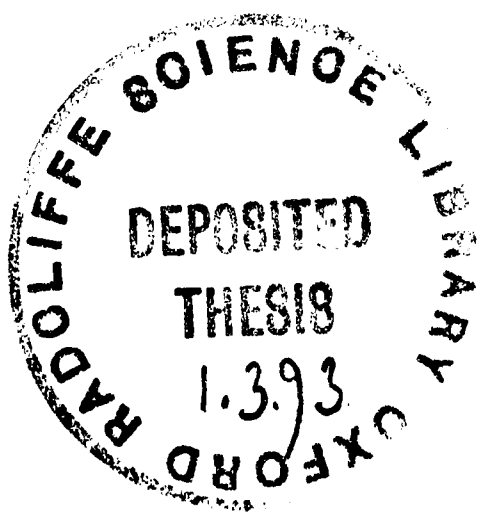
PROTEIN STRUCTURES FROM NMR DATA

A thesis submitted to the University of Oxford in partial fulfilment of the requirements for the degree of Doctor of Philosophy.

Lorna J. Smith.

Lady Margaret Hall.
Michaelmas Term, 1992.



To my parents.

ABSTRACT

Protein Structures from NMR Data.

A thesis submitted for the degree of Doctor of Philosophy

by Lorna J. Smith.

Lady Margaret Hall, Oxford. Michaelmas Term, 1992.

This thesis describes the use of nuclear magnetic resonance techniques to determine the structures of two proteins in solution, hen egg-white lysozyme and human interleukin-4.

Using 2D ^{1}H methods an extensive set of structural data has been collected for hen lysozyme (1158 NOE distance restraints, 68 ϕ and 24 χ_1 dihedral angle restraints) and these data have been used in distance geometry-dynamical simulated annealing calculations to determine an ensemble of NMR structures for the protein. The overall Cα RMSD from the average for a set of 16 calculated structures is 1.8 ± 0.2 Å but, excluding 14 residues in exposed disordered regions, this value reduces to 1.3 ± 0.2 Å. Regions of secondary structure, and particularly the four α helices, are well defined (Cα RMSD 0.8 ± 0.3 Å for helices). Detailed comparisons of the NMR structures with crystal structures of the protein have shown the close similarity of the main chain fold and the conformation of interior side chains in solution and in crystals. $^{3}J_{\alpha\beta}$ coupling constant measurement have indicated, however, that the conformational mobility of the side chains of many surface residues is significantly more pronounced than an individual crystal structure would suggest.

For human interleukin-4, a strategy involving ^{15}N and ^{13}C labelled recombinant protein together with heteronuclear 3D NMR techniques has been employed to determine the structure of the protein. The work has provided the first structure for this protein, a left-handed four helix bundle with an up-up-down-down connectivity. For an ensemble of 10 final calculated NMR structures there is a Cα RMSD from the average of 1.6 ± 0.4 Å, the definition of the helical core of the protein being particularly good (0.8 ± 0.2 Å). There is, however, some disorder in the long overhand loops of the structure; this reflects the unusually high conformational mobility of these regions. Comparison of the interleukin-4 structure with proteins with related folds, particularly members of the haemopoietic cytokine superfamily, suggests that the fold found here for interleukin-4 may be the adopted structure throughout this cytokine superfamily.

In a postscript to this thesis the NMR structure of human interleukin-4 is shown to have a very similar conformation to a crystal structure of the protein which has been solved very recently.

CONTENTS

Acknowledgements.

I would first like to thank my supervisor, Dr. Chris Dobson, for his enthusiastic advice and encouragement throughout the course of my research. I have very much valued and enjoyed the opportunity of working as a graduate student in his research group.

I would also like to thank Dr. Christina Redfield, with whom I collaborated for much of the work described in this thesis, for all her help and advice; I have learnt a great deal from her clear scientific thinking. Thanks are also due to:

Dr. Jonathan Boyd, who helped me set up, record and process the 3D NMR experiments on human IL-4;

SmithKline Beecham Pharmaceuticals, and in particular Dr. Richard Smith, who provided the samples of human IL-4;

Dr. Mike Sutcliffe, who gave advice on protein structure calculations;

Drs. Fernando Bazan and Chris Bystroff at UCSF, who helped me understand some of the details of four helix bundles;

Sheena, Kim, Chris M., Jeremy and Andrew M. who proof read this thesis, and Jocelyn who helped with the ChemDraw diagrams.

In addition, I would like to thank all the members and associates of the Dobson Group, particularly Dan, Kevin, Sheena, Mike, Andrew, Kristin, Ursula, Jocelyn, Li, Jeremy, Kim, James, Steve, Chris, Matthew, Adrian, Caroline and everyone else in F21c, CG1 and F23b, who have provided such a friendly environment in which to work and in the past few months have sympathetically understood my thesis hassles.

I am very grateful for the funding I have received during my course: a studentship from the Science and Engineering Research Council, a graduate Scholarship from Lady Margaret Hall, a Scientist's Scholarship from ICI and, in recent weeks, a Junior Research Fellowship from St. Cross College, Oxford.

Many thanks to all the residents, past and present, of No. 11 Norham Gardens, especially Shirley, Judith, Philippa, Nigel and Kok Choong, for making it a very enjoyable place to live. I would also like to thank Fiona, Janet, Annabel, Alison, Chen-Chen and all the Ichthus Group at Wesley Memorial Church for their friendship.

Finally, I would like to thank my parents and the rest of my family for their love, care and understanding.

Abbreviations.

Å	Angstrom (1Å = 10^{-10}m)
COSY	J-correlated spectroscopy
CD	circular dichroism
2D	two-dimensional
3D	three-dimensional
DQF-COSY	double quantum filtered correlation spectroscopy
E. COSY	exclusive correlation spectroscopy
GM-CSF	granulocyte-macrophage colony-stimulating factor.
GRH	growth hormone
HMQC	heteronuclear multiple quantum coherence
HSQC	heteronuclear single quantum coherence
Hz	Hertz
IL	interleukin
IL-4	interleukin-4
NMR	nuclear magnetic resonance
NOE	nuclear Overhauser enhancement
NOESY	nuclear Overhauser enhancement spectroscopy
ppm	parts per million
RMSD	root mean square deviation
TOCSY	total correlation spectroscopy
τ_R	rotational correlation time

For amino acids the IUPAC-IUB one-letter or three-letter symbols are used.

CHAPTER 1.

INTRODUCTION.

Proteins play a wide variety of important roles in biological systems and consequently a detailed knowledge of their structures and dynamics is needed if biochemical processes are to be understood at a molecular level (Creighton, 1984). There has been much interest, therefore, in the recent developments in nuclear magnetic resonance spectroscopy (NMR) which are providing methods, complementary to diffraction techniques, for the determination of the structure of proteins (Wüthrich, 1989; Wagner, 1990; Clore & Gronenborn, 1991a).

2D NMR spectroscopy was first successfully used in protein structure determination in the mid-1980's, when the structures of micelle-bound glucagon (29 residues) (Braun et al., 1983) and then the lac repressor headpiece (51 residues) (Kaptein et al., 1985) were solved. Since then, there have been rapid advances in the scope and capabilities of protein structure determination by 2D NMR techniques. High field NMR spectrometers have become readily available, and a large array of NMR experiments for the assignment of protein spectra and the collection of structural data have been developed (Bax, 1989; Markley, 1989). In addition, a number of algorithms which can compute the protein structures that satisfy an NMR data set have been produced (Braun, 1987; Clore & Gronenborn, 1989; Kuntz et al., 1989; Scheek et al., 1989; Havel, 1991). The result of these advances is that 2D NMR techniques can now be used to determine the structures of small proteins of up to approximately 100 residues with an accuracy equivalent to that of a 2.0-2.5 Å resolution crystal structure (Clore & Gronenborn, 1991a). Recent examples of the use of 2D NMR techniques in structure determinations include the proteins eglin c, 70 residues (Hyberts et al., 1992), human

thioredoxin, 105 residues (Forman-Kay et al., 1991), plastocyanin, 99 residues (Moore et al., 1991), and porcine calbindin D_{9k}, 78 residues (Akke et al., 1992).

Under favourable conditions, structure determinations can also be carried out for larger proteins by 2D NMR techniques although often at a lower resolution e.g. ribonuclease A, 124 residues (Rico et al., 1991). In general, however, problems arise due to the greatly increased chemical shift degeneracy, and consequent overlap in the 2D spectra, as well as the larger line widths of the protein resonances. Some progress has been made with 3D homonuclear NMR techniques (Oschkinat et al., 1988) and also with selective isotopic labelling (Torchia et al., 1989; McIntosh & Dahlquist, 1990). Indeed, by employing selective ^{15}N-labelling and deuteration, a low resolution structure for the E. coli. trp repressor has been determined (25kDa. dimer, 107 residues/monomer) (Arrowsmith et al., 1991). However, in the last two years a major breakthrough for the structure determination of larger proteins has come from the development of 3D and 4D heteronuclear NMR experiments for uniformly ^{15}N and/or ^{13}C-labelled proteins (Fesik & Zuiderweg, 1990; Clore & Gronenborn, 1991c). These experiments lead to great spectral simplification and improved resolution as the information that would be contained in a conventional 2D spectrum is spread out into a third, or even a fourth, dimension. In addition, many of these experiments exploit large one-bond heteronuclear couplings for the magnetization transfer steps, which make them less sensitive to the increased line widths associated with larger proteins. Using these heteronuclear techniques the structures of larger proteins, as first shown for interleukin-1β with 153 residues (Clore et al., 1991b), can be determined with an accuracy comparable to that of the NMR structures of small proteins.

The structures provided by NMR methods are for proteins in solution and there has been much interest in the comparison of these structures with those of protein crystals. For globular proteins, where both NMR and crystal

structures have been determined, it has been generally found that the two structures are very similar, at least for atoms in the interior of the protein, but that the exposed regions are more disordered in solution e.g. comparisons for tendamistat (Billeter et al., 1989), interleukin-8 (Clore & Gronenborn, 1991b) and plastocyanin (Moore et al., 1991). As the exposed areas of proteins will be most directly involved in function and specificity, it is important to attempt to understand the extent and nature of the dynamics and disorder of surface regions in solution. In addition to giving structural information, NMR can also provide insight into conformational disorder of this type (Campbell et al., 1985; Williams, 1989). The rate of dynamic processes governs the effect they will have on the NMR spectrum (Jardetzky & Roberts, 1981). If the rate of interconversion between the conformational states adopted is slow on the NMR time scale, then separate resonances will be observed for the different conformations. By contrast, if the rate of interconversion is rapid, the observed NMR parameters will be an average of those for the individual conformations. Although the interpretation of motionally averaged parameters such as chemical shifts and NOE's is often difficult, as shown in Chapter 3 of this thesis, coupling constant measurements can give clear evidence for conformational disorder. In addition, NMR can also be used to probe dynamics through a study of hydrogen exchange rates (Wagner & Wüthrich, 1986; Pedersen et al., 1991) and, as has been shown recently, ^{15}N relaxation studies can give valuable information about backbone dynamics (Kay et al., 1989).

In this thesis structure determinations by NMR spectroscopy for two 'larger' proteins, hen egg-white lysozyme and human interleukin-4, are presented. The size of these proteins is very similar; hen lysozyme has 129 residues while recombinant human interleukin-4 has 130 residues. Hen lysozyme, considering its size, gives excellent 2D spectra with considerable

chemical shift dispersion. It has been possible, therefore, to use 2D ^{1}H NMR techniques alone to determine successfully the NMR structure for this protein in solution. Human IL-4, by contrast, has a predominantly helical content and consequently a much smaller chemical shift dispersion than lysozyme. The 2D ^{1}H NMR spectra of this protein are extremely crowded, with many peaks overlapping, and 3D heteronuclear NMR techniques using uniformly labelled samples of the protein have had to be used in order to obtain a high resolution structure.

As discussed further below, hen lysozyme has already been studied extensively by a variety of structural techniques and a number of high resolution crystal structures for the protein exist (Mason et al., 1984; Handoll, 1985). The results of the NMR structure determination could, therefore, be compared extensively with crystallographic data. In contrast, prior to the start of this work very little was known about the structure of human IL-4; indeed the project has been able to provide the first high resolution structure for this protein (Smith et al., 1992a).

Biologically, both proteins are of great interest, hen lysozyme as an enzyme which is being used as a model system for studies of catalysis and protein folding and dynamics, and human IL-4 as a member of the cytokine family which has important potential pharmaceutical properties. A brief review of these two proteins, their biological roles and their previous structural studies is given below.

1.1 Hen egg-white lysozyme.

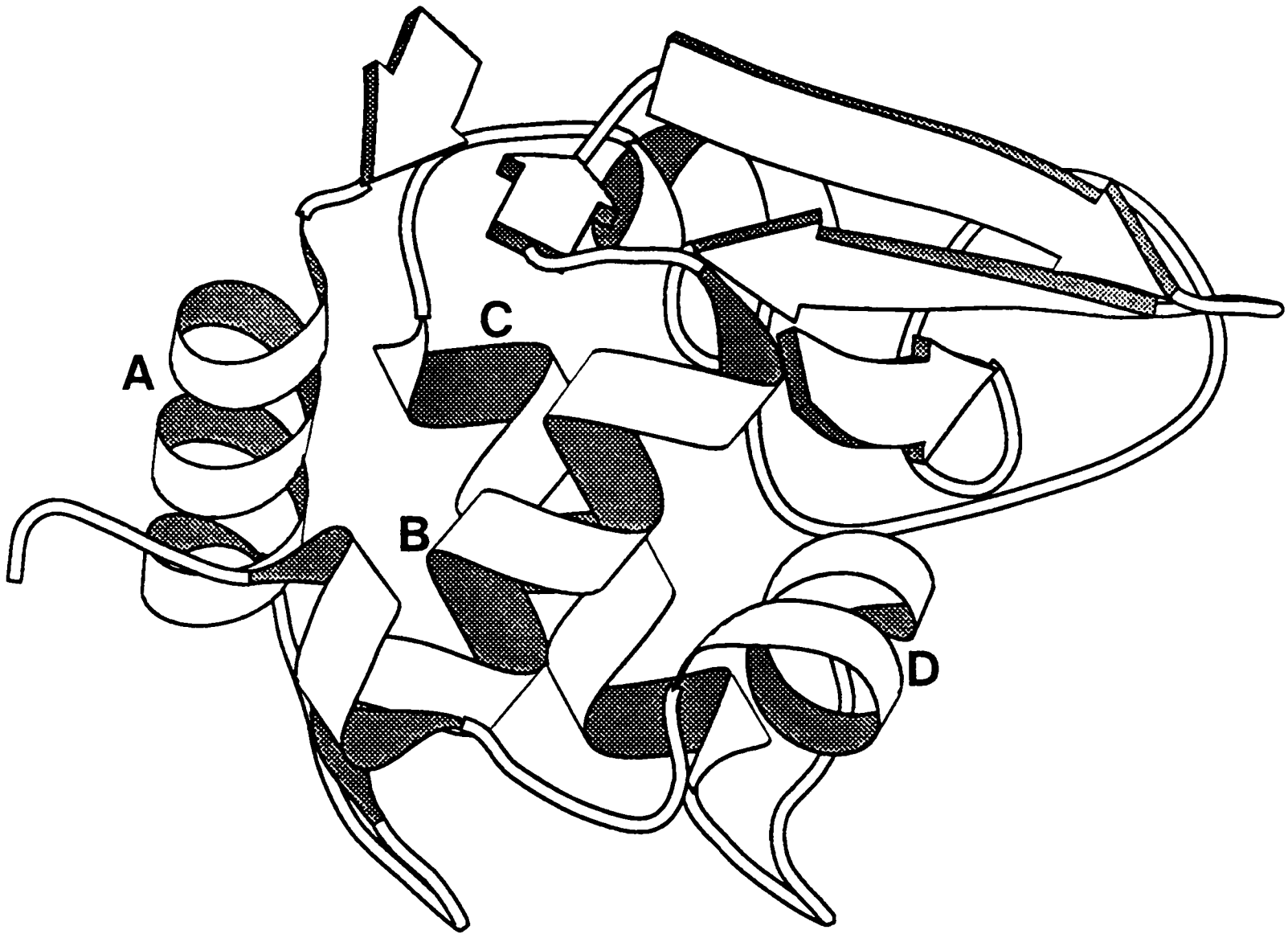
Hen lysozyme is an enzyme that acts as a glycoside hydrolase cleaving the glycosidic bond between C-1 of N-acetylmuramic acid and the C-4 of N-acetylglucosamine in the bacterial cell wall peptidoglycan (Imoto et al., 1972). It was the first enzyme to be submitted to a complete X-ray analysis, a 2 Å resolution structure of the protein being reported in 1965 (Blake et al., 1965).

