

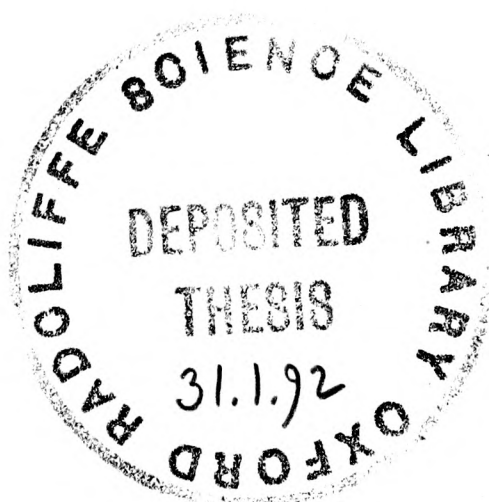
STRUCTURAL STUDIES OF METALLOPROTEINS BY NMR

Stefan Bagby

Hertford College
University of Oxford

A thesis submitted in partial fulfilment of the requirements
for the degree of Doctor of Philosophy

Michaelmas Term 1991



Abstract

Stefan Bagby
Hertford College

Degree of Doctor of Philosophy
Michaelmas Term 1991

Structural Studies of Metalloproteins by NMR

This thesis describes the application of one and two dimensional proton NMR spectroscopy and associated computational methods to the study of structure in solution of two metalloproteins: the cytochrome subunit (cytPCMH) of *Pseudomonas putida para*-cresol methylhydroxylase (PCMH) and parsley plastocyanin.

Initial work on cytPCMH focused on the haem prosthetic group. Haem *meso* proton resonances were assigned by combined analysis of NOE data and the PCMH crystal structure. Interproton distances derived from these assignments indicated that one of the axial ligands to the haem Fe, Met 50, has a different orientation in cytPCMH in solution relative to that in crystalline PCMH. Further evidence for this difference between the unbound and bound states of the cytochrome subunit was provided by differences in reduction potentials, and by differences in the pattern of haem methyl group resonances of the oxidised species. Precedents for different ligand orientations among *c*-type cytochromes have been discussed.

Aromatic and methyl group spin system assignment in cytPCMH was carried out as the initial stage of sequential resonance assignment and structure determination. Aromatic side chain assignments, combined with analysis of NOE data, provided further evidence for a local conformational change in cytPCMH relative to PCMH. A potentially important residue in the conformational change was identified.

Essentially complete sequential resonance assignments have been obtained for parsley plastocyanin, which has important differences in sequence from other higher plant plastocyanins, and is closely related to some algal plastocyanins. Assignments were initially used in location and characterisation of elements of regular secondary structure. Distance and dihedral angle restraints were then accumulated and used in the calculation of a high resolution three dimensional structure of parsley plastocyanin. This structure has been analysed in detail and compared to reported structures of other plastocyanins. Features of interest include a β -bulge, a *cis*-Pro bend and a carboxyl...carboxylate hydrogen bond which accounts for parsley plastocyanin's anomalously high pK_a . Residue deletions cause elimination of a turn found in most higher plant plastocyanins. This turn is located in an acidic patch binding site, which may be further disrupted in parsley plastocyanin by non-conservative substitution of two charged residues. Calculations show that these sequence differences result in diminution of the negative electrostatic field of parsley plastocyanin relative to that of poplar plastocyanin.

Acknowledgements

Thanks to past and present members of the HAOH group, including Paul B., Gordon, Kati, Liang, Marcia, Alex, Jude, Paul H. and Nick for things too numerous to try to remember and to bother typing. A deserving case for honorary membership of the group is George McLendon, an upstanding citizen and world class tennis player. Thanks to Bill McIntire for his gracious hospitality during my stay in San Francisco. Also for his expert and extremely hard work on the PCMH prep. Scott Mathews was kind enough to send sets of coordinates for PCMH.

Initial guidance in the operation of NMR spectrometers and in data processing was provided by Christina Redfield. Jonathan Boyd was always ready to help when experiments weren't going to plan, and he and Nick Soffe maintained the spectrometers — not an easy task with people like myself using them. Many thanks to Christina, Jonathan and Nick.

Special mention goes to Paul Driscoll and Tim Harvey. Paul has made available his considerable knowledge of NMR and resonance assignment over the past two years, particularly in the early stages of the parsley plastocyanin project. Tim has been a great help with the computational aspects of this work, notably in dealing with the vagaries of L^AT_EX and providing programs for all sorts of things.

Finally, thanks to my supervisor, Allen Hill, for providing a very agreeable working environment, for facilitating attendance at several conferences, and for all round support, advice and encouragement.

Abbreviations and Acronyms	
1D	One dimensional
2D	Two dimensional
3D	Three dimensional
4D	Four dimensional
ACES	N-(2-Acetamido)-2-aminoethanesulfonic acid
ADP	Adenosine diphosphate
ATP	Adenosine triphosphate
BPTI	Basic pancreatic trypsin inhibitor from bovine organs
BTP	Bis-tris propane
COSY	Correlated spectroscopy
CYCLOPS	Cyclically ordered phase sequence
DEAE-cellulose	Diethylaminoethyl-cellulose
DQF	Double quantum filter
E.COSY	Exclusive COSY
FAD	Flavin adenine dinucleotide
FPLC	Fast protein liquid chromatography
FID	Free induction decay
FT	Fourier transform
HEPES	(N-[2-Hydroxyethyl]piperazine-N'-[2-ethanesulfonic acid])
HOHAHA	Homonuclear Hartmann-Hahn
MES	2-(N-Morpholino)ethanesulfonic acid
MLEV	Malcolm Levitt
MQF	Multiple quantum filter
NOE	Nuclear Overhauser effect
NOESY	Nuclear Overhauser enhancement and exchange spectroscopy
Pc	Plastocyanin
P.E.COSY	Primitive E.COSY
PS1	Photosystem 1
PTFE	Polytetrafluoroethylene
rmsd	Root mean square deviation
ROESY	Rotating frame nuclear Overhauser effect spectroscopy
SQC	Single quantum coherence
TES	N-tris(Hydroxymethyl)methyl-2-aminoethanesulfonic acid
TPPI	Time proportional phase incrementation
TQF	Triple quantum filter
Tris	Tris[hydroxymethyl]-aminomethane
WALTZ	Wideband alternating phase low power technique for zero residue splitting
ZQC	Zero quantum coherence

Contents

1	Introduction to Plastocyanin and <i>p</i>-Cresol Methylhydroxylase	1
1.1	The Function of Plastocyanin	1
1.2	Primary Structure of Plastocyanin	1
1.3	Three Dimensional Structure of Plastocyanin	5
1.3.1	Crystal Structure of Poplar Cu ^{II} Plastocyanin	5
1.3.2	Crystal Structure of Poplar Cu ^I Plastocyanin	13
1.3.3	Crystal Structure of <i>Enteromorpha prolifera</i> Pc	19
1.4	Interactions of Plastocyanin with Redox Partners	20
1.4.1	Cytochrome <i>f</i>	20
1.4.2	Photosystem I	24
1.4.3	Inorganic Complexes	24
1.5	Mutant Plastocyanins	28
1.6	Previous 2D NMR Studies of Plastocyanin	28
1.7	Other Blue Cu Proteins	28
1.8	<i>p</i> -Cresol Methylhydroxylase (PCMH)	29
1.9	Kinetic and Mechanistic Studies of PCMH	30
1.10	Three Dimensional Structure of PCMH	31
1.10.1	The Cytochrome Subunit	31
1.10.2	The Flavoprotein Subunit	33
1.10.3	Cytochrome-Flavoprotein Interaction	33
1.10.4	Comparisons with Other Proteins	35
2	Experimental Procedures: Isolation and Purification of Proteins	45
2.1	Parsley Plastocyanin	45

2.2	<i>p</i> -Cresol Methylhydroxylase	47
2.2.1	Resolution of PCMH into its Subunits by Isoelectric Focusing	49
2.3	Preparation of Protein Samples for NMR Experiments	50
3	NMR Spectroscopy and its Application to Protein Structure Determination	55
3.1	NMR Spectroscopy and X-Ray Crystallography: Complementary Techniques for Study of Protein Structure	55
3.2	One Dimensional NMR Spectroscopy	57
3.2.1	Quadrature Detection	57
3.3	Product Operator Formalism	58
3.4	2D NMR Spectroscopy	59
3.4.1	Scalar Coupling in the Product Operator Formalism	60
3.4.2	COSY	60
3.4.3	Dipolar Coupling	61
3.4.4	NOESY	62
3.4.5	Coherence Orders and Phase Cycles	63
3.4.6	Phase Sensitive 2D NMR Spectra	65
3.5	Survey of Relevant 2D NMR Techniques	67
3.5.1	COSY and Variants	67
3.5.2	TOCSY or HOHAHA	71
3.5.3	NOESY	74
3.6	Experimental Details of NMR Spectroscopy	75
3.6.1	Parsley Pc Experiments and Data Processing	75
3.6.2	CytPCMH Experiments and Data Processing	77
4	NMR-Derived Data for Protein Structure Determination	85
4.1	Sequential Resonance Assignment	86
4.2	Distance Restraints from the Nuclear Overhauser Effect	86
4.2.1	Initial Slopes of NOE Build Up Curves	86
4.2.2	Zero Quantum Contributions	87
4.2.3	Spin Diffusion	88

4.2.4	Protein Mobility	88
4.2.5	Definition of the Fold Using Approximate Distance Restraints	89
4.2.6	Distribution of NOEs	90
4.2.7	Effect of Approximations on the Accuracy of Derived Structures	90
4.3	Torsion Angle Restraints from Coupling Constants	91
4.3.1	ϕ Backbone Torsion Angle Restraints	91
4.3.2	Side Chain Torsion Angle Restraints and Stereospecific Assignment of Methylene and Methyl Resonances	91
4.4	Identification of Regular Secondary Structure Elements	93
4.4.1	NOEs and Secondary Structure	93
4.4.2	Coupling Constants and Secondary Structure	94
4.4.3	Hydrogen Exchange and Secondary Structure	95
4.4.4	Effect of Secondary Structure Identification on Structure Calculation	95
4.5	Identification of Tight Turns	95
4.6	Methods of Structure Calculation	95
4.6.1	Metric Matrix Distance Geometry	96
4.6.2	Restrained Molecular Dynamics	97
4.6.3	Dynamical Simulated Annealing	98

5 Haem Geometry and Resonance Assignment in the Cytochrome Subunit of PCMH 103

5.1	Assignment of Haem <i>Meso</i> Protons	105
5.2	Proposed Rearrangement of Met Ligand	108
5.3	Polypeptide Resonance Assignments	113
5.3.1	Delineation of Aromatic Spin Systems	113
5.3.2	Assignment of $C^\alpha H-C^\beta H_2$ of Aromatic residues	119
5.3.3	Methyl Group Region: 1. Val, Leu and Ile	120
5.3.4	Methyl Group Region: 2. Ala and Thr	124
5.3.5	Glycines	124
5.4	Overview	127

6	Sequential Resonance Assignment and Analysis of Secondary Structure of Parsley Plastocyanin	131
6.1	Residue Numbering	134
6.2	Spin System Identification	134
6.2.1	Methyl Group Region: 1. Alanines and Threonines	134
6.2.2	Methyl Group Region: 2. Isoleucines, Leucines and Valines . .	134
6.2.3	Identification of Phe, Tyr and His Spin Systems	137
6.2.4	Identification of Proline Spin Systems	137
6.2.5	Fingerprint Region	140
6.3	Further Identification of Spin Systems and Attachment of Backbone Amide Protons	140
6.3.1	Glycines	140
6.3.2	Methyl Group-Containing Residues	143
6.3.3	Non-unique Spin Systems	143
6.4	Second Stage Assignment	146
6.4.1	Fragments with Extended Conformation	149
6.4.2	Helical Segment	154
6.4.3	Alignment and Packing of β -Strands	159
6.4.4	Turns and Loops	161
7	High Resolution Solution Structure of Parsley Plastocyanin	167
7.1	Experimental Restraints for Parsley Plastocyanin Structure Calculation	167
7.1.1	Approximate Interproton Distance Restraints	167
7.1.2	Dihedral Angle Restraints and Stereospecific Assignment of β -Methylene Protons from Qualitative Data Analysis	168
7.1.3	Stereospecific Assignments and Dihedral Angle Restraints from STEREOSEARCH	169
7.1.4	Stereospecific Assignment of Methyl Groups of Val and Leu .	171
7.1.5	Hydrogen Bond Restraints	171
7.1.6	Peptide Bond Restraints	171
7.1.7	Cu Restraints	172

7.2	Structure Calculations	172
7.2.1	Overall Strategy	172
7.2.2	The Hybrid Method	173
7.2.3	Folding an Extended Strand by Dynamical Simulated Annealing	175
7.2.4	The Quick Method	177
7.3	The Converged Structures	177
7.3.1	Aromatic Ring Alignment	180
7.3.2	Fitting and Calculation of Average Structure	183
7.3.3	Accuracy and Precision of the Structures	183
7.4	Description of the Molecule	190
7.4.1	Overall Topology	190
7.4.2	β -Strands and β -Sheets	191
7.4.3	Hydrogen Bonding Network	191
7.4.4	The Cu Site	194
7.4.5	Loops and Tight Turns	197
7.4.6	β -Bulge	202
7.4.7	Helical Fragment	204
7.4.8	Hydrogen Bonding Involving Residues 41, 52 and 56	205
7.4.9	Aromatic Residues	208
7.4.10	Acidic Patch	210
7.4.11	Acidic Patch Hydrogen Bond	215
7.5	Electrostatic Field Calculations using DelPhi	215
7.5.1	Description of the Program and Parameters Used	217
7.5.2	Preliminary Results	218
7.6	Overview	219

Corrections

1. Page 11, line 5: ‘...NH of Asp 61 ...’ should read ‘...CO of Asp 61 ...’
2. Page 20, line 18: ‘...cytcochrome c_1 ...’ should read ‘...cytochrome c_1 ...’
3. Page 29, legend Table 1.4: ‘...contains of 40% carbohydrate.’ should read ‘...contains 40% carbohydrate.’
4. Page 55, line 6 and page 59, line 6: reference 33 should be reference 2.
5. Page 60: equation 3.7 should read

$$2I_{1x}I_{2z} \xrightarrow{\pi J_{12}t} 2I_{1x}I_{2z} \cos \pi J_{12}t + I_{1y} \sin \pi J_{12}t$$
6. Page 60: equation 3.8 should read

$$2I_{1y}I_{2z} \xrightarrow{\pi J_{12}t} 2I_{1y}I_{2z} \cos \pi J_{12}t - I_{1x} \sin \pi J_{12}t$$
7. Page 60: equation 3.10 should read

$$I_{1z} \xrightarrow{90_y^\circ} I_{1x} \xrightarrow{t_1} -2I_{1x}I_{2z} \xrightarrow{90_y^\circ} 2I_{1z}I_{2x} \xrightarrow{t_2} I_{2x}$$
8. Page 98, line 12: references 10,14,16 should be 29,30,31.
9. Page 114, line 17: ‘...protons 4 with 7 and 5 with 7.’ should read ‘...protons 4 with 7, 4 with 6 and 5 with 7.’
10. Page 120, legend to Table 5.2: ‘..., 35°).’ should read ‘..., 35°C).’
11. Page 134, line 5: parentheses has been misspelt as parantheses.
12. Page 146, line 2: ‘...spin sytems ...’ should read ‘...spin systems ...’
13. Page 172, line 26: ‘...the Hybrid method.’ should read ‘...a modification of the Hybrid method, excluding the distance geometry stage.’
14. Page 174, line 21: k_{tor} should have units of $\text{kcal mol}^{-1} \text{rad}^{-2}$.
15. Page 184, Table 7.3: number of Cu restraints should be 14 instead of 16.
16. Page 192, legend Table 7.4: ‘...(E. *prolif.* ...’ should read ‘...(E. *prolif.*) ...’
17. Page 195, line 8: ‘...less than 93° ...’ should read ‘...greater than 93° ...’
18. Page 196, Table 7.6: in the final column, θ_2 should read θ_1 .
19. Page 197, line 9: ‘...and 73(71)–74(74).’ should read ‘...and 73(71)–76(74).’
20. Page 208, line 13 and page 215, subsection 7.4.12: Asp 53 should read Glu 53.

Chapter 1

Introduction to Plastocyanin and p-Cresol Methylhydroxylase

1.1 The Function of Plastocyanin

Plastocyanin [1, 2, 3, 4] (Pc) is a single polypeptide protein of between ninety-seven and one hundred and five residues, with a blue or type 1 Cu active site. It occurs in all higher plants, many green algae, and some blue-green algae (cyanobacteria). Plastocyanin is part of the photosynthetic electron transport chain of these organisms, transferring electrons from membrane-bound [5] cytochrome *f* to P700⁺, the double chlorophyll pigment of photosystem I (PSI) (Figure 1.1). The release of plastocyanin after mechanical damage of the membrane, without detergent, is consistent with it not being membrane-bound.

Photosynthesis occurs at highly convoluted thylakoid membranes inside the chloroplast. Vectorial electron transport from the interior to the exterior of the membrane is linked with the generation of protons (by H₂O oxidation) inside the thylakoid and with plastoquinol transport of protons from the exterior. According to Mitchell's chemiosmotic theory [6], this proton-translocating system and ATP synthetase are linked to form a phosphorylating proton circuit.

1.2 Primary Structure of Plastocyanin

Plastocyanins from eighteen higher plants [7], four green algae (*Chlorella fusca* [8], *Enteromorpha prolifera* [9], *Scenedesmus obliquus* [10], *Ulva arasakii* [11]) and one

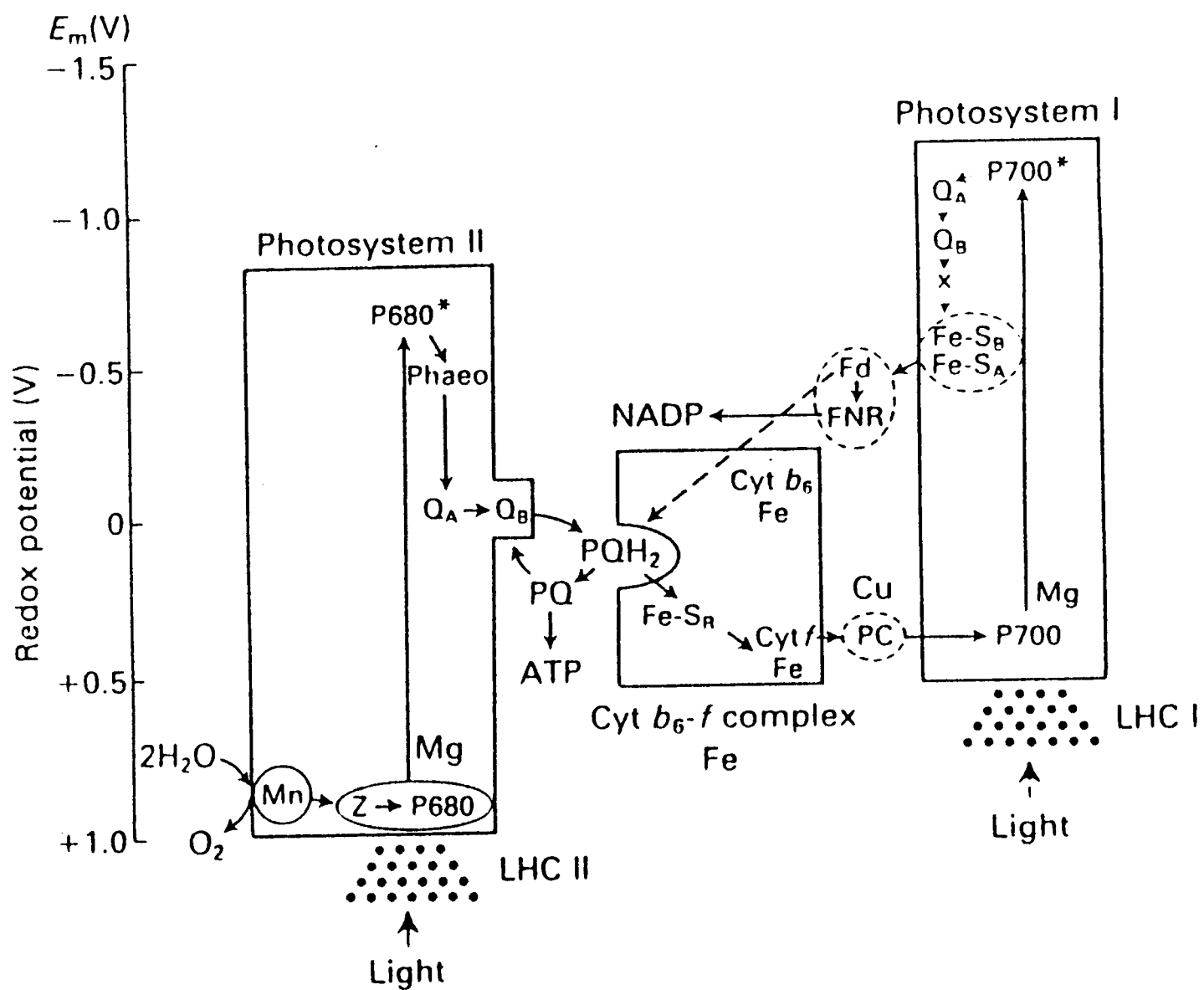


Figure 1.1: Pathway of electron flow from H_2O to $NADP^+$ in photosynthesis. Abbreviations: Z, tyrosine radical intermediate between Mn centre and P680; P and P*, ground and excited states of reaction centre chlorophyll; Phaeo, pheophytin; Q_A and Q_B , plastoquinone-binding proteins; PQ, plastoquinone; PC, plastocyanin (high potential one electron shuttle); Fd, ferredoxin (low potential one electron shuttle); FNR, ferredoxin $NADP^+$ reductase; Fe-S_A, Fe-S_B and Fe-S_R, iron-sulphur proteins; LHC, light harvesting complex.

blue-green alga (*Anabaena variabilis* [12]) have been completely sequenced. Seven of these sequences are shown in Table 1.1, including plastocyanins for which there are extensive crystallographic and NMR data. In Table 1.1, ‡ denotes a residue which is invariant in all known sequences, § denotes a residue which is invariant in higher plant plastocyanins. Among the complete higher plant sequences, forty-five residues are invariant if parsley plastocyanin is included, fifty-two if it is not. Invariant or highly conserved residues are particularly common to the 31–45 and 82–94 portions, including the Cu ligands, Ala 33, Phe 35, Pro 36, Ile/Val 39, Pro 86 and Ala 90 located in the north hydrophobic pocket, and residues 42–44 of the acidic patch (see later for explanation of the terms hydrophobic pocket and acidic patch). Considerable variability occurs in functionally less significant regions. The homology is reduced if the plastocyanins from green algae are included (twenty-four invariant residues). The high degree of conservation has two important implications: that plastocyanins from different sources have similar structures, and that those residues which are conserved are important for function or structure.

A striking characteristic of all higher plant and green algal plastocyanins is the high overall negative charge of the molecule, estimated to be between -8 and -10 for Cu^IPc at pH 7.0. The plastocyanin from *A. variabilis* has an estimated charge of $+2$ and a $\text{pI} > 7$, and is unique among the known plastocyanins. Common to virtually all other plastocyanins is a surface concentration of negative charge due to residues 42–45 and 59–61. The only higher plant plastocyanin for which the Asp-Glu-Asp-Glu sequence at 42–45 is broken is poplar plastocyanin (45 is serine), while for parsley plastocyanin negative charge is not conserved at 59–61. Parsley plastocyanin is also unusual in having deletions at positions 57 and 58, a property it shares with *C. fusca* and *S. obliquus* plastocyanins. *E. prolifera* plastocyanin is slightly different, having *-Glu-* at positions 58–60 instead of the Ser-Glu-Glu of poplar plastocyanin (alignment based on superposition of structures [13]). The homologies of parsley and *E. prolifera* plastocyanins to poplar plastocyanin are 64% and 56%, with sixty-three and fifty-six homologous residues respectively. It should be noted here that the plastocyanins from barley [14], rice [7], and carrot (Shoji et al., cited in [7]), have been shown to have similar ‘strangeness’ [15] to parsley plastocyanin. For example, barley plastocyanin

	0	5	10	15	20	25
Poplar	Ile	Leu	Gly	Phe	Gln	Gly
Spinach	Val	Leu	Asp	Phe	Ser	Gly
P'r. bean	Leu	Leu	Gly	Phe	Ser	Gly
Parsley	Ala	Leu	Asp	Phe	Ser	Gly
S. obliquus	Ala	Leu	Gly	Phe	Ala	Gly
P. prolifera	Ala	Leu	Asp	Phe	Ala	Asp
A. variabilis	Tyr	Leu	Gly	Phe	Ala	Gly
Conservation		§	Gly	†	lys	†
Poplar	lys	Phe	Gly	Ile	Ser	Gly
Spinach	Gln	Phe	Gly	Val	Gln	Gly
P'r. bean	lys	Phe	Gly	Val	Gln	Gly
Parsley	lys	Phe	Gly	Val	Gln	Gly
S. obliquus	Ser	Trp	Gly	Ile	Ala	Gly
P. prolifera	Ser	Phe	Gly	Ile	Ala	Gly
A. variabilis	Tyr	Phe	Val	Val	Ala	lys
Conservation		§	Val	†	†	§
Poplar	Asp	lys	Gln	Asn	Phe	Leu
Spinach	Asp	lys	Gln	Asn	Trp	Leu
P'r. bean	Asp	lys	Gln	Asn	Trp	Leu
Parsley	Asn	lys	Gln	Asn	Trp	Leu
S. obliquus	Asn	Ala	His	Gln	Trp	Phe
P. prolifera	Asp	Ala	His	Gln	Val	Trp
A. variabilis	Ala	lys	His	Met	Ser	Phe
Conservation		§	His	†	§	§
Poplar	Asn	Gln	Cys	Gln	lys	Asn
Spinach	Gln	Trp	Cys	Gln	lys	Asn
P'r. bean	Trp	Trp	Cys	Gln	lys	Asn
Parsley	Gln	Trp	Cys	Gln	Gln	Asn
S. obliquus	Trp	Gln	Cys	Gln	lys	Gln
P. prolifera	Trp	Trp	Cys	Gln	Trp	Gln
A. variabilis	Ala	Gln	Cys	Gln	lys	Ala
Conservation		†	†	†	†	†

has deletions at positions 57 and 58, a tyrosine at position 62 and a glutamate at position 85.

Interestingly, different amino acids have been found at the same sequence position in some plastocyanins [16]. For example, two different plastocyanins, designated PCa and PCb, have been isolated and characterised in poplar, *Populus nigra*, var. *italica* [17]. The PCb and PCa sequences differ at twelve positions. Heterogeneity also occurs in the parsley plastocyanin sequence, Glu 53 in PCa being replaced by Asp in PCb [15]. The physiological significance of the plastocyanin polymorphism has not yet been elucidated.

1.3 Three Dimensional Structure of Plastocyanin

1.3.1 Crystal Structure of Poplar Cu^{II} Plastocyanin

The Cu atoms found in most Cu proteins have spectroscopic and redox properties which are unusual compared to those of Cu^I and Cu^{II} in low M_r complexes. Cu electron transfer proteins contain only one Cu centre, belonging to the type 1 class. There also exist type 2 and type 3 centres. Characteristic properties of type 1 Cu centres include:

1. An absorption maximum at $\approx 600\text{nm}$ with ϵ_{max} of the order of $5 \times 10^3 \text{ M}^{-1}\text{cm}^{-1}$, approximately two orders of magnitude greater than the ϵ_{max} of a typical low M_r Cu^{II} complex.
2. A hyperfine splitting constant, A_{||}, of $0.003\text{--}0.008\text{cm}^{-1}$, compared with $0.012\text{--}0.20\text{cm}^{-1}$ for normal Cu^{II}.
3. A redox potential of $0.3\text{--}0.8\text{V}$, compared with 0.17V for the aqueous Cu^I/Cu^{II} couple.

It was to account for these unusual properties that Freeman and colleagues at the University of Sydney undertook the determination of the three dimensional structure of plastocyanin. After trying unsuccessfully to crystallise plastocyanins from many sources, the protein from poplar (*Populus nigra*, var. *italica*) leaves was found to give

suitable crystals [18]. The structure of Cu^{II}Pc was determined at a resolution of $2.7\text{\AA}$ [19], and subsequently refined using data from crystals at pH 6.0 and pH 4.2 at resolutions of $1.6\text{\AA}$ and $1.9\text{\AA}$ [20].

Plastocyanin is a β -sandwich with a slightly flattened cylindrical shape, consisting of eight strands of polypeptide connected by seven loops at the ends of the cylinder. Schematic representations of the plastocyanin structure are given in Figures 1.2 and 1.3. The topology shown in the former, which is adapted from a figure in reference [20], is particular to poplar plastocyanin. Figure 1.3 illustrates the general folding pattern of all small blue Cu proteins. Seven of the strands consist largely of β secondary structure, the exception being strand 5 which contains the only helical fragment. One face of the sandwich is formed by strands 2 (residues 10–14), 1 (residues 1–7), 3 (residues 25–32) and 6 (residues 67–74), and the other by strands 2 (residues 17–21), 8 (residues 92–98), 7 (residues 78–85) and 4 (residues 36–47). Hydrogen bonding produces stretches of β -sheet between strands 1 and 3 (parallel), 3 and 6 (anti-parallel), 7 and 8 (anti-parallel) and 8 and 2 (parallel). Strand 5 is not part of either face, and interrupts the continuous network of hydrogen bonds that would be required for formation of a β -barrel. Strand 5 contains about 1.5 turns of helix (51–56) followed by a β -turn (58–61).

Poplar plastocyanin contains nine β -turns. Four of them (7–10, 65–68, 84–87 and 88–91), plus a loop at 32–36, constitute the northern end of the molecule. (North and south are defined by the position of Cu with respect to the long axis of the protein, Cu being located to the north). The southern end consists of two β -turns (22–25 and 47–50) and a loop at 75–79. Three mid-strand turns occur at 14–17, 42–45 and 58–61. The first of these, in strand 2, is a *cis*-Pro bend [21]. A similar turn is found in strand 2 of azurin [22]. One consequence of this turn is that the $\text{C}^\alpha\text{-C}^\beta$ bond of invariant Phe 14 is directed towards the protein interior. This residue is in contact with Met 92, a Cu ligand. It seems that this phenylalanine side chain is an essential component of the Cu site. A second *cis*-peptide group occurs at Pro 36, adjacent to the Cu ligand His 37.

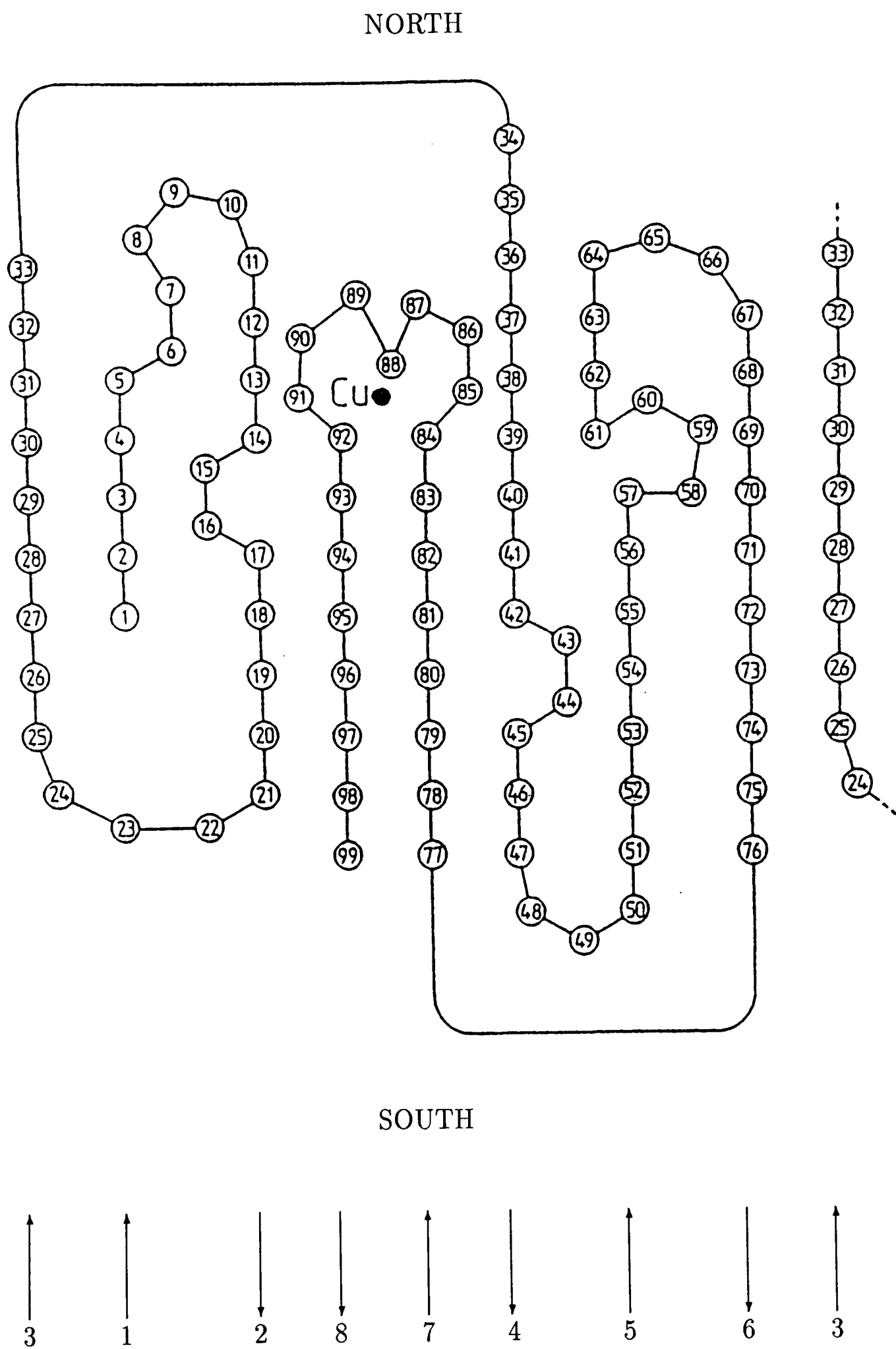
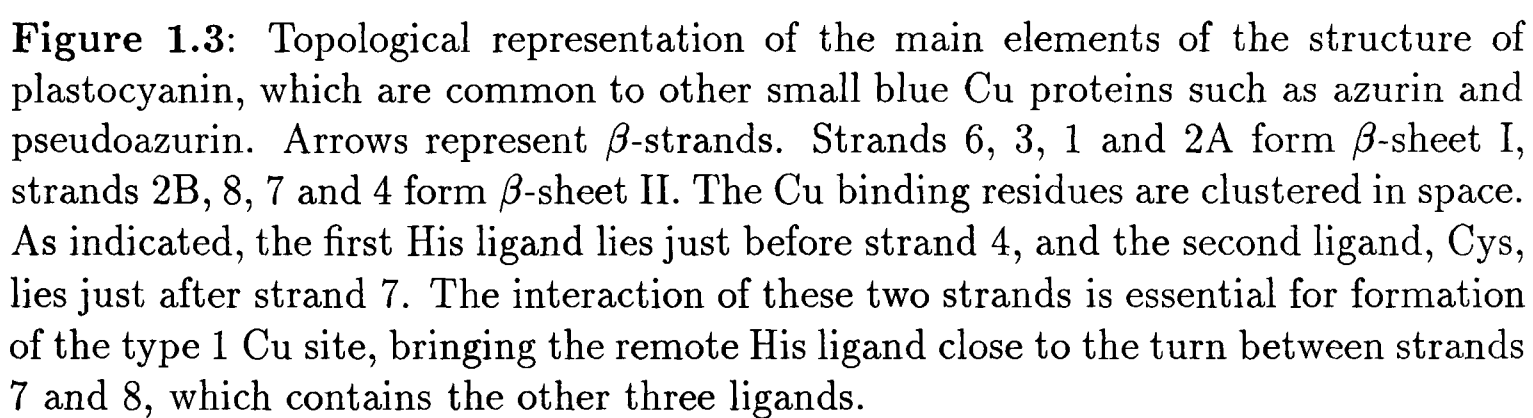


Figure 1.2: Schematic representation of the polypeptide fold of poplar plastocyanin.



The Cu Site

The Cu site lies at the northern end of the molecule in a depression or pocket between three loops. Cu is coordinated by His 37, Cys 84, His 87 and Met 92, the two N and two S donor atoms forming a distorted tetrahedron (elongated C_{3v} effective symmetry with rhombic distortions; Figure 1.4 and Table 1.2). Two S donor atoms is ideal for ‘soft’ Cu^I and acceptable for ‘intermediate’ Cu^{II} . The reverse applies for the two rather basic imidazole N atoms.

The Cu atom lies 0.72Å from the plane of N(His 37), S(Cys 84) and S(Met 92). When compared to low M_r complexes with similar ligands but more conventional coordination geometries, the Cu-N bond lengths are normal, the Cu-S(Cys) bond is short and the Cu-S(Met) bond is exceptionally long. Indeed, the Cu-S(Met) bond was not detected by EXAFs measurements in a study by Scott et al. [24], but has recently been clearly identified by EXAFS at pH 8.0 in spinach plastocyanin [25]. There is no evidence for H_2O molecules at the Cu site or anywhere else inside the plastocyanin molecule [20].

The intense blue band near 600nm is a cysteine S $p\pi \rightarrow Cu d_{x^2-y^2}$ transition which probably derives its intensity from very good overlap between ground and excited state wavefunctions [23]. The thiolate- Cu^{II} bond makes a dominant contribution to the electronic structure of the blue Cu site which can be strongly influenced by orientation of Cys 84 by the protein backbone [23].

There has been some consideration of whether the Cu-S(Met) interaction is sufficiently strong to be described as a bond. The S(Met 92) atom makes a small but definite contribution to the ligand field at the Cu centre, but the analysis of Penfield et al. [26] reveals no S(Met 92)→Cu charge transfer component in the electronic spectrum of $Cu^{II}Pc$. Comparison with the Cu-O distances in square-pyramidal or tetragonally distorted octahedral Cu^{II} complexes provides evidence to support the view that the Cu-S(Met) interaction should be described as a bond. A similar rationalisation has been adopted for very long axial Cu-S bonds in some octahedral Cu^{II} complexes [27, 28]. Further, theoretical analysis [23] indicates that there is a reasonable interaction between Cu and the thioether S at 2.9Å.

Several highly conserved residues are involved in the formation of hydrogen bonds

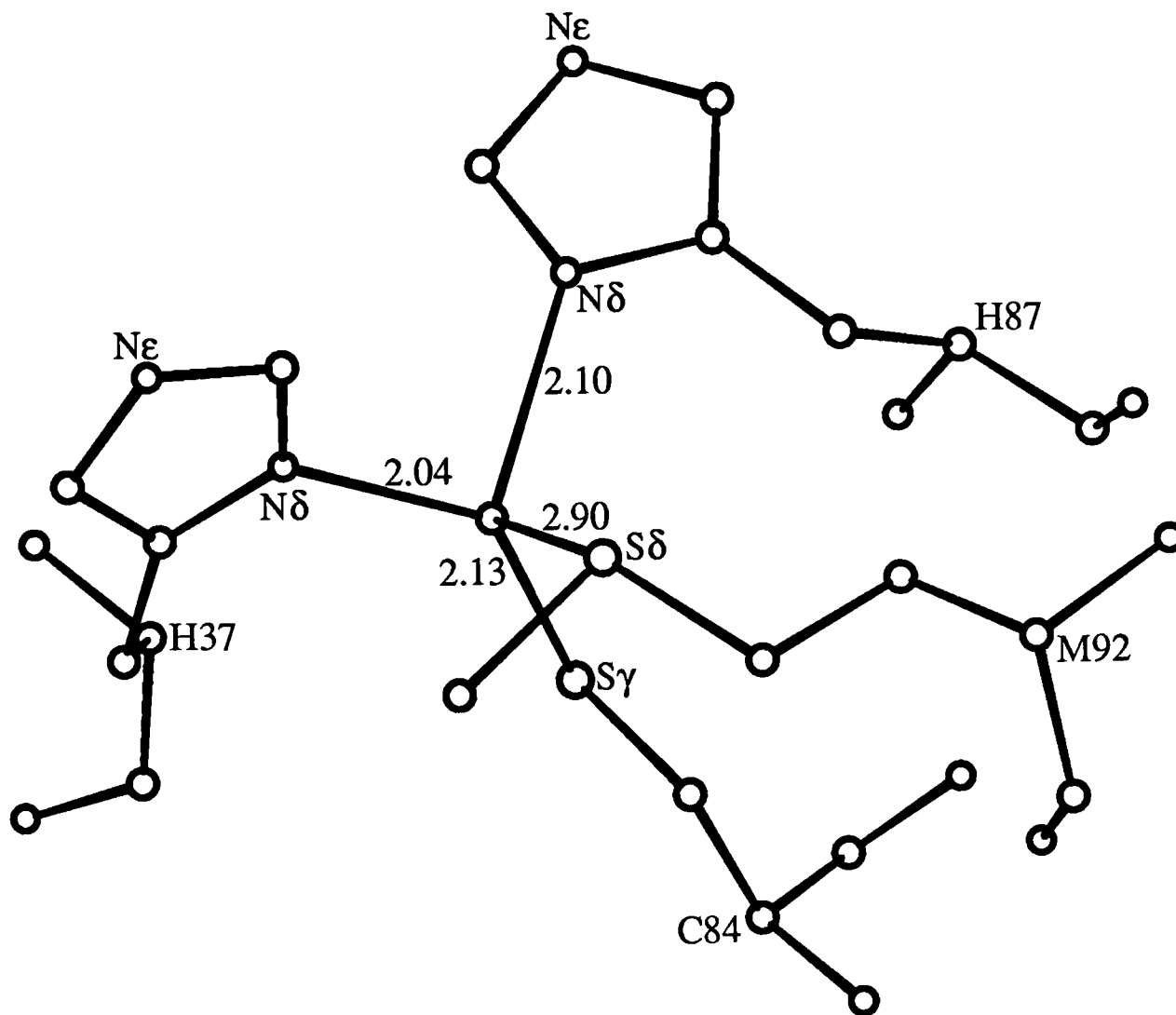


Figure 1.4: The Cu site in poplar $\text{Cu}^{\text{II}}\text{Pc}$ at pH 6.0. The non-hydrogen atoms of each ligand residue are shown.

	Poplar	Poplar	<i>E. prolifer</i>
Cu-ligand bond lengths (Å)	pH 4.2	pH 6.0	pH 5.8
Cu-N(His 37)	2.07	2.04	1.89
Cu-N(His 87)	2.17	2.10	2.17
Cu-S(Cys 84)	2.11	2.13	2.12
Cu-S(Met 92)	2.87	2.90	2.92
Bond angles at Cu atom (deg)			
N(His 37)-Cu-N(His 87)	96	97	104
N(His 37)-Cu-S(Cys 84)	133	132	125
N(His 37)-Cu-S(Met 92)	87	85	90
N(His 87)-Cu-S(Cys 84)	122	123	120
N(His 87)-Cu-S(Met 92)	103	103	102
S(Cys 84)-Cu-S(Met 92)	109	108	108

Table 1.2: Dimensions of the Cu site in poplar $\text{Cu}^{\text{II}}\text{Pc}$ and in *E. prolifer* $\text{Cu}^{\text{II}}\text{Pc}$.

that stabilise the conformation at and around the Cu site. The orientation of the ring of His 37 is stabilised by a $\text{NH}\cdots\text{O}$ bond from the ring to the backbone amide group of residue 33. Also noteworthy is the role of invariant Asn 38. Its backbone NH is hydrogen bonded to the thiolate sulphur of Cys 84, and its side chain interacts with the backbone NH and side chain of Ser 85 and with the backbone NH of Asp 61. Asn 38 possibly regulates the separation between His 37 and Cys 84 or the position of strand 4 relative to strands 7 and 5. An analogous invariant residue (Asn 47) occurs in another type 1 Cu protein, *Pseudomonas aeruginosa* azurin. This is one of a number of shared features brought out in comparisons of different blue copper proteins [29, 30, 31] (see also section on blue copper proteins).

Hydrophobic Pocket and Hydrophobic Patch

The pocket in which the Cu site lies is lined with hydrophobic side chains: Leu 12, Ala 33, Pro 36, Leu 62, Pro 86 and Ala 90 at the rim; Leu 5, Pro 36, Leu 63 and Pro 86 along parts of the wall; Phe 14 (in contact with Met 92) and Val 39 at the base. Only Ala 33 is variable, the rest being highly conserved or invariant. Those residues at the rim have solvent exposed side chains and form a hydrophobic patch surrounding the exposed edge of the imidazole ring of His 87. The latter is a logical candidate as a potential electron transfer pathway, since it is solvent exposed on one side and coordinated to Cu on the other, and has the hydrophobic patch as a nearby recognition site (Figure 1.5). Supporting evidence has been provided by using negatively charged complexes as relaxation probes in NMR experiments [32, 33].

Three glycine residues located at the northern hydrophobic surface are either invariant (Gly 10 and Gly 89), or highly conserved (Gly 34). This is possibly because the distance (6Å) from the molecular boundary to the Cu is thereby minimised, an important consideration if the pseudo-aromatic imidazole ring of His 87 is an electron transfer pathway.

The Acidic Patch

The overall charge of the oxidised poplar plastocyanin molecule, simply counting charges, is -8 . This figure is derived from a ratio of fifteen:six negatively to positively

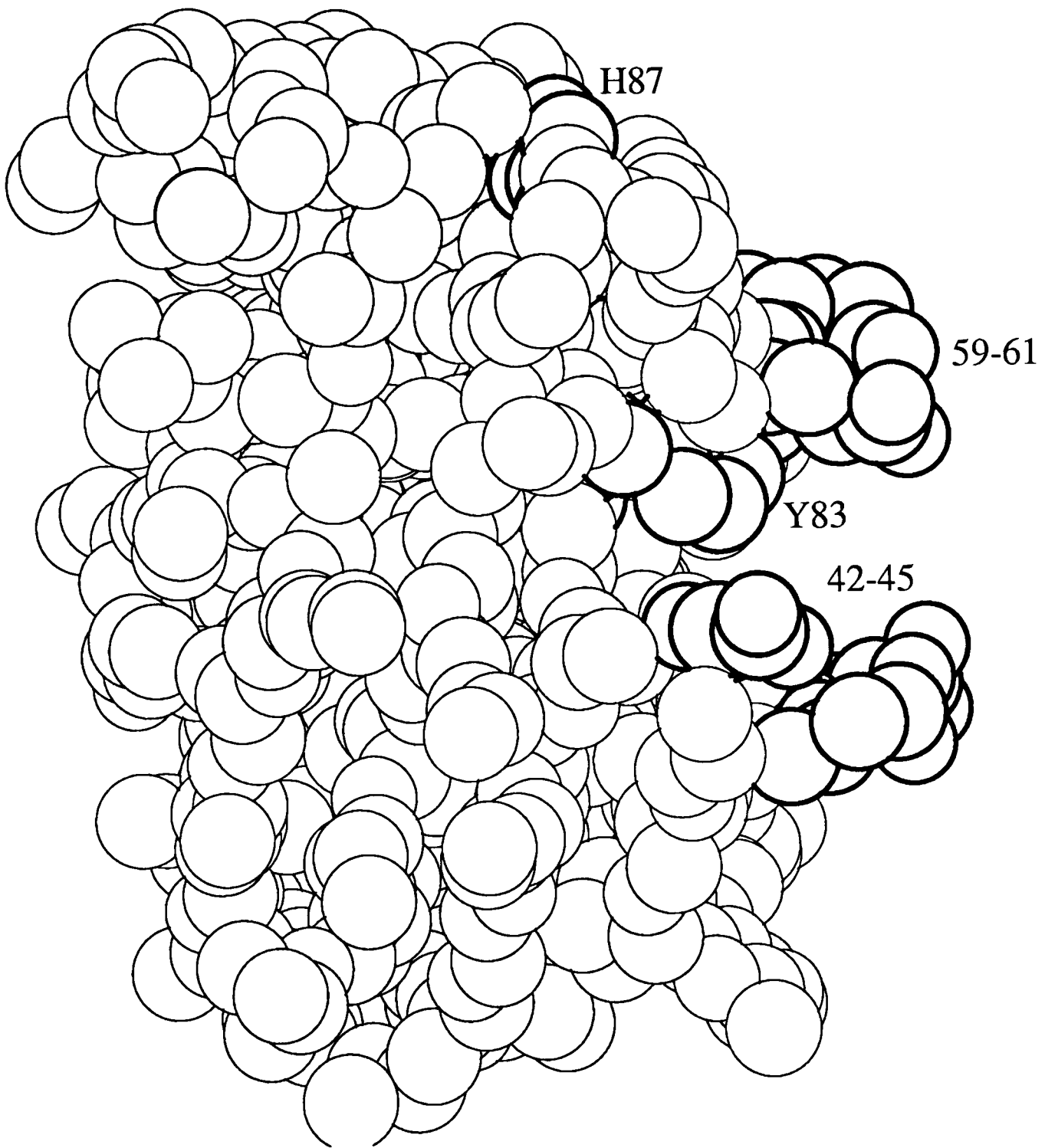


Figure 1.5: Regions of probable functional importance on the surface of poplar plastocyanin. His 87, Tyr 83 and acidic patch residues in bold.

charged residues, a charge of +2 on Cu, and one of -1 on the thiolate of Cys 84. There is a striking imbalance in the distribution of the ten charged residues which are conserved in plant plastocyanins. None is found in the northern quarter of the molecule, and several are concentrated on the two mid-strand β turns at positions 42–45 and 58–61. The conformations of these turns are such that the acidic side chains protrude from the protein surface to form an elongated patch of negative charge, surrounding the solvent exposed side chain of Tyr 83 (Figures 1.5 and 1.6). This patch has been proposed [19] as a second recognition site for physiological partners, with Tyr 83 having a possible role in electron transfer.

Aromatic Residues

Of the nine aromatic residues, only Phe 35 lies north of the Cu site (Figure 1.7). The sequence Phe 14...Phe 82...Phe 19...Phe 41...Tyr 80 extends from just below the side chain of Cu ligand Met 92 to the southern end of the molecule. Phe 14, Phe 41 and Tyr 80 are invariant in all plastocyanins, Phe 82 is replaced by Tyr in *S. obliquus* plastocyanin, and Phe 19 is conservatively substituted by other hydrophobic residues.

1.3.2 Crystal Structure of Poplar Cu^I Plastocyanin

There are no functionally important changes in the structure of Cu^{II}Pc on changing the pH from 6.0 to 4.2. This indicates that the pH-dependence of the redox properties of plastocyanin [34, 35] is associated with structural changes in Cu^IPc . The crystal structure of Cu^IPc has been determined at pH values of 3.8, 4.4, 5.1, 5.9, 7.0 and 7.8 [36]. At the highest of these pH values, the Cu coordination geometry is essentially the same as that in Cu^{II}Pc . At pH 3.8, however, the geometry is trigonal planar, the Cu-S(Met 92) and Cu-N $^{\delta}$ (His 87) bond lengths have changed from 2.9Å to 2.5Å and from 2.3Å to 3.2Å, and Cu has moved into the plane of N(His 37), S(Cys 84) and S(Met 92). The changes appear to occur as a smooth function of pH, but the apparent structure at each pH is most probably a weighted mean of the two structures at either end of the pH range.

The low pH form of reduced plastocyanin, which will be termed HCu^IPc , has a conformer in which the imidazole ring of His 87 has rotated by 180° about the

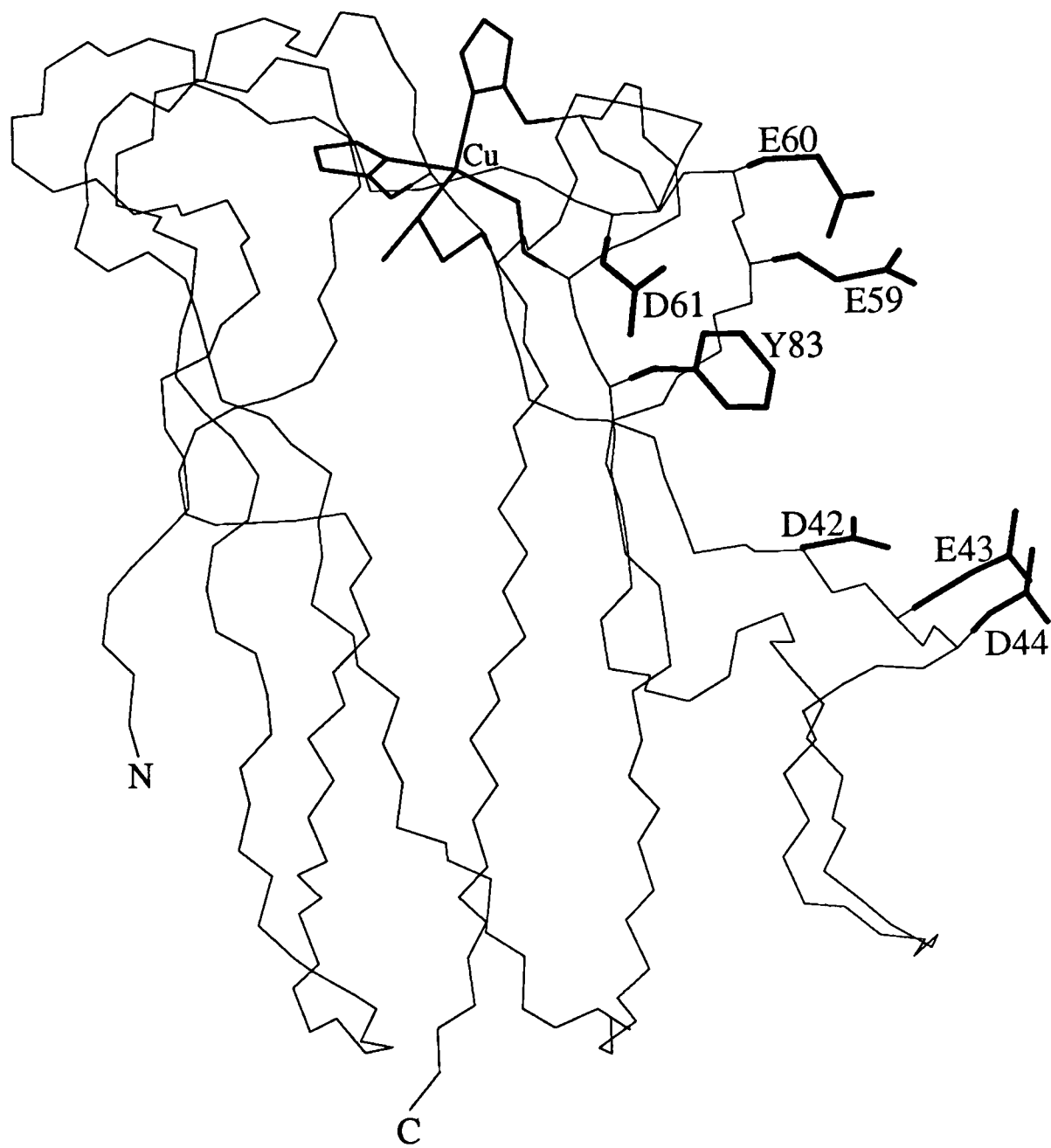


Figure 1.6: Poplar plastocyanin with acidic patch residues 42–44, 59–61 and 83 in bold. The Cu site is also indicated. Tyr 83 has been implicated as an important residue in electron transmission to Cu.

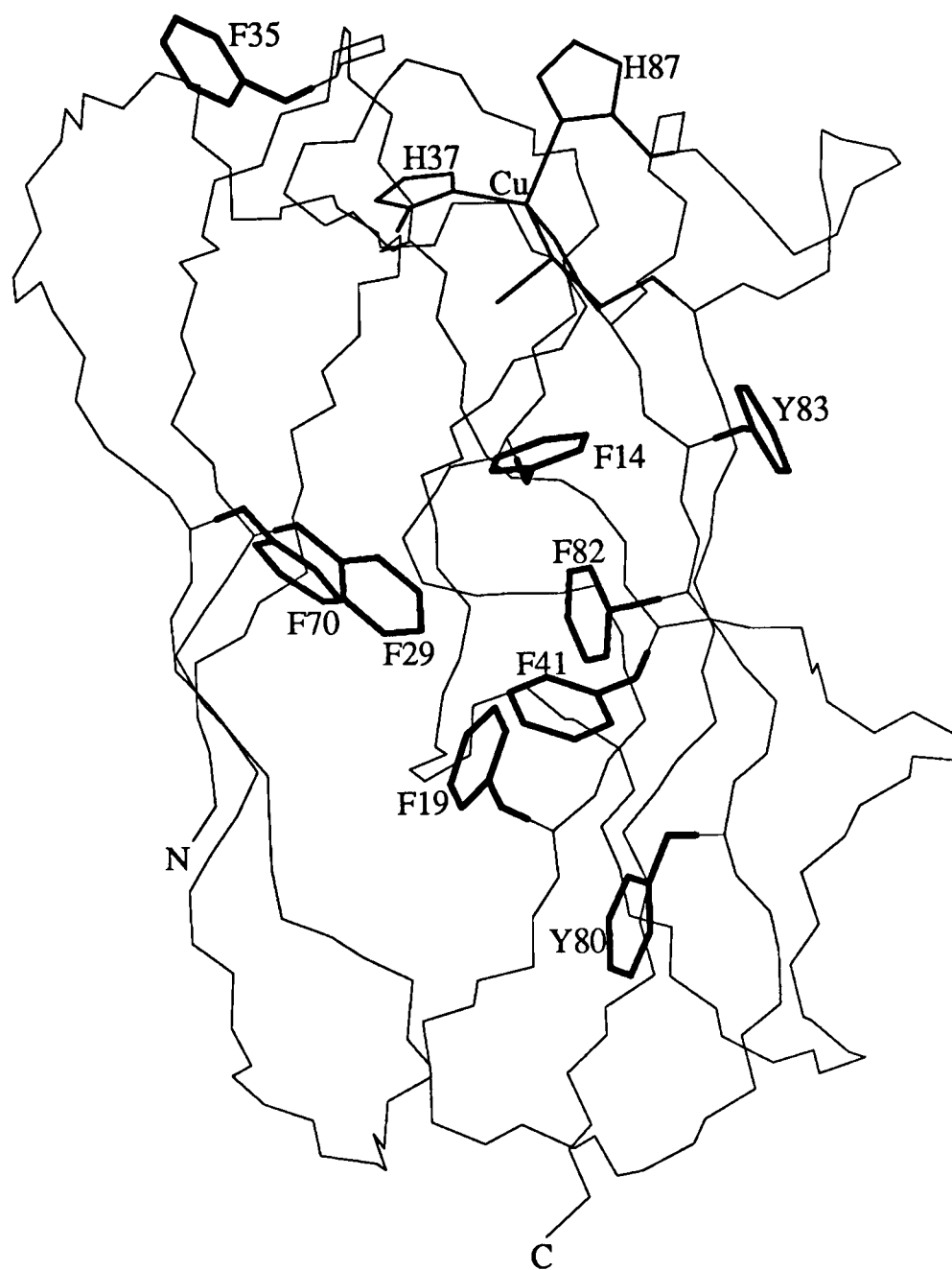


Figure 1.7: The aromatic residues in the poplar plastocyanin molecule.

C^β - C^γ bond (Figure 1.8). There is now a hydrogen bond between N^δ (His 87) and a water molecule, and the atom in closest contact with Cu is C^δ (His 87) instead of N^δ (His 87). Thus, at low pH, the Cu- N^δ (His 87) bond is severed and is replaced by a Cu $\cdots$ H-C(His 87) van der Waals contact. An additional hydrogen bond between $N^{\epsilon 2}$ (His 87)-H and O(Pro 36) confirms that the orientation of the His 87 imidazole ring has changed. It is not possible to tell whether the crystal contains a mixture of two conformers or consists exclusively of this second conformer.

The close similarity of the Cu site geometries in $Cu^{II}Pc$ and Cu^IPc is consistent with a low reorganisation energy (Franck-Condon barrier) and hence optimisation of the electron transfer process [37]. Distortion from the square-planar geometry preferred by Cu^{II} towards the tetrahedral arrangement preferred by Cu^I increases the tendency of $Cu^{II}Pc$ to accept an electron, i. e. increases the redox potential. On the other hand, the irregular geometry is not ideal for either Cu^{II} or Cu^I , so the protein is poised for change in both oxidation states. This is an indication of how the Cu site is elegantly designed to suit plastocyanin's role as a link between two high potential components of the photosynthetic chain.

There are no precedents from model compounds to explain the retention of the long Cu-S(Met 92) bond in Cu^IPc . However, the change in this interaction from weak covalent bond (2.9Å) in Cu^IPc to moderately strong covalent bond (2.5Å) in HCu^IPc indicates that its importance lies with the latter. The Cu-S(Met 92) bond may be present in order to prevent loss of Cu from HCu^IPc , since dissociation of His 87 would otherwise leave the metal strongly coordinated by only two ligands. This should be an important consideration since the physiological pH is 4–6, so that a large proportion of reduced plastocyanin will be present as HCu^IPc . However, this reasoning does not explain the occurrence of highly conserved residues corresponding to Met 92 in other blue Cu proteins. Protonation and dissociation of one of the His ligands have not been shown to occur in any other blue Cu protein [38]. An alternative explanation is that the fourth Cu ligand helps to tune the reduction potential to an appropriate value. This is supported, for example, by the fact that stellacyanin, which lacks methionine, has a reduction potential (184mV) which is much lower than that of any other blue Cu protein (300–780mV). Further support for this tuning effect is provided by the results

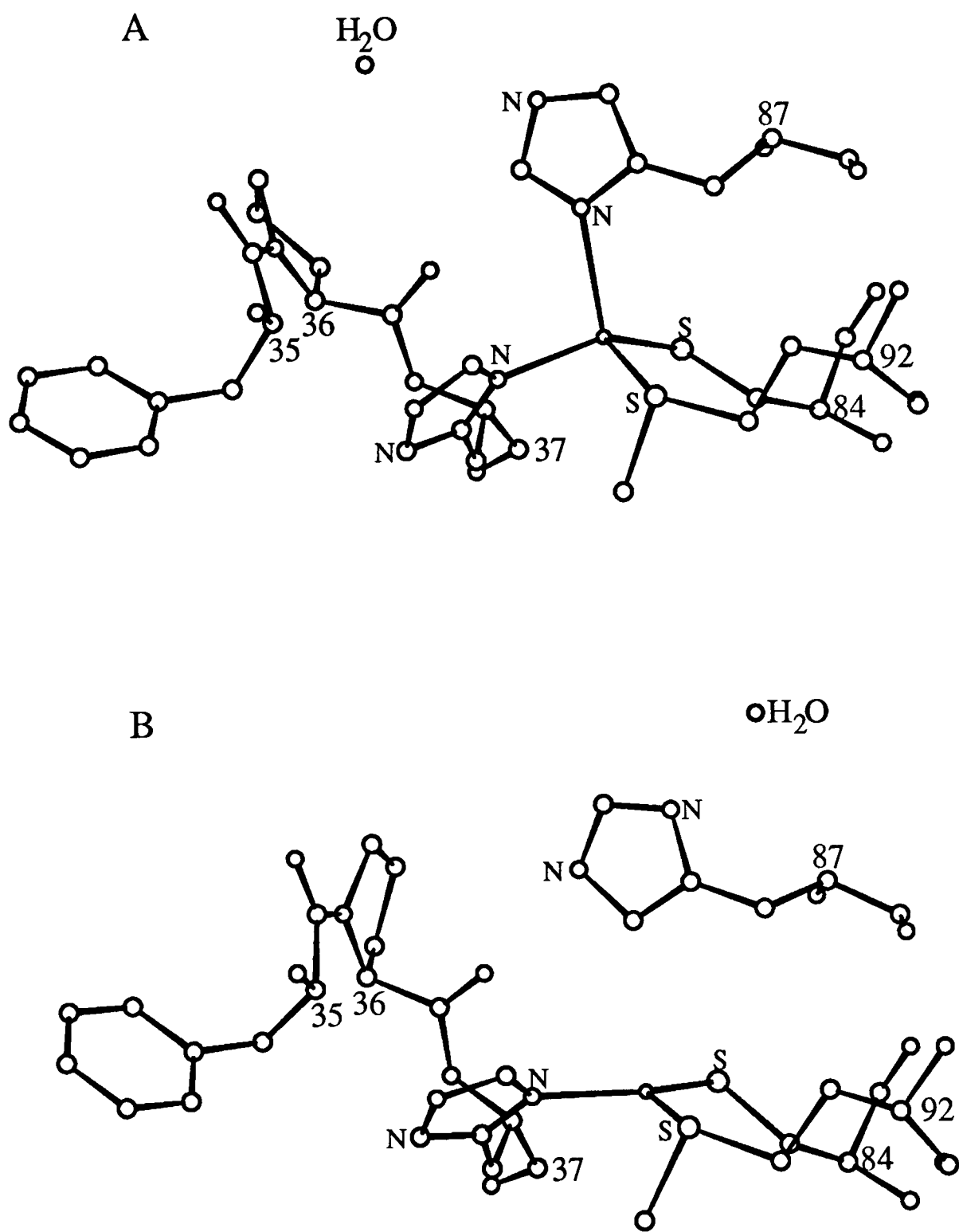


Figure 1.8: The Cu site in reduced plastocyanin (A) at pH 7.0 and (B) at pH 3.8. At pH 3.8, a water molecule is observed which can only be accommodated if the imidazole ring of His 87 is rotated by 180° about the C^β - C^γ bond relative to its orientation at pH 7.0 and in $\text{Cu}^{\text{II}}\text{Pc}$. Note also the flip of the side chain of Pro 36 from C^γ -*exo* conformation in $\text{Cu}^{\text{I}}\text{Pc}$ to C^γ -*endo* conformation in $\text{HCu}^{\text{I}}\text{Pc}$.

Protein	Cu-Met distance	E _{1/2}
Azurin (<i>A. denitrificans</i>)	3.12	276
Azurin (<i>P. aeruginosa</i>)	3.10	310
Plastocyanin (Poplar, pH 6.0)	2.90	380
Plastocyanin (Poplar, pH 8.0)	2.90	380
Plastocyanin (Algal)	2.90	390
Pseudoazurin (<i>A. faecalis</i>)	2.69	?
Rusticyanin	2.55	680

Table 1.3: Cu-Met bond lengths in some blue Cu proteins: a possible tuning effect of reduction potentials. Data taken from a poster presented at ICBIC 5 [25].

of a recent EXAFS study [25], in which rusticyanin (reduction potential 680mV) was found to have a Cu-Met bond length of 2.55Å. This is the shortest Cu-Met bond known to date in a blue Cu protein (Table 1.3).

Apart from the Cu site, the most striking structural change is the flip of the side chain of Pro 36 from *C^γ-exo* conformation in Cu^IPc to *C^γ-endo* conformation in HCu^IPc (Figure 1.8). A similar change occurs when Cu^{II} is replaced by Hg^{II} to give Hg^{II}Pc [39], and when Cu is removed to produce apo-plastocyanin [40]. Pro 36, which forms part of the rim and wall of the hydrophobic pocket, is highly flexible. This may explain why, despite the tight packing of side chains at the north site, Cu can be replaced by Hg²⁺ and extracted by CN⁻. Gaps may be created during fluctuations from the average structure. Thus, access of water molecules to the Cu site or to the protein interior cannot be ruled out. The structural evidence is that such access is unlikely to be other than transitory.

Reduced plastocyanin behaves as though it has a more exposed Cu site than does oxidised plastocyanin. For instance, Cu is displaced by Hg²⁺ more rapidly from reduced plastocyanin than from oxidised plastocyanin. The difference is probably due to the equilibrium between Cu^IPc and HCu^IPc. The latter has a Cu site which is exposed by the mobility of His 87.

1.3.3 Crystal Structure of *Enteromorpha prolifera* Pc

The structure of *E. prolifera* plastocyanin has been determined [13] by molecular replacement using models derived from the poplar plastocyanin structure. Alignment of sequences shows that *E. prolifera* plastocyanin has an additional residue at the N-terminus, but two fewer residues between positions 57 and 61 (see Table 1.1). There is homology at fifty-six residues. In an evolutionary sense, differences are more likely to represent changes from the algal to the higher plant protein than *vice versa*.

The poplar Cu^{II}Pc structure [20] is used for purposes of comparison (Table 1.2). Superposition of the two structures shows that there is a close correspondence between them. The Cu sites of the two proteins are essentially the same, but the kink in the backbone at positions 43–44 and the helical segment at 51–56 are significantly displaced. Deletion of two residues means that the β -bend at 58–60 in poplar plastocyanin is absent in *E. prolifera* plastocyanin. Glu 59 of *E. prolifera* plastocyanin has no structural equivalent in the poplar protein.

With the exception of 55 NH $\cdots$ O 52, 61 NH $\cdots$ O 58, 68 NH $\cdots$ O 65 and 88 NH $\cdots$ O 84, all of the intramolecular hydrogen bonds between backbone atoms in poplar plastocyanin also occur in *E. prolifera* plastocyanin. Two hydrogen bonds in *E. prolifera* plastocyanin correspond to distances $> 3.6\text{\AA}$ in poplar plastocyanin. These are 39 NH $\cdots$ O 57 and 88 NH $\cdots$ O 85.

Changes in tertiary structure can be related to sequence differences. For example, replacement of Ile 46 in poplar plastocyanin by Val 46 in *E. prolifera* plastocyanin allows the helical segment (51–56) to move closer to strand 4 until the side chain of Ala 52 is in van der Waals contact with the backbone at residue 43. Hydrogen bonds which cannot be formed in poplar plastocyanin then cause the helix to move closer to sheet II and away from sheet I. Strands 1 (sheet I) and 2B (sheet II) approach each other more closely in *E. prolifera* plastocyanin due to differences at residues 1, 19 and 21. Replacement of Pro at position 23 by Ala in *E. prolifera* plastocyanin makes possible a hydrogen bond to Asn 99.

The poplar and *E. prolifera* crystals have different space groups, allowing an investigation of the effects of intermolecular contacts. Differences in intermolecular hydrogen bonds induce a significant change in backbone conformation in the turn at 7–

10 and in the 32–35 loop, both at the northern end of the molecule. An intermolecular van der Waals contact involving Pro 36 and Gln 68, which is not found in poplar plastocyanin, accounts for a change to a C^γ -*endo* conformation of the Pro 36 side chain. This conformation has been observed in the poplar proteins HCu^IPc [36], Hg^{II}Pc [39] and apo-plastocyanin [40].

One conformer of the disordered residue Glu 43 has a side chain carboxylate which is close to the side chain carboxylate of Asp 53. Since the crystals are at pH_{meas} of 5.8, it would be expected that both carboxylates are deprotonated. The short (2.8Å) $\text{O}\cdots\text{O}$ contact makes this unlikely, so a $-\text{COOH}\cdots-\text{OOC}-$ hydrogen bond may be present. Such interactions, which are characteristic of aspartic proteinases [41, 42, 43], are associated with anomalously high pK_a values [44]. It is consistent with a pK_a of ≈ 6 that about half the *E. proliferans* plastocyanin molecules contain this hydrogen bond at pH 5.8.

1.4 Interactions of Plastocyanin with Redox Partners

1.4.1 Cytochrome *f*

Cytochrome *f* (cyt. *f*) and plastocyanin have been described [45] as functional analogues of cytochrome c_1 and cytochrome *c*, which function in the mitochondrial electron transport chain. Cyt. *f* has been purified from a number of plant, algal and cyanobacterial sources. It has a molecular weight of $\approx 33\,000$ and a redox potential of 360mV [46]. Cyt. *f* has lysine as its sixth ligand [47, 48], unlike other *c*-type cytochromes which have methionine in this position.

Electron transfer from cyt. *f* to Cu^{II}Pc is very rapid, with rate constants in excess of $10^7\text{ M}^{-1}\text{ s}^{-1}$ having been reported [49, 50, 51]. Cu^{II}Pc is thirty times more reactive with $\text{Fe}^{II}\text{cyt. } f$ than with other cytochromes [49], consistent with high specificity of interaction between cyt. *f* and plastocyanin. Dependence of the rate on ionic strength [50] and on the extent of chemical modification [52] of plastocyanin indicate that electrostatic interactions are important in the reaction between these two proteins.

In order to determine the binding site on plastocyanin for cyt. *f*, Sykes and co-workers [53, 54] have employed Cr^{III}-modification of plastocyanin [55], competitive inhibition with such complexes as (NH₃)₅CoNH₂Co(NH₃)₅⁵⁺, and pH effects. The results from each of these approaches indicate involvement of the acidic patch around Tyr 83 on plastocyanin in its reaction with cyt. *f*. This is despite the fact that cyt. *f* has an overall negative charge at pH > 5.5 [56]. The pH profile of rate constants for the reduction of Cu^{II}Pc with Fe^{II}cyt. *f* gives a p*K*_a of ≈5, believed to correspond to protonation of a residue in the acidic patch. Not all reductants exhibit such a pH dependence. For example, reduction with Fe(CN)₆⁴⁻ shows no pronounced dependence in the pH 4.5–7.5 range, consistent with reaction at the hydrophobic north site of plastocyanin.

