Multidimensional, Heteronuclear NMR Spectroscopy of High Molecular Weight Proteins in Solution: Application to a Bacterial Dimethylarginine Dimethylaminohydrolase

Michael John Plevin
University College London
For the Degree of Doctor of Philosophy
February 2003
To my Family

and Cristina
Abstract

In higher mammals, the enzyme dimethylarginine dimethylaminohydrolase (DDAH) is responsible for regulating the levels of the asymmetric methylarginine family of nitric oxide synthase (NOS) inhibitors. In this thesis a bacterial DDAH from *Pseudomonas aeruginosa* (PaDDAH) is characterised by multidimensional heteronuclear NMR spectroscopy.

A model system was developed to mimic heteronuclear NMR investigations of larger proteins. $^{15}\text{N}$ $R_1$, $^{15}\text{N}$ $R_2$ and heteronuclear NOE experiments were conducted on a sample of isotope-labelled perdeuterated ubiquitin in 50% (v/v) glycerol at five different temperatures ($17.5^\circ\text{C} < T < 34.3^\circ\text{C}$). Isotropic rotational correlation times ($\tau_c$) were estimated at each temperature. Between $17.5^\circ\text{C}$ and $34.3^\circ\text{C}$ a 21.2 ns range of $\tau_c$ was observed. TROSY-based 3D HNCA and 3D HN(CA)CB spectra were recorded of [$^2\text{H}$, $^{13}\text{C}$, $^{15}\text{N}$]-labelled ubiquitin in 50% (v/v) glycerol at $25^\circ\text{C}$ ($\tau_c = 34.6$ ns). Sequence specific backbone $^1\text{H}^N$, $^{15}\text{N}$, $^{13}\text{C}_\alpha$ and $^{13}\text{C}_\beta$ resonance assignment of ubiquitin was performed under these conditions.

2D [$^1\text{H}$, $^{15}\text{N}$]-TROSY spectra of [$^{15}\text{N}$]-labelled WT PaDDAH showed the protein to be folded with superior resolution and signal-to-noise obtained on perdeuteration. The
hydrodynamic properties of WT PaDDAH were analysed by solution NMR spectroscopy. Estimates of the translational diffusion coefficient $D_z$ ($7.6 \times 10^{-7}$ cm s$^{-1}$) and $\tau_c$ (35 ns) indicated the enzyme to be a homodimer in solution. A series of TROSY-based 3D triple resonance spectra were recorded of [$^2$H, $^{13}$C, $^{15}$N]-labelled WT PaDDAH. A full complement of $^1$H, $^{15}$N cross peaks and $^{13}$C correlations was not observed in these spectra. Therefore, unambiguous resonance assignments were not obtained.

The self-association properties of WT PaDDAH were investigated using analytical size exclusion chromatography (SEC). The results suggested WT PaDDAH exists in a dynamic equilibrium between monomer and homodimer species in solution. A dissociation constant $K_d$ for this equilibrium was estimated ($50 \text{nM} < K_d < 250 \text{nM}$). Residues contributing to the homodimer interface were identified from the previously reported crystal structure of PaDDAH. A series of point mutants at the protomer-interface were produced and each assayed for solubility, enzymatic activity and hydrodynamic properties. Each soluble PaDDAH mutant retained enzymatic activity but exhibited modified solution properties in comparison to the wild-type enzyme. A subset of PaDDAH mutants were further analysed by 1D $^1$H NMR and 2D heteronuclear NMR methods. 2D NH correlation spectra of N36W, R40E and R98H PaDDAH were highly comparable to WT PaDDAH. Estimations of $\tau_c$ suggested N36W PaDDAH is a strongly self-associated homodimer variant ($\tau_c = 31$ ns) and R40E PaDDAH tends towards a monomeric species ($\tau_c = 21.9$ ns). The lower $\tau_c$ for R40E PaDDAH suggests this protein may be more amenable for future attempts at sequence-specific backbone resonance assignment.

The investigations and results presented in this thesis are discussed with reference to recent advances in heteronuclear NMR studies of higher molecular weight proteins.
Acknowledgements

I would like to take this opportunity to thank all of my colleagues in the UCL/LICR NMR laboratory. Every member of the group has helped me at some stage along the way with insightful answers to my many questions. There are several people who deserve individual acknowledgement: Dr. Richard Harris for his patient and invaluable assistance in running NMR experiments and analysing data; Andrew Sankar for making many of the protein samples in the later periods of my project; and Drs Mark Williams and Diego Esposito for their many helpful discussions on analysing experimental data. I would also like to mention Dr Andrew Dingley for critical reading of some sections of this manuscript and Dr. Snezana Djordjevic, my academic mentor, for support throughout my time at UCL. In addition I would like to wish Beatriz good fortunes in her future endeavours with PaDDAH.

I would like to express my gratitude to my peers in the UCL BSM group who have all eased my way through this eventful journey with their smiling friendly, faces, words of encouragement and penchant for Friday evening relaxations. In addition, I would like to dearly thank my friends outside the department for helping me to keep things in perspective, preventing me from losing touch with reality and being ever polite when asking what I do. In particular I think Cristina Viola deserves an individual mention, for she has borne the brunt of this over the last two years. I would also like to thank my family for their support, encouragement and interest from the beginning.

I would like to acknowledge the Medical Research Council, the Royal Society and the Bogue Foundation for providing financial support during this project.

Finally, I would like to wholeheartedly thank my supervisor and mentor Dr. Paul Driscoll for his unwavering support, enthusiasm and guidance throughout this project, especially during the final months.
# Table of Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract</td>
<td>3</td>
</tr>
<tr>
<td>Acknowledgements</td>
<td>5</td>
</tr>
<tr>
<td>Table of Contents</td>
<td>6</td>
</tr>
<tr>
<td>List of Figures</td>
<td>12</td>
</tr>
<tr>
<td>List of Tables</td>
<td>16</td>
</tr>
<tr>
<td>Abbreviations</td>
<td>18</td>
</tr>
<tr>
<td>Chapter I</td>
<td>21</td>
</tr>
<tr>
<td>1. Introduction.</td>
<td>21</td>
</tr>
<tr>
<td>1.1. Problems Associated with Studying Large Proteins by NMR.</td>
<td>23</td>
</tr>
<tr>
<td>1.1.1 – Signal overlap</td>
<td>23</td>
</tr>
</tbody>
</table>
1.1.2 – Increases in signal linewidths with molecular weight
1.1.3 – Magnetisation transfer schemes employed in NMR pulse sequences
1.1.4 – Loss of phase coherence through rapid transverse relaxation.

1.2. Recent Technical and Experimental Developments.
1.2.1 – Isotopic enrichment of proteins
1.2.2 – Improvements in NMR experimental methodology
1.2.3 – Development of NMR instrumentation

1.3. Summary

1.4. An Outline of this Thesis.

Chapter II

Chapter Summary

2.1. Physiology and Pathophysiology of Methylarginines.
2.1.1 – Protein Arginine N-Methylation
2.1.2 – Guanidino-Methylated Arginines
2.1.3 – Nitric Oxide Synthase
2.1.4 – Asymmetric Methylarginines Inhibit Nitric Oxide Synthase (NOS)
2.1.5 – Metabolism of Asymmetric Methylarginines by DDAH

2.2. Structure-Function relationship of PaDDAH.
2.2.1 – Global description of the structure of PaDDAH
2.2.2 – Protein sequences and structures that are homologous to PaDDAH
2.2.3 – Description of the active site of PaDDAH
2.2.4 – Similarity of β-propellers
2.2.5 – Classifying the PaDDAH fold
2.2.6 – Is DDAH a metalloenzyme?

2.3. Summary.

Chapter III

3.1. General Chemicals and Laboratory Equipment.
3.2. General Protocols.
3.2.1 – Expression vector for PaDDAH
3.2.3 – Transformation of competent cells 74
3.2.4 – Large-scale expression of unlabelled PaDDAH from LB medium 74
3.2.5 – Large-scale expression of isotope-labelled PaDDAH from minimal medium 75
3.2.6 – Large-scale purification of PaDDAH 75
3.2.7 – Mutation of WT PaDDAH 77
3.2.8 – Small-scale expression tests of WT and mutant PaDDAH 78
3.2.9 – Measurement of enzymatic activity of WT and mutant PaDDAH samples 79
3.2.10 – Analytical size exclusion chromatography 80
3.2.11 – Protein samples prepared by other people 81

3.3. Nuclear Magnetic Resonance Spectroscopy. 81
3.3.1 – Heteronuclear single quantum coherence experiments (HSQC) 82
3.3.2 – Transverse relaxation optimised spectroscopy (TROSY) 82
3.3.3 – Measuring proton transverse relaxation rates 82
3.3.4 – Measuring translational diffusion coefficients using 1D $^1$H NMR 82
3.3.5 – Measuring $^{15}$N $T_1$, $T_2$ and $^{15}$N($^1$H) heteronuclear-NOE 82
3.3.6 – NMR experiments recorded to facilitate sequence-specific resonance assignment 85
3.3.7 – 3D $[^1$H, $^{15}$N] NOESY HSQC 85
3.3.8 – Processing NMR data 85
3.3.9 – Estimation of amide proton solvent exchange rates 86

3.4. Analysis of Data. 86
3.4.1 – Calculating signal-to-noise (S/N) ratios 86
3.4.2 – Derivation of $^{15}$N $R_1$, $^{15}$N $R_2$, $^{15}$N($^1$H) heteronuclear-NOE and estimation of $\tau_c$ 86
3.4.3 – Calculating proton transverse relaxation rates 88
3.4.4 – Calculating translational diffusion rates 89
3.4.5 – Estimating the dissociation constants of WT PaDDAH from analytical SEC 89

Chapter IV 91

Chapter Summary 91
4. Introduction.

4.1. A Model for Studying Large Proteins by NMR.

4.2. Analysis of $^{15}$N relaxation parameters: [$^2$H, $^{15}$N]-labelled ubiquitin in 50% (v/v) glycerol as a model for larger proteins.

4.2.1 - Extracting $^{15}$N relaxation parameters from NMR data

4.2.2 - $^{15}$N relaxation parameters from [$^{15}$N]- and [$^2$H, $^{15}$N]-labelled ubiquitin

4.2.3 - $^{15}$N relaxation parameters in the absence and presence of 50% (v/v) glycerol

4.2.4 - The effect of changing the acquisition temperature on $^{15}$N relaxation rates

4.2.5 - Secondary structure affects NH cross peak S/N ratio

4.3. Resonance Assignment of [$^2$H, $^{13}$C, $^{15}$N]-labelled Ubiquitin in 50% (v/v) Glycerol.

4.5. Discussion.

Chapter V

Chapter Summary

5. Introduction.

5.1. Expression and Purification of PaDDAH.

5.2. Analysis of [$^{15}$N]-labelled WT PaDDAH by NMR.

5.3. Analysis of [$^2$H, $^{15}$N]-labelled WT PaDDAH by NMR.

5.3.1 - Acquisition of HSQC-type spectra from perdeuterated WT PaDDAH

5.3.2 - Amide protonation of WT PaDDAH following expression in DM9.

5.4. Analysis of the solution and heteronuclear relaxation properties of WT PaDDAH by NMR.

5.4.1 - $^1$H transverse relaxation rates

5.4.2 - Pulse-field gradient 1D NMR translational diffusion experiments

5.4.3 - Analysis of $^{15}$N spin relaxation parameters for WT PaDDAH at 25°C

5.4.4 - The effect of increased temperature on $R_2$ and signal-to-noise values
5.5. Acquisition of spectra at different magnetic field strengths. 148
5.6. Towards residue specific resonance assignment of WT PaDDAH. 153
  5.6.1 – Acquisition of backbone $^1H$N, $^{15}N$, $^{13}Ca$ and $^{13}C\beta$ correlations 153
  5.6.2 – Initial efforts towards sequential resonance assignment 153
5.6. Discussion. 160

Chapter VI __________________________________________________166

Chapter Summary. 166
6.1. Introduction. 166
6.2. Analysis of the Self Association Properties of WT PaDDAH. 168
  6.2.1 – Analytical size exclusion chromatography: WT PaDDAH 168
  6.2.2 – Analytical size exclusion chromatography: ovalbumin and GST 168
  6.2.3 – Conversion of elution volume into apparent molecular weight 171
  6.2.4 – Estimation of the dissociation constant of WT PaDDAH using analytical SEC 171
6.3. Structural Description of Residues Contributing to the Homodimer Interface of WT PaDDAH. 177
  6.3.1 – A description of the protomer interface of WT PaDDAH 177
  6.3.2 – Analysis of the interface of WT PaDDAH using accessible surface areas 181
  6.3.3 – Selection of Residues for Mutation 183
6.4. Solubility and Activity Assays of WT and Mutant PaDDAH. 185
  6.4.1 – Activity Assays of WT PaDDAH 185
  6.4.2 – Activity Assays of PaDDAH Mutants 189
6.5. Analysis of the Self Association of PaDDAH Mutants. 189
6.6. Estimation of Equilibrium Dissociation Constant for PaDDAH. 194
6.7. Discussion. 195
Chapter VII

Chapter Summary

7.1. Introduction.

7.2. Analysis of R40E and R98H PaDDAH by NMR.

7.2.1 - \(^{1}H,\,^{15}N\)-TROSY spectra of \(^{15}N\)-labelled WT, R40E and R98H PaDDAH

7.2.2 - Measurement of the translational diffusion rates of R40E and R98H PaDDAH

7.3. Comparison of \(^{1}H,\,^{15}N\)-labelled WT, N36W and R40E PaDDAH.

7.3.1 - \(^{1}H,\,^{15}N\)-TROSY spectra of \(^{2}H,\,^{15}N\)-labelled WT, N36W and R40E PaDDAH

7.3.2 - Comparison of \(^{15}N\) relaxation rates of WT, N36W and R40E PaDDAH

7.4. Discussion.

Chapter VIII

Appendix


A.2. Evaluation of the Minimisation Function used to Optimise the Fit of Equation 6.12.

References
# List of Figures

## Chapter I

<table>
<thead>
<tr>
<th>Figure</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>24</td>
</tr>
<tr>
<td>1.2</td>
<td>24</td>
</tr>
<tr>
<td>1.3</td>
<td>28</td>
</tr>
<tr>
<td>1.4</td>
<td>34</td>
</tr>
<tr>
<td>1.5</td>
<td>35</td>
</tr>
<tr>
<td>1.6</td>
<td>40</td>
</tr>
</tbody>
</table>

## Chapter II

<table>
<thead>
<tr>
<th>Figure</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>49</td>
</tr>
<tr>
<td>2.2</td>
<td>54</td>
</tr>
<tr>
<td>2.3</td>
<td>55</td>
</tr>
<tr>
<td>2.4</td>
<td>59</td>
</tr>
<tr>
<td>2.5</td>
<td>61</td>
</tr>
<tr>
<td>Figure 5.10</td>
<td>149</td>
</tr>
<tr>
<td>Figure 5.11</td>
<td>150</td>
</tr>
<tr>
<td>Figure 5.12</td>
<td>151</td>
</tr>
<tr>
<td>Figure 5.13</td>
<td>152</td>
</tr>
<tr>
<td>Figure 5.14</td>
<td>154</td>
</tr>
<tr>
<td>Figure 5.15</td>
<td>156</td>
</tr>
<tr>
<td>Figure 5.16</td>
<td>158</td>
</tr>
<tr>
<td>Figure 5.17</td>
<td>159</td>
</tr>
</tbody>
</table>

**Chapter VI**

| Figure 6.1 | 166 |
| Figure 6.2 | 169 |
| Figure 6.3 | 170 |
| Figure 6.4 | 172 |
| Figure 6.5 | 175 |
| Figure 6.6 | 178 |
| Figure 6.7 | 179 |
| Figure 6.8 | 187 |
| Figure 6.9 | 188 |
| Figure 6.10 | 190 |
| Figure 6.11 | 193 |

**Chapter VII**

| Figure 7.1 | 202 |
| Figure 7.2 | 205 |
| Figure 7.3 | 207 |
| Figure 7.4 | 209 |
| Figure 7.5 | 211 |
List of Tables

Chapter I
Table 1.1 32

Chapter III
Table 3.1 77
Table 3.2 79
Table 3.3 83
Table 3.4 84

Chapter IV
Table 4.1 112
Table 4.2 116
Table 4.3 118

Chapter V
Table 5.1 146
Table 5.2 146
Table 5.3 155
Chapter VI
Table 6.1  176
Table 6.2  181
Table 6.3  184
Table 6.4  185
Table 6.5  195

Chapter VII
Table 7.1  206
Table 7.2  210

Appendix
Table A.1  228
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>1D</td>
<td>One-Dimensional</td>
</tr>
<tr>
<td>2D</td>
<td>Two-Dimensional</td>
</tr>
<tr>
<td>3D</td>
<td>Three-Dimensional</td>
</tr>
<tr>
<td>A.U.</td>
<td>Absorbance Units</td>
</tr>
<tr>
<td>ADI</td>
<td>Arginine Deiminase</td>
</tr>
<tr>
<td>ADMA</td>
<td>Asymmetric-L-$N^1$, $N^2$-dimethylarginine</td>
</tr>
<tr>
<td>AGAT</td>
<td>L-Arginine;Glycine amidinotransferase</td>
</tr>
<tr>
<td>ASA</td>
<td>Accessible Surface Area</td>
</tr>
<tr>
<td>AT</td>
<td>Amidinotransferase</td>
</tr>
<tr>
<td>βME</td>
<td>β-Mercaptoethanol</td>
</tr>
<tr>
<td>COLDER</td>
<td>Colour Developing Agent</td>
</tr>
<tr>
<td>COSY</td>
<td>Correlated Spectroscopy</td>
</tr>
<tr>
<td>CPMG</td>
<td>Carr-Purcell-Meiboom-Gill</td>
</tr>
<tr>
<td>CRINEPT</td>
<td>Cross Relaxation-Enhanced Polarisation Transfer</td>
</tr>
<tr>
<td>CRIPFT</td>
<td>Cross Relaxation Induced Polarisation Transfer</td>
</tr>
<tr>
<td>CSA</td>
<td>Chemical Shift Anisotropy</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Definition</td>
</tr>
<tr>
<td>--------------</td>
<td>------------</td>
</tr>
<tr>
<td>CT</td>
<td>Constant Time</td>
</tr>
<tr>
<td>CV</td>
<td>Column Volume</td>
</tr>
<tr>
<td>DDAH</td>
<td>Dimethylarginine Dimethylaminohydrolase</td>
</tr>
<tr>
<td>DI</td>
<td>Deiminase</td>
</tr>
<tr>
<td>eNOS</td>
<td>Endothelial NOS</td>
</tr>
<tr>
<td>FID</td>
<td>Free Induction Decay</td>
</tr>
<tr>
<td>GST</td>
<td>Glutathione-S-Transferase</td>
</tr>
<tr>
<td>HSQC</td>
<td>Heteronuclear Single Quantum Coherence</td>
</tr>
<tr>
<td>INEPT</td>
<td>Insensitive Nucleus Enhanced by Polarisation Transfer</td>
</tr>
<tr>
<td>iNOS</td>
<td>Inducible NOS</td>
</tr>
<tr>
<td>kD</td>
<td>kilo Daltons</td>
</tr>
<tr>
<td>LB</td>
<td>Luria Bertani</td>
</tr>
<tr>
<td>L-NMMA</td>
<td>L-\textsuperscript{\text{N}}-monomethylarginine</td>
</tr>
<tr>
<td>mAb</td>
<td>Monoclonal Antibody</td>
</tr>
<tr>
<td>MBP</td>
<td>Maltodextrin Binding Protein</td>
</tr>
<tr>
<td>MCS</td>
<td>Multiple Cloning Site</td>
</tr>
<tr>
<td>mRNA</td>
<td>Messenger Ribonucleic Acid</td>
</tr>
<tr>
<td>MSG</td>
<td>Malate Synthase G</td>
</tr>
<tr>
<td>Ni-NTA</td>
<td>Nickel-charged Nitriloacetic Acid</td>
</tr>
<tr>
<td>(nJ_{xy})</td>
<td>n-bond scalar coupling between nuclei x and y</td>
</tr>
<tr>
<td>NMR</td>
<td>Nuclear Magnetic Resonance</td>
</tr>
<tr>
<td>nNOS</td>
<td>Neuronal NOS</td>
</tr>
<tr>
<td>NO</td>
<td>Nitric Oxide</td>
</tr>
<tr>
<td>NOE</td>
<td>Nuclear Overhauser Effect</td>
</tr>
<tr>
<td>NOESY</td>
<td>Nuclear Overhauser Effect Spectroscopy</td>
</tr>
<tr>
<td>NOS</td>
<td>Nitric Oxide Synthase</td>
</tr>
<tr>
<td>PaDDAH</td>
<td>\textit{Pseudomonas aeruginosa} DDAH</td>
</tr>
<tr>
<td>PCR</td>
<td>Polymerase Chain Reaction</td>
</tr>
<tr>
<td>PDB</td>
<td>Protein Data Bank</td>
</tr>
<tr>
<td>PAD</td>
<td>Peptidyl Arginine Deiminase</td>
</tr>
<tr>
<td>PFG</td>
<td>Pulsed Field Gradient</td>
</tr>
<tr>
<td>PP2Ac</td>
<td>Catalytic domain of Protein Phosphotase 2A</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Definition</td>
</tr>
<tr>
<td>--------------</td>
<td>------------</td>
</tr>
<tr>
<td>PRMT</td>
<td>Protein Arginine Methyltransferase</td>
</tr>
<tr>
<td>RF</td>
<td>Radio-Frequency</td>
</tr>
<tr>
<td>S/N</td>
<td>Signal-to-noise</td>
</tr>
<tr>
<td>S-AdoMet</td>
<td>S-Adenosylmethionine</td>
</tr>
<tr>
<td>SDMA</td>
<td>Symmetric-L- $N^1, N^2$-dimethylarginine</td>
</tr>
<tr>
<td>SDS-PAGE</td>
<td>SDS-Polyacrylamide Gel Electrophoresis</td>
</tr>
<tr>
<td>SEC</td>
<td>Size Exclusion Chromatography</td>
</tr>
<tr>
<td>StrBl</td>
<td>L-Arginine,Inosamine Amidinotransferase</td>
</tr>
<tr>
<td>SW</td>
<td>Spectral Width</td>
</tr>
<tr>
<td>$T_1$</td>
<td>Spin-Lattice Relaxation Time Constant</td>
</tr>
<tr>
<td>$T_2$</td>
<td>Spin-Spin Relaxation Time Constant</td>
</tr>
<tr>
<td>$\tau_c$</td>
<td>Rotational Correlation time</td>
</tr>
<tr>
<td>TEV</td>
<td>Tobacco Etch Virus</td>
</tr>
<tr>
<td>TOCSY</td>
<td>Total Correlated Spectroscopy</td>
</tr>
<tr>
<td>TROSY</td>
<td>Transverse Relaxation-Optimised Spectroscopy</td>
</tr>
<tr>
<td>Y2H</td>
<td>Yeast Two Hybrid</td>
</tr>
</tbody>
</table>

**Amino acid nomenclature used in this thesis**

Cys-249 Residue type using the standard three letter code (e.g. cystiene) and residue number/position

C249S Mutation of Cys-249 to Ser-249
1. Introduction.

Nuclear magnetic resonance (NMR) spectroscopy is an established technique for probing three-dimensional (3D) structure and dynamical processes in biological macromolecules. Since the publication of the initial protein conformation elucidated using NMR (Williamson et al. 1985), great advances in both technology and experimental processes have occurred making NMR an alternative and complementary technique to X-ray crystallography for 3D structure determination. At the time of writing, a total of 3046 NMR-derived 3D structures, including 2446 proteins, 87 protein-nucleic acid complexes and 464 nucleic acids, had been deposited in the protein data bank (PDB; Berman et al. 2002). For comparison the total number of 3D structures present in the PDB was 19469 of which almost 85% were derived from X-ray crystallographic data. Although elucidating protein structures using crystallisation and X-ray diffraction was established some 25 years prior to the first protein NMR structure, the disparity between the numbers of 3D structures in the PDB is not solely due to the relative availability of the two techniques. The practice of NMR spectroscopy for 3D structure determination is subject to many fundamental limitations. These include a limited number of atomic elements with nuclei that possess the necessary quantum physical properties, low intrinsic sensitivity and a
restricted bandwidth of resonance frequencies. Further to these limitations on resolution and sensitivity, the molecular size and shape of the target molecule affects the longevity of the excited NMR signal and thereby again impacts upon the lineshape, sensitivity and resolution of the NMR spectrum.

In the history of the application of NMR, the 3D structure and dynamics of proteins were initially probed by $^1$H NMR techniques (Wüthrich 1986). Because of the relatively large number of signals observed, even in small proteins, and the limited bandwidth available, two-dimensional (2D) experiments were devised to ease the problem of signal overlap. The combined application of 2D-$^1$H-correlated spectroscopy (COSY), nuclear Overhauser spectroscopy (NOESY) and total correlation spectroscopy (TOCSY) experiments permitted the assignment of $^1$H resonance frequencies to individual atoms of small proteins (Bax 1989). Using interproton distance restraints derived from the intensity of NOESY cross peaks and torsion angle constraints derived from the quantitative estimation of scalar couplings it was possible to determine the 3D structure. Even with two frequency dimensions, $^1$H signal overlap is still problematic and proteins larger than 10 kD often yield spectra that are too crowded to resolve many features. As a technique, NMR has been noted for its inability to extract structural information from large proteins. X-ray crystallography, however, has been used to obtain 3D structures in excess of 1000 kD, including those of complete viruses.

In 2002 two laboratories published results showing how the molecular weight limitations previously hampering the progress of NMR have been, to some extent, surmounted. Kay and co-workers published nearly complete backbone assignments with accompanying information on backbone dynamics of malate Synthase G (MSG), a 723 residue 81.4 kD protein (Tugarinov et al. 2002). Wüthrich and colleagues used recently developed NMR pulse sequences to record interpretable NMR spectra from a 900 kD chaperone complex (Fiaux et al. 2002). In the decade preceding these publications, significant improvements in both experimental methodology and instrument specification had occurred that enabled NMR to be applied to proteins of these proportions. In the remainder of this chapter the reasons that larger molecular weight proteins are difficult to characterise by NMR will be discussed along with a
summary of the most recent developments in sample preparation, NMR experimental methods and instrumentation.

1.1. Problems Associated with Studying Large Proteins by NMR.

Figure 1.1 shows the distribution of molecular weights of 3D protein structures solved by NMR in the PDB as of December 2002 (Berman et al. 2002). As can be seen, there is a heavy bias towards the lower molecular weight proteins with relatively few solution structures of proteins of more than 200 amino acid residues. The intrinsic factors responsible for the molecular weight limitation on NMR spectroscopy are discussed in the following sections.

1.1.1 – Signal overlap

Signal overlap in NMR spectroscopy can be problematic. Each type of observable nucleus has only a limited bandwidth of resonance frequencies. This limitation is a consequence of the range of electronic environments that a particular moiety experiences in a folded polypeptide. The chemical shift range for the amide proton, for example, is normally confined to a range of approximately 5 ppm (approximately between 6 ppm and 11 ppm). However, each amino acid residue in a protein (with the exception of proline) has an amide proton and therefore the number of N\(^1\)H\(^N\) signals in a spectrum will increase with the size of the protein. This can be seen in Figure 1.2. The one-dimensional (1D) N\(^1\)H NMR spectrum of a small peptide presented in Figure 1.2(a) shows well-resolved amide proton signals. In higher molecular weight proteins, however, which have a greater number of N\(^1\)H\(^N\) nuclei, signal overlap occurs and the individual linewidths are broader (Figures 1.2(b) and 1.2(c)). In the case of the large protein homodimer (Figure 1.2(c)), it is difficult to discern separate signals corresponding to individual N\(^1\)H\(^N\) resonances. The problem of signal overlap is not confined to proton NMR spectroscopy. Other NMR-visible nuclei also have limited resonance bandwidths.

1.1.2 – Increases in signal linewidths with molecular weight

Magnetization precessing in the transverse plane with angular frequency \(\omega_0\) decays at a rate of \(1/T_2\), where \(T_2\) is the spin-spin or transverse relaxation time constant (see Chapter 4, Section 4.2.1 for further discussion). Fourier transformation of the time
Figure 1.1. Histogram showing the frequency of 3D NMR solution structures in the protein data bank (PDB) compared to the length of the amino acid sequence. Correct as of December 2002.

Figure 1.2. 1D $^1$H NMR spectra of (a) 1.8 kD intrinsically helical peptide with 21 residues, (b) a monomeric 76 residue, 8.6 kD protein and (c) a homodimeric 2 × 254 residue 60 kD protein. The regions shown correspond to amide and aromatic proton resonance frequencies. Spectra are not drawn to scale in the vertical axis.
domain free induction decay (FID) of a NMR signal generates a Lorentzian peak centred at $\omega_0$. A convenient measure of the linewidth of this signal is the full width at half height ($\Delta v_{\text{FWHH}}$), which is reported in Hertz (Equation 1.1).

$$\Delta v_{\text{FWHH}} = \frac{1}{\pi T_2}$$  
Equation 1.1

In experimental reality, the $T_2$ term should be replaced by $T_2^*$ to take account of the effects of inhomogeneity of the magnetic field, $T_2_{\text{inhom}}$, such that $1/\pi T_2^* = 1/\pi T_2 + 1/\pi T_{2_{\text{inhom}}}$ (Cavanagh et al. 1995). As will be discussed later, the $T_2$ value of a NMR signal is approximately inversely proportional to the rate of molecular rotational reorientation and consequently is strongly dependent on the molecular size (see Chapter 4, Section 4.2.1). The importance of the latter statement is that in general larger molecules have shorter $T_2$ times and therefore broader linewidths. In the case of large proteins, the problem of broad linewidth can be very pronounced (Figure 1.2(c)).

1.1.3 – Magnetisation transfer schemes employed in NMR pulse sequences

In NMR spectroscopy magnetisation can be transferred between coupled nuclei. Two types of coupling exist, scalar and dipolar. Scalar coupling can occur between directly or indirectly bonded nuclei where polarisation of the covalent bonding electrons communicates the magnetic character of one nucleus to another. An example is the three bond scalar coupling that occurs between the amide and carbon alpha protons, $^3J_{HNHa}$, in an amino acid residue of a protein. The spin state of the amide proton (i.e. aligned or against the external magnetic field) influences the transition frequency of the Hα proton, such that the Hα resonance is split in two, forming a doublet. The difference in resonance frequency between the doublet components is independent of the external magnetic field and described by the coupling constant $J$. For the general scalar coupling, the size of the coupling constant is dependent on the number of bonds between the coupled nuclei and their gyromagnetic ratios. In general, the magnitude of one bond scalar couplings between a heteronucleus and a proton (e.g. $^1J_{NH}$ equals 90 Hz) tend to be larger than proton-proton scalar couplings (e.g. $^3J_{HNHa} < 10$ Hz, dependent on polypeptide structure).
Dipolar interactions do not require the two nuclei to be bonded. In simple terms, the orientation of a magnetic moment of one nucleus (i.e. aligned or against the applied magnetic field) influences that of another nucleus. A more extensive description of dipolar coupling is presented in Chapter 4, Section 4.2.1.

The most commonly used pulse sequence that utilizes cross-relaxation between dipolar-coupled spins to establish through-space correlations is the NOESY experiment. In the basic 2D homonuclear NOESY experiment, chemical shift encoded Z-magnetization from one proton nucleus, I, is transferred through-space to a dipolar-coupled partner, S. The NOESY experiment yields a 2D spectrum, with frequency dimensions $F_1$ and $F_2$ in which both diagonal peaks and cross peaks are observed. The cross peaks represent the through-space correlation between nuclei I and S that occurs as a result of dipolar coupling.

Coherence can be transferred between pairs of nuclei with absolute values of scalar couplings greater than zero. Spin-spin coherence can be transferred as either in-phase or anti-phase magnetization. Two well known NMR experiments develop these ‘through-bond’ correlations. COSY experiments use evolution of scalar couplings in combination with a 90° RF mixing pulse to transfer antiphase magnetization of one nucleus to antiphase magnetization of a scalar coupled nucleus. A limitation of the COSY experiment is that in general only nuclei that are separated by no more than three covalent bonds have scalar couplings of sufficient magnitude to be exploited. TOCSY experiments make use of composite-pulse mixing sequences to effect transfer of in-phase transverse magnetization around a J-coupled spin system. For example, a homonuclear $^1$H TOCSY experiment can generate correlations between all the protons in a single amino acid. The theory behind such TOCSY transfers is complex and beyond the scope of the chapter.

The remainder of this section will deal exclusively with correlated spectroscopy of heteronuclear spin systems. The standard product operator notation will be used to describe the effects of radiofrequency (RF) pulses and evolution of chemical shift and scalar couplings (Cavanagh et al. 1995). The practical considerations of enrichment of proteins with heteronuclei are discussed in more detail in Section 1.2.1.
In heteronuclear protein NMR spectroscopy the building block of many pulse sequences is the INEPT (Insensitive Nuclei Enhanced by Polarisation Transfer) pulse sequence (Figure 1.3(a)). In the following description of the INEPT pulse sequence the heteronuclear $^{15}$N-$^1$H spin system will be used as an example. The principles can be equally applied to other scalar coupled nuclei (e.g. $^1$H and $^{13}$C). The function of the INEPT sequence is use the superior polarisation of the higher $\gamma$ nucleus to enhance the signal for the lower $\gamma$ nucleus. The sensitivity of a particular nucleus is related to the Boltzmann distribution of spins between the low and high energy states. The parameter determining the NMR sensitivity of a nucleus is the gyromagnetic ratio, $\gamma$. When a sample of spin-\(\frac{1}{2}\) nuclei is placed in an applied magnetic field, the energy difference created between the lower and upper energy levels is proportional to the gyromagnetic ratio of the nucleus (see Table 1.1 and Section 1.2.3). The gyromagnetic ratio of the proton, $\gamma_H$, is approximately 10 times larger than that of the $^{15}$N, $\gamma_N$, and will therefore have a larger magnetic moment and stronger NMR signal.

The INEPT sequence enhances sensitivity by excitation of the high $\gamma$ nucleus (e.g. $^1$H) to create transverse magnetisation and then allowing the heteronuclear scalar coupling to effect evolution to heteronuclear antiphase magnetisation. A brief description of the effect of each pulse and time delay period in the INEPT pulse sequence is given below and in Figure 1.3 for the $^1$H-$^{15}$N moiety. For brevity, the modulation of each product operator by chemical shift and scalar coupling evolution will not be presented in full. The first 90° RF pulse converts equilibrium $^1$H magnetisation, $aH_x$ (Figure 1.3(b), time ‘1’), into coherent transverse magnetisation $-aH_y$ (Figure 1.3(b), time ‘2’). In this nomenclature $(-)aX_i$ represents net $X$ magnetisation along the $(-)i$-axis. The coefficient ‘$n$’ describes the strength of $X$ magnetisation (a is proportional to $\gamma_H$ and b is proportional to $\gamma_N$). During the period $2\tau$ both the chemical shift of the proton, $\Omega_H$, and the scalar coupling of the H and N nuclei, $J_{HN}$, evolve. The 180° RF pulse applied to both $^{15}$N and $^1$H nuclei situated in the middle of $2\tau$ period causes only the evolution due to the chemical shift $\Omega_H$, to refocus. The scalar coupling is permitted to evolve over the whole $2\tau$ period. Evolution of $-aH_y$ magnetisation created by the first 90° RF pulse through the influence of the HN scalar coupling produces antiphase $a2H_xN_z$ magnetisation (Figure 1.3(c)). Setting the $2\tau$ period to $1/(2 \times J_{HN})$, where $J_{HN}$ is the one-bond scalar coupling between bonded $^{15}$N and $^1$H nuclei, allows maximisation of
Figure 1.3. The INEPT pulse sequence. (a) Schematic of the INEPT pulse sequence. Each black box represents a 90° (thin) or 180° (thick) RF pulse applied along the indicated axis. The upper line corresponds to pulse at the frequency of the $^{1}H$ nucleus and lower line at the frequency $^{15}N$ nucleus. The period $\tau$ is set to $1/(4\times J_{HN})$. (b) A simplified description of the INEPT experiment using the standard product operator formulation. The state of $H$ and $N$ magnetisation is indicated for each position in (a). For additional descriptions see text. (c) Transformation of the $-aH_y$ operator produced by the $^1J_{HN}$ scalar coupling.
the antiphase a2H,N term – i.e. a 90° rotation around the transverse xy plane (Figure 1.3(b) time, '3'). Application of the final pair of 90° RF pulses converts the antiphase a2H,N term into one with antiphase transverse N magnetisation, a2H,N (Figure 1.3(a) time ‘4’). In this situation the N magnetisation has the intensity defined by the γH-containing coefficient a. Therefore, in the example above involving 1H and 15N nuclei, the INEPT sequence generates 15N antiphase transverse magnetisation with the intensity derived from the attached proton. From their respective γ values, |γH/γN|, this represents an increase in 15N signal intensity of approximately 10. The remaining small and undesired -bN_y term, which has not been enhanced by the proton, is typically removed by phase cycling.

The basic version of the commonly used HSQC (Heteronuclear Single Quantum Coherence) pulse sequence (see Materials and Methods' Section 3.3.1) employs two INEPT sequences. The first INEPT sequence generates antiphase transverse 15N magnetisation, as described above. In between the INEPT sequences in the HSQC experiment, a delay period is introduced to allow chemical shift encoding of 15N magnetisation. The second INEPT sequence, which is really a reverse INEPT, converts the frequency-labelled 15N magnetisation back into 1H for detection. A variation of the INEPT sequence is termed a refocused INEPT. In this pulse sequence, an additional [-Δ-180°(1H, 15N) -Δ-] element is applied after the INEPT sequence. This additional elements allows the 1H, 15N J-coupling to be refocused producing in-phase 15N magnetisation. For that reason refocused INEPT sequences are employed in experiments that measure 15N longitudinal and transverse relaxation rates which require in-phase N_L and N_xy magnetisation, respectively (see Chapter 4, Section 4.2.1). As with the INEPT sequence, the reverse of a refocused INEPT sequence produces in-phase 1H magnetisation for detection from in-phase 15N magnetisation. In reality, both the HSQC and 15N relaxation experiments are more complicated than the simplified explanations presented here. There are many additional features that are now employed as standard which are outside the scope of this review. These include selective pulse water suppression techniques, pulse field gradients coherence selection, phase cycling and additional pulses to improve sensitivity.
1.1.4 – Loss of phase coherence through rapid transverse relaxation.

To achieve a NMR signal transverse magnetisation has to be created. Specifically, to detect a signal in NMR the net magnetisation that arises from placing a sample of spin-$\frac{1}{2}$ nuclei into a magnetic field $B_0$ has to be rotated into a plane that is transverse to the direction of $B_0$. As with other forms of spectroscopy, following the initial excitation energy is released as the system returns to the equilibrium state. In NMR spectroscopy the energy difference between the two energy levels is comparatively small and corresponds to radio-frequency electromagnetic radiation (in the range 100 MHz to 1 GHz). In comparison to alternative spectroscopic excited states, the rate at which magnetisation relaxes in NMR is very slow. Phase coherence in NMR can remain for a sufficiently long period for it to be transferred between spins (see Section 1.1.3). However, on this transfer time scale, the finite rate of relaxation of the NMR signal can present problems.

In many multi-dimensional NMR experiments phase coherence is transferred between scalar-coupled nuclei (see Section 1.1.3). The optimal duration of the magnetisation transfer elements employed is determined by the magnitude of the scalar coupling between the two subject nuclei. In the case of the $^{15}$N–$^1$H bond, the INEPT sequence (see Section 1.1.3) employs two delay periods of $1/(4\timesJ_{\text{NH}})$, where $J_{\text{NH}} \approx 90$ Hz, to facilitate magnetisation transfer from $^1$H$^N$ to $^{15}$N. Therefore, each INEPT element in a pulse sequence adds approximately 5.6 ms to the length of an experiment. For small proteins where $^{15}$N $T_2$ values are greater than 100 ms the duration of such INEPT transfer steps is not a limiting factor. However, for larger proteins the $^{15}$N $T_2$ time is severely shortened and extended periods of transverse magnetisation are accompanied by substantial loss of signal.

The rapid relaxation of $^{13}$C transverse magnetisation is a particularly limiting factor in the acquisition of multi-dimensional triple-resonance NMR data for larger proteins (Gardner and Kay 1998). The scalar couplings between aliphatic (e.g. $J_{\text{CaCp}}$) and backbone ($J_{\text{COC}}$) carbon nuclei are smaller than the NH coupling: $J_{\text{CaCp}}$ and $J_{\text{COC}}$ equals 35 Hz and 55 Hz respectively, whereas $J_{\text{NH}}$ equals 90 Hz. Therefore, longer delay periods are required to facilitate coherence transfer scalar coupled $^{13}$C nuclei, which in turn will be subject to greater signal loss through transverse relaxation. In
addition, the short \(^{13}\text{C}\) \(T_2\) time limits the maximum length of any \(^{13}\text{C}\) frequency encoding periods and therefore the spectral resolution available in that dimension. A recent study of the 37 kD trp repressor/operator complex, which has a reported isotropic tumbling time of 15 ns, provided an illustration of how the \(^{13}\text{C}\) \(T_2\) times of larger proteins can prohibit the application of 3D NMR experiments (Gardner and Kay 1998). In this system, the average \(^{13}\text{Ca}\) \(T_2\) time measured was 16 ms which limited both the number of COSY-type magnetisation transfer steps that could be applied and the optimum length of \(^{13}\text{C}\) frequency encoding periods.

Many of the multidimensional heteronuclear NMR experiments that correlate protein backbone \(^1\text{H}\), \(^{13}\text{C}\) and \(^{15}\text{N}\) nuclei are based on through-bond transfer of antiphase coherence using INEPT style pulse elements (such experiments are discussed in Section 1.2.2). Transfer of magnetisation from one type of nucleus to another requires an INEPT element with the correct delay periods. The experiments commonly used to correlate protein backbone nuclei start and finish with proton magnetisation. This requires transferring magnetisation from the proton to the heteronuclei of interest via one or more scalar coupling and then back again. Each INEPT sequence is therefore used twice — once forward on the way ‘out’ and then in reverse on the way ‘back’. In larger proteins, certain experiments are not feasible due to the loss of signal that occurs during the numerous INEPT elements as a result of magnetisation losses due to transverse relaxation.

1.2. Recent Technical and Experimental Developments.

1.2.1 — Isotopic enrichment of proteins

\(^1\text{H}\) is the most naturally abundant spin-½ nucleus. However, other spin-½ nuclei are available in nature, albeit at lower natural abundances. Spin-½ \(^{12}\text{C}\) and \(^{14}\text{N}\) isotopes are now routinely incorporated into proteins using uniform or selective labelling strategies (Gronenbom and Clore 1995). With these additional nuclei, the problem of \(^1\text{H}\) signal overlap can be very much relieved by acquiring data in two or more frequency dimensions. Following uniform \([^{15}\text{N}]\)-enrichment of a protein, NMR experiments can be recorded that correlate the \(^1\text{H}\) and \(^{15}\text{N}\) resonance frequencies of amide NH moieties. The addition of the second, \(^{15}\text{N}\) dimension allows individual NH
cross peaks to be separated. The resulting spectrum shows a 'fingerprint' of the protein with a NH correlation observed for each backbone amide group (see, for example, Figure 4.1(a)). One drawback is the relative insensitivity of the $^{15}$N as determined by its gyromagnetic ratio (Table 1.1), but great increases in sensitivity can be afforded through the use of INEPT-based experiments (see Section 1.1.3).

<table>
<thead>
<tr>
<th>Nucleus</th>
<th>$\gamma$ (T s)$^{-1}$</th>
<th>Natural Abundance (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$^1$H</td>
<td>$2.6752 \times 10^8$</td>
<td>99.98</td>
</tr>
<tr>
<td>$^2$H</td>
<td>$4.107 \times 10^7$</td>
<td>0.02</td>
</tr>
<tr>
<td>$^{13}$C</td>
<td>$6.728 \times 10^7$</td>
<td>1.11</td>
</tr>
<tr>
<td>$^{15}$N</td>
<td>$-2.712 \times 10^7$</td>
<td>0.36</td>
</tr>
</tbody>
</table>

Table 1.1. Properties of selected nuclei relevant to biomolecular NMR spectroscopy

Taken from Cavanagh et al. 1995.

Enriching a protein with $^{13}$C as well as $^{15}$N permits the recording of NMR experiments with even more dimensions. Many NMR pulse sequences have been developed that correlate backbone and side chain $^1$H, $^{13}$C and $^{15}$N spins in three and four dimensions (Gronenborn and Clore 1995; Cavanagh et al. 1995). The acquisition of such multi-dimensional data sets permits numerous inter-atomic through-bond and through-space interactions to be individually identified. This is a significant improvement on the information content that can be extracted from 1D NMR.

Although relieving the overlap problems associated with studying larger and more complex molecules, isotope labelling with $^{13}$C and $^{15}$N presents another problem: the issue of rapid transverse relaxation of $^{13}$C and $^{15}$N nuclei and the limits in resolution that these properties impose on multi-dimensional data (see Section 1.1.4). The relaxation of transverse magnetisation of a heteronucleus with a directly bonded proton, such as $^{13}$C–$^1$H or $^{15}$N–$^1$H, is governed partly by the dipole-dipole interaction between the two bonded atoms (see Chapter 4, Section 4.2.1). The rate of dipolar-mediated relaxation is dependent on the product of the gyromagnetic ratios, $\gamma$, of the two bonded nuclei (Table 1.1). Simplistically, the larger the product of the two
gyromagnetic ratios the greater the rate of dipolar relaxation. Inspection of Table 1.1 shows that deuterium has a smaller $\gamma$ than normal hydrogen: $\gamma_D/\gamma_H = 0.153$. Substitution of the proton bonded to heteronucleus with a deuteron – deuteration – will therefore decrease the relaxation rate of the heteronucleus proportional to $(\gamma_D/\gamma_H)^2$ (Sattler and Fesik 1996). This type of substitution is now commonly performed to increase $^{13}$C $T_2$ times in higher molecular weight proteins (Gardner and Kay 1998; Kay 2001).

Long-range homonuclear dipolar interactions are responsible for the flow of magnetisation around the extensive network of dipolar-coupled protons present in proteins (Figure 1.4(a)). These spin diffusion pathways result in the efficient and rapid relaxation of proton transverse magnetisation. Furthermore, heteronuclear antiphase magnetisation relaxes more rapidly than in-phase heteronuclear magnetisation because of the efficient relaxation of proton magnetisation around the spin diffusion network (Section 1.1.3). In a fully deuterated, amide-protonated protein these pathways are almost eliminated (Figure 1.4(b)). In this case $^{15}$N relaxation is almost totally dependent on the isolated NH interaction.

Several strategies exist for the preparation of $[^2\text{H}]$-enriched proteins which vary in the level and specificity of substitution achieved (Sattler and Fesik 1996; Goto and Kay 2000). As NMR of $^{13}$C and $^{15}$N is intrinsically less sensitive than $^1\text{H}$, a subset of protons are usually retained for the purpose of exciting and recording magnetisation. In most deuteration strategies, the protein is purified in H$_2$O-based solvents, which promotes the protonation of any exchangeable deuterated groups (e.g. hydroxyl, thiol, and backbone and side chain amide, guanidino and amine groups). As each amino acid residue contains an amide NH group, a generic starting point for NMR experiments is regained.

For NMR studies of globular proteins in excess of 20 kD, it is now relatively common practice to employ random fractional deuterium labelling (Figure 1.5(a)). This strategy involves the expression of a protein in a medium containing D$_2$O. During cell growth, the expression host incorporates deuterons from the medium into its amino acid reservoir and consequently into any translated protein products. The final level of
Figure 1.4. The effect of deuteration on $^1$H spin diffusion pathways. (a) schematic view of homonuclear dipolar couplings (white arrows) in protonated proteins (top) and a 3D structure of a protein showing all the protons present (bottom). All backbone amide protons are coloured light blue and the remainder are coloured yellow; (b) As with (a) but for a uniformly 100% deuterated, amide protonated protein. The 3D structure in (b) shows only amide protons.
Figure 1.5.
**Figure 1.5.** Common methods for isotopic enrichment. In the schematic representations of isotopically enriched proteins (bottom panels), all NMR-visible spin-½ nuclei ($^1$H, $^{13}$C and $^{15}$N) are coloured black, deuterons are coloured dark grey and all other NMR-invisible nuclei ($^{12}$C, $^{15}$N, $^{16}$O) are coloured light yellow. (a) Random fractional deuteration. The protein is expressed in [$^2$H,$^{13}$C,$^{15}$N]-enriched medium (top) and purified in H$_2$O-based buffers (middle). The result (bottom) is uniform [$^{13}$C,$^{15}$N]-labelling with [$^2$H]-enrichment at all non-labile sites (Kay and Gardner, 1998); (b) Methyl protonation in an otherwise [$^2$H,$^{13}$C,$^{15}$N]-labelled protein. The expression protocol is as described as (a) but with [3,3-$^3$H] $^{13}$C-$\alpha$-ketobutyrate and [3-$^2$H] $^{13}$C-$\alpha$-ketoisovalerate supplemented 1 hour prior to induction (top). Following purification in H$_2$O (middle), the protein is uniformly deuterated except $\gamma$-methyl groups of valine, $\delta$-$\gamma$ methyl group of isoleucine and $\delta$-methyl groups of leucine (bottom; Goto *et al.*, 1999); (c) Amino acid specific labelling using *in vitro* protein synthesis. Protein is expressed in a concentrated bacterial cell extract containing, for example, [$^{13}$C,$^{15}$N]-labelled alanine with the other 19 amino acids unlabelled (top). Following purification in H$_2$O (middle) the protein contains only [$^{13}$C,$^{15}$N]-labelled alanine residues (bottom; Kigawa *et al.*, 1995).
deuteron enrichment achieved in a protein varies from site to site. For example, the level of Ca deuteration is approximately equivalent to the H/D ratio used in the expression medium (Rosen et al. 1996; McCallum et al. 1999). The level of incorporation achieved at other aliphatic carbon positions tends to be lower. In most bacteria, the use of protonated glucose as the carbon source leads to a high proportion of protonated aromatic carbons (Gardner and Kay 1998). This is because the biosynthetic pathways that generate aromatic amino acids directly utilise the protonated carbons of glucose in the manufacture of the aromatic groups. More uniform side chain deuteration can therefore be achieved through the use of $[^2H, ^{13}C]$-labelled glucose in the expression medium, but at greater financial expense.

