High resolution in 2D NMR using spectral aliasing : application to the determination of pKa's by automated NMR titration and the study of 13C-enriched cholesterol

SHIVAPURKAR, Rupali

Abstract
La méthodologie RMN, basée sur le repliement spectral pour augmenter la résolution en dimension indirecte des spectres 2D, a été appliquée à la détermination des constantes d'acidité de mélanges d'acides carboxyliques. Avec l'aide de spectres HSQC à haute résolution obtenus dans des conditions de repliement, nous avons suivi très finement les variations de déplacements chimiques carbone-13. La méthode de titration n'est pas seulement applicable au pKa mais à d'autres constantes chimiques. Le second projet a consisté en la production de cholestérol marqué au 13C par une souche de levure modifiée. Nous avons obtenu un marquage à 92% avec du glucose u-13C6, 99%. Le cholestérol enrichi au 13C devrait trouver de nombreuses applications en RMN liquide et solide.

Reference

URN : urn:nbn:ch:unige-173007
DOI : 10.13097/archive-ouverte/unige:17300

Available at:
http://archive-ouverte.unige.ch/unige:17300

Disclaimer: layout of this document may differ from the published version.
High resolution in 2D NMR using spectral aliasing.

Application to the determination of pKₐ’s by automated NMR titration and the study of ¹³C-enriched cholesterol

THÈSE

présentée à la Faculté des sciences de l’Université de Genève
pour obtenir le grade de Docteur ès sciences, mention chimie

par

Rupali SHIVAPURKAR

de

Solapur (Inde)

Thèse N° 4358

GENÈVE

Atelier d'impression ReproMail

2011
Doctorat ès sciences
Mention chimie

Thèse de Madame Rupali SHIVAPURKAR
intitulée :

"High Resolution in 2D NMR Using Spectral Aliasing.
Application to the Determination of pKa’s by Automated NMR
Titration and the Study of $^{13}$C-Enriched Cholesterol"

La Faculté des sciences, sur le préavis de Messieurs D. JEANNERAT, docteur et directeur de
thèse (Département de chimie organique), S. MATILE, professeur ordinaire (Département
de chimie organique), K. PERRON, docteur (Département de botanique et biologie
végétale et Section des sciences pharmaceutiques) et J.-M. NUZILLARD, professeur
(Institut de chimie moléculaire, Université de Reims Champagne-Ardenne, France),
autorise l’impression de la présente thèse, sans exprimer d’opinion sur les propositions qui y
sont énoncées.

Genève, le 10 octobre 2011

Thèse - 4358 -

Le Doyen, Jean-Marc TRISCONE

N.B.- La thèse doit porter la déclaration précédente et remplir les conditions énumérées
dans les "Informations relatives aux thèses de doctorat à l'Université de Genève".
Dedicated to my loving aunt Babi
Acknowledgements

All the work behind this thesis could not have been possible without the help and support of many people. I would like to express my gratitude to all those people who contributed one way or another, to this work.

First of all, my thanks go to my supervisor Dr. Damien Jeannerat for his support, encouragement, motivation, and confidence through these three years. I would also thank him for giving me the opportunity to carry out my doctoral studies in his group.

It is a great pleasure to have Dr. Jean-Marc Nuzillard (University of Reims), Prof. Stefan Matile (University of Geneva), and Dr. Karl Perron (University of Geneva) as the thesis jury. I would also like to thank them for reading my thesis.

I am also grateful to Prof. Riezman and his group members for allowing me to work in their laboratory for cholesterol biosynthesis and specially to Thomas for teaching me the lipid extraction.

I would like to thank Andre Pinto, Patrick Romanens and Stéphane Grass for the technical support during all these years, especially to André Pinto for the coffee with more sugar. I am also thankful to my colleagues Bruno, Mohammadali and Gaëtan for making wonderful environment in the lab. I convey my sincere thanks to Bruno and Ali specially for the constant support, cooperation and making work enjoyable during hard times.

I am thankful to all the past and present colleagues from the Department of organic chemistry and also friends from Physical chemistry and Pharmacy department. Thanks to group of Prof. Matile and Dr. Zumbühl, I enjoyed the journal clubs sessions. A special thanks to Monique and Sonia for the administrative work and Rémy Marclay for the construction of the pipette holder.

A special thanks to Ankit and Mohammadali for the corrections and formatting of the thesis.

Last but not least, I want to thank my family for the constant support and faith in me which is the source of encouragement throughout my life.
Abbreviations

1D, 2D: One-, two-dimensional
b.p.: Boiling point
BBHD: BroadBand Homo-Decoupling
COSY: COrrelation SpectroscopY
d: Doublet
D$_2$O: Deuterium oxide
DCI: Deuterium chloride
DFT: Density Functional Theory
DMF: Dimethyformamide
DMSO: Dimethylsulfoxide
FID: Free Induction Decay
FT: Fourier Transform
GCMS: Gas Chromatography-Mass Spectroscopy
GIAO: Gauge Invariant Atomic Orbitals
HCN: Hydrogen cyanide
HMBC: Heteronuclear Multiple-Bond Correlation
HPLC: High Pressure Liquid Chromatography
HSQC: Heteronuclear Single-Quantum Correlation
INEPT: Insensitive Nuclei Enhanced by Polarization Transfer
MRI: Magnetic Resonance Imaging
N.A.: Natural Abundance
NaOD: Sodium deuteroxide
NMR: Nuclear Magnetic Resonance
NOESY: Nuclear Overhauser Effect SpectroscopY
<table>
<thead>
<tr>
<th>Symbol</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>q:</td>
<td>Quartet</td>
</tr>
<tr>
<td>QNP:</td>
<td>Quadruple Nucleus Probe. Dual (switchable X1 or X2 or X3) and $^1$H</td>
</tr>
<tr>
<td>s:</td>
<td>Singlet</td>
</tr>
<tr>
<td>SNR:</td>
<td>Signal-to-noise ratio</td>
</tr>
<tr>
<td>SW:</td>
<td>Spectral width</td>
</tr>
<tr>
<td>t:</td>
<td>Triplet</td>
</tr>
<tr>
<td>TCI:</td>
<td>Triple resonance Carbon observe in proton Inverse detection</td>
</tr>
<tr>
<td>TD:</td>
<td>Time Domain</td>
</tr>
<tr>
<td>THF:</td>
<td>Tetrahydrofurane</td>
</tr>
<tr>
<td>TMS:</td>
<td>Tetramethylsilane</td>
</tr>
<tr>
<td>TOCSY:</td>
<td>TOtal Correlation SpectroscopY</td>
</tr>
<tr>
<td>TPPI:</td>
<td>Time Proportional Phase Incrementation</td>
</tr>
<tr>
<td>YE:</td>
<td>Yeast Extract</td>
</tr>
<tr>
<td>YNB:</td>
<td>Yeast Nitrogen Base</td>
</tr>
</tbody>
</table>

**Symbols**

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>%:</td>
<td>percentage</td>
</tr>
<tr>
<td>°:</td>
<td>degree</td>
</tr>
<tr>
<td>$B_0$:</td>
<td>Static magnetic field</td>
</tr>
<tr>
<td>$E$:</td>
<td>Energy</td>
</tr>
<tr>
<td>$h$:</td>
<td>Planck’s constant</td>
</tr>
<tr>
<td>$I$:</td>
<td>Spin quantum number</td>
</tr>
<tr>
<td>$J$:</td>
<td>Scalar coupling constant</td>
</tr>
<tr>
<td>$\gamma$:</td>
<td>Magnetogyric ratio</td>
</tr>
<tr>
<td>$\delta$:</td>
<td>Chemical shift</td>
</tr>
<tr>
<td>$\mu$:</td>
<td>Magnetic moment</td>
</tr>
</tbody>
</table>
v: Frequency

ω: Angular frequency

Units

Hz, MHz: Hertz, Megahertz

kDa: Kilodalton

M: molar

mg, g: milligram, gram

min: minute

ms, s: millisecond, second

ppm: parts per million

ul, ml: micro, milliliter

um, mm: micro, millimeter

v: volume
Summary in French

La résonance magnétique nucléaire (RMN) est fréquemment utilisée par les chimistes et biochimistes pour résoudre les structures moléculaires. Elle peut aussi donner des informations détaillées sur la dynamique, l’état des réactions et de l’environnement chimique des molécules organiques. La méthodologie RMN, basée sur le repliement spectral pour augmenter la résolution en dimension indirecte des spectres 2D, a été développée dans le groupe du Dr. Damien Jeannerat. Par le passé, la technique a été appliquée à la résolution de structures de produits naturels, l’étude de cinétiques et la mesure de vitesses de diffusion.

Dans ce travail, nous présenterons l’application du repliement spectral à la détermination des constantes d’acidité de mélanges d’acides carboxyliques. Les déplacements chimiques carbones ont été choisis pour leurs plus grande sensibilité aux variations de densité électronique des groupes fonctionnels auxquels ils sont liés. Avec l’aide de spectres HSQC à haute résolution obtenus dans des conditions de repliement, nous avons été capables de suivre très finement les variations de déplacements chimiques carbone-13. Des séries de spectres HSQC repliés ont pu être enregistrés en une nuit. La méthode de titration n’est pas seulement applicable au pKa mais aussi à d’autres constantes chimiques et à l’étude de différents type d’interactions chimiques.

Illustration de l’utilisation de la dimension carbone de spectres HSQC repliés pour évaluer avec précision le degré de protonation d’un acide carboxylique et en déduire la différence de pKa avec les autres composés présents dans le mélange. En présence d’un composé de référence servant de “pH-mètre” interne on peut calculer les pKa absolus.
Le second projet a consisté en la production de cholestérol marqué au $^{13}$C. Le cholestérol est un des plus importants lipides et a été l’objet de beaucoup d’attention à cause de son implication dans les maladies cardiovasculaires. Il joue aussi un rôle central dans les micro-domaines membranaires. Nous avons utilisé une levure *Saccharomyces cervisiae* modifiée pour produire du cholestérol marqué $^{13}$C. Nous avons obtenu un marquage à 92% avec du glucose $u$-$^{13}$C$_6$, 99%, et, à 10% par dilution avec du glucose à abondance naturelle. Pour simplifier les structures de couplage $^{13}$C nous avons réalisé un marquage partiel à partir de glucose $1$-$^{13}$C, 99%. Cette levure modifiée produit environ 1 mg de cholestérol par gramme de glucose dans 100 ml de milieu de culture. Le protocole peut être appliqué à d’autres stérols produits par la levure, comme par exemple l’ergostérol. La disponibilité du cholestérol enrichi au $^{13}$C devrait trouver de nombreuses applications en RMN liquide et solide.

Spectres 1D $^{13}$C du cholestérol enrichi au carbone-13 produit par une levure modifiée et utilisant comme source de carbone du glucose 92% d’enrichissement uniforme (haut) 10% d’enrichissement uniforme (milieux) et 16% d’enrichissement semi-selectif des carbones (en bas).
Summary in English

Nuclear Magnetic Resonance (NMR) spectroscopy is frequently used by chemists and biochemists for solving the structure of molecules. It provides detailed information about the structure, dynamic, reaction state and chemical environment. The NMR methodology based on spectral aliasing to increase the resolution in the indirect dimension of 2D NMR spectra was developed in the group of Dr. Damien Jeannerat. The technique was already applied for the structure determination of natural products, kinetic studies, and diffusion measurements.

Here we demonstrate the application of spectral aliasing for the determination of pK$_a$ of mixtures of organic acids. The carbon chemical shifts were selected to study the pK$_a$ of acids because they are more sensitive towards the changes in the electronic densities of the functional group than the proton they are bound to. With the help of high-resolution 2D-HSQC spectra recorded under aliasing conditions we were able to track the changes in carbon chemical shifts. Series of two-dimensional aliased HSQC spectra were recorded in overnight experiments. This titration method is not only applicable for pK$_a$ determination but can be also applied to the determination of other chemical constants, and many other studies of interacting molecules.

Illustration of the use of aliased HSQC spectra to accurately assess the degree of protonation of a carboxylic acid and deduce the difference in pK$_a$ with other compounds present in the mixture. In the presence of a reference compound used as internal pH meter the absolute pK$_a$ can be calculated.
The second project consisted in the production of $^{13}$C-labeled cholesterol. Cholesterol is one of the most important lipids and has attracted a lot of attention due to its involvement in cardiovascular diseases in humans. It also plays an important role in membrane microdomains. We used a genetically engineered *Saccharomyces cervisiae* yeast strain to produce $^{13}$C-cholesterol. We are able to produce 92% labeled cholesterol, 10% labeled cholesterol by dilution of $\nu^{13}$C$_6$, 99% glucose. We also produced semi-selectively labeled cholesterol from $1^{13}$C, 99% glucose. This method produces approximately 1 mg of cholesterol per gram of glucose in 100 ml of culture medium. The protocol can be extended to other sterols produced in yeast, for example, ergosterol. The availability of enriched cholesterol for NMR experiments should find many applications in liquid and solid-state NMR.

Comparison of the 1D carbon spectra of cholesterol produced by modified yeast cells fed with labeled glucose. 92% uniformly enriched (top) 10% uniformly enriched (middle) and 16% semi-selective enrichment of red carbons (bottom).
This research work led to the following presentations:

**JOURNAL ARTICLES**


**ORAL COMMUNICATIONS**

1. “NMR titration experiments based on aliased HSQC spectra”, 18\(^\text{th}\) Swiss NMR Symposium (Geneva, Switzerland, 9\(^\text{th}\) September 2009).


**POSTERS**

1. “An automatic titration device for the determination of chemical constants using 2D NMR. Application to the determination of pK\(_a\)’s in complex mixtures”, Swiss Chemical Society Meeting (Zurich, Switzerland, 11\(^\text{th}\) September 2008), 40\(^\text{th}\) CUSO Summer school on Supramolecular Chemistry (Villars, Switzerland, 23-27 August 2009).


