PAW — the Protein Analysis Workshop for 2D Nuclear Magnetic Resonance Spectroscopy

VOLUME I

Principles, Methods and Results

A thesis presented in partial fulfilment of the
requirements for the degree of
Doctor of Philosophy
in Physics
at Massey University
New Zealand

Wilford Lie

1999

The study described in this thesis has been supported by AgResearch, the New Zealand Pastoral Agriculture Research Institute, a division of the Crown Research Institute of New Zealand.

"Love is all there is."

— To my mum, my wife and daughters

Abstract

An X Window-based software package for SGI workstations has been developed to process and assign NMR spectra. Special consideration has been given to the assignment of two-dimensional ^{1}H NMR spectra of proteins.

The program combines features from the packages PROSPA [Eccles 1995], EASY [Eccles 1991] and FELIX [Biosym 1995] as well as having its own capabilities. It allows simultaneous display of multiple toolboxes and spectra, which can be flexibly manipulated by mouse operations, command entries, and user-editable macros.

NMR spectra can be processed either interactively or with macros containing commands with parameters. A unique filter that combines the exponential and sine-bell functions has been frequently used. A water suppression technique based on fitting averaged time-domain data, as well as an efficient algorithm for calculating fast Fourier transform and Hilbert transform [Eccles 1995] are discussed and implemented.

NMR spectral assignment is done interactively in three steps: peak picking, spin-system identification, and sequence-specific assignment. The process utilises three peak lists: a *raw-peak list* that contains records of all possible peaks in a NOESY spectrum, a *diagonal peak list* that contains records of peaks that define a curve about which the spectrum is symmetric, and a *cross-peak list* that contains records of peaks that are assigned. Details of the peak-picking methods are discussed.

By reference to a list of diagonal peaks, a common calibration problem caused by Bloch-Siegert shifts [Bloch and Siegert 1940, Ernst 1987] has been minimised. Automatically produced NOE summaries allow a quick identification of peaks that are unassigned or incorrectly assigned. The peak position and integration parameters can be calculated through non-linear curve fitting with Gaussians.

NMR data processing and spectral assignment using the package has been completed for Caerin 4.1, a 23-residue protein. Linear-prediction has been applied to increase the spectral resolution. Detailed results for this protein are presented. The NOE summary of the sequential assignments indicates a well-defined secondary structure that is different from Caerin 1.1 [Wong 1996, 1997].

Acknowledgements

The author would like to thank the following organisations and persons for their support and involvement in the study.

- AgResearch, the New Zealand Pastoral Agriculture Research Institute, for providing scholarship, office, and access to computing facilities and the VARIAN 500 NMR spectrometer.
- The Physics Department and the Institute of Fundamental Sciences, Massey University, for providing a PhD candidacy, a study space, and access to computing facilities.
- Dr Craig Eccles, the chief supervisor for the study, for his marvellous patience and the enormous time spent in supervising every aspect of the study.
- Dr Paul Atkinson of AgResearch, a co-supervisor for the study, for his initiation of the project, helpful supervision, and endless support.
- Dr Herbert Wong of Industrial Research Limited, another co-supervisor, for his great help with NMR experiments, theory and spectral assignment.
- Professor Paul Callaghan, for his wonderful teaching of advanced quantum mechanics and other support.
- Dr Mick Roberts of AgResearch, for initiating the study, continual support, and helpful discussions in the area of advanced mathematics.
- Dr Wayne Severn and Miss Pamela Kawakami of AgResearch, for their help with NMR sample preparation.
- Professor John Carver of the Wollongong University, Australia, for providing the Caerin 4.1 NMR data.
- Professor John Bowie and Mr Brian Chia of the University of Adelaide, Australia, for providing another set of the Caerin 4.1 NMR data.
- Professor David Parry of Massey University, for enthusiastic discussions and efforts in finding suitable proteins.
- Professor Peter Steinert of the National Institute of Health, USA, for providing the 54-residue human trichohyalin peptide.
- Dr Mark Smith of the Dairy Research Institute, for valuable discussions concerning protein NMR and for providing the glycomacropeptide of κ-casein.
- Lilian Morrison, for proof-reading.
- The staff and post-graduate students at AgResearch Wallaceville and Massey University, for their marvellous friendship and help with various issues.
- Celia and Lucia, his daughters, for their amazing understanding and marvellous forgiveness. Hai Yan Tan, his wife, in particular, for her infinite love, boundless support, and unrepayable sacrifice. Kee Fong How, his mum, for her eternal love and greatest support even during times of serious sickness and injury.

Table of Contents

List of Tables

Chapter 1:

Introduction

1.1 Project Justification

1.1.1 NMR spectroscopy and proteins

The applications of multi-dimensional NMR spectroscopy have developed rapidly since 2D NMR spectroscopy was proposed [Jeener 1971] and realised [Ernst 1974, 1975, 1976] in the 1970's. With its fascinating ability and potential to investigate structural information of molecules, NMR spectroscopy has become one of the most rapidly developed and widely used tools in molecular research. This is marked by an upsurge of multi-dimensional NMR experimental methods and applications, resulting in thousands of papers being published over the last 25 years.

Among the many researchers from different disciplines, a number of pioneers have made significant contributions to NMR spectroscopy of proteins and macro-molecules with many influential publications [e.g., Wüthrich 1976 & 1986, Jardetzky & Roberts 1981, Bax 1982, Redfield 1988, Roberts *et al.* 1993]. It has become clear that NMR spectroscopy is a powerful tool with which to investigate the 3D molecular structure of proteins in solution. In some cases it is complimentary to X-ray crystallography. Research on molecular interactions such as protein-chemical, protein-protein, protein-DNA and protein-RNA interactions have also been conducted using NMR spectroscopy. These techniques have enabled the investigation of some important properties of bio-molecules at the 3D molecular structure level, such as the functions of bio-molecules, disease mechanism, and drug design.

1.1.2 NMR software

Because of the complexity of NMR data processing and spectral analysis, a number of software packages have been developed to support the rapid progress of NMR research. The software development has been further stimulated by the rapid development of affordable high-capacity and high-performance computers over the last 20 years.

Comprehensive NMR data-processing programs have been available since the 1980's. A few of them have been developed as part of the spectrometer consoles by NMR spectrometer manufacturers, such as *VARIAN*, *BRUKER* and *JEOL*. Others are spectrometer-independent. Some of them are better known in NMR circles, such as the *FTNMR 5.1* [Hare Research 1988], *GIFA V.1-3* [Delsuc 1989], *FELIX 1.0* [Hare Research 1990], *FELIX 2.0* [Hare Research 1991], *PROSA* [Güntert 1992], *FELIX 2.1* [Biosym 1993], *FELIX 2.3* [Biosym 1994], *NMRView* [Johnson 1994], *NMRPipe* [Delaglio 1995], and *PROSPA* [Eccles 1995].

The programs *ANSIG* [Kraulis 1989], *EASY* [Eccles 1991], *CLAIRE* [Keywegt 1991], and *Pronto (2D)* [KjÆR 1991] were the first few comprehensive software packages written specifically for 2D NMR spectral analysis. These were followed by several other packages of the same kind, such as *EASY3D* [C. Bartels and K. Wüthrich, unpublished results], *XEASY* [Xia and Christian Bartels 1994], *Pronto (4D)* [KjÆR 1994], *FELIX* 95 [Biosym/MSI 1995], and *GIFA V.4* [Pons 1996].

1.1.3 Project outline

This PhD project was planned in late 1993 and started in April 1994. At that time, there had been some demand for an NMR spectral assignment package to run under the X-Window system. The situation became critical when SUN stopped supporting its SUNVIEW operating system (for which EASY had been written) in favour of the X-Window based Solaris-2.

As a comprehensive and yet strictly time-constrained training process, this project was planned to include four aspects as follows:

- the study of protein NMR theory,
- practice in protein NMR sample preparation and experiments, as well as NMR data processing and spectral analysis using other commercially available packages,
- development of an X-Window-based package that allows useful features for protein NMR study to be easily implemented in the future, and
- NMR spectral assignment of a small protein using the program mentioned above as a process to design, improve and test the various features in the package.

3D NMR spectroscopy was not included in the project, because:

- It was believed that concentrating on 2D NMR data processing and spectral assignment would be sufficient for a PhD project.
- Since our spectrometer had no gradient capability, obtaining high quality 3D NMR spectra was difficult.
- We could not envisage the possibility of obtaining a labelled protein-NMR sample that was suitable for 3D NMR studies.

1.1.4 Software development process for the project

As the project also embraced studies in quantum mechanics and NMR theory, as well as the exercises in NMR sample preparation and data acquisition, serious programming did not start until August 1995 when an SGI Indy was bought for this purpose. The processing and display modules were completed in early 1997. The final development stage was completed in conjunction with the assignment of a small protein called the *Caerin 4.1*[1].

There have been a number of difficulties that had to be faced during the project, such as:

- the complexity of the X-Window programming environment,
- the delays in obtaining a suitable protein for the project, and

[1] The Caerin 4.1 NMR data set was kindly provided by Prof. John Carver at the Wollongong University of Australia in late February 1997. Another data set collected by Mr Brian Chia at the University of Adelaide in September 1998 was also kindly given.

- the additional requirements of writing a comprehensive user guide for the software package, which has become Volume II of this thesis.

The decision to add data processing ability to the package has also resulted in tremendous effort and significant time being spent in the development process. Nevertheless, the first version of the package was finally completed in mid 1997 with the name PAW — the X Window-based Protein Analysis Workshop for NMR Spectroscopy [Lie 1997]. After that, it has been continually modified to implement more useful features and improve run-time performance.

1.1.5 An overview of PAW

PAW is a Motif[2] application based on the standard C programming language. It is an event-driven and multi-window system equipped with a compact macro language and three-letter keyboard-commands.

The package provides utilities for NMR data processing and spectral analysis. Many useful features have been implemented to ease NMR data processing, data display, peak picking, spin-system identification, spectral assignment, and peak display. Special consideration has been given to the spin-system and sequential assignments of two-dimensional ^{1}H NMR spectra of proteins, which are usually very difficult and cumbersome.

1.2 Thesis Organisation

The thesis is organised in two volumes, as described below.

1.2.1 Volume One

This is the thesis, proper. It has been organised so that Chapter 2, 3 and 4 contain the theory, Chapter 5 to 7, the methods, Chapter 8 to 11, the results, and Chapter 12, the conclusion, followed by appendices, references and index. In brief, the chapter contents are as follows:

- Chapter 2 introduces basic NMR theory.
- Chapter 3 introduces product operator formalism and pulse analysis.
- Chapter 4 introduces the NMR experiments.
- Chapter 5 introduces the NMR software and PAW.
- Chapter 6 introduces the NMR data processing with PAW.
- Chapter 7 describes the NMR spectral assignments with PAW.
- Chapter 8 describes the NMR data processing process and results for Caerin 4.1.

[2] Motif is a portable GUI toolkit developed by MIT in the United States. See also Chapter 5.

- Chapter 9 describes the peak-picking process and results obtained for the Caerin 4.1 NMR spectra.
- Chapter 10 describes the spin-system identification process and results for the Caerin 4.1 NMR spectra.
- Chapter 11 describes the NMR spectral assignment process and results for Caerin 4.1.
- Chapter 12 summarises the project and discusses the future development plan.
- The appendix includes the background material required to understand various parts of the thesis.

1.2.2 Volume Two

The second volume is the spectral processing and assignment guide for PAW.

- Chapter 1 contains abbreviations, notation and conventions in this chapter, followed by an installation guide.
- Chapter 2 describes the user interface, including macro programming.
- Chapter 3 describes how to set up a working environment.
- Chapter 4 describes macros and NMR data files.
- Chapter 5 describes how to handle 1D NMR data.
- Chapter 6 describes 1D-processing methods.
- Chapter 7 describes how to handle 2D NMR data.
- Chapter 8 describes the 2D-processing process as applied to the Caerin 4.1 NMR data.
- Chapter 9 describes the peak-picking process as applied to the Caerin 4.1 NMR spectra.
- Chapter 10 describes the spin-system identification process as applied to the Caerin 4.1 NMR spectra.
- Chapter 11 describes the peak-assignment process as applied to the Caerin 4.1 NMR spectra.
- Chapter 12 describes macro variables and commands with syntaxes and examples, as well as keyboard commands.

1.3 Abbreviation and Notation

1.3.1 Abbreviations

- nD n-dimensional, where n is a number, e.g., 1D, 2D, 3D
- COSY 2D correlated spectroscopy
- Dn Dimension-n, e.g., D1, D2, D3
- DQF-COSY Double-quantum filtered COSY
- f-domain frequency domain
- FFT fast Fourier Transform
- FID (or fid) free induction decay
- HN amide proton
- HA α-proton
- HB β-proton
- HG γ-proton
- HD δ-proton
- HE ε-proton
- HZ ζ-proton
- HH η-proton
- HFT Hilbert Fourier Transform
- IFT Inverse Fourier Transform
- MsBtn#1 the left *mouse button*
- MsBtn#2 the middle mouse button
- MsBtn#3 the right mouse button
- NMR Nuclear magnetic resonance
- NOE nuclear Overhauser effect
- NOESY 2D NOE spectroscopy
- PAW the Protein Analysis Workshop, this package
- ppm part(s) per million
- S/N ratio signal-to-noise ratio
- SD standard deviation
- t-domain time domain

- TOCSY 2D total correlation spectroscopy

1.3.2 Notation

- A_{kl} relative intensity corresponding to transition $|\psi_k(t)\rangle \rightarrow |\psi_l(t)\rangle$
- $\mathbf{B}_0$ static magnetic field, with a magnitude B_0
- $\mathbf{B}_1$ rf-field, with a magnitude B_1
- E_1, E_2 energy levels, with energy difference being ΔE
- k Boltzmann's constant
- **I, S** nuclear spin operator of spin **I** and spin **S**, respectively
- $\mathbf{I}_x, \mathbf{I}_y, \mathbf{I}_z$ Lie group generators of the nuclear spin **I**; likewise $\mathbf{S}_x, \mathbf{S}_y, \mathbf{S}_z$
- $\mathbf{I}_+, \mathbf{I}_-$ raising and lowering operators of spin **I**; likewise $\mathbf{S}_+, \mathbf{S}_-$
- $\mathbf{I}_{kx}, \mathbf{I}_{ky}, \mathbf{I}_{kz}$ Lie group generators of the k^{th} nuclear spin **I**
- J_{IS} coupling constant between spin **I** and spin **S**
- **J** nuclear angular momentum vector of a free spin
- $\mathbf{J}_i$ nuclear angular momentum vector of the i^{th} spin
- J_{iz} z-component of $\mathbf{J}_i$
- $\mathbf{M}_0$ ensemble nuclear magnetic moment vector, with a magnitude M_0
- $\mathbf{M}(t)$ time-dependent net magnetisation vector
- $M_x(t), M_y(t), M_z(t)$ x-, y- and z-components of $\mathbf{M}(t)$
- $\mathbf{M}_{xy}(t)$ time-dependent net magnetisation vector on the transverse plane
- $M_{xy}(t)$ the magnitude of $\mathbf{M}_{xy}(t)$
- P_{kl} transition probability corresponding to $|\psi_k(t)\rangle \dashrightarrow |\psi_l(t)\rangle$
- I nuclear spin quantum number for a spin
- M (total) magnetic quantum number for a spin system
- t_1 evolution period in a 2D NMR experiment
- t_2 acquisition period in a 2D NMR experiment
- t1,t2,t3,t4 The time-domains corresponding to D1, D2, D3 and D4
- T temperature in Kelvin
- T_1 longitudinal relaxation time

- T_2 transverse relaxation time
- $|i\rangle$, $|j\rangle$ basis kets representing eigen-states of a single-spin system
- $|r\rangle$, $|s\rangle$ basis kets representing eigen-states of a multi-spin system
- $|½\rangle$, $|-½\rangle$ basis kets of a spin-½ system
- $|\alpha\rangle$, $|\beta\rangle$ alternative basis kets of a spin-½ system
- $|1\rangle$, $|2\rangle$, ... basis kets of a spin system, either single or multiple
- $|\psi(t)\rangle$ eigen-ket of total Hamiltonian; corresponding bra is $\langle\psi(t)|$.
- $|\psi_k(t)\rangle$ the k^{th} eigen-ket of total Hamiltonian; corresponding bra is $\langle\psi_k(t)|$.
- $(\pi/2)_x$ 90° x-pulse; likewise $(\pi)_x$, $(\pi/2)_y$, $(\pi)_y$, etc.
- $\langle\mathbf{A}\rangle$ expectation value of observable **A**
- $\langle\mathbf{A}\rangle_k$ the k^{th} expectation value of observable **A**
- γ gyro-magnetic ratio
- $\Delta\omega$ offset of ω_{rotate} from the Larmor frequency ω_0
- δ chemical shift
- $\hbar$ Dirac's symbol (= $h/2\pi$, where h is the *Planck's constant*)
- $\mathcal{H}$ total Hamiltonian operator, with components $\mathcal{H}_{rs}$
- $(\mathcal{H}_{rs})^*$ complex conjugate of $\mathcal{H}_{rs}$
- $\mathcal{H}^\dagger$ adjoint of $\mathcal{H}$, with components $(\mathcal{H}^\dagger)_{rs}$
- $\mathcal{H}^{[Z]}$ Zeeman Hamiltonian
- $\mathcal{H}^{[C]}$ chemical shift Hamiltonian
- $\mathcal{H}^{[J]}$ J-coupling Hamiltonian
- $\mathcal{H}_1$ perturbation Hamiltonian
- $\mathcal{L}_\xi$ 3D operator subspace (ξ_α, ξ_β, ξ_γ) of Liouville operator space, which is isomorphic to Lie group generator space ($\mathbf{I}_x$, $\mathbf{I}_y$, $\mathbf{I}_z$)
- μ nuclear magnetic momentum, with a magnitude μ
- $\sigma(t)$ density operator
- $\sigma_0(t)$ equilibrium density operator
- $\sigma_k(t)$ the k^{th} state density operator (= $|\psi_k(t)\rangle\langle\psi_k(t)|$)
- τ_m mixing time

- ν_0 Larmor frequency in Hz
- ω_0 Larmor frequency vector
- ω_0 Larmor frequency in rad/sec
- ω_{rotate} angular frequency of a rotating frame
- Ω angular frequency offset of a Lamor frequency in a rotating frame
- Ω_I Ω of spin **I**
- Ω_S Ω of spin **S**

1.4 Conventions

1.4.1 Conventions for formatted text

- Text like *this* signals a special term.
- Text like this is a command, a keyboard entry or a header in a dialog.
- `Text like this is a macro- or program-statement.`
- Text in a macro or program headed with a hash like `#this` starts a note.
- Text enclosed by square brackets such as [Ernst *et al.* 1987] is a reference.
- Highlighted text enclosed by a pair of square brackets such as [Load] is a button in a menu, toolbox or a dialog.
- Two numbers separated by a comma enclosed by a pair of square brackets like [2000, 2 l00] defines a 1D region.
- Two words separated by a slash (/) within square brackets like [yes/no] is an exclusive-OR option for an entry.
- A highlighted character-string like Ctrl represents a key on the keyboard.
- A compound character-string like DataDir is a variable name used in PAW.
- A boldface character-string containing a dot such as **PAW.Init** is a filename.
- A boldface character-string ended with a slash (/) such as **~/PAW/** is a directory name, where ~/ is the user root directory.

home

1.4.2 Conventions for mouse operations

- To point or to position means to move the mouse cursor to a specific screen position.
- To click means to position the mouse cursor, then press and release a mouse button without moving the mouse.
- To double-click means to click a mouse button twice in rapid succession without moving the mouse.

- To click-and-drag means to press a mouse button, hold it down while the mouse is moving, and then release the button at a desirable position.
- To choose means to click on a command button with MsBtn#1, e.g., 'choose [Convert]'.
- To pick means to drag the mouse cursor with MsBtn#2, then release the button after pointing to a desirable position, e.g., 'pick a peak'.
- To select an item means to click an item in a name list with MsBtn#1, e.g., 'select **data7noesy150.2Ddata**.
- To block or select a region means to define a rectangle region in a plot by dragging the mouse pointer with the MsBtn#1 after clicking on a position in a window then release the button at a desirable position. During the process, an expanding (or contracting) rubber rectangle is drawn while the mouse is moving. Note that the edges of a rubber rectangle must not be less than five pixels; otherwise, the operation is cancelled.
- To load a file means to select a filename in the file list of a file-selection dialog, and then choose [Load] in the dialog, e.g., 'load data7noesy150.pdt'. If the filename is not in the current list, you may need to select a sub-directory in the directory list then choose [Filter] in the dialog to change the directory. (See Chapter 4, Volume II.)
- To turn on (or off) a button means clicking on the button with MsBtn#1 so that it looks as if it has been (or not been) pressed down after the mouse button is released. For example, 'turn off [DspBufs] in the *1D-display Toolbox*'.

In addition, if no button number is specified for a mouse operation, the task is by default performed by MsBtn#1.

1.4.3 Conventions for keyboard entries

- The key names used in this booklet match the standard keyboard labels, e.g., F1, Enter, Ctrl, Alt, #, Space.
- A comma between two key-names indicates that two keys are to be pressed sequentially, e.g., F3,F7.
- A plus sign between two key-names indicates that the first key must be pressed and held down while pressing the second key, e.g., Ctrl+F1.
- To *enter* a command means to type the command then press the Enter key while the mouse pointer (or cursor) is in a draw-window. (See also Cahpter 2, Volume II.)

Chapter 2:

Basic Theory of Nuclear Magnetic Resonance Spectroscopy

2.1 Introduction

This chapter describes the basic theory of NMR spectroscopy. Section 2.2 and 2.3 introduce nuclear magnetism and NMR phenomena. Section 2.4 to 2.8 describe basic quantum mechanics theory for NMR, followed by a typical example for indirect coupling of a two-spin system.

2.2 Nuclear Magnetism

2.2.1 Nuclear magnetic moment and Larmor frequency

Most nuclear species possess a *nuclear magnetic dipole* with an associated *magnetic moment* ($\boldsymbol{\mu}$) and an *angular momentum* ($\mathbf{J}$):

$$\boldsymbol{\mu} = \gamma \mathbf{J} \tag{2.2.1.1}$$

γ, the *gyro-magnetic ratio* or *magneto-gyric ratio*, varies with different nuclear species. In the presence of an external *static magnetic field* ($\mathbf{B}_0$), a nuclear magnetic dipole precesses about an axis parallel to $\mathbf{B}_0$ along a conical surface with an angular frequency (ω_0 rad/sec), known as the *Larmor frequency*, mimicking the precession of a spinning top in the earth's gravitational field. The corresponding *angular frequency vector* ($\boldsymbol{\omega}_0$) is linearly dependent on and anti-parallel to $\mathbf{B}_0$:

$$\boldsymbol{\omega}_0 = -\gamma \mathbf{B}_0 \tag{2.2.1.2}$$

The Larmor frequency ω_0 is more often represented by ν_0 in Hertz such that

$$\nu_0 = \omega_0/2\pi = \gamma B_0/2\pi \tag{2.2.1.3}$$

A list of physical parameters for different nuclear species can be found in standard references such as [Weast *et al.* 1964-] and [Fuller 1976]. For example, the nuclear magnetic moment of the proton (^{1}H) in water is observed to precess at ν_0 = 42.5759 MHz if the intensity of B_0 is one Tesla. Accordingly, $\gamma_{proton} = 2\pi\nu_0/B_0 = 26.75\times10^7$ rad·sec^{-1}·Tesla^{-1}, which is a figure quoted in many textbooks.

2.2.2 Nuclear magnetic equilibrium and quantisation

An NMR sample usually contains a very large number of nuclei in the molecules to be studied. A simple calculation with *Avogadro's number* shows that, if each of the molecules has 1000 protons, then, every millilitre of a 1 mM solution contains approximately 6.02×10^{20} protons of the molecules.

In the presence of an external field $\mathbf{B}_0$ with no other perturbing field, these nuclei will reach a *thermal equilibrium* state. According to quantum physics, at equilibrium, the nuclear magnetic moments in the field are '*quantised*' into a number of *allowable energy states* or *eigen-states* — a fact that cannot be explained by classical physics. As a consequence of the quantisation, an *ensemble nuclear magnetic moment* is formed, which is often represented by a vector $\mathbf{M}_0$ that points in the same direction as $\mathbf{B}_0$.

2.2.3 Allowable energy states

Depending on the nuclear species, the number of allowable energy states ranges from 1 to 17[1]. Each of the energy states of an isolated nucleus corresponds to an *allowable angular momentum* $\mathbf{J}_i$ that is independent of any electronic orbital motion[2], and hence is *intrinsic*. The subscript i indicates that $\mathbf{J}_i$ is associated with the i^{th} energy state.

In NMR theory, the orientation of $\mathbf{B}_0$ defines the z-axis of the Cartesian coordinate system in the laboratory frame of reference. The z-component of the i^{th} allowable angular momentum (J_{iz}) is an *eigen-value* of the allowable energy states. It is quantised in units of $\hbar$ ($= h/2\pi$, where h is the *Planck's constant*) such that

$$J_{iz} = m_i \hbar \tag{2.2.3.1}$$

where

$$m_i = I, I-1, I-2, \ldots -I \tag{2.2.3.2}$$

Here, I, the *nuclear spin quantum number* or *azimuthal quantum number*, is a half-integer number that characterises the nuclear spin:

$$I = 1/2, 1, 3/2, \ldots \tag{2.2.3.3}$$

For the above reason, $\hbar$ is also known as the *angular momentum unit* for quantum mechanics.

In terms of nuclear magnetisation, nuclear species are often referred to by their nuclear spin quantum number such as spin-½ nuclei, spin-1 nuclei, and so on. Spin-½ nuclei such as ^{1}H, ^{13}C and ^{15}N can only have two energy states, whereas spin-1 nuclei such as ^{2}H, ^{32}P, ^{14}N have three, and so on. For the spin-½ nuclei, the two energy states correspond to $m_i \in \{½, -½\}$; whereas for the spin-1 nuclei, $m_i \in \{1, 0, -1\}$. In this thesis, only spin-½ nuclei in liquid samples will be discussed because of the nature of the project.

The two eigen-states of a spin-½ nucleus are called the spin-up state and spin-down state. They are symbolised in quantum mechanics by two orthonormal *basis kets* $|1\rangle$ and $|2\rangle$ (or $|\alpha\rangle$ and $|\beta\rangle$)[3]. Here, $|1\rangle$ (or $|\alpha\rangle$) represents the lower energy state (or the ground state) that corresponds to $m_i = ½$, and $|2\rangle$ (or $|\beta\rangle$), for $m_i = -½$. In column vector notation, the orthonormal basis of the state space is simply comprised of two vectors:

$$|1\rangle = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \text{ and } |2\rangle = \begin{bmatrix} 0 \\ 1 \end{bmatrix}. \tag{2.2.3.4}$$

In this thesis, the general symbol $|i\rangle$ represents any of the basis kets of a single-spin system.

---

[1] The highest spin number observed so far is 8 for ^{134}Cs at a meta-stable excited state. [See CRC Handbook of Chemistry and Physics, published yearly.]

[2] Here it is referring to an isolated nucleus without being coupled to any other nuclei. The energy levels of coupled nuclei will be discussed later.

[3] Also seen as $|\uparrow\rangle$ and $|\downarrow\rangle$, or $|½\rangle$ and $|-½\rangle$, in different articles.

2.2.4 The Boltzmann distribution law

At equilibrium, the populations (N_i) of nuclei at different energy states obey a well-known relationship called the *Boltzmann distribution law*:

$$N_1 : N_2 : ... = \exp(-E_1 / kT) : \exp(-E_2 / kT) : ... \qquad (2.2.4.1)$$

In the expression, k is *Boltzmann's constant*. T is the temperature in Kelvin. E_i, where i=1,2,..., are the *energy levels* associated with the i^{th} energy state such that

$$E_i = -\hat{\mu}_i \bullet \mathbf{B}_0 = -\hat{\gamma}\, \mathbf{J}_i \bullet \mathbf{B}_0 = +\mathbf{J}_i \bullet \boldsymbol{\omega}_0$$
$$= -m_i \omega_0 \hbar = -m_i \nu_0 h \qquad (2.2.4.2)$$

For example, the two allowable energy levels for spin-½ nuclei are

$$E_1 = -\tfrac{1}{2} \omega_0 \hbar = -\tfrac{1}{2} \nu_0 h \qquad (2.2.4.3)$$

$$E_2 = +\tfrac{1}{2} \omega_0 \hbar = +\tfrac{1}{2} \nu_0 h \qquad (2.2.4.4)$$

as illustrated in Figure 2.1.

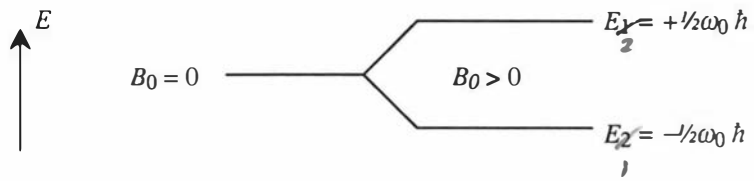


Figure 2.1: The energy level splitting for a spin-1/2 nucleus

The split in energy levels for a nuclear species in the presence of an external magnetic field is called the *Zeeman effect* in quantum mechanics.

The energy difference ΔE between the two energy levels expressed by the last two equations is

$$\Delta E = \omega_0 \hbar = \nu_0 h \qquad (2.2.4.5)$$

Note that ΔE is associated with an energy transition between the two energy levels. If the energy is measured in the units of $\hbar$, then the value of ΔE is identical to the observable resonance frequency. It will be shown later that the intensity of a peak observed in a frequency-domain spectrum is related to the transition probability between energy levels.

2.2.5 Factors related to M_0

Based on the Boltzmann distribution law, it can be shown that at equilibrium the macroscopic intensity of the *ensemble nuclear magnetic moment* (M_0) for spin-½ nuclei is related to three factors, namely, B_0, N and T [see e.g., Abragam 1961]:

$$M_0 = B_0 N \gamma^2 \hbar^2 / 4kT \qquad (2.2.5.1)$$

Here N is the total number of resonant nuclei in the sample and B_0 is the magnitude of $\mathbf{B}_0$. For example, suppose that a 5 mM liquid sample is used in an NMR experiment with a 500 MHz NMR spectrometer. Then, a simple calculation

shows that the magnitude of M_0 at room temperature is only about 2.35×10^{-16} Am^{-1}. The use of bulk samples in NMR experiments is therefore necessary.

For NMR spectroscopy experiments, (2.2.5.1) implies that one can either increase B_0 or N, or decrease T in order to increase the magnitude of M_0. However, for protein NMR experiments, increasing N or decreasing T too much may reduce the tumbling rate of the molecules, thereby broadening the peaks in a spectrum, or even change the physical state of the sample, which may be undesirable.

2.3 Nuclear Magnetic Resonance and Free Induction Decay

2.3.1 Nuclear magnetic resonance and the rf-pulse

Energy-level transitions[4] of all nuclear-spins with a Larmor frequency ω_0 occur when a linearly *oscillating magnetic field* ($\mathbf{B}_{osc}$) at the same frequency ω_0 is applied perpendicularly to $\mathbf{M}_0$. This process is called *nuclear magnetic resonance* (NMR). For this reason, the Larmor frequency is also called the *nuclear magnetic resonance frequency*.

$\mathbf{B}_{osc}$ is often called *the rf-field* because its frequency is usually within the range of radio frequencies. In practice, $\mathbf{B}_{osc}$ is always applied perpendicularly to the field $\mathbf{B}_0$ in order to manipulate nuclear magnetisation in NMR experiments. With $\mathbf{B}_0$ parallel to the z-axis of a Cartesian coordinate system, the transverse plane is then the xy-plane. In this thesis, $\mathbf{B}_{osc}$ is always placed on the negative y-axis of the Cartesian coordinate system. Mathematically, $\mathbf{B}_{osc}$ can be expressed as follows:

$$\mathbf{B}_{osc} = 2B_1 \cos(\omega_0 t)\, \mathbf{e}_y \qquad (2.3.1.1)$$

Here, $2B_1$ is the amplitude of $\mathbf{B}_{osc}$, and $\mathbf{e}_y$ is the unit vector along the y-axis. To see why $\mathbf{B}_{osc}$ induces nuclear magnetic resonance, it is usually decomposed into two counter-rotating components:

$$\mathbf{B}_{osc} = \mathbf{B}_1 + \mathbf{B}_2 \qquad (2.3.1.2)$$

where

$$\mathbf{B}_1 = B_1 (-\sin(\omega_0 t)\, \mathbf{e}_x + \cos(\omega_0 t)\, \mathbf{e}_y) \qquad (2.3.1.3)$$

$$\mathbf{B}_2 = B_1 (+\sin(\omega_0 t)\, \mathbf{e}_x + \cos(\omega_0 t)\, \mathbf{e}_y) \qquad (2.3.1.4)$$

These represent two magnetic fields rotating clockwise and anti-clockwise about the z-axis at ω_0. The component $\mathbf{B}_1$ rotates at the frequency $-\omega_0$, which is the same as the Larmor frequency of the sample nuclei. The component $\mathbf{B}_2$ rotates at twice the Lamor frequency and hence can be ignored because it does not meet the resonance condition.

Therefore, only $\mathbf{B}_1$, which appears to be static relative to every $\boldsymbol{\mu}_i$ during the excitation process, has a significant effect on the nuclear spins. Just as the static field $\mathbf{B}_0$ forces

[4] Note the energy level transitions here refer to continual changes of individual spin energy levels, which is different from the so-called flip-flop energy exchange between different spins. The speed of the energy level transitions here depends on the magnitude of $\mathbf{B}_{osc}$, as explained next.

all nuclear magnetic dipoles to precess about it, the existence of $\mathbf{B}_1$ at ω_0 forces every μ_i to precess about $\mathbf{B}_1$. Consequently, the net magnetic moment $\mathbf{M}_0$ is forced to rotate about the negative y-axis with an angular velocity $\omega_1 = -\gamma B_1$. Therefore, if t_p is the duration of the rf-field, then $\mathbf{M}_0$ will be tipped from its equilibrium orientation through an angle β such that

$$\beta = -\gamma B_1 t_p \qquad (2.3.1.5)$$

As a result, the thermal equilibrium of the system is destroyed and $\mathbf{M}_0$ starts to precess about the z-axis while moving gradually towards the xy-plane, as shown in Figure 2.2.

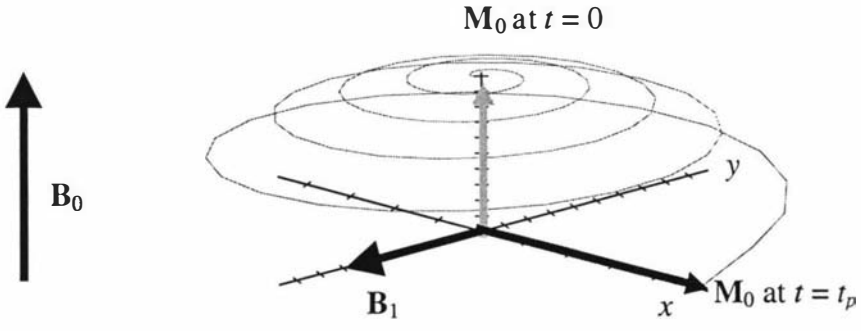


Figure 2.2: The precession of $\mathbf{M}_0$ about the z-axis under the effect of $\mathbf{B}_1$. The spiral curve shows how the end of $\mathbf{M}_0$ moves gradually towards the xy-plane.

In order to maximise the magnetisation signals, most NMR experiments aim to tip $\mathbf{M}_0$ into the transverse plane before data collection. Thus, $\mathbf{M}_0$ will end up lying on the x-axis, perpendicular to the $\mathbf{B}_1$. The time required to tip $\mathbf{M}_0$ by 90° is usually of the order of microseconds; and hence the rf-field is often called the *rf-pulse* instead. An rf-pulse is also referred to by the tipping angle it produces, e.g., a 90°-pulse, a 180°-pulse, etc. The process of manipulating nuclear spins with rf-pulses is called *nuclear spin excitation.*

2.3.2 Nuclear magnetic relaxation

As soon as the rf-field is switched off, each of the nuclear magnetic moments start to realign again to one of their allowable energy states. This process is called *nuclear magnetic relaxation*, which plays an important role in the observation of the resonance phenomena.

Note that the relaxation is a dynamic process in which the orientation of each nuclear moment μ_i can be changed significantly, due to the rapid movement of and collision between molecules, as well as the changes in local magnetic field. As a consequence, a nuclear spin may or may not resume its original energy level at the end of the

relaxation. Nevertheless, the ensemble magnetisation vector $\mathbf{M}_0$ will resume its initial magnitude and orientation along the z-axis.

2.3.3 The Bloch equations in the laboratory frame

During the relaxation process, the evolution of the time-dependent *net magnetisation vector* $\mathbf{M}(t)$ can be described empirically by a set of ordinary differential equations known as the *Bloch equations* [Bloch 1946]. In a frame of reference rotating at an angular frequency ω_{rotate} about the z-axis, these are usually presented as follows:

$$\frac{d}{dt}\begin{bmatrix} M_x \\ M_y \\ M_z \end{bmatrix} = \begin{bmatrix} -1/T_2 & -\Delta_\omega & 0 \\ \Delta_\omega & -1/T_2 & 0 \\ 0 & 0 & -1/T_1 \end{bmatrix}\begin{bmatrix} M_x \\ M_y \\ M_z \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ M_0/T_1 \end{bmatrix} \tag{2.3.3.1}$$

In the expression, M_x, M_y and M_z are three scalar components of $\mathbf{M}(t)$. M_0 is the amplitude of $\mathbf{M}$ at equilibrium. T_1, the *spin-lattice relaxation time*, is the characteristic time required for $\mathbf{M}$ to relax back to its original equilibrium orientation and intensity. T_2, the *spin-spin relaxation time*, is the characteristic time required for the transverse magnetisation vector to vanish from the transverse plane. Δ_ω is the offset of the Larmor frequency ω_0 from ω_{rotate} such that

$$\Delta_\omega = \omega_0 - \omega_{rotate} \tag{2.3.3.2}$$

In this thesis, $\mathbf{M}(t)$ is assumed to rotate clockwise in the rotating frame if $\Delta_\omega > 0$. This is mainly for the consistency with the discussion on Fourier transformation. Correspondingly, the following format of Bloch equations will be considered:

$$\frac{d}{dt}\begin{bmatrix} M_x \\ -M_y \\ M_z \end{bmatrix} = \begin{bmatrix} -1/T_2 & -\Delta_\omega & 0 \\ \Delta_\omega & -1/T_2 & 0 \\ 0 & 0 & -1/T_1 \end{bmatrix}\begin{bmatrix} M_x \\ -M_y \\ M_z \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ M_0/T_1 \end{bmatrix} \tag{2.3.3.3}$$

Note that T_2 is always less than or equal to T_1. Primarily, this is because of the differences in the precession rates[5] of the magnetic dipoles in the system. These differences in precession rates cause the dipoles to quickly fan out from the net transverse magnetisation vector $\mathbf{M}_{xy}$ after rotating for a number of cycles. Consequently, the magnitude of $\mathbf{M}_{xy}$ is reduced to zero quicker, because the opposite components of the dipoles that fan out from $\mathbf{M}_{xy}$ cancel each other, as shown in the next diagram.

[5] This is related to chemical shift, a concept that will be discussed shortly.

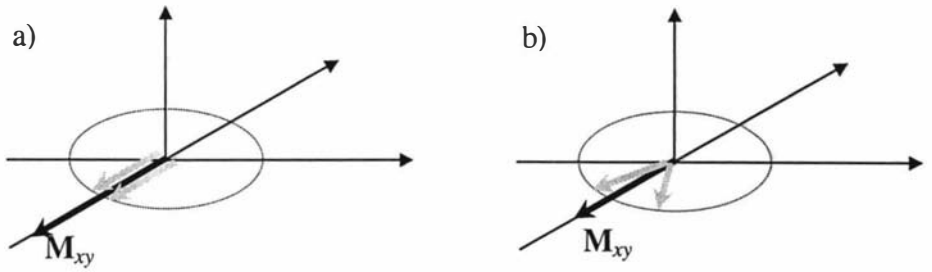


Figure 2.3: a) The magnitude of $\mathbf{M}_{xy}$ is maximised at t=0. b) The magnitude of $\mathbf{M}_{xy}$ is reduced because the opposite components of the dipoles that fan out from $\mathbf{M}_{xy}$ cancel each other.

Assuming that $\mathbf{M}(t)$ is tipped through an angle β by an rf-pulse, then the initial conditions are

$$M_x(0) = M_0 \sin(\beta) \tag{2.3.3.4}$$

$$M_y(0) = 0 \tag{2.3.3.5}$$

$$M_z(0) = M_0 \cos(\beta) \tag{2.3.3.6}$$

With these, the solutions to the Bloch equations can be found:

$$M_x(t) = M_0 \sin(\beta) \cos(-\Delta_\omega t) \exp(-t/T_2) \tag{2.3.3.7}$$

$$M_y(t) = M_0 \sin(\beta) \sin(-\Delta_\omega t) \exp(-t/T_2) \tag{2.3.3.8}$$

2.3.4 Transverse magnetisation vector and free induction decay

In the transverse plane of the rotating frame, the pair $\mathbf{M}_x$ and $\mathbf{M}_y$ form a *transverse magnetisation vector* $\mathbf{M}_{xy}$ that decays exponentially while rotating about $\mathbf{B}_0$ at an angular frequency Δ_ω. Their magnitudes $M_x(t)$ and $M_y(t)$ can be combined into a complex function[6] $M_{xy}(t)$:

$$M_{xy}(t) = M_0 \sin(\beta) \exp(-i\Delta_\omega t) \exp(-t/T_2) \tag{2.3.4.1}$$

This rotating magnetisation is detected by an NMR spectrometer via a coil that is wound around the sample. The induced voltage that is detected in quadrature (i.e., as x- and y- components) is referred to as the *free induction decay* (FID). In most NMR experiments, the receiver is switched on to collect signals almost immediately after a 90°-pulse. Hence, the collected complex FID is simply

$$s(t) = M_0 \exp(-i\Delta_\omega t) \exp(-t/T_2) \text{ in the rotating frame} \tag{2.3.4.2}$$

[6] Note: This is how a complex expression of FID is introduced into spectroscopy.

2.3.5 A n ideal FID and its Fourier transform

A plot of an ideal FID containing a single exponentially decaying frequency may look like this[7]:

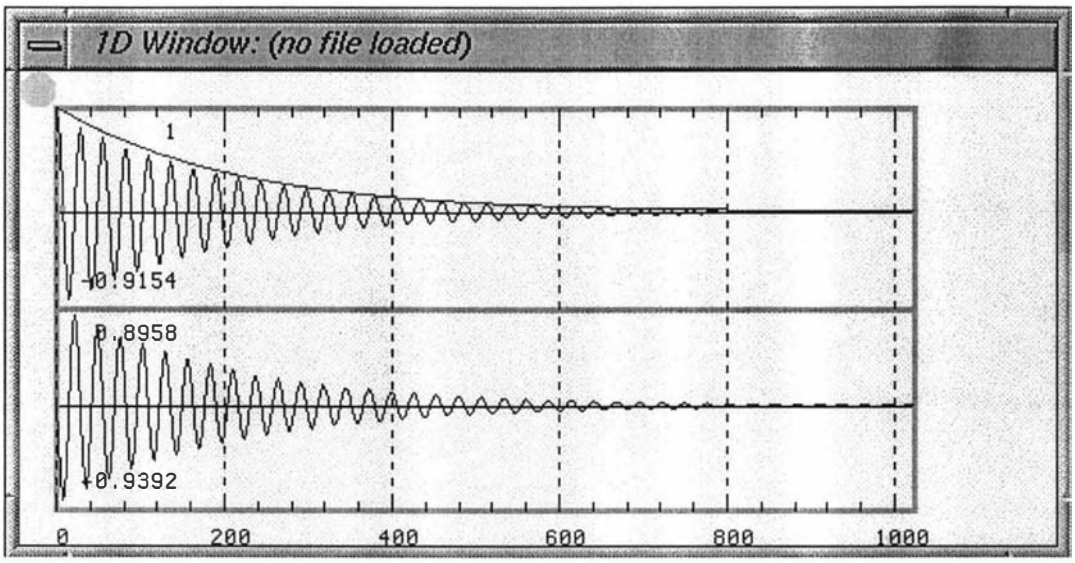


Figure 2.4: The real (top) and imaginary (bottom) parts of an ideal FID.

Here, the real and imaginary parts of the complex FID are plotted separately. The smooth curves in the diagram were drawn from the exponentially decaying function:

$$y(t) = \exp(-t/T_2). \qquad (2.3.5.1)$$

In practice, the observed decay-rate is normally greater than $1/T_2$.

The Fourier transform of an ideal FID gives a frequency-domain spectrum with the line-shapes in the real and imaginary parts being in *pure absorption* and *pure dispersion*, respectively, as illustrated in the next diagram.

[7] All diagrams of the spectra presented in this thesis are created using the program PAW.

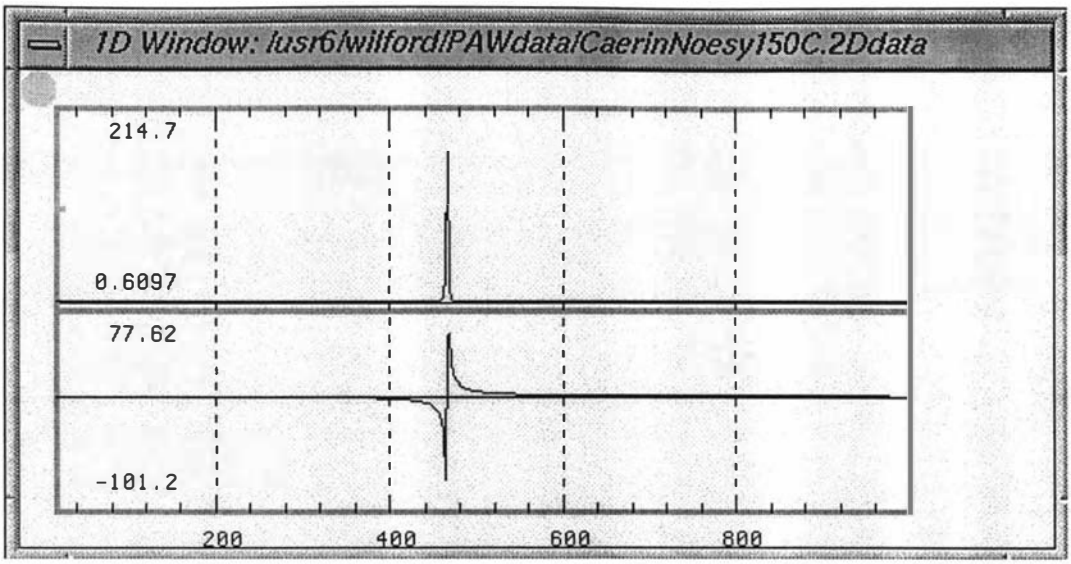


Figure 2.5: A pure absorption signal (top) and dispersion signal (bottom) obtained from the Fourier transform of an ideal FID.

2.3.6 Multi-frequency FID

An FID collected from an NMR sample of organic compounds or bio-molecules is a superposition of many frequencies, as shown in the next diagram, which may contain hundreds of proton resonance frequencies.

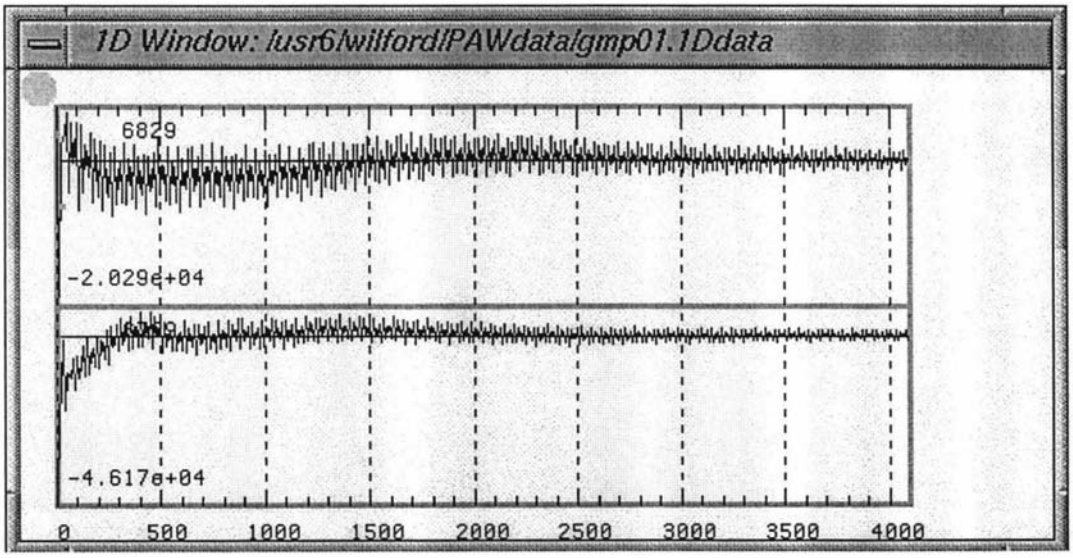


Figure 2.6: The real (top) and imaginary (bottom) parts of an FID collected from a protein. (A frequency-domain spectrum is shown in Figure 2.7.)

2.3.7 Chemical shifts

There are different frequencies observed in NMR spectra obtained from a sample containing many nuclei in each molecule. The differences are caused by perturbations in $\mathbf{B}_0$ at different locations in the molecular structure. At every location, the

perturbation is related to the local electron cloud in the chemical bond, which slightly changes the local magnetic fields seen by a nearby nucleus and hence induces an offset (υ_{offset}) in the resonance frequency:

$$\upsilon_{perturbed} = \upsilon_0 + \upsilon_{offset} \tag{2.3.7.1}$$

Here $\upsilon_{perturbed}$ and υ_0 are the perturbed and unperturbed resonance frequencies.

The offset resonance frequency depends linearly on the magnitude of $\mathbf{B}_0$, usually varying from a few hundred to a few thousand Hz for different spectrometers. A dimensionless quantity δ, has been used so that the offset frequency can be consistently represented for all spectrometers. This quantity, measured in parts per million (*ppm*), is called the *chemical shift* :

$$\delta = \frac{\upsilon_{perturbed} - \upsilon_0}{\upsilon_0} \times 10^6 \tag{2.3.7.2}$$

In protein NMR data, an observed frequency is usually expressed as an offset of the resonance frequency from the carrier frequency of a spectrometer. Water protons, for example, are often set experimentally to induce a peak at a frequency close to zero. When the data are processed, the chemical shift of the water protons is often set to around 4.7 ppm, providing a secondary reference frequency. Other protons in the protein have resonance frequencies that are distributed around the water proton's, ranging approximately from 0.6 ppm to 10.5 ppm [Wüthrich 1986, Wishart 1991]. An example showing this distribution may be found in the next diagram.

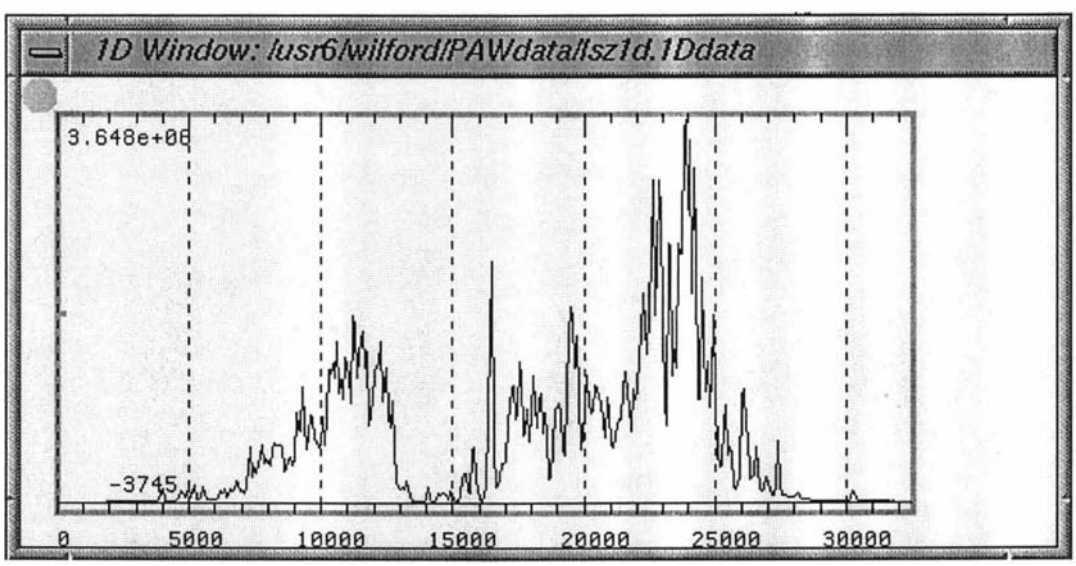


Figure 2.7: A 1D spectrum of hen-egg lysozyme.

The variations in chemical shifts allow different nuclear spins to be identified and characterised, as will be explained later.

2.3.8 Nuclear spin systems and spin-spin coupling

If a group of nuclear spins in a molecule are observed to couple with each other via through-bond connectivities, they are said to form a *spin system.* This kind of through-bond connection is called *spin-spin coupling* or *scalar coupling.* In some NMR experiments such as COSY, changes of energy levels (or *Zeeman effects*) in the individual nuclei of a spin system, due to spin-spin coupling, can be detected. These manifest themselves as a splitting of peaks in an NMR spectrum. The magnitude of the splitting is independent of $\mathbf{B}_0$. It is measured in Hz and is called the *spin-spin coupling constant.*

2.4 The Schrödinger Equation in NMR

2.4.1 The Schrödinger equation

To discuss the interaction between nuclear spins, the density operator [see, e.g., Abragam 1961 or Ernst *et al.* 1987] and product operator formalism [Sørensen *et al.* 1983] are often used. In NMR applications, the evolution of a spin-system's energy states is characterised by the *Schrödinger equation* (SE) [Cohen-Tannoudji *et al.* 1973] with a time-dependent *wave function* (or *state function*) $|\psi(t)\rangle$:

$$i\frac{d}{dt}|\psi(t)\rangle = \mathcal{H}(t)\,|\psi(t)\rangle \qquad (2.4.1.1)$$

Here, $\mathcal{H}(t)$, measured in angular momentum units $\hbar$, is the *total-energy operator* (or *spin Hamiltonian*) applied to a spin system.

The reason for introducing $\hbar$ as a unit for angular momentum is simply to get rid of $\hbar$ in many formulae. Using this unit, the angular momenta (J_i) and energy levels (E_i) are also measured in $\hbar$. Hence, the energy difference (ΔE) between two energy levels of a free spin becomes equal to its Larmor frequency (ω_0). In addition, the angular momentum vector **J** of a free spin becomes identical to its nuclear-spin vector **I**. They will be represented by matrices and termed the *spin operators* in density-operator formalism.

Equation (2.4.1.1) describes changes of energy states under the influence of a time-dependent total-energy operator $\mathcal{H}(t)$. The reason that the imaginary symbol i is involved in the equation may not be clear at this stage until the solution to the problem is seen. Simply speaking, it is due to the wave-like nature of quantum-mechanical systems.

2.4.2 Magnetic quantum levels and states

The Schrödinger equation is often used to study the behaviour of a *statistical mixture of states* for a multi-spin system. Being a first-order differential equation with respect to time, its solution is related to the eigen-value problem:

$$\mathcal{H}(t)\,|\psi_k(t)\rangle = E_k\,|\psi_k(t)\rangle \qquad k=1,2,...,n \qquad (2.4.2.1)$$

Here E_k, the k^{th} eigen-value of $\mathcal{H}$, is one of the *magnetic quantum energy levels* of the system. The set of all E_k is called the *spectrum* of $\mathcal{H}$. The resulting *eigen vectors* (or *eigen kets*), $|\psi_k(t)\rangle$, form a state space $\mathcal{E}$ of the system. This space is a closed inner-product space (or *Hilbert space*) that contains only orthonormal vectors. At a fixed time t_0, the state of the system is defined by specifying an eigen-ket $|\psi_k(t_0)\rangle$ belonging to $\mathcal{E}$— this is *the first postulate of quantum mechanics* [see, e.g., Cohen-Tannoudji *et al.* 1973]. In a vector format, an eigen-ket is represented by a linear combination of *basis state* (see the next subsection) that are symbolised by a set of *basis kets* $|r\rangle$, r=1,2,...,n:

$$|\psi_k(t)\rangle = \sum_r c_{k,r}(t)\,|r\rangle \tag{2.4.2.2}$$

Here $c_{k,r}(t)$ satisfy the normalisation and orthogonality conditions for the eigen-kets $|\psi_k(t)\rangle$:

$$\langle\psi_k(t)\,|\,\psi_k(t)\rangle = \sum_r c_{k,r}c_{k,r}^{*} = \sum_r |\,c_{k,r}\,|^2 = 1 \tag{2.4.2.3}$$

$$\psi_k(t) \bullet \psi_l(t) = \sum_r c_{k,r}c_{l,r} = 0, \qquad k \neq l \tag{2.4.2.4}$$

Here, $|\,c_{k,r}\,|^2$ is the probability of finding an individual spin system in the r^{th} basis state $|r\rangle$ while the *system ensemble*[8] is in the k^{th} magnetic quantum state $|\psi_k\,(t)\rangle$.

2.4.3 Basis state space for a spin system

A basis ket $|r\rangle$ for a multi-spin system of m nuclei is simply the direct product (or tensor product) of m basis kets of every nuclear spin in the system. Therefore, the dimension of the basis kets is a multiplication of the numbers of allowable energy states of every nucleus in the system. For example, for a two-spin-½ system, m_i = 2, the orthonormal basis ket space (or *basis state space*) is comprised of

$$|1\rangle = |\alpha\alpha\rangle = \begin{bmatrix}1\\0\end{bmatrix}\otimes\begin{bmatrix}1\\0\end{bmatrix} = \begin{bmatrix}1\\0\\0\\0\end{bmatrix}, \qquad |2\rangle = |\alpha\beta\rangle = \begin{bmatrix}1\\0\end{bmatrix}\otimes\begin{bmatrix}0\\1\end{bmatrix} = \begin{bmatrix}0\\1\\0\\0\end{bmatrix},$$

$$|3\rangle = |\beta\alpha\rangle = \begin{bmatrix}0\\1\end{bmatrix}\otimes\begin{bmatrix}1\\0\end{bmatrix} = \begin{bmatrix}0\\0\\1\\0\end{bmatrix}, \qquad |4\rangle = |\beta\beta\rangle = \begin{bmatrix}0\\1\end{bmatrix}\otimes\begin{bmatrix}0\\1\end{bmatrix} = \begin{bmatrix}0\\0\\0\\1\end{bmatrix}. \tag{2.4.3.1}$$

Here, the symbols α and β correspond to two allowable energy states of a spin-½ particle $|\alpha\rangle$ and $|\beta\rangle$:

$$|\alpha\rangle = |½\rangle, \qquad |\beta\rangle = |-½\rangle. \tag{2.4.3.2}$$

For a three-spin-½ system, there are eight basis kets:

[8] This refers to the entire body of all spin systems of the same kind in a sample.

$$|1\rangle = |\alpha\alpha\alpha\rangle, \quad |2\rangle = |\alpha\alpha\beta\rangle, \quad |3\rangle = |\alpha\beta\alpha\rangle, \quad |4\rangle = |\alpha\beta\beta\rangle,$$
$$|5\rangle = |\beta\alpha\alpha\rangle, \quad |6\rangle = |\beta\alpha\beta\rangle, \quad |7\rangle = |\beta\beta\alpha\rangle, \quad |8\rangle = |\beta\beta\beta\rangle. \qquad (2.4.3.3)$$

Here, the ground states for all spin systems, including a single- and multi-spin systems, are denoted by the same notation $|1\rangle$.

Note: In the following sections, the index k applies to eigen-kets, and r, to the basis kets of a multi-spin system. Also, the letter n is the dimension of the eigen-kets, whereas m is a number defined locally in each subsection.

2.5 The Spin Hamiltonian

2.5.1 General spin Hamiltonian

Unlike the complete Hamiltonian for a molecular system, which is usually an extremely complex differential operator, a nuclear spin Hamiltonian is an $n \times n$ matrix acting on a Hilbert space that is spanned by n independent state-functions. Hence, it permits a closed solution for a rather complex system. Spin Hamiltonians in NMR are real, and, in general, contain the sum of various nuclear spin operators:

$$\mathcal{H} = \mathcal{H}^Z + \mathcal{H}^Q + \mathcal{H}^D + \mathcal{H}^C + \mathcal{H}^K + \mathcal{H}^J \qquad (2.5.1.1)$$

Here,

- $\mathcal{H}^Z$, the *Zeeman Hamiltonian*, is a *Zeeman interaction operator* that corresponds to the Zeeman splitting of the potential energy (or *Zeeman energy*) of nuclear magnetic moments in a static magnetic field, and is independent of local field;
- $\mathcal{H}^Q$ is a *quadrupole interaction operator* that corresponds to the coupling of nuclear quadrupole moments with local electric field gradients, which differs from one lattice site to another in a random fashion, causing a spread of Larmor frequencies and hence inhomogeneous broadening [Abragam 1961];
- $\mathcal{H}^D$ is a *direct dipole-dipole coupling interaction operator* that corresponds to dipole-dipole interactions between spins in a rigid lattice, and is due to a local field $\mathbf{B}_{ij}$ produced by spin i at the site of spin j;
- $\mathcal{H}^C$, the *chemical-shift Hamiltonian*, is a *magnetic shielding interaction operator* that corresponds to a small local field due to the electron motion around the nuclei, and is directly proportional to the applied field $\mathbf{B}_0$;
- $\mathcal{H}^K$ is a *Knight shift operator* corresponding to the coupling of conduction electrons in metals[9] with nuclear spins [see Knight 1956 or Abragam 1961];
- $\mathcal{H}^J$, the *J-coupling Hamiltonian* is an *indirect spin-spin coupling interaction operator* that corresponds to bilinear spin-spin couplings (e.g., $\mathbf{I} \bullet \mathbf{S}$) between spins[10]. The couplings induce multiple resonance frequencies in liquid samples, with the separations between the components of a multiplet measured in Hertz and independent of the applied field $\mathbf{B}_0$ [p.183-4, Abragam 1961].

[9] This is due to paramagnetic susceptibility.

[10] This is due to the rapid rotation of molecules.

2.5.2 Spin Hamiltonian for liquid samples

In isotropic liquids, direct dipole-dipole interactions and quadrupole interactions are cancelled by motional averaging because molecules and ions reorient rapidly on the NMR time-scale. Also, metals are often not involved in protein NMR spectra. Therefore, the Hamiltonian for the system is reduced to

$$\mathcal{H} = \mathcal{H}^Z + \mathcal{H}^C + \mathcal{H}^J \qquad (2.5.2.1)$$

Consequently, the determination of chemical shifts (δ) and indirect couplings (J) are much easier in liquid samples than in solids.

2.5.3 Properties of the spin Hamiltonian

The way a Hamiltonian is constructed determines that it is *self-adjoint* (or *Hermitian*):

$$\mathcal{H} = \mathcal{H}^\dagger \qquad (2.5.3.1)$$

The *adjoint* of $\mathcal{H}$ is defined as

$$(\mathcal{H}^\dagger)_{rs} = (\mathcal{H}_{sr})^* \qquad (2.5.3.2)$$

This can be seen in the Hamiltonian constructed for a two-spin-½ system, which will be given as an example later. This property ensures that eigen-vectors corresponding to distinct eigen-values are mutually orthogonal. (See, e.g., p.361, Greenberg 1978.)

2.6 Density Operators in NMR

2.6.1 Spin density operators

The *ensemble density operator*, $\sigma(t)$, is defined by

$$\sigma(t) = \sum_k p_k |\psi_k(t)\rangle \langle\psi_k(t)| = \sum_k p_k \sigma_k(t) \qquad (2.6.1.1)$$

where p_k is the probability of finding a system in the k^{th} magnetic quantum state $|\psi_k(t)\rangle$. Hence,

$$\sum_k p_k = 1 \qquad \text{and} \qquad 0 \le p_1, p_2, ..., p_n \le 1 \qquad (2.6.1.2)$$

$\sigma_k(t)$ is the *state density operator* associated with the k^{th} magnetic quantum state of a system such that

$$\sigma_k(t) = |\psi_k(t)\rangle \langle\psi_k(t)| = \sum_r \sum_s \overline{c_{k,r} c_{k,s}^*} |r\rangle \langle s| \qquad (2.6.1.3)$$

Here, $\overline{c_{k,r} c_{k,s}^*}$ is an average of the product $c_{k,r} c_{k,s}^*$ taken over the system ensemble. Two other alternative equations are

$$\sigma_{k,rs}(t) = \langle r|\sigma_k(t)|s\rangle = \overline{c_{k,r} c_{k,s}^*} \qquad (2.6.1.4)$$

$$\sigma_{rs}(t) = \langle r|\sigma(t)|s\rangle = \sum_k p_k \langle r|\sigma_k(t)|s\rangle = \sum_k p_k \sigma_{k,rs}(t) \qquad (2.6.1.5)$$

2.7 Observable Operators

2.7.1 Observable operators and their expectation values

By definition, a Hermitian operator **A** for a quantum mechanics system is an *observable* (*operator*) if its orthonormal eigen-vectors form a basis state space. According to *the third postulate of quantum mechanics*, the only possible measurement of a physical quantity is one of the eigen-values of an observable operator of the system.

The *mean value of an observable* **A** for a magnetic quantum state $|\psi_k(t)\rangle$ is the k^{th} *expectation value* of **A**, denoted by $\langle \mathbf{A} \rangle_k$. It is defined as an average of the results obtained when a large number of measurements of this observable are performed on a system ensemble that is purely in state $|\psi_k(t)\rangle$. It is well-know that, if $|\psi_k(t)\rangle$ is normalised, then

$$\langle \mathbf{A} \rangle_k = \langle \psi_k(t)| \mathbf{A} |\psi_k(t)\rangle = \mathrm{tr}\{\mathbf{A}\sigma_k(t)\} \tag{2.7.1.1}$$

This concept is easy to understand if there exists only one magnetic quantum state for the system ensemble, which is called a *free spin system*. A discussion for this can be found in p.24-6, Abragam 1961.

If a system ensemble contains more than one magnetic quantum state, then such a 'pure state' cannot exist. Nevertheless, one can imagine its existence in order to understand the next formula regarding the mean value of **A** for the entire spin system. It says that the *expectation value* $\langle \mathbf{A} \rangle$ for a multi-spin system can be evaluated by the trace of the product of **A** and $\sigma(t)$:

$$\langle \mathbf{A} \rangle = \sum_k p_k \langle \psi_k(t)| \mathbf{A} |\psi_k(t)\rangle = \mathrm{tr}\{\mathbf{A}\sigma(t)\} \tag{2.7.1.2}$$

2.7.2 Spin operator and finite rotation operator

One of the most important observable operators in quantum mechanics is the *spin operator* **I**, of which the x, y and z components are generators of a Lie group that obey the *Lie algebra*:

$$[\mathbf{I}_\alpha, \mathbf{I}_\beta] = i\,\mathbf{I}_\gamma \tag{2.7.2.1}$$

where the subscript set (α, β, γ) symbolises a cyclic permutation of (x, y, z). The Lie algebra relationship can be derived by successive system rotations with *infinitesimal rotation operators* for eigen-kets by an infinitesimal angle $\delta\varphi$ about the α-axes, $\alpha \in \{x, y, z\}$. (See, e.g., p.698, Cohen-Tannoudji *et al.*1973.)

For a free spin system of a single spin-½ nucleus, the three components of **I** are

$$\mathbf{I}_x = \frac{1}{2}\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad \mathbf{I}_y = \frac{1}{2}\begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}, \quad \mathbf{I}_z = \frac{1}{2}\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \tag{2.7.2.2}$$

The spin operator for a multi-spin system is called the *total spin operator*, which is defined as a tensor sum of the corresponding components of all spins in the system. For example, the x-component of the total spin operator of a two-spin-½ system, $\mathbf{I}_x$, is

$$\mathbf{I}_x = \mathbf{I}_{x,1} \oplus \mathbf{I}_{x,2} = \frac{1}{2}\begin{bmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{bmatrix} \tag{2.7.2.3}$$

Note that there are two important properties regarding spin operators:

1. The expectation values of the components of the spin operators, $\langle \mathbf{I}_x \rangle$, $\langle \mathbf{I}_y \rangle$ and $\langle \mathbf{I}_z \rangle$, are directly proportional to the measurable components of magnetic moment **M**. (See, e.g., p.25, Abragam 1961.)

2. The k^{th} (total) *magnetic quantum number* M for the k^{th} eigen-state $|\psi_k(t)\rangle$ of a spin system is defined by the k^{th} expectation value of $\mathbf{I}_z$.

$$M_k = \langle \psi_k(t) | \mathbf{I}_z | \psi_k(t) \rangle \tag{2.7.2.4}$$

where $\mathbf{I}_z$ is the z-component of the total spin operator of the system.

A set of total magnetic quantum numbers is denoted by M:

$$M = \{ M_1, M_2, \ldots \} \tag{2.7.2.5}$$

An example can be found in Section 2.9.

2.7.3 Raising and lowering operators

The *raising operator* $\mathbf{I}_+$ and *lowering operator* $\mathbf{I}_-$ are defined by

$$\mathbf{I}_+ = \mathbf{I}_x + i\,\mathbf{I}_y \qquad \text{and} \qquad \mathbf{I}_- = \mathbf{I}_x - i\,\mathbf{I}_y \tag{2.7.3.1}$$

For a single-spin-½ system, they are

$$\mathbf{I}_+ = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \qquad \text{and} \qquad \mathbf{I}_- = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} \tag{2.7.3.2}$$

For a two-spin-½ system, the raising and lowering operators are, respectively, a tensor sum of two corresponding spin operators:

$$\mathbf{I}_+ = \mathbf{I}_{+,1} \oplus \mathbf{I}_{+,2} = \begin{bmatrix} 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix} \tag{2.7.3.3}$$

$$\mathbf{I}_- = \mathbf{I}_{-,1} \oplus \mathbf{I}_{-,2} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \end{bmatrix} \tag{2.7.3.4}$$

Note: One of the important properties of raising operators is that the raising *transition probabilities* per unit time P_{lk} from states $|\psi_l(t)\rangle$ to $|\psi_k(t)\rangle$, where $l > k$, are directly proportional to $|\langle \psi_k(t) | \mathbf{I}_+ | \psi_l(t) \rangle|^2$. In other words.

$$P_{lk} \propto |\langle \psi_k(t) | \mathbf{I}_+ | \psi_l(t) \rangle|^2 = A_{lk} \tag{2.7.3.5}$$

This probability is also equal to the lowering transition probabilities P_{kl} from states $|\psi_k(t)\rangle$ to $|\psi_l(t)\rangle$, where, again, $l > k$:

$$P_{kl} \propto |\langle\psi_l(t)|\, \mathbf{I}_-\, |\psi_k(t)\rangle|^2 = A_{kl} \tag{2.7.3.6}$$

A_{lk} and A_{kl} are called the *relative intensities*. For example,

$$A_{42} = |\langle\psi_2(t)|\, \mathbf{I}_+\, |\psi_4(t)\rangle|^2 = |\langle\psi_4(t)|\, \mathbf{I}_-\, |\psi_2(t)\rangle|^2 = A_{24} \tag{2.7.3.7}$$

The results for a two-spin-½ system are presented in Section 2.8.

These relationships are closely related to the well-known result from time-independent perturbation theory:

$$P_{lk} = |\langle\psi_k(t)|\, \mathcal{H}_1\, |\psi_l(t)\rangle|^2\, g(\nu_0) \tag{2.7.3.8}$$

Here, $g(\nu_0)$ is a distribution function around the spin system's central value at ν_0. $\mathcal{H}_1$ is a perturbation Hamiltonian:

$$\mathcal{H}_1 = -\gamma(B_x \mathbf{I}_x + B_y \mathbf{I}_y) \tag{2.7.3.9}$$

$$= -\gamma B_1\, (\mathbf{I}_x \cos(\omega t) + \mathbf{I}_y \sin(\omega t))$$

$$= -\frac{1}{2}\gamma B_1\, (\mathbf{I}_+ \exp(-i\omega t) + \mathbf{I}_- \exp(i\omega t))$$

(See, e.g., p.28, Abragam 1961. Here, the factor $1/\hbar^2$ has been removed for the reason mentioned in the beginning of this section.) Note: only one term in the brackets is involved in a transition: either the first term for a raising transition or the second term for a lowering transition.

2.8 Density Operator Equations in NMR

2.8.1 Density operator equation for unperturbed systems

Taking into account that $\mathcal{H}(t)$ is Hermitian, it can be derived from the Schrödinger equation that [see, e.g., p.299, Chapter 3, Cohen-Tannoudji *et al.* 1973]

$$i\frac{d}{dt}\sigma(t) = [\mathcal{H}(t), \sigma(t)] \tag{2.8.1.1}$$

This is called the *density operator equation* or *Liouville-von Neuman equation.* In particular, if $\mathcal{H}$ is independent of time, the general solution is simply

$$\sigma(t) = \mathbf{U}(t, t_0)\, \sigma(t_0)\, \mathbf{U}^{-1}(t, t_0) \tag{2.8.1.2}$$

$\mathbf{U}(t, t_0)$, called the *evolution operator* or *propagator*, is a unitary operator defined by an *exponential operator* such that

$$\mathbf{U}(t, t_0) = \exp\{-(i\mathcal{H})(t - t_0)\} \tag{2.8.1.3}$$

$$= \mathbf{I}_n - i\mathcal{H}(t - t_0) + \frac{1}{2!}(i\mathcal{H})^2(t - t_0)^2 - \frac{1}{3!}(i\mathcal{H})^3(t - t_0)^3 + \ldots$$

where $\mathbf{I}_n$ is a unity operator.

The result (2.8.1.2) expresses a transformation of the density operator. It represents a rotational transformation from the initial density operator $\sigma(t_0)$ to a new operator $\sigma(t)$ under the influence of the time-independent Hamiltonian $\mathcal{H}$. Expressing a time-

dependent operator as a 'sandwich' of its initial operator by a pair of exponential operators is frequently seen in quantum mechanics.

Note: The evolution operator $\mathbf{U}(t, t_0)$ changes the eigen-state from $|\psi(t_0)\rangle$ to $|\psi(t)\rangle$; whereas its inversion $\mathbf{U}^{-1}(t,t_0)$ changes the state $|\psi(t)\rangle$ back to $|\psi(t_0)\rangle$. In other words,

$$\mathbf{U}(t, t_0)\,|\psi(t_0)\rangle = |\psi(t)\rangle \tag{2.8.1.4}$$

$$\mathbf{U}^{-1}(t, t_0)\,|\psi(t)\rangle = |\psi(t_0)\rangle \tag{2.8.1.5}$$

Clearly, if t_0 is always chosen to be the origin of time, then, the notation can be simplified to

$$\sigma(t) = \mathbf{U}(t)\,\sigma_0\,\mathbf{U}^{-1}(t) \tag{2.8.1.6}$$

where

$$\sigma_0 = \sigma(t_0) \tag{2.8.1.7}$$

$$\mathbf{U}(t) = \exp\{-i\mathcal{H}t\} = \mathbb{I} - i\mathcal{H}t + \frac{1}{2!}(i\mathcal{H}t)^2 - \frac{1}{3!}(i\mathcal{H}t)^3 + \ldots \tag{2.8.1.8}$$

2.8.2 Hamiltonian super operator

A *super operator* is an operator acting on other operator(s), resulting in an operator. In fact, the exponential operator is a super operator. The idea of defining a super operator in quantum mechanics can be further seen in the next example.

Let $\hat{\mathcal{H}}$ be a *Hamiltonian super operator* that deals with a commutator such that

$$\hat{\mathcal{H}}\sigma(t) = [\mathcal{H}(t), \sigma(t)] \tag{2.8.2.1}$$

Then, the density operator equation for an unperturbed system becomes similar to the SE:

$$i\frac{d}{dt}\sigma(t) = \hat{\mathcal{H}}\,\sigma(t\} \tag{2.8.2.2}$$

Accordingly, the problem becomes an eigen-value problem in which the search is for the *eigen-operators* $[\sigma(t)]_{kl}$ corresponding to the eigen-values λ_{kl} of the super operator $\hat{\mathcal{H}}$. If $\hat{\mathcal{H}}$ is Hermitian and time-independent, the general solution to the differential equation in operator format can be expressed by a super exponential operator acting on the initial-value operator $\sigma(t_0)$:

$$\sigma(t) = \exp\{-i\,\hat{\mathcal{H}}t\}\sigma(t_0) \tag{2.8.2.3}$$

Not surprisingly, it looks similar to the solution of an ordinary differential equation; and indeed, this is a good way to remember it.

Note: In general, an operator $\mathbf{A}(t)$ satisfying $\mathbf{A}(t) = \exp\{-i\,\hat{\mathcal{H}}t\}\mathbf{A}(t_0\}$ is a nontrivial solution to the eigen-value problem $i\frac{d}{dt}\mathbf{A}(t) = \hat{\mathcal{H}}\,\mathbf{A}(t)$ of the time-independent self-adjoint super-operator $\hat{\mathcal{H}}$. Therefore, it is an eigen-operator.

Comparing with the outcome obtained in the last subsection, it can be shown that

$$\exp\{-i\,\hat{\mathcal{H}}t\}\sigma(t_0) = \mathbf{U}(t, t_0)\,\sigma(t_0)\,\mathbf{U}^{-1}(t, t_0) \tag{2.8.2.4}$$

where

$$\mathbf{U}(t, t_0) = \exp\{-(i\mathcal{H})(t - t_0)\} \tag{2.8.2.5}$$

2.9 Indirect Coupling of a Two-Spin-½ System

This section describes a famous case in NMR theory: a two-spin-½ system under the influence of a chemical shift Hamiltonian and a *J*-coupling Hamiltonian. The results presented in this section are based on the so-called traditional 'indirect method' [Banwell 1962]. It is very useful for relating many concepts to a quantum mechanics system and the product operator formalism.

2.9.1 Spin Hamiltonian of a two-spin-½ system

Let the resonance frequencies of the two spins, **I** and **S**, be

$$\omega_I = \gamma B_0 (1-\delta_I) \quad \text{and} \quad \omega_S = \gamma B_0 (1-\delta_S). \tag{2.9.1.1}$$

Here δ_I and δ_S are the chemical shifts of the two spins.

The total-energy spin Hamiltonian for the system is then

$$\mathcal{H} = \mathcal{H}^C + \mathcal{H}^J \tag{2.9.1.2}$$

$$= (-\omega_I \mathbf{I}_z - \omega_S \mathbf{S}_z) + J(\mathbf{I} \bullet \mathbf{S})$$

$$= -\omega_0 (\mathbf{I}_z \oplus \mathbf{S}_z) - \frac{1}{2}\Delta(\mathbf{I}_z \oplus \mathbf{S}_z) + J(\mathbf{I}_x \otimes \mathbf{S}_x + \mathbf{I}_y \otimes \mathbf{S}_y + \mathbf{I}_z \otimes \mathbf{S}_z)$$

$$= \begin{bmatrix} -\omega_0 + \frac{1}{4}J & 0 & 0 & 0 \\ 0 & -\frac{1}{2}\Delta - \frac{1}{4}J & \frac{1}{2}J & 0 \\ 0 & \frac{1}{2}J & \frac{1}{2}\Delta - \frac{1}{4}J & 0 \\ 0 & 0 & 0 & \omega_0 + \frac{1}{4}J \end{bmatrix}$$

where J is the scalar coupling constant between the two spins, and

$$\omega_0 = \frac{1}{2}(\omega_I + \omega_S) \tag{2.9.1.3}$$

$$\Delta = (\omega_I - \omega_S) \tag{2.9.1.4}$$

Note: The fact that the spin Hamiltonian consists only of the terms containing spin operators and their products is a starting point for the product operator formalism presented in the next chapter.

2.9.2 Eigen-values and eigen-vectors of a two-spin-½ system

Referring back to Eq. (2.7.2.1), the eigen-values corresponding to the Hamiltonian can be determined to be

$$E_1 = -\omega_0 + \frac{1}{4}J \tag{2.9.2.1}$$

$$E_2 = -\frac{1}{2}R - \frac{1}{4}J \tag{2.9.2.2}$$

$$E_3 = +\frac{1}{2}R - \frac{1}{4}J \tag{2.9.2.3}$$

$$E_4 = +\omega_0 + \frac{1}{4}J \tag{2.9.2.4}$$

where

$$R = \sqrt{J^2 + \Delta^2} \tag{2.9.2.5}$$

The corresponding normalised eigen-vectors are

$$|\psi_1\rangle = |\alpha\alpha\rangle = |½,½\rangle \tag{2.9.2.6}$$

$$|\psi_2\rangle = -\cos(\phi)\,|\alpha\beta\rangle + \sin(\phi)\,|\beta\alpha\rangle \tag{2.9.2.7}$$

$$|\psi_3\rangle = +\sin(\phi)\,|\alpha\beta\rangle + \cos(\phi)\,|\beta\alpha\rangle \tag{2.9.2.8}$$

$$|\psi_4\rangle = |\beta\beta\rangle = |-½,-½\rangle \tag{2.9.2.9}$$

where ϕ is such that

$$\tan(2\phi) = J/\Delta \quad \text{or} \quad \sin(2\phi) = J/R \tag{2.9.2.10}$$

2.9.3 Magnetic quantum numbers of a two-spin-½ system

The corresponding *magnetic quantum numbers* M_k, calculated from Eq.(2.7.2.5), are

$$M \in \{1, 0, 0, -1\} \tag{2.9.3.1}$$

2.9.4 Calculated spectrum of a two-spin-½ system

The observable spectral frequencies associated with energy transitions between levels differing by ±1 in magnetic quantum numbers are

$$\omega_{21} = \omega_0 - \frac{1}{2}(R + J) \tag{2.9.4.1}$$

$$\omega_{43} = \omega_0 - \frac{1}{2}(R - J) \tag{2.9.4.2}$$

$$\omega_{31} = \omega_0 + \frac{1}{2}(R - J) \tag{2.9.4.3}$$

$$\omega_{42} = \omega_0 + \frac{1}{2}(R + J) \tag{2.9.4.4}$$

From Eq. (2.7.3.5), the relative spectral intensities of the system are

$$A_{21} = A_{42} = 1 - J/R \tag{2.9.4.5}$$

$$A_{31} = A_{43} = 1 + J/R \tag{2.9.4.6}$$

The following diagrams show four calculated spectra corresponding to the so-called AM-, AX-, AB- and AA- (or A_2) system.

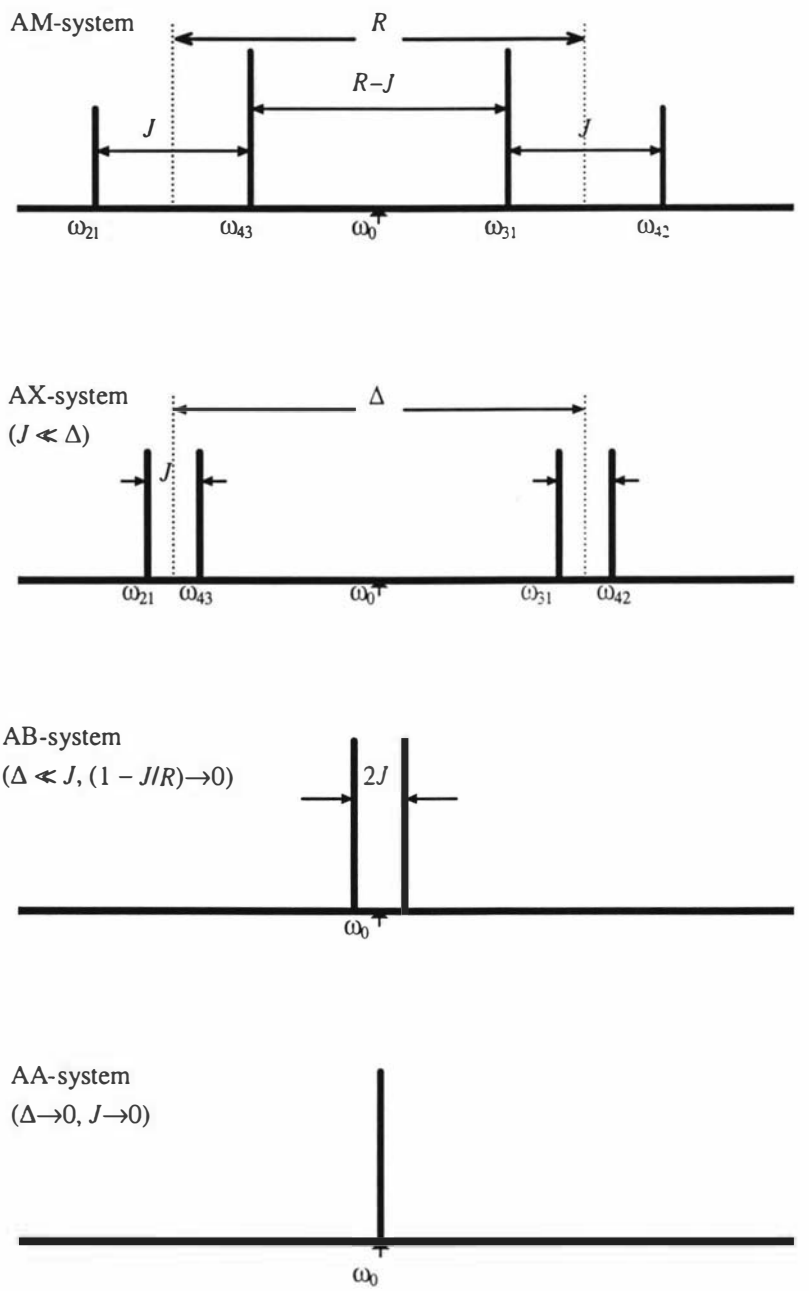


Figure 2.8: The calculated spectra corresponding to, from top to bottom, the AM-, AX-, AB- and AA-system.

Chapter 3:

Product Operator Formalism and Pulse Analysis

3.1 Introduction

This chapter describes the product operator formalism and its application to the analysis of pulse sequences for a number of different NMR experiments.

3.2 Product Operator Formalism

3.2.1 Liouville operator Space

In general, the density operator $\sigma(t)$ can be expressed as a combination of independent *basis operators* ξ_s that span a *Liouville operator space* of dimension n^2:

$$\sigma(t) = \sum_s b_s(t)\, \xi_s \qquad (3.2.1.1)$$

Each of the ξ_s is an $n \times n$ matrix acting on an n-dimensional Hilbert space.

Note: Since spin Hamiltonians consist only of a summation of terms containing spin operators and their product operators[1], the basis operators ξ_s can therefore be expressed in terms of the same set of spin operators and their product.
This can be easily justified by looking at the Schrödinger equation and the definition of density operator presented in the previous chapter:

$$\sigma(t) = \sum_k p_k\, |\psi_k(t)\rangle \langle\psi_k(t)| \qquad (3.2.1.2)$$

$$i\frac{d}{dt}|\psi(t)\rangle = \mathcal{H}(t)\,|\psi(t)\rangle \qquad (3.2.1.3)$$

For a two-spin-½ system with spin operators **I** and **S**, the basis state operators $|r\rangle$ are in a Hilbert space of dimension four. Hence, there are 16 basis operators ξ_s, each of them an independent operator or product of operators [Sørensen *et al.* 1983]:

$$\tfrac{1}{2}\mathbf{I}_n,$$

$$\mathbf{I}_x, \mathbf{I}_y, \mathbf{I}_z, \mathbf{S}_x, \mathbf{S}_y, \mathbf{S}_z,$$

$$2\mathbf{I}_x\mathbf{S}_x, 2\mathbf{I}_x\mathbf{S}_y, 2\mathbf{I}_x\mathbf{S}_z,$$

$$2\mathbf{I}_y\mathbf{S}_x, 2\mathbf{I}_y\mathbf{S}_y, 2\mathbf{I}_y\mathbf{S}_z,$$

$$2\mathbf{I}_z\mathbf{S}_x, 2\mathbf{I}_z\mathbf{S}_y, 2\mathbf{I}_z\mathbf{S}_z. \qquad (3.2.1.4)$$

where $\mathbf{I}_n$ is an $n \times n$ unity operator.

It is well known that each of these basis products is associated with a distinct physical meaning. Some of them can be represented either by a magnetisation vector, an energy level diagram, or a sketch of stick spectrum.

The following concepts are used frequently in describing the physical meanings of basis operators:

[1] See the remarks in the last section of previous chapter.

- *X- and y-magnetisation* corresponds to single-quantum transition between basis states, which are observable along the *x*- and *y*-axes, respectively.
- *Polarisation* is the *z*-magnetisation parallel to $\mathbf{B}_0$, which is not observable.
- *Coherence* refers to transverse single- or multi-quantum transition $|r\rangle \leftrightarrow |s\rangle$ between any pair of basis states, and is represented by a non-vanishing off-diagonal element σ_{rs} of the density operator $\sigma(t)$.
- *In-phase coherence* corresponds to a multiplet with individual components having the same phase along the *x*- or *y*-axes.
- *Anti-phase coherence* corresponds to a multiplet with individual components having opposite phase along the *x*- or *y*-axes.
- *Order of coherence* refers to the difference in the magnetic quantum number $(M_r - M_s)$ between two magnetic quantum states.

For example,

- $\mathbf{I}_x$ or $\mathbf{I}_y$ correspond to an in-phase *x*- or *y*-magnetisation of spin **I**.
- $\mathbf{I}_z$ corresponds to the polarisation of spin **I**.
- $2\mathbf{I}_x\mathbf{S}_z$ or $2\mathbf{I}_y\mathbf{S}_z$ correspond to an *x*- or *y*-magnetisation of spin **I** that is anti-phase with respect to spin **S**, which can be represented by an up-down doublet along the *x*- or *y*-axis.
- $2\mathbf{I}_z\mathbf{S}_z$ corresponds to *longitudinal spin-order* (or *J-order*) of spin **I** and **S**, which refers to a non-equilibrium population distribution without observable net magnetisation.
- $2\mathbf{I}_x\mathbf{S}_x$, $2\mathbf{I}_x\mathbf{S}_y$, $2\mathbf{I}_y\mathbf{S}_x$ and $2\mathbf{I}_y\mathbf{S}_y$ corresponds to two-spin coherence of spin **I** and **S**, which consist of a superposition of zero-quantum coherence operators ($\mathbf{I}_+\mathbf{S}_-$ and $\mathbf{I}_-\mathbf{S}_+$) and double-quantum coherence operators ($\mathbf{I}_+\mathbf{S}_+$ and $\mathbf{I}_-\mathbf{S}_-$), as may be seen from the following identities:

$$\mathbf{I}_x\mathbf{S}_x = \tfrac{1}{4}(\mathbf{I}_+\mathbf{S}_+ + \mathbf{I}_+\mathbf{S}_- + \mathbf{I}_-\mathbf{S}_+ + \mathbf{I}_-\mathbf{S}_-) \qquad (3.2.1.5)$$

$$i\mathbf{I}_x\mathbf{S}_y = \tfrac{1}{4}(\mathbf{I}_+\mathbf{S}_+ - \mathbf{I}_+\mathbf{S}_- + \mathbf{I}_-\mathbf{S}_+ - \mathbf{I}_-\mathbf{S}_-) \qquad (3.2.1.6)$$

$$i\mathbf{I}_y\mathbf{S}_x = \tfrac{1}{4}(\mathbf{I}_+\mathbf{S}_+ + \mathbf{I}_+\mathbf{S}_- - \mathbf{I}_-\mathbf{S}_+ - \mathbf{I}_-\mathbf{S}_-) \qquad (3.2.1.7)$$

$$\mathbf{I}_y\mathbf{S}_y = \tfrac{1}{4}(-\mathbf{I}_+\mathbf{S}_+ + \mathbf{I}_+\mathbf{S}_- + \mathbf{I}_-\mathbf{S}_+ - \mathbf{I}_-\mathbf{S}_-) \qquad (3.2.1.8)$$

where $\mathbf{I}_+$ and $\mathbf{S}_+$ are the raising operators, and $\mathbf{I}_-$ and $\mathbf{S}_-$, the lowering operators, as described in Chapter 2.

3.2.2 Evolution of spin systems

As noted in Chapter 2, the density operator equation is derived from the Schrödinger equation, to which the solution is a rotation from equilibrium density operator σ_0 to a new operator $\sigma(t)$ under the effect of time-independent Hamiltonian $\mathcal{H}$:

$$\sigma(t) = \mathbf{U}(t)\,\sigma_0\,\mathbf{U}^{-1}(t) \qquad (3.2.2.1)$$

$$\mathbf{U}(t) = \exp\{-i\mathcal{H}t\} \qquad (3.2.2.2)$$

This implies that simply focusing on the rotational transformation of the density operator can characterise the evolution of a spin system under the effect of various terms in $\mathcal{H}$.

Symbolically, a rotational transformation of a density operator is denoted as follows:

$$\sigma(t_0) \xrightarrow{\mathcal{H}t} \sigma(t) \tag{3.2.2.3}$$

Because $\sigma(t)$ can be expressed as a composition of terms containing various operators and product operators in Liouville space, the effect of $\mathcal{H}$ can be evaluated as a cascade of successive rotational transformations:

$$\sigma(t_0) \xrightarrow{\mathcal{H}_1 t} \sigma(t_1) \xrightarrow{\mathcal{H}_2 t} \sigma(t_2) \quad ... \quad \xrightarrow{\mathcal{H}_{n\times n} t} \sigma(t) \tag{3.2.2.4}$$

Note that the order of evaluation is also unimportant. These two properties rely simply on the fact that

$$\begin{aligned} \exp\{-i\mathcal{H}t\} &= \exp\{-i(\mathcal{H}_1 + \mathcal{H}_2 + ... + \mathcal{H}_{n\times 1})t\} \\ &= \exp\{-i(\mathcal{H}_{n\times 1} + ... + \mathcal{H}_2 + \mathcal{H}_1)t\} \\ &= \exp\{-i\mathcal{H}_{n\times 1} t\} ... \exp\{-i\mathcal{H}_2 t\} \exp\{-i\mathcal{H}_1 t\} \end{aligned} \tag{3.2.2.5}$$

Therefore,

$$\sigma(t) = ... (e^{-i\mathcal{H}_2 t} (e^{-i\mathcal{H}_1 t} \sigma_0 e^{i\mathcal{H}_1 t}) e^{i\mathcal{H}_2 t}) ... \tag{3.2.2.6}$$

In fact,

- each of the successive transformations corresponds to a rotation of a component of the density operator within a 3D operator subspace in Liouville space, and
- the solutions to the rotational transformations of all ξ_s under the effect of different Hamiltonian operators are known to obey a few simple rules, as will be described in the next few sections.

3.2.3 The basis of rotational transformations

Let $\mathcal{L}_\xi = (\xi_\alpha, \xi_\beta, \xi_\gamma)$ be a 3D operator subspace of Liouville operator space that is isomorphic to Lie group generator space (S_x, S_y, S_z) such that

$$[\xi_\alpha, \xi_\beta] = i\,\xi_\gamma \tag{3.2.3.1}$$

Here (α, β, γ) may take any cyclic permutation. Then, a clockwise rotation of the system about an axis parallel to say, ξ_γ, by an angle φ_ξ transforms the operators within $\mathcal{L}_\xi$ in the following way:

$$\xi_\alpha \xrightarrow{\varphi_\xi \xi_\gamma} \cos(\varphi)\,\xi_\alpha + \sin(\varphi)\,\xi_\beta \tag{3.2.3.2}$$

$$\xi_\beta \xrightarrow{\varphi_\xi \xi_\gamma} \cos(\varphi)\,\xi_\beta - \sin(\varphi)\,\xi_\alpha \tag{3.2.3.3}$$

$$\xi_\gamma \xrightarrow{\varphi_\xi \xi_\gamma} \xi_\gamma \tag{3.2.3.4}$$

The expressions denote the initial operator before transformation on the LHS and the resulting operators on the RHS, where $\xrightarrow{\varphi_\xi\, \xi_\gamma}$ indicates the conditions of transformation, which will be clearer after the discussion in the next few sections.

Note: These expressions form the basis of *rotational transformations* in the product operator formalism of NMR spectroscopy.

The last expression simply says that operators are unaffected by rotations about themselves.

As a matter of fact, the induction of the last terms in (3.2.3.2) and (3.2.3.3) hinges on the non-commutative property of Lie group generator ($\mathbf{S}_x$, $\mathbf{S}_y$, $\mathbf{S}_z$). For example, corresponding to the last two transformations, the Lie algebra relationships are respectively

$$[\mathbf{S}_z, \mathbf{S}_x] = i\,\mathbf{S}_y \tag{3.2.3.5}$$

$$[\mathbf{S}_z, \mathbf{S}_y] = -i\,\mathbf{S}_x \tag{3.2.3.6}$$

In other words, $[\mathbf{S}_z, \mathbf{S}_x]$, with $\mathbf{S}_z$ now an operator component in the Hamiltonian acting on $\mathbf{S}_x$, induces an imaginary term $\mathbf{S}_y$. On the other hand, $[\mathbf{S}_z, \mathbf{S}_y]$ induces a negative imaginary term $\mathbf{S}_x$. Lie algebra, however, does not indicate the important factors φ, $\cos(\varphi)$ and $\sin(\varphi)$ in the formulae.

3.2.4 Generalised cyclic notation

The formulation method that describes a transformation of a basis operator under the effect of a component of the total Hamiltonian is called the *product operator formalism* in NMR. A generalised cyclic notation for the first two expressions of rotational transformations within a three-dimensional operator subspace $\mathcal{L}_\xi = (\xi_\alpha, \xi_\beta, \xi_\gamma)$ can be given as follows:

$$\xi_\alpha \xrightarrow{\varphi_\xi\, \xi_\gamma} \cos(\varphi)\,\xi_\alpha + \eta_{\alpha\beta}\sin(\varphi)\,\xi_\beta \tag{3.2.4.1}$$

where (α, β, γ) symbolises a cyclic permutation of (x, y, z), and

$$\eta_{\alpha\beta} = \begin{cases} +1 & \text{if the permutation circle is clockwise;} \\ -1 & \text{otherwise.} \end{cases} \tag{3.2.4.2}$$

For example, if an operator ξ_x is transformed to ξ_y, then, $\eta_{\alpha\beta} = \eta_{xy} = +1$ because the permutation circle of (x, y, z) is $x \rightarrow y \rightarrow z$. On the other hand, if the permutation (α, β, γ) is (y, x, z), then, $\eta_{\alpha\beta} = \eta_{yx} = -1$, as shown in Figure 3.1.

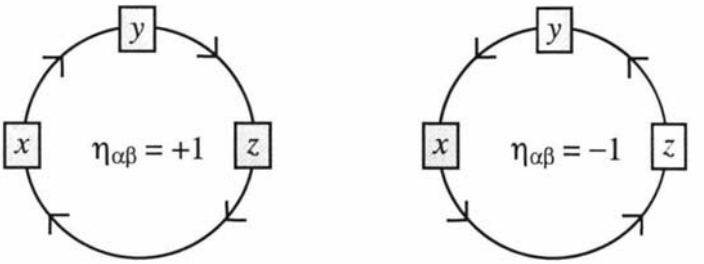


Figure 3.1: The relationship between the permutation circle and $\eta_{\alpha\beta}$ in Equation (3.2.4.1), where (α, β, γ) symbolises a cyclic permutation of (x, y, z) in the figure, either clockwise (left) or anti-clockwise (right). Two examples can be found in Equation (3.2.5.2) and (3.2.5.3).

Note that all other ways to describe rotational transformations are unnecessarily complicated if you can remember this simple rule.

For simplicity, the following shorthand notations for sine and cosine related to Zeeman and chemical-shift Hamiltonians of spin **I** and **S** are often used:

$$c_I = \cos(\varphi_I), \qquad s_I = \sin(\varphi_I). \tag{3.2.4.3}$$

$$c_S = \cos(\varphi_S), \qquad s_S = \sin(\varphi_S). \tag{3.2.4.4}$$

Similar equations related to J-coupling of two spins are

$$c_J = \cos(\varphi_J), \qquad s_J = \sin(\varphi_J) \tag{3.2.4.5}$$

The definitions of φ_I, φ_S and φ_J will be specified whenever necessary.

3.2.5 Evolution under Zeeman and chemical shift Hamiltonian

Consider a two-spin-½ system with spin operators **I** and **S**. Then, the Zeeman and chemical shift Hamiltonian $\mathcal{H}$ of the system will induce only a rotation within one of the following two $\mathcal{L}_\xi$ subspaces:

$$(\mathbf{I}_x, \mathbf{I}_y, \mathbf{I}_z), (\mathbf{S}_x, \mathbf{S}_y, \mathbf{S}_z) \tag{3.2.5.1}$$

For example, the evolution of $\mathbf{S}_x$ and $\mathbf{S}_y$ under the effect of $\mathcal{H} = \Omega_S \mathbf{S}_z$ over a duration τ is characterised by the following two rotational transformations about the z-axis of a Cartesian coordinate system by an angle $\varphi_S = \Omega_S \tau$:

$$\mathbf{S}_x \xrightarrow{\varphi_S \mathbf{S}_z} \cos(\varphi_S)\, \mathbf{S}_x + \sin(\varphi_S)\, \mathbf{S}_y \tag{3.2.5.2}$$

$$\mathbf{S}_y \xrightarrow{\varphi_S \mathbf{S}_z} \cos(\varphi_S)\, \mathbf{S}_y - \sin(\varphi_S)\, \mathbf{S}_x \tag{3.2.5.3}$$

Here,

$$\varphi_S \mathbf{S}_z = \mathcal{H}\tau \tag{3.2.5.4}$$

is the major factor in the exponent of the rotation operator **U**(t).

3.2.6 Evolution under pulses

Just like the Zeeman and chemical shift Hamiltonian, a pulse Hamiltonian also induces only a rotation within one of the following two operator subspaces:

$$(\mathbf{I}_x, \mathbf{I}_y, \mathbf{I}_z), (\mathbf{S}_x, \mathbf{S}_y, \mathbf{S}_z) \tag{3.2.6.1}$$

For example, the evolution of $\mathbf{I}_y$ and $\mathbf{I}_z$ under the effect of an x-pulse $\Omega_I \mathbf{I}_x$ over a duration τ is characterised by the following two rotational transformations about the x-axis by an angle $\varphi_I = \Omega_I\tau$:

$$\mathbf{I}_y \xrightarrow{\varphi_I \mathbf{I}_x} c_I \mathbf{I}_y + s_I \mathbf{I}_z \tag{3.2.6.2}$$

$$\mathbf{I}_z \xrightarrow{\varphi_I \mathbf{I}_x} c_I \mathbf{I}_z - s_I \mathbf{I}_y \tag{3.2.6.3}$$

Similar equations for the evolutions of $\mathbf{I}_z$ and $\mathbf{I}_x$ under the effect of a y-pulse $\Omega_I \mathbf{I}_y$ are

$$\mathbf{I}_z \xrightarrow{\varphi_I \mathbf{I}_y} c_I \mathbf{I}_z + s_I \mathbf{I}_x \tag{3.2.6.4}$$

$$\mathbf{I}_x \xrightarrow{\varphi_I \mathbf{I}_y} c_I \mathbf{I}_x - s_I \mathbf{I}_z \tag{3.2.6.5}$$

Under the effect of an x-pulse $\Omega_S\mathbf{S}_x$ that is applied only to the S spin over a duration τ, the evolution of product operators (such as $2\mathbf{I}_x\mathbf{S}_y$ and $2\mathbf{I}_x\mathbf{S}_z$) is characterised by the following two rotational transformations about the x-axis by an angle $\varphi_S = \Omega_S\tau$:

$$2\mathbf{I}_x\mathbf{S}_y \xrightarrow{\varphi_S\mathbf{S}_x} 2\mathbf{I}_x\mathbf{S}_y c_S + 2\mathbf{I}_x\mathbf{S}_z s_S \tag{3.2.6.6}$$

$$2\mathbf{I}_x\mathbf{S}_z \xrightarrow{\varphi_S\mathbf{S}_x} 2\mathbf{I}\,\mathbf{S}_z c_S - 2\mathbf{I}_x\mathbf{S}_y s_S \tag{3.2.6.7}$$

Note how $2\mathbf{I}_x$ is unaffected in the last two transformations.

When a 90° pulse is applied, the first term on the RHS disappears because cos(90°) is zero. For example,

$$2\mathbf{I}_x\mathbf{S}_y \xrightarrow{(\pi/2)\mathbf{S}_x} + 2\mathbf{I}_x\mathbf{S}_z \tag{3.2.6.8}$$

$$2\mathbf{I}_x\mathbf{S}_z \xrightarrow{(\pi/2)\mathbf{S}_x} - 2\mathbf{I}_x\mathbf{S}_y \tag{3.2.6.9}$$

3.2.7 Evolution under *J*-couplings

Different from the cases in the last two sections, a *J*-coupling Hamiltonian contains the coupling term $\mathbf{I}_z\mathbf{S}_z$, and induces a rotation within one of the following four operator subspaces:

$$(2\mathbf{I}_z\mathbf{S}_z, \mathbf{I}_x, 2\mathbf{I}_y\mathbf{S}_z),$$

$$(2\mathbf{I}_z\mathbf{S}_z, 2\mathbf{I}_x\mathbf{S}_z, \mathbf{I}_y),$$

$$(2\mathbf{I}_z\mathbf{S}_z, \mathbf{S}_x, 2\mathbf{I}_z\mathbf{S}_y),$$

$$(2\mathbf{I}_z\mathbf{S}_z, 2\mathbf{I}_z\mathbf{S}_x, \mathbf{S}_y). \tag{3.2.7.1}$$

For example, the evolution of $\mathbf{I}_x$ and $2\mathbf{I}_y\mathbf{S}_z$ under the effect of a J-coupling Hamiltonian $J_{IS}\mathbf{I}_z\mathbf{S}_z$ over a duration τ is characterised by the following two rotational transformations about the z-axis by an angle $\varphi_J = (J_{IS}/2)\tau$:

$$\mathbf{I}_x \xrightarrow{2\varphi_J\mathbf{I}_z\mathbf{S}_z} \mathbf{I}_x c_J + 2\mathbf{I}_y\mathbf{S}_z s_J \tag{3.2.7.2}$$

$$2\mathbf{I}_y\mathbf{S}_z \xrightarrow{2\varphi_J\mathbf{I}_z\mathbf{S}_z} 2\mathbf{I}_y\mathbf{S}_z c_J - \mathbf{I}_x s_J \tag{3.2.7.3}$$

where J_{IS} is in Hertz.

Note that these results represent evolution processes between in-phase x-magnetisation and anti-phase y-magnetisation.

Similar equations for $2\mathbf{I}_x\mathbf{S}_z$ and $\mathbf{I}_y$ in the second operator subspace are

$$2\mathbf{I}_x\mathbf{S}_z \xrightarrow{2\varphi_J\mathbf{I}_z\mathbf{S}_z} 2\mathbf{I}_x\mathbf{S}_z c_J + \mathbf{I}_y s_J \tag{3.2.7.4}$$

$$\mathbf{I}_y \xrightarrow{2\varphi_J\mathbf{I}_z\mathbf{S}_z} \mathbf{I}_y c_J - 2\mathbf{I}_x\mathbf{S}_z s_J \tag{3.2.7.5}$$

For $\varphi_J = (J_{IS}/2)\tau = \pi/2$ or $\tau = \pi / J_{IS}$, the last four transformations become

$$\mathbf{I}_x \xrightarrow{2\varphi_J\mathbf{I}_z\mathbf{S}_z} +2\mathbf{I}_y\mathbf{S}_z \tag{3.2.7.6}$$

$$2\mathbf{I}_y\mathbf{S}_z \xrightarrow{2\varphi_J\mathbf{I}_z\mathbf{S}_z} -\mathbf{I}_x \tag{3.2.7.7}$$

$$2\mathbf{I}_x\mathbf{S}_z \xrightarrow{2\varphi_J\mathbf{I}_z\mathbf{S}_z} +\mathbf{I}_y \tag{3.2.7.8}$$

$$\mathbf{I}_y \xrightarrow{2\varphi_J\mathbf{I}_z\mathbf{S}_z} -2\mathbf{I}_x\mathbf{S}_z \tag{3.2.7.9}$$

Note: This means that, in a coupled IS system, a free evolution for a time $\tau = \pi / J_{IS}$ results in a *coherence transfer* between in-phase and anti-phase magnetisation.

3.2.8 More on coherence transfer

Suppose that a basis operator $2\mathbf{I}_x\mathbf{S}_z$ has been formed in a coupled **I-S** spin system, as described in the last subsection. A 90° y-pulse applied to both spins gives

$$2\mathbf{I}_x\mathbf{S}_z \xrightarrow{90°\mathbf{I}_y} -2\mathbf{I}_z\mathbf{S}_z \xrightarrow{90°\mathbf{S}_y} -2\mathbf{I}_z\mathbf{S}_x \tag{3.2.8.1}$$

Thus, the anti-phase magnetisation $2\mathbf{I}_x\mathbf{S}_z$ on spin **I** is entirely transferred to anti-phase magnetisation on spin **S**.

3.3 NMR Experimental Pulse Analysis

3.3.1 The spin echo experiment

This is one of the simplest 1D experiments. The principle behind it is the basis for a spin-lock pulse sequence that will be discussed shortly. The pulse sequence is

$$(\pi/2)_x - \tau - (\pi)_x - \tau - \text{(acquisition)} \qquad (3.3.1.1)$$

Here, the de-coupling process $\tau - (\pi)_x - \tau$ has been divided into a *de-phasing process* and a *refocusing process* before and after the 180° pulse. As a consequence, the magnetic inhomogeneity of the field is removed and will not be included in the following derivation. Using the product operator formalism and with $\varphi_\tau = (J_{IS}/2)\tau$, it can be shown that

$$\begin{aligned}
\mathbf{I}_z &\xrightarrow{(\pi/2)\mathbf{I}_x} -\mathbf{I}_y \\
&\xrightarrow{\varphi_\tau 2\mathbf{I}_z\mathbf{S}_z} -(c_J\mathbf{I}_y - s_J 2\mathbf{I}_x\mathbf{S}_z) \\
&\xrightarrow{\pi\mathbf{I}_x} +c_J\mathbf{I}_y + s_J 2\mathbf{I}_x\mathbf{S}_z \\
&\xrightarrow{\pi\mathbf{S}_x} +c_J\mathbf{I}_y - s_J 2\mathbf{I}_x\mathbf{S}_z \\
&\xrightarrow{\varphi_\tau 2\mathbf{I}_z\mathbf{S}_z} +c_J(c_J\mathbf{I}_y - s_J 2\mathbf{I}_x\mathbf{S}_z) - s_J(c_J 2\mathbf{I}_x\mathbf{S}_z + s_J\mathbf{I}_y) \\
&= \cos(2\varphi_\tau)\,\mathbf{I}_y - \sin(2\varphi_\tau)\,2\mathbf{I}_x\mathbf{S}_z
\end{aligned} \qquad (3.3.1.2)$$

Note that the repetition of the last four transformations under the effect of the combined Hamiltonian $(\varphi_J 2\mathbf{I}_z\mathbf{S}_z + \pi\mathbf{I}_x + \pi\mathbf{S}_x + \varphi_J 2\mathbf{I}_z\mathbf{S}_z)$ is often used to de-couple spins within a spin system.

For simplicity, the last four transformations shall be combined into one expression and denoted by $\mathcal{H}^{[\tau,\pi,\tau]}$:

$$-\mathbf{I}_y \xrightarrow{(2\tau)\mathcal{H}^{[\tau,\pi,\tau]}} \cos(2\varphi_\tau)\,\mathbf{I}_y - \sin(2\varphi_\tau)\,2\mathbf{I}_x\mathbf{S}_z \qquad (3.3.1.3)$$

By setting $2\varphi_\tau = \pi$, or $\tau = \pi/J_{IS}$, the transformation leads to a pure in-phase spectrum:

$$-\mathbf{I}_y \xrightarrow{(2\pi/J_{IS})\mathcal{H}^{[\tau,\pi,\tau]}} \mathbf{I}_y \qquad (3.3.1.4)$$

Similarly, a pure anti-phase spectrum can be obtained by setting $2\varphi_\tau$ to $\pi/2$.

The following derivation shows how $\mathcal{H}^{[\tau,\pi,\tau]}$ affects an anti-coherence operator $2\mathbf{I}_x\mathbf{S}_z$:

$$\begin{aligned}
2\mathbf{I}_x\mathbf{S}_z &\xrightarrow{\varphi_\tau 2\mathbf{I}_z\mathbf{S}_z} c_\tau 2\mathbf{I}_x\mathbf{S}_z + s_\tau\mathbf{I}_y \\
&\xrightarrow{\pi\mathbf{I}_x} c_\tau 2\mathbf{I}_x\mathbf{S}_z - s_\tau\mathbf{I}_y \\
&\xrightarrow{\pi\mathbf{S}_x} -c_\tau 2\mathbf{I}_x\mathbf{S}_z - s_\tau\mathbf{I}_y \\
&\xrightarrow{\varphi_\tau 2\mathbf{I}_z\mathbf{S}_z} -c_\tau(c_\tau 2\mathbf{I}_x\mathbf{S}_z + s_\tau\mathbf{I}_y) - s_\tau(c_\tau\mathbf{I}_y - s_\tau 2\mathbf{I}_x\mathbf{S}_z) \\
&= -\sin(2\varphi_\tau)\,\mathbf{I}_y - \cos(2\varphi_\tau)\,2\mathbf{I}_x\mathbf{S}_z
\end{aligned}$$

that is,

$$2\mathbf{I}_x\mathbf{S}_z \xrightarrow{(2\tau)\mathcal{H}^{[\tau,\pi,\tau]}} -\sin(2\varphi_\tau)\,\mathbf{I}_y - \cos(2\varphi_\tau)\,2\mathbf{I}_x\mathbf{S}_z \qquad (3.3.1.5)$$

Therefore, by setting $2\varphi_\tau = \pi$, or $\tau = \pi/J_{IS}$, the transformation leads to a pure anti-phase peak:

$$2\mathbf{I}_x\mathbf{S}_z \xrightarrow{(2\pi/J_{IS})\mathcal{H}^{I\tau,\pi,\tau I}} -2\mathbf{I}_x\mathbf{S}_z \qquad (3.3.1.6)$$

The important fact is: $\mathcal{H}^{I\tau,\pi,\tau I}$ can transfer a coherence from in-phase to in-phase, or anti-phase to anti-phase.

3.3.2 The COSY experiment

This is short for 'homonuclear 2D correlation spectroscopy'. It is the simplest 2D experiment proposed by Jeener [Jeener 1971] and demonstrated by Ernst's group [Aue *et al.* 1976]. The experiment is based on the coherence transfer from one transition to another due to the existence of scalar coupling. The resulting COSY spectrum shows different cross-peak connectivity patterns for different spin systems, thereby allowing the chemical shifts of coupling partners to be identified.

The pulse sequence is

$$(\pi/2)_x - t_1 - (\pi/2)_x - t_2 \qquad (3.3.2.1)$$

During the evolution period t_1, the system undergoes two rotational transformations of spin $\mathbf{I}$ about $\mathbf{I}_z$: one is by an angle $\varphi_I = \Omega_I t_1$ under the effect of the chemical shift Hamiltonian $\Omega_I\mathbf{I}_z$; and one is by an angle $\varphi_J = (J_{IS}/2)t_1$ under the effect of the scalar coupling Hamiltonian $J_{IS}\,\mathbf{I}_z\mathbf{S}_z$. Then, under the effect of the second 90° pulse, it induces an observable anti-phase coherence of spin $\mathbf{S}$ that precesses at Ω_S during t_2. A series of FIDs are collected, each of them[2] with a different value of t_1, varying incrementally from zero to $t_{1\max}$.

Using product operator formalism as well as the abbreviations c_I, c_J, s_I and s_J for $\cos(\varphi_I)$, $\cos(\varphi_S)$, $\sin(\varphi_I)$ and $\sin(\varphi_J)$, respectively, it can be shown that

~~$\mathbf{I}_z$ and $\mathbf{S}_z$ $\xrightarrow{(\pi/2)\,\mathbf{S}_x \text{ and } (\pi/2)\,\mathbf{I}_x}$~~

$$\begin{aligned}
\mathbf{I}_z &\xrightarrow{(\pi/2)\,\mathbf{I}_x} -\mathbf{I}_y \\
&\xrightarrow{\varphi_I \mathbf{I}_z} -c_I\mathbf{I}_y + s_I\mathbf{I}_x \\
&\xrightarrow{\varphi_J 2\mathbf{I}_z\mathbf{S}_z} -c_I(c_J\mathbf{I}_y - s_J 2\mathbf{I}_x\mathbf{S}_z) + s_I(c_J\mathbf{I}_x + s_J 2\mathbf{I}_y\mathbf{S}_z) \\
&\xrightarrow{(\pi/2)\mathbf{I}_x} -c_I(c_J\mathbf{I}_z - s_J 2\mathbf{I}_x\mathbf{S}_z) + s_I(c_J\mathbf{I}_x + s_J 2\mathbf{I}_z\mathbf{S}_z) \\
&\xrightarrow{(\pi/2)\mathbf{S}_x} -c_I(c_J\mathbf{I}_z + s_J 2\mathbf{I}_x\mathbf{S}_y) + s_I(c_J\mathbf{I}_x - s_J 2\mathbf{I}_z\mathbf{S}_y).
\end{aligned} \qquad (3.3.2.2)$$

Only the last two terms contribute to observable signals. The term containing $\mathbf{I}_x$ gives rise to an in-phase doublet from the x-magnetisation of spin $\mathbf{I}$ at resonance frequency

[2] Each pair of the FIDs will be collected with the same t_1 if the States-Haberkorn-Ruben method is used, as will be explained in Chapter 5.

($\Omega_I \pm J_{IS}/2$) during both t_1 and t_2. The term containing $2\mathbf{I}_z\mathbf{S}_y$ gives rise to an anti-phase doublet from the y-magnetisation of spin **S** at frequency ($\Omega_I \pm J_{IS}/2$) during t_1 and ($\Omega_S \pm J_{IS}/2$) during t_2. In a 2D NMR spectrum, the former appears to be a *diagonal peak* centred at (Ω_I, Ω_I), and the latter, an off-diagonal *cross peak* centred at (Ω_I, Ω_S), as shown in Figure 3.2.

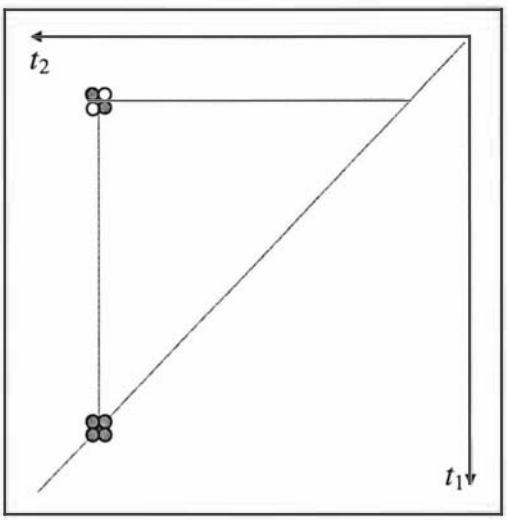


Figure 3.2: The 2D COSY peaks due initially to spin **I**. The in-phase diagonal peak corresponds to the term containing $\mathbf{I}_x$ in Equation (3.3.2.2), while the anti-phase cross peak corresponds to the term containing $2\mathbf{I}_z\mathbf{S}_y$.

In fact, the appearances of frequencies and phases of the peaks formed during t_1 can be clearly understood if the modulations of the two terms are expanded as follows:

$$+2s_I c_J = \sin((\Omega_I + J_{IS}/2)t_1) + \sin((\Omega_I - J_{IS}/2)t_1) \quad (3.3.2.3)$$

$$-2s_I s_J = \cos((\Omega_I + J_{IS}/2)t_1) - \cos((\Omega_I - J_{IS}/2)t_1). \quad (3.3.2.4)$$

Here, the signs for the two terms on the right indicate whether the two resonance frequencies represented by them are in-phase (3.3.2.3) or anti-phase (3.3.2.4).

Likewise, the four terms due initially to spin **S** are

$$\mathbf{S}_z \xrightarrow{(\pi/2)\mathbf{S}_x} -\mathbf{S}_y$$

$$\ldots \xrightarrow{(\pi/2)\mathbf{I}_x} -c_S(c_J\mathbf{S}_z + s_J 2\mathbf{I}_y\mathbf{S}_x) + s_S(c_J\mathbf{S}_x - s_J 2\mathbf{I}_y\mathbf{S}_z). \quad (3.3.2.5)$$

Again, only the last two terms contribute to observable signals. They give rise to a diagonal peak centred at (Ω_S, Ω_S) and an off-diagonal *cross* peak centred at (Ω_S, Ω_I), as shown in Figure 3.3.

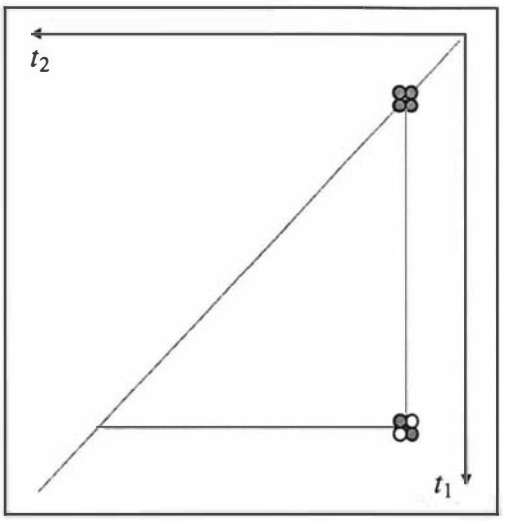


Figure 3.3: The 2D COSY peaks due initially to spin S. The in-phase diagonal peak corresponds to the term containing S_x in Equation (3.3.2.5), while the anti-phase cross peak corresponds to the term containing $2I_yS_z$.

Note: In both diagrams, the on-diagonal peaks arise from x-magnetisation, whereas the off-diagonal peaks from the y-magnetisation.

Therefore, the on-diagonal peaks are always 90° out of phase with the off-diagonal peaks in this simple COSY experiment. When processing the data, either the diagonal or off-diagonal peaks must be dispersive and hence undesirable.

3.3.3 Double-quantum filtered (DQF) COSY

The pulse sequence is

$$(\pi/2)_x - t_1 - (\pi/2)_x - \Delta - (\pi/2)_x - t_2 \qquad (3.3.3.1)$$

Here, Δ is a short delay, during which phase-cycling is applied, so that only double-quantum coherence present after the second pulse can contribute to observable signals. Thus, following the results in Equation (3.3.2.2) and (3.3.2.5), it can be derived that the resulting operators due initially to both spins **I** and **S** are as follows:

$$\begin{aligned}
\mathbf{I}_z \text{ and } \mathbf{S}_z &\xrightarrow{(\pi/2)\mathbf{S}_x \text{ and } (\pi/2)\mathbf{I}_x} -s_J\,(c_I\,2\mathbf{I}_x\mathbf{S}_y + c_S\,2\mathbf{I}_y\mathbf{S}_x) + \ldots \\
&= -\tfrac{1}{2}s_J\,\{c_I\,[(2\mathbf{I}_x\mathbf{S}_y + 2\mathbf{I}_y\mathbf{S}_x) - (2\mathbf{I}_y\mathbf{S}_x - 2\mathbf{I}_x\mathbf{S}_y)] \\
&\quad + c_S\,[(2\mathbf{I}_x\mathbf{S}_y + 2\mathbf{I}_y\mathbf{S}_x) + (2\mathbf{I}_y\mathbf{S}_x - 2\mathbf{I}_x\mathbf{S}_y)]\} + \ldots \\
&\xrightarrow{\Delta} -\tfrac{1}{2}s_J\,\{c_I(2\mathbf{I}_x\mathbf{S}_y + 2\mathbf{I}_y\mathbf{S}_x) + c_S(2\mathbf{I}_x\mathbf{S}_y + 2\mathbf{I}_y\mathbf{S}_x)\} \\
&\xrightarrow{(\pi/2)\mathbf{I}_x} -\tfrac{1}{2}s_J\,\{c_I(2\mathbf{I}_x\mathbf{S}_y + 2\mathbf{I}_z\mathbf{S}_x) + c_S(2\mathbf{I}_x\mathbf{S}_y + 2\mathbf{I}_z\mathbf{S}_x)\} \\
&\xrightarrow{(\pi/2)\mathbf{S}_x} -\tfrac{1}{2}s_J\,\{c_I(2\mathbf{I}_x\mathbf{S}_z + 2\mathbf{I}_z\mathbf{S}_x) + c_S(2\mathbf{I}_x\mathbf{S}_z + 2\mathbf{I}_z\mathbf{S}_x)\} \qquad (3.3.3.2)
\end{aligned}$$

Since

$$2c_I s_J = \sin((\Omega_I + J_{IS}/2)t_1) - \sin((\Omega_I - J_{IS}/2)t_1) \quad (3.3.3.3)$$

$$2c_S s_J = \sin((\Omega_S + J_{IS}/2)t_1) - \sin((\Omega_S - J_{IS}/2)t_1) \quad (3.3.3.4)$$

each of the resulting four operators give rise to an anti-phase doublet. The appearances of peaks corresponding to the four terms in the last two equations are:

- The first term gives rise to an anti-phase doublet of spin **I** at frequency $(\Omega_I \pm J_{IS}/2)$ during both t_1 and t_2.
- The second term gives rise to an anti-phase doublet of spin **S** at frequency $(\Omega_I \pm J_{IS}/2)$ during t_1 and $(\Omega_S \pm J_{IS}/2)$ during t_2.
- The third term gives rise to an anti-phase doublet of spin **I** at frequency $(\Omega_S \pm J_{IS}/2)$ during t_1 and $(\Omega_I \pm J_{IS}/2)$ during t_2.
- The fourth term gives rise to an anti-phase doublet of spin **S** at frequency $(\Omega_S \pm J_{IS}/2)$ during both t_1 and t_2.

They correspond to the 2D spectrum shown in Figure 3.4.

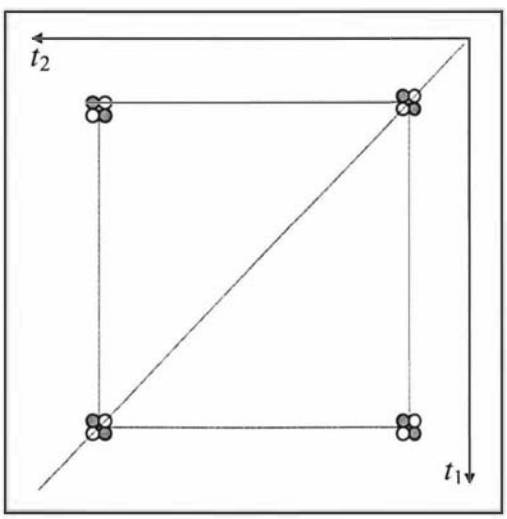


Figure 3.4: The anti-phase 2D DQF-COSY peaks due initially to both spins I and S. The diagonal peaks at the bottom-left and top-right corners correspond respectively to the first and fourth terms in Equation (3.3.3.2), while the cross peaks at the bottom-right and top-left corners correspond to the second and third terms.

Note: Both the on- and off-diagonal peaks arise from x-magnetisation from this experiment.

3.3.4 The NOESY experiments

This is an acronym for 2D *nuclear Overhauser effect spectroscopy* [Jeener *et al.* 1979, Macura and Ernst 1980, Kumar *et al.* 1980, Montelione *et al.* 1989]. The experiments are based on magnetisation transfer between dipolar-coupled spins in a cross-relaxation process rather than coherence transfer. The *cross* relaxation is induced by random motional processes of molecules in a liquid system, causing mutual spin flips

in pairs of dipolar-coupled spins [see e.g., Noggle and Schirmir 1971, or Campbell and Freeman 1973].

The pulse sequence is

$$(\pi/2)_x - t_1 - (\pi/2)_x - \tau_m - (\pi/2)_x - t_2 \qquad (3.3.4.1)$$

where τ_m is a mixing time during which phase cycling is used to suppress all magnetisation other than z-components, leaving them to exchange under cross relaxation. The scalar coupling is in this case not under consideration because transverse magnetisation has been suppressed. Supposing that dipolar coupling takes place during τ_m with a *cross-relaxation rate* $R^{[C]}$ and a *cross-relaxation leakage rate* $R^{[L]}$, then,

$$\begin{aligned}
\mathbf{I}_z &\xrightarrow{(\pi/2)\mathbf{I}_x} -\mathbf{I}_y \\
&\xrightarrow{\varphi_I \mathbf{I}_z} -c_I \mathbf{I}_y + s_I \mathbf{I}_x \\
&\xrightarrow{(\pi/2)\mathbf{I}_x} -c_I \mathbf{I}_z + s_I \mathbf{I}_x \\
&\xrightarrow{\tau_m} -c_I (R^{[L]} \mathbf{I}_z + R^{[C]} \mathbf{S}_z) \\
&\xrightarrow{(\pi/2)\mathbf{I}_x} c_I (R^{[L]} \mathbf{I}_y - R^{[C]} \mathbf{S}_z) \\
&\xrightarrow{(\pi/2)\mathbf{S}_x} c_I (R^{[L]} \mathbf{I}_y + R^{[C]} \mathbf{S}_y)
\end{aligned} \qquad (3.3.4.2)$$

The last expression indicates that each of the resulting operators give rise to an in-phase peak. The first term gives rise to a diagonal peak of spin **I** at frequency Ω_I during both t_1 and t_2. The second term gives rise to a cross peak correlated to spin **S** at the frequencies Ω_I during t_1 and Ω_S during t_2.

Similar analysis for spin **S** leads to another pair of diagonal and cross peaks centred at (Ω_S, Ω_S) and (Ω_S, Ω_I). These are shown in a 2D NMR below.

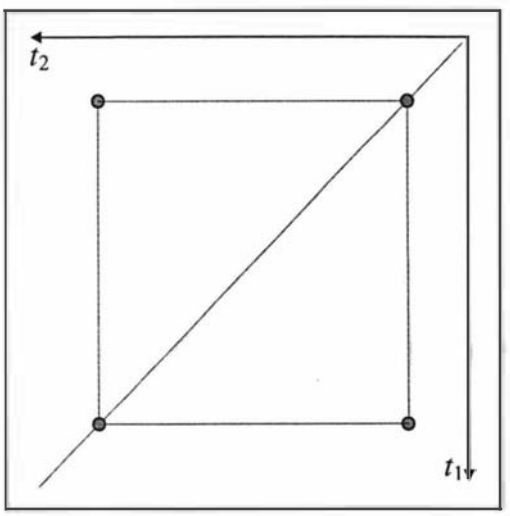


Figure 3.5: The in-phase 2D NOESY peaks due initially to both spins **I** and **S**. Spin **I** leads to a pair of peaks centred at (Ω_I, Ω_I) and (Ω_I, Ω_S). The diagonal peak at the bottom-left corner corresponds to the first term in Equation (3.3.4.2),

while the cross peak at the top-left corner corresponds to the second term. Similarly, spin S leads to another pair of diagonal and cross peaks centred at (Ω_S, Ω_S) and (Ω_S, Ω_I).

Note: NOESY signals are the results of dipole-dipole coupling in the mixing period. Therefore, cross peaks in NOESY provide useful information on the spatial connections of nearby nuclei within molecules.

It is understood that the cross-relaxation rate is proportional to $1/r^6$, where r is the inter-nuclear distance between two spins. Hence, NOESY cross peaks can be used to estimate inter-nuclear distances.

3.3.5 The TOCSY and HOHAHA experiments

TOCSY is an acronym for total correlation spectroscopy [Braunschweiler and Ernst 1983, Edison *et al.* 1991]. It is a general name for experiments that rely on any spin-locking technique to produce cross polarisation between neighbouring coupled spins, thereby providing information on the connection between nuclei within a spin system. Another similar experiment is *HOHAHA*, short for the homonuclear Hartmann-Hahn spectroscopy [Davis and Bax 1985]. Theoretical analysis for the experiments can be found in many publications [e.g., Kessler *et al.* 1988].

A common expression of the pulse sequence for both the experiments is

$$(\pi/2)_x - t_1 - \tau_m - t_2 \qquad (3.3.5.1)$$

where τ_m is the mixing time when a spin-locking sequence is applied.

In the case of TOCSY, it is an isotropic mixing induced by applying a strong $\mathbf{B}_1$ field parallel to $\mathbf{I}_x$ (thereby forcing all spins to precess about $\mathbf{B}_1$). In HOHAHA, however, the spin-locking mechanism is called 'Hartmann-Hahn mixing' induced by an $[\text{MLEV17}]_n$ sequence [Davis and Bax 1985, Bax and Davis 1985]:

$$(\pi/2)_x - t_1 - [\text{SL}_y\text{-MLEV-SL}_y] - t_2 \qquad (3.3.5.2)$$

SL_y is the y-pulse that produces the spin-locking in the mixing period.

Another method uses the $(\pi/2)_x$ — WALTZ — $(\pi/2)_x$ sequence [Rance 1987, Rucker 1989] for the spin-locking.

$$(\pi/2)_x - t_1 - (\pi/2)_x - [\text{WALTZ}] - (\pi/2)_x - t_2 \qquad (3.3.5.3)$$

The latter method is less difficult to apply because the decoupler of the spectrometer can be use for spin locking, and the transmitter for the pulses; whereas the former requires the frequency source used for the mixing sequence be coherent with that for the pulses. The phase cycling in (3.3.5.2) is as described in DQF-COSY, and (3.3.5.3), as in NOESY.

The Hamiltonian operators in both the spin-locking processes contain only the scalar coupling terms. For TOCSY, it is called the *isotropic mixing Hamiltonian*:

$$\mathcal{H}^{[s]} = \sum_{k<l} J_{kl}\, \mathbf{I}_k \bullet \mathbf{I}_l \qquad (3.3.5.4)$$

Detailed treatment for a two-spin-½ system can be found in Ernst 1987. Here, only the transformations corresponding to rotations about $\mathbf{B}_1$ will be considered. With a treatment similar to that for NOESY, the final operators will be

$$c_I (P^{[L]} I_y + P^{[C]} S_y) \tag{3.3.5.5}$$

where $P^{[C]}$ and $P^{[L]}$ are respectively the *cross-polarisation rate* and *cross-polarisation leakage rate*.

Note: Different from those of NOESY, the TOCSY signals are the results of scalar couplings in the spin-locking period. Therefore, cross peaks in TOCSY provide useful information on through-bond connectivity of nuclei within spin systems.

Chapter 4:

Protein NMR Experiments

4.1 Introduction

This chapter describes some technical details of protein NMR experiments, including sample selection, sample preparation, and experimental settings. It contains mainly summaries of useful notes collected from the author's practice in NMR experiments.

4.2 Sample Selection

NMR study of macromolecules is expensive and labour intensive. Besides evaluating the importance of the protein to be investigated, there are a number of other factors that should be considered when selecting samples. This section described four aspects of sample selection: feasibility, availability, solubility and stability.

4.2.1 Feasibility

The molecular weight should be the first thing to be considered. For a large molecule, an alternative can be to investigate a segment or segments that may be structurally or biologically significant.

Up until 1991, the largest protein NMR spectra assigned was a 25 kDa dimer [Arrowsmith *et al.* 1991], and the largest monomer for which a 2D NOESY was nearly totally assigned was a 19 kDa protein with 169 residues [Chubb *et al.* 1991]. Although NMR spectra assignment has become relatively easy today with the assistance of 3D NMR technique [see, e.g., Uhrinova *et al.* 1998, Carvanagh *et al.* 1996], it is still expected to be difficult and often not feasible for big molecules. In 1997, it was believed that, without better technology, around 40 kDa was the limit of molecular size with which the NMR spectroscopy might work[1]. However, the limit was suddenly reported [Borman 1998] to have increased by many times since a high-resolution[2] NMR spectroscopy technique called TROSY[3] was conceived and developed by researchers in Wüthrich's lab [Pervushin *et al.* 1997 and 1998, Czisch and Boelens 1998, Weigelt 1998].

The other thing to investigate is the amino sequence of the molecule. If more than two amino acids of the same kind are frequently chained together, it is likely that the NMR spectra will contain heavily overlapping peaks that cannot be resolved for structural calculation.

[1] This conclusion is drawn from the conversation between NMR experts in ISMAR95 and ANZMAG97, Australia.

[2] In NMR spectroscopy, resolution is defined as the ability to resolve two closely spaced peaks.

[3] Short for the transverse relaxation-optimised spectroscopy.

4.2.2 Availability

To obtain acceptable 2D NMR spectra, about 400 μL of sample at a minimum concentration of 2 mM is usually required.

Let M_r be the relative molecular mass. The formula to work out the amount required for every 100 μL of 1 mM sample is

$$W = M_r \times 1.05 \times 10^{-4} \text{ mg} \qquad (4.2.2.1)$$

Here, only 5% loss in sample preparation process is assumed. For instance, about 42 mg of hen-egg lysozyme (M_r 14,300) will be needed to make 400 μL of 7 mM sample.

4.2.3 Solubility

The protein used is usually required to be soluble in a water-based buffer that often contains 10% D_2O to establish a lock signal in experiments. A table of the most commonly used NMR buffers with useful pH ranges can be found in Primrose 1993. Different kinds of salt can also be added for better solubility [Arakawa 1985].

Deuterated 2,2,2-trifluoro-ethanol (TFE-d_2) has also been used [Wong *et al.* 1997] as a helix-enhancing co-solvent in NMR. It stabilises helices in regions with some α-helical propensity [Sonnichsen *et al.* 1992], which in turn can be fairly accurately determined by secondary structure prediction. In addition, TFE may provide better solubility for some proteins because of the alcohol in it. Our experiments with 25% TFE-d_2 have also resulted in spectra with better water suppression, possibly due to a better locked signal throughout the experiments. (Note that a preferred conformation of a protein segment in TFE may not necessarily correlate with the actual situation in water, in intact protein, or upon protein-protein interaction. For instance, in segments with both β-sheet and helical propensity, TFE is most likely to induce helix formation at least at a higher TFE concentration [Reutimann *et al.* 1981, Stone *et al.* 1985, Martenson *et al.* 1985, Sonnichsen *et al.* 1992].)

In some cases, pH levels may affect the solubility of proteins. For example, lower pH may help clear the viscosity of some samples taken out from a freezer. The samples may then be re-adjusted to a preferred pH value without precipitation. A plot for the pH dependence on the exchange rates of labile protons[4] in proteins can be found in Primrose 1993 (adapted from Wüthrich & Wagner 1979 as well as Wüthrich 1986).

4.2.4 Stability

If possible, a small quantity of sample at preferred concentration should be made first and left in a fridge for 24 to 72 hours so that the stability can be observed. Air oxidation, microbial contamination, and hydrolytic breakdown must be eliminated. For example, adding preservative (e.g. cyanine[5]) to a sample, keeping it sealed and

[4] See Chapter 7.

[5] This results in a sharp peak at about 3 ppm.

cold before and after experiments may help maximise the lifetime of a sample. Some samples may be kept in a freezer if they are not going to be used for a long time. Refer to the pH treatment in the last subsection if they become viscous later at room temperature.

In practice, the comparison of 1D spectra obtained before and after a period of time can give a reliable measurement of sample stability.

4.3 Sample Preparation

This section describes a few aspects of NMR sample preparation, including tube cleaning, filtration and dialysis, concentration, as well as pH adjustment and final purification. Details on sample preparation are also described very well in Primrose 1993.

4.3.1 Cleaning the NMR tubes

In general, cleaning agents can be NaOH, ethanol, or other cleaning solutions, as long as the tubes are rinsed with plenty of water. A final rinse with acetone or distilled water is frequently used to remove the last organic contents from the tube. To remove traces of water from the surface of the tube, it is recommended that tubes be placed on a perfectly flat tray at 125° C for about 30-45 minutes [Wilmad Glass 1998].

For serious cleaning, the tubes can be boiled in concentrated nitric acid. This will oxidise all peptide bonds and dissolves most inorganic materials. Finally, rinse the tubes with water and dry them at 125° C.

Never use abrasive materials that may cause scratches to clean NMR tubes, because the portion of sample that fills a scratch may experience a different magnetic field than the rest of the sample, which will result in spectral line-broadening.

4.3.2 Filtration and dialysis

Filtration and dialysis are two standard procedures to filter out small molecule impurities in NMR samples. Filtration can be done with various filters [see, e.g., Amicon 1998]. For easily obtained bulk samples like hen-egg lysozyme, the standard dialysis technique can be used, as described in Appendix 4a.

4.3.3 Concentrating NMR samples

A number of methods can be used to increase the concentration of NMR samples solution, including ultra-filtration, spin-drying, and lyophilising.

Ultra-filtration can be done effectively with a CENTRIPLUS® disposable centrifugal concentrator. This device is designed for concentrating and desalting 2-15 ml of macromolecular solution by ultra-filtration. It is a 110-mm height plastic unit (29 mm

in diameter) with two detachable components: a sample reservoir (with a lid on top and an isotropic membrane at the bottom), and a filtrate vial. Operation begins by adding sample solution to the sample reservoir. The unit is then centrifuged to drive small molecules, including solvents, through the membrane into the filtrate vial. Different MW cut-off units are available, ranging from 3 to 100 kDa. For the concentrations required in protein NMR, the recovery rates are 80-95%. For chemical compatibility and other technical parameters, see Amicon 1998.

Spin-drying is a technique that depends upon centrifugal spinning in a low temperature vacuum compartment. The centrifugal force retains the sample solution in a container and drives heavy molecules to the bottom. The vacuum environment helps speed up the drying process of water surface in low temperature. It can be done with a spin-drying device like SpeedVac®. In our practice, about 700 μL of sample solution is kept in a 1 mL Eppendorf® centrifuge container, and covered with perforated parafilm. It is then put into the SpeedVac and balanced. The temperature was first set to low then medium. This technique is very simple yet efficient, which may reduce the volume of a sample by 200 μL in 20 minutes. Theoretically, there should not be any loss of protein in the process.

