Kinetic Measurements Using Time-Resolved NMR Spectroscopy

A thesis submitted to The University of Manchester for the degree of Doctor of Philosophy in the Faculty of Engineering and Physical Sciences

2010

Maryam Khajeh

School of Chemistry

Contents

List of Figures

Chapter 3

Chapter 4

Chapter 5

List of Tables

Chapter 2

List of Abbreviations and Symbols

NMR: Nuclear Magnetic Resonance
CW: Continuous Wave
RF: Radiofrequency
I : Spin quantum number
m_I : Magnetic quantum number
B_0 : External magnetic field
B_1: Oscillating magnetic field
ΔE : Energy difference
γ : Gyromagnetic ratio
$\hbar$: Planck's constant divided by 2π
$\bar{\mu}$: Magnetic moment
$M_\circ$: Bulk magnetization
ω_0 : Larmor frequency
FID: Free Induction Decay
ω_{ref} : Reference frequency
ADCs: Analogue-to-Digital Convertors
TMS: Tetramethylsilane
TSP: 3-trimethylsilylpropionic acid sodium salt
J: Coupling constant
T_1: Spin-lattice relaxation
T_2: Spin-spin relaxation
FT^-: Fourier transformation
FT^+: Inverse Fourier transformation
lb: Lorentzian line width parameter
gf : Time domain Gaussian parameter
CTP: Coherence Transfer Pathway
LS: Least Squares
2D: Two-dimensional
PCA: Principal Component Analysis
DOSY: Diffusion-Ordered SpectroscopY
PARAFAC: PARAllel FACtor analysis
ALS: Alternating Least Squares
RMS: Root Mean Square
PFG: Pulsed Field Gradient
PFGSE: Pulsed Field Gradient Spin Echo
PFGSTE: Pulsed Field Gradient STimulated Echo
BPPSTE: Bipolar Pulse Pair Stimulated Echo
SNR (S/N): Signal-to-Noise Ratio
ID: Inside Diameter
OD: Outside Diameter

Abstract

Many thousands of pounds are spent every year by pharmaceutical companies on understanding the mechanisms and kinetics of chemical reactions involved in drug discovery and production. NMR spectroscopy is often at the core of these studies as it is a powerful, non-destructive method for structure elucidation. As such investigations can be time-consuming and cost-inefficient, AstraZeneca, the project sponsor, is interested in more efficient methods for studying the kinetics of pharmaceutical reactions. In this work a number of different techniques have been devised, studied, and implemented to study the kinetics of chemical reactions by time-resolved NMR spectroscopy, in which every species in a reaction can be monitored simultaneously. These novel techniques allow the study of reactions which are difficult or impossible to study by conventional NMR methods (such as heterogeneous reactions), or which are complicated by having overlapping signals.

It is possible to monitor the kinetics of a reaction very simply by acquiring a series of ^{1}H spectra, and obtaining the integrals of the signals by least squares fitting. This technique has been used for kinetic studies of static and on-flow reactions. In the static systems the reaction mixture was placed in the normal NMR tube in the magnet, while in the flow system the reaction mixture was placed outside of the magnet, and the solution flowed through an NMR tube placed in the magnet. The novel flow system designed, constructed and tested here has been used for kinetic studies of illustrative homogeneous and heterogeneous reactions, and is suitable for use in a wide range of NMR instrumentation.

Kinetic studies have also been carried out by acquiring a series of DOSY datasets, analysing the results using the multi-way method PARAFAC (PARAllel FACtor analysis). A series of DOSY datasets contains multivariate information on spectrum, time evolution and diffusion. Without providing any predetermined model, the data can be decomposed by PARAFAC to yield the spectrum, kinetics, and diffusion profiles for each of the components. It has also been shown that PARAFAC is remarkably robust to low signal-to-noise ratio data, significantly below the level at which conventional methods would fail.

Declaration

The author hereby declares that no portion of the work referred to in this thesis has been submitted of an application for another degree or qualification of this or any other university or other institute of learning.

Copyright Statement

i. The author of this thesis (including any appendices and/or schedules to this thesis) owns certain copyright or related rights in it (the "Copyright") and s/he has given The University of Manchester certain rights to use such Copyright, including for administrative purposes.

ii. Copies of this thesis, either in full or in extracts and whether in hard or electronic copy, may be made **only** in accordance with the Copyright, Designs and Patents Act 1988 (as amended) and regulations issued under it or, where appropriate, in accordance with licensing agreements which the University has from time to time. This page must from part of any such copies made.

iii. The ownership of certain Copyright, patents, designs, trade marks and other intellectual property (the "Intellectual property") and any reproductions of copyright works in the thesis, for example graphs and tables ("Reproductions"), which may be described in this thesis, may not be owned by the author and may be owned by third parties. Such Intellectual Property and Reproductions cannot and must not be made available for use without the prior written permission of owner(s) of the relevant Intellectual Property and/or Reproductions.

iv. Further information on the conditions under which disclosure, publication and commercialisation of this thesis, the Copyright and any Intellectual Property and/or Reproductions describe in it may take place is available in the University IP Policy (see http://www.campus.manchester.ac.uk/medialibrary/policies/intellectual-property.pdf), in any relevant Thesis restriction declarations deposited in the University Library, The University Library's regulations (see http://www.manchester.ac.uk/library/aboutus/regulations) and in The University's policy on presentation of Theses.

Acknowledgments

First and foremost, I would like to thank my supervisor, Professor Gareth A. Morris, who has been enormously helpful and always available during the last three and a half years. I have learnt a lot from Gareth and I take away with me, not only lessons in science, but also plenty of lessons in life. I am grateful for all of his advice, patience and support, and it has been an honour to work with him. This thesis would have not been completed without his supervision. Gareth has been more than a supervisor and I cannot thank him enough in words. I wish him and his lovely family all the happiness in the world.

I would like to thank my industrial supervisor, Dr. Michael A. Bernstein, for all of his support, and for being such a great host, making my time at AstraZeneca in Loughborough an enjoyable experience.

Thanks should also go to the School of Chemistry of The University of Manchester for the fantastic facilities and equipment, and to EPSRC and AstraZeneca for funding this project.

I would also like to thank everyone in the group and in the workshops for being so friendly, supportive, helpful, and making it a nice environment in which to work. I have enjoyed working with Mathias Nilsson, Adolfo Botana, Penny Hubbard, Andy McLachlan, Juan Aguilar, Rob Evans, Christopher Slann, Peter Wilde, Ian Goodbody, and Barbara Gore. At this stage I would like to take the opportunity to especially thank my good friend Adolfo. Adolfo has been extremely helpful, more than any colleague or friend could be. I count myself as being lucky to have met such a bright and loving person and wish him the best of luck for his future which, I am sure, will be bright.

I want to thank my loving family, particularly my parents for believing in me, and raising me in such a way so that I had the resilience to overcome all the difficulties and the hard times during the last few years; and my sisters for their support and encouragement.

It is my pleasure to thank my husband, Afshin, for remaining sympathetic, so understanding, helpful and supportive. Without his cooperation I would have not been able to complete this work.

And finally, last but not least, I would love to thank with all my heart a very special person. My little 6 year old son, Daniel, who has brightened my life and given me hope during the times I was struggling. He has coped so many times with the absence of his mum. Recently Daniel has been so understanding and has given his mum the time she needed to finish this thesis. So ‘Thank You’ my darling.

To My Family

Chapter 1

Introduction

1. Introduction

After the first detection of nuclear magnetic resonance (NMR) in 1945, this phenomenon became an extremely powerful technique in a short time. Bloch [1.1] and Purcell [1.2] in two different research groups discovered NMR almost simultaneously, and won the joint Nobel Prize at 1952 [1.3]. At first, the technique was based on continuous wave (CW) method, in which the radio frequency (RF) signal was slowly varied over a frequency range, to measure the absorption signals of all the different spin sites sequentially in the frequency domain. This method is quite time-consuming, as it has to sweep the whole frequency range slowly. Since 1966 [1.4] CW NMR has been largely replaced by Fourier transform (FT) NMR. In this method a radio frequency pulse is applied, which excites all spins in a given frequency range. FT NMR is much faster in comparison with CW, and the data produced have much better quality.

NMR spectroscopy has been widely used in different subjects such as physics, biology, chemistry, biochemistry, environmental control and clinical medicine. The comprehensive capability of NMR spectroscopy in clarifying molecular conformation and structure has turned the technique into an outstanding tool in academic and industrial laboratories. The great chemical specificity of NMR makes it useful in a wide variety of applications such as structure determination, kinetic and chemical studies of compounds or mixtures of compounds.

An understanding of the basic chemical and physical processes of a reaction, its mechanism and kinetics, allows industrial processes to be improved, optimizing product production and eliminating side-products. AstraZeneca, a pharmaceutical

company, is interested in novel methods to understand and study the kinetics of a variety of chemical reactions by methods that are more efficient and are less costly to manufacturer, environment, and society. This project was sponsored by AstraZeneca; the aim was to be able to characterize the kinetics of different species in a solution mixture, including intermediates, by implementation of novel methods in time-resolved NMR spectroscopy.

NMR is frequently used in the monitoring of chemical reactions. However, it is not always simple to interpret the data, especially when signals overlap. This common problem complicates analysis, sometimes to the point where the identification of individual components is not possible. Some NMR techniques such as DOSY and 2D experiments can be utilized to overcome this problem. It can also be possible to resolve this puzzle by obtaining 3-dimensional experimental data [1.5].

This thesis is divided into six chapters. This chapter and chapter 2 provide a basic introduction and a brief summary of the theory involved. Chapter 3 describes preliminary work on studying the kinetics of chemical reactions by time-resolved NMR spectroscopy, and contains 4 sections (introduction, theory, experimental work, and results and discussion). Chapters 4 and 5 each have their own sub-sections (introduction, theory, experimental work and data processing, results and discussion, conclusion (in chapter 4), future work (in chapter 5)), and describe different novel techniques.

Chapter 4 illustrates kinetic studies using trilinear data (obtained from time-resolved 2-dimensional experiments) as input to PARAFAC, a multi-way analysis method.

PARAFAC will decompose the data to yield the spectra, concentration and diffusional profile of each of the mixture components.

Chapter 5 shows the design and demonstration of a simple, cheap, robust flow system for monitoring homogeneous and heterogeneous reactions. This system can be employed with any conventional NMR probe and magnet, and no modification of the NMR instrument is needed.

Chapter 6 summarizes the conclusions drawn from the project.

Chapter 2

Theory

2. Theory

2.1 Basic Theory

The NMR phenomenon is based upon the behaviour of the nuclei of certain atomic isotopes in the presence of a magnetic field. In quantum mechanics the spin quantum number (I) can take zero, positive integer, and positive half-integer values. Any nucleus with a non-zero spin quantum number is magnetic. If a magnetic nucleus is placed in an external magnetic field the energy states of spins split into different energy levels. Each nucleus produces $2I+1$ energy levels, which are represented by magnetic quantum numbers m_I ($-I < m_I < I$). In an external magnetic field (B_0), a proton nucleus with I =1/2 splits into two energy levels containing low (α) and high (β) energy states. The energy difference (ΔE) between these two energy states depends upon the gyromagnetic ratio (γ) of the nucleus and the strength of the magnetic field (B_0):

$$\Delta E = \hbar \gamma B_0 \qquad \text{(equation 2.1)}$$

($\hbar$: Planck's constant divided by 2π)

The ratio of spin population between the energy states at equilibrium is described by the Boltzmann distribution (equation 2.2), in which the population of spins at the lower level is slightly higher than at the higher level.

$$\frac{N_\beta}{N_\alpha} = \exp(-\Delta E / k_B T) \qquad \text{(equation 2.2)}$$

where $k_B, N_\beta, N_\alpha, \Delta E, T$ are the Boltzmann constant, spin population at higher energy level, spin population at lower energy level, energy difference between the energy states and temperature respectively. Radiofrequency (RF) irradiation with a frequency corresponding to the energy difference between the energy levels can cause transitions between the spin states [2.1, 2].

Spins possess a small magnetic moment $(\overline{\mu})$ due to the positive charge of the nucleus and its motion. In the absence of an external magnetic field the spins are distributed evenly over all orientations. On applying a magnetic field (B_0), each spin starts a precessional motion around the field that is slightly influenced by the much weaker local magnetic fields generated by thermal motion of the molecules. This will eventually produce a net magnetization along the external magnetic field (z-axis). This is known as the equilibrium or bulk magnetization (M_0), which for a spin quantum number *I* has magnitude:

$$M_0 = \frac{N\gamma^2\hbar^2 I(I+1)B_0}{3k_B T} \qquad \text{(equation 2.3)}$$

where N is the number of spins.

At equilibrium only the z-component of magnetization exists, while x- and y-components of spins cancel one another. In order to rotate the magnetization away from z-axis, a radiofrequency (RF) pulse with frequency close to the Larmor

frequency is applied to the coil, which results in an oscillating current that causes an oscillating magnetic field, known as the B_1 field or the radiofrequency field [2.2]. This field causes the net magnetization to tilt away from the z-direction, and rotates around the z-axis (precesses) with a frequency known as the Larmor frequency (ω_0).

$$\nu_0 = -\frac{\gamma B_0}{2\pi} \quad \text{or} \quad \omega_0 = -\gamma B_0 \qquad \text{(equations 2.4)}$$

where the Larmor frequency ν_0 is in Hz, and ω_0 is in rad s^{-1}.

The precession of magnetization produces measurable components of magnetization in the xy plane (M_x, M_y). If a transverse coil is wrapped around the sample, the precessing magnetization induces a voltage in the coil. This small induced voltage is detected, amplified and recorded as the NMR signal or free induction decay (FID).

2.2 Rotating Frame of Reference and Bloch Equations

The applied magnetic field (B_1) rotates in the xy plane at the transmitter frequency (ω_{ref}). In order to simplify analysis and remove the time dependence of the rotation, it is assumed that the rotation of the B_1 field occurs in a rotating frame, in which the xy plane rotates about the z-axis. If the rotating frame has the same frequency as the transmitter frequency, the B_1 field appears to be static. Therefore the rotating frame of reference makes it simpler to study the magnetizations.

The magnetization of a sample is represented by the three magnetization components M_x, M_y, and M_z. The motions of the three magnetization components under the influence of relaxation and RF field were first described by Bloch in 1946 [2.3]. The Bloch equations:

$$\frac{dM_z}{dt} = \gamma B_1 M_y - \frac{(M_z - M_0)}{T_1}$$
$$\frac{dM_x}{dt} = \Delta\omega M_y - \frac{M_x}{T_2} \qquad \text{(equations 2.5)}$$
$$\frac{dM_Y}{dt} = -\Delta\omega M_x - \gamma B_1 M_z - \frac{M_y}{T_2}$$

where $\Delta\omega$, M_0 are the difference between Larmor frequency and the rotating frame frequency (offset), and magnetization at equilibrium respectively. These equations describe the exponential relaxation of the three magnetization components back to equilibrium with two relaxation time constants of T_1 and T_2.

2.3 Coherence

At equilibrium spins are randomly distributed in different directions, while the bulk magnetization is along the z-axis. According to the Boltzmann distribution the numbers of spins in different energy levels α and β are not equal. This produces a polarization along the z direction. By applying an RF pulse this polarization is rotated from the z direction to the transverse plane. The correlation between the phases of spins in the transverse plane while they are precessing, represents a type of coherence. There are different kinds of coherences: some result in observable signals and some are non-observable, some are wanted and some unwanted.

The total magnetization in a sample is the sum of all the contributions from the individual spins. Due to the very large number of spins it is not possible to follow their evolution individually. However, the quantum mechanical behaviour of a sample can be calculated without specifying the motions of individual spins by using the density operator. The density operator for a one-spin system can be written as follows:

$$\hat{\rho}(t) = a_x(t)\hat{I}_x + a_y(t)\hat{I}_y + a_z(t)\hat{I}_z \qquad \text{(equation 2.6)}$$

where the spin operators $\hat{I}_x$, $\hat{I}_y$ and $\hat{I}_z$ represent the components of the spin angular momentum in the x, y and z directions, and $a_x(t)$, $a_y(t)$ and $a_z(t)$ are coefficients which are related to the variation of M_x, M_y and M_z with time respectively. Therefore for a one-spin system a total of three operators is needed.

By applying different pulses, spin operators can be interconverted. For instance a pulse about the *x* and *y* axis with flip angle $\pi/2$ radians produces the following transformations:

$$\hat{I}_y \xrightarrow{(\pi/2)\hat{I}_x} \hat{I}_z$$

$$\hat{I}_z \xrightarrow{(\pi/2)\hat{I}_x} -\hat{I}_y$$

$$\hat{I}_x \xrightarrow{(\pi/2)\hat{I}_y} -\hat{I}_z$$

$$\hat{I}_z \xrightarrow{(\pi/2)\hat{I}_y} \hat{I}_x$$

In these rotations the density operator is given by:

$$\hat{\rho}(t) \equiv \cos\theta \hat{I}_n + \sin\theta \hat{I}_{n'} \qquad \text{(equation 2.7)}$$

where n represents the starting operator and n' the operator generated by the rotation.

For describing a two-spin system a total of sixteen (4^2) operators are constructed from the four single spin operators ($\hat{E}_1, \hat{I}_{1x}, \hat{I}_{1y}, \hat{I}_{1z}; \hat{E}_2, \hat{I}_{2x}, \hat{I}_{2y}, \hat{I}_{2z}$), where $\hat{E}$ is the identity operator. In this system the density operator is defined as in equation 2.6, but with a total of sixteen operators. In the same way for a three-spin system, 64 (4^3) operators are constructed.

The order of a coherence determines the effect of rotation about the z-axis through an angle ϕ on the coherence. The operator $\hat{I}_z$ is not affected by rotation around the z-axis, therefore it has zero order. The following table describes the rotations for different coherence orders.

Coherence order	Rotation about z-axis through ϕ_z
0	0
+1	$-\phi$
+2	-2ϕ
-1	$+\phi$
-2	$+2\phi$

Table 2.1 Relationship between coherence order and rotation angle about the z-axis (coherence order 0=zero-quantum coherence, coherence order ±1=single-quantum coherence, coherence ±2=double-quantum coherence).

Raising ($\hat{I}_{i+}$) and lowering ($\hat{I}_{i-}$) operators, having +1 and -1 coherence orders respectively, can be used to define the coherence order of spin *i* operators according to the following equations.

$$\hat{I}_{ix} \equiv \frac{1}{2}(\hat{I}_{i+} + \hat{I}_{i-})$$
$$\hat{I}_{iy} \equiv \frac{1}{2i}(\hat{I}_{i+} - \hat{I}_{i-})$$

(equations 2.8)

Therefore, $\hat{I}_{ix}$ and $\hat{I}_{iy}$ have equal mixtures of +1 and -1 coherence orders, known as single-quantum coherence. The $\hat{I}_{iz}$ operator is unaffected by a z pulse and has coherence order of 0, zero-quantum coherence. For example a two spin system with product operator $2\hat{I}_{1y}\hat{I}_{2y}$ has an equal mixture of coherence orders +2 and -2, double-quantum coherence, and order 0, zero-quantum coherence as follows:

$$2\hat{I}_{1y}\hat{I}_{2y} \equiv 2\times\frac{1}{2i}(\hat{I}_{1+} - \hat{I}_{1-})\times\frac{1}{2i}(\hat{I}_{2+} - \hat{I}_{2-})$$
$$\equiv \frac{1}{2i^2}(\hat{I}_{1+}\hat{I}_{2+} - \hat{I}_{1+}\hat{I}_{2-} - \hat{I}_{1-}\hat{I}_{2+} + \hat{I}_{1-}\hat{I}_{2-})$$

(equations 2.9)

where the coherence orders of $\hat{I}_{1+}\hat{I}_{2+}, \hat{I}_{1+}\hat{I}_{2-}, \hat{I}_{1-}\hat{I}_{2+}, \hat{I}_{1-}\hat{I}_{2-}$ are +2, 0, 0, and -2 respectively.

In a spin system comprised of *N* spins, coherence orders of +*N* to –*N* in integer steps are possible. Therefore, any one-spin system has coherence orders of -1, 0, +1 and a two-spin system has coherence orders of -2, -1, 0 +1, and +2. Only transverse

magnetizations ($\hat{I}_{ix}$ and $\hat{I}_{iy}$), which are single-quantum coherences, produce observable signals. By convention, only -1 coherence order is detected by the receiver, so this is the wanted coherence order. The coherence orders can be represented in a chart, in which each horizontal line represents an order. This diagram, a coherence transfer pathway (CTP), is used to show the changes in orders during a sequence of pulses, ending in the desired -1 coherence order. During free precession, spins maintain their coherence order, while an RF pulse can change the order. Phase cycling and gradient pulses are the two methods that are commonly used to select wanted coherence pathways and reject others [2.2].

2.3.1 Selecting a CTP Using Phase Cycling

An RF pulse can change the coherence order from p_1 to p_2. Therefore, the phase of a given pulse can be shifted to select a desired coherence pathway. When a pulse with phase ϕ_1 is applied, a coherence transfer pathway with coherence difference Δp (p_2 - p_1) experiences a phase shift $\phi_2 = -\Delta p \times \phi_1$. In phase cycling, a pulse sequence is repeated multiple times while changing the phase ϕ_1 of the pulse. Each time the phase of the receiver is made to follow the phase of the pulse so as to have a phase change of ϕ_2. In this case the signals from the desired coherence add up (because the phases of the pulse and receiver follow each other), while the signals from the other coherences cancel each other (because the phase of the receiver and the coherence change by different amounts). Common pulse phase (ϕ_1) shifts are 0°, 90°, 180°, 270°. For example a required coherence pathway of $+1 \rightarrow -1$, which has Δp of -2, experiences a phase change of $2\phi_1$ ($\phi_2 = -\Delta p \times \phi_1$). This means that by applying each of the four above RF phase shifts, this pathway experiences shifts of 0°, 180°, 360° (equivalent to

0°), and 540° (equivalent to 180°). So, if the phase of the receiver follows the sequence 0°, 180°, 0°, 180°, the signals of the wanted coherence will add up and the signal from the rest of the coherences will cancel. Therefore phase cycling can be used to choose the desired coherence history of a pulse sequence [2.2, 4].

2.3.2 Selecting a CTP Using Gradients

It is desirable to have a homogeneous magnetic field, to optimise the resolution of the spectrum. However, in some circumstances the magnetic field needs to be deliberately changed linearly along the z-axis for a short period. This variation of the magnetic field -a field gradient pulse- can be generated using an extra, gradient, coil in the NMR probe. Different sizes and directions of current are passed through the gradient coil to produce the gradient pulses. The consequence of having a field gradient in a sample is that the spins at different layers precess at different rates. In the presence of a gradient, the magnetic field can be described as follows:

$$B_z = B_0 + Gz \qquad \text{(equation 2.10)}$$

where B_z, B_0, G, z are the magnetic field in the presence of gradient, magnetic field without a gradient, strength of gradient, and distance along the field direction from the centre of the sample respectively. Such magnetic field gradients can be used to dephase and rephase the spins and select a specific coherence transfer pathway.

A gradient pulse dephases the spins as follows:

$$\phi_1(z) = -p_1\gamma G_1 z\tau_1 \qquad \text{(equation 2.11)}$$

where $p_1, \gamma, G_1, z, \tau_1$ are the coherence order at the time the gradient is applied, gyromagnetic ratio, strength of gradient, distance along the field direction from the centre of the sample, and duration of the gradient pulse respectively. After applying an RF pulse a variety of coherences are produced. By applying a second gradient pulse, the phase distribution caused by the first gradient pulse can be either refocused, or perturbed further, for specific coherence orders. The phase after the second gradient pulse, $\phi_2(z)$, can be refocused (rephased) if the total phase distribution is zero:

$$\begin{aligned} &\phi_1(z) + \phi_2(z) = 0 \\ &-p_1\gamma G_1 z\tau_1 - p_2\gamma G_2 z\tau_2 = 0 \\ &\frac{G_1\tau_1}{G_2\tau_2} = -\frac{p_2}{p_1} \end{aligned} \qquad \text{(equations 2.12)}$$

Equation 2.12 shows the refocusing condition for coherence orders p_1 and p_2, while other coherence orders are dephased and lost. For example if a coherence order change of $+1 \rightarrow -1$ is desired, according to equation 2.12 for the same duration of gradient pulses ($\tau_1 = \tau_2$), the strengths of the gradients must be equal ($G_1 = G_2$) [2.2, 5, 6]

2.4 Spectrometer Anatomy

An NMR spectrometer is a complex instrument constructed of different parts. Some of the main components are briefly explained as follows.

2.4.1 The Magnet

Modern NMR spectrometers are constructed around superconducting magnets. These magnets are assembled from a coil of wire typically made of copper, niobium, and tin, in which a current passes through the coil generating a magnetic field (B_0). This field is stable and remains forever without applying any additional electrical power. The superconductivity of the coils is protected by constructing a cold environment, placing them in a bath of liquid helium, in which the helium is surrounded by liquid nitrogen (which is cheaper), all surrounded by a vacuum flask to protect from the outside heat. A sample can be placed in the magnet by going though the bore, a vertical tube, into the probe, where the maximum magnetic field exists. This field needs to be made homogeneous in the active volume by the use of additional coils known as shim coils (see section 2.4.6). The temperature of the sample is not affected by the magnet, as the bore is insulated from the superconducting coil [2.7].

2.4.2 The Probe

When a sample is inserted into a magnet it is placed in the probe, which is mounted in the bore. A probe is one of the most expensive parts of the NMR instrument, and can be exchanged for different experiments. In a basic probe the different parts can be simply described as an electronic circuit of a coil wound perpendicular to $B_{0,}$ and two capacitors, tuning and matching capacitors. The coil surrounds a small region, known as the active volume, (about 1.5 cm) of the bottom of the sample. Radiofrequency currents (RF pulses) are applied, generating an oscillating magnetic field (B_1) in the sample, which is much smaller than the original magnetic field (B_0), but through resonance can affect the magnetic moments presented in the sample.

There are different frequency channels for different isotopes, the desired channel will be tuned to focus on the frequency region of the studied isotope. During the tuning of the probe, the tune and match capacitors are adjusted until the frequency of the tuning circuit matches the Larmor frequency of the desired isotope and the impedance of the resonant circuit matches that of the spectrometer electronics (RF amplifier, preamplifier) to which it is connected. The coil in the probe produces thermal noise, caused by the thermal motion of the electrons, which can be reduced by cooling the coil.

2.4.3 The Transmitter

The transmitter of a spectrometer is constructed of different parts to produce an RF pulse with the desired amplitude and phase. An RF synthesizer generates a radiofrequency wave oscillating at the reference frequency, ω_{ref}, (transmitter frequency: the frequency of the middle of the spectral window). At the output of the RF synthesizer a pulse programmer controls the phase and duration of the pulse. The duration of a pulse is controlled by a gate, open to allow the radiofrequency to pass and shut to terminate the pulse. The generated pulse has a low power (mW) and needs to be amplified (to about 100 W) by an amplifier which has a fixed gain. If the power of a pulse needs to be varied (a non-selective pulse is more powerful than a selective pulse) an attenuator placed before the amplifier controls the change.

2.4.4 The Receiver

The RF pulse produced travels into the duplexer from the transmitter. The duplexer is the device that sends and receives the radiofrequency to and from the probe. The high voltage RF pulse passes through the duplexer towards the probe and ends up in the coil around the sample, producing the B_1 field. The voltage induced by the precession of the nuclear spins is received by the same coil and travels back towards the duplexer. This weak current is passed through an amplifier moving towards the receiver. These currents (NMR signals) oscillate at hundreds of MHz, too fast for digitization. Therefore, the NMR signal oscillating at the Larmor frequency (ω_0) is combined with the spectrometer reference signal oscillating at frequency of ω_{ref} as in equation 2.13 to produce a detectable signal which has a frequency, of the order of 1 MHz or less, known as the offset.

$$\Omega = \omega_0 - \omega_{ref} \qquad \text{(equation 2.13)}$$

NMR signals can have positive or negative offsets, depending on the position of the signal relative to the spectrometer reference frequency. These two possibilities can be distinguished by quadrature detection, where the NMR signals received from the probe are split into two, and transferred to two separate mixers. In one of the mixers the NMR signal represented by $\cos(\omega_0 t)$ is multiplied by the receiver reference, $\cos(\omega_{ref} t)$, as follows:

$$\cos(\omega_0 t) \times \cos(\omega_{ref} t) = \frac{1}{2}\left[\cos(\omega_0 + \omega_{ref})t + \cos(\omega_0 - \omega_{ref})t\right] \qquad \text{(equation 2.14)}$$

The other mixer multiplies the NMR signal by a 90° phase shifted receiver signal, $\cos(\omega_{ref} t + \pi/2)$, which can be written as $-\sin(\omega_{ref} t)$.

$$\cos(\omega_0 t) \times \left(-\sin(\omega_{ref} t)\right) = \frac{1}{2}\left[-\sin(\omega_0 + \omega_{ref})t + \sin(\omega_0 - \omega_{ref})t\right] \qquad \text{(equation 2.15)}$$

The low frequency signals in equations 2.14 and 2.15, $\cos(\omega_0 - \omega_{ref})t$ and $\sin(\omega_0 - \omega_{ref})t$ are separated, so the orthogonal x (real) and y (imaginary) components of the magnetization have been detected. The signals are passed to separate analogue-to-digital converters (ADCs). An ADC converts the continuous wave of an NMR signal to a discrete set of numbers to be stored in the computer. A three-bit ADC (three binary bits) contains eight levels. The input wave signal of an ADC changes

continuously, but the ADC with eight levels only picks the level that is closest to the input. Therefore, the output is an approximation of the original signal. As the number of bits of an ADC increases, the number of levels will increase, resulting in more accurate digitization [2.2].

2.4.5 Field-Frequency Lock

The magnetic field in the sample can change slowly due to temperature fluctuations, movement of a metal material close to magnet, small imperfections in the superconducting magnet, etc. Such magnetic shifts disturb the NMR signals. Therefore, a device known as the field-frequency lock is used to compensate any changes and keep the magnetic field constant. The field-frequency lock is a spectrometer within the main spectrometer which is tuned to the resonance frequency of the deuterium nuclei (^{2}H) within the sample, and constantly monitors their frequency. The lock monitors the intensity of the dispersion mode of the deuterium signal, which should be zero when it is on resonance. If there is a change in the field, the intensity of the dispersion mode becomes positive or negative; therefore a positive or negative current is added to the z0 shim coil to compensate the field variation and restore the initial lock frequency.

2.4.6 Shim Coils

The B_0 magnetic field has minor spatial inhomogeneity; reasons include magnet design, ferromagnetic objects close to the magnet, and the magnetic susceptibility [2.8, 9] of the sample itself. It is essential to correct this inhomogeneity at the position of the sample solution to obtain high quality spectra. The depth of the solution is usually about 5 cm from the bottom of the NMR tube. This region is surrounded by a set of coils called shim coils. The current in shim coils produce a magnetic field opposite to the inhomogeneity, to cancel the magnetic field inhomogeneity in the active volume. As the inhomogeneities can be in different directions, the shim coils have different functional forms, typically spherical harmonics. The shimming process is a regular task in every NMR experiment which can be difficult and time-consuming. Therefore attempts have been made to make this process automatic [2.10].

2.5 Pulse Sequences

The simplest pulse sequence can be described as the application of a single RF pulse and a signal detection period. By placing a sample in the magnet the nuclear spins build up a bulk magnetization along the z-direction, reaching the equilibrium state. This remains until the system is perturbed by applying an RF pulse, in which the produced oscillating magnetic field (B_1) rotates the magnetization into the transverse plane through an angle depending on the power and duration of the pulse. The spins precess about B_0 at their Larmor frequencies (depending on their gyromagnetic ratio and the magnetic field) returning back to the equilibrium position, while producing a current in the surrounded coil. Different types of experiments have various pulse sequences with multiple pulses and in some cases magnetic field gradient pulses.

2.6 Chemical Shift

Different parts of a molecule experience different magnetic fields due to the small magnetic fields generated by the electronic structure, which can add or subtract from the applied magnetic field. This dependence is known as the chemical shift. Therefore, the precession frequency of a nucleus is not only dependent on the magnitude of the external magnetic field and gyromagnetic ratio, but also on the molecular structure. Electronegative atoms withdraw electron density from neighbouring groups, and hence deshield their nuclei from the local magnetic field. Therefore, these deshielded nuclei result in higher chemical shifts.

As was shown in equation 2.4, the frequencies of NMR signals are dependent on the strength of the magnetic field (B_0). This means that if NMR spectra are represented in frequency scale, the output spectra of the same sample would be different in different magnetic fields. To overcome this problem another scale was introduced which is independent of the external magnetic field. In this scale, known as the chemical shift scale, the frequencies of peaks in a spectrum are assessed in relation to the resonance frequency of a reference signal.

$$\delta(ppm) = \frac{\Delta\nu(Hz)}{\nu_0(MHz)} \qquad \text{(equation 2.16)}$$

where $\Delta\nu$ is the difference in Hz between the frequency of a given peak and a reference compound (usually TMS), and ν_0 stands for the reference signal frequency. TMS (Tetramethylsilane) is a common reference molecule as it is soluble in most organic solvents and it shows a sharp single signal far from the rest of the signals. TSP

(3-trimethylsilylpropionic acid sodium salt) is a common reference compound for polar solutions.

2.7 Spin-Spin Coupling

Spin-spin couplings in a sample can be divided into two categories: direct and indirect dipole-dipole couplings. Each nuclear spin generates a magnetic field, which interacts with the magnetic fields of the neighbouring spins through space. This intermolecular and intramolecular coupling is known as direct dipole-dipole coupling. This coupling depends on the distance, and on the angle between a line joining the spins and the magnetic field direction. As a result of rapid molecular motions these couplings cancel one another or are so weak that they may be ignored in liquids, and do not appear in spectra.

Intramolecular spin-spin coupling mediated by electrons is known as indirect dipole-dipole coupling (scalar coupling or *J*-coupling). When spins are coupled through one or more chemical bonds, the correlations between spins of bonding electrons and other nuclear spins produce higher or lower energy states. These energy states result in multiple patterns. The magnitude of the splitting is determined by the coupling constant (*J*), in Hz, and in contrast to the chemical shift it is not dependent on the field strength. Scalar couplings cause multiplet structure of signals (doublet, triplet, etc.), which provides information about the neighbouring environment of a spin and helps define the chemical structure of a sample [2.1, 2].

Coupling constants can be determined experimentally by measuring the splitting in a multiplet signal. In poorly resolved data where splittings are not observed due to a

non-uniform magnetic field, resolution enhancement (see 2.9.2) can assist. There are various theoretical methods [2.11, 12] to calculate the *J* coupling. One of the most commonly used methods is the Karplus relationship [2.13], in which a three bond coupling constant is calculated as follows:

$$^{3}J = A + B\cos\phi + C\cos 2\phi \qquad \text{(equation 2.17)}$$

where ϕ is the dihedral angle, and *A*, *B*, *C* are empirical parameters which are determined from fitting to a Karplus equation of a series of *J* measurements for which the dihedral angle is known.

If the chemical shift difference between two nuclei is large enough in comparison with their coupling constant ($|\delta_1 - \delta_2| >> |J_{12}|$), the coupling is known as a weak coupling. In spectra in which strong coupling (the difference between chemical shifts is comparable with the *J*-coupling) exists, line intensities are perturbed and very complicated spectra can result [2.14].

Scalar couplings can exist between any close bonded nuclei. For some nuclei such as ^{13}C which can be coupled to many protons, signal splitting can produce a complicated crowded spectrum. Therefore, it is desirable to suppress the couplings. This is done by broadband decoupling, in which the proton nuclei are irradiated continuously by intense radiofrequency. The resulted decoupled spectrum shows better signal-to-noise ratio, as all the intensity appears in a single signal rather than in a multiplet.

2.8 Relaxation

As mentioned in section 2.1, at equilibrium the population difference between spins in different energy levels is described by the Boltzmann distribution. If a sample is suddenly placed into a magnetic field, or a long RF pulse is applied, the population difference between the energy levels is zero. This state is known as the sample being saturated. The process that takes the spins back to equilibrium (higher population of spins at the lower energy level), in which M_z returns to $M_{0,}$ is known as spin-lattice relaxation (T_1). This relaxation occurs due to random interactions between the individual nuclei and the surrounding environment, where the magnetic field fluctuates. If this fluctuation is in the right frequency range (Larmor frequency), it can cause relaxation. Several parameters can cause the fluctuation of the local magnetic field, such as dipole-dipole intermolecular and intramolecular interactions, scalar coupling, paramagnetic interactions, quadrupolar interactions, spin rotation, chemical shift anisotropy, and cause spin- lattice relaxation [2.15] .

In addition, due to local magnetic field fluctuations the Larmor frequencies of spins are perturbed, causing the loss of phase coherences. This process is known as spin-spin relaxation (T_2) [2.14], in which the transverse magnetization decays towards zero.

2.9 Data Processing Techniques

2.9.1 Fourier Transformation

In NMR spectroscopy the signals are obtained by measuring the net transverse magnetization as a function of time. The x and y components of magnetization are detected simultaneously, producing S_x and S_y signals as follows:

$$S_x(t) = S_0 \cos \Omega t \exp\left(\frac{-t}{T_2}\right)$$
$$S_y(t) = S_0 \sin \Omega t \exp\left(\frac{-t}{T_2}\right) \qquad \text{(equations 2.18)}$$

where S_0, Ω, T_2 represent the signal intensity, offset, and spin-spin relaxation time. These equations can be combined to show the complex form of the signal:

$$S(t) = S_x + iS_y \qquad \text{(equation 2.19)}$$

Due to the simplicity of interpreting information in the frequency domain, the time-domain signal obtained ($S(t)$), known as a FID (Free Induction Decay) is transformed to the frequency domain by a mathematical procedure known as the Fourier transform (FT^-). This operation is a reversible procedure which can also convert the frequency domain to time domain (inverse Fourier transformed, FT^+). The Fourier transform can be viewed as multiplying the time domain function by different trial sine and cosine waves and integrating. The mathematical formulation of the real part of the Fourier transform is:

$$S(f) = \Sigma_{i=1}^{i=N} S_{FID}(t_i)\cos(2\pi f t_i) \qquad \text{(equation 2.20)}$$

where $S(f), S_{FID}(t_i), t_i, N$ are the spectrum as a function of frequency, intensity of the FID as a function of time, time corresponding to the i^{th} data point, and the total number of data points respectively.

2.9.2 Sensitivity and Resolution Enhancement

During the acquisition time of an experiment, the free induction decay and noise generated by the RF receiver, coils, probe, ADC etc. are recorded simultaneously. It is important for a spectroscopist to obtain a spectrum with high signal-to-noise ratio, so care is needed in data processing.

A FID contains stronger signals at the beginning, which decay over time resulting in weaker signals at the end, with noise at all times. In order to reduce the impact of the noise, attenuating the weak parts of a FID and leaving the strong parts unchanged, the FID is multiplied by a function known as a weighting function. This is typically an exponential function which begins with the value one and gradually reduces to zero over time. If a FID is not multiplied by a weighting function the frequency domain resulting from the Fourier transform is a spectrum with a noisy baseline. This is due to the noise in the original time domain data. If a weighting function decays rapidly, the weak part of the FID will be eliminated and the noise will also attenuate. This elimination results in a reduction in peak height and the line becoming broader. If the weighting function decays too slowly, the spectrum produced will be noisy. A weighting function changes the signal-to-noise ratio (S/N) and the linewidth of a

spectrum. If the increment in linewidth is equal to the linewidth of the original spectrum, the S/N ratio will reach its highest value. This kind of a weighting function is known as a matched filter.

The natural linewidth of a signal is determined by spin-spin relaxation. The faster the spin-spin relaxation is, the shorter the FID will be, which results in line broadening. Homonuclear dipolar interactions, temperature, inhomogeneity of the magnetic field all influence line broadening [2.16, 17]. The connection between S/N ratio, linewidth and weighting function determines the choice of parameters for a weighting function [2.2].

2.9.3 Reference Deconvolution

Despite many instrumental developments and new technologies, NMR experiments are still imperfect because of instrumental limitations. Lineshape distortions, spinning sidebands, amplitude, phase and frequency errors are typical effects. It is important to be able to obtain high quality experimental data from imperfect NMR experiments. A simple and fast data processing method, reference deconvolution, provides lineshape correction without distortion of the information contained in a spectrum. The majority of experimental lineshape imperfections affect all the signals of spectrum in the same way. The reference deconvolution method uses several steps to remove these distortions. The basic explanation of this method is as follows.

The original time-domain signal (FID), is zero filled, Fourier transformed (FT^{-}) and phase corrected to produce the experimental spectrum. This spectrum is inverse Fourier transformed (FT^{+}), and transformed by indirect Hilbert method to produce the time-shifted full experimental free induction decay $S_e(t)$. In the experimental

spectrum a signal is selected as the reference for determination of the necessary correction, so that the latter can be applied to all the other signals in the spectrum. A suitable frequency region (ω_L to ω_R) is selected for the reference signal, wide enough to contain all the reference signal intensity, including any spinning sideband signals. By applying a base line correction, the selected region is lowered to zero at both sides. The remaining signals outside of the selected region and the imaginary components are zeroed; thus the inverse Fourier transformed (FT^+) signal contains the time-domain reference signal only. The inverse Fourier transformation of the result produces a symmetrical FID consisting of real and imaginary parts. The second half of this FID is discarded, producing the complete FID of the reference signal. The ideal reference time domain signal can be obtained either by explicit calculations, or by inverse Fourier transform of the ideal reference spectrum. Then the ideal reference signal is multiplied by a weighting function to produce the ideal FID. By multiplying the free induction decay of the original spectrum (including all the signals), $S_e(t)$, by the complex ratio of the ideal FID of the reference signal ($S_i(t)$) to the extracted experimental FID of the reference signal ($S_r(t)$), a corrected FID ($S_c(t)$) is generated. This FID is then Fourier transformed to produce a corrected spectrum $S_c(\omega)$ without instrumental lineshape errors [2.18-21].

$$S_c(\omega) = FT^{-}[S_e(t) \times \frac{S_i(t)}{S_r(t)}] \qquad \text{(equation 2.21)}$$

In practice, reference deconvolution can be carried out by defining a reference signal region, base line correction and applying the '*fiddle*' command (developed for Varian

Vnmr software) with appropriate values for *lb* (Lorentzian line width parameter) and *gf* (time constant Gaussian).

Chapter 3

Analysis of NMR Timecourse Data

3.1 Introduction

NMR spectroscopy is widely used in kinetic studies, in which the rates of reactions are measured and interpreted. NMR provides information about both intermolecular and intramolecular dynamics [3.1]. Different methods for studying the kinetics of a system by NMR include band-shape analysis [3.2, 3], magnetization exchange [3.4, 5], and time-resolved NMR methods [3.6].

Line-shape analysis and magnetization exchange methods can provide kinetic information about motion, binding or exchange in a sample [3.7]. In band-shape analysis the exchange of nuclear spins in a system can be investigated by monitoring its effect on the line-shapes. Any molecular dynamics that alter the magnetic environment of a nucleus can influence the line shape of its signal, depending on the timescale of the dynamics [3.8]. If slow exchange occurs between two inequivalent sites, different signals are observed. As the exchange becomes faster, the signals become broader until they merge and become an average signal. This method depends on comparison of signals intensities, shapes, linewidths and chemical shift differences of signals at different temperatures [3.3]. In magnetization exchange methods (eg. selective inversion recovery) the quicker the rate exchange between two spins is, the more rapidly the magnetizations exchange. This method indicates how rapidly the magnetization is transmitted from the inverted peak to a non-inverted peak during magnetization recovery [3.4, 9, 10].

Time-resolved NMR spectroscopy has been utilized for kinetic studies in a variety of areas in dynamics and chemical reactions, including solvent-protein interactions [3.11], RNA structural transformations [3.12], conformational changes and intermediate states of

folding/unfolding proteins [3.11, 13, 14], drug structure elucidation and drug degradation studies [3.15], depolymerization and de-N-acetylation reactions [3.16], alcoholysis of vegetable oils [3.17], ionic liquids [3.18], and in flow studies [3.19].

In this research, the kinetics of chemical reactions have been monitored by using time-resolved NMR, in which a series of ^{1}H spectra of a reaction are acquired in a timecourse. This method allows simultaneous monitoring of reactant degradation and product formation. Time-resolved NMR can provide the basis of a detailed understanding of reaction kinetics and mechanism. However, it is not always simple to interpret the resultant NMR data, for example where signals overlap or where chemical shifts change in the course of a reaction. The first studied reactions were reductive amination of aldehydes/ketones. To be able to monitor signal changes during the time course of a reaction, the rate of a reaction should be neither too fast nor too slow. Here the kinetics of a Ugi-Smiles reaction and an ester hydrolysis have been monitored. Changes of signal intensity can be followed by manual integration or peak height measurement, but only if signals do not overlap. Least squares [3.20] fitting can give optimum estimation of line positions and integrals, if the lineshape is known, even in the presence of overlap. Fitting to the exact shape of the NMR signals has been done here by simple statistical non-linear least squares methods with Mathematica 5.1 and 6.0 software [3.21].

3.2 Theory

3.2.1 Least Squares Methods

Numerical data can be analysed by models consisting of a dependent variable, one or more independent variables, and unknown adjustable parameter/s, in a statistical technique known as regression analysis [3.22]. One class of such methods is known as the least squares (*LS*) technique, first published in 1805 by Legendre [3.20]. This method estimates the parameters of a model in such a way as to obtain the minimum sum of squares of residuals (*r*) (difference between the experimental and the fitted data) (equation 3.1).

$$LS = \sum_{i=1}^{i=n} r_i^2 \qquad \text{(equation 3.1)}$$

Least squares methods fall into two categories, linear and non-linear. In linear regression (equation 3.2) the dependent variable (y_n) is a linear function of the unknown parameters ($P = p_1 + p_2 + ...p_n$) which will be estimated from the data. The dependent variable can have linear or nonlinear dependence on the independent variables (x_n) [3.23].

$$\begin{aligned} y_n &= p_1 x_{n1} + p_2 x_{n2} + ... + \xi_n \\ &= (x_{n1} + x_{n2} + ...)P + \xi_n \end{aligned} \qquad n = 1,2,..,N \qquad \text{(equation 3.2)}$$

where ξ_n is the error.

In non-linear regression, the model is not a linear function of the parameter/s and the dependent variable (y_n) is a function of both the independent variables (x_n) and parameters (P) [3.23].

$$y_n = f(x_n, P) + \xi_n \qquad n = 1,2,..,N \qquad \text{(equation 3.3)}$$

The two important differences between linear and non-linear least squares fitting, apart from the different types of function, are the necessity of estimating initial parameters and the iterative procedure needed in the non-linear method [3.24]. Non-linear least squares fitting methods can fit a broad range of models. As shown in equation 3.4, they iterate to find the parameters that define the best fit between the model and the dataset.

$$p_n^{k+1} = p_n^k + \Delta p_n \qquad \text{(equation 3.4)}$$

where k is the iteration number, and Δp_n is a vector of increments (shift vector). Providing initial parameter values can be a problem with some datasets. These values need to be reasonably close to the true parameters, or the procedure may not converge, or may converge to a local minimum. The decision as to when the iteration should be terminated - convergence criteria - can also be an issue in some cases. Both linear and non-linear methods can estimate the parameters in relative small data sets, and both are sensitive to outliers. The existence of one or two outliers in data can seriously change the output result.

The non-linear least squares fitting method used in this work is the iterative Levenberg-Marquardt algorithm [3.24]. During iteration the algorithm changes the direction and length of the shift vector (Δp_n in equation 3.4) towards the direction of steepest descent [3.25].

3.2.2 Two-Dimensional J-Resolved Spectroscopy

A 2D (two-dimensional) NMR experiment can be simply described by a sequence containing four time intervals: preparation, evolution (t_1), mixing, and detection (t_2) which is repeated in a series of experiments. The preparation time changes the normal status of a spin system, for example, by applying a pulse to generate transverse magnetization. During the evolution time (t_1) the spins are allowed to evolve depending on their chemical shift and scalar spin-spin coupling. In the last time interval, t_2, the signal is detected (the voltage generated by the precession). A 2D spectrum is obtained by measuring a series of FIDs, $S(t_2)$, for incremented values of t_1, resulting in a two dimensional data matrix $S(t_1, t_2)$. These data are Fourier transformed twice, generating a 2D spectrum as a function of two frequencies $S(F_1, F_2)$. Having a large number of FIDs in 2D experiments makes them more time consuming than 1D NMR experiments.

One two-dimensional NMR method, known as 2DJ spectroscopy [3.26-28], is a popular method to distinguish between overlapping signals which have similar chemical shifts. This is done by separating the chemical shift and spin-spin coupling parameters, which improves the resolution of individual signals. 2DJ spectroscopy is

used to analyze crowded spectra of complex spin systems, and to distinguish between heteronuclear and homonuclear couplings.

The homonuclear J-resolved 2D experiment with the following pulse sequence is applicable to proton NMR and any nuclei with homonuclear couplings.

90°_{x} - $t_{1/2}$ - 180°_{x} - $t_{1/2}$ - acquisition (t_2)

After initial excitation of spins, the individual components of spin multiplets precess throughout the first evolution time ($t_{1/2}$) under the influence of chemical shift and spin-spin coupling. The 180°_{x} pulse rotates the spins around the x-axis and interchanges the spin states of passive spins. After the second delay the spins have different phases depending on the average precession rate over the evolution time (t_1); at time t_1 the chemical shifts and effects of magnetic field inhomogeneity are refocused, but not the effects of couplings. The incremented evolution time (t_1) generates phase modulated signals, which are Fourier transformed to produce the frequency domain F_1. As the chemical shift effect is suppressed during t_1, the F_1 domain only displays the multiplet structure caused by the spin-spin coupling.

The decaying signal in t_2 is detected and Fourier transformed to the frequency domain F_2. This domain contains both the chemical shift and the spin-spin coupling frequencies. The two-dimensional spectrum is typically displayed as a contour diagram, with multiplets being aligned along diagonals tilted 45° from the F_2 axis.

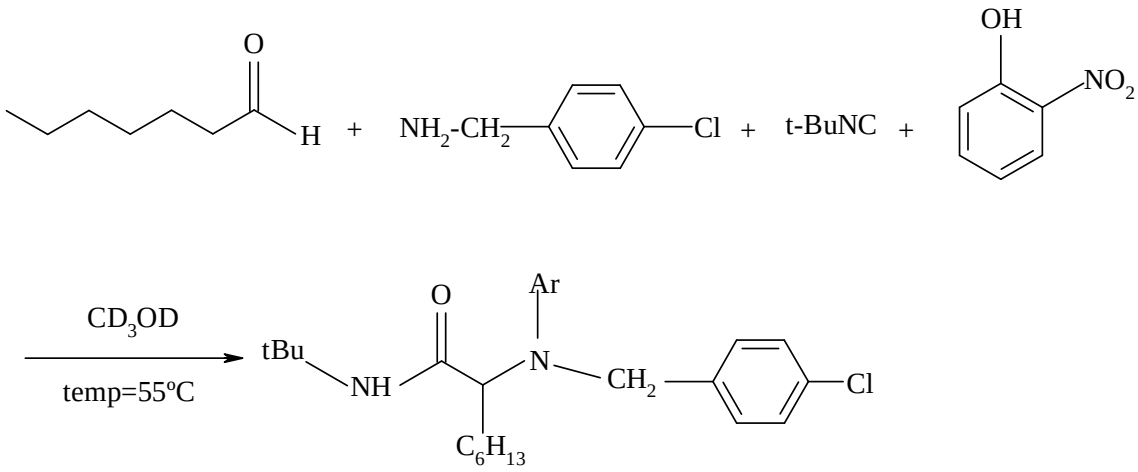
In many 2D experiments, Fourier transformation of the phase-modulated signals (obtained from arrayed t_1) yields a complex mixture of absorption and dispersion lineshapes, known as ‘phasetwist’ lineshapes. At the exact resonance offset of a signal

the lineshape is purely absorption mode, but moving further from the exact resonance the lineshape shows more and more dispersion mode. There are two problems with these lineshapes. First, the superposition of 2D absorption and dispersion shapes has the disadvantage of producing broad signals, in comparison to those signals from the absorption mode, which interfere with neighbouring signals. The second problem occurs on 45° projection of the phasetwist signals. The negative and positive parts superimpose, and thus cancel one another, so that the peak vanishes. This problem can be avoided by using the absolute value mode. Overlap problems can be reduced by multiplying the time-domain signal by an appropriate weighting function, such as a pseudo-echo [3.29] or sine-bell [3.30], to give an FID which is stronger at the middle and decays symmetrically on either side. Such data processing greatly affects the sensitivity, due to the loss of early parts of the original FID that contribute to high signal-to-noise ratio. The Fourier transformation of the weighted FID gives a symmetric peak with no apparent dispersion mode "tails". However, absolute-value mode still needs to be applied. The 2DJ spectrum can show distorted multiplet intensities as a result of these signal manipulations [3.28, 31-35].

3.3 Experimental Work

3.3.1 Ugi-Smiles Rearrangement Reaction

One of the studied reactions was an Ugi-Smiles rearrangement, involving four components: heptanal, 4-chlorobenzylamine, tert-butyl isocyanide, and 2-nitrophenol [3.36]. The solution was made by mixing equal concentrations (1 M) of each of the components in methanol-d_4 as solvent, using TMS as the reference compound.



Scheme 3.1 Ugi-Smiles rearrangement of mixture of heptanal, 4-chlorobenzylamine, tert-butyl isocyanide, and 2-nitrophenol in methanol-d_4 at 55 °C.

The experiment was performed in a 500 MHz Varian Unity spectrometer with a 5 mm $^{1}H/^{13}C/^{15}N$ triple probe. The spectra obtained were processed using Varian VNMR software on a Sun workstation. ^{1}H spectra of each of the four components, and the mixture of them in a time course array, were obtained. The total duration of the latter experiment was 66 h, with a recycle delay of 32 s, 90° pulse width of 10.4 μs, and acquisition time of 3.9 s for each ^{1}H spectrum. The timecourse was obtained by arraying the number of transients; the spectra were recorded at gradually increasing intervals in order to obtain the right time resolution for each stage of reaction with

optimum S/N ratio and without obtaining too many spectra. At the beginning of a reaction, where the rate of a reaction is usually fastest, the first 50 spectra are obtained rapidly using a single transient each. For a typical experiment with a recycle delay of 30 s, the first 50 spectra each take 30 s to be recorded (number of transients=1). The second block of 50 spectra has 4 transients, the third, and fourth blocks of spectra each has 16, and 64 transients respectively, while the fifth block of 12 spectra has 256 transients each.

The array of ^{1}H spectra for this reaction was complicated and indicated many changes with overlapping and non-overlapping signals (figure 3.1). A variety of experiments such as 2DJ, DOSY, and COSY were used to assist assignment.