3. “Competition experiments followed by aliased 2D-NMR spectra. Affinity study of small library of carbohydrate for a borohydrate”, Swiss Chemical Society Meeting (Lausanne, Switzerland, 4\(^\text{th}\) September 2009).

4. “NMR characterization of \(^{13}\)C Enriched cholesterol Biosynthesised by Yeast”, EUROMAR (Florence, Italy, 4-9 July 2010).

5. “Evaluation of 10-ppm HSQC NMR spectra to study complex system at the fast exchange regime. Application to the study of the association constants of xanthines for paracetemol”, Swiss Chemical Society Meeting (Zurich, Switzerland, 16\(^\text{th}\) September 2010).


7. “\(^{13}\)C, \(^2\)H, Enrichment of cholesterol and ergosterol for NMR applications”, Swiss Chemical Society Meeting (Lausanne, Switzerland, 9\(^\text{th}\) September 2011), SMASH 2011, (Chamonix, France, 18-21 September 2011).
Table of Contents

1 Nuclear Magnetic Spectroscopy 15
  1.1 Introduction 15
    1.1.1 Nuclear spin 15
    1.1.2 Absorption of energy 15
    1.1.3 Mechanism of resonance 17
  1.2 Chemical shift and chemical environment 17
  1.3 Coupling constants 19
  1.4 2D NMR 19
    1.4.1 2D $^1$J_{HC} correlation spectrum (HQSC sequence) 20
  1.5 High resolution in F1 22
    1.5.1 Spectral aliasing 22
    1.5.2 10-ppm HSQC spectra 23
    1.5.3 Assignment of signals in 10-ppm HSQC 24

2 Application of aliased HSQC to study chemical equilibria 27
  2.1 NMR analysis of complex systems 27
    2.1.1 Slow exchange 27
    2.1.2 Fast exchange 28
    2.1.3 Intermediate exchange 28
  2.2 NMR titrations 29
    2.2.1 Direct access to the states of compounds 29
    2.2.2 The problem of sensitivity 30
    2.2.3 Flexibility of NMR with respect to the solvent 30
    2.2.4 The ability to study mixtures 30
    2.2.5 The advantages of $^{13}$C NMR 31
    2.2.6 The advantages of 2D over 1D NMR 31
    2.2.7 The automatic NMR titration setup 33
  2.3 Application of aliased HSQC to the determination of pK_a’s 34
    2.3.1 The acid/base equilibrium 34
    2.3.2 HSQC-based titration experiments 37
    2.3.3 Determination of pK_a in D$_2$O 43
    2.3.4 Determination of pK_a in D$_2$O/DMSO mixtures 45
    2.3.5 Experimental details 46
    2.3.6 Relation between $\Delta$pK_a, errors and pK_a scale 52
3 Production of isotopically-enriched cholesterol

3.1 Isotope labeling in NMR
   3.1.1 Introduction
   3.1.2 Different organisms used for enrichment

3.2 Cholesterol
   3.2.1 Biochemical pathways of cholesterol
   3.2.2 Related sterols
   3.2.3 Production of isotopically-enriched cholesterol
   3.2.4 $^{13}$C enrichment for NMR experiments

3.3 Experimental part
   3.3.1 Sample preparation
   3.3.2 NMR acquisition parameters
   3.3.3 DFT-GIAO calculations

4 Bibliography
1 Nuclear Magnetic Spectroscopy

1.1 Introduction

Since its discovery in 1946 by Bloch\(^1\) and Purcell\(^2\) (Nobel Prize for physics in 1952) Nuclear Magnetic Resonance (NMR) has developed into major spectroscopic and imaging techniques with important applications in physics, chemistry, biology, and medicine. This is acknowledged by three Nobel Prizes awarded to R. R. Ernst in 1991, K. Wüthrich in 2002 for chemistry and biology applications and P. C. Lauterbur together with P. Mansfield in 2003 for medicine. Liquid-state and solid-state NMR has a great success in the study of molecular structure and dynamic and the technique of Magnetic Resonance Imaging (MRI) has become an essential part of day-to-day life in medicine to obtain three dimensional images of biological tissues.

1.1.1 Nuclear spin

The Spin is one of the fundamental properties of atomic nuclei. Spins come in multiple of \(\frac{1}{2}\) and can be + or –. The common nuclei which posses a nuclear spin include \(^1\text{H}\), \(^2\text{H}\), \(^13\text{C}\), \(^14\text{N}\), \(^17\text{O}\), and \(^19\text{F}\). The number of allowed spin states can be determined by its nuclear spin quantum number \(I\). For each nucleus the number \(I\) is a physical constant and there are \(2I+1\) allowed spin states with integral differences ranging from \(+I\) to \(-I\). For example, protons, with a spin quantum number \(I = \frac{1}{2}\) have two spin states: \(\frac{1}{2}\) and \(-\frac{1}{2}\).

1.1.2 Absorption of energy

In the presence of a static magnetic field, the spin state \(+1/2\) of a hydrogen nucleus is aligned with the field and has a lower energy than the spin state \(-1/2\), which, is opposed to the field. Nuclear magnetic resonance occurs when the nuclei aligned with the applied field are induced to absorb energy and change their spin orientations with respect to the applied field. Figure 1.1. illustrate this process for a hydrogen nucleus.\(^3\)
This energy difference is function of the strength of the static magnetic field $B_0$ as illustrated in Figure 1.2.

\[
\Delta E = f(B_0).
\]  \hspace{1cm} (1.1)

The stronger the magnetic field, the greater the energy difference between the possible spin states:

\[
\Delta E = f(B_0) = h\nu.
\]  \hspace{1cm} (1.2)

The magnitude of the energy level separation also depends on the nucleus involved. Each nucleus has a different ratio of magnetic moment, because they have a different charge and mass. This ratio is called “magnetogyric ratio” $\gamma$ which is a constant for each nucleus and determines the energy dependence on the magnetic field.

Since the angular momentum of the nucleus is quantized in units of $h/2\pi$, the final equation takes the form

\[
\Delta E = \gamma \left( \frac{h}{2\pi} \right) B_0 = h\nu.
\]  \hspace{1cm} (1.3)

Solving the frequency of the absorbed energy,
\[ \nu = \left( \frac{\gamma}{2\pi} \right) B_0, \]  

(1.4)

we obtain the resonance frequency \( \nu \) called Larmor frequency of a given isotope as a function of the magnetic field.

1.1.3 Mechanism of resonance

To illustrate the nature of nuclear spin evolution, one can consider the behavior of child’s spinning top. When applying an external magnetic field to, say, a proton, it absorbs it and begins to precess. The frequency at which the proton processes is directly proportional to the strength of magnetic field; the stronger the field the higher the rate of precession. For a proton, if \( B_0 \) is 1.41 Tesla, the frequency of precession is approximately 60 MHz.

![Figure 1.3](image)

**Figure 1.3.** (a) A top precessing in the earth’s gravitational field; (b) the precession of a spinning nucleus resulting from the influence of an static magnetic field.

When the frequency of the oscillating electric field component of the incoming radiation just matches the frequency of the electric field generated by the precessing nucleus, the two fields can be couple, and energy can be transferred from the incoming radiation to the nucleus, thus causing spin states to change. This condition is called as resonance, and the nucleus is said to resonate with the incoming electromagnetic wave.

1.2 Chemical shift and chemical environment

NMR spectroscopy would not be useful if all the nuclei were resonating at the exactly same resonance frequency \( \nu \). In fact “chemical properties” causes carbon and proton to absorb at slightly different frequencies. The frequency of absorption for a
nucleus of interest relative to the frequency of absorption of molecular standard is called the chemical shift of the nucleus. Normally tetramethylsilane (TMS) is used as the molecular standard for proton and carbon NMR spectroscopy. The chemical shift of a particular nucleus in a molecule gives chemists information about where the atom with that nucleus is located within the molecule.

In the NMR spectrum, the scale corresponds to the absorption energy. The chemical shift $\delta$ is defined as

$$\delta = (\nu - \nu_{TMS})/\nu_0$$  \hspace{1cm} (1.5)$$

Where $\nu - \nu_{TMS}$ is the frequency difference between the resonance of the signal of interest and the resonance of the TMS. Note that $(\nu - \nu_{TMS})$ is given in Hz. (typically between 0 Hz to a few thousand Hz) we call $\nu_0$ as the frequency of spectrometer which depends on the static magnetic field of the $B_0$ and is given megahertz (MHz). For example for $B_0 = 7.05$ T, $\nu_0 = 300$ MHz for $^1$H nuclei but only 75 MHz for $^{13}$C nuclei because the latter has a four times smaller magnetogyric ratio. At higher field, for example, $B_0 = 11.75$ T, $\nu_0 = 500$ MHz for $^1$H nuclei or 125 MHz for $^{13}$C nuclei.

The variation of the chemical shifts can be illustrated with methyl acetate. The $^1$H NMR spectrum shows two peaks: one at $\delta = 2.1$ ppm for the three equivalent protons of the methyl in $\alpha$ relative to the carbonyl and one at $\delta = 3.7$ ppm for the three equivalent protons of the methoxy. The latter protons are said to resonate downfield relative to the methyl protons. The three methyl protons are equivalent and the three methoxy protons are equivalent because of the rapid rotation about the carbon-carbon and carbon-oxygen single bonds. $^1$H or $^{13}$C nuclei are said to be chemically equivalent if they have exactly the same chemical shifts; in other words they experience the same magnetic field. This equivalence most commonly occurs as a result of molecular symmetry coupled with rapid rotations about single bonds, but can also occur by coincidence.
1.3 Coupling constants

NMR signals often have a structure due to spin-spin interactions. These couplings are due to the mutual interactions between the different nuclei within the molecule. They are transmitted via the electrons involved in the bonding of the atoms. If the two nuclei are coupled, they show a splitting in the NMR spectra which depends on the quantum number to the partner spin. If two spin $\frac{1}{2}$ nuclei are coupled to each other and they are chemically shifted from each other then each one of them should be a doublet. The distances between the individual peaks of a multiplet are called the coupling constant “$J$” and are reported in Hz to be independent of the field strength. The coupling constant gives information about the relative positions of the nuclei involved, for example through the Karplus relationship.

1.4 2D NMR

Conventional NMR spectra (one-dimensional spectra) are plots of intensity vs. frequency, but in two-dimensional spectroscopy intensity is plotted as a function of two frequencies, usually called $F_1$ and $F_2$. The signals found in these spectra identify pairs of interacting spins. In most cases the interactions are due to scalar or dipolar interactions. The two-dimensional spectroscopy includes experiments such as COSY (COrrelation SpectroscopY), HSQC (Heteronuclear Single Quantum Correlation), NOESY (Nuclear Overhauser Effect Spectroscopy), etc. The first two-dimensional experiment, the COSY, was proposed by Jean Jeener in 1971 but has been only implemented later by Walter P. Aue, Enrico Bartholdi and Richard R. Ernst, who published their work in 1976. There are various ways of representing such a spectrum on paper, but the one most usually used is to draw “contour plots” in which the intensity of the peaks is represented by contour lines drawn at suitable intervals, in the same way as in a topographical map. The position of each peak is specified by two frequency coordinates corresponding to $F_1$ and $F_2$. Two-dimensional NMR spectra are arranged so that the $F_1$ and $F_2$ coordinates of the peaks correspond to those found in the normal one-dimensional spectrum, and this relation is often emphasized by plotting the one-dimensional spectra alongside the $F_1$ and $F_2$ axis. An example of heteronuclear correlation spectrum is shown in Figure 1.4. The HSQC spectrum shows, for example, the correlation of the methyl carbon (20 ppm) with its directly-attached proton (1.12 ppm).
1.4.1 2D $^{1}J_{HC}$ correlation spectra (HQSC sequence)

The two-dimensional experiments can be broken down into four basic intervals: the preparation, the evolution, the mixing (A-C in Figure 1.5.) and the detection periods. The HSQC (Heteronuclear Single-Quantum Correlation) experiment is widely used for recording one-bond correlation spectra between $^{1}H$ and $^{13}C$ or, in protein NMR, between $^{1}H$ and $^{15}N$.

In the HSQC pulse sequence if the X nucleus is carbon then

A. The preparation period is an INEPT sequence ($^{1}H$ to $^{13}C$)
B. The $t_{1}$ evolution period allows for indirect $^{13}C$ chemical shift evolution
C. The mixing period is a reverse-INEPT sequence ($^{13}C$ to $^{1}H$)
D. The $t_{2}$ evolution period allows for direct $^{1}H$ chemical shift detection

![Figure 1.5. The pulse sequence of the HSQC experiment. Narrow lines represent 90° pulses and broad lines represent 180° pulses. The delay $\tau$ is set to $1/(2J_{CH})$; all phases are x unless otherwise indicated.](image)
We shall briefly discuss each element of the sequence with respect to chemical shift and scalar coupling evolution.

Periods $A$ and $C$ are spin echoes in which $180^\circ$ pulses are applied to both partners of coupling. It therefore follows that the offsets of spins $I$ and $S$ will be refocused, but the coupling between them will evolve throughout the entire period. As the total delay in the spin echo is $1/(2^1 J_{HC})$ the result will be the complete conversion of in-phase into anti-phase magnetization.

Period $B$ is a spin echo in which a $180^\circ$ pulse is applied only to spin $I$. Thus, the offset of spin $I$ is refocused, as is the coupling between spins $I$ and $S$. Only the offset of spin $S$ affects the evolution, which is indirectly, detected by incrementation of the $t_1$ delay.

