experiments measure distances within molecules or between them. Design of NMR experiments plays an important role in this thesis, therefore important concepts for the understanding of solution-state NMR experiments are outlined in this chapter.

### 7.1 Spins and Spectrometer

Using NMR, we look at an intrinsic nuclear property of specific isotopes (atoms with different numbers of neutrons), their nuclear spin. The nuclear spin quantum number ( $I$ ) describes how the atomic nucleus behaves in an external magnetic field; only isotopes with  $I \neq 0$  are observable, and from these isotopes,  $I = \frac{1}{2}$  are preferred in biomolecular NMR, because they normally lead to the best spectra. In the magnetic field of a spectrometer, each spin has a magnetic moment ( $\mu$ ), which is determined by the isotope's gyromagnetic ratio ( $\gamma$ ). This gyromagnetic ratio determines the NMR sensitivity of each isotope and the isotopes most commonly used in biomolecular NMR are  $^1\text{H}$ ,  $^{13}\text{C}$ ,  $^{15}\text{N}$  and  $^{31}\text{P}$ .  $^1\text{H}$  has a natural abundance of 99.98%, whereas  $^{13}\text{C}$  (1.11%) and  $^{15}\text{N}$  (0.37%) are only abundant in small fractions. Therefore artificial enrichment of the latter isotopes – called labeling – vastly increases the information content of biomolecular NMR spectra, and allows highlighting the signals of desired molecules. The heavy hydrogen isotope deuterium (D,  $I=1$ ) is also NMR-observable, but its signal is completely separated from the signal of the  $^1\text{H}$  isotope and therefore D is not visible in standard NMR spectra. Consequently substitution with D in the solvent can be used to suppress unwanted  $^1\text{H}$  signals, often needed for the strong signal of water ( $\text{H}_2\text{O}$ ). Furthermore, selective D labeling of biomolecules is a powerful tool to reduce spectral crowding (Levitt 2005). The work presented in this thesis uses liquid-state NMR, where in solution, molecules of interest normally tumble freely, which eliminates several phenomena which influence NMR signals in the solid state. For example dipolar couplings are averaged to zero, but can be re-introduced if desired.

To create an NMR signal, the sample is placed in a strong magnetic field, generated in the spectrometer using superconductive coils. This field, called  $B_0$ , is desired to be homogeneous in space and constant in time, and can reach up to 23.5 Tesla, corresponding to a  $^1\text{H}$  frequency of 1000 MHz. The  $B_0$  field is defined as z-axis and creates two states (caused by the Zeeman effect), spin up ( $m = 1/2$ ) and spin down ( $m = -1/2$ ) (Fig. 6). These states differ in energy proportional to the  $B_0$  field and therefore their occupancies differ slightly (Boltzmann distribution). The occupancy difference between the two states results in net/bulk magnetization  $M_z$ , which is also called longitudinal magnetization and is used to create NMR signals. Since the energy difference is very small,  $M_z$  is very small too; therefore NMR is a rather (in)sensitive. Because  $M_z$  increases with  $B_0$ , as strong magnets as possible are used for NMR (Keeler 2005).

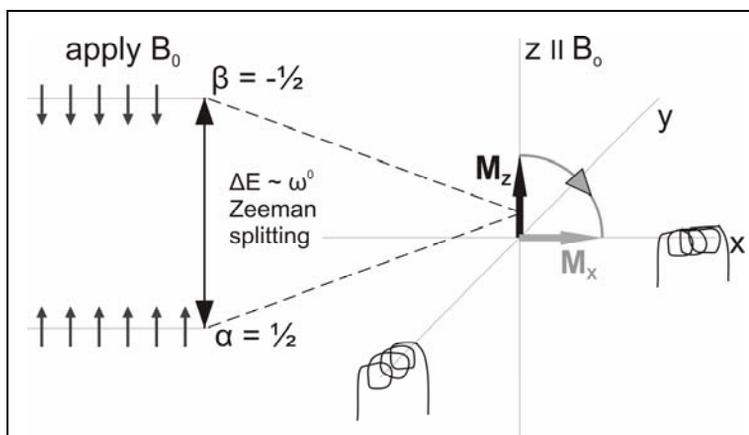


Figure 6: To the left: In a static magnetic field ( $B_0$ ) the nuclear magnetic moments are aligned along  $B_0$  and create two energy levels  $\alpha$  and  $\beta$ , which exist for spin- $1/2$  nuclei. According to the Boltzmann distribution, the energy difference  $\Delta E$  between the states leads to a population difference, such that slightly more spins occupy the lower-energy state. For the whole ensemble of spins in an NMR sample, this creates a bulk magnetization  $M_z$  along the z axis. To the right: The bulk magnetization in a laboratory frame along the z axis (black:  $M_z$  magnetization) is manipulated by a radiofrequency field ( $B_{RF}$ ) transmitted along the x or y axis, symbolized by two coils. The gray round arrow depicts a 90 degree pulse around the y axis. After this “excitation”, the resulting transverse magnetization  $M_x$  (grey arrow) creates a detectable NMR signal. This radiofrequency signal is detected by a coil and used to create the NMR spectrum

The equilibrium magnetization  $M_z$  can be tilted by a radiofrequency field into the x-y plane (Fig. 6), such perturbation and the response of the spin system are the basis of NMR. The magnetic moment of the aligned spins experiences precession with the Larmor frequency, which depends on the gyromagnetic ratio of the spin and the strength of the  $B_0$  field. To simplify,

one creates a new reference frame (“rotating frame”) which rotates with the reference Larmor frequency, and obtains therefore a static picture of the magnetic moment – the net magnetization.

## 7.2 Manipulation of spins

The net magnetization created in the  $B_0$  field can be manipulated by radio frequency pulses ( $B_{RF}$ ) that match the Larmor frequency of the spin. Viewed in the rotating frame, the  $M_Z$  magnetization rotates around a  $B_{RF}$  field applied along the  $y$  (or  $x$ ) axis, creating transverse magnetization (magnetization in the  $x$ - $y$  plane), which is called coherence (Fig. 6). After creating transverse magnetization one can monitor the behavior of this excited state and link it to different processes, such as the influence of its chemical environment on the spin, called the chemical shift, or the magnetization’s return to equilibrium, called relaxation. Using a  $90^\circ$  ( $\pi/2$ ) pulse, one can flip magnetization from  $z$  into the  $x$ - $y$  plane or vice versa and a  $180^\circ$  ( $\pi$ ) pulse inverts the orientation (e.g.  $z$  to  $-z$ ). Because each nucleus has a particular gyromagnetic ratio, the pulses needed to influence its magnetization must have a specific frequency. Therefore the magnetization of selected nuclei ( $^1\text{H}$ ,  $^{13}\text{C}$ ,  $^{31}\text{P}$  ...) can be manipulated separately with these different frequencies. Other tools to manipulate the magnetization are gradient pulses, defined inhomogeneities of the  $B_0$  field over the sample volume, and spin locks which arrest the magnetization at a certain direction within the rotating frame of reference. By combining pulses, delays, gradient pulses and spinlocks in a planned order, one creates something called a pulse sequence (examples chapter 7.4 & Fig. 10) and defines, which information one wants to receive from the NMR experiment. The description of the state of the magnetization before and after a pulse or during delays can be obtained by using the product operator formalism. This is a simplified version of quantum-mechanical descriptions of the evolution of spin systems (Keeler 2005; Levitt 2005).

At the end of each pulse sequence, the resulting coherence can be observed as radiofrequency signal emitted from the sample, using the same coil that is used to transmit pulses to the nuclei. The recorded signal is called free induction decay (FID) and describes the evolution of the coherence over time. The FID is then Fourier transformed to obtain the NMR spectrum, which contains signals of various atoms within the molecule investigated along a frequency axis.

## 7.3 Spin interaction

Having the possibility of manipulating spins, and therefore a signal to measure, one can extract specific information for every spin in a molecule. Each nuclear spin interacts with other nuclear spins of the same and different atom types (J couplings, dipolar couplings), with its electronic environment (chemical shift), the  $B_0$  field, and random electromagnetic fields (relaxation). These interactions give rise to several NMR parameters which give important information on the molecular environment of each spin.