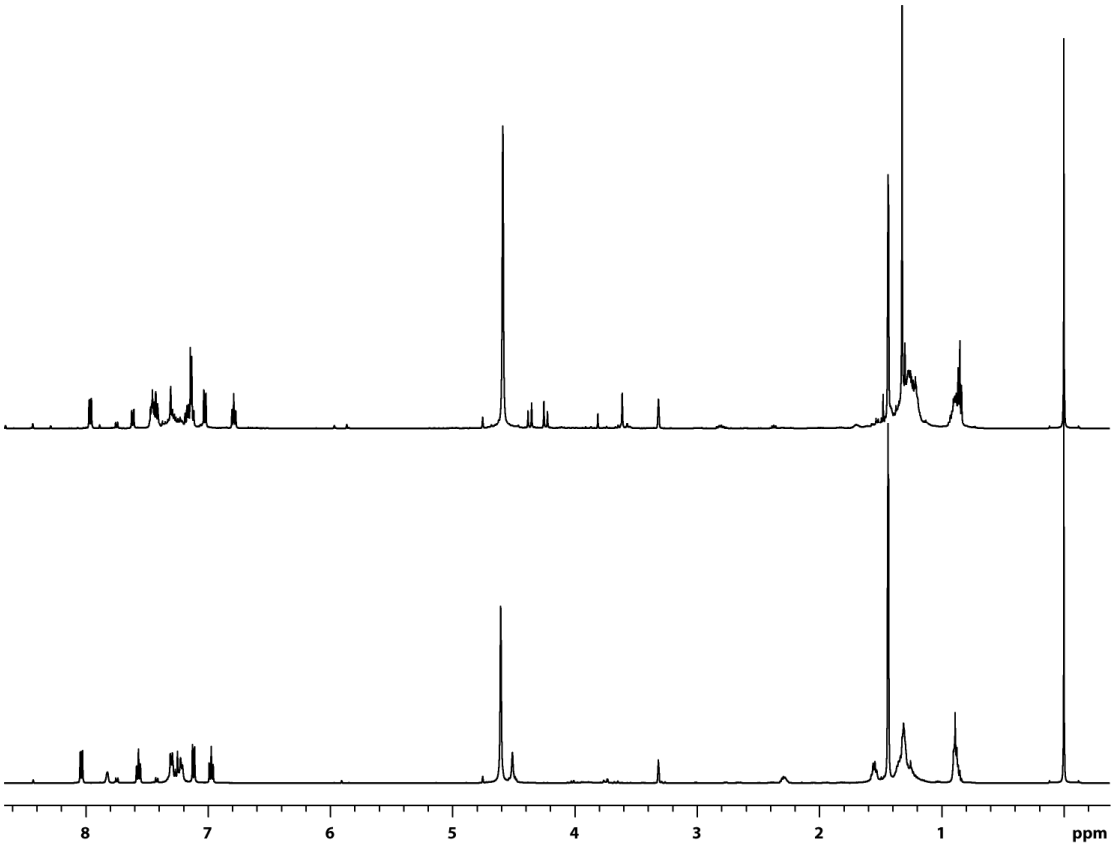


Figure 3.1 ^{1}H spectra of the mixture of four components of heptanal, 4-chlorobenzylamine, tert-butyl isocyanide and 2-nitrophenol in methanol-d_4 at temperature of 55˚C. The first spectrum (bottom) was obtained after half a minute from acquisition, and the last spectrum (top) was obtained after 66 hours.

A 2DJ spectrum of the Ugi-Smiles mixture was obtained after the reaction was completed. This spectrum was used to estimate the number of overlapping signals. The first objective was to determine the number of overlapping signals in the methyl region (0.82 – 0.95 ppm in figure 3.2) for further data processing. The number of overlapping signals for this region was observed to be 8 triplets.

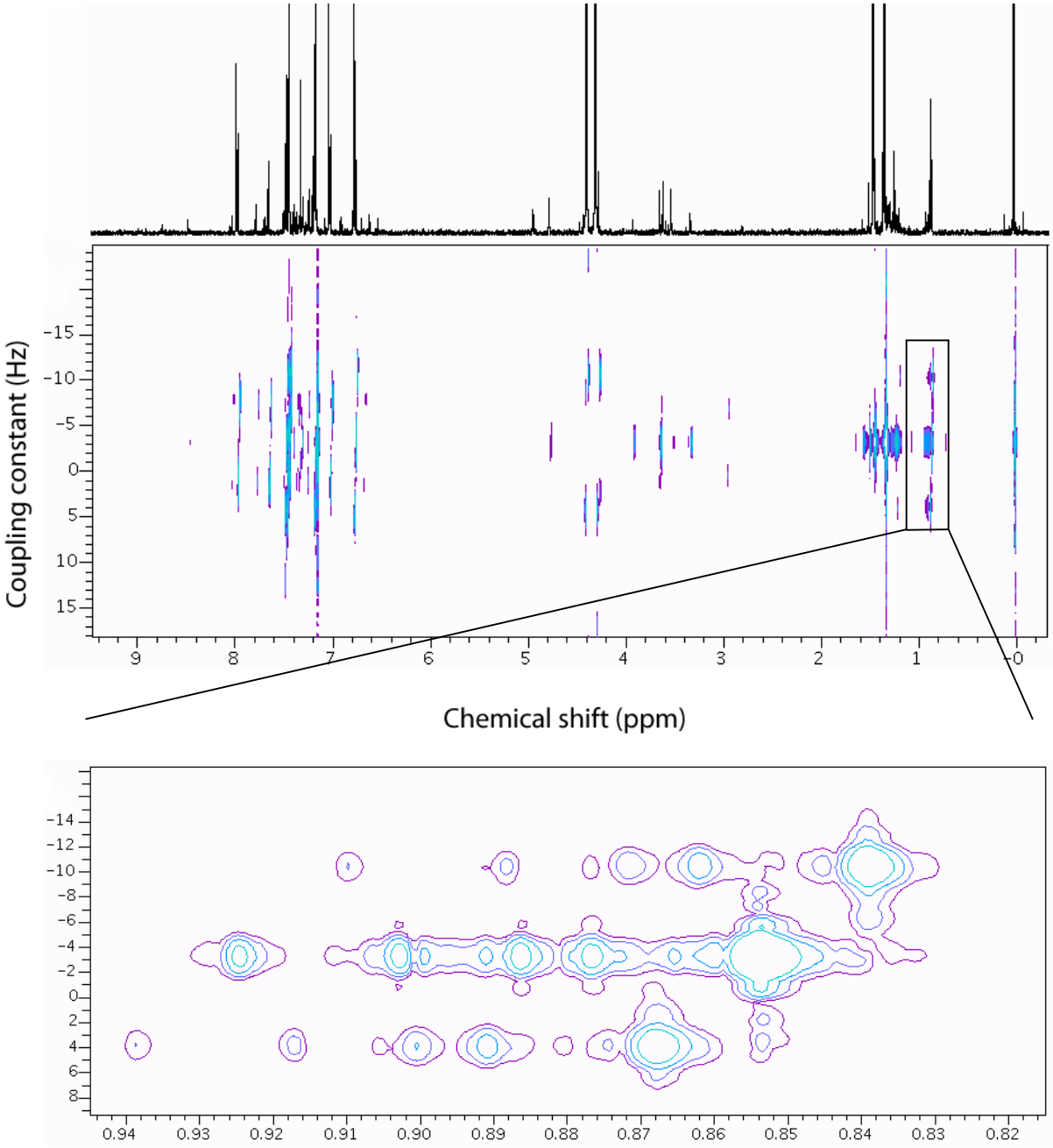


Figure 3.2 2DJ spectrum of the Ugi-Smiles reaction, expanding the methyl region.

3.3.1.1 Data Processing

All the ^{1}H spectra obtained for the Smiles reaction were Fourier transformed, reference deconvoluted using a 1.5 Hz Lorentzian line broadening, and baseline

corrected in Vnmr 6.1C. Data were then exported to Mathematica as a text file. All statistical analyses and fitting processes were performed in Mathematica version 5.1. The array of ^{1}H spectra included many regions with overlapping signals. Various data processing methods were used to attempt to fit the overlapping methyl region (figure 3.3) of the spectra.

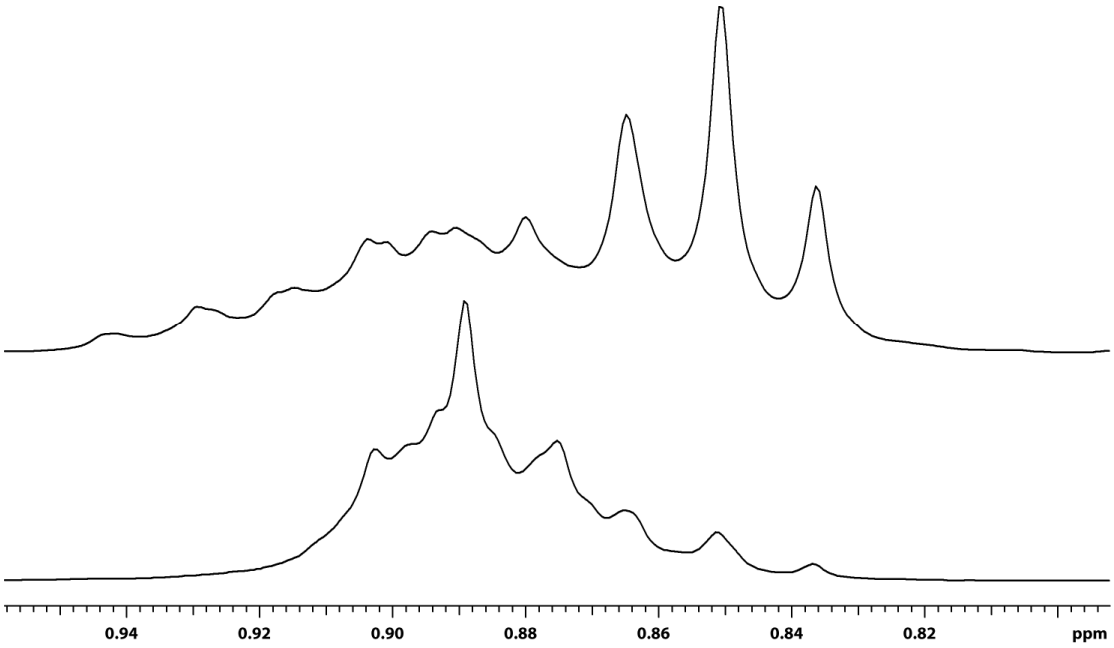


Figure 3.3 First spectrum of the methyl region of Smiles reaction (bottom) and the last spectrum (top).

The data from the 2DJ spectrum (figure 3.2) suggested 8 overlapping triplets in this region. An attempt was made to find the individual triplets by non-linear least square fitting to the Lorentzian lineshape function, as follows:

$$L(f) = \left(\frac{aw_1}{2\pi(f - f_0)^2 + \left(\frac{w}{2}\right)^2}\right) + \left(\frac{0.5aw_2}{2\pi(f - f_0 - j)^2 + \left(\frac{w}{2}\right)^2}\right) + \left(\frac{0.5aw_2}{2\pi(f - f_0 + j)^2 + \left(\frac{w}{2}\right)^2}\right)$$

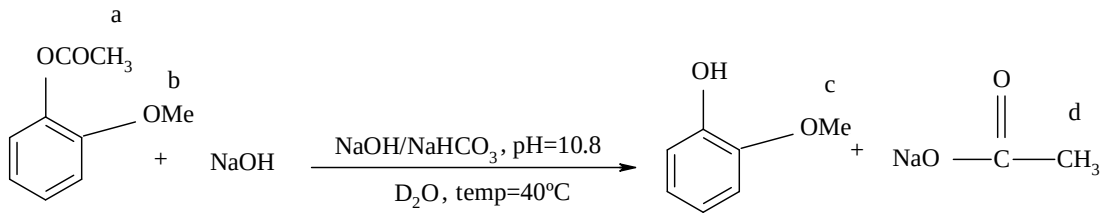
(equation 3.5)

where *a*, w_1, w_2, f_0 *and j* are the amplitude, width of the middle peak, width of the side peaks, frequency of the middle peak and the coupling constant between the middle peak and the side peaks respectively.

Because it was not possible to separate the overlapping signals by least squares fitting (figure 3.5 in section 3.4.1), another method was attempted. This was to identify one or more triplet signals within the structure of the overlapped region, and to subtract the guessed triplets from the raw data. This process was repeated each time by guessing another set of triplets and subtracting them from the preceding spectrum. It was hoped that this continued process would result in the identification of all the overlapped triplets. However, it did not succeed, due to the complexity of the overlapping region.

3.3.2 Ester Hydrolysis Reaction

The next reaction investigated was an ester hydrolysis. An array of ^1H spectra of a solution (0.03 M) of 2-methoxyphenylacetate in pH 10.8 sodium hydroxide/sodium hydrocarbonate buffer (0.08 M) in D_2O was monitored for 5 h at 40˚C using a 500 MHz Varian Unity spectrometer with a 5 mm $^1H/^{13}C/^{15}N$ triple probe.



Scheme 3.2 Ester hydrolysis of 2-methoxyphenyl acetate with buffer solution of sodium hydroxide/sodium hydrocarbonate in D_2O. The lower-case letters are for later NMR assignments.

The total duration of the experiment was 5 h, with a recycle delay of 21 s, a 90° pulse width of 10.4 μs, and an acquisition time of 3.9 s for each ^{1}H spectrum. The timecourse was obtained by arraying the number of transients. The first and second block of 50 spectra each had 1 and 4 transients respectively. The rest of the 30 spectra had 16 transients each.

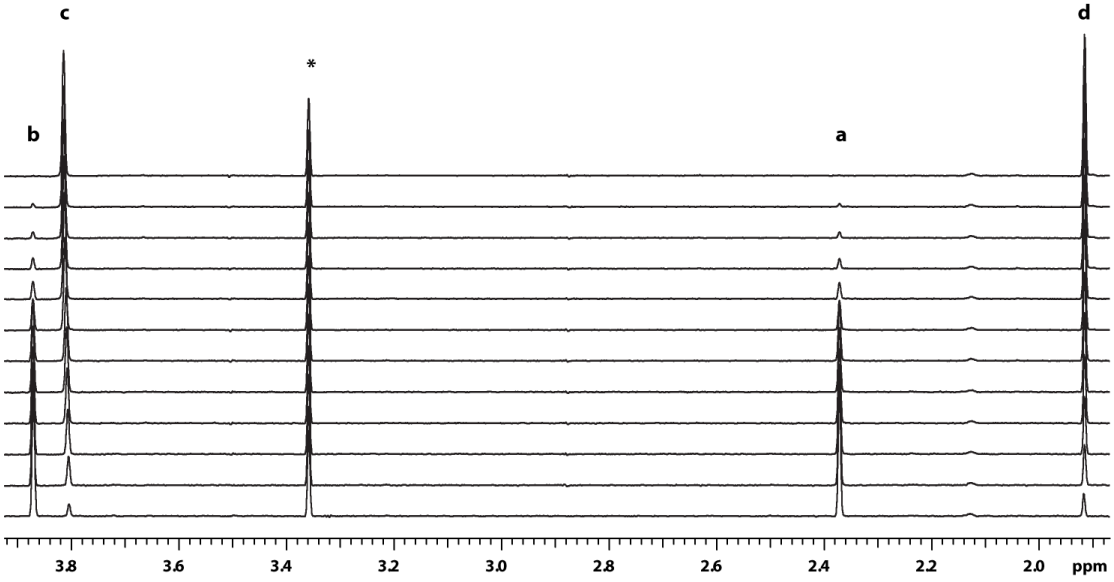


Figure 3.4 Array showing every 10th spectrum from the first (bottom) to the 120th (top). The signals from left to right are reactant methoxy (b), product methoxy (c), impurity (*), reactant acetate (a) and product acetate (d).

3.3.2.1 Data Processing

The array of ^{1}H spectra for the ester hydrolysis reaction was reference deconvoluted to give Gaussian line shapes with a linewidth of 3.9 Hz, to minimize the overlap between the methoxy signals of reactant and product. The data were baseline corrected in Vnmr 6.1C and exported to Mathematica as a text file. All statistical analyses and fitting processes were performed in Mathematica version 5.1. A program was written to fit the methoxy and acetate signals of the successive spectra in Mathematica by nonlinear least squares to the following Gaussian lineshape and linear baseline model:

$$S(f) = a_1 e^{-4\log_e 2(f-f_1)^2/w_1^2} + a_2 e^{-4\log_e 2(f-f_2)^2/w_2^2} + b_1 f + b_0 \qquad \text{(equation 3.6)}$$

where a_1, a_2, f_1, f_2, w_1, w_2 are the amplitudes, frequencies, and the widths of the two signals respectively, and b_1 is the slope and b_0 is the intercept of the baseline.

The initial parameter values for the Gaussian function were estimated from the original data and given to the program. Then during the iteration process the initial parameter values for the Gaussian function of one spectrum were obtained from fitting the preceding spectrum. This sequential process allows information on parameter changes to be used for the next fitting step where signals move. This has to be done because the algorithm is sensitive to initial values. The fitted and experimental spectral data are plotted together in figures 3.6 and 3.7.

The integral values ($a \cdot w$) for each of the peaks were obtained from the fitting process. In order to convert the integral values to concentrations, the integral of the reactant

signals at time zero is related to the initial concentration (0.03 M) used in the reaction to obtain the conversion factor. The initial integral value of the reactant was obtained by extrapolation of the fitted integrals as a function of time to the reactant ($R(t)$) formula of the following 1st order kinetic model:

$$R \xrightarrow{k} P$$

$$\frac{d[R]}{dt} = -k[R] \rightarrow R(t) = R_0 e^{-k(t+t_0)}$$
$$\frac{d[P]}{dt} = k[P] \rightarrow P(t) = (1 - e^{-k(t+t_0)})$$

(equations 3.7)

where $[R]$, $[P]$, $R(t), P(t), R_0$, *k, t, and* t_0 are the concentration of the reactant, concentration of the product, concentration of reactant as a function of time, concentration of product as a function of time, initial concentration of the reactant, rate constant, time, and the initial time respectively. As the initial experimental time is not zero (due to the delay, for sample loading, locking and shimming, between mixing the reactants and acquiring the first spectrum), the t_0 is added to the model to shift the time to the real experimental starting time. The initial integral value of the reactant (R_0) is obtained from the fitting process by nonlinear regression. The R_0 was equalled to the initial concentration value of the reactants in the reaction mixture (0.03 M), and all experimental integral values were scaled using this conversion factor. The concentration changes of the signals were plotted as a function of time (figure 3.8). These reactant and product data were fitted simultaneously to the first order kinetic model of equations 3.7.

3.4 Results and Discussion

3.4.1 Ugi-Smiles Reaction

Despite the application of different fitting methods for the Ugi-Smiles experiment, it was not possible to find the individual triplet peaks of the methyl region. One of the results (figure 3.5) suggests that there may be a broad hump under this region, which limits the fitting.

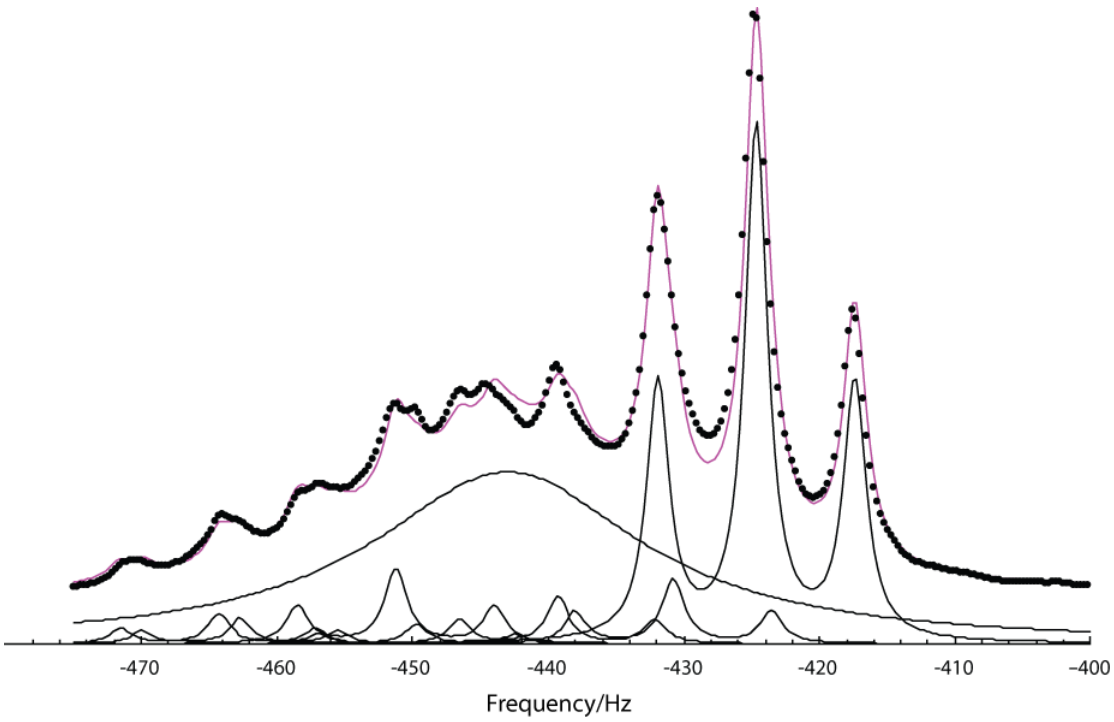


Figure 3.5 The result of fitting the overlapping signals of the methylene region (0.82-0.95 ppm in figure 3.1). The dots show the experimental data and the solid lines on them are the total fit of the overlapped region. The individual triplets and the background hump are shown at the bottom.

3.4.2 Ester Hydrolysis Reaction

The experimental and fitted reactant and product methoxy and acetate signals of the first and the last signals in figure 3.4 are shown in figures 3.6 and 3.7.

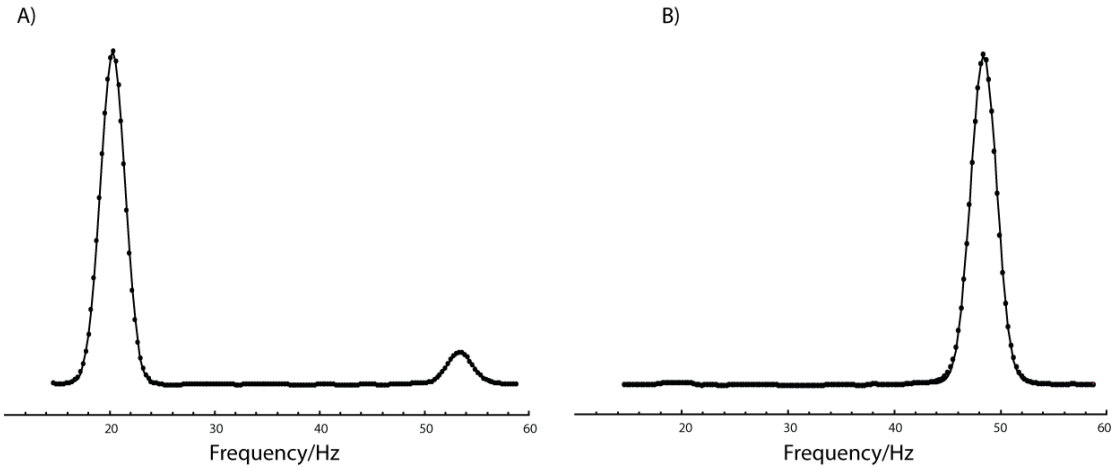


Figure 3.6 The first (A) and the last (B) experimental (dots) and fitted data (solid lines) for the methoxy region of the ester hydrolysis (figure 3.4), with arbitrary frequency scales. The reactant methoxy is at 20 Hz, while the chemical shift of the product methoxy (OMe) on the frequency scale used varies between 53 and 48 Hz during the reaction. As the reaction proceeds, the change in the ratio of Ar-OH to ArO- for the 2-methoxy phenol product causes the chemical shift of the OMe group to change.

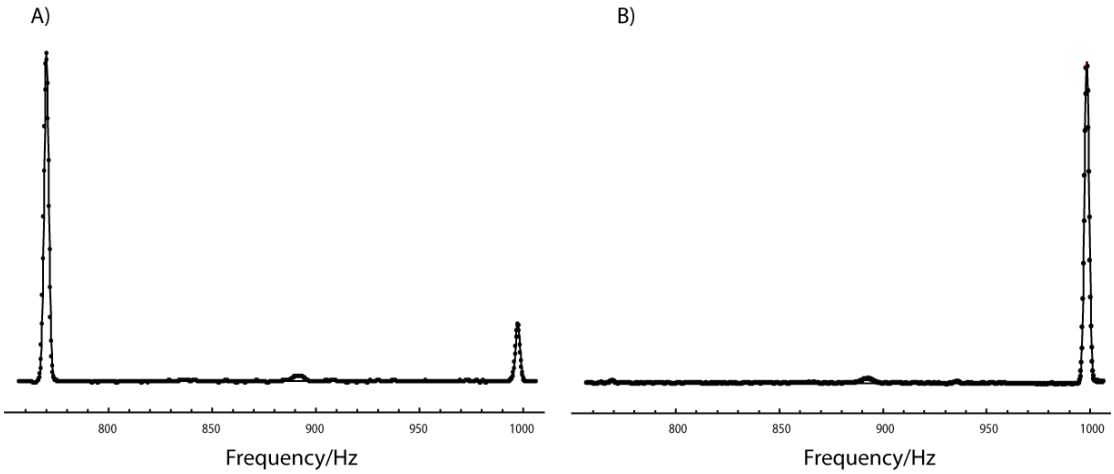


Figure 3.7 Experimental (dots) and fitted (solid line) spectra for the reactant and product acetate signals in the ester hydrolysis reaction (figure 3.4); the first spectrum is shown at the left and the last on the right. On the arbitrary frequency scale used, the reactant acetate is at 740 Hz and product is at about 1000 Hz.

The concentrations of the products and reactants obtained for the methoxy and acetate signals are plotted as a function of time in figure 3.8, together with fit to a first order kinetic model.

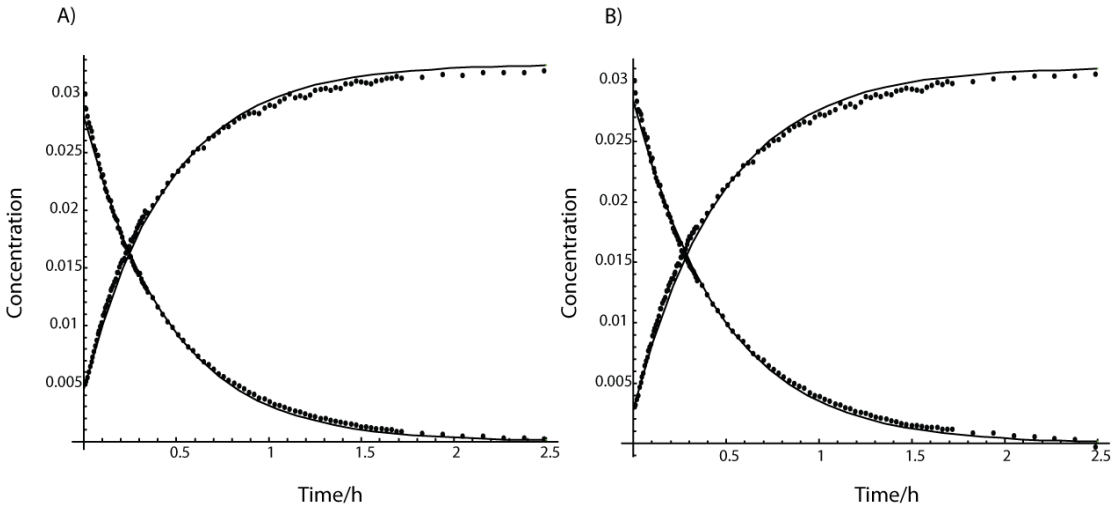


Figure 3.8 A) Kinetic plot of acetate signals. The dots represent the concentration values of the fitted signals as a function of time, which are fitted (solid lines) to a 1st order kinetic model, giving a rate constant of $6.08 \pm 0.03 \times 10^{-4}$ s^{-1}. B) Kinetic plot of methoxy signals, giving a rate constant of $5.81 \pm 0.03 \times 10^{-4}$ s^{-1}.

The results show that the kinetics of chemical reactions can be studied by applying statistical methods to an array of NMR spectra. However, analysis requires that the spectra be assigned, so that the signals of individual species can be identified. Where reactions and/or spectra are complex, for example in the case of the Ugi-Smiles rearrangement studied, analysis may not be possible with ^{1}H spectra alone and further techniques (e.g. ^{13}C, HMBC, HMQC) may be needed to resolve and assign signals of all the reaction products and intermediates.

In the ester hydrolysis reaction, the signals were fully resolved and could be identified easily, providing a simple example of a kinetic study using statistical methods to fit spectra and concentration timecourses. The ability to fit a signal with unstable chemical shift (figure 3.6) confirms the possibility of studying the kinetics of mobile NMR peaks.

A further kinetic study involving overlapped signals and more complicated kinetics, in which more sophisticated statistical methods allow the separation and identification of overlapping signals, is presented in chapter 5.

Chapter 4

Kinetic Studies Using PARAFAC Analysis

4.1 Introduction

Reaction monitoring by quantitative NMR is simplest when each component in a reaction mixture has at least one well-resolved resonance; the change in peak integral can then be used directly to determine the kinetic behaviour. When no resolved peaks are available, as is quite common, the extraction of kinetic data becomes much more challenging, and it can become extremely difficult to identify individual reaction components. Different 2D statistical methods such as PCA [4.1] (Principal Component Analysis) have been used for the analysis of complicated overlapped datasets [4.2-4]. However these analyses can have the problem of rotational ambiguity [4.5, 6]. This means that there is no unique set of spectra that can fit the experimental data, but there are arbitrary mixtures of different spectra that can fit the data just as well as the true spectra of the species. To overcome this problem it is necessary to make some assumptions about the bilinear experimental data (e.g. non-negativity constraint) prior to analysis.

As was shown in the previous chapter, the identification of signals in overlapping regions can be a difficult task. Assumptions had to be applied (e.g. number of signals, widths and frequencies), while sometimes spectra could not be satisfactorily fitted even with prior assumptions (section 3.4.1 in chapter 3). Therefore, it would be ideal to have a model-free analysis method that would not require any assumptions about the experimental data, and would avoid the rotational ambiguity problem. This can be done by adding another dimension to the 2D dataset, to produce a trilinear dataset [4.7-9]. In this way, no assumptions need to be made about the data prior to analysis, and the risk of applying constraints that might not be valid is avoided.

DOSY [4.10] (Diffusion-Ordered SpectroscopY) is a powerful method which can distinguish between the components in a mixture, if they have different diffusion behaviour. DOSY and timecourse ^{1}H experiments produce bilinear data, in which in DOSY, components can be separated according to their diffusion, and in timecourse ^{1}H spectra, the kinetics of components can be monitored as a function of time.

A new method is introduced here for the study of kinetics while simultaneously identifying the spectra and diffusion coefficients of the species involved. It relies on recording NMR spectra as a function of time and of gradient amplitude, i.e. measuring a timecourse of DOSY spectra. Such a timecourse of DOSY spectra produces a trilinear dataset for PARAFAC (PARAllel FACtor) [4.11, 12] analysis. No prior knowledge of the component spectra, diffusion behaviour or kinetics is needed; the only requirement is that the spectrum, diffusional attenuation as a function of gradient, and concentration profile of each species be independent of each other.

The initial systems studied were carbohydrate hydrolysis reactions, such as the hydrolyses of sucrose and maltose. These reactions are good candidates as they have the challenging problem of severely overlapping spectra, have convenient kinetic and concentration regime, have some resolved signals for comparison purposes, and have spectra that are insensitive to pH changes, so the problem of pH-dependent chemical shifts is avoided. Although these reactions are in principle complicated by the presence of multiple anomers, in practice the kinetics remain simple because anomerisation is rapid compared with hydrolysis under the conditions used. Later, maltotriose hydrolysis was studied to examine the capability of PARAFAC to

distinguish between components with very similar diffusion coefficients. The final reaction studied was the hydrolysis of clarithromycin, a relatively complex case.

4.2 Theory

4.2.1 PARAFAC

It is conventional to describe the datasets used for multivariate methods [4.13] in matrix format where the data are organized in rows and columns. For example, a bilinear dataset such as a timecourse array of ^{1}H spectra can be described by a set of spectra and concentration functions for the N different components:

$$\underline{\underline{D}} = \sum_{i=1}^{N} \underline{C}_i \underline{S}_i + \underline{\underline{R}} \qquad \text{(equation 4.1)}$$

where the matrix $\underline{\underline{D}}$ represents the experimental dataset, with $\underline{C}_i$ and $\underline{S}_i$ being the *i*th vectors of the matrices $\underline{\underline{C}}$ (concentration) and $\underline{\underline{S}}$ (spectrum) respectively, and $\underline{\underline{R}}$ the residuals matrix. In multivariate analysis the fitting process discovers the matrices $\underline{\underline{C}}$ and $\underline{\underline{S}}$ that best represent the experimental data ($\underline{\underline{D}}$). However, in 2D analysis the rotational ambiguity problem described earlier can be a problem; therefore, here a multivariate statistical model such as PARAFAC will be utilized for 3D analysis, in which another dimension is added to the 2D dataset.

PARAFAC (PARAllel FACtor) analysis is a multi-way analysis tool which was first implemented in psychometrics for data analysis [4.14-16]. Since then it has found extensive application in different areas such as chemometrics[4.17-19], chromatography [4.20], analytical chemistry [4.21], fluorescence and NMR spectroscopy[4.8, 22-25]. PARAFAC can be understood as an extension of PCA to more than two-dimensional arrays[4.16]. A PARAFAC model [4.26-28] can be defined as a collection of multi-linear data. In

trilinear data this model can be represented as a box ($\underline{\underline{X}}$) which includes three loading matrices ($\underline{\underline{A}}, \underline{\underline{C}}, \underline{\underline{S}}$), and the residuals ($\underline{\underline{R}}$) (figure 4.1), in which $\underline{\underline{X}}$ varies linearly with each of the loading matrices as a function of gradient (*g*), time (*t*), and frequency (*f*) for each of the i components: $X_i(g,t,f) = A_i(g)C_i(t)S_i(f)$. The trilinear PARAFAC model can be described as in equation 4.2.

$$\underline{\underline{X}} = \sum_{i=1}^{N} \underline{A}_i \otimes \underline{S}_i \otimes \underline{C}_i + \underline{\underline{R}} \qquad \text{(equation 4.2)}$$

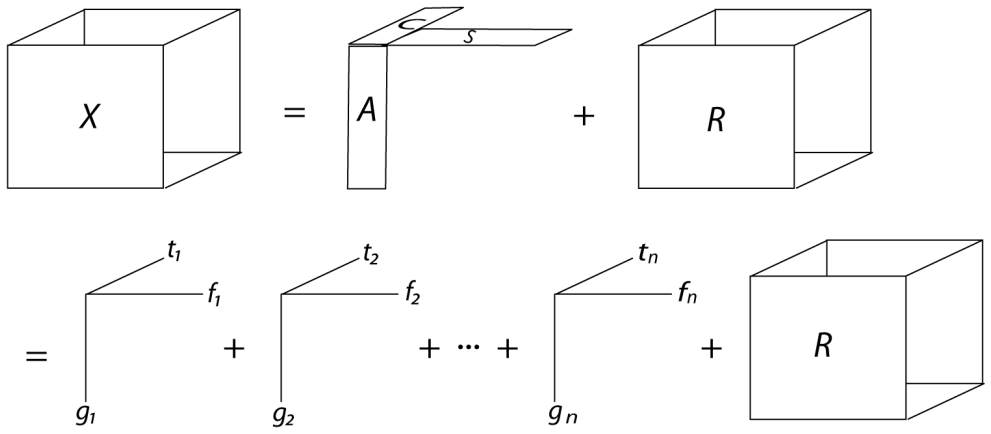


Figure 4.1 Graphical decomposition of multi-component PARAFAC model. $\underline{\underline{X}}$ is the initial trilinear array. $\underline{\underline{A}}, \underline{\underline{C}}$, and $\underline{\underline{S}}$ are the loading matrices of the three modes, and $\underline{\underline{R}}$ is the residuals. The three independent variables for different number of components are assigned as *t*, *f*, and *g*.

PARAFAC uses the Alternating Least Squares method (ALS) or algorithms such as GRAM-DTLD, ASD, PMF3, etc.[4.29, 30] to minimise the sum of squares of variation between the input and output data. In most cases this analysis begins with random values. In a trilinear analysis the ALS assumes known values for two matrices and

estimates the third matrix by least squares regression to find the best fit to the model. This iteration continues until the global minimum is reached (best fit to the model is obtained). Optimum results are obtained from PARAFAC when trilinearity holds and the correct number of components is used [4.31].

There are several ways to determine the right number of components: 1) evaluation of the residuals. This can be done by: a) evaluating the difference between the input data and the output of PARAFAC, or b) calculating the root-mean-square (RMS) of the residuals of the model. Plotting the RMS values vs. number of components shows a significant reduction of RMS for each true component added [4.31, 32]. 2) Splitting the data into half and comparing the output models. If the two models are not equal it suggests the wrong number of components has been used. If the number of components is chosen correctly, the same results should be obtained for the two halves of the data due to the uniqueness of the PARAFAC model [4.33] . 3) Compare the result with prior knowledge of the input data being modelled (e.g. in spectroscopy the output component spectra from PARAFAC can be compared with the known spectra of the pure materials) [4.33]. 4) Physical interpretation of the output, e.g. positive spectra and Gaussian decay in the diffusion dimension.

4.2.2 Diffusion and Diffusion-Ordered Spectroscopy

Molecules in solutions show different motions, such as rotation and translation. The random translational motion of particles due to their thermal energy is known as self-diffusion, and is characterized by a diffusion coefficient D (m^2 s^{-1}) [4.34]. In a homogeneous system, the probability of finding at position $\underline{r}$ a particle with self-diffusion coefficient D which was originally at position $\underline{r}_0$ after time t_d, can be defined [4.35] by a Gaussian distribution as in equation 4.3.

$$P(\underline{r}_0, \underline{r}, t_d) = (4\pi D t_d)^{-\frac{3}{2}} e^{-\frac{(\underline{r}-\underline{r}_0)^2}{4Dt_d}} \qquad \text{(equation 4.3)}$$

Einstein described the root mean square displacement of a particle by diffusion coefficient as:

$$< x >= \sqrt{nDt_d} \qquad \text{(equation 4.4)}$$

where n is 2,4, or 6 for one, two, or three dimensional diffusion respectively.

The Einstein-Smoluchowski equation (equation 4.5) defines the self-diffusion coefficient for uncharged particles in dilute solutions as follows [4.34]:

$$D = \frac{k_b T}{f} = \frac{RT}{Nf} \qquad \text{(equation 4.5)}$$

where k_b, T, f, R, and N are the Boltzmann constant, temperature, hydrodynamic frictional coefficient, gas constant ($R = Nk_b$), and Avogadro's number respectively. The hydrodynamic frictional coefficient (f) of a spherical molecule in laminar flow is defined by Stokes equation [4.34] as:

$$f = 6\pi\eta r_s \qquad \text{(equation 4.6)}$$

where η and r_s are the viscosity and hydrodynamic radius of a molecule respectively. Species with different geometries have different f. Combination of equations 4.5 and 4.6 gives the famous Stokes-Einstein equation:

$$D = \frac{k_b T}{6\pi\eta r_s} \qquad \text{(equation 4.7)}$$

Diffusion coefficients of molecules thus depend on their molecular weights. For two spherical molecules of equal density with diffusion coefficients D and molecular weights M [4.34]:

$$\frac{D_1}{D_2} = \sqrt[3]{\frac{M_2}{M_1}} \qquad \text{(equation 4.8)}$$

The measurement of self-diffusion coefficient is now possible using NMR spectroscopy. Diffusion-Ordered SpectroscopY (DOSY) is a pulsed field gradient NMR method that discriminates between the signals of different components on the basis of their diffusion behaviour. In this method an array of spectra is measured as a

function of Pulsed Field Gradient (PFG). The signals in each spectrum decay according to the strength of the gradient. Lower gradient strength causes less signal attenuation, and as the gradient strength increases the signals attenuate more. Then, each individual signal attenuation as a function of pulsed field gradient is fitted to Stejskal-Tanner equation (4.9) to obtain the apparent diffusion coefficient of the particles related to that signal.

$$S = S_0 e^{-D\gamma^2\delta^2 g^2 \Delta'} \qquad \text{(equation 4.9)}$$

where $S, S_0, D, \gamma, \delta, g, \Delta'$ are the amount of signal received after application of gradient pulse, the signal in the absence of gradient, diffusion coefficient, gyromagnetic ratio, gradient duration, gradient amplitude, and the corrected time difference between the midpoints of the two gradient pulses respectively. The diffusion coefficient values obtained from the fitting process are plotted along one axis, with the other axis being the chemical shift. Peaks are centred on the diffusion coefficients, with widths being determined by estimated errors of the fits. This 2D plot is known as a DOSY plot. Heavier molecules correspond to smaller values of diffusion coefficient, and smaller molecules show greater diffusion coefficients [4.10, 36-38].

The Larmor frequencies of spins in absence (ω_0) and presence (ω_z) of a magnetic field gradient are described in equations 4.10 and 4.11 respectively

$$\omega_0 = -\gamma B_0 \qquad \text{(equation 4.10)}$$

$$\omega_z = -\gamma B_0 - \gamma g_z z \qquad \text{(equation 4.11)}$$

where $\gamma, B_0, g_{z,} z$ are the gyromagnetic ratio, main magnetic field in the z direction, gradient in the z direction, and position of the spins in the z direction respectively [4.39].

Over the past few decades numerous DOSY sequences have been used. The simplest of these sequences, known as the Pulsed Field Gradient Spin Echo (PFGSE, figure 4.2 A) is the combination of a spin echo [4.40] sequence with pulsed field gradients. The PFGSE is made of a single 90˚ pulse, a 180˚ pulse and two equivalent magnetic field gradient pulses which are at either side of the 180˚ pulse, with length of δ. The 90˚ pulse rotates all the spin magnetization from the z direction to the xy plane. During the first delay time (τ) a gradient pulse with duration δ and magnitude g is applied to label the spins. The effect of labelling is to encode each spin with a phase proportional to its position. This causes the spins to experience the following phase shift in the laboratory frame of reference at the end of the first delay time:

$$\phi(\tau) = -\gamma B_0 \tau - \gamma g \int_0^{\delta} z(t) dt \qquad \text{(equation 4.12)}$$

The first term in equation 4.12 shows the phase shift of a spin due to the static magnetic field, and the second term shows the phase shift due to the gradient field. Every spin in a plane perpendicular to the z direction is affected by the z gradient in the same way. Therefore, the phase shift in each spin due to the gradient reflects the position of the spin. After the 180˚ pulse the sign of the shifted phase is inverted. A second gradient pulse, equivalent to the first, affects the spins in the same way as the

first gradient (decodes the spins). If there has been no diffusion during the experiment, all the spins get refocused, which leads to maximum signal intensity. The effect of the gradients on spins that have diffused is to distribute their phases, which leads to attenuation of signal intensity. The random phase shifts of particular nuclei at different positions of the sample are averaged, and determine the net amplitude of the NMR signal. Thus if a series of gradient pulses of incremented intensity is applied, the rate of attenuation of the signal can be used to find the diffusion coefficient from the Stejskal-Tanner equation (equation 4.9) [4.41, 42].

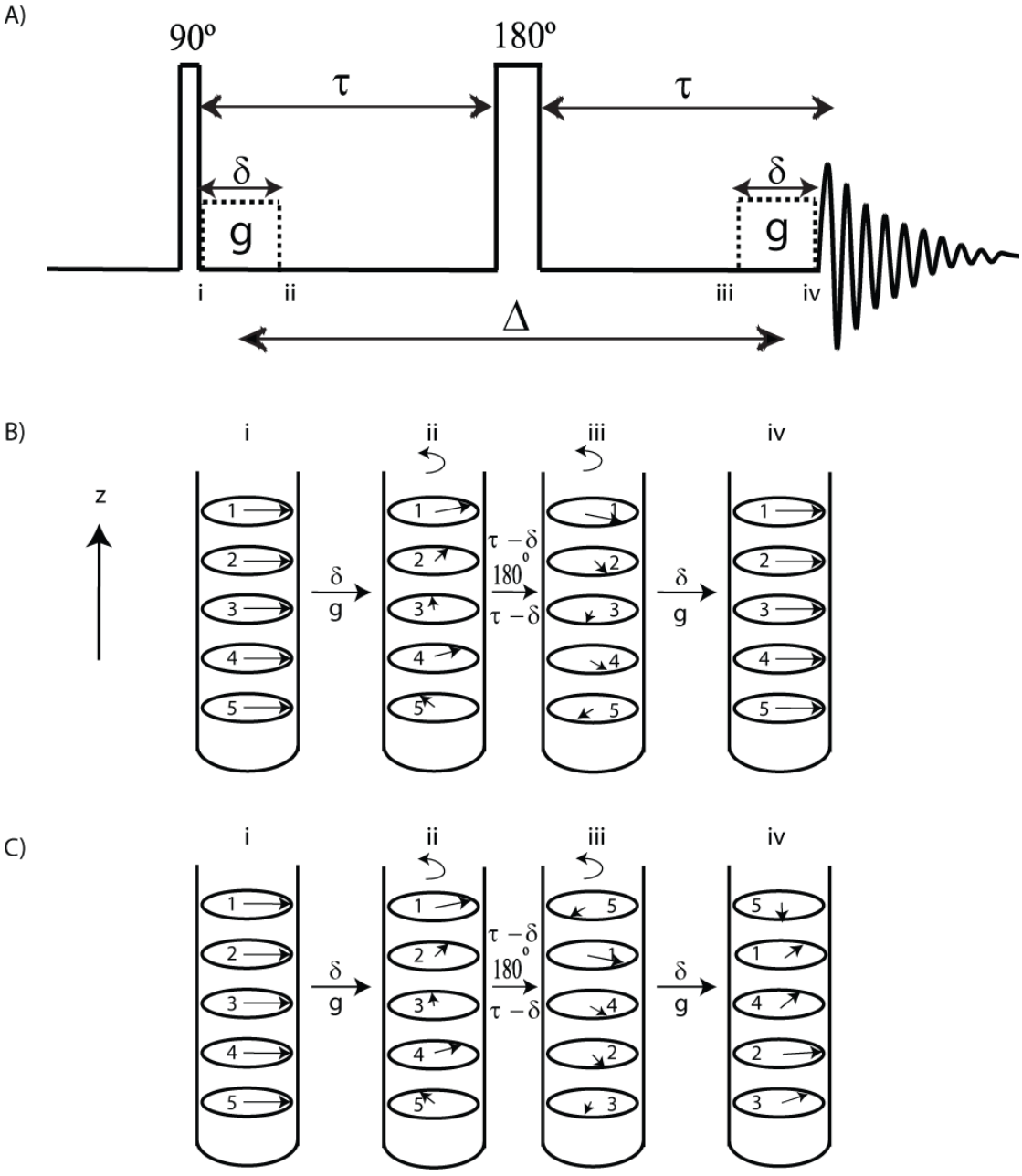


Figure 4.2 A schematic representation of a pulsed field gradient spin echo (PFGSE) experiment (A), and the behaviour (B, C) of magnetization vectors during the sequence. (B) The behaviour of spins during the PFGSE sequence if there is no diffusion. After the 90° pulse all the spins are rotated into the xy plane (i). After the first gradient pulse the spins are phase shifted (ii). During the time $\tau - \delta$ (time after the first gradient and before the 180° pulse) the spins precess (without diffusing). A 180° pulse rotates the spins into mirror image positions and they carry on evolving (iii). The second gradient pulse affects the spins in the same way as the first gradient, so the vectors are refocused and maximum signal intensity is observed (iv). (C) The behaviour of spins during the PFGSE sequence if there is diffusion in the sample. As in A, the vectors are transformed by the 90° and the first gradient pulse (i and ii). As the spins diffuse, their vertical positions change while they evolve. The 180° pulse rotates the spins into mirror image positions and they again carry on evolving and

diffusing (iii). The second gradient pulse affects each position in the same way as before, but as the spins are in different locations, the effect is different from that of the first gradient pulse. Therefore, the vectors are not completely refocused and the total signal intensity is reduced (iv). The greater the gradient amplitude and the greater the diffusion, the greater will be the signal attenuation.

Throughout the diffusion period of the PFGSE sequence the magnetization is in the transverse (xy) plane, and therefore experiences J-modulation [4.43]. This disadvantage is reduced in another pulse sequence known as the Pulsed Field Gradient STimulated Echo (PFGSTE) [4.44, 45]. In the PFGSTE sequence (figure 4.3 A) after the first 90° pulse the transverse magnetization is produced, then the first gradient pulse distributes the phases. The second 90° returns some of the transverse magnetization to the z-direction and the rest remains in the transverse plane. A gradient pulse between the last two 90° pulses (ii in figure 4.3 A) can be used to dephase permanently the remaining magnetization in the transverse plane and/or phase cycling can be used to select the required coherence transfer pathway. The final 90° pulse returns the stored z magnetization to the xy plane and the final gradient pulse decodes it.

The PFGs disturb the homogeneity of the magnetic field; this perturbs the deuterium lock signal and can thus affect the quality of the spectrum. To reduce this problem, the Bipolar Pulse Pair STimulated Echo, BPPSTE, (figure 4.3 B) combines a bipolar PFG with a 180° pulse instead of using a single gradient pulse [4.46, 47]. If the gradients and durations of the bipolar gradients (i and ii, iii and iv in figure 4.3 B) are the same, then the effect of the combination of a 180° and gradient pulse ii/iv (in figure 4.3 B) on spins is the same as the effect of a single i/iii pulse (in figure 4.3 B). So by applying the bipolar gradient pulses and the 180°, proton spins are being affected in the same way as in PFGSTE, but the effects in deuterium are refocused. The ^{1}H 180° pulse does

not affect ^{2}H, and the effect of the gradient pulse before the 180° on the lock is cancelled by the gradient pulse after the 180°.

The Oneshot sequence [4.48] is a further DOSY sequence with more improvements. The first bipolar pulse pair (i an ii gradient pulses in figure 4.3 C) are optional pulses, which decrease in strength as the strength of the other gradient pulses increases. This keeps a constant net heat produced by the gradient coils. In this sequence the bipolar PFGs iv, v, vi and vii, ix, x (figure 4.3 C) are unbalanced in the ratio of $1+\alpha:1-\alpha$ to dephase any magnetizations that are not refocused by the 180° pulse, therefore the phase cycling process is minimized. The pulse vii dephases unwanted coherences that have non-zero order during the polarisation storage delay.

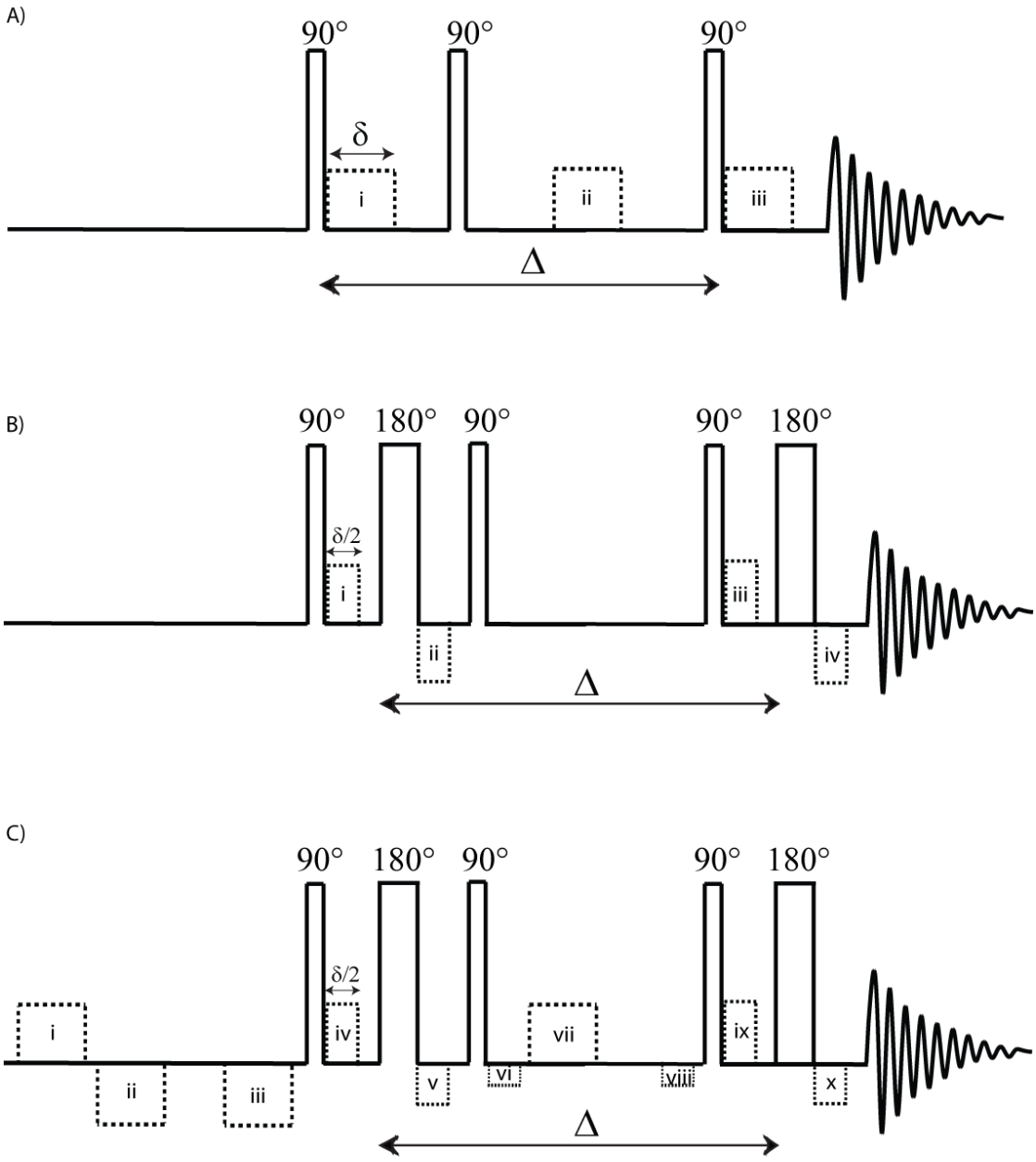


Figure 4.3 Schematic graph of PFGSTE sequence (A), BPPSTE sequence (B), and Oneshot sequence (C).

4.3 Experimental Work and Data Processing

Kinetic studies using PARAFAC were carried out for hydrolysis of the sugars sucrose, maltose & maltotriose, and the macrolide antibiotic clarithromycin.

All of the hydrolysis experiments were recorded on a 400 MHz Varian spectrometer, using a 5 mm diameter indirect detection probe equipped with a z-gradient coil allowing gradient pulses up to 30 G cm^{-1}. Prior to acquisition, the magnetic field was locked, and shimming was carried out to obtain a homogenous magnetic field. All experiments were performed without sample spinning. A timecourse series of DOSY experiments using the Oneshot sequence was acquired to obtain the NMR spectra as a function of time and gradient amplitude.

For each DOSY experiment, 6 gradient levels in equal steps of gradient squared, ranging from 3.0 to 27.3 G cm^{-1}, were used, to obtain a maximum signal attenuation of approximately 70 % (figure 4.4). A total diffusion-encoding gradient pulse duration δ of 3 ms, diffusion delay Δ of 100 ms, and a 90° pulse width of 9 μs were used. A normal 5 mm outside diameter tube (4.2 mm inside diameter) was used for the first experiment, whereas for the rest of the experiments a thick-walled NMR tube (outside diameter 5 mm, inside diameter 2.2 mm) was utilized to prevent convection.

