

Copyright is owned by the Author of the thesis. Permission is given for a copy to be downloaded by an individual for the purpose of research and private study only. The thesis may not be reproduced elsewhere without the permission of the Author.

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

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 ^1H 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

Chapter 1: Introduction	1
1.1 Project Justification	2
1.1.1 NMR spectroscopy and proteins	2
1.1.2 NMR software	2
1.1.3 Project outline	3
1.1.4 Software development process for the project	3
1.1.5 An overview of PAW	4
1.2 Thesis Organisation	4
1.2.1 Volume One	4
1.2.2 Volume Two	5
1.3 Abbreviation and Notation	6
1.3.1 Abbreviations	6
1.3.2 Notation	7
1.4 Conventions	9
1.4.1 Conventions for formatted text	9
1.4.2 Conventions for mouse operations	9
1.4.3 Conventions for keyboard entries	10
Chapter 2: Basic Theory of Nuclear Magnetic Resonance Spectroscopy	11
2.1 Introduction	12
2.2 Nuclear Magnetism	12
2.2.1 Nuclear magnetic moment and Larmor frequency	12
2.2.2 Nuclear magnetic equilibrium and quantisation	12
2.2.3 Allowable energy states	13
2.2.4 The Boltzmann distribution law	14
2.2.5 Factors related to M_0	14
2.3 Nuclear Magnetic Resonance and Free Induction Decay	15
2.3.1 Nuclear magnetic resonance and the rf-pulse	15
2.3.2 Nuclear magnetic relaxation	16
2.3.3 The Bloch equations in the laboratory frame	17
2.3.4 Transverse magnetisation vector and free induction decay	18
2.3.5 An ideal FID and its Fourier transform	19
2.3.6 Multi-frequency FID	20
2.3.7 Chemical shifts	20
2.3.8 Nuclear spin systems and spin-spin coupling	22
2.4 The Schrödinger Equation in NMR	22
2.4.1 The Schrödinger equation	22
2.4.2 Magnetic quantum levels and states	22
2.4.3 Basis state space for a spin system	23
2.5 The Spin Hamiltonian	24
2.5.1 General spin Hamiltonian	24
2.5.2 Spin Hamiltonian for liquid samples	25
2.5.3 Properties of the spin Hamiltonian	25
2.6 Density Operators in NMR	25
2.6.1 Spin density operators	25
2.7 Observable Operators	26
2.7.1 Observable operators and their expectation values	26
2.7.2 Spin operator and finite rotation operator	26
2.7.3 Raising and lowering operators	27
2.8 Density Operator Equations in NMR	28
2.8.1 Density operator equation for unperturbed systems	28
2.8.2 Hamiltonian super operator	29
2.9 Indirect Coupling of a Two-Spin-$\frac{1}{2}$ System	30

2.9.1 Spin Hamiltonian of a two-spin- $\frac{1}{2}$ system.....	30
2.9.2 Eigen-values and eigen-vectors of a two-spin- $\frac{1}{2}$ system	31
2.9.3 Magnetic quantum numbers of a two-spin- $\frac{1}{2}$ system	31
2.9.4 Calculated spectrum of a two-spin- $\frac{1}{2}$ system	31
Chapter 3: Product Operator Formalism and Pulse Analysis	35
3.1 Introduction	36
3.2 Product Operator Formalism.....	36
3.2.1 Liouville operator Space	36
3.2.2 Evolution of spin systems.....	37
3.2.3 The basis of rotational transformations	38
3.2.4 Generalised cyclic notation	39
3.2.5 Evolution under Zeeman and chemical shift Hamiltonian.....	40
3.2.6 Evolution under pulses	41
3.2.7 Evolution under J -couplings	41
3.2.8 More on coherence transfer.....	42
3.3 NMR Experimental Pulse Analysis	42
3.3.1 The spin echo experiment	42
3.3.2 The COSY experiment.....	44
3.3.3 Double-quantum filtered (DQF) COSY	46
3.3.4 The NOESY experiments.....	47
3.3.5 The TOCSY and HOHAHA experiments.....	49
Chapter 4: Protein NMR Experiments	51
4.1 Introduction	52
4.2 Sample Selection	52
4.2.1 Feasibility.....	52
4.2.2 Availability.....	53
4.2.3 Solubility.....	53
4.2.4 Stability.....	53
4.3 Sample Preparation	54
4.3.1 Cleaning the NMR tubes	54
4.3.2 Filtration and dialysis	54
4.3.3 Concentrating NMR samples	54
4.3.4 pH adjustment and final purification.....	55
4.4 NMR Experiments	56
4.4.1 NMR spectrometer system	56
4.4.2 Setting up 1D protein NMR experiments.....	57
4.4.3 Setting up 2D protein NMR experiments.....	59
4.4.4 Temperature selection	60
4.4.5 Mixing-time selection.....	60
Chapter 5: NMR Software and PAW.....	63
5.1 Introduction	64
5.2 A Review of NMR Software.....	64
5.2.1 Existing NMR data processing and analysis software.....	64
5.3 An Overview of PAW	66
5.3.1 Functionality	66
5.3.2 Development model	67
5.3.3 Project constraints and objectives	68
5.3.4 Development Strategy	69
5.3.5 System and software design	70