### 7.3.1 Chemical Shift

The chemical shift reports on the local chemical environment of each nuclear spin. The chemical shift effect occurs when electron clouds circulate in the magnetic field  $B_0$ , inducing small magnetic fields which slightly change the  $B_0$  field experienced by the nuclear spin, and therefore its Larmor frequency. The size of this effect depends on e.g. the electro-negativity of directly bound atoms. Thus, the chemical shift creates a specific frequency position for each signal, which reflects the atom's chemical environment in the molecule (Fig. 7). The induced field is very small and the chemical shift is therefore denoted as parts per million (ppm) deviation from a reference frequency. Because the electron environment is different for every atom (e.g.  $\text{CH}_3$  protons experience a higher electron density than a  $\text{COOH}$  proton), one gains a unique marker of the local chemical surrounding of each atom, which can be helpful for identifying the chemical group that the atom belongs to.

The chemical shift also gives rise to an interesting phenomenon when several structural states, such as conformers, exchange on a time scale of milliseconds. In this case, the line shapes of the involved signals become sensitive indicators of the rate of the exchange process. If two resolved signals are observed for two exchanging states, one can conclude that the exchange rate is small compared to the chemical shift difference between the signals. If a single signal for two states is observed, but it is known that exchange is occurring, the exchange rate is large compared to the chemical shift difference between the individual states, but in this case their chemical shifts are usually not well known. Finally, if the exchange rate is of the same magnitude as the chemical shift difference, a single broad signal is observed (exchange broadening) and, if so the exchange rate can be derived from the lineshape.

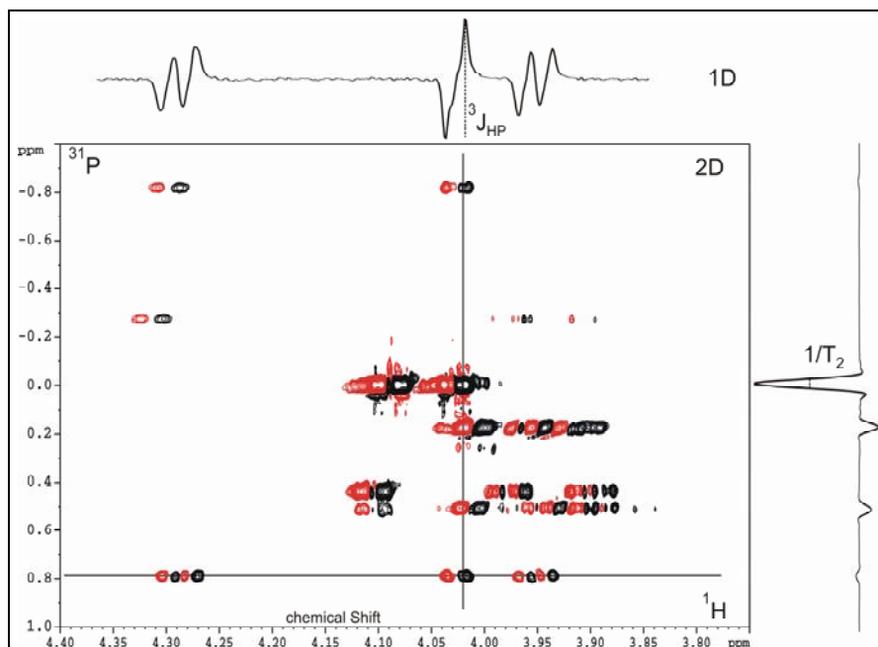


Figure 7: Example for a two-dimensional NMR spectrum, obtained with a P,H-COSY experiment (see pulse sequence in Fig. 10). The pulse sequence of the experiment is designed to correlate  $^{31}\text{P}$  atoms with  $^1\text{H}$  atoms. On the horizontal axis  $^1\text{H}$  chemical shift is displayed and on the vertical axis  $^{31}\text{P}$  chemical shift. Each signal in the contour plot shows that a  $^1\text{H}$  atom with a particular chemical shift is linked to a  $^{31}\text{P}$  atom with a certain chemical shift. On the top one sees a slice from the  $^1\text{H}$  dimension, where one can see anti-phase signals (up-and-down patterns). In these anti-phase signals one can measure the  $^3J_{\text{HP}}$  coupling, which reflects the interaction between a  $^{31}\text{P}$  atom and a  $^1\text{H}$  atom separated by three bonds in a phospholipid (see sketch Fig. 5). On the right hand side, one can see a slice from the  $^{31}\text{P}$  dimension, where positive and negative in-phase signals can be observed. From the width of these signals one can estimate the  $^{31}\text{P}$  nuclei's transverse relaxation times  $T_2$ .

Chemical shift anisotropy (CSA) describes the orientation dependence of the chemical shift, which is caused by an asymmetric distribution of electron density around the nucleus, e.g. in a chemical bond between two nuclei. In solid-state NMR spectra, the CSA can be directly observed as the shape of a signal (powder pattern). In molecules which tumble isotropically in solution, the CSA is averaged to the chemical shift, but it still contributes to relaxation.

### 7.3.2 J coupling

J couplings (scalar or indirect spin-spin couplings) are mediated between nonequivalent nuclear spins by the magnetic fields of bonding electrons and are therefore limited to nuclear spin systems that are connected via at most three bonds ( $^3J$  coupling). J coupling can be used to transfer magnetization from one spin to another, which forms the basis of the two-dimensional NMR experiments of the COSY and TOCSY type. Because these experiments connect the signals of chemically bonded atoms, they are essential tools for structure elucidation by NMR. Each J coupling that a spin experiences causes a splitting of its signal according to the number of coupling partners and therefore a reduction in intensity (Fig. 7). In some cases the transfer of magnetization or the splitting of signals is undesirable; one can therefore decouple the interacting spin using strong radio-frequency field.  $^3J$  couplings depend on the dihedral angle between the interacting atoms according to Karplus relationships; this fact is frequently used to determine local conformation. Because of their relation to covalent structure and to conformation, J couplings are extremely useful for structure determination.

If a conformational equilibrium exists in a molecule, the observed J coupling is a population-weighted average of the J couplings that would be observed for the individual conformers. While this at first hinders the interpretation of J couplings in terms of local conformation, it can in favorable cases be used to determine the populations of the conformers from J couplings.

### 7.3.3 Dipolar interactions

Dipolar interaction describes the coupling between different nuclear spins, which are close in space and is called direct dipole-dipole coupling (DD coupling). The coupling constant depends only on the distance between the nuclei ( $r^{-3}$ ) and is therefore important for structure determination. As for J couplings, also direct dipole-dipole couplings lead to splitting of signals, unless they are averaged out by molecular tumbling. In solution NMR, dipole-dipole couplings can be deliberately re-introduced if the molecules are partially aligned by an anisotropic environment. A weak alignment is chosen so that the scaled-down “residual” dipolar couplings (RDCs) are conveniently measurable (Getz et al. 2007). Although the direct effect of DD couplings on a spectrum is averaged to zero for isotropically tumbling molecules, dipole-dipole coupling still influences the relaxation behavior strongly. An important application of this effect is the through-space transfer of magnetization between nuclei, which strongly depends on the internuclear distance. This cross relaxation effect is the basis for the two-dimensional NOESY and ROESY experiments, which are essential for distance measurements in structure determination.