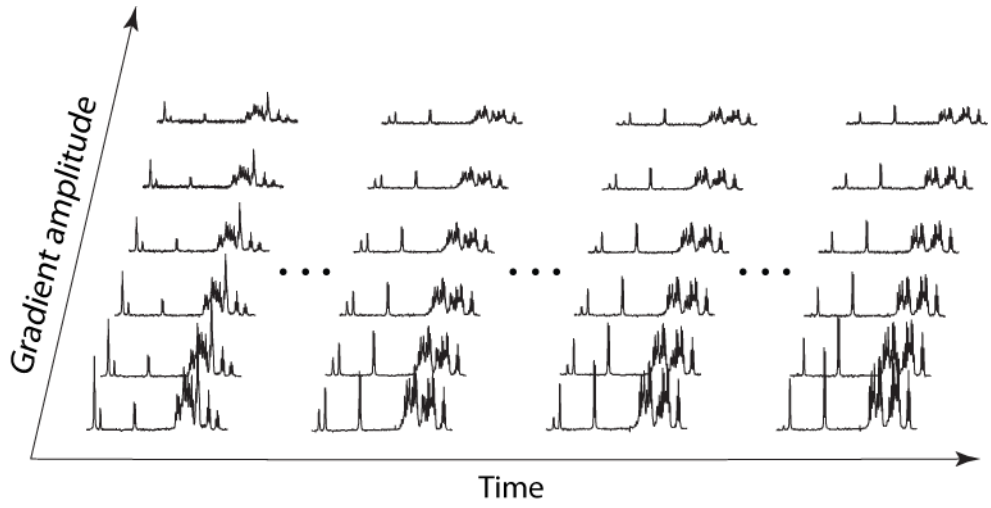
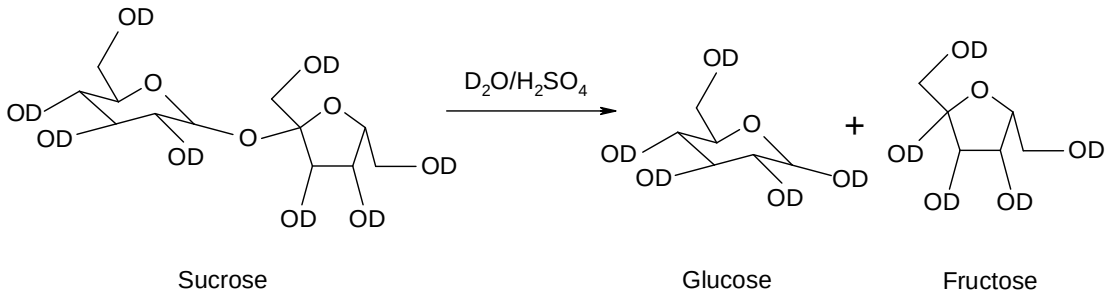


Figure 4.4 A subset of the raw experimental data, showing the attenuation of the signals for 6 gradient levels in one dimension, and the evolution of the signals during the timecourse of the experiment in the other dimension.

4.3.1 Sucrose Hydrolysis

Initially, the acid hydrolysis of sucrose [4.49] to glucose and fructose (scheme 4.1) was studied by mixing 900 µL of 0.6 M sucrose in D_2O, 22 mg of concentrated sulphuric acid (18 M), with 100 µL of 1.4 M tert-butanol solution in D_2O as reference. This reaction was carried out at 25 °C for a total duration of 6 h, with 16 transients for each of 46 DOSY experiments and a 3.6 s recovery delay. All spectra were Fourier transformed, reference deconvoluted [4.50], using 2.2 Hz line broadening, and baseline corrected in the spectrometer software Vnmr 6.1C. Data were then exported to Matlab [4.51] as a text file for PARAFAC analysis with the Matlab N-way toolbox [4.26, 33]. Small variations in receiver sensitivity over the course of the experiment were corrected for by normalising the integral of the spectrum for each gradient level using the average area of the tert-butanol reference peak.



Scheme 4.1 Sucrose hydrolysis.

PARAFAC processing with Matlab requires the 3D data (as a function of gradient, time, and frequency) to be loaded, the number of components specified, and any constraints applied. No prior knowledge of diffusion behaviour, kinetics, or component spectra is needed. All the 6 gradient levels, 46 spectra, the frequency region between 3.2 ppm and 5.7 ppm, and two components were used as the input data; the water region between 4.95 ppm and 5.35 ppm was discarded from the analysis.

The analysis yields statistical components for the diffusional attenuation $A_i(g)$, concentration evolution $C_i(t)$, and spectral mode $S_i(f)$ (figure 4.5), and here accounted for 99.9% of the observed data variation. In figure 4.5 the diffusion mode was fitted to the Stejskal-Tanner equation (equation 4.9), and the evolution mode was fitted to first order kinetics (shown in figure 4.6, presented in the Results section 4.4.1), both by least squares fitting.

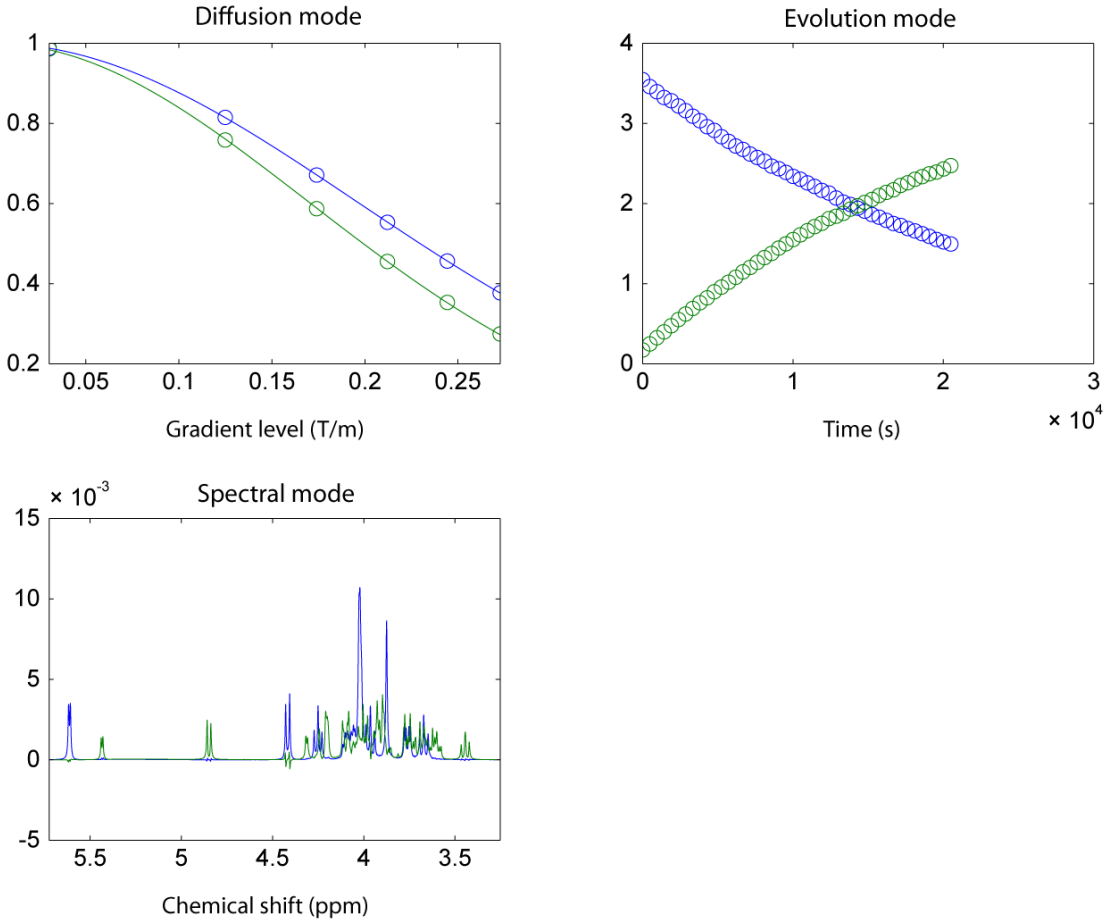


Figure 4.5 Normalised PARAFAC output for concentrated sucrose hydrolysis (0.6 M). In each of the modes the sucrose is shown in blue and the mixture of glucose and fructose is in green. In the diffusion and evolution modes the dots show the gradient levels and concentration data points respectively. The diffusion mode fits to the Stejskal-Tanner equation are shown as solid lines.

Each diffusion mode $A_i(g)$ is normalised so that it extrapolates back to unity at zero gradient, and each spectral mode $S_i(f)$ is normalised to have an integral proportional to the number of protons involved. The remaining mode (evolution mode) is multiplied by the normalising factors by which $S_i(f)$ and $A_i(g)$ were divided, to give a result which is proportional to concentration.

^{1}H spectra of a 0.6 M sucrose solution in D_2O, and of an equimolar 0.6 M solutions of glucose and fructose, were acquired separately to compare with the PARAFAC spectra obtained, as a validation test for PARAFAC (figure 4.7, in section 4.4.1).

To investigate the sources of disagreement between the PARAFAC and experimental spectra, residue spectra (the difference between PARAFAC and experimental spectra) were produced and analysed. For this purpose the 1D PARAFAC spectra at given time points and gradient levels were obtained by summing the product of the 3 PARAFAC outputs $(\sum_{i=1}^{N} A_i(g)C_i(t)S_i(f))$ for each component (N being the number of components). The difference was calculated for every 5th spectrum at the first gradient level and is shown in figure 4.8 (in section 4.4.1). The amplitudes of the residuals: $R(f) = S^{PARAFAC}(f) - S^{Raw}(f)$ for a given value of gradient and time are plotted as a percentage error of the average of the maximum points in $S^{PARAFAC}(f)$ (spectrum obtained by PARAFAC), and $S^{Raw}(f)$ (experimental spectrum):

$$\varepsilon(f) = \frac{2R(f)}{\max(S^{PARAFAC}(f)) + \max(S^{Raw}(f))} \times 100\% \qquad \text{(equation 4.13)}$$

The result of the 0.6 M sucrose hydrolysis showed some composition-dependent chemical shifts, in which a few out of phase signals are observed due to intermolecular interactions favoured by the high concentration. Therefore, the sucrose hydrolysis reaction was monitored for a second time using a lower concentration. 370 µL of 0.06 M sucrose solution in D_2O was mixed with 18 mg concentrated sulphuric acid (18 M), using 30 µL of 1.4 M tert-butanol solution in D_2O as the reference. The

reaction was carried out at 25 °C for a total duration of 16 h with 16 transients and a recovery delay of 4 s for each of 110 DOSY datasets.

These data were processed in the same manner as for the previous sucrose hydrolysis, and exported to Matlab for PARAFAC data processing. All of the 110 DOSY spectra, using gradient levels 2 to 6 (excluded the first gradient level as the intense water signal had disturbed the baseline), between 3 ppm and 5.8 ppm (with the water signal excluded as before) were used for PARAFAC data processing, again yielding 99.9 % of data variation accounted for.

As the relative amplitude of the residuals (figure 4.10) was significantly reduced by reducing the concentration of the reactant, the experiment was repeated, further reducing the concentration of the sucrose solution by a factor of 10. 370 µL of 0.006 M sucrose solution in D_2O was mixed with 30 mg concentrated sulphuric acid and 30 µL of 1.4 M tert-butanol solution in D_2O used as the reference. This reaction was carried out at 25 °C for a total duration of 12 h with 16 transients, and a recovery delay of 4 s for each of 80 DOSY datasets.

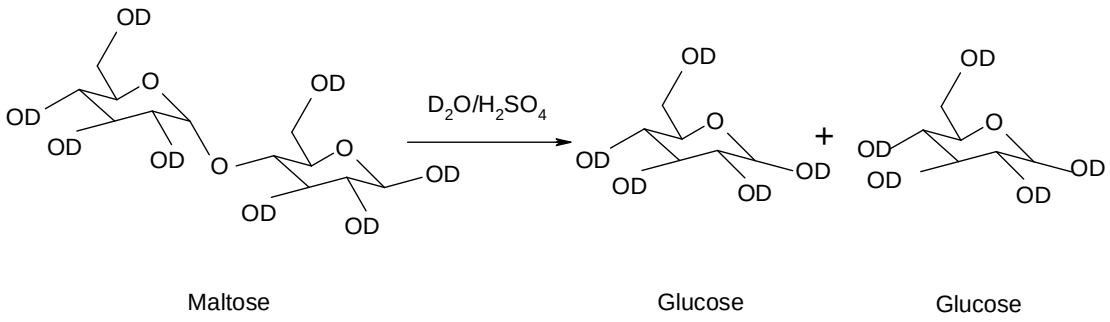
4.3.1.1 Application of Non-Negativity Constraint

In the series of sucrose hydrolysis reactions with concentrations 0.6 M, 0.06 M and 0.006 M, PARAFAC was able to produce high quality spectra although the signal to noise ratio (S/N or SNR) was different for each dataset. This raised the question of how sensitive PARAFAC is to the S/N of the experimental data. Noisy experimental data were simulated by adding different amplitudes of Gaussian white noise to the 0.06 M sucrose dataset.

The success of PARAFAC handling data with simulated low SNR led to a test of its capabilities using real experimental data with low SNR. 400 µL of 1.1 mM sucrose/D_2O/tert-butanol solution was mixed with 24 mg concentrated sulphuric acid in a thick-walled NMR tube. This reaction was carried out at 25 °C for a total duration of 15 h with 4 transients, and a recovery delay of 4 s for each of 350 DOSY datasets. This experiment was repeated with a lower receiver gain to produce noisier data. The data were processed in the same manner as for previous experiments.

4.3.2 Maltose Hydrolysis

The second reaction studied was the acid hydrolysis of maltose to glucose [4.52], in which one maltose is hydrolysed to two glucose units (scheme 4.2). For this purpose 900 µL of 0.27 M maltose in D_2O was mixed with 505 mg of concentrated sulphuric acid (18 M), with 100 µL of 0.23 M pivalic acid in D_2O as reference. Hydrolysis was carried out at 50 °C for a total duration of 41 h 49 min, with 32 transients for each of 98 DOSY experiments and a 6.5 s recovery delay. These data were processed in the same manner as for the sucrose hydrolysis (reference deconvoluted using 2.5 Hz line broadening) and exported to Matlab for PARAFAC data processing. All of the 98 DOSY spectra (figure 4.4), 6 different gradient levels, the frequency region between 3.1 ppm and 5.5 ppm (with the water signal excluded as before), and two components were used for PARAFAC data processing. The resultant fit accounted for 99.8% of the variance in the data, yielding statistical components for the diffusional-attenuation $A_i(g)$, concentration evolution $C_i(t)$, and spectral mode $S_i(f)$. The normalization of each mode was done in the same manner as for the sucrose hydrolysis.

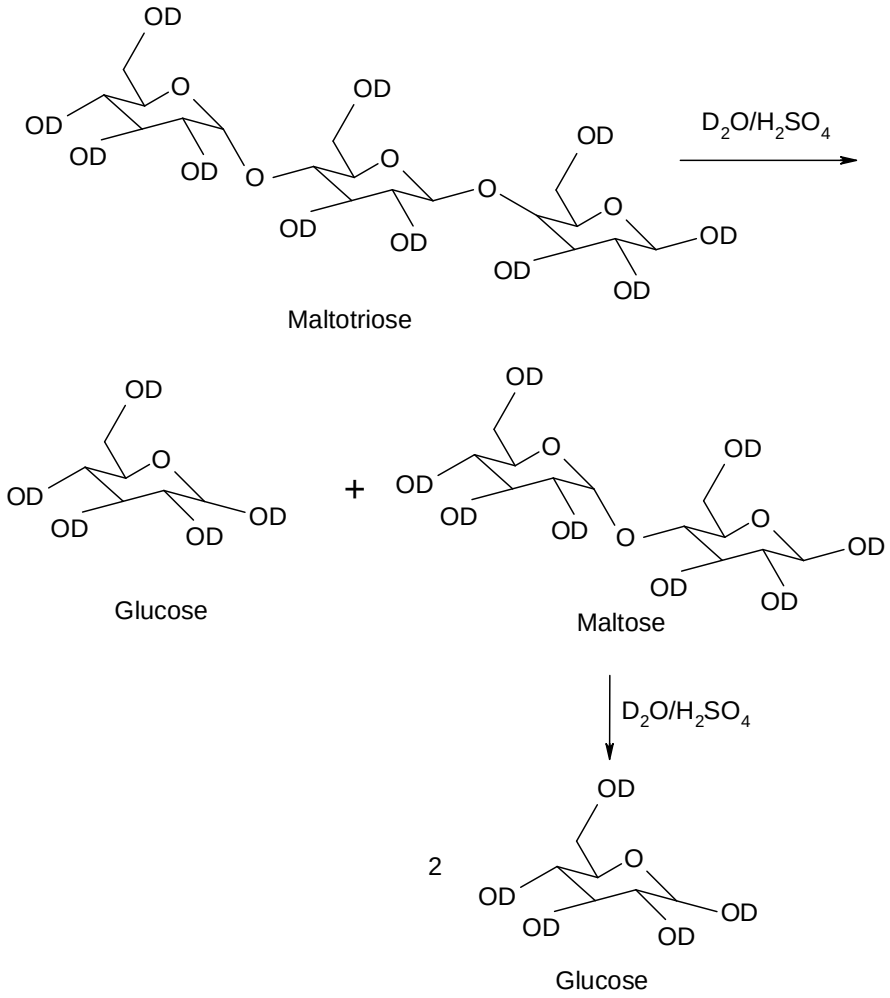


Scheme 4.2 Maltose hydrolysis.

^{1}H spectra of a 0.27 M maltose solution in D_2O, and of 0.27 M solution of glucose in D_2O, were acquired separately to compare with the PARAFAC spectra obtained, as a validation test for PARAFAC (figure 4.18).

4.3.3 Maltotriose Hydrolysis

The third reaction studied was the acid hydrolysis of maltotriose[4.53], in which one maltotriose is hydrolysed to one unit of maltose and one unit of glucose, and then the maltose is hydrolysed further to 2 units of glucose (scheme 4.3). This reaction was chosen as the ability of PARAFAC to separate 3 components, where an intermediate (maltose) is produced and then dies away, could be evaluated. It was also a challenging experiment as two of the components, maltose and maltotriose, have very similar diffusion coefficients (because of similar hydrodynamic radii) and similar spectra.



Scheme 4.3 Maltotriose hydrolysis.

For this purpose 320 µL of 0.02 M maltotriose in D_2O was mixed with 133 mg of concentrated sulphuric acid (18 M), with 40 µL of 0.23 M pivalic acid in D_2O as reference. Hydrolysis was carried out at 50 °C for a total duration of 52 h, with 32 transients for each of 130 DOSY experiments and a 6 s recovery delay, a total diffusion-encoding gradient pulse duration δ of 2 ms, diffusion delay Δ of 200 ms and 6 gradient levels. These data were processed in the same manner as the previous hydrolyses (reference deconvoluted using 2.2 Hz line broadening), and exported to Matlab for PARAFAC data processing. The frequency region between 3.0 ppm and 5.7 ppm (with the water signal excluded as before) and three components were used

for PARAFAC data processing. However, this analysis did not decompose the data successfully. It showed spectra with negative signals, and regions which did not match the reference spectra of maltotriose, maltose, and glucose. As the diffusion coefficients and spectra of 2 of the components (maltose and maltotriose) are relatively similar, a realistic constraint had to be applied to be able to decompose the data. For this purpose, non-negativity [4.33, 54] was used. Applying the constraint to all the three modes yielded realistic outputs for the three components. The best results were obtained when only the first two gradient levels were used for analysis. Further investigation showed that this was because of 2 °C temperature drift during the experiment. Decomposition while using all of the 130 DOSY spectra, the first 2 gradient levels, and applying non-negativity constraint accounted for 99.9% of the variance in the data.

Maltotriose hydrolysis experiments with lower concentrations were not successful, showing the sensitivity of non-negativity constraint to lower signal-to-noise ratio. Because there are just a few regions in the spectra where there is signal from only one component. Any noise perturbs the analysis: while in the overlapping region it can only change the apparent intensities of signals, in the non-overlapping regions, it can mask negative signals and therefore stop non-negativity constraint from discriminating between different possible decompositions of the data.

The reference spectra of glucose and maltotriose, for comparison with the output spectra of PARAFAC as a validation test, were obtained respectively from the last and the first spectra of the maltotriose hydrolysis. The comparison of maltose spectrum was done by acquiring a ^{1}H spectrum of a mixture of 320 µL of 0.02 M maltose

solution in D_2O mixed with 133 mg of concentrated sulphuric acid and 40 µL of 0.23 M pivalic acid in D_2O (the same condition as for the hydrolysis of maltotriose). As the maltose hydrolysis reaction is rapid under this condition, the reference spectrum was obtained by back-extrapolation to the required spectrum (figure 4.20).

The kinetic plot obtained by PARAFAC was fitted by non-linear least squares to the analytical solution (found using Matlab) of the coupled differential equations 4.14, which describe the successive first order kinetics of maltotriose hydrolysis.

$$ma3 \xrightarrow{k1} m + g$$
$$m \xrightarrow{k2} 2g$$

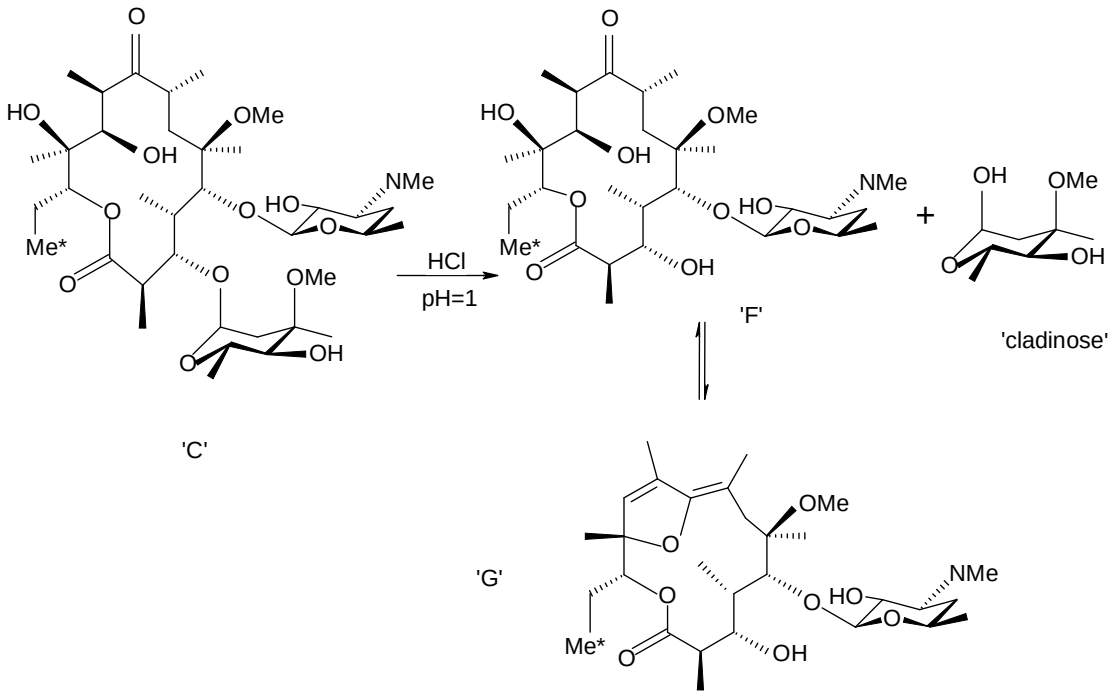
$$\frac{d[ma3]}{dt} = -k1[ma3]$$
$$\frac{d[m]}{dt} = k1[ma3] - k2[m] \qquad \text{(equation 4.14)}$$
$$\frac{d[g]}{dt} = k1[ma3] + 2k2[m]$$

where *ma3*, *m*, and *g* represent maltotriose, maltose, and glucose respectively.

4.3.4 Clarithromycin Hydrolysis

The last reaction studied was the acid-catalyzed degradation of clarithromycin [4.55, 56], an antibacterial molecule, where clarithromycin (C in scheme 4.4) is decomposed to 5-*O*-desosaminyl-6-*O*-methylerythronolide A (F in scheme 4.4) and cladinose, and later F equilibrates with G (in scheme 4.4). The purpose was to apply PARAFAC to the DOSY timecourse data of this reaction to try to decompose C, F, G, and cladinose.

300 µL of 0.1 M HCl in D_2O (pH~1), was mixed with 1.2 mg of clarithromycin and 100 µL of 0.24 M TSP in D_2O as reference. Reaction was carried out at 25 °C for a total duration of 85 h, with 4 transients for each of the first 60 DOSY experiments, where the reaction was faster, and 64 transients for the rest of the 120 DOSYdatasets, and a 6.37 s recovery delay, a total diffusion-encoding gradient pulse duration δ of 3 ms, diffusion delay Δ of 100 ms and 6 gradient levels. These data were processed in the same manner as the previous hydrolysis experiments (reference deconvoluted using 1.5 Hz line broadening) and exported to Matlab for PARAFAC data processing.



Scheme 4.4 Clarithromycin (C) acid-catalyzed degradation by HCl aqueous solution to F, cladinose, and G.

The frequency region between 0.78 ppm and 0.9 ppm, which corresponds to methyl (Me* in scheme 4.4), and 3 components for C, F, and G were used for PARAFAC data processing. Also, the whole region 0.78 ppm to 4 ppm was used with 4 components to try to decompose C, F, cladinose and G.

4.4 Results and Discussion

4.4.1 Sucrose Hydrolysis

The great advantage of PARAFAC fitting is that if the assumption of trilinearity holds, the fitted components obtained should have physical relevance, i.e. for these data should be the proton spectrum, diffusion coefficient of the species, and reaction kinetics. Therefore the quality of the PARAFAC decomposition was assessed by:

- Comparing the kinetics of the integrals of the anomeric signals from the experimental data, from regions with no overlap, with the kinetics obtained by PARAFAC.
- Comparing the proton spectra of sucrose and the product mixture (glucose/fructose) obtained by PARAFAC with the spectra of the pure materials.
- Comparing the diffusion coefficients of sucrose and glucose/fructose with the diffusion coefficients obtained by PARAFAC.

The normalised kinetic plot obtained from PARAFAC was fitted to first order kinetics by non-linear least squares fitting, yielding a rate constant of $k = 4.39 \pm 0.02 \times 10^{-5}$ s^{-1} (figure 4.6 A). This result was assessed by comparing it with the kinetics obtained from the integrals of resolved signals. The integrals of the anomeric sucrose (5.53 ppm) and glucose/fructose (alpha 5.47 ppm and beta at 4.9 ppm) protons were calculated by summing the amplitudes of the relevant data points of the signals for a given gradient level (gradient level 2 was used). The integral plot was fitted to first

order kinetics by non-linear least squares fitting and yielded a rate constant of $k = 4.38 \pm 0.03 \times 10^{-5}\ s^{-1}$ (figure 4.6 B), confirming the quality of the kinetic data obtained by PARAFAC.

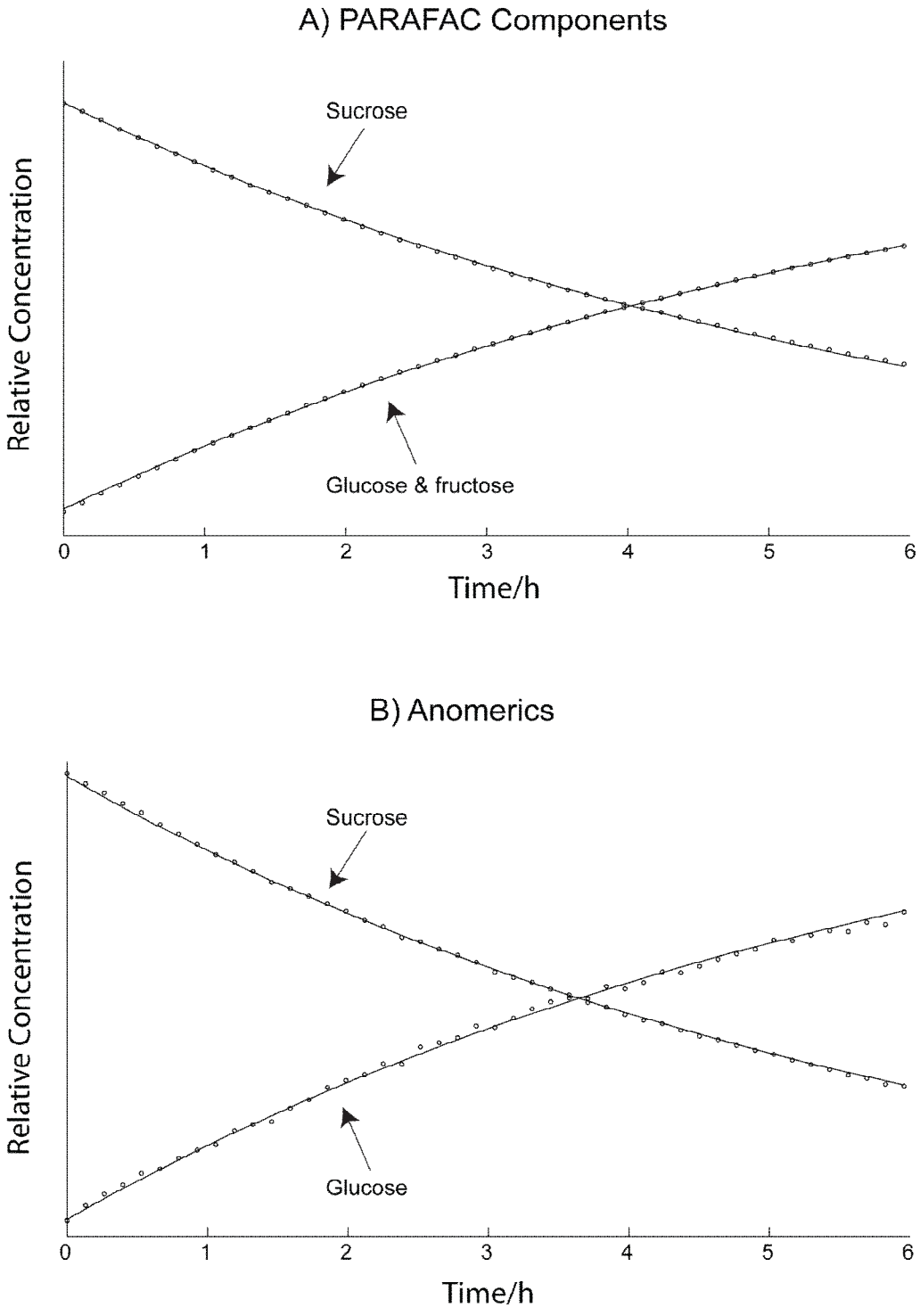


Figure 4.6 Calculated relative concentrations of reactants (downward curve) and products (rising curve) obtained by PARAFAC (A) and by integration of the anomeric signals (B) together with non-linear least squares fits (solid lines) to first order kinetics, for the acid hydrolysis of 0.6 M sucrose to glucose and fructose. A) Normalised PARAFAC components $C_i(t)$; fitting yielded a rate constant k = 4.39 ± 0.02 × 10^{-5} s^{-1}. B) Relative concentrations of sucrose and of glucose/fructose obtained by integration of the anomeric signals of sucrose (5.53 ppm) and of glucose (alpha at 5.47 ppm and beta at 4.9 ppm). Fitting to first order kinetics yielded k = 4.38 ± 0.03 × 10^{-5} s^{-1}.

In the PARAFAC analysis, the two glucose and fructose products are equivalent as one component. These two molecules are degenerate in diffusion coefficient and concentration throughout the timecourse of the reaction, so PARAFAC is not able to distinguish them.

As is shown in figure 4.7, the PARAFAC components spectra are almost, but not quite, identical to the spectra of pure sucrose and glucose/fructose mixture. There are a few cross-talk signals (marked with numbered arrows) in the PARAFAC spectra that do not match the spectrum of the pure material.

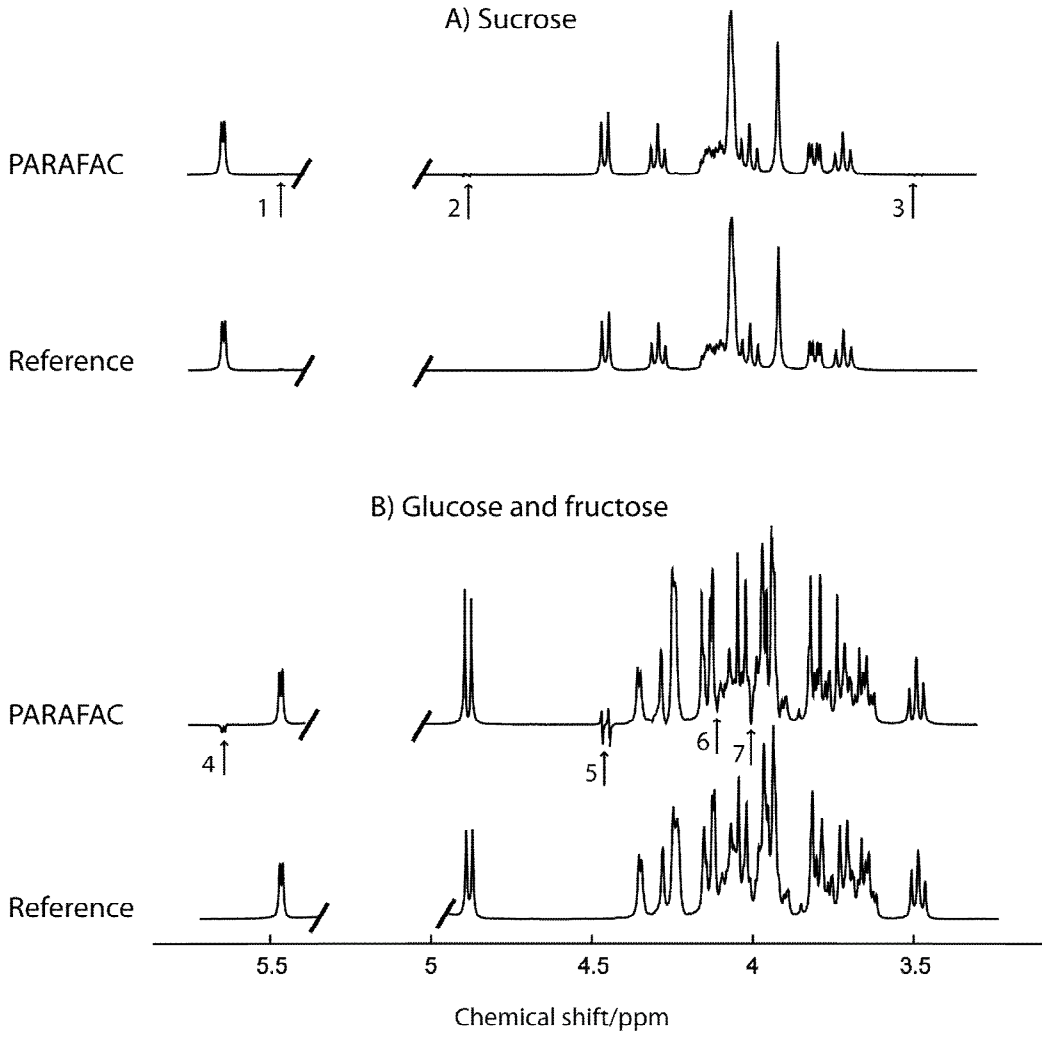


Figure 4.7 Spectra for 0.6 M sucrose hydrolysis. A) Comparison of the sucrose spectrum obtained by PARAFAC and the reference spectrum for the pure material. B) Comparison of mixed spectrum for glucose and fructose obtained by PARAFAC and the experimental (reference) spectrum of an equivalent mixture of the two sugars. The two anomeric signals at 5.47 ppm and 4.9 ppm are the signals of the alpha and beta anomeric protons of glucose respectively. The gap in the spectra is where the water signal appears; this region was discarded from PARAFAC analysis. The numbered arrows show the key differences between the PARAFAC spectra and the pure material spectra.

The diffusion mode of the PARAFAC output fitted to the Stejskal-Tanner equation (in figure 4.5) shows Gaussian decay of the diffusional attenuation of sucrose and glucose/fructose, with diffusion coefficients of $D = 2.09 \pm 0.02 \times 10^{-10}$ m^2 s^{-1} for sucrose and of $D = 2.76 \pm 0.01 \times 10^{-10}$ m^2 s^{-1} for glucose/fructose. The quality of the diffusion mode of the PARAFAC output was tested by comparing the above diffusion coefficients with the experimental diffusion coefficients. The experimental diffusion coefficients obtained for pure sucrose $D = 2.08 \pm 0.01 \times 10^{-10}$ m^2 s^{-1} and glucose/fructose $D = 2.77 \pm 0.08 \times 10^{-10}$ m^2 s^{-1} show excellent agreement.

The residuals obtained from the high concentration sucrose hydrolysis datasets (figure 4.8) show a significant difference between the PARAFAC and raw spectra not only at the position arrowed in figure 4.7, where the cross-talk in one spectrum originates from a signal at the same chemical shift in the other, but also in the rest of the chemical shifts.

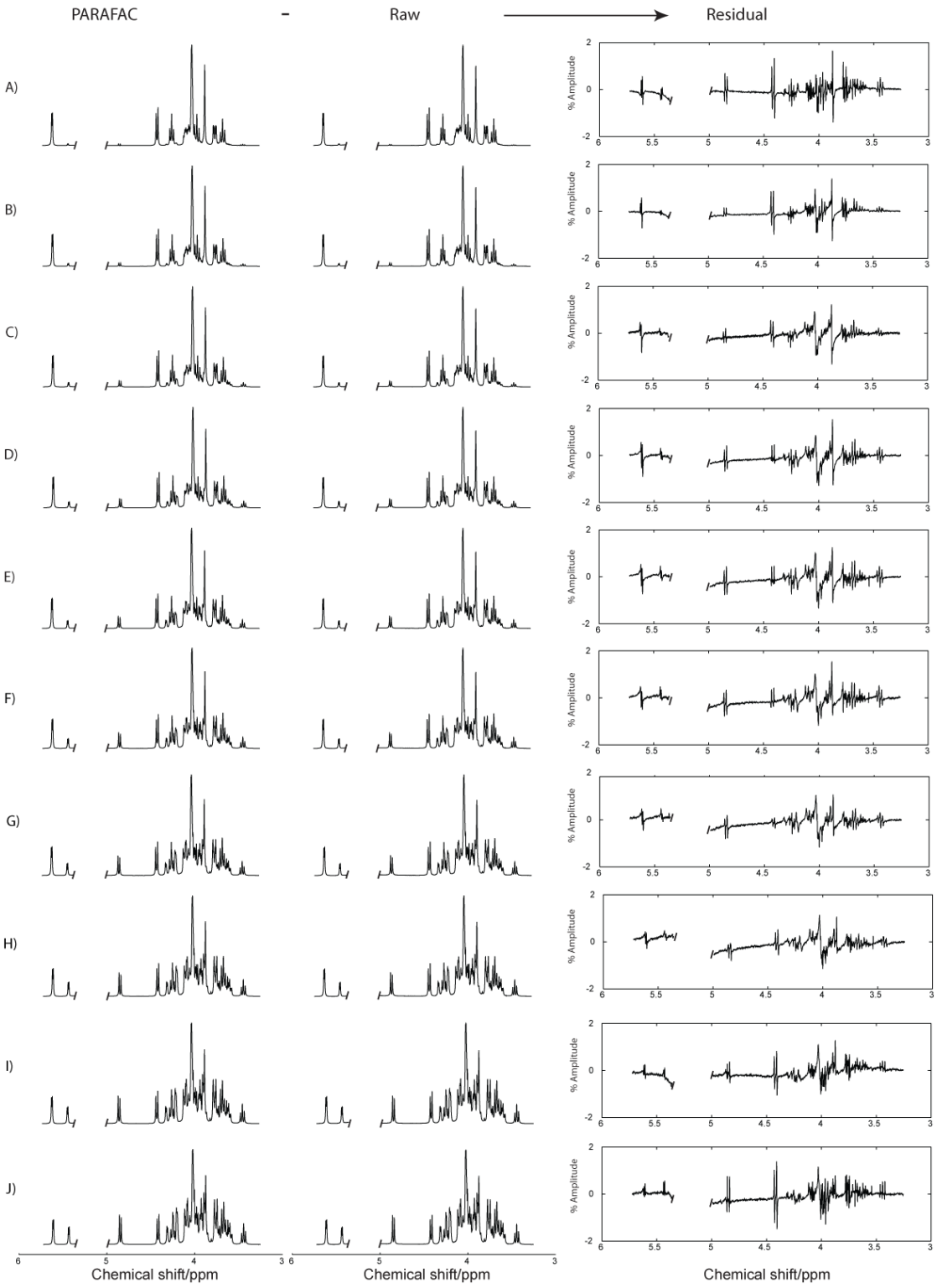


Figure 4.8 PARAFAC reconstruction and experimental spectra, with their percentage difference (PARAFAC reconstructed spectra - experimental spectra) for the first gradient level of the 1st, 5th 10th, 15th, 20th, 25th, 30th, 35th, 40th, and 45th spectra of acid hydrolysis of high concentration sucrose (0.6 M) are shown in A to J respectively.

The frequency, line width, and amplitude variation of each of the peaks showing cross-talk were analysed over the whole timecourse of the experiment. The results revealed frequency variation during the time course of the experiment in these regions. A lower concentration sucrose hydrolysis (0.06 M) reaction was therefore performed. From figure 4.9 it is clear that the amplitudes of the cross-talk peaks have been reduced in comparison with the spectra obtained for high concentration sucrose hydrolysis (figure 4.7).

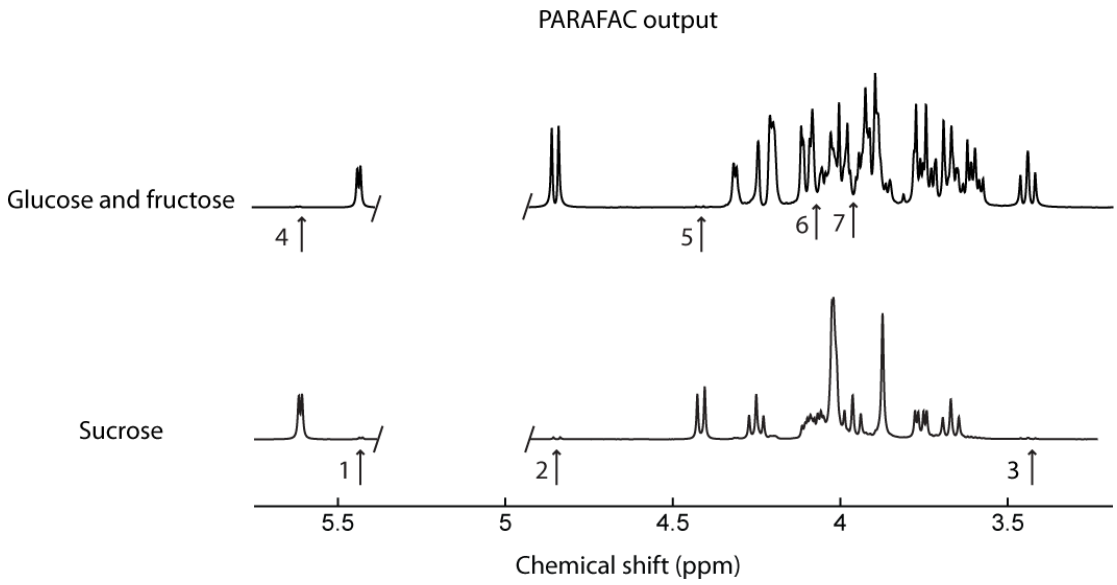


Figure 4.9 Spectra obtained by PARAFAC for less concentrated sucrose hydrolysis (0.06 M). The arrows point to the regions showing cross-talk in figure 4.7; cross-talk is greatly reduced but still just visible for most.

A series of residuals for every 10^{th} spectrum for the first gradient level used (gradient 2) of the lower concentration dataset is shown in figure 4.10. These residuals are very small, with noise being dominant. However, a curved baseline is seen in the residuals at some points (between 3 ppm to 4 ppm in B, C, D, E, G and I plots in figure 4.10).

This could be due to magnetic field disturbance (ΔB), which can be caused by the slight gradient field disturbance at an early stage of the FID.

In order to investigate further the origin of the cross-talk, the amplitude of the residual for each cross-talk region (arrowed numbers in figure 4.9) was evaluated. Amplitudes of the residual at 4.8 ppm, corresponding to cross-talk arrowed 2 in figure 4.9, are shown as a function of gradient level and time in figure 4.11. The lack of features in the residual shows the lack of any systematic drift related to the concentration variation during the time course.

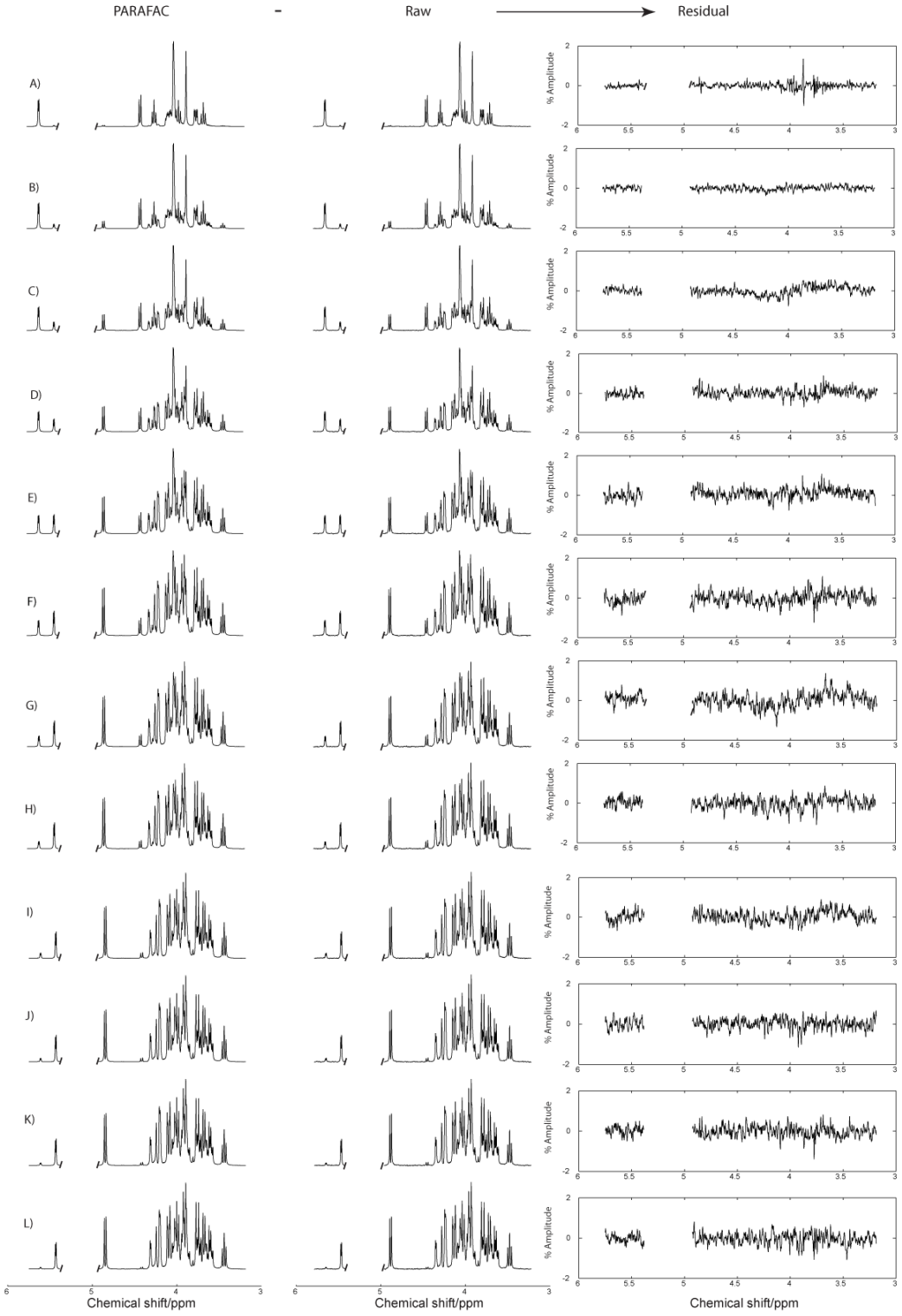


Figure 4.10 PARAFAC reconstruction and experimental spectra, with their difference for the first gradient level of every 10th spectrum from spectrum 1 to 110 acid hydrolysis of the lower concentration sucrose (0.06 M) are shown from A to L respectively.

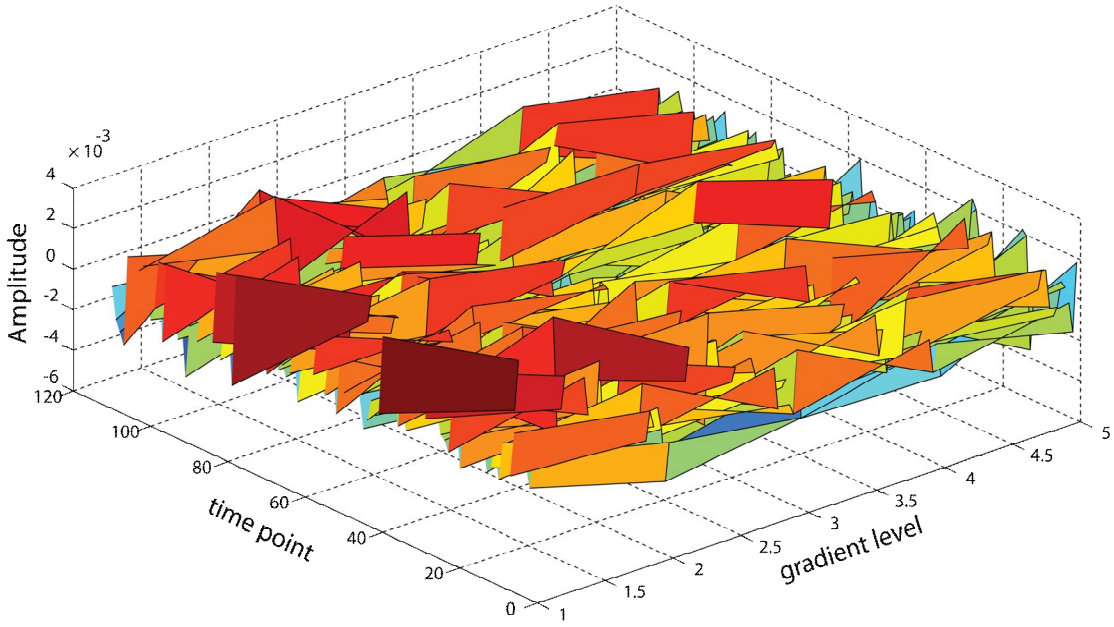


Figure 4.11 The difference between PARAFAC and experimental spectra at 4.8 ppm (arrow 2 in figure 4.9) as a function of gradient level and time course of the experiment in a 3D plot.

The PARAFAC data processing of the lowest concentration sucrose hydrolysis reaction still shows slight cross-talk (figure 4.12), which could be the result of slight temperature variation and/or the change of concentration profile during the course of the reaction. Both of these effects change the diffusion behaviour of the components and perturb the trilinearity of the data. Therefore, the component spectra calculated by PARAFAC differ only slightly from the reference spectra, showing a small amount of cross-talk.

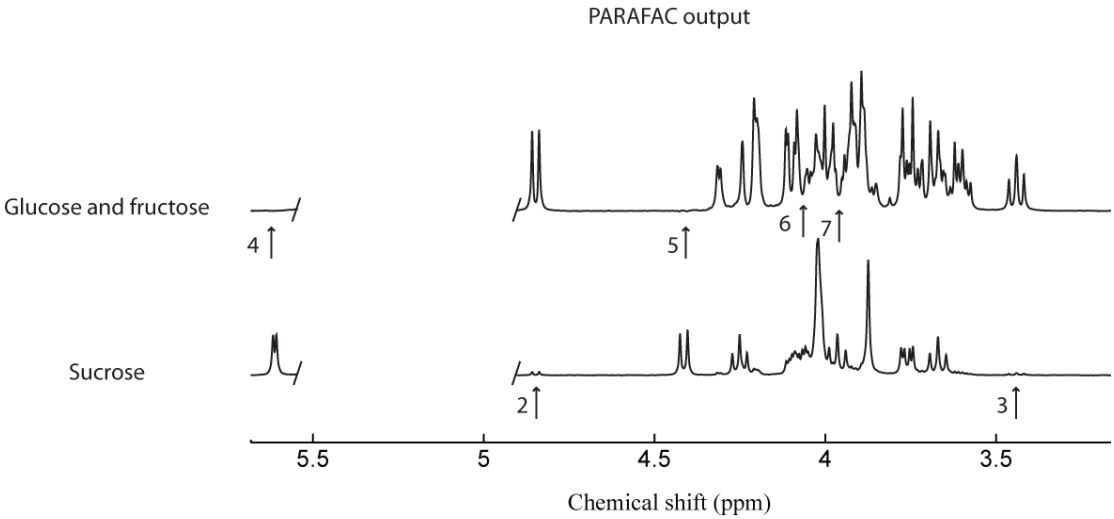


Figure 4.12 Spectra obtained by PARAFAC for less concentrated sucrose hydrolysis (0.006 M). The arrows show the existence of the cross talk, although the concentration has been reduced significantly in comparison with the spectra obtained in figure 4.9.

4.4.1.1 Application of Non-Negativity Constraint

The minimum SNR that PARAFAC could handle and still decompose the data successfully was 3.4:1 in the simulated dataset consisting of a total of 550 spectra (110 spectra × 5 gradient levels). This suggests that PARAFAC can decompose the components when the signal-to-noise ratio of the sum of all the experimental spectra is more than 80 ($\sqrt{550} \times SNR$). With lower values of SNR, PARAFAC can still decompose the data if a non-negativity constraint is applied. The minimum SNR found where PARAFAC can still decompose the data when applying non-negativity was 2:1 (figure 4.13).

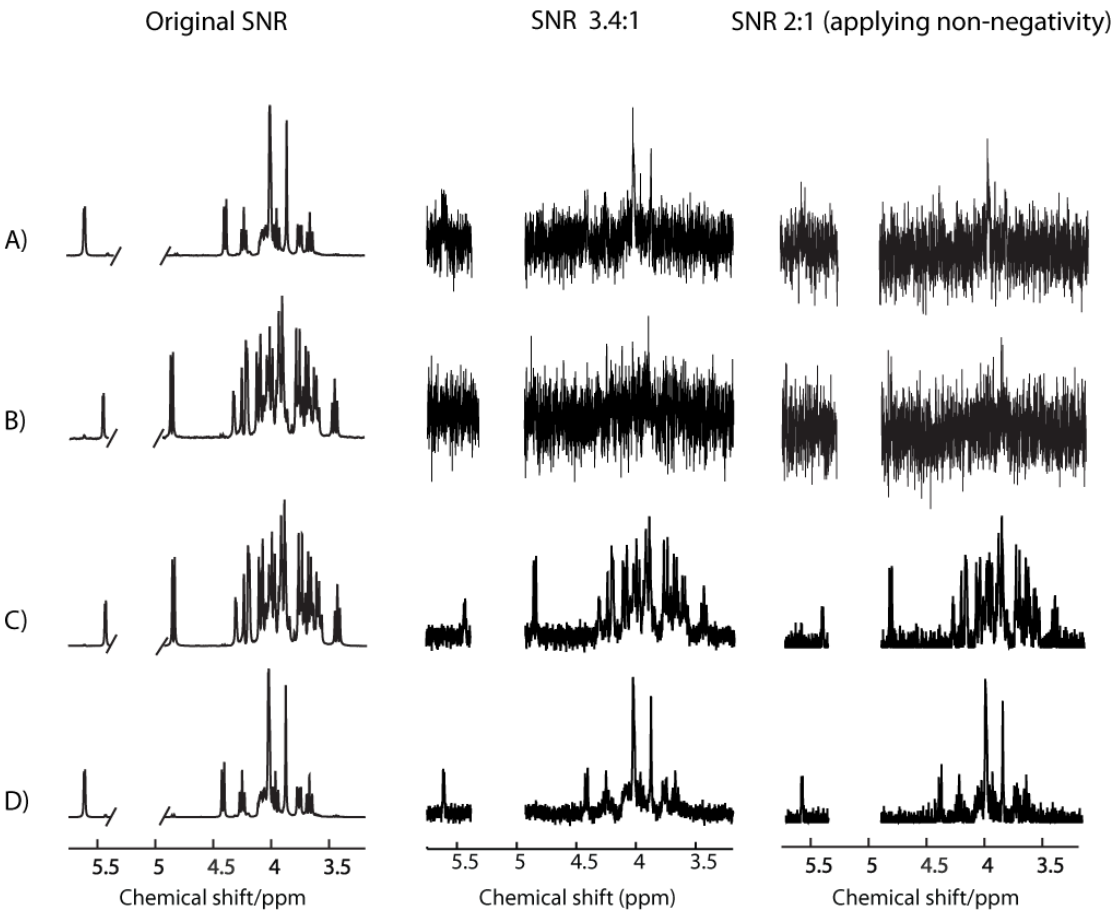


Figure 4.13 Comparison of robustness of PARAFAC to noisy simulated experimental data (water signal is discarded in all spectra). The left column shows the first (A) and the last (B) experimental data; the PARAFAC output, glucose/fructose (C) and sucrose (D) spectrum. The middle column shows the same data as the first column with noise added to give a SNR of 3.4:1. In the right column the SNR is 2:1, for which PARAFAC can not decompose the components without applying a non-negativity constraint. By applying non-negativity, even for such poor signal-to-noise data, PARAFAC can identify the glucose/fructose (C in the right column), and sucrose (D in the right column) spectra.