In terms of quantum operators, the first pulse generates $-I_y$; during period $A$ this then becomes $-2I_zS_z$. The $90^\circ(y)$ pulse to spin $I$ turns this to $2I_zS_x$ and the $90^\circ(x)$ pulse to spin $S$ turns it to $-2I_zS_y$. The evolution during period $B$ is simply under the influence of the chemical shift of spin $S$:

$$-2I_zS_y \rightarrow +2I_yS_x\cos(\omega_s t_1) - 2I_zS_y\sin(\omega_s t_1).$$

(1.5)

The next two $90^\circ$ pulses transfer the first term to spin $I$, the second term is rotated into multiple quantum and is not observed

$$\rightarrow -2I_yS_x\cos(\omega_s t_1) + 2I_yS_x\sin(\omega_s t_1)$$

(1.6)

The first term on the right evolves during the period $C$ into in-phase magnetization (the evolution of offsets is refocused). So the final observable term $I_x\cos(\omega_s t_1)$ is modulated by the chemical shift and coupling.

$$\rightarrow +I_x\cos(\omega_s t_1) - \exp(i\omega t_2)\cos(\pi J_{CH} t_2)$$

(1.7)

The resulting spectrum is therefore an in-phase doublet in $F_2$, centered at the offset of spin $I$, and these peaks will both have the same frequency in $F_1$, namely the offset of spin $S$. In most cases a broadband carbon decoupling is applied during the acquisition to simplify the multiplet structure. Alternative sequences provide one-bond correlation spectra, but the HSQC has certain advantage over the HMBC. The HSQC pulse sequence is often embedded in much more complex sequences which are used to record two and three dimensional spectra of carbon-13 and nitrogen-15 labeled
proteins.

1.5 High resolution in F1

The resolution in heteronuclear two-dimensional spectra is not identical in the two axes. In the direct proton dimension it is fine but in the indirect F₁ carbon-13 dimension the signals are typically 100 Hz broad for 256 time increments. One can increase the number of time increments but this increase the experiment time. Spectral aliasing allows increasing F₁ resolution without increasing the experiment time.

1.5.1 Spectral aliasing

The low resolution observed in the carbon dimension of heteronuclear two-dimensional NMR experiments can be improved by reducing the spectral width in F₁ and increase the resolution by a factor of 20-25 with no significant increase in acquisition time. “Spectral Aliasing” is the phenomenon in which NMR peaks appear at frequencies that are different from their real frequencies. These peaks are called “aliased” and, depending on the acquisition scheme, they have different signs. This causes ambiguities in the chemical shifts.

![Diagram](image)

**Figure 1.4.** (A) When the spectral window is set below the full spectral width (SW₀), signals are either aliased (B) or folded (C) depending on the quadrature detection used in the indirect dimension. The results will look like cutting and overlapping the slices from a 2D spectrum.

Aliasing arise because of a property of FT-NMR, and applies to signals that are outside the selected spectral boundaries. By changing the spectral boundaries, for example by moving the carrier frequency (center of spectrum) and/or changing the
spectral width, different peaks can be selected for aliasing. When applying aliasing, it is important to tune the acquisitions parameters such that folded or aliased peaks do not overlap with other peaks and can easily be distinguished from non-aliased peaks.

In the time domain, the effect of the violation of the Nyquist condition is to increase the dwell time

$$DW = \frac{1}{2} SW$$  \hspace{1cm} (1.8)

which in turns increases the acquisition time

$$AQ = TD \cdot DW.$$  \hspace{1cm} (1.9)

Close resonances therefore have more time to distinguish themselves (see Figure 1.5).

**Figure 1.5.** (A) Schematic representation of the signals produced by two resonances with similar frequencies. The sampling of the data points for a full spectral window is quite short (B). Spectral aliasing extends the acquisition time in the indirect dimension (C).

### 1.5.2 10-ppm HSQC spectra

For obtaining aliased spectra it is only necessary to change the spectral window, i.e. replace the standard carbon spectral window (typically 200 ppm) with 10 ppm and sometimes adjust the number of time increments or carrier frequency. We have used 128-512 data points for the acquisition depending on the experiment and sample complexity. The choice of 10 ppm simplifies the determination of the real chemical shifts as will be illustrated in the next section.
1.5.3 Assignment of signals in 10-ppm HSQC

Figure 1.7 shows the full and 10-ppm spectra of α and β cubebin a natural product obtained from *Drymis winterii*. In the full spectrum, signals overlap near 39 ppm in the carbon dimension. The 10-ppm spectrum resolves the signals but one needs to know where to look for the corresponding signals in the 10-ppm spectrum. This assignment of the signals in the latter is straightforward. The correspondence of the signals between the full HSQC and the 10-ppm HSQC spectra can be determined by looking for signals with equal proton chemical shifts and compatible carbon chemical shifts. The "compatibility" has to explained: in the full HSQC the last figure of the chemical shift are unknown because of the width of signals in the carbon dimension. In the 10-ppm HSQC, one or two more figures can be directly read on the right of the period. The chemical shift value in the full HSQC (left side) of the spectrum close to 50 ppm is 52.00 ppm. To find out the chemical shift value of corresponding signal in the 10-ppm HSQC (right side) one has to consider the central digits of 52.0 ie "2.0" and try to find the signal which also has "2.0" as the second digit in the 10-ppm HSQC. The signals close to 72.04 ppm in the 10-ppm HSQC has these "2.0" digits so by combining these two chemical shifts one will get a real full chemical shift value of 52.04 ppm. For confirmation of the value with the 1D carbon spectrum can be also useful but it is normally not necessary.
Figure 1.7. Signal assignment in full and 10-ppm HSQC of the two epimers of cubebin.

The high-resolution of aliased spectra will be exploited to follow fine changes in $^{13}$C chemical shifts to determine pK$_a$’s (Chapter 2) and observe carbon-carbon coupling structures in 2D spectra of $^{13}$C-enriched cholesterol (Chapter 3).
2 Application of aliased HSQC to study chemical equilibria

2.1 NMR analysis of complex systems

In the recent years, the development of dynamic libraries and combinatorial chemistry\(^5\) highlighted the need of analytical tools to study the details of the processes involved in complex systems. The power of aliased 2D HSQC spectra to study complex systems will be illustrated in this Chapter. We should show that NMR could be used to analyze a set of equilibria schematized in Figure 2.1.

![Figure 2.1](image)

**Figure 2.1.** Schematic representation of a sample containing a mixture of compounds (colored shapes) competing for a titrant (black square).

Because the time scale of NMR experiments is relatively slow we have to distinguish two main cases depending on the kinetic of the reactions.

2.1.1 Slow exchange

Slow chemical exchange corresponds to the situations where the equilibria of Figure 2.1. are slow relative to the NMR time scale that is when the exchange rate \(k \ll \Delta \delta\), where \(\Delta \delta\) is the difference in resonance frequency between the mutually exchanging spins. The relative quantities of the compounds involved in the equilibria can be simply determined by integration of the NMR signals. The precision is not always very satisfactory, except in the favorable cases of 1D proton spectra, but following some modifications makes it possible to quantify signal in HSQC spectra.\(^6\)

In the context of collaboration with the group of Prof. Prins (University of
Padova, Italy) a former student of our group studied a set of equilibria illustrated in Figure 2.2.  

![Figure 2.2](image)

**Figure 2.2.** Schematic representation of a kinetic study of the slow equilibrium reactions of compounds competing for a hydrazone.

2.1.2 Fast exchange

Fast exchange means that the reaction takes place on a time scale smaller than the acquisition of the NMR signal, *i.e.* when the exchange rate $k \gg \Delta \delta$. In the case of the protonation/deprotonation of an acid or base, the exchange rates are quite fast and the change of the electronic environment about the protonation site almost instantaneous. Only one signal is therefore observed for any pair of exchange in molecules and the chemical shifts, scalar couplings and other properties of the molecules will be the average of the values of the two forms weighted by their relative populations. We should see that determining the relative proportions of the two exchanging forms can be quite precise because chemical shifts can be measured with high precision.

2.1.3 Intermediate exchange

In the intermediate regime the observed nuclei are present in two or more magnetically non-equivalent states and the fit of the broadening NMR signals makes
it possible to determine the exchange rates and obtain thermodynamic and kinetic exchange constants. Such dynamic studies have first been investigated in 1953 by Gutowsky et. al,\textsuperscript{8} and it has been used since mostly in 1D $^1$H spectra. In this work we shall consider only fast processes at room temperature.

### 2.2 NMR titrations

Titrations are among the most important experiments in chemistry. Most high-school students used color indicators reporting a change in the composition of a solution. In most cases, they observe the changes in the pH when studying acid/base properties of compounds, because the pKa is one of the most fundamental concepts in chemistry. The determination of pK\textsubscript{a} may seem old-fashioned, but it still plays an important role in protein structure determination,\textsuperscript{9} determination of the affinity for ligands\textsuperscript{10} and enzymatic activity,\textsuperscript{11} etc. NMR is extremely selective for distinguishing different species in solution, making it one of the most powerful methods for monitoring the extent of ionization of organic compounds.

#### 2.2.1 Direct access to the states of compounds

The main advantage of NMR is to directly observe the states of the molecules of interest instead of indirectly observe their effects through the solvent as for most other methods. NMR spectroscopy is being used more and more for the determination of acid-base equilibrium constants (pK\textsubscript{a}) of substances like amino acids,\textsuperscript{12} peptides,\textsuperscript{13} antibiotics\textsuperscript{14} or other substances containing acidic or basic groups such as alcohols, acids, amides, esters. We should see that the variations of NMR chemical shifts during the titration of a single solution allows one to determine relative pK\textsubscript{a}’s of mixtures of organic compounds without having to measure the pH of the solutions.\textsuperscript{15}

An additional advantage of NMR, when compared to electrode potential measurements, is to be able to study a particular substance present in a solution of physiological interest such as blood, plasma which is complex mixture often containing other acids and bases, without being disturbed by the other components of the matrix.
2.2.2 The problem of sensitivity

On the other hand, the NMR measurements have much lower sensitivity than potentiometric and UV-based methods. The sensitivity of NMR depends on the concentration of species, the particular isotope observed, and the pulse sequence used. But sensitivity is often a serious limitation because biological species are sometimes present in the micro molar concentration in the sample of interest, making measurements difficult even when using protons, the most sensitive and readily available isotope for NMR experiments.

2.2.3 Flexibility of NMR with respect to the solvent

The classical methods have difficulties to determine a pKₐ in the presence of salts or when trying to determine extreme values. In most cases, this is due to the difficulty in deducing the degree of protonation of the compounds of interest from a measure of the pH of the solution. NMR measurements directly report the degree of protonation of the compounds and have no particular problem with the determination at high and low pH, or in non-aqueous solvents like THF acetonitrile-water mixtures or other non-water containing binary solvent mixtures. The ability to determine pKₐ in non-aqueous conditions is very important for non water-soluble pharmaceutical drugs. NMR can also be used for in vivo measurements where one wishes to avoid adding electrodes or molecular sensors like dyes for UV detection. These advantages of NMR for titrations are not limited to the study of pKₐ and have been exploited to determine other chemical constants.

2.2.4 The ability to study mixtures

Another advantage of NMR is to be able to determine simultaneously the pKₐ’s of mixtures of acids or bases or all the pKₐ’s of the acid/base active side-chains of macromolecules. This is very interesting when studying small libraries or systems at dynamic exchange where separation of the components is impossible or when the goal is to study the influence of the medium on the pKₐ’s. The complexity of the samples is only limited by the ability to resolve and assign the signals to the individual components.
2.2.5 The advantages of $^{13}$C NMR

The $pK_a$ can be determined using $^1$H, $^{13}$C, $^{15}$N, $^{31}$P, or $^{19}$F spectroscopy.\textsuperscript{12b, 26} One-dimensional $^1$H NMR is at the front line of all NMR studies and provides the information about chemical shifts and coupling constants. The main disadvantage of 1D $^1$H NMR is the signal overlap due to the narrow dispersion of the $^1$H chemical shifts. Carbon-13 NMR overcomes the problem of resolution with a typical chemical shift range of $\sim$200 ppm compared to 15 ppm for $^1$H spectrum. It allows to study all carbons, including quaternary carbons (carbonyl, carboxylic acids, substituted sp2 carbons) while heteronuclear one-bound 2D experiments (see below) only detects carbons bond to one or more hydrogen atom. The major disadvantage of $^{13}$C NMR spectroscopy is its low sensitivity as compared to $^1$H NMR. This is because of the low natural abundance of the $^{13}$C nuclei (1.1\%) and four-fold lower magnetogyric ratio.

2.2.6 The advantages of 2D over 1D NMR

The determination of $pK_a$ by NMR is based on the analysis of the variation of the chemical shifts of spins located in the vicinity of the acidic or basic site in the molecule. It was originally focused on proton chemical shifts but other isotopes are potentially more interesting. The $^1$H is a quite sensitive nucleus and the most frequently used, but the overlap due to the small dispersion of signals, scalar coupling structures and strong coupling makes it very difficult to study mixture by proton NMR. (see Figure 2.3.)
COSY and TOCSY spectra can resolve signals in the dimension of a coupling partner when determining the pK<sub>a</sub> in peptide side-chain functional groups<sup>27</sup> but the other problems remain.

The development of superconducting magnets and FT-NMR made it possible to improve signal-to-noise ratio. Among the nuclei that benefitted the most of these developments, 13C is certainly the most important. Nowadays one can study a few milligrams of compounds at natural 13C abundance. The carbon chemical shifts were used to determine pK<sub>a</sub> of carboxylic acids<sup>28</sup>, amines<sup>28b, c, 29</sup>, amino acids, peptides and proteins. The very high dependence of 13C nucleus to structural parameters makes it attractive for pK<sub>a</sub> studies. As a consequence, 13C chemical shifts are more interesting because they are more sensitive to changes in the electronic densities of the directly bound functional groups than protons located one bond farther.