### 7.3.4 Relaxation

By excitation of the net magnetization one perturbs the thermal equilibrium of the nuclear spins and creates non-equilibrium states. Relaxation describes the process which leads back to the thermal equilibrium by interactions of the spins with their molecular environment ( $T_1$  – spin-lattice relaxation time) or the loss of coherence by interaction with other spins ( $T_2$  – spin-spin-relaxation time). Relaxation is induced by fluctuations of the magnetic field surrounding the nuclear spin, which can be caused by dipole-dipole relaxation, chemical shift anisotropy, or by imperfections of the  $B_0$  field. In solution, molecules rotate randomly, and the speed of the rotation is described by a rotational correlation time  $\tau_c$  which is in the range of nanoseconds for biomolecules.  $T_1$  and  $T_2$  depend on  $\tau_c$ , each in their own way, and on motions within the molecule.  $T_1$  and  $T_2$  relaxation occur at the same time, but  $T_1$  is usually longer than  $T_2$  ( $T_2 \leq T_1$ ), because magnetization can dephase in the x-y plane without relaxing to z. However, if the z magnetization is restored, no transverse magnetization can be left. For biomolecules,  $T_1$  is typically of the order of seconds, which means that one must wait seconds between consecutive repeats of an NMR pulse sequence.  $T_2$  is typically 50 milliseconds, this limits how long pulse sequences may be, without losing too much signal by  $T_2$  relaxation. The relaxation rates  $R_1$  and  $R_2$  ( $T_1^{-1}$  and  $T_2^{-1}$ , respectively) are each the sum of an auto-relaxation rate constant ( $R_{\text{auto}}$ , relaxation constant for the spin itself) and a cross-relaxation rate constant ( $R_{\text{cross}}$ , relaxation caused by interaction with surrounded spins).  $R_{\text{cross}}$  and  $R_{\text{auto}}$  are both dependent on the correlation time  $\tau_c$  but the dependence differs between  $R_1$  and  $R_2$ . The rates contributing to  $R_1$  are called longitudinal rates  $R^L$ ; and the respective rates for  $R_2$  are called transverse rates  $R^T$ . As function of the molecular tumbling (and therefore the rotational correlation time  $\tau_c$ ),  $R^{L/T}_{\text{auto}}$  are always positive,  $R^T_{\text{cross}}$  is always negative but  $R^L_{\text{cross}}$  has a zero transition (Fig. 11) (Levitt 2005). Besides their dependences on  $\tau_c$ , the rates are differently influenced by intramolecular motions and can therefore be used to analyze these.

The cross relaxation rates  $R^L_{\text{cross}}$  and  $R^T_{\text{cross}}$  are due to the through-space dipole-dipole interaction between two spins. The dipole-dipole relaxation leads to the Nuclear Overhauser Effect (NOE) for the longitudinal magnetization or the Rotating frame Overhauser Effect (ROE) for transversal magnetization (chapter 7.4.2). Because cross relaxation rates depend on the distance between the spins ( $\sim r^{-6}$ ), they are used to measure intramolecular distances for structure determination. The nuclear Overhauser effect occurs also between  $^1\text{H}$  and a heteronucleus (e.g.  $^{15}\text{N}$ - $\{^1\text{H}\}$  hetNOE), this changes the z magnetization of the heteronucleus to a degree that also reflects intramolecular motion.

Besides the dipole-dipole interaction, the shape of the electron cloud in a chemical bond also gives rise to a relaxation mechanism, based on the chemical shift anisotropy (CSA). The CSA depends on the hybridization of the atomic orbitals and its relaxation influence increases with  $B_0^2$ , therefore CSA relaxation becomes more important with increasing  $B_0$  fields. For RNA  $^{13}\text{C}$  relaxation studies, the CSA contributes up to 25% to the total relaxation rate.

$T_1$  and  $T_2$  report on local motions of the studied vectors (e.g. CH groups in RNA) (Fig. 8), but the molecular tumbling time  $\tau_c$  sets a limit for the motion that can be detected: The motion of the whole molecule randomly reorients all bond vectors after  $\tau_c$ , therefore intramolecular motion that is slower than  $\tau_c$  cannot be detected (Levitt 2005).

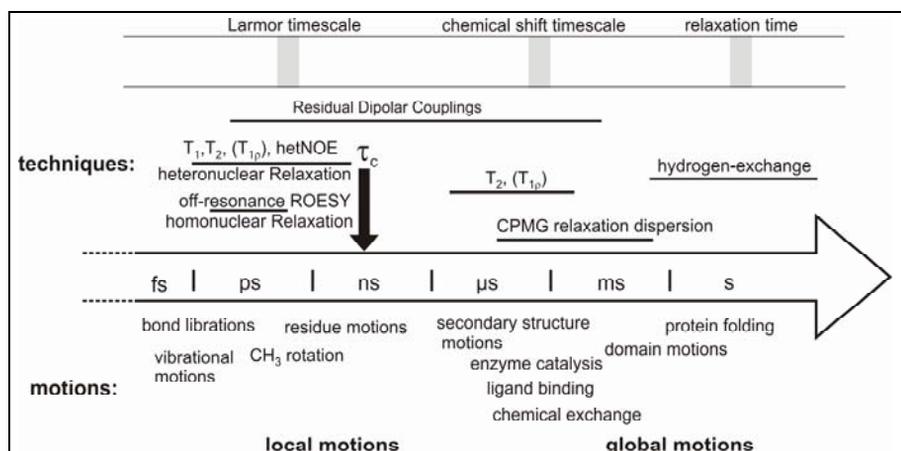


Figure 8: Overview over motions, their timescales and relevant NMR quantities (Levitt 2005). The correlation time  $\tau_c$  marks a critical border, such that most quantities cannot detect motion slower than  $\tau_c$ , because the random reorientation of all bond vectors by the tumbling of the molecule causes a loss of motional memory. Except for special cases, only residual dipolar couplings overcome this difficulty. While fast motions are not directly related to biological function, they form the basis for slower global motions which can directly represent functional transitions, such as domain flipping and folding. Motion on all time scales is relevant for NMR structure determination and for the interpretation of structures.

Relaxation rates depend on motion and are therefore good indicators for it. Motions on the timescale of  $10^{-12}$  s to  $10^2$  s (picoseconds to minutes) can be monitored and distance changes smaller than an Å (e.g. motion of a single residue) to many Å (movement of whole domains) can be observed (Fig. 8) (Al-Hashimi 2007). Motions that are faster than the rotational correlation time of the molecule ( $\tau_c$ ) are considered fast and can be captured by heteronuclear relaxation measurements (chapter 7.4.3) (Korzhnev et al. 2001). Motions much slower than  $\tau_c$  can be attributed to conformational

exchange or domain motions, and can be captured by relaxation dispersion/CPMG experiments (Palmer et al. 2006). Very slow motions can be accessed by exchange spectroscopy, real-time NMR or hydrogen exchange. For a long time, NMR was blind for motions on timescales between  $\tau_c$  and microseconds; recently residual dipolar coupling (RDC) measurements have filled this gap. Measuring relaxation is important for the structure determination process, because many structure parameters measured by NMR (such as NOE, J couplings, chemical shift and even RDCs) can be affected by motion (motional averaging), and this affected structural restraints should therefore be excluded from structure calculation (Getz et al. 2007).

## 7.4 Design of NMR experiments

One-dimensional (1D) NMR spectra often do not allow access to all information required about biomolecular systems. Therefore the development of two- to four-dimensional (2D - 4D) NMR experiments is a central task of biomolecular NMR (Sattler et al. 1999). The key to access the information is to increase chemical shift resolution, by using the combination of two chemical shifts to identify a signal. The two axes of a two-dimensional NMR spectrum (Fig. 7) reflect the chemical shift of one isotope each, and the plane of the spectrum contains so-called cross-peaks, which are usually displayed as a contour map. Each cross-peak depicts the correlation of a pair of spins by a defined interaction. Multidimensional experiments vastly enhance the power of NMR, because the added dimensions increase the chemical shift resolution that can be obtained, and the interactions creating a cross-peak even carry molecular information about connectivity (through bond or space) and distance between these spins. The interaction originates in transfer of magnetization between the two spins either via J coupling or cross-correlated relaxation, originated from dipole-dipole couplings. In this chapter the designs of the two-dimensional (2D) experiments of papers I-III are described.