The experimental data and the kinetics obtained from PARAFAC analysis for the sucrose hydrolysis with low concentration (1.1 mM) and SNR of 8:1 is shown in figure 4.14.

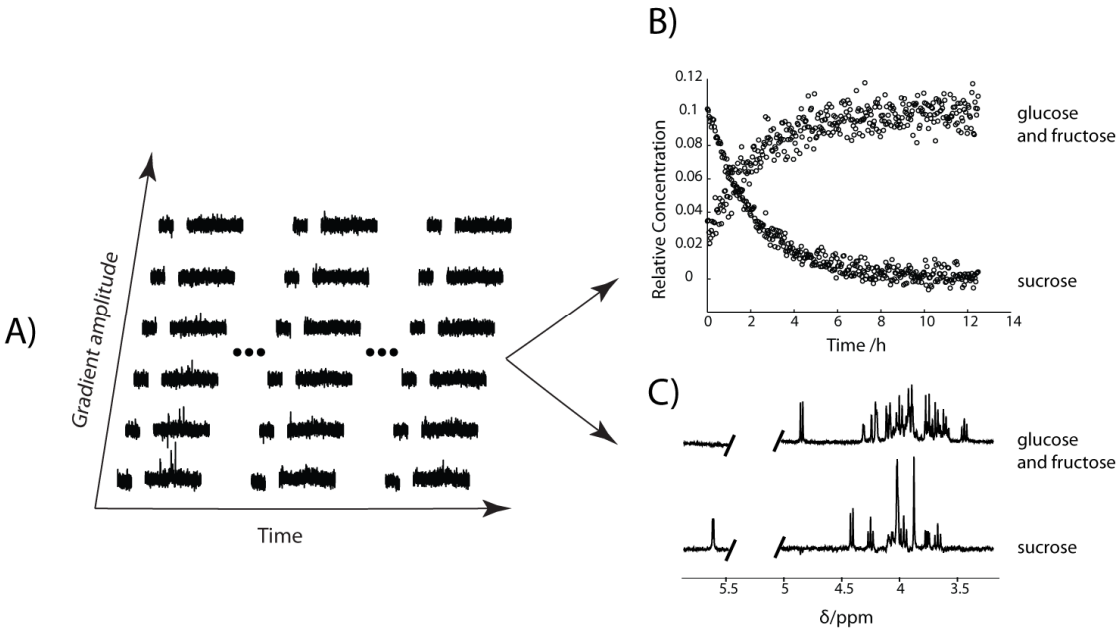


Figure 4.14 A) A subset of the experimental data (spectra as a function of pulsed field gradient and reaction time) obtained for low concentration sucrose (1.1 mM) showing data for the start, midpoint, and end of the period monitored, together with the time evolution (B) of the components corresponding to sucrose and to glucose/fructose, and the fitted spectra (C).

The low concentration sucrose (1.1 mM) experiment was repeated twice, each time with different SNR (varying the gain of the spectrometer). The result in figure 4.15 shows that PARAFAC is able to distinguish the 2 components for both levels of S/N, although the noisier data required the application of non-negativity constraint. PARAFAC cannot distinguish the components for these data if the non-negativity constraint is not applied (figure 4.16).

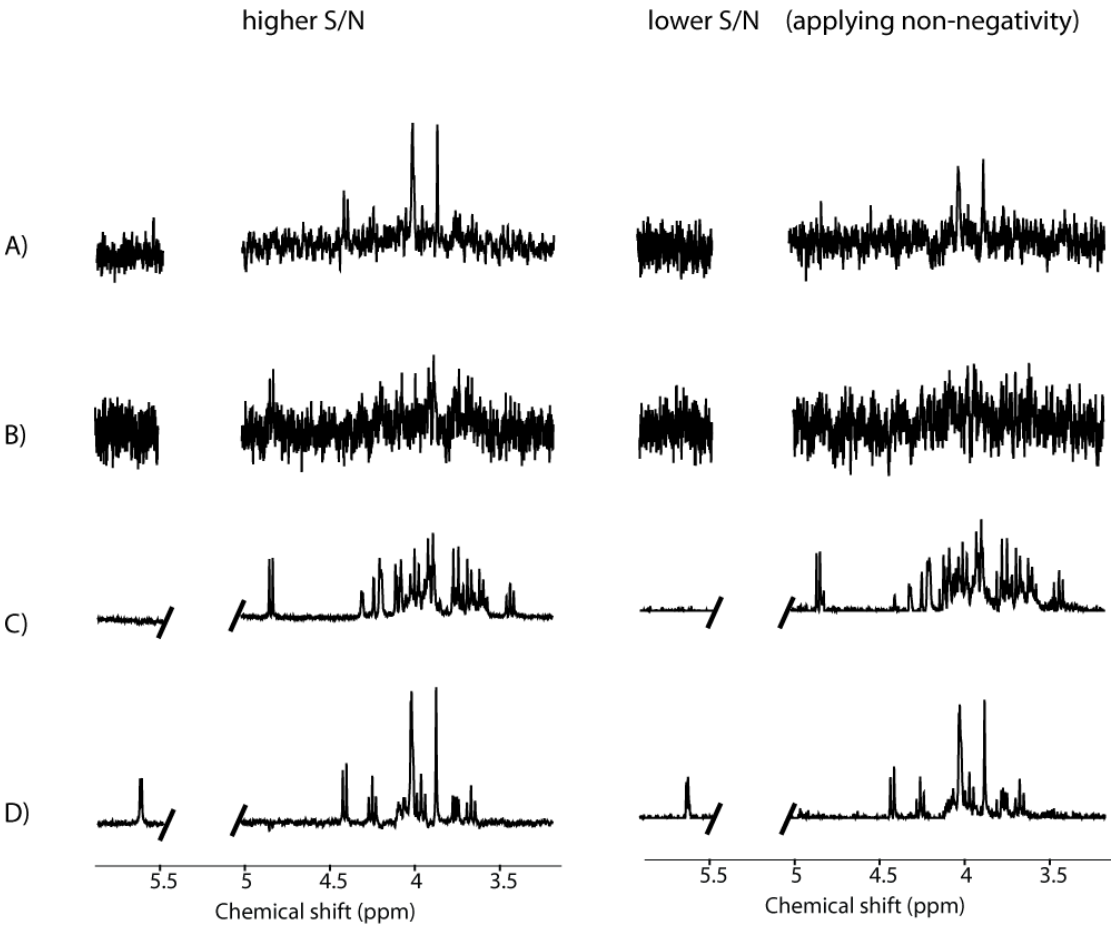


Figure 4.15 A and B in the left column are experimental data (spectrometer gain of 50), with A being the first (S/N of 8:1) and B the last spectrum. PARAFAC can easily distinguish between the components (C and D in the left column) without applying any constraints. A and B in the right column are experimental data (spectrometer gain of 26), with A being the first (S/N of 3.5:1) and B the last spectrum. By applying non-negativity constraint, PARAFAC can still find the components (C and D in the right column). The data processing of the lower S/N data was done using gradient levels 4 to 6 only.

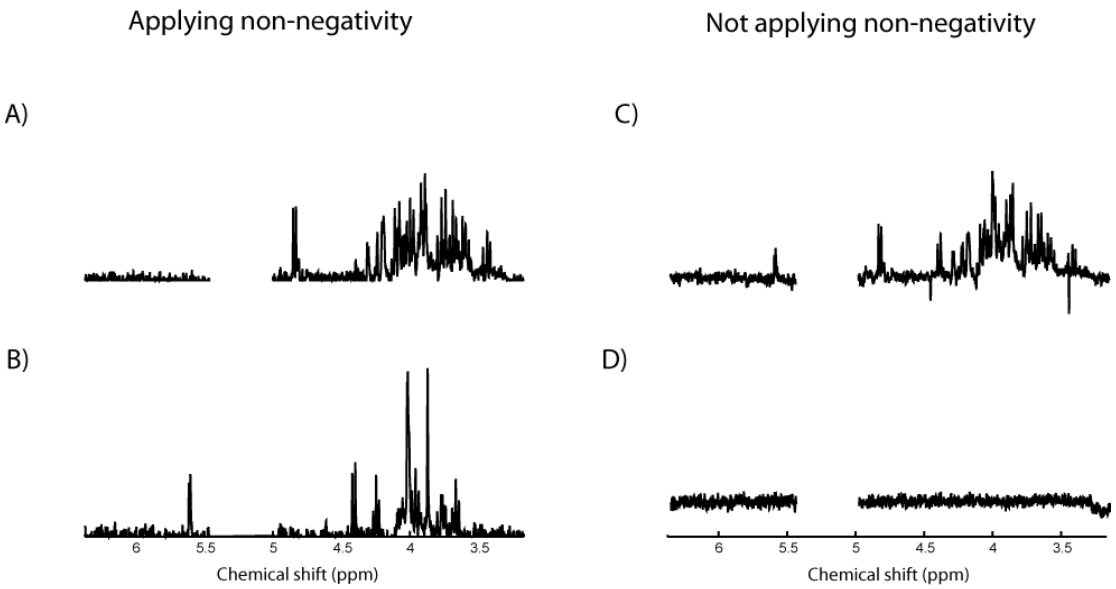


Figure 4.16 The effect of applying and not applying non-negativity constraint to the analysis of the dataset shown in the right column of figure 4.15. By applying non-negativity constraint PARAFAC can distinguish between the glucose/fructose spectrum (A), and sucrose spectrum (B). Using the same data but without the constraint applied, PARAFAC did not distinguish between the components, and yields a spectrum which is a mixture of both components (C).

Initial PARAFAC analysis results of sucrose hydrolysis showed spectra that were comparable with the reference spectra (figure 4.7), but anomalies were observed for a small number of signals. These anomalies were shown to be caused by composition-dependent chemical shifts attributable to the high solution concentrations initially used, which causes significant intermolecular interactions. Experiments performed at much lower concentration showed excellent results. The results also show that PARAFAC is able to analyse very low signal-to-noise (S/N or SNR) datasets.

4.4.2 Maltose Hydrolysis

The normalised kinetic plot obtained from PARAFAC was fitted to first order kinetics by non-linear least squares fitting, yielding a rate constant of $k = 1.40 \pm 0.01 \times 10^{-5}$ s^{-1} (figure 4.17 A). This result was assessed by comparing it with the kinetics obtained directly from the experimental integrals (figure 4.17 B). The integral of the inner anomeric maltose signal (5.4 ppm) of the experimental data was calculated by summing the amplitudes of the relevant data points of the signal for a given gradient level (gradient level 2 was used). In the same manner the integral of the anomeric glucose signals were calculated by summing the integrals of the β and α peaks between 4.6 and 5.3 ppm minus the integral of the inner anomeric maltose at 5.4 ppm (this subtraction has to be done because the β and α peaks of maltose and glucose overlap. In order to be able to calculate the anomeric signals (β and α) of glucose only, one proton worth intensity of maltose (at 5.4 ppm) needs to be subtracted from the mixture of β and α glucose and maltose). The integral plot was fitted to first order kinetics by non-linear least squares fitting and yielded a rate constant of $k = 1.36 \pm 0.02 \times 10^{-5}$ s^{-1} (figure 4.17 B), confirming the quality of kinetic studies obtained by PARAFAC.

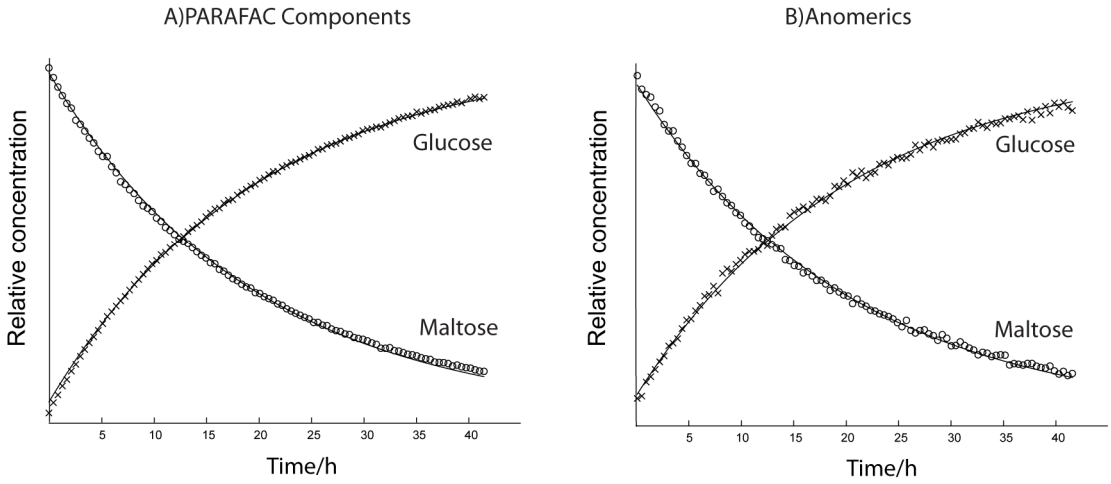


Figure 4.17 Kinetic plots of acid hydrolysis of maltose to glucose with non-linear least squares fitting to a first order kinetic model. Kinetics obtained from PARAFAC (A) with $k = 1.40 \pm 0.01 \times 10^{-5}$ s^{-1}, and from integration of anomeric signals (B) with $k = 1.36 \pm 0.02 \times 10^{-5}$ s^{-1}.

As it is shown in figure 4.18 the PARAFAC components spectra are virtually identical to the spectra of pure maltose and glucose. This confirms the trilinear decomposition into two components was successful.

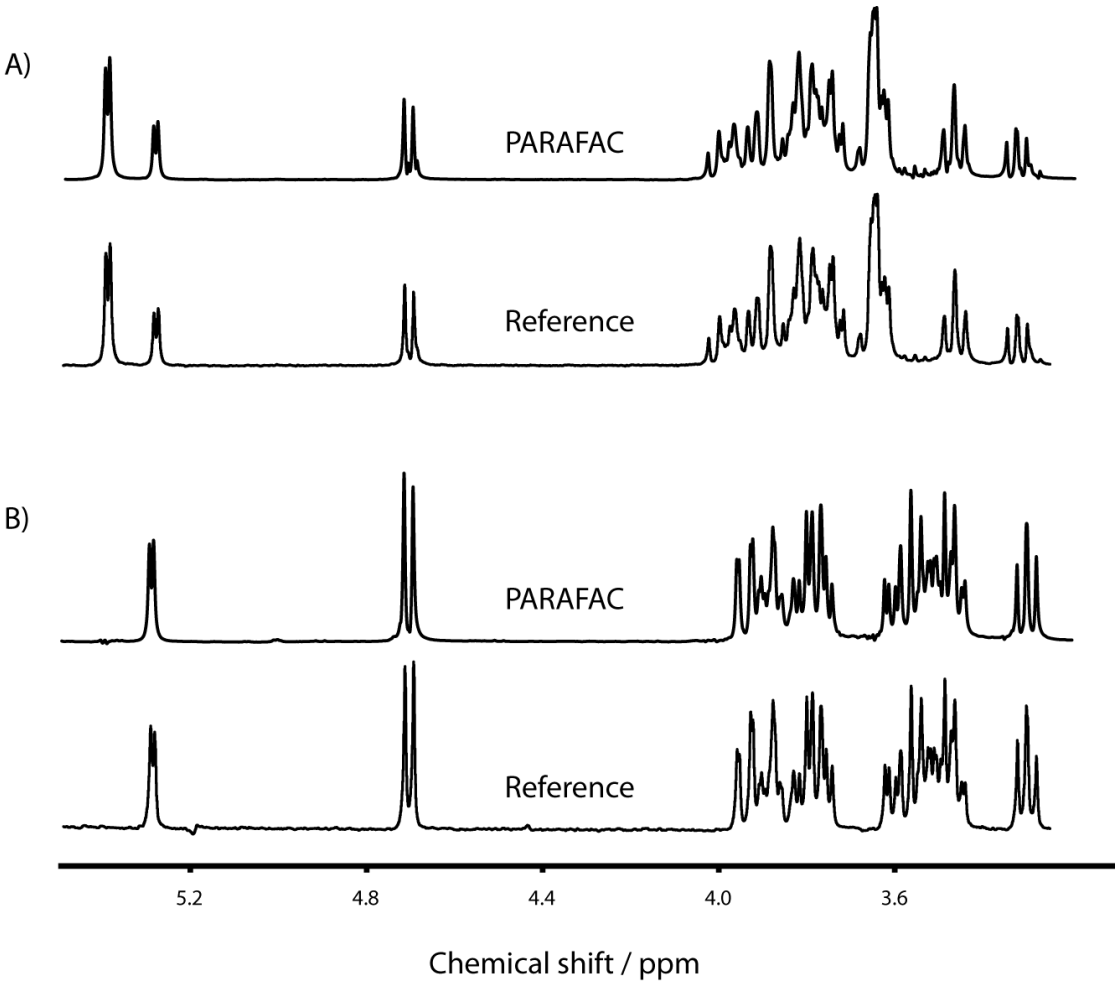


Figure 4.18 Spectra obtained by PARAFAC and reference spectra of pure materials for maltose (A) and glucose (B).

The diffusion mode of the PARAFAC output fitted to the Stejskal-Tanner equation gave diffusion coefficients of 5.4 ± 0.2 × 10^{-10} m^2 s^{-1} for maltose and 6.5 ± 0.2 × 10^{-10} m^2 s^{-1} for glucose. These values are comparable with diffusion coefficients obtained for pure maltose of 5.3 ± 0.2 × 10^{-10} m^2 s^{-1} and 6.5 ± 0.3 × 10^{-10} m^2 s^{-1} for glucose.

4.4.3 Maltotriose Hydrolysis

Concentration timecourses obtained by PARAFAC were fitted to successive first order kinetics, giving rate constants of $k_1 = 5.68 \pm 0.07 \times 10^{-5}$ s^{-1} for the hydrolysis of maltotriose to maltose and glucose, and $k_2 = 2.44 \pm 0.02 \times 10^{-5}$ s^{-1} for the hydrolysis of maltose to glucose (figure 4.19).

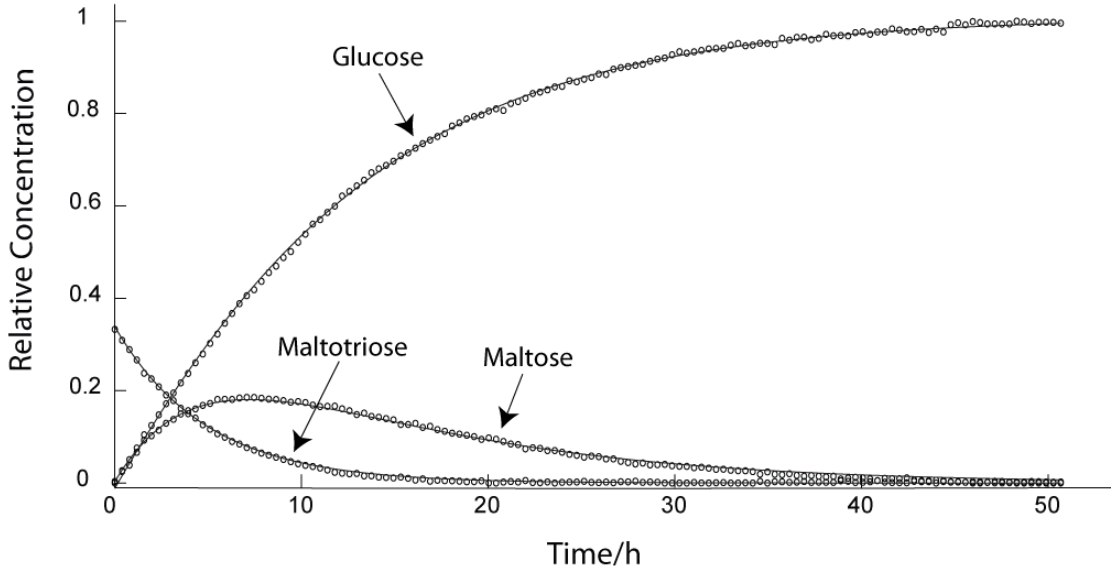


Figure 4.19 Concentration timecourses (circles) obtained from the PARAFAC fit, together with fits to sequential first order kinetics (solid lines). Estimated rate constants were $k_1 = 5.68 \pm 0.07 \times 10^{-5}$ s^{-1} and $k_2 = 2.44 \pm 0.02 \times 10^{-5}$ s^{-1} for the maltotriose and maltose hydrolyses respectively.

As shown in figure 4.20, the PARAFAC component spectra are virtually identical to the spectra of pure maltotriose, maltose and glucose.

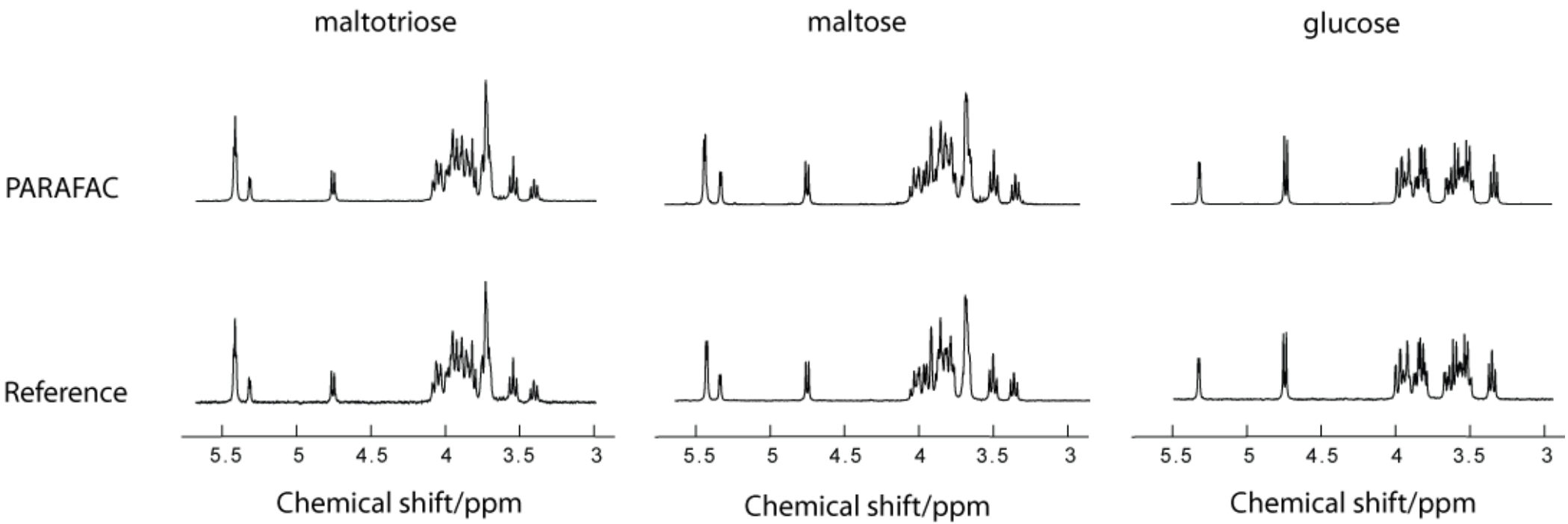


Figure 4.20 Spectra obtained from a non-negativity constrained three-component PARAFAC are shown at the top, and the pure reference spectra of the components are at the bottom.

The diffusion mode of the PARAFAC output fitted to the Stejskal-Tanner equation gave apparent diffusion coefficients of $5.00 \pm 0.01 \times 10^{-10}$ m^2 s^{-1} for maltotriose, $4.44 \pm 0.02 \times 10^{-10}$ m^2 s^{-1} for maltose, and $6.02 \pm 0.04 \times 10^{-10}$ m^2 s^{-1} for glucose. The greater apparent diffusion coefficient for maltotriose than for maltose was an unexpected, and unphysical, result, so more data analysis was done. Examination of the raw time course data showed that the diffusion coefficients of the components were changing with time. This diagnosis was confirmed by measuring the reference chemical shift position as a function of time. The apparent reference shift changes with temperature because of the temperature dependence of the deuterium lock resonance (ca. 9.4 ppb K^{-1}). The reference signal position showed that the temperature of the sample drifted approximately 2 °C in the early stages of the reaction. Thus in the PARAFAC analysis, the maltotriose, which was only present early in the reaction, shows faster diffusion than the maltose, which only appeared later, when the sample was cooler. Restricting the analysis to only two gradient values reduced but could not eliminate the impact of the temperature changes on the PARAFAC analysis. The measurements have since been repeated by a colleague under more stable conditions; this time the maximum temperature variation was only 0.5°, and PARAFAC analysis using all 6 gradient levels gave the expected diffusion results as well as the correct component spectra.

4.4.4 Clarithromycin Hydrolysis

Different timepoints of the clarithromycin hydrolysis data were used to decompose the components. Attempts were made to decompose C, F, and cladinose (scheme 4.4) using the early timepoints of the reaction, which yields signals of these components

only. The diffusion coefficients of C and F are similar, and also the concentrations of F and cladinose are the same. Due to these degeneracies PARAFAC was not able to identify the three components, despite the non-negativity constraint. The later part of the data, where C has vanished and only F, G, and cladinose are present, was also used for PARAFAC analysis to identify the relevant components. However, due to degeneracy between diffusion coefficients of F and G the decomposition failed. The methyl region corresponding to Me* (scheme 4.4) of C, F, and G was also used for PARAFAC analysis to distinguish between C, F, and G; again without any success. The whole dataset was also used for PARAFAC analysis in an attempt to decompose the 4 components, but was not successful. These results illustrate the expected failure of PARAFAC analysis for datasets with multiple degeneracies.

4.5 Conclusions

Timecourse DOSY and PARAFAC can be a very powerful combination for studying the kinetics of chemical reactions. The method is very useful when signals overlap and it is not easy to distinguish between the signals of different components. At very low concentrations, where noise is dominant, PARAFAC can still decompose the data, showing that this method copes well with low signal-to-noise ratio. This could be particularly useful for the studying of reactions where the reactants are expensive, have low solubility, or have low availability.

Problems with spectral inconsistencies during the reaction that are of instrumental origin can in many cases be alleviated by using reference deconvolution, while changes in spectra due to concentration-dependent shifts can be minimised by making measurements at relatively low concentration. In systems with near degeneracy (e.g. very similar NMR spectra or diffusion coefficients), accurate kinetic and spectral data can be obtained by imposing physically appropriate constraints such as non-negativity. However, as was shown in the clarithromycin study, where there is extensive degeneracy applying constraints can not always solve the problem.

The best results with PARAFAC are obtained if the data have no deviation from trilinearity, and the behaviour of different components is non-degenerate, i.e. no two components have one of spectrum, diffusion coefficient or time course in common. (If two modes are the same, the components become indistinguishable). One potential source of deviation from trilinearity is the non-instantaneous acquisition of the DOSY dataset. This means that any changes in concentration of the different components during the recording of a given DOSY dataset will cause non-trilinear variation. In

this investigation, however, such effects can safely be ignored as the time taken to record one set of diffusion-weighted spectra is negligible compared to the total reaction time. The second source of deviation from trilinearity is composition-dependent chemical shifts. However, as was shown in the Results section, this problem can be considerably reduced by using much less concentrated solutions.

Chapter 5

A Novel Flow Cell for Kinetic Studies by NMR

5.1 Introduction

The application of NMR spectroscopy to kinetic and mechanism studies of chemical reactions is not limited to static methods, where a solution is placed in an NMR tube and inserted into the magnet for measurements, but also includes stopped [5.1] or continuous flow methods [5.2]. Flow NMR techniques have been used in a variety of areas, such as studies of protein folding [5.3, 4], food science [5.5], and HPLC-NMR [5.6]. These techniques have found considerable application since 1951 [5.7] in the investigation of flow rates, and in monitoring species during chemical reactions [5.8-10].

Producing a sample in an NMR tube, placing it in the NMR probe, and acquiring a spectrum can hardly be performed faster than a few seconds. This length of time can result in a loss of data in fast chemical reactions. This time limitation can be prevented by fast mixing process of solutions by flow NMR and rapid data acquisition which allows short lifetime intermediate species to be monitored [5.3, 11].

Flow NMR has a number of advantages. It reduces the apparent T_1 and T_2 values of nuclei, so shorter recycle delays can be used, improving sensitivity. As fresh premagnetized spins enter the detection region (region surrounded by the RF coils), saturated spins leave it, causing a reduction in the apparent relaxation times [5.12] (and hence an increase in line broadening [5.13]). The faster the flow rate, the greater the reduction in apparent relaxation times due to the increase in the motions between the spins [5.5]. If rapid pulsing is used (recycle delay << $5T_1$), increasing the flow rate increases the average signal intensity as long as the spin lifetime ($T_{p,}$) in the premagnetization region (the region in the magnet before the detection) is large

enough for the spins to be fully relaxed ($T_p > 5T_1$). If the flow rate is increased further, the lifetime of the spins in the premagnetization region will become too short for full relaxation and the signal intensity is reduced [5.5, 14].

The faster apparent relaxation caused by flow is particularly favourable for nuclei with long spin-lattice relaxation times and low sensitivity (e.g. ^{13}C). Data acquisition for such nuclei requires time-consuming signal averaging. This problem can be reduced by the addition of paramagnetic species, as these "relaxation reagents" considerably reduce T_1. However they contaminate the sample and can perturb chemical shifts and line shapes [5.12].

Spin equilibration in the premagnetization region is needed to provide magnetization to be detected, but increases the dead volume of the flow system. The extra volume needed can be minimised by using immobilized free radical species to reduce T_1 [5.12, 15], but again this risks sample contamination and is not often used.

Here a novel flow system for the study of reaction kinetics is developed. No modification of the NMR instrument is needed, allowing the system to be employed with any conventional NMR probe and magnet. The system is cheap, robust, economical, and can be used for studying both homogeneous and heterogeneous reactions. The reaction vessel is placed outside of the magnet; therefore the composition of the sample can be varied during the experiment, if any reagents need to be added. This technique can be used for either continuous or stopped flow measurements, and the results obtained are comparable with those from conventional high resolution probes.

5.2 Theory

5.2.1 Flow and Magnetization

NMR signals obtained from flowing nuclei are affected by different parameters such as the degree of premagnetization, relaxation, flow rate, residence time, velocity, and detection region dimension. For quantitative measurements, the sample should ideally be completely premagnetized prior to detection. The flow rate can be expressed as [5.5]:

$$f = \frac{V_p}{T_p} \qquad \text{(equation 5.1)}$$

where $T_p > 5T_1$; f, V_p, T_p are the flow rate, and volume of and time in the premagnetization region respectively. To understand the interplay of the different parameters, it is helpful to analyze how they influence the results of a saturation recovery experiment performed in a flowing system.

In a flowing system the detection region contains a combination of nuclei that have entered since the last radiofrequency pulse was applied (new nuclei), and ones that have previously been saturated (old nuclei). For plug flow, where the flow velocity is constant over the cross-section of the flow, the total magnetization ($M_z(t)$) obtained from the two groups of nuclei is:

$$M_z(t) = M_z^{new} + M_z^{old}$$

$$M_z(t) = \alpha V^{new} + \alpha(V - V^{new})(1 - e^{-\frac{t}{T_1}}) \qquad \text{(equations 5.2)}$$

$$M_z(t) = \alpha \frac{t}{\tau} V + \alpha \frac{\tau - t}{\tau} V(1 - e^{-\frac{t}{T_1}})$$

$$for \;\; 0 < t < \tau; \;\; for \;\; t > \tau: \; M_z(t) = \alpha V$$

where M_z^{new}, M_z^{old}, α, V^{new}, V, t, T_1, and τ are the new magnetization, old magnetization, magnetization per unit volume, volume occupied by new magnetization, volume of detection region, time, spin-lattice relaxation time, and total time spent in the detection region respectively.

The plot of z-magnetization variation against time according to equations 5.2 is shown in figure 5.1. Graphs ‘a’ to ‘e’ show the effect of increasing flow rate. At low flow rates recovery is exponential, dominated by relaxation; at high flow rates it is linear, dominated by influx.

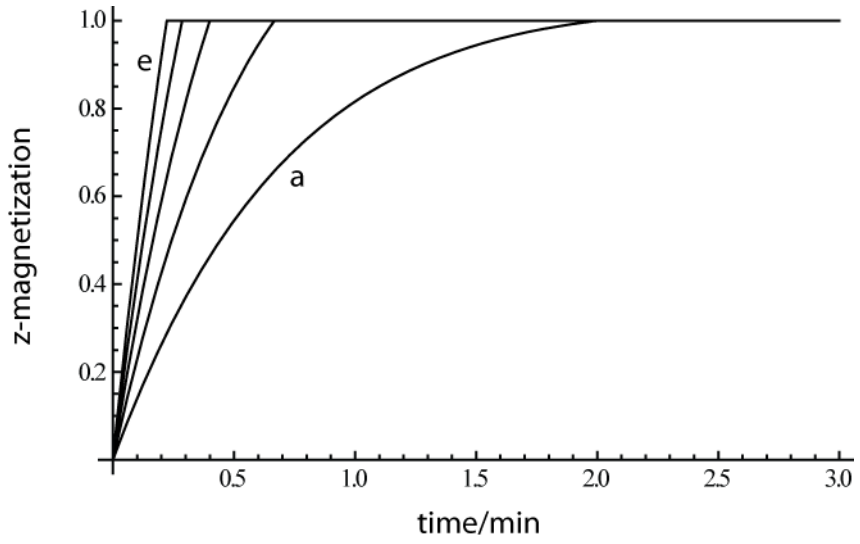


Figure 5.1 Variation of z-magnetization with time for saturation recovery in a plug flow system. The z-magnetization is calculated from equations 5.2, where α, V and T_1 are assumed to have values of 1. The graphs 'a' to 'e' are for different flow rates from 0.5 to 4.5 mL/min in steps of 1 mL/min.

For laminar flow in a cylinder of radius r_0, the distribution of the magnitude of the velocity is parabolic with a maximum velocity v_{max}:

$$v = v_{max}\left[1-\left(\frac{r}{r_0}\right)^2\right] \qquad \text{(equation 5.3)}$$

The signal contributed by spins at a radius r is proportional to the area (dA) of a ring, $2\pi r dr$, at that radius. The relative contribution made by spins with a velocity v is then:

$$S_v = \frac{\frac{dA}{dv}}{\int_0^{v_{max}} \frac{dA}{dv}\,dv} = \frac{\frac{dA}{dr}\frac{dr}{dv}}{\int_0^{v_{max}} \frac{dA}{dr}\frac{dr}{dv}\,dv} \qquad \text{(equations 5.4)}$$

Substituting in:

$$\frac{dA}{dr} = 2\pi r$$
$$\frac{dr}{dv} = \left(\frac{dv}{dr}\right)^{-1} = -\frac{r_0^2}{2r v_{\max}}$$
$$\int_0^{v_{\max}} \frac{dA}{dr}\frac{dr}{dv}\,dv = -\pi r_0^{\,2}$$

(equations 5.5)

Gives

$$S_v = \frac{1}{v_{\max}}$$

(equation 5.6)

Combining equations 5.2 and 5.6, for laminar flow the magnetization as a function of time and velocity is:

$$M_z(t,v) = \frac{\alpha t V \frac{v}{L} + \alpha V(1-\frac{vt}{L})(1-e^{-\frac{t}{T_1}})}{v_{\max}}$$

(equation 5.7)

for $0 < t < \tau$, and

$$M_z(t,v) = \alpha V$$

(equation 5.8)

for t > τ, where

$$v_{max} = \frac{2FL}{V} \qquad \text{(equation 5.9)}$$

The net magnetization as a function of time is then found by integrating equations 5.7 and 5.8 over the range $v = 0$ to $v = v_{max}$, with the results shown in figure 5.2.

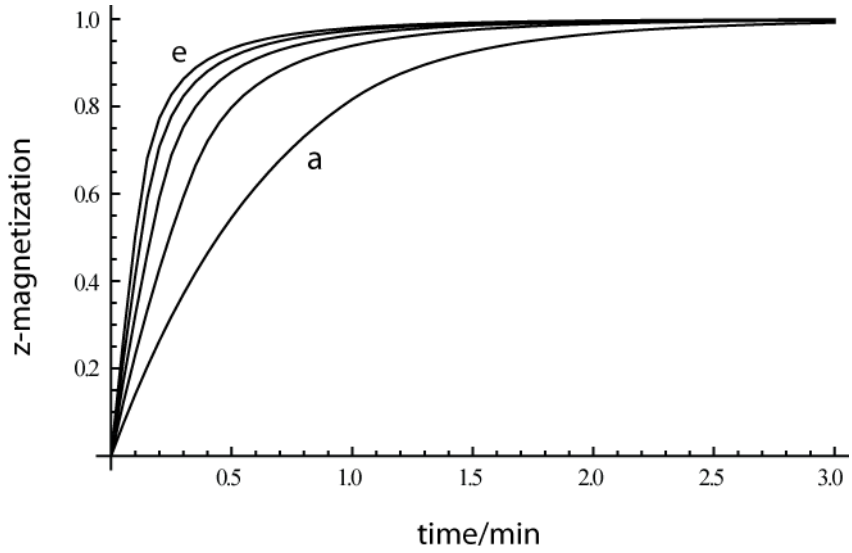


Figure 5.2 The variation of z-magnetization with time for a laminar flow system. The z-magnetization is calculated from equations 5.7 and 5.8, where α, V and T_1 are assumed to have values of 1. The graphs 'a' to 'e' are for flow rates from 0.5 to 4.5 mL/min in steps of 1 mL/min.

In literature reviews [5.5, 14, 15] the combined effect of flow and relaxation on saturation recovery is often expressed as follows:

$$M_z(t) = \alpha V(1 - e^{-\frac{t}{T_1^{flow}}})$$

(equations 5.10)

$$T_1^{flow} = \frac{1}{\frac{1}{T_1^{static}} + \frac{1}{\tau}}$$

where T_1^{flow} is the (apparent) nuclear spin-lattice relaxation time on flow, and T_1^{static} is the true value. According to equations 5.10 the variation of z-magnetization with time is exponential; graphs for various flow rates are shown in figure 5.3. As the preceding analysis shows, equations 5.10 are not correct either for plug or for laminar flow, but the reasonably close agreement between figures 5.2 and 5.3 shows that it is a useful approximation for the laminar flow case.

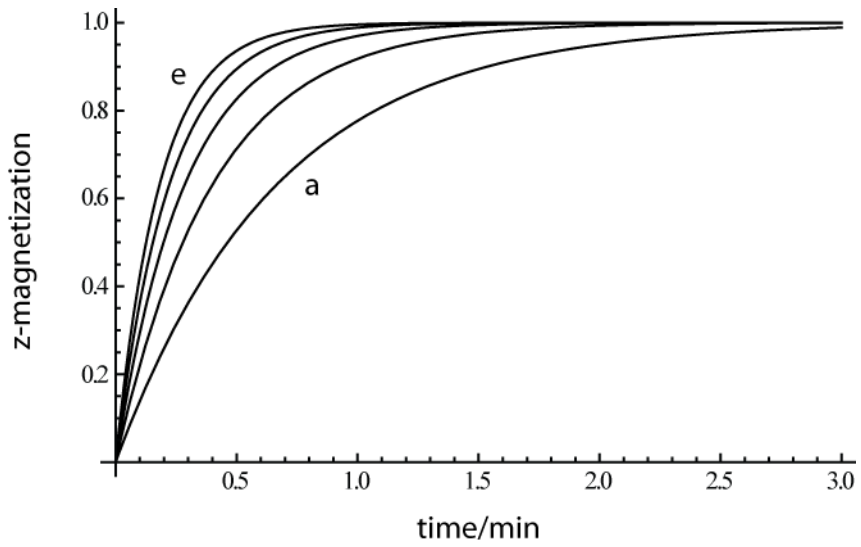


Figure 5.3 The variation of z-magnetization with time determined by some literature for the flow condition. The z-magnetization is calculated from equations 5.10, where α, V and T_1 are assumed to have values of 1. The graphs 'a' to 'e' are for flow rates from 0.5 to 4.5 mL/min in steps of 1 mL/min.

5.2.2 Initial Structure of the Flow System

The basis of the designed novel flow system is to have a reaction mixture outside of the magnet, and using a pump to transfer the reaction mixture into and out of an NMR tube which is placed in the magnet. This system is constructed of the following pieces: a reaction vessel, HPLC PEEK and PTFE pipes, three unions, peristaltic pump, and a 5 mm outside diameter screw thread NMR tube. The first inlet pipe (L_{inlet1}) (2 in figure 5.4) (HPLC PTFE, L_{inlet1} = 30 cm, ID = 1.6 mm, OD = 3.2 mm) is inserted into the reaction vessel (1), and is connected to the peristaltic pump pipe (L_{inlet2}) (5) (HPLC PTFE, L_{inlet2} = 38 cm, ID = 2 mm, OD = 4 mm) by a union (3). The third inlet pipe (L_{inlet3}) (7) (HPLC PEEK, L_{inlet3} = 3 m, ID = 0.5 mm, OD = 1.6 mm) is attached to the pump's pipe (5) by the second union (6) and ends at the very bottom of the NMR tube (8). To keep the input tubing concentric in the NMR tube (4.15 mm ID), two sections of a PTFE vortex plug (9) are used, one at the bottom of the NMR tube/input pipe, and the other 6 cm above that. Grooves are cut into these pipe holders to allow the solution to pass. A fine hole is made through the middle of the rubber seal of the NMR tube cap (10) to pass the input pipe (7) through. A second hole is made about 2 mm from the first for the output pipe (12). The output pipe extends about 1 mm below the cap seal. A thick plug of epoxy resin (Araldite) glue (11) is formed to hold the pipes firmly to the cap. The first output pipe ($L_{outlet1}$) (12) attached to the cap is a 7 cm HPLC PEEK pipe (ID of 0.5 mm, OD of 1.6 mm). To minimize the back pressure of the mixture in the cell at the output pipe, a 1/8"-1/16" union (13) is used to attach a second outlet pipe ($L_{outlet2}$) (14) which is a 3 m HPLC PTFE pipe (ID of 1.6 mm and OD of 3.2 mm) to the first. The solution is passed through the outlet2 (14)

back to the reaction vessel (1). The total volume of reaction mixture required to fill the pipes, pump and the NMR cell is 12.3 mL.

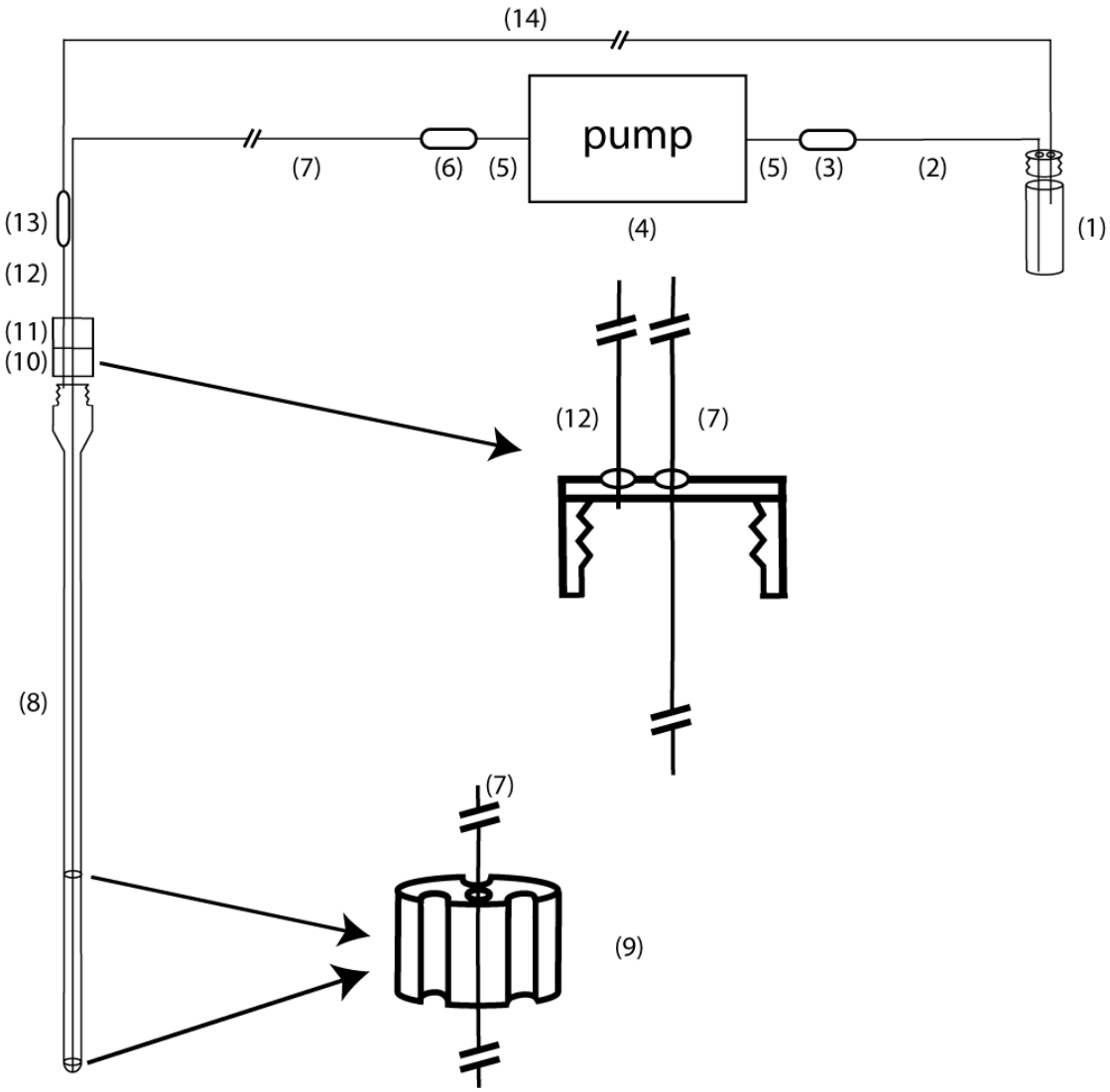


Figure 5.4 Earlier structure of the flow system. (1) Reactor vessel, (2) first inlet pipe, (3, 6, 13) unions, (4) peristaltic pump, (5) pump's pipe, (7) third inlet pipe, (8) NMR tube, (9) vortex plugs, (10) NMR screw cap with 2 holes, (11) epoxy resin, (12) first outlet pipe, (14) second outlet pipe

5.2.3 Further Structure of the Flow System

A disadvantage of the initial constructed flow system was the large volume required to fill the system (12.3 mL). Therefore this system was manipulated to be adapted for less reaction volumes. For this purpose in the new flow system a normal 5 mm OD NMR tube (11 in figure 5.5) was used instead of using a 5 mm OD screw thread NMR tube which contains a dead volume at the top. A home made PEEK head piece (9), constructed from 3 threaded inserts (16, 22), 3 'O' rings (17, 21), and a cap makes a firm connection to the NMR tube while the third inlet pipe (7) passes through its middle. Here the outlet pipe (23) is an HPLC PEEK with L_{inlet3} = 3 m, ID = 0.5 mm, OD = 1.6 mm, which is inserted into the top of one side of the head piece (9) and ends back to the reaction vessel (1). In order to minimize the dead volume of the NMR tube, four long PTFE vortex plug (13) containing grooves, are placed inside the NMR tube to fill in the dead volume. The total volume of reaction mixture required to fill the pipes, pump and the NMR cell of this new design is 3.5 mL.

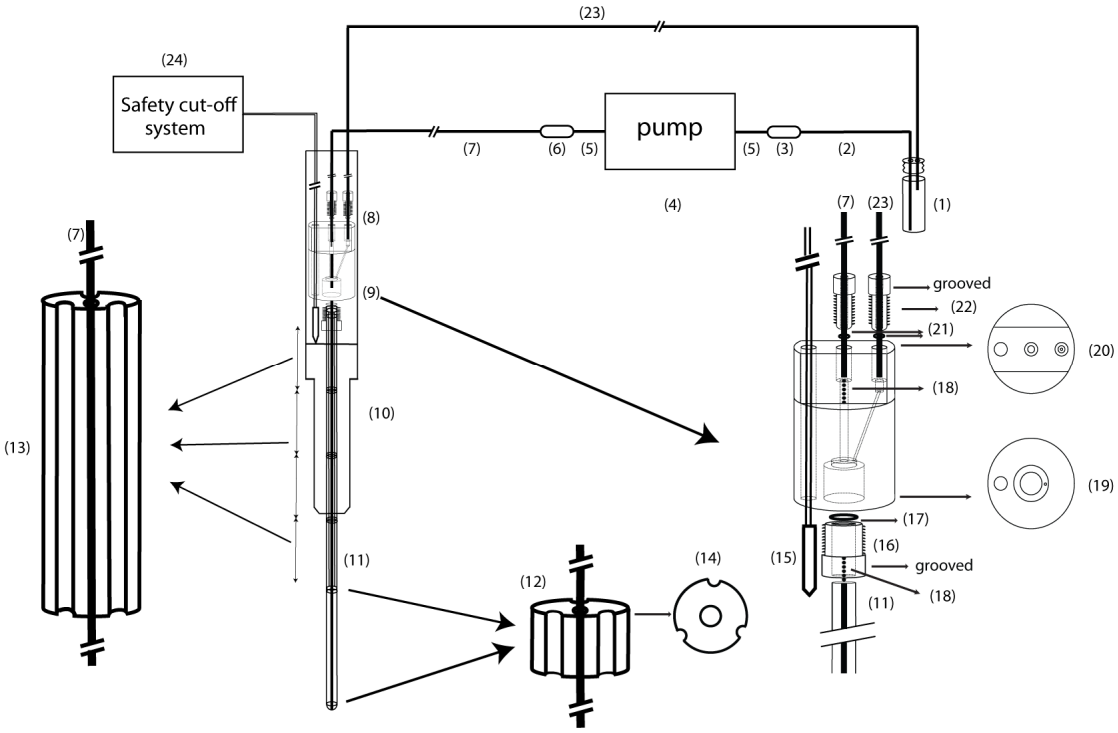


Figure 5.5 Latest structure of the flow system. (1) Reactor vessel, (2) first inlet pipe, (3, 6) unions, (4) peristaltic pump, (5) pump's pipe, (7) third inlet pipe, (8) glass container, (9) head piece, (10) NMR turbine, (11) NMR tube, (12) small vortex plug, (13) Long vortex plug, (14) vortex plug from top, (15) optical liquid level sensor, (16) bigger threaded insert which fits the NMR tube to the head piece, (17) bigger o-ring, (18) diagonal path for the solution inserting the NMR tube, (19) head piece from bottom, (20) head piece from top, (21) smaller o-rings, (22) smaller threaded inserts which fit the input and output pipes to the head piece, (23) outlet pipe, (24) safety cut-off system.

5.2.3.1 Safety Issues

If the pressure of the flow system increases too far, it will fail, either by leakage or by breakage. This could occur at any point in the system, for example at one of the unions, or at the head piece. A pressure increase could be due to a blockage in the system, use of viscous solutions, sudden temperature change of the solution, fast flow rates, etc.

Detection of any leak within the magnet is crucial, as it could cause damage to the probe. An optical liquid level sensor (15 in figure 5.5) was used to detect any leak in the magnet above the level of the turbine. This sensor was placed inside an 8 cm diameter glass cylinder (8) glued by epoxy resin to the NMR turbine which holds the NMR tube. It uses the refractive index difference between air and liquid to detect the latter. This sensor is attached to a cut-off system (24), so that in the event of liquid being detected the pump will be disabled.

Figure 5.6 Photograph of the entire latest flow system. At right the reaction vessel is placed on an aluminium stand, and is based on a heater/stirrer. In the middle, the safety cut-off system is located on the peristaltic pump. At left, the NMR tube, head piece, and the turbine are shown.

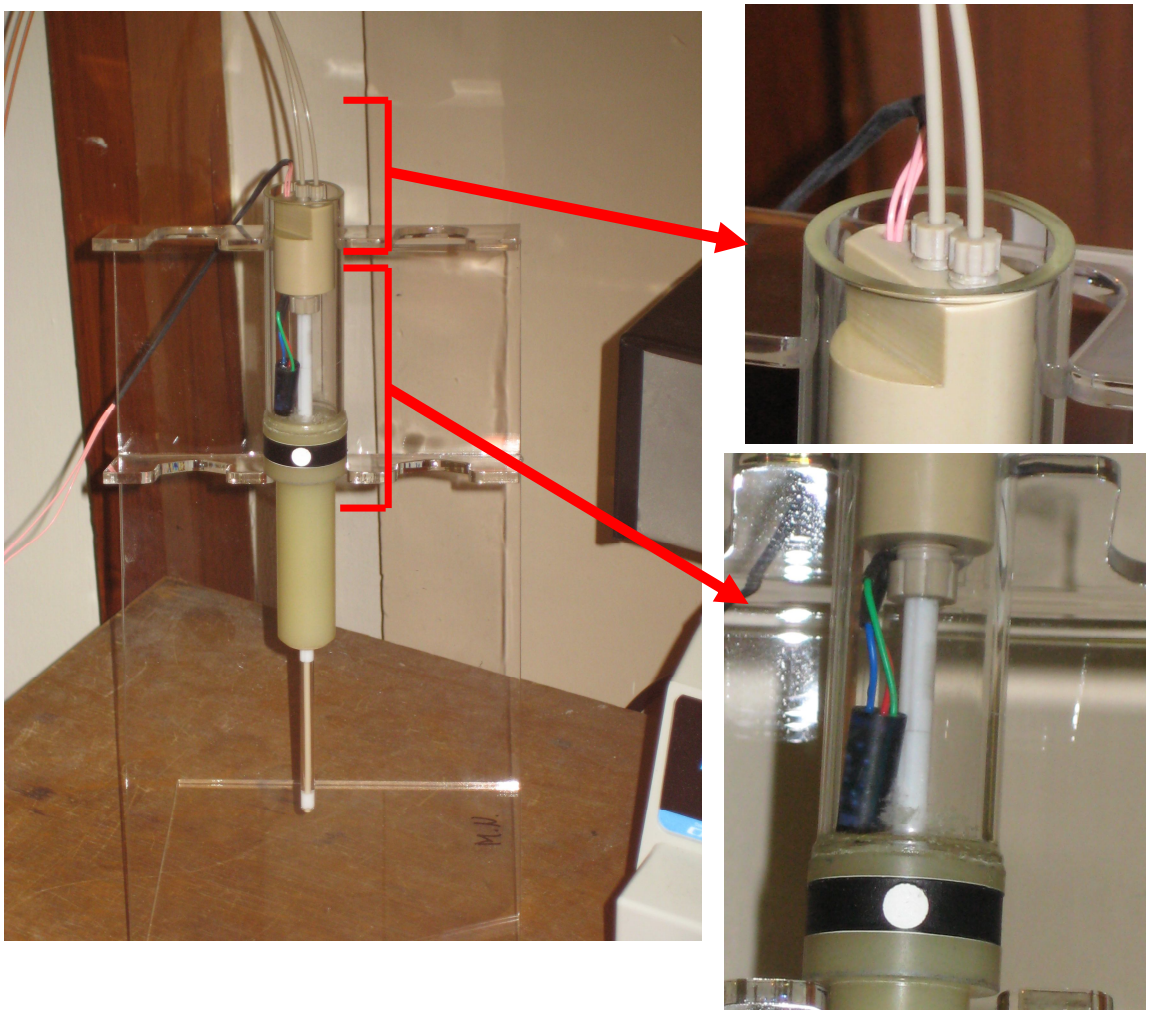


Figure 5.7 Photographs of the head piece, inlet and outlet pipes, sensor, glass container, NMR tube, and the turbine illustrated in figure 5.5.