Lyophilisation depends upon a vacuum freeze-drying process in a refrigerator with the assistance of a condenser. Before starting the process, the refrigerator and condenser must be turned on for at least 30 minutes. Meanwhile, samples are shell-frozen with dry ice and ethanol. Strict operation procedure must be followed, which may vary with different devices and sample quantity. This technique is recommended only for inexpensive samples, because the recovery rates are low for the small quantity required in NMR spectroscopy. Our experience showed only around 32% recovery rate from 500 mg of hen-egg lysozyme after lyophilising for 12 hours.

4.3.4 pH adjustment and final purification

This is the same as standard pH adjustment procedure for solutions. Usually, 0.1 and 0.5 mM of HCl and NaOH are used for the adjustment. The concentration of the sample should be higher with less volume at the beginning so that the final volume is correct.

A pH meter fitted with an electrode of 3-5 mm is recommended for preparing protein NMR samples, because the sample is usually contained in a 1-mL Eppendorf tube. Commence the process carefully and quickly, and do not wash the electrode during the process to avoid unnecessary loss in the process.

Finally, spin the sample for 5-10 minutes to pellet impure heavy material in the sample and transfer it to an NMR tube with a long glass pipette, as described in Primrose 1993.

4.4 NMR Experiments

This section describes some practical knowledge of NMR spectrometers and experiments, including a general description of an NMR spectrometer and how to set up NMR experiments.

4.4.1 NMR spectrometer system

The instrumental information on modern NMR spectrometers (Figure 4.1) can be found in many textbooks, manuals and papers [e.g., Frenkiel 1993].

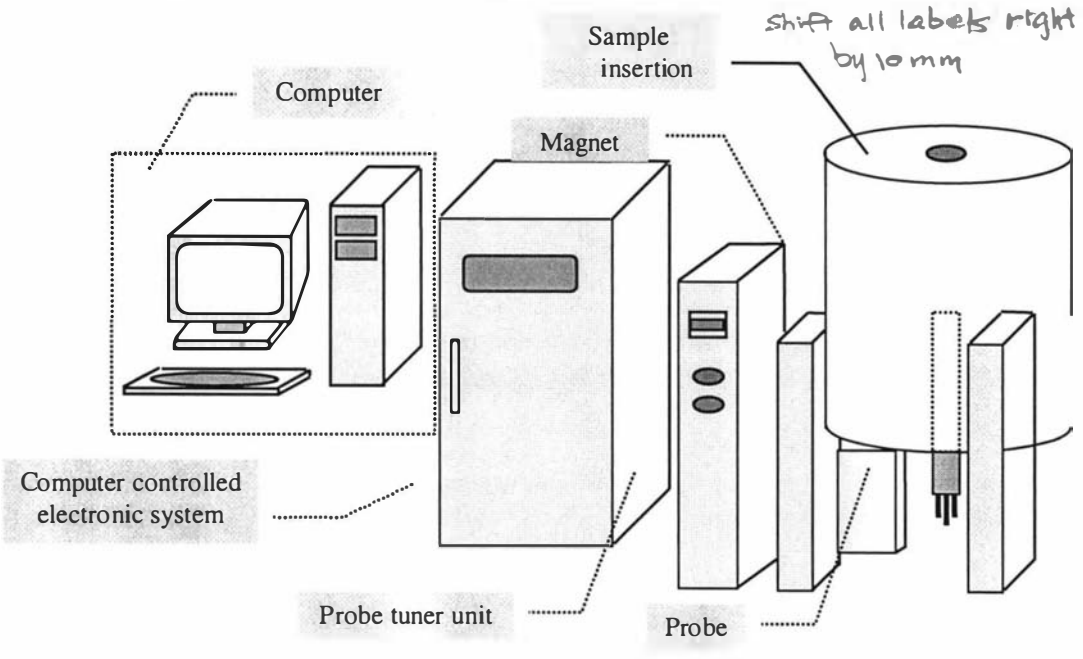


Figure 4.1: Basic units of an NMR spectrometer. The computer is a console to operate the spectrometer. The computer-controlled electronic system contains electronic devices to produce pulses, acquire data, adjust various parameters such as temperature, field gradients, decoupler and transmitter frequencies, etc. The probe tuner is to manually tune the observation and decoupling channels of the probe. The super-conductor magnet provides the main magnetic field B_0.

In brief, there are four basic units:

1. a host computer with an operating console,
2. a tower containing computer-controlled electronic system,
3. a super-conductor magnet, and
4. a sample probe with a tuner unit.

The electronic system consists of a number of subsystems, including

- a broadband transmitter that generates excitation pulses,
- a broadband decoupler that provides a second frequency source for pulsing or spin-spin couplings,
- a high-band amplifier and a low-band amplifier that boost the transmitter and decoupler outputs to higher levels needed for hard pulses[6],

[6] The high-band amplifier high Larmor frequency pulses such as those for ^{1}H and ^{19}F, and the low-band one is used for low frequency pulses such as those for ^{13}C and ^{15}N.

- a lock transmitter and a lock receiver to ensure that any drift in reference frequency is compensated so that the magnetic field experienced by the sample is constant throughout an experiment,
- an observation receiver that detects and digitises the free induction decay signal,
- a variable temperature (VT) controller that ensures a constant sample temperature throughout an experiment, and
- a shimming controller that sets a handful of shimming parameters to form a homogeneous field around the sample.

4.4.2 Setting up 1D protein NMR experiments

Unless a spectrometer is used exclusively for routine tasks, setting up an NMR experiment can be very time-consuming.

A typical step-by-step procedure for performing a 1D ^{1}H experiment could be as follows:

1. Load sample into the probe.
2. Set up sample temperature.
3. Tune sample probe.
4. Load a 1D pulse-sequence.
5. Set up experimental parameters.
6. Search for the best locking parameters.
7. Search for the best decoupler or saturation frequency.
8. Search for the best 90° pulse-width.
9. Shim for field homogeneity.
10. Search for best parameters for switching-on receiver and data-acquisition.
11. Acquire 1D data.
12. Repeat steps 6 to 11 if necessary.

In the following context of this section, some VNMR commands [Varian 1989-92b] are also presented for NMR researchers who have dealt with them.

Searching for the best lock parameters for the deuterium signal should not be hard. It is done with the number of transients (*nt*) set to one. If there is some difficulty in the process, check with the following factors in sequence: sample loading, probe tuning, temperature stability, lock gain, lock power and lock phase. Once lock signal is found, make sure that the lock power is not too high; otherwise the field stability may be reduced.

In terms of duration, 1D NMR pulse sequences can be typically divided into a maximum of three periods between two pulses, as shown in Figure 4.2. Sequentially, these are the preparation, mixing, and acquisition periods.

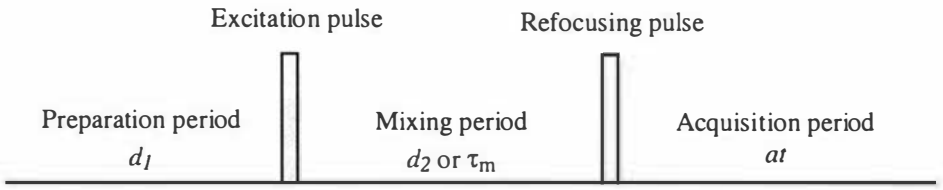


Figure 4.2: A typical 1D pulse-sequence. Various experiments may have different operations applied during each period. For example, the preparation period may simply consist of a delay to allow all spins to return to equilibrium. It may also contain an operation to suppress the signal of the sample solvent. The excitation pulse is usually a 90° pulse to tip the magnetisation vector to the transverse plane. The mixing period may contain operations that select a specific coherence to be observed, during which the magnetisation vector may not remain in the transverse plane. The refocusing pulse is usually a single pulse to tip the magnetisation vector back to the transverse plane for observation. The acquisition period is time during which the receiver is collecting data. Note that variations occur for different experiments [see, e.g., Cavanagh *et al.* 1996, or Frenkiel 1993].

Different parameters must be set up for each period. For example, a 1D NOESY pulse-sequence in protein NMR starts with around 200 milliseconds of preparation delay (*d1*) and then a few micro-seconds for an excitation pulse (*p1*). It is followed by a mixing delay (*d2*) that may vary from 100 to a few hundred milliseconds, and then continued with a few microseconds of refocusing pulse (*pw*). Finally, the process is finished with a few seconds of data acquisition time (*at*).

In addition, there are many other parameters such as the number of complex data points per FID (*np*), spectral width (*sw*), transmitter offset frequency (*tof*), decoupler offset frequency (*dof*), saturation power (*satpwr*), saturation delay (*satdly*), etc. For Varian spectrometers, the data acquisition time *at* is automatically set to *nc*/*sw*, where $nc = np/2$ is the number of complex points per FID.

The decoupler frequency (or saturation frequency, depending on experiment type) can be sought initially by setting an array of frequencies around the solvent's resonance frequency. This process is done repeatedly and the range of frequencies is gradually narrowed down until a satisfactory frequency is found.

The 90° pulse-width can be found by the null method, in which an array of *pw* values around 360° pulse-width is set, typically between 40 to 50 micro-seconds with the Varian 500. A graduated change of spectral peak intensity, from entirely negative to entirely positive, should be seen with the *pw* increased in steps. The 360° pulse can be obtained from the one that minimises the spectral intensity. This process is repeated with the range of *pw* values that are gradually narrowed down until a satisfactory value is found, which is then divided by four to get the 90° pulse-width.

Shimming, a technique used to improve the field homogeneity, is often the most difficult and time-consuming task. It searches for an optimal combination of field-gradient parameters that often interact with each other. The shimming parameters are divided into two groups: the axial and transverse gradients. The former are adjusted while a sample is spinning, and the latter, non-spinning. In practice, the aims are to minimise the underlying hump of a solvent peak as well as to eliminate any split in the peak. One may often achieve this through adjusting the shimming parameters to maximise the level of the lock signal or the trailing parts of single-scanned FIDs [Varian 1989-92]. The latter method is based on the fact that increasing the intensity of the trailing part of an FID will effectively narrow the peaks in the corresponding spectrum, thereby indicating an optimal condition. However, this method will work

well only if the differences in the trailing parts can be judged by eye. Often, the auto-shimming procedure of VNMR 4.1 can give a satisfactory result, though it may not be the best.

The next difficult task is to search for the best parameters for switching on the receiver and starting data-acquisition. In Varian VNMR 4.1, these are called *rof2* and *alfa*, both of them are in the order of microseconds. Here, *rof2* is a dead time after the final pulse *pw* with receiver off, and *alfa* is a delay before acquisition started with receiver on. If these two parameters are not set properly, the first few data points may be distorted, resulting in a baseline problem. Adjusting them is often difficult, and their values are also pulse-sequence dependent.

4.4.3 Setting up 2D protein NMR experiments

Setting up 2D protein NMR experiments requires considerable experience in order to obtain quality spectra with minimal effort.

A typical step-by-step procedure for performing a 2D ^{1}H experiment can be as follows:

1. Roughly set up the 1D experimental parameters by quickly following a procedure similar to those for 1D ^{1}H experiments.
2. Load a 2D pulse-sequence.
3. Set up experimental and acquisition parameters for collecting only one FID per experiment.
4. Repeat step 6 to 11 in the procedures for setting 1D ^{1}H experiment (as described previously, usually without spinning).
5. Set up experimental and acquisition parameters for collecting a set of FIDs that form a 2D NMR data set.
6. Start the 2D experiment.

Note that only a rough result is needed from step 1, because everything may change after a 2D pulse-sequence is loaded. Therefore, it is better to spend more time in optimising all parameters with a pulse-sequence loaded.

In terms of duration, a 2D NMR pulse sequence can be typically divided into four periods separated by three pulses. These are the preparation, evolution, mixing, and acquisition periods, as shown in Figure 4.3.

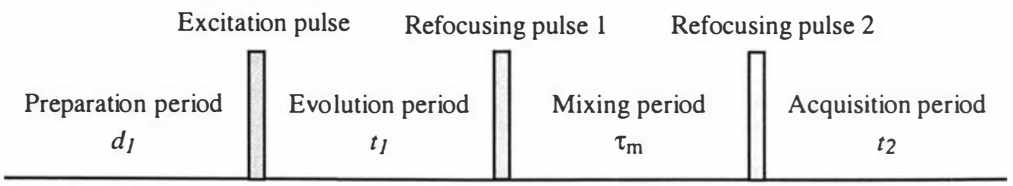


Figure 4.3: A typical 2D pulse-sequence. Here, the evolution period contains incremental delays (each for an FID) during which the nuclear spins are effected by the *J*-coupling Hamiltonian (see Chapter 3). The refocusing pulse 1 prepares the magnetisation for the mixing operation. The remaining operations are the same as those in a typical 1D pulse-sequence. Note that variations occur for different experiments [see. e.g., Cavanagh *et al.* 1996, or Frenkiel 1993].

Again, every parameter used in an experiment must be properly set up. For example, the saturation delay should be set between 1.5 to 2.0 times the longitudinal relaxation time T_1, which can be easily worked out using the famous inverse recovery experiment [see e.g., Günther 1995]. A severe first-order phase error may indicate an insufficient delay time, as will be explained in Chapter 6.

When setting up parameters, make sure that they work with the conditions required for the 2D experiment. For instance, shimming without spinning is important because there will be no spinning during the experiment. Also, acquisition with a certain number of transients is important, especially when phase cycling [see e.g., Derome 1987] is used in data acquisition. Even the number of points per FID may affect other parameters, such as *rof2* and *alfa*.

After a 2D experiment is started, the acquisition of the first few FIDs should be monitored for possible receiver-overflow error. Sometimes the receiver-overflow error happens after collecting about three-quarters of FIDs, because the intensity of solvent signals can increase due to a drift in its resonance frequency with time. This can be eliminated by setting a lower receiver gain and saturation power in an experiment. Shimming periodically every time after acquiring a number of FIDs in a 2D experiment did not give any better results from our experience, because it produced rows of noise parallel to the diagonal of 2D spectra.

It is perhaps a good idea to run a small number of experiments with d2 varied from 0 to t_{1max}, namely, *ni* / *sw*, where *ni* is the number of t_1 increments in the 2D experiment. For a protein sample, typically *ni* = 512 and *sw* = 5000. Therefore, t_{1max} = 0.1024 seconds, and the t_1 *increment* is 200 microseconds.

4.4.4 Temperature selection

For bio-molecules, the ranges of NMR sample temperature are often selected in between room temperature and body temperature. It is recommended that spectra be collected with two temperature values differing by 10° C or more in order to observe peaks that are covered by water or affected by water suppression, because the water resonance frequency is more temperature dependent.

In practice, a higher temperature is often used unless there is a major concern about sample stability. This is based on the fact that the secondary structures of molecules are more expanded and that the molecular motions are faster at higher temperature. The former is supposed to result in better dispersion of 2D spectra, and the latter, narrower line-shape.

4.4.5 Mixing-time selection

TOCSY mixing times

The mixing times (τ_m) for TOCSY experiments are usually around 70 to 110 milliseconds. A number of TOCSY NMR data sets should be collected with different mixing times. These data can be processed separately and their spectra can be co-

added to a single spectrum. Since TOCSY is used only as a reference in the spin system assignment, the single criterion is whether a large number of cross-peaks that correlate amide and side-chain protons can be observed.

Short mixing times give better results in the amide-alpha region of a spectrum, thereby providing a useful reference for evaluating a project. The number of peaks seen in this area must in theory be equal to or more than the number of residues in the peptide, depending on whether there are serine and/or histidine in the peptide, as will be explained in Chapter 7.

Longer mixing times gives better signals for long side-chain residues. This is useful for identifying residues like lysine, isoleucine and arginine.

NOESY mixing times

When a nuclear spin-system has a long correlation time (which is the case for macromolecules at high magnetic field), strong positive diagonal and cross-peaks can be seen in 2D NOESY. The *correlation time* (τ_c) is the duration of an isotropic random process that modulates the orientation of the inter-nuclear vector between two spins. It depends on the character of the motional process and the separation of the interacting spins. Faster motion and further separation both result in faster reorientation of interacting spins and hence a shorter correlation time.

The study of τ_m-dependence of the diagonal and off-diagonal peaks in a NOESY spectrum of an AB spin system can be seen in Macura and Ernst 1980. It shows that the intensity of cross-peaks increases sharply with small mixing time, quickly reaches a maximum value at about 50 ms, then gradually reduces with larger mixing times. Applied to macromolecules, the τ_m-dependence of 2D NOE cross-peak intensity (I_{AB}) can be expressed as follows: [Eq. (9.7.20), Ernst 1987]

$$I_{AB}(\tau_m) \propto \tau_c \tau_m (1 - \eta_{ext} \tau_m) / r^6 \qquad (4.4.5.1)$$

Here, η_{ext} is the *external relaxation factor* that takes into account the interactions with further spins. The term is effectively an *external relaxation rate.*

Kumar *et al.* (1981b) reported the τ_m-dependence of a few diagonal and cross-peaks in the 2D NOE spectra of BPTI. Most of the curves of the cross-peak intensities look like parabolas with maximum values in the range τ_m = 150 to 200 ms, which can be fitted reasonably well to (4.4.5.1). Using the same experimental data from BPTI, τ_c can be calculated as being between 10 and 30 microseconds, with η_{ext} between 0.2 and 0.3 %.

If the external relaxation is neglected, equation (4.3.5.1) becomes

$$I_{AB}(\tau_m) \propto \tau_c \tau_m / r^6 \qquad (4.4.5.2)$$

Note that this linear approximation is the relationship used in most structural calculations. Strictly speaking, it is only valid for the linear part of the intensity curve, e.g., with τ_m =100 ms for a protein like BPTI.

A mixing time that is too short cannot produce sufficient number of peaks for sequential assignment; whereas a mixing time that is too long can produce false peaks

resulting from spin diffusion. Christine Redfield suggested a looser criterion based on inspection [Redfield 1993].

In practise, the NOESY mixing times used are around 100 to 150 ms for small proteins [see, e.g., Clubb *et al.* 1991, or Redfield & Dobson 1988]. A set of NOESY spectra with a small number of transients and different mixing times ranging from 70 to 250 ms should be collected at the beginning. The number of peaks seen usually increase with longer mixing time — unless an extremely long mixing time has been used, in which case the external relaxation effect becomes significant.

From these spectra, a number of isolated cross-peaks are chosen for the assessment of the best mixing time for the sample. The criteria should ideally be a time that is at the end of the linear region of the intensity curve — this gives the strongest average intensities for conventional structure calculation.

Chapter 5:

NMR Software and PAW

5.1 Introduction

This chapter gives a brief review of the NMR data-processing and spectral-analysis software, followed by a description of PAW's functionality and some program development details.

5.2 A Review of NMR Software

Stimulated by the rapid development of computer hardware capacity and performance, there has been a rapid evolution in software engineering [see, e.g., Booch 1991, Aho & Ullman 1992, Pressman 1992] since the 1980's. This development has a significant impact to the development of NMR software packages, which often deal with large data sets, as well as very sophisticated, tedious and time-consuming operations. In addition, the appearance of many new techniques in NMR spectroscopy has also contributed to the remarkable development.

An incomplete list collected by Lundberg (1997) contains 29 programs for NMR data-processing and spectral-analysis data, as well as many others. A review shows that NMR software packages are becoming more user-friendly, powerful, versatile, and portable. In terms of how these achievements have been made, it can be summarised as follows:

a) The user-friendly feature is achieved by the application of the powerful window-programming technique.
b) The powerful functionality is achieved by the embodiment of the latest multidimensional data-processing and spectral-analysis methods.
c) The versatility is achieved by adding flexible keyboard-commands and a macro-programming language.
d) The portability is achieved by the use of a hardware-independent programming language.

5.2.1 Existing NMR data processing and analysis software

The list below briefly describes some of the better-known programs in the NMR community:

- CAMRA (Computer Aided Magnetic Resonance Assignment) [Gronwald *et al.* 1998] is designed for the residue-specific assignments of proteins. It consists of three units: ORB, CAPTURE and PROCESS. The ORB predicts NMR chemical shifts for unassigned proteins using two databases. The first database contains chemical shifts of previously assigned homologous proteins, and the second, the statistically derived chemical shifts that are categorised according to their residue, atom and secondary structure type. The CAPTURE generates a list of valid peaks from NMR spectra by filtering out noise peaks and other artefacts, and then separating the derived peak list into distinct spin systems. The PROCESS

combines the chemical shift predictions from ORB with the spin systems identified by CAPTURE to obtain residue specific assignments.

- EASY [Eccles *et al.* 1991] is designed for the spectral analysis of bio-macromolecular 2D NMR data. It provides a window-based environment for interactive spectral interpretation. The program includes a number of automated routines for peak picking, spin-system identification, polypeptide sequence-specific resonance assignment, and cross-peak integration. It has a database system to store all resulting parameter lists on disk. EASY accepts the frequency-domain data sets as input, and after a combination use of the automated and interactive routines it can yield a list of conformational constraints in the format required as input for the calculation of the 3D structure.

- Felix95 [Biosym/MSI 1995] is a set of commercial programs for NMR data processing, analysis and assignment. It is based on a powerful macro language that permits sophisticated and versatile processing.

- GIFA (General Iterative Fixed-point Algorithm) [Pons *et al.* 1996] is a program for processing and analysis of 1D and multidimensional NMR data sets. Various NMR data-file formats are supported, including those of Varian and Bruker (AM, AMX, DMX) spectrometers, ASCII format, Matlab files, etc. A macro language permits sophisticated processing. The interface is user-designable and programmable. The program currently runs on several UNIX X-Windows platforms (SGI, HP; also later Sun, and RS6000).

- NUTS is a NMR data-processing program (1D and 2D) developed by Acorn NMR for PCs and Macintosh. (The program also includes a 10-spin simulation routine and performs spectral comparison and simplex optimisation to match experimental spectra with simulated ones.)

- PRONTO (4D) (4D version of the PROtein Nmr TOol) [KjÆR *et al.* 1994] is an NMR spectral-analysis software developed at Carlsberg Laboratory, Denmark. It has four sets of functions: a system for displaying 1D to 4D data, a database for storing structural determination information, a set of tools for spectral analysis, and a report utility that includes also an input facility for distance and dihedral angle restraints for structure calculation programs.

- PROSA (processing algorithm) [Güntert *et al.* 1992] is an implementation of common NMR data processing steps, such as linear prediction, digital filtering, Fourier transformation, automatic phase-correction and baseline correction. It is designed for high efficiency by avoiding disk storage of intermediate data in an automated process without any graphics display, and enables all calculations to be performed in batch mode on a variety of different computer systems, including supercomputers.

- XEASY [Xia & Bartels 1994] is an interactive program to analyse 2D, 3D and 4D NMR spectra for 3D structure-determination of bio-macromolecules. It includes many of the 2D functions of EASY, such as peak picking, sequence-specific assignments, cross-peak assignments, cross-peak integration, and rate-constant determination for dynamic processes. The program is developed on the X-window

system with the Motif[1] library, and hence is portable on a wide range of UNIX workstations. It provides maximal computer support for spectral analysis while giving users a complete control over the final resonance assignments. XEASY uses a flexible display of "strips", namely, 2D regions of 3D or 4D NMR spectra. Its automated sorting routines can narrow down the selection of strips that need to be interactively considered in a particular assignment step. It can transfer resonance assignments between spectra of different types and dimensions, including projected and reduced-dimensionality triple-resonance experiments. The output is compatible with DIANA, a program to calculate 3D structure from NMR restraints by Güntert (1995).

5.3 An Overview of PAW

PAW is a software system that contains a collection of interrelated components that work together to provide utilities for NMR data processing and spectral analysis.

This section gives an overview of the software development details on PAW. The author apologises for the software engineering jargon used, which may not be familiar to NMR specialists.

5.3.1 Functionality

There are six utilities provided by PAW. Together they constitute the functionality of PAW, as illustrated in Figure 5.1.

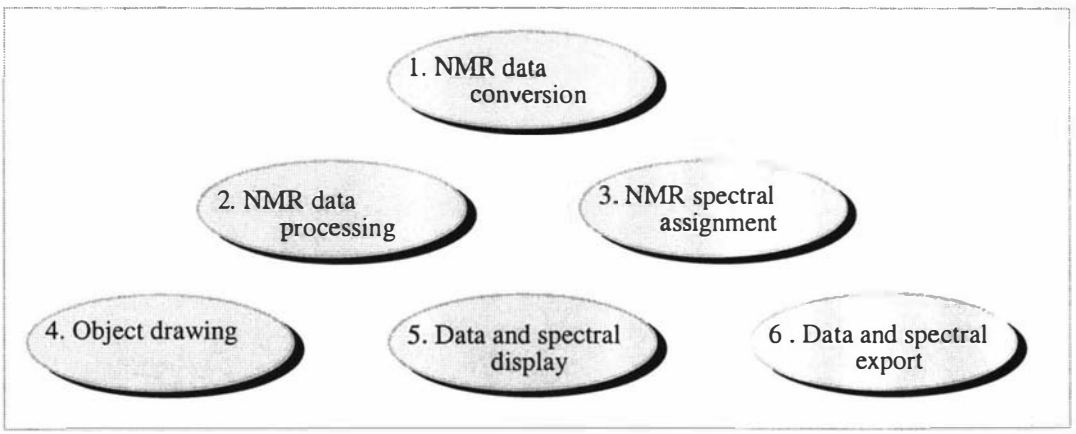


Figure 5.1: The functionality of PAW. The six utilities provide a system of tools for 1D and 2D NMR data processing and spectral assignment.

The data-conversion utility provides a tool to convert NMR data in other formats (such as Varian, Bruker, and others) into the format used by PAW. It also creates a processing-parameter macro for data processing and spectral display. (See also Volume II.)

[1] Motif is a portable language developed at MIT in the United States

The data-processing utility provides tools for processing 1D and 2D NMR data. This includes linear prediction, data filtering, water-signal suppression, Fourier transformation, phase correction, and baseline correction.

The spectral assignment utility provides tools for peak picking, assignments, labelling, and integration.

The object-drawing utility provides tools for drawing lines, rectangles and text strings.

The data and spectral display utility provides tools for displaying 1D and 2D NMR data, as well as peak symbols, labels and connectivities. This includes a variety of options for spectral zooming and display, as well as those for selective cross-peak and label display.

The data and spectral export utility provides tools for storing and exporting various types of data and processed spectra. These include parameters for spectral processing and display, sorted peak lists, NOE summary plots of assigned cross-peaks, 1D line-plots, and 2D contour plots (with or without labelled cross-peaks and annotations).

5.3.2 Development model

The software development model for PAW is shown in Figure 5.2.

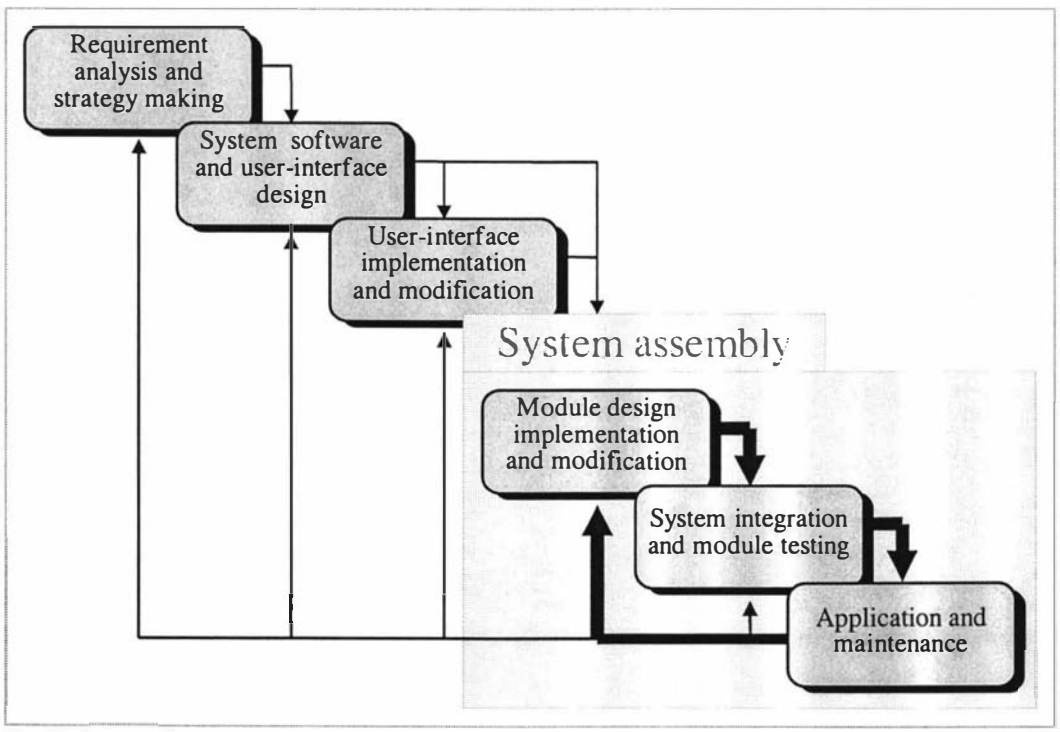


Figure 5.2: The software development model for PAW. In short, it is a combination of the traditional waterfall model [Royce 1970] and the contemporary system-assembly model with reuse-oriented program components [Martin 1993]. The much thicker arrow within the system-assembly block indicates the higher frequency of the programming activity in the process.

Further details on each of the sub-processes are given in the next few sub-sections.

5.3.3 Project constraints and objectives

The constraints and objectives are illustrated in Figure 5.

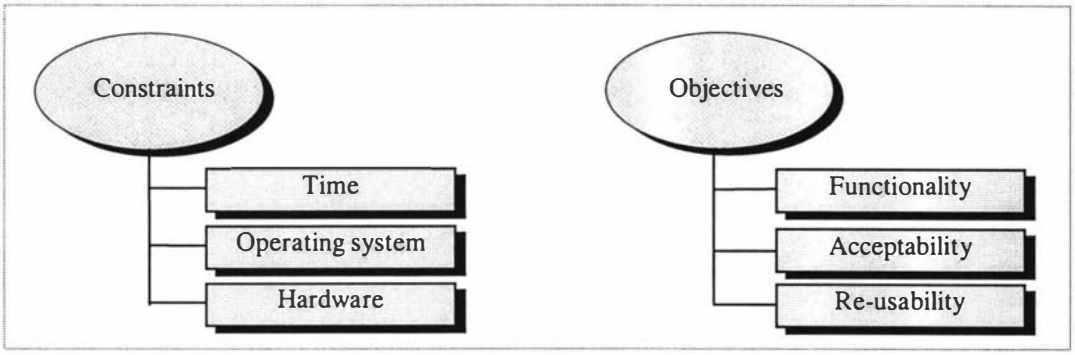


Figure 5.3: Project constraints and objectives.

The three aspects of constraints for the project are the time, operating system, and hardware:

1. It was proposed that the project could only address two-dimensional problems because of the available development time.
2. The *X Window Operating System* (or simply X) was chosen as a platform, because X was at that time the most popular network-based windowing system that was independent of hardware and can be run on a remote computer. It was developed by MIT and the *Digital Equipment Corporation* and has been adopted by the computer industry as a standard for computer graphic applications.
3. The SGI computers were chosen to be the hardware because of their popularity among researchers working in this field and their profound graphic performance at that time (in comparison with PC and Macintosh).

The three objectives for the project are functionality, acceptability, and re-usability:

1. The functionality was initially set to be a 2D protein-NMR spectral assignment package similar to EASY. This was later expanded to include NMR data processing, which turned out to be a complex task and hence required additional time for program development and documentation.
2. The acceptability was initially aimed at developing a user-friendly package driven by menus and commands. This macro interface was added later, mainly as an aid to multidimensional data processing.
3. The re-usability feature aided in the cyclic assembly of the various modules, as will be explained later.

5.3.4 Development Strategy

The initially software development strategy is shown in Figure 5.4.

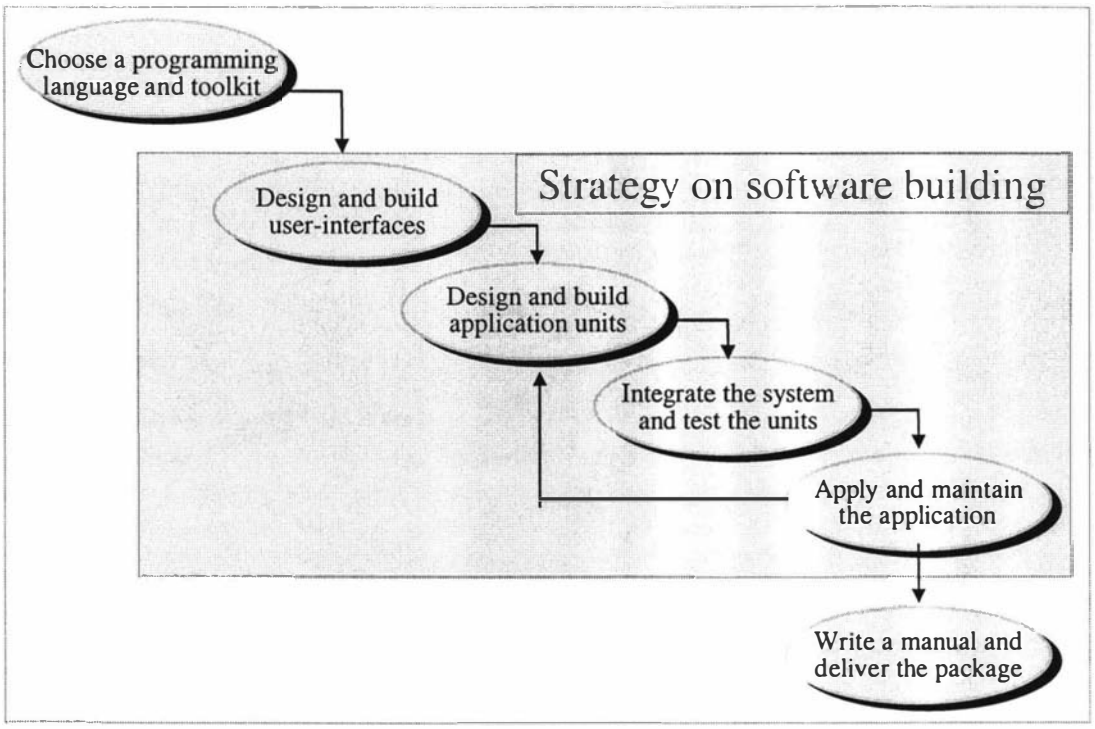


Figure 5.4: Initial software development strategy for PAW. It leads to the actual development model described in Section 5.3.2. Note that the term 'units' was loosely used here instead of 'modules' — this was because they were not planned to be modules at the beginning. Examining the differences between Figure 5.2 and this shows the relationship between the original plan and the actual development process.

The programming language for PAW was selected to be the standard C due to the lack of familiarity with C++.[2]

The strategy schedules the design and implementation of user-interfaces ahead of other software building processes, because the functionality that the software package provides often relies heavily on the interaction between users and the program. For the standard graphic user-interface, the *Motif toolkit* (or simply Motif)[3] from the Open Software Foundation (OSF) was chosen.

Once the user-interfaces were built, the process then moved on to a cyclic process of designing and implementing individual application units, integrating the units into the entire system, and then repeatedly testing the application with NMR spectroscopy data.

---

[2] An attempt to convert the program into C++ had seemed to be very encouraging at the beginning. About 30 thousand lines of the program had been converted and compiled. However, the author had to give up in the end because the C++ compiler available to him did not properly support the variable argument-list commands that allow a variable number of parameters to be transferred when calling a routine.

[3] Motif is based on the *X Toolkit Intrinsics Library (Xt)* — a standard on which many of the toolkits written for X are based. *Xt* is a standard X toolkit established by the *X Consortium* that provides an object-oriented programming framework in the C language for creating reusable and configurable user-interface objects called *widgets*. Motif provides many widgets for common user-interface, such as buttons, menus, dialog boxes etc. Collectively, Xt and Motif (or any other Motif-like products) are called the *X toolkit*.

There are two design-and-build processes in the strategy (see Figure 5.4), because it was not intended to put everything in place right from the beginning. This has proved to be a practicable and efficient strategy for a big project, because there were just too many things to be put together in a single design process and it was hard to predict how much could be done in the available time. Indeed, the actual development process turned out to be quite unexpected. For example, the design and implementation for the menu- and toolbox-interface was done at a much earlier stage than those for the macro- and hotkey-interface, forcing the development process to branch along the path for system integration and module testing.

5.3.5 System and software design

Being a very large event-driven X-Window program, PAW is a complex entity developed with different design approaches for various parts of the system. In fact, the design of PAW had evolved with time. It started from a function-oriented design with a single menu-event-driven system, and finished with module- and object-oriented designs with multiple event-driven systems. In addition, PAW's architectural design is a dynamic process that includes system structuring, control-hierarchy modelling and module decomposition.

5.3.5.1 System structuring

Since PAW is designed to efficiently handle extremely large data sets, the system structuring follows a repository model [see, e.g., Sommerville, 1995], as illustrated by Figure 5.5.

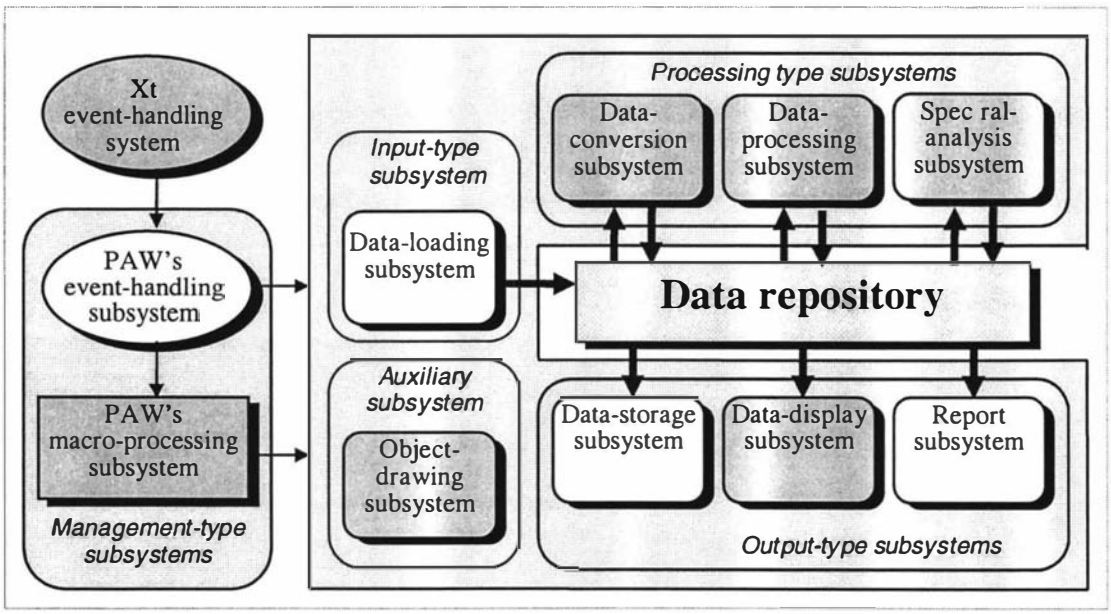


Figure 5.5: System structuring for PAW. It is a repository model based on a shared database as a repository that can be accessed by all subsystems. The model provides an efficient way of sharing large amounts of data. Each of the subsystems is a stand-alone processor and there is no need to transfer data explicitly from one subsystem to another.

A total of ten subsystems are available, and they are divided into five types of subsystems enclosed in the round boxes, as described in this subsection.

There are two management-type subsystems: the event-handling and macro-processing subsystems. Only the data-loading subsystem belongs to the input type. The data-conversion, data-processing and spectral-analysis subsystems belong to the processing-type. The data-storage, data-display and report subsystems belong to the output type. Finally, the object-drawing subsystem is an auxiliary system that has nothing to do with data repository.

The repository consists of a number of *buffers* (i.e., float-type arrays for storing spectral data) and structured data (such as the peak lists) that are shared by the seven subsystems. These data are loaded into computer memory and kept as public variables that can be modified by the processing-type subsystems.

Any keyboard- or mouse-events are initially processed by the Xt event-handling system that calls a registered function (called the '*callback*' in Motif programming) or an action routine in PAW's event-handing subsystem, which in turn calls another routine(s) in other subsystems to access the repository.

Note that a thin arrow indicates a routine- or function-call operation, and a thick arrow, a data-accessing operation. An arrow pointing towards the repository indicates a possible modification to the data, and those pointing from the repository indicates an access to the data without modification.

In addition, a subsystem in PAW may not be encoded as an entity but as a group of similar routines that may appear at different places in the program. In other words, a number of routines having similar functionality are grouped together as a subsystem for the purpose of illustration only. The four subsystems like this are highlighted in Figure 5.5. For example, PAW's event-handling routines are scattered all over the program. They are included in each dialog and module (except the macro-processing module).

Figure 5.6 shows the contents of the event-handling subsystem in PAW.

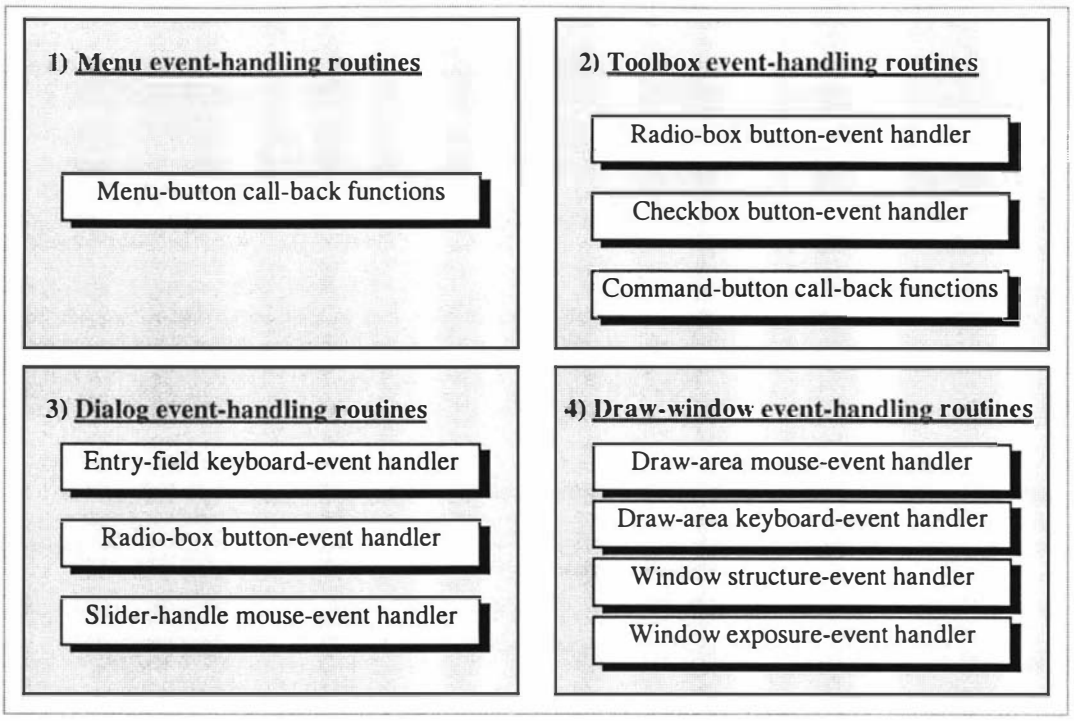


Figure 5.6: PAW's event-handling subsystem. It consists of a group of similar routines that are scattered all over the program. The mechanism and advantage of this model of scattered subsystems will be seen clearly in the next two sub-sections.

5.3.5.2 Modules

Four of PAW's major subsystems are decomposed into a total of thirteen modules called *toolboxes*. (See Volume II for details.) Each of the toolboxes are designed to implement and maintain a number of push-buttons, which may include radio-box buttons, checkbox buttons, and command buttons. Figure 5.7 is an example illustrating the *Common-display Toolbox* of PAW.

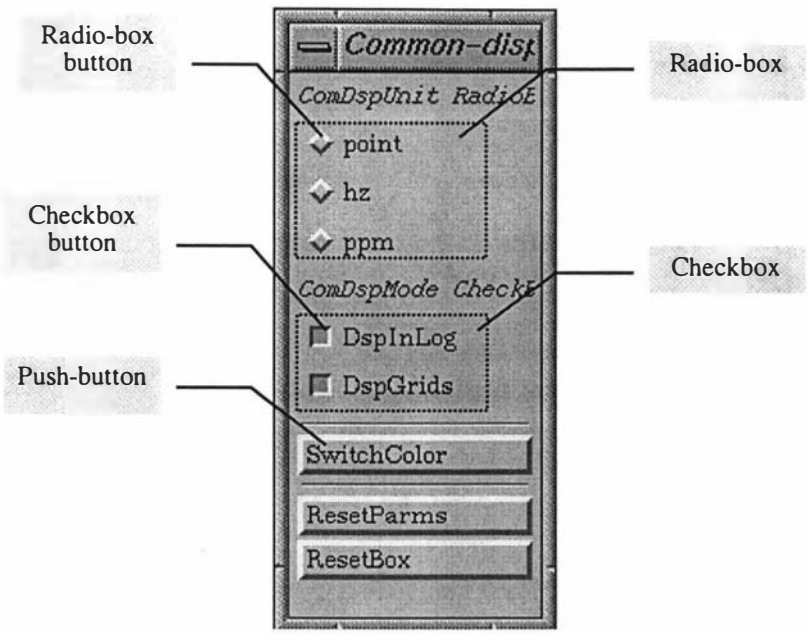


Figure 5.7: A toolbox that contains various types of push buttons. Buttons in a radio-box specify an exclusive OR-condition, while those in a checkbox specified an AND-condition. The push buttons are used to perform various commands.

Figure 5.8 shows the toolbox modules within PAW.

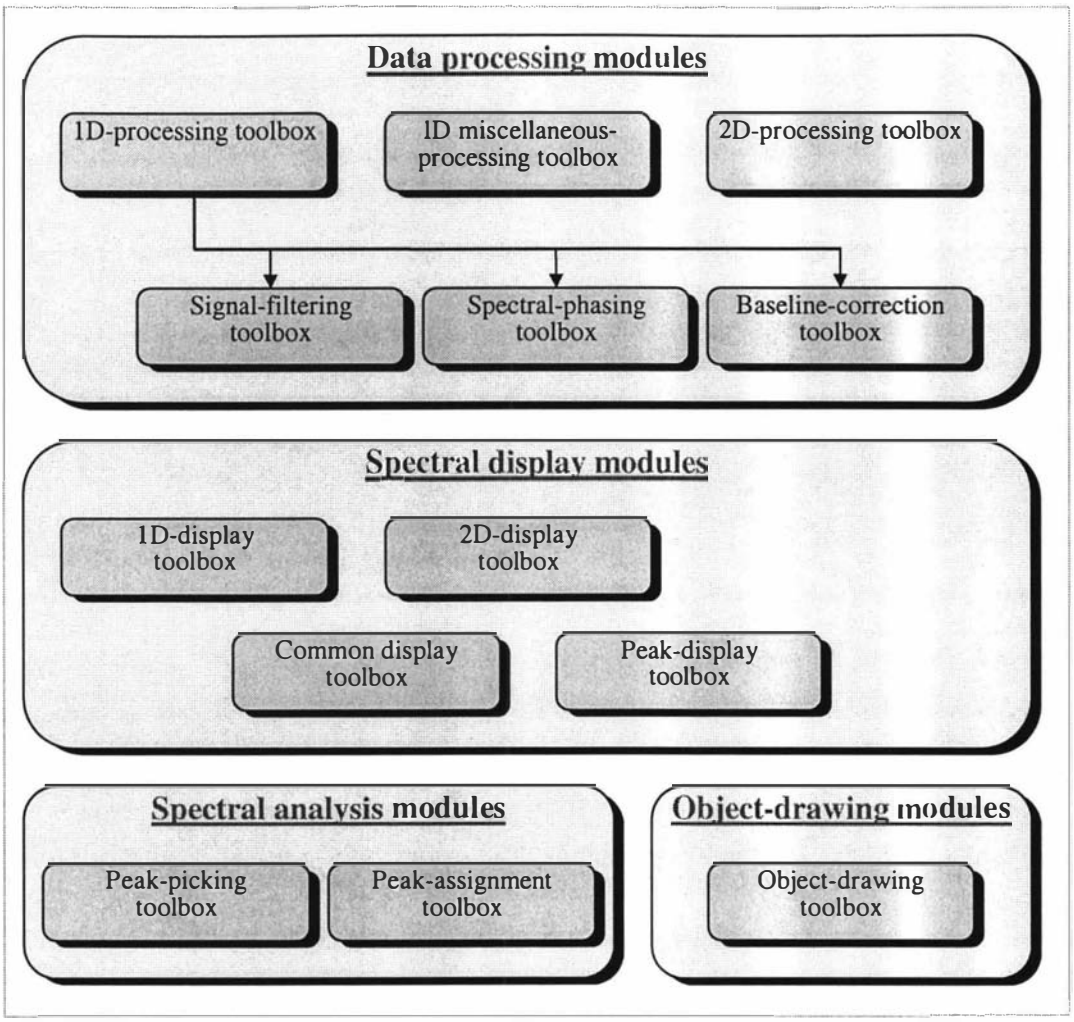


Figure 5.8: The modules within PAW. Each model produces and maintains a toolbox that provides utilities for various tasks.

As can be seen from the diagram, there are six modules for data processing, four for data display, two for spectral analysis, and one for object drawing.

There are three general features of the toolbox modules. First, they access data independently (except for the object-drawing toolbox module, which does not access data). Second, there is no structural dependency between modules. Thirdly, they are generically similar, thereby allowing a toolbox to be created with minimal effort. The last two features are achieved by the object-oriented design of the toolbox modules, as explained later.

The three modules arrowed (see Figure 5.8) can be invoked by pressing buttons in the *1D-processing Toolbox*, thereby providing extra convenience to users. These buttons do not have to be there, because PAW also provides other ways to invoke them, including keyboard commands, user-definable menu buttons and hotkey accesses.

5.3.5.3 Widget family

In Motif programming [Heller & Ferguson 1991], *widgets* are reusable and configurable objects. They are grouped in different *object classes* according to their structural variables and functions. Many of them are provided by Motif as standard user-interface objects that are pre-built with a number of variables and functions in their structure. (See also section 5.3.~~3.5~~4 for more information.)

One of the main features in object-oriented programming is the *inheritance hierarchy of object* classes. The hierarchy organises widgets into different levels. The relationship between any two adjacent widget levels is termed parent and child. Accordingly, a *parent widget* can have a number of child widgets; and a child widget inherits all re-definable object variables and object functions from their parent's structure in addition to their own variables and functions. Figure 5.9 shows the family tree of PAW's mainstream widgets, where the ones in grey are widgets that possess at least one child widget. (For simplicity, there are still many less-important widgets that are not shown in the diagram.)

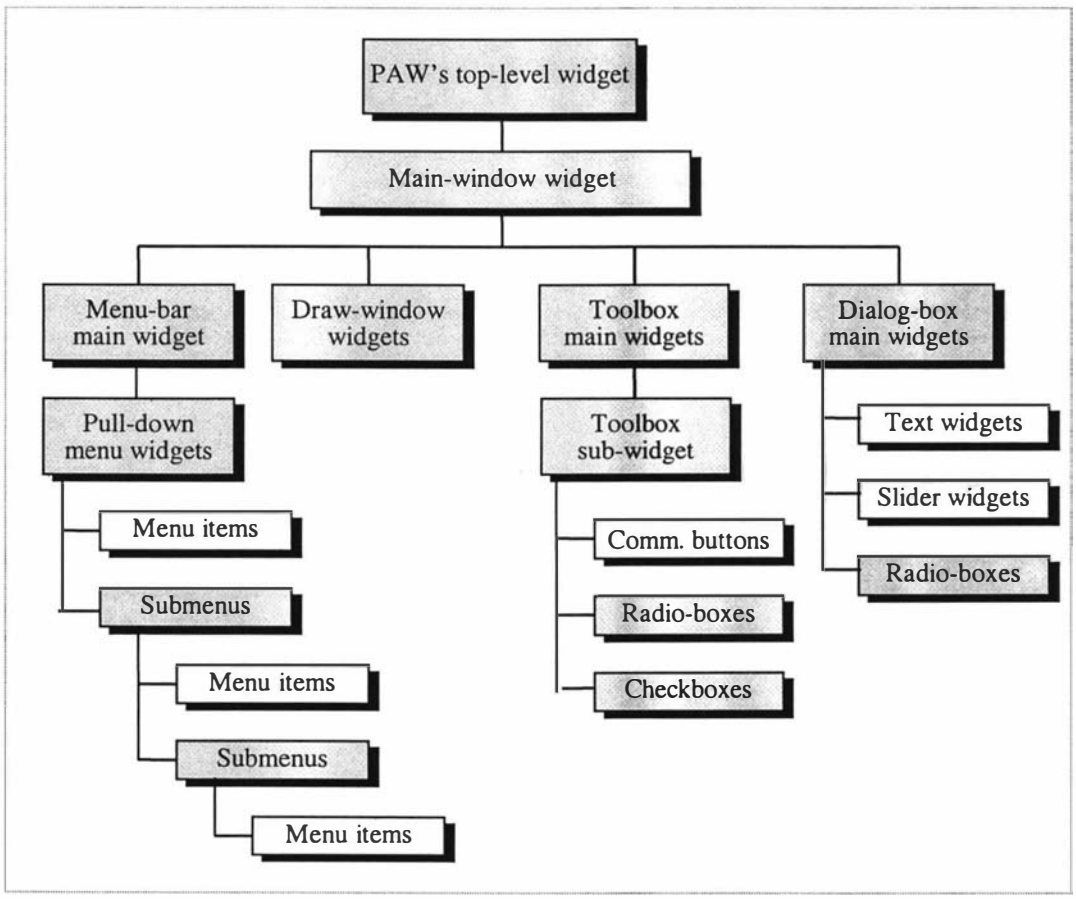


Figure 5.9: The widget family tree in PAW. In Motif programming, widgets are reusable and configurable objects [see, e.g., Heller 1991]. The family-tree diagram indicates a parent-child relationship in widget structure by a line joining two widgets.

In the diagram, the *top-level widget* is typically the highest-level parent widget for all other widgets created in an Xt application, and is designed to interact directly with the window manager.

The *main-window widget* is the only child of the top-level widget and the parent of the three main widgets and the draw-window widgets. On the screen, this is a window that provides a canvas for the entire application.

The *menu-bar* main widget is the parent of all pull-down-menu widgets and grandparent of the menu and sub-menu items.

A *draw-window* widget provides an area for drawings. Up to nine draw-windows can be opened at one time.

A *toolbox* main widget is the parent of a toolbox sub-widget that in turn is the parent of a number of command buttons and other widgets.

A *dialog* may contain a number of text-entry widgets, slider widgets, and at least one radio-box widget for execution and cancellation (see Figure 5.7). This is true for all three types of dialogs in PAW, including the keyboard-entry, dual-entry, and file-selection dialogs. (See Chapter 2 of Volume II for more details.)

The widgets for the top level, the main window and menu-bar are built as soon as the application starts. Other widgets are built dynamically.

Note that a child-widget only inherits certain behaviour and all configurable parameters from its parent and ancestor widgets. The management of widgets, however, has nothing to do with the structural inheritance in most cases, except when a widget is destroyed or unmanaged, as will be explained next.

5.3.5.4 Control model

To work as a system, a proper control model must be set up so that operations can be performed correctly.

As can be seen from information given in the last sub-section, the skeleton of PAW is composed of widgets that are directly managed by both Xt and PAW's event-handling routines. Like other Motif applications, each of the event-handling routines in PAW is registered as a pointer stored either in a *translation table* or *call-back resource* for a specific widget event.

Over all, PAW has an event-based control system. The way that Xt controls PAW is illustrated in Figure 5.10.

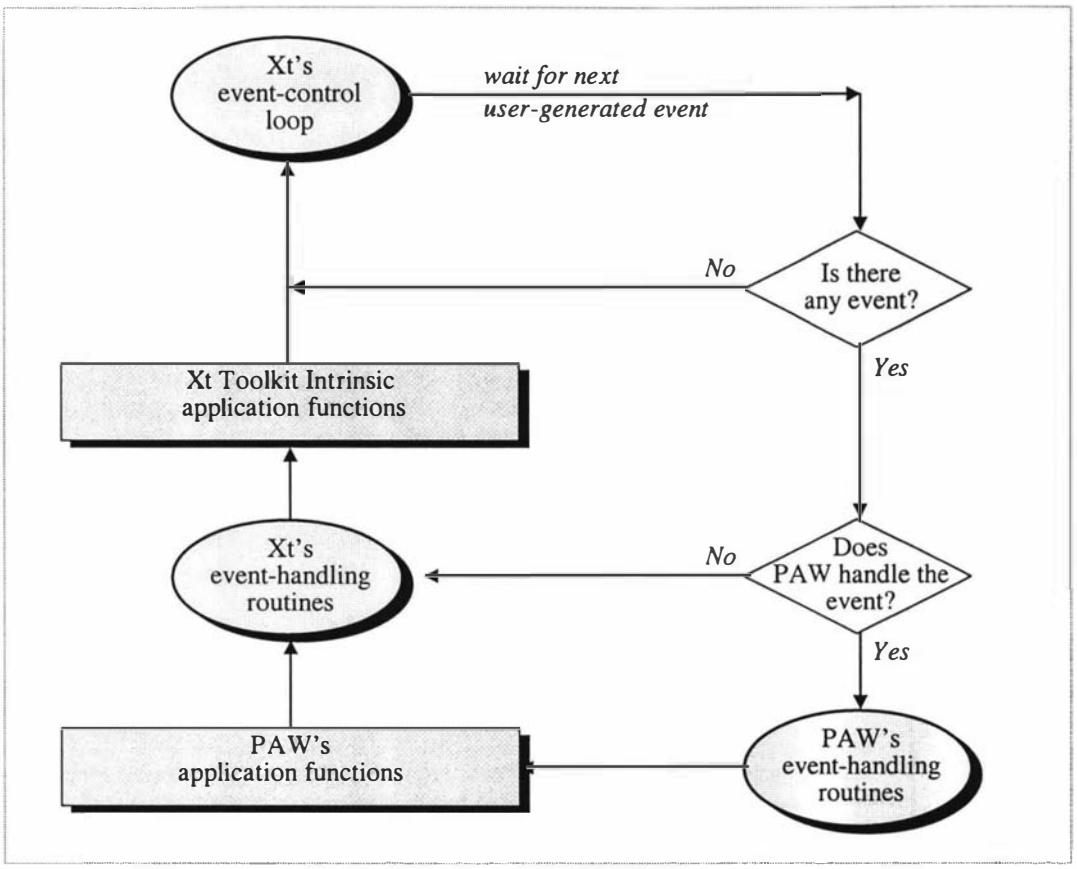


Figure 5.10: A flow chart that shows how Xt handles PAW's events. Xt's event-control loop detects all mouse- and keyboard-events and dispatches them to the event-handling routine of an appropriate widget from which the event is initiated. If PAW has been set up to handle the event, it will be passed to PAW's registered event-handling routine. Otherwise, the event will be handled by Xt's event-handling routines.

In general, PAW only handles those events that are relevant to its own application, such as changes in size and position of the toolboxes and draw-windows, redrawing a newly exposed draw-area, popping up a dialog box for an operation, executing a command when a command-button is pressed, etc.

Figure 5.11 and 5.12 show how Xt dispatches the mouse- and keyboard-events for PAW's widgets. Also illustrated in the figures are PAW's routine passing pathways. The rectangles in grey are PAW's application functions; the rectangles in white are PAW's event-handling routines; and the ovals in grey (both small and big) are Xt's event-handling routines.

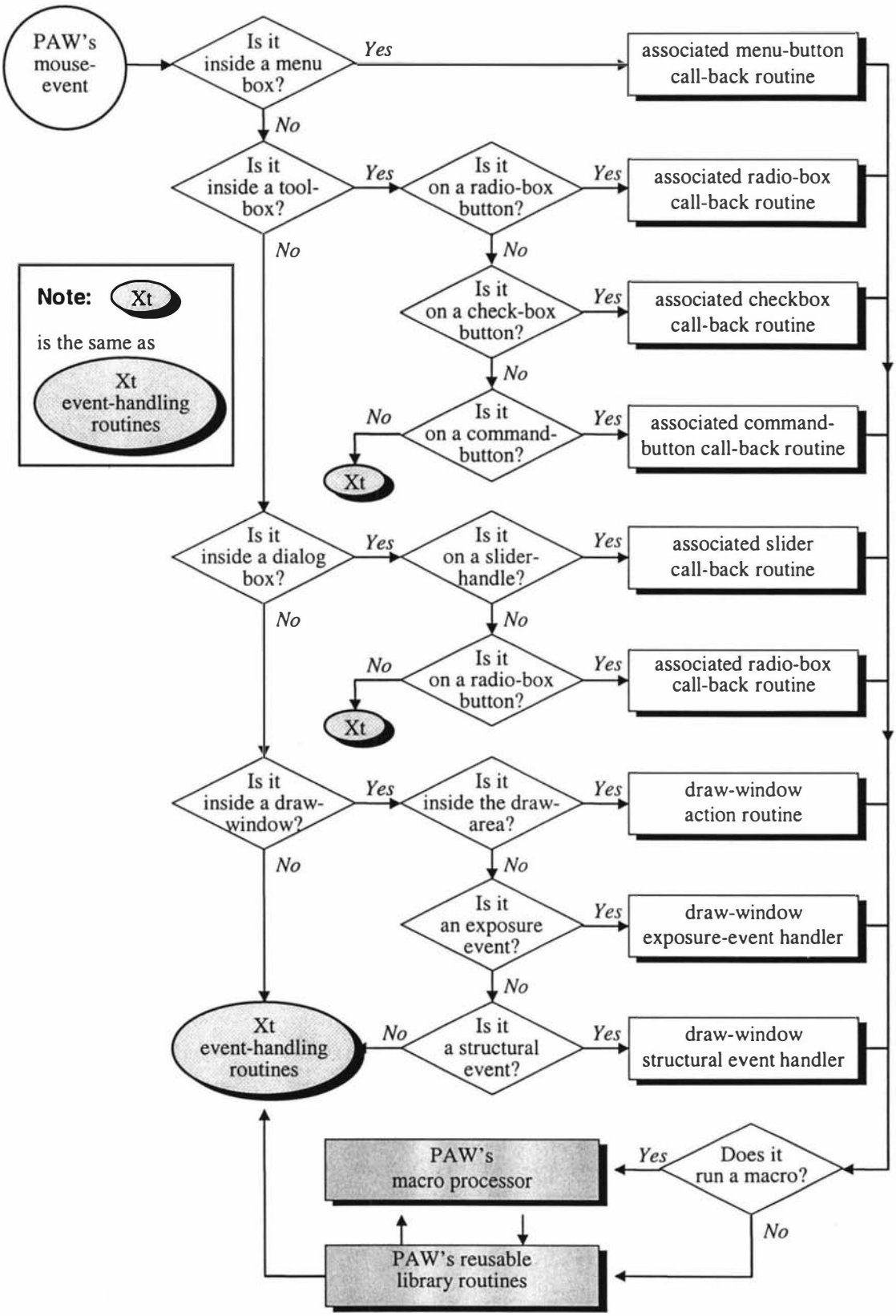


Figure 5.11: Xt's event-handling procedure for a PAW-related mouse event. The rectangles in white are either PAW's call-back functions or event-handling routines, the rectangles in grey are PAW's application functions, and the ovals in grey (small or big) are Xt's event-handling routines.

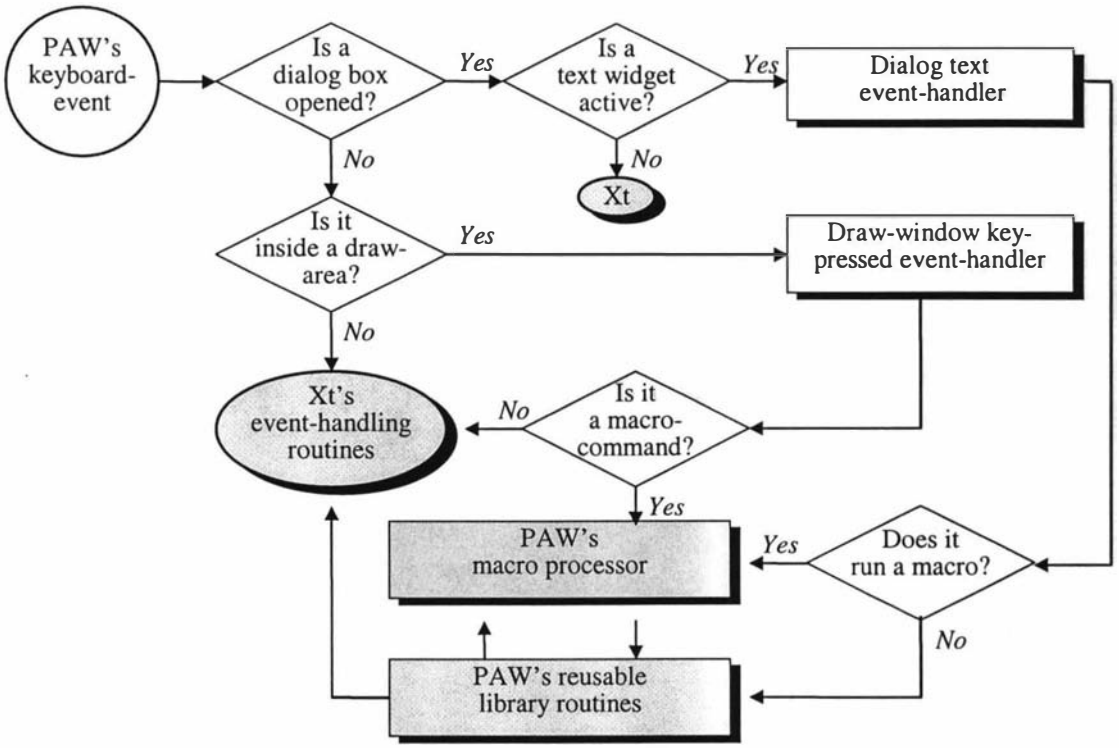


Figure 5.12: Xt's event-handling procedure for a PAW-related keyboard event. The rectangles in white are either PAW's call-back functions or event-handling routines, the rectangles in grey are PAW's application functions, and the ovals in grey (small or big) are Xt's event-handling routines.

Note that the event-handling routines are often called the 'action routines' in the Motif programming manual. In PAW, however, the call-back routines, action routines, exposure-event handlers, and structure-event handlers are all at the same control-level. These routines and event handlers are the intermediate components that receive calls from Xt and dispatch them to PAW's macro processor and reusable library routines.

The *macro processor* is an interpreter, which contains routines that process macro files and keyboard commands. Users can access the macro processor in two ways: either clicking on a menu button or entering a macro command while the mouse cursor is within a draw-window, as shown in the last two figures.

PAW's *reusable library routines* include those for spectral processing and analysis, such as signal filtering, phase correction, baseline correction, peak picking, etc. Some of the routines also perform macro commands. Together, they form a core that provides PAW's functionality. Other application routines, including toolbox builders, event handlers and macro processor, are included to support PAW's functionality.

In short, PAW's software structure is about the hierarchy of widget management and event handling with a variety of intermediate program components that dispatch various tasks to PAW's reusable library routines.

5.3.6 User-interface design and implementation

5.3.6.1 Available user-interfaces

To achieve maximum flexibility, PAW is equipped with a number of different user-interfaces. Figure 5.13 shows the available user-interfaces in PAW.

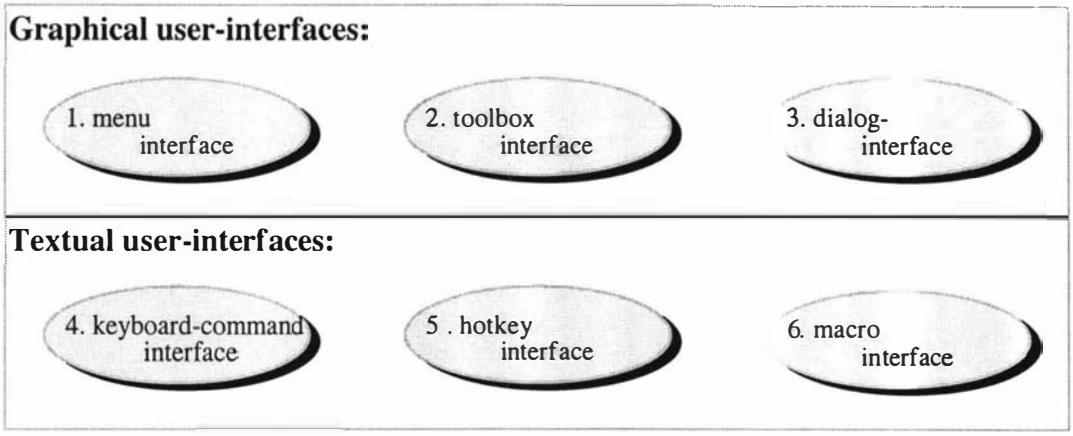


Figure 5.13: The user-interface methods available in PAW. They can be divided into two types: *graphical user interface* (*GUI*) and *textual user interface* (*TUI*) methods. Details on how each user-interface method works are very straightforward and can be found in Volume II.

5.3.6.2 Interface design and consideration

The first principle in the interface design takes into account the needs and capability of users at all levels, which include experienced or inexperienced, regular or casual, professional or non-professional. For this reason, a variety of user-interfaces are provided so that the users are not forced to accept only one user-interface. This requires that considerable effort be put into a series of design processes for a software package that is developed entirely from scratch.

As can be anticipated, most NMR spectroscopy software users do have their preferred interface methods in practice. Figure 5.14 shows from the acceptability point of view how various interface methods are designed for different users.

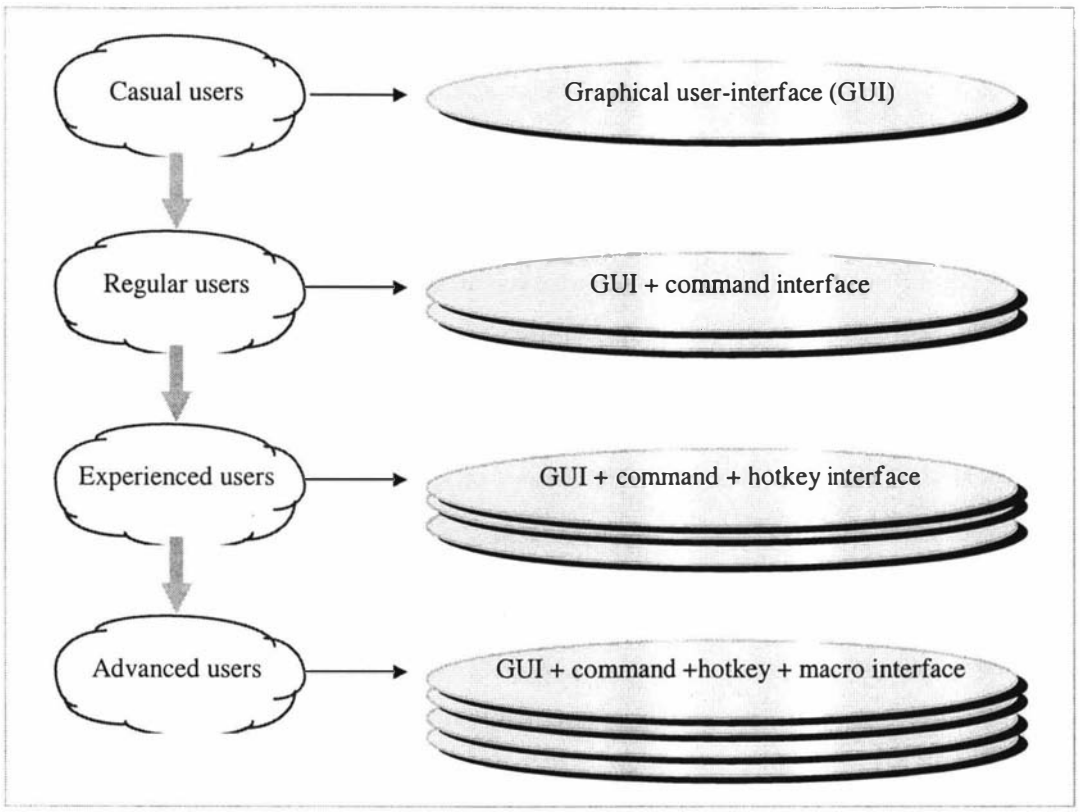


Figure 5.14: The user-oriented interface-design for PAW. Casual users are expected to use the GUI only. As they become regular users, typing simple commands rather than searching for buttons to click on can often increase the operation speed. Experienced users may like to define their own set of hotkeys to reduce the need for multi-key typing. Once becoming advanced users, they can then write macros for various processing tasks and make full use of the user interface methods provided by PAW.

Another principle in the interface design is to make it easy to learn, operate and be customised. The implementation of user-interface methods in PAW is mostly based on this consideration. The following two figures show the design consideration for the graphic user-interface (GUI) and text user-interface (TUI) methods in PAW.

User activities	Design consideration
Learning	—— Users who have no experience with the package can learn to operate it without training or with minimum of training or reading.
	—— Labels for widget items are relatively clear. Interface methods are consistent for comparable operations to reduce learning time.
	—— Online help is relatively easy to get.
Operating	—— Allow multiple draw windows and toolboxes to open separately from the menu.
	—— Allow users to select any pre-set window and toolbox layouts for performing specific tasks.
	—— Provide some mechanism that allows users to recover from their unwanted processing results or casual operational mistakes.
	—— Provide warning for potentially destructive operations.
Interface design	—— Allow users to design and build their own menus.
	—— Allow users to design and build their own window and toolbox layouts.

Figure 5.15: The design consideration for the GUI of PAW. It facilitates a Motif style windowing navigation model that is designed mainly for casual users, but not necessary beginners. There is little learning effort needed for anyone familiar with window applications.

User activities	Design consideration
Learning	—— Users can learn to utilise the TUI systems with little reading and practice.
	—— Each keyboard command is identical to a corresponding macro command to avoid excessive learning effort.
	—— Commands are named using a regular method. Some commands are identical or similar to those in some existing NMR software packages.
	—— The macro language is relatively easy to understand with a small number of simple rules. The rules are also consistent for similar tasks.
Operating	—— All keyboard commands are only two- to three-characters long to avoid lengthy typing, and require no parameters.
	—— Allow hotkey access for all commands, with or without parameters.
	—— Macros are relatively short with some clear examples to follow. A number of macros can be embedded into a macro.
	—— Provide warning of invalid commands and a switching mechanism to correct casual typing mistakes.
Interface design	— Users can define their own hotkeys with little effort.

Figure 5.16: The design consideration for the TUI of PAW. It is is designed for experienced and advanced users who prefer typing and/or programming for faster and more versatile multi-dimensional processing than the GUI that depends largely on mouse operations.

5.3.7 Module design and implementation

5.3.7.1 Object-oriented design (OOD)

An *object* in programming is an entity that possesses data as well as functions to manipulate the data. It can be created with any appropriate programming technique using many different languages such as C and C++ [see, e.g., p.13, Martin 1993]. The

aim of object oriented analysis and design is to achieve massive reusability of object classes.

In PAW, the object-oriented design is powered by Motif widget-class libraries. In addition to making use of the Motif widget-class, PAW also has its own objects that possess and manage a number of Motif widget-objects. Also, PAW has a number of other objects for frequently used tools such as the keyboard-entry box, slider-entry box, file-selection box and draw-window. Figure 5.17 shows the Window Object and an example of a typical toolbox object.

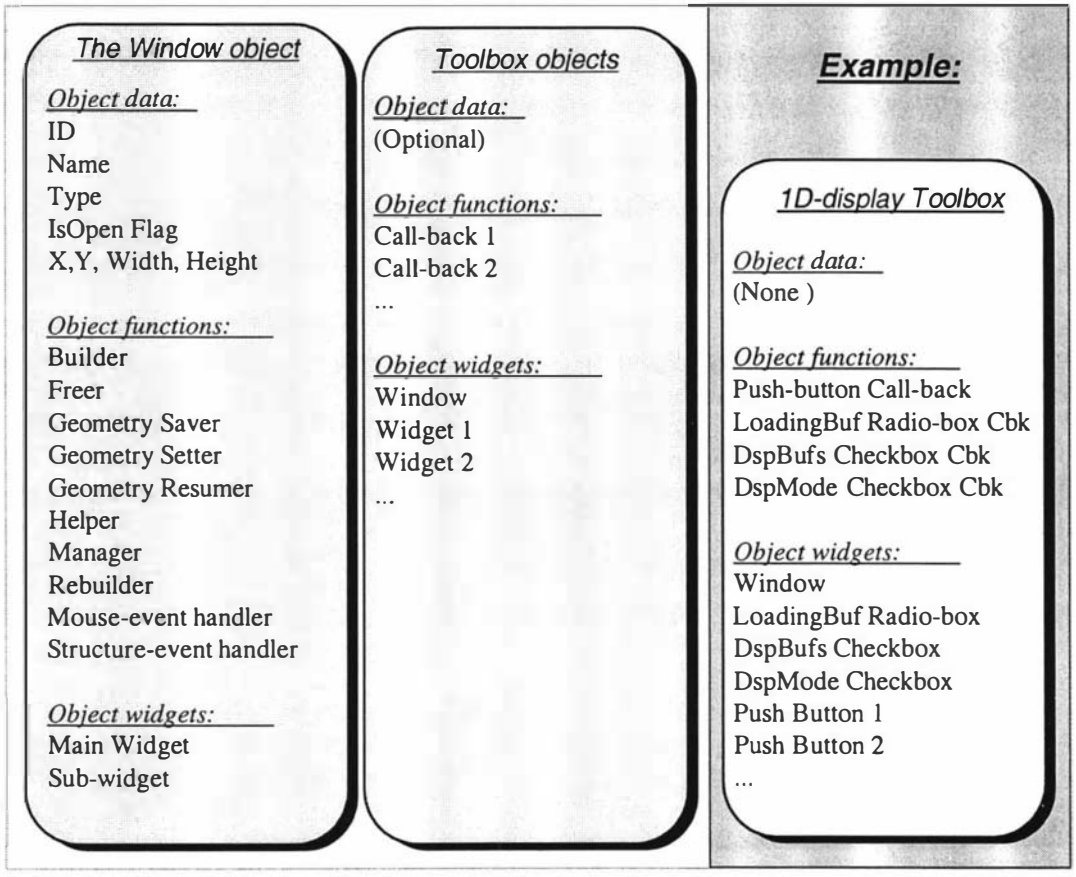


Figure 5.17: A typical object structure of the toolboxes in PAW. All toolbox objects (middle) contain a Window Object (left), thereby inheriting the object data, functions and widgets from the Window object. Additional functions to perform various operations and widgets to construct the toolbox are usually required, and the additional data are optional. An example is given on the right.

The *Window Object* in PAW is a parent-object for any objects built with a Motif window[4]. It contains a set of parameters to keep basic data (such as the window ID, name, type, etc.) and functions to build the window (such as the window builder, destroyer (or 'freer'), geometric-parameter saver, etc.). In addition, there are two widgets in the object. The main widget provides a Motif window in which all other widgets of a module are located. The sub-widget is the only child of the Main Widget. It contains a matrix of objects (called the xmRowColumnWidgetClass), and is a parent for all other widgets in a toolbox.

[4] This is often called a shell in Motif.

In PAW, noun-type names are used for object-functions rather than verb-type. For the Window Object, the roles of the object functions are as follows:

- the Builder builds the window;
- the Freer destroys the window and frees up the memory;
- the Geometry Saver saves the geometric parameters of the window;
- the Geometry Setter sets the geometric parameters of the window;
- the Geometry Resumer resumes the window to its original size and location;
- the Helper provides the help message about the window;
- the Manager manages the window;
- the Mouse-event Handler handles any mouse-events related to the window; and
- the Structure-event Handler handles any geometry changing-events related to the window.

Every toolbox object contains at least a number of dynamically built command-type push-buttons (or *command buttons*) that are managed by a *single* registered call-back function, which in turn calls various routines to perform different tasks. This design allows the program to handle only one call-back function for each toolbox, thereby avoiding excessive object structure referencing and maintenance, which are two of the disadvantages when programming with object-oriented approach [see, e.g., p.239, Sommerville 1995].

A toolbox may contain a number of radio-boxes and checkboxes. Each of them is also managed by registering a call-back function. For example, the *1D-display Toolbox* contains three extra widgets: the Loading-buffer Radio-box, Display-buffer Checkbox and Display-mode Checkbox. Correspondingly, there are three registered call-back functions, as shown in Figure 5.17.

5.3.7.2 Routine structuring in PAW's modules

A module consists of a number of routines that are arranged in a structure to perform various tasks. Figure 5.18 shows the routine-structure of the *1D-processing Toolbox*.

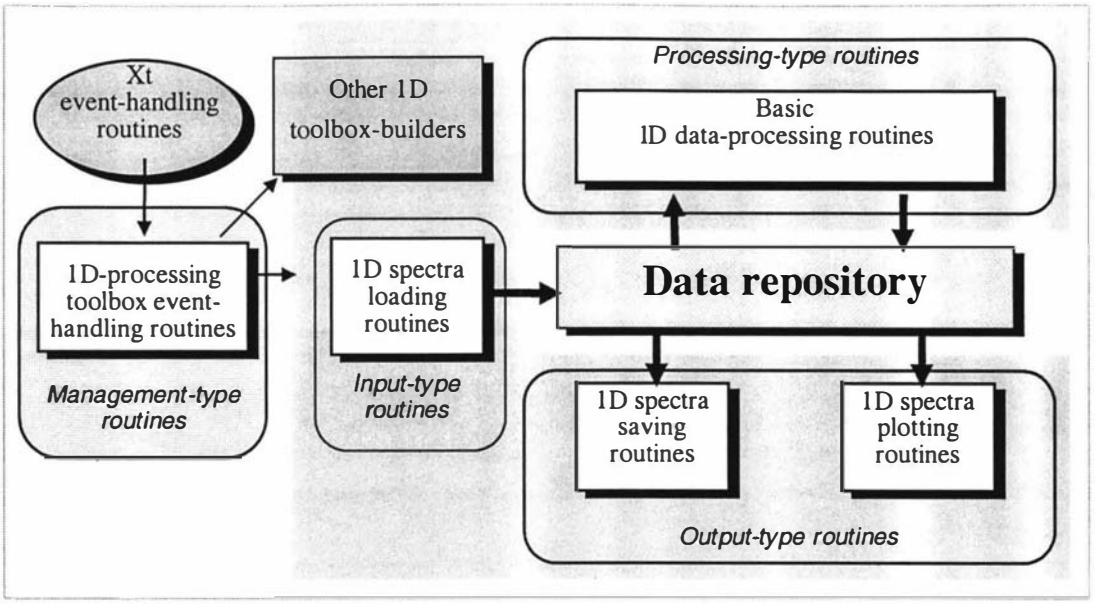


Figure 5.18: Routine structuring for the *1D-processing Toolbox*. When the toolbox's event-handling routine is invoked by the Xt event-handling routine, it will call either a toolbox builder to open one of the other *1D-processing Toolboxes* (see previous sub-section on the modules), or a routine that modifies or exports data in the repository.

The routine structuring designs for other processing modules in PAW follow a similar model and will not be discussed further.

5.3.7.3 Module implementation

The source code of PAW contains more than 30 thousand statements, most of them are highly optimised. This code is partitioned into more than 30 program files that are compiled individually and linked together. To ease the programming process, PAW is developed with good modularity. Each *module file* encapsulates a group of highly localised routines to implement and maintain a single user interface such as a menu bar, a toolbox, an entry box, or a dialog box. Each module is loosely coupled to the outside world with only one or two connection pathways.

There are currently 18 modules in PAW, consisting of one macro processing module, four dialog modules and thirteen toolbox modules:

The *macro-processing module* encapsulates a group of routines to process all macros. PAW's macros are not only used for data processing but also for other purposes such as setting initial values, menus, hotkeys, help messages, and processing parameters. In this way, PAW can draw information from ASCII files with a simple yet powerful macro language that can be easily expanded for any other purposes. (See also Volume II.)

A *dialog module* encapsulates a group of routines to implement and maintain dialogs, which are frequently called by many other routines. In PAW, a dialog contains a number of widgets, including keyboard-entry boxes, slides, buttons, etc. The four dialog modules currently in PAW are: the colour-map module, file-selection module,

keyboard-entry module, and slider-entry module. Among them, the number and contents of the slider-entry dialogs and keyboard-entry dialogs are dynamically implemented for different purposes.

A *toolbox module* encapsulates a group of highly localised routines to implement and maintain a toolbox that is called by a keyboard- or button-command. Each toolbox contains a number of buttons or groups of buttons. The 13 toolboxes currently in PAW are:

1. the *Common-display Toolbox*,
2. the *1D-display Toolbox*,
3. the *1D-processing Toolbox*,
4. the *1D-MiscProc Toolbox*,
5. the *Filtering Toolbox*,
6. the *Phasing Toolbox*,
7. the *Baseline-correction Toolbox*,
8. the *2D-display Toolbox*,
9. the *2D-processing Toolbox*,
10. the *Peak-picking Toolbox*,
11. the *Peak-assignment Toolbox*,
12. the *Peak-display Toolbox*, and
13. the *Drawing Toolbox*.

More details about the toolboxes can be found in Volume II.

Chapter 6:

NMR Data Processing with PAW

6.1 Introduction

This chapter describes NMR data processing with PAW, including the theory, general procedures, and examples. The interactive operation loops used during data processing with PAW are briefly described in Section 6.2. The processing of 1D data is discussed extensively in Section 6.3, because it is the core of NMR data processing and is complicated. In contrast, the processing of 2D data is only briefly discussed because it is essentially a series of 1D processing operations.

Note that, in this thesis, the term 'NMR data' is used for data in both the time and frequency domains. Unprocessed NMR time-domain data (sets) collected from NMR spectrometers are also called the *NMR raw data* (sets), whereas processed frequency-domain data sets are often called the *NMR spectra*.

6.2 Interactive Loops

Because of the variability of NMR experiments, NMR data are often processed with different parameters and methods. For this reason, many of the data-processing routines are designed to work interactively, although fully automatic macro processing is also available. Two types of interactive looping are used when processing NMR data with PAW, as shown in Figure 6.1.

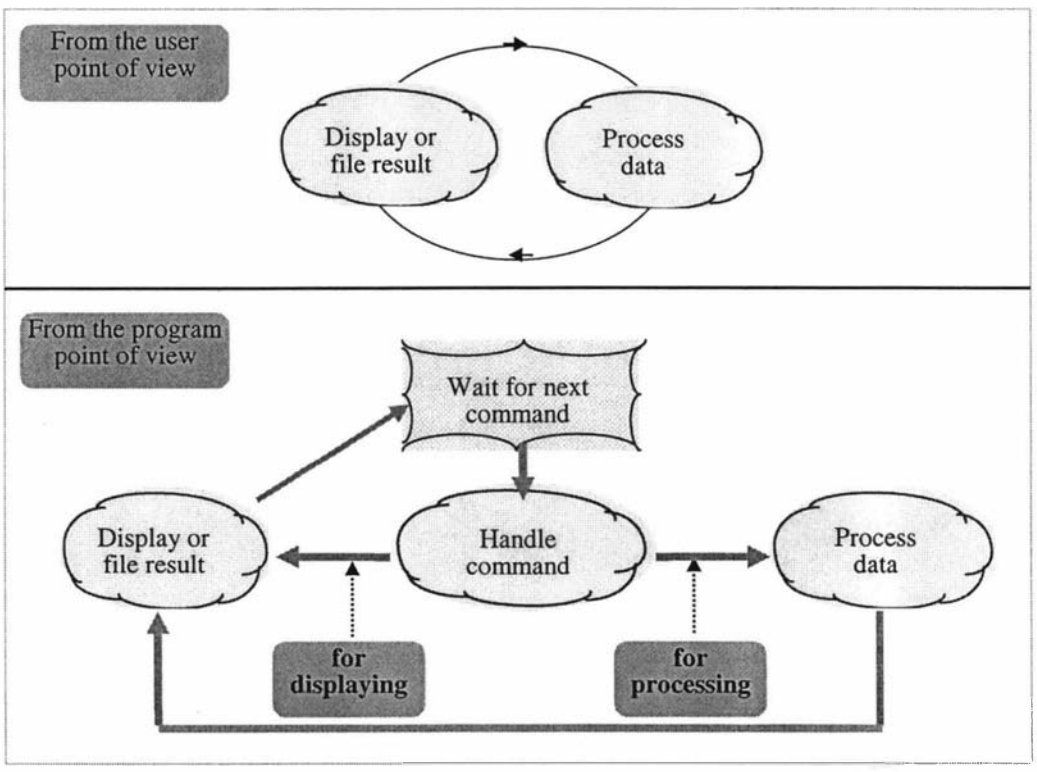


Figure 6.1: The interactive loops used when processing NMR data with PAW. From the user point of view, the loop is a simple cycle alternating between processing and display operations. The program has two interactive loops. For displaying, the loop is simply a command-handling and data-displaying cycle. For processing, the loop is a cycle of three steps: command handling, data processing, and automatic data displaying.

For example, following a phase correction operation, there is often a need to zoom into a small region to see if the results are satisfactory. This may be followed by a further phase correction or other operation.

Note that, in an interactive process, a series of processing operations can be done successively without having to re-display the data (unless a different display pattern is required), because PAW always upgrades the display whenever there is a change in data after an interactive operation. However, processing operations performed by executing a macro will not automatically result in any display update unless a display command is explicitly presented in the macro. This is to prevent unnecessary display operations. Otherwise they would be thousands of 1D-display updates in multidimensional data processing.

6.3 One-dimensional NMR Data Processing

Since multi-dimensional NMR data sets are usually made up of a large number of 1D data sets, 1D data processing is an important task that has attracted a lot of attention in the literature. This section discusses major aspects of 1D NMR data processing in detail, including theory and program implementation.

6.3.1 Convolution character of NMR data

An ideal 1D NMR data set can be viewed as the sum of a series of unperturbed harmonic signals multiplied by exponentially decaying functions. Because of the Fourier connection between the time- and frequency-domain representations of NMR data, the mathematical characteristics of NMR data can be explained using the *inverse convolution theorem*:

$$\mathcal{F}[s(t)w(t),\omega] = S(\omega) * W(\omega) \tag{6.3.1.1}$$

The notation means that the Fourier transform of a product of two functions is the convolution of their corresponding Fourier transforms. This theorem can be proved as follows:

$$\begin{aligned}
&\mathcal{F}^{-1}[S(\omega) * W(\omega);t] \\
&= \int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty} S(\xi)W(\omega-\xi)d\xi\right)e^{-i\omega t}d\omega \\
&= \int_{-\infty}^{\infty} S(\xi)\left(\int_{-\infty}^{\infty} W(\omega-\xi)e^{-it(\omega-\xi)}d\omega\right)e^{-it\xi}d\xi \\
&= w(t)\int_{-\infty}^{\infty} S(\xi)e^{-it\xi}d\xi \\
&= s(t)w(t)
\end{aligned} \tag{6.3.1.2}$$

Note that there is a subtle difference between the inverse convolution theorem and the forward *Fourier convolution theorem* (or simply, the *convolution theorem)*:

$$\mathcal{F}[g(t) * h(t),\omega] = G(\omega)H(\omega) \tag{6.3.1.3}$$

The notation used in (6.3.1.3) means that the Fourier transform (FT) of the convolution of two functions equals the product of their corresponding Fourier transforms. Likewise, the inverse Fourier transform of a product of two functions is the convolution of their corresponding inverse Fourier transform, which is not

of direct concern in NMR Fourier spectroscopy. Swapping the domain variables in (6.3.1.3) cannot result in (6.3.1.1).

The inverse convolution theorem is useful for understanding NMR spectra, as shown in the two examples below.

Example 1: Consider a set of exponentially decaying data consisting of a product of an unperturbed harmonic signal $s_u(t)$ with an exponentially decaying function $w(t)$:

$$s_u(t) = M_0 \exp(i\omega_0 t) \tag{6.3.1.4}$$

$$w(t) = \exp(-at) \quad a > 0 \tag{6.3.1.5}$$

Here, only a single resonance frequency is considered — cases for more frequencies can be treated as a sum. After Fourier transformation, a single spike is obtained in the real part, and nothing in the imaginary part, because

$$S_u(\omega) = \mathcal{F}[M_0 \exp(-i\omega_0 t); \omega] = \sqrt{2\pi} M_0 \delta(\omega - \omega_0) \tag{6.3.1.6}$$

The notation $\delta(\omega - \omega_0)$ is the Dirac delta function such that

$$\delta(\omega - \omega_0) = \begin{cases} \infty & at \ \ \omega = \omega_0 \\ 0 & elsewhere \end{cases} \tag{6.3.1.7}$$

and

$$\int_0^\infty \delta(\omega - \omega_0) d\omega = 1 \tag{6.3.1.8}$$

This is a standard formula in Fourier transform theory [see, e.g., Zwillinger 1989 or Gradshteyn and Ryzhik 1983]. The proof can be obtained by integrating the Green's function for a Sturm-Liouville eigenvalue problem. For more details, see [Davies 1985] or [Stakgold 1979].

The corresponding real spectrum is the convolution of a spike at a point ω_0 with a Lorentzian function:

$$S_{real}(\omega) = S_u(\omega) * \sqrt{\frac{2}{\pi}} \frac{a}{\left(a^2 + (\omega - \omega_0)^2\right)} \tag{6.3.1.9}$$

This explains the occurrence of Lorentzian line-shapes in the spectra of exponentially decaying NMR signals.

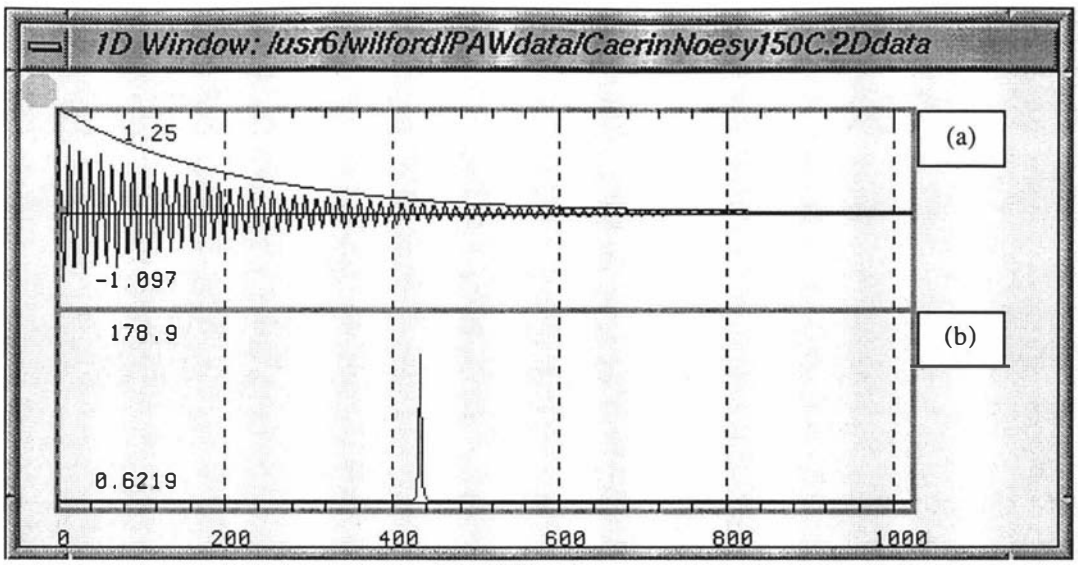


Figure 6.2: (a) An exponentially decaying signal produced by taking the product of an unperturbed harmonic signal with an exponentially decaying function. (b) The Lorentzian peak in the real part of the NMR spectrum of the exponentially decaying signal.

Example 2: Consider a truncated time-domain data set that is equivalent to the product of an exponentially decaying signal $s_{real}(t)$ with a hat function $w(t)$:

$$s_{trunc}(t) = s_{real}(t)w(t) \tag{6.3.1.10}$$

$$w(t) = \begin{cases} 1 & |t| \leq t_{trunc} \\ 0 & elsewhere \end{cases} \tag{6.3.1.11}$$

Hence, the frequency-domain spectrum is a convolution of a Lorentzian function with a sinc function:

$$S_{trunc}(\omega) = S_{real}(\omega) * \sqrt{\frac{2}{\pi}} \frac{\sin(t_{trunc}\omega)}{\omega} \tag{6.3.1.12}$$

This explains the reason for the occurrence of ripples in the spectra of truncated signals.

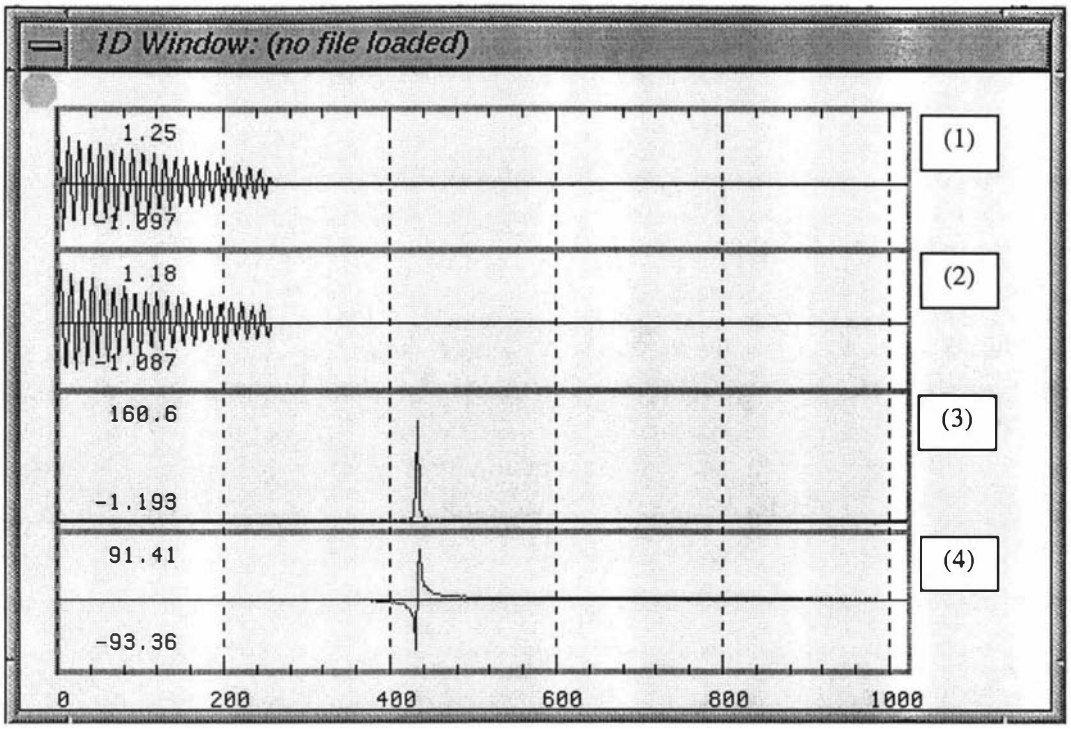


Figure 6.3: A truncated FID and the sinc-ripple problem. Plot (1) and (2): The real and imaginary parts of a truncated FID produced by taking the product of an exponentially decaying signal with a hat function. Plot (3) and (4): The Lorentzian peaks with sinc-ripples in the real and imaginary parts of the corresponding NMR spectrum.

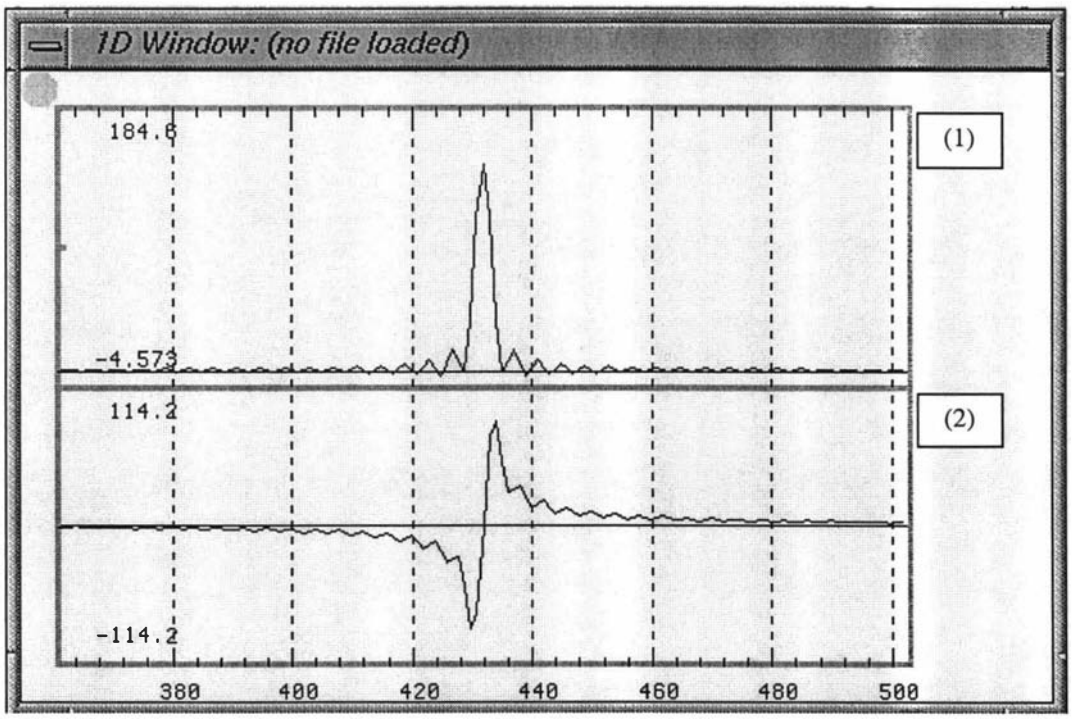


Figure 6.4: An expanded view of plot (3) and (4) in the last figure.

6.3.2 General problems

If NMR data are collected under perfect conditions, only one processing step is needed — the Fourier transform (FT). This converts time-domain data sets into frequency-domain spectra with *pure absorption* peaks in the real-part and *pure dispersion* peaks in the imaginary-part, as shown in Chapter 2.

In reality, NMR data sets are not only always mixed with noise, but are also often distorted. The imperfection generally results in five problems that must be corrected by 1D spectral processing. These are the sinc-ripples, noise, excessive line-width, phase error, and baseline distortion.

The *sinc-ripple problem* manifests itself as strong ripples on both sides of every peak in frequency domain, as already shown in Figure 6.4. This is caused by a truncation in the time-domain data, often due to adding zeros to expand the data size without proper apodisation. By truncation, we mean a sudden drop to zero at and from a point in a data set. (Data sets not ending with zero are not necessarily truncated, even if their end values are large.)

The *noise problem* manifests itself as Gaussian noise that is added to the detected signal, thereby making it difficult to identify weak peaks. The degree of this problem is measured by the *signal-to-noise ratio.* This is defined as S/σ_N, where S is the peak height of a reference signal, and σ_N is the root-mean-square of the noise amplitude. In data processing, the problem is exacerbated when the time-domain data is over-enhanced near the end of a data set, as shown in Figure 6.5.

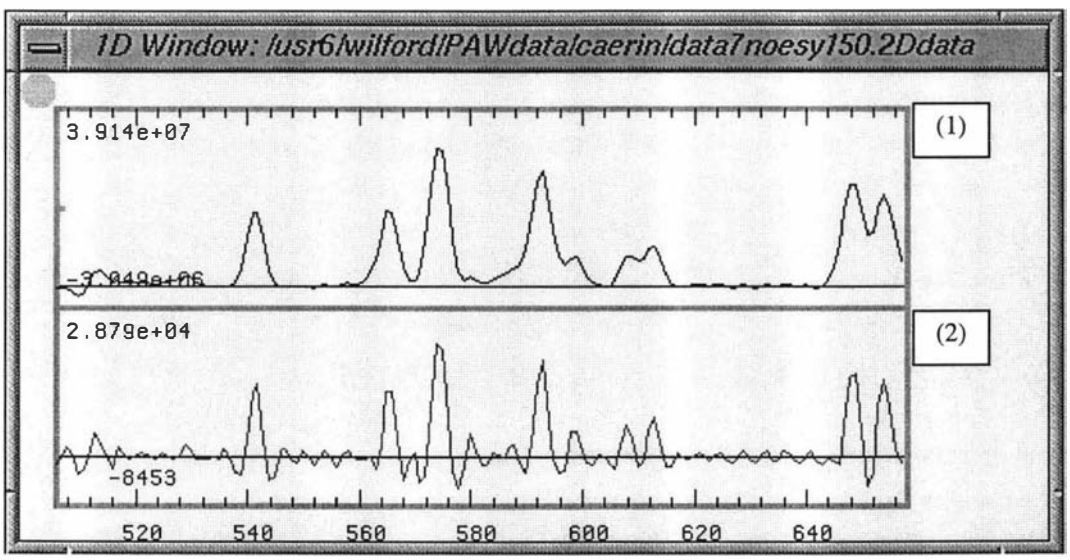


Figure 6.5: The noise problem due to over-enhancement at the end of a time-domain data set. (1) A region of a spectrum processed with little enhancement. (2) The same region of the spectrum processed with over-enhancement, showing much a higher noise level.

The *line-broadening problem* manifests itself as unacceptably broad peaks in a frequency-domain spectrum. This is often due to sample conditions and field in-homogeneity that cause the time-domain data to decay rapidly. Figure 6.6 shows the line-broadening problem due to rapid decay of the time-domain signal.

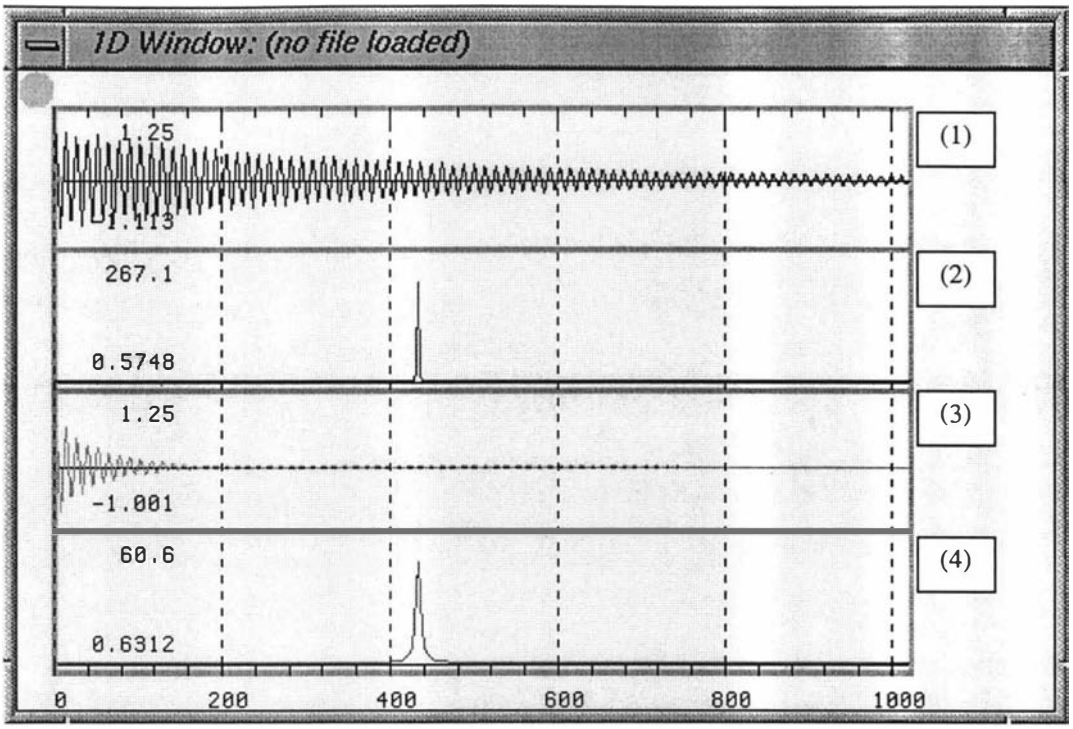


Figure 6.6: Line-broadening problem due to rapid decay of the time-domain signal. Plot (1) and (2): A time-domain data set with little decay and the corresponding spectrum. Plot (3) and (4):): A time-domain data set with rapid decay and the corresponding spectrum.

The *phase problem* manifests itself as a phase shift $e^{i\varphi}$ that may or may not depend upon resonance frequency. The problem is caused by a number of factors, including initial conditions, *off-resonance effects* and *transverse interference*[1] [see, e.g., Ernst *et al.* 1987], as well as the electronics between transmitter and digitiser. Almost every NMR data set shows some degree of phase error, as shown in the next diagram.

[1] This is a result from incomplete recovery of transverse magnetisation between pulses in repetitive pulse experiments.

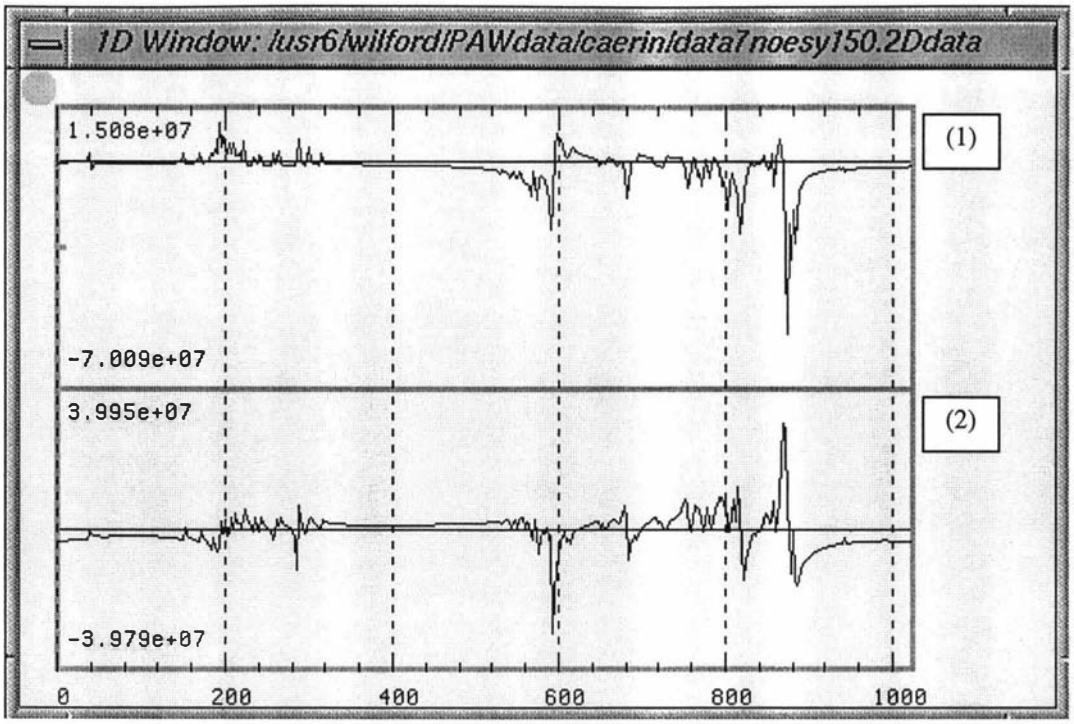


Figure 6.7: Phase error in the real and imaginary parts of an NMR spectrum. Plot (1): The real part shows peaks on the left are mainly positive, whereas those on the right are mainly negative. Plot (2): The imaginary part always shows a 90-degree phase shift with respect to the real part.

The *baseline problem* manifests itself as a poor spectral baseline, usually caused by distortion of the first few data points in an FID. An example is shown in the next diagram.

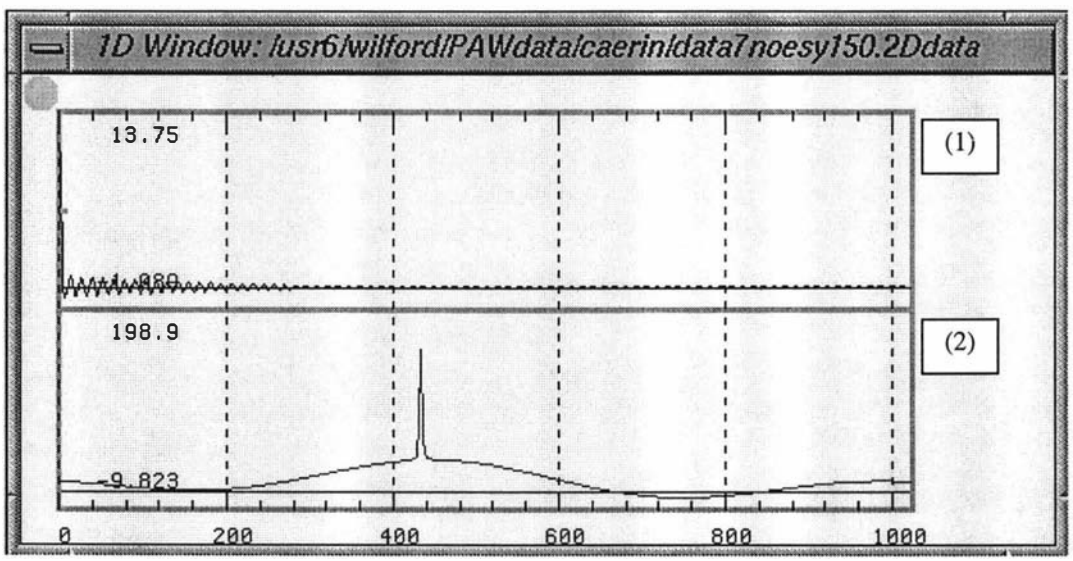


Figure 6.8: (1) An FID with the first two points being distorted. (2) The corresponding spectrum showing a distorted baseline.

6.3.3 General procedure

Because of the diversity of NMR data, different techniques are required to solve the problems mentioned in the last subsection. Figure 6.9 illustrates the possible steps taken during 1D NMR data processing.

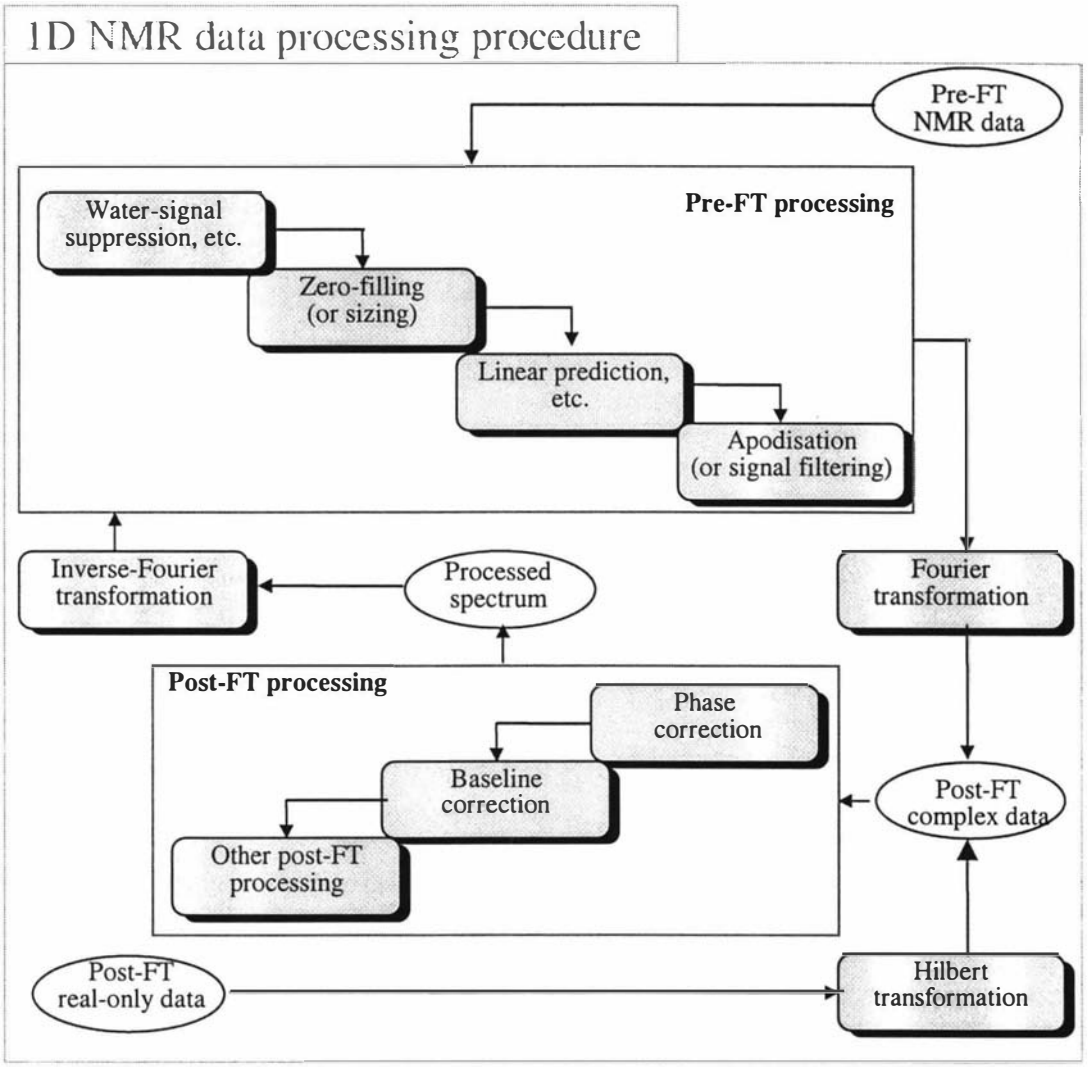


Figure 6.9: A possible 1D NMR data processing procedure. NMR time-domain data are usually processed with a number of operations before applying a Fourier transform (FT) that converts the data into complex frequency-domain data. The post-FT processing is then applied to improve the processed spectrum. If necessary, the inverse Fourier transform can be applied to convert the spectrum back into time-domain data for further processing. The post-FT real-only data are often available in multi-dimensional NMR data processing, and are converted into complex form before performing the post-FT processing.

In the figure, there are four pre-Fourier-transform (FT) operations. The time-domain *water-signal suppression* removes predominant water signals; *zero-filling* expands the size of 1D data sets; *linear prediction* (LP) fills the expanded part of 1D data sets using an extrapolation technique; and *apodisation* applies various filter functions to the data. In addition, a number of other pre-FT processing techniques have also been occasionally employed. These include a technique based on Jaynes' *principle of maximum entropy* [Jaynes 1957, Hoch 1984, Hore 1985, Hore & Denille 1986, Laue

1986, Martin 1985, Levy & Scheraga 1986], and one based on the *criteria of minimum area* [Newman 1988].

The *Fourier transform* converts time-domain real or complex signals into complex frequency-domain spectra. The *Inverse Fourier transform*, on the other hand, converts complex spectra into time domain data. The *Hilbert transform* converts real-only data sets into complex.

The two major post-FT processing steps are *phase correction* and *baseline correction*. The former corrects the zeroth and 1st order phase errors, and requires the data be complex; while the latter fits a number of sampled baseline points with a polynomial or a spline function. Other post-FT processing operations may include post-FT water-signal suppression and noise elimination.

Note that although a number of pre- and post-FT processing steps are presented in the two water-fall diagrams, not all the steps are required for processing every set of 1D data. The actual operations depend upon the size and quality of the data. For example, zero filling may not be necessary for sufficiently sampled data sets. Linear prediction may not be necessary for low distortion and un-truncated data sets. Water-signal suppression may not be required if data are collected with an adequate water-suppression technique during experiments. Likewise, the post-FT steps may be ignored if the quality of the processed spectra is acceptable. Finally, the inverse Fourier transform is seldom used after the baseline correction is performed.

Except for zero filling, each of the above steps is complicated in terms of mathematics and programming. Steps such as apodisation, phase correction and baseline correction often require special attention and different degrees of human interaction.

6.3.4 General interfaces for one-dimensional data processing

In addition to keyboard commands and macros, PAW provides five toolboxes for 1D data processing, as shown in the next two diagrams.

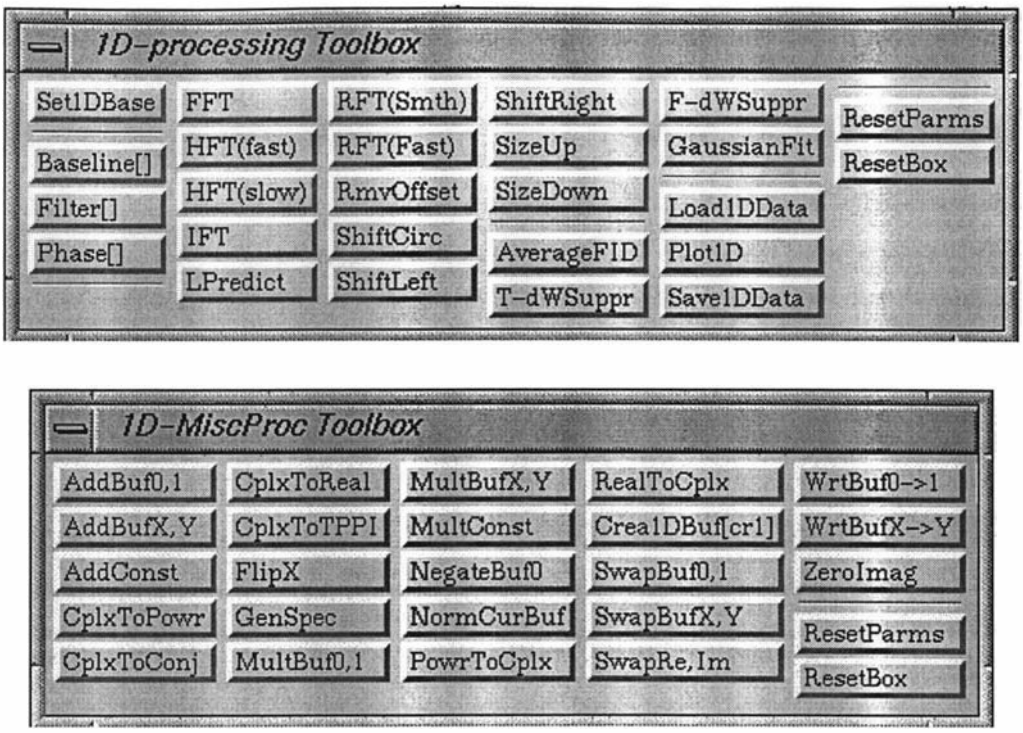


Figure 6.10: Two of the *1D-processing Toolboxes* available in PAW. The *1D-processing Toolbox* (top) contains the most frequently used 1D-processing command buttons, in which the first three buttons provide accesses to three other toolboxes for apodisation, phasing, and baseline correction. The *1D-MiscProc Toolbox* (bottom) contains buttons that are seldom used.

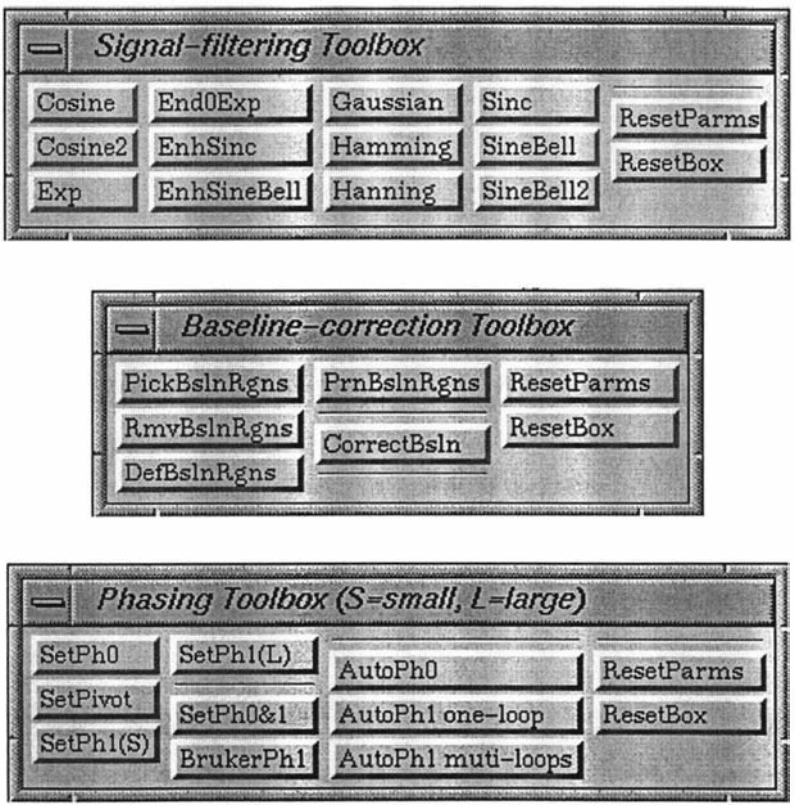


Figure 6.11: Other *1D-processing Toolboxes* in PAW. The *Signal-filtering Toolbox* (top) contains command buttons for accessing the various filters use in the apodisation process. The *Phasing Toolbox* (middle) contains command buttons to correct different types of phase errors. The *Baseline-correction Toolbox* (bottom) contains command buttons to perform the various operations required in baseline correction.

6.3.5 Zero-filling

Zero-filling is only applied to time-domain data. It expands the data size by filling zeros to the end so as to obtain Fourier spectra with smoother peaks.

6.3.5.1 Theory

A zero-filled data set is equivalent to an expanded data set weighted with a hat function. Consequently, the frequency-domain spectrum is a convolution of the corresponding spectra of the expanded data set and the sinc function. (See section 6.3.1.) This ensures the smoothness of Fourier spectra obtained from zero-filled time-domain data because convolution can smooth data.

However, zero-filling also induces sinc-ripples as a result of truncation. For this reason, it is always followed by at least an apodisation to help eliminate ripples. The next figure shows the peak-shape differences in two spectra processed with or without zero filling but filtered with the same apodisation function.

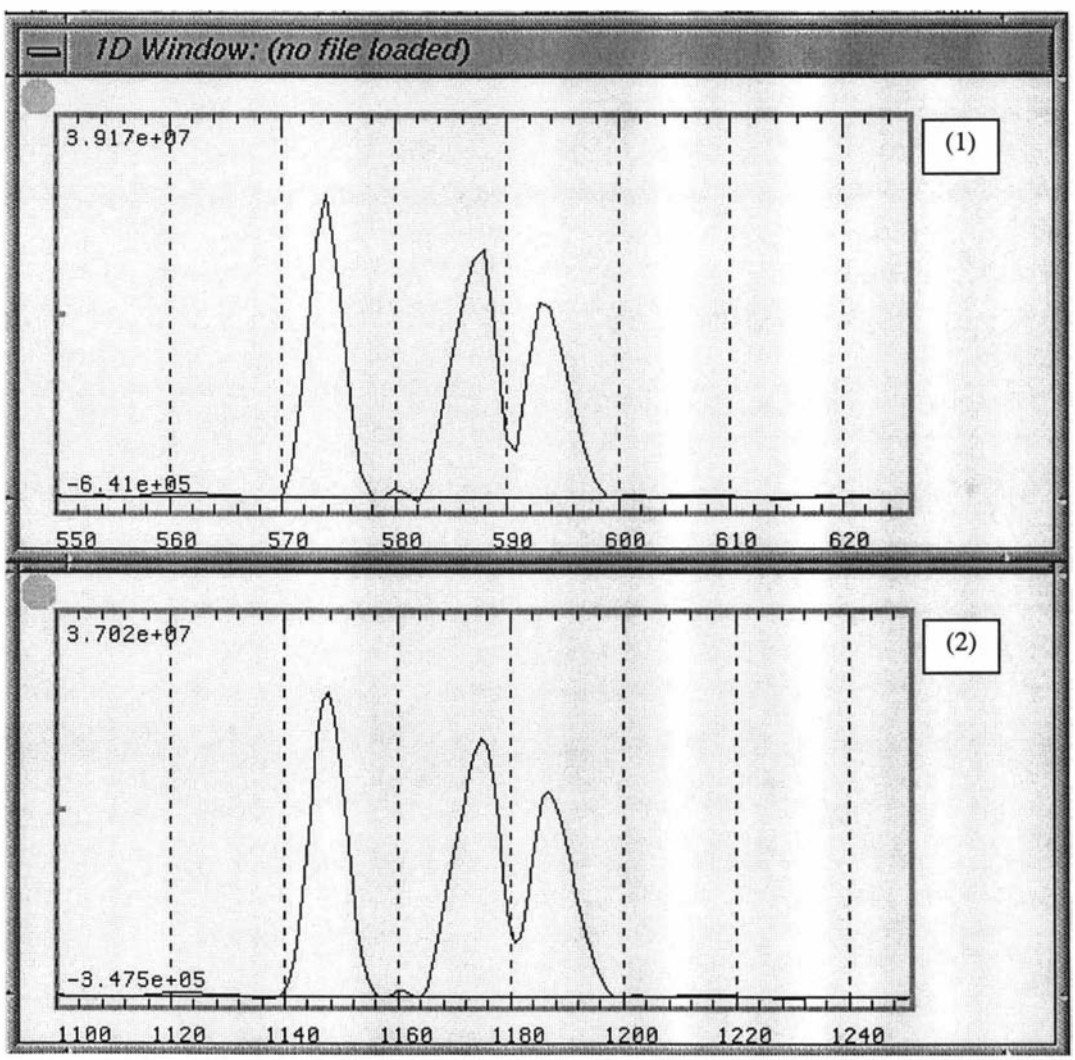


Figure 6.12: The difference in peak shapes with or without zero-filling. The same filter has been used in both cases.

Because of these advantages, zero filling is frequently used. The operation is extremely important in multi-dimensional NMR data processing, especially in the second, third and forth dimensions. This is because the initial sizes in these dimensions are the numbers of FIDs collected for each dimension (or half of these number in the case of complex data), which can be as small as 32.

The practical limitation to zero-filling is the size of the resultant data, especially for multidimensional data sets. This is further hampered by a stringent restriction that requires the size of NMR data set be a power of two, as will be explained later.

6.3.5.2 Program implementation

In PAW, the zero-filling command button is [SizeUp] (or szu as a keyboard command). The operation doubles the size of a data set and fills the expanded part with zeros. This is because the fast Fourier transform (FFT) algorithm employed requires the data size be an integer-power of two.

The corresponding command button for the reverse operation is [SizeDown] (or szd as a keyboard command). The command halves the data size, retaining only the first half of the data.

6.3.6 Apodisation

Apodisation, also called *filtering* or *windowing*, is one of the most useful steps in spectral processing. This technique is used to eliminate ripples, reduce noise, narrow peak-shapes, and improve the spectral baseline.

6.3.6.1 Theory

Mathematically, apodisation is simply a process that shapes the envelope of a data set. This is performed by multiplying an FID with a weighting function (also called *filter functions*, *window functions*, or *apodisation functions*). In practise, an apodisation operation is the search for a weighting function with parameters that optimise the spectral peak-shapes and signal-to-noise ratio after FT.

Consider a general case in which the data set is the product of an unperturbed harmonic signal s_u with a filter $w(t)$ such that

$$s_u(t) = \begin{cases} M_0 \exp(i\omega_0 t) & t < t_{acq} \\ 0 & elsewhere \end{cases} \qquad (6.3.6.1.1)$$

$$s(t) = s_u(t)w(t) \qquad (6.3.6.1.2)$$

The corresponding NMR spectrum is

$$S(\omega) = \frac{M_0}{\sqrt{2\pi}} \int_0^{t_{acq}} w(t) e^{i(\omega-\omega_0)t} dt \qquad (6.3.6.1.3)$$

The real spectra resulting from the application of three simple filters are as follows:

$$w(t) = \begin{cases} 1 & |t| \leq t_{trunc} \\ 0 & elsewhere \end{cases} \Rightarrow S_{real}(\omega) = \frac{M_0}{\sqrt{2\pi}} \frac{\sin(t_{trunc}(\omega - \omega_0))}{(\omega - \omega_0)} \qquad (6.3.6.1.4)$$

$$w(t) = e^{-at} \quad \Rightarrow \quad S_{real}(\omega) = \frac{M_0}{\sqrt{2\pi}} \frac{a}{a^2 + (\omega - \omega_0)^2} \tag{6.3.6.1.5}$$

$$w(t) = 1 \quad \Rightarrow \quad S_{real}(\omega) = \sqrt{2\pi} M_0 \delta(\omega - \omega_0) \tag{6.3.6.1.6}$$

The equations on the right represent, respectively, a sinc-function, a Lorentzian, and a spike. All of them are centred at ω_0. As a matter of fact, the results are simply the convolutions of the Fourier transforms of $s_u(t)$ and $w(t)$, as discussed previously.

It is useful to look at the implications that can be drawn from these equations. For example, the second equation implies that, to obtain a Lorentzian peak (see Appendix 6d), the time-domain data set must be filtered to be exponentially decaying. Likewise, if expecting a spike, the time-domain data must be filtered so that it does not decay. This means filtering can produce an expected line-shape as long as the time-domain data set is shaped correspondingly. Understanding this principle of filter searching will free us from a labyrinth of guessing. (Despite that expecting to obtain a spectrum containing only spikes by applying a filter is unrealistic because of the noise.)

Many weighting functions have been proposed in the past [see, e.g., Ernst *et al.* 1987]; some of them are quite complicated. In practice, it is not necessary to use complicated functions for the best results. This is because small differences in data envelopes do not induce big differences in corresponding spectra, and a sequence of simple weighting functions can effectively produce similar envelopes to those produced by complicated ones.

6.3.6.2 Program implementation

The filter functions listed below are ones that are implemented in PAW; most of them are used frequently. In these expressions, *a*, *b* and *c* are user-adjustable parameters; t^{trunc} is the time from which data is truncated or the point where data is smoothed down to zero.

- *cosine filter*:

$$W(t) = \begin{cases} \cos(\pi t / t^{trunc}) & 0 \le t \le t^{trunc} \\ 0 & elsewhere \end{cases} \tag{6.3.6.2.1}$$

- *cosine squared filter*:

$$W(t) = \begin{cases} \cos^2(\pi t / t^{trunc}) & 0 \le t \le t^{trunc} \\ 0 & elsewhere \end{cases} \tag{6.3.6.2.2}$$

- *exponential-broadening filter*:

$$W(t) = \begin{cases} \exp(-at) & 0 \le t \le t^{trunc} \\ 0 & elsewhere \end{cases} \tag{6.3.6.2.3}$$

- *shifted exponential-broadening filter*:

$$W(t) = \begin{cases} \exp(-at) - \exp(at^{trunc}) & 0 \le t \le t^{trunc} \\ 0 & elsewhere \end{cases} \quad (6.3.6.2.4)$$

- *Hanning filter*:

$$W(t) = \begin{cases} 0.5 \times (1 + \cos(at + b)) & 0 \le t \le t^{trunc} \\ 0 & elsewhere \end{cases} \quad (6.3.6.2.5)$$

- *Hanning square filter*:

$$W(t) = \begin{cases} (0.5 \times (1 + \cos(at + b)))^2 & 0 \le t \le t^{trunc} \\ 0 & elsewhere \end{cases} \quad (6.3.6.2.6)$$

- *Hamming filter*:

$$W(t) = \begin{cases} 0.54 + 0.46\cos(\pi t / t^{trunc}) & 0 \le t \le t^{trunc} \\ 0 & elsewhere \end{cases} \quad (6.3.6.2.7)$$

- *Lorentz-Gauss transform filter*:

$$W(t) = \begin{cases} \exp(at - b^2 t^2 / 2) & 0 \le t \le t^{trunc} \\ 0 & elsewhere \end{cases} \quad (6.3.6.2.8)$$

- *sine-bell filter*:

$$W(t) = \begin{cases} \sin(at + b) & 0 \le t \le t^{trunc} \\ 0 & elsewhere \end{cases} \quad (6.3.6.2.9)$$

- *sine-bell squared filter:*

$$W(t) = \begin{cases} \sin^2(at + b) & 0 \le t \le t^{trunc} \\ 0 & elsewhere \end{cases} \quad (6.3.6.2.10)$$

- *enhanced sine-bell filter*:

$$W(t) = \begin{cases} \sin^n(at + b) \times \exp(ct) & 0 \le t \le t^{trunc} \\ 0 & elsewhere \end{cases} \quad (6.3.6.2.11)$$

6.3.6.3 User-interface

The user-interface in PAW is similar for all apodisation operations. In short, users simply issue a filter command or click on a button in the *Filtering Toolbox* to open a filter dialog. A typical example of the filter dialog is shown in the next figure. (See Volume II for more details.)

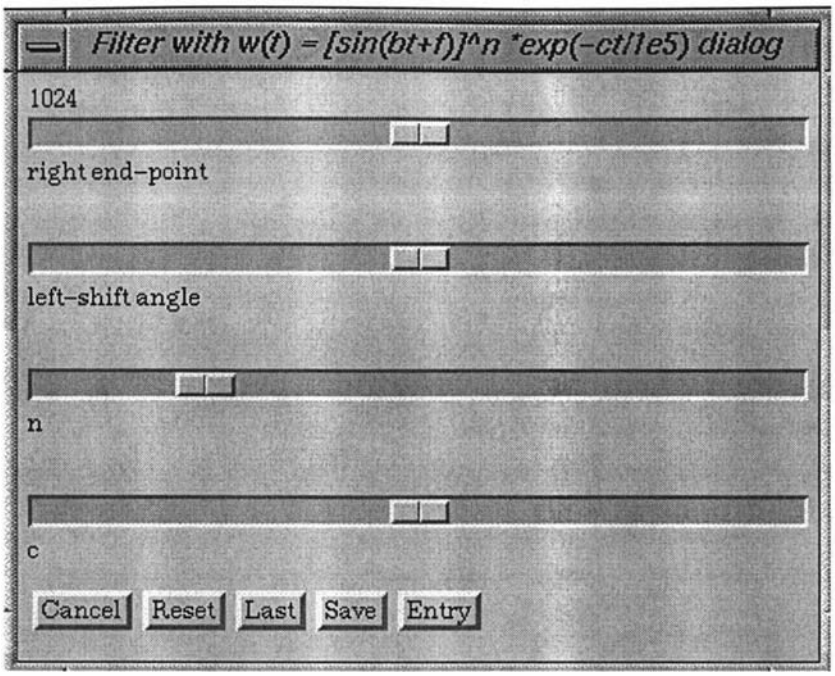


Figure 6.13: A dialog for an apodisation operation. The sliders can be adjusted to set the filter parameters in real time. Alternatively, a keyboard-entry dialog can be opened by choosing [Entry].

6.3.6.4 Choosing a filter function

There is no single formula that predicts what filters must be used for what type of spectra. An optimal filter for a signal should give the narrowest peak shape with an acceptable noise level in the corresponding NMR spectrum. Yet there must be a compromise between narrowing the peak shape and lowering the noise level, as will be discussed later. In general, a filter is chosen to enhance and suppress different parts of a data set.

The *Lorentz-Gauss transform filter* is considered one of the best for NMR spectral processing, because applying it can produce reasonably narrow peak shapes that are elliptically symmetric (see Appendix 6c). This means that every 2D peak-contour is an ellipse that is symmetric about the two mutually perpendicular axes passing through the centre of the contour. Many people, however, still use squared sine-bell or Gaussian filters in processing NOESY and TOCSY spectra because they suppress the initial part of the data and produce acceptable peak shapes in the Fourier spectra. For the COSY spectra, which start from zero, a 90° right-shifted cosine filter is often used.

The enhanced sine-bell filter introduced in PAW has worked reasonably well for all spectra. It has the combined advantages of exponential narrowing and different degrees of sine-bell enhancement. Depending on the parameters chosen, it can also be reduced to one of the conventional filters, such as the cosine, cosine-squared, sine-bell, sine-bell-squared, exponential-broadening or -narrowing filter. With a set of properly chosen parameters, it can work better than the Lorentzian-Gaussian filter. (See also Appendix 6e for the peak-shapes.)

6.3.6.5 Eliminating ripples by apodisation

To eliminate ripples from a spectrum, truncation in the time-domain data must be prevented. This can be achieved simply by applying apodisation to gradually reduce data values before the truncation point to zero. Not surprisingly, this ripple-suppression technique must be accompanied by a certain degree of peak broadening because it increases the decay rate.

6.3.6.6 Enhancing signal-to-noise ratio

The signal-to-noise ratio in an NMR spectrum can be increased by suppressing the tail part of the time-domain data. Unfortunately, the side effect is inevitable peak broadening, which is contradictory to a more important goal in spectral processing, as explained next.

6.3.6.7 Narrowing line-width

In the laboratory frame, an *ideal FID* from a single nuclear species can be expressed as

$$s(t) = M_0 \exp(-t/T_2) \exp(i\omega_0 t) \qquad (6.3.6.7.1)$$

Multiplying this by $\exp(t/T_2)$ results in an unperturbed harmonic signal, as discussed previously. The corresponding NMR spectrum is a spike, which is the narrowest peak one can obtain. Therefore, the principle of narrowing line-width is to minimise the signal decay rate. This can be achieved by applying filters to enhance the tail part of time-domain data.

Different filters can be used to narrow peaks. For example, the three peaks in Figure 6.14 were obtained by processing an ideal FID with exponential-narrowing, Lorentzian-Gaussian, and enhanced sine-bell squared filters, respectively. In this case, the exponential narrowing filter produced the broadest peak, while the enhanced sine-bell, the narrowest. Here, the parameters for the latter two filters are chosen so that the 'ditches' on either side of the peak are nearly the same.