A caveat of protein expression in deuterated media is incomplete protonation of backbone amide groups during purification. If a protein is expressed and purified in the native state from a medium containing a high level of $^2H$, a number of backbone amide groups can remain deuterated for a considerable period of time (Salzmann et al. 2000; Tugarinov et al. 2002). This is particular common for those amide groups within the core of the protein. For example, following the deuteration of human glutathione-S-transferase (GST), 15 % of the backbone amide groups remained deuterated for up to four months (McCallum et al. 1998). If possible, it is preferable to express the protein in unfolded state (e.g. into inclusion bodies) or unfold and refold the natively expressed protein during the purification (Tugarinov et al. 2002).

Although likely to benefit backbone resonance assignment, the substitution of the majority of non-labile protons with deuterons can lead to difficulties later on in efforts to determine proteins structures. The primary distance restraint extracted from NMR data is derived from the inter-proton NOE. Ideally, a large number of NOE restraints per residue are required with a good proportion being between protons distant in the primary structure. In the general case, one can divide the possible yield of NOE measurements into three different categories: those that connect amide protons; those that connect amide and aliphatic protons; and those that connect aliphatic protons (Sattler and Fesik 1996). In a highly deuterated, amide-protonated protein NOEs will only be observed between amide protons. The derived distant restraints can typically only be used to determine the secondary structure elements of the protein and perhaps the overall backbone conformation at low resolution (Kay 2001). To overcome this
problem, several strategies have been developed to specifically introduce side chain protons whilst retaining a high level of deuteration at the Ca and Cβ positions (Goto et al. 1999; Goto and Kay 2000). Kay and co-workers initially developed an expression system that protonated selected amino acid methyl groups in an overall \([^2\text{H}, ^{13}\text{C}, ^{15}\text{N}]\)-labelled protein using \(^{13}\text{C}\)-labelled pyruvate as the sole carbon source. However the resulting \([^1\text{H}, ^{13}\text{C}]\)-spectra yielded asymmetric methyl cross peaks resulting from a distribution of isotopomers (i.e. \(\text{CH}_3\), \(\text{CH}_2\text{D}\) and \(\text{CHD}_2\); \(\text{CD}_3\) would also be present but would not be observed) and \(^2\text{H}\) isotope shifts (Rosen et al. 1996).

An improved protocol for the protonation of valine, leucine and isoleucine methyl groups in otherwise \([^2\text{H}, ^{13}\text{C}, ^{15}\text{N}]\)-labelled proteins that did not lead to H/D isotopomers was later reported (Goto et al. 1999). Protonation of valine \(\gamma\), leucine \(\delta\)- and isoleucine \(\delta\)-methyl groups was achieved by addition of \([3,3-^2\text{H}_2]\) \(^{13}\text{C}\)-\(\alpha\)-ketobutyrate and \([3,^2\text{H}]\) \(^{13}\text{C}\)-\(\alpha\)-ketoisovalerate to a growth medium containing \([^2\text{H}, ^{13}\text{C}]\)-labelled glucose one hour prior to the induction of protein expression (Figure 1.5(b)). This methodology was used to obtain long-range NOE distance constraints for the 370 residue maltodextrin binding protein (MBP). With these distance constraints and with the aid of residual dipolar couplings (Tjandra and Bax 1997) a 3D solution structure of this protein was elucidated to a precision of 2.2 Å (Mueller et al. 2000).

In highly deuterated proteins, proton spin diffusion pathways are effectively diluted, which can lead to a dramatic increase in the \(T_1\) time for the remaining protons. It is therefore sensible to increase the relaxation delay to ensure more complete recovery of \(^1\text{H}\) magnetisation between transients. Tugarinov et al. measured average proton \(T_1\) times of 2.03 ± 0.84 s for 81.4 kD MSG (Tugarinov et al. 2002). If the recycle delay is set to three times the proton \(T_1\) time then only 95 % of \(^1\text{H}\) magnetisation has fully relaxed. To achieve a value greater than 99 %, the delay has to be five times the \(T_1\) time. In the case of MSG, this would require a delay of ten seconds for the average \(T_1\) and longer still for more slowly relaxing amide protons. However, a commonly used rule of thumb is to set the recycle delay between transients to around \(1.3 \times T_1\) (Freeman 1987). This method optimises the competing factors of recovery of equilibrium magnetisation and experiment duration. A beneficial by-product of the methyl protonation strategy of Goto et al. (Goto et al. 1999) is that the introduction of
aliphatic protons into an otherwise perdeuterated, amide protonated proteins increases the efficiency of proton longitudinal relaxation, thereby avoiding this problem.

In addition to the uniform and partial deuteration strategies outlined above, specific incorporation of certain isotope-labelled amino acids is also possible. Isotope-labelled amino acids can be added to the expression medium prior to induction. However, unless bacteria that are auxotrophic for the chosen amino acid type(s) are used, metabolic scrambling of the isotope labels is likely to occur. One way to circumvent this problem is the use of in vitro protein synthesis (Figure 1.5 (c); Kigawa et al. 1999). For example, Kigawa et al. showed in an elegant series of experiments, samples of human c-Ha-Ras could be expressed at high yield using *E. coli* cell-free expression systems with uniform or residue-specific $[^{15}\text{N}]$-labelling (Kigawa et al. 1995). $[^1\text{H}, \, ^{15}\text{N}]$-HSQC spectra were recorded to confirm that only $[^{15}\text{N}]$-serine or $[^{15}\text{N}]$-aspartate had been incorporated at the correct locations and that no metabolic scrambling had occurred.

1.2.2 – Improvements in NMR experimental methodology

The development of triple resonance experiments for $^{13}\text{C}, \, ^{15}\text{N}$ enriched proteins greatly reduces the problem of signal overlap (Bax 1994). This class of experiments is now reasonably familiar to practitioners of biomolecular NMR spectroscopists. A detailed description of the theory and practice of these methods has been discussed elsewhere (Gardner and Kay 1998; Sattler et al. 1999) and therefore is it not necessary to summarise this material at any length in this thesis.

An extensive range of ‘out-and-back’ triple resonance NMR pulse sequences have been reported that correlate backbone $^1\text{H}^N$, $^{15}\text{N}$ and $^{13}\text{C}$ spins in proteins (Cavanagh et al. 1995). These experiments can be used for sequence-specific backbone resonance assignment (Gardner and Kay 1998). This type of experiment relies upon initiating polarisation of the amide proton, transfer of coherence to the desired heteronucleus to allow encoding of the X-nucleus chemical shift, coherence transfer back to the more sensitive amide proton before concluding with detection of a $[^{15}\text{N}]$-decoupled $^1\text{H}^N$ FID (see Section 1.1.3). The number of coherence transfer and frequency-encoding steps are varied to permit different nuclei to be correlated and their resonance frequencies recorded. Due to sensitivity considerations, $^{13}\text{C}$ transverse relaxation rates
generally impose a limit on the molecular weight of a protein that this type of experiment can be applied to (see Section 1.1.3). Deuteration at Ca and Cβ positions can suppress $^{13}$C transverse relaxation rates sufficiently to permit larger protein systems to be investigated (see Section 1.2.1). However, as 'out-and-back’ experiments require protonated amides, the rate of $^{15}$N relaxation becomes one of the major determinants of sensitivity. A significant experimental advance came with the development of transverse relaxation optimised spectroscopy (TROSY; Pervushin et al. 1997). In conventional 2D [$^1$H, $^{15}$N]-heteronuclear spectroscopy, the four-fold multiplet structure of a NH correlation is collapsed to give a single cross peak through the use of 180° RF pulses on $^1$H ($F_1$) and broad band $^{15}$N decoupling during acquisition ($F_2$) (Figure 1.6(a); Riek et al. 2000). In [$^1$H, $^{15}$N]-TROSY pulse sequences, the NH signal remains in the multiplet state with resonance frequencies of $(\omega_H \pm \pi J_{NH})$ and $(\omega_N \pm \pi J_{NH})$, specified by the $^1J_{NH}$ scalar coupling between H and N nuclei (Figure 1.6(b)). Importantly, in general and at any magnetic field strength the transverse relaxation rates of each of the individual multiplet components are different. The TROSY principle exploits the interference between dipolar and chemical shift anisotropy (CSA) mediated transverse relaxation that gives rise to the different linewidth (Figure 1.6; for a description of dipolar and CSA interactions see Chapter 4, Section 4.2.1). For the $^1$H signal, the $(\omega_H + \pi J_{NH})$ component of the scalar coupled multiplet has a slower transverse relaxation rate than that with $(\omega_H - \pi J_{NH})$ (Figure 1.6(b)). The reverse is true for the $^{15}$N signal where the $(\omega_N - \pi J_{NH})$ component relaxes more slowly than $(\omega_N + \pi J_{NH})$ (Figure 1.6(b)). Overall, in two dimensions, the component of the four-peak multiplet at $(\omega_H + \pi J_{NH}, \omega_N - \pi J_{NH})$ will exhibit slow transverse relaxation in both dimensions even for large proteins (Wider and Wüthrich 1999; Pervushin 2000; Kay 2001). The original TROSY pulse sequence was designed to actively select only the multiplet component with the slowest transverse relaxation rate and suppress the other components through phase cycling (Figure 1.6(c); Pervushin et al. 1997). The expectation that the TROSY principle yields a sharp component is based upon the assumption that for each of the $^{15}$N and $^1$H nuclei in a NH bond the dominant relaxation mechanisms are the heteronuclear NH dipolar coupling and the $^{15}$N (or to a lesser extent the $^1$H) CSA. For the nitrogen nucleus this assumption is valid. In the case of the amide proton, the possible contributions of homonuclear proton-proton couplings have to be considered. The
Figure 1.6. The appearance of NH crosspeaks in theoretical $[^1\text{H}, ^{15}\text{N}]$-spectra. (a) A single NH resonance in an HSQC spectra with refocusing of HN scalar coupling during $t_1$ and broad band $^{15}\text{N}$ decoupling during $t_2$; (b) The multiplet structure of the NH scalar coupling in an HSQC without refocusing of HN scalar coupling during $t_1$ or broad band $^{15}\text{N}$ decoupling during $t_2$; (c) A TROSY spectrum with active selection of the narrowest component in (b). The resonance frequency ($\Omega$) of $^1\text{H}$ and $^{15}\text{N}$ nuclei, and the scalar coupled components ($\Omega \pm \pi/\hbar_{\text{HN}}$) are indicated in each case. 1D cross sections for each grey dashed line are also shown.
presence of such couplings undermines the overall benefit of the TROSY selection principle. Therefore TROSY works most efficiently in fully deuterated, amide-protonated proteins where the homonuclear dipolar couplings between amide and other backbone and side chain protons are effectively suppressed (Pervushin et al. 1997, Salzmann et al. 2000).

Existing ‘out-and-back’ experiments are relatively easily converted to employ TROSY selection. Use of this principle has been demonstrated to enable the resonance assignment of several large proteins (Salzmann et al. 1998; Yang and Kay 1999b; Pervushin 2000). Wüthrich and co-workers have reported backbone assignments from several high molecular weight proteins including a 110 kD homooolctameric protein (Salzmann et al. 2000) and the bacterial membrane protein OmpX solubilised in dihexanoyl phosphatidylcholine micelles (Fernandez et al. 2001), which is a complex of protein in lipid with an effective molecular weight of around 60 kD. Kay and colleagues have developed a series of four-dimensional (4D) TROSY-based pulse sequences that are extensions of existing HNC-type sequences (Yang and Kay 1999b; Konrat et al. 1999) that have further served to reduce the problem of signal overlap. These pulse sequences employ a variation of TROSY selection where both of the $^{15}$N multiplet components can be kept (Yang and Kay 1999a). In sufficiently large proteins and at higher magnetic field strengths, the unwanted fast decaying component is lost through passive relaxation. This scheme can result in a 10 to 20 % sensitivity increase over the originally proposed sequence based upon active selection. The pulse sequences proposed by the Kay group have been successfully applied to some very large proteins including a 42 kD complex between MBP and β-cyclodextrin at both 37°C and 4°C (Yang and Kay 1999b), a 67 kD dimeric fragment of p53 (Mulder et al. 2000) and to 81.4 kD MSG (Tugarinov et al. 2002).

1.2.3 – Development of NMR instrumentation

The intrinsic sensitivity of a NMR signal is dependent on the energy difference between the two population levels (Cavanagh et al. 1995). When a sample of spin-$\frac{1}{2}$ nuclei is placed into a magnetic field the nuclear magnetic moments can adopt two orientations with respect to the direction of the external magnetic field $B_0$: parallel ($\alpha$
or ↑) and antiparallel (β or ↓). The energy difference between these two states is proportional to $B_0$ (Equation 1.2).

\[ \Delta E = \frac{h \gamma B_0}{2\pi} \]  

Equation 1.2

where $h$ is Planck's constant ($6.626 \times 10^{-34}$ J s) and $\gamma$ is the gyromagnetic ratio of the nucleus in question (see Table 1.1). The frequency of radiation, $\nu$ (Hz), required to induce a transition from the lower to upper level is thus $\gamma B_0/2\pi$. Because for NMR $\Delta E$ is modest, the ratio of the $\alpha$ and $\beta$ state populations is small (approximately $1 - 10^{-5}$) meaning that only a tiny fraction of the nuclei present in a macroscopic sample are available for excitation. One way of increasing the sensitivity of NMR is to simply increase the applied magnetic field. Over the past 15 years the maximum available spectrometer magnetic field strength (in $^1$H resonance frequency) has more than doubled. The largest static field currently commercially available is 900 MHz. Although higher values of $B_0$ improve polarisation, this improvement is undermined by the effect of CSA. For conventional spectroscopy the CSA contribution to transverse relaxation scales with the square of $B_0$ ($B_0^2$) leading to increases in signal linewidths at higher magnetic field strengths. This effect can be very dramatic in nuclei whose transverse relaxation is dominated by the CSA interaction, e.g. the $^{13}$C nucleus in the carbonyl group.

An interesting feature of the TROSY phenomenon (see Section 1.2.2) is that there is a theoretical field strength at which suppression of transverse relaxation of the sharpest component is predicted to be optimal (Pervushin et al. 1997). This results from the fact that the CSA interaction is dependent on $B_0^2$ but the dipolar interaction is not (see
Chapter 4, Section 4.2.1 for further discussion). Therefore at the optimal value of $B_0$, the two relaxation mechanisms effectively cancel each other out for the sharp component. In reality it is unlikely that the levels of interference between dipolar and CSA mediated relaxation will cancel completely for the multiplet component at $(\omega_H + \pi J_{NH}, \omega_N - \pi J_{NH})$. There are several explanations for this. First, the theory assumes that the unique axis of the $^{15}$N CSA tensor is co-linear with the NH bond and equal for all NH groups (Wüthrich 1998). Both assumptions have been shown experimentally not to be valid (Fushman et al. 1998). Second, in theoretical calculations of the linewidths obtained using the TROSY principle, the NH group is considered as an isolated entity. In reality this is not true (see Chapter 4, Section 4.5 for further discussion). Third, theoretical calculations show that the optimum magnetic field strength for $^1$H and $^{15}$N nuclei is slightly different (Wüthrich 1998). Overall, for $^1$H–$^{15}$N groups the theoretical optimum value of $B_0$ is expected to correspond to a proton resonance frequency of 950 to 1050 MHz (Pervushin 2000). NMR spectrometers with this size of magnetic field are not currently available so these theoretical estimations have yet to be experimentally verified.

1.3. Summary

The problems associated with NMR analysis of large proteins discussed in Section 1.1 are extensive resonance signal overlap (Section 1.1.1), increasing signal linewidth with molecular weight (Section 1.1.2) and the effects of relaxation of transverse magnetisation during coherence transfer (Sections 1.1.3 and 1.1.4). The enrichment of proteins with $^{13}$C and $^{15}$N isotopes allows additional frequency dimensions to be recorded, to some extent resolving the signal overlap problem. Two methods have increased the signal longevity of $^{13}$C and $^{15}$N transverse magnetisation. Deuteration of proteins can slow the otherwise prohibitive rate at which $^{13}$C transverse magnetisation relaxes. In the TROSY experiment the slowest relaxing component of the NH multiplet is selected giving great improvement in linewidth for larger proteins. Other notable advances in experimental technique and machine specification are the CRINEPT (cross-relaxation enhanced polarisation transfer) experiment of Wüthrich and co-workers (Riek et al. 1999) and the advent of cryogenically-cooled probes. The scope of this review, however, was limited to those advances in NMR that will be employed in this thesis.
1.4. An Outline of this Thesis.

Dimethylarginine dimethylaminohydrolase (DDAH) from *Pseudomonas aeruginosa* (PaDDAH) is a bacterial homologue of mammalian DDAH enzymes. As will be discussed in greater detail in Chapter 2, mammalian DDAH isoforms regulate the cellular levels of asymmetrically methylated arginines, which are themselves inhibitors of nitric oxide synthases. The 3D structure of PaDDAH was recently elucidated by X-ray crystallography and was shown to be a homodimer of approximately 60 kD (Murray-Rust *et al.* 2001). It was therefore proposed that NMR-based methods could be used to characterise the molecular dynamics and ligand-binding properties of PaDDAH. Because the molecular weight of PaDDAH is high, many of the problems facing NMR (e.g. low S/N and extensive signal overlap) that were described earlier in this chapter were encountered. This thesis describes how these problems were surmounted.

In Chapter 4, a model protein system will be established that mimics the solution behaviour of high molecular weight proteins. $^{15}$N nuclear spin relaxation rates will be used to describe the tumbling of $[^2H, ^{15}N]$-labelled ubiquitin in 50 % glycerol. A range of isotropic rotational correlation times was achieved by making adjustments in the sample temperature. 3D triple resonance correlation experiments using TROSY selection were tested on $[^2H, ^{13}C, ^{15}N]$-labelled ubiquitin in 50 % glycerol at 25°C. Backbone $^1H^N$, $^{13}C$ (α and β) and $^{15}N$ resonance assignments were obtained under these sample conditions, where ubiquitin is shown to have a 35 ns rotational correlation time. The utility of this system as a model for larger protein systems is discussed.

In Chapter 5, expression and purification protocols used in the manufacture of both unlabelled and multiply isotope-labelled wild-type PaDDAH will be presented. $[^1H, ^{15}N]$-HSQC and $[^1H, ^{15}N]$-TROSY spectra recorded from $[^{15}N]$-labelled WT PaDDAH will be shown which exhibit low signal intensity and low spectral resolution. The improvements and pitfalls of perdeuteration of this protein will be exemplified with relevant results. An investigation into the hydrodynamic properties and quaternary structure of WT PaDDAH is discussed, with particular focus on the
analysis of $^{15}\text{N}$ nuclear spin relaxation rates. Lastly, the difficulties faced and currently preventing unambiguous backbone resonance assignment are discussed.

In Chapter 6, evidence of a dynamic homodimer-monomer equilibrium for WT PaDDAH is presented and a dissociation constant is estimated following extensive investigations using analytical size exclusion chromatography. The 3D structure of the homodimer interface of PaDDAH is discussed in detail. A number of amino acid residues were selected for mutation. The effects of each mutation on the solubility, activity and oligomeric state of the enzyme are investigated.

In Chapter 7, the hydrodynamic properties of three PaDDAH mutants are analysed using homo- and heteronuclear NMR spectroscopy. Conclusions about the respective oligomeric states and amenability for future NMR analyses will be drawn.
Chapter II
Structure and Function of Dimethylarginine Dimethylaminohydrolase

Chapter Summary
In the first half of Chapter 2, the biology of asymmetric methylarginines will be discussed. This discussion will focus on the production of asymmetric methylarginines and their regulation through excretion and the action of DDAH enzymes. The interplay between asymmetric methylarginine metabolism and nitric oxide production will also be highlighted. In the second half of this chapter, the 3D structure of a bacterial DDAH enzyme will be described in detail and compared and contrasted to several proteins with functional or structural homology.

2.1. Physiology and Pathophysiology of Methylarginines.
There are only twenty amino acids that can be naturally incorporated into proteins. Post-translational modification of amino acid residues in proteins can extend the repertoire of chemical properties available. Common post-translational modifications include phosphorylation of serine, threonine or tyrosine side chain hydroxyl groups,
demethylation of arginine and lysine side chains, demethylation of methionine side chains and methylation of arginine, lysine and histidine side chains.

Methylation of proteins is a common and long established post-translational modification (Aletta et al. 1998). In certain systems the effect of methylation is analogous to protein phosphorylation and serves to modulate the activity of a protein. In motile bacteria, for example, the adaptive capacity of chemotactic responses is dictated by regulated methylation of transmembrane chemoreceptors (Djordjevic and Stock 1998). The level of methylation and hence adaptation of this system is determined by the antagonistic activities of a methyltransferase CheR and a methylesterase CheB. CheR transfers a methyl group from S-adenosyl-methionine (S-AdoMet) onto the γ-carboxyl group of glutamate residues at specific sites on methyl-accepting chemotaxis proteins (Levit et al. 1998). Post-translational methylation is not limited to prokaryotic organisms. In eukaroytes several proteins, including G-proteins and the catalytic domain of protein phosphatase 2A (PP2Ac), are methylated on the C-terminal α-carboxylate group (Aletta et al. 1998). PP2Ac is reversibly methylated at the carboxy-group of the conserved C-terminal residue leucine (Janssens and Goris 2001). In this system, the methylation state is determined by the opposing actions of leucine carboxyl methyltransferase 1 (De Baere et al. 1999) and protein phosphatase methylesterase I (Ogris et al. 1999). The exact function of carboxymethylation in PP2Ac remains unclear.

Protein N-methylation of L-arginine residues, as well as lysine and histidine, is also common (Aletta et al. 1998). Arginine N-methylation is observed in a variety of proteins and has been well represented in the literature. However, the exact function of this post-translational modification remains unclear. The addition of one or more methyl groups to the guanidino group of arginine residues does not alter the charge of the side chain but the modification may limit the scope for hydrogen bonding to the η-amino group(s). The enzymes responsible for protein arginine N-methylation have long been established but as yet no specialised protein arginine N-demethylase has been identified. The focus of the remainder of Section 2.1 will be the effects and results of asymmetric methylation of arginine residues.
2.1.1 - Protein Arginine N-Methylation

Three methylated derivatives of arginines are found in eukaryotic proteins: monomethylarginine (L-NMMA; Figures 2.1(b) and 2.1(f)), and symmetric dimethylarginine (SDMA; Figures 2.1(c) and 2.1(g)) and asymmetric dimethylarginine (ADMA; Figures 2.1(d) and 2.1(h)) (Leiper and Vallance 1999). In each case methylation occurs on one or both of the η-nitrogens of the guanidine group of arginine (atom nomenclature as described in Figure 2.1(a) and legend). These modifications are catalysed by a class of enzymes called protein arginine N-methyltransferases (PRMTs; Gary and Clarke 1998). Two sub-classes of PRMT are responsible for generation of methylated arginines. Type I PRMTs catalyse the creation of L-NMMA and ADMA. Substrates of Type I enzymes include heat shock proteins, and nuclear and nucleolar proteins. Type II or myelin basic protein-specific PRMT catalyses the production of SDMA and L-NMMA.

Five active Type I PRMT enzymes have been identified: four mammalian enzymes, PRMT1, PRMT3, CARM1 (or PRMT4) and PRMT6, and RMT1, which is a yeast homologue of PRMT1 (Tang et al. 2000a; Frankel et al. 2002). The original enzyme discovered was PRMT1, which was detected in a yeast two-hybrid (Y2H) screen using two proteins thought to be involved in negative regulation of cell growth, TIS12 and BTG1 (Lin et al. 1996). Utilising the amino acid sequence of PRMT1, two other PRMT enzymes have been identified (Leiper and Vallance 1999): PRMT2 from sequence similarity to PRMT1 and PRMT3, which interacts with PRMT1 in Y2H screens. Although both PRMT2 and PRMT3 enzymes show sequence homology to PRMT1 (33 % and 46 % identity respectively), only PRMT3 has been shown to have methyltransferase activity.

PRMT1, the main protein arginine N-methyltransferase in mammalian cells (Tang et al. 2000a), is a 41 kD predominantly nuclear protein that exists in large 300 to 400 kD complexes (Lin et al. 1996). Further to interactions with TIS21 and BTG1 already mentioned, PRMT1 has been shown to interact with and be regulated by interleukin enhancer-binding factor 3 (ILF3) by Y2H screens and co-immunoprecipitation. ILF3 has been proposed to function as factors in the regulation of transcription and DNA repair. Deletion of the entire arginine, serine and glycine rich C-terminal domain of ILF3 or of specific amino acids within this domain attenuates the interaction with
Figure 2.1. Chemical structure of L-arginine and \(^{N^m}\)-methylated derivatives. (a,e) L-arginine, (b,f) L-\(^{N^m}\)-monomethylarginine (L-NMMA), (c,g) asymmetric dimethylarginine (ADMA) and (d,h) symmetric dimethylarginine (SDMA). Methyl groups are represented by bold text (b-d) and pink atoms (f-h). Figures e-h generated using Rasmol. The standard greek numbering of the side chain carbon and nitrogen atoms used throughout this thesis is indicated on (a).
PRMT1 (Tang et al. 2000b). NF90, a protein with high sequence identity to ILF3 and similar proposed functions lacks the arginine, glycine and serine rich C-terminus and has been shown not to interact with or modulate the activity of PRMT1 (Tang et al. 2000b).

A third class of PRMT has been hypothesised, denoted Type III PRMT. In *Saccharomyces cerevisiae* RMT1 was believed to be solely responsible for PRMT activity. When cells lacking the RMT1 gene were supplied with $[^{14}\text{C}]$-labelled S-AdoMet an unexpected methylated arginine residue was detected (Zobel-Thropp et al. 1998). However, unlike $N^\text{o}$-methylated arginine derivatives, this compound did not decompose to form $[^{14}\text{C}]$-labelled urea in the presence of sodium hydroxide indicating that methylation of the $N^\text{o}$-nitrogen of arginine.

Type I PRMTs produce asymmetric methylarginines in many targets including signal transduction proteins, RNA processing and binding proteins, nuclear matrix proteins and cytokines (Leiper and Vallance 1999; Tang et al. 2000b). Heterogeneous nuclear ribonucleoproteins A (hnRNP A) have an arginine- and glycine-rich (RGG) C-terminal domain that facilitates RNA binding. Methylation of the RGG domain of hnRNP A2 determines the protein's cellular location. Inhibition of PRMT activity or deletion of the RGG domain localises the protein in the cytoplasm indicating that only methylated forms are found in the nucleus (Nichols et al. 2000).

2.1.2 - Guanidino-Methylated Arginines

Free methylarginines are released into the cytosol and circulatory system following proteolysis of methylated proteins (Kakimoto and Akazawa 1970). SDMA and ADMA are found at levels of 0.5 $\mu$M to 1.0 $\mu$M in the plasma of healthy patients (Leiper and Vallance 1999; Teerlink et al. 2002). Plasma levels of L-NMMA are considerably lower. No clear measurement of cellular levels of methylarginines has been established although estimates are in the region of 10 $\mu$M for ADMA (Leiper and Vallance 1999).

ADMA was originally identified in the plasma of patients suffering from chronic renal disease (Vallance et al. 1992b). In healthy individuals ADMA is excreted in the urine at a rate of 60 $\mu$mol/day (Leiper and Vallance 1999). In renal failure, however,
when urine production is almost non-existent, plasma ADMA levels are severely elevated. Second to age, plasma ADMA levels provide a good predictor of all-cause mortality in patients with end-stage renal failure (Zoccali et al. 2001). In particular, high plasma ADMA levels were linked to patient mortality related to cardiovascular complications. Elevation of plasma ADMA levels in disease is not limited to renal dysfunction, where removal of ADMA by excretion is prevented. Plasma ADMA is also found at elevated levels in hypercholesterolemia (Jang et al. 2000; Boger et al. 2000), hyperhomocysteinemia (Stuhlunger et al. 2001), atherosclerosis (Miyazaki et al. 1999), hypertension (Matsuoka et al. 1997) and insulin resistance (Stuhlunger et al. 2002). Furthermore, elevated levels of ADMA are associated with endothelial (Ito et al. 1999) and cardiovascular dysfunction (Zoccali et al. 2001), and as a risk factor for coronary disease (Valkonen et al. 2001).

2.1.3 – Nitric Oxide Synthase

Nitric Oxide (NO) is a signalling molecule that is critical for regulation of blood pressure and aspects of neuronal signal transduction and the immune response (Knowles and Moncada 1992). Endothelial cells synthesise NO in response to additions of L-arginine and L-citrulline but not D-arginine suggesting an enzyme-catalysed reaction pathway (Palmer et al. 1988). NO release from endothelial cells has been shown to relax strips of vascular tissue in vitro (Palmer et al. 1988). The enzyme responsible for catalysing this reaction is nitric oxide synthase (NOS). Three isoforms of NOS have been identified. Two of the isoforms, which are each found predominately in either neuronal (nNOS) or endothelial (eNOS) tissues, are expressed constitutively and regulated by calcium ion concentration (Titheradge 1999). Expression of the third isoform, inducible NOS (iNOS), is up-regulated by pro-inflammatory molecules and cytokines and is involved in immune system responses (Knowles and Moncada 1992).

Several observations support the theory that extracellular arginine is the preferred substrate for NO production: (1) L-arginine induced production of NO is short-lived (Palmer et al. 1988); (2) administration of L-arginine improves endothelial dysfuction, hypercholesterolemia and atherosclerosis which are all conditions associated with impaired NOS activity (Boger and Bode-Boger 2001); (3) nitrogen-15 labelled L-arginine is rapidly converted into [15N]-labelled NO (Palmer et al. 1988);
and (4) supplementation of L-arginine in vivo increases NO production (Tsikas et al. 2000; Jang et al. 2000). It is therefore counterintuitive that cellular levels of L-arginine – which have been measured at up to 2 mM (Tsikas et al. 2000) – far exceed the $K_m$ values of purified NOS isoforms, which fall in the range 1 to 10 μM (Wiesinger 2001). This situation is commonly referred to as the Arginine Paradox. Current explanations of this phenomenon favour sequestration of cellular L-arginine within intracellular reservoirs and the presence of endogenous inhibitors of NOS (Tsikas et al. 2000).

### 2.1.4 - Asymmetric Methylarginines Inhibit Nitric Oxide Synthase (NOS)

Both mono- and dimethylated asymmetric N$^\gamma$-methylarginines are endogenous inhibitors of nitric oxide synthase (Vallance et al. 1992a). SDMA is not an inhibitor of NOS. L-NMMA and ADMA have an approximately equal inhibitory effect towards NOS isoforms with IC$_{50}$ values – the concentration of inhibitor required to achieve 50% inhibition – in the range 2 to 5 μM (Leiper and Vallance 1999).

The complications associated with elevation of ADMA levels observed in many pathological conditions, most of which are associated with endothelial dysfunction, may be manifested as a result of impaired NO biosynthesis. Patients suffering from renal failure have plasma methylarginine levels that are two to six times higher than those observed in healthy patients (Zoccali et al. 2001). At these concentrations ADMA significantly affects NOS activity and vasodilation in isolated vascular rings (Vallance et al. 1992a). In hypercholesterolemia ADMA levels increase and impair angiogenesis through attenuation of NO production (Jang et al. 2000). In short, elevation of ADMA has the opposite effect of NO. Whereas NO is a vasodilatory molecule which also inhibits platelet activation and retards atherosclerosis, ADMA causes vasoconstriction and an increase in systemic vascular resistance.

Although SDMA has no direct inhibitory effect on NOS it employs the same membrane transport systems as L-arginine and asymmetric methylarginines and consequentially may affect cellular levels of these molecules (Teerlink et al. 2002; Wiesinger 2001).
2.1.5 – Metabolism of Asymmetric Methylarginines by DDAH

Two paths for the removal of asymmetric methylarginines exist. The first, as already indicated, is via excretion in the urine. The second is catabolism of asymmetric methylarginines to L-citrulline and mono- or dimethylamine by the enzyme dimethylarginine dimethylaminohydrolase (DDAH; Ogawa et al. 1989; Figures 2.2 and 2.3). Using radiolabelled substrates, Ogawa and co-workers showed first that oxygen-18 labelled water in combination with isolated DDAH produced \( ^{18}\text{O} \)-labelled L-citrulline (Ogawa et al. 1989) and later that carbon-14 labelled L-NMMA yields \( ^{14}\text{C} \)-labelled L-citrulline (MacAllister et al. 1996). The presence of an active site cysteine in DDAH was also proposed as some thiol blocking reagents inhibited the reaction (Ogawa et al. 1989).

Since the original observation of DDAH activity, two isforms of the enzyme have been discovered (Leiper et al. 1999). DDAH I, which was the originally characterised enzyme, has been purified from rat kidney (Ogawa et al. 1989; Kimoto et al. 1997), bovine brain (Fundel et al. 1996; Bogumil et al. 1998) and human liver (Kimoto et al. 1998). DDAH II was discovered by database searching using the amino acid sequence of DDAH I; the two isoforms have 62% sequence identity (Leiper et al. 1999). When expressed as a recombinant protein, DDAH II shows methylaminohydrolase activity towards ADMA but not SDMA. Investigations into the expression patterns of DDAH I and DDAH II mRNAs has shown that each isoform has distinct tissue distributions: DDAH II expression is low in brain tissues compared to the DDAH I isoform; DDAH I is only found at low levels in the heart and placenta whereas DDAH II is well expressed (Leiper et al. 1999). Furthermore, the tissue distribution of DDAH I expression is similar to that of nNOS, and that for DDAH II is similar to eNOS. Both isoforms of DDAH, when expressed with the myc-epitope as C-terminal fusion proteins in human umbilical vein endothelial cells (ECV304), localise to the cytoplasm and appear to have no specific association with the plasma membrane (Birdsey et al. 2000).

The enzymatic activity of DDAH I and II isoforms has been measured from protein purified either from tissue extracts or recombinant expression sources. When originally isolated from rat kidney the \( K_m \) of DDAH I for ADMA and L-NMMA was measured at 0.18 mM and 0.36 mM respectively (Ogawa et al. 1989). Purified DDAH
Figure 2.2. Schematic of the reactions catalysed by dimethylarginine dimethylaminohydrolase (DDAH) with (a) L-$N_m$-monomethylarginine (L-NMMA) and (b) asymmetric dimethylarginine (ADMA) as substrates.
Figure 2.3. The metabolism of asymmetric methylarginines with emphasis on interrelation with nitric oxide metabolism. Further explanation in text. Individual reactions of asymmetric methylarginines shown in figure 1.1.2. ADMA, asymmetric dimethylarginine; DDAH, dimethylarginine dimethylaminohydrolase; L-NMMA, monomethylarginine; NO, nitric oxide; NOS, nitric oxide synthase; PRMT, protein arginine methyltransferase.
I from bovine brain has been characterised as a zinc containing protein with $K_m$ values of 1.6 mM and 1.31 mM for ADMA and L-NMMA respectively (Bogumil et al. 1998). The activity of the apo form for ADMA was determined to be lower than that of the zinc-bound form ($K_m$(apo) = 0.81 mM). This difference, however, was reversed in a subsequent study where the authors claimed that the bound zinc ion had an inhibitory effect (Knipp et al. 2001). The $K_m$ of human DDAH II for L-NMMA has been determined as 0.51 mM (Leiper et al. 1999). It is believed that the intracellular concentrations of ADMA are in the range of 10 μM (Leiper and Vallance 1999).

DDAH I and DDAH II are found on chromosomes 1 (1p22) and 6 (6p21.3) respectively (Tran et al. 2000) with DDAH II located within the major histocompatibility III region (Ribas et al. 1999). A number of gene families map to these regions within chromosomes 1 and 6 and also on chromosome 9 (Tran et al. 2000). It is believed that sequential chromosomal duplication occurred from chromosome 6 to 9 and then to chromosome 1. This pattern suggests the potential existence of a third isoform of DDAH on chromosome 9. The gene structure of both isoforms is highly conserved with only small deviations in the position of intron/exon boundaries (Tran et al. 2000). However, alternative splicing at the 5’ terminus of the mRNA transcript of DDAH II can occur (Tran et al. 2000); the exact function of this phenomenon is currently unclear.

Problems associated with elevated ADMA levels in renal disease or in hypercholesterolemia or hyperhomocysteinemia, which are not associated with renal failure, show that both excretory and metabolic routes are required for the regulation of ADMA levels. There is growing evidence that the reported increase in ADMA levels in patients that do not suffer from renal disease is due to dysfunction of DDAH activity. Elevation of ADMA levels in hypercholesterolemia, hyperlipidemia and atherosclerosis is likely due to decreases in metabolic turnover of ADMA by DDAH (Ito et al. 1999). The authors of the Ito study concluded that oxidised low-density lipoprotein decreases DDAH I activity but does not alter DDAH expression levels. However, the authors used monoclonal antibodies (mAbs) raised against rat kidney DDAH – later shown to be DDAH I (Leiper et al. 1999) – to detect changes in DDAH expression in ECV304 cells. The more recently revealed existence of two isoforms of
DDAH suggests the need for future investigation. ECV304 cells are derived from human bladder where DDAH II is the predominantly expressed isoform (Tran et al. 2000). Changes in expression levels of DDAH II, which were not measured, could account for the reduced overall DDAH activity observed. Nevertheless, the effect of increases in lipid levels on DDAH activity remains an important observation.

Hyperhomocysteinemia is characterised by increased levels of homocysteine (Stuhlinger et al. 2001) that are sufficient to affect endothelium-dependent vasodilation. Increases in homocysteine are accompanied by increases in ADMA but not SDMA. Stuhlinger and co-workers showed that homocysteine inhibits DDAH activity. The resulting increase in ADMA levels would impair NO production (Stuhlinger et al. 2001). Specifically, homocysteine was suggested to form a disulphide bond to a putative active site cysteine residue of DDAH. The crystal structure of a bacterial isoform of DDAH later confirmed the presence of a critical cysteine residue in the active site that could account for these observations (Murray-Rust et al. 2001). The structure of this bacterial homologue will be discussed in Section 2.2.

Disruption of the removal of asymmetric arginines by blocking excretion or DDAH activity leads to impaired NO generation. In certain conditions, NO production can be unregulated. Septic shock, for example, is a condition characterised by severe hypotension and decreased responsiveness to vasoconstrictors caused by the presence of bacterial endotoxins in the blood (Bhagat et al. 1996; Titheradge 1999). iNOS expression is upregulated as part of the immune response, which leads to uncontrolled and sustained NO production that is dependent only on substrate availability (Nava et al. 1991; Wolkow 1998). High doses of L-NMMA have been shown to overcome vasodilation and increase vascular tone in vivo (Nava et al. 1991).

The intimate involvement of DDAH in the regulation of asymmetric methylarginine levels (Figure 2.3) makes this enzyme an attractive therapeutic target. Inhibition of DDAH by a structural analogue of L-NMMA (compound 4124W) causes contraction of rat aorta in vitro, an effect that is reversed by supplementation of L-arginine (MacAllister et al. 1996). In vivo, 4124W caused a 20% reversal of bradykinin-
induced relaxation in human saphenous veins (MacAllister et al. 1996). Inhibition of DDAH leads to raised levels of ADMA but not SDMA.

There are two routes for the elimination of asymmetric methylarginine: catabolism and excretion. Impairment of either route gives rise to increases in ADMA (and to a lesser extent L-NMMA) and causes severe clinical complications. The role of DDAH in the breakdown of asymmetric methylarginine levels has been established and through this activity it appears to act as an indirect regulator of NO biosynthesis. DDAH activity has a two-fold relief on NOS inhibition. First, DDAH breaks down methylarginine NOS inhibitors. Second, the product of the reaction catalysed by DDAH is L-citrulline, which can be metabolically recycled to produce further L-arginine NOS substrate (Figure 2.3; Wiesinger 2001; Boger and Bode-Boger 2001).

DDAH isoforms I and II have now been identified in the genomes of several higher eukaryotic organisms allowing some elucidation of the enzyme’s evolutionary origins (Tran et al. 2000). Interestingly, DDAH isoforms have also been reported in bacterial genomes that lack NOS (Santa Maria et al. 1999). This raises a question about the function of free methylarginines in bacteria and maybe the possibility that asymmetric methylarginines predate NO. Recently, the 3D structure of a bacterial DDAH was reported (Murray-Rust et al. 2001). The second half of this chapter will discuss the structure of this bacterial homologue and how the structure mediates its function.

2.2. Structure-Function relationship of PaDDAH.

2.2.1 – Global description of the structure of PaDDAH

The crystal structure of dimethylarginine dimethylaminohydrolase (DDAH) from Pseudomonas aeruginosa (PaDDAH; Murray-Rust et al. 2001) reveals five modules arranged in a circular fashion related by pseudo five-fold rotational symmetry. The fold architecture resembles a pentagonal prism (Figures 2.4(a) and 2.4(b)). Each module consists of a three-stranded mixed β-sheet with an α-helix inserted between β-strands 2 and 3. This motif is thus described as a ββαβ motif (Figures 2.4(c) to 2.4(f)). In modules 2 to 5 the N-terminal β-strand of the mixed β-sheet forms a central
Figure 2.4. 3D structure of *Pseudomonas aeruginosa* dimethylarginine dimethylaminohydrolase (PaDDAH; PDB code: 1h70; Murray-Rust *et al.*, 2001). (a) Side view of the 3D structure of a single protomer of PaDDAH with modules coloured individually (Module 1: residues 1-67 and 249-254, dark red; 2: 68-117, green; 3: 118-166, dark blue; 4: 167-205, red; 5: 206-248, light blue). (b) Plan view of a single subunit of PaDDAH coloured as (a). 3D structure (c) and topology (d) of module 2 coloured as (a). 3D structure (e) and topology (f) of module 1. Residues 249-254, which include strand-β1A, have been coloured a different shade of red in (e) and (f). The orientations of modules 1 and 2 in (c) and (e) are not the same as those in (a) and (b). 3D structural images created using MOLSCRIPT (Kraulis 1991).
channel and the β-sheet progresses to the C-terminal β-strand at the perimeter. Module 1 is a slight exception as it is composed of residues 249-254 and 1-67. The centre-most β-strand of module 1, which lines the central channel, is formed by residues 252 to 254 while the remainder of the sheet is composed from two β-strands situated at the N-terminus of the amino acid sequence (4-8 and 48-52). This method of ‘closing the chain’ is often termed ‘Velcro closure’ and is commonly seen in β-propeller proteins. A description of the β-propeller fold topology and a comparison between the structure of PaDDAH and β-propellers are presented in Sections 2.2.4 and 2.2.5 respectively. Hereafter the annotation of secondary structure will follow a scheme employed to describe β-propellers: the modules will be numbered 1 to 5 from the N-terminus and the β-strands labelled A to C from the innermost β-strand. Thus residues 4-8 and 28-42 constitute strand-β1B and helix-α1 respectively (Figures 2.4(e) and 2.4(f)).

PaDDAH exists as a homodimeric protein in the crystallographic unit cell. Evidence that this oligomeric state occurs in solution was provided by size exclusion chromatography (SEC) and dynamic light scattering results, with the observation being true for both apo- and ligand-bound forms (Murray-Rust et al. 2001). The oligomerisation interface is located on the side of the enzyme, between helices α1 and α2 and forming an extension of the β-sheet in module 1 across the interface. A more detailed description of the structure of this interface is given in Chapter 6 Section 6.3.1.

2.2.2–Protein sequences and structures that are homologous to PaDDAH
Several authors have published sequence alignments between PaDDAH, eukaryotic DDAH isoforms and homologous proteins (Murray-Rust et al. 2001; Leiper et al. 1999; Santa Maria et al. 1999). There is moderate sequence similarity between bacterial and human isoforms of DDAH; PaDDAH exhibits 44% sequence homology to human DDAH (hDDAH) over 260 residues (Santa Maria et al. 1999). PSI-BLAST (Altschul et al. 1997) sequence searches using either bacterial or mammalian DDAH sequences identify two further families of related enzymes that are involved in arginine metabolism: arginine amidinotransferases (EC 2.1.4) and arginine deiminases (Figure.
Figure 2.5. (a) Secondary structure-based alignment of amino acids sequences from *P. aeruginosa* dimethylarginine dimethylaminohydrolase (PaDDAH), human L-arginine:glycine amidinotransferase (AT) and L-arginine:inosamine-phosphate amidinotransferase (StrBl) with structural annotations showing ββββ motifs. Module 1 is coloured navy; module 2, green; module 3, red; module 4, blue; and module 5, orange (adapted from Murray-Rust et al., 2001). (b) Diagram showing the scissile bonds in L-arginine substrates cleaved (red bar) in the reactions catalysed by (clockwise from top left) DDAH, arginine deiminases and amidinotransferases AT and StrBl.
2.5(a)). Both these classes of enzymes also catalyse reactions in which the guanidino group of arginine is modified (Figure 2.5(b)).

L-arginine:glycine amidinotransferase (AGAT; EC 2.1.4.1) catalyses the transfer of an amidino group from L-arginine to glycine. This reaction is an important step in creatine biosynthesis. Under acidic conditions a covalent S-amidinocysteine intermediate can be observed in this reaction (Humm et al. 1997a) suggesting the presence of a reduced cysteine in the active site. \[^{14}C\] -labelling techniques in concert with pepsin cleavage failed to identify the specific cysteine residue involved. A mutation based approach revealed that Cys-407 was crucial for enzyme activity (Humm et al. 1997a). The subsequently determined 3D crystal structure of the wild-type (WT) enzyme complexed with L-ornithine, a product inhibitor, and the structure of a mutant enzyme (C407A) bound to L-arginine placed the guanidino group of the arginine substrate proximate to Cys-407. From these structures, the authors proposed a two-stage nucleophilic attack based reaction mechanism (Figure 2.6(a)). The structure depicts a catalytic triad comprising residues Asp-254, His-303 and Cys-407. A network of residues within the active site orient the guanidino group of the substrate so that the \(\gamma\)-sulphur atom of Cys-407, the guanidino \(\zeta\)-carbon atom and the \(\delta_1\)-nitrogen of His-303 are orthogonal to the plane of the guanidino group. On binding of a substrate molecule, the \(e\)-nitrogen lone-pair of the substrate co-ordinates the \(S\gamma\) proton of Cys-407. The \(S\gamma\) lone pair attacks the electrophilic \(C\zeta\) centre of the substrate producing a tetrahedral, amidino-cysteine intermediate. This intermediate is stabilised at low pH (Humm et al. 1997a). Cleavage of the arginine \(C^5 - N^e\) bond releases L-ornithine and leaves a covalently bound thio-imino group. In the second stage, glycine enters the active site and the proton from the \(\alpha-NH_3^+\) group is donated to the \(N\delta\) of His-303. The \(\alpha-NH_2\) electron lone pair attacks the amidino carbon atom of the tetrahedral adduct producing a guanidinoacetate intermediate. The \(S\gamma\) lone pair attacks the \(N_eH\) proton causing scission of the \(S^7 - C\) bond, release of the product, guanidinoacetic acid and recovery of Cys-407.

The alignment of the amino acid sequence of AGAT with another amidinotransferase, L-arginine:inosamine-phosphate amidinotransferase (StrB1; EC 2.1.4.2), suggested a similar reaction mechanism and this hypothesis was reinforced on the solution of the
Figure 2.6. (a) Proposed reaction mechanism for L-Arginine;Glycine amidinotransferase (from Shirai et al, 2001), (b) Proposed reaction mechanism for dimethylarginine dimethylaminohydrolase (from Murray-Rust et al, 2001)
enzymes 3D structure (Fritsche et al. 1998). Again a catalytic triad comprising
cysteine, histidine and aspartate residues, and a comparable hydrogen bonding
network was observed. Further to exhibiting functional homology, the structures of
both AGAT and StrB1 show the same ββαβ modular array with pseudo five-fold
symmetry as that seen in the structure of PaDDAH. Both amidinotransferases employ
the same 'Velcro' closure arrangement seen in PaDDAH. The root-mean-squared
deviation of core structural elements between AGAT and PaDDAH is 1.8 Å for 216
equivalent Cα atoms (Murray-Rust et al. 2001). In all three structures the array of
ββαβ motifs acts as a scaffold with the residues involved in catalysis and ligand
binding located on variable loop regions.

Two classes of arginine deiminases exist which have either free arginine or peptidyl-
arginine substrates. Arginine deiminase (ADI; EC 3.5.3.6), an enzyme in the arginine
hydrolase pathway, catalyses the irreversible conversion of free L-arginine to L-
citrulline and ammonia (Knodler et al. 1998). This pathway (and thus ADIs) are
primarily confined to prokaryotic genomes. Only three primitive eukaryotes have
been shown to exhibit ADI activity: Giardia intestinalis, Trichomonas vaginalis and
Tritrichromonas fetus (Knodler et al. 1998). Peptidyl-arginine deiminase (PAD; EC
3.5.3.15) enzymes catalyse the deimination of carboxy-terminal peptidyl-arginine
residues again producing ammonia and peptidyl-citrulline. Despite catalysing similar
reactions there is only low sequence similarity between PAD and ADI families.