5.3.6 User-interface design and implementation	80
5.3.7 Module design and implementation.....	83
 Chapter 6: NMR Data Processing with PAW.....	 89
6.1 Introduction.....	90
6.2 Interactive Loops	90
6.3 One-dimensional NMR Data Processing.....	91
6.3.1 Convolution character of NMR data.....	91
6.3.2 General problems	95
6.3.3 General procedure	98
6.3.4 General interfaces for one-dimensional data processing	99
6.3.5 Zero-filling	101
6.3.6 Apodisation	102
6.3.7 Linear prediction	109
6.3.8 Water signal suppression.....	112
6.3.9 Fast Fourier transform (FFT) and its inverse (IFT)	117
6.3.10 Real Fourier transform (RFT).....	118
6.3.11 Hilbert transforms (HT).....	119
6.3.12 Phase correction	124
6.3.13 Baseline correction	129
6.4 Two-dimensional NMR Data Processing	137
6.4.1 2D NMR data sets	137
6.4.2 General procedure	138
 Chapter 7: NMR Spectral Assignment of Proteins Using PAW.....	 141
7.1 Introduction.....	142
7.2 Protein Structure and Proton Codes.....	142
7.2.1 Protein primary structure	142
7.2.2 Side chains of the 20 common amino acids.....	143
7.2.3 Proton codes used in PAW	145
7.3 Spin Systems in Protein NMR.....	146
7.3.1 ^1H spin systems in amino-acid residues	146
7.3.2 PAW's spin-system strings.....	148
7.3.3 ^1H chemical shifts in amino acid residues.....	149
7.3.4 Chemical shift distribution plot	149
7.3.5 NOESY connectivity diagrams of protein spin systems	150
7.3.6 Chemical shift-groups of ^1H spins in amino acids	152
7.3.7 Notation for spins in different chemical shift groups.....	153
7.3.8 Notation for ^1H spins in different residues.....	153
7.4 Two-dimensional ^1H NMR Spectra of Proteins.....	153
7.4.1 Notation for peaks in 2D ^1H NMR spectra	153
7.4.2 Distribution of peaks in 2D ^1H NMR spectra of proteins	153
7.5 Two-dimensional NMR Spectral Assignment of Proteins	156
7.5.1 General procedure of 2D NMR spectral assignment	157
7.5.2 Selection of spectra for assignment	158
7.5.3 Peak picking and refinement	158
7.5.4 Group assignment.....	163
7.5.5 Spin system identification and assignment	164
7.5.6 Inter-residue connectivity identification and assignment.....	166
 Chapter 8: NMR Spectral Processing of Caerin 4.1 Using PAW	 169
8.1 Introduction.....	170

8.2 Processing the Caerin 4.1 DQF-COSY NMR data	170
8.2.1 The raw data.....	170
8.2.2 High-resolution processing operations and results	174
8.2.3 Conventional processing operations and results.....	188
8.3 Processing the Caerin 4.1 TOCSY070 NMR data set	194
8.3.1 The raw data.....	194
8.3.2 High-resolution processing operations and results	195
8.4 Processing the Caerin 4.1 NOESY150 NMR data set	205
8.4.1 The raw data.....	205
8.4.2 High-resolution processing operations and results	206