The haem-containing portion of cyt. *f* (residues 1–250, 22% of which are charged) is thought [5] to be located in the intrathylakoid space where it can interact directly with plastocyanin (Figure 1.9). The overall charge on this fragment of the protein is small and negative. It has been speculated [5] that five pairs of positively charged residues observed in the sequence are brought together in the fold to form a recognition site for plastocyanin.

Plastocyanin and cyt. *f* have been covalently linked in the presence of the zero-length chemical cross-linker EDC [57, 58, 59]. The adducts have 1:1 stoichiometry. The reduction potential of the plastocyanin component is unchanged, whereas that of the cyt. *f* component is shifted by –20mV relative to free cyt. *f*, in the direction favouring electron transfer from cyt. *f* to plastocyanin. The degree of haem exposure is found to be decreased in the adduct relative to that in free cyt. *f*, indicating that the exposed portion of the haem is important in electron transfer. The plastocyanin/cyt. *f* adduct is incapable of donating electrons to PSI, so that a common region or feature may be involved in plastocyanin's interactions with both of its physiological partners. Intracomplex electron transfer has been observed in one study [58], but not in another [59]. Two sites of covalent linkage have been identified [57]: Asp 44 to Lys 187 and Glu 60 (plus perhaps Glu 59) to an unidentified lysine (the plastocyanin residue is given first in each of these pairs). The former occurs in a highly conserved part of the cyt. *f* sequence, and is in one of the three fragments (45–69, 150–166 and 180–190)

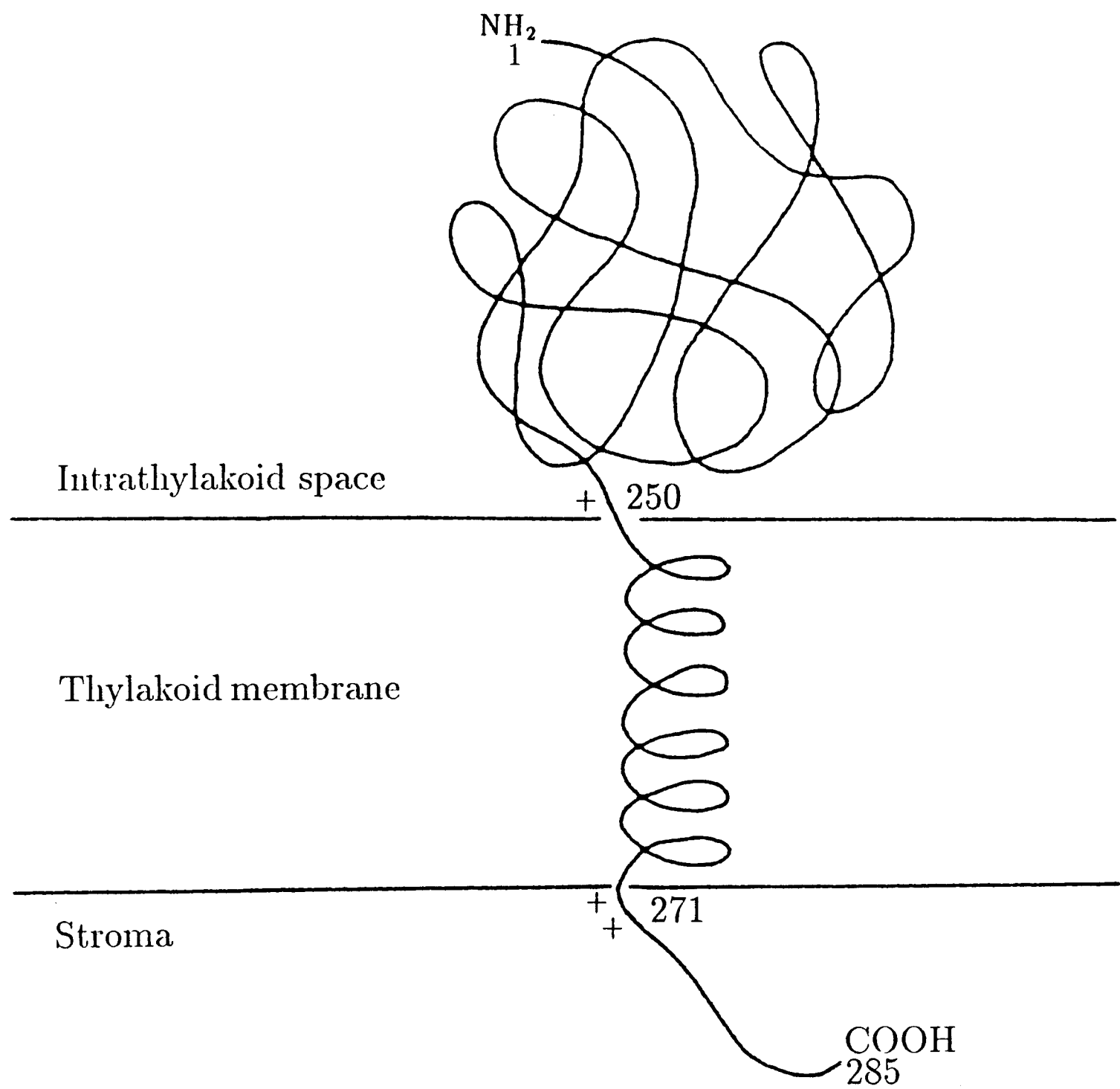


Figure 1.9: Model of the orientation of cyt. *f* in the thylakoid membrane.

suggested as potential candidates for binding sites [60]. The plastocyanin residues identified are supported by chemical modification studies [61], which indicate that both 42–45 and 59–61 fragments of the acidic patch are involved in interaction with cyt. *f*. Involvement of the 59–61 portion may explain the lack of reaction with PSI [59], since modification of this region altered plastocyanin’s interaction with PSI as well as with cyt. *f* [61]. Another study [62], involving modification of basic residues of cyt. *f* in the presence and absence of plastocyanin, identified Arg 88 and Arg 154 as being involved in interaction with plastocyanin. It should be noted that attempts to identify specific residues were limited to arginines.

It has been suggested [63] that the primary site of recognition for cyt. *f* on plastocyanin is the acidic patch, but that the final site of electron transfer may be at the hydrophobic north site. Thus initial docking of plastocyanin is followed by diffusion or rotation [64]. A similar proposal has been made for the interaction between cyt. *c* and plastocyanin [65]. Alternatively, cyt. *f* is able to react with both binding sites simultaneously. A more recent report from the same laboratory [66] proposes a new model for binding of plastocyanin to cyt. *f*. In this, the initial binding is not directly at the Tyr 83 site of plastocyanin, but is more towards the backside of the molecule, nearer Glu 68. In this case, cyt. *f* could donate electrons at either the east or north sites.

A very recent report [67] on the plastocyanin/cyt. *f* interaction contrasts with those [63, 66] considered above, and agrees more closely with the interpretation of Sykes and co-workers [53, 54]. This study [67] employs stopped flow spectrophotometrical analysis of reactions involving $\text{Cu}^{II}\text{Pc}/\text{Fe}^{II}\text{cyt. } f$, $\text{Cu}^I\text{Pc}/\text{Fe}^{III}\text{cyt. } f$ and $\text{NO}_2\text{-Y83-Pc}/\text{cyt. } f$ pairs. The results are compared with those predicted by theory, and indicate that electron transfer takes place *via* the remote Tyr 83 site on plastocyanin. The Tyr 83-Cys 84-Cu route is postulated as a facile through bond pathway for electrons. Further support for this electron transmission route is derived from analysis of the electrostatic orientation of the plastocyanin/cyt. *c* complex: multiple precollision orientations appear to converge onto the Tyr 83 electron transfer pathway [68].

1.4.2 Photosystem I

Both plastocyanin and PSI have a net negative charge at pH 7. Divalent cations, which probably provide electrostatic shielding, are required for reaction of plastocyanin with the P700 subunit of the PSI complex [69, 70]. Chemical modification studies [71, 61] indicate that residues 59–61 and 68, and not residues 42–45, of plastocyanin are involved in binding PSI. This suggests that electron transfer takes place at the hydrophobic north site *via* the Cu ligand His 87. It is possible that plastocyanin can bind cyt. *f* and PSI simultaneously, although this would seem an unlikely occurrence as PSI and the cytochrome *b*₆/*f* complex are probably separated by a considerable distance in the thylakoid membrane [72], consistent with the view that plastocyanin is a mobile electron shuttle [70] (see 1.1).

Plastocyanin can be cross-linked to PSI [73], blocking further binding of non-cross-linked plastocyanin to PSI. Cross-linking takes place *via* a single specific 19kDa subunit of PSI called PSI-200. This subunit does not contain any prosthetic groups [69] but may serve to neutralise the negative charge repulsion between plastocyanin and PSI by providing a docking site for plastocyanin. From current available data [73, 74], PSI is organised such that ferredoxin and plastocyanin are surface-exposed on opposite sides of the thylakoid membranes (Figure 1.10).

1.4.3 Inorganic Complexes

In an extension of the kinetic studies involving higher plant plastocyanins [4, 75, 76, 77], the reactivities of algal plastocyanins have been investigated [78, 79, 80]. These will be considered here since they give an indication of the effect of sequence on reactivity, and one of them, *S. obliquus* plastocyanin, is related in terms of primary structure to parsley plastocyanin.

The plastocyanin from the blue-green alga *A. variabilis* has one hundred and five residues, an overall charge of +2 and a pI of 7.8. Only one of the highly conserved acidic residues of higher plants at positions 42–45 and 59–61 is retained. The north site is essentially non-polar but has some positive residues not too distant. *S. obliquus* plastocyanin has an overall charge of –9 at pH 7, excluding His 59, and residue 45 is

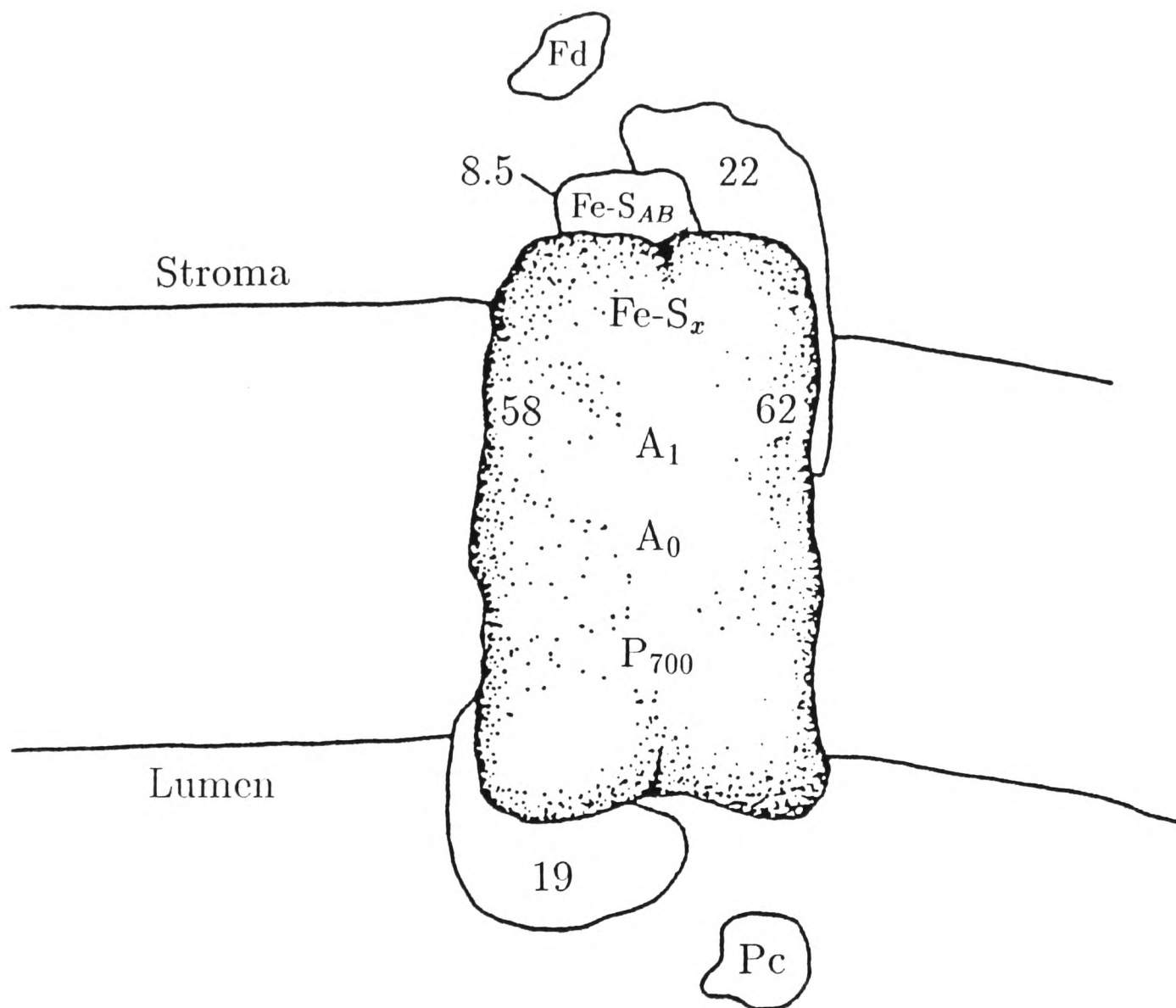


Figure 1.10: Model for the structural and functional organisation of subunits in the PSI complex. 19 refers to a 19kDa subunit which has been implicated in binding plastocyanin. Ferredoxin interacts with the 22kDa subunit (labelled 22). Electrons are probably donated from Fe-S centres A and B to ferredoxin.

uncharged (Ala).

The reduction potential of *A. variabilis* plastocyanin (343mV at pH 7.5) is reasonably similar to that of higher plant plastocyanins (365mV). Also, like higher plant plastocyanins, it becomes redox-inactive (at least with inorganic complexes) at low pH. The pH profile of its reaction with $\text{Fe}(\text{CN})_6^{3-}$ gives a pK_a of 5.0, assigned to protonation and dissociation of His 87. At $\text{pH} < 7$, *S. obliquus* plastocyanin shows some similarities with parsley plastocyanin. For example, its active site pK_a is higher than the typical higher plant value, though not as high as the anomalous parsley value. Both proteins have deletions at positions 57 and 58, with Pro at position 60 of parsley plastocyanin perhaps further influencing this protein's unusual behaviour. In terms of pK_a values, *A. variabilis* plastocyanin more closely resembles spinach and French bean plastocyanins than does parsley plastocyanin.

The pK_a s for reaction of plastocyanins with $\text{Co}(\text{phen})_3^{3+}$ require interpretation in terms of a second protonation (pK'_a) in addition to the active site protonation [4] (see Table 7.15 for pK_a values of several plastocyanins). For higher plant plastocyanins, the most acceptable explanation for pK'_a is protonation of carboxylate(s) at the acidic patch. This may not hold for *A. variabilis* plastocyanin, which retains only Asp 42 as an acidic residue at this site, but which has Glu 85 close by. Most higher plant plastocyanins do not have glutamate at position 85. Parsley plastocyanin and barley plastocyanin are two exceptions. In *A. variabilis* plastocyanin, Asp 42 and Glu 85 could together provide a binding site. However, NMR experiments with $\text{Cr}(\text{phen})_3^{3+}$ show no line broadening, implying very weak, if any, binding. No more than 45% of $\text{Co}(\text{phen})_3^{3+}$ reaction takes place at the east site, compared with 72% and 63% for spinach plastocyanin and parsley plastocyanins respectively [75]. The remainder of the reaction may be at the hydrophobic north site.

S. obliquus plastocyanin behaves similarly to other plastocyanins in its reaction with $\text{Co}(\text{phen})_3^{3+}$. In terms of sequence variation, it is noted that poplar plastocyanin with a -3 charge in the 42–45 fragment, has the lowest pK'_a , parsley plastocyanin with its deletions at 57 and 58 has the highest, and *S. obliquus* plastocyanin, with both of these features, has a similar pK'_a to spinach and French bean plastocyanins.

The observed oxidation state dependence of pK'_a may be related to conformational

changes at the east face upon reduction [81]. This could be part of a mechanism for enhancing the binding of $\text{Cu}^{\text{II}}\text{Pc}$ to $\text{Fe}^{\text{II}}\text{cyt. } f$ prior to electron transfer, followed by a reduction in affinity after electron transfer [4, 76].

NMR studies with $\text{Cr}(\text{phen})_3^{3+}$ show that *S. obliquus* plastocyanin has two main interaction regions: 42–44 and Asp 60–Asp 61, with weaker binding at acidic residues 75 and/or 79. It is difficult to tell whether 42–44 and 60–61 form distinct sites or a diffuse region of charge. A similar point has been made with respect to spinach plastocyanin in this laboratory [82]. One possible effect of the deletions in *S. obliquus* (and parsley) plastocyanin is to bring 59–61 closer to 42–45. This point is addressed in Chapter 7 of this thesis.

S. obliquus plastocyanin exhibits an additional pH effect at $\text{pH} > 7$ in reactions with $\text{Fe}(\text{CN})_6^{3-}$ and $\text{Co}(\text{phen})_3^{3+}$. This has been attributed to deprotonation of His 59, which has an NMR-determined pK_a of 7.8. For $\text{Co}(\text{phen})_3^{3+}$ the effect can partly be put down to increased affinity for positively charged species. This would be consistent with the closer proximity of 42–44 and 59–61, as a result of deletion of 57 and 58, to give a significant combined effect. There is evidence [79] that the state of protonation of His 59 influences binding over a broad region of the surface of *S. obliquus* plastocyanin, perhaps even taking in Glu 85. This is an additional acidic residue found in parsley, barley and *S. obliquus* plastocyanins, as mentioned above. Glu 85 may also be relevant to the reactivity of *A. variabilis* plastocyanin.

The explanation for the increase in reaction rate with $\text{Fe}(\text{CN})_6^{3-}$ on deprotonation of His 59 is less obvious. His 59 has negative charge around it, and NMR experiments indicate no specific binding of $\text{Cr}(\text{CN})_6^{3-}$ anywhere on the surface of *S. obliquus* plastocyanin. However, NMR titration experiments do indicate a sensitivity of His 59 to Cu site protonation and *vice versa*. It is not known how these effects are transmitted to one another, but it seems that deprotonation of His 59 decreases the reduction potential of the active site by 7–15mV.

Finally, the interchange of Phe and Tyr at positions 82 and 83 seems to have little or no effect on the reactivity of *S. obliquus* plastocyanin.

1.5 Mutant Plastocyanins

The prerequisite for mutant construction is the production of a functional, copper containing recombinant plastocyanin. This has been achieved using an expression vector designed for overexpression of plastocyanin in the periplasmic space of *E. coli* [83]. More recently, genes for *P. aeruginosa* azurin and poplar plastocyanin have been constructed and site saturation mutagenesis has been used to alter two Cu binding residues of azurin (Met 121 and His 46) and Met 92 of plastocyanin [84]. The azurin variants all bind Cu, whereas the plastocyanin mutants do not. Extension of this preliminary work should allow improved understanding of properties such as electronic spectra, redox potentials, rates of electron transfer, and the pathways by which electrons are transmitted to and from partner proteins and through the protein interior.

1.6 Previous 2D NMR Studies of Plastocyanin

Sequential assignments for the plastocyanins from spinach [85], French bean [86, 87] and *S. obliquus* [88] have been obtained. Partial assignments of the protein from parsley were also available in this laboratory [89]. These previous NMR studies are considered in Chapters 6 and 7.

1.7 Other Blue Cu Proteins

The characteristic properties of the blue Cu centre have been described earlier in this chapter. Proteins containing a type 1 Cu site constitute a diverse class which includes small proteins (like plastocyanin) and multicopper oxidases. The small blue proteins are all monomeric, contain a single type 1 Cu centre, and are involved in electron transport chains in bacteria, algae and plants. The small blue Cu proteins have the same general fold, with variations, as plastocyanin. Thus, they are all single domain β -sheet sandwiches, composed of eight strands in two sheets with mainly antiparallel topology (Figure 1.3). It has been suggested that there are several sub-categories of blue Cu protein structure [3]. Sequence alignment studies [29, 31] have

Protein	Source	M_r	No. of a.a.	pI	E^0 (mV)	λ_{max}/ϵ (nm)/(M ⁻¹ cm ⁻¹)
Plastocyanin	Plants, algae	10500	99	4.2	370	597/4500
Azurin	<i>P. aeruginosa</i>	14000	128	5.4	330	625/4800
Azurin	<i>P. putida</i>	14400	133	7.3	?	620/4200
Pseudoazurin	Bacteria	14000	123	?	250	?
Stellacyanin	Lacquer tree	20000	107	9.9	184	608/4080
Rusticyanin	<i>T. ferro-oxidans</i>	16000	159	9.1	680	597/1950
Umecyanin	Horse-radish	14600	125	5.9	283	610/3400
Amicyanin	<i>T. versutus</i>	11000	105	?	?	?

Table 1.4: A comparison of properties of small blue Cu proteins. The reported plastocyanin pI is that of the protein from spinach. Stellacyanin contains of 40% carbohydrate.

shown that the Cu binding residues are almost invariant among the blue Cu proteins. Comparison of poplar plastocyanin and *Pseudomonas aeruginosa* azurin [29] resulted in identification of five residues, apart from the Cu ligands, which are invariant in these two proteins, and which are highly conserved in the class as a whole. These residues are Asn 38 (Asn 47 in azurin), Val 40 (Val 49 in azurin), Tyr 80 (Tyr 108 in azurin), Pro 86 (Pro 115 in azurin) and Gly 94 (Gly 123 in azurin). Some of the properties of a number of small blue Cu proteins are presented in Table 1.4.

Multicopper oxidases are large, complex proteins which contain several Cu atoms. All appear to have at least one blue Cu site. This family of proteins includes ascorbate oxidase, laccase, ceruloplasmin, and coagulation factors, factor V/factor VII. Multicopper oxidases reduce O₂ to water, with accompanying one electron transfer from the substrate. Determination of the crystal structure of ascorbate oxidase from zucchini [90] confirmed that there is an evolutionary relationship between multicopper oxidases and small blue Cu proteins.

1.8 *p*-Cresol Methylhydroxylase (PCMH)

PCMH [4-cresol:(acceptor)oxidoreductase (methyl-hydroxylating), EC 1.17.99.1] [91], a flavocytochrome *c* found in the periplasmic space of certain pseudomonads, catalyses dehydrogenation and hydration of *p*-cresol and its homologues to the corresponding alcohols. The latter are then further dehydrogenated to the corresponding

aldehydes or ketones. Thus, oxidation of *p*-cresol yields *p*-hydroxybenzyl alcohol and *p*-hydroxybenzaldehyde [92].

Seven forms of PCMH have been isolated from six bacterial sources [93, 94]. *Pseudomonas putida* strain NCIB 9869 produces two forms of PCMH [95]. The subject of this study, *P. putida* PCMH form A, has a molecular weight of approximately 116 000. It contains a covalently linked flavin with tryosine as the substituent on C-8 α of FAD [96]. The enzyme can be reversibly resolved into its flavoprotein (M_r 49 000) and cytochrome *c* (M_r 8780) subunits [97], permitting the demonstration that the cytochrome subunit regulates the catalytic activity of the FAD in the flavin subunit.

In common with PCMH, numerous other bacterial enzymes contain covalently bound FAD or FMN as the redox cofactor. Also in this class are the mammalian mitochondrial enzymes succinate dehydrogenase (heart), monoamine oxidase (liver, brain, placenta) and sarcosine dehydrogenase (liver). Although these enzymes catalyse different reactions, there may be common biosynthetic pathways by which flavin is attached. The mode of transport of the enzymes to the proper cellular compartments may also be similar. PCMH and *Pseudomonas putida* may thus constitute a good model system in providing insights into the properties of other enzymes in this class.

1.9 Kinetic and Mechanistic Studies of PCMH

On anaerobic titration of PCMH with either dithionite or *p*-cresol, three reaction phases are observed [91]. First, the haem is reduced, then an anionic flavin radical is formed, and finally the flavin is fully reduced. The observation of three distinct one electron steps with the obligatory two electron reductant *p*-cresol indicates that intraenzymic electron transfer between flavin and haem groups occurs. The cytochrome-free α_2 (see discussion of structure below) flavoprotein is also reduced by dithionite and *p*-cresol. Again, the anionic flavin radical is observed [91].

PCMH has been shown to obey a ping-pong type steady state kinetic mechanism [91, 98]. The isolated flavoprotein dimer retains only 2% of the activity of the flavocytochrome, and functions *via* either a partial equilibrium random, or a steady state

random mechanism [98].

The overall mechanism involves hydroxylation of *p*-cresol by abstraction of two hydrogen atoms by the flavoprotein followed by attack by water on the resulting intermediate. The two electrons are transferred separately from the fully reduced flavin to the cytochrome subunit (cytPCMH) and then to an acceptor protein which is postulated to be an azurin [91].

1.10 Three Dimensional Structure of PCMH

The structure of PCMH was first determined at 6.0Å resolution [99] and is now available at 3.0Å resolution [100]. PCMH is an $\alpha_2\beta_2$ tetramer with the flavoprotein subunits tightly packed about the molecular 2-fold axis, sharing a common interface. The cytochrome subunits are located on the outside of the flavoprotein dimer, each in a depression on the surface of a flavoprotein subunit. The polypeptide backbone and active sites of the PCMH $\alpha\beta$ dimer are shown in Figure 1.11.

1.10.1 The Cytochrome Subunit

CytPCMH (Figure 1.12) has a number of features characteristic of the cytochrome *c* fold [101]. There are three helical segments formed by residues 8–14, 36–45 and 67–72, and three reverse turns at 2–5, 47–50 and 55–58. The remainder is in irregular or extended conformation. The haem iron is ligated by His 19 and Met 50, while the haem itself is covalently attached to the polypeptide at Cys 15 and Cys 18 through the vinyl groups at positions 2 and 4 (see Chapter 5). Even in the isolated subunit most of the haem is shielded from solvent.

The cytochrome has ten acidic and seven basic side chains (Table 1.5). These, together with the two haem propionates and N- and C-termini, give an overall total at neutral pH of thirteen negative charges and eight positive charges. These figures are consistent with the pI of 4.55–4.70 [98]. There is a slight concentration of acidic residues at the ‘back’ of the molecule. Acidic and basic residues are evenly mixed on the two sides. The ‘front’, which faces the flavoprotein, is essentially neutral except for Arg 44 and Arg 48, which seem to have their side chains oriented away from the

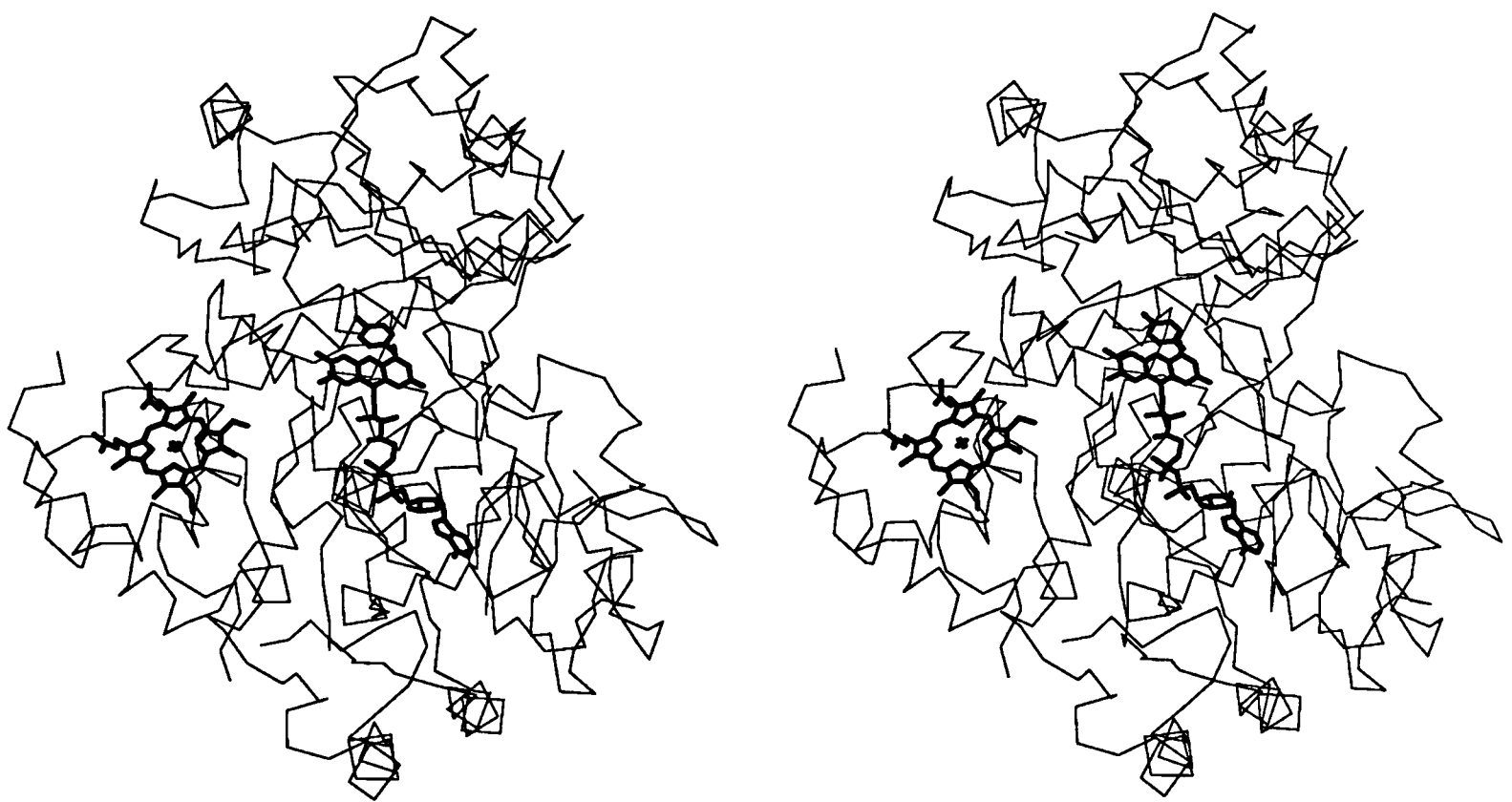


Figure 1.11: Stereo view of the $\alpha\beta$ dimer of PCMH. The FAD and tyrosine substituent of the flavoprotein subunit, and the haem prosthetic group of the cytochrome subunit, are shown in bold.

flavoprotein.

1.10.2 The Flavoprotein Subunit

This subunit is folded into three domains. The N-terminal domain (first two hundred and fifteen residues) wraps around most of the FAD cofactor. The central domain (next two hundred and eight residues) covers the flavin ring and the C-terminal domain (last thirty-nine residues), which lies against the N-terminal domain, covers the adenosine part of FAD.

The N-terminal domain contains four helices and the first eleven β -strands. The remainder consists of turns and extended chains. The ADP-ribose fragment of FAD is extended and lies in a groove in the N-terminal domain. This groove is covered by the small C-terminal domain. The central domain is a seven-stranded anti-parallel β -sheet which forms a dome over the isoalloxazine ring of the flavin. An eighth strand lies on the edge of this structure and would form an eighth antiparallel strand if it were not oblique to, and formed more than one or two hydrogen bonds with, the seventh strand. It is, however, convenient to consider this strand to be part of an eight-stranded antiparallel sheet. On the outside of the dome are four antiparallel α -helices arranged side-by-side, and there is one helix inside the dome.

The basic four-stranded antiparallel motif of the central domain is rare. It has been observed in the eight-iron ferredoxin [103] and in bacterial ribulose-1,5-biphosphate carboxylase [104]. The eight-stranded form of the dome of PCMH, produced by duplication of the basic motif, appears to be unique.

1.10.3 Cytochrome-Flavoprotein Interaction

Domain 1 makes the major contribution to the cytochrome binding site on the flavoprotein, with some contribution from domain 2. The segments of the cytochrome involved in the interface are 10–19, 23–25 and 44–57, with 17–18, 23–25, 44–45, 48–49 and 56–57 in closest contact (Figure 1.12).

The haem iron and the centre of the flavin ring are separated by $\approx 18\text{\AA}$, while the closest approach of the prosthetic groups is $8\text{\AA}$. The benzenoid portion of the isoalloxazine ring is directed toward the thioether-containing edge of the haem. The

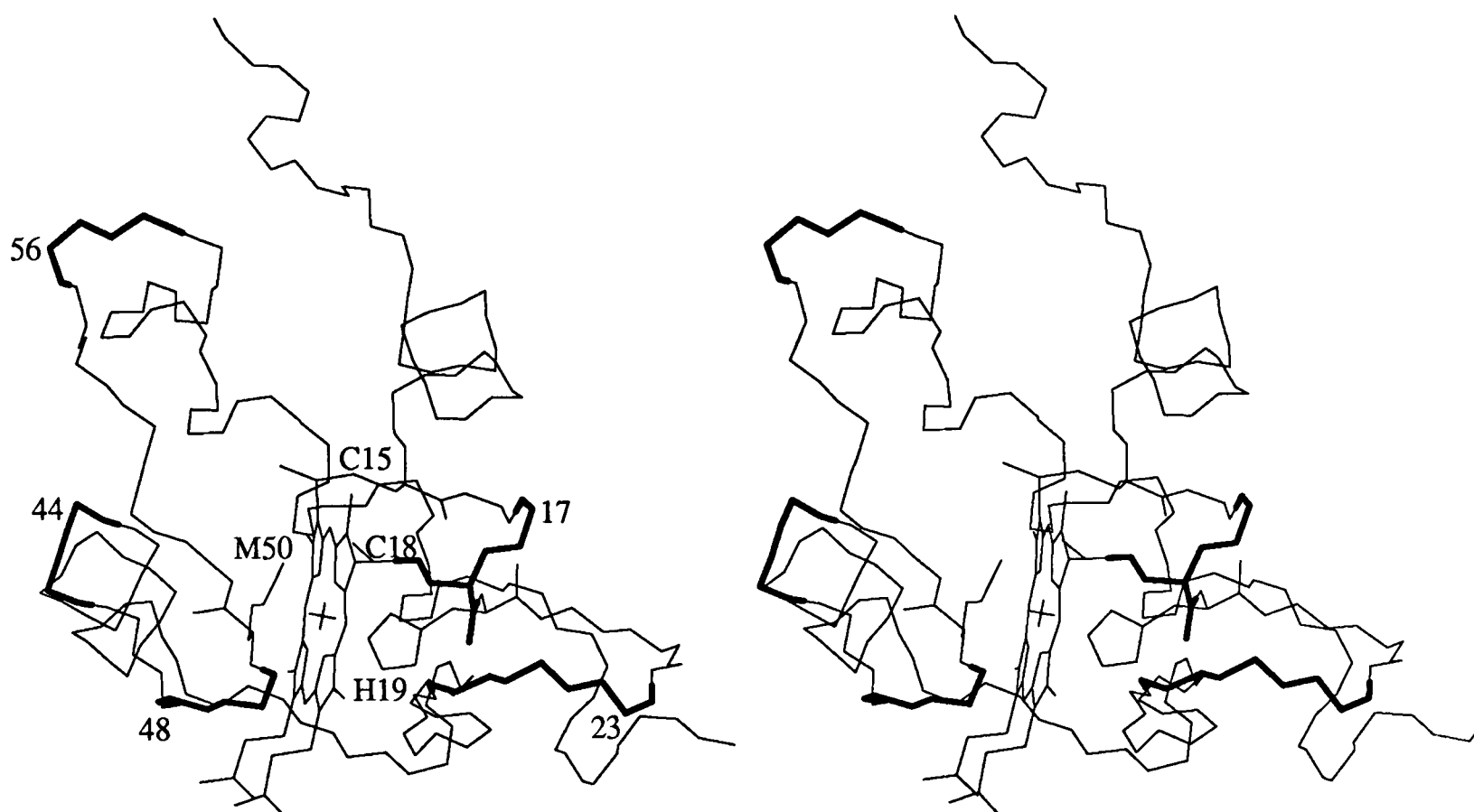


Figure 1.12: Backbone tracing of the cytochrome subunit of PCMH. The N- and C-termini are indicated. The side chains of the Fe ligands, His 19 and Met 50, are shown, as are those of Cys 15 and Cys 18, through which the haem is covalently bound to the polypeptide. The polypeptide fragments in close contact with the flavoprotein subunit, 17–18, 23–25, 44–45, 48–49 and 56–57, are in bold. The numbering of the first C α in each fragment is indicated.

Asp	Ser	Gln	Trp	Gly	Ser	Gly	Lys	Asn	Leu	Tyr	Asp	Lys	Val	<i>Cys</i>
Gly	<u>His</u>	<u>Cys</u>	<i>His</i>	Lys	Pro	Glu	<u>Val</u>	<u>Gly</u>	<u>Val</u>	Gly	Pro	Val	Leu	Glu
Gly	Arg	Gly	Leu	Pro	Glu	Ala	Tyr	Ile	Lys	Asp	Ile	Val	<u>Arg</u>	<u>Asn</u>
Gly	Phe	<u>Arg</u>	<u>Ala</u>	<i>Met</i>	Pro	Ala	Phe	Pro	Ala	<u>Ser</u>	<u>Tyr</u>	Val	Asp	Asp
Glu	Ser	Leu	Thr	Gln	Val	Ala	Glu	Tyr	Leu	Ser	Ser	Leu	Pro	Ala
Pro	Ala	Gln												

Table 1.5: Amino acid sequence of cytPCMH. The four ligands are italicised. Residues making up the three helical segments are in boldface. The five segments in closest contact with the flavoprotein are underlined.

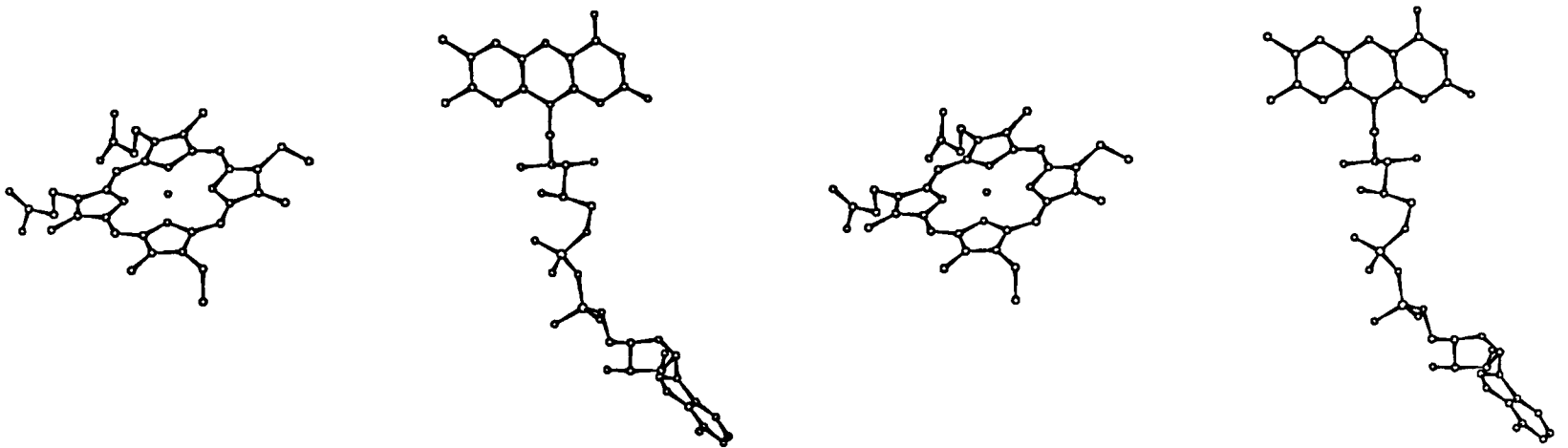


Figure 1.13: Stereo view of the relationship between the FAD and haem groups in the $\alpha\beta$ dimer of PCMH.

flavin and haem planes make an angle of 65° to each other (Figure 1.13).

The most direct electron transfer route is from the benzene ring of the flavin to pyrrole ring II of the haem. This is similar to pathways proposed for other flavodoxin-cytochrome interactions [105, 106]. As for electron transfer to azurin, the most likely area would seem to be around the propionate groups of cytPCMH. This face of the cytochrome remains partially exposed in the complex with the flavoprotein subunit, so that the haem is here most accessible (Figure 1.11).

1.10.4 Comparisons with Other Proteins

The conformation of the flavoprotein subunit of PCMH is significantly different from that of other structurally characterised flavoproteins. Also, in all flavoproteins of known structure except acyl-CoA dehydrogenase and PCMH, the phosphate or pyrophosphate moiety of FAD lies at the N-terminal end of an α -helix. The helix positive dipole [107] is thought to stabilise binding of the negatively charged phosphate. The lack of this stabilisation may be compensated in PCMH by the fact that FAD is covalently bound, although this is open to question.

CytPCMH is a member of the bacterial subclass of cytochromes [108], which differ from their eukaryotic counterparts in two main respects: they are acidic ($pI < 5.0$) and have a major deletion at the bottom of the molecule near the propionates.

High resolution structures are available for two other small bacterial cytochromes.

These are cytochrome c_5 (cyt. c_5) from *Azotobacter vinlandii* [109] and cytochrome c_{551} (cyt. c_{551}) from *Pseudomonas aeruginosa* [110]. CytPCMH has fewer residues (30) separating the iron ligands than either cyt. c_5 (39) or cyt. c_{551} (43). CytPCMH has only one helix in this region compared to two in both of the other cytochromes. The main effect of this is to leave the propionates more exposed to solvent in PCMH. From alignment of backbone structures, it would appear that the interaction regions of the three proteins with their respective reductases are in corresponding positions, and that the interaction of cyt. c_5 with its (unknown) partner is geometrically closer to that of cytPCMH than that of cyt. c_{551} .

As mentioned above, the most likely site of interaction of PCMH with its reoxidant is in the region of the propionate groups. The interaction of the other cytochromes with their reoxidants may be quite different, since they have less exposed haems. Since cyt. c_{551} is monomeric, both redox processes may occur *via* the same face. The same may apply to cyt. c_5 , although this protein may be membrane-bound [109].

Bibliography

- [1] D. Boulter, B. G. Haslett, D. Peacock, J. A. M. Ramshaw & M. D. Scawen (1977) *Int. Rev. Biochem* 13, 1–40.
- [2] H. C. Freeman (1981) in *Coordination Chemistry-21* (J. L. Laurent, Ed.) pp. 29–51, Pergamon Press.
- [3] E. T. Adman (1985) in *Metalloproteins* (P. M. Harrison, Ed.) Part I, pp. 1–42, Macmillan.
- [4] A. G. Sykes (1985) *Chem. Soc. Rev.* 14, 283–315.
- [5] D. L. Willey, A. D. Auffret & J. C. Gray (1984) *Cell* 36, 555–562.
- [6] P. Mitchell (1979) *Science* 206, 1148–1159, and references cited therein.
- [7] H. Yano, M. Kamo, A. Tsugita, K. Aso & Y. Nosu (1989) *Protein Seq. Data Anal.* 2, 385–389, and references cited therein.
- [8] J. Kelly & R. P. Ambler (1974) *Biochem. J.* 143, 681–690.
- [9] R. J. Simpson, R. L. Moritz, E. C. Nice, B. Grego, F. Yoshizaki, Y. Sugimura, H. C. Freeman & M. Murata (1986) *Eur. J. Biochem.* 157, 497–506.
- [10] J. M. Kelly & R. P. Ambler, unpublished work cited in [4].
- [11] F. Yoshizaki, T. Fukazawa, Y. Mishina & Y. Sugimura (1989) *J. Biochem.* 106, 282–288.
- [12] A. Aitken (1975) *Biochem. J.* 149, 675–683.

- [13] C. A. Collyer, J. M. Guss, Y. Sugimura, F. Yoshizaki & H. C. Freeman (1990) *J. Mol. Biol.* **211**, 617–632.
- [14] P. S. Nielsen & K. Gausing (1987) *FEBS Lett.* **225**, 159–162.
- [15] M. I. Dimitrov, A. A. Donchev & T. A. Egorov (1990) *FEBS Lett.* **265**, 141–145.
- [16] D. Boulter, J. T. Gleaves, B. G. Haslett, D. Peacock & U. Jensen (1978) *Phytochemistry* **17**, 1585–1589.
- [17] A. A. Donchev & M. I. Dimitrov (1988) *Cr Acad. Bulg. Sci.* **41**, 69–72.
- [18] B. V. Chapman, P. M. Colman, H. C. Freeman, J. M. Guss, M. Murata, V. A. Norris, J. A. M. Ramshaw & M. P. Venkatappa (1977) *J. Mol. Biol.* **110**, 187–189.
- [19] P. M. Colman, H. C. Freeman, J. M. Guss, M. Murata, V. A. Norris, J. A. M. Ramshaw & M. P. Venkatappa (1978) *Nature* **272**, 319–324.
- [20] J. M. Guss & H. C. Freeman (1983) *J. Mol. Biol.* **169**, 521–563.
- [21] R. Huber & W. Steigemann (1974) *FEBS Lett.* **48**, 235–237.
- [22] E. T. Adman & L. H. Jensen (1981) *Isr. J. Chem.* **21**, 8–12.
- [23] A. A. Gewirth & E. I. Solomon (1988) *J. Am. Chem. Soc.* **110**, 3811–3819.
- [24] R. A. Scott, J. E. Hahn, S. Doniach, H. C. Freeman & K. O. Hodgson (1982) *J. Am. Chem. Soc.* **104**, 5364–5369.
- [25] L. M. Murphy, S. S. Hasnain, I. Harvey, R. W. Strange, W. J. Ingledew & E. T. Adman (1991) *J. Inorg. Biochem.* **43**, 176. (Abstract for Poster C015, ICBIC 5, Oxford, U.K.).
- [26] K. W. Penfield, R. R. Gay, R. S. Himmelwright, N. C. Eickman, V. A. Norris, H. C. Freeman & E. I. Solomon (1981) *J. Am. Chem. Soc.* **103**, 4382–4388.

- [27] J. A. Thich, D. Mastropaolo, J. A. Potenza & H. J. Schugar (1974) *J. Am. Chem. Soc.* **96**, 726–731.
- [28] H. J. Prochaska, W. F. Schwindinger, M. Schwartz, M. J. Burk, E. Bernarducci, R. A. Lalancette, J. A. Potenza & H. J. Schugar (1981) *J. Am. Chem. Soc.* **103**, 3346–3355.
- [29] C. Chothia & A. M. Lesk (1982) *J. Mol. Biol.* **160**, 309–323.
- [30] J. M. Guss, E. A. Merritt, R. P. Phizackerley, B. Hedman, M. Murata, K. O. Hodgson & H. C. Freeman (1988) *Science* **241**, 806–811.
- [31] J. Van Beeumen, S. Van Bun, G. W. Canters, A. Lommen & C. Chothia (1991) *J. Biol. Chem.* **266**, 4869–4877.
- [32] D. J. Cookson, M. T. Hayes & P. E. Wright (1980) *Nature* **283**, 682–683.
- [33] P. M. Handford, H. A. O. Hill, R. W.-K. Lee, R. A. Henderson & A. G. Sykes (1980) *J. Inorg. Biochem.* **13**, 83–88.
- [34] M. G. Segal & A. G. Sykes (1978) *J. Am. Chem. Soc.* **100**, 4585–4592.
- [35] S. K. Chapman, I. Sanemasa & A. G. Sykes (1983) *J. Chem. Soc., Dalton Trans.*, 2549–2553.
- [36] J. M. Guss, P. R. Harrowell, M. Murata, V. A. Norris & H. C. Freeman (1986) *J. Mol. Biol.* **192**, 361–387.
- [37] R. A. Marcus & N. Sutin (1985) *Biochim. Biophys. Acta* **811**, 265–322.
- [38] J. McGinnis, J. D. Sinclair-Day & A. G. Sykes (1985) in *Biochemical and Inorganic Aspects of Copper Coordination Chemistry* (K. D. Karlin & J. Zubieta, Eds.) pp. 11–22, Adenine Press.
- [39] W. B. Church, J. M. Guss, J. J. Potter & H. C. Freeman (1986) *J. Biol. Chem.* **261**, 234–237.

- [40] T. P. J. Garrett, J. M. Guss, S. J. Rogers & H. C. Freeman (1984) *J. Biol. Chem.* **259**, 2822–2825.
- [41] M. N. G. James & A. R. Sielecki (1983) *J. Mol. Biol.* **163**, 299–361.
- [42] L. Pearl & T. Blundell (1984) *FEBS Lett.* **174**, 96–101.
- [43] K. Suguna, R. R. Bott, E. A. Padlan, E. Subramanian, S. Sherriff, G. H. Cohen & D. R. Davies (1987) *J. Mol. Biol.* **196**, 877–900.
- [44] L. Sawyer & M. N. G. James (1982) *Nature* **295**, 79–80.
- [45] P. M. Wood (1977) *Eur. J. Biochem.* **72**, 605–612.
- [46] A. G. Lippin (1981) in *Metal Ions in Biological Systems*, vol. 13, (H. Sigel, Ed.) pp. 1–71, Dekker.
- [47] D. J. Davis, M. K. Frame & D. A. Johnson (1988) *Biochim. Biophys. Acta* **936**, 61–66.
- [48] S. E. J. Rigby, G. R. Moore, J. C. Gray, P. M. A. Gadsby, S. J. George & A. J. Thompson (1988) *Biochem. J.* **256**, 571–577.
- [49] P. M. Wood (1974) *Biochim. Biophys. Acta* **357**, 370–379.
- [50] S. Niwa, H. Ishikawa, S. Nikai & T. Takabe (1980) *J. Biochem.* **88**, 1177–1183.
- [51] K. Tanaka, M. Takahashi & K. Asada (1981) *Plant Cell Physiol.* **22**, 33–39.
- [52] T. Takabe, H. Ishikawa, S. Niwa & Y. Tanaka (1984) *J. Biochem.* **96**, 385–393.
- [53] D. Beoku-Betts, S. K. Chapman, C. V. Knox & A. G. Sykes (1983) *J. Chem. Soc., Chem. Commun.*, 1150–1152.
- [54] D. Beoku-Betts, S. K. Chapman, C. V. Knox & A. G. Sykes (1985) *Inorg. Chem.* **24**, 1677–1681.
- [55] O. Farver, Y. Shahak & I. Pecht (1982) *Biochemistry* **21**, 1885–1890.

- [56] J. C. Gray (1978) *Eur. J. Biochem.* 82, 133–141.
- [57] D. J. Davis & K. Hough (1983) *Biochem. Biophys. Res. Commun.* 116, 1000–1006.
- [58] T. Takabe & H. Ishikawa (1989) *J. Biochem.* 105, 98–102.
- [59] L. Z. Morand, M. K. Frame, K. K. Colvert, D. A. Johnson, D. W. Krogmann & D. J. Davis (1989) *Biochemistry* 28, 8039–8047.
- [60] D. J. Davis (1987) *Prog. Photosynth. Res.* 2, *Proc. 7th Int. Congr. Photosynth.* 1986, 473–476.
- [61] G. P. Anderson, D. G. Sanderson, C. H. Lee, S. Durell, L. B. Anderson & E. L. Gross (1987) *Biochim. Biophys. Acta* 894, 386–398.
- [62] Z. Adam & R. Malkin (1989) *Biochim. Biophys. Acta* 975, 158–163.
- [63] E. L. Gross, A. Curtiss, S. R. Durell & D. White (1990) *Biochim. Biophys. Acta* 1016, 107–114.
- [64] S. H. Northrup, J. O. Boles & J. C. L. Reynolds (1988) *Science* 241, 67–70.
- [65] J. D. Rush, F. Levine & W. H. Koppenol (1988) *Biochemistry* 27, 5876–5884.
- [66] E. L. Gross & A. Curtiss (1991) *Biochim. Biophys. Acta* 1056, 166–172.
- [67] H. E. M. Christensen, L. S. Conrad & J. Ulstrup (1991) *J. Inorg. Biochem.* 43, 94. (Abstract for Poster A007, ICBIC 5, Oxford, U.K.).
- [68] V. A. Roberts, H. C. Freeman, A. J. Olson, J. A. Tainer & E. D. Getzoff (1991) *J. Biol. Chem.* 266, 13431–13441.
- [69] N. Haehnel, A. Präper & J. Krause (1980) *Biochim. Biophys. Acta* 593, 384–399.
- [70] T. Takabe, H. Ishikawa, S. Niwa & S. Itoh (1983) *J. Biochem.* 94, 1901–1911.
- [71] K. O. Burkey & E. L. Gross (1982) *Biochemistry* 21, 5886–5890.

- [72] J. M. Anderson (1982) *FEBS Lett.* 138, 62–66.
- [73] R. M. Wynn & R. Malkin (1988) *Biochemistry* 27, 5863–5869.
- [74] R. M. Wynn & R. Malkin (1988) *FEBS Lett.* 229, 293–297.
- [75] J. McGinnis, J. D. Sinclair-Day & A. G. Sykes (1986) *J. Chem. Soc., Dalton Trans.*, 2007–2009.
- [76] J. McGinnis, J. D. Sinclair-Day & A. G. Sykes (1986) *J. Chem. Soc., Dalton Trans.*, 2011–2012.
- [77] J. D. Sinclair-Day & A. G. Sykes (1986) *J. Chem. Soc., Dalton Trans.*, 2069–2073.
- [78] M. P. Jackman, J. D. Sinclair-Day, M. J. Sisley, A. G. Sykes, L. A. Denys & P. E. Wright (1987) *J. Am. Chem. Soc.* 109, 6443–6449.
- [79] J. McGinnis, J. D. Sinclair-Day, A. G. Sykes, R. Powls, J. Moore & P. E. Wright (1988) *Inorg. Chem.* 27, 2306–2312.
- [80] M. P. Jackman, J. McGinnis, R. Powls, G. A. Salmon & A. G. Sykes (1988) *J. Am. Chem. Soc.* 110, 5880–5887.
- [81] E. L. Gross, G. P. Anderson, S. L. Ketchner & J. E. Draheim (1985) *Biochim. Biophys. Acta* 808, 437–447.
- [82] F. A. Armstrong, P. C. Driscoll, H. A. O. Hill & C. Redfield (1986) *J. Inorg. Biochem.* 28, 171–180.
- [83] M. Nordling, T. Olausson & L. G. Lundberg (1990) *FEBS Lett.* 276, 98–102.
- [84] T. K. Chang, S. A. Iverson, C. G. Rodrigues, C. N. Kiser, A. Y. C. Lew, J. P. Germanas & J. H. Richards (1991) *Proc. Natl. Acad. Sci. USA* 88, 1325–1329.
- [85] P. C. Driscoll, H. A. O. Hill & C. Redfield (1987) *Eur. J. Biochem.* 170, 279–292.

- [86] W. J. Chazin, M. Rance & P. E. Wright (1988) *J. Mol. Biol.* 202, 603–622.
- [87] W. J. Chazin & P. E. Wright (1988) *J. Mol. Biol.* 202, 623–636.
- [88] J. M. Moore, W. J. Chazin, R. Powls & P. E. Wright (1988) *Biochemistry* 27, 7806–7816.
- [89] P. C. Driscoll (1987) D. Phil. thesis, University of Oxford.
- [90] A. Messerschmidt, A. Rossi, R. Ladenstein, R. Huber, M. Bolognesi, G. Gatti, A. Marchesini, R. Petruzzelli & A. Finazzi-Agro (1989) *J. Mol. Biol.* 206, 513–529.
- [91] W. McIntire, D. J. Hopper & T. P. Singer (1985) *Biochem. J.* 228, 325–335.
- [92] D. J. Hopper & D. G. Taylor (1977) *Biochem. J.* 167, 155–162.
- [93] S. C. Koerber, W. McIntire, C. W. Bohmont & T. P. Singer (1985) *Biochemistry* 24, 5276–5280.
- [94] I. D. Bossert, G. Whited, D. Y. Gibson & L. Y. Young (1989) *J. Bacteriol.* 171, 2956–2962.
- [95] M. H. Keat & D. J. Hopper (1978) *Biochem. J.* 175, 649–658.
- [96] W. McIntire, D. E. Edmondson, D. J. Hopper & T. P. Singer (1981) *Biochemistry* 20, 3068–3075.
- [97] W. McIntire & T. P. Singer (1982) *FEBS Lett.* 143, 316–318.
- [98] W. McIntire, D. J. Hopper & T. P. Singer (1987) *Biochemistry*, 26, 4107–4117.
- [99] N. Shamala, L. W. Lim, F. S. Mathews, W. McIntire, T. P. Singer & D. J. Hopper (1986) *Proc. Natl. Acad. Sci. U.S.A.* 83, 4626–4630.
- [100] F. S. Mathews, Z. Chen, H. D. Bellamy & W. S. McIntire (1991) *Biochemistry* 30, 238–247.

- [101] R. J. Almassy & R. E. Dickerson (1978) *Proc. Natl. Acad. Sci. U.S.A.* 75, 2674–2678.
- [102] W. S. McIntire, S. C. Koerber, C. W. Bohmont & T. P. Singer (1984) in *Flavins and Flavoproteins* (R. C. Bray, P. C. Engel & S. G. Mayhew, Eds.) pp. 643–654, DeGruyter, New York.
- [103] E. T. Adman, L. C. Sieker & L. H. Jensen (1973) *J. Biol. Chem.* 248, 3987–3996.
- [104] G. Schneider, Y. Lindqvist, C.-I. Brande & G. Lorimer (1986) *EMBO J.* 5, 3409–3415.
- [105] R. P. Simonsen, P. C. Weber, F. R. Salemme & G. Tollin (1982) *Biochemistry* 21, 6366–6375.
- [106] D. E. Stewart, J. LeGall, I. Moura, J. J. G. Moura, H. D. Peck, A. V. Xavier, P. K. Weiner & J. E. Wampler (1988) *Biochemistry* 27, 2444–2450.
- [107] W. G. J. Hol, P. T. van Duijnen & H. J. C. Berendsen (1978) *Nature* 273, 443–446.
- [108] R. E. Dickerson (1980) *Sci. Am.* 242, 98–110.
- [109] D. C. Carter, K. A. Melis, S. E. O'Donnell, B. A. Burgess, W. F. Furey, Jr., B. C. Wang & C. D. Stout (1985) *J. Mol. Biol.* 184, 279–295.
- [110] Y. Matsuura, T. Takano & R. E. Dickerson (1982) *J. Mol. Biol.* 156, 389–409.

Chapter 2

Experimental Procedures: Isolation and Purification of Proteins

2.1 Parsley Plastocyanin

Parsley plastocyanin was isolated and purified by a modification of the method developed by Plesnicar & Bendall [1]. Ten kg of frozen dry parsley was split into 0.5kg batches. To each 0.5kg of parsley 450ml of deionised water, 50ml of 1M Tris buffer (pH 7.6) and 500ml of cold (-20°C) acetone were added. The mixture was then mashed, blended for two minutes in an ATO mixer, homogenised with a Silverson tissue homogeniser and finally blended in a Kenwood liquidiser for a further 1 minute, producing a thin green suspension. This material was then filtered through four layers of muslin and the resulting green filtrate was centrifuged at 1000g (MSE Coolspin 2) for thirty minutes at 0°C. Supernatants were decanted, combined and cooled, in 1.7 litre aliquots, to 0°C in an ice-salt bath before slowly adding 1.16 volumes (1.92 litres) of cold (-20°C) acetone with vigorous mechanical stirring. Proteins, including plastocyanin, precipitated and were allowed to settle, forming an oily light brown slurry at the bottom of the beaker. Most of the supernatant was decanted and the protein residue was centrifuged at 1000g (MSE Coolspin 2) for twenty minutes. The protein pellets were suspended in 50ml of 0.2M Tris buffer, pH 7.6, and the suspension was dialysed overnight against 0.06M Tris buffer, pH 7.6. Ten kg of parsley was processed over three days. Each day the final dialysate was recovered and centrifuged at 1000g

(Beckman J2-21 centrifuge). The supernatant was decanted and kept at 2°C ready to apply to the first column.

The supernatant, containing plastocyanin, ferredoxin and ferredoxin NADP reductase, was applied to a 5cm × 25cm column of DEAE-cellulose (Whatman DE23) previously equilibrated with 0.06M Tris buffer, pH 7.6. The loaded column was washed with 1 litre of 0.06M Tris buffer to remove most of the yellow protein. Subsequent elution with 0.31M Tris buffer, pH 7.6, produced a flavoprotein as the first eluate, followed by plastocyanin. The latter was identified from the characteristic UV absorption spectrum of its oxidised form.

The crude plastocyanin was applied to a second anion exchange column (2cm × 20cm) which was then washed with 0.02M phosphate buffer (pH 7.3) for six hours at 3ml/minute followed by 0.05M phosphate for four hours. The plastocyanin was obtained by performing a gradient elution from 0.05M to 0.2M phosphate. Potassium ferricyanide was added to the collected fractions and their UV spectra were measured. Those containing plastocyanin were pooled and concentrated using an Amicon 8MC ultrafiltration apparatus fitted with a YM5 membrane. The concentrated plastocyanin was then applied to a gel filtration column (Pharmacia Sephacryl S100 HR, 2.5cm × 2m), equilibrated with 0.01M phosphate/0.1M KCl, pH 7.3. Slow elution separated the blue plastocyanin band from the slower moving colourless and faster moving yellow bands. The purity ratio $A_{278}/\Delta A_{597}$ (where ΔA represents the difference in absorption between oxidised and reduced protein) of the plastocyanin fractions was then checked. Fractions with a ratio of 1.7 or lower were collected, concentrated as described previously, and further purified by fast protein liquid chromatography (FPLC). For this a Mono-Q anion exchange column (HR5/5) was equilibrated in 20mM BTP, pH 7.4. Plastocyanin was eluted using a gradient of 0–500mM KCl at 12mM ml⁻¹. Oxidised and reduced protein eluted at 150mM and 180mM KCl respectively. After elution the gradient was increased to 40mM ml⁻¹. 120mg of protein with a purity ratio of 1.5 was obtained.

2.2 *p*-Cresol Methylhydroxylase

Seven forms of PCMH have been isolated from six bacterial sources [2, 3]. *Pseudomonas putida* N. C. I. B. 9869 contains two forms of the enzyme [4]: this study is concerned with isozyme A of *Pseudomonas putida* (N. C. I. B. 9869), produced from cells grown on 3,5-xyleneol. It is for this form that the abbreviation PCMH will subsequently be used, unless otherwise stated.

The extraction of PCMH, and the resolution of PCMH into its constituent subunits, was carried out in collaboration with Dr. W. S. McIntire in the Molecular Biology Division of the Veterans' Administration Medical Center, San Francisco, California, U. S. A.

Pseudomonas putida N. C. I. B. 9869 was maintained by growth on nutrient-agar slants for 24h at 30°C, followed by storage at 4°C [5, 4]. In order to avoid the heat denaturation step of the original extraction procedure [5, 4], and to improve yield and purity, a modified protocol [2] was adopted.

Cells were harvested in a continuous-flow centrifuge at 4°C, subjected to a freeze-thaw cycle, and suspended in 16mM potassium phosphate buffer, pH 7.0, to give 2.5mg of protein/ml. MgCl₂ and deoxyribonuclease (1mg of each per 120g of wet cell paste) were added. The suspended cells were then disrupted by three passes through a French press. Cell debris was removed by centrifugation at 39000g and 4°C for fifteen minutes. The supernatant was collected, ready for the next centrifugation step, while the pellets were re-suspended in 16mM potassium phosphate buffer and centrifuged at 35000g for one hour. The total supernatant was then centrifuged at 160000g for seventeen hours. The viscous red solution between pellet and supernatant was removed and saved. The supernatant, which contains the bulk of the azurin, was kept for processing at a later date. The pellets were collected in a homogenising vessel and re-suspended, to be centrifuged at 270000g for 4.5 hours. In subsequent purification steps, it was decided to keep separate the supernatant resulting from this step and the initial supernatant. The latter was dialysed twice against a fifteen-fold excess of the same buffer in order to remove molecules and ions arising from the original cell debris.

The solution ($\approx 120\text{ml}$) was applied to a $5\text{cm} \times 50\text{cm}$ column of DEAE-cellulose (Whatman DE53). Fractions were collected immediately as some azurin comes straight through. Elution was carried out with a linear potassium chloride gradient ($0\text{--}0.5\text{M}$) in 16mM potassium phosphate buffer, pH 7.0. Azurin and cytochrome fractions were collected and frozen for future processing. Two bands having enzymic activity were obtained, with the first representing about 20% of the total activity recovered; the bands were kept separate for subsequent processing. In order to remove turbidity prior to concentration, the fractions having enzymic activity were centrifuged at 28000rpm in a TY30 rotor for thirty minutes. These fractions were then pooled, concentrated to $\approx 70\text{ml}$ using a large Amicon unit fitted with a YM30 membrane, and dialysed against 16mM potassium phosphate, pH 7.6. The protein was then applied to a second DEAE column ($2\text{cm} \times 120\text{cm}$), and elution was carried out with a linear potassium chloride gradient ($0\text{--}0.5\text{M}$) in 16mM potassium phosphate, pH 7.6. Fractions possessing enzymic activity were pooled, concentrated and dialysed against 16mM potassium phosphate, pH 7.0.

The solution was then applied to a $1.5\text{cm} \times 35\text{cm}$ column of hydroxylapatite (Bio-Rad, Bio-Gel HTP) and eluted with a linear phosphate gradient of $16\text{--}250\text{mM}$ phosphate, pH 7.0. Fractions having enzymic activity were pooled, concentrated and dialysed against glass-distilled water. The purity ratio A_{280}/A_{412} was then measured. A value of less than 1.5 is acceptable and 1.4 indicates the presence of only trace amounts of impurity.

The reason for the appearance of two peaks with activity on the first DEAE-cellulose column is not known. After purification by identical procedures, the enzymes isolated from the two peaks have essentially the same specific activity and properties [2]. It has been suggested (W. S. McIntire, personal communication) that this phenomenon implies the presence of two species with a significant difference in charge, and that this is more likely to be due to heterogeneity in the flavin subunit of the enzyme than in the cytochrome subunit (cytPCMH), since the latter always produces only one band during isoelectric focusing.

2.2.1 Resolution of PCMH into its Subunits by Isoelectric Focusing

In the protocol used for separation of the flavoprotein and cytochrome subunits, the ampholyte mixture used in the original method [6] was replaced by a small-molecule amphoteric buffer described by Nguyen [7]. Sephadex G-200 Superfine (Pharmacia) was washed with ≈ 40 l of distilled and deionised water to remove low molecular weight contaminants, then dried by very slow addition of anhydrous methanol. One gram of this pre-treated Sephadex was slowly added in small portions to 34.5ml of a solution which was 20mM in each of the following: taurine, MES, glycine, ACES, γ -amino-*n*-butyric acid, TES, HEPES, tricine, bicine, acetic acid and propionic acid. The gel was then left to hydrate overnight.

A portion of the slurry was smoothed onto a glass plate ($7.1 \times 12.8 \times 0.15$ cm) and, using a commercial dryer to blow unheated air across the surface, the slurry was dried to 75% of its original weight.

A Pharmacia FBE 3000 flat-bed apparatus was then set up by pipetting a layer of mineral oil onto the bed. On this was placed a glass plate, followed by a second layer of mineral oil and then the focusing plate. Excess mineral oil was removed, and two $2.5\text{mm} \times 6.5\text{mm}$ Pharmacia electrode strips were cut to the width of the gel on the plate, moistened and placed at either end of the plate. Several drops of 40mM histidine were pipetted on to the cathodic wick and 40mM glutamic acid on to the anodic wick.

A concentrated solution of the native enzyme (typically $200\mu\text{l}$ of a 20mg/ml solution) was centrifuged by Eppendorf and then pipetted onto the gel 1–1.5cm from the cathode. The initial focusing power was set to 10W and was reduced if the apparatus began to overheat (indicated by the formation of condensation on the lid).

The cytochrome was sharply focused after ≈ 2.5 hours, and the final gradient was approximately linear in the range pH 3.0–6.0. The gel containing the cytochrome was scooped off the plate and washed with 50mM sodium acetate, pH 5.0. The gel and protein were separated by repeated centrifugation until the gel pellet was colourless. Centricon-10 tubes (Amicon) were then used to wash the cytochrome twice in the sodium acetate buffer, followed by several washes in 50mM potassium phosphate, pH

2.3 Preparation of Protein Samples for NMR Experiments

Plastocyanin samples were prepared by exchanging the protein solution into 20mM phosphate solution in a miniature diafiltration cell (Amicon 8MC with microvolume accessory) fitted with a YM5 membrane. Fully reduced protein was obtained by adding a slight excess of sodium ascorbate prior to diafiltration. When fully exchanged into the appropriate buffer, 0.5ml of argon-saturated (BOC oxygen-free white spot argon) plastocyanin solution was transferred to a sealed NMR tube which had previously been flushed with argon gas. This procedure was adopted as a precaution against atmospheric oxidation. pH adjustments were carried out on the buffer only. Samples prepared in this way remained reduced for several months. For D₂O solutions, pH values are reported as direct meter readings pH*, uncorrected for the deuterium isotope effect.

Since cytPCMH oxidises more readily than plastocyanin, NMR samples of the former were prepared under more rigorous conditions of oxygen exclusion using standard Schlenk-line techniques. The protein was exchanged into the appropriate buffer in the same way as plastocyanin. Prior to transfer of the protein solution, the NMR tube was sealed with PTFE tape and flushed with argon (BOC oxygen-free white spot) for ≈ 20 minutes. 0.495ml of argon-saturated cytPCMH solution was then introduced with a Scientific Glass Engineering (SGE) gas-tight syringe which had previously been flushed with argon. While still passing argon into the NMR tube, the dithionite (reductant) solution was flushed with argon. Again, an SGE gas-tight syringe which had previously been flushed with argon was used to add 0.005ml of this dithionite solution to the protein in the NMR tube, giving a slight excess of reductant over cytPCMH. Finally, the tube was further flushed for ≈ 20 minutes before it was sealed with more PTFE tape.

Phosphate buffer was prepared at 100mM concentration in 99.8% D₂O, then freeze-dried at least three times from aliquots of D₂O to exchange all protons for

deuterons. Buffer solutions for D₂O NMR experiments were prepared by diluting the stock phosphate solution with 99.8% D₂O. Experiments for the study of labile protons were carried out in a buffer solution consisting of the phosphate stock diluted with H₂O to give a ratio of 90% H₂O/10% D₂O.

Bibliography

- [1] M. Plesnicar & D. S. Bendall (1970) *Biochim. Biophys. Acta* 216, 192–199.
- [2] S. C. Koerber, W. McIntire, C. Bohmont & T. P. Singer (1985) *Biochemistry* 24, 5276–5280.
- [3] I. D. Bossert, G. Whited, D. T. Gibson & L. Y. Young (1989) *J. Bacteriol.* 171, 2956–2962.
- [4] M. J. Keat & D. J. Hopper (1978) *Biochem. J.* 175, 649–658.
- [5] D. J. Hopper & D. G. Taylor (1977) *Biochem. J.* 167, 155–162.
- [6] W. McIntire & T. P. Singer (1982) *FEBS Lett.* 143, 316–318.
- [7] N. Y. Nguyen & A. Chrambach (1976) *Anal. Biochem.* 74, 145–153.

Chapter 3

NMR Spectroscopy and its Application to Protein Structure Determination

3.1 NMR Spectroscopy and X-Ray Crystallography: Complementary Techniques for Study of Protein Structure

NMR spectroscopy has become a viable alternative to X-ray crystallography for the determination of the detailed structure of small proteins. Striking advances in instrumentation and methodology have been responsible for this progression. The major development in instrumentation has been that of high field spectrometers, which employ superconducting magnets and which have high sensitivity. The introduction of two dimensional (2D) techniques [1, 33] has been the seminal advance in NMR methodology: the recent development of three dimensional (3D) [3, 4, 5, 6, 7] and four dimensional (4D) [8, 7] experiments is, conceptually, a corollary of Jeener's original proposal [1]. This is not to diminish the efforts of those involved in the development of 3D and 4D techniques: such experiments require sophisticated pulse sequences and are more demanding of instrument and operator. Other important advances have been the conception of a semi-systematic assignment procedure for proteins [9], and algorithms for structure determination which use NMR-derived distance and angle constraints. These include distance geometry [10, 11, 12] and restrained molecular dynamics [13, 14, 15, 16].

NMR and X-ray diffraction methods are, in fact, complementary techniques and should ideally be used in conjunction with one another. Their complementarity arises from their significantly different timescales, and from the different media in which the respective measurements are made: NMR spectroscopy can be used to study proteins in solution, whereas diffraction methods are restricted to the solid state and in particular single crystals. Solution conditions can be varied so that conformational changes due to changes in pH, temperature or ionic strength can be monitored, or so that physiological conditions are simulated. NMR furnishes quantitative and semiquantitative information on internal dynamics [17], whilst crystallography provides an indication of the conformation space covered by static disorder and high frequency motions.

Like any other technique, the application of NMR to the study of proteins in solution has limitations. It requires that the protein should be soluble and not aggregate up to 1mM. The dependence of linewidth on the rotational correlation time τ_c sets an upper limit on the molecular weight of proteins that can be studied. For molecules of up to about one hundred and thirty residues, homonuclear 2D techniques may be sufficient to determine the structure. For larger proteins, spectral overlap and low efficiency of magnetisation transfer through small ^1H - ^1H three bond couplings become insurmountable problems in homonuclear 2D spectra. Using the most sophisticated 3D and 4D heteronuclear experiments currently available, structures may be determined for molecules of up to two hundred amino acid residues.

Among the advantages of NMR for the study of proteins are that the protein does not have to be crystallised, and that dynamic processes can be studied over a wide timescale range: from picosecond to second.