Shortly after this Phillips proposed a mechanism for the action of this enzyme based on its structure (Phillips, 1966), and a model for the structure of the complex of hen lysozyme with its substrate hexasaccharide was also proposed on the basis of a hen lysozyme-$(GlcNAc)_3$ crystal structure (Blake et al., 1967a,b). Since this time, lysozyme has continued to be studied actively (Jollès & Jollès, 1984; McKenzie & White, 1991; Strynadka & James, 1991). It is still a system of great interest, particularly as the mechanism of the enzyme is still under debate (e.g. Post & Karplus, 1986) and as lysozyme is being used as a model protein in attempts to understand the principles of protein folding and dynamics in general.

The crystal structure of hen lysozyme has been extensively refined since 1965, and high resolution structures for a number of crystalline forms are now available; all have essentially the same conformation (Mason et al., 1984; Handoll, 1985; Strynadka & James, 1991). In addition, the protein fold has been found to be highly conserved in several different c-type lysozymes including human, turkey and tortoise lysozymes (Aschaffenburg et al., 1980; Artymiuk & Blake, 1981; Parsons, 1988). The protein has a globular fold and its structure can be divided into two lobes or domains with the active site of the enzyme being situated in the cleft between them; residues 1-35 and 85-129 form a domain with 4 α helices and a short 3_{10} helix near the C-terminus, while the other domain (residues 36-84) contains a long loop, a triple stranded antiparallel β sheet and a 3_{10} helix. Another short region of β sheet connects residues 2-3 to residues 39-40. The overall fold of the protein is shown schematically in Figure 1.1.

Lysozyme was also one of the first proteins to be studied in detail by NMR (McDonald & Phillips, 1967; Campbell et al., 1973) and since then it has been studied extensively (Redfield & Dobson, 1988; Petros & Kopple, 1990; Ohkubo et al., 1991; Ueda et al., 1991). The protein gives excellent 2D ^1H NMR spectra, and ^1H resonance assignments have been made for all the

Figure 1.1 A schematic diagram showing the main chain fold of the tetragonal type 2 crystal structure of hen lysozyme.



main chain and 70% of the side chain protons of the protein (Redfield & Dobson, 1988). Identification of short range NOE effects has also enabled the secondary structure of the protein in solution to be defined; the major elements seen in the crystal structures are preserved in solution (Redfield & Dobson, 1988). In addition to structural studies, NMR spectroscopy has also been used extensively to study the binding of substrates and inhibitors, and the dynamics and folding of the protein (Dobson & Evans, 1984; Dobson, 1991; Miranker et al., 1991; Pedersen et al., 1991; Lumb et al., 1992; Radford et al., 1992). The two structural domains of the protein have been found to be distinct folding domains, the helical domain folding in advance of the domain containing the β sheet (Miranker et al., 1991; Radford et al., 1992).

Theoretical methods have been used to probe the dynamics of hen lysozyme both in the presence and absence of substrate (by molecular dynamics simulations) and also to investigate the possibility that hen lysozyme may exhibit hinge bending (i.e. changes in the relative orientations of the two domains) (McCammon et al., 1976). Hinge bending has been found to occur in T4 lysozyme (Faber & Matthews, 1990), although little experimental evidence has been found to support its existence in hen lysozyme.

1.2 Human Interleukin-4.

Interleukin-4 was first identified in studies of murine B cell differentiation, where it was found to be capable of co-stimulating with anti-IgM antibody the proliferation of murine B cells (Howard et al., 1982). It was hence originally known as B cell stimulating factor-1. The human protein was later isolated on the basis of homology with the cDNA of murine IL-4 (Yokota et al., 1986). Mature human IL-4 has 129 residues and two N-glycosylation sites (Asn 38 and Asn 105. Le et al., 1988), although studies have shown that the carbohydrate is not required for biological activity (Park et al.,

1987). In the work presented here recombinant human IL-4 from *Escherichia coli* has been used; this protein has an N-terminal methionine initiator and is nonglycosylated, but it has also been shown to be biologically fully active (van Kimmenade et al., 1988).

IL-4 is a cytokine, one of a family of proteins that are coordinators of the immune and inflammatory responses (Balkwill & Burke, 1989; Arai et al., 1990). Other cytokines include all the interleukins, growth factors such as epidermal growth factor (EGF), interferons (e.g. IFN-α) and colony stimulating factors (e.g. granulocyte-macrophage colony-stimulating factor, GM-CSF). Studies of natural and recombinant cytokines have shown that most cytokines are pleiotropic and have multiple biological activities; those of human IL-4 are summarized in Table 1.1.

Table 1.1 Range of activities of human interleukin-4.

Promotes growth of T cells (Spits et al., 1987) B cells (DeFrance et al., 1987)
Enhances expression, on resting B cells, of class II major histocompatibility molecules low-affinity receptor for IgE (Rousett et al., 1988)
Induces synthesis of IgE (Vercelli et al., 1989) IgG and IgM by activated B cells (DeFrance et al., 1988)
Counteracts IL-2 induced proliferation of chronic lymphocytic leukemia B cells (Karray et al., 1988).

The effects of IL-4 are mediated by the binding of the cytokine to a receptor on the cell surface but little is known about the intracellular signal transduction pathway triggered within the cell by the binding. The receptor for IL-4 has been cloned; it has a 207 residue external domain (to which the IL-4 ligand binds), a short 24 residue transmembrane section and a 569 residue cytoplasmic domain (Idzerda et al., 1990). The receptor belongs to the recently recognised superfamily of haemopoietic cytokine receptors the members of which have significant homology in their extracellular domains (Bazan, 1990a, b; Cosman et al., 1990). Other members of this superfamily, which is discussed further in Chapter 7, include the receptors for growth hormone, prolactin, erythropoietin, GM-CSF, IL-2 (β-subunit), IL-3, IL-6 and IL-7.

At the start of this work very little was known about the structure of human IL-4. Preliminary crystallographic results had shown that the molecule forms well ordered crystals that diffract to beyond 2.8 Å resolution and are suitable for detailed diffraction analysis (Cook et al., 1991). In addition, CD studies of the protein in solution had shown that it has a high percentage of helical structure and has a high stability over a wide range of solvent conditions (Windsor et al., 1991). In the past year, however, the situation has radically changed. A structure prediction has been reported for IL-4 which proposes a right-handed four helix bundle fold for the protein (Curtis et al., 1991) and, subsequent to the publication of the work described in this thesis, a second NMR structure has been reported (Powers et al., 1992). In addition, a 2.25 Å resolution crystal structure for the protein has just been solved (Wlodawer et al., 1992).

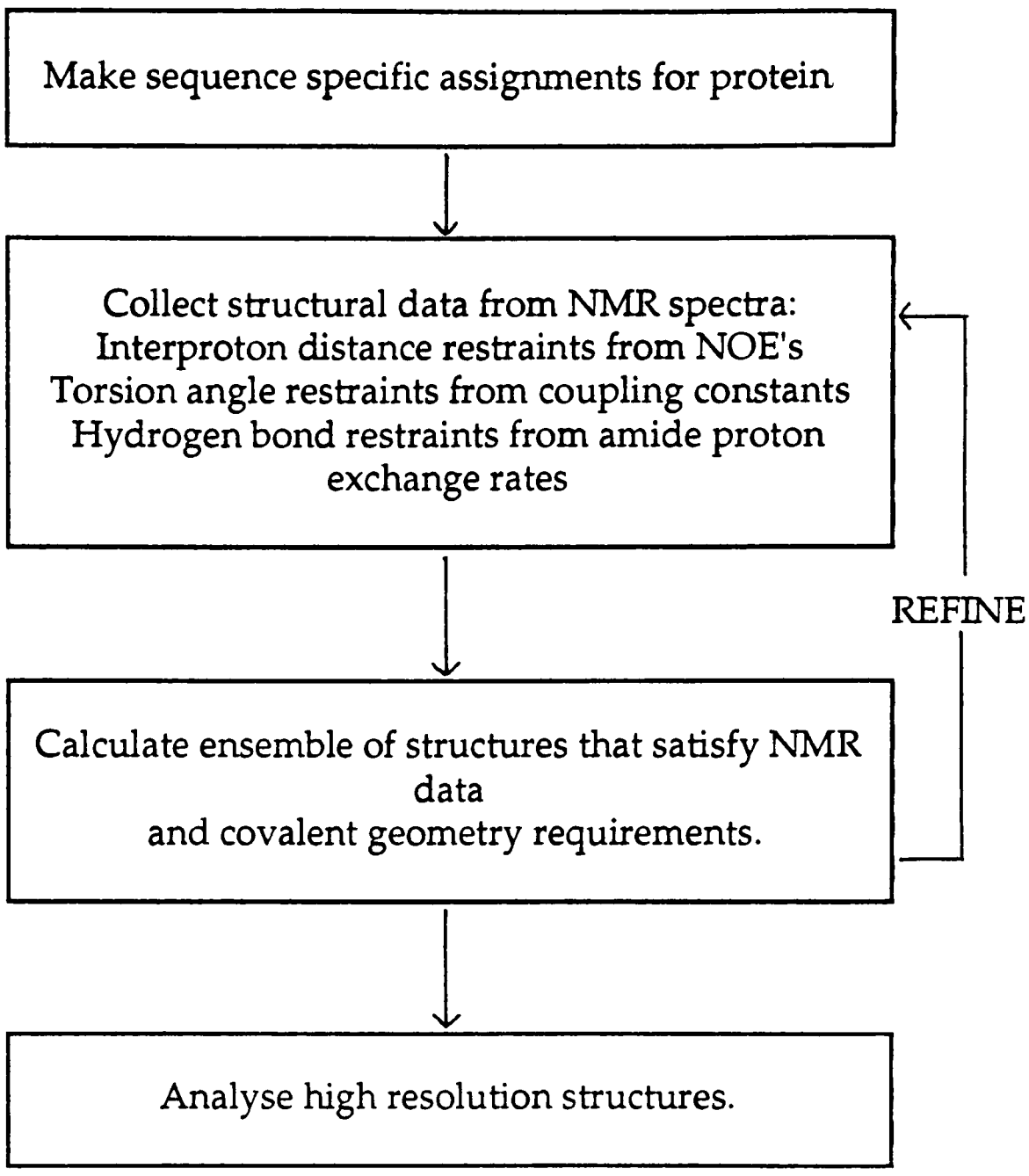
1.3 Outline of the thesis.

The basic procedure used in an NMR structure determination is summarized in Figure 1.2 (Wright, 1989). An initial requirement is that sequence specific assignments must be made for the great majority of the resonances of the protein. Structural information can then be collected from the NMR spectra and calculations performed which search conformational space and find structures which satisfy the NMR data set. The major structural information that is provided by NMR is the nuclear Overhauser effect (NOE), a relaxation effect that depends upon dipolar coupling between nuclear spins. At the limit of short mixing times the effect has an r^{-6} dependency on the distance (r) between the nuclei involved and so can be used to provide a set of distance restraints between pairs of protons in the protein. Additional structural restraints are obtained from spin-spin coupling constants which give information about dihedral angles (particularly ϕ and χ_1 torsion angles).

For hen lysozyme, sequence specific ^1H NMR assignments had been made prior to this project, so the work started with the collection of structural data from the 2D ^1H NMR spectra of the protein. This is described in Chapter 3 which also examines the structural data in the light of the crystal structures of the protein. In addition, this chapter discusses evidence for the conformational disorder of various side chains of the protein from $^3J_{\alpha\beta}$ coupling constant measurements. In Chapter 4 the computation of protein structures for lysozyme from the experimental data set is described. The ensemble of calculated structures is compared extensively both with the crystal structures and with other NMR data about the protein. Overall this work enables a comprehensive description of the structure of hen lysozyme in solution to be given (Smith et al., 1991, 1992b).

Chapters 5, 6 and 7 describe the work carried out on the protein IL-4 which gave the first three-dimensional structure for this protein, hence

Figure 1.2. Steps involved in the determination of the 3D structures of proteins in solution by NMR.



giving insight into the biological activities of this important cytokine. In particular, Chapter 5 gives details of the procedures used to determine the structure of human IL-4 using heteronuclear 3D NMR techniques on isotopically labelled samples of the protein. The resulting structures for IL-4 are then described in detail in Chapter 6, and Chapter 7 discusses the significance of the IL-4 structure and compares it with a structure prediction (Curtis et al., 1991) and with proteins with related folds.

A postscript to this thesis gives an initial comparison of the NMR structure of IL-4 with the new crystal structure for the protein (Wlodawer et al., 1992).

CHAPTER 2.

MATERIALS AND METHODS.

2.1 Sample source and preparation.

Hen lysozyme was obtained from the Sigma Chemical Co. as a three times crystallised, dialysed and lyophilised powder; prior to use it was dialysed extensively against distilled water at pH 3.0. Samples of the protein were prepared in a number of ways so as to exploit the range of amide proton exchange rates observed in the protein and hence gain spectral simplification (Redfield & Dobson, 1988). Non exchanged lysozyme was prepared by dissolving the protein in 90%H_2O/10%D_2O while fully exchanged lysozyme was prepared by dissolving the protein in 99.9% D_2O and heating at 80°C for 10 min. The spectra of these samples contained resonances from all and from none of the amide protons respectively. Partly exchanged lysozyme was prepared by freshly dissolving the protein in 99.9% D_2O; the spectrum was recorded immediately and only contained resonances from slowly exchanging amide protons. Samples of reverse-exchanged lysozyme were prepared by redissolving a lyophilised fully exchanged sample in 90%H_2O/10%D_2O. In this case the spectrum only contained resonances from rapidly exchanging amide protons. In all cases the samples contained 7mM protein at pH 3.8, and spectra were recorded at temperatures between at 35°C and 55°C.

Fully enriched isotopically labelled samples of recombinant human interleukin-4 were produced by SmithKline Beecham Pharmaceuticals. Further details are given in Chapter 5. Sample of the protein for the NMR experiments were prepared by dissolving the protein in 90%H_2O/10%D_2O or in 99% D_2O and adjusting to pH 4.5 or 5.3. All samples were approximately 2mM in protein, and spectra were recorded at temperatures of 20°C and 35°C.

2.2 NMR Spectroscopy.

All the NMR experiments were performed on the home-built NMR spectrometer of the Oxford Centre for Molecular Sciences which operates at 500.1 MHz for ^{1}H, 50.7 MHz for ^{15}N and 125.7 MHz for ^{13}C. The spectrometer is equipped with an Oxford Instruments Co. superconducting magnet, GE/Nicolet software and digital control equipment, and a Bruker reverse probehead.

For hen lysozyme, double quantum filtered correlation spectroscopy (DQF COSY) (Rance et al.,1983), exclusive correlation spectroscopy (E.COSY) (Griesinger et al.,1987), and nuclear Overhauser enhancement spectroscopy (NOESY) (Jeener et al.,1979; Anil Kumar et al.,1980) experiments were performed according to the method of States et al.(1982) with standard phase cycling schemes. Data sets consisting of 512 t_1 increments of 2K complex data points were collected for the DQF COSY and E. COSY experiments using a sweep width of 5405.4 Hz in both cases. For the NOESY experiments, 256 t_1 increments of 1K complex data points were collected with a sweep width of 7042 Hz in both dimensions. Mixing times of 50, 100, 200 and 500 ms were used, randomly varied by 10% (Macura et al.,1981).

The data were processed on a Satellite data processing station using GEN software and were plotted using a Zeta plotter. All spectra were resolution enhanced in t_2 by trapezoidal and double exponential multiplication and in t_1 by trapezoidal multiplication, and were zero filled once in t_2 and twice in t_1. After zero filling the digital resolution of the E. COSY and DQF COSY spectra in F2 was 1.32 Hz/point while the digital resolution of the NOESY spectra was 3.4 and 6.9 Hz/point in F2 and F1, respectively.