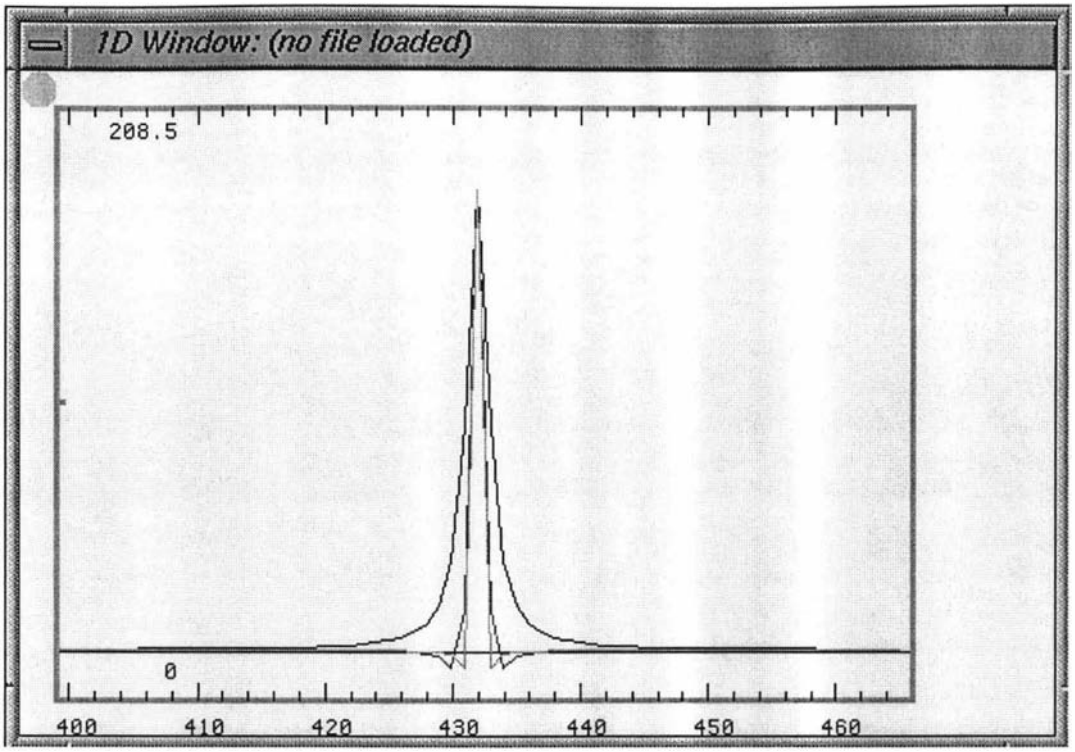


Figure 6.14: Spectra of an FID weighted by different filter functions. These were obtained by processing an ideal FID with exponential-narrowing, Lorentzian-Gaussian, and enhanced sine-bell square filters, respectively. In this case, the exponential narrowing filter produced the broadest peak, while the enhanced sine-bell, the narrowest.

Line-width narrowing using apodisation requires a larger weighting on the tail of an FID, where true signals are vanishing but the noise level remains. Therefore, it can only be achieved at the expense of a reduction in the signal-to-noise ratio, because the noise is also enhanced.

6.3.6.8 Obtaining ideal Gaussian peaks

If the FID is ideal, the Lorentz-Gauss transform filter (6.3.2.8) can produce an ideal Gaussian peak. For simplicity, consider an ideal FID for a spin that resonates at a frequency of ω_0:

$$s(t) = M_0 \exp(-i\omega_0 t) \exp(-t/T_2) \qquad (6.3.6.8.1)$$

Multiplying this with (6.3.2.8), a filtered signal is obtained:

$$s(t)W(t) = M_0 \exp(-i\omega_0 t))\exp(ct - \tfrac{1}{2}b^2t^2) \qquad (6.3.6.8.2)$$

where

$$c = a - 1/T_2 \qquad (6.3.6.8.3)$$

The Fourier transform of $s(t)W(t)$ gives

$$S(\Omega) = \frac{1}{\sqrt{2\pi}} M_0 \int_{-\infty}^{\infty} \exp(ct - \tfrac{1}{2}b^2t^2)\exp(i\Omega t)dt \qquad (6.3.6.8.4)$$

where

$$\Omega = \omega - \omega_0 \qquad (6.3.6.8.5)$$

If the parameter a is chosen so that $c = 0$, then

$$S(\Omega) = \frac{1}{b} M_0 \exp(-\Omega^2 / 2b^2) \qquad (6.3.6.8.6)$$

This represents an ideal Gaussian peak centred at $\Omega = 0$, that is, at $\omega = \omega_0$. The significant part of a Gaussian peak is symmetrically distributed within a close distance of its peak centre, as shown in Figure 6.15 (a).

If c is not zero, then,

$$S(\Omega) = \frac{1}{b} M_0 \exp\left(\frac{c^2}{2b^2}\right) \exp(-\frac{\Omega^2}{2b^2}) \exp\left(-i\frac{c\Omega}{b^2}\right) \qquad (6.3.6.8.7)$$

The real part is

$$S_{real}(\Omega) = \frac{1}{b} M_0 \exp\left(\frac{c^2}{2b^2}\right) \exp(-\frac{\Omega^2}{2b^2}) \cos\left(\frac{c\Omega}{b^2}\right) \qquad (6.3.6.8.8)$$

The sign of the cosine factor in the equation is negative in different parts:

$$\cos\left(\frac{c\Omega}{b^2}\right) < 0 \quad \text{for} \quad (2n + \tfrac{1}{2})\pi < \left|\frac{c\Omega}{b^2}\right| < (2n + \tfrac{3}{2})\pi, \quad n = 0,1,2,\ldots \qquad (6.3.6.8.9)$$

Therefore, the signs on both sides of the peak centred at ω_0 are changed periodically to negative. If the negative parts of the cosine factor are within the significant part of the Gaussian peak $\exp(-\Omega^2 / 2b^2)$, then, negative lobes can be seen symmetrically on both sides of the peak, as shown in Figure 6.15 (b).

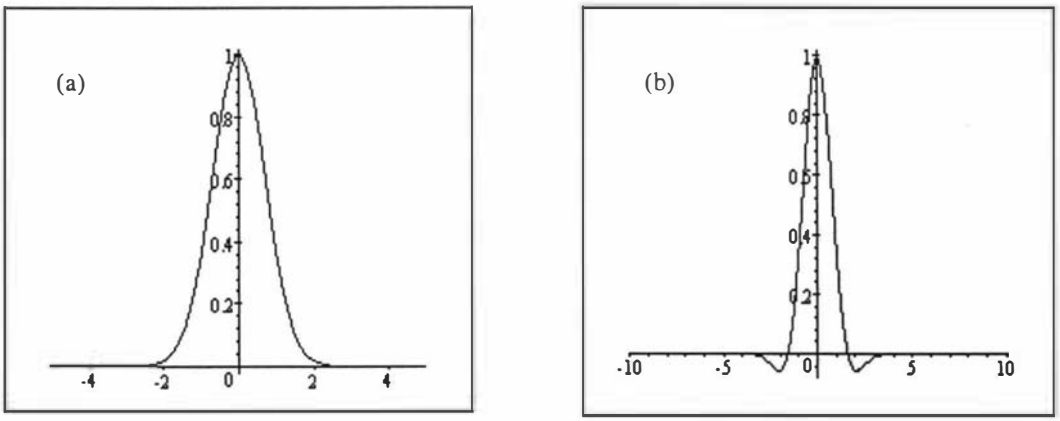


Figure 6.15: (a) An ideal Gaussian peak. (b) A peak narrowed by a Lorentzian-Gaussian filter, where the negative lobes are caused by a large parameter a.

To conclude, spectral filtering is about eliminating sinc-ripple (if any), suppressing noise, and narrowing line-width. However, resolution enhancement and signal-to-noise ratio improvement are mutually exclusive goals, and hence a compromise must be reached. Because different data sets have different peak shapes, there is no single 'best' filter.

6.3.7 Linear prediction

Traditional 1D NMR data processing primarily relies on apodisation to eliminate ripples, narrow peak-shapes and suppress the first few unreliable data points so as to improve spectral baselines. Although very useful, apodisation often introduces problems such as noise and undesired peak-shapes, as explained in the next subsection.

To understand why, a device-related problem in NMR data must be investigated. For an FID that is collected with sufficient acquisition time, the head and tail part are the most susceptible to perturbation. The former is often distorted due to improper timing of device switching, while the latter is dominated by noise because of diminishing signal with time. For an FID that is collected without sufficient acquisition time, such as those for the D_2, D_3 and D_4 dimensions in multi-dimensional NMR data, the first few points are the most susceptible to perturbation. (Here D_2, D_3 and D_4 refer to the time domains in the second, third and fourth dimensions of an mD NMR data set.) The most reliable information resides in the central part where the signals are relatively strong. It would therefore be desirable to use a reliable part of a data set to predict and replace an unreliable part.

In recent years, an extrapolation technique called *linear prediction* (LP) has become increasingly attractive in spectral data processing [Haacke *et al.* 1990, Hess and Liang 1996]. The technique is capable of predicting the future (or past) of a time series from a set of known records varying with a regular pattern, and is particularly successful at extrapolating smooth and oscillatory signals. Therefore, it is suitable for predicting time-domain NMR data because generically they are related to a number of resonance frequencies and hence must follow a regular pattern. Clearly, being able to predict data is important for processing 1D NMR data sets in the D_2, D_3 and D_4 dimensions that often contain only a limited number of data points and must be significantly expanded.

6.3.7.1 Theory

The mathematics behind LP cannot be covered in a few pages [Strobach 1998], and is beyond the scope of this thesis. In brief, LP extrapolates the signals using a linear equation consisting of M available data points:

$$y_n = \sum_{k=1}^{M} \alpha_k y_{n-k} \qquad n = N+1, N+2, \ldots \qquad (6.3.7.1.1)$$

Here y_n is the n^{th} predicted value, and α_k *is* the k^{th} *prediction coefficient*. There are several methods to obtain the prediction coefficients. The one implemented in PAW embraces a concept called the *maximum entropy method* (MEM) [Burg 1967, 1968, 1972, Anderson 1977, Anderson and Mentz 1980].

In reality, there must be some discrepancy between the prediction and the true value. For this reason, if an FID data set has not significantly decayed and contains only a few resonance frequencies, one can expect a better result from applying linear prediction [Delsuc 1989]. Otherwise, no magical results can be produced using LP because any prediction based on significantly decayed data cannot give much more information.

6.3.7.2 Program implementation

In PAW, a group of routines were implemented to calculate the prediction coefficients using a very fast recursive algorithm that utilised the symmetry of the matrix [Anderson 1977]. The results for α_k are then used to calculate the future values using Equation (6.3.7.1.1). To predict values for the first few points in a time series, PAW simply flips the data set back to front, predict the 'future' values, then flips the data back again.

6.3.7.3 User-interface

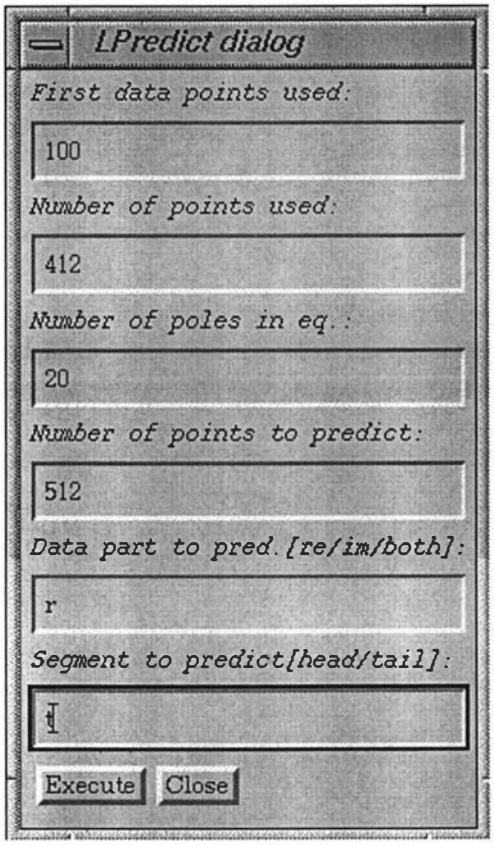


Figure 6.16: The dialog for linear prediction.

In PAW, linear prediction can be performed by issuing the command lp or choosing [LPredict] in the *1D-processing Toolbox* to open a dialog, as shown on the left.

With this dialog, one can specify how the LP is to be performed. Six parameters are required for the operation; these are explained by the labels used in the dialog.

(See Volume II for more details.)

6.3.7.4 Effectiveness of PAW's LP command

The next two diagrams provide a comparison of results produced by processing with LP and by applying conventional apodisation techniques. The time-domain data set prior to applying LP in Figure 6.17 was drawn from a column in a 2D NMR data set obtained after FT in the first dimension.

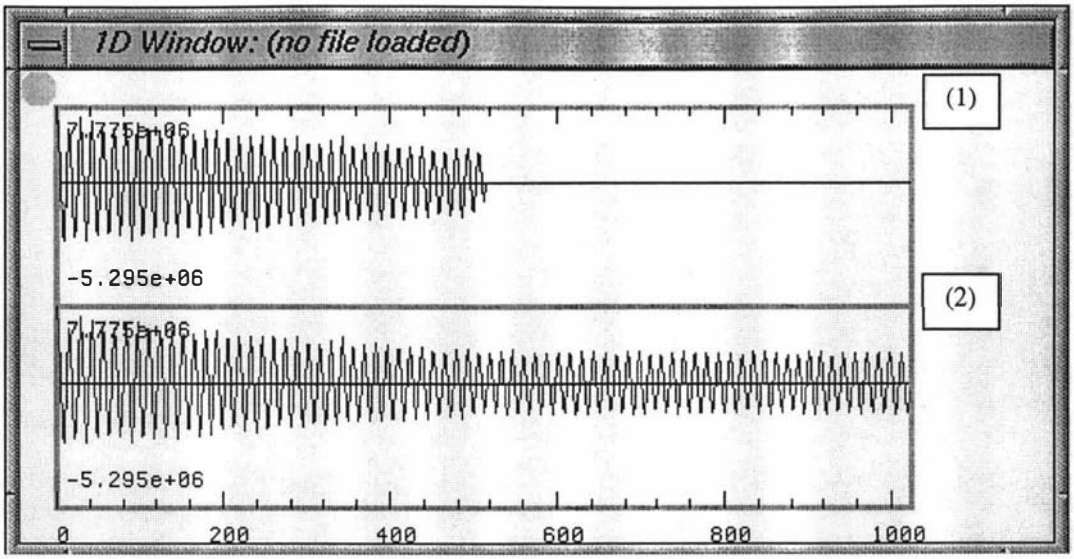


Figure 6.17: A signal with and without applying LP. (1) A truncated time-domain data set due to zero filling. (2) The result after linear prediction has been applied.

Unexpectedly, the predicted part seems to have stopped decaying, indicating a discrepancy in the prediction. Not surprisingly, a comparison of the NMR spectrum of the two data sets still shows a better resolution in the spectra processed with the LP technique (Figure 6.18).

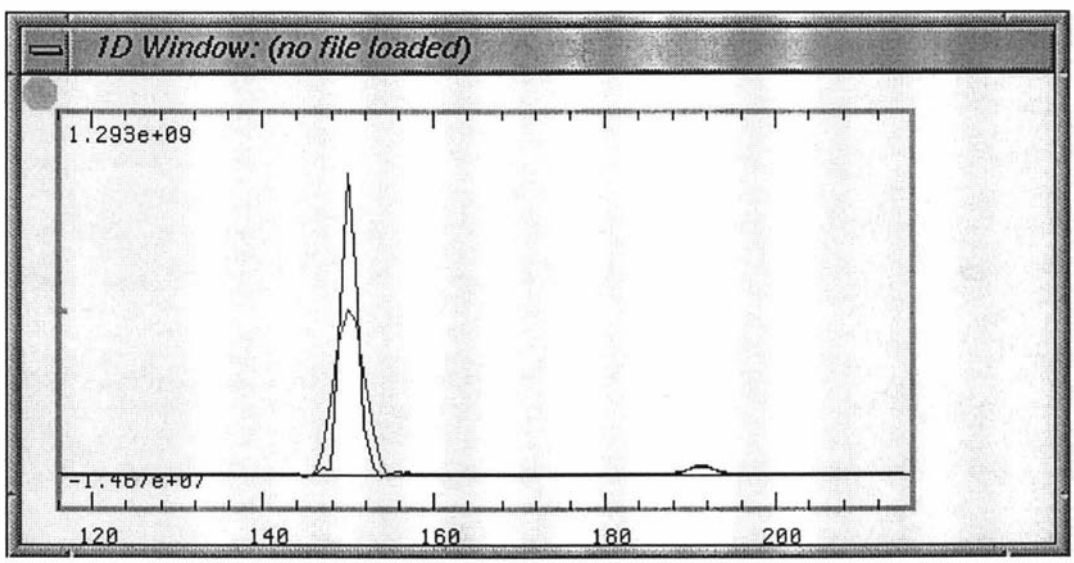


Figure 6.18: A figure that shows the advantage of linear prediction. The spectrum with narrower peaks was produced using the LP technique, and that with broader peaks was processed using conventional apodisation technique. Similar differences can be shown for the peaks on the right when zooming into a smaller region around them.

However, LP cannot provide a universal solution to predict every part successfully, especially for those data that are sensitive to perturbation. The prediction of the first few points, for example, may or may not be advantageous, depending on the case and how LP is applied. In fact, if the baseline is acceptable, LP may only make it worse. An example of this effect is shown in the next diagram.

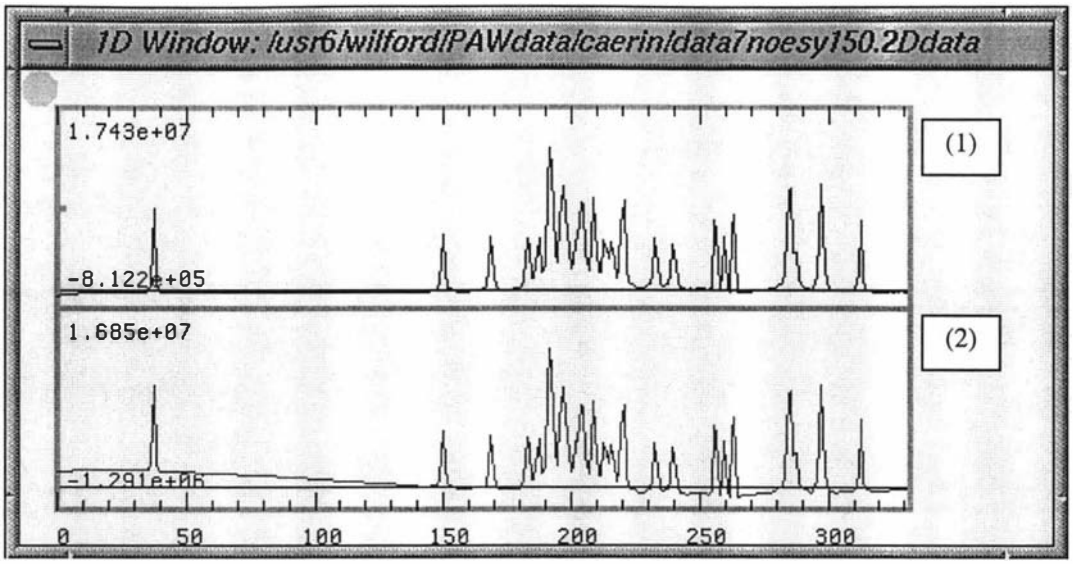


Figure 6.19: A figure that shows a negative effect of LP. (1): A quality spectrum obtained without applying LP. (2): A spectrum produced by incorrect application of LP to predict the first two points of the time-domain data. This is because spectral baselines are very sensitive to any discrepancy in the first few points, even if small.

In short, LP is more successful at predicting the end of time-domain data rather than the beginning. It will only be useful for predicting the first few points if they are badly distorted.

6.3.8 Water signal suppression

Samples used in protein NMR spectroscopy often contain around 90% H_2O, i.e., 50 M in terms of concentration. Assume that the sample concentration is 5 mM. Then, the water concentration is 10,000 times larger. Therefore, many protein NMR experiments rely heavily on water suppression techniques to obtain good quality spectra.

If water suppression fails in an experiment, the time-domain signal will contain a dominant low frequency water signal, as shown in Figure 6.20.

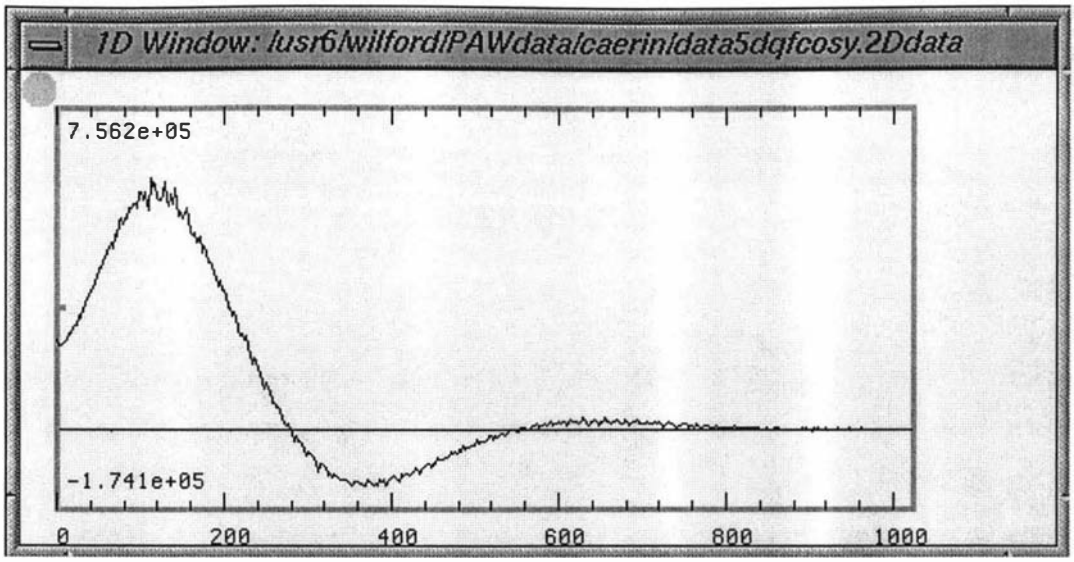


Figure 6.20: An FID with predominant water signal. The low-frequency curve is the water signal. The oscillating NMR data of the protein (carried by the water curve) are around ten times weaker in this case.

The pattern of the water signal was observed empirically to obey a curve only slightly different from the simple exponential decay function:

$$y(t) = A\cos(\omega t+\varphi)\exp(-\eta t^m) \qquad (6.3.8.1)$$

Here A, ω, φ, η and m are the amplitude, frequency, phase shift, decay rate and degree, respectively. These parameters can be determined by a non-linear fitting process. Our work shows that m is between 1.2 and 1.5.

Fortunately, the water signal can often be removed in data processing, as explained in the next section.

6.3.8.1 Principle and method

The principle of time-domain water signal suppression is to remove the water signal from the data. The focus of this task is therefore to search for a curve that best fits the water signal.

Unfortunately, it is often the case that the water signal cannot be fitted with a simple function. For example, at the beginning of the signal, the decay can be much slower than expected. This is probably because of radiation damping, a phenomenon caused an additional magnetic field induced by the rapid rotation of the ensemble water magnetisation vector.

Some data are even corrupted as a result of receiver overflow due to strong water signals (Figure 6.21).

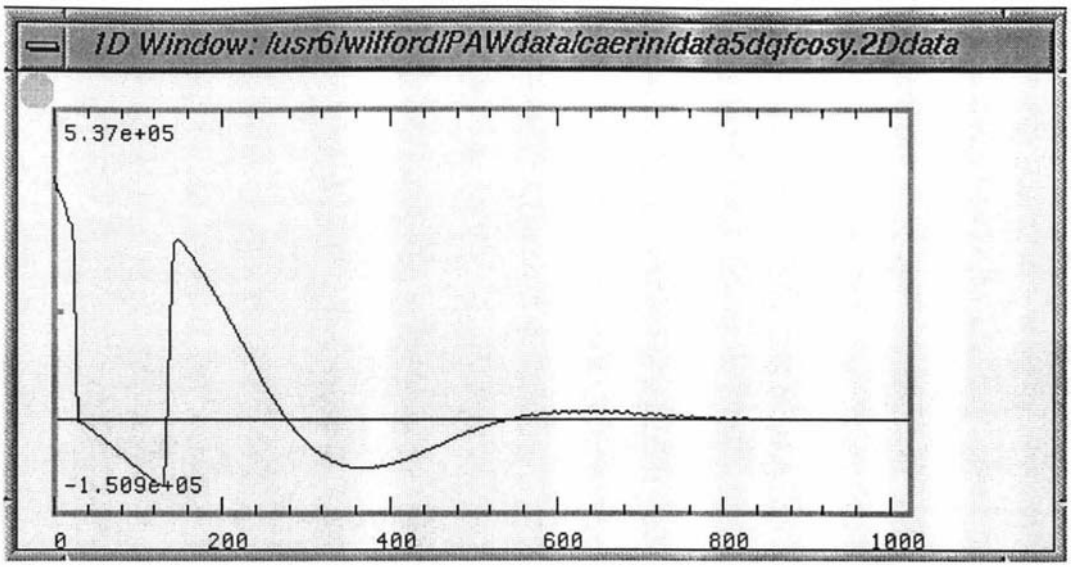


Figure 6.21: A corrupted FID due to a strong water signal that resulted in receiver overflow.

Unfortunately, there is no 'best solution' to eliminate water signals — every method is susceptible to failure. The most reliable fitting method is therefore to work directly with the data rather than any proposed models for the data. For this reason, the water suppression method of PAW is based on a piece-wise spline that is fitted to a smooth curve obtained by averaging the FID signal locally, as described next.

6.3.8.2 Program implementation

In PAW, the time-domain water suppression routine suppresses water signals by following these steps:

1. Construct a curve $y_1(t)$ for which every point is obtained by averaging a number of surrounding points (say, 20) in the FID. (Note that this does not replace the original FID.)
2. Construct a second curve $y_2(t)$ by fitting the first 5% of the original FID with a polynomial using the singular-value decomposition (SVD) method [Press *et al.* 1986].[2]
3. Calculate the first derivative of $y_2(t)$ at about 3% of t_{max}.
4. Sample a number of points, say 100, from $y_1(t)$ to a buffer, started from about 3% and finished at about 97% of t_{max}.
5. Construct a third curve $y_3(t)$ for the points sampled in Step 4 using a cubic spline-function,(with the first derivative for the first point being that obtained in Step 3, and the last point, zero).
6. Fill the beginning part of $y_3(t)$ with that of $y_2(t)$, and the end part with the value of $y_3(t)$ at 97% of t_{max}.
7. Subtract the original FID from $y_3(t)$ to obtain a filtered signal $y_4(t)$.

[2] This is because neither end can be averaged properly.

6.3.8.3 User-interface

To perform water suppression with PAW, users only need to issue the command wsp or click on the [WaterSup] button in the *1D-processing Toolbox*.

6.3.8.4 Effectiveness of PAW's time-domain water-suppression technique

The following diagrams show a water-dominated FID and the real parts of two spectra processed with and without water signal suppression.

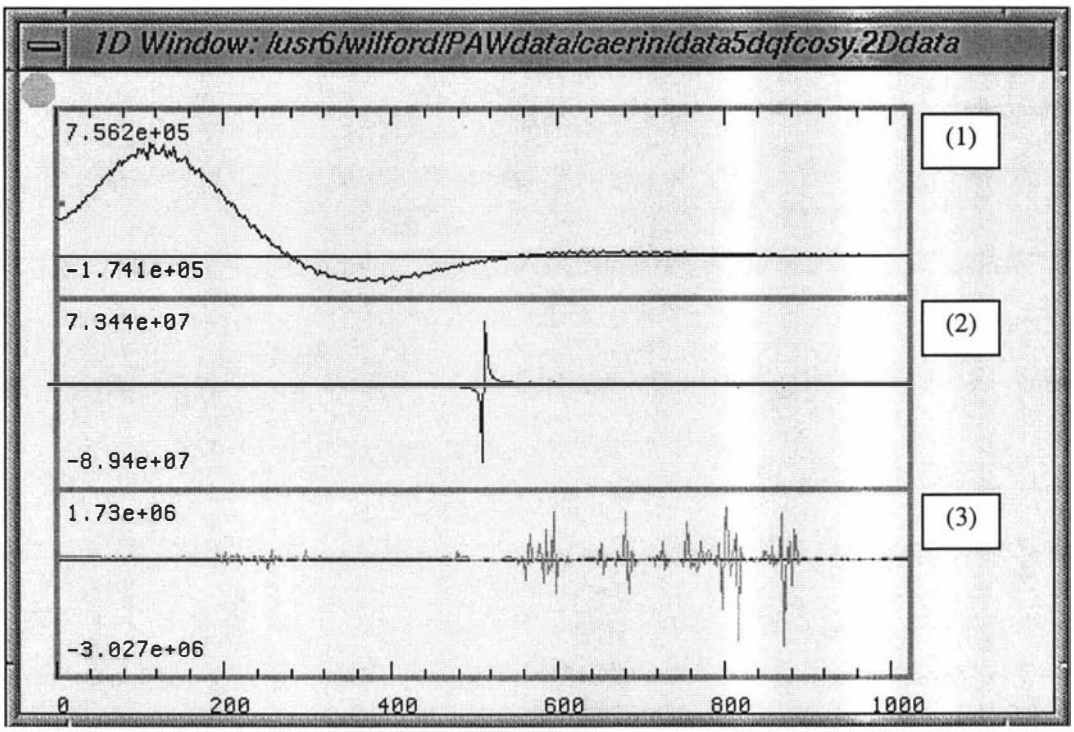


Figure 6.22: The effect of time-domain water-signal suppression. (1) A water-dominated FID. (2) The real part of a spectrum processed without water suppression. (3) The real part of a spectrum obtained by applying the water suppression command three times.

The next figure shows an expanded region of Plot 2 and 3 in Figure 6.22.

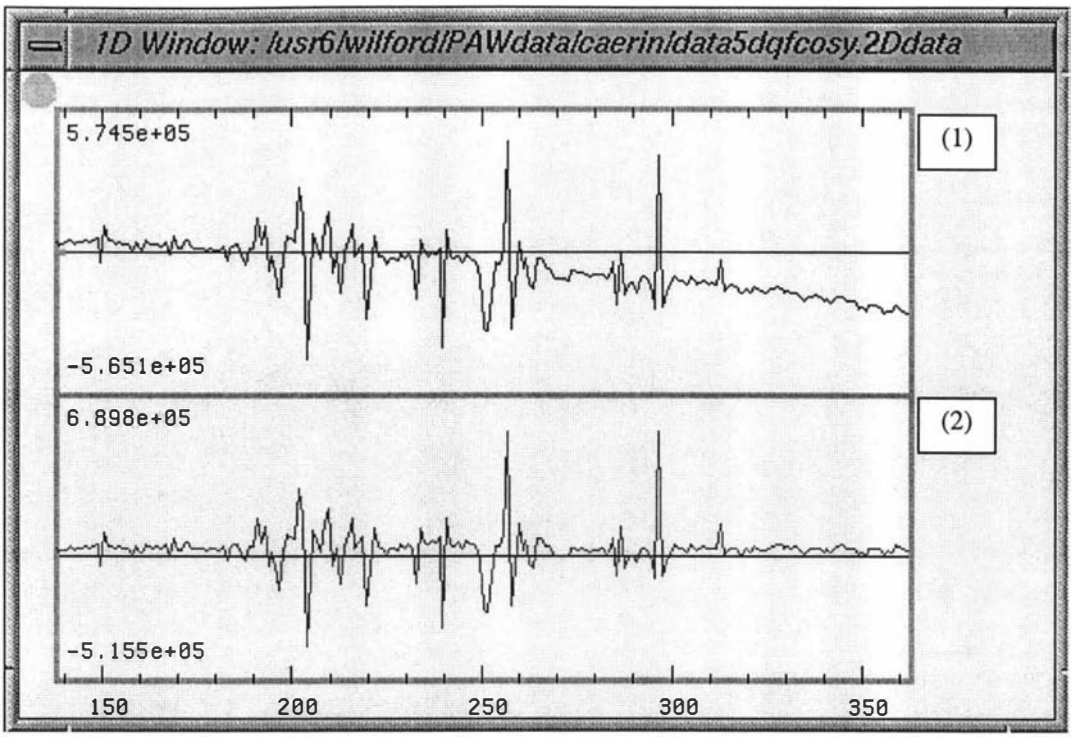


Figure 6.23: An expanded region of Plot 2 and 3 in Figure 6.22 (water region not included). Note how the baseline has also been corrected in the spectrum treated with water suppression (bottom).

Note that PAW also provides a command to suppress the water signal from the frequency domain spectrum. The difficulty is that fitting a peak in the frequency domain requires that the peak be properly phased and the mathematical expression for the peak be exactly known. Otherwise, the suppression will result in a positive peak surrounded by two small negative peaks, as illustrated in the next two diagrams.

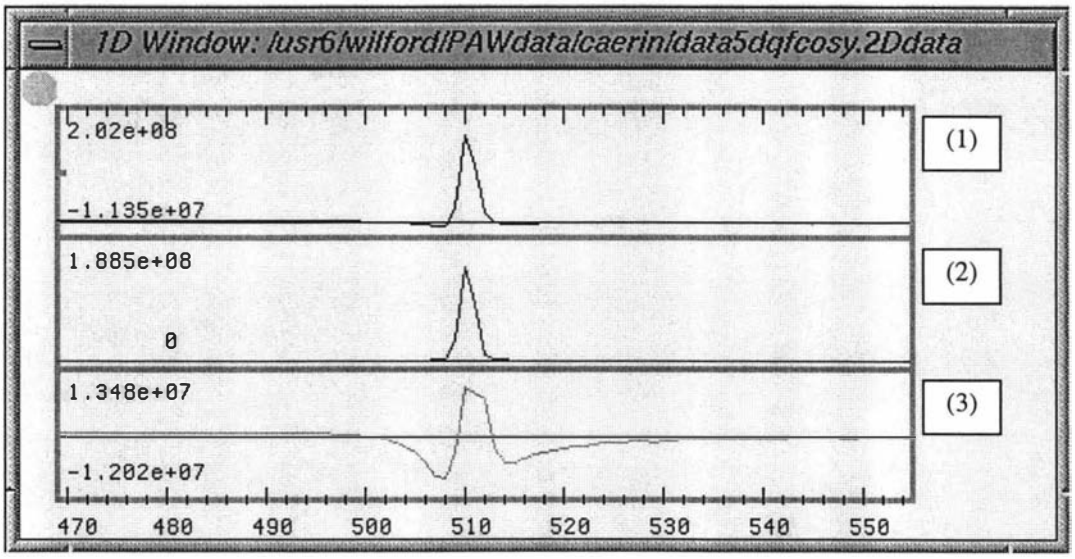


Figure 6.24: Water suppression in the frequency domain. (1) A water-peak. (2) A Gaussian peak fitted to the water peak. (3) The difference between the last two peaks.

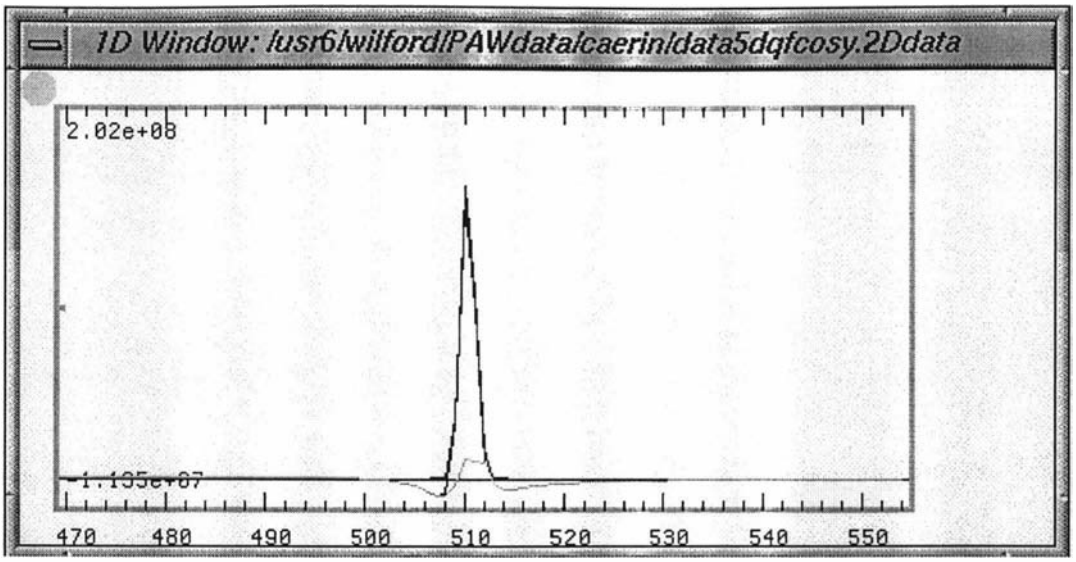


Figure 6.25: An overlapping view of the three plots in the last figure. Although the positive part of the calculated peak fits well with the water peak, the subtraction result (underneath) still shows a clear discrepancy between the two.

6.3.9 Fast Fourier transform (FFT) and its inverse (IFT)

The fast Fourier transform (FFT) is at the core of all spectral processing today. Perhaps the FFT can be considered as Wonder #1 in computational mathematics, not simply because it is extremely fast in comparison with the normal discrete FT, but also because of the elegant and concise programming used.

Understanding the FFT algorithm, however, requires a good background in computational mathematics, and is much more complicated than one might think by just looking at the simplicity of the mathematical derivation found in many textbooks.

In terms of speed, it is astonishingly fast. For example, with a 100 MHz (about 10 mega-flops[3]) computer, it can be shown that the time required for an FFT of a 1024-point 1D data set is about 3.8 milliseconds; whereas the normal FT will be about 1000 times slower. If the FFT were not available, the time required to transform a 1024×1024 2D data set with a 100MHz computer would be 7,600 seconds!

6.3.9.1 Theory

The theory of FFT and various algorithms has been discussed extensively in many articles [see e.g., Bracewell 1965, Brigham 1974, Press *et al.* 1986, Nussbaumer 1982, or Eccles 1987]. In brief, the FFT is a discrete Fourier transform obtained using the formula:

$$F(k) = \sum_{j=0}^{N-1} W^{jk} s_j$$

[3] Number of floating-point operations per second.

$$
\begin{aligned}
&= \sum_{j=0}^{N/2-1} W^{2jk} s_{2j} + \sum_{j=0}^{N/2-1} W^{(2j+1)k} s_{2j+1} \\
&= \sum_{j=0}^{N/2-1} W^{2jk} s_{2j} + W^k \sum_{j=0}^{N/2-1} W^{2jk} s_{2j+1} \\
&= F_{even}(k) + W^k F_{odd}(k)
\end{aligned}
\qquad (6.3.9.1.1),
$$

where W is a complex number:

$$W = \cos(2\pi / N) + i\sin(2\pi / N) = e^{2\pi i / N} \qquad (6.3.9.1.2).$$

6.3.9.2 Program implementation

When processing a 2D NMR data set, the FFT is often applied thousands of times. For this reason, the program PROSPA [Eccles 1995] further optimised the FFT algorithm by constructing two trigonometric look-up tables that depend only on the data size. This idea is also adopted by PAW, which gives a 20% speed improvement.

6.3.9.3 User interface

The command buttons for the fast Fourier transform and its inverse are [FFT] and [IFT], which are in the *1D-processing Toolbox*. The corresponding commands are fft and ift.

6.3.10 Real Fourier transform (RFT)

The RFT converts a single real function from the t-domain into the f-domain [Press 1986]. In NMR spectroscopy, it is used to process data collected using the *time-proportional phase increment* (TPPI) method [Redfield & Kunz 1975; Bodenhausen *et al.* 1977, Drobny *et al.* 1979, Bodenhausen *et al.* 1980, Marion & Wüthrich 1983].

6.3.10.1 Theory

The collection of a single real function requires special treatment in NMR spectroscopy. For example, the TPPI method collects successive real values in the t-domain such that

$$s(t) = \{s_x(0), s_y(\Delta_t), -s_x(2\Delta_t), -s_y(3\Delta_t), s_x(4\Delta_t), s_y(5\Delta_t), -s_x(6\Delta_t), -s_y(7\Delta_t), \ldots\} \qquad (6.3.10.1.1),$$

where Δ_t is the sampling interval. This real series consists of the real data obtained by applying a 90° phase-increment to subsequent complex values in a time series. In complex analysis, a −90° phase-increment of a complex number is equivalent to multiplying the number by $-i$. For example, multiplying a complex element $a(t)+ib(t)$ by $-i$ results in the next three consecutive elements of a time series:

$$-i\,a(t+\Delta_t)+b(t+\Delta_t), \quad -a(t+2\Delta_t)-ib(t+2\Delta_t), \quad \text{and} \quad ia(t+3\Delta_t)-b(t+3\Delta_t).$$

Hence, the four real parts are

$a(t)$, $b(t+\Delta_t)$, $-a(t+2\Delta_t)$, and $-b(t+3\Delta_t)$.

When $t = 0$, these correspond to the first four components in (6.3.10.1.1).

Note that a single real-function is not simply the real part of a complex function. If a normal complex FID $s_{complex}(t) = s_x(t) + is_y(t)$ is converted into a real signal $s_{real}(t)$ such that

$$s_{real}(t) = \{s_x(0), s_y(\Delta_t), -s_x(2\Delta_t), -s_y(3\Delta_t), \ldots\} \quad (6.3.10.1.2).$$

Then, applying an RFT will give an 'expanded' spectrum with each peak located twice as far from the centre as they normally would be. This is because the use of the RFT in NMR spectroscopy requires the sampling rate be increased by a factor of two.

6.3.10.2 Program implementation

The program is similar to that described for the FFT of a single real function in [Press 1986].

6.3.10.3 User interface

The command buttons for the real Fourier transform is [RFT], which is in the *1D-processing Toolbox*. The corresponding command is rft.

6.3.11 Hilbert transforms (HT)

In NMR spectral processing, some processes require a complex data set. If only the real part is available, the imaginary part must be generated with a Hilbert transform (HT). This is frequently applied during the phasing of 2D data sets for which only the real parts are kept.

6.3.11.1 Theory

There are two formulae for the Hilbert transform (HT) [see, e.g., Bracewell 1986], generating the real part of a complex data set from its imaginary part, and the imaginary part from its real part:

$$\mathrm{Re}\{H(\omega)\} = \frac{1}{\pi}\int_{-\infty}^{\infty}\frac{\mathrm{Im}\{H(\omega')\}}{\omega-\omega'}d\omega' \quad (6.3.11.1.1).$$

$$\mathrm{Im}\{H(\omega)\} = -\frac{1}{\pi}\int_{-\infty}^{\infty}\frac{\mathrm{Re}\{H(\omega')\}}{\omega-\omega'}d\omega' \quad (6.3.11.1.2).$$

where

$$H(\omega) = \int_{-\infty}^{\infty} S(t)\exp(i\omega t)dt \quad (6.3.11.1.3).$$

The formulae imply that, to obtain one value in a data set, a discrete integration is required. Calculating a full set of data must therefore be relatively slow; yet the results

cannot be accurate because the integral is improper. This will be called the HT(slow) in this thesis.

Fortunately, there is a must faster way to obtain approximate results using Fourier transform [Eccles, 1995], which will be called the HT(fast) in this thesis.

Before understanding the mathematics, it is helpful to have a look at the 1D Fourier pairs in the next three figures. In each of the figures, Plot 1 and 2 are the real and imaginary parts of a complex NMR spectrum, respectively; Plot 3 and 4 are the corresponding inverse Fourier transform (IFT) results. The first figure was drawn from a normal 1D NMR data set, the second and third, from the complex conjugate and the real-only data of the complex NMR spectrum in the first figure. The most obvious differences are the distribution of the significant parts in the IFT results. Those in the first figure are on the left, those corresponding to the complex conjugate, on the right, and those corresponding to the real-only data, on both sides.

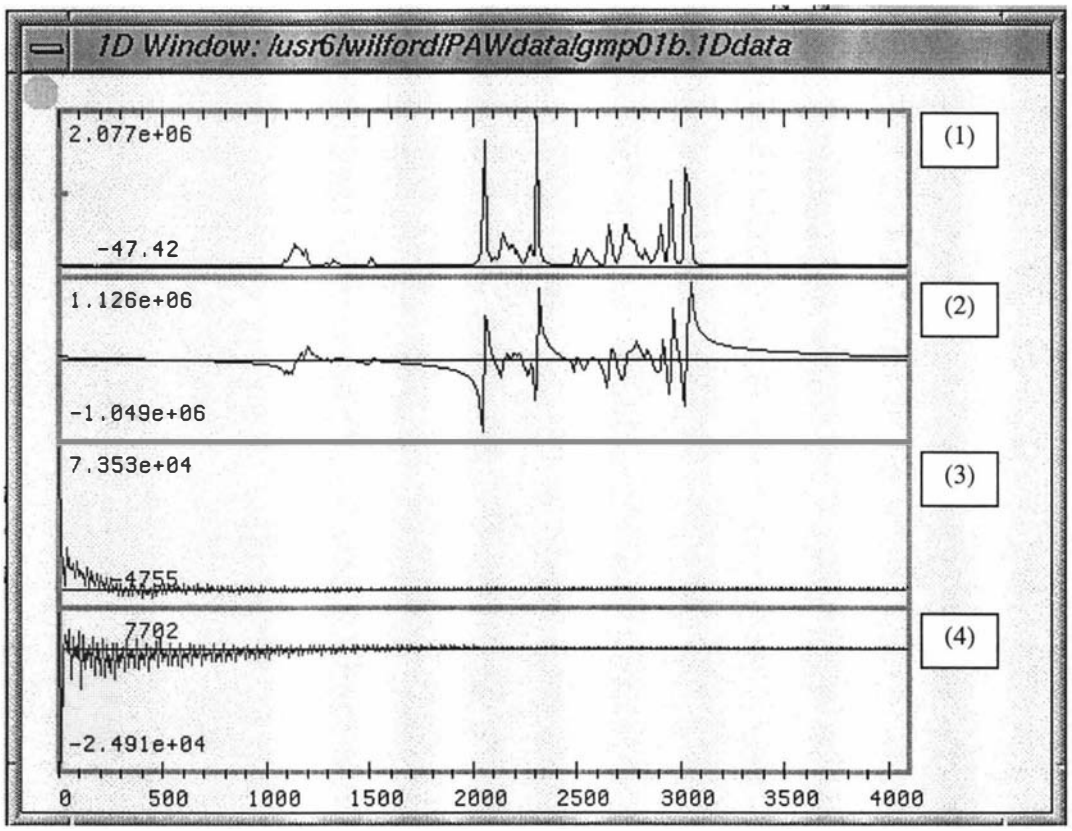


Figure 6.26: The real and imaginary parts (Plot 3 and 4) obtained by applying IFT to a complex NMR spectrum (Plot 1 and 2).

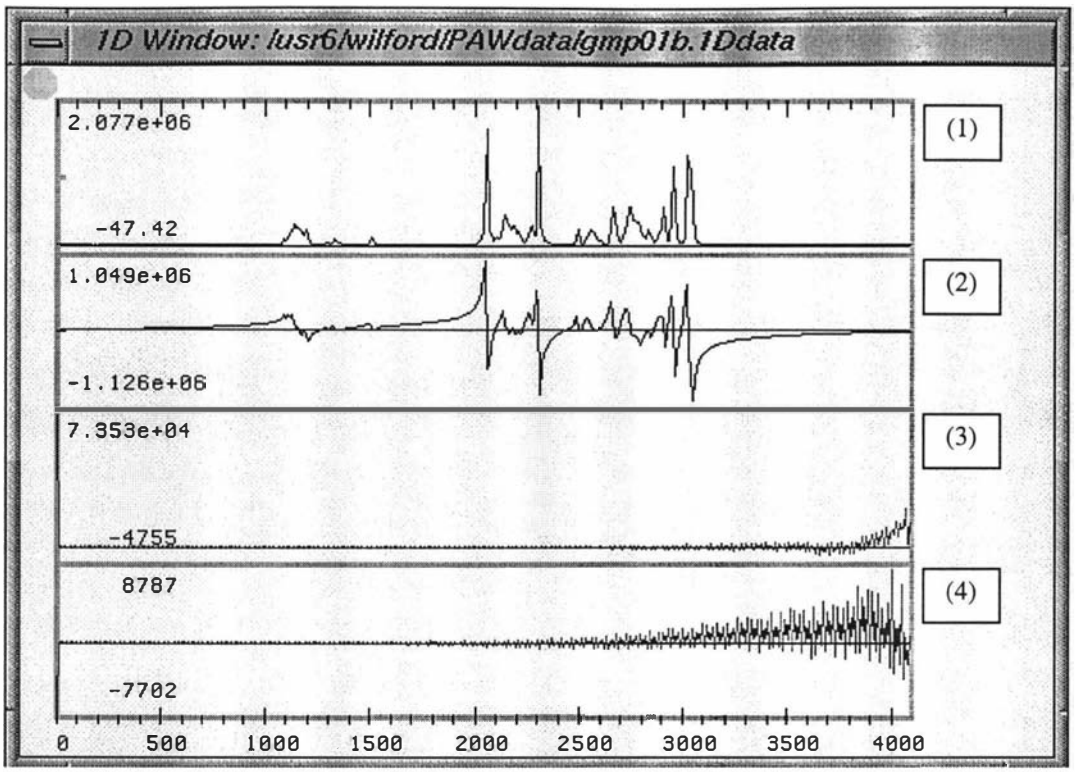


Figure 6.27: The real and imaginary parts (Plot 3 and 4) obtained by applying IFT to the complex conjugate (Plot 1 and 2) of the spectrum in Figure 6.26.

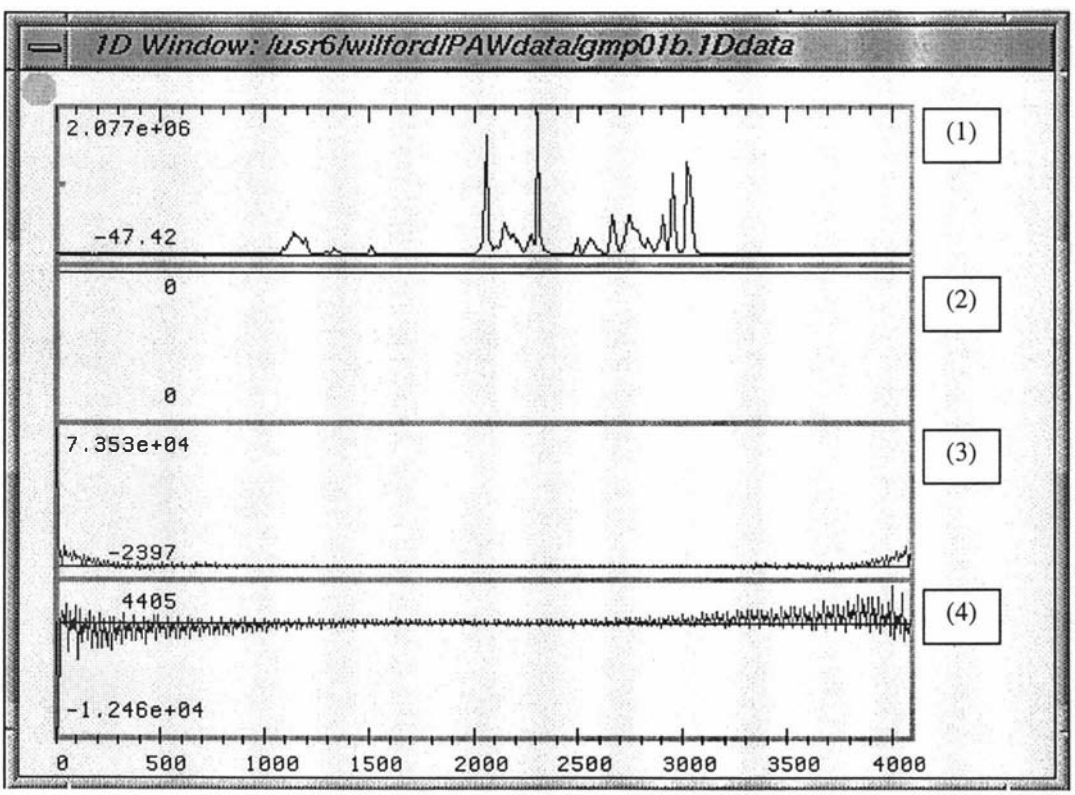


Figure 6.28: The real and imaginary parts (Plot 3 and 4) obtained by applying IFT to the real-only data (Plot 1 and 2) of the spectrum in Figure 6.26.

Mathematically, there is a simple relation as follows:

$$H(\omega) + [H(\omega)]^* = 2\,\mathrm{Re}\{H(\omega)\} \qquad (6.3.11.1.4).$$

That is, the summation of a complex function and its conjugate equals twice its real part. Hence,

$$\mathcal{F}^{-1}\{2\,\mathrm{Re}\{H(\omega)\} = \mathcal{F}^{-1}\{H(\omega)\} + \mathcal{F}^{-1}\{[H(\omega)]^*\} \qquad (6.3.11.1.5).$$

A typical characteristic of most NMR data is that the significant part of $\mathcal{F}^{-1}\{H(\omega)\}$ lies on the left, and that of $\mathcal{F}^{-1}\{[H(\omega)]^*\}$ lies on the right, as shown in the first two figures. Therefore, the $\mathcal{F}^{-1}\{H(\omega)\}$ can be obtained approximately by first interpolating $2\mathrm{Re}\{H(\omega)\}$ to double its size (with any interpolation technique) before applying an IFT, and then retaining only the first half of the $\mathcal{F}^{-1}\{2\,\mathrm{Re}\{H(\omega)\}$. This approximation has proved to be sufficient for phasing. Theoretically, if the second half of $\mathcal{F}^{-1}\{H(\omega)\}$ contains only zeros, this method will give an exact result.

6.3.11.2 Program implementation

PAW provides two routines for applying the Hilbert transform to convert real-only data into complex. One of them performs the HT using the faster transform method described in the last subsection, and the other one, the slow method.

To avoid unexpected operational mistakes, PAW also automatically applies an HT before performing a phasing operation if the first few data points are detected to be zeros.

6.3.11.3 User interface

The command buttons for the two methods of the HT are [HT(fast)] and [HT(slow)], which may be found in the *1D-processing Toolbox*. The corresponding commands are ht and hts.

6.3.11.4 Hilbert transform errors

As mentioned previously, an inevitable discrepancy arises when performing the HT because of the improper integral involved. The figure below is an overlap of a fast HT transform result and the original data set.

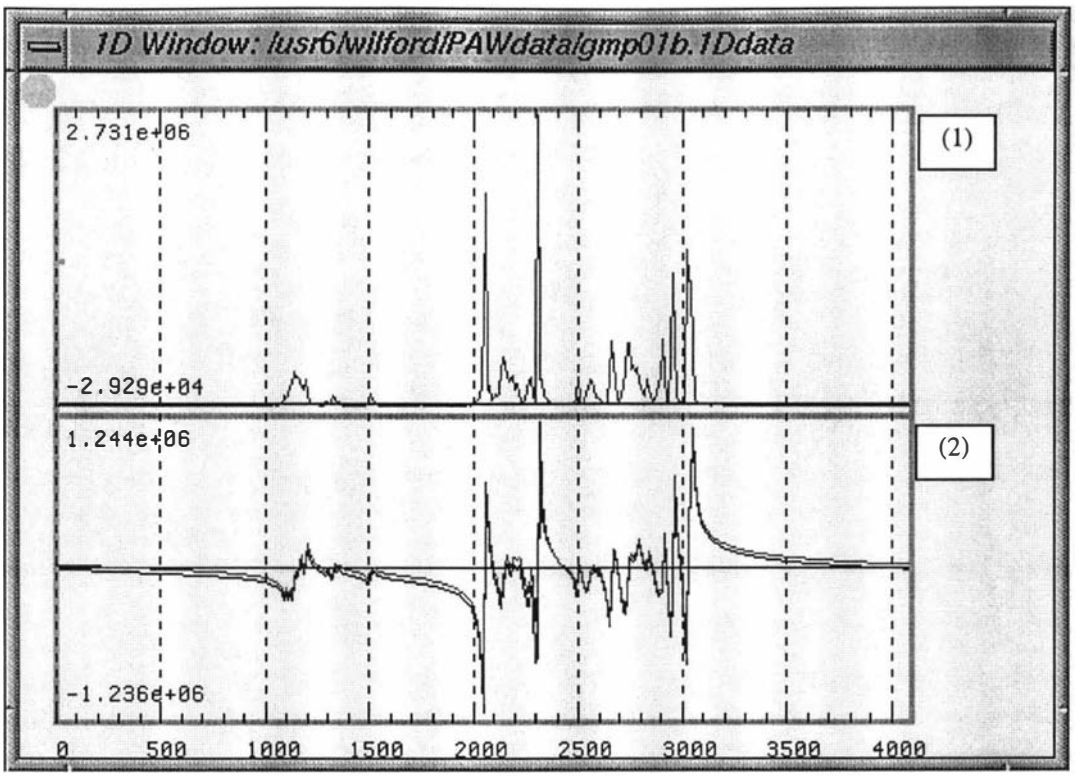


Figure 6.29: An overlapping view of the HT result and the original complex spectrum. (1) The real part is exactly the same in both cases. (2) The imaginary part of the HT result is only slightly higher in this case due to the significant decay of the time-domain data.

Not surprisingly, there are some differences between the imaginary parts. The differences, however, can be more significant if there is little decay in the signal as is shown in Figure 6.30.

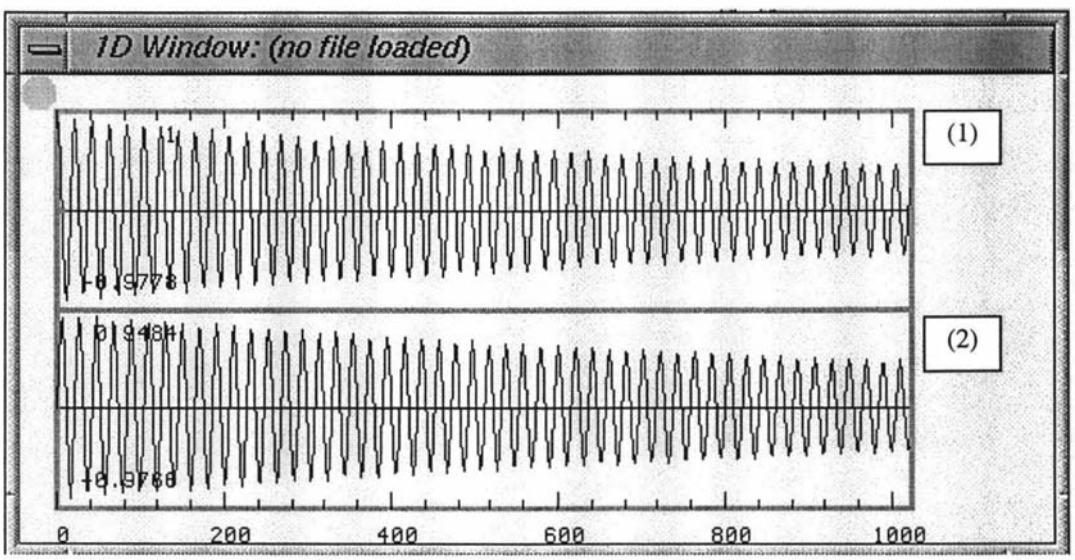


Figure 6.30: A generated 1D data set with little decay.

For comparison, the next figure shows an expanded view of three overlapping complex spectra.

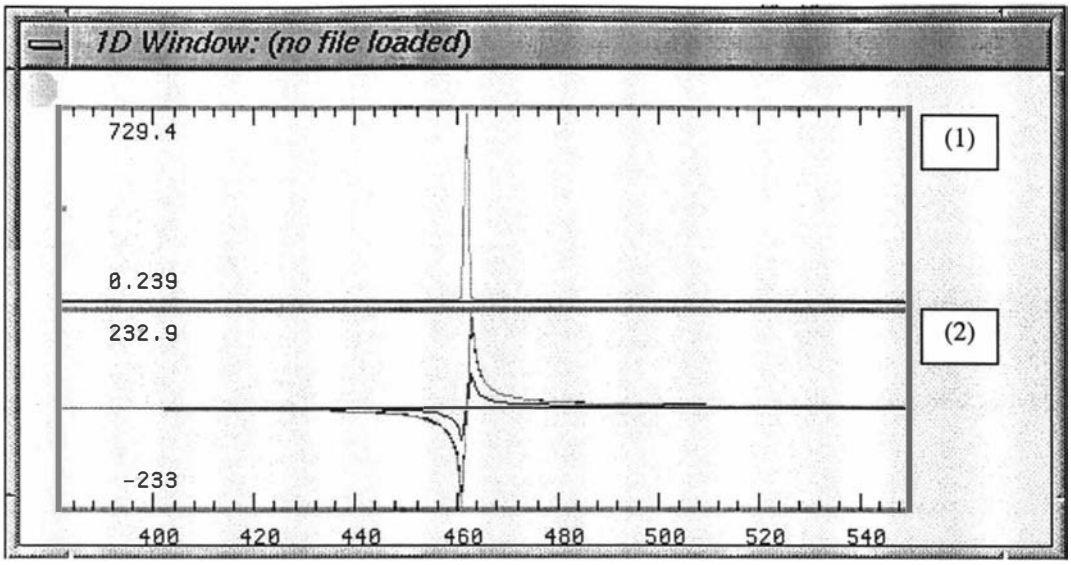


Figure 6.31: The FFT and HT results of a signal with little decay. (1) The real parts show no difference. (2) In the imaginary parts, the one with the smallest amplitude is the original data; the other two, which are virtually superimposed, were obtained by applying the ht and hts commands to the time-domain real data, respectively.

The result shows that, due to the limited computational accuracy, the HT(slow) does not seem to be better even for a data set with little decay. To summarise, HT(fast) is an efficient alternative for the HT, and is ideal when applied to a rapidly decaying data set.

6.3.12 Phase correction

In NMR data processing, only two kinds of phase problems are normally corrected: the zeroth- and first-order phase errors. Nonlinear phase shifts such as those due to the transverse interference originating from insufficient delay-time between experiments [Ernst *et al.* 1987] are usually not seen in NMR data and hence can be ignored.

6.3.12.1 Frequency-independent phase shifts

Frequency-independent phase shifts are very common in NMR data, and are primarily due to the electronics between probe, transmitter and digitiser.

As a consequence, the FID collected becomes

$$s(t) = M_0 \exp(i\alpha_0) \exp(-i\Omega_0 t) \exp(-t/T_2) \qquad (6.3.12.1.1)$$

In other words,

$$s(t) = s_{ideal}(t) \exp(i\alpha_0). \qquad (6.3.12.1.2)$$

The equation shows a constant phase-shift of α_0 from the ideal solution.

6.3.12.2 Frequency-dependent phase shifts

A common reason for a frequency-dependent phase-shift is the incorrect timing of device switching. Assuming that the time delay is τ and the signal collected is

$$s(t) = M_0 \exp(-i\Omega_0(t - \tau)) \exp(-(t - \tau)/T_2) \quad (6.3.12.2.1)$$

i.e.,

$$s(t) = s_{ideal}(t) \exp(i\Omega_0\tau) \exp(\tau/T_2). \quad (6.3.12.2.2)$$

This corresponds to a change in amplitude from the ideal solution by a factor of $\exp(\tau/T_2)$ and a phase-shift of α_1:

$$\alpha_1 = \Omega_0\tau \quad (6.3.12.2.3)$$

Here, α_1 is linearly dependent on Ω_0, thereby depending on the Larmor frequency ω_0 because $\Omega_0 = \omega_0 - \omega_{rotate}$.

For a system containing spins that have more than one Larmor frequency, the following general notation can be used

$$\alpha_{1,j} = \Omega_{0,j}\tau \quad (6.3.12.2.4)$$

Accordingly, $s(t)$ and $s_{ideal}(t)$ must also be changed:

$$s_j(t) = s_{ideal,j}(t) \exp(i\Omega_{0,j}\tau) \exp(\tau/T_{2,j}). \quad (6.3.12.2.5)$$

Another possible reason for the frequency-dependent phase shift is the so-called *off-resonance effect*. In the rotating frame, the off-resonance frequency Ω_0 corresponds to an offset field $\mathbf{B}_{offset}$:

$$\mathbf{B}_{offset} = -\Omega_0/\gamma = \mathbf{B}_0 + \omega_{rotate}/\gamma \quad (6.3.12.2.6)$$

Here, Ω_0/γ is often comparable to the applied r.f. field $\mathbf{B}_1$. In a frame rotating at ω_0 rad/sec where $\mathbf{B}_1$ becomes static, there is an effective field $\mathbf{B}_{eff}$ acting on the nuclear spins in the system (Figure 6.32):

$$\mathbf{B}_{eff} = \mathbf{B}_{offset} + \mathbf{B}_1 \quad (6.3.12.2.7)$$

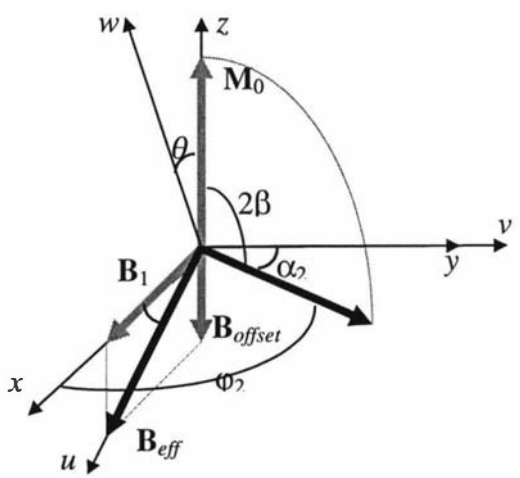


Figure 6.32: B_{eff} in a frame rotating at ω_0 rad/sec. The (u, v, w) system is obtained by a $\theta°$ rotation about the y-axis from the (x, y, z) system. The $\mathbf{B}_{eff}$ is a combined vector of $\mathbf{B}_1$ and $\mathbf{B}_{offset}$. The tip angle is 2β. The frequency-dependent phase shift due to the off-resonance effect is α_2 or $(90° - \varphi_2)$.

The effect of $\mathbf{B}_{eff}$ on $\mathbf{M}_0$ can be calculated by three consecutive transformations of $\mathbf{M}_0$:

1) a coordinate rotation about the y-axis by θ^o to a (u, v, w) system in which $\mathbf{B}_{eff}$ sits on the u-axis (Figure 6.32),
2) a vector rotation about the u-axis by $2\beta^o$ that corresponds to the effect of the pulse Hamiltonian operator in the (u, v, w) system, and
3) a reverse co-ordinate rotation about the y-axis by $-\theta^o$ to the (x, y, z) system.

The equation for the process is

$$\mathbf{M}_{0+} = \mathbf{R}_y^{-1}(\theta)\,\mathbf{R}_u(2\beta)\,\mathbf{R}_y(\theta)\,\mathbf{M}_{0-} \tag{6.3.12.2.8}$$

Here, $\mathbf{R}_y^{-1}(\theta)$, $\mathbf{R}_u(2\beta)$ and $\mathbf{R}_y(\theta)$ are rotational transform operators such that

$$\mathbf{R}_y(\theta) = \begin{bmatrix} \cos\theta & 0 & \sin\theta \\ 0 & 1 & 0 \\ -\sin\theta & 0 & \cos\theta \end{bmatrix} \tag{6.3.12.2.9}$$

$$\mathbf{R}_u(2\beta) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos(2\beta) & -\sin(2\beta) \\ 0 & \sin(2\beta) & \cos(2\beta) \end{bmatrix} \tag{6.3.12.2.10}$$

Therefore,

$$\mathbf{M}_{0+} = M_0 \begin{bmatrix} \sin\theta\cos\theta(1-\cos 2\beta) \\ -\cos\theta\sin 2\beta \\ \sin^2\theta + \cos^2\theta\cos 2\beta \end{bmatrix} \tag{6.3.12.2.11}$$

This result is consistent with that obtained using the product operator formalism. Note that the angle $2\beta^o$ is negative because the rotation of $\mathbf{M}_0$ about $\mathbf{B}_{eff}$ is left-handed, and hence $\sin(2\beta)$ is negative.

As a consequence,

$$\tan(\alpha_2) = \frac{M_x}{M_y} = \frac{(1-\cos 2\beta)}{-\sin 2\beta}\sin\theta = -(\tan\beta)\times\frac{\Omega_0}{\sqrt{\omega_1^2 + \Omega_0^2}} \tag{6.3.12.2.12}$$

Here ω_1 is the angular velocity of $\mathbf{M}_0$ rotating about $\mathbf{x}$, and α_2 (see Figure 6.32) is the phase shift that depends on Ω_0. The last expression in the last equation was obtained by applying a trigonometric identity.

For $2\beta \simeq \pi/2$, it can be shown by the Taylor's expansion of an arctan function that

$$|\alpha_2| \simeq (\tan\beta)\times\frac{\Omega_0}{\omega_1}\,\text{rad} \tag{6.3.12.2.13}$$

For example, if the 90° pulse width is 11 μsec, then, $\omega_1 = 2\beta/t_p = (\pi/2)/(11\times10^{-6})$ rad/sec. If the spectral width is 8 kHz (i.e., 4 kHz on each side of the spectrum), it can be calculated that α_2 may vary from 0 to $(4\times11\times10^{-6})\times4000$ rad (or 10.084°) following a 90° pulse, such an angle is by no means insignificant.

The solution to the Bloch equation due to the above changes in the initial conditions is

$$s(t) = M_0\exp(i\varphi_2)\exp(-i\Omega_0 t)\exp(-t/T_2) \tag{6.3.12.2.14}$$

that is,

$$s(t) = s_{ideal}(t)\exp(i\varphi_2). \quad (6.3.12.2.15)$$

Here, as shown in Figure 6.32,

$$\varphi_2 = \pi/2 - \alpha_2 \quad (6.3.12.2.16)$$

Like α_2, the phase shift φ_2 also depends linearly on Ω_0. Therefore, in a case that involves n resonance frequencies Ω_1, Ω_2, ... Ω_n, the off-resonance effect induces phase shifts that depend linearly on each Ω_j for each peak.

6.3.12.3 Zeroth- and first-order phase corrections

The *combined phase shift* from the results in the last two subsections is

$$\varphi = \alpha_0 + \alpha_1 + \pi/2 - \alpha_2 \quad (6.3.12.3.1)$$

Here, $(\alpha_0 + \pi/2)$ is frequency-independent and $(\alpha_1 - \alpha_2)$ is frequency-dependent. Hence,

$$s(t) = \exp[i\varphi]\exp[\tau/T_2]\, s_{ideal}(t) \quad (6.3.12.3.2)$$

After Fourier transform, it becomes

$$S(\Omega) = \exp[i\varphi]\exp[\tau/T_2]\, S_{ideal}(\Omega) \quad (6.3.12.3.3)$$

In both expressions, the phase shifts to be corrected are embedded in $\exp[i\varphi]$.

The way to correct the *combined* phase shift is therefore to multiply the signal $s(t)$ or the corresponding spectrum $S(\Omega)$ with $\exp[-i\varphi]$ so that the factor $\exp[i\varphi]$ is cancelled.

6.3.12.4 Program implementation

In practice, phase errors are usually corrected in two steps after FT.

The constant phase shift $(\alpha_0 + \pi/2)$ is corrected by multiplying the frequency-domain data with $\exp[-i(\alpha_0 + \pi/2)]$. This process is called *zeroth-order phase correction.* Mathematically, it can be express as follows:

$$\exp[i(\alpha_0 + \alpha_1 + \pi/2 - \alpha_2)]\exp[-i(\alpha_0 + \pi/2)] = \exp[i(\alpha_1 - \alpha_2)] \quad (6.3.12.4.1)$$

The correction of the frequency-dependent phase shift is done empirically by multiplying the frequency-domain data with $\exp[i\beta_1(\Omega - \Omega_{pivot})]$, where both β_1 and Ω_{pivot} can be dynamically set by users. This process is called *first-order phase correction.* The quantity Ω_{pivot}, called the *pivot frequency*, is a point at which the phase remains the same during the operation.

Mathematically, the condition for 1st-order phase correction is as follows:

$$\exp[i(\alpha_1 - \alpha_2)]\exp[i\beta_1(\Omega - \Omega_{pivot})] = 1 \quad (6.3.12.4.2)$$

The best β_1 to eliminate the frequency-dependent phase shift $(\alpha_1 - \alpha_2)$ should therefore be related to α_1 and α_2 such that

$$\beta_1\Omega = -(\alpha_1 - \alpha_2) \simeq \left(\frac{\tan\beta}{\omega_1} - \tau\right)\Omega \quad (6.3.12.4.3)$$

namely,

$$\beta_1 \simeq \frac{\tan\beta}{\omega_1} - \tau \quad (6.3.12.4.4)$$

However, selecting this value of β_1 also induces a frequency-independent phase shift ($-\beta_1\Omega_{pivot}$) unless Ω_{pivot} is zero because

$$\exp[i\beta_1(\Omega - \Omega_{pivot})] \exp[i(\alpha_1 + \alpha_2)] = \exp[-i\beta_1\Omega_{pivot}] \quad (6.3.12.4.5).$$

The additional phase shift must then be eliminated by a further zeroth-order phase correction. This is why phasing is sometimes difficult, because the results given in the first two steps cannot correct the phase entirely if the pivot is not zero.

As a consequence of phase correction, the spectrum becomes

$$S(\Omega) = \exp[\tau/T_2]\, S_{ideal}(\Omega) \quad (6.3.12.4.6)$$

where $\exp[\tau/T_2]$ is a factor introduced by a switching delay.

First-order phase shifts are often small. However, some digitally filtered and over-sampling data sets collected recently using BRUKER spectrometers have been headed by a number of artificial points (Figure 6.33). In this case the error is definitely not caused by an off-resonance effect but a significant delay before sampling begins. The corresponding spectrum can be treated with an extremely large first-order phase correction that is about 100 times bigger than usual.

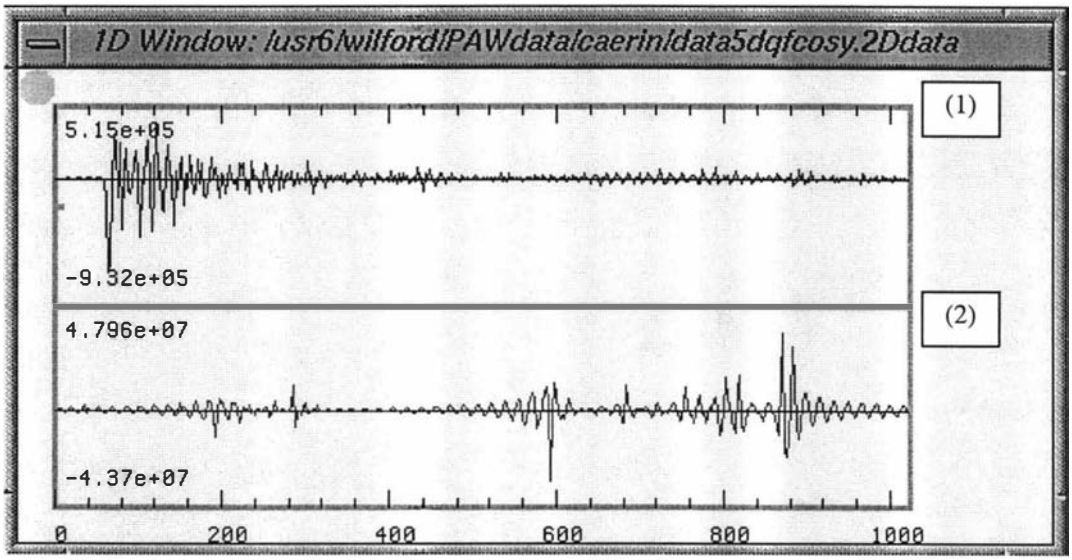


Figure 6.33: A digitally filtered, over-sampling data set (top) and its corresponding spectrum (bottom).

6.3.12.5 User-interface

The user-interface for the zeroth- and first-order phase correction is designed so that users can simply issue a phase-correction command or click on a button in the *Phasing Toolbox* to open a dialog. An example of the phasing dialog is shown in the next figure.

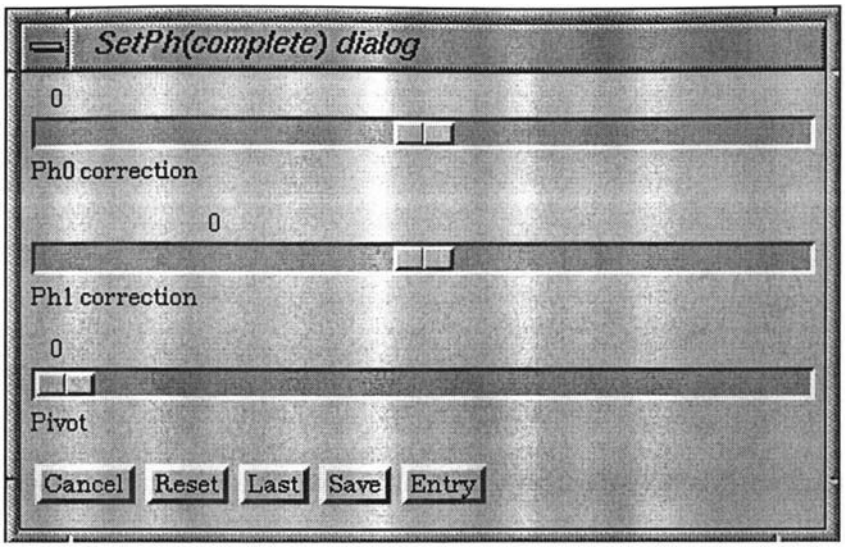


Figure 6.34: A phase-correction dialog. The sliders can be used to adjust the phase-correction parameters in real time. They can also be entered by choosing [Entry] in the dialog to open a keyboard-entry dialog.

Note that, PAW also allows users to interactively select the pivot frequency. In addition, PAW implements a separate option for correcting very large first-order phase errors. (See Volume II for more details.)

6.3.13 Baseline correction

Almost all NMR spectra have some degree of baseline problem. It gets worse as an increasing number of initial time-domain data points are distorted. Figure 6.35 shows different baseline patterns resulting from adding a constant to the values of the first 1, 2, 3, 4, 5, 10 and 15 data points.

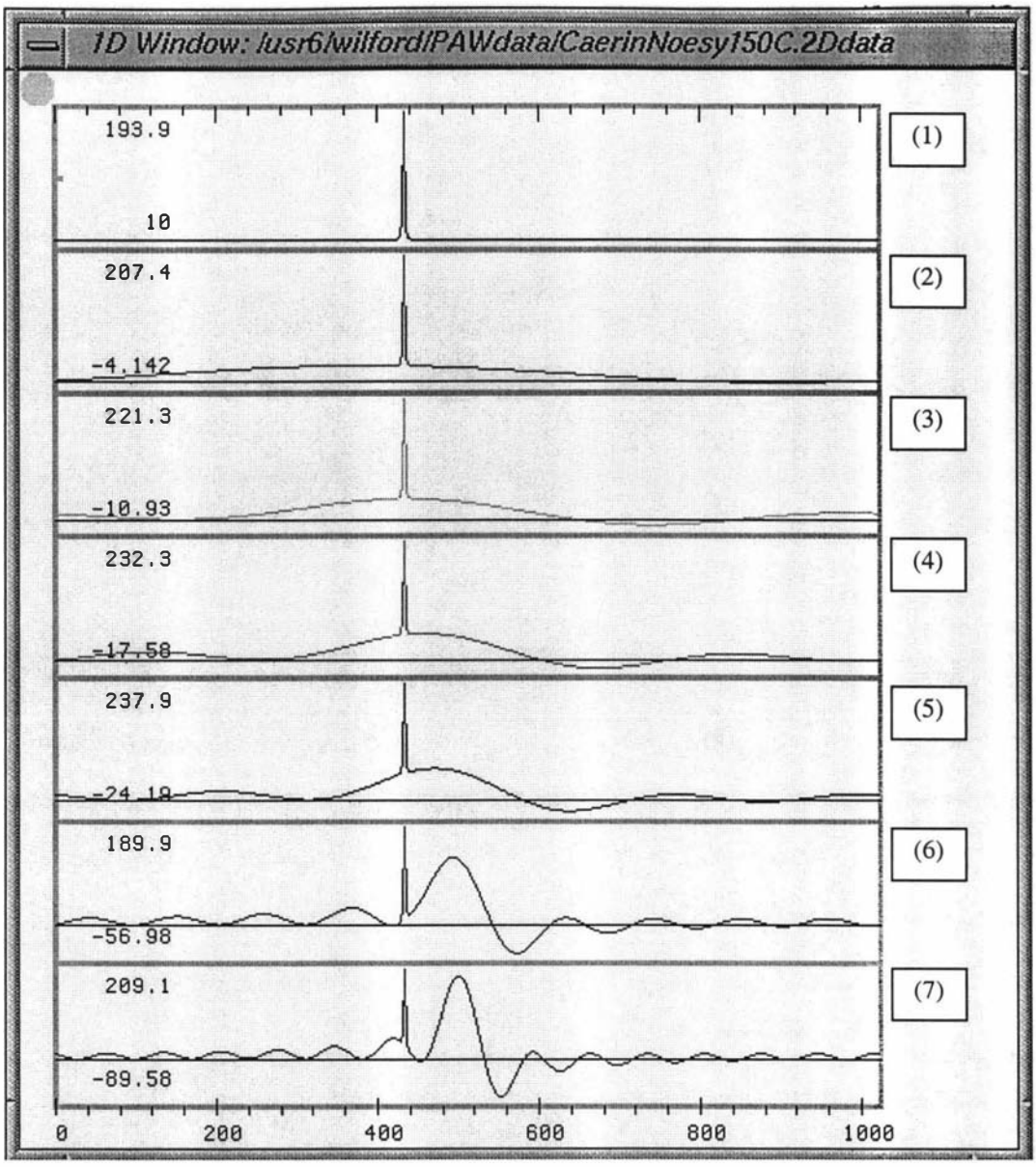


Figure 6.35: From top to bottom, different baseline patterns due to adding a constant to the values of the first 1, 2, 3, 4, 5, 10 and 15 time-domain data points. Note that the number of maxima appearing in the baseline is equal to the number of distorted points minus one.

The time-domain data of a spectrum with a distorted baseline is equivalent to the superposition of an ideal FID with another data set that produces the distorted baseline after applying Fourier transform. In the examples above, the distortion is simply a number of constant values added to the initial part of the FID. Actual NMR data, however, are not distorted in such a simple way. Figure 6.36 shows an example of spectral baseline distortion of an experimental data set.

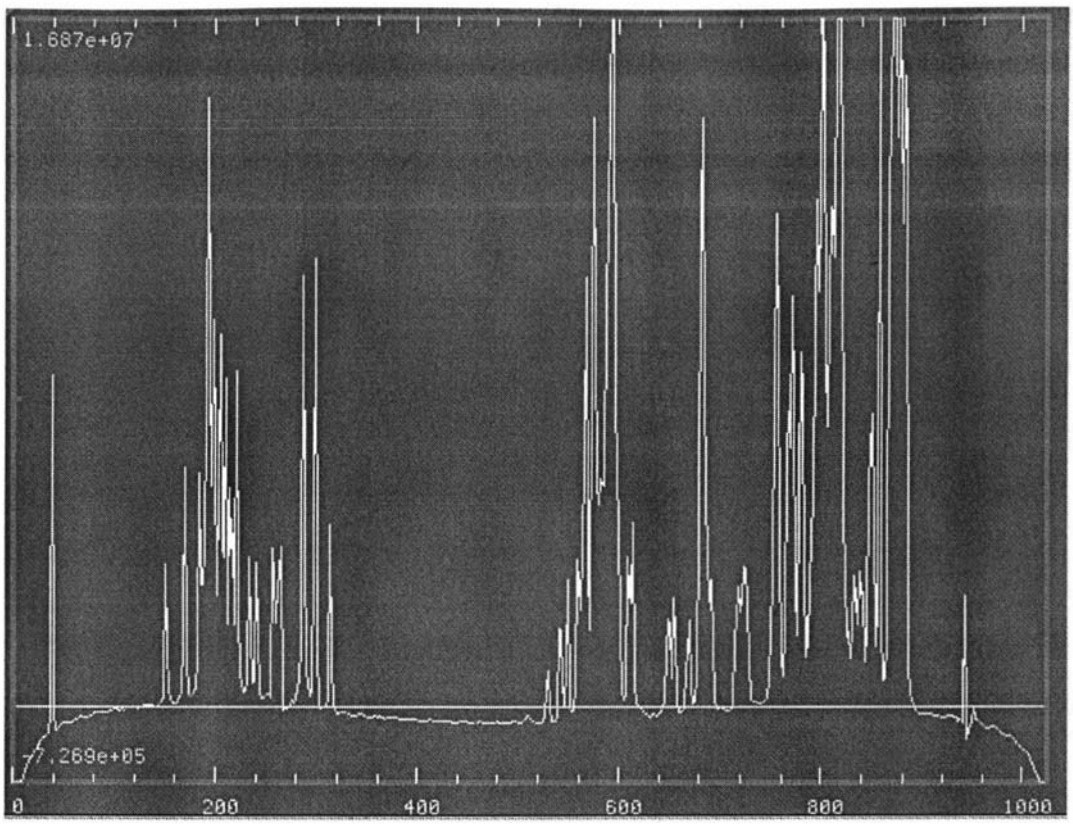


Figure 6.36: An example of experimental data with a distorted baseline. Here, the distorted baseline is due to non-linear distortion of the first few points.

6.3.13.1 Theory

Theoretically, baseline errors can be corrected if the distortion in the time-domain data can be determined and subtracted from the original signal. This can perhaps be done by a computer search for the distorted data points. However, the task is very time-consuming if more than two points are distorted, because a large number of Fourier transforms must be performed in order to optimise a curve fitted to an observed baseline.

In practice, given an FID signal, the first step to reduce baseline distortion is to choose a filter that can reasonably suppress the initial part of the time-domain data set. For instance, a 70° left-shifted sine-bell filter (Plot 1, Figure 6.37) is often used to process the NOESY or TOCSY signals, for which the maximum value is at the beginning of a data set. On the other hand, a non-shifted sine-bell filter is often used (Plot 2, Figure 6.37) to process COSY signals because M_{xy} is supposed to be zero at the beginning of a COSY data set. These operations, however, cannot completely correct the baseline problem because a number of data points may be affected.

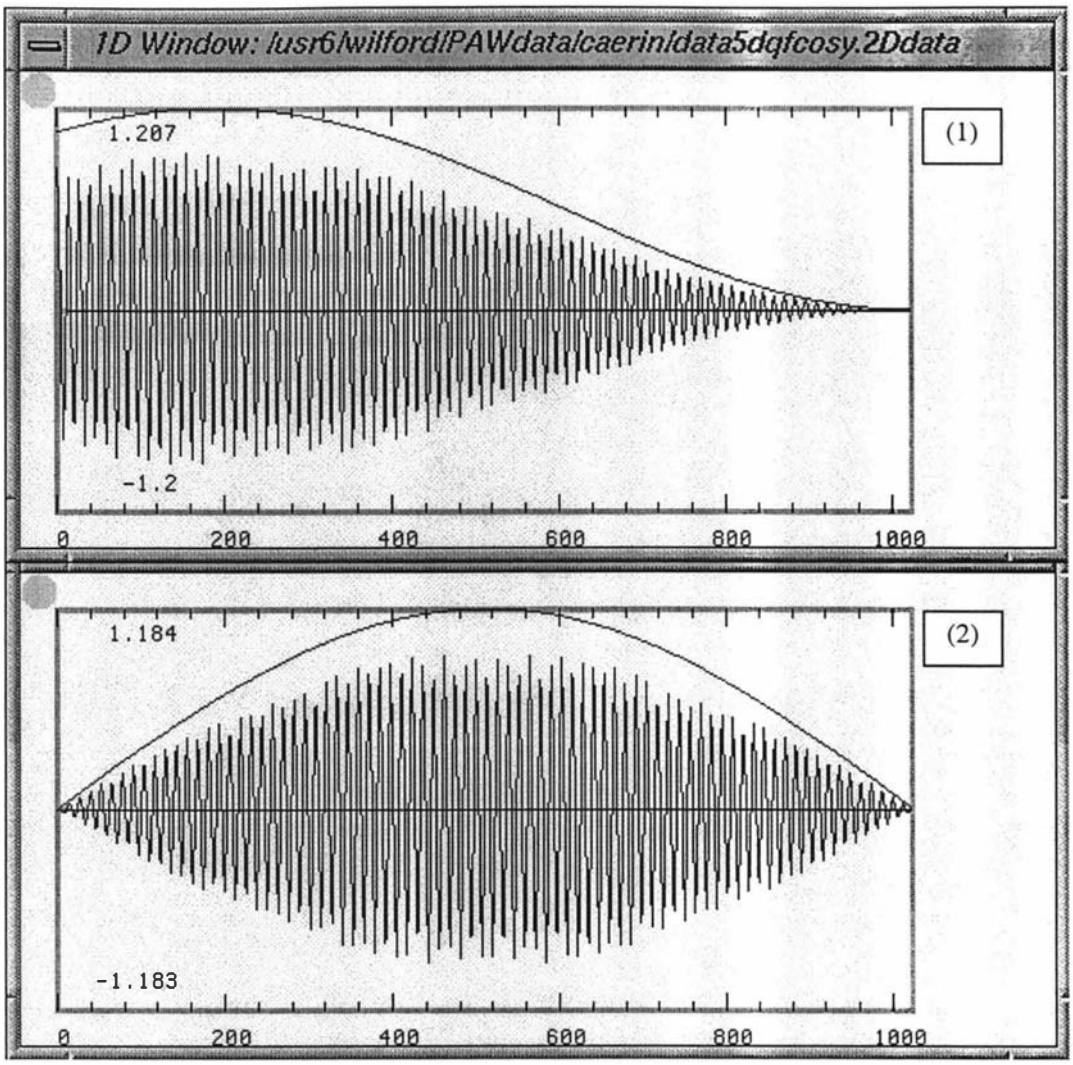


Figure 6.37: (1) A 70° left-shifted sine-bell filter. (2) A 90° right-shifted cosine filter.

The remaining baseline problems are then fixed by curve-fitting after FT. In many NMR packages, selected baseline points are fitted with a polynomial. This is often based on a curve-fitting theory that relies on the singular decomposition method to solve a linear least-square problem [Press 1986]: Given three sets of data, $\{x_i\}$, $\{y_i\}$[4] and the measurement errors $\{\sigma_i\}$, find $\{a_k\}$ that minimises

$$\chi^2 = \sum_{i=1}^{n} \left[\frac{y_i - \sum_{k=1}^{m} a_k x_i^{k-1}}{\sigma_i} \right]^2 \qquad (6.3.13.1.1).$$

Although it may not be the best, this is by far the most popular method for fitting baselines because it is relatively robust and easy to implement. In general, polynomial fitting works well if only a few points are distorted. For severe distortion involving many data points, polynomial fitting may fail.

[4] e.g., x_i is the time of the i^{th} data point and y_i is its amplitude.

An alternative to fitting polynomials to spectral baselines is to use spline function interpolation. The most popular spline interpolation uses a *cubic spline function* that uses a set of first or second derivatives as boundary conditions. Mathematically, the problem can be simply stated as follows:

> *Given two points and the first or second derivatives of the two points, find a cubic polynomial function that joints the two points.*

However, this only works if the derivatives at both ends of every baseline segment can correctly define curves that connect every two adjacent segments. For protein NMR data, applying the interpolation directly often does not give satisfactory results because the derivatives calculated locally are so susceptible to distortion. Figure 6.38 shows an overlap of a curve poorly fitted to the baseline of the spectrum in Figure 6.36 by a cubic-spline fitting technique.

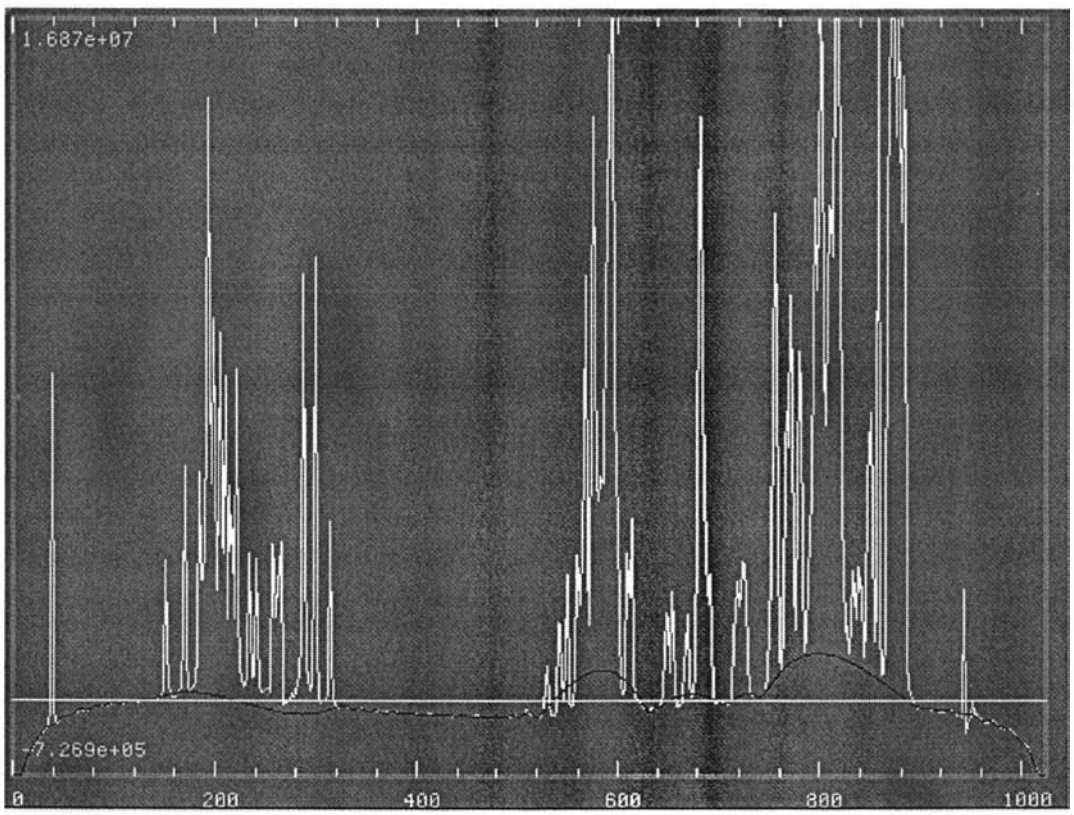


Figure 6.38: An example of experimental data poorly fitted with a cubic spline. Due to incorrect derivatives used in the fitting process, some regions are poorly fitted, such as those around points #200, 600 and 800.

Nevertheless, using spline functions to fill the gaps between baseline points is feasible because spline interpolation can produce curves with both continuity and smoothness. In fact, given a spectrum, one can often draw a reasonable baseline by hand. A reliable way to realise this hand-drawing-like process in computing is to use a cubic spline with a set of properly calculated first derivatives (Appendix 6a), as described in the next sub-subsection.

6.3.13.2 Program implementation

PAW has a number of routines to carry out every step required for manual or automatic baseline correction process, including those for picking, defining, removing, displaying and printing baseline segments, as well as those for correcting baselines with either polynomial functions or cubic spline functions.

The data input required for the correction routines is a set of pre-defined baseline segments. PAW samples a number of points x_j from the segments. To minimise local noise errors, PAW takes an average of five values from adjacent points for each y_j.

The routines for polynomial fitting are based on a curve fitting theory that relies on the singular decomposition method to solve the least-square problem described in the last sub-subsection. This method has been widely used and understood [see, e.g., Press 1986].

The routines for spline fitting are very sensitive to the calculation of first derivatives. To minimise local slope errors, the first derivative of each point x_j is calculated from a cubic polynomial that is fitted to a group of six points around x_j, where x_j is either the second or fifth point in a group, depending on where it is. With the derivatives, parameters for piecewise cubic polynomials are calculated using the formulae in Appendix 6a. A baseline curve is formed accordingly. This curve is finally subtracted from the original spectrum as a correction.

6.3.13.3 User-interface

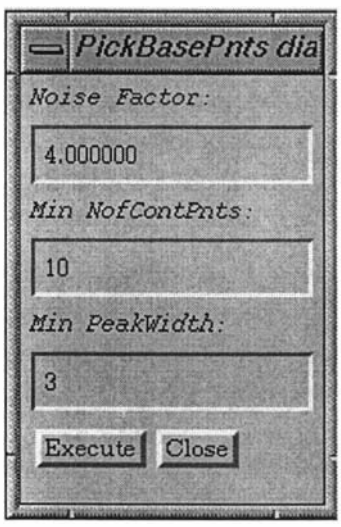
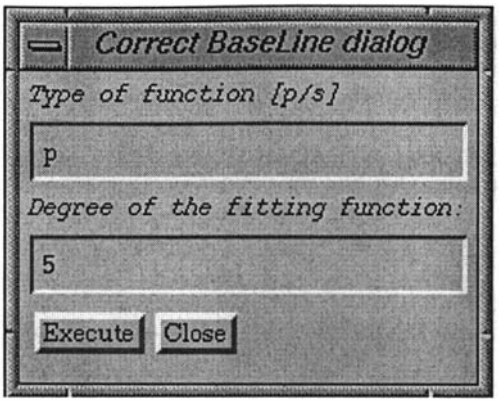


Figure 6.39: The baseline-picking dialog

PAW provides simple interfaces needed for picking, removing and defining baseline segments, either interactively or in macros. For example, in the process of automatic baseline-segment picking, users can specify the parameters using a dialog window, as shown on the left.

The interactive baseline-picking process provides an immediate display of the baseline segments. Details for the segments can also be printed so that the information can be used for macro programming.



The final correction step is a command that opens the baseline-correction dialog, as shown on the right. The dialog is used for specifying a fitting function (either a p for polynomial or an s for cubic spline) and the degree of the function if a polynomial is to be used.

Figure 6.40: The baseline-correction dialog

PAW requires that the baseline segments rather than points be defined, and the definitions can then be put into a macro for multi-dimensional data processing. This would not have been possible if baseline points were to be specified. In general, a baseline fitting technique based on defined segments is more stable than those based on auto-pick routines, because the latter may pick up incorrect baseline points. Since multi-dimensional NMR data sets always have a number of common baseline segments, the fitting method that uses defined baseline segments can often be applied with satisfactory results.

6.3.13.4 Effectiveness of PAW's baseline correction routines

The two plots in Figure 6.41 show a spectrum before and after being corrected by PAW's baseline correction routine.

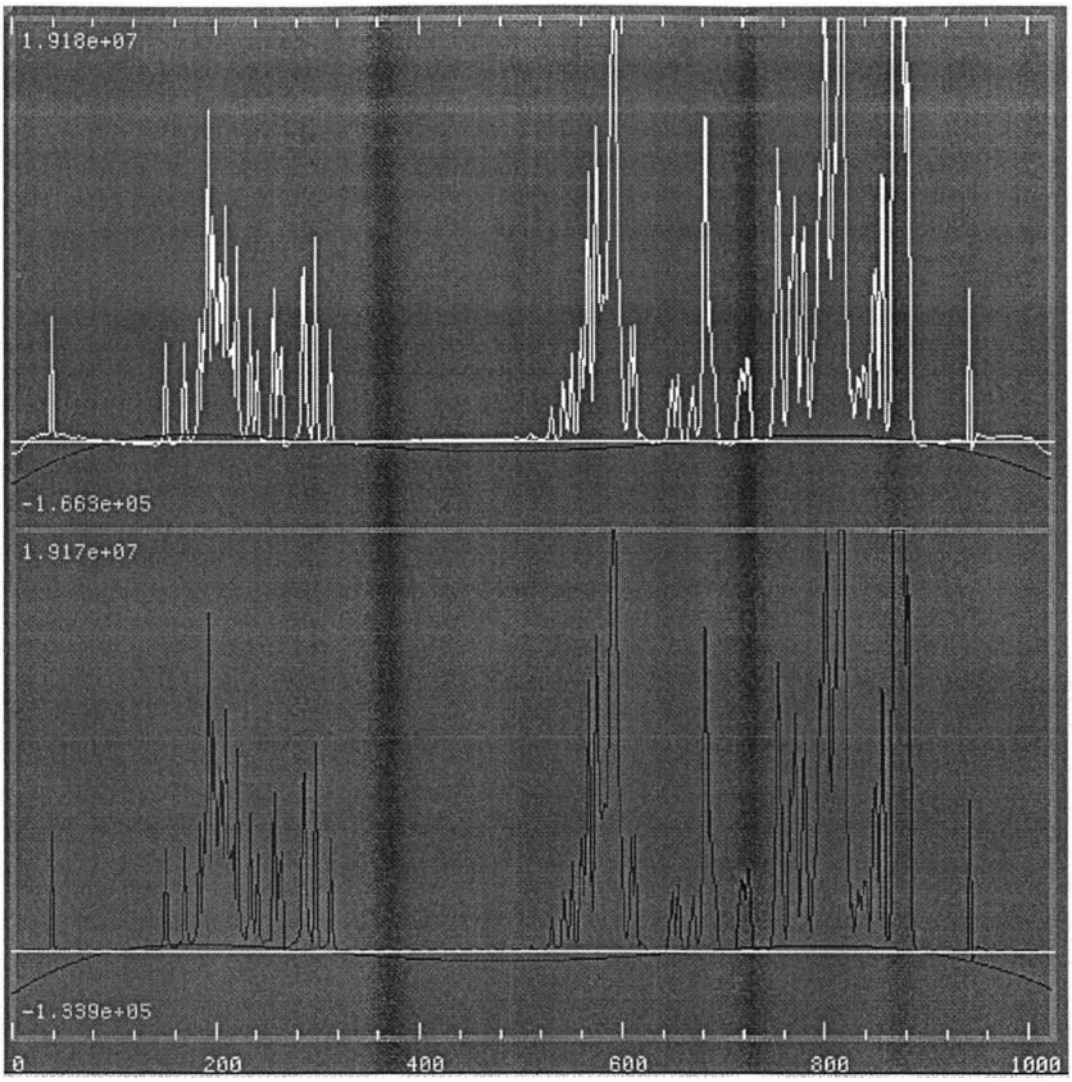


Figure 6.41: The same spectrum (see also Figure 3.36 and 6.38) after being corrected with polynomial (top) and cubic spline (bottom) fitting routines in PAW. In both plots, the smooth curves are the baselines fitted. The resulting spectra show that the cubic spline fitting gives a better result in this case.

In general, the spline fitting of PAW can work better than polynomial fitting. This is because piecewise cubic-spline functions are much more flexible for fitting curves, provided that the gaps between points can be fitted with cubic polynomials and that the derivatives between two ends of each gap are properly calculated. In addition, the baseline segments must be selected with great care; otherwise, some severe problems can still occur.

6.4 Two-dimensional NMR Data Processing

6.4.1 2D NMR data sets

A 2D time-domain data set consists of a number of FID signals (or FIDs). Usually, the data set is organised into a matrix of complex data, with each row being a complete FID signal. The real and imaginary parts of the data set can be displayed separately by 2D intensity plots with the two axes being D_1 horizontally from left to right, and D_2 vertically from bottom to top, as shown in the next diagram.

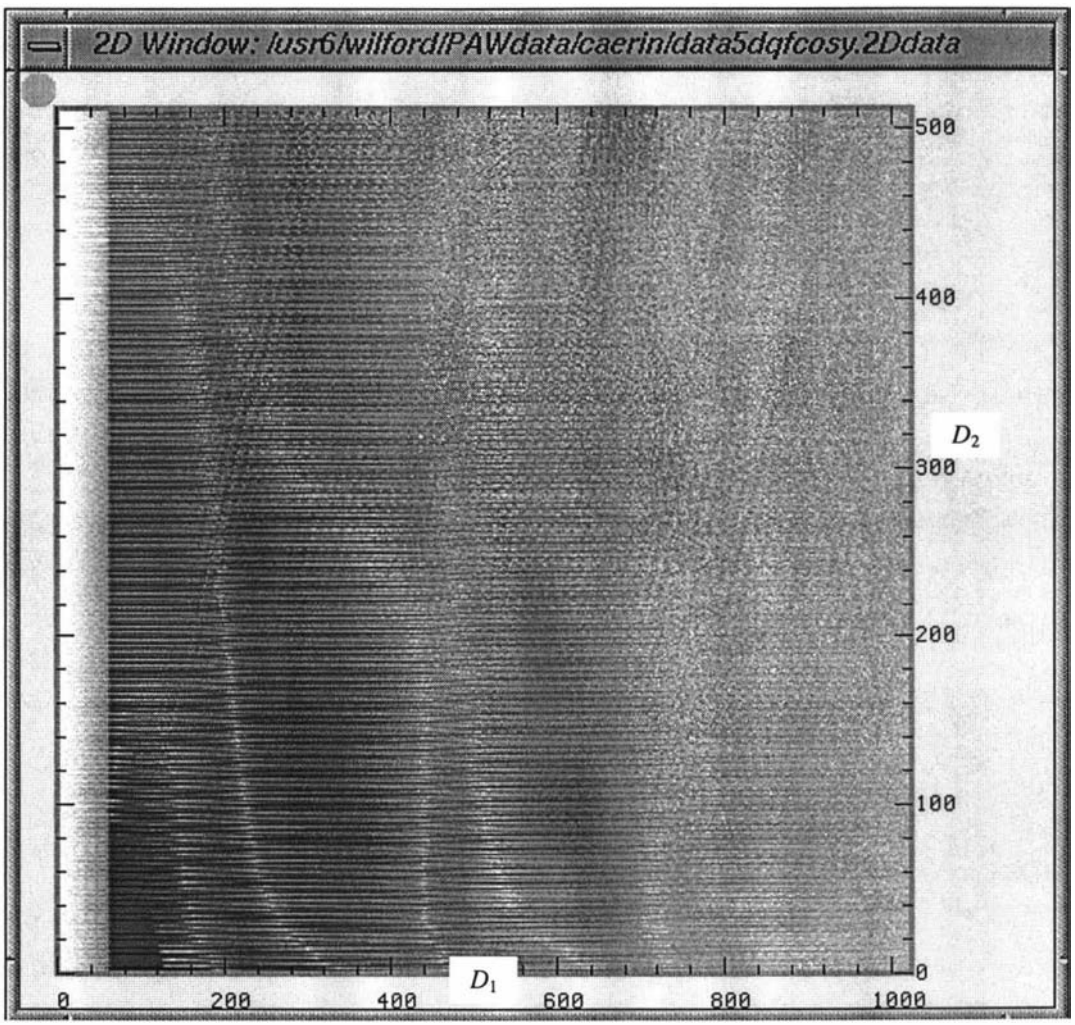


Figure 6.42: An intensity plot of a 2D time-domain data set. Note that the subscripts for the D_1 and D_2 are opposite to the t_1 and t_2 normally used in 2D NMR theory. In other words, D_1 corresponds to the data-acquisition time t_2, and D_2 corresponds to the evolution time t_1. (See also chapter 3.)

After being processed and transformed into frequency-domain data, only the real part is retained in the computer memory. The real data can be displayed using an intensity plot or a contour plot, with the two axes D_1 and D_2 becoming Ω_1 and Ω_2, respectively. They can have units of points, Hz or ppm.

2D NMR time-domain data can be collected using different methods. The most commonly used ones are the *States-Haberkorn-Ruben method* (Varian 1989-92b) and the time-proportional phase-incrementation (TPPI) method (Drobny 1979,

Bodenhausen 1980). Both methods usually[5] collect a series of complex FIDs. The difference appears after each row of data has been Fourier transformed in D_1. The former method results in columns of complex 1D data in D_2, and the latter, real.

6.4.2 General procedure

There are slightly different procedures for processing a 2D NMR data set. In general, the order is to process rows first and then columns. Figure 6.43 and 6.44 show two of the procedures that are often used during 2D processing with PAW.

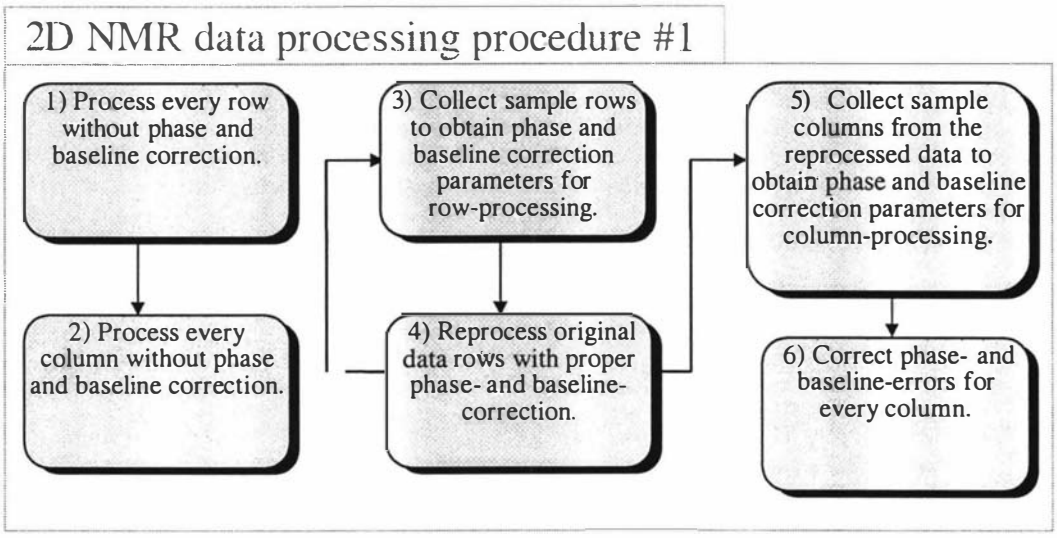


Figure 6.43: A procedure for processing a 2D NMR data set. The phase correction parameters are obtained after 2D Fourier transform, and hence the Hilbert transform is required if only the real part of the 2D spectrum is kept. This procedure requires the 2D Fourier transform be performed at least two to three times.

The first two steps can be performed with or without zero-filling. In both steps, only the real parts are retained to reduce the memory space required for the process. Step 1 results in a 2D interferogram data set, whereas step 2 results in a 2D NMR spectrum. Note that neither the phase nor the baseline correction is applied in the first two steps. Hence, the 2D spectrum at this stage is usually quite poor.

In step 3, at least two rows of data from the 2D spectrum are collected. They are then added together to form a combined real-only 1D spectrum. In order to obtain a set of phasing parameters that are generally suitable for every row of the 2D spectrum, the 1D combined spectrum usually contains at least one peak at each end. To work out the phase parameters, the imaginary part of the 1D combined spectrum has to be created by applying a Hilbert transform. (Also, the baseline segments selected for the correction must be common to all rows and columns.) The parameters thus obtained will be applied to every row in step 4 and hence must be written down or printed out for future reference.

[5] Some TPPI experiments also collect real FIDs instead of complex.

Step 5 and 6 are just a repeat of the last two steps for the phase and baseline correction of columns.

This procedure, however, can be changed in a few different ways. For example, step 5 and 6 can be done before step 3. Also, multiple rows in D_1 can be used to work out row-wise processing parameters before performing step 1, as shown in Figure 6.44.

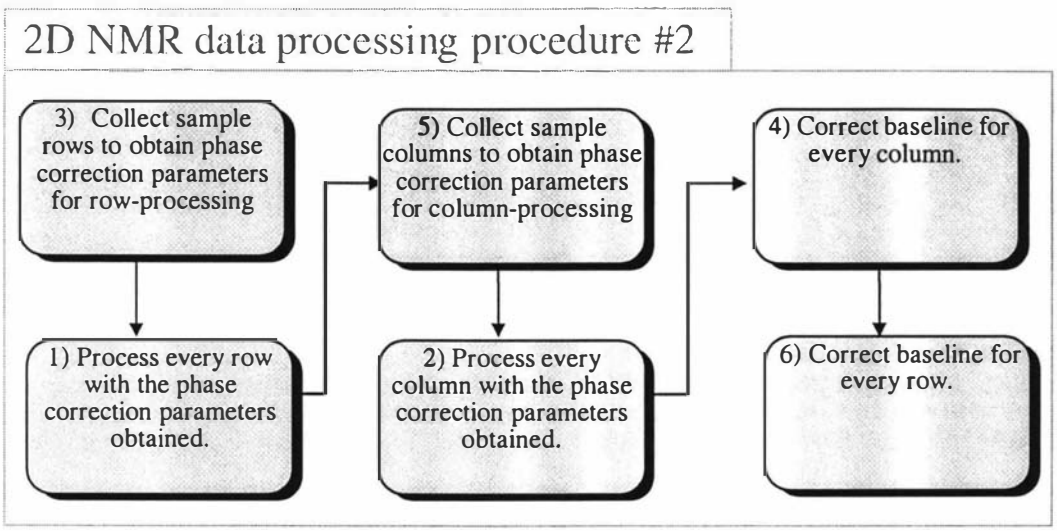


Figure 6.44: An alternative procedure for processing a 2D NMR data set. The phase correction parameters are obtained before Fourier transform, which is only performed once in this procedure, and hence is more efficient.

Examples of how 2D processing is performed using procedure #2 may be found in Chapter 8 of Volume II.

Chapter 7:

NMR Spectral Assignment of Proteins Using PAW

7.1 Introduction

This chapter discusses the principles and methods used in protein NMR spectral assignment with PAW. The methods are the summaries from our study and practice based on the characteristics of 2D protein NMR spectra. The chapter also contains a number of useful tables and diagrams used in the assignment processes.

Section 7.2 describes the protein primary structures and the proton codes used in PAW. Section 7.3 describes the spin systems in proteins and the notation for ^{1}H spins and spin systems in protein NMR spectra. Section 7.4 describes 2D ^{1}H NMR spectra of proteins and notation for 2D cross-peaks in PAW. Section 7.5 describes 2D NMR spectral assignment of proteins, including spin-system identification and sequence-specific assignment. Example of spectral assignment using PAW can be found in Chapters 10 and 11.

7.2 Protein Structure and Proton Codes

7.2.1 Protein primary structure

Proteins are polymers of amino acids. There are 20 common *amino acids* that form the basic building blocks of proteins. Every amino acid contains a distinct group of atoms called the *residue group* (R) and two common functional groups: the *amino group* ($NH_3^{\oplus}$) and *carboxyl group* ($COO^{\ominus}$), as shown in Figure 7.1.

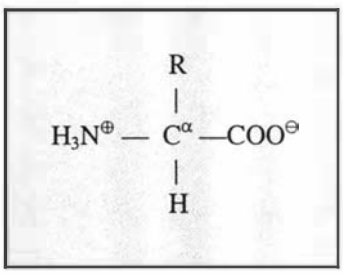


Figure 7.1: General structure of amino acids, where R represents a side-chain residue.

In a protein molecule, the carboxyl group of one amino acid is covalently attached to the amino group of its neighbour by a *peptide bond* (indicated by the symbol ⊷ in Figure 7.2).

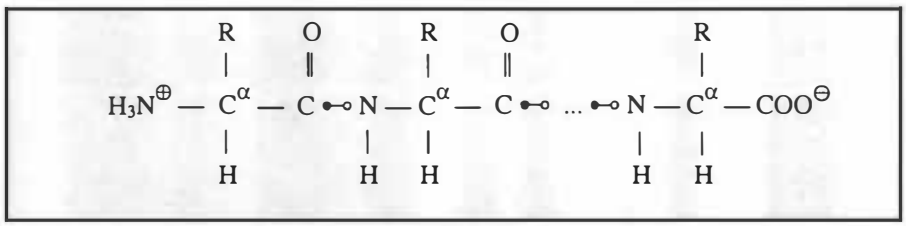


Figure 7.2: Protein structure, where the symbol ⊷ represents a peptide bond. The left end is called the N-terminal, and the right end, C-terminal. Every two adjacent amino acids are covalently attached to each other with the elimination of a water molecule between them.

7.2.2 Side chains of the 20 common amino acids

The primary structures of the 20 common amino-acid residues found in proteins, as well as their single- and three-letter codes (e.g., G, Gly), are listed in Table 7.1. To avoid confusion, the symbols for atoms in the residues follow the rules recommended by IUPAC-IUB Commission on Biochemical Nomenclature [IUPAC-IUB Commission 1970], e.g., N^{ζ}, $C^{\gamma 2}$. Different symbols used in Wüthrich 1986 are also attached to the bottom-left corners of some heavy atoms as cross-references, e.g., ${}_{\varepsilon}N^{\zeta}$, ${}_{\gamma}C^{\gamma 2}$, because they have also been used in many articles.

Table 7.1. Primary structures of the 20 common amino-acid residues in proteins. In the brackets are the single- and three-letter codes for the amino acids. IUPAC-IUB Commission's codes are used throughout the table. The symbols at the bottom-left corner attached to some Cs and Ns are found in Wüthrich 1986.

(a) Unbranched side-chains

Glycine	(G, Gly)	$-H$
Alanine	(A, Ala)	$-C^{\beta}H_3$
Serine	(S, Ser)	$-C^{\beta}H_2-O^{\gamma}$
Cysteine	(C, Cys)	$-C^{\beta}H_2-S^{\gamma}$
Methionine	(M, Met)	$-C^{\beta}H_2-C^{\gamma}H_2-S^{\delta}-C^{\epsilon}H_3$
Lysine	(K, Lys)	$-C^{\beta}H_2-C^{\gamma}H_2-C^{\delta}H_2-C^{\epsilon}H_2-{}_{\epsilon}N^{\zeta}H_3^{\oplus}$

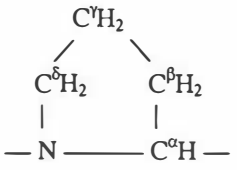
(b) Branched side-chains

Valine	(V, Val)	$-C^{\beta}H(-{}_{\gamma}C^{\gamma 1}H_3)-{}_{\gamma}C^{\gamma 2}H_3$
Threonine	(T, Thr)	$-C^{\beta}H(-O^{\gamma 1})-{}_{\gamma}C^{\gamma 2}H_3$
Isoleucine	(I, Ile)	$-C^{\beta}H(-{}_{\gamma}C^{\gamma 1}H_2-C^{\delta}H_3)-{}_{\gamma}C^{\gamma 2}H_3$
Leucine	(L, Leu)	$-C^{\beta}H_2-C^{\gamma}H(-{}_{\delta}C^{\delta 1}H_3)-{}_{\delta}C^{\delta 2}H_3$
Aspartic acid	(D, Asp)	$-C^{\beta}H_2-C^{\gamma}(=O^{\delta 1})-O^{\delta 2}H$
Asparagine	(N, Asn)	$-C^{\beta}H_2-C^{\gamma}(=O^{\delta 1})-{}_{\gamma}N^{\delta 2}H_2$
Glutamic acid	(E, Glu)	$-C^{\beta}H_2-C^{\gamma}H_2-C^{\delta}(=O^{\epsilon 1})-O^{\epsilon 2}H$
Glutamine	(Q, Gln)	$-C^{\beta}H_2-C^{\gamma}H_2-C^{\delta}(=O^{\epsilon 1})-{}_{\delta}N^{\epsilon 2}H_2$
Arginine	(R, Arg)	$-C^{\beta}H_2-C^{\gamma}H_2-C^{\delta}H_2-N^{\epsilon}H-C^{\zeta}(={}_{N}N^{\eta 1}H_2^{\oplus})-{}_{N}N^{\eta 2}H_2$

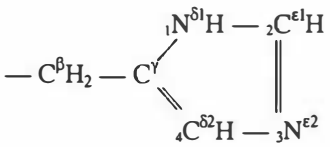
Table 7.1. Primary structures of the 20 common amino-acid residues in proteins. (Continued)

(c) Cyclic side chains

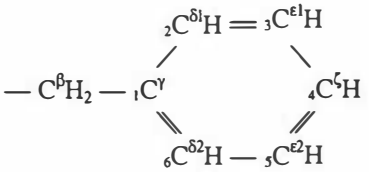
Proline (P, Pro)



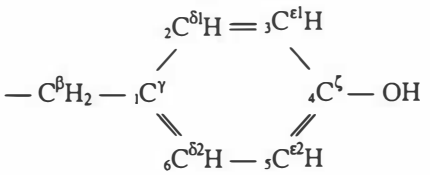
Histidine (H, His)



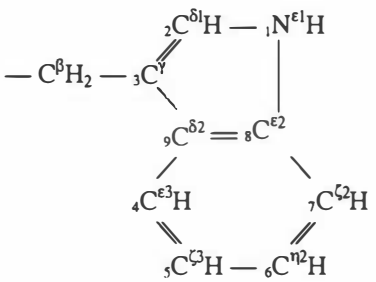
Phenylalanine (F, Phe)



Tyrosine (Y, Tyr)



Tryptophan (W, Trp)



7.2.3 Proton codes used in PAW

The proton codes used in PAW are similar to those seen in many articles on NMR spectroscopy. The codes are composed by applying the following rules:

1. Nuclear proton codes are closely related to the atom codes in Table 7.1.
2. Only symbols in regular format are used to denote nuclear spins, neither superscript nor subscript.
3. The letters A, B, G, D, E, Z and H (also recommended by IUPAC-IUB Commission) substitute for α, β, γ, δ, ε, ζ and η because of the difficulty in typing and displaying Greek characters.
4. Nuclear spins of heavy atoms are denoted by the position of the corresponding atoms, e.g., CA, CB.

5. HN is used exclusively for spins of ^{1}H protons that are covalently attached to the backbone nitrogen atoms. Other spins of ^{1}H protons are identified by the position of heavy atoms to which the protons are covalently attached, e.g., HA, HB.

6. A different number is added to denote different spins in the same chemical bond. For example, HB1 and HB2 denote the spins of the two protons in $C^{\beta}H_2$; HH21 and HH22 denote those found in $C^{\eta 2}H_2$.

7. No number is added to the proton codes for the spins of ethyl and methyl ^{1}H protons with identical chemical shifts, e.g., HB, HG and HD. (The letter Q is used to identify these spins in the programs XEASY [Xia *et al.* 1994] and DIANA [Güntert 1995] developed by Wüthrich's group, e.g., QB, QG and QD.)

7.3 Spin Systems in Protein NMR

There are many different spin systems in protein NMR. Some experiments based on different coherence transfer pathways may be used to identify coherence transfers[1] between nuclear spins of the same spin system[2], either homo- or hetero-nuclear.

7.3.1 ^{1}H spin systems in amino-acid residues

The homo-nuclear ^{1}H spin-systems in amino-acid residues are the most important for protein structural research in NMR. They are within an amino-acid block, separated by heavy atoms to which no hydrogen atom is attached, and can be observed in various NMR spectra. DQF-COSY spectra, for example, correlate spins of adjacent ^{1}H protons that are covalently attached through nitrogen and carbon atoms, such as HN-Hα, Hα-Hβ and Hβ-Hγ. TOCSY spectra, on the other hand, correlate all spins of all ^{1}H protons in a spin system, such as HN-Hα, HN-Hβ and HN-Hγ. (Note: All ^{1}H protons attached to a nitrogen atom in the 20 common amino acid residues are *labile protons* that are exchangeable with deuterium.)

Each of the spin systems in the 20 common amino-acid residues is often characterised by a string of letters adapted from Pople [Pople *et al.* 1959], such as the AMX system, AM(PT)X system, A_3B_3MX system, AMX+AA′XX′ system, etc. A list of all spin systems in the 20 common amino-acid residues is given in column three of Table 7.2.

The rules for forming a *spin-system string* is as follows:
1. Different capital letters represent spins with different chemical shifts.
2. Neighbouring letters that are next to each other in the alphabet like AB indicates strong coupling.
3. Letters with a numeric subscript indicate a number of strongly coupled spins that have the same chemical shifts, such as A_2 for the two spins of ethyl protons and A_3 for the three spins of methyl protons.

[1] See Chapter 3.

[2] See Chapter 2 and 3.

4. All letters for the spins of side-chain residues are listed from periphery position towards Hα. Letters in parenthesis such as (PT) indicate that the chemical shifts of the spins they represent are higher than those listed before them.
5. A string following a plus sign indicates a separate spin system in a residue.
6. A pair of characters with an apostrophe such as AA' represents the spins of a pair of ^{1}H protons in an aromatic ring that have the same chemical shifts due to symmetrical structure and identical chemical environment.

Table 7.2. Spin-systems in 20 common amino acids and PAW's spin-system strings

BRIEF CODES		SPIN SYSTEMS	PAW'S SPIN-SYSTEM STRINGS
G	Gly	AX	[N-α-α]
A	Ala	A_3X	[N-α-β_3]
S	Ser	AMX	[N-α-β-β]
C	Cys	AMX	[N-α-β-β]
M	Met	AM(PT)X+A_3	[N-α-γ-γ-β-β] + [ε_3]
K	Lys	$A_2(F_2T_2)$MPX	[N-(ζ_3)-α-ε_2-β-β-δ_2-γ_2]
V	Val	A_3B_3MX	[N-α-β-$\gamma 1_3$-$\gamma 2_3$]
T	Thr	A_3MX	[N-α-β-$\gamma 2_3$]
I	Ile	A_3MP(B_3)TX[3]	[N-α-β-γ1-γ1-$\gamma 2_3$-δ_3]
L	Leu	A_3B_3MPTX	[N-α-β-β-γ-$\delta 1_3$-$\delta 2_3$]
D	Asp	AMX	[N-α-β-β]
N	Asn	AMX	[N-(δ2-δ2-)α-β-β]
E	Glu	AM(PT)X	[N-α-γ-γ-β-β]
Q	Gln	AM(PT)X	[N-(ε2-ε2-)α-γ-γ-β-β]
R	Arg	$A_2(T_2)$MPX	[N-($\eta 1_2$-$\eta 2_2$-)α-δ_2-β-β-γ_2]
P	Pro	$A_2(T_2)$MPX	[α-δ_2-β-γ_2-β] or [α-δ-δ-β-γ-γ-β]
H	His	AMX+AX	[ε1-(δ1)] +δ2] +[N-α-β-β]
F	Phe	AMX+AMM′ XX′	[ε1ε2-ζ-δ1δ2] +[N-α-β-β]
Y	Tyr	AMX+AA′ XX′	[ε1-δ1 +ε2-δ2] +[N-α-β-β]
W	Trp	AMX+A(X)MP+A	[(ε1)-δ1] +[ε3-ζ2-η2-ζ3] +[N-α-β-β]

Because some residues contain more than one spin system, the total number of spin systems in the list is 26. Among them there are eight AMX systems, three AM(PT)X systems and two $A_2(T_2)$MPX systems.

7.3.2 PAW's spin-system strings

Also in column four of Table 7.2 are PAW's *spin-system strings* for the spin-systems in the 20 common amino-acid residues. The strings, listed in reverse order corresponding to Pople's spin-system strings, are constructed using the following five rules:

1. All strings are formed by proton codes, sorted by their chemical shifts.
2. Proton codes with different chemical shifts are separated by a hyphen, e.g., α-γ-γ-β-β. Otherwise, they are either grouped with a subscript (e.g., γ_2 for the spins of two ethyl protons with identical chemical shifts) or listed without separation (e.g., ε1ε2 for the spins of two symmetric ^{1}H protons in a ring structure).

[3] This is listed as A_3MPT(B_3)X in Wüthrich 1986, which doesn't comply with rule 4. The rearrangement follows the rule strictly.

3. Strings enclosed in round brackets represent the spins of *labile protons* that can be exchanged by deuterium, e.g., (ε1).
4. Proton codes enclosed in squared brackets and separated with a plus sign belong to the same spin systems, e.g., +[ε2-δ2].

Be aware that there is a difference between numeric letters in regular format and subscript format such as those in δ2 and δ_2. The former represents a spin of a proton attached to a carbon at the δ2 position in a residue; the latter represents the spins of two ^{1}H protons at the δ position. Hence, the symbol $\gamma 1_3$ represents the spins of three ^{1}H protons at the γ1 position.

A table of PAW's spin-system strings in Latin letters can be displayed at anytime during spectral assignment with PAW, as explained in Volume II. Because PAW's spin-system strings clearly contain proton codes that are sorted by statistical chemical shift but not by atom position in the primary structure, the table can be used as a quick reference to identified spin systems in the spectral assignment process.

7.3.3 ^{1}H chemical shifts in amino acid residues

The conformation effects on chemical shifts of amino acids in polymer systems have been discussed in many articles, such as Sternlicht & Wilson 1967, Markley *et at.* 1967, Tigelaar & Flygare 1972, Perkins & Wüthrich 1979, Clayden & Williams 1982, Pardi *et al.* 1983, Dalgarno *et al.* 1983, Szilagyi & Jardertzky 1989, as well as Pastore & Saudek 1990. The ^{1}H chemical shifts of random coiled tetra-peptides H-Gly-Gly-Xxx-Ala-OH, where Xxx is one of the 20 common amino acid, can be found in Bundi & Wüthrich 1979, Wüthrich 1986, as well as Gross and Kalbitzer (1988).

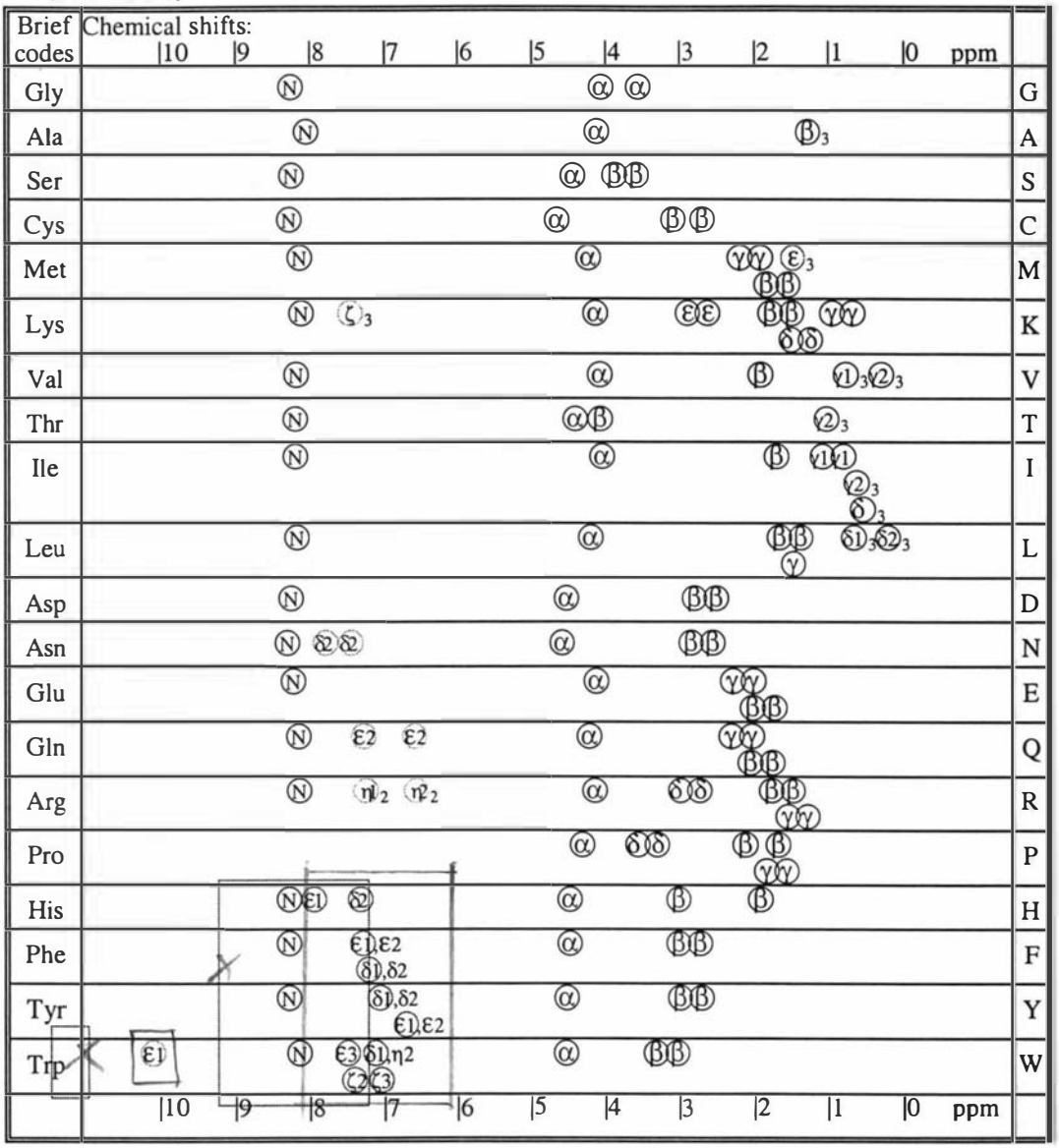
Statistical analysis on ^{1}H chemical shifts obtained from NMR spectral assignment for over 70 proteins revealed some important relationships between NMR chemical shifts and protein secondary structure [Wishart 1991, 1995, 1998]. It was found that all residues experience up-field shifts in α-helical configurations, and down-field shifts in β-strand configurations. The ranges of shifts are both between 0.15 and 0.60 ppm from the random coil values. (The study for ^{13}C and ^{15}N also revealed similar relationships.) The findings on ^{1}H chemical shift distribution of the 20 common amino acids are particularly useful for spin-system identifications and sequence-specific assignment in protein NMR.

7.3.4 Chemical shift distribution plot

Table 7.3 shows the distribution of ^{1}H chemical shifts[4] of the 20 common amino acids using proton codes recommended by IUPAC-IUB Commission. With the proton codes pictorially placed at positions that are proportional to their average chemical shifts, the table is a very useful reference during the assignment process.

[4] Based on the statistics presented in Wishart *et al.* 1991.

Table 7.3. Distribution of ^1H chemical shifts of the 20 common amino acids using proton codes recommended by IUPAC-IUB Commission. Protons in a crowded area are shifted to another row. Those in the separate spin systems of aromatic rings are enclosed in rectangles. Mobile protons are in dotted circles.



Note that the actual locations of peaks are subject to variations ranging from ±0.15 to ±0.60 ppm, but their distribution patterns should be similar in a 2D spectrum.

The same distribution table using the notation found in Wüthrich 1986 is also presented in Appendix 7a.

7.3.5 NOESY connectivity diagrams of protein spin systems

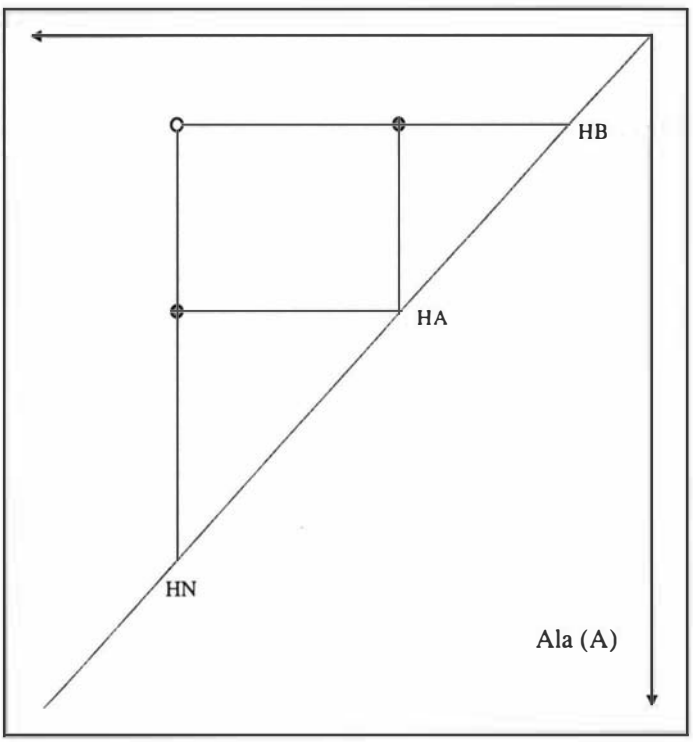
Spin-system connectivity diagrams illustrate the distribution patterns of peaks within a 2D spectrum. The COSY, relayed COSY and double-relayed COSY connectivity diagrams for the spin systems of non-labile protons in the 20 common amino acid residues can be found in Wüthrich 1986.

In protein structure determination, it is often useful to reference the TOCSY connectivity diagrams of the ^1H spin-systems. Figure 7.3 shows the spin-system connectivity diagrams of the Ala and Trp residues.

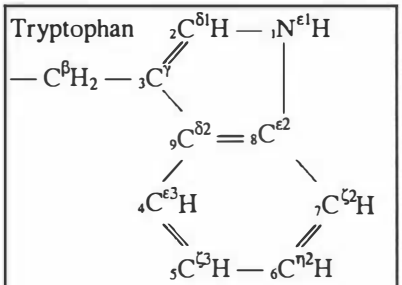
Alanine $-C^{\beta}H_3$

Spin system:

$[N\text{-}\alpha\text{-}\beta_3]$



Tryptophan



Spin systems:

$[(\varepsilon 1)\text{-}\delta 1]$

$+[\varepsilon 3\text{-}\zeta 2\text{-}\eta 2\text{-}\zeta 3]$

$+[N\text{-}\alpha\text{-}\beta\text{-}\beta]$

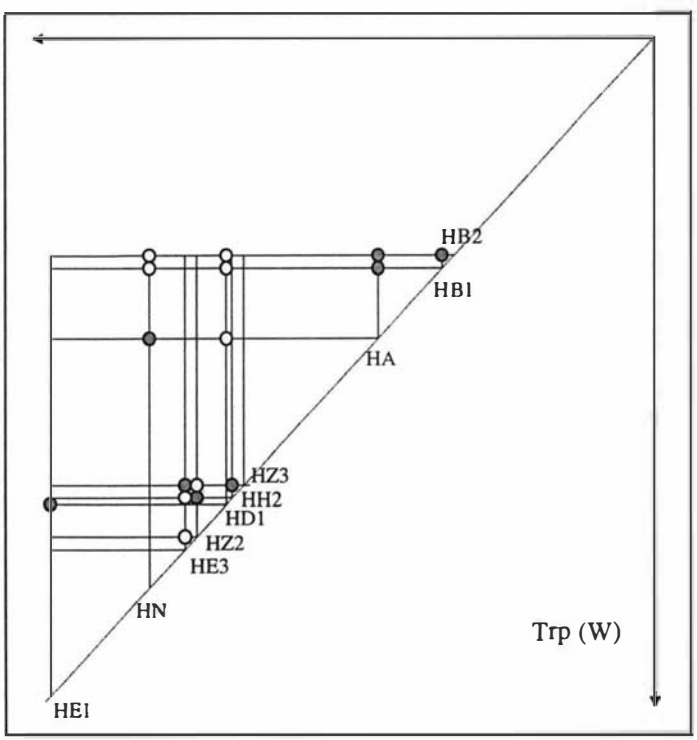


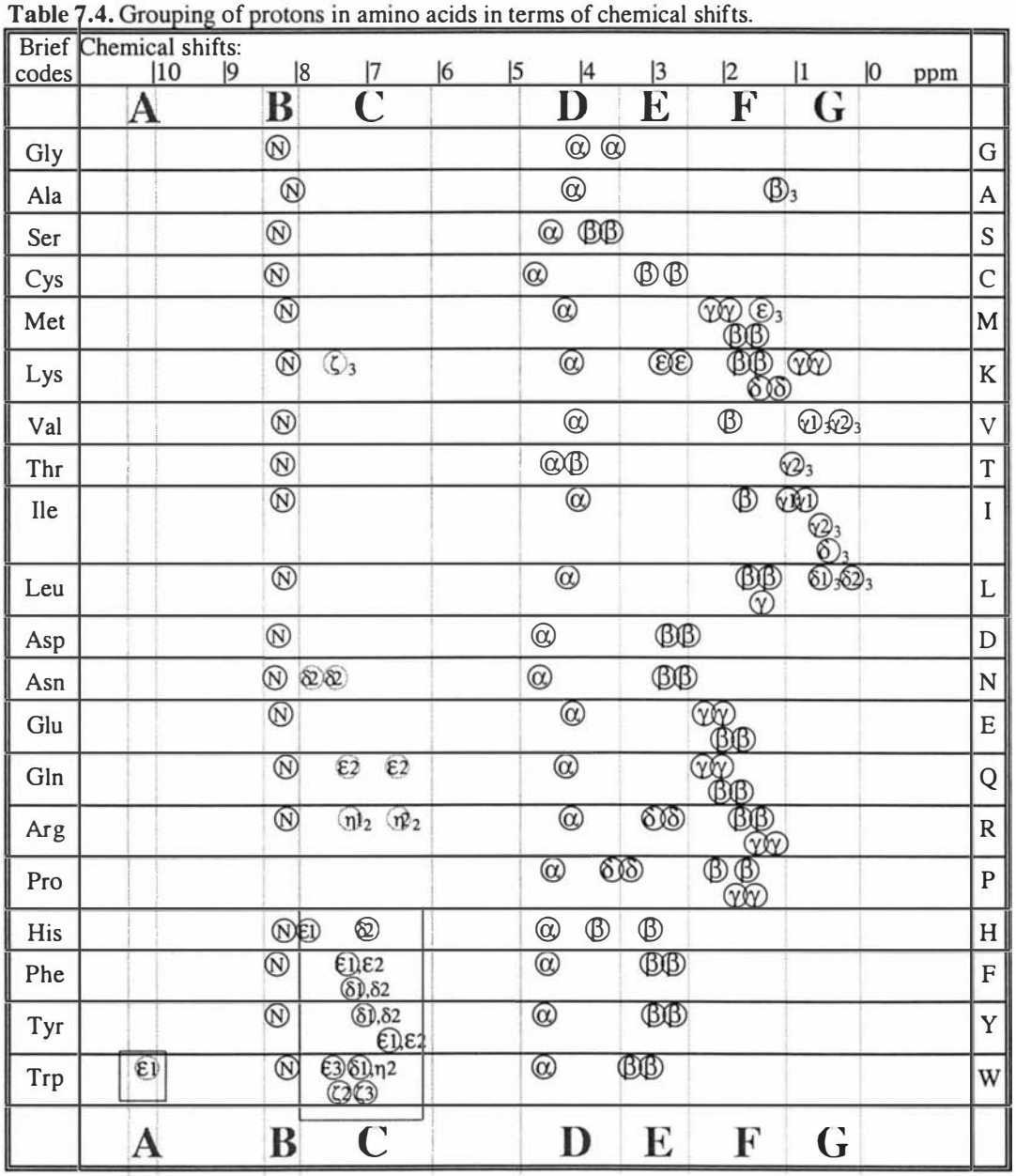
Figure 7.3: The ^1H TOCSY connectivity diagrams of Alanine (top) and Tryptophan (bottom) spin-systems. Solid circles represent neighbouring-proton connectivities like those in a COSY spectrum. Open cycles represent other TOCSY connectivities.