In a two-dimensional experiment the chemical shift of one nucleus is measured at the end of the pulse sequence during a time called  $t_2$  in the same way as in a one-dimensional experiment. Then the duration in the pulse sequence called  $t_1$  is incremented to record the chemical shift of the second spin.

### 7.4.1 COSY (correlation spectroscopy)

COSY (correlation spectroscopy) experiments transfer coherence from one spin to another via J coupling, implying that the nuclei that are separated by at most three bonds can be linked ( $^3J$  coupling). COSY-type experiments are therefore immensely valuable to assign NMR signals, because they reveal connectivity of chemically bonded nuclei. While the original COSY is a homonuclear  $^1H, ^1H$  correlation experiment, COSY can also be applied in heteronuclear spin systems and this case is considered here (Fig. 10). COSY experiments rely on the formation of so-called antiphase operators, which can be created in INEPT (insensitive nuclei enhanced by polarization transfer) building blocks (Fig. 9). In this example, the first  $90^\circ$  pulse on  $^{31}P$  creates  $P_Y$  coherence (inphase), which evolves during the two durations of  $\Delta$  into the desired  $2P_XH_Z$  operator (antiphase). Full evolution of the desired coupling occurs if  $2\Delta = 1/2J$ , so  $\Delta$  should be chosen close to  $1/4J$ . Between the two  $\Delta$  delays, a pair of  $180^\circ$  pulses ascertains that only the J coupling is active and the chemical shift is refocused. After the J coupling evolution time of  $2\Delta$ ,  $90^\circ$  pulses on both nuclei species convert the  $2P_XH_Z$  operator into  $2P_ZH_Y$ , which means that coherence has been transferred from  $^{31}P$  to  $^1H$ , and therefore signal intensity that originates from  $^{31}P$  can now be detected on  $^1H$ . The signal of the type  $2P_ZH_Y$  gives a positive-and-negative peak pattern in the spectrum (Fig. 7/paper III), which is characteristic of antiphase operators.

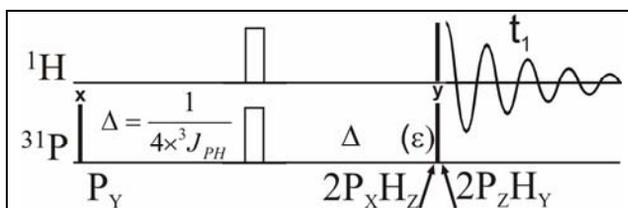


Figure 9: Description of the 1D INEPT pulse sequence for the example of  $^{31}P, ^1H$  transfer. Vertical lines and open boxes represent  $90^\circ$  and  $180^\circ$  pulses, respectively. The spacing of pulses and delays along the horizontal axis represents their chronological succession.  $P_Y$  (inphase) magnetization is created by a  $90^\circ$  pulse. The evolution of  $^{31}P, ^1H$  J couplings during two  $\Delta$  creates antiphase magnetization ( $2P_XH_Z$ ). This is transferred by the final  $90^\circ$  pulses into  $2P_ZH_Y$ , which is observed as a  $^1H$  FID, symbolized by the oscillating signal during  $t_1$ . \*( $\epsilon$ ) after  $\Delta$ : Possible expansion to a 2D experiment by inserting an evolution time for the  $^{31}P$  chemical shift.

The INEPT sequence described in figure 9 could directly be turned into a two-dimensional  $^{31}P, ^1H$  correlation experiment by inserting an incremented delay after  $2\Delta$ , which would serve as evolution time for the  $^{31}P$  chemical shift. By convention, the  $^{31}P$  evolution time would then become  $t_1$ , and the

detection time of the FID would be called  $t_2$ . In practice, this approach is very unfavorable, because the combined  $2\Delta + t_1$  delay would become very long and the signal would decay before its transfer to  $^1\text{H}$ . In the semi-constant-time (Grzesiek et al. 1993) P,H-COSY (Fig. 10 & paper III), the  $180^\circ$  pulses move in a customized fashion, and the first  $\Delta$  delay is decremented with incrementation of  $t_1$  instead of being constant. In this way, the  $2\Delta$  delay can be used simultaneously to allow the P,H coherence transfer, and to evolve the  $^{31}\text{P}$  chemical shift and thus less magnetization decays through relaxation.

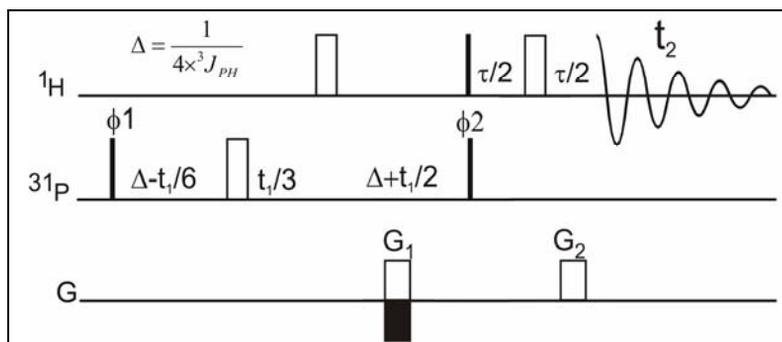


Figure 10: Pulse sequence of the P,H-COSY experiment. In the first part the INEPT step is recognizable, although the  $180^\circ$  pulses are shifted. The semi-constant time chemical shift evolution is presented for the  $^{31}\text{P}$  nucleus with a suitable incrementation and decrementation of delays together creating  $t_1$ .

The final  $\tau/2$   $180^\circ$  ( $^1\text{H}$ )  $\tau/2$  element in the P,H-COSY sequence is not critical for coherence transfer, but is in practice essential to obtain clean spectra, because together with  $G_1$  and  $G_2$  it suppresses all signals that have not undergone the desired P,H transfer (Sattler et al. 1999).

#### 7.4.2 ROESY: (Rotating frame Overhauser Effect spectroscopy)

ROESY and NOESY are homonuclear 2D type experiments that are used to determine internuclear distances for structure determination by NMR. Both experiments depend on cross-relaxation for magnetization transfer. In the classical NOESY experiment (Nuclear Overhauser Effect Spectroscopy) longitudinal magnetization is transferred, in the newer ROESY experiment (Rotating frame Overhauser Effect spectroscopy) it is transversal magnetization instead. Because cross relaxation rates ( $R_{\text{cross}}$ ) depend on internuclear distances, these experiments are excellent to measure distance through space and therefore they are essential in structure determination. As described in chapter 7.3.4 and Fig. 11,  $R_{\text{cross}}^T$  and  $R_{\text{cross}}^L$  have different dependences on the rotational correlation time  $\tau_c$  which is mainly determined

by molecular weight. Because  $R_{\text{cross}}^L$  changes sign as function of  $\tau_c$ ,  $R_{\text{cross}}^L$  experience a zero transition for medium-sized molecules, and the NOESY experiment cannot be used for these molecules. A further problem for the NOESY experiment is the impossibility - for large molecules - to distinguish between cross-relaxation and chemical exchange, which both give rise to positive cross-peaks.

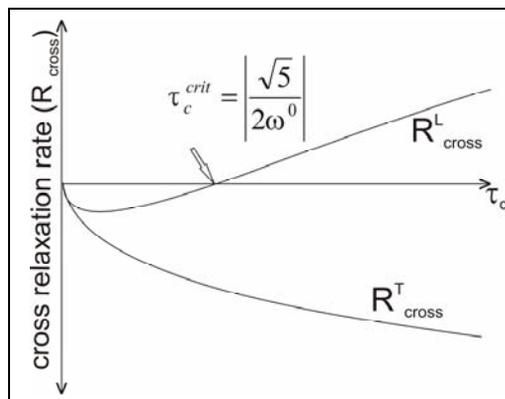


Figure 11: Background of ROESY. DD cross-relaxation rates are plotted as function of the rotational correlation time  $\tau_c$ , which is strongly related to molecular weight and describes the molecular tumbling. The indices L and T denote longitudinal and transverse cross-relaxation rate (dipole-dipole coupling effect), respectively.  $R_{\text{cross}}^L$  becomes zero for a certain  $\tau_c$ , which usually corresponds to medium-sized molecules of about 1 kD.