For the double labelled sample of human IL-4, 3D ^{13}C-^{1}H HCCH-COSY, ^{13}C-^{1}H HCCH-TOCSY and ^{13}C-^{1}H NOESY-HMQC experiments were recorded (Bax et al., 1990b; Ikura et al., 1990b; Kay et al., 1990). These experiments are

described further in Appendix I. For the HCCH-COSY experiment a data set of 128 x 32 x 512 complex points was collected using sweep widths of ± 2325.6 Hz, ± 1388.9 Hz and ± 2702.7 Hz and acquisition times of 27.52 ms, 11.52 ms and 97.72 ms in the t_1, t_2 and t_3 dimensions respectively. The HCCH-TOCSY was acquired as 128 x 32 x 1024 complex points with spectral widths of ± 2325.6 Hz, ± 1388.9 Hz and ± 5434.8 Hz and acquisition times of 27.52 ms, 11.52 ms and 94.21 ms for t_1, t_2 and t_3. An isotropic mixing time of 23.4 ms was used in this experiment. For the NOESY-HMQC 128 x 60 x 1024 complex points were acquired with sweep widths of ± 2325.6 Hz, ± 2777.7 Hz and ± 5434.8 Hz and acquisition times of 27.52 ms, 11.52 ms and 94.21 ms in the t_1, t_2, t_3 dimensions respectively. A NOESY mixing time of 125 ms was used. The increased sweep width and number of points in t_3 in the HCCH-TOCSY and NOESY-HMQC experiments resulted in a flatter baseline. During the processing of these two experiments only the central 1024 points, which contained all the regions of interest in the spectra, were stored in t_3 after zero filling. Each 3D spectrum was recorded over a period of about 3 days.

The 3D NMR data sets were processed on a Sun Sparc station using the FELIX software package (Hare Research Inc.). The spectra were processed first in t_3, then in t_1 and finally in t_2. In all cases the spectra were zero filled once in each dimension. All spectra were resolution enhanced with double exponential multiplication and a 45°-shifted sine bell function in t_3, with a 45°-shifted sine bell function in t_1 and with a 60°-shifted sine bell squared function in t_2.

2.3 Protein coordinate sets.

Crystal structure coordinates for hen lysozyme were taken from the 2.0 Å resolution tetragonal type 2 crystal structure obtained by X-ray diffraction (Blake et al., 1967b; Handoll, 1985) and the 1.4 Å resolution triclinic crystal structure obtained by neutron diffraction and provided by Dr. S. Mason (Grenoble). The R factors for these structures are 0.22 and 0.18 respectively.

For comparison with the NMR structure of human interleukin-4, the coordinates for the structure prediction of human IL-4 (Curtis et al., 1991) were provided by Prof. F. E. Cohen (University of California, San Francisco), and the 2.4 Å resolution crystal structure coordinates of GM-CSF (Diederichs et al., 1991) were provided by Prof. P. A. Karplus (Cornell University). The 2.25 Å resolution crystal structure coordinates of human IL-4 (Wlodawer et al., 1992) have been recently provided by Dr. A. Wlodawer (NCI-Frederick Cancer Research and Development Center, Maryland).

2.4 Calculation of NMR parameters.

2.4.1 Coupling constants.

Predicted coupling constant values were calculated from torsion angles using the equation (Karplus, 1959):

$$^3J = A\cos^2\theta + B\cos\theta + C$$

where $\theta = \phi - 60^o$ for $^3J_{HN\alpha}$, $\theta = \chi_1 - 120^o$ for $H\beta_2$ and $\theta = \chi_1$ for $H\beta_3$ for $^3J_{\alpha\beta}$ ($H\beta_2$ and $H\beta_3$ are defined according to the IUPAC-IUB conventions (1970)). Values for A, B and C of 6.4, -1.4 and 1.9 for $^3J_{HN\alpha}$ (Pardi et al.,1984) and 9.5, -1.6 and 1.8 for $^3J_{\alpha\beta}$ (de Marco et al.,1978) were used. It is well recognised, however, that the electronegativity of substituents affects the observed coupling constant values. In proteins this is particularly important for the $^3J_{\alpha\beta}$ values of the amino acid residues serine and threonine where the high electronegativity of the oxygen substituent leads to a reduction in the observed values. There have been various methods proposed to take this

into account. For example, Kopple et al. (1973) suggested the calculated $^3J_{\alpha\beta}$ values for serine and threonine should be divided by 1.08, whilst Haasnoot et al. (1980) have determined coefficients for an equation involving terms incorporating the electonegativity of the substituents. It is difficult to assess which method gives the best correlation as the extent of time averaging of the coupling constants due to small fluctuations about a single χ_1 is not known. However, from a comparison of the calculated and experimental $^3J_{\alpha\beta}$ values in this work, the division of the calculated values for serine and threonine by 1.08 appears to give a simple but adequate correction.

2.4.2 Chemical shift index.

The chemical shift index of each residue of the protein was determined from the Cα proton resonance assignments by the method of Wishart et al. (1992). This method enables prediction of the type and location of secondary structure elements in the protein. Using the range of chemical shift values given in Table 2.1, each residue is assigned a chemical shift index of +1 if the Cα proton chemical shift is greater (more positive) than the range, -1 if the Cα proton chemical shift is less (more negative) than the range, and 0 if the Cα chemical shift is within the chemical shift range for the relevant residue type. From the chemical shift indices secondary structure can then be identified, a dense grouping of four or more -1's not interrupted by a +1 indicating a helix, and a dense grouping of three or more +1's not interrupted by a -1 indicating a β strand.

Table 2.1 Ranges of chemical shift values of Cα protons used in the determination of chemical shift indices (from Wishart et al., 1992).

Residue	Cα ^{1}H range (ppm)	Residue	Cα ^{1}H range (ppm)
Ala	4.35 ± 0.10	Met	4.52 ± 0.10
Cys	4.65 ± 0.10	Asn	4.75 ± 0.10
Asp	4.76 ± 0.10	Pro	4.44 ± 0.10
Glu	4.29 ± 0.10	Gln	4.37 ± 0.10
Phe	4.66 ± 0.10	Arg	4.38 ± 0.10
Gly	3.97 ± 0.10	Ser	4.50 ± 0.10
His	4.63 ± 0.10	Thr	4.35 ± 0.10
Ile	3.95 ± 0.10	Val	3.95 ± 0.10
Lys	4.36 ± 0.10	Trp	4.70 ± 0.10
Leu	4.17 ± 0.10	Tyr	4.60 ± 0.10

2.4.3 Ring current shifts.

Ring current shift calculations were performed using the program VNMR (Hoch et al., 1982; Hoch, 1983; Redfield & Dobson, 1990), the shifts being calculated by the method of Johnson and Bovey (1958) using the ring current intensity factors of Giessner-Prettre & Pullman (1969). Experimental secondary shifts were calculated by subtracting the observed chemical shifts from the random coil chemical shift values obtained from tetrapeptides (Wüthrich, 1986).

2.5 Calculation of protein structures.

Protein structure calculations were performed on a Convex C210 mini supercomputer using a hybrid distance geometry-dynamical simulated annealing approach (Nilges et al., 1988). The calculations used a distance geometry program (Havel & Wüthrich, 1984) from Dr. R. Scheek, Groningen

and the program XPLOR (Brünger, 1988). The protocol, which is described in Table 2.2, is based on that of Sutcliffe & Dobson (1991).

The distance geometry stage of the structure calculations used only about one third of the atoms in the full protein structure hence reducing the CPU time required. After optimization, the full side chains were then fitted onto the substructure using a least squares procedure to fit the N, Cα and C atoms. The side chains were positioned on the basis of the corresponding dihedral angles in the substructure where possible, the remaining atoms being built in the most probable conformer for the residue in question (Sutcliffe et al., 1987).

In the dynamical simulated annealing the non-bonded interactions in the protein were represented in the potential energy function by a single van der Waals' repulsive term (Nilges et al., 1988). The initial high temperature and low value of k_{REPEL} (force constant for repulsive term) used in the protocol enabled major structural rearrangements, including regions of the structure passing through each other, to occur when necessary. As the dynamics continued and the violations of the experimental restraints reduced, k_{REPEL} was increased in order to ensure that there were no close bonded contacts. The system was then gradually cooled to 300K and the structure energy minimized. The dynamical simulated annealing had a tendency to produce expanded structures due to the fact that the only non-bonded attractive forces were generated by the NOE's. This was overcome to a large extent by the full potential dynamics stage of the protocol (the simple repulsive term is replaced by full van der Waals', electrostatic and hydrogen bond terms).

Table 2.2 Hybrid distance geometry-dynamical simulated annealing protocol used in this work.

1. Distance geometry calculations generate substructure containing only N, C, Cα, Cβ, non-terminal Cγ and Cδ atoms and pseudoatoms for aromatic rings.

2. Structure optimized and the full side chains built on to the substructure. Then energy minimized to ensure correct covalent geometry.

3. Simulated annealing performed on full structure.

a) 1.5 ps of dynamics at 1000K.

High values for the NOE and phi angle restraining force constants (k_{NOE} = 50 kcal/mol $Å^2$, k_ϕ = 200 kcal/mol rad^2). Low value for k_{REPEL} (0.01 kcal/mol $Å^{4)}$. High force constants for covalent geometry (k_{BOND} =500 kcal/mol $Å^2$, k_{ANGLE} =500 kcal/mol rad^2, k_{IMPR}=500 kcal/mol rad^2). Mass of all atoms 10 amu.

b) 3 ps of dynamics at 1000K.

k_{REPEL} increased from 0.01 kcal/mol $Å^4$ to 10 kcal/mol $Å^4$ by a factor of $1000^{1/40}$ in 40 steps of 0.075 ps. Covalent force constants relaxed to those in XPLOR parameter set PARMALLH3X. Atomic masses set to correct values. k_{NOE}= 10 kcal/mol $Å^2$, k_ϕ = 20 kcal/mol $Å^2$

c) 1.4 ps of steady cooling from 1000K to 300K followed by 200 steps of Powell minimization.

4. Full potential dynamics (i.e. full van der Waals', electrostatic and hydrogen bonding potentials)

3 ps of the dynamics at 1000K; 2.8 ps slow cooling to 300K; 3 ps of dynamics at 300K.

5. Final structures obtained by averaging over the last 2ps of the trajectory at 300K and minimizing the structure using 200 steps of restrained Powell minimization.

Each structure took approximately 5 h of CPU time to generate. For hen lysozyme 20 structures were calculated and the 16 of lowest energy were selected, while for human IL-4 for each set of calculations an ensemble of 15 structures were calculated and the 10 of lowest energy were selected. An average coordinate set was produced from an ensemble of calculated structures by best fit superimposing the set of structures (N, C, Cα, Cβ used in superposition) and averaging their coordinates. These averaged coordinates then underwent a two step restrained Powell minimization.

Protein structures were displayed on a Silicon graphics Personal Iris using the program INSIGHT (Biosym). Ribbon diagrams were generated using the program MOLSCRIPT (Kraulis, 1991).

CHAPTER 3.

THE STRUCTURE OF HEN LYSOZYME IN SOLUTION.

NMR data collection and comparison with crystal structures.

The NMR spectra of a protein contain a wealth of information about the structure and dynamics of the protein but many of the parameters, such as chemical shifts and line widths, are difficult to interpret without detailed structural models. Two sources of information, however, the nuclear Overhauser effect (NOE) and three bond spin-spin coupling constants (particularly $^{3}J_{HN\alpha}$ and $^{3}J_{\alpha\beta}$) can be readily converted into structural restraints and these data form the basis of an NMR structure determination. In this chapter the collection of these structural data for hen lysozyme from 2D ^{1}H spectra of the protein and their conversion into structural restraints are presented. The experimental data are then compared with two high resolution structures of the protein (2.0 Å resolution tetragonal type 2 and 1.4 Å resolution triclinic) to assess how well these structures represent the structure of the protein in solution. This comparison has been performed in two different ways. For the coupling constants, the torsion angles from the crystallographic coordinates are used to predict coupling constant values (using the Karplus relationship) and the experimental and calculated values are compared. In contrast, the NOE data are converted into interproton distance restraints which are then compared with the interproton distances in the crystal structures. This method is only approximate as it neglects the effects of multispin relaxation or differences in correlation time on the observed NOE intensities as is discussed in more detail later, but it does provide a simple method of comparison provided the limitations of the approach are recognised.

3.1 NMR structural data for hen lysozyme.

3.1.1 NOE distance restraints.

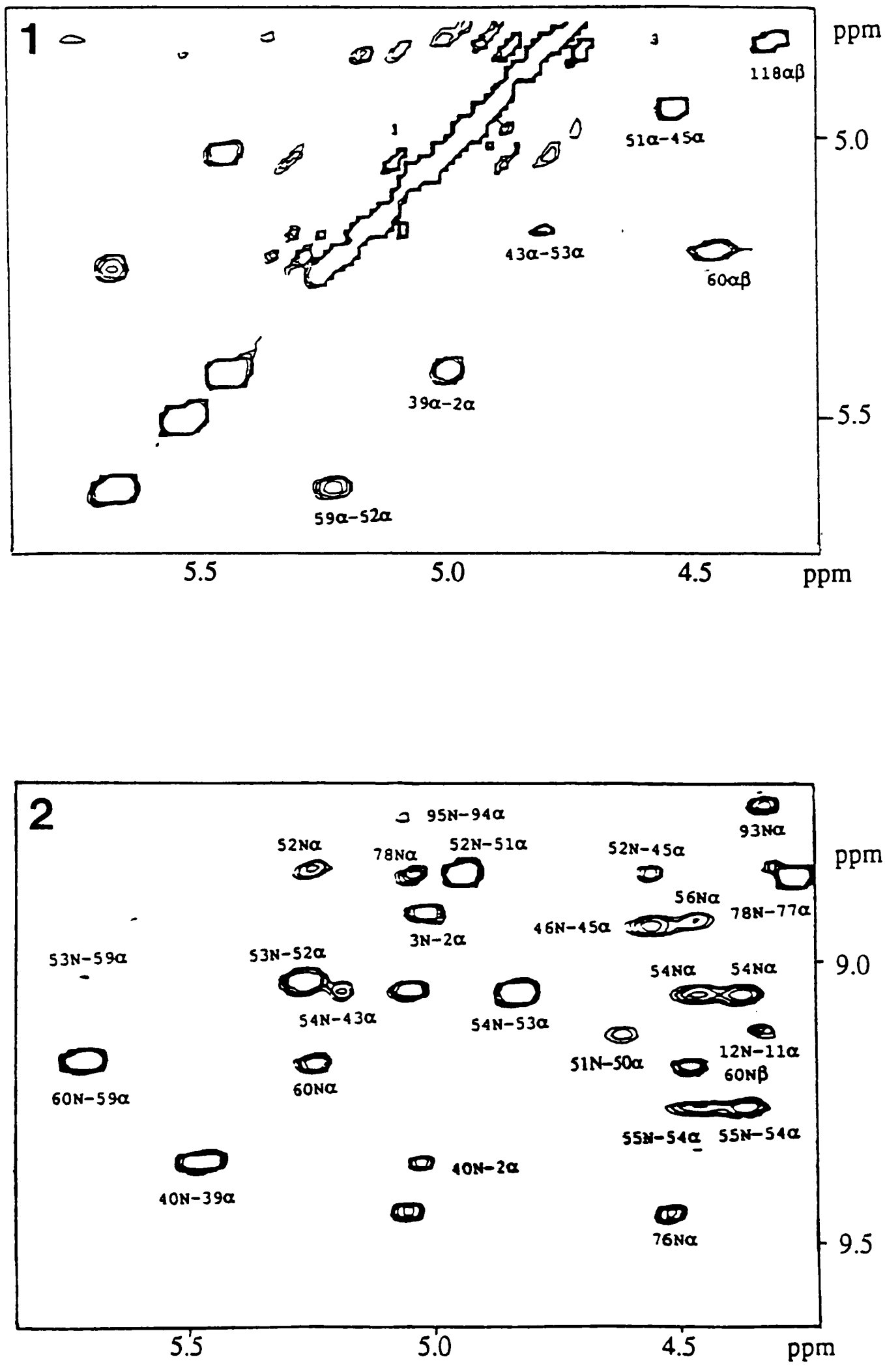
NOESY spectra were recorded with mixing times of between 50 ms and 500 ms. Study of the build up of NOE intensities (intensities estimated by measuring peak heights) showed clear evidence for spin diffusion as the mixing time increased; the major set of distance restraints was therefore derived from NOESY spectra recorded with a short 50 ms mixing time. The spectra were analysed using the sequence specific assignments for hen lysozyme that have been made previously (Redfield & Dobson, 1988), including stereospecific assignments for 23 residues (Smith et al., 1991). To overcome overlap and ambiguity problems in the analysis, the spectra were simplified by exploiting the range of amide proton exchange rates observed for lysozyme. Use of samples of protein freshly dissolved in D_2O (whose spectra contain resonances from the slowly exchanging amide protons) and fully exchanged in D_2O and then redissolved in 90% H_2O/ 10% D_2O (whose spectra contain resonances from rapidly exchanging amide protons (Redfield & Dobson, 1988)), in addition to nonexchanged samples (freshly dissolved in 90%H_2O/10%D_2O) and samples fully exchanged in D_2O, enabled many ambiguities to be solved and a preliminary set of ~800 NOE's was collected[1]; an example of the spectra used is given in Figure 3.1 This set of data was used to calculate low resolution structures which were then used to solve many of the remaining ambiguities in the NOESY cross peak assignments in an iterative manner (Klevit et al., 1990). By these methods a total of 994 NOE cross peaks from spectra recorded with a 50 ms mixing time were identified. The NOESY cross peaks were classified as strong, medium or weak by counting the number of contour levels in the spectral plots.