Figure 2.3. 1D proton spectra obtained during an acid/base titration experiment. Overlap (☆) and strong coupling (★) makes it very difficult to determine chemical shifts.
Table 2.1.1. Comparison of the variations of $^1$H and $^{13}$C chemical shifts of CH’s located in the vicinity of carboxylic groups changing protonation states

<table>
<thead>
<tr>
<th>Acid</th>
<th>CH</th>
<th>$^1$H ppm</th>
<th>$^{13}$C Hz&lt;sup&gt;a&lt;/sup&gt; ppm</th>
<th>$^1$H Hz&lt;sup&gt;a&lt;/sup&gt;</th>
<th>$^{13}$C Hz&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acetic acid</td>
<td>Me</td>
<td>0.188</td>
<td>94.0</td>
<td>2.897</td>
<td>362.1</td>
</tr>
<tr>
<td>Benzoic acid</td>
<td>o-</td>
<td>0.129</td>
<td>64.5</td>
<td>0.825</td>
<td>103.1</td>
</tr>
<tr>
<td></td>
<td>p-</td>
<td>0.124</td>
<td>62.0</td>
<td>2.549</td>
<td>318.7</td>
</tr>
<tr>
<td></td>
<td>m-</td>
<td>0.047</td>
<td>23.5</td>
<td>0.382</td>
<td>47.8</td>
</tr>
<tr>
<td>Camphamic acid</td>
<td>7-CH&lt;sub&gt;exo&lt;/sub&gt;&lt;sup&gt;b&lt;/sup&gt;</td>
<td>0.124</td>
<td>62.0</td>
<td>0.003</td>
<td>4.4</td>
</tr>
<tr>
<td>Me ($\gamma$)</td>
<td></td>
<td>0.023</td>
<td>11.5</td>
<td>0.008</td>
<td>10.0</td>
</tr>
</tbody>
</table>

<sup>a</sup> Measured at 500 MHz $^1$H Larmor frequency. <sup>b</sup> H<sub>exo</sub> of the CH<sub>2</sub> in \(\beta\) position relative to the carboxylic group.

Carbons typically present larger changes in chemical shifts (see the Table 2.1.1) except when the hydrogen atoms are in close contact with the protonation site. Moreover the carbon chemical shifts are less dependant on the concentration, ionic strength, solvent conditions that may change during the titration. Ultimately indirect measurements of carbon chemical shifts using HSQC experiments combine the advantages of proton NMR for sensitivity and carbon for signal dispersion provided the resolution is sufficiently high, a condition we satisfy with spectral aliasing.

2.2.7 The automatic NMR titration setup

We developed a computer-controlled automatic titration system to facilitate the acquisition of series of 2D NMR spectra. (Figure 2.4.). The titration setup controls the addition of the titrant inside the NMR tube and saves the time of manual addition as each addition can be made using an automatic syringe. The additions are made in between consecutive NMR measurements. Typically 25 additions of 20 ul of titrant are made in-field and each addition/measurement cycle takes approximately 40 minutes allowing a full titration to run overnight. In contrast with the usual NMR titration setup<sup>24b</sup>,<sup>30</sup> based on flow-NMR probes, our reactor consists in a simple elongated Pastor pipette fixed above the tube holder so that its tip reaches to the bottom of a standard 5 mm tube<sup>31</sup> (see Figure 2.4). The reactor (Figure 2.4.b) is
descended in the magnet and remains there until the end of the titration. The titrant is added into the pipette through a Teflon line by a computer-controlled push-syringe and mixed with the solution present in the NMR tube.

![Diagram](image)

**Figure 2.4.** (a) Scheme of the connections of the push syringe and peristaltic pump to the pipette (b) Plexiglas holder.

### 2.3 Application of aliased HSQC to the determination of pKₐ’s

#### 2.3.1 The acid/base equilibrium

In 1887 Arrhenius gave the first reasonable account of the property of acids and bases, namely that an "acid" gives rise to excess of H⁺ in aqueous solutions whereas "bases" gives rise to excess of OH⁻ in solution. This theory was refined by Bronsted and Lowry in 1923 defining acids as proton donor and bases as proton acceptors. They also introduced the concept of the conjugated acids and bases.

The pKₐ are very important. For example, the strength of hydrogen bonding has a direct relationship with the pKₐ’s. In biochemistry the pKₐ of protein side-chains play a major role in the activity of enzymes and stability of proteins. They also play an important role in analytical chemistry. In acid-base extraction, the efficiency of extraction of a compound into an organic phase, such as ether, can be optimized by adjusting the pH of the aqueous phase using an appropriate buffer.
2.3.1.1 The acid/base equilibrium

For every acid AH, there is a conjugate base, A⁻ and for every base B, there is a conjugate acid BH⁺

\[ \text{AH} \rightleftharpoons \text{H}^+ + \text{A}^- \]  \hspace{1cm} (2.1)

\[ \text{BH}^+ \rightleftharpoons \text{H}^+ + \text{B} \]

If AH (BH⁺) has a great tendency to lose protons, it follows that its conjugate, A⁻ (or B), has only a small tendency to accept protons. In aqueous solution, acids react with water acting as a base:

\[ \text{AH (or BH}^+ \text{)} + \text{H}_2\text{O (base)} \rightleftharpoons \text{H}_3\text{O}^+ + \text{A}^- \text{ (or B)} \]  \hspace{1cm} (2.2)

and bases react with water acting as an acid:

\[ \text{A}^- \text{(or B)} + \text{H}_2\text{O (acid)} \rightleftharpoons \text{AH (or BH}^+) \text{ + OH}^- \]  \hspace{1cm} (2.3)

The strength of AH, or BH⁺, relative to the base strength of water is given by the equilibrium constant

\[ K = [\text{H}_3\text{O}^+] [\text{A}^- \text{ or B}] / [\text{H}_2\text{O}] [\text{AH or BH}^+]. \]  \hspace{1cm} (2.4)

In the case of acid base reaction the acid dissociation constant \( K_a \) measures the strength of acid in a solution. The chemical species HA, A⁻ and H⁺ are said to be in equilibrium when their concentrations do not change with the passing of time. This equilibrium being quite fast we are clearly in the “fast regime” relative to NMR experiments. The dissociation constant is usually written as a quotient of the equilibrium concentrations (in mol/L), denoted by [HA], [A⁻] and [H⁺]:

\[ K_a = \frac{[\text{A}^-][\text{H}^+]}{[\text{HA}]} \]  \hspace{1cm} (2.5)

Here \( \text{H}^+ \) is solvated proton. The p\( K_a \) value is defined from \( K_a \), as

\[ \text{p}K_a = -\log_{10} K_a \]  \hspace{1cm} (2.6)

where \( p \) denotes the negative log of 10. This value is equal to the pH at which the concentrations of A and B are equal. Equation (2.6) expresses the strength of conjugate acids of an organic base. The p\( K_a \) values of common acids are shown in the
Table 2.1.2.

<table>
<thead>
<tr>
<th>Name of Compounds</th>
<th>Typical pK&lt;sub&gt;a&lt;/sub&gt; value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Propanoic acid CH&lt;sub&gt;3&lt;/sub&gt;CH&lt;sub&gt;2&lt;/sub&gt;CO&lt;sub&gt;2&lt;/sub&gt;H</td>
<td>4.9</td>
</tr>
<tr>
<td>Ethanoic acid CH&lt;sub&gt;3&lt;/sub&gt;CO&lt;sub&gt;2&lt;/sub&gt;H</td>
<td>4.7</td>
</tr>
<tr>
<td>Chloroethanoic acid CH&lt;sub&gt;2&lt;/sub&gt;ClCO&lt;sub&gt;2&lt;/sub&gt;H</td>
<td>2.9</td>
</tr>
<tr>
<td>Nitroethanoic acid O&lt;sub&gt;2&lt;/sub&gt;NCH&lt;sub&gt;2&lt;/sub&gt;CO&lt;sub&gt;2&lt;/sub&gt;H</td>
<td>1.7</td>
</tr>
</tbody>
</table>

2.3.1.2  Determination of pK<sub>a</sub>

There are several methods to determine the acidity constants. Among them potentiometric titration is the most widely used method. It has some drawbacks: it required glass electrodes and it need physical contact with the solution. Other commonly used methods are based on the spectrophotometric titrations. It is very sensible technique but requires a pH indicator in it. Capillary electrophoresis is the third important method for pK<sub>a</sub> determination. Different methods for determining acidity constants are discussed by Cookson.

2.3.1.3  pK<sub>a</sub> in non-aqueous solvents

Normally all the standard pK<sub>a</sub> values are determined in water. But the solvents have dramatic affects on the acidity of a compound.

Table 2.1.3. pK<sub>a</sub>’s of various acids in different solvents

<table>
<thead>
<tr>
<th>Acid</th>
<th>Solvent</th>
<th>H&lt;sub&gt;2&lt;/sub&gt;O</th>
<th>CH&lt;sub&gt;3&lt;/sub&gt;OH</th>
<th>DMSO</th>
<th>DMF</th>
<th>CH&lt;sub&gt;3&lt;/sub&gt;CN</th>
</tr>
</thead>
<tbody>
<tr>
<td>CH&lt;sub&gt;3&lt;/sub&gt;COOH</td>
<td></td>
<td>4.76</td>
<td>9.5</td>
<td>12.6</td>
<td>13.5</td>
<td></td>
</tr>
<tr>
<td>p-NO&lt;sub&gt;2&lt;/sub&gt;C&lt;sub&gt;6&lt;/sub&gt;H&lt;sub&gt;4&lt;/sub&gt;OH</td>
<td></td>
<td>7.15</td>
<td>11.4</td>
<td>11.0</td>
<td>12.6</td>
<td>21.0</td>
</tr>
<tr>
<td>HCN</td>
<td></td>
<td>9.3</td>
<td>-</td>
<td>12.9</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>malonitrile</td>
<td></td>
<td>11.0</td>
<td>-</td>
<td>11.0</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>PhNH&lt;sub&gt;2&lt;/sub&gt;&lt;sup&gt;-&lt;/sup&gt;</td>
<td></td>
<td>4.6</td>
<td>-</td>
<td>3.2</td>
<td>4.2</td>
<td>-</td>
</tr>
</tbody>
</table>

The pK<sub>a</sub> of a compound in water is quite different from that in DMSO or other solvent. But most of the time pK<sub>a</sub> values change of a similar quantity when changing solvents but they are many exceptions. In water HCN is a stronger acid than
malononitrile [CH$_2$(CN)$_2$], but in DMSO it is the opposite. Still, in organic solvents the pK$_a$ values are almost always larger because they are not as effective at supporting the charges that develop upon creating the positively charged ion and tend to spread over a larger pK$_a$ range. The effect of DMSO on the pK$_a$ of carboxylic acid was demonstrated by Ostrow. The dissociation constants of carboxylic acids were studied by $^{13}$C NMR spectroscopy by McDonagh in DMSO/water mixtures. In some cases, the pK$_a$ in an pure organic solvents has been interpolated from values obtained in mixtures with 5-20% water content.

2.3.2 HSQC-based titration experiments

We determined the pK$_a$ of the mixture of compounds of Figure 2.5.

![Figure 2.5](image)

Figure 2.5. Components of the mixture used for the NMR determination of their pK$_a$ in D$_2$O. 1 acetic acid, 2 camphoric acid, 3 2,3-dimethoxy benzoic acid, 4 benzoic acid, 5 caffeic acid, 6 coumaric acid, 7 4-hydroxy benzoic acid, 8 propionic acid.

The 1D proton spectra recorded during the acid/base titration of a mixture containing 1-6 (see Figure 2.3.) show severe overlap and signals are difficult to assign to individual components. As mentioned above, 1D carbon-13 spectra could be also used but the problem is low sensitivity, but we have shown that 2D $^1$H-$^{13}$C HSQC are much more sensitive.

In order to benefit of the larger changes in $^{13}$C chemical shift one has to be able to measure them with the same precision as in proton spectra. This is easy to do in 1D spectra where the signals are directly detected and the FID can be recorded with tens of thousands of points. But in the indirect $^{13}$C dimension of HSQC experiments, the desired precision would require excessively long experiments with standard
acquisition parameters. As discussed earlier Spectral aliasing\textsuperscript{42} is a straightforward solution to this problem. When signals are expected to change positions we recommend an approach consisting in setting the carbon spectral window to a non-optimized value.\textsuperscript{43} The goal is to obtain spectra with the highest possible resolution in the carbon dimension. Table 2.1.4. shows that 10-ppm spectra recorded with 256 time increments are quite convenient\textsuperscript{44} but it results to signals that are still about 8 Hz broad. When studying small molecules with T\textsubscript{2} > 1 s, one can push this resolution to the limit of field in homogeneity which is about 1 Hz. In such a case, (entry 5 in table 2.1.4. ) the signal truncation is negligible and one can refrain from using window functions except for SNR considerations. In spectra recorded with a maximal evolution time of 0.5 s, signals have a width at half-height of about 2 Hz which is quite similar to what is observed in 1D carbon spectra with 1 Hz line broadening. Aliased HSQC can therefore provide spectra where the measure of the $^{13}$C chemical shifts has the same precision as when using 1D carbon spectra and additionally benefit from the resolution of signals in the proton dimension and a higher sensitivity.

When sensitivity imposes a long acquisition, one should increase the number of time increments rather than the number of times each increment is recorded because it allows to increase the spectral window by the same factor without changing t\textsubscript{1,max}. Large windows reduce the probability that signals cross the boundaries of the spectrum and reduce the probability of accidental overlap. Neither of these events is problematic but reducing their instances facilitates the exploitation of the spectra.\textsuperscript{45}

**Table 2.1.4.** Signal width in the $^{13}$C dimension of HSQC experiment

<table>
<thead>
<tr>
<th>SW ppm</th>
<th>TD pt.</th>
<th>t\textsubscript{1, max} s</th>
<th>Line width/Hz\cos\alpha</th>
<th>No.proc.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 220$^a$</td>
<td>256</td>
<td>0.0047</td>
<td>214</td>
<td>176</td>
</tr>
<tr>
<td>2 10</td>
<td>256</td>
<td>0.1024</td>
<td>9.77</td>
<td>8.01</td>
</tr>
<tr>
<td>3 5</td>
<td>256</td>
<td>0.2048</td>
<td>4.88</td>
<td>4.00</td>
</tr>
<tr>
<td>4 5</td>
<td>625</td>
<td>0.4096</td>
<td>2.44</td>
<td>2.00</td>
</tr>
<tr>
<td>5 2.5</td>
<td>512</td>
<td>0.4096</td>
<td>2.44</td>
<td>2.00</td>
</tr>
<tr>
<td>6 2.5</td>
<td>312.5</td>
<td>0.8192</td>
<td>1.22</td>
<td>1.00</td>
</tr>
</tbody>
</table>

\textsuperscript{a} ppm. Hz\textsuperscript{3} TD pt. c s. Line width/Hz\cos\alpha. No.proc. d
a Contribution of the processing to the full line width at half height. The broadening is due to field inhomogenieties is not included.

b At 125 MHz $^{13}$C Larmor frequency.

c Number of time increments in the carbon dimension.

d Measured on the main signal. Sinc wiggles due to signal truncation may be quite strong when signals do not decay significantly (for $t_{max} < 200$ ms).

e Typical conditions for a full-width HSQC spectrum.