This chapter will cover some of the basic theory of the 2D ^1H NMR techniques commonly applied to proteins in solution. More extensive treatment of the principles of NMR can be found in numerous texts and review articles, for example [18, 19, 20, 21, 22, 23, 24]. The following most closely parallels the review by Kessler *et al.* [22]. Details of the actual experiments carried out will be given. In the next chapter, the nature of the data furnished by those experiments will be described, as will the methodology used to determine protein structure from such NMR-derived data.

3.2 One Dimensional NMR Spectroscopy

This will be discussed briefly since the same principles also apply to 2D NMR. The macroscopic magnetisation of a system of nuclear spins $1/2$ can be expressed in terms of the Bloch equations [25]. The effect of applying a short radiofrequency pulse, β_y , orthogonal to the static magnetisation, M_z , can be expressed as:

$$M_z \xrightarrow{\beta_y} M_z \cos\beta + M_x \sin\beta \quad (3.1)$$

The transverse component of the resulting magnetisation, M_x , precesses about the z axis at a frequency $\omega = \gamma\beta_{eff}$. In the rotating frame coordinate system, the observer (detector) is also rotating about the z axis at the ‘carrier frequency’ ω_0 of the pulse, such that the detector sees only the difference frequencies $\omega - \omega_0$. These are termed precession frequencies or chemical shifts:

$$M_x \xrightarrow{\Omega t} M_x \cos\Omega t + M_y \sin\Omega t \quad (3.2)$$

A signal oscillating at frequency Ω is obtained. This is the *free induction decay* (FID). Fourier transformation converts the FID into a spectrum with a single peak.

3.2.1 Quadrature Detection

The function $\cos\Omega t$ in equation 3.2 contains the frequencies $\omega = +\Omega$ and $\omega = -\Omega$. Quadrature detection [26, 27] permits distinction of these frequencies so that the transmitter frequency can be set in the centre of the spectrum (with resulting increase in sensitivity and in efficiency of use of the radiofrequency field). This scheme involves two orthogonally positioned detectors, such that a signal $\cos\Omega t$ is constructed in the x channel (K_x) from M_x in equation 3.2, and a signal $\sin\Omega t$ in the y channel from M_y in equation 3.2. The addition $K_x + iK_y$ is carried out to yield $e^{i\Omega t}$, which, after complex FT, gives a single line at frequency $+\Omega$.

A single channel spectrometer can be made to behave as a quadrature detection system using the scheme of Redfield [28].

$E/2$	Where E is the unity operator
I_{1x}, I_{2x}	In-phase x -magnetisation; SQC
I_{1y}, I_{2y}	In-phase y -magnetisation; SQC
I_{1z}, I_{2z}	Longitudinal magnetisation
$2I_{1x}I_{2z}, 2I_{1y}I_{2z}$	Transverse magnetisation of 1 antiphase with respect to 2; SQC
$2I_{1z}I_{2x}, 2I_{1z}I_{2y}$	Transverse magnetisation of 2 antiphase with respect to 1; SQC
$2I_{1x}I_{2x}, 2I_{1y}I_{2y}, 2I_{1x}I_{2y}, 2I_{1y}I_{2x}$	Two-spin coherence of 1 and 2; Superposition of ZQC and DQC
$2I_{1z}I_{2z}$	Longitudinal two-spin order of 1 and 2

Table 3.1: The sixteen product operators for two weakly coupled spins 1 and 2. ZQC, SQC and DQC are zero, single and double quantum coherence respectively.

3.3 Product Operator Formalism

The general theory to describe the response of a system to an arbitrary pulse sequence is the rather cumbersome density matrix approach [29, 30]. A more pictorial approach [31] is based on the decomposition of the density matrix into a linear combination of products of spin angular momentum operators [32].

The sixteen product operators for two weakly coupled spins, 1 and 2, are given in Table 3.1.

The factors of $1/2$ and 2 are normalisation constants to ensure that all the product operators give $1/4E$ when multiplied by themselves. Antiphase magnetisation represents multiplets with individual components that have opposite phases. For instance, $2I_{1x}I_{2z}$ corresponds to the x -component of the 1-spin magnetisation split into $+x$ and $-x$ components depending on the polarisation of spin 2. An antiphase multiplet has zero integrated intensity. Longitudinal spin order is a specific disturbance of populations, having no net polarisation and no observable magnetisation.

Using this formalism, the evolution of magnetisation during a simple 1D experiment is given by:

$$I_z \xrightarrow{90_y^{\circ} \text{ pulse}} I_x \xrightarrow{\text{Evolution during detection}} I_x \cos \Omega t_2 + I_y \sin \Omega t_2 \quad (3.3)$$

The state I_x produced by the pulse is termed the *coherence*. Coherences describe transitions between energy levels, and are characterised by the coherence order p ,

which is the change in the quantum number m_z . Only those transitions (coherences) for which $\Delta m_z = \pm 1$ are associated with directly observable magnetisations, and of these SQCs, only those in which the state of one spin is changed are detectable.

3.4 2D NMR Spectroscopy

The concept of multidimensional NMR was first proposed by Jeener [1], but appears to have been neglected until the seminal paper of Ernst and co-workers [33]. Any 2D NMR experiment can be conveniently divided into four stages: a preparation pulse, evolution, mixing and detection. During evolution, the spins are labelled according to their chemical shifts. During mixing, the spins are correlated with one another. The second dimension is introduced by collecting a large number of FIDs, each corresponding to a different value of the evolution time t_1 , which is successively linearly incremented. A signal $s(t_1, t_2)$ is obtained. Information about the frequencies with which the magnetisation evolved during t_1 are encoded into the amplitude observed during t_2 . The encoding is unravelled by a 2D FT. The time constant of the ‘memory’ of nuclear spins is the spin-lattice relaxation time T_2 .

Generally, the Fourier transforms in each dimension are carried out separately, that in t_2 being first:

$$t_2 \xrightarrow{FT} F_2 \quad (3.4)$$

This gives n 1D spectra with intensities modulated as a function of the t_1 duration. Data points corresponding to a particular F_2 frequency are then selected from spectra for $t_1 = 0$, $t_1 = \Delta_1$, $t_1 = 2\Delta_1$, etc. giving an interferogram $s(t_1)$. FT of these interferograms produces a spectrum as a function of two frequencies.

A 2D spectrum may be viewed as a representation of the transfer processes of coherences. In high resolution NMR, there are essentially two spin-spin interactions which can be used to effect coherence transfer: through-bond scalar spin-spin coupling (J coupling) and through-space dipolar coupling. The former is utilised in COSY and its variants, the latter in NOESY-type experiments. Coherence transfer can also occur by exchange processes, but these do not involve multispin systems.

3.4.1 Scalar Coupling in the Product Operator Formalism

Scalar coupling between two nuclei 1 and 2 may be described by the following transformations:

$$I_{1x} \xrightarrow{\pi J_{12} t 2I_{1z} I_{2z}} I_{1x} \cos \pi J_{12} t + I_{1y} I_{2z} \sin \pi J_{12} t \quad (3.5)$$

$$I_{1y} \xrightarrow{\pi J_{12} t 2I_{1z} I_{2z}} I_{1y} \cos \pi J_{12} t - I_{1x} I_{2z} \sin \pi J_{12} t \quad (3.6)$$

$$I_{1x} I_{2z} \xrightarrow{\pi J_{12} t 2I_{1z} I_{2z}} 2I_{1x} I_{2z} \cos \pi J_{12} t + I_{1y} \sin \pi J_{12} t \quad (3.7)$$

$$I_{1y} I_{2z} \xrightarrow{\pi J_{12} t 2I_{1z} I_{2z}} 2I_{1y} I_{2z} \cos \pi J_{12} t - I_{1x} \sin \pi J_{12} t \quad (3.8)$$

3.4.2 COSY

Consider a system of three spins, 1, 2 and 3, at equilibrium. The first 90_y° pulse (Figure 3.1) generates $I_{1x} + I_{2x} + I_{3x}$. Attention will be focused on I_{2x} , since the transfer $I_{1x} \rightarrow I_{2x}$ gives a cross peak at Ω_1 in ω_1 and Ω_2 in ω_2 . (Terms which do not lead to signals are ignored). Spin-spin coupling occurs during t_1 :

$$I_{1x} \longrightarrow 2I_{1x} I_{2z} \quad (3.9)$$

The full series of transformations for the 1,2 cross signal is given by:

$$I_{1z} \xrightarrow{90_y^\circ} I_x \xrightarrow{t_1} -2I_{1x} I_{2z} \xrightarrow{90_y^\circ} 2I_{1z} I_{2x} \xrightarrow{t_2} I_{2x} \quad (3.10)$$

There is no other way to transfer magnetisation from spin 1 to spin 2 in this pulse sequence. Note that antiphase magnetisation exists before and after the coherence transfer, caused by J_{12} coupling. The J_{12} coupling is termed *active coupling*, whereas couplings to other nuclei such as spin 3 are described as *passive* and appear in-phase. The full term after evolution under chemical shift and coupling during t_1 is:

$$-2I_{1x} I_{2z} \sin \Omega_1 t_1 \sin \pi J_{12} t_1 \cos \pi J_{13} t_1 \quad (3.11)$$

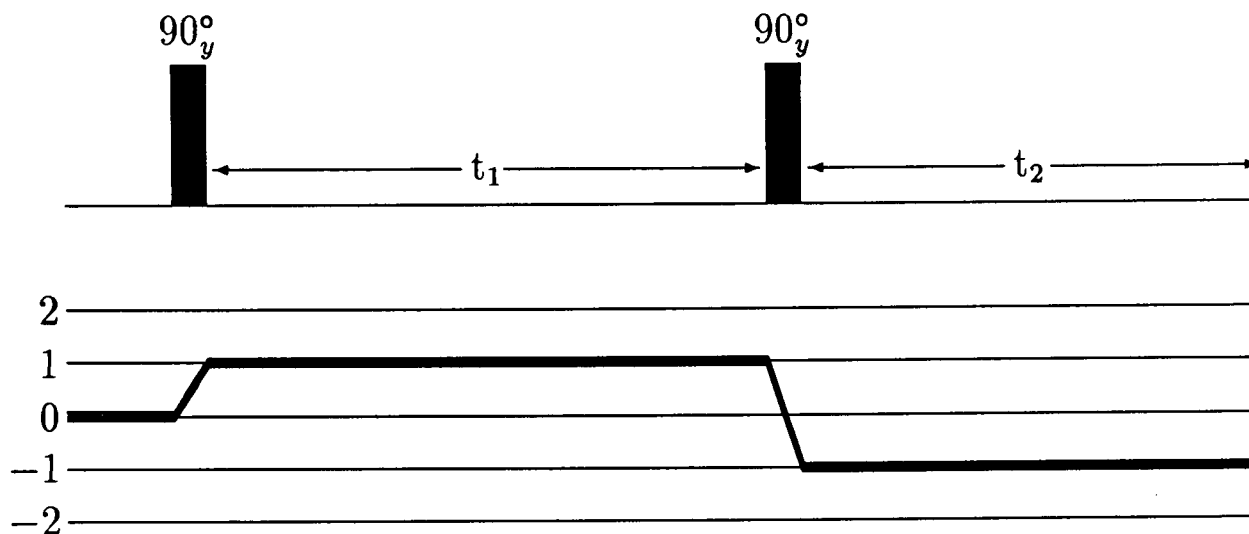


Figure 3.1: Pulse sequence and coherence transfer pathway (see 3.4.5) of the basic COSY experiment. The coherence transfer pathway is represented in the form of a ‘staff’. This example shows positive SQC during t_1 and negative SQC during t_2 . Horizontal lines indicate conservation of coherence order during free evolution.

Polarisation transfer occurs through the second 90_y pulse, giving

$$2I_{1z}I_{2x}\sin\Omega_1t_1\sin\pi J_{12}t_1\cos\pi J_{13}t_1 \quad (3.12)$$

which evolves into observable magnetisation according to equations 3.5 and 3.7. Other cross peaks are obtained similarly, for example for spins 2 and 3 by starting with I_{2x} after a 90° pulse and detecting I_{3x} .

3.4.3 Dipolar Coupling

Dipolar coupling [30, 34, 35, 36] is the interaction of the magnetic moments of two spins through space. It is the major mechanism of spin-lattice relaxation in liquids for $I = 1/2$ nuclei. With respect to coherence transfer, it is different from scalar coupling in that scalar coupling is operative during the evolution period, whereas dipolar coupling must be effective during the mixing period:

$$I_{1x} \xrightarrow{t_1} I_{1z} \xrightarrow{NOE} I_{2z} \xrightarrow{t_2} I_{2x} \quad (3.13)$$

Dipolar coupling is caused by the rapid reorientation of molecules in solution, which produces oscillating magnetic fields at neighbouring nuclei. The magnetic field B_D^1

which is generated at spin 1 by spin 2 is given by:

$$B_D^1 = \gamma_2(3\cos^2\Theta_{12} - 1)r_{12}^{-3} \quad (3.14)$$

where r_{12} is the internuclear distance and Θ_{12} is the angle between the internuclear vector and the static field.

These oscillating fields induce exchange of polarisation (z magnetisation) and cause relaxation of nearby spin systems which have non-Boltzmann population distributions back to equilibrium states. This latter process of cross-relaxation is responsible for the nuclear Overhauser effect (NOE), which can be defined as the change in magnetisation of proton 1 upon perturbation of the magnetisation of proton 2.

When molecular reorientation is slow, such that $\tau_{eff} > \omega_0^{-1}$, cross-relaxation occurs mainly through W_0 transitions. These are ones in which the two spin system changes between $\alpha\beta$ and $\beta\alpha$ states. For these transitions, the change in the quantum number M_z (which is the sum of the individual quantum numbers for the two nuclei) is 0. Cross-relaxation occurs through W_2 processes only when reorientation is fast. W_2 transitions are ones in which the two spin system moves between $\alpha\alpha$ and $\beta\beta$ states, and which involve a change in M_z of 2. $W_2 - W_0$ may be positive or negative. By convention, a positive cross-relaxation rate and increased intensity of the 2-spin signal correspond to a positive NOE.

3.4.4 NOESY

The NOESY experiment [37] (Figure 3.2) is the most important for incoherent magnetisation transfer. It requires longitudinal magnetisation before and after mixing:

$$I_{1z} \xrightarrow{90_x^\circ} -I_{1y} \xrightarrow{t_1} -I_{1y} \xrightarrow{90_x^\circ} I_{1z} \xrightarrow{NOE} -I_{2z} \xrightarrow{90_x^\circ} I_{2y} \xrightarrow{t_2} -I_{2x} \quad (3.15)$$

If $W_2 - W_0 > 0$, I_{1z} evolves into $-I_{2z}$, so that the diagonal and cross signals have opposite sign. If $W_2 - W_0 < 0$, I_{1z} evolves into I_{2z} and the diagonal and cross signals have the same sign.

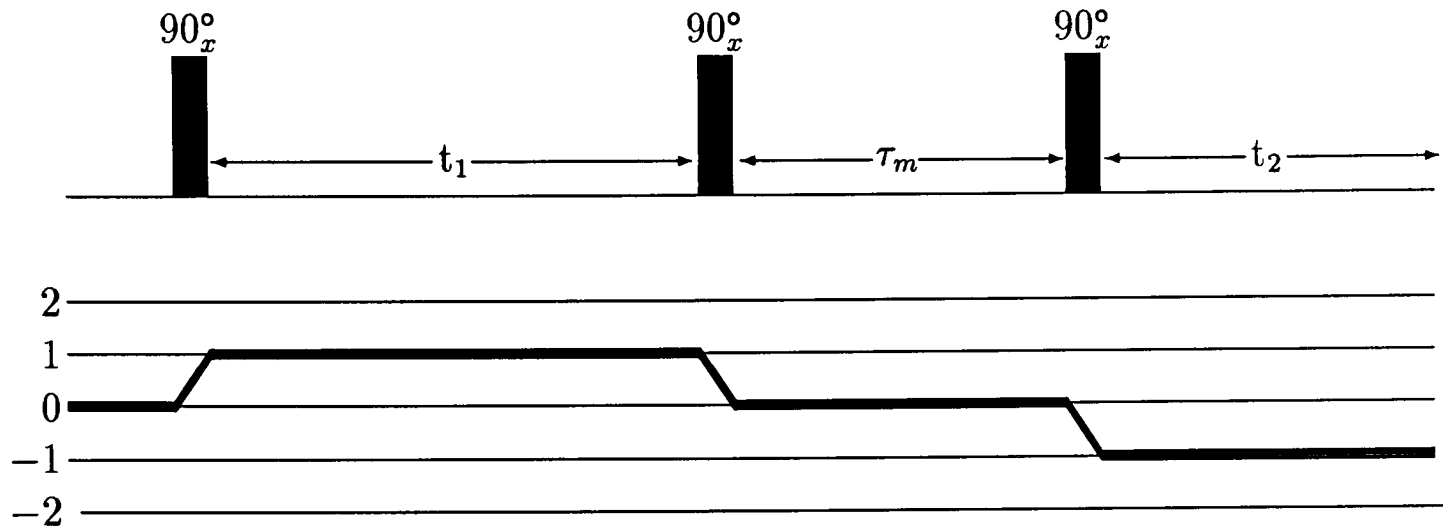


Figure 3.2: Pulse sequence and coherence transfer pathway of the NOESY experiment.

3.4.5 Coherence Orders and Phase Cycles

In a phase cycle, a pulse sequence is repeated, systematically varying the phases of the single pulses (transmitter) and of the receiver, keeping all other parameters constant. All FIDs obtained for one t_1 increment are added. The purpose of a phase cycle is to suppress undesired signal components, while the signals of interest are improved by time averaging. The principle can be illustrated by the simple example of suppression of axial peaks (common artifacts in most 2D experiments) in the COSY experiment. Such peaks arise when longitudinal magnetisation is transformed into detectable transverse magnetisation during the mixing sequence. These signals are not modulated during t_1 and therefore resonate at $\omega_1 = 0$. Axial signals transform as:

$$I_{1z} \xrightarrow{90_y^\circ} I_{1x} \quad (3.16)$$

$$I_{1z} \xrightarrow{90_{-y}^\circ} -I_{1x} \quad (3.17)$$

If equations 3.16 and 3.17 are compared with the transfers responsible for the cross peaks in COSY spectra:

$$2I_{1x}I_{2z} \xrightarrow{90_y^\circ} -2I_{1z}I_{2x} \quad (3.18)$$

$$2I_{1x}I_{2z} \xrightarrow{90_{-y}^\circ} -2I_{1z}I_{2x} \quad (3.19)$$

it can be seen that axial signals are suppressed by addition of the two experiments $90_y^\circ - t_1 - 90_y^\circ - t_2$ and $90_y^\circ - t_1 - 90_{-y}^\circ - t_2$.

In order to design a phase cycle, it is necessary to consider the *coherence transfer pathway* of the experiment. The number of transverse operators n determines the highest ($p = +n$) and lowest ($p = -n$) coherence orders present in a product operator. All values from $p = -n$ to $p = +n$ in increments of two are valid coherence orders. It is helpful to display the coherence orders during the experiment in the form of a ‘staff’ [38], as exemplified in Figure 3.1.

Pulses generate jumps in the coherence orders, exciting all the coherence orders between $p = +n$ and $p = -n$. A phase cycle allows only the coherence order jumps $\Delta p \pm kN$, where N is the selectivity of the phase cycle, Δp is the coherence order jump and $k=0, 1, 2$, etc. There must be a compromise between selectivity and time requirements, since increasing N increases the necessary number of accumulations for each t_1 increment. N is related to the phase increment $\Delta\varphi$: $\Delta\varphi = 2\pi/N$. If the pulse phase is increased by $\Delta\varphi$, the receiver phase must also be increased by $-\Delta p\Delta\varphi$. Desired pathways are followed by the receiver and unwanted pathways are turned off. The general formula for the phase cycle of the i -th pulse is:

$$\varphi_{det} = - \sum_i \Delta p_i \Delta\varphi_i \quad (3.20)$$

It is essential to know two further facts in order to design a phase cycle:

1. The detector conventionally selects negative single quantum coherences.
2. The first pulse can create only single quantum coherence from the thermal equilibrium state.

Generally, pulse and phase imperfections are best averaged out by using a large phase cycle. A common example, used in both 1D and 2D spectroscopy to reduce quadrature images and the quadrature glitch, is the sixteen step CYCLOPS (Cyclically Ordered Phase Sequence) introduced by Hoult and Richards [39].

3.4.6 Phase Sensitive 2D NMR Spectra

Phase sensitive spectra require amplitude modulation of the signal in t_1 , as is characteristic of those 2D experiments (for example, COSY, NOESY, and HOHAHA) which involve polarisation transfer from one spin to another. Amplitude modulation can be represented as two equal counter-rotating phase modulations, such that the sense of precession during t_1 is ambiguous. This could be overcome by placing the transmitter at one end of the spectrum, so that all resonances have offsets of the same sign. However, this wastes data storage.

An alternative is to impose a phase modulation, rather than amplitude modulation, during t_1 . This produces a phase-twisted lineshape, one in which the phase rotates from pure dispersion through pure absorption to dispersion in the other sense as successive parallel sections are viewed. This could be avoided by plotting the absolute value, which is given by $\sqrt{(abs^2 + disp^2)}$. Overlapping of cross peaks could then be overcome by multiplication by a weighting function such as a sine-bell. This would, however, incur a severe sensitivity penalty.

Overall, then, it is desirable to achieve sign discrimination in F_1 while retaining a pure absorption mode lineshape. This can be accomplished in two ways. The first to be discussed was developed by States and co-workers [40].

Consider a signal which is amplitude modulated in t_1 :

$$C(t_1, t_2) = \cos\Omega_1 t_1 (\cos\Omega_2 t_2 + i \sin\Omega_2 t_2) E_1 E_2 \quad (3.21)$$

where $E_1 = \exp(-t_1/T_2)$ and $E_2 = \exp(-t_2/T_2)$ are the decay constants. The amplitude modulated signal in t_1 can be represented as two counter-rotating phase modulations $\cos\Omega_1 t_1 + i \sin\Omega_1 t_1$ and $\cos\Omega_1 t_1 - i \sin\Omega_1 t_1$. Complex FT in t_2 (quadrature detection) produces an absorption mode and a dispersion mode spectrum:

$$C_a(t_1, F_2) = \cos(\Omega_1 t_1) A_2 E_1 \quad (3.22)$$

$$C_d(t_1, F_2) = \sin(\Omega_1 t_1) D_2 E_1 \quad (3.23)$$

where A_2 and D_2 represent Lorentzian absorption and dispersion mode signals respectively. For the second transform, cosine and sine transforms can be defined:

$$C^{cos}(F_1, F_2) = \int C(t_1, F_2) \cos(2\pi F_1 t_1) dt_1 \quad (3.24)$$

$$C^{sin}(F_1, F_2) = \int C(t_1, F_2) \sin(2\pi F_1 t_1) dt_1 \quad (3.25)$$

such that the cosine and sine transforms of equations 3.22 and 3.23 can be written as follows to emphasise the lack of sign discrimination in the F_1 dimension:

$$C_a^{cos}(F_1, F_2) = A_1^+ A_2 + A_1^- A_2 \quad (3.26)$$

$$C_d^{cos}(F_1, F_2) = A_1^+ D_2 + A_1^- D_2 \quad (3.27)$$

$$C_a^{sin}(F_1, F_2) = D_1^+ A_2 + D_1^- A_2 \quad (3.28)$$

$$C_d^{sin}(F_1, F_2) = D_1^+ D_2 + D_1^- D_2 \quad (3.29)$$

The extra information needed to determine unambiguously the phase angle of a vector precessing in t_1 is provided by a second experiment in which the modulation during t_1 follows a sine rather than a cosine wave:

$$S(t_1, t_2) = \sin\Omega_1 t_1 (\cos\Omega_2 t_2 + i \sin\Omega_2 t_2) E_1 E_2 \quad (3.30)$$

For example, this would be achieved in a COSY experiment by a phase shift of the initial pulse by 90° . As before, quadrature detection in t_2 produces an absorption and a dispersion mode spectrum:

$$S_a(t_1, F_2) = \sin(\Omega_1 t_1) A_2 E_1 \quad (3.31)$$

$$S_d(t_1, F_2) = \sin(\Omega_1 t_1) D_2 E_1 \quad (3.32)$$

The trick of the States method is to discard both the dispersion mode spectrum, equation 3.32, and S_a^{cos} , leaving only

$$S_a^{sin}(F_1, F_2) = A_1^+ A_2 - A_1^- A_2 \quad (3.33)$$

in which the minus sign reflects the fact that $\sin\Omega_1 t_1$ reverses sign for negative frequencies. Combination with equation 3.26 from the first experiment yields

$$C_a^{cos}(F_1, F_2) + S_a^{sin}(F_1, F_2) = 2A_1^+ A_2 \quad (3.34)$$

This is a signal with positive frequency in t_1 , the component at negative frequency having been cancelled.

The second method was developed [41] for spectrometers without a complex FT facility. It involves application in t_1 of the Redfield scheme [28] already referred to in the section on quadrature detection. The sampling rate is doubled (the t_1 increment is halved) and the receiver phase is shifted by 90° after each t_1 increment. This method is known as time proportional phase incrementation (TPPI), and is the one used on Bruker spectrometers. It has the effect of increasing the apparent frequencies in F_1 by $\Omega_C = (SW/2)$, as if the reference frame were rotating at $\Omega_0 - \Omega_C$ instead of Ω_0 . For a negative coherence order $-n$, the frequency shift would be in the opposite sense, so that F_1 spectra with opposite signs of n are separated. Both the States method and TPPI require twice as many data points as for a non-phase sensitive spectrum of the same resolution (and equal measuring time).

3.5 Survey of Relevant 2D NMR Techniques

3.5.1 COSY and Variants

DQF COSY

The basic COSY experiment has been discussed in 3.4.2. One of the problems in applying either the hypercomplex FT approach [40] or TPPI [41] to COSY is that the diagonal and cross peaks have different phases, making it impossible to simultaneously phase all peaks to absorption mode. As a result, cross signals near the diagonal are often distorted and difficult to observe due to dispersive extensions from diagonal signals. This is overcome using the pulse sequence shown in Figure 3.3, which is that of an MQF COSY experiment [42, 43, 44, 45].

In the COSY technique, the product operator responsible for diagonal signals contains before the mixing pulse a single operator (I_y), while that for cross signals contains two operators ($2I_{1x}I_{2z}$). A 90_x° pulse generates DQ excitation of the latter term, giving $-2I_{1x}I_{2y}$ which is selected by phase cycling. A 90_y° pulse converts this back to SQC. In this way, single spin terms are filtered, only two operator terms contribute, and the desired transfer, equation 3.35, is selected.

$$2I_{1x}I_{2z} \rightarrow 1/2(2I_{1z}I_{2y}) \quad (3.35)$$

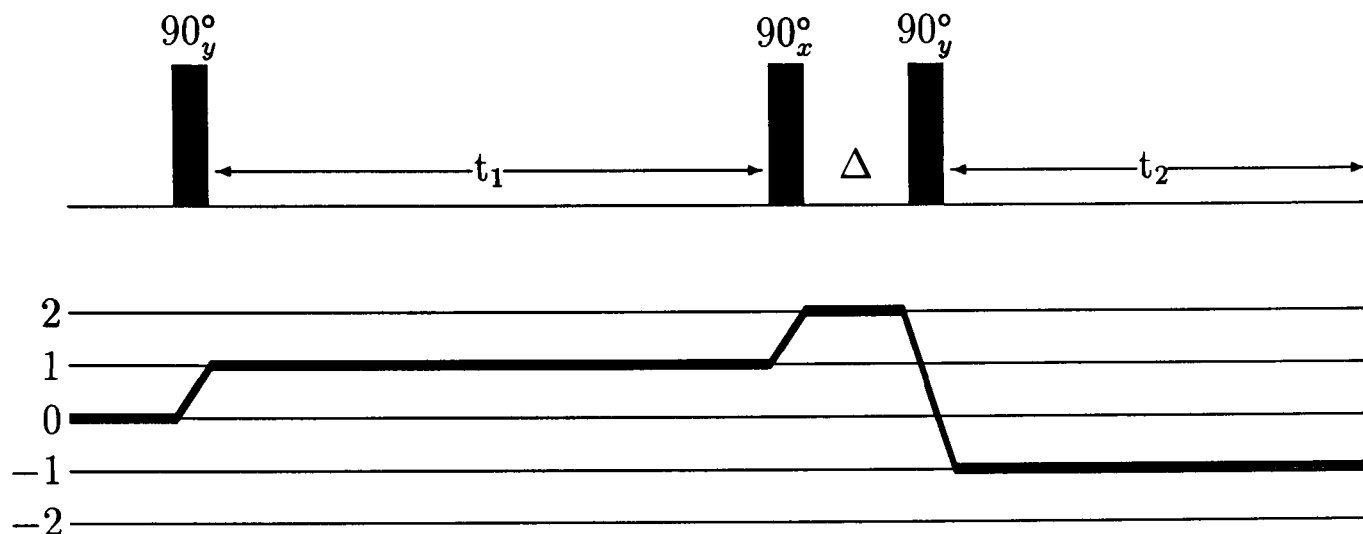


Figure 3.3: Pulse sequence of a COSY experiment with a multiple quantum filter. The order of the filter is determined by the phase cycle.

The factor $1/2$ indicates that the transfer is half as efficient as in a standard COSY. Occurring at the same time is the transfer 3.36 which gives rise to the diagonal peaks. The latter are now antiphase and absorptive, such that their net integrated intensity is zero and they contain no dispersive components.

$$2I_{1x}I_{2z} \rightarrow 1/2(2I_{1y}I_{2z}) \quad (3.36)$$

Coupling constants can be obtained from the multiplet structure of COSY cross peaks, as described in the following chapter.

E.COSY

E.COSY [46, 47] cross peaks have relatively simple multiplet structure, facilitating determination of coupling constants and automated assignment of spectra by the pattern recognition technique [48, 49]. The E.COSY spectrum is obtained, theoretically, by weighted addition of a DQF COSY and a TQF COSY spectrum. The principle of the latter is the same as the former: the pulse sequence is the same, only the phase cycle is different. Cross peaks are generated by the transfer $I_{1x}I_{2z}I_{3z} \rightarrow I_{1z}I_{2x}I_{3z}$. The sensitivity of the TQF COSY technique is a factor four lower than unfiltered COSY.

Data acquisition in the E.COSY experiment is complex, requiring linear combination of several phase-shifted data sets. In principle, E.COSY-type spectra can be obtained using the much simpler β -COSY technique [50], which is a regular COSY with a small flip angle (β) mixing pulse. Magnetisation transfer is then restricted to occur between connected transitions. However, diagonal and cross signals cannot be properly phased simultaneously as both are out of phase in both frequency dimensions.

P.E.COSY

P.E.COSY [51, 46, 52] provides a simple alternative to the powerful but complex E.COSY outlined above. The aim is to modify the COSY experiment such that both diagonal and cross peaks are in absorption mode. This is accomplished by subtracting a reference signal $S_{ref}(t_1, t_2)$ recorded without a mixing pulse from the 2D FID $S(t_1, t_2)$ recorded with a small flip angle mixing pulse (Figure 3.4). Thus the signal that is in phase with the mixing pulse β is eliminated, so that all remaining signal is 90° out of phase with β . The whole spectrum can then be phased to pure absorption mode.

A mixing pulse of about 35° gives a suppression of nonconnected transitions of at least 10:1 [33], but incurs a 40% loss in sensitivity. Recording the reference signal, which contains no connectivity information, causes a further 33% loss in sensitivity.

P.COSY

P.COSY [53] involves a variation of the P.E.COSY technique and is used to purge the dispersive character of the diagonal in the standard COSY experiment. Instead of recording a 2D matrix for the reference signal, as is done in P.E.COSY, a similar result can be obtained from one FID [54]. By left-shifting the data of this FID, the time-domain data for successive t_1 increments of the reference experiment are obtained. The single FID is usually recorded with N times as many scans ($N = 8$ or 16) as the 2D experiment, so that the sensitivity of the difference spectrum is reduced by only $\sqrt{(N^2 + 1)}/N$.

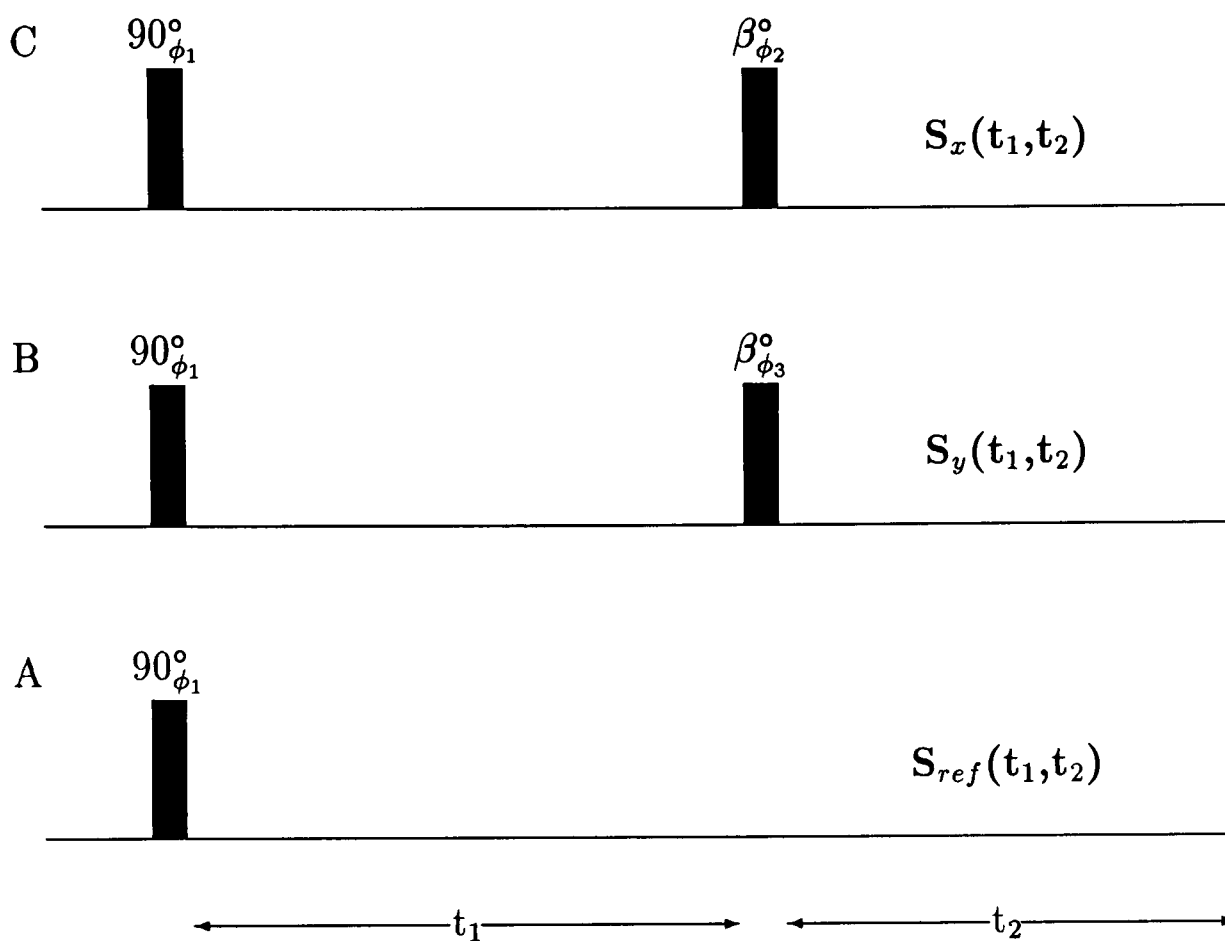


Figure 3.4: Pulse sequence used to generate P.E.COSY spectra.

This methodology can be applied to regular COSY experiments, as well as to small flip angle variants. Indeed, the P.COSY technique provides a more sensitive alternative to DQF COSY. The reason for its greater sensitivity can be understood by considering the DQF COSY pulse sequence:

$$90^\circ_\phi - t_1 - 90^\circ_x 90^\circ_\psi - Acq(t_2) \quad (3.37)$$

Filtering is achieved by incrementing the phase ψ by 90° in four successive experiments. Two of these ($\psi = x, -x$, corresponding to mixing pulses of 180° and 0°), do not contribute signal to the cross peaks [55] but add noise and subtract the dispersive character from the diagonal.

Other differences from DQF COSY are that P.COSY does not remove singlet signals from the diagonal (this makes little difference), and that phasing is simpler in P.COSY since the integrated intensity of the diagonal is non-zero. This means that F_2 phasing can be carried out using the first t_1 increment, whereas a reference spectrum is needed for F_2 phasing in DQF COSY.

3.5.2 TOCSY or HOHAHA

Use of this type of experiment for proton correlation was first proposed by Braunschweiler & Ernst [56], and was ‘rediscovered’ through the analysis of artifacts [57] in ROESY [58] spectra which were shown to be due to the homonuclear Hartmann-Hahn (HOHAHA) effect. HOHAHA spectroscopy is now the most useful 2D technique for spectral assignment of proteins since it permits correlation of all protons in a scalar coupling network, so that complete amino acid spin systems can be delineated. Adjustment of the mixing time allows direct, single, double and multiple relayed connectivities to be successively obtained. There have been several theoretical discussions of this experiment [60, 61, 62, 63, 64, 65].

The original Hartmann-Hahn experiment [59] was applied to the solid state and involved the detection of the response of low abundance or low magnetogyric ratio nuclei S surrounded by abundant spins I with a strong NMR response. The S spins are irradiated at resonance with a field B_2 , generating z magnetisation which oscillates at $\gamma_S B_2 / 2\pi$ Hz. Contact between the two types of spins occurs by dipole-dipole coupling, such that the oscillating magnetic field induces I spin transitions if it matches the precession frequency of the I spins. This Hartmann-Hahn matching is achieved by applying a field B_1 to the I spins such that

$$\gamma_S B_2 / 2\pi = \gamma_I B_1 / 2\pi \quad (3.38)$$

In isotropic liquids, the static dipole-dipole interaction averages to zero so the spin systems are scalar coupled and polarisation transfer can only occur within the molecule. Under conditions of isotropic mixing, magnetisation of spin 1 is periodically converted into magnetisation of spin 2, if 1 and 2 are scalar coupled. This can be expressed as:

$$I_{1x} \xrightarrow{J\tau} I_{1x}(1 + \cos 2\pi J\tau)/2 + I_{2x}(1 - \cos 2\pi J\tau)/2 + (I_{1y}I_{2z} - I_{1z}I_{2y})\sin 2\pi J\tau \quad (3.39)$$

In a two-spin system, all 1 spin magnetisation can be transferred to 2 (in the absence of relaxation) if τ is set to $1/2J$. Diagonal peaks would then disappear and all intensity would be in the cross peaks. However, short relaxation times in proteins prohibit use of long mixing times.

Equation 3.39 indicates that in-phase 1 spin magnetisation can be transferred to in-phase 2 spin magnetisation. The spectrum can be recorded in absorption mode with cross peaks having the same sign as diagonal peaks. Overlapping antiphase components ($I_{1y}I_{2z}$ and $I_{1z}I_{2y}$) largely cancel one another since protein linewidths are often large relative to J . Since the multiple components of the cross peaks are in-phase, the HOHAHA experiment is generally more sensitive and gives better resolution than COSY-type experiments.

The condition for magnetisation transfer via the Hartmann-Hahn effect is that the difference in the magnitudes of the effective rf fields at the coupled spins 1 and 2 is smaller than the scalar interaction between the spins. The effective rf field is the vector sum of the applied rf field and the resonance offset:

$$\nu_{eff} = \sqrt{(\delta^2 + \nu^2)} \quad (3.40)$$

For the case where $\nu \gg \delta$, the difference is given by:

$$\Delta = \nu_{eff1} - \nu_{eff2} \approx (\delta_1^2 - \delta_2^2)/2\nu \quad (3.41)$$

The condition for Hartmann-Hahn transfer is $\Delta \ll J$. It is insufficient to continuously apply a strong rf field for wide-band cross-polarisation, since too much power is required. Also, transverse NOE effects, which have opposite sign to the Hartmann-Hahn transfers, may lead to cancellation. The original TOCSY sequence [56] gives a significant reduction in the required power, but transverse NOEs are still a problem and the bandwidth covered effectively is still narrow.

Composite Pulse Decoupling Sequences

Any scheme for achieving wide-band cross-polarisation in proteins must minimise the effects of transverse NOEs, use minimal rf power during cross-polarisation to avoid sample heating, and be amenable to routine use. Broad-band heteronuclear decoupling schemes [66, 67] are suitable. MLEV-16 [68] and WALTZ-16 [69] decoupling schemes have been adapted for homonuclear cross-polarisation, and these are the sequences used in the HOHAHA experiments described in this thesis.

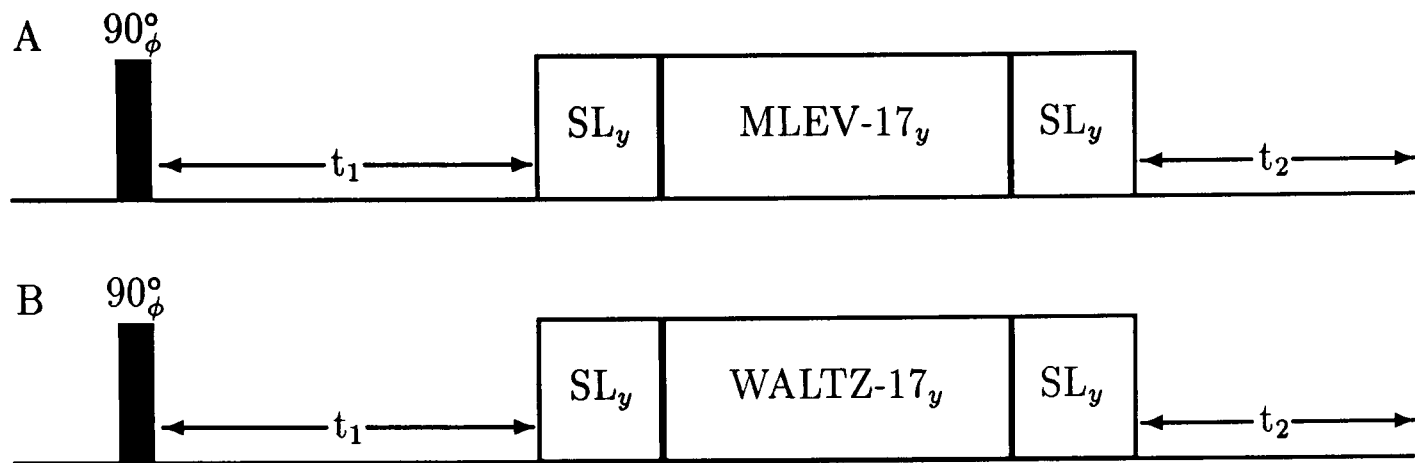


Figure 3.5: (A) Pulse sequence of HOHAHA experiment with $MLEV-17_y$ mixing scheme sandwiched by trim pulses. $MLEV-17_y$ mixing consists of an integral number of repetitions of the sequence: ABBA BBAA BAAB AAB 60_y , where $A = 90_x 180_y 90_x$ and B is the inverse of A. (B) Mixing using a WALTZ- 17_y cycle, which consists of an integral number of repetitions of ABBA - α_y , where $A = 270_{-x} 360_x 180_{-x} 270_x 90_{-x} 180_x 360_{-x} 180_x 270_{-x}$ and $B = 270_x 360_{-x} 180_x 270_{-x} 90_x 180_{-x} 360_x 180_{-x} 270_x$; flip angle α is adjusted between 0 and 90° . The phase cycling for TPPI type experiments (used throughout this work) is $\phi = x, -x$ and $Acq = x, -x$.

$MLEV-16$ is designed to accomplish isotropic mixing over a wide bandwidth, meaning that the magnetisation of any isolated spin finishes in its initial orientation after the $MLEV-16$ cycle. However, this scheme has been shown to be theoretically and practically imperfect, such that I_y may finish after many cycles as I_z and so go undetected. The solution has been to add a 60_y° pulse which prevents y magnetisation from rotating away from the y axis. The resulting scheme is $MLEV-17$ [70] (Figure 3.5).

A second fault in the $MLEV-16$ sequence is the retention of both x and y components during the mixing period. These have different relaxation rates, resulting in quadrature artifacts and phase distortions. Either x or y component can be removed by sandwiching the mixing sequence between trim pulses. In the inhomogeneous rf field, magnetisation which is not parallel to the effective field vector is quickly defocused.

The WALTZ-16 cycle [69] provides more efficient mixing than MLEV-16, with up to a 40% reduction in rf power. A seventeenth pulse is added, as in the case of MLEV-16 and for a similar reason, to give WALTZ-17 (Figure 3.5).

The residual H₂O signal is much more troublesome in HOHAHA than in COSY or NOESY due to the long trim pulses. These excite H₂O magnetisation that is well removed from the coil and in a region of the field with low homogeneity.

Advantages of the MLEV and WALTZ schemes over earlier spin-lock sequences include reduced relaxation during mixing and less interference from transverse NOE effects. In the MLEV-17 sequence, magnetisation spends half its time along the z axis (relaxing with constant T_1) and half in the x - y plane (relaxing with T_2). In macromolecules, proton T_1 s are generally $\gg T_2$ s, so that magnetisation decay is significantly reduced relative to that in less sophisticated schemes, such as the original TOCSY experiment [56]. Similarly for relaxation in the WALTZ-17 sequence. This advantage only applies to in-phase spin-locked magnetisation. When a spin-lock field is applied, NOE effects are positive, such that the cross peaks are of opposite sign to the diagonal. In the MLEV and WALTZ schemes, the magnetisation spends half the time along the z axis, where the NOE is negative for macromolecules. Thus, the spin-locked and regular NOEs largely cancel.

Recording HOHAHA Spectra Without Presaturation

After the mixing period in the pulse sequence [71] of Figure 3.6, spin-locked magnetisation is stored along the z axis through the 90°_ψ flip-back pulse. This magnetisation is then ‘read’ by the 1–1 jump-and-return [72] sequence. Inversion of the water signal leads to severe radiation damping. Therefore, the first excitation pulse is subjected to a two step cycle rather than the conventional four step cycle.

3.5.3 NOESY

The essential details of this technique have already been covered in 3.4.3 and 3.4.4. Points concerning solvent suppression and short mixing time experiments are found in the next section.

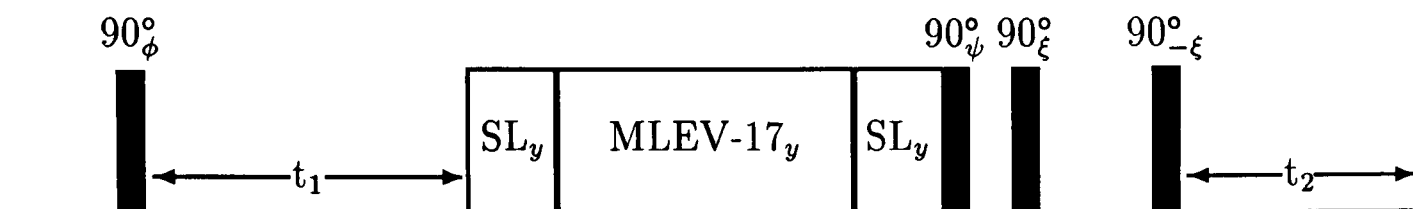


Figure 3.6: Version of the HOHAHA experiment for recording spectra in H_2O solution without presaturation.

3.6 Experimental Details of NMR Spectroscopy

Most NMR spectra were recorded on the Bruker AM600 spectrometer of the Oxford Centre for Molecular Sciences, employing an Oxford Instruments magnet and Aspect 3000 computer. All 2D spectra were recorded in the pure phase absorption mode according to the time proportional phase incrementation (TPPI) method [28, 73] as described by Marion & Wüthrich, 1983 [41].

3.6.1 Parsley Pc Experiments and Data Processing

All data processing was carried out using the Hare Research software packages FT-NMR and FELIX. Data from earlier experiments were processed using FTNMR, the older of the two packages. FELIX was used to re-process these data, and to process data from all subsequent experiments.

NOESY [37, 74] (Figure 3.2) spectra were recorded in D_2O at temperatures of $35^\circ C$ and $20^\circ C$, and in H_2O at $35^\circ C$, $20^\circ C$ and $15^\circ C$. Mixing times of 200ms, 175ms, 150ms, 125ms and 60ms were used. In the case of the experiments with 60ms mixing time, zero quantum coherence effects were minimised with a 10% (D_2O) or 5% (H_2O) random variation of the mixing time. HOHAHA [56, 75, 57, 65] (Figure 3.5) spectra were recorded in D_2O at $35^\circ C$ and $20^\circ C$, and in H_2O at $35^\circ C$, $20^\circ C$ and $15^\circ C$. Early HOHAHA spectra in D_2O were recorded with a $MLEV-17_y$ spin-lock sequence sandwiched between 2.5ms trim pulses. Subsequent experiments in D_2O and H_2O utilised a $WALTZ-17_y$ anisotropic mixing sequence bracketed by 1.5ms and 3ms trim pulses. Mixing times ranged from 34ms to 54ms. D_2O spectra were recorded at pH^* 7.3,

those in H₂O at pH 7.3 and 6.5.

For most NOESY spectra in H₂O, a semiselective ‘jump and return’ sequence ($90_x^\circ - \tau - 90_{-x}^\circ$) [72] in place of the final 90° pulse, was used to suppress the water resonance. For most HOHAHA spectra in H₂O, suppression of the water resonance was achieved using the sequence of Bax et al., 1987 [71] (Figure 3.6). Other NOESY and HOHAHA spectra in H₂O were recorded with presaturation of the water resonance. Spectra in D₂O were recorded with presaturation of the residual HOD resonance. In all cases, baseline distortions in F₂ were minimised by adjusting the reference phase of the spectrometer receiver and the delay time between the detection pulse and the acquisition period so that the zero and first order phase correction parameters were 90° and 180° [76]. Typically, 700–1024 increments of 2K data points were collected per experiment, with thirty-two scans per increment. Digital resolution was approximately 4Hz/pt in each dimension after zero filling. For all NOESY and HOHAHA spectra, a line-broadening parameter of -15Hz to -25Hz and a Lorentz-to-Gaussian resolution enhancement parameter of 0.15, 0.20 or 0.25 were applied in F2 and a 60° or 70° phase-shifted sine-squared bell was applied in F1.

DQF COSY (Figure 3.3) spectra [44] were recorded at 35°C in both D₂O and H₂O, with a phase cycle for removal of artifacts due to rapid pulsing [77]. Selective irradiation during the relaxation delay period was used to suppress the solvent resonance. 1024 increments of 4K data points were acquired. A 0° phase-shifted sine bell window function was applied in both dimensions. For the purpose of accurate measurement of $^3J_{\alpha N}$ and $^3J_{\alpha\beta}$ (valyl) coupling constants, zero filling was employed to increase the digital resolution to 0.86Hz/pt in F2 and 3.44Hz/pt in F1.

P.E.COSY (Figure 3.4) spectra [52] were recorded at 35°C in D₂O and H₂O, and at 42°C in D₂O. 1024 increments of 4K data points were collected, with a spectral width of 7.35kHz. A 0° phase-shifted sine bell window function was applied in both dimensions. Zero filling yielded a digital resolution of 0.90Hz/pt in F2 and 3.6Hz/pt in F1.

3.6.2 CytPCMH Experiments and Data Processing

NOESY [37, 74] (Figure 3.2) and HOHAHA (Figure 3.5) [56, 75, 57, 65] spectra were recorded in D₂O and in H₂O at temperatures of 35°C and 22°C. All NOESY spectra were recorded with a mixing time of 150ms apart from the 35°C H₂O experiment which used a 200ms mixing time. The 35°C HOHAHA experiment in D₂O utilised a MLEV-17_y spin-lock sequence sandwiched between 2.5ms trim pulses and a mixing time of 38ms, whereas all subsequent HOHAHA experiments employed a WALTZ-17_y spin-lock sequence bracketed by trim pulses of 1.5ms and 3ms with a mixing time ranging between 40ms and 44ms. The D₂O spectra were recorded at pH* 7.0, those in H₂O at pH 7.0 and 5.7. Baseline distortions in F₂ were minimised in all spectra using the method described in the previous subsection [76]. Typically, 700–1024 increments of 2K data points were collected per experiment, with thirty-two or forty-eight scans per increment. Digital resolution was approximately 5Hz/pt in each dimension after zero filling. Solvent suppression schemes for the various types of experiment were as described above for the parsley plastocyanin experiments, as were the processing parameters.

A P.COSY [53] spectrum was recorded at 35°C in D₂O. 1024 increments of 2K data points were collected, with thirty-two scans per increment. Selective irradiation during the relaxation delay period was used to suppress the solvent resonance. A 0° phase-shifted sine bell window function was applied in both dimensions.

Bibliography

- [1] J. Jeener, Ampère International Summer School, Basko Polje, Yugoslavia, 1971.
- [2] W. P. Aue, E. Bartholdi & R. R. Ernst (1976) *J. Chem. Phys.* 64, 2229-2246.
- [3] H. D. Plant, T. H. Mareci, M. D. Cockman & W. S. Brey (1986) 27th ENC Conference, Poster A23, Baltimore, MD.
- [4] G. W. Vuister & R. Boelens (1987) *J. Magn. Reson.* 73, 328–333.
- [5] H. Oschkinat, C. Griesinger, P. J. Kraulis, O. W. Sørensen, R. R. Ernst, A. M. Gronenborn & G. M. Clore (1988) *Nature* 332, 374–376.
- [6] A. Bax (1989) *Chemtracts* 1, 215–222.
- [7] G. M. Clore & A. M. Gronenborn (1991) *Science* 252, 1390–1399.
- [8] L. E. Kay, G. M. Clore, A. Bax & A. M. Gronenborn (1990) *Science* 249, 411–414.
- [9] K. Wüthrich, ‘NMR of Proteins and Nucleic Acids’, Wiley, 1986.
- [10] T. F. Havel, G. M. Crippen & I. D. Kuntz (1979) *Biopolymers* 18, 73–81.
- [11] T. F. Havel, I. D. Kuntz & G. M. Crippen (1983) *Bull. Math. Biol.* 45, 665–720.
- [12] W. Braun & N. Gō (1985) *J. Mol. Biol.* 186, 611–626.

- [13] W. F. van Gunsteren, R. Kaptein & E. R. P. Zuiderweg (1983) in *Nucleic Acid Conformation and Dynamics* (W. K. Olson, Ed.) Report of NATO/CECAM Workshop, pp.79–92, Orsay, France.
- [14] R. Kaptein, E. R. P. Zuiderweg, R. M. Scheek, R. Boelens & W. F. van Gunsteren (1985) *J. Mol. Biol.* 182, 179–182.
- [15] G. M. Clore, A. M. Gronenborn, A. T. Brünger & M. Karplus (1985) *J. Mol. Biol.* 186, 435–455.
- [16] G. M. Clore, M. Nilges, D. K. Sukumaran, A. T. Brünger, M. Karplus & A. M. Gronenborn (1986) *EMBO J.* 5, 2729–2735.
- [17] G. Wagner & K. Wüthrich (1986) *Methods Enzymol.* 131, 307–326.
- [18] J. K. M. Sanders & B. K. Hunter, ‘Modern NMR Spectroscopy: A Guide for Chemists’, OUP, 1987.
- [19] A. E. Derome, ‘Modern NMR Techniques for Chemistry Research’, Pergamon Press, Oxford, 1987.
- [20] A. Bax, ‘Two Dimensional Nuclear Magnetic Resonance in Liquids’, Delft University Press, 1982.
- [21] N. Chandrakumar & S. Subramanian, ‘Modern Techniques in High-Resolution FT-NMR’, Springer-Verlag, New York, 1987.
- [22] H. Kessler, M. Gehrke & C. Griesinger (1988) *Angew. Chem. Int. Ed. Engl.* 27, 490–536.
- [23] R. R. Ernst, G. Bodenhausen & A. Wokaun, ‘Principles of Nuclear Magnetic Resonance in One and Two Dimensions’, OUP(Clarendon), 1987.
- [24] M. Goldman, ‘Quantum Description of High-Resolution NMR in Liquids’, OUP, 1988.
- [25] F. Bloch (1946) *Phys. Rev.* 70, 460–474.

- [26] E. O. Stejskal & J. Shaefer (1974) *J. Magn. Reson.* 13, 249–251.
- [27] E. O. Stejskal & J. Shaefer (1974) *J. Magn. Reson.* 14, 160–169.
- [28] A. G. Redfield & S. D. Kunz (1975) *J. Magn. Reson.* 19, 250–254.
- [29] A. Abragam, ‘The Principles of Nuclear Magnetism’, OUP(Clarendon), 1961.
- [30] C. P. Slichter, ‘Principles of Magnetic Resonance, 2nd ed.’, Springer-Verlag, New York, 1980.
- [31] O. W. Sørensen, G. W. Eich, M. H. Levitt, G. Bodenhausen & R. R. Ernst (1983) *Prog. Nucl. Magn. Reson. Spectrosc.* 16, 163–192.
- [32] U. Fano (1957) *Rev. Mod. Phys.* 29, 74–93.
- [33] W. P. Aue, E. Bartholdi & R. R. Ernst (1976) *J. Chem. Phys.* 64, 2229–2246.
- [34] I. Solomon (1955) *Phys. Rev.* 99, 559–565.
- [35] J. H. Noggle & R. E. Schirmer, ‘The Nuclear Overhauser Effect: Chemical Applications’, Academic Press, New York, 1971.
- [36] C. P. Poole & H. A. Farach, ‘Relaxation in Magnetic Resonance: Dielectric and Mössbauer Applications’, Academic Press, New York, 1971.
- [37] J. Jeener, B. H. Meier, P. Bachmann & R. R. Ernst (1979) *J. Chem. Phys.* 71, 4546–4553.
- [38] G. Bodenhausen, H. Kogler & R. R. Ernst (1984) *J. Magn. Reson.* 58, 370–388.
- [39] D. I. Hoult & R. E. Richards (1975) *Proc. Roy. Soc. Lond. A* 344, 311–340.
- [40] D. J. States, R. A. Haberkorn & D. J. Ruben (1982) *J. Magn. Reson.* 48, 286–292.

- [41] D. Marion & K. Wüthrich (1983) *Biochem. Biophys. Res. Commun.* 113, 967–974.
- [42] U. Piantini, O. W. Sørensen & R. R. Ernst (1982) *J. Am. Chem. Soc.* 104, 6800–6801.
- [43] A. J. Shaka & R. Freeman (1983) *J. Magn. Reson.* 51, 169–173.
- [44] M. Rance, O. W. Sørensen, G. Bodenhausen, G. Wagner, R. R. Ernst & K. Wüthrich (1983) *Biochem. Biophys. Res. Commun.* 117, 458–479.
- [45] N. Müller, R. R. Ernst & K. Wüthrich (1986) *J. Am. Chem. Soc.* 108, 6482–6492.
- [46] C. Griesinger, O. W. Sørensen & R. R. Ernst (1985) *J. Am. Chem. Soc.* 107, 6394–6396.
- [47] C. Griesinger, O. W. Sørensen & R. R. Ernst (1987) *J. Magn. Reson.* 75, 474–492.
- [48] B. U. Meier, G. Bodenhausen & R. R. Ernst (1984) *J. Magn. Reson.* 60, 161–163.
- [49] P. Pfändler, G. Bodenhausen, B. U. Meier & R. R. Ernst (1985) *Anal. Chem.* 57, 2510–2516.
- [50] A. Bax & R. Freeman (1981) *J. Magn. Reson.* 44, 542–561.
- [51] A. Bax, R. H. Griffey & B. L. Hawkins (1983) *J. Magn. Reson.* 55, 301–315.
- [52] L. Müller (1987) *J. Magn. Reson.* 72, 191–196.
- [53] D. Marion & A. Bax (1988) *J. Magn. Reson.* 80, 528–533.
- [54] A. Bax & L. Lerner (1988) *J. Magn. Reson.* 79, 429–438.
- [55] J. Keeler & D. Neuhaus (1985) *J. Magn. Reson.* 63, 454–472.
- [56] L. Braunschweiler & R. R. Ernst (1983) *J. Magn. Reson.* 53, 521–528.

- [57] A. Bax & D. G. Davis (1985) *J. Magn. Reson.* 63, 207–213.
- [58] A. A. Bothner-By, R. L. Stephens, J. T. Lee, C. D. Warren & R. W. Jeanloz (1984) *J. Am. Chem. Soc.* 106, 811–813.
- [59] S. R. Hartmann & E. L. Hahn (1962) *Phys. Rev.* 128, 2042–2053.
- [60] A. Bax & D. G. Davis (1986) in *Advanced Magnetic Resonance Techniques in Systems of High Molecular Complexity* (N. Niccolai, G. Valensin, Eds.) pp. 21–48, Birkhaeuser, Basel.
- [61] J. S. Waugh (1986) *J. Magn. Reson.* 68, 189–192.
- [62] R. Bazzo & J. Boyd (1987) *J. Magn. Reson.* 75, 452–466.
- [63] R. R. Ernst, G. Bodenhausen & A. Wokaun, ‘Principles of NMR in One and Two Dimensions’, p.444, OUP(Clarendon), 1987.
- [64] A. Bax (1988) *J. Magn. Reson.* 77, 134–147.
- [65] A. Bax (1989) *Methods Enzymol.* 176, 151–168.
- [66] T. A. Frenkiel, M. H. Levitt & R. Freeman (1983) *Adv. Magn. Reson.* 11, 47–110.
- [67] A. J. Shaka & J. Keeler (1987) *Prog. Nucl. Magn. Reson. Spectrosc.* 19, 47–129.
- [68] M. H. Levitt, R. Freeman & T. A. Frenkiel (1982) *J. Magn. Reson.* 47, 328–330.
- [69] A. J. Shaka, J. Keeler, T. A. Frenkiel & R. Freeman (1983) *J. Magn. Reson.* 52, 335–338.
- [70] A. Bax & D. G. Davis (1985) *J. Magn. Reson.* 65, 355–366.
- [71] A. Bax, V. Sklenář, A. M. Gronenborn & G. M. Clore (1987) *J. Am. Chem. Soc.* 109, 6511–6513.

- [72] P. Plateau & M. Guéron (1982) *J. Am. Chem. Soc.* 104, 7310–7311.
- [73] G. Bodenhausen, R. L. Vold & R. R Vold (1980) *J. Magn. Reson.* 37, 93–106.
- [74] C. Macura, Y. Huang, D. Suter & R. R. Ernst (1981) *J. Magn. Reson.* 43, 259–281.
- [75] D. G. Davis & A. Bax (1985) *J. Am. Chem. Soc.* 107, 2821–2822.
- [76] D. Marion & A. Bax (1988) *J. Magn. Reson.* 79, 532–356.
- [77] A. E. Derome & M. P. Williamson (1990) *J. Magn. Reson.* 88, 177–185.

Chapter 4

NMR-Derived Data for Protein Structure Determination

The process of determining the solution structure of a protein using NMR-derived data consists of the following stages:

1. Sequential resonance assignment:
 - Delineation of amino acid spin systems using experiments which demonstrate through-bond connectivities.
 - Identification of neighbouring amino acids using experiments which demonstrate through-space ($\leq 5\text{\AA}$) short range ($|i - j| \leq 5$) interproton connectivities.
2. Identification of regular secondary structure elements.
3. Measurement of three-bond coupling constants.
4. Assignment of long range ($|i - j| \geq 5$) NOE distance restraints.
5. Classification of NOEs according to intensity.
6. Calculation of three dimensional structure on the basis of distance and torsion angle restraints.

4.1 Sequential Resonance Assignment

The sequential resonance assignment procedure was developed by Wüthrich and co-workers [1]. It is described in some detail in this thesis with respect to the assignment of the ^1H NMR spectrum of parsley plastocyanin.

4.2 Distance Restraints from the Nuclear Overhauser Effect

Of fundamental importance, both in the assignment process and in the identification of interproton distance restraints, is the NOE. The basis of the NOESY experiment has been considered briefly in sections 3.4.3 and 3.4.4. Aspects of the NOE relevant to structure determination are dealt with here.

The rate constant for exchange of magnetisation between two protons 1 and 2 (cross relaxation rate), σ_{12} , is given by

$$\sigma_{12} = f(\tau_{12}) r_{12}^{-6} \quad (4.1)$$

where r_{12} is the distance between the two protons and $f(\tau_{12})$ is a function of the correlation time of the interproton vector τ_{12} .

4.2.1 Initial Slopes of NOE Build Up Curves

If the molecule is assumed to be rigid, the initial build up rate of cross peaks in 2D NOE spectra recorded with short mixing times τ_m is proportional to σ_{12} . Thus, σ_{12} can be measured from a single 2D spectrum or, more accurately, from a number of experiments recorded with a range of mixing times. Obstacles to the accurate determination of interproton distances from the initial slopes of NOE build up curves include:

1. The linear range of the build up curves is short. At longer mixing times the slope decreases and becomes negative due to relaxation. This problem could be overcome by dividing cross peak intensities by diagonal peak intensities at every mixing time [2]. In practice, resonance overlap often prevents accurate

integration of diagonal peaks. Overlap can be removed in the case of amide protons if the sample is ^{15}N -labelled, by performing a ^{15}N relayed NOESY.

2. One of the possible methods for obtaining quantitative interproton distances is to use a known calibration distance:

$$r_{12} = r_{cal}(\sigma_{cal}/\sigma_{12})^{1/6} \quad (4.2)$$

A possible standard distance in proteins is that between methylene protons (1.8Å). The distances between aromatic ring protons are also fixed, but NOE intensities involving corresponding pairs of protons in different aromatic side chains are variable. This suggests that the assumption of rigidity is inappropriate for longer side chains. Protons with a fixed separation tend to have similar chemical shifts in proteins, so that the cross peaks are near to the diagonal and are difficult to integrate. A further difficulty arises from the fact that the protons are J-coupled, so that the zero quantum contribution (see below) to the NOESY cross peak must be suppressed.

3. Spectrometer performance may change with time, complicating the comparison of cross peak intensities obtained with different mixing times.
4. Time restrictions may prevent collection of a sufficient number of data points to cover the time course of the NOE.

4.2.2 Zero Quantum Contributions

There are other problems associated with accurate determination of interproton distances which apply to NOE experiments in general. That of zero quantum contributions to NOESY cross peaks of scalar coupled nuclei has been alluded to above. The J-peaks have antiphase patterns which are wider than the inphase NOESY cross peaks and have perturbing sidelobes. They may interfere with neighbouring NOESY cross peaks. Attempts to eliminate zero quantum contributions have included random variation of the mixing time [3, 4] or incrementation of the mixing time proportional to t_1 [3, 4]. A more recent proposal involves co-addition of NOESY spectra in which

the mixing times have been increased in equal increments [5]. The increments depend on the frequency range over which suppression is required.

4.2.3 Spin Diffusion

The major errors in distances obtained by assuming an isolated spin pair are caused by spin diffusion, which is the indirect transfer of magnetisation between two spins *via* one or more other spins. This may cause NOEs to be more intense than expected for a direct dipole-dipole interaction, and may lead to the appearance of spurious NOE peaks. Consequently, erroneous distance restraints may be used in the structure calculation. This problem is most serious for NOEs involving methylene protons, and may be avoided by recording spectra with a range of mixing times: cross peaks due to spin diffusion are absent from spectra recorded with short mixing times, and build up as the mixing time is increased. Another possible solution to the problems associated with spin diffusion is to obtain distance restraints from ROESY spectra. Spin diffusion effects are then easily identified since second order effects have opposite sign to first order effects. Hartmann-Hahn contribution to ROEs between scalar coupled protons does, however, make quantification difficult. If a three dimensional structure is available from inaccurate NMR data, the effects of spin diffusion can be calculated using a full relaxation matrix analysis, permitting refinement of the structure [6, 7].

4.2.4 Protein Mobility

The effect of protein mobility on NOE data has been largely neglected. Internal motions which reorient the internuclear vector relative to the external field faster than the tumbling of the whole molecule reduce cross relaxation rates. If $\tau_{int} \ll \tau_c$, where τ_{int} is the correlation time for internal motion and τ_c is that for overall tumbling of the molecule, the measured NOE build up rate reflects $\langle r^{-3} \rangle^2$. If $\tau_{int} \gg \tau_c$, the averaged distance appears as $\langle r^{-6} \rangle$ in equation 4.1 [9]. In either case, neglect of internal motion results in an apparent distance shorter than the average distance $\langle r \rangle$. It has been indicated that the effects of distance fluctuations are largely compensated by angular fluctuations, and that a $\langle r^{-6} \rangle$ model may be suitable in

the presence of fast local motion [8].

As a result of internal motions, a protein may exist as a number of conformers interconverting rapidly on the NMR timescale. The observed NOEs will then reflect an average of the conformations, such that the NOE pattern may be inconsistent with any single conformation. In order to take into account the time-dependence of the structure, the restraining potential might be based on the time-averaged distance between atoms [10]. This will allow sufficient freedom for each conformation to occur, and should result in greater accuracy of NMR-derived structures. Many such structures are likely to be subject to dynamic averaging, particularly in longer side chains at the protein surface.

4.2.5 Definition of the Fold Using Approximate Distance Restraints

Clore and Gronenborn [3] have pointed out that because of the $\langle r^{-6} \rangle$ dependence of the NOE, approximate distance restraints can be obtained in spite of large variations in effective correlation times and other difficulties referred to above. This is a consequence of their classification system: strong, medium and weak NOEs correspond to distance ranges of 1.8–2.7, 1.8–3.5 and 1.8–5.0Å. (The lower limit of 1.8Å is the sum of the van der Waals radii of two protons. The same value is used in each of the classes since local motion could lead to attenuation of NOE cross peaks). A decrease in the effective correlation time of an interproton vector is manifested as a reduction in the intensity of the NOE, so that the corresponding distance is defined less precisely, rather than erroneously. Further, it has been shown [15, 13] that an accurate definition of a protein fold can be obtained using a large number of relatively imprecisely defined NOE-observed distances. This is attributed to the fact that the polypeptide folds back on itself, producing long range restraints. It is more important to maximise the number of restraints than to include a smaller number of accurately determined restraints. With this in mind, it becomes advisable to record NOESY spectra with relatively long mixing times, even though some of the NOEs will be associated with only loose distance restraints and some will be subject to spin diffusion contributions.

4.2.6 Distribution of NOEs

One of the strengths of NMR for structure determination lies in the ability to identify many distance constraints defining the overall topology. Current high resolution structure determinations employ an average of fifteen to twenty distance restraints per residue. The restraints are not uniformly distributed through the protein: the conformation of interior residues may be defined by forty or more interresidue NOEs, whereas a surface residue may be characterised by as few as two or three NOEs, none of them long range. NOEs are most appropriate for definition of overall fold and secondary structure. Other information, including coupling constants and hydrogen bonds, is required to characterise protein surfaces. This aspect of NMR studies of proteins should become increasingly prominent, reflecting the importance of the protein surface for function.

4.2.7 Effect of Approximations on the Accuracy of Derived Structures

In a recent report [14], hypothetical NOESY spectra generated from a high resolution crystal structure were used to produce sets of distance constraints by three different methods. Two methods involved the isolated spin pair approximation (ISPA), with restrictive and conservative assignment of distance bounds. The third method used complete relaxation matrix analysis with the program MARDIGRAS [15], which accounts for spin diffusion and will be modified to take into account internal motions. The accuracy of the resulting distance geometry-generated structures was assessed partly by back calculation of NOESY spectra [16] (computed spectra are compared with experimental spectra) with subsequent determination of a factor similar to the crystallographic R factor. The most significant trend was that structural accuracy improved with increasingly restrictive restraints. Also, accumulation of distance errors in local structure regions produced systematic errors in the ISPA structures. The MARDIGRAS structures showed none of the systematic errors.

α -helix ($\phi = -57^\circ$)	$^3J_{HN\alpha} = 3.9\text{Hz}$
3_{10} helix ($\phi = -60^\circ$)	$^3J_{HN\alpha} = 4.2\text{Hz}$
Antiparallel β -sheet ($\phi = -139^\circ$)	$^3J_{HN\alpha} = 8.9\text{Hz}$
Parallel β -sheet ($\phi = -119^\circ$)	$^3J_{HN\alpha} = 9.7\text{Hz}$

Table 4.1: Values of the ϕ backbone torsion angles, and associated $^3J_{HN\alpha}$ coupling constants, for different types of regular secondary structure.

4.3 Torsion Angle Restraints from Coupling Constants

4.3.1 ϕ Backbone Torsion Angle Restraints

Coupling constants between protons three bonds apart have a characteristic dependence on the intervening torsion angle θ [17]. Polypeptide $^3J_{HN\alpha}$ coupling constants in Hertz are related to the torsion angle ϕ by:

$$^3J_{HN\alpha} = 6.4\cos^2\theta - 1.4\cos\theta + 1.9 \quad (4.3)$$

where $\theta = |\phi - 60^\circ|$. The coefficients in equation 4.3 were obtained by correlating the ϕ angles in the crystal structure of BPTI and the $^3J_{HN\alpha}$ values measured in solution [6]. In proteins, values of ϕ are concentrated in the range -30 to -180° for all amino acids except glycine [27] (Figure 4.1). $^3J_{HN\alpha}$ takes the values shown in Table 4.1 in regular secondary structure elements.