---

1 Preliminary work on the analysis of the NOESY spectrum of lysozyme and the measurement and analysis of $^3J_{\alpha\beta}$ coupling constants of the protein was presented in L. J. Smith, Part II thesis, University of Oxford.

Figure 3.1.

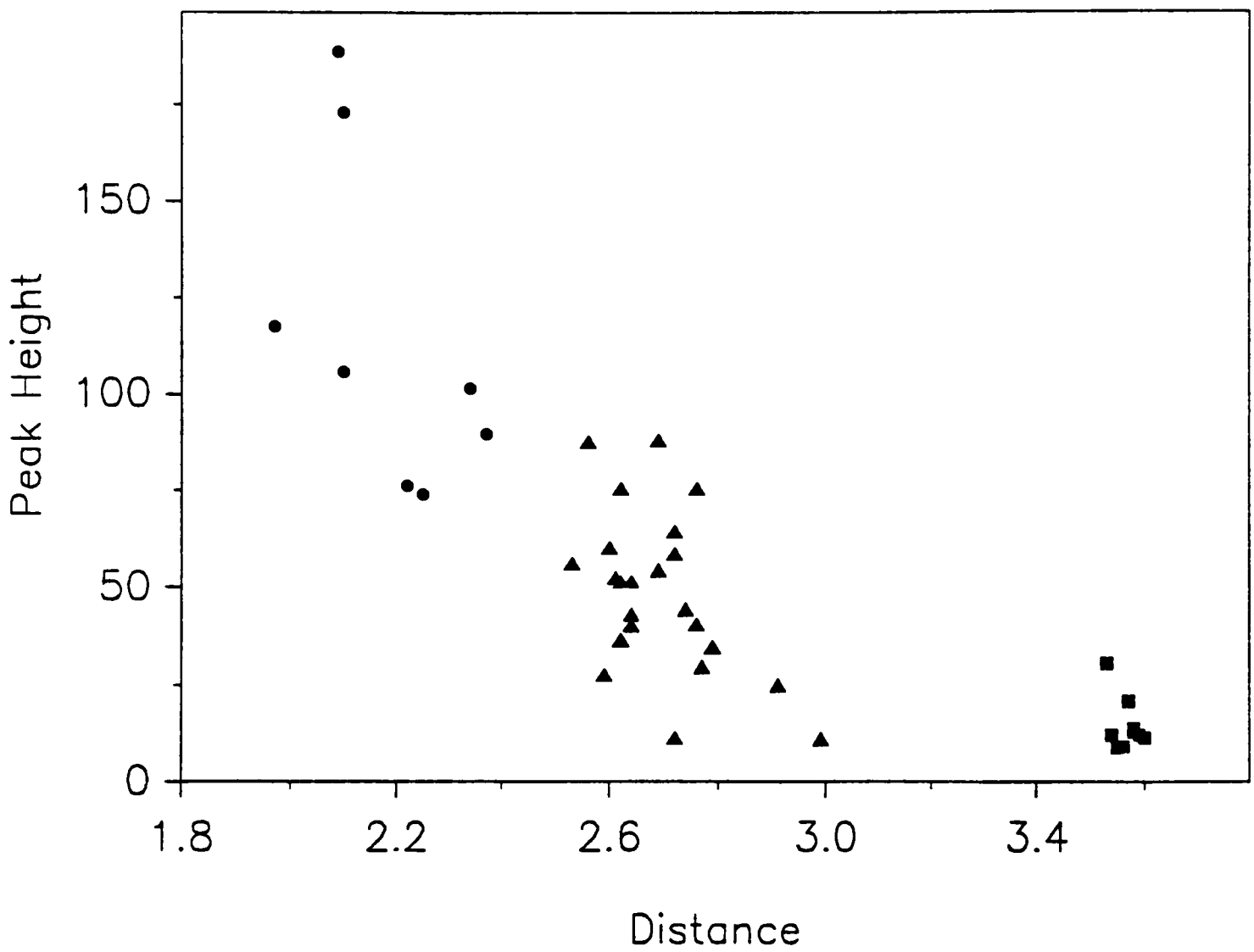
Regions of the phase-sensitive 500-MHz NOESY spectrum of partly exchanged lysozyme recorded with a 100ms mixing time showing cross peaks arising from the triple stranded antiparallel β sheet. A shows CαH-CαH cross peaks and B CαH-NH cross peaks.



In order to define the limits on interproton distance restraints the variation in the intensities of NOE's corresponding to distances between specific main chain protons in secondary structure elements was examined. This variation is shown in Figure 3.2; there is more than a three-fold range in intensity of the NH-NH NOE's in the four α helices in lysozyme and similar variations are found when other secondary structure NOE's are examined (e.g CαH-NH NOE's both in the β sheet and in the α helices). Indeed in a small number of cases variations of up to a factor of ten are seen. Some of these differences result from the fact that only peak heights, which are affected by line widths and coupling constants, rather than volume integrals were used as a measure of intensity in the crowded 2D spectra. Examination of the integrals of various resolved peaks indicated, however, that this was not the dominant contribution to the variation except for resonances with very broad lines. An additional factor is that resolution enhancement techniques have been used in the processing of the NOESY spectra; this causes significant intensity differences in the spectra for resonances with differing line widths even if volume integrals are used. Other contributions to the differences could arise from the effects of spin diffusion, even at a short 50 ms mixing time, for a protein the size of lysozyme, and from irregularities or the presence of some internal motion in the secondary structure elements. Experiments suggest that the effects of internal motions on cross-relaxation rates involving side chains may be significant even in the interior of lysozyme (Olejniczak et al., 1981), although for the main chain these effects are expected to be smaller (Post et al.,1989; Post, 1992). A variation of up to a factor of three for NOE intensities at a given distance represents, however, only an approximate 10% variation in distance. It was therefore still possible to select distance restraints for the three categories of cross peaks: 1.8-2.5 Å (strong), 1.8-3.0 Å (medium) and

Figure 3.2.

Plot of the distances (in Å) between specific main chain protons in secondary structure elements in the tetragonal type 2 crystal structure of hen lysozyme versus the peak height (in arbitrary units), for the NOE's corresponding to these distances, in a 50 ms NOESY spectrum. Only NOE's that are well resolved in the spectra have been included. A circle indicates sequential CαH-NH NOE's in β sheet, a triangle sequential NH-NH NOE's in α helices and a square sequential CαH-NH NOE's in α helices.



1.8-4.5 Å (weak). For the NOE's involving main chain distances in the secondary structure of the protein, this categorization placed 85 % of the observed effects in the correct distance range. For the other 15 % of the NOE's the intensities were, with one exception, less than expected.

An additional set of 164 NOE distance restraints was obtained from NOESY spectra recorded with longer mixing times of 200 and 500 ms. Weak distance restraints of 1.8-5.5 Å (for cross peaks first seen in spectra recorded with a 200 ms mixing time) and 1.8-7.5 Å (for cross peaks only seen in spectra recorded with a 500 ms mixing time) were used, so as to take the possibility of extensive spin diffusion into account; 5.5 and 7.5 Å were estimated to be the maximum interproton separations for which an NOE would be seen at these longer mixing times.

For all distance restraints pseudoatoms were used where no stereospecific assignments had been made, the necessary correction being added to the distance range (Wüthrich et al., 1983). In addition, 0.5 Å was added to the upper limits of the distance restraints when the NOE involved protons of a methyl group (Wüthrich, 1986). The total list of the 1158 NOE's identified for hen lysozyme is given in Appendix III.

3.1.2 Coupling constants and dihedral angle restraints.

For a protein the size of lysozyme the line widths of ^1H NMR resonances are largely in excess of 5 Hz, and frequently comparable to the value of the three-bond spin-spin coupling constants. This means that accurate values for coupling constants cannot be obtained directly from the splitting in conventional 2D COSY spectra as the observed splittings between anti-phase components becomes increasingly different from the true splittings with increasing line widths (Neuhaus et al.,1985). For the $^3J_{HN\alpha}$ coupling constants of hen lysozyme the true values of the coupling constants were therefore extracted from anti-phase splittings in COSY cross sections by

the use of spectral simulations (Smith et al., 1991). The values are listed in Table 3.1; these measurements were made by Dr. C. Redfield.

Table 3.1 $^3J_{HN\alpha}$ coupling constants for hen lysozyme, ϕ torsion angles from the tetragonal type 2 and triclinic crystal structures and minimum calculated ϕ differences between experimental and crystal data.[1]

Residue	$^3J_{HN\alpha}$ (Hz)	ϕ in tetragonal	ϕ in triclinic	ϕ_{exp}-ϕ_{tet}	ϕ_{exp}-ϕ_{tric}
Val 2	10.0	-102.7	-105.2	-17.3	-14.8
Phe 3	7.4	-76.4	-79.3	-9.1	-6.2
Gly 4	6.1, 8.0	-93.7	-77.3		
Cys 6	5.8	-66.4	-74.1	-6.4	1.3
Glu 7	4.5	-59.6	-58.3	-2.9	-4.2
Leu 8	5.5	-67.6	-71.3	-2.9	0.8
Ala 9	3.7	-55.7	-50.5	0.0	-5.1
Ala 10	3.9	-65.8	-69.4	8.4	12.0
Ala 11	4.8	-70.6	-67.8	5.6	2.9
Met 12	4.6	-67.2	-64.8	3.9	1.5
Lys 13	4.2	-58.4	-69.0	-1.6	9.0
Arg 14	4.4	-61.5	-69.7	-0.2	8.0
His 15	9.2	-88.7	-87.2	-16.0	-17.5
Gly 16	6.1, 6.2	81.1	72.9		
Leu 17	7.6	-82.0	-84.9	-5.2	-2.3
Asp 18	5.7	-64.7	-65.3	-7.3	-6.7
Asn 19	7.0	69.4	70.3	-9.4	-10.3
Tyr 20	5.5	-61.8	-70.0	-8.6	-0.4
Arg 21	6.8	60.4	51.3	7.2	1.1
Gly 22	6.0, 6.8	85.4	79.8		
Tyr 23	8.6	-99.0	-109.8	2.2	12.9
Gly 26	3.3, 6.2	-58.2	-57.9		
Asn 27	5.4	-65.7	-58.9	-3.9	-10.8
Trp 28	6.0	-61.0	-67.3	-13.3	-7.0
Val 29	5.9	-68.1	-67.9	-5.4	-5.6
Cys 30	3.8	-32.1	-61.4	5.5	4.9
Ala 31	3.8	-60.6	-63.6	4.1	7.1
Ala 32	4.8	-62.4	-63.4	-2.5	-1.6
Lys 33	3.6	-54.3	-60.1	-0.4	5.4
Phe 34	7.6	-78.3	-89.3	-8.9	2.1
Glu 35	6.4	-80.1	-79.9	2.7	2.5
Ser 36	8.3	-135.5	-134.7	-10.8	-11.6
Phe 38	6.3	69.4	68.7	9.6	10.3
Asn 39	8.8	-100.4	-93.0	1.2	-6.1
Thr 40	5.4	-66.8	-67.3	-2.9	-2.4
Gln 41	9.2	-88.6	-96.9	-16.1	-7.7

Ala 42	4.5	-62.5	-62.9	0.0	0.4
Thr 43	9.3	-139.1	-129.5	5.4	-4.2
Asn 44	9.4	-154.0	-126.2	22.1	-5.7
Arg 45	7.7	-93.6	-91.1	5.5	3.0
Asn 46	8.8	-105.7	-92.2	6.5	-7.0
Thr 47	4.4	-62.9	-80.7	1.2	19.0
Asp 48	7.7	-74.8	-87.8	-13.3	-0.3
Gly 49	5.3, 8.0	91.3	91.3		
Ser 50	7.8	-82.9	-97.2	-6.1	8.2
Thr 51	9.8	-139.6	-131.3	19.6	11.3
Asp 52	9.6	-100.8	-98.6	-12.4	-14.6
Tyr 53	9.6	-126.6	-177.1	-0.2	3.9
Leu 56	9.7	-111.1	-113.5	-8.9	-6.5
Gln 57	6.3	45.2	54.5	-4.2	-13.5
Ile 58	8.0	-78.2	-84.9	-12.6	-5.9
Ser 60	5.1	-74.5	-83.1	7.2	15.8
Arg 61	6.2	-69.5	-75.9	-6.4	0.0
Cys 64	8.8	-151.8	-129.8	11.0	-11.0
Asn 65	9.4	-103.0	-98.3	-5.1	-9.8
Asp 66	10.0	-166.9	-109.2	-3.1	-10.8
Gly 67	5.5, 6.2	65.4	85.2		
Arg 68	9.7	-139.5	-115.4	19.5	-4.6
Thr 69	9.3	-120.3	-98.2	-13.7	-8.0
Gly 71	5.7, 5.9	54.1	73.7		
Cys 76	8.8	-88.6	-86.7	-10.6	-12.5
Asn 77	7.4	52.9	60.4	7.1	-0.4
Ile 78	8.0	-152.9	-145.4	3.7	-3.8
Cys 80	3.6	-53.0	-59.4	-1.6	4.8
Ser 81	3.6	-60.6	-68.6	6.0	13.9
Ala 82	5.4	-63.4	-63.8	-6.2	-5.9
Leu 83	7.2	-84.6	-86.4	0.8	2.6
Leu 84	9.2	-104.4	-107.4	-0.2	2.7
Ser 85	5.8	-72.4	-72.2	-0.4	-0.5
Ser 86	5.8	-64.5	-72.0	-8.3	-0.7
Asp 87	8.9	-84.6	-82.3	-15.8	-18.1
Ile 88	6.5	-76.0	-86.9	-2.2	8.7
Ala 90	4.2	-60.7	-62.7	0.7	2.7
Ser 91	5.45	-62.7	-68.6	-7.7	-1.9
Val 92	5.6	-60.9	-67.9	-10.3	-3.4
Asn 93	4.4	-59.8	-57.4	-1.8	-4.3
Cys 94	6.3	-74.1	-71.1	-2.5	-5.5
Lys 96	4.4	-56.7	-64.2	-5.0	2.5
Lys 97	6.5	-67.6	-65.8	-10.6	-12.4
Val 99	5.2	-60.7	-72.4	-7.4	4.3
Asp 101	7.0	-65.4	-94.7	-16.8	12.5
Gly 102	6.2, 6.4	161.5	67.0		
Asn 103	8.2	-113.1	-107.7	20.4	14.9
Gly 104	3.9, 6.4	57.5	-83.6		
Met 105	7.4	-70.5	-77.2	-15.0	-8.3

Ala 107	4.2	-60.7	-55.7	0.7	-4.3
Trp 108	9.6	-108.0	-93.9	-5.2	-19.3
Val 109	4.0	-48.3	-61.1	-10.0	2.8
Trp 111	7.1	-69.3	-61.0	-13.7	-22.0
Arg 112	4.5	-58.9	-62.9	-3.6	0.4
Asn 113	5.8	-83.0	-89.9	10.3	17.2
Arg 114	9.6	-121.8	-127.8	-5.0	1.0
Cys 115	9.8	-119.0	-118.8	-1.0	-1.2
Gly 117	6.2, 6.5	73.2	84.7		
Thr 118	9.8	-107.6	-90.8	-12.4	-29.2
Asp 119	6.7	-85.7	-79.4	5.9	-0.4
Val 120	4.6	-60.0	-58.4	-3.3	-4.9
Gln 121	5.0	-59.1	-53.9	-7.4	-12.6
Ala 122	3.7	-60.7	-53.9	5.1	-1.7
Trp 123	5.4	-65.9	-70.1	-3.7	0.4
Ile 124	10.6	-114.6	-113.0	-5.4	-7.0
Arg 125	4.4	-62.7	-68.1	1.0	6.5
Gly 126	5.6, 6.8	81.5	81.4		
Cys 127	7.7	-85.0	-101.0	-3.1	12.9
Arg 128	8.0	-95.1	-71.8	4.3	-19.0
Leu 129	9.0	-124.0	-165.5	-14.3	27.2

[1] The coupling constants values given are the average of the values at 45 and 55°C (in Hz). The torsion angles are given in degrees.