For the HSQC-based NMR titration experiment, the spectral windows were set to 2.5 ppm and resulted to the spectra of Figures 2.6. and 2.7. (see the characteristics of the spectra in Table 2.1.4. in entry 5 and the experimental section).

Figure 2.6. Superposition’s of 24 aliased HSQC spectra.

One difficulty of aliased spectra is to follow the track specifically when signals cross a spectral boundary (See Figure 2.7).
2.3.2.1 *Determination of ΔpKₐ’s*

In NMR spectroscopy the pKₐ is usually determined using the chemical shifts to calculate the ratio between the protonated and deprotonated forms of acid/base pair. Any reporter spin located in the neighboring of the site of exchange can be used to calculate the relative proportions of the two forms present, provided it is sensitive to the change.

The chemical shift of particular signal during the titrations and the chemical shifts at the beginning and the end of the titration must be known to accurately determine the ratio of acidity constants using this methodology. The degree of protonation $p$ can be determined from observed average chemical shift

$$
\delta^{\text{obs}} = p\delta^{\text{B-H}} + (1 - p)\delta^{\text{B-}},
$$

(2.7)

where $\delta^{\text{B-H}}$ and $\delta^{\text{B-}}$ are the chemical shifts of protonated and deprotonated forms and $p$ is the proportion of protonated form. The measurement of precise chemical shifts is important as the degree of protonation is inversely proportional to $|\delta^{\text{B-H}} - \delta^{\text{B-}}|$. The protonated/deprotonated ratio

$$
r = p/(1-p)
$$

(2.8)

can be inserted into the equation of the acidity constants,
to express the acidity constant as a function of \( r \):

\[
K_a = \frac{|B^-|}{|H^+|} \frac{1}{|BH|} \tag{2.9}
\]

In the case of mixtures, the equalities

\[
K_{a_i}r_i = K_{a_j}r_j \tag{2.11}
\]

hold for the pair of acids or bases \( i \) and \( j \) and result to:

\[
\frac{K_{a_i}}{K_{a_j}} = \frac{(\delta^B_j - \delta^B_i)(\delta^B_i - \delta^B_D)}{(\delta^D_i - \delta^B_i)(\delta^B_j - \delta^B_i)}. \tag{2.12}
\]

This gives the difference in acidity constants:

\[
\Delta pK_{a_{ij}} = \log\left(\frac{K_{a_i}}{K_{a_j}}\right). \tag{2.13}
\]

In principle six chemical shifts are necessary for each acid to determine the \( \Delta pK_{a_{ij}} \). Two are the fully protonated (\( \delta^{B^+} \), \( \delta^{B+} \)) and deprotonated (\( \delta^{B^-} \), \( \delta^{B-} \)) values and the others when the \( i \) and \( j \) are both are partially protonated (\( \delta^{obs}_i \), \( \delta^{obs}_j \)), but in practice we measure data for many partial protonations. When plot together, the points of coordinates

\[
x = (\delta^{obs}_j - \delta^{B^+}_j)(\delta^{B+}_i - \delta^{obs}_i),
\]

\[
y = (\delta^{obs}_i - \delta^{B^-}_i)(\delta^{B^-}_j - \delta^{obs}_j) \tag{2.14}
\]

should be on the straight line with a slope \( K_{a_j}/K_{a_i} \).

2.3.2.2 Normalization and linearization of \(^{13}C \) chemical shifts

For the normalization of the chemical shifts one requires the chemical shifts in protonated (\( \delta^{B^+} \)) and deprotonated (\( \delta^{B^-} \)) forms. The precise values of these boundary chemical shifts are fundamental because the coordinates of the equation (2.14) gives a straight line only when the values of \( \delta^{B^-} \) and \( \delta^{B+} \) were correct.
Figure 2.8. (A) The simulation of normalized chemical shifts of four acids when there is difference in $pK_a$ as 0.1, 1 and 2 units. (B) X/Y plot of the linearized chemical shifts. (C) Deviation effect.

The precision of $\Delta pK_a$ is much higher if the $K_{ai}$ and $K_{aj}$ are close. If one acid terminates its change of protonation before the other starts then the $K_{ai}/K_{aj}$ are not defined as two of the parenthesis of equation 2.12 becomes zero. The limit situation occurs when the $pK_a$ of one acid is two units away from the one as illustrated in the Figure 2.8-A with the dotted and continues lines. Here the precision for the $|\Delta pK_a| \approx 1$ is $\pm 0.01$. Sometimes it is difficult to determine the boundary values of protonated and deprotonated forms. If one of these values is wrong, the data points deviate from a straight line but this property allows a program to adjust the problematic boundary values until the points perfectly align. (see Figure 2.9-B)

2.3.2.3 Gran’s method

It is important to adjust the pH of the starting solution so that the protonation or deprotonation of the molecule starts only after two or three additions of the titrant. The titration should continue after the molecule of interest has completely changed the protonation state. This allows to use Gran’s method to improves the values of boundary chemical shifts (see Figure 2.9-A).
Figure 2.9. Normalization of chemical shifts for (A) β-caffeic and (B) ortho 2,3 dimethoxy benzoic acid. (A) Grant’s method was used to determine the lower boundary highlighted in the circle. The origin was set to the vertical coordinate of the intersection of the dotted straight lines running respectively through the first three and 4-6th experimental chemical shifts of the titration. (B) Determination of the upper boundary based on the quality of linearization. (Left inset) Poor alignment of the X/Y plot caused by a completely wrong value of δB-H (extremum of the series of points available). (Right inset) Best alignment obtained after optimization of the value of δB-H. ΔpK_a(4, 3) = −0.5085 with 95 % confidence for +/- 0.0016.

If some of the experimental points are missing, the alignment method can be applied to determine the boundary value (see Figure 2.9-B). It appeared that the ΔpK_a values obtained by this method have the same precision as the ones measured with the intersect method.

2.3.3 Determination of pK_a in D_2O

The pK_a values were calculated in two steps detailed in the Experimental section. In short, the first step consists in using the ΔpK_a determined with the best precision (Figure 2.10.) to calculate an “initial” set of pK_a values. The second step consists of a refinement procedure taking into account all the available data.
Figure 2.10. (A-E) Linearization plots of the pairs of acids with the highest precision in the determination of the ΔpK_a.

The values obtained using the data of Figure 2.10 were used to position each acid on a pK_a scale. In principle one should only work with relative pK_a (see the “initial” values in Table 2.1.5, and the scale above the axis in Figure 2.11.) but absolute values assuming the pK_a of acetic acid to be 4.75 can be used if one has confidence that the reference compound really has this pK_a in the conditions of the titration. The errors given in Table 2.1.5 only reflect the precision obtained using the linearization process. As expected small differences in pK_a are observed.

Figure 2.11. Graphical representation of the determination of the pK_a of each acid using the minimal set of ΔpK_a.

Note that in order to be rigorous, one should also correct the pK_a to take into account the difference between H_2O and D_2O. We expect these corrections to be negligible because the use of an internal reference cancels out the zero order corrections and those of first order are much smaller.
Table 2.1.5. \( \Delta pK_a \) and \( pK_a \) of acids 1-6 in \( D_2O \)

<table>
<thead>
<tr>
<th>Acid</th>
<th>Initial values</th>
<th>Refined values(^b)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( \Delta pK_a )</td>
<td>Error(^a)</td>
</tr>
<tr>
<td>1 Acetic</td>
<td>0(ref.)</td>
<td>0.001</td>
</tr>
<tr>
<td>2 Camphanic</td>
<td>-2.171</td>
<td>0.012</td>
</tr>
<tr>
<td>3 2,3-Dimethoxybenzoic</td>
<td>-1.011</td>
<td>0.004</td>
</tr>
<tr>
<td>4 Benzoic</td>
<td>-0.502</td>
<td>0.003</td>
</tr>
<tr>
<td>5 Caffeic</td>
<td>-0.039</td>
<td>0.002</td>
</tr>
<tr>
<td>6 Coumaric</td>
<td>-0.027</td>
<td>0.001</td>
</tr>
</tbody>
</table>

\(^a\) For 95% confidence interval. When determined in multiple steps, errors were summed.

\(^b\) After integrating all the available data the \( \Delta pK_a \) were refined.

2.3.4 Determination of \( pK_a \) in \( D_2O/DMSO \) mixtures

The \( pK_a \) of acid [1-4, 6-8] were also calculated in the \( D_2O/DMSO \) mixtures in the different proportions 10%, 20%. We observed an increase in the \( pK_a \) value as the portions of DMSO increases. This was expected, as the \( pK_a \) in DMSO are generally higher than water. It is observed that propionic acid is weakest acid and camphanic acid is strongest acid in 10%, 20% DMSO/\( D_2O \) mixture. Table 2.1.6 and Figure 2.11. provide the relative and absolute \( pK_a \) values.

Table 2.1.6. \( \Delta pK_a \) and \( pK_a \) of acids [1-8] in DMSO/\( D_2O \) mixtures

<table>
<thead>
<tr>
<th>Acid</th>
<th>20%(^a)</th>
<th>10%(^a)</th>
<th>0%</th>
<th>Litt. (^{47})</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Acetic</td>
<td>5.15(^b)</td>
<td>4.93(^b)</td>
<td>4.75(^b)</td>
<td>4.75</td>
</tr>
<tr>
<td>2 Camphanic</td>
<td>3.15</td>
<td>2.80</td>
<td>2.58</td>
<td>-</td>
</tr>
<tr>
<td>3 2,3-Dimethoxy benzoic</td>
<td>4.23</td>
<td>3.87</td>
<td>3.70</td>
<td>3.98</td>
</tr>
<tr>
<td>4 Benzoic</td>
<td>4.64</td>
<td>4.38</td>
<td>4.25</td>
<td>4.204</td>
</tr>
<tr>
<td>5 Caffeic</td>
<td>-</td>
<td>-</td>
<td>4.79</td>
<td>4.43</td>
</tr>
<tr>
<td>6 Coumaric</td>
<td>5.08</td>
<td>4.82</td>
<td>4.72</td>
<td>4.36</td>
</tr>
<tr>
<td>7 4-Hydroxy benzoic</td>
<td>5.04</td>
<td>4.77</td>
<td>-</td>
<td>4.57</td>
</tr>
<tr>
<td>8 Propionic</td>
<td>5.52</td>
<td>5.08</td>
<td>-</td>
<td>4.875</td>
</tr>
</tbody>
</table>

\(^a\) % (vol.) corresponds to 2.78 % (mol) and 6.04 % (mol) respectively.

\(^b\) Using linear interpolation based on \( \Delta pK_a = 6.6441^{38} \) an \( H_2O \) value of 4.75.
2.3.5 Experimental details

2.3.5.1 Materials

Acetic acid (99%), propionic acid (99%), 4-hydroxy benzoic acid (99%), 2, 3 dimethoxy benzoic acid (99%) were obtained from Aldrich and (-)-camphanic acid (98%), p-coumaric acid (98%) from Fluka. Benzoic acid (99%) and caffeic acid (99%) were obtained from Acros Organic and Sigma respectively. All products were used as received. For the titrations, NaOH (98.5%, Acros organic) tablets were dissolved in D$_2$O while DCl was obtained as 38% D$_2$O solution from Fluka. All NMR samples were prepared in D$_2$O (D-99.8 %, Armar) and mixture with dimethyl sulfoxide-D6 (D-99.9 %, CIL).

2.3.5.2 Sample preparation

The D$_2$O sample was prepared using 150 µl stock solutions of basic solutions of 1 at c.a. 0.50 M and 2-6 at c.a. 0.25 M. The total volume was therefore 900 µl. The basifications of the solutions were done separately by small additions of 0.5 M NaOD and followed on pH paper until pH > 9. The titrations consisted in 10 µl additions of a 1.56 M solution of DCl prepared using 642 µl of concentrated DCL (38%) completed to 5 ml with D$_2$O. The concentration of 1 was used as 0.5 M because of its low signal.
intensity possibly due to a solubility problem.

The two D$_2$O/DMSO titrations (10 and 20% DMSO) were run at constant concentration. Each compound was present at 0.007 M both in the starting solution and the titrant. Stock 5 ml solutions containing 1-4, 6-8 0.05 M were prepared. The 0.5 ml starting solution was prepared using 70 µl stock solutions of 1-4, 6-8 and completed with solvent mixtures and a 10 µl of 0.816 M DCl. The titrant was prepared in a similar fashion except that it was completed with 26.13 mg of solid NaOH. We did not use any calibration compound, but an external chemical shift reference to align spectra is advisable.