$^3J_{HN\alpha}$ values can be obtained by measuring the peak-to-peak separation of antiphase components of NH-C $^\alpha$ H cross peaks in a DQF COSY spectrum. Couplings smaller than the linewidth (usually $\geq 5\text{Hz}$) cannot be reliably measured due to cancellation effects in antiphase multiplets [20].

4.3.2 Side Chain Torsion Angle Restraints and Stereospecific Assignment of Methylene and Methyl Resonances

Stereospecific assignment of prochiral β -methylene protons and χ_1 torsion angle restraints can be obtained from values of $^3J_{\alpha\beta}$ coupling constants (from P.E.COSY spectra; see Chapter 7) combined with intraresidue NOEs (from short mixing time experiments) involving backbone amide, C $^\alpha$ and C $^\beta$ protons. These parameters have

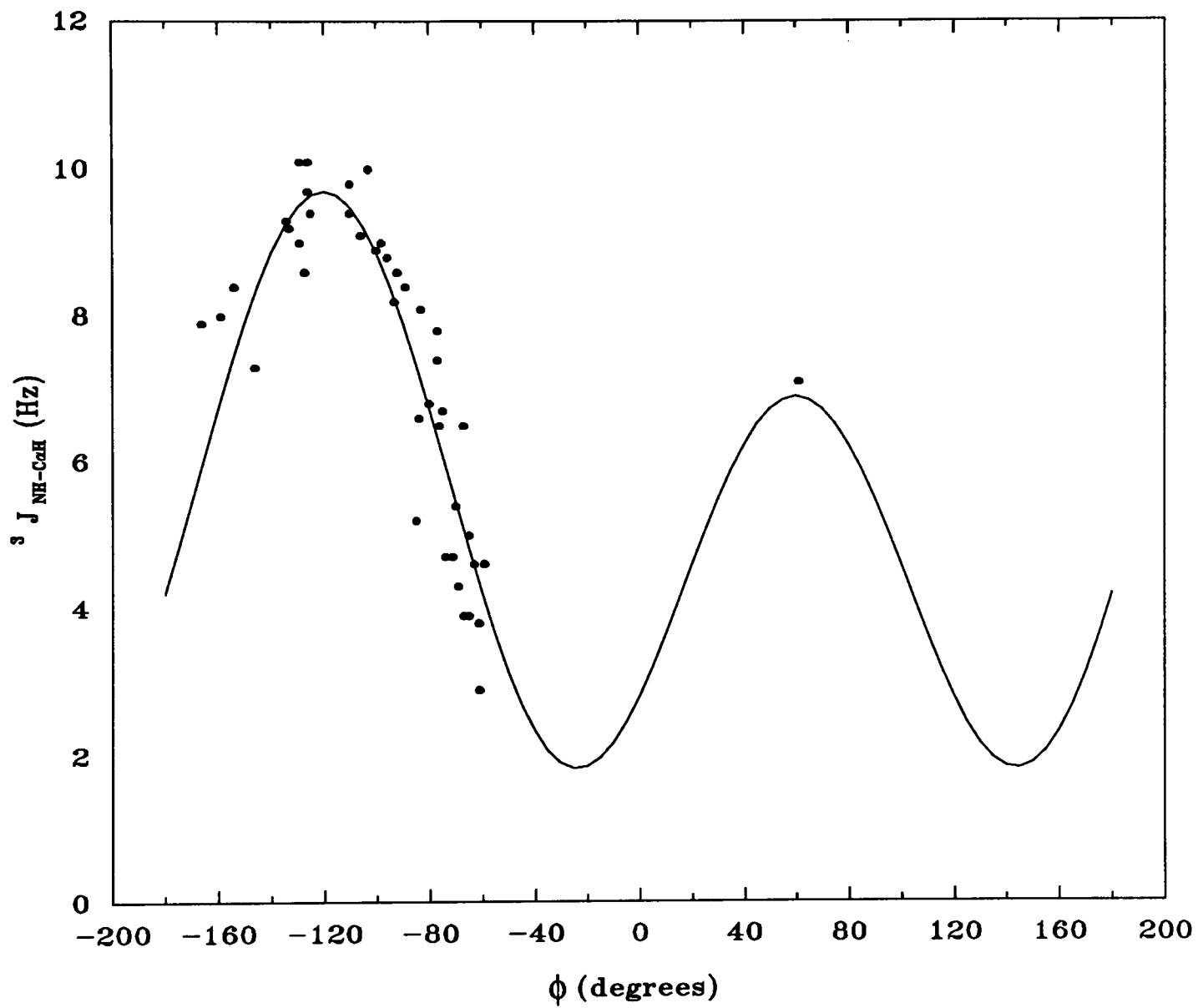


Figure 4.1: Plot of $^3J_{HN\alpha}$ versus the torsion angle ϕ in BPTI. The points represent values of $^3J_{HN\alpha}$ for individual residues. ϕ was obtained from the crystal structure. The curve corresponds to the best fit obtained with the Karplus equation.

a characteristic pattern for each of the three energetically favourable staggered rotamer positions, corresponding to $\chi_1 = 60^\circ$, 180° or -60° [2]. A similar approach can be taken to obtain stereospecific assignment of methyl groups of valine residues [9].

Analysis of high resolution crystal structures has shown that 95% of all χ_1 angles lie within $\pm 15^\circ$ of the staggered rotamer conformations [24, 25]. It also seems that the better refined the structure, the closer the χ_1 angles to the staggered rotamer values.

Rotational averaging (i.e. a mixture of rotamer populations) is indicated if both $^3J_{\alpha\beta}$ coupling constants are approximately 6–7Hz. No stereospecific assignment can be obtained in such cases. Degeneracy of $C^\beta H$ or $C^\gamma H_3$ chemical shifts can also prevent stereospecific assignment.

The so called pseudo atom approximation is used in the absence of a stereospecific assignment. The distance to a methylene proton is entered as the distance to the geometric centre between the two methylene protons, and a $1\text{\AA}$ correction is added to the upper bound [23]. The same approach is used for nonstereospecifically assigned methyl groups and for protons of rapidly flipping aromatic rings, with a $2.4\text{\AA}$ and a $2\text{\AA}$ correction in each case.

The influence of stereospecific assignment of β -methylene protons and inclusion of χ_1 torsion angle restraints on three dimensional structure determination has been studied [8]. It was shown that the definition of the backbone, and especially of side chains, is significantly improved by the inclusion of restraints involving stereospecifically assigned protons. This improvement can extend to side chains remote from the stereospecifically assigned β -methylene groups.

4.4 Identification of Regular Secondary Structure Elements

4.4.1 NOEs and Secondary Structure

An element of secondary structure can be identified by qualitative analysis of the pattern of short range ($|i - j| \leq 5$) NOEs. The reliability of such qualitative methodology has been assessed using a statistical analysis of high resolution crystal structures [1].

It was found, for example, that 98% of all helical residues are located in a search with $d_{NN} \leq 3.6\text{\AA}$. However, the *uniqueness* (percentage of all distance values $d_{NN} \leq 3.6\text{\AA}$ in the whole molecule located in helical regions) was only 49%. A similar result was found for β -sheet identification.

This low uniqueness arises from the fact that one of the distances $d_{\alpha N}$ and d_{NN} is short in all local conformations corresponding to allowed regions in the Ramachandran ϕ versus ψ plot. The uniqueness is improved if several short sequential distances are identified.

Helices can be discerned from a string of intense or medium d_{NN} NOEs ($2.8\text{\AA}$), whereas β -strands are characterised by strong $d_{\alpha N}$ ($2.2\text{\AA}$) connectivities and by the absence of other short range NOEs involving the amide and C^α protons. In practice, these sequential NOEs provide a strong indication of the location of helices and strands. Confirmation or otherwise may be obtained using medium range NOEs for helices and interstrand long range NOEs for β -sheets. Thus, the location of an α -helix can be more firmly identified from a series of medium or weak $d_{\alpha N}(i, i+3)$ ($3.4\text{\AA}$) and $d_{\alpha\beta}(i, i+3)$ ($2.5\text{\AA}$ and $4.4\text{\AA}$) NOEs. In theory, observation of weak $d_{\alpha N}(i, i+4)$ NOEs distinguishes an α -helix from a 3_{10} -helix. $d_{\alpha N}(i, i+2)$ NOEs may be observed for the latter. However, the $d_{\alpha N}(i, i+4)$ and $d_{\alpha N}(i, i+2)$ NOEs are normally only observed at longer mixing times, so that the distinction may be unreliable due to spin diffusion. Unambiguous distinction of helix types is made after calculation of structure.

β -sheets can be identified and aligned from interstrand NOEs involving amide, C^α and C^β protons. For example, antiparallel β -sheets can be distinguished from parallel β -sheets by the appearance of intense $d_{\alpha\alpha}(i, j)$ ($2.3\text{\AA}$) NOEs. Other interstrand distances in antiparallel sheets are $d_{\alpha N}(i, j) = 3.2\text{\AA}$ and $d_{NN}(i, j) = 3.3\text{\AA}$. For parallel β -sheets, $d_{\alpha\alpha}(i, j) = 4.8\text{\AA}$, $d_{\alpha N}(i, j) = 3.0\text{\AA}$ and $d_{NN}(i, j) = 4.0\text{\AA}$.

4.4.2 Coupling Constants and Secondary Structure

The use of $^3J_{HN\alpha}$ values for the identification of secondary structure elements has been described in 4.3.1. These data provide important evidence to support the identifications made on the basis of short interproton distances.

4.4.3 Hydrogen Exchange and Secondary Structure

Slowly exchanging amide protons are usually involved in hydrogen bonding. This fact can be used to provide supporting evidence for regular secondary structure elements identified by a combination of NOE and $^3J_{HN\alpha}$ data. For example, it can be used to confirm the presence of the bonds CO_i-NH_{i+4} in an α -helix, and CO_i-NH_{i+3} in a 3_{10} -helix. All amide protons in regular secondary structures are involved in hydrogen bonding, except the first four residues in an α -helix, the first three residues in a 3_{10} -helix and alternate residues in the peripheral strands of a β -sheet. Hydrogen exchange rates are thus especially useful in the location of the C-terminal end of helices, and in the distinction between peripheral and internal strands in β -sheets.

4.4.4 Effect of Secondary Structure Identification on Structure Calculation

Knowledge of secondary structure facilitates location of hydrogen bonds. Also, improved convergence may be obtained in early calculations due to inclusion of dihedral angle restraints for elements of regular secondary structure.

4.5 Identification of Tight Turns

Tight turns are more difficult to identify reliably using NMR-derived data than helices and β -sheets. The reasons for this, and the procedures used to identify and classify turns, are described in Chapter 6. Discussion of the turns and loops in parsley plastocyanin, based on detailed structural information, rather than NOE and coupling constant data, is presented in Chapter 7.

4.6 Methods of Structure Calculation

There is no single, universally accepted method for the determination of protein tertiary structure from NMR-derived data. All methods have as their main aim the characterisation of the conformational space compatible with NMR-derived restraints. An ensemble of structures is produced, each a particular representation of the allowed

conformational space, and together showing the range of conformations sampled by the protein.

The most widely used methods have in common a conformational search to locate the global minimum of a target function that consists of stereochemical and experimental NMR restraints. The target function is characterised by many false local minima, such that the descent to the global minimum region is not straightforward.

The two general classes of methods are (i) real space methods, which include restrained molecular dynamics [7] and dynamical simulated annealing [10, 14, 16] in cartesian coordinate space, and (ii) distance space methods, generally known as distance geometry [32]. Real space methods require initial structures, which can be any one of four types: random structures with the correct covalent geometry, structures very far from the final structure, a random array of atoms, or structures generated by distance space methods. Critical to the success of real space methods is the avoidance of incorrect folding. This problem is circumvented in methods developed by Nilges [10, 14, 16], as outlined later in this chapter, and as described in some detail in Chapter 7. The standard general procedure for structure calculation is to begin with a set of distance geometry type structures and to refine these with molecular dynamics.

4.6.1 Metric Matrix Distance Geometry

The metric matrix distance geometry algorithm was developed long before protein structure determination by NMR became possible. Distance geometry is a method for converting a set of bounds on distances between atoms into a conformation consistent with those bounds [32]. The first step is generation of upper and lower bound matrices $\mathbf{U}$ and $\mathbf{L}$ for all atom-atom distances of the molecule. Some of the elements r_{ij}^u and r_{ij}^l follow from standard covalent geometry, and some from NMR data. A bound smoothening procedure using triangle inequalities extends the constraints to all elements of $\mathbf{U}$ and $\mathbf{L}$. A distance matrix $\mathbf{R}$ is then set up, consisting of randomly chosen distances such that $l_{ij} \leq r_{ij} \leq u_{ij}$. This matrix is then converted into a structure in cartesian space using the embedding algorithm. The deviations of the initial structures with respect to the data are then minimised to produce the final structures. The random step of choosing the distance matrix $\mathbf{R}$ means that repeti-

tion of the calculation produces different structures. This allows assessment of how uniquely the structure is defined by the constraints.

Model calculations have shown that in regions of the protein that are relatively poorly constrained by NMR data, distance geometry tends to produce slightly expanded structures [15]. This implies that conformational space is not being sampled in a strictly random manner. The final structures tend to have relatively poor stereochemistry, especially in terms of nonbonded contacts. Computational times are long and rise rapidly with increase in molecular size. A further drawback is that an energy function cannot be incorporated directly in distance geometry type calculations, but must be converted into a function of atom-atom distances. This involves substantial loss of information from the energy function, and hinders the characterisation of the distribution of conformations. That is, whether the conformations populate low energy regions of conformational space according to a Boltzmann weighting. If this is not the case, the generated conformations do not represent the range of conformations sampled by the protein.

4.6.2 Restrained Molecular Dynamics

Restrained molecular dynamics involves simultaneous integration of Newton's equations of motion for all atoms of the system at a temperature that is directly proportional to the kinetic energy of the system. It differs from regular molecular dynamics in the inclusion of an extra energy term, F_{expt} , which represents the experimental NMR data [33]. F_{expt} pulls protons to within the distance range determined from the experimental data. Its value depends on the extent to which a structure violates the data. The potential energy is defined by:

$$F_{tot} = F_{bond} + F_{angle} + F_{dihed} + F_{vdW} + F_{coulomb} + F_{expt} \quad (4.4)$$

The first two terms tend to keep bond lengths and bond angles at equilibrium values. F_{dihedr} is a sinusoidal potential describing rotations about bonds. A Lennard-Jones potential is generally used for F_{vdW} , and $F_{coulomb}$ describes electrostatic interactions. Interproton distance restraints and torsion angle restraints are incorporated as square well potentials (see Chapter 7).

The force constants for the restraints are set sufficiently high to ensure that the experimental restraints are satisfied within the precision of the measurements. The empirical energy function should be set up such that deviations from idealised covalent geometry are small, with no bad nonbonded contacts. The latter will be evidenced by a large negative Lennard-Jones energy, as is referred to in Chapter 7 with respect to simulated annealing structures of parsley plastocyanin.

Restrained molecular dynamics involves a large computational requirement, mainly because of the inclusion of a full nonbonded interaction potential. It is thus best used for refinement. For example, molecular dynamics refinement of distance geometry structures has become a standard procedure, as mentioned above.

4.6.3 Dynamical Simulated Annealing

This method [10, 14, 16] is conceptually similar to restrained molecular dynamics, and differs most significantly in the replacement of the van der Waals, electrostatic and hydrogen bonding potentials by a simple van der Waals repulsion term. This reduces calculation time and allows greater flexibility in the type of protocol to be used.

Dynamical simulated annealing efficiently overcomes potential energy barriers on the path to the global minimum region by initially raising the temperature of the system followed by slow cooling. In the initial high kinetic energy phase, atoms are free to move through one another, ensuring that the correct fold is obtained. This circumvents one of the most critical problems of real space methods.

Three simulated annealing protocols have been used in this work to calculate solution structures of parsley plastocyanin. In one method, dynamical simulated annealing is used to fold an extended polypeptide strand with random dihedral angles [14]. The second is a hybrid of the metric matrix distance geometry and simulated annealing methods [10]. The third protocol is a variation on the second, using restriction of the number of degrees of freedom of the system in order to reduce computation time. These protocols will be described in detail in Chapter 7.

Bibliography

- [1] K. Wüthrich, ‘NMR of Proteins and Nucleic Acids’, Wiley, 1986.
- [2] S. Macura, B. T. Farmer & L. R. Brown (1986) *J. Magn. Reson.* 70, 493–499.
- [3] S. Macura, K. Wüthrich & R. R. Ernst (1982) *J. Magn. Reson.* 47, 351–357.
- [4] S. Macura, Y. Huang, D. Suter & R. R. Ernst (1981) *J. Magn. Reson.* 43, 259–281.
- [5] M. Rance, G. Bodenhausen, G. Wagner, K. Wüthrich & R. R. Ernst (1985) *J. Magn. Reson.* 62, 497–510.
- [6] B. A. Borgias & T. L. James (1988) *J. Magn. Reson.* 79, 493–512.
- [7] B. A. Borgias, M. Gochin, D. J. Kerwood & T. L. James (1990) *Prog. NMR Spectrosc.* 22, 83–100.
- [8] D. M. LeMaster, L. E. Kay, A. T. Brünger & J. H. Prestegard (1988) *FEBS Lett.* 236, 71–76.
- [9] D. Neuhaus & M. Williamson, ‘The Nuclear Overhauser Effect in Structural and Conformational Analysis’, VCH Publishers, 1989.
- [10] A. E. Torda, R. M. Scheek & W. F. van Gunsteren (1990) *J. Mol. Biol.* 214, 223–256.
- [11] G. M. Clore & A. M. Gronenborn (1989) *CRC Crit. Rev. Biochem.* 24, 479–564.
- [12] T. F. Havel & K. Wüthrich (1985) *J. Mol. Biol.* 182, 281–294.

- [13] W. Braun & N. Gō (1985) *J. Mol. Biol.* 186, 611–626.
- [14] P. D. Thomas, V. J. Basus & T. L. James (1991) *Proc. Natl. Acad. Sci. USA* 88, 1237–1241.
- [15] B. A. Borgias & T. L. James (1990) *J. Magn. Reson.* 87, 475–487.
- [16] E.-I. Suzuki, N. Pattabiraman, G. Zon & T. L. James (1986) *Biochemistry* 25, 6854–6865.
- [17] M. Karplus (1963) *J. Am. Chem. Soc.* 85, 2870–2871.
- [18] A. Pardi, M. Billeter & K. Wüthrich (1984) *J. Mol. Biol.* 180, 741–751.
- [19] J. S. Richardson (1981) *Advan. Protein Chem.* 34, 167–339.
- [20] D. Neuhaus, G. Wagner, M. Vašák, J. H. R. Kägi & K. Wüthrich (1985) *Eur. J. Biochem.* 151, 257–273.
- [21] G. Wagner, W. Braun, T. F. Havel, T. Schaumann, N. Gō & K. Wüthrich (1987) *J. Mol. Biol.* 196, 611–639.
- [22] E. R. P. Zuiderweg, R. Boelens & R. Kaptein (1985) *Biopolymers* 24, 601–611.
- [23] K. Wüthrich, M. Billeter & W. Braun (1983) *J. Mol. Biol.* 169, 949–961.
- [24] J. W. Ponder & F. M. Richards (1987) *J. Mol. Biol.* 193, 775–791.
- [25] M. J. McGregor, S. A. Islam & M. J. E. Sternberg (1987) *J. Mol. Biol.* 198, 295–310.
- [26] P. C. Driscoll, A. M. Gronenborn & G. M. Clore (1989) *FEBS Lett.* 243, 223–233.
- [27] K. Wüthrich, M. Billeter & W. Braun (1984) *J. Mol. Biol.* 180, 715–740.
- [28] G. M. Clore, A. M. Gronenborn, M. Nilges & C. A. Ryan (1987) *Biochemistry* 26, 8012–8023.

- [29] M. Nilges, A. M. Gronenborn & G. M. Clore (1988) *FEBS Lett.* 229, 317–324.
- [30] M. Nilges, A. M. Gronenborn, A. T. Brünger & G. M. Clore (1988) *Protein Eng.* 2, 27–38.
- [31] M. Nilges, A. M. Gronenborn & G. M. Clore (1988) *FEBS Lett.* 239, 129–136.
- [32] T. F. Havel, I. D. Kuntz & G. M. Crippen (1983) *Bull. Math. Biol.* 45, 665–720.
- [33] R. Kaptein, E. R. P. Zuiderweg, R. M. Scheek, R. Boelens & W. F. van Gunsteren (1985) *J. Mol. Biol.* 182, 179–182.

Chapter 5

Haem Geometry and Resonance Assignment in the Cytochrome Subunit of PCMH

PCMH has a molecular weight of approximately 115 000, and is thus too large to provide useful 2D ^1H NMR spectra. There are two main reasons for this: cross peak overlap and low efficiency of magnetisation transfer as linewidths become comparable to or greater than ^1H - ^1H coupling constants, which are usually $\leq 14\text{Hz}$. Larger linewidths result from an increase in rotational correlation time of the molecule. PCMH can, however, be reversibly dissociated into its constituent subunits [2]. One of these subunits, a *c*-type cytochrome (cytPCMH), has a molecular weight of 8780. The small size of this protein makes it amenable to study by 2D NMR methods.

The 1D ^1H NMR spectrum of cytPCMH confirms that it retains a highly folded structure in solution (Figure 5.1). The observed chemical shift dispersion can only be obtained for a folded (native-like) structure. This relates particularly to the upfield ring current shifted methyl group resonances.

The work presented in this chapter was carried out in collaboration with Professor George McLendon of the University of Rochester and Jude Charman, a part II student [1]. The former was visiting Oxford as a Guggenheim Fellow.

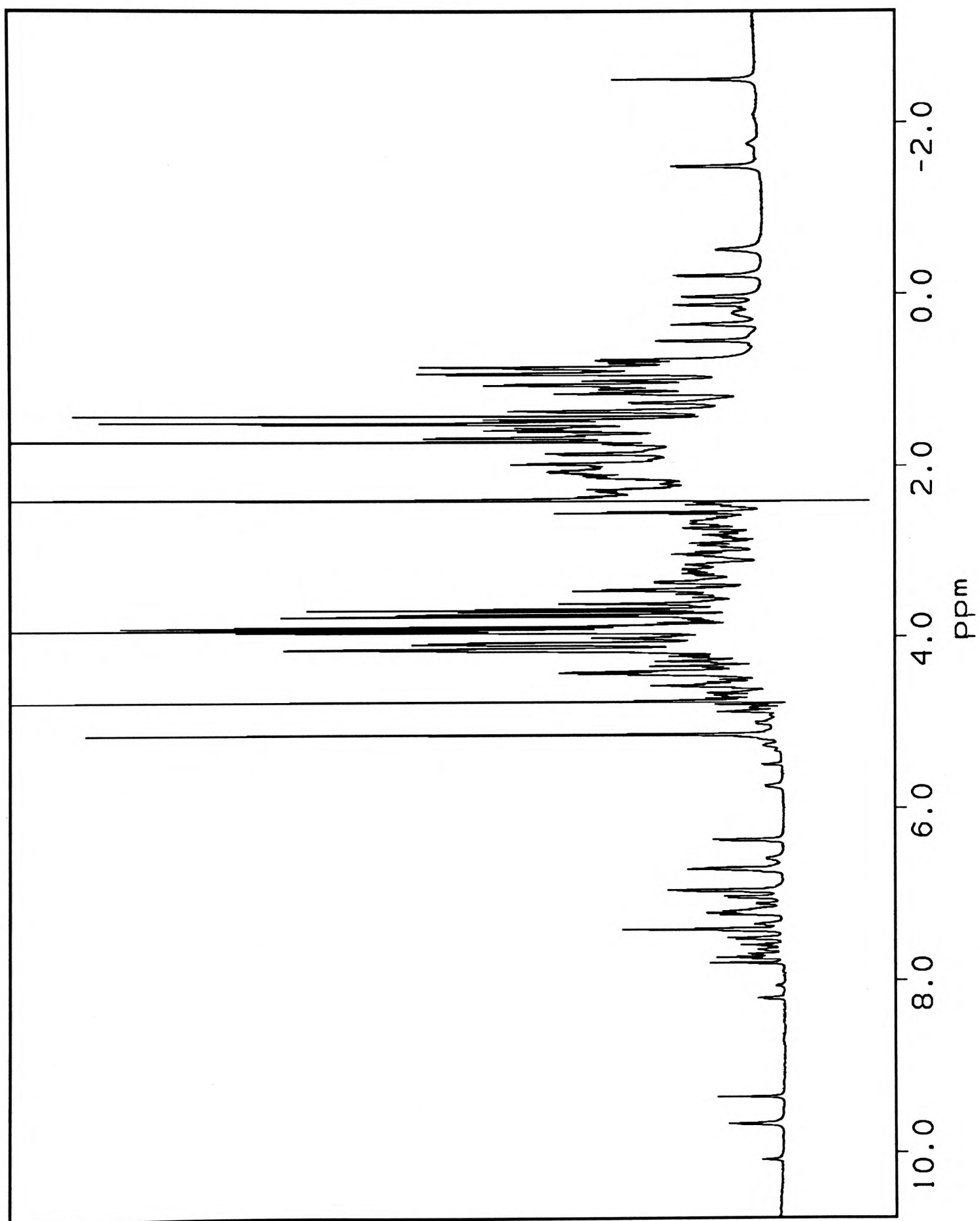


Figure 5.1: 500 MHz ^1H NMR spectrum of ferrous cytPCMH in $\text{D}_2\text{O}/20\text{mM } \text{P}_i$, $\text{pH}^* 7.0$ recorded at 37°C .

5.1 Assignment of Haem *Meso* Protons

The *meso* protons of the haem *c* prosthetic group are located at positions labelled α , β , γ and δ in Figure 5.2. These protons give resonances with characteristic downfield chemical shifts such that cross peaks involving them are easily identified (Figure 5.3). Thus, the haem *meso* protons were designated as suitable starting points for assignment of resonances due to other haem groups in the spectrum of diamagnetic ferrous cytPCMH.

Previous NMR studies [3, 4, 5, 6] of *c*-type cytochromes have employed NOE connectivities for the assignment of haem group resonances in a manner similar to the analysis presented here, which is most closely related to that of Moore [6]. Such methods are primarily applicable to diamagnetic, low spin ferrous haems. The axial ligands in both *b*- and *c*-type ferrous haems enforce the low spin state. NOE methods have also been applied to paramagnetic ferric haem groups [7].

Individual *meso* protons can be identified using predicted connectivity patterns, employing the nomenclature outlined in Figure 5.2. The predictions herein are based simply on the relative positions of groups in the haem and on the three dimensional structure of PCMH as determined by X-ray crystallography [8]. The α *meso* proton is close to haem methyl group 3, while the β *meso* proton is close to haem methyl group 5. Each of these *meso* protons should therefore show at least one intense NOE. The γ *meso* proton should not show a strong NOE to any methyl group since the closest is more than 5Å away. Instead, NOEs to (at least) four different single proton resonances from the propionate groups are expected around 2–3ppm. Finally, the δ *meso* proton should show strong NOEs to haem methyl groups 1 and 8.

With these predicted patterns in mind and using analogies with previous NMR studies of cytochromes, the haem *meso* proton resonances of cytPCMH may be assigned. For convenience, the *meso* proton resonances in the NOESY spectrum have been labelled 1–4 (Figures 5.3 and 5.4). Meso proton 1 at 10.18ppm does not exhibit any NOE to a haem methyl group, but does connect with four single proton peaks. This pattern clearly identifies the γ *meso* proton. Further, the only resonance for which there are two three proton (methyl) NOEs of roughly equal intensity is that of

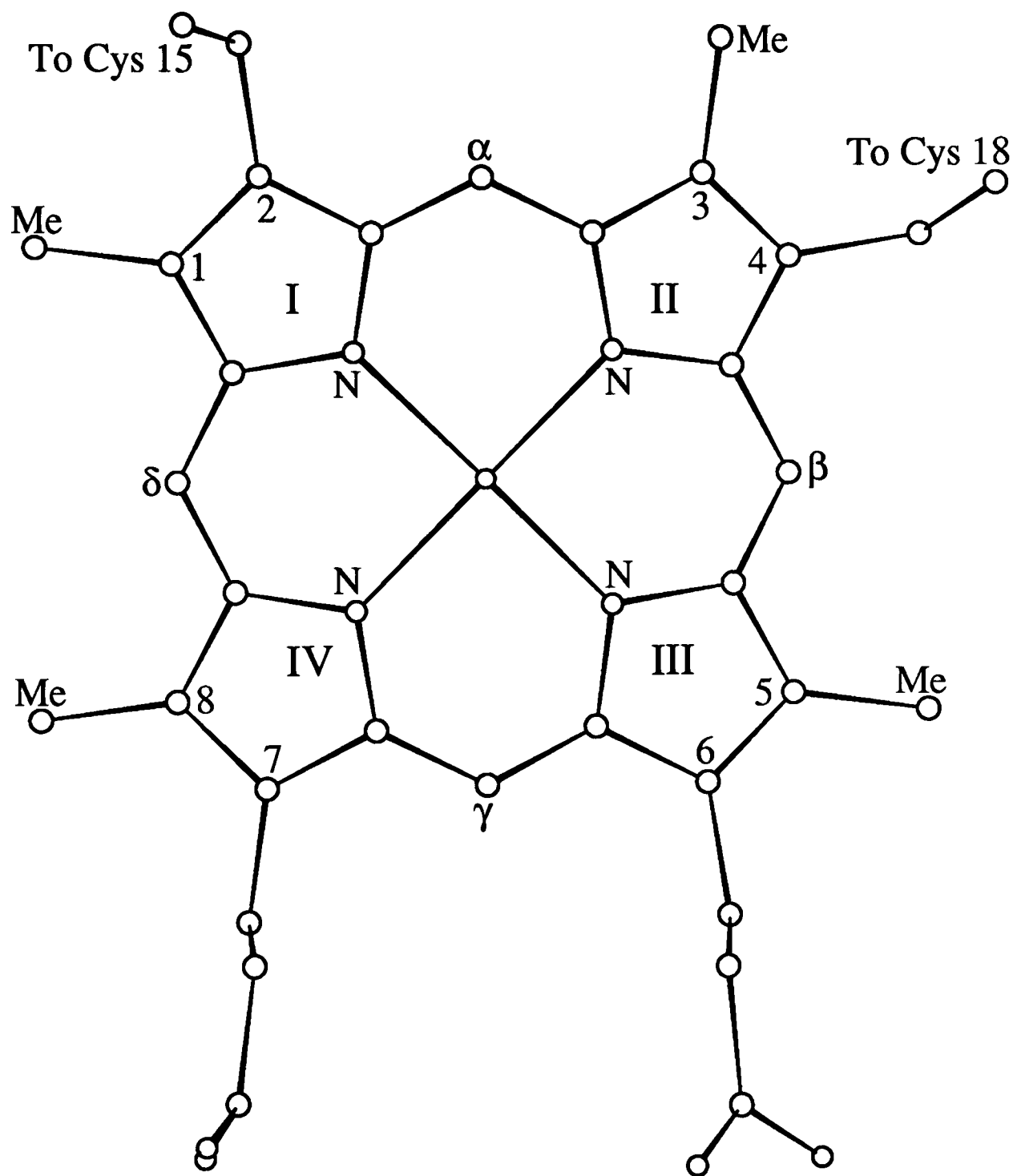


Figure 5.2: Nomenclature for the haem prosthetic group of cytPCMH.

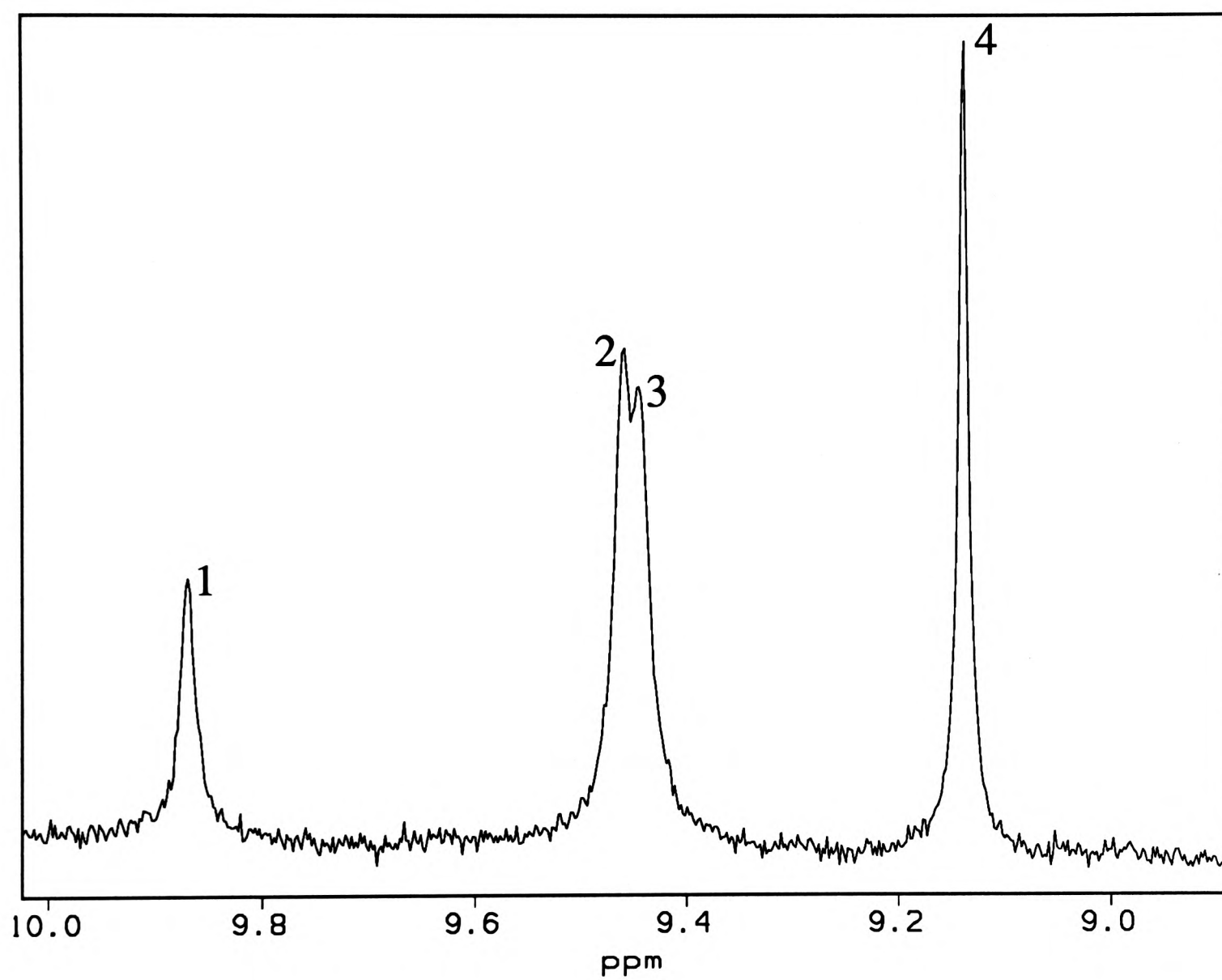


Figure 5.3: Haem *meso* proton region of the 500 MHz ^1H NMR spectrum of ferrous cytPCMH in $\text{D}_2\text{O}/20\text{mM } \text{P}_i$, $\text{pH}^* 7.0$ recorded at 37°C .

meso proton 2 at 9.48ppm. This is then tentatively assigned as the δ *meso* proton. That the cross peak at 2.5ppm is slightly less intense than that at 3.4ppm is consistent with more rapid relaxation, as is evident from the relatively large linewidth of the peak at 2.5ppm in the 1D spectrum.

The remaining signals at 9.15ppm and 9.46ppm may possibly be distinguished by the presence of a unique NOE involving *meso* proton 4 and a three proton ring current shifted resonance at -2.7 ppm. The latter is probably due to the $C^{\epsilon}H_3$ group of the axial ligand Met 50, since it shows no intraresidue coupling in the HOHAHA spectrum. In addition, this $C^{\epsilon}H_3$ group exhibits an NOE connectivity with the 3.25ppm methyl group resonance which is associated with *meso* proton 4. Therefore, the α and β *meso* protons might be assigned by determining which more closely approaches Met 50 $C^{\epsilon}H_3$.

The expected NOEs involving Met 50 $C^{\epsilon}H_3$ can be established from the crystal structure [8] of the intact flavocytochrome. Hydrogen atoms were added to the crystallographic coordinates of the cytochrome subunit, using the HBUILD feature of the program X-PLOR [9], by Dr. P. C. Driscoll. Only one intense NOE between Met 50 $C^{\epsilon}H_3$ and a *meso* proton is predicted, and this involves the γ *meso* proton. Moreover, it is expected from the crystal structure [8] of PCMH that no NOE will be observed between Met 50 $C^{\epsilon}H_3$ and any of the haem methyl groups: no haem methyl group is within $5\text{\AA}$ of Met 50 $C^{\epsilon}H_3$. Both predictions are inconsistent with the NMR data. In the solution phase, it appears that Met 50 $C^{\epsilon}H_3$ is close to (i.e. $\leq 5\text{\AA}$) *meso* proton 4 (α or β) and its associated methyl group, whereas in the crystal structure [8], Met 50 $C^{\epsilon}H_3$ is close to the γ *meso* proton. It is concluded that the axial ligand Met 50 has an orientation which is different in free, solution phase cytPCMH to that in intact PCMH in the crystalline phase.

5.2 Proposed Rearrangement of Met Ligand

Analysis of the NOESY spectrum of cytPCMH in conjunction with the crystal structure of PCMH provides a simple possible mechanism to account for the proposed difference in axial ligand orientation [10, 11]. Changing the torsion angle about the

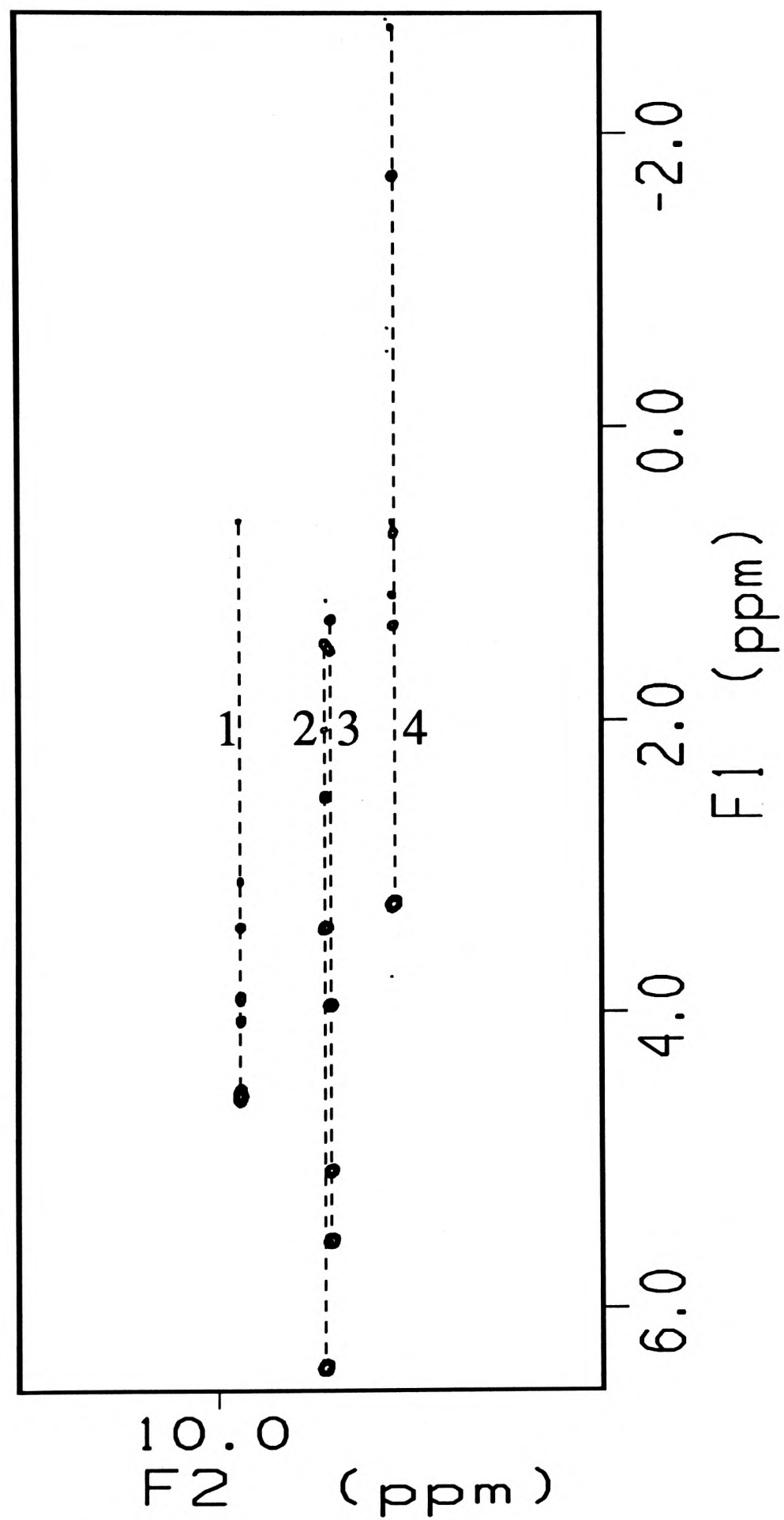


Figure 5.4: Contour plot of the haem *meso* proton region of the NOESY spectrum of ferrous cytPCMH in D₂O/20mM P_i, pH* 7.0 recorded at 35°C.

$C^\gamma-S^\delta$ bond of Met 50 causes the Met 50 $C^\epsilon H_3$ group to swing across the haem from a position close to the γ *meso* proton (where it is located in the crystal structure of the intact enzyme [8]) to a position near to the α *meso* proton (Figure 5.5). This new orientation is nearly 180° from that in the crystal structure, and leads to close contacts of Met 50 $C^\epsilon H_3$ with the α *meso* proton and haem methyl 1, in agreement with the NOE data. No simple change in torsion angle leads to the alternative Met 50 orientation in which the $C^\epsilon H_3$ group is close to the β *meso* proton. This suggests that *meso* proton resonance 4 should be assigned to the α *meso* proton.

The proposed change in orientation of axial ligand Met 50 on association or dissociation of the subunits of PCMH has a readily observable consequence: a large difference between ferric cytPCMH and ferric PCMH in the resonance frequencies of the paramagnetically shifted haem methyl groups. Previous results have shown that the haem methyl resonances are probes of the electron distribution about the central metal ion [12]. It is also known that in *c*-type cytochromes, the spatial arrangement of the Met ligand exerts strong influence on the electronic structure of the haem [13, 14, 15, 16, 17, 18]. The significant differences observed in the pattern of haem methyl resonances of ferric cytPCMH and ferric PCMH thus provide further evidence for a difference in orientation of Met 50 with consequent shift in distribution of unpaired electron spin density. Although other factors might contribute to the change in haem methyl chemical shifts, the large magnitude (up to 8ppm) of the differences suggests that there is a shift in the position of the haem *g*-tensors due to a change in Fe-ligand geometry.

A precedent for the proposed difference in axial ligand orientation exists in the comparison of eukaryotic cytochrome *c* and bacterial cytochrome c_{551} [5]. The axial methionine has a different spatial arrangement in these two cytochromes. One of the consequences is a difference in the pattern of haem methyl group hyperfine shifts in the respective ferricytochromes (Table 5.1) [3]. The pattern of shifts in ferricytochrome *c* indicates the presence of greater spin density on haem pyrrole rings II and IV [19]. This distribution has been rationalised [19] on the basis that it facilitates electron transfer, since pyrrole II is the most exposed in cytochrome *c*. In contrast, the more deshielded pair of methyl groups in ferricytochrome c_{551} is methyl 1 and methyl 5.

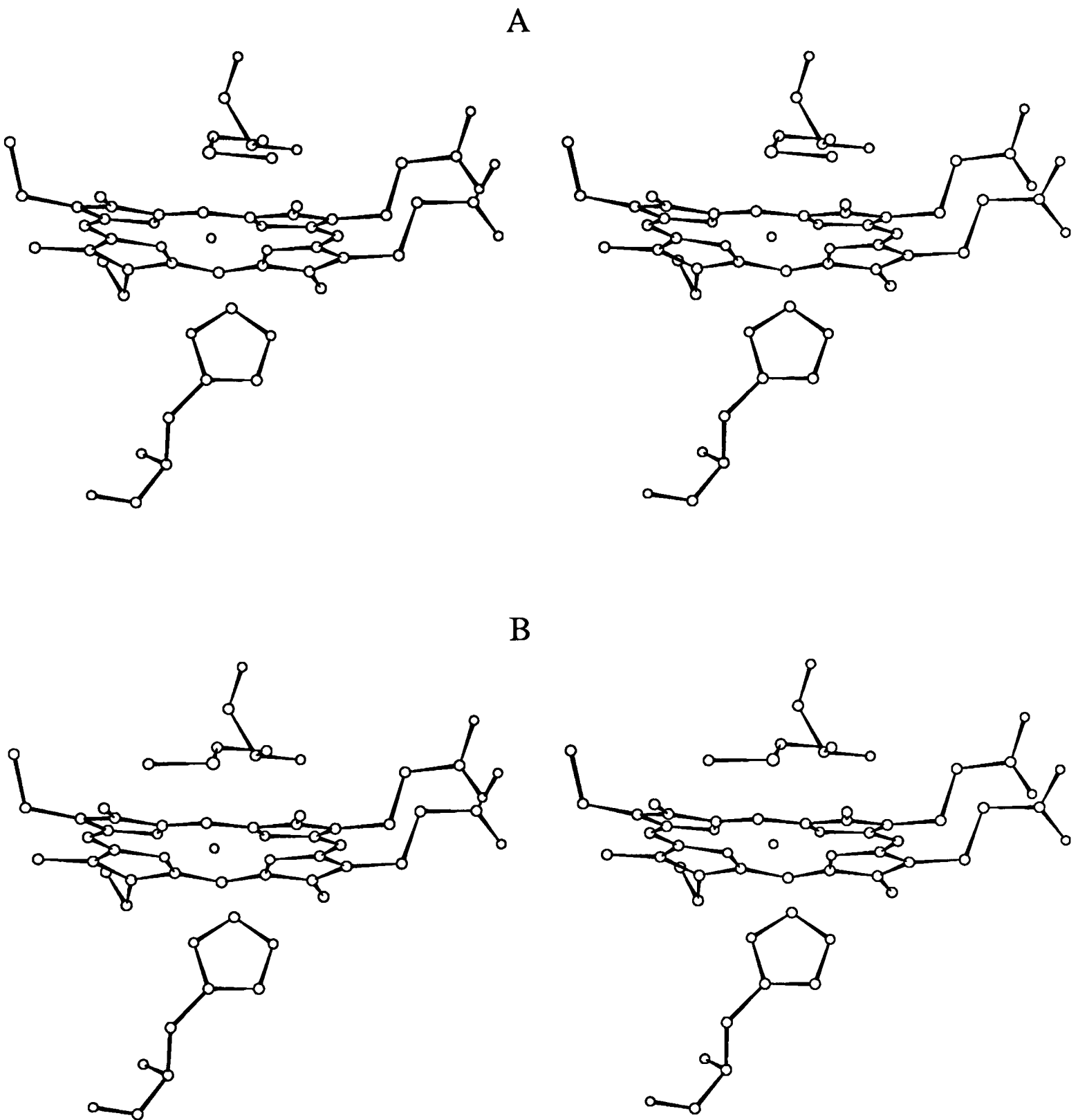


Figure 5.5: Stereo view of the haem group and ligands of cytPCMH, showing the proposed reorientation of Met 50, which lies above the haem in this figure. (A) Met 50 oriented as in the crystal structure of intact PCMH; (B) proposed orientation of Met 50 in unbound, solution phase cytPCMH.

Haem group	Chemical shift (ppm)	
	Horse heart cytochrome <i>c</i>	<i>P. aeruginosa</i> cytochrome <i>c</i> ₅₅₁
Ring methyl 1	3.9	21.1
Ring methyl 3	27.3	9.6
Ring methyl 5	6.9	28.8
Ring methyl 8	31.7	14.4

Table 5.1: The hyperfine shifts of the haem methyl groups in horse heart ferricytochrome *c* (35°C) and *Pseudomonas aeruginosa* ferricytochrome *c*₅₅₁ (27°C).

This indicates that pyrroles I and III receive the most unpaired electron spin density. This difference in electronic structure relative to that in eukaryotic cytochrome *c* has been ascribed to a 90° rotation in the orientation of the principal axes of the *g*-tensor [3]. Noting that the crystal structure of *P. aeruginosa* cytochrome *c*₅₅₁ indicates that pyrrole III is the most exposed [20], the functional interpretation of the ferricytochrome *c*₅₅₁ haem electronic structure is the same as that made for eukaryotic cytochrome *c* — that electron density is directed towards the most exposed haem edge to facilitate reaction with redox partners. A similar functional shift in electron density may accompany the proposed reorientation of the axial methionine which occurs when the cytochrome and flavoprotein subunits come together to form PCMH.

The chirality of the axial methionine ligand can be established by a combination of NOE and circular dichroism (CD) experiments. The methionine ligands of over twelve *c*-type cytochromes have been characterised in this way. The results show a strong correlation between methionine chirality and the identity of the high frequency haem methyl resonance pair: R chirality is associated with the 8,3 methyl groups as the more deshielded pair, S chirality with the 1,5 methyl groups as the more deshielded pair. This reflects the influence of the spatial structure of the axial methionine on the electronic structure of *c*-type haems.

A second, related effect of the change in ligand orientation can be predicted. As pointed out by Moore *et al.* [21], a change in haem electronic structure will have an effect on the cytochrome redox potential. Such an effect is indeed observed: cytPCMH in the intact flavocytochrome reduces at 254mV *vs.* NHE [22], whereas isolated cytPCMH is more difficult to reduce, having a reduction potential of −179mV *vs.* NHE [23]. This shift in reduction potential favouring Fe^{III} indicates that the methion-

ine orientation in isolated cytPCMH is such as to diminish Fe-S back bonding. The cytochrome reduction potential will depend on many other factors. An example is solvation, which probably changes on dissociation of the subunits of PCMH. The portion of the observed shift in reduction potential to be ascribed to the change in Fe-ligand bonding is therefore uncertain.

In summary, assembly of PCMH from its subunits leads to a directly observable change in active site stereochemistry, with consequent effects on bonding and reactivity. Further data are needed in order to delineate the mechanism by which subunit interactions lead to change in the spatial arrangement of axial ligand Met 50. It is noteworthy that in the crystal structure of PCMH [8], the haem, and Met 50 in particular, are located close to the subunit interface. This has significance in that, as Stryer has noted [24], interfaces between different polypeptide chains play a critical role as molecular switches in mediating allosteric effects in such proteins as hemoglobin and ATCase. It is hoped that a complete three dimensional structure determination of cytPCMH will be possible in the future. This has been precluded by the low quality of spectra so far obtained in H₂O.

5.3 Polypeptide Resonance Assignments

The first half of this section is concerned with the identification of the resonances due to the aromatic ring protons of the tyrosine, phenylalanine and tryptophan residues of cytPCMH. The second half of the section focuses on the assignment of the AMX spin systems ($C^\alpha H-C^\beta H_2$) of the side chains of aromatic residues. Subsequent identification of backbone amide protons *via* NH- $C^\alpha H$ couplings will provide a starting point for full sequential assignment.

5.3.1 Delineation of Aromatic Spin Systems

A full analysis (with the exception of histidine residues) of the aromatic region of the spectrum in D₂O has been carried out.

CytPCMH contains nine aromatic amino acid residues: two histidines, four tyrosines, two phenylalanines and one tryptophan. One of the histidines is an axial

ligand of the haem iron and will have an upfield shifted resonance. The cross peak of the imidazole ring of the remaining histidine is not observed in the P.COSY or HOHAHA spectra. This is probably because the scalar coupling between the C^εH and C^δH of the ring is smaller than the linewidth.

In the 1D spectrum of cytPCMH, peaks in the aromatic chemical shift range have been labelled A1 to A15 (Figure 5.6). The fine structure of these signals, caused by scalar coupling, has been used to provide confirmatory evidence for the assignments. Thus, the C^δ and C^ε protons of tyrosine give rise to a doublet resonance in a 1D spectrum. The C^δ protons of phenylalanine give rise to a doublet, whereas the C^ε protons and C^ζ proton resonances are two and one proton triplets respectively. In tryptophan, protons 4 and 7 give rise to one proton doublets and protons 5 and 6 give rise to one proton triplets.

The HOHAHA spectrum contains all of the cross peaks observed in the P.COSY spectrum, plus five additional cross peaks, marked * (Figure 5.7). These relay peaks are associated with the Phe and Trp spin systems. In the relay peaks of the Phe spin system, the C^δ protons are correlated with the C^ζ proton, whilst in the Trp spin system, relay cross peaks correlate protons 4 with 7 and 5 with 7.

Two of the four tyrosine spin systems can be identified from the P.COSY spectrum. These are (A1,A5) and (A3,A4), both of which show the expected two proton doublet-two proton doublet pattern in the 1D spectrum. Distinction of the C^δ and C^ε protons cannot be made at this stage.

Analysis of the 1D spectra recorded at different temperatures indicates that certain of the aromatic resonances are degenerate at the temperature (35°C) at which the 2D spectra were recorded. For example, peak A3 is resolved into two separate peaks at 47°C — a doublet at 6.50ppm and a triplet at 6.55ppm. A3 thus turns out to be a four proton peak arising from both Phe and Tyr protons. Once this becomes clear, the completion of the assignment of this Phe spin system becomes trivial in the HOHAHA spectrum. It corresponds to peaks A4, A3 and A13 in the 1D spectrum (Figure 5.6). The two remaining tyrosines are easily identified in the HOHAHA spectrum where distinction between cross peaks is clearer than in the P.COSY spectrum. These tyrosine cross peaks correspond to the pairs (A8,A10) and (A8,A11) (Figure 5.6).

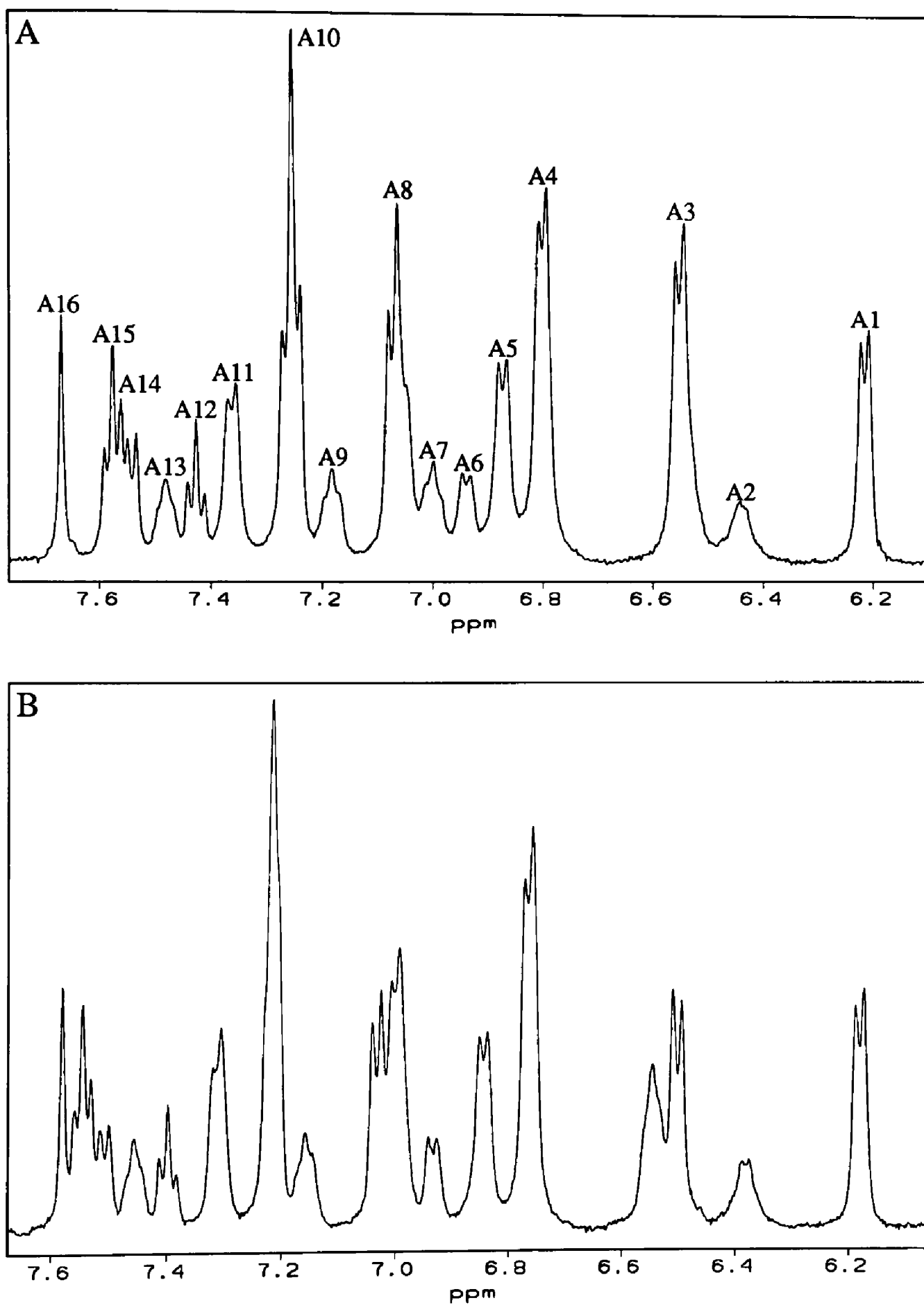


Figure 5.6: Aromatic region of the 500 MHz spectrum of ferrous cytPCMH in D₂O/20mM P_i, pH* 7.0 recorded at (A) 35°C and (B) 47°C. The resonance labels referred to in the text are shown.

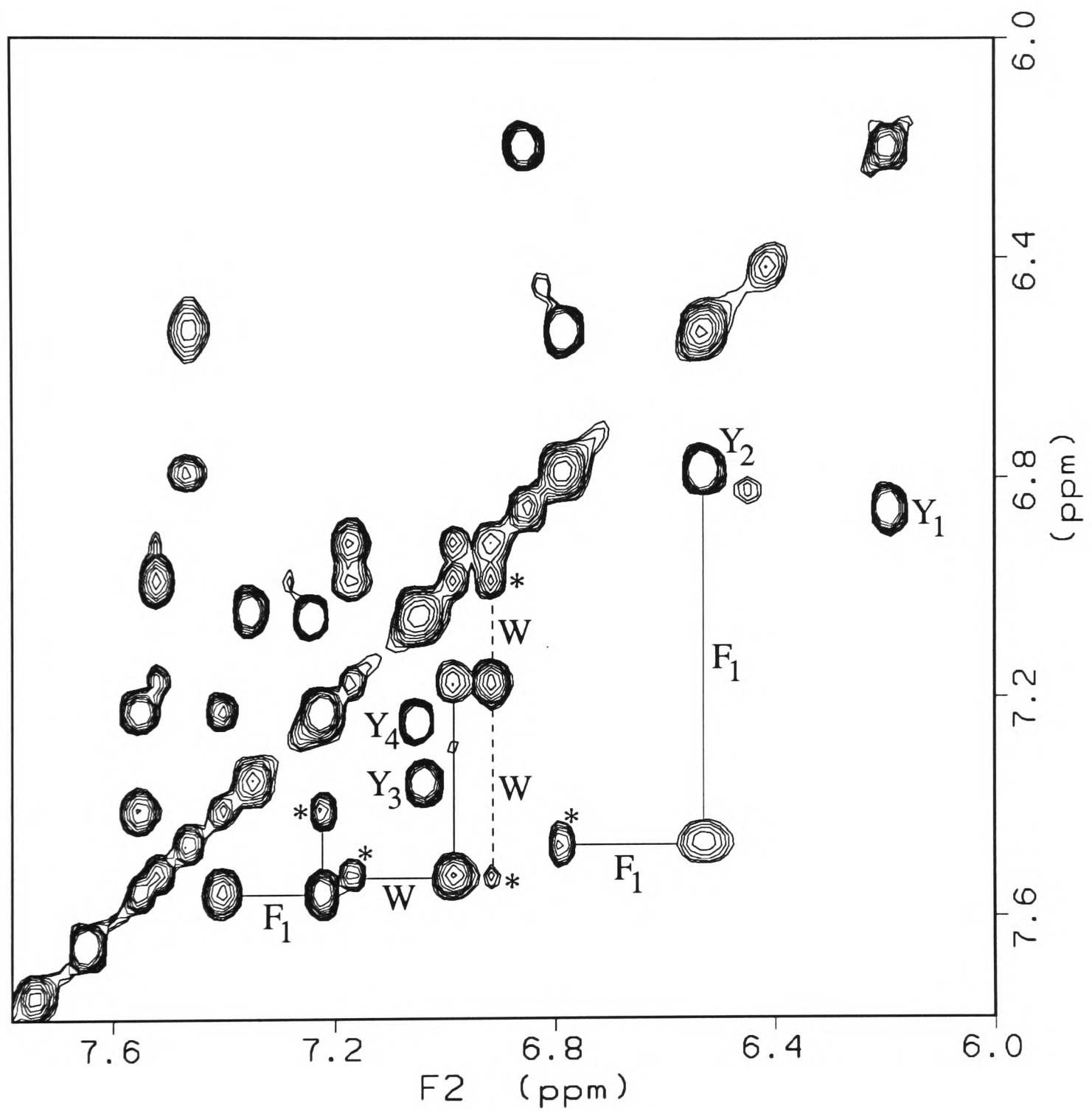


Figure 5.7: Aromatic region of the HOHAHA spectrum of ferrous cytPCMH in $D_2O/20mM P_i$, $pH^* 7.0$ recorded at $35^\circ C$. The spin system labels referred to in the text are shown.

Thus, a second case of chemical shift degeneracy at 35°C is observed. In the 1D spectrum recorded at 47°C, A8 is resolved into two doublet peaks at 7.01ppm and 7.03ppm respectively. In order to distinguish between the tyrosines, the nomenclature Y₁ for (A1,A5), Y₂ for (A3,A4), Y₃ for (A8,A11) and Y₄ for (A8,A10) is adopted.

Of the remaining aromatic spin systems, the distinctive coupling pattern of the single tryptophan residue is easily discerned in the HOHAHA spectrum. These cross peaks correspond to A14, A7, A9 and A6 in the 1D spectrum (Figure 5.6), which show the expected multiplicities. The second Phe spin system is also readily identifiable in the HOHAHA spectrum and corresponds to peaks A15, A10 and A12 in the 1D spectrum (Figure 5.6). This Phe spin system is labelled F₂, while that assigned above (corresponding to A4, A3 and A13 in Figure 5.6) is labelled F₁.

These assignments permit tentative inferences concerning structure. Analysis of the NOESY spectrum indicates that a number of aromatic residues lie in close proximity (<4Å) to one another. The following interresidue NOEs have been identified:

- A5 of Tyr(Y₁) with A3 of Tyr(Y₂)
- A1 of Tyr(Y₁) with A3 of Tyr(Y₂)
- A11 of Tyr(Y₃) with A4 of Phe(F₁)
- A8 of Tyr(Y₃) with A4 of Phe(F₁) (C^δ protons)
- A8 of Tyr(Y₃) with A3 of Phe(F₁) (C^ε protons)

In the three dimensional structure of intact PCMH [8], tyrosines 11 and 69 of the cytochrome subunit are in close proximity (Figure 5.8). On the basis of the first two NOEs in the above list, therefore, it might be expected that Y₁ and Y₂ correspond to these two residues. The last three NOEs in the list would not be expected on the basis of the crystal structure of intact PCMH, since no Tyr and Phe are in sufficiently close proximity to one another to observe NOEs between them. This furnishes evidence, additional to that from the pattern of haem *meso* proton resonances, for a local conformational change in the isolated cytochrome subunit relative to the conformation observed in the crystalline enzyme.

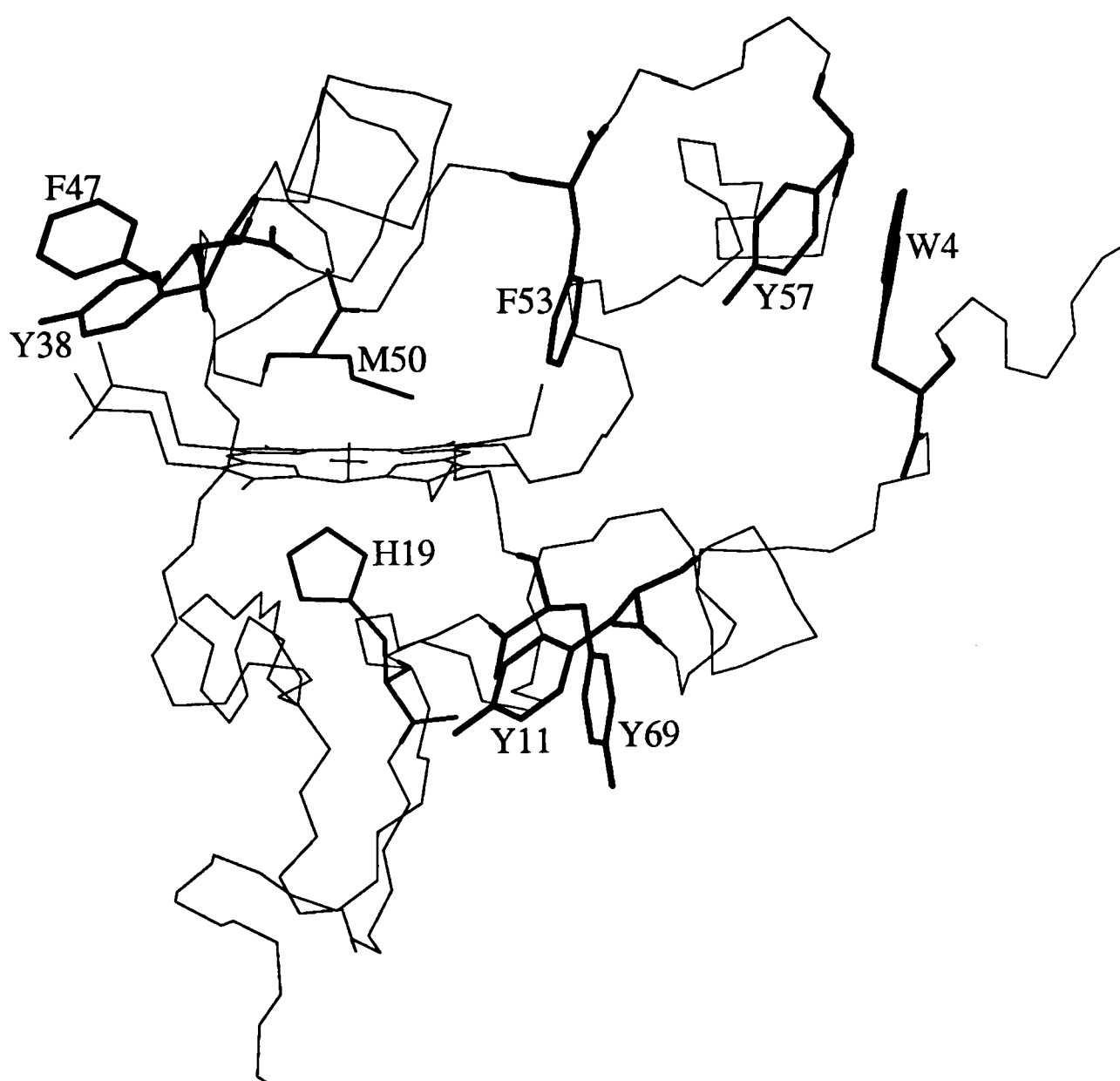


Figure 5.8: The polypeptide backbone of the cytochrome subunit of PCMH. Aromatic side chains and the Fe ligands are highlighted.

Speculative investigation of the three dimensional structure of PCMH identifies Phe 53 of the cytochrome subunit as a potentially important residue in any conformational change on binding/dissociation of the subunits. In intact, crystalline PCMH, Phe 53 lies at the interface with the flavoprotein, near haem *meso* proton α and on the same side of the haem as Met 50, the axial ligand. The proposed reorientation of Met 50 to a position close to *meso* proton α on dissociation of the subunits will consequently require a change in position of Phe 53. The presence of an NOE between Phe(F₁) and Met 50 C $^{\epsilon}$ H₃ indicates that the Phe(F₁) spin system is that of Phe 53, and further supports the proposed reorientation of the axial methionine ligand. Consideration of the above NOE list indicates that Tyr(Y₃) may be assigned to Tyr 57, since no other tyrosine residue could possibly be sufficiently close to Phe 53 for the observation of NOEs — Tyr 38 lies across the haem from Phe 53. Additional evidence for the change in position of Phe 53 accrues from the presence of a weak NOE between it and *meso* proton β . This would not be expected from the crystallographic coordinates of PCMH, and indicates that in cytPCMH, Phe 53 is located in the vicinity of both α and β *meso* protons.

The remaining tyrosine and phenylalanine residues may now be assigned: Phe(F₂) is Phe 47, and Tyr(Y₄) is Tyr 38. The assignments of the aromatic residues of cytPCMH are given in Table 5.2.

5.3.2 Assignment of C $^{\alpha}$ H-C $^{\beta}$ H₂ of Aromatic residues

The ring protons of aromatic residues form essentially isolated spin systems. Scalar coupling of the C $^{\delta}$ H₂ ring protons with C $^{\beta}$ H₂ is only rarely observed. Attachment of the ring to the rest of the side chain can be made using through space connectivities. For example, plots of the intraresidue $d_{\beta ring}^2$ ^1H - ^1H distances as a function of the torsion angle χ_i^2 [25] show that in phenylalanine, tyrosine and tryptophan at least one ^1H - ^1H distance between C $^{\beta}$ H₂ and the ring is less than 3.0Å for all values of χ_i^2 . Generally, the C $^{\alpha}$ H-C $^{\beta}$ H₂ fragments of aromatic spin systems can be distinguished from other AMX spin systems by the observation of strong NOEs between aromatic ring protons and at least two of backbone NH, C $^{\alpha}$ H and C $^{\beta}$ H₂. In tryptophan the distinguishing NOE may involve ring proton 4 or 2 depending on the torsion angle,

Residue	Spin system	C ^α H	C ^β H	Ring
Trp 4				C ₄ H 7.51; C ₅ H 6.97 C ₆ H 7.15; C ₇ H 6.91
Tyr 11/69	Y ₁	3.79	2.85,3.06	C ^δ H 6.84,6.84 C ^ε H 6.18,6.18
Tyr 11/69	Y ₂	4.26/4.82	2.43,2.60/2.88,3.80	C ^δ H 6.76,6.76 C ^ε H 6.53,6.53
Tyr 38	Y ₄	4.50/4.60	1.19,2.11/3.15,3.41	C ^δ H 7.22,7.22 C ^ε H 7.13,7.13
Phe 47	F ₂	4.50/4.60	1.19,2.11/3.15,3.41	C ^δ H 7.22,7.22 C ^ε H 7.55,7.55 C ^ζ H 7.39
Phe 53	F ₁	4.26/4.82	2.43,2.60/2.88,3.80	C ^δ H 6.76,6.76 C ^ε H 6.53,6.53 C ^ζ H 7.45
Tyr 57	Y ₃	3.9	0.62,1.32	C ^δ H 7.13,7.13 C ^ε H 7.33,7.33

Table 5.2: ¹H NMR chemical shifts of aromatic residues of ferrous cytPCMH (20mM P_i, pH* 7.0, 35°).

and in phenylalanine and tyrosine both C^δ protons are within 3Å of C^βH₂. In these two amino acids, the C^δH₂-C^βH₂ NOEs will be more intense than the corresponding C^εH₂-C^βH₂ NOEs. This fact can be used to distinguish the C^δ and C^ε protons, which is particularly useful in assignment of tyrosine spin systems. Location of the C^βH₂ resonances in this way facilitates subsequent establishment of C^αH and NH chemical shifts using P.COSY and HOHAHA spectra. Confirmation of the correct attachment of ring proton to C^αH-C^βH₂ fragments comes in the second stage assignment. The assignments of the aromatic spin systems of cytPCMH are given in Table 5.2.

5.3.3 Methyl Group Region: 1. Val, Leu and Ile

The upfield region of the 1D spectrum of cytPCMH contains several relatively intense resonances due to ring current shifted methyl groups (Figure 5.9). These provide a convenient point for starting the assignment of the aliphatic region of the spectrum since they are shifted away from the main resonance envelope.