The ^{1}H TOCSY connectivity-diagrams for all the 20 common amino acid residues can be found in Appendix 7b.

7.3.6 Chemical shift-groups of ^{1}H spins in amino acids

In terms of the chemical shift values, ^{1}H spins in amino acids can be divided into seven chemical-shift groups, from A to G, as shown in Table 7.4.

Table 7.4. Grouping of protons in amino acids in terms of chemical shifts.



Note that the boundaries for each group are loosely defined and often overlap between two neighbouring groups. The chemical-shift range of spins in Group B, for example, may span from 7.8 to 8.8 ppm and overlaps with those in Group C in a 2D spectrum.

7.3.7 Notation for spins in different chemical shift groups

The following notation will be used in our discussion to denote spins in different chemical-shift groups:

- H_{GrpA} to H_{GrpG} denote spins in each group separated by the vertical lines in Table 7.5.
- $H_{GrpC\text{-}rings}$ denotes the spins within the aromatic rings of Group C.
- $H_{GrpC\text{-}others}$ denotes other spins of Group C that are not in aromatic rings.
- H_N denotes spins of Group B.
-

7.3.8 Notation for ^{1}H spins in different residues

PAW uses the following naming rules to denote ^{1}H spins in different residues:

- A spin in a residue is denoted by the residue code plus a colon, then the proton code, e.g., K5:HN denotes the spin of the amide proton in Lys 5.
- A subscript *i* following a residue code is a general notation for any residues in the sequence, e.g., W_i:HE1 denotes the HE1 spin of any Trp in the sequence.
- X_i and X_j are general notation for the i^{th} and j^{th} residues of any type.

7.4 Two-dimensional ^{1}H NMR Spectra of Proteins

7.4.1 Notation for peaks in 2D ^{1}H NMR spectra

To denote 2D cross-peaks, PAW uses two proton codes connected by a hyphen. For example,

- K5:HN-Q4:HA denotes an inter-residue cross-peak that correlates the spins of the K5 amide-proton and Q4 alpha-proton;
- K5:HN-K5:HA (or simply K5:HN-HA) denotes an intra-residue cross-peak that correlates the spins of the K5 amide-proton and alpha-proton;
- X_i:HN-X_j:HN denotes an inter-residue cross-peak that correlates the spins of two amide-protons of the i^{th} and j^{th} residues;
- X_i:HN-X_i:HN (or simply X_i:HN-HN) denotes a diagonal-peak corresponding to the spin of amide-proton in the i^{th} residue;
- X_i:HN-X_i:HA (or simply X_i:HN-HA) denotes an intra-residue cross-peak that correlates the spins of amide-proton and alpha-proton of the i^{th} residue;
- X_i:H_{GrpA}-X_j:H_{GrpG} denotes an inter-residue cross-peak that correlates the spins of a proton in Group A of the i^{th} residue and a proton in Group G of the j^{th} residue.
- X_i:H_{GrpA}-X_i:H_{GrpG} (or simply X_i:H_{GrpA}-H_{GrpG}) denotes an intra-residue cross-peak that correlates the spins of a proton in Group A and a proton in Group G of the i^{th} residue.

7.4.2 Distribution of peaks in 2D ^{1}H NMR spectra of proteins

Because of the topic of this thesis, the following discussion will focus on three types of 2D ^{1}H NMR spectra of proteins: DQF-COSY, TOCSY and NOESY.

In terms of chemical-shift distribution of 2D cross-peaks, 2D ^{1}H NOESY spectra can be divided into 49 regions, from A1, A2, ... A7, to G1, G2, ... G7 (Figure 7.4). Again, the boundaries of any 2D regions in the figure can only be loosely defined because they are often overlapped with those of their neighbouring regions. Owing to the (ideal) symmetry of 2D NMR spectra about their diagonal, the number of distinct regions is 28. For this reason, only half of 2D NMR spectra is discussed in this thesis.

The NOESY spectrum in Figure 7.4 is from a 23-residue protein called Caerin 4.1, and has been processed with PAW. Since NOESY spectra contain intra- and inter-residue cross-peaks due to through-space dipole-dipole interactions, there are usually some peaks in almost every region, even for a small protein like Caerin 4.1. The cross-peaks in regions A4, A5, A6 and A7 are often very weak.

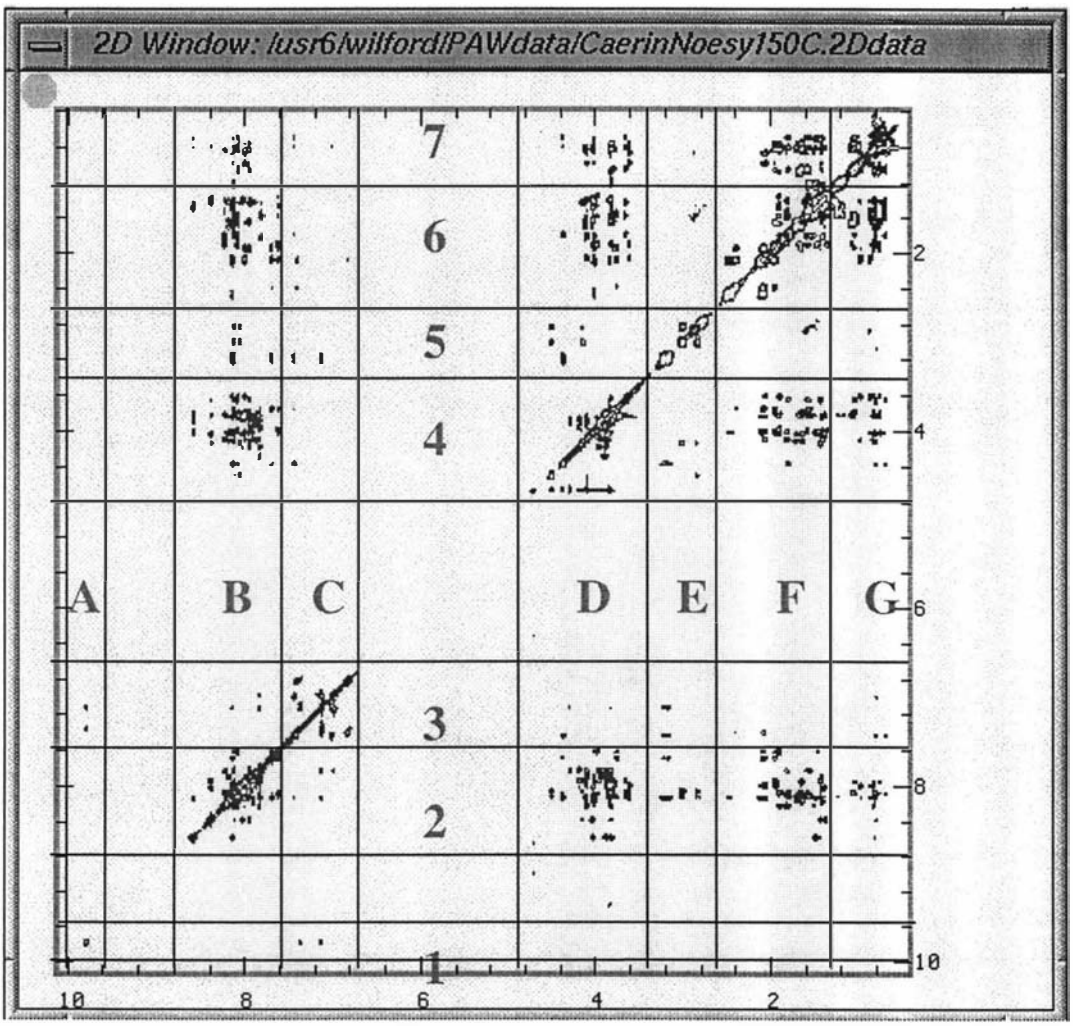


Figure 7.4: Regional division scheme for a protein ^{1}H NOESY spectrum

TOCSY spectra contain only intra-residue cross-peaks. This also means that some of the regions may not have any cross-peaks, such as the regions A2, B2 and A4 to A7 in Figure 7.5.

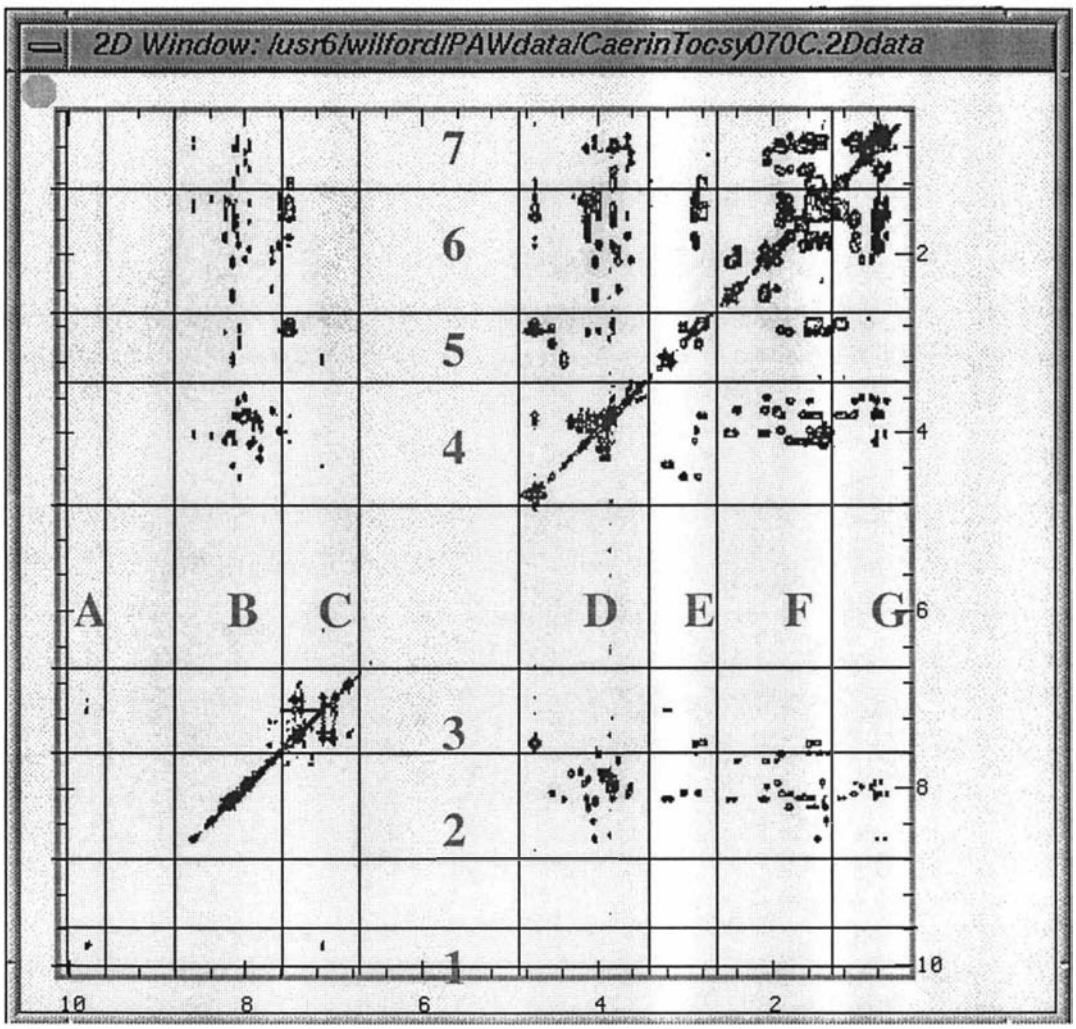


Figure 7.5: Regional division scheme for a protein ^{1}H TOCSY spectrum

DQF-COSY spectra show even fewer peaks because they contain only cross-peaks that correlate spins between adjacent ^{1}H protons that are covalently attached through nitrogen and carbon within an amino-acid block. Therefore, DQF-COSY cross-peaks should only be seen in regions A3, B4, B5, C5, D3, D4, D5, E5, E6, F6, F7, G7 and their symmetric regions about the diagonal, as indicated in Figure 7.6.

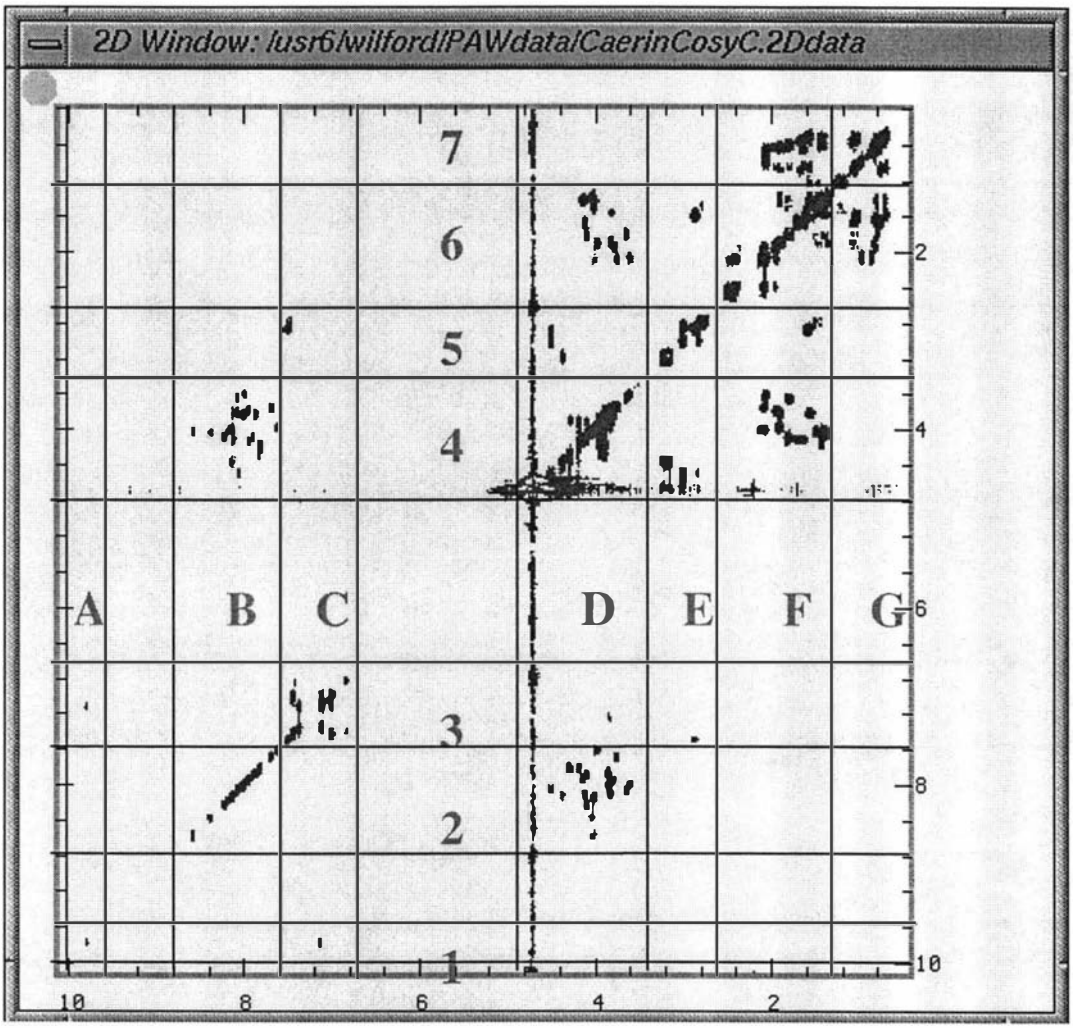


Figure 7.6: Regional division scheme for a protein ^{1}H DQF-COSY spectrum.

A summary of peaks in each region can be drawn from Table 7.4, as described in Appendix 7c.

7.5 Two-dimensional NMR Spectral Assignment of Proteins

It has been recommended [Wüthrich 1986 and Redfield 1993] that spin-system assignment be done by reference to a 2D ^{1}H DQF-COSY, mainly because many amino acids have unique spin system topologies that give rise to unique patterns of cross-peaks in 2D ^{1}H DQF-COSY spectra.

Ideally, if there is little overlap in a spectrum, then spin systems can be identified via ^{1}H DQF-COSY alone. Unfortunately, protein spectra often contain many overlapping cross-peaks. In addition, the mutiplet nature of DQF-COSY anti-phase cross-peaks often results in poor resolution, especially in a crowded area.

2D TOCSY spectra, in contrast, can give extra information with less overlap and ambiguity because their in-phase cross-peaks are much narrower. With the assistance of PAW, a more efficient strategy has been used, which relies heavily on TOCSY but not DQF-COSY for nuclear spin system identification.

7.5.1 General procedure of 2D NMR spectral assignment

Using PAW, the general procedure of 2D NMR spectral assignment consists of a few major steps, as illustrated in the next diagram.

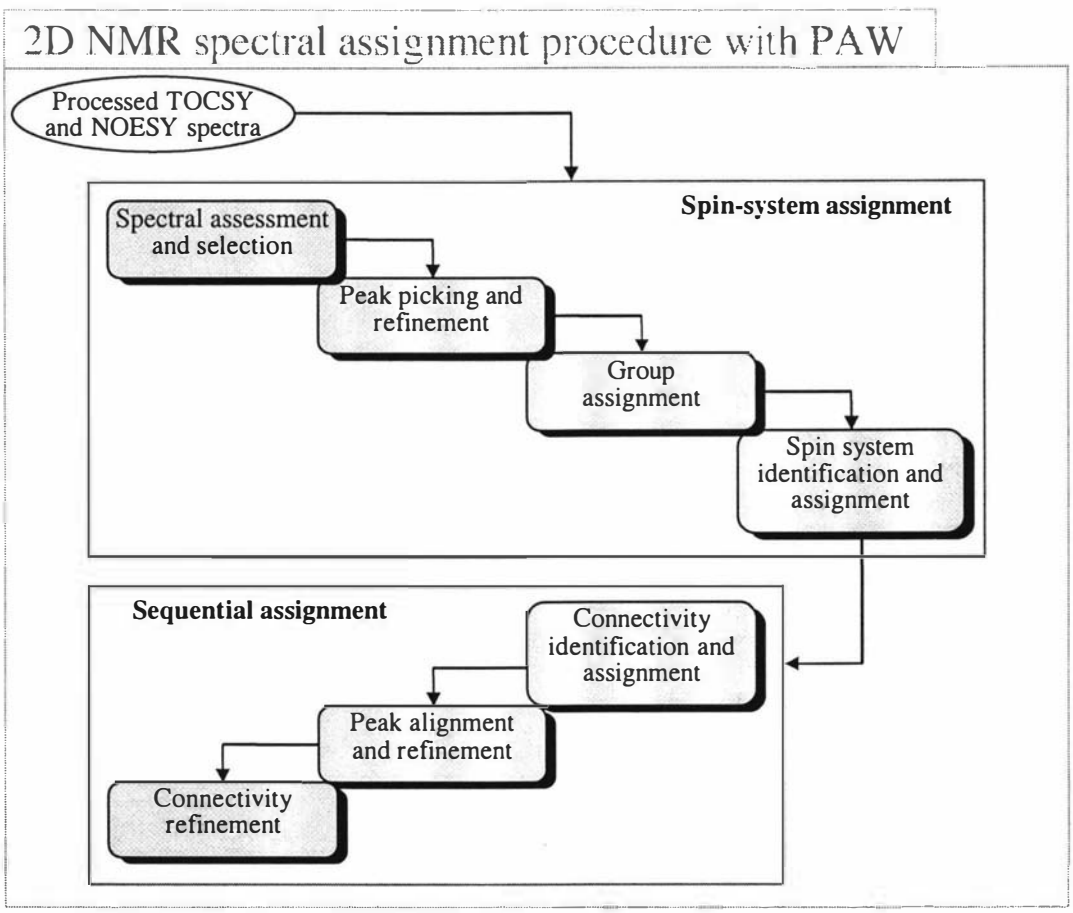


Figure 7.7: The general procedure of 2D NMR spectral assignment using PAW.

Each of the steps is described separately in the following subsections.

7.5.2 Selection of spectra for assignment

PAW is deliberately designed to assign only NOESY cross-peaks by reference to a good TOCSY spectrum. Therefore, both the NOESY and TOCSY spectra must be carefully selected. The selection strategies can be expressed as follows:

1. Select a TOCSY spectrum in which all X_i:H_N-HA and X_i:H_N-$H_{long\text{-}side\text{-}chain}$ cross-peaks can be seen.
    - If any Lys system can be identified, look for its X_i:HN-HE1 and X_i:HN-HE2 cross-peaks in Region B5. If the cross-peaks cannot be identified, the mixing time may be too short.
    - If any Leu and/or Ile system can be identified, look for its X_i:H_N-HD1 and X_i:H_N-HD2 cross-peaks in Region B7. If the cross-peaks cannot be identified, the mixing time may be too short.
    - Look at the resolution of peaks in regions B4 and B7.
    - If none of the TOCSY spectra meet these criteria individually, select a number of processed TOCSY spectra from data collected with different mixing times and combined them into one.
2. Select a NOESY spectrum by comparing it with a selected TOCSY spectrum.[5]
    - All intra-residue X_i:HN-HA cross-peaks in Region B4 must be seen and should not be badly overlapped.
    - A considerable number of inter-residue X_i:HN-X_j:HA cross-peaks in Region B4 must be seen and should be not badly overlapped.
    - The number of inter-residue X_i:HN-X_{i+1}:HN cross-peaks in Region B2 must be consistent with the helical structure determined by other methods (see section 7.5.6).
    - The number of inter-residue X_i:HN-X_j:HA cross-peaks in Region B4 must be consistent with the sheet-structure determined by other methods.

7.5.3 Peak picking and refinement

With PAW, a peak list can be created automatically and interactively using a properly processed spectrum. Because PAW has provided a number of commands (including selective peak-display commands and a flexible calibration command) that can eliminate the need to assign DQF-COSY and TOCSY spectra, it is strongly recommended that only NOESY peak-lists be produced.

7.5.3.1 The methods

The automatic peak-picking routine in PAW's program is capable of picking positive in-phase and anti-phase 2D peaks. Negative in-phase peaks are simply converted to

[5] Note that, there could be some cases in which a number of DQF-COSY peaks are missing owing to a small coupling constant or a large line-width.

positive before picking. The routine is quite complicated and requires a base level be either specified or calculated previously from a region sampled by the user. In short, it consists of the following steps:

- The user selects a region in which the peaks are to be picked, and then opens a dialog for parameter entry.
- For the selected region, the intensity of every data point inside the region is checked to see if it is greater than a base level.
- If the intensity of a data value is found to be greater than the base level, it is compared with those of the eight points surrounding it, as shown in Figure 7.8.

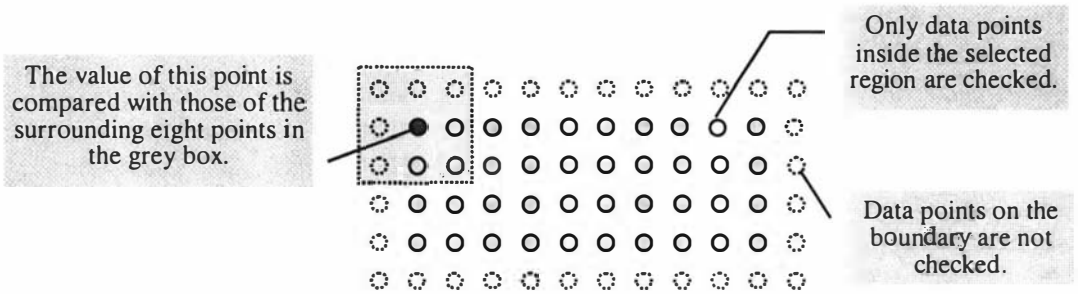


Figure 7.8: PAW's peak-picking scheme.

- If the value is higher than those of the surrounding points, this point is considered as part of a peak. The routine then calculates peak widths at four different angles (horizontal, vertical and two diagonals) and checks to see if they satisfy the minimum and maximum peak widths specified in the dialog.
- If the peak-width restrictions are satisfactory, the centre of peak is then calculated over the significant data surrounding the centre as follows:[6]

$$X_{centre} = \frac{\int xydx}{\int ydx} \quad \text{and} \quad Y_{centre} = \frac{\int yxdy}{\int xdy} \tag{7.5.3.1.1}$$

- If the peak to be picked is an in-phase peak, continue the last three steps with the next data point until all points inside the region are checked.
- If the peak to be picked is an anti-phase peak, three other adjacent peaks forming an anti-phase quartet (Figure 7.9) are sought so that the vertical and horizontal distances between their centres are about the same. Figure 7.9 shows a group of four numbered peaks that form an anti-phase quartet, where peaks 0 and 3 are positive, but peaks 1 and 2 are negative.

[6] The formula is similar to that for the calculation of the centre of mass. For example, $y_i dx_i$ in the first equation can be considered as a particle mass m_i. When performing the integration, the independent variables are always shifted to the origin so that x_i are always small, thereby putting more weight onto the higher values that have better signal-to-noise ratios.

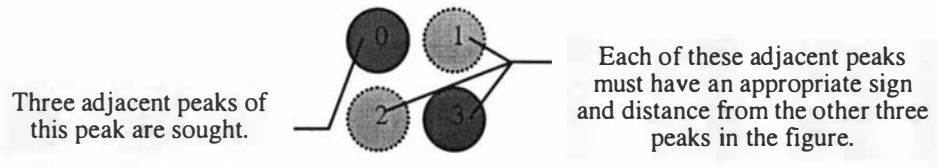


Figure 7.9: A group of four peaks that may form an anti-phase peak, where peaks 0 and 3 are positive, and peaks 1 and 2 are negative.

- The absolute intensity of each peak is then weighted with the minimum and maximum absolute intensities and then rounded to an integer. From this process, two sets of rounded weights are obtained.
- If there are more than two 1's in any set of the rounded weights, these four peaks are considered to form an anti-phase peak. This selection criterion can effectively exclude any less intense satellite peaks surrounding a DQF-COSY anti-phase peak, as shown in Figure 7.10. If the number of 1's in either set is three, the quartet is assumed to be overlapped with another anti-phase quartet and will still be picked. Heavily overlapped peaks that cannot result in three 1's in any set of the rounded weights, however, cannot be picked by this method and will be excluded.

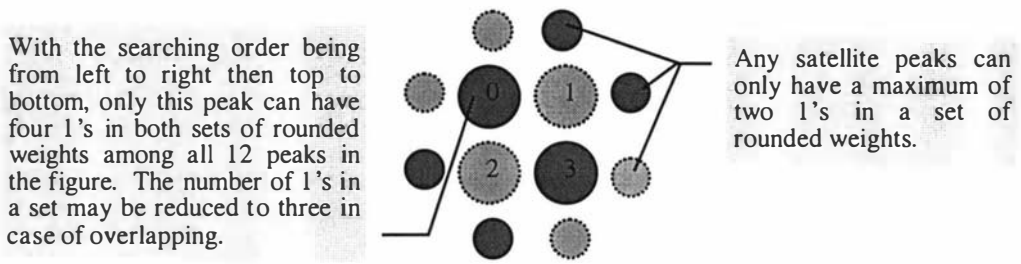


Figure 7.10: A group of 12 peaks that may form an anti-phase peak, where peaks 0 and 3 are positive, whereas peaks 1 and 2 are negative. The annotation explains how the criterion works in anti-phase peak-picking process.

- To determine the centre of the quartet, the mass-centres of the four peaks for the quartet are first calculated by equation (7.5.3.1). The mid-point coordinates (X_m, Y_m) between each pair of adjacent peaks (X_i and X_j) are also calculated as follows:

$$X_m = (X_i + X_j)/2 \quad \text{and} \quad Y_m = (Y_i + Y_j)/2 \qquad (7.5.3.1.2)$$

- The x-coordinate of the quartet centre is then determined by the mass-centres of either peaks 0 and 1 or peaks 2 and 3, depending on which pair gives a mass-centre that is closer to the mid-point between the peaks of the pair (Figure 7.11). Likewise, the y-coordinate of the quartet is determined by the mass centre of either peaks 0 and 2 or peaks 1 and 3.

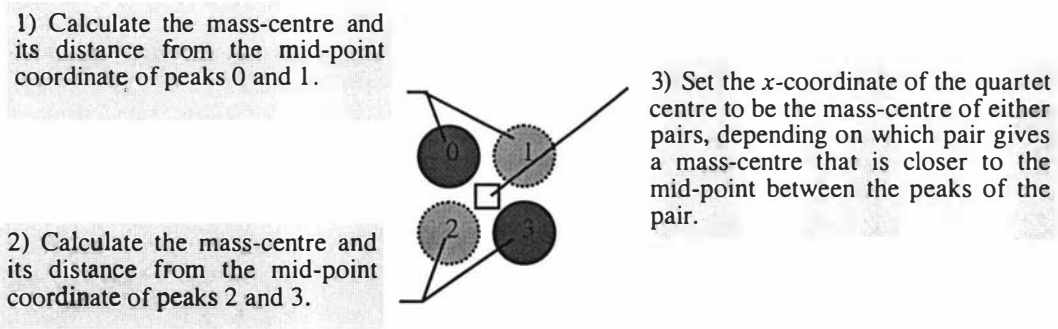


Figure 7.11: The calculation for the x-coordinate of the quartet centre.

7.5.3.2 The peak lists

To provide better references in the peak-picking and assignment process, PAW's assignment process requires that three separate peak lists be created. These are the lists for *raw peaks*, *cross-peaks*, and *diagonal peaks*. Correspondingly, there are three peak-picking commands for the different peak lists. The peaks that are picked by the methods described above are put into a raw-peak list. The cross-peaks and diagonal peaks can only be picked from one or a group of raw peaks in a raw-peak list, as described in Volume II.

The functions of the three peak lists are as follows:

- A raw-peak list contains the coordinates (in ppm) of all possible peaks in a spectrum. It is used to produce the other two peak lists and cannot be assigned.
- A diagonal peak list contains the coordinates of carefully selected raw peaks that define a curve about which the spectrum is symmetric. It is used to calculate the locations of transposed peaks, and help define the topology of transposed regions. Therefore, this list is only needed for spectra that contain diagonal peaks. Like the raw peaks, the diagonal peaks cannot be assigned.
- A cross-peak list initially contains the records of all peaks that are to be assigned. These peaks are then assigned and refined in the assignment process.

Note that, in NMR spectroscopy, the word 'transposed' refers to a mirror position about the top-right to bottom-left diagonal in a 2D spectrum.

From a practical point of view, the strategies for picking raw peaks can be as follows:

1. Pick the raw peaks region-by-region with automatic picking routine and then manually remove those peaks that have been incorrectly picked.
2. Manually add any raw peaks that cannot be picked properly — this often happens if peaks are badly overlapped.
3. Try a few picking operations with different base levels by repeatedly performing the peak-picking and removing operations. Select a relatively low base level that allows all raw peaks to be picked, including the weakest one. For in-phase peak picking, the region for the calculation of a base level can be a sub-region that

encloses an isolated weak peak (or part of any isolated peaks) and some noise. For anti-phase peak picking, it is better to specify a base level that is higher than the maximum intensities of all satellite peaks.

4. Save each satisfactory result frequently with different filenames so that you can always get back to the previous best result if anything goes wrong.

The strategies for picking the diagonal peaks and cross-peaks can be expressed as follows:

1. If there are diagonal peaks in the region, pick them immediately after picking the raw peaks because they are used for zooming into transposed regions and picking transposed cross-peaks. This is done by selecting two properly located diagonal peaks to pick up others along a line that connects the two.
2. Perform the cross-peak picking only after the other two types of peaks are completely picked.
3. If there are diagonal peaks in the spectrum, specify to pick also transposed peaks when picking cross-peaks. Zoom into transposed regions to work with each raw peak that does not have a transposed element. If it is a true peak, draw a transpose rectangle by clicking on the peak. Then, locate its transposed element and manually add a raw peak there. Keep doing this until all missing transposed peaks are located and then pick the transposed cross-peak again.
4. The coordinates of the cross-peaks can be identical to those in the raw-peak list at this stage and will be slightly different when adjustments are made in the assignment process.
5. Save each satisfactory result frequently with different filenames so that it is possible to retrieve the previous best result if anything goes wrong.
6. Finally, assess the quality of cross-peaks picked to see if they are suitable for assignment.

Note that:

- PAW allows three peak-list sets to be loaded: a main peak-list set and two reference peak-list sets. Each of the sets contains a raw-peak list, a diagonal-peak list and a cross-peak list. To avoid confusion, only the cross-peaks in the main peak-list set can be assigned.
- Loading an existing peak list as a main peak list will replace the main peak list contents in the computer memory. Also, any changes made to the peak lists will only be kept when they are saved into a file.
- When creating a peak list, it is better to have more peaks than less. Many weak peaks may be useful for spin system and connectivity identification.
- The efficiency of the peak-picking routine depends heavily on the appropriateness of the phases and baselines of a spectrum. This is more critical for anti-phase peak picking, which, fortunately, is less important in the assignment procedure used by PAW. If in any regions the raw peaks cannot be properly picked, it may be necessary to further correct the phases and baselines.
- Although manually lining-up peaks by removing and picking individual peaks is possible, it is not recommended. With PAW, precise peak positions can be

obtained using an automatic peak-alignment routine at the later stage of assignment. This step relies on a well-isolated peak in a group. PAW also has a relatively reliable routine to calculate the centre of a peak even if they are not well isolated, as will be explained later.

- PAW provides different colours and symbols by which the user may identify different types of peaks. It also has a checkbox for the user to specify the type(s) of peaks to be displayed.
- Whether a raw peak is truly a cross-peak can often be judged by the possession of a transposed element. To calculate the location of a transposed element of a peak, PAW uses diagonal peaks that are closest to the x and y-coordinates of the peak. Therefore, the search for the transposed peaks is often quite effective, even in a region heavily affected by the Bloch-Siegert shifts (Note 1 and 4, Figure 7.12). If the diagonal peak corresponding to Peak A is in the diagonal-peak list, the y-coordinate of Peak B will be obtained exactly. Likewise, if the diagonal peak of Peak B (i.e., Peak E) is in the diagonal-peak list, the exact x-coordinate of Peak B will be obtained (Note 3, Figure 7.12). For this reason, it is strongly recommended that the diagonal peaks be picked carefully in each region.

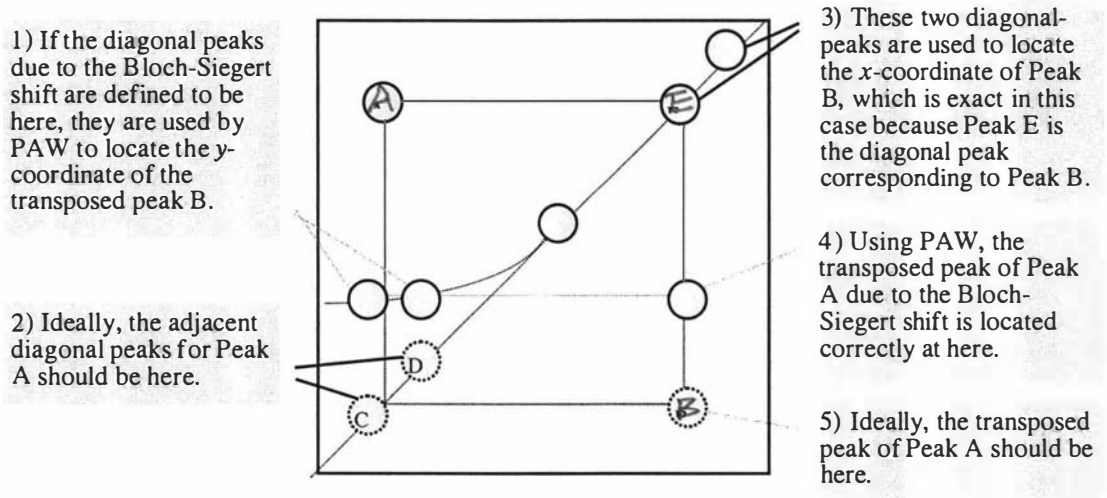


Figure 7.12: PAW's method for locating a transposed peak. This method ensures a precise search for a transposed peak even in a region heavily affected by Bloch-Siegert shifts.

7.5.4 Group assignment

Once all the possible cross-peaks are carefully picked, assignment can be started by assigning the proton codes of cross-peaks in groups. For example, all proton codes in the vertical direction for cross-peaks in column A and B (see Figure 7.4) can be assigned to HN. Similarly, all proton codes in the vertical direction for cross-peaks in column D can be assigned to HA — of course, some of them are HB but they can be easily corrected once they are identified.

PAW provides a command for group assignment. By selecting the command, the user will be urged to select a region, and then fill in a dialog for the proton code and direction in which all cross-peaks in the region are to be assigned.

From the practical point of view, the following procedure is recommended:

1. Assign the proton codes in the vertical direction for all cross-peaks in columns A and B to be HN, D to be HA, E to be HB, and G to be HG (see Figure 7.4).
2. Repeat operations in Step 1 for corresponding regions in the horizontal direction.
3. Save the peak list with a different name.

Note that, to avoid duplicated effort, it is strongly recommended that spectral assignment be started only after the cross-peak list is satisfactory.

7.5.5 Spin system identification and assignment

The spin-system identification is done mainly by observing the cross-peak distribution in a TOCSY spectrum, especially in regions B4, B5, B6 and B7[7]. There are a number of strategies that can be used to identify protein spin systems. Using PAW, the following procedure is recommended:

1. Prepare an amino acid list and write down the number of occurrences of each amino acid type in the peptide sequence as a reference. For example, the protein sequence of Caerin 4.1 is:

   GLWQKIKSAA GDLASGIVEG IKS-NH_2

   Accordingly, the amino acid list is: (The last column is used in the assignment process for the number of amino acids that have been and are still to be identified.)

Amino acid codes	Number in the sequence	Number identified
G	4	
L	2	
W	1	
Q	1	
K	3	
I	3	
S	3	
A	3	
D	1	
V	1	
E	1	
Total	**23**	

   Except for the total in the last row, each number in the table corresponds to the number of observable TOCSY cross-peaks for each pair of correlated nuclear spins in a spin system. For example, in the B4 region, one can expect to see four HN-HA cross-peaks for residue G, two for L, one for W, three for K, and so on. In addition,

[7] Alternatively, the region D2, E2, F2 and G2 can be used.

there are two HN-HB peaks for each S; and hence a total of 29 peaks must be seen in the B4 region of a TOCSY spectrum.

2. Target less-crowded regions first. For example,

    - Trp intra-residue cross-peaks can be identified in regions A1 to A5 of DQF-COSY, TOCSY and NOESY spectra, especially A1 and A3.

    - Lys and Arg intra-residue cross-peaks are uniquely seen in Region E6. Their long side-chain structures also contribute to columns of TOCSY cross-peak patterns in regions B4 to B7, and are often easily identified. To separate Lys from Arg, Region F7 of a TOCSY spectrum can be closely inspected, since no Arg cross-peak should be seen there.

    - The rest of the high-field intra-residue cross-peaks in Region B7 should belong to either Val, Ile or Leu systems, which will be called the *VIL group* in this thesis. Often, they can be unambiguously assigned during the sequence-specific assignment.

3. Target unique intra-residue cross-peaks patterns of some spin systems. For example,

    - Proline is the only system that has no cross-peak in all regions beyond 4.7 ppm. In other regions, the distribution pattern of the Pro spin system is similar to that of Arg.

    - Glycine intra-residue cross-peaks can only be seen in regions B4 and D4. Their HN-HA DQF-COSY cross-peaks have a prominent split doublet arising from a geminal coupling constant of about 15 Hz, as reported by Wüthrich (1986).

    - Alanine and threonine intra-residue cross-peaks can only be seen in region B4, B6, D4 and D6. They can usually be easily identified by their relatively strong intra-residue cross-peaks in region B6. Then, look for HN-HB and HA-HB of Thr systems in regions B4 and D4 respectively to separate Thr systems from Ala systems.

    - HA-HB intra-residue cross-peaks for Thr, His and Ser systems (or the *THS group*) can be sought in Region D4, which can be further confirmed by looking at the well aligned HN-HA and HN-HB cross-peaks in Region B4. Thr systems can be singled out easily with the tip given in the last step. Hopefully, His systems can be singled out by the presence of their HN-HD1 or HN-HE1 cross-peaks in the B3 region; otherwise, the identification has to rely on sequence-specific assignment.

    - Cys, Asp and Asn are the only non-aromatic systems that have two HB related cross-peaks in Region B5 and D5, one HB-HB cross-peak in Region E5, and nothing else beyond these regions except D4.

    - Other non-aromatic cross-peaks belong either to the Met, Glu or Gln systems (or the *MEQ group*). With some luck, Gln can be separated from them by

inspecting their HE2-related cross-peaks; however, they are often too weak to be identified.

- The rest of the unidentified aromatic cross-peaks should belong to either Phe or Tyr systems (or the *FY group*).

In summary, it is possible to identify all spin systems in 2D NMR ^{1}H spectra of small proteins. Difficulty may occur for unambiguous identification of some spin systems. They can be separated into groups, such as the VIL, THS, MEQ and FY groups. If the assignment is for homonuclear ^{1}H spectra only, the identification has to rely on the sequence-specific assignment.

PAW provides a number of commands for spin system assignment. For example, the assign-one-peak command urges the user to select a peak to be assigned, and then open a dialog for the user to specify a number of parameters. The command for lining up peaks allows a number of cross-peaks to be aligned and assigned in one operation.

From the user point of view, the following procedure is recommend:

1. Display the NOESY peak-list on a TOCSY spectrum and assign the spin system and proton codes for any peaks that can be easily identified. If necessary, re-calibrate the TOCSY spectrum by reference to a NOESY peak in the region, which is not necessarily the water peak. (The NOESY spectrum can also be re-calibrated — just make sure that the chemical shifts of all other peaks in the peak-list be changed accordingly.)
2. Assign the amino-acid code and proton codes for the cross-peaks in column D once a Ser, Thr, Pro or His system is identified. Perform the same operation for the cross-peaks in column E for any identified Lys and Arg systems, as well as those in column G for any identified Ile and Leu systems.
3. Repeat operations in Step 2 for corresponding regions in horizontal direction.
4. Save the peak list with a different name.

Note that:

- Using PAW, any peak list and assigned results can be displayed on any spectrum and hence it does not matter which spectrum is used as a reference for the assignment. For example, cross-peaks in a peak list can be assigned while displayed on top of a TOCSY or a DQF-COSY spectrum. However, do not expect the locations of peaks to be exactly the same between spectra, because many factors may affect the chemical shifts of nuclear spins in different experiments.
- There is nothing to stop a user from picking and assigning peaks on TOCSY, DQF-COSY, and any other spectra. In fact, they are often needed for the presentation of final assignment results.

7.5.6 Inter-residue connectivity identification and assignment

Protein NMR sequence-specific assignment assigns the NOE cross-peaks of each identified spin system to a specific residue in the protein sequence. This is achieved by identifying NOE cross-peaks due to through-space dipole-dipole magnetisation transfer between nuclear spins of adjacent residues in the sequence. The assignment

results can give a 3D picture of through-space connectivity between ^{1}H protons of different residues, and often provide further evidence and clues for the assigned and unassigned spin systems from the confirmed primary and secondary structure of the protein.

It is well known that the observable NOE cross-peaks correspond to three short range (or sequential), five medium-range and three long-range ^{1}H-^{1}H distances [Schulz and Schirmer 1979, Richardson 1981, Wüthrich 1986]. Table 7.5, for which the data were drawn from Wüthrich 1986, shows the statistical ^{1}H-^{1}H distances in proteins. The distances are measured in Å. Those less than 3.0 Å are printed in shadow. The long-range distances are for those between backbone ^{1}H protons in non-twisted, regular β-sheet. Those distances larger than 4.0 Å correspond to weak NOE cross-peaks, whereas those less than 3.0 Å correspond to strong NOE cross-peaks.

Table 7.5. ^{1}H-^{1}H distances in proteins, where the long-range distances correspond to those between backbone protons in non-twisted, regular β-sheet. The distances are measured in Å. Those less than 3.0 Å are printed in shadow. Note that a tight turn has four residues; the proton distances between residues 2 and 3, 3 and 4, 2 and 4, 1 and 4 are indicated with superscript a, b, c and d, respectively. (The data shown here are from Wüthrich 1986.)

distances	β-sheet	β_p-sheet	α-helix	3_{10}-helix	Type-I turn		Type-II turn	
Sequential distances								
$d_{NN}(i,i+1)$	4.3	4.2	2.8	2.6	2.6[a]	2.4[b]	4.5[a]	2.4[b]
$d_{\alpha N}(i,i+1)$	2.2	2.2	3.5	3.4	3.4	3.2	2.2	3.2
$d_{\beta N}(i,i+1)$	3.2-4.5	3.7-4.7	2.5-4.1	2.9-4.4	2.9-4.4	3.6-4.6	3.6-4.6	3.6-4.6
Medium-range distances								
$d_{NN}(i,i+2)$			4.8	3.8	3.6[c]		3.3[c]	
$d_{\alpha N}(i,i+2)$			4.2	4.1	3.8		4.3	
$d_{\alpha N}(i,i+3)$			3.4	3.3	3.1-4.2[d]		3.8-4.7[d]	
$d_{\alpha N}(i,i+4)$			4.2					
$d_{\alpha\beta}(i,i+3)$			2.5-4.4	3.1-5.1				
Long-range distances (j > i+4)								
$d_{NN}(i, j)$	3.3	4.0						
$d_{\alpha N}(i, j)$	3.2	3.0						
$d_{\alpha\alpha}(i, j)$	2.3	4.8						

The table can be used to help recognise the secondary structure of proteins. For example, β-sheet structures give rise to NOE cross-peaks corresponding to sequential distances $d_{NN}(i, i+1)$, $d_{\alpha N}(i, i+1)$ and $d_{\beta N}(i, i+1)$, for which the intensity of $d_{\alpha N}(i, i+1)$ should be higher. Helical structure, in addition, also gives rise to NOE cross-peaks corresponding to medium-range distances, as well as intense cross-peaks corresponding to $d_{\alpha N}(i, i+3)$ and $d_{\alpha\beta}(i, i+3)$.

With PAW, the sequence-specific assignment is done using a cross-peak list for a NOESY spectrum. A number of strategies can be used for the sequence-specific assignment. For example, a strategy can be as follows:

1. Start the assignment from the cross-peaks of a residue that is easily identified. Usually the peaks belong to a residue type that is the least common in the sequence. Assign as many neighbouring residues as possible by identifying the inter-residue connectivities shown in the NOESY spectrum. This is done primarily in region B2 and then B4 by extending the assignment from both ends of any identified spin systems to build up identified peptide segments. (See Chapter 11 for detailed examples.)
2. Compare the identified segments with the protein sequence to identify the segments' location in the sequence. Mark the identified segments in the protein sequence with a highlighter.
3. Repeat the last two steps with the next least common residue type. Keep doing this until nothing can be done further with this method.
4. For all residues that show a strong $d_{NN}(i,i+1)$ in B2 region, assign any cross-peaks related to medium-range connectivity (see Table 7.5). These cross-peaks must be related to a helical or tight-turn secondary structure.
5. Produce an NOE summary for the peaks assigned. (PAW can create this automatically.) Use this as a reference to figure out the secondary structure, then predict any unassigned peaks and correct any inconsistent assignment according to the structure. Keep in mind that protein secondary structures can also be predicted by methods other than NMR study, such as *circular dichroism measurement* [Sonnichsen *et al.* 1992] and *bioinformatics* [Gaëta 1998]. Any published results can also be use as a reference in the NMR spectral assignment.

Note that :

- The NOESY and TOCSY spectra used in the sequence-specific assignment must be acquired with the same experimental parameters such as temperature, spectral width, etc. It is strongly recommended that the two spectra be processed to the same data size and calibrated with the same reference.

Chapter 8:

NMR Spectral Processing of Caerin 4.1 Using PAW

8.1 Introduction

Six 2D NMR raw data sets of the Caerin 4.1 protein were collected and provided by Professor John Carver (Wollongong University, Australia). These include two TOCSY [Davis and Bax, 1985], one double-quantum filtered DQF-COSY [Rance *et al.* 1983] and three NOESY [Jeener *et al.* 1979] spectra. They were collected at the University of New South Wales on a 600.13 MHz Bruker DMX600 spectrometer at 25°C in a 5 mm triple resonance probe using the TPPI method [Drobny 1979, Bodenhausen 1980, Marion & Wüthrich 1983] and the digitally filtered and over-sampling technique [Smallcombe *et al.* 1996]. Each of the 2D data sets consists of 512 rows of 1024 point time-domain complex data.

Only the processing results from three of the Caerin 4.1 NMR raw data sets will be discussed in this chapter. These are the DQF-COSY, the TOCSY collected with 70 ms mixing time, and the NOESY collected with 150 ms mixing time. They will be called the Caerin 4.1 DQF-COSY, TOCSY070, and NOESY150 in the thesis.

A high-resolution processing technique was consistently used to process the 2D data sets. In short, the resolution improvement was achieved by the application of large first-order phase correction to the time-domain data in D1 and linear prediction in the time-domain data in D2. The former operation allowed the use of filters that minimised unnecessary suppression of the initial parts of the time-domain data sets in D1 for the TOCSY070 and NOESY150 data, thereby increasing the peak intensities in the NMR spectra. The latter allowed the use of filters that maximised the useful information in the trailing parts of the time-domain data sets in D2 for all the data sets, thereby narrowing the peak width.

Each row in each of the data sets was processed with an enhanced sine-bell filter, transformed with a fast Fourier transform, then followed by proper phasing and baseline correction. After the Fourier transform in D1, 1024 columns of 512 point time-domain real data in D2 were obtained. Each column of this real data set was extended with 512 zeros, filled with data obtained by a linear-prediction operation, filtered with an enhanced sine-bell function, transformed with a real Fourier transform, then followed by proper phasing and baseline correction.

8.2 Processing the Caerin 4.1 DQF-COSY NMR data

8.2.1 The raw data

Figure 8.1 shows an intensity plot of the Caerin 4.1 2D DQF-COSY raw data.

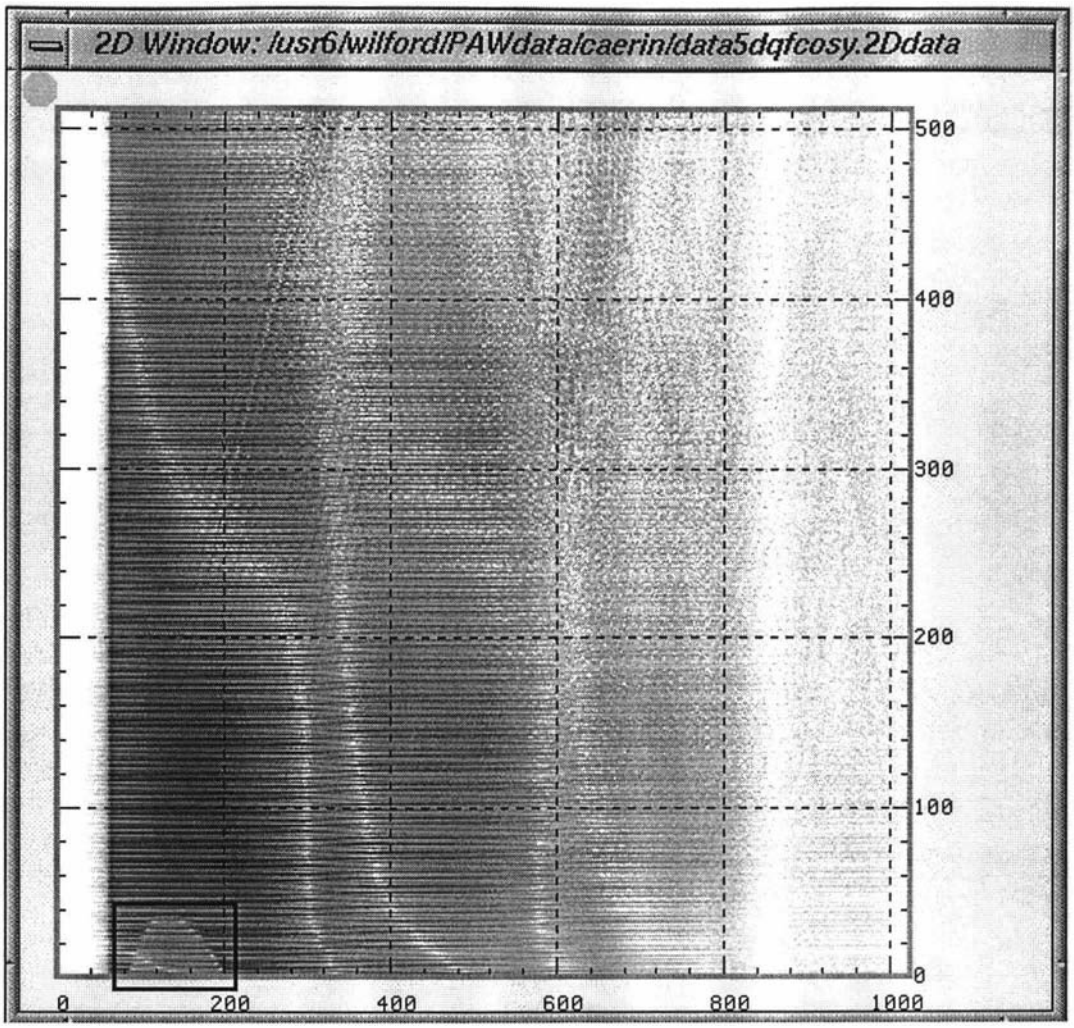


Figure 8.1: The intensity plot of the Caerin 4.1 DQF-COSY raw data. The white area on the left contains weak data added by the digitally filtered and over-sampled method used in the experiment. The blocked area at the lower-left corner contains cut-off values due to receiver overflow.

Each row of the data set is dominated by the water signal due to the failure of solvent-signal suppression during the experiment. Figure 8.2 shows the 1D real and imaginary plots of four typical rows from the 2D data set.

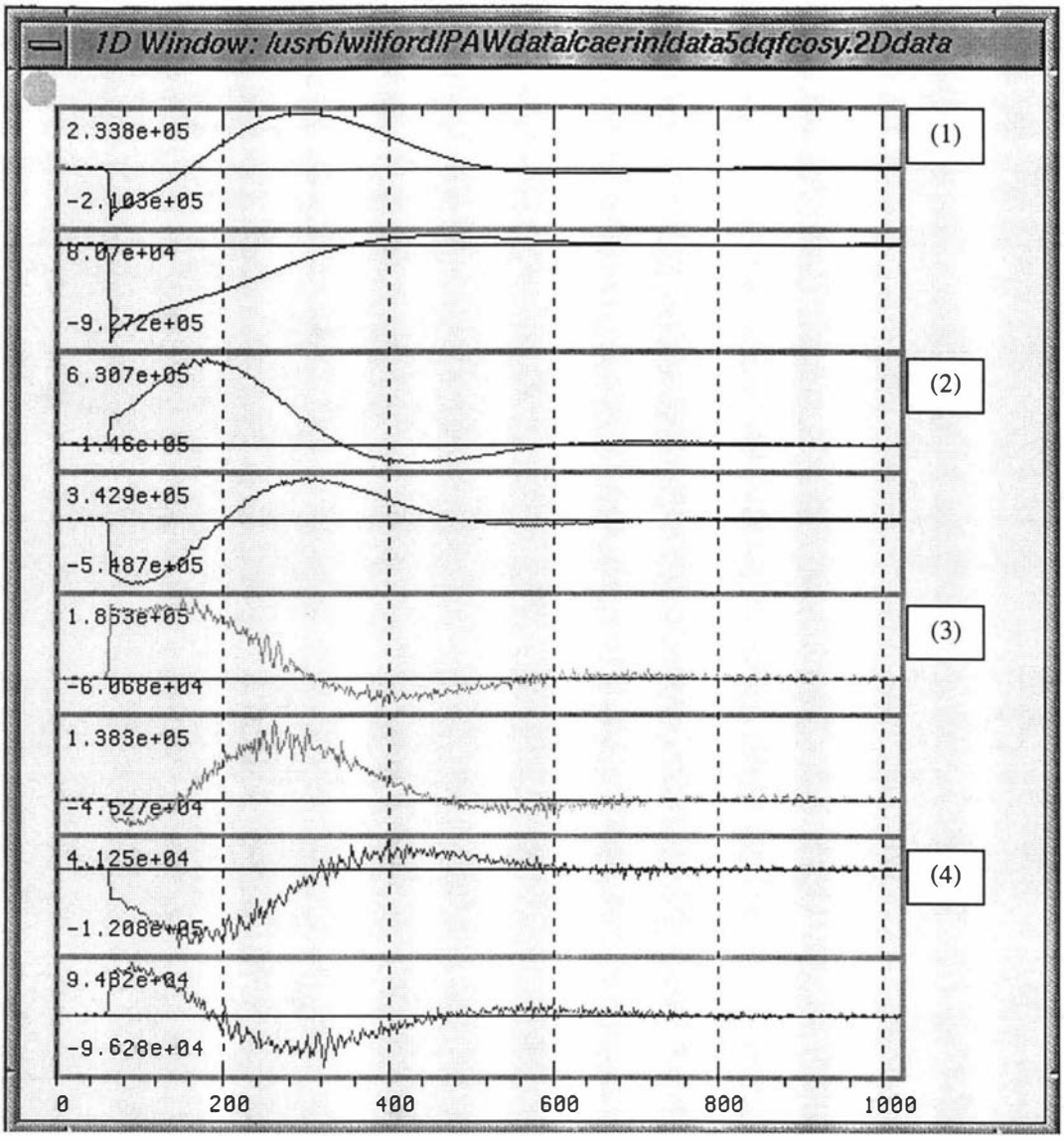


Figure 8.2: The real and imaginary parts of Row 0, 200, 400 and 511 of the Caerin 4.1 DQF-COSY.

Each row of the data set contains a relatively weak filter-response build-up curve [Smallcombe *et al.* 1996] at the beginning, as shown in the expanded view of a 1D region in Figure 8.3.

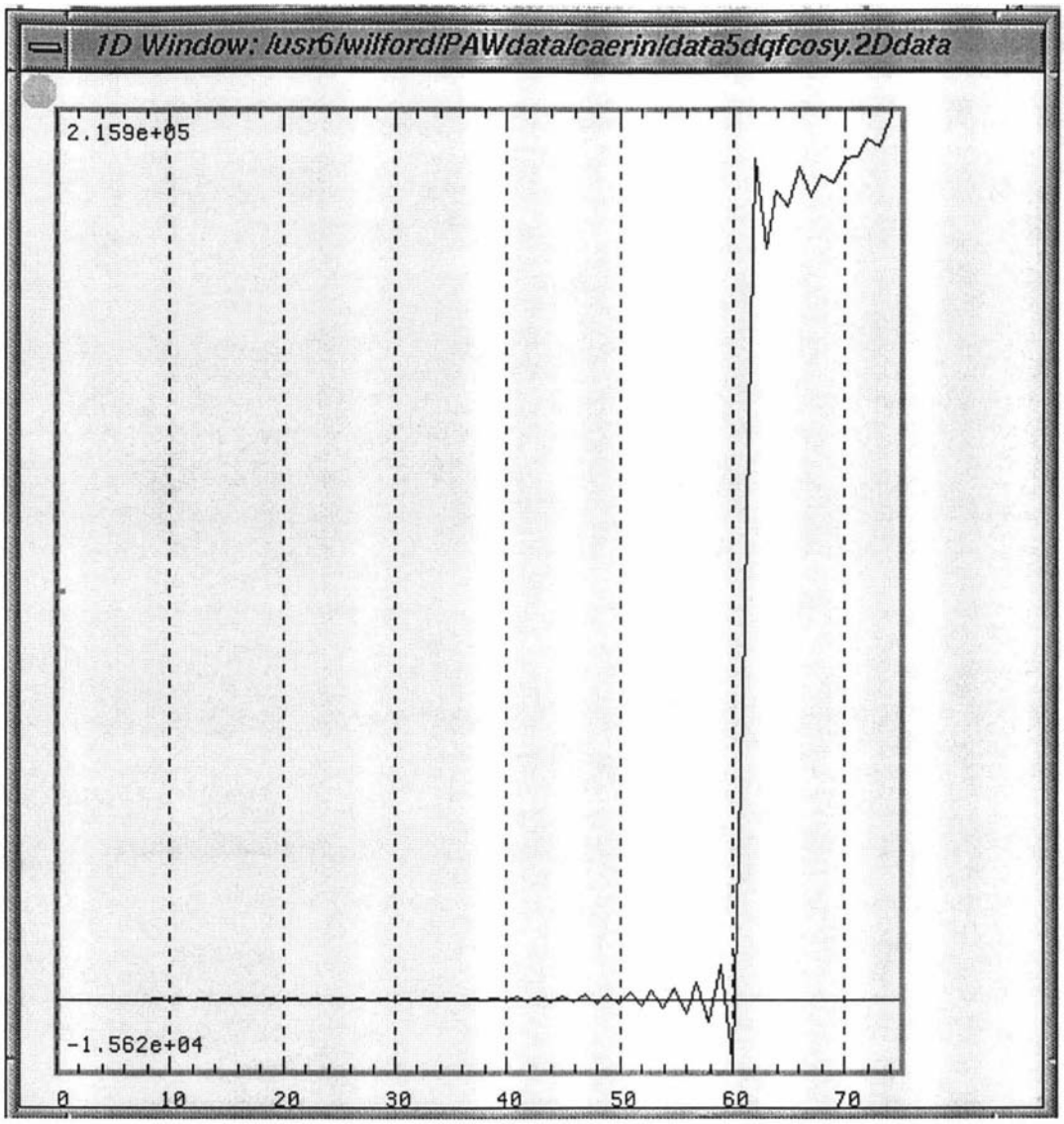


Figure 8.3: An expanded view of a time-domain data set characterised by a filter response build-up curve at the beginning.

The blocked area at the lower-left corner of Figure 8.1 contains spurious values due to receiver overflow. These happened in rows 1, 3, 5, ... 89. Four of them are shown in Figure 8.4.

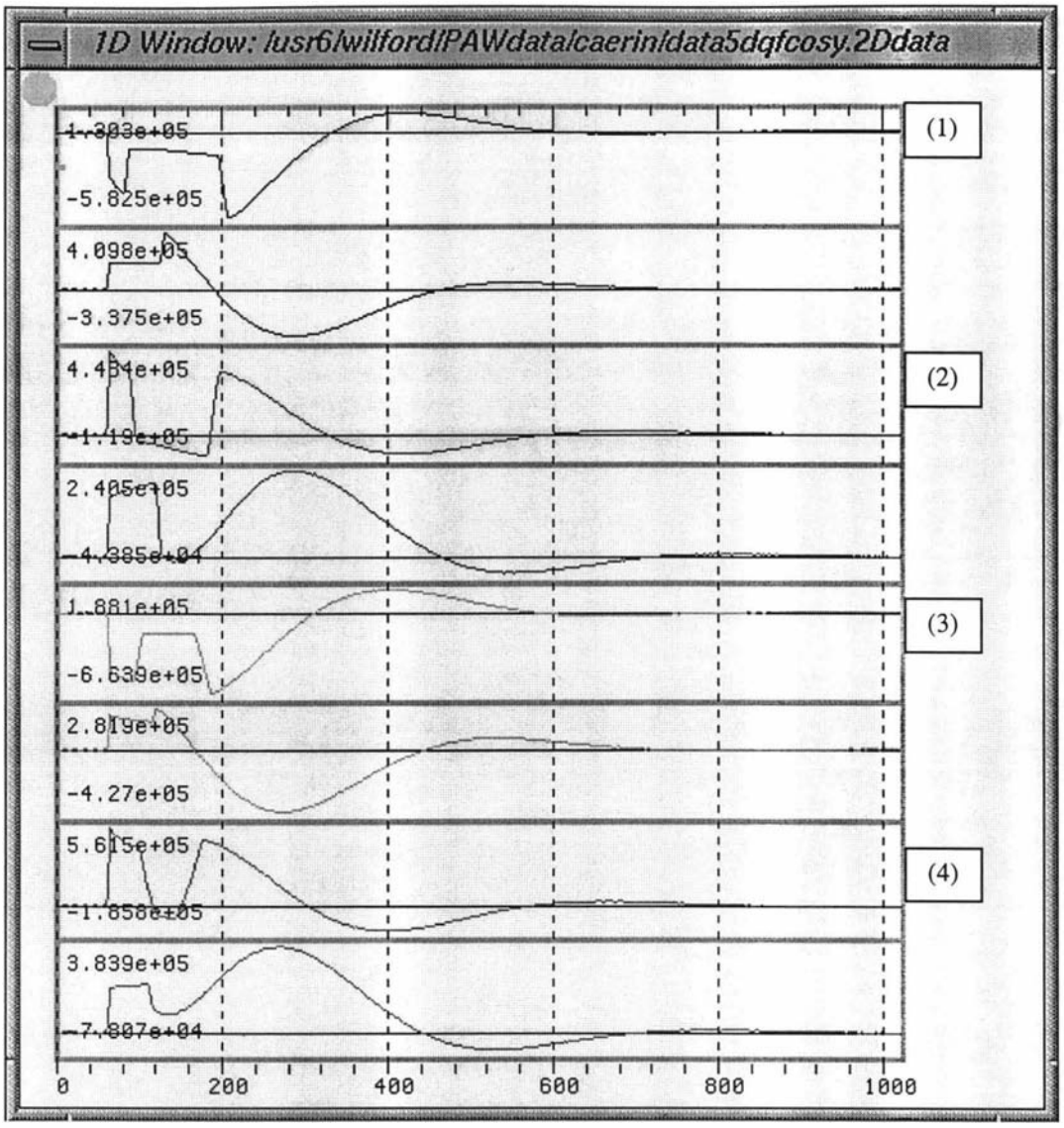


Figure 8.4: The real and imaginary parts of rows 1, 11, 21 and 31 of the Caerin 4.1 DQF-COSY. The problems due to receiver-overflow are characterised by a platform-like region in each of the 1D time-domain plots.

Apparently, the signals collapsed once their intensities reached a certain level. In this case, no critical intensity could be deduced because the problem areas did not show a consistent intensity at which the overflow started.

8.2.2 High-resolution processing operations and results

8.2.2.1 D1-processing

A digitally filtered and over-sampled 1D data set with build-up at the beginning of the FID can be processed with a very large first-order phase correction.

Since this is a DQF-COSY NMR data, the FIDs collected at around 20% of D2 usually contain the most intense signals. This can be justified by comparing the maximum intensities of the FIDs or inspecting a number of columns after processing

in D1. For this reason, row 100 was randomly selected to work out the processing parameters in D1. Figure 8.5 shows the time-domain data of the row and the processed results.

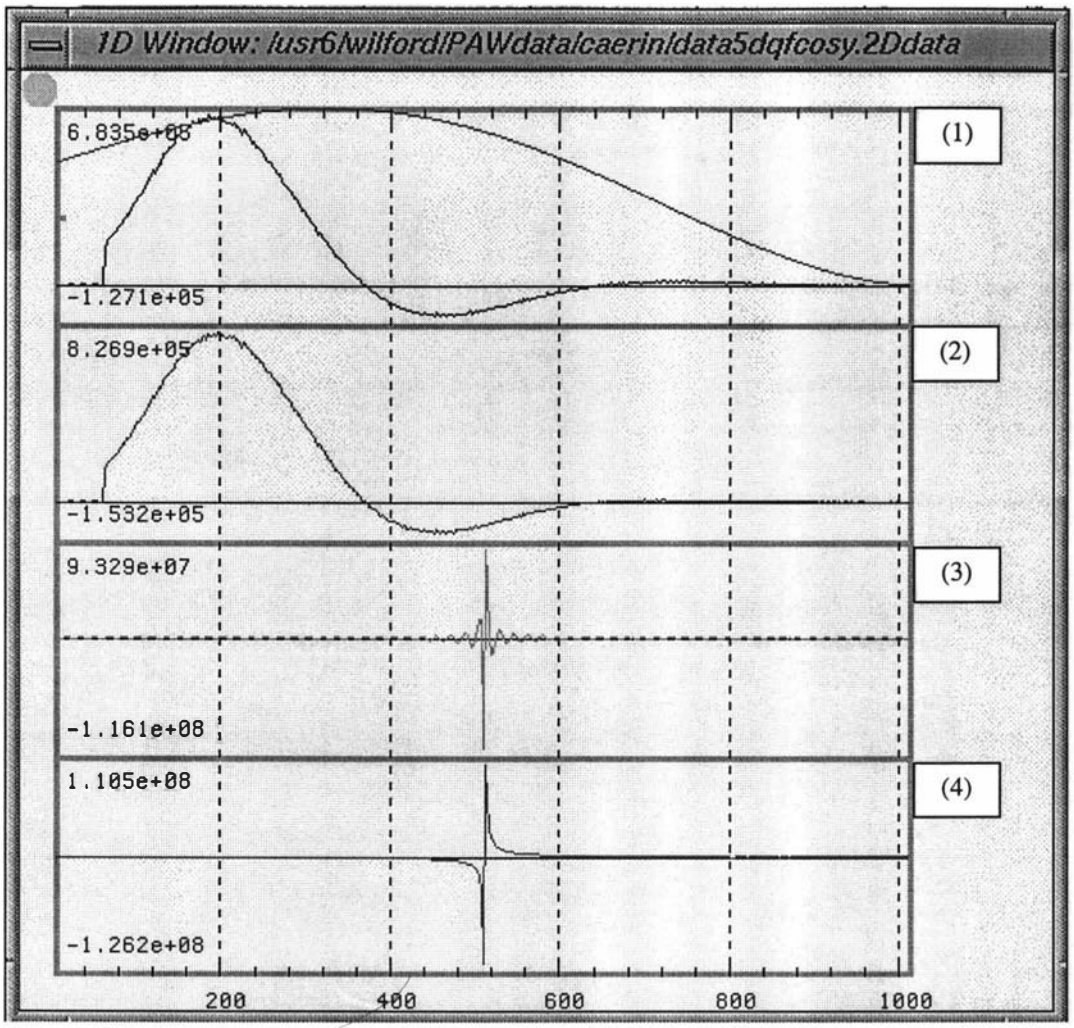


Figure 8.5: Row 100 of the Caerin 4.1 DQF-COSY and the processed results. The first plot is the raw data with the filter to be applied. The second plot is the data after applying the filter. The third plot is the frequency-domain NMR spectrum. The last plot is the spectrum after phase correction.

The raw data was filtered with the enhanced sine-bell filtering command esb(1024,70,2,-100). In the command, the first three parameters in the brackets are, from left to right, the end-point from which the filter becomes zero, the angle by which the sine function is phase-shifted, and the degree of the sine-function. The last parameter is the degree of exponential enhancement. With the shift-angle set to 70°, the filter applies a bit of suppression to the initial region. To increase the resolution further, an exponential enhancement of −100 units was applied to increase the weight of the middle region. It also forces the data at the end of each row to gradually reduce to zeros, thereby suppressing noise.

DQF-COSY spectra are usually presented with anti-phase peaks. According to the first-value theorem of the Fourier transform, the volume of a peak in an NMR spectrum is equal to the initial value (M_0) of its time-domain data [Bracewell, 1986]. For a fast Fourier transform from 0 to N, the resulting peak volume is $0.5NM_0$ [see

Appendix 11b, 11c and 11d], which is still directly proportional to M_0. Because a DQF-COSY time-domain data set always starts from zero, having both positive and negative components in the peaks is inevitable because the peak volume must be zero. Presenting a DQF-COSY spectrum with in-phase peaks gives big negative lobes around every peak centre, and hence is not recommended.

Conventional processing of DQF-COSY data sets often uses filters that are zero at the beginning. Often, a first-degree cosine function that starts with zero is chosen. This is a natural choice as it fits the profile of the DQF-COSY time-domain data.

For the digitally filtered Caerin 4.1 DQF-COSY NMR data that are filled with weak signals at the beginning, there was no need to use a filter that starts with zero during D1-processing. Instead, a shifted cosine filter has been applied to provide some suppression to the damaged parts of the data set.

Figure 8.6 shows the vertically scaled frequency-domain plot of the row. As expected, oscillations are observed due to the very large first-order phase-shift. (The additional 62 initial points of weak values correspond to a distortion that result in 61 maxima in the spectrum. See also Chapter 6 on the baseline correction.)

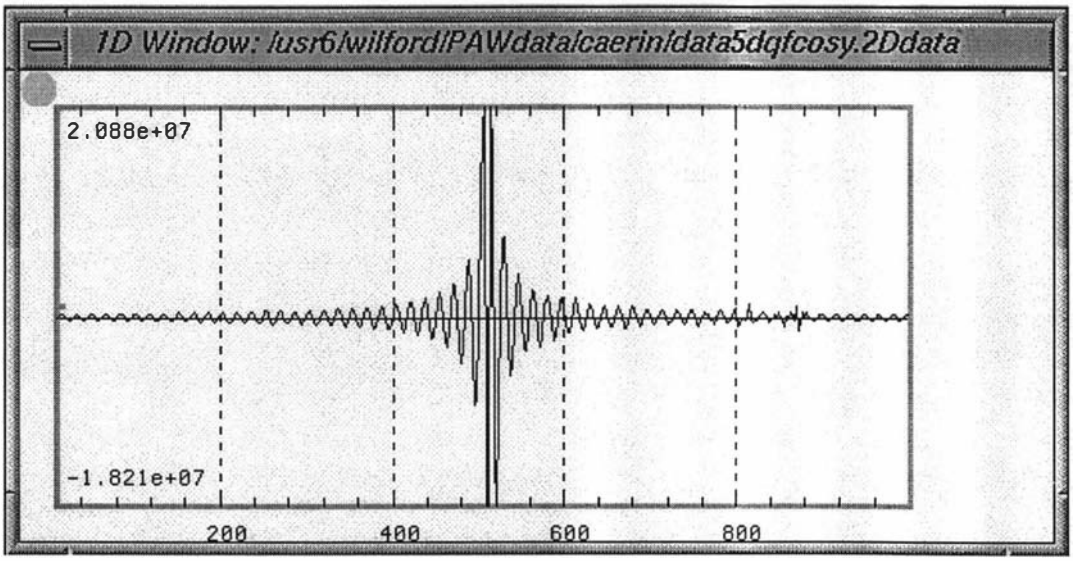


Figure 8.6: The vertically scaled frequency-domain spectrum of Row 100 of the Caerin 4.1 DQF-COSY before phasing.

With the pivot set to 0 and base region set to [340,480], PAW's auto-phase routine resulted in −53.6° for the zeroth-order phase correction and −22054° for the first-order phase correction in D1 (Figure 8.7).

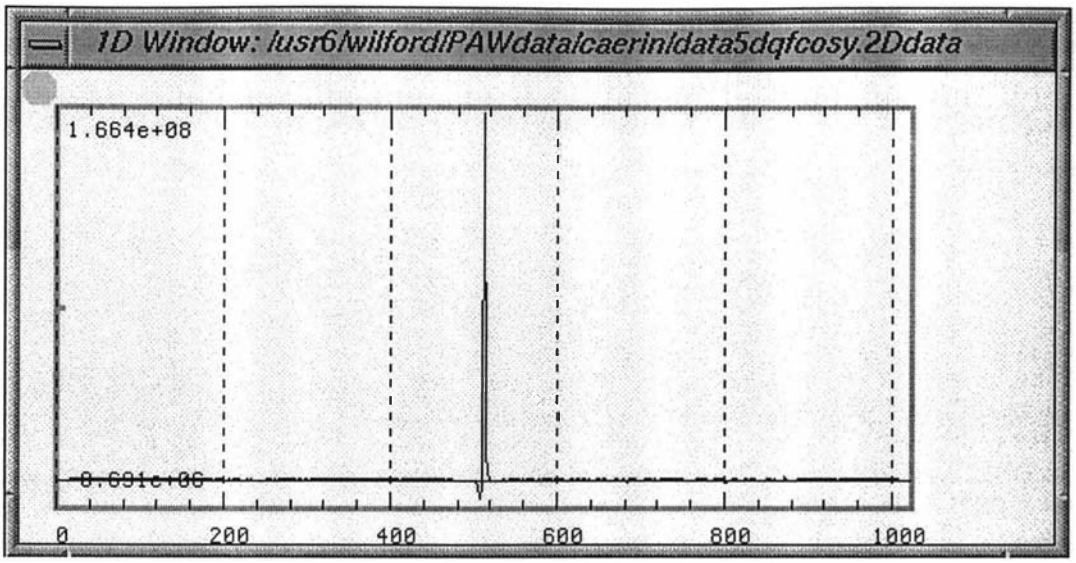


Figure 8.7: The auto-phased result of Row 100 of the Caerin 4.1 DQF-COSY.

These phasing parameters were only used as starting points for phase correction because the auto-phase routine was not designed for correcting anti-phase spectra. An expanded view of the first 200 points showed that the spectrum was not correctly phased, as indicated in Plot 1 of Figure 8.8. Nevertheless, the resulting spectrum was sufficient for obtaining proper anti-phase correction parameters. The adequate zeroth-order phase correction for the DQF-COSY spectrum was found to be either –4 or –184 degree, as shown in Plots 2 and 3 of Figure 8.8. Without an approximate first-order phase correction to start with, the search for the phasing parameters could be extremely cumbersome and time-consuming.

Note that PAW's auto-phase routine uses the minimisation of the baseline slopes on two sides of a spectrum as a criterion for the best phase. Therefore, using the most intense FID in a 2D data set becomes particularly useful for working out the large first-order phase shift required.

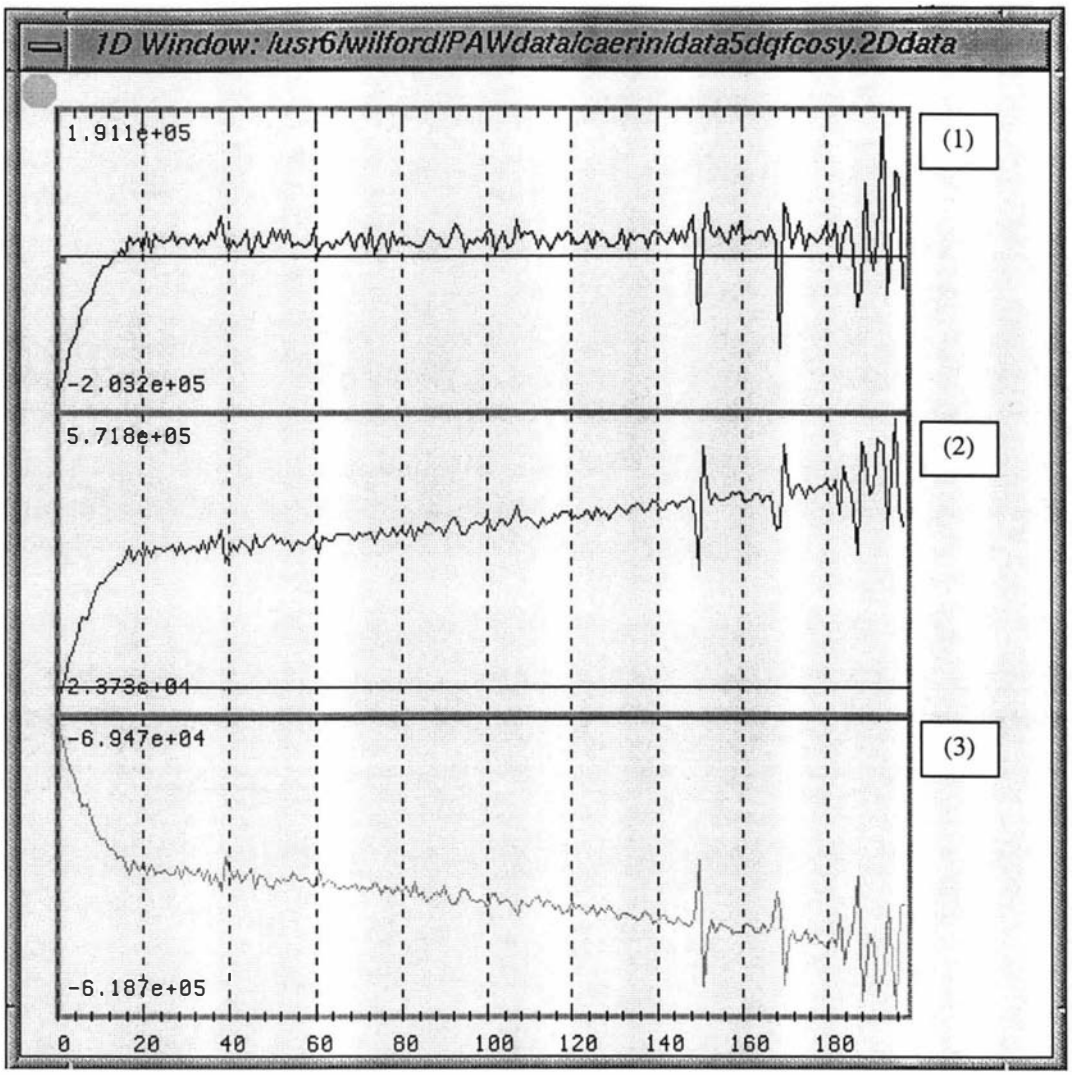


Figure 8.8: The zeroth-order phase correction of Row 100 of the Caerin 4.1 DQF-COSY. Plot 1 shows a region on the left of the row after auto-phase correction. Plot 2 and 3 shows two possible phase corrections, both of which produce an anti-phase spectrum.

To choose between the two zeroth-order phase values, the DQF-COSY was processed twice, once with each value. Following this procedure, it was clear that the proper phasing command was ph(−184,−21972,0), where the parameters in the brackets are the zeroth-order phase correction, first-order phase correction, and the pivot point at which no phase was changed while the first-order correction was carried out.

The macro used to process the DQF-COSY NMR data set was as follows (commands for D2-processing will be explained in the next subsection):

```
# (datadqDQF-COSY.2Dproc)
MD1  = 1024    # Matrix D1
MD2  = 1024    # Matrix D2

lpCom1    = "n"                          # D1 linear prediction
lpCom2    = "lp (100,412,20,288, r,t)" # D2 linear prediction

fltCom1   = "esb (1024,70,2,-100)"     # D1 filtering
fltCom2   = "esb (800,  0,2,-100)"     # D2 filtering
```

```
phCom1     = "ph (-184,-21972,0)"  # D1 phasing
phCom2     = "ph (-88, 0, 0)"      # D2 phasing

bcCom1     = "n"      # D1 baseline correction
bcCom2     = "n"      # D2 baseline correction

DimNoToTrans = b    # Dimension to transform
ExprmtType   = t    # Experiment type

sbs(4, 340,480, 44,140, 900,930, 960,1020) # Set baseline segments
```

In the sbs command, four baseline segments were set for the calculation of 1D base-levels and baseline correction. They were obtained by zooming into different regions of the processed Row 100. The first region, [340,480], was used to work out 1D base levels.

Figure 8.9 shows the intensity plot of the 2D intermediate data after processing in D1 with the above parameters and zero-filling to 1024 points for each column.

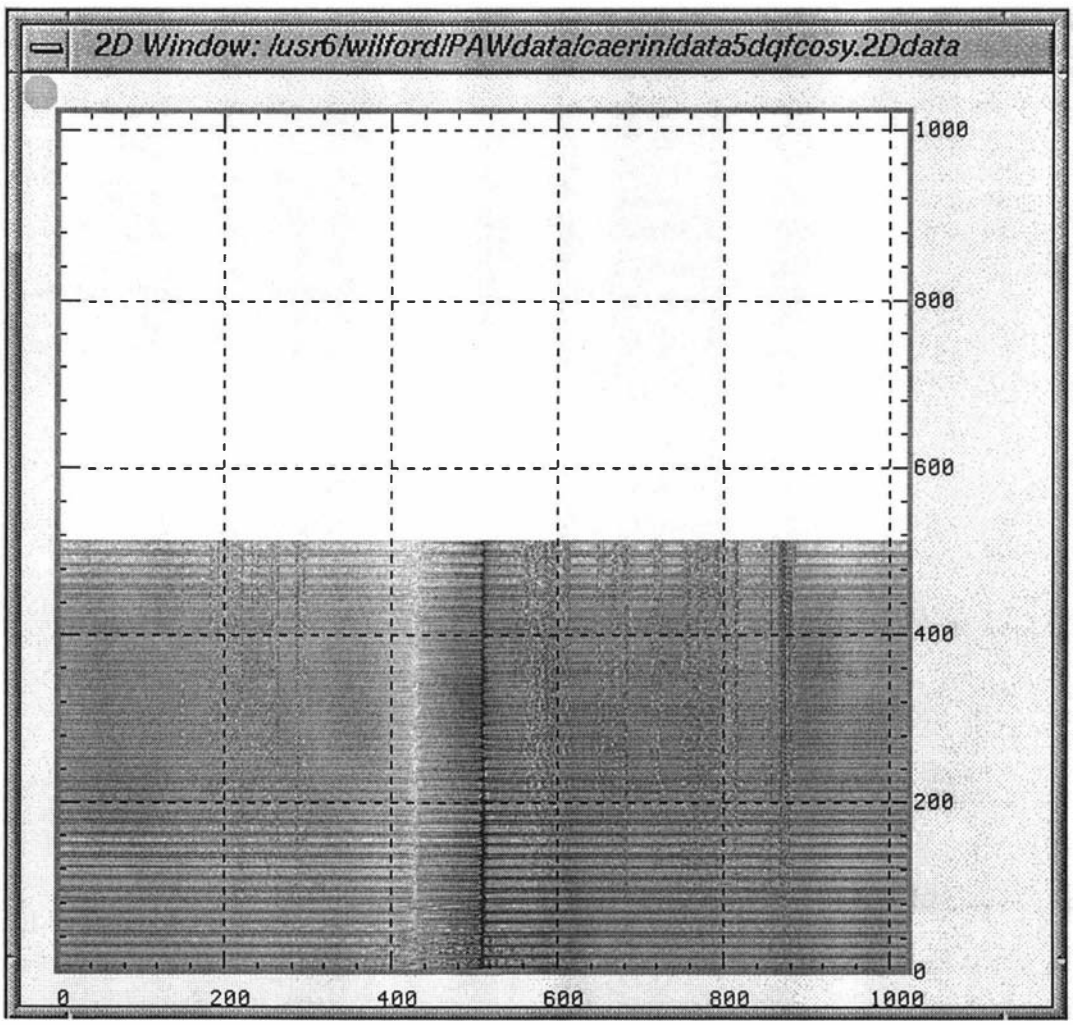


Figure 8.9: The 2D the Caerin 4.1 DQF-COSY intermediate data after processing in D1. The white region at the top is the zero-filled data. The white column at around 430 is where the data in many rows change their sign.

A quick check with a number of rows shows very good phasing and baseline in each row. This is also characterised by the even pattern in various regions of the data set.

8.2.2.2 D2-processing

As is indicated in the macro, linear prediction was applied to each column during D2-processing to narrow the peaks, thereby improving the resolution of the spectrum. The operation used 412 points starting from point 100 to predict 288 points of the real part with a 20^{th}-order prediction model. This resulted in a total of 800 significant data points for each column of data. The reason for predicting only 288 points and not the entire zero-filled region was to avoid some possible noise caused by incorrect prediction.

The enhanced sine-bell filter used in D2 was esb(800,0,2,–100). For each column, the filter applies a full suppression to the initial part in order to minimise the influence of the overflow data in the first 89 rows. To increase the resolution further, an exponential enhancement of –100 units was applied to increase the weight in the central regions, thereby slowing down the decay rate. The filter also caused the data to gradually reduce to zero, thereby reducing the noise and avoiding sinc-wiggles in the frequency-domain spectrum.

To obtain a set of phasing parameters that would be suitable for all columns, Column 38 and 885 were added together to form a combined FID. These were the two most significant columns near the two ends of the 2D intermediate data set. The phasing command determined from phasing the combined FID was ph(–88,0,0).

Figure 8.10 shows the combined FID and the processed results for the FID at various stages.

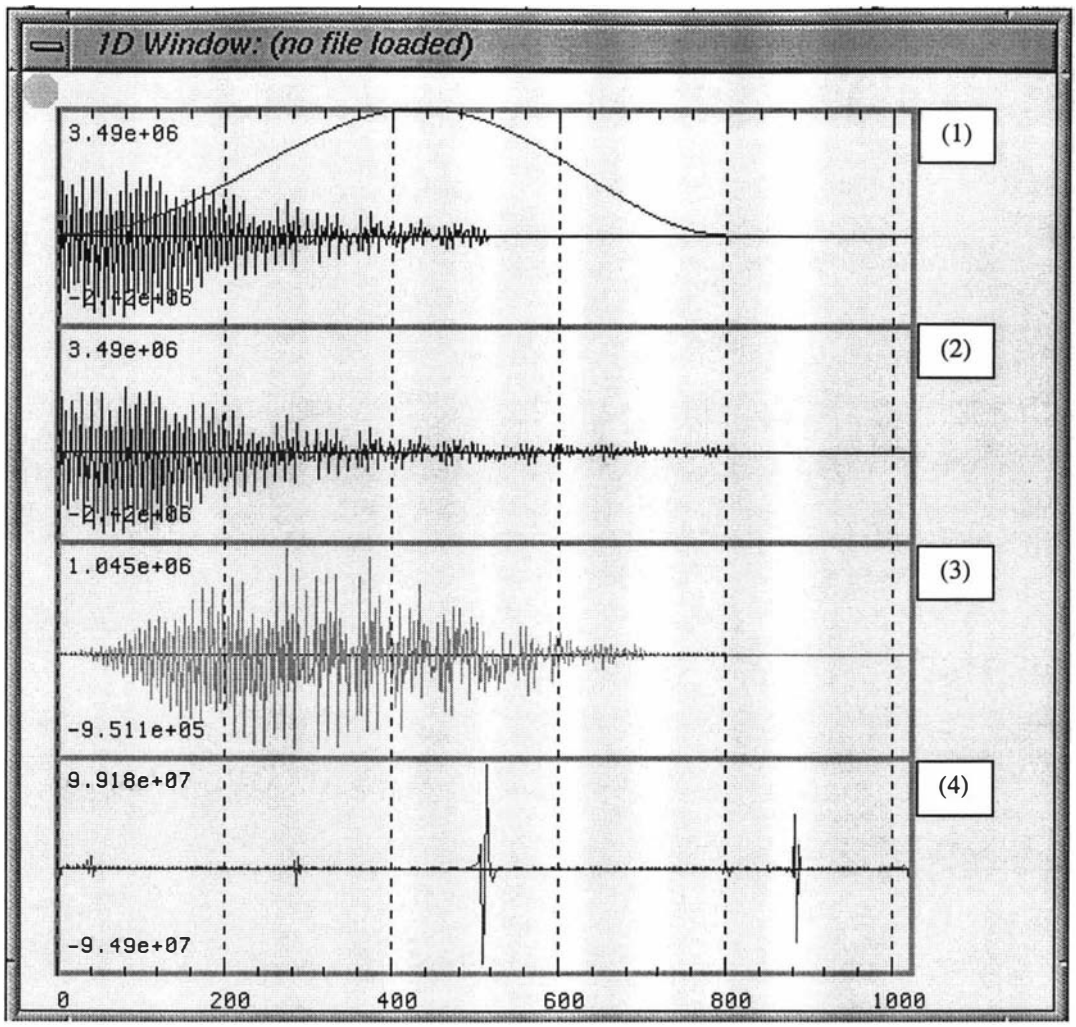


Figure 8.10: The processed results of a combined column of the intermediate data set. The first plot is the raw data and the filter to be applied. The second plot is the data after applying the linear-prediction. The third plot is the data after applying the filter. The last plot is the result after applying the real Fourier transform and performing the phase correction.

A total of five bad linear prediction results were reported. Four of the columns causing the error contain only noise. The failures in prediction for noise were expected because noise would not have a pattern that could be predicted. PAW automatically retained the original data if the predicted data was poor. It would be a problem if the original data were seriously truncated at point 512, because the filter is set to end at point 800 for the predicted data set. However, a careful check made by zooming into a few small regions did not show any column that was particularly bad due to the failure.

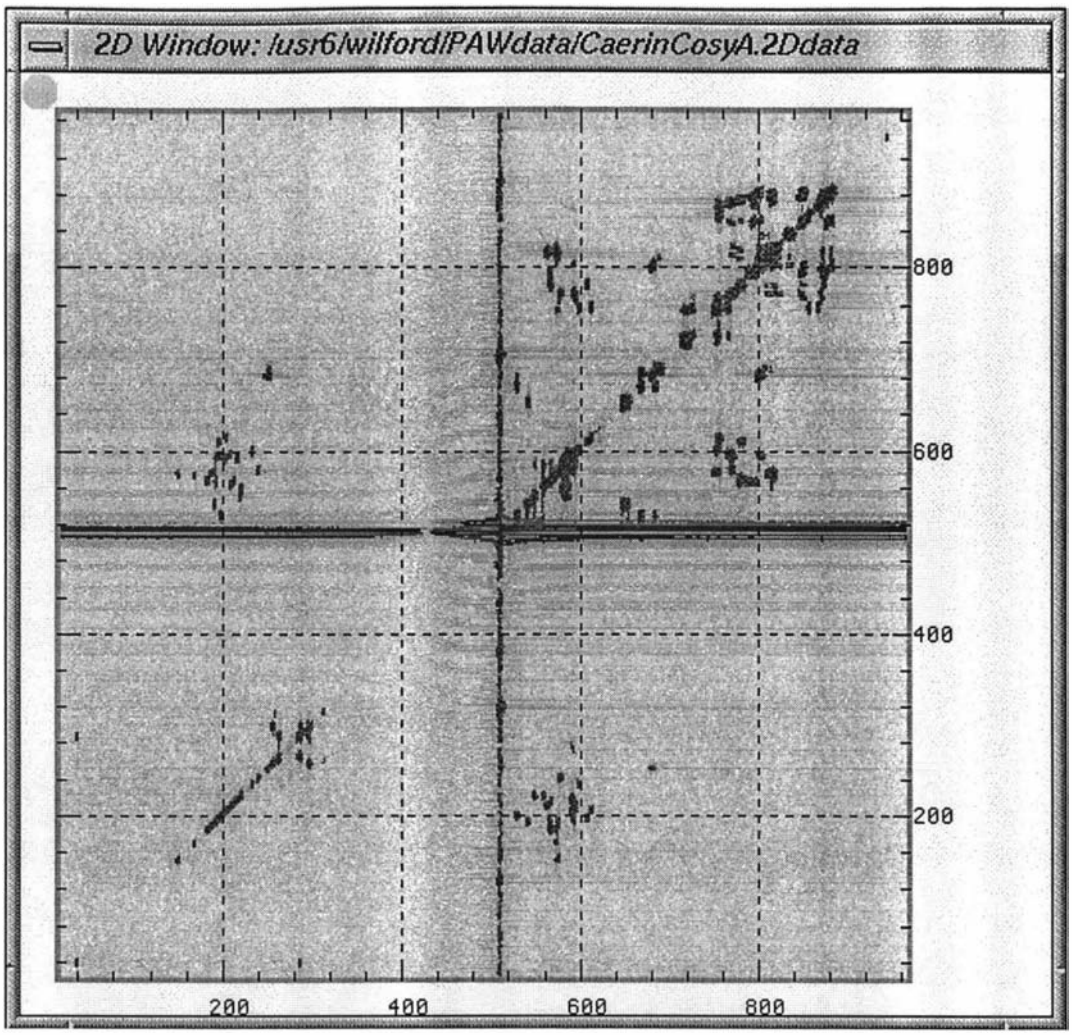


Figure 8.11: The Caerin 4.1 DQF-COSY spectrum before baseline correction in D2.

Surprisingly, the water peak in each row is quite narrow and does not impact on the adjacent peaks, except for those very close to the intense horizontal streaks in the middle. This indicates a proper phasing-operation during processing in D1.

The horizontal streaks in the 2D spectrum indicate some baseline problem due to the strong water signal. This was corrected with the following baseline correction macro:

```
# (CorrectBaseline.mcr)
D1    = 1024              # Dimension 2
D2    = 1024              # Dimension 1
D1Struct  = r             # D1 structure
D2Struct  = r             # D2 structure
DrawOn    = F             # Do not draw
sbs (4, 340,480, 44,140, 900,936, 960,1024) #Set baseline segments
for (%I=0, 1023, 1)       # Start the for-loop
  ldr (%I, r)             # Load row %I as real data
  bc (s, 5)               # Correct baseline
  wrr (%I, r)             # Save row %I as real data
next                      # Continue the for-loop
DrawOn = T                # Draw now
dr1                       # Draw the 1D plot
dr2                       # Draw the 2d plot
```

Figure 8.12 shows the DQF-COSY spectrum after baseline correction. A comparison with the previous spectrum shows that the correction was successful.

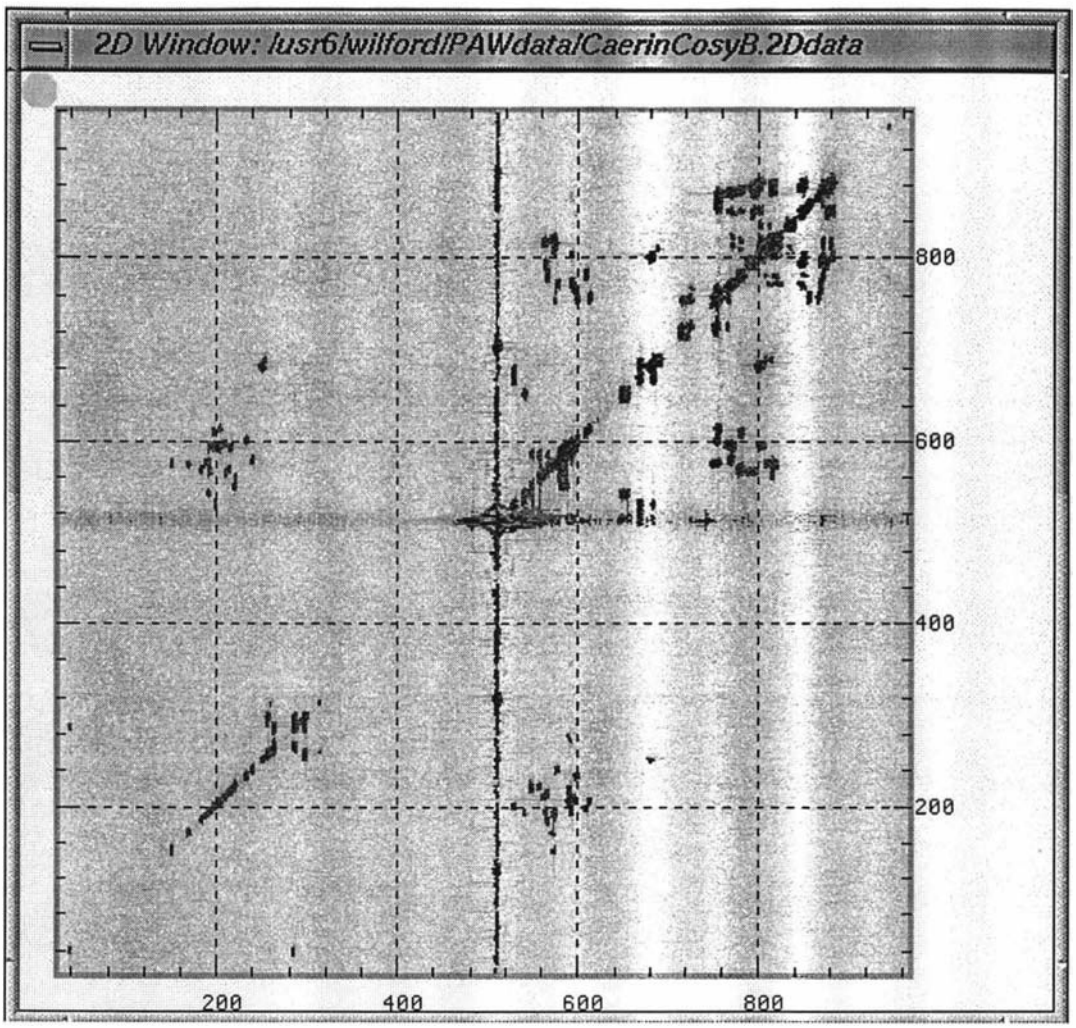


Figure 8.12: The Caerin 4.1 DQF-COSY spectrum after baseline correction.

Figures 8.13 to 8.16 show all the significant regions in the processed DQF-COSY spectrum.

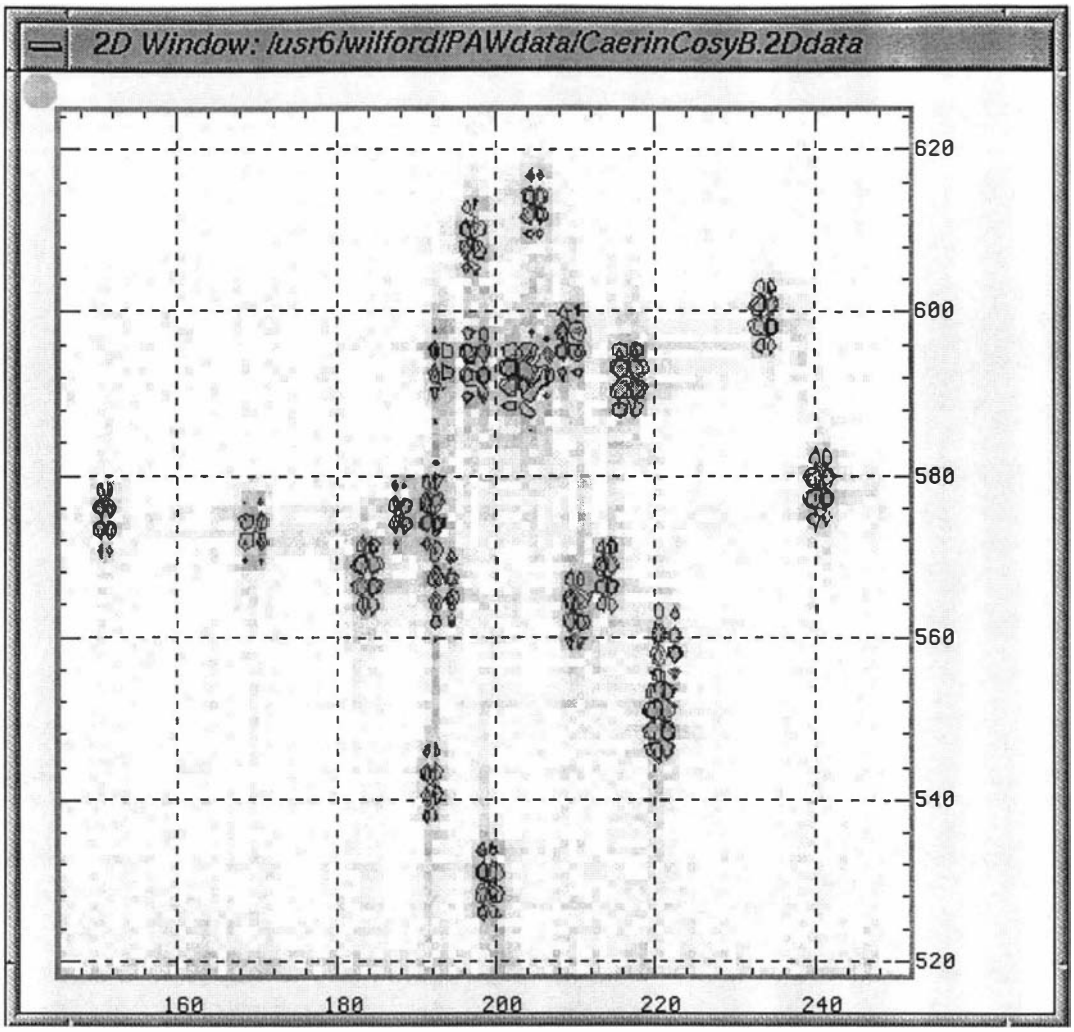


Figure 8.13: The upper-left (fingerprint) region of the processed Caerin 4.1 DQF-COSY spectrum.

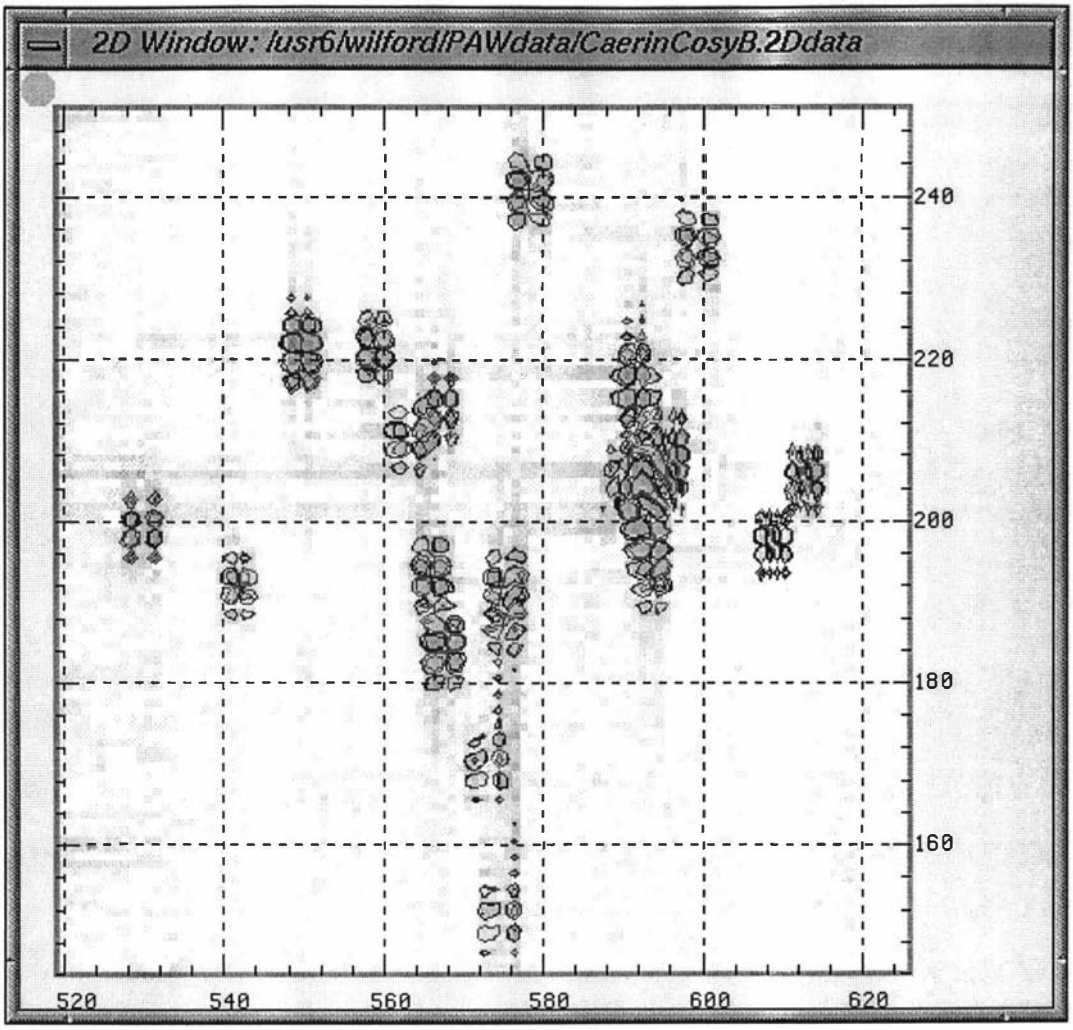


Figure 8.14: The lower-right (fingerprint) region of the processed Caerin 4.1 DQF-COSY spectrum.

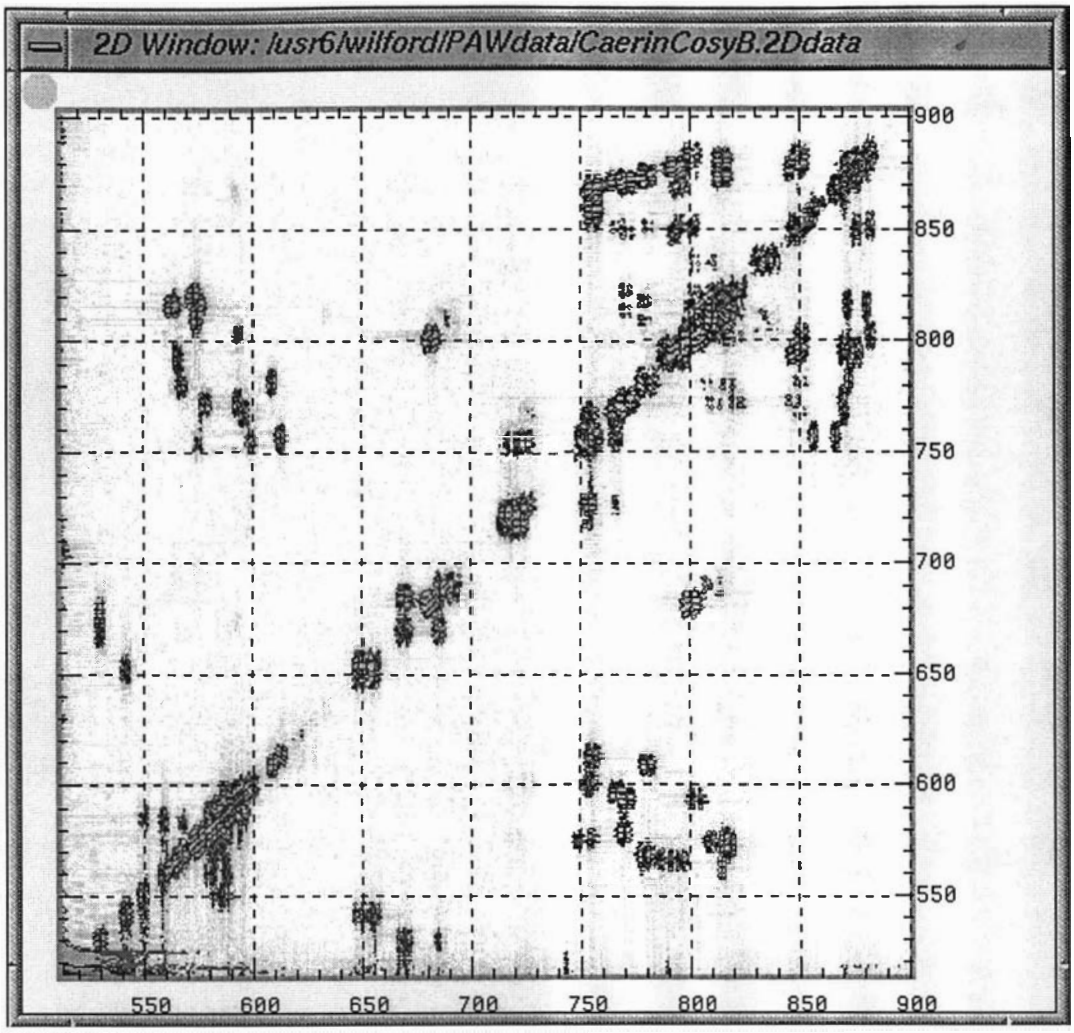


Figure 8.15: The upper-right (aliphatic) region of the processed Caerin 4.1 DQF-COSY spectrum.

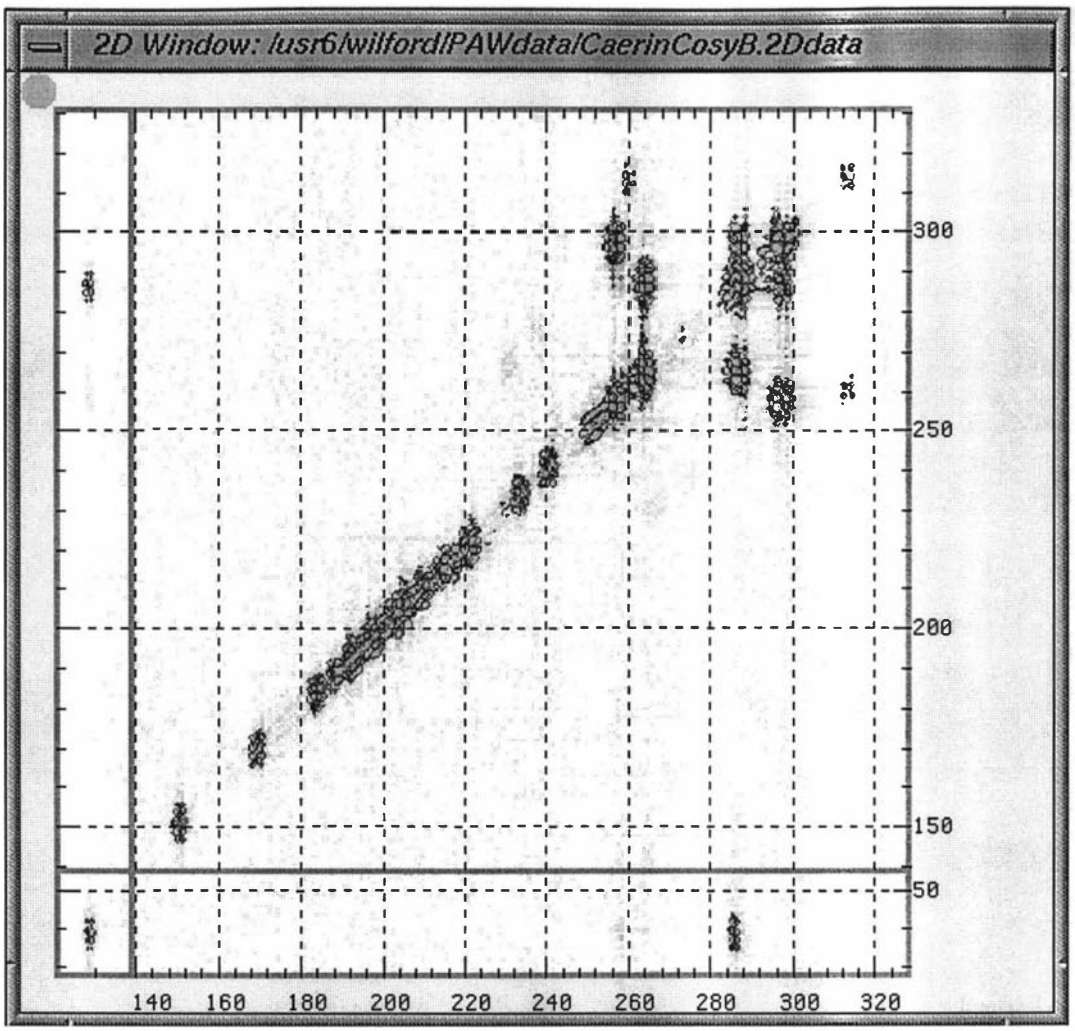


Figure 8.16: The lower-left (aromatic) region of the processed Caerin 4.1 DQF-COSY spectrum.

From the four figures, it can be seen that all peaks are correctly phased. The figures also show that the resolution in the (horizontal) D1-direction is much better than that in the D2-direction. This is because the filters ended at point 1024 for the complex FIDs in D1, and 800 for the real-only FIDs in D2, as indicated in the macro. In terms of the resolution that can be achieved, a 1D NMR spectrum obtained from transforming a 512 point real-only data set is equivalent to that obtained by transforming a 256 point complex data set.

Many attempts to further increase the resolution in D2 have failed because of this problem. In theory, if linear prediction could be successfully applied in D2 to properly predict the extended trailing part, then, the peak-shape in both dimensions would be the same. This, unfortunately, would introduce too much noise into the 2D DQF-COSY spectrum because there were not sufficient data points for the linear prediction routine to precisely predict so many points.

Applying linear prediction to repair the receiver-overflow time-domain data in D1 has also been fruitless. This was because predicting around 150 initial points for each FID induced a variety of phase shifts, and hence made it impossible to phase the resulting

2D spectrum with a general macro. Nevertheless, the DQF-COSY spectrum finally obtained was sufficient to be used as a reference for NMR spin-system identification.

8.2.3 Conventional processing operations and results

Conventional processing operations were also applied in order to assess the differences between the conventional and the high-resolution processing methods.

To get rid of the large first-order phase shifts, the first 62 points in each row were simply deleted using PAW's left-shift command. The operation resulted in a normal-looking 2D data set as shown in Figure 8.17.

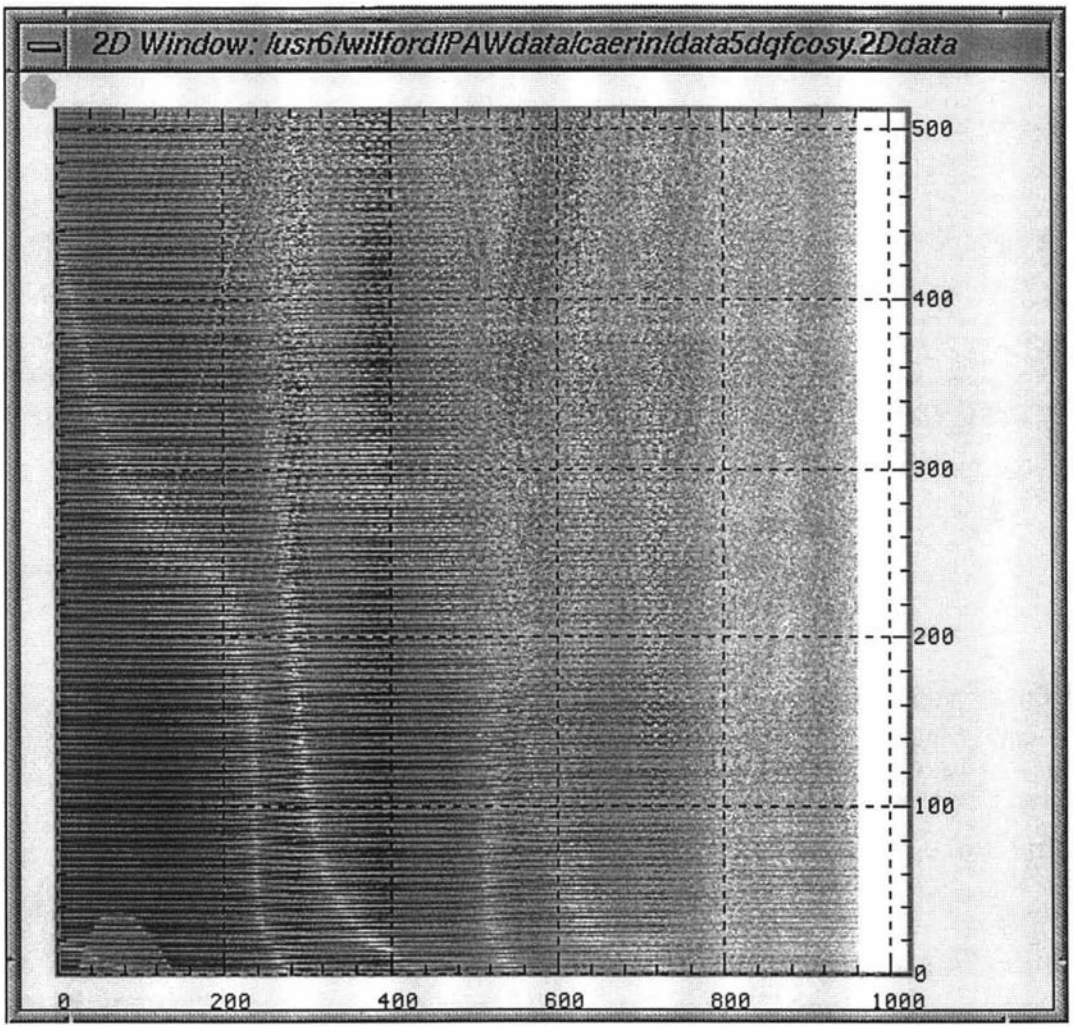


Figure 8.17: A whole view of the Caerin 4.1 DQF-COSY after left-shifting 62 points for each row.

8.2.3.1 D1-processing

Like the treatment in the last section, Row 100 was used to obtain the processing parameters in D1. However, this time the row was processed with the conventional cosine filter and normal phase-correction parameters. In PAW's commands, they were esb(962,0,1,0) and ph(−10,−200,510).

Figure 8.18 shows the raw data from Row 100 and the processed results.

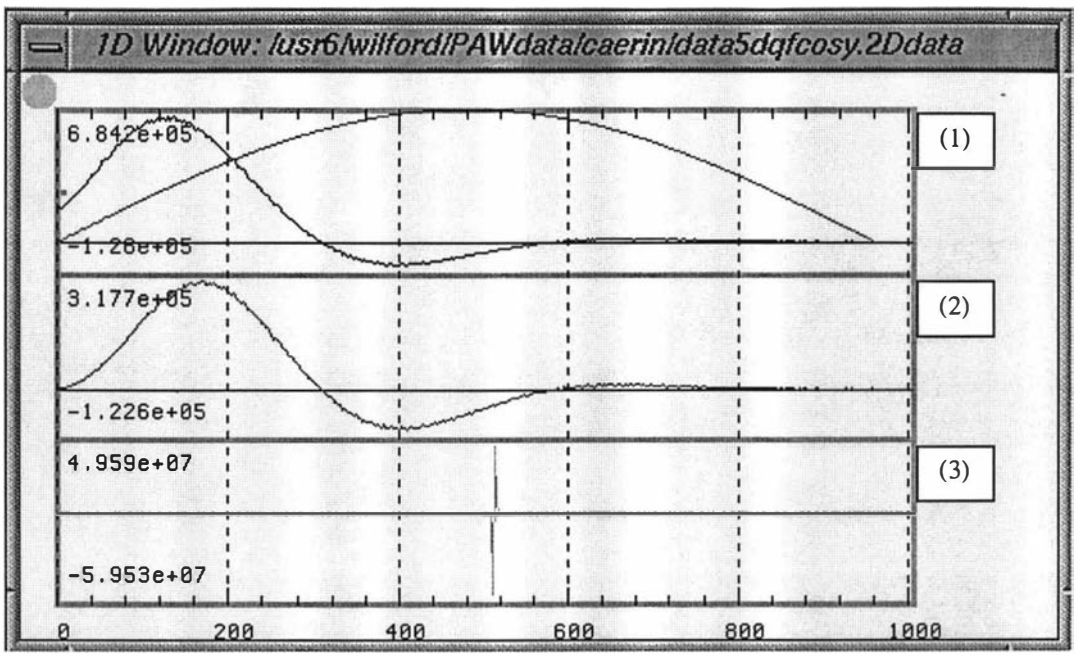


Figure 8.18: Row 100 of the Caerin 4.1 DQF-COSY and the processed results. The first plot is the raw data and the filter to be applied. The second plot is the data after applying the filter. The last plot is the spectrum after FFT and phase correction.

The macro used for the conventional 2D processing of the DQF-COSY NMR data set was as follows:

```
# (CrnCosyShifted.2Dproc)
MD1   = 1024          # Matrix D1
MD2   = 1024          # Matrix D2

lpCom1     = "n"      # D1 linear prediction
lpCom2     = "n"      # D2 linear prediction

fltCom1    = "esb (962, 0, 1, 0)"       # D1 filtering
fltCom2    = "esb (512, 0, 1, 0)"       # D2 filtering


phCom1     = "ph (-10, -200, 510)"      # D1 phasing
phCom2     = "ph ( 88, -6,   512)"      # D2 phasing

bcCom1     = "n"      # D1 baseline correction
bcCom2     = "n"      # D2 baseline correction

DimNoToTrans = b    # Dimension to transform
ExprmtType   = t    # Experiment type

# Set baseline segments
sbs(4, 340,480, 44,140, 900,930, 960,1020)
```

Figure 8.19 shows the 2D intermediate data set after processing in D1 with the above parameters.

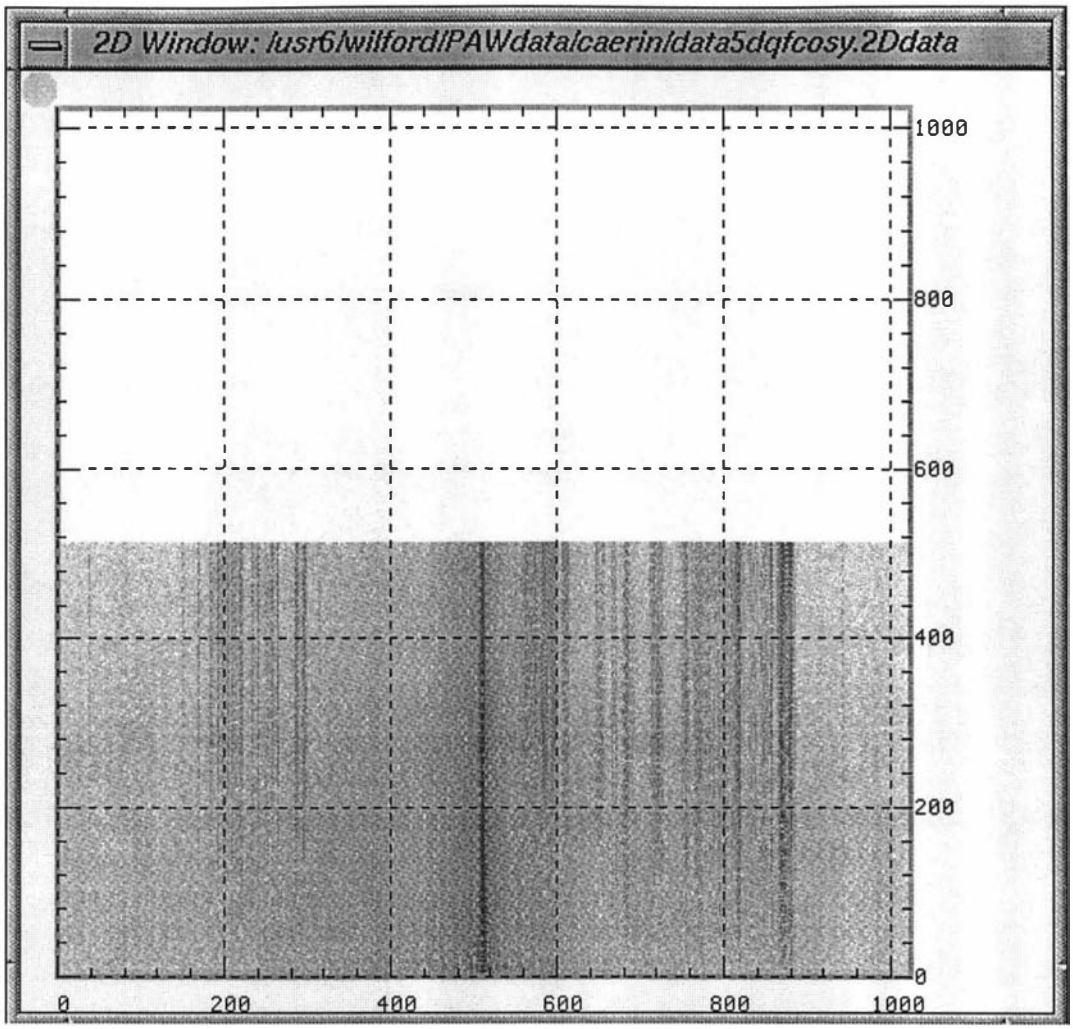


Figure 8.19: A whole view of the Caerin 4.1 DQF-COSY intermediate data after D1-processing with a conventional cosine filter.

8.2.3.2 D2-processing

The D2-processing parameters were obtained by processing a combined FID formed by adding columns 37 and 885 together, which again were two significant columns near the ends of the 2D intermediate data set. As indicated in the macro, the enhanced sine-bell filter used in D2 was esb(512,0,1,0), and the phase correction command was ph(88,–6,512).

Figure 8.20 shows the combined FID and the processed results for the data at various stages.

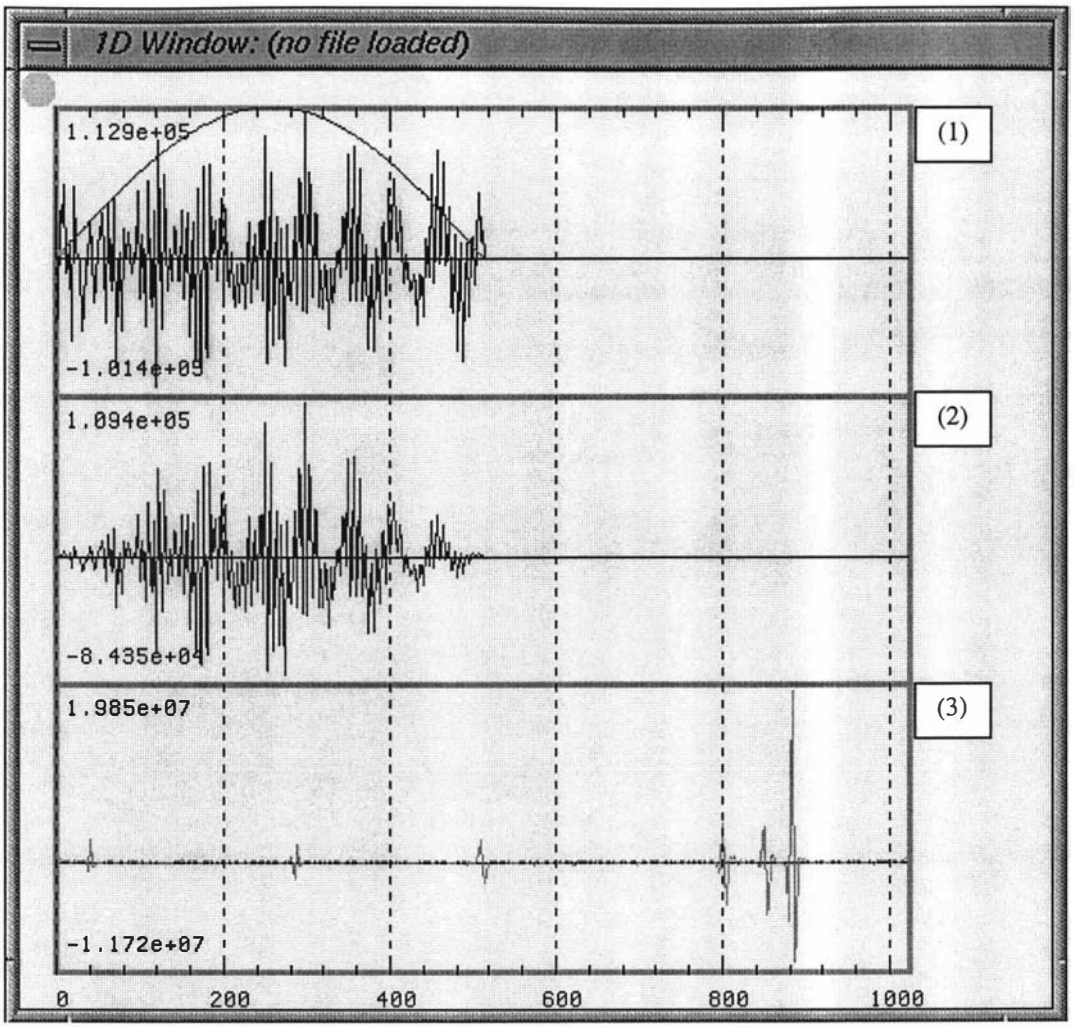


Figure 8.20: The processed results of a combined column of the intermediate data set. The first plot is the combined data and the applied filter. The second plot is the data after applying the filter. The last plot is the result after real Fourier transformation and phase correction.

Figure 8.21 shows the 2D spectrum after processing in D2 with the above parameters.

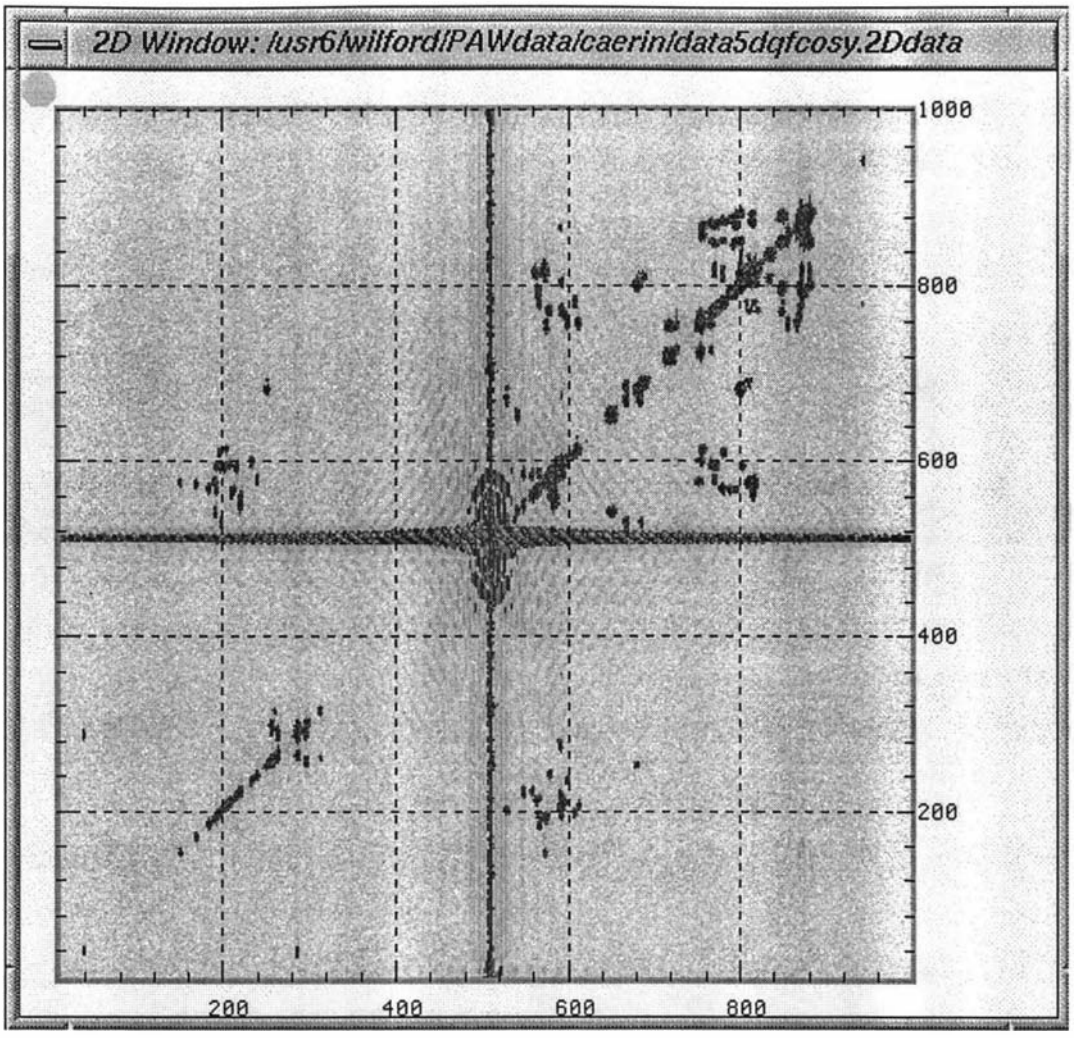


Figure 8.21: A full view of the Caerin 4.1 DQF-COSY processed with a conventional sine-bell filter. The water signals appear to be particularly bad at the centre of the spectrum. This was because the filter applied during D2-processing did not properly suppress the damaged data at the beginning of each column.

Figure 8.22 shows the contour plots of the upper-left (fingerprint) region of Figure 8.21.

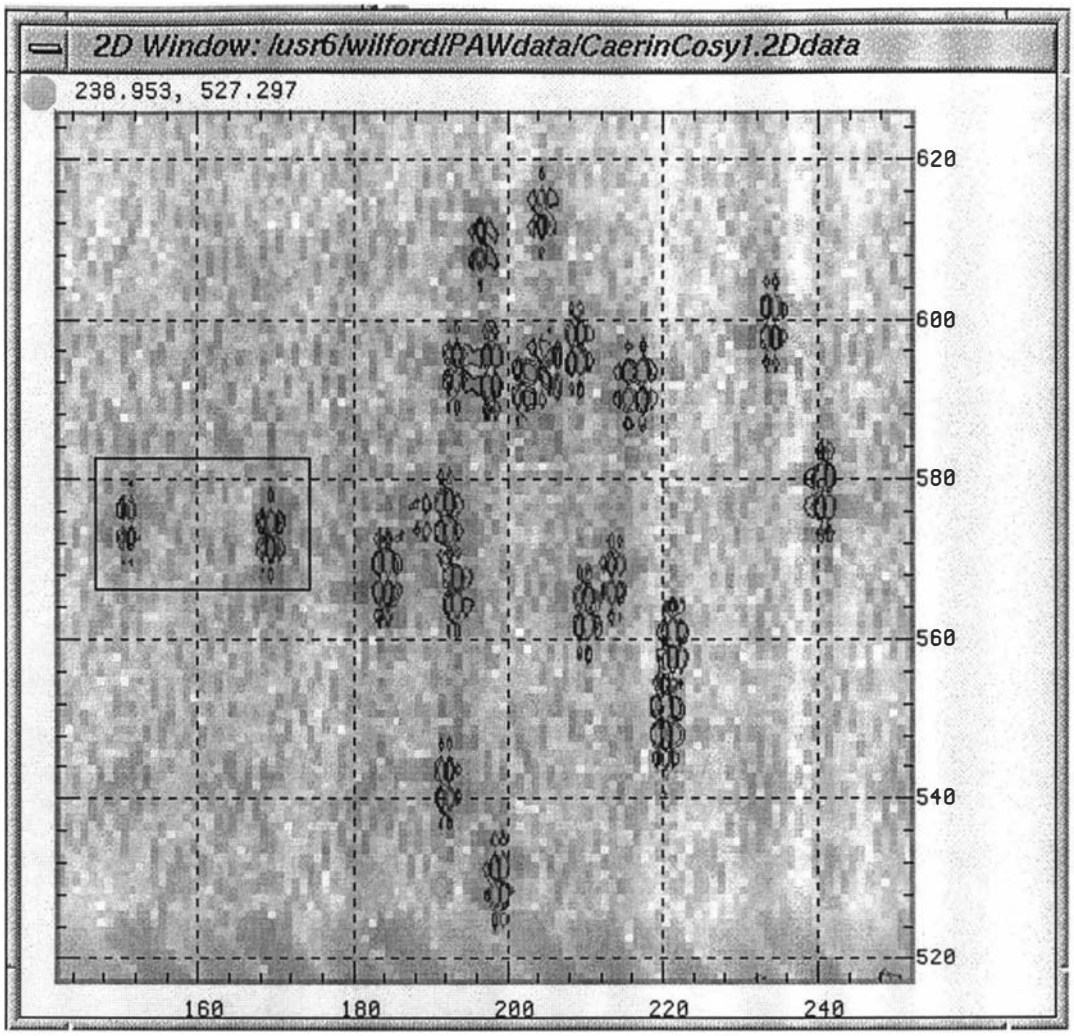


Figure 8.22: The upper-left (fingerprint) region of the Caerin 4.1 DQF-COSY spectrum processed with a conventional method.

As expected, the conventional processing method produced peaks surrounded by undesirable satellite peaks in both dimensions. The biggest problem encountered during processing the Caerin 4.1 DQF-COSY data with the conventional method, however, was the difficulty in phasing. None of the regions could be properly phased in D1 because the phase shifts were different in each row after a left-shift operation. This can be seen from the peaks inside the rectangle in Figure 8.22. Replacing the parameters in any processing operation could not solve this problem, including those for the water-signal suppression, apodisation, linear prediction, phase correction, and baseline correction. The damaged data also has nothing to do with the phase errors, because applying any higher-degree sine-bell filter to suppress them cannot solve the problem. Clearly, the initial part of each row did contain important information that was lost during the left-shift process.

8.3 Processing the Caerin 4.1 TOCSY070 NMR data set

8.3.1 The raw data

The intensity plot of the Caerin 4.1 TOCSY070 time-domain data is shown in Figure 8.23.