Both problems are solved with the ROESY experiment, because  $R_{\text{cross}}^T$  does not become zero for any value of  $\tau_c$ , and therefore ROESY cross-peaks are negative for molecules of all sizes, so that cross relaxation and chemical exchange can always be distinguished. These advantages of the ROESY experiment rely on the transfer of transverse magnetization, which is only possible if a spinlock is applied. However, spinlocks can also transfer magnetization via J couplings through bonds (Hartmann-Hahn transfer as used in the TOCSY (Total Correlation Spectroscopy) experiment). This is a problem one cannot avoid, but minimize its influence by applying the spinlock outside of the spectrum (off-resonant), so that transfer via J couplings becomes inefficient (Schleucher et al. 1995). In paper II, an improved ROESY pulse sequence is presented, which simplifies the setup of the ROESY experiments. Furthermore, it is shown that the new EASY-ROESY experiment can be reliably integrated to measure internuclear distances.

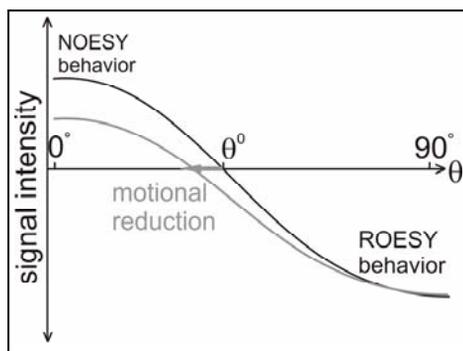


Figure 12: In off-resonance ROESY experiments, the magnetization can be locked along an experimentally tunable angle  $\theta$ , measured off the  $z$  axis. As a function of  $\theta$ , cross relaxation occurs with a weighted average of  $R_{\text{cross}}^L$  and  $R_{\text{cross}}^T$ . Because  $R_{\text{cross}}^L$  and  $R_{\text{cross}}^T$  have opposite signs for large molecules, the two different relaxation rates cancel each other out for an angle  $\theta^0$ . The black line represents the cross-peak intensity of  $\theta$  for completely rigid  $^1\text{H}$ - $^1\text{H}$  vectors, the grey line shows a reduction of  $\theta^0$  by motion. This reduction is a measure for the extent of motion.

Besides their dependence on internuclear distance,  $R_{\text{cross}}^L$  and  $R_{\text{cross}}^T$  are influenced by motion. While this effect is usually small, it can be detected and used to identify motion (Fig. 12, paper I and (Schleucher et al. 2002)).

### 7.4.3 Heteronuclear relaxation reveals fast intramolecular motion

NMR can detect intramolecular motion on a vast range of time scales and at atomic resolution. It is therefore the leading technique to establish relationships between intramolecular motion and biological function.

Different NMR parameters are sensitive to motion on various time scales. A classical approach to detect motion by NMR relies on the influence of motion on the relaxation properties of  $^{15}\text{N}$ - $^1\text{H}$  spin pairs in amide groups. In this approach,  $T_1$  and  $T_2$  of the  $^{15}\text{N}$  spin and the  $^{15}\text{N}$ - $\{^1\text{H}\}$  NOE are measured. These parameters reflect the time the  $^{15}\text{N}$  spin takes to relax to equilibrium ( $T_1$ ), the loss of coherence of the  $^{15}\text{N}$  signal ( $T_2$ , related to the  $^{15}\text{N}$  line width), and the enhancement of the  $^{15}\text{N}$  signal if the attached  $^1\text{H}$  magnetization is perturbed (heteronuclear NOE). Most importantly, each parameter depends in a particular way on motion on the picosecond-nanosecond time scale. Therefore this combination of experimental observables reveals if the  $^{15}\text{N}$ - $^1\text{H}$  bond exhibits motion on this timescale. Using the most common “ModelFree” analysis, the relaxation parameters are interpreted as time scale of the motion (internal correlation time  $\tau_i$ ) and a description of its amplitude (the order parameter  $S^2$ ). An order parameter of 1 means that the  $^{15}\text{N}$ - $^1\text{H}$  bond is completely rigid within the molecule and an order parameter of 0 means that the  $^{15}\text{N}$ - $^1\text{H}$  bond has full motional freedom.

However, there are always many motional models consistent with the relaxation parameters, and it is therefore not possible to derive a motional model from relaxation data alone.

Relaxation measurements are well established for  $^{15}\text{N}$ - $^1\text{H}$  bonds in the protein backbone, where they reveal motion of flexible loops, for example. In nucleic acids, detectable  $^{15}\text{N}$ - $^1\text{H}$  groups are only present in hydrogen bonds of base pairs, amides, and consequently do not reveal interesting motions. Therefore  $^{13}\text{C}$ - $^1\text{H}$  groups must be studied in nucleic acids. During structure determination, all carbons are usually labeled with  $^{13}\text{C}$ , which creates problems for  $^{13}\text{C}$  relaxation studies, because relaxation of any  $^{13}\text{C}$  is also influenced by its neighboring  $^{13}\text{C}$  atom and not only by the directly bound  $^1\text{H}$ . Therefore care has to be taken that true  $T_1$ ,  $T_2$  relaxation times are measured, and that the relaxation parameters may be interpreted as motion of the  $^{13}\text{C}$ - $^1\text{H}$  bond. Because of the complexity of the measurement and the high chemical shift overlap of the resonances, not many RNA relaxation studies have been reported so far but methodology is advancing to overcome these problems (Korzhnev et al. 2001; Duchardt et al. 2005; Shajani et al. 2007).

The conventional relaxation analysis of  $^{15}\text{N}$ - $^1\text{H}$  or  $^{13}\text{C}$ - $^1\text{H}$  spin pairs has several limitations: First, relaxation data cannot give a description of the actual movement. To overcome this limitation, relaxation measurements can be complemented with molecular dynamics simulations of intramolecular motion (Lindorff-Larsen et al. 2005). Second, heteronuclear relaxation reports only on the local motion of directly bonded atoms. An approach to overcome this limitation, called off-resonance ROESY, has been employed in paper I. Third, many functionally important motions occur on time scales too long to be analyzed by heteronuclear relaxation. The most promising approach to study such motions is the use of residual dipolar couplings (RDCs) (Bouvignies et al. 2005; Getz et al. 2007; Lange et al. 2008).

## 8 Conclusion and Outlook

### Paper I:

#### Relaxation measurements of the *HBV* apical loop

**Aim:** The apical loop RNA of *HBV* interacts with the viral reverse transcriptase in the priming complex; therefore the apical loop's 3D structure had previously been determined in our group. Compared with other species, the apical loop is quite stable, well structured and highly conserved. However, the structure determination suggested that some structural parameters might be affected by motion. Here we examined possible motions, because a motional fingerprint might reveal the mechanism of the interaction between the apical loop and the reverse transcriptase, a highly relevant target for drug design.

**Conclusion:** We applied  $^{13}\text{C}$  relaxation measurements ( $T_1$ ,  $T_{1\rho}$  and heteronuclear NOE) and homonuclear off-resonance ROESY to the apical loop RNA to identify areas and patterns of motion on the sub-nanosecond timescale. Using off-resonance ROESY we accessed a variety of vectors which report for local and secondary structure motion, e.g. intra-ribose contacts or sequential ribose-ribose contacts.  $^1\text{H}$ ,  $^1\text{H}$  motions found were in good agreement with  $^{13}\text{C}$  relaxation results, showing that off-resonance ROESY is a good tool to measure motions in RNA. Comparing several nucleotides, we identified different patterns of motion, which led to the conclusion that RNA can independently be flexible in the ribose or the base moiety, in contrast to the situation found in proteins. We found considerable motion in the pseudo-triloop ( $\text{U}_{12}\text{-U}_{14}$ ), as well as for the bulged nucleotides  $\text{C}_{16}$  and  $\text{U}_{23}$ , while the canonical helical stems are rigid. We also recorded experiments to detect motion on the microsecond- and millisecond time scales. No significant motions on these time scales were observed, implying that fast motions must be relevant for the biological function. Because nucleotides observed to be flexible are also highly conserved, we concluded that the observed motion is relevant for the binding process, by allowing the RNA to sample several different conformations, one of which could be trapped by the reverse transcriptase.