Cross peak patterns corresponding to the unique spin systems of valine and leucine

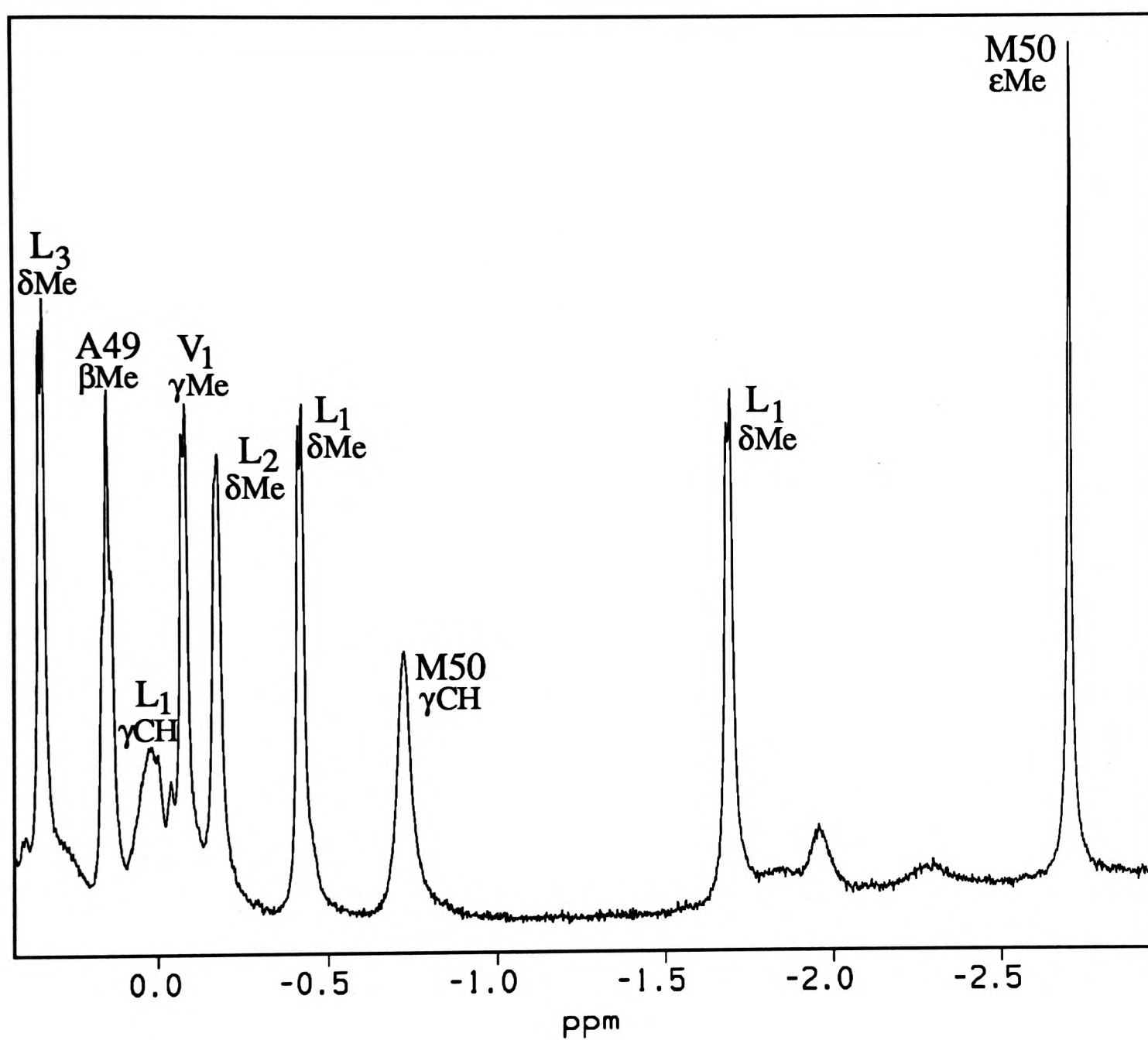


Figure 5.9: Upfield region of the 500 MHz spectrum of ferrous cytPCMH in D₂O/20mM P_i, pH* 7.0 recorded at 37°C.

have been located by combined use of P.COSY and HOHAHA spectra. As in the case of aromatic spin systems of cytPCMH discussed above, corroborating evidence for aliphatic assignments has been sought from multiplicity patterns in the 1D spectra (methyl groups of valine and leucine give rise to doublet signals). Splittings due to spin-spin coupling are often unresolved at 35°C, but are clearly observable at higher temperatures.

Valine and leucine spin systems were identified by first locating pairs of cross peaks corresponding to the methyl groups. In the P.COSY spectrum the peak shape can be used as an aid in identification, since peaks arising from methyl groups are lengthened in one direction compared with those from single proton groups. The two types of spin system are then most readily distinguished by the observation of $C^\alpha H-C^\gamma H_3$ relay cross peaks of valines in the HOHAHA spectrum. These have not been observed in the currently available HOHAHA spectra of cytPCMH. Instead, pairs of $C^\beta H-C^\gamma H$ cross peaks of leucine spin systems were sought. From these, the $C^\alpha H-C^\beta H$ pair of signals can often be located, resulting in identification of the complete non-labile part of the leucine spin system. In this way, tracing the spin systems back from the methyl group cross peaks, the non-labile protons of three of the six leucines and two of the seven valines have been assigned (Figure 5.10). The resonances of the other valine and leucine residues are in the crowded region of the spectrum between 0.5 and 2.4ppm and remain to be assigned.

The two isoleucine residues of cytPCMH have not been assigned. This is despite the fact that this amino acid has a unique spin system and its $C^\delta H_3$ group gives a triplet resonance, unlike the methyl groups of valine and leucine residues, which give doublets.

As mentioned previously, the three proton singlet at -2.7ppm has been assigned to Met 50 $C^\epsilon H_3$. This group does not give rise to any cross peaks in the P.COSY or HOHAHA spectra, but there is an NOE connectivity involving it and a resonance at -0.7ppm. The latter has been tentatively assigned to $C^\gamma H$ of Met 50.

The single remaining unassigned ring current shifted resonance is at 0.18ppm. This is a three proton signal which begins to show splitting into a doublet. In both the P.COSY and HOHAHA spectra, the spin system to which this resonance belongs

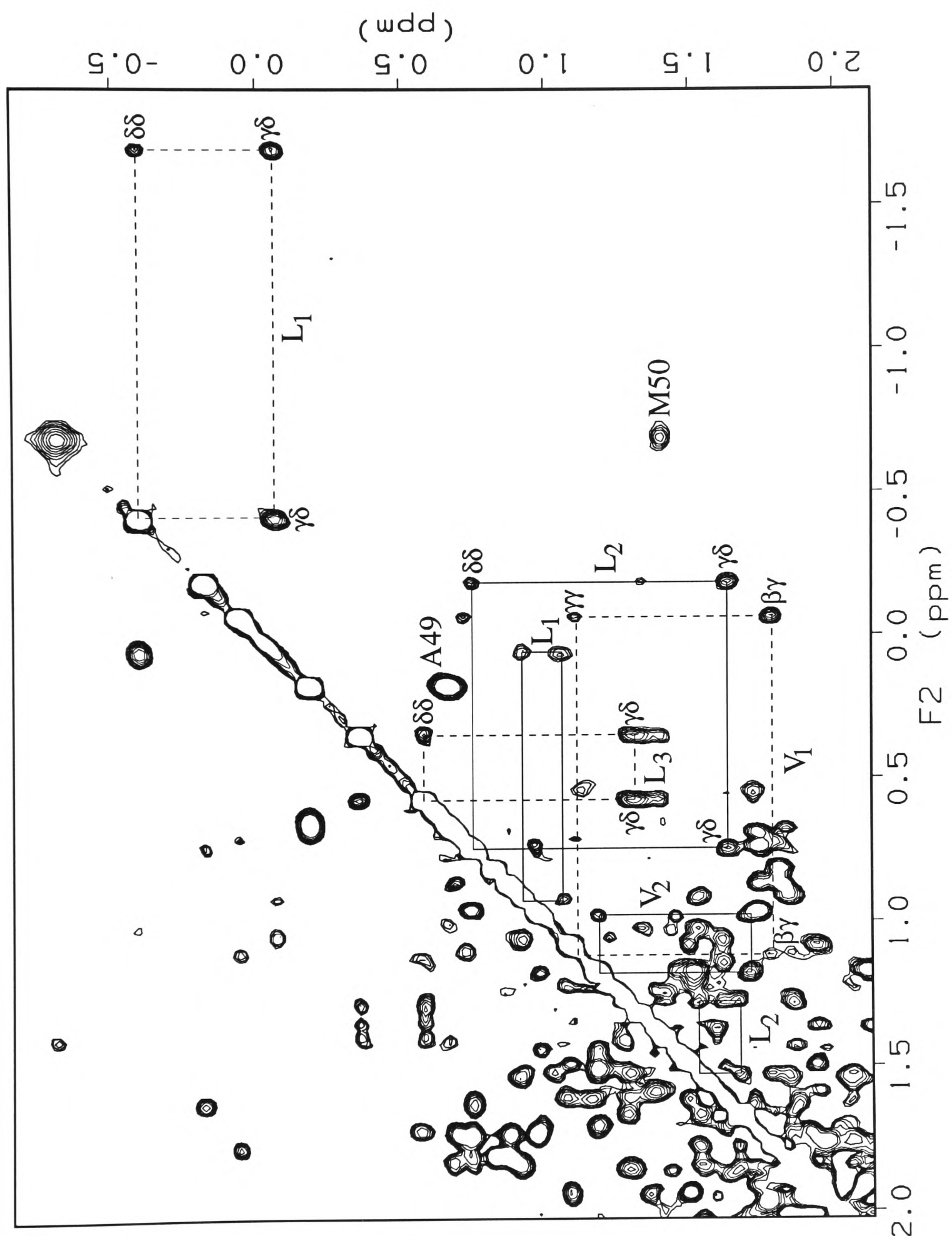


Figure 5.10: Methyl group region of the HOHAHA spectrum of ferrous cytPCM in $D_2O/20mM P_i$, $pH^* 7.0$ recorded at $35^\circ C$. The spin system labels referred to in the text are shown.

consists of a single, intense, cross peak at (0.65ppm,0.18ppm). This spin system is consequently assigned as an upfield shifted alanine residue. Analysis of the crystal structure of PCMH [8] indicates that Ala 49 is a candidate for this residue, since it lies close to the haem ring and would therefore experience the observed large ring current shift. This assignment is further supported by the presence of NOEs at 0.65ppm with Met 50 C ϵ H₃ and C γ H₂ groups.

5.3.4 Methyl Group Region: 2. Ala and Thr

The haem ring current shifted C α H and C β H₃ resonances of Ala 49 have been assigned above. CytPCMH contains six more alanine residues, plus one threonine. Six cross peaks which possess the characteristic intensity of alanine C α H-C β H₃ resonances are indicated in the D₂O HOHAHA spectrum (Figure 5.11). These are tentatively assigned as belonging to alanine spin systems. The peak at (1.28ppm,4.4ppm) seems to consist of two near degenerate signals superimposed on one another. These were initially labelled as cross peaks 6 and 7. Further support for these spin system assignments is provided by analysis of the HOHAHA spectrum recorded in H₂O (Figure 5.12). This contains HN-C α H cross peaks, along with probable HN-C β H₃ relays, at C α H chemical shifts corresponding to those of the C α H-C β H₃ cross peaks identified in Figure 5.11. A candidate for the single threonine spin system is also observed: this is one of the spin systems previously labelled as 6 and 7. It possesses a possible HN-C β H cross peak in the expected chemical shift range, with its HN-C γ H₃ cross peak degenerate with that of alanine spin system 6.

5.3.5 Glycines

Six of the eight glycines have been tentatively assigned on the basis of the usual characteristics in COSY-type spectra: square C α H-C α H cross peaks, lack of passive couplings and the large geminal coupling constants ($^2J_{\alpha\alpha}\approx 15\text{Hz}$).

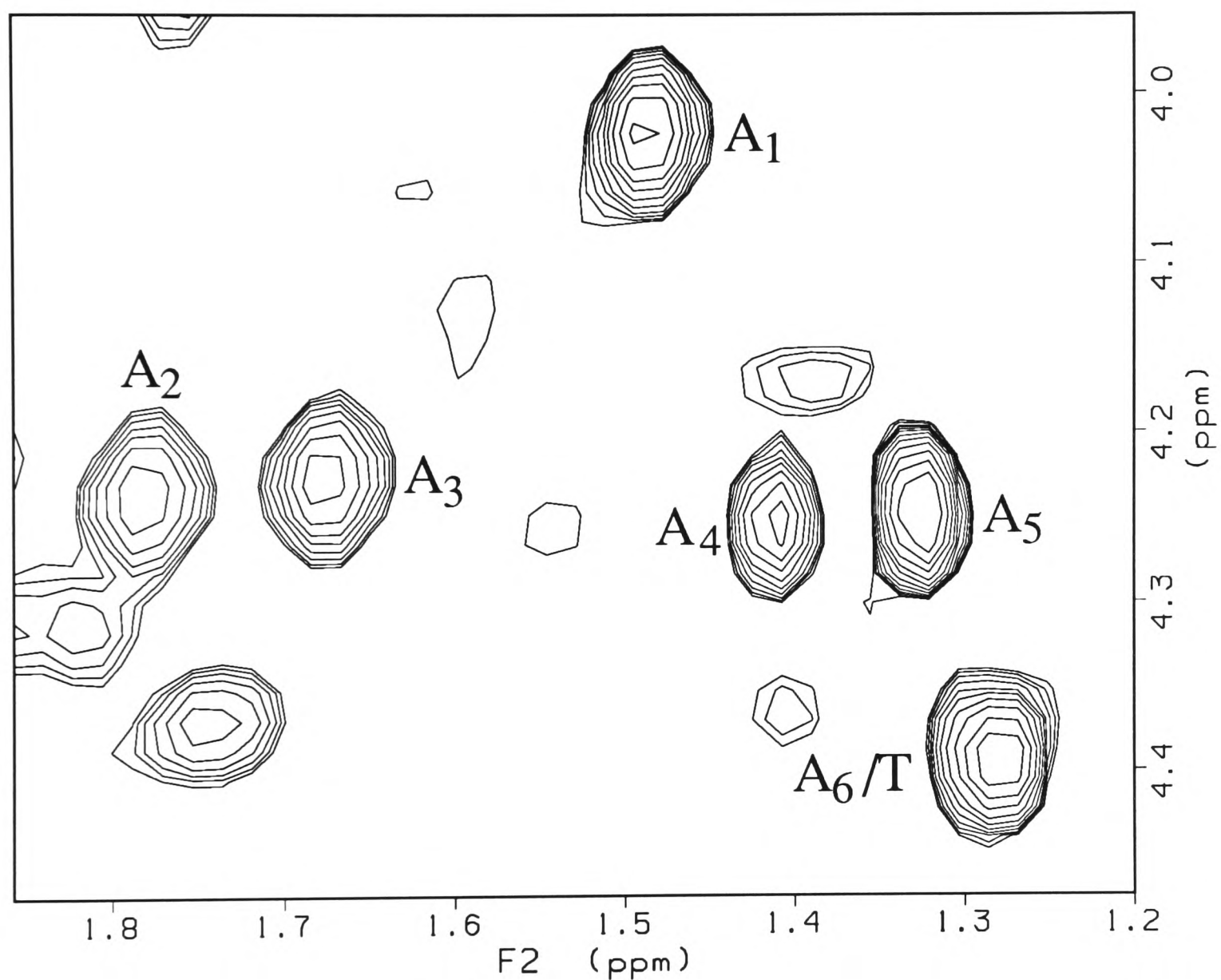


Figure 5.11: Ala/Thr region of the HOHAHA spectrum of ferrous cytPCMH in D₂O/20mM P_i, pH* 7.0 recorded at 35°C. The spin system labels referred to in the text are shown.

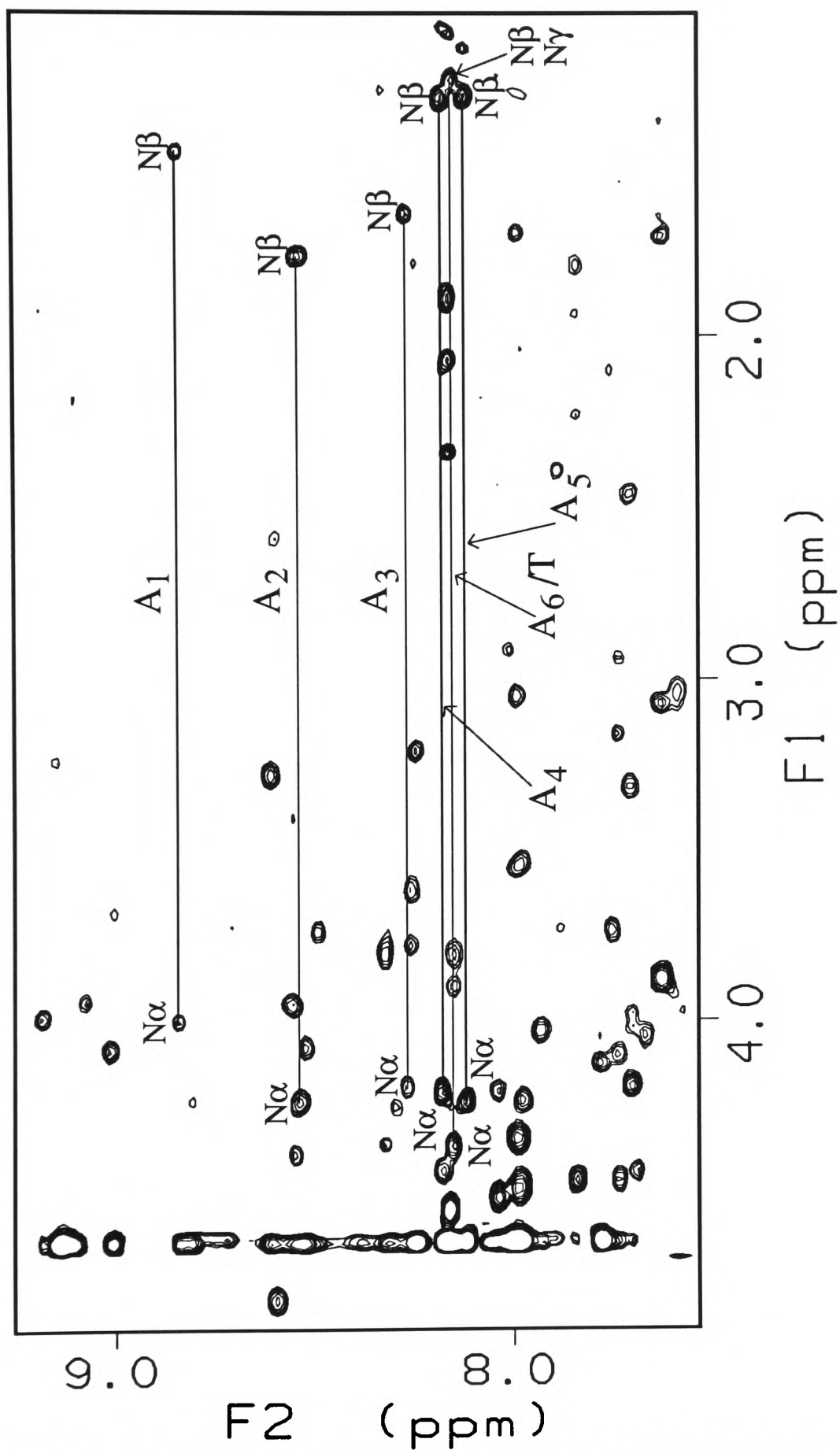


Figure 5.12: Part of a HOHAHA spectrum of ferrous cytPCMH in H₂O/20mM P_i, pH* 7.0 recorded at 35°C. Three bond NH-C^αH correlations and relayed NH-C^βH₃/NH-C^γH₃ correlations are indicated using the spin system labels referred to in the text.

5.4 Overview

The results presented in this chapter constitute the initial steps in the determination of the three dimensional structure of the cytochrome subunit of PCMH. Future efforts must be concentrated on improving the quality of H_2O spectra, so that full sequential assignments can be obtained.

While the cytPCMH structure will have some intrinsic interest, its main importance may belong in the wider context of understanding the cooperativity between the subunits of PCMH. Given the high M_r of PCMH, NMR by itself will be insufficient to fully characterise subunit interactions. X-ray crystallography and site directed mutagenesis studies are either underway or in preparation in the laboratories of McIntire (VA Medical Center, San Francisco) and Mathews (Washington University Medical Center, St. Louis).

Future NMR experiments may be directed towards the study of the azurin which is postulated to be the redox partner of PCMH in *P. putida*. Eventually, it may become feasible to investigate the interaction between PCMH and azurin. This would probably require a detailed knowledge of the structures of both proteins and specific amino acid labelling in PCMH. In the absence of the latter, the hyperfine shifted haem resonances of oxidised PCMH could provide a handle for studying this protein-enzyme complex.

Beyond the scope of this thesis, there are many intriguing questions relating to PCMH. For example, what structural elements are necessary for substrate oxidation? What is the pathway for electron transmission between the flavin and haem groups? Does covalent attachment of the flavin require another enzyme or is it self-catalytic? It can be seen that the work presented in this chapter has only begun to answer one of a battery of questions on PCMH.

Bibliography

- [1] J. A. Charman (1990) Part II thesis, University of Oxford.
- [2] W. McIntire & T. P. Singer (1982) *FEBS Lett.* 143, 316–318.
- [3] R. M. Keller & K. Wüthrich (1978) *Biochem. Biophys. Res. Commun.* 83, 1132–1139.
- [4] M. Smith & G. L. McLendon (1981) *J. Am. Chem. Soc.* 103, 4912–4921.
- [5] H. Senn, M. Billeter & K. Wüthrich (1984) *Eur. Biophys. J.* 11, 3–15.
- [6] G. R. Moore (1985) *Biochim. Biophys. Acta* 829, 425–429.
- [7] S. J. McLachlan, G. N. La Mar & K.-B. Lee (1988) *Biochim. Biophys. Acta* 957, 430–445.
- [8] F. S. Mathews, Z. Chen, H. D. Bellamy & W. S. McIntire (1991) *Biochemistry* 30, 238–247.
- [9] A. T. Brünger (1988) *X-PLOR Manual*, Yale University, New Haven, CT.
- [10] G. L. McLendon (1990) personal communication.
- [11] G. L. McLendon, S. Bagby, J. A. Charman, P. C. Driscoll, W. S. McIntire, F. S. Mathews & H. A. O. Hill (1991) *Proc. Natl. Acad. Sci. USA*, in press.
- [12] A. Mayer & H. Eicher (1984) *J. Mol. Catal.* 23, 151–161.
- [13] H. Senn, R. M. Keller & K. Wüthrich (1980) *Biochem. Biophys. Res. Commun.* 92, 1362–1369.

- [14] H. Senn, A. Eugster & K. Wüthrich (1983) *Biochim. Biophys. Acta* 743, 58–68.
- [15] H. Senn, F. Guerlesquin, M. Bruschi & K. Wüthrich (1983) *Biochim. Biophys. Acta* 748, 194–204.
- [16] H. Senn & K. Wüthrich (1983) *Biochim. Biophys. Acta* 743, 69–81.
- [17] H. Senn & K. Wüthrich (1983) *Biochim. Biophys. Acta* 746, 48–60.
- [18] H. Senn & K. Wüthrich (1983) *Biochim. Biophys. Acta* 747, 16–25.
- [19] R. M. Keller & K. Wüthrich (1978) *Biochim. Biophys. Acta* 533, 195–208.
- [20] Y. Matsuura, T. Takano & R. E. Dickerson (1982) *J. Mol. Biol.* 156, 389–409.
- [21] G. R. Moore, C. G. S. Eley & R. J. P. Williams (1984) in *Advances in Inorganic and Bioinorganic Mechanisms 3* (A. G. Sykes, Ed.) pp. 1–96, Academic Press, New York.
- [22] D. J. Hopper (1983) *FEBS Lett.* 161, 100–102.
- [23] L.-H. Guo, H. A. O. Hill, G. A. Lawrance, G. S. Sanghera & D. J. Hopper (1989) *J. Electroanal. Chem.* 266, 379–396.
- [24] L. Stryer, ‘Biochemistry’, Freeman, 1988.
- [25] M. Billeter, W. Braun & K. Wüthrich (1982) *J. Mol. Biol.* 155, 321–346.

Chapter 6

Sequential Resonance Assignment and Analysis of Secondary Structure of Parsley Plastocyanin

Essentially complete assignment of the ^1H NMR spectrum of parsley Cu^I plastocyanin has been obtained using standard methodology [1, 2]. Through-bond scalar connectivities were identified principally from HOHAHA spectra, which demonstrate both direct and multiple relayed connectivities. DQF COSY spectra were used to distinguish direct and relayed couplings in the HOHAHA spectra. In this way, the spin systems of individual amino acid residues were identified. The position in the sequence of individual residues was subsequently determined *via* identification of strings of sequential NOEs involving backbone amide, C^α and C^β protons.

It should be noted that the two stages of spin system identification and sequential assignment were not mutually exclusive, but overlapped in the sense that sequential assignments for some spin systems could be made before other portions of the spectrum could be resolved in terms of first stage assignment. This simplifies the assignment task by elimination of cross peaks from further consideration. Chemical shifts are given in Tables 6.1, 6.2 and 6.3.

Res.	NH	C $^{\alpha}$ H	C $^{\beta}$ H	C $^{\gamma}$ H	C $^{\delta}$ H	Other
A1	–	4.07	1.33			
E2 ^a	8.63	5.04	1.85,1.93	2.10,2.10		
V3	8.80	4.27	1.38	1.07,–0.15		
K4	9.04	4.72	1.66,1.90	1.29,1.29		C $^{\epsilon}$ H 3.67,3.67
L5	8.68	4.03	0.91,2.35	1.54	0.81,0.91	
G6	8.05	4.21,5.10				
S7	8.80	4.51	3.67,4.24			
D8 ^a	8.82	4.20	2.58,2.76			
D9 ^a	7.94	4.55	2.60,2.80			
G10	8.04	3.28,4.15				
G11	7.86	3.60,3.84				
L12	8.48	4.23	0.94,1.71	1.57	0.25,0.63	
V13	7.11	4.78	2.23	0.56,0.85		
F14	8.91	4.80	2.63,3.12		7.08,7.08	C $^{\epsilon}$ H 7.22,7.22; C $^{\zeta}$ H 7.22
S15	8.82	5.04	3.34,3.80			
P16	–	5.05	2.28,2.59	1.53,1.92	3.53,3.70	
S17	8.32	4.60	4.43,4.46			
S18	7.52	5.80	3.80,3.83			
F19	7.88	5.11	3.14,3.36		6.83,6.83	C $^{\epsilon}$ H 6.93,6.93; C $^{\zeta}$ H 6.65
T20	8.50	5.50	3.99	1.13		
V21	8.89	4.71	2.17	0.71,0.88		
A22	7.92	4.78	1.39			
A23	8.52	3.93	1.27			
G24	8.66	3.17,4.25				
E25	8.17	4.22	2.10,2.25	2.21,2.35		
K26	8.32	4.21	1.55,1.58	1.09,1.16	1.48,1.48	
I27	9.13	4.28	1.77	0.62,1.85	0.79	C $^{\gamma}$ H ₃ –0.01
T28	8.31	4.65	3.88	1.12		
F29	9.59	5.26	2.82,3.31		7.18,7.18	C $^{\epsilon}$ H 6.76,6.76; C $^{\zeta}$ H 6.00
K30	9.13	4.86	1.44,1.69	1.20,1.45	1.55,1.55	C $^{\epsilon}$ H 2.78,2.78
N31	9.13	5.12	2.72,2.78		N $^{\delta}$ H 5.29,6.80	
N32	9.13	5.19	2.35,3.48		N $^{\delta}$ H 7.66,7.83	
A33	9.05	5.19	1.33			
G34 ^a	8.80	3.45,3.49				
F35	5.66	3.16	3.07,3.15		6.79,6.79	C $^{\epsilon}$ H 7.24,7.24; C $^{\zeta}$ H 7.39

Table 6.1: ^1H NMR chemical shifts of residues 1–35 of parsley Cu^I plastocyanin (pH 7.3, 35°C unless otherwise stated). ^a NH-C $^{\alpha}$ H cross peak was not observed at 35°C; NH shifts for these residues were taken from 20°C spectra.

Res.	NH	C $^{\alpha}$ H	C $^{\beta}$ H	C $^{\gamma}$ H	C $^{\delta}$ H	Other
P36	–	5.02	1.53,2.39	1.71,1.93	3.52,3.70	
H37	7.24	5.70	2.61,3.53			C $^{\delta}$ H 7.70; C $^{\epsilon}$ H 7.06; N $^{\epsilon}$ H 11.50
N38	9.98	4.41	3.70,3.73		6.84,7.25	
I39	6.69	4.25	1.25	0.13,1.08	0.43	C $^{\gamma}$ H ₃ –0.28
V40	9.02	3.96	1.09	0.58,0.68		
F41	8.54	4.59	2.72,2.80		7.23,7.23	C $^{\epsilon}$ H 7.00,7.00; C $^{\zeta}$ H 6.61
D42	8.49	4.57	2.44,2.95			
E43	8.51	3.97	2.08,2.20	2.30,2.37		
D44	8.60	4.83	2.70,2.84			
E45	8.16	4.70	1.69,2.36	1.98,2.22		
V46	6.93	4.82	1.76	0.80,1.28		
P47	–	4.24	1.43,2.33	1.22,1.35	2.10,2.50	
A48	8.15	4.12	1.36			
G49	8.52	3.61,4.21				
V50	7.66	3.85	2.11	0.78,0.83		
N51 ^a	8.55	4.95	2.70,2.97		N $^{\delta}$ H 7.20,7.72	
A52	9.12	3.80	1.43			
E53	8.55	4.40	2.78,2.82	1.76,1.80		
K54	7.48	4.28	1.90,2.02	1.47,1.58	1.72,1.72	C $^{\epsilon}$ H 2.93,2.93
I55	6.95	4.47	1.70	0.99,1.07	–0.12	C $^{\gamma}$ H ₃ 0.73
S56	7.33	4.68	4.10,4.65	O $^{\gamma}$ H 5.84		

Q59	9.23	4.75	1.06,1.15	–0.03,2.50		
P60	–	4.20	2.18,2.46	1.90,1.98	3.20,3.78	
E61	7.74	4.00	1.77,1.87	2.18,2.25		
Y62	7.81	4.36	2.72,2.91		7.00,7.00	C $^{\epsilon}$ H 6.75,6.75
L63	9.20	4.46	1.30,1.68	1.68	0.66,0.74	
N64	8.37	4.61	2.65,2.73		N $^{\delta}$ H 6.69,7.46	
G65	8.08	3.49,3.97				
A66	8.42	3.44	1.42			
G67	8.38	3.64,4.17				
E68	7.10	4.30	2.03,2.11	1.91,2.28		
T69	8.28	5.74	4.07	1.14		
Y70	8.88	5.04	2.59,3.28		6.74,6.74	C $^{\epsilon}$ H 7.12,7.12
E71	7.80	5.50	1.63,1.77	2.08,2.13		
V72	8.75	4.37	1.90	0.93,1.03		

Table 6.2: ¹H NMR chemical shifts of residues 36–72(70) of parsley Cu^I plastocyanin (pH 7.3, 35°C unless otherwise stated). ^a NH-C $^{\alpha}$ H cross peak was not observed at 35°C; NH shifts for these residues were taken from 20°C spectra.

6.1 Residue Numbering

Parsley plastocyanin residues will generally be referred to by two numbers: the first corresponds to the poplar plastocyanin sequence, such that the deletions at positions 57 and 58 in the parsley plastocyanin sequence are counted as residues. All spectra are labelled according to this scheme. The second number in parentheses is obtained by omitting the deletions from the count, so that the total number of positions in the sequence is reduced by two.

6.2 Spin System Identification

All spectra presented in this section, and subsequent sections of this chapter, are of parsley Cu^I plastocyanin.

6.2.1 Methyl Group Region: 1. Alanines and Threonines

Parsley plastocyanin contains nine Ala and seven Thr residues. In the D₂O HO-HAHA experiment, all seven Thr residues could be distinguished from Ala residues by the presence of a relayed C^αH-C^γH₃ connectivity in addition to a direct C^αH-C^βH connectivity. Each of the nine Ala residues gave a characteristically intense direct C^αH-C^βH₃ connectivity in the D₂O spectra.

6.2.2 Methyl Group Region: 2. Isoleucines, Leucines and Valines

Parsley plastocyanin contains three Ile, nine Val and four Leu residues giving rise to intense methyl group cross peaks as shown in Figure 6.1. Pairs of C^βH-C^γH₃ and C^γH-C^δH₃ methyl group cross peaks corresponding to Val and Leu residues, and single C^βH-C^γH₃ cross peaks of Ile, were readily identified. Val and Leu residues could be distinguished on the basis of their unique spin systems by analysis of, for example, a HOHAHA spectrum recorded in D₂O.

Res.	NH	C ^α H	C ^β H	C ^γ H	C ^δ H	Other
T73	8.02	4.48	3.75	0.81		
L74	8.93	4.57	0.92,1.56	1.30	−0.28,0.31	
T75	8.58	4.21	4.12	1.00		
E76	7.78	4.39	1.92,2.24	1.96,2.06		
K77	8.59	4.01	1.72,1.90	1.30,1.60	1.70,1.70	C ^ε H 3.01,3.01
G78	8.59	3.86,4.63				
T79	8.38	5.17	3.85	1.12		
Y80	9.60	5.59	2.97,3.62		7.12,7.12	C ^ε H 6.57,6.57; O ^η H 10.32
K81	8.26	5.11	1.90,2.12	1.38,1.61	1.70,1.70	C ^ε H 2.94,2.94
F82	7.83	5.38	1.57,1.76		6.18,6.18	C ^ε H 5.87,5.87; C ^ζ H 6.89
Y83	9.24	5.21	3.12,3.20		6.77,6.77	C ^ε H 6.63,6.63
C84	7.75	5.21	2.93,3.24			
E85	9.84	4.39	2.24,2.43	2.34,2.59		
P86	–	4.17	0.99,1.74	1.09,1.37	3.29,5.16	
H87	8.35	5.18	3.40,3.79			C ^δ H 7.00; C ^ε H 7.59
A88 ^a	8.55	4.18	1.59			
G89 ^a	9.14	3.89,3.93				
A90	7.58	4.61	1.62			
G91	8.01	3.79,4.46				
M92	7.68	4.65	1.81,2.28	1.65,2.07		C ^ε H ₃ 0.51
K93	8.06	5.16	1.87,2.14	1.51,1.84	1.66,1.66	
G94	8.42	2.90,4.58				
E95	8.29	5.09	1.78,1.90	1.75,2.10		
V96	9.08	4.95	1.22	0.30,0.82		
T97	8.22	5.00	3.94	1.05		
V98	9.25	4.66	2.43	0.62,0.66		
N99	8.92	4.70	2.63,2.70		N ^δ H 6.47,7.23	

Table 6.3: ¹H NMR chemical shifts of residues 73(71)–99(97) of parsley Cu^I plastocyanin (pH 7.3, 35°C unless otherwise stated). ^a NH-C^αH cross peak was not observed at 35°C; NH shifts for these residues were taken from 20°C spectra.

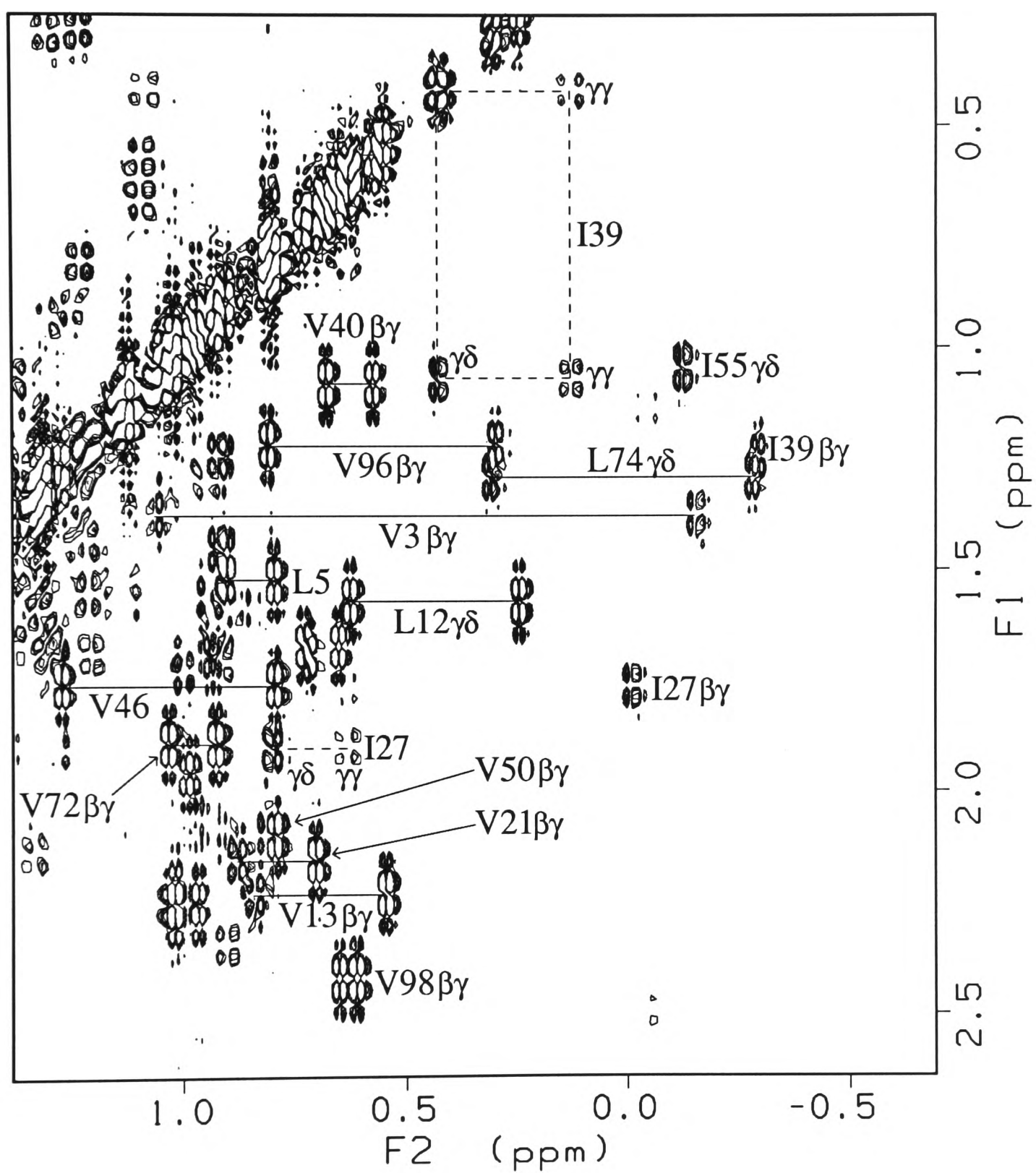


Figure 6.1: Ile, Val and Leu methyl group region of the DQF COSY spectrum of parsley Cu^I plastocyanin recorded in D_2O at 35°C , $\text{pH}^* 7.3$.

6.2.3 Identification of Phe, Tyr and His Spin Systems

The protons of an aromatic ring form an essentially isolated spin system, with no resolvable coupling to the backbone $\text{NH-C}^\alpha\text{H-C}^\beta\text{H}_2$ group. Since aromatic ring protons are intrinsically associated with ring currents, their resonances occur in a downfield chemical shift range such that the aromatic region of the spectrum can be viewed as a 'self-contained' sub-spectrum.

Figure 6.2 shows the aromatic region of a HOHAHA spectrum recorded in H_2O . The spin systems of the two His residues were immediately identifiable by their weak $\text{C}^\delta\text{H-C}^\epsilon\text{H}$ cross peaks in both DQF COSY and HOHAHA spectra. The characteristic coupling patterns of five of the six Phe spin systems could be discerned, complete identification of the sixth being prevented by the degeneracy of the $\text{C}^\epsilon\text{H}$ and C^ζH resonances. The four remaining cross peaks are due to Tyr ring protons. The distinction of the C^δH and $\text{C}^\epsilon\text{H}$ resonances of Tyr 62(60) and Tyr 70(68) is less reliable than in other aromatic spin systems due to the lack of Ar type NOEs for these residues.

6.2.4 Identification of Proline Spin Systems

The lack of a fingerprint $\text{NH-C}^\alpha\text{H}$ cross peak makes identification of proline spin systems particularly difficult. In addition the signals associated with these residues are often weak due to the torsion angles within the ring resulting in low coupling constants. Two pieces of evidence can often be used in the elucidation of proline spin systems. The first is the appearance of an NOE from the proline $\text{C}^\delta\text{H}_2$ group to the C^αH of the preceding residue. Examples are shown in Figure 6.3, where the C^δH protons of Pro 16 and Pro 60(58) are seen to give rise to NOEs to the C^αH of Ser 15 and Gln 59(57) respectively. This provided a marker for assignment of two of the five prolines present in parsley plastocyanin. The second characteristic is the appearance of a strong intraresidue $\text{C}^\delta\text{H-C}^\delta\text{H}$ cross peak, often near the diagonal, in COSY, HOHAHA and NOESY spectra. This feature was most difficult to identify in the case of Pro 86(84), due to the unusual downfield chemical shift of one of its δ protons. Examples of proline spin systems in the D_2O HOHAHA spectrum are shown in Figure 6.4.

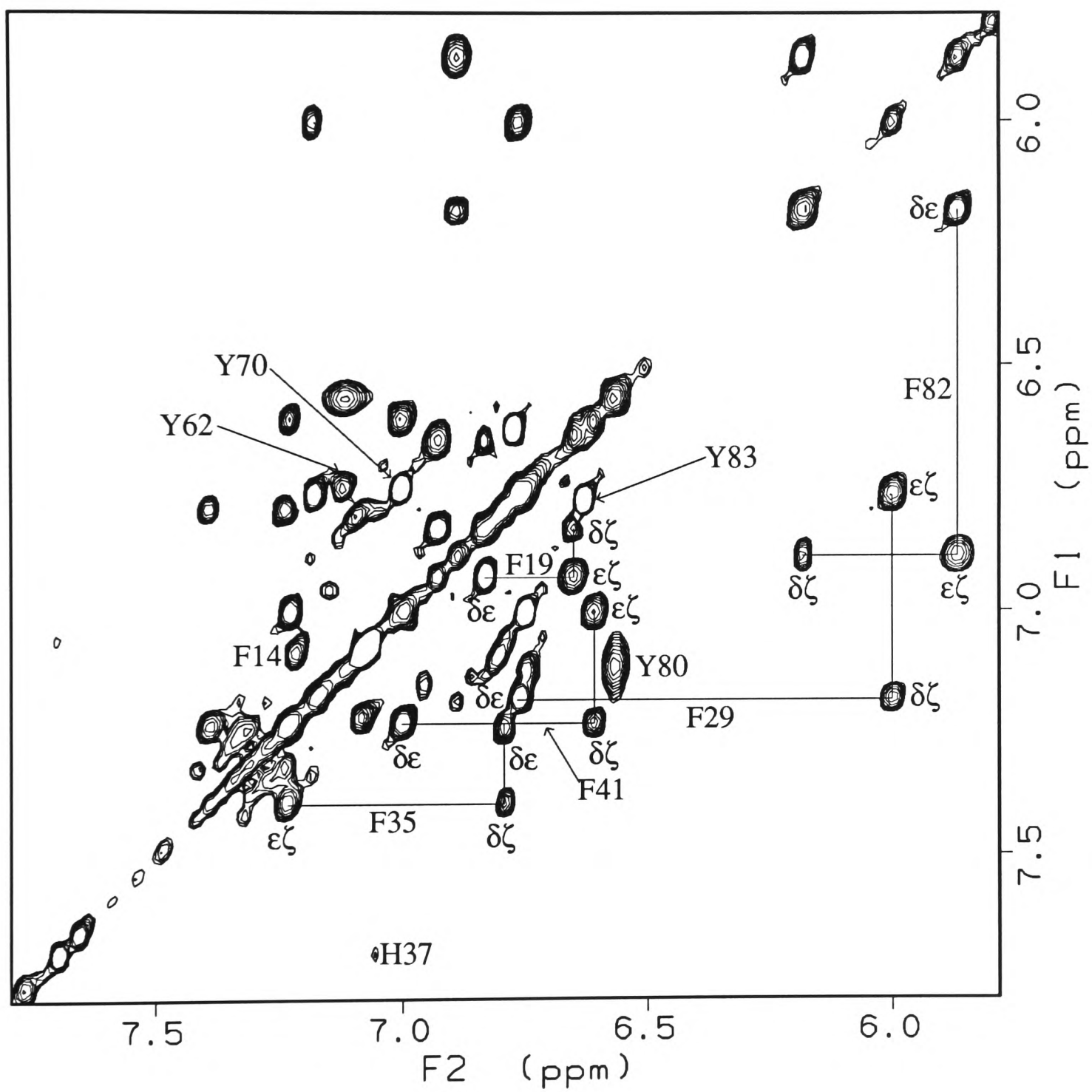


Figure 6.2: Aromatic region of the HOHAHA spectrum of parsley Cu^I plastocyanin recorded in D_2O at 35°C , $\text{pH}^* 7.3$ with a 34ms MLEV17 mixing sequence.

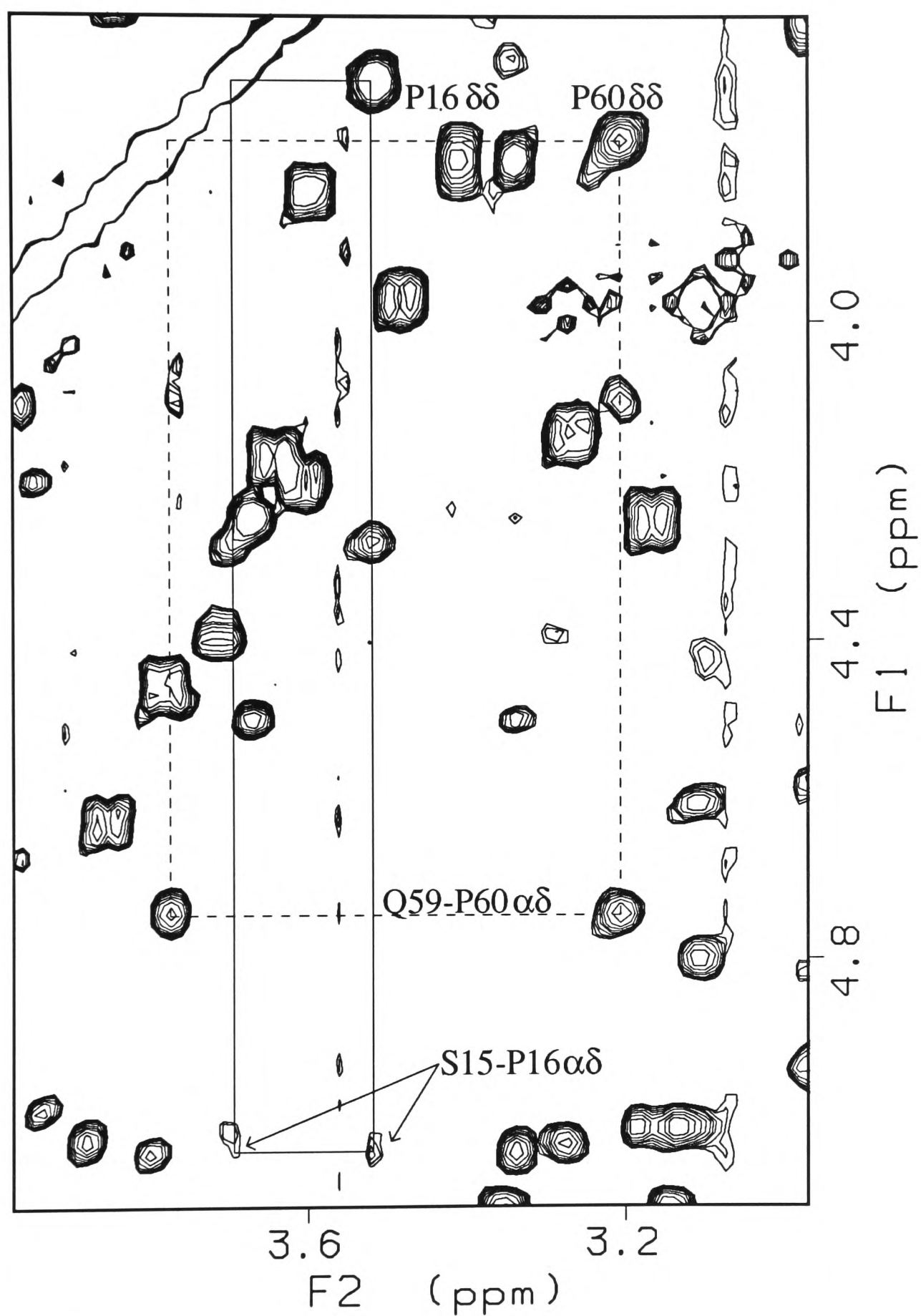


Figure 6.3: Part of the NOESY spectrum of parsley Cu^I plastocyanin recorded in D_2O at 35°C , $\text{pH}^* 7.3$ with a mixing time of 200ms. The intrasidue $\text{C}^\delta\text{H}-\text{C}^\delta\text{H}$ cross peaks of Pro 16 and Pro 60(58), and corresponding sequential $\text{C}^\alpha\text{H}-\text{C}^\delta\text{H}_2$ cross peaks, are indicated.

The $C^\alpha H$ chemical shifts were obtained by looking for $C^\alpha H-C^\beta H$ and $C^\alpha H-C^\gamma H$ cross peaks which appear with low intensity in the HOHAHA spectrum. The $C^\alpha H$ chemical shifts of Pro 47 and Pro 60 were determined during the second stage of the assignment process by the identification of sequential $d_{\alpha N}$ NOEs involving Ala 48 and Glu 61(59).

6.2.5 Fingerprint Region

Figure 6.5 shows the fingerprint region of the HOHAHA spectrum recorded in H_2O at $20^\circ C$. This spectrum contains all the fingerprint $NH-C^\alpha H$ cross peaks that are expected for parsley plastocyanin. Those due to Glu 2, Asp 8, Asp 9, Gly 34, Asn 51, Ala 88(86) and Gly 89(87) did not appear in the spectrum recorded at $35^\circ C$. Not shown in the figure are the cross peaks for which the chemical shift of the exchangeable proton is relatively high field, i.e. the $O^\gamma H-C^\beta H$ cross peak of Ser 56 and the $NH-C^\alpha H$ cross peak of Phe 35. The assignment of the fingerprint region cross peaks is now described.

6.3 Further Identification of Spin Systems and Attachment of Backbone Amide Protons

6.3.1 Glycines

Ten of the twelve Gly spin systems of parsley plastocyanin were identified by their square $C^\alpha H-C^\alpha H$ cross peaks in the D_2O DQF COSY spectrum. Identifying characteristics are the lack of passive couplings and the large geminal coupling constants ($^2J_{\alpha\alpha} \approx 15 Hz$). For these ten spin systems, pairs of $NH-C^\alpha H$ cross peaks were easily identified in the H_2O HOHAHA spectrum recorded at $35^\circ C$, whilst pairs of weak $NH-C^\alpha H$ cross peaks for the remaining two Gly spin systems appeared in the H_2O HOHAHA spectrum recorded at $20^\circ C$ (Figure 6.5). The $C^\alpha H-C^\alpha H$ cross peaks for these two residues could subsequently be identified as lying close to the diagonal in the D_2O DQF COSY spectrum.

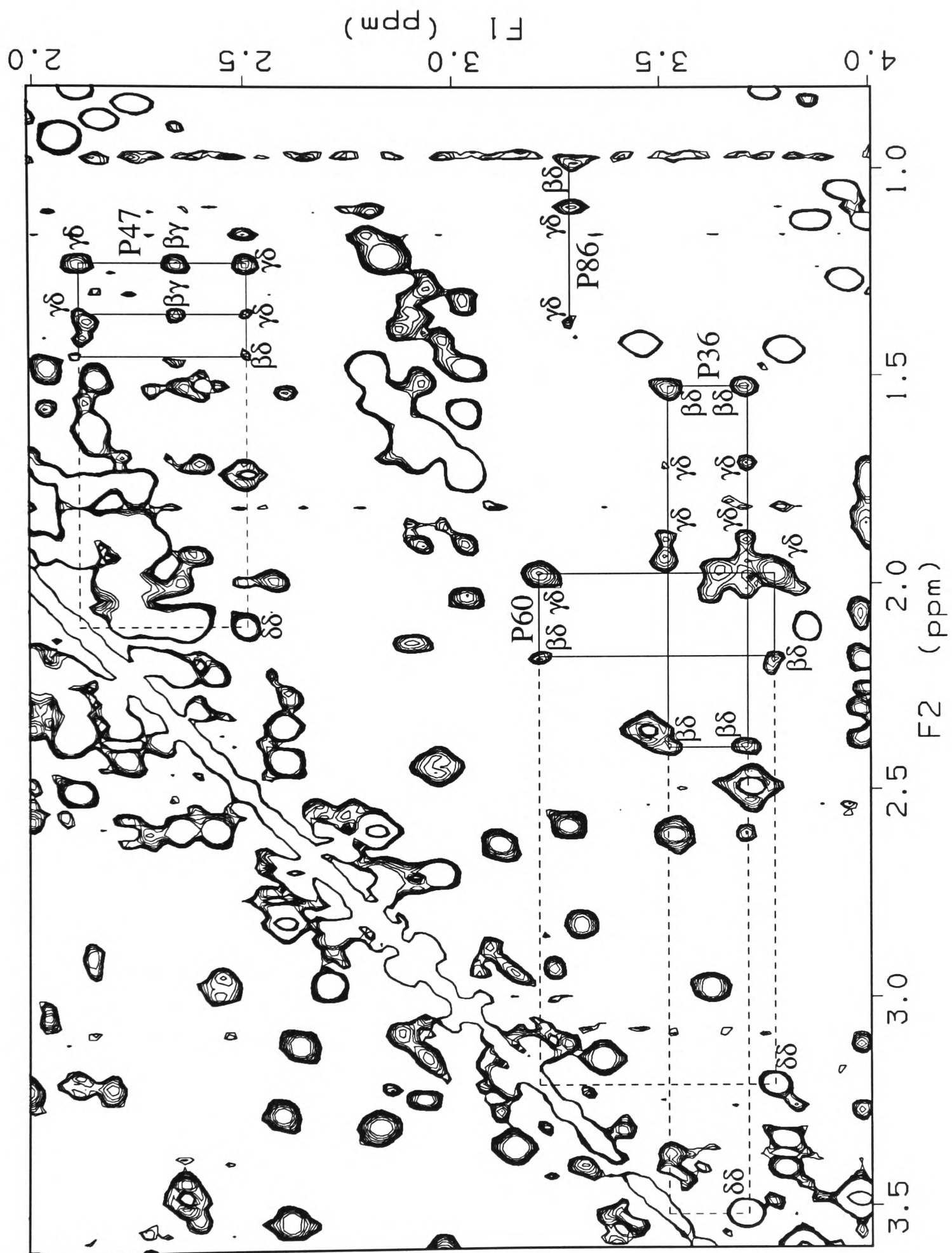


Figure 6.4: Region of the HOHAHA spectrum of parsley Cu^I plastocyanin recorded in D_2O at 35°C , pH^* 7.3 with a 34ms MLEV17 mixing sequence bracketed by 2.5ms trim pulses. Proline spin systems are indicated.

6.3.2 Methyl Group-Containing Residues

The identification of the major non-labile parts of these spin systems has been described above. Peaks delineating the complete spin systems from the amide proton for all Ala and Thr residues could be identified in the H₂O HOHAHA spectrum recorded at 35°C (Figures 6.6 and 6.7). Attachment of amide NH protons to the four Leu spin systems was made using C^αH chemical shifts previously obtained by location of C^αH-C^δH₃ NOEs. The NH proton was typically correlated with the C^αH and C^βH protons in HOHAHA spectra, the NH-C^βH cross peaks all having type U character (i. e. C^βH chemical shift upfield of 2.5ppm) [2].

Five complete Val spin systems were discerned in the H₂O HOHAHA spectrum recorded at 35°C (Figures 6.6 and 6.7), while only a NH-C^γH₃ cross peak was missing for two of the remaining four Val spin systems. The final two Val residues gave rise to just a direct NH-C^αH cross peak, and their identification awaited second stage assignment of neighbouring residues.

6.3.3 Non-unique Spin Systems

The remaining fingerprint region cross peaks were assigned to one of type J, U or X on the basis of the chemical shift of the corresponding β protons. Type J spin systems are those with C^βH chemical shifts of greater than 2.5ppm, which are likely to be any of Asn, Asp, Cys, His, Phe, Ser and Tyr. Those spin systems with C^βH chemical shifts less than 2.5ppm are designated type U, and are usually any of Arg, Glu, Gln, Leu, Lys, Met or Pro. Overall, twenty-six fingerprint region cross peaks were observed from type J spin systems (this total includes the fingerprint cross peak of Glu 53, which belongs to the group of residues expected to have type U character, but for which the C^βH chemical shifts are shifted downfield), twenty-three from type U spin systems (this total includes the fingerprint cross peak of Phe 82(80), which belongs to the group of residues expected to have type J character, but for which the C^βH chemical shifts are shifted upfield). In addition to the latter total are the five Pro spin systems which do not, of course, involve fingerprint cross peaks. The remaining spin systems are designated as type X.

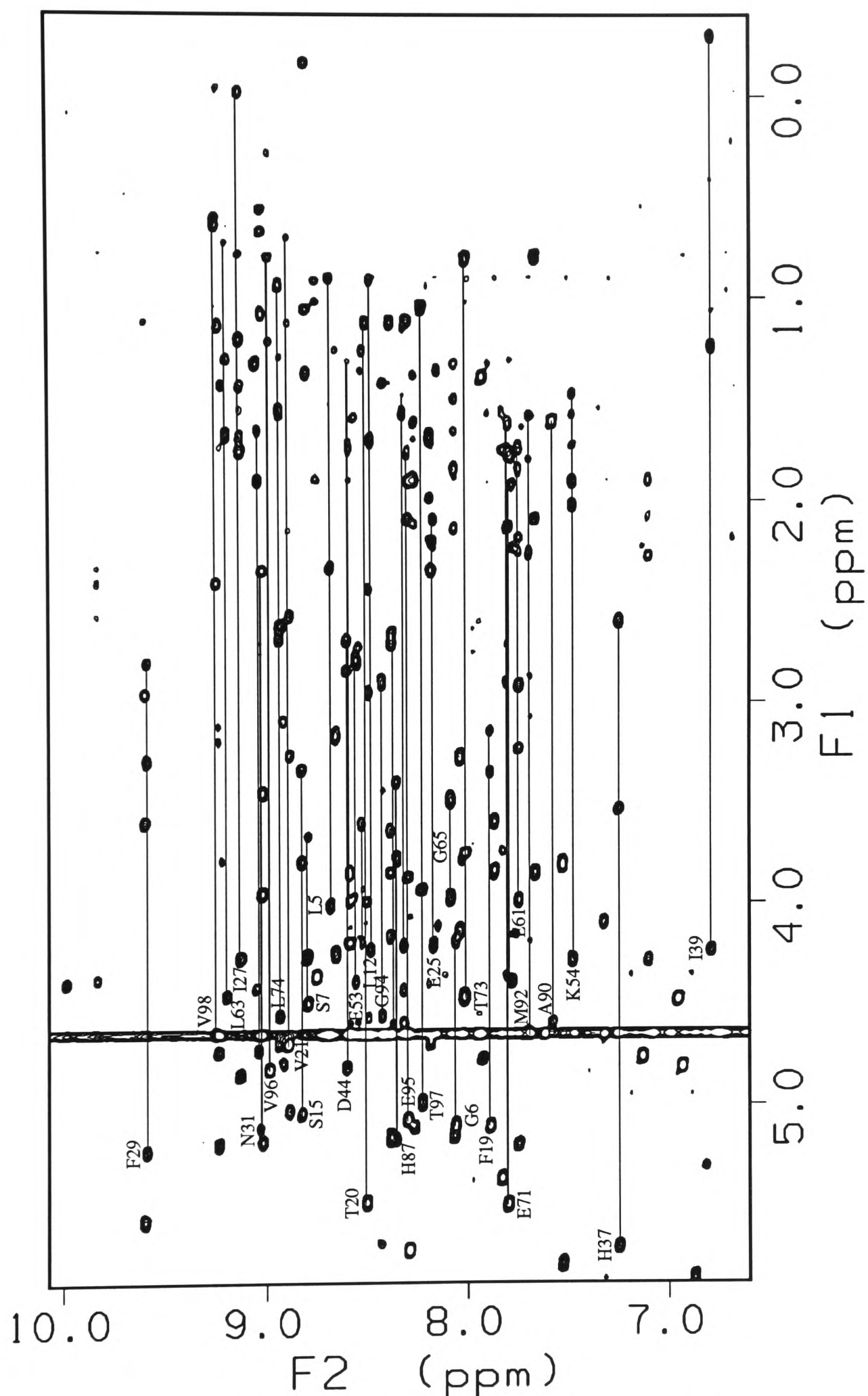


Figure 6.6: Part of the HOHAHA spectrum of parsley Cu^I plastocyanin recorded in H_2O (35°C , pH 7.3; 55ms WALTZ17 mixing scheme bracketed by 1.5ms and 3ms trim pulses). Approximately half of the plastocyanin spin systems are labelled at the positions of the direct $\text{NH}-\text{C}^\alpha\text{H}$ connectivities, with NH -side chain cross peaks lying along the horizontal lines.

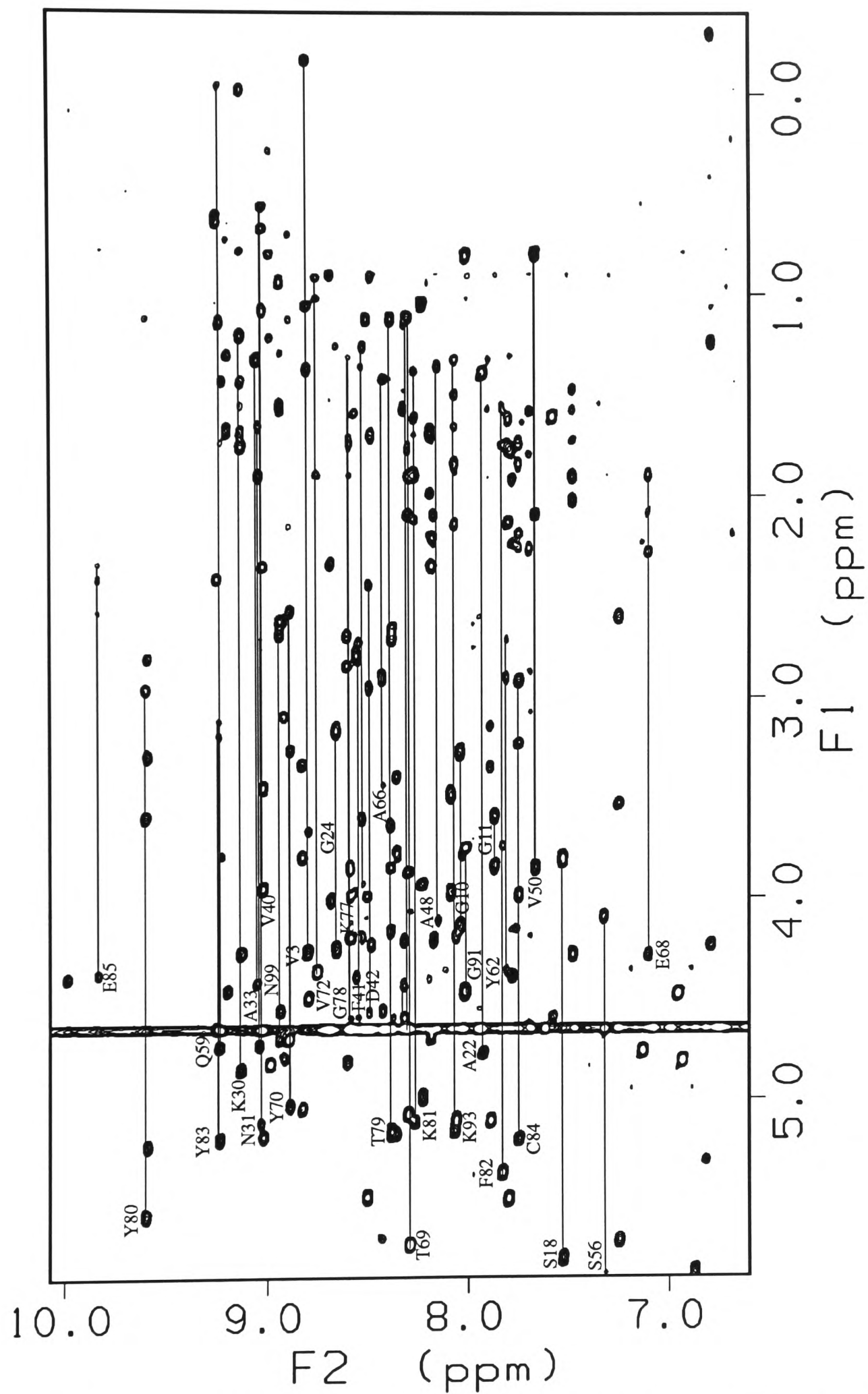


Figure 6.7: Same region of the same HOHAHA spectrum as shown in the previous figure indicating the positions of the remaining plastocyanin spin systems.

The type J spin systems have an Ar (indicating aromatic character) subclass, to which spin systems are assigned on the basis of the observation of one or more of $C^\delta H-C^\beta H$, $C^\delta H-C^\alpha H$ and $C^\delta H-NH$ intraresidue NOEs. Using this criterion, two His, five Phe and two Tyr spin systems could be identified. In addition, a type U spin system could be seen to have aromatic character (see previous paragraph), and this was assigned as Phe 82(80) in the second stage analysis.

Two aromatic spin systems remained to be assigned at this stage. Later analysis would show that a type J spin system which gave rise to none of the NOEs used to determine aromatic character is Tyr 70(68). The final aromatic residue, for which no $NH-C^\beta H$ scalar couplings and no characteristic NOEs were observed, was shown to be Tyr 62(60) during the sequential assignment process.

6.4 Second Stage Assignment

Sequence specific resonance assignments of parsley Cu^I plastocyanin have been made by identification of short $^1H-^1H$ distances between sequentially neighbouring residues. Whether the observed connectivities are of the $d_{\alpha N}$ or d_{NN} type is dependent on the nature of the secondary structure (see Chapter 4). Spin systems are allocated to particular positions in the sequence by identifying groups of NMR signals corresponding to peptide segments which are unique in the primary structure of the protein. Such peptide segments constitute anchoring points from either end of which further assignments can be made until the various segments are linked.

A cursory inspection of the sequential NOEs observed for parsley Cu^I plastocyanin (Figures 6.8 and 6.9) reveals the predominance of the $d_{\alpha N}$ over d_{NN} variety, indicating that much of the protein is in extended conformation. Additional evidence from large $^3J_{HN\alpha}$ coupling constants and slow backbone NH exchange indicate that the protein consists mainly of β -secondary structure.

In the remainder of this chapter, the sequential NOE data are described, organised according to secondary structure type. The packing of the β -strands is then elucidated by analysis of interstrand NOEs. Finally, evidence is sought for the location and type of turns and loops from NOE and coupling constant data.

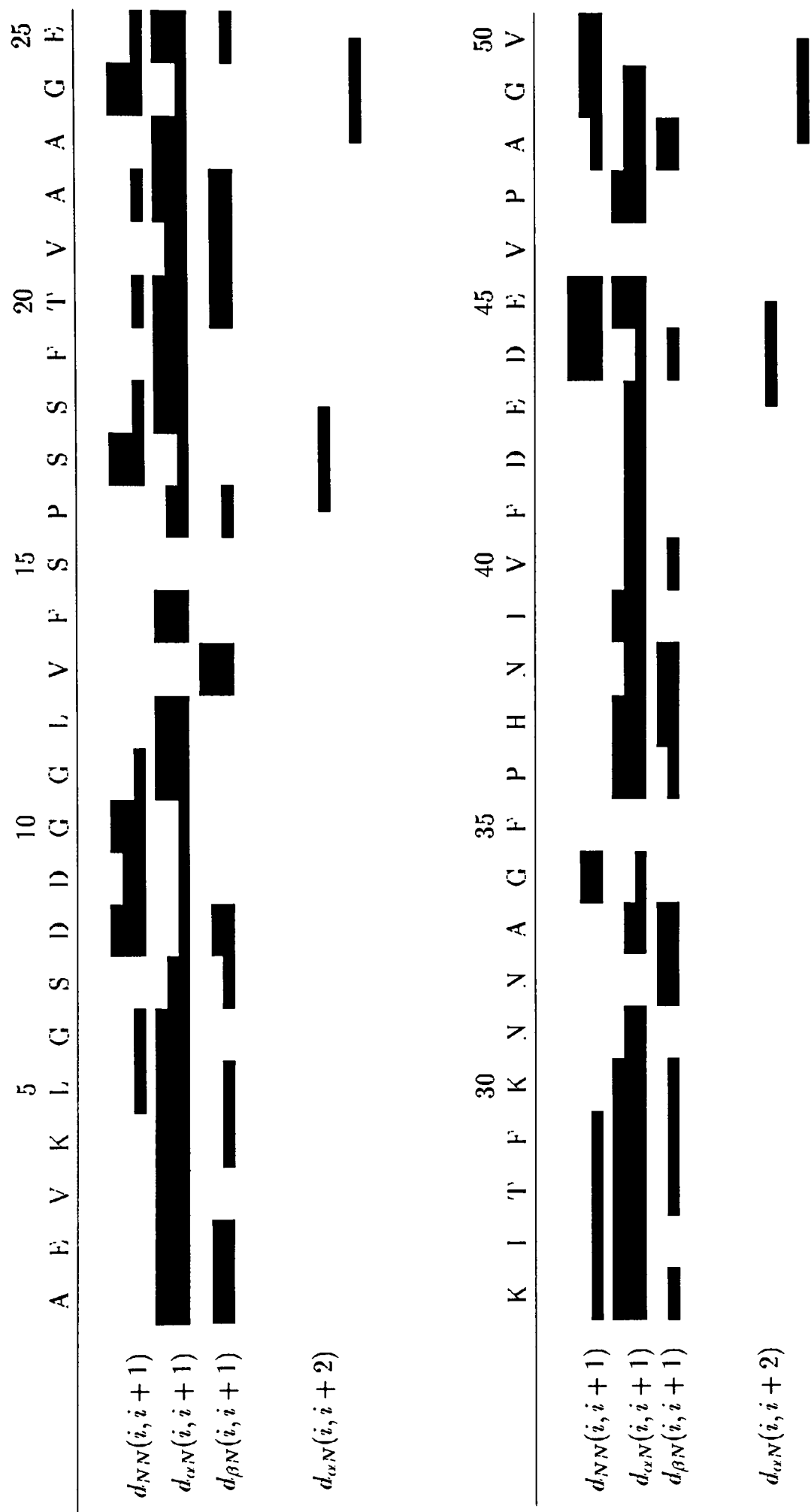


Figure 6.8: The first half of the primary structure of parsley plastocyanin with a survey of the sequential NOEs used in determining sequence specific assignments.

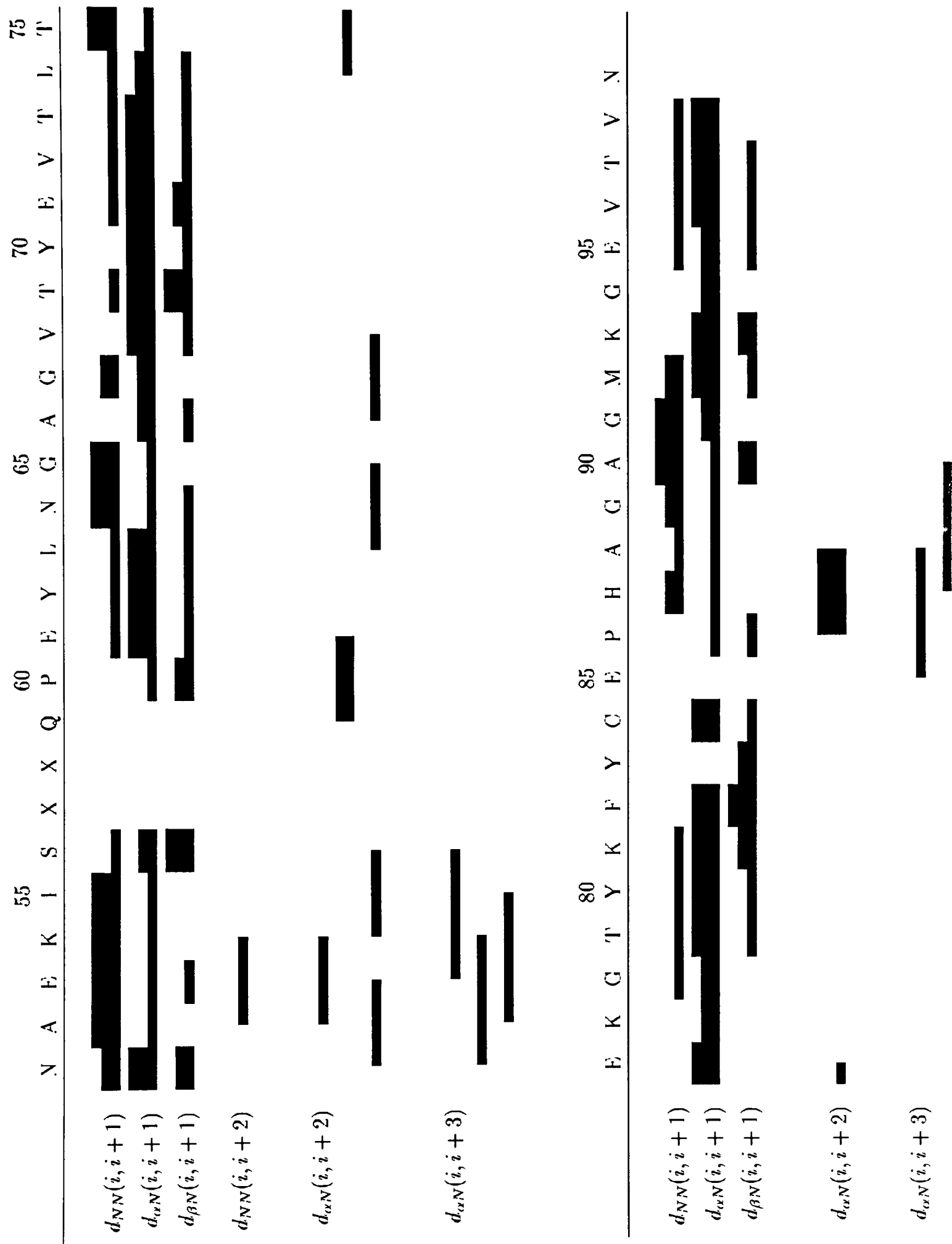


Figure 6.9: The second half of the primary structure of parsley plastocyanin with a survey of the sequential NOEs used in determining sequence specific assignments. The 'X' at positions 57 and 58 denotes deletions relative to the poplar plastocyanin sequence.