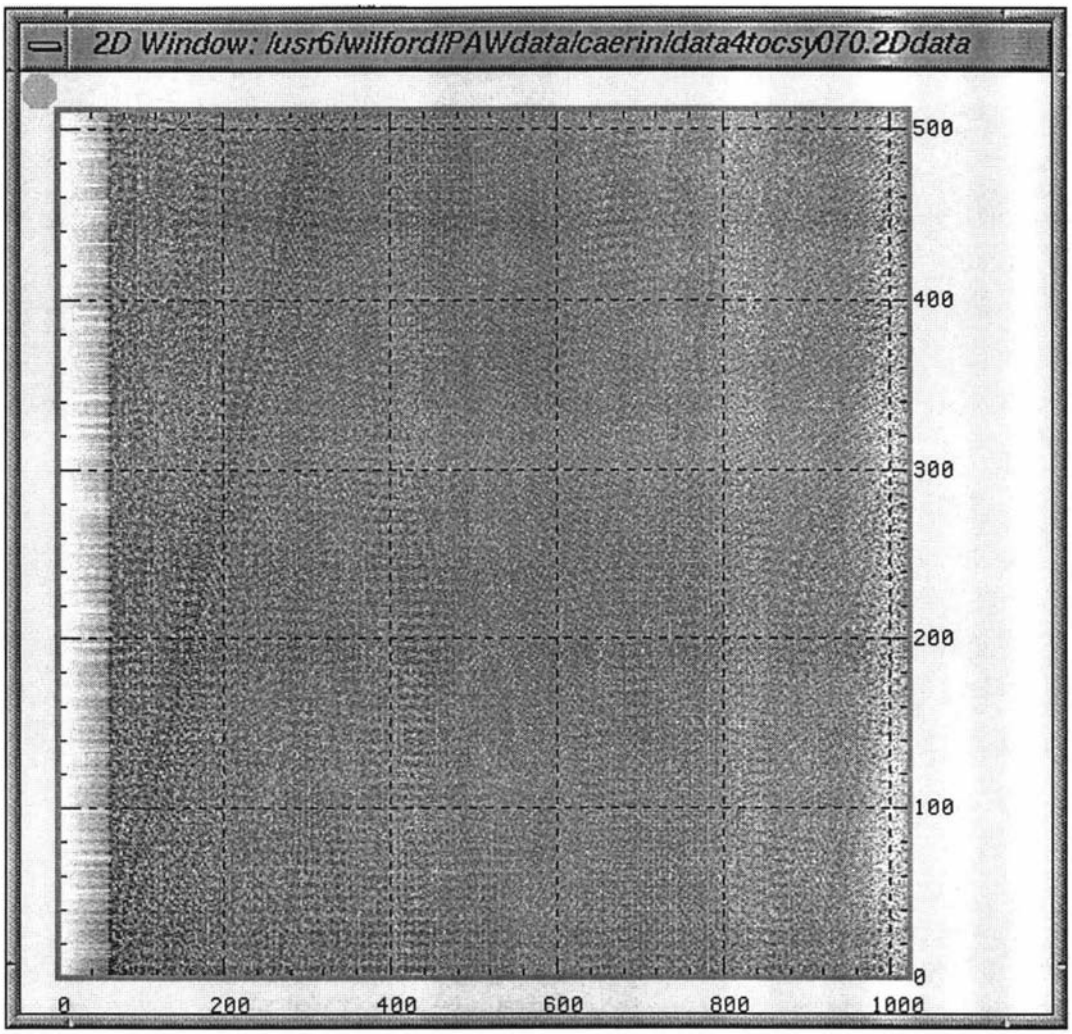


Figure 8.23: A whole view of the Caerin 4.1 TOCSY070 raw data. The evenly distributed pattern indicates no receiver-overflow problem in the data collection process. The white part on the left is due to the digitally filtered over-sampling technique.

The water signals are much less intense than in the DQF-COSY. In fact, they are not even visible in the 1D spectra drawn from a number of rows (Figure 8.24).

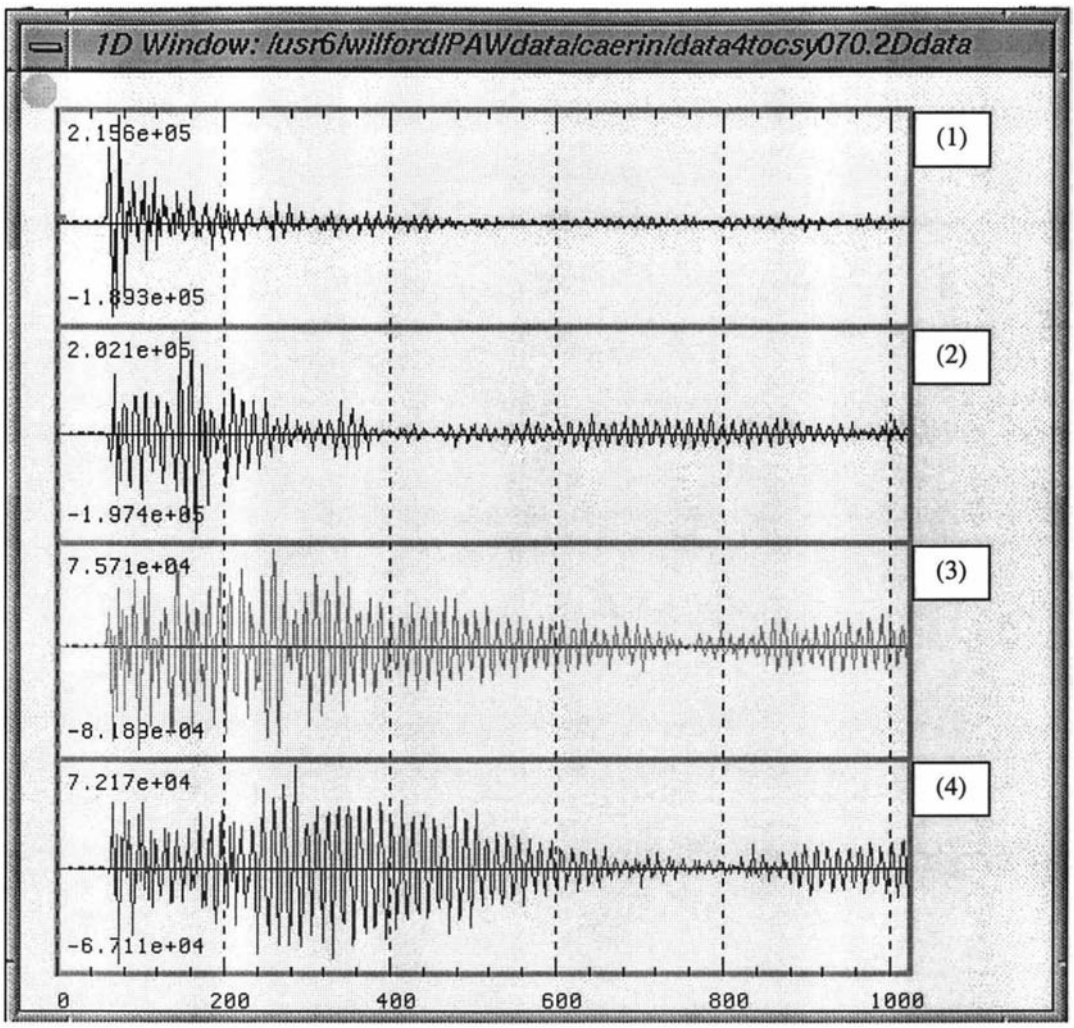


Figure 8.24: Row 0, 200, 400 and 511 of the Caerin 4.1 TOCSY070 raw data.

The significant intensities at the ends of plots 3 and 4 indicate that the acquisition time was again shorter than the average T_2 of the protons in the sample. Nevertheless, it looked appropriate from the first plot. In D2, however, it will be shown later that there was a significant truncation of each column.

Overall, the quality of the TOCSY070 data set appears to be quite good. However, it would have been better if 1024 FIDs had been collected with a slightly longer acquisition time.

8.3.2 High-resolution processing operations and results

8.3.2.1 D1-processing

Since this is a TOCSY070 NMR data set, the first row contains the most intense signal and hence can be used to determine the processing parameters in D1. Figure 8.25 shows the raw data from the first row and the processed results.

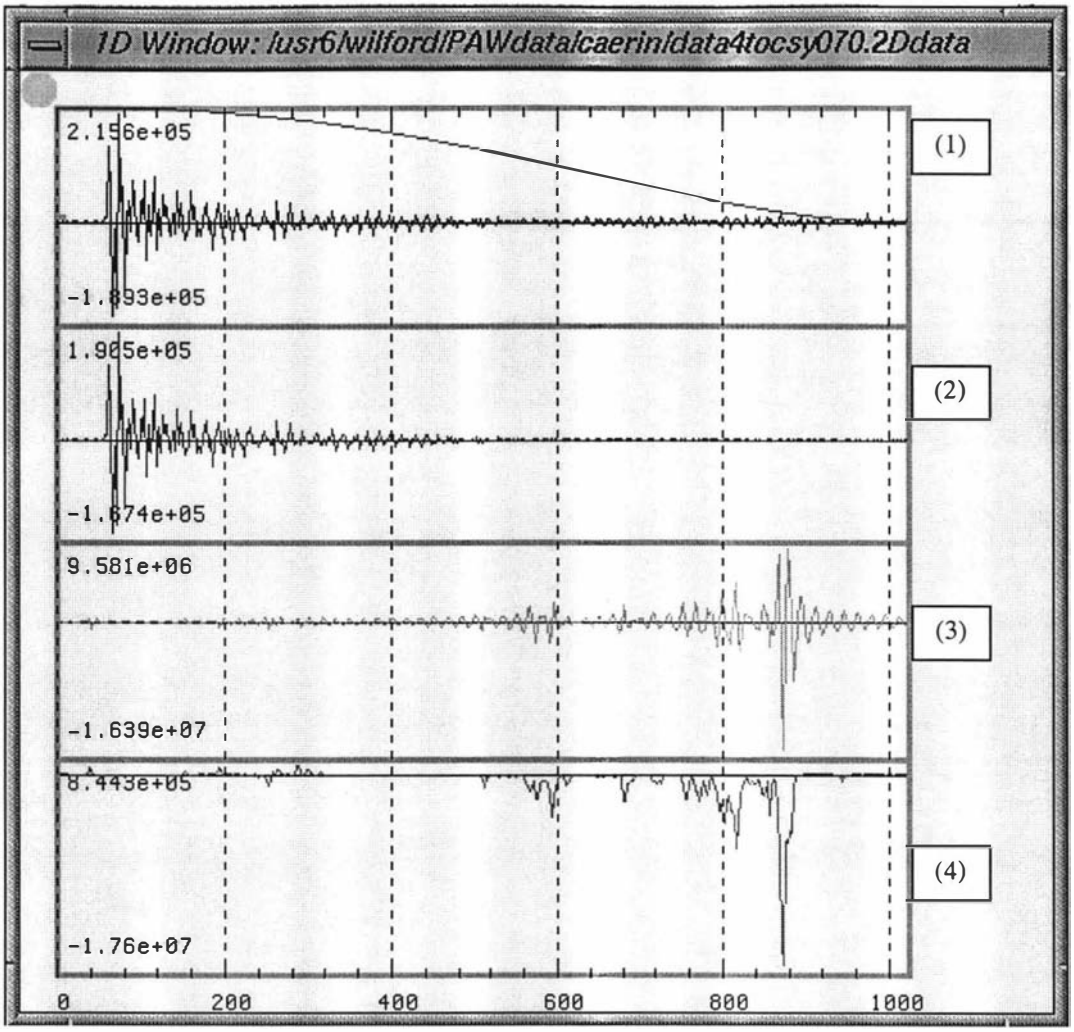


Figure 8.25: The processed results of the first row of the Caerin 4.1 TOCSY070 NMR data set. The first plot is the raw data and the applied filter. The second plot is the data after applying the filter. The third plot is the spectrum after FFT. The last plot is the spectrum after phase correction.

The enhanced sine-bell filter used in D1 was esb(1024,80,2, –100), as shown in Plot 1 of Figure 8.25.

Again, in the NMR spectrum, the baseline distortion due to the large first-order phase shift is significant (Plot 3, Figure 8.25). With the pivot set to 0 and the base region set to [340,480], PAW's auto-phase routine resulted in 10.8° for the zeroth- and –21982° for the first-order phase corrections. These parameters were adjusted manually to 16 and –21988 to make the peak at point 38 correctly phased. Applying the same operations to rows 8, 18, 28 and 38 all resulted in similar first-order phase correction.

The macro for processing the TOCSY070 data was as follows:

```
# (data4tocsy070.2Dproc)
MD1  = 1024    # Matrix D1
MD2  = 1024    # Matrix D2

lpCom1    = "n"                              # D1 linear prediction
lpCom2    = "lp (100,412,20,288, r,t )" # D2 linear prediction
```

```
fltCom1    = "esb (1024, 80, 2, -100)"  # D1 filtering
fltCom2    = "esb (800,  80, 2, 0)"     # D2 filtering

phCom1     = "ph (16, -21988, 0)"       # D1 phasing
phCom2     = "ph (64, 62, 0)"           # D2 phasing

bcCom1     = "n"      # D1 baseline correction
bcCom2     = "n"      # D2 baseline correction

DimNoToTrans = b    # Dimension to transform
ExprmtType   = t    # Experiment type

sbs(4, 340,480, 44,140, 900,930, 960,1020) # Set baseline segments
```

Figure 8.26 shows the intensity of the 2D data after processing in D1 with the above parameters.

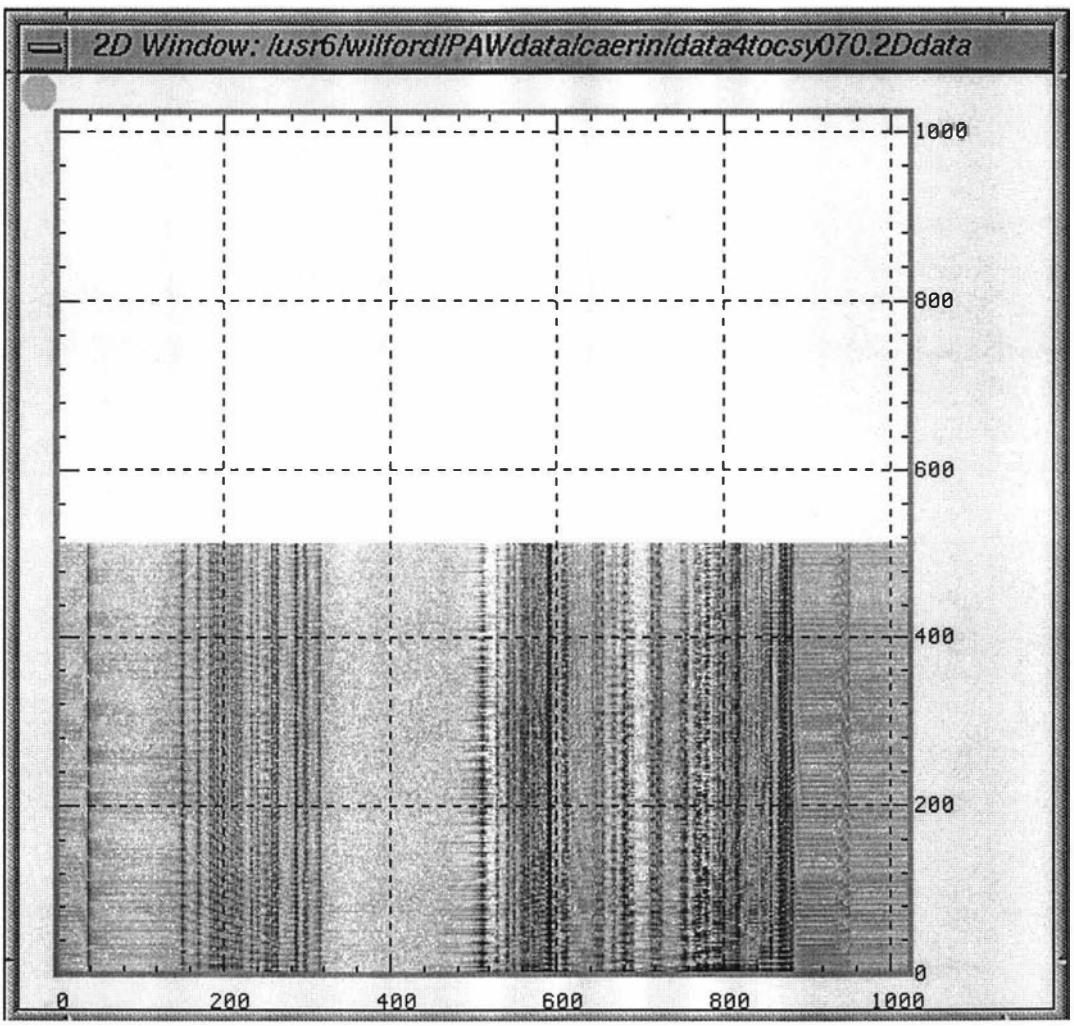


Figure 8.26: The Caerin 4.1 TOCSY070 intermediate data after processing in D1.

8.3.2.2 D2-processing

This proceeded in a similar fashion to that described in section 8.22.

Figure 8.27 shows the combined FID from columns 38 and 885, as well as the processed results for the FID at various stages.

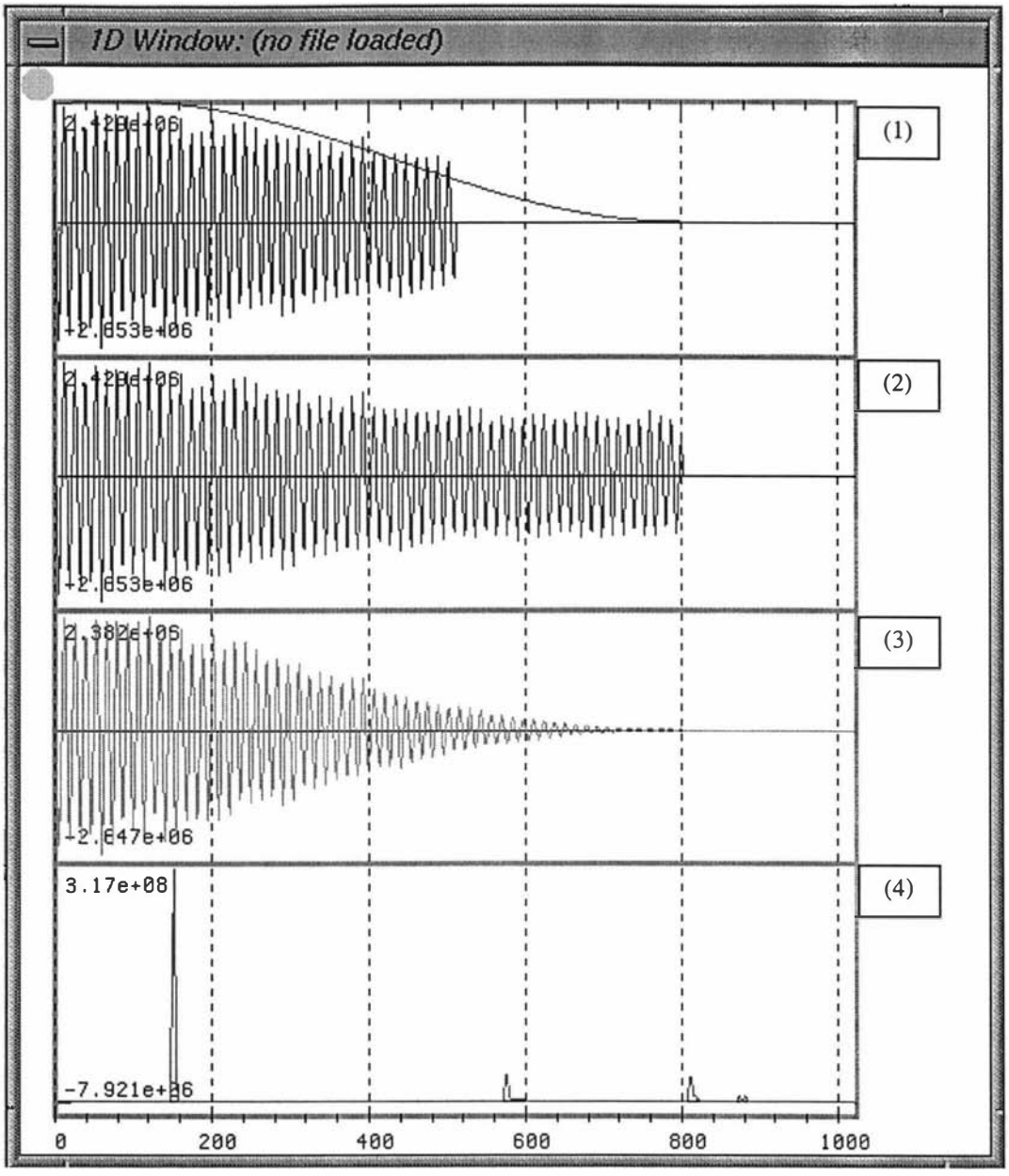


Figure 8.27: The processed results of a combined column of the intermediate data set. The first plot is the raw data and the filter to be applied. The second plot is the data after applying the linear-prediction. The third plot is the data after applying the filter. The last plot is the result after applying the real Fourier transform and the phase correction.

Figure 8.28 shows the 2D spectrum after processing in D2.

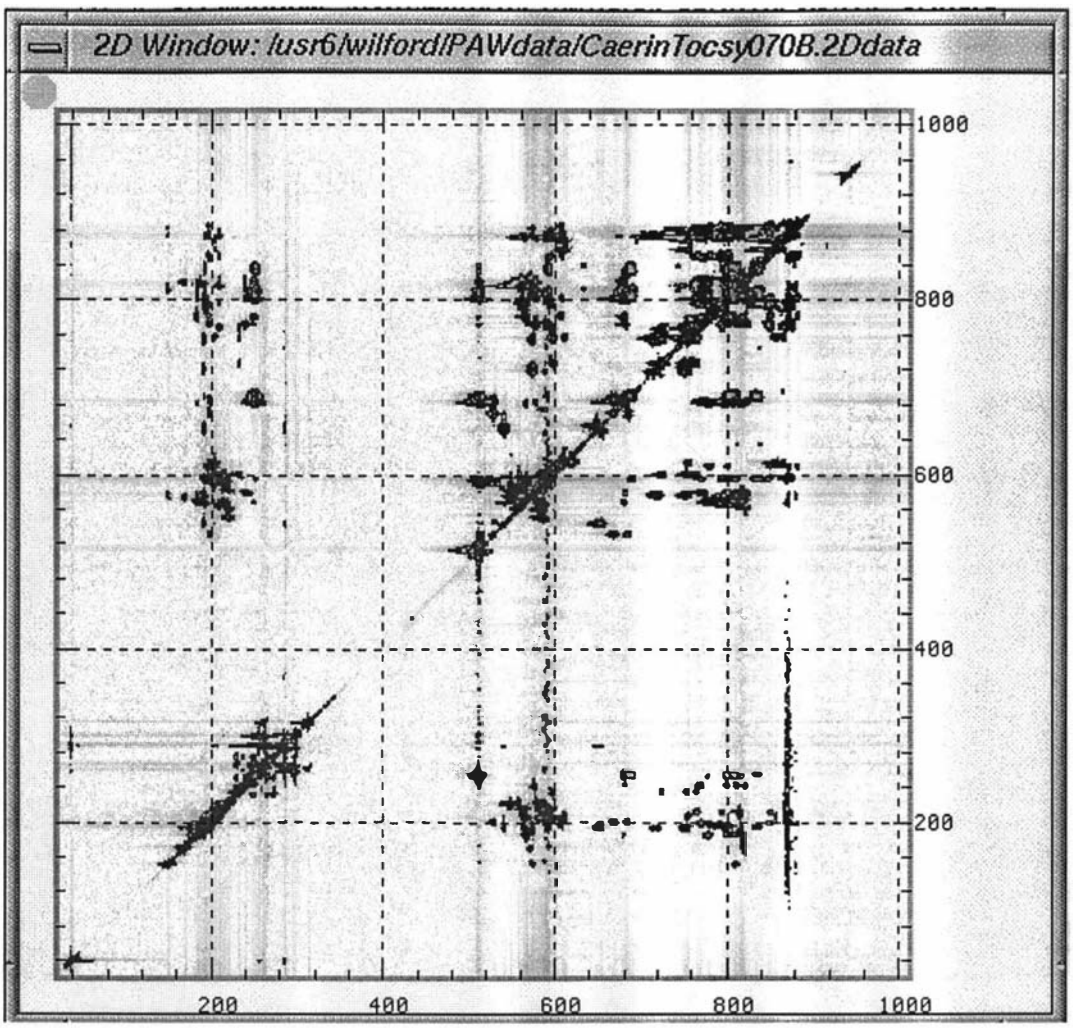


Figure 8.28: The Caerin 4.1 TOCSY070 spectrum after processing in D2.

A number of columns returned a linear prediction error due the lack of signal and so were ignored. The vertical streaks in the 2D spectrum indicate some baseline problem, which was corrected with the following baseline correction macro:

```
# (CorrectBaseline.mcr)
D1    = 1024              # Dimension 2
D2    = 1024              # Dimension 1
D1Struct  = r             # D1 structure
D2Struct  = r             # D2 structure
DrawOn    = F             # Not to draw
sbs (4, 340,480, 44,140, 900,936, 960,1024) #Set baseline segments
for (%I=0, 1023, 1)       # Start the for-loop
  ldc (%I, r)             # Load column %I as real data
  bc (s, 5)               # Correct baseline
  wrc (%I, r)             # Save column %I as real data
next                      # Continue the for-loop
for (%I=0, 1023, 1)       # Start the for-loop
  ldr (%I, r)             # Load row %I as real data
  bc (s, 5)               # Correct baseline
  wrr (%I, r)             # Save row %I as real data
next                      # Continue the for-loop
DrawOn = T                # Draw now
dr1                       # Draw the 1D plot
```

```
dr2                          # Draw the 2d plot
```

Figure 8.29 shows the TOCSY070 spectrum after baseline correction.

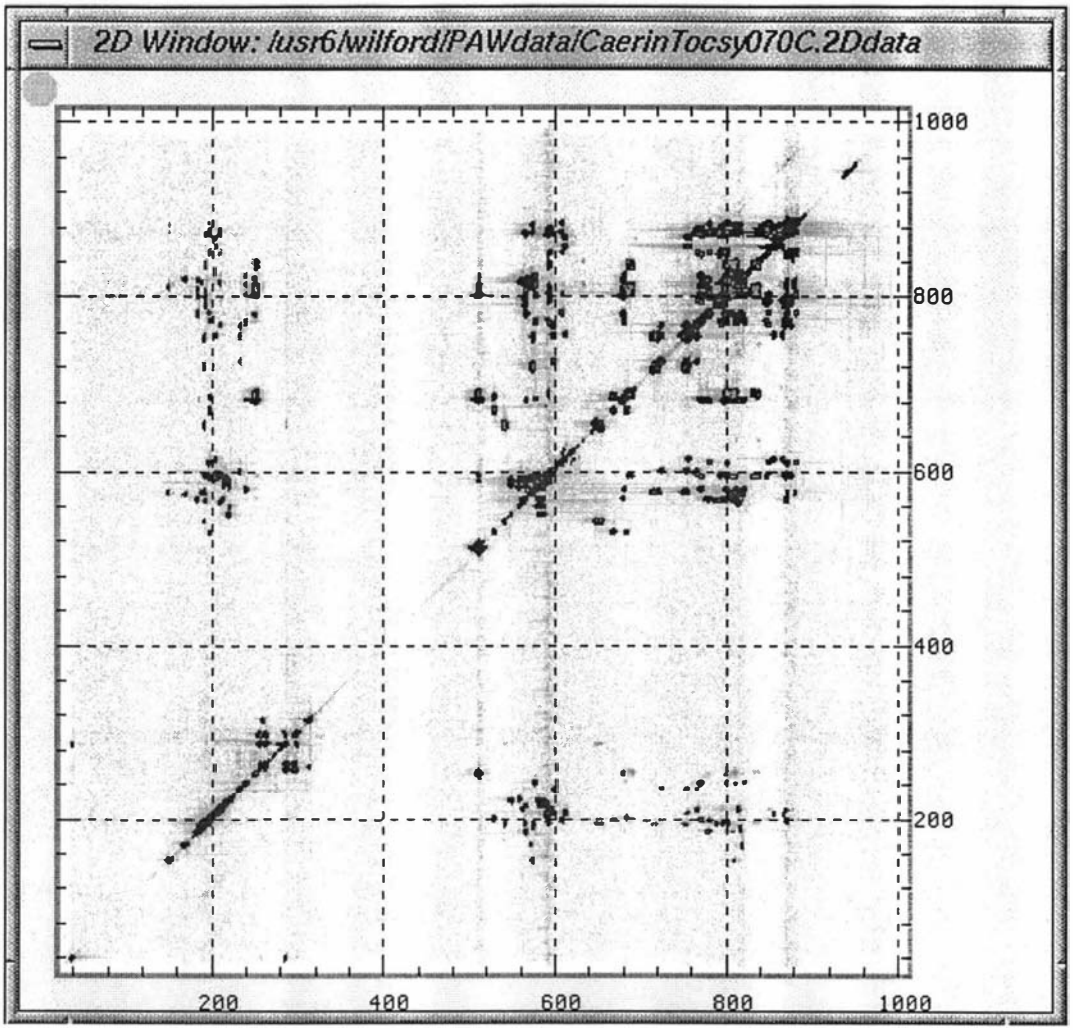


Figure 8.29: The Caerin 4.1 TOCSY070 spectrum after baseline correction. The white parts surrounding the spectrum are the zeros resulting from PAW's baseline-correction routine.

A comparison with the previous spectrum shows that the correction was largely successful.

Figures 8.30 to 8.33 show contour plots of all the significant regions of the processed TOCSY070 spectrum.

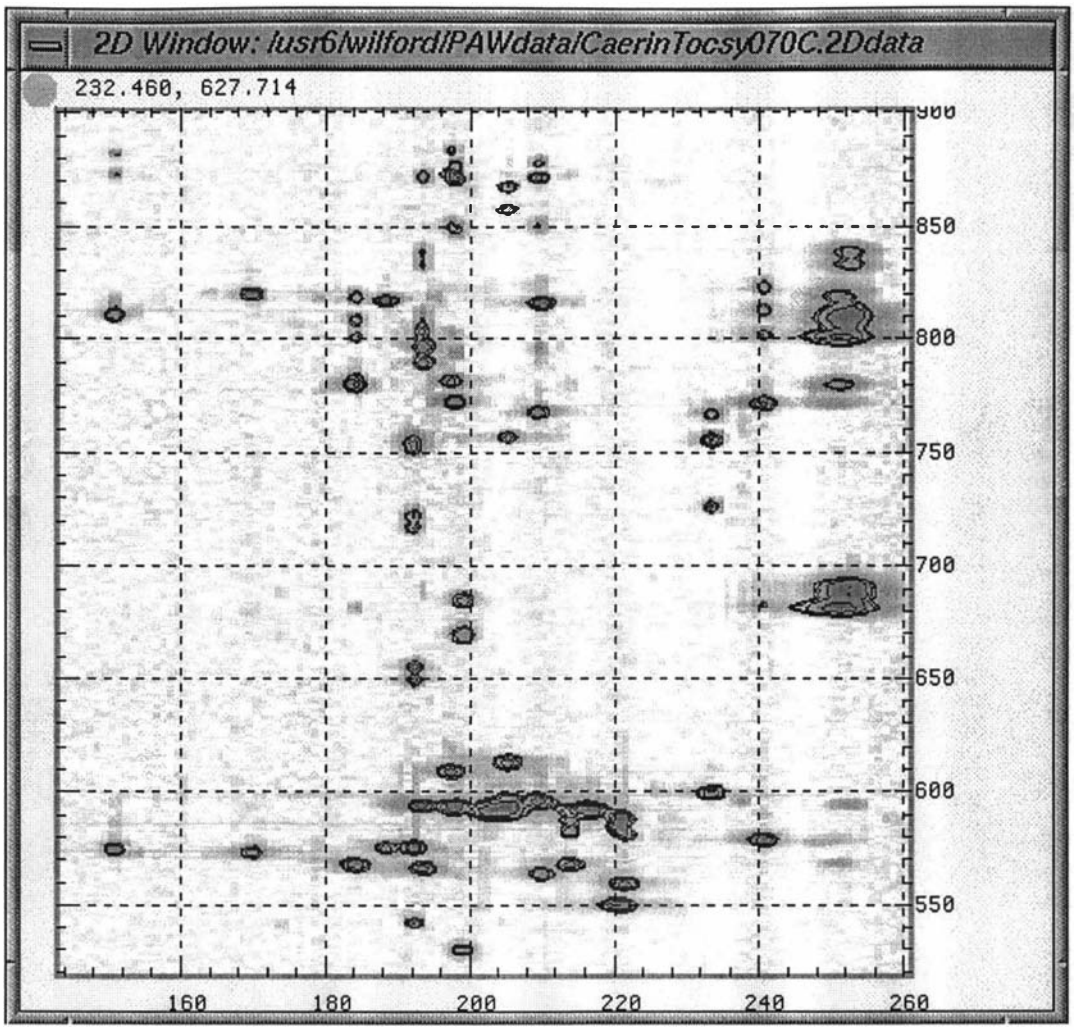


Figure 8.30: The upper-left (fingerprint) region of the processed Caerin 4.1 TOCSY070 spectrum.

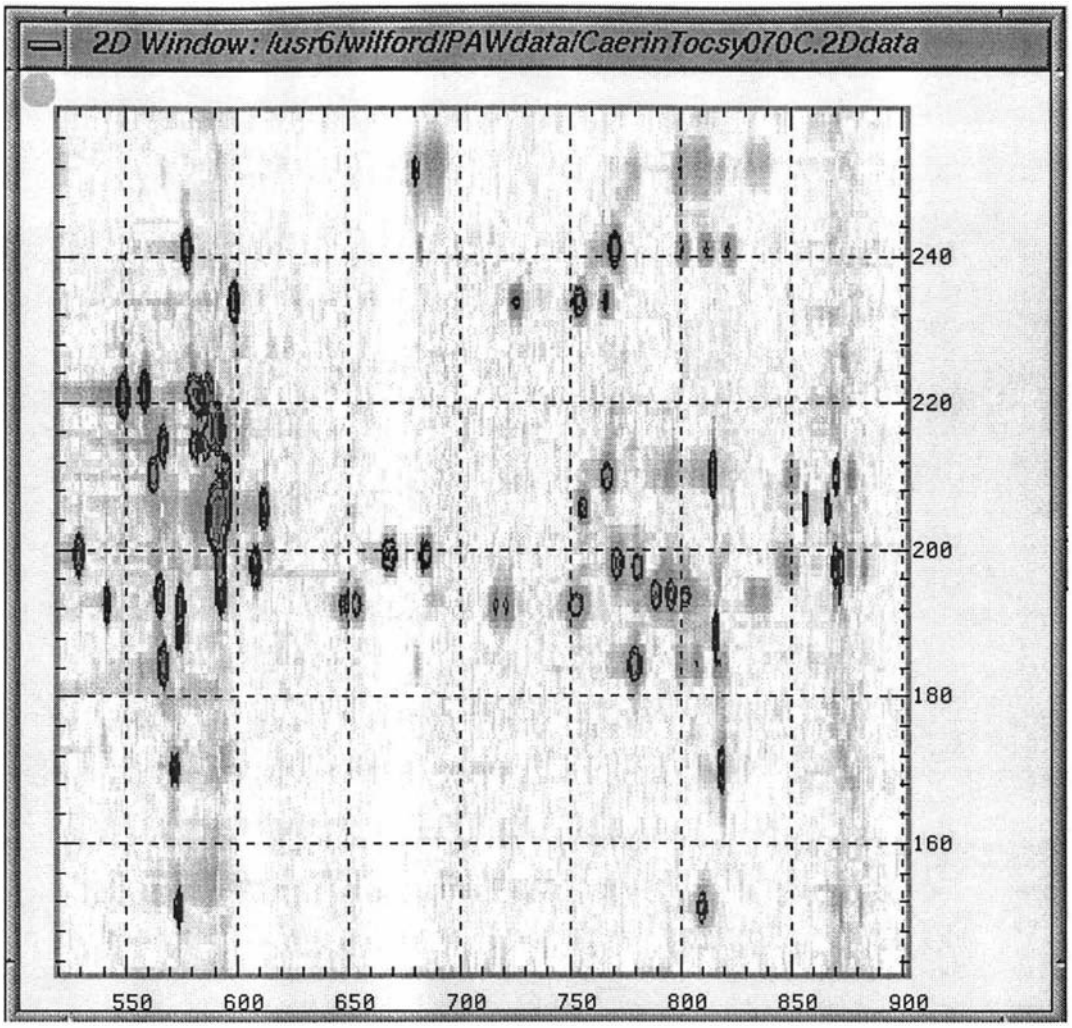


Figure 8.31: The lower-right region of the processed Caerin 4.1 TOCSY070 spectrum.

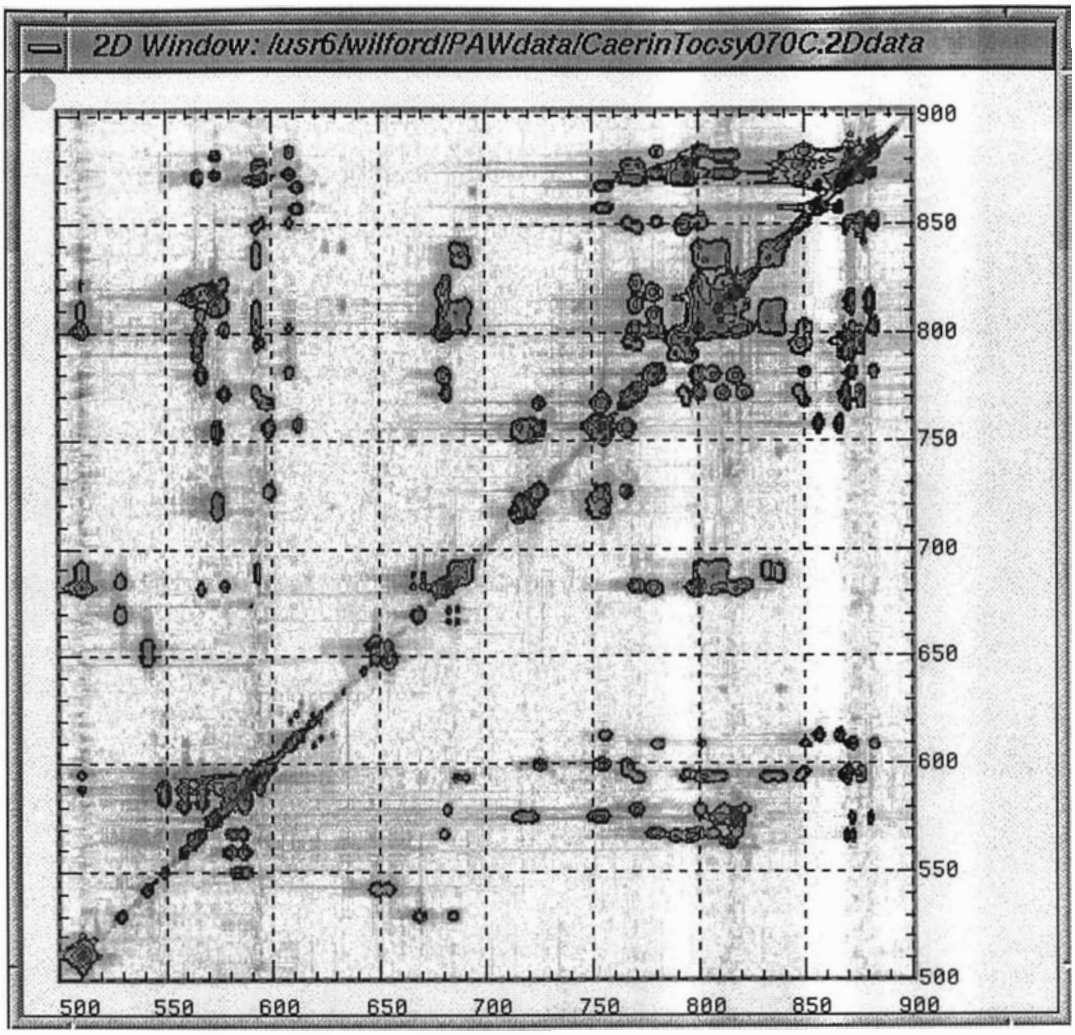


Figure 8.32: The upper-right (aliphatic) region of the processed Caerin 4.1 TOCSY070 spectrum.

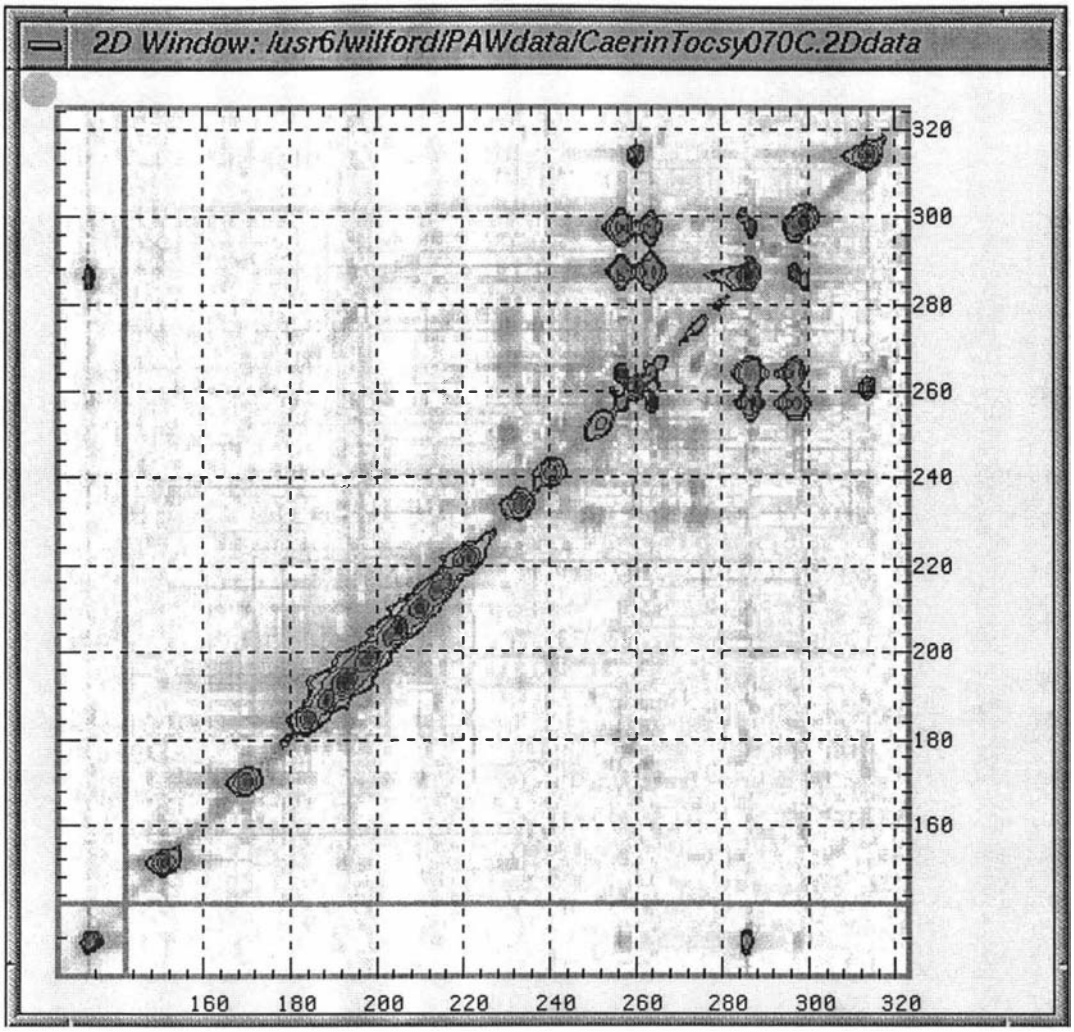


Figure 8.33: A multi-region view of the lower-left area of the processed Caerin 4.1 TOCSY070 spectrum.

8.4 Processing the Caerin 4.1 NOESY150 NMR data set

8.4.1 The raw data

Figure 8.34 shows the intensity plot of the Caerin 4.1 NOESY150 time-domain data.

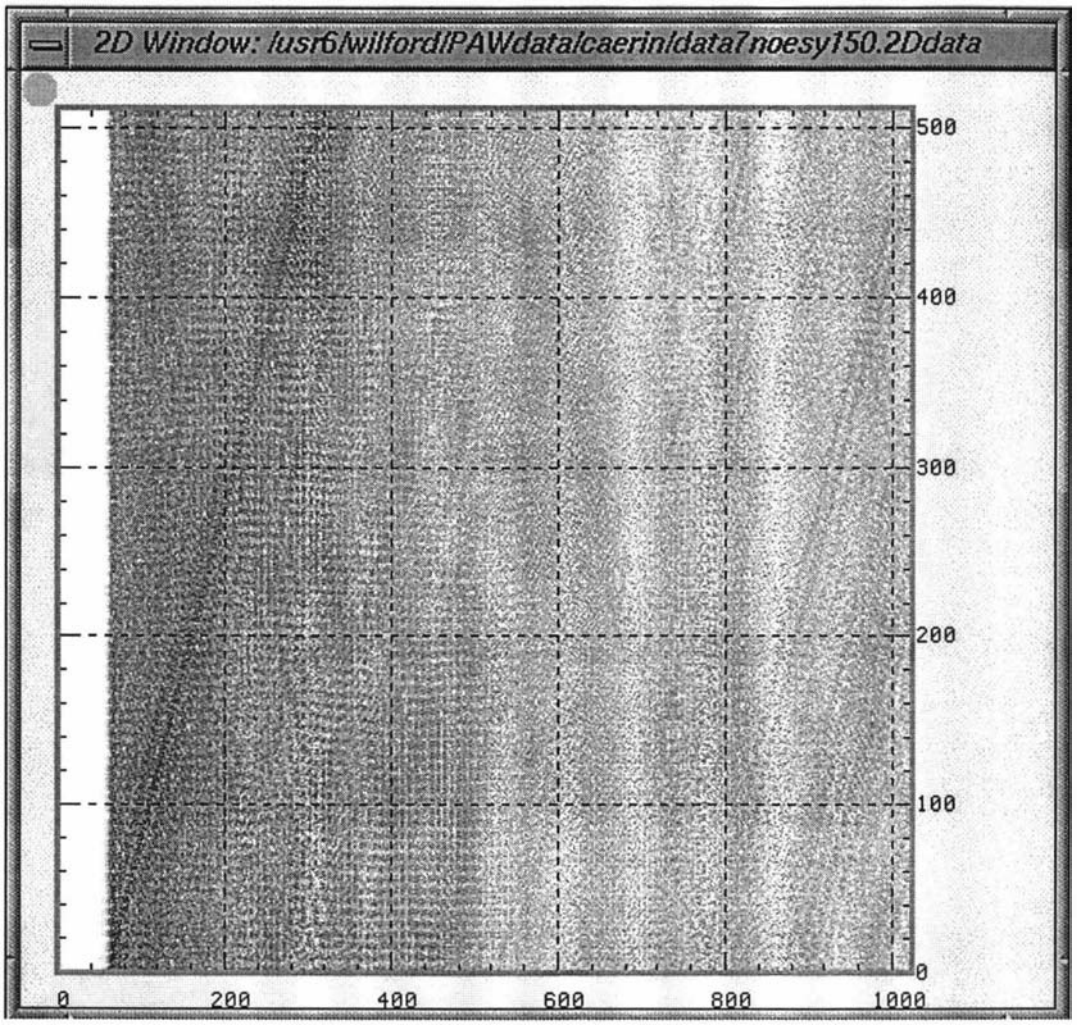


Figure 8.34: The intensity plot of the Caerin 4.1 NOESY150 raw data.

Like the TOCSY, the quality of the raw NOESY150 data set was very good, as shown by the real plots of four typical FIDs in Figure 8.35.

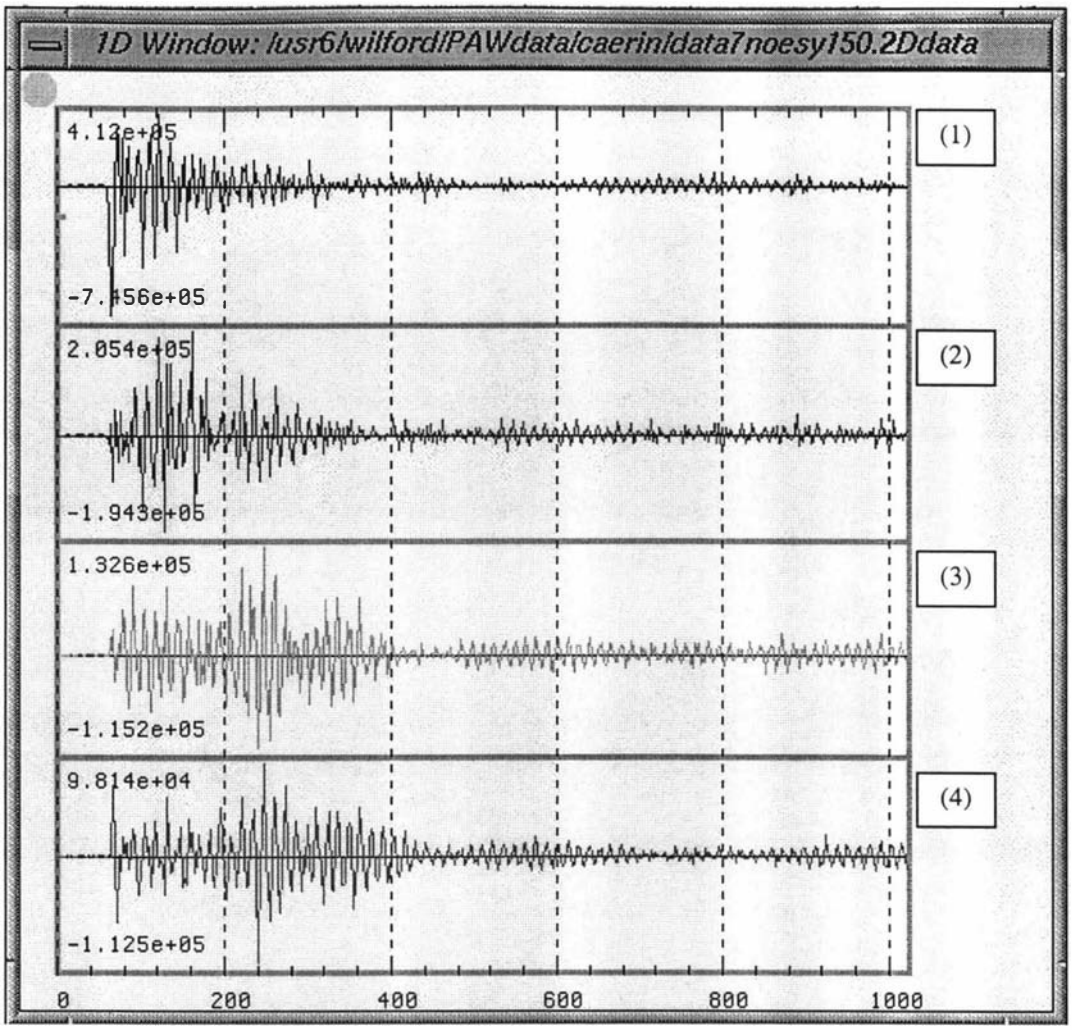


Figure 8.35: Row 0, 200, 400 and 511 of the Caerin 4.1 NOESY150 raw data.

Again, the acquisition time looks appropriate from the first plot.

8.4.2 High-resolution processing operations and results

8.4.2.1 D1-processing

As with the 2D TOCSY070 NMR data set, the first row of the 2D NOESY150 NMR data set contains the most intense signal. For the same reason stated in the previous section, it was used to work out the processing parameters in D1.

Figure 8.36 shows the raw data of the first row and the processed results.

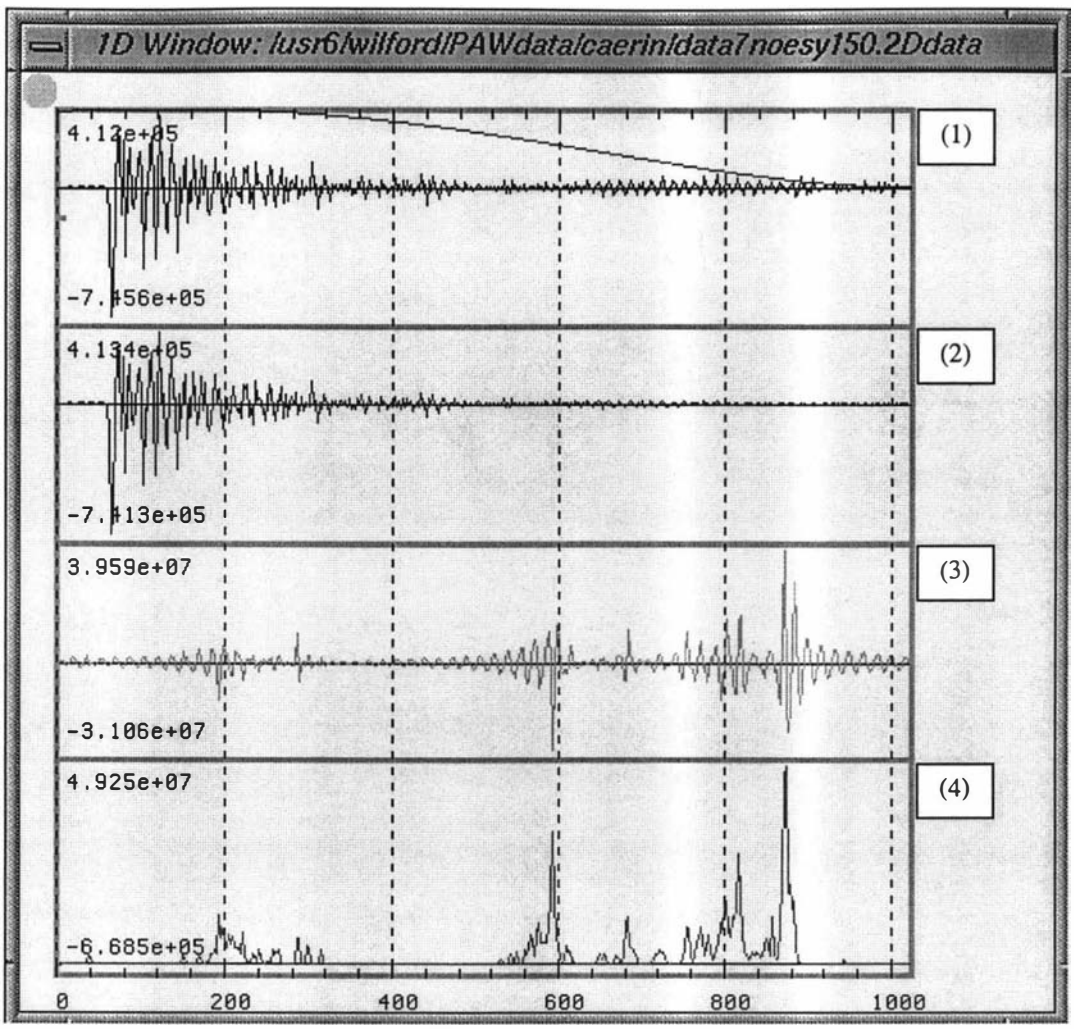


Figure 8.36: The processed results of the first row of the Caerin 4.1 NOESY150 data. The first plot is the raw data and the filter to be applied. The second plot is the data after applying the filter. The third plot is the frequency-domain spectrum after FFT. The last plot is the spectrum after phase correction.

Based on the 1D processing operations for the first row, the following macro was produced:

```
# (data4noesy150.2Dproc)
MD1  = 1024    # Matrix D1
MD2  = 1024    # Matrix D2

lpCom1    = "n"                              # D1 linear prediction
lpCom2    = "lp (100,412,20,288, r,t )"      # D2 linear prediction

fltCom1   = "esb (1024, 80, 2, -100)"        # D1 filtering
fltCom2   = "esb (800,  80, 2, 0)"           # D2 filtering

phCom1    = "ph (-78, -21978, 0)"            # D1 phasing
phCom2    = "ph (-12, 27, 37.8)"             # D2 phasing

bcCom1    = "n"     # D1 baseline correction
bcCom2    = "n"     # D2 baseline correction

DimNoToTrans = b    # Dimension to transform
ExprmtType   = t    # Experiment type
```

```
sbs(4, 340,480, 44,140, 900,930, 960,1020) # Set baseline segments
```

Figure 8.37 shows the 2D intermediate data set after processing in D1 with the above parameters.

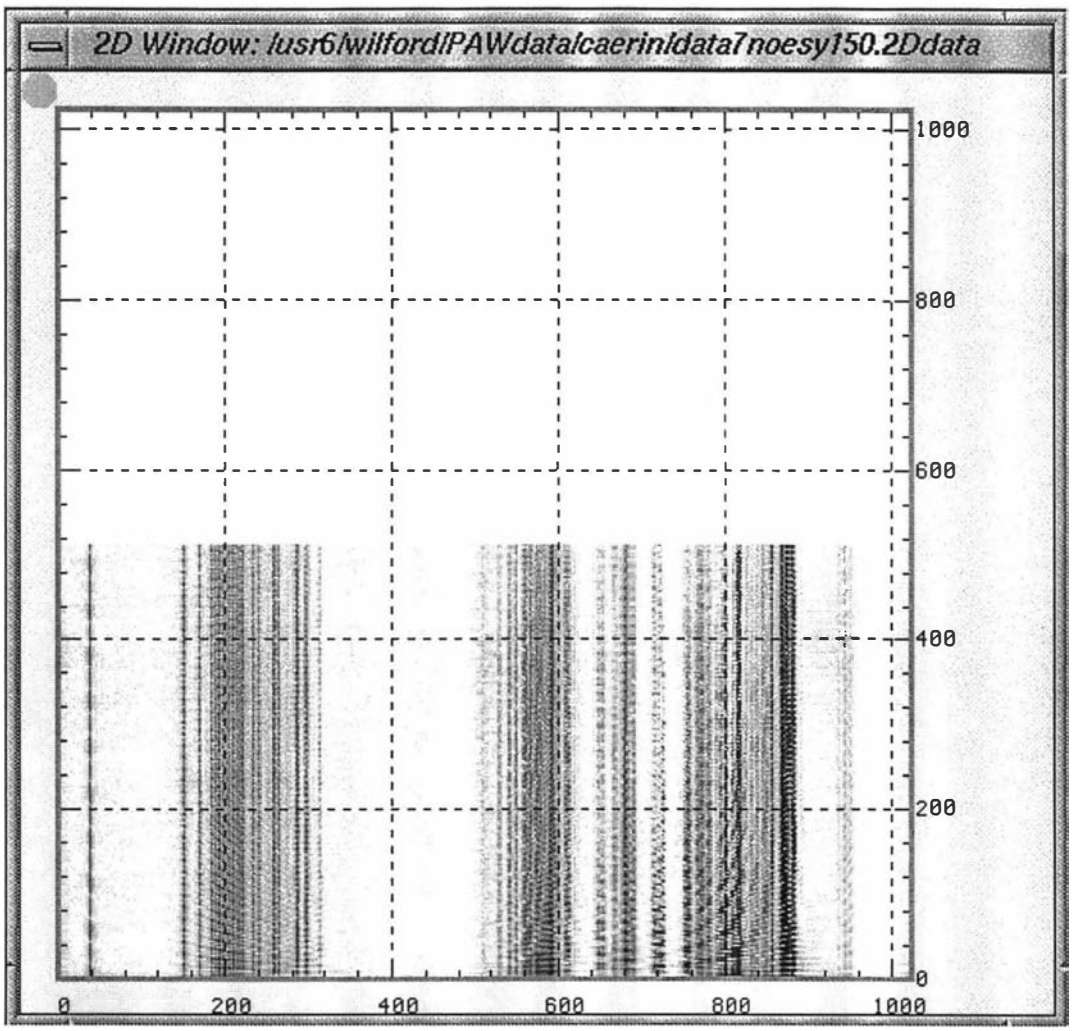


Figure 8.37: The Caerin 4.1 TOCSY070 intermediate data after processing in D1.

8.4.2.2 D2-processing

Again, linear prediction was applied during D2-processing to narrow the peaks and improve the resolution of the spectrum. Figure 8.38 shows the combined FID from columns 38 and 877, as well as the processed results at various stages.

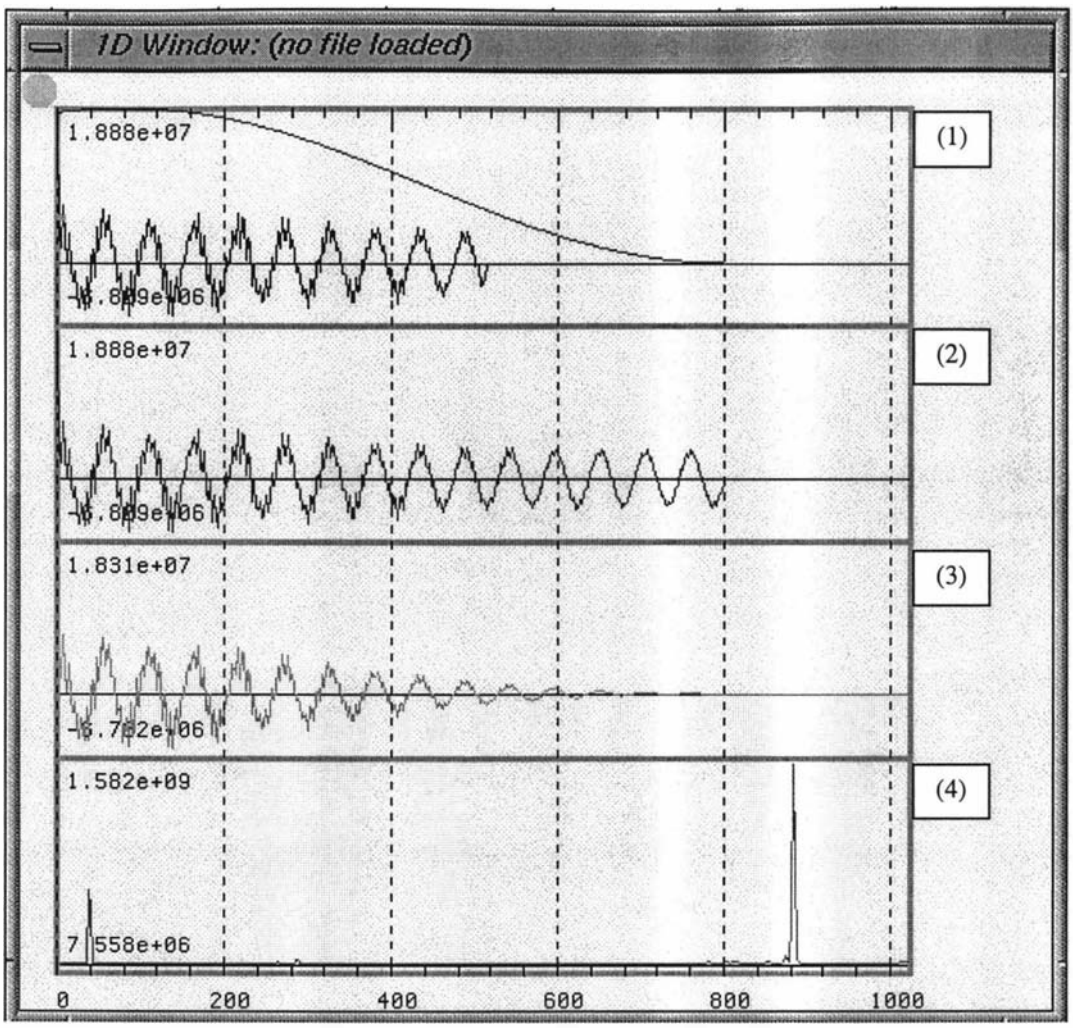


Figure 8.38: The processed results of a combined column from the intermediate data set. The first plot is the time-domain data and the filter to be applied. The second plot is the data after applying a linear prediction. The third plot is the data after applying the filter. The last plot is the final result after applying the real Fourier transform and the phase correction.

Figure 8.39 shows the 2D spectrum after processing in D2.

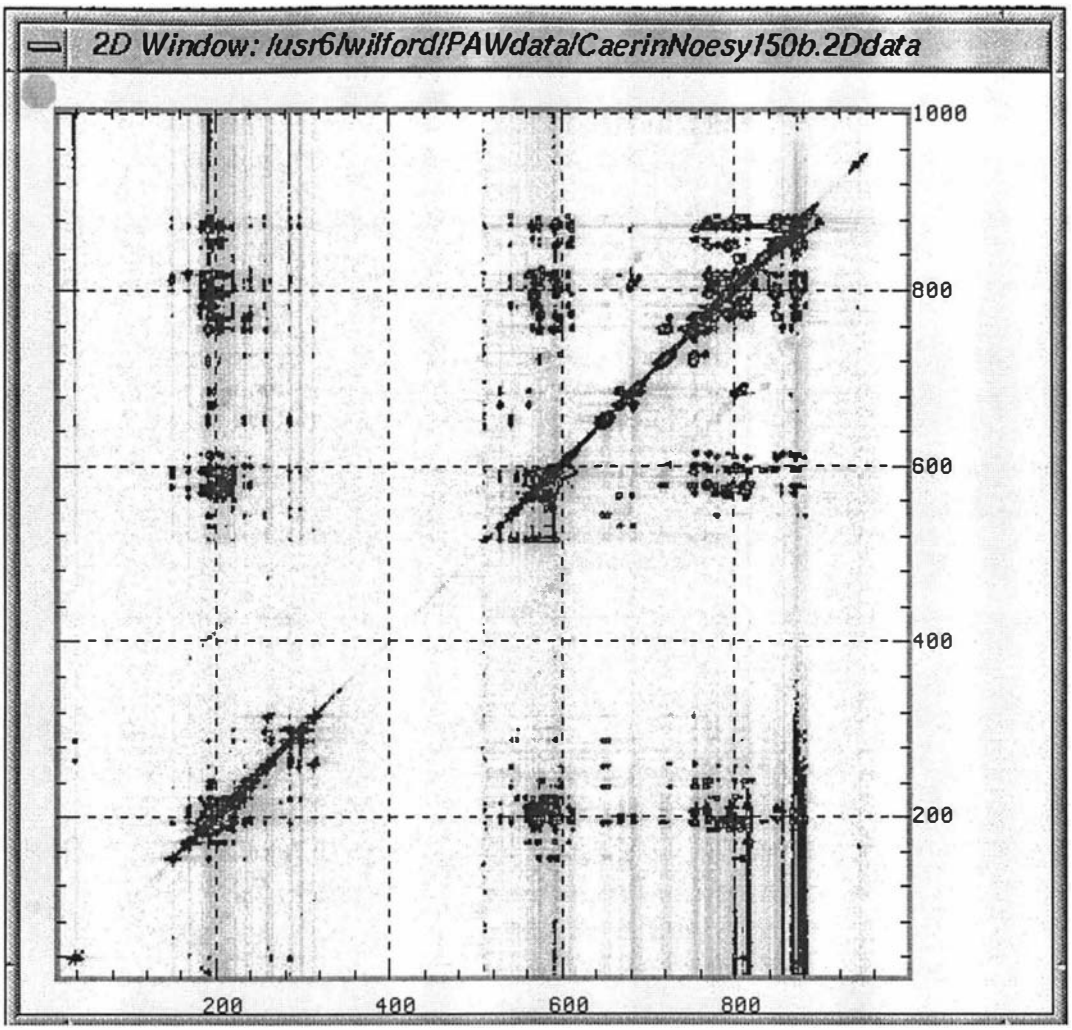


Figure 8.39: The Caerin 4.1 NOESY150 spectrum after processing in D2.

In this case, a total of 16 columns returned linear-prediction errors.

The vertical streaks in the 2D spectrum indicate some baseline problem, which was corrected with the same macro used for refining the Caerin 4.1 TOCSY070 described in the previous section.

Figure 8.40 shows the NOESY150 spectrum after baseline correction. A comparison with the previous spectrum shows that the correction was successful.

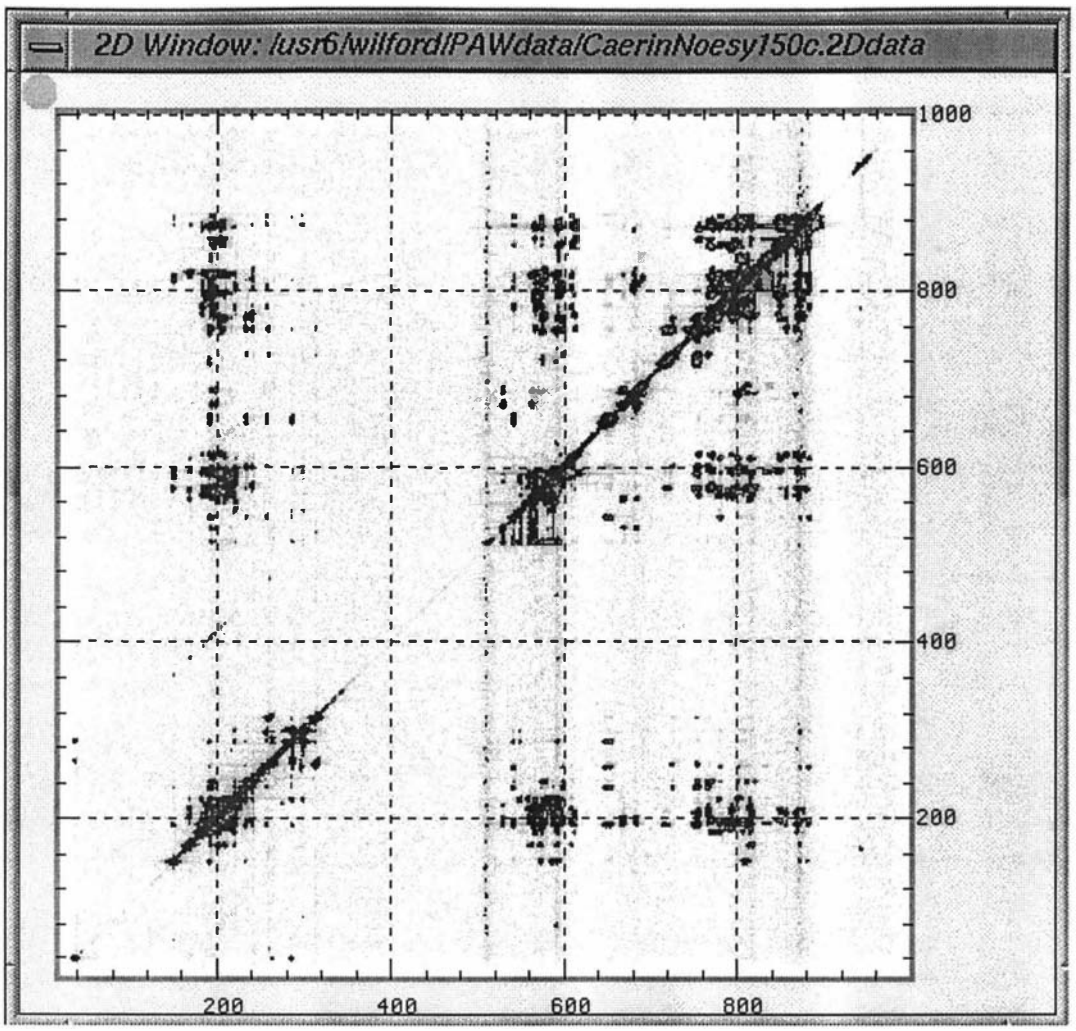


Figure 8.40: The Caerin 4.1 NOESY150 spectrum after baseline correction.

Figures 8.41 to 8.44 show the contour plots of all the significant regions of the 2D NOESY150 spectrum. Note that the weak peaks that do not appear in the contour plots are still clearly visible in the 1D column and row plots in the next two figures.

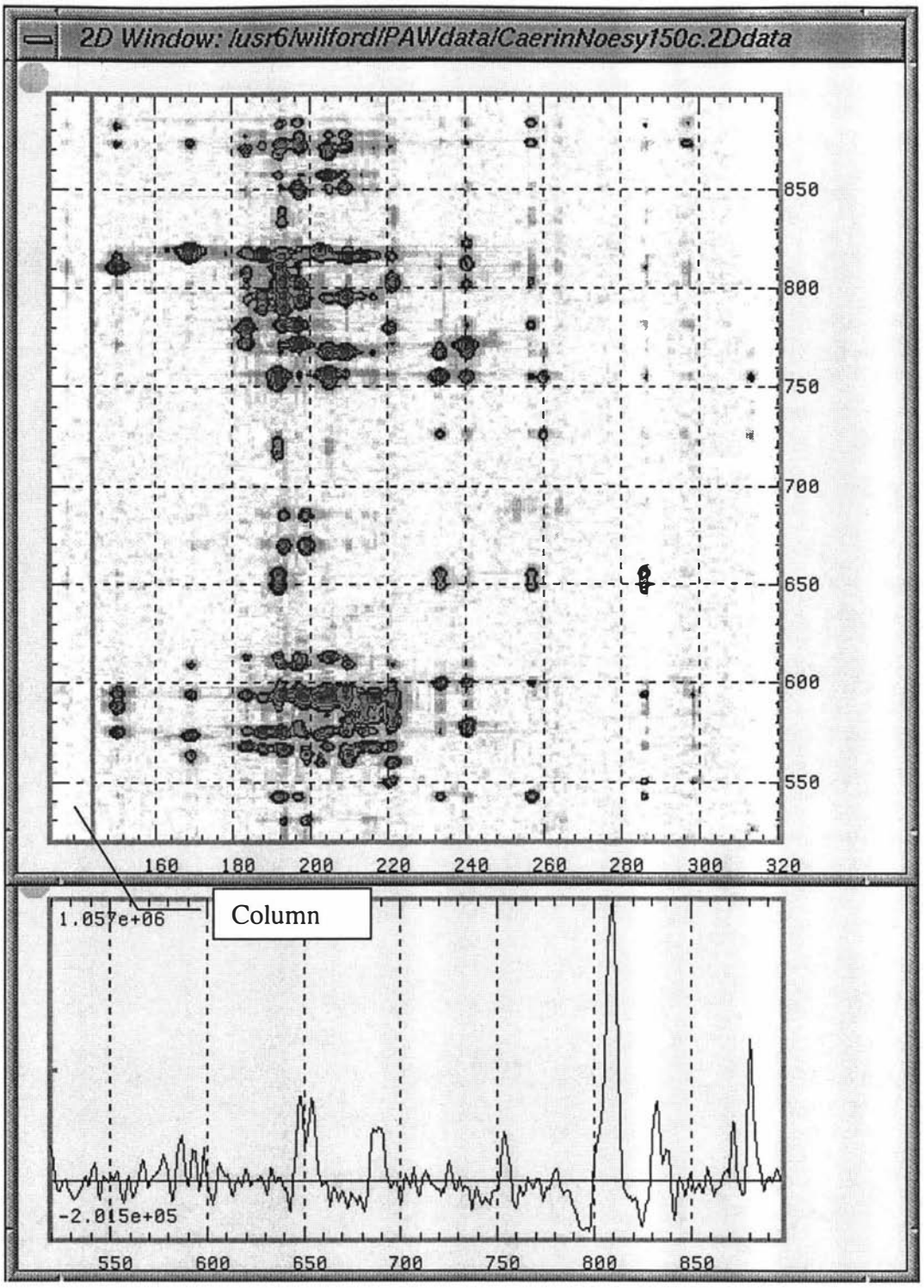


Figure 8.41: The upper-left region of the Caerin 4.1 NOESY150 spectrum. The 1D plot underneath is column 38 that contains a number of very weak peaks.

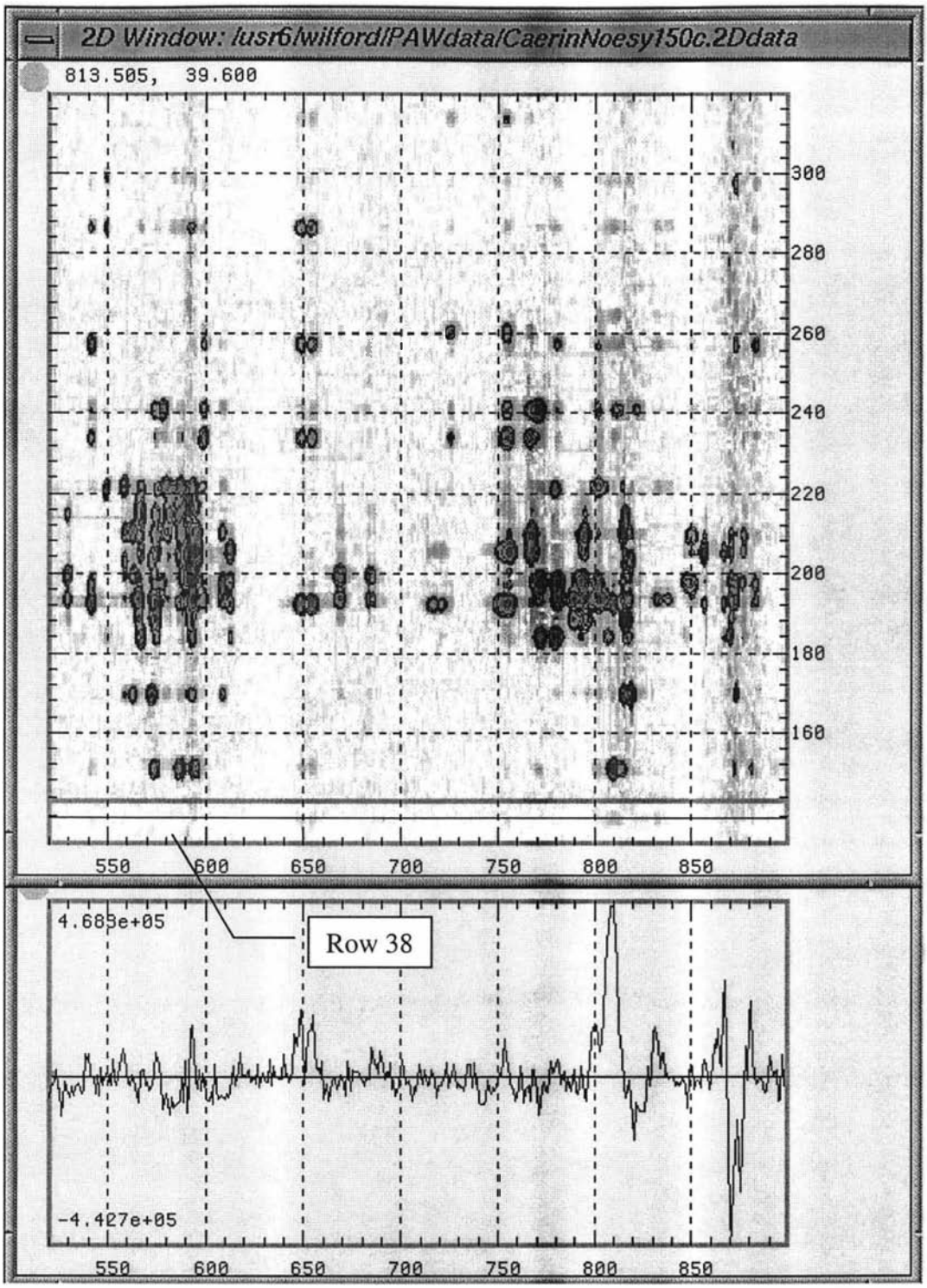


Figure 8.42: The lower-right region of the Caerin 4.1 NOESY150 spectrum. The 1D plot underneath is row 38 that contains a number of very weak peaks.

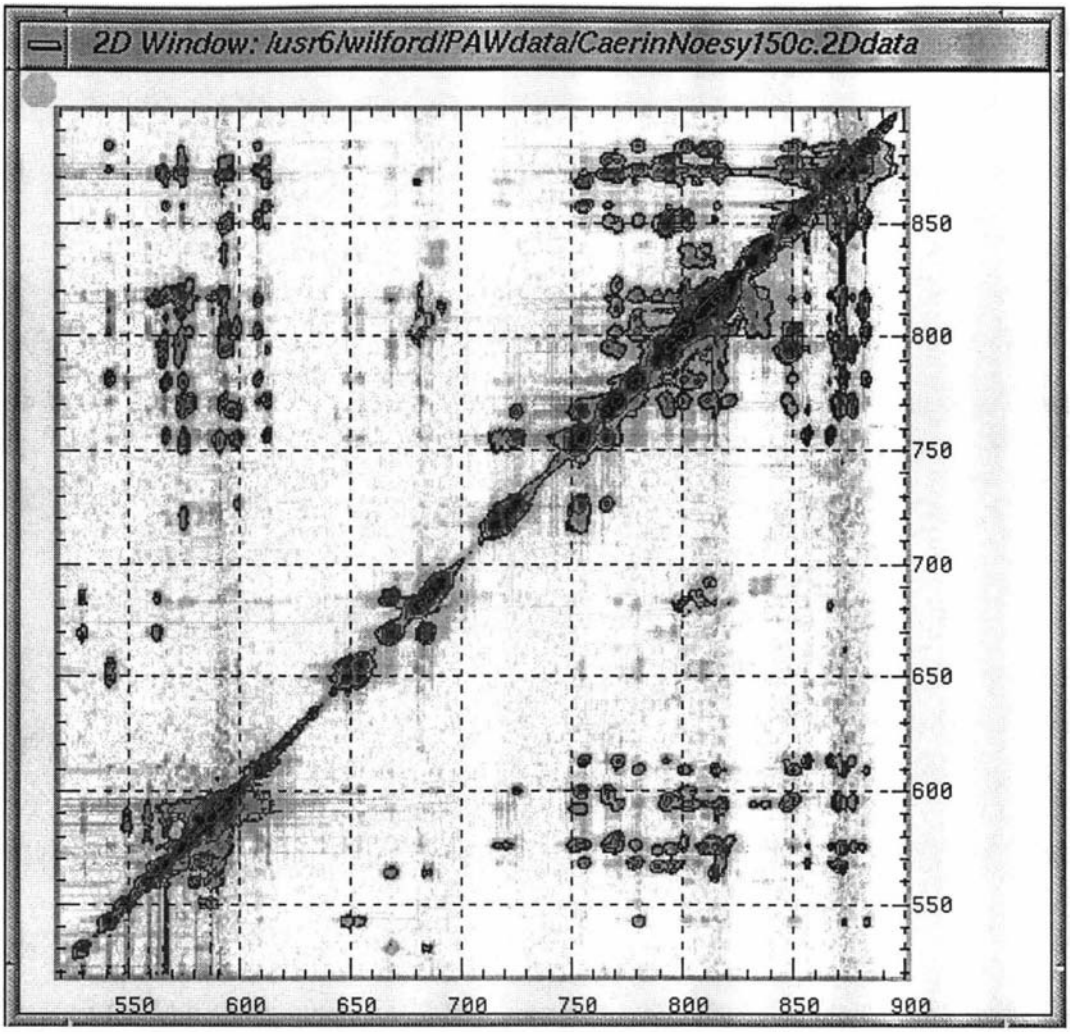


Figure 8.43: The upper-right region of the processed Caerin 4.1 NOESY150 spectrum.

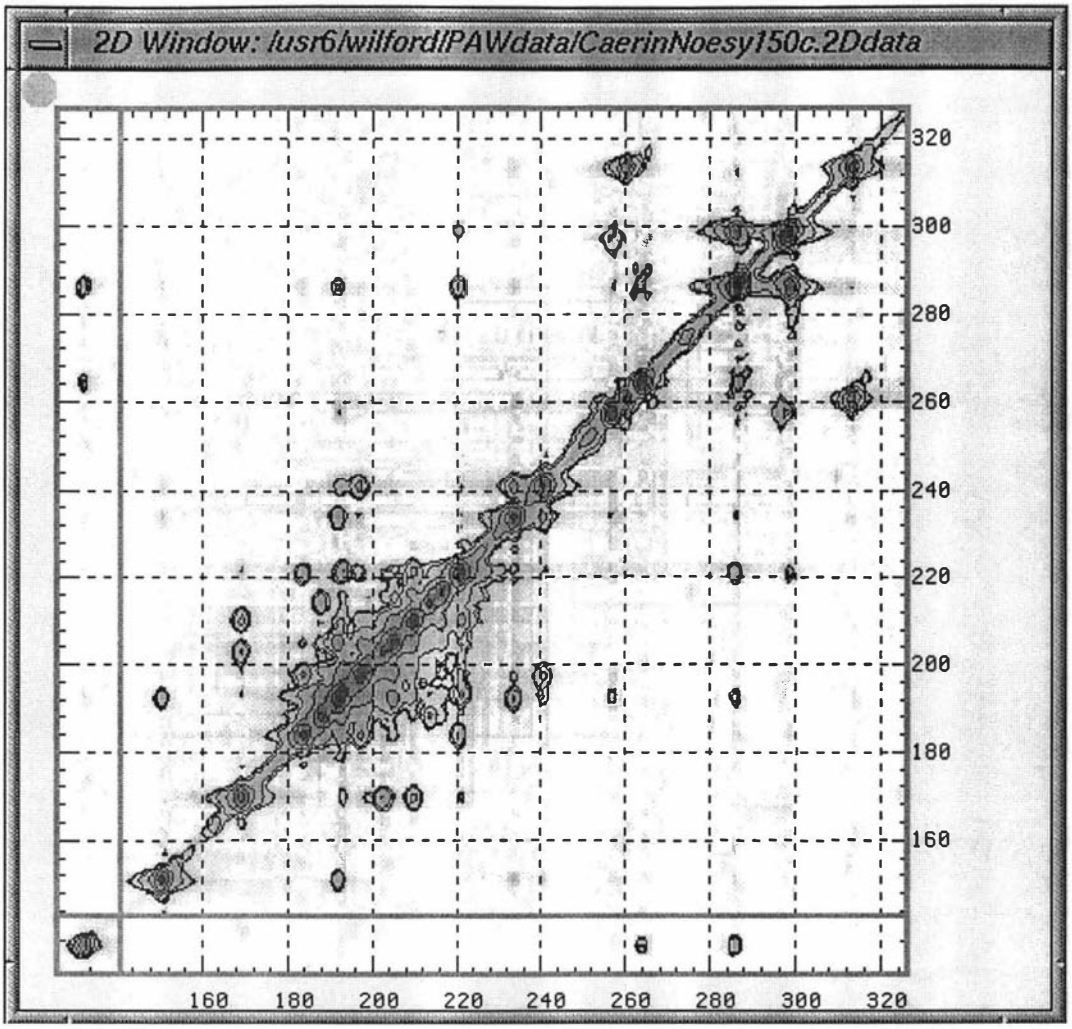


Figure 8.44: A multi-region view of the lower-left region of the processed Caerin 4.1 NOESY150 spectrum.

Chapter 9:

Peak-picking Process for Caerin 4.1 NMR Spectra Using PAW

9.1 Introduction

As mentioned in Chapter 7, PAW uses three kinds of peak lists during spectral assignment: a raw-peak list, a diagonal peak list, and a cross-peak list. In the spectral assignment of Caerin 4.1, the raw peak list contains the coordinates of peaks that are picked by PAW's automatic peak-picking routine. The other two peak-lists are extracted from the raw-peak list and then refined. Accordingly, the peak-picking process for Caerin 4.1 consists of three major operations to create and maintain the three peak lists.

PAW's peak-picking routine is capable of picking in-phase and anti-phase peaks in any kind of NMR spectra. To avoid duplication of work, only the NOESY peaks were picked for the spectral assignment of Caerin 4.1. These peaks were saved at each stage of the picking process in order to avoid unexpected mistakes or problems. The methods for picking DQF-COSY anti-phase peaks are the same and will not be presented.

This chapter discusses some major results of the peak-picking operations for the Caerin 4.1 NOESY150 spectrum using the automatic and interactive peak-picking routines of PAW. More details of the picking operations can be found in Chapter 9 of Volume II.

To be consistent with the graphic titles, the combined word **CaerinNoesy150C** will be used to represent the processed Caerin 4.1 NOESY150 spectrum.

9.2 Raw-peak Picking

The raw peaks were picked from four regions of the **CaerinNoesy150C** spectrum. In each region, a base level used in the picking process was calculated using the data within a small area that contained a medium-intensity cross-peak. The parameter values in the *Peak-picking Dialog* shown in Figure 9.1 were used for all raw-peak picking operations.

This chapter describes the peak-picking operations to create the peak lists used for the NMR spectral assignment of Caerin 4.1. All the three high-resolution spectra obtained in Chapter 8 were used during the peak picking.

To avoid unexpected errors, it is recommended that PAW is restarted and the same workbench as that for the operations in Chapter 7 is used.

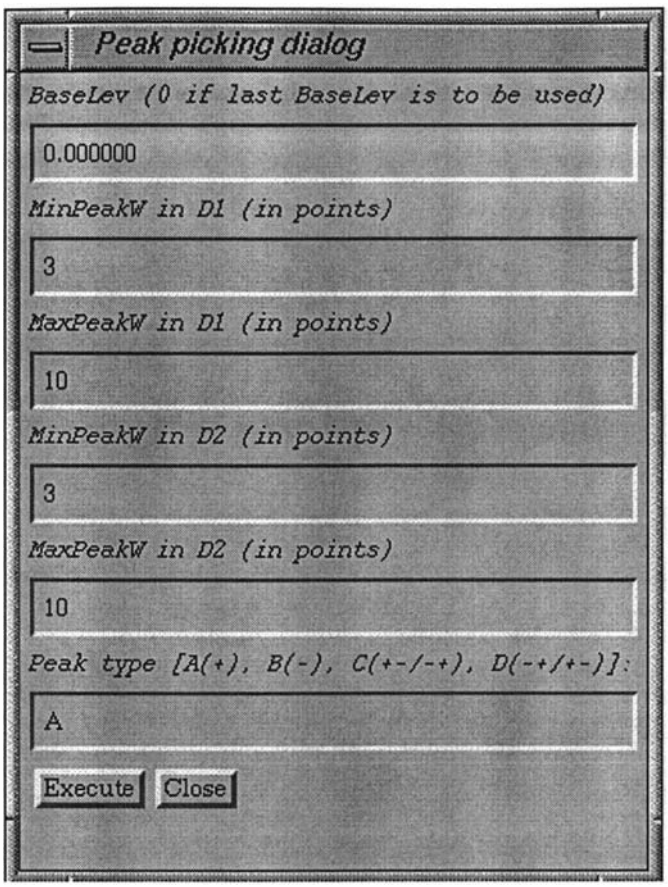


Figure 9.1: The peak-picking dialog for raw-peak picking operation.

To preserve potentially useful information in the spectral assignment process, the coordinate of these peaks were kept unmodified at the initial stage. Only minor modifications were made during the assignment process whenever it was necessary and clear, as described in Chapter 10 and 11 of Volume II.

9.2.1 Raw-peak picking in the upper-left region

The raw peaks in this region were picked with the base level set to around 1.08e+06, which was calculated from a used-defined area that enclosed a small peak at around [6.828,2.078][1]. Figure 9.2 shows the enclosed peak and the picking result.

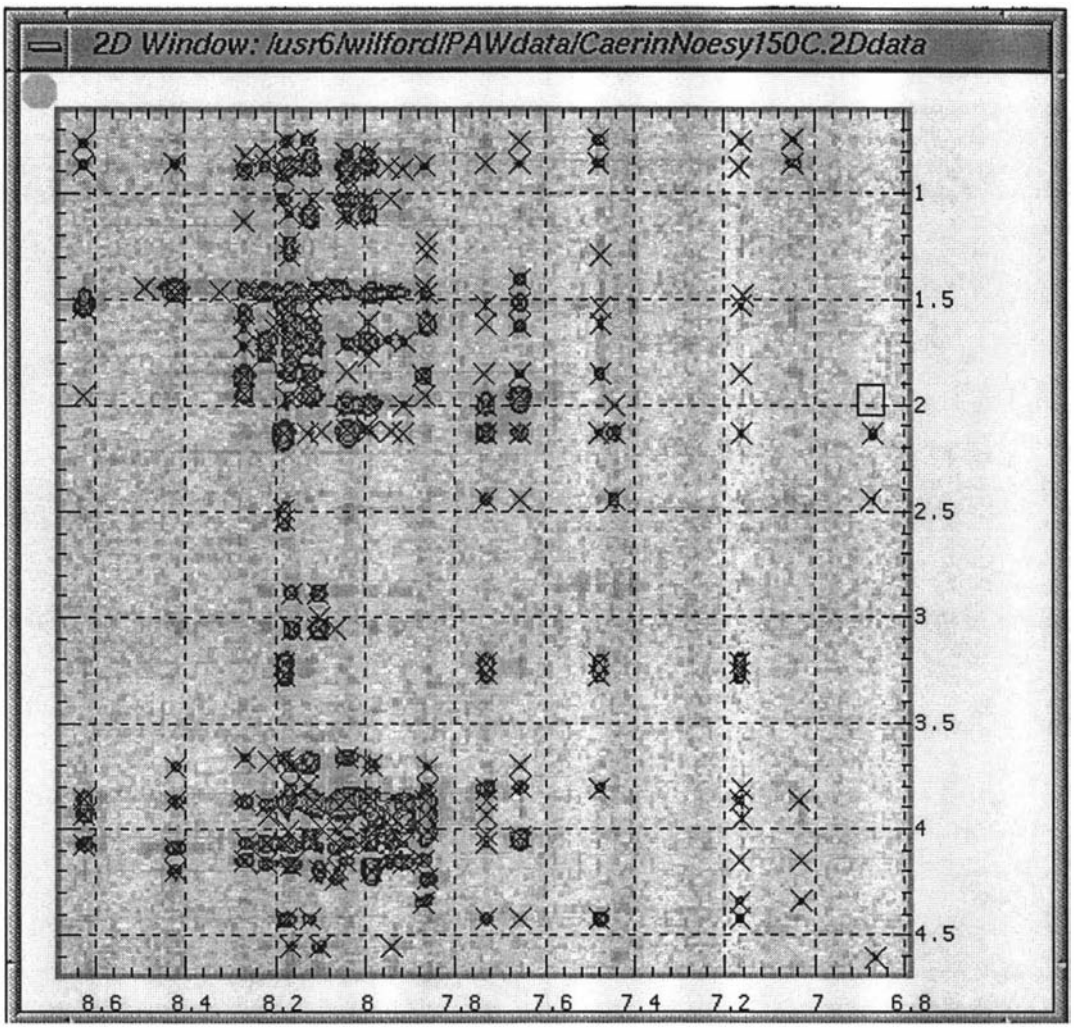


Figure 9.2: The raw peaks in the fingerprint region. The small rectangle on the right is the area used to define the base level. The crosses are the peaks picked.

[1] For the rest of this thesis, ppm units will be used for all spectral locations.

9.2.2 Raw-peak picking in the lower-right region

The raw peaks in this region were picked with the same parameter values used in the last subsection, including the base level. The result is shown in Figure 9.3.

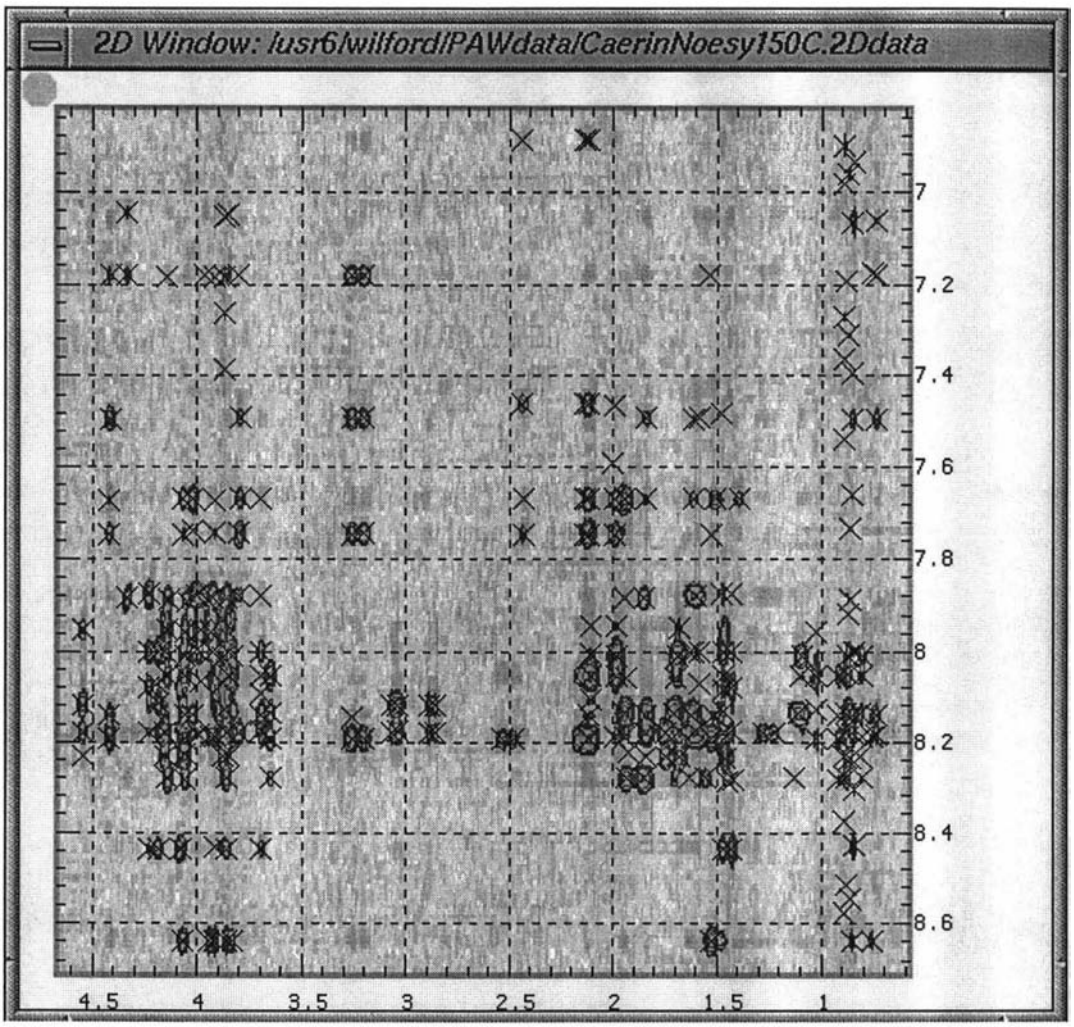


Figure 9.3: The raw peaks in the lower-right region. These peaks were picked with the same base level used for the last section.

9.2.3 Raw-peak picking in the upper-right region

Because of the higher level of noise, a higher base level at around 3.66e+06 was used to pick the raw peaks in this region. This level was calculated by selecting a small region that encloses a small peak at around [4.368,0.696], which is the most intense peak at the top-left corner in Figure 9.4 that shows the picking result.

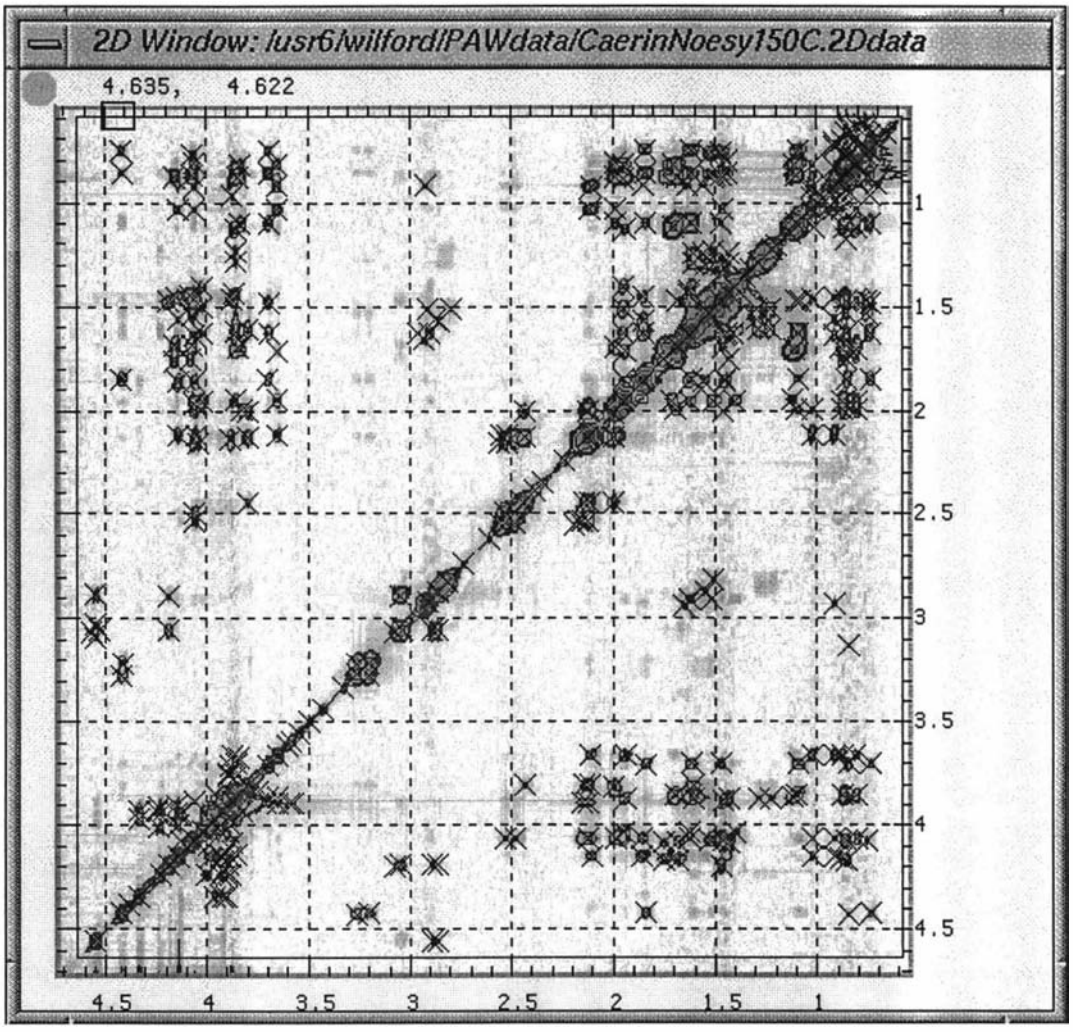


Figure 9.4: The raw peaks in the aliphatic region. The small rectangle at the top-left corner is the area from which the base level was set.

It can be seen that some weak peaks were not picked with this base level. These peaks were considered at the later stage of the spectral assignment.

9.2.4 Raw-peak picking in the lower-left region

A different base level at around 2.17e+06 was used for the peak-picking operation in this region. This level was calculated by selecting a small region that encloses a small peak at around [8.134,7.124]. This peak and the picking results are shown in the next figure.

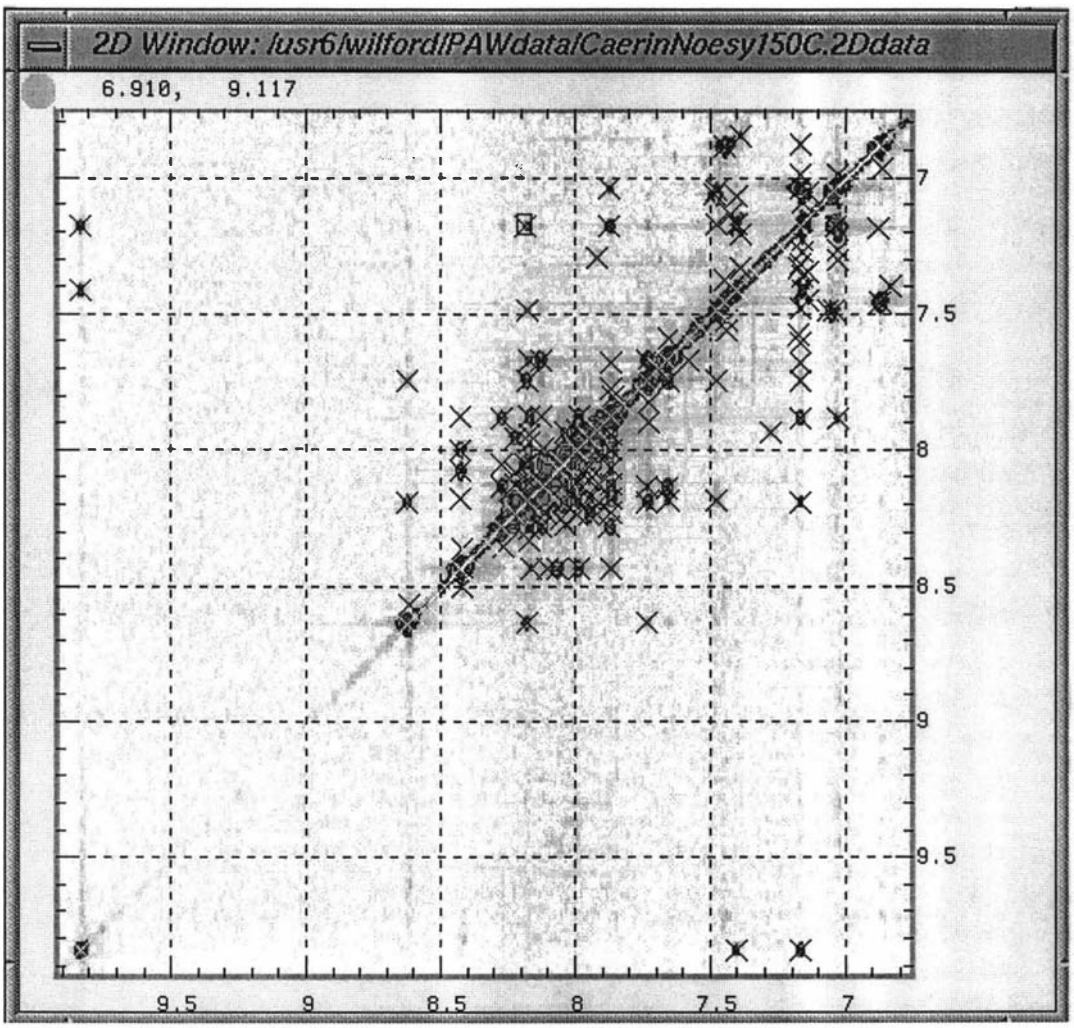


Figure 9.5: The raw peaks in the lower-left region. The small rectangle at the centre-top is the area from which the base level was set.

9.3 Diagonal-peak Picking

PAW uses diagonal peaks to locate the transposed partner of a peak. This technique effectively avoids serious mistakes in locating transposed peaks, especially when the Bloch-Siegert effect is significant.

9.3.1 Diagonal-peak picking in the lower-left region

The diagonal peaks in this region were picked by PAW's diagonal-peak picking routine using the two peaks at the diagonal ends in the next figure. The diagonal peaks in the entire region had been closely checked and no correction was found to be necessary.

All the diagonal peaks in the plot were picked, as shown in Figure 9.6.

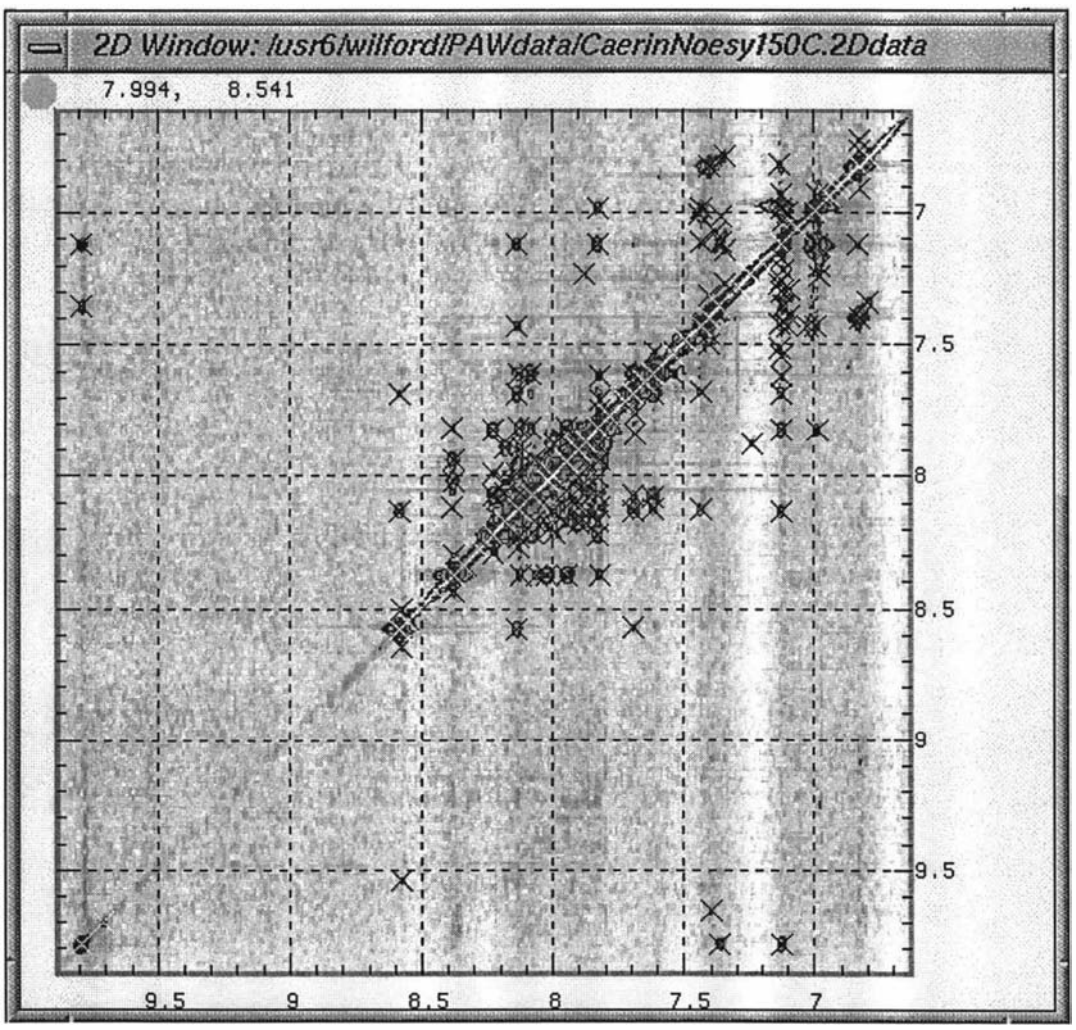


Figure 9.6: The diagonal peaks in the lower-left region. They are shown with a brighter colour along the diagonal.

9.3.2 Diagonal-peak picking in the upper-right region

The diagonal peaks in this region were picked by the same method as described in the last subsection. The peaks are shown in Figure 9.7.

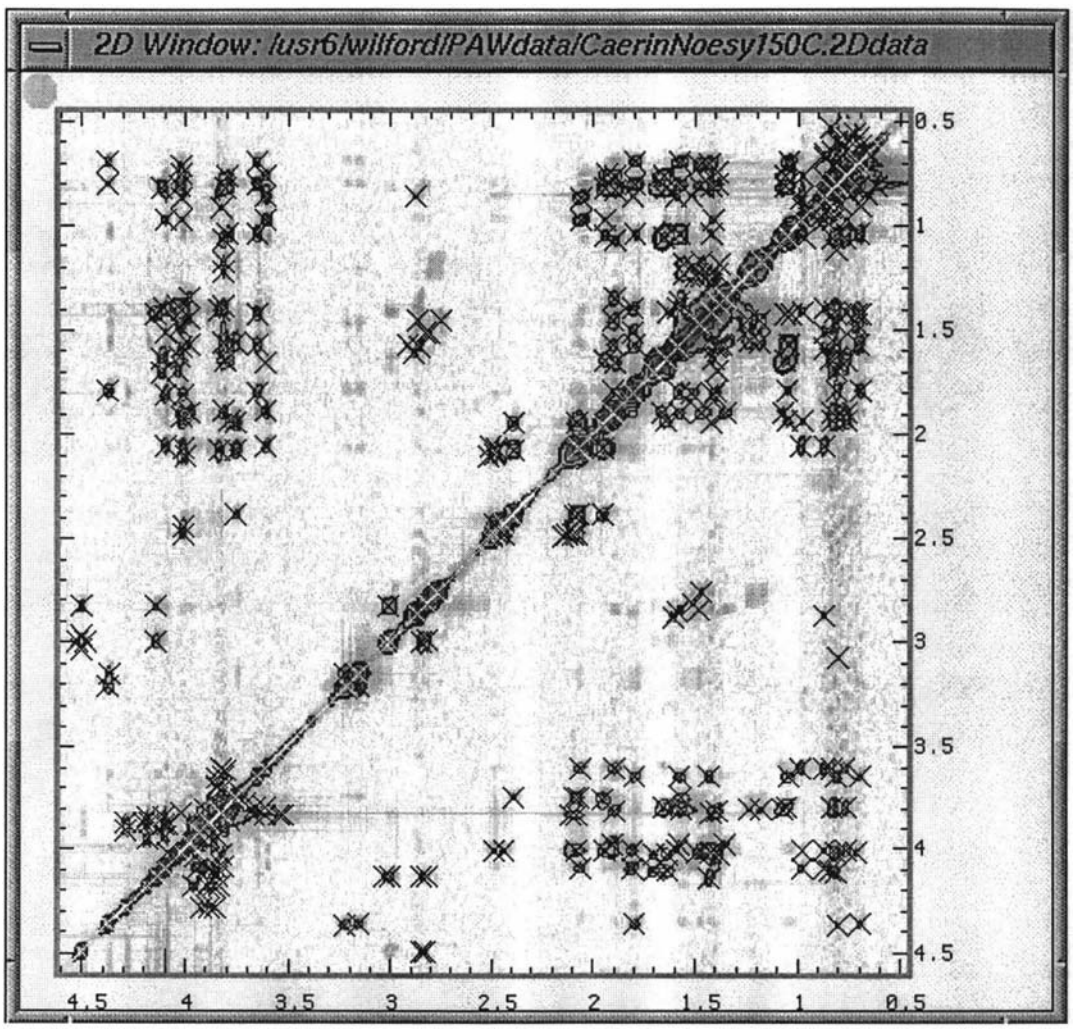


Figure 9.7: The diagonal peaks in the upper-left region. They are shown with a brighter colour along the diagonal.

9.3.3 Diagonal-peak refinement in the upper-right region

Four of the diagonal peaks that were picked using the method described in the last subsection were found to be improperly located because of overlap. To correct this, it was necessary to zoom into three smaller regions in which all the diagonal peaks were clearly seen. A diagonal line was drawn in each of the small regions in order to assess the alignment of the diagonal peaks, as shown in Figure 9.8.

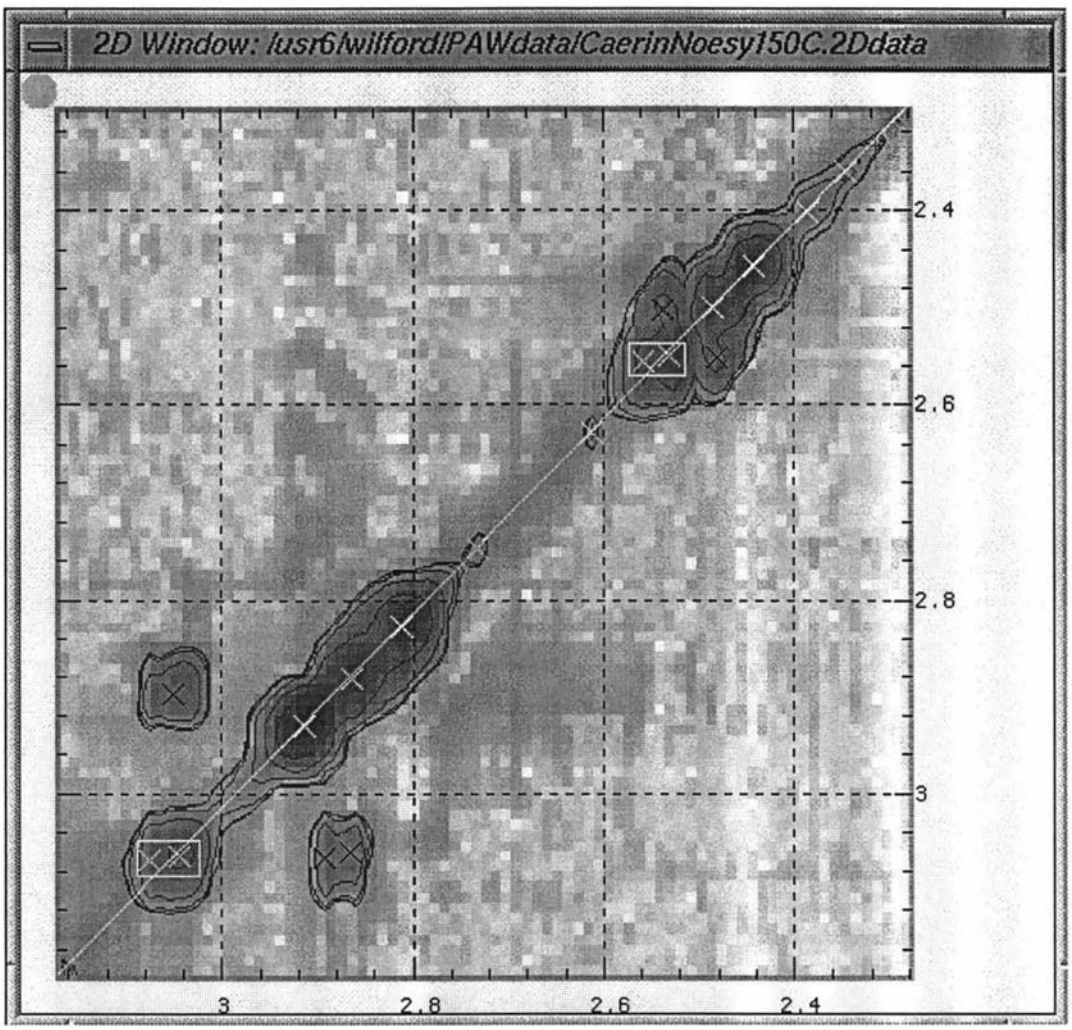


Figure 9.8: A diagonal line drawn in a small region for assessing the locality of the diagonal peaks that are automatically picked.

The rectangles enclosed four peaks that were furthest away from the diagonal. These peaks were removed from the diagonal-peak list because the remaining diagonal peaks were sufficient to define the diagonal, as shown in the next plot.

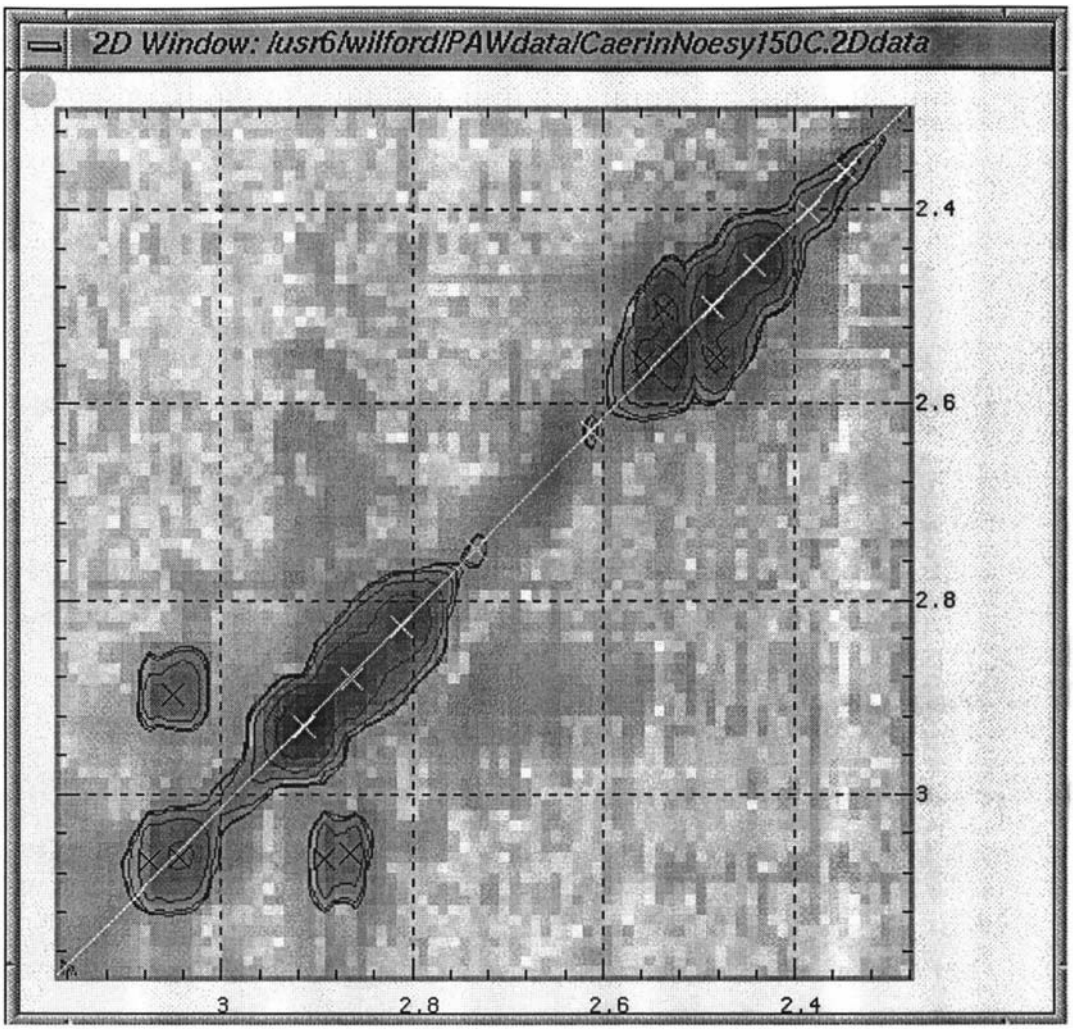


Figure 9.9: A set of carefully selected diagonal peaks after removing four peaks in the initial list. The remaining diagonal peaks were sufficient to define the diagonal.

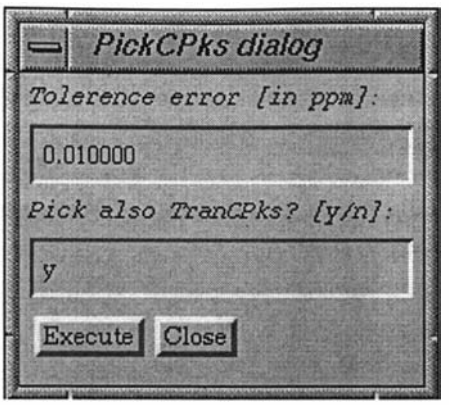
9.4 Cross-peak Picking

This section describes the results of cross-peak picking. The majority of the cross-peaks in the **CaerinNoesy150C** spectrum were initially picked from four regions by PAW's cross-peak picking routine. In practice, cross-peaks are picked in groups and pairs at the initial stage, as described in the next two subsections. Care must be taken when they are picked individually, as described in Section 9.4.3 and 9.4.4.

More details can be found in Chapter 9 of Volume II.

9.4.1 Cross-peak picking in the lower-left region

The operation was performed by selecting the entire region with the entry set to y for the second item in the *PickCPks Dialog* (Figure 9.10). The setting ensures only peaks that have a transposed element in the raw-peak list will be picked.



There were no complications in the peak-picking operation for this region. Figure 9.11 shows the result of the operation. In PAW, an unassigned cross-peak is displayed with a symbol that looks like an hour-glass, different from the simple cross used for a raw peak.

Figure 9.10: The dialog for picking cross-peaks.

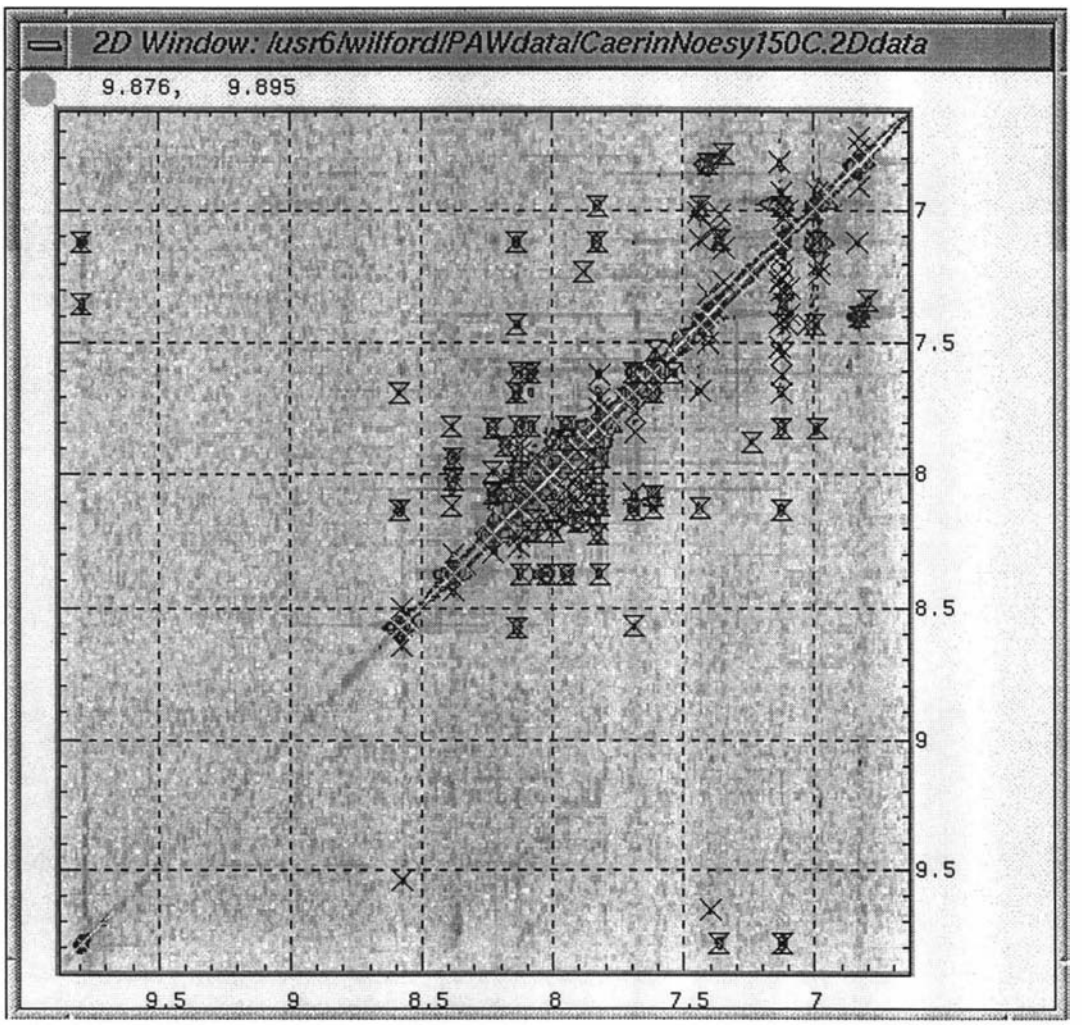


Figure 9.11: The cross-peaks in the lower-left region of the **CaerinNoesy150C** spectrum. Each of them is displayed by a symbol that looks like an hour-glass.

9.4.2 Cross-peak picking in the other regions

Cross-peak picking operations in other regions were the same as that described in the last subsection. The following three figures shows the results of the operations.

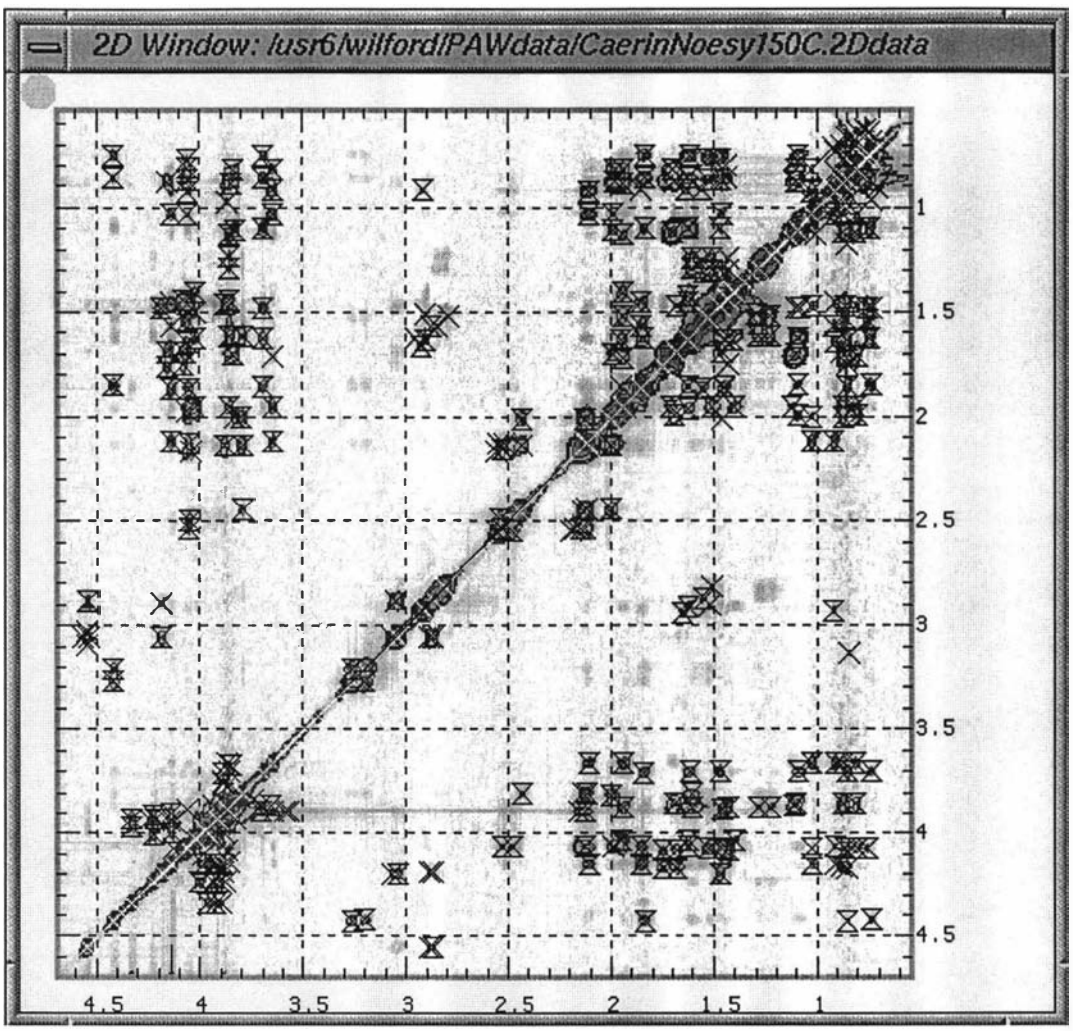


Figure 9.12: The cross-peaks in the upper-right region of the **CaerinNoesy150C** spectrum.

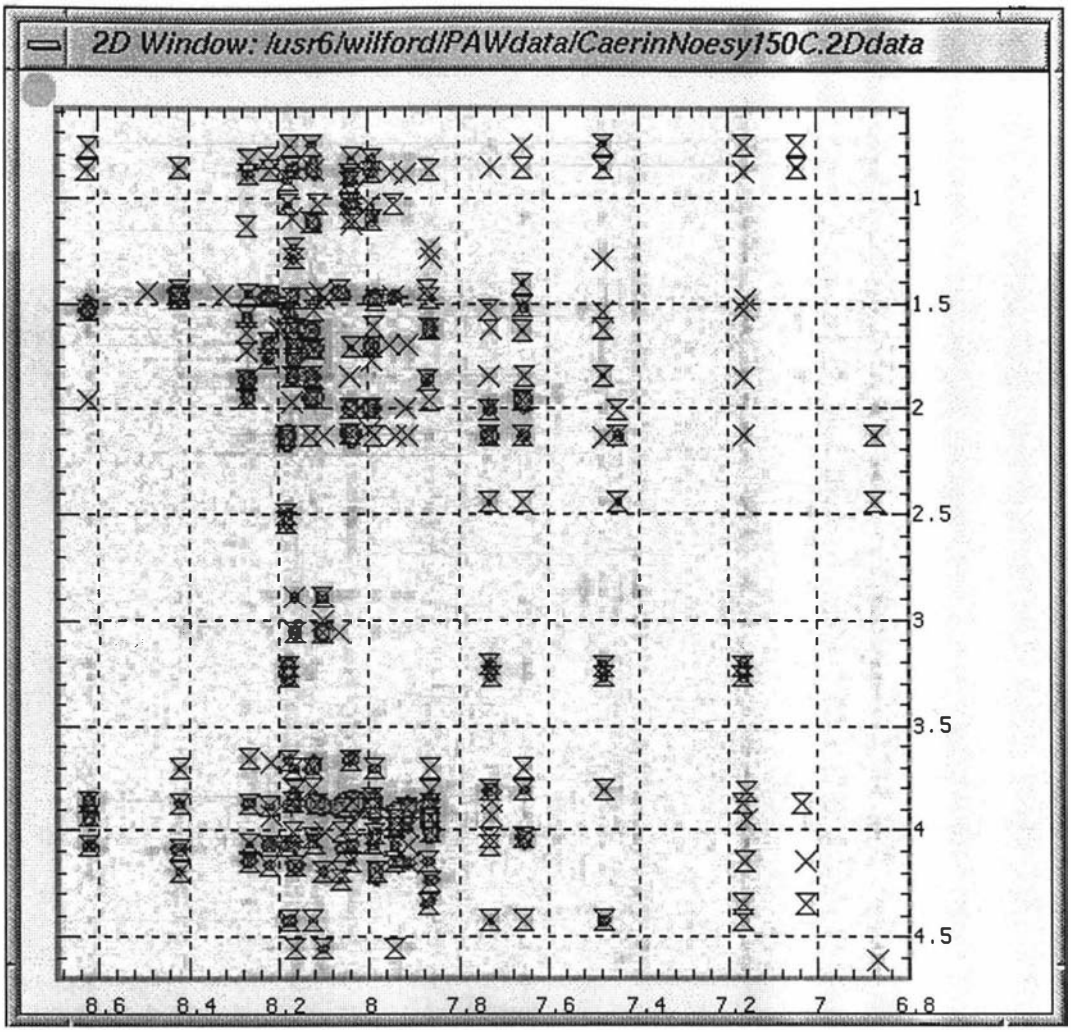


Figure 9.13: The cross-peaks in the upper-left region of the **CaerinNoesy150C** spectrum.

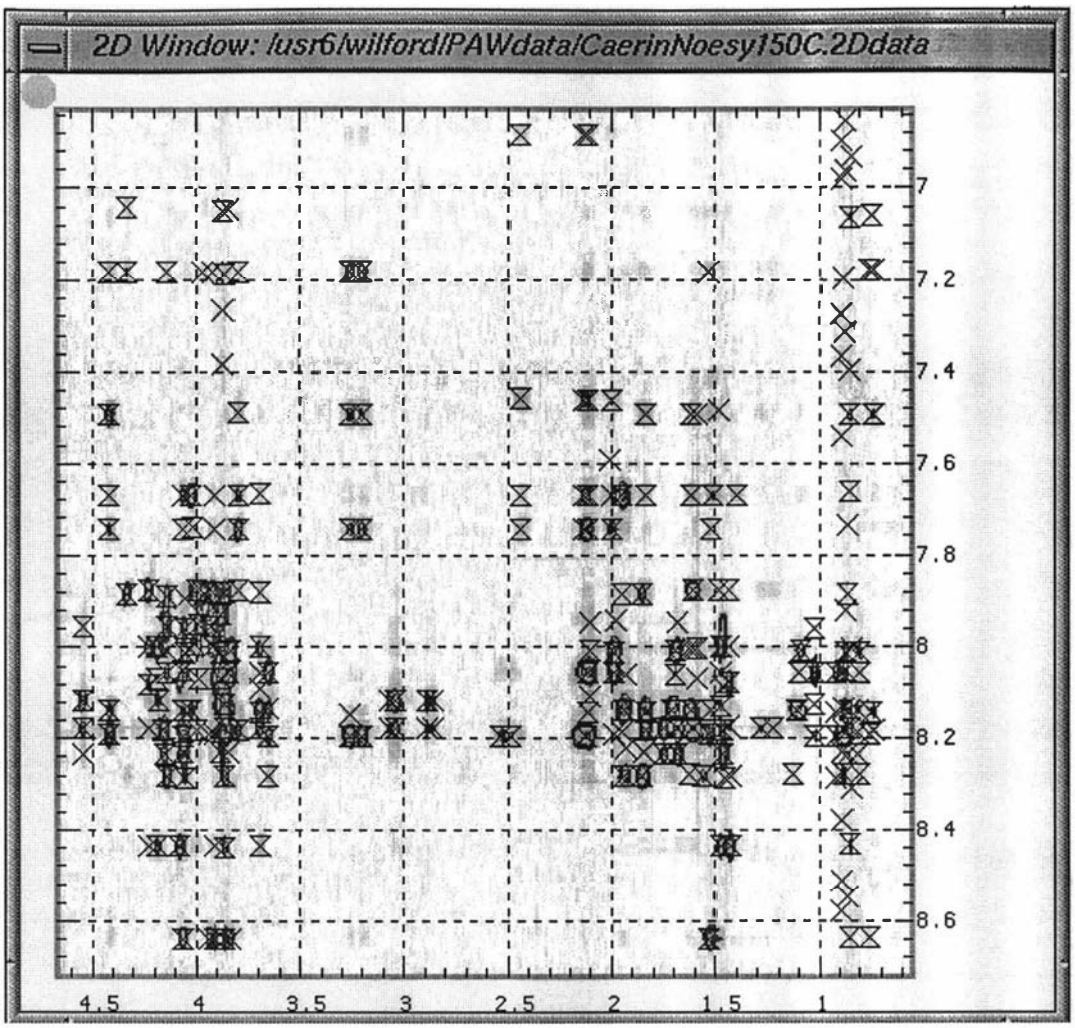


Figure 9.14: The cross-peaks in the lower-right region of the **CaerinNoesy150C** spectrum.

9.4.3 Assessing missing cross-peaks

A number of unpaired raw peaks were not picked by the automatic peak-picking operations. Some of them were real cross-peaks, and some of them were noise. To help assess the existence of missing transposed peaks in a region, a number of transposed rectangles associated with the missing peaks were drawn. The corners of the rectangles indicate the approximate locations of the possible transposed peaks, as shown in the next two figures.

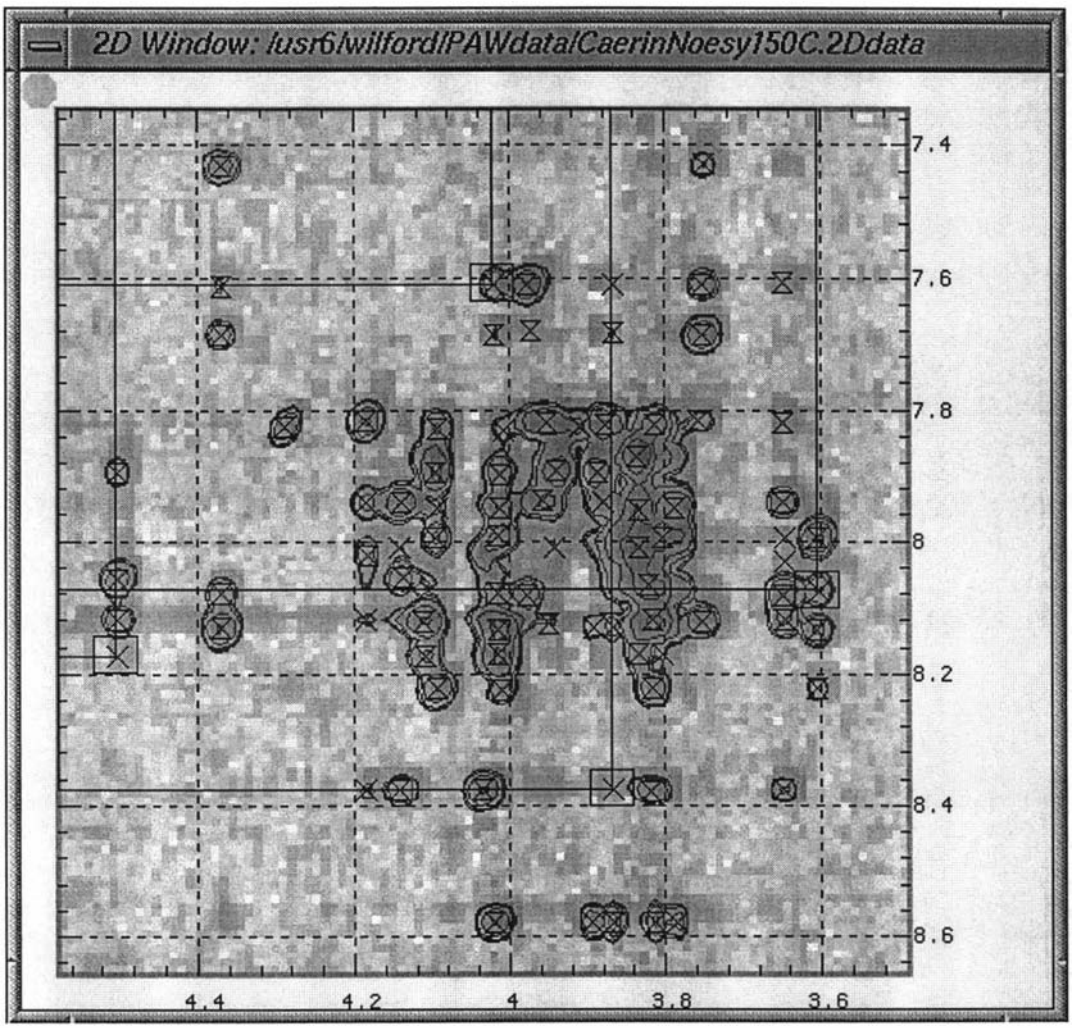


Figure 9.15: A smaller region with four transposed rectangles drawn for the assessment of possible missing transposed peaks.