Chapter 9: Peak-picking Process for Caerin 4.1 NMR Spectra Using PAW..... 217

9.1 Introduction	218
9.2 Raw-peak Picking	218
9.2.1 Raw-peak picking in the upper-left region	220
9.2.2 Raw-peak picking in the lower-right region	221
9.2.3 Raw-peak picking in the upper-right region	222
9.2.4 Raw-peak picking in the lower-left region	223
9.3 Diagonal-peak Picking	224
9.3.1 Diagonal-peak picking in the lower-left region	224
9.3.2 Diagonal-peak picking in the upper-right region.....	225
9.3.3 Diagonal-peak refinement in the upper-right region	226
9.4 Cross-peak Picking	227
9.4.1 Cross-peak picking in the lower-left region	227
9.4.2 Cross-peak picking in the other regions	229
9.4.3 Assessing missing cross-peaks	232
9.4.4 Adding missing cross-peaks	234

Chapter 10: Spin-system Identification for Caerin 4.1 NMR Spectra Using PAW.... 237

10.1 Introduction	238
10.2 Identification of the Ala residues	239
10.3 Identification of the Ser residues	244
10.4 Identification of the Lys, Asp and Trp residues	246
10.5 Identification of the Gly residues	254
10.6 Identification of the Val, Glu and Gln residues	256
10.7 Identification of the Leu and Ile residues	261

Chapter 11: NMR Spectral Assignment for Caerin 4.1 Using PAW..... 265

11.1 Introduction	266
11.2 Proton-code Assignment in Groups	266
11.2.1 Proton-code assignment in the lower-left region.....	267
11.2.2 Proton-code assignment in the upper-left region	271
11.2.3 Proton-code assignment in the lower-right region.....	273
11.2.4 Proton-code assignment in the upper-right region.....	274
11.3 Adding and Editing Cross-peak Records	275
11.3.1 Cross-peak analysis using multiple spectrum display.....	275
11.3.2 Editing cross-peak records	277
11.3.3 Editing transposed cross-peak records	279
11.3.4 Editing cross-peak positions.....	279
11.3.5 Finding precise cross-peak positions	280
11.3.6 Lining up cross-peaks.....	281
11.3.7 Moving cross-peak labels.....	283

11.4 Sequence-specific Assignment	285
11.4.1 The symbols for the assigned cross-peaks	286
11.4.2 Assignment of the unambiguous residues	286
11.4.3 Identification of the adjacent residues	288
11.4.4 Sequential assignment using various peak-display options	292
11.5 Exporting Spectral Assignment Results	302
11.5.1 The PostScript plots of the final assignment results	302
11.5.2 The assigned cross-peak list	309
Chapter 12: Summary and Future Development Plan	327
12.1 Summary 12.1	328
11.2 Future Development 12.2	329
Appendixes	331
Appendix 4a: Dialysis of Hen-egg Lysozyme	332
Appendix 6a: Cubic spline fitting	333
Appendix 6b: The Discrete Fourier Transform	335
Appendix 6c: Gaussian Peaks in NMR Spectra	338
Appendix 6d: Lorentzian Peaks in NMR Spectra	341
Appendix 6e: Sine-bell Filtered Peaks in NMR Spectra	343
Appendix 7a: Protein ¹H Chemical-shifts (Table 2)	348
Appendix 7b: TOCSY Connectivity Diagrams of ¹H Spin Systems in Proteins	349
Appendix 7c: Cross-peak Distribution in 2D NMR Spectra	357
References	359

List of Tables

Table 7.1. Primary structures of the 20 common amino-acid residues in proteins	144
Table 7.2. Spin-systems in 20 common amino acids and PAW's spin-system strings	148
Table 7.3. Distribution of ¹ H chemical shifts of the 20 common amino acids	150
Table 7.4. Grouping of protons in amino acids in terms of chemical shifts	152
Table 7.5. ¹ H- ¹ H distances in proteins	167