6.4.1 Fragments with Extended Conformation

Sequential assignments in these fragments were made largely on the basis of intense $d_{\alpha N}$ NOEs between neighbouring residues.

The sequence Val-Lys-Leu-Gly at positions 3–6 was recognised as a $d_{\alpha N}$ -connected Val-Type U-Leu-Gly group of spin systems (Figure 6.10), in which the methyl group resonances of Val 3 have characteristic upfield and downfield ring current shifts, as in the spectra of other plastocyanins, for example, those from spinach [3] and French bean [4, 5]. Extension of this sequence to Ala 1 in one direction, and to Asp 8 in the other, is made through $d_{\alpha N}$ NOEs. Here it should be noted that the fingerprint cross peaks of Glu 2 and Asp 8 were not observed in spectra recorded at 35°C, but appeared in spectra recorded at 20°C. Moreover, the $d_{\alpha N}$ (Ser 7, Asp 8) NOE was less intense than others in this fragment.

Intense $d_{\alpha N}$ NOEs were found to connect residues 11–13 and 14–15 (Figure 6.11). Any $d_{\alpha N}$ connectivity between residues 13 and 14 was obscured due to the degeneracy of the $C^\alpha H$ chemical shifts for these two residues at 35°C, and due to overlap with the fingerprint cross peak of Val 96(94) at 20°C and 15°C.

The next group of residues containing uniquely identifiable spin systems is Phe-Thr-Val at positions 19–21 ($d_{\alpha N}$ -connected Type J(Ar=Phe)-Thr-Val) (Figure 6.12). A sequence of strong $d_{\alpha N}$ NOEs from Val 21 amide NH allows the sequential assignment of Thr 20, Phe 19 and Ser 18. In the other direction, a string of mainly intense $d_{\alpha N}$ NOEs extends from Val 21 to Gly 24.

The sequence Ile-Thr-Phe at positions 27–29 was recognised as a $d_{\alpha N}$ -connected Ile-Thr-Type J(Ar=Phe) group of spin systems (Figure 6.13). Ile 27 is connected to 26 and Glu 25 *via* a strong and then a less intense $d_{\alpha N}$ NOE. In the opposite direction, a string of intense $d_{\alpha N}$ NOEs extends from Ile 27 to Asn 31. Medium strength $d_{\alpha N}$ NOEs connect the pairs (31,32) and (33,34), but the expected sequential connectivity between Asn 32 and Ala 33 is not observed (their NH chemical shifts are the same at 35°C, 20°C and 15°C).

Residues 36–44 are readily connected through $d_{\alpha N}$ NOEs (Figure 6.14). The $d_{\alpha N}$ -connected pair Ile 39-Val 40 provides the starting point for sequential assignment of this fragment. Extension to residue 46 can be achieved by recognising the threesome

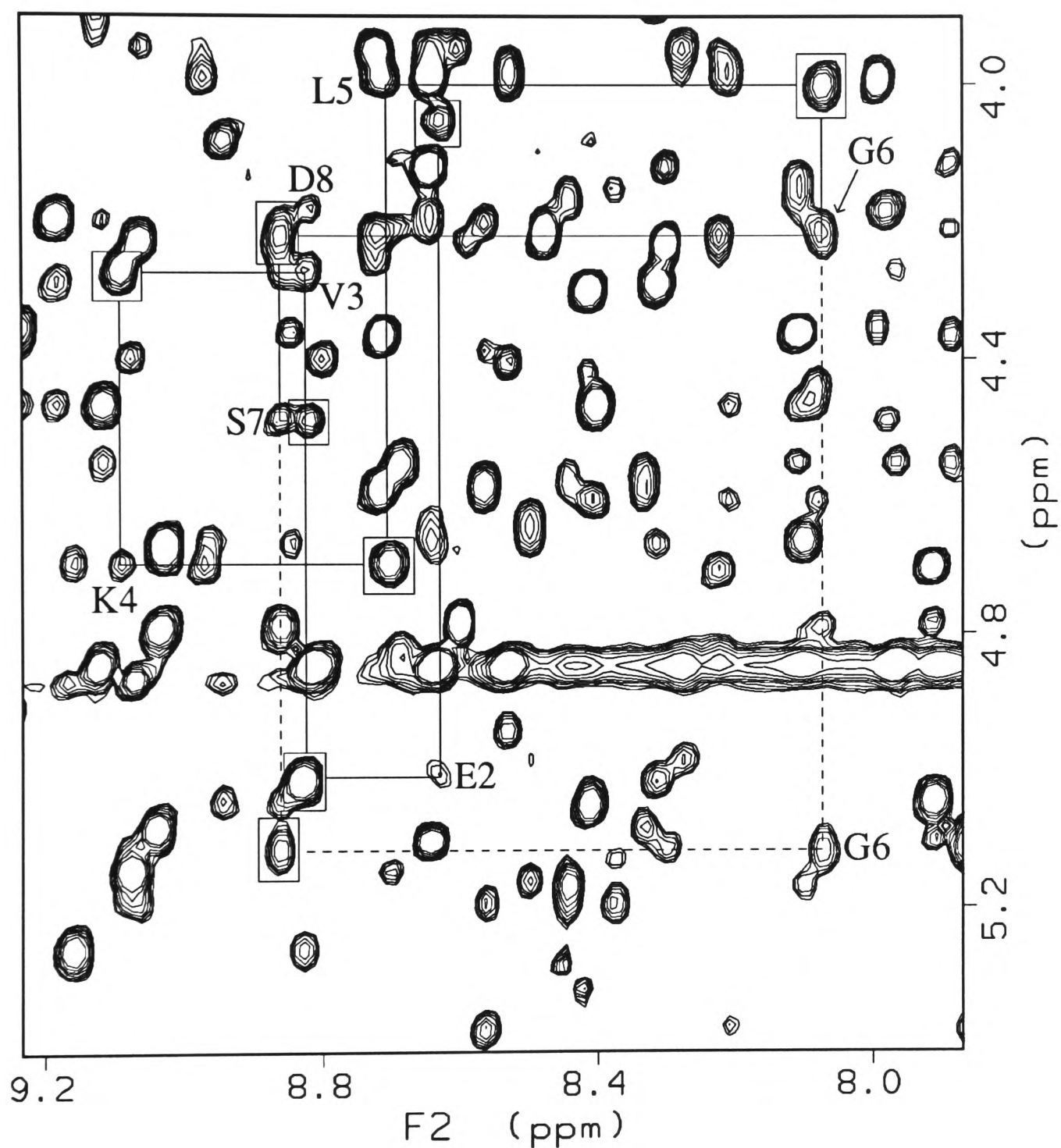


Figure 6.10: Sequential resonance assignments for polypeptide segments in extended conformation I. Part of the NOESY spectrum of parsley Cu^I plastocyanin recorded in H_2O at 35°C , pH 7.3 with a mixing time of 150ms. Sequential $d_{\alpha N}$ connectivities for residues 1–7 are shown.

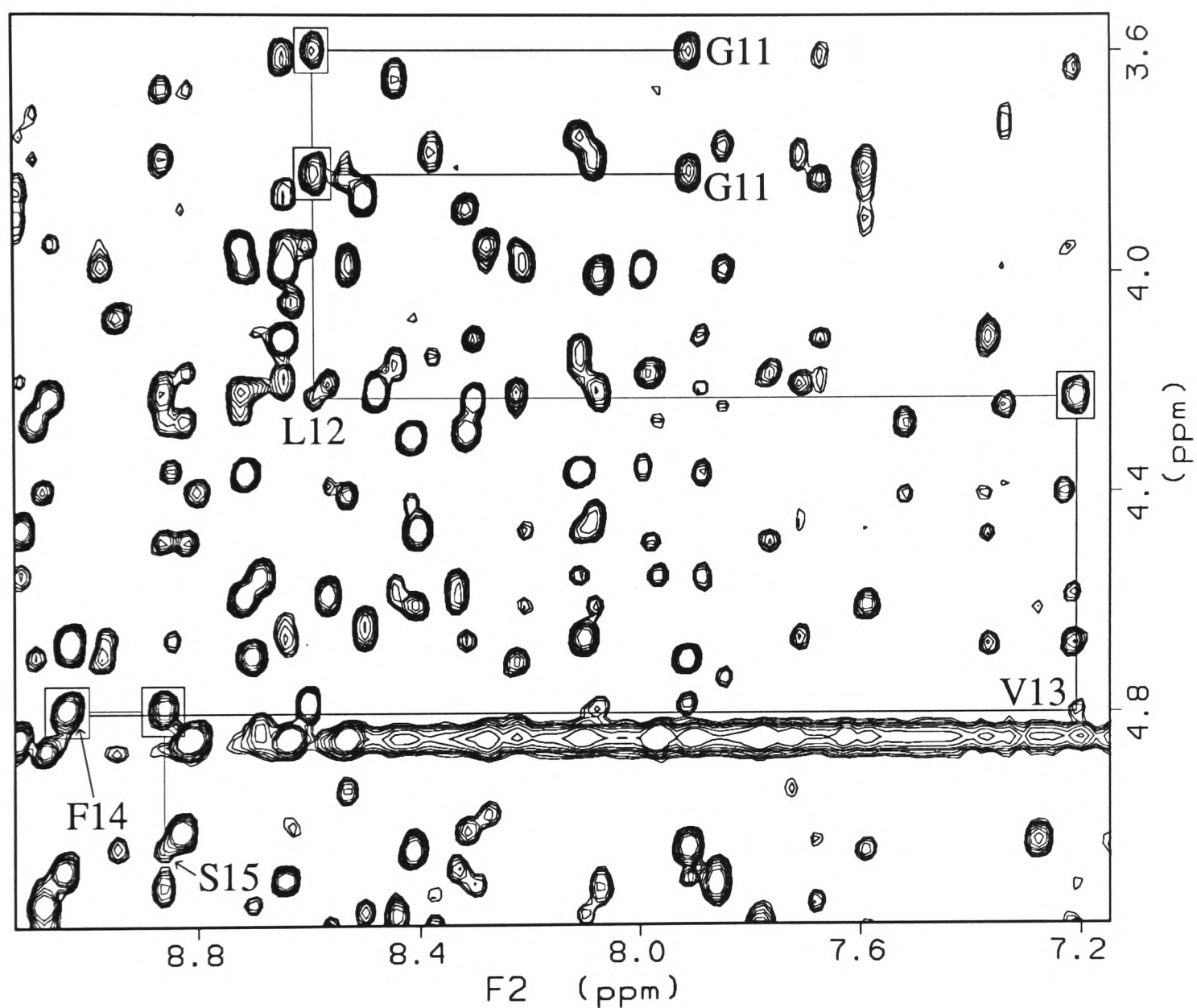


Figure 6.11: Sequential resonance assignments for polypeptide segments in extended conformation II. Part of the NOESY spectrum of parsley Cu^I plastocyanin recorded in H_2O at 35°C , pH 7.3 with a mixing time of 150ms. Sequential $d_{\alpha N}$ connectivities for residues 11–15 are shown.

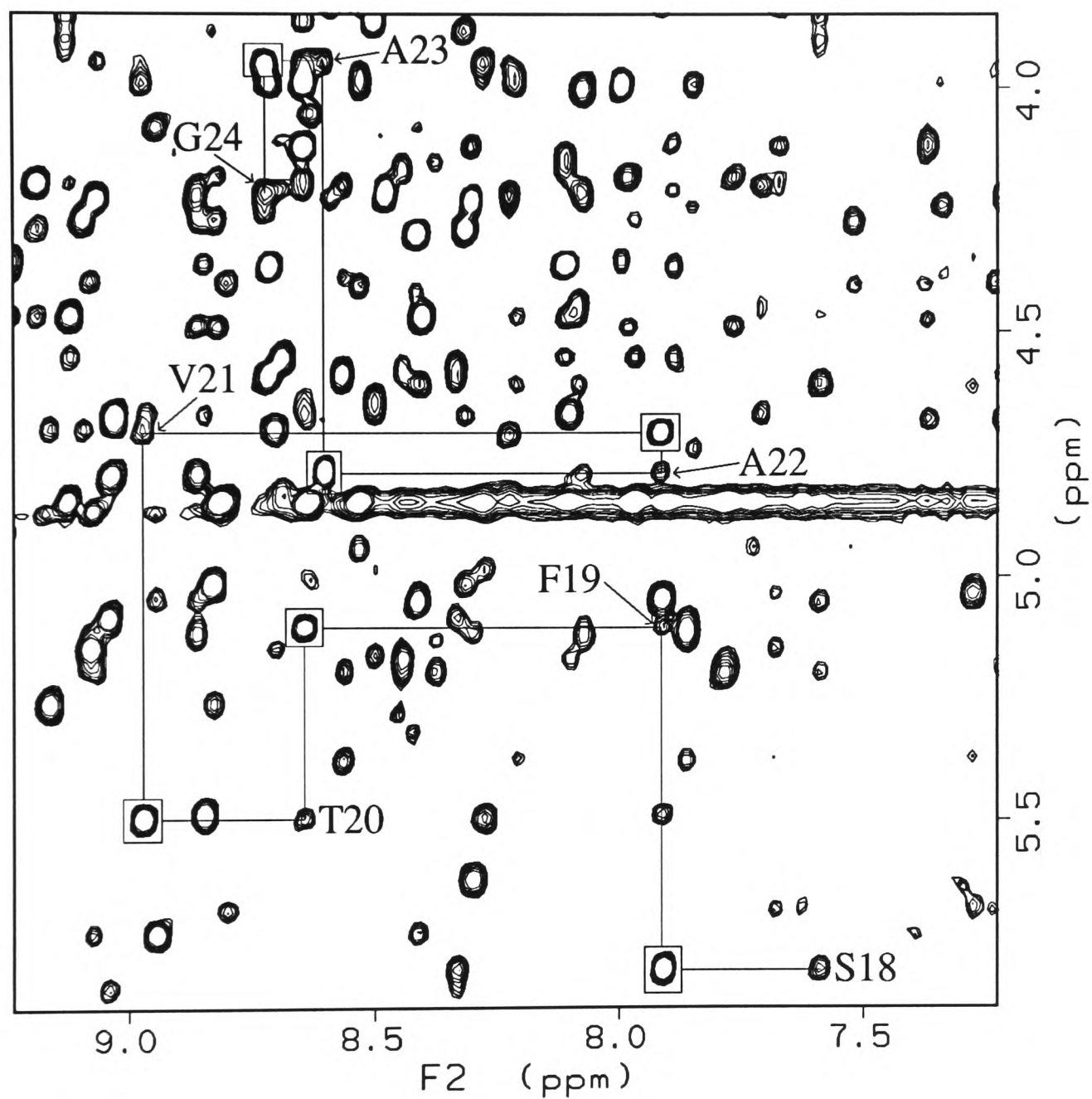


Figure 6.12: Sequential resonance assignments for polypeptide segments in extended conformation III. Part of the NOESY spectrum of parsley Cu^I plastocyanin recorded in H_2O at 35°C , pH 7.3 with a mixing time of 150ms. Sequential $d_{\alpha N}$ connectivities for residues 18–24 are shown.

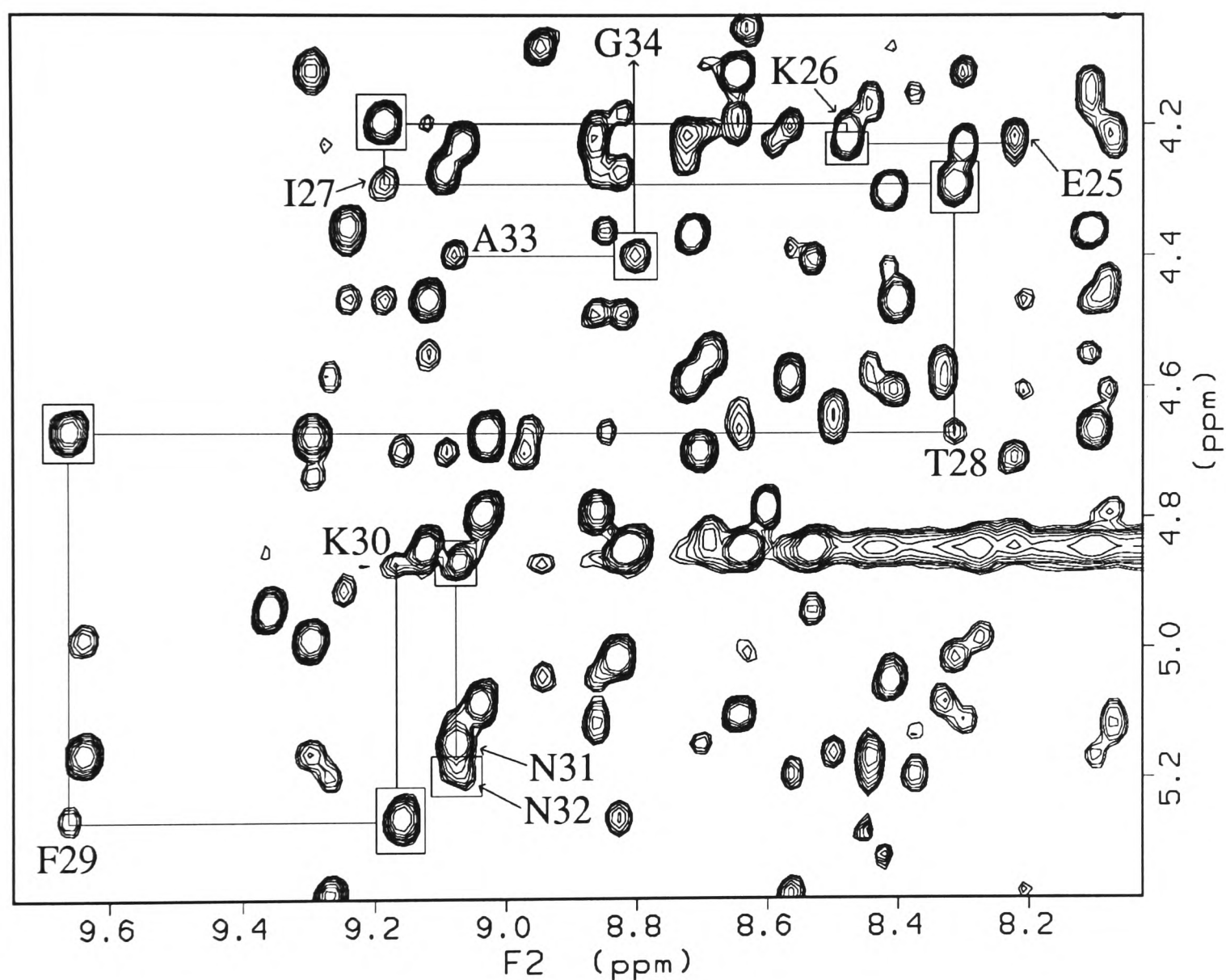


Figure 6.13: Sequential resonance assignments for polypeptide segments in extended conformation IV. Part of the NOESY spectrum of parsley Cu^I plastocyanin recorded in H_2O at 35°C , pH 7.3 with a mixing time of 150ms. Sequential $d_{\alpha N}$ connectivities for residues 25–32 and 33–34 are shown.

Asp 44-Glu 45-Val 46 as a d_{NN} -connected Type J-Type U-Val group of spin systems (noting that these three residues are not in an extended conformation). Pro 47 is linked *via* an intense $d_{\alpha N}$ NOE with Ala 48, and with Val 46 *via* strong $d_{\alpha P\delta}$ NOEs, which are characteristic of Pro residues.

β -type secondary structure in the short sequence of residues 61(59)–64(62) is indicated by the presence of intense $d_{\alpha N}$ connectivities. The next fragment with extended conformation contains another of the groups of residues that constitute anchoring points in the sequential assignment process: Val-Thr-Leu-Thr at positions 72(70)–75(73) was picked out as a sequence of unique spin systems. From Val 72(70), a series of intense $d_{\alpha N}$ NOEs extends back to Glu 68(66) (Figure 6.15). In the forward direction from Thr 75(73), medium and strong $d_{\alpha N}$ NOEs occur through to Glu 85(83) with two breaks: Thr 75(73) and Glu 76(74) are strongly d_{NN} -connected, and the sequential NOE between Tyr 83(81) and Cys 84(82) is obscured due to degeneracy of $C^\alpha H$ chemical shifts (Figure 6.16).

The C-terminal strand of parsley plastocyanin contains the $d_{\alpha N}$ -connected sequence Val-Thr-Val at positions 96(94)–98(96). This threesome was recognised as a group of residues with unique spin systems in the sequential assignment process. Val 98(96) is linked to the C-terminal residue, Asn 99(97), by a strong $d_{\alpha N}$ NOE, and a series of medium and strong $d_{\alpha N}$ NOEs extends back from Val 96(94) to the Cu ligand Met 92(90) (Figure 6.17).

6.4.2 Helical Segment

Intense sequential d_{NN} and intraresidue $d_{\alpha N}$ connectivities indicate that residues 52–56 adopt a helical conformation (Figure 6.18). This is further confirmed by the observation of 2 diagnostic $d_{\alpha N}(i, i+3)$ NOEs [6] between residues 51 and 54, and 52 and 55. The first two residues in the helical segment of parsley plastocyanin have $^3J_{HN\alpha}$ values $<6\text{Hz}$, but the next two have values $>6\text{Hz}$. For Ser 56, the near-degeneracy of the $C^\alpha H$ and one of the $C^\beta H$ chemical shifts prevents determination of the $^3J_{HN\alpha}$ value for this residue. A similar pattern of $^3J_{HN\alpha}$ values is observed in the helical portion of French bean plastocyanin [5], for which $^3J_{HN\alpha}$ coupling constants $>6\text{Hz}$ for residues Lys 54, Ile 55 and Ser 56 indicate possible fraying or distortion at

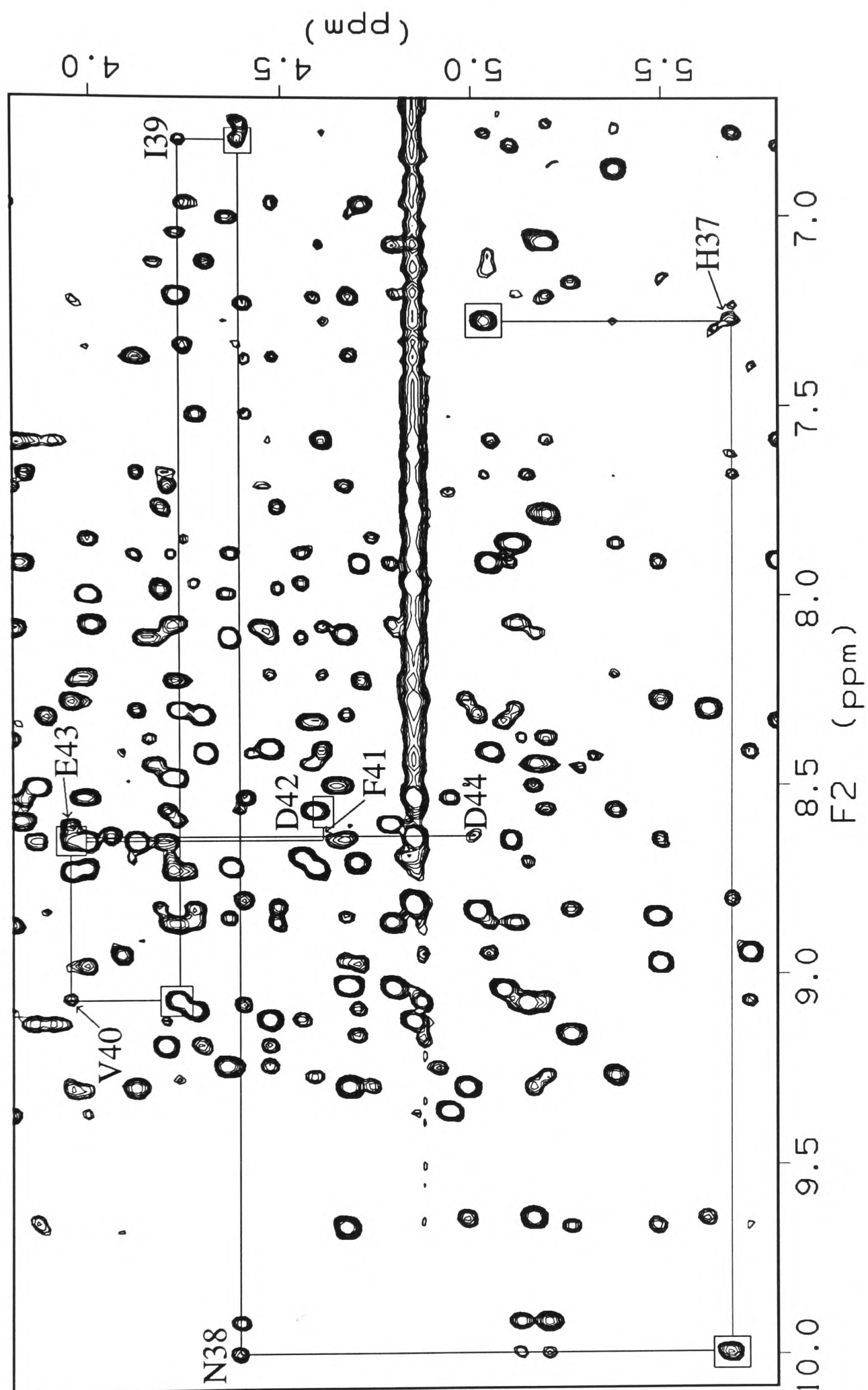


Figure 6.14: Sequential resonance assignments for polypeptide segments in extended conformation V. Part of the NOESY spectrum of parsley Cu^I plastocyanin recorded in H_2O at 35°C , pH 7.3 with a mixing time of 150ms. Sequential $d_{\alpha N}$ connectivities for residues 36–44 are shown.

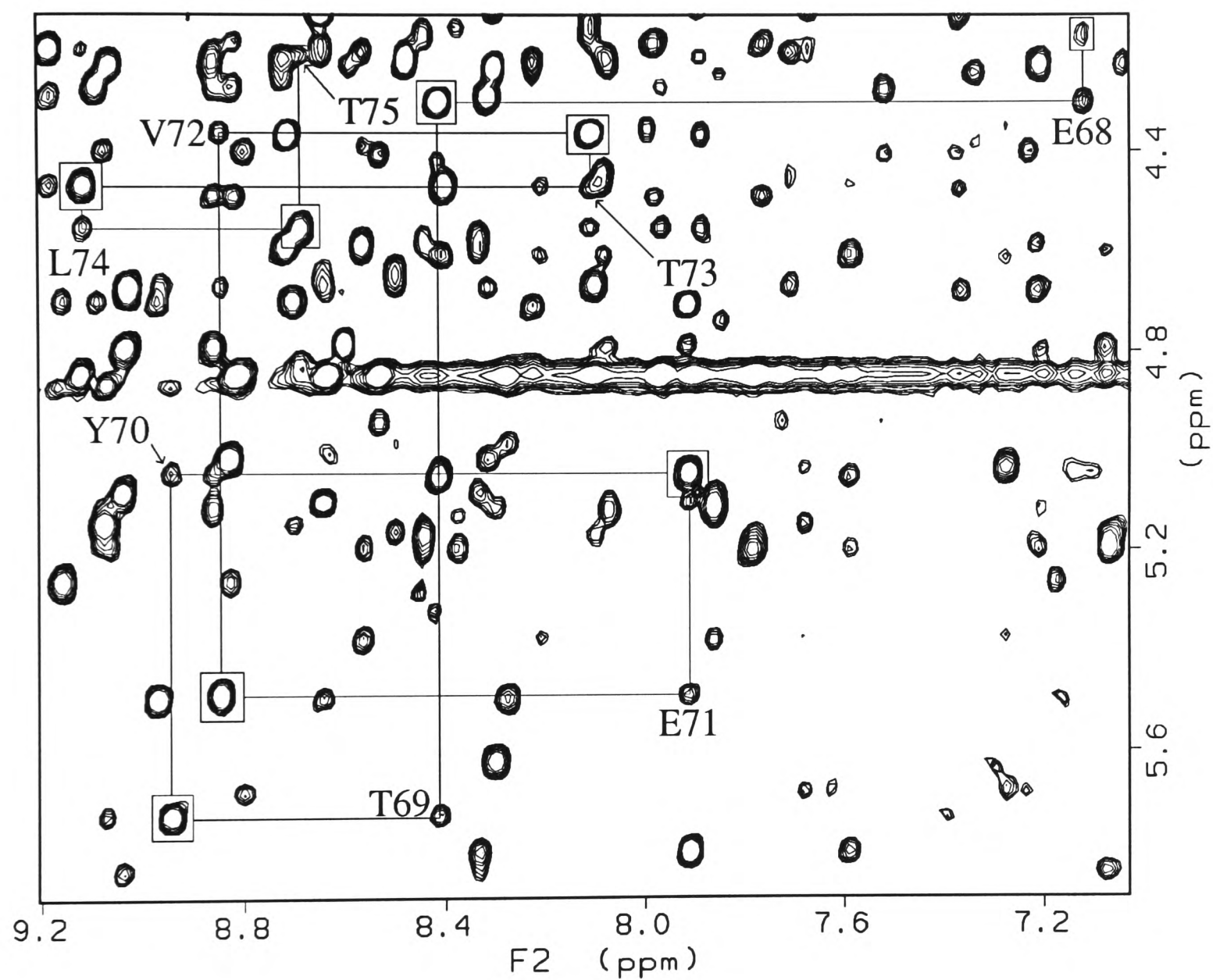


Figure 6.15: Sequential resonance assignments for polypeptide segments in extended conformation VI. Part of the NOESY spectrum of parsley Cu^I plastocyanin recorded in H_2O at 35°C , pH 7.3 with a mixing time of 150ms. Sequential $d_{\alpha N}$ connectivities for residues 68(66)–75(73) are shown.

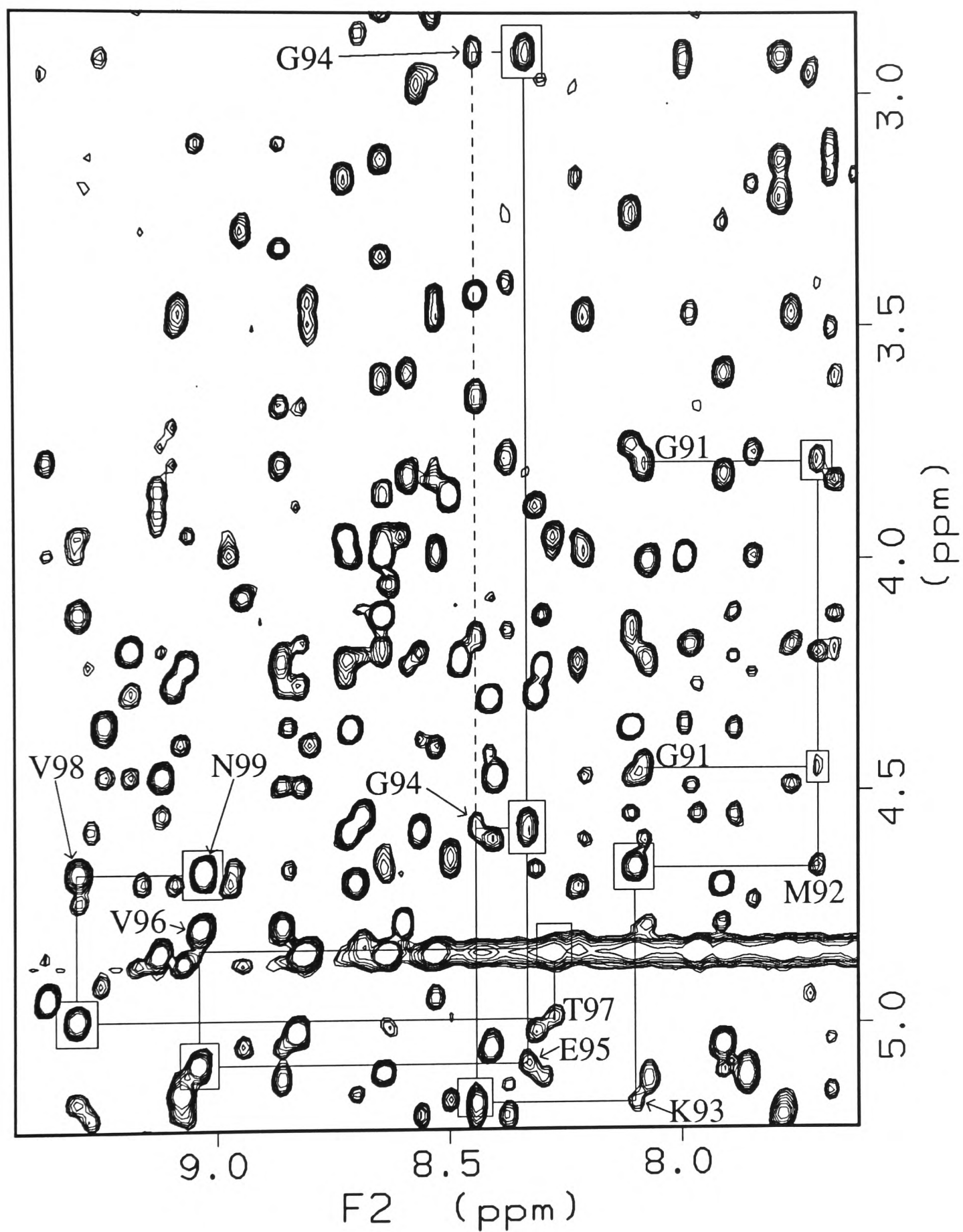


Figure 6.17: Sequential resonance assignments for polypeptide segments in extended conformation VIII. Part of the NOESY spectrum of parsley Cu^I plastocyanin recorded in H₂O at 35°C, pH 7.3 with a mixing time of 150ms. Sequential $d_{\alpha N}$ connectivities for residues 92(90)–99(97) are shown.

the C-terminal end of the helix [5]. There is evidence for similar distortion in the helical segment of parsley plastocyanin.

The appearance in the spectra of the resonance of the O γ H group of Ser 56 is a rare phenomenon and indicates low solvent accessibility. That this proton should also give a cross peak in HOHAHA spectra, and so be involved in coherence transfer, is even more unusual and suggests participation in a very stable hydrogen bond [5]. Such a bond is observed in the crystal structure of poplar plastocyanin [7], where O γ H of Ser 56 has low solvent accessibility and is hydrogen bonded to the peptide carbonyl of Ala 52.

6.4.3 Alignment and Packing of β -Strands

Alignment of β -strands can be deduced from cross-sheet NOEs between protons on adjacent strands. Both parallel and antiparallel packing of β -strands are observed in plastocyanin. These two arrangements have been shown to have distinctive patterns of NOEs [6]. Since the cross-sheet $d_{\alpha N}(i, j)$ distances of 3.2Å (antiparallel β -sheet) and 3.0Å (parallel β -sheet) obtained from model structures [6] are greater than the typical $d_{\alpha N}(i, i + 1)$ distance of 2.2Å, long range and sequential $d_{\alpha N}$ connectivities can usually be distinguished on the basis of cross peak intensity. This is particularly reliable if a short mixing time NOESY is used, so that spin diffusion does not lead to false intensities.

The pattern of interstrand NOEs observed in parsley plastocyanin is indicated in Chapter 7. It is essentially the same as those in the plastocyanins from spinach [8], French bean [5] and *Scenedesmus obliquus* [9], indicating that the overall folds of these proteins are very similar.

The pattern of NOEs around the contact between residues 7 and 14 suggests the presence of a β -bulge [10] in this region of both French bean plastocyanin [5] and *S. obliquus* plastocyanin [9]. It seems that parsley plastocyanin may also contain such an element of structure, which is not observed in poplar plastocyanin. The observation of intense $d_{\alpha N}(i, i + 1)$ connectivities and slow backbone amide proton exchange rates for residues 4–6 and 13–15 indicate regular antiparallel packing of β -sheets. As in French bean [5] and *S. obliquus* [9] plastocyanins, however, a cross-

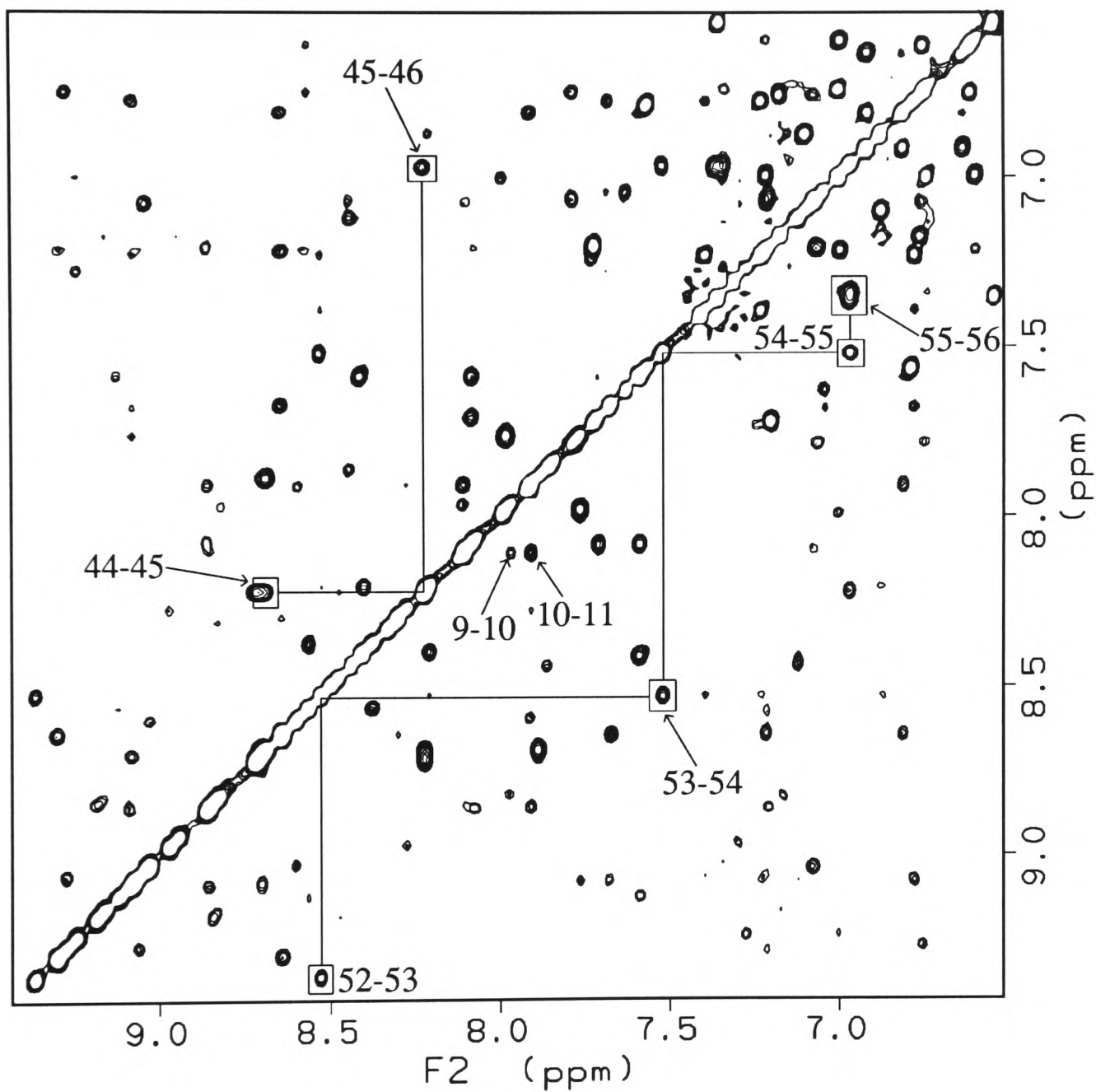


Figure 6.18: NH-NH region of the NOESY spectrum of parsley Cu^I plastocyanin recorded in H_2O at 35°C , pH 7.3 with a mixing time of 150ms. The string of sequential d_{NN} NOEs which demark the single helical segment of parsley plastocyanin is labelled (below the diagonal). Also indicated are the shorter series of sequential d_{NN} NOEs for residues 9–11 and 44–46.

sheet d_{NN} connectivity is observed between residues 7 and 13. This is instead of the $d_{\alpha N}(7,13)$ connectivity expected if the regular antiparallel sheet continued through the next residue. Further, the backbone amide proton of Val 13 is observed to have a slow exchange rate, implying hydrogen bonding to the peptide carbonyl of Ser 7. This, together with the NOE pattern, evidences the presence of a β -bulge [10] at residues Gly 6 and Ser 7, with the Ser 7 amide proton twisted to the same side of the strand as the Gly 6 amide proton (see Chapter 7 for more detailed analysis of the structure in this region of the protein).

6.4.4 Turns and Loops

Tight turns are the most prevalent element of nonrepetitive protein structure that has been recognised. They involve a peptide segment of four residues [10], with residues 2 and 3 forming the actual turn [11]. The conformation of such turns is defined by the torsion angles ϕ_2 , ψ_2 , ϕ_3 and ψ_3 . Eight types of turn have been identified [10]. The commonest are type I and type II, and their mirror images type I' and type II', where the sign of the torsion angles is reversed.

Criteria for the identification and classification of reverse turns using NMR parameters have been set out [6, 12]. Observation of intense $d_{NN}(3,4)$ and medium $d_{\alpha N}(2,4)$ NOEs distinguishes turns from α -helical, extended and irregular conformations. In type II turns, the torsion angle ϕ_3 is hardly constrained by the NOE data: the distances $d_{NN}(3,4)$ and $d_{\alpha N}(2,4)$ change little if ϕ_3 is rotated from $+90^\circ$ through 0° to -90° . This has led to the distinction of a true type II turn with $\phi_3 \approx +90^\circ$ from a half-turn with $\phi_3 \approx -90^\circ$ [12]. Type II and half-turn can be experimentally distinguished by the fact that $^3J_{HN\alpha}$ for residue 3 should be about 5Hz for type II and about 8Hz for half-turn.

It has been cautioned [2] that the identification of turns using NMR parameters is less reliable than for other secondary structures, and that, with the exception of hairpin turns in antiparallel β -sheet, the location of tight turns should be delayed until a tertiary structure is available. There are three reasons for this: the isolated nature of the d_{NN} and $d_{\alpha N}$ connectivities, the similarity of the distance constraints to those in helical segments, especially for type I turns and 3_{10} helices, and the generally

rapid exchange of hydrogen bonded amide protons due to the tendency of tight turns to be surface exposed in globular proteins. The last of these factors hinders the identification of hydrogen bonds in turns. With these points in mind, the NMR data have been analysed to see if any turns can be reliably defined, without consideration of the detailed structure of parsley plastocyanin presented in the next chapter.

The first step in the identification of turns was to look for isolated, intense d_{NN} NOEs, which were assumed to correspond to residues 3 and 4 of the turn. The important $d_{\alpha N}(i, i+2)$ connectivity was then sought. Once located, the type of a turn can theoretically be determined from the nature and intensity of the sequential NOE involving the two middle residues. Confirmation of the type may be obtained from the $^3J_{HN\alpha}$ value of residue 3 of the turn. For example, a value of $\approx 9\text{Hz}$ indicates a type I turn, one of $\approx 5\text{Hz}$ a type II turn. Residue 2 of both these types, and of the half-turn, has a distinctively low $^3J_{HN\alpha}$ coupling constant of around 4Hz . This is useful in the location of these types of turn.

In order of sequence, the first intense d_{NN} connectivity is that between residues 10 and 11. A medium d_{NN} connectivity is observed between residues 9 and 10. A similar pattern of sequential connectivities has been observed in French bean plastocyanin [5] for residues 8–11. Lack of a $d_{\alpha N}(2,4)$ connectivity means that a regular β -turn cannot be assigned with any confidence in this region of parsley plastocyanin. This contrasts with the assignment of a regular type I turn in the poplar plastocyanin crystal structure [7].

The turn at residues 14–17 of French bean and *S. obliquus* plastocyanins has been designated type VIa [10]. This is based on a strong $d_{\alpha\alpha}$ NOE, characteristic of a *cis*-peptide linkage, between residues 15 and 16, and a $d_{\alpha N}(2,4)$ connectivity involving residues 15 and 17. Such a turn cannot be assigned in parsley plastocyanin, mainly because the $C^\alpha\text{H}$ chemical shifts of residues 15 and 16 are essentially the same. This prevents observation of a $d_{\alpha\alpha}$ NOE between these residues.

The presence of a turn is indicated by an intense d_{NN} connectivity between residues 24 and 25 in the middle of a long sequence of $d_{\alpha N}$ NOEs. This is confirmed through a weak $d_{\alpha N}(2,4)$ connectivity for residues 23 and 25. Ala 23 has $^3J_{HN\alpha} < 4\text{Hz}$, and there is a strong $d_{\alpha N}$ connectivity between Ala 23 and Gly 24.

These data are not inconsistent with the assignment of this turn to type II or half-turn. Distinction of these alternatives cannot be made in this case since the $^3J_{HN\alpha}$ coupling constants of residue 3 (Gly 24) have not been determined.

Residues Gly 34 and Phe 35 are connected by a medium d_{NN} NOE, Ala 33 and Gly 34 by a medium $d_{\alpha N}$ NOE. The latter connectivity indicates that this is not a type I turn. Ala 33 (residue 2 of the putative turn) has a $^3J_{HN\alpha}$ value of 6.9Hz, suggesting that this turn is neither type II nor half-turn. In *S. obliquus* and poplar plastocyanins, the 32–35 fragment is described as a loop [9, 7]. However, the available NMR data for parsley plastocyanin are not inconsistent with this fragment being a type I' turn.

Consecutive intense d_{NN} connectivities are observed between residue pairs 44,45 and 45,46. A $d_{\alpha N}(43,45)$ NOE is also present, implying a β -turn at residues 42–45. This may be assigned as type I on the basis of a $^3J_{HN\alpha}$ for Asp 44 of 9.7Hz. However, lack of further information, such as a reliable value for the $^3J_{HN\alpha}$ coupling constant of Glu 43, means that this assignment must be tentative at present. The observation of medium intensity $d_{\alpha N}(42,43)$ and $d_{\alpha N}(43,44)$ NOEs casts further doubt on the veracity of the assignment.

An intense $d_{NN}(49,50)$ NOE and weak $d_{\alpha N}(48,50)$ NOE suggest the presence of a turn at residues 47–50. Supporting evidence is provided by the $^3J_{HN\alpha}$ value for Ala 48 of 4.6Hz. The turn is cautiously described as a type II or half-turn on the basis of a strong $d_{\alpha N}(48,49)$ connectivity. It is not possible to distinguish the two alternatives, since residue 3 of the turn is a glycine (as in the case of the 22–25 turn described above).

The next turn in the sequence is located from the reasonably intense d_{NN} NOE between Glu 68(66) and Gly 67(65), and a $d_{\alpha N}(i, i+2)$ connectivity involving Ala 66(64) and Glu 68(66). The middle two residues of this turn are linked by a $d_{\alpha N}$ NOE of medium intensity, favouring assignment to type II or half-turn. This cannot be further supported by coupling constant data due to the low intensity of the fingerprint cross peak of Ala 66(64), and because residue 67(65) is a glycine.

Two consecutive long strings of $d_{\alpha N}$ connectivities are interrupted by an intense d_{NN} NOE linking Glu 76(74) and Thr 75(73). Taking these two residues as 3 and 4 of

a putative turn, other relevant NOEs are a medium and a strong $d_{\alpha N}$ linking the pairs Leu 74(72), Thr 75(73) and Thr 73(71), Leu 74(72) respectively. The critical $d_{\alpha N}(2,4)$ connectivity is also present between Leu 74(72) and Glu 76(74). Nevertheless, $^3J_{HN\alpha}$ coupling constants for the four residues of 9.9Hz, 9.1Hz, 10.5Hz and 6.2Hz indicate that this fragment can by no means be classified as a regular β -turn.

The final portion of the plastocyanin molecule containing turns is the Cu-binding loop, so called because it contains three of the four Cu ligands in a nine-residue sequence from Cys 84(82) to Met 92(90). There are neither d_{NN} nor $d_{\alpha N}(i, i+2)$ connectivities in the 84(82)–87(85) fragment, and obviously there is no $^3J_{HN\alpha}$ value for Pro 86(84). Two strong d_{NN} NOEs can, however, be identified elsewhere in the Cu-binding loop. These link Ala 90(88) with Gly 91(89) and Gly 91(89) with Met 92(90). Taking the first of these pairs as residues 3 and 4 of a β -turn, the medium d_{NN} NOE between residues 90(88) and 89(87) and the $^3J_{HN\alpha}$ value of 9Hz for Ala 90(88) indicate that residues 88(86)–91(89) form a type I turn.

More definitive analysis of the β -turns in parsley plastocyanin will be given in Chapter 7, which includes the determination and detailed description of the three dimensional structure of parsley plastocyanin.

Bibliography

- [1] K. Wüthrich, G. Wider, G. Wagner & W. Braun (1982) *J. Mol. Biol.* 155, 311–319.
- [2] K. Wüthrich, ‘NMR of Proteins and Nucleic Acids’, Wiley, New York, 1986.
- [3] P. C. Driscoll, H. A. O. Hill & C. Redfield (1987) *Eur. J. Biochem.* 170, 279–292.
- [4] W. J. Chazin, M. Rance & P. E. Wright (1988) *J. Mol. Biol.* 202, 603–622.
- [5] W. J. Chazin & P. E. Wright (1988) *J. Mol. Biol.* 202, 623–636.
- [6] K. Wüthrich, M. Billeter & W. Braun (1984) *J. Mol. Biol.* 180, 715–740.
- [7] J. M. Guss & H. C. Freeman (1983) *J. Mol. Biol.* 169, 521–563.
- [8] P. C. Driscoll (1987) D. Phil. thesis, University of Oxford.
- [9] J. M. Moore, W. J. Chazin, R. Powls & P. E. Wright (1988) *Biochemistry* 27, 7806–7816.
- [10] J. S. Richardson (1981) *Advan. Protein Chem.* 34, 167–339.
- [11] B. L. Sibanda & J. M. Thornton (1985) *Nature* 316, 170–174.
- [12] G. Wagner, D. Neuhaus, E. Wörgötter, M. Vašák, J. H. R. Kägi & K. Wüthrich (1986) *J. Mol. Biol.* 187, 131–135.

Chapter 7

High Resolution Solution Structure of Parsley Plastocyanin

7.1 Experimental Restraints for Parsley Plastocyanin Structure Calculation

7.1.1 Approximate Interproton Distance Restraints

Interproton distance restraints were obtained from the NOESY experiments detailed in Chapter 3. Experiments were carried out under a variety of conditions, with resulting small changes in resonance frequency, in order to overcome the problem of spectral overlap. Ambiguities in assignments and consequently in distance restraints could thus be minimised. Lower temperature H₂O experiments also allowed the observation of fingerprint cross peaks for those residues with more rapidly exchanging backbone amide protons. The close similarity in the pattern and intensity of NOEs observed under this range of experimental conditions indicates that there are no significant conformational changes.

All NOEs were classified as strong, medium or weak, corresponding to the ranges 1.8–2.7Å, 1.8–3.5Å and 1.8–5.0Å in the interproton distance restraints. Upper limits of distances involving methyl groups, nonstereospecifically assigned methylene protons and C^δ and C^ε protons of phenylalanine and tyrosine, were corrected for centre averaging [1] (equivalent to the use of pseudoatoms). An additional 0.5Å was added to the upper limits for distances involving methyl groups to account for the greater apparent intensity of methyl resonances [2, 3].

7.1.2 Dihedral Angle Restraints and Stereospecific Assignment of β -Methylene Protons from Qualitative Data Analysis

ϕ torsion angle restraints were obtained on the basis of $^3J_{HN\alpha}$ coupling constants. The $^3J_{\alpha N}$ coupling constants of all non-glycine residues were measured in high digital resolution DQF COSY spectra using spectral simulations [4, 5], thus correcting for distortions arising from finite linewidths. A $^3J_{HN\alpha} < 7\text{Hz}$ was taken to indicate the presence of α secondary structure, and the ϕ torsion angle assigned a value of $-65^\circ \pm 25^\circ$. β secondary structure was inferred for $^3J_{HN\alpha} > 9\text{Hz}$, and the ϕ torsion angle was then assigned a value of $-120^\circ \pm 40^\circ$ [6] (Figure 4.1).

χ_1 torsion angle restraints and stereospecific assignment of prochiral β -methylene protons were obtained on the basis of $^3J_{\alpha\beta}$ coupling constants and intraresidue NOEs involving the backbone amide, C^α and C^β protons. (Values of the $^3J_{\alpha\beta}$ coupling constants of non-valyl residues were determined as passive coupling constants, where possible, from the relative displacement of cross peak components in P.E.COSY spectra recorded in D_2O). These coupling constant and distance parameters have a characteristic pattern for each of the three energetically favourable staggered rotamer positions, corresponding to $\chi_1 = 60^\circ, 180^\circ$ or -60° [2]. When a preferred rotamer position of a side chain was identified, the torsion angle restraint used in the structure calculation was described by a square well effective potential function [7] of width $\pm 20\text{--}40^\circ$ about the rotamer positions [8]. Based on this qualitative analysis, stereospecific assignments were obtained for twenty-three of the forty-nine non-degenerate β -methylene proton pairs in parsley plastocyanin.

In cases where both $^3J_{\alpha\beta}$ coupling constants had values of 6–9Hz, indicative of motional averaging, conformational heterogeneity (i.e. a mixture of rotamer populations) about the $C^\alpha\text{--}C^\beta$ bond was assumed and neither stereospecific assignment nor χ_1 torsion angle restraint was obtained. This was the case for six of the pairs of β -methylene protons for which $^3J_{\alpha\beta}$ coupling constants could be determined (Lys 30, Asn 51, Glu 53, Glu 68(66), Lys 93(91), and Glu 95(93)). For residues Lys 4, Asp 42, Glu 43, Lys 54, Gln 59(57), Leu 63(61), Lys 77(75) and Met 92(90), no $^3J_{\alpha\beta}$ coupling constants could be obtained, due to low cross peak intensity or complete absence of

one or both $\alpha\beta$ cross peaks. For other residues, the $^3J_{\alpha\beta}$ coupling constants could not be measured due to chemical shift degeneracy or near degeneracy of the C^β protons, or due to overlap with cross peaks from other residues. This was the case for Ser 17, Ser 18, Asn 31, Asn 38, Phe 41 and Glu 53.

7.1.3 Stereospecific Assignments and Dihedral Angle Restraints from STEREOSEARCH

Stereospecific assignments of β -methylene protons and ϕ and χ_1 dihedral angle restraints were initially obtained by qualitative analysis of 3J coupling constant and intraresidue NOE data, as described above. This method is only applicable to unambiguous situations and provides no quantitative assessment of the accuracy of the assignment.

In the later stages of the structure calculation process, the program STEREOSEARCH [10, 11] became available. This can be used to obtain stereospecific assignments of β -methylene protons and ϕ , ψ and χ_1 dihedral angle restraints. Experimental $^3J_{HN\alpha}$ and $^3J_{\alpha\beta}$ values, obtained as described above, and sequential and intraresidue NOEs involving NH, $C^\alpha H$ and $C^\beta H$ protons, are matched with those calculated for conformations present in a data base. The data base employed was a systematic one with idealised geometry in which the ϕ , ψ and χ_1 angles of the central residue of a tripeptide fragment were varied in 10° steps. For each of the two possible assignments, an indication is given of the number of conformations in the data base which conform to the experimental data. The results were mostly in agreement with the previous qualitatively based stereospecific assignments. For Phe 19 and Phe 82(80), however, the numbers of conformational matches indicated that the initial assignments should be reversed. The minimum ranges employed for ϕ , ψ and χ_1 dihedral angle restraints were $\pm 30^\circ$, $\pm 50^\circ$ and $\pm 20^\circ$ [11]. The final list of stereospecifically assigned β -methylene protons and corresponding χ_1 restraints are shown in Table 7.1.

Residue	Assignments		χ_1 restraint (deg)
	β_1 (ppm)	β_2 (ppm)	
Glu 2	1.85	1.93	-60 $\pm$ 30
Leu 5	0.91	2.35	-60 $\pm$ 20
Ser 7	3.67	4.24	60 $\pm$ 20
Leu 12	0.94	1.71	-60 $\pm$ 20
Phe 14	2.63	3.12	-70 $\pm$ 25
Ser 15	3.34	3.80	-30 $\pm$ 30
Phe 19	3.14	3.36	60 $\pm$ 20
Glu 25	2.25	2.10	170 $\pm$ 30
Phe 29	2.82	3.31	-65 $\pm$ 20
His 37	3.53	2.61	-60 $\pm$ 20
Phe 41	2.72	2.80	-60 $\pm$ 30
Asp 44	2.84	2.70	-60 $\pm$ 20
Glu 45	2.36	1.69	-90 $\pm$ 30
Tyr 62(60)	2.72	2.91	-80 $\pm$ 30
Leu 63(61)	1.68	1.30	75 $\pm$ 40
Glu 71(69)	1.63	1.77	-45 $\pm$ 30
Leu 74(72)	0.92	1.56	-50 $\pm$ 20
Tyr 80(78)	2.97	3.62	-70 $\pm$ 30
Lys 81(79)	2.12	1.90	-50 $\pm$ 30
Phe 82(80)	1.57	1.76	40 $\pm$ 25
Tyr 83(81)	3.20	3.14	90 $\pm$ 30
Cys 84(82)	2.93	3.24	180 $\pm$ 30
His 87(85)	3.79	3.40	-60 $\pm$ 20

Table 7.1: Stereospecific β -methylene proton assignments and corresponding χ_1 restraints used in the calculation of parsley Cu^I plastocyanin structures.

7.1.4 Stereospecific Assignment of Methyl Groups of Val and Leu

Stereospecific assignments were obtained for the methyl groups of eight of the nine valine residues in parsley plastocyanin. The basis for these assignments were $^3J_{\alpha\beta}$ coupling constants (from high digital resolution DQF COSY spectra by spectral simulation [4, 5]) and the relative intensities of intraresidue NH-C γ H and C α H-C γ H NOEs [9]. Stereospecific assignments were obtained for the methyl groups of all four leucine residues during the later stages of the calculation process, when the side chains exhibited well defined conformations.

7.1.5 Hydrogen Bond Restraints

Sixty distance restraints corresponding to backbone NH-CO hydrogen bonds were included in the input for parsley plastocyanin structure calculation. Fifty-four of these were for interstrand hydrogen bonds, six for hydrogen bonds in turns. Also included were four distance restraints for the hydrogen bonds between the hydroxyl proton of Ser 56 and the backbone carbonyl of Ala 52, and between the cysteinyl sulphur of Cys 84(82) and the backbone NH of Asn 38. Only those hydrogen bonds were included which could be assigned unambiguously on the basis of qualitative interpretation of NOE and amide exchange data. Two distance restraints were used for each hydrogen bond, one between the donor heavy atom and the acceptor atom of 1.8–3.5Å, and one between the hydrogen and the acceptor atom of 1.8–2.3Å.

7.1.6 Peptide Bond Restraints

All peptide bonds were restrained to be planar and *trans*, with the exception of those between Ser 15 and Pro 16 and between Phe 35 and Pro 36, which were restrained to be planar and *cis*. For these the ω torsion angle was constrained to $0 \pm 30^\circ$, and for *trans*-prolines to $180 \pm 30^\circ$.

7.1.7 Cu Restraints

Cu was incorporated by following the example for Fe in the X-PLOR manual [12]. Cu-ligand bond lengths and distances between pairs of ligating atoms were restrained to values taken from the crystal structure of reduced poplar plastocyanin at pH 7.0 [13], with a range of $\pm 0.2\text{\AA}$.

7.2 Structure Calculations

7.2.1 Overall Strategy

Structure calculations were carried out iteratively, with each successive round of calculations employing a larger experimental data set. Ambiguities among distance restraints could often be resolved by analysis of earlier low resolution structures. During the later stages of the calculation process, additional stereospecific assignments could be made for side chains with well defined conformations. This was possible for the α -methylenes of nine of the twelve glycine residues, and, as mentioned above, for the methyl groups of all four leucine residues.

Three protocols were employed during the course of the calculations. The method used to generate initial low resolution structures, termed the Quick method, employed the SHAKE algorithm (see below) to minimise calculation time. The main aim of these early calculations was to assess whether distance restraints were correct and to resolve ambiguities and uncertainties among NOE cross peaks. Subsequent calculations used hybrid metric matrix distance geometry-dynamical simulated annealing [14], henceforth referred to as the Hybrid method. This requires a longer calculation time per structure than the Quick method, and gives a more accurate representation of the conformational space sampled by the protein. A combination of two protocols was employed in the third stage of the calculation process. Firstly, simulated annealing was used to fold polypeptide chains with random backbone dihedral angles (Yasap method) [16, 17]. The resulting structures were then refined using the Hybrid method. Structures were displayed on a Silicon Graphics system using the program INSIGHT (Biosym Technology).

All three protocols used the program X-PLOR [12, 18] which is derived from

CHARMM [19] and has been adapted for dynamical simulated annealing calculations with NMR restraints [20, 7].

The general philosophy of these methods was described in Chapter 4. Specific details are given below.

7.2.2 The Hybrid Method

This protocol [14] comprises three stages. The first is the generation using DISGEO [15] of a set of substructures which contain only about a third of the atoms and have approximately the correct fold. (DISGEO, employing distance geometry (Chapter 4) and the EMBED algorithm [15], was the first program written specifically for calculation of protein structure from NMR-derived data). This set is embedded from n -dimensional distance space into Cartesian coordinate space without checking triangle inequalities. Included are the N, C, C $^\alpha$, C $^\alpha$ H, C $^\beta$ and non-terminal C $^\gamma$ and C $^\delta$ atoms, plus a pseudoatom for aromatic rings. In the second stage, all remaining atoms are added with the side chains in extended conformation. Finally, the resulting structures are subjected to dynamical simulated annealing.

The annealing phase consists of simulating the system at a temperature T by solving Newton's equations of motion using X-PLOR. The total target function consists of the following terms:

$$F_{tot} = F_{covalent} + F_{repel} + F_{NOE} + F_{tor} \quad (7.1)$$

$F_{covalent}$ maintains correct bond lengths, angles, chirality and planes and comprises quadratic harmonic terms for these:

$$F_{covalent} = \sum_{bonds} k_b(r - r_0)^2 + \sum_{angles} k_\phi(\phi - \phi_0)^2 + \sum_{impropers} k_\varphi(\varphi - \varphi_0)^2 + \sum_{\omega} k_\omega(\omega - \omega_0)^2 \quad (7.2)$$

The force constants for bonds (k_b), angles (k_ϕ), improper torsion angles (k_φ) and peptide bond dihedral angles (k_ω) are set to 500 kcal mol $^{-1}$ Å $^{-2}$ and 500 kcal mol $^{-1}$ rad $^{-2}$. These high values ensure near-perfect stereochemistry throughout the calculations.

Nonbonded interactions are represented by a simple van der Waals repulsion term with a variable force constant:

$$F_{repel} = 0 \quad \text{if} \quad r \geq s.r_{min}$$

$$F_{repel} = k_{vdW}(s^2 r_{min}^2 - r^2)^2 \quad \text{if } r < s.r_{min} \quad (7.3)$$

Values of r_{min} are given by the sum of the standard van der Waals radii between two atoms. The scale factor s is set to 0.8 during the later stages of the calculation.

Interproton distance restraints are incorporated as square well potentials:

$$\begin{aligned} F_{NOE} &= k_{NOE}(r_{ij} - r_{ij}^u)^2 \quad \text{if } r_{ij} > r_{ij}^u \\ F_{NOE} &= 0 \quad \text{if } r_{ij}^l \leq r_{ij} \leq r_{ij}^u \\ F_{NOE} &= k_{NOE}(r_{ij} - r_{ij}^l)^2 \quad \text{if } r_{ij} < r_{ij}^l \end{aligned} \quad (7.4)$$

where r_{ij}^u and r_{ij}^l are the upper and lower bounds of the target distances, and k_{NOE} is the variable NOE force constant. Torsion angle restraints are included in a similar form:

$$\begin{aligned} F_{tor} &= k_{tor}(\varphi_i - \varphi_i^u)^2 \quad \text{if } \varphi_i > \varphi_i^u \\ F_{tor} &= 0 \quad \text{if } \varphi_i^l \leq \varphi_i \leq \varphi_i^u \\ F_{tor} &= k_{tor}(\varphi_i - \varphi_i^l)^2 \quad \text{if } \varphi_i < \varphi_i^l \end{aligned} \quad (7.5)$$

where φ_i is the calculated value of a torsion angle.

The DISGEO substructures are subjected to two hundred cycles of Powell minimisation with no experimental potential, the hard sphere atomic radii are set to their standard Lennard-Jones values ($s = 1.0$), and $k_{vdW} = 0.001 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$. This improves covalent geometry prior to simulated annealing. The experimental restraints are then included and the system is subjected to fifty cycles of 75fs dynamics with 1fs time steps. Initial velocities are chosen from a Maxwell distribution at 1000K, and are rescaled to 1000K at the start of each cycle. k_{NOE} and k_{tor} are each initially $0.5 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$, and are doubled at the beginning of each cycle up to maximum values of $50 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$ and $200 \text{ kcal mol}^{-1} \text{ rad}^{-2}$. The initial value of k_{vdW} is very low ($0.001 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$), such that atoms are ‘soft’ and the barriers between different folds are of the order of the kinetic energy of the system. Atoms can then pass freely through each other. This ensures that the experimental restraints are satisfied and that an incorrect fold is avoided. k_{vdW} is increased slowly by multiplication by 1.125 before each cycle up to a final value of $0.25 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$. The second step of the

simulated annealing process consists of 1.5ps of dynamics (reduced to 0.5ps in the later rounds of calculation) at 300K, with a time step of 1fs (0.001ps). k_{NOE} and k_{tor} have the values they had at the end of the first stage of simulated annealing, and k_{vdW} is set to 4 kcal mol⁻¹ Å⁻² such that the atoms are harder and smaller. s is set to 0.8 to account for the fact that interatomic separations slightly smaller than the sum of the hard sphere van der Waals radii can occur. The final part of the protocol consists of two hundred cycles of restrained Powell minimisation. No electrostatic, hydrogen bonding, or 6-12 Lennard-Jones van der Waals terms are used in this protocol.

Calculation times using the Hybrid method were ≈ 3.5 hours per structure (Sun 4/110) or ≈ 1.5 hours per structure (Sparcstation2) using the final data set.

7.2.3 Folding an Extended Strand by Dynamical Simulated Annealing

The Yasap protocol [16, 17] is designed for structure calculation, as opposed to structure refinement. Starting structures consist of plastocyanin polypeptide chains with random backbone dihedral angles. These starting structures must be unfolded because an incorrectly folded chain cannot converge to the correct global minimum region.

The total target function, F_{tot} , $F_{covalent}$ and F_{repel} have the same form as those used in the Hybrid method (Equations 7.1, 7.2 and 7.3). The initial NOE target function is, however, more complex. It consists of a square well potential with harmonic walls at the upper and lower bounds of the distance restraints. The higher distance side has a soft asymptote with a harmonic/linear switching function at 0.5Å above the upper bound. This means that the energy of any distance restraint which is violated by more than 0.5Å will have a ‘soft’ dependence on $(r - r_0)$. This NOE potential function is designed for gradual incorporation of interproton distances between protons far apart in the sequence. This is so that secondary structure elements are formed prior to tertiary structure folding.

There are five steps in the Yasap method. The calculations are initiated with twenty steps of Powell minimisation to remove a number of bad nonbonded contacts. As in the Hybrid method, the force constants for bonds (k_b), angles (k_ϕ) and improper

torsions (k_ϕ) are set to 500 kcal mol⁻¹ Å⁻² and 500 kcal mol⁻¹ rad⁻². k_{NOE} and k_{tor} are set to 50 kcal mol⁻¹ Å⁻² and 50 kcal mol⁻¹ rad⁻². The last two force constants retain these values throughout the calculation. Note that the ‘normal’ dihedral angle potential is switched off ($k_\omega = 0$).

Next is the conformational search phase. This consists of 15ps (reduced to 9ps in later rounds of calculation) of high temperature dynamics. In order to reduce computation time, the SHAKE algorithm [21] was used in this phase. The principle behind SHAKE is that the length of the time step in a molecular dynamics simulation is limited by the highest frequency occurring in the molecular system of interest. Thus, the time step can be lengthened by constraining the degrees of freedom with the highest frequencies. Conventionally, SHAKE is used to keep bonds rigid. This permits an increase to a time step of 3fs rather than 1fs. Initial velocities are chosen from a Maxwell distribution at 1000K. The force constants k_ϕ and k_ψ must be reduced to 200 kcal mol⁻¹ rad⁻² and 50 kcal mol⁻¹ rad⁻² in order for SHAKE and a 3fs time step to work at 1000K. The force constant of the repulsive potential k_{vdW} is set to a very low value (0.002 kcal mol⁻¹ Å⁻²) to allow atoms to move freely through each other, as in the Hybrid method. The slope of the linear part of F_{NOE} is 0.1 kcal mol⁻¹ Å⁻¹.

The annealing phase consists of two stages. The first is 1ps of high temperature dynamics with a time step of 3fs, using the velocities from the previous phase. k_{vdW} is increased to 0.1 kcal mol⁻¹ Å⁻², and the slope of the asymptote of F_{NOE} is incremented.

For the second phase of annealing, during which the system is cooled to 300K, a normal square well potential is adopted for F_{NOE} . Spheres are made harder and smaller ($k_{vdW} = 4$ kcal mol⁻¹ Å⁻² and scale factor $s = 0.8$). Thirty cycles of dynamics are performed with a time step of 3fs.

In the fifth and final phase, the SHAKE constraints are switched off, and k_ϕ and k_ψ are each reset to 500 kcal mol⁻¹ rad⁻². Two hundred steps of steepest descent minimisation are carried out.

Calculation times using the Yasap method were ≈ 6 hours per structure (Sun 4/110) or ≈ 2.5 hours per structure (Sparcstation2) using the final data set.

7.2.4 The Quick Method

This is a variation on the Hybrid method in which the SHAKE algorithm is used to allow an increase in the time steps of the dynamical simulated annealing stages from 1fs to 3fs. The basic format of the method is the same as the Hybrid: extended side chains fitted to a DISGEO substructure, unrestrained minimisation prior to simulated annealing at 1000K, followed by cooling to 300K and restrained minimisation. k_{bond} is kept uniformly high at 500 kcal mol⁻¹ Å⁻².

Calculation times using the Quick method were $\approx$ 1 hour per structure (Sun 4/110 only) using relatively small data sets in the initial calculations.

7.3 The Converged Structures

The final structures were calculated using a total of 1403 approximate interproton distance restraints, comprising 64 hydrogen bond restraints, 14 Cu restraints, 280 intraresidue restraints, 340 sequential ($|i-j| = 1$), 120 medium range ($1 < |i-j| \leq 5$), and 585 long range ($|i-j| > 5$) interresidue restraints. An analysis of the NOE restraints per residue is presented in Figure 7.1, and a list of strong or medium long range ($|i-j| > 5$) $d_{\alpha N}$, d_{NN} and $d_{\alpha\alpha}$ NOEs is presented in Table 7.2. The location of these NOEs in the parsley plastocyanin fold can be ascertained from Figure 7.2, which is a schematic representation of the protein's topology. Figure 7.2 was taken from reference [22] and adapted to fit the primary structure of parsley plastocyanin. The final total of 127 dihedral angle restraints comprised 55 ϕ , 5 proline ω , 41 ψ and 26 χ_1 restraints.