Shirai et al. (Shirai et al. 2001) employed PSI-BLAST and FUGUE searches using
PAD, ADI and AT amino acid sequences and proposed the existence of a superfamily
of enzymes that catalyse modification reactions of arginine guanidino groups. The
catalytic residues observed in AGAT (Asp-170, Asp-254, His-303, Asp-305 and Cys-
407) are mostly conserved across the whole of this putative superfamily. With this
conservation and the 3D structure of AGAT in mind, the authors of this survey
modelled the 3D structure of ADI and proposed a mechanism for arginine deimination
that is essentially similar to that proposed for AGAT.
2.2.3 – Description of the active site of PaDDAH

The active site of PaDDAH is located in the negatively charged, central channel of the protein which has an entrance located at the “top” of the protein (orientations as used in Murray-Rust et al. 2001; Figure 2.4). Residues 14 to 25, between strand-β1B and helix-α1, form a loop that is presumed to close over the active site upon ligand binding. The diffraction data from the apo-enzyme shows this region to be disordered suggesting the loop may be mobile in the absence of substrate. In the ligand-bound structure, however, this region shows increased diffraction order. The loop makes only one direct contact to the ligand (L-citrulline or ADMA) with a hydrogen bond between the α-amino group of the ligand and the carbonyl of Leu-18. A similar loop region, residues 298-301 (the “300-flap”), is observed in the 3D structure of AGAT. Two distinct conformations of AGAT are observed which are dependent on the nature of the bound ligand. Analogues of glycine, the amidino acceptor, stabilise the open state whereas amidino donor analogues stabilise the closed conformation (Fritsche et al. 1999). In the open state, which has the same conformation as the apo-enzyme, there is no steric clash between the carbonyl group of the ligand and Asn-300. On substrate (L-arginine) or product inhibitor (L-ornithine) binding a large structural displacement of up to 5 Å is observed in the 300-flap and also in helix-9 (Humm et al. 1997b). Thus, AGAT demonstrates induced fit ligand binding – the carboxyl and α-amino groups of the ligand clash with the side chain of Asn-300, which displaces the 300-flap and causes helix-9 to reposition. The Cα of Pro-299 is shifted by 5 Å and other Cα atoms of helix-9 by 3.3 Å.

The amino acid sequences of StrB1 and AGAT are markedly different in the region of the 300-flap and helix-9. StrB1 lacks a counterpart helix-9 resulting in a single open structure, which suggests the enzyme does not undergo conformational changes during catalysis (Fritsche et al. 1998).

The catalytic triad of PaDDAH is slightly altered when compared to that observed in AGAT. It comprises Cys-249, His-162 and a glutamate rather than an aspartate residue at position 114. The carboxylate and α-amino groups of the ligand are hydrogen-bonded to the carbonyl groups of Leu-18 and Ile-243 and the side chains of Arg-132, Arg-85 and Asp-60. Leu-18 is located in the loop that undergoes a conformational
change on substrate binding. The aliphatic part of the methylarginine substrate is protected from solvent through interactions with Phe-63 and Leu-161. The guanidino group forms hydrogen bonds to the side chain carboxylate groups of residues Asp-66 and Glu-65, correctly orientating the C⁵ – N⁶ bond with respect to the nucleophilic thiol of Cys-249. Arginine amidinotransferases and dimethylarginine dimethylaminohydrolases catalyse similar reactions (Figure 2.5(b)). The active site of AGAT is structured to achieve cleavage of the N⁵ – C⁵ bond of the arginine substrate. In DDAH it is the C⁵ – N⁶ bond that is broken and consequentially the architecture of the active site is modified. The residues that interact with asymmetric methylarginines orient the ligand such that when the position is compared to the ligand of AGAT, the methylarginine is rotated 180° about the C⁵ – N⁶ bond. The reaction method proposed for PaDDAH (Figure 2.6(b)) is initiated on proton donation from the thiol group of Cys-249 to the guanidino group of the substrate. This is followed by nucleophilic attack by the negatively charged sulphur on the C⁵ nucleus bond forming a tetrahedral thiolate-imidazoliunum ion. The intermediate, whose existence is only speculated, is broken down by hydrolysis to produce L-citrulline and methylamine.

The PaDDAH structure accounts for substrate specificity and the ability of the enzyme to differentiate between asymmetric methylarginines, which are substrates, and SDMA, which is not. The non-methylated amidino nitrogen, N⁶', of asymmetric substrates sits in a negatively charged pocket produced by the side chains of Asp-66 and Glu-65. Binding of SDMA, which has symmetric N⁶-methylation, would require placing a methyl group in this negatively charged pocket. Such an interaction would be both electrostatically and sterically unfavourable and thereby provide a mechanism to differentiate between asymmetric and symmetric N⁶-methylated arginines.

2.2.4 – Similarity of β-propellers

A major class of protein fold topology is the β-propeller (Paoli 2001b). These folds are composed of repeated, four-stranded, anti-parallel β-sheet motifs. The β-sheet motifs form modules which arrange in a circular array such that the number of modules determines the order of rotational symmetry. Each modular β-sheet is twisted such that, when viewed from above, the structure resembles a propeller. The β-sheet motif approximates to the earlier described topologies of PaDDAH, AGAT and
In β-propellers, however, the third structural element is a β-strand rather than an α-helix. In both classes of protein this third structural element fits the overall modular antiparallel topology. The CATH protein structure classification database (Orengo et al. 1997) catalogues β-propeller structures with between four and eight blades. Only one structure, however, has been elucidated with a five-bladed β-propeller: tachylectin (PDB code: 1tl2; Kairies et al. 2001). The 3D structures of β-propellers in the PDB employ several methods of chain 'closure'. In prolyl oligopeptidase (PDB code: 1qfm; Fulop et al. 1998) the modules at both N and C-termini are folded into distinct intact blades. The ring is closed through hydrophobic interactions between modules 1 and 7. Alternatively, as seen in four-bladed propellers, the N- and C-terminal modules can be linked by disulphide bonds (Fulop and Jones 1999). The ‘Velcro’ method can occur in three varieties which are classified as 1+3, 2+2 or 3+1 depending on the composition of the closure module; i.e. 1+3 describes a four stranded β-sheet comprised of one β-strand from the N-terminus and three β-strands from the C-terminus (Fulop and Jones 1999). With this classification in mind, the closure arrangement of PaDDAH, AGAT and StrB1 can be described at 3+1 (Figures 2.4(e) and 2.4(f)). Interestingly, this type of closure is also observed in tachylectin.

Many β-propellers exhibit high sequence homology between component β-modules even though there is often poor similarity between different β-propeller proteins. There are several recognised motifs that recur in β-propellers. The WD motif was first seen in the β-subunit of a heterotrimeric G-protein. Each WD motif is a tandem sequence repeat of 44 to 66 amino acids that, when aligned to the structure, is not contained entirely within a single module. Instead the motif crosses the module boundary starting with the last β-strand of one motif and completing with the first three of the next. Other prominent sequences include the Aspartate box and the kelch and YWTD motifs (Fulop and Jones 1999). The amino acid sequences of AGAT and PaDDAH show little internal sequence homology and no obvious repeated motifs conferring this structure (Paoli 2001b; Murray-Rust et al. 2001, Figure 2.5(a)). This factor may complicate accurate predictions of other examples of this fold based on sequence data alone.
2.2.5 – Classifying the PaDDAH fold

The structure of PaDDAH is described by the authors as a barrel consisting of five weakly conserved ββαβ modules whilst Huber and co-authors (Humm et al. 1997b) chose to characterise the architecture of both AGAT and StrB1 as a basket with the loops as its handles. An additional structure reported by Burley and co-workers Groft et al. 2000) is topologically homologous to AGAT and PaDDAH but exhibits no sequence or functional similarities. The structure of the ribosome anti-association factor IF6 was solved by crystallography as a result of a structural proteomics initiative whose aims include the elucidation of the structure of all protein folds. IF6 was chosen as a target as it was believed to adopt a unique fold that the authors, on solution of the structure, named a pentein. Figure 2.7 compares the core ββαβ motif structures of PaDDAH, AGAT, eIF6 and tachylectin viewed from above. Paoli (Paoli 2001a) comments that the fold of IF6 is comparable to that seen in AGAT and further proposes the name β/α-propeller. Considering the comparisons to β-propeller topologies, this represents a more accurate classification.

2.2.6 – Is DDAH a metalloenzyme?

Human DDAH-I has been suggested to contain a Zn(II) site (Fundel et al. 1996). Preparations of this enzyme in zinc-bound and apo-forms have yielded contradictory conclusions to the role of the metal. Vašák and co-workers originally stated that a novel protein with methylarginase activity purified from bovine brain tissue, later shown to be DDAH-I, contained one bound Zn²⁺ ion per polypeptide chain (Fundel et al. 1996). Further, the zinc-bound enzyme is claimed to be approximately three times more active than the metal-ion free form (Bogumil et al. 1998). However, this conclusion has more recently been reversed (Knipp et al. 2001) such that Zn(II) has been assigned an inhibitory function with the apo-enzyme accounting for dimethylargininase activity. Results from X-ray absorption spectroscopy suggest that the zinc co-ordination site contains two sulphur atoms and two nitrogen (or maybe two oxygen) atoms (Knipp et al. 2001). This arrangement is similar to Zn(II)Cys₂His₂ motifs seen in DNA-binding zinc finger domains. However, no back scattering characteristic of histidine imidazole groups was observed leaving no firm conclusion on the structure of this site. The dissociation constant for Zn(II) binding to DDAH-I
Figure 2.7. Comparison of secondary structure topologies of from *P. aeruginosa* dimethylarginine dimethylaminohydrolase (PaDDAH; Murray-Rust et al., 2001), L-arginine:glycine amidinotransferase (AGAT; Humm et al. 1997a), eukaryotic initiation factor-6 (IF6; Groft et al. 2000) and tachylectin (Kairies et al. 2001). (a) PaDDAH, pdb code: 1h70; (b) AT, 1jdw; (c) IF6, 1g61; (d) tachylectin, 1tl2. All figures drawn with MOLSCRIPT (Kraulis, 1991). The 'Velcro' module has the C-terminus strand in white with the edge coloured as the rest of the module (as labeled in (b)).
has been calculated as 4.2 nM using a fluorine-19 labelled chelator and $^{19}$F NMR (Knipp et al. 2001).

No bound zinc ion was observed in the crystal structure of PaDDAH. PaDDAH was, however, expressed as a fusion protein with a poly-histidine tag for affinity purification (Murray-Rust et al. 2001). The high concentrations of imidazole used to elute bound his-tagged proteins could have abstracted any bound cation as shown with DDAH-I and histidine derivatives (Knipp et al. 2001).

2.3. Summary.
DDAH plays an important role in regulating NO synthesis through the catalytic removal of asymmetric methylarginines. These arginine derivatives have been shown to be endogenous inhibitors of NOS isoforms. The recent elucidation of the 3D structure of a bacterial homologue of human DDAH enzymes has provided great insight into how this family of enzymes function. With the information divulged by the 3D structure of the bacterial enzyme it may be possible to design small ligands that inhibit the activity of both bacterial and human enzymes. It has already been shown that PaDDAH can be obtained in a pure form at high concentration and that it is suitable for structural analysis (Murray-Rust et al. 2001). Heteronuclear NMR spectroscopy is a technique capable of providing invaluable information on molecular dynamics of proteins that, in this example, would complement the existing 3D structure of PaDDAH. Furthermore, NMR spectroscopy is a proven tool for performing ligand binding assays and screening of compound libraries. However, the homodimer molecular weight of PaDDAH is approximately 60 kD which is considered towards the upper molecular weight limit for NMR spectroscopy (see Chapter 1). Therefore, in Chapter 4 a model system is presented that is designed to mimic the solution behaviour of high molecular weight proteins. This model will be used to test the current NMR methods aimed at larger proteins that will later be applied to PaDDAH.
Chapter III
Materials and Methods

The experimental protocols detailed in this chapter are designed to complement the explanations given elsewhere in the text. Techniques that are to be discussed in more depth in later are only briefly described in this chapter and the reader is directed to the relevant section when appropriate. Conversely, if a technique is not discussed elsewhere a more thorough description of the methodology is presented herein.

3.1. General Chemicals and Laboratory Equipment.
Apart from where stated general chemicals used in this thesis were purchased from Sigma Aldrich or BDH. All solutions were made with Elga Maxima Ultra pure water purification system.

Centrifugation steps were performed using MSE microcentaur (all microfuge tubes), Eppendorf 5810 or Heraeus Labofuge 400R (15 and 50 ml Falcon tubes) and Sorvall RC5B centrifuges. SS34 and GS3 rotors were used for centrifugation in Sorvall RC5B centrifuges.
SDS-polyacrylamide gel electrophoresis (SDS-PAGE) was performed using BioRad equipment. All SDS-polyacrylamide gels were made following published protocols (see Sambrook et al. 1989, pages 18.52 and 18.53) using BioRad self-assembly gel construction kits. Gels were digitally photographed using a UVi DOC gel documentation system.

Protein concentrations were established using absorbance at 280 nm. The extinction coefficients used were calculated from the amino acid sequence (Creighton 1992). Absorbance measurements were obtained from Varian 100Bio and Hitachi U-1800 UV/Vis spectrophotometers. For GST and WT PaDDAH, protein concentrations stated in the text refer to the concentration of protomer chains.

3.2. General Protocols.

3.2.1 – Expression vector for PaDDAH
The cDNA insert of wild-type (WT) dimethylarginine dimethylaminohydrolase from Pseudomonas aeruginosa (PaDDAH; Figure 3.1(a)) had previously been cloned into the bacterial expression plasmid pPROEX Hta (Invitrogen; Figure 3.1(b)) between the 5’ EcoRI site and the 3’ HindIII site. Induction of PaDDAH expression was achieved by supplementing the medium with 50 µg/ml isopropyl-β-D-thiogalactopyranoside (IPTG; Melford Laboratories, UK). The promoter region controlling expression of the target protein contains the lac repressor target sequence. pPROEX Hta constitutively expresses the lac repressor protein which binds the promoter region preventing expression of the recombinant protein. Induction of protein expression is achieved by adding IPTG, a non-metabolisable analogue of lactose. IPTG bound lac repressor falls off the promoter allowing transcription to occur. Thus protein expression occurs only when desired. PaDDAH was expressed as a fusion protein with a N-terminal polyhistidine affinity tag. A tobacco etch virus (TEV) protease cleavage site was located N-terminal to the PaDDAH insert to facilitate removal of the affinity tag (Figure 3.1). pPROEX Hta contains the β-lactamase gene to confer ampicillin and carbenicillin resistance.
Figure 3.1. PaDDAH was expressed from the pPROEX Hta bacterial expression plasmid (Invitrogen). (a) DNA and amino acid sequence of the multiple cloning site (MCS) of pPROEX Hta with the polyhistidine tag and TEV cleavage site indicated. The PaDDAH insert (in grey) is situated between EcoRI and HindIII restriction endonuclease cleavage sites (the excised section of the MCS has been omitted). (b) The complete pPROEX Hta expression vector showing the position of restriction endonuclease cleavage sites.
3.2.2 - Producing chemically competent cells

Competent BL21, BL21(DE3) and BL21(DE3)PlysS expression strains of *E. coli* were freshly prepared in-house. Competent *E. coli* DH5α were purchased from Invitrogen.

To manufacture competent BL21-type cells, 100 ml of Luria-Bertani (LB) medium (per litre: 10 g bacterial peptone; 5 g yeast extract; 5 g NaCl) was inoculated with non-competent BL21-type cells and grown overnight at 37°C at 200 rpm. The cells were collected by centrifugation (2000 rpm, 10 minutes) and resuspended gently in 20 ml pre-chilled, sterile 0.1 M CaCl₂ solution using a sterile colony picker. The suspension was then left on ice for ten minutes before pelleting the cells by centrifugation (1500 rpm, 5 minutes). These steps were repeated a further two times. On the final time the cell pellet was gently resuspended into 4 ml sterile, pre-chilled 0.1 M CaCl₂ supplemented with 10 % v/v glycerol. Competent cells were snap frozen in liquid nitrogen as 250 μl aliquots and stored at -80°C.

3.2.3 - Transformation of competent cells

A suitable amount of plasmid was added to 50 μl aliquots of fresh or gently thawed competent cells in pre-chilled sterile 1.5 ml microfuge tubes and left on ice for 30 minutes. The competent cell solution was then subjected to a heat shock that was dependent on cell type. For the DH5α strain cells were incubated at 37°C for 30 seconds before adding 300 μl nutrient rich SOC medium (see Sambrook *et al.* 1989, page A3). Cell growth was permitted for a further hour in a 37°C water bath before plating 100 μl of cell mixture onto a carbenicillin (Duchefa, Netherlands) selective 25 ml LB-agar plate. BL21-type cells were heat shocked for 90 seconds at 42°C before being chilled on ice for two minutes. 250 μl of SOC medium was then added and the cells allowed to grow for one hour in a 37°C water bath. 50 μl of cell mixture was aliquotted onto a carbenicillin selective 25 ml LB-agar plate. For both cell types plates were left overnight in a 37°C incubator.

3.2.4 - Large-scale expression of unlabelled PaDDAH from LB medium

10-15 colonies from a plate of BL21-type cells freshly transformed with pPROEX-Hta PaDDAH were used to inoculate 10 ml of LB medium with carbenicillin (50 μg/ml). The cultures were then incubated (37°C, 250 rpm) until the optical density at
600 nm (OD$_{600}$) had reached 0.4 absorbance units (A.U.). At this stage all 10 ml of this culture was used to further inoculate 100 ml LB-carbenicillin. The second culture was incubated (37°C, 250 rpm) until OD$_{600}$ had reached 0.4 A.U. 50 ml aliquots were then used to inoculate 500 ml batches of LB-carbenicillin. After the OD$_{600}$ had reached 0.6 to 0.8 A.U., protein expression was induced with 50 μg/ml IPTG for three hours at 30°C, 250 rpm. After the induction period cells were harvested by centrifugation (30 minutes, 4000 rpm) and stored at −20°C until required.

3.2.5 – Large-scale expression of isotope-labelled PaDDAH from minimal medium

Minimal medium (M9) was used for the expression of isotopically enriched PaDDAH (per litre: 6.5 g NaH$_2$PO$_4$; 3.0 g KH$_2$PO$_4$; 0.5 g NaCl; 1.0 g (NH$_4$)$_2$SO$_4$; 5 g glucose; 100 mg carbenicillin; 10 mg thiamine; 10 mg biotin; all filter sterilised and then supplemented with 100 μl 1 M CaCl$_2$ and 200 μl 1 M MgSO$_4$). For labelling with nitrogen-15, uniformly [^{15}N]-labelled (NH$_4$)$_2$SO$_4$ (Cambridge Isotope Ltd) was used at 1.0 g/L; for carbon-13 isotope-labelling, uniformly [^{13}C]-labelled glucose (Cambridge Isotope Ltd) at 2 g/L; and in the case of deuterium labelling the medium was made with 99.5 % $^2$H$_2$O (Goss Scientific Instruments).

10-15 colonies from a plate of BL21-type cells freshly transformed with pPROEX-Hta PaDDAH were used to inoculate 50 ml of LB medium with carbenicillin (50 μg/ml). The cultures were then incubated (37°C, 250 rpm) until OD$_{600}$ had reached 0.4 A.U. The cells were pelleted gently by centrifugation (1800 rpm, 5 minutes). The pellet was used to inoculate 100 ml of M9 medium containing the desired isotope enrichment. After the culture OD$_{600}$ had reached 0.4 A.U., a 50 ml aliquot of the culture was used to inoculate 1 L of M9. When the final culture had reached an OD$_{600}$ of 0.6 to 0.8 A.U. protein expression was induced with IPTG (50 μg/ml) and incubation continued at 30°C. The period of incubation depended on the type of isotope labelling: nitrogen-15 only, 4-6 hours; deuterium, carbon-13 and nitrogen-15, 8-12 hours. An example of the bacterial growth curves for BL21(DE3) cells in $[^2$H, $^{13}$C, $^{15}$N]-labelled M9 medium is presented in Chapter 5, Figure 5.3.

3.2.6 – Large-scale purification of PaDDAH

Cell pellets were thawed on ice and resuspended in 20 to 30 ml Buffer A (50 mM NaH$_2$PO$_4$, 300 mM NaCl, pH 8.0) and the suspension passed over a French press
Soluble protein in the lysis supernatant was mixed with pre-equilibrated (Buffer A) nickel charged nitriloacetic acid His-Bind® resin (Ni-NTA; approximately 2.5 ml resin/500 ml bacterial culture; Novagen) for one hour at 4°C. Separation of the resin from the solution was achieved by centrifugation at 2000 rpm with no brake. The resin was washed four times with Buffer A (4 × 4 column volumes; CV) and twice with Buffer A plus 10 mM imidazole (2 × 4 CV). Bound protein was then eluted by washing with Buffer B (Buffer A plus 250 mM imidazole; 5 × 4 CV). Fractions containing PaDDAH were identified by SDS-PAGE and concentrated to a volume below 5 mls. All protein concentration steps were performed with Amicon® Centriprep® and Centricon®, and Vivascience Vivaspin® and Vivaspin20® sample concentrators.

If required the polyhistidine tag was removed by cleavage with recombinant TEV protease (rTEV). rTEV was added to 1/20 the concentration of PaDDAH. The reaction mixture was then dialysed against 1 to 2 litres of Buffer A for 48 hours. The course of the reaction was monitored by SDS-PAGE. Cleaved PaDDAH was separated from the affinity tag, uncleaved PaDDAH and rTEV by a second affinity chromatography step. rTEV, which itself has a poly-histidine tag, as well as the cleaved tag and any uncleaved PaDDAH will be absorbed onto the Ni-NTA resin, while PaDDAH will be present in the flow-through. The purity of PaDDAH in the flow-through was analysed by SDS-PAGE.

His-Bind-purified samples of PaDDAH were loaded onto size exclusion chromatography (SEC) columns (16/600 home-packed Superdex 200, S200, or 16/700 home-packed Superdex-75, S75; Amersham Biosciences, UK) pre-equilibrated with 20 mM sodium phosphate buffer, pH 7.0 (100 mM NaCl). SEC columns were attached to a Pharmacia FPLC (Amersham Biosciences, UK) or a BIOcad Sprint perfusion chromatography system (Applied Biosystems) and washed at a flow rate of 1 ml/min. Protein elution was followed by absorbance at 280 nm. Fractions containing PaDDAH were confirmed by SDS-PAGE. Unless otherwise
stated in the text, NMR samples were prepared in the same buffer used for gel filtration.

This protocol is further discussed with examples of the SDS-PAGE analyses of each step of the purification protocol in Chapter 5, Section 5.1.

3.2.7 - Mutation of WT PaDDAH

Site directed mutations of the primary sequence of WT PaDDAH were created using the Quikchange system (Stratagene). Reaction protocols were obtained from the product manual. All mutation reactions used the WT PaDDAH insert located in a pPROEX-Hta expression plasmid as a template.

<table>
<thead>
<tr>
<th>Mutation</th>
<th>Direction</th>
<th>Primer Sequence (5' to 3')</th>
</tr>
</thead>
<tbody>
<tr>
<td>E33H</td>
<td>Forward</td>
<td>TACGCCAAGGCCCTGCAACAGCAACGCCTAC</td>
</tr>
<tr>
<td></td>
<td>Reverse</td>
<td>GTAGGCGTTGTGCTGCTGCAAGCTTGGGGTA</td>
</tr>
<tr>
<td>E33Q</td>
<td>Forward</td>
<td>TACGCCAAGGCCCTGCAACAGCAACGCCTAC</td>
</tr>
<tr>
<td></td>
<td>Reverse</td>
<td>GTAGGCGTTGTGCTGCTGCAAGCTTGGGGTA</td>
</tr>
<tr>
<td>N36D</td>
<td>Forward</td>
<td>CCCTGGAGCAACACGGCTGACATCGAGCCTAC</td>
</tr>
<tr>
<td></td>
<td>Reverse</td>
<td>GGATGTAGGGCTGCTGCTGCAGGGGAGG</td>
</tr>
<tr>
<td>N36H</td>
<td>Forward</td>
<td>GCCCTGGAGCAACACGGCTGACATCGAGCCTAC</td>
</tr>
<tr>
<td></td>
<td>Reverse</td>
<td>GGATGTAGGGCTGCTGCTGCAGGGGAGG</td>
</tr>
<tr>
<td>N36W</td>
<td>Forward</td>
<td>GCCCTGGAGCAACACGGCTGACATCGAGCCTAC</td>
</tr>
<tr>
<td></td>
<td>Reverse</td>
<td>GGATGTAGGGCTGCTGCTGCAGGGGAGG</td>
</tr>
<tr>
<td>R40E</td>
<td>Forward</td>
<td>GCCAAGCGCTCAGATCGAGTCTGACATCGAGCCTAC</td>
</tr>
<tr>
<td></td>
<td>Reverse</td>
<td>CGCAGGTTCTGCAAGGCTGCTGCTGCTCCAGGG</td>
</tr>
<tr>
<td>R40W</td>
<td>Forward</td>
<td>GCCAAGCGCTCAGATCGAGTCTGACATCGAGCCTAC</td>
</tr>
<tr>
<td></td>
<td>Reverse</td>
<td>CGCAGGTTCTGCAAGGCTGCTGCTGCTCCAGGG</td>
</tr>
<tr>
<td>Q43H</td>
<td>Forward</td>
<td>CCTACATCCCGCCCTTGACACCTGACCTGACCTG</td>
</tr>
<tr>
<td></td>
<td>Reverse</td>
<td>CCACGTCCAGGCTGCAAGGCTGCTGCTGCTCCAGGG</td>
</tr>
<tr>
<td>Q43R</td>
<td>Forward</td>
<td>CCTACATCCCGCCCTTGACACCTGACCTGACCTG</td>
</tr>
<tr>
<td></td>
<td>Reverse</td>
<td>CCACGTCCAGGCTGCAAGGCTGCTGCTGCTCCAGGG</td>
</tr>
<tr>
<td>R98H</td>
<td>Forward</td>
<td>GCCAAGCGCTCAGATCGAGTCTGACATCGAGCCTAC</td>
</tr>
<tr>
<td></td>
<td>Reverse</td>
<td>GCCCGGATAGAAGTTGCTGCAAGGCTGCTGCTGCTCCAGGG</td>
</tr>
<tr>
<td>R98N</td>
<td>Forward</td>
<td>GCCAAGCGCTCAGATCGAGTCTGACATCGAGCCTAC</td>
</tr>
<tr>
<td></td>
<td>Reverse</td>
<td>GCCCGGATAGAAGTTGCTGCAAGGCTGCTGCTGCTCCAGGG</td>
</tr>
</tbody>
</table>
Primers for the polymerase chain reaction (PCR) based mutation steps of the Quikchange system are detailed in Table 3.1. Oligonucleotides were obtained from MWG-Biotech AG (Ebersberg, Germany). All PCR steps were performed on a MWG Primus thermocycler.

The DNA reaction mixture was transformed into XL21-Blue supercompetent cells (Stratagene). Single colonies, showing successfully transformed cells, were used to inoculate 10 ml LB-carbenicillin medium and incubated overnight at 37°C, 250 rpm. After pelleting the cells by centrifugation (2000 rpm, 10 minutes) plasmid DNA was extracted using Qiagen miniprep kits as instructed in the product manual. DNA sequencing was performed using ABI PRISM compatible sequencing reactions as outlined in the product manual. Following the reaction, excess dye was removed using Qiagen Dye-Ex spin kits and the resultant solution sent for automated sequencing.

3.2.8 - Small-scale expression tests of WT and mutant PaDDAH

The following procedure was performed for each of the mutant constructs listed in Table 3.1, plus WT PaDDAH, an active site mutant C249S and unmodified pPROEX-Hta vector control. Each plasmid was separately transformed into competent E. coli BL21(DE3) cells. 10-15 colonies from the resulting agar plate were used to inoculate 50 ml aliquots of LB medium with carbenicillin (50 μg/ml). When the cell density reached approximately 0.8 A.U protein expression was induced for three hours at 30°C, 200 rpm. Cells were harvested by centrifugation (3500 rpm, 30 minutes), the supernatant discarded and the pellet stored at -20°C. The cell pellet was gently thawed on ice and resuspended in 10 ml Buffer A and then lysed with a French press (three passes). The cell debris was removed by centrifugation (40 minutes, 15,000 rpm). 250 μl of pre-equilibrated (Buffer A) Ni-NTA resin was added and mixed by gentle rotation for 1 hour at 4°C. After each wash the resin was separated from the solution by centrifugation (2 minutes, 2000 rpm with no brake). The resin was washed with 5 ml Buffer A and eluted in a single step with 5 ml Buffer B. Protein expression was checked by SDS-PAGE (see for example Chapter 6, Figure 6.9). DDAH activity assays were conducted on samples from the 5 ml elution with Buffer B. The above procedure was performed for WT PaDDAH, each of the PaDDAH mutants (including C249S PaDDAH) and the vector control.
For analysis of PaDDAH expression 5 μl samples of each fraction were mixed with 5 μl gel loading buffer (see Sambrook et al. 1989, pages 18.52 and 18.53) and temperature denatured at 100°C. 5 μl of the denatured sample was loaded onto a 11% tris-glycine polyacrylamide gel and analysed by SDS-PAGE (see for example Chapter 6, Figure 6.9). The resulting gel was photographed and the gel image (TIFF format) analysed using the software Scion Image® (ScionCorp, USA). Boxes from individual lanes were selected from each gel image that surrounded the PaDDAH band. The intensity (by peak area) in arbitrary units for each peak was estimated and normalised to the intensity of the WT peak.

3.2.9 – Measurement of enzymatic activity of WT and mutant PaDDAH samples
The activity of PaDDAH was assayed by colorimetric detection of L-citrulline production following the addition of L-N-monomethylarginine (L-NMMA; Sigma Aldrich) using an adaptation of the method outlined by Knipp and Vašák (Knipp and Vašák 2000).

| Table 3.2. The composition of solutions required for the L-citrulline colour-developing assay. |
|---|---|---|
| Solution | Chemical | Amount |
| 1 | Diacetylmonoxime (DAMO) | 80 mM |
| | Thiosemicarbazide | 2 mM |
| 2 | 85 % Phosphoric Acid | 3 M |
| | c. Sulphuric Acid | 6 M |
| | NH₄Fe(SO₄)₂ | 2 mM |
| COLDER | Solution A | 1 part |
| | Solution B | 3 parts |

* care should be taking when preparing solution B – consult the original publication (Knipp and Vašák 2000).

A standard curve graph to calibrate L-citrulline concentration and optical absorbance at 540 nm (A₅₄₀) was generated (see Chapter 6, Section 6.4.1). 200 μl samples of L-citrulline at 0, 1, 2.5, 5, 10, 20, 30, 40, 50, 75, 100, 200, 400, 1000, 1500 and 2000 μM were prepared (5 samples per L-citrulline concentration). 666 μl of colour
developing agent (COLDER; Table 3.2) was added and the mixture heated at 95°C for 15 minutes and A_{540} recorded.

A time course of WT DDAH activity in comparison to a negative control (elution of pPROEX-Hta as detailed in section 3.2.8) was established (see Chapter 6, Section 6.4.1). 20 µl of sample eluant was diluted to 200 µl with Buffer A and pre-incubated for five minutes at 37°C. At t = 0, 25 µl of 50 mM L-NMMA was added. At each time point 20 µl samples were taken and added to freshly aliquotted 180 µl Buffer A and 666 µl of the COLDER (Table 3.2) solution. This solution was incubated at 95°C for 15 minutes before measuring the sample absorption at 540 nm, A_{540}. Time points were recorded at 0, 1, 2, 5, 10, 15 and 20 minutes after adding substrate for WT PaDDAH and 0, 5, 10, 15 and 20 minutes after adding substrate for the negative control. The activity at each time point was recorded in triplicate.

The activity of the mutant PaDDAH samples was compared to WT in a single time point assay. A time of 10 minutes after adding substrate was chosen as the time point. At this position in the time course assay production of L-citrulline was still in the linear regime for the WT Protein. For each WT or mutant PaDDAH sample, 20 µl of the elution from small-scale purification detailed in section 3.2.8 was added to 180 µl Buffer A and pre-incubated for ten minutes. At t = 0 minutes, 25 µl of 50 mM L-NMMA was added. At t = 10 minutes, a 20 µl aliquot was removed and added to 180 µl Buffer A and 666 µl COLDER solution, both freshly aliquotted. This solution was incubated at 95°C for 15 minutes before measuring A_{540}. The recorded enzymatic activity of each sample was normalised to the estimated level of protein present (see Section 3.2.8 and Chapter 6, Section 6.4.2).

3.2.10 – Analytical size exclusion chromatography

Stock solutions of WT and mutant PaDDAH were prepared with greater than 10 mg/ml protein (when samples permitted). 100 µl samples of protein at 10, 5, 1, 0.5, 0.25, 0.1 or 0.05 mg/ml were loaded onto a pre-equilibrated (2 column volumes of 20 mM sodium phosphate buffer, pH 7.0, see Section 3.2.6; 1 ml/min flow rate) 10/30 Superdex 75 analytical gel filtration column (Amersham Biosciences) attached to a BioCAD Sprint perfusion (Applied Biosystems) or ÄKTA Purifier10 (Amersham Biosciences) chromatography system. Protein elution was followed by absorbance at
280 nm (see Chapter 6, Figure 6.1). Molecular weight standards (Gel Filtration Low Molecular Weight Calibration Kit; Amersham Biosciences) for calibration of elution volumes were loaded at 4 mg/ml in 100 µl (see Chapter 6, Figure 6.3). Samples of ovalbumin (Amersham Biosciences) and glutathione-S-transferase (GST) at the same concentrations as PaDDAH samples (in mg/ml) were used for comparison (see Chapter 6, Figure 6.2). GST samples were produced as by-products of GST-tagged fusion proteins expressed in our laboratory. For samples evaluated under reducing conditions (see Chapter 6, Section 6.2.1) the column was pre-equilibrated with 20 mM sodium phosphate buffer, pH 7.0 supplemented with 2 mM β-mercaptoethanol (βME). Samples were reduced before being applied to the column by the addition of βME to a final concentration 2 mM.

3.2.11 – Protein samples prepared by other people

Some of the samples of [15N]- and [2H,15N]-labelled WT and mutant PaDDAH were prepared by Andrew Sankar (University College London) using the protocols outlined in this thesis (see Sections 3.2.5 and 3.2.6).

[2H,15N]- and [2H, 13C, 15N]-labelled samples of ubiquitin were obtained from Andrew Sankar (University College London).

3.3. Nuclear Magnetic Resonance Spectroscopy.

NMR spectra recorded at 500 MHz and 600 MHz proton frequency were recorded in the UCL/Ludwig Institute NMR laboratory on Varian UnityPlus spectrometers equipped with triple resonance Z-axis pulse field gradient (PFG) probes with, respectively, four and three radio frequency (RF) channels. All experiments performed on UCL spectrometers had been previously adapted from the cited pulse sequences by staff members and were set up by the author alone or under guidance of Dr R Harris. Spectra obtained at 800 MHz proton frequency were recorded at the MRC Biomedical NMR center, Mill Hill, London. The 800 MHz Varian spectrometer was equipped with four RF channels and single-axis pulsed field gradients and is equipped with a triple-resonance PFG probe. Experiments recorded at NIRM were performed by Dr. G. Kelly.
3.3.1 – Heteronuclear single quantum coherence experiments (HSQC)

\[^1\text{H}, \, ^{15}\text{N}\]-HSQC pulse sequences were performed with gradient coherence selection, sensitivity enhancement and a water flip back pulse (Zhang et al. 1994).

3.3.2 – Transverse relaxation optimised spectroscopy (TROSY)

\[^1\text{H}, \, ^{15}\text{N}\]-TROSY experiments encompassing WATERGATE solvent suppression were downloaded with permission from Prof. Lewis Kay at the University of Toronto, Canada. (http://pound.med.utoronto.ca/pulse_reg.html) and modified in-house by Dr R. Harris.

3.3.3 – Measuring proton transverse relaxation rates

Proton transverse relaxation rates were measured using a 1D 1-1 spin echo experiment similar to that described (Sklenář and Bax 1987). The delay period T (see Figure 5.7(a)) was set to \( \frac{1}{4A} \), where \( \Delta \) was the difference in Hz between the water signal and the amide signals. \( \Delta_A \) and \( \Delta_B \) were set at 0.1 and 2.9 ms respectively. All experiments were recorded at 500 MHz and 25°C. The data were analysed as described in Section 3.4.3. A description of the principles of this experiment is given in Chapter 5, Section 5.4.1.

3.3.4 – Measuring translational diffusion coefficients using 1D \(^1\text{H} NMR\)

Translational diffusion coefficients were extracted from 1D pulse field gradient experiments performed as described by Byrd and co-workers (Altieri et al. 1995) with the addition of a WATERGATE water suppression sequences prior to acquisition. Experiments were performed at gradient strengths between 6.5 and 51.7 G cm\(^{-1}\) and always at 500 MHz proton frequency and 25°C. The data were analysed as described in Section 3.4.4. A description of the principles of this experiment is given in Chapter 5, Section 5.4.2.

3.3.5 – Measuring \(^{15}\text{N} T_1, T_2\) and \(^{15}\text{N}\{^1\text{H}\} \) heteronuclear-NOE

\(^{15}\text{N} T_1, T_2\) and \(^{15}\text{N}\{^1\text{H}\}\) NOE experiments were based on published pulse sequences (Kay et al. 1989) and incorporated gradient coherence selection and sensitivity enhancement (Zhang et al. 1994). All relaxation delay and pre-saturation periods are detailed in Tables 3.3 and 3.4. All \(^{15}\text{N} \) relaxation data was analysed as described in
Table 3.3. Delay times for $^{15}$N relaxation experiments conducted on ubiquitin samples (see Chapter 4).

<table>
<thead>
<tr>
<th>Isotope labelled?</th>
<th>Temperature (°C)</th>
<th>Experiment</th>
<th>% Glycerol (v/v)</th>
<th>Relaxation Time (s) $^a$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$[^{15}N]$</td>
<td>25</td>
<td>$^{15}$N T₁</td>
<td>0</td>
<td>0.010 0.060 0.151 0.301 0.452 0.753 1.004 1.406 0.151</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>$^{15}$N T₂</td>
<td>0</td>
<td>0.015 0.045 0.075 0.105 0.135 0.165 0.196 0.226 0.271 0.316 0.361 0.075</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>$[^{1}H]$-$^{15}$N NOE</td>
<td>0</td>
<td>0.000 2.997 2.997 0.000</td>
</tr>
<tr>
<td>$[^{2}H,^{15}N]$</td>
<td>25</td>
<td>$^{15}$N T₁</td>
<td>0</td>
<td>0.010 0.060 0.151 0.301 0.452 0.753 1.004 1.406 0.151</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>$^{15}$N T₂</td>
<td>0</td>
<td>0.017 0.050 0.083 0.116 0.149 0.182 0.215 0.248 0.298 0.347 0.397</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>$[^{1}H]$-$^{15}$N NOE</td>
<td>0</td>
<td>0.000 2.997</td>
</tr>
</tbody>
</table>

|                  |                  |            |                  |                          |
|                  | 15               | $^{15}$N T₁ | 50               | 0.126 0.251 0.502 1.506 2.008 3.012 4.016 0.502 |
|                  | 15               | $^{15}$N T₂ | 50               | 0.004 0.008 0.012 0.016 0.023 0.031 0.039 0.047 0.070 0.016 |
|                  | 20               | $^{15}$N T₁ | 50               | 0.126 0.251 0.502 0.753 1.004 1.506 2.008 3.012 4.016 0.502 1.004 3.012 |
|                  | 20               | $^{15}$N T₂ | 50               | 0.004 0.008 0.016 0.023 0.031 0.039 0.047 0.070 0.093 0.016 |
|                  | 25               | $^{15}$N T₁ | 50               | 0.126 0.251 0.502 0.753 1.004 1.506 2.008 3.012 4.016 0.502 |
|                  | 25               | $^{15}$N T₂ | 50               | 0.004 0.008 0.016 0.023 0.031 0.039 0.047 0.070 0.097 0.136 0.016 |
|                  | 25               | $[^{1}H]$-$^{15}$N NOE | 50 | 0.000 2.997 2.997 0.000 |
|                  | 30               | $^{15}$N T₁ | 50               | 0.126 0.251 0.502 0.753 1.004 1.506 2.008 3.012 4.016 0.502 |
|                  | 30               | $^{15}$N T₂ | 50               | 0.004 0.008 0.016 0.023 0.031 0.039 0.047 0.070 0.097 0.136 0.016 |
|                  | 35               | $^{15}$N T₁ | 50               | 0.126 0.251 0.502 0.753 1.004 1.506 2.008 2.510 3.012 0.502 |
|                  | 35               | $^{15}$N T₂ | 50               | 0.004 0.008 0.016 0.023 0.031 0.047 0.051 0.093 0.117 0.016 |
|                  | 35               | $[^{1}H]$-$^{15}$N NOE | 50 | 0.000 2.997 0.000 2.997 |

$^a$ In $^{15}$N T₁ and $^{15}$N T₂ experiments, the time periods stated refer to the length of time permitted for relaxation in the pulse sequence. In the case of $[^{1}H]$-$^{15}$N NOE experiments the time stated refers to the length of the proton saturation period.
Table 3.4. Delay times for $^{15}$N relaxation experiments conducted on PaDDAH samples (see Chapters 5 and 7)

<table>
<thead>
<tr>
<th>Sample</th>
<th>Temperature (°C)</th>
<th>Experiment</th>
<th>Relaxation Delay Period (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$^{15}$N T$_1$</td>
<td>0.000 0.126 0.251 0.502 0.753 1.004 1.255 1.506 1.757 2.008 2.510</td>
</tr>
<tr>
<td>WT</td>
<td>25</td>
<td>$^{15}$N T$_2$</td>
<td>0.000 0.008 0.012 0.016 0.023 0.031 0.051 0.070 0.098 0.125</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$^{1}$H-$^{15}$N NOE</td>
<td>0.000 3.011</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>$^{15}$N T$_2$</td>
<td>0.000 0.008 0.012 0.016 0.023 0.031 0.051 0.070 0.098 0.125</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>$^{15}$N T$_2$</td>
<td>0.000 0.008 0.012 0.016 0.023 0.031 0.051 0.070 0.098 0.125</td>
</tr>
<tr>
<td></td>
<td>35</td>
<td>$^{15}$N T$_2$</td>
<td>0.000 0.008 0.012 0.016 0.023 0.031 0.051 0.070 0.098 0.125</td>
</tr>
</tbody>
</table>

|        |                  | $^{15}$N T$_1$     | 0.126 0.502 1.004 1.506 2.008 2.510 3.012 4.016 0.502 |
| N36W   | 25               | $^{15}$N T$_2$     | 0.004 0.008 0.016 0.023 0.031 0.039 0.047 0.062 0.078 0.097 0.031 |
|        |                  | $^{1}$H-$^{15}$N NOE | 0.000 2.997                  |
| R40E   | 25               | $^{15}$N T$_1$     | 0.126 0.251 0.502 1.004 2.008 3.012 0.251 |
|        | 25               | $^{15}$N T$_2$     | 0.004 0.008 0.016 0.031 0.047 0.070 0.078 0.124 0.016 |
|        |                  | $^{1}$H-$^{15}$N NOE | 0.000 2.997                  |

*In $^{15}$N T$_1$ and $^{15}$N T$_2$ experiments, the time periods stated refer to the length of time permitted for relaxation in the pulse sequence. In the case of $^{1}$H-$^{15}$N NOE experiments the time stated refers to the length of the proton saturation period.*
Section 3.4.2. A brief description of the theory of NMR relaxation and the principles of these experiments are given in Chapter 4, Section 4.2.1.

3.3.6 – NMR experiments recorded to facilitate sequence-specific resonance assignment

All 3D triple resonance pulse sequences used (see Chapter 4, Table 4.2 and Chapter 5, Table 5.3) were based on the sequences published by Kay and co-workers and employed TROSY selection in $^{15}$N and $^1$H dimensions (Yang and Kay 1999a; Yang and Kay 1999b). In the case of WT PaDDAH, active suppression of the undesired $^{15}$NH cross peak components was performed (Yang and Kay 1999b) due to the presence of intense signals from the highly flexible polyhistidine purification tag. The length of the recycle delay (ideally ca. 5 seconds for deuterated proteins) was optimised for experiment time against the numbers of scans acquired. All experiments were recorded at 600 MHz and the temperatures stated in the text.

3.3.7 – 3D $[^1]H$, $^{15}N$ NOESY HSQC

A 3D $[^1]H$, $^{15}N$ NOESY HSQC experiment (Zhang et al. 1994) was recorded with gradient coherence selection and sensitivity enhancement and a mixing time of 100 ms.

3.3.8 – Processing NMR data

All raw NMR data were processed using the nmrPipe program of Delaglio and co-workers (Delaglio et al. 1995). Standard manipulations of data sets included zero-filling to the nearest $2^n$ points, application of window functions, base line corrections, linear prediction of indirect dimensions where necessary, and first and second order phase corrections. The precise processing steps varied between data sets. Spectra were initially visualised in nmrDraw (Delaglio et al. 1995).

Following processing, spectra were exported into AZARA (Boucher 2002) format using the PIPE2AZARA command. Multiple spectra were visualised and cross-referenced in Plot2 (Boucher 2002). Sequential assignment was performed using the ANSIG analysis package (Kraulis 1991) with contour and cross peak files generated by AZARA.
3.3.9 – Estimation of amide proton solvent exchange rates

A $[^1H, ^{15}N]$-TROSY spectrum (see Section 3.3.2; see Figure 5.6 for experiment details) was recorded from a 600 μl, 1 mM sample of $[^{15}N]$-labelled WT PaDDAH in 20 mM sodium phosphate buffer, pH 7.0, 100mM NaCl, 10 % D$_2$O. Following acquisition of the spectrum, the sample was diluted 1:10 with 20 mM sodium phosphate buffer, pH 7.0, 100 mM NaCl prepared in D$_2$O and concentrated to 600 μl. An identical $[^1H, ^{15}N]$-TROSY spectrum with identical parameters was recorded. Again, following acquisition of the spectrum, the sample was diluted 1:10 with 20 mM sodium phosphate buffer, pH 7.0, 100 mM NaCl prepared in D$_2$O an concentrated to 600 μl and a final identical $[^1H, ^{15}N]$-TROSY spectrum recorded.

3.4. Analysis of Data.

3.4.1 – Calculating signal-to-noise (S/N) ratios

S/N ratios for amide cross peaks in $[^1H, ^{15}N]$-HSQC and $[^1H, ^{15}N]$-TROSY spectra were measured as follows. For each selected amide cross peak, a signal intensity value was extracted using ANSIG scripts written in-house (Pfuhl et al. 1999). For spectra of ubiquitin, NH cross peaks were assigned in a residue specific fashion. In the case of PaDDAH, selected NH cross peaks were assigned numerically. The signal intensity values and corresponding peak ‘assignments’ were output as ASCII text files. Noise values were calculated using Plot2 (Boucher 2002) from a region of the spectrum devoid of cross peaks.

S/N ratios were initially analysed for all selected cross peaks. The mean and standard deviation of this data set are strongly biased by the cross peaks from extremely flexible NH groups, such as those at the polypeptide termini, which had extraordinarily high S/N values. The data was, therefore, re-analysed using a standard deviation filter to omit any cross peak with a S/N value outside the original mean plus or minus 1.5 times the original standard deviation. These latter statistics have been presented throughout.

3.4.2 – Derivation of $^{15}N R_1, ^{15}N R_2, ^{15}N T_1 \& ^{15}N T_2$ heteronuclear-NOE and estimation of $\tau_c$

$^{15}N T_1$ and $^{15}N T_2$ experiments were acquired with incremented relaxation delay periods (see Section 3.3.5). Effectively for both experiments, a 2D NH correlation
spectrum was recorded for each relaxation delay period (Tables 3.3 and 3.4). $^{15}\text{N}$ $T_1$ and $^{15}\text{N}$ $T_2$ spectra could therefore be processed as pseudo 3D-spectra with the third dimension comprising 2D correlation planes with incremented relaxation delays. For each set of relaxation data a reference 2D $[^1\text{H}, ^{15}\text{N}]$-correlation spectrum was recorded; either $[^1\text{H}, ^{15}\text{N}]$-HSQC or $[^1\text{H}, ^{15}\text{N}]$-TROSY experiments depending on the sample being studied. The relaxation spectra were processed with an artificial third dimension consisting of a single point per relaxation delay increment using nmrPipe. In the analysis of $^{15}\text{N}$ relaxation data from ubiquitin samples, residue specific assignment of amide cross peaks in the 2D NH reference spectrum were available. For PaDDAH, selected NH cross peaks were assigned numerically. Only well-resolved cross peaks were selected for relaxation analysis. For each selected amide cross peak in the reference spectrum a signal intensity value was extracted from each of the 2D planes in the pseudo 3D relaxation spectrum using ANSIG scripts written in-house (Pfuhl et al. 1999). These values and corresponding peak assignments were output as text files. To measure the $[^1\text{H}, ^{15}\text{N}]$ heteronuclear NOE, 2D NH correlation spectra were recorded in the presence and absence of proton saturation (see Tables 3.3 and 3.4). The data sets could there be processed as a pseudo 3D experiments. Thus 3D spectra had two or four planes in the third dimension depending on the number of repeated 2D spectra. Signal intensity values were extracted as described for $T_1$ and $T_2$ spectra and output as text files. Signal intensity values for all relaxation experiments were corrected for noise, which was calculated using Plot2 (Boucher 2002) from a region of the spectrum devoid of cross peaks.

