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 Philsophy 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