**Outlook:** With these data we now try to simulate underlying motion with Molecular Dynamics (MD) – restraining with relaxation data, to obtain an ensemble of structures present in the average apical loop. We then might be able to dock the RNA ensemble to the RT-protein and see which binding mechanism is preferred.

## **Paper II:**

### **EASY ROESY – distance measurement for medium-sized molecules**

**Aim:** For molecules of medium size, ROESY is the only possibility to derive distance information. ROESY spectra frequently contain artifacts caused by undesired interactions between spins, and several pulse sequences have been suggested to minimize Hartmann-Hahn effects and offset dependence. However, these pulse sequences are either cumbersome to set-up or prone to other artifacts.

**Conclusion:** In paper II, we describe a new ROESY experiment, with the aim to eliminate the described problems. We equipped a published ROESY sequence with recently introduced adiabatic pulses, which do not require exact calibration. Therefore the new pulse experiment is easy to set up, and retains the high performance of the original sequence. We tested the experiment on an organometallic complex and a small protein. The protein results showed that distance estimates obtained with the new experiment agree with the known structure and that cross-peaks in the spectra can be translated into distances without a need for correction factors. The results on the organometallic complex showed that the experiment can be used to quantify chemical exchange where other methods fail.

**Outlook:** We expect that the new experiment will be widely used in organic chemistry and structure determination of natural products.

## **Paper III & IV:**

### **Design of the SCT P,H-COSY experiment and characterization of *H. pylori* phospholipids**

**Aim:** *H. pylori* adheres via membrane proteins to the human stomach mucosa. The activity of these adhesion proteins depends on the age of the cells and on the environment, and is therefore related to the infection process. Because membrane proteins require a functional membrane matrix, we aimed to characterize the phospholipid composition of *H. pylori*. Specifically we address if this composition varies with infectious-related variables, such as adhesion protein activity or release of vesicles, and if a specific membrane composition is required for protein function. Using NMR to analyze phospholipids avoids cumbersome and error-prone separation of lipids and can identify unknown lipid components by structural information. However, the small  $^{31}\text{P}$  chemical shift range and the strong solvent dependence of  $^{31}\text{P}$  chemical shifts hinders NMR analysis of phospholipids. To overcome these problems we developed a new experiment, the semi-constant-time  $^{31}\text{P},^1\text{H}$ -COSY, and proved its applicability (paper III), and characterized phospholipids of different membranes with focus on the vesicles shed by *H. pylori* (paper IV).

### Paper III

**Conclusion:** To design the P,H-COSY pulse sequence, we adopted concepts from multidimensional protein NMR. The semi-constant-time evolution central in the experiment yields high  $^{31}\text{P}$  resolution, similar efficiency in magnetization transfer for different phospholipids, and minimizes relaxation losses. Therefore cross-peak intensities in the two-dimensional spectrum can be measured to quantify phospholipids. A heteronuclear gradient echo exclusively selects  $^{31}\text{P},^1\text{H}$  correlations so that phospholipids can be observed in crude lipid extracts. To increase sensitivity, the magnetization is detected on  $^1\text{H}$  as anti-phase signals transferred via  $^3\text{J}_{\text{HP}}$  couplings. Thus, a characteristic multiplet pattern is observed for different phospholipids in the  $^1\text{H}$  dimension, which allows for a reliable identification of known phospholipids and yields structural information for unknown species. Applied to lipid extracts of *H. pylori*, the experiment allowed the reliable quantification of all phospholipids and the identification of several unusual phospholipids.

### Paper IV

**Conclusion:** Here, the experiment developed in paper III was used to characterize vesicles shed by *H. pylori* by analyzing the phospholipid composition of different membranes. We showed that the phospholipid composition of vesicles closely resembles that of the outer membrane, with PE and CL as main components. Furthermore, inner membrane samples showed depletion in CL and PG and an absence of PC, which was present in the outer membrane, whole cell and vesicle samples. This result is in agreement with the observation that the vesicles contain the virulence related adherence proteins BabA and SabA, which are also located in the outer membrane of whole cells. The study suggests that vesicles are important carriers of virulence factors, such as adherence proteins, and that the activities of these proteins in whole cells were similar to that in large vesicles, but lower in small vesicles. While the reason for this difference remains elusive, it may be related to a variation in phospholipid content associated to the size of the vesicles. By quantification of  $^{13}\text{C},^1\text{H}$  correlation spectra, we also estimated that *H. pylori* membranes contain about 10% cholesterol, which must be taken up by the bacteria from the host.

**Outlook:** To further characterize the dependence of the activity of adherence proteins on phospholipid composition, we will analyze membrane samples from different bacterial strains, grown under different conditions, and therefore differing in virulence. Characterization of the phospholipid requirements of outer membrane proteins will aid the reconstitution of these proteins and therefore their further characterization. It may also point to specific lipid-protein interactions or the requisite of specific membrane structures for adhesion, such as phospholipid micro-domains.

## 9 Acknowledgment

I firstly want to thank very much the whole department of Medical Biochemistry and Biophysics, affectionate called “**medchem**”: I think this is the greatest department I have ever seen, the best working atmosphere one can imagine!

As the most important, I want to thank very much “*meinem Doktorvater*”/supervisor **Jürgen Schleucher** for giving me the opportunity to work in his group with NMR. Thank you for your everlasting support and interesting discussion and for the many exciting projects. Thanks for the encouragement and freedom to follow my own ideas and guidance in the right direction, when I lost my track. I really learned what it means to do research by working with you. I also want to thank you for a relaxed and great working atmosphere and a good friendship over the years.

I furthermore would like to thank the NMR people with **Gerhard Gröbner**, **Göran Larsson**, and **Tobias Sparmann**, who are great PIs, always a helping and listening to all problems I found, and a lot of good company.

I want to thank my office mates **Miriam Larsson** and specially **Peter Gimeson** and **Linus Malm**, with whom I have had great fun, thanks for many good advices and deep discussions, sometimes with a bit too much alcohol ☺ – I will never forget the good old times, when the disco ball was blinking on Fridays. Thank you for your friendship!

I also want to thank the former group members **Elke Duchardt**, **Angela Augusti**, **Tatjana Betson**, **Aster Gebrekirstos** and **Tobias Tengel** for a nice time working together and last but not least, the new ones in the group: **Louise Rundqvist** and **Andrea Vincent** for many happy hours spent together.

Special thanks to **Ingrid Råberg** and **Clas Wikström** for keeping the whole medchem crowd going and for creating such a great working place.

Thanks to my summer students, **Maria Forsell** and **Pär Johansson**, who always makes me and other people smile.

I want to thank my second supervisor **Jörgen Johansson**. Big thanks to my cooperation partner in Nijmegen **Sybren Wijmenga**, **Frank Nelisson**: I learned a lot from you and it was always so easy! Thanks to people helping me with the simulations, especially **Kresten Lindorff-Larssen**, who took great time to explain me all my questions. Also **Janusz Zdunek** and **Xaviar Salvatelle** for help with the code. Thanks also to the InHouse cooperation on the world spanning *Helicobacter* project to **Anna Arnqvist** and **Thomas Borén** and their groups

## Acknowledgment

And of course thanks to the organization for financing much of the traveling and exchange with other labs: the “Stiftelsen JC Kempes Minne Stipendiefond”, Svenska Sällskapet för Medicinsk Forskning (SSMF), Marie Curie and the European Molecular Biology Organization (EMBO).

I want to thank the IKSU beach people and IKSU volley for all the nice years – my first Swedish words were “gratis” and “nätet”, and here especially **Charlotta Sundin**, who was a great partner on the net the last year!