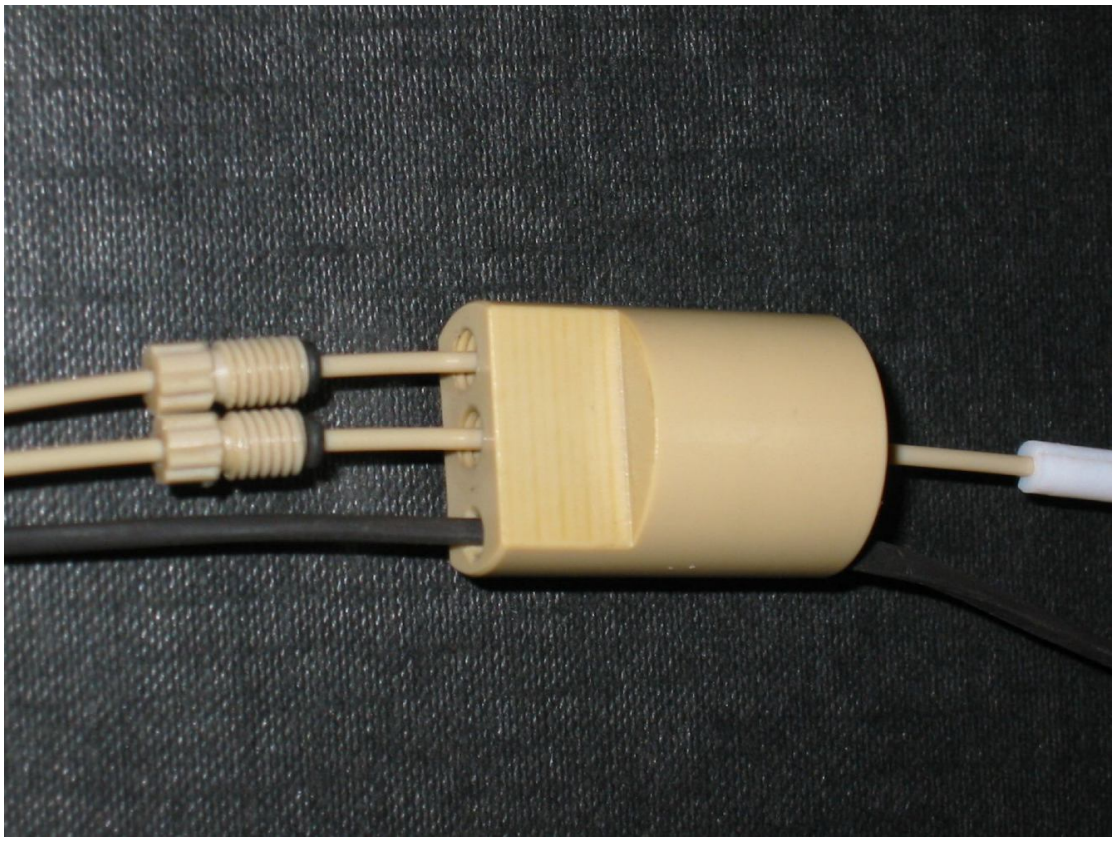


Figure 5.8 Flow cell cap, with the screw caps and the o-rings.

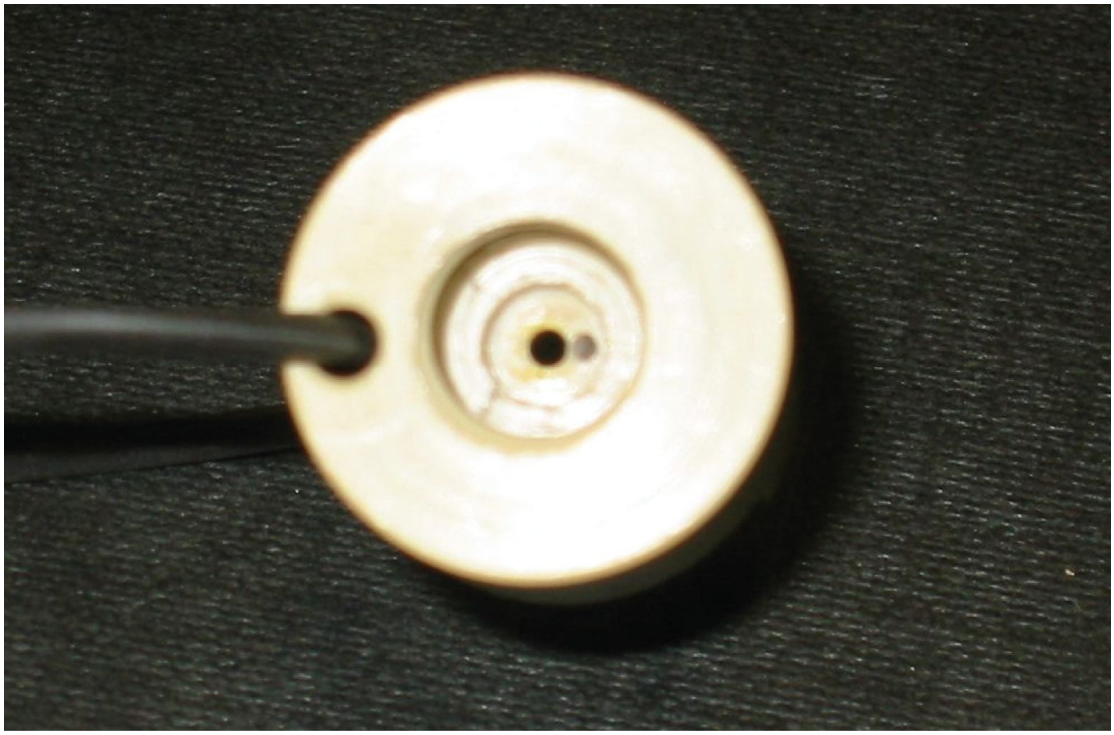


Figure 5.9 Flow cell cap bottom.

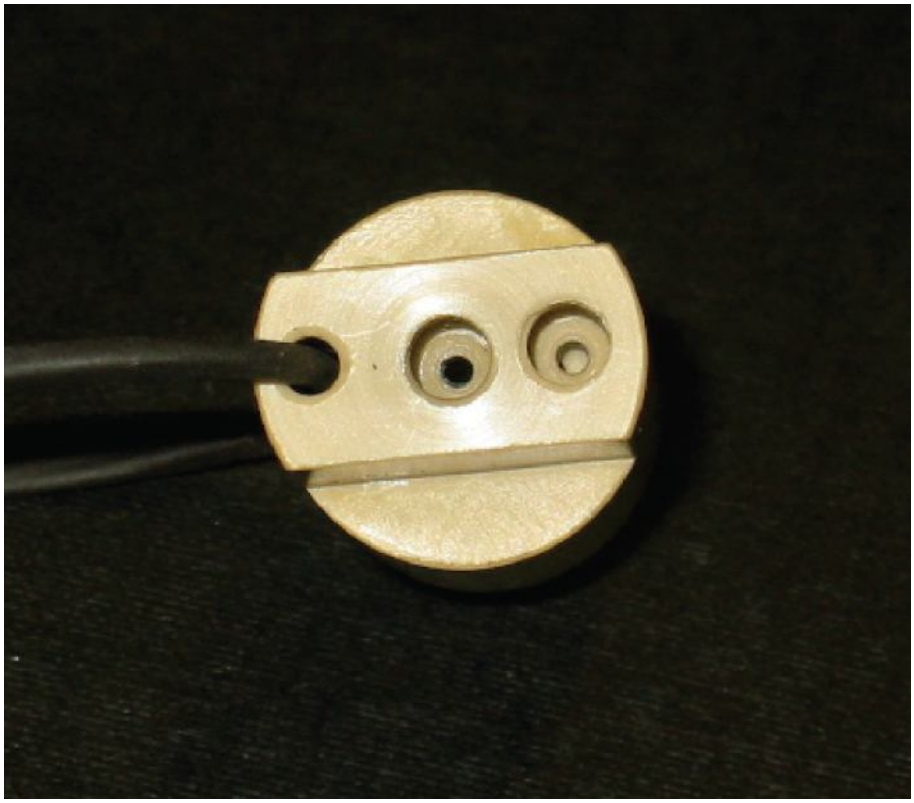


Figure 5.10 Flow cell cap top.

5.2.4 Flow and Pressure

5.2.4.1 Flow and Pressure of the Initial Flow System

The pressure difference (ΔP) between any two points in a viscous fluid with laminar flow in a long cylindrical pipe is determined by Poiseuille's law:

$$\Delta P = \frac{8\eta FL}{\pi r^4} \qquad \text{(equation 5.11)}$$

where η, F, L, r are the liquid viscosity, flow rate, length and radius of the pipe respectively.

The pressure difference in the flow system described here can be represented as:

$$\Delta P_1 = P_p - P_c$$
$$\Delta P_2 = P_c - P_{atm} \qquad \text{(equations 5.12)}$$

where P_p, P_c, P_{atm}, ΔP_1 and ΔP_2 are pressures at the pump outlet, NMR cell, atmosphere, pressure differences between pump and NMR cell, and between NMR cell and the reaction vessel, respectively. Therefore, the pump and the NMR cell pressures are:

$$P_c = \frac{8\eta FL_{outlet1}}{\pi r_{outlet1}^{4}} + \frac{8\eta FL_{outlet2}}{\pi r_{outlet2}^{4}} + P_{atm} \quad \text{(equation 5.13)}$$

$$P_p = \frac{8\eta FL_{inlet3}}{\pi r_{inlet3}^{4}} + \frac{8\eta FL_{outlet1}}{\pi r_{outlet1}^{4}} + \frac{8\eta FL_{outlet2}}{\pi r_{outlet2}^{4}} + P_{atm} \quad \text{(equation 5.14)}$$

For a flow rate of 1 mL/min with water as solvent (η =0.001 Pa s), and the lengths and radius defined in 5.2.2, the pressures at cell and pump are 1.02 bar and 1.35 bar respectively. These low pressures minimize the risk of leaks or of damage to the NMR tube.

5.2.4.2 Flow and Pressure of the Latest Flow System

In the latest design of the flow system, the two outlet pipes that were originally used in the initial flow design were replaced by a single, narrower outlet pipe to reduce the total volume of the system. The pressures of cell and pump outlet for this new design are defined as follows:

$$P_c = \frac{8\eta FL_{outlet}}{\pi r_{outlet}^{4}} + P_{atm} \quad \text{(equation 5.15)}$$

$$P_p = \frac{8\eta FL_{inlet3}}{\pi r_{inlet3}^{4}} + \frac{8\eta FL_{oulet}}{\pi r_{outlet}^{4}} + P_{atm} \quad \text{(equation 5.16)}$$

According to equations 5.15 and 5.16 the pressures of cell and pump are now 1.33 bar and 1.66 bar respectively. This increase of pressure difference in the latest flow system is expected due to the smaller radius of the outlet pipe. However this pressure

is still low enough not to cause any risk. The maximum pressure a short cylindrical tube can handle is given by [5.16]:

$$P_{\max} = \frac{WT}{OD} K \qquad \text{(equation 5.17)}$$

where *WT*, *OD* and *K* are wall thickness, outside diameter, and constant related to the tensile strength of the tube material respectively. In the case of glass *K* is about 2,000 psi (1374 bar). Therefore the maximum pressure that can be applied to a 5 mm OD NMR tube with wall thickness of 0.44 mm is about 12 bar. This pressure capacity is far greater than the pressure caused in the tube by the flow system.

A lower limit on the pressure that the join between the head piece of the latest flow system and the NMR tube can withstand was obtained by hanging a 1.2 kg weight from a normal NMR tube (ID of 4.15 mm) held in the head piece. No movement was seen, showing that such a joint should withstand a minimum pressure of $1.2 \times 9.8/(\pi (0.002075)^2) \approx 8.7$ bar.

5.3 Experimental Work and Data Processing

In evaluating the NMR flow system, the first question was whether having a pipe in an NMR tube would result in a high resolution spectrum, as discontinuities in magnetic susceptibility can ruin the shimming. Therefore the first experiment was under static conditions (no flow), trying to produce a homogeneous magnetic environment with concentric HPLC tubing in the NMR tube, to obtain the maximum spectral resolution. Later experiments to study the suitability of the novel flow system for kinetic studies monitored four set of different chemical reactions. Three homogeneous reactions (one at high temperature), and a heterogeneous reaction were monitored. All of the experiments were conducted on a 300 MHz Varian spectrometer, using a standard 5 mm diameter indirect detection probe.

5.3.1 Initial Experiment

The ^{1}H spectrum of a static 0.05 M solution of quinine in approximately 90 % CH_3OH and 10 % acetone-d_6 with TMS was obtained, with PEEK HPLC tubing (ID 0.5 mm, OD 1.6 mm) held concentrically in the normal NMR tube (4.15 mm ID), using two sections of a PTFE vortex plug. The highly resolved spectrum obtained (figure 5.11) confirms the feasibility of using HPLC PEEK tubing in an NMR tube for the flow system. Figure 5.12 shows the result of processing the free induction decay, using reference deconvolution with a target lineshape of a 1 Hz Lorentzian.

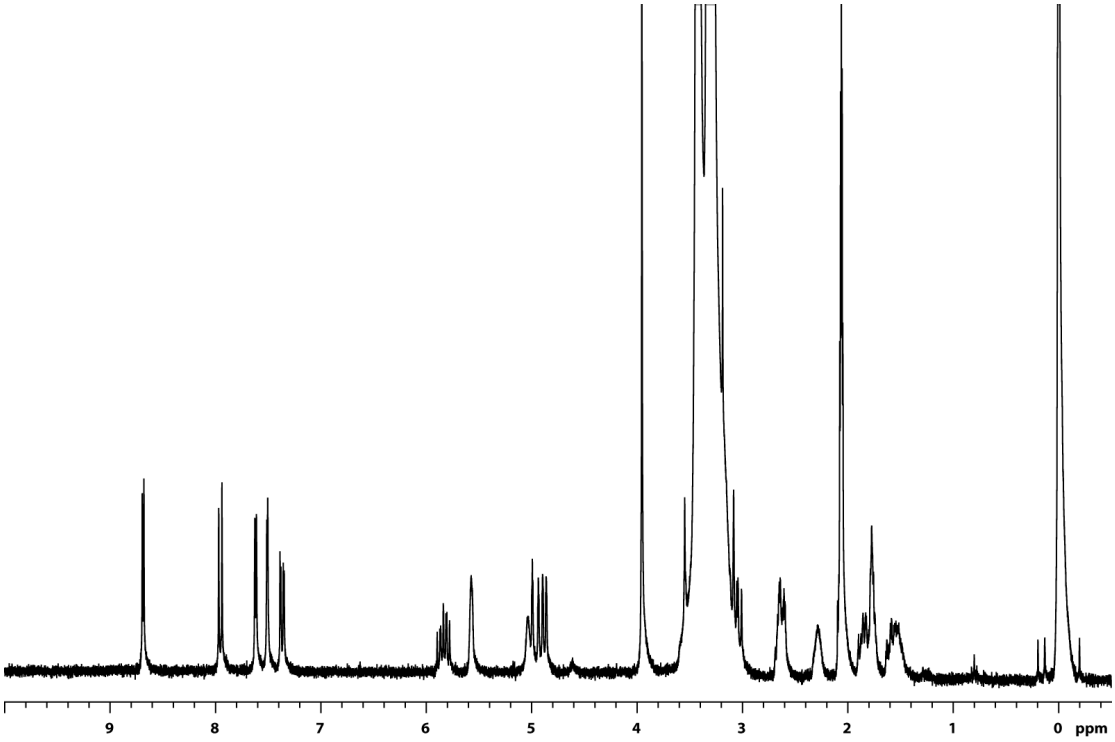


Figure 5.11 The raw ^{1}H spectrum of 0.05 M quinine/acetone-d_6/CH_3OH/TMS in flow cell under static conditions.

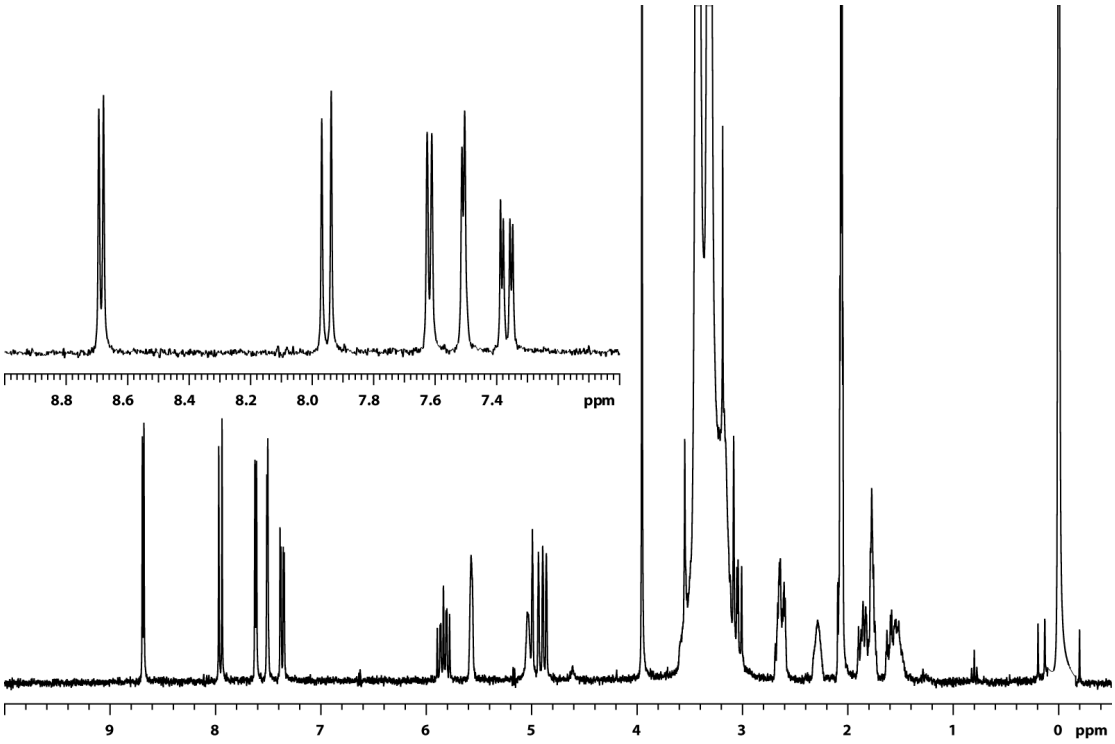
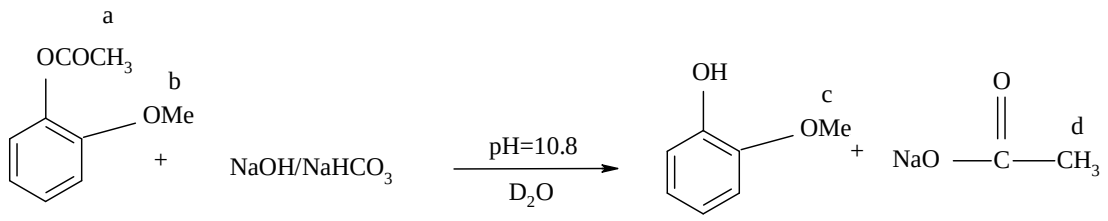


Figure 5.12 ^{1}H spectrum of 0.05 M quinine/acetone-d_6/CH_3OH/TMS in flow cell under static conditions; reference deconvoluted to 1 Hz Lorentzian.

5.3.2 First Homogeneous Reaction

The first chemical reaction implemented in the flow system was an ester hydrolysis (scheme 5.1). This experiment was done using the initial flow system where the total volume was 12.3 mL. A mixture of 11 mL of 0.08 M buffer solution of sodium hydroxide/sodium bicarbonate (pH = 10.8) in 1.5 mL D_2O with 240 mM TSP as the reference compound was pumped into the empty system at a flow rate of 1 mL/min. After filling the system, the flow was stopped to lock and shim the magnetic field (a linewidth of 1.1 Hz was obtained for the TSP signal). Then the parameters to acquire a timecourse array of ^1H spectra were set up. The total duration of the experiment was 12 h, with a recycle delay of 20 s, 4 transients, a 90° pulse width of 9 μs, and an acquisition time of 3.4 s for each ^1H spectrum. The timecourse was obtained by arraying the recycle delay. The next step was to add the reactant (1.5 mL of 0.08 M 2-methoxyphenyl acetate) to the reaction vessel. The solution was then stirred and the acquisition started. Parts of the array of ^1H spectra obtained are shown in figure 5.13.



Scheme 5.1 Ester hydrolysis of 2-methoxyphenyl acetate with buffer solution of sodium hydroxide/sodium bicarbonate in D_2O (lower-case letters indicate the NMR assignments in figure 5.13).

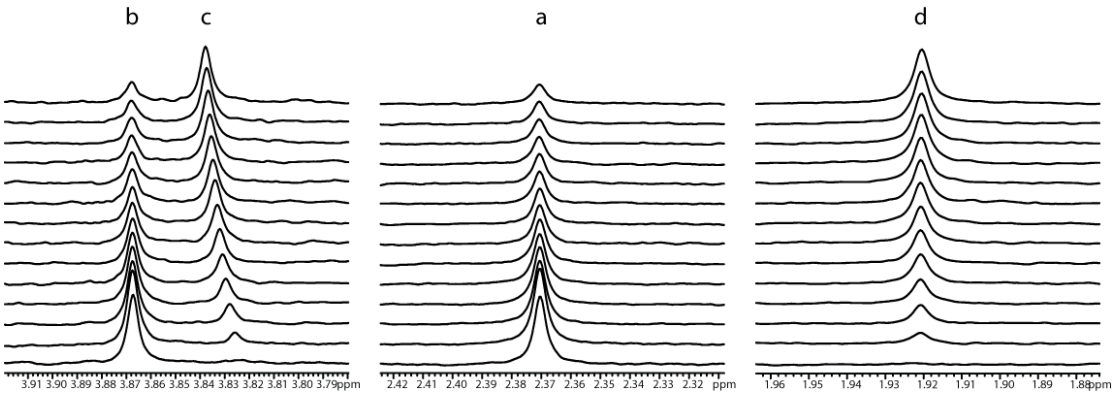


Figure 5.13 Array showing every 35th spectrum from the third (obtained after 3 min from acquisition) to the last for a 12 h measurement on the reaction of scheme 5.1. The signals a, b, c, and d represent the Hs in scheme 5.1.

5.3.2.1 Data Processing

All the spectra obtained were Fourier transformed, reference deconvoluted to a 1.5 Hz Lorentzian lineshape, and baseline corrected in Vnmr 6.1C. Data were then exported to Mathematica as a text file. All statistical analysis and fitting was performed in Mathematica version 6.0 [5.17]. All the peaks shown in figure 5.13 were fitted to mixed absorption and dispersion Lorentzian functions:

$$L(f) = \cos\left(\frac{\phi\pi}{180}\right)\left(\frac{aw}{2\pi(f-f_0)^2+(\frac{w}{2})^2}\right) - \sin\left(\frac{\phi\pi}{180}\right)\left(\frac{a(f-f_0)}{(f-f_0)^2+(\frac{w}{2})^2}\right)$$

(equation 5.18)

on a linear baseline, where ϕ, a, w, f_0 are the phase, amplitude, width at half height, and frequency of the signal respectively.

Iterative fitting of peaks in each spectrum of the ^{1}H timecourse array was done using the Nonlinear Regression function in Mathematica 6.0. For the first spectrum in the series, initial parameter values for the Lorentzian functions were estimated by inspection. From then on, each successive spectrum was fitted using starting parameters extrapolated from the preceding fits. This minimises the number of iterations required, reduces problems with signals that drift in frequency, and allows signal assignments to be retained even when the chemical shift ordering changes during a reaction.

Peak integral values (parameter *a* in equation 5.18) for each signal were obtained from the fitting process. In order to convert the integral values to concentrations, the integral of the reactant signals at time zero should be equalized to the initial concentration (0.009 M) used in the reaction to obtain the converting factor. The initial integral value for the reactant (R_0) was obtained by extrapolation of the fitted integrals as a function of time to the reactant formula of the following 1st order kinetic model:

$$R \xrightarrow{k} P$$

$$\frac{d[R]}{dt} = -k[R] \;\rightarrow R(t) = R_0 e^{-k(t+t_0)}$$
$$\frac{d[P]}{dt} = k[P] \;\rightarrow P(t) = (1 - e^{-k(t+t_0)})$$

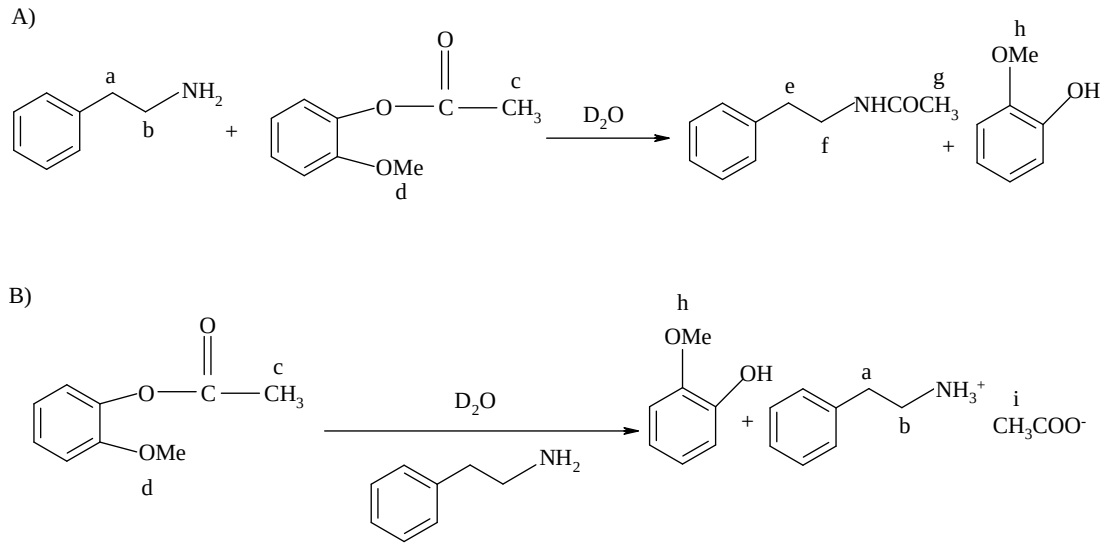
(equations 5.19)

As the initial experimental time is not zero, the t_0 is added to the model to shift the time to the real experimental starting time. The obtained integral R_0 was equated to the initial concentration value of the reactants in the reaction mixture (0.009 M) to convert

the experimental integral values to concentrations. The concentrations of the signals are plotted as a function of time in figure 5.18 (in section 5.4.1), and were fitted simultaneously to the reactant and product models of equations 5.19 to yield the rate constant.

5.3.3 Second Homogeneous Reaction

A mixture of 2 mL of 0.03 M 2-methoxyphenyl acetate in D_2O, 2 mL of 0.03 M phenylethylamine in D_2O, and 300 µL of 0.24 M TSP as the reference compound was prepared as a test reaction to evaluate the latest flow system with a lower total volume (section 5.2.3). As shown in scheme 5.2, two parallel reactions occur in this mixture: transesterification (scheme 5.2 A), and hydrolysis (scheme 5.2 B).



Scheme 5.2 Reactions of 2-methoxyphenyl acetate with phenylethylamine in D_2O. Reaction A is an amide synthesis, and B a hydrolysis, with reaction B being the faster. The methylene and methyl protons are assigned with lower-case letters from 'a' to 'i' for later NMR assignments.

The flow system was initially filled by pumping phenylethylamine and TSP solution (3.5 mL of 0.03 M phenethylamine in D_2O, and 300 µL of 0.24 M TSP) into the

system. After filling the system, the flow was stopped to lock and shim the magnetic field (a linewidth of 1.8 Hz was obtained for the TSP signal). Parameters were set to acquire a timecourse array of ^{1}H spectra with presaturation of the water signal. Saturation of the intense water signal allows the gain of the receiver to be increased, giving better signal-to-noise ratio for the remaining signals. The total duration of the experiment was 5 h, with a recycle delay of 7 s, 4 transients, a saturation delay of 4 s, a saturation power of 9 dB, and a 90° pulse width of 9 μs for each ^{1}H spectrum. The timecourse was obtained by arraying the number of transients. Once the experimental parameters were chosen, the solution inside the flow system was substituted with the reacting solution (mixture of phenylethylamine, 2-methoxyphenyl acetate, and TSP solutions in D_2O as described above). For this purpose the main solution (4.3 mL) was prepared in the reaction vessel and the input pipe was introduced into it. The solution was pumped at a flow rate of 1mL/min for 3.5 min (the total volume of the system is 3.5 mL) with the output sent to waste to empty the system of the mixture of phenylethylamine and TSP that was initially used for setting up the experiment. After 3.5 min the output pipe was transferred to the main vessel, and data acquisition started. The spectral timecourse array for the reaction is shown in figure 5.14.

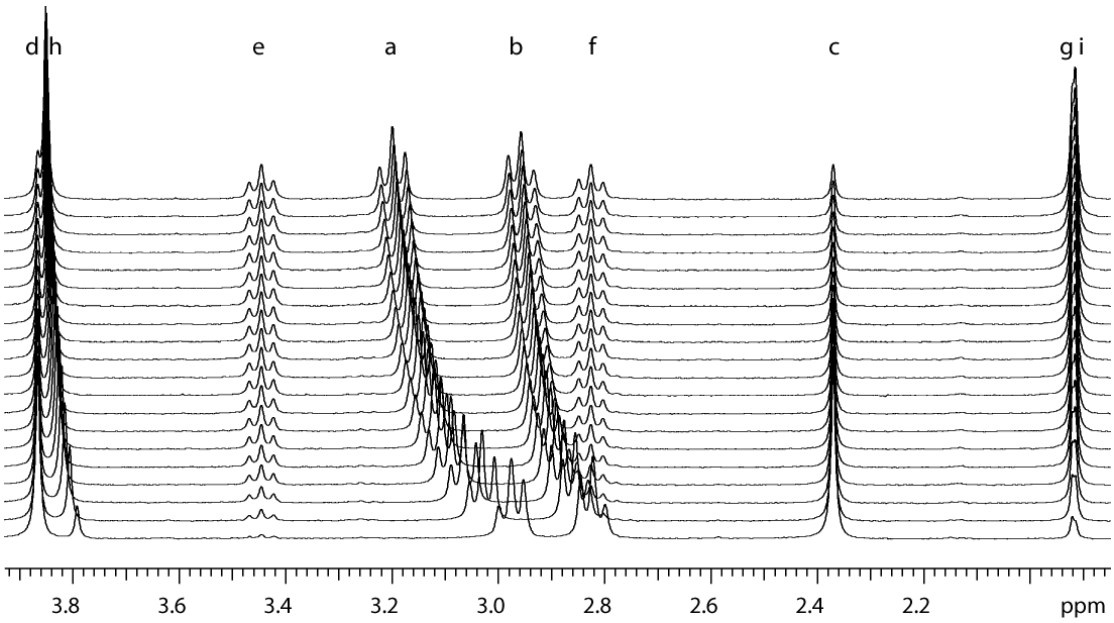


Figure 5.14 Timecourse array of the mixture of phenylethylamine and 2-methoxyphenyl acetate solution, shown for every 35th spectrum from the first (bottom) until the last (top). The lower-case letters from 'a' to 'i' correspond to the methyl and methylene signals highlighted in scheme 5.2. During the course of the reaction the chemical shifts of signals 'a', 'b' and 'h' change due to the change in pH as hydrolysis proceeds.

5.3.3.1 Data Processing

All the data obtained were Fourier transformed, reference deconvoluted using a 2.4 Hz Lorentzian target lineshape, normalised to correct systematic drifts in receiver gain, and baseline corrected in Vnmr 6.1C. Data were then exported to Mathematica as a text file. All statistical analysis and fitting processes were performed in Mathematica version 6.0.

All the peaks shown in figure 5.14 were fitted to a phased Lorentzian function. The singlet signals were fitted to a linear baseline plus equation 5.18 and the triplets were fitted to a linear baseline plus the following equation:

$$L(f)=\left[\cos\left(\frac{\phi\pi}{180}\right)\left(\frac{aw}{2\pi(f-f_0)^2+(\frac{w}{2})^2}\right)-\sin\left(\frac{\phi\pi}{180}\right)\left(\frac{a(f-f_0)}{(f-f_0)^2+(\frac{w}{2})^2}\right)\right]+$$
$$\left[\cos\left(\frac{\phi\pi}{180}\right)\left(\frac{0.5a(1+\alpha)w}{2\pi(f-f_0-j)^2+(\frac{w}{2})^2}\right)-\sin\left(\frac{\phi\pi}{180}\right)\left(\frac{0.5a(1+\alpha)(f-f_0-j)}{(f-f_0-j)^2+(\frac{w}{2})^2}\right)\right]+$$
$$\left[\cos\left(\frac{\phi\pi}{180}\right)\left(\frac{0.5a(1-\alpha)w}{2\pi(f-f_0+j)^2+(\frac{w}{2})^2}\right)-\sin\left(\frac{\phi\pi}{180}\right)\left(\frac{0.5a(1-\alpha)(f-f_0+j)}{(f-f_0+j)^2+(\frac{w}{2})^2}\right)\right]$$

(equation 5.20)

where $\phi, a, w, f_0, \alpha, j$ are the phase, amplitude, width at half height, frequency, balancing factor, and coupling constants for the signals respectively. In the triplet peaks (e, a, b, f in figure 5.14) the outer signals do not have equal amplitudes, because of strong coupling. The sum of the integrals of the outer signals in a triplet is approximately equal to the integral of the inner signal for a mild-strong coupling, so the amplitudes of the outer signals of each of the triplets can be defined as $0.5a(1+\alpha)$ and $0.5a(1-\alpha)$, where a is the amplitude of the inner signal. The outer signals are assumed to be spaced *j* Hz at either side of the central component, i.e. it is assumed that the strong coupling is sufficiently mild to affect the amplitudes of the triplet signals but not their positions.

The four triplet signals in figure 5.14 were fitted simultaneously from the last to the first spectrum. This reversal of order was done because at the beginning of the experiment the triplet f is under triplet b (figure 5.14), so it was simpler to start the fitting from the last spectrum where the triplet f is completely separate. As the

behaviour of f is exactly the same as e, the amplitude, width, and asymmetry of f was forced to be the same as e.

The integral values (amplitude in the Lorentzian equation for the singlet signals and $2\times$amplitude for the triplet signals) of each of the peaks were obtained from the fitting process. The initial integral values of the reactants would normally be obtained by extrapolating the fitted integrals as a function of time to an appropriate kinetic model. However, because an analytical solution for parallel 2nd order kinetics was not available, the integrals were fitted to the following 2nd order kinetics model for reactant (R) and product (P) to obtain the initial integral values:

$$2R \xrightarrow{k} 2P$$

$$\frac{d[R]}{dt} = -k[R]^2 \rightarrow R(t) = \frac{R_0}{1 + R_0 k(t + t_0)} \qquad \text{(equation 5.21)}$$

$$\frac{d[P]}{dt} = k[P]^2 \rightarrow P(t) = \frac{{R_0}^2 kt}{1 + R_0 k(t + t_0)} \qquad \text{(equation 5.22)}$$

The kinetics of the two 2nd order parallel reactions in scheme 5.2 can be described as follows:

$$A_1 + B \xrightarrow{k_1} C + D$$

$$B \xrightarrow[A_1\text{:catalyst}]{k_2} C + A_2 E \qquad \text{(scheme 5.3)}$$

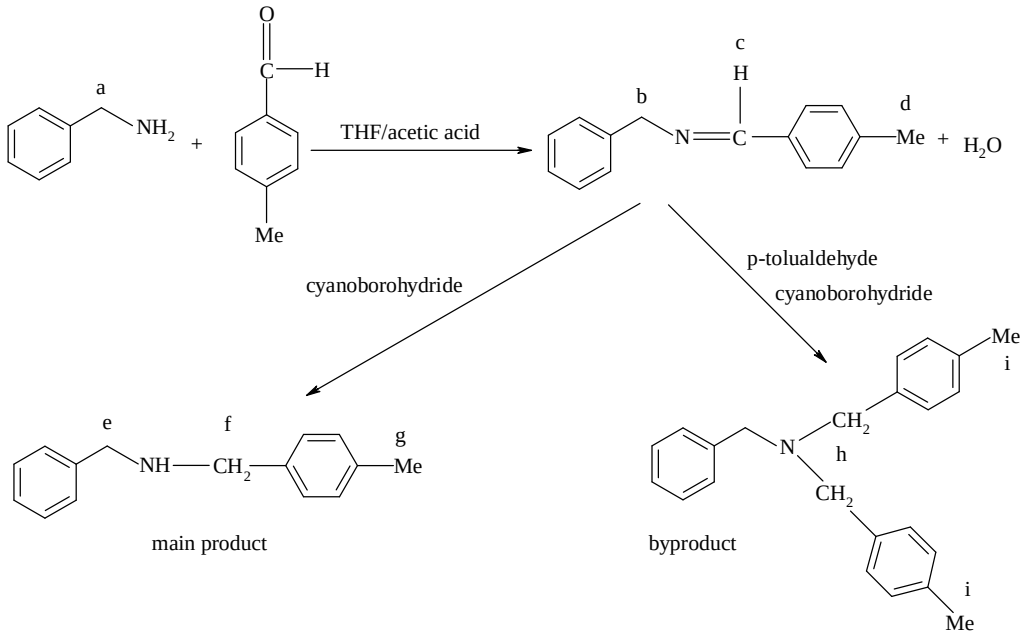
A_1: amine, A_2: protonated amine, B: ester, C: phenol, D: amide, E: acetate

Equations 5.23 describe the kinetic model for the 2nd order parallel reactions of scheme 5.3. As Mathematica 6.0 is unable to give an analytical solution to these equations, the data were fitted to this model numerically. The concentration of amine (A_1) shown in scheme 5.3 is equal to the initial concentration of the amine minus the concentration of the aminium ion (A_2): $A_1 = A - A_2$. The concentration of A_2 is assumed to be equal to the concentration of E (acetate in scheme 5.2), so $A_1 = A - E$.

$$\frac{d[A]}{dt} = -k_1([A]-[E])[B]$$
$$\frac{d[B]}{dt} = -k_1([A]-[E])[B] - k_2([A]-[E])[B]$$
$$\frac{d[C]}{dt} = k_1([A]-[E])[B] + k_2([A]-[E])[B] \qquad \text{(equations 5.23)}$$
$$\frac{d[D]}{dt} = k_1([A]-[E])[B]$$
$$\frac{d[E]}{dt} = k_2([A]-[E])[B]$$

5.3.4 A Heterogeneous Reaction

Heterogeneous reactions are difficult to study in static conditions in an NMR tube. As the magnetic susceptibilities of solids and liquids are normally different, the homogeneity of the magnetic field is disturbed. One of the main advantages of the flow system is the ability to study heterogeneous reactions while avoiding having solid particles in the detection region of the probe. A reductive amination of an aldehyde was studied as a test heterogeneous reaction. As shown in scheme 5.4, the reaction was between benzylamine and p-tolualdehyde in THF/acetic acid with cyanoborohydride on a polymer support. This reaction was repeated several times with the amine in excess, in an attempt to minimize the side product of dialkylated amine [5.18]. However, the production of side products could not be completely avoided. The final reaction studied was a mixture of 2.43 mL of 0.26 M benzylamine in THF, 2.43 mL of 0.13 M p-tolualdehyde in THF, 0.34 mL acetic acid, 1.82 mL THF, 434 mg cyanoborohydride and 20 drops TMS (scheme 5.4).



Scheme 5.4 Reductive amination of benzylamine with p-tolualdehyde, involving para-dimethyl benzyl benzylimine as an intermediate. The imine is reduced to para-dimethyl benzyl benzylamine as the main product and 2-(para-dimethyl benzyl) benzylamine as the byproduct (but this was not confirmed). The protons for which signals are assigned in the spectra below are marked with lower-case letters.

As the ^{13}C satellites of the THF signals overlap with other signals, broadband ^{13}C decoupling was used during acquisition. The parameters used to acquire a timecourse array of ^{1}H spectra while presaturating one of the THF signals and decoupling the ^{13}C satellites are as follows. The total duration of the experiment was 6 h 19 min, with a recycle delay of 8 s, number of transients of 16, saturation delay 3 s, saturation power 5 dB, and a 30° flip angle (pulse width of 3 μs) was used to minimise signal saturation. The timecourse was obtained by arraying the number of transients. Once the experimental parameters were set, the reaction solution (7 mL) was pumped into the emptied and dried flow system at a flow rate of 1 mL/min. A frit was attached to the input pipe to avoid the solid reducing agent entering the flow system, while the output pipe was placed in the main reaction vessel. The latter was placed on a

magnetic stirrer to stir the solids. The acquisition was started when the first drop came out of the output pipe (ensuring that the flow system was full of reaction solution). The timecourse array of the reaction is shown in figure 5.15.

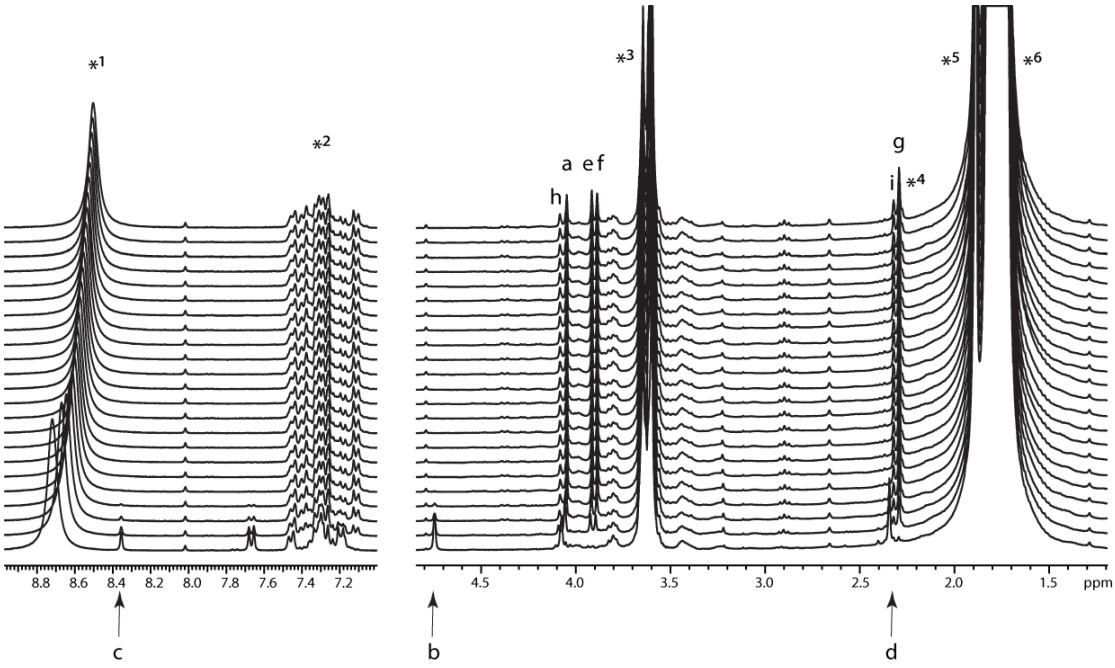


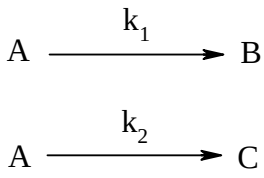
Figure 5.15 Timecourse array of the mixture of benzylamine and p-tolualdehyde solution; spectra are shown for every 10th spectrum from the first (bottom) until the last (top) spectrum. The lower-case letters from 'a' to 'i' correspond to proton signals highlighted in scheme 5.4. The star signals are: *^{1}: water in exchange with carboxyl proton of acetic acid, *^{2}: aromatics, *^{3}: partially saturated THF, *^{4}: impurity, *^{5}: methyl of acetic acid, *^{6}: unsaturated THF signal. The signals from the reactant aldehyde are not observed as the initial step of the reaction is very fast. One of the methylenes of the byproduct is not assigned as it is probably overlapped by another signal. There are lots of impurity signals observed in this timecourse. The effect of initial temperature equilibration can be seen in the moving signal of *^{1}; the water signal is particularly temperature-sensitive because of the comparatively low strength of the hydrogen bond.

The water is observed at unexpectedly high chemical shift, because it is in rapid exchange with the carboxylic acid proton of acetic acid. At the beginning of the

timecourse of the reaction the shape and frequency of signal *[1] changes rapidly as the sample temperature and temperature gradient change, but stabilises later.

5.3.4.1 Data Processing

All the initial data processing and exporting was done as described in section 5.3.3.1. The methyl peaks (d, g, i) were fitted to baseline-corrected Lorentzian lineshapes (equation 5.19), and the integral values obtained were plotted as a function of time. The integrals were fitted to first order parallel kinetics.



(Scheme 5.5)

A: imine, B: main product, C: byproduct

Analytical solution of equations 5.24 produces a model describing the changes of A, B, and C with respect to time (equation 5.25).

$$\frac{d[A]}{dt} = -k_1[A] - k_2[A]$$

$$\frac{d[B]}{dt} = k_1[A] \qquad \text{(equations 5.24)}$$

$$\frac{d[C]}{dt} = k_2[A]$$

$$A = A_0 e^{-(k_1+k_2)t}$$
$$B = A_0\left(-\frac{k_1(-1+e^{-(k_1+k_2)t})}{k_1+k_2}\right) \qquad \text{(equations 5.25)}$$
$$C = A_0\left(-\frac{k_2(-1+e^{-(k_1+k_2)t})}{k_1+k_2}\right)$$

5.3.5 A High Temperature Reaction

The flow system has been tested with homogeneous and heterogeneous reactions at ambient temperature, therefore, the possibility of investigating a higher temperature reaction using room temperature flow sampling was investigated. The reaction studied was the acid hydrolysis of clarithromycin ($C_{38}H_{69}NO_{13}$, a macrolide antibiotic) [5.19], with only the reaction vessel heated. 7 mL of a 0.1 M HCl / KCl buffer solution (pH=1.2) in D_2O, and 0.31 mL of 0.24 M TSP in D_2O was prepared in the reaction vessel. The locking and shimming process was carried out with the buffer and TSP solution in the system. The total duration of the experiment was set to 8 h, with a recycle delay of 6.7 s, the number of transients for the first 50 spectra was set to 4, 16 for the second 50 spectra, 64 for the third 50 spectra, and 256 for the last two spectra (allowing more rapid data acquisition at the beginning of the reaction, which is faster), saturation delay of 3.7 s, saturation power of 5 dB, and a 90° pulse width of 9 μs for each ^1H spectrum. Then, the system was emptied of the buffer and the TSP solution. This solution was added to 21 mg solid clarithromycin in the reaction vessel, stirred and was placed on a hot plate at 40°C. The solution was pumped (1 mL/min) into the system, with the output pipe in the reaction vessel. As the first drop of the solution came out of the output pipe, the acquisition was started. The timecourse array of this reaction is shown in figure 5.16.

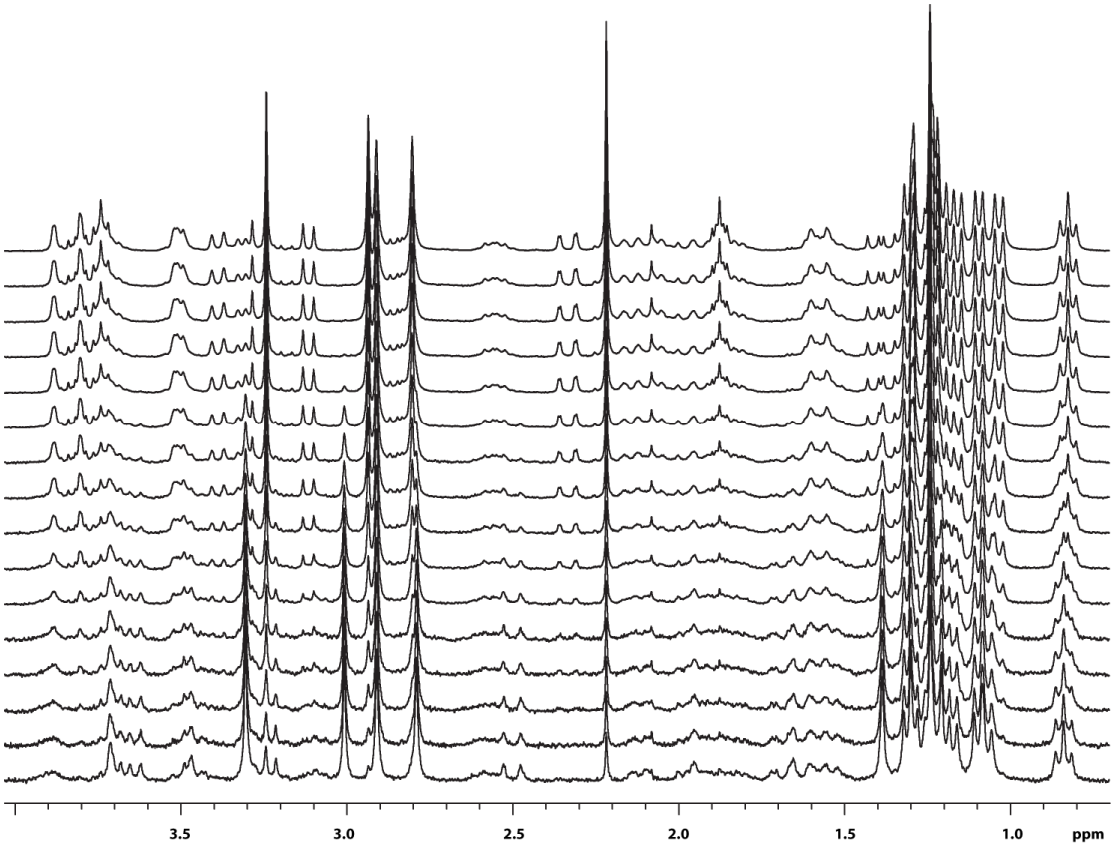


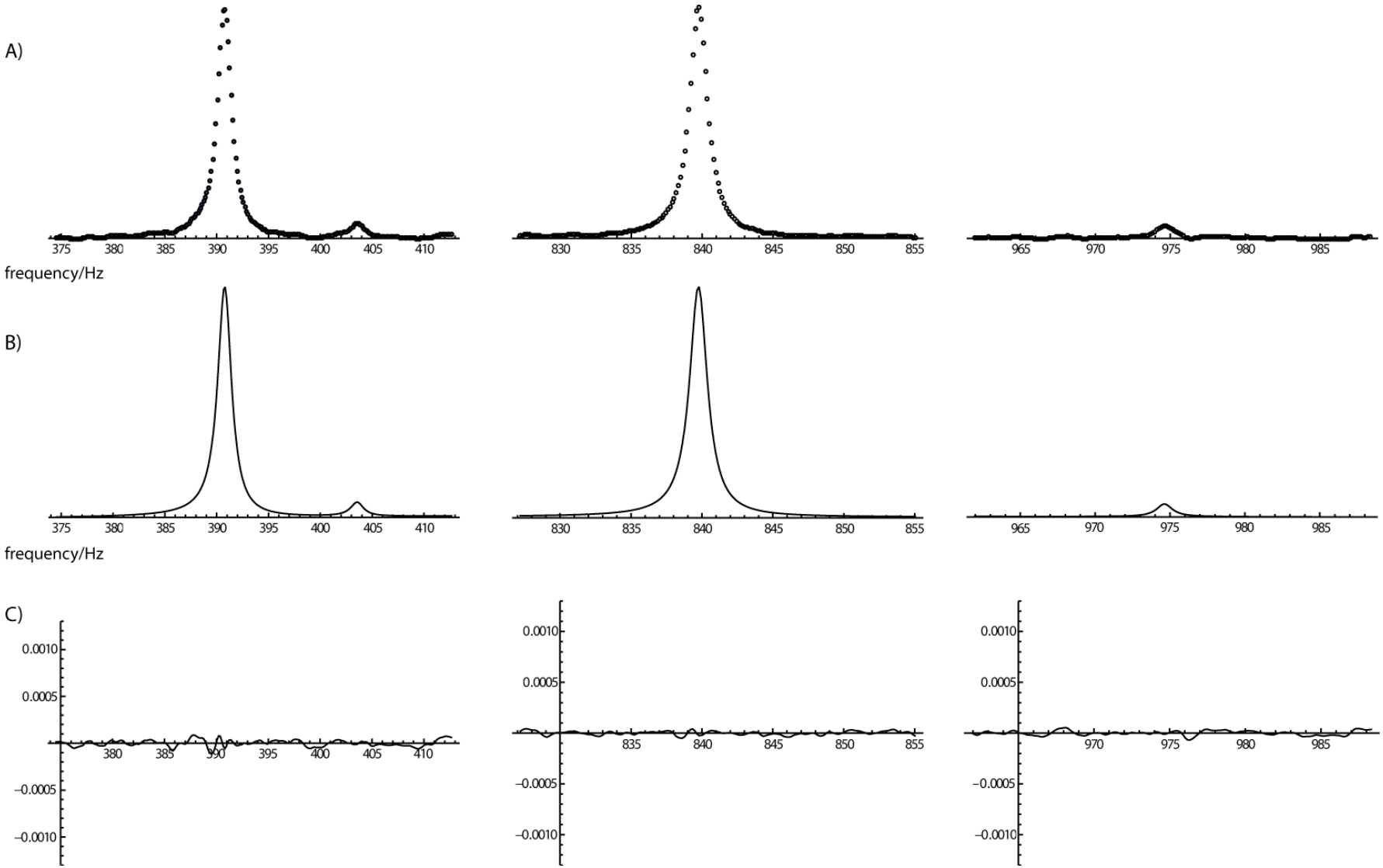
Figure 5.16 Timecourse array for clarithromycin hydrolysis, showing every 10th spectrum from the first (bottom) to the last (top).

5.4 Results and Discussion

The spectrum (figure 5.11 in section 5.3.1) resulting from the first experiment done with flow NMR, where a concentric PEEK tube was used in a normal NMR tube proved the possibility of obtaining an acceptably homogeneous magnetic field for a flowing sample in the NMR cell. The spectral quality achieved was comparable to that routinely obtained with samples in normal NMR tubes.

5.4.1 First Homogeneous Reaction

The signals in figure 5.13 (in section 5.3.2) were fitted to a Lorentzian lineshape (equation 5.18). The experimental and the fitted signals, and the difference between them (residuals) are shown for the first and the last spectrum in figure 5.17. The integrals of the fitted signals were converted to concentrations. The plots of these data as a function of time (kinetic plots) are shown in figure 5.18, with fits to first order kinetics. The concentrations of the reactants in about the first 40 minutes of the experiment are continuously oscillating. In this experiment the flow system was initially filled with the solvent/buffer, and then the reactant (2-methoxyphenyl acetate) was added to the reaction vessel. Initially the reactant was not homogeneously mixed with the solution in the system, causing oscillations as higher and lower concentrations flow alternately through the probe active volume. This inefficient mixing can be prevented either by discarding the initial solution in the system, allowing the composition of a homogeneous solution to be monitored from the beginning, or by filling the system with the reacting solution from empty.