Fifty eight structures were computed using the Yasap protocol. These were then refined using the Hybrid method. None of the structures was discarded on the basis of relatively high total energies or relatively large violations of restraints. All of the structures showed small deviations from idealised covalent geometry and had good nonbonded contacts, as evidenced by a low value of the van der Waals repulsion energy (ranging from 17 to 26 kcal mol⁻¹) and large negative values of the Lennard-Jones van der Waals energy (which is not used in the target function for simulated annealing calculations) (Table 7.3). Total energies ranged from 1647 to 1681 kcal

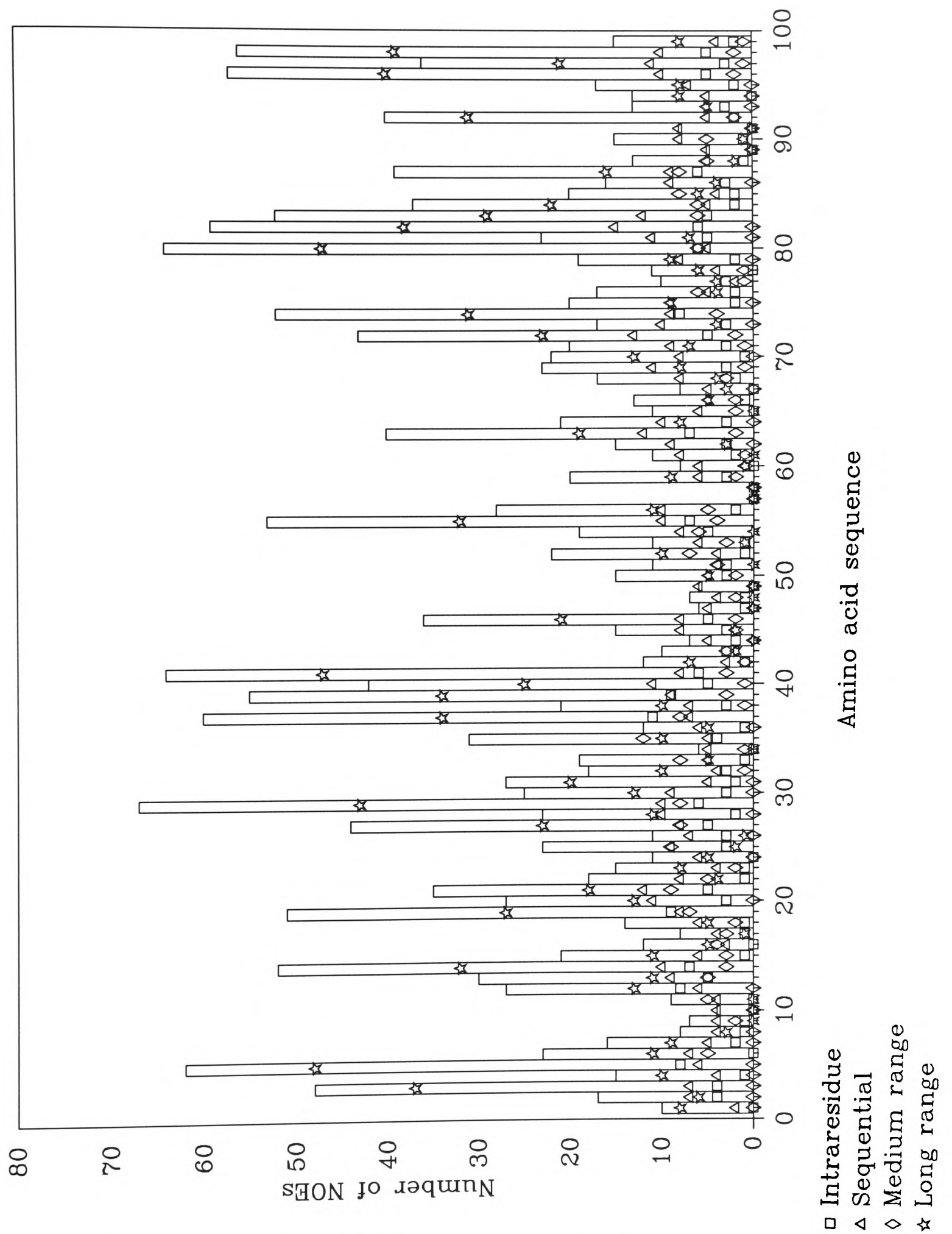


Figure 7.1: Plot of the number of NOE restraints obtained per residue *versus* sequence for parsley plastocyanin. The numbers of each type of restraint are indicated. The height of the histogram is the total number of distance restraints for the residue. The deletions at positions 57 and 58 are included in the sequence.

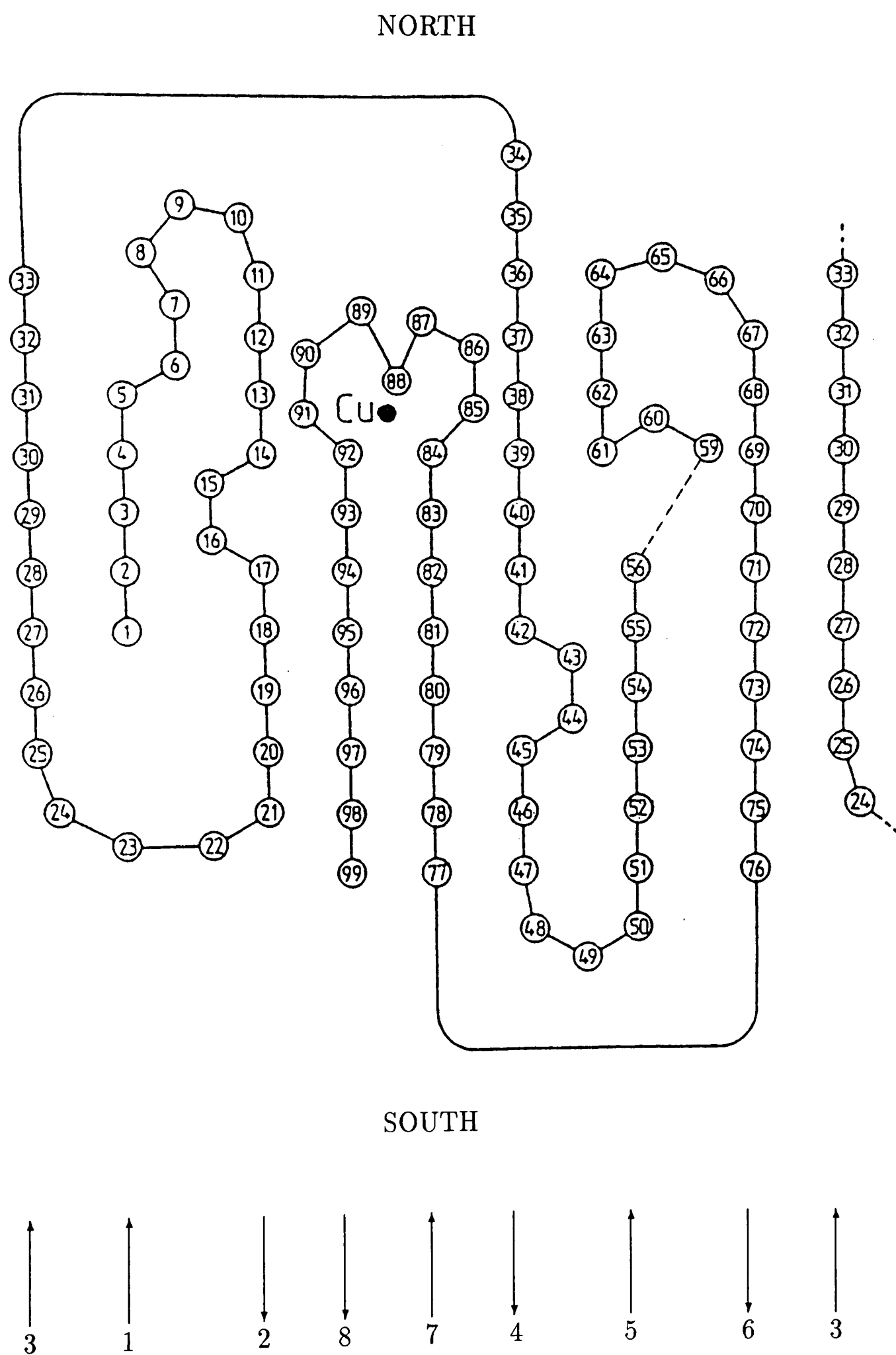


Figure 7.2: Schematic representation of the polypeptide fold of parsley plastocyanin.

$d_{N\alpha}$	d_{NN}	$d_{\alpha\alpha}$
3-29	4-15	5-14
5-31	5-30	28-71(69)
13-6	5-32	30-69(67)
19-96(94)	7-11	36-64(62)
21-98(96)	7-13	41-82(80)
28-2	19-97(95)	46-80(78)
29-71(69)	23-99(97)	79(77)-97(95)
30-4	27-72(70)	
31-69(67)	29-70(68)	
33-6 ₁	37-63(61)	
42-82(80)	40-83(81)	
94(92)-83(81)	78(76)-98(96)	
95(93)-18	80(78)-96(94)	
97(95)-20	82(80)-94(92)	
99(97)-22		

Table 7.2: Long range ($|i - j| > 5$) NOEs obtained for parsley Cu^I plastocyanin.

mol⁻¹. Different stereo views of the superposition of the backbone atoms are shown in Figures 7.3 and 7.4.

7.3.1 Aromatic Ring Alignment

In aromatic rings which are flipping rapidly on the NMR time scale, the two C^δ protons are magnetically equivalent to each other, and so are the two C^ε protons. This means that the two orientations of the ring related by a 180° rotation about the C^α-C^β bond are energetically equivalent. Since no stereospecific assignments of aromatic ring protons were used in the parsley plastocyanin structure calculations, each of the two orientations of aromatic rings will occur in the computed structures. This leads to difficulty in the calculation of an average structure as the rings take up an orientation approximately perpendicular to the original orientations. Subsequent to structure calculation, therefore, a program was used to flip aromatic rings so that the ring of each residue has the same orientation in each of the fifty-eight structures.

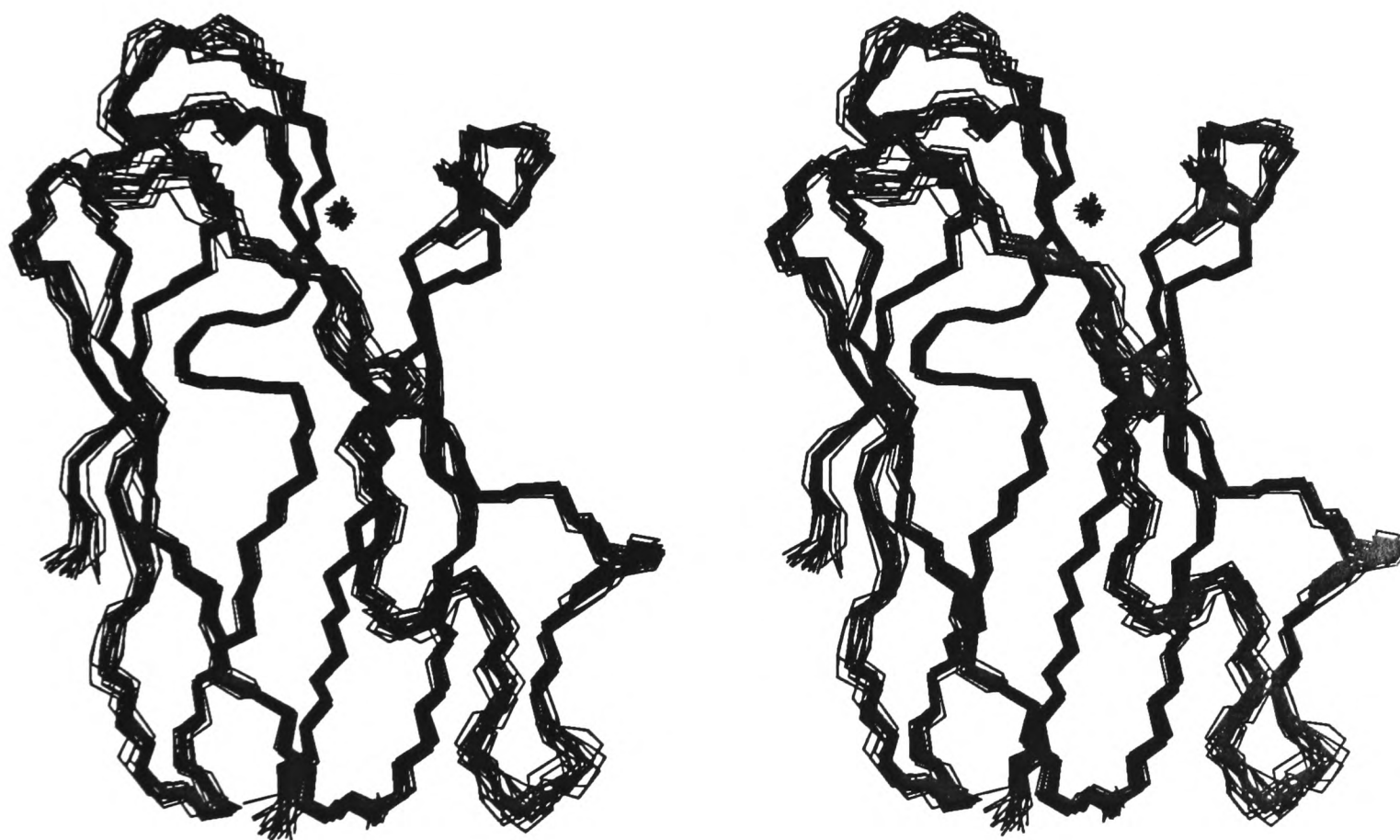


Figure 7.3: Stereo view showing best fit superpositions of the backbone (N, C^α and C) atoms of the yasap/simulated annealing structures of parsley Cu^I plastocyanin.

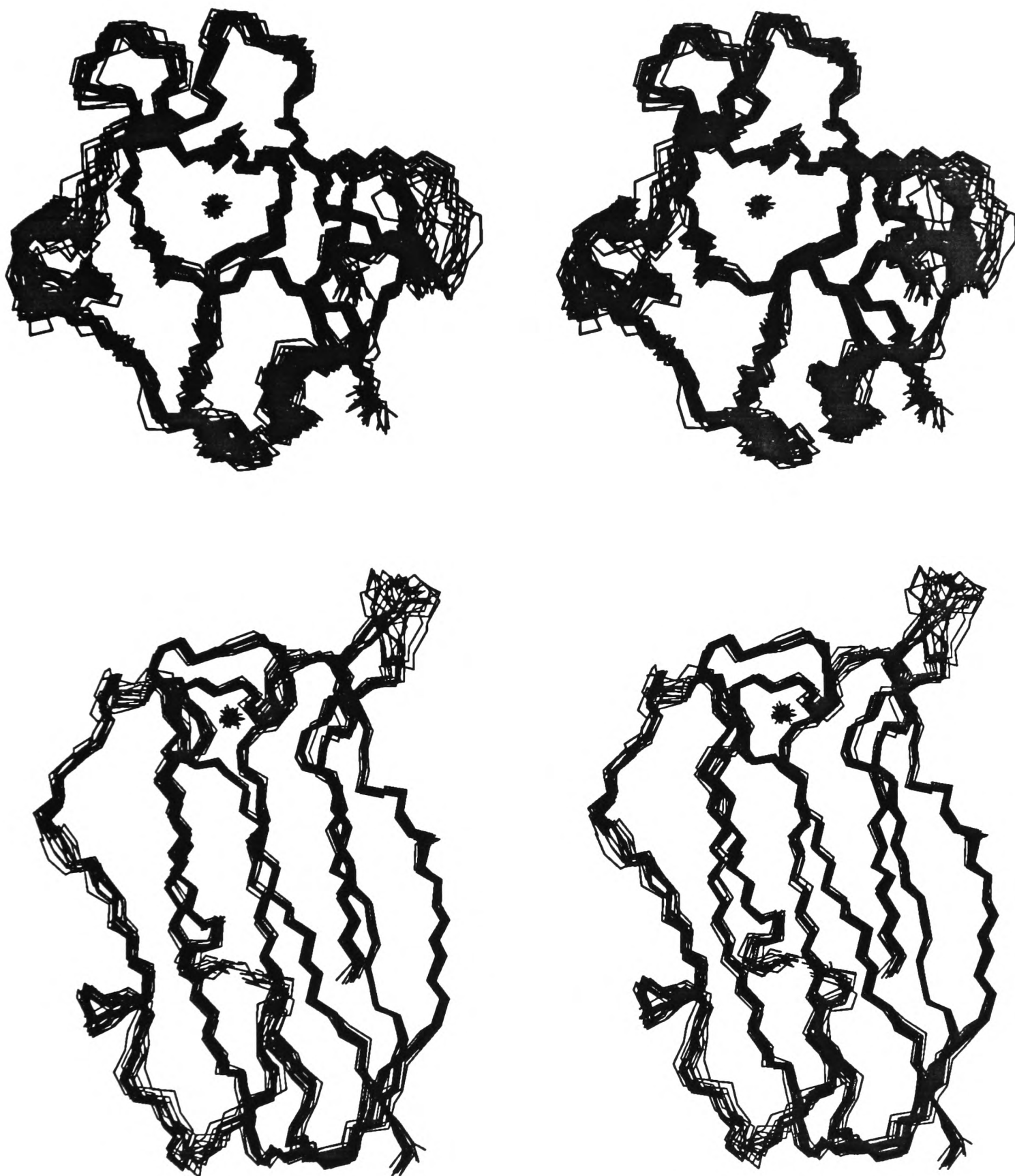


Figure 7.4: Stereo views showing best fit superpositions of the backbone (N, C $^{\alpha}$ and C) atoms of the yasap/simulated annealing structures of parsley Cu I plastocyanin. The top view looks down on the northern end of the molecule.

7.3.2 Fitting and Calculation of Average Structure

Following aromatic ring alignment, the structures were subjected to iterative fitting of the backbone atoms, with the most variable regions being identified and discarded in subsequent fits [23]. Residues 8–10 and 47–50 were ignored in the final fit. The coordinates were then averaged, producing an average structure, termed $\overline{SA}$, to which the individual structures were refitted. The average structure was subjected to 1000 cycles of restrained Powell minimisation to optimise geometry and nonbonded contacts. The resulting minimised average structure is termed $(\overline{SA})_r$. Structural statistics are presented in Table 7.3.

7.3.3 Accuracy and Precision of the Structures

The fifty-eight final structures of reduced parsley plastocyanin exhibit an atomic rms distribution about the mean coordinate positions of 0.61 ± 0.04 Å for backbone atoms. This figure excludes residues 8–10 and 47–50, which are poorly defined. One of the criteria by which the accuracy of the structures can be assessed is the extent to which they fit the experimental data. No residual distance restraint violation exceeded 0.32 Å, and no residual dihedral angle restraint violations were greater than 3°. These figures indicate that the parsley plastocyanin structures are acceptably accurate by NMR standards.

As is common in NMR-derived protein structures, regions of the backbone exhibiting regular secondary structure are particularly well defined. Loops and turns tend to show a greater variation in conformation. This is particularly true of the turns at residues 7–10 and 47–50, of which, only residue 7 was included in the fitting procedure. Of the side chains, those in the molecular interior tend to be better defined than those at the surface (Figures 7.5, 7.6 and 7.7). In many cases where stereospecific assignment of β -methylene protons could not be made, the side chains are reasonably well defined up to C^β . This is true, for example, of Lys 26, Lys 30, Asn 31, Asn 32, Asn 38, Asp 42, Glu 43, Lys 54, Gln 59(57), Lys 77(75), Lys 81(79), Lys 93(91) and Glu 95(93). Beyond C^β , most of these side chains are highly disordered. In other examples, where stereospecific assignment of β -methylene protons could be

Structural Statistics		
	$\langle SA \rangle$	$(\overline{SA})_r$
rms deviations from exptl. distance restraints (Å)		
all(1403)	0.018	0.019
sequential ($ i - j = 1$) (340)	0.011	0.016
medium range ($ i - j \leq 5$) (120)	0.017	0.019
long range ($ i - j > 5$) (585)	0.021	0.019
intraresidue (280)	0.018	0.019
H-bond (64)	0.018	0.022
Cu (16)	0.030	0.024
rms deviations from exptl. dihedral restraints (deg) (127)	0.223 ± 0.041	0.209
F_{NOE} (kcal mol ⁻¹)	15 ± 3	14
F_{tor} (kcal mol ⁻¹)	0.52 ± 0.42	0.70
F_{repel} (kcal mol ⁻¹)	21 ± 2	21
E_{L-J} (kcal mol ⁻¹)	-325 ± 13	-314
Deviations from idealised covalent geometry		
bonds (Å)	0.005 ± 0	0.005
cdih (deg)	0.246 ± 0.006	0.253
angles (deg)	1.904 ± 0.006	2.006
impropers (deg)	0.498 ± 0.005	0.501

Table 7.3: Structural statistics. $\langle SA \rangle$ are the fifty-eight final simulated annealing structures. $(\overline{SA})_r$ is the restrained minimised mean structure. The number of terms for the various restraints is given in parantheses. E_{L-J} is the Lennard-Jones van der Waals energy.

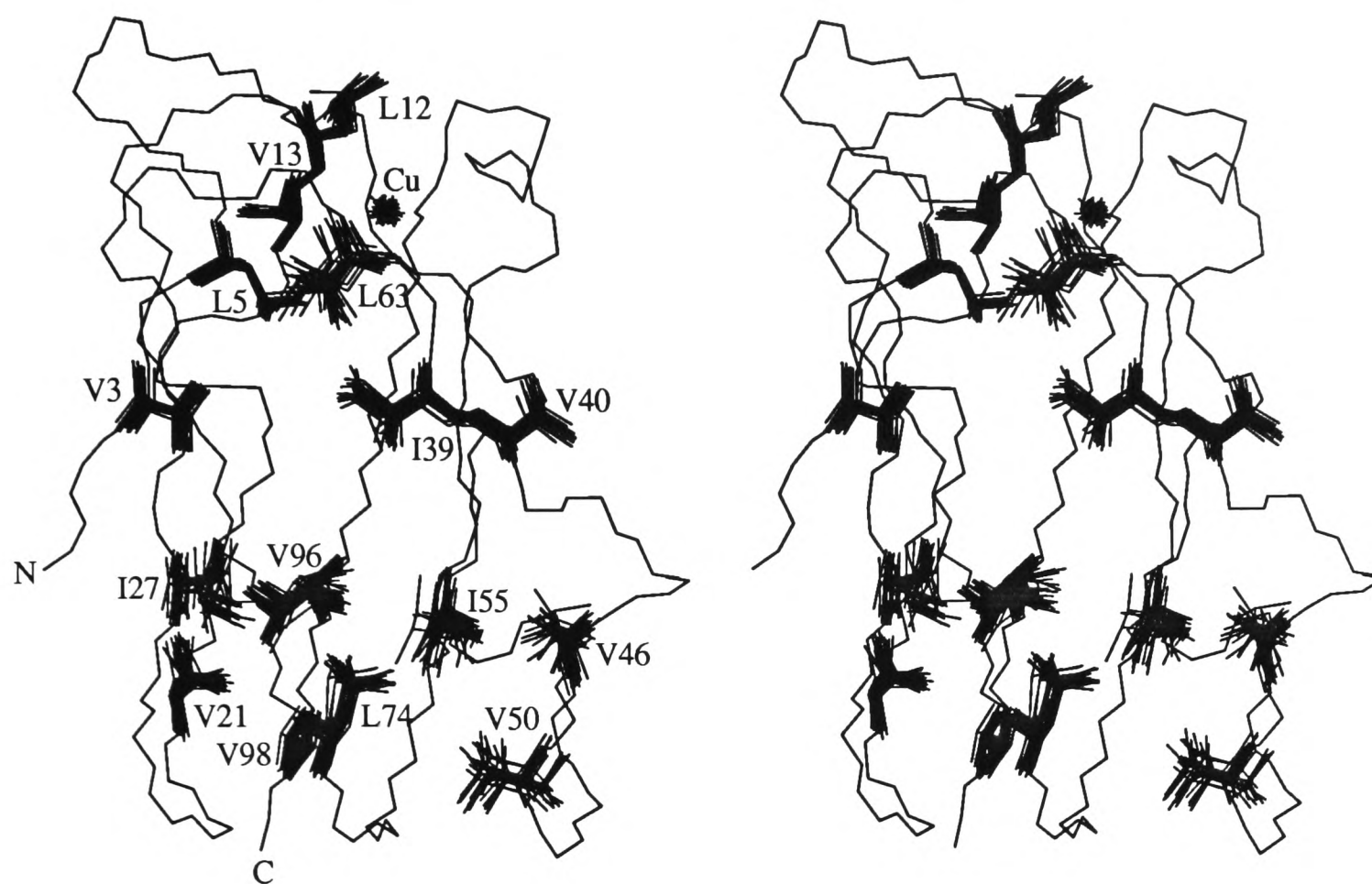


Figure 7.5: Stereo view of the backbone of one of the calculated structures with hydrophobic side chains of the calculated structures. Backbone (N, C $^{\alpha}$ and C) atoms were superposed.

obtained, the side chains are relatively poorly defined. This applies, for example, to Asp 44 and Glu 45. The disorder in such side chains is probably due to paucity of NOEs, a characteristic feature of surface residues.

Parts of the structure which are relatively poorly defined are not necessarily inaccurate. Rather, they are likely to be accurate representations of the conformational space sampled, reflecting conformational flexibility.

Backbone dihedral angles largely fall within the generally allowed regions of the ϕ, ψ plot (Figure 7.9). Apart from glycines, the residues which can have positive ϕ are Asp 8, Ala 33 and Ala 66(64). Of these, only Ala 33 consistently has positive ϕ . All three residues are involved in tight turns or loops (see below).

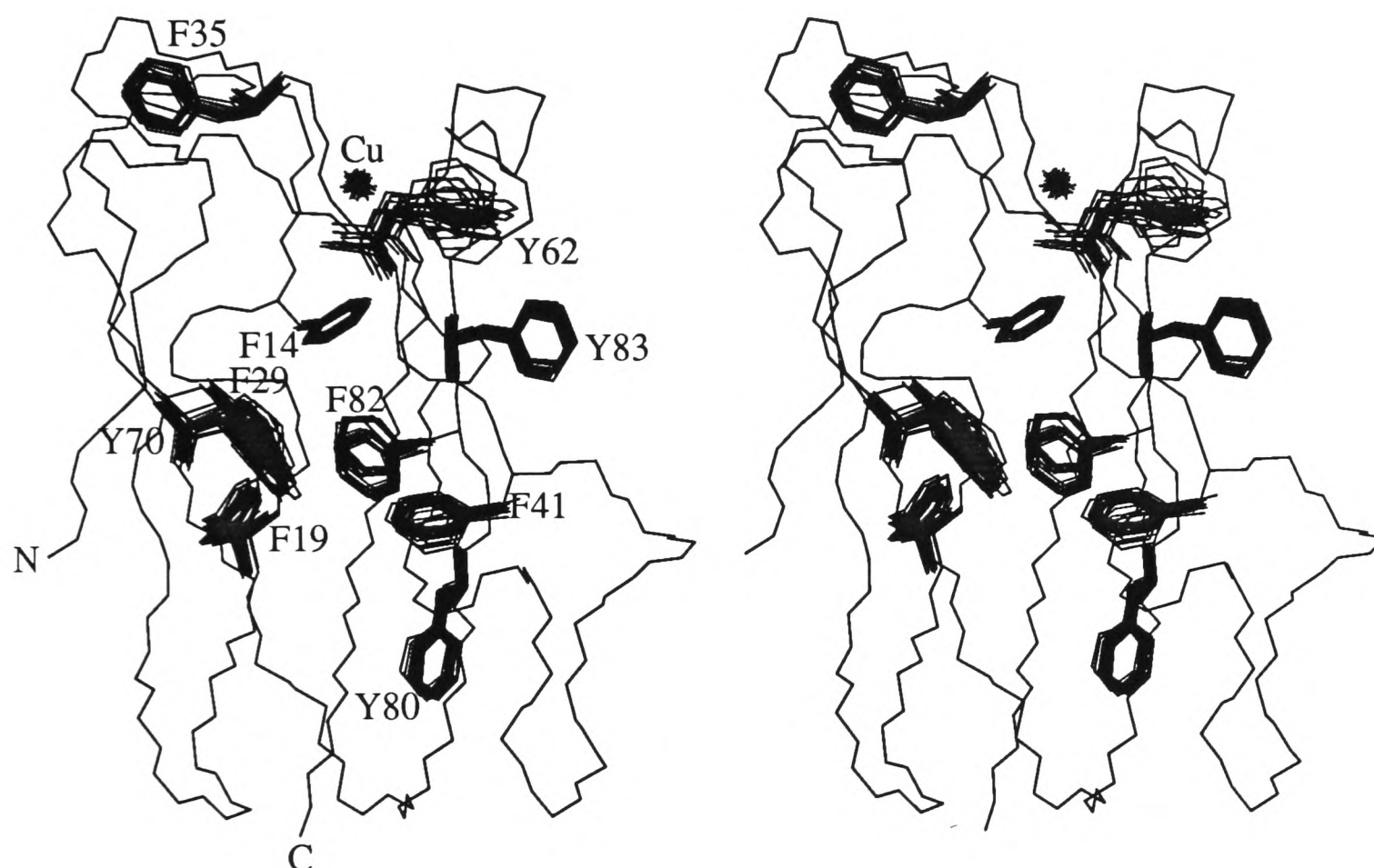


Figure 7.6: Stereo view of the backbone of one of the calculated structures with aromatic side chains of all fifty-eight calculated structures. Note, for example, the contrast in definition of the surface exposed side chains of Tyr 62(60) and Tyr 83(81). The latter side chain has a particularly low B factor in the poplar plastocyanin crystal structure. The high definition of this side chain may have functional significance since Tyr 83(81) has been postulated to be part of the electron transmission route from the acidic binding site to the Cu (see Chapter 1, 1.3 and 1.5 and the section later in this chapter on aromatic residues). Also to note is the location of Phe 35 as the only aromatic residue north of the Cu site, and Phe 14 as the first residue of a *cis*-Pro bend which directs its C^α - C^β bond towards the molecular interior (see section on turns and loops later in this chapter). Backbone (N, C^α and C) atoms were superposed.

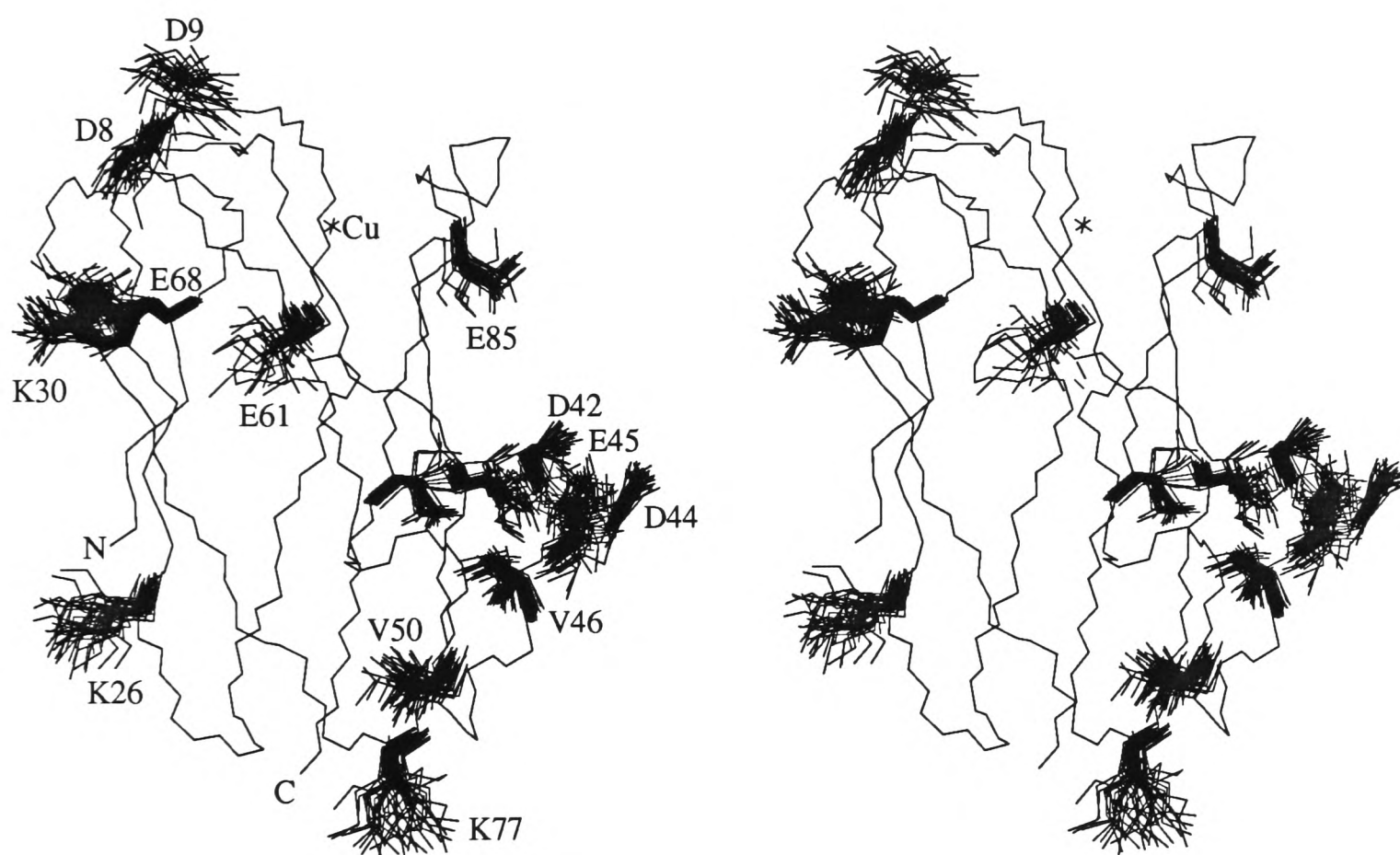


Figure 7.7: Stereo view of the backbone of one of the calculated structures showing some of the residues with surface exposed side chains. Backbone (N, C $^{\alpha}$ and C) atoms were superposed.

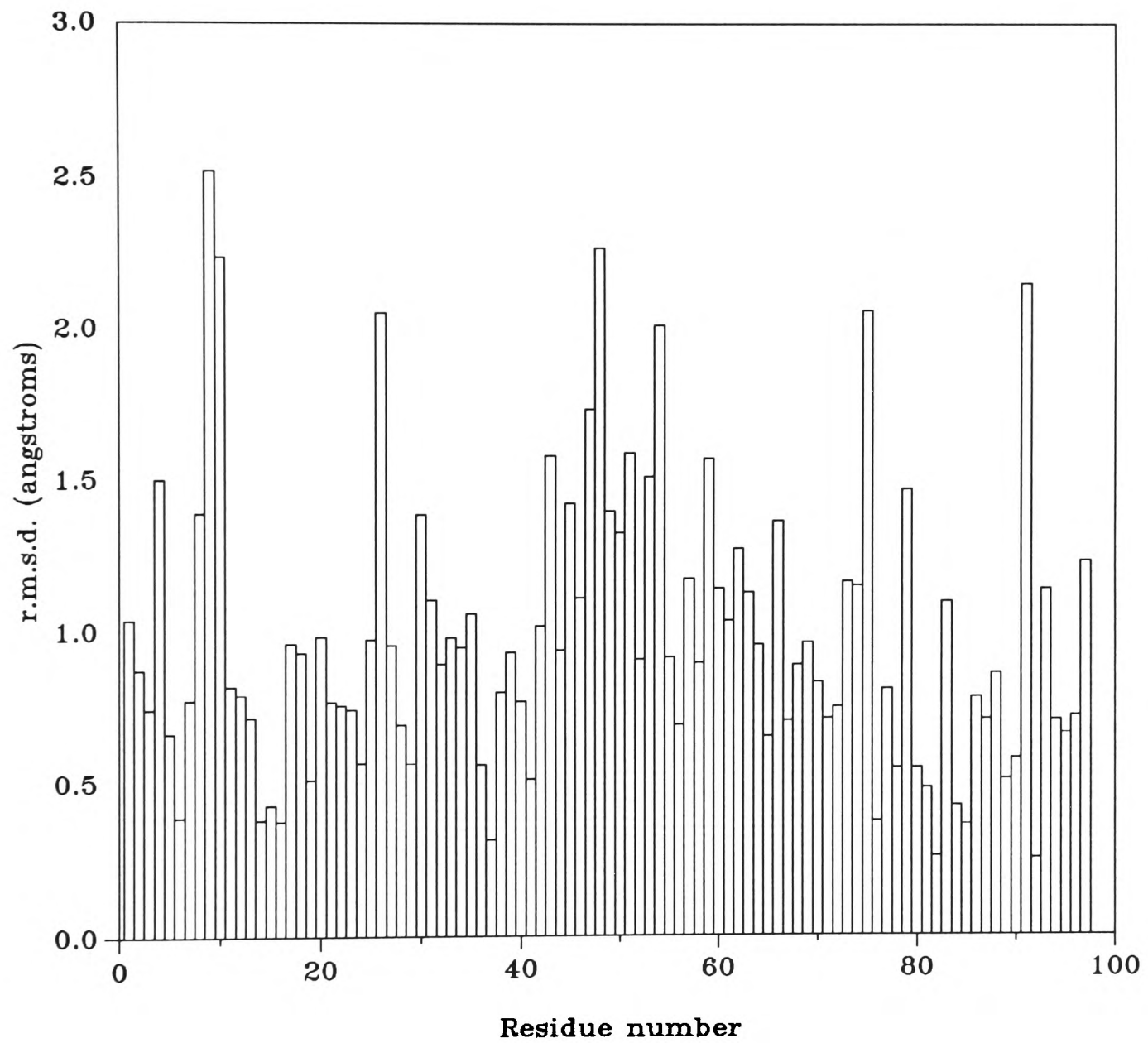


Figure 7.8: Atomic rms distribution of the fifty-eight calculated structures about the mean structure as a function of residue number.

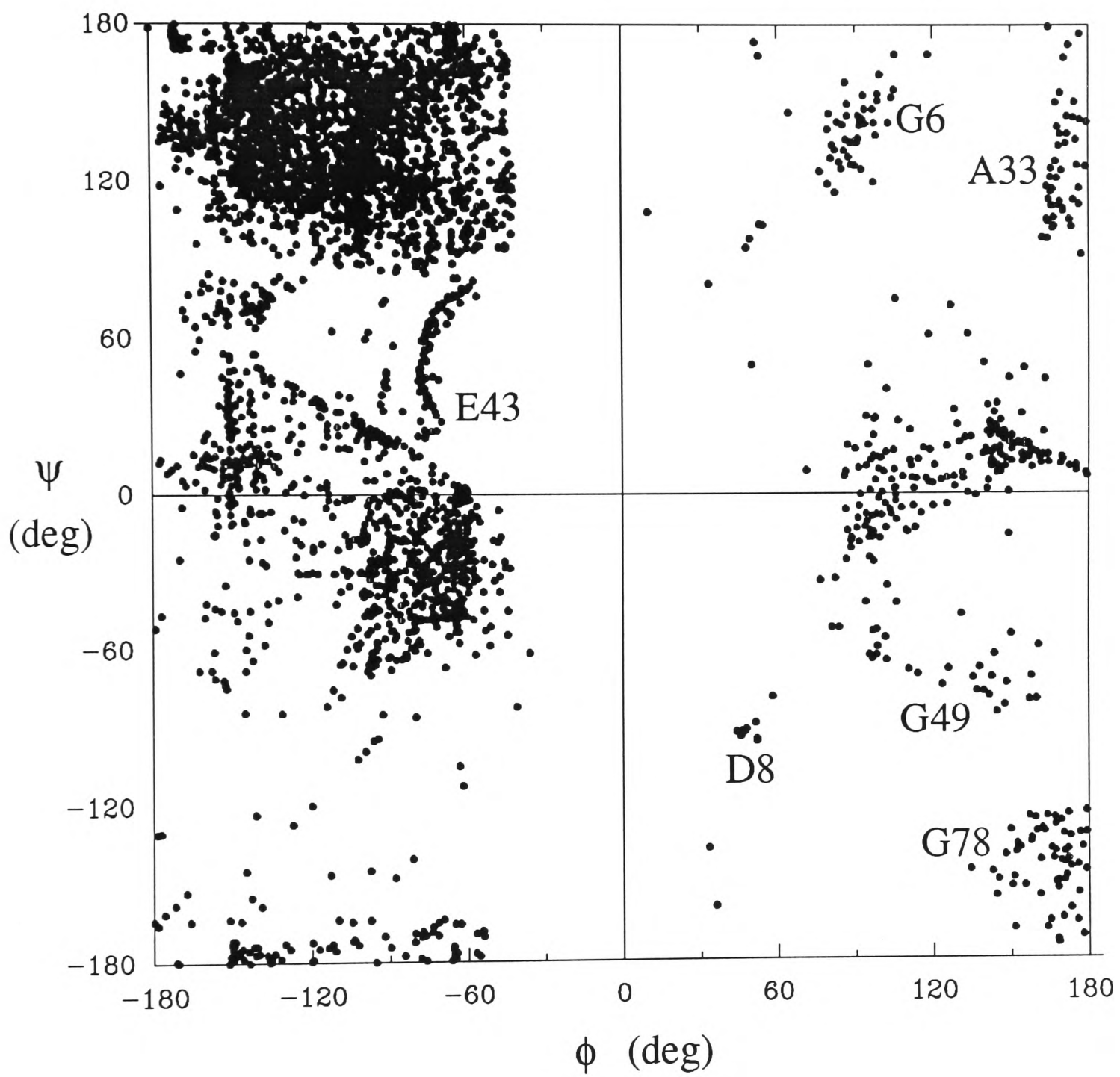


Figure 7.9: Ramachandran ϕ, ψ plot for the fifty-eight parsley plastocyanin structures.

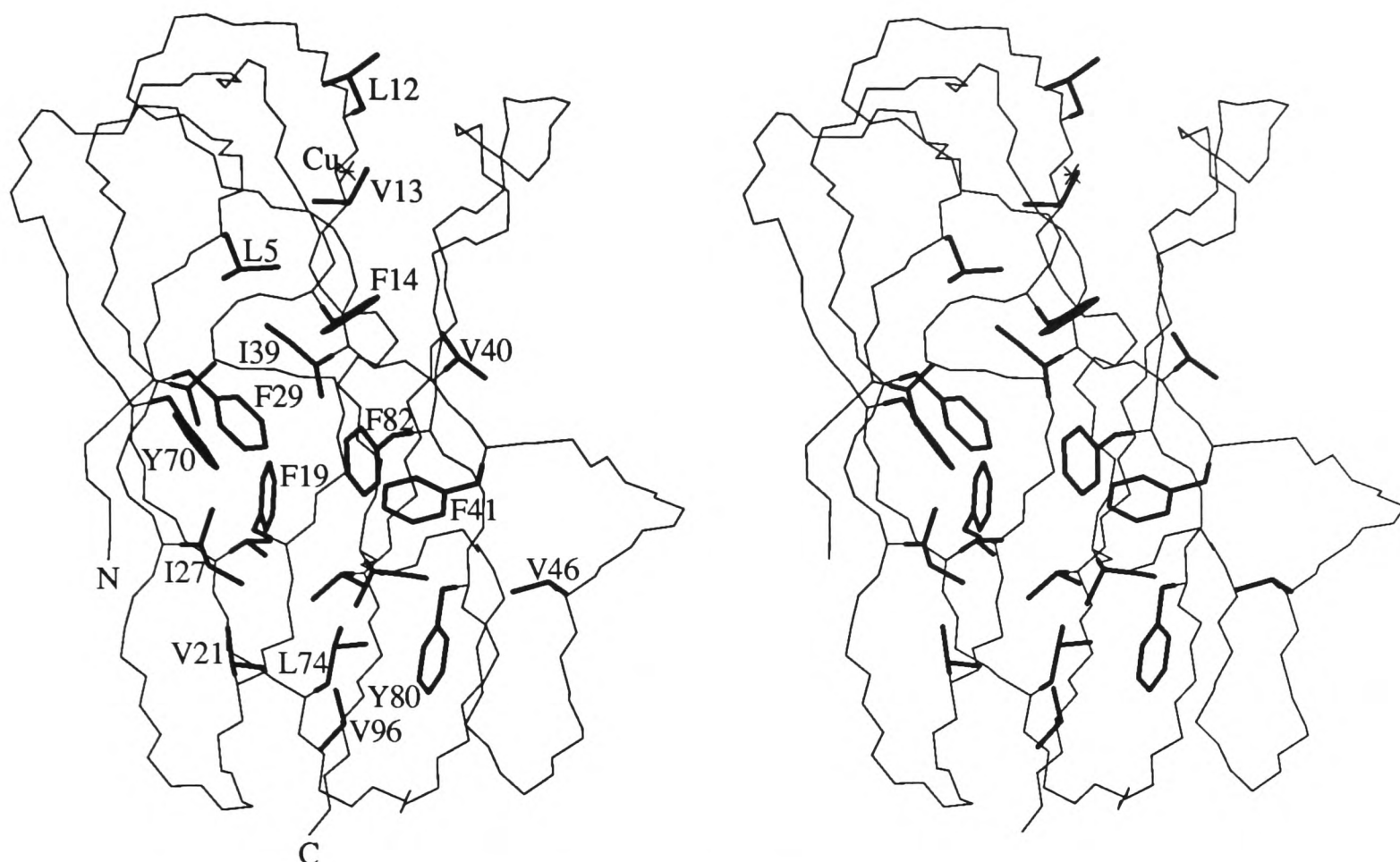


Figure 7.10: Some of the hydrophobic residues in parsley plastocyanin, many of which make up the hydrophobic core.

7.4 Description of the Molecule

7.4.1 Overall Topology

The global conformation of parsley plastocyanin is very similar to those of other structurally characterised plastocyanins [22, 13, 24]. Thus, the molecule can be described as a β -sandwich, the two faces of which are β -sheets made up by a total of eight β -strands. Pairs of strands are connected by loops or tight turns, and the β -sheets are separated by a hydrophobic core (Figure 7.10). This global topology has been described in some detail with respect to poplar plastocyanin in Chapter 1. (See Figures 1.3 and 7.2 for a topological illustration of the plastocyanin fold). The intention here is to highlight similarities and differences which exist among the structurally characterised plastocyanins. The structural details quoted for parsley plastocyanin will be those of the minimised average structure, unless otherwise stated.

7.4.2 β -Strands and β -Sheets

Sheet I is formed by strands 2A (residues 11–14), 1 (residues 1–7), 3 (residues 25–32) and 6 (residues 67(65)–74(72)). Sheet II consists of strands 2B (residues 17–21), 8 (residues 92(90)–99(97)), 7 (residues 78(76)–85(83)) and 4 (residues 36–47) (Figure 7.2).

7.4.3 Hydrogen Bonding Network

Table 7.4 shows the hydrogen bonds involving polypeptide backbone atoms of the plastocyanins from poplar, *E. prolifera* and parsley. Those included for the parsley protein are the ones used as restraints in the structure calculations. The general pattern of hydrogen bonds is retained in the three proteins, indicating that they have similar supersecondary structures. Some of the more easily explainable differences occur in turns, where the lack of a hydrogen bond in parsley plastocyanin is related to the change from a tight turn in the other two proteins to an open loop in parsley plastocyanin. Examples of this are the hydrogen bonds involving residue pairs 7-10 and 47-50. The hydrogen bond involving residues 61 and 58 of poplar plastocyanin is absent due to residue deletions in both other proteins. A series of potentially significant absences in parsley plastocyanin are in the vicinity of the Cu site. This is a highly conserved region, and it might be expected that the general hydrogen bonding pattern would be retained. This is investigated in the discussion of the Cu site below. Three more hydrogen bonds missing from the parsley plastocyanin restraint list are in the single helical region of the protein, which is located in strand 5. Evidence from $^3J_{HN\alpha}$ values that the helix of parsley plastocyanin is distorted was adduced in Chapter 6. The helical segment of the protein is discussed in more detail below.

At the southern end of the molecule, replacement of Pro 23 in poplar plastocyanin by Ala 23 in parsley plastocyanin makes possible the formation of a new 23NH-99(97)COO⁻ hydrogen bond (Figure 7.11). Distance and angle criteria (for which, see Cu site section below) are satisfied for this hydrogen bond, which was not included as a restraint in the calculation of the parsley plastocyanin structure.

D-A	Pop.	<i>E. prolif.</i>	Pars.	D-A	Pop.	<i>E. prolif.</i>	Pars.
3NH-28O	Yes	Yes	Yes	56NH-52O	Yes	Yes	No
4NH-15O	Yes	Yes	Yes	57NH-39O	Yes	Yes	No
5NH-30O	Yes	Yes	Yes	61NH-58O	Yes	No	No
6NH-13O	Yes	Yes	Yes	63NH-37O	Yes	Yes	No
7NH-11O	Yes	Yes	Yes	67NH-31O	Yes	Yes	No
10NH-7O	Yes	Yes	No	68NH-65O	Yes	No	No
15NH-4O	Yes	Yes	Yes	70NH-29O	Yes	Yes	Yes
19NH-95O	Yes	Yes	Yes	72NH-27O	Yes	Yes	Yes
21NH-97O	Yes	Yes	Yes	78NH-98O	Yes	Yes	Yes
24NH-74O	Yes	Yes	No	80NH-96O	Yes	Yes	Yes
25NH-22O	Yes	Yes	Yes	81NH-45O	Yes	Yes	No
27NH-72O	Yes	Yes	Yes	82NH-94O	Yes	Yes	Yes
28NH-1O	Yes	Yes	Yes	83NH-40O	Yes	Yes	Yes
29NH-70O	Yes	Yes	Yes	84NH-92O	Yes	Yes	Yes
30NH-3O	Yes	Yes	Yes	87NH-84O	Yes	Yes	No
31NH-68O	Yes	Yes	No	88NH-84O	Yes	No	No
32NH-5O	Yes	Yes	No	88NH-85O	No	Yes	No
33NH-5O	Yes	Yes	No	90NH-87O	Yes	Yes	No
37NH-63O	Yes	Yes	Yes	91NH-88O	Yes	Yes	No
39NH-57O	No	Yes	No	92NH-87O	Yes	Yes	No
40NH-83O	Yes	Yes	Yes	94NH-82O	Yes	Yes	Yes
42NH-81O	Yes	Yes	Yes	95NH-17O	Yes	Yes	Yes
45NH-42O	Yes	Yes	Yes	96NH-80O	Yes	Yes	Yes
50NH-47O	Yes	Yes	No	97NH-19O	Yes	Yes	Yes
54NH-51O	Yes	Yes	No	98NH-78O	Yes	Yes	Yes
55NH-52O	Yes	No	No	99NH-21O	Yes	Yes	No

Table 7.4: Hydrogen bonds observed in poplar (Pop.), *Enteromorpha prolifera* (*E. prolif.* and parsley (Pars.) plastocyanins. The structures of the first two proteins have been determined by X-ray crystallography. The numbering system is such that sequences are aligned to the poplar sequence (the deletions in *E. prolifera* and parsley plastocyanins are thus included in the count).

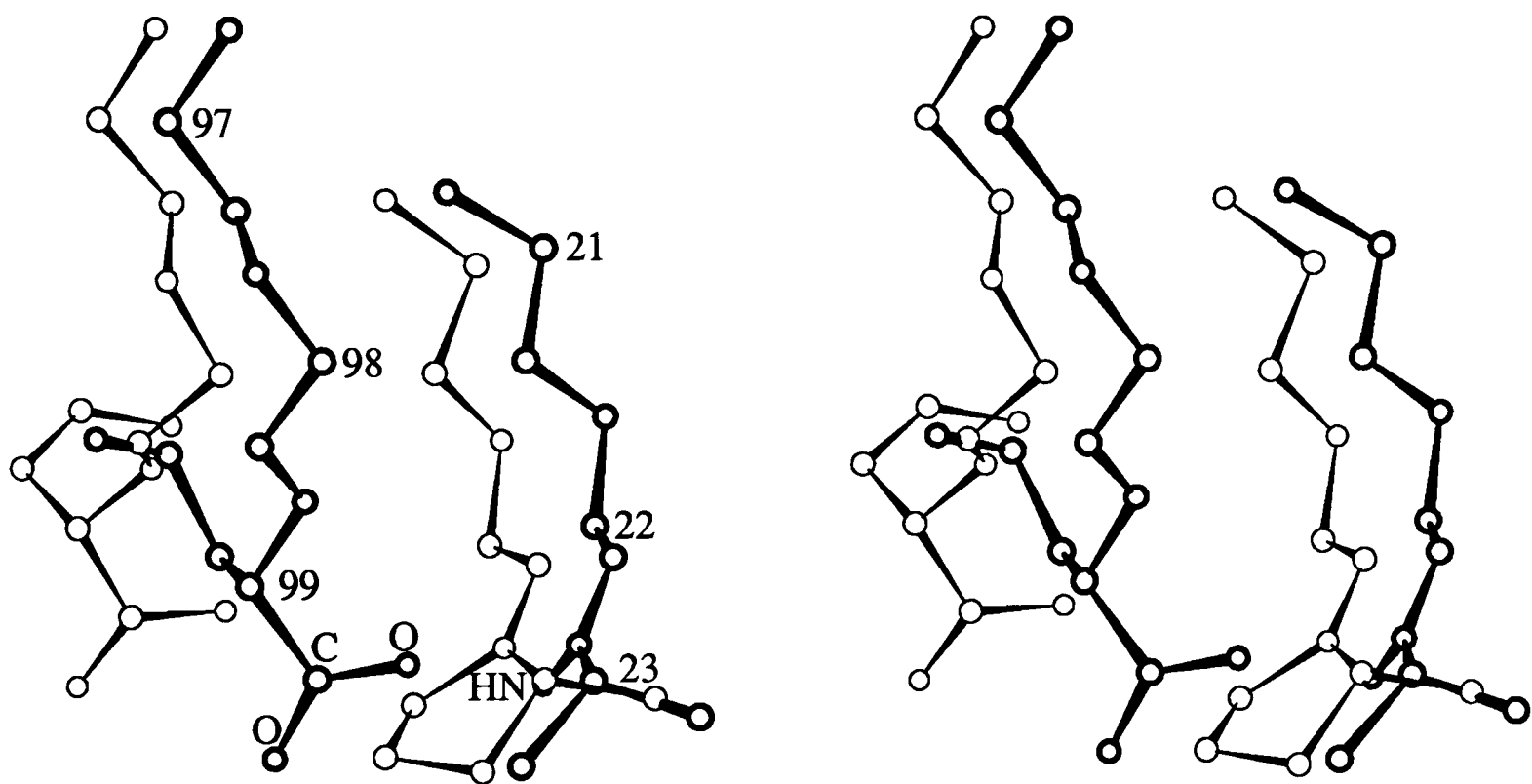


Figure 7.11: Comparison of the C-termini of poplar and parsley (bold) Cu^I plastocyanins. A hydrogen bond in parsley plastocyanin involving Ala 23NH and Asn 99(97)COO⁻ cannot occur in poplar plastocyanin because residue 23 is a proline.

	Parsley	Poplar	Poplar
Cu-ligand bond lengths (Å)	Cu ^I , pH 7.3	Cu ^I , pH 7.0	Cu ^{II} , pH 6.0
Cu-N(H37)	2.24	2.13	2.04
Cu-N(H87)	2.33	2.39	2.10
Cu-S(C84)	2.22	2.17	2.13
Cu-S(M92)	2.86	2.87	2.90
Bond angles at Cu atom (deg)			
N(His 37)–Cu–N(His 87)	108	99	97
N(His 37)–Cu–S(Cys 84)	134	136	132
N(His 37)–Cu–S(Met 92)	94	88	85
N(His 87)–Cu–S(Cys 84)	112	110	123
N(His 87)–Cu–S(Met 92)	101	106	103
S(Cys 84)–Cu–S(Met 92)	102	113	108
Ligand dihedral angles (deg)			
H87 χ_1	-64	-60	-61
H87 χ_2	136	128	138
H37 χ_1	-80	-76	-67
H37 χ_2	-51	-52	-62
C84 χ_1	-152	175	170
M92 χ_1	-128	-164	-169
M92 χ_2	-179	169	165
M92 χ_3	180	168	173

Table 7.5: Dimensions of the Cu site in parsley and poplar Cu^I, and poplar Cu^{II}, plastocyanins. The ligand residue numbering scheme is such that the deleted residues at positions 57 and 58 in parsley plastocyanin are counted.

7.4.4 The Cu Site

The geometries of the Cu ligands of parsley and poplar plastocyanins are not significantly different (Table 7.5; c. f. Table 1.2). For purposes of comparison, Cu site data from the crystal structure of oxidised poplar plastocyanin have been included. One of the more obvious trends is a general lengthening of the Cu-ligand bonds on going from the oxidised to reduced protein. The Cu-S(Met 92) bond length is exceptional in this respect. Superposition of the ligand side chains and Cu atoms of the calculated parsley plastocyanin structures are shown in Figure 7.12.

Several hydrogen bonds in the vicinity of the Cu site of poplar plastocyanin could not be unambiguously identified for parsley plastocyanin, and so were not included

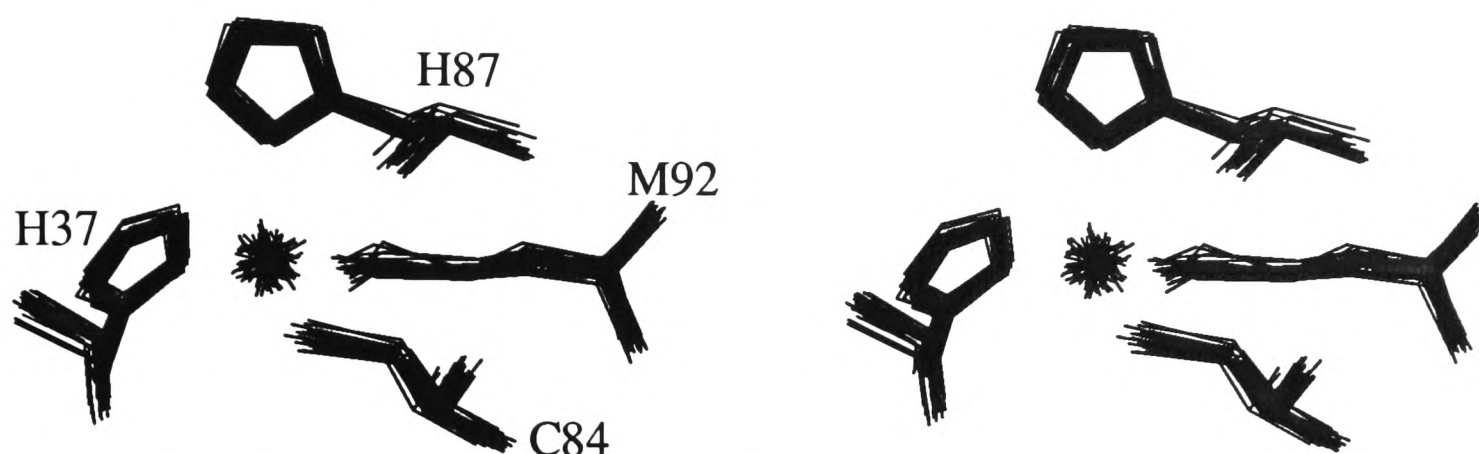


Figure 7.12: Stereo view of the Cu ligands in the calculated structures.

as restraints in the structure calculations (see Table 7.4 for those hydrogen bonds involving backbone atoms only). The structure in the vicinity of the Cu site was examined to determine the likelihood of retention of the poplar plastocyanin hydrogen bonds in parsley plastocyanin (Table 7.6).

Distance criteria used for inclusion of hydrogen bonds in the crystal structure of poplar Cu^{II} plastocyanin [22] were $\text{N}\cdots\text{O}$ distances of between 2.7Å and 3.2Å, and $\text{N}\cdots\text{S}$ distances of less than 3.4Å. Of the three angle criteria used, one has been measured here: θ ($\text{C}^\alpha\text{-N}\cdots\text{O}$) was required to be less than 93° . The authors [22] do not seem to have adhered strictly to this requirement, and perhaps included the hydrogen bond if the other angle criteria were satisfied. Some of the plastocyanin hydrogen bonds do not satisfy another published hydrogen bond angle requirement, that θ ($\text{N-H}\cdots\text{O}$) be greater than 135° [25]. The data presented in Table 7.6 indicate that most of the hydrogen bonds around the Cu site of poplar plastocyanin are retained in parsley plastocyanin. The retention of two of the hydrogen bonds involving invariant Asn 38 seems less certain. This is slightly surprising given the structural importance attached to this residue in poplar plastocyanin [22] (Chapter 1, 1.3). It is possible that the hydrogen bond involving residue 85(83) is present in parsley plastocyanin, its formation pulling the surface exposed Glu 85(83) side chain closer to the polypeptide backbone. In the second case, the peptide oxygen atom of Glu 61(59) is on the wrong side of a short section of strand (61(59)–64(62)) to be involved in a hydrogen bond with the N^δ protons of Asn 38 (Figure 7.13). It seems that this possible disruption in the hydrogen bonding network is a result of the deletion of residues at positions

Bond	Parsley Cu ^I			Poplar Cu ^I			Poplar Cu ^{II} θ_2
	N...O (Å)	θ_1	θ_2	N...O (Å)	θ_1	θ_2	
84NH-92O	2.28	129	166	2.14	122	177	122
87NH-84O	3.24	100	96	2.57	103	136	102
88NH-84O	2.59	90	91	2.58	94	96	98
90NH-87O	2.13	108	143	2.44	101	120	93
91NH-88O	3.22	85	100	1.99	95	152	99
92NH-87O	2.66	111	167	2.26	105	157	105
37N ^ε H-33O	3.37	-	159	1.71	-	161	-
38NH-84S	2.33	129	152	2.49	129	169	127
38N ^δ H-61O	4.48*	-	-	2.85*	-	-	-
85O ^γ H-38O ^δ	4.60*	-	-	3.67	-	-	-

Table 7.6: Details of the hydrogen bonds in the vicinity of the Cu site in poplar Cu^I and Cu^{II} plastocyanin crystal structures. Distances and both angles can only be determined for poplar Cu^I plastocyanin, onto the heavy atoms of which protons have been built using the HBUILD feature of X-PLOR. Corresponding data are presented for parsley plastocyanin, for which only one (38NH-84S) of these hydrogen bonds was included as a distance restraint. The angles θ_1 and θ_2 are θ (C^α-N...O) and θ (N-H...O). Starred distances are the shorter in cases where the donor group has two protons. The numbering system is such that sequences are aligned to the poplar sequence (the deletions in parsley plastocyanin are thus included in the residue count). Note that residue 85 in poplar plastocyanin is a serine, that in parsley plastocyanin is a glutamate.

57 and 58 in parsley plastocyanin relative to poplar plastocyanin.

7.4.5 Loops and Tight Turns

There are eleven turns and loops in parsley plastocyanin, one fewer than in the poplar protein. This disparity arises from the absence of residues at positions 57 and 58 in the parsley plastocyanin sequence. The possible further significance of these deletions will be addressed later. The turns and loops making up the northern end of parsley plastocyanin are at residues 7–11, 32–35, 65(63)–68(66), 84(82)–87(85) and 88(86)–91(89). The southern end of the molecule is defined by turns at residues 22–25, 47–50 and 73(71)–74(74).

Preliminary analysis of turns, based on NOE and 3J data, was made in the previous chapter, wherein the main turn types were also introduced. More rigorous analysis of the turns in parsley plastocyanin is now possible. The values of the ϕ and ψ backbone dihedral angles for the two central residues in each of the turns of the minimised average structure of parsley plastocyanin, along with typical values for types I and II, are presented in Table 7.7. Angle Vi in Table 7.7 is the virtual bond dihedral angle defined by the four α carbons of the turn residues. For purposes of comparison, Table 7.8 shows conformational data for the turns in reduced poplar plastocyanin (crystal structure at pH 7.0 [13]). A point to note is that in both proteins, all the turns which have been classified as type II possess a glycine in the third position. This is characteristic of type II turns. In types II' and I' also, the dihedral angles are such that for one or both of the central positions glycine is preferred.

Apart from turn types I and II and their mirror images, five other types of turn have been identified. Type III has repeating ϕ, ψ values of $-60^\circ, -30^\circ$, and is indistinguishable from a 3_{10} helix. The miscellaneous type IV category includes any example with two of the dihedral angles different by more than more than 40° from ideal values for any of the other types. A type VI turn is described below.

The first piece of nonrepetitive structure in the parsley plastocyanin sequence (residues 7–10) (Table 7.7) cannot be classified as a regular turn, although the configuration of the central peptide linkage is closer to that of a type I turn than a type II turn. This fragment is probably best described as an open turn or loop which

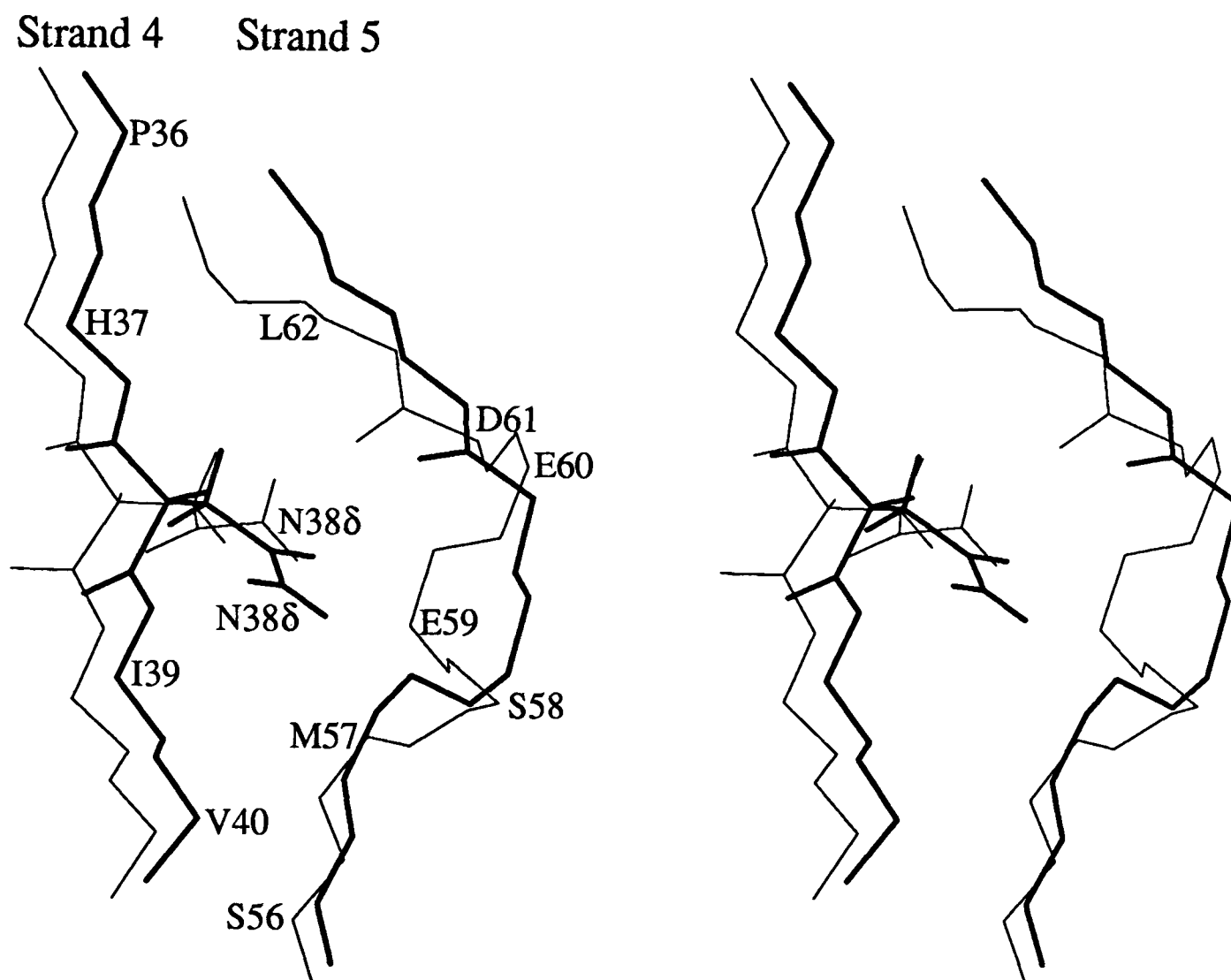


Figure 7.13: Comparison of poplar (pH 7.0) and parsley (pH 7.3; in bold) Cu^I plastocyanins in the region of the poplar plastocyanin $38\text{N}^{\delta}\text{H}\cdots 61\text{O}$ hydrogen bond. This portion of the protein contains the type I turn at residues 58–61 of poplar plastocyanin. This turn is eliminated from parsley plastocyanin due to the deletion of residues at positions 57 and 58. The C^{α} atoms labelled in strand 4 are those of parsley plastocyanin (same amino acids in the poplar protein). In strand 5, the C^{α} atoms of poplar plastocyanin are labelled in order to indicate the turn. The peptide carbonyl group of Glu 61(59) of parsley plastocyanin faces into the page, away from the NH_2 group of Asn 38.

Residues	Dihedral angles (deg)					Distances (Å)		Type
	ϕ_2	ψ_2	ϕ_3	ψ_3	Vi	O(1)–N(4)	C $^\alpha$ (1)–C $^\alpha$ (4)	
7–10	-45	-42	-172	79	-21	4.2	5.6	Loop
14–17	-118	113	-65	179	26	3.1	4.8	VIb
22–25	-77	58	171	11	35	3.3	5.5	II
32–35	177	110	143	19	64	6.9	8.2	Loop
42–45	-74	39	-154	6	50	3.3	4.7	I
47–50	-91	-165	88	-59	68	4.9	7.5	Loop
65(63)–68(66)	-75	110	98	-1	5	3.2	5.1	II
73(71)–76(74)	-105	170	-150	25	-172	7.1	9.9	Loop
84(82)–87(85)	-54	-49	-62	-58	45	3.5	5.4	I
88(86)–91(89)	-66	-28	-80	-52	46	3.5	5.2	I
Type I	-60	-30	-90	0	45	3–3.5	4.5–5.5	
Type II	-60	120	90	0	0	3–3.5	5.0–6.0	

Table 7.7: Conformational data for the turns and loops of parsley plastocyanin. The virtual bond angle Vi is that defined by the four α carbons of the turn residues. Included at the bottom of the table are typical values for type I and type II turns. Residue numbers in brackets are produced by omitting the deletions from the residue count.

Residues	Dihedral angles (deg)					Distances (Å)		Type
	ϕ_2	ψ_2	ϕ_3	ψ_3	Vi	O(1)–N(4)	C $^\alpha$ (1)–C $^\alpha$ (4)	
7–10	-66	-12	-102	13	-9	2.9	5.1	I
14–17	-127	113	-71	165	17	3.5	4.7	VIb
22–25	-60	138	70	5	13	3.1	5.6	II
32–35	-156	147	75	15	29	6.3	7.3	Loop
42–45	-56	-31	-91	6	48	2.8	5.1	I
47–50	-58	137	83	-6	20	3.2	5.9	II
65–68	-56	134	81	-4	18	3.2	5.8	II
73–74	-106	142	-116	-41	173	6.9	9.8	Loop
84–87	-63	-50	-67	-23	42	3.3	5.4	I
88–91	-68	-30	-70	-9	61	2.9	5.3	I
Type I	-60	-30	-90	0	45	3–3.5	4.5–5.5	
Type II	-60	120	90	0	0	3–3.5	5.0–6.0	

Table 7.8: Conformational data for the turns and loops of poplar plastocyanin, taken from that crystal structure which most closely reproduces the conditions used for the parsley plastocyanin NMR experiments: reduced poplar plastocyanin at pH 7.0.

includes Gly 11 in addition to residues 7–10. This contrasts with the occurrence of a regular type I turn in the poplar plastocyanin crystal structure (Table 7.8).

The unit of nonrepetitive structure at residues 22–25 approximates to a type II turn (Table 7.7). The large discrepancy between the typical value of ϕ_3 and the one observed may be reconciled by the fact that in a type II turn, the distances $d_{NN}(3,4)$ and $d_{\alpha N}(2,4)$ change little if ϕ_3 is rotated from $+90^\circ$ through 0° to -90° .

The fragment at positions 32–35 is immediately identified as an open loop rather than a tight turn by the fact that the α -carbons of the first and fourth residues are more than $7\text{\AA}$ apart (Tables 7.7 and 7.8). The 42–45 turn approximates to type I, with an atypical value for ϕ_3 . The two remaining turns at the southern end of parsley plastocyanin, at residues 47–50 and 73(71)–76(74), are both irregular. The same consideration applies to the latter as that mentioned above for residues 32–35, the loop at residues 73(71)–76(74) having an even more open structure.

In contrast to the turns so far described, those at residues 65(63)–68(66), 84(82)–87(85) and 88(86)–91(89) can be confidently classified as one of the regular turn types. The first of these is a type II turn, whereas the remaining two are type I turns (Table 7.7).

An interesting feature of all currently available plastocyanin structures, including that of parsley plastocyanin, is the mid-strand turn at residues 14–17 (see also Chapter 1, 1.3). This is a type VI turn or *cis*-Pro bend [26]. Turns involving *cis*-peptide bonds have been subdivided into types VIa and VIb: the former have a 1→4 hydrogen bond, the latter do not. Unambiguous distinction between these possibilities in the case of parsley plastocyanin cannot be made with the available data. However, the ϕ_2 angle of -118° identifies the parsley plastocyanin turn as type VIb. (The ϕ_2 angle is $> -60^\circ$ for type VIa turns and $< -90^\circ$ for type VIb turns [27]). A similar turn is also present in strand 2 of the related blue Cu protein azurin, indicating that such a turn in this position probably has structural and/or functional significance. In poplar plastocyanin, one effect of the turn is to direct the C^α - C^β bond of Phe 14 towards the molecular interior [22]. In this position it is close to the side chain of Met 92(90), one of the Cu ligands. Phe 14 has a very similar location in parsley plastocyanin (Figure 7.17). An analogous juxtaposition of side chains is found in azurin. In both of these

blue Cu proteins, the phenylalanine involved is an invariant residue. It seems that the phenylalanine is an essential part of the Cu site in plastocyanin and azurin.