For each selected NH cross peak in the $T_1$ or $T_2$ spectrum, the decay of magnetisation was fit to Equation 4.2 using Mathematica® (Wolfram 2002). The cross peak signal intensity and relaxation delay time data sets were fit against Equation 4.2 (see Chapter 4, Section 4.2.1) using a two parameter ($I_0$, the intial signal intensity, and $R_i$, the relaxation rate constant where $i = 1$ or 2) least squares regression minimisation (Levenberg-Marquardt method; as described in Kay et al. 1992; Peng and Wagner 1994). A standard deviation for each experimental signal was calculated from the difference between two separate experiments with identical relaxation delay values or, in the absence of repeated data points, the spectral noise value. The uncertainty in the fit for $I_0$ and $R_i$ was determined using a Monte Carlo simulation. For each Monte Carlo iteration, a random signal intensity for each relaxation delay time was generated
from a normal distribution about the experimental mean signal intensity for that relaxation delay. Each simulated data set was separately used to optimise Equation 4.2 (see Chapter 4, Section 4.2.1), using the protocol described above for the initial fit. The relaxation rates and standard deviations stated in text correspond to the mean relaxation rate and standard deviation of the distribution of 200 such Monte Carlo iterations.

The \({^1\text{H}}-{^{15}\text{N}}\) heteronuclear NOE value for each selected NH cross peak was the ratio of noise-corrected signal intensities from 2D \({^1\text{H}}-{^{15}\text{N}}\) heteronuclear correlation spectra with and without proton saturation. If repeated experiments were conducted the ratio was derived using the mean of the saturated and non-saturated signal intensities. Uncertainties in the estimation of the NOE values were propagated from the experimental reproducibility of saturated and non-saturated signal intensity values. In the absence of repeated experiments, the error was calculated from the spectral noise.

The isotropic rotational correlation time, \(\tau_c\), for a protein was derived from a \(R_2/R_1\) data set which had been filtered for internal mobility and chemical exchange contributions to \(R_1, R_2\) and the \({^1\text{H}}-{^{15}\text{N}}\) heteronuclear NOE values (described in Chapter 4, Section 4.2.1). A Monte Carlo procedure using the experimental mean \(R_2/R_1\) ratio and standard deviation of the filtered data set was used to generate a simulated distribution of 1000 \(R_2/R_1\) ratios. \(\tau_c\) values were determined (see Chapter 4, Section 4.2.1) for each simulated \(R_2/R_1\) ratio (see Chapter 4, Equation 4.4). The values of \(\tau_c\) stated in the text refer to the mean and standard deviation of the resultant distribution of artificial \(\tau_c\) values.

3.4.3 - Calculating proton transverse relaxation rates

The signal intensity in a selected region in the 1D spectrum of a 1-1 spin echo experiment was extracted using nmrPipe scripts written in-house and output as ASCII text files. Signals in the amide proton region of the 1D proton spectrum were selected. Five identical experiments were recorded per sample and the proton \(T_2\) calculated in each case using Equation 5.1 (see Chapter 5, Section 5.4.1). \({^1\text{H}}\) \(T_2\) values stated in the text refer to the mean and standard deviation of the repeated experiments.
3.4.4 - Calculating translational diffusion rates

Multiple 1D PSG diffusion experiments were recorded per sample with incremented gradient strengths between 6.5 and 51.7 G cm\(^{-1}\). The exact gradient strengths employed varied between experiments. The mean log(signal intensity) and standard deviation per gradient strength were fit against Equation 5.5 (see Chapter 5, Section 5.4.2) in Mathematica® using a two parameter (\(I_0\), the intial signal intensity, and \(D_z\), the translation diffusion coefficient) least squares regression minimisation (Levenberg-Marquardt method). The errors in the fit for \(I_0\) and \(D_z\) were determined using a Monte Carlo simulation. For each Monte Carlo iteration, a random signal intensity for each gradient strength was generated from a normal distribution about the experimental mean signal intensity for that gradient strength. The simulated data set was then used to optimise Equation 5.5 (see Chapter 5, Section 5.4.2), using the protocol described above for the initial fit. The values and standard deviations of \(D_z\) stated in text correspond to the mean \(D_z\) and standard deviation of the distribution of 1000 Monte Carlo iterations.

3.4.5 - Estimating the dissociation constants of WT PaDDAH from analytical SEC

The elution volume of each peak was measured using the software packages accompanying the chromatography systems used (see Section 3.2.10). 8 separate SEC experiments were run at each sample concentration. Apparent molecular weight (app. MW) values were derived from the peak elution volumes using Equation 6.1 (see Chapter 6, Section 6.2.3). The experimental data set (comprising an average app. MW and standard deviation for each of the six sample concentrations used) was used to fit against Equation 6.12 (see Chapter 6, Section 6.2.4) in Mathematica® using a three parameter (\(K_d\), the homodimer dissociation constant, \(MW_M\), the app. MW of a monomeric species and \(MW_D\), the app. MW of a homodimer) least squares regression minimisation (Levenberg-Marquardt method). Loading protein concentrations were adjusted for dilution on the column using the rationale of Manning and co-workers (Manning et al. 1996). The dilution factor for WT PaDDAH data was derived from the ratio of the average elution peak width at half-height (in ml) and the sample loading volume (always 100 \(\mu\)l).

In instances where \(MW_M\) and \(MW_D\) were defined, the error of the fit for \(K_d\) was determined using a Monte Carlo simulation. For each Monte Carlo iteration, a random
app. MW for each sample concentration was generated from a normal distribution about the experimental mean app. MW at that concentration (using the standard error of each experimental data set). The simulated data set was used to optimise Equation 6.12 (see Chapter 6, Section 6.2.4), using the protocol described above for the initial fit. The values and standard deviations of $K_d$ given in text correspond to the mean and standard deviation of the distribution of values generated from 200 Monte Carlo iterations.
Chapter IV
Studying Large Proteins by NMR: Ubiquitin in glycerol as a model

Chapter Summary
In this chapter a model that mimics the solution behaviour of larger proteins will be presented. The hydrodynamic properties of $[^2]H, ^{15}N$-labelled ubiquitin in 50% (v/v) glycerol are explored over a wide range of sample temperatures. 3D triple resonance TROSY-based experiments are recorded on a sample of $[^2]H, ^{13}C, ^{15}N$-labelled ubiquitin in 50% (v/v) glycerol at 25°C to obtain sequence-specific resonance assignments of the protein under these solution conditions.

4. Introduction.
There are several contributory factors that impose an upper molecular weight limit in the investigation of protein properties by solution NMR spectroscopy. These include rapid relaxation of transverse magnetisation, broad signal linewidths and the extent of resonance overlap (see Chapter 1, Section 1.1.2). Ultimately, the practical upper size boundary is dependent on the type of analysis that is to be performed. For example, it is possible to assess the hydrodynamic properties of proteins ranging from a few
thousand Daltons to several hundred kilodaltons using simple one-dimensional (1D) proton NMR experiments (see, for example, Chapter 5, Sections 5.3.1 and 5.3.2). However if the objective is to obtain the solution structure of a protein or characterise backbone and side chain dynamics then, invariably, this is limited to relatively small proteins. Improvements in both NMR technology and application methods are continually raising the size limit for the latter type of analysis (see Chapter 1). However, for a given project objective the number of amino acid residues in a protein and hence the number of NMR correlations in a particular spectrum – both intrinsic properties of the system being studied – can rarely be reduced. In special cases the protein can be successfully dissected into smaller modular fragments that are more amenable to NMR spectroscopy, but this is not always biologically desirable or experimentally feasible.

The subject of this project, WT PaDDAH, is reportedly a homodimeric protein with an oligomeric molecular weight of around 60 kD (Murray-Rust et al. 2001). As already discussed (see Chapter 1), there are several reports of backbone chemical shift assignments of proteins of this size and larger (Kay 2001). To date, however, there are no complete solution structures of proteins much greater than ca. 40 kD. This is most probably due to the inherent decrease in signal intensity observed for higher molecular weight species combined with the sheer complexity of the NMR spectra that are required to obtain distance constraints. An alternative explanation derives from the necessary use of deuteration for NMR experiments that enable sequence-specific backbone resonance assignments to be obtained. Although deuteration can be beneficial in terms of improving transverse relaxation rates (see Chapter 1, Section 1.2.1), the resultant absence of side chain protons dramatically reduces the number of inter-proton NOE distance restraints that can be measured. As the size of WT PaDDAH is far greater than any protein previously studied in our laboratory, a suitable model system was sought with which to test the NMR experimental and analytical processes that were intended for use in its structural characterisation.

4.1. A Model for Studying Large Proteins by NMR.

Human ubiquitin is a small, monomeric protein with a molecular weight of approximately 8.6 kD for which a crystal structure (PDB code: 1ubq; Vijay-Kumar et al. 1987) and a high definition ensemble of solution NMR structures (Cornilescu et al. 1987)
1999) have been deposited in the PDB. Ubiquitin has been extensively characterised by solution NMR spectroscopy with information on backbone and side chain resonance assignments, solution structures and dynamic properties present in the literature. Due to this extensive characterisation, ubiquitin has been frequently used as the protein subject when developing novel NMR techniques. In the present investigation human ubiquitin was prepared in a manner in order that its hydrodynamic behaviour should mimic that of larger proteins.

The rotational correlation time, $\tau_c$, of a protein in solution is affected by the viscosity of the solution as described by Stoke's Law (Equation 4.1).

$$\tau_c = \frac{4\pi \eta_w r^3}{3 k_B T}$$  \text{Equation 4.1}

where $\eta_w$ is the viscosity of solvent, $r$ the hydrodynamic radius of the molecule, $k_B$ the Boltzman constant and $T$ the sample temperature (Cavanagh et al. 1995). Increasing the viscosity of the solution, for example by preparing a sample with a high percentage of glycerol as a co-solvent, leads to increases in the correlation time of the protein solute.

Samples of isotope-labelled ubiquitin were prepared in solutions containing $[^2\mathrm{H}]$-labelled glycerol at 50% (v/v) (see 'Materials and Methods' Section 3.2.12). Deuterated glycerol was used in order to eliminate the presence of strong proton signals in NMR spectra recorded of ubiquitin. The viscosity of the solution and hence tumbling rate of the protein could be further 'tuned' by adjusting the sample temperature; a more preferential manipulation to varying the percentage of glycerol in the solution. As ubiquitin has only 76 amino acid residues, of which only 70 yield amide NH cross peaks, the problems of resonance overlap that would be inevitable when studying higher molecular weight proteins were largely circumvented. Figure 4.1 shows a comparison of NH correlation spectra recorded from $[^2\mathrm{H}, ^{15}\mathrm{N}]$-labelled ubiquitin in the absence (Figure 4.1(a)) and presence (Figure 4.1(b)) of 50% (v/v) glycerol. A clear increase in the linewidth of amide NH cross peaks was observed on the addition of glycerol. Sequential resonance assignment of the backbone amide cross peaks observed in the HSQC spectrum of $[^{15}\mathrm{N}]$-labelled ubiquitin in the absence of glycerol has previously been performed in this laboratory (Dr. R. Harris personal...
Figure 4.1. NH correlation spectra of [²H, ¹⁵N]-labelled ubiquitin in the absence and presence of 50% (v/v) glycerol. (a) [¹H, ¹⁵N]-HSQC spectrum recorded at 25°C without glycerol. The data matrix consisted of 512* (t₁) x 512* (t₂) data points (where n* refers to complex points) with acquisition times of 128 ms (t₁) and 64 ms (t₂). A total 16 scans were recorded per t₁ increment; (b) [¹H, ¹⁵N] TROSY spectrum recorded at 25°C with 50% (v/v) glycerol. The data matrix consisted of 256* (t₁) x 2560* (t₂) data points (where n* refers to complex points) with acquisition times of 64 ms (t₁) and 128 ms (t₂). A total 32 scans were recorded per t₁ increment. Both spectra were recorded at 600 MHz ¹H resonance frequency.
communication) and elsewhere (Lee and Wand 1999). The NH cross peak assignments used here for \(^{2}\text{H},^{15}\text{N}\)-labelled ubiquitin were determined using 3D triple resonance NMR methods (see Section 4.3).

Only limited chemical shift differences were noted on the superposition of a NH correlation spectrum recorded of \(^{15}\text{N}\)-labelled ubiquitin in the absence of glycerol and one recorded of \(^{2}\text{H},^{15}\text{N}\)-labelled ubiquitin in the presence of glycerol. This is a good indication that the structure of the protein has not been greatly perturbed by the presence of the co-solvent.

From Equation 4.1, adjustments in the temperature of the sample should affect the viscosity of the 50\% (v/v) glycerol solution and the tumbling rate of ubiquitin. As will be discussed in more depth in Section 4.2.1, NMR signal intensities and linewidths are dependent on the rate at which a protein tumbles in solution (see also Chapter 1, Section 1.1.2). By comparing signal-to-noise (S/N) ratios for spectra recorded at different temperatures it may be possible to make a rudimentary estimate of the change in tumbling rate of ubiquitin. Accordingly, \(^{1}\text{H},^{15}\text{N}\)-TROSY spectra were recorded from a sample of \(^{2}\text{H},^{15}\text{N}\)-labelled ubiquitin in 50\% (v/v) glycerol at six temperatures from 10.1 to 34.3°C (Figure 4.2). As the acquisition temperature was raised the distribution of signal-to-noise (S/N) ratios of NH cross peaks showed a marked shift towards higher values (see 'Materials and Methods' Section 3.4.1 for a description of calculating S/N ratios). The improvements observed in both spectral resolution and S/N between 10°C (Figure 4.2(a)) and 34.3°C (Figure 4.2(f)) were substantial; the average S/N (⟨S/N⟩) ratio increased approximately 20-fold over this temperature range. The ⟨S/N⟩ ratio calculated for \(^{2}\text{H},^{15}\text{N}\)-labelled ubiquitin at 25°C in the absence of glycerol was approximately 400 (data not shown). On initial inspection and using S/N ratios as a crude estimate, it was clear that this model system could be used to mimic a large range of higher molecular weight systems. Furthermore, the benefits of using a small protein model with a relatively uncomplicated NH correlation spectrum are immediately apparent. Even at lower temperatures, where the S/N ratio tended to be low and the linewidths very broad, a large number of non-overlapped NH cross peaks could be readily observed and quantified.
Figure 4.2. $[^1H, ^{15}N]$ TROSY and amide cross peak signal-to-noise (S/N) ratio histograms at (a) 10°C; (b) 17.5°C; (c) 21.5°C; (d, overleaf) 25.7°C; (e, overleaf) 29.9°C; and (f, overleaf) 34.3°C. $[^1H, ^{15}N]$ TROSY spectra were recorded at 600 MHz $^1$H resonance frequency. Each data matrix consisted of 256* ($t_1$) and 2560* ($t_2$) data points (where $n^*$ refers to complex data points) with acquisition times of 64 ms ($t_1$) and 128 ms ($t_2$). A total of 32 transients were recorded per $t_1$ increment. Average S/N ratios, $\langle S/N \rangle$, and standard deviations (in brackets) are given in each histogram. These parameters do not include extreme values (See ‘Materials and Methods’ Section 3.4.1 for details).
Figure 4.2. Continued.
4.2. Analysis of $^{15}$N relaxation parameters: [$^2$H, $^{15}$N]-labelled ubiquitin in 50% (v/v) glycerol as a model for larger proteins.

For a more comprehensive analysis of the hydrodynamic properties of ubiquitin in the presence of glycerol $^{15}$N nuclear relaxation parameters were measured. These parameters provide information of the rotational reorientation rate of a protein as well as details of molecular dynamics. Although a complete account of nuclear spin relaxation is outside the scope of this thesis, the key issues, particularly concerning the underlying causes of $^{15}$N relaxation, are discussed briefly below.

4.2.1 – Extracting $^{15}$N relaxation parameters from NMR data

Information on dynamic processes in proteins can be obtained by measuring the relaxation rates of nuclear magnetisation (Kay et al. 1989; Peng and Wagner 1994; Palmer 2001). Analysis of the rate at which magnetisation relaxes can report on dynamic processes occurring over a wide range of time scales encompassing such molecular motions as molecular tumbling, segmental motion and conformational exchange (Figure 4.3(a)). The characterisation of protein backbone dynamics on the pico- to nanosecond (ps - ns) time scale and rotational diffusion is commonly achieved by the extraction of amide $^{15}$N auto- and $^{15}$N–$^1$H cross-relaxation rates.

When a sample of spin-$\frac{1}{2}$ nuclei is placed into a magnetic field, $B_0$, their magnetic moments align with or against the direction of $B_0$. These two orientations have different energies and are termed Zeeman states. A magnetic moment is aligned with $B_0$ is in the lower energy Zeeman state. A $90^\circ$ RF pulse on equilibrium magnetisation effects a condition termed coherence. This term describes a situation where the populations of spins in the lower and upper Zeeman states have been equalised such that net ‘transverse’ magnetisation is aligned along an axis perpendicular to $B_0$. Such coherence represents a non-equilibrium state of the spin system and as such is an excited state in NMR spectroscopy. Transfer of energy between nuclear spins induces relaxation from this excited state allowing the system to return to equilibrium. For a spin-$\frac{1}{2}$ nucleus, the interaction between its nuclear magnetic moment and local magnetic fields fluctuating at the correct resonance frequency allows for such relaxation. Local magnetic fields can be generated by the relative motion between nuclei or currents of bonding electrons. The precessional frequency of a nuclear
Figure 4.3. (a) Approximate time and frequency ranges of biological protein motions (top) and the NMR parameters which can report on them (bottom); (b) Theoretical variation of $^{15}$N $T_1$ (---), $T_2$ (—) and $^{1}$H$^{15}$N NOE (---) with the isotropic rotational correlation time, $\tau_c$; (c) Theoretical relationship between $^{15}$N $R_2/R_1$ ratio and $\tau_c$. Definitions of $^{15}$N $T_1$, $T_2$ and heteronuclear NOE from Palmer 2001 (see also Appendix A.1). Functions evaluated for a $^1$H frequency ($\omega_H$) of 599.9 MHz and a $^{15}$N frequency ($\omega_N$) of 60.797 MHz using the spectral density function describing a spherical top and considering only isotropic rotational diffusion (see Appendix A.1, Equation A.3).
magnetic moment (i.e. the Larmor frequency, \( \omega_0 \)) is relatively slow compared to the frequencies of molecular motion, e.g. bond vibrations and rotations. Only slower molecular motions, such as molecular reorientation, occur at a directly comparable frequency to NMR transitions. Tumbling of a molecule in solution or bond motions of an appropriate frequency cause the strength and direction of local magnetic fields to constantly fluctuate.

There are two principal origins of local magnetic fields: interactions between nuclear dipoles and the interaction of the nuclear dipole with anisotropic distributions of electrons (the chemical shift anisotropy (CSA)). A full explanation of each of these concepts and how they induce relaxation is non-trivial. Only a simplified account will be presented here.

Consider two spin-\( \frac{1}{2} \) nuclei at finite separation. The magnetic moment of one nucleus affects the magnetic moment of another in a reciprocal fashion. This interaction is termed a dipolar coupling. The magnitude of the dipolar coupling is strongly dependent on the internuclear distance and is dominated by short range interactions. A dipolar interaction can occur between both bonded and non-bonded pairs of nuclei. Instantaneous dipolar couplings between two nuclei will average to zero over time if the tumbling of the molecule in solution is random with respect to the direction of \( B_0 \).

At any given time the magnetic moment of a dipolar-coupled nucleus effectively experiences a combination of magnetic fields: the external field (i.e. \( B_0 \)) and the local field due to nearby nuclear dipoles. Although the effect of the local magnetic field in comparison to \( B_0 \) is slight, a magnetic moment will reorient in order to be parallel to the combined magnetic field it experiences. Furthermore, a magnetic moment is essentially unaffected by molecular rotation and will always align parallel to the magnetic field it experiences, which is strongly dominated by \( B_0 \). An analogy commonly used to describe the phenomenon of the nuclear magnetic moment being uncoupled from the motion of a molecule is the behaviour of a ship's compass whose orientation is unaffected by a stormy sea (Freeman 1987, Levitt 2001). However, the action of a molecule tumbling in solution does cause the magnitude and direction of the dipolar interaction between two coupled nuclei to fluctuate. This in turn causes the contribution of the local magnetic field to vary with time. For a dipolar interaction to
mediate nuclear spin relaxation, fluctuation of the local magnetic fields due to molecular motion has to occur at the correct frequency, $\omega_0$. If the rate of tumbling is appropriate, i.e. $\tau_c = \omega_0$, the component of fluctuations that is at the same resonance frequency as $\omega_0$ will effect nuclear spin transitions that lead to recovery of the equilibrium Boltzman populations. The magnitude of the dipolar interaction is dependent on the gyromagnetic ratios of the dipolar coupled nuclei and the internuclear distance (see Appendix A.1, Equation A.1).

The CSA interaction involves magnetic fields generated by electrons rather than nuclei. The application of an external magnetic field can induce currents in electronic orbitals. As these electron currents circulate throughout their bonding network a magnetic field is generated. Such local electron-induced magnetic fields can oppose or enhance the external $B_0$ magnetic field and are responsible for the chemical shifts seen in NMR spectra. Molecular motions modulate the strength and direction of the electron-induced field and therefore its contribution to the total magnetic field experienced by a nucleus. If this modulation is of an appropriate frequency, relaxation is induced. The direction of the electron induced magnetic field is determined by the orientation of the bonding network with respect to $B_0$. For example, in the study of amide $^{15}$N spin relaxation, the CSA of the $^{15}$N nucleus is an important consideration. The CSA of the $^{15}$N nucleus is approximately axially symmetric aligned along the direction of the NH bond vector. The contribution of this nuclear-electron interaction to nuclear relaxation is dependent on the strength of the applied magnetic field and the degree of anisotropy of the CSA tensor of the nucleus in question (see Appendix A.1, Equation A.2).

In the example of an amide group, relaxation of the amide $^{15}$N nucleus is governed by the heteronuclear dipole-dipole interaction with the bonded proton and the $^{15}$N CSA. At 600 MHz proton resonance frequency the dipolar interaction is dominant in terms of relaxation, with $^{15}$N CSA contributing approximately 10%. Further to global molecular motion, such as tumbling, local fluctuations in the NH bond orientation (i.e. so-called internal motion) of appropriate frequency can also elicit relaxation of magnetisation.
From the perspective of the $^{15}$N nucleus, dynamics of the amide $^{15}$N–$^1$H bond as a result of molecular reorientation causes fluctuations in the local NH dipolar interaction and in the $^{15}$N CSA magnetic field. The dipolar and $^{15}$N CSA contributions to the magnetic fields experienced by a $^{15}$N nucleus fluctuate over time. The effect of the local magnetic fields is to introduce a small positive or negative transverse component to the overall magnetic field, the magnitude of which averages to zero over time (Levitt 2001). It is not possible to fully predict a priori the frequency range of fluctuations that occur in a given protein. It is therefore necessary to consider a simplified model. When considering $^{15}$N relaxation the amide bond is assumed to be within a spherical top. That is a hard sphere with isotropic rotational and inertia characteristics. The probability of the small transverse magnetic field component being identical at time $t$ and at time $t + \tau$ decays exponentially. A Fourier transform of this autocorrelation function yields the spectral density function, $J(\omega)$, which takes the form of a Lorentzian distribution centred at $\omega = 0$ (see Appendix A.1, Equation A.3). This function describes the probability distribution of the frequency of the fluctuations of the local magnetic fields caused by isotropic rotational diffusion. In the realistic situation of a protein, the complex molecular dynamics that drive magnetic field fluctuations are unknown. Therefore, the autocorrelation function and thus the appearance of $J(\omega)$ are substantially more complex than suggested by Equation A.3 (Appendix A.1). In general, the fluctuations of the $^{15}$N CSA and heteronuclear dipolar magnetic fields influencing amide $^{15}$N magnetisation in a protein cannot be accurately described by considering only isotropic rotational tumbling. The effects of local bond motions on various time scales and anisotropic rotational diffusion also have to be considered. These factors have been cast in alternative analytical forms of the spectral density function. The so-called ‘model free’ formalism of Lipari and Szabo is a commonly invoked strategy to simplify the analysis of $J(\omega)$ (Lipari and Szabo 1982a; Lipari and Szabo 1982b). This treatment describes motions occurring on two independent time scales: isotropic molecular tumbling characterised by an overall correlation time $\tau_c$, and internal motion described by generalised order parameter $S^2$ and a second correlation time $\tau_e$ (see Appendix A.1, Equation A.6). This form of the spectral density function is at best an approximation, but has proved of general utility in the first order analysis of protein molecular dynamics in a large number of systems. There are many extensions to this simple treatment of the description of molecular motions that describe the observed nuclear relaxation properties, but the use of these
more complex analyses is often difficult to justify on the quality of the experimental data. In this thesis only this simplest of model is invoked at any stage.

There are two types of NMR signal relaxation: longitudinal (spin-lattice) or transverse (spin-spin) which are described by the time constants $T_1$ and $T_2$ respectively. Spin-lattice relaxation describes the recovery of Boltzman equilibrium magnetisation, that is $H_z$ or $N_z$ using the product operator formalism (see Chapter 1, Section 1.1.3). For example, a 90° RF pulse on $H_z$ magnetisation creates coherent transverse magnetisation (e.g. $H_x$ or $H_y$) where the upper and lower Zeeman state populations are equalised. $T_1$ describes the time taken to recover pure $H_z$ magnetisation. In qualitative terms, excited state nuclei relax to equilibrium by exchanging energy with the surroundings (also termed the lattice). In the case of spin-$\frac{1}{2}$ nuclei this is mediated purely by magnetic field interactions. The lattice consists of many magnetic dipoles undergoing thermal induced fluctuations that span a wide range of frequencies. Therefore, there is only a weak component of the lattice fluctuating at the Larmor frequency of a given nucleus. Therefore, the rate of spin-lattice relaxation is slow. Both dipolar and CSA interactions contribute to the rate of longitudinal relaxation. Plotting the function that describes $^{15}\text{N} \ T_1$ relaxation against the isotropic rotational correlation time, $\tau_c$, shows a minimum (i.e. when the rate of relaxation is most rapid) when $\tau_c = \omega_0^{-1}$ (Figure 4.3(b)). Simplistically this is because the modulation of local magnetic field fluctuations by overall molecular tumbling is occurring at the same frequency as the energy differences between the nuclear spin levels. The analytical function describing the $^{15}\text{N}$ longitudinal relaxation rate is given in Equation A.4 (Appendix A.1). The three transition frequencies described by the terms that are arise from considerations of dipole-dipole induced single, double and zero quantum transitions and hence are multiplied by the dipolar constant (see Appendix A.1, Equation A.1). In addition the $^{15}\text{N}$ CSA contribution to spin-lattice relaxation can be seen (see Appendix A.1, Equation A.4).

The $R_1$ rate, where $R_1$ equals $1/T_1$, is ordinarily measured by an ‘inversion-recovery’ experiment. Equilibrium magnetisation, e.g. $H_z$, is rotated onto the negative z-axis $-H_z$ by a 180° RF pulse from where it relaxes back to the positive z-axis. The rate of decay can be monitored by recording a series of experiments that use a second 90° RF pulse after an incremented relaxation delay to flip the longitudinal magnetisation into the
transverse plane for detection (Freeman 1987). For the measurement of $^{15}$N $R_1$, an inversion recovery sequence is inserted into a two-dimensional heteronuclear correlation experiment. A refocused INEPT sequence creates polarisation enhanced $^{15}$N magnetisation aligned along the negative z-axis (see Chapter 1, Section 1.1.3). The magnetisation is then permitted to relax along the z-axis before being rotated into the transverse plane for frequency labelling of the $^{15}$N chemical shift. Finally, chemical shift encoded $^{15}$N magnetisation is transferred back to $^1$H for detection. The benefits of constructing the experiment in this fashion are two-fold: first, individual NH resonances can be resolved through the use of two dimensions and second, INEPT-style pulse sequences can be used to enhance the low sensitivity $^{15}$N magnetisation (see Chapter 1, Section 1.1.3).

Transverse relaxation is defined by the intrinsic time dependent loss of phase coherence. A 90° RF pulse on a sample of spin-$\frac{1}{2}$ nuclei at equilibrium equalises the populations of the two energy levels and rotates net magnetisation from the positive z-axis onto one of the transverse axes of the rotating frame. This excited state is one of phase coherence. Whereas spin-lattice relaxation describes the recovery of the Boltzman equilibrium populations, spin-spin relaxation describes the loss of phase coherence in the transverse plane. The analytical function describing $^{15}$N transverse relaxation is plotted against the rotational correlation time in Figure 4.3(b)). The graph shows that in the extreme narrowing limit (describing fast molecular motion), when $\omega_c \ll 1$, the two relaxation mechanisms are essentially identical (Figure 4.3(b)). Under these conditions the mechanisms that promote $T_2$ relaxation are the same as those already discussed for $T_1$ relaxation. This can be seen in the analytical function describing the rate of $^{15}$N transverse relaxation (see Appendix A.1, Equation A.5). The main difference between the equations describing the two forms of relaxation are the $J(0)$ terms in both the dipolar and $^{15}$N CSA elements that arise for transverse relaxation. However, when the rate of tumbling ($\tau_c^{-1}$) is slower than the Larmor frequency, the $T_2$ parameter continues to decrease while $T_1$ increases (Figure 4.3(b)). This is due to the contributions of the $J(0)$ term in Equation A.5 (Appendix A.1) At these lower frequencies additional mechanisms, which only effect transverse relaxation, have to be considered. These include the contributions of chemical exchange and scalar relaxation (Freeman 1987). An additional experimental effect also has to be considered. Inhomogeneity of the external magnetic field across the
sample produces isochromats, small regions of the sample that experience the same $B_0$. The rate of precession of a nucleus is therefore dependent on its location in the sample. This mechanism causes dephasing of transverse magnetisation across the sample.

The spin-spin relaxation rate constant, $R_2$ (equals $1/T_2$), is ordinarily measured using an experiment based on the simple spin echo pulse sequence ($90^\circ$–$\Delta$–$180^\circ$–$\Delta$–Acquire). In this pulse sequence a $90^\circ$ RF pulse on a sample of isolated spin-$\frac{1}{2}$ nuclei generates transverse phase coherence, which decays during the period $\Delta$. The $180^\circ$ RF pulse reverses the sense of nuclear precession due to chemical shift offsets such that after a second period of duration $\Delta$ a signal ‘echo’ is observed. The $180^\circ$ pulse refocusses the evolution due to chemical shift and additionally suppresses the effects of $B_0$ magnetic field inhomogeneity. The spin-echo sequence is the building block of the Carr-Purcell-Meiboom-Gill (CPMG) pulse train, $[\Delta$–$180^\circ$–$\Delta]_{2n}$ (Carr and Purcell 1954; Meiboom and Gill 1958). This sequence performs the same functions as the spin echo sequence but also serves to reduce the problem of systematic errors in pulse length, which would lead to off-resonance effects. In addition, the effects of the finite diffusion rates of molecules through the sample during the relaxation delay are also reduced. In experiments that measure $^{15}$N $R_2$, the period during which transverse relaxation occurs is normally inserted into a 2D heteronuclear correlation experiment with indirect detection using proton magnetisation. In the $R_2$ experiment, because $^{15}$N magnetisation is left in the transverse plane, it is important to ensure that no evolution of the NH scalar coupling occurs. Evolution of in-phase $^{15}$N transverse relaxation, e.g. $N_y$, as a result of the scalar coupling $J_{HN}$ (see Chapter 1, Section 1.1.3), generates antiphase transverse magnetisation $-2H_xN_x$ (see for example Chapter 1, Figure 1.3(c)). Relaxation of the in-phase $N_x$ term is dictated only by the heteronuclear HN dipolar and $^{15}$N CSA interactions. The presence of $^1$H magnetisation in the antiphase $-2H_xN_x$ term means that the effect of homonuclear proton-proton dipolar couplings, which cause efficient relaxation, has to be considered (Peng and Wagner 1994). Consequently the antiphase term relaxes more quickly than the pure in-phase term. If the scalar coupling is allowed to evolve, the experimental relaxation rate determined will be a product of the relaxation of both of these operator terms. To prevent this from occurring the delay period $\Delta$ in the CPMG sequence is kept much shorter that the inverse of the $^{15}$N–$^1$H coupling constant, $J_{NH}$. In addition, $180^\circ$ pulses on the amide
protons are applied between pairs of CPMG sequence elements to prevent cross relaxation between the CSA and dipolar interactions. In summary, the CPMG pulse trains serves to ensure that any signal attenuation occurring during its application is a result of the intrinsic decay of phase coherence due to transverse relaxation. A caveat with the measurement of $R_2$ values is the effect of chemical exchange. As $^{15}$N magnetisation is kept in the transverse plane during the relaxation period any nucleus exchanging between two chemical environments will relax more quickly and give rise to an additional contribution to the measured $R_2$ value.

To determine each $^{15}$N relaxation rate (i.e. $R_1$ and $R_2$) experiments, a series of 2D NH correlation spectra are recorded differing only by the length of the appropriate relaxation delay period (e.g. by varying $n$ in the CPMG pulse train in $R_2$ experiments). To estimate the $^{15}$N relaxation rate, the signal intensity of amide cross peak in each 2D NH correlation spectrum is measured. Over a series of incremented relaxation delays the signal intensity is a function of the duration of the delay period, $t$ (Equation 4.2)

$$I(t) = I_0 \exp \left[ -\frac{t}{T_i} \right]$$  \hspace{1cm} \text{Equation 4.2}$$

where $I(t)$ is the measured signal intensity at time $t$, $I_0$ is the initial signal intensity at 0 seconds, $t$ is the delay period and $T_i$ is the relaxation time where $i = 1$ or 2. The measured signal intensities can then be fitted to Equation 4.2 to estimate the respective relaxation rate, $R_i$ on a cross peak by cross peak (i.e. residue by residue) basis.

The third commonly determined heteronuclear relaxation parameter, the $^{1}{H}^{-^{15}}$N NOE, describes the enhancement of the $^{15}$N spin when the nuclear energy levels of the bonded proton are saturated. The NOE value is then, simply, the ratio of the $^{15}$N signal intensity with ($I_{\text{sat}}$) and without ($I_{\text{nosat}}$) proton saturation. The heteronuclear NOE describes cross-relaxation rate between $^1$H and $^{15}$N nuclei in the amide NH group. Following saturation of spin states of the proton, magnetisation is transferred to the $^{15}$N nucleus via the heteronuclear dipolar coupling. The magnitude of the heteronuclear NOE is a function of the effective rate of NH bond vector rotation (Figure 4.3(b)). The sign of the heteronuclear NOE value ($I_{\text{sat}}/I_{\text{nosat}}$) is sensitive to the
frequency of motion the NH bond experiences. If the NH group only experiences motion of the same frequency as the global tumbling rate then the heteronuclear NOE value is positive and close to a typical maximum value around 0.8 (Kay et al., 1989; this value is dependent on the magnitude of $B_0$). Smaller values of the heteronuclear NOE, which may approach zero or be negative indicate that a NH group is undergoing large amplitude motion at a frequency greater than the rate of overall protein tumbling. The heteronuclear NOE value is therefore a useful indicator of the rate of motion a particular NH group is experiencing. As will be seen later, this parameter plays an important role when estimating the rotational correlation time of a molecule.

The steady-state $^1$H-$^{15}$N NOE is ordinarily measured with the aid of a 2D heteronuclear correlation experiment that does not include the first refocused INEPT element. In the absence of proton saturation, $^{15}$N magnetisation is excited by a 90° RF pulse, frequency encoded during the $t_1$ period and transferred onto the proton for detection by a reverse refocused INEPT element. The saturated spectrum is recorded using an essentially identical pulse sequence but with a long period (typically several seconds) of proton saturation at the start of the experiment prior to $^{15}$N excitation.

Within the application of the Lipari and Szabo model-free formulism, the rate at which a protein tumbles isotropically in solution can be estimated from using the average $^{15}$N $R_2/R_1$ ratio, $\langle R_2/R_1 \rangle$ (Kay et al., 1989). In the application of this mode of analysis it is necessary to exclude those residues that are undergoing chemical exchange (i.e. those with shortened $T_2$ times) or extensive rapid internal motion (i.e. those with low heteronuclear NOE values). The use of such a filter permits the assumption that the frequencies of fluctuations of local magnetic fields are modulated only by the tumbling rate of the protein. In these circumstances the NH bond motions can be assumed to mimic that of a spherical top. The numerical values of the heteronuclear NOE is used to filter NH groups who have additional, faster motions to overall molecular tumbling. Only NH groups that have $I_{\text{sat}}/I_{\text{fnosat}}$ values greater than a cut off value (e.g. $> 0.65$) are included in the final data set. The contribution of chemical exchange to transverse relaxation rates can lead to an overestimation of $\langle R_2/R_1 \rangle$ and consequently an erroneously long correlation time. A statistical filter is
normally applied to exclude residues with \( R_2/R_1 \) values significantly different from the sample mean value (Equation 4.3)

\[
\left| \frac{\langle R_1 \rangle - R_{1,i}}{\langle R_1 \rangle} - \frac{\langle R_2 \rangle - R_{2,i}}{\langle R_2 \rangle} \right| > 1.5 \times SD \tag{Equation 4.3}
\]

where \( \langle R_1 \rangle \) and \( \langle R_2 \rangle \) are the sample average \(^{15}\)N longitudinal and \(^{15}\)N transverse relaxation rates respectively, \( R_{1,i} \) and \( R_{2,i} \) are the individual NH group relaxation rates and SD is the standard deviation of the left hand side of Equation 4.3 over the complete data set of \(^{15}\)N \( R_2/R_1 \) values (Tjandra et al. 1995). This filter reliably serves to remove those residues undergoing chemical exchange. Only \(^{15}\)N \( R_2/R_1 \) values that satisfy the standard deviation and heteronuclear NOE filters are used in the estimation of \( \tau_c \). The application of these filters allows the use of the simplest possible definition of the spectral density function to be used as given in Equation A.3 (Appendix A.1). If residues undergoing motion on a different time scale to molecular tumbling are not excluded a more advanced definition of \( J(\omega) \) would be required, such as that proposed by Lipari and Szabo (Lipari and Szabo 1982a; Lipari and Szabo 1982b).

\(^{15}\)N \( R_2/R_1 \) increases as a function of \( \tau_c \) (Figure 4.3(c)). With a suitable data set filtered as described above, \( \tau_c \) can be estimated by minimisation of Equation 4.4

\[
\chi^2 = \left( \frac{R_2(\tau_c)}{R_1(\tau_c)} - \langle R_2/R_1 \rangle \right)^2 \tag{Equation 4.4}
\]

where \( R_1(\tau_c) \) and \( R_2(\tau_c) \) are functions describing the effect of molecular tumbling frequency on the respective relaxation rates (see Appendix A.1) and \( \langle R_2/R_1 \rangle \) is the average experimental rates ratio.

4.2.2 - \(^{15}\)N relaxation parameters from \([^{15}\text{N}]\)- and \([^{2}\text{H},^{15}\text{N}]\)-labelled ubiquitin

\(^{15}\)N \( T_1 \), \(^{15}\)N \( T_2 \) and \( \{^{1}\text{H}\}^{15}\text{N} \) heteronuclear NOE experiments were recorded on samples of \([^{15}\text{N}]\)- and \([^{2}\text{H},^{15}\text{N}]\)-labelled ubiquitin without glycerol at 25°C and at 600 MHz proton frequency (see 'Materials and Methods' Section 3.3.5 for the experimental details and Section 3.4.2 for details of the processing of the \(^{15}\)N relaxation data). The results are presented in Figure 4.4. A good correlation was observed between the \(^{15}\)N \( R_2/R_1 \) ratios of \([^{15}\text{N}]\)- and \([^{2}\text{H},^{15}\text{N}]\)-labelled samples on a
Figure 4.4. Comparison of backbone NH $^{15}\text{N} R_2/R_1$ ratio for $^{15}\text{N}$-labelled (●) and $^{2}\text{H}, ^{15}\text{N}$-labelled ubiquitin (●). Error bars indicate plus and minus one standard deviation from the mean. All $^{15}\text{N}$ relaxation experiments were performed at 600 MHz (See ‘Material and Methods’ Section 3.3.5).
residue-by-residue basis. However, for the majority of residues the $^{15}\text{N} R_2/R_1$ ratio was lower in the deuterated sample.

4.2.3 - $^{15}\text{N}$ relaxation parameters in the absence and presence of 50% (v/v) glycerol

$^{15}\text{N}$ T$_1$, $^{15}\text{N}$ T$_2$ and $^1\text{H}$-$^{15}\text{N}$ heteronuclear NOE experiments were recorded on $^2\text{H}$, $^{15}\text{N}$]-labelled ubiquitin in 50% (v/v) glycerol at 25°C and at 600 MHz proton frequency (see 'Materials and Methods' Sections 3.3.5 and 3.4.2). The results are presented in Figure 4.5 and Table 4.1. As was crudely observed with the measurement of cross peak S/N ratios in Section 4.1, the addition of glycerol had a dramatic effect on the $^{15}\text{N}$ relaxation properties of ubiquitin (Figure 4.5). A substantial increase in the residue-by-residue $^{15}\text{N} R_2/R_1$ ratios was observed. A less desirable but inevitable side effect of the addition of glycerol was an increase the level of experimental uncertainty of the derived relaxation parameters: the average residue experimental standard deviation rose from 3 % of the mean value in the absence of glycerol to 13.5 % of the mean with 50% glycerol. However, the profile of $^{15}\text{N} R_2/R_1$ ratios across the protein as a whole appeared similar for both sample conditions (Figure 4.5). $^{15}\text{N} R_2/R_1$ ratios were used to calculate $\tau_c$ for both samples (see 'Materials and Methods' Section 3.4.2). In the absence of glycerol, ubiquitin had an isotropic rotational correlation time of 4.03 ns ± 0.2 ns at 25°C which increased considerably to 34.6 ns ± 0.7 ns on the addition of glycerol to 50% (v/v) at the same temperature (Table 4.1). The value of $\tau_c$ obtained for ubiquitin in the absence of 50% (v/v) glycerol is in good agreement with values from other studies (Tjandra et al., 1995; Lee and Wand, 1999).

4.2.4 - The effect of changing the acquisition temperature on $^{15}\text{N}$ relaxation rates

The addition of glycerol to a solution of $^2\text{H}$, $^{15}\text{N}$]-labelled ubiquitin exacted a coarse adjustment of the hydrodynamic properties of the protein. A more subtle variation was achieved by varying the temperature of the sample. $^{15}\text{N} R_1$, $^{15}\text{N} R_2$ and heteronuclear NOE parameters were estimated for amide $^{15}\text{N}$ nuclei of $^2\text{H}$, $^{15}\text{N}$]-labelled ubiquitin in 50 % (v/v) glycerol at five temperatures ranging from 17.5 to 34.3°C (Table 4.1).
Figure 4.5. Backbone NH $^{15}\text{N} \ R_2/R_1$ ratios of $[^2\text{H}, ^{15}\text{N}]$-labelled ubiquitin at 25°C in the absence (●; left hand y-axis) and presence (●; right hand y-axis) of 50% (v/v) glycerol. Error bars indicate plus and minus one standard deviation from the mean. $^{15}\text{N} \ R_2/R_1$ ratios recorded from the same sample of $[^2\text{H}, ^{15}\text{N}]$-labelled ubiquitin at 600 MHz (See ‘Materials and Methods’ Section 3.3.5).
Table 4.1. Estimated $\tau_c$ times for [^{2}H, ^{15}N]-labelled ubiquitin under different sample conditions

<table>
<thead>
<tr>
<th>Temperature (°C)</th>
<th>Glycerol (%)</th>
<th>$\langle R_2/R_1 \rangle^a$</th>
<th>$\tau_c$ (ns)$^b$</th>
</tr>
</thead>
<tbody>
<tr>
<td>25.7</td>
<td>0</td>
<td>2.79 ± 0.3</td>
<td>4.03 ± 0.17</td>
</tr>
<tr>
<td>17.5</td>
<td>50</td>
<td>195.2 ± 41.1</td>
<td>44.5 ± 1.36</td>
</tr>
<tr>
<td>21.5</td>
<td>50</td>
<td>133.4 ± 18.7</td>
<td>37.7 ± 0.78</td>
</tr>
<tr>
<td>25.7</td>
<td>50</td>
<td>115.6 ± 15.8</td>
<td>34.6 ± 0.7</td>
</tr>
<tr>
<td>29.9</td>
<td>50</td>
<td>82.7 ± 7</td>
<td>29.3 ± 0.45</td>
</tr>
<tr>
<td>34.3</td>
<td>50</td>
<td>51.6 ± 5.9</td>
<td>23.3 ± 0.29</td>
</tr>
</tbody>
</table>

$^a$ average $^{15}N$ $R_2/R_1$ ratio values and standard deviations over the $^{15}N$ $R_2/R_1$ data set (see 'Materials and Methods' Section 3.4.2); $^b$ average isotropic rotational correlation time and Monte Carlo standard deviation (see 'Materials and Methods' Section 3.4.2).

Attempts to record relaxation experiments at temperatures lower than 17.5°C proved unsuccessful as extremely low S/N ratios and severe line broadening were observed in the first 2D plane (i.e. with the relaxation delay set to 0 μs) of [^{1}H, ^{15}N]-HSQC based relaxation experiments (data not shown). However, between 17.5 and 34.4°C it proved possible to extract signal decay curves and heteronuclear NOE ratios for the majority of the NH cross peaks of ubiquitin. The results of this survey are summarised in Figure 4.6 and Table 4.1. The rate of protein tumbling is proportional to temperature (Equation 4.1). The average $^{15}N$ $R_1$ value, $\langle R_1 \rangle$, increased and the average $^{15}N$ $R_2$ value, $\langle R_2 \rangle$, decreased as the sample temperature was raised (Figures 4.6(a) and 4.6(b)). At lower temperatures the observed signal linewidth increased, resulting in lower S/N ratios. Consequently, the degree of experimental uncertainty was even more pronounced at lower temperatures. This was particularly noticeable in the change in the average $^{15}N$ $R_2/R_1$ ratio, $\langle R_2/R_1 \rangle$, with temperature (Table 4.1 and Figure 4.6(c)). Broader distributions of relaxation parameter values were generally seen at lower temperatures. Finally, the range of $\langle R_2/R_1 \rangle$ values corresponded to a 21.2 ns decrease in $\tau_c$ over the temperature range studied (Table 4.1 and Figure 4.6(d)).
Figure 4.6. Backbone $^{15}$N relaxation parameters for $[^{2}$H, $^{15}$N]-labelled ubiquitin in 50% (v/v) glycerol between 17.5 °C and 34.3 °C. Overall averages of (a) $^{15}$N $R_1$, (b) $^{15}$N $R_2$, (c) $^{15}$N $R_2/R_1$ and (d) $\tau_c$. The error bars indicate plus and minus one standard deviation from the mean (See 'Materials and Methods' Section 3.4.2). Each graph shows a linear regression line (—) and a correlation coefficient, $r$, for the data.
4.2.5 – Secondary structure affects NH cross peak S/N ratio

Figure 4.7(a) shows the differences in (S/N) ratios measured from a [$^1$H, $^{15}$N]-TROSY spectrum for amide groups in different types of secondary structure (see ‘Materials and Methods’ Section 3.4.1). The difference between the (S/N) ratio of amide groups in α-helices and β-strands compared to that of all the residues is particularly striking. No such differences were observed between $R_2/R_1$ values when classed in the same structural subsets (Figure 4.7(b)).

4.3. Resonance Assignment of [$^2$H, $^{13}$C, $^{15}$N]-labelled Ubiquitin in 50% (v/v) Glycerol.

As was shown in Section 4.2, [$^2$H, $^{15}$N]-labelled ubiquitin in 50% (v/v) glycerol performs well as a temperature ‘tunable’ model for studying larger proteins. A selection of the three-dimensional (3D) triple resonance NMR experiments that would be used to facilitate resonance assignment of WT PaDDAH were tested on a sample of [$^2$H, $^{13}$C, $^{15}$N]-labelled ubiquitin in 50% (v/v) glycerol at 25°C. These solution conditions correspond to a protein with an isotropic rotational correlation time of approximately 35 ns (Table 4.1).

A minimum of three 3D triple resonance experiments are required to assign backbone $^1$H, $^{15}$N and $^{13}$Cα, $^{13}$Cβ and $^{13}$C carbonyl resonance frequencies: HNCA (Figure 5.14(a)), HN(CA)CB (Figure 5.14(b)) and HN(CA)CO. In this instance [$^1$H, $^{13}$C, $^{15}$N]-HNCA and [$^1$H, $^{13}$C, $^{15}$N]-HN(CA)CB experiments, both employing TROSY coherence selection, were recorded on a sample of [$^2$H, $^{13}$C, $^{15}$N]-labelled ubiquitin in 50% (v/v) glycerol at 25°C and 600 MHz proton frequency (Table 4.2). Both of these experiments provide intra- and inter-residue connectivities and should be sufficient to resolve any chemical shift degeneracy, especially considering the relatively low number of correlations expected for ubiquitin (approximately 70).

The number of correlations observed in each of the 3D spectra recorded was compared to the number expected from the amino acid sequence of ubiquitin (Figure 4.8(a)). The statistics presented in Figure 4.8(a) showed that all of intra-residue
Figure 4.7. Comparison of average S/N (a) and $^{15}$N $R_2/R_1$ (b) for amide NH cross peaks in different secondary structure types between 17.5 °C and 34.3 °C. Structural subsets were determined from the crystal structure of ubiquitin (1ubq; Vijay-Kumar et al., 1987). Each graph shows the average for all NH groups (●), NH groups in all secondary structure (O), α-helical (■), β-strands (▲) and non-structured (▼) regions. (S/N) and $\langle R_2/R_1 \rangle$ values calculated as described in the text from data sets filtered for $^{15}$N $R_2/R_1$ (Equation 4.3 with $n = 1.5$ and heteronuclear NOE > 0.65).
Table 4.2. Summary of NMR experiments conducted to facilitate backbone chemical shift assignment of $[^2\text{H}, ^{13}\text{C}, ^{15}\text{N}]$-labelled ubiquitin in 50% (v/v) glycerol.