A)
375 380 385 390 395 400 405 410
830 835 840 845 850 855
965 970 975 980 985
frequency/Hz
B)
375 380 385 390 395 400 405 410
830 835 840 845 850 855
965 970 975 980 985
frequency/Hz
C)
0.0010
0.0005
−0.0005
−0.0010
380 385 390 395 400 405 410
835 840 845 850 855
970 975 980 985

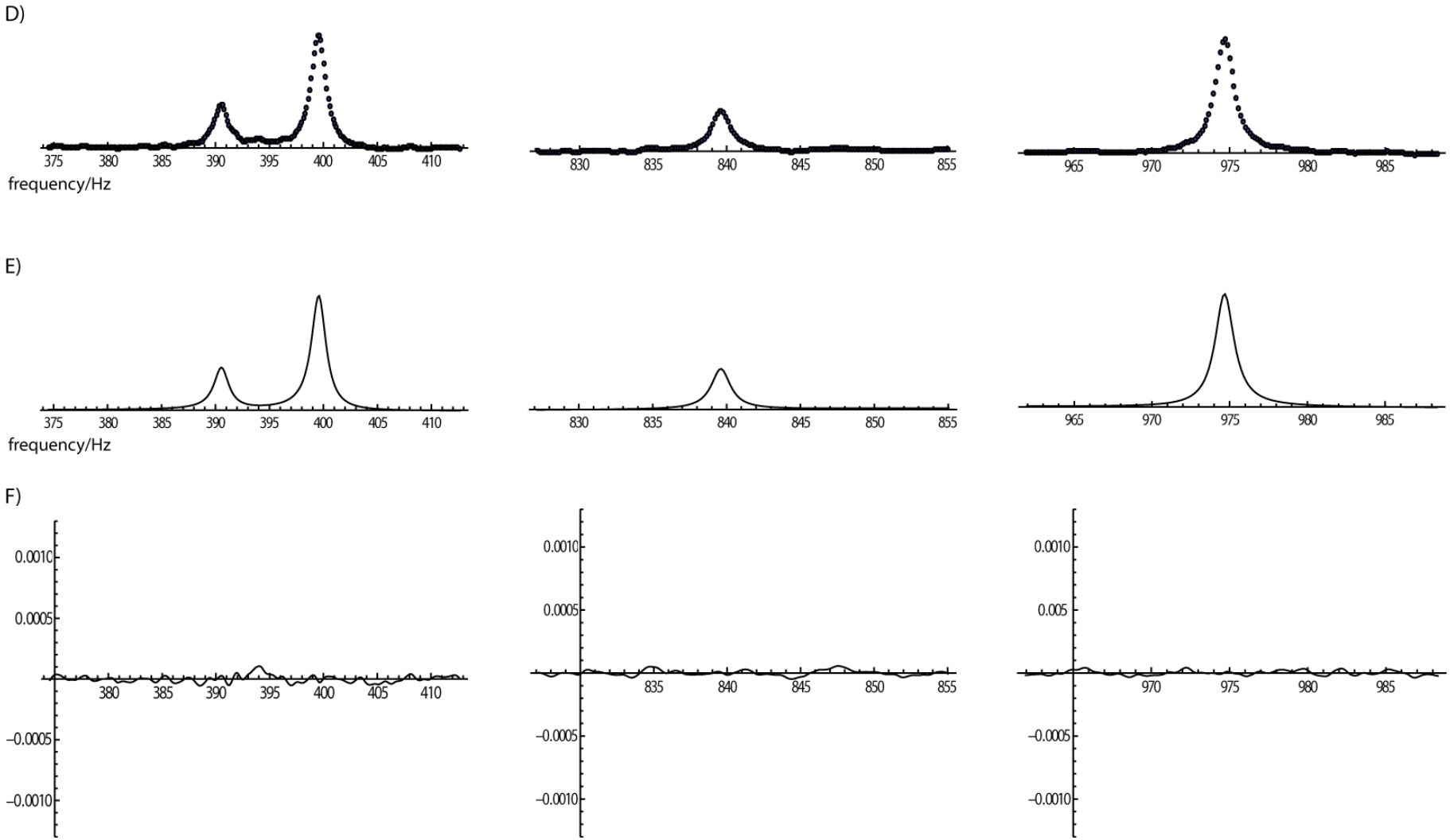


Figure 5.17 Segmented plots for the data of figure 5.13. Rows 'A' to 'C' are the 1st experimental spectrum, the fit to Lorentzian lineshapes, and the residuals (vertical scale showing the relative difference in amplitude) respectively. Rows 'D' to 'F' are the corresponding plots for the last experimental spectrum.

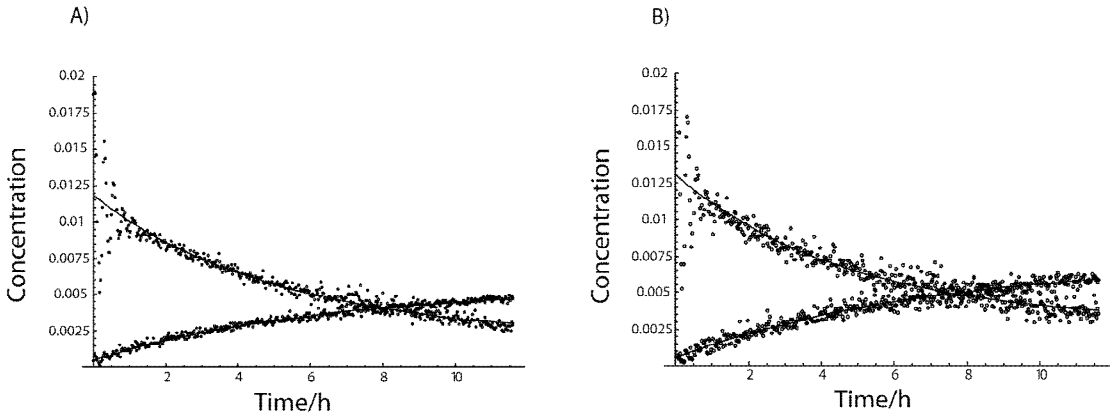
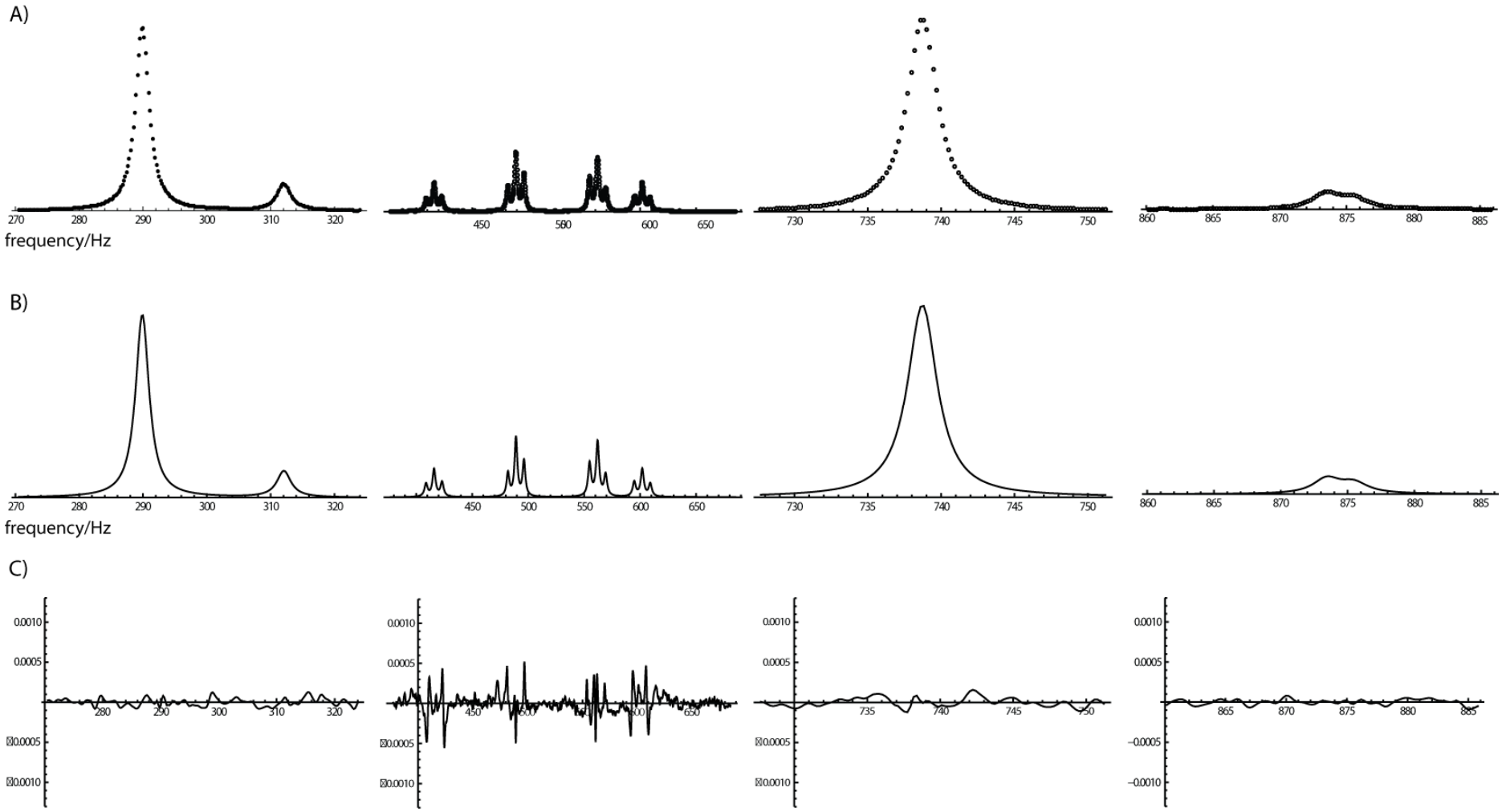


Figure 5.18 Experimental (dots) and fitted (solid lines) kinetic plots of reactant and product acetate (A) and methoxy (B) signals in scheme 5.1. Experimental data are fitted to 1st order kinetics (equations 5.19), yielding a rate constant of $2.22 \pm 0.02 \times 10^{-5}$ s^{-1}. The oscillations at the beginning of the reaction are due to inefficient mixing; these data were omitted from the fitting.

5.4.2 Second Homogeneous Reaction

As was explained in section 5.3.3.1, all signals of figure 5.14 were fitted to Lorentzian lineshapes. The experimental peaks, fitted peaks, and residuals of the first and the last spectrum of the ^1H timecourse array are shown in figure 5.19 in rows A to F respectively. As is shown in the residual plots (rows C and F), the residuals for the singlet peaks are small and random, showing the quality of the fitting. However, the residual plots for the four triplet signals in the range 400-650 Hz show a bigger systematic error between the experimental data and the fitted data in this region. For such an A_2B_2 spin system it is only an approximation to represent the spectrum as asymmetric triplets, hence the larger residuals for this region are seen.


A)
270
280
290
300
310
320
450
580
600
650
730
735
740
745
750
860
865
870
875
880
885
frequency/Hz
B)
270
280
290
300
310
320
450
500
550
600
650
730
735
740
745
750
860
865
870
875
880
885
frequency/Hz
C)
0.0010
0.0005
280
290
300
310
320
0.0005
0.0010
0.0010
0.0005
450
650
0.0005
0.0010
0.0010
0.0005
735
740
745
750
0.0005
0.0010
0.0010
0.0005
865
870
875
880
885
−0.0005
−0.0010

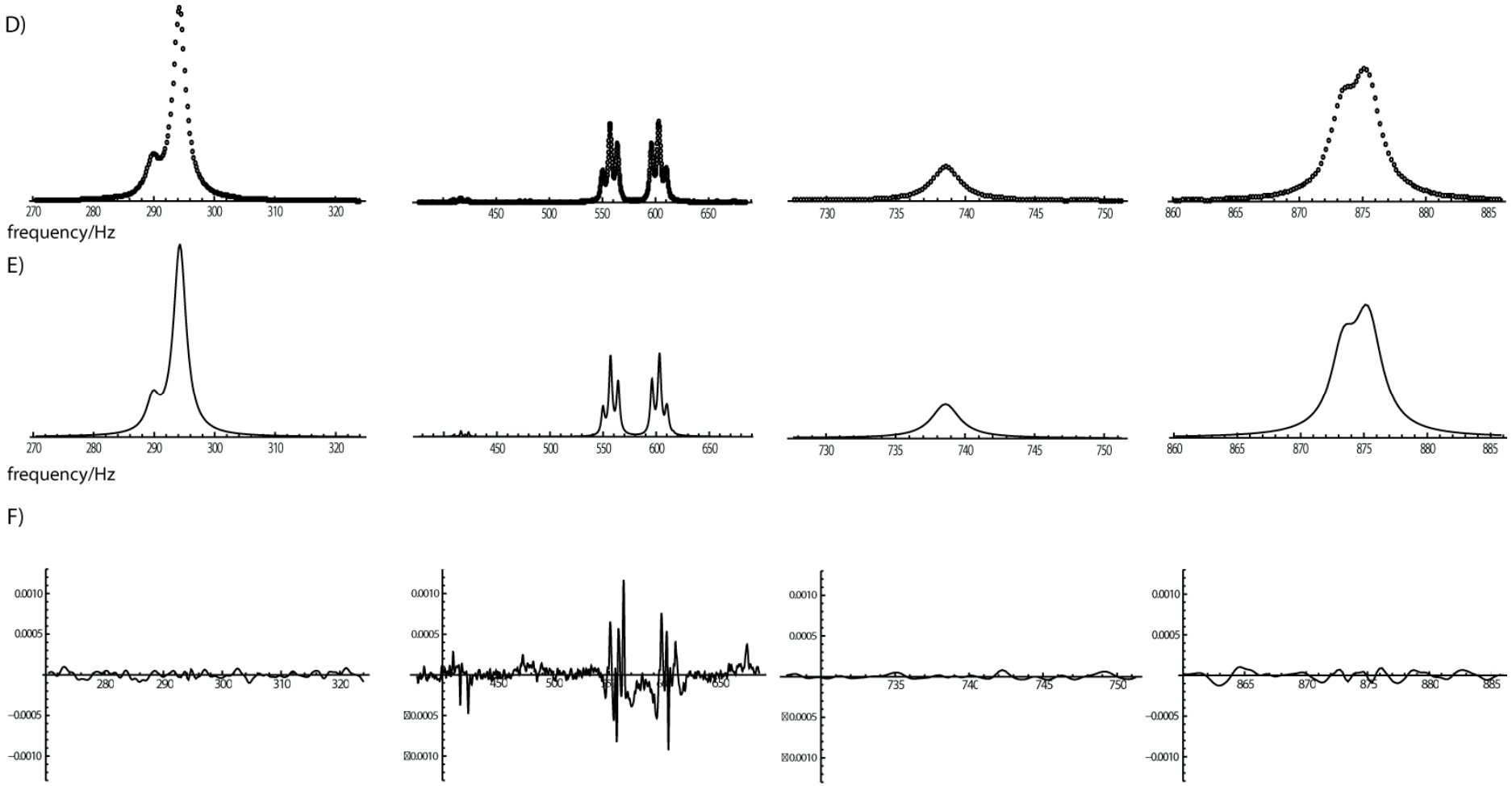


Figure 5.19 Segmented plots of figure 5.14. Rows 'A' to 'C' are the 1st experimental spectrum, their fitting to Lorentzian lineshape, and their residuals (vertical scale showing the relative difference in amplitude) respectively. Rows 'D' to 'F' are the last experimental spectrum, their fitting to Lorentzian lineshape, and their residuals respectively.

The plots of concentration values as a function of time (kinetic plots) with the fitted kinetics to a numerical 2^{nd} order parallel model (equations 5.23) for signals ‘d’, ‘h’, ‘b’, ‘f’, and ‘i’ of figure 5.14 are shown in figure 5.20.

At the beginning of the reaction the output pipe was out of the reaction vessel for 3.5 min, with the flow rate of 1 mL/min (total volume of the system is 3.5 mL), to empty the flow system from the solution used for the experiment initialization. During the first 3.5 min, while the initial solution is being removed from the system, it is also being mixed with the inflowing reaction solution. Due to diffusion, laminar flow, and dead volumes, the initial period of 3.5 min is not enough to completely flush the original solution from the system. Therefore, some residual oscillation in the data points for the first 10 min of the reaction can be seen in figure 5.20. This problem can be prevented by increase the discarding time, or starting the reaction with an empty flow system.

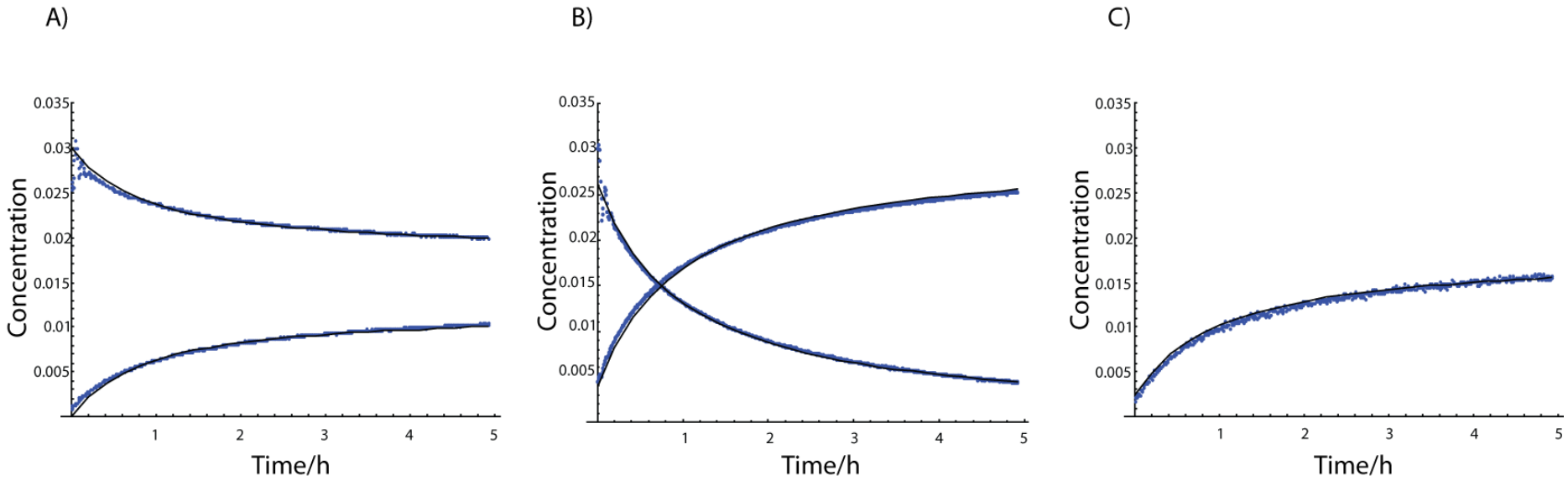


Figure 5.20 Experimental (dots) and fitted (solid lines) kinetic plots for reactants and products in scheme 5.2. A) Kinetic plot for phenethylamine (reactant) and N-phenethyl acetamide (product), marked as signals 'b' and 'f' in figure 5.14. B) Kinetic plot of 2-methoxyphenyl acetate (reactant) and 2-methoxy phenol (product), marked as signals 'd' and 'h' in figure 5.14. C) Kinetic plot of acetate marked as signals 'i' in figure 5.14. Optimised values of k_1 and k_2 were 4.0×10^{-3} L mol^{-1} s^{-1} and 5.9×10^{-3} L mol^{-1} s^{-1} respectively. The effects of incomplete mixing are observed in the oscillations in the reactant curves A and B for about the first 10 min of the reaction; these data were omitted from the fitting.

5.4.3 A Heterogeneous Reaction

Data from the region 2.25-2.48 ppm in figure 5.15 were fitted to 4 Lorentzian lineshapes (figure 5.21) as explained in section 5.3.4.1. The impurity signal in this region (*[4] signal in figure 5.15) was also fitted, as it overlaps with the methyl signals (d, i, g in figure 5.15). The integrals of the methyl signals were plotted against time (figure 5.22).

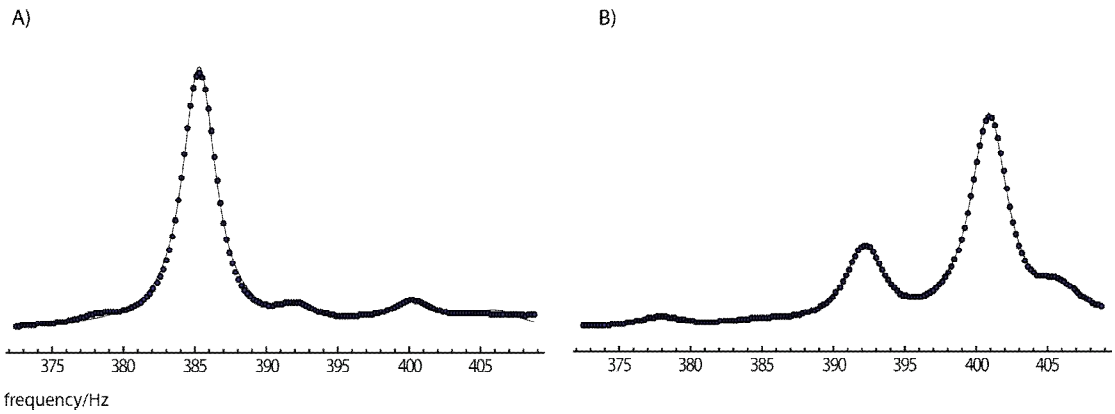


Figure 5.21 Fitting showing only the first (A) and the last (B) methyl region spectra (signals 'd', 'i', 'g', *[4]) of figure 5.15. The dots represent the experimental data, and the solid lines the fits to 4 Lorentzian lineshapes.

The data analysis for this reaction is complicated by the occurrence of byproducts in the reaction mixture and by the overlapping signals. The result of fitting the methyl signal to first order parallel kinetics (equation 5.25) is shown in figure 5.22.

As is shown in figure 5.22, the initial concentration of the reactant is not oscillating. This proves that by starting the reaction with an empty flow system the problem of

incomplete mixing that was observed in figures 5.18 and, to a lesser extent, 5.20 can be avoided.

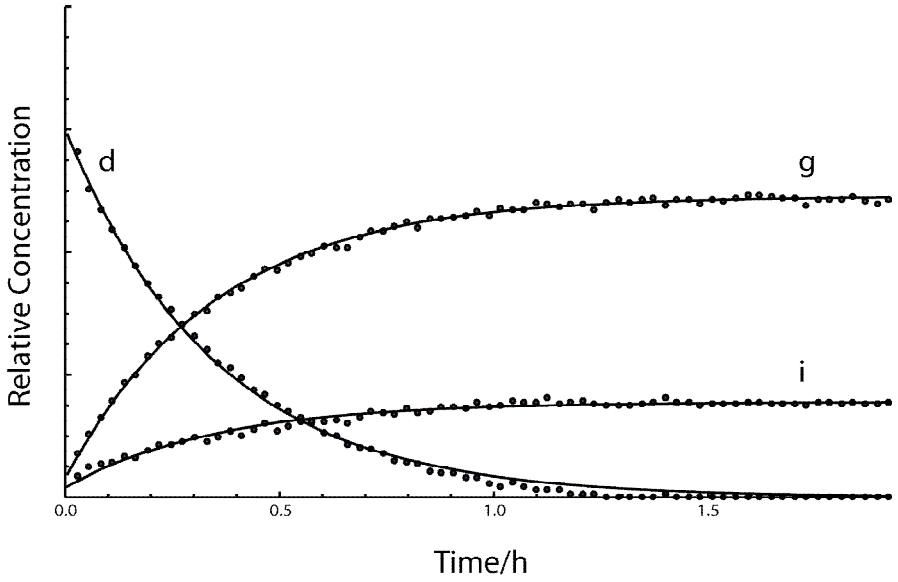


Figure 5.22 Kinetic plot for the signals 'd', 'i', 'g' in figure 5.15. The dots represent the experimental data, which are fitted to a 1st order parallel kinetic model (solid lines), yielding a k_1 of 1.86 ± 0.03 × 10^{-4} L mol^{-1} s^{-1} and k_2 of 6.08 ± 0.04 × 10^{-4} L mol^{-1} s^{-1}.

5.4.4 A High Temperature Experiment

The timecourse data obtained with a heated reaction vessel but the flow system at ambient (figure 5.16) show that this system can be used for reactions that require high temperature, albeit at the price of perturbing the reaction kinetics if the sampled reaction mixture is recycled into the reaction vessel. The mixture in the reaction vessel is at 40°C, but is cooled to ambient temperature while entering the magnet.

5.5 Future Work

A variety of different reactions were studied with the flow system. However, there are still different experiments and improvements that could be implemented, such as:

- Study a reaction mixture at various flow rates to investigate the effect of flow on signal intensity.
- Measure T_1 values of nuclei in non-flow and flow experiments at different rates. This can allow finding the effect of the different flow rates on the relaxation time.
- Investigate the pressure limits of the system.
- Limit the pressure increase by using a small pressure relief valve, resistant to organic solutions. Connect the valve to the input pipe on top of the NMR tube outside of the magnet by a long T connection (3 way union); under normal conditions there will be no contact between the solution and the valve. The release pressure of the valve is set higher than the maximum pressure expected at the NMR tube and lower than the pressure that the tube can handle. In the case of the pressure limit being exceeded, the valve opens and the liquid exits through the pressure relief valve, limiting the pressure in the NMR tube.
- Measure spectra of nuclei other than ^{1}H (e.g. ^{19}F) where the large signal from solvent would not interfere with the rest of the signals (solvent suppression techniques would not be required).
- Lag or thermostat the transfer pipes outside the magnet, to make a homogeneous temperature system.

Chapter 6

Summary

Summary

The work presented in this thesis describes kinetic studies of chemical reactions by novel techniques using time-resolved NMR spectroscopy. The aim of the work was to be able to study the kinetics of reactions by sophisticated and cost-efficient methods. This involved studying kinetics by least squares fitting of experimental data for reactions under static and flow condition, and by multi-way PARAFAC (PARAllel FACtor) analysis.

Kinetic studies of chemical reactions were performed by acquiring a series of ^{1}H spectra and fitting the signals by an iterative least squares method to a model which represents the shape of the signals (e.g. Lorentzian or Gaussian). For this purpose, the initial parameters of the model are given for the individual signals in the first spectrum, and during the iteration process the parameters required to fit subsequent signals are obtained by extrapolation. This continuity allows monitoring mobile signals whose frequencies vary during the course of a reaction. Peak integrals can be fitted to the appropriate kinetic model to obtain the rate constant of the reaction. If there are overlapping signals in the spectra, it can be difficult to distinguish between individual peaks; other NMR experiments (e.g. 2DJ spectroscopy) can be used to assist assignments.

A time course of DOSY experiments, yielding amplitudes as a function of diffusion weighting, frequency, and time, produces a three-dimensional trilinear dataset, as each of the dimensions is independent from the others. These data and the number of components can be used as an input for PARAFAC analysis. Without making any assumptions about the diffusion behaviour, frequency or time dependence of signals,

PARAFAC is able to identify the diffusion profile, spectrum, and kinetics for each of the components. In this work, PARAFAC used alternating least squares fitting to find the minimum difference between the input and output data. Kinetic studies by this method are particularly useful as the challenge of identification of overlapping signals is avoided. No matter how many signals overlap, and how many components are present, as long as the trilinearity holds within an acceptable signal-noise-ratio and the individual component modes (e.g. diffusion, timecourse) are sufficiently different, PARAFAC can decompose the data. Analysing data with very poor signal-to-noise ratio (SNR) proved the remarkable robustness of PARAFAC to SNR values at which conventional methods would fail. It was shown that slight deviations from trilinearity of the data, for example due to imperfect experimental conditions (e.g. frequency changes due to high concentration of the reaction mixture, or slight temperature variation during the reaction) can affect the output data by showing additional signals, cross-talk, in the spectra, or incorrect diffusion coefficients. In some reactions, near-degeneracy in one dimension (e.g. the similar values of diffusion coefficients for maltose and maltotriose) can perturb the analysis, in which case applying constraints such as non-negativity can provide sufficient additional information for the algorithm to solve the problem. However, if the degeneracy is observed in more than one dimension PARAFAC may not be able to decompose data even when constraints are applied.

Resolution in NMR spectra is dependent on the homogeneity of the static magnetic field in the sample. The presence of materials in the sample with different magnetic susceptibilities will create internal field gradients that can result in broadening of the signals, with the consequent losses in resolution. Therefore, it is undesirable to study

the kinetics of heterogeneous reactions by conventional static methods in which an NMR tube is placed in a magnet. The flow system constructed in this work uses concentric HPLC tubing positioned concentrically in a normal NMR tube, allowing solution to enter and leave the active volume. In this system any solid particles in a heterogeneous reaction remain in the reaction vessel outside of the magnet and do not enter the NMR tube to perturb the magnetic homogeneity, allowing heterogeneous reactions to be studied with high resolution NMR. This flow system can also be used for homogeneous reactions in flow studies, and/or for experiments in which the reaction composition needs to be manipulated during the reaction.

References

Chapter 1

1.1. Bloch, F., Nuclear Induction. *Physical Review* **1946,** 70, (7-8), 460.

1.2. Purcell, E. M.; Torrey, H. C.; Pound, R. V., Resonance Absorption by Nuclear Magnetic Moments in a Solid. *Physical Review* **1946,** 69, (1-2), 37.

1.3. Andrew, E. R., A historical review of NMR and its clinical applications. *British Medical Bulletin* **1984,** 40, (2), 115-119.

1.4. Levitt , M., *Spin Dynamics: Basics of Nuclear Magnetic Resonance* Wiley: Chichester, 2001.

1.5. Nilsson, M.; Khajeh, M.; Botana, A.; Bernstein, M. A.; Morris, G. A., Diffusion NMR and trilinear analysis in the study of reaction kinetics. *Chemical Communications* **2009**, 1252-1254.

Chapter 2

2.1. Levitt, M. H., *Spin Dynamics: Basics of Nuclear Magnetic Resonance*. John Wiley & Sons Ltd: 2001.

2.2. Keeler, J., *Understanding NMR Spectroscopy* John Wiley & Sons Ltd: Chichester, 2005.

2.3. Bloch, F., Nuclear Induction. *Physical Review* **1946,** 70, (7-8), 460.

2.4. Bodenhausen, G.; Kogler, H.; Ernst, R. R., Selection of coherence-transfer pathways in NMR pulse experiments. *Journal of Magnetic Resonance*, **1984,** 58, (3), 370-388.

2.5. Keeler, J.; Clowes, R. T.; Davis, A. L.; Laue, E. D.; Thomas, L. J.; Norman, J. O., [4] Pulsed-field gradients: Theory and practice. *Methods in Enzymology* **1994,** 239, 145-207.

2.6. Parella, T., Pulsed field gradients: A new tool for routine NMR. *Magnetic Resonance in Chemistry* **1998,** 36, (7), 467-495.

2.7. Darbeau, R. W., Nuclear Magnetic Resonance (NMR) Spectroscopy: A Review and a Look at Its Use as a Probative Tool in Deamination Chemistry. *Applied Spectroscopy Reviews* **2006,** 41, (4), 401 - 425.

2.8. Kuchel, P. W.; Chapman, B. E.; Bubb, W. A.; Hansen, P. E.; Durrant, C. J.; Hertzberg, M. P., Magnetic susceptibility: Solutions, emulsions, and cells. *Concepts in Magnetic Resonance Part A* **2003,** 18A, (1), 56-71.

2.9. Durrant, C. J.; Hertzberg, M. P.; Kuchel, P. W., Magnetic susceptibility: Further insights into macroscopic and microscopic fields and the sphere of Lorentz. *Concepts in Magnetic Resonance Part A* **2003,** 18A, (1), 72-95.

2.10. Barjat, H.; Chilvers, P. B.; Fetler, B. K.; Horne, T. J.; Morris, G. A., A Practical Method for Automated Shimming with Normal Spectrometer Hardware. *Journal of Magnetic Resonance* **1997,** 125, (1), 197-201.

2.11. Edison, A. S.; Markley, J. L.; Weinhold, F., Calculation of nuclear spin-spin coupling constants with ab initio molecular orbital wave functions. *Journal of Physical Chemistry* **1993,** 97, (45), 11657-11665.

2.12. Wu, A.; Cremer, D., Extension of the Karplus Relationship for NMR Spin-Spin Coupling Constants to Nonplanar Ring Systems: Pseudorotation of Tetrahydrofuran. *International Journal of Molecular Sciences* **2003,** 4, (4), 158-192.

2.13. Karplus, M., Contact Electron-Spin Coupling of Nuclear Magnetic Moments. *Journal of Chemical Physics* **1959,** 30, (1), 11-15.

2.14. Freeman, R., *Magnetic Resonance in Chemistry and Medicine*. Oxford University Press: 2003.

2.15. Wittebort, R. J.; Szabo, A., Theory of NMR relaxation in macromolecules: Restricted diffusion and jump models for multiple internal rotations in amino acid side chains. *Journal of Chemical Physics* **1978,** 69, (4), 1722-1736.

2.16. Garner, J.; Hill, T.; Odell, L.; Keller, P.; Morgan, J.; McCluskey, A., Identification of Aminopyrimidine Regioisomers via Line Broadening Effects in ^{1}H and ^{13}C NMR Spectroscopy. *Australian Journal of Chemistry* **2004,** 57, (11), 1079-1083.

2.17. Mikhin, N. P., NMR line broadening in ^{3}He-^{4}He heterophase solid mixtures below 50 mK. *Low Temperature Physics* **2004,** 30, (6), 429-431.

2.18. Morris, G. A.; Barjat, H.; Home, T. J., Reference deconvolution methods. *Progress in Nuclear Magnetic Resonance Spectroscopy* **1997,** 31, (2-3), 197-257.

2.19. Gibbs, A.; Morris, G. A., Reference deconvolution. Elimination of distortions arising from reference line truncation. *Journal of Magnetic Resonance*, **1991,** 91, (1), 77-83.

2.20. Morris, G. A., Compensation of instrumental imperfections by deconvolution using an internal reference signal. *Journal of Magnetic Resonance*, **1988,** 80, (3), 547-552.

2.21. Metz, K. R.; Lam, M. M.; Webb, A. G., Reference deconvolution: A simple and effective method for resolution enhancement in nuclear magnetic resonance spectroscopy. *Concepts in Magnetic Resonance* **2000,** 12, (1), 21-42.

Chapter 3

3.1. Lacey, M. E.; Subramanian, R.; Olson, D. L.; Webb, A. G.; Sweedler, J. V., High-Resolution NMR Spectroscopy of Sample Volumes from 1 nL to 10 µL. *Chemical Reviews* **1999,** 99, (10), 3133-3152.

3.2. Apperley, D. C.; Chaloner, P. A.; Crowe, L. A.; Harris, R. K.; Harrison, R. M.; Hitchcock, P. B.; Lagier, C. M., X-Ray diffraction and phosphorus-31 NMR studies of the dynamically disordered 3:2 phenol–triphenylphosphine oxide complex. *Physical Chemistry Chemical Physics* **2000,** 2, 3511 - 3518.

3.3. Borsato, G.; Negra, F. D.; Gasparrini, F.; Misiti, D.; Lucchini, V.; Possamai, G.; Villani, C.; Zambon, A., Internal Motions in a Fulleropyrrolidine Tertiary Amide with Axial Chirality. *Journal of Organic Chemistry* **2004,** 69, (17), 5785-5788.

3.4. Tarkanyi, G.; Jude, H.; Palinkas, G.; Stang, P. J., Dynamic NMR Study of the Hindered Pt−N(bipyridine) Rotation in Metal-Directed Self-Assembled Macrocycles. *Organic Letters* **2005,** 7, (22), 4971-4973.

3.5. Babailov, S. P.; Krieger, Y. G., Application of Nuclear Magnetic Double Resonance to Kinetic Studies of Photoinduced Chemical Exchange in Solution. Some Theoretical and Methodological Problems. *Journal of Structural Chemistry* **2001,** 42, (2), 305-308.

3.6. Bock, K.; Sigurskjold, B. W., Mechanism and binding specificity of β-glucosidase-catalyzed hydrolysis of cellobiose analogues studies by competition enzyme kinetics monitored by ^{1}H-NMR spectroscopy. *European Journal of Biochemistry* **1989,** 178, 711-720.

3.7. Mittag, T.; Schaffhausen, B.; Gunther, U. L., Tracing Kinetic Intermediates during Ligand Binding. *Journal of the American Chemical Society* **2004,** 126, (29), 9017-9023.

3.8. Lane, A. N., NMR Studies of Dynamics in Nucleic Acids. *Progress in NMR Spectroscopy* **1993,** 25, 481-505.

3.9. Bain, A. D., Chemical exchange in NMR. *Progress in Nuclear Magnetic Resonance Spectroscopy* **2003,** 43, (3-4), 63-103.

3.10. MacKinnon, N.; Ridgway, J.; Crowell, K. J.; Macdonald, P. M., Aluminum binding to phosphatidylcholine lipid bilayer membranes: aluminum exchange lifetimes from ^{31}P NMR spectroscopy. *Chemistry and Physics of Lipids* **2006,** 139, (2), 85-95.

3.11. Harper, S. M.; Neil, L. C.; Day, I. J.; Hore, P. J.; Gardner, K. H., Conformational Changes in a Photosensory LOV Domain Monitored by Time-Resolved NMR Spectroscopy. *Journal of the American Chemical Society* **2004,** 126, (11), 3390-3391.

3.12. Wenter, P.; Fürtig, B.; Hainard, A.; Schwalbe, H.; Pitsch, S., A Caged Uridine for the Selective Preparation of an RNA Fold and Determination of its Refolding Kinetics by Real-Time NMR. *ChemBioChem* **2006,** 7, (3), 417-420.

3.13. Klaus, B.; Bent, W. S., Mechanism and binding specificity of β-glucosidase-catalyzed hydrolysis of cellobiose analogues studied by competition enzyme kinetics monitored by ^{1}H-NMR spectroscopy. *European Journal of Biochemistry* **1989,** 178, (3), 711-720.

3.14. Kuhn, T.; Schwalbe, H., Monitoring the Kinetics of Ion-Dependent Protein Folding by Time-Resolved NMR Spectroscopy at Atomic Resolution. *Journal of the American Chemical Society* **2000,** 122, (26), 6169-6174.

3.15. Bhadra, P. K.; Morris, G. A.; Barber, J., Design, Synthesis, and Evaluation of Stable and Taste-Free Erythromycin Proprodrugs. *Journal of Medicinal Chemistry* **2005,** 48, (11), 3878-3884.

3.16. Einbu, A.; Varum, K. M., Depolymerization and De-N-acetylation of Chitin Oligomers in Hydrochloric Acid. *Biomacromolecules* **2007,** 8, (1), 309-314.

3.17. Morgenstern, M.; Cline, J.; Meyer, S.; Cataldo, S., Determination of the Kinetics of Biodiesel Production Using Proton Nuclear Magnetic Resonance Spectroscopy (^{1}H NMR). *Energy & Fuels* **2006,** 20, (4), 1350-1353.

3.18. Durazo, A.; Abu-Omar, M. M., Deuterium NMR spectroscopy is a versatile and economical tool for monitoring reaction kinetics in ionic liquids. *Chemical Communications* **2002**, 66-67.

3.19. Fyfe, C. A.; Cocivera, M.; Damji, S. W. H., Flow and stopped-flow nuclear magnetic resonance investigations of intermediates in chemical reactions. *Accounts of Chemical Research* **1978,** 11, (7), 277-282.

3.20. Bunke, H.; Bunke, O., *Nonlinear Regression, Functional Relations and Robust Methods*. Wiley: New York, 1989.

3.21. Wolfram Research, Mathematica, http://www.wolfram.com/. 2004.

3.22. Allen, M. P., *Understanding Regression Analysis*. Plenum Press: 1997.

3.23. Bates, D. M.; Watts, D. G., *Nonlinear Regression Analysis and Its Applications*. Wiley: New York, 1988.

3.24. Gans, P., *Data Fitting in the Chemical Sciences:By the Method of Least Squares*. Wiley: New York, 1992.

3.25. Seber, G. A. F.; Wild, C. J., *Nonlinear Regression*. Wiley: New Jersey, 1989.

3.26. Christian, L.; Mark, R. V., Two-dimensional J-resolved NMR spectroscopy: review of a key methodology in the metabolomics toolbox. *Phytochemical Analysis* **2010,** 21, (1), 22-32.

3.27. Wider, G.; Baumann, R.; Nagayama, K.; Ernst, R. R.; Wüthrich, K., Strong spin-spin coupling in the two-dimensional J-resolved 360-MHz ^{1}H NMR spectra of the common amino acids. *Journal of Magnetic Resonance* **1981,** 42, (1), 73-87.

3.28. Morris, G. A., J-Resolved 2D Spectroscopy. *Encyclopedia of Nuclear Magnetic Resonance, ed. D.M. Grant and R.K. Harris, J. Wiley and Sons*, Chichester, **1996,** 8, 4845-4849

3.29. Bax, A.; Freeman, R.; Morris, G. A., A Simple Method for Suppressing Dispersion-Mode Contributions in NMR Spectra: The "Pseudo Echo". *Journal of Magnetic Resonance* **1981,** 43, 333-338.

3.30. Tiziani, S.; Lodi, A.; Ludwig, C.; Parsons, H. M.; Viant, M. R., Effects of the application of different window functions and projection methods on processing of ^{1}H J-resolved nuclear magnetic resonance spectra for metabolomics. *Analytica Chimica Acta* **2008,** 610, (1), 80-88.

3.31. Günther, H., *NMR Spectroscopy Basic principles, concepts, and applications in chemistry*. Wiley: Chichester, 1992.

3.32. Freeman, R., *A Handbook of Nuclear Magnetic Resonance*. Longman Publishing Group: 1997.

3.33. Bax, A.; Mehlkopf, A. F.; Smidt, J., Absorption Spectra From Phase-Modulated Spin Echoes. *Journal of Magnetic Resonance* **1979,** 35, 373-377.

3.34. Guéron, M., Line narrowing and line broadening using trigonometric functions. *Journal of Magnetic Resonance* **1978,** 30, (3), 515-520.

3.35. Lindon, J. C.; Ferrige, A. G., Digitisation and Data Processing in Fourier Transform NMR. *Progress in NMR spectroscopy* **1979,** 14, 27-66.

3.36. Laurent El, K.; Laurence, G.; Julie, O., Phenol Ugi-Smiles Systems: Strategies for the Multicomponent N-Arylation of Primary Amines with Isocyanides, Aldehydes, and Phenols. *Angewandte Chemie International Edition* **2005,** 44, (48), 7961-7964.

Chapter 4

4.1. Wold, S.; Esbensen, K.; Geladi, P., Principal Component Analysis. *Chemometrics and Intelligent Laboratory Systems* **1987,** 2, 37-52.

4.2. Nilsson, M.; Morris, G. A., Speedy Component Resolution: An Improved Tool for Processing Diffusion-Ordered Spectroscopy Data. *Analytical Chemistry* **2008,** 80, (10), 3777-3782.

4.3. Binstead, R. A.; Stultz, L. K.; Meyer, T. J., Proton-Coupled Electron Transfer in Acetonitrile Solution. Irreversible Disproportionation of $[Ru^{III}(bpy)_2(py)(OH)]^{2+}$. *Inorganic Chemistry* **1995,** 34, (3), 546-551.

4.4. Nordon, A.; Gemperline, P. J.; McGill, C. A.; Littlejohn, D., Quantitative Analysis of Low-Field NMR Signals in the Time Domain. *Analytical Chemistry* **2001,** 73, (17), 4286-4294.

4.5. Harman, H. H., *Modern Factor Analysis* University Of Chicago Press: 1976.

4.6. Vosough, M.; Mason, C.; Tauler, R.; Jalali-Heravi, M.; Maeder, M., On rotational ambiguity in model-free analyses of multivariate data. *Journal of Chemometrics* **2006,** 20, (6-7), 302-310.

4.7. Dyrby, M.; Petersen, M.; Whittaker, A. K.; Lambert, L.; Nørgaard, L.; Bro, R.; Engelsen, S. B., Analysis of lipoproteins using 2D diffusion-edited NMR spectroscopy and multi-way chemometrics. *Analytica Chimica Acta* **2005,** 531, (2), 209-216.

4.8. Nilsson, M.; Khajeh, M.; Botana, A.; Bernstein, M. A.; Morris, G. A., Diffusion NMR and trilinear analysis in the study of reaction kinetics. *Chemical Communications* **2009**, 1252-1254.

4.9. Cattell, R., "Parallel proportional profiles" and other principles for determining the choice of factors by rotation. *Psychometrika* **1944,** 9, (4), 267-283.

4.10. Morris, G. A., Diffusion-ordered spectroscopy (DOSY). *Encyclopedia of Nuclear Magnetic Resonance, ed. D.M. Grant and R.K. Harris, J. Wiley and Sons,* Chichester, **2002,** 9.

4.11. Harshman, R. A.; Lundy, M. E., PARAFAC: Parallel factor analysis. *Computational Statistics & Data Analysis* **1994,** 18, (1), 39-72.

4.12. Bro, R.; Hansen, P. I.; Viereck, N.; Dyrby, M.; Pedersen, H. T.; Engelsen, S. B., A New Principle for Unique Spectral Decomposition of 2D NMR Data. *Magnetic resonance in Food Science: The Multivariate Challenge* **2005,** 299, 195-203

4.13. Srivastava, M. S., *Methods of Multivariate Statistics* Wiley-Interscience: New York, 2002.

4.14. J. D. Carroll; J. Chang, Analysis of individual differences in multidimensional scaling via an n-way generalization of "Eckart-Young" decomposition. *Psychometrika* **1970,** 35, (3), 283-319.

4.15. Harshman, R. A., Foundation of the PARAFAC procedure: models and conditions for an "explanatory" multimodal factor analysis. *UCLA Working Papers in Phonetics* **1970,** 16, 1-84.

4.16. Smilde, A.; Bro, R.; Geladi, P., *Multi-way Analysis with Applications in the Chemical Sciences*. John Wiley & Sons: Chichester, UK, 2004.

4.17. Mitchell, B. C.; Burdick, D. S., An empirical comparison of resolution methods for three-way arrays. *Chemometrics and Intelligent Laboratory Systems* **1993,** 20, 149-161.

4.18. Zhang, X.; Xu, D., Blind PARAFAC signal detection for polarization sensitive array. *EURASIP Journal on Advances in Signal Processing* **2007** , vol. 2007, Article ID 12025, 7 pages. doi:10.1155/2007/12025.

4.19. Rutledge, D. N.; Bouveresse, D. J., Multi-way analysis of outer product arrays using PARAFAC. *Chemometrics and Intelligent Laboratory Systems* **2007,** 85, 170-178.

4.20. Bylund, D.; Danielsson, R.; Malmquist, G.; Markides, K. E., Chromatographic alignment by warping and dynamic programming as a pre-processing tool for PARAFAC modelling of liquid chromatography–mass spectrometry data. *Journal of Chromatography* **2002,** 961, 237-244.

4.21. Kiers, H. A. L.; Smilde, A. K., Some theoretical results on second-order calibration methods for data with and without rank overlap. *Chemometrics* **1995,** 9, 179-195.

4.22. Nørgaard, L., Spectral resolution and prediction of slit widths in fluorescence spectroscopy by two-and three-way methods. *Chemometrics* **1996,** 10, 615-630.

4.23. Rinnan, Å.; Anderson, C. M., Handling of first-order Rayleigh scatter in PARAFAC modelling of fluorescence excitation-emissin data. *Chemometrics and Intelligent Laboratory Systems* **2005,** 76, 91-99.

4.24. Forshed, J.; Stolt, R.; Idborg, H.; Jacobsson, S. P., Enhanced multivariate analysis by correlation scaling and fusion of LC/MS and ^{1}H NMR data. *Chemometrics and intelligent laboratory systems* **2007,** 85, 179-185.

4.25. Engelen, S.; Frosch, S.; Jørgensen, B. M., A fully robust PARAFAC method for analyzing fluorescence data. *Journal of Chemometrics* **2009,** 23, (3), 124-131.

4.26. Anderson, C. A.; Bro, R., The *N*-way Toolbox for MATLAB. *Chemometrics and Intelligent Laboratory Systems* **2000,** 52, 1-4.

4.27. Sidiropoulos, N. D.; Bro, R., On the uniqueness of multilinear decomposition of *N*-way arrays. *Chemometrics* **2000,** 14, 229-239.

4.28. Bro, R.; Kiers, H. A. L., A new efficient method for determining the number of components in PARAFAC models. *Chemometrics* **2003,** 17, 274-286.

4.29. Tomasi, G.; Bro, R., A comparison of algorithms for fitting the PARAFAC model. *Computational Statistics & Data Analysis* **2006,** 50, 1700-1734.

4.30. Hopke, P. H.; Paatero, P.; Jia, H.; Ross, R. T.; Harshman, R. A., Three-way (PARAFAC) factor analysis: examination and comparison of alternative

computational methods as applied to ill-conditioned data. *Chemometrics and Intelligent Laboratory Systems* **1998,** 43, 25-42.

4.31. Durell, S. R.; Lee, C.; Ross, R. T.; Gross, E. L., Factor Analysis of the Near-Ultraviolet Absorption Spectrum of Plastocyanin Using Bilinear, Trilinear, and Quadrilinear Models. *Archives of Biochemistry and Biophysics* **1990,** 278, (1), 148-160.

4.32. Ross, R. T.; Leurgans, S., Component Resolution Using Multilinear Models. *Methods in Enzymology* **1995,** 246, 679-700.

4.33. Bro, R., PARAFAC. Tutorial and applications. *Chemometrics and Intelligent Laborary Systems* **1997,** 38, 149-171.

4.34. Waldeck, A. R.; Kuchel, P. W.; Lennon, A. J.; Chapman, B. E., NMR diffusion measurements to characterise membrane transport and solute binding. *Progress in Nuclear Magnetic Resonance Spectroscopy* **1997,** 30, (1-2), 39-68.

4.35. Yoram, C.; Liat, A.; Limor, F., Diffusion NMR Spectroscopy in Supramolecular and Combinatorial Chemistry: An Old Parameter - New Insights. *Angewandte Chemie International Edition* **2005,** 44, (4), 520-554.

4.36. Nilsson, M.; Morris, G. A., Improved DECRA processing of DOSY data: correcting for non-uniform field gradients. *Magnetic Resonance in Chemistry* **2007,** 45, (8), 656-660.

4.37. Morris, K. F.; Johnson, C. S., Diffusion-ordered two-dimensional nuclear magnetic resonance spectroscopy. *Journal of the American Chemical Society* **1992,** 114, (8), 3139-3141.

4.38. Nilsson, M.; Connell, M. A.; Davis, A. L.; Morris, G. A., Biexponential Fitting of Diffusion-Ordered NMR Data: Practicalities and Limitations. *Analytical Chemistry* **2006,** 78, (9), 3040-3045.

4.39. Antalek, B., Using pulsed gradient spin echo NMR for chemical mixture analysis: how to obtain optimum results. *Concepts in Magnetic Resonance* **2002,** 14, (4), 225-258.

4.40. Hahn, E. L., Spin Echoes. *Physical Review* **1950,** 80, (4), 580.

4.41. Price, W. S., Pulsed-field gradient nuclear magnetic resonance as a tool for studying translational diffusion: Part 1. Basic theory. *Magnetic Resonance* **1997,** 9, (5), 299-336.

4.42. Price, W. S., Pulsed-field gradient nuclear magnetic resonance as a tool for studying translational diffusion: Part II. Experimental aspects. *Concepts in Magnetic Resonance* **1998,** 10, (4), 197-237.

4.43. Johnson Jr, C. S., Diffusion ordered nuclear magnetic resonance spectroscopy: principles and applications. *Progress in Nuclear Magnetic Resonance Spectroscopy* **1999,** 34, (3-4), 203-256.

4.44. Sørland, G. H.; Hafskjold, B.; Herstad, O., A Stimulated-Echo Method for Diffusion Measurements in Heterogeneous Media Using Pulsed Field Gradients. *Journal of Magnetic Resonance* **1997,** 124, (1), 172-176.

4.45. Cotts, R. M.; Hoch, M. J. R.; Sun, T.; Markert, J. T., Pulsed field gradient stimulated echo methods for improved NMR diffusion measurements in heterogeneous systems. *Journal of Magnetic Resonance*, **1989,** 83, (2), 252-266.

4.46. Wu, D. H.; Chen, A. D.; Johnson, C. S., An Improved Diffusion-Ordered Spectroscopy Experiment Incorporating Bipolar-Gradient Pulses. *Journal of Magnetic Resonance, Series A* **1995,** 115, (2), 260-264.

4.47. Pelta, M. D.; Barjat, H.; Morris, G. A.; Davis, A. L.; Hammond, S. J., Pulse sequences for high-resolution diffusion-ordered spectroscopy (HR-DOSY). *Magnetic Resonance in Chemistry* **1998,** 36, (10), 706-714.

4.48. Pelta, M. D.; Morris, G. A.; Stchedroff, M. J.; Hammond, S. J., A one-shot sequence for high-resolution diffusion-ordered spectroscopy. *Magnetic Resonance in Chemistry* **2002,** 40, (13), S147-S152.

4.49. Echeverria, E., Developmental Transition from Enzymatic to Acid Hydrolysis of Sucrose in Acid Limes (Citrus aurantifolia). *Plant Physiology* **1990,** 92, 168-171.

4.50. Morris, G. A.; Barjat, H.; Home, T. J., Reference deconvolution methods. *Progress in Nuclear Magnetic Resonance Spectroscopy* **1997,** 31, (2-3), 197-257.

4.51. Mathworks, Matlab, http://www.mathworks.com/,2004.

4.52. BeMiller, J. N.; Mann, R. K., Acid-catalyzed hydrolysis of maltose and selected maltose derivatives. *Carbohydrate Research* **1996,** 2, (1), 70-79.

4.53. Koyama, I.; Komine, S.-i.; Yakushijin, M.; Hokari, S.; Komoda, T., Glycosylated salivary alpha-amylases are capable of maltotriose hydrolysis and glucose formation. *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology* **2000,** 126, (4), 553-560.

4.54. Bro, R., Exploratory study of sugar production using fluorescence spectroscopy and multi-way analysis. *Chemometrics and Intelligent Laboratory Systems* **1999,** 46, (2), 133-147.

4.55. Mordi, M. N.; Pelta, M. D.; Boote, V.; Morris, G. A.; Barber, J., Acid-Catalyzed Degradation of Clarithromycin and Erythromycin B: A Comparative Study Using NMR Spectroscopy. *Journal of Medicinal Chemistry* **2000,** 43, (3), 467-474.

4.56. Nakagawa, Y.; Itai, S.; Yoshida, T.; Nagai, T., Physicochemical properties and stability in the acid solution of a new macrolide antibiotic clarithromycin, in a

comparison with erythromycin. *Chemical & Pharmaceutical Bulletin* **1992,** 40, 725-728.

Chapter 5

5.1. Green, D. B.; Lane, J.; Wing, R. M., A Standard Session Stopped-Flow NMR Tube. *Applied Spectroscopy* **1987,** 41, 847-851.

5.2. Bayer, E.; Albert, K., Continuous-flow carbon-13 nucler magnetic resonance spectroscopy. *Journal of Chromatography A* **1984,** 312, 91-97.

5.3. Hamang, M.; Sanson, A.; Liagre, L.; Forge, V.; Berthault, P., Fast mixing device inside a nuclear magnetic resonance magnet: A tool for observing early steps in protein folding. *Review of Scientific Instruments* **2000,** 71, (5), 2180-2183.