For the measurement of $^3J_{\alpha\beta}$ coupling constants, for residues with an AMX type spin system (residues where the two protons attached to the β carbon are not spin-spin coupled to any protons in the γ group, e.g.Asn, Ser,Tyr) and certain residues (e.g.Arg, Lys) with more complex spin systems the values were determined as passive coupling constants, where possible, from the relative displacement of the cross peak components in E.COSY NH-CαH, CαH-CβH or CβH-CβH cross peaks (Griesinger et al.,1987). This relative displacement is independent of the line-width and so can be used to give the coupling constant values directly. In some cases where only one of the CαH-CβH cross peaks could be observed (the other CαH-CβH cross peak overlapping with other peaks or being too close to the diagonal to be studied) one of the coupling constants could only be measured as an active coupling constant. In these cases, the true coupling constant value was then

determined from the observed splitting by using spectral simulations. For threonine, valine and isoleucine residues where only one proton is attached to the β carbon, the $^3J_{\alpha\beta}$ coupling constant values were extracted by spectral simulation from the observed antiphase splittings in CαH-CβH cross peaks in a DQF COSY spectrum. For hen lysozyme a total of 35 pairs of $^3J_{\alpha\beta}$ values for residues with an AMX type spin system, 8 pairs for residues with more complex spin systems and 14 values for threonine, valine and isoleucine residues have been determined. Details are given in Table 3.2 and examples of the experimental cross sections indicating how the measurements were made are shown in Figure 3.3.

The measured $^3J_{HN\alpha}$ and $^3J_{\alpha\beta}$ values were then used to give ϕ and χ_1 dihedral angle restraints. Ranges of ϕ were chosen that included all possible sterically allowed ϕ angles that could give rise to the observed $^3J_{HN\alpha}$ (assuming an accuracy of ± 1 Hz for the experimental values). For $^3J_{HN\alpha}$ values in the range 5.5-7.0 Hz no restraint was used because of the number of possible corresponding ϕ angles. χ_1 restraints (of 60±15°, 180±15° or -60±15°) were included for residues where the rotamer adopted by the side chain had been identified during the stereospecific assignment procedure (Smith et al., 1991). Using these criteria a total of 68 ϕ and 24 χ_1 restraints were determined for lysozyme. These are listed in Appendix IV.

Overleaf:

Figure 3.3. CαH-CβH cross peaks from the 500 MHz E.COSY spectrum of fully exchanged hen lysozyme with cross sections through the cross peaks showing how the $^3J_{\alpha\beta}$ values were measured for cases with A) Two small $^3J_{\alpha\beta}$ values; B) One large and one small $^3J_{\alpha\beta}$ value; C) $^3J_{\alpha\beta}$ values averaged by large scale motions.
1 and 2 show the cross sections through the cross peak for the Cβ proton with the lower chemical shift, and 3 and 4 show the cross sections through the cross peak for the Cβ proton with the higher chemical shifts.

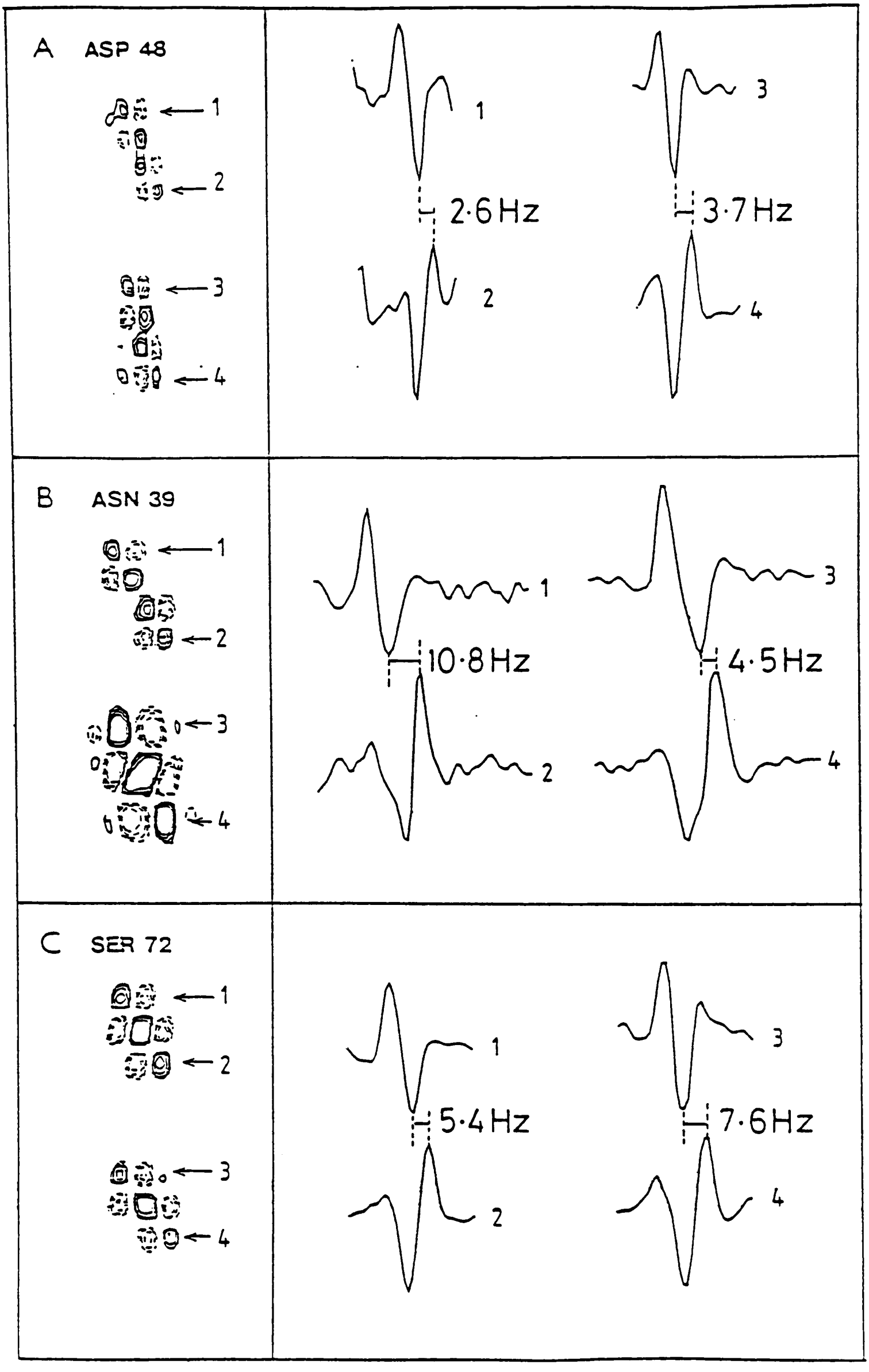

A ASP 48
1
2
3
4
1
2·6Hz
2
3
3·7Hz
4
B ASN 39
1
2
3
4
1
10·8Hz
2
3
4·5Hz
4
C SER 72
1
2
3
4
1
5·4Hz
2
3
7·6Hz
4

Table 3.2 $^3J_{\alpha\beta}$ coupling constants for hen lysozyme and χ_1 torsion angles from the tetragonal type 2 and the triclinic crystal structures.[a]

Residue	$^3J_{\alpha\beta}A$[b]	$^3J_{\alpha\beta}B$	Method [c]	χ_1 in tetragonal	χ_1 in triclinic
Val 2	10.8		C	178.5	179.6
Phe 3	(3.0	10.0)[d]	B	-84.9	-82.8
Cys 6	3.5	11.5	B	-68.2	-67.5
Glu 7[e]	6.7	6.4	A	-177.6	-65.7
Lys 13[e]	9.2	5.1	A	-157.8	-173.4
His 15	2.6	11.2	A	-84.0	-72.0
Asp 18	11.0	4.2	A	-166.2	-170.5
Asn 19[e]	6.4	7.3	A	-79.5	-161.1
Tyr 20	2.3	11.7	B	-172.8	179.7
Tyr 23	10.9	2.7	B	-70.6	-66.4
Asn 27	10.3	2.4	A	-78.9	-75.9
Trp 28	10.7	4.1	A	-64.5	-60.9
Val 29	11.1		C	175.0	178.6
Cys 30	10.8	5.3	B	-177.4	-173.5
Phe 34	5.0	10.7	B	-58.0	-67.1
Asn 37[e]	8.1	4.2	A	175.0	178.6
Asn 39	10.8	4.5	A	-172.3	-173.5
Thr 40	4.5		C	67.3	67.1
Thr 43	3.7		C	62.7	60.5
Arg 45[e]	6.9	6.7	A	-175.1	-66.4
Asn 46	4.7	11.2	A	-49.8	-60.9
Thr 47	2.6		C	52.4	61.5
Asp 48	2.6	3.7	A	70.0	70.3
Thr 51	9.3		C	-55.8	-62.1
Asp 52	11.6	3.6	A	-61.4	-55.5
Tyr 53	10.4	3.1	B	+63.6	-69.9
Asn 59	5.4	11.3	A	-168.8	-179.1
Arg 61	10.8	5.7	B	-175.5	-168.3
Cys 64	2.7	4.6	A	63.6	67.3
Asn 65	11.4	4.5	A	177.5	-170.7
Asp 66	5.1	4.5	A	65.5	64.2
Arg 68[e]	4.8	6.5	A	68.1	-66.0
Thr 69	9.3		C	-43.7	-57.8
Ser 72[e]	5.4	7.6	A	-152.0	-165.0
Asn 74	10.5	3.9	A	-154.9	-168.6
Leu 75	12.4	2.1	A	-87.5	-63.6
Asn 77[e]	8.3	5.9	A	-168.2	-162.2
Ser 85[e]	5.7	7.4	A	131.4	-166.8
Ser 86[e]	6.4	4.1	A	-69.5	-54.6
Asp 87	11.5	5.1	B	-177.6	150.4
Ile 88	4.5		C	74.5	66.0
Thr 89	9.5		C	-54.6	-63.5

Val 92	10.1		C	162.3	167.1
Asn 93	10.8	3.5	A	-74.3	-67.4
Cys 94	4.0	12.2	A	-173.6	-174.5
Val 99[e]	6.3		C	80.3	171.1
Ser 100[e]	7.7	4.0	A	-52.4	-47.3
Asp 101[e]	6.6	5.6	A	-151.4	-97.9
Asn 106	10.5	3.6	A	-68.7	-70.4
Val 109[e]	8.0		C	67.5	38.2
Thr 118	4.2		C	69.8	70.9
Asp 119	11.7	4.9	A	173.5	-175.6
Trp 123	2.9	10.6	A	-69.0	-68.9
Ile 124	4.6		C	57.2	64.3
Arg 125[e]	7.9	6.1	A	-52.0	166.4
Cys 127	4.8	11.6	B	-53.5	-58.2
Arg 128[e]	7.2	7.9	A	170.9	-86.3

[a] Coupling constant values are in Hertz (Error approximately ±1 Hz) and the χ_1 angles are given in degrees. For Val, Thr and Ile the $^3J_{\alpha\beta}$ values were determined from spectra run at 55°C. For all other residues the values were determined from spectra run at 35°C.

[b] In this work βA and βB are defined as the proton attached to the β carbon whose resonance has the higher and lower chemical shift respectively. For residues with only one β proton the values are listed under βA.

[c] Method of coupling constant measurement:

A: Values measured as passive coupling constants from E.COSY spectrum.

B: One or both values determined by simulations using active coupling constant values from the E.COSY spectrum.

C: Value measured from splitting in DQF COSY cross peaks and corrected for line width effects using simulations.

[d] Values are only approximate.

[e] Residues identified as not occupying a single staggered conformation about χ_1 using the criteria listed in the text.

3.2 Comparison of the NMR data with crystal structures of the protein.

3.2.1 Comparison with interproton distances.

The complete set of NOE distance restraints (corresponding to known distances in secondary structure and to variable distance between protons) was compared with distances calculated from the crystal structures of the protein. Figure 3.4 shows the range of interproton distances in the tetragonal type 2 crystal structure for which NOE's (those not involving methyl groups) have been observed for the different intensity categories. The majority of the NOE's fall into the category expected from the crystal structure distances and the restraints chosen from a study of the secondary structure NOE's. Over 30% of the NOE's between protons whose distances are not fixed by the secondary structure or the covalent structure of the protein, are, however, either significantly more or significantly less intense than would be predicted from the crystal structure. These NOE's are considered in more detail next.

Of the 1158 assigned NOE's, 96 violate, i.e. are more intense than expected, the tetragonal type 2 crystal structure by more than 0.1 Å with the distance restraints defined above. 59 of these violations are greater than 0.5 Å while 31 of them are greater than 1.0.Å. The NOE violations are distributed throughout the structure (see Figure 3.5) both in terms of the sequence and the three-dimensional fold. The results for the triclinic crystal structure are very similar (there are 105 NOE violations greater than 0.1 Å, 69 greater than 0.5 Å and 31 greater than 1 Å); 80% of the observed violations occur in both crystal structures. The majority of violations, including all except 6 greater than 0.5 Å, involve at least one side chain proton. The similarity in violations from the two crystal structures, that were solved independently for the protein in different space groups, suggests that they are unlikely to represent systematic differences between the average structure of the protein in a given crystalline state and in solution, but they could well reflect the effects of motional averaging, particularly for side chains, for certain regions

Figure 3.4.

Histograms showing the range of interproton distances in the tetragonal type 2 crystal structure for which NOE's (those not involving methyl groups) have been observed for the different intensity categories:

A, strong intensity NOE's; B, medium intensity NOE's; C, weak intensity NOE's; D, NOE's only observed in the longer mixing time experiments (200 or 500 ms).