<table>
<thead>
<tr>
<th>Experiment $^a$</th>
<th>Scans/FID</th>
<th>$t_1$</th>
<th>$t_2$</th>
<th>$t_3$</th>
<th>Correlations $^b$</th>
<th>Reference $^c$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$[^1\text{H}, ^{15}\text{N}]$ TROSY $^d$</td>
<td>8</td>
<td>128 ($^{15}\text{N}$)</td>
<td>1024 ($^1\text{H}$)</td>
<td>-</td>
<td>N, H$^N$</td>
<td>Pervushin et al., 1997</td>
</tr>
<tr>
<td>$[^1\text{H}, ^{13}\text{C}, ^{15}\text{N}]$ HNCA $^e$</td>
<td>8</td>
<td>64 ($^{13}\text{C}$)</td>
<td>36 ($^{15}\text{N}$)</td>
<td>512 ($^1\text{H}$)</td>
<td>N, H$^N$, C$\alpha_i$, C$\alpha_{i-1}$</td>
<td>Yang and Kay (1999)</td>
</tr>
<tr>
<td>$[^1\text{H}, ^{13}\text{C}, ^{15}\text{N}]$ HN(CA)CB $^e$</td>
<td>8</td>
<td>64 ($^{13}\text{C}$)</td>
<td>36 ($^{15}\text{N}$)</td>
<td>512 ($^1\text{H}$)</td>
<td>N, H$^N$, C$\beta_i$, C$\beta_{i-1}$</td>
<td>Yang and Kay (1999)</td>
</tr>
</tbody>
</table>

* The brackets indicate correlations not observed; $^b$ 'Correlations' lists the expected correlations generated by the experiment; $^c$ most experiments were based on those cited but optimized in house; $^d$ $^{13}\text{C}$ decoupling was employed during $t_1$ and proton acquisition. $^e$ all 3D pulse sequences employed TROSY selection as described in Yang and Kay 1999. No deuterium decoupling was performed on any experiment. All experiments were performed at 25°C and at 600 MHz proton resonance frequency (See 'Materials and Methods' Section 3.3.6). The acquisition times for both HNCA and HN(CA)CB were approximately 36 hours. Scans/FID refers to the number of transients recorded per $t_i$ transient.
Table 4.1: Expected and observed number of cross peaks for HNCA and HN(CA)CB experiments.

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Correlation</th>
<th>Residue</th>
<th>Number of cross peaks</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>($^{1}\text{H}^{N},^{15}\text{N}$)$\rightarrow^{13}\text{C}$</td>
<td></td>
<td>Expected a</td>
</tr>
<tr>
<td>HNCA</td>
<td>$\alpha$ i</td>
<td></td>
<td>70</td>
</tr>
<tr>
<td></td>
<td>$\alpha$ i - 1</td>
<td></td>
<td>70</td>
</tr>
<tr>
<td>HN(CA)CB</td>
<td>$\beta$ i</td>
<td></td>
<td>64</td>
</tr>
<tr>
<td></td>
<td>$\beta$ i - 1</td>
<td></td>
<td>64</td>
</tr>
</tbody>
</table>

\textsuperscript{a} number of expected cross peaks was calculated from the primary sequence of ubiquitin accounting for the N-terminus, three prolines (residues 19, 37 and 38) and three missing backbone NH cross peaks in ($^{1}\text{H},^{15}\text{N}$)-TROSY spectra: 24, 53 and 46. In the HNCA CB 6 fewer cross peaks are expected because of the 6 glycine residues; \textsuperscript{b} Cross peaks observed in 3D carbon strip plots. Inter- and intra-residue cross peak overlap was not considered when counting cross peaks and, thus, may account for some of the discrepancies between the numbers of observed and expected cross peaks.

Figure 4.8: Observed $^{1}\text{H}^{15}\text{N}_{i} \rightarrow^{13}\text{C}$ correlations in 3D spectra. (a) Table showing the number of expected and observed carbon correlations per spectrum (see Table 4.1 for details of the experiments conducted). i indicates an intra-residue correlation and i - 1 indicates an inter-residue correlation to the preceding residue; (b) A pie chart showing the observed number of carbon correlations per amide NH cross peak (minimum: 0; maximum: 4) and the percentage of NH cross peaks that yield that number.
(HN$_i$$\rightarrow$C$_i$) and the majority of inter-residue (HN$_i$$\rightarrow$C$_{i+1}$) connectivities were observed in these experiments. The discrepancy between the observed and expected number of intra- and inter-residue connectivities observed in both 3D spectra may be due to overlap of the intra- and inter-residue cross peaks and decrease in the efficiency of inter-residue magnetisation transfer. The number of carbon correlations observed for each amide cross peak in the [$^1$H, $^{13}$N]-TROSY reference spectrum recorded at 25°C was also totaled (Figure 4.8(b)). The maximum expected number of $^{13}$C correlations per NH group from the two experiments recorded was four. As can be seen (Figure 4.8(b)), over three quarters of the residues with observable amide cross peaks in a NH correlation spectrum yielded the maximum number of correlations.

The $^{13}$Ca and $^{13}$Cb resonance frequencies of alanine, glycine and serine and threonine residues make these residue types relatively easy to identify in 3D triple resonance NMR spectra (see Chapter 5, Section 5.5.2 and Figure 5.15(a) for more details). The primary sequence of ubiquitin contains 2 alanine, 6 glycine, 3 serine and 7 threonine residues (Table 4.3). The $^{13}$C planes corresponding to each NH cross peak in the reference [$^1$H, $^{15}$N]-TROSY spectrum were inspected and the number of identifiable $^{13}$Ca and $^{13}$Cb connectivities recorded (Table 4.3).

<table>
<thead>
<tr>
<th>Amino Acid</th>
<th>Number in Sequence</th>
<th>Number Observed $^b$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>i</td>
</tr>
<tr>
<td>Alanine ($\beta$)</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Glycine ($\alpha$)</td>
<td>6</td>
<td>5</td>
</tr>
<tr>
<td>Serine and Threonine ($\beta$)</td>
<td>10</td>
<td>9</td>
</tr>
</tbody>
</table>

$^a$ The $\alpha$ or $\beta$ terms in parentheses refers to the $^{13}$C correlation that allows identification of that residue type; $^b$ 'i' refers to characteristic intra-residue $\alpha$ or $\beta$ cross peak in the i spin system whereas 'i - 1' refers to the inter-residue $\alpha$ or $\beta$ cross peak in the i + 1 spin system; $^c$ it is not possible to differentiate between serine and threonine residues using $^{13}$C chemical shifts alone (see Chapter 5, Section 5.6).

The number of carbon planes identifiable as corresponding to either alanine, glycine or serine and threonine residues on the basis of $^{13}$C chemical shift alone was close to that expected from the amino acid sequence. Several of the discrepancies had simple explanations. The linewidth of the NH cross peak of Gly-53 is broadened due to
chemical exchange (Lee and Wand 1999). Hence no $^{13}$C strip was observed for this residue, but the inter-residue $(i - 1)$ correlation to Ca of Gly-53 was observed in the carbon plane from the NH cross peak of Arg-54. Gly-76 is the C-terminal residue and obviously no inter-residue cross peak can be observed in the spin system of the following residue. Only three of the expected residue-characteristic $^{13}$C correlations were not observed. These were the intra-residue $\beta$ cross peak of Thr-9 and the inter-residue $\beta$ cross peaks for Thr-9, Ala-28 and Ser-57 in the carbon plane of the $i + 1$ residue. In each instance the reason was a missing correlation and not overlap of the intra- and inter-residue $^{13}$C cross peaks.

The 3D triple resonance spectra recorded here permitted near complete sequence-specific resonance assignment of the [$^1$H, $^{15}$N]-TROSY spectrum of [$^2$H, $^{13}$C, $^{15}$N]-labelled ubiquitin in 50% glycerol at 25°C. There were six exceptions: Glu-24 and Gly-53 (chemical exchange-broadened cross peaks), Met-1 and the three proline residues, 19, 37 and 38. Examples of the inter-residue connectivities for residues Thr-66 to His-68 are presented in Figure 4.9.

As was seen earlier with $^{15}$N relaxation spectra of [$^2$H, $^{15}$N]-labelled ubiquitin in 50% (v/v) glycerol, the [$^1$H, $^{15}$N]-TROSY spectrum shows broad linewidths but good signal resolution due to the low number of NH cross peaks (Figure 4.9(a)). The spectral quality of the carbon dimensions, in terms of S/N ratio and signal resolution, in both experiments was also good (Figure 4.9(b)).

4.5. Discussion.

Ubiquitin in a viscous solvent was chosen as a model system to mimic the heteronuclear NMR study of large proteins. Ubiquitin has been extensively characterised by solution NMR and yields a well-resolved NH correlation spectrum. Furthermore, ubiquitin is known to be stable in solution for long periods of time and can withstand elevated temperatures. In this study, [$^1$H, $^{15}$N]-HSQC and TROSY spectra of [$^2$H, $^{15}$N]-labelled ubiquitin in H$_2$O (with 10 % D$_2$O) revealed all the expected backbone amide NH correlations as well as a full complement of asparagine and glutamine side chain NH$_2$ cross peaks (Figure 4.1(a)). However, in 50% glycerol, [$^1$H, $^{15}$N]-TROSY spectra revealed only seven out of the eight expected asparagine and glutamine side chain NH$_2$ groups (Figure 4.1(b)). There are many published
Figure 4.9. Examples of $^{1}H^{15}N_i \rightarrow ^{13}C$ correlations visualised using ANSIG (Kraulis 1989). The NH cross peaks indicated in the $[^{1}H, ^{15}N]$-TROSY spectrum shown in (a) have the corresponding carbon spin systems shown in (b). (b) ANSIG session strips showing Ca (purple; from HNCA data set) and Cβ (blue; HN(CA)CB) 3D cross peaks. The vertical and horizontal axes of the carbon strips are $^{13}C$ chemical shift (ppm) and $^{1}H$ chemical shift (ppm) respectively. The vertical dashed line corresponds to the $^{1}H$ chemical shift of the NH cross peak selected in the TROSY spectrum (left). Inter-residue connectivities are indicated with the orange (Ca) and green (Cβ) lines.
accounts of incomplete protonation of amide groups of proteins expressed in deuterated media (McCallum et al. 1999; Salzmann et al. 2000). Ubiquitin was purified in the native state in buffers close to pH 7 ± 1. No unfolding and refolding steps were required to achieve full backbone amide protonation. The overall level of aliphatic deuteration as judged by 1D $^1$H spectroscopy was greater than 70%. The protein was expressed from a bacterial culture prepared in 99.5 % D$_2$O and thus the level of Cα protonation, which is believed to reflect the ratio of D/H in the medium, was assumed also to be of the order of 99 %.

The use of glycerol to modify the viscosity of a solution and thus the tumbling rate of a protein is not novel. For example, Kay and co-workers tested the benefits of gradient coherence selection and sensitivity enhancement on NH cross peak S/N ratios with a SH2 domain in 0, 15 or 30 % glycerol (Shan et al. 1996). For ubiquitin in this study, a clear deterioration in $[^1$H, $^{15}$N]-TROSY spectral quality (Figure 4.1) and S/N ratios (Figure 4.2) was observed on the addition of glycerol to 50 % (v/v). This was indicative of a decrease in molecular tumbling rate. Adjustments in sample temperature also affected the S/N ratios (Figure 4.2). This could be seen most dramatically by comparing the spectrum recorded at 10°C in 50 % (v/v) glycerol with that recorded at 25°C in the absence of glycerol. There was a 20-fold reduction in the average NH cross peak S/N value on the addition of glycerol and a 20-fold variation over the temperature range that was explored. $^{15}$N spin relaxation measurements provided a more robust measurement of the behaviour of ubiquitin in glycerol. Estimation of $^{15}$N relaxation parameters showed a considerable change in apparent $\tau_c$ on the addition of glycerol to 50 % (v/v) and a further 21.2 ns over the temperature range studied (Table 4.1 and Figure 4.6). Therefore this model system has the capacity to be ‘tuned’ to mimic the solution behaviour of a wide range of molecular weight systems.

At 25°C in 50 % glycerol the ratio of the average S/N values for residues in α-helices and β-sheets is 0.46 (Figure 4.7(a)). The explanation for this observation is straightforward: in a deuterated protein the amide proton of a NH group in a β-sheet is effectively isolated from other protons. The relaxation of both $^{15}$N and $^1$H nuclei is exclusively the result of the heteronuclear dipolar interaction and the $^{15}$N (and to a
lesser extent $^1$H) CSA interaction. However, the architecture of the $\alpha$-helix results in each amide NH group being in close proximity to the preceding and following amide NH group (Wüthrich 1986). In a $\beta$-sheet in a perdeuterated protein three remote amide protons, placed at distances of 4.3 Å (residue $i - 1$), 4.3 Å ($i + 1$) and 3.3 Å (long range contact across the $\beta$-sheet), are considered in homonuclear $^1$H dipolar interactions. Whereas in an $\alpha$-helix four possible dipolar coupled amide protons are present at distances of 2.8 Å ($i - 1$), 2.8 Å ($i + 1$), 4.2 Å ($i - 2$) and 4.2 Å ($i + 2$; Wüthrich 1986; Salzmann et al. 1998). Consequently, in $\alpha$-helices but to a lesser extent in $\beta$-sheets, $^1$H homonuclear auto- and cross-relaxation can occur between pairs of proximate amide protons, which in turn accelerates the relaxation of $^{15}$N magnetisation. In a [$^1$H, $^{15}$N]-TROSY pulse sequence no decoupling of the amide proton occurs during $^{15}$N frequency encoding. This permits evolution of the $J_{\text{NH}}$ scalar coupling. As discussed in Section 4.2.1, the $HN$ antiphase term (such as $2H_2N_X$ or $2H_2N_Y$) contains both $^1$H and $^{15}$N magnetisation terms. The overall relaxation of the antiphase term is therefore dependent on relaxation of both $^{15}$N and $^1$H magnetisation. As the magnitude of the dipolar interaction is strongly dependent on the inter-nuclear distance of the two spins as a result of the $(r^{-3})^2$ terms in Equations A.1, A.4 and A.5 (Appendix A.1), amide protons in $\alpha$-helices will experience stronger dipolar couplings than amide protons in $\beta$-sheets. As a consequence $^{15}$N nuclei from $\alpha$-helical amide groups will have a faster $R_2$ rate during the $t_1$ period than those in $\beta$-sheets and ultimately a broader linewidth. The S/N enhancement for TROSY-based triple-resonance experiments compared to HSQC-based counterparts has been shown to be dependent on secondary structure (Salzmann et al. 2000). In ubiquitin this phenomenon of S/N ratio reflecting the underlying secondary structure was particularly apparent. Only 57 of a possible 70 amide cross peaks were observed in the [$^1$H, $^{15}$N]-TROSY spectrum recorded at 10°C in 50% glycerol (Figure 4.2(a)). Of the 16 $\alpha$-helical residues in ubiquitin (23-34 and 56-59; Vijay-Kumar et al. 1987), only half yielded observable amide NH cross peaks at this temperature. At 25°C in 50% glycerol, where all 70 expected NH peaks were seen, the amide cross peaks of 15 of the 16 $\alpha$-helical residues were observed. This observation has important implications when studying large proteins with TROSY-based experiments. Clearly high molecular weight proteins with a large proportion of $\alpha$-helical secondary structure would be expected to yield poorer TROSY spectra compared to those with mainly $\beta$-sheet secondary structure.
The NMR experiments used in this study to measure $^{15}\text{N}$ auto-relaxation rates decoupled the amide proton during the $^{15}\text{N}$ relaxation period (see Section 4.2.1). Accordingly no differences were observed in the $\langle R_2/R_1 \rangle$ of amide groups in $\alpha$-helices or the $\beta$-sheet (Figure 4.7(b)).

The dramatic decrease in S/N observed on the addition of 50% (v/v) glycerol was echoed in the quality of the $^{15}\text{N}$ relaxation data recorded. The experiments employed here were based on the [$^1\text{H}$, $^{15}\text{N}$]-HSQC sequence. In each relaxation experiment conducted, the 2D NH spectra recorded for each relaxation delay time suffered from the low resolution this HSQC sequence affords for larger proteins. This was particularly acute at temperatures below 25°C. Attempts to record relaxation at 10°C failed because the resolution and S/N ratio of test spectra recorded with no relaxation period was poor. At each temperature studied it was necessary to discard the relaxation data extracted for several NH cross peaks because of compromising signal overlap. In instances where two peaks were partially overlapping the $^{15}\text{N} R_1$ and $^{15}\text{N} R_2$ values were only kept if decay curves appeared monoexponential. As ubiquitin was [$^2\text{H}$, $^{15}\text{N}$]-labelled, it is possible that the use of [$^1\text{H}$, $^{15}\text{N}$]-TROSY based relaxation experiments may have partially alleviated the problem of signal overlap (Zhu et al. 2000). The potential enhancement in S/N ratio and improvement in signal linewidth obtained using the TROSY principle may permit accurate extraction of relaxation data below 17.5°C. This would serve to further expand the range of molecular weights this model could be used to imitate on a quantitative basis.

Backbone $^1\text{H}^N$, $^{13}\text{C}$ ($\alpha$ and $\beta$) and $^{15}\text{N}$ resonance assignment of [$^2\text{H}$, $^{13}\text{C}$, $^{15}\text{N}$]-labelled ubiquitin in 50% (v/v) at 25°C was performed relatively easily and required the acquisition of only two TROSY-based 3D triple-resonance experiments. Under these sample conditions ubiquitin was shown to have an isotropic rotational correlation time of approximately 35 ns (Section 4.2.3). All alanine, glycine and serine or threonine residues were accounted for in the NH correlation map using $^{13}\text{C}\alpha$ and $^{13}\text{C}\beta$ chemical shifts only (Section 4.3). Only three cross peaks remained unaccounted for in these $^{13}\text{C}$ spin systems. Using these very characteristic $^{13}\text{C}$ chemical shifts as starting points,
sequential resonance assignment was efficiently completed and the results agreed well with those previously reported (Lee and Wand 1999).

There were advantages and disadvantages with the choice of ubiquitin in 50 % (v/v) glycerol as a model system for studying large proteins by heteronuclear NMR. The analyses of $^{15}$N relaxation parameters and estimation of $\tau_c$ indicated that under these solution conditions and over the temperature range studied ubiquitin had relaxation properties characteristic of a protein described by a 23.3 to 44.5 ns rotational correlation time. A further benefit of choosing ubiquitin is that it is a small monomeric protein and as such would not generate complex 2D and 3D correlation spectra that contain a large number of correlations. Although this simplified the task of sequence-specific resonance assignment (see Section 4.3), such uncrowded spectra are not representative of the similar spectra expected of proteins with $\tau_c$ equal to 35 ns in H$_2$O. A near full complement of $^{13}$C correlations were observed in 3D triple-resonance spectra of ubiquitin in 50 % (v/v) glycerol at 25°C indicating that magnetisation transfer efficiency in these experiments was high. The magnetisation transfers statistics presented for ubiquitin in 50 % (v/v) glycerol at 25°C (Figure 4.8) are superior to those reported elsewhere for proteins of a similar $\tau_c$ using similar experiments (Shan et al. 1996). In this respect, ubiquitin under these solution conditions might be unrepresentative of higher molecular weight proteins. However, the intended function of this model system was to test the molecular weight limits of NMR experiments and analyses that were later to be implemented with PaDDAH. For these purposes a small, well characterised protein in viscous solution is more preferable subject than a large, poorly characterised protein in H$_2$O. The results presented here show both the utility of the ubiquitin model and that the experimental protocols tested work. Furthermore, the applicability of the ubiquitin in glycerol model system is not limited to the experiments performed here. The model is predicted to have many future uses in testing and development of NMR experiments intended for use on higher molecular weight protein systems.

In summary, isotope-labelled ubiquitin in solutions containing glycerol was proposed as a simple model for high molecular weight proteins. The coarse adjustment in hydrodynamic properties exacted on the addition of glycerol could be further fine tuned with adjustments in the sample temperature. With 50% (v/v) glycerol a range of
rotational correlation times spanning more than 20 ns from approximately 23 ns to approximately 45 ns was achieved between sample temperatures of 17.5 and 34.3°C (Table 4.1). As ubiquitin is a small monomeric protein, all the spectra recorded were uncomplicated and showed good resolution with little signal overlap.

In the following chapter, the large homodimeric protein PaDDAH will be characterised using solution NMR techniques. The results presented in this chapter suggest that it should be possible to obtain backbone resonance assignments of proteins with a rotational correlation time of approximately 35 ns using the 3D triple resonance experiments available in our laboratory. Details of the application of these experiments to PaDDAH and attempts to obtain unambiguous backbone assignment will also be detailed in Chapter 5.
Chapter V
Studying Large Proteins by NMR: WT PaDDAH

Chapter Summary
In this chapter WT PaDDAH is characterised using solution NMR spectroscopy. 2D $^1$H$^N$, $^{15}$N correlation spectra recorded from both $[^{15}$N]- and $[^2$H, $^{15}$N]-labelled WT PaDDAH are presented and discussed. $^1$H and $^{15}$N nuclear spin relaxation, and translational diffusion data are used to investigate the quaternary structure and hydrodynamic properties of WT PaDDAH. NMR data is analysed at multiple sample temperatures and magnetic field strengths. Finally, the acquisition and analysis of triple resonance 3D NMR experiments recorded for $[^2$H, $^{13}$C, $^{15}$N]-labelled WT PaDDAH is presented.

5. Introduction.
PaDDAH is a member of the DDAH family, which has many reported representatives in both bacterial and eukaryotic genomes (Santa Maria et al. 1999). Because of their intimate involvement in the regulation of NOS activity, mammalian DDAH enzymes are attractive therapeutic targets (see Chapter 2, Section 2.2). NMR spectroscopy can provide unique information about protein-ligand interactions, particularly the importance of molecular flexibility and dynamics in these processes. Several methods for rapidly screening protein-binding small molecule ligands using NMR
spectroscopic assays have been reported. The initial level of information required for a given system is dependent on the nature of the NMR assay to be performed. A ligand library can be screened against a protein target using simple 2D heteronuclear NMR spectroscopy. In this paradigm, sometimes referred to as 'SAR-by-NMR', a 2D correlation map of an isotope labelled protein is recorded in the absence and presence of a candidate ligand. Changes in cross peak chemical shifts of the protein in the presence of the ligand are then used to identify the structural specificity and sometimes affinity of a binding interaction. To maximize the information obtained from this protocol, a resonance assigned 2D heteronuclear correlation map and the 3D structure of the target protein are required. Changes in heteronuclear chemical shifts, e.g. \(^1\)H, \(^15\)N, can then be mapped onto the 3D structure of the protein in question.

Fesik and co-workers have exploited this strategy to identify small molecular fragments with low binding affinity that can be chemically linked to form novel, high affinity ligands (Shuker et al. 1996). In the current example of DDAH, a crystal structure of a bacterial homologue has already been solved (Murray-Rust et al. 2001). Therefore, to implement this type of NMR screening protocol would in principle only require resonance assignment of the ‘fingerprint’ \(^1\)H, \(^15\)N correlation spectrum of PaDDAH. Although the crystal structure of PaDDAH shows a 60 kD homodimer (Murray-Rust et al. 2001), proteins of this magnitude have been successfully tackled using recent developments in heteronuclear NMR methodologies (see Chapter 1; McCallum et al. 1999; Salzmann et al. 2000; Kay 2001). Furthermore, the investigations presented in Chapter 4 on ubiquitin in a viscous solution have demonstrated that the equipment, protocols and pulse sequences available in this laboratory should be capable of successfully studying high molecular weight proteins with size of the order of homodimeric PaDDAH.

Previous investigations demonstrated that it was not possible to express sufficient amounts of soluble mammalian DDAH to allow 3D-structural characterisation by either NMR or X-ray crystallographic methods (Dr. M. McAlister personal communication). However, the bacterial homologue, PaDDAH can be expressed and purified to high concentration (Murray-Rust et al. 2001). PaDDAH has 40 % sequence identity to human DDAH isoforms (Santa Maria et al. 1999) and therefore should function as a suitable model for the mammalian enzymes.
In this chapter, WT PaDDAH is characterised by homonuclear 1D and multidimensional heteronuclear NMR methods. Details of ongoing attempts to obtain sequence specific backbone resonance assignments of WT PaDDAH will be discussed.

5.1. Expression and Purification of PaDDAH.

WT PaDDAH was expressed in good yield from both unlabelled LB medium (Figure 5.1(a)) and minimal (M9) medium enriched with [15N]-labelled ammonium sulphate. WT PaDDAH was also expressed from M9 medium enriched with [15N]-labelled ammonium sulphate alone or in concert with uniformly [13C]-labelled glucose prepared in 100 % deuterium oxide (D2O; data not shown). Again, a good yield of labelled WT PaDDAH was obtained but, as would be expected from a less rich medium, the expression levels were lower. PaDDAH was expressed fused to a N-terminal polyhistidine tag. A 20 amino acid ‘linker region’ separated the first residue of the WT PaDDAH sequence and the hexahistidine affinity tag. The 7 residue tobacco etch virus (TEV) protease cleavage site was located within the linker region (see ‘Materials and Methods’ Section 3.2.1). Following cell lysis by French press, WT PaDDAH was initially purified by affinity chromatography using a nickel nitrilotriacetic acid (Ni-NTA) resin (Figure 5.1(a); see ‘Materials and Methods’ Section 3.2.6). PaDDAH was then further purified by preparative scale gel filtration chromatography (Figures 5.1(b) and 5.1(c)). To remove the polyhistidine tag, if required, positive elution fractions from the initial affinity purification step were incubated with recombinant TEV (rTEV) protease (Figure 5.1(d)). Cleaved PaDDAH was then separated from rTEV, the affinity tag and any uncleaved PaDDAH by a further affinity chromatography step (Figure 5.1(e)) before being finally purified by gel filtration chromatography.

The effect of the polyhistidine tag on the hydrodynamic properties of WT PaDDAH was assessed using analytical size exclusion chromatography (SEC). Samples of PaDDAH with the affinity tag either present or removed by proteolytic cleavage were applied to an analytical SEC column at 1 mg/ml loading concentration. A comparison of the elution profiles showed no discernible differences in either the shape of the elution profile or the elution volume (data not shown). Furthermore, the presence of
Figure 5.1. The purification strategy employed for PaDDAH (These results are typical but do not represent one complete purification). (a) SDS-PAGE analysis of affinity chromatography of PaDDAH using His-Bind® resin (Novagen). Lanes from left to right: Before – post lysis; FT – non-binding eluant; W1 to W4 – washes in Buffer A (See ‘Materials and Methods’ Section 3.2.6); W5 and W6 – washes with Buffer A plus 10 mM imidazole; E1 to E4 – elution with Buffer B (250 mM imidazole); M12 – Mark12® protein molecular weight standards (Novex). Poly-His tag fused PaDDAH elutes between 36 and 31.5 kD markers (indicated). (b) An elution profile following preparative scale gel filtration chromatography using a preparative scale S200 gel filtration column (See ‘Materials and Methods’ Section 3.2.6). Vertical grey lines indicate 5 ml fractions collected. (c) SDS-PAGE analysis of elution fractions from (b). (d) Removal of the polyhistidine tag by rTEV protease cleavage. From left to right: M12 – as in (a); Before – E1 to E4 from (a) pooled and concentrated; 48 and 72 – 48 and 72 hour time points. (e) SDS-PAGE analysis of affinity chromatography of the cleavage reaction in (d) using His-Bind® resin. From left to right: M12 – as in (a); Before – results from affinity tag cleavage reaction; FT – non-binding fraction (contains cleaved PaDDAH); E1 – elution with Buffer B. De-tagged PaDDAH runs to approximately 32kD in SDS-PAGE analysis.
the affinity tag did not affect the enzymatic activity of WT PaDDAH (data not shown).

5.2. Analysis of $[^{15}\text{N}]$-labelled WT PaDDAH by NMR.

$[^{1}\text{H}, ^{15}\text{N}]$-HSQC (see ‘Materials and Methods’ Section 3.3.1) and $[^{1}\text{H}, ^{15}\text{N}]$-TROSY (see ‘Materials and Methods’ Section 3.3.2) were recorded on $[^{15}\text{N}]$-labelled samples (see ‘Materials and Methods’ Section 3.2.5). The resulting spectra indicated that WT PaDDAH was folded as was evidenced by the good signal dispersion observed in the proton dimension. The $[^{1}\text{H}, ^{15}\text{N}]$-HSQC spectrum (Figure 5.2 (a)) showed varied peak intensity, severe line broadening and signal overlap in the centre. As was expected for a 60 kD homodimer, the implementation of the $[^{1}\text{H}, ^{15}\text{N}]$-TROSY experiment dramatically improved the quality of the spectrum (Figure 5.2(b)). In particular, the crowded central region of the spectrum showed promising degree of improved resolution. However, a wide range of peak intensities still persisted. This was most noticeable in the region of the TROSY spectrum between proton chemical shifts of 9.5 ppm and 10.5 ppm (Figure 5.2(b)). There are several weak signals that become observable due to the use of the TROSY principle and an increased number of scans per FID. These factors suggested that it would be profitable to express PaDDAH from deuterium enriched medium. A description of the benefits in NMR analysis of perdeuterating non-labile sites in proteins is given in Chapter 1, Section 1.2.1.

5.3. Analysis of $[^{2}\text{H}, ^{15}\text{N}]$-labelled WT PaDDAH by NMR.

5.3.1 – Acquisition of HSQC-type spectra from perdeuterated WT PaDDAH

WT PaDDAH is a 60 kD homodimer and as such represents a challenge to current NMR methodologies. Although NMR has successfully tackled proteins of this size and larger, most resonance assignment strategies have employed deuteration protocols (see Chapter 1, Section 1.2.1).

The expression of $[^{2}\text{H}, ^{15}\text{N}]$-labelled PaDDAH was performed routinely using the protocol outlined in ‘Materials and Methods’ Section 3.2.5. Typical growth curves for each step in the deuterated culture protocol are presented in Figure 5.3.
Figure 5.2. [\textsuperscript{1}H, \textsuperscript{15}N]-HSQC (a) and [\textsuperscript{1}H, \textsuperscript{15}N]-TROSY (b) spectra recorded from [\textsuperscript{1}H,\textsuperscript{15}N]-labelled WT PaDDAH with the affinity tag still presence at 0.8 mM and 25°C. Both data matrices consists of 512* (t\textsubscript{1}) and 1024* (t\textsubscript{2}) points (where n* refers to complex points) with acquisition times of 71.1 ms (t\textsubscript{1}) and 102.4 ms (t\textsubscript{2}). A large indirect spectral width (3600 Hz) was recorded to prevent aliasing of arginine N\textsuperscript{6}H \textsuperscript{15}N resonances. The [\textsuperscript{1}H, \textsuperscript{15}N]-HSQC experiment was recorded with 16 transients per t\textsubscript{1} increment and the [\textsuperscript{1}H, \textsuperscript{15}N]-TROSY experiment with 48 transients per t\textsubscript{1} increment. The inset in (b) shows the region 6.7 to 8.3 ppm (\textsuperscript{1}H) and 81 to 89 ppm (\textsuperscript{15}N) and cross peaks from arginine N\textsuperscript{6}H groups.
Figure 5.3. Typical culture growth curves for *E. coli* BL21(DE3) cells transformed with pPROEX Hta--WT PaDDAH in minimal media prepared in deuterium oxide (See 'Materials and Methods' Section 3.2.5). Bacterial growth was followed by absorbance at 600 nm (OD$_{600}$). (a) Growth data and curve for the 100 ml initial DM9 culture. The value in bold font states the OD$_{600}$ at the point of inoculating the final culture. (b) Growth data and curve for the 500 ml final expression DM9 culture. Protein expression was induced with IPTG, in this instance, at OD$_{600}$ = 0.71 (bold text in table).
Figure 5.4. [\(^1H, ^{15}N\])-HSQC (a) and [\(^1H, ^{15}N\)]-TROSY (b) spectra recorded of [\(^2H, ^{15}N\)]-labelled WT PaDDAH with the affinity tag still present at 0.8 mM and 25°C. Both [\(^1H, ^{12}N\)]-HSQC and [\(^1H, ^{15}N\)]-TROSY data matrices consisted of 400* (t\(_1\)) and 2048* (t\(_2\)) and points (where n* refers to complex points) with acquisition times of 100 ms (t\(_1\)) and 85 ms (t\(_2\)). A total of 32 transients were recorded per t\(_1\) increment. Both spectra were processed with NMRPipe (Delaglio et al., 1995). The proposed correlation of the N\(^{\alpha}\)H group of Trp-207 is indicated in (b).
Both \([1H, ^{15}N]\)-HSQC (Figure 5.4 (a)) and \([1H, ^{15}N]\)-TROSY (Figure 5.4 (b)) spectra of WT PaDDAH showed significant improvement in comparison to their counterparts recorded with \(^{15}N\)-labelled samples (shown in Figure 5.2). However, each spectrum appeared to contain fewer cross peaks than expected. As WT PaDDAH forms a symmetric homodimer, the environment of corresponding nuclei on opposite subunits should be identical. Therefore, a NH correlation spectrum of the homodimer should contain the same number of signals as that of a single protomer. The spectra shown here were recorded from a sample of WT PaDDAH with the affinity tag still present. The primary sequence of this construct (see Figure 3.1) is 284 amino acids in length. Therefore, it would be predicted that a \([1H, ^{15}N]\) correlation map of this construct should contain 268 cross peaks; one for each of the amide NH groups in the polypeptide backbone (i.e. the total complement of amino acid residues corrected for the number of proline residues and the N-terminal \(\alpha\)-amino group). One would also expect to observe NH correlations of certain amino acid side chains including the \(N^2\) and \(N^\%\) groups of asparagine and glutamine residues, respectively, the \(N^{^^H}\) group of tryptophans and the \(N^{1H}\) group of arginines. The maximum number of peaks expected for PaDDAH is therefore 302. With the exception of the \(N^2\) group of arginine side chains, these NH groups have \(1H\) and \(^{15}N\) resonance frequencies similar to backbone amide groups. Consequently, it can prove difficult to differentiate between backbone amide and side chain resonances. WT PaDDAH has only a single tryptophan residue, Trp-207. Using a combination of a 3D-[\(1H, ^{15}N\)]-NOESY HSQC (see ‘Materials and Methods’ Section 3.3.7) and 3D backbone HNC-correlating experiments (see Section 5.6.1), the \(N^{1H}\) group of Trp-207 was tentatively assigned (indicated in Figure 5.4(b)). The side chain amide groups of asparagines and glutamine residues display characteristic ‘aligned’ cross peak patterns in \([1H, ^{15}N]\)-HSQC experiments. As each \(^{15}N\) nucleus has two attached protons, two cross peaks with an identical \(^{15}N\) resonance frequency but typically different \(1H\) frequencies are observed. Such peaks were observed in both \([1H, ^{15}N]\)-HSQC and \([1H, ^{15}N]\)-TROSY spectra (see Figure 5.4 in the region around 7.5 ppm \((1H)\) and 115 ppm \((^{15}N)\)). Although expected in \([1H, ^{15}N]\)-HSQC experiments, cross peaks originating from \(^{15}N\) nuclei with two attached protons are not expected in \([1H, ^{12}N]\)-TROSY spectra (Pervushin et al. 1997). However, such characteristic signals were observed in the \([1H, ^{15}N]\)-TROSY spectra presented in this work. It is possible that these cross peaks result from the presence of NHD groups in glutamine and asparagine side chains. A
maximum of 12 pairs of cross peaks from asparagine and glutamine residues were expected and six such pairs could be readily identified in $[^1\text{H}, ^{15}\text{N}]-\text{HSQC}$ and $[^1\text{H}, ^{15}\text{N}]-\text{TROSY}$ spectra. In this instance it is likely that the reason for only observing half the expected number of cross peaks is due to exchange of the amide proton with the solvent.

Of the remaining side chains that contain nitrogen (arginine, lysine and histidine) only the N$^\text{tH}$ group of arginine will affect the number of peaks observed in the backbone amide region of a NH correlation map. The N$^\text{tH}$ group of arginine has approximately $^1\text{H}$ and $^{15}\text{N}$ chemical shifts of 7.3 ppm and 87 ppm, respectively (values from BioMagResBank). Generally, for WT PaDDAH, the spectral width (SW) in the indirect dimension of NH correlation spectra was set in the region of 2000 Hz. As the $^{15}\text{N}$ nuclei of arginine N$^\text{tH}$ groups have resonance frequencies greater than SW/2, their signals appear at an aliased F$_1$ frequency in the Fourier transformed 2D spectrum. Therefore, it is important to account for the number of arginine N$^\text{tH}$ groups. This can be achieved in two ways. First, a HSQC-type experiment can be acquired with SW set to a value sufficiently large to record the $^{15}\text{N}$ resonance of the N$^\text{tH}$ group without aliasing (Figure 5.2(b, inset)). Second, two identical spectra can be recorded but with different indirect sampling frequencies – i.e. two difference spectral widths. Both methods were employed to account for the arginine N$^\text{tH}$ cross peaks in the tally of cross peaks observed in $[^1\text{H}, ^{15}\text{N}]-\text{TROSY}$ spectra. For WT PaDDAH, eleven arginine N$^\text{tH}$ groups were observed from an expected total of 22 (Figure 5.2(b, inset)).

In many cases NH cross peaks from residues in the affinity tag are not observed in $[^1\text{H}, ^{15}\text{N}]-\text{HSQC}$ or TROSY spectra. The inherent conformational flexibility of residues within the affinity tag is likely to result in exchange broadening of their NMR signals. In addition, due to the high solvent exposure of the amide protons of residues in the tag are likely to be in exchange with the solvent. If the assumption is made that no residue from the affinity tag yields an observable NMR signal then a lower estimate of the number of the observed backbone amide cross peaks can be made. In the case of WT PaDDAH, a full complement of backbone amide cross peaks in a NH correlation spectrum is predicted to be 246 in number.
In many proteins studied by NMR, amide cross peaks are ‘missing’ from NH correlation spectra. Conformational exchange or exchange of amide protons with the solvent are common explanations for unobserved correlations. The NH correlation spectrum of ubiquitin, for example, is ‘missing’ amide cross peaks for residues 24 and 53 (see Chapter 4 Section 4.3). On the basis of such an empirical observation, it is therefore expected to be unlikely that the \([^{1}H, \, ^{15}N]-TROSY\) spectrum of WT PaDDAH would contain a cross peak for every amide group in the protein.

A spectrum of WT PaDDAH should contain 240 backbone amide cross peaks with the polyhistidine tag adding a further maximum of 28. However, the total number of NH correlations counted in Figures 5.4(a) and 5.4(b) believed to correspond to backbone amide NH groups was in the region of 160, a number very significantly fewer than expected.

5.3.2 – Amide protonation of WT PaDDAH following expression in DM9.

Proteins expressed in deuterium enriched media and purified in water-based buffers often yield NMR spectra that do not possess the expected number of amide cross peaks. When the protein is expressed and purified in the native state, certain amide groups with slow solvent exchange rates can remain deuterated, particularly those buried in core regions of the protein. Slow ‘back-exchanging’ amide groups would therefore remain deuterated or fractionally protonated. The time required for complete protonation is dependent on the solvent exchange rate of a particular amide group and can be on the order of days or even weeks (see Chapter 1, Section 1.2.1).

In this investigation it was observed that the quality (in terms of numbers of cross peaks observed) of NH correlation spectra recorded for WT PaDDAH was particularly dependent on the conditions that the sample underwent both during and after purification. An example is shown when the purification of the sample was performed rapidly at 4°C (Figure 5.5). The initial TROSY spectrum recorded at 25°C showed some extremely strong cross peaks in the centre of the spectrum as well as a small number of additional dispersed, but weak intensity cross peaks. The sample was then kept overnight at 25°C, 30°C or 35°C during the acquisition of additional spectra (data not shown). Finally, after eight days a spectrum was recorded at 25°C that
Figure 5.5. Protonation of amide groups following expression of $[^2H,^{15}N]$-labelled PaDDAH from deuterated minimal medium. The scale at the top of the page indicates time from 0 to 8 days. The sample (0.7 mM protein concentration, in 20 mM sodium phosphate buffer, pH 6.0, 100 mM NaCl in H$_2$O) was kept for periods at either 4°C (blue), 25°C (yellow), 30°C (light orange) or 35°C (orange). The $[^1H,^{15}N]$-TROSY experiments presented were recorded after (a) zero days and (b) eight days. ‘zero days’ is defined as the first acquisition of NMR data. Both spectral data matrices consisted of 360* ($t_1$) and 1152* ($t_2$) data points (where n* refers to complex points) with acquisition times of 90 ms ($t_1$) and 63.9 ms ($t_2$). A total of 32 (a) or 16 (b) transients were recorded per $t_1$ increment. Experiments were recorded at 600 MHz proton frequency.
showed considerable improvement in terms of both signal-to-noise and the number of peaks observed (Figure 5.5(b)).

The presence of amide groups in WT PaDDAH with slow solvent exchange rates was further investigated. Figure 5.6(a) shows a \(^1\text{H}, \text{\(^{15}\text{N}\)}\)-TROSY spectrum of \(^{15}\text{N}\)-labelled WT PaDDAH at 1 mM in sodium phosphate buffer prepared with H\(_2\)O. The sample was then diluted ten-fold into an identical sodium phosphate buffer, but prepared with D\(_2\)O rather than H\(_2\)O, and concentrated back to 600 µl. A second \(^1\text{H}, \text{\(^{15}\text{N}\)}\)-TROSY spectrum was recorded with identical parameters. This step was repeated once more to produce a sample containing effectively 1 % H\(_2\)O, 99 % D\(_2\)O (see ‘Materials and Methods’ Section 3.3.9). The initial \(^1\text{H}, \text{\(^{15}\text{N}\)}\)-TROSY spectrum with only 10 % D\(_2\)O (Figure 5.6(a)) showed clearly resolved cross peaks with good signal intensity. On changing the buffer to 1:9 H\(_2\)O:D\(_2\)O, a clear decrease in signal intensity was observed (Figure 5.6(b)). A further ten-fold dilution of the residual H\(_2\)O produced only a slight decrease in signal intensity (Figure 5.6(c)). After a further 14 days, a significant number of NH resonances persisted (Figure 5.6(d)). The data presented in Figure 5.5 suggested a significant proportion of the amide groups of WT PaDDAH were resistant to exchange with the solvent under these conditions.

In the example of slow ‘back protonation’ presented above, incubating the PaDDAH sample at temperatures up to 35°C appeared to promote amide protonation (Figure 5.5). However, this strategy also resulted in noticeable precipitation of the sample. Therefore, a method that was not dependent on high temperature incubation was sought. During the purification of PaDDAH, the lysis and the elution buffers used were both at pH 8.0 but the buffer used in gel filtration, and also that used for the preparation of samples for NMR, was at pH 6.0 or pH 7.0 (see ‘Materials and Methods’ Section 3.2.6). The exchange rate of the proton (or deuteron) of a backbone amide group with a water-based solvent increases at higher pH values (Cavanagh et al. 1995). In protein NMR spectroscopy, the amide group is frequently the focus of attention and therefore it is necessary to perform experiments at a lower pH (e.g. pH < 7.5). At elevated solution pH, the intensity of a NH signal may be decreased as a result of the amide proton undergoing exchange with the solvent. If the solvent exchange rate is in an intermediate regime with respect to the chemical shift.
Figure 5.6. Exchange of amide protons with deuterons (See ‘Materials and Methods’ Section 3.2.11). [\textsuperscript{1H, 15N]}-TROSY spectra of [\textsuperscript{15N}]-labelled PaDDAH recorded from samples in (a) 10% D\textsubscript{2}O; (b) 90% D\textsubscript{2}O; (c) and (d) 99% D\textsubscript{2}O. The time line at the top indicates the age and storage state of the sample. Each wide box represents 24 hours and each narrow box the acquisition of a spectrum (12 hours). Blue indicates the sample was stored at 4°C and yellow, 25°C. Each spectral data matrix consisted of 200\footnote{(1)} \( t_1 \) and 2370\footnote{(1)} \( t_2 \) data points (where \( n^* \) refers to complex points) with acquisition times of 50 ms \( t_1 \) and 118.5 ms \( t_2 \). All spectra were recorded at 600 MHz proton frequency with 160 transients per \( t_1 \) increment.
time scale, the NH signal can broaden to such an extent that it is barely observable above the noise. Furthermore, the methods implemented to suppress the water resonance in NMR spectra produce a large reservoir of water proton magnetisation (Grzesiek and Bax 1993b). Exchange of solvent and amide protons can therefore lead to cross saturation of the amide signals. In the example given above (Figure 5.5), the sample was rapidly purified and spent less than sixteen hours at pH 8.0 and 4°C. Incubation of the sample at higher temperatures was required to improve the spectrum in terms of numbers of amide cross peaks observed. It was found that increasing the period the sample was kept at pH 8.0 considerably improved the appearance of subsequent NH correlation spectra (data not shown). For example, purification protocols that included the removal of the affinity tag (a 72 hour cleavage reaction at pH 8.0) generated samples that yielded high quality spectra at the first acquisition (data not shown).

5.4. Analysis of the solution and heteronuclear relaxation properties of WT PaDDAH by NMR.

The rate of relaxation of transverse magnetisation can limit the success of the experiments used to correlate backbone nuclei (see Chapter 1 Section 1.1.3). It is important, therefore, to gauge an impression of the relaxation properties of a protein before embarking on resonance assignment strategies. The principal driving force of nuclear spin relaxation is the rate of molecular tumbling (see Chapter 4, Section 4.2.1). In the case of a typically isotropic globular protein, the rotational reorientation rate is proportional to its molecular volume. Additional factors, including the degree of structural anisotropy, segmental flexibility, radius of gyration and the hydration shell of the 3D structure, have to be considered (Dingley et al. 1995; Cavanagh et al. 1995). Several NMR experiments have been developed to assess the size, shape and relaxation properties of proteins in solution. Alternative biophysical techniques that assay these properties do exist but they tend to report on protein samples that are too dilute for direct comparison with NMR observations. By exploiting NMR-based methods to assess the hydrodynamic characteristics of a protein, it should be possible to ascertain the hydrodynamic properties of WT PaDDAH under the same solution conditions that future 3D NMR experiments would be recorded at.
5.4.1 - $^1H$ transverse relaxation rates

A crude, qualitative assessment of the rotational correlation time and oligomeric molecular weight of protein can be obtained through the estimation of the average amide proton transverse relaxation time. The amide region of the one dimensional (1D) proton spectrum was chosen for this purpose as the relaxation rates of these protons should better represent the global dynamics of PaDDAH. Side chain aliphatic protons, for example, are considerably more flexible and thus unrepresentative of the global properties.

$^1H$ T$_2$ times were estimated for a sample of undeuterated WT PaDDAH using a 1-1 spin echo 1D NMR experiment (Figure 5.7 (a); see 'Materials and Methods' Section 3.3.3). Two spectra were recorded with the delay time, $\Delta$, set to 0.1 ms and 2.9 ms (Figure 5.7(b)). An average $^1H$ T$_2$ time for the amide proton region of the spectrum was calculated from the proportional decrease in signal intensity between the two spectra using Equation 5.1.

$$\frac{1}{2} (A_B - A_A) = \frac{I_A}{I_B}$$

where $A_A$ and $A_B$ were set to 0.1 ms and 2.9 ms respectively. $I_A$ and $I_B$ are the recorded signal intensities of the amide region with the delays $A_A$ and $A_B$ respectively. Using the $^1H$ T$_2$ time it is also possible to estimate the rotational correlation time ($\tau_c$; Equation 5.2) and from that the apparent molecular weight of a protein (MW; Equation 5.3).

$$\tau_c (\text{ns}) \approx \frac{1}{5 \: T_2 (\text{s})}$$

$$\text{MW (kD)} \approx 2\tau_c (\text{ns})$$

WT PaDDAH yielded a $^1H$ T$_2$ time of 7.35 ms ± 0.14 ms (Figure 5.7(c)). This value corresponds to an apparent rotational correlation time of 27.2 ns and an apparent molecular weight of 54.5 kD. Values of $^1H$ T$_2$, $\tau_c$ and molecular weight obtained from this approach are only approximate but provide a reasonable method of qualitatively determining NMR relaxation parameters under the conditions employed in protein NMR spectroscopy. For example, a value of $\tau_c$ in excess of 15 ns can be used as an upper limit for studying proteins by NMR without opting for perdeuteration. For WT
Figure 5.7. Estimating amide proton transverse relaxation rates using a 1-1 echo 1D NMR experiment. (a) The 1-1 echo pulse sequence (See ‘Materials and Methods’ Section 3.3.3 for complete description): vertical bars indicate 90° $^1$H radiofrequency pulses; $T \sim 100 \mu$s and $\Delta = 0.1$ or 2.9 ms. (b) Example of 1-1 echo spectrum for WT PaDDAH (6 ppm to 10 ppm) recorded at 500 MHz. The upper spectrum is recorded with $\Delta_A = 0.1$ ms and the lower with $\Delta_B = 2.9$ ms (as indicated). (c) Table showing average proportional signal intensity decrease, $I_A/I_B$, and the values and standard deviations calculated using Equations 4.1 to 4.3 (See ‘Materials and Methods’ Section 3.4.1).

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>$I_A/I_B$</td>
<td>2.14</td>
<td>0.03</td>
</tr>
<tr>
<td>$T_2$</td>
<td>7.35 ms</td>
<td>0.14 ms</td>
</tr>
<tr>
<td>$\tau_c$</td>
<td>27.2 ns</td>
<td>0.51 ns</td>
</tr>
<tr>
<td>Apparent MW</td>
<td>54.5 kD</td>
<td>1.03 kD</td>
</tr>
</tbody>
</table>
PaDDAH, $\tau$ is estimated at 27.2 ns indicating that deuteration would be required in order to obtain high quality NMR data. This conclusion is in agreement with the 2D heteronuclear correlation spectra already presented (Figure 5.4).

The values obtained from the 1-1 echo experiment permitted an assessment of the oligomeric state of WT PaDDAH. The crystal structure of C249S PaDDAH reports the enzyme as a homodimer. The apparent molecular weight for WT PaDDAH estimated from Equation 5.3 was 54.5 kD. As this value is closer to the predicted molecular weight of a homodimer (62 kD with affinity tag) than a monomer (31 kD), it is likely that WT PaDDAH is homodimeric under the solution conditions adopted for the NMR measurement. A more detailed analysis of the oligomeric state of PaDDAH is presented in Chapter 6.