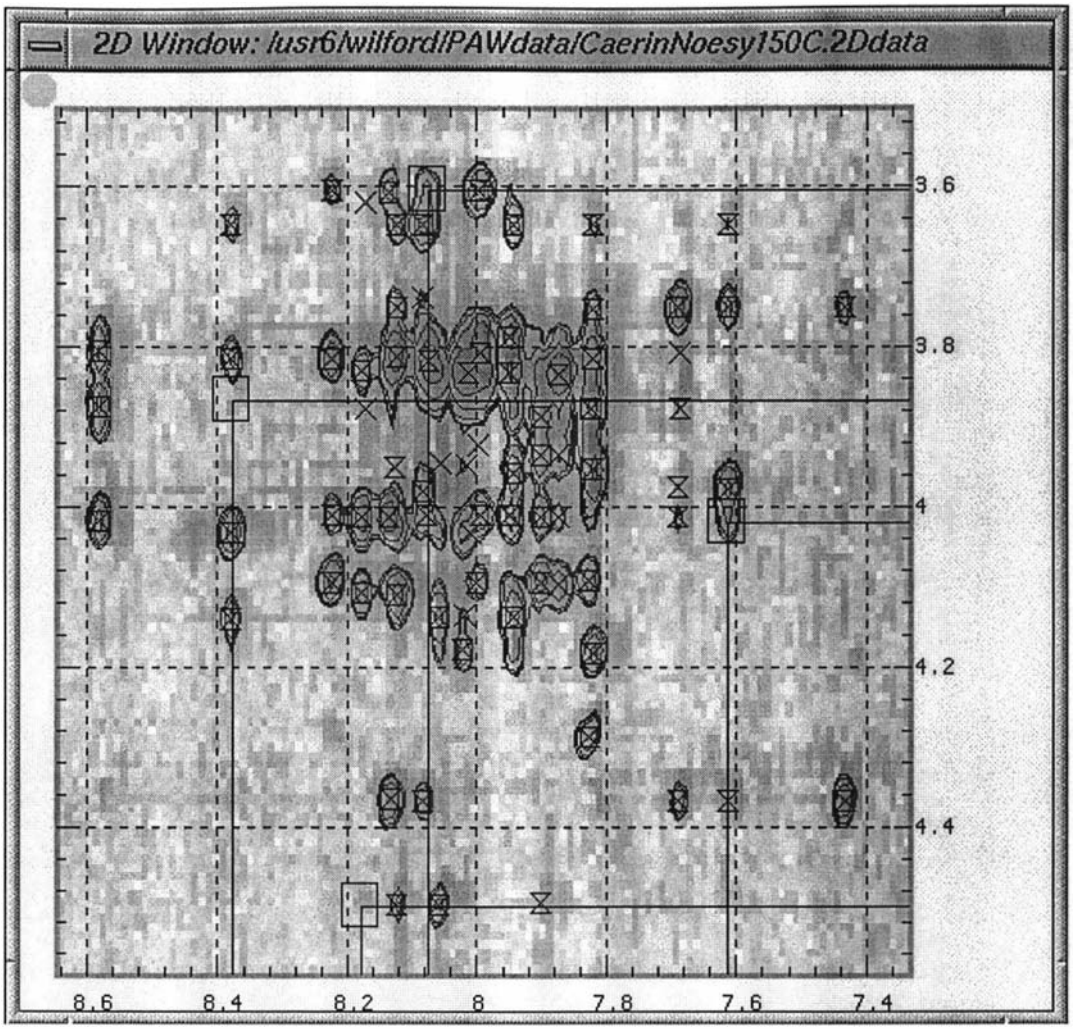


Figure 9.16: The transposed region of the last plot.

Often, additional evidence was obtained by looking at a row or column of data passing through the transposed rectangle corners. For example, the next 1D plot confirmed the existence of a cross peak within the small rectangle at the bottom of Figure 9.16. In contrast, the column at around 8.35 ppm (see Figure 9.16) did not show a significant peak (Figure 9.18).

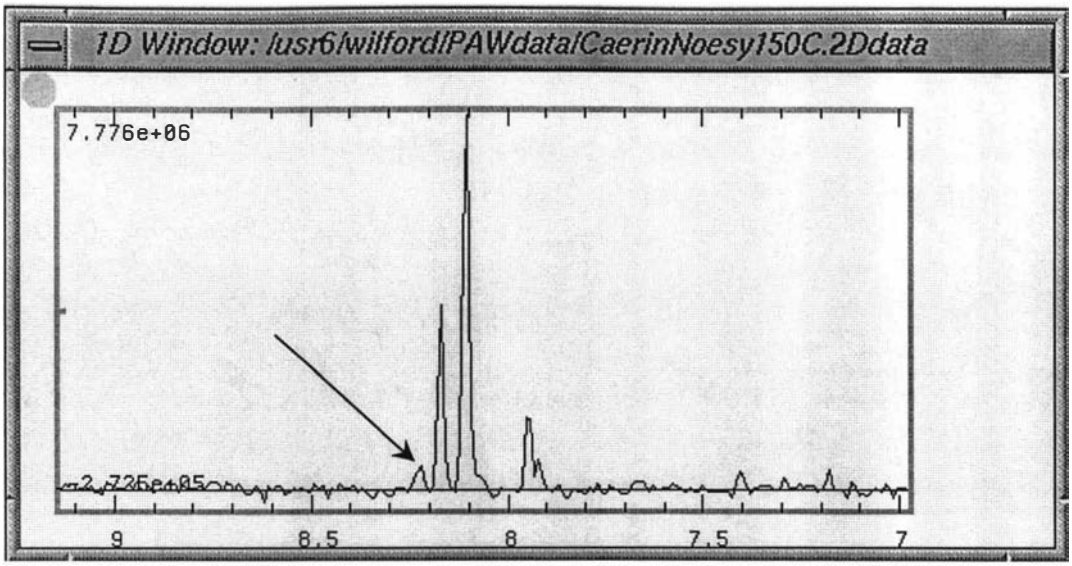


Figure 9.17: An expanded plot of a row drawn from Figure 9.16 at around 4.56 ppm. The arrow points to a peak that is inside the rectangle at the bottom of Figure 9.16.

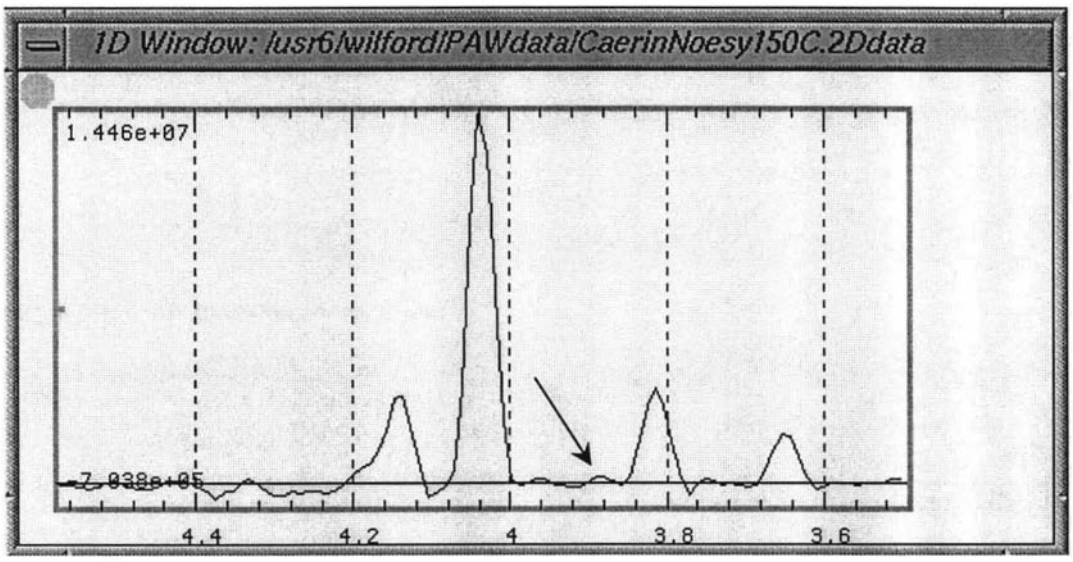


Figure 9.18: An expanded plot of a column drawn from Figure 9.16 at around 8.35 ppm. The arrow points to the insignificant peak that is inside the rectangle on the left of Figure 9.16.

9.4.4 Adding missing cross-peaks

In general, there are at least two circumstances when cross-peaks need to be added:

1. when a transposed raw peak is missing due to a reduced resolution in one dimension, and
2. when both are missing due to heavy overlap.

Note that, a transposed cross-peak pair can be picked from one or two existing raw peaks. If both peaks are missing, two operations are required: picking a raw peak, and then finding a transposed cross-peak. Otherwise, only the latter operation is needed.

The missing cross-peaks that were not picked in the initial stage have been frequently added whenever there was strong evidence for their existence. This has been done before and during the spectral assignment. Figure 9.19 shows three cross-peaks that were added.

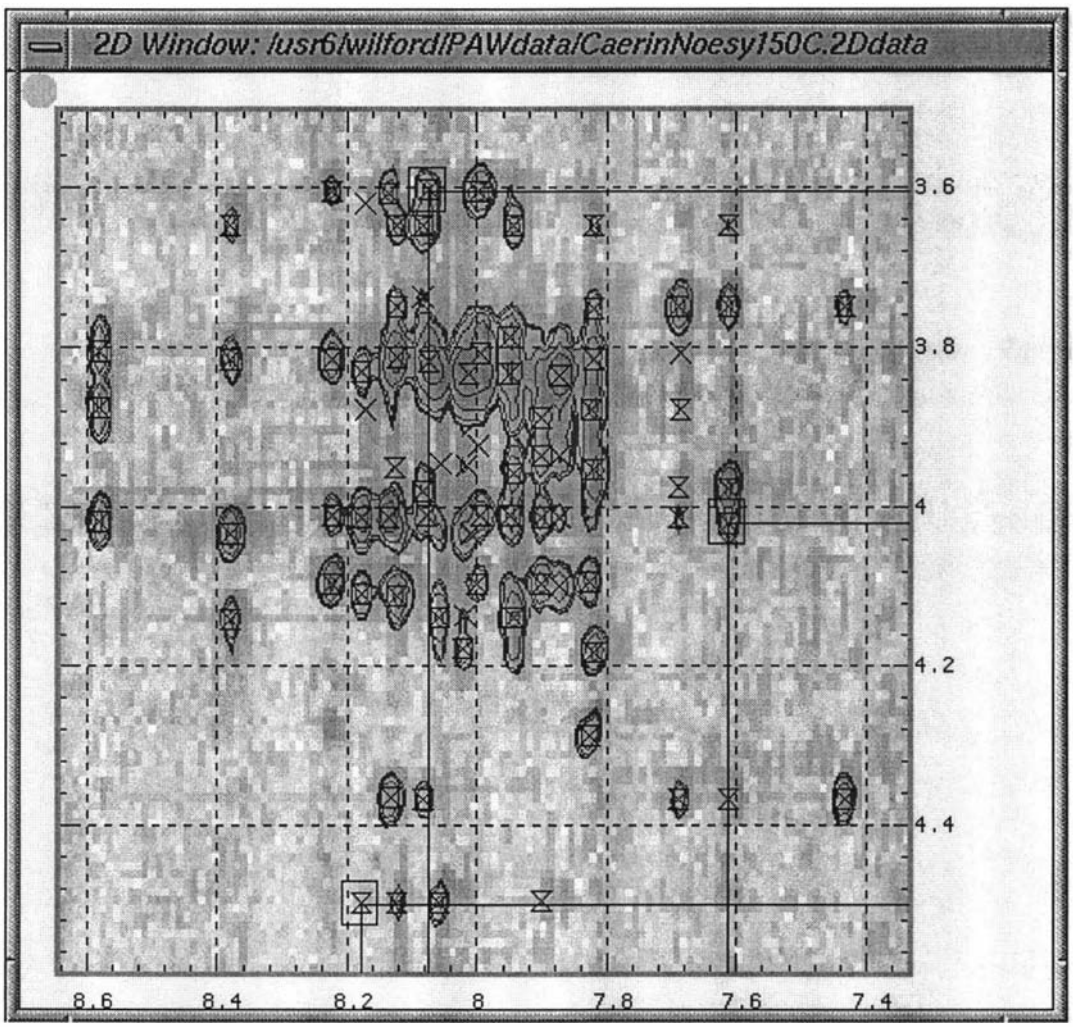


Figure 9.19: An example of three cross-peaks that were found and added.

A few examples on finding cross-peaks using transposed raw peaks can be found in Chapter 9 of Volume II. More examples on finding precise positions of missing cross-peaks can also be found in Chapter 11 of Volume II.

Note that transposed cross-peaks can be absent for a number of reasons, including the significantly lower resolution in D2, improperly phased data in one dimension, improperly picked diagonal peaks, heavily overlapped peaks, and local noise. The time that is required for refining a cross-peak list depends on the number of transposed-peaks that are missing or not properly picked. This is usually the most difficult and time-consuming part of the entire picking process because each case must be judged individually. A better option is perhaps to pick only unambiguous ones at

the start, and then consider the ambiguous ones at a later stage of the spectral assignment.

Chapter 10:

Spin-system Identification for Caerin 4.1 NMR Spectra Using PAW

10.1 Introduction

The identification of the spin systems in Caerin 4.1 was achieved by displaying NOESY peak symbols onto multi-region plots of the **CaerinTocsy070C** spectrum. This permits most of the inter-residue cross-peaks be clearly separated from those of intra-residues, because the background **CaerinTocsy070C** spectrum contains only intra-residue cross-peaks. Figure 10.1 shows the NOESY peak symbols displayed on a plot that contains both the upper-left and upper-right regions of the spectrum.

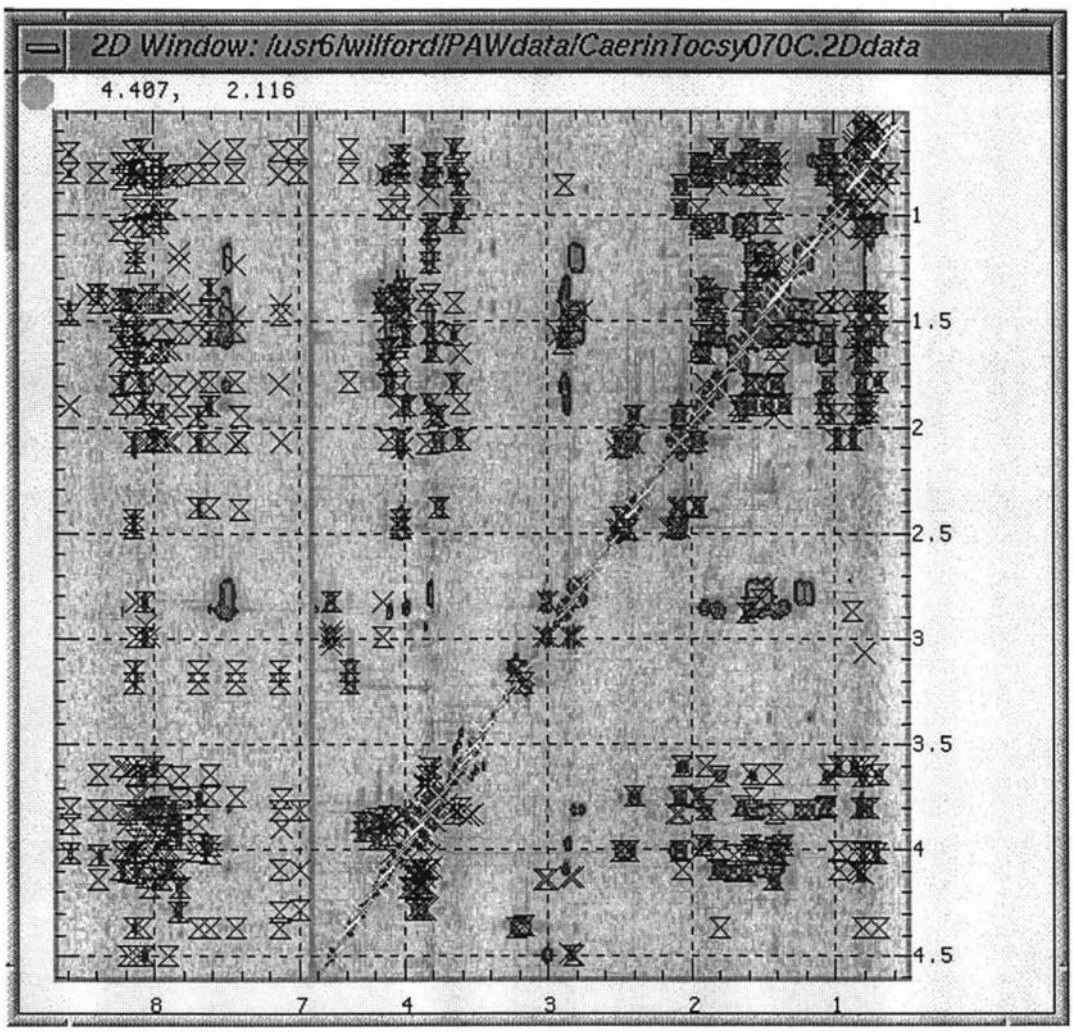


Figure 10.1: A multi-region plot of the **CaerinTocsy070C** spectrum, on which the NOESY peak symbols were plotted.

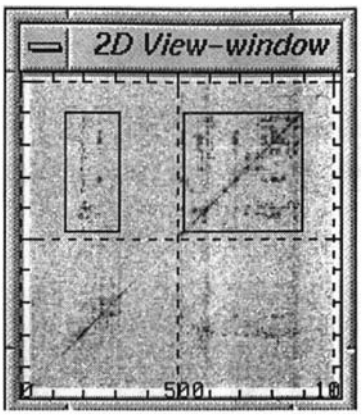


Figure 10.2: A 2D-view window for the previous plot.

To identify the cross-peaks that belong to a spin system, crosshairs were often drawn from peak centres to find the connections between cross-peaks. In addition, Tables 7.1, 7.2, 7.4 and Appendix 7.b have been frequently referenced throughout the spin-system identification process. These references give the primary-structures, spin-systems, chemical-shifts and 2D cross-peak distributions of the 20 common amino-acids.

Details of the spin-system identification operations can be found in Chapter 10 of Volume II.

10.2 Identification of the Ala residues

An Ala residue in ^{1}H TOCSY spectra is characterised by a 2D cross-peak distribution pattern shown in the next diagram:

Alanine — $C^{\beta}H_3$

Spin system:
[N-α-β_3]

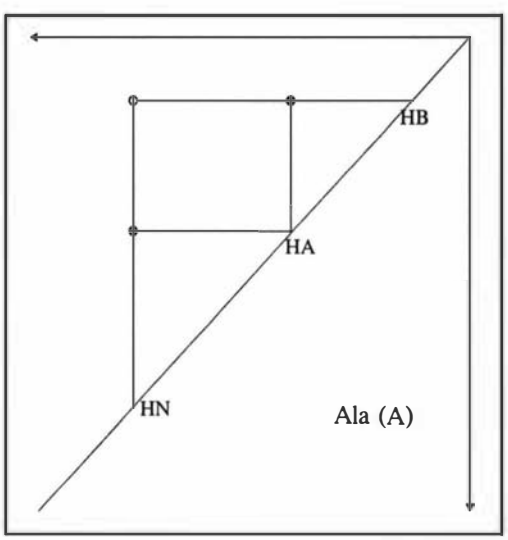


Figure 10.3: The ^{1}H TOCSY spin-system and connectivity-diagram of Ala.

Note that this thesis uses the solid circles in the connectivity diagrams to represent neighbouring-proton connections like those in a ^{1}H COSY spectrum, while the open cycles represent other ^{1}H TOCSY connections.

The three Ala-residue spin-systems were identified by the unique distribution of their HN:HB and HA:HB cross-peaks in the upper regions of the **CaerinTocsy070C** spectrum, as shown in the following three figures.

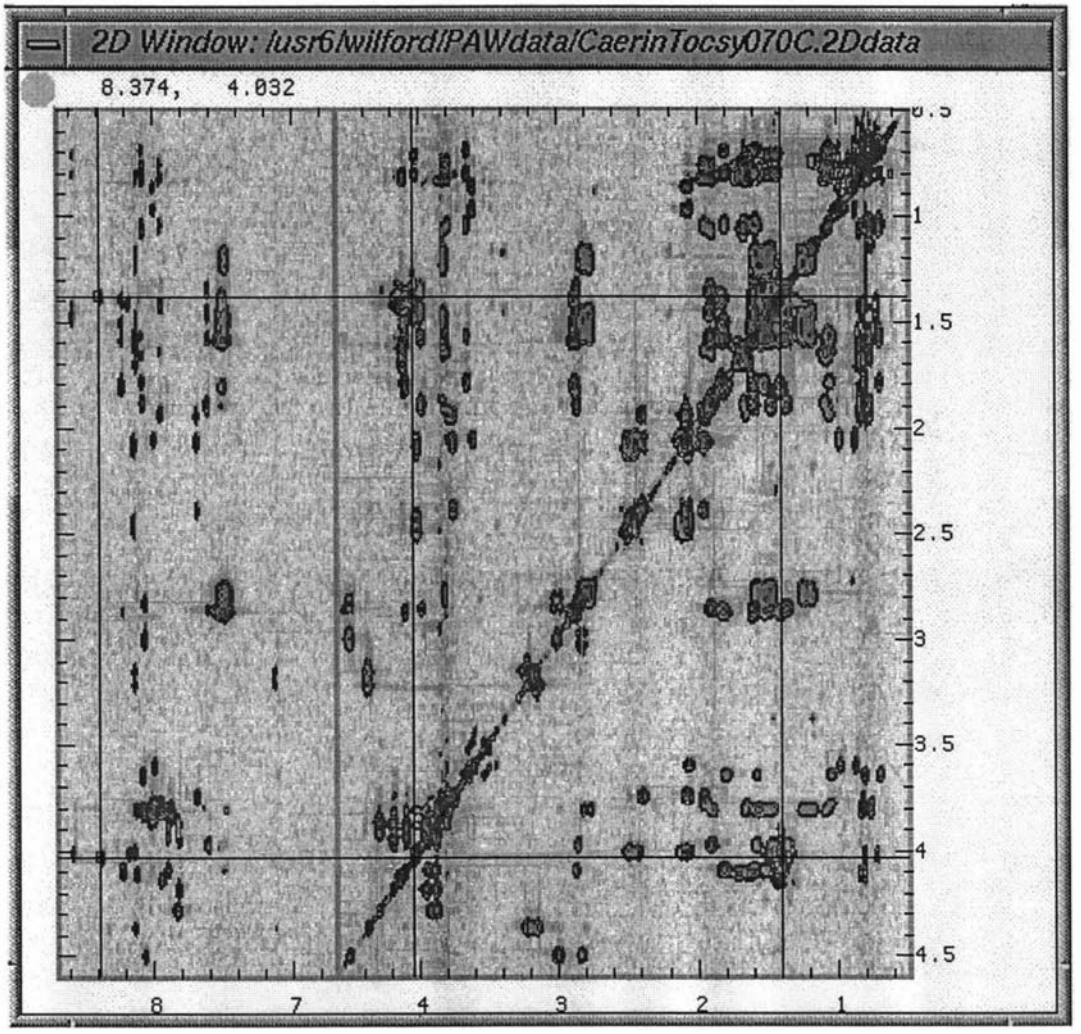


Figure 10.4: A multi-region plot that shows how one of the Ala spin-systems in Caerin 4.1 was identified.

Note that the cross-peaks of a spin system must be located at the intersections of the lines. This applies to all figures in this chapter that show the connectivity of spin systems.

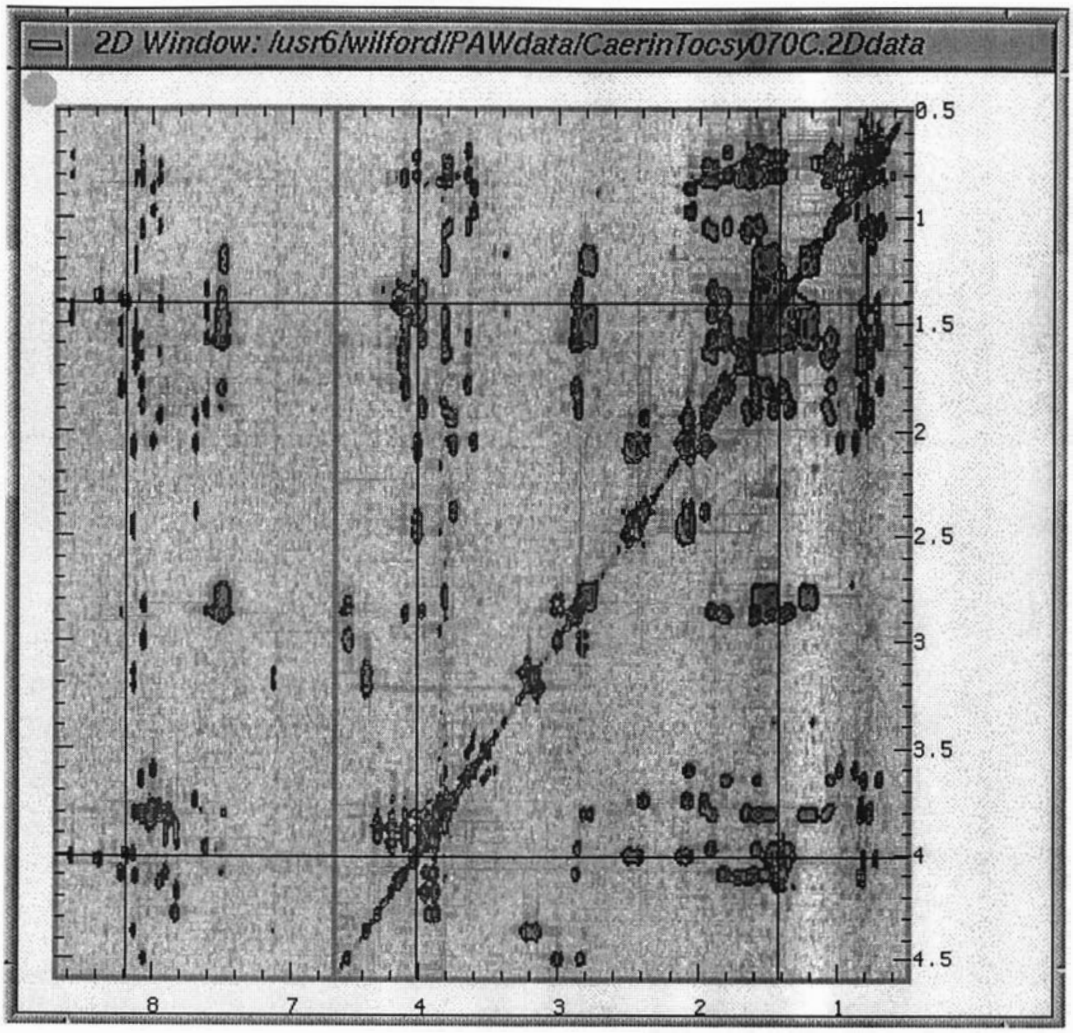


Figure 10.5: A multi-region plot that shows how the second Ala spin-system in Caerin 4.1 was identified.

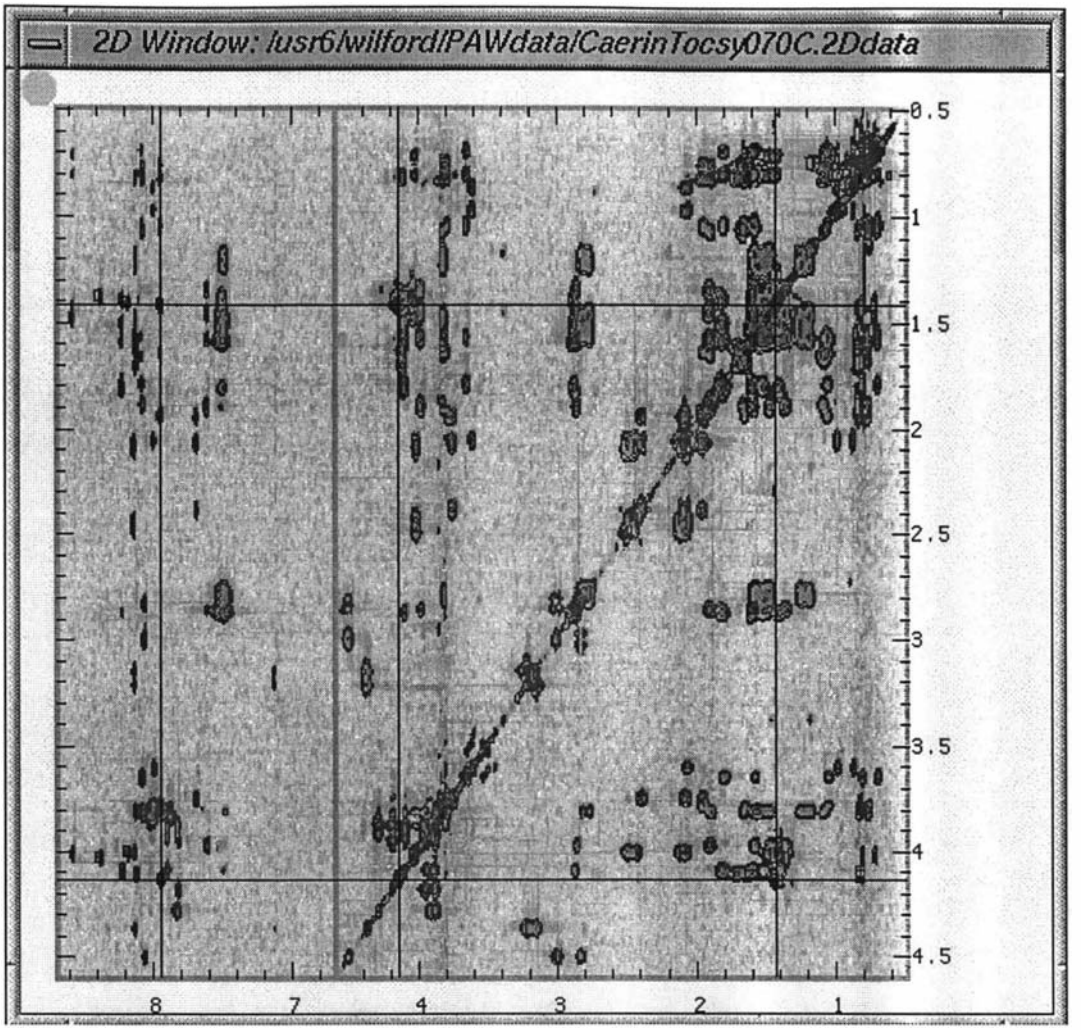


Figure 10.6: A multi-region plot that shows how the third Ala spin-system in Caerin 4.1 was identified.

Assignment of the third Alanine was obvious after zooming into a much smaller region and comparing the 1D peak-shapes with five other intra-residue TOCSY peaks (Figure 10.7).

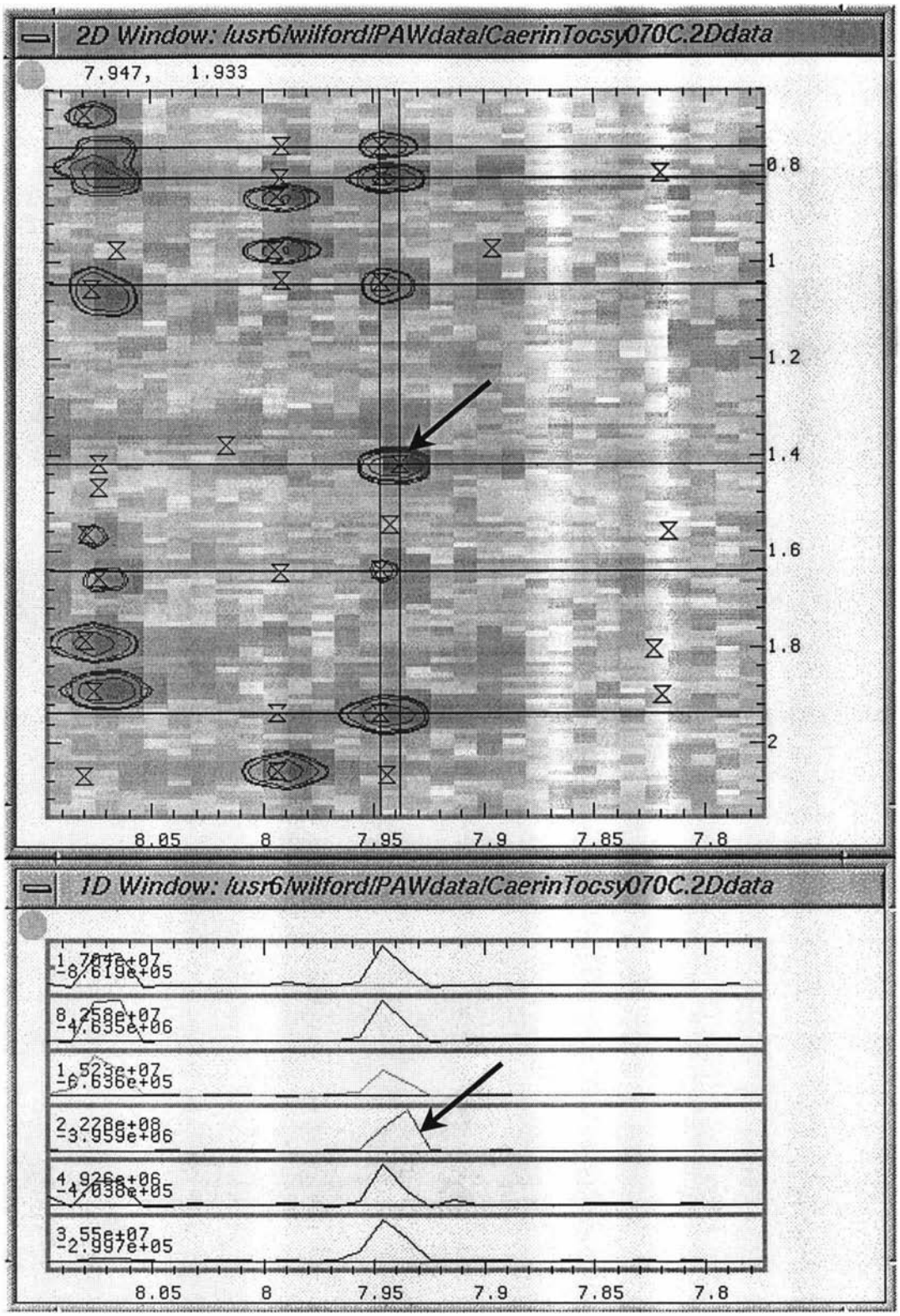


Figure 10.7: A expanded 2D and 1D view that shows clearly the HB cross-peak of an Ala residue, as arrowed in the plots. The six 1D plots show the six rows indicated by the horizontal lines in the 2D plot.

10.3 Identification of the Ser residues

A Ser residue in ^{1}H TOCSY spectra is characterised by the 2D cross-peak distribution pattern shown in the next diagram:

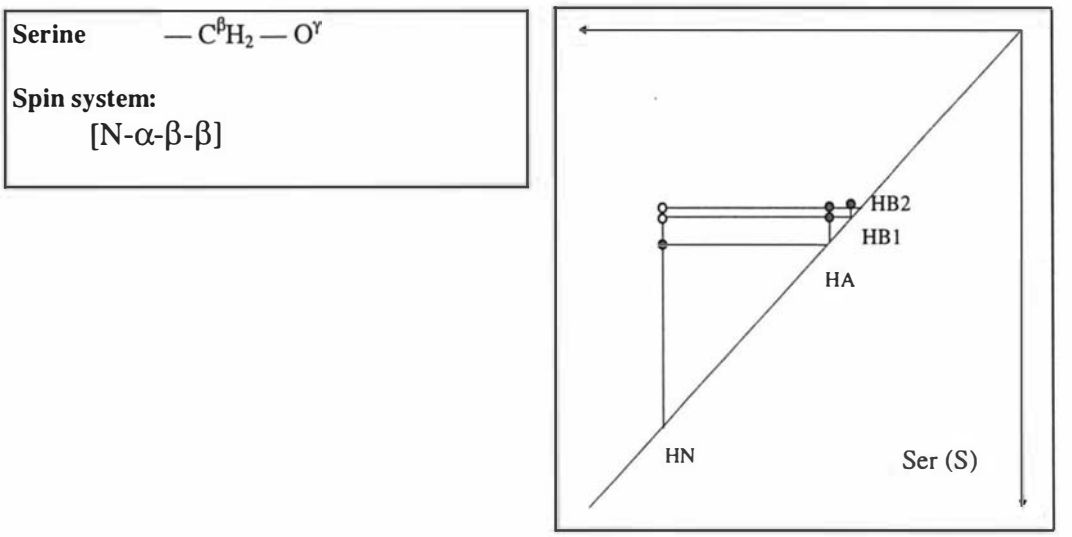


Figure 10.8: The ^{1}H TOCSY spin-system and connectivity-diagram of a Ser residue.

The Ser-related cross-peaks were identified by zooming into another multi-region pattern, as shown in Figure 10.9.

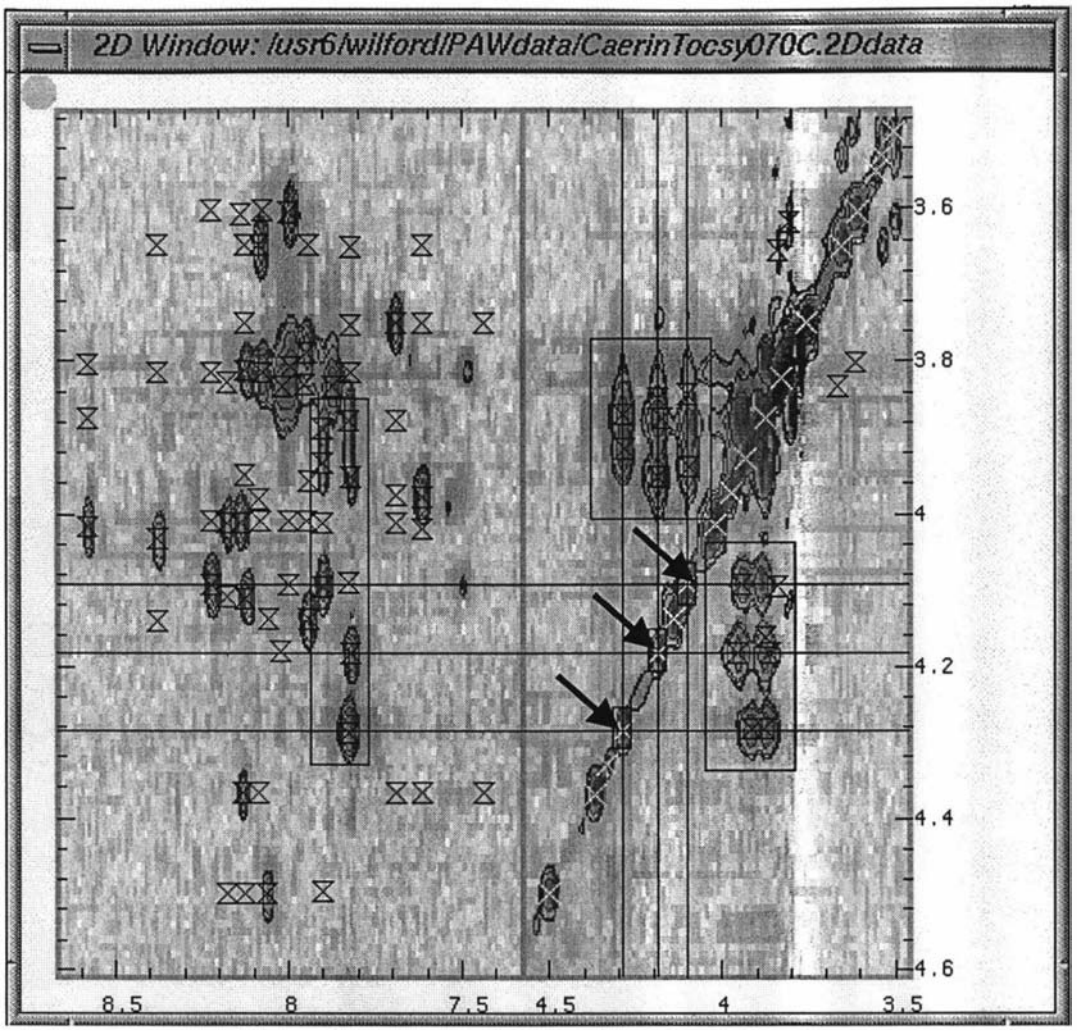


Figure 10.9: An expanded view for the identification of Ser spin systems. Three sets of crosshairs were drawn from the three diagonal peaks (arrowed). They are associated with the twelve HA-HB and HB-HA cross-peaks inside the two rectangles on the right. As a consequence, the Serine-related cross-peaks in the HN-HA region were located (see the rectangle on the left). Note that the peaks were picked from a NOESY spectrum, and some adjustments of the cross-peak locations were performed at a later stage.

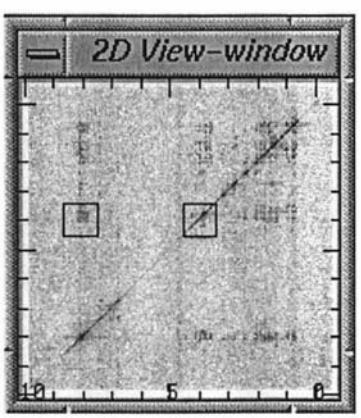


Figure 10.10: The overview window for the multi-region plot in the last figures.

10.4 Identification of the Lys, Asp and Trp residues

The Lys, Asp and Trp residues in ^{1}H TOCSY spectra are characterised by the 2D cross-peak distribution patterns shown in the next three diagrams:

Lysine

$$— C^{\beta}H_2 — C^{\gamma}H_2 — C^{\delta}H_2 — C^{\varepsilon}H_2 — {}_{\varepsilon}N^{\zeta}H_3^{\oplus}$$

Spin system:

$[N\text{-}(\zeta_3)\text{-}\alpha\text{-}\varepsilon_2\text{-}\beta\text{-}\beta\text{-}\delta_2\text{-}\gamma_2]$

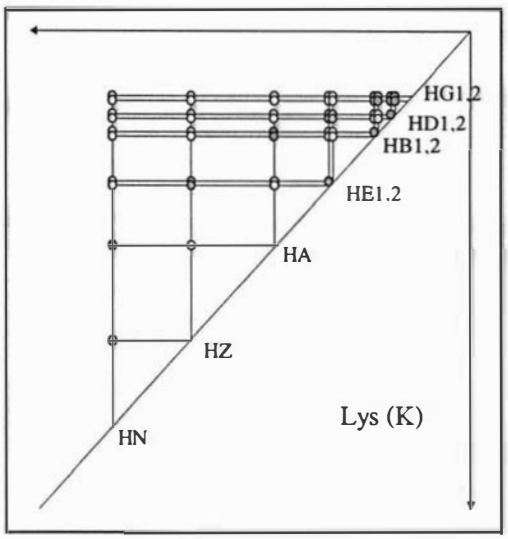


Figure 10.11: The ^{1}H TOCSY spin-system and connectivity-diagram of a Lys residue.

Aspartic acid

$$— C^{\beta}H_2 — C^{\gamma}(=O^{\delta 1}) — O^{\delta 2}H$$

Spin system:

$[N\text{-}\alpha\text{-}\beta\text{-}\beta]$

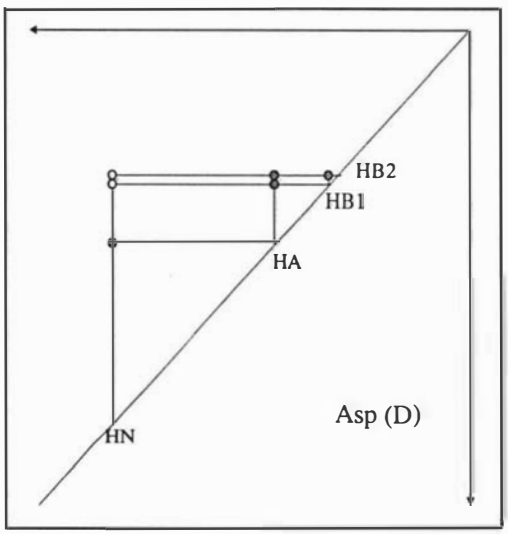


Figure 10.12: The ^{1}H TOCSY spin-system and connectivity-diagram of a Asp residue.

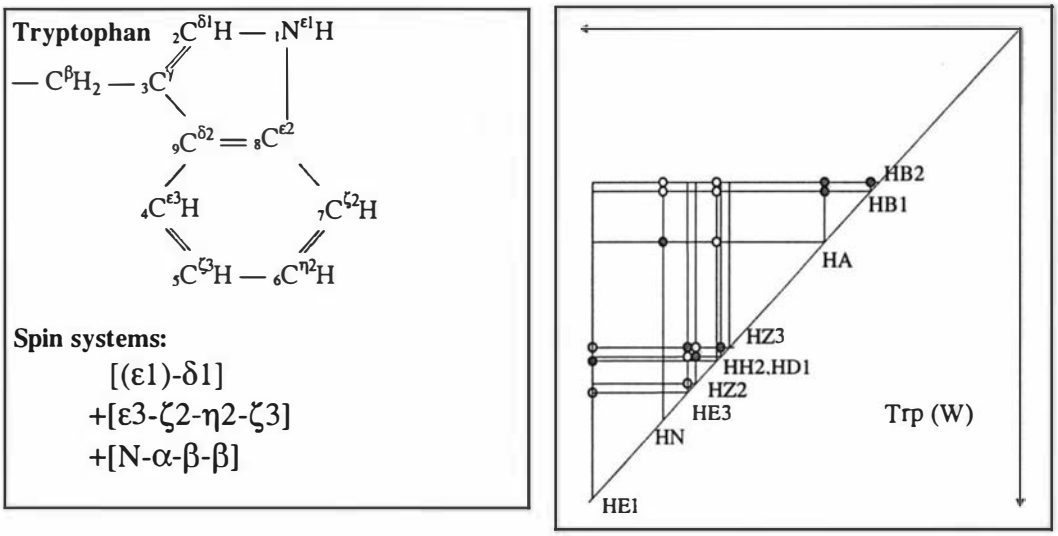


Figure 10.13: The ^{1}H TOCSY spin-system and connectivity-diagram of a Trp residue.

The HE-related cross-peaks for the Lys residues and the HB-related cross-peaks for the Asp and Trp residues were found within the two rectangles in Figure 10.14.

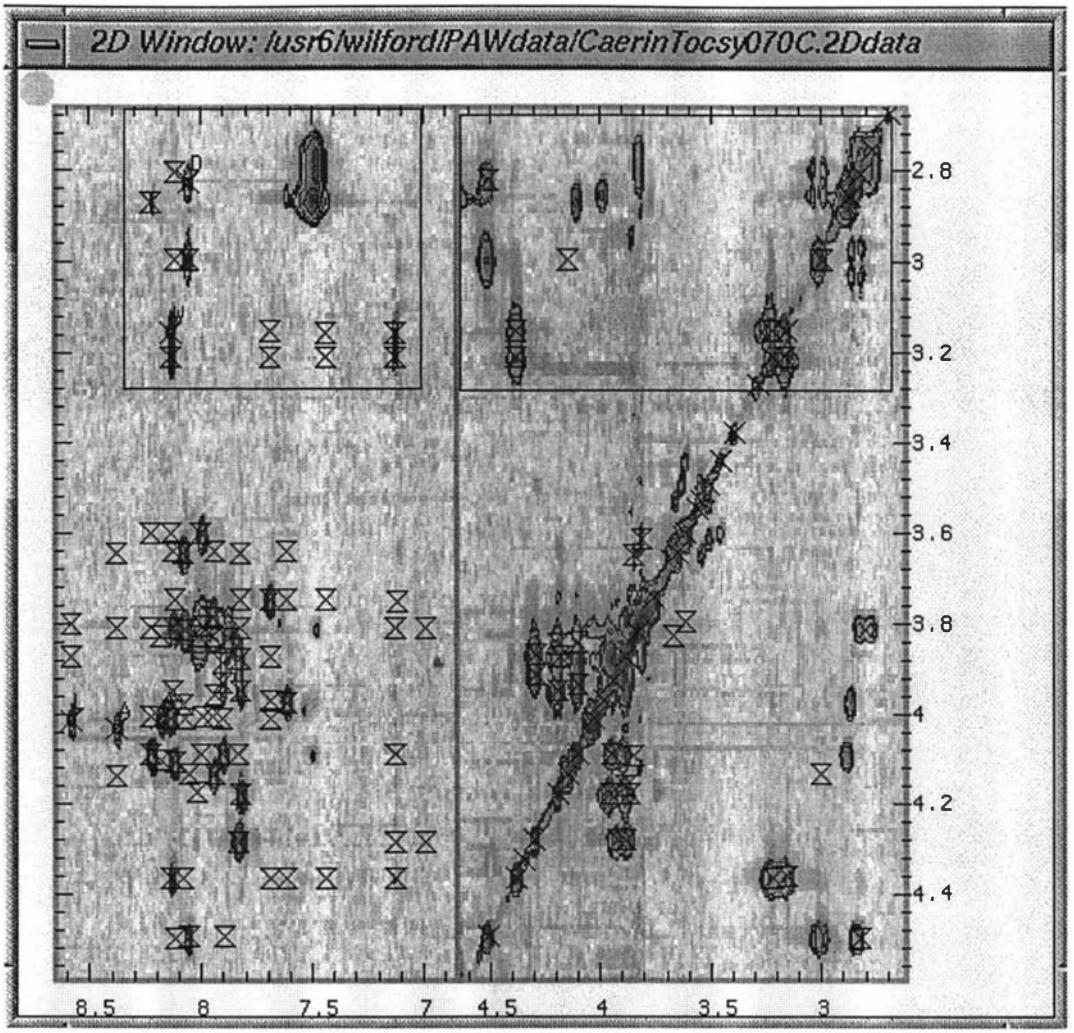


Figure 10.14: A multi-region plot that shows the HE-related cross-peaks for the Lys residues and the HB-related cross-peaks for the Asp and Trp residues. They are to be identified individually in the following discussion.

Among this group of residues, Trp-related cross-peaks are identified by their obvious connection to the aromatic cross-peaks, as shown in Figure 10.15.

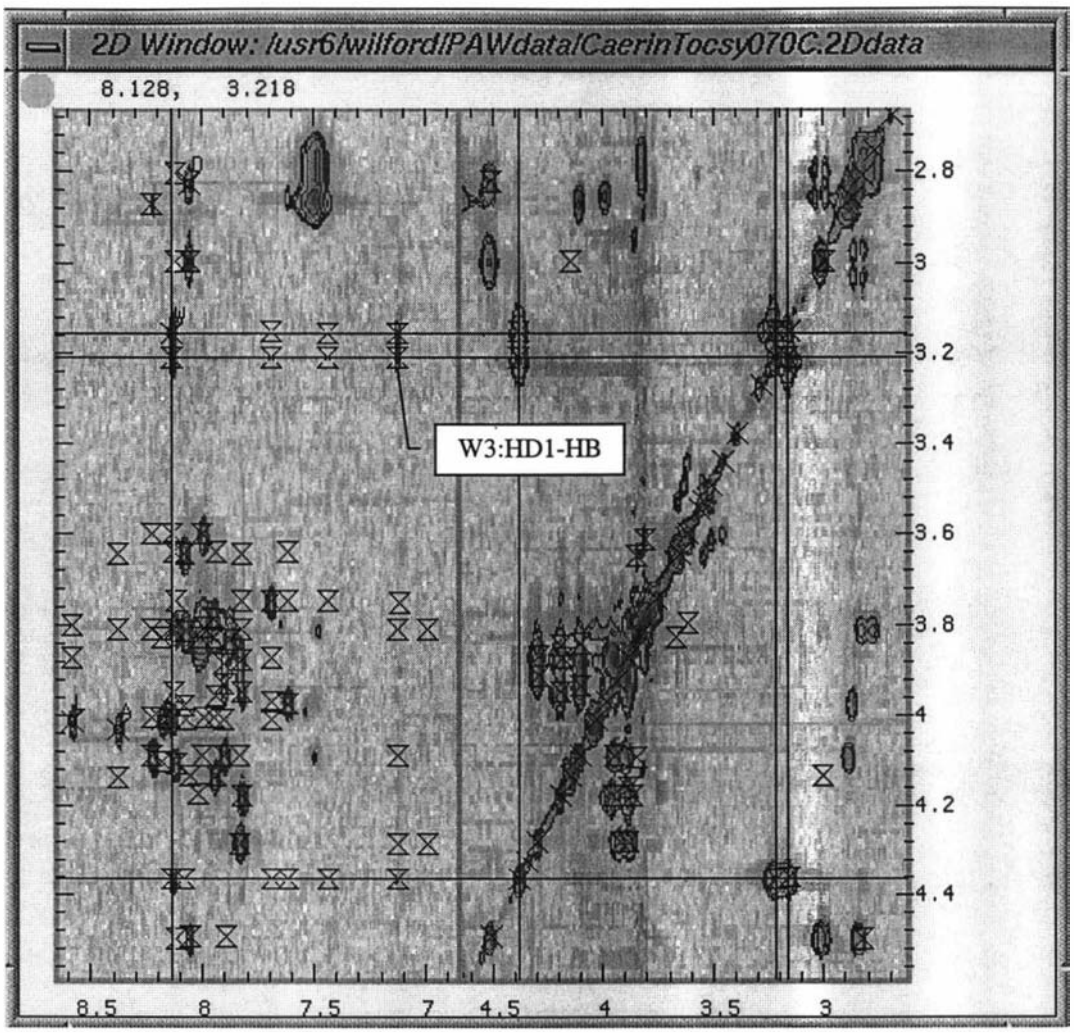


Figure 10.15: The same regions as the last plot. The crosshairs enable the unique Trp residue to be identified by the connection with the aromatic cross-peaks (such as W3:HD1-HB) in Tryptophan..

The other two intense cross-peaks at the upper-left corner defined the Asp-related cross-peaks, as shown in Figure 10.16.

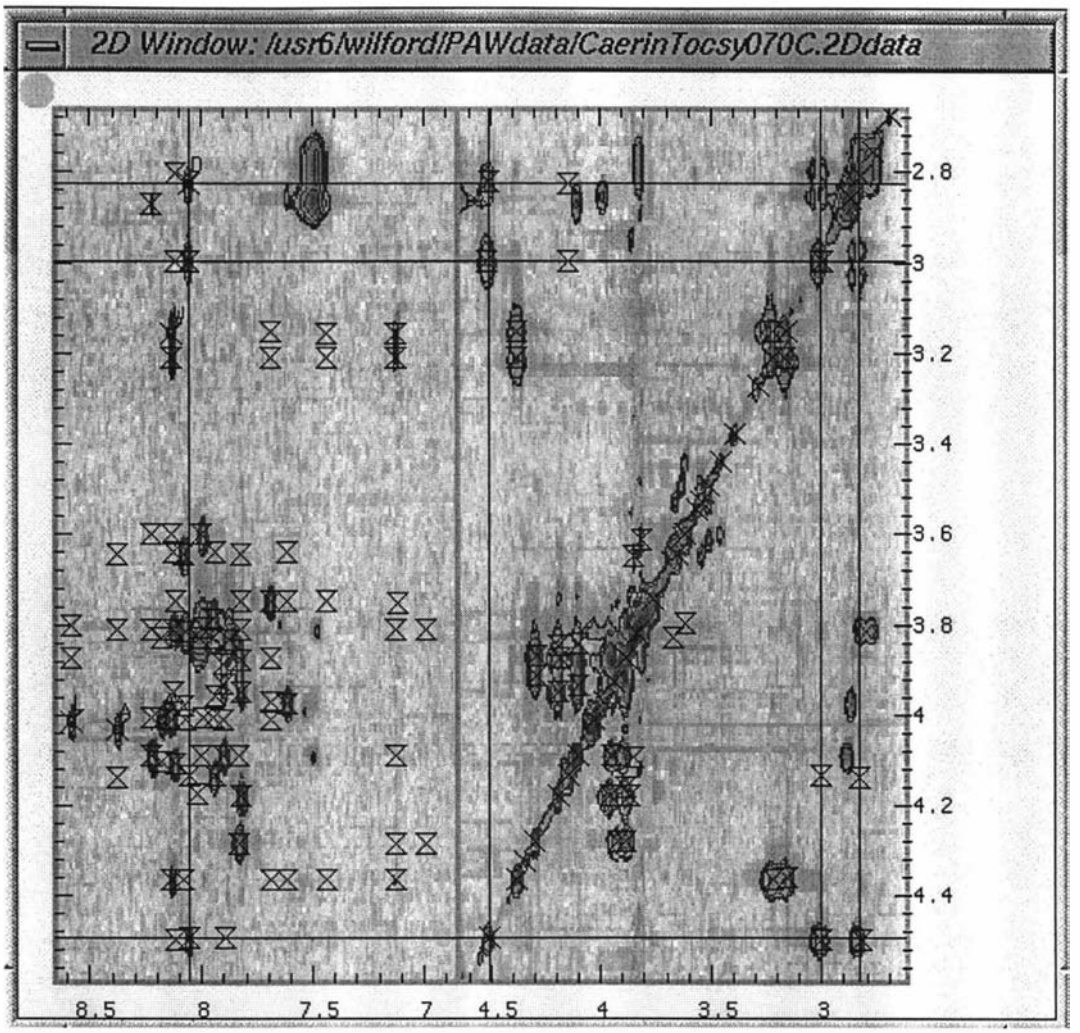


Figure 10.16: The same regions as the last plot. The crosshairs enable the unique Asp residue to be identified, because the Asp cross peaks are much more intense than the Lys HN-HE cross-peaks and they have no connection with any aromatic cross-peaks.

The remaining weak peaks in the upper area of the plot define the three Lys–related cross-peaks, as shown in Figures 10.17 to 10.19.

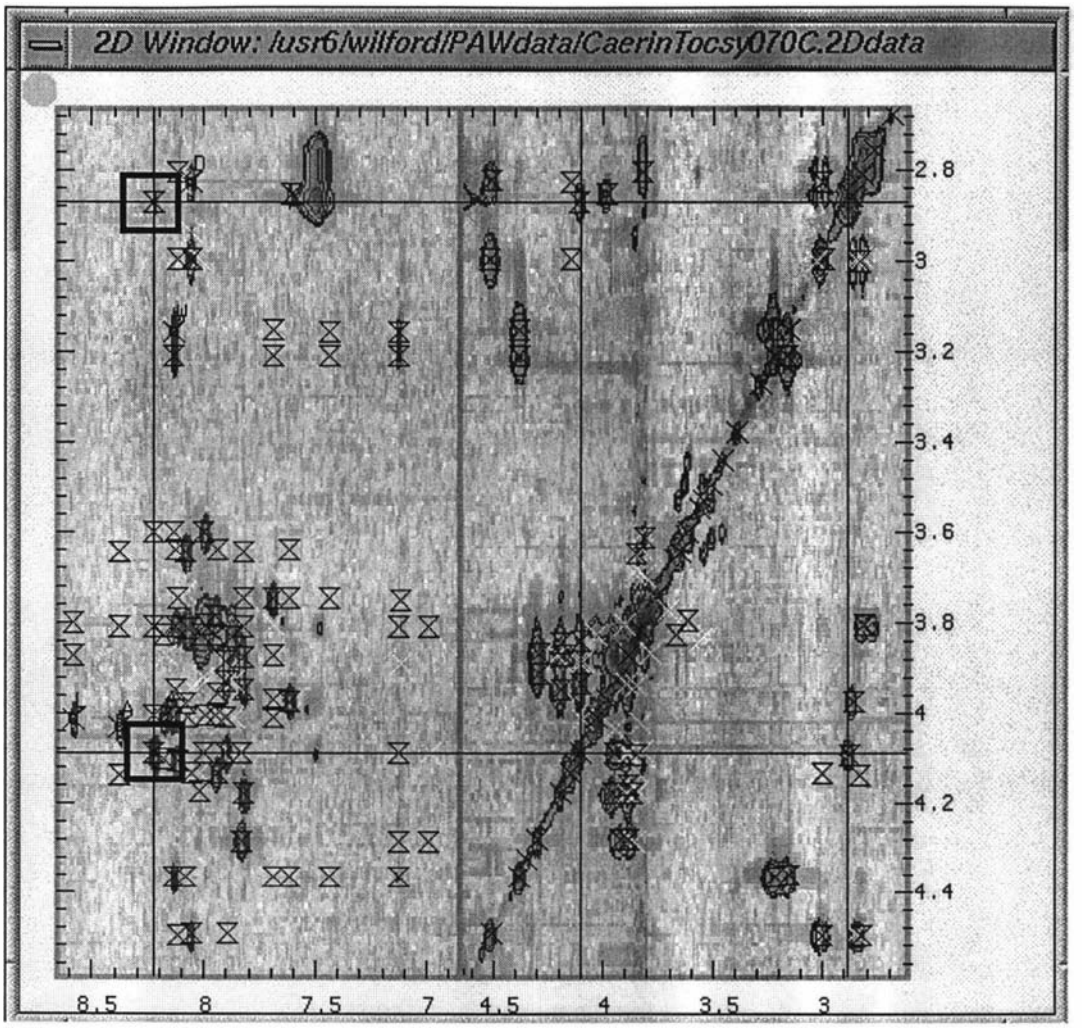


Figure 10.17: A multi-region plot that shows how one of the Lys spin-systems was identified. Some cross-peaks in the plots were picked manually, as described in Chapter 10 of Volume II. The cross-peaks in the bold rectangles correspond to the HN:HA and HN:HE cross-peaks in Figure 10.11.

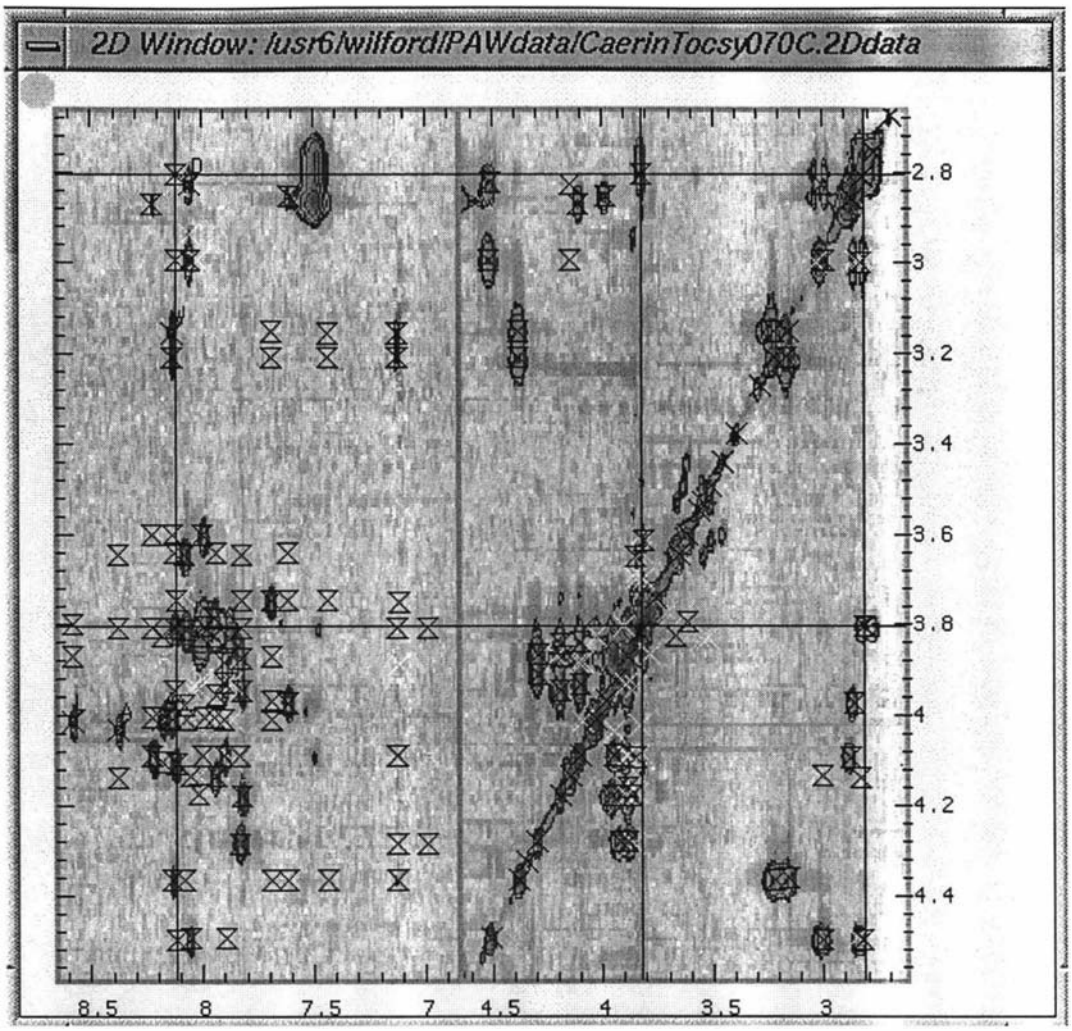


Figure 10.18: A multi-region plot that shows how the second Lys spin-system was identified.

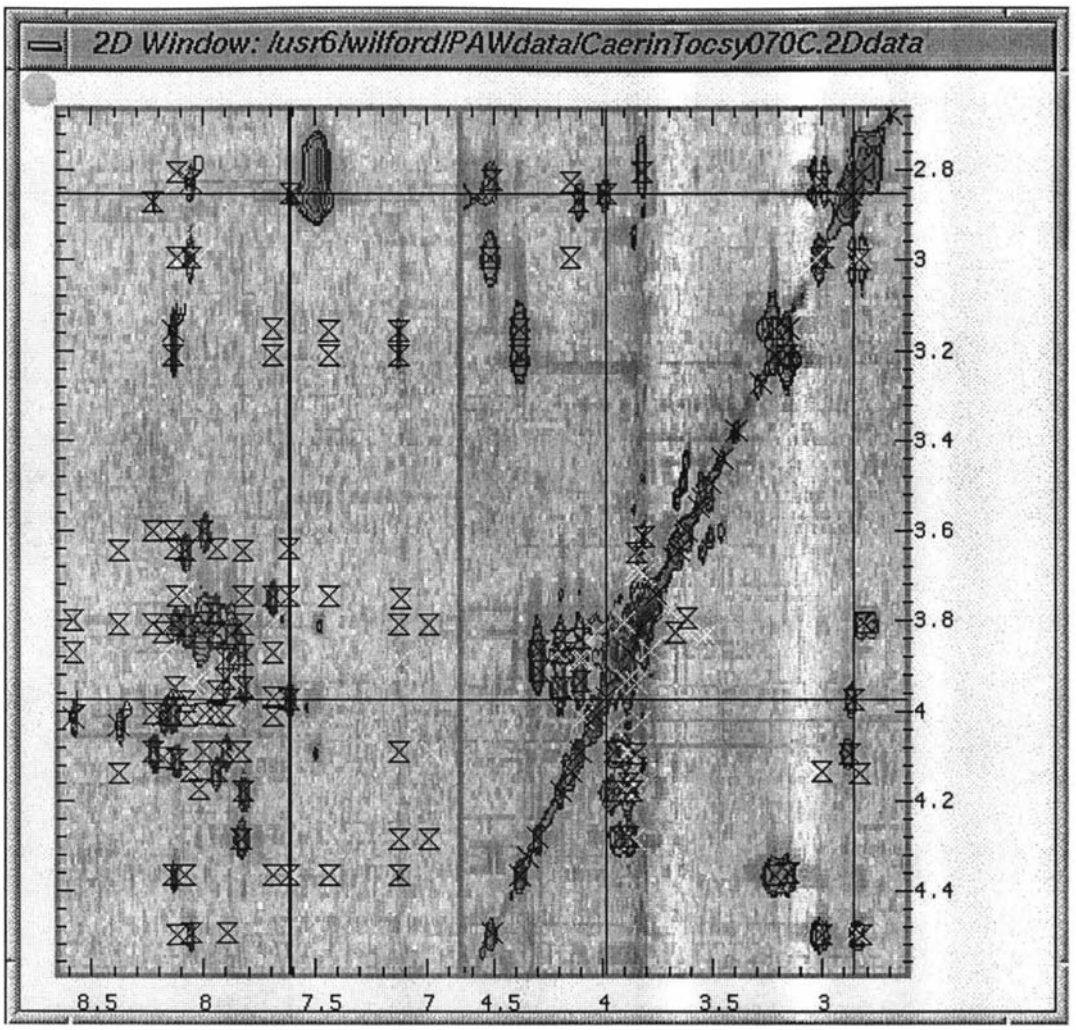


Figure 10.19: A multi-region plot that shows how the third Lys spin-system was identified. The first vertical bold line indicates an imperfect alignment between two cross-peaks used for drawing the cross-hairs.

Note that some cross-peaks in the plots were picked manually, as described in Chapter 10 of Volume II. The approximate locations of the manually picked cross-peaks were obtained by drawing crosshairs from some of the intense cross-peaks of the three spin-systems.

10.5 Identification of the Gly residues

A Gly residue in ^{1}H TOCSY spectra is characterised by a 2D cross-peak distribution pattern shown in the next diagram:

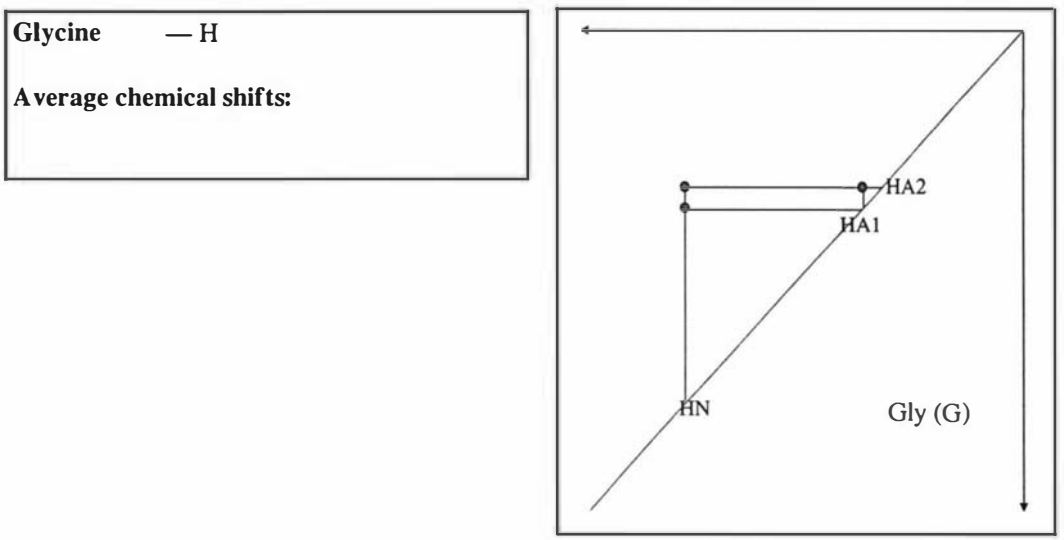


Figure 10.20: The ^{1}H TOCSY spin-system and connectivity-diagram of a Gly residue.

The HN-HA cross peaks of two Gly spin-systems were unambiguously identified, because there were no other HN-related cross-peaks connected to them in the upper area of the fingerprint region, as shown in the next figure. The HA of all Gly spin-systems, however, could not be resolved in all three spectra until a later stage in the spectral assignment, during which the other two Gly spin-systems were also identified. In addition, all of the HN-HA1 and HA2 cross-peaks are degenerated.

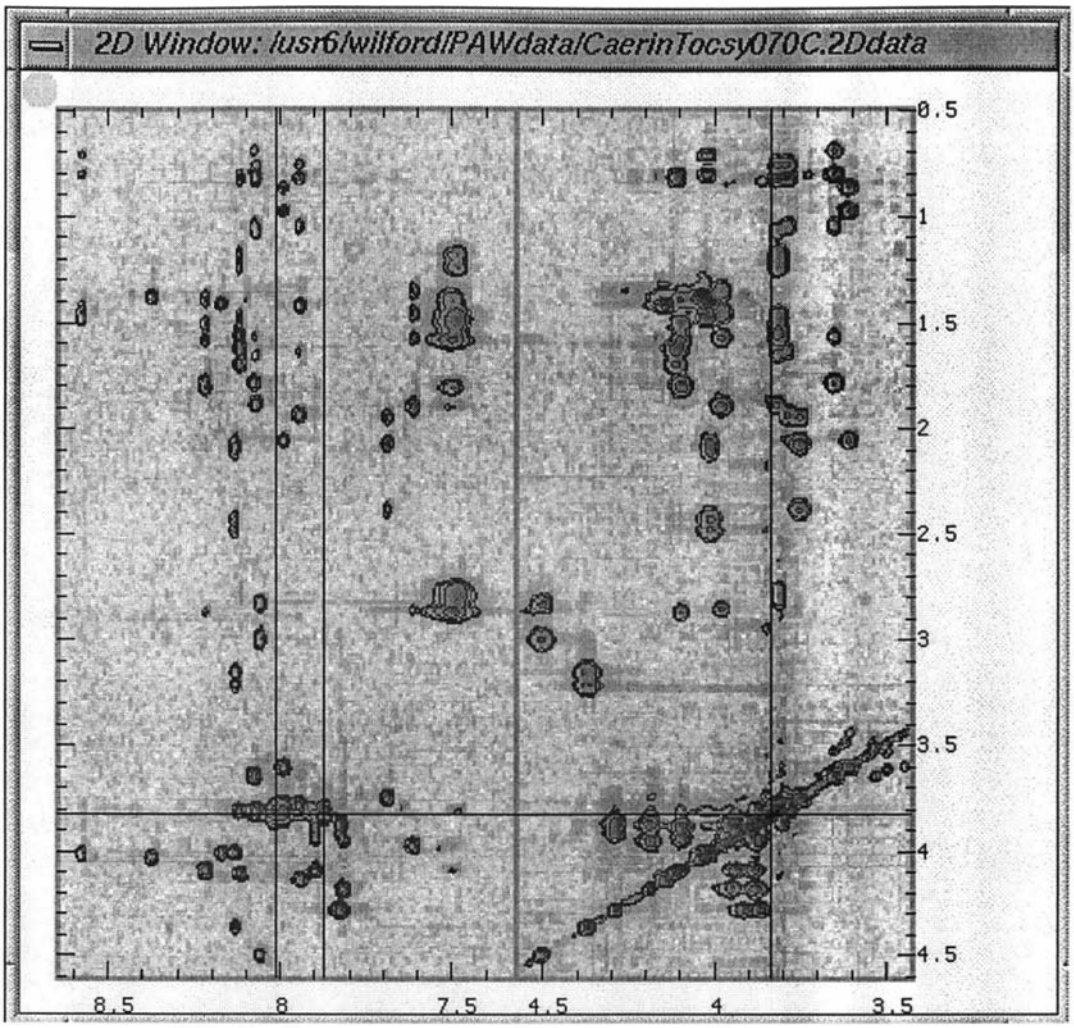
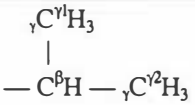


Figure 10.21: A multi-region plot that shows how two of the Gly spin-systems were identified.

10.6 Identification of the Val, Glu and Gln residues

The Val, Glu and Gln residues in ^{1}H TOCSY spectra are characterised by the 2D cross-peak distribution patterns shown in the next three diagrams:

Valine



Spin system:

[N-α-β-γ1$_3$-γ2$_3$]

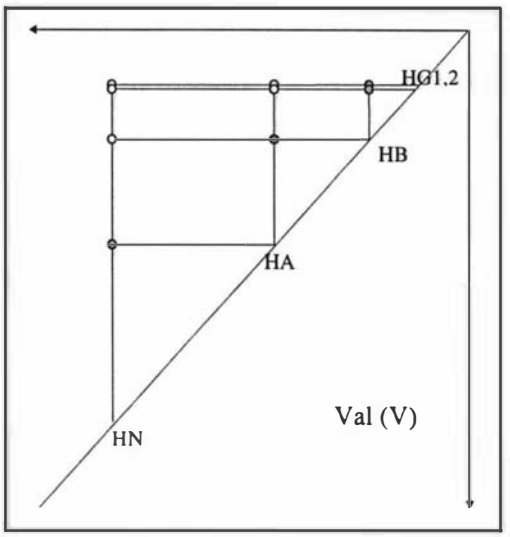
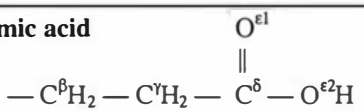


Figure 10.22: The ^{1}H TOCSY spin-system and connectivity-diagram of a Val residue.

Glutamic acid



Spin system:

[N-α-γ-γ-β-β]

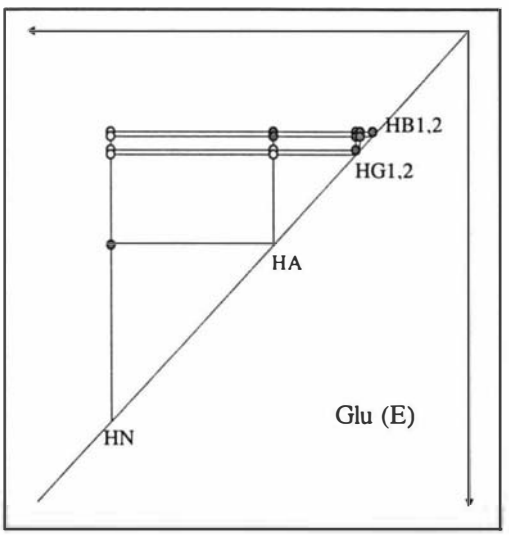
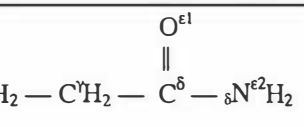


Figure 10.23: The ^{1}H TOCSY spin-system and connectivity-diagram of a Glu residue.

Glutamine



Spin system:

[N-(ε2-ε2-)α-γ-γ-β-β]

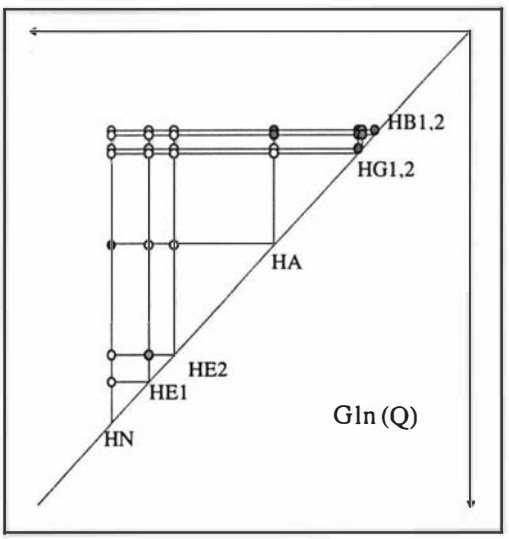


Figure 10.24: The ^{1}H TOCSY spin-system and connectivity-diagram of a Gln residue.

Each of these spin-systems only occurs once in the Caerin 4.1 sequence.

The Val spin-system was identified by its distinct peak-distribution pattern in the **CaerinTocsy070C** spectrum (Figure 10.25).

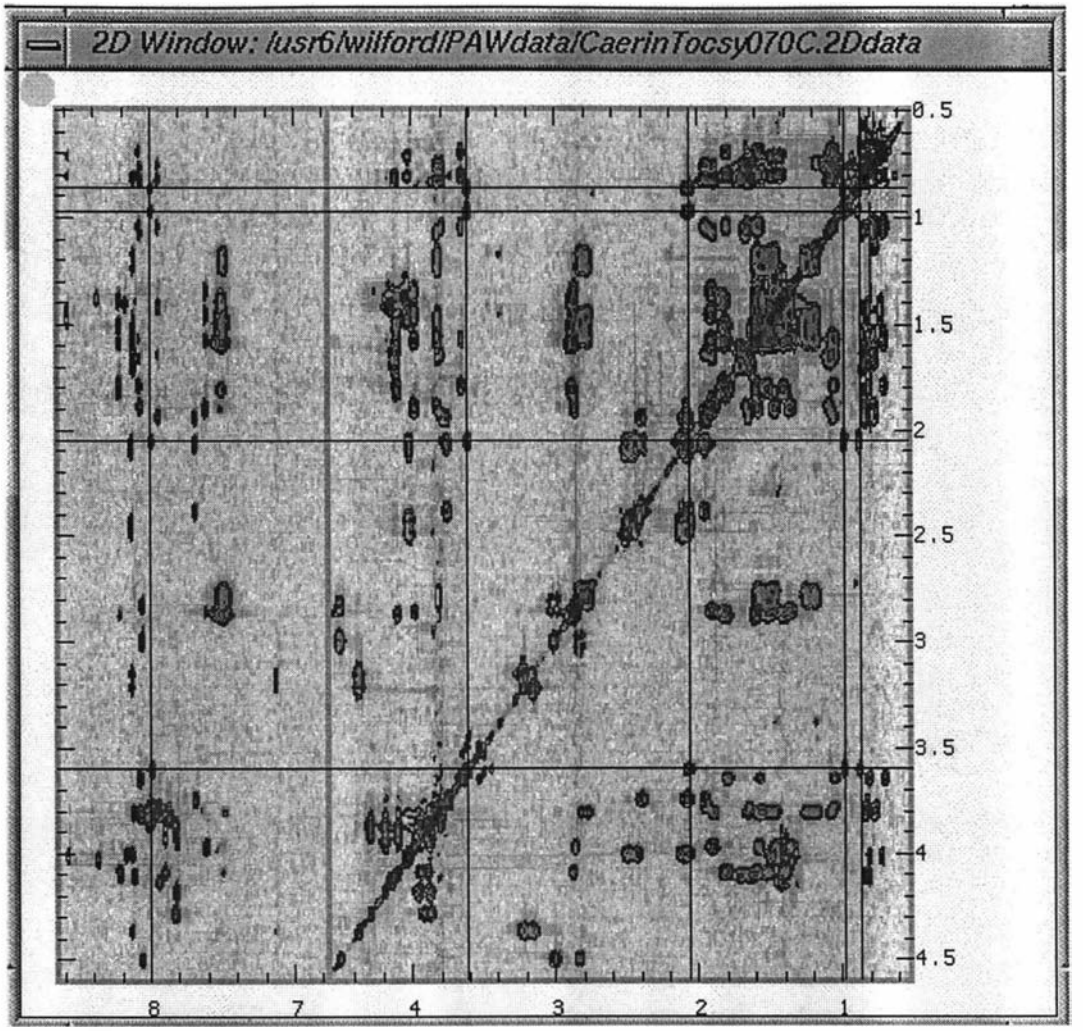


Figure 10.25: A multi-region plot that shows how the Val spin-system in Caerin 4.1 was identified.

The Gln system was identified by the unique peak patterns from the HE2 protons (Figure 10.26).

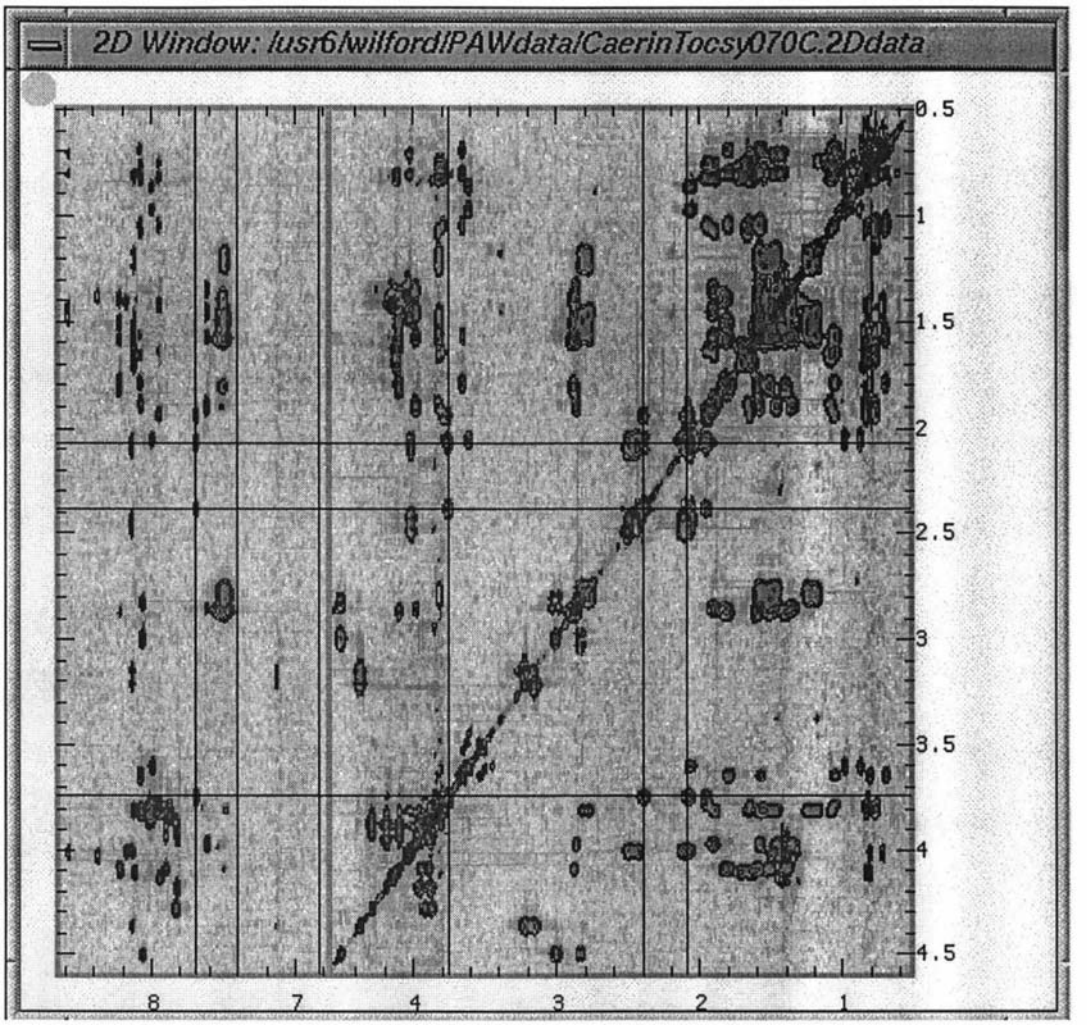


Figure 10.26: A multi-region plot that shows how the Gln spin-system in Caerin 4.1 was identified.

The Glu system was identified by its characteristic peak-distribution pattern (Figure 10.23 and 10.27).

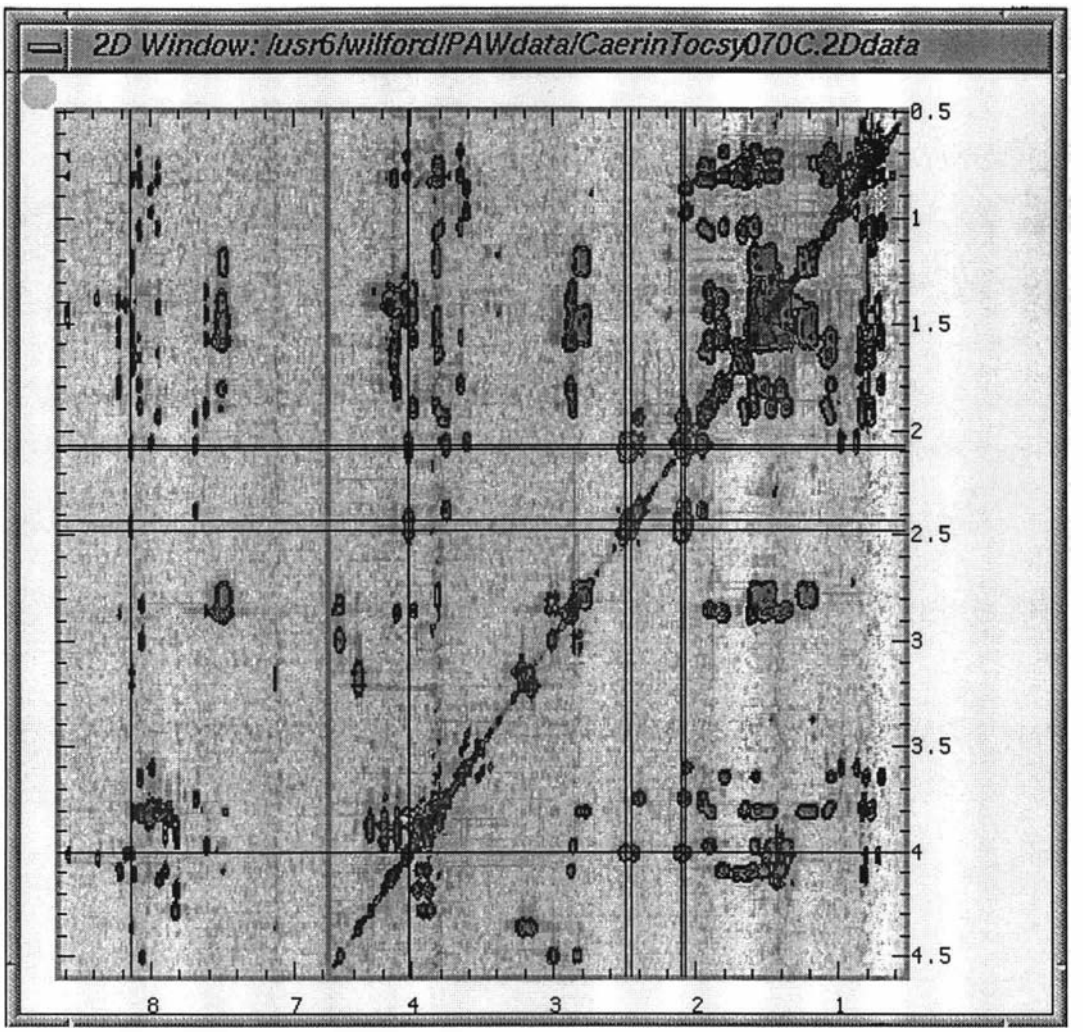
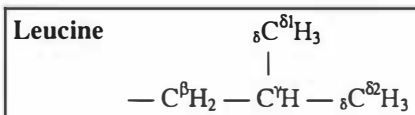


Figure 10.27: A multi-region plot that shows how the Glu spin-system in Caerin 4.1 was identified.

10.7 Identification of the Leu and Ile residues

The Leu and Ile residues in ^{1}H TOCSY spectra are characterised by the 2D cross-peak distribution patterns shown in the next two diagrams:

Leucine



Spin system:

[N-α-β-β-γ-δ1$_3$-δ2$_3$]

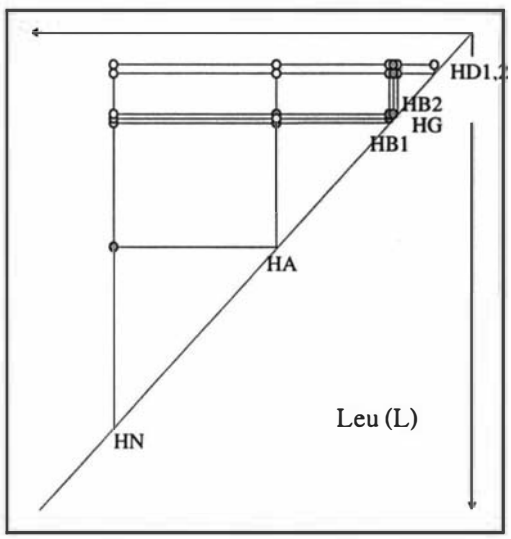
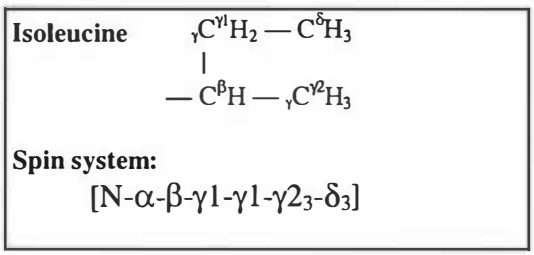


Figure 10.28: The ^{1}H TOCSY spin-system and connectivity-diagram of a Leu residue.

Isoleucine



Spin system:

[N-α-β-γ1-γ1-γ2$_3$-δ$_3$]

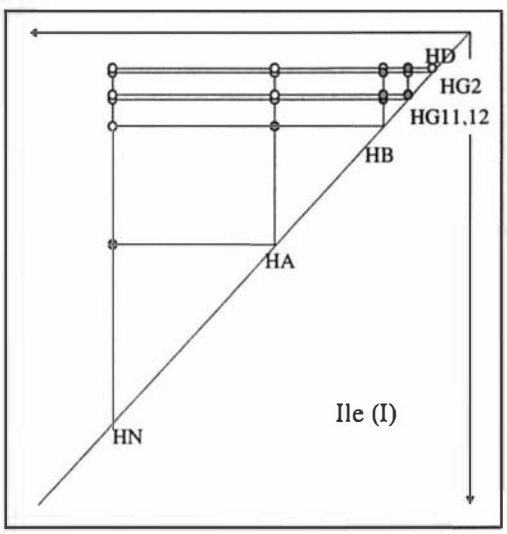


Figure 10.29: The ^{1}H TOCSY spin-system and connectivity-diagram of a Ile residue.

The two Leu residues in the sequence were identified from the remaining unidentified spin-systems by looking for their heavily overlapping peak-patterns that contain HB and HG cross-peaks. (Figure 10.30 and 10.31).

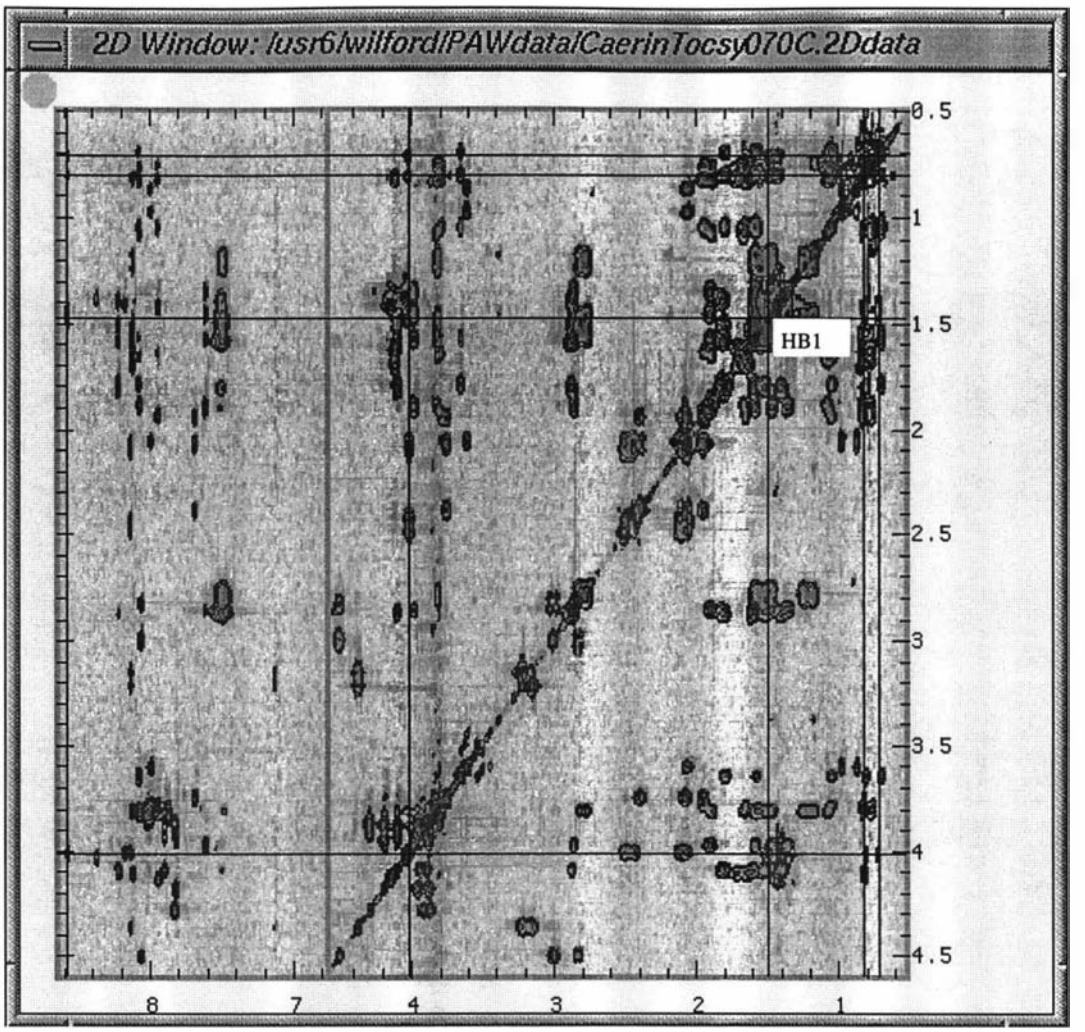


Figure 10.30: A multi-region plot that shows how one of the two Leu spin-systems in Caerin 4.1 was identified.

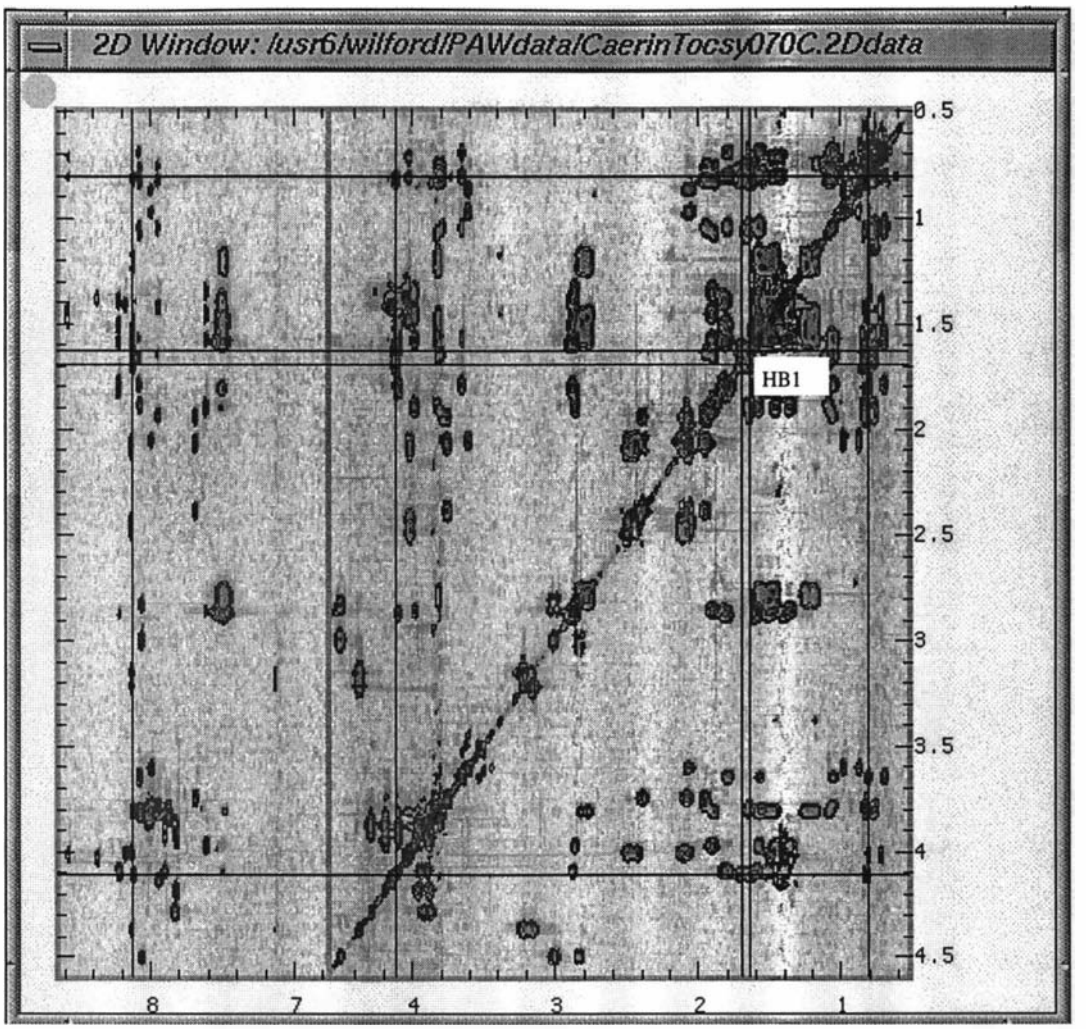


Figure 10.31: A multi-region plot that shows how the second Leu spin-system in Caerin 4.1 was identified.

After this, the three Ile were identified by their long side-chain structure, as shown in a typical example in Figure 10.32.

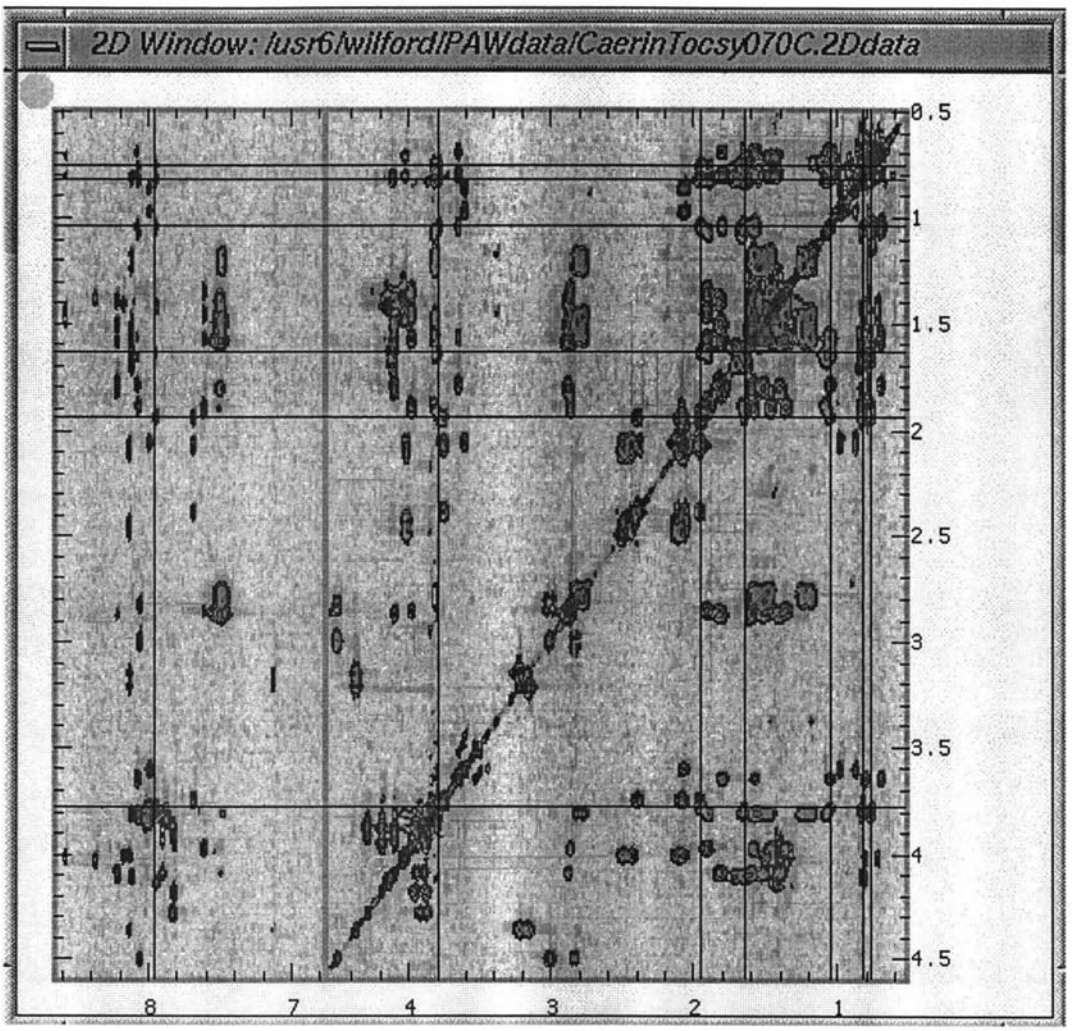


Figure 10.32: A multi-region plot that shows how one of the Ile spin-system in Caerin 4.1 was identified.

Chapter 11:

NMR Spectral Assignment for Caerin 4.1 Using PAW

11.1 Introduction

Once spin-systems have been identified and classified into one of the residue types described in Chapter 7, the next step is to assign them to specific amino-acid residues. If the spin-system is unique to a specific amino-acid residue in the sequence, the problem is simple. However, usually it is necessary to use the amino-acid sequence and NOE connectivities to determine which residue a particular spin system corresponds to.

During this process tables 7.1, 7.2 and 7.4 have been frequently referenced throughout the sequential assignment process. These tables list the primary-structure, spin-system and chemical-shift information of the 20 common amino-acid residues.

At the first stage in the assignment process, only the cross-peaks that were picked with the auto-pick routine were considered, and ambiguous ones were ignored. Around 50% of cross-peak proton codes were assigned to just a few groups, as described in Section 11.2.

More raw peaks and cross-peaks were then added, either interactively or automatically by setting a lower base level for a peak-picking operation in a small region. A number of peak-editing commands have been frequently used. Among them the most powerful and frequently used command was the lining-up-peak command that assigns a number of closely aligned peaks in a single operation, as described in Section 11.3.

During the sequential assignment, transposed rectangles and crosshairs were frequently drawn from selected peak centres to identify the connections between intra- and inter-residue cross-peaks. These have been drawn either manually or automatically with the aid of various peak-display options, as described in section 11.4.

Details of the operations used during the sequential assignment process can be found in Chapter 11 of Volume II.

11.2 Proton-code Assignment in Groups

The group assignment of cross-peaks for the proton codes has been performed by repeated use of PAW's *AssignCPks* command. The command requires that a dialog be filled in and a group of cross-peaks be selected with a user-drawn rubber rectangle on a plot. The dialog, as shown in Figure 11.1, contains an entry to specify a proton code and an entry to specify the direction in which the atom code is to be assigned for the cross-peaks enclosed in the rectangle.

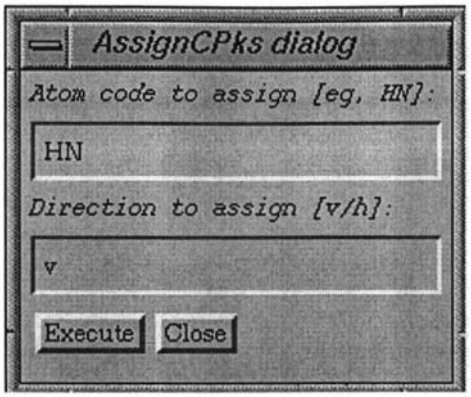


Figure 11.1: The dialog for assigning proton codes of cross-peaks in a group.

11.2.1 Proton-code assignment in the lower-left region

This process has been applied to the **CaerinCosyC** spectrum, in which every cross-peak correlates immediate-neighbour protons in a spin-system.

Figure 11.2 shows the lower-left region of the **CaerinCosyC** spectrum. Transposed rectangles have been drawn from the centres of three Trp-related peaks. Since there is only one Trp residue at location 3 in the sequence, they can be assigned unambiguously to W3.

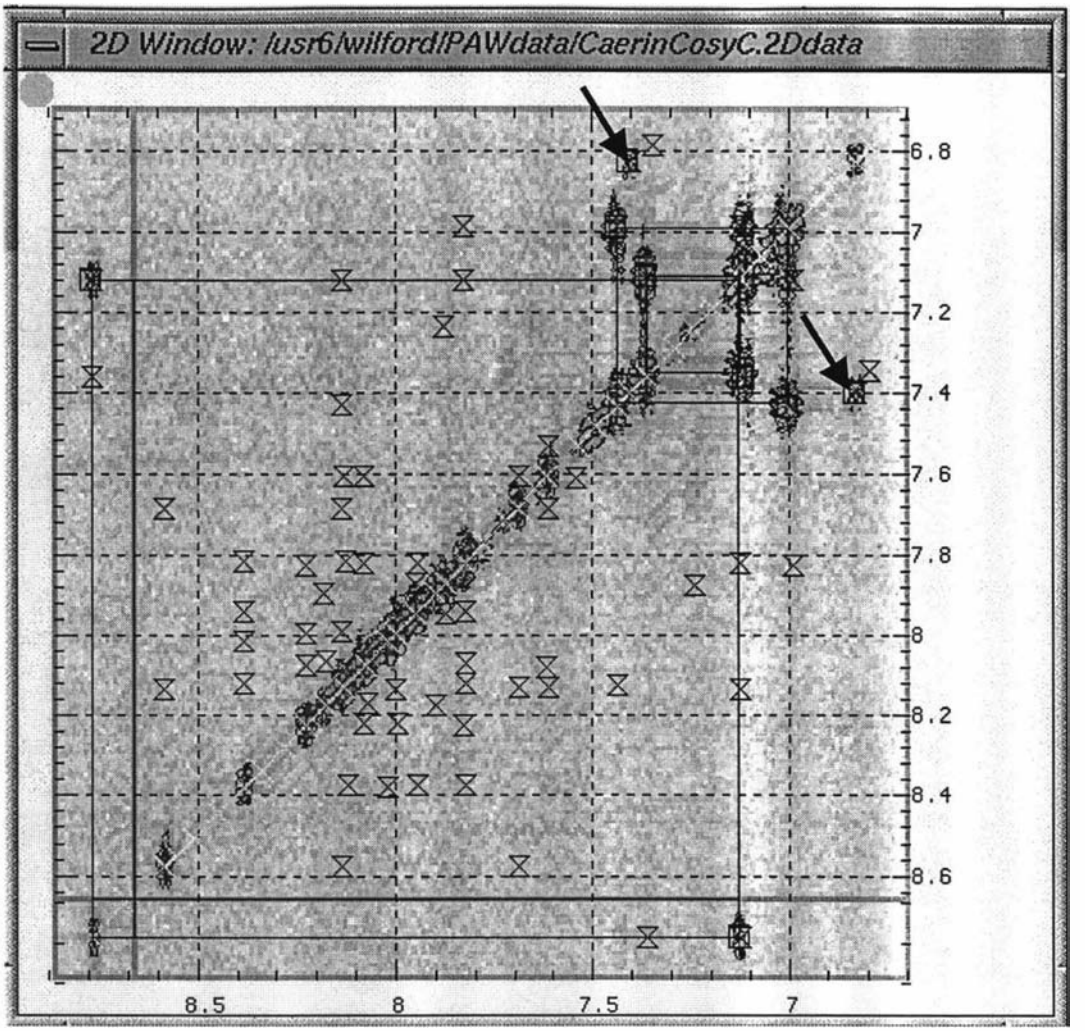


Figure 11.2: The lower-left regions of the **CaerinCosyC** spectrum, where the transposed rectangles have been drawn from the centres of three W3-related peaks.

Assignment for peaks in this region were made by reference to Table 7.2 and by noting the following points:

- W3:HD1 and W3:HE1 form a two-spin system in the **CaerinCosyC** spectrum (see Table 7.2). In addition, W3:HE1 must have the highest chemical shift in the spectrum (see Table 7.3). Therefore, the four NOESY cross-peaks at the upper-left and lower-right regions were recognised to be W3:HE1-related and were correspondingly assigned. Of these, the two cross-peaks at the corners of the largest rectangle were assigned as W3:HD1, because they coincided with DQF-COSY cross-peaks.

- The peaks linked to the diagonal peaks at [7.43,7.43], [7.36,7.36], [7.12,7.12] and [6.99,6.99][1] were assigned as W3:HE3, W3:HZ3, W3:HH2, and W3:HZ3, respectively, according to the chemical shift relations shown in Table 7.2. These

[1] In units of ppm.

are the cross-peaks at the corners of the other two transposed rectangles in Figure 11.2.

- The two peaks (arrowed) were assigned as Q4:HE2. These assignment results were confirmed later.

The remaining peaks (i.e., those shown inside the three rectangles in the next plot) were assigned as follows:

- The peaks enclosed inside the largest rectangle were assigned generally as the HN:HN cross-peaks using PAW's group-assignment command, as described in Chapter 11 of Volume II.
- The peaks inside the other two rectangles were assigned to either W3:HE3, W3:HZ3, W3:HH2, W3:HZ3 or X24:HN, where X24 signifies the terminal HN_2. They were assigned using PAW's lining-up-peak command by reference to previously assigned peaks, as described in Chapter 11 of Volume II.

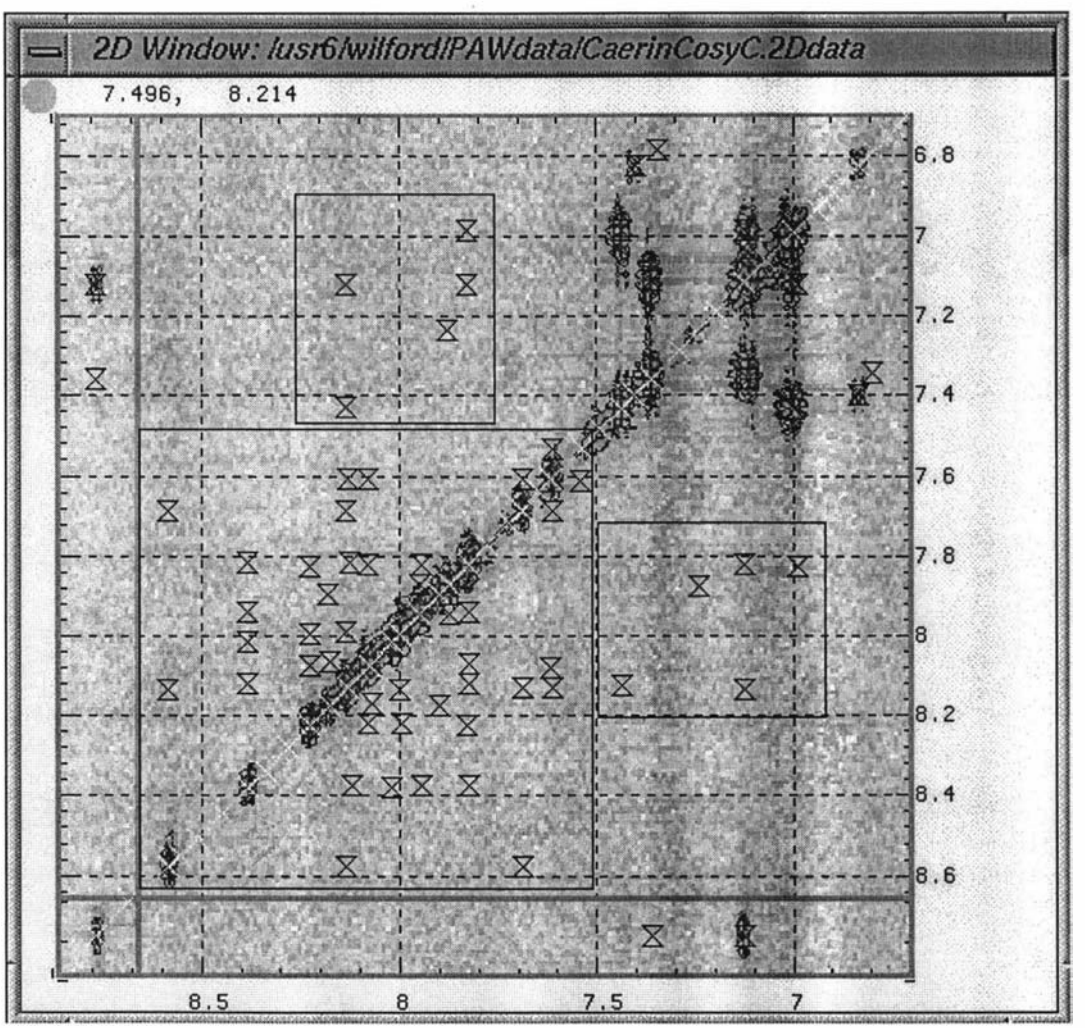


Figure 11.3: The lower-left region of the CaerinCosyC spectrum. The largest rectangles show the HN-HN cross-peaks. The peaks that are enclosed inside the other two rectangles were assigned to either W3:HE3, W3:HZ3, W3:HH2, W3:HZ3 or X24:HN, where X24 signifies the terminal HN_2.

Figure 11.4 shows the results of the assignment. The proton codes of all peaks in the lower-left region were assigned apart from two. These are located at the two corners of the transposed-rectangle in the middle of the figure, and they are not linked with any of the significant diagonal peaks.

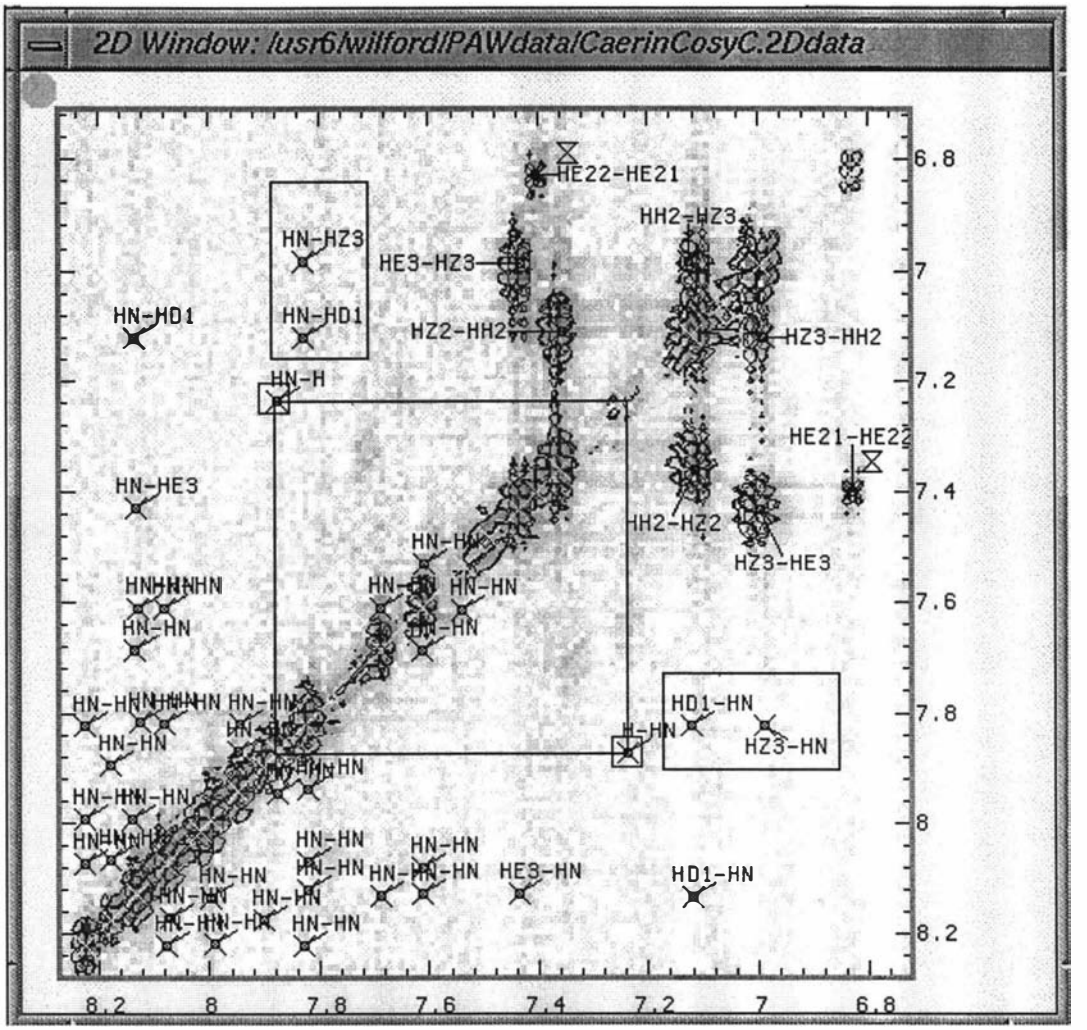


Figure 11.4: The proton codes of the NOESY cross-peaks in the lower-left region. The assignment was performed using the **CaerinCosyC** spectrum.

Note:

- There was an ambiguity regarding the assignment of the W3:HD1-related peaks because its chemical shift is almost identical with that of the W3:HH2. This was resolved by considering the fact that, within the Trp residue, HH2 and HZ2 are far away from the other protons beyond the ring, and hence their connectivity with others is less likely to be seen in the **CaerinNoesy150C** spectrum.

- Because of overlap, there was insufficient evidence to identify X24-related peaks at this initial stage. Four of the assignments, as enclosed by the two smaller rectangles in the above figure, are in fact incorrect and will be changed later. Initial mistakes like this occur frequently during spectral assignment process.

11.2.2 Proton-code assignment in the upper-left region

The group assignment of proton codes in this region was performed using the **CaerinTocsy070C** spectrum (Figure 11.5), in which only the intra-residue cross-peaks that relate protons in the same spin-system occur. When the NOESY cross-peak symbols are superimposed over the **CaerinTocsy070C** spectrum, most of the inter-residue NOESY cross-peaks could be clearly identified.

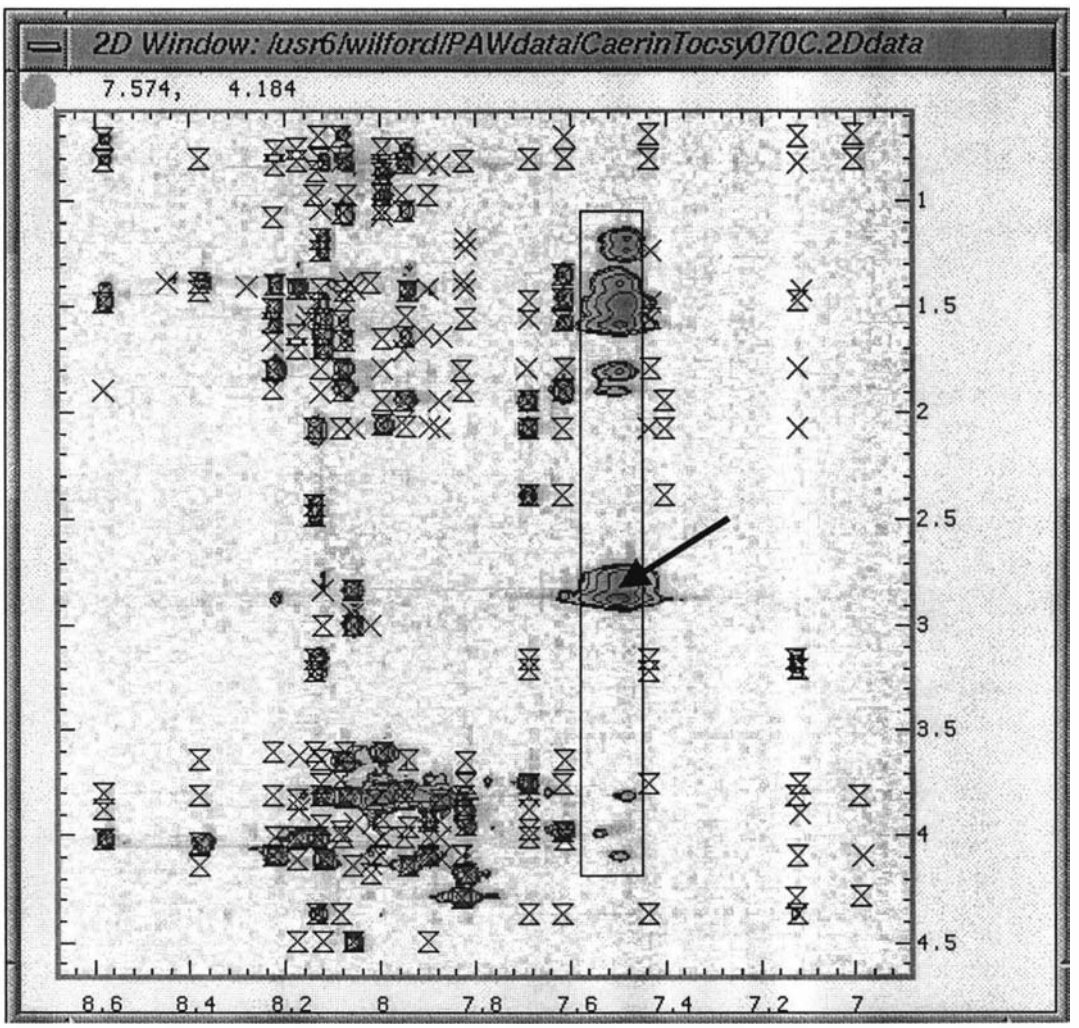


Figure 11.5: The upper-left region of the **CaerinTocsy070C** spectrum with the NOESY cross-peaks superimposed.

As is evident from Figure 11.5, the NOESY cross-peaks fall into two categories. Each of those associated with an intra-residue connectivity have a corresponding TOCSY peak, whereas those associated with an inter-residue connectivity do not.

In can also be seen that those peaks enclosed within the rectangle at 7.5 ppm in D2 are not related to any of the NOESY peaks, because they are located at the far end of the Lys side-chain and hence the NOEs are too weak to be seen. They are the HZ3-related peaks of the Lys residues in the protein. Also, only the HZ3-HE (arrowed) cross-peaks can be seen in the **CaerinCosyC** spectrum.

The proton codes in the vertical direction were all assigned using the group assignment operations. Figure 11.6 shows the result of the assignment.

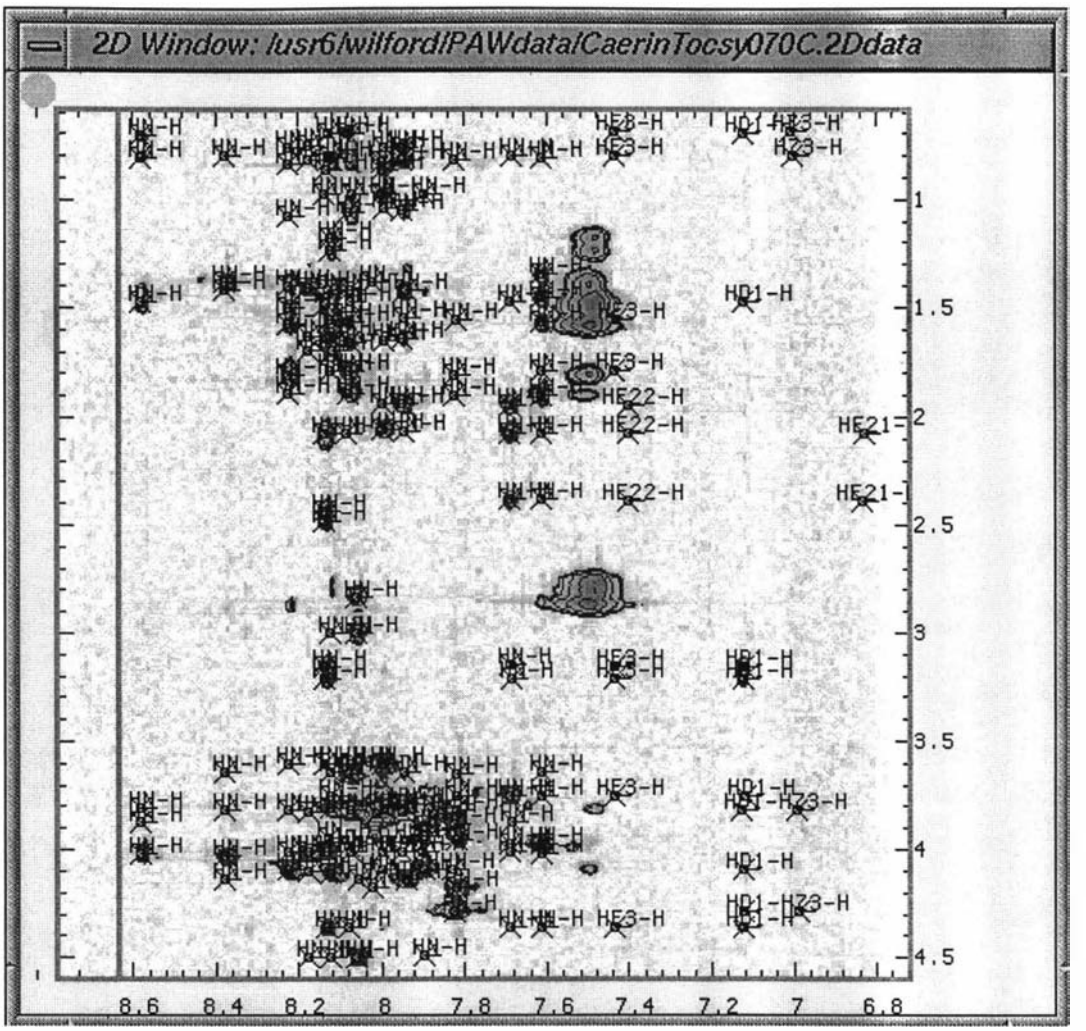


Figure 11.6: The group assignment result for all proton codes along D1 in the upper-left region of the **CaerinTocsy070C** spectrum.

Initially, the proton codes in D2 for all peaks in the upper-left region which had chemical-shifts larger than 3.6 ppm were assigned to HA (Figure 11.7), despite the fact that a small fraction of them were HB.

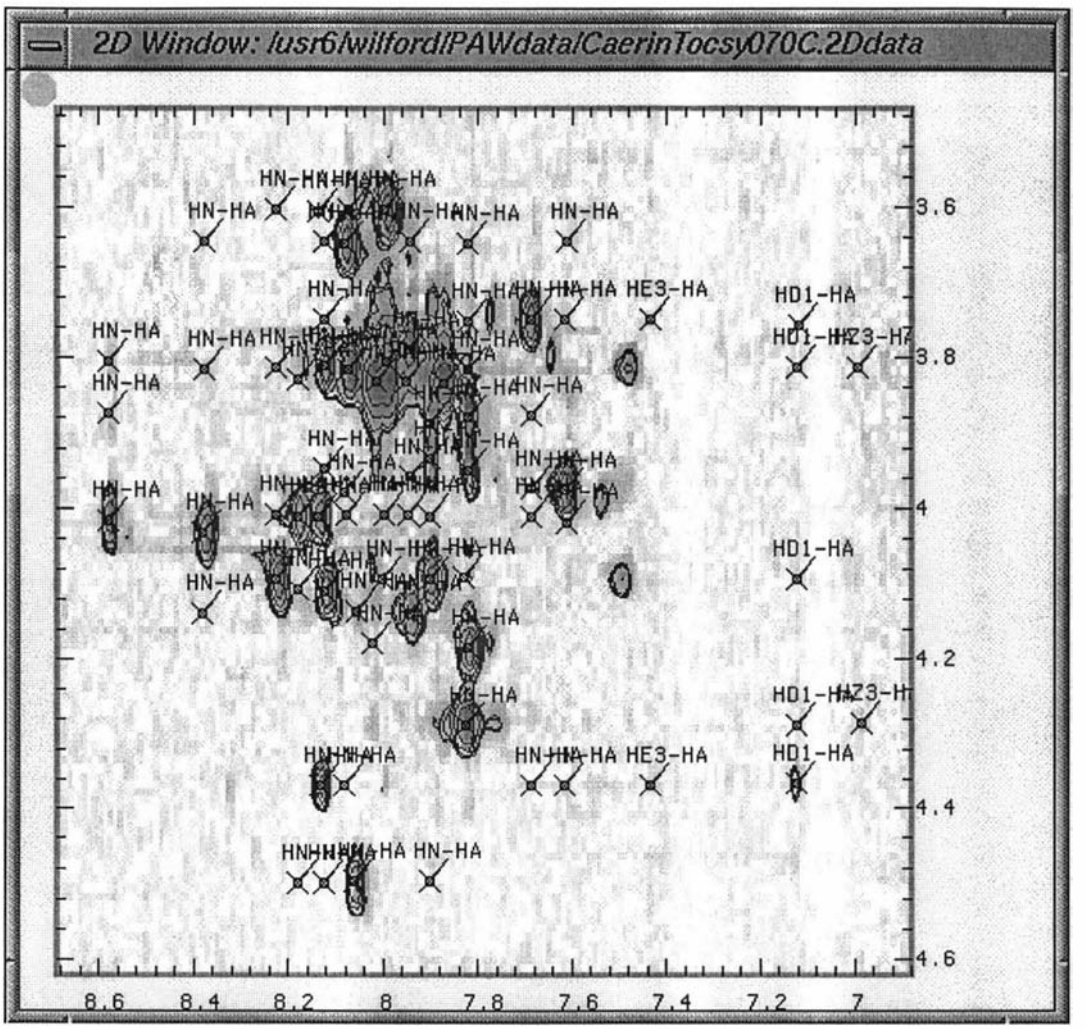


Figure 11.7: The group assignment result for all proton codes along D2 in the lower part of the upper-left region of the **CaerinTocsy070C** spectrum.

The remaining peaks had their D2 proton codes assigned later.

11.2.3 Proton-code assignment in the lower-right region

The assignment of cross-peak proton-codes in the lower-right region is optional. It was done at the last stage of the assignment in order to verify the previous assignments. For the Caerin 4.1 spectra, there was another reason: some peaks appeared to split in D1 due to better resolution that enables the spin-spin coulpling to be resolved.

11.2.4 Proton-code assignment in the upper-right region

The assignment of this region was also performed on the **CaerinTocsy070C** spectrum but with a higher *First-level Threshold* for the contour plot because of the higher peak-intensities. The **CaerinTocsy070C** plot with the NOESY cross-peak symbols superimposed is shown in Figure 11.8.

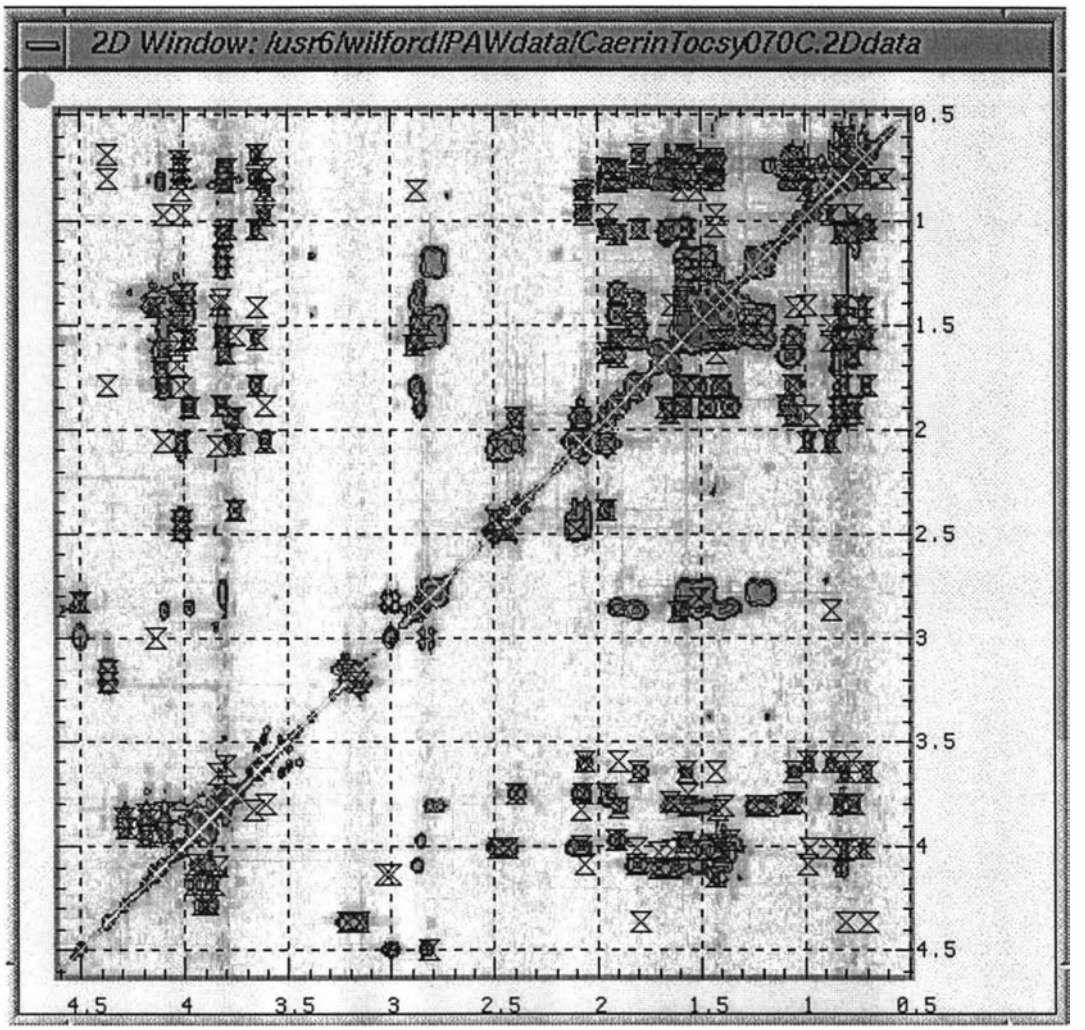


Figure 11.8: The upper-right region of the **CaerinTocsy070C** spectrum, with the NOESY cross-peak symbols superimposed.

Cross-peaks with chemical shifts larger than 3.6 in the upper-right (or aliphatic) region were again assigned as HA, despite the fact that some of them should in fact be HB. (This is corrected later.) Figure 11.9 shows the result, where the peak assignment below the diagonal is optional.

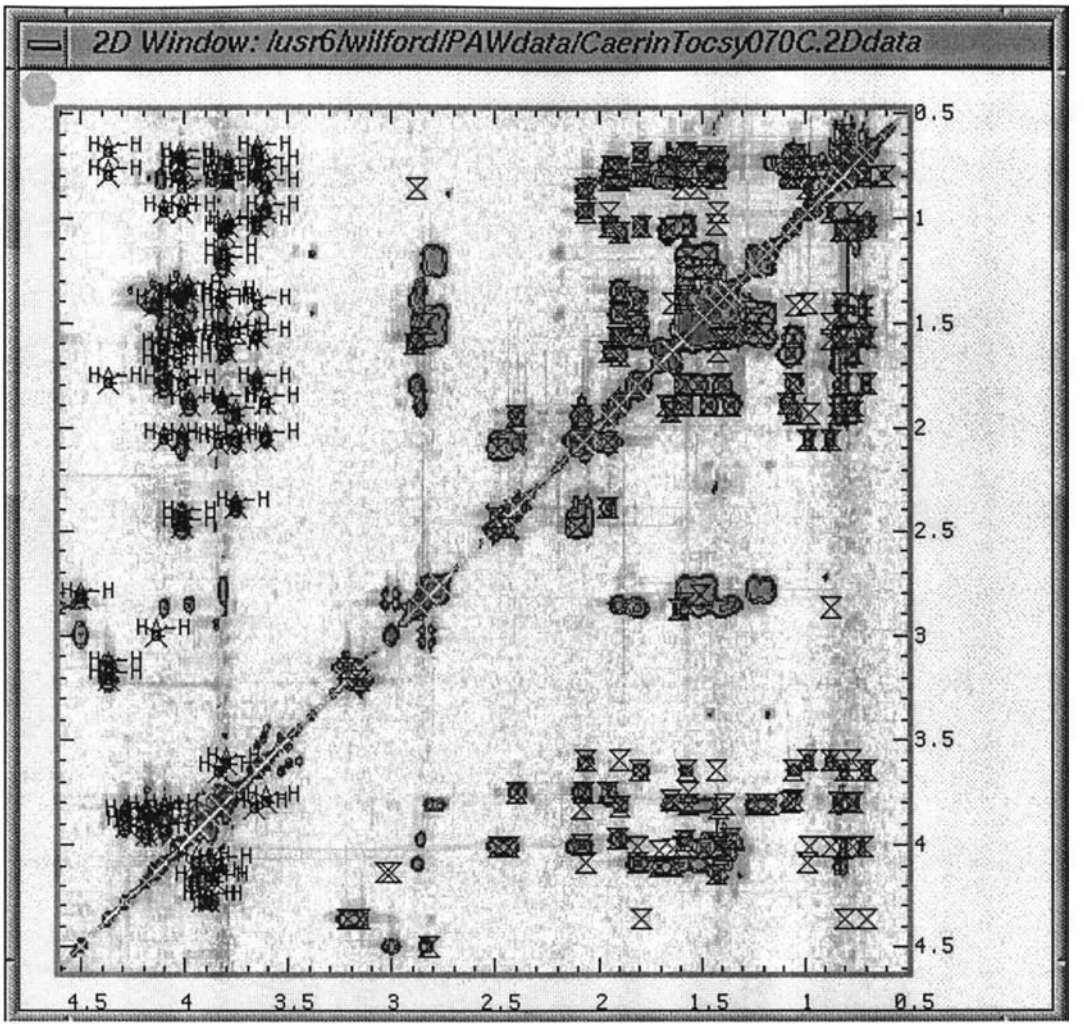


Figure 11.9: The assignment results for HA in D1 in the upper-right region. Some of them should in fact be HB and are corrected later.

11.3 Adding and Editing Cross-peak Records

This section describes a few more operations that change the information in a cross-peak list, including adding, copying and modifying the cross-peak records.

For the reasons mentioned in the Section 11.2, this process was mainly performed on the **CaerinTocsy070C** spectrum. Transposed-rectangles and crosshairs were frequently drawn throughout this process. A few examples detailing the cross-peak analysis and different editing operations can be found in Chapter 11 of Volume II.

11.3.1 Cross-peak analysis using multiple spectrum display

The analysis for the cross-peaks has been performed by cross-referencing the plots of the same small region in different spectra. For example, the three expanded plots in Figures 11.10 were used to unambiguously assign the HE2-related cross-peaks of Q4

and all cross-peaks that are associated with the ring structure of W3. Both W and Q appear only once in the sequence.