2.3.5.3 The titration device

The push syringe (KDS310 plus) used for additions was obtained from KD scientific and the additions were controlled by an I/O device (USB-6501, National Instruments) with the help of 500 µl syringe from Hamilton (81220-1750 TL). The peristaltic pump (77120-62, Masterflex, 10–60 rpm) controls the addition and mixing. For additions the Teflon line of 0.3 mm diameter (HAM-86510, Hamilton) was used. The tubing used for mixing (EW-07632-26, 0.89 mm diameter) come from Cole Parmer. The additions were done in the open NMR tube (Wilmad, 5 mm). The actions of the push-syringe (used for the additions) and the peristaltic pump (insuring the mixing of the solutions) were controlled by the I/O device when the presence of files characteristic of the termination of NMR experiments was identified. The push syringe (set on auto-start mode) was controlled by switching on and off the power supply with an optotriac. The amount of product injected was determined by counting the changes of sign of the voltage of one of the control lines of the step motor via an optocoupler. The peristaltic pump was modified to control the start/stop button and sense of rotation with relays. An optical sensor detecting the passage of a flag fixed on the axis of the motor was used to count the number of turns of the pump. The additions were made trough a Teflon line at 10 ml.min$^{-1}$ into an elongated Pasteur pipette reaching the bottom of the open NMR tube and fixed to the spinner with a plastic holder (see Figure 2.4.b). The top of the pipette was sealed with a teflon cap with two holes. The tubing used for the mixing was connected to the peristaltic pump and open at the other end. Nitrogen or other gas source could be used, would oxygen be a concern. The mixing was insured by tree in/out cycles bringing the solution to
the pipette, sending it back to the tube and bubbling for a few seconds. A final short “in” flow insured that the pipette was filled with solution at the level of the NMR detection coil. Finally, a small and controlled leak in the mixing line allowed the level of the solution to equilibrate between the tube and the pipette during a two-minute stabilization time. An “au” program controlling that the width of the solvent signal in a 1D deuterium spectrum was not exceeding a value indicating the presence of bubbles was used for some experiments. When identifying homogeneity problems, it would run an additional in/out cycle.

2.3.5.4 Titration times 1D, 2D

The acquisition times of 1D and 2D HSQC experiments were about 1 minute and 44 minutes respectively. Complete titration were obtained overnight with the automatic titration setup. During the mixing, pseudo 2D experiments composed of a series of single-scan deuterium spectra were recorded. Displaying the stack of spectra gave the possibility to observe the evolution of the solvent signal during the addition and mixing operations.

All spectra were recorded on a Bruker 500 MHz spectrometer equipped with a ($^{13}$C/$^{31}$P/$^{19}$F) switchable detection probe (QNP) optimized for detection of X. The probe was tune/matched on the first spectrum but possible mismatch due to changes in salt concentration were not significant enough to result to observable reductions of signals amplitude in the following HSQC’s.

2.3.5.4.1 Acquisition parameters

For each point of the titration, a 1D proton and a 2D HSQC spectrum was recorded using the sensitivity-enhanced HSQC sequence (hsqcepgpsi) with an INEPT delay optimized for $J_{CH} = 145$ Hz. The carbon dimension was recorded with 512 increments on a spectral width of 2.5 ppm centered at 120 ppm resulting to a maximal $t_1 = 819$ ms. In the proton dimension 64 k points were recorded with a spectral width of 20 ppm. The duration of the pulses were 12.5 and 9.3 us for the proton and carbon channels respectively. For the D$_2$O solutions, the HSQC spectra were recorded with 2 scans per increment (44 min. per HSQC) while the two others used 4 scans (1 h. 11 min.).

2.3.5.4.2 Processing

The processing based on NMRpipe$^{48}$ involved a $\pi/2$ shifted sine bell in the
direct dimension. Zero filling was extended 8x in the indirect dimension to insure excellent resolution before the peak-picking realized by the NMRpipe "pkFindROI" command. Different matlab functions were used to process the data. First the chemical shifts obtained from the aliased HSQC were normalized. Different matlab scripts were used to read the bruker data (readBrukerPeaklist.m), peak picking (panel_peak_picking.m), plotting the chemical shifts with the graph (plotBrukerSpectrum.m).

2.3.5.5  Spectral analysis

2.3.5.5.1  $^1$H and $^{13}$C NMR data of carboxylic acids

The $^1$H and $^{13}$C chemical shifts values of all the carboxylic acids were as follows,

**Acetic acid** ($C_2H_4O_2$) $^1$H NMR (500 MHz, D$_2$O): $\delta = 1.84$ (s, 3H);

$^{13}$C NMR (125 MHz, D$_2$O) $\delta = 23.34$.

**Camphanic acid** ($C_{10}H_{14}O_4$) $^1$H NMR (500 MHz, D$_2$O): $\delta = 2.36$ (ddd, $J = 13.6$, $J = 10.7$, $J = 4.2$ Hz, 1H), $\delta = 2-20-2.05$ (m, 2H), $\delta = 1.57$ (ddd, $J = 13.4$, $J = 9.4$, $J = 4.2$ Hz, 1H), $\delta = 1.02$ (s, 3H), $\delta = 1.00$ (s, 3H), $\delta = 0.83$ (s, 3H);

$^{13}$C NMR (125 MHz, D$_2$O): $\delta = 184.04$, 174.20, 96.56, 55.52, 53.76, 30.10, 29.17, 16.56, 16.25, 8.80.

**2, 3 dimethoxy benzoic acid** ($C_9H_{10}O_4$) $^1$H NMR (500 MHz, D$_2$O): $\delta = 7.06$ (dd, $J = 7.5$ Hz, $J = 8.2$ Hz, 1H), $\delta = 7.01$ (dd, $J = 8.2$ Hz, $J = 1.5$ Hz, 1H), $\delta = 6.90$ (dd, $J = 7.5$ Hz, $J = 1.6$ Hz, 1H), $\delta = 3.80$ (s, 3H), $\delta = 3.74$ (s, 3H);

$^{13}$C NMR (125 MHz, D$_2$O): $\delta = 176.12$, 151.90, 144.52, 134.43, 134.82, 119.69, 113.26, 61.53, 56.00.

**Benzoic acid** ($C_7H_6O_2$) $^1$H NMR (500 MHz, D$_2$O): $\delta = 7.80$ (dt, $J = 7.5$ Hz, $J = 1.5$ Hz, 2H), $\delta = 7.47$ (tt, $J = 7.4$ Hz, $J = 1.5$ Hz, 1H), $\delta = 7.40$ (tt, $J = 7.8$ Hz, $J = 1.6$ Hz, 2H);

$^{13}$C NMR (125 MHz, D$_2$O): $\delta = 175.88$, 136.35, 131.33, 128.86, 128.40.

**Caffeic acid** ($C_9H_8O_4$) $^1$H NMR (500 MHz, D$_2$O): $\delta = 7.17$ (d, $J = 15.9$ Hz, 1H), $\delta = 7.00$ (d, $J = 2.0$ Hz, 1H), $\delta = 6.90$ (dd, $J = 8.2$ Hz, $J = 1.9$ Hz, 1H), $\delta = 6.75$ (d, $J = 8.2$ Hz, 1H).
Hz, 1H), \( \delta = 6.20 \) (d, \( J = 16 \text{ Hz}, 1\text{H} \));

\(^{13}\text{C} \) NMR (125 MHz, D\(_2\)O): \( \delta = 176.32, 147.93, 145.33, 141.16, 127.11, 121.53, 121.07, 116.38, 114.16 \).

**Coumaric acid** (C\(_9\)H\(_8\)O\(_3\)) \(^{1}\text{H} \) NMR (500 MHz, D\(_2\)O): \( \delta = 7.60 \) (d, \( J = 16.0 \text{ Hz}, 1\text{H} \), \( \delta = 7.53 \) (d, \( J = 8.7 \text{ Hz}, 2\text{H} \), \( \delta = 6.90 \) (d, \( J = 8.7 \text{ Hz}, 2\text{H} \), \( \delta = 6.33 \) (d, \( J = 15.9 \text{ Hz}, 1\text{H} \));

\(^{13}\text{C} \) NMR (125 MHz, D\(_2\)O): \( \delta = 172.55, 161.14, 148.30, 132.94, 132.92, 128.30, 118.44, 116.96, 116.94 \).

**4-Hydroxy benzoic acid** (C\(_7\)H\(_6\)O\(_3\)) \(^{1}\text{H} \) NMR (500 MHz, D\(_2\)O): \( \delta = 7.83 \) (d, \( J = 8.9 \text{ Hz}, 2\text{H} \), \( \delta = 6.86 \) (d, \( J = 8.9 \text{ Hz}, 2\text{H} \));

\(^{13}\text{C} \) NMR (125 MHz, D\(_2\)O): \( \delta = 170.00, 161.04, 132.45, 121.50, 115.60 \).

**Propionic acid** (C\(_3\)H\(_6\)O\(_2\)) \(^{1}\text{H} \) NMR (500 MHz, D\(_2\)O): \( \delta = 2.30 \) (q, \( J = 10 \text{ Hz}, 2\text{H} \), \( \delta = 1.00 \) (t, \( J = 10 \text{ Hz}, 3\text{H} \));

\(^{13}\text{C} \) NMR (125 MHz, D\(_2\)O): \( \delta = 179.70, 22.16, 8.32 \).

**2.3.5.5.2 Pick-picking /matlab**

The \(^{13}\text{C} \) chemical shifts from the aliased HSQC were peak picked from the Matlab based program. After selecting the first signal in the first spectrum the program automatically searches the next signals in the series of the spectra. In case of missing points one can manually select the peaks in the series.

**2.3.5.5.3 Normalization**

The second Matlab application facilitates the normalization and determines the boundary value of \( \delta^{\text{B}} \) and \( \delta^{\text{B-}} \). First the chemical shifts were corrected because they jump across the window boundary due to aliasing. In the second step the user can select a method for the determination of \( \delta^{\text{B}} \) and \( \delta^{\text{B-}} \). For the normalization all the chemical shifts in the protonated and deprotonated forms were used. We used the intersect method most of the times. But in case of caffeic acid and coumaric acid it is difficult to apply the intersect method because of initial chemical shifts values. In this case the protonation of the phenol starts early before starting the protonation of the carboxylic acid. (see the first 3 points shown in Figure 2.9-A). The alignment method (Gran’s method) has been applied for the missing boundary points. (see Figure 2.9-B).
The ΔpKₐ values obtained with this method were as precise as the values from the intersect method.

2.3.5.4 Combination of the ΔpKₐ

Finally the third Matlab program combines and displays the best pairs of normalized data. There were two main steps applied to determine the ΔpKₐ scale. The first step takes as input the tables of ΔpKₐᵢⱼ and error Eᵢⱼ, and a list associating the elements of the array to individual components of the mixture. A recursive procedure takes, as a starting point, the compound used as a reference and finds in the table the ΔpKₐ associated with the smallest value of Eᵢⱼ where i is any element corresponding to the reference compound and j any element of the other compounds. This ΔpKₐ positions the compound determined with the highest precision on the relative scale. The procedure is repeated recursively, except that the set of values of i is extended to include the data of all compounds positioned on the scale while the set of values of j corresponds to the rest of the data. This procedure stops after the last compound has been positioned on the scale. Such an “initial” scale takes into account only the minimal set of the best data available but ignore the rest, which is statistically not satisfactory and justifies a refinement procedure.

The second step consists in a refinement of the scales of relative pKₐ determined with the minimal set of ΔpKₐ so that all the available data are taken into account. We applied a force-field optimization similar to the ones used in chemical structure optimization but other statistical treatment may be more appropriate.

For each element of the tables of ΔpKₐᵢⱼ a gaussian probability density function

\[ f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \]  

(2.15)

is associated where \( \mu = (\Delta pK_{a i j} - \Delta pK_{a 0 i j} ) \), pKaₐ₀ᵢⱼ is the current value of ΔpKₐ for the account for 95% of the gaussian distribution). In other words, the error Eᵢⱼ which was calculated for a 95% confidence interval is converted into a gaussian distribution with the appropriate variance. The surface between the y-axis and the middle of the gaussian
\[ S = \int_{0}^{\mu} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}} \, dx \]  
\hspace{1cm} \text{(2.16)}

is used to determine the force

\[ F_{ij} = \frac{1}{E_{ij}} \int_{0}^{b} e^{-t^2} \, dt, \]  
\hspace{1cm} \text{(2.17)}

where \( b = (\Delta pK_{aij} - \Delta pK_{ai0})/\sqrt{2E_{ij}} \). The sum of all \(|F_{ij}|\) is minimized by adjustment of \( pK_{ai0} \) of the individual components.

This method gives comparable importance to the \( \Delta pK_a \) obtained with the best precision (e.g. when they are based on pairs of CH’s with large chemical shifts changes) and almost completely ignores the data with large errors making it unnecessary to select the data to integrate. This is an important property of such a calculation because the amount of data may be quite large and a criterion for manual or automatic selection may not be easy to determine.

2.3.6 Relation between \( \Delta pK_a \), errors and \( pK_a \) scale

For any pair of normalized chemical shifts, the \( \Delta pK_{aij} \) is determined by adjusting the variable \( a \) of the function \( y = ax \), until it fits the best the data point of coordinates derived from equation (2.14):

\[ x = (\delta_j^{\text{norm}})(1 - \delta_i^{\text{norm}}), \]  
\[ y = (\delta_i^{\text{norm}})(1 - \delta_j^{\text{norm}}). \]  
\hspace{1cm} \text{(2.18)}

Only the points for which \( \delta^{\text{norm}} \) was in the range 0.01:0.99 were taken into consideration (red circles in Figure 2.10.) because the others do not contribute to the line and cause numerical instabilities. In order to give more importance to points determined with the best precisions, that is the ones that are the farthest away relative to the origin (see symbols in Figure 2.10-B), they were associated with a weight

\[ w_{ij} = \left[ (\delta_j^{\text{norm}}\delta_i^{\text{norm}})^{-1} + (\delta_j^{\text{norm}}\delta_i^{\text{norm}})^{-1} + (\delta_i^{\text{norm}}\delta_j^{\text{norm}})^{-1} + (\delta_i^{\text{norm}}\delta_j^{\text{norm}})^{-1} \right]^{-1}. \]  
\hspace{1cm} \text{(2.19)}
Calculated using the expression of the propagation of standard deviation upon multiplication and division of variables. The fit procedure (polyfit function of Matlab) returns the value of $a$ and the 95% confidence interval $a_{\text{min}}$-$a_{\text{max}}$, $\Delta pK_a = -\log_{10}(a)$ and the error (shown in Figure 2.10 A-E) were set to $(\log_{10}(a_{\text{max}}) - \log_{10}(a_{\text{min}}))/2.$
3 Production of isotopically-enriched cholesterol

3.1 Isotope labeling in NMR

The structure determination of molecules by Nuclear Magnetic Resonance has long been limited by spectral complexity and low sensitivity. But the development in the hardware and methodology contributed to solve this problem to some extend. Over the past decades the developments of innovative methods for the isotope labeling of samples were also addressing this problem in good ways. The main aim of the isotope labeling is to replace inactive or detrimental isotopes with NMR-active isotopes with spin 1/2. This increases the sensitivity of the signals and makes it possible to use additional dimensions to simplify complex spectra and, in some cases, narrow the line width.