5.4. van Nuland, N. A. J.; Forge, V.; Balbach, J.; Dobson, C. M., Real-Time NMR Studies of Protein Folding. *Accounts of Chemical Research* **1998,** 31, (11), 773-780.

5.5. Tellier, C.; Mariette, F.; G.A. Webb, P. S. B.; Mccarthy, M. J., On-line Applications in Food Science. In *Annual Reports on NMR Spectroscopy*, Academic Press: 1995; Vol. Volume 31, pp 105-122.

5.6. Watanabe, N.; Niki, E., Direct-Coupling of FI-NMR to High Performance Liquid Chromatography. *Proceedings of the Japan Academy, Ser. B, Physical and Biological Sciences* **1978,** 54, 194-199.

5.7. Suryan, G., Nuclear Resonance in Flowing Liquids. *Proceedings Indian Academy Science* **1951,** 33, 107-111.

5.8. Bernstein, M. A.; Scarontefinovi, M.; Sleigh, C. J., Optimising reaction performance in the pharmaceutical industry by monitoring with NMR. *Magnetic Resonance in Chemistry* **2007,** 45, (7), 564-571.

5.9. Maiwald, M.; Fischer, H. H.; Kim, Y.-K.; Albert, K.; Hasse, H., Quantitative high-resolution on-line NMR spectroscopy in reaction and process monitoring. *Journal of Magnetic Resonance* **2004,** 166, (2), 135-146.

5.10. Mix, A.; Jutzi, P.; Rummel, B.; Hagedorn, K., A Simple Double-Chamber NMR Tube for the Monitoring of Chemical Reactions by NMR Spectroscopy. *Organometallics* **2010,** 29, (2), 442-447.

5.11. Yushmanov, P. V.; Furó, I., A rapid-mixing design for conventional NMR probes. *Journal of Magnetic Resonance* **2005,** 175, (2), 264-270.

5.12. Laude, D. A.; Lee, R. W. K.; Wilkins, C. L., Signal enhancement of long-relaxing ^{13}C nuclei by flow NMR. *Journal of Magnetic Resonance* **1984,** 60, (3), 453-459.

5.13. Fyfe, C. A.; Cocivera, M.; Damji, S. W. H., Flow and stopped-flow nuclear magnetic resonance investigations of intermediates in chemical reactions. *Accounts of Chemical Research* **1978,** 11, (7), 277-282.

5.14. Nordon, A.; McGill, C. A.; Littlejohn, D., Process NMR spectrometry. *Analyst* **2001,** 126, 260-272.

5.15. Laude, D. A.; Lee, R. W. K.; Wilkins, C. L., Analytical applications of a recycled flow nuclear magnetic resonance system: signal enhancement of slowly relaxing nuclei. *Analytical Chemistry* **1985,** 57, (7), 1281-1286.

5.16. Wilmad-Labglass. NMR-003: Pressure Performance of NMR & EPR Sample Tubes. http://www.wilmad-labglass.com/services/NMR_003.jsp. Last access 19/10/2009.

5.17. Wolfram Research, Mathematica, http://www.wolfram.com/. **2004**.

5.18. Abdel-Magid, A. F.; Carson, K. G.; Harris, B. D.; Maryanoff, C. A.; Shah, R. D., Reductive Amination of Aldehydes and Ketones with Sodium Triacetoxyborohydride. Studies on Direct and Indirect Reductive Amination Procedures 1. *Journal of Organic Chemistry* **1996,** 61, (11), 3849-3862.

5.19. Mordi, M. N.; Pelta, M. D.; Boote, V.; Morris, G. A.; Barber, J., Acid-Catalyzed Degradation of Clarithromycin and Erythromycin B: A Comparative Study Using NMR Spectroscopy. *Journal of Medicinal Chemistry* **2000,** 43, (3), 467-474.

Appendices

Published Articles

Appendix A

Chemical Communications, 2009, pp 1252-1254

Diffusion NMR and trilinear analysis in the study of reaction kinetics

Mathias Nilsson,[*a] *Maryam Khajeh,*[a] *Adolfo Botana Alcalde*[a] *Michael A. Bernstein,*[b] *and Gareth A. Morris*[a]

Received (in Cambridge, UK) 20th November 2008, Accepted 10th December 2008
First published on the web 15th January 2009

DOI: 10.1039/b820813a

Measurement of diffusion-weighted NMR spectra as a function of time allows the time-dependence of concentration and the isolated spectrum to be found for each component in a reaction, without prior assumptions about spectra, kinetics or diffusion behaviour, by data decomposition using the PARAFAC algorithm.

Nuclear Magnetic Resonance (NMR) is frequently employed to study reaction kinetics. NMR can provide detailed structural information about (and often identify) the chemical entities involved in a reaction, and as it is non-invasive and non-destructive, the kinetics of an intact mixture can be studied in real time directly in the NMR tube[1] (there are also a number of alternatives for the study of reaction conditions that cannot be duplicated in an NMR probe[2]). Reaction monitoring by NMR works best when each component in a reaction mixture has at least one well-resolved resonance; the change in peak integral can then be used directly to determine the kinetic behaviour[3]. When no such resolved peaks are available, as is quite common, the extraction of kinetic data becomes much more challenging, and it is frequently impossible to identify individual reaction components, let alone determine their concentrations. In this investigation, we demonstrate that by adding diffusion information to the NMR experiments, the spectrum, time evolution and diffusion data

can be recovered for each component in the reaction mixture. Because the data are trilinear (i.e. they vary independently in three dimensions, here diffusional attenuation, time evolution and chemical shift) they can be decomposed using a PARAFAC[4] algorithm, and it is therefore possible to analyse the data without the need for fitting to a predetermined model, and without having to constrain the data to fit either the reaction kinetics or the diffusional attenuation.

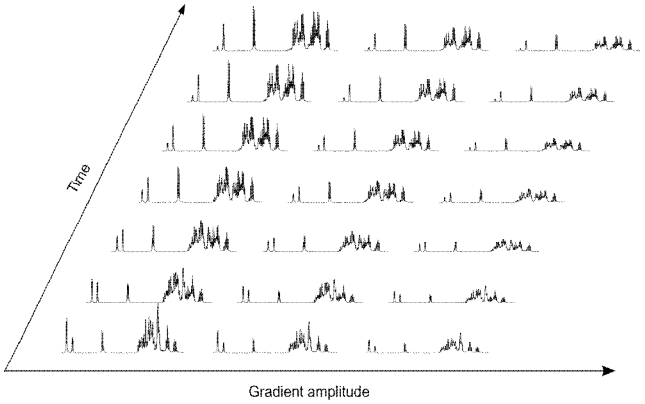


Figure 1. A subset of the raw experimental data. For the time evolution every 16th spectrum is shown, and for the decay with gradient amplitude (caused by diffusion) the first three gradient levels are shown.

The study of reactions is an example of the general case of mixture analysis by NMR. It is well known that it can be frustrating to study intact mixtures by NMR, as it is often difficult to assign resonances unambiguously to given mixture components. It is expensive, tedious and time-consuming to separate components physically (e.g. by chromatography) before submitting them to NMR, and frequently it is the study of the intact mixture itself that is of interest (as for reaction monitoring). Therefore it is highly desirable to develop NMR methods that can recover the required information from intact mixtures. Some of the most powerful NMR methods currently available, commonly referred to as DOSY (diffusion-ordered spectroscopy) experiments, are based on diffusion[5-9]; these are most effective where each component in a mixture has

a unique rate of diffusion. The diffusion of molecules can be measured by recording the signal attenuation in a pulsed field gradient NMR experiment[10], typically by incrementing the gradient strength in a pulsed field gradient spin or stimulated echo. It was recognised early on that the results of such experiments can be used to distinguish the signals from different molecular species[11]. The decays of individual NMR signals are typically fitted to a model function, and the fitted diffusion coefficient is then used to correlate the signals of individual molecular species. In high resolution DOSY[6, 5], this is done by fitting each peak individually (implicitly assuming that there is no spectral overlap), while in multivariate methods the whole bandwidth is fitted simultaneously[8, 9, 7]. The model function used is typically some form of the Stejskal-Tanner equation[10], which describes the effect of pulsed field gradient on signal amplitude for free (unbounded) diffusion; for best results, the equation can be extended to include the effects of imperfect field gradient uniformity[12, 6].

Diffusion-ordered spectroscopy and kinetic studies by NMR have a good deal in common: both rely on fitting variations in NMR signal amplitude to suitable model functions, and in both cases it is far easier to analyse experimental data when the NMR signals of individual species are well resolved. DOSY data and timecourse spectra are bilinear: signal intensity I is measured as a function of two variables, frequency and gradient amplitude, and frequency and time respectively. In a bilinear dataset, the theoretical intensity I_i for a given signal i is the product of the signal variation as a function of two different variables, $I_i(p,q) = P_i(p)\ Q_i(q)$. Thus in DOSY, if the spectrum of component i is $S_i(f)$ and its signals attenuate as a function of gradient amplitude g according to $A_i\ (g)$, then $I_i(f,g)$ is the product of $S_i(f)$ and $A_i(g)$. The experimental dataset is a tensor of rank 2, and may be represented as a sum over N components i of outer products of two vectors $\mathbf{S}_i$ and $\mathbf{A}_i$, plus some residual $\mathbf{E}$:

$$I = \sum_{i=1}^{N} S_i \otimes A_i + E \qquad (1)$$

In analysing bilinear data with spectral overlap it is common to employ multivariate methods to help resolve the component spectra (and diffusion/kinetics)[13, 14, 8, 9, 7]. Unfortunately such analyses suffer from the problem of *rotational ambiguity*: any linear combination of the true functions P_i, or the true functions Q_i, gives an equally good fit to the experimental data. For bilinear analysis it is therefore necessary to apply constraints, for example non-negativity and/or known/hypothesised kinetic models, to allow the true solutions to be selected out from the infinite range of linear combinations. This problem can be avoided, and a model-free fit obtained by PARAFAC[4] decomposition, if trilinear data $I_i(p,q,r) = P_i(p)\ Q_i(q)\ R_i(r)$ can be measured. Adding a diffusion dimension to a bilinear dataset can create a trilinear structure[15]. Recording NMR spectra as a function both of time and of gradient amplitude, i.e. measuring a timecourse of DOSY spectra, gives just such a dataset. No prior knowledge of the component spectra, diffusion behaviour or kinetics is required for its analysis; the only requirement is that the spectrum $S_i(f)$, diffusional attenuation $A_i(g)$ as a function of gradient g, and concentration profile $C_i(t)$ of each species be independent of each other, so that the signal intensity $I_i(f,g,t) = S_i(f)\ A_i(g)\ C_i(t)$. The experimental dataset is now a rank 3 tensor:

$$I = \sum_{i=1}^{N} S_i \otimes A_i \otimes C_i + E \qquad (2)$$

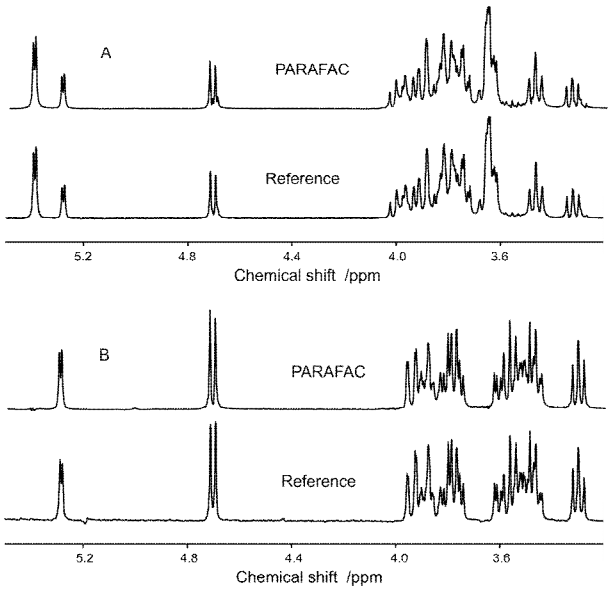


Figure 2. Reference spectra of pure materials and spectra obtained from the data of Fig. 1 by PARAFAC for maltose (A) and glucose (B).

To demonstrate the value of using diffusion encoding in the NMR study of a reacting mixture we have chosen the well-known acid hydrolysis of maltose to glucose[16], in which one maltose is hydrolysed to two glucose units.

A aqueous solution of maltose (5.5 % w/w) in 33 % (w/w) sulfuric acid was prepared, with 0.15 % (w/w) pivalic acid as a reference compound. Hydrolysis was carried out at 50 °C in a thick-walled NMR tube (to prevent convection; i.d. 2.2 mm) in a 400 MHz Varian Inova instrument, using a 5 mm diameter indirect detection probe equipped with a *z*-gradient coil allowing gradient pulses up to 30 G cm^{-1}. A series of 98 DOSY experiments was carried out over the course of the reaction, in a total of 41 h 49 min. Each DOSY experiment used the oneshot sequence[17] with 32 transients at each of 6 gradient levels, with equal steps in gradient squared, ranging from 3.0 to 27.3 G cm^{-1}. The data were then Fourier transformed, phase corrected, baseline corrected, reference deconvoluted[18] using the pivalic acid signal, and the solvent (HOD) peak was removed by digital filtering, all using the manufacturer's VnmrJ

software, before export to MATLAB for further analysis. All PARAFAC analysis was performed with the MATLAB *N*-Way toolbox[19, 20] Fig. 1 shows a subset of the experimental spectra as a function of time and gradient level. Small variations in receiver sensitivity over the course of the experiment were corrected for by normalising the integral of the spectrum for each gradient level using the average area of the pivalic acid reference peak.

PARAFAC fitting was carried out, assuming two components, for the spectral region 3.1-5.5 ppm. The resultant fit accounted for 99.8% of the variance in the data, yielding statistical components $S_i(f)$, $A_i(g)$ and $C_i(t)$ representing the spectrum, signal decay as a function of gradient strength, and time evolution for the reactant and product respectively.

One great advantage of PARAFAC fitting is that, if the assumption of trilinearity holds, the fitted components obtained should have physical relevance, i.e. should in this case be the true spectrum, diffusional attenuation and concentration timecourse. Where prior information exists, therefore, it is possible to assess directly the quality of the PARAFAC decomposition, for example by comparing the spectra of reaction components extracted by fitting the experimental data with the known spectra of the pure materials. As can be seen in Fig. 2, in this case the fitted spectra are virtually identical to the spectra of pure maltose and pure glucose, confirming that trilinear decomposition into two components was successful.

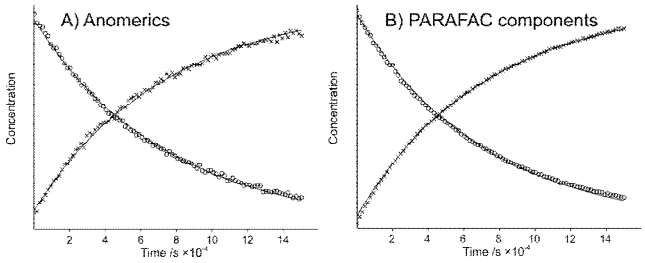


Figure 3. Non-linear least squares fits to first order kinetics for the acid hydrolysis of maltose to glucose. A) relative concentrations of maltose (integral of the inner anomeric signal at 5.4 ppm, decaying curve) and glucose (sum of the integrals of the peaks between 4.6 and 5.3 ppm minus the integral of that at 5.4 ppm, rising curve). Estimated rate constant $1.36 \pm 0.02 \times 10^{-5}$ s^{-1}. B) Normalised PARAFAC components $C_i(t)$ for maltose (decaying) and glucose (rising); $k = 1.40 \pm 0.01 \times 10^{-5}$ s^{-1}.

Because the relative scaling of the three multiplicands in the trilinear model is arbitrary, to obtain true relative concentrations from the analysis it is necessary to ensure that the other two modes in the model, $S_i(f)$ and $A_i(g)$, are normalised. Where, as here, the structures of the reaction components are known this is straightforward: each PARAFAC spectral mode output $S_i(f)$ is normalised to have an integral proportional to the number of protons involved, and each diffusion mode $A_i(g)$ is normalised so that it extrapolates back to unity at zero gradient g. Multiplying the remaining raw modes $C_i(t)$ by the normalisation factors by which the $S_i(f)$ and $A_i(g)$ were divided then gives $C_i(t)$ modes which are directly proportional to concentration. The net result for the experimental data of Fig. 1 is shown in Fig. 3B, and as expected gives an excellent fit to first order kinetics. The PARAFAC results can, for this model system in which well-resolved anomeric signals are available, be compared with relative concentration profiles obtained by direct integration of the respective anomeric signals (Fig. 3A); while there is excellent agreement, the signal-to-noise ratio for the PARAFAC result is, as expected, superior. Although for this example the full range of chemical shifts, including the well-resolved anomeric signals, was used,

essentially identical results were obtained when only the highly overlapped region between 3.1 and 4.1 ppm was fitted, confirming that both fully-resolved spectra and kinetic information can be recovered even when the experimental data contain no resolved peaks. The PARAFAC decomposition is remarkably robust; similar results can be obtained using only two of the six gradient increments measured, and/or with many fewer time points.

From this model study it is clear that by adding diffusion information to an experimental time course study and using multi-way methods to decompose the results into the underlying structure of NMR spectrum, diffusional attenuation and time evolution, it is possible not only to obtain good estimate kinetic data irrespective of whether any resolved signals are available, but also to recover the NMR spectra of individual reaction components. In principle, one could obtain by this method the NMR spectra of intermediates that are difficult or impossible to isolate. The fundamental requirement is that each reaction component shows a different diffusion coefficient and a different time course; even where these requirements are not fully met, a hybrid analysis in which PARAFAC is constrained using prior knowledge (e.g. spectral non-negativity) can still succeed. It should however be stressed that for a trilinear PARAFAC decomposition, no assumptions are needed about the form of the spectra, the diffusional attenuation, or the kinetics. Thus decomposition is just as simple even when no prior information is available, as for example in a heterogeneous reaction where the functional form of the attenuation caused by (bounded) diffusion is unknown.

Notes and references

[a] *School of Chemistry, University of Manchester, Oxford Road, Manchester M13 9PL, UK*
[b] *AstraZeneca R&D Charnwood, Bakewell Rd, Loughborough, Leics. UK LE11 5RH*
mathias.nilsson@manchester.ac.uk

1 D. Kaufman, C. Sterner, B. Masek, R. Svenningsen, and G. Samuelson, *J. Chem. Educ.*, 1982, 59, 885-886.

2 M. A. Bernstein, M. Stefinovic, and C. J. Sleigh, *Magn. Reson. Chem.*, 2007, 45, 564-571.

3 J. Chrastil, *Comput. Chem.*, 1988, 12, 289-292.

4 R. Harshman, *UCLA Working Papers in Phonetics.*, 1970, 16, 1-84.

5 C. S. Johnson, *Prog. Nucl. Magn. Reson. Spectrosc.*, 1999, 34, 203-256.

6 G. A. Morris, in *Encyclopedia of Nuclear Magnetic Resonance*, eds. D. M. Grant and R. K. Harris, John Wiley & Sons Ltd, Chichester, 2002, vol. 9 : Advances in NMR, pp. 35-44.

7 M. Nilsson and G. A. Morris, *Anal. Chem.*, 2008, 80, 3777-3782.

8 L. C. Van Gorkom and T. M. Hancewicz, *J. Magn. Reson.*, 1998, 130, 125-130.

9 W. Windig and B. Antalek, *Chemom. Intell. Lab.*, 1997, 37, 241-254.

10 E. O. Stejskal and J. E. Tanner, *J. Chem. Phys.*, 1965, 42, 288-292.

11 P. Stilbs, *Anal. Chem.*, 1981, 53, 2135-2137.

12 P. Damberg, J. Jarvet, and A. Gräslund, *Journal of Magnetic Resonance*, 2001, 148, 343-348.

13 E. Bezemer and S. Rutan, *Analytica Chimica Acta*, 2002, 459, 277-289.

14 A. de Juan and R. Tauler, *Analytica Chimica Acta*, 2003, 500, 195-210.

15 H. T. Pedersen, M. Dyrby, S. B. Engelsen, R. Bro, and G. A. Webb, in *Annual Reports On Nmr Spectroscopy*, Academic Press, 2006, vol. Volume 59, pp. 207-233.

16 J. N. BeMiller and R. K. Mann, *Carbohydrate Research*, 1966, 2, 70-79.

17 M. D. Pelta, G. A. Morris, M. J. Stchedroff, and S. J. Hammond, *Magnetic Resonance in Chemistry*, 2002, 40, S147-S152.

18 G. A. Morris, H. Barjat, and T. J. Horne, *Progress in Nuclear Magnetic Resonance Spectroscopy*, 1997, 31, 197-257.

19 C. A. Andersson and R. Bro, *Chemometrics and Intelligent Laboratory Systems*, 2000, 52, 1-4.

20 R. Bro, *Chemometrics and Intelligent Laboratory Systems*, 1997, 38, 149-171.

Appendix B

Analytical Chemistry, 2010, 82 (5), pp 2102-2108

Reaction Kinetics Studied Using Diffusion-Ordered Spectroscopy and Multi-Way Chemometrics

Maryam Khajeh[a], *Adolfo Botana*[a], *Michael A. Bernstein*[b], *Mathias Nilsson*[a*], *and Gareth A. Morris*[a]

[a]School of Chemistry, University of Manchester, Oxford Road, Manchester M13 9PL, UK.

[b]AstraZeneca R&D Charnwood, Bakewell Rd, Loughborough, Leics, LE11 5RH, UK.
[*]Corresponding author:
E-mail:mathias.nilsson@manchester.ac.uk
DOI: 10.1021/ac100110m

ABSTRACT

Nuclear magnetic resonance (NMR) spectroscopy is frequently used in the monitoring of reaction kinetics, due to its non-destructive nature and to the wealth of chemical information that can be obtained. However, when spectra of different mixture components overlap, as is common, the information available is greatly reduced, sometimes to the point where the identification of individual chemical species is not

possible. In such cases the resolution of component spectra and their concentration timecourses can be greatly improved by recording DOSY (Diffusion-Ordered Spectroscopy) data for each time point during the reaction. Adding this additional degree of freedom to the experimental data, allowing the signals of different species to be distinguished through their different rates of diffusion, makes the data trilinear, and therefore susceptible to analysis by powerful multi-way (here more specifically multilinear) model-free decomposition methods such as PARAFAC (parallel factor analysis). This approach is shown to produce high quality data even for species with near-degenerate spectra. Another important limitation of NMR is its inherently low sensitivity. Here we show that the combination of DOSY and PARAFAC is surprisingly robust with respect to input data with low signal-to-noise ratio. High quality component spectra and kinetic profiles are obtained from a dataset in which the signal-to-noise ratios of the reaction components in the spectra for individual time points are below the detection level.

KEYWORDS: NMR, DOSY, PARAFAC, PFGSTE, MATLAB, kinetics, diffusion, pulsed field gradients (PFG), multivariate analysis, multi-way, multilinear, trilinear, chemometrics.

The monitoring of chemical reactions is an important task in both industry and academia, for example in process development and degradation studies. The non-invasive and non-destructive method of Nuclear Magnetic Resonance (NMR) can be a particularly effective tool due to the wealth of information it provides, for example on chemical structure, and on intermolecular and intramolecular dynamics. Where the signals of individual species are well-resolved, reaction monitoring is straightforward and peak integrals can be used directly to determine changes in concentration[1]. However, a major obstacle in the study of mixtures by NMR is that when component spectra overlap it is often impossible to assign signals unambiguously to a specific molecular species. In such cases, methods that aim to fit entire spectra (i.e. multivariate methods) can often be helpful in resolving component spectra and determining concentrations[2-8]. Much of the power of multi-way (or in this case more specifically multi-linear) methods stems from the fact that multilinearity offers relief from the so-called rotational ambiguity limitation[9] that bilinear methods suffer from (*vide infra*). The combination of diffusion-ordered spectroscopy (DOSY[10, 11]) experiments and multivariate multi-way chemometrics[12] has recently been shown to be an excellent tool both for mixture analysis in general[13-15], and more specifically for reaction kinetics[16]. Here the practicalities of this combination of methods are explored, and its remarkable robustness when analysing data with limited signal-to-noise ratio (S/N) is illustrated.

The most powerful general methods for mixture analysis by NMR are LC-NMR[17] and, as used in this investigation, diffusion-based methods such as DOSY[10, 11]. Although the term DOSY strictly refers only to a processing and display method used with diffusion-weighted NMR data, it is commonly used more loosely to describe both the analysis method and the experimental data, and will be used in that sense here

despite actual DOSY spectra not being constructed. A central advantage of DOSY methods is that they do not require any physical separation of a mixture, and therefore are particularly helpful when a reaction is followed directly in a NMR tube[16, 18]. The measurement of diffusion-weighted spectra allows the signals from different species to be distinguished, by virtue of their different effective hydrodynamic radii. A DOSY experiment is typically performed by recording a set of NMR spectra in a pulsed field gradient spin or stimulated echo (PFGS[T]E) as a function of gradient amplitude. The NMR signals are ideally attenuated according to the Stejskal-Tanner equation[19]:

$$S(G) = S_0 e^{-D\gamma^2\delta^2G^2\Delta'} \quad (1)$$

where S is the signal amplitude, S_0 is the amplitude in absence of diffusion, D is the diffusion coefficient, δ is the diffusion encoding gradient pulse duration, γ is the magnetogyric ratio, G is the gradient amplitude, and Δ' is the diffusion time (Δ) corrected for the effects of the non-zero gradient pulse duration. The experimental data are normally fitted either to the Stejskal-Tanner equation or to a version of that equation modified to take into account instrumental imperfections such as non-uniform field gradients[20],to obtain the spectra and diffusion coefficients of the mixture components. Most commonly this is done by fitting the decay of each signal individually, using univariate methods[21-23], and displaying the result in a 2D DOSY plot. However, as all the signals for each component decay identically (in the absence of exchange[24, 25]), it is possible to exploit this covariance to separate the component spectra even where individual signals overlap[2-4, 26, 27].

In multivariate methods it is conventional to describe the data analysis in matrix form. For example, in the decomposition of a DOSY dataset into component spectra and attenuation functions for N different components the experimental data are modeled by the equation

$$\mathbf{D}=\mathbf{A}\mathbf{S}^{T} + \mathbf{E} \qquad (2)$$

where the experimental dataset **D** is a matrix in which successive rows contain the measured mixture spectrum for different values of the diffusion-encoding gradient strength. This matrix is represented in Equation (2) as the product of two row matrices containing the spectral amplitude profiles **A** (for a DOSY dataset these are the signals decays as a function of gradient strength for the different components) and the component spectra **S**, where ^{T} denotes the transpose. The error matrix **E** contains unfitted data, which ideally consist solely of unavoidable random experimental errors such as noise. In an alternative notation this can be written as:

$$\mathbf{D} = \sum_{i=1}^{N} \mathbf{s}_i \otimes \mathbf{a}_i + \mathbf{E} \qquad (3)$$

where $\mathbf{s}_i$ and $\mathbf{a}_i$ are the *i*th vectors of the matrices **S** and **A** and $\otimes$ is the Kronecker product. The same notation can be used to describe the data obtained in a 'standard' kinetics NMR experiment where a simple proton spectrum is recorded for a succession of time points, with the matrix **A** describing the variation in concentration of each of the components with time**.**

Multivariate methods set out to find the matrices **A** and **S** that best represent the experimental dataset. Unfortunately this is far from straightforward, as an infinite number of solutions exists in the form of linear combinations of the actual spectra and signal amplitude profiles. This is known as the rotational ambiguity problem for bilinear data[9]. This ambiguity can be resolved by imposing appropriate constraints on the analysis, for example non-negativity, or a predetermined form of concentration profile[2-8, 26, 27]. However, this requires detailed prior knowledge about the system under study, and there is a considerable risk of applying constraints that are not valid.

In contrast to the bilinear case, for multilinear data (varying independently in n dimensions; $n > 2$) rotational ambiguity does not arise. This has the very useful consequence that no prior knowledge about the variation of data in any of the n dimensions is necessary; fitting data to such a PARAFAC[28, 29] (parallel factor analysis) model, the form of the variation in the different dimensions is obtained directly.

Trilinear data may be obtained by recording a DOSY dataset for each time point in a chemical reaction[16]. In the notation of Eq (3) a trilinear data set can be written as:

$$\mathbf{D} = \sum_{i=1}^{N} \mathbf{s}_i \otimes \mathbf{a}_i \otimes \mathbf{b}_i + \mathbf{E} \qquad (4)$$

where $\mathbf{b}_i$ is the ith column of a matrix **B** containing the relative concentration as a function of time for each of the reaction components. For a given number N of mixture components the matrices **S**, **A** and **B** (often referred to as modes; e.g. **S** is the spectral mode) can be obtained directly from decomposition of **D** using one of the many algorithms to fit the PARAFAC model[30]; no further assumptions are needed in such 'model free' fitting. In the present investigation the matrices **S**, **A** and **B** contain the N spectra, diffusion decays and concentration timecourses, respectively; each column i corresponds to the spectrum, diffusion decay and concentration timecourse for a given mixture component. When the data are truly trilinear, the PARAFAC decomposition is straightforward, but with experimental data it is common to find small (or not so small) additional systematic variation that does not fit a trilinear model. In the present investigation we acquire approximately trilinear data by recording a DOSY spectrum for each time point in the reaction. One obvious source of deviation from trilinearity is that the acquisition of a DOSY spectrum is not instantaneous, so any changes in concentration of the different components during the

recording of a given DOSY dataset will cause a non-trilinear variation. In this investigation, however, such effects can safely be ignored as the time taken to record one set of diffusion-weighted spectra is negligible compared to the total reaction time. A second, more insidious, source of deviation from the trilinear model is variation in the component spectra, for example because of composition-dependent chemical shifts. For PARAFAC analysis to succeed, it is vital both that the component spectra remain identical throughout the reaction, and that the behavior of the different components is non-degenerate, i.e. that no two components have one of a spectrum, diffusion coefficient or timecourse in common. (If two modes are the same, the components become indistinguishable). In this work, some common deviations from trilinearity such as frequency shifts and (near-)degeneracy are investigated, and potential solutions to these problems are examined, and the robustness of the combination of DOSY and PARAFAC with respect to signal-to-noise ratio is investigated. The illustrative reactions used are the acid hydrolyses of sucrose and of maltotriose, the former showing the effects on PARAFAC decomposition of chemical shift variation and signal-to-noise ratio, and the latter those of near-degeneracy. (Although these reactions are in principle complicated considerably by the presence of two anomers for most of the species, in practice the kinetics remain simple because anomerisation is rapid compared with hydrolysis under the conditions used.)

Experimental

Four samples were used for the acid hydrolysis experiments, all with D_2O as solvent: sample 1 containing maltotriose (Fisher Scientific, 18 mM), pivalic acid (Sigma-Aldrich, 25 mM), and sulphuric acid (90 mM); sample 2 containing sucrose (Silver Spoon, 540 mM), tert-butanol (Fluka, 135 mM), and sulphuric acid (224 mM);

sample 3 containing sucrose (50 mM), tert-butanol (105 mM), and sulphuric acid (464 mM); and sample 4 containing sucrose (1.1 mM), tert-butanol (82 mM), and sulphuric acid (592 mM). All hydrolyses were carried out in thick walled 5 mm NMR tubes (i.d. 2.2 mm, to discourage convection) in a 400 MHz Varian Inova spectrometer, using an indirect detection probe equipped with a *z*-gradient coil allowing gradient pulses up to 30 G cm^{-1}. The hydrolysis of sample 1 was carried out at 50 °C, and that of samples 2-4 without temperature control in a room air-conditioned at a nominal 20 °C. All DOSY experiments were recorded using the Oneshot sequence[31] using 6 gradient levels with equal steps in gradient squared, ranging from 3.0 to 27.3 G cm^{-1}.

Data were acquired during hydrolysis using the following sample-specific parameters:

sample 1) 130 DOSY experiments with a recycle time of 6 s, a total diffusion-encoding gradient duration (δ) of 2 ms and a diffusion time (Δ) of 0.2 s, using 32 transients of 16384 complex data points in a total time of 52 h;

sample 2) 46 DOSY experiments with a recycle time of 3.6 s, a total diffusion-encoding gradient duration (δ) of 3 ms and a diffusion time (Δ) of 0.1 s, using 16 transients of 16384 complex data points in a total time of 6 h;

sample 3) 110 DOSY experiments with a recycle time of 4 s, a total diffusion-encoding gradient duration (δ) of 3 ms and a diffusion time (Δ) of 0.1 s, using 16 transients of 16384 complex data points, in a total time of 16 h; and

sample 4) 350 DOSY experiments with a recycle time 4 s, a total diffusion-encoding gradient duration (δ) of 3 ms and a diffusion time (Δ) of 0.1 s, using 4 transients of 16384 complex data points in a total time of 15 h.

An illustrative excerpt from the data acquired for sample 2 is shown in Figure 1. Reference spectra for reactants, intermediates and products were obtained as follows,

to allow comparison between the spectra obtained by PARAFAC fitting and the authentic spectra. The reference spectrum for glucose was acquired from sample 1 after the reaction had proceeded to termination, and that for maltotriose was taken as the first spectrum in the hydrolysis timecourse. Because the hydrolysis of maltose is relatively rapid under the experimental conditions used, the reference spectrum for maltose was obtained by back extrapolation of spectra acquired under the same conditions as for the hydrolysis of sample 1, using maltose (Fisher Scientific) as starting material. The reference spectra for sucrose, and for the mixture of glucose and fructose, were acquired using a 0.6 M sucrose solution, and an equimolar 0.6 M solution of glucose (Sigma-Aldrich) and fructose (Alfa Aesar), respectively, in sulphuric acid and D_2O.

All spectra were Fourier transformed, phase corrected, baseline corrected, and reference deconvoluted[32, 33] using the pivalic acid (sample 1) or tert-butanol (samples 2-4) signal, all using the manufacturer's VnmrJ software, before exporting to MATLAB (www.mathworks.com) for further analysis. Reference deconvolution was used to minimise the effects of changes in instrumental lineshape, signal phase, and signal frequency over the course of a reaction, and greatly improves the quality of the final results. PARAFAC analysis was performed with the MATLAB *N*-way Toolbox[34]. The DOSY Toolbox[35] (open source free software for DOSY processing), also contains an interface with the *N*-way Toolbox. Small variations in receiver sensitivity over the course of the experiment were corrected for by normalising the amplitude of every individual DOSY dataset to show the same total reference signal integral for the first gradient level for each DOSY experiment. Where the water signal lay within the spectral region containing signals of interest, it was excluded from analysis by setting the data points to zero for 0.25 ppm either side of the water signal.

(Note that this appears at higher chemical shifts than usual because of the effects of high acid concentration and of temperature). The relevant spectral region (3.0-5.7 ppm for sample 1, 3.2-5.7 ppm for sample 2, and 3-5.8 ppm for samples 3 and 4) was then selected and fitted to the PARAFAC model.

PARAFAC analysis produces output modes with arbitrary scaling. Each PARAFAC spectral mode, $S_i(f)$, was therefore normalized to have an integral proportional to the number of protons involved, and each diffusion mode $A_i(g)$ was normalized so that the extrapolated amplitude at zero gradient for each component was unity. The remaining mode, concentration $C_i(t)$, was then multiplied by the factors that $S_i(f)$ and $A_i(g)$ were divided by, to make it proportional to the true concentrations of the components. The modes representing the timecourses were fit using the non-linear least squares fitting in the Optimization Toolbox (MATLAB), to first order kinetics. Rate constants are reported with error margins corresponding to twice the standard error of the fit.

Results and discussion

One of the central limitations of the DOSY-PARAFAC approach is that degeneracy in one of the modes can cause the clean separation of components to fail. The acid hydrolysis of aqueous maltotriose, which proceeds via maltose to glucose, was therefore investigated as a challenging test case. Here both the spectral modes of maltose and maltotriose are similar, because of the extensive commonality in chemical shifts, and the diffusion modes are similar, because of the relatively small difference in diffusion coefficient. It has been noted previously that imposing physically realistic constraints like non-negativity can be helpful in such cases[28, 36]. An unconstrained three component PARAFAC fit did reproduce the spectra, diffusion decays and

timecourses fairly faithfully, but significant deviation from a physically realistic result was evident early in the timecourse, where the ‘maltose’ component showed a small negative initial concentration. These deviations are probably attributable to small temperature disturbances (see figure caption) causing minor violation of trilinearity. However, when non-negativity constraints were applied to all three modes, no anomalies were seen in any of the modes (explained variance 99.9 %). The fitted component spectra were virtually identical to the reference spectra, and the concentration timecourses followed the form expected for a sequential first order reaction (Figure 2).

In this test case, both the number of components and the component spectra were already known, so definitive validation could be performed by comparing the spectral modes to the model spectra. However split-half analysis, jack-knifing and core consistency checks were all consistent with this prior knowledge; split-half analysis with three components, for example, gave essentially identical results to the full analysis in all three modes. Non-linear least squares fitting of the concentration timecourses to successive first order kinetics gave estimated rate constants of $k_1 = 7.03 \pm 0.15 \times 10^{-5}$ s^{-1} for the hydrolysis of maltotriose to maltose under the conditions used, and $k_2 = 1.58 \pm 0.02 \times 10^{-5}$ s^{-1} for the hydrolysis of maltose to glucose. Experiments with lower signal-to-noise ratio data were significantly less successful, reflecting the importance of the non-negativity constraints in disentangling the component spectra. Because there are relatively few points in the maltotriose spectrum at which there are no maltose or glucose signals, the decomposition is very vulnerable to interference from noise at these points.

The second reaction studied was the simpler case of the acid hydrolysis of sucrose to glucose and fructose. Although at first sight this is a three-component problem, here

the concentration timecourses of glucose and fructose are identical, and the diffusional attenuations very nearly so. The result is that the two products behave in the two-component PARAFAC decomposition (explained variance 99.9%) as a single species with a composite spectrum, so the data may be analysed as a simple two-component problem. At first sight the fitted spectra obtained by PARAFAC decomposition of the experimental data for sample 2 are quite promising (Figure 3a), but close examination shows the presence of a small number of anomalies at the positions shown by vertical arrows. These anomalies represent cross-talk between the two component spectra, with signals appearing in the sucrose spectrum at the chemical shifts of glucose/fructose signals, and vice versa, and arise from deviations from strict trilinearity in the experimental data. Their “dispersion mode”-like appearance suggests frequency shifts as the origin of the anomalies. Detailed analysis confirms that this is the case; the individual component spectra change subtly as a function of the overall composition of the reaction mixture, with chemical shifts changing by a few parts per billion as the relative proportions of reactant and products change. Interestingly, it is not the case that the signals for which anomalies are seen are the only ones whose chemical shifts change with composition; rather, all the sugar signals shift relative to the tert-butanol reference signal, but the signals that yield anomalies are the few signals that behave differently from the majority.

The diagnosis that differential medium effects on chemical shifts were the cause of the anomalous signals prompted measurements on sample 3, where the initial concentration was reduced by an order of magnitude and hence the effects of composition on chemical shifts were expected to be much smaller. As predicted, the anomalies were greatly reduced and an excellent match was seen between the fitted spectra yielded by PARAFAC (Figure 3b) and the reference spectra (Figure 3c); the

two component fit explained 99.9 % of the variance. The concentration timecourses obtained were fitted to first order kinetics, as shown in Figure 4, yielding a rate constant of $7.48 \pm 0.03 \times 10^{-5}$ s^{-1}, and the diffusion decays were fitted to the Stejskal-Tanner equation corrected for non-uniform field gradients[20] (Figure 4) giving $D_{sucrose} = 1.28 \pm 0.01 \times 10^{-10}$ $m^2 s^{-1}$ and $D_{glucose,\ fructose} = 1.37 \pm 0.01 \times 10^{-10}$ $m^2 s^{-1}$.

Much larger chemical shift changes can be seen, for example, in the spectra of reactions in which pH changes. In such systems it would be necessary either to exclude itinerant signals from analysis, or to measure spectra under fixed chemical conditions, for example by buffering and/or using diluted sampled aliquots. Differential chemical shift changes cannot corrected by reference deconvolution[32, 33] (although the latter plays a vital role in minimising the effects of instrumental irreproducibility, as has been shown previously for both uni- and multivariate processing of DOSY data[23, 37]).

In order to study the usefulness of the DOSY-PARAFAC approach for low concentration/low signal-to-noise ratio samples, a third hydrolysis of sucrose was carried out at a very low concentration (1.1 mM; sample 4). Here, the signal-to-noise ratio of approximately 2:1 in the least attenuated (lowest gradient) spectrum (Figure 5a) was below the detection level (normally taken as 5:1). The classical approach of following the intensity changes of individual peaks would here be totally impractical. Remarkably, the PARAFAC decomposition (Figure 5c) produces excellent component spectra, once again showing a very good match with the reference spectra of Figure 3c for sucrose and glucose/fructose (explained variance 16 %, reflecting the extremely low signal-to-noise ratio). The reason why PARAFAC is so successful here is that the entire dataset is fitted, so the signals from all the 350 experiments contribute to the fitted spectra, and all the signals in each spectrum contribute to the

fitted timecourses. The latter are also perfectly serviceable (Figure 5b), fitting to first order kinetics yielding a rate constant of $1.15 \pm 0.05 \times 10^{-4}$ s^{-1}.

Conclusion

The combination of NMR diffusion measurements and trilinear data analysis (in the form of PARAFAC), can be very powerful for the study of reaction kinetics. The method can be used for data with extremely low signal-to-noise ratio, significantly below the level at which conventional methods would fail. Problems with spectral inconsistencies during the reaction that are of instrumental origin can in many cases be alleviated by using reference deconvolution, while changes in spectra due to concentration dependent shifts can be minimised by making measurements at relatively low concentration. While the relatively slow reactions studied here allowed DOSY datasets with many transients and gradient levels to be acquired, where appropriate the method can be used with as little as one transient and two gradient levels per time point, allowing reactions to be studied on timescales as low as 1 min. In systems with near degeneracy in one or several modes (e.g. very similar NMR spectra or diffusion coefficients), accurate kinetic and spectral data can be obtained by imposing physically appropriate constraints such as non-negativity. In contrast to standard NMR methods, the combination of NMR diffusion measurements and trilinear data analysis can be used in situations with severe spectral overlap and at very low concentrations.

ACKNOWLEDGMENTS

Support from the Engineering and Physical Sciences Research Council (grant references EP/D05592X, EP/E057888/1 and EP/E05899X) is gratefully acknowledged. MK thanks the EPSRC and AstraZeneca for an Industrial CASE studentship, and AB thanks the EPSRC for a project studentship.

FIGURE CAPTIONS

Figure 1. A subset of the raw data acquired for sample 2. The diffusion-weighted experimental spectra for the 3rd, 17th, 34th, and 50th time points are shown.

Figure 2. (a) Component spectra obtained from a non-negativity constrained three-component PARAFAC fit of the data acquired for sample 1 (top), and reference spectra (bottom); (b) concentration timecourses (circles) obtained from the PARAFAC fit, together with fits to sequential first order kinetics (solid lines). Estimated rate constants were $k_1 = 7.03 \pm 0.15 \times 10^{-5}$ s^{-1} and $k_2 = 1.58 \pm 0.02 \times 10^{-5}$ s^{-1} for the maltotriose and maltose hydrolyses respectively. The small deviations from the lines of best fit correlate in time with temperature disturbances of the order of 0.1 °C between 0 and 10 h and between 20 and 25 h.

Figure 3. (a) Component spectra obtained from a PARAFAC fit of the data acquired for sample 2 (540 mM initial sucrose concentration); (b) component spectra obtained by fitting the data acquired for sample 3 (50 mM initial sucrose concentration); (c) reference spectra of sucrose and of an equimolar mixture of glucose and fructose.

Figure 4. The diffusional decay and time evolution components obtained from PARAFAC fitting of the data originating from sample 3 (circles), together with (solid lines) non-linear least squares fits to the non-uniform gradient compensated Stejskal-Tanner equation (Eq. 1) and to first order kinetics respectively ("a" corresponds to fructose + glucose while "b" corresponds to sucrose). The fitted diffusion coefficients were $1.28 \pm 0.01 \times 10^{-10}$ m^2 s^{-1} for sucrose and $D = 1.37 \pm 0.01 \times 10^{-10}$ m^2 s^{-1} for the

product component (glucose and fructose), while the first order rate constant was 7.48 $\pm$ 0.03$\times10^{-5}$ s^{-1}.

Figure 5. (a) A subset of the experimental data (spectra as a function of pulsed field gradient and reaction time) obtained for sample 4 (1.1 mM initial sucrose concentration) showing data for the start, midpoint, and end of the period monitored, together (b) with the time evolution of the components corresponding to sucrose and to glucose/fructose, and (c) the fitted spectra. The fitted diffusion coefficients were 1.15 $\pm$ 0.05$\times10^{-10}$ m^2 s^{-1} for sucrose and 1.38 $\pm$ 0.03$\times10^{-10}$ m^2 s^{-1} for glucose/fructose, while the first order rate constant was 1.29 $\pm$ 0.02$\times10^{-4}$ s^{-1}.

REFERENCES

(1) Chrastil, J. *Computers & Chemistry* **1988**, *12*, 289-292
(2) Windig, W.; Antalek, B. *Chemometrics Intellig. Lab. Syst.* **1997**, *37*, 241-254
(3) Van Gorkom, L. C. M.; Hancewicz, T. M. *J. Magn. Reson.* **1998**, *130*, 125-130.
(4) Nilsson, M.; Morris, G. A. *Anal. Chem.* **2008**, *80*, 3777-3782.
(5) de Juan, A.; Tauler, R. *Anal. Chim. Acta* **2003**, *500*, 195-210.
(6) Binstead, R. A.; Stultz, L. K.; Meyer, T. J. *Inorg. Chem.* **1995**, *34*, 546-551.
(7) Binstead, R. A.; Chronister, C. W.; Ni, J. F.; Hartshorn, C. M.; Meyer, T. J. *J. Am. Chem. Soc.* **2000**, *122*, 8464-8473.
(8) Bezemer, E.; Rutan, S. *Anal. Chim. Acta* **2002**, *459*, 277-289.
(9) Harman, H. H. *Modern Factor Analysis*; University Of Chicago Press, 1976.
(10) Morris, G. A. In *Encyclopedia of Nuclear Magnetic Resonance*; Grant, D. M., Harris, R. K., Eds.; John Wiley & Sons Ltd: Chichester, 2002; Vol. 9 : Advances in NMR, pp 35-44.
(11) Johnson Jr, C. S. *Prog. Nucl. Magn. Reson. Spectrosc.* **1999**, *34*, 203-256
(12) Smilde, A.; Bro, R.; Geladi, P. *Multi-way Analysis: Applications in the Chemical Sciences*; John Wiley & Sons Chichester, UK, 2004.
(13) Pedersen, H. T.; Dyrby, M.; Engelsen, S. B.; Bro, R.; Webb, G. A. In *Annual Reports on NMR Spectroscopy*; Academic Press, 2006; Vol. Volume 59, pp 207-233.
(14) Nilsson, M.; Botana, A.; Morris, G. A. *Anal. Chem.* **2009**, *In Press*. DOI: 10.1021/ac901321w
(15) Dyrby, M.; Petersen, M.; Whittaker, A. K.; Lambert, L.; Nørgaard, L.; Bro, R.; Engelsen, S. B. *Anal. Chim. Acta* **2005**, *531*, 209-216
(16) Nilsson, M.; Khajeh, M.; Botana, A.; Bernstein, M. A.; Morris, G. A. *Chem. Commun.* **2009**, 1252-1254.
(17) Albert, K. *On-line LC-NMR and related techniques*; John Wiley and Sons, 2002.
(18) Kaufman, D.; Sterner, C.; Masek, B.; Svenningsen, R.; Samuelson, G. *J. Chem. Educ.* **1982**, *59*, 885-886.
(19) Stejskal, E. O.; Tanner, J. E. *J. Phys. Chem.* **1965**, *42*, 288-292
(20) Connell, M. A.; Bowyer, P. J.; Adam Bone, P.; Davis, A. L.; Swanson, A. G.; Nilsson, M.; Morris, G. A. *J. Magn. Reson.* **2009**, *198*, 121-131.
(21) Barjat, H.; Morris, G. A.; Smart, S.; Swanson, A. G.; Williams, S. C. R. *J. Magn. Reson., Ser B* **1995**, *108*, 170-172.
(22) Morris, K. F.; Johnson, C. S. *J. Am. Chem. Soc.* **1993**, *115*, 4291-4299.
(23) Nilsson, M.; Connell, M. A.; Davis, A. L.; Morris, G. A. *Anal. Chem.* **2006**, *78*, 3040-3045.
(24) Brand, T.; Cabrita, E. J.; Morris, G. A.; Günther, R.; Hofmann, H. J.; Berger, S. *J. Magn. Reson.* **2007**, *187*, 97-104.
(25) Johnson, C. S. *J. Magn. Reson., Ser A* **1993**, *102*, 214-218.
(26) Stilbs, P.; Paulsen, K. *Rev. Sci. Instrum.* **1996**, *67*, 4380-4386.
(27) Stilbs, P.; Paulsen, K.; Griffiths, P. C. *J. Phys. Chem.* **1996**, *100*, 8180-8189
(28) Bro, R. *Chemometrics Intellig. Lab. Syst.* **1997**, *38*, 149-171.
(29) Harshman, R. A.; Lundy, M. E. *Comput. Stat. Data Anal.* **1994**, *18*, 39-72.

(30) Tomasi, G.; Bro, R. *Comput. Stat. Data Anal.* **2006**, *50*, 1700-1734
(31) Pelta, M. D.; Morris, G. A.; Stchedroff, M. J.; Hammond, S. J. *Magn. Reson. Chem.* **2002**, *40*, S147-S152.
(32) Morris, G. A. In *Encyclopedia of Nuclear Magnetic Resonance*; Grant, D. M., Harris, R. K., Eds.; John Wiley & Sons Ltd: Chichester, 2002; Vol. 9 : Advances in NMR, pp 125-131.
(33) Morris, G. A.; Barjat, H.; Horne, T. J. *Prog. Nucl. Magn. Reson. Spectrosc.* **1997**, *31*, 197-257
(34) Andersson, C. A.; Bro, R. *Chemometrics Intellig. Lab. Syst.* **2000**, *52*, 1-4
(35) Nilsson, M. *J. Magn. Reson.* **2009**, *200*, 296-302.
(36) Bro, R. *Chemometrics Intellig. Lab. Syst.* **1999**, *46*, 133-147.
(37) Nilsson, M.; Morris, G. A. *Magn. Reson. Chem.* **2006**, *44*, 655-660.

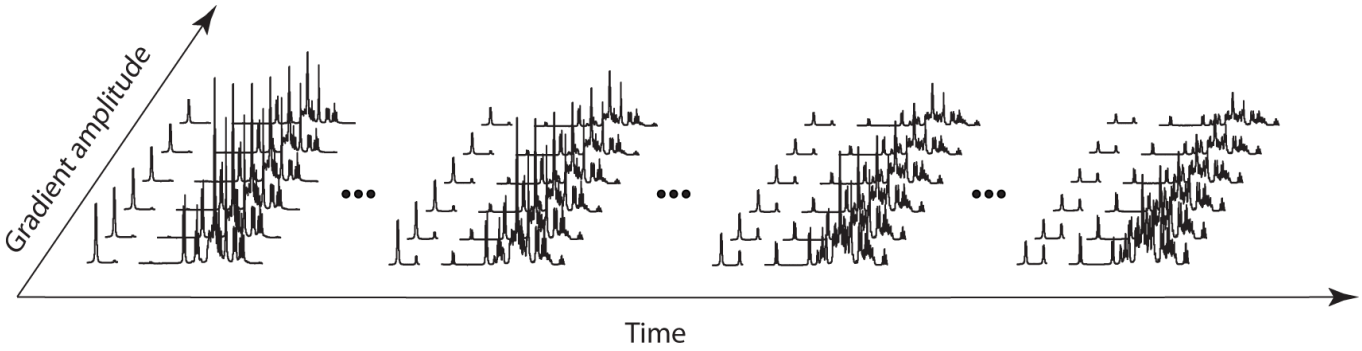
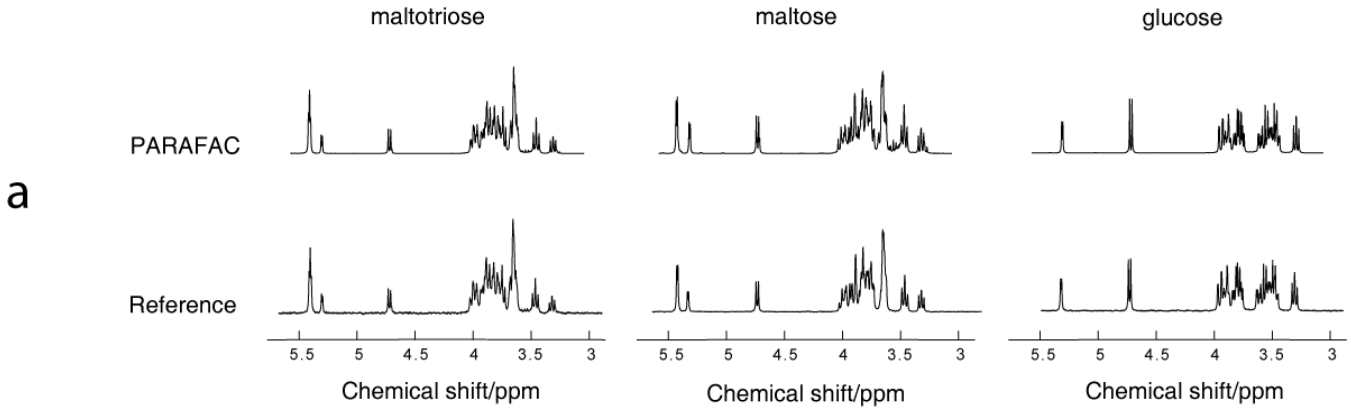


Figure 1. A subset of the raw data acquired for sample 2. The diffusion-weighted experimental spectra for the 3rd, 17th, 34th, and 50th time points are shown.



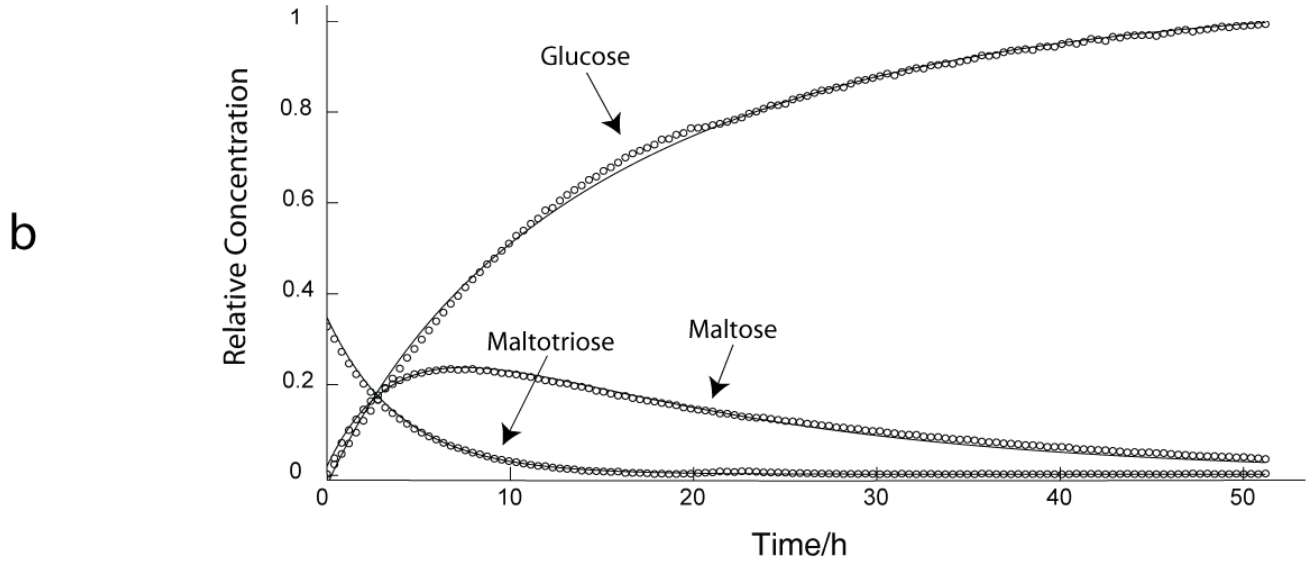


Figure 2. (a) Component spectra obtained from a non-negativity constrained three-component PARAFAC fit of the data acquired for sample 1 (top), and reference spectra (bottom); (b) concentration timecourses (circles) obtained from the PARAFAC fit, together with fits to sequential first order kinetics (solid lines). Estimated rate constants were $k_1 = 7.03 \pm 0.15 \times 10^{-5}$ s^{-1} and $k_2 = 1.58 \pm 0.02 \times 10^{-5}$ s^{-1} for the maltotriose and maltose hydrolyses respectively. The small deviations from the

lines of best fit correlate in time with temperature disturbances of the order of 0.1 °C between 0 and 10 h and between 20 and 25 h.