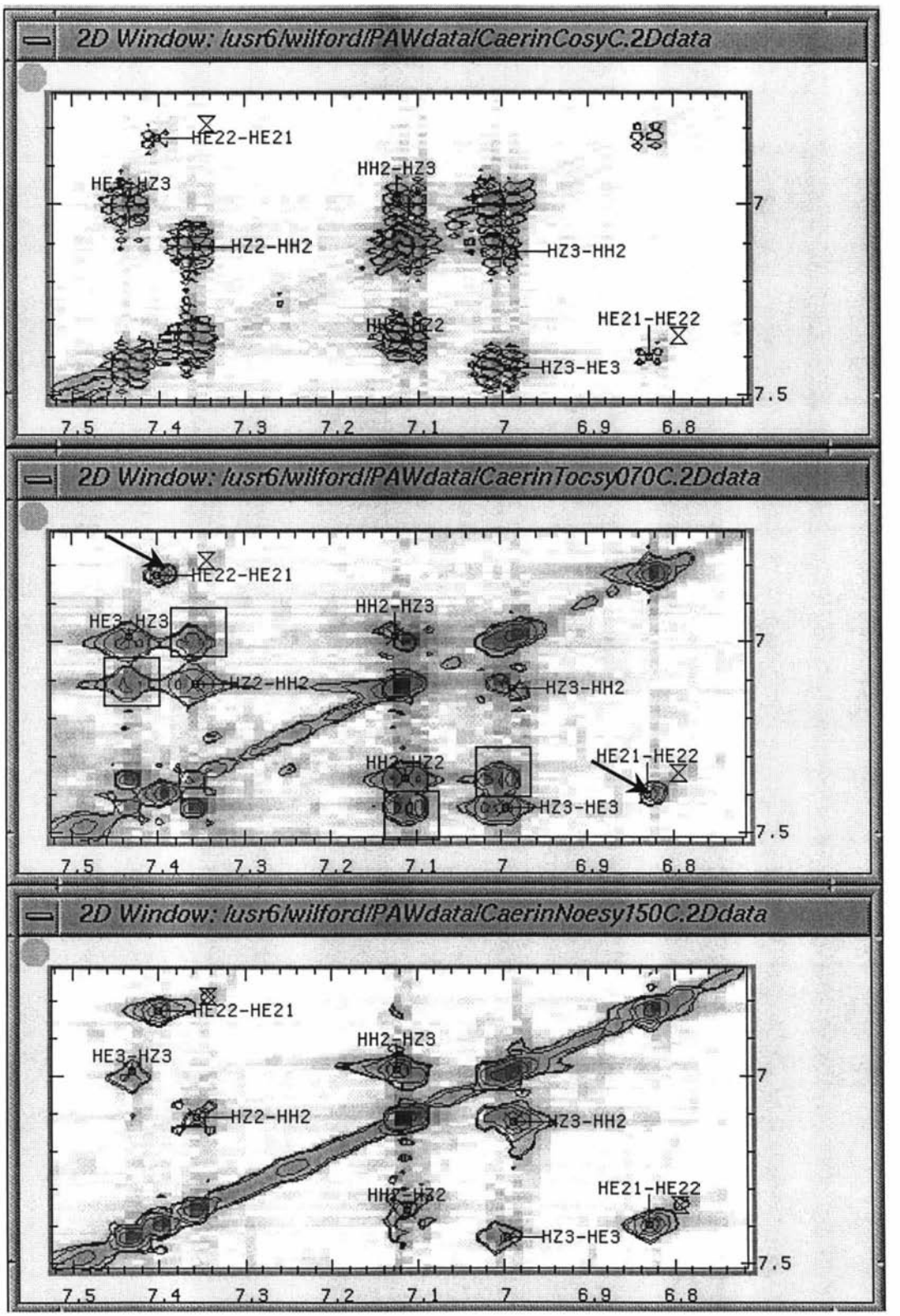


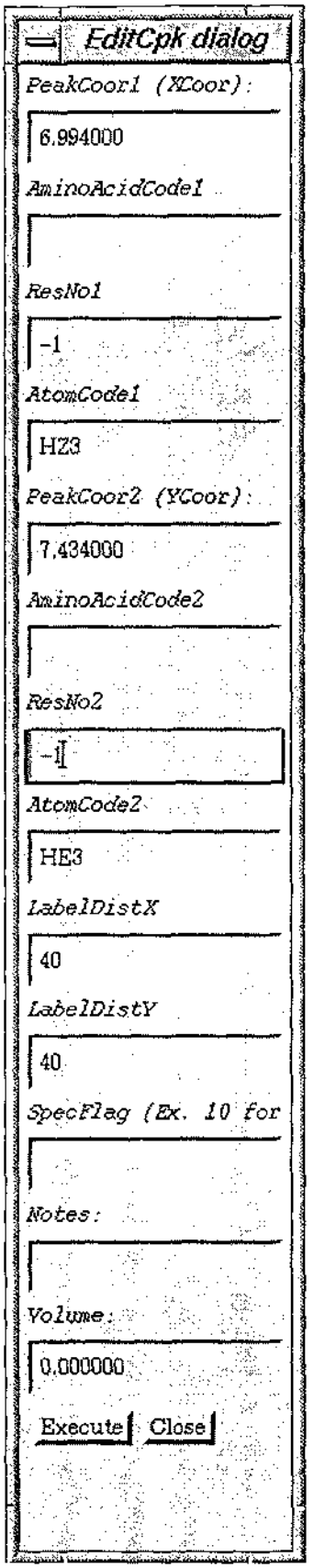
Figure 11.10: An example of a multiple spectrum display used in the cross-peak analysis.

From the plots, it can be seen that the four TOCSY cross-peaks enclosed in the small rectangles are apparently missing in the **CaerinNoesy150C** and **CaerinCosyC**. This is a good indication of a Trp ring-structure, because HE3 and HZ3 are far apart from each other in the ring. (See Table 7.1 for the primary structure.) As a result, the remaining two cross-peaks arrowed can only belong to either Q4 or the terminal NH_2. Later, this was confirmed to be Q4.

11.3.2 Editing cross-peak records

Around 15% of the Caerin 4.1 cross-peak records were initially modified using PAW's *EdtOneCPk* command. The command requires that a peak be selected and a dialog (Figure 11.11) be filled. The meaning of each field in the dialog is self-explanatory. In many cases, only four fields were changed initially, as shown in Figure 11.11 (b).

(a)



(b)

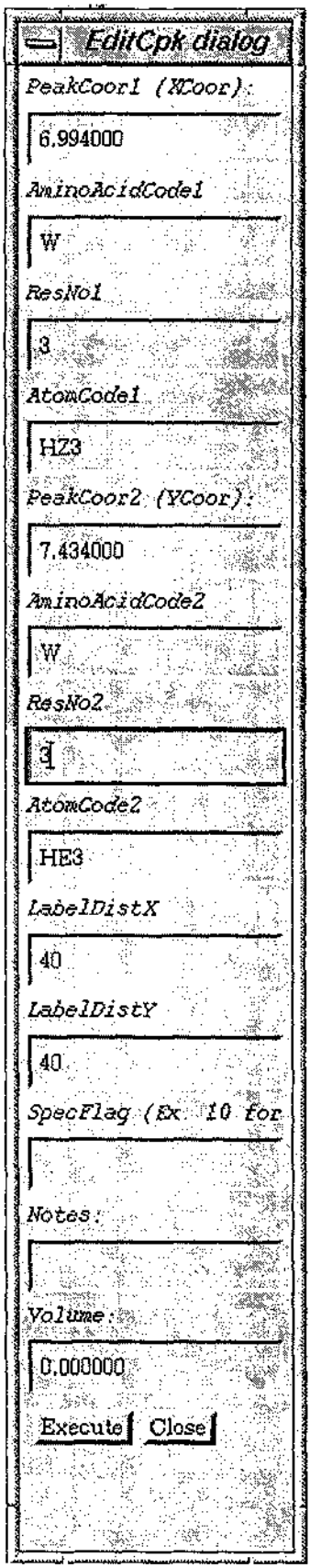


Figure 11.11: (a) The dialog for editing a cross-peak record with automatic assigned values, where −1 indicates an unassigned residue number. (b) The dialog for editing a cross-peak record with additional assigned values.

11.3.3 Editing transposed cross-peak records

Some cross-peaks were assigned using PAW's *EditTranCPk* command. The command requires two peaks be selected, one to copy and one to edit. The same dialog for editing a cross-peak record is used, with six of the entry values copied from the first peak selected. These include the *AminoAcidCode*, *ResNo* and *AtomCode* in both dimensions. In most of the cases, no changes were made, and the [Execute] button could be simply selected.

11.3.4 Editing cross-peak positions

Some of the cross-peak positions were shifted slightly because of overlap. For example, it can be seen from Figure 11.12 that the two W3:HZ3-related cross-peaks on the right do not vertically align with each other. Neither do the two on the top in horizontal direction.

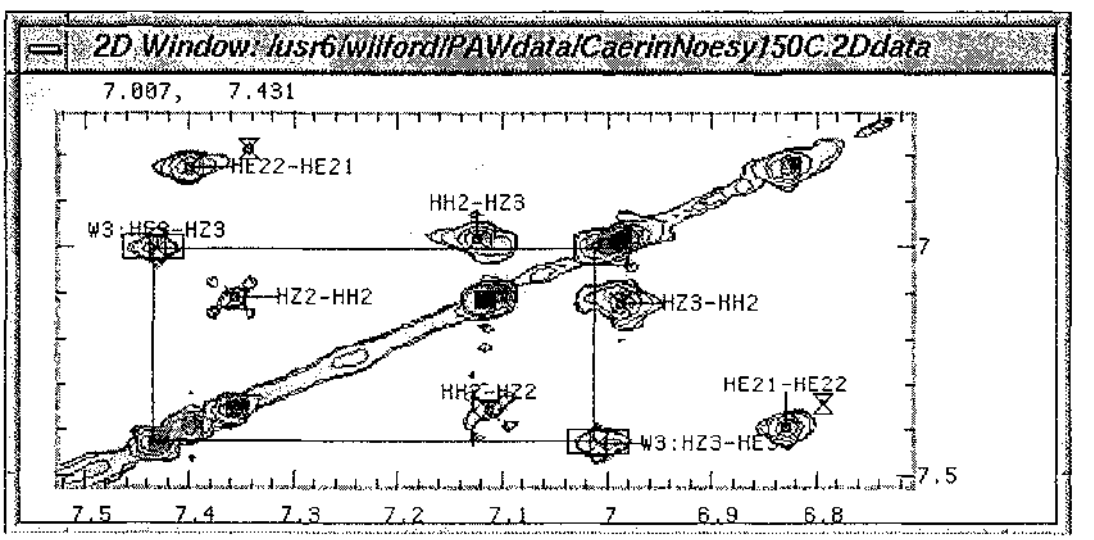


Figure 11.12: The result of a cross-peak assignment of W3:HZ3-HE3 after correction by using 1D spectral slices. (See Figure 11.10 for comparison.)

It was found later that this was because of an overlap with the terminal NH_2 cross-peaks, symbolised respectively as X24:NH2 and X24:HN1.

This problem could usually be solved by obtaining one-dimensional spectral slices through the peaks. The corrected chemical shift was then applied to the cross-peak records using the *EdtOneCPk* command described previously.

For example, the chemical shift of W3:HZ3 was set to 7.007 ppm, which was obtained from two 1D spectra: one was drawn from a row that passes through W3:HZ3-HE3, and the other from a column that passes through W3:HE3-HZ3. The corrected results are shown in Figure 11.14.

11.3.5 Finding precise cross-peak positions

Many new peaks were added and the positions of cross-peaks were improved during the assignment. In most cases, the positions were obtained by drawing crosshairs from other well-defined cross-peaks that relate to them. For example, Figure 11.13 shows two sets of crosshairs that were drawn from the two peaks enclosed by the rectangles on the left in order to find the precise location of the cross-peak pair HH2-HZ3 and HZ3-HH2.

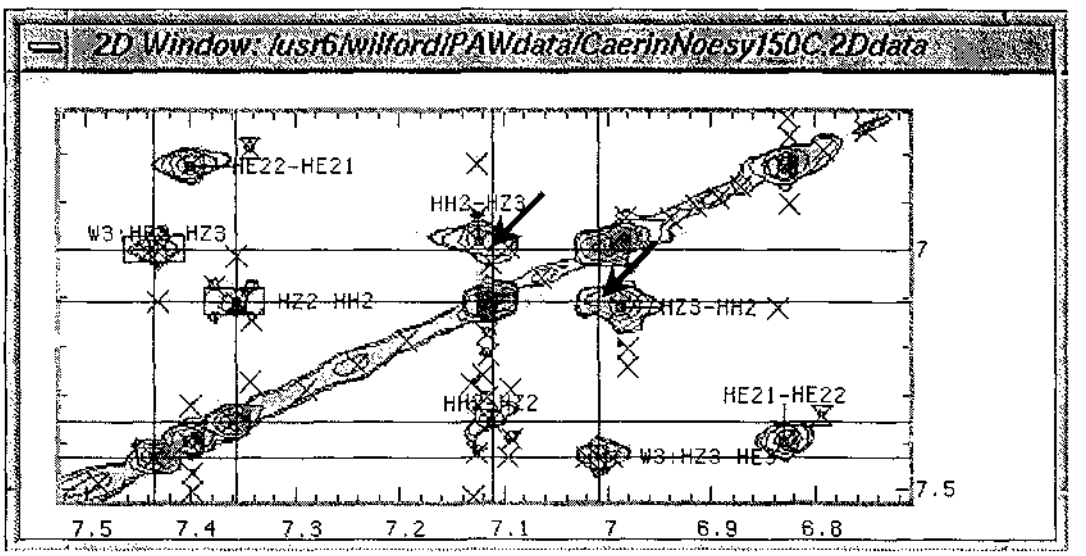


Figure 11.13: A plot that shows two sets of crosshairs that were drawn from the two peaks enclosed by the rectangles on the left.

With the help of the crosshairs, it was concluded that the two intersections of the transposed rectangle at the top-left define precisely the peak positions, arrowed in the plot. A raw peak was then picked manually at one of the intersections with PAW's *PickOneCpk* command, followed by the *FindTransPk* command to add a pair of transposed cross-peaks. (See the next plot.)

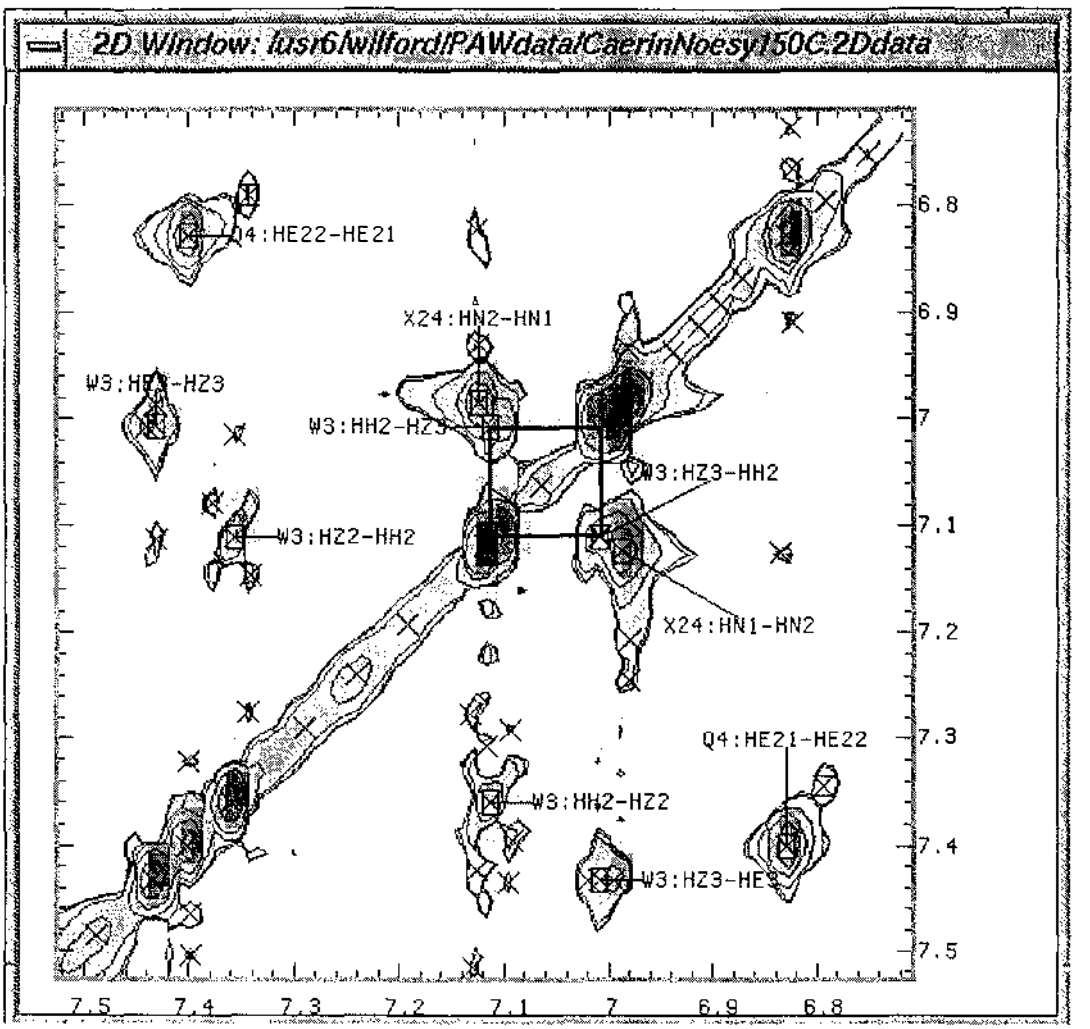


Figure 11.14: A plot that shows a new pair of missing cross-peaks that was picked and assigned based on the positions of the W3:HE3-HZ3 and W3:HZ2-HH2 cross-peaks.

11.3.6 Lining up cross-peaks

The majority of the cross-peaks were assigned using the *LineUpCpk* command. The operation lines up and assigns a group of cross-peaks in one dimension using the correct position and assignment of a cross-peak that is well defined. This process largely reduces the necessity of entering individual records.

There are two ways that a group of cross-peaks can be lined up. One of them lines up a group of selected cross-peaks, and the other one lines up all the cross-peaks in one dimension. The former requires a number of cross-peaks be selected manually, with the first one being the reference peak. The latter requires only one reference peak be selected. Both of them use the same *LineUpCpk Dialog* (Figure 11.15).

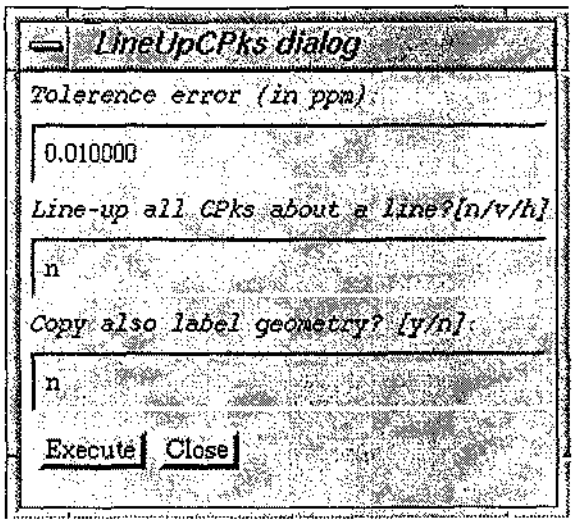


Figure 11.15: The dialog for lining up cross-peaks.

Note that

- When the second entry value is n, only peaks that are selected by clicking are aligned. When v (or h) is entered, all peaks within a tolerant distance from the vertical (or horizontal) line drawn from the centre of the reference peak are aligned.
- When the third entry value is n, the label geometry of the reference peak is not copied to any of the aligned peaks; otherwise, it will be copied for all of them.

Figure 11.16 shows three columns of peaks that have been properly lined up and assigned in the upper-left region of the NOESY spectrum.

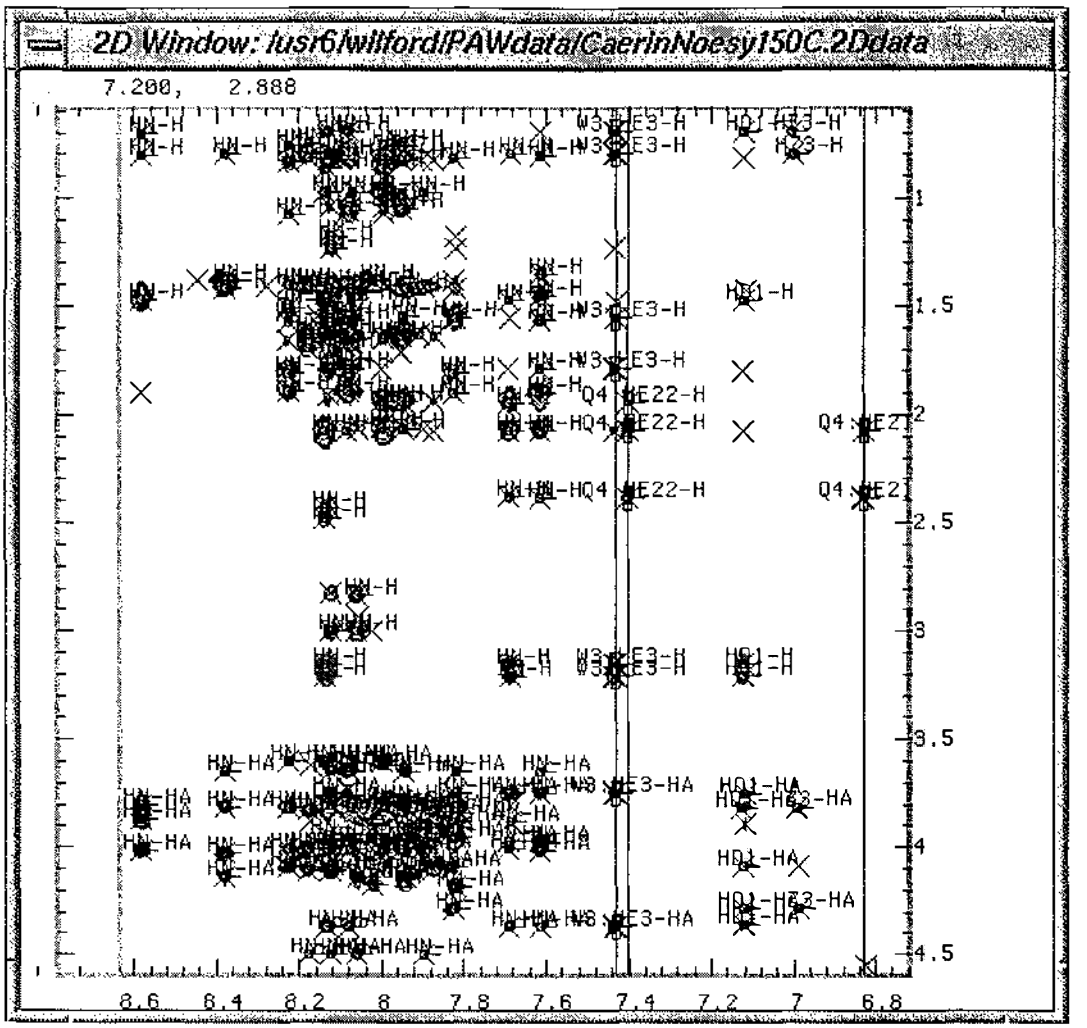


Figure 11.16: A plot that shows three columns of peaks that have been properly lined up and assigned in the upper-left region of the **CaerinNoesy150C** spectrum.

11.3.7 Moving cross-peak labels

By default, the label for an assigned peak is centred at the upper-right corner of the peak. This arrangement may not be the best, especially in a crowded region. For example, the label arrangement in the last plot is really not suitable for publication. The positions of the labels in Figure 11.14 are acceptable, and they have in fact been shifted.

One way to shift the labels is to follow the peak editing procedure described previously, and change the values for the *LabelDistX* and *LabelDistY* from the *EditCpk Dialog*. The majority of label positions in the assignment for Caerin 4.1, however, were adjusted using PAW's *MovePkLbl* command that provides a graphic-interface method to move the peak labels. The command draws a big cross from the current peak centre as a reference for the new position, as shown in Figure 11.17.

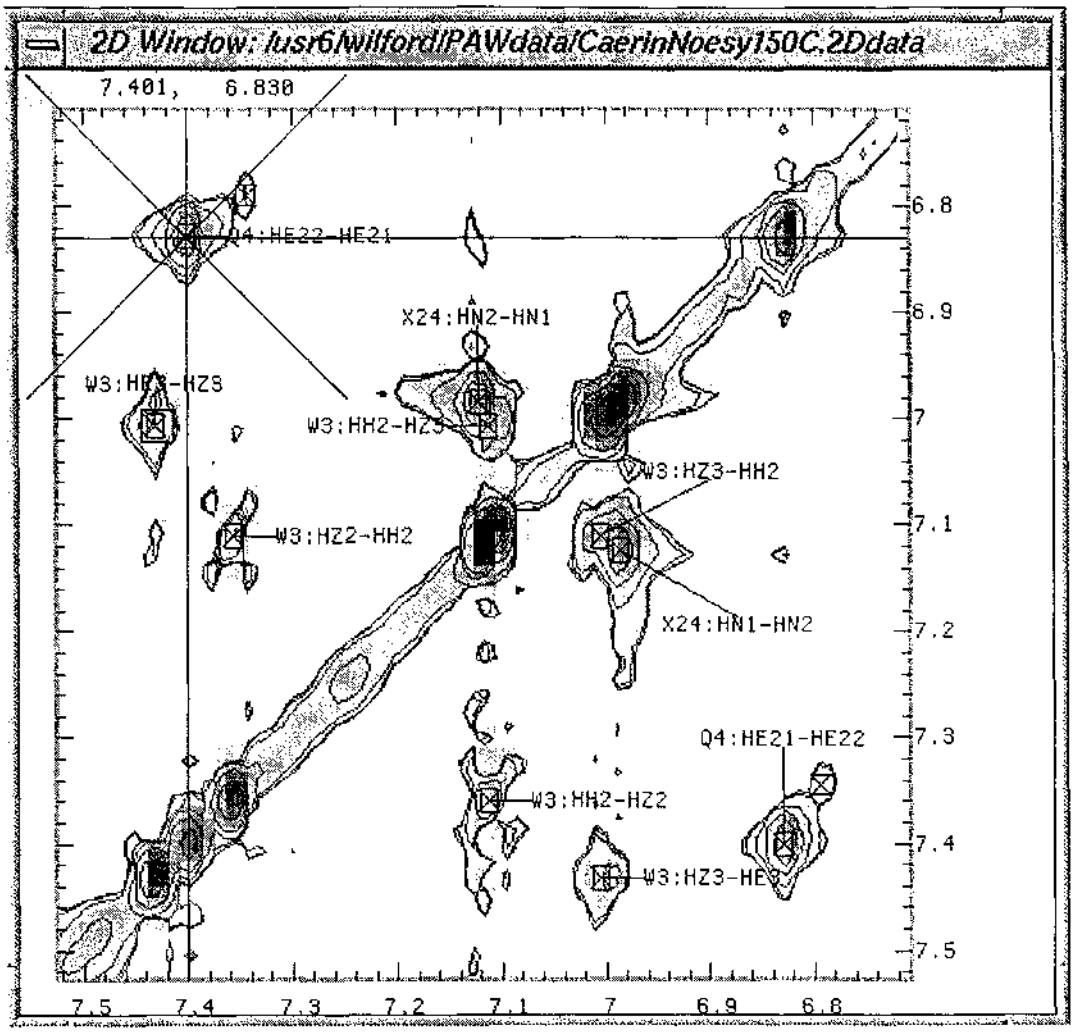


Figure 11.17: A big cross that was drawn on a plot for the reference to select a new peak-label position.

And then, with a second click on any position, the new label position is set.

Figure 11.18 shows a few labels that were moved using the graphic-interface method.

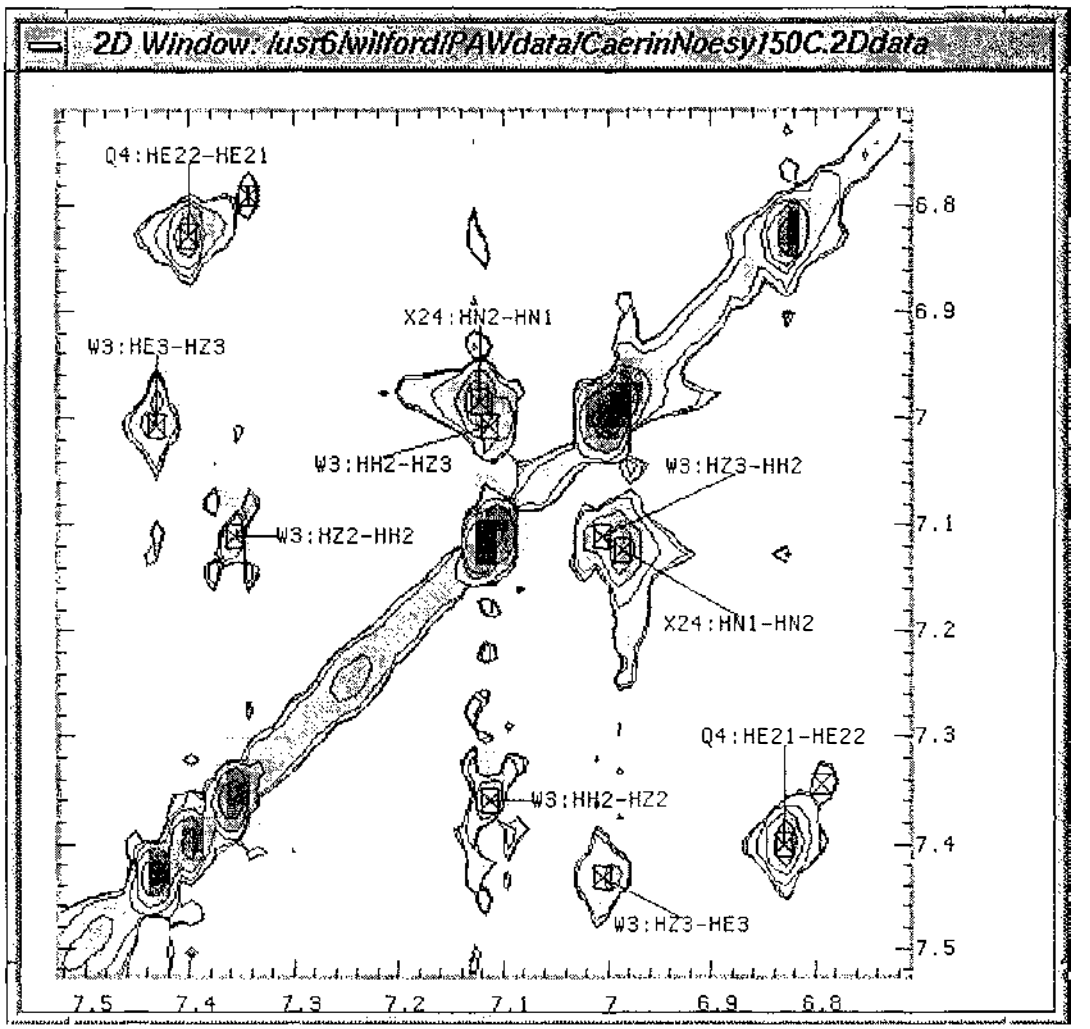


Figure 11.18: Plot with a better label-display layout. The labels were moved using the graphic-interface method.

11.4 Sequence-specific Assignment

The last stage of the assignment process is to identify which residue (or spin-system if still ambiguous) goes with which sequence-specific residue (e.g., which amino acid in the sequence does the AMX spin-system *n* belong to). This process, called the sequence-specific assignment is a very time-consuming process, even for a small molecule like Caerin 4.1. This is due to the poor peak-resolution in most NMR spectra. The peak-editing operations are the same as those described previously. This section briefly describes how the sequential assignment for Caerin 4.1 has been done with PAW. For more details, see Chapter 11 of Volume II.

Both the **CaerinNoesy150C** and **CaerinTocsy070C** spectra were used in this process. Most of the inter-residue NOESY cross-peaks were clearly identified by displaying the NOESY peaks on top of the **CaerinTocsy070C** spectrum. The existence of certain

inter-residue cross-peaks and their intensities indicate that the secondary structure of Caerin 4.1 is prominently helical.

Table 7.5, which contains important information about the distances between protons in the basic protein secondary structures, has been used frequently throughout the assignment.

Initially, the assignment was focused on the fingerprint and HN-HN regions. Starting with the assignment of the five unambiguous residues, the next neighbour residue connectivity were elucidated. This has been performed with the aid of various peak-display options.

11.4.1 The symbols for the assigned cross-peaks

As mentioned in the last chapter, unassigned cross-peaks are indicated by symbols like hour-glasses. In addition, PAW indicates assigned cross-peaks using different symbols according to the completion of assignment for the cross-peak.

There are six fields to be assigned for each cross-peak. These are the amino-acid codes, residue numbers and proton codes in D1 and D2. The rules for the cross-peak symbol formation are:

- Partially assigned cross-peaks are indicated by different shaped rectangles.
- A small square indicates one field is assigned in each direction (see Figure 11.4). A medium-sized square indicates two fields are assigned in each direction.
- A vertical oblong indicates the number of fields assigned in D1 is more than that in D2. Likewise, A horizontal oblong indicates the number of fields assigned in D2 is more than that in D1 (see Figure 11.18).
- Fully assigned cross-peaks are displayed with the biggest squares(see Figure 11.25).

11.4.2 Assignment of the unambiguous residues

Five HN-HA cross-peaks belong to five amino acids that appear only once in the sequence. A set of crosshairs was drawn from each of these cross-peaks to identify other cross-peaks that are related to them, as shown in Figure 11.19. The five HN-HA cross-peaks assigned are, from top to bottom, V18, Q4, E19, W3 and D12.

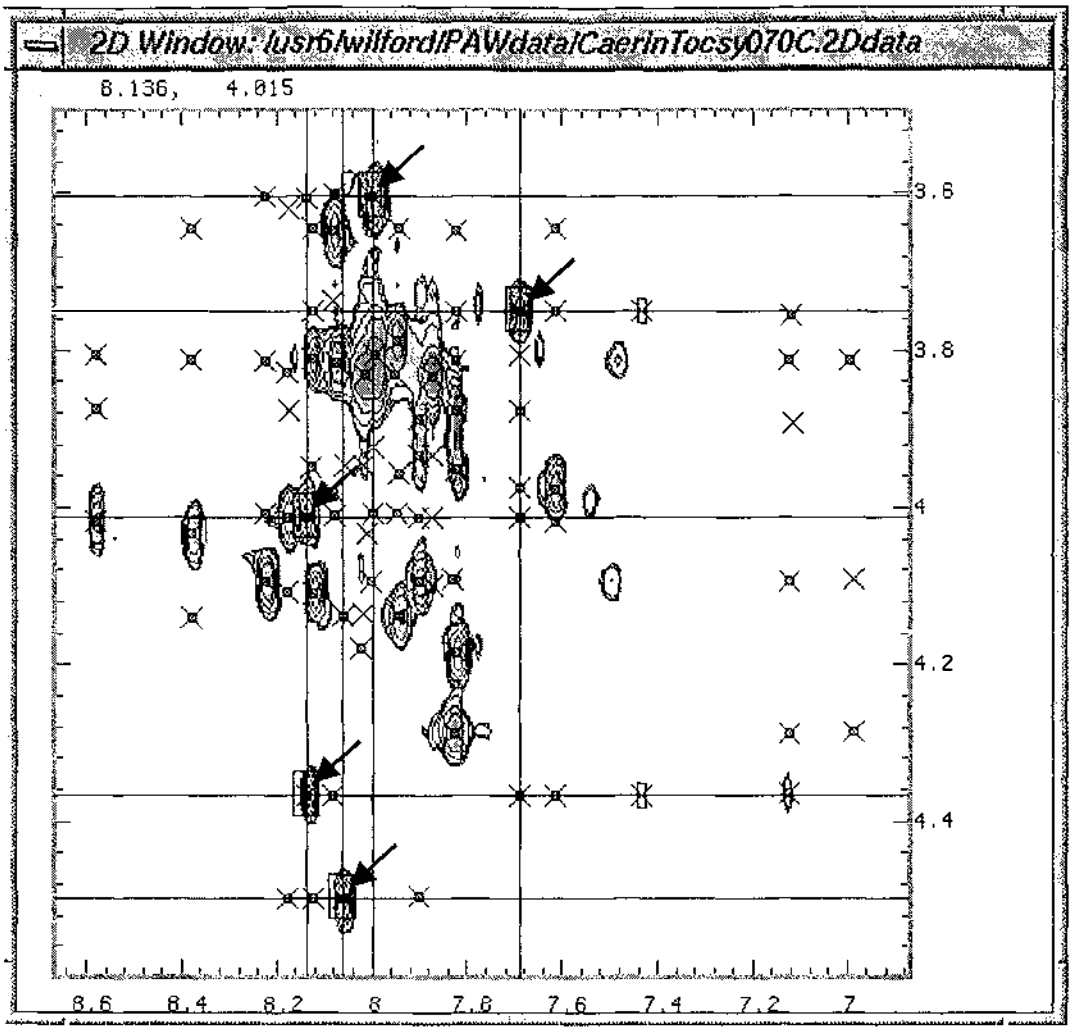


Figure 11.19: The HN-HA cross-peaks of five unambiguously known residues in the sequence. The crosshairs were drawn to identify cross-peaks that are related to them.

All cross-peaks that can be clearly separated vertically and horizontally were then assigned by PAW's lining-up cross-peak command, including those in the columns for V18, Q4, W3 and D12, as well as those in the rows for D12 and Q4.

Figure 11.20 shows the upper-left region of the **CaerinNoesy150C** spectrum after the operation, with the peak symbols plotted differently according to the assignment results.

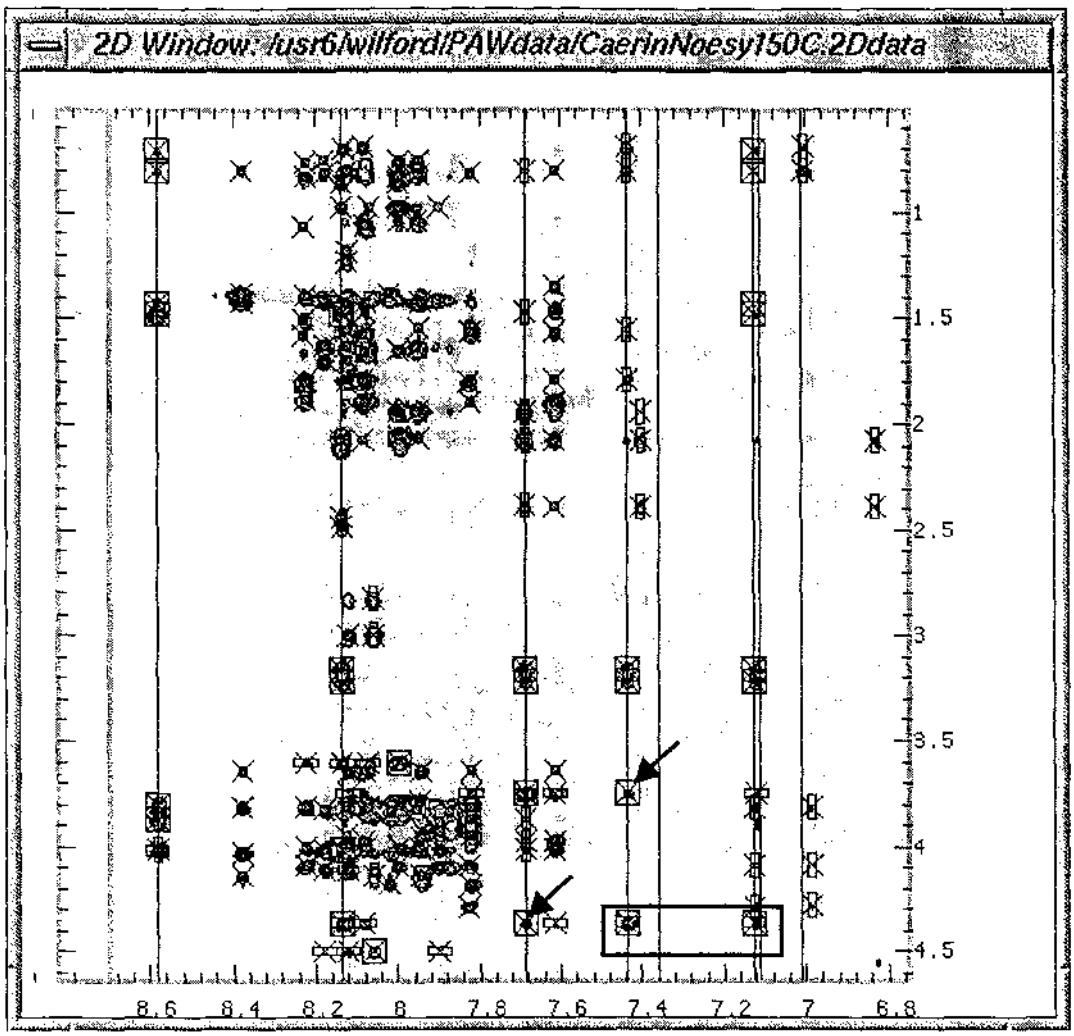


Figure 11.20: The cross-peaks in the upper-left region of the **CaerinNoesy150C** spectrum, with the symbols for the assigned cross-peaks plotted differently. The two fully-assigned cross-peaks that are blocked are W3:HE3-HA (left) and W3:HD1-HA (right), and two inter-residue cross-peaks arrowed are W3:HE3-Q4:HA (left) and Q4:HN-W3:HA (right). The peak labels are plotted later in larger plots.

As can be seen from the symbols in the last plot, some peaks are partly assigned, and some are fully assigned. Apart from the five intra-residue HN-HA peaks that have been assigned individually, there are four other cross-peaks that are fully assigned by following lining-up operations. These are the two intra-residue cross-peaks W3:HE3-HA and W3:HD1-HA that are on the right of the bottom row, as well as two inter-residue cross-peaks W3:HE3-Q4:HA and Q4:HN-W3:HA, as arrowed in the figure.

11.4.3 Identification of the adjacent residues

Starting from the cross-peaks of a known residue, most of the adjacent residues have been found by drawing crosshairs from a number of peaks in the HN-HA and HN-HN regions in the spectra. For example, the L2-related cross-peaks were identified from the intersections of the crosshairs drawn from W3-HN-HA and one of the HN-HA cross-peaks that belong to the Leu residues, as shown in Figure 11.21.

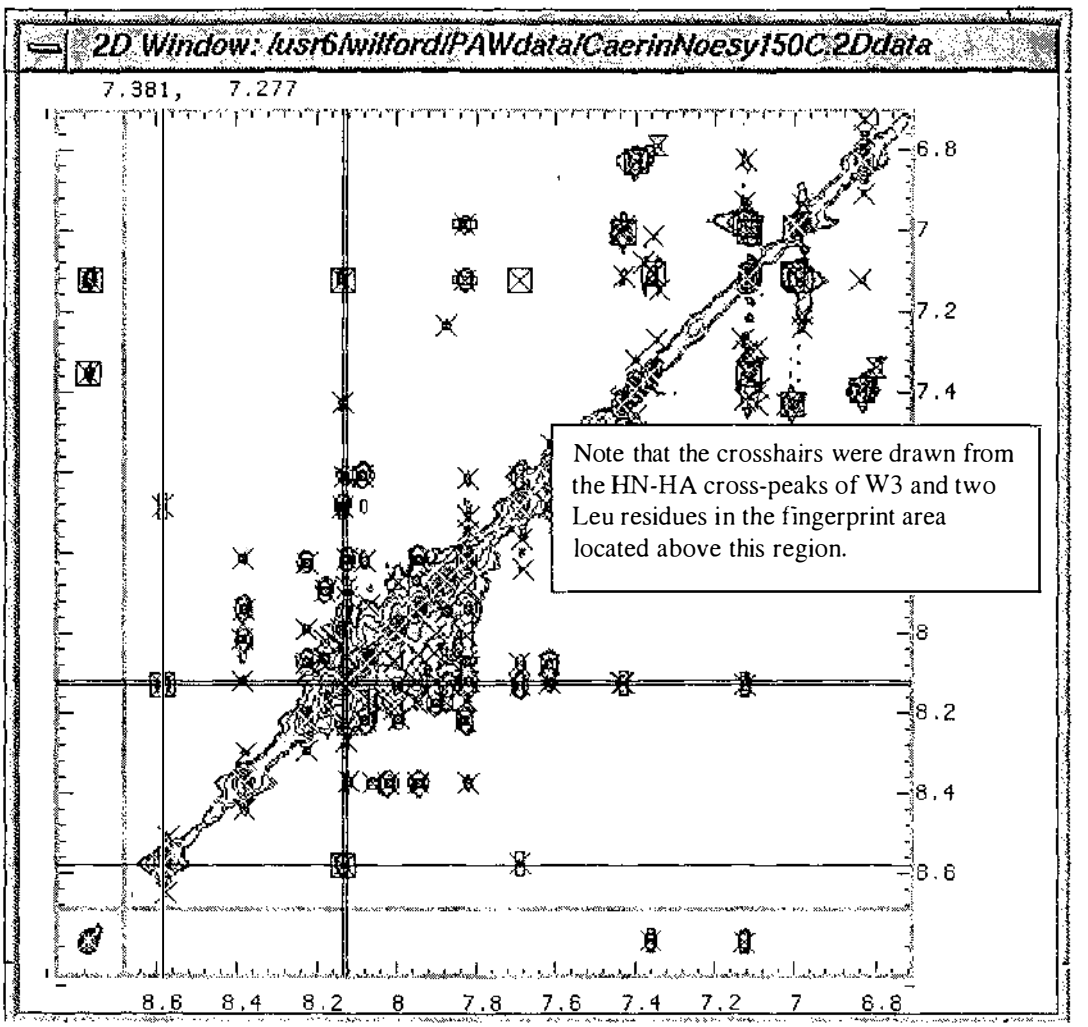


Figure 11.21: Three sets of crosshairs used to identify L2-related cross-peaks. The apparent cross-peaks at the intersections indicate clearly that the chemical shift of L2:HN is the one at around 8.6 ppm.

Only the lower-left region is presented here because the inter-residue cross-peaks between W3:HN-HA and either of the two HN-HA cross-peaks of Leu residues could not be unambiguously identified in the HN-HA region. The apparent cross-peaks at the intersections clearly indicate that the chemical shift of L2:HN is the one at around 8.6 ppm. (Hence, the chemical shift of L13:HN is also identified to be one around 8.12 ppm.)

The crosshairs in Figure 11.22 and 11.23 highlight all the L2 and W3-related cross-peaks in the lower-left and upper-left regions. Most of them are assigned by the lining-up operation.

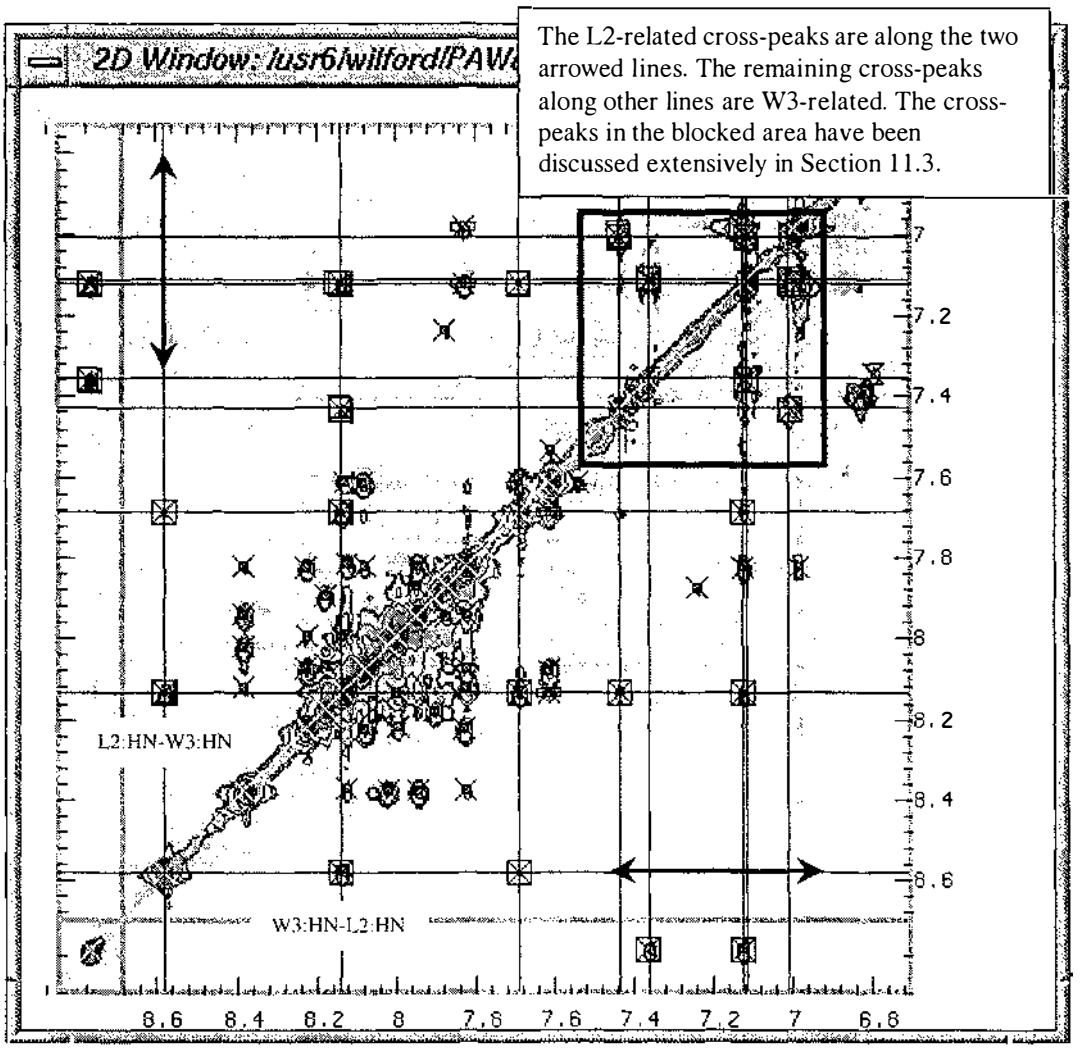


Figure 11.22: The crosshairs that highlight all the L2 and W3-related cross-peaks in the lower-left region.

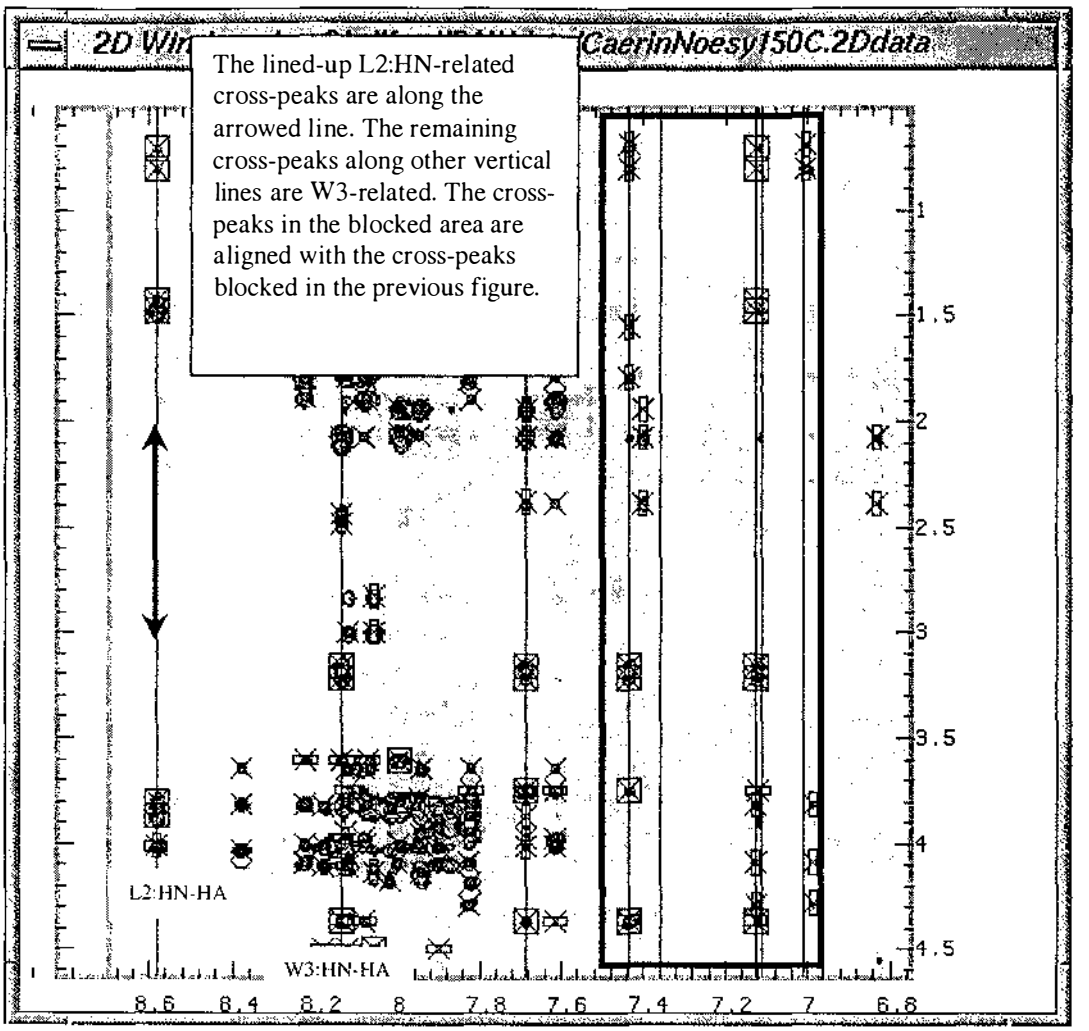


Figure 11.23: The crosshairs that highlight all the L2 and W3-related cross-peaks in the upper-left region.

Other inter-residue cross-peaks were found and assigned in the same way.

11.4.4 Sequential assignment using various peak-display options

Various peak-display options have been used in the sequential assignment of the Caerin 4.1 spectrum to assist the search for the inter-residue cross-peaks. These include six drawing-object display modes, ten label display formats, as well as the capability of displaying any of the 21 spin-systems and three peak-lists, as shown in the *Peak-display Toolbox* (Figure 11.24). In the toolbox, the different label display formats and peak-lists are in two separate exclusive-OR radio boxes. The code Z in the *SpinSystem Checkbox* is reserved for any unassigned spin system, and is not a standard amino-acid code.

Various plots displayed with a combination of peak-display formats and types have been used in the sequential assignment process to assist the search for the residue connectivity.

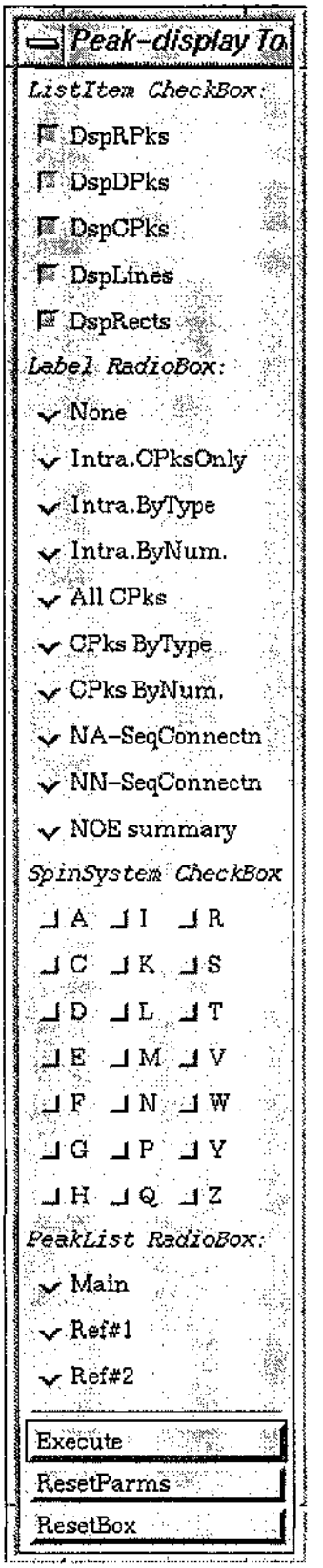


Figure 11.24: The *Peak-display Toolbox*.

The next figure shows the upper-left region of the **CaerinNoesy150C** spectrum. It is displayed with the symbols of raw peaks (in simple crosses) and cross-peaks (in crosses with rectangles). As can be seen from the symbols, the cross-peaks are nearly fully assigned. No labels were displayed in order to have a better overview of peak symbols.

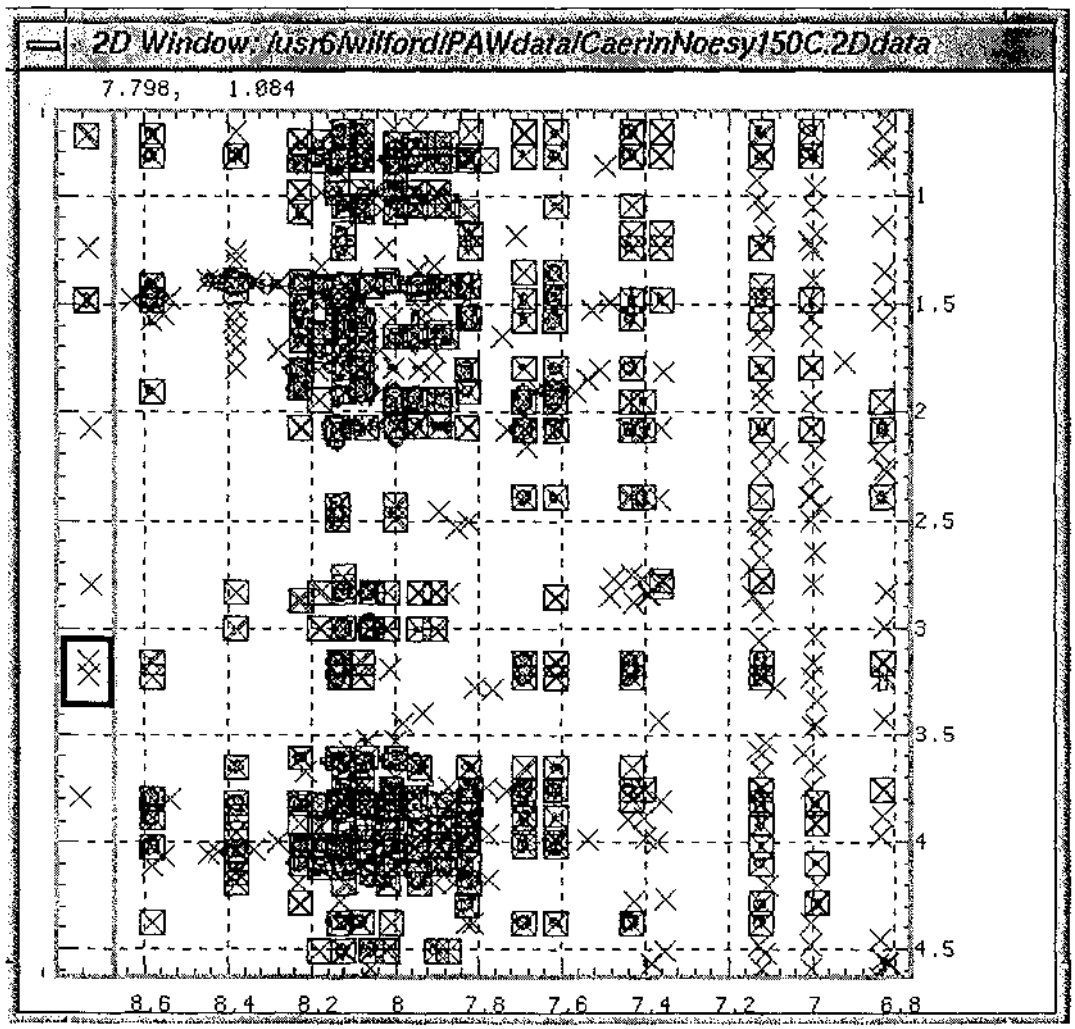


Figure 11.25: The upper-left region of the **CaerinNoesy150C** spectrum. This is displayed with a nearly fully assigned cross-peak list and a raw peak list from which the cross-peaks were selected. The label format for both plots was set to "none" in order to have a better view of all peak symbols.

There are many more raw peaks in this plot than in previous ones, because the peak lists used in the final assignment were picked with a much lower base level. It can be seen that a number of raw peaks were not in the cross-peak list. Most of them are very weak and are just noise. However, some of them were real and hence were added to the cross-peak list and assigned.

For example, the two raw peaks on the left at about D2=3.15 ppm were added to the final cross-peak list because they are indeed real, although they were very weak and less important for structure determination.

The next figure is similar to the previous plot, but without raw peaks drawn.

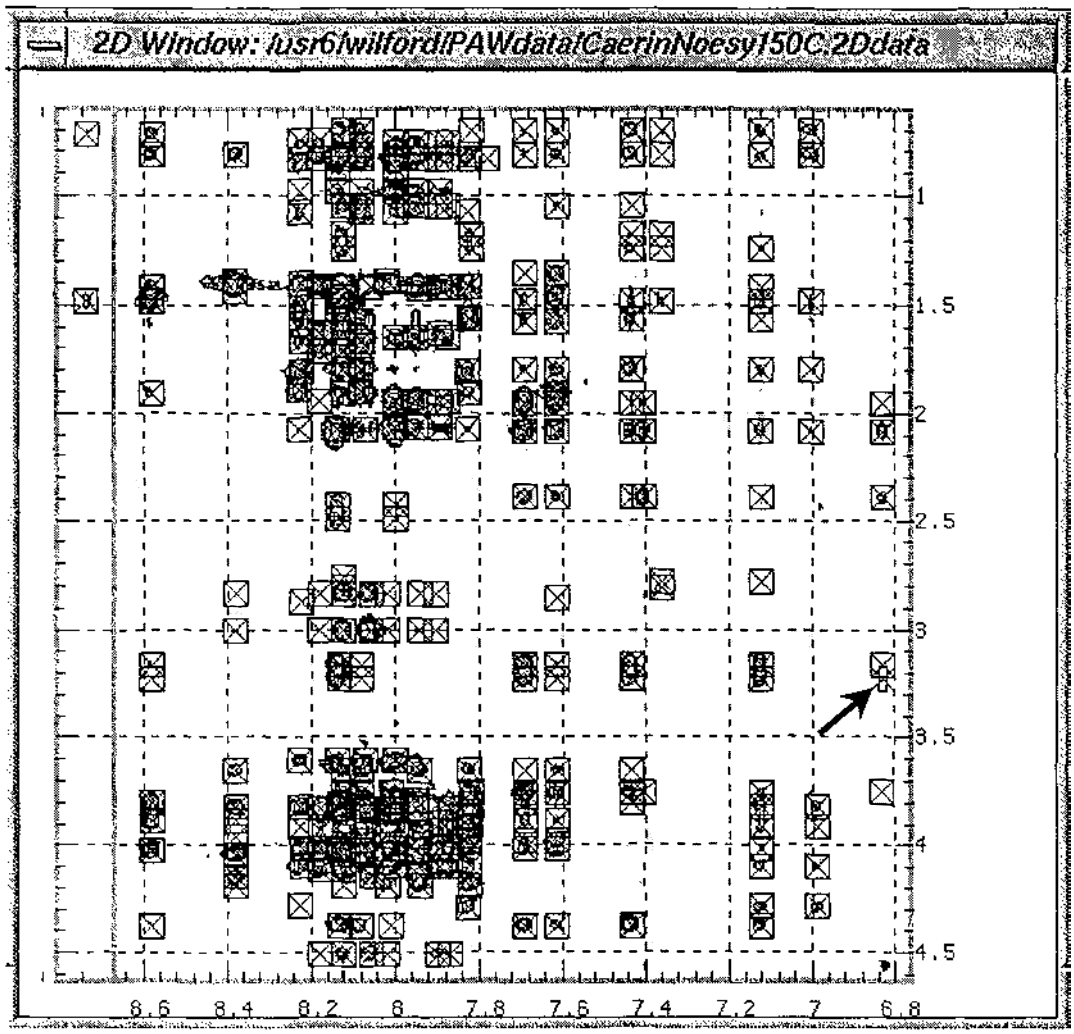


Figure 11.26: The same region as in the last figure, but with raw-peak symbols suppressed. The plot clearly shows that all of the strong cross-peaks have been fully assigned, except for one arrowed at around 3.2 ppm.

The plot gives a tidy view of the assigned cross-peaks in the region. It can be seen that all of the strong cross-peaks have been fully assigned, except for one arrowed at around 3.2 ppm. The peak was subsequently fully assigned by lining it up with other peaks in the horizontal direction.

The next figure labels only intra-residue cross-peaks in the HN-HA region of the **CaerinNoesy150C** spectrum. No raw peak symbols are displayed in the plot. The simple crosses without rectangles represent the inter-residue cross-peaks.

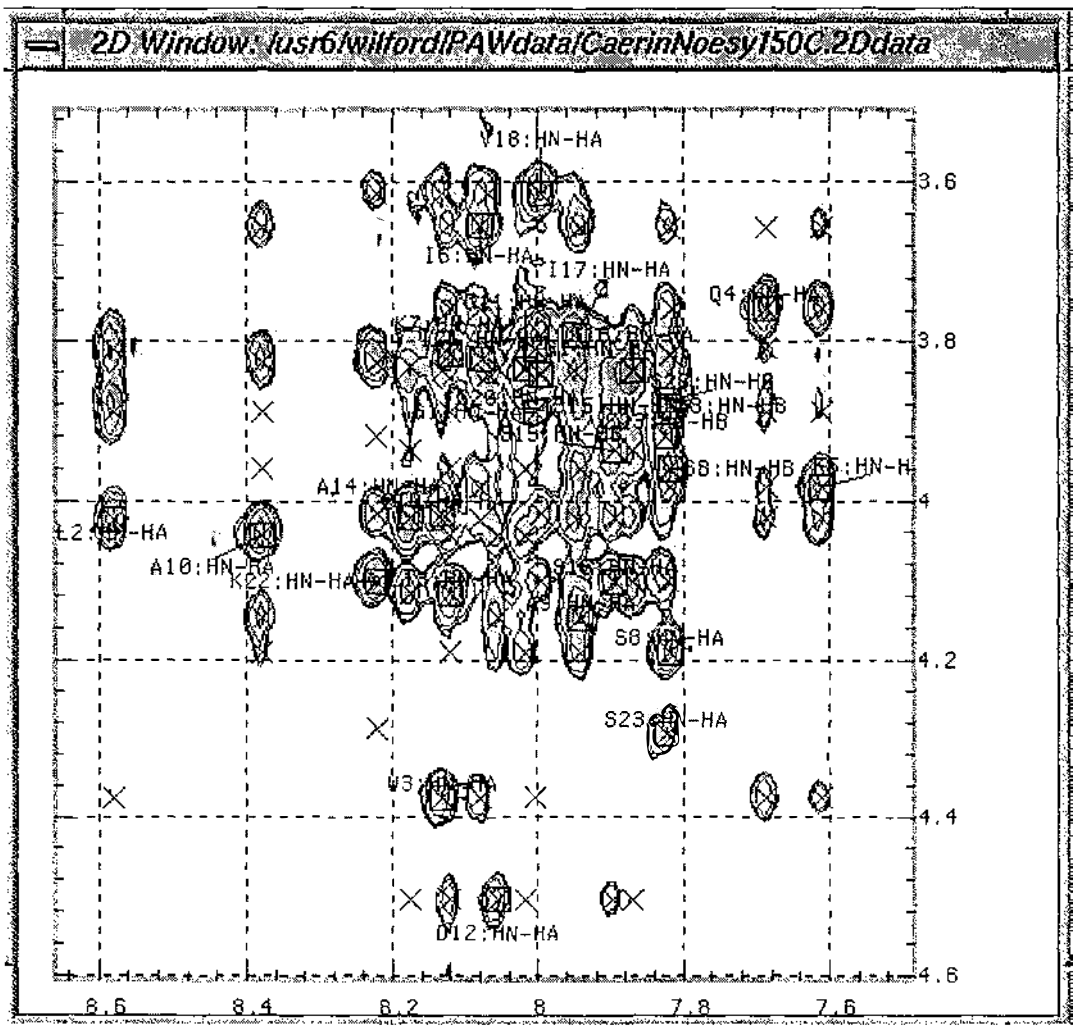


Figure 11.27: The intra-residue cross-peak labels displayed in the HN-HA region of the **CaerinNoesy150C** spectrum. No raw peak is displayed in the plot. The simple crosses without rectangles represent the inter-residue cross-peaks.

This display mode clearly distinguished the intra-residue cross-peaks in a **CaerinNoesy150C** spectrum from the inter-residue ones, especially when the plot is displayed with a larger draw window.

The next figure displays the same intra-residue NOE cross-peaks labels on top of the **CaerinTocsy070C** spectrum.

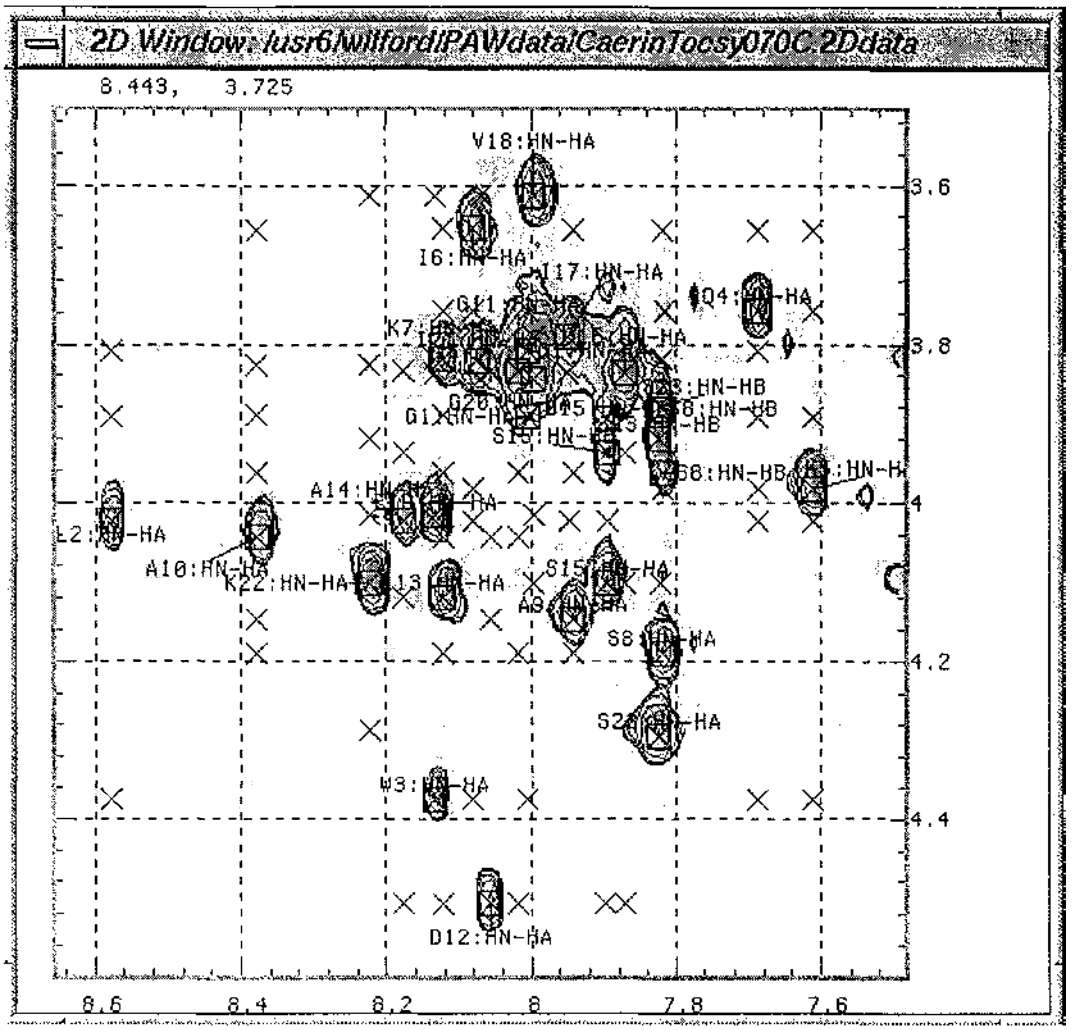


Figure 11.28: The intra-residue cross-peak labels displayed in the HN-HA region of the **CaerinTocsy070C** spectrum. No raw peak is displayed in the plot. The simple crosses without rectangles represent the inter-residue cross-peaks.

This plot provides a cleaner display for finding any mistakes in the assignment. Every intra-residue cross-peak must be seen with a TOCSY peak at the background. The plots displayed in this mode have been used during the sequential assignment to ensure the correctness of cross-peak assignment.

The next figure displays the intra-residue cross-peaks of Q and K in the HN-HA region of the **CaerinNoesy150C** spectrum. Three of the cross-peaks are K-related.

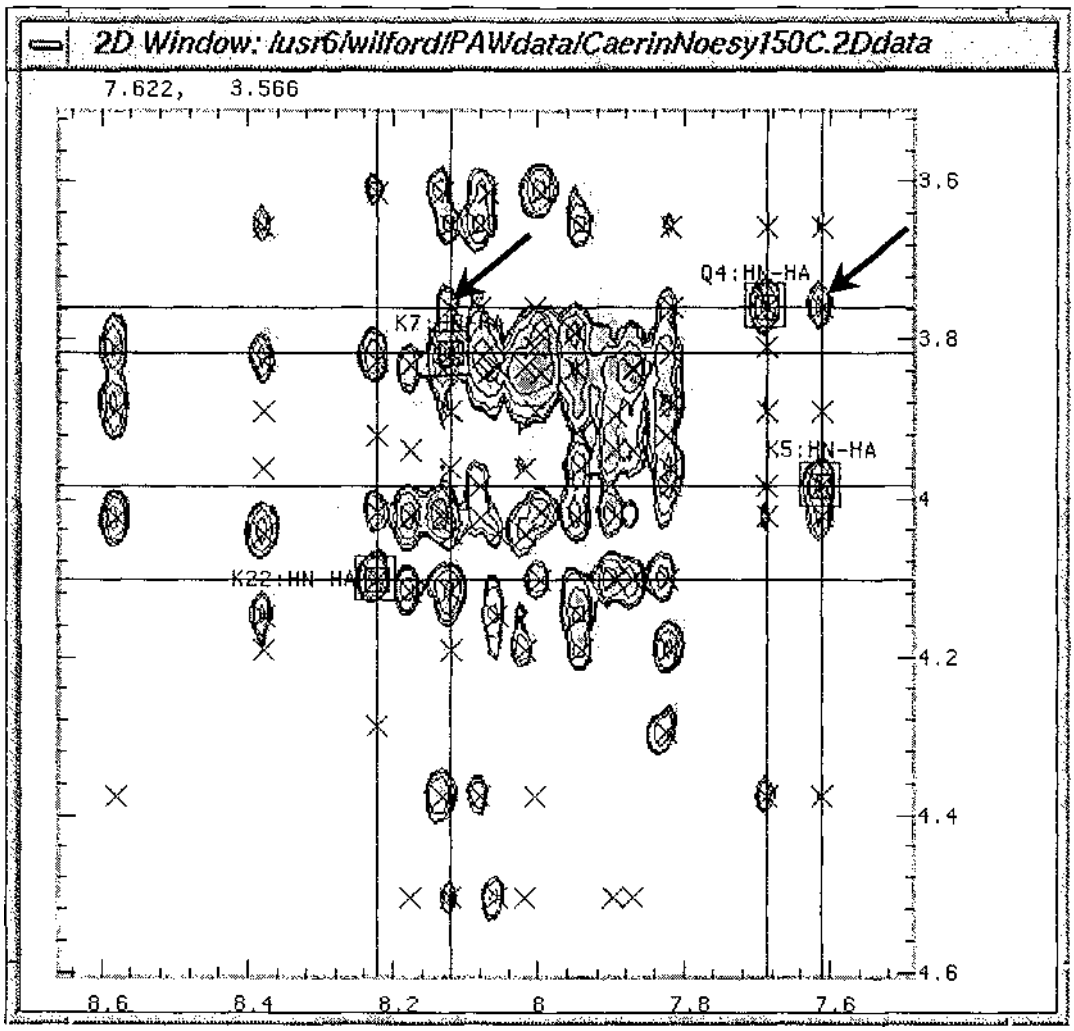


Figure 11.29: The intra-residue cross-peaks of Q and K residues in the HN-HA region of the **CaerinNoesy150C** spectrum. (Note that this is plotted with a higher first contour level in order to reduce the overlap.)

This plot distinguishes K22:HN-HA from K5:HN-HA, because there is an inter-residue cross-peak that connects K5:HN-HA with Q4:HN-HA. Either of the other two K-related HN-HA cross-peaks could be K5:HN-HA, because Q4:HN-HA can be connected to both, as arrowed. Eventually, it was the unambiguous connection between I6:HN-HA and K7:HN:HA that clarified the assignment of K7:HN-HA, leaving K5:HN-HA unambiguously identified, as shown in Figure 11.31.

The next figure displays all K5-related cross-peaks in the HN-HA region of the **CaerinNoesy150C** spectrum.

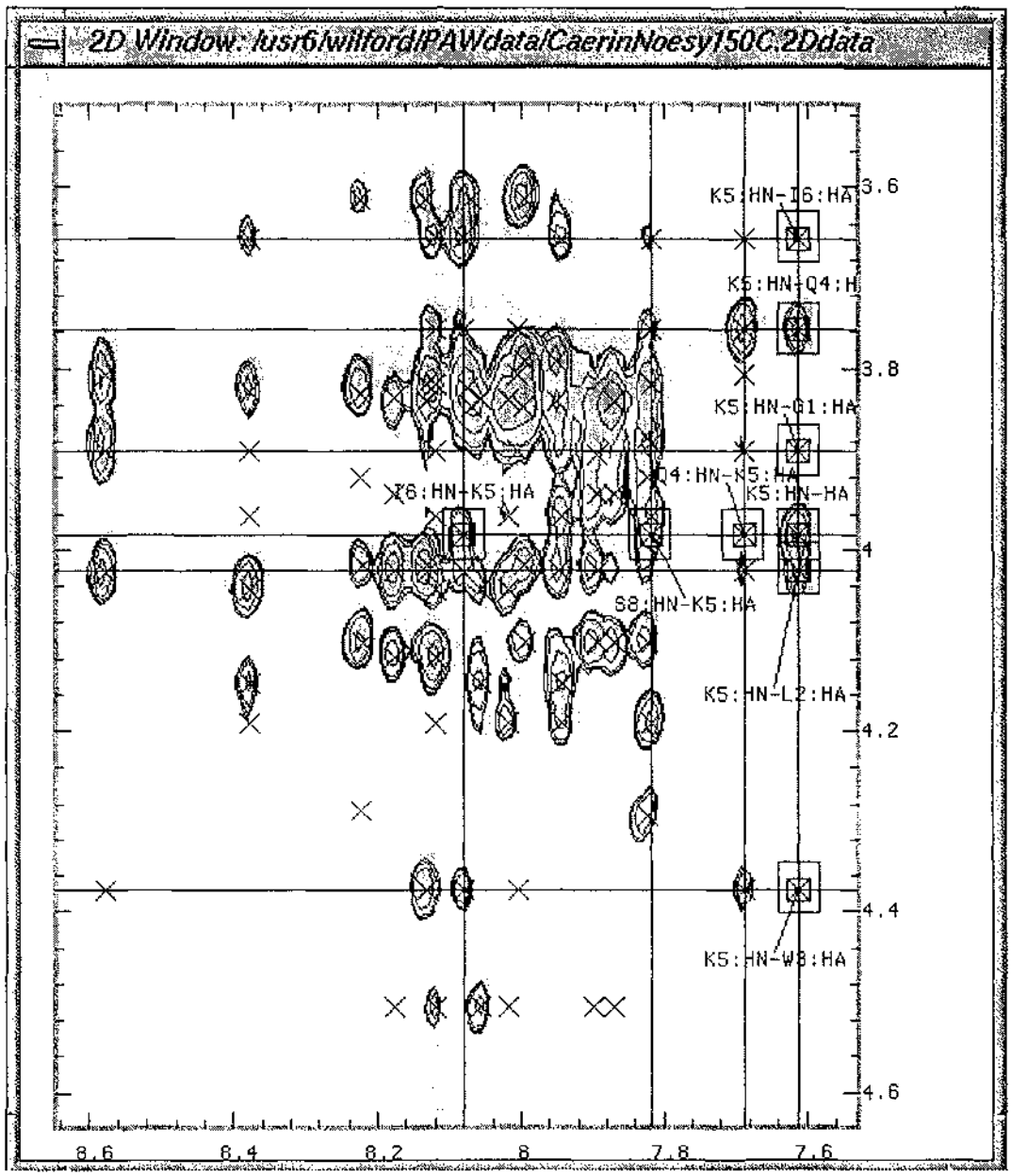


Figure 11.30: All K5-related cross-peaks in the HN-HA region of the **CaerinNoesy150C** spectrum.

This plot provides a clear view of the K5-connectivity and the helical structure between K5 and S8 (see Table 7.5). The crosshairs and rectangles in the plot were drawn automatically to show all intra-residue cross peaks from residue 5 to 8.

The next figure displays the NOE connectivity from L2 to S8 in the HN-HA region of the **CaerinNoesy150C** spectrum.

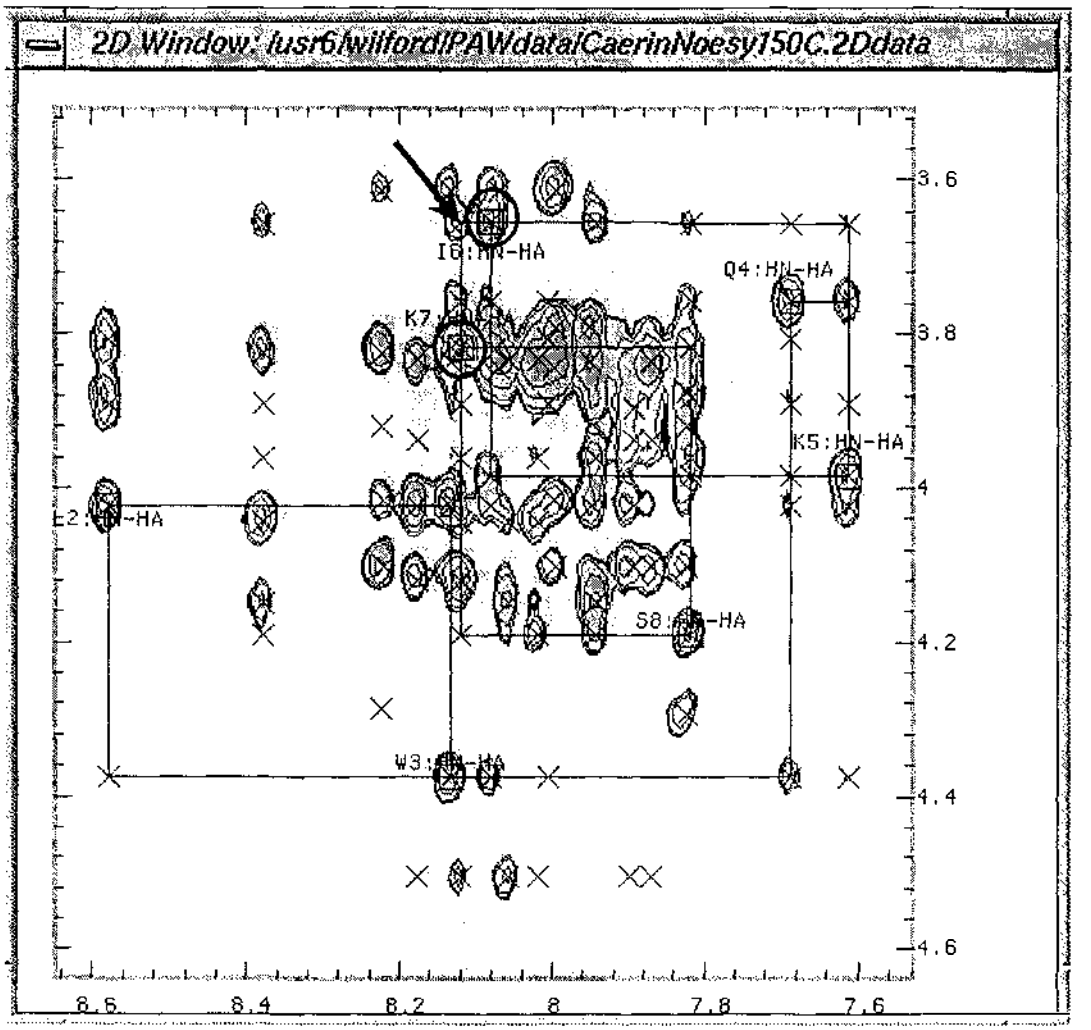


Figure 11.31: The NOE connectivity from L2 to S8 in the HN-HA region of the **CaerinNoesy150C** spectrum. The arrow shows the connectivity between I6:HN-HA and K7:HN-HA (circled).

This plot provides a clear view of the connectivities from L2 to S8 in the HN-HA region. The lines in the plot were drawn automatically to highlight the cross-peaks involved. (See Chapter 11, Volume II.)

The next figure displays the NOE connectivity from L2 to S8 in the HN-HN region of the **CaerinNoesy150C** spectrum.

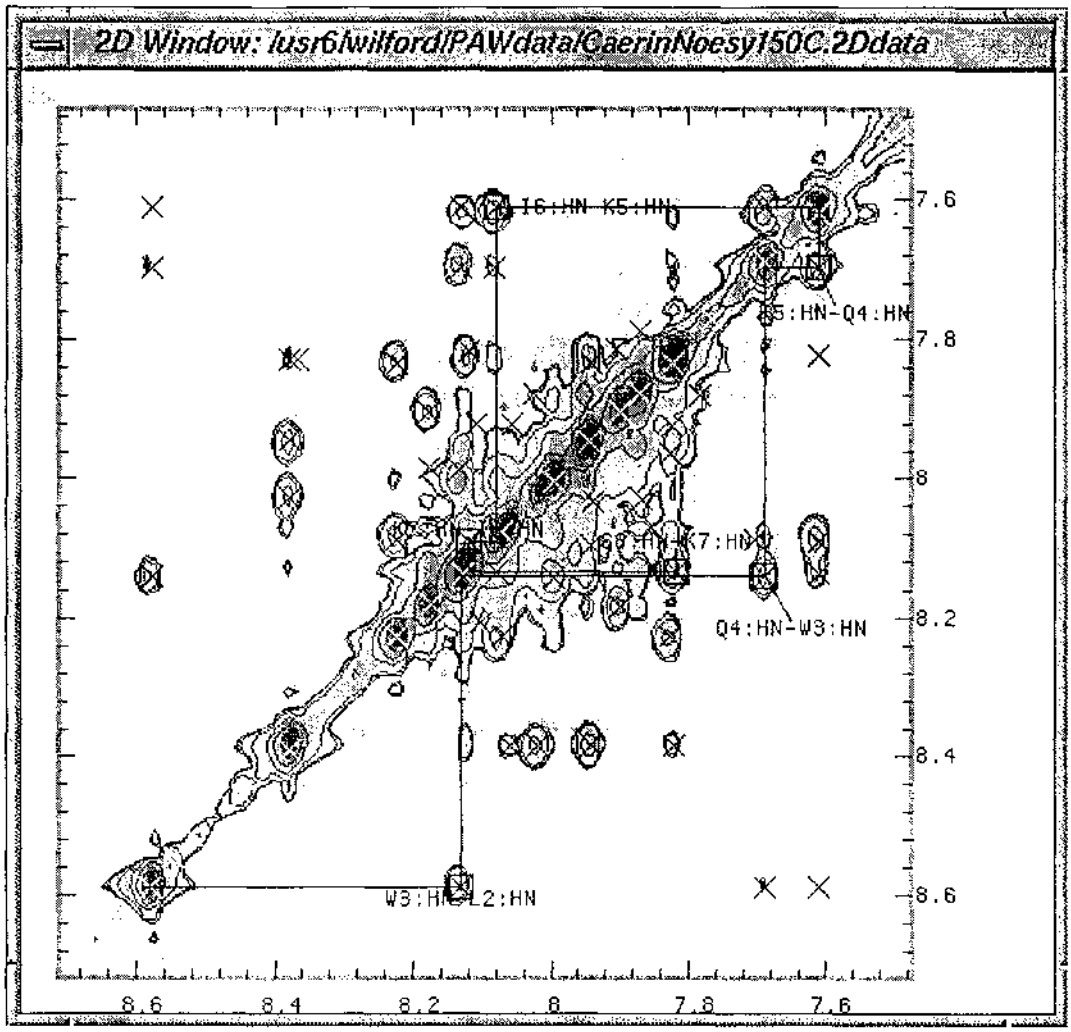


Figure 11.32: The NOE connectivity from L2 to S8 in the HN-HN region of the **CaerinNoesy150C** spectrum.

This plot provides a clear view of the connectivity from L2 to S8 in the HN-HN region. Again, the lines in the plot were drawn automatically.

The next figure displays a brief NOE summary from G1 to S24.

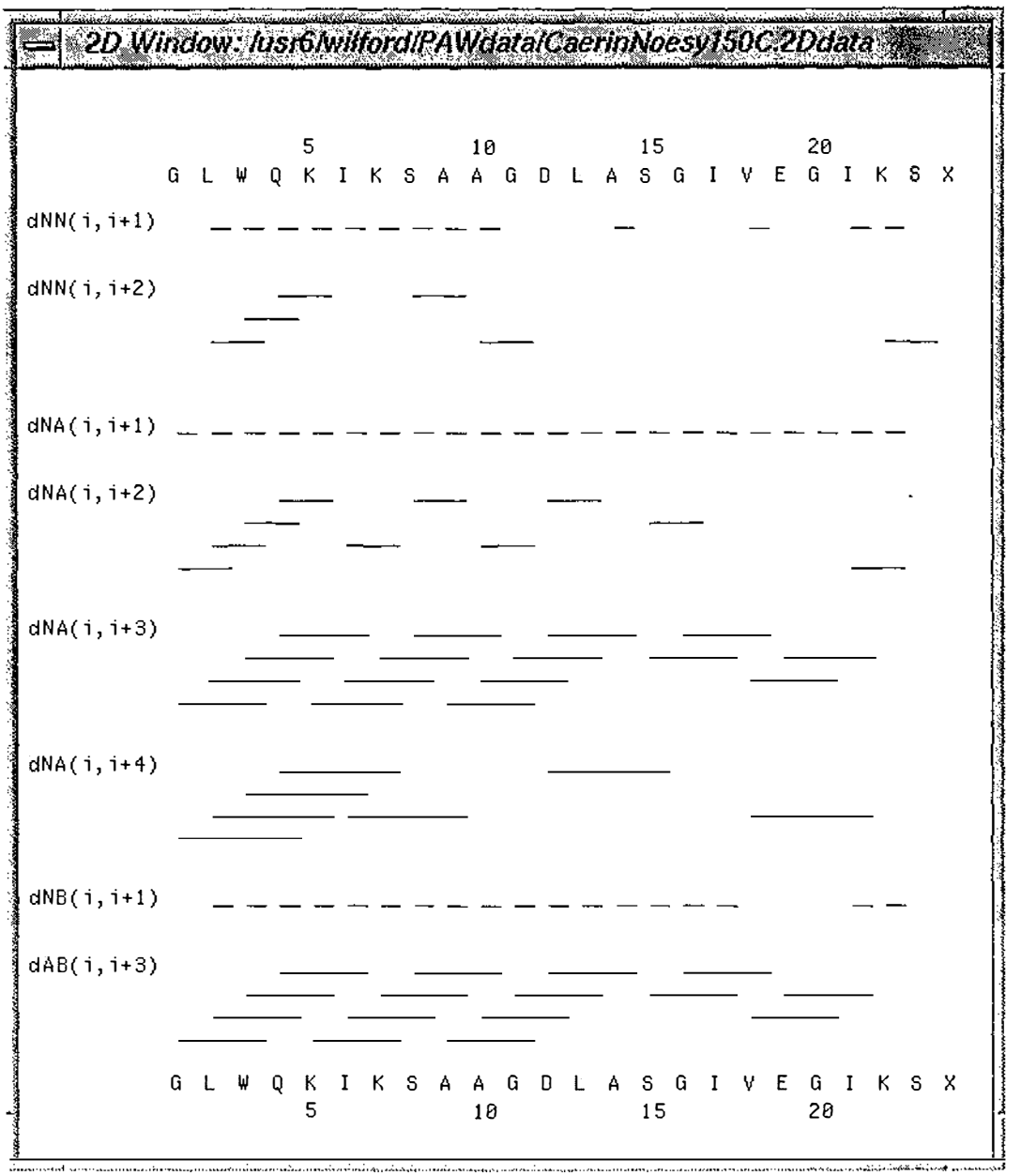


Figure 11.33: The NOE summary from G1 to S24. The notations on the right correspond to NOE distances in Table 7.5.

This plot provides a clear view of the secondary structure. Again, the lines in the plot were drawn automatically from the assignment data. Each of the lines in the summary indicates the existence of an inter-residue cross-peak that connect residues at the end of each line. The summary clearly indicates the presence of a long helical structure from G1 to D12 (see Table 7.5), followed by two single loop helixes separated by two turns at around S15 and E19.

Note that because the summary is designed to be created at any stage during the assignment, the information on peak intensities is not shown. (Peak intensity can only be obtained at the final stage after all cross-peaks are assigned.).

11.5 Exporting Spectral Assignment Results

This section presents the assignment results with several PostScript plots and a complete list of the assigned cross-peaks.

11.5.1 The PostScript plots of the final assignment results

The rectangles in the **CaerinNoesy150** full spectrum in Figure 11.34 identify the regions plotted in the next six figures.

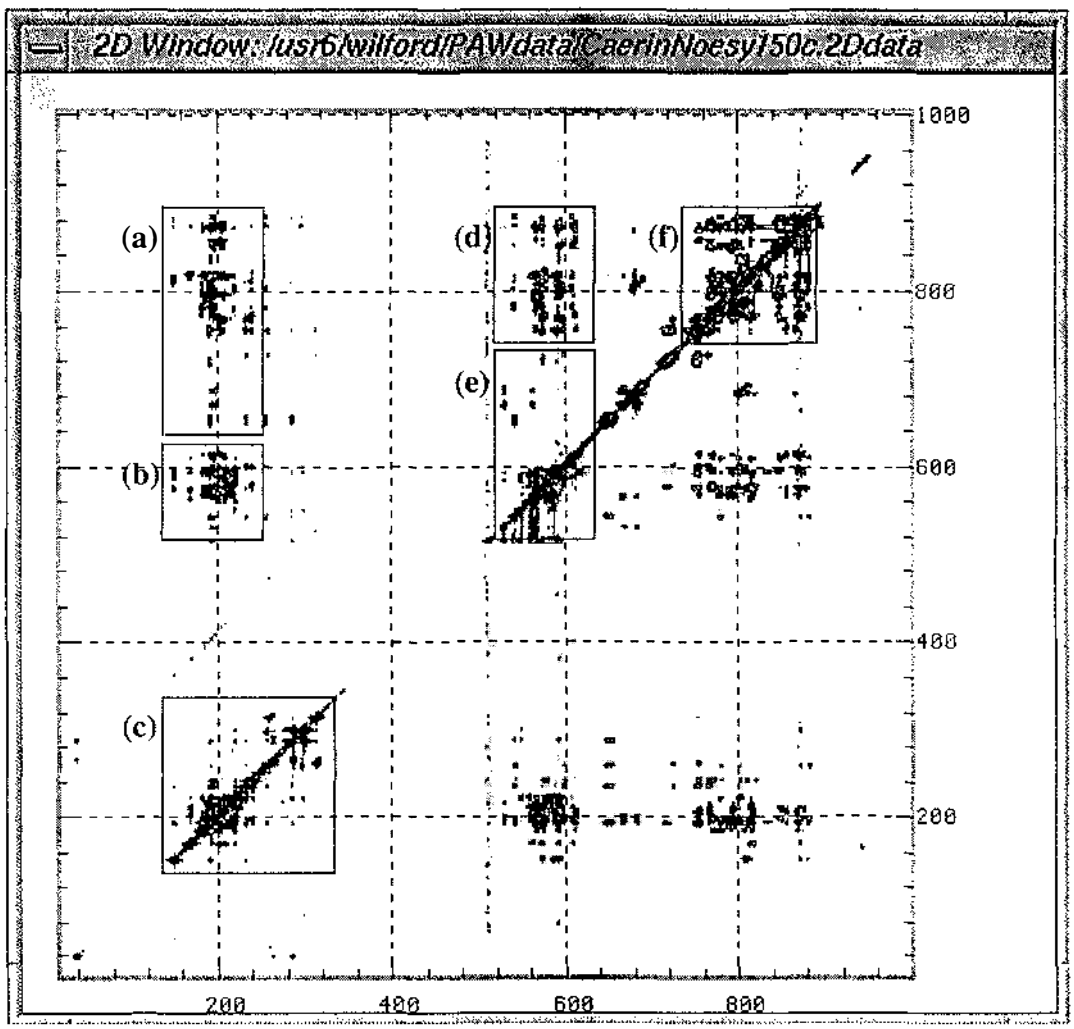


Figure 11.34: The corresponding regions of the next six plots on the **CaerinNoesy150C** full spectrum.

The contents of the plots are explained in the footnotes to each plot. To avoid confusion, only the labels for the intra-residue cross-peaks were plotted, where the simple crosses without labels are inter-residue cross-peaks. For Figure 11.38, the labels of all cross-peaks were plotted in order to show the HN-HN connectivity. The hourglass-like symbols indicate unassigned cross-peaks, which are either too weak or meaningless.

The region from W3:HN-HB to L2:HN-HD of Caerin Noesy150

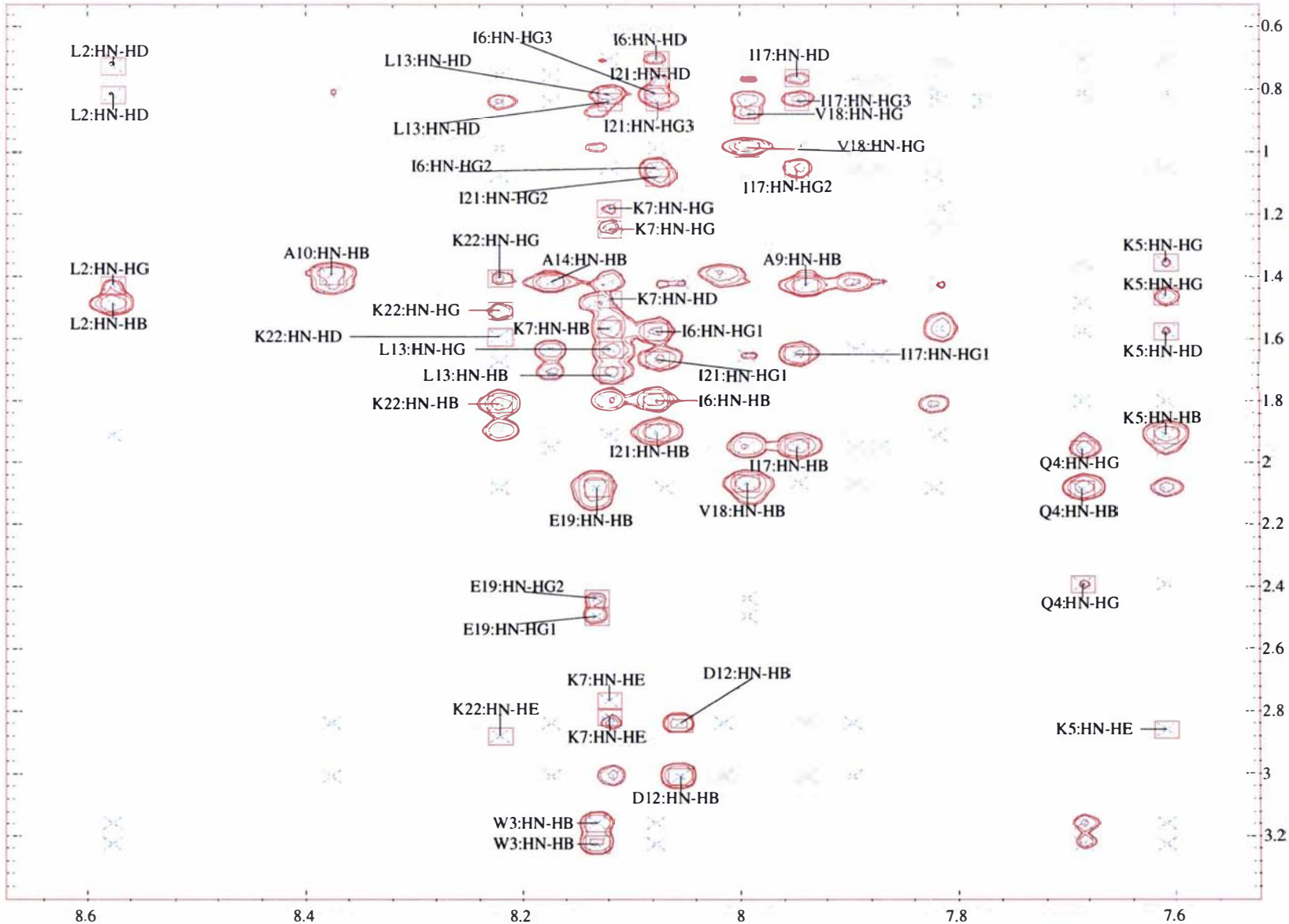


Figure 11.35

The region from Q4_HG-HB to L2:HD-HD of Caerin Noesy150

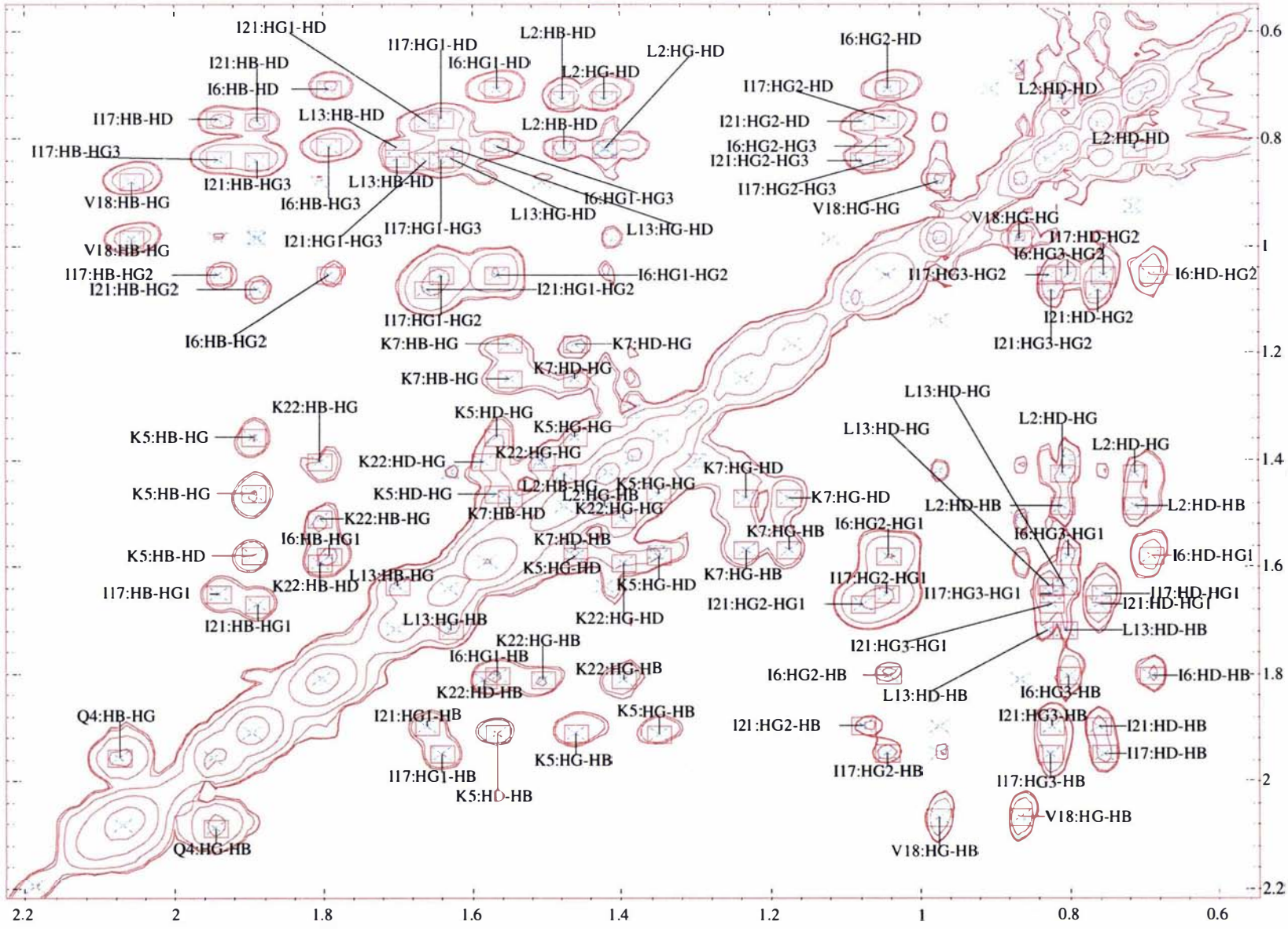


Figure 11.36

The region from Q4:HA-HB to L2:HA-HD of Caerin Noesy150

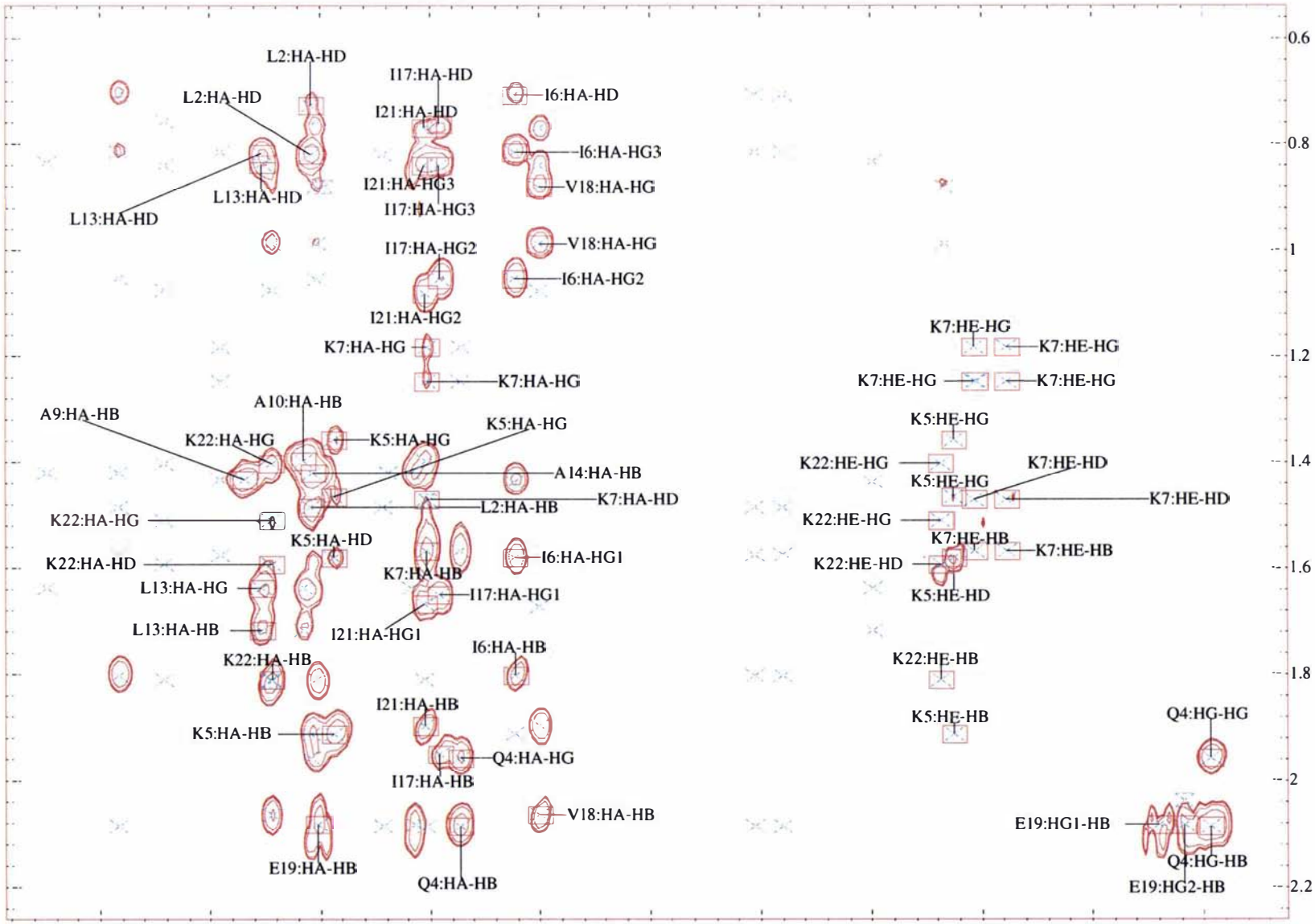


Figure 11.37

The HN-HN and aromatic region of Caerin Noesy150

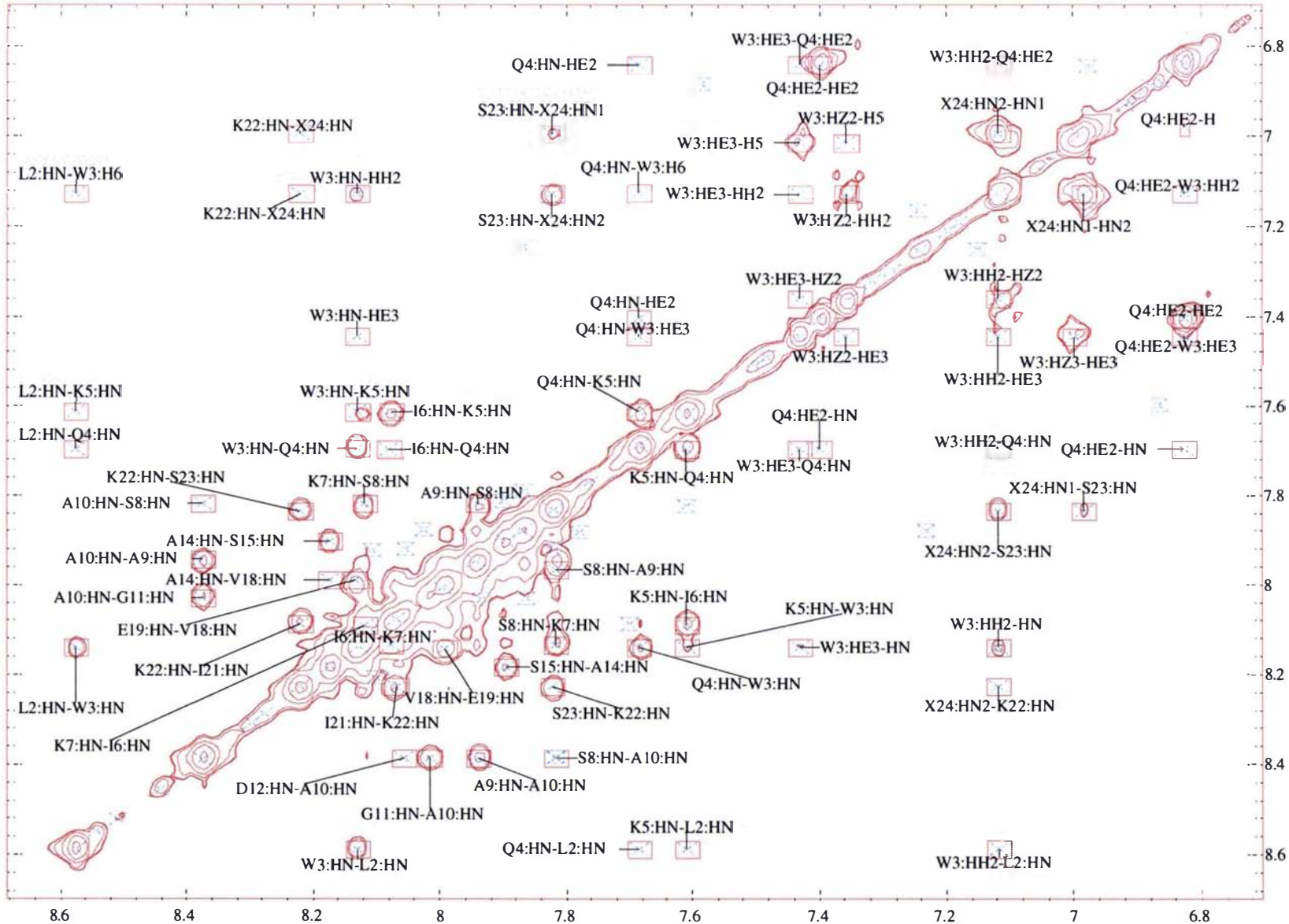


Figure 11.38

The fingerprint region of Caerin Noesy150

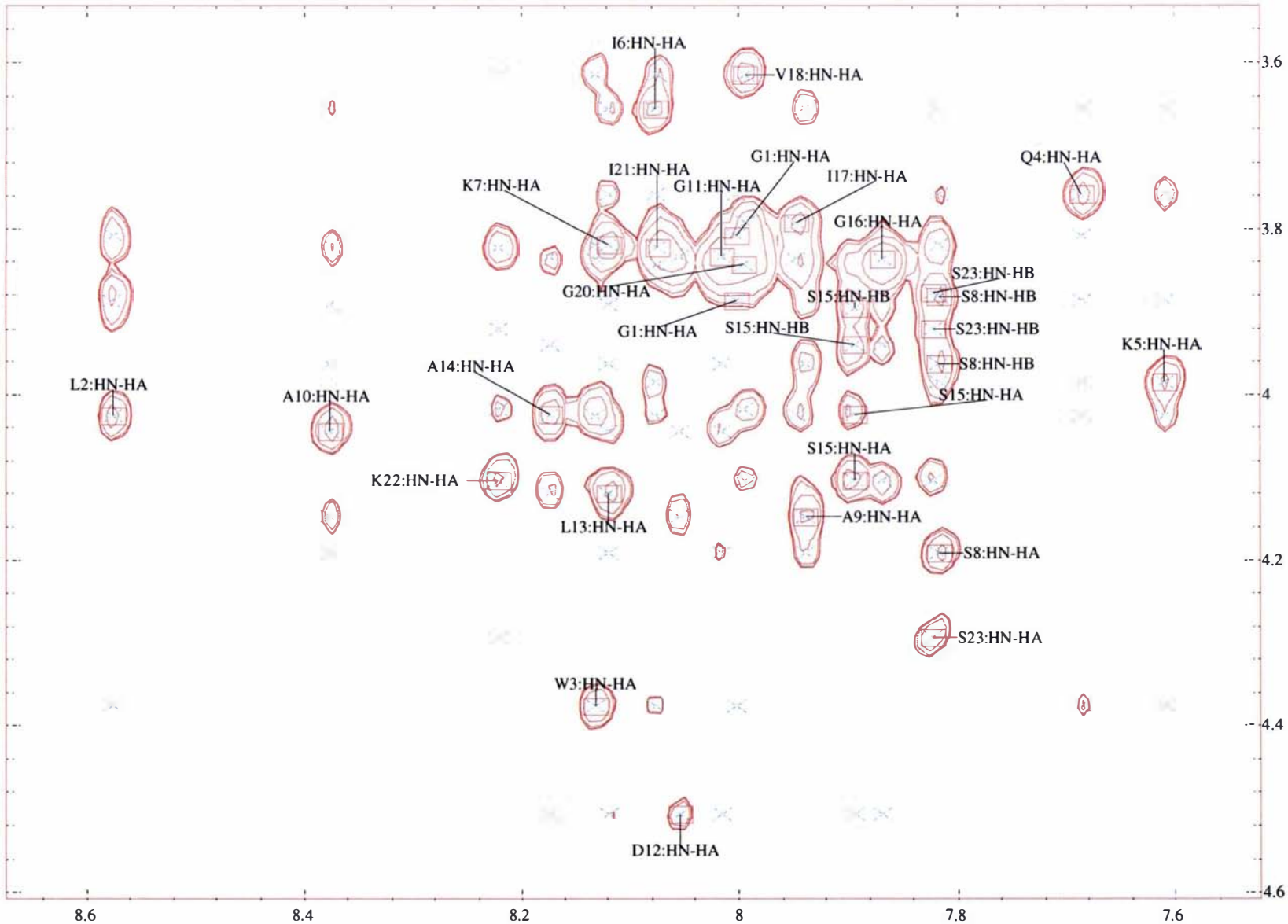


Figure 11.39

The region from D12:HA-HB to Q4:HG-HA of Caerin Noesy150

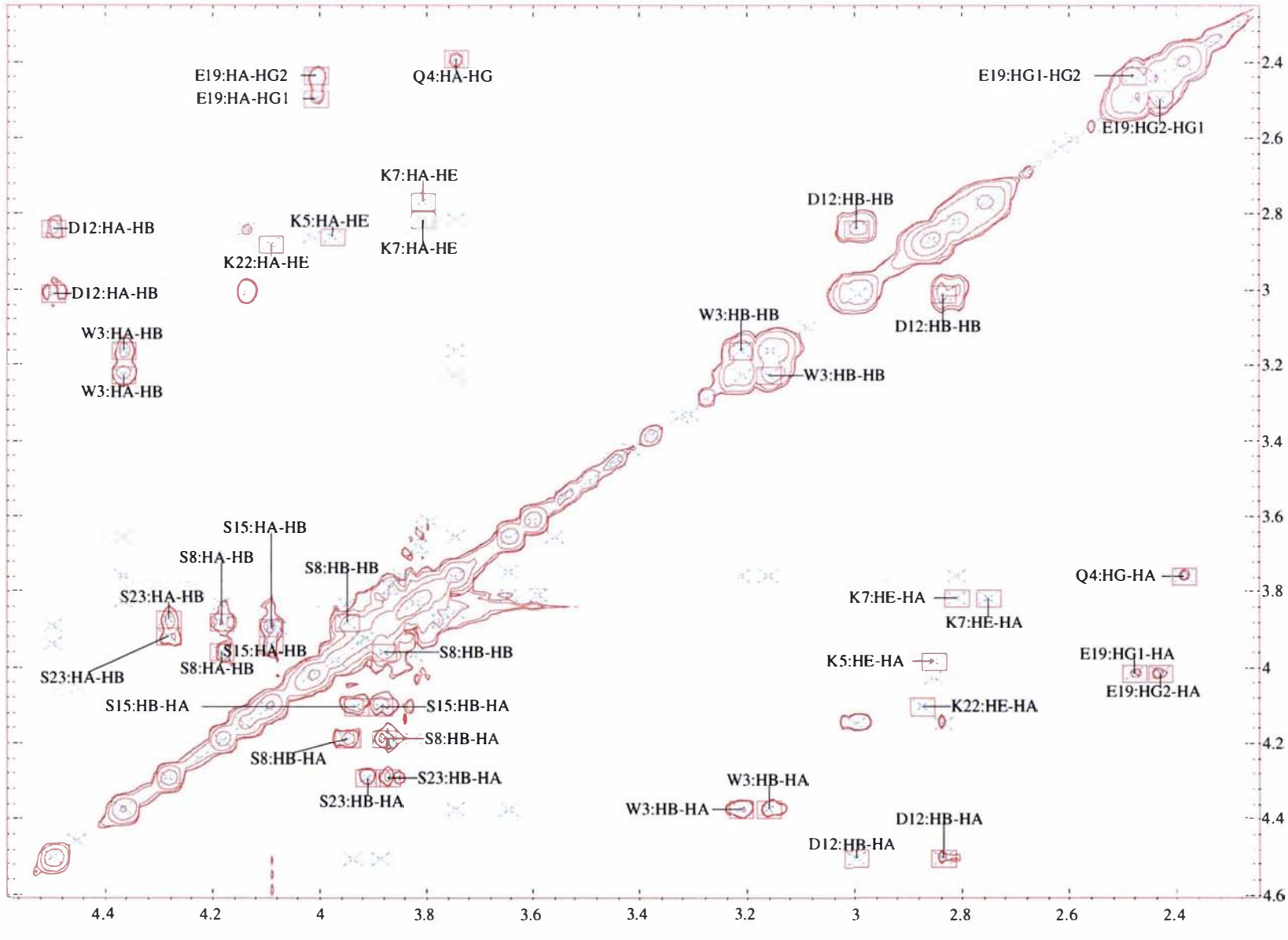


Figure 11.40

11.5.2 The assigned cross-peak list

The following cross-peak list was created by PAW as the final result of the spectral assignment for the **CaerinNoesy150C** spectrum. The first two columns are the chemical shifts[2] in D1 and D2; and the last two columns are the assignment in D1 and D2.

δ in D1	δ in D2	Assignment in D1	Assignment in D2
3.809,	8.013,	G1:HA2,	G1:HN
3.887,	8.012,	G1:HA1,	G1:HN
8.003,	3.809,	G1:HN,	G1:HA2
8.003,	3.887,	G1:HN,	G1:HA1
3.887,	1.487,	G1:HA1,	L2:HB
3.887,	0.819,	G1:HA1,	L2:HD
3.809,	8.588,	G1:HA2,	L2:HN
3.809,	7.446,	G1:HA2,	W3:H4
8.003,	4.377,	G1:HN,	W3:HA
3.809,	2.087,	G1:HA2,	Q4:HB
3.887,	2.087,	G1:HA1,	Q4:HB
3.809,	7.697,	G1:HA2,	Q4:HN
3.887,	7.697,	G1:HA1,	Q4:HN
8.003,	3.759,	G1:HN,	Q4:HA
3.887,	7.613,	G1:HA1,	K5:HN
8.577,	3.887,	L2:HN,	G1:HA1
8.577,	3.809,	L2:HN,	G1:HA2
4.017,	1.487,	L2:HA,	L2:HB
4.017,	0.727,	L2:HA,	L2:HD
4.017,	0.819,	L2:HA,	L2:HD
4.017,	8.589,	L2:HA,	L2:HN
1.477,	4.026,	L2:HB,	L2:HA
1.477,	0.819,	L2:HB,	L2:HD
1.477,	0.727,	L2:HB,	L2:HD
1.477,	1.427,	L2:HB,	L2:HG
1.477,	8.589,	L2:HB,	L2:HN
0.714,	4.026,	L2:HD2,	L2:HA
0.810,	4.026,	L2:HD1,	L2:HA
0.714,	1.487,	L2:HD2,	L2:HB
0.810,	1.487,	L2:HD1,	L2:HB
0.714,	0.819,	L2:HD2,	L2:HD
0.810,	0.727,	L2:HD1,	L2:HD
0.714,	1.427,	L2:HD2,	L2:HG
0.810,	1.427,	L2:HD1,	L2:HG
0.714,	8.589,	L2:HD2,	L2:HN
0.810,	8.589,	L2:HD1,	L2:HN
1.421,	4.026,	L2:HG,	L2:HA
1.421,	1.487,	L2:HG,	L2:HB
1.421,	0.727,	L2:HG,	L2:HD
1.421,	0.819,	L2:HG,	L2:HD
1.421,	8.589,	L2:HG,	L2:HN
8.577,	4.026,	L2:HN,	L2:HA
8.577,	1.487,	L2:HN,	L2:HB
8.577,	0.819,	L2:HN,	L2:HD
8.577,	0.727,	L2:HN,	L2:HD
8.577,	1.427,	L2:HN,	L2:HG
8.577,	7.128,	L2:HN,	W3:H6
8.577,	4.377,	L2:HN,	W3:HA
8.577,	3.163,	L2:HN,	W3:HB2
8.577,	3.229,	L2:HN,	W3:HB1
8.577,	8.140,	L2:HN,	W3:HN
4.017,	1.958,	L2:HA,	Q4:HG2