3.1.1 Introduction

In the early 1960’s, the use of isotope labeling in NMR experiment brought a big change in the nature of many biochemical problems. The assignment of carbon-13 resonances in 1D spectra, particularly of fungal metabolites and the availability of more sensitive high-field spectrometers brought this spectroscopic method into the limelight. During the 1970’s and 80’s many experiments were carried out by Tanabe, Simpson, Staunton, Cane, Vederas, to exploit this methodology in polyketide and terpenoid biosynthesis. For this, deuterium and carbon-13 labeled substrate were used. In the twentieth century the study carried out by Baldwin on penicillin and Scott and Batters on vitamin B_{12} required the synthesis of a wide range of stereo-specifically labeled substrates in order to understand their biosynthesis mechanisms. The application of isotopic labeling to the identification of specific NMR signal and their variation in substrate enzyme interaction can afford insights into enzyme catalysis. The various non-invasive techniques developed for medical imaging also had a significant impact on the development of organic chemistry synthesis for labeling compounds. Positron Emission Tomography (PET) employs short-lived radioactive tracers decaying with the emission of positrons. This technique was used to map the metabolic activity of cells within the body. These applications first time developed in 1970’s require synthesis of labeled substrates. In 1960 the detection of site of label by $^{13}$C NMR had a significant impact on the biosynthetic
The far better resolution and sensitivity in NMR was achieved by the introduction of Fourier-Transform techniques\textsuperscript{53} (FT-NMR). The introduction of superconducting magnets increased the sensitivity and resolution of signals, which is helpful for resolving the structures of peptides and other biological molecules. The labeling of particular carbon atom in a molecule with $^{13}$C is sometimes useful for signal assignment because the $^{13}$C-enriched positions produce correlation signals. Due to the enhancement of the signal with few milligram samples, many biological pathways have been solved. At natural abundance it is usually not possible to measure the $^{13}$C-$^{13}$C coupling because of the low probability to encounter pairs of coupled $^{13}$C in natural-abundance compounds.

The other commonly used isotope for labeling is nitrogen. The $^{14}$N has a natural abundance of 99.6\%, but the quadrupolar interactions of $^{14}$N makes the signals very broad and cannot be used to relay magnetization. That is why efforts were made to replace it with $^{15}$N which has a spin $\frac{1}{2}$. Thus with the use of isotope labeling it is possible to address the challenges in the structural biology.

3.1.2 Different organisms used for enrichment

3.1.2.1 Photosynthetic bacteria

Photosynthetic bacteria were used to produce many labeled proteins such as LH2 (Light harvesting proteins).\textsuperscript{54} These organisms can grow on the media containing $^{13}$CO$_2$ or K$^{15}$NO$_3$, $^{15}$NH$_4$Cl and this method is very economical. The Ferredoxin\textsuperscript{55} and plastocyanin obtained from cyanobacterium \textit{Anabaena variabilis} were the first two examples of $^{13}$C enrichment. Photosynthetic bacteria were also used to prepare the $^2$H-labeled proteins containing different labeling of the amino acids. The proteins enriched by using this method has been limited to those produced naturally and in high quantities by the photosynthetic bacteria such as photosynthetic electron transport proteins. This method is normally not applicable to other proteins.

3.1.2.2 Bacteria

Bacteria are the most commonly used organisms for preparing selectively and uniformly labeled proteins. Bacteria can also incorporate the isotope with the high
efficiency (~98%). Bacterial cells can grow on the minimal essential medium that contain $^{13}$C-glucose, $^{15}$NH$_4$Cl, and D$_2$O as carbon, nitrogen, and deuterium sources. E.coli is the most commonly used Gram-negative bacteria for protein enrichment. Growth media for NMR labeling in E.coli are simple and easy to prepare and, depending on the type of labeling, it is relatively cheap. The amino acid pathways of bacteria are well understood which allows the development of different strategies. E.coli can easily take up and incorporate the protein exogenously to supply isotopically labeled compounds. The yield of protein per liter of medium is often in tens of milligram, which allows to produce enough material for NMR studies.

### 3.1.2.3 Fungi

Yeast is a single-celled eukaryotic fungal organism. They are an inexpensive, yet efficient alternative to prokaryotes. Simple handling, rapid growth and powerful genetic tools contributed to their popularity. One of the major advantages is that yeast cultures can be grown to very high densities, which makes them especially useful for the production of isotope labeled protein for NMR. Yeast are often used to produce recombinant proteins. Pichia pastoris yeast can be used to express large quantities of isotopically labeled proteins such as thrombomodulin (TM) by fermentation. Apart from the well-known advantages for protein expression, such as eukaryotic proteolytic processing, protein folding, disulfide bond formation and glycosylation, another factor contributing to Pichia pastoris outstanding popularity is its ability to secrete heterologous proteins while secreting very low levels of endogenous proteins. The first report of application of NMR to proteins produced using these microorganisms was published by Eakin on yeast Candida utilis. MacKenzie et al. efficiently produced the HEWL (Hen Egg-White Lysozyme) by using $^{13}$C and $^{15}$N labeled Isotope in Aspergillus niger.

### 3.1.2.4 Mammalian cells

Many proteins can be functionally produced in prokaryotic systems such as E.coli. But most of the eukaryotic membrane proteins, including most drug targets cannot be produced in such systems in sufficiently functional quality and purity. Mammalian cell cultures are particularly useful because the proteins are often made in properly folded and glycosylated forms, thus eliminating the need to renature them. That is why it is important to develop the culture media for mammalian cells, which
allows cost effective labeling with stable isotope for the production of labeled glycoproteins\textsuperscript{60} and other recombinant proteins such as urokinase\textsuperscript{61} produced by CHO cells (Chinese Hamster Ovary).

The first reports of stable-isotope labeling in mammalian cells described the use of stable-isotope cell culture media consisting of amino acids isolated from \textit{E. coli} and extract of lyophilized algae and supplemented with cysteine and glutamine.\textsuperscript{61-62}

The advantage to use mammalian cell lines is the ability to obtain correctly processed mammalian proteins when that is the only one method available to express the protein. But it has some great disadvantage as it takes longer time to establish the stable cell lines. The target gene responsible for particular proteins need to be transfected into cell line and the transfected cells need to be selected for recombinant protein. Mammalian cells also require vitamins, growth factors, amino acids, and in many cases serum as a supplementary growth factor.

3.1.2.5 Plants

Plants can be considered as better than microbes in terms of cost, protein complexity, storage and distribution. Labeling of the plants was mainly done by supplying the labeled nutrients to the plant. Labeling the whole plant\textsuperscript{63} with \textsuperscript{15}NO$_3$ and \textsuperscript{15}NH$_4$ is the efficient way. For the \textsuperscript{15}N labeling the solution of K\textsuperscript{15}NO$_3$ is used. The \textsuperscript{15}N-labeled seedlings can be obtained from the \textsuperscript{15}N-labeled plants. The \textsuperscript{13}C$_6$-glucose is the usual source of the \textsuperscript{13}C labeling in plants. The biosynthesis of gallic acid\textsuperscript{64} can be studied by using \textsuperscript{13}C$_6$-labeled glucose supplied to the \textit{Rhus typhina} plants supplied through the leaf stem.

3.2 Cholesterol

Cholesterol is a very important lipid present in the membranes of vertebrate animals. The high level of cholesterol in blood and its involvement in cardiovascular diseases makes it medically important. It plays an important role in the structure of membranes and is the precursor of steroid hormones and bile acids.
Cholesterol has been the object of very intense study during the 20th century. In 1926 the Heinrich Wieland received the Nobel Prize in chemistry for the structure of cholesterol and bile acids. Konrad Bloch and David Rittenberg showed that the ring structure and side chain of cholesterol were derived from acetate, and they identified the intermediates in the pathway. Further work by Bloch, John Cornforth and George Popjak succeeded in establishing the biosynthetic origin of all 27 carbons of cholesterol. For his elegant work, Bloch was awarded the Nobel Prize in chemistry in 1964.\textsuperscript{65}

3.2.1 Biochemical pathways of cholesterol

Sterols are synthesized from the two-carbon building block, acetyl-CoA. In most cases, the acetyl-CoA entering the sterol synthesis is originating from the glycolysis, the process where glucose is converted into two molecules of pyruvate and further into acetyl-CoA. The two main steps are illustrated in Figure 3.2.

The biosynthesis of cholesterol is illustrated in Figure 3.3. First, the enzyme acetoactyl-CoA thiolase interconvert’s acetyl CoA and acetoactyl-CoA. The latter is then condensed by 3-hydroxy-3-methylglutaryl (HMG)-CoA synthase to form HMG-CoA. HMG-CoA reductase catalyzes the reduction of HMG-CoA to mevalonate. Mevalonate is phosphorylated by two sequential phosphate transfers from ATP,
yielding the pyrophosphate derivative. Pyrophosphomevolonate Decarboxylase catalyzes ATP-dependant decarboxylation with dehydration to yield isopentenyl pyrophosphate. Isopentenyl Pyrophosphate Isomerase interconvert isopentenyl pyrophosphate and dimethylallyl pyrophosphate. Prenyl Transferase catalyzes a series of head and tail reactions. First, dimethylallyl pyrophosphate reacts with isopentenyl pyrophosphate to form geranyl pyrophosphate. Subsequently the condensation with another isopentyl pyrophosphate forms farnesyl pyrophosphate. Squalene synthase is the 47-kDa protein of endoplasmic reticulum which condenses two molecules of farnesyl pyrophosphate (head-to-head condensation) to yield squalene. Squalene epoxidase catalyzes the oxidation of squalene to form 2,3-oxidosequalene. This requires NADPH as reductant and O₂ as oxidant. One atom of oxygen is incorporated into the substrate and the other oxygen atom is reduced to water.

Squalene oxidocyclase then catalyzes the cyclization of this structure by protonating the epoxide and leaving an electron deficient carbon. Migration of the resulting electron deficiency sequentially cyclizes the rings, leaving a carbocation on the Protosterol cation. Then this carbocation drives a series of hydride and methyl migrations, finally resulting in Lanosterol. Lanosterol is finally converted to cholesterol in a 19-step process in which three methyl groups are removed by sequential oxidations.
3.2.2 Related sterols

Sterols are essential components of eukaryotic cell membranes. The yeast *Saccharomyces cerevisiae* accumulates ergosterol, the sterol equivalent of cholesterol, which is found in animals and sitosterol is found in plants. But the group of Prof. Howard Riezman genetically modified yeast strains to make them synthesize...
cholesterol instead of ergosterol.

![Figure 3.4. Structures of ergosterol and sitosterol](image)

3.2.3 Production of isotopically-enriched cholesterol

Cholesterol has been also produced by using mammal cells. But only small amounts of low-enrichment cholesterol were obtained in vivo by skin injection of enriched mevalonate in rat or feeding mammals and humans with $^{13}$C enriched precursors. But much a more effective method consist in using the engineered *Saccharomyces cerevisiae* strain where gene deletions and heterologous genes on plasmids makes it capable of synthesizing cholesterol as major sterol.

3.2.4 $^{13}$C enrichment for NMR experiments

3.2.4.1 High and uniform enrichment

The uniform enrichment allows all carbons of cholesterol to be replaced at the same time. In principle, the mass of cholesterol can therefore increase from 386 to 413 when all carbons are $^{13}$C. The comparison of experimental and simulated GCMS spectra (Figure 3.5.) indicate that ($u^{13}$C$_6$, 99%) is only slightly diluted by the presence of natural-abundance (N.A.) yeast extract at about 1:1 equivalent in mass relative to glucose.
Only a small amount (c.a. 2%) of cholesterol is unlabeled (see the peaks at 386/387) and probably originates from cholesterol produced before the main cell culture started. The distribution of high masses (400-413) indicate that 6% of the carbons present in cholesterol are $^{12}$C, probably originating from the atoms present in the non-enriched yeast extract. The most abundant isotopomer is nevertheless the one corresponding to the fully $^{13}$C-enriched cholesterol with a mass of 413.

The 1D carbon spectrum of 92%-enriched cholesterol shows (Figure 3.6.) the multiplicity expected for such enrichment level. Carbons come out as doublet, triplet, quartet, just like protons in proton AX, AX2, AX3 spin systems, except that the coupling constants $^{1}J_{CC} \approx 35$ Hz. For example C(18), which is methyl, comes out as a doublet because it is coupling with the C(13) (AX). Note the presence of a minor singlet signal between the two main lines of the doublet corresponding to the 6% abundance of C(18) bound to C(13) with a $^{12}$C isotope. The carbon 11 shows an AMX triplet structure due to coupling with C(12) and C(9) while the C(9) has the quartet-like structure as it is coupling with carbons 8, 10 and 11 (AKMX system).
Figure 3.6. (a) 1D $^1$H-decoupled $^{13}$C NMR spectra of 92% $u$-labeled cholesterol with expansion of quartet (9), triplet (11), doublet (18) protons and (b) Natural abundance cholesterol.