At the end I want to thank some people with whom I have been working over the years, but which most importantly became great friends to me:

**Annelie Olofsson**, thanks for all the kind words and the care you have in your heart, stay as you are!

I want to thank five women who have tremendously influenced my life with their friendship over this long time:

**Nasim Sabouri**, thanks for the beach time (silver in Övik I will never forget), your honesty and your great heart.

**Rima Sulniute**, you are the most energetic happy person I have met and our ways go long back with sooo many laughs, I am very grateful for your friendship.

**Anna Vallström**, my writing body, you are my ideal Swedish person, direct, honest and you have an absolutely lovely personality.

**Ana Virel**, even if you have left quite a while ago already, Umeå without you would have never been like this. You were the first person I really got to know here and I immediately felt like home and well taken care off; you are amazing.

And last but definitely not least, **Mascha Thurm**, your heart is bigger than the universe! I have never met a person like you and being your friend made my life so much richer! I hope we will always stay in contact, all of you!

Ein großen Dank für die Unterstützung, Hilfe and beruhigenden Worte geht an meine **Familie**, ohne die ich sicher heute nicht wäre was ich bin! Ein ganz spezielles Dankeschön geht dabei an meine Eltern **Petra und Hans-Jürgen Höllige** und an meinem Bruder **Alexander Höllige**, sowie an meine Großeltern **Rita und Günter Rüger**.

My final thanks goes to **Magnus Rehn**, thank you for being there my love, There are no words to describe what you mean to me, you are „unbezahlbar“!

## 10 References

- Al-Hashimi, H. M., **2005**. Dynamics-based amplification of RNA function and its characterization by using NMR spectroscopy. *ChemBiochem* 6: 1506-1519.
- Al-Hashimi, H. M., **2007**. Beyond static structures of RNA by NMR: folding, refolding, and dynamics at atomic resolution. *Biopolymers* 86: 345-347.
- Alm, R. A., Bina, J., Andrews, B. M., Doig, P., Hancock, R. E. and Trust, T. J., **2000**. Comparative genomics of *Helicobacter pylori*: analysis of the outer membrane protein families. *Infect Immun* 68: 4155-4168.
- Atherton, J. C., **2006**. The pathogenesis of *Helicobacter pylori*-induced gastro-duodenal diseases. *Annu Rev Pathol* 1: 63-96.
- Aue, W. P., Bartholdi, E. and Ernst, R. R., **1976**. Two-dimensional spectroscopy. Application to nuclear magnetic resonance. *J. Chem. Phys* 64: 2229.
- Bäckstrom, A., Lundberg, C., Kersulyte, D., Berg, D. E., Boren, T. and Arnqvist, A., **2004**. Metastability of *Helicobacter pylori* bab adhesin genes and dynamics in Lewis b antigen binding. *Proc Natl Acad Sci USA* 101: 16923-16928.
- Beck, J. and Nassal, M., **2007**. Hepatitis B virus replication. *World J Gastroenterol* 13: 48-64.
- Bloch, F., Hansen, W. W. and Packard, M. E., **1946**. Nuclear Induction. *Physical Review* 69: 680.
- Bouvignies, G., Bernado, P., Meier, S., Cho, K., Grzesiek, S., Bruschweiler, R. and Blackledge, M., **2005**. Identification of slow correlated motions in proteins using residual dipolar and hydrogen-bond scalar couplings. *Proc Natl Acad Sci USA* 102: 13885-13890.
- Bugaytsova, J., Vikström, S., Henriksson, S., Björnham, O., Mendez, M., Aisenbray, C., Bylund, G., Mahdavi, J., Ögren, J., Ilver, D., Sjöström, R., Gilman, R. H., Chowdhury, A., Asish, M. K., Engstrand, L., et al., **2009**. pH Regulated H. Pylori Adherence: Implications for Persistent Infection and Disease. Manuscript
- Bukholm, G., Tannaes, T., Nedenskov, P., Esbensen, Y., Grav, H. J., Hovig, T., Ariansen, S. and Guldvog, I., **1997**. Colony variation of *Helicobacter pylori*: pathogenic potential is correlated to cell wall lipid composition. *Scand J Gastroenterol* 32: 445-454.
- Chitcholtan, K., Hampton, M. B. and Keenan, J. I., **2008**. Outer membrane vesicles enhance the carcinogenic potential of *Helicobacter pylori*. *Carcinogenesis* 29: 2400-2405.
- Cronan, J. E., **2003**. Bacterial membrane lipids: where do we stand? *Annu Rev Microbiol* 57: 203-224.
- Dickinson, W. C., **1950**. Dependence of the F19 Nuclear Resonance Position on Chemical Compound. *Phys.Rev.* 77: 736-737.

- Doerrler, W. T., **2006**. Lipid trafficking to the outer membrane of Gram-negative bacteria. *Mol Microbiol* 60: 542-552.
- Duchardt, E. and Schwalbe, H., **2005**. Residue Specific Ribose and Nucleobase Dynamics of the cUUCGg RNA Tetraloop Motif by MNMR (13)C Relaxation. *J Biomol NMR* 32: 295-308.
- Engelman, D. M., **2005**. Membranes are more mosaic than fluid. *Nature* 438: 578-580.
- Epand, R. M. and Epand, R. F., **2009**. Lipid domains in bacterial membranes and the action of antimicrobial agents. *Biochim Biophys Acta* 1788: 289-294.
- Ernst, R. R. and Anderson, W. A., **1966**. Application of Fourier transform spectroscopy to magnetic resonance. *Rev. Sci. Instr.* 37: 93-102.
- Flodell, S., Petersen, M., Girard, F., Zdunek, J., Kidd-Ljunggren, K., Schleucher, J. and Wijmenga, S., **2006**. Solution structure of the apical stem-loop of the human hepatitis B virus encapsidation signal. *Nucleic Acids Res* 34: 4449-4457.
- Flodell, S., Schleucher, J., Cromsigt, J., Ippel, H., Kidd-Ljunggren, K. and Wijmenga, S., **2002**. The apical stem-loop of the hepatitis B virus encapsidation signal folds into a stable tri-loop with two underlying pyrimidine bulges. *Nucleic Acids Res* 30: 4803-4811.
- Freeman, R., **1995**. A short history of NMR. *Khim. Geterotsikl. Soedin.*: 1157-1158.
- Fuchs, B. and Schiller, J., **2009**. Lysophospholipids: their generation, physiological role and detection. Are they important disease markers? *Mini Rev Med Chem* 9: 368-378.
- Furtig, B., Richter, C., Wohnert, J. and Schwalbe, H., **2003**. NMR spectroscopy of RNA. *Chembiochem* 4: 936-962.
- Getz, M., Sun, X., Casiano-Negroni, A., Zhang, Q. and Al-Hashimi, H. M., **2007**. NMR studies of RNA dynamics and structural plasticity using NMR residual dipolar couplings. *Biopolymers* 86: 384-402.
- Ghany, M. and Liang, T. J., **2007**. Drug targets and molecular mechanisms of drug resistance in chronic hepatitis B. *Gastroenterology* 132: 1574-1585.
- Girard, F. C., Ottink, O. M., Ampt, K. A., Tessari, M. and Wijmenga, S. S., **2007**. Thermodynamics and NMR studies on Duck, Heron and Human HBV encapsidation signals. *Nucleic Acids Res* 35: 2800-2811.
- Glebe, D. and Urban, S., **2007**. Viral and cellular determinants involved in hepadnaviral entry. *World J Gastroenterol* 13: 22-38.
- Grzesiek, S. and Bax, A., **1993**. Amino acid type determination in the sequential assignment procedure of uniformly <sup>13</sup>C/<sup>15</sup>N-enriched proteins. *J Biomol NMR* 3: 185-204.
- Haest, C. W., de Gier, J., den Kamp, J. O., Bartels, P. and van Deenen, L. L., **1972**. Changes in permeability of *Staphylococcus aureus* and