[2] In units of ppm.

```
4.017,   7.697,   L2:HA,      Q4:HN
1.477,   7.697,   L2:HB,      Q4:HN
8.577,   7.697,   L2:HN,      Q4:HN
4.017,   1.914,   L2:HA,      K5:HB1
4.017,   2.862,   L2:HA,      K5:HE
4.017,   7.613,   L2:HA,      K5:HN
8.577,   1.914,   L2:HN,      K5:HB1
8.577,   7.613,   L2:HN,      K5:HN
7.431,   3.809,   W3:HE3,     C1:HA2
4.367,   1.487,   W3:HA,      L2:HB
4.367,   8.589,   W3:HA,      L2:HN
3.160,   1.487,   W3:HB2,     L2:HB
3.213,   1.487,   W3:HB1,     L2:HB
3.160,   8.589,   W3:HB2,     L2:HN
3.213,   8.589,   W3:HB1,     L2:HN
7.121,   1.487,   W3:HD1,     L2:HB
7.121,   0.819,   W3:HD1,     L2:HD
7.121,   1.427,   W3:HD1,     L2:HG
7.431,   1.487,   W3:HE3,     L2:HB
9.790,   1.487,   W3:HE1,     L2:HB
9.790,   0.727,   W3:HE1,     L2:HD
7.121,   8.589,   W3:HH2,     L2:HN
8.131,   4.026,   W3:HN,      L2:HA
8.131,   1.487,   W3:HN,      L2:HB
8.131,   8.589,   W3:HN,      L2:HN
6.999,   1.487,   W3:HZ3,     L2:HB
7.358,   1.487,   W3:HZ2,     L2:HB
7.358,   0.819,   W3:HZ2,     L2:HD
4.367,   3.163,   W3:HA,      W3:HB2
4.367,   3.229,   W3:HA,      W3:HB1
4.367,   7.446,   W3:HA,      W3:HE3
4.367,   7.128,   W3:HA,      W3:HH2
4.367,   8.140,   W3:HA,      W3:HN
3.213,   4.377,   W3:HB1,     W3:HA
3.160,   4.377,   W3:HB2,     W3:HA
3.160,   3.229,   W3:HB2,     W3:HB1
3.213,   3.163,   W3:HB1,     W3:HB2
3.213,   7.446,   W3:HB1,     W3:HE3
3.160,   7.446,   W3:HB2,     W3:HE3
3.160,   7.128,   W3:HB2,     W3:HH2
3.213,   7.128,   W3:HB1,     W3:HH2
3.213,   8.140,   W3:HB1,     W3:HN
3.160,   8.140,   W3:HB2,     W3:HN
7.121,   4.377,   W3:HD1,     W3:HA
7.121,   3.229,   W3:HD1,     W3:HB1
7.121,   3.163,   W3:HD1,     W3:HB2
7.431,   7.016,   W3:HE3,     W3:H5
7.431,   4.377,   W3:HE3,     W3:HA
7.431,   3.163,   W3:HE3,     W3:HB2
7.431,   3.229,   W3:HE3,     W3:HB1
7.431,   9.788,   W3:HE3,     W3:HE1
9.790,   7.446,   W3:HE1,     W3:HE3
7.431,   7.128,   W3:HE3,     W3:HH2
9.790,   7.128,   W3:HE1,     W3:HH2
7.431,   8.140,   W3:HE3,     W3:HN
9.790,   8.140,   W3:HE1,     W3:HN
7.431,   7.363,   W3:HE3,     W3:HZ2
9.790,   7.363,   W3:HE1,     W3:HZ2
7.121,   7.446,   W3:HH2,     W3:HE3
7.121,   9.788,   W3:HH2,     W3:HE1
7.121,   8.140,   W3:HH2,     W3:HN
7.121,   7.363,   W3:HH2,     W3:HZ2
8.131,   4.377,   W3:HN,      W3:HA
8.131,   3.229,   W3:HN,      W3:HB1
```

```
8.131,   3.163,   W3:HN,      W3:HB2
8.131,   9.788,   W3:HN,      W3:HE1
8.131,   7.446,   W3:HN,      W3:HZ3
8.131,   7.128,   W3:HN,      W3:HH2
7.358,   7.016,   W3:HZ2,     W3:H5
6.999,   7.446,   W3:HZ3,     W3:HZ3
7.358,   9.788,   W3:HZ2,     W3:HE1
7.358,   7.446,   W3:HZ2,     W3:HZ3
7.358,   7.128,   W3:HZ2,     W3:HH2
4.367,   3.759,   W3:HA,      Q4:HA
4.367,   2.087,   W3:HA,      Q4:HB
4.367,   7.697,   W3:HA,      Q4:HN
3.160,   3.759,   W3:HB2,     Q4:HA
3.213,   3.759,   W3:HB1,     Q4:HA
3.160,   2.087,   W3:HB2,     Q4:HB
3.213,   2.087,   W3:HB1,     Q4:HB
3.160,   7.697,   W3:HB2,     Q4:HN
3.213,   7.697,   W3:HB1,     Q4:HN
7.121,   3.759,   W3:HD1,     Q4:HA
7.121,   2.087,   W3:HD1,     Q4:HB
7.121,   2.393,   W3:HD1,     Q4:HG1
7.431,   3.759,   W3:HE3,     Q4:HA
7.431,   2.087,   W3:HE3,     Q4:HB
7.431,   6.842,   W3:HE3,     Q4:HE22
7.431,   2.393,   W3:HE3,     Q4:HG1
7.431,   1.958,   W3:HE3,     Q4:HG2
7.431,   7.697,   W3:HE3,     Q4:HN
7.121,   6.842,   W3:HH2,     Q4:HE22
7.121,   7.697,   W3:HH2,     Q4:HN
8.131,   7.697,   W3:HN,      Q4:HN
6.999,   2.087,   W3:HZ3,     Q4:HB
4.367,   7.613,   W3:HA,      K5:HN
3.160,   7.613,   W3:HB2,     K5:HN
3.213,   7.613,   W3:HB1,     K5:HN
7.121,   1.575,   W3:HD1,     K5:HB2
7.431,   1.575,   W3:HE3,     K5:HB2
8.131,   7.613,   W3:HN,      K5:HN
4.367,   3.658,   W3:HA,      I6:HA
4.367,   1.805,   W3:HA,      I6:HB
4.367,   0.707,   W3:HA,      I6:HD
4.367,   1.580,   W3:HA,      I6:HG11
4.367,   1.054,   W3:HA,      I6:HG12
4.367,   0.816,   W3:HA,      I6:HG2
4.367,   8.094,   W3:HA,      I6:HN
3.160,   1.805,   W3:HB2,     I6:HB
3.213,   1.805,   W3:HB1,     I6:HB
3.160,   0.707,   W3:HB2,     I6:HD
3.213,   0.707,   W3:HB1,     I6:HD
3.213,   0.816,   W3:HB1,     I6:HG2
3.160,   0.816,   W3:HB2,     I6:HG2
3.160,   1.580,   W3:HB2,     I6:HG11
3.213,   1.580,   W3:HB1,     I6:HG11
3.160,   8.094,   W3:HB2,     I6:HN
3.213,   8.094,   W3:HB1,     I6:HN
7.121,   1.805,   W3:HD1,     I6:HB
7.121,   0.707,   W3:HD1,     I6:HD
7.431,   3.658,   W3:HE3,     I6:HA
7.431,   1.805,   W3:HE3,     I6:HB
7.431,   0.707,   W3:HE3,     I6:HD
7.431,   1.054,   W3:HE3,     I6:HG12
7.431,   0.816,   W3:HE3,     I6:HG2
6.999,   1.805,   W3:HZ3,     I6:HB
6.999,   0.707,   W3:HZ3,     I6:HD
7.358,   0.707,   W3:HZ2,     I6:HD
6.999,   0.816,   W3:HZ3,     I6:HG2
```

```
7.121,    2.768,    W3:HD1,      K7:HE2
0.692,    8.140,    W3:HD,       K7:HN
7.431,    1.248,    W3:HE3,      K7:HD2
7.430,    1.183,    W3:HE3,      K7:HG2
7.121,    1.248,    W3:HH2,      K7:HD2
8.131,    3.879,    W3:HN,       K7:HA
7.358,    1.248,    W3:HZ2,      K7:HD2
7.358,    2.818,    W3:HZ2,      K7:HE1
7.358,    2.768,    W3:HZ2,      K7:HE2
7.358,    1.183,    W3:HZ2,      K7:HG2
7.121,    4.018,    W3:HD1,      E19:HA
7.121,    3.823,    W3:HH2,      I21:HA
7.121,    4.104,    W3:HH2,      K22:HA
3.748,    8.012,    Q4:HA,       G1:HN
2.075,    3.809,    Q4:HB,       G1:HA2
7.686,    3.887,    Q4:HN,       G1:HA1
7.686,    3.809,    Q4:HN,       G1:HA2
1.947,    4.026,    Q4:HG2,      L2:HA
7.686,    4.026,    Q4:HN,       L2:HA
7.686,    1.487,    Q4:HN,       L2:HB
7.686,    8.089,    Q4:HN,       L2:HN
3.748,    7.446,    Q4:HA,       W3:H4
3.748,    7.128,    Q4:HA,       W3:H6
3.748,    4.377,    Q4:HA,       W3:HA
3.748,    3.329,    Q4:HA,       W3:HB1
3.748,    3.163,    Q4:HA,       W3:HB2
2.075,    7.446,    Q4:HB,       W3:H4
6.827,    7.446,    Q4:HE22,     W3:HE3
6.827,    7.128,    Q4:HE22,     W3:HH2
1.947,    7.446,    Q4:HG2,      W3:H4
2.389,    7.446,    Q4:HG1,      W3:H4
7.686,    7.128,    Q4:HN,       W3:H6
7.686,    4.377,    Q4:HN,       W3:HA
6.827,    3.163,    Q4:HE1,      W3:HB2
7.686,    3.329,    Q4:HN,       W3:HB1
7.686,    3.163,    Q4:HN,       W3:HB2
7.686,    7.446,    Q4:HN,       W3:HE3
7.686,    8.140,    Q4:HN,       W3:HN
3.748,    2.087,    Q4:HA,       Q4:HB
3.748,    6.842,    Q4:HA,       Q4:HE22
3.748,    7.407,    Q4:HA,       Q4:HE21
3.748,    2.393,    Q4:HA,       Q4:HG1
3.748,    1.958,    Q4:HA,       Q4:HG2
3.748,    7.697,    Q4:HA,       Q4:HN
2.075,    3.759,    Q4:HB,       Q4:HA
2.075,    7.407,    Q4:HB,       Q4:HE21
2.075,    6.842,    Q4:HB,       Q4:HE22
2.075,    2.393,    Q4:HB,       Q4:HG1
2.075,    1.958,    Q4:HB,       Q4:HG2
2.075,    7.697,    Q4:HB,       Q4:HN
6.827,    3.759,    Q4:HE22,     Q4:HA
7.399,    3.759,    Q4:HE21,     Q4:HA
7.399,    2.087,    Q4:HE21,     Q4:HB
6.827,    2.087,    Q4:HE22,     Q4:HB
7.399,    6.842,    Q4:HE21,     Q4:HE22
6.827,    7.407,    Q4:HE22,     Q4:HE21
6.827,    2.393,    Q4:HE22,     Q4:HG1
7.399,    2.393,    Q4:HE21,     Q4:HG1
6.827,    1.958,    Q4:HE22,     Q4:HG2
7.399,    1.958,    Q4:HE21,     Q4:HG2
7.399,    7.697,    Q4:HE21,     Q4:HN
6.827,    7.697,    Q4:HE22,     Q4:HN
1.947,    3.759,    Q4:HG2,      Q4:HA
2.389,    3.759,    Q4:HG1,      Q4:HA
```

```
1.947,    2.087,    Q4:HG2,      Q4:HB
2.389,    2.087,    Q4:HG1,      Q4:HB
1.947,    6.842,    Q4:HG2,      Q4:HE22
1.947,    7.407,    Q4:HG2,      Q4:HE21
2.389,    6.842,    Q4:HG1,      Q4:HE22
2.389,    7.407,    Q4:HG1,      Q4:HE21
1.947,    2.393,    Q4:HG2,      Q4:HG1
2.389,    1.958,    Q4:HG1,      Q4:HG2
1.947,    7.697,    Q4:HG2,      Q4:HN
2.389,    7.697,    Q4:HG1,      Q4:HN
7.686,    3.759,    Q4:HN,       Q4:HA
7.686,    2.087,    Q4:HN,       Q4:HB
7.686,    6.842,    Q4:HN,       Q4:HE22
7.686,    7.407,    Q4:HN,       Q4:HE21
7.686,    2.393,    Q4:HN,       Q4:HG1
7.686,    1.958,    Q4:HN,       Q4:HG2
3.748,    7.613,    Q4:HA,       K5:HN
2.075,    7.613,    Q4:HB,       K5:HN
1.947,    7.613,    Q4:HG2,      K5:HN
2.389,    7.613,    Q4:HG1,      K5:HN
7.686,    3.985,    Q4:HN,       K5:HA
7.686,    1.575,    Q4:HN,       K5:HB2
7.686,    1.358,    Q4:HN,       K5:HG2
7.686,    7.613,    Q4:HN,       K5:HN
3.748,    3.658,    Q4:HA,       I6:HA
7.686,    3.658,    Q4:HN,       I6:HA
7.686,    1.805,    Q4:HN,       I6:HB
7.686,    0.707,    Q4:HN,       I6:HD
7.686,    0.816,    Q4:HN,       I6:HG2
3.748,    1.368,    Q4:HA,       K7:HB
3.748,    1.248,    Q4:HA,       K7:HD2
3.748,    2.818,    Q4:HA,       K7:HE1
3.748,    1.183,    Q4:HA,       K7:HG2
3.748,    8.129,    Q4:HA,       K7:HN
3.748,    7.823,    Q4:HA,       S8:HN
7.610,    3.887,    K5:HN,       G1:HA1
1.895,    4.026,    K5:HB,       L2:HA
2.253,    4.026,    K5:HE,       L2:HA
7.610,    4.026,    K5:HN,       L2:HA
7.610,    8.589,    K5:HN,       L2:HN
7.610,    4.377,    K5:HN,       W3:HA
7.610,    3.229,    K5:HN,       W3:HB1
7.610,    3.163,    K5:HN,       W3:HB2
7.610,    8.140,    K5:HN,       W3:HN
3.976,    7.697,    K5:HA,       Q4:HN
1.569,    7.697,    K5:HD,       Q4:HN
1.350,    7.697,    K5:HG2,      Q4:HN
7.610,    3.759,    K5:HN,       Q4:HA
7.610,    2.087,    K5:HN,       Q4:HB
7.610,    2.393,    K5:HN,       Q4:HG1
7.610,    1.958,    K5:HN,       Q4:HG2
7.610,    7.697,    K5:HN,       Q4:HN
3.976,    1.914,    K5:HA,       K5:HB1
3.976,    1.580,    K5:HA,       K5:HD
3.976,    2.862,    K5:HA,       K5:HE
3.976,    1.466,    K5:HA,       K5:HG1
3.976,    1.358,    K5:HA,       K5:HG2
3.976,    7.613,    K5:HA,       K5:HN
1.895,    3.985,    K5:HB,       K5:HA
1.895,    1.580,    K5:HB,       K5:HD
1.895,    2.862,    K5:HB,       K5:HE
1.895,    1.466,    K5:HB,       K5:HG1
1.895,    1.358,    K5:HB,       K5:HG2
1.895,    7.613,    K5:HB,       K5:HN
1.569,    3.985,    K5:HD,       K5:HA
```

```
1.569,    1.914,    K5:HD,       K5:HB1
1.569,    2.862,    K5:HD,       K5:HE
1.569,    1.466,    K5:HD,       K5:HG1
1.569,    1.358,    K5:HD,       K5:HG2
1.569,    7.613,    K5:HD,       K5:HN
2.853,    3.985,    K5:HE,       K5:HA
2.853,    1.914,    K5:HE,       K5:HB1
2.853,    1.580,    K5:HE,       K5:HD
2.853,    1.466,    K5:HE,       K5:HG1
2.853,    1.358,    K5:HE,       K5:HG2
2.853,    7.613,    K5:HE,       K5:HN
1.350,    3.985,    K5:HG2,      K5:HA
1.461,    3.985,    K5:HG1,      K5:HA
1.350,    1.914,    K5:HG2,      K5:HB1
1.461,    1.914,    K5:HG1,      K5:HB1
1.350,    1.580,    K5:HG2,      K5:HD
1.461,    1.580,    K5:HG1,      K5:HD
1.350,    2.862,    K5:HG2,      K5:HE
1.461,    2.862,    K5:HG1,      K5:HE
1.350,    1.466,    K5:HG2,      K5:HG1
1.461,    1.358,    K5:HG1,      K5:HG2
1.350,    7.613,    K5:HG2,      K5:HN
1.461,    7.613,    K5:HG1,      K5:HN
7.610,    3.985,    K5:HN,       K5:HA
7.610,    1.914,    K5:HN,       K5:HB1
7.610,    1.580,    K5:HN,       K5:HD
7.610,    2.862,    K5:HN,       K5:HE
7.610,    1.358,    K5:HN,       K5:HG2
7.610,    1.466,    K5:HN,       K5:HG1
3.976,    8.094,    K5:HA,       I6:HN
1.895,    3.658,    K5:HB,       I6:HA
7.610,    3.658,    K5:HN,       I6:HA
7.610,    1.805,    K5:HN,       I6:HB
7.610,    0.707,    K5:HN,       I6:HD
7.610,    1.054,    K5:HN,       I6:HG12
7.610,    0.816,    K5:HN,       I6:HG2
7.610,    8.094,    K5:HN,       I6:HN
1.895,    8.139,    K5:HB,       K7:HN
8.078,    4.026,    I6:HN,       L2:HA
3.646,    7.446,    I6:HA,       W3:H4
3.646,    4.377,    I6:HA,       W3:HA
0.692,    7.446,    I6:HD,       W3:H4
0.692,    4.377,    I6:HD,       W3:HA
0.803,    7.446,    I6:HG2,      W3:H4
8.078,    4.377,    I6:HN,       W3:HA
8.078,    3.229,    I6:HN,       W3:HB1
8.078,    3.163,    I6:HN,       W3:HB2
3.646,    7.697,    I6:HA,       Q4:HN
0.692,    7.697,    I6:HD,       Q4:HN
0.803,    7.697,    I6:HG2,      Q4:HN
8.078,    3.759,    I6:HN,       Q4:HA
8.078,    7.697,    I6:HN,       Q4:HN
3.646,    1.914,    I6:HA,       K5:HB1
3.646,    7.613,    I6:HA,       K5:HN
0.692,    7.613,    I6:HD,       K5:HN
0.803,    7.613,    I6:HG2,      K5:HN
1.044,    7.613,    I6:HG12,     K5:HN
8.078,    3.985,    I6:HN,       K5:HA
8.078,    7.613,    I6:HN,       K5:HN
3.646,    1.805,    I6:HA,       I6:HB
3.646,    0.707,    I6:HA,       I6:HD
3.646,    1.580,    I6:HA,       I6:HG11
3.646,    0.816,    I6:HA,       I6:HG2
3.646,    1.054,    I6:HA,       I6:HG12
```

```
3.646,    8.094,    I6:HA,       I6:HN
1.793,    3.658,    I6:HB,       I6:HA
1.793,    0.707,    I6:HB,       I6:HD
1.793,    1.580,    I6:HB,       I6:HG11
1.793,    1.054,    I6:HB,       I6:HG12
1.793,    0.816,    I6:HB,       I6:HG2
1.793,    8.094,    I6:HB,       I6:HN
0.692,    3.658,    I6:HD,       I6:HA
0.692,    1.805,    I6:HD,       I6:HB
0.692,    1.580,    I6:HD,       I6:HG11
0.692,    1.054,    I6:HD,       I6:HG12
0.692,    8.094,    I6:HD,       I6:HN
0.803,    3.658,    I6:HG2,      I6:HA
1.044,    3.658,    I6:HG12,     I6:HA
1.569,    3.658,    I6:HG11,     I6:HA
0.803,    1.805,    I6:HG2,      I6:HB
1.044,    1.805,    I6:HG12,     I6:HB
1.569,    1.805,    I6:HG11,     I6:HB
1.044,    0.707,    I6:HG12,     I6:HD
1.569,    0.707,    I6:HG11,     I6:HD
0.803,    1.580,    I6:HG2,      I6:HG11
0.803,    1.054,    I6:HG2,      I6:HG12
1.044,    1.580,    I6:HG12,     I6:HG11
1.044,    0.816,    I6:HG12,     I6:HG2
1.569,    1.054,    I6:HG11,     I6:HG12
1.569,    0.816,    I6:HG11,     I6:HG2
0.803,    8.094,    I6:HG2,      I6:HN
1.044,    8.094,    I6:HG12,     I6:HN
1.569,    8.094,    I6:HG11,     I6:HN
8.078,    3.658,    I6:HN,       I6:HA
8.078,    1.805,    I6:HN,       I6:HB
8.078,    0.707,    I6:HN,       I6:HD
8.078,    1.580,    I6:HN,       I6:HG11
8.078,    1.054,    I6:HN,       I6:HG12
8.078,    0.816,    I6:HN,       I6:HG2
3.646,    8.129,    I6:HA,       K7:HN
1.793,    8.129,    I6:HB,       K7:HN
1.044,    8.129,    I6:HG12,     K7:HN
8.078,    8.129,    I6:HN,       K7:HN
3.646,    7.823,    I6:HA,       S8:HN
0.692,    7.823,    I6:HD,       S8:HN
3.646,    1.433,    I6:HA,       A9:HB
3.646,    7.946,    I6:HA,       A9:HN
3.646,    8.387,    I6:HA,       A10:HN
1.234,    7.446,    K7:HD2,      W3:H4
1.234,    7.128,    K7:HD2,      W3:H6
1.234,    3.759,    K7:HD2,      Q4:HA
2.813,    3.759,    K7:HE1,      Q4:HA
1.176,    3.759,    K7:HG2,      Q4:HA
8.119,    3.759,    K7:HN,       Q4:HA
8.119,    1.914,    K7:HN,       K5:HB1
8.119,    3.658,    K7:HN,       I6:HA
8.119,    1.805,    K7:HN,       I6:HB
8.119,    0.707,    K7:HN,       I6:HD
8.119,    1.054,    K7:HN,       I6:HG12
8.119,    8.094,    K7:HN,       I6:HN
3.807,    1.568,    K7:HA,       K7:HB
3.807,    1.472,    K7:HA,       K7:HD1
3.807,    2.818,    K7:HA,       K7:HE1
3.807,    2.768,    K7:HA,       K7:HE2
3.807,    1.248,    K7:HA,       K7:HG1
3.807,    1.183,    K7:HA,       K7:HG2
3.807,    8.129,    K7:HA,       K7:HN
1.552,    3.819,    K7:HB,       K7:HA
1.552,    1.472,    K7:HB,       K7:HD1
```

```
1.552,    2.818,    K7:HB,        K7:HE1
1.552,    2.768,    K7:HB,        K7:HE2
1.552,    1.248,    K7:HB,        K7:HG1
1.552,    1.183,    K7:HB,        K7:HG2
1.552,    8.129,    K7:HB,        K7:HN
1.462,    3.819,    K7:HD1,       K7:HA
1.462,    1.568,    K7:HD1,       K7:HB
1.462,    2.818,    K7:HD1,       K7:HE1
1.462,    2.768,    K7:HD1,       K7:HE2
1.462,    1.248,    K7:HD1,       K7:HG1
1.462,    1.183,    K7:HD1,       K7:HG2
1.462,    8.129,    K7:HD1,       K7:HN
2.756,    3.819,    K7:HE2,       K7:HA
2.756,    1.568,    K7:HE2,       K7:HB
2.813,    3.819,    K7:HE1,       K7:HA
2.813,    1.568,    K7:HE1,       K7:HB
2.756,    1.472,    K7:HE2,       K7:HD1
2.813,    1.472,    K7:HE1,       K7:HD1
2.756,    1.248,    K7:HE2,       K7:HG1
2.756,    1.183,    K7:HE2,       K7:HG2
2.813,    1.248,    K7:HE1,       K7:HG1
2.813,    1.183,    K7:HE1,       K7:HG2
2.756,    8.129,    K7:HE2,       K7:HN
2.813,    8.129,    K7:HE1,       K7:HN
1.176,    3.819,    K7:HG2,       K7:HA
1.176,    1.568,    K7:HG2,       K7:HB
1.234,    3.819,    K7:HG1,       K7:HA
1.234,    1.568,    K7:HG1,       K7:HB
1.176,    1.472,    K7:HG2,       K7:HD1
1.234,    1.472,    K7:HG1,       K7:HD1
1.176,    2.818,    K7:HG2,       K7:HE1
1.176,    2.768,    K7:HG2,       K7:HE2
1.234,    2.818,    K7:HG1,       K7:HE1
1.234,    2.768,    K7:HG1,       K7:HE2
1.176,    8.129,    K7:HG2,       K7:HN
1.234,    8.129,    K7:HG1,       K7:HN
8.119,    3.819,    K7:HN,        K7:HA
8.119,    1.568,    K7:HN,        K7:HB
8.119,    1.472,    K7:HN,        K7:HD1
8.119,    2.818,    K7:HN,        K7:HE1
8.119,    2.768,    K7:HN,        K7:HE2
8.119,    1.248,    K7:HN,        K7:HG1
8.119,    1.183,    K7:HN,        K7:HG2
1.234,    7.823,    K7:HD1,       S8:HN
1.176,    7.823,    K7:HG2,       S8:HN
8.119,    4.192,    K7:HN,        S8:HA
8.119,    3.963,    K7:HN,        S8:HB1
8.119,    3.883,    K7:HN,        S8:HB2
8.119,    7.823,    K7:HN,        S8:HN
3.807,    1.398,    K7:HA,        A10:HB
3.807,    8.387,    K7:HA,        A10:HN
7.818,    3.759,    S8:HN,        Q4:HA
4.183,    1.575,    S8:HA,        K5:HB2
7.818,    3.985,    S8:HN,        K5:HA
7.818,    1.914,    S8:HN,        K5:HB1
4.183,    0.816,    S8:HA,        I6:HG2
7.818,    3.658,    S8:HN,        I6:HA
7.818,    0.707,    S8:HN,        I6:HD
7.818,    0.816,    S8:HN,        I6:HG2
4.183,    1.248,    S8:HA,        K7:HD2
4.183,    1.183,    S8:HA,        K7:HG2
4.183,    8.129,    S8:HA,        K7:HN
3.950,    8.129,    S8:HB1,       K7:HN
3.881,    8.129,    S8:HB,        K7:HN
```

```
7.818,    3.819,    S8:HN,      K7:HA
7.818,    1.568,    S8:HN,      K7:HB
7.818,    1.248,    S8:HN,      K7:HD2
7.818,    1.193,    S8:HN,      K7:HG2
7.818,    8.129,    S8:HN,      K7:HN
4.183,    3.963,    S8:HA,      S8:HB1
4.183,    3.883,    S8:HA,      S8:HB2
4.183,    7.823,    S8:HA,      S8:HN
3.950,    4.192,    S8:HB1,     S8:HA
3.881,    4.192,    S8:HB,      S8:HA
3.950,    3.883,    S8:HB1,     S8:HB2
3.881,    3.963,    S8:HB,      S8:HB1
3.950,    7.823,    S8:HB1,     S8:HN
3.881,    7.823,    S8:HB,      S8:HN
7.818,    4.192,    S8:HN,      S8:HA
7.818,    3.963,    S8:HN,      S8:HB1
7.818,    3.883,    S8:HN,      S8:HB2
4.183,    1.433,    S8:HA,      A9:HB
4.183,    7.946,    S8:HA,      A9:HN
3.950,    7.946,    S8:HB1,     A9:HN
7.818,    1.433,    S8:HN,      A9:HB
7.818,    7.946,    S8:HN,      A9:HN
4.183,    8.387,    S8:HA,      A10:HN
3.950,    8.387,    S8:HB1,     A10:HN
3.881,    8.387,    S8:HB,      A10:HN
7.818,    8.387,    S8:HN,      A10:HN
4.183,    3.834,    S8:HA,      G11:HA
4.183,    8.033,    S8:HA,      G11:HN
3.950,    3.834,    S8:HB1,     G11:HA
3.950,    8.032,    S8:HB1,     G11:HN
7.940,    3.658,    A9:HN,      I6:HA
1.424,    4.192,    A9:HB,      S8:HA
1.424,    7.823,    A9:HB,      S8:HN
7.940,    4.192,    A9:HN,      S8:HA
7.940,    3.963,    A9:HN,      S8:HB1
7.940,    7.823,    A9:HN,      S8:HN
4.137,    1.433,    A9:HA,      A9:HB
4.137,    7.946,    A9:HA,      A9:HN
1.424,    4.149,    A9:HB,      A9:HA
1.424,    7.946,    A9:HB,      A9:HN
7.940,    4.149,    A9:HN,      A9:HA
7.940,    1.433,    A9:HN,      A9:HB
4.137,    8.387,    A9:HA,      A10:HN
1.424,    8.387,    A9:HB,      A10:HN
7.940,    8.387,    A9:HN,      A10:HN
4.137,    3.012,    A9:HA,      D12:HB1
4.137,    2.842,    A9:HA,      D12:HB2
4.137,    8.069,    A9:HA,      D12:HN
7.940,    3.012,    A9:HN,      D12:HB1
7.940,    2.842,    A9:HN,      D12:HB2
8.376,    3.658,    A10:HN,     I6:HA
8.376,    3.819,    A10:HN,     K7:HA
8.376,    4.192,    A10:HN,     S8:HA
8.376,    3.963,    A10:HN,     S8:HB1
8.376,    3.883,    A10:HN,     S8:HB2
8.376,    7.823,    A10:HN,     S8:HN
8.376,    4.149,    A10:HN,     A9:HA
8.376,    1.433,    A10:HN,     A9:HB
8.376,    7.946,    A10:HN,     A9:HN
4.033,    1.398,    A10:HA,     A10:HB
4.033,    8.387,    A10:HA,     A10:HN
1.381,    4.045,    A10:HB,     A10:HA
1.381,    8.387,    A10:HB,     A10:HN
8.376,    4.045,    A10:HN,     A10:HA
8.376,    1.398,    A10:HN,     A10:HB
```

```
4.033,    8.032,    A10:HA,     G11:HN
8.376,    8.032,    A10:HN,     G11:HN
4.033,    8.069,    A10:HA,     D12:HN
8.376,    3.012,    A10:HN,     D12:HB1
8.376,    2.842,    A10:HN,     D12:HB2
4.033,    1.719,    A10:HA,     L13:HB
4.033,    1.638,    A10:HA,     L13:HG
8.376,    0.818,    A10:HN,     L13:HD2
3.833,    4.192,    G11:HA2,    S8:HA
3.833,    3.963,    G11:HA2,    S8:HB1
8.017,    4.192,    G11:HN,     S8:HA
8.017,    3.963,    G11:HN,     S8:HB1
8.017,    4.045,    G11:HN,     A10:HA
8.017,    1.398,    G11:HN,     A10:HB
8.017,    8.387,    G11:HN,     A10:HN
3.833,    8.032,    G11:HA2,    G11:HN
8.017,    3.834,    G11:HN,     G11:HA
8.017,    4.507,    G11:HN,     D12:HA
8.017,    3.012,    G11:HN,     D12:HB1
8.017,    2.842,    G11:HN,     D12:HB2
3.833,    1.638,    G11:HA2,    L13:HG
3.833,    1.421,    G11:HA2,    A14:HB
3.876,    1.421,    G11:HA1,    A14:HB
3.833,    8.183,    G11:HA2,    A14:HN
2.999,    4.149,    D12:HB1,    A9:HA
2.838,    4.149,    D12:HB2,    A9:HA
2.999,    1.433,    D12:HB1,    A9:HB
8.055,    4.149,    D12:HN,     A9:HA
2.999,    8.387,    D12:HB1,    A10:HN
8.055,    4.045,    D12:HN,     A10:HA
8.055,    8.387,    D12:HN,     A10:HN
8.055,    3.834,    D12:HN,     G11:HA
4.498,    3.012,    D12:HA,     D12:HB1
4.498,    2.842,    D12:HA,     D12:HB2
4.498,    8.069,    D12:HA,     D12:HN
2.838,    4.507,    D12:HB2,    D12:HA
2.999,    4.507,    D12:HB1,    D12:HA
2.838,    3.012,    D12:HB2,    D12:HB1
2.999,    2.842,    D12:HB1,    D12:HB2
2.838,    8.069,    D12:HB2,    D12:HN
2.999,    8.069,    D12:HB1,    D12:HN
8.055,    4.507,    D12:HN,     D12:HA
8.055,    3.012,    D12:HN,     D12:HB1
8.055,    2.842,    D12:HN,     D12:HB2
4.498,    1.638,    D12:HA,     L13:HG
4.498,    8.130,    D12:HA,     L13:HN
2.999,    1.719,    D12:HB1,    L13:HB
2.999,    1.638,    D12:HB1,    L13:HG
2.838,    8.130,    D12:HB2,    L13:HN
2.999,    8.130,    D12:HB1,    L13:HN
4.498,    1.421,    D12:HA,     A14:HB
4.498,    8.183,    D12:HA,     A14:HN
2.999,    8.183,    D12:HB1,    A14:HN
8.055,    1.421,    D12:HN,     A14:HB
4.498,    3.896,    D12:HA,     S15:HB
4.498,    3.940,    D12:HA,     S15:HB
4.498,    7.903,    D12:HA,     S15:HN
8.118,    4.045,    L13:HN,     A10:HA
8.118,    4.507,    L13:HN,     D12:HA
8.118,    3.012,    L13:HN,     D12:HB1
8.118,    2.842,    L13:HN,     D12:HB2
4.108,    1.719,    L13:HA,     L13:HB
4.108,    0.837,    L13:HA,     L13:HD1
4.108,    0.818,    L13:HA,     L13:HD2
```

```
4.108,    1.638,    L13:HA,     L13:HG
4.108,    8.130,    L13:HA,     L13:HN
1.701,    4.120,    L13:HB,     L13:HA
1.701,    0.837,    L13:HB,     L13:HD1
1.701,    0.818,    L13:HB,     L13:HD2
1.701,    1.638,    L13:HB,     L13:HG
1.701,    8.130,    L13:HB,     L13:HN
0.807,    4.120,    L13:HD2,    L13:HA
0.829,    4.120,    L13:HD1,    L13:HA
0.807,    1.719,    L13:HD2,    L13:HB
0.829,    1.719,    L13:HD1,    L13:HB
0.807,    1.638,    L13:HD2,    L13:HG
0.829,    1.638,    L13:HD1,    L13:HG
0.807,    8.130,    L13:HD2,    L13:HN
0.829,    8.130,    L13:HD1,    L13:HN
1.630,    4.120,    L13:HG,     L13:HA
1.630,    1.719,    L13:HG,     L13:HB
1.630,    0.837,    L13:HG,     L13:HD1
1.630,    0.818,    L13:HG,     L13:HD2
1.630,    8.130,    L13:HG,     L13:HN
8.118,    4.120,    L13:HN,     L13:HA
8.118,    1.719,    L13:HN,     L13:HB
8.118,    0.837,    L13:HN,     L13:HD1
8.118,    0.818,    L13:HN,     L13:HD2
8.118,    1.638,    L13:HN,     L13:HG
4.106,    8.183,    L13:HA,     A14:HN
1.701,    8.183,    L13:HB,     A14:HN
0.829,    8.183,    L13:HD1,    A14:HN
1.630,    8.183,    L13:HG,     A14:HN
8.118,    1.421,    L13:HN,     A14:HB
8.175,    3.834,    A14:HN,     G11:HA
8.175,    4.507,    A14:HN,     D12:HA
8.175,    3.012,    A14:HN,     D12:HB1
8.175,    2.842,    A14:HN,     D12:HB2
8.175,    4.120,    A14:HN,     L13:HA
8.175,    1.719,    A14:HN,     L13:HB
8.175,    0.837,    A14:HN,     L13:HD1
8.175,    1.638,    A14:HN,     L13:HG
4.014,    1.421,    A14:HA,     A14:HB
4.014,    8.183,    A14:HA,     A14:HN
1.412,    4.024,    A14:HB,     A14:HA
1.412,    8.183,    A14:HB,     A14:HN
8.175,    4.024,    A14:HN,     A14:HA
8.175,    1.421,    A14:HN,     A14:HB
4.014,    7.903,    A14:HA,     S15:HN
8.175,    3.940,    A14:HN,     S15:HB
8.175,    7.903,    A14:HN,     S15:HN
8.175,    1.951,    A14:HN,     I17:HB
8.175,    0.763,    A14:HN,     I17:HD
8.175,    8.007,    A14:HN,     V18:HN
3.887,    4.507,    S15:HB2,    D12:HA
3.932,    4.507,    S15:HB1,    D12:HA
7.893,    4.507,    S15:HN,     D12:HA
7.893,    3.012,    S15:HN,     D12:HB1
7.893,    2.842,    S15:HN,     D12:HB2
7.893,    1.638,    S15:HN,     L13:HG
7.893,    1.421,    S15:HN,     A14:HB
7.893,    8.183,    S15:HN,     A14:HN
4.092,    3.896,    S15:HA,     S15:HB
4.092,    3.940,    S15:HA,     S15:HB
4.092,    7.903,    S15:HA,     S15:HN
3.887,    4.104,    S15:HB2,    S15:HA
3.932,    4.104,    S15:HB1,    S15:HA
3.887,    7.903,    S15:HB2,    S15:HN
3.932,    7.903,    S15:HB1,    S15:HN
```

```
7.893,    4.104,    S15:HN,     S15:HA
7.893,    3.896,    S15:HN,     S15:HB
7.893,    3.940,    S15:HN,     S15:HB
4.092,    7.879,    S15:HA,     G16:HN
3.932,    7.879,    S15:HB1,    G16:HN
7.893,    1.951,    S15:HN,     T17:HB
7.893,    0.763,    S15:HN,     T17:HD
7.893,    1.056,    S15:HN,     I17:HG12
7.893,    0.838,    S15:HN,     I17:HG2
4.092,    2.068,    S15:HA,     V18:HB
4.092,    0.988,    S15:HA,     V18:HG1
4.092,    8.007,    S15:HA,     V18:HN
7.893,    2.068,    S15:HN,     V18:HB
7.893,    0.988,    S15:HN,     V18:HG1
7.868,    4.507,    G16:HN2,    D12:HA
7.871,    1.421,    G16:HN1,    A14:HB
7.868,    4.104,    G16:HN2,    S15:HA
7.868,    3.940,    G16:HN2,    S15:HB
3.833,    7.879,    G16:HA,     G16:HN
7.868,    3.837,    G16:HN2,    G16:HA
7.871,    1.951,    G16:HN1,    I17:HB
7.871,    0.763,    G16:HN1,    I17:HD
7.871,    1.652,    G16:HN1,    I17:HG11
7.871,    1.056,    G16:HN1,    I17:HG12
7.871,    0.838,    G16:HN1,    I17:HG2
3.833,    2.085,    G16:HA,     F19:HB
0.755,    8.183,    I17:HD,     A14:HN
7.947,    7.903,    I17:HB,     S15:HN
0.755,    7.903,    I17:HD,     S15:HN
0.828,    7.903,    I17:HG2,    S15:HN
1.046,    7.903,    I17:HG12,   S15:HN
1.642,    7.903,    I17:HG11,   S15:HN
7.948,    4.104,    I17:HN,     S15:HA
1.941,    7.879,    I17:HB,     G16:HN
0.755,    7.879,    I17:HD,     G16:HN
0.828,    7.879,    I17:HG2,    G16:HN
1.046,    7.879,    I17:HG12,   G16:HN
1.642,    7.879,    I17:HG11,   G16:HN
7.948,    3.837,    I17:HN,     G16:HA
3.784,    1.951,    T17:HA,     T17:HB
3.784,    0.763,    T17:HA,     T17:HD
3.784,    1.652,    T17:HA,     I17:HG11
3.784,    1.056,    T17:HA,     T17:HG12
3.784,    0.838,    T17:HA,     T17:HG2
3.784,    7.952,    T17:HA,     T17:HN
7.947,    3.793,    T17:HB,     T17:HA
1.941,    0.763,    T17:HB,     T17:HD
1.941,    1.652,    I17:HB,     I17:HG11
1.941,    1.056,    I17:HB,     I17:HG12
1.941,    0.838,    I17:HB,     I17:HG2
1.941,    7.952,    I17:HB,     I17:HN
0.755,    3.793,    I17:HD,     I17:HA
0.755,    1.951,    I17:HD,     I17:HB
0.755,    1.652,    I17:HD,     I17:HG11
0.755,    1.056,    I17:HD,     T17:HG12
0.755,    7.952,    T17:HD,     I17:HN
0.838,    1.951,    I17:HG2,    I17:HB
1.046,    3.793,    I17:HG12,   I17:HA
1.046,    1.951,    I17:HG12,   I17:HB
0.828,    3.793,    I17:HG2,    I17:HA
1.642,    3.793,    I17:HG11,   I17:HA
1.642,    1.951,    I17:HG11,   I17:HB
0.828,    1.652,    I17:HG2,    I17:HG11
0.828,    1.056,    I17:HG2,    T17:HG12
```

```
1.046,    0.763,    I17:HG12,    I17:HD
1.046,    1.652,    I17:HG12,    I17:HG11
1.046,    0.838,    I17:HG12,    I17:HG2
1.642,    0.763,    I17:HG11,    I17:HD
1.642,    1.056,    I17:HG11,    I17:HG12
1.642,    0.838,    I17:HG11,    I17:HG2
0.828,    7.963,    I17:HG2,     I17:HN
1.046,    7.952,    I17:HG12,    I17:HN
1.642,    7.952,    I17:HG11,    I17:HN
7.948,    3.793,    I17:HN,      I17:HA
7.948,    1.951,    I17:HN,      I17:HB
7.948,    0.763,    I17:HN,      I17:HD
7.948,    1.652,    I17:HN,      I17:HG11
7.948,    1.056,    I17:HN,      I17:HG12
7.948,    0.838,    I17:HN,      I17:HG2
1.941,    0.988,    I17:HB,      V18:HG1
0.828,    8.007,    I17:HG2,     V18:HN
1.046,    8.007,    I17:HG12,    V18:HN
7.948,    2.068,    I17:HN,      V18:HB
7.948,    0.988,    I17:HN,      V18:HG1
7.994,    4.104,    V18:HN,      S15:HA
2.058,    7.952,    V18:HB,      I17:HN
0.975,    1.951,    V18:HG,      I17:HB
0.975,    7.952,    V18:HG,      I17:HN
7.994,    3.793,    V18:HN,      I17:HA
7.994,    1.951,    V18:HN,      I17:HB
7.994,    0.763,    V18:HN,      I17:HD
7.994,    0.838,    V18:HN,      I17:HG2
7.994,    1.652,    V18:HN,      I17:HG11
3.602,    2.068,    V18:HA,      V18:HB
3.602,    0.988,    V18:HA,      V18:HG1
3.602,    0.880,    V18:HA,      V18:HG2
3.602,    8.007,    V18:HA,      V18:HN
2.058,    3.615,    V18:HB,      V18:HA
2.058,    0.988,    V18:HB,      V18:HG1
2.058,    0.880,    V18:HB,      V18:HG2
2.058,    8.007,    V18:HB,      V18:HN
0.975,    3.615,    V18:HG,      V18:HA
0.866,    3.615,    V18:HG,      V18:HA
0.975,    2.068,    V18:HG,      V18:HB
0.866,    2.068,    V18:HG,      V18:HB
0.975,    0.880,    V18:HG1,     V18:HG2
0.866,    0.988,    V18:HG,      V18:HG1
0.975,    8.007,    V18:HG,      V18:HN
0.866,    8.007,    V18:HG,      V18:HN
7.994,    3.615,    V18:HN,      V18:HA
7.994,    2.068,    V18:HN,      V18:HB
7.994,    0.988,    V18:HN,      V18:HG1
7.994,    0.880,    V18:HN,      V18:HG2
3.602,    8.145,    V18:HA,      E19:HN
0.975,    4.018,    V18:HG,      E19:HA
0.866,    4.018,    V18:HG,      E19:HA
0.975,    8.145,    V18:HG,      E19:HN
0.866,    8.145,    V18:HG,      E19:HN
7.994,    2.437,    V18:HN,      E19:HG
7.994,    2.497,    V18:HN,      E19:HG
7.994,    8.145,    V18:HN,      E19:HN
3.602,    1.899,    V18:HA,      I21:HB
3.602,    0.769,    V18:HA,      I21:HD1
3.602,    1.670,    V18:HA,      I21:HG11
3.602,    1.082,    V18:HA,      I21:HG12
3.602,    0.841,    V18:HA,      I21:HG2
3.602,    8.088,    V18:HA,      I21:HN
0.975,    1.899,    V18:HG,      I21:HB
0.975,    8.088,    V18:HG,      I21:HN
```

```
7.994,    1.082,    V18:HN,      I21:HG12
3.603,    8.227,    V18:HA,      I22:HN
0.866,    1.811,    V18:HG,      I22:HN
1.975,    8.227,    V18:HG,      I22:HN
8.132,    3.837,    R19:HN,      G16:HA
4.007,    1.056,    R19:HA,      I17:HG12
4.007,    0.988,    R19:HA,      V18:HG1
4.007,    0.880,    R19:HA,      V18:HG2
8.132,    3.615,    R19:HN,      V18:HA
8.132,    0.989,    R19:HN,      V18:HG1
8.132,    0.880,    E19:HN,      V18:HG2
8.132,    8.007,    E19:HN,      V18:HN
4.007,    2.085,    E19:HA,      E19:HB
4.007,    2.497,    E19:HA,      E19:HG1
4.007,    2.437,    E19:HA,      E19:HG2
4.007,    8.145,    R19:HA,      E19:HN
2.071,    4.018,    E19:HB,      E19:HA
2.071,    2.497,    E19:HB,      E19:HG1
2.071,    2.437,    E19:HB,      E19:HG2
2.071,    8.145,    E19:HB,      E19:HN
2.435,    4.018,    R19:HG2,     E19:HA
2.435,    2.085,    E19:HG2,     E19:HB
2.482,    4.018,    E19:HG1,     E19:HA
2.482,    2.085,    R19:HG1,     E19:HB
2.435,    2.497,    E19:HG2,     E19:HG1
2.482,    2.437,    E19:HG1,     E19:HG2
2.435,    8.145,    R19:HG2,     E19:HN
2.482,    8.145,    R19:HG1,     E19:HN
8.132,    2.085,    R19:HN,      E19:HB
8.132,    2.497,    R19:HN,      E19:HG1
8.132,    2.437,    E19:HN,      E19:HG2
4.007,    0.757,    R19:HA,      I21:HD2
2.071,    8.088,    R19:HB,      I21:HN
4.007,    1.811,    R19:HA,      I22:HN
7.995,    4.018,    G20:HN,      E19:HA
7.995,    3.844,    G20:HN,      G20:HA
1.889,    3.615,    I21:HB,      V18:HA
1.889,    0.988,    I21:HB,      V18:HG1
0.764,    3.615,    I21:HD,      V18:HA
0.762,    8.007,    I21:HD,      V18:HN
0.835,    3.615,    I21:HG2,     V18:HA
1.077,    3.615,    I21:HG12,    V18:HA
1.661,    3.615,    I21:HG11,    V18:HA
1.661,    8.007,    I21:HG11,    V18:HN
8.075,    3.615,    I21:HN,      V18:HA
8.075,    0.988,    I21:HN,      V18:HG1
0.757,    4.018,    I21:HD,      E19:HA
8.075,    2.085,    I21:HN,      E19:HB
8.075,    3.844,    I21:HN,      G20:HA
3.811,    1.899,    I21:HA,      I21:HB
3.811,    0.769,    I21:HA,      I21:HD1
3.811,    1.670,    I21:HA,      I21:HG11
3.811,    1.082,    I21:HA,      I21:HG12
3.811,    0.841,    I21:HA,      I21:HG2
3.811,    8.088,    I21:HA,      I21:HN
1.899,    3.823,    I21:HB,      I21:HA
1.889,    0.769,    I21:HB,      I21:HD1
1.889,    1.670,    I21:HB,      I21:HG11
1.889,    1.082,    I21:HB,      I21:HG12
1.889,    0.841,    I21:HB,      I21:HG2
1.889,    8.088,    I21:HB,      I21:HN
0.762,    3.823,    I21:HD,      I21:HA
0.762,    1.899,    I21:HD,      I21:HB
0.762,    1.670,    I21:HD,      I21:HG11
```

```
0.762,    1.082,    I21:HD,      I21:HG12
0.762,    8.088,    I21:HD,      I21:HN
0.825,    3.823,    I21:HG2,     I21:HA
0.825,    1.899,    I21:HG2,     I21:HB
1.077,    3.823,    I21:HG12,    I21:HA
1.077,    1.899,    I21:HG12,    I21:HB
1.661,    3.823,    I21:HG11,    I21:HA
1.661,    1.899,    I21:HG11,    I21:HB
0.825,    1.670,    I21:HG2,     I21:HG11
0.825,    1.082,    I21:HG2,     I21:HG12
1.077,    0.769,    I21:HG12,    I21:HD1
1.077,    1.670,    I21:HG12,    I21:HG11
1.077,    0.841,    I21:HG12,    I21:HG2
1.661,    0.769,    I21:HG11,    I21:HD1
1.661,    1.082,    I21:HG11,    I21:HG12
1.661,    0.841,    I21:HG11,    I21:HG2
0.825,    8.088,    I21:HG2,     I21:HN
1.077,    8.088,    I21:HG12,    I21:HN
1.661,    8.088,    I21:HG11,    I21:HN
8.075,    3.823,    I21:HN,      I21:HA
8.075,    1.899,    I21:HN,      I21:HB
8.075,    0.769,    I21:HN,      I21:HD1
8.075,    1.670,    I21:HN,      I21:HG11
8.075,    1.082,    I21:HN,      I21:HG12
8.075,    0.841,    I21:HN,      I21:HG2
3.811,    1.811,    I21:HA,      K22:HB
3.811,    8.227,    I21:HA,      K22:HN
0.762,    8.227,    I21:HD,      K22:HN
0.825,    8.227,    I21:HG2,     K22:HN
1.077,    8.227,    I21:HG12,    K22:HN
8.075,    8.227,    I21:HN,      K22:HN
3.811,    7.832,    I21:HA,      S23:HN
1.806,    0.880,    K22:HB,      V18:HG2
2.868,    0.988,    K22:HE2,     V18:HG1
2.868,    0.880,    K22:HE2,     V18:HG2
1.506,    0.880,    K22:HG1,     V18:HG2
8.221,    3.615,    K22:HN,      V18:HA
8.221,    0.988,    K22:HN,      V18:HG1
1.806,    4.018,    K22:HB,      E19:HA
8.221,    2.085,    K22:HN,      E19:HB
8.221,    4.018,    K22:HN,      E19:HA
4.090,    1.082,    K22:HA,      T21:HG12
8.221,    3.823,    K22:HN,      T21:HA
8.221,    1.899,    K22:HN,      T21:HB
8.221,    0.757,    K22:HN,      T21:HD2
8.221,    1.670,    K22:HN,      I21:HG11
8.221,    1.082,    K22:HN,      I21:HG12
8.221,    0.841,    K22:HN,      I21:HG2
8.221,    8.088,    K22:HN,      I21:HN
4.090,    1.811,    K22:HA,      K22:HB
4.090,    1.594,    K22:HA,      K22:HD1
4.090,    2.883,    K22:HA,      K22:HE
4.090,    1.511,    K22:HA,      K22:HG1
4.090,    1.404,    K22:HA,      K22:HG2
4.090,    8.227,    K22:HA,      K22:HN
1.806,    4.104,    K22:HB,      K22:HA
1.806,    1.594,    K22:HB,      K22:HD1
1.806,    2.883,    K22:HB,      K22:HE
1.806,    1.511,    K22:HB,      K22:HG1
1.806,    1.404,    K22:HB,      K22:HG2
1.806,    8.227,    K22:HB,      K22:HN
1.586,    4.104,    K22:HD,      K22:HA
1.586,    1.811,    K22:HD,      K22:HB
1.586,    2.883,    K22:HD,      K22:HE
1.586,    1.404,    K22:HD,      K22:HG2
```

```
1.586,    8.227,    K22:HD,      K22:HN
2.875,    4.104,    K22:HB*,     K22:HA
2.875,    1.811,    K22:HB*,     K22:HB
2.875,    1.594,    K22:HB*,     K22:HD1
2.875,    1.511,    K22:HB*,     K22:HG1
2.875,    1.404,    K22:HB*,     K22:HG2
2.875,    8.227,    K22:HB*,     K22:HN
1.396,    4.104,    K22:HG2,     K22:HA
1.396,    1.811,    K22:HG2,     K22:HB
1.396,    1.594,    K22:HG2,     K22:HD1
1.506,    4.104,    K22:HG1,     K22:HA
1.506,    1.811,    K22:HG1,     K22:HB
1.396,    2.883,    K22:HG2,     K22:HE
1.396,    1.511,    K22:HG2,     K22:HG1
1.506,    2.883,    K22:HG1,     K22:HE
1.506,    1.404,    K22:HG1,     K22:HG2
1.396,    8.227,    K22:HG2,     K22:HN
1.506,    8.227,    K22:HG1,     K22:HN
8.221,    4.104,    K22:HN,      K22:HA
8.221,    1.811,    K22:HN,      K22:HB
8.221,    1.594,    K22:HN,      K22:HD1
8.221,    2.883,    K22:HN,      K22:HE
8.221,    1.511,    K22:HN,      K22:HG1
8.221,    1.404,    K22:HN,      K22:HG2
4.098,    7.832,    K22:HA,      S23:HN
8.221,    4.295,    K22:HN,      S23:HA
8.221,    3.921,    K22:HN,      S23:HB1
8.221,    7.832,    K22:HN,      S23:HN
8.221,    6.995,    K22:HN,      X24:HN1
8.221,    7.128,    K22:HN,      X24:HN2
4.285,    7.128,    S23:HA,      W3:H6
7.823,    2.085,    S23:HN,      E19:HB
4.285,    0.757*,   S23:HA,      I21:HD2
4.285,    1.020,    S23:HA,      I21:HG12
4.285,    0.847*,   S23:HA,      I21:HG2
7.823,    1.062,    S23:HN,      I21:HG12
7.823,    0.847*,   S23:HN,      I21:HG2
4.285,    1.811,    S23:HA,      K22:HB
4.285,    1.594,    S23:HA,      K22:HD1
4.285,    1.511,    S23:HA,      K22:HG1
4.285,    1.404,    S23:HA,      K22:HG2
4.285,    8.227,    S23:HA,      K22:HN
7.823,    4.104,    S23:HN,      K22:HA
7.823,    1.811,    S23:HN,      K22:HB
7.823,    8.227,    S23:HN,      K22:HN
4.285,    3.921,    S23:HA,      S23:HB1
4.285,    3.817,    S23:HA,      S23:HB2
4.285,    7.832,    S23:HA,      S23:HN
3.874,    4.295,    S23:HB2,     S23:HA
3.913,    4.295,    S23:HB1,     S23:HA
3.874,    7.832,    S23:HB2,     S23:HN
3.913,    7.832,    S23:HB1,     S23:HN
7.823,    4.295,    S23:HN,      S23:HA
7.823,    3.921,    S23:HN,      S23:HB1
7.823,    3.817,    S23:HN,      S23:HB2
7.823,    6.995,    S23:HN,      X24:HN1
7.823,    7.128,    S23:HN,      X24:HN2
6.985,    3.823,    X24:HN2,     I21:HA
6.985,    4.104,    X24:HN2,     K22:HA
7.121,    8.227,    X24:HN1,     K22:HN
6.985,    4.295,    X24:HN2,     S23:HA
6.985,    3.921,    X24:HN2,     S23:HB1
7.121,    4.295,    X24:HN1,     S23:HA
7.121,    3.921,    X24:HN1,     S23:HB1
```

```
6.985,    7.832,    X24:HN2,     S23:HN
7.121,    7.832,    X24:HN1,     S23:HN
6.985,    7.128,    X24:HN2,     X24:H1
7.121,    6.992,    X24:HN1,     X24:H1
```

Chapter 12:
Summary and Future Development Plan

12.1 Summary

The primary result of this project has been the development of an NMR software package called *the Protein Analysis Workshop for NMR Spectroscopy* (or PAW) and its application to the NMR data processing and spectral assignment of a small protein, Caerin *4.1*.

The software has been designed for 2D NMR data processing and spectral analysis. It has been developed under the UNIX-based X-window operating system using a Silicon Graphics computer. The computer language used has been standard C with the *Motif Toolkit* from the *Open Software Foundation* (OSF). Many useful features and functions have been implemented into the software package. The user-friendly mouse- and command-interface systems, as well as the macro-definable working environment, provide users with lots of flexibility.

The data processing for the Caerin 4.1 NMR spectra has been rather difficult but fruitful. More than one-sixth of the DQF-COSY data set contained corrupted (receiver-overflow) values, and the entire data set was dominated by a water signal. In addition, the data sets were collected using a digitally filtered over-sampling technique that produced unusual FID curves. The data has been successfully processed with a high resolution processing method with the aid of linear prediction, filtering, automatic phasing, and baseline correction operations.

The spectral assignment for the Caerin 4.1 NOESY spectrum has been carried out by applying a systematic process. Three sets of peaks have been picked. These were the raw peaks, diagonal peaks and cross peaks. All the significant cross peaks have been fully assigned with the aid of various line-drawing, rectangle-drawing and peak-editing commands, as well as various spectral and peak-display options. The results have been summarised in an NOE-connectivity diagram, an assigned peak list, and PostScript plots for six significant regions in the NOESY spectrum.

11.2 Future Development

To make the package more valuable, more features could be added. These may include:

- An efficient object oriented programming for the entire package using C++. This may provide better programming environment.
- A system of 3D NMR data processing macro-commands and toolboxes. This will make the package even more powerful.
- A faster macro processing algorithm. This will reduce the time required for multi-dimensional data processing.
- Database referencing of ^{1}H, ^{13}C and ^{15}N chemical shifts for the 20 common amino acids. This may be used for predicting protein spin-systems and secondary structure prediction.

Appendixes:

Note:

The first letter of an appendix number indicates the chapter in which the appendix is first referred to, and the second letter, the order of the appendix's appearance in a chapter.

Appendix 4a:

Dialysis of Hen-egg Lysozyme

This appendix details the lab-notes for the dialysis of hen-egg lysozyme. It is presented because the sample obtained by this method has provided good quality NMR spectra of the lysozyme for the initial stage of the project.

Date: 19 August 1994
Task: Dialyse 500 mg of hen egg-white lysozyme in distilled water at pH 3.0.
Material:

3 to 4 litres of distilled water at pH 3.0
hen-egg lysozyme (product L6876 from SIGMA, MW 14,300)
1 M HCl, 0.1 M NaOH, ethanol
12 mm sterilised dialysis tubing, tubing clips, parafilm

Instrument:

10 ml universal bottle (can hold a total volume of 20 ml)
pH meter
200 ml beaker, 200 μL pipette, and rubber bulb

Method:

- Weigh the 10 ml universal bottle and zero the scales.
- Weigh 500 mg of the lysozyme into the bottle.
- Add about 3 ml of distilled water and 200 μL of 1M HCl to help dissolve the lysozyme. Shake the bottle so that the lysozyme is completely dissolved.
- With a pH meter, adjust the pH to 3 by gradually adding 1M HCl to it.
- Pick up the sterilised dialysis tubing and wash it with distilled water or the buffer. Also watch carefully for possible leakage. Clip one end.
- Open the unclipped end of the tubing by slightly twisting it with two fingers. Put the lysozyme solution into the tubing with a glass pipette. Clip the top end of the tubing. (Allow some space inside the clipped tubing so that water can get into it during dialysis.)
- Put the tubing into the buffer and seal the buffer container with parafilm.
- Put the buffer on top of a magnetic stirrer with a magnetic bar inside, then start the stirrer with medium speed.
- Change the buffer every 24 hours for three days. (According to a suggestion from Christina Redfield, the buffer can be changed three times in 24 hours.[1])

[1] Private communication.

Appendix 6a:

Cubic spline

A cubic spline is made up of piecewise cubic polynomials that connect together smoothly. It is used mathematically as a simple model to fit a curve for which the expression is either unknown or too complicated for numerical or analytical calculation. The method can be found in many books on numerical analysis (e.g., Press et al. 1986). Many of them describe cubic spline in a way that is unnecessarily complicated.

There are several conditions that can be applied to achieve the necessary smoothness of the connection. This appendix describes a cubic spline method used in PAW in a simple way. It requires the consistency in the first derivatives at each sampled point. Hence, a set of properly calculated first derivatives must be provided, as already mentioned in Section 6.3.13.

The problem can be stated as follows:

Given two points (t_1, y_1) and (t_2, y_2), as well as the first derivatives of the cubic polynomial at the two points, k_1 and k_2, find a cubic polynomial

$$y = a(t-t_1)^3 + b(t-t_1)^2 + c(t-t_1) + d \tag{6a.1}$$

that passes through the two points such that

$$y'(t_1) = k_1 \qquad \text{and} \qquad y'(t_2) = k_2 \tag{6a.2}$$

where

$$y' = 3a(t-t_1)^2 + 2b(t-t_1) + c \tag{6a.3}$$

Substitute the paired values (t_1, y_1) and (t_2, y_2) into (6a.1) and (6a.3) yields

$$\begin{aligned} y_1 &= d \\ y_2 &= a\Delta^3 + b\Delta^2 + c\Delta + d \\ k_1 &= c \\ k_2 &= 3a\Delta^2 + 2b\Delta + c \end{aligned} \tag{6a.4}$$

where

$$\Delta = (t_2 - t_1) \tag{6a.5}$$

Therefore, the solution is the set of equations:

$$
\begin{aligned}
a &= \frac{k_2 + k_1 - 2m_1}{\Delta^2} \\
b &= -\frac{k_2 + 2k_1 - 3m_1}{\Delta} \\
c &= k_1 \\
d &= y_1
\end{aligned}
\tag{6a.6}
$$

where

$$
m_1 = \frac{y_2 - y_1}{\Delta} \tag{6a.7}
$$

This technique does not require knowledge of the second derivatives. A spline-fitting program based on this solution has proven to be satisfactory.

Appendix 6b:

The Discrete Fourier Transform

Fourier transformation reveals a wonderful relationship between data in time domain and in frequency domain:

$$\mathcal{F}[f(t);\omega] = \int_{-\infty}^{\infty} f(t)\exp(i\omega t)dt = F(\omega) \tag{6b.1}$$

The discrete Fourier transform (see, e.g., Press *et al.* 1986) estimates the Fourier transform of a discrete data set h_k, $k = 0, 1, ..., N-1$, with a discrete sum:

$$H_\omega = \Delta_t \sum_{k=0}^{N-1} h_k \exp(i2\pi\omega k/N) \tag{6b.2}$$

Here,

$$\omega = -\frac{N}{2}, -\frac{N}{2}+1, ..., \frac{N}{2} \tag{6b.3}$$

In programming, the formulae containing parameters related to Fourier transformation are often quite confusing. To obtain correct formulae, it is useful to investigate the analytical results based on the integrals that correspond to the discrete Fourier transform. If N is very large, the analogue of Equation (6b.2) is either

$$H(\omega) = \int_0^{N-1} h(t)\exp\left(i\frac{2\pi}{N}\omega t\right)dt \tag{6b.4}$$

or

$$H(\omega) = \int_0^{\infty} h(t)\exp\left(i\frac{2\pi}{N}\omega t\right)dt \tag{6b.5}$$

Note that the integration is no longer over the interval from $-\infty$ but 0. Also, there is an additional factor $\frac{2\pi}{N}$ in the exponent.

In theory, the ideal NMR signal and corresponding spectrum are often presented as

$$s(t) = \exp(-at)\exp(-i\omega_0 t), \quad t, a \geqslant 0 \tag{6b.6}$$

$$S(\Omega) = \int_0^{\infty} \exp(-|a|t)\exp(i|\Omega|t)dt = \frac{|a| + i\Omega}{a^2 + \Omega^2} \tag{6b.7}$$

where

$$\Omega = \omega - \omega_0 \tag{6b.8}$$

In reality, NMR data sets inherent a factor of $\frac{2\pi}{N}$ for the independent variable t if the increments Δ_t is taken as 1 unit. Typically,

$$g(t) = \exp\left(-a \times \frac{2\pi t}{N}\right)\exp\left(-i\omega_0 \times \frac{2\pi t}{N}\right) \tag{6b.9}$$

Hence, its discrete Fourier transform corresponds to the following integration:

$$G(\Omega) = \int_0^{\infty} \exp\left(-|a|\frac{2\pi t}{|N|}\right)\exp\left(i\Omega\frac{2\pi t}{|N|}\right)dt = \frac{|N|}{2\pi}\frac{|a| + i\Omega}{a^2 + \Omega^2} \qquad (6b.10)$$

For example, for $a = 0.5$ and $\omega_0 = 10$, then,

$$g(t) = \exp\left(-0.5 \times \frac{2\pi t}{1024}\right)\exp\left(-i10 \times \frac{2\pi t}{1024}\right) \qquad (6b.11)$$

It has 10 sinusoidal maxima within 1024 points and decays exponentially with $a = 0.5$, as shown in Figure 6b.1.

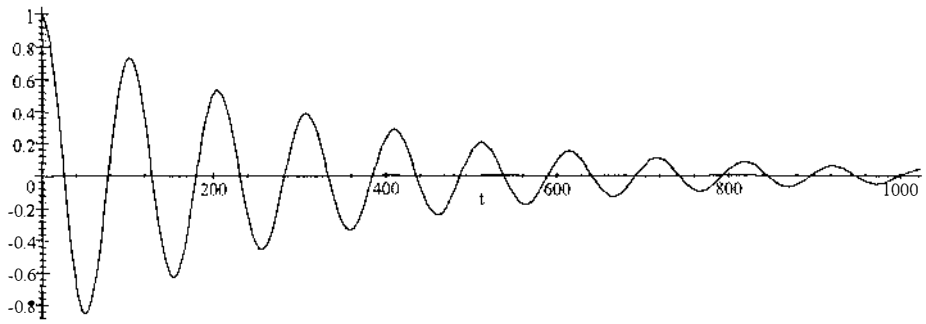


Figure 6b.1: The real part of $g(t) = \exp\left(-0.5 \times \frac{2\pi t}{1024}\right)\exp\left(-i10 \times \frac{2\pi t}{1024}\right)$.

The 1024-point sampled values for the plot are

$$g_k \in \{1,\ e^{-0.5\eta}\cos(10\eta),\ e^{-0.5\times 2\eta}\cos(10 \times 2\eta),\ldots,\ e^{-0.5\times 1023\eta}\cos(10 \times 1023\eta)\} \qquad (6b.12)$$

where

$$\eta = \frac{2\pi}{1024} \qquad (6b.13)$$

The corresponding discrete Fourier transform is

$$\begin{aligned} G(\Omega)_{\text{real}} &= \int_0^{\infty} \exp\left(-0.5 \times \frac{2\pi|t|}{1024}\right)\exp\left(i\Omega \times \frac{2\pi|t|}{1024}\right)dt \\ &= \frac{1024}{2\pi}\frac{0.5}{0.5^2 + (\omega - 10)^2} = \frac{325.95}{1.0 - 2.0i\Omega} \end{aligned} \qquad (6b.14)$$

where

$$\Omega = \omega - 10 \qquad (6b.15)$$

It represents a peak that is 10 points from the spectral centre, as shown in Figure 6b.2.

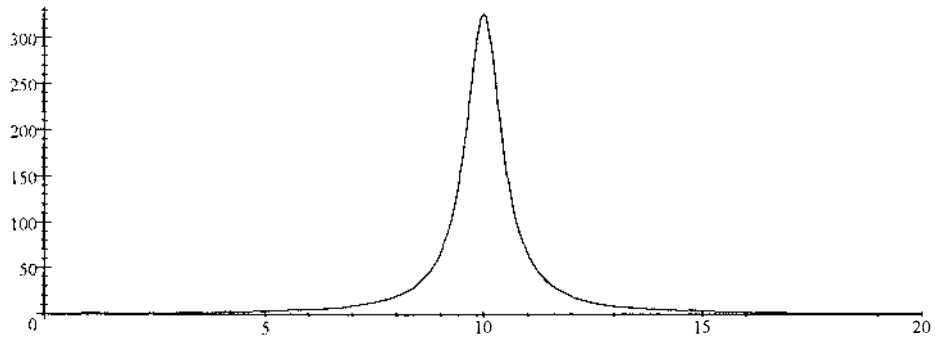


Figure 6b.2: The real plot of $G(\omega) = \frac{1024}{2\pi}\frac{0.5}{0.5^2+(\omega-10)^2}$.

It can be seen that the peak width at half height is $2a = 1$, which is also derived in Appendix 6d.

The last two figures can be produced with PAW as a varification of the theory. The steps involved involved are:

- Generate a spectrum of 10 Hz,
- Filter with exp(500), which is equivalent to $\exp\left(-0.5 \times \frac{2\pi t}{1024}\right)$.
- Apply a Fourier transform.

The common format of Eq.(6b.10) and (6b.14) indicates the following general integration associated with the discrete Fourier transform in NMR spectroscopy:

$$\mathcal{F}_d[f(t);\Omega] = \int_0^\infty f\left(\frac{2\pi t}{N}\right)\exp\left(i\Omega\frac{2\pi t}{N}\right)dt \qquad (6b.16)$$

Since $\exp\left(i\Omega\frac{2\pi t}{N}\right) = \cos\left(\Omega\frac{2\pi t}{N}\right) + i\sin\left(\Omega\frac{2\pi t}{N}\right)$, the discussion can also be extended to and integration from $-\infty$ to $+\infty$ if $g(t)$ is an even or odd funtion. In this cases, the result are:

$$\mathcal{F}_d[f(t);\Omega]_{\text{real}} = \frac{1}{2}\int_{-\infty}^{\infty} g\left(\frac{2\pi t}{N}\right)\exp\left(i\Omega\frac{2\pi t}{N}\right)dt \quad \text{if } g(t) \text{ is an even funtion} \qquad (6b.17)$$

and

$$\mathcal{F}_d[f(t);\Omega]_{\text{imag}} = \frac{1}{2}\int_{-\infty}^{\infty} g\left(\frac{2\pi t}{N}\right)\exp\left(i\Omega\frac{2\pi t}{N}\right)dt \quad \text{if } g(t) \text{ is an odd funtion} \qquad (6b.18)$$

Simply because the integration of any odd function over the inteval is 0.

Comparing with Equation (6b.1), it can be shown that the right-hand side of the last two equations only differ from the Fourier transform by a factor of $\frac{N}{\pi}$. That is,

$$\frac{1}{2}\int_{-\infty}^{\infty} f\left(\frac{2\pi t}{N}\right)\exp\left(i\Omega\frac{2\pi t}{N}\right)dt = \frac{N}{\pi}\int_{-\infty}^{\infty} f(t)\exp(i\Omega t)dt = \frac{N}{\pi}\mathcal{F}[f(t);\Omega] \qquad (6b.19)$$

Some examples of these integrations can be found in Appendix 6.c to 6.e.

Appendix 6c:

Gaussian Peaks in NMR Spectra

This appendix discusses the properties of the Gaussian peaks in NMR spectra using the general integration associated with the discrete Fourier transform in NMR spectroscopy. (See Appendix 6b.)

1D Gaussian peak in NMR spectroscopy

Let a signal be

$$g(t) = M_0 \exp\left(-|b|\left(\frac{2\pi t}{N}\right)^2\right)\exp\left(-i\omega_0\left(\frac{2\pi t}{N}\right)\right), \quad t = 0 \text{ to } \infty \qquad (6c.1)$$

Let also the resonance frequency in the rotating frame Ω be:

$$\Omega = \omega - \omega_0 \qquad (6c.2)$$

Since $g(t)$ is an even function, the real part of the corresponding 1D Gaussian peak is:

$$\begin{aligned} G(\Omega)_{\text{real}} &= \frac{1}{2}\int_{-\infty}^{\infty} M_0 \exp\left(-|b|^2\left(\frac{2\pi t}{N}\right)^2\right)\exp\left(i\Omega\frac{2\pi t}{N}\right)dt \\ &= \frac{NM_0}{4b\sqrt{\pi}}\exp\left(-\frac{1}{4b^2}\Omega^2\right) \\ &= G_0 \exp\left(-\frac{1}{4b^2}\Omega^2\right) \end{aligned} \qquad (6c.3)$$

Here, G_0 is the peak height of the real part at $\Omega = 0$:

$$G_0 = \frac{|N|M_0}{4|b|\sqrt{\pi}} \qquad (6c.4)$$

The Ω at which $G(\Omega) = \frac{1}{2}G_0$ is

$$\Omega_{\frac{1}{2}} = \pm\sqrt{-4b^2 \ln 0.5} \qquad (6c.5)$$

Therefore, the peak-width at half height is

$$W_{\frac{1}{2}} = 2\sqrt{-4b^2 \ln 0.5} \qquad (6c.6)$$

Examples

Suppose that the real part of a signal is

$$g_{\text{real}}(t) = \exp\left(-|0.9|^2\left(\frac{2\pi t}{1024}\right)^2\right)\cos\left(25 \times \frac{2\pi t}{1024}\right) \qquad (6c.7)$$

Then,

$$G_{\text{real}}(\Omega) = \frac{1024}{4 \times 0.9\sqrt{\pi}}\exp\left(-\frac{1}{4 \times 0.9^2}\Omega^2\right) \qquad (6c.8)$$

$$W_{\frac{1}{2}} = 2\sqrt{-4 \times 0.9^2 \times \ln 0.5} = 2.171 \qquad (6c.9)$$

The following two figures show the plots of $g_{\text{real}}(t)$ and $G_{\text{real}}(\Omega)$.

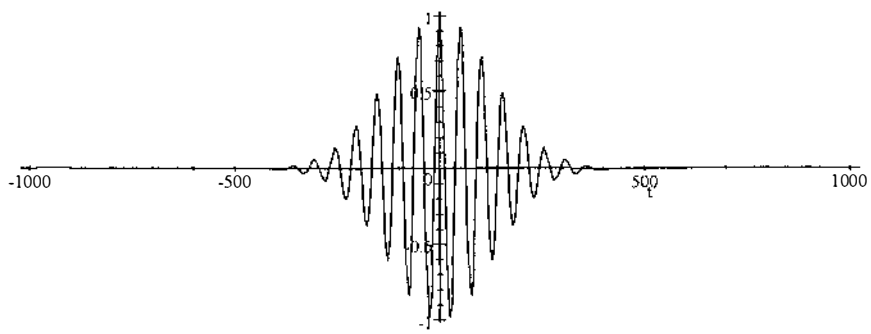


Figure 6c.1: The plot of $g_{\text{real}}(t) = \exp\left(-|0.9|^2\left(\frac{2\pi t}{1024}\right)^2\right)\cos\left(25 \times \frac{2\pi t}{1024}\right)$.

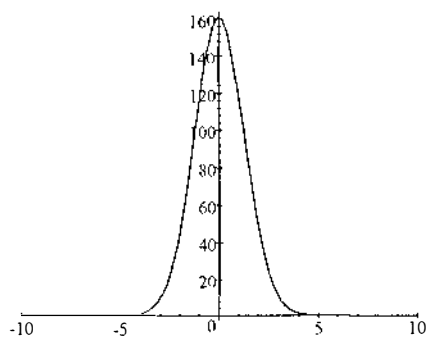


Figure 6c.2: The plot of $G_{\text{real}}(\Omega) = \frac{1024}{4\times 0.9\sqrt{\pi}}\exp\left(-\frac{1}{4\times 0.9^2}\Omega^2\right)$.

The area of 1D Gaussian peaks in NMR spectroscopy

The area of the 1D Gaussian peak, A_{Gauss}, resulting from the Fourier transform is:

$$\begin{aligned} A_{\text{Gauss}} &= \int_{-\infty}^{\infty} G_0 \exp\left(-\frac{1}{4|b|^2}\Omega^2\right) d\Omega \\ &= 2|b|\sqrt{\pi}\, G_0 \\ &= \frac{1}{2} N M_0 \end{aligned} \qquad (6c.10)$$

which is different from the first-point theorem of normal Fourier transform by a factor of $\frac{1}{2}N$. This is the formula used in PAW for the calculation of 1D-peak areas.

In practice, G_0 is known if the peak is isolated. If an FID contains a number of resonance frequencies, the parameter b for each peak can be obtained by curve-fitting with multiple Gaussians. Then, M_0 can be obtained from the values of G_0 and b.

The volume of 2D Gaussian peaks in NMR spectroscopy

For a 2D Gaussian peak, the equation is

$$G(\Omega_1,\Omega_2) = G_0 \exp\left(-\frac{1}{4|b_1|^2}\Omega_1^2\right)\exp\left(-\frac{1}{4|b_2|^2}\Omega_2^2\right) \qquad (6c.11)$$

The peak integration then yields

$$V_{\text{Gauss}} = G_0 \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \exp\left(-\frac{1}{4|b_1|^2}\Omega_1^2\right)\exp\left(-\frac{1}{4|b_2|^2}\Omega_2^2\right)d\Omega_1 d\Omega_2$$
$$= 4\pi b_1 b_2 G_0 \qquad (6c.12)$$

Therefore, three parameters are required for the calculation of a peak volume. This is the formula used in PAW for the calculation of 2D-peak volumes.

The 3D and countour plots of a 2D Gaussian peak are shown in the next figure.

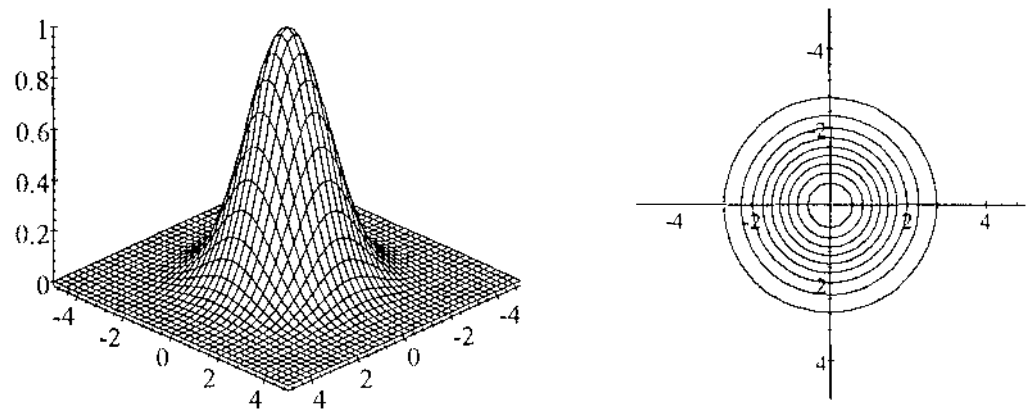


Figure 6c.3: The 3D (left) and contour (right) plot of $S(\Omega_1,\Omega_2) = \exp\left(-\frac{1}{4\times 0.9^2}\Omega_1^2\right)\exp\left(-\frac{1}{4\times 0.9^2}\Omega_2^2\right)$

Appendix 6d:

Lorentzian Peaks in NMR Spectra

This appendix discusses the properties of the Lorentzian peaks in NMR spectra using the general integration associated with the discrete Fourier transform in NMR spectroscopy. (See Appendix 6b.)

1D Lorentzian peak

Let a signal $g(t)$ be

$$g(t) = M_0 \exp\left(-a\left(\frac{2\pi t}{N}\right)\right) \exp\left(-i\omega_0\left(\frac{2\pi t}{N}\right)\right) \tag{6d.1}$$

Let also the resonance frequency in the rotating frame Ω be:

$$\Omega = \omega - \omega_0 \tag{6d.2}$$

The corresponding 1D Gaussian peak is:

$$\begin{aligned} G(\Omega) &= M_0 \int_0^\infty \exp\left(-|a|\left(\frac{2\pi t}{|N|}\right)\right) \exp\left(i\Omega \frac{2\pi t}{|N|}\right) dt \\ &= \frac{|N|M_0}{2\pi} \frac{|a| + i\Omega}{|a|^2 + \Omega^2} \\ &= G_0 \frac{1 + i(\Omega/a)}{1 + (\Omega/a)^2} \end{aligned} \tag{6d.3}$$

For the real part, G_0 is the peak height at $\Omega = 0$:

$$G_0 = \frac{NM_0}{2\pi a} \tag{6d.4}$$

The Ω at which $G(\Omega) = \frac{1}{2}G_0$ can be obtained from the following equation:

$$\frac{1}{2}G_0 = G_0 \frac{1}{1 + (\Omega/a)^2} \tag{6d.5}$$

The solution is

$$\Omega_{\frac{1}{2}} = \pm a \tag{6d.6}$$

Therefore, the peak-width at half height is

$$W_{\frac{1}{2}} = 2a \tag{6d.7}$$

An example of exponentially decaying 1D FID and the corresponding Lorentzian peak can be found in ASppendix 6.b.

The area of 1D Lorentzian peak in NMR spectroscopy

The 1D peak area A_{1D} is $\frac{1}{2}NM_0$, which can be verified by the peak integration:

$$A_{\text{Lorentzian}} = \frac{NM_0}{2\pi}\int_{-\infty}^{\infty}\frac{|a|}{|a|^2+\Omega^2}d\Omega$$
$$= \frac{1}{2}NM_0$$
$$= \pi a G_0 \qquad (6d.8)$$

This is the formula used in PAW for the calculation of 1D-peak areas.

In practice, G_0 is known if the peak is isolated. If an FID contains a number of resonance frequencies, the parameter a for each peak can be obtained by curve-fitting with multiple Lorentzians. Then, M_0 can be obtained from the values of G_0 and a.

2D Lorentzian peaks and peak-volumes in NMR spectroscopy

For a 2D Lorentzian peak, the equation is

$$G(\Omega_1,\Omega_2) = G_0\frac{1}{1+(\Omega_1/a_1)^2}\frac{1}{1+(\Omega_2/a_2)^2} \qquad (6d.9)$$

The peak integration then yields

$$G_0\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{1}{1+(\Omega_1/|a_1|)^2}\frac{1}{1+(\Omega_2/|a_2|)^2}d\Omega_1 d\Omega_2 \qquad (6d.10)$$

$$V_{\text{Lorentzian}} = G_0\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{1}{1+(\Omega_1/|a_1|)^2}\frac{1}{1+(\Omega_2/|a_2|)^2}d\Omega_1 d\Omega_2$$
$$= \pi^2 a_1 a_2 G_0 \qquad (6d.11)$$

The 3D and contour plots of a 2D Lorentzian peak are shown in the next figure.

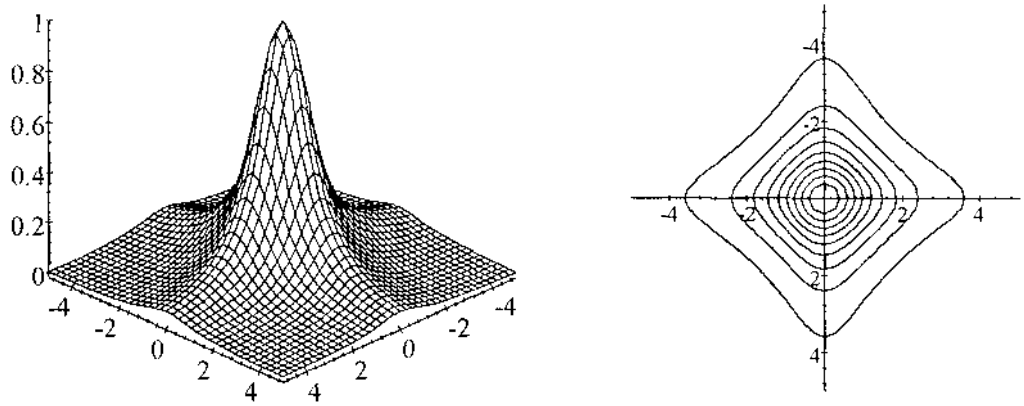


Figure 6c.3: The 3D (left) and contour (right) plots of
$S(\Omega_1,\Omega_2) = \frac{1}{1+(\Omega_1/1.2)^2}\frac{1}{1+(\Omega_2/1.2)^2}$

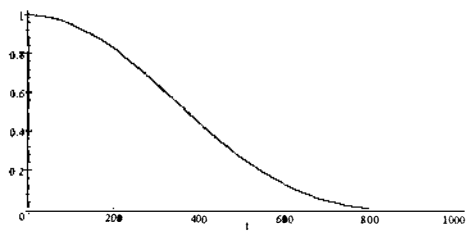
Appendix 6e:

Sine-bell Filtered Peaks in NMR Spectra

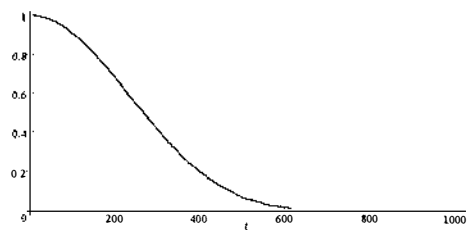
This appendix conducts a discussion similar to that presented in Appendix 6.b, but is for the sine-bell filtered peaks in NMR spectra.

Sine-bell filters

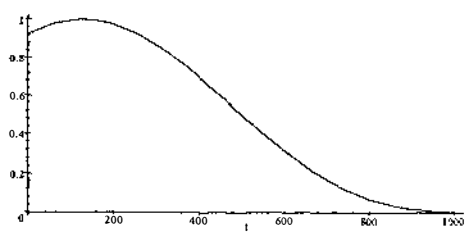
A sine-bell filter in spectral processing contains part of a (shifted) sine function that spans the entire time domain. Since cosine functions are shifted sine functions, they are also considered as sine-bell filters. Figure 6e.1 shows a few examples of the sine-bell filters:



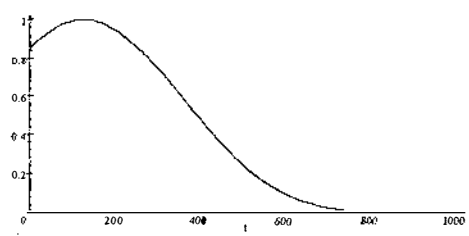
(1) $W(t) = \frac{1}{4}\left(\cos\left(\frac{\pi t}{1024}\right) + 1\right)^2$



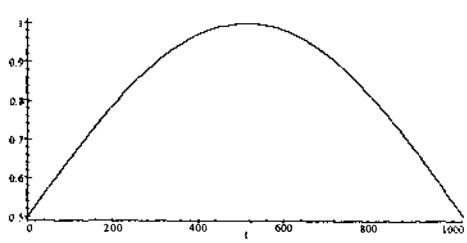
(2) $W(t) = \frac{1}{16}\left(\cos\left(\frac{\pi t}{1024}\right) + 1\right)^4$.



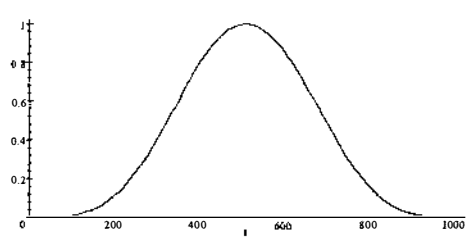
(3) $W(t) = \frac{1}{4}\left(\cos\left(\frac{\pi t}{1024} - \frac{\pi}{8}\right) + 1\right)^2$



(4) $W(t) = \frac{1}{16}\left(\cos\left(\frac{\pi t}{1024} - \frac{\pi}{8}\right) + 1\right)^4$.



(5) $W(t) = \frac{1}{2}\left(\sin\left(\frac{\pi t}{1024}\right) + 1\right)$



(6) $W(t) = \frac{1}{4}\left(\sin\left(\frac{2\pi t}{1024} - \frac{\pi}{2}\right) + 1\right)^2$.

Figure 6e.1: Various sine-bell filters.

Here, the fractions leading the expressions of each $W(t)$ normalise the filters. Filter 6 is not a good filter for NMR data processing because it suppresses the initial part too

much.

1D sine-bell-filtered peak

In NMR spectroscopy, a filtered signal can be expressed as:

$$g(t) = M_0 W(\frac{2\pi t}{N}) \exp\left(-a\left(\frac{2\pi t}{N}\right)\right) \exp\left(-i\omega_0\left(\frac{2\pi t}{N}\right)\right) \qquad (6e.1)$$

The following two figures show the profiles of an NMR signal before and after filtering by an enhanced sine-bell function.

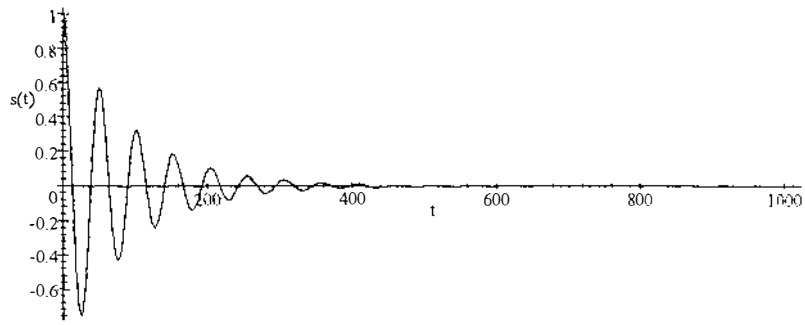


Figure 6e.2: The profiles of an NMR data set before being filtered, where $g(t) = \exp\left(-1.8 \times \frac{2\pi t}{1024}\right)\cos\left(20 \times \frac{2\pi t}{1024}\right)$.

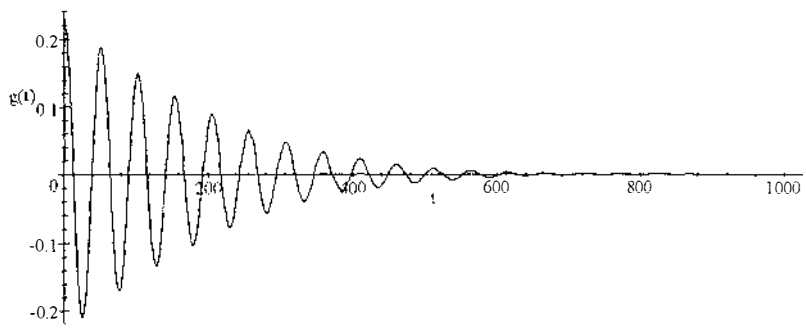


Figure 6e.3: The profiles of an NMR data set after filtering by a partially enhanced sine-bell function, where $g(t) = \frac{1}{16}\left(\cos\left(\frac{\pi t}{1024} - \frac{\pi}{8}\right) + 1\right)^2 \exp\left(-0.8 \times \frac{2\pi t}{1024}\right)\cos\left(20 \times \frac{2\pi t}{1024}\right)$.

The Fourier transform of real signal can be very complicated in theoretical derivation. To simplify the problem, only the following profile of fully enhanced signals will be discussed.

$$g(t) = \frac{M_0}{2^m}\left(\cos\left(\frac{\pi t}{N}\right) + 1\right)^m \exp\left(-i\omega_0 \frac{\pi t}{N}\right) \qquad (6e.2)$$

where m is the degree of the sine function. The result can be achieved by applying PAW's enhanced sine-bell filter to an FID. Figure 6e.4 shows the profile of a fully enhanced signal.

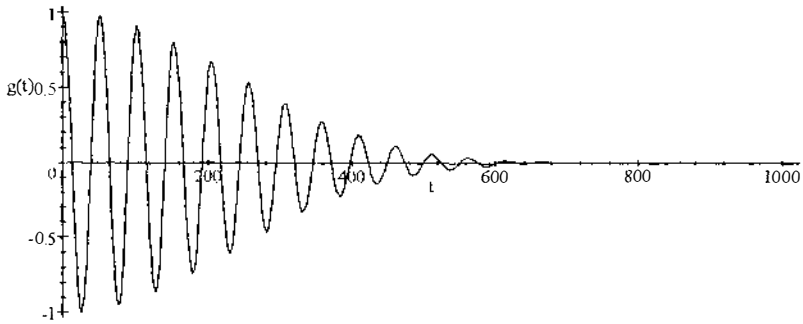


Figure 6e.4: The profiles of an ideal FID filtered by a fully enhanced sine-bell function of order 4, where $g(t) = \frac{1}{16}\left(\cos\left(\frac{\pi t}{1024}\right) + 1\right)^4 \cos\left(20 \times \frac{2\pi t}{1024}\right)$.

Also, only the definite integration from 0 to N will be discussed because sinusoidal functions do not converge.

Let the resonance frequency in the rotating frame Ω be:

$$\Omega = \omega - \omega_0 \tag{6e.3}$$

Then, for $m = 4$, the integration leads to

$$\begin{aligned} G(\Omega) &= \frac{M_0}{16}\int_0^{|N|}\left(\cos\left(\frac{\pi t}{|N|}\right)+1\right)^4 \exp\left(i\Omega\frac{2\pi t}{|N|}\right)dt \\ &= \frac{NM_0}{64} i\frac{-315e^{2i\Omega\pi} + 512\Omega^8 - 3584\Omega^6 + 7168\Omega^4 - 4096\Omega^2 + 315}{16\pi\Omega^9 - 120\pi\Omega^7 + 273\pi\Omega^5 - 205\pi\Omega^3 + 36\pi\Omega} \end{aligned} \tag{6e.4}$$

Hence, the real part $G_{\text{real}}(\Omega)$, the peak height G_0, and the peak area A_{sb4} are

$$G_{\text{real}}(\Omega) = \frac{1}{64}\frac{315NM_0\sin(2\pi\Omega)}{16\pi\Omega^9 - 120\pi\Omega^7 + 273\pi\Omega^5 - 205\pi\Omega^3 + 36\pi\Omega} \tag{6e.5}$$

$$G_0 = \frac{1}{64}\lim_{\Omega\to 0}\frac{315NM_0\frac{d}{d\Omega}(\sin(2\pi\Omega))}{\frac{d}{d\Omega}(16\pi\Omega^9 - 120\pi\Omega^7 + 273\pi\Omega^5 - 205\pi\Omega^3 + 36\pi\Omega)} = \frac{35}{128}NM_0 \tag{6e.6}$$

$$\begin{aligned} A_{\text{sb4}} &= \frac{1}{64}\int_{-\infty}^{\infty}\frac{315NM_0\sin(2\pi\Omega)}{16\pi\Omega^9 - 120\pi\Omega^7 + 273\pi\Omega^5 - 205\pi\Omega^3 + 36\pi\Omega}d\Omega \\ &= \frac{1}{2}NM_0 = \frac{64}{35}G_0 \end{aligned} \tag{6e.7}$$

By equalising $G_{\text{real}}(\Omega)$ and $\frac{1}{2}G_0$, it can be found numerically that $\Omega_{\frac{1}{2}}$ is independent of N and M_0:

$$\Omega_{\frac{1}{2}} = \pm 0.873\,48 \tag{6e.8}$$

Therefore, the peak width at half height is always $2\Omega_{\frac{1}{2}} = 1.747$ for $m = 4$.

Figure 6e.5 shows a typical 1D peak resulting from a fully enhanced sine-bell filter of degree 4.

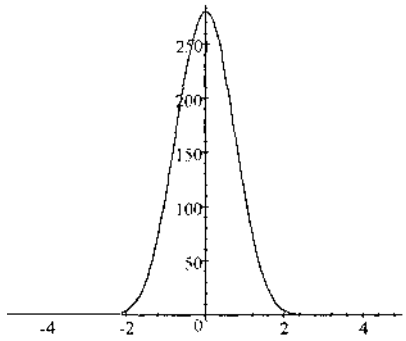


Figure 6e.5: A typical 1D peak resulting from a fully enhanced sine-bell filter of degree 4.

Figure 6e.6 shows the typical 3D and contour plots of a 2D peak resulting from a fully enhanced sine-bell filter of degree 4.

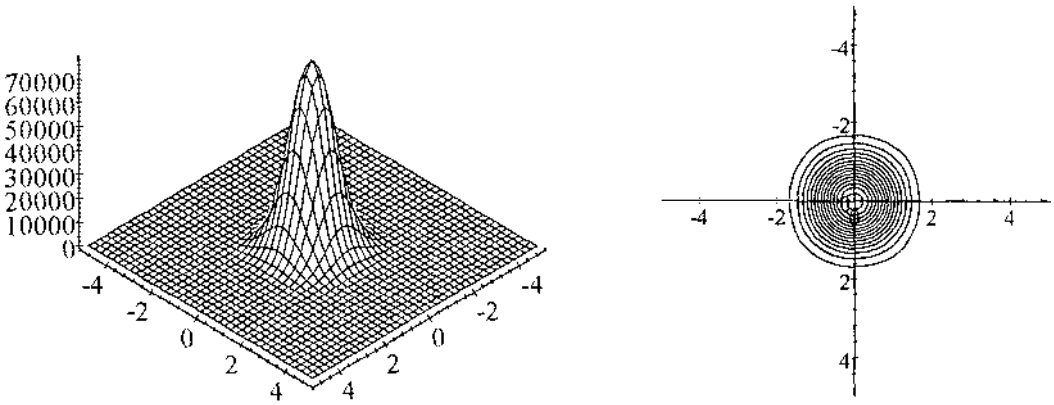


Figure 6e.6: The typical 3D and contour plots of a 2D peak resulting from a fully enhanced sine-bell filter of degree 4.

For $m = 2$, the corresponding results are

$$G(\Omega) = \frac{M_0}{4}\int_0^{|N|}\left(\cos\left(\frac{\pi t}{|N|}\right)+1\right)^2 \exp\left(i\Omega\frac{2\pi t}{|N|}\right)dt$$
$$= \frac{NM_0}{16} i\frac{-3e^{2i\Omega\pi}+32\Omega^4-32\Omega^2+3}{4\pi\Omega^5-5\pi\Omega^3+\pi\Omega} \qquad (6e.9)$$

$$G_{\text{real}}(\Omega) = \frac{3NM_0}{16}\frac{\sin(2\pi\Omega)}{4\pi\Omega^5-5\pi\Omega^3+\pi\Omega} \qquad (6e.10)$$

$$G_0 = \frac{3NM_0}{16}\lim_{\Omega\to 0}\frac{\frac{d}{d\Omega}\sin(2\pi\Omega)}{\frac{d}{d\Omega}(4\pi\Omega^5-5\pi\Omega^3+\pi\Omega)} = \frac{3}{8}NM_0 \qquad (6e.11)$$

$$A_{sb2} = \frac{3NM_0}{16}\int_{-\infty}^{\infty} \frac{\sin(2\pi\Omega)}{4\pi\Omega^5 - 5\pi\Omega^3 + \pi\Omega} d\Omega$$

$$= \frac{1}{2}NM_0 = \frac{4}{3}G_0 \qquad (6e.12)$$

$$\Omega_{\frac{1}{2}} = \pm 0.64744 \qquad (6e.13)$$

Therefore, the peak width at half height is always $2\Omega_{\frac{1}{2}} = 1.2949$ for $m = 2$.

Figure 6e.7 shows a typical 1D peak resulting from a fully enhanced sine-bell filter of degree 2.

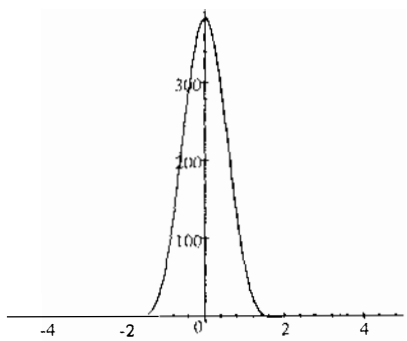


Figure 6e.7: A typical 1D peak resulting from a fully enhanced sine-bell filter of degree 2.

Figure 6e.8 shows the 3D and contour plots of a 2D peak resulting from a fully enhanced sine-bell filter of degree 2.

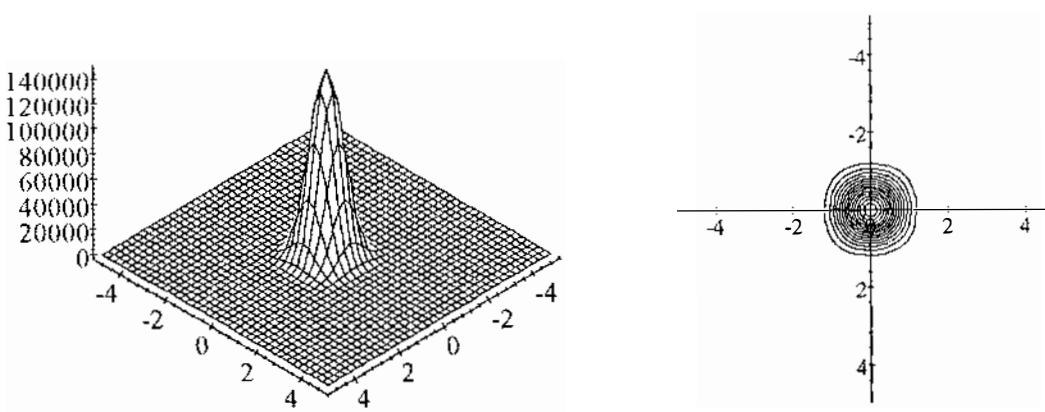


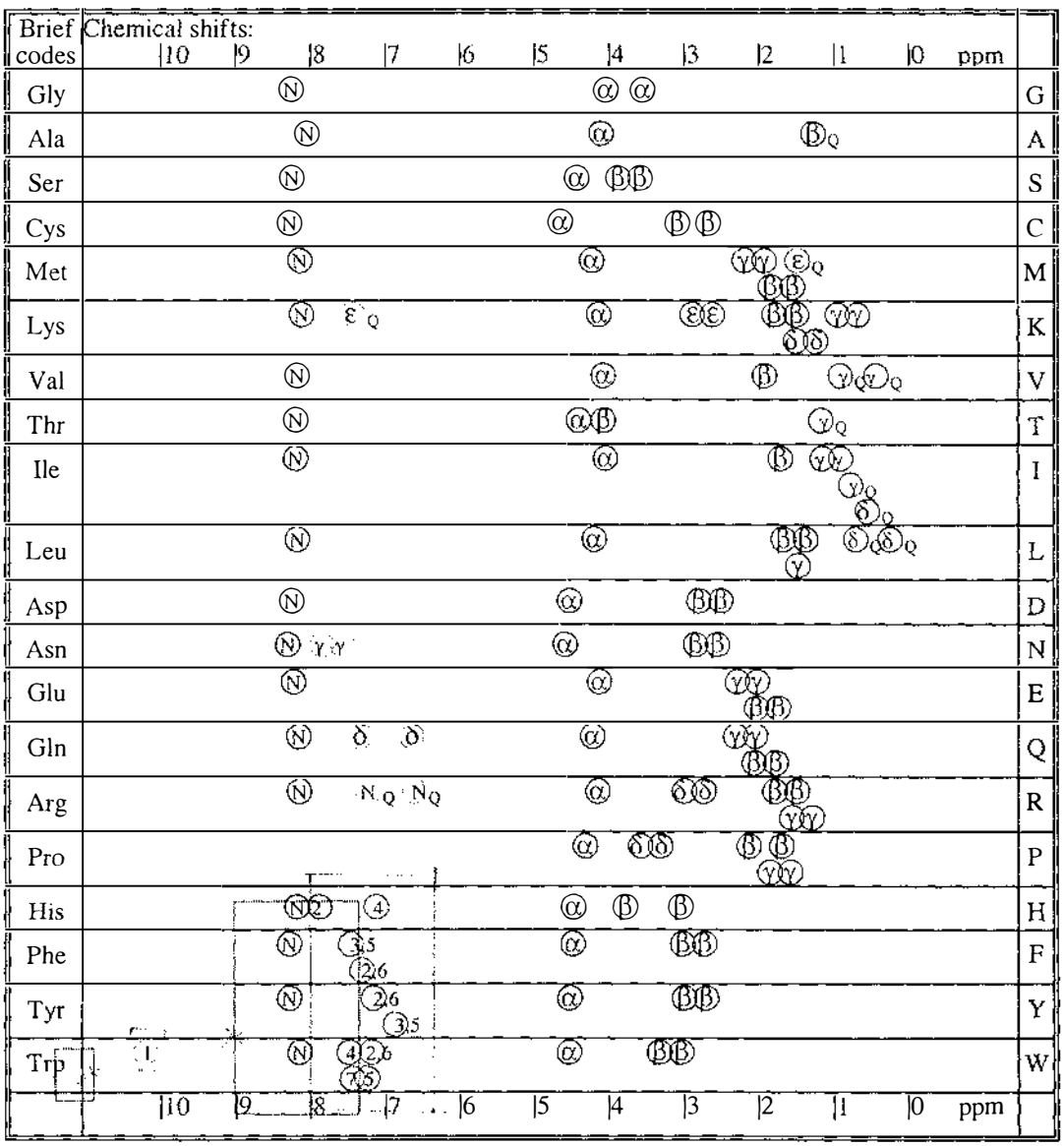
Figure 6e.8: The 3D and contour plots of a 2D peak resulting from a fully enhanced sine-bell filter of degree 2.

The results above have been verified using the mathematical-derivation software called Maple with Scientific Workplace.

Appendix 7a:

Protein ^1H Chemical-shifts (table2)

Distribution of ^1H chemical shifts of the 20 common amino acids using spin codes found in [Wüthrich 1986]. A subscript Q represents a number of ethyl (or methyl) spins that have an identical chemical shift. The spins of ^1H protons are shifted to another row if it is too crowded to be in the same row. Separate spin systems in aromatic rings are enclosed in rectangles. Spins of exchangeable protons are in dotted circles.



Appendix 7b:

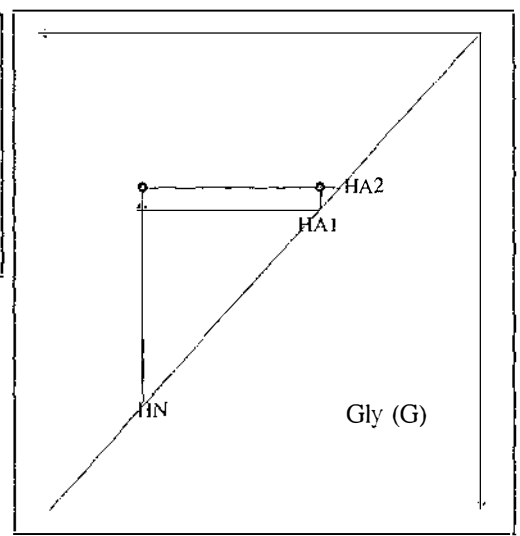
TOCSY Connectivity Diagrams of ^1H Spin-systems in Proteins

The following figures show the ^1H TOCSY spin-system connectivity diagrams of the 20 common amino acid residues.

Glycine — H

Average chemical shifts:

Nuclear	All	Helical	β-strand	Coil
HN	8.33	8.07	8.44	8.36
HA1	4.14	3.88	4.47	4.11
HA2	3.64	3.57	3.69	3.65



Alanine — $C^{\beta}H_3$

Average chemical shifts:

Nuclear	All	Helical	β-strand	Coil
HN	8.12	8.08	8.25	8.12
HA	4.20	4.05	4.67	4.20
HB	1.36	1.40	1.28	1.34

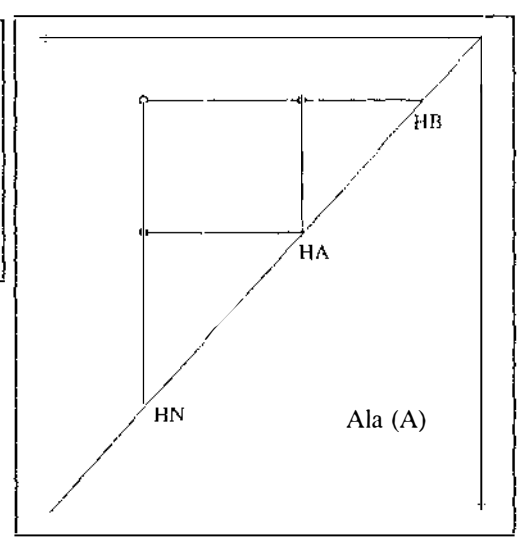
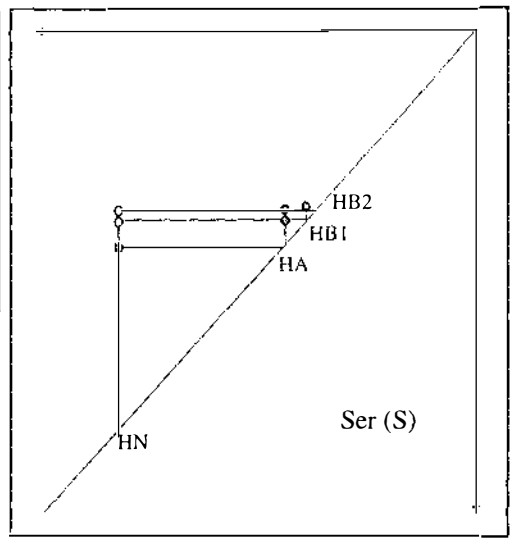


Figure 7b.1: The average chemical shifts of the 20 common amino acid residues. On the right are the connectivity diagrams of the ^1H TOCSY spin-systems. Solid circles represent neighbouring-proton connectivities like those in a COSY spectrum. Open cycles represent other TOCSY connectivities. A symbol * indicates that the chemical shift is not measured. Data for the protons in ring structures are taken from Wüthrich 1986; others are from Wishart et al. 1991.

Serine — $C^{\beta}H_2$ — O^{γ}

Average chemical shifts:

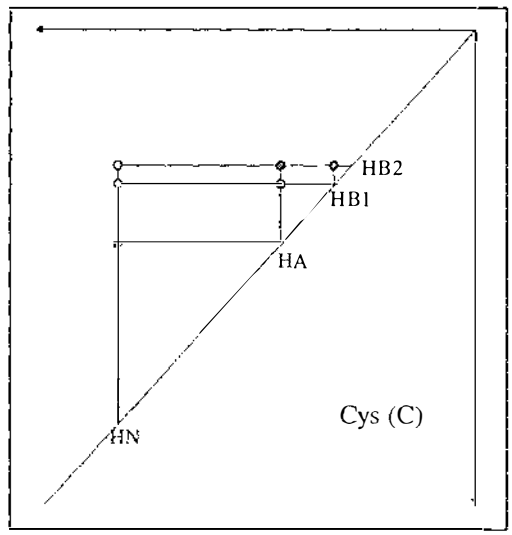
Nuclear	All	Helical	β-strand	Coil
HN	8.32	8.07	8.56	8.30
HA	4.52	4.25	4.96	4.38
HB1	3.95	3.97	3.90	3.95
HB2	3.79	3.81	3.72	3.83



Cysteine — $C^{\beta}H_2$ — S^{γ}

Average chemical shifts:

Nuclear	All	Helical	β-strand	Coil
HN	8.35	7.90	8.62	8.18
HA	4.74	4.15	5.10	4.52
HB1	3.16	3.01	3.32	3.17
HB2	2.76	2.76	2.82	2.80



Methionine — $C^{\beta}H_2$ — $C^{\gamma}H_2$ — S^{δ} — $C^{\varepsilon}H_3$

Average chemical shifts:

Nuclear	All	Helical	β-strand	Coil
HN	8.21	8.03	8.72	8.12
HA	4.32	4.07	4.88	4.32
HB1	1.92	1.88	2.07	1.84
HB2	1.66	1.61	1.87	1.57
HG1	2.25	2.18	2.46	2.20
HG2	2.00	1.92	2.39	1.87
HG2	1.53	1.54	1.53	1.47

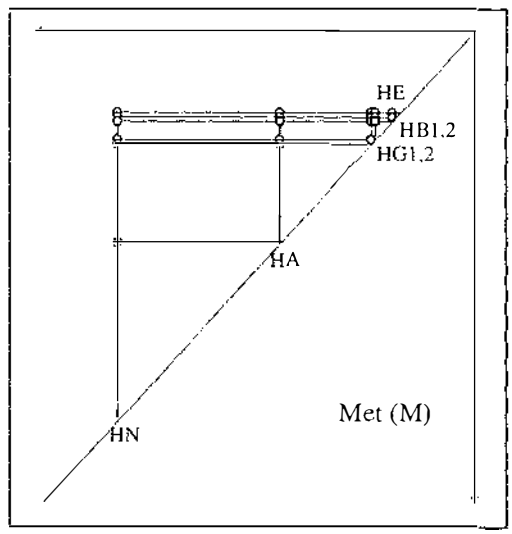


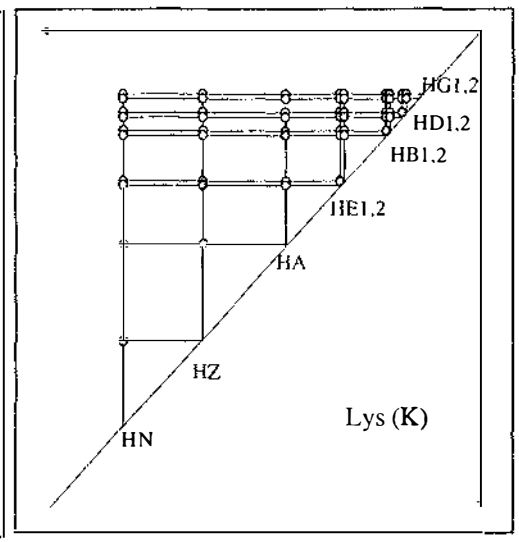
Figure 7b.1: The average chemical shifts and the spin-system connectivity of the 20 common amino acid residues. (Continued)

Lysine

$$-C^{\beta}H_2-C^{\gamma}H_2-C^{\delta}H_2-C^{\varepsilon}H_2-{}_{\varepsilon}N^{\zeta}H_3^{\oplus}$$

Average chemical shifts:

Nuclear	All	Helical	β-strand	Coil
HN	8.19	7.91	8.55	8.29
HA	4.23	3.95	4.72	4.23
HB1	1.87	1.88	1.79	1.91
HB2	1.72	1.76	1.64	1.73
HG1	1.42	1.42	1.39	1.43
HG2	1.34	1.33	1.29	1.37
HD1	1.63	1.66	1.60	1.61
HD2	1.60	1.62	1.56	1.58
HE1	2.95	2.94	2.94	2.95
HE2	2.89	2.88	2.86	2.93
HZ	7.51	7.62	7.74	7.49

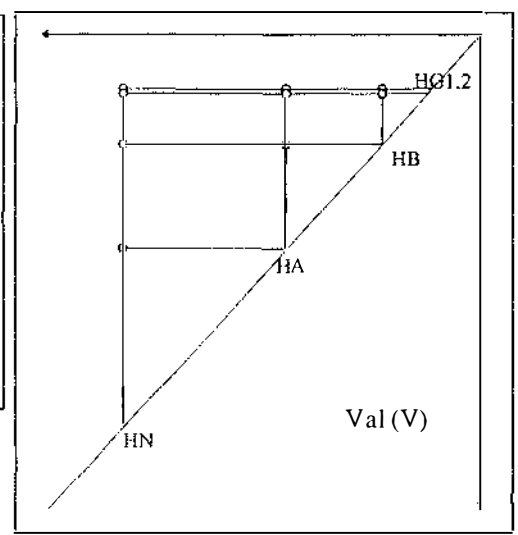


Valine

$$\begin{array}{l} {}_{\gamma}C^{\gamma 1}H_3 \\ | \\ -C^{\beta}H-{}_{\gamma}C^{\gamma 2}H_3 \end{array}$$

Average chemical shifts:

Nuclear	All	Helical	β-strand	Coil
HN	8.26	7.96	8.54	8.08
HA	4.13	3.58	4.56	4.11
HB	2.00	2.07	1.92	2.06
HG1	0.91	0.96	0.86	0.95
HG2	0.72	0.72	0.69	0.81



Threonine

$$\begin{array}{l} O^{\gamma 1} \\ | \\ -C^{\beta}H-{}_{\gamma}C^{\gamma 2}H_3 \end{array}$$

Average chemical shifts:

Nuclear	All	Helical	β-strand	Coil
HN	8.27	8.10	8.44	8.17
HA	4.50	4.02	4.85	4.37
HB	4.15	4.21	4.07	4.23
HG	1.17	1.18	1.15	1.20

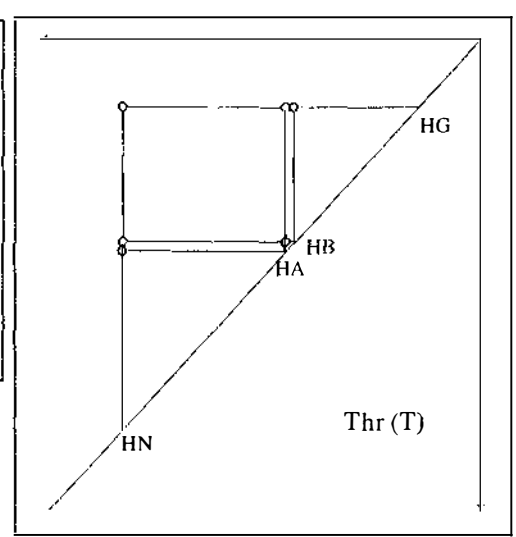
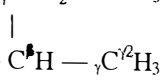


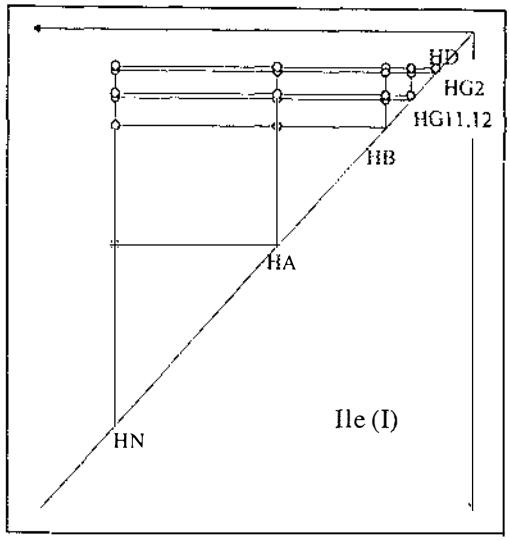
Figure 7b.1: The average chemical shifts and the spin-system connectivity of the 20 common amino acid residues. (Continued)

Isoleucine

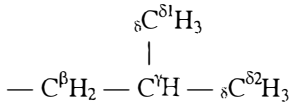


Average chemical shifts:

Nuclear	All	Helical	β-strand	Coil
HN	8.27	8.06	8.58	7.99
HA	4.14	3.67	4.61	4.09
HB	1.79	1.86	1.76	1.70
HG11	1.22	1.27	1.26	1.21
HG12	1.16	1.18	1.10	1.11
HG2	0.77	0.77	0.77	0.75
HD	0.67	0.72	0.61	0.70

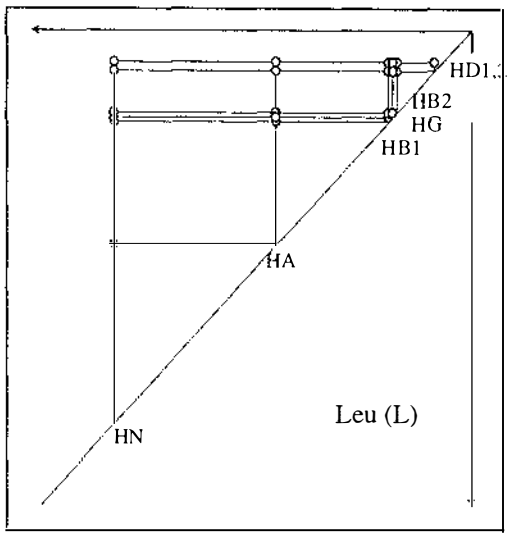


Leucine

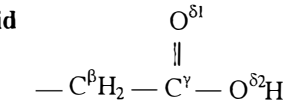


Average chemical shifts:

Nuclear	All	Helical	β-strand	Coil
HN	8.25	8.11	8.58	7.99
HA	4.29	3.98	4.69	4.35
HB1	1.75	1.80	1.70	1.73
HB2	1.50	1.52	1.45	1.57
HG	1.54	1.57	1.49	1.55
HD1	0.79	0.77	0.78	0.79
HD2	0.67	0.65	0.66	0.67



Aspartic acid



Average chemical shifts:

Nuclear	All	Helical	β-strand	Coil
HN	8.33	8.16	8.56	8.38
HA	4.62	4.45	5.07	4.63
HB1	2.88	2.83	2.89	2.91
HB2	2.62	2.63	2.53	2.65

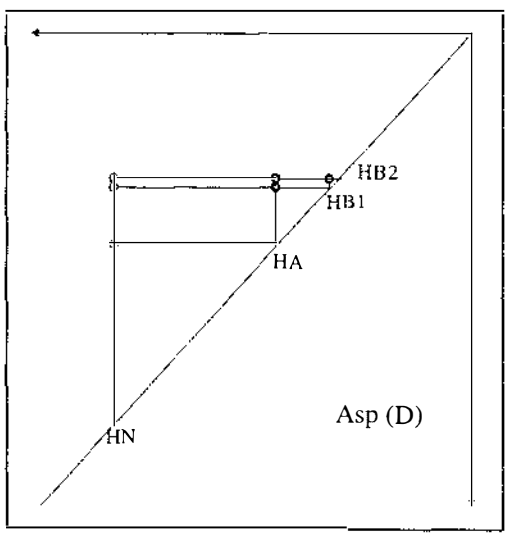
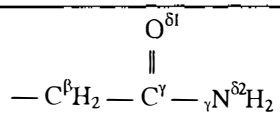


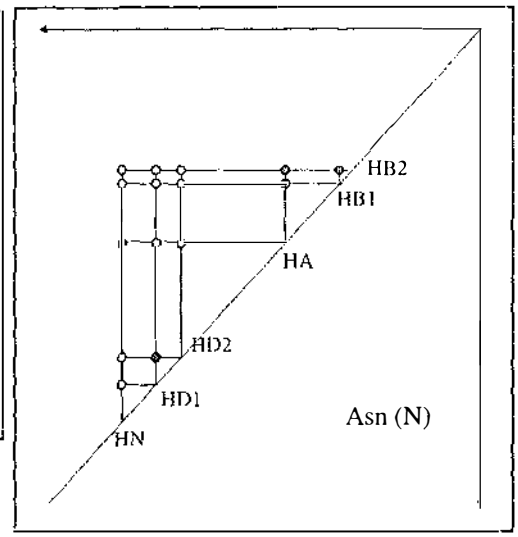
Figure 7b.1: The average chemical shifts and the spin-system connectivity of the 20 common amino acid residues. (Continued)

Asparagine

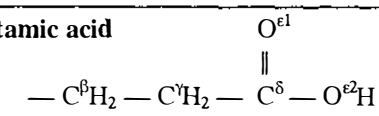


Average chemical shifts:

Nuclear	All	Helical	β-strand	Coil
HN	8.38	8.20	8.62	8.33
HA	4.68	4.41	5.03	4.62
HB1	2.93	2.87	2.98	2.92
HB2	2.64	2.68	2.66	2.59
HD1	7.61	7.60	7.55	7.67
HD2	6.91	6.89	6.85	6.99

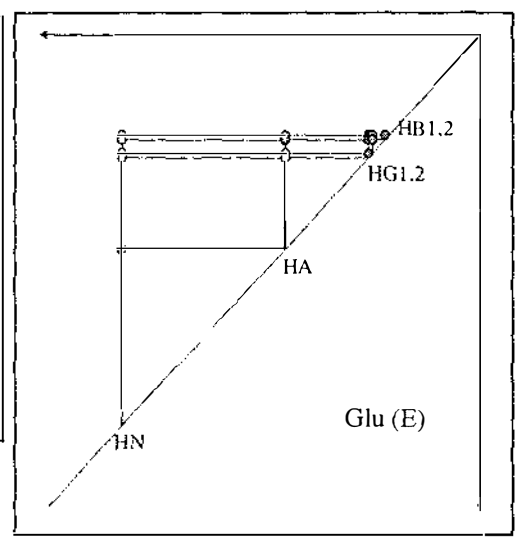


Glutamic acid

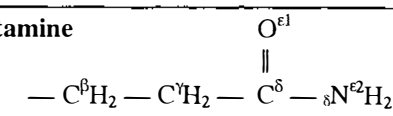


Average chemical shifts:

Nuclear	All	Helical	β-strand	Coil
HN	8.30	8.20	8.49	8.40
HA	4.22	4.00	4.85	4.24
HB1	2.11	2.15	2.03	2.08
HB2	2.00	2.03	1.93	1.97
HG1	2.35	2.39	2.26	2.35
HG2	2.28	2.29	2.19	2.29



Glutamine



Average chemical shifts:

Nuclear	All	Helical	β-strand	Coil
HN	8.23	8.11	8.51	8.19
HA	4.31	4.02	4.88	4.28
HB1	2.14	2.18	2.07	2.12
HB2	1.97	2.03	1.88	1.95
HG1	2.37	2.44	2.25	2.35
HG2	2.25	2.33	2.10	2.25
HE1	7.40	7.27	7.43	7.60
HE2	6.71	6.44	6.77	6.79

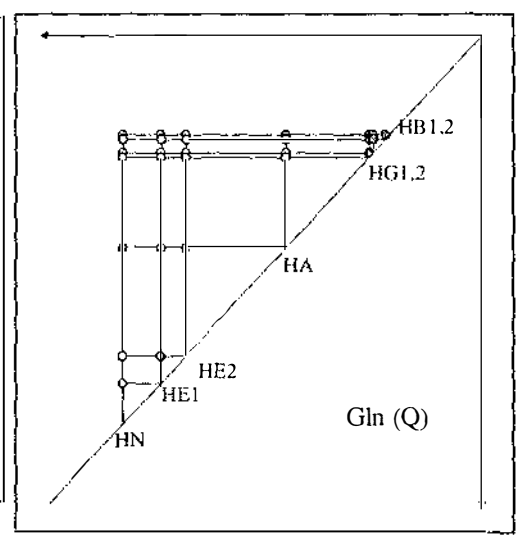
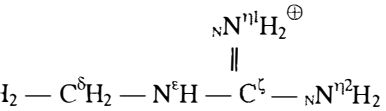


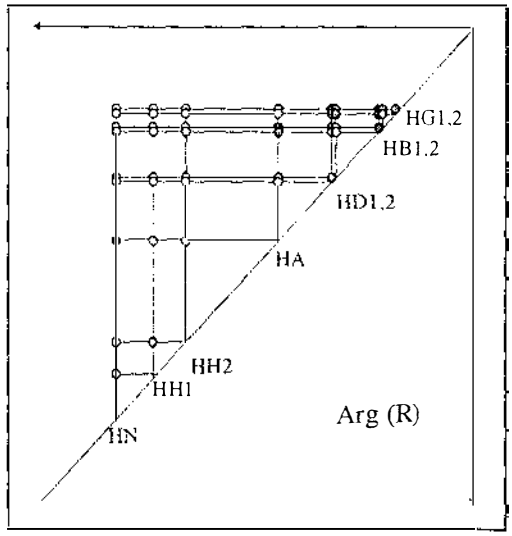
Figure 7b.1: The average chemical shifts and the spin-system connectivity of the 20 common amino acid residues. (Continued)

Arginine

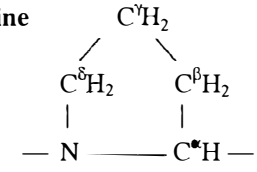


Average chemical shifts:

Nuclear	All	Helical	β-strand	Coil
HN	8.22	8.09	8.50	8.23
HA	4.24	3.99	4.65	4.32
HB1	1.86	1.90	1.78	1.86
HB2	1.71	1.79	1.58	1.71
HG1	1.62	1.65	1.60	1.61
HG2	1.52	1.51	1.51	1.53
HD1	3.12	3.09	3.14	3.14
HD2	3.08	3.07	3.07	3.10
HH1	7.30	7.27	7.49	7.30
HH2	6.62	6.62	6.62	6.62

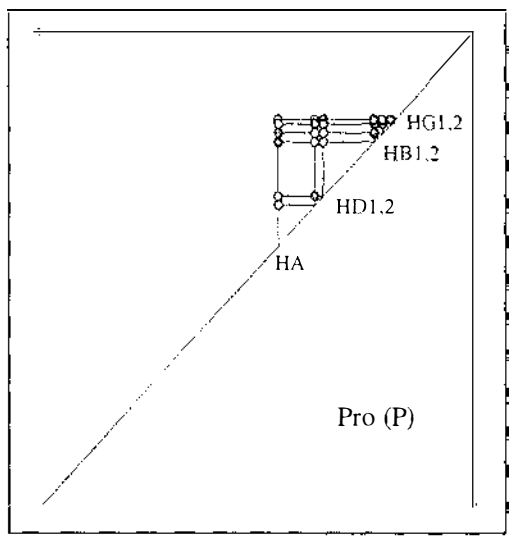


Proline

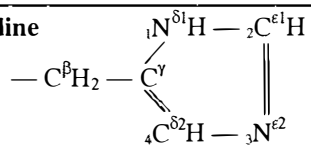


Average chemical shifts:

Nuclear	All	Helical	β-strand	Coil
HA	4.41	4.12	4.79	4.33
HB1	2.21	2.01	2.30	2.21
HB2	1.84	1.73	1.93	1.83
HG1	1.91	1.79	1.95	1.91
HG2	1.77	1.58	1.81	1.77
HD1	3.70	3.47	3.86	3.67
HD2	3.46	3.31	3.61	3.43



Histidine



Average chemical shifts:

Nuclear	All	Helical	β-strand	Coil
HN	8.26	7.90	8.57	8.36
HA	4.57	4.20	4.97	4.59
HB1	3.12	2.83	3.33	3.24
HB2	2.86	2.55	3.03	3.01
HD2	7.14	*	*	*
HE1	8.12	*	*	*

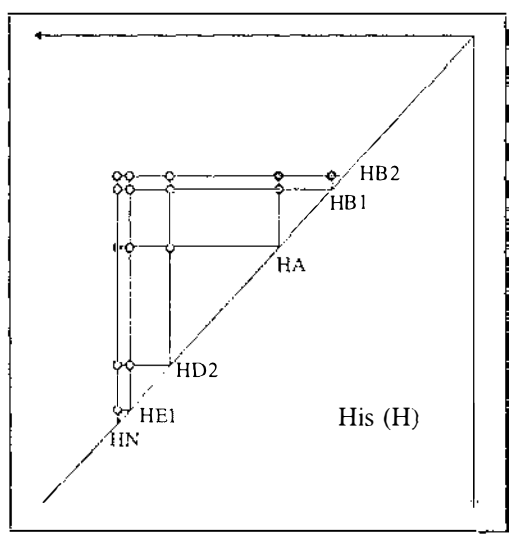
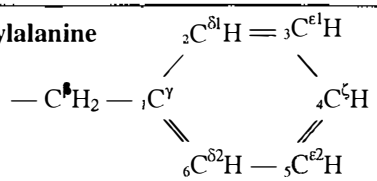


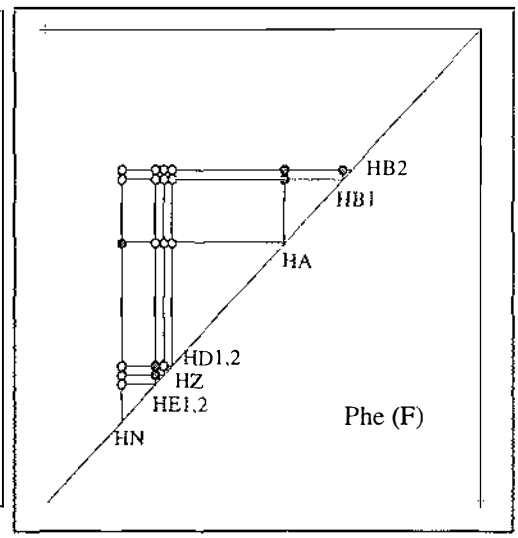
Figure 7b.1: The average chemical shifts and the spin-system connectivity of the 20 common amino acid residues. (Continued)

Phenylalanine

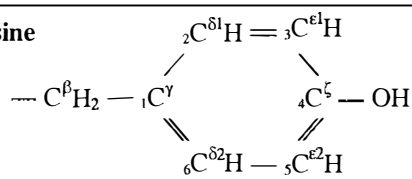


Average chemical shifts:

Nuclear	All	Helical	β-strand	Coil
HN	8.36	8.19	8.78	7.93
HA	4.58	4.16	5.08	4.42
HB1	3.10	3.22	3.01	3.09
HB2	2.81	2.94	2.71	2.79
HD1,2	7.30	*	*	*
HE1,2	7.39	*	*	*
HZ	7.34	*	*	*



Tyrosine



Average chemical shifts:

Nuclear	All	Helical	β-strand	Coil
HN	8.36	8.04	8.68	8.10
HA	4.62	4.15	5.03	4.43
HB1	3.08	3.12	3.04	3.11
HB2	2.77	2.82	2.72	2.79
HD1,2	7.15	*	*	*
HE1,2	6.86	*	*	*

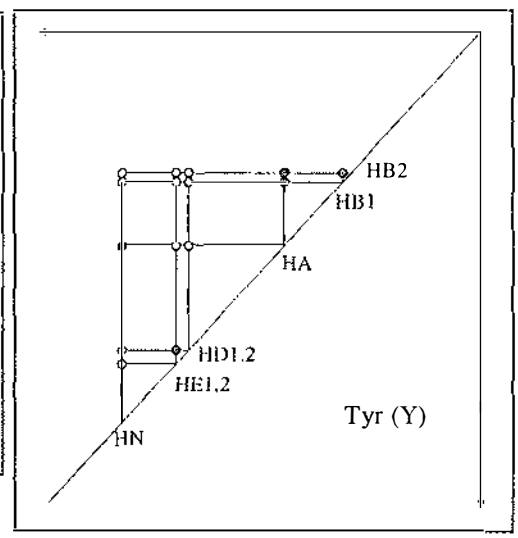


Figure 7b.1: The average chemical shifts and the spin-system connectivity of the 20 common amino acid residues. (Continued)

Tryptophan $_2C^{\delta1}H - {}_1N^{\varepsilon1}H$
$— C^{\beta}H_2 — {}_3C^{\gamma}$
$_9C^{\delta2} = {}_8C^{\varepsilon2}$
$_4C^{\varepsilon3}H \qquad {}_7C^{\zeta2}H$
$_5C^{\zeta3}H — {}_6C^{\eta2}H$

Average chemical shifts:

Nuclear	All	Helical	β-strand	Coil
HN	8.22	7.94	8.56	8.03
HA	4.63	4.29	5.03	4.42
HB1	3.41	3.41	3.30	3.57
HB2	3.11	3.16	3.02	3.19
HD1	7.24	*	*	*
HE1	10.22	*	*	*
HE3	7.65	*	*	*
HH2	7.24	*	*	*
HZ2	7.50	*	*	*
HZ3	7.17	*	*	*

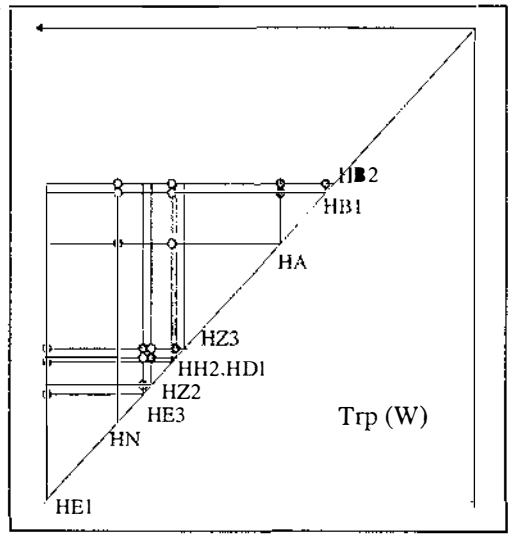


Figure 7b.1: The average chemical shifts and the spin-system connectivity of the 20 common amino acid residues. (Continued)

Appendix 7c:

Cross-peak Distribution in 2D NMR Spectra

The following summary of peaks in each of the 2D NMR spectral regions are drawn from Table 7.4 for the convenience of peak recognition in spectral assignment.

The summary is organised to explain the meaning of peaks from column A to G. Note that overlaps between neighbouring regions are often seen and will not be mentioned.

- Regions in column A contains only peaks that are related to HE1 of Trp residues W_i:HE1. Region A1 contains only diagonal peaks, namely, W_i:HE1-W_i:HE1. Region A2 to A7 contain cross-peaks that correlate W_i:HE1 and other ^{1}H spins. Except for those in Region A3, cross-peaks in these regions are usually weak. In COSY and TOCSY spectra, W_i:HE1-W_i:HD1 in Region A3 are the only cross-peaks that can be seen in these regions because W_i:HE1 and W_i:HD1 are the only members of the spin system that are related to W_i:HE1.

- Region B2 contains only X_i:HN-X_i:HN diagonal peaks in both COSY and TOCSY spectra. For NOESY spectra, this region also contains X_i:HN-X_j:HN inter-residue cross-peaks.

- Region B3 may contain weak X_i:HN-X_i:$H_{GrpC\text{-}others}$ intra-residue cross-peaks. In NOESY spectra, it may also contain X_i:HN-X_j:$H_{GrpC\text{-}rings}$ inter-residue cross-peaks. (Note that W_i:HE1 do not belong to $H_{GrpC\text{-}rings}$.)

- Region B4 contains all X_i:HN-X_i:H_{GrpD} intra-residue cross-peaks, including all X_i:HN-X_i:HA and X_i:HN-X_i:HB of Ser, Thr and His spin systems. The number of cross peaks should be about the same in this region for both COSY and TOCSY spectra. In NOESY spectra, the region also contains X_i:HN-X_j:H_{GrpD} inter-residue cross-peaks. (This is the fingerprint region.)

- Region B5 contains all X_i:HN-X_i:H_{GrpE} intra-residue cross-peaks in TOCSY spectra. It also contains X_i:HN-X_j:H_{GrpE} inter-residue cross-peaks in NOESY spectra. There is no COSY peak in this region. Similarly, regions B6 and B7 contain cross-peaks that correlate HN-H_{GrpF} and HN-H_{GrpG} of all spin systems.

- Region C3 contains strong X_i:$H_{GrpC\text{-}rings}$-X_i:$H_{GrpC\text{-}rings}$ diagonal and cross-peaks in both COSY and TOCSY spectra. For NOESY spectra, this region also contains X_i:$H_{GrpC\text{-}rings}$-X_j:$H_{GrpC\text{-}rings}$ inter-residue cross-peaks.

- Region C4 may contain X_i:H_{GrpC}-X_i:H_{GrpD} inter- and intra-residue cross-peaks. In some COSY spectra of very small proteins, this region can contain few or no peaks. Many COSY spectra, however, often contain a number of X_i:HN-X_i:HA peaks belonging to residues in helical configurations. Many TOCSY peaks and strong intra-residue X_i:H_{GrpC}-X_i:H_{GrpD} NOESY cross peaks can definitely be seen in this region.

- Region C5 contains all X_i:H_{GrpC}-X_i:H_{GrpE} intra-residue cross-peaks in TOCSY spectra. It also contains X_i:H_{GrpC}-X_j:H_{GrpE} inter-residue cross-peaks in NOESY spectra. There are no COSY peaks in this region. Similarly, Region C6 and C7 contains cross-peaks that correlate H_{GrpC}-H_{GrpF} and H_{GrpC}-H_{GrpG} of all spin systems.

- Region D4 contains strong X_i:H_{GrpD}-X_i:H_{GrpD} diagonal and cross-peaks in both COSY and TOCSY spectra. For NOESY spectra, this region also contains X_i:H_{GrpD}-X_j:H_{GrpD} inter-residue cross-peaks.

- Region D5 contains X_i:H_{GrpD}-X_i:H_{GrpE} intra-residue cross-peaks in COSY and TOCSY spectra. It also contains X_i:H_{GrpD}-X_j:H_{GrpE} inter-residue cross-peaks in NOESY spectra. Similarly, Region D6 and D7 contain cross-peaks that correlate H_{GrpD}-H_{GrpF} and H_{GrpD}-H_{GrpG} of all spin systems.

- Region E5 contains strong X_i:H_{GrpE}-X_i:H_{GrpE} diagonal and cross-peaks in both COSY and TOCSY spectra. For NOESY spectra, this region also contains X_i:H_{GrpE}-X_j:H_{GrpE} inter-residue cross-peaks.

- Region E6 contains X_i:H_{GrpE}-X_i:H_{GrpF} intra-residue cross-peaks of Lys and Arg residues in COSY and TOCSY spectra. It also contains X_i:H_{GrpE}-X_j:H_{GrpF} inter-residue cross-peaks in NOESY spectra. Similarly, region E7 contains cross-peaks that correlate H_{GrpE}-H_{GrpG}.

- Region F6 contains strong X_i:H_{GrpF}-X_i:H_{GrpF} diagonal and cross-peaks in both COSY and TOCSY spectra. For NOESY spectra, this region also contains X_i:H_{GrpF}-X_j:H_{GrpF} inter-residue cross-peaks.

- Region F7 contains X_i:H_{GrpF}-X_i:H_{GrpG} intra-residue cross-peaks of Lys, Val, Thr, Ile and Lue residues in COSY and TOCSY spectra. It also contains X_i:H_{GrpF}-X_j:H_{GrpG} inter-residue cross-peaks in NOESY spectra.

- Region G7 contains strong X_i:H_{GrpG}-X_i:H_{GrpG} diagonal and cross-peaks in both COSY and TOCSY spectra. For NOESY spectra, this region also contains X_i:H_{GrpG}-X_j:H_{GrpG} inter-residue cross-peaks.

References

Abragam, A. (1961). *The Principles of Nuclear Magnetism.* Clarendon, Oxford.

Aho, A. V. and J. D. Ullman (1992). *Foundations of Computer Science.* W. H. Freeman, New York.

Amicon Inc. (1998). Centriplus Concentrators Operating Manual, Amicon, Inc. Beverly, MA 01915, USA, or http://www.amicon.com

Anil-Kumar A., R. R. Ernst, and K. Wüthrich (1980). *Biochem. Biophys. Res. Commun.* **95**, 1.

Anil-Kumar A., G. Wagner, R. R. Ernst, and K. Wüthrich (1981). *J. Am. Chem. Soc.* **103**, 3654.

Anderson, T.W. (1977). *Annals of Statistics* **5**, 842.

Anderson, T.W., and R.P. Mentz (1980). *J. Time Series Analysis* **1**, 83.

Arakawa, T. and S. N. Timasheff (1985). *Methods in Enzymology* **114**, 49.

Arrowsmith, C., R. Pachter, R. Altman and O. Jardetzky (1991). *Eur. J. Biochem.* **202**, 53.

Aue W. P., E. Barthholdi, and R. R. Ernst (1976). *J. Chem. Phys.* **64**, 2229.

Bajpai, A. C., L. R. Mustoe and D. Walker (1977). *Advanced Engineering Mathematics.* John Wiley & Sons, New York.

Bartels, C. H., T-H. Xia, M. Billeter, P. Güntert and K. Wüthrich (1995). *J. Biomol. NMR* **5**, 1.

Bax, A. (1982). *Two-dimensional Nuclear Magnetic Resonance in Liquids.* Reidel, London.

Bax, A. and D. G. Davis (1985). *J. Magn. Reson.* **65**, 355.

Banwell, C. N. and H. Primas (1962). *J. Mol. Phys.* **6**, 225.

Berard, E. V. (1993). *Essays on Object-Oriented Software Engineering.* Englewood Cliffs, Prentice-Hall, New Jersey.

Biosym (1993). *Felix User Guide, Version 2.1.* Biosym Technologies, San Diego.

Biosym (1994). *Felix User Guide, Version 2.3.* Biosym Technologies, San Diego.

Biosym/Molecular Simulations (1995). *Felix 95.0 — NMR Data Processing, Analysis & Assignment User Guide*, Biosym/MSI, San Diego, USA.

Bloch, F., W. W. Hansen and M. Packard (1946). *Phys. Rev.* **70**, 474.

Bodenhausen G., R. Freeman, G. A. Morris, R. Niedermeyer and D. L. Turner (1977). *J. Magn. Reson.* **25**, 559.

Bodenhausen G., R. L. Vold and R. R. Vold (1980). *J. Magn. Reson.* **37**, 93.

Booch, G. (1991) *Object Oriented Design with Applications.* Benjamin Cummings, Redwood City, CA.

Borman, S. (1998). Advances in NMR of Macromolecules. *C&En August 10, 1998*, 55.

Bracewell, R. N. (1986). *The Fourier Transform and its Applications*, 2nd edition, McGraw-Hill, New York.

Braunschweiler, L. and R. R. Ernst (1983), *J. Magn. Reson.* **53**, 521.

Brigham E. O. (1974). *The Fast Fourier Transform.* Englewood Cliffs, Prentice-Hall, New Jersey.

Gaëta, B. A. (1997). *ANGIS Bioinformatics Handbooks* (four volumes). CSIRO Publishing, Australia.

Bundi, A and K. Wüthrich (1979a). *Biopolymers* **18**, 285.

Bundi, A. and K. Wuthrich (1979b). *Biopolymers* **18**, 299.

Burg, J.P. (1967). *Maximum Entropy Spectral Analysis*, Proceedings of the 37th Meeting of the Society of Exploration Geophysicists, Oklahoma City. Reprinted in Modern Spectral Analysis, Childers, D.G. (ed.) IEEE Press, New York.

Burg, J.P. (1968). *A New Analysis Technique for Time Series Data*, NATO Advanced Study Institute on Signal Processing, Reprinted in Modern Spectral Analysis, Childers, D.G. (ed.), IEEE Press, New York

Burg, J.P. (1972). Geophysics **37**, 375-376.

Campbell, I. D. and R. Freeman (1973). *J. Magn. Reson.* **11**, 143.

Cavanagh, J. and M. Rance (1993). *Annu. Rep. NMR Spectrosc.* **27**, 1.

Cavanagh, J., W. J. Fairbrother, A. G. Palmer III, and N. J. Skelton (1996). *Protein NMR Spectroscopy – Principles and Practice*, Academic Press, Toronto.

Chubb, R. T., V. Thanabel, C. Osborne and G. Wagner (1991). *Biochemistry* **30**, 7718.

Clayden, N. J. and R. J. P. Williams (1982). *J. Magn. Reson.* **49**, 383.

Cohen-Tannoudji, C., B. Diu and F. Laloë (1973). *Quantum Mechanics* (Translated from French by S. R. Hemly, N. Ostrowsky and D. Ostrowsky in 1977), John Willey & Sons, New York.

Czisch, M. and R. Boelens (1998). *J. Magn. Reson.* **134**, 158.

Dale, N. and S. Lily (1995) *Pascal Plus Data Structures, Algorithms, and Advanced Programming.* Lexington, DC Heath, MA.

Dalgarno, D. C., B. A. Levine, and R. J. P Williams (1983). *Biosci. Rep.* **3**, 443.

Davies, B. (1985). *Integral Functions and Their Applications.* 2nd edition, Springer-Verlag, New York.

Delaglio, F., S. Grzesiek, G. Vuister, G. Zhu, J. Pfeifer and A. Bax (1995). *J. Biomol. NMR* **5**, 277.

Davis, D. G. and A. Bax (1985). *J. Am. Chem. Soc.* **107**, 2820.

Delsuc, M. A. (1989). In *Maximum Entropy and Bayesian Methods.* Cambridge 1988. (Ed. by Skilling, J.), Kluwer, Dordrecht, The Nederlands, 285.

Delsuc, M. A. (1995). *The Gifa Program Version 4.0.* E-mail: mad@cbs.univ-montp1.fr. Web site: http://metal.foodsci.unibo.it/GIFA/basic_index.html

Derome, A. E. (1987). *Modern NMR Techniques for Chemistry Research*, Pergamon Press, Oxford.

Drobny G., A. Pines, S. Sinton, D. Weitekamp and D. Wemmer (1979), Faraday Div. Chem. Soc. Symp. **13**, 49.

Eccles C. D. (1995). Private communication on Hilbert transform.

Eccles C. D. (1987). *Microscopic NMR Imaging*. PhD Thesis, Massey University.

Eccles, C. D., P. Güntert, M. Billeter and K. Wüthrich (1991). *J. Biomol. NMR* **1**, 111-130.

Edison, A. S., W. M. Westler and J. L. Markley (1991). *J. Magn. Reson.* **9**, 434.

Ernst R., B. Bodenhausen and A. Wokaun (1987, 1994). *Principles of Nuclear Magnetic Resonance in One and Two Dimensions*, Clarendon Press, Oxford.

Ernst R. R. (1974). VIth International Conference on Magnetic Resonance in Biological Systems, Kandersteg, Switzerland.

Frenkiel, T.A. (1993). *Instrumentation and Pulse Sequences.* . Collected in [Roberts 1993], Oxford University Press, Oxford.

Fairley, R. E. (1985) *Software Engineering Concepts.* McGraw-Hill, New York.

Fuller, G. H. (1976). *J. Phys. Chem. Ref. Data.* **5**, No. 4.

Goldman S. (1953). *Information Theory*. Englewood Cliffs, Prentice-Hall, New Jersey.

Gradshteyn, I. S. and I. M. Ryzhik (1983). *Tables of Integrals, Series and Products.* Academic Press, New York.

Gronwald W, L. Willard, T. Jellard, R. F. Boyko, K. Rajarathnam, D. S. Wishart, F. D. Sonnichsen, B. D. Sykes (1998). *J Biomol. NMR* **12**, 395.

Greenberg, M. D. (1978). *Foundations of Applied Mathematics*. Prentice-Hall, New Jersey.

Gross, K. H. and H. R. Kalbitzer (1988). *J. Magn. Reson.* **76**, 87.

Güntert, P., V. Dötsch, G. Wider and K. Wüthrich (1992). *J. Biomol. NMR* **2**, 619.

Güntert, P. (1995). *DIANA Program Version 2.8 User's Manual and Instructions*, Institute of Molecular biology and Biology, Zürich, Switzerland.

Günther, H. (1995). *NMR Spectroscopy – Basic Prociples, Concepts, and Applications in Chemistry.* John Wiley & Sons, Toronto.

Haacke, E. M., Z. P. Liang and F. Boada (1990). *Optical Engineering.* **29**, 555.

Hare Research (1988). *FTNMR User Documentation (V.5.1)*, Hare Research Inc.

Hare Research (1990). *FELIX User Documentation (V.1.0)*, Hare Research Inc.

Hare Research (1990). *FELIX User Documentation (V.1.1)*, Hare Research Inc.

Hare Research (1991). *FELIX User Documentation (V.2.0)*, Hare Research Inc.

Heller, D. and P. M. Ferguson (1991). *Motif Programming Manual for OSF/Motif Release 1.2*. O'Reilly & Associates, Inc. USA.

Hess, C. P. and Z. P. Liang (1996). *Int. J. Imaging Syst. Techn.* **6**, 136.

Hoch, J. C. (1984). *J. Magn. Reson.* **64**, 436.

Hore, P. J. (1985). *J. Magn. Reson.* **62**, 561.

Hore, P. J. and G. J. Danille (1986). *J. Magn. Reson.* **69**, 386.

IUPAC-IUB Commission on Biochemical Nomenclature (1970). *Abbreviations and symbols for Description of the Conformation of Polypeptide Chains*, *J. Mol. Biol.* **52**, 1.

Jardetzky, O. and G. C. K. Roberts (1981). *NMR in Molecular Biology.* Academic Press, Inc.

Jaynes, E. T. (1957). *Phys. Rev.* **106**, 620.

Jeener, J. (1971). Ampere International Summer School, Basko Polje, Yugoslavia.

Jeener J., B. H. Meier, P. Bachmann and R. R. Ernst (1979). *J. Chem. Phys.* **71**, 4546.

Johnson, B. and R. Blevins (1994). *J. Biomol. NMR* **4**, 603.

Kay, L. E., P. Keifer and T. Saarinen (1992). *J. Am. Chem. Soc.* **114**, 10663.

Kessler, H., M. Gehrke and C. Griesinger (1988). *Two-dimensional NMR Spectroscopy: Background and Overview of the Experiments. Angew. Chem. Int. Ed. Engl.* **27**, 490.

Keywegt, G, R. Boelens, M. Cox, M. Llinas and R. Kaptein (1991). *J. Biomol. NMR* **1**, 23.

Kiczales, G. (1992). *Towards a New Model of Abstraction in the Engineering of Software.* IMSA'92 Proceedings -Workshop on Reflection and Meta-level Architectures.

KjÆR, M., K. V. Andersen, A. Ludvigson, H. Shen, D. Windekilde, B. Sorensen and F. M. Poulsen (1991). Collected in *Computational aspects of the study of biological macromolecules by NMR spectroscopy* (edited by J.C. Hoch, C. Redfield and F.M. Poulsen), Plenum, New York.

KjÆR, M., K. V. Andersen, and F. M. Poulsen (1994). *Methods in Enzymology,* Vol. **239**, 288.

Knight, W. D. (1956). Solid State Physics **2**, 93, 1956.

Kraulis, P. (1989). *J. Magn. Reson.* **84**, 627.

Kumar --- See Anil-Kumar.

Levy, F. N. G. C., and H. A. Scheraga (1986). *J. Magn. Reson.* **66**, 385.

Lie, W, P. Atkinson, H. Wong and C. Eccles (1997). *A new XWindow-based protein analysis software package.* ANZMAG 97, Queensland, Australia.

Lundberg, P. (1997) *Compilation of Educational NMR Software* (version 1.6.2beta), http://www.york.ac.uk/depts/chem/services/nmr/, University of York, U.K.

Macura S., and R. R. Ernst (1980). *Mol. Phys.* **41**, 95

Marion D. and K. Wüthrich (1983). Biochem. Biophys. Res. Comm. **113**, 967.

Martin, J. f (1985). *J. Magn. Reson.* **65**, 291.

Marquardt D. W. (1963). *J. Soc. Ind. Appl. Math.* **11**, 431.

Markley, J. L., D. H. Meadows, and O. Jardertzky (1967). *J. Mol. Biol.* **27**, 25.

Markley, J. L. and M. Kainosho (1993). *Stable Isotope Labelling and Resonance Assignments in Larger Proteins.* Collected in [Roberts 1993], Oxford University Press, Oxford.

Martenson, R. E., J. Y. Park, and Stone, A. L. (1985). *Biochemistry* **24**, 7689.

Martin, J. (1993). *Principles of Object-oriented Analysis and Design.* Prentice-Hall, New Jersey.

Montelione, G. T., M. E. Winkler, P. Rauenbeuhler, and G. Wagner (1989). *J. Magn. Reson.* **82**, 198.

Muller, L., A. Kumar, and R. R. Ernst (1975). *J. Chem. Phys.* **63**, 5490.

Noggle J. H. and R.E. Schirmer (1971). *The Nuclear Overhauser Effect, Chemical Applications.* Academic Press, New York.

Newman, R. H. (1988). *J. Magn. Reson.* **79**, 448.

Nussbaumer H. J. (1982). *Fast Fourier Transform and Convolution Algorithms.* Springer-Verlag, New York.

Oh, B. H., and J. L. Markley (1990a). *Biochemistry* **29**, 3993.

Oh, B. H., E. S. Mooberry and J. L. Markley (1990b). *Biochemistry* **29**, 4004.

Oh, B. H., and J. L. Markley (1990c). *Biochemistry* **29**, 4012.

Pardi, A., G. Wagner and K. Wüthrich (1983). *Eur. J. Biochem.* **137**, 445

Pastore, A. and V. Saudek (1990). *J. Magn. Reson.* **90**, 165.

Perkin, S. J. and K. Wüthrich (1979). *Biochem. Biophys. Acta* **576**, 409

Pervushin, K., R. Riek, G. Wider and K. Wüthrich (1997). *Proc. Natl. Acad. Sci.* **94**, 12366.

Pervushin, K., R. Riek, G. Wider and K. Wüthrich (1998). *J. Am. Soc.* **120**, 6394.

Pons, J-L, T.E. Mallianvin and M.A. Delsuc (1996). *Jour. Biomol. NMR.* **8**, 445.

Pople, J. A., W. G. Schneider and H. J. Bernstein (1959). *High Resolution NMR.* McGraw-Hill, New York.

Press W. H., B. P. Flannery, S. A. Teukolsky and W. T. Vetterling (1986). *Numerical Recipes.* Cambridge University Press, Cambridge.

Pressman, R. S. (1992). *Software Engineering: A Practitioner's Approach.* McGraw-Hill, Inc, New York.

Primrose, W. U. (1993). *Sample Preparation.* Collected in Roberts 1993, Oxford University Press, Oxford.

Purcell, E. M., H. C. Torrey and R.V. Pound (1946). ibid. **69**, 37.

Rance, M. (1987). *J. Magn. Reson.* **74**, 557.

Rance, M. et al. (1990). *J. Magn. Reson.* **87**, 336. (This is to be completed.)

Redfield, A. G. (1965). Adv. Mag. Reson. **1**, 1.

Redfield A. G., S. D. Kunz (1975). *J. Magn. Reson.* **19**, 250.

Redfield, C. and C. Dobson (1988). *Biochemistry* **27**, 122.

Redfield, C. (1993). *Resonance Assignment Strategies for Small Proteins.* Collected in Roberts 1993, Oxford University Press, Oxford.

Reutimann, H., B. Straub, P. L. Luisi, and A. Holmgren (1981). *J. Biol. Chem.* **256**, 6796.

Richardson, J. (1981). *Adv. Protein Chem.* **34**, 167.

Roberts G. C. K. *et al* (1993). *NMR of Macromolecules: A Practical Approach.* Oxford University Press, Oxford.

Royce, W. W. (1970). *Managing the Development of Large Software System: Concepts and Techniques. Proc. IEEE WESTCON*, 1, Los Angeles.

Rucker, S. P. and A. J. Shaka (1989). *Mol. Phys.* **68**, 509.

Schulz, G.E. and R.H. Schirmer (1979). *Principles of Protein Structure*, Springer, Berlin.

Smallcombe, S, S. Patt, J Wurl, and J Provost (1996). (Varian) Magnetic Moments **8**, 1.

Sommerville, I. (1995). *Software Engineering* (5th edition). Addison-Wesley Publishing Company, New York.

Sonnichsen, F. D., J. E. Van Eyk, R. S. Hodges and B. D. Sykes (1992). *Biochemistry* **31**, 8790.

Sørensen, O. W., G. W. Eich, M. H. Levitt, G. Bodenhausen and R. R. Ernst (1983). Prog. NMR Spectrosc. **16**, 163.

Stakgold, I. (1979). *Green's Functions and Boundary Value Problems.* John Wiley & Sons, New York.

Sternlicht, H. and D. Wilson (1967). *Biochemistry* **6**, 2881.

Stone, A. L., J. Y. Park, and R. E. Martenson (1985). *Biochemistry* **24**, 6666.

Strobach, J. D. (1998). *Linear Prediction Theory: A Mathematical Basis for Adaptive Systems. Springer Series in Information Science* **21**. Springer-Verlag, New York.

Szilagyi, L. and O. Jardertzky (1989). *J. Magn. Reson.* 83, 441.

Tigelaar, H. L. and W. H. Flygare (1972). J. *Amer. Chem. Soc.* **94**, 343.

Uhrinova, S., D. Uhrín, H. Denton, M. Smith, L. Sawyer, and P. N. Barlow (1998). *J. Biomol. NMR* **12**, 89.

Varian NMR Instruments (1989-92). *System Operation Manual for NMR Spectrometer Systems (with VNMR ver. 4.1 software installed)*, Varian NMR Instruments, California.

Varian NMR Instruments (1989-92b). *Command and Parameter Reference for VNMR ver. 4.1 software*, Varian NMR Instruments, California.

Weast, R. C., G. L. Tuve and many collaborators and contributors (1964-1997). *Table of nuclear spins*, Handbook of Chemistry and Physics, CRC Press, Ohio.

Weigelt, J. (1998). *Single scan, sensitivity- and gradient-enhanced TROSY for multidimensional NMR experiments. J. Am. Chem. Soc.* **120**, 10778

Wilmad Glass (1998). *Proper Cleaning Procedures for NMR Sample Tubes*, http://www.wilmad.com/html/nf/Welcome.html.

Wishart, D. S., B. D. Sykes and F. M. Richards (1991). *J. Mol. Biol.* **222**, 311.

Wishart DS, Bigam CG, Yao J, Abildgaard F, Dyson HJ, Oldfield E, Markley JL, Sykes BD (1995). *J Biomol NMR* **6**, 135.

Wishart D. S. R. F. Boyko, L. Willard, F. M. Richards, B. D. Sykes (1994). *Comput. Appl. Biosci.* **10**, 121.

Wishart DS, Nip AM (1998). *Biochem Cell Biol.* **76**,153.

Wong H. (1990). *Density Operator and Super-operator Methods in Multiple Pulse NMR*. DSIR, Lower Hutt, New Zealand.

Wong H. (1996). *The determination of the Helical Secondary Structure of a Caerin Peptide by High Field NMR Spectroscopy*. Industrial Research Limited, Lower Hutt, New Zealand.

Wong H., J. H. Bowie, J. A. Carver (1997). *Eur J Biochem.* **247**, 545.

Wüthrich, K. (1976). *NMR in Biological Researches: Peptides and Proteins.* North Holland, Amsterdam.

Wüthrich, K. and G. Wagner (1979). *J. Mol. Biol.* **130**, 1.

Wüthrich, K. (1986). *NMR of Proteins and Nucleic Acids*, John Wiley & Son, New York.

Xia, T. and C. Bartels (1994). *XEASY User Manual.* Institute of Molecular Biology and Biophysics, ETH-Hönggerberg, CH-8093 Zurich, Switzerland.

Zwillinger, D. (1989). *Handbook of Differential Equations.* Academic Press, New York.