5.4.2 – Pulse-field gradient 1D NMR translational diffusion experiments

The translational diffusion coefficient ($D_z$) of a macromolecule in solution can be estimated using NMR methodologies (Altieri et al. 1995; Dingley et al. 1995). Pulse field gradient (PFG) diffusion experiments, which are based on a spin-echo pulse sequence, use gradient pulses to phase-encode nuclei with respect to their z-axis coordinate within the NMR sample tube. A linear magnetic field gradient pulse along the z-axis dephases spin transverse magnetisation depending on the z-axis position. The application of a second, identical pulse after a delay period ($\Delta$) will rephase the transverse magnetisation producing a signal 'echo' after a second delay of $\Delta$. Translational diffusion along the z-axis between gradient pulses will effect overall signal attenuation due to the incomplete rephasing of nuclei by the second gradient pulse. $D_z$ is related to signal attenuation, $A$, by

$$A(t) = A_0 \exp\left[-\gamma_H^2 G^2 D_z \delta^2 (\Delta - \frac{5}{3})\right]$$

Equation 5.4

where $A(t)$ is the signal attenuation at time $t$ and $A_0$ at a reference time, $\gamma_H$ is the gyromagnetic ratio of a proton, $\delta$ is the duration of the gradient pulse, $G$ is the gradient strength (T cm$^{-1}$; Altieri et al. 1995).

A series of 1D PFG diffusion experiments (Figure 5.8(a)) were recorded from a sample of undeuterated WT PaDDAH (0.5 mM) with increasing gradient strengths. Signal attenuation was measured from the resulting 1D proton spectra (Figure 5.8(b);
Figure 5.8. Measuring the translation diffusion coefficient of WT PaDDAH. (a) 1D NMR diffusion stimulated echo experiment incorporating WATERGATE water suppression (enclosed in the dashed box). Experiments were conducted at 500 MHz, 25°C, with 256 scans per FID. Phase cycling is as described in (Altieri, et al., 1995). (b) Example of signal decay in 1D stimulated echo experiment with increasing gradient strength. (c) A graph of signal intensity (between 1.0 and 0.5 ppm) as a function of gradient strength for WT PaDDAH (□) and the calculated fit (—) of Equation 5.4. (d) Residuals of experimental data from the minimised solution of Equation 5.4: residual = ŷᵢ - yᵢ.
see ‘Materials and Methods’ Section 3.3.4). The translational diffusion coefficient \( D_z \) and the initial signal intensity \( A_0 \) were estimated by performing a two parameter least squares regression minimisation of Equation 5.5 against the experimental data (see ‘Materials and Methods’ Section 3.3.4).

\[
\ln(A(t)) = \ln(A_0) - \left( \frac{\Gamma}{2} G^2 D_z \right) \tag{Equation 5.5}
\]

where the constant \( \Pi = \gamma_h^2 \delta^2 (\Delta - \delta/3) \), \( G \) is defined as above \( (G \text{ cm}^{-1}) \), the diffusion delay, \( \Delta \), was set to 70 ms and the duration of the gradient pulse, \( \delta \), was set 5 ms. An evaluation of the quality of this fit was established using a Monte Carlo procedure with 1000 iterations (see ‘Materials and Methods’ Section 3.4.3). Signal intensity was measured at eleven gradient strengths between 8.6 G cm\(^{-1}\) and 51.65 G cm\(^{-1}\) with a step size of 4.30 G cm\(^{-1}\). A plot of the natural log of the signal intensity, \( \ln(A) \), versus the square of the gradient strength, \( G^2 \), did not yield a linear relationship over the full range of gradient strengths (data not shown). At lower gradient strengths the signal appeared to decay more rapidly, which suggested the presence of two diffusing species with markedly different diffusion rates. This could have been due to the presence of glycerol in the sample as a result of the methods used in protein concentration. Consequently, analysis concentrated on higher gradient strength values where a linear relationship was observed for WT PaDDAH and two control proteins ubiquitin and glutathione S-transferase (GST). The truncated data set was fitted against Equation 5.5 and yielded an estimate for the translational diffusion coefficient of WT PaDDAH of \( 7.6 \times 10^{-7} \text{ cm s}^{-1} \) with a standard deviation of \( 1.01 \times 10^{-8} \text{ cm s}^{-1} \) (Table 5.1). Figure 5.5(c) shows the experimental data together with Equation 5.5 plotted using the values of \( D_z \) and \( A_0 \) estimated for WT PaDDAH (Table 5.1). Translational diffusion coefficients were also estimated for ubiquitin and GST using the same methodology (Table 5.1).

5.4.3 — Analysis of \(^{15}\text{N}\) spin relaxation parameters for WT PaDDAH at 25°C

Analysis of spin-lattice \((R_1)\) and spin-spin relaxation \((R_2)\) rates of nitrogen-15 nuclei in the amide groups of a protein provides extremely useful information about the dynamics of the molecule in solution (see Chapter 4, Section 4.2.1).
Table 5.1. Estimated diffusion coefficients.

<table>
<thead>
<tr>
<th>Protein</th>
<th>MW (kD)</th>
<th>Δ (ms)</th>
<th>$D_z$ ($\times 10^6$ cm s$^{-1}$)</th>
<th>Standard Deviation ($\times 10^{-6}$ cm s$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT PaDDAH</td>
<td>64.4</td>
<td>70</td>
<td>0.76</td>
<td>0.01</td>
</tr>
<tr>
<td>Ubiquitin</td>
<td>8.6</td>
<td>60</td>
<td>1.62</td>
<td>0.031</td>
</tr>
<tr>
<td>GST</td>
<td>47.1</td>
<td>120</td>
<td>0.66</td>
<td>0.009</td>
</tr>
</tbody>
</table>

* protein molecular weight calculated from the amino acid sequence. WT PaDDAH and GST have been considered as homodimers; $^b$ experimental delay between gradient pulses; $^c$ diffusion coefficient calculated by fitting the experimental data to Equation 5.5.

$^{15}$N $T_1$, $^{15}$N $T_2$ and $^{1}{^1}$H-$^{15}$N heteronuclear-NOE experiments were recorded on $[^2$H, $^{15}$N]-labelled WT PaDDAH at 25°C using the methods described in Chapter 4, Sections 4.2.1 and 4.2.2. 109 well resolved cross peaks were selected for analysis of relaxation parameters and the results are summarised in Table 5.2. The distribution of cross peak $^{15}$N $R_2/R_1$ values was used to estimate the overall isotropic rotational correlation time of WT PaDDAH (Table 5.2; Figure 5.9; see Chapter 4, Section 4.2.1).

Table 5.2. Average $^{15}$N heteronuclear spin relaxation parameters for WT PaDDAH.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>All $^a$</th>
<th>Filtered $^b$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$^{15}$N $\langle R_1 \rangle$ (s$^{-1}$)</td>
<td>0.48 (0.15)</td>
<td>0.43 (0.05)</td>
</tr>
<tr>
<td>$^{15}$N $\langle R_2 \rangle$ (s$^{-1}$)</td>
<td>49.6 (15.6)</td>
<td>50.9 (6.26)</td>
</tr>
<tr>
<td>$^{1}{^1}$H-$^{15}$N NOE</td>
<td>0.70 (0.27)</td>
<td>0.80 (0.11)</td>
</tr>
<tr>
<td>$^{15}$N $\langle R_2/R_1 \rangle$</td>
<td>111.2 (40.4)</td>
<td>118.9 (19.9)</td>
</tr>
<tr>
<td>$\tau_e$ (ns)</td>
<td>-</td>
<td>35.0 (1.13)</td>
</tr>
</tbody>
</table>

The values in brackets correspond to the standard deviation of each parameter set. $^a$ Average values from all 109 cross peaks; $^b$ average data selected using a NOE and $(R_2/R_1)$ filter (83 cross peaks; Equation 4.3). $^c$ calculated assuming isotropic rotational diffusion and using only filtered data.

5.4.4 - The effect of increased temperature on $R_2$ and signal-to-noise values
The rate of at which magnetisation relaxes is governed by the rate at which a molecule tumbles in solution. The tumbling rate is, in turn, influenced by factors including the shape of the molecule and the viscosity and temperature of the solution.
Figure 5.9. $^{15}$N spin relaxation parameters recorded from [$^{2}$H,$^{15}$N]-labelled WT PaDDAH (0.4 mM). $^{15}$N $R_1$ (a), $^{15}$N $R_2$ (b), $^{15}$N $R_2/R_1$ (c) values were calculated from 109 well resolved NH cross peaks (See 'Materials and Methods' Section 3.4.3 for description of calculations). $^{15}$N $R_2/R_1$ ratios for [$^{2}$H,$^{15}$N]-labelled WT PaDDAH (grey bars) at 25°C and [$^{2}$H,$^{15}$N]-labelled ubiquitin in 50% glycerol at 25°C (white bars) are compared in (c) (see Chapter 4, Section 4.2.4).
(see Chapter 4 Section 4.1). Assuming no temperature-induced aggregation or unfolding, by increasing the temperature of the solution it should be possible to increase the molecular reorientation rate.

\[^{1}\text{H}, \text{ }^{15}\text{N}\]-TROSY reference spectra and \(^{15}\text{N}\) \(T_2\) experiments were recorded on a sample of \(^{1}\text{H}, \text{ }^{15}\text{N}\)-labelled WT PaDDAH incubated at 25°C, 30°C and 35°C. S/N from each \(^{1}\text{H}, \text{ }^{15}\text{N}\)-TROSY spectrum (see 'Materials and Methods' Sections 3.4.1) and \(^{15}\text{N}\) \(R_2\) rates (see 'Materials and Methods' Sections 3.3.5 and 3.4.2) were calculated. Where possible the same cross peaks were used for the calculations of these quantities. However, variations in the 2D NH correlation spectrum at different sample temperature sometimes prevented this (Figure 5.10). As would be expected, the average \(^{15}\text{N}\) \(R_2\) value decreased as the acquisition temperature was increased (Figures 5.11(a) to 5.11(c), left panel). The average \(^{15}\text{N}\) \(R_2\) decreased approximately 20% on changing the sample temperature from 25°C to 35°C. This trend was also generally observed with individual cross peaks (data not shown). The spread of \(^{15}\text{N}\) \(R_2\) values, as indicated by the standard deviations of each \(^{15}\text{N}\) \(R_2\) data set, remained similar over the temperature range studied. The S/N ratio of each cross peak also showed improvement as the temperature was increased (Figures 5.11(a) to 5.11(c), right hand panel).

5.5. Acquisition of spectra at different magnetic field strengths.

The intrinsic sensitivity of the NMR phenomenon is a positive function of the static magnetic field strength (see Chapter 1, Section 1.2.3). \(^{1}\text{H}, \text{ }^{15}\text{N}\)-TROSY experiments were recorded for \(^{15}\text{N}\)-labelled sample of WT PaDDAH at proton frequencies of 500 MHz, 600 MHz and 800 MHz to compare S/N ratios (Figure 5.12; see ‘Materials and Methods’ Section 3.4.1). S/N ratios of 50 well resolved cross peaks in the TROSY spectrum recorded at 500 MHz (Figure 5.12 (a)) were compared to values calculated from identical spectra recorded at the higher magnetic field strengths (Figure 5.13). A significant improvement in signal intensity is achieved when acquiring spectra at 800 MHz (Figure 5.13(a)). The negative values seen for NH cross peaks numbered 22, 25 and 26 are the result of errors during the acquisition of data at 800 MHz (Figure 5.12(c)). An overall improvement in S/N was also observed when data was acquired at 600 MHz in comparison to that obtained at 500 MHz.
Figure 5.10. Comparison of \([\text{H}, \text{N}^\text{15}]-\text{TROSY}\) spectra of \([\text{H}, \text{N}^\text{15}]-\text{labelled WT PaDDAH}\) at (a) 25°C, (b) 30°C and (c) 35°C. Each data matrix consisted of 256\(^*\) \((t_1)\) and 2048\(^*\) \((t_2)\) data points (where \(n^*\) refers to complex points) with acquisitions times of 75.3 ms \((t_1)\) and 128 ms \((t_2)\). All spectra were recorded at 600 MHz with 16 transients per \(t_1\) increment.
Figure 5.11. Comparison of $^{15}$N transverse relaxation rates ($R_2$; left) and signal-to-noise ratios (S/N; right) of WT PaDDAH at (a) 25°C, (b) 30°C and (c) 35°C. The average $^{15}$N $R_2$, $\langle R_2 \rangle$ and signal-to-noise, $\langle S/N \rangle$ are given with standard deviation (in brackets in each case). Only values within 1.5 standard deviations of $\langle R_2 \rangle$ and 1 standard deviation of $\langle S/N \rangle$ were used in the calculation of these averages.
Figure 5.12. [\(^1\text{H},^{15}\text{N}\)]-TROSY spectra recorded on [\(^1\text{H},^{15}\text{N}\)]-labelled WT PaDDAH (25°C, 0.5 mM) at three magnet field strengths: (a) 500 MHz, (b) 600 MHz and (c) 800 MHz proton resonance frequency. All spectra were recorded with 128 transients per t\(_1\) increment in the proton dimension and within five days of each other. The data matrix for (a) consists of 256* (t\(_1\)) and 2134 (t\(_2\)) data points (where n* refers to complex points) with acquisition times of 64 ms (t\(_1\)) and 128 ms (t\(_2\)). The data matrix for (b) consists of 256* (t\(_1\)) and 2560 (t\(_2\)) complex data points with acquisition times of 64 ms (t\(_1\)) and 128 ms (t\(_2\)). The data matrix for (c) consists of 256* (t\(_1\)) and 3414 (t\(_2\)) complex data points with acquisition times of 48 ms (t\(_1\)) and 128 ms (t\(_2\)). The cross peaks labelled on (a) were those used in the calculation of signal-to-noise ratios in all three spectra.
Figure 5.13. Comparison of amide crosspeak signal-to-noise (S/N) values calculated from $[^{1}H,^{15}N]$-TROSY spectra recorded on $[^{15}N]$-labelled WT PaDDAH (0.5 mM) at 500 MHz, 600 MHz and 800 MHz proton resonance frequency. (a) Comparison of S/N at 500 MHz and 800 MHz; (b) Comparison of S/N at 500 MHz and 600 MHz. Relative S/N = $(S/N_X - S/N_{500}) / (S/N_{500})$ where X = 600 MHz or 800 MHz. NH crosspeak The dashed black line in each graph shows the average relative S/N for each compared data set. S/N values were calculated using ANSIG (Kraulis 1989) from well resolved NH cross peaks selected from the TROSY spectrum recorded at 500MHz (indicated in Figure 5.12(a); see ‘Materials and Methods’ Section 3.4.3).
5.6. Towards residue specific resonance assignment of WT PaDDAH.

5.6.1 Acquisition of backbone $^1\text{H}$, $^{15}\text{N}$, $^{13}\text{C}_\alpha$ and $^{13}\text{C}_\beta$ correlations

The experiments conducted thus far, using both protonated and deuterated samples, have shown that deuteration in concert with TROSY methodologies is critical for achieving good signal intensity and cross peak resolution in 2D NH correlation maps of WT PaDDAH. Thus, a sample of [3H, $^{13}$C, $^{15}$N]-labelled WT PaDDAH with affinity tag still present was produced using the protocol presented in ‘Materials and Methods’ Section 3.2.5. A final sample concentration of 0.7 mM (monomer concentration) was achieved in 300 μL. A series of TROSY-based 3D NMR experiments were recorded that correlate $^1\text{H}$, $^{15}\text{N}$ resonance frequencies to backbone $^{13}\text{C}$ chemical shifts (Figure 5.14; Table 5.3; ‘Materials and Methods’ Section 3.3.6).

5.6.2 Initial efforts towards sequential resonance assignment

Following processing of the 3D data sets (see ‘Materials and Methods’ Section 3.3.8), backbone chemical assignment was initiated using AZARA version 2.7 (Boucher 2002) and ANSIG (Kraulis 1991) software packages.

In proteins, glycine, alanine, serine and threonine residues have distinctive and easily identifiable $^{13}\text{C}_\alpha$ or $^{13}\text{C}_\beta$ chemical shifts (Figure 5.15(a); Grzesiek and Bax 1993a; Wishart and Sykes 1994). WT PaDDAH contains several of each of these residues and close scrutiny of the primary sequence reveals numerous di- and tripeptide groupings of these residues types (Figures 5.15(b) and 5.15(c)). Such short sequences should provide a convenient starting point for sequence-specific chemical shift assignment. It should be possible to assign residue types (e.g. glycine, alanine or serine or threonine) to certain cross peaks in the NH correlation spectrum of WT PaDDAH on the basis of $^{13}\text{C}$ cross peak frequencies. It should be noted, however, that it is not possible to unambiguously differentiate between serine and threonine spin systems using $^{13}\text{C}_\alpha$ and $^{13}\text{C}_\beta$ chemical shifts alone (Grzesiek and Bax 1993a). Analysis of the 3D correlation spectra revealed considerably less than the full complement of spin systems characteristic of these types of residue (Figure 5.15(c)).
Figure 5.14. Experimental strategy for the resonance assignment of backbone $^1$H, $^{15}$N and $^{13}$Ca, and $^{13}$Cβ nuclei. (a) Magnetisation transfer pathways in HNCA (top) and HN(CO)CA (bottom) experiments and a schematic spectrum. (b) Magnetisation transfer pathways in HN(CA)CB (top) and HN(COCA)CB (bottom) experiments and a schematic spectrum. All four experiments shown are of the 'out-and-back' variety with magnetization originating and ending on the amide proton. Only the outward transfer pathways are shown by arrows in each case.
Table 5.3. Summary of NMR experiments conducted to facilitate backbone chemical shift assignment.

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Transients per $t_1$ increment</th>
<th>Number of complex points acquired $^b$</th>
<th>Correlations $^c$</th>
<th>Reference $^d$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$[^1H, ^{15}N]$ TROSY $^e$</td>
<td>64</td>
<td>256 ($^{15}N$) 2048 ($^1H$) -</td>
<td>$N_H, H_N$</td>
<td>Pervushin et al., 1997</td>
</tr>
<tr>
<td>$[^1H, ^{13}C, ^{15}N]$ HNCA $^f$</td>
<td>24</td>
<td>136 ($^{13}C$) 64 ($^{15}N$) 2048 ($^1H$)</td>
<td>$N_H, H_N, C_{\alpha}, C_{\alpha+1}$</td>
<td>Yang and Kay (1999a)</td>
</tr>
<tr>
<td>$[^1H, ^{13}C, ^{15}N]$ HN(CO)CA $^f$</td>
<td>32</td>
<td>140 ($^{13}C$) 64 ($^{15}N$) 2048 ($^1H$)</td>
<td>$N_H, H_N, C_{\alpha}, C_{\alpha+1}$</td>
<td>Yang and Kay (1999a)</td>
</tr>
<tr>
<td>$[^1H, ^{13}C, ^{15}N]$ HN(CA)CB $^f$</td>
<td>32</td>
<td>136 ($^{13}C$) 64 ($^{15}N$) 2048 ($^1H$)</td>
<td>$N_H, H_N, C_{\beta}, C_{\beta+1}$</td>
<td>Yang and Kay (1999a)</td>
</tr>
<tr>
<td>$[^1H, ^{13}C, ^{15}N]$ HN(COCA)CB $^f$</td>
<td>32</td>
<td>136 ($^{13}C$) 64 ($^{15}N$) 2048 ($^1H$)</td>
<td>$N_H, H_N, C_{\beta}, C_{\beta+1}$</td>
<td>Yang and Kay (1999a)</td>
</tr>
<tr>
<td>$[^1H, ^{13}C, ^{15}N]$ HNCO $^f$</td>
<td>24</td>
<td>96 ($^{13}C$) 64 ($^{15}N$) 2048 ($^1H$)</td>
<td>$N_H, H_N, CO_{\beta+1}$</td>
<td>Yang and Kay (1999a)</td>
</tr>
</tbody>
</table>

$^a$ The curved parentheses indicate correlations not observed; $^b$ during spectral processing of 3D data sets the resolution in all indirect dimensions was increased by linear predicting to twice the number of acquired points; $^c$ 'Correlations' lists the expected correlations created by the experiment; $^d$ most experiments were based on those cited but optimized in-house; $^e$ $^{13}C$ decoupling was employed during $t_1$ and proton acquisition. $^f$ all 3D pulse sequences employed TROSY selection. No deuterium decoupling was performed on any experiment. All experiments were performed at 25°C and at 600 MHz proton resonance frequency (See 'Materials and Methods' Section 3.3.6).
Figure 5.15. Amino acids with characteristic $^{13}$Cα and $^{13}$Cβ chemical shifts in proteins. (a) distribution of $^{13}$Cα and $^{13}$Cβ chemical shifts of glycine, alanine, serine and threonine residues in proteins (Taken from BioMagResBank chemical shift database); (b) position and groupings of glycine (grey shading), alanine (green), serine (yellow) and threonine (red) residues in the amino acid sequence of WT PaDDAH (residues 1 to 254) and polyhistidine tag (residues -30 to -1; upper line); (c) table detailing the numbers of glycine, alanine, and serine or threonine in the primary sequence and the number of intra- and inter-residue crosspeaks observed.
A more comprehensive investigation of the number of carbon correlations realised for each NH cross peak in the reference TROSY spectrum also showed considerably fewer than the number expected (Figure 5.16).

Efforts towards resonance assignment persisted with some NH cross peaks giving rise to identifiable carbon spin systems (Figure 5.17(a)). Five NH cross peaks have been selected from a region of a $[^1\text{H}, ^{15}\text{N}]$ TROSY spectrum of WT PaDDAH. ANSIG session strips showing the $^{13}\text{C}$ dimension from the $^{15}\text{N}, ^1\text{H}$ frequency coordinate are also presented. For cross peaks 1, 2, 4 and 5 it is possible to identify either the residue itself or the one that precedes it solely on the basis of $^{13}\text{C}$ chemical shift (Figure 5.17(b)). It is noteworthy that some amide cross peaks with equal S/N ratios to those exampled in Figure 5.17 yield no carbon correlations (data not shown).

In many cases the $^{13}\text{C}_\alpha$ or $^{13}\text{C}_\beta$ chemical shifts correlated to a NH cross peak do not individually allow unambiguous identification of the residue type. In this situation, a combination of the $^{13}\text{C}_\alpha$ or $^{13}\text{C}_\beta$ chemical shift values can be used to generate a list of potential residue types (Grzesiek and Bax 1993a). On a single residue basis the probability statistics returned are unlikely to positively identify the residue type. However, when residue spin systems are connected in extended sequences, it is possible to 'add up' the residue type probabilities. Cross-referencing these statistics to the amino acid sequence of the protein can generate a high level of confidence regarding sequential resonance assignment. Increasing the number of sequentially connected spin systems further increases the probability of identifying a unique sequence within the protein.

Where carbon correlations were available, resonance assignment of the primary sequence of WT PaDDAH was attempted using the $^{13}\text{C}$ chemical shift-based methods detailed above. In some instances 'unambiguous' assignment of short stretches of the amino acid sequence of WT PaDDAH to specific NH cross peaks was possible (data not shown). Invariably, however, the low percentage of NH cross peaks with a full complement of carbon correlations prevented the assignment of extended regions of the amino acid sequence. Under these circumstances, it appeared sensible to conclude the confidence level that could be applied to these relatively small numbers of 'unambiguous' assignments is low, and therefore these resonance assignments have
### Table 5.16. Observed $^{1}H^{N}, ^{15}N_{i} \rightarrow ^{13}C$ correlations in 3D data spectra.

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Correlation $^{a}$</th>
<th>Residue</th>
<th>Number of cross peaks</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>($^{1}H^{15}N \rightarrow ^{13}C$)</td>
<td></td>
<td>Expected $^{b}$</td>
</tr>
<tr>
<td>HNCA</td>
<td>$\alpha$</td>
<td>$i$</td>
<td>268</td>
</tr>
<tr>
<td></td>
<td>$\alpha$</td>
<td>$i - 1$</td>
<td>268</td>
</tr>
<tr>
<td>HN(CO)CA</td>
<td>$\alpha$</td>
<td>$i - 1$</td>
<td>268</td>
</tr>
<tr>
<td>HN(CA)CB</td>
<td>$\beta$</td>
<td>$i$</td>
<td>249</td>
</tr>
<tr>
<td></td>
<td>$\beta$</td>
<td>$i - 1$</td>
<td>249</td>
</tr>
<tr>
<td>HN(COCA)CB</td>
<td>$\beta$</td>
<td>$i - 1$</td>
<td>249</td>
</tr>
<tr>
<td>HNCO</td>
<td>$i - 1$</td>
<td></td>
<td>268</td>
</tr>
</tbody>
</table>

$^{a}$ $\alpha$ or $\beta$ refers to $^{13}C\alpha$ or $^{13}C\beta$ correlations; $^{b}$ number of expected cross peaks was calculated from the primary sequence of WT PaDDAH accounting for the N-terminus and proline residues; $^{c}$ The number of cross peaks observed in 3D carbon strip plots. Inter- and intra-residue cross peak overlap was not considered when counting cross peaks and, thus, may account for some of the discrepancies between the numbers of observed and expected cross peaks.

**Figure 5.16.** Observed $^{1}H^{N}, ^{15}N_{i} \rightarrow ^{13}C$ correlations in 3D data spectra. (a) table showing the number of expected and observed carbon correlations per spectrum (see Table 5.3 for details of experiments conducted). $i$ indicates an intra-residue correlation and $i - 1$ indicates an inter-residue correlation to the preceding residue; (b) a pie chart showing the observed number of carbon correlations (minimum: 0; maximum: 6) and the percentage of NH cross peaks that yield that number. Note this does not include $^{1}H^{N}, ^{15}N_{i} \rightarrow ^{13}C_{i+1}$ correlations from the HNCO experiment.
Figure 5.17. Examples of $^1$H, $^{15}$N$_i$ → $^{13}$C correlations visualised using ANSIG (Kraulis 1989). NH cross peaks numbered in the region of the [$^1$H, $^{15}$N]-TROSY spectrum shown (Left) give rise to corresponding carbon spin systems (Right). The vertical axis of the carbon strips are $^{13}$C chemical shift (ppm) and the horizontal $^1$H chemical shift (ppm). The vertical dashed line corresponds to the $^1$H chemical shift of the NH cross peak selected in the [$^1$H, $^{15}$N]-TROSY spectrum (left). The carbon correlations (right hand panels) are labelled with carbon atom ($\alpha$ or $\beta$) and for the intra- ($i$) and inter-residue ($i-1$) cross peak; (b) table showing the residue types identified (where possible) by $^{13}$C chemical shifts alone. A '?' indicates this was not possible.
not been reported here. The explanation behind the 'missing' correlations in both 2D NH and 3D triple resonance correlation spectra is the subject of on-going investigation.

5.6. Discussion.

A substantial improvement in both spectral resolution and S/N was observed in \(^1\text{H}, ^{15}\text{N}\)-TROSY spectra of WT PaDDAH on perdeuteration. However, the principal concern arising from deuteration was the low number of cross peaks observed in \(^1\text{H}, ^{15}\text{N}\)-correlation spectra compared to the number expected. The available explanations for this observation also hamper the search for a solution. Missing amide cross peaks in NMR spectra of deuterated proteins can be attributed to two possible causes. First, if the protein has been purified in the native state, amide groups buried in the core of the protein often will remain deuterated long after purification in H\(_2\)O-based buffers. Second, chemical or conformational exchange occurring at an intermediate rate with respect to chemical shift time scale will cause acute line broadening. In the worse case scenario, the resonance signal of a nucleus experiencing a dynamic chemical environment could disappear completely.

The problem of incomplete protonation or 'back exchange' in proteins expressed in D\(_2\)O but purified in H\(_2\)O is well documented (see Chapter 1, Section 1.2.1 for references). \(^1\text{H}, ^{15}\text{N}\)-TROSY spectra of WT PaDDAH recorded at different times following purification showed an appreciable increase in the number of NH correlations observed (Figure 5.5). Keeping the protein at pH 8 also increased the numbers of peaks observed (data not shown). In a sample of \(^{15}\text{N}\)-labelled WT PaDDAH exchanged into deuterated buffer, a considerable number of NH cross peaks were present after two weeks (see Figure 5.6). Therefore it is apparent that incomplete amide protonation is a partial explanation for the low number of amide cross peaks observed in the perdeuterated sample. The size of WT PaDDAH means that it is difficult to accurately count the number of observed amide cross peaks in a \(^1\text{H}, ^{15}\text{N}\)-correlation spectra recorded from a \(^{15}\text{N}\)-labelled sample. This is because a sample enriched only with \(^{15}\text{N}\) will suffer from rapid relaxation of \(^{15}\text{N}\) and \(^1\text{H}\) magnetisation through spin diffusion pathways leading to broadened signal linewidths (Figure 5.2). Although \(^1\text{H}, ^{15}\text{N}\)-TROSY experiments can be applied to non-deuterated proteins,
only relaxation caused by heteronuclear interaction between $^{15}$N and $^1$H nuclei is suppressed. Homonuclear dipolar interactions between neighbouring protons are not suppressed and will therefore reduce the signal intensity of the selected component (Pervushin et al. 1997). It is therefore expected that a [$^1$H, $^{15}$N]-HSQC experiment would generate a spectrum with greater NH signal intensity than an equivalent [$^1$H, $^{15}$N]-TROSY experiment, but analysis of the spectrum would be impeded by broad NH linewidths. In short, neither TROSY nor HSQC experiments of [$^{15}$N]-labelled WT PaDDAH produced spectra of a sufficient quality to permit accurate determination of the number of correlations observed. At a best approximation a [$^1$H, $^{15}$N]-TROSY of [$^{15}$N]-labelled WT PaDDAH contains around 220 amide correlations. This value is still below that expected from the amino acid sequence of the protein.

Mass spectrometry (MS) is an analytical technique that is well suited to detect small increases in the molecular mass of a protein. However, MS cannot be used to identify how many amide groups have remained deuterated. The level of deuteron incorporation into proteins that are expressed in D$_2$O-based media varies greatly and is dependent on the particular atoms in the polypeptide chain (Gardner and Kay 1998). This phenomenon arises particularly when protonated glucose is used as the sole carbon source, as is typically the case for economical reasons. Metabolic processes in the host cell will incorporate protonated aliphatic groups into the amino acid reservoir, which will subsequently be utilised in protein expression. This is particularly noticeable for aromatic groups in certain residues. As a result, the level of deuteration varies greatly both between and within amino acids. The Ca position, for example, is usually deuterated at the same percentage as the medium the protein is expressed in. However, the percentage of deuteration in the side chain decreases at carbon positions further from the Ca, resulting in the presence of a large number of isotopomers in the protein sample. Consequently, it is statistically improbable that two protein molecules in a sample will have identical isotope incorporation. Because of this distribution of isotopomers and the small mass difference between a fully amide protonated and a fractionally amide deuterated protein, MS cannot be used to ascertain the number of deuterated amide groups.

The structure of PaDDAH was reported as a homodimer (Murray-Rust et al. 2001). Furthermore, X-ray diffraction data from crystals of apo WT PaDDAH showed a loop
region (residues 14 to 25) with weak electron density. A similar observation for the homologous region in Arg:Gly amidinotransferase was also noted (Humm et al. 1997b). The crystal structure of PaDDAH published is that of a C249S mutant in complex with L-citrulline. The reported presence of a flexible loop and the observation that WT PaDDAH is in dynamic equilibrium between homodimer and monomer (see Chapter 6) suggest that some residues may be undergoing chemical exchange under the conditions of the experiment. If a nucleus is in exchange between two different chemical environments with resonance frequencies that differ by $\Delta \nu$ the NMR spectrum observed will be dependent on the exchange rate, $k_{ex}$, with respective signal intensities proportional to the population of each species. If $k_{ex} \ll \Delta \nu$, i.e. slow exchange on the chemical shift time scale, the spectrum will show signals at both resonance frequencies. When $k_{ex} \gg \Delta \nu$, i.e. fast exchange, a single signal is seen at the population weighted average chemical shift of the two sites. However, when $k_{ex} \approx \Delta \nu$, i.e. the intermediate exchange regime, the signal broadens. In certain cases this can result in the complete loss of the signal (Cavanagh et al. 1995). In the case of PaDDAH where residues are possibly undergoing conformational exchange in flexible loops or at the site of the homodimer interface, extreme line broadening could occur if the exchange rate is in the appropriate regime. Because $\Delta \nu$ is dependent on the magnetic field strength, comparing spectra recorded at different $B_0$ could indicate the existence of chemical exchange. S/N ratios calculated from identical [^H, ^15N]-TROSY experiments of [^15N]-labelled WT PaDDAH recorded at three magnetic fields did not reveal significant differences between cross peak intensities (Figures 5.12 and 5.13). However, the S/N ratio is not the best parameter for comparison between different spectrometers. A better comparison would have been permitted if $R_2$ values were recorded at different magnetic field strengths. The equation describing the rate of transverse relaxation has both frequency dependent and independent components (see Chapter 4, Section 4.2.1). Therefore in the absence of chemical exchange the $R_2$ value of a given $^{15}$N nucleus should scale proportionally with $B_0$, whereas a nucleus undergoing chemical exchange should not. An alternative strategy involves measuring $R_{1p}$ values where an off-resonance spin-lock period is applied during the relaxation delay (Akke and Palmer 1996; Korzhnev et al. 2002). A value of $R_2$ can be calculated from the experimental $R_{1p}$ value. If experimental and $R_{1p}$-derived values of $R_2$ for a
particular nucleus are not equal there is evidence of chemical exchange occurring on the μs to ms timescale.

Magnetisation transfer efficiency in the triple resonance experiments was poor for triple labelled PaDDAH. The number of HN→C correlations observed in each spectrum was prohibitively less than the number expected (Figure 5.16(a)). Only 65 % of the expected \(^{13}\)C correlations were observed in the \(^{1}\)H, \(^{13}\)C, \(^{15}\)N]-TROSY HNCO, 45 % in the \(^{1}\)H, \(^{13}\)C, \(^{15}\)N]-TROSY HNCA (both i and i – 1) and 32 % in the \(^{1}\)H, \(^{13}\)C, \(^{15}\)N]-TROSY HN(CA)CB (i and i – 1). The \(^{1}\)H, \(^{13}\)C] plane of a \(^{1}\)H, \(^{13}\)C, \(^{15}\)N]-TROSY HN(CA)CO was also recorded but showed only very weak signal intensity. Of the amide cross peaks present in the \(^{1}\)H, \(^{15}\)N]-TROSY spectrum used for resonance assignment, only 20 % gave the maximum possible number of \(^{13}\)C correlations (Figure 4.16(b)). Although the NH correlation spectrum did not show a full complement of amide cross peaks, this does not explain why such a small number of amide cross peaks yielded the maximum number of \(^{13}\)C correlations. Triple labelled WT PaDDAH samples were expressed from media containing protonated carbon \(^{13}\)C]-labelled glucose as the sole carbon source. Therefore varying levels of aliphatic deuteration would have been present as a result of the metabolic incorporation of the glucose protons into amino acid side chains. The low \(^{13}\)C sensitivity observed in the 3D experiments may result from this incomplete aliphatic deuteration. Therefore it may be possible to decrease \(^{13}\)C transverse relaxation rates further still by using \(^{2}\)H, \(^{13}\)C]-labelled glucose. It is possible that missing amide cross peaks, possibly the result of incomplete re-protonation or chemical exchange broadening, can explain why so few alanine, glycine and serine or threonine residues were observed (Figure 5.15(c)).

Shan and colleagues reported low numbers of \(^{13}\)C correlations in triple resonance data recorded from a 64 kD trp repressor/operator complex, but still achieved greater than 90 % backbone \(^{1}\)H, \(^{13}\)C, \(^{15}\)N resonance assignment (Shan et al. 1996). With a full complement of amide cross peaks it may have been possible to proceed with resonance assignment of PaDDAH using the data presented. However, with such a large percentage of HNC spin systems missing, this was not the case.
There are several strategies that could be applied to resolve the problem of missing amide cross peaks. The issue of incomplete ‘back protonation’ would be resolved if perdeuterated WT PaDDAH was unfolded and refolded in a H₂O-based buffer during purification. Kay and co-workers unfolded and refolded MSG prior to NMR analysis to ensure full amide protonation (Tugarinov et al. 2002). Initial trials in our laboratory suggest that WT PaDDAH can be denatured in 6 M guanidinium hydrochloride and refolded under certain buffer conditions. Refolded WT PaDDAH samples have been shown to be enzymatically active (B. Simas Magalhaes personal communication). Alternatively, the protein could be expressed in an unfolded state into inclusion bodies followed by refolding during purification. This latter method was employed by Salzmann and co-workers (Salzmann et al. 2000). It is also possible that complete unfolding of WT PaDDAH is not required to effect more efficient protonation. Only a relatively low concentration of denaturant may be needed to relax the structure sufficiently to allow an increase amide group solvent accessibility. As has been shown here, WT PaDDAH cannot simply be expressed and purified in the native state and the issue of complete amide group protonation needs to be addressed prior to future attempts at resonance assignment.

As mentioned above, a proportion of the NH cross peaks could be exchange broadened due to conformational mobility of a flexible loop or structural plasticity of the homodimer interface. The flexible loop mentioned by Murray-Rust et al. (residues 14-25) covers the active site of PaDDAH and is believed to become ordered on ligand binding (Murray-Rust et al. 2001). This occurrence is analogous to the situation with amidinotransferases (Humm et al. 1997b; Fritsche et al. 1998). Therefore NH correlation maps of ligand-bound PaDDAH may reveal additional amide cross peaks. There are several ligands that could potentially be used with the WT enzyme: L-citrulline, the enzymatic reaction product; 4124W, an inhibitor of human DDAH (MacAllister et al. 1996); L-arginine, a substrate analogue; SDMA, a non-substrate methylarginine. Alternatively, the enzymatically inactive mutant C249S PaDDAH could be used in conjunction with those compounds listed above or the two asymmetric methylarginine substrates of the WT enzyme.

PaDDAH displays a concentration dependent elution volume in analytical size exclusion chromatography indicating the existence of a dynamic equilibrium between
homodimer and monomer forms (Murray-Rust et al. 2001). With knowledge of the structure of the homodimer interface, it may be possible to design mutants that either stabilise the homodimer or generate a non-associating monomer. A stable monomeric species of PaDDAH would not only alleviate the problems of oligomer exchange but also reduce the effective size of the protein: an 'isotropic' 30 kD monomeric protein represents a far less challenging target for NMR than a 60 kD anisotropic homodimer. With this concept in mind, the following chapter details the effects of introducing mutations at the homodimer interface of PaDDAH.
Chapter VI
Exploring the Homodimer-Monomer Equilibrium of PaDDAH

Chapter Summary.
In the current chapter the solution and self-association properties of WT PaDDAH are investigated using analytical size exclusion chromatography (SEC). A dynamic equilibrium between homodimer and monomer and an approximate dissociation constant are reported for WT PaDDAH. Amino acid residues that potentially participate in the homodimer interaction are identified from the 3D structure of PaDDAH and assessed by site directed mutagenesis. The enzymatic activity, solubility and self-association properties of eleven homodimer interface mutants are explored. Conclusions are drawn about the oligomeric state of these mutants in comparison to the WT enzyme.

6.1. Introduction.
On the publication of the crystal structure of WT PaDDAH, the authors briefly mentioned an observation that the elution volume of the protein in analytical SEC experiments was dependent on the concentration of the sample applied to the column
(Murray-Rust et al. 2001). They attributed this phenomenon to the existence of an equilibrium between homodimeric and monomeric forms of WT PaDDAH.

Analytical SEC is a technique commonly used to characterise the hydrodynamic properties of a protein. In SEC a protein sample is passed over a chemically inert porous matrix that separates the component molecules with respect to their size or hydrodynamic radius. The separation matrix employed later in this chapter is Superdex 75 (S75; see ‘Materials and Methods’ Section 3.2.10), which is composed of a small beads (average diameter of 13 μm) made of a highly cross-linked agarose matrix with covalently linked dextran. The volume available for diffusion inside the SEC matrix to small molecules is greater than that available for larger ones. Therefore, larger molecules will pass through the column more quickly; they will have a faster elution time or smaller elution volume. The S75 gel medium has an optimum separating capacity for proteins between 3 and 70 kD in mass. It is important to note that the relationship between protein mass and hydrodynamic radius is only approximate. Many proteins have non-isotropic 3D structures and will therefore not behave during SEC as predicted from the molecular mass alone.

Though the use of gel filtration it is relatively straightforward to establish the apparent oligomeric state of a protein under a variety of solution conditions. In addition, there are several reported examples where analytical SEC has been used to estimate the dissociation constant $K_d$ for a reversibly self-associating system. For example, Manning and co-workers described the use of an analytical Superose-12 gel filtration column to estimate $K_d$ for the subunit dissociation of haemoglobin homotetramers into homodimers (Manning et al. 1996, Manning et al. 1999). The $K_d$ values published from studies using analytical SEC are comparable to those obtained by other techniques such as sedimentation equilibrium analytical ultracentrifugation and fluorescence depolarisation (Manning et al. 1996, Dorfman and Walsh 2001).
6.2. Analysis of the Self Association Properties of WT PaDDAH.

6.2.1 - Analytical size exclusion chromatography: WT PaDDAH
A high resolution 10/30 Superdex75 (S75) analytical gel filtration column was employed to investigate the oligomeric state of WT PaDDAH. 100 µl samples of WT PaDDAH over a 100-fold concentration range (5 mg/ml to 50 µg/ml) were applied to a pre-equilibrated S75 column and the elution of protein followed by absorbance at 280 nm (see ‘Materials and Methods’ Section 3.2.10). At each concentration, the elution profile of WT PaDDAH showed a single peak. As shown in Figures 6.1(a) and 6.1(b) a clear increase in elution volume from 10.23 ± 0.02 ml to 10.64 ± 0.02 ml was observed as the protein concentration is decreased. Over the course of this study, a total of eight separate SEC experiments were recorded per WT PaDDAH concentration and this behaviour was completely reproducible (Figure 6.1(c)).

The crystal structure of PaDDAH shows numerous free cysteine residues at the surface of the protein. To ensure the behaviour observed in SEC analyses was not due to the formation of transient inter-protomer disulphide bonds, analytical SEC runs were repeated in the presence of a reducing agent, β-mercaptoethanol (βME; see ‘Materials and Methods’ Section 3.2.10). The elution volume of WT PaDDAH in reducing solution conditions did not differ from that in non-reducing conditions (data not shown).

6.2.2 - Analytical size exclusion chromatography: ovalbumin and GST
The proteins ovalbumin (43.5 kD), a monomer, and glutathione S-transferase (GST; 51 kD), an obligate homodimer, were chosen for comparison with WT PaDDAH (see ‘Materials and Methods’ Section 3.2.10). 100 µl samples of each protein over a 100-fold concentration range (5 mg/ml to 50 µg/ml) were applied to a pre-equilibrated S75 column and protein elution followed by absorbance at 280 nm. The elution profiles of both ovalbumin and GST displayed two peaks (Figures 6.2(a) and 6.2(b) respectively). In both cases, the ratio of peak areas for the major and minor peaks was invariant (data not shown). The subsequent analysis of the hydrodynamic properties of ovalbumin and GST deals only with the major peak in each elution profile. The elution volume of both Ovalbumin and GST, over a similar range of sample
Figure 6.1. Analytical size exclusion chromatography of WT PaDDAH. (a) Elution profile of WT PaDDAH samples with loading concentrations of 5, 1, 0.5, 0.25, 0.1 and 0.05 mg/ml; (b) expansion of (a) with loading concentrations (mg/ml) annotated; (c) Summary of elution volume compared to loading concentration (mg/ml) for WT PaDDAH. The error bars indicate plus and minus one standard error from the mean with n = 8.
Figure 6.2. Analytical size exclusion chromatography of ovalbumin and glutathione S-transferase (GST). Elution profiles of ovalbumin (a) and GST (b) samples with loading concentrations of 5, 1, 0.5, 0.25, 0.1 and 0.05 mg/ml; Summary of elution volume (ml) compared to loading concentration (mg/ml) for ovalbumin (c) and GST (d). The error bars indicate plus and minus one standard error from the mean with n = 5 for ovalbumin and n = 2 for GST.
concentrations and under identical sample conditions to WT PaDDAH, were not dependent on the loading concentration (Figures 6.2(a) to 6.2(d)).

6.2.3 – Conversion of elution volume into apparent molecular weight

Molecular weight gel filtration standards were applied to a pre-equilibrated S75 column and protein elution followed by absorbance at 280 nm. A graph of peak elution volume versus log of the species molecular weight was linear ($r = -0.99$) between molecular weights of 88 kD and 14.8 kD (Figure 6.3(a) and 6.3(b)). Using these data, an empirical relationship between elution volume ($EV; \text{in ml}$) and molecular weight ($MW; \text{in kD}$) was calculated (Equation 6.1).

\[
MW = 10 \left( \frac{(21.1 - EV)}{6.37} \right) \tag{Equation 6.1}
\]

Equation 6.1 was used to convert the elution volumes of WT PaDDAH, ovalbumin and GST into apparent molecular weight values (app. MW). For WT PaDDAH, a 100-fold decrease in loading concentration gave a decrease in average app. MW of $7.2 \pm 0.1 \text{kD}$ from $51 \pm 0.91 \text{kD}$ to $43.8 \pm 0.32 \text{kD}$ (Figure 6.3(c)). No such decrease was seen for either ovalbumin or GST (Figure 6.3(d)), which have respective average app. MWs of $47.2 \pm 0.2 \text{kD}$ and $48.4 \pm 0.2 \text{kD}$ over the concentration range studied.

6.2.4 – Estimation of the dissociation constant of WT PaDDAH using analytical SEC

During analytical SEC analysis WT PaDDAH eluted as a single peak at all concentrations. Analysis of the peak shape revealed a slight asymmetry around the peak maximum. This asymmetry was probably of experimental origin as the degree was not dependent on concentration (data not shown; Manning et al. 1996).

If the exchange rate between homodimeric and monomeric species is fast with respect to the rate of molecular separation then a single peak is expected in the elution profile (Manning et al. 1996). The app. MW, $\langle MW \rangle$, of this peak is a function of the mole fraction of homodimer and the mole fraction of monomer and the app. MW of the dimeric ($MW_D$) and monomeric ($MW_M$) species (Equation 6.2)

\[
\langle MW \rangle = p_M(MW_M) + p_D(MW_D) \tag{Equation 6.2}
\]

where $p_M$, the mole fraction of monomer, is defined as
Figure 6.3. (a) Elution profile of protein molecular weight standard samples with loading concentration of 4 mg/ml (see ‘Materials and Methods’ Section 3.2.10; (b) Plot of elution volume (EV) versus log molecular weight for elution volumes of molecular weight standards in (a) with calculated regression line; $r = -0.99$. Graph showing apparent molecular weight (MW) versus loading concentration for (c) WT PADDAH and (d) ovalbumin (black symbols) and GST (grey symbols).
By definition for a homodimer-monomer equilibrium (i.e. with no higher oligomer species)

\[ p_M + p_D = 1 \tag{Equation 6.5} \]

and the total concentration of protein, \( A_0 \), is given by

\[ A_0 = 2[D] + [M] \tag{Equation 6.6} \]

where \([M]\) is the concentration of free monomer and \([D]\) the concentration of homodimer. \( A_0 \) was defined as such because the concentration of protein in the solution was calculated using an extinction coefficient predicted from the amino acids sequence of a PaDDAH monomer.

Elimination of \( p_M \) from Equation 6.2 yields

\[ \langle \text{MW} \rangle = (1 - p_D)(\text{MW}_M) + p_D(\text{MW}_D) \tag{Equation 6.7} \]

Defining \( p_D \) in terms of \( A_0 \) and \([M]\) (Equation 6.6) and rearranging now allows \( \langle \text{EV} \rangle \) to be solved in terms of \([M]\), \( \text{MW}_M \) and \( \text{MW}_D \).

\[ \langle \text{MW} \rangle = \text{MW}_M + \left[ \frac{A_0 - [M]}{A_0} \right] (\text{MW}_D - \text{MW}_M) \tag{Equation 6.8} \]

The equilibrium dissociation constant \( K_d \), is defined

\[ D \rightleftharpoons M + M \]

\[ \therefore K_d = \frac{[M]^2}{[D]} \tag{Equation 6.9} \]

Combining Equations 6.6 and 6.9 and eliminating \([D]\) gives,

\[ 2[M]^2 + K_d[M] - A_0K_d = 0 \tag{Equation 6.10} \]

from which \([M]\) can be solved in terms of \( K_d \) and \( A_0 \)

\[ [M] = -\frac{K_d \pm (K_d^2 + 8A_0K_d)^{\frac{1}{2}}}{4} \tag{Equation 6.11} \]
[M] can be eliminated from Equation 6.8 with the correct solution of Equation 6.11. Now \( \langle \text{MW} \rangle \) is defined in terms of \( A_o \), which is known, and \( \text{MW}_M, \text{MW}_D, \) and \( K_d \)

\[
\langle \text{MW} \rangle = \text{MW}_M + \left[ \frac{4A_o - K_d - (K_d^2 + 8A_o K_d)^{1/2}}{4A_o} \right] (\text{MW}_D - \text{MW}_M)
\]

Equation 6.12

An apparent \( K_d \) can be estimated by fitting the experimental data to Equation 6.12. If the app. MW of the monomeric or homodimeric species are also unknown these can also be estimated by fitting.