- derived liposomes with varying lipid composition. *Biochim Biophys Acta* 255: 720-733.
- Hall, K. B., **2008**. RNA in motion. *Curr Opin Chem Biol* 12: 612-618.
- Hirai, Y., Haque, M., Yoshida, T., Yokota, K., Yasuda, T. and Oguma, K., **1995**. Unique cholesteryl glucosides in *Helicobacter pylori*: composition and structural analysis. *J Bacteriol* 177: 5327-5333.
- Hu, J. and Lin, L., **2009**. RNA-protein interactions in hepadnavirus reverse transcription. *Front Biosci* 14: 1606-1618.
- Hu, K., Beck, J. and Nassal, M., **2004**. SELEX-derived aptamers of the duck hepatitis B virus RNA encapsidation signal distinguish critical and non-critical residues for productive initiation of reverse transcription. *Nucleic Acids Res* 32: 4377-4389.
- Ilver, D., Arnqvist, A., Ogren, J., Frick, I. M., Kersulyte, D., Incecik, E. T., Berg, D. E., Covacci, A., Engstrand, L. and Boren, T., **1998**. *Helicobacter pylori* adhesin binding fucosylated histo-blood group antigens revealed by retagging. *Science* 279: 373-377.
- Keeler, J., **2005**. Understanding NMR Spectroscopy (Wiley).
- Kelly, S. M. and Price, N. C., **2000**. The use of circular dichroism in the investigation of protein structure and function. *Curr Protein Pept Sci* 1: 349-384.
- Korzhnev, D. M., Billeter, M., Arseniev, A. S. and Orekhov, V. Y., **2001**. NMR studies of Brownian tumbling and internal motions in proteins. *Prog Nucl Mag Res Sp* 38: 197-266.
- Lange, O. F., Lakomek, N. A., Fares, C., Schroder, G. F., Walter, K. F., Becker, S., Meiler, J., Grubmuller, H., Griesinger, C. and de Groot, B. L., **2008**. Recognition dynamics up to microseconds revealed from an RDC-derived ubiquitin ensemble in solution. *Science* 320: 1471-1475.
- Lavanchy, D., **2008**. Chronic viral hepatitis as a public health issue in the world. *Best Pract Res Clin Gastroenterol* 22: 991-1008.
- Leontis, N. B. and Westhof, E., **2003**. Analysis of RNA motifs. *Curr Opin Struct Biol* 13: 300-308.
- Levitt, M. H., **2005**. spin dynamics: basics of nuclear magnetic resonance (John Wiley & Sons Ltd., Chichester).
- Lindorff-Larsen, K., Best, R. B., Depristo, M. A., Dobson, C. M. and Vendruscolo, M., **2005**. Simultaneous determination of protein structure and dynamics. *Nature* 433: 128-132.
- Mahdavi, J., Sonden, B., Hurtig, M., Olfat, F. O., Forsberg, L., Roche, N., Angstrom, J., Larsson, T., Teneberg, S., Karlsson, K. A., Altraja, S., Wadstrom, T., Kersulyte, D., Berg, D. E., Dubois, A., et al., **2002**. *Helicobacter pylori* SabA adhesin in persistent infection and chronic inflammation. *Science* 297: 573-578.
- Matsumoto, K., Kusaka, J., Nishibori, A. and Hara, H., **2006**. Lipid domains in bacterial membranes. *Mol Microbiol* 61: 1110-1117.

- Mitra, K. and Frank, J., **2006**. Ribosome dynamics: Insights from atomic structure modeling into cryo-electron microscopy maps. *Annu Rev Bioph Biom* 35: 299-317.
- Nagayama, K. and Wuthrich, K., **1977**. Two-dimensional NMR spectroscopy. A powerful tool for the investigation of biopolymers in solution. *Naturwissenschaften* 64: 581-583.
- Narberhaus, F., Waldminghaus, T. and Chowdhury, S., **2006**. RNA thermometers. *FEMS Microbiol Rev* 30: 3-16.
- Nassal, M., **2008**. Hepatitis B viruses: reverse transcription a different way. *Virus Res* 134: 235-249.
- Odenbreit, S., **2005**. Adherence properties of *Helicobacter pylori*: impact on pathogenesis and adaptation to the host. *Int J Med Microbiol* 295: 317-324.
- Palmer, A. G., 3rd and Massi, F., **2006**. Characterization of the dynamics of biomacromolecules using rotating-frame spin relaxation NMR spectroscopy. *Chem Rev* 106: 1700-1719.
- Parkin, D. M., **2004**. International variation. *Oncogene* 23: 6329-6340.
- Reichow, S. and Varani, G., **2006**. Structural biology: RNA switches function. *Nature* 441: 1054-1055.
- Sachs, J. N. and Engelman, D. M., **2006**. Introduction to the membrane protein reviews: the interplay of structure, dynamics, and environment in membrane protein function. *Annu Rev Biochem* 75: 707-712.
- Sattler, M., Schleucher, J. and Griesinger, C., **1999**. Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. *Prog Nucl Mag Res Sp* 34: 93-158.
- Schleucher, J., Quant, J., Glaser, S. J. and Griesinger, C., **1995**. A Theorem Relating Cross-Relaxation and Hartmann-Hahn Transfer in Multiple-Pulse Sequences - Optimal Suppression of TOCSY Transfer in ROESY. *J Magn Reson Ser A* 112: 144-151.
- Schleucher, J. and Wijmenga, S. S., **2002**. How to detect internal motion by homonuclear NMR. *J Am Chem Soc* 124: 5881-5889.
- Schwalbe, H., Buck, J., Furtig, B., Noeske, J. and Wohnert, J., **2007**. Structures of RNA switches: insight into molecular recognition and tertiary structure. *Angew Chem Int Ed Engl* 46: 1212-1219.
- Shajani, Z. and Varani, G., **2007**. NMR studies of dynamics in RNA and DNA by <sup>13</sup>C relaxation. *Biopolymers* 86: 348-359.
- Shao, S. H., Wang, H., Chai, S. G. and Liu, L. M., **2005**. Research progress on *Helicobacter pylori* outer membrane protein. *World J Gastroenterol* 11: 3011-3013.
- Shimomura, H., Hayashi, S., Yokota, K., Oguma, K. and Hirai, Y., **2004**. Alteration in the composition of cholesteryl glucosides and other

## References

- lipids in *Helicobacter pylori* undergoing morphological change from spiral to coccoid form. *FEMS Microbiol. Lett.* 237: 407-413.
- Stryer, L., Berg, J. M. and Tymoczko, J. L., **2002**. *Biochemistry* (W.H. Freeman and Company, New York).
- Tannaes, T., Bukholm, I. K. and Bukholm, G., **2005**. High relative content of lysophospholipids of *Helicobacter pylori* mediates increased risk for ulcer disease. *FEMS Immunol. Med. Microbiol.* 44: 17-23.
- Tannaes, T., Dekker, N., Bukholm, G., Bijlsma, J. J. and Appelmelk, B. J., **2001**. Phase variation in the *Helicobacter pylori* phospholipase A gene and its role in acid adaptation. *Infect Immun* 69: 7334-7340.
- van Meer, G., Voelker, D. R. and Feigenson, G. W., **2008**. Membrane lipids: where they are and how they behave. *Nat Rev Mol Cell Biol* 9: 112-124.
- Waters, L. S. and Storz, G., **2009**. Regulatory RNAs in bacteria. *Cell* 136: 615-628.
- Yeagle, P. L., Bentz, J., Boesze-Battaglia, K., Cherry, R. J., Edidin, M., Epand, R. M., Gawrisch, K., Gruner, S. M., Hauser, H., Huang, C. H., Jennings, M. L., Lewis, R. N. A. H., McElhaney, R. N., Mittal, A., Ohki, S., et al., **2005**. *The Structure of Biological Membranes* (CRC Press LLC, Boca Raton).
- Zuiderweg, E. R., **2002**. Mapping protein-protein interactions in solution by NMR spectroscopy. *Biochemistry* 41: 1-7.