The dominant structure of C(5) (enlarged in Figure 3.7.) shows a doublet of triplet ($dt$) structure because of the coupling with carbons 4, 6 and 10. Note that the values of the carbon-carbon coupling differ from the usual 35 Hz observed when both carbons have sp3 hybridations. The C(sp2)-C(sp2) coupling $J_{5,6}$ is 71.25 Hz while the $J_{5,4}$ and $J_{5,10}$ are both equal to 38.75 Hz. These values are in good agreement with the values calculated using DFT/GIAO techniques (see the experimental section).
Figure 3.7. Enlargements of the multiplet of carbon 5 coupling with carbons 4, 6 and 10.

We also observed the minor coupling structures of the carbons bound to two $^{13}$C and one $^{12}$C. The four minor lines with a doublet of doublet structure ($dd, J = 78$ Hz, $J = 39$ Hz) observed between the lines of the main structure are due to the coupling of 5 with 6 and 10 which overlaps with the minor doublet due to the coupling with 6 and 4. The minor signal expected to have a triplet-like structure due to nearly equal couplings of 5 with 10 and 4 ($t, J = 39$ Hz) is not directly observed because the lines are hidden by the dominant multiplet.

In some cases, second order effects are observed because of a ABKMX coupling system where A and B are the nearly degenerated carbons 7 and 8, and the three additional partners being carbons 6, 9 and 14. In principle, C(6) should be a simple doublet of doublet because of its coupling with 5, but minor lines appear between the main lines. These lines, that are too intense to be due to the minor $^{12}$C multiplets, are in fact lines appearing because of the virtual coupling with 8, an effect rarely observed in the X multiplet when second-order effects are very strong in ABX systems.
Figure 3.8. Sections of the spectrum where second-order effects can be observed because of nearly degenerated carbons 7 and 8.

Figure 3.9. HSQC of 92% labeled cholesterol.
In the HSQC (Figure 3.10.), a couple of signals are highlighted. They correspond to the correlation between the C(16) and its directly-bound protons. The carbon is a CH$_2$ this is why it has two different signals along the proton dimension. In the normal HSQC spectrum (Figure 3.9.) the resolution is not sufficient to observe the J$_{CC}$ coupling structures but become clearly visible in the 10-ppm spectrum of Figure 3.11.

![Figure 3.10. HSQC of 92% labeled cholesterol.](image)
The disadvantage of uniform enrichment are that $^{13}$C-$^{13}$C J-couplings complicate resonance line shapes, (see Figure 3.6.) which decreases the spectral resolution and sensitivity. In case of uniform enrichment, it is also difficult to measure the relaxation parameters or other parameters measured from series of HSQC because of the complex coupling patterns. There are two possibilities to simplify the coupling structures. The first is to apply the broadband homodecoupling (see Figure 3.12.) introduced by Foroozandeh et al. One problem with this homodecoupling technique is that the gradient encoding significantly decreases the sensitivity.

**Figure 3.11.** 10-ppm HSQC of 92% labeled cholesterol
The alternative is to reduce the labeling and find a compromise between sensitivity and multiplet complexity (see next section 3.2.4.2)

Finally we recorded a $^{13}$C J-resolved spectrum, an experiment usually applied to protons. The J-resolved spectroscopy is usually applied when the 1D spectra are crowded and where proton multiplets severely overlap. J-resolved spectroscopy can be used to separate the multiplets structures, which gives the way for the determination of spin-spin coupling constants and chemical shifts. In J-resolved spectra, the coupling appears in the $F_1$ domain and the chemical shift is confined to the $F_2$ domain. But here we used J-resolved spectroscopy to carbon 13 (Figure 3.13.).
Figure 3.14. Zoom into the signal of carbons 19 into the J-resolved spectrum. It shows the measurement of the $J_{CC} = 34.37$ Hz coupling constant. The small coupling observed in the dd of C21 is probably due to the $J_{21,16}$. 

Figure 3.13. $^{13}$C J-resolved spectrum of 92% labeled cholesterol
3.2.4.2 Uniform 10% enrichment

We produced labeled cholesterol with moderate enrichment to simplify signal multiplicities, an approach which is also cost effective because it facilitates less labeled glucose. For HSQC and related experiments (HSQC-NOESY, etc.) one usually wants to obtain singlet’s with maximal intensities and minimize additional doublet or triplet due to $^{13}$C coupling partners. For these experiments the maximal proportion of singlet (7-15%) can be obtained with 20-30% average enrichment using diluted ($^{13}$C$_6$, 25%) glucose. We produced 10% uniformly-enriched cholesterol by mixing ($^{13}$C$_6$, 24.0%) with 1.5 equivalents of N.A. glucose. On average 6.8% of the carbons come as singlet’s while the proportion of doublets due to $^{1}$J$_{CC}$ is 2.45%, which is not disturbing as these doublets spread their intensities over pairs of peaks.

The MS spectrum confirmed the level of enrichment since the distribution of masses matches to predictions.

![Figure 3.15](image)

**Figure 3.15.** Experimental blue and simulated (red, green) MS spectra of cholesterol produced 2:3 ($^{13}$C$_6$, 24%)/(N.A.) glucose. The proportions of natural abundance cholesterol (in green) were estimated based on the difference of 386 signals in experimental and the simulated spectra.

When compared to fully enriched cholesterol, the NMR spectrum of 10% average enrichment (Figures 3.16 -3.18) is simplified and similar to the one of N.A. cholesterol, except that it has about 10x the signal intensity at equal concentration. A second difference is the presence of weak doublets due to c.a. 1% probability of simultaneous presence of pairs of $^{13}$C.
Figure 3.16. (a) 1D $^{13}$C NMR of 10% labeled cholesterol (b) 1D $^{13}$C NMR of N.A. cholesterol.

Figure 3.17. HSQC of 10% labeled cholesterol.
Figure 3.18. 10-ppm HSQC of 10% labeled cholesterol.

The ADEQUATE\textsuperscript{71} is a $^{13}\text{C}-^{13}\text{C}$ double-quantum experiment facilitating the connectivity of the carbons of organic compounds. It directly shows the carbon-carbon coupling between the pairs of directly-bound carbons. The spectrum correlates the proton frequency on the $F_2$ axis with the double-quantum carbon frequency on the $F_1$ axis making the spectrum slightly more complicated to interpret than other 2D spectra. This experiment is not sensitive enough to be routinely applied because the probability of occurrence of pairs of directly-bound carbons is only about 0.01\% at natural abundance, but the 10\%-\textsuperscript{13}\text{C} enrichment (Figure 3.19.) increases the sensitivity by a factor 100.
3.2.4.3 Semi-selective enrichment using (1-$^{13}$C$_1$ glucose)

The alternation of carbons stemming from the methyl and carbonyl of acetyl-CoA in the precursor 5 (see Figure 3.20.) and the availability of (1-$^{13}$C$_1$, 99%) and (2-$^{13}$C$_1$, 99%) glucose makes it possible to significantly increase the proportion of singlet’s by selectively labeling either the filled or the empty carbons.

![Figure 3.19. Adequate spectrum of the 10% labeled cholesterol.](image)

**Figure 3.19.** Adequate spectrum of the 10% labeled cholesterol.

![Figure 3.20. Biosynthetic pathway of cholesterol produced by yeast grown on a glucose medium.](image)

**Figure 3.20.** Biosynthetic pathway of cholesterol produced by yeast grown on a glucose medium. Rounded frames indicate the 11 pairs of carbons stemming from glucose through acetyl CoA. Filled and empty dots correspond to the carbonyl and the Me of acetyl-CoA respectively.

With singly-labeled glucose, the maximum level of enrichment can only reach c.a. 50% because carbons 5 and 6 of glucose accounting for half of the Acetyl-CoA
are not enriched (see Figure 3.20.). We produced cholesterol with 9% average enrichment diluting (1\textsuperscript{-13}C, 99\%) with 1:2 natural-abundance glucose. The MS spectrum confirmed the predicted distribution of isotopomers.

**Figure 3.21.** Experimental blue and simulated (red, green) MS spectra of cholesterol produced with 1:2 (1\textsuperscript{-13}C, 99\%)/N.A. glucose. The enrichment levels were set according to the values certified by the producer. The simulations match the experimental data with the values calculated based on the quantity of added N.A. glucose. The proportions of N.A. cholesterol (in green) were estimated based on the difference of 386 signals in experimental and simulated spectra.

The NMR spectra show only about half the carbons. They appear as singlets because \( ^1J_{CC} \) partners are quite rare (only carbons 15, 16 and 18 are enriched and directly bound, see the empty circles in Figure 3.20.) and the long-range coupling are not too problematic since the dilution with N.A. glucose decreases their probabilities. Note the presence of minor doublet structures for carbons 13, 17, 18 because they have about 33\% probabilities to be directly bound to enriched positions.
Figure 3.22. (a) 1D $^{13}$C NMR of 16% labeled cholesterol (b) 1D $^{13}$C NMR of N.A. cholesterol.

Figure 3.23. 10-ppm HSQC of 16% labeled cholesterol.
3.3 Experimental part

3.3.1 Sample preparation

3.3.1.1 Materials

Yeast nitrogen base (YNB) and yeast extracts (YE) were obtained from US Biological and Difco respectively. D-glucose (99%), leucine (95%), uracil (99%) were obtained from Aristar, Fluka and Sigma respectively. The pyrogallol and petroleum ether with high-boiling points (b.p. 40-60) come from Sigma while potassium hydroxide (85%) and methanol were obtained from Acros. All enriched sources of glucose and chloroform-D (99.8%) were purchased from Cambridge Isotope Laboratory (CIL).

3.3.1.2 Media preparation

Three samples of cholesterol were prepared. The 94%-enriched cholesterol was prepared in a 450 ml medium containing 0.7% YNB, 0.5% YE and 1.5% (6.75 g) (u-13C6, 99%) glucose. It was mixed under magnetic stirring, and filter sterilized. A 4.5 ml solution of leucine and uracil was added from a 10x stock solution at 0.4 mg/l. The production medium was inoculated with 50µl (1:10000) yeast culture (Saccharomyces cervisiae strain RH6829; Souza et al.68) and incubated at 30°C in a 3-liter flask for two days with shaking at 230 rpm. We obtained approximately 5.5mg of cholesterol after lipid extraction (see below). The 10%-enriched cholesterol was prepared from 2.5 l medium where 9.90 g of (u-13C6, 24.0%) were diluted with 15.0 g of natural-abundance glucose. For growth the medium was split into three flasks. We obtained approximately 46 mg of cholesterol.

The partially labeled cholesterol was prepared in a 300 ml medium. 0.98 g of (1-13C1, 98.1%) were diluted with 2 g of natural-abundance glucose and resulted in 1.7 mg of cholesterol.

3.3.1.3 Lipid extraction

Cholesterol extraction and purification was performed as previously described (Souza et al.68). Briefly, the culture media were centrifuged at 3'000 rpm for 5 min and the cells harvested and washed with water in glass tubes. The cells were resuspended with 1 ml 60% KOH, 1 ml of 0.5% pyrogallol in methanol and 1.5 ml
methanol in screw-cap glass tubes and incubated for 2 hr at 85°C, tightly closed. Sterols were extracted three times with 2 ml fractions of petroleum ether. The petroleum ether phases were evaporated under nitrogen flow and the samples stored at -80°C.

The purification involved a C18 reversed-phase HPLC. The extracts were dissolved in acetone at 45°C and the soluble extracts were applied to an Uptisphere 120A 5 µm ODB column (Laubscher Labs, Switzerland) and eluted isocratically with 70% acetonitrile/30% ethanol (v/v). The cholesterol peak was identified by spotting 5 µl onto a TLC plate. The TLC plate was immersed in staining solution (50 mg FeCl$_3$:6H$_2$O, 90 ml H$_2$O, 5 ml glacial acetic acid, 5 ml sulfuric acid) for 1 min, then heated for 3 min at 110°C and scanned. The cholesterol peak fractions were pooled and dried under nitrogen.

3.3.1.4 MS analysis

Samples were injected into a VARIAN CP-3800 Gas Chromatograph equipped with a Factor Four Capillary Column VF-5ms 30 m x 0.25 mm ID DF = 0.25 and analysed by a Varian 320 MS triple quadruple with electron energy set to -70 Ev at 200°C. Sterols were eluted with a linear gradient from 195-230°C at 4°C per min.$^{72}$ The simulated spectra were calculated on a desktop PC using a brute-force program written by C language taking the enrichment level and dilution of glucose as input and determining the probability of each 2 isotopomers.$^{29}$

3.3.2 NMR acquisition parameters

All the NMR measurements were performed at 25°C on a Bruker Avance-III 500 MHz spectrometer equipped with a TCI low-temperature probe. The proton-decoupled $^{13}$C experiment were recorded in 512 scans with a 236 ppm window, 3 s recycling delay and 10 µs hard pulses. The 2D HSQC spectrum was recorded using the sensitivity-enhanced HSQC sequence (hsqcetgpsi) with an INEPT delay optimized for $^{1}$J$_{CH}$= 145 Hz. The carbon dimension was recorded with 512 increments on a spectral width of 200 ppm t$_1$= 110 ms. In the proton dimension 64 k points were recorded with a spectral width of 20 ppm. The duration of the pulses were 12.5 and 9.3 us for the proton and carbon channels respectively.
3.3.3 DFT-GIAO calculations

DFT-GIAO calculations of a model structure of cholesterol were used in order to compare experimental coupling constants with prediction. Because of the mobility of the side chain, only the 22 carbons of the main polycyclic part were used for these calculations.

Figure 3.24. (top) B3LYP/6-31G*-minimized structure of cholesterol and (bottom) B3LYP/6-311+G** GIAO shielding and carbon-carbon scalar coupling constants of a model structure of cholesterol.
4 Bibliography


54. (a) Sundström, V.; Pullerits, T. n.; van Grondelle, R., Photosynthetic Light-Harvesting: Reconciling Dynamics and Structure of Purple Bacterial LH2