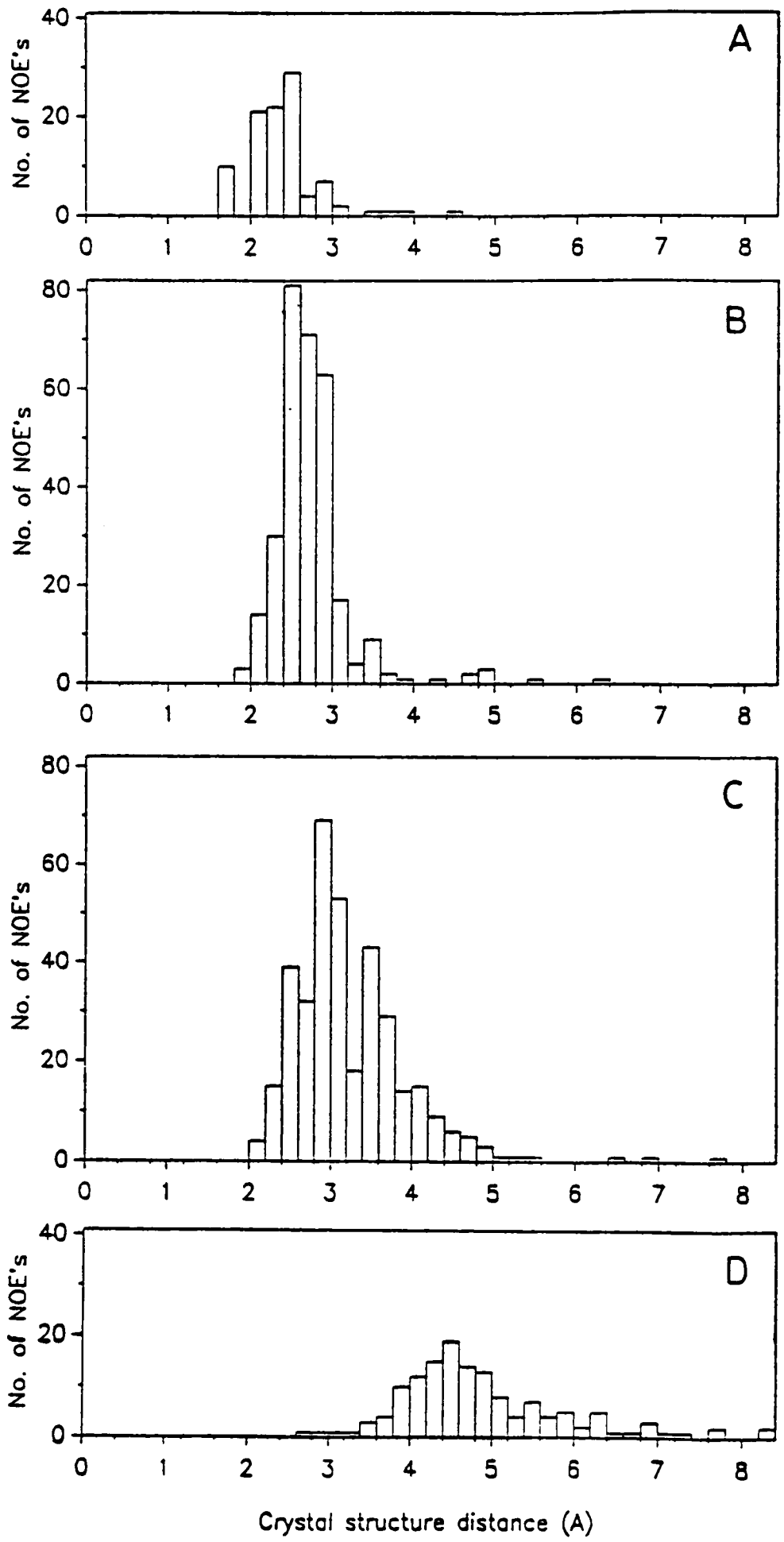
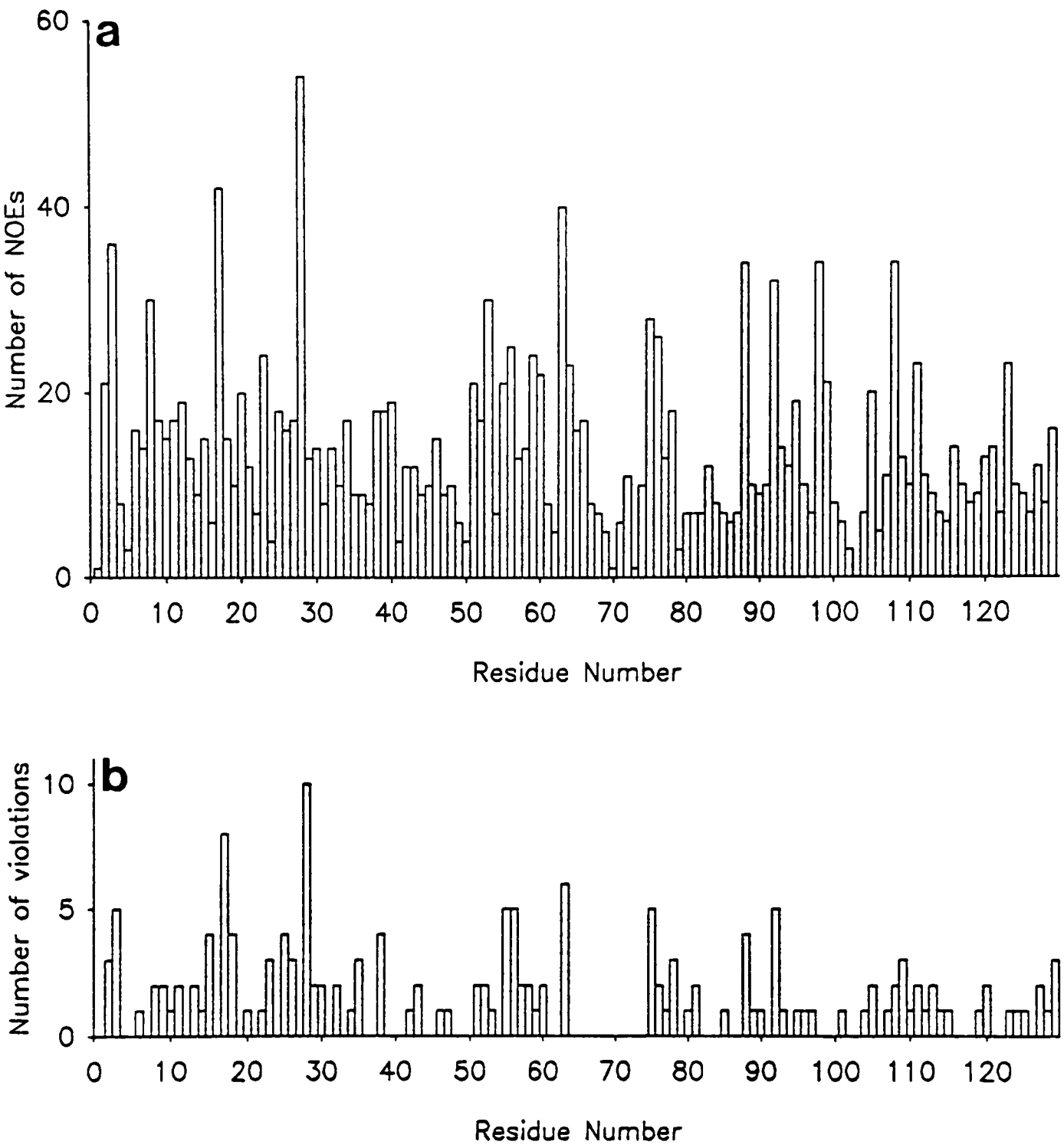


Figure 3.5

Plot a shows the total number of assigned NOE effects per residue. Plot b shows the number of violations per residue (> 0.1 Å) of the experimental NOE data, by the tetragonal type 2 crystal structure. Each NOE has been counted twice, once for each contributing residue.



of the structure in solution (Olejniczak et al., 1981; Dobson & Karplus, 1986; Post, 1992), as well as other limitations in the conversion of NOE intensities into simple distance ranges discussed in section 3.1.1.

It is also of interest to determine which NOE's are not observed, even though they lie within the distance ranges for which NOE's are expected, and which NOE's are observed to be weaker than expected. A complete search of the NMR spectra for such NOE effects was not possible because of chemical shift degeneracy and severe crowding and overlap in the 2D spectra. A selection of resolved resonances was therefore studied. In general, this analysis showed that most anticipated NOE's are in fact seen. In certain cases, however, marked discrepancies exist. For example, there are 14 protons in the tetragonal type 2 crystal structure within 4.5 Å of the Cα proton of Cys 64. It might be expected that all the NOE cross peaks between the well resolved 64 CαH and these protons would be seen in the NOESY spectrum recorded with a 50 ms mixing time. However only six of them are observed; six of the others can, however, be seen in spectra recorded with a 200 ms mixing time while the other two only appear in spectra recorded with a 500 ms mixing time. (Of these eight peaks seen at longer mixing times only four could be unambiguously assigned due to overlap and chemical shift degeneracy). The NOE intensities are therefore smaller than expected in these cases, but the effects are not absent from the spectra. In some cases the intensity reductions may be caused by large line widths for the resonances involved. For example, the NH of Asp 66 in the long loop region of the protein has a broad resonance with a line width of 14.4 Hz at 35°C (compared to an average NH line width of 9.8 Hz for the 69 residues whose line widths have been measured accurately (C. Redfield, unpublished results).) and the NOE Asp 66 NH-Thr 69 NH is only observed in 500 ms mixing time NOESY experiments despite the short interproton distance of 3.4 Å.

Although there are marked NOE intensity variations which cannot

yet be explained, the overall good agreement between the NMR data and the crystal structures clearly shows the close similarity between these crystal structures and the structure of the protein in solution. Furthermore, comparison of the data with two different crystal structures enabled any discrepancies to be identified as being likely to arise from the NOE data (and its interpretation) rather than from true differences between the structure of the protein in solution and in crystals. Many of the observed discrepancies are likely to arise from multispin relaxation or segmental motions which have not been considered in this simple analysis rather than from true differences between the average structures.

3.2.2 Comparison with ϕ angles.

$^3J_{HN\alpha}$ values calculated for lysozyme, using the ϕ angles from the tetragonal type 2 and triclinic crystal structures, are summarised in Figure 3.6 which shows plots of the coupling constants calculated from these two crystal structure versus the experimental $^3J_{HN\alpha}$ values. The agreement between the experimental and calculated data is excellent; an RMSD of 0.87Hz with a correlation coefficient of 0.93 is obtained for the tetragonal structure and an RMSD of 0.93Hz with a correlation coefficient of 0.96 for the triclinic structure. The slope of the least-squares line fitted to the data is 0.99 for tetragonal and 0.98 for triclinic lysozyme. For 82 residues in the tetragonal structure and 78 in the triclinic structure the agreement between the experimental and calculated coupling constants is better than 1Hz. The residues for which worse agreement is obtained are scattered throughout the sequence and vary to some extent between the two crystal structures. For 24 residues in the tetragonal structure and 28 in the triclinic structure the difference between the experimental and calculated coupling constants is greater than 1Hz. For only 11 of these residues (~40%) are differences of greater than 1Hz found for both the tetragonal and triclinic structures.

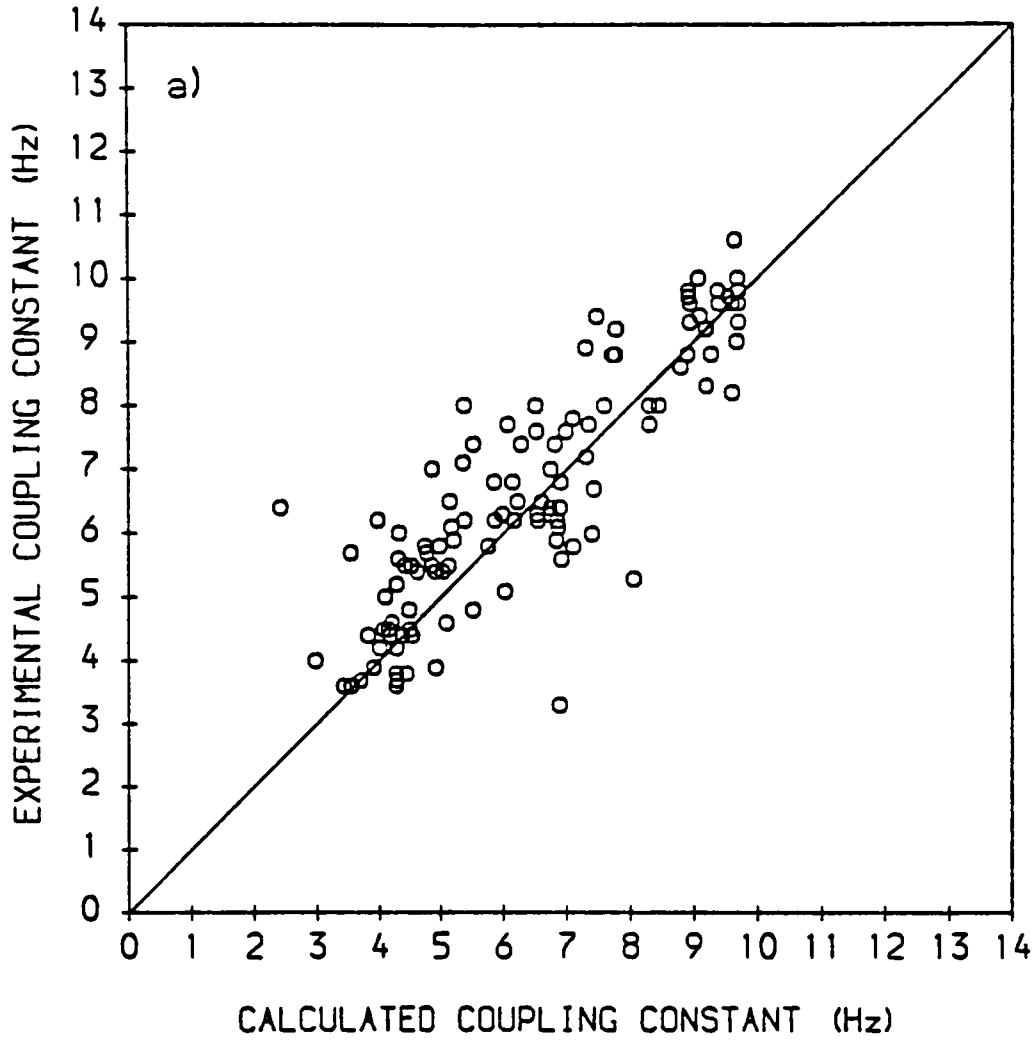
Figure 3.6.

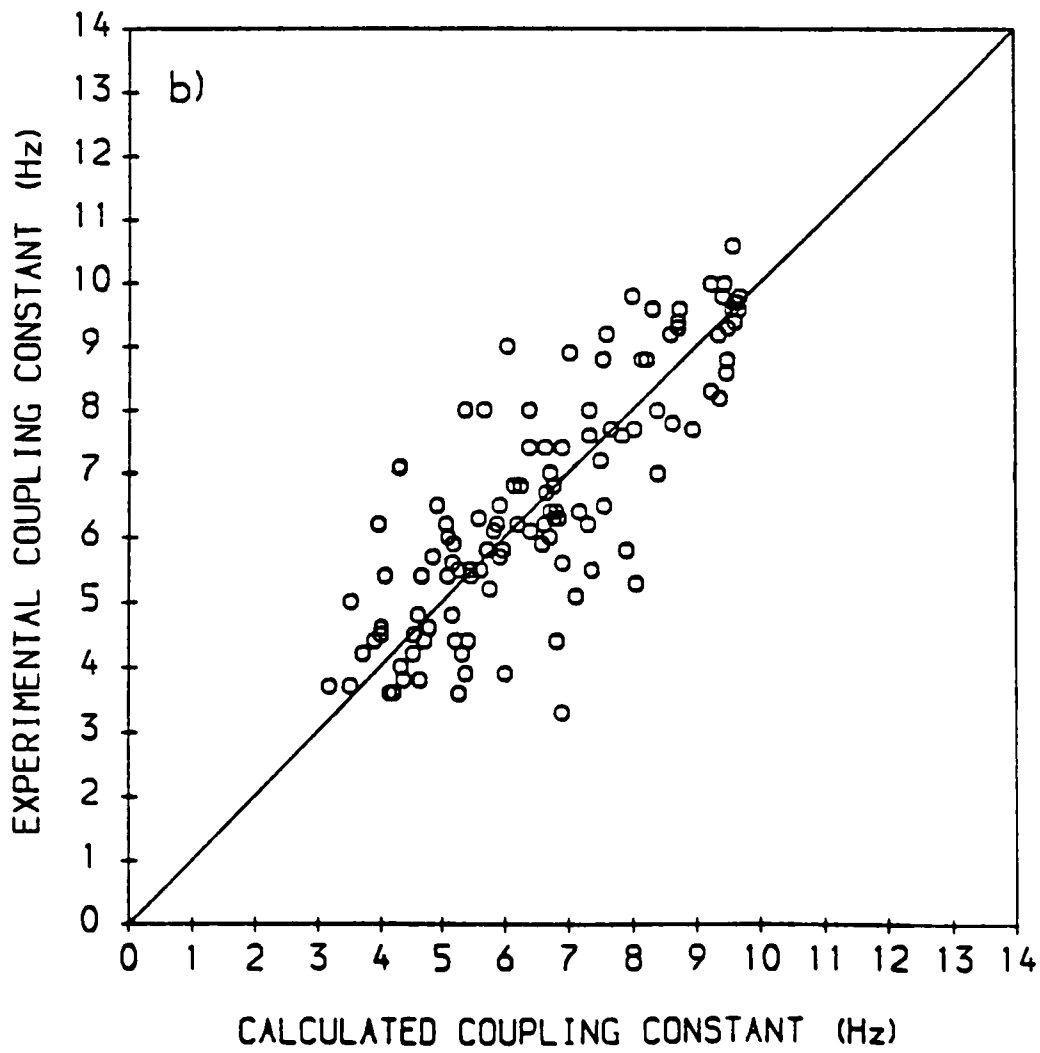
Plots of the experimental $^3J_{HN\alpha}$ coupling constants versus the calculated coupling constants for $^3J_{HN\alpha}$ values calculated using

a) ϕ torsion angles from the tetragonal type 2 crystal structure.

b) ϕ torsion angles from the triclinic crystal structure of hen lysozyme.

Stereospecific assignments for glycine CαH have not been made. The calculated glycine coupling constants are plotted against the experimental coupling constants so as to give the best agreement.





In many cases good agreement is found between the experimental and calculated coupling constant for a particular residue in one structure and a less good agreement for that residue in the other structure. For example, a difference of -1.9Hz between the calculated and experimental coupling constants of Asn 44 is found when the tetragonal structure is used whereas a difference of only 0.2Hz is found when the triclinic structure is used. For Arg 128 a difference of -2.3Hz is found for the triclinic structure while a difference of only 0.4Hz is found for the tetragonal structure. Such examples may indicate that the average structure of lysozyme in solution resembles more closely that of the tetragonal crystal structure in some regions of the protein and that of the triclinic crystal structure in other regions.