In order to estimate \( K_d \) from WT PaDDAH analytical SEC data it was necessary to correct \( A_o \) to account for the dilution of the sample that occurred during each experiment. The area of the elution peak of a protein should be directly proportional to the loading concentration such that graph of \( \log(\text{peak area}) \) versus loading \( \log(\text{sample concentration}) \) should yield a positive linear relationship with a slope of unity (Manning et al. 1996). For WT PaDDAH, such a graph, gave a positive linear relationship \( r = 0.998 \) with a slope of 1.0 when the relationship was constrained to go through the origin. (Figure 6.4(a)). A dilution factor of \( 6.4 \pm 0.16 \) was calculated from the peak width at half-height (see ‘Materials and Methods’ Section 3.4.5; Manning et al. 1996).

The app. MW data (Section 6.2.3), corrected for the 6.4-fold dilution factor, was fit to Equation 6.12 (see ‘Materials and Methods’ Section 3.4.5). The best-fit values of \( \text{MW}_M, \text{MW}_D \) and \( K_d \) parameters are presented in Table 6.1 with their associated asymptotic standard errors. The 95% confidence intervals calculated from the standard errors of the best-fit values revealed a good level of confidence for \( \text{MW}_D \) (52.2 ± 1.02 kD) but unacceptable levels for \( \text{MW}_M \) (36.3 ± 10.28 kD) and \( K_d \) (235 ± 489 nM). Because of the high levels of uncertainty for \( \text{MW}_M \) and \( K_d \), attempts to evaluate the fit of Equation 6.12 to the experimental data using a Monte-Carlo procedure failed (see ‘Materials and Methods’ Section 3.4.5). Both 3-parameter \( (\text{MW}_M, \text{MW}_D \) and \( K_d \) \) and 2-parameter \( (\text{MW}_M \) and \( K_d, \) with \( \text{MW}_D \) fixed as 52.2) Monte-Carlo simulations were unsuccessful. Using the values of \( \text{MW}_M, \text{MW}_D \) and \( K_d \)
Figure 6.4. Fitting WT PaDDAH SEC data. (a) A plot showing log (peak area) versus loading log(WT PaDDAH concentration). The line of best fit is constrained to go through the origin, $r = 0.998$; (b) WT PaDDAH analytical SEC data best-fit to Equation 6.12 yielding fitting parameters of $K_D = 235$ nM, $MW_M = 36.3$ and $MW_D = 52.2$ (see Table 6.1); (c) the residuals of experimental data from simulated data (residual = $\hat{y}_i - y_i$).
in Table 6.1, the function used to optimise the fit of Equation 6.12 to the experimental data was evaluated (see Appendix A.1). Of the three parameters, \( MW_D \) was the best defined. When \( MW_M \) and \( K_d \) were floated and \( EV_D \) fixed, a large range of possible solutions of \( MW_M \) and \( K_d \) were observed (see Appendix A.2). It was not possible therefore to independently fix upon \( MW_M \) and \( K_d \) estimates.

<table>
<thead>
<tr>
<th>Table 6.1. Results of fitting WT PaDDAH analytical SEC data.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Parameter</strong> (^a)</td>
</tr>
<tr>
<td>( MW_M )</td>
</tr>
<tr>
<td>( MW_D )</td>
</tr>
<tr>
<td>( K_d )</td>
</tr>
</tbody>
</table>

\(^a\) \( MW_M \) and \( MW_D \) are the app MW of PaDDAH monomer and homodimer species, and \( K_d \) is the equilibrium dissociation constant. \(^b\) Asymptotic standard errors; \(^c\) 95 % confidence interval (C.I.) calculated using asymptotic standard errors.

To obtain a more satisfactory fit required the constraint of both \( MW_M \) and \( MW_D \). In the absence of experimentally derived values of \( MW_D \) and \( MW_M \), two alternatives were tried. One possible way was to define these parameters as the values returned from the original fit. With \( MW_M \) and \( MW_D \) defined as their respective values in Table 6.1, \( K_d \) was estimated to be \(218 \pm 7\) nM following 200 Monte-Carlo simulations (Figure 6.4(b)). Alternatively, \( MW_M \) and \( MW_D \) can be estimated using the predicted molecular weight of each PaDDAH species. The protomer molecular weight of affinity tagged PaDDAH from the amino acid sequence is 32.2 kD. With these definitions, a single parameter fit of Equation 6.12 was performed and \( K_d \) estimated to be \(1.3 \pm 0.03\) µM.

In summary, using a three parameter fit of Equation 6.12 against the experimental WT PaDDAH SEC data resulted in low levels of confidence for \( K_d \). This uncertainty related to the poor estimations of \( MW_M \) due to no experimental estimate of the elution volume of a PaDDAH monomer.
The results presented in Section 6.2 validate the previous observation of a concentration dependent elution volume for PaDDAH and suggestion of a dynamic homodimer-monomer equilibrium. In the following sections, the potential to manipulate this equilibrium is explored by using site-directed point mutations of residues located at the protomer interface to modulate the strength of the dimer interaction. Ultimately, through mutation, it may be possible to manipulate the equilibrium in such a way as to produce a monomeric variant of PaDDAH. Such as species would likely have solution properties more amenable to NMR studies.

6.3. Structural Description of Residues Contributing to the Homodimer Interface of WT PaDDAH.

6.3.1 – A description of the protomer interface of WT PaDDAH

The crystal structure of PaDDAH reveals a homodimer (Murray-Rust et al. 2001). Each protomer is built from five structurally homologous modules that adopt a five-fold pseudosymmetrical β/α-propeller fold. (see Figure 2.4; Paoli 2001a; Murray-Rust et al. 2001). The interaction surface between the two protomers is composed of three segments of each polypeptide chain (as defined in Jones and Thornton 1996) located at the edges of propeller blades 1 and 2. The major secondary structure contact points occur between helix-α1, strand-β1C (both in module 1) and helix-α2 (module 2) and their equivalent counterparts on the opposing protomer (Figures 6.5(a) and 6.6(a)).

The interaction between corresponding β1C strands (denoted β1C_A to indicate strand-β1C in subunit A and β1C_B, likewise in subunit B) forms a five residue antiparallel β-sheet at the interface that links the module 1 β-sheets (Figure 6.5; (a)). Strand-β1C comprises residues Asp-48, Ile-49, Thr-50, Leu-51 and Leu-52. The side chains of these residues fan out in an essentially perpendicular fashion to the direction of the β-strand producing a small hydrophobic patch at the centre of the homodimer interface (Figure 6.6(b)). Asp-48 appears to form an interaction with Arg-98 of helix-α2 from the opposite protomer, possibly producing an inter-protomer salt-bridge (Figure 6.5(b)).
Figure 6.5. The 3D structure of the protomer interface of WT PaDDAH and comparisons to AGAT and StrB1. (a) The backbone Ca trace of each subunit and the interface helices (α1 and α2) coloured red and interprotomer β-sheet coloured blue; helix-α1 is on the right of the β-sheet and helix-α1 on the left (b) the interaction of R98A (helix-α2A) and D48B (strand-β1C_B), and vice-versa; (c) the side chain interactions occurring between counterpart helices α1_A and α1_B. (d) 3D structure of AGAT (PDB: 1jdw) highlighting equivalent structural features as with WT PaDDAH in (a); (e) 3D structure of StrB1 (PDB: 1bwd) coloured as (a). Images created using Molscript (Kraulis 1991).
Figure 6.6. 3D structural analysis of the molecular surface of WT PaDDAH. (a) Worm diagram of the protein backbone of WT PaDDAH indicating the secondary structure elements involved in the homodimer interface. The arrows indicate the direction of the polypeptide chain. This orientation of WT PaDDAH is used in parts (b) to (d); (b) WT PaDDAH with a 3D molecular surface coloured according to residue hydrophobicity. The scale represent hydrophobicity value according to Eisenberg et al., 1984. The more positive (i.e. yellow) a value the more hydrophobic; (c) WT PaDDAH with a molecular surface coloured according to fractional ASA, (see Section 6.3.2); (d) WT PaDDAH with a molecular surface coloured according to absolute ASA, (see Section 6.3.2). The residues discussed in Figure 6.5 and the text are indicated in (d). All images generated using GRASP (Nicholls et al., 1991)
There are several side chain-side chain interactions between the α-helices situated at the homodimer interface (Figure 6.5(a) and 6.5(b)). Helices α1A (to indicate helix-α1 in subunit A) and α1B (similarly in subunit B) adopt an anti-parallel arrangement that is centred about the reciprocal interaction between Asn-36 residues in both A and B chains. Four residues of helix-α1 (Glu-33, Asn-36, Arg-40 and Gln-43), whose side chains are aligned along one face, interdigitate with the same set of residues from the opposing protomer. This antiparallel arrangement permits Arg-40 of subunit A to form a salt-bridge with Glu-33 from subunit B and vice versa.

The crystal structures of two arginine amidinotransferases (AGAT and StrB1) that have homologous 3D polypeptide folds to PaDDAH both suggest homodimer formation (Humm et al. 1997b). The nature of the interaction between the protomers in AGAT and StrB1 are strikingly similar to each other (Figure 6.5(d) and 6.5(e)). Interestingly, however, there are differences between the interface topologies in the amidinotransferases and that observed in PaDDAH. The homodimer interfaces of both AGAT and StrB1 are composed by residues from β-propeller modules 1, 2 and 3 and are centred about the 3-stranded β-sheet in module 2. In PaDDAH, however, the interface is limited only to the surface of modules 1 and 2. The extended β-sheet formed at the interface of PaDDAH is not formed in either AGAT or StrB1. In all three cases, the 3-stranded β-sheet of module 2 is involved in the interface but, in AGAT and StrB1 the opposing β-sheets are arranged at an angle to each other and not joined. In the arginine amidinotransferases, the contributions of modules 2 and 3 are homologous to those from modules 1 and 2 in PaDDAH. But in both AGAT and StrB1 there is an additional contribution from a third module – module 1 – between two opposing 2-stranded β-sheet loops.

The results of analytical SEC presented above for WT PaDDAH indicated the existence of a dynamic equilibrium between dimeric and monomeric species. It was therefore possible that minor alterations of the composition of the interface, such as those resulting from site-directed amino acid substitution, would have an appreciable effect on the position of the equilibrium. A thorough investigation into the structure of the protomer interface of PaDDAH was undertaken to identify the residues that are key in the interaction.
6.3.2 – Analysis of the interface of WT PaDDAH using accessible surface areas

Analysis of the homodimer interface of WT PaDDAH using the Protein-Protein Interaction (PPI) server (Jones and Thornton 1996) revealed that the area occluded from a solvent molecule with a radius of 1.4 Å on dimer formation was approximately 860 Å² per protomer subunit (Table 6.2).

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Protein</th>
</tr>
</thead>
<tbody>
<tr>
<td>PDB Code</td>
<td>WT PaDDAH</td>
</tr>
<tr>
<td>Number of Residues</td>
<td>254</td>
</tr>
<tr>
<td>Protomer Mass (kD)</td>
<td>28.5</td>
</tr>
<tr>
<td>ΔASA (Å²)</td>
<td>857.4</td>
</tr>
<tr>
<td>ΔASA as % of total ASA</td>
<td>7.8</td>
</tr>
<tr>
<td>Residues with ΔASA &gt; 0.0 Å²</td>
<td>23</td>
</tr>
<tr>
<td>Gap Volume (Å³)</td>
<td>5709.6</td>
</tr>
<tr>
<td>Gap Index (Å)</td>
<td>3.33</td>
</tr>
</tbody>
</table>

Table 6.2. Analysis of the homodimer interfaces from the crystal structures of WT PaDDAH, AGAT, StrBl and GST using the protein-protein interaction server.

Investigations into homodimeric protein-protein interfaces by Thornton and co-workers (Jones and Thornton 1996) yielded a positive correlation between mass of the monomer (in kD) and interface area (in Å²) for sets of homodimers selected from the PDB. In their most recent study, a set of 72 homodimers was generated which displayed a weak correlation coefficient of 0.51 between protomer mass (Daltons) and interface surface area (Å²; T. Kabir personal communication). The average solvent accessible surface area (ASA) buried per subunit in these homodimers was 1860 Å² with a standard deviation of 1170 Å² and minimum and maximum values of 540 Å².
and 7150 Å² respectively. From these data, a homodimer composed of 29 kD monomers was predicted to have an interface surface area of ca. 1650 Å² per subunit. GST is a homodimeric enzyme of comparable molecular mass to homodimeric PaDDAH (Table 6.2). Analysis of the area concealed on formation of a GST homodimer showed a larger interface than for PaDDAH: only 7.8 % of the surface of a PaDDAH subunit is buried on homodimer formation compared to 13.5 % with GST. The buried ASA for GST was 1505.2 Å² compared to a predicted value of 1400 Å². The interface of the two structurally homologous amidinotransferases, AGAT and StrB1, both involve more segments of the polypeptide chain than PaDDAH (Section 6.3.1). Consequently, the percentage of monomer ASA buried in AGAT or StrB1 homodimers was larger than that of PaDDAH: 10.9 % and 11.0 %, for AGAT and StrB1 respectively, in comparison to 7.8 % for PaDDAH (Table 6.2). Clearly, for the mass of the protomer subunits, the dimer interface of WT PaDDAH was small when compared with the data set of Kabir (T. Kabir personal communication).

Jones and Thornton also used a ‘Gap Index’ value to describe the complementarity between the interacting surfaces of multimeric proteins (Equation 6.13).

\[
\text{Gap Index (Å)} = \frac{\text{Gap Volume (Å}^3\text{)}}{\text{Total } \Delta \text{ASA (Å}^2\text{)}}
\]

Equation 6.13

where the ‘Gap Volume’ is the ‘empty’ volume between the two protomers calculated by the program SURFNET (Laskowski 1995) and ΔASA is the total interface surface area buried on oligomerisation. An average Gap Index value of 2.2 Å (standard deviation = 0.87 Å and minimum and maximum values of 0.57 Å and 4.43 Å respectively) was calculated from a set of 32 homodimers (Jones and Thornton 1996). The calculated Gap Index for WT PaDDAH was 3.3 Å (Table 6.2), which suggests the interface displays lower complementarity than the average. The Gap Index for GST, 2.91, was also greater than the average value (Table 6.2), again suggesting a weaker interaction. AGAT and StrB1, however, have Gap Indices below the average, 1.63 and 1.76 respectively, indicating a high level of interface complementarity.

The contribution of individual residues to the PaDDAH homodimer interface was assessed in two ways. The buried surface area of each residue was determined using the program NACCESS (Hubbard 1992). A value for ASA per residue \(i\) for both the
monomeric (ASA$_{iM}$) and homodimeric (ASA$_{iD}$) species was derived using NACCESS from which the change in ASA per residue (ΔASA$_{i}$) on going from monomer to homodimer can be calculated (Table 6.3). For each residue, ΔASA$_{i}$ can be expressed as either an absolute value (in Å$^2$) or as a fraction of ASA$_{iM}$ (fASA). Both of these values highlight different aspects of each residues involvement at the interface. Colouring residues in a 3D protein structure by fASA will indicate the position of the intermolecular interaction on the surface of the protomer (e.g. Figure 6.6(c)). Figure 6.6(c) shows the surface of a PaDDAH protomer with individual residues coloured according to fASA (Table 6.3). From this representation, one can clearly identify those residues which are centremost in the interface (coloured orange and red), those which reside at the periphery (yellow) and those not involved (white). The residues that form the central β-strand all have values of fASA greater than 0.75 indicating that more than 75% of each of these residues ASA is concealed upon homodimerisation. Figure 6.6(d) shows the same face of WT PaDDAH but coloured according to absolute ΔASA$_{i}$ in Å$^2$. This representation identifies which residues contribute most surface area in absolute terms to the interface (coloured red). Residues which have no contribution are coloured white. The residues with highest absolute ΔASA$_{i}$ are Arg-40 (102.5 Å$^2$) and Gln-43 (89.7 Å$^2$) from helix-α1, leucine-51 (84.9 Å$^2$), which is the centre residue of strand-β3, and Arg-98 (108.6 Å$^2$) from helix-α2.

6.3.3 – Selection of Residues for Mutation

WT PaDDAH has been shown to exist in a homodimer-monomer equilibrium (Section 6.2.1) with an estimated K$_d$ value of the order of 10$^{-7}$ M (Section 6.2.4). Analysis of the structure of WT PaDDAH has identified several residues that appear to play important roles in the homodimer interface. Residues from either helices α1 (33, 36, 40 or 43) or α2 (94 or 98), or from the interface β-strand, β1C (48 to 52), are all identified above by surface area contributions (see Section 6.3). Residues Glu-33, Asn-36, Arg-40, Gln-43 and Arg-98 were selected for site-directed mutagenesis. In the crystal structure of WT PaDDAH, the side chains of these residues interact predominantly with other residues at the interface and particularly residues from the opposite subunit. The structural disruption caused by substitutions at these positions would likely to be localised to the interface of WT PaDDAH, with only limited perturbation of the global fold.
<table>
<thead>
<tr>
<th>Residue</th>
<th>Number</th>
<th>Segment</th>
<th>$ASA_{i,M}$</th>
<th>$ASA_{i,D}$</th>
<th>$\Delta ASA_i$</th>
<th>$f_{ASA_i}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>LYS</td>
<td>3</td>
<td>1</td>
<td>147.06</td>
<td>139.16</td>
<td>7.90</td>
<td>0.05</td>
</tr>
<tr>
<td>HIS</td>
<td>4</td>
<td>1</td>
<td>45.51</td>
<td>18.87</td>
<td>26.64</td>
<td>0.59</td>
</tr>
<tr>
<td>THR</td>
<td>9</td>
<td>1</td>
<td>32.29</td>
<td>7.41</td>
<td>24.88</td>
<td>0.77</td>
</tr>
<tr>
<td>LEU</td>
<td>32</td>
<td>2</td>
<td>71.89</td>
<td>56.49</td>
<td>15.40</td>
<td>0.21</td>
</tr>
<tr>
<td>GLU</td>
<td>33</td>
<td>2</td>
<td>117.93</td>
<td>95.92</td>
<td>22.01</td>
<td>0.19</td>
</tr>
<tr>
<td>HIS</td>
<td>35</td>
<td>2</td>
<td>6.92</td>
<td>2.54</td>
<td>4.38</td>
<td>0.63</td>
</tr>
<tr>
<td>ASN</td>
<td>36</td>
<td>2</td>
<td>77.65</td>
<td>16.27</td>
<td>61.38</td>
<td>0.79</td>
</tr>
<tr>
<td>ILE</td>
<td>39</td>
<td>2</td>
<td>15.29</td>
<td>5.69</td>
<td>9.60</td>
<td>0.63</td>
</tr>
<tr>
<td>ARG</td>
<td>40</td>
<td>2</td>
<td>158.61</td>
<td>56.10</td>
<td>102.51</td>
<td>0.65</td>
</tr>
<tr>
<td>GLN</td>
<td>43</td>
<td>2</td>
<td>112.62</td>
<td>22.92</td>
<td>89.70</td>
<td>0.80</td>
</tr>
<tr>
<td>ASP</td>
<td>46</td>
<td>2</td>
<td>106.30</td>
<td>86.81</td>
<td>19.49</td>
<td>0.18</td>
</tr>
<tr>
<td>VAL</td>
<td>47</td>
<td>2</td>
<td>19.08</td>
<td>2.92</td>
<td>16.16</td>
<td>0.85</td>
</tr>
<tr>
<td>ASP</td>
<td>48</td>
<td>2</td>
<td>81.54</td>
<td>19.49</td>
<td>62.05</td>
<td>0.76</td>
</tr>
<tr>
<td>ILE</td>
<td>49</td>
<td>2</td>
<td>46.66</td>
<td>0.97</td>
<td>45.69</td>
<td>0.98</td>
</tr>
<tr>
<td>THR</td>
<td>50</td>
<td>2</td>
<td>30.37</td>
<td>2.15</td>
<td>28.22</td>
<td>0.93</td>
</tr>
<tr>
<td>LEU</td>
<td>51</td>
<td>2</td>
<td>86.62</td>
<td>1.69</td>
<td>84.93</td>
<td>0.98</td>
</tr>
<tr>
<td>LEU</td>
<td>52</td>
<td>2</td>
<td>22.77</td>
<td>5.78</td>
<td>16.99</td>
<td>0.75</td>
</tr>
<tr>
<td>PRO</td>
<td>53</td>
<td>2</td>
<td>89.25</td>
<td>37.26</td>
<td>51.99</td>
<td>0.58</td>
</tr>
<tr>
<td>PRO</td>
<td>54</td>
<td>2</td>
<td>64.78</td>
<td>49.37</td>
<td>15.41</td>
<td>0.24</td>
</tr>
<tr>
<td>GLU</td>
<td>94</td>
<td>3</td>
<td>133.50</td>
<td>119.34</td>
<td>14.16</td>
<td>0.11</td>
</tr>
<tr>
<td>THR</td>
<td>95</td>
<td>3</td>
<td>5.86</td>
<td>5.12</td>
<td>0.74</td>
<td>0.13</td>
</tr>
<tr>
<td>ARG</td>
<td>98</td>
<td>3</td>
<td>148.36</td>
<td>39.79</td>
<td>108.57</td>
<td>0.73</td>
</tr>
<tr>
<td>PHE</td>
<td>99</td>
<td>3</td>
<td>56.23</td>
<td>27.65</td>
<td>28.58</td>
<td>0.51</td>
</tr>
</tbody>
</table>

* from the Protein-Protein Interaction server (see Table 6.2); b defined in Section 6.3.2; c ASA of residue i in monomer; d ASA of residue i in dimer; e difference between $ASA_{i,M}$ and $ASA_{i,D}$; f fractional ASA calculated where $f_{ASA_i} = (ASA_{i,M} - ASA_{i,D})/ASA_{i,M}$. Values of $ASA_{i,M}$, $ASA_{i,D}$ and $\Delta ASA_i$ are in Å².

Single amino acid substitutions were created at the positions listed in Table 6.4 (see 'Materials and Methods' Section 3.2.7). As PaDDAH is a symmetric homodimer, the
result of mutating a single residue at the DNA level would effectively produce a double substitution in the final molecule.

### Table 6.4. Summary of PaDDAH mutations produced.

<table>
<thead>
<tr>
<th>Position</th>
<th>WT PaDDAH Residue</th>
<th>Codon</th>
<th>Mutant Residue</th>
<th>Codon</th>
</tr>
</thead>
<tbody>
<tr>
<td>33</td>
<td>Glu</td>
<td>GAG</td>
<td>His</td>
<td>CAC</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Gln</td>
<td>CAG</td>
</tr>
<tr>
<td>36</td>
<td>Asn</td>
<td>AAC</td>
<td>Asp</td>
<td>AAG</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>His</td>
<td>CAC</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Trp</td>
<td>TTG</td>
</tr>
<tr>
<td>40</td>
<td>Arg</td>
<td>CGC</td>
<td>Glu</td>
<td>GAG</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Trp</td>
<td>TTG</td>
</tr>
<tr>
<td>43</td>
<td>Gln</td>
<td>CAG</td>
<td>His</td>
<td>CAC</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Arg</td>
<td>CGG</td>
</tr>
<tr>
<td>98</td>
<td>Arg</td>
<td>CGC</td>
<td>His</td>
<td>CAC</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Asn</td>
<td>AAC</td>
</tr>
</tbody>
</table>

See also Table 3.1 for details of the primers used in the production of these mutations. The active site mutant C249S PaDDAH was donated by Dr. M. MacAllister.

### 6.4. Solubility and Activity Assays of WT and Mutant PaDDAH.

#### 6.4.1 – Activity Assays of WT PaDDAH

For each protein PaDDAH sample (i.e. WT, mutant or negative control), a 50 ml LB culture of *E. coli* BL21(DE3) was grown (see ‘Materials and Methods’ Section 3.2.8). After a basic affinity purification procedure (see ‘Materials and Methods’ Section 3.2.8), protein expression was analysed using SDS-PAGE (Figure 6.9(a)). The SDS-PAGE based assay employed here only allowed a qualitative assessment of the expression and solubility level of each PaDDAH sample. However, it was adequate for the purpose of allowing the activity of mutant forms of PaDDAH to be compared to the activity of the wild-type enzyme. The use of an affinity purification step served to remove the majority of the background bacterial proteins that might interfere in the assay.

DDAH enzymes catalyse the breakdown of asymmetric N-methylarginines into L-citrulline and methylamine. The activity of WT PaDDAH was established using a
colorimetric assay that detected L-citrulline (see ‘Materials and Methods’ Section 3.2.9). The assay employed here was adapted from that presented by Ogawa et al. (1989) and Knipp and Vašák (2000).

A standard curve was produced for L-citrulline between 0 μM and 400 μM (Figure 6.7(a)) which showed a positive linear correlation between absorbance at 540 nm ($A_{540}$) and L-citrulline concentration over the range studied ($r = 0.999$). From the graph, a relationship between L-citrulline concentration and $A_{540}$ was derived (Equation 6.14).

$$[\text{L-citrulline}] \, (\mu\text{M}) = \frac{(A_{540} - 0.0372)}{0.0058}$$

Equation 6.14

There appeared to be an upper limit to the linear response assay for L-citrulline concentrations above 400 μM (Figure 6.7(b)). Between 0 μM and 400 μM L-citrulline, however, the results obtained were highly reproducible and compared extremely well with those previously reported (Knipp and Vašák 2000).

The activity of WT PaDDAH was measured over a time course (Figure 6.8). $A_{540}$ increased linearly with time for 10 minutes indicating that, over that period, the presence of substrate had not become a limiting factor. In addition, the negative control (expression product from unmodified pPROEX Hta vector; see ‘Materials and Methods’ Section 3.2.8) showed no increase in $A_{540}$ indicating no basal level of L-citrulline production. Values of $A_{540}$ were converted to L-citrulline concentration in the reaction mixture using Equation 6.14 and adjusted for the 1 in 10 dilution of the reaction mixture (see ‘Materials and Methods’ Section 3.2.9). Assuming a linear relationship between L-citrulline concentration and time (for $t < 10$ mins $r = 0.99$), the slope of the regression line gives an approximate rate of L-citrulline production of 3 μM s$^{-1}$. As the protein concentration of this sample was not accurately determined it was not possible to convert this value into enzyme specific activity units. However, this method allows a suitable means for the comparison between the activity of WT PaDDAH and different mutant forms of the enzyme.
Figure 6.7. L-citrulline calibration plots. (a) absorbance at 540 nm for L-citrulline samples of 0 μM to 400 μM. The regression line (—) has $r = 0.999$. INSET: expansion between 0 μM and 30 μM; (b) absorbance at 540 nm for L-citrulline samples of 0 mM to 2.5 mM. INSET: a superposition between data the data in (a), •, with those in (b), *, for data points obtained between 0 μM and 450 μM. In both parts the error bars indicate plus and minus one standard deviation from the mean with $n = 5$. 
Figure 6.8. A time course of WT PaDDAH activity compared to a negative control. Activity was measured using a colorimetric assay which detects L-citrulline (see ‘Materials and Methods’ Section 3.2.9). The observed absorbance ($A_{540}$) was converted into L-citrulline concentration (mM) in the reaction mixture using Equation 6.14 (see Section 6.4.1). The error bars indicate plus and minus one standard deviation from the mean with $n = 3$. The nature of the negative control is described in Section 6.4.1.
6.4.2 - Activity Assays of PaDDAH Mutants

The results of the time course experiment (see Section 6.4.1) showed that the production of L-citrulline in the presence of WT PaDDAH is linear up to a total reaction time of twenty minutes. A single time point assay that measured L-citrulline concentration after ten minutes was used to compare the activity of WT and mutant PaDDAH samples. The raw $A_{540}$ data was converted into L-citrulline concentration ($\mu$M) using Equation 6.14 and then normalised to the relative amount of protein present in each sample as estimated from SDS-PAGE analysis (see 'Materials and Methods' Sections 3.2.8 and 3.2.9) using Equation 6.15.

$$\text{Normalised [L-citrulline]} = \frac{[\text{L-citrulline}]}{\text{WT}}$$  \text{Equation 6.15}

where $X_i$ is the estimated amount of protein present in a gel lane and WT is the amount of WT PaDDAH present. A comparison of WT and mutant PaDDAH expression and normalised activities is presented in Figure 6.9.

6.5. Analysis of the Self Association of PaDDAH Mutants.

Analytical gel filtration was performed to ascertain the effect of the amino acid substitutions on the self-association phenomenon observed for WT PaDDAH (see Section 6.2.1). All PaDDAH mutants were expressed and purified on a large scale as described for WT (see Chapter 5, Section 5.1). Only R40W and R98N could not be produced in sufficient quantity to permit characterisation. As with the analytical SEC experiments performed on WT PaDDAH, 100 µl samples of mutant PaDDAH at protein concentrations of 5 mg/ml to 50 µg/ml were applied to an analytical S75 column and their elution profiles followed by absorbance at 280 nm. A direct comparison of the elution chromatogram of WT PaDDAH and those of nine soluble interface mutants, all at 1 mg/ml, is shown in Figure 6.10. As can be seen (Figure 6.10), the effects of the residue substitutions on the elution profile of PaDDAH vary considerably. At 1 mg/ml WT PaDDAH eluted as a single peak at 10.27 ml. Two mutants, N36D and R40E, yielded double peaked elution profiles. The profiles of the Q43R and R98H mutants both showed a major peak with an elution volume of approximately 11.8 ml, a volume considerably larger than WT PaDDAH at 1 mg/ml. Comparison of the elution volumes of Q43R and R98H PaDDAH with the elution
Figure 6.9. Solubility and activity of mutant PaDDAH compared to WT. (a) SDS-PAGE showing WT and mutant PaDDAH samples following batch purification from small scale expression cultures; (b) Assay of WT and mutant PaDDAH activity. The observed data ($A_{540}$) was converted into L-citrulline concentration (mM) in the reaction mixture using Equation 6.14 and normalised for approximate protein concentration using Equation 6.15. The error bars indicate plus and minus one standard deviation from the mean with $n = 3$. 

190
Figure 6.10. Elution profiles of WT PaDDAH (top) compared with the mutations (from second top downwards, as labelled) E33H, E33Q, N36D, N36H, N36W, R40E, Q43H, Q43R and R98H. All samples were run with a loading concentration of 1 mg/ml (30 μM). The elution volume (ml) of each major peak is provided. The bottom profile shows the elution profile of molecular weight standards at 4 mg/ml (See ‘Materials and Methods’ Section 3.2.10). The molecular weight of each standard protein is given in kD. The y-axis of all the profiles represents absorbance at 280 nm in arbitrary units (A.U.) and is not to scale.
volumes of molecular weight standards (Figure 6.10, bottom profile) indicated that both mutant peaks corresponded to molecules of approximately 30 kD in mass. The remaining elution profiles (for mutants E33H, E33Q, N36H, N36W and Q43H) still showed a single peak but the elution volume (10.43 ml, 10.55 ml, 10.50 ml, 10.18 ml and 10.88 ml respectively) differed in comparison to WT PaDDAH. Again, samples were run in the presence of βME to preclude any effects of free cysteine thiol groups. The elution profiles of these samples did not differ from those obtained in the absence of βME (data not shown).

As the concentration of WT PaDDAH samples applied to the column was decreased 100-fold from 5 mg/ml, the elution volume increased from 10.23 ml to 10.64 ml. This change converted into a decrease in apparent molecular weight of 7.2 kD from 51.0 kD to 43.8 kD (Section 6.2.1). Further effects of the residue substitutions on the self-associative properties of PaDDAH were observed when comparing the change in apparent molecular weight as the concentration of the samples loaded was decreased (Figure 6.11(a)). The E33H, E33Q and Q43H mutants eluted as single peaks at 5 mg/ml with elution volumes that gave app. MWs in the range 46.8 kD to 48.6 kD. These values were comparable to WT PaDDAH, which gave an app. MW of 51.0 kD at 5 mg/ml, and suggested that the E33H, E33Q and Q43H mutants were predominantly dimeric at higher sample concentrations. However over the concentration range studied, all three mutants showed a considerably larger decrease in apparent molecular weight (11.1 kD, 15.2 kD and 17.0 kD respectively) compared to WT PaDDAH (Figure 6.11(a)).

Both Q43R and R98H mutants had decreased elution volumes in comparison to WT PaDDAH (Figure 6.10). At the highest concentration applied, the elution volumes of Q43R and R98H were 11.46 ml and 11.64 ml respectively, decreasing to 12.03 ml and 11.96 ml at 50 μg/ml. As can be seen in Figure 6.11(a), these elution volumes values tend towards an app. MW in the region of 30 kD and approximated to the value predicted for a monomer of WT PaDDAH.

Ovalbumin and GST (Section 6.2.2), and more interestingly the PaDDAH mutant N36W showed no change in app. MW when the sample concentration was varied in
Figure 6.11. (a) Elution volume of WT and mutant PaDDAH at varying loading concentrations. For the elution profiles that show two peaks in Figure 6.10 (e.g. N36D and R40E) the peak with the smallest elution volumes has been labelled ‘1’ and the largest with ‘2’; (b) Analytical SEC elution profiles of N36D with loading concentrations of 10, 5, 1, 0.5, 0.25 and 0.1 mg/ml; (c) expansion of (b) with loading concentrations (mg/ml) annotated. 1 mg/ml and 0.25 mg/ml samples were repeated.
this experiment (Figure 6.11(a)). The mutation N36W in PaDDAH appears to have removed the concentration dependence of the measured elution volume. The mutants N36D and R40E both yielded an elution profile with two peaks when applied at a concentration of 1 mg/ml (Figure 6.10). Over the concentration range studied, the peak elution volumes in the profile of R40E PaDDAH did not change as a function of protein concentration (Figure 6.11(a)). However, large changes were observed with the N36D mutant (Figure 6.11(b)). Initially, at 10 mg/ml a single peak was observed, but as the loading concentration was decreased, two separate elution peaks were resolved. The elution volume of the larger molecular weight species (10.17 ml at 2.5 mg/ml) remained constant, at approximately 10.2 ml, at all sample concentrations. The species that eluted at 10.49 ml at 2.5 mg/ml showed an increase in elution volume as the loading concentration was decreased (Figures 6.11(b) and 6.11(c)). In terms of app. MW the peak with the smallest elution volume (denoted N36D1 in Figure 6.11(a)) gave values in the region expected for a dimer of PaDDAH. Whereas, the second, more slowly eluting peak (denoted N36D2 in Figure 6.11(a)) showed a decrease in app. MW tending towards a monomeric species.

6.6. Estimation of Equilibrium Dissociation Constant for PaDDAH.

In section 6.2.4, an attempt was made to fit the analytical SEC profile of WT PaDDAH to Equation 6.12. However, the absence of an experimental elution volume (and therefore app. MW) for a monomeric species of PaDDAH prevented a reliable estimate of $K_d$ being made. The elution profiles of several PaDDAH mutants suggested the formation of a either monomeric or dimeric species (Figures 6.10 and 6.11). In particular, N36W displayed a concentration-independent elution volume averaging 10.18 ml, which was suggestive of a homodimer, and Q43R and R98H had elution volumes converging on 12.01 ml and 11.94 ml at the lowest loading concentrations. The mutants R40E and N36D gave double peaked elution profiles. Both molecular species in the profiles of R40E PaDDAH had relatively invariant elution volumes with average values of 10.19 ± 0.03 ml and 11.48 ± 0.03 ml for the slow and fast eluting species respectively. The profile of N36D PaDDAH showed a single peaked elution profile at higher concentrations that bifurcated as the sample concentration was lowered (Figures 6.12(b) and 6.12(c)). The respective elution
volumes of the slow and fast migrating species were 10.19 ml and 11.56 ml at the lowest concentration applied. In summary, it appeared from studies of mutant PaDDAH samples that a minimum elution volume, i.e. that resulting from a PaDDAH homodimer, was approximately 10.2 ml. But, two possibilities existed for the elution volume of a potential monomer, approximately 11.5 (N36D and R40E) or approximately 12 ml (from Q43R and R98H). In terms of app. MW these values correspond to 51.9 kD for a PaDDAH homodimer and 32.1 kD or 26.8 kD for a PaDDAH monomer.

The best-fit value of $MW_D$ estimated from fitting Equation 6.12 to the experimental data was 52.2 kD (Table 6.1). This value appeared to be fit with high precision and is well reflected in the app. MW of both WT PaDDAH at higher concentration and a subset of the PaDDAH point mutations. N36W and the faster eluting species in the elution profiles of N36D and R40E all had elution volumes that gave corresponding app. MW of approximately 52 kD. However, in the original fitting substantial imprecision remained in the estimated app. MW of monomeric PaDDAH, $MW_M$. Therefore, $K_d$ was estimated again by re-fitting Equation 6.12 to the analytical SEC data using as constraints $MW_D$ defined as 51.9 kD and $MW_M$ fixed as either 32.1 kD or 26.8 kD (Table 6.5). Standard deviations for each $K_d$ were determined using a Monte-Carlo evaluation of each fit (see 'Materials and Methods', Section 3.4.5).

<table>
<thead>
<tr>
<th>$MW_D$ (ml)</th>
<th>$MW_M$ (ml)</th>
<th>$K_d$ (nM)</th>
<th>Std. Dev. (nM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>52</td>
<td>32.1</td>
<td>115</td>
<td>3.2</td>
</tr>
<tr>
<td>52</td>
<td>26.8</td>
<td>64</td>
<td>1.9</td>
</tr>
</tbody>
</table>

$^a$ constrained to limiting values obtained in experimental SEC profiles of WT and mutant PaDDAH – as described in text; $^b$ mean and $^c$ standard deviation of 200 Monte Carlo simulations (see 'Materials and Methods' Section 3.4.5).

### 6.7. Discussion.

The data presented here clearly show that, in solution, WT PaDDAH is in dynamic equilibrium between homodimer and monomer species. This property is evidenced by the observation of concentration dependent elution volumes in analytical SEC.
experiments, under conditions where control monomer (ovalbumin) and obligate homodimer (GST) proteins exhibit invariant elution characteristics. That PaDDAH exhibits this property is not entirely unsurprising given the relatively small intermolecular surface contact (ca. 860 Å²) revealed in the 3D structure determined by X-ray crystallography (Murray-Rust et al. 2001). Moreover the qualitative and quantitative characteristics of this varying hydrodynamic behaviour were found to be modulated by site directed mutagenesis of residues that are within the inter-protomer interface.

The variation of elution volume with protein concentration in the SEC experiments permits attempts to obtain an estimate of the PaDDAH homodimer dissociation constant. This was first attempted by fitting of Equation 6.12 which casts the observed apparent molecular weights in terms of the limiting values corresponding to intact monomer and homodimer species (MWₘ and MWₐ respectively) and the dissociation constant Kₐ. Attempts to fit all three parameters simultaneously yielded a well-defined estimate of MWₐ but rather imprecise estimates of MWₘ and Kₐ (Table 6.1). Whilst reasonably sensible values of MWₘ and Kₐ were returned by the fitting procedure, the estimated uncertainty in these parameters was unacceptably high. It can be shown (see Appendix A.2) that equally good fits to the experimental SEC data can be obtained for many different combinations of MWₘ and Kₐ. A limitation on the estimation of the magnitude of Kₐ is therefore the extrapolation of the analytical SEC elution volume of WT PaDDAH to infinite dilution (which would by definition yield an estimate of MWₘ). Intrinsically it is difficult to obtain a reliable estimate of MWₘ because by the nature of the experiment the sensitivity and reliability of the detection of an SEC elution peak is degraded by the low concentration of the eluting protein.

In other cases where analytical SEC measurements have been used to derive oligomer dissociation constants advantage has been taken of prior knowledge of mutants that have altered hydrodynamic properties. In two recently published examples, analytical SEC was used to calculate Kₐ for reversibly associating systems (Manning et al. 1996, Dorfman and Walsh 2001). Both studies used the method described by Manning and co-workers (Manning et al. 1996, Manning et al. 1999). In their original publication, Manning and co-workers studied equilibrium dissociation of haemoglobin tetramers into homodimers (Manning et al. 1996). In this case the limiting elution volumes of
the two species were well characterised. The homotetramer elution volume was established from a sample of chemically cross-linked homotetramers and the homodimer elution volume from a naturally occurring homodimeric mutant. In the other example, the elution profile of the apple 4 domain of Factor XI at certain sample concentrations gave two peaks corresponding to the elution of the homodimer and of the monomer (Dorfman and Walsh 2001). In both of these systems, the elution volume of each oligomer was known and allowed the experimental data to be fit using a single-parameter optimisation. To give a more accurate estimation of the $K_d$ for the homodimer-monomer equilibrium of WT PaDDAH reported here, the elution volume of a PaDDAH monomer will have to be known.

A set of point mutated PaDDAH variants has been characterised with a view to modulation of the dimerisation property that makes NMR analysis of the protein particularly challenging. The analytical SEC data for these proteins usefully provide the potential to validate the value of $MW_d$ and estimate the limiting value of $MW_m$.

For example, taken as a whole the elution volumes of all forms of PaDDAH applied to SEC varied of a range of approximately 1.5 ml, which is significantly greater than that observed for the WT protein (Figure 6.11(a)). In addition, no variant of PaDDAH eluted significantly faster than the WT protein at its highest concentration (Figure 6.11(a)). As concluded above, the data suggest that a subset of the elution profiles is consistent with invariant dimer forms of PaDDAH (e.g. N36W), whereas others appear to tend towards a limiting value of the elution time at low concentration between 11.5 ml (e.g. R40E1) and 12.0 ml (e.g. R98H). These elution volumes would appear to be consistent with monomer forms of PaDDAH ($MW \sim 30$ kD) based upon the prior calibration of the column with molecular weight standards. Taking example from the work of Manning and co-workers (Manning et al. 1996), each of these elution volume values has been used to estimate of $MW_m$ using Equation 6.1, together with the reasonable constraint that $MW_d = 51.9$ kD, to refit the data for the concentration dependent elution profile for WT PaDDAH. Under these conditions the problem is reduced to an optimisation of Equation 6.12 with constrained $MW_d$ and $MW_m$ values to yield a fitted value of $K_d$. The values of $K_d$ returned in this procedure are detailed in Table 6.5. It would therefore appear that the homodimer dissociation constant of WT PaDDAH lies within the interval of 50 to 250 nM.
The crystal structure of WT PaDDAH revealed that the interaction surface between
the subunits was relatively small: 860 Å² per protomer, which represents only 7.8 %
of the total surface area of a monomer. The points of interaction between the two
protomers appears to be limited to the central antiparallel β-sheet, formed between
opposing β1C strands, and to a small number of peripheral side chains located in
helices α1 and α2. Only 23 residues in the protein had decreased ASA on the
formation of the homodimer (Table 6.3). In this study, five of those residues (Glu-33,
Ans-36, Arg-40, Gln-43 and Arg-98) were targeted for mutagenesis. In the analysis of
the mutant proteins, there were only two cases (R40W and R98N) when
characterisation by analytical SEC was not possible, in both cases due to insufficient
yield of soluble protein from a large-scale expressions. Of the soluble mutants, all
showed some degree of deviation from the hydrodynamic behaviour of the wild-type
enzyme. In addition to the mutations reported here, several were also produced with
the intention of disrupting the strand-β1C interaction. None of these mutants were
expressed in the soluble fraction of the bacterial expression host (data not shown).

In analytical SEC experiments, WT PaDDAH eluted as a single peak at 10.27 ml at a
loading concentration of 1 mg/ml (Figure 6.10). Two of the PaDDAH mutants had
longer elution times at this sample concentration: Q43R, 11.72 ml, and R98H, 11.86
ml. In both cases, the oligomer equilibrium appeared to have shifted towards a smaller
molecular weight species (Figure 6.11(a)). The respective elution volumes of these
samples correspond to molecules with a molecular weight comparable to that
predicted for monomeric PaDDAH. From study of the 3D structure, it may be
possible to rationalise the observed changes in elution volume, which themselves
support a change in oligomeric state, with the effect each mutation would have the
structure of the homodimer interface. Specifically, the mutations Q43R, R98H and
N36W will be focused upon.

In the 3D structure of PaDDAH, Gln-43 resides at the periphery of the helix-α1
interactions. The change in ASA for Gln-43, in terms of both $f_{ASA}$ and absolute
difference, is high (Table 6.3). Clearly this residue has important cross-interface
contacts. Furthermore, the side chain of Gln-43 packs closely alongside the aliphatic
side chain methylenes of Arg-40 (Figure 6.5(b)). The mutation Q43R would introduce
an arginine at this position, a residue with a longer, bulkier side chain and a positively
charged terminal guanidino group. The high surface area contribution of Gln-43 at the interface suggests a tight interaction at this position and that the larger arginine side chain would not be easily accommodated. Furthermore, an important cross-interface interaction is the formation of a hydrogen bond between Arg-40 and Glu-33. It is therefore possible that the presence of arginine at position 43 might disrupt this important stabilising interaction. In combination, these effects would likely reduce stability of the interactions between helices $\alpha_1A$ and $\alpha_1B$ and of the homodimer interface as a whole. Another mutation of Gln-43, in this case to histidine, also had a marked effect on the solution behaviour of PaDDAH (Figure 6.11(a)). Again, histidine has a larger side chain than glutamine and would therefore effect similar structural disruptions as those described for arginine above.

Arg-98, at the C-terminus of helix-$\alpha_2$, forms a salt bridge with Asp-48, located on strand-$\beta 1C$, from the opposite subunit (Figure 6.5(c)). In the amino acid arginine, the guanidino functional group is separated from the backbone by a three carbon aliphatic chain. The salt bridge interaction occurs between the guanidino group of Arg-98 and the carboxylate group of Asp-48. The introduction of a histidine at position 98 would prevent this interaction occurring. Although histidine has a positively charged imidazole ring under certain solution conditions, there is only a single methylene group separating the functional group from the backbone. Therefore, the charged groups of His-98 and Asp-48 would likely not be in sufficient proximity to form a similar stabilising interaction. In addition, the single methylene side chain of histidine would place the imidazole ring in a region previously occupied by the $C\beta$, $C\gamma$ and $C\delta$ CH$_2$ groups of arginine, causing local steric clashes. Finally, of the 23 residues involved directly in the interface, Arg-98 has the largest surface area contribution, approximately 12.5% of the total area concealed on the formation of the homodimer. Due to the smaller side chain, His-98 would have a reduced surface area contribution in this region of the interface and not necessarily provide the stabilising dispersive contacts of Arg-98.

The elution volume of N36W PaDDAH was not affected by changes in the loading concentration. This behaviour was also seen with the obligate homodimer GST. The average elution volume of N36W PaDDAH over all concentrations was 10.18 ± 0.03
ml, which corresponds to an average app. MW of 51.9 ± 0.5 kD. This value compares well with the estimated elution volume of dimeric PaDDAH (Table 6.1) and strongly suggests that the N36W mutation shifted the oligomer equilibrium towards a fixed homodimer. The crystal structure of PaDDAH clearly shows the importance of Asn-36 in the interaction between helices α1A and α1B. It appears from analytical SEC results that the introduction of a tryptophan at position 36, rather than causing massive local steric hindrance, actually increases the stability of the helix-α1 interaction. The residues exposed to solvent in a monomer of WT PaDDAH from helix-α1 are predominantly hydrophilic with only a small hydrophobic patch, comprising Tyr-28, Ala-29 and Leu-32 at the N-terminus of helix-α1 (Figure 6.6(b)). The substitution of a hydrophilic asparagine at position 36 with a large and strongly hydrophobic tryptophan would increase the hydrophobic character of this region of the interface. The results of analytical SEC suggest the N36W mutation forms a hydrophobic core within the interacting side chains of helices α1A and α1B. Such a region would be preferentially concealed from the polar solvent and strengthen the inter-protomer interaction.

The results of analytical SEC studies presented here confirm that WT PaDDAH is a homodimer in solution at higher protein concentrations. Each mutation in this study was designed using the interface predicted in the crystal structure of PaDDAH. As was shown in Section 6.5, the majority of the interface mutations produced a significant alteration in the hydrodynamic behaviour of PaDDAH. It is therefore possible to conclude that the protomer interaction described by Murray-Rust and co-workers exists in the solution state, at least in vitro and at high protein concentration, as well as in the crystalline state.

Each of the PaDDAH mutants were analysed for enzyme activity. In these investigations, only a crude estimate of the amount of protein purified was obtained (see Section 6.4). All the interface mutant proteins produced retained some degree of DDAH activity (Figure 6.9 (b)). The majority of mutants had greater than 50 % of the WT enzyme activity. These samples were also characterised by near WT expression levels (Figure 6.9 (a)). Only R40W and Q43R had significantly reduced expression levels under these conditions and both also had less than 10 % WT activity. The
retention of DDAH activity in all the interface mutations, albeit low in some cases, indicates that the overall fold of the enzyme was never dramatically altered. The side chain of each residue selected in this study appears to interact primarily with residues involved in the inter-subunit interaction. It was therefore expected that each mutation would only have local structural consequences. The measurement of activity for each mutant, especially those with greater than 50 % WT activity, supports this prediction. It is possible with R40W PaDDAH that the mutation may have produced larger structural changes, which would account for the low level of solubility and activity observed.

Combining the conclusions of both the analytical SEC and activity assay investigations reported above suggests that both monomeric and dimeric forms of WT PaDDAH are active. This raises a question about the biological significance of WT PaDDAH existing as a dimeric species. Analytical SEC results (Section 6.5) suggest that the mutations Q43R and R98H forms monomers while activity assays showed both retained DDAH activity. Furthermore, analysis of the oligomer equilibrium of WT PaDDAH revealed a dissociation constant of in the region of 50 to 250 nM. Although the levels of PaDDAH present in vivo are unknown they are unlikely to exceed this value. It could be the PaDDAH homodimer is an artefact of the high concentration of proteins used in vitro. It would be interesting to determine if the monomeric or dimeric mutants produced in this study would allow normal DDAH function in vivo.

Several of the mutant PaDDAH proteins characterised in this chapter gave interesting analytical SEC chromatograms that suggested a change in the oligomer equilibrium position. In the following chapter, Chapter 7, three of the mutants described above are further characterised by solution NMR spectroscopy and the results compared to those obtained for the WT protein.
Chapter VII
Analysis of PaDDAH mutations by NMR

Chapter Summary
In this chapter three of the PaDDAH mutants produced in Chapter 6 will be further analysed using NMR methodologies. NH correlation spectra will be used to assess whether these mutant proteins are similar in structure to WT PaDDAH. The hydrodynamic properties of N36W, R40E and R98H PaDDAH will be investigated. The results from these various analyses of the PaDDAH mutants will be compared to those reported earlier for the wild-type protein.