In common with the other structurally characterised plastocyanins, the parsley protein possesses a second *cis*-peptide linkage at invariant Pro 36. This residue is adjacent to His 37, a Cu ligating residue. This *cis*-peptide group is unusual in that Pro 36 is not part of a turn. Most *cis*-peptide bonds are found at the third position of reverse turns.

A comparison of the turns and loops of four plastocyanins is presented in Table 7.9. In general, the turns in the poplar crystal structure are more regular or classical than those in parsley plastocyanin. Cases where well defined, regular turns are present in the crystal structure, but irregular turns or loops are present in parsley plastocyanin, may reflect local conformational flexibility in solution. Examples are the 7–10 region and part of the Cu-binding loop at residues 84(82)–87(85). The former is described as a loop in all three NMR studies but as a type I turn in the crystal structure. The latter is an irregular type I turn in parsley plastocyanin, a loop in French bean [28] and *S. obliquus* plastocyanins [29], and a regular type I turn in the poplar plastocyanin crystal structure. No obvious reasons for these variations can be found from differences in sequence. Indeed, the 84(82)–87(85) fragment is one of the most highly conserved in plastocyanin. The parsley protein differs from most of its higher plant counterparts in having a glutamate in position 85(83), a characteristic it shares with barley and algal plastocyanins (see Section 1.2 and Table 1.1). The general trend of turn regularity in crystal structures and turn irregularity in NMR-derived structures is reversed in one case: NMR evidence indicates that residues 32–35 in French bean plastocyanin form a type II or half turn [28], whereas the corresponding section in the poplar plastocyanin crystal structure is an open loop.

The most obvious, and probably most significant feature in Table 7.9, is the elimination of a turn from parsley and *S. obliquus* plastocyanins due to the deletion of residues at positions 57 and 58 (Figure 7.14). Both poplar and French bean plastocyanins contain regular turns at residues 58–61: a type I in the former protein, a type II in the latter. Extra significance is added by the fact that this difference in backbone conformation occurs in the acidic patch, which consists mainly of residues

Position	Parsley		<i>S. obliquus</i>		French bean		Poplar	
	Seq.	Type	Seq.	Type	Seq.	Type	Seq.	Type
7–10	SDDG	L	ADSG	L	SGDG	L	ADDG	I
14–17	FSPS	VIb	FEPA	VIb	FVPS	VIb	FVPS	VIb
22–25	AAGE	II	KAGD	II/HT	PSGE	II	SPGE	II
32–35	NAGF	L	NAGF	L	NAGF	II/HT	NAGF	L
42–45	DEDE	I	DEDA	I	DEDE	I	DEDS	I
47–50	PAGV	L	PAGV	II/HT	PAGV	II/HT	PSGV	II
58–61	**QE	No turn	**DD	No turn	PEEE	II	SEED	I
65–68	GAGE	II	APGE	II/HT	APGE	II	AKGE	II
73–76	TLTE	L	KFDT	L	TLDT	L	ALSN	L
84–87	CEPH	I	CEPH	L	CSPH	L	CSPH	I
88–91	AGAG	I	QGAG	L	QGAG	I	QGAG	I

Table 7.9: Comparison of turns and loops in various plastocyanins. Half turn is abbreviated as HT. Parsley, *Scenedesmus obliquus* and French bean plastocyanins have all been characterised by NMR in solution. At present, no three dimensional structures have been published for *S. obliquus* and French bean plastocyanins. The turns in these proteins have been classified on the basis of NOE and 3J data only. The data for poplar plastocyanin are from X-ray crystallographic analysis.

42–45 and 59(57)–61(59). This patch is one of two postulated recognition sites on the surface of plastocyanin. The possible effects on the functional properties of parsley and similarly bereft plastocyanins, will be considered later in this chapter.

7.4.6 β -Bulge

Parsley plastocyanin contains a β -bulge [30] involving residues Gly 6 and Ser 7 of strand 1 and residue Val 13 of strand 2A (Figure 7.15). The existence of such a unit of structure was evidenced by the NOE and hydrogen exchange data described in the previous chapter. The β -bulge manifested in the calculated structures appears to be of the G1, rather than the classic, type. G1, the second most common type of bulge, is so called because most examples have a glycine in position 1 (position 1 of the bulge in parsley plastocyanin is occupied by Gly 6). The ϕ, ψ angles of position 1 of the G1 bulge are centred around $85^\circ, 0^\circ$, which is favourable only for glycine. The corresponding values for Gly 6 in parsley plastocyanin are $88^\circ, 130^\circ$. Position 2 of G1 bulges is within the usual range for β structure, but is centred around ϕ, ψ

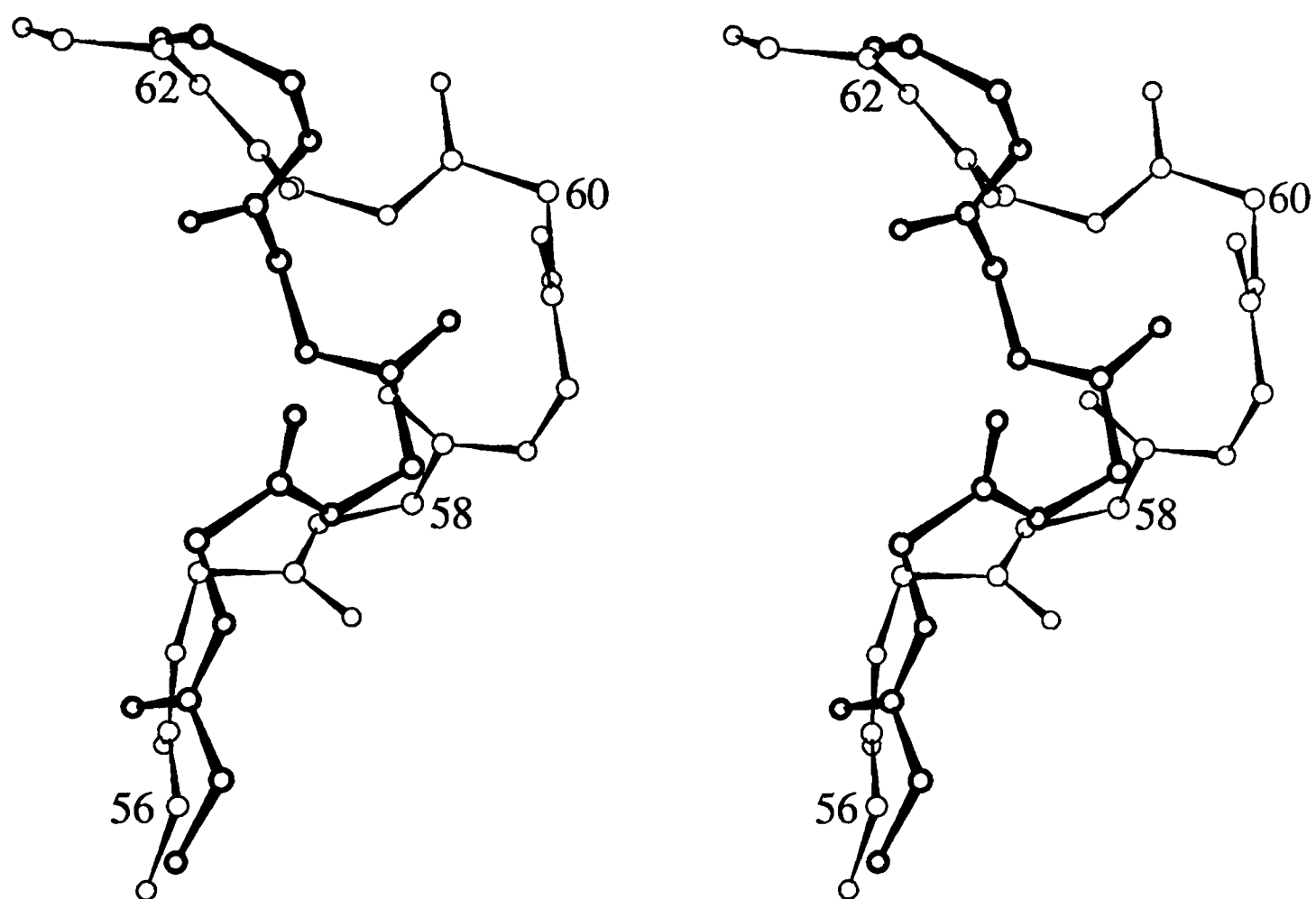


Figure 7.14: Stereo view comparing the polypeptide backbones of parsley (pH 7.3; thicker bonds) and poplar (pH 7.0) Cu^I plastocyanins near residues 57–61. The labelled positions are C^α atoms of poplar plastocyanin. C^α atoms of residues 1–56 and 59–97 were superposed.

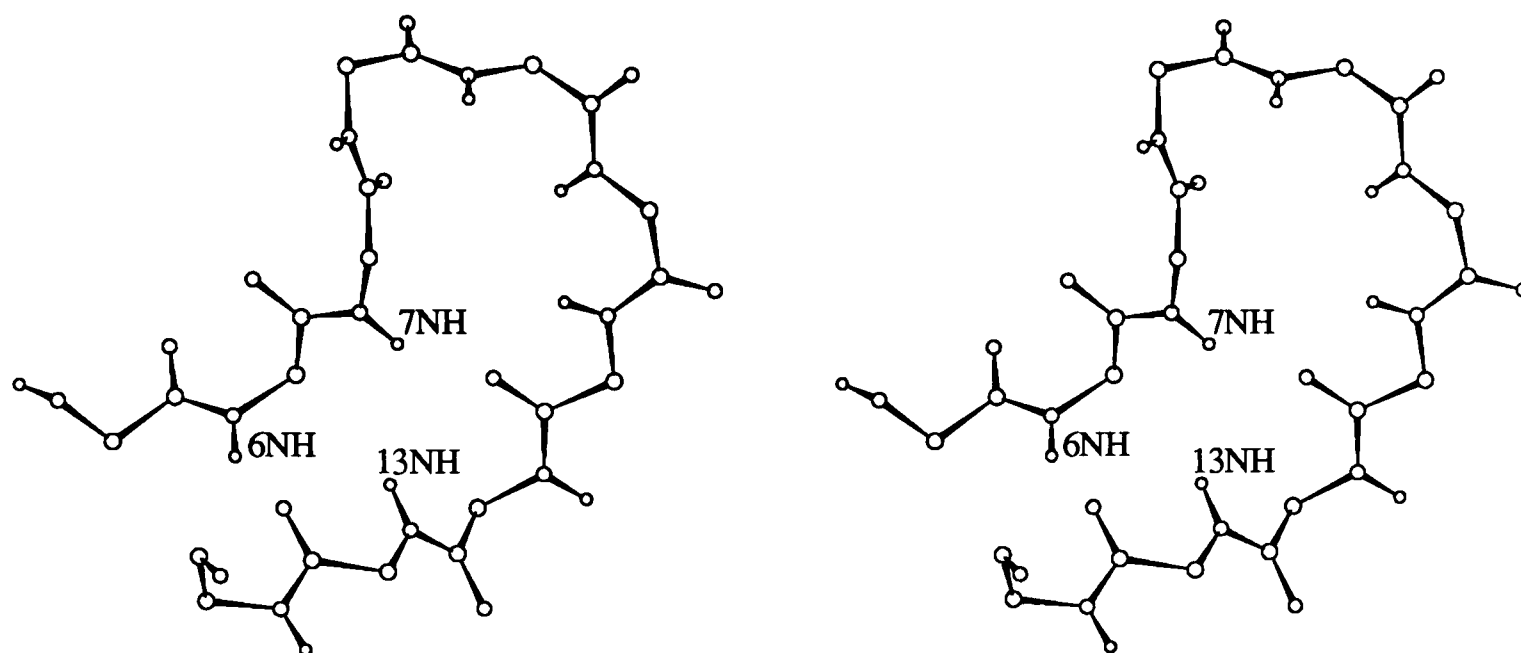


Figure 7.15: Stereo view of the β -bulge of parsley plastocyanin involving residues G6, S7 and V13.

values of $-90^\circ, 150^\circ$. Ser 7 has ϕ, ψ values of $-68^\circ, 155^\circ$. As is characteristic of β -bulges, strand 1 is twisted so that the amide protons of Gly 6 and Ser 7 are on the same side of the β -sheet. The G1 bulge in parsley plastocyanin is unusual in that it occurs in combination with an open loop at residues 7–11: most G1 bulges occur in combination with a type II tight turn [30]. NOE data evidence the existence of a β -bulge in a similar location in the plastocyanins from French bean [28] and *S. obliquus* [29]. Poplar plastocyanin does not, however, possess such a unit of structure (at least in the crystalline phase).

7.4.7 Helical Fragment

The only helical portion of plastocyanin occurs in strand 5, that which causes disruption of the continuous hydrogen bonding network that would be required for plastocyanin to be described as a β -barrel (see Chapter 1). In most higher plant plastocyanins, this strand also contains the turn at residues 58–61 which is eliminated in

parsley and algal plastocyanins. The helical segment of poplar plastocyanin extends from residue 51 to residue 56. NOE data discussed in Chapter 6 indicated that the helix in parsley plastocyanin starts at residue 52. Coupling constant data pointed to a fraying or distortion towards the C-terminal end of the helix. These features can now be examined in greater detail (Tables 7.10 and 7.11). The helices of parsley and poplar plastocyanins are compared in Figure 7.16.

The data presented in Tables 7.10 and 7.11 indicate that the helical section of parsley plastocyanin is indeed less regular than its poplar plastocyanin counterpart. In neither protein, however, do the ϕ, ψ angles conform to those of right-handed α -helix ($-60^\circ, -60^\circ$). The angles are closer to those of the generally less favourable 3_{10} -helix, which has backbone conformational angles of approximately $-60^\circ, -30^\circ$. The $n + 3$ hydrogen bonding pattern in poplar plastocyanin is also that of a 3_{10} -helix, α -helix being characterised by $n + 4$ hydrogen bonds. It is debatable whether there is sufficient cause to regard Asn 51 of parsley plastocyanin as being in a non-helical conformation. Its ψ angle is certainly not that found in regular helix, but then neither is that of Asp 51 of poplar plastocyanin.

The hydrogen bond data (Table 7.11) support the contention that the helix of parsley plastocyanin is loose or frayed, although it is probable that inclusion of the expected hydrogen bond restraints would have pulled the conformation into one more closely approaching regular helix. However, no reason to include these restraints was obtained from slow exchange of amide protons.

7.4.8 Hydrogen Bonding Involving Residues 41, 52 and 56

The hydrogen bond between 56 O γ H and 52 O was used as a restraint in the calculation of the parsley plastocyanin structure, as mentioned in the caption to Table 7.11. Closer inspection of the vicinity of the helix revealed that there is a second potential hydrogen bond involving the hydroxyl group of Ser 56. In this case, Ser 56 O γ is the acceptor and Phe 41 NH is the donor. The existence of this interaction is evidenced by the observation that Phe 41 backbone amide proton is slowly exchanging. Distance and angle criteria are satisfied for this hydrogen bond (Table 7.12).

It is postulated that both 41NH-56O γ and 56O γ H-52O hydrogen bonds occur in

Residue		Parsley Cu ^I , pH 7.3		Poplar Cu ^I , pH 7.0	
Parsley	Poplar	ϕ (deg)	ψ (deg)	ϕ (deg)	ψ (deg)
Asn 51	Asp 51	-55	158	-83	109
Ala 52	Ala 52	-96	-6	-64	-24
Ser 53	Glu 53	-89	-10	-69	-28
Lys 54	Lys 54	-96	-25	-72	-25
Ile 55	Ile 55	-103	20	-119	-1
Ser 56	Ser 56	-115	-180	-120	-178

Table 7.10: Backbone dihedral angles in the helical section of parsley and poplar Cu^I plastocyanins.

Hydrogen Bond	Distance (Å)	
	Parsley Cu ^I , pH 7.3	Poplar Cu ^I , pH 7.0
54NH-51O	3.05	2.41
55NH-52O	3.13	2.22
56NH-52O	3.50	2.43
56O ^γ H-52O	1.85	2.27

Table 7.11: Hydrogen bond distances in the helical section of parsley and poplar Cu^I plastocyanins. Only 56O^γH-52O was entered as a restraint in the calculation of the structure of parsley plastocyanin. The existence of this hydrogen bond was evidenced by the appearance of a cross peak corresponding to Ser 56 O^γH in HOHAHA spectra (see the discussion of the helical fragment in Chapter 6). Cross peaks for Ser O^γ protons are rarely observed. None of the other putative hydrogen bonds were evidenced by slowly exchanging amide protons.

Bond	Parsley Cu ^I			Poplar Cu ^I		
	N...O/O...O (Å)	θ_1	θ_2	N...O (Å)	θ_1	θ_2
41NH-56O ^γ	2.70	124	121	3.20	109	162
56O ^γ H-52O	2.82	-	107	2.73	-	152

Table 7.12: Comparison of angle and distance data for two hydrogen bonds involving Ser 56 O^γ in parsley and poplar plastocyanins. The angles θ_1 and θ_2 are θ (C^α-N...O) and θ (N-H...O).

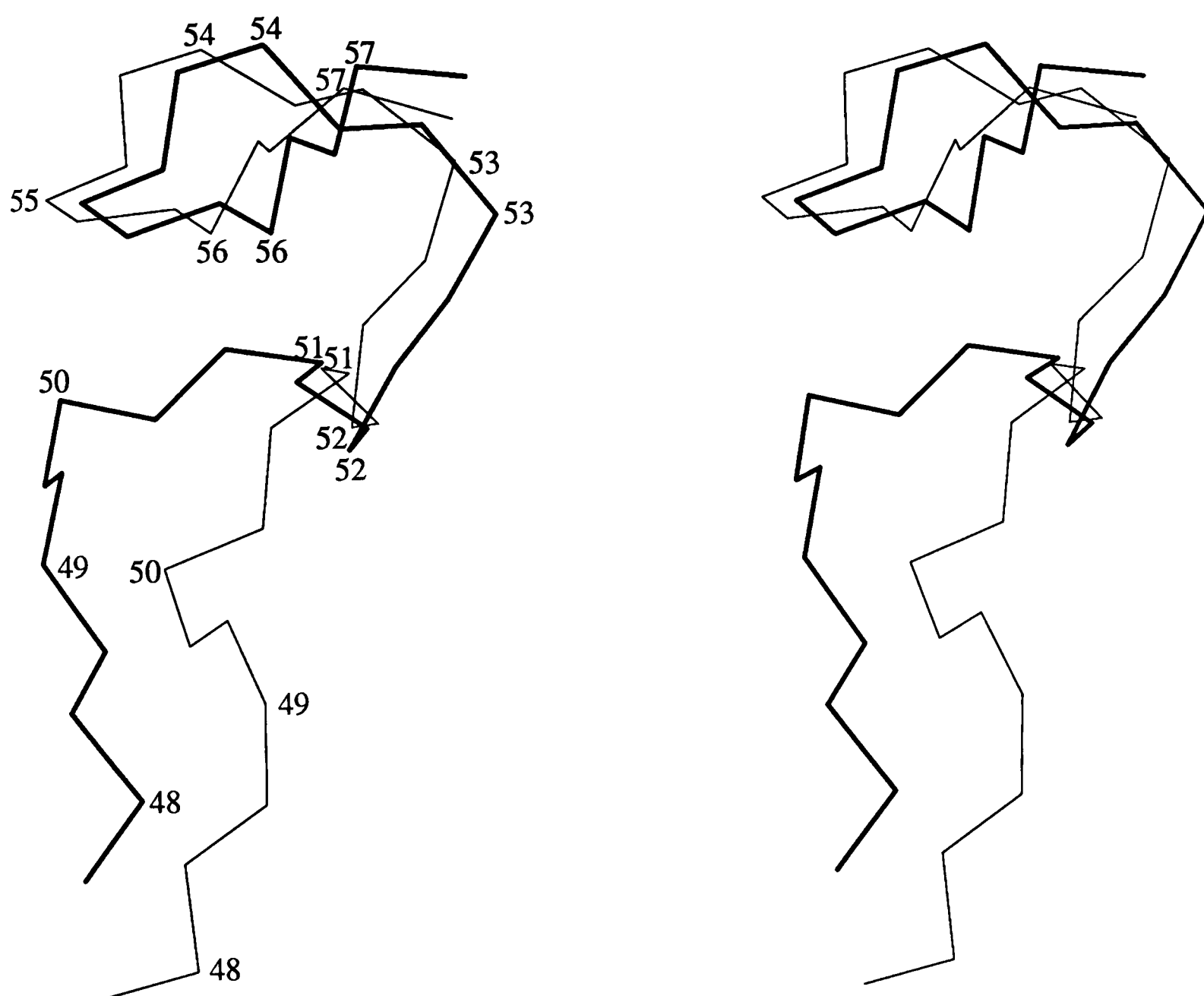


Figure 7.16: Comparison of the helical segments of poplar and parsley (bold) Cu^I plastocyanins. C^α atoms are labelled. C^α atoms of residues 1–56 and 59–97 were superposed.

Poplar	Parsley	<i>E. prolifera</i>
Ile 46	Val 46	Val 46
Ser 53	Glu 53	Asp 53
Met 57	Gln 57	Ala 57
Phe 70	Phe 70	Val 70
Val 72	Val 72	Arg 72

Table 7.13: Changes in primary structure which are postulated to be responsible for a shift in the position of the helix towards strand 4 and away from strand 6 in *E. prolifera* plastocyanin relative to poplar plastocyanin. The corresponding residues in parsley plastocyanin are included.

parsley plastocyanin in solution, with a certain percentage of molecules containing one or the other interaction. The 41NH-56O γ hydrogen bond was not observed in the crystal structure of poplar plastocyanin [22], but is present in the *E. prolifera* plastocyanin crystal structure [24]. This difference in the two solid state structures has been ascribed to changes in the sequence, most importantly the replacement of Ile 46, Phe 70 and Val 72 in the poplar protein by Val 46, Val 70 and Arg 72 in *E. prolifera* plastocyanin (Table 7.13). Resulting changes in side chain volumes allow the helix to move closer to strand 4, permitting the formation of a number of hydrogen bonds, one of which is 41NH-56O γ . The side chain changes between parsley and poplar plastocyanins (Table 7.13) are perhaps less significant than those observed for *E. prolifera* plastocyanin. Nevertheless, certain side chains are in closer proximity than in poplar plastocyanin. A particularly important potential interaction is a hydrogen bond from O ϵ of Glu 43 (strand 4) to O δ of Asp 53 (helix). These two groups are much closer than in poplar plastocyanin (5.4Å *versus* 10Å). The chemical consequences of this proposed hydrogen bond are discussed in the section on the acidic patch below.

7.4.9 Aromatic Residues

Parsley plastocyanin has aromatic residues in sequence positions which correspond to those in poplar plastocyanin. In addition, parsley plastocyanin has a tyrosine at position 62(60), a characteristic it shares with barley plastocyanin and algal plastocyanins such as those from *Scenedesmus obliquus* and *Enteromorpha prolifera*. Tyr 62(60) is surface exposed and is located in the northern half of the molecule near the Cu site

(Figure 7.17). The aromatic side chains of parsley plastocyanin occupy very similar spatial positions to those in poplar plastocyanin (Figure 7.17). The importance of the side chain of Phe 14 as a component of the Cu site has already been discussed in this chapter under ‘Loops and Tight Turns’. Some of the other aromatic residues are considered below.

Tyrosine 80(78)

The environment of invariant Tyr 80(78) is illustrated, along with some other features of the southern half of parsley and poplar plastocyanins, in Figure 7.18. Sequence considerations indicate that the environment of Tyr 80(78) is highly conserved. This is supported by the general similarity of tertiary structure in this region of poplar and parsley plastocyanins. In both proteins, the orientation of the ring of Tyr 80(78) is controlled by the formation of a hydrogen bond between the hydroxyl group of the ring and the peptide carbonyl of residue 76(74). In poplar plastocyanin, the side chain of Pro 47 makes a close nonbonded contact with the methylene group of Gly 78(76). The C^β and C^γ protons of Pro 47 are, on average, 1.3Å further from the methylene group of Gly 78(76) in parsley plastocyanin than in poplar plastocyanin. The former cannot therefore be said to contain a similarly close nonbonded contact. Both Pro 47 and Gly 78(76) are invariant. It has been proposed that Pro 47 may be essential for the initiation of the turn at residues 47–50 [22]. The invariance of Gly 78(76) is more difficult to explain. The same authors [22] suggest that the spacing between strands 4 and 7 in this region is critical. However, the retention of a glycine in position 78(76) of the parsley protein is still accompanied by an increase in the strand 4-7 separation relative to that in poplar plastocyanin of between 0.3Å and 1.6Å.

Tyrosine 70(68)

The value of the thermal parameter B of Tyr 70 indicates that this residue is ordered in the crystal structures of both poplar and *E. proliferans* plastocyanins [22, 24]. Among the simulated annealing structures of parsley plastocyanin, however, there is some disorder in the Tyr 70(68) side chain. This may be correlated with the obser-

vation that the ring of Tyr 70(68) is significantly more exposed to solvent in parsley plastocyanin than in poplar plastocyanin.

Tyrosine 83(81)

In contrast to the case of Tyr 70(68), the solvent exposed side chain of Tyr 83(81) is extremely well defined (Figure 7.6). There are a total of fifty two distance restraints involving Tyr 83(81), compared with seventeen involving Tyr 70(68). This precise definition of the side chain of Tyr 83(81) has possible functional significance, given its postulated role in the transmission of electrons to and/or from the Cu site. The implication is that a reasonably fixed orientation is required for the reception and/or delivery of electrons. It is noted that the Tyr 83(81) ring of parsley plastocyanin has an orientation approximately perpendicular to that of poplar plastocyanin. Whether this would have a significant effect on reactivity is not known.

7.4.10 Acidic Patch

This recognition site was described in Chapter 1 with respect to poplar plastocyanin. It is helpful to reconsider the nature of the acidic patch here. It consists of several negatively charged residues located on mid-strand β -turns at positions 42–45 and 58–61. The acidic side chains project into solvent, together forming an elongated patch of negative charge which encompasses the solvent exposed side chain of Tyr 83. Ironically, the only higher plant plastocyanin for which the Asp-Glu-Asp-Glu sequence at 42–45 is broken is poplar plastocyanin (45 is serine). Most higher plant plastocyanins have the sequence Glu-Glu-Asp/Glu at positions 59–61.

As indicated in several previous instances, parsley plastocyanin is exceptional among the higher plant plastocyanins in that the turn at positions 58–61 is eliminated by the deletion of two residues (other higher plant plastocyanins, for example that from barley, have more recently been shown to have deletions in the same sequence positions). Further, the sequence Gln-Pro-Glu in parsley plastocyanin replaces Glu-Glu-Asp/Glu at positions corresponding to 59–61. The resulting differences in the acidic patches of poplar and parsley plastocyanins are shown in Figure 7.19. The acidic patch characteristics described for parsley plastocyanin are common in algal



Figure 7.17: Superposition of the polypeptide backbones of poplar (pH 7.0) and parsley (pH 7.3; in bold) Cu^I plastocyanins, showing the aromatic side chains of each. The C^α atoms of residues 1-56 and 59-97 were superposed.

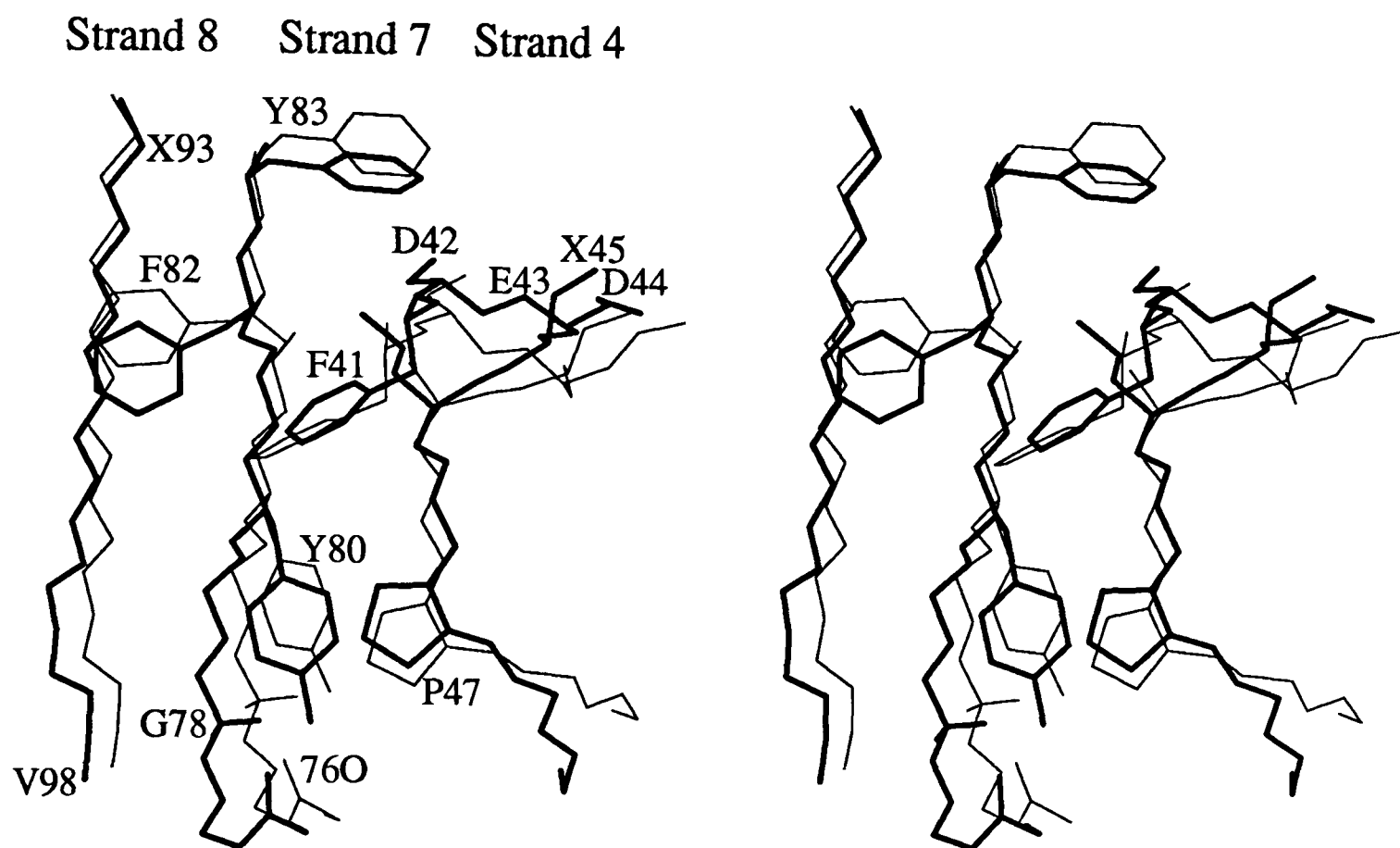


Figure 7.18: Stereo view of part of the southern half of parsley (bold) and poplar Cu^I plastocyanins. Part of the east binding site is shown, with the side chains of acidic residues 42–45 protruding into the solvent. Note that residue 45 is a serine in poplar plastocyanin, glutamate in parsley plastocyanin. Other differences among the highlighted amino acids are at positions 76(74) (asparagine in poplar plastocyanin, glutamate in parsley plastocyanin) and 93(91) (valine in poplar plastocyanin, lysine in parsley plastocyanin).

Source	$k_{Fe}, M^{-1}s^{-1}$	$k_{Co}, M^{-1}s^{-1}$	k_{Fe}/k_{Co}
Parsley	94000	3000	31
Spinach	85000	2500	34
French bean	58000	4700	12
Poplar	69000	2900	25
<i>S. obliquus</i>	90000	1850	25
<i>A. variabilis</i>	670000	630	1060

Table 7.14: Comparison of rate constant parameters for plastocyanins from different sources, taken from the papers of Sykes and co-workers. Measurements at 25°C, $I = 0.10M$ for the $Fe(CN)_6^{3-}$ (k_{Fe}) and $Co(phen)_3^{3+}$ (k_{Co}) oxidation of Cu^I plastocyanin.

plastocyanins, although the composition of the 58–61 sequence is variable. His-Asp-Asp and Ala-Glu-Asp are the triplets occupying positions 59–61 of *S. obliquus* and *E. prolifera* plastocyanins, for example.

Figure 7.19 shows that the conformations of poplar and parsley plastocyanin are closely similar at residues 40–46. In the more northerly portion of the acidic patch, elimination from parsley plastocyanin of the prominent kink at residues 58–61 has two major effects: the side chain of Gln 59(57) is directed more towards the molecular interior, and the two sets of residues (42–45 and 59(57)–61(59)) constituting the acidic patch are closer to one another than in poplar plastocyanin. Corresponding pairs of atoms, one each in the 42–45 and 59(57)–61(59) sections, are separated by 2–5Å less in parsley than in poplar plastocyanin.

It would be expected that a reduction in the number of acidic residues and absence of a turn lead to a diminished contribution of residues 59(57)–61(59) to the acidic patch. As a result, the negative electrostatic field would be expected to be less extensive and of reduced strength in parsley plastocyanin. This might be expected to lead to reduced reactivity. The results of the extensive kinetic studies of Sykes and co-workers [31, 32, 33, 34], which were discussed in Chapter 1, indicate that the reactivity of parsley plastocyanin is not significantly reduced relative to the reactivities of its higher plant counterparts (Table 7.14).

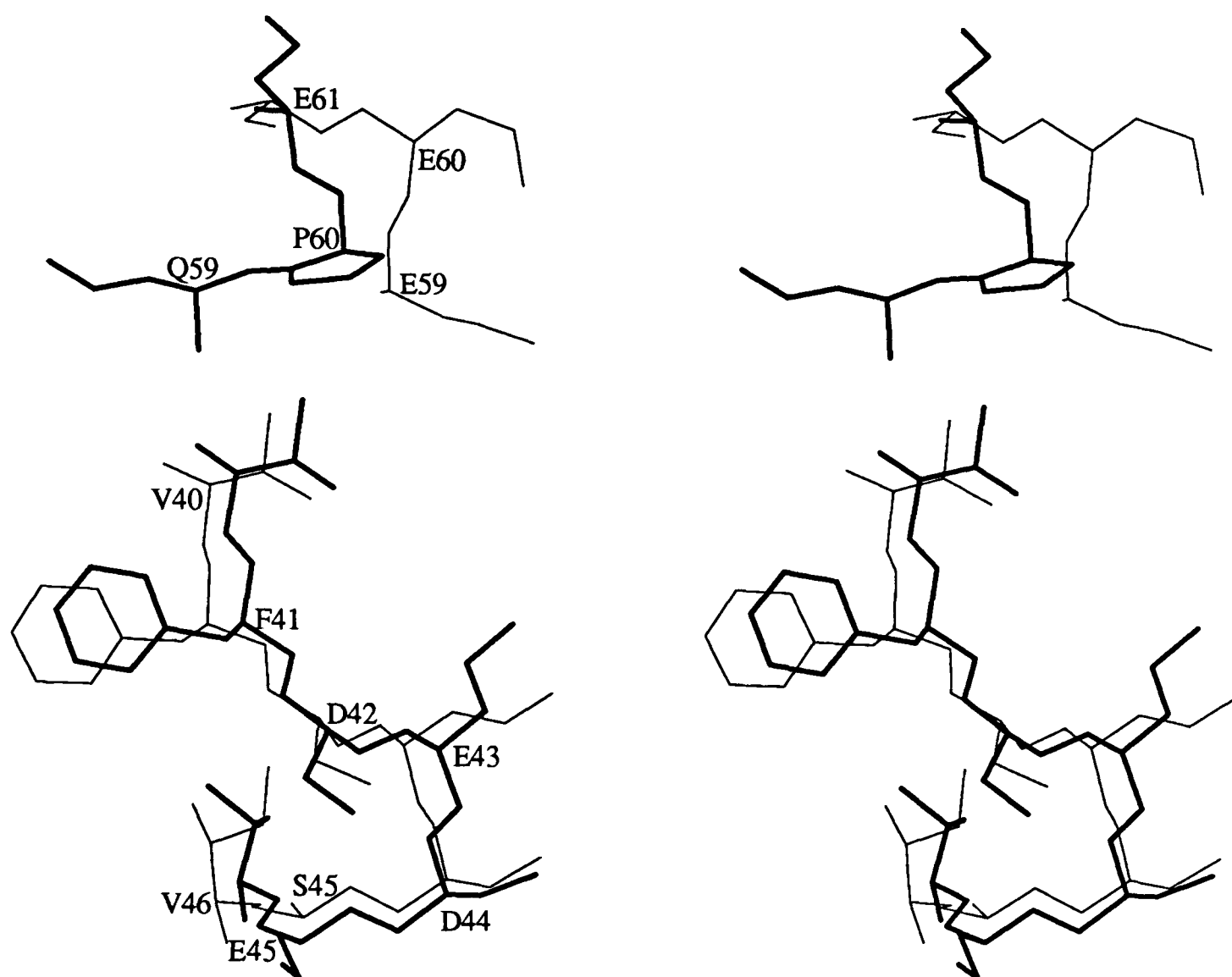


Figure 7.19: Residues 40–46 and 59(57)–61(59) of poplar and parsley Cu^I plastocyanins. Latter shown in bold. For residues 59(57)–61(59), C^α atoms of both proteins are labelled, since sequence and conformation are significantly different. For residues 40–46, only one C^α label is used for each residue, since the sequences and conformation of the two proteins are similar. The exception is residue 45, which is serine in poplar plastocyanin and glutamate in parsley plastocyanin.

7.4.11 Acidic Patch Hydrogen Bond

It has been mentioned above (in the section ‘Hydrogen Bonding Involving Residues 41, 52 and 56’) that a potential hydrogen bond exists between the side chain carboxylate groups of residues 43 and 53. The distance between O^ε of Glu 43 (strand 4) and O^δ of Asp 53 (helix) in the minimised average structure of parsley plastocyanin is 5.4 Å. This separation is well outside the normal range for consideration of the interaction as a hydrogen bond. However, the two oxygen atoms are much closer than in poplar plastocyanin, where they are 10 Å apart (Figure 7.20). Also, both Glu 43 and Asp 53 have poorly defined side chain conformations in the fifty-eight calculated structures. Such disorder is often observed for surface side chains in protein structures determined from NMR-derived data. This conformational flexibility makes it conceivable that in a certain percentage of parsley plastocyanin molecules at least, the -COOH ... -OOC hydrogen bond may indeed occur. The plastocyanin from *E. proliferans* contains just such a hydrogen bond [24].

In the discussion of the reactions of various plastocyanins with inorganic reagents (Chapter 1), it was stated that, in addition to the protonation and dissociation of His 87(85), an extra protonation is required to account for the reactions of plastocyanins with Co(phen)₃³⁺. For higher plant plastocyanins, it was said that the most acceptable process is protonation of carboxylate(s) at the acidic patch. The existence of the Glu 43 O^ε ... Asp 53 O^δ hydrogen bond is consistent with this hypothesis, and would account for the anomalously high pK_a of parsley plastocyanin (Table 7.15; data taken from [33, 34]). Further, an unusually high pK_a has been implicated for *E. proliferans* plastocyanin [24] on the basis of the occurrence of the -COOH ... OOC⁻ hydrogen bond (see Chapter 1, 1.3).

7.5 Electrostatic Field Calculations using DelPhi

The program DelPhi [35, 36] has been used to calculate the electrostatic fields of poplar and parsley Cu^I plastocyanins.

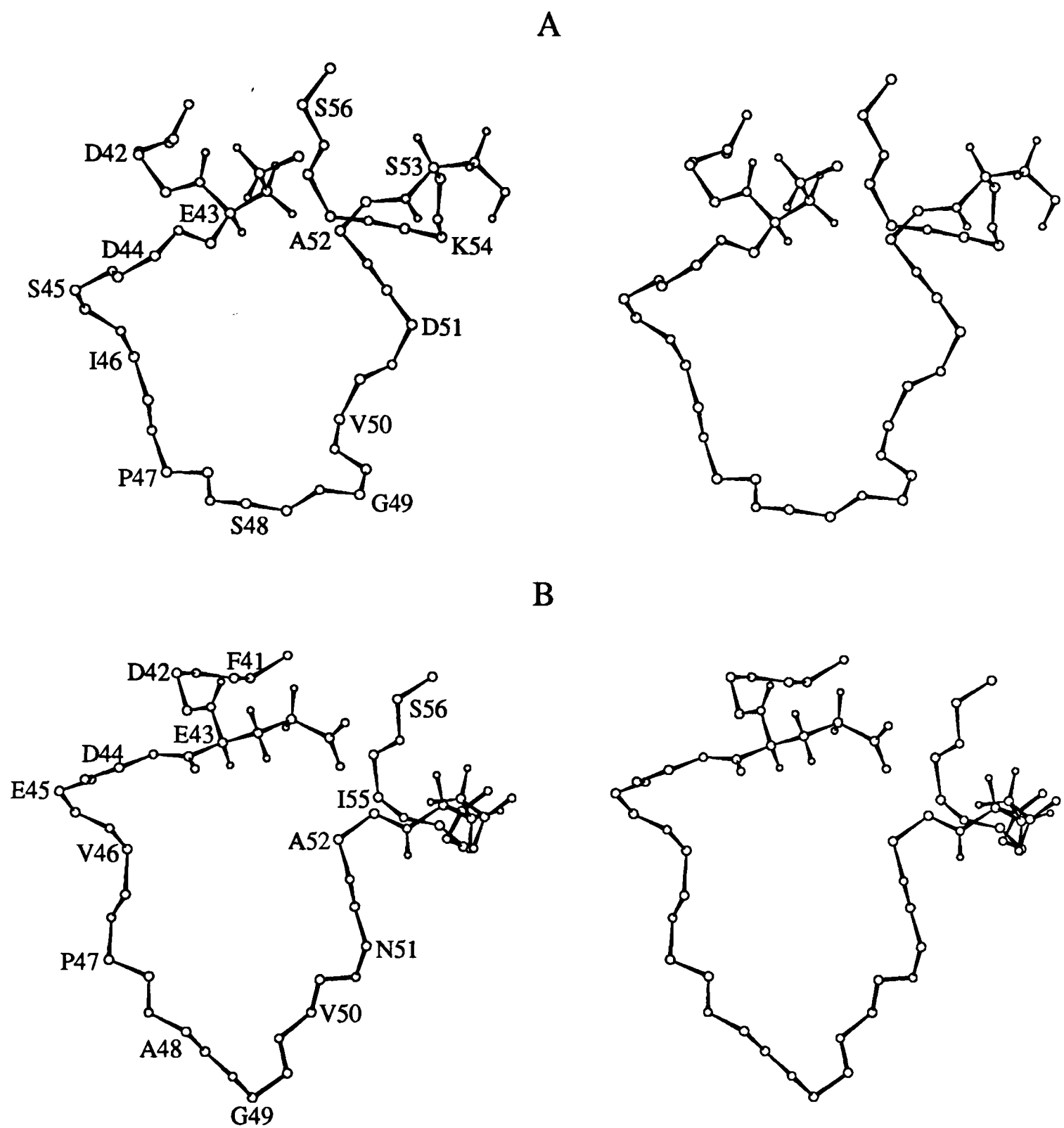


Figure 7.20: Stereoview comparing the side chain orientations of residues 43 and 53 in (A) poplar and (B) parsley Cu^I plastocyanins.

Cu ^I Pc source	pK _a			pK' _a
	NMR	Fe(CN) ₆ ³⁻	Co(phen) ₃ ³⁺	
Parsley	5.7	5.5	6.1	5.8
Spinach	4.9	4.8	5.6	5.6
French bean	4.7	4.6	5.4	5.5
Poplar	N/A	4.7	5.2	5.3
<i>S. obliquus</i>	5.4	5.0	5.6	5.5
<i>A. variabilis</i>	5.1	5.0	5.4	5.7

Table 7.15: Summary of acid dissociation constants as determined by NMR and kinetic studies with Fe(CN)₆³⁻ and Co(phen)₃³⁺ as oxidants at 25°C. The values in the last column assume that two processes pK_a (determined by NMR or with Fe(CN)₆³⁻ as oxidant) and pK'_a contribute with Co(phen)₃³⁺ as oxidant.

7.5.1 Description of the Program and Parameters Used

DelPhi uses a macroscopic continuum approach to model the protein and solvent materials. The atomic coordinates of the protein are mapped on to a three dimensional grid. United atom van der Waals radii are assigned to the heavy atoms to define the surface of the protein. Different dielectric constants can be assigned to the protein and solvent: in this case, a dielectric constant of two was used for the protein, and one of eighty for the solvent. These are typical values. The protein surface defines the dielectric boundary. The Stern layer thickness (closest approach of solvated ions to the protein surface) was set to 2Å. Fractional electrostatic charges are mapped on to the grid such as to provide spherically symmetric distributions around the charged atoms of the protein. The electrostatic potential field of the protein is then calculated by numerical solution of the Poisson-Boltzmann equation by the finite difference method.

A cubic grid of side length 65 units was used. The plastocyanin molecule was mapped on to the grid such that the geometric centre of the molecule was at the centre of the grid. In order to fit the plastocyanin electrostatic field into the grid, the scale was adjusted so that the longest dimension of the molecule was 57% of the side length of the grid. All calculations were carried out with ionic strength set to 0.145 moles per litre.

7.5.2 Preliminary Results

Notwithstanding the fact that this investigation is in its early stages, the pictures shown in Figures 7.21, 7.22 and 7.23 provide an interesting insight into the effect of the sequence differences between the two plastocyanins. The pictures show the electrostatic field with the polypeptide backbone in three orientations:

1. Figure 7.21 shows the field with the protein in the standard orientation. That is, the orientation shown in Figures 7.3 and 7.10. Thus, Cu lies towards the north of the molecule, the acidic patch is to the east, the N-terminus to the west and the C-terminus to the left of centre at the southern end of the molecule. The largest negative part of the field is due to the easterly acidic patch residues. The region of negative field in the north-western corner is associated with Asp 8 and Asp 9, that below it with Glu 25. A patch of positive field at the hydrophobic north recognition site of the molecule is observed, projecting into the space between the acidic patch and Asp 8/Asp 9 negative regions. The positive region at the middle of the southern end of the molecule can be ascribed in part to Lys 77(75), whilst that projecting to the west is due to Lys 26 and Lys 30. The negative field of parsley plastocyanin is clearly less extensive than that of poplar plastocyanin.
2. Figure 7.22 shows the electrostatic field looking directly on to the acidic patch. Whereas the negative field of the poplar protein forms a single patch, that of parsley plastocyanin is broken up by regions of positive potential. The three small patches of positive potential in the parsley plastocyanin field are associated with Lys 81(79) (central; residue 81 is a serine in poplar plastocyanin), Lys 26 (lower east side) and Lys 77(75) (south). The positive hydrophobic patch to the north is again evident.
3. Figure 7.23 shows the electrostatic field from the rear of the molecule, i. e. with the protein rotated through approximately 180° about the vertical axis from the standard orientation described above. The acidic patch is now on the left, Asp 8 and Asp 9 are in the top right corner. This, together with the two previous orientations, illustrates how the negative field of poplar plastocyanin forms

a complete belt around the molecule, whereas that of the parsley protein is incomplete. In the case of the latter, lysines 81(79), 26 and 77(75) are responsible for the three positive regions, as described above for the second molecular orientation.

Figures 7.21, 7.22 and 7.23 show that the negative electrostatic fields of both poplar and parsley plastocyanins are extensive. That of poplar plastocyanin essentially forms a belt around the molecule, whereas that of parsley plastocyanin contains breaks where regions of positive field are observed.

7.6 Overview

The most significant differences between poplar and parsley plastocyanins occur in the acidic patch. This region of the protein has been implicated as a means of facilitating electrostatic recognition of a redox partner, and is thus important for function.

It is postulated that parsley plastocyanin contains a $\text{-COOH} \cdots \text{-OOC}$ hydrogen bond involving the side chains of residues 43 and 53. The existence of this interaction would account for the anomalously high pK_a of parsley plastocyanin. Such a hydrogen bond is not observed in poplar plastocyanin.

Deletion of two residues deprives the parsley protein of a prominent mid-strand kink at residues 58–61. In poplar and most other higher plant plastocyanins, this kink is oriented such as to project negatively charged side chains into solvent. The deletions also result in closer proximity of residue groups 42–45 and 59–61 in parsley plastocyanin. Further distinction of the parsley protein from most of its higher plant counterparts derives from replacement at residues 59–61 of the common Glu-Glu-Asp/Glu sequence by Gln-Pro-Glu. These non-conservative residue substitutions and loss of the 58–61 kink have been shown to result in diminution of the negative electrostatic field of parsley plastocyanin.

The acidic patch features described for parsley plastocyanin are common to algal plastocyanins, such as those from *Scenedesmus obliquus* and *Enteromorpha prolifera*. A few other higher plant plastocyanins which have been sequenced more recently share some of the unusual sequence characteristics with parsley plastocyanin. For example,

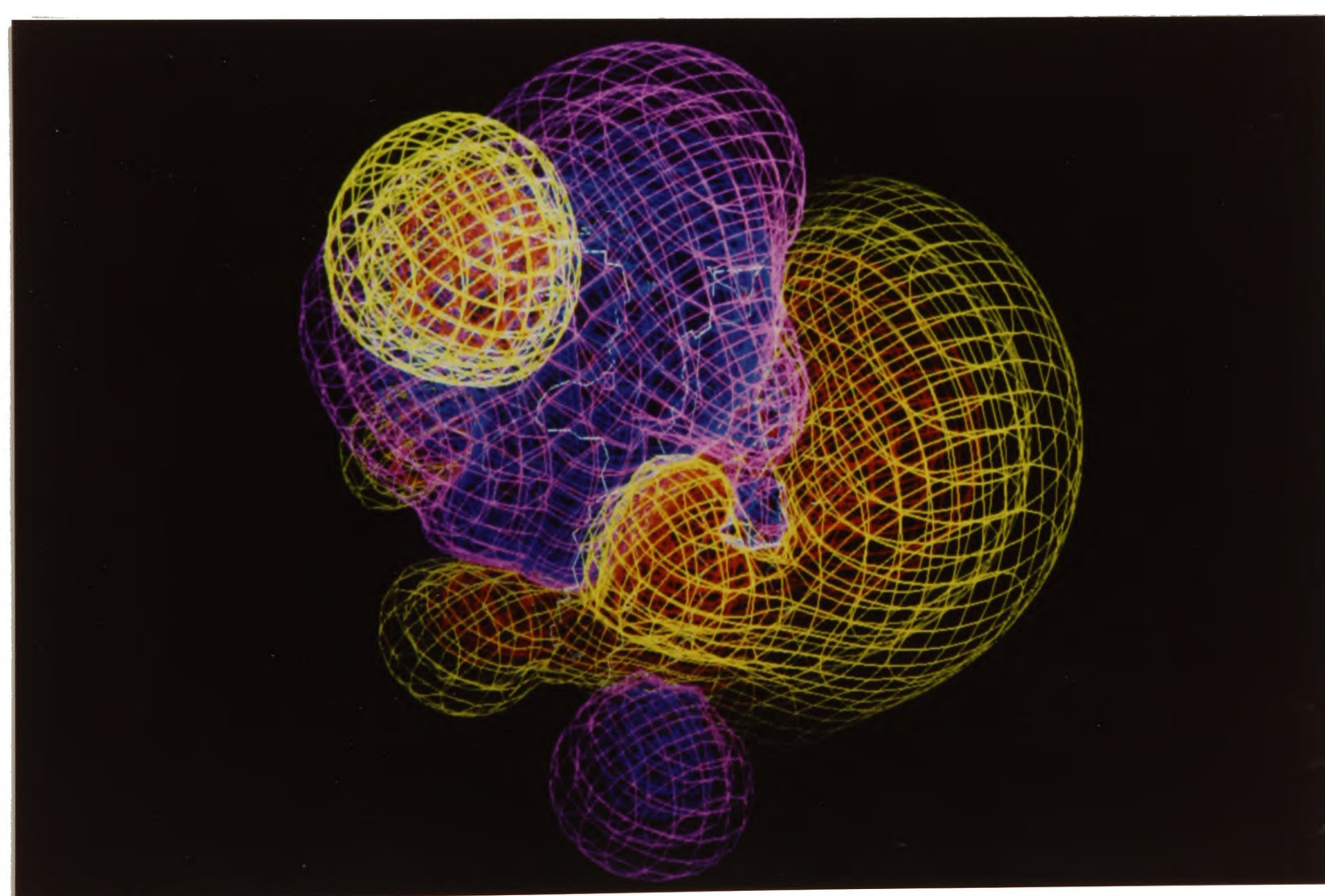
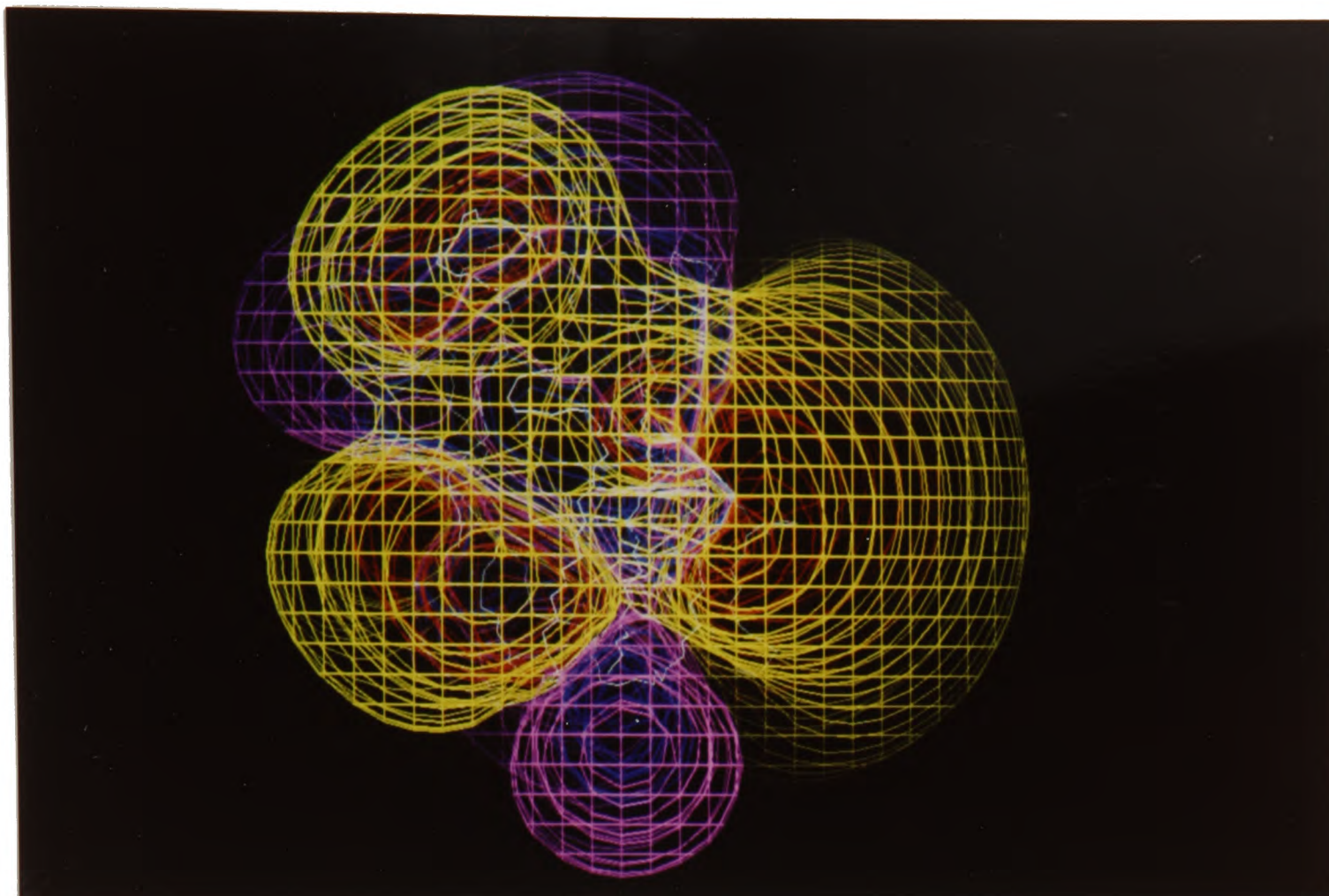


Figure 7.21: The electrostatic field around poplar (top) and parsley Cu^I plastocyanins in the standard orientation. Negative regions are yellow, positive magenta.

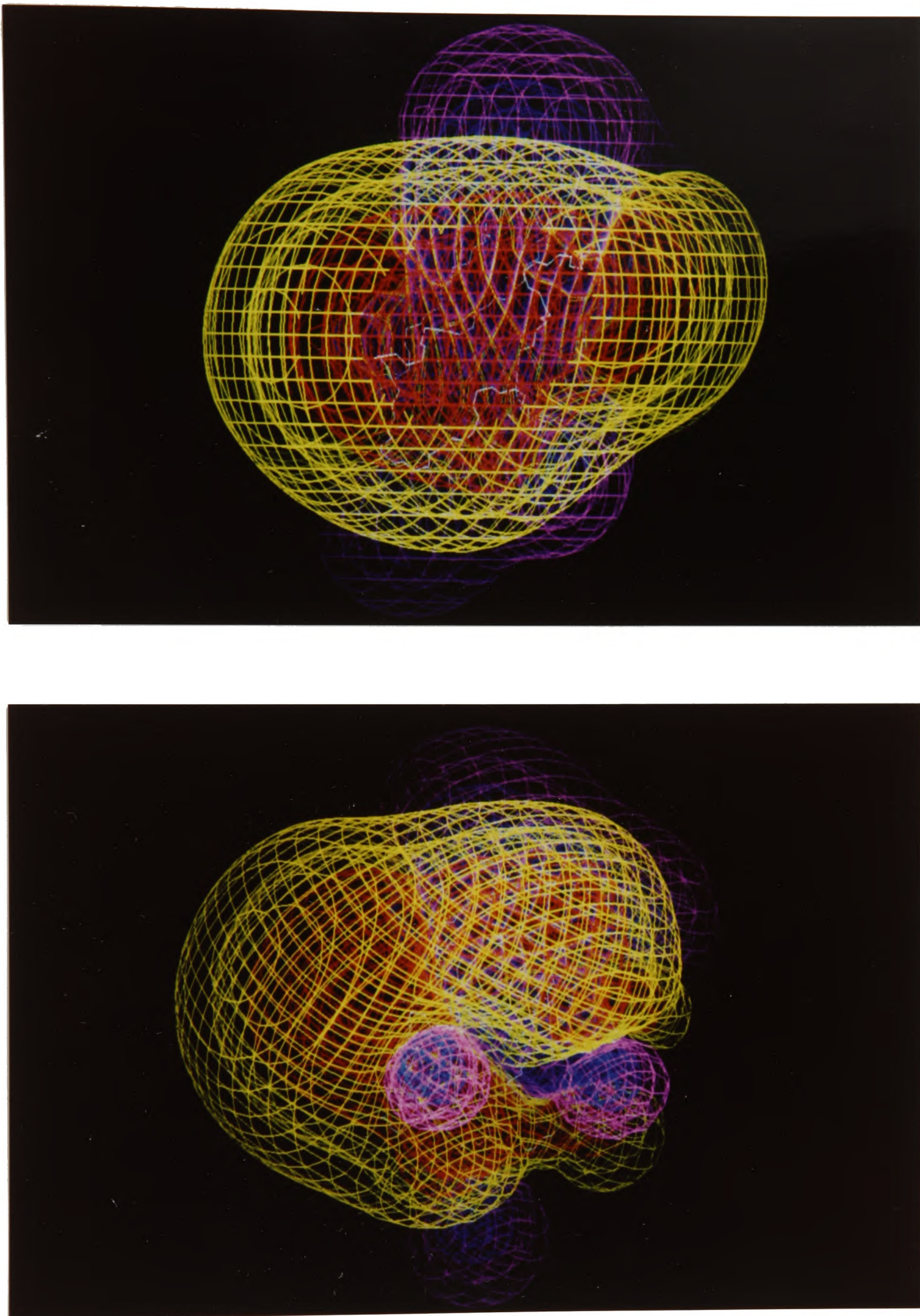


Figure 7.22: The electrostatic field around poplar (top) and parsley Cu^I plastocyanins looking directly on to the acidic patch. Negative regions are yellow, positive magenta.

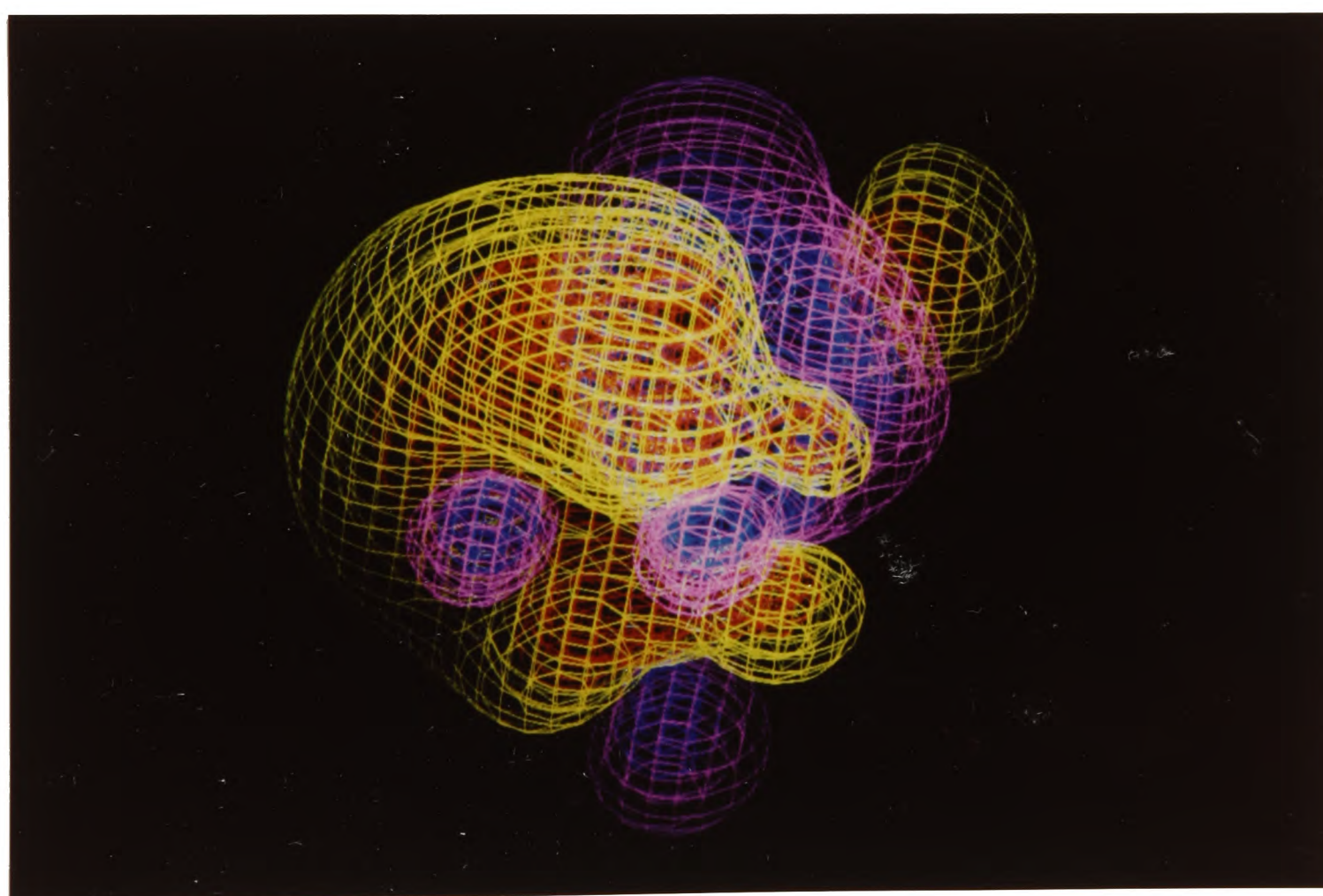
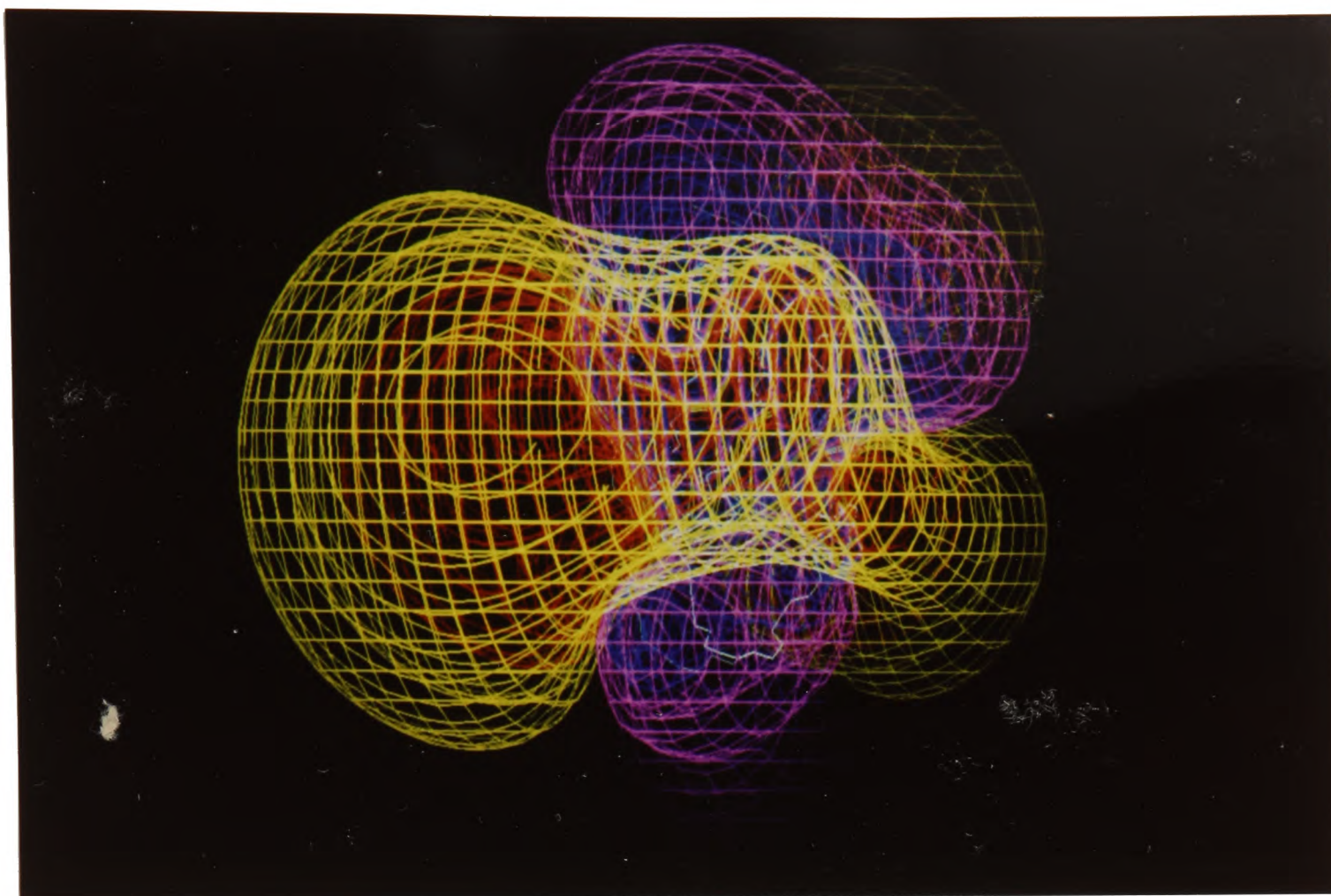


Figure 7.23: The electrostatic field around poplar (top) and parsley Cu^I plastocyanins viewing the molecule from the rear. That is, the molecule has been rotated through approximately 180° about the vertical axis.

barley plastocyanin has deletions at positions 57 and 58, a tyrosine at position 62 and a glutamate at position 85. Unique among the plastocyanins is that from the blue-green alga *Anabaena variabilis*, in which most of the acidic patch residues are replaced by neutral or positive residues. Thus, it seems that during evolution from the blue-green algae through the green algae to higher plants, the acidic patch has developed as a structurally distinctive and functionally important feature of plastocyanin. Parsley and the other anomalous higher plant plastocyanins are then placed as a link between the green algal and remaining higher plant plastocyanins.

Once the structure of cytochrome *f* is determined, facilitating more definitive study of the plastocyanin/cytochrome *f* interaction than has hitherto been possible, the functional consequences of acidic patch diminution in parsley and other plastocyanins can be investigated.

Bibliography

- [1] K. Wüthrich, M. Billeter & W. Braun (1983) *J. Mol. Biol.* 169, 949–961.
- [2] G. Wagner, W. Braun, T. F. Havel, T. Schaumann, N. Gō & K. Wüthrich (1987) *J. Mol. Biol.* 196, 611–639.
- [3] G. M. Clore, A. M. Gronenborn, M. Nilges & C. Ryan (1987) *Biochemistry* 26, 8012–8023.
- [4] C. Redfield & C. M. Dobson (1990) *Biochemistry* 29, 7201–7214.
- [5] L. J. Smith, M. J. Sutcliffe, C. Redfield & C. M. Dobson (1991) *Biochemistry* 30, 986–996.
- [6] A. Pardi, M. Billeter & K. Wüthrich (1984) *J. Mol. Biol.* 180, 741–751.
- [7] G. M. Clore, M. Nilges, D. K. Sukumaran, A. T. Brünger, M. Karplus & A. M. Gronenborn (1986) *EMBO J.* 5, 2729–2735.
- [8] P. C. Driscoll, A. M. Gronenborn & G. M. Clore (1989) *FEBS Lett.* 243, 223–233.
- [9] E. R. P. Zuiderweg, R. Boelens & R. Kaptein (1985) *Biopolymers* 24, 601–611.
- [10] M. Nilges, G. M. Clore & A. M. Gronenborn (1990) *Biopolymers* 29, 813–822.
- [11] P. J. Kraulis, G. M. Clore, M. Nilges, T. A. Jones, G. Pettersson, J. Knowles & A. M. Gronenborn (1989) *Biochemistry* 28, 7241–7257.

- [12] A. T. Brünger (1988) *X-PLOR Manual*, Yale University, New Haven, CT.
- [13] J. M. Guss, P. R. Harrowell, M. Murata, V. A. Norris & H. C. Freeman (1986) *J. Mol. Biol.* 192, 361–387.
- [14] M. Nilges, G. M. Clore & A. M. Gronenborn (1988) *FEBS Lett.* 229, 317–324.
- [15] T. F. Havel (1986) *DISGEO, Quantum Chemistry Program Exchange No. 507*, Indiana University, Bloomington, IN.
- [16] M. Nilges, G. M. Clore & A. M. Gronenborn (1988) *Protein Eng.* 2, 27–38.
- [17] M. Nilges, Yale University, Feb. 1990.
- [18] A. T. Brünger, G. M. Clore, A. M. Gronenborn & M. Karplus (1987) *Protein Eng.* 1, 399–406.
- [19] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan & M. Karplus (1983) *J. Comput. Chem.* 4, 187–217.
- [20] G. M. Clore, A. M. Gronenborn, A. T. Brünger & M. Karplus (1985) *J. Mol. Biol.* 186, 435–455.
- [21] J. P. Ryckaert, G. Cicotti & H. J. C. Berendsen (1977) *J. Comp. Phys.* 23, 327–341.
- [22] J. M. Guss & H. C. Freeman (1983) *J. Mol. Biol.* 169, 521–563.
- [23] M. Nilges, G. M. Clore & A. M. Gronenborn (1987) *FEBS Lett.* 219, 11–16.
- [24] C. A. Collyer, J. M. Guss, Y. Sugimura, F. Yoshizaki & H. C. Freeman (1990) *J. Mol. Biol.* 211, 617–632.
- [25] E. N. Baker & R. E. Hubbard (1984) *Prog. Biophys. Mol. Biol.* 44, 97–179.
- [26] R. Huber & W. Steigemann (1974) *FEBS Lett.* 48, 235–237.
- [27] J. S. Richardson (1981) *Adv. Protein Chem.* 34, 167–339.

- [28] W. J. Chazin & P. E. Wright (1988) *J. Mol. Biol.* 202, 623–636.
- [29] J. M. Moore, W. J. Chazin, R. Powls & P. E. Wright (1988) *Biochemistry* 27, 7806–7816.
- [30] J. S. Richardson, E. D. Getzoff & D. C. Richardson (1978) *Proc. Natl. Acad. Sci. USA* 75, 2574–2578.
- [31] S. K. Chapman, I. Sanemasa & A. G. Sykes (1983) *J. Chem. Soc. Dalton Trans.*, 2549–2553.
- [32] J. D. Sinclair-Day & A. G. Sykes (1986) *J. Chem. Soc. Dalton Trans.*, 2069–2073.
- [33] M. P. Jackman, J. D. Sinclair-Day, M. J. Sisley, A. G. Sykes, L. A. Denys & P. E. Wright (1987) *J. Am. Chem. Soc.* 109, 6443–6449.
- [34] J. McGinnis, J. D. Sinclair-Day, A. G. Sykes, R. Powls, J. Moore & P. E. Wright (1988) *Inorg. Chem.* 27, 2306–2312.
- [35] I. Klapper, R. Hagstrom, R. Fine, K. Sharp & B. Honig (1986) *Proteins* 1, 47–59.
- [36] M. K. Gilson, K. A. Sharp, & B. H. Honig (1987) *J. Comput. Chem.* 9, 327–335.