In a few cases the $^3J_{HN\alpha}$ values do not agree well with either structure. For Asp 101, for example, a difference of -2.1Hz between experimental and calculated coupling constants is found for the tetragonal structure and a difference of +1.4Hz is found for the triclinic crystal structure; the experimental coupling constant of 7.0Hz is closer to the average calculated coupling constant of 6.6Hz than it is to the values obtained from either the tetragonal (4.9Hz) or the triclinic (8.4Hz) structures. If the coupling constants calculated from the tetragonal and the triclinic crystal structures are averaged and compared with the experimental values an RMSD of 0.73Hz is obtained. It is interesting to note that this RMSD value is smaller than either of the values obtained using the tetragonal (0.87Hz) or the triclinic (0.93Hz) crystal structures alone. Although this may indicate the limitations of the present resolution of crystal structures, it is possible that for some residues the differences between the experimental and calculated coupling constants may result from some averaging of the ϕ torsion angle in solution. In this respect it is interesting that residues 100-102 in tetragonal lysozyme are characterised by high temperature factor for backbone atoms.

The differences between the experimental and calculated coupling

constants could in general arise from errors in the measurement of the coupling constants, from errors in the X-ray structures or from real differences between the solution and X-ray structures. In order to explore the latter, the minimum difference in torsion angles needed to bring the experimental and calculated coupling constants into exact agreement have been determined; these are also listed in Table 3.1 for both the tetragonal and the triclinic structures. For 70 of the non-glycine residues in tetragonal lysozyme (66 in triclinic) the required difference in torsion angle is less than 10^{o} and for only two residues (3 in triclinic) is the difference more than 20^{o} [10 residues > 15^{o} in the tetragonal and 13 residues > 15^{o} in the triclinic]. Thus, even without consideration of experimental errors only very small differences need to exist between the average backbone structure of hen lysozyme in the solution and crystalline states to explain the deviations from a perfect correlation in Figure 3.6.

3.2.3 Comparison with χ_1 angles.

$^{3}J_{\alpha\beta}$ values were calculated in a similar way using the Karplus relationship from the χ_1 torsion angles in the tetragonal type 2 and the triclinic crystal structures. Figure 3.7 shows plots of calculated versus experimental $^{3}J_{\alpha\beta}$ values using the tetragonal and triclinic crystal structure χ_1 data. For 29 residues, 67% of those studied, where a pair of $^{3}J_{\alpha\beta}$ values have been measured, the experimental values are either both less than 5 Hz or differ from each other by more than 5 Hz. Using the Karplus equation, the predicted $^{3}J_{\alpha\beta}$ values for a staggered conformation about the $\alpha\beta$ bond are either both 3.4 Hz (for $\chi_1=60^{o}$) or 12.9 and 3.4 Hz (for $\chi_1=180^{o},-60^{o}$). Given the accuracy of the experimental measurements (of the order of 1Hz), the assumptions made in the calculation of the $^{3}J_{\alpha\beta}$ values, and the fact that even slight fluctuations around the average χ_1 value will reduce the observed value for large coupling constants and increase the observed value

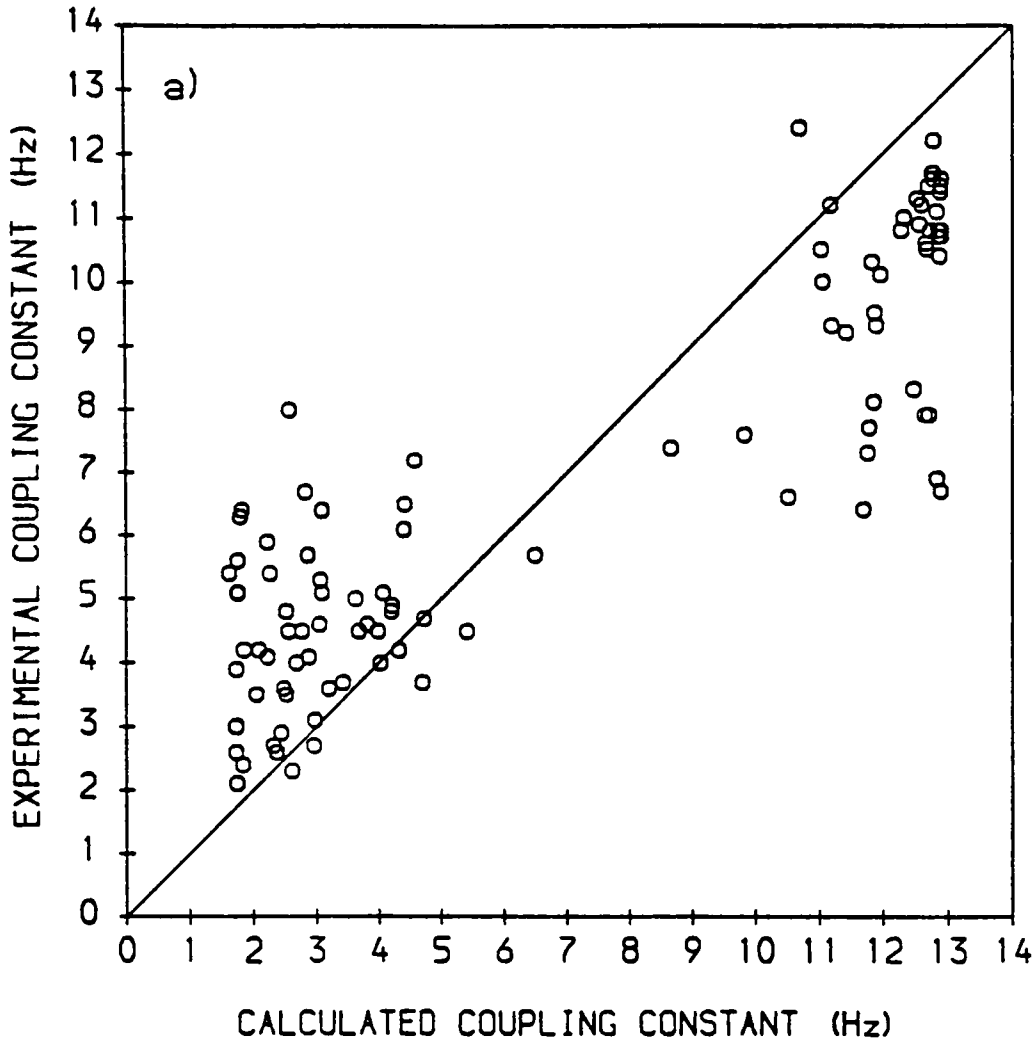
Figure 3.7

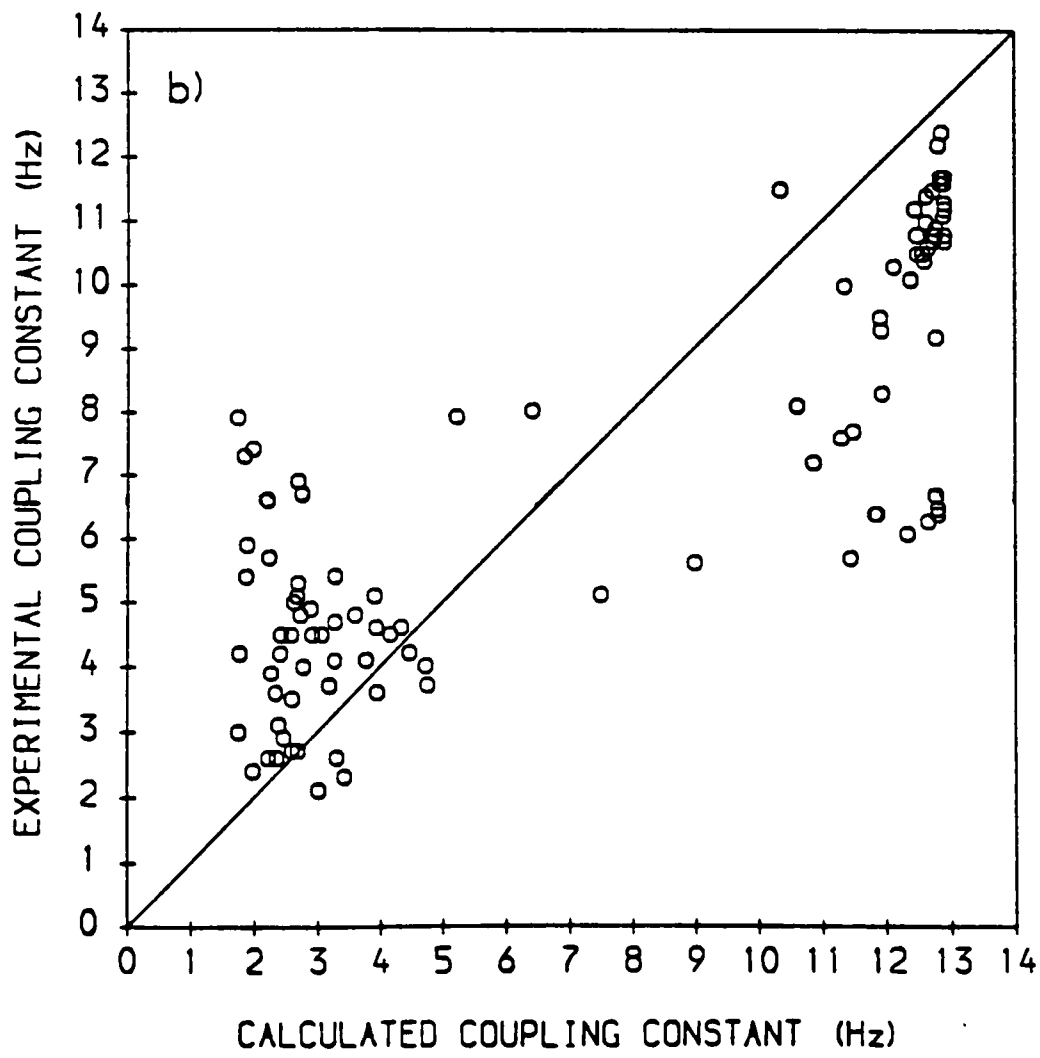
Plots of the experimental $^3J_{\alpha\beta}$ coupling constants versus the calculated coupling constants for $^3J_{\alpha\beta}$ values calculated using

a) χ_1 torsion angles from the tetragonal type 2 crystal structure.

b) χ_1 torsion angles from the triclinic crystal structure of hen lysozyme.

For residues where the β-methylene protons have not been stereospecifically assigned the calculated coupling constants are plotted against the experimental coupling constants so as to give the best agreement.





for small coupling constants (Hoch et al.,1985), we have assumed here that for these 29 residues the experimental results are consistent with a single staggered conformation. Similarly, for valine, threonine and isoleucine residues, where only one proton is attached to the β carbon, the side chain is assumed to adopt a single staggered conformation when the measured coupling constant is greater than 10 (greater than 9 for threonine because of the effects of the high electronegativity of the oxygen substituent) or less than 5 Hz. 12 out of the 14 residues studied fit into this category.

For all the residues where the experimental data are indicative of a single staggered conformation, the agreement between the experimental and calculated values is good for both structures. In addition, in every case where stereospecific assignments have been made (Smith et al., 1991), the same staggered rotamer is found to be adopted in solution and in both the crystal structures. If we consider just the data for these residues recognised to be occupying a single staggered conformation and compare the experimental and calculated $^3J_{\alpha\beta}$ values the RMSD is 1.48Hz with a correlation coefficient of 0.97 for the tetragonal data and 1.59Hz with a correlation coefficient of 0.97 for the triclinic data. The slope of the least-squares line fitted to the data is, however, only 0.74 for the tetragonal structure (0.73 for the triclinic) compared with a value of 0.99 for the $^3J_{HN\alpha}$ data. This point is discussed later.

The values of the measured $^3J_{\alpha\beta}$ coupling constants for the remaining 16 residues studied here are not consistent with the side chains of these residues being present in a single staggered conformation in solution. For the majority of these cases the measured coupling constants are in the range 5-8 Hz. For residues with two Cβ protons this requires that the coupling constants are averaged due to motion between two or more conformations populated by the side chain in solution (Nagayama and Wüthrich,1981; Hoch et al.,1985). For completely free rotation about the αβ bond both coupling

constant values are predicted (from the Karplus equation) to be ~6.6 Hz. For residues with only one β proton (2 residues here, Val 99 and Val 109) the results could either represent occupation of a non-staggered conformation or population of multiple conformations. For the residues with significantly averaged coupling constants the crystal structure χ_1 values are however generally close to the values for staggered rotamers (χ_1=60°,-60°,180°) (see Table 3.2); in all but two cases, Ser 85 in the tetragonal and Asp 101 in the triclinic, the χ_1 values are within ±30° of a staggered value. As the experimental coupling constant values are averaged as a consequence of the population of multiple conformations the agreement between the experimental and predicted set of $^3J_{\alpha\beta}$ values is poor. It is important to note here that the observation of a single resonance and $^3J_{\alpha\beta}$ coupling constant for each Cβ proton shows that fast exchange conditions are satisfied on the NMR time-scale, indicating that the interconversion between the different conformations adopted by the side chain must occur at a rate of at least 20 Hz.

It is interesting to compare the χ_1 values found in the tetragonal type 2 and the triclinic crystal structures (listed in Table 3.2). For the residues that have been identified as having a well defined χ_1 angle in solution the χ_1 values are closely similar, for 95% of the residues the difference being less than 15°. By contrast, for 8 of the 16 residues with averaged coupling constants there are major differences, the side chain adopting different rotameric forms in the two structures. For Glu 7, for example, χ_1=-177.6° in the tetragonal structure but χ_1=-65.7° in the triclinic crystal structure.

Attempts to extend the coupling constant calculations and χ_1 angle comparisons to additional crystal structures of a lower resolution (e.g. the monoclinic 2.5 Å resolution structure (Rao et al.,1983)) showed that there are major differences between the χ_1 angles in these and the tetragonal and triclinic crystal structures for residues with both averaged and non-averaged

$^3J_{\alpha\beta}$ values. For example, in the comparison of the monoclinic and type 2 tetragonal crystal structures, 62% of the side chains with nonaveraged $^3J_{\alpha\beta}$ values occupy different rotamers in the two structures. The lower resolution of these structures is, therefore, not sufficient to define the side chain conformations sufficiently accurately for the purpose of comparison with the NMR data.