7.1. Introduction.
WT PaDDAH, with a homodimeric molecular weight of 60 kD, would have previously been considered a large protein in terms of the application of high resolution solution NMR spectroscopy. However, recent advances in experimental procedure and instrument technology mean that NMR spectroscopy can be used to characterise the 3D structure and dynamics of similar sized proteins to PaDDAH (see Chapter 1, Section 1.2 for a more detailed discussion). For many purposes and particularly in cases where the 3D structure is already known, only a rudimentary
level of resonance assignment is required – e.g. limited to sequential resonance assignment of NH groups in a \([^{1}\text{H}, {^{15}\text{N}}]\)-correlation spectrum. Upon achieving this level of resonance assignment, information on structural or dynamic processes can be readily assessed by NMR under different sample conditions and changes in these properties mapped onto the known 3D structure. Therefore, a major initial goal of this project was to obtain backbone NH resonance assignments of WT PaDDAH. As discussed in Chapter 5, backbone \(^{1}\text{H}^{N}\) and \(^{15}\text{N}\) (and \(^{13}\text{C}\)) resonance assignments of WT PaDDAH have not yet been realised (see Chapter 5, Section 5.6).

Investigations that have focused upon disruption of the homodimer interface of WT PaDDAH appear to have yielded promising results (see Chapter 6, Section 6.5). Several mutations of PaDDAH apparently modify the self-association properties of the wild-type protein without any significant impact on enzymatic activity. In particular, three PaDDAH mutants gave interesting analytical SEC results: N36W PaDDAH appeared to have an small elution volume that was not dependent on the loading concentration, which is suggestive of an obligate homodimer (Figure 6.11(a)); R40E PaDDAH yielded a double peaked elution profile with peaks that potentially corresponded to homodimer and monomer species (Figure 6.10); and R98H PaDDAH yielded a single peak with an elution volume tending towards a 30 kD species at the lowest concentration tested (Figures 6.10 and 6.11(a)). In this chapter, these three mutants (N36W, R40E and R98H) of PaDDAH are characterised by NMR spectroscopy and the results are compared to the wild-type protein.

### 7.2. Analysis of R40E and R98H PaDDAH by NMR.

\([^{15}\text{N}]\)-labelled NMR samples of R40E and R98H PaDDAH mutants were expressed and purified following the same protocols used for WT PaDDAH (see Chapter 5, Section 5.1 and references therein). The final purification step was preparatory-scale SEC (see ‘Materials and Methods’ Section 3.2.6). Elution fractions containing PaDDAH protein were confirmed by SDS-PAGE analysis, as with WT samples (e.g. see Chapter 5, Figure 5.1). Due to the low separating capacity of the preparatory-scale SEC column used, it was not possible to separate the two peaks that were observed for R40E PaDDAH at an analytical resolution. Therefore this sample must be considered as containing both species.
7.2.1 - $[^1\text{H},^{15}\text{N}]-\text{TROSY}$ spectra of $[^{15}\text{N}]-\text{labelled}$ WT, R40E and R98H PaDDAH

$[^1\text{H},^{15}\text{N}]-\text{TROSY}$ spectra were recorded using 0.5 mM samples of $[^{15}\text{N}]-\text{labelled}$ WT, R40E and R98H PaDDAH (Figure 7.1). As could immediately be seen upon visual inspection of the three spectra, neither of the mutations appeared to have disrupted the PaDDAH fold. The chemical shifts of the majority of the observed NH cross peaks were unaffected. Differences were only seen for a small number of cross peaks when the three spectra were superposed (data not shown). The approximate numbers of backbone NH cross peaks observed in each spectrum were: WT, 221 cross peaks (Figure 7.1(a)); R40E, 220 cross peaks (Figure 7.1(b)); and R98H, 235 cross peaks (Figure 7.1(c)).

In Chapter 1, Section 1.1.2 the theoretical relationship between molecular weight and signal linewidth was introduced. The linewidth of a NMR signal from a larger molecule with a slower tumbling rate will be broader than that from a smaller molecule. It can be expected that a PaDDAH monomer would have sharper, more intense NMR signals than a PaDDAH homodimer. Given that the spectra presented in Figure 7.1 were recorded from samples with equal protein concentration, any differences in signal intensity (i.e. cross peak height) between the spectrum of WT PaDDAH and those of the PaDDAH mutants might be attributed to changes in the oligomeric state. A crude comparison of the signal intensity observed was afforded by measuring the average NH cross peak S/N ratio in each of the three spectra (see ‘Materials and Methods’ Section 3.4.1). The mean amide cross peak S/N ratio from the $[^1\text{H},^{15}\text{N}]-\text{TROSY}$ spectrum of WT PaDDAH was $29.2 \pm 14.7$, whereas for R40E and R98H PaDDAH the (S/N) was $37.6 \pm 15.8$ and $36.3 \pm 16.1$, respectively. The average cross peak S/N values measured from the spectra of each mutant were higher than that measured in the WT spectrum. However, the large standard deviations obtained for each S/N data set meant that few definitive conclusions could be drawn from these measurements. Therefore, further investigation of the hydrodynamic properties of each sample would be required to verify the changes in oligomeric state.
Figure 7.1. Comparison of [$^1$H, $^{15}$N]-TROSY spectra recorded of [$^1$H, $^{15}$N]-labelled (a) WT PaDDAH, (b) PaDDAH R40E and (b) R98H. Each spectrum was recorded under identical sample conditions at a protein concentration of 0.5 mM. Each data matrix consisted of $256^* (t_1)$ and $2560^* (t_2)$ data points (where $n^*$ refers to complex points) with acquisition times of 64 ($t_1$) and 128 ms ($t_2$). A total of 128 transients were recorded per $t_1$ transient.
of R40E and R98H PaDDAH suggested by analytical SEC (see Chapter 6, Section 6.5).

7.2.2 - Measurement of the translational diffusion rates of R40E and R98H PaDDAH

1D $^1$H NMR diffusion experiments were performed using 0.5 mM samples of R40E and R98H PaDDAH for comparison with WT PaDDAH (see Chapter 5, Section 5.4.2 for an estimation of $D_z$ for WT PaDDAH). All experiments were performed and analysed as previously described (see Chapter 5, Section 5.4.2). For each mutant, the parameters of Equation 5.5 (see Chapter 5, Section 5.4.2) were optimised against the experimental data (see 'Materials and Methods' Section 3.4.4). The results of these fits are presented in Table 7.1 and Figure 7.2. Both R40E (Figure 7.2(a)) and R98H PaDDAH (Figure 7.2(b)) yielded distinctively faster $D_z$ estimates than the wild-type protein (Table 7.1). In Figure 7.2, the fitted curve from Equation 5.5 is plotted using the values presented in Table 7.1 along with the experimental data. For each mutant, the fit of Equation 5.5 using the $D_z$ of WT PaDDAH, but the estimated value of $A_0$ for that mutant, is also plotted (Figures 7.2(a) and (b)). For both mutant cases, the experimental signal attenuation does not fit the curve describing the estimated translational diffusion rate of the WT species.

<table>
<thead>
<tr>
<th>PaDDAH</th>
<th>MW (kD)$^a$</th>
<th>$\Delta$ (ms)$^b$</th>
<th>$D_z \times 10^{-7}$ cm s$^{-1}$$^c$</th>
<th>Standard Deviation $\times 10^{-7}$ cm s$^{-1}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT</td>
<td>64.4</td>
<td>70</td>
<td>7.6</td>
<td>0.1</td>
</tr>
<tr>
<td>R40E</td>
<td>64.4</td>
<td>70</td>
<td>8.9</td>
<td>0.09</td>
</tr>
<tr>
<td>R98H</td>
<td>64.4</td>
<td>70</td>
<td>9.2</td>
<td>0.16</td>
</tr>
</tbody>
</table>

$^a$ predicted protein molecular weight calculated for a homodimer from the amino acid sequence with the affinity tag still present. $^b$ experimental diffusion delay between pulse field gradient pulses; $^c$ diffusion coefficient calculated by fitting Equation 5.5 (see Chapter 5, Section 5.2.4) to the experimental data; $^d$ WT results taken from Chapter 5, Table 5.1.

Mean amide proton $T_2$ times were also measured for 0.5 mM samples R40E and R98H PaDDAH (see Chapter 5, Section 5.4.1). The estimated $^1$H $T_2$ time of R40E PaDDAH (12.4 ms) was greater than that estimated for R98H (8.9 ms). A larger value of this measurement indicates more slowly relaxing transverse proton magnetisation.
Figure 7.2. Estimation of the diffusion coefficient of (a) R40E and (b) R98H PaDDAH. Each experiment was recorded with identical solution conditions and with protein concentrations of 0.5 mM. The top panels in (a) and (b) show the experimental data for each protein (□) and the calculated fit of Equation 5.5 (see Chapter 5, Section 5.4.2) for WT PaDDAH ( — ) and that mutant ( — ). The WT PaDDAH fit was calculated with $D_z = 7.6 \times 10^{-7}$ cm$^2$ s$^{-1}$ (see Chapter 5, Table 5.1) and with the same $A_0$ as the given mutant. The calculated translational diffusion coefficient, $D_z$, and standard deviation (in parentheses) are given for each mutant. The lower panels show the residuals of experimental data from simulated data (residual = $\tilde{y}_i - y_i$). The simulated data was calculated from Equation 5.5 using the stated values of $D_z$ and $A_0$ for each mutant.
7.3. Comparison of $[^2H, ^{15}N]$-labelled WT, N36W and R40E PaDDAH.

In section 7.2, the hydrodynamic and NMR properties of two PaDDAH mutants were analysed. R40E and R98H PaDDAH presented comparable (S/N) values and numbers of NH cross peaks in $[^1H, ^{15}N]$-TROSY spectra (Figure 7.1). In terms of hydrodynamic behaviour, both yielded similar estimates of $D_z$, though R40E had a more favourable average amide $^1H$ T$_2$ time. These findings support the conclusions from Chapter 6 that R98H PaDDAH is a monomer and suggest a similar condition for the R40E mutant. The longer $^1H$ T$_2$ time, however, suggests R40E PaDDAH has more amenable NMR relaxation properties.

In Chapter 5 a potential reason for not observing a full complement of NH cross peaks in $[^1H, ^{15}N]$-spectra of WT PaDDAH was structural plasticity of the protomer-protomer interaction. The existence of a homodimer-monomer equilibrium for WT PaDDAH was validated in analytical SEC examination presented in Chapter 6. In analytical SEC experiments, the N36W mutation appeared to produce a more strongly associated PaDDAH homodimer. This mutant protein was therefore chosen for further analysis by NMR.

$[^2H, ^{15}N]$-labelled NMR samples of N36W and R40E PaDDAH were expressed and purified following the same protocols used for $[^2H, ^{15}N]$-labelled WT PaDDAH samples (see Chapter 5, Section 5.1 and references therein). As mentioned in Section 7.2, the low separating capacity of the preparatory-scale SEC column used in purification prevented separation of the two elution peaks that were observed for R40E PaDDAH at an analytical resolution. The affinity tag was removed from all $[^2H, ^{15}N]$-labelled PaDDAH samples during purification (see ‘Materials and Methods’ Section 3.2.6).

7.3.1 - $[^1H, ^{15}N]$-TROSY spectra of $[^2H, ^{15}N]$-labelled WT, N36W and R40E PaDDAH

$[^1H, ^{15}N]$-TROSY spectra were recorded of 0.5 mM samples of $[^2H, ^{15}N]$-labelled WT, N36W and R40E PaDDAH (Figure 7.3). The $[^1H, ^{15}N]$-TROSY spectra of $[^2H, ^{15}N]$-labelled N36W and R40E PaDDAH were extremely similar to the spectrum of the wild-type protein in that they each showed good dispersion and essentially
Figure 7.3. Comparison of [1H, 15N]-TROSY spectra recorded of [2H, 15N]-labelled (a) WT PaDDAH, (b) N36W PaDDAH and (c) R40E PaDDAH. Each experiment was performed under identical sample conditions and at a protein concentration of 0.5 mM. Each data matrix consisted of 256* (t1) and 2560* (t2) data points (where n* refers to complex points) with acquisition times of 64 (t1) and 128 ms (t2). A total of 16 transients were recorded per t1 transient.
identical patterns of chemical shifts. However the approximate number of amide cross peaks was still below the predicted number from the amino acid sequence. The [¹H, ¹⁵N]-TROSY spectrum of N36W PaDDAH contained approximately 180 observable NH cross peaks, while R40E had approximately 200 cross peaks.

7.3.2 - Comparison of ¹⁵N relaxation rates of WT, N36W and R40E PaDDAH

¹⁵N T₁, ¹⁵N T₂ and {¹H}¹⁵N-NOE experiments were recorded for 0.5 mM samples of [²H, ¹⁵N]-labelled N36W and R40E PaDDAH (see ‘Materials and Methods’ Section 3.3.5). For each mutant, ¹⁵N R₁, ¹⁵N R₂ and heteronuclear NOE values were derived for a sample of well resolved cross peaks in the reference [¹H, ¹⁵N]-TROSY spectrum. These values were analysed to extract mean ¹⁵N R₂/R₁ ratios and an estimate of the isotropic rotational correlation time τₑ using the protocols outlined in Section 4.2 (see also ‘Materials and Methods’ Section 3.4.2). The results of this analysis are presented in Table 7.2.

<table>
<thead>
<tr>
<th>PaDDAH sample</th>
<th>Number of Peaks Analysed</th>
<th>R₂/R₁</th>
<th>τₑ (ns)</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT</td>
<td>83</td>
<td>118.9 (19.9)</td>
<td>35.2 (0.1)</td>
</tr>
<tr>
<td>N36W</td>
<td>37</td>
<td>99.5 (29.3)</td>
<td>31.1 (2.8)</td>
</tr>
<tr>
<td>R40E</td>
<td>34</td>
<td>48.2 (10.9)</td>
<td>21.9 (0.9)</td>
</tr>
</tbody>
</table>

* refers to the number of selected peaks remaining following a statistical (n > 1.5 in Equation 4.3) and NOE (> 0.65) filter (see Chapter 4, Section 4.2.1); b mean R₂/R₁ values plus standard deviations in parentheses; * estimated isotropic rotational correlation time (see ‘Materials and Methods’ Section 3.4.2 for descriptions of calculations). d WT values from Chapter 5, Table 5.2.

Figure 7.4 shows the distribution of ¹⁵N R₂/R₁ ratios of N36W and R40E PaDDAH plotted with that observed for the wild-type protein (WT PaDDAH data taken from Chapter 5, Figure 5.9). The distribution of ¹⁵N R₂/R₁ ratios for N36W and WT PaDDAH clearly overlapped (Figure 7.4(a)). The difference in τₑ for these two proteins (see Table 7.2) falls within the uncertainty described by the standard deviation of each value. The mean of the distribution of ¹⁵N R₂/R₁ ratios for R40E PaDDAH, 48.2 ± 10.9, was significantly lower than that for the wild-type protein R₂/R₁ data set (Table 7.2). Figure 7.4(b) shows, unmistakably, that there is no overlap of the ¹⁵N R₂/R₁ data sets for R40E and WT PaDDAH.
Figure 7.4. $^{15}$N $R_2/R_1$ ratios (grey bars in both cases) recorded from [$^2$H,$^{15}$N]-labelled (a) N36W PaDDAH and (b) R40E PaDDAH. Values were calculated from well resolved NH cross peaks (See ‘Materials and Methods’ Section 3.4.3 for description of calculations). $^{15}$N $R_2/R_1$ ratios for WT [$^2$H,$^{15}$N]-labelled PaDDAH (white bars) at 25°C are also shown (see Chapter 5, Section 5.3.3).
For each PaDDAH mutant, the mean $^{15}$N $R_2/R_1$ ratio was used to estimate the isotropic rotational correlation time using the protocol outlined in Chapter 4, Section 4.2. The values of $\tau_c$ estimated for N36W and R40E PaDDAH are compared to that obtained for the wild-type protein in Table 7.2. As was predicted from the similarities observed between the distribution of $^{15}$N $R_2/R_1$ ratios (Figure 7.4(a)), the derived rotational correlation times of WT and N36W PaDDAH are similar: $35.2 \pm 0.1$ and $31.1 \pm 2.8$ ns respectively. However, the isotropic correlation time of R40E PaDDAH, $21.9 \pm 0.9$ ns, was almost 40% smaller than that of the wild-type protein.

7.4. Discussion.

Extensive unambiguous sequence specific backbone $^1$H$^N$, $^{15}$N resonance assignments for WT PaDDAH have not yet been obtained (see Chapter 5, Section 5.5). This was primarily due to an incomplete complement of cross peaks in both NH correlation and triple-resonance NMR experiments. Several hypotheses that may account for these missing correlations are discussed in Chapter 5, Section 5.7. One potential explanation concerns the loss of signals due to chemical exchange at the oligomer interface of PaDDAH, such as might arise due to a dynamic homodimer-monomer equilibrium. The existence of such an equilibrium was demonstrated for WT PaDDAH in Chapter 6, Section 6.2.1. A series of point mutations of residues at the protomer interface were produced with the aim of artificially manipulating the equilibrium observed for the wild-type protein. In this chapter the PaDDAH mutants N36W, R40E and R98H were further characterised using NMR methodologies.

$[^1$H, $^{15}$N]-TROSY spectra were recorded for $[^{15}$N]- or $[^2$H, $^{15}$N]-labelled mutant samples and compared to an equivalent spectrum of WT PaDDAH (Figures 7.1 and 7.3). In each case on superposition of corresponding WT and mutant spectra, changes in chemical shift were noted for only a few NH cross peaks with the remainder being effectively equivalent (data not shown). The NMR spectrum of a protein is highly sensitive to changes in the 3D structure. Changes in the 3D structure will alter the local chemical environment causing a change in chemical shift. As changes in
chemical shifts were only observed for a very limited number of NH cross peaks in the spectra presented in Figures 7.1 and 7.3, it can be safely concluded that the overall structure of each mutant analysed by NMR spectroscopy in this chapter closely resembles that of the wild-type protein. These conclusions are reinforced by the results of the enzyme activity assays performed in Chapter 6. Enzyme activity is sensitive to perturbation of the 3D structure and particularly the active site architecture. Therefore, the fact that all three mutants analysed by NMR in this chapter retained DDAH activity to some degree (see Chapter 6, Figure 6.9(b)) strongly suggests that each mutation provoked only minor changes in the 3D structure of PaDDAH, most likely localised to the site of the mutated residue.

The number of cross peaks observed in \([^1\text{H}, ^{15}\text{N}]-\text{TROSY spectra of}[^{15}\text{N}]-\text{labelled WT, R40E and R98H PaDDAH} was approximately the same (Section 7.2.1). However, a full complement of amide NH cross peaks for tagged PaDDAH was still not observed. In the case of \([^2\text{H}, ^{15}\text{N}]-\text{labelled samples, which were not affinity tagged, the}[^1\text{H}, ^{15}\text{N}]-\text{TROSY spectra of R40E PaDDAH} contained significantly more NH cross peaks than both WT and N36W PaDDAH (Section 7.2.2). In Chapter 5 only 160 amide NH cross peaks were reported in the \([^1\text{H}, ^{15}\text{N}]-\text{TROSY spectra of}[^2\text{H}, ^{12}\text{N}]-\text{labelled WT PaDDAH. The total number of cross peaks counted for R40E PaDDAH (ca. 200) represents an increase of 25 \% over the number seen in the spectrum of the wild-type protein. However, again a full complement of 246 amide cross peaks for de-tagged PaDDAH was still not observed.}

The gain in sensitivity observed in \([^1\text{H}, ^{15}\text{N}]-\text{TROSY spectra achieved on perdeuteration of PaDDAH allows the application of more advanced NMR experiments and pulse sequences. However, the potential problem of incomplete 'back protonation' can still persist. In the case of R40E PaDDAH, more NH cross peaks were observed in the spectrum recorded from the \([^{15}\text{N}]-\text{labelled sample (Figure 7.1(b)) than in that from a}[^{2}\text{H}, ^{15}\text{N}]-\text{labelled sample (Figure 7.3(c)). However, each spectrum of the}[^{12}\text{N}]-\text{labelled samples presented in Figure 7.1 required a total acquisition time of 16 hours with 128 transients recorded per} t_1 \text{ increment. In comparison, the spectra of the perdeuterated samples in Figure 7.3 each required less than two hours to record with only 16 transients recorded per} t_1 \text{ increment. In the characterisation of protein structure and function using NMR many experiments, from}
ligand-binding titrations to sequential resonance assignment, necessitate that a high sensitivity 2D correlation map with good signal resolution can be obtained in a relatively short period. Therefore, although more NH cross peaks were observed in spectra from $[^{15}\text{N}]$-labelled samples, it is considerably more practical to utilise the benefits in sensitivity afforded by deuteration. For this reason analysis of $^{15}\text{N}$ nuclear spin relaxation was performed on $[^{2}\text{H},^{15}\text{N}]$-labelled samples.

In Chapter 5, the solution properties of WT PaDDAH were described by an average amide $^1\text{H} T_2$ time, a translational diffusion coefficient and an estimate of the rotational correlation time derived from $^{15}\text{N}$ relaxation terms. Each parameter indicated that the wild-type protein behaved as a 60 kD species in solution. Identical investigations into the translational diffusion rates of R40E and R98H PaDDAH were conducted for comparison (Section 7.2.2). Although a complete description of the hydrodynamic properties of WT PaDDAH is unavailable, one can compare the values of $D_z$ obtained for mutant and WT forms of the enzyme. Both R40E and R98H PaDDAH presented faster translational diffusion rates than the wild-type protein suggesting that both mutants behave as smaller molecules (Table 7.1). In the absence of more extensive knowledge of the parameters that describe the hydrodynamic behaviour of PaDDAH there is no accurate way of directly correlating values of $D_z$ with apparent molecular mass. However, these results suggest these two mutants are less highly self-associated than the wild-type proteins, consistent with the observations of hydrodynamic behaviour by analytical SEC (see Chapter 6, Section 6.5).

A more comprehensive hydrodynamic analysis was performed on $[^{2}\text{H},^{15}\text{N}]$-labelled R40E and N36W PaDDAH using $^{15}\text{N}$ nuclear spin relaxation experiments. The estimated rotational correlation times for N36W and WT PaDDAH are approximately equal. A considerable decrease in the estimated $\tau_c$ for R40E PaDDAH was observed in comparison to that established for the wild-type protein.

Similar relaxation analyses of $[^{2}\text{H},^{15}\text{N}]$-ubiquitin in 50 % (v/v) glycerol (see Chapter 4, Section 4.3.4) showed that $^{15}\text{N}$ relaxation parameters estimated for proteins with larger $\tau_c$ values tended to exhibit a greater level of experimental uncertainty. This imprecision was also noted in the analysis of $^{15}\text{N}$ relaxation parameters for N36W and R40E PaDDAH here and for WT PaDDAH in Chapter 5. The source of the increased
uncertainty is most likely the decrease in NH cross peak S/N in larger proteins coupled with the more rapid relaxation of transverse magnetisation in $^{15}$N $T_2$ experiments. To extract a meaningful relaxation rate for a given $^{15}$N nucleus, a smooth decay in signal intensity is required over the range of delay periods recorded. Furthermore, the relative decay in signal intensity should preferably be of around one order of magnitude over the series. When attempting to extract an accurate relaxation rate it is also important to ensure that the cross peak signal intensity measured at each time point originates from a single NH group. The presence of overlap of cross peaks from two NH groups with different $^{15}$N relaxation rates can lead to erroneous derivation of the rate parameter. The use of such peaks should thus be avoided. Only those amide cross peaks that met these requirements and had well fitted relaxation decay curves were included in the $^{15}$N $R_2/R_1$ data set used in the estimation of the isotropic rotational correlation times presented here. It should also be noted that $^{15}$N $R_2/R_1$ ratios from a fewer amide groups were used in the estimation of $\tau_c$ for the PaDDAH mutants compared to wild-type enzyme (Table 7.2). $^{15}$N relaxation spectra of both $[^2H,^{15}N]$-labelled mutants suffered from low S/N ratio and cross peak overlap and therefore it was only possible to extract signal intensities from a relatively restricted number of NH cross peaks. This was not the case for corresponding experiments with the wild-type protein (see Chapter 5, Section 5.4.3). The implementation of the TROSY principle may have afforded better cross peak resolution and permitted the inclusion of $^{15}$N $R_1$ and $R_2$ rates from a larger portion of amide NH groups.

The S/N ratio of a NH cross peak is inherently dependent on the transverse relaxation rates of the $^{15}$N and $^1H^N$ nuclei (see Chapter 1, Section 1.1.2). For the reasons discussed above, cross peaks were only selected for analysis if they had suitable signal intensity. Therefore it is possible that the resultant $^{15}$N $R_2/R_1$ data set may not be properly representative and biased towards cross peaks with low $R_2$ rates. Ideally, relaxation data should be calculated for each amide $^{15}$N nucleus to prevent the introduction of bias into the estimation of $\tau_c$.

In summary, the work presented in this chapter further supports the results and conclusions of analytical SEC experiments in Chapter 6. The combination of the results obtained through analytical SEC, translation diffusion and $^{15}$N relaxation
experiments strongly suggest that R40E PaDDAH behaves as a significantly smaller species than the wild-type protein. These findings strongly indicate that the R40E mutation has generated a monomeric form of the enzyme. R98H PaDDAH has also yielded promising analytical SEC and translational diffusion results. It would appear that this mutation also caused the formation of a stable more monomeric variant of PaDDAH. The extraction of $^{15}$N relaxation parameters from N36W PaDDAH supported the earlier conclusion that this mutant is a more strongly-associated PaDDAH homodimer.
The interplay between the roles of NOS and DDAH enzymes in humans make both classes of protein important focal points for biochemical research. The regulation of NOS isoforms by asymmetric methylarginines has been studied intensively (see Chapter 2, Section 2.1.4). However, the role of DDAH isoforms in controlling asymmetric methylarginine levels and therefore the production of NO is less widely characterised. In 1996 it was shown that inhibition of DDAH induced contraction of aorta sections in vitro (MacAllister et al. 1996). Since that observation the significance of the role of DDAH in NO regulation has become increasingly apparent (Leiper and Vallance 1999). A recent report suggested a mechanism of negative feedback for DDAH (Leiper et al. 2002). The authors proposed that DDAH is inhibited by nitrosylation of the active site cysteine by free NO. Presumably, under conditions of elevated NO, DDAH activity would be inhibited. As a result, the levels of asymmetric methylarginines would increase thereby inhibiting NOS and further NO production. Such a regulatory mechanism would increase the complexity of the 'cross talk' between these two metabolic pathways (see Chapter 2, Figure 2.3).
The work presented in this thesis forms part of a multidisciplinary, cross-department collaboration that is utilising the complementary analyses of X-ray crystallography, NMR spectroscopy and clinical pharmacology. The interest in studying DDAH from both a structural biological and a pharmacological perspective is obvious. A combined objective of the DDAH project is the development of novel ligands that inhibit human DDAH isoforms. Such ligands could potentially be used as a chemical ‘knock-out’ with which to probe the function of DDAH in vivo. Inhibitors of this nature may also represent a potential lead in the development of novel therapeutics that exploit the biology of DDAH. Two isoforms of human DDAH exist with differing tissue specific expression patterns (see Chapter 2, Section 2.1.5). Therefore, it may even be possible to develop isoform-specific and therefore tissue-specific inhibitors of DDAH.

There are many ways in which NMR analysis can aid in such a project, especially when the structure of the protein is already known. In the case of DDAH, the 3D structure of a bacterial isoform was elucidated using X-ray crystallography by a team of collaborators (Murray-Rust et al. 2001). The analysis of dynamics by NMR spectroscopy allows a unique insight into the mobility of individual bonds in a 3D structure of the protein. No other technique can couple such molecular dynamics to 3D structure in this way. Several biologically relevant questions remain unanswered by the determination of the 3D structure of PaDDAH. Most notably, the 3D structure of a 12-residue loop that is believed to close over the active site of the enzyme on substrate binding was not resolved in the apo-enzyme.

As discussed in the introduction to Chapter 5, solution NMR spectroscopy can be an extremely useful tool for the development of high affinity ligands. The insight available from NMR-based screening methods is substantially increased when the 3D structure of the target protein is known, which is the case for PaDDAH. One of the advantages of using 2D heteronuclear NMR methods to characterise ligand binding is that the changes in chemical shift observed can be readily mapped onto the 3D structure of the target protein. Therefore, 2D heteronuclear NMR can identify novel ligands whilst simultaneously providing information about the nature of the interaction. Alternative strategies require ligand identification and structural characterisation of the resulting complex(es) to occur separately. However, to gain
this level of insight from heteronuclear NMR, resonance assignment of the spectrum is a primary requirement.

The remainder of this chapter will discuss how the work presented in this thesis has moved the DDAH project closer to its goals. Figure 8.1 provides the reader with a short chapter-by-chapter summary of the results presented.

PaDDAH is a large protein with a homodimeric mass of approximately 60 kD. As such, it is a challenging protein to investigate with solution NMR techniques. There are many intrinsic problems encountered when studying larger proteins by NMR (these are discussed in Chapter 1, Section 1.1). Two major improvements in experimental technique have recently combined to raise the upper molecular weight limit in solution NMR spectroscopy of proteins. These are the expression of perdeuterated, amide protonated proteins and the use of TROSY-based pulse sequences (see Chapter 1, Section 1.2). Both of these strategies lead to the improvement of signal linewidths of larger proteins. In summary, deuteration can reduce the transverse relaxation rate of $^{13}$C and $^1$H nuclei (see Chapter 1, Section 1.2.1), whereas TROSY-based pulse sequences select the component of the four-fold NH multiplet for which the combination of dipolar and CSA relaxation is optimally suppressed (see Chapter 1, Section 1.2.2). The combined application of these methods can make resonance assignment of large proteins possible where traditional spectroscopic NMR performed on non-deuterated proteins would probably have failed.

To investigate the likelihood of recording tractable NMR data for PaDDAH, a model protein system was developed to mimic high molecular weight proteins (see Chapter 4, Section 4.1). The model consisted of isotope-labelled ubiquitin, a well-characterised small protein, in 50% (v/v) glycerol. The tumbling rate of ubiquitin in this solvent was shown to be dependent on the temperature of the sample (see Chapter 4, Figure 4.6). It was also demonstrated that perdeuteration and the use of TROSY-based NMR experiments are essential when studying proteins with slow $\tau_c$ times. For example, $^{15}$N relaxation experiments performed on the ubiquitin model highlighted the limitations of using HSQC-based experiments on proteins with longer $\tau_c$ times (see Chapter 4, Section 4.2.4). In general, isotope-labelled perdeuterated ubiquitin
Chapter IV
Isotope labelled ubiquitin in 50% glycerol proved a 'tunable' model for larger proteins. A ~ 20 ns range of $\tau_c$ values was recorded over 5 sample temperatures (17.5°C to 34.3°C). Backbone resonance assignment of ubiquitin in 50% (v/v) glycerol at 25°C was performed using just TROSY-HNCA and TROSY-HN(CA)CB experiments.

Chapter V
WT PaDDAH was characterised by solution NMR. [$^1$H, $^{15}$N]-TROSY spectra of [$^2$H, $^{15}$N]-labelled samples looked promising although were missing some NH cross peaks. Evaluation of hydrodynamic properties supported the published symmetric homodimer structure ($\tau_c \sim 35$ ns). Unambiguous backbone assignments were not obtained for WT PaDDAH.

Chapter VI
A dynamic oligomer equilibrium was observed for WT PaDDAH in analytical SEC experiments (50 nM $<$ $K_d$ $<$ 250 nM). Analysis of homodimer interface structure identified key residues in protomer interaction. A selection of residues were separately mutated and the derived mutant proteins analysed for solubility, enzyme activity and hydrodynamic behaviour. Several of the mutants had modified SEC behaviour whilst retaining enzymatic activity.

Chapter VII
Three PaDDAH mutants were further analysed by NMR. NH correlation spectra of each mutant strongly suggested little change in the overall 3D structure of the protomer. Estimates of $D_2$ for R40E and R98H PaDDAH suggested these mutants were smaller molecular species than WT. $^{15}$N relaxation measurements suggested that N36W PaDDAH is a strongly associated dimer ($\tau_c \sim 31$ ns) and R40E PaDDAH a monomer ($\tau_c \sim 22$ ns).

Figure 8.1. Brief summary of the results presented in Chapters 4 to 7 of this thesis. This figure is not intended to provide an exhaustive summary of the results presented in each chapter.
satisfied the original objective of producing a model of larger proteins with adjustable hydrodynamic properties. The successful application of 3D triple-resonance experiments employing TROSY selection to the ubiquitin model (see Chapter 4, Section 4.3) suggested that similar experiments could also be successfully applied to WT PaDDAH.

The use of the TROSY principle in concert with deuteration indeed benefited NMR spectra of WT PaDDAH (see Chapter 5, Sections 5.2 and 5.3), which is not unexpected when considering the predicted size of the protein as a homodimer. Even so, it was not possible to obtain extensive unambiguous resonance assignments of WT PaDDAH with the recorded spectra (see Chapter 5, Section 5.6). The isotropic rotational correlation time established for WT PaDDAH is ca. 35 ns at 25°C (see Chapter 5, Section 5.4.3), which is approximately equal to that of ubiquitin under the conditions in which sequential resonance assignment was successfully performed. This outcome poses an interesting quandary: is the ubiquitin model presented in Chapter 4 unrepresentative of larger proteins or are the problems encountered when attempting resonance assignment of WT PaDDAH specific to this protein? In terms of \(^{15}\)N relaxation measurements, both proteins were shown to have similar properties. For example, the distribution of experimental \(^{15}\)N \(R_2/R_1\) ratios of ubiquitin in 50 % glycerol at 25°C and WT PaDDAH overlap (see Chapter 5, Figure 5.9). An important determinant of the success of 3D triple resonance experiments is the rate of relaxation of \(^{13}\)C transverse magnetisation. Although \(^{13}\)C relaxation rates were not derived experimentally for either protein, the presence of \(^{13}\)C cross peaks in 3D triple resonance spectra provides some anecdotal evidence of these properties. The two 3D spectra recorded of triple-labelled ubiquitin (shown in Chapter 4, Figure 4.9(a)) showed close to a full complement of \(^{13}\)C intra- and inter-residue correlations (see Chapter 4, Figure 4.8). All of the alanine, glycine and serine and threonine spin systems were easily identified (see Chapter 4, Table 4.3). In addition, 55 out of 70 amide NH cross peaks observed in a NH correlation spectrum of triple-labelled ubiquitin yielded the maximum number of \(^{13}\)C correlations in the 3D data sets (see Chapter 4, Figure 4.8). That the majority of expected \(^{13}\)C cross peaks were observed in 3D spectra of ubiquitin indicates that magnetisation transfer levels were good. This further suggests that \(^{13}\)C \(T_2\) times were not a limiting factor for these experiments. This was not the case for WT PaDDAH. \(^{13}\)C correlation statistics in 3D spectra of WT
PaDDAH were much less encouraging. $[^1H, ^{15}N]$-TROSY spectra of perdeuterated WT PaDDAH did not contain the full complement of amide NH cross peaks. Therefore, a suitable comparison between the 3D spectra of WT PaDDAH and ubiquitin is the number of $^{13}C$ cross peaks observed per amide NH group. Unlike the case for ubiquitin in 50 % glycerol at 25°C, only 20 % of the amide NH cross peaks in a $[^1H, ^{15}N]$-TROSY spectrum of WT PaDDAH yielded the maximum number of $^{13}C$ correlations (see Chapter 5, Figure 5.16). Furthermore, many amide NH cross peaks with good S/N ratios did not yield $^{13}C$ correlations in 3D spectra of PaDDAH. In summary, ubiquitin in 50 % glycerol at 25°C and WT PaDDAH at 25°C had comparable $^{15}N$ relaxation properties. However, there appeared to be differences in $^{13}C T_2$ times between the two proteins as evidenced by stark differences in the number of $^{13}C$ correlations observed in each set of 3D spectra. Although the problem of chemical shift degeneracy does increase with the size of the protein, it is highly improbable that overlap of cross peaks can solely account for the correlations not observed in spectra of WT PaDDAH.

The lack of a full complement of NH peaks in $[^1H, ^{15}N]$-TROSY spectra of perdeuterated WT PaDDAH plus low magnetisation transfer efficiency in 3D experiments prevented unambiguous resonance assignments being obtained. There are two potential explanations for these ‘missing’ correlations. As was noted, a large portion of the expected amide NH cross peaks were not observed in $[^1H, ^{15}N]$-TROSY spectra of perdeuterated WT PaDDAH samples (see Chapter 5, Section 5.3). During the purification of deuterated proteins it is possible that full protonation of amide groups will not be achieved (see Chapter 1, Section 1.2.1). Both the 2D and 3D experiments recorded on WT PaDDAH were of the ‘out-and-back’ variety. These pulse sequences necessitate the presence of an amide proton from which to initiate magnetisation transfer and detect signal from. Evidently, unexchanged amide ND groups would not yield cross peaks in such sequences. Therefore, both NH and $^{13}C$ cross peaks would be missing from subsequent spectra. This does not explain why observable amide NH cross peaks with good S/N ratio did not show $^{13}C$ cross peaks in 3D triple-resonance spectra. Another factor that may contribute to missing amide NH cross peaks is the result of line broadening due to intermediate exchange. Several sources of conformational exchange have been suggested for WT PaDDAH. These include the presence of a mobile loop (residues 14 to 25) near the active site and
existence of a dynamic equilibrium between homodimer and monomer species. It is likely that failure to observe the expected number of amide NH cross peaks is due to a combination of factors including contributions from incomplete 'back-protonation' and chemical exchange.

In response to the earlier question regarding the utility of the ubiquitin model, it appears that many of the problems that have prevented resonance assignment of WT PaDDAH are inherent to this protein. The major obstacles of potential incomplete protonation of amide groups, low numbers of $^{13}$C cross peaks and the potential for chemical exchange did not pose any difficulties during the analysis of ubiquitin in glycerol. However, a drawback of this model is that ubiquitin is a small, monomeric protein and consequently NMR spectra are relatively uncrowded and well-resolved in comparison to large proteins with similar tumbling rates.

One possible source of chemical exchange in WT PaDDAH arises from changes in the oligomeric state of the protein as a result of a dynamic equilibrium. Analytical SEC experiments shows that such an equilibrium between monomer and homodimer does exist for the wild-type protein. The dissociation constant for this equilibrium was estimated to be of the order of 50 to 250 nM. If the exchange rate between monomer and homodimer species is on the appropriate time scale, the signal linewidths of many of the nuclei involved in the protomer interaction could be broadened as a result. The protomer interface constitutes approximately 8% of the overall accessible surface area of a PaDDAH monomer. Therefore, a substantial number of heteronuclear cross peaks could be affected. With this possibility in mind a series of interface mutants were generated with the objective making the formation of the homodimer less favourable. The potential benefits of this strategy are two-fold. Not only might a monomer mutant of PaDDAH be free from the effects of exchange broadening, but the effective molecular weight of the protein would also be dramatically reduced. The analysis of several interface mutants highlighted the sensitivity of the WT protomer interaction to small changes in the primary structure (Chapter 6, Section 6.5). All of the soluble PaDDAH interface mutants analysed by analytical SEC showed a deviation from the wild-type behaviour. One of the point mutants, N36W PaDDAH, appeared to abolish the concentration dependence of the experimental elution volume (see Chapter 6, Figure 6.11(a)). This suggested the formation of a higher affinity
interface. Several other mutants, namely R40E, Q43R and R98H PaDDAH, appeared to drive the equilibrium towards a monomeric species (see Chapter 6, Figure 6.11(a)). N36W, R40E and R98H PaDDAH were also analysed using NMR methodologies (see Chapter 7). The results of these studies largely confirmed conclusions derived from analytical SEC results.

The initial objectives of this project were adventurous considering the size of the protein being investigated. However, the current position is a promising one and the results presented in this thesis will provide the foundations for many potentially fruitful lines of research. The ubiquitin model that was produced and tested in Chapter 4 has been shown to be a suitable system for testing the utility of heteronuclear NMR experiments over a range of protein $\tau_c$ times. With the information presented here, NMR pulse sequences can be tested for proteins described by $22 \text{ ns} < \tau_c < 45 \text{ ns}$. The range of the model can be extended to include longer $\tau_c$ times by characterising the tumbling rate of ubiquitin at lower temperatures. Such an extension would appear to require using TROSY-based $^{15}$N relaxation experiments rather than HSQC-based ones.

In the case of PaDDAH, many of the problems that have prevented resonance assignment of this protein have been identified and can now be resolved. For example, since identifying the problem of incomplete protonation of deuterated PaDDAH samples, research is currently being undertaken in our laboratory to explore the possibility of unfolding PaDDAH samples during the purification protocol. This line of investigation is yielding encouraging results. In addition, a more definitive estimate of the $K_d$ for the homodimer-monomer equilibrium is presently being sought through the use of analytical ultracentrifugation techniques. Of the mutants analysed by NMR in Chapter 7, R40E PaDDAH appears a very promising lead for the generation of a monomeric variant. The difference in $\tau_c$ for this protein in comparison to WT PaDDAH should mean tangible improvements in the rate of $^{13}$C transverse relaxation and therefore increase the S/N ratio in 3D triple resonance experiments spectra.

There are several extensions of the work presented in this thesis that immediately spring to mind. $^{15}$N relaxation analyses were only performed on a small subset of the
soluble mutant forms of PaDDAH generated in Chapter 6. Other notable PaDDAH mutants of interest that could have been analysed in this manner include N36D PaDDAH (showed a double peaked elution profile; Figure 6.11(b)), Q43R (had elution volumes tending towards a monomer molecular weight) and R98H PaDDAH (has a faster $D_z$ value to the WT; Figure 7.2(b)). Alternatively, a more extensive trial of mutations could be undertaken. This could involve either trying different substitutions of those residues already targeted in this thesis or selecting alternative residues at the interface that have not been yet been mutated. A further extension, and perhaps more promising, would be to capitalise on the behaviour of the mutations presented here to design double point mutations. Due to the symmetric structure of the interface, each individual mutation in the amino acid sequence effectively produces two at the level of the quaternary structure. Therefore two mutations of the primary structure would have an even greater impact on the protomer interaction. From the results presented in Chapter 7, it would be predicted that the double mutant R40E-R98H PaDDAH should form a stable, monomeric protein. Such a mutant protein may represent a even more suitable candidate for future NMR analyses.

\[ d = \frac{\mu_0 \gamma_H \gamma_N h}{8\pi^2 r^3} \]  \hspace{1cm} \text{Equation A.1}

\[ c = \frac{\Delta \sigma \omega_N}{\sqrt{3}} \]  \hspace{1cm} \text{Equation A.2}

\[ J(\omega) = \frac{2}{5} \frac{\tau_e}{1 + (\omega \tau_e)^2} \]  \hspace{1cm} \text{Equation A.3}

\[ R_1 = \frac{d^2}{4} \left[ 4J(0) + J(\omega_H - \omega_N) + 3J(\omega_N) + 6J(\omega_H + \omega_N) \right] + c^2 J(\omega_N) \]  \hspace{1cm} \text{Equation A.4}

\[ R_2 = \frac{d^2}{8} \left[ 4J(0) + J(\omega_H - \omega_N) + 3J(\omega_N) + 6J(\omega_H) + 6J(\omega_H + \omega_N) \right] \]
\[ + \frac{c^2}{6} \left[ 4J(0) + 6J(\omega_N) \right] \]  \hspace{1cm} \text{Equation A.5}

\[ J(\omega) = \frac{2}{5} \left[ \frac{\Delta S^2 \tau_e}{1 + (\omega \tau_e)^2} + \frac{(1 - \Delta S^2)\tau}{1 + (\omega \tau)^2} \right] \]  \hspace{1cm} \text{Equation A.6}

where \( \mu_0 \) is the permeability of free space, \( \gamma_H \) and \( \gamma_N \) are the gyromagnetic ratios of the \(^1\)H and \(^{15}\)N nuclei respectively, \( h \) is the Planck’s constant, \( \Delta \sigma \) is the chemical shift anisotropy of \(^{15}\)N, \( \tau_e \) is the isotropic rotational correlation time, \( r \) is length of the NH bond, \( J(\omega) \) is the spectral density at the frequency in parentheses and \( \omega_H \) and \( \omega_N \) are the Larmor frequencies of the \(^1\)H and \(^{15}\)N nuclei respectively. The simplified form of the spectral density function is given in Equation A.3. The spectral density function according to the ‘model-free’ formulism is given in Equation A.6. \( S^2 \) is a generalised order parameter and \( \tau^{-1} = \tau_c^{-1} + \tau_e^{-1} \) where \( \tau_e \) is an effective correlation time describing rapid internal motions. Equations taken from Palmer 2001 and Lipari and Szabo 1982a,b.
A.2. Evaluation of the Minimisation Function used to Optimise the Fit of Equation 6.12.

In Chapter 6, Section 6.2.4, Equation 6.12 was used to describe the change in apparent molecular weight of WT PaDDAH sample at different concentration eluting from an analytical SEC column. The equation was cast in terms of protein concentration, $A_0$, app. MW of homodimer and monomer species, $MW_D$ and $MW_M$ respectively, and the equilibrium dissociation constant $K_d$. To estimate the three unknown parameters ($MW_D$, $MW_M$ and $K_d$) Equation 6.12 was optimised against the experimental analytical SEC data using a 3-parameter least squares minimisation (see 'Materials and Methods' Section 3.4.5 for further details). The fit returned unsatisfactory confidence levels for both $MW_M$ and $K_d$ (Table 6.1).

Analysis of the $\chi^2$ minimisation function used to optimise Equation 6.12 is presented here. The $\chi^2$ function is analysed by way of a 2D contour plot where, in turn, one of the three unknown three parameters was fixed while the other two varied. The evaluation of $\chi^2$ under the definitions and ranges specified in Table A.1 is presented in Figure A.1. Each contour plot demonstrates the minimum of the $\chi^2$ function over two dimensions. For example, in Figure A.1(b) $MW_M$ is fixed at 36.3 kD (Table A.1) and $MW_D$ and $K_d$ varied over the ranges stated in Table A.1. A minimum is observed for values of $MW_D$ approaching 52 kD and value of $K_d$ tending towards 240 nM. The ranges of the values chosen for each of the three unknown parameters used in the evaluation of $\chi^2$ were constrained around those estimated in the original fit (Table 6.1).

<table>
<thead>
<tr>
<th>Defined Parameter</th>
<th>Value *</th>
<th>Floated Parameter</th>
<th>Range</th>
<th>Figure</th>
</tr>
</thead>
<tbody>
<tr>
<td>$MW_D$</td>
<td>52.2</td>
<td>$MW_M$</td>
<td>25 - 45 kD</td>
<td>A.1 (a)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$K_d$</td>
<td>10 - 500 nM</td>
<td></td>
</tr>
<tr>
<td>$MW_M$</td>
<td>36.3 ml</td>
<td>$MW_D$</td>
<td>42 - 62 kD</td>
<td>A.1 (b)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$K_d$</td>
<td>10 - 500 nM</td>
<td></td>
</tr>
<tr>
<td>$K_d$</td>
<td>235 nM</td>
<td>$MW_M$</td>
<td>25 - 45 kD</td>
<td>A.1 (c)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$MW_D$</td>
<td>42 - 62 kD</td>
<td></td>
</tr>
</tbody>
</table>

* Value obtained from initial fit of Equation 6.12 to experimental analytical SEC data (Table 6.1).
Figure A.1. Evaluation of the fit of Equation 6.12 to the experimental data. 3D contour plots showing the variation in $\chi^2$ over the ranges specified, where the minimisation function $\chi^2 = [(\text{experimental} \{x\}) - (\text{fit} \{x\})]^2$ has been used. In each instance, one of the three unknown parameters in Equation 6.12 was fixed to its value in Table 6.1 and the remaining two optimised against the experimental data. (a) App. MW$_D$ defined as 52.2 kD with app. MW$_M$ varied between 25 and 45 kD and $K_d$ between $1 \times 10^{-8}$ to $5 \times 10^{-7}$ M; (b) App. MW$_M$ defined as 36.3 kD with app. MW$_D$ varied between 42 and 62 kD and $K_d$ between $1 \times 10^{-8}$ to $5 \times 10^{-7}$ M; (c) $K_d$ defined as $2.35 \times 10^{-7}$ M with app. MW$_D$ varied between 42 and 62 kD and app. MW$_M$ between 25 and 45 kD.
References


S. Hubbard (1992). NACCESS.


E. Ogris, X. Du, K. C. Nelson, E. K. Mak, X. X. Yu, W. S. Lane, and D. C. Pallas (1999). A protein phosphatase methylesterase (PME-1) is one of several novel


G. Ribas, M. Neville, J. L. Wixon, J. Cheng, and R. D. Campbell (1999). Genes encoding three new members of the leukocyte antigen 6 superfamily and a novel member of Ig superfamily, together with genes encoding the regulatory nuclear chloride ion channel protein (hRNCC) and an N^+ - N^+dimethylarginine dimethylaminohydrolase homologue, are found in a 30-kb segment of the MHC class III region. *J. Immunol.* **163**: 278-287.


D. Yang and L. E. Kay (1999a). Improved \textsuperscript{1}HN-detected triple resonance TROSY-based experiments. \textit{J. Biomol. NMR} \textbf{13}: 3-10.


O. Zhang, L. E. Kay, J. P. Olivier, and J. D. Forman-Kay (1994). Backbone \textsuperscript{1}H and \textsuperscript{15}N resonance assignments of the N-terminal SH3 domain of drk in folded and unfolded states using enhanced sensitivity pulse field gradient NMR techniques. \textit{J. Biomol. NMR} \textbf{4}: 845-858.