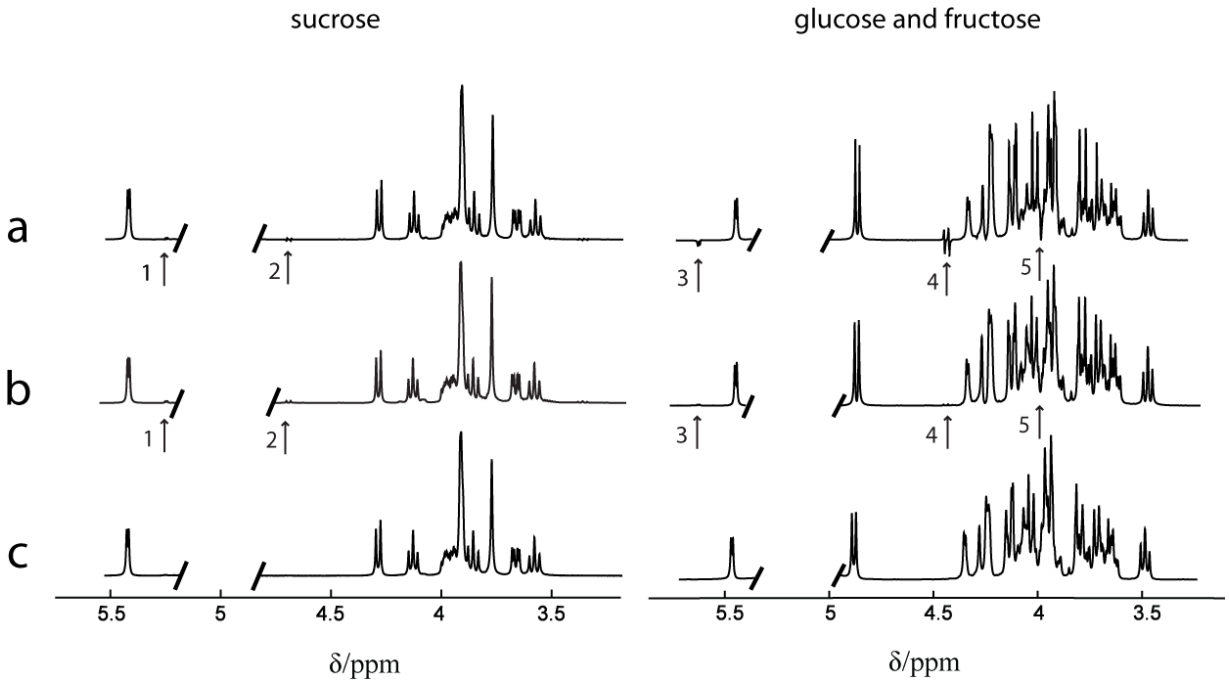


Figure 3. (a) Component spectra obtained from a PARAFAC fit of the data acquired for sample 2 (540 mM initial sucrose concentration); (b) component spectra obtained by fitting the data acquired for sample 3 (50 mM initial sucrose concentration); (c) reference spectra of sucrose and of an equimolar mixture of glucose and fructose.

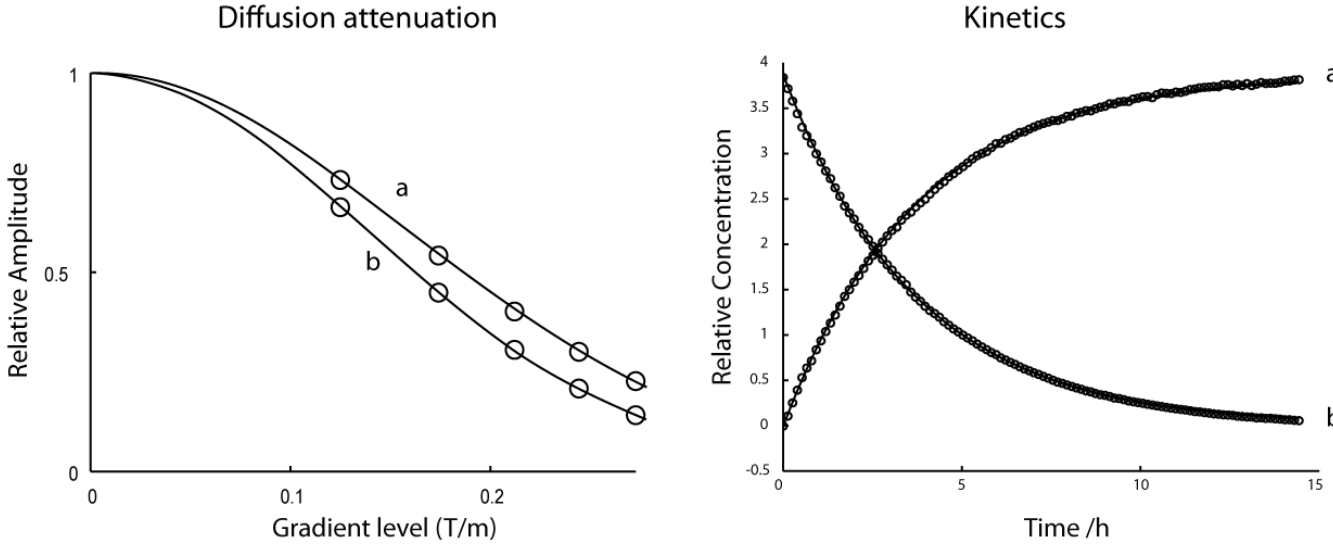


Figure 4. The diffusional decay and time evolution components obtained from PARAFAC fitting of the data originating from sample 3 (circles), together with (solid lines) non-linear least squares fits to the non-uniform gradient compensated Stejskal-

Tanner equation (Eq. 1) and to first order kinetics respectively (a corresponds to fructose + glucose while b corresponds to sucrose). The fitted diffusion coefficients were 1.28 ± 0.01×10^{-10} m^2 s^{-1} for sucrose and D = 1.37 ± 0.01×10^{-10} m^2 s^{-1} for the product component (glucose and fructose), while the first order rate constant was 7.48 ± 0.03×10^{-5} s^{-1}.

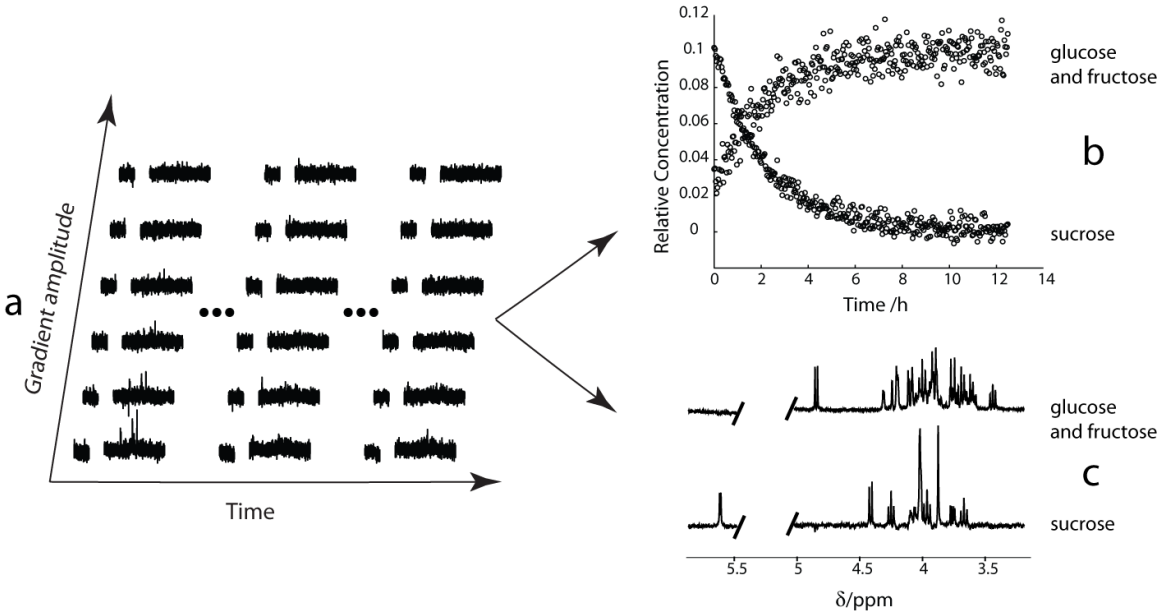


Figure 5. (a) A subset of the experimental data (spectra as a function of pulsed field gradient and reaction time) obtained for sample 4 (1.1 mM initial sucrose concentration) showing data for the start, midpoint, and end of the period monitored, together (b) with the time evolution of the components corresponding to sucrose and to glucose/fructose, and (c) the fitted spectra. The fitted diffusion coefficients were 1.15 ± 0.05×10^{-10} m^2 s^{-1} for sucrose and 1.38 ± 0.03×10^{-10} m^2 s^{-1} for glucose/fructose, while the first order rate constant was 1.29 ± 0.02×10^{-4} s^{-1}.

Appendix C

In press: Magnetic Resonance in Chemistry, 2010

A Simple Flowcell for Reaction Monitoring by NMR

M. Khajeh[a], M.A. Bernstein[b] and G.A. Morris[a]*

[a] School of Chemistry, The University of Manchester, Oxford Road, Manchester M13 9PL, UK

[b] AstraZeneca R&D Charnwood, Bakewell Rd, Loughborough, Leics. UK LE11 5RH

Sponsors: AstraZeneca, EPSRC (grant refs EP/D05592X, EP/E057888)

Short title: A Simple Flowcell for Reaction Monitoring

Proofs to: Gareth.Morris@manchester.ac.uk tel +44 (0)161 275 4665, fax +44 (0)161 275 4598

*Correspondence to: G.A. Morris, School of Chemistry, The University of Manchester, Oxford Road, Manchester, M13 9PL, UK. Email: Gareth.Morris@manchester.ac.uk

Abstract

A simple, cheap and flexible flowcell based on a standard 5 mm NMR tube, designed for the monitoring of reactions but of wide applicability, is described. No modification of the NMR instrument is needed, allowing the system to be employed with any conventional NMR probe and magnet. The system is robust, economical in use of reagents, and can be used for studying both homogeneous and heterogeneous reactions.

Keywords

NMR, ^{1}H, flowcell, kinetics, reaction monitoring

Introduction

The application of NMR spectroscopy to kinetic and mechanistic studies of chemical reactions is not limited to static methods, in which a solution is placed in an NMR tube and inserted into a magnet for measurements, but also includes stopped [1] or continuous flow methods [2]. Flow NMR techniques have been used since 1951 [3] in a variety of areas, such as studies of protein folding [4, 5], food science [6], and HPLC-NMR [7]; recent examples of the use of flow NMR for reaction monitoring include references [8-10].

Most applications of flow NMR use dedicated flow probes. These allow the balance between sample volume, sensitivity and resolution to be optimised, but are expensive and are relatively inflexible, typically being optimised for a single nucleus or group of nuclei; they also generally have a limited temperature range and are very vulnerable to blockage. This paper describes a simple, cheap and flexible flowcell based on a standard 5 mm NMR tube. Several designs for rapid mixing cells based on 5 mm NMR tubes have recently been published [9, 11]; in contrast, the cell described here is designed for continuous or stopped-flow use (but not for reactions on sub-minute timescales). No modification of the NMR instrument is needed, allowing the system to be employed with any conventional 5 mm or greater NMR probe and magnet. The system is robust, economical in use of reagents, and can be used for studying both homogeneous and heterogeneous reactions. The only materials in contact with the reaction mixture are glass, PTFE and PEEK, accommodating a wide range of solvents and reactants. The reaction vessel is placed outside of the magnet; therefore the composition of the sample can be varied during the experiment, if any reagents need to be added. It is also relatively straightforward to perform reactions at high or low temperature, or under an inert atmosphere The spectral quality obtainable is comparable to that available from conventional high resolution probes.

Experimental

Flowcell

Fig. 1 here

The flowcell (Figs. 1 and 2) consists of a standard 5 mm NMR tube (6) fitted with a PEEK headpiece (1), through which a length of PEEK HPLC tubing (8) runs concentrically to the base of the tube. The PEEK tubing is held concentric by grooved PTFE spacers at the bottom of the tube and above the active volume, and by four PTFE grooved sleeves (18) that also serve to reduce the dead volume of the cell. The NMR tube is held in the headpiece by an M8 threaded bush (4) that compresses an O-ring (5). The inlet HPLC tubing runs continuously through the headpiece, and the outlet tube (9) is mated to the headpiece, both held by O-rings (3) compressed by M3 threaded bushes (2). The flowcell is held in the appropriate turbine (17), at the appropriate depth, for the probe used. As a precaution against probe damage from reactant leaks, a liquid sensor (7; Sensortechnics, type OLP01B0F3) sits below the headpiece. If the turbine design permits (as with Varian turbines), the active element of the sensor sits directly on the turbine, and a glass tube (16) is glued to the turbine with epoxy resin to contain any leaks; if not (as with Bruker turbines), the sensor sits on a PEEK disc through which the NMR tube passes and which is held within the glass tube, in both cases sealed with O-rings. The liquid sensor is connected to an electronic interlock (21) that switches off the pump in the event of a leak being detected. (To date the only occasion on which the safety system was activated was when a loose union above the headpiece leaked as a result of a reactant precipitating in the system).

A number of features of the design are noteworthy. First, the use of a concentric inlet pipe minimises disturbance of the magnetic field homogeneity by maintaining cylindrical symmetry, while the use of narrow bore tubing ensures that the frequency-shifted signal from within the tubing is very small. Any deviation from concentricity will lead to significant degradation in lineshape and difficulty in shimming. Second, the cell geometry is intended for relatively low flow rates (< 2 mL min^{-1}), since prepolarization takes place largely in the portion of the NMR tube between the base and the active volume. If higher flow rates are required, for which the residence time in the prepolarization volume is less than the spin-lattice relaxation time, an extra reservoir could be used above the flowcell headpiece, or the direction of flow could be reversed. Third, the system is designed for low pressure, fail-safe operation; if the pressure rises too much (significantly above 1 MN m^{-2}), the O-ring holding the NMR tube to the headpiece should yield well before the pressure limit of the NMR tube

itself, leaking into the space above the turbine and triggering the liquid detector. The system may be used with HPLC pumps, but in this case a low pressure limit should be used. Fourth, the PEEK HPLC tubing used may easily be replaced with different inside diameters, to accommodate solvents of different viscosities within the pressure constraints of the system.

Flow system

Fig. 2 here

The flowcell is designed to be used with an external reaction vessel, outside of the magnet, with reactant mixture circulated by a low pressure pump. The system used here, illustrated in Fig. 2, comprises a reaction vessel (10), HPLC PEEK and PTFE piping, unions, peristaltic pump (14), flowcell headpiece (1), and NMR tube (6). The first inlet pipe (11; PTFE tubing, length 30 cm, ID 1.6 mm, OD 3.2 mm) leads from the reaction vessel (10) to the peristaltic pump (14; Cole-Parmer, type KH-77912-07), connected by a union (12). The pump uses 4 mm OD, 2 mm ID PTFE peristaltic tubing (13; Cole-Parmer, type WW 77390-50). The output of the pump is connected by a union (15) to the flowcell inlet tube (8) (PEEK, length 3 m, ID 0.5 mm, OD 1.6 mm) that leads directly to the bottom of the NMR tube (6). The outlet pipe (9; PEEK tubing, length 3 m, ID 0.5 mm, OD 1.6 mm leads from the headpiece (1) back to the reaction vessel. The total volume of reaction mixture required to fill the pipes, pump and flowcell is approximately 3.5 mL; this could be reduced somewhat by the use of lower bore and shorter peristaltic tubing.

The flowcell was placed in the turbine (17) and lowered by its input and output tubing into the standard 5 mm diameter indirect detection probe of a 300 MHz Varian INOVA spectrometer. All experiments were conducted at room temperature (20 ± 2 °C).

Results

Spectra

Spectra obtained with the flow cell typically showed good resolution and acceptable lineshape, with a full width at half height of around 1 Hz both under static conditions and at low (ca. 1 mL min^{-1}) flow rate. Figure 3 shows the spectrum obtained for a lineshape test solution of chloroform in acetone-d_6 containing a small amount of ethanol stabiliser. The full width at half height of the TMS signal was 0.75 Hz, only slightly greater than that routinely obtained with conventional non-spinning samples.

Fig. 3 here

A sample homogeneous reaction

In the reaction of 2-methoxyphenyl acetate with phenylethylamine in aqueous solution, transesterification (Scheme 1a) and hydrolysis (Scheme 1b) proceed in parallel. The hydrolysis is catalysed by the amine, which is gradually consumed, complicating kinetic analysis.

Scheme 1 here

The two 2nd order parallel reactions in scheme 1 can be summarised as follows:

$$\begin{aligned} A + B &\xrightarrow{k_1} C + D \\ A + B &\xrightarrow{k_2} C + AH^+ + E \end{aligned} \qquad (1)$$

where A denotes phenylethylamine, AH^+ the phenylethylammonium ion, B 2-methoxyphenyl acetate, C 2-methoxyphenol, D phenylethylacetamide, and E acetate ion. The rates of change of species A to E are then

$$\frac{d[A]}{dt} = -k_1([A]-[E])[B]$$
$$\frac{d[B]}{dt} = -k_1([A]-[E])[B] - k_2([A]-[E])[B]$$
$$\frac{d[C]}{dt} = k_1([A]-[E])[B] + k_2([A]-[E])[B]$$
$$\frac{d[D]}{dt} = k_1([A]-[E])[B]$$
$$\frac{d[E]}{dt} = k_2([A]-[E])[B] \qquad (2).$$

To follow this reaction, the flow system was initially filled by pumping a solution of phenylethylamine and TSP (3.5 mL of 0.03 M phenethylamine in D_2O, and 300 µL of 0.24 M TSP in D_2O) into the system. After filling, flow was stopped to allow locking and shimming of the magnetic field and setup of presaturation of the residual HDO signal. A mixture of 2 mL of 0.03 M 2-methoxyphenyl acetate (Sigma) in D_2O, 2 mL of 0.03 M phenylethylamine (Alfa Aesar) in D_2O, and 300 µL of 0.24 M TSP (sodium 3-(trimethylsilyl)-propionate-2,2,3,3,-d_4; Goss Scientific) as reference was then prepared, and immediately placed in the reaction vessel. The input pipe to the peristaltic pump was placed in the reaction vessel, pumping initiated at 1 mL/min, and acquisition of a timecourse of ^{1}H spectra initiated. For the first 3.5 min of the timecourse the output from the flowcell was diverted to waste, to discard the initial contents of the system; the output was then returned to the reaction vessel to recycle the reacting mixture. The total duration of the experiment was 5 h, with a recycle delay of 7 s, 4 transients per acquisition, a saturation delay of 4 s, and a 90° pulse of duration 9 µs, for each ^{1}H spectrum. The spectral timecourse array for the reaction is illustrated in Figure 4.

Fig. 4 here

All the data obtained were Fourier transformed, reference deconvoluted using a 2.4 Hz Lorentzian target lineshape, normalised to correct systematic drifts in receiver gain, and baseline corrected in Vnmr 6.1C. Data were then exported to Mathematica version 6.0[12] as a text file for statistical analysis and fitting. The peaks marked in Fig. 4 were

fitted to Lorentzian functions, with width, amplitude, frequency, phase, and zero and first order baseline corrections as variable parameters. In the case of the triplet signals a, b, e and f the relative intensities of the inner and outer triplet components were constrained to equal that of the central component, and the triplet splittings were held constant; in both cases this is a good approximation for the degree of strong coupling present. For the first spectrum in the series, the initial parameter values for the singlet Lorentzian functions were estimated by inspection. From then on, each successive spectrum was fitted using starting parameters extrapolated from the preceding fits. This minimises the number of iterations required, reduces problems with signals that drift in frequency, and allows signal assignments to be retained even when the chemical shift ordering changes during a reaction. For the four triplet signals, fitting was performed in reverse order, starting with the last spectrum, because of the overlap between triplets b and f at the start of the reaction. In the very early spectra, where f is completely buried under b, the intensities of the f signals were constrained to equal those of e, allowing b and f to be separated in the fitting.

The integrals obtained were converted to reactant concentrations using the known initial total concentration. Signals a and b reflect total amine concentration, as a result of rapid exchange between free amine (A) and ammonium (AH^+) species, so the free amine concentration was obtained by subtraction of the acetate concentration known from signal i. The concentration time courses were fitted to numerical solutions of equations (2), with the results shown in Fig. 5. The slight oscillations in reactant concentration seen in Figs. 5a and 5b are caused by imperfect mixing in the pipework, pump and flowcell at the start of the experiment; not all of the initial contents of the system were flushed out by the reactant mixture, with the result that small damped oscillations in reactant concentration occur with a period equal to the transit time for the system. This problem can be reduced (at the expense of an increase in the total volume of reactants required) by increasing the time for which the flowcell output is discarded, or (at the risk of trapping bubbles) by starting the reaction with the flow system empty.

Fig. 5 here

A sample heterogeneous reaction

Heterogeneous reactions are difficult to study in static conditions in an NMR tube. As the magnetic susceptibilities of solids and liquids are normally different, the homogeneity of the magnetic field is severely perturbed, and it is difficult to maintain the solid component in suspension. One of the advantages of the flow system described is the ability to study heterogeneous reactions while avoiding having solid particles in the detection region of the probe. As a test heterogeneous reaction, the reductive amination of an aldehyde was studied. As shown in scheme 3, the reaction chosen was between benzylamine (Alfa Aesar) and p-tolualdehyde (Sigma) in (protio-)THF/acetic acid with cyanoborohydride on a polymer support (Sigma, product no. 17337). The reaction cell containing the solid phase and reactant mixture was continuously stirred to keep the polymer-supported cyanoborohydride in suspension, while a stainless steel frit on the inlet of the pipe leading to the pump was used to retain the solid in the reaction vessel and keep it from blocking the pipework or being pumped into the probe. An excess of primary amine was used in an attempt to side products such as dialkylated amine [13], but these could not be completely avoided. The reaction mixture consisted of 2.43 mL of 0.26 M benzylamine in tetrahydrofuran (THF), 2.43 mL of 0.13 M p-tolualdehyde in THF, 0.34 mL glacial acetic acid, 1.82 mL THF, 434 mg polymer-supported cyanoborohydride and 250 μL TMS (scheme 3).

Scheme 2 here

Spectra were run unlocked, with the THF resonance at 3.6 ppm presaturated; broadband ^{13}C decoupling was used during acquisition to avoid ^{13}C satellites of THF signals overlapping with signals of interest. The total duration of the experiment was 6 h 19 min, with a recycle delay of 8 s, 16 transients per spectrum, a saturation delay of 3 s, saturation power 5 dB, and a 30° flip angle (3 μs pulse width). Immediately after mixing, the reaction solution (7 mL) was pumped into the dry, empty flow system at a flow rate of 1 mL/min. Data acquisition was started when the first drop emerged from the output pipe; representative spectra from the timecourse are shown in Fig. 6.

Fig. 6 here

Initial data processing, export and fitting to Lorentzian lineshapes were carried out as described for the homogeneous reactions above. The integrals of the methyl peaks d, g

and i were converted to concentrations and their timecourses fitted to the analytical solutions for first order parallel kinetics (Scheme 3)

Scheme 3 here

with the results shown in Fig. 7.

Fig. 7 here

Conclusions

The flowcell described offers a simple and economical route to the acquisition of spectra under stopped or continuous flow. While primarily designed for reaction monitoring, it is adaptable to a wide range of uses; one example would be to allow hydrogen gas to be bubbled through a sample in para-hydrogen induced polarization (PHIP) experiments, The resolution, lineshape and sensitivity obtainable are only slightly poorer than for a standard 5 mm NMR tube, and the system is compatible with a wide range of probes and nuclei. It is compatible with standard NMR instrumentation, and requires only small (mL) volumes of solution.

Acknowledgments

Support from the Engineering and Physical Sciences Research Council (grant references EP/D05592X and EP/E057888/1) and from AstraZeneca (studentship to MK) is gratefully acknowledged.

References

[1] D. B. Green, J. Lane, R. M. Wing, *Appl. Spectrosc*. **1987**, *41*, 847.

[2] E. Bayer, K. Albert, *J. Chromatogr. A* **1984**, *312*, 91.

[3] G. Suryan, *Proc. Indian Acad. Sci.* **1951**, *33*, 107.

[4] M. Hamang, A. Sanson, L. Liagre, V. Forge, P. Berthault, *Rev. Sci. Instrum.* **2000**, *71*, 2180.

[5] N. A. J. van Nuland, V. Forge, J. Balbach, C. M. Dobson, *Acc. Chem. Res.* **1998**, *31*, 773.

[6] C. Tellier, F. Mariette, P. S. B. G.A. Webb, M. J. Mccarthy, *Ann. Rep. NMR Spectrosc.*, **1995**, *31*, 105.

[7] N. Watanabe, E. Niki, *Proc. Jpn. Acad. Ser. B* **1978**, *54*, 194.

[8] M. A. Bernstein, M. Scarontefinovi, C. J. Sleigh, *Magn. Reson. Chem,* **2007**, *45*, 564.

[9] A. Mix, P. Jutzi, B. Rummel, K. Hagedorn, *Organometallics, in press, DOI:10.1021/om900919f* **2009**.

[10] M. Maiwald, H. H. Fischer, Y.-K. Kim, K. Albert, H. Hasse, *J. Magn. Reson*, ***2004***, *166*, 135.

[11] P. V. Yushmanov, I. Furó, *J. Magn. Reson.* **2005**, *175*, 264.

[12] Wolfram Research, Inc., Mathematica, Version 6.0, Champaign, IL (2007)

[13] A. F. Abdel-Magid, K. G. Carson, B. D. Harris, C. A. Maryanoff, R. D. Shah, *J. Org. Chem.* **1996**, *61*, 3849.

Figure Captions

1. (Left) scale drawing, and (right) photographs of the flowcell headpiece assembly. (1) headpiece; (2) inlet and outlet pipe bushes; (3), (5) O-rings; (4) NMR tube retaining bush. The left part of the composite photograph shows the inlet and outlet pipes and their respective bushes and O-rings, together with the liquid sensor cable, the latter and the inlet pipe running continuously through the headpiece. The top right shows the top view, and the bottom right the bottom view, of the headpiece.

2. Structure of the flow system used. (1) flowcell headpiece; (2) inlet and outlet pipe bushes; (3), (5) O-rings; (4) NMR tube retaining bush; (6) NMR tube; (7) optical liquid sensor; (8) inlet pipe; (9) outlet pipe; (10) reaction vessel; (11) reaction vessel outlet pipe; (12), (15) unions; (13) PTFE peristaltic tubing; (14) peristaltic pump; (16) glass tube, glued to turbine (17); (18) long grooved PTFE spacers; (19) short grooved PTFE spacers; (20) cross-section of spacers; (21) safety cutoff.

3. ^{1}H spectrum obtained with the flow cell under static conditions for a solution containing protiochloroform, ethanol and TMS in deuterioacetone.

4. Timecourse array for the solution of phenylethylamine and 2-methoxyphenyl acetate in D_2O, with every 35^{th} spectrum from the first (bottom) to the last (top) shown, from 0 to 4 h reaction time. The lower case letters from ‘a’ to ‘i’ correspond to the methyl and methylene signals highlighted in Scheme 1. During the course of the reaction the chemical shifts of signals ‘a’, ‘b’ and ‘h’ change due to the change in pH as hydrolysis proceeds.

5. Experimental (dots) and calculated (solid lines) kinetic plots for reactants and products in the reactions of scheme 1. (a) Kinetic plot for phenethylamine (reactant) and N-phenylethylacetamide (product), signals ‘b’ and ‘f’ in figure 4. (b) Kinetic plot of 2-methoxyphenyl acetate (reactant) and 2-methoxy phenol (product), signals ‘d’ and ‘h’ in figure 4. (c) Kinetic plot of acetate, signal ‘i’ in figure 4. Optimised values of k_1 and k_2 were 4.0×10^{-3} L mol^{-1} s^{-1} and 5.9×10^{-3} L mol^{-1} s^{-1} respectively. The

effects of incomplete mixing are seen in the oscillations in the reactant curves (a) and (b) for about the first 10 min of the reaction; these data were omitted from the fitting.

6. Timecourse array for a solution of benzylamine and p-tolualdehyde in THF; every 10^{th} spectrum is shown the first (bottom) to the last (top). The lower-case letters from ‘a’ to ‘i’ correspond to proton signals highlighted in scheme 2. The starred signals are *1: carboxylic proton of acetic acid, in exchange with residual water; *2 overlapped aromatic signals of reactants and products; *3 THF signal, partially saturated; *4 impurity; *5 methyl of acetic acid; *6 THF signal, not saturated. The signals from the reactant aldehyde are not observed as the initial imine formation is very fast. One of the methylene signals of the putative byproduct is not assigned, and is assumed to be overlapped by another signal. Multiple signals of minor species are seen. The effect of small temperature changes during initial equilibration can be seen in the moving COOH signal *1.

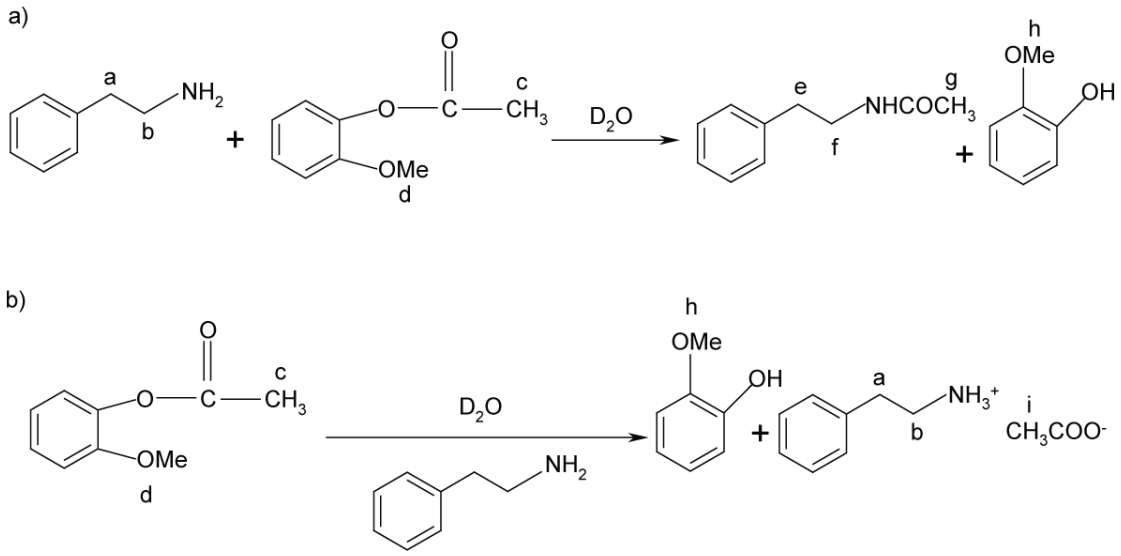
7. Kinetic plot for the signals ‘d’, ‘i’, ‘g’ in Fig. 6. The dots represent the experimental data, which were fitted to first order parallel kinetics (solid lines), yielding a k_1 of $1.86 \pm 0.08 \times 10^{-4}$ L mol^{-1} s^{-1} and k_2 of $6.08 \pm 0.011 \times 10^{-4}$ L mol^{-1} s^{-1}.

Scheme captions

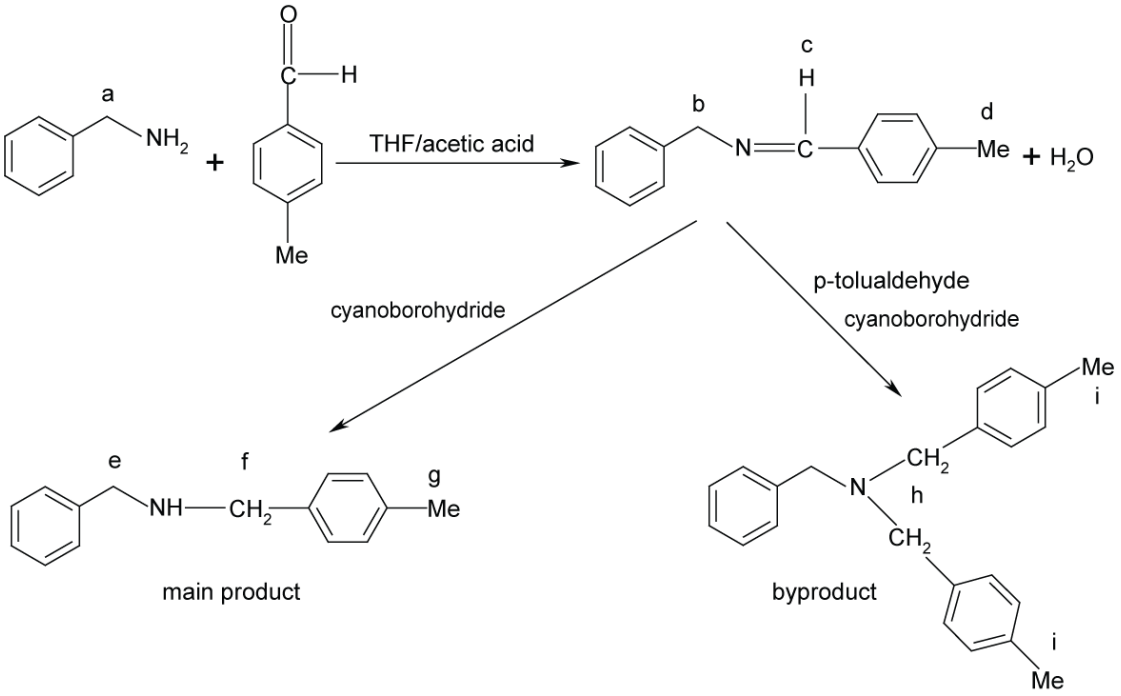
Scheme 1. Reactions of 2-methoxyphenyl acetate with phenylethylamine in D_2O. Lower case letters denote assignments for methylene and methyl protons in the spectra of Figure 4.

Scheme 2. Reductive amination of benzylamine with p-tolualdehyde, involving para-dimethyl benzyl benzylimine as an intermediate. The imine is reduced to para-dimethyl benzyl benzylamine as the main product; the byproduct (signals h, i) was tentatively identified as 2-(para-dimethyl benzyl) benzylamine but this was not confirmed. The protons for which signals are assigned in the spectra below are marked with lower-case letters.

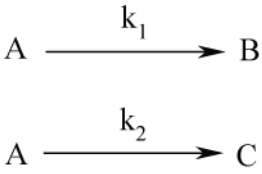
Scheme 3. Parallel first order kinetics for the conversion of imine (A) to product (B) and byproduct (C).



Scheme 1. Reactions of 2-methoxyphenyl acetate with phenylethylamine in D_2O. Lower case letters denote assignments for methylene and methyl protons in the spectra of Figure 4.



Scheme 2. Reductive amination of benzylamine with p-tolualdehyde, involving para-dimethyl benzyl benzylimine as an intermediate. The imine is reduced to para-dimethyl benzyl benzylamine as the main product; the byproduct (signals h, i) was tentatively identified as 2-(para-dimethyl benzyl) benzylamine but this was not confirmed. The protons for which signals are assigned in the spectra below are marked with lower-case letters.



Scheme 3. Parallel first order kinetics for the conversion of imine (A) to product (B) and byproduct (C).

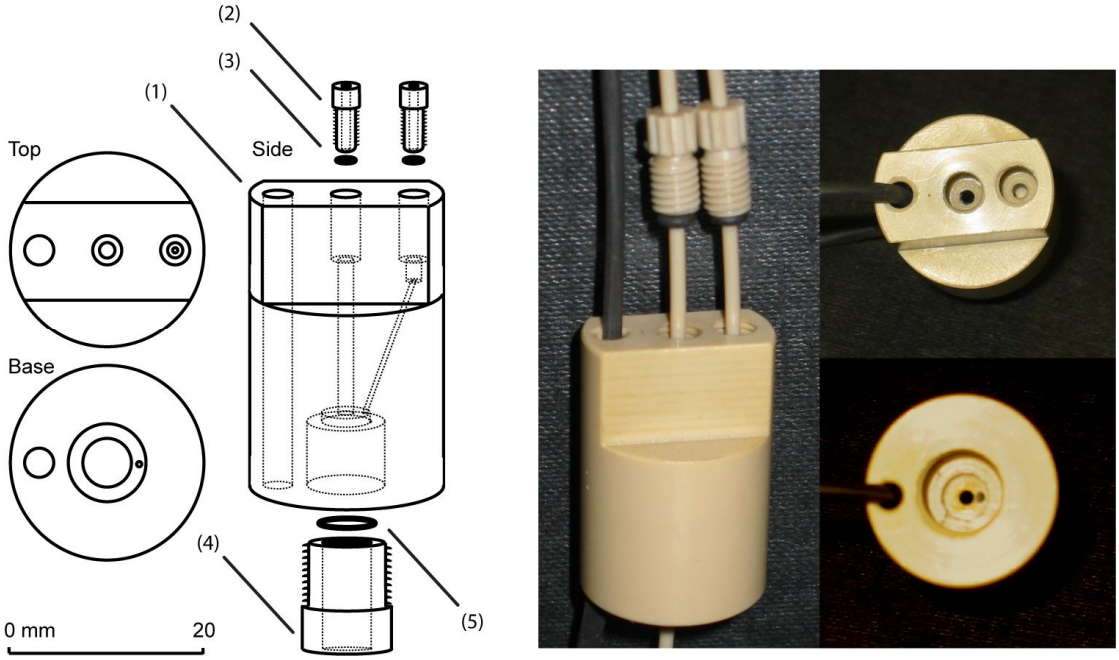


Figure 1. (Left) scale drawing, and (right) photographs of the flowcell headpiece assembly. (1) headpiece; (2) inlet and outlet pipe bushes; (3), (5) O-rings; (4) NMR tube retaining bush. The left part of the composite photograph shows the inlet and outlet pipes and their respective bushes and O-rings, together with the liquid sensor cable, the latter and the inlet pipe running continuously through the headpiece. The top right shows the top view, and the bottom right the bottom view, of the headpiece.

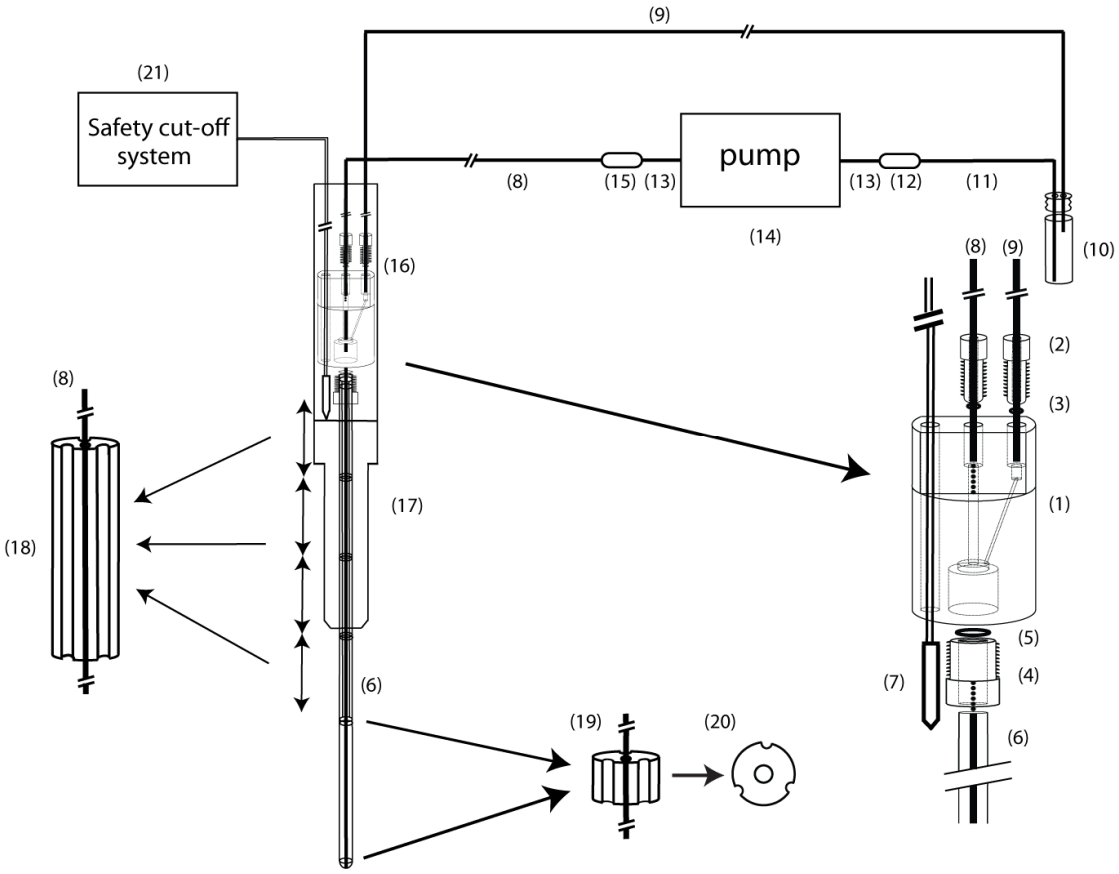


Figure 2. Structure of the flow system used. (1) flowcell headpiece; (2) inlet and outlet pipe bushes; (3), (5) O-rings; (4) NMR tube retaining bush; (6) NMR tube; (7) optical liquid sensor; (8) inlet pipe; (9) outlet pipe; (10) reaction vessel; (11) reaction vessel outlet pipe; (12), (15) unions; (13) PTFE peristaltic tubing; (14) peristaltic pump; (16) glass tube, glued to turbine (17); (18) long grooved PTFE spacers; (19) short grooved PTFE spacers; (20) cross-section of spacers; (21) safety cutoff.

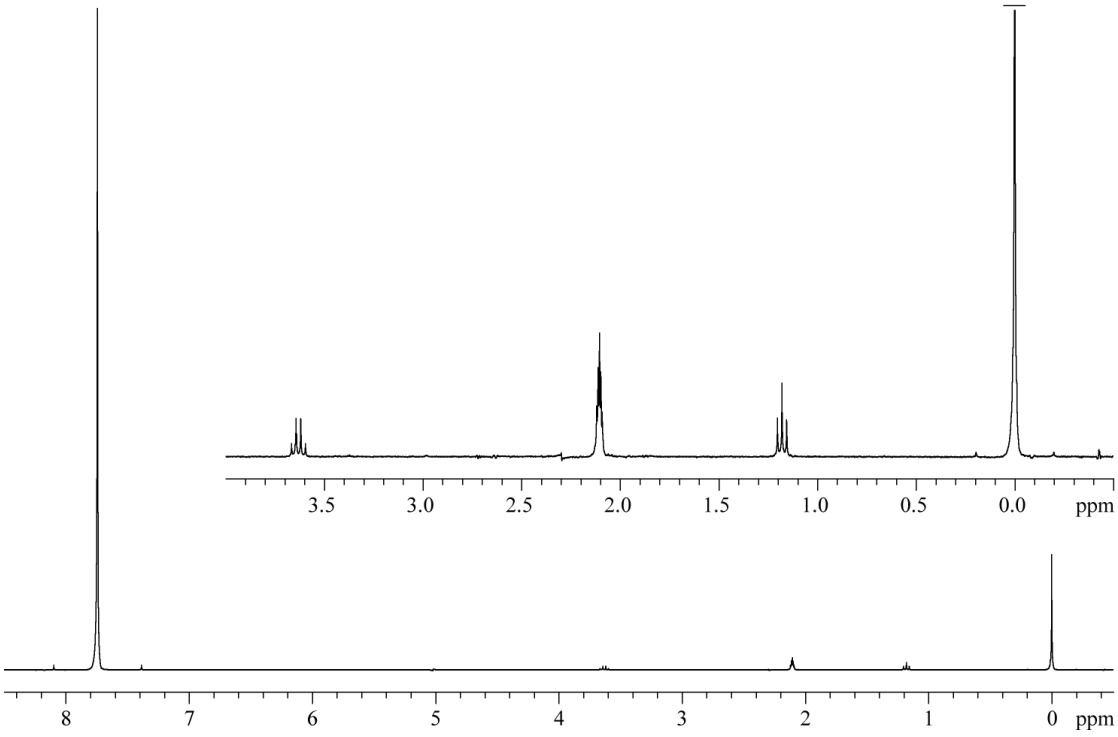


Figure 3. ^{1}H spectrum obtained with the flow cell under static conditions for a solution containing protiochloroform, ethanol and TMS in deuterioacetone.

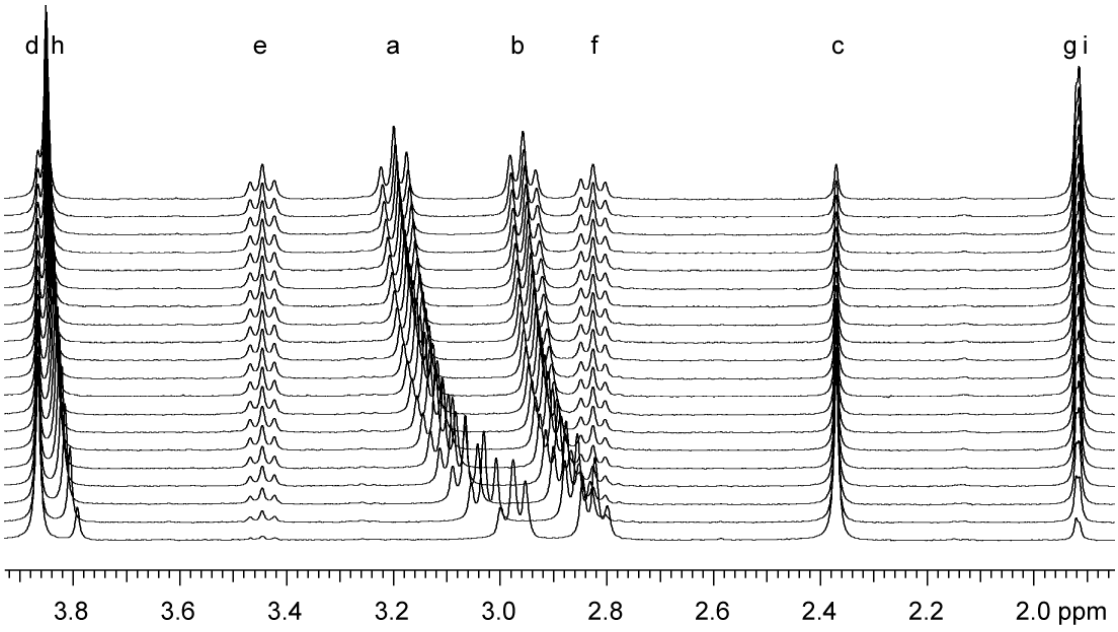


Figure 4. Timecourse array for the solution of phenylethylamine and 2-methoxyphenyl acetate in D_2O, with every 35th spectrum from the first (bottom) to the last (top) shown, from 0 to 4 h reaction time. The lower case letters from 'a' to 'i' correspond to the methyl and methylene signals highlighted in Scheme 1. During the course of the reaction the chemical shifts of signals 'a', 'b' and 'h' change due to the change in pH as hydrolysis proceeds.

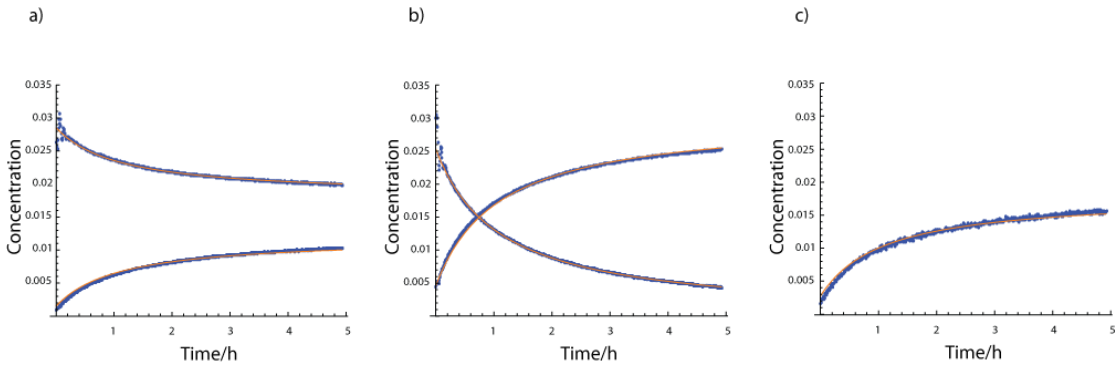


Figure 5. Experimental (dots) and calculated (solid lines) kinetic plots for reactants and products in the reactions of scheme 1. (a) Kinetic plot for phenethylamine (reactant) and N-phenylethylacetamide (product), signals 'b' and 'f' in figure 4. (b) Kinetic plot of 2-methoxyphenyl acetate (reactant) and 2-methoxy phenol (product), signals 'd' and 'h' in figure 4. (c) Kinetic plot of acetate, signal 'i' in figure 4. Optimised values of k_1 and k_2 were 4.0×10^{-3} L mol^{-1} s^{-1} and 5.9×10^{-3} L mol^{-1} s^{-1} respectively. The effects of incomplete mixing are seen in the oscillations in the reactant curves (a) and (b) for about the first 10 min of the reaction; these data were omitted from the fitting.

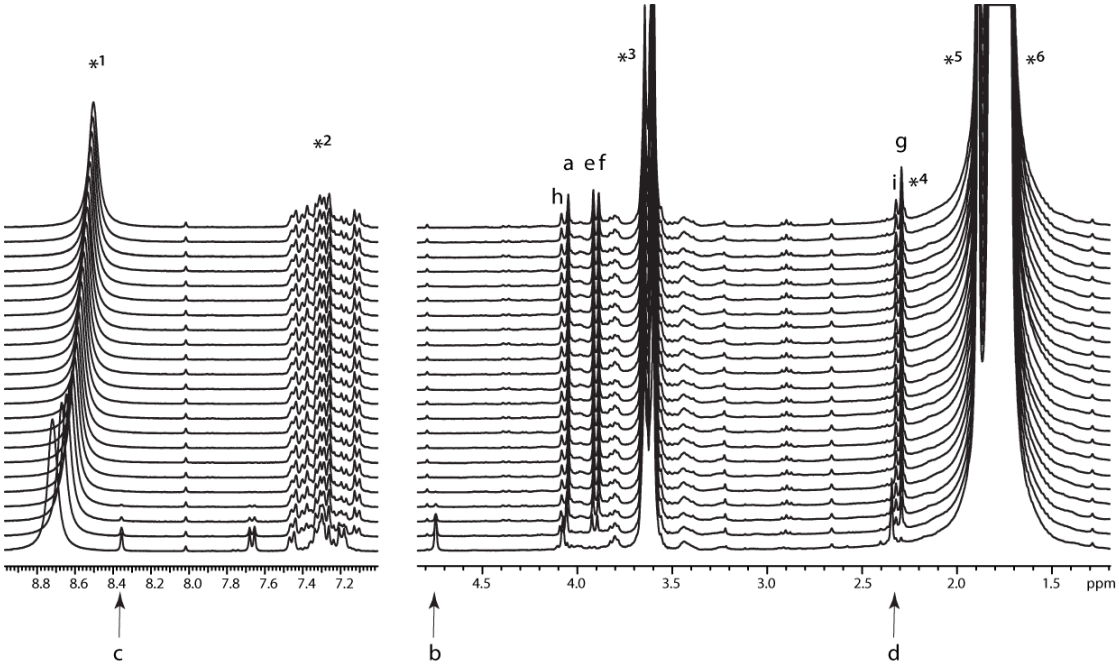


Figure 6. Timecourse array for a solution of benzylamine and p-tolualdehyde in THF; every 10th spectrum is shown the first (bottom) to the last (top). The lower-case letters from 'a' to 'i' correspond to proton signals highlighted in scheme 2. The starred signals are *1: carboxylic proton of acetic acid, in exchange with residual water; *2 overlapped aromatic signals of reactants and products; *3 THF signal, partially saturated; *4 impurity; *5 methyl of acetic acid; *6 THF signal, not saturated. The signals from the reactant aldehyde are not observed as the initial imine formation is very fast. One of the methylene signals of the putative byproduct is not assigned, and is assumed to be overlapped by another signal. Multiple signals of minor species are seen. The effect of small temperature changes during initial equilibration can be seen in the moving COOH signal *1.

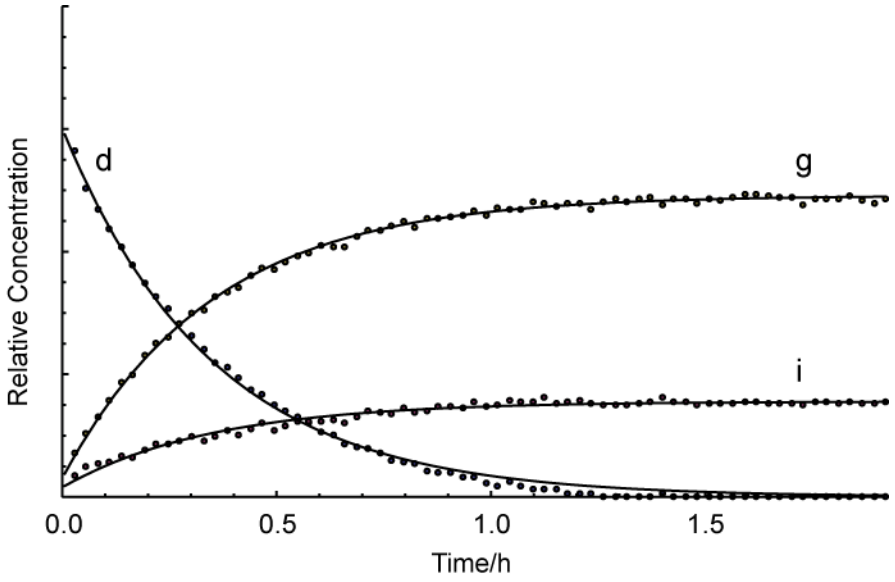


Figure 7. Kinetic plot for the signals ‘d’, ‘i’, ‘g’ in Fig. 6. The dots represent the experimental data, which were fitted to first order parallel kinetics (solid lines), yielding a k_1 of $1.86 \pm 0.08 \times 10^{-4}$ L mol^{-1} s^{-1} and k_2 of $6.08 \pm 0.011 \times 10^{-4}$ L mol^{-1} s^{-1}.