3.2.4 Comparison with crystallographic temperature factors.

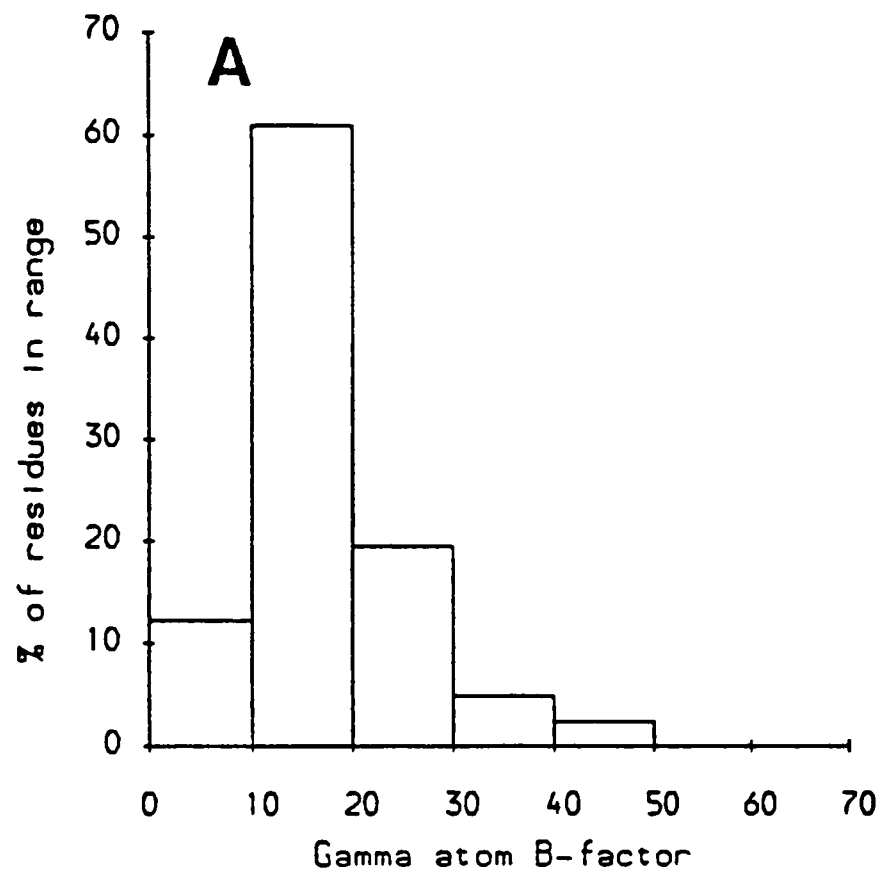
Unless multiple occupancies are specifically identified in a crystal structure (Smith et al., 1986) information about distributions of side chain conformers comes from crystallographic temperature factors. Examination of the available temperature factors for tetragonal type 2 crystal structure shows that for residues identified as having a single conformational state about χ_1 in solution, the temperature factors of the γ atoms are in general low and comparable in size with the temperature factors for main chain atoms. For the residues with motionally averaged $^3J_{\alpha\beta}$ coupling constants a significant proportion of the residues have substantially higher γ atom temperature factors; 50% of the residues, compared with only 7% for residues occupying a single conformation, have γ atom temperature factor values greater than 30Å^2 (see Figure 3.8). In these cases there is therefore evidence for rotational disorder about χ_1 in the crystal structure as well as in the solution structure. Rotational disorder for side chains in crystals of lysozyme has been previously identified specifically from crystallographic data for four residues (Arg 73, Val 109, Arg 125 and Arg 128) (Grace, 1980). Three of these residues have been investigated here, Val 109, Arg 125 and Arg 128; in all cases the $^3J_{\alpha\beta}$ values are found to be motionally averaged. It is, however, interesting to look at the cases with averaged $^3J_{\alpha\beta}$ coupling constant values but low temperature factors for the γ position. In three of the five cases with temperature factors for the γ atom less than 20Å^2 (Lys 13, Asn 37 and Arg 68)

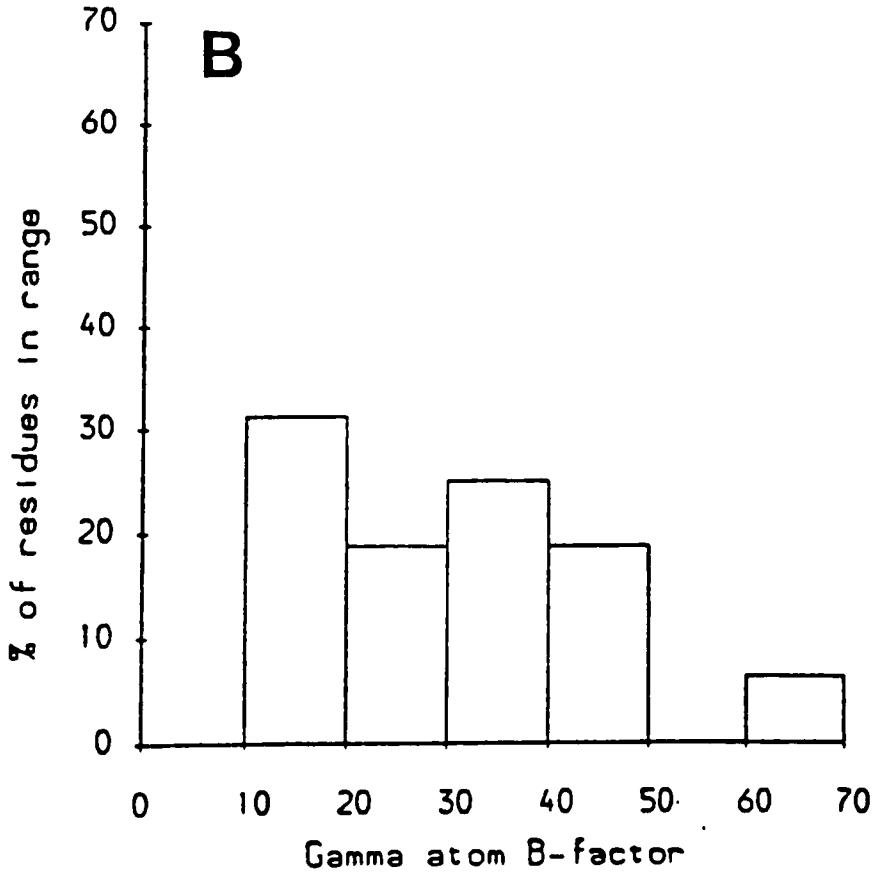
Figure 3.8.

Histograms showing the percentage of the residues studied with a given temperature factor for the γ position in the tetragonal type 2 crystal structure for :

A) The 41 residues where $^3J_{\alpha\beta}$ values are consistent with a single staggered conformation about χ_1.

B) The 16 residues with $^3J_{\alpha\beta}$ values averaged by the population of multiple conformations.



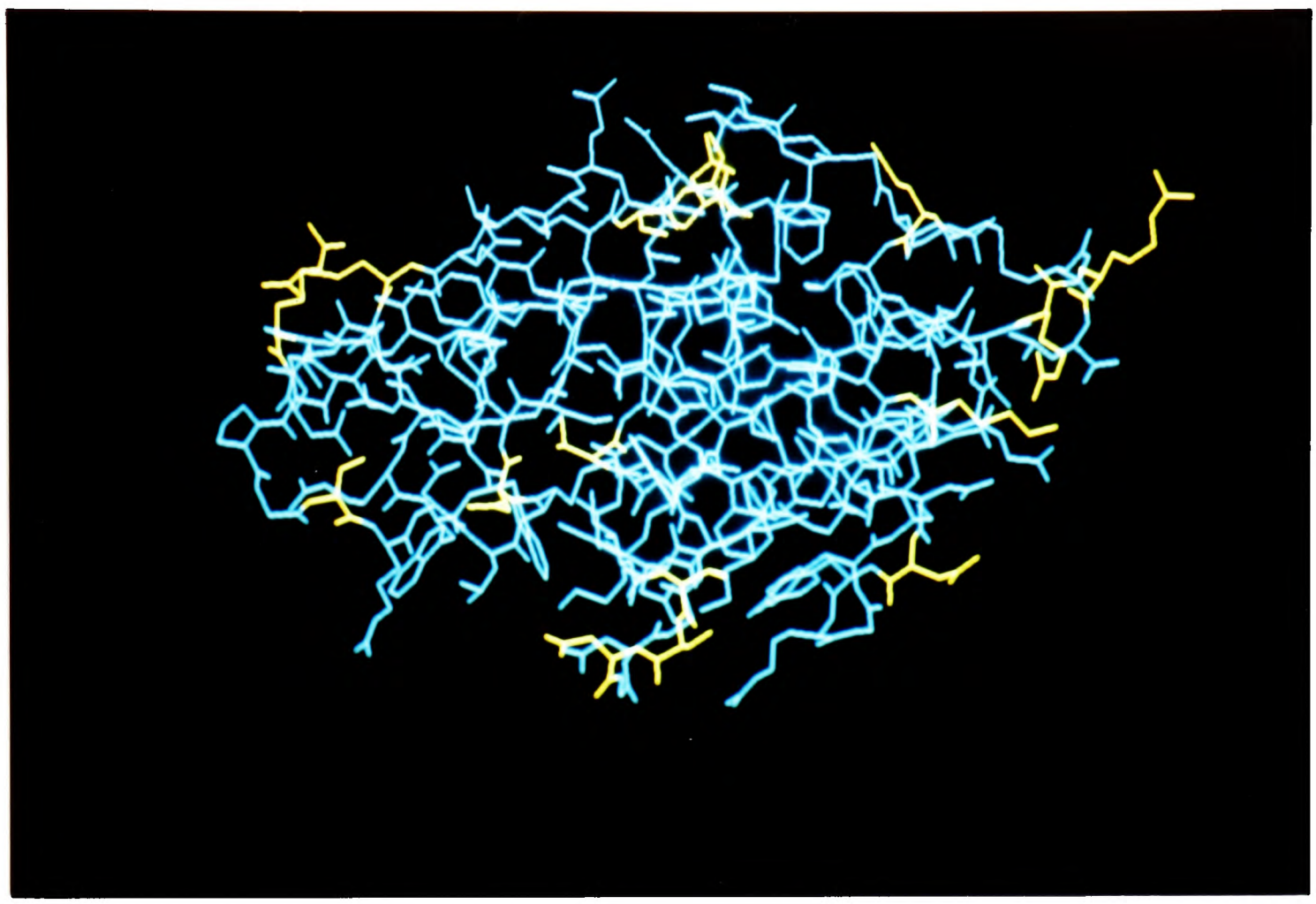


there are high temperature factors (greater than 30Å^2) for atoms further out along the side chains suggesting that even in these cases rotational disorder does occur in the crystal but that, in comparison to the solution case, it is restricted to the extremities of side chains.

3.2.5 Side chain environments.

An examination of the environments of the side chains of the residues for which $^3J_{\alpha\beta}$ values are available in the crystal structures shows that almost all the residues found to have large discrepancies between the calculated and experimental $^3J_{\alpha\beta}$ coupling constants have exposed positions at the protein surface, with the side chains clearly extending out into the surrounding medium (see Figure 3.9). The only exception is Val 99 which has an interior position in the protein structure; this residue is discussed further below. Interestingly, the side chains of two of the residues with multiple conformations in solution, Glu 7 and Arg 68, are found in the tetragonal, but not the triclinic, crystal structure to be hydrogen bonded, to Gly 4 and Thr 51 respectively. If these hydrogen bonds are present in solution they do not constrain the side chain conformations to a single minimum. It seems likely, therefore, that either they are not present or that they are weak and are only present in one of the multiple side chain conformations adopted in the solution structure. The residues where we do not see large scale averaging of coupling constants have a range of positions in the structure. Many are quite buried, but a few such as Asp 18 and Asn 65, are in positions which are as exposed as the majority of those in the earlier category. In these cases interactions with neighbouring groups are presumably sufficient to stabilize one conformation significantly relative to the others. There is no apparent difference, however, in the correlation between the experimental and calculated $^3J_{HN\alpha}$ data for residues with or without rotationally disordered side chains.

Figure 3.9
The tetragonal type 2 crystal structure of hen lysozyme showing the residues with motionally averaged $^3J_{\alpha\beta}$ values in yellow.



In order to probe the environment of the side chains more systematically, solvent accessibility calculations were performed for the tetragonal and triclinic crystal structures using the method of Richmond and Richards (1978). The percentage exposure of the side chains (the surface area of the side chain atoms accessible to water in the lysozyme molecule compared to the accessible surface area of the same side chain in a fully extended conformation) were calculated for all the residues. The results for the tetragonal crystal structure are shown in Figure 3.10; the data for the triclinic structure are closely similar. The distribution of values confirms that the side chains where we see extensive motional averaging generally have a high percentage solvent exposure (62% of the side chains in both crystal structures having values greater than 60%) while lower percentage exposure values dominate for side chains where we see no large scale averaging of coupling constants (only 14% of these residues in the tetragonal and 17% in the triclinic structures having values above 60%).

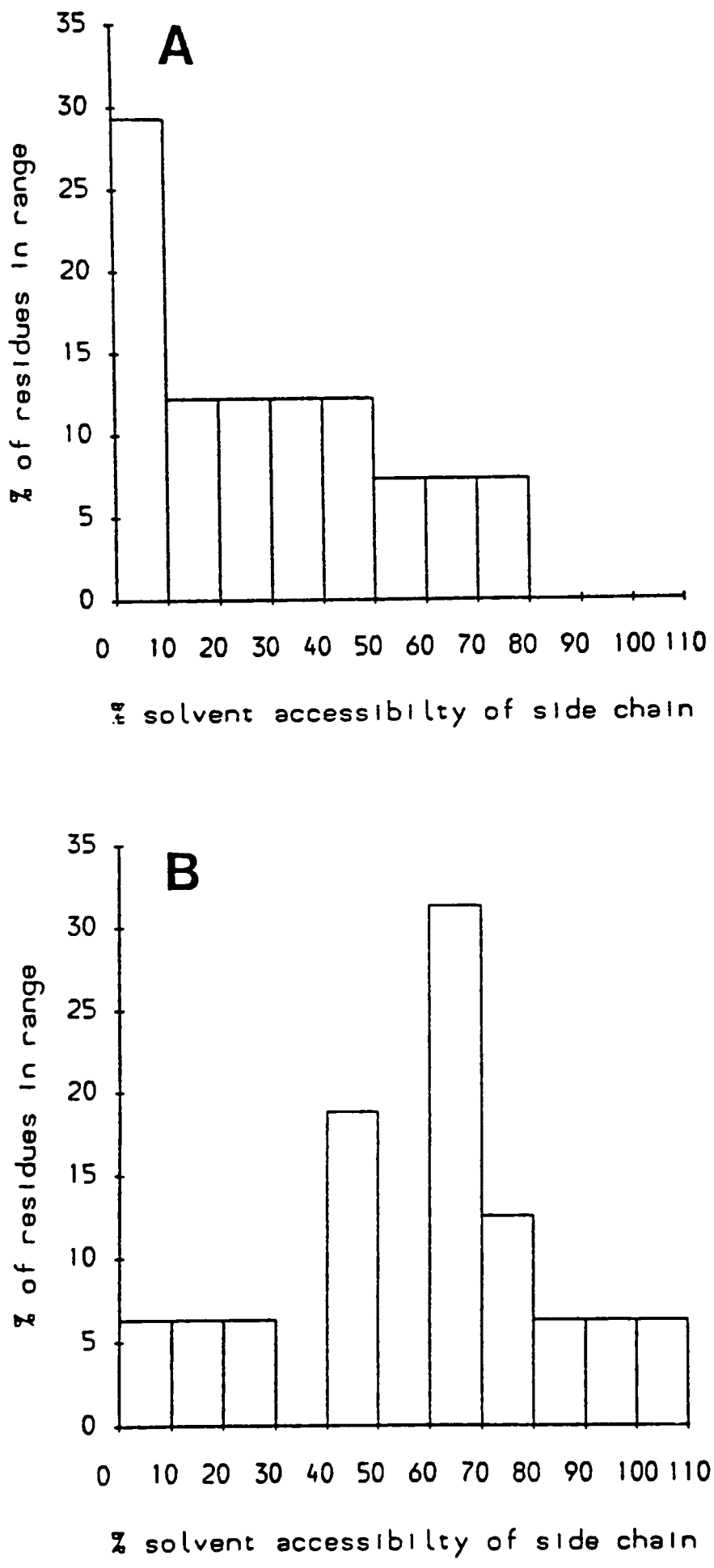
This analysis indicates that in the interior of the protein the majority of side chains have a well defined conformation about χ_1 in solution which is accurately represented by the crystallographic χ_1 values. In contrast, 55% of the residues studied with a side chain solvent accessibility greater than 60% adopt multiple conformations and hence are undergoing large scale motions in solution. An exception to this generalisation appears to be Val 99 which has an experimental $^3J_{\alpha\beta}$ value of 6.3 Hz but an interior position in the structure and a side chain solvent accessibility of zero. Here, as there is only one proton attached to the β carbon, the experimental coupling constant value could represent the presence of an unusual conformation about χ_1 rather than motional averaging. A comparison of the χ_1 values in the two crystal structures shows however, that the residue occupies different staggered conformations in the tetragonal and triclinic crystal forms (χ_1 =

Figure 3.10

Histograms showing the distribution of the residues studied in categories of side chain solvent accessibility in the tetragonal type 2 crystal structure for :

A) The 41 residues where $^3J_{\alpha\beta}$ values are consistent with a single staggered conformation about χ_1.

B) The 16 residues with $^3J_{\alpha\beta}$ values averaged by the population of multiple conformations.



171.1° in the triclinic and 80.3° in the tetragonal). It is probable therefore that the side chain of this residue adopts multiple conformations in solution; 50% population of these two conformers would give, using the Karplus equation, a predicted $^3J_{\alpha\beta}$ value of 7.3Hz. Additional evidence in support of this interpretation of the coupling constant value comes from the observation that the two NH-CγH_3 NOE's are found to have closely similar intensities. Distances from the tetragonal crystal structure predict that the NH-C$\gamma_1 H_3$ NOE should be more intense than the NH-C$\gamma_2 H_3$ NOE (NH to CH_3 distance 2.9 Å for γ1 and 4.4 Å for γ2) while the triclinic crystal structure predicts the converse (i.e. NH-C$\gamma_2 H_3$ should be more intense than NH-C$\gamma_1 H_3$; distances 4.6 Å for γ1 and 2.7 Å for γ2). In addition, evidence for the rotational mobility of the side chain of Val 109 has been obtained previously from the broadening of resonances on the binding of Gd(III) to the protein (Campbell et al.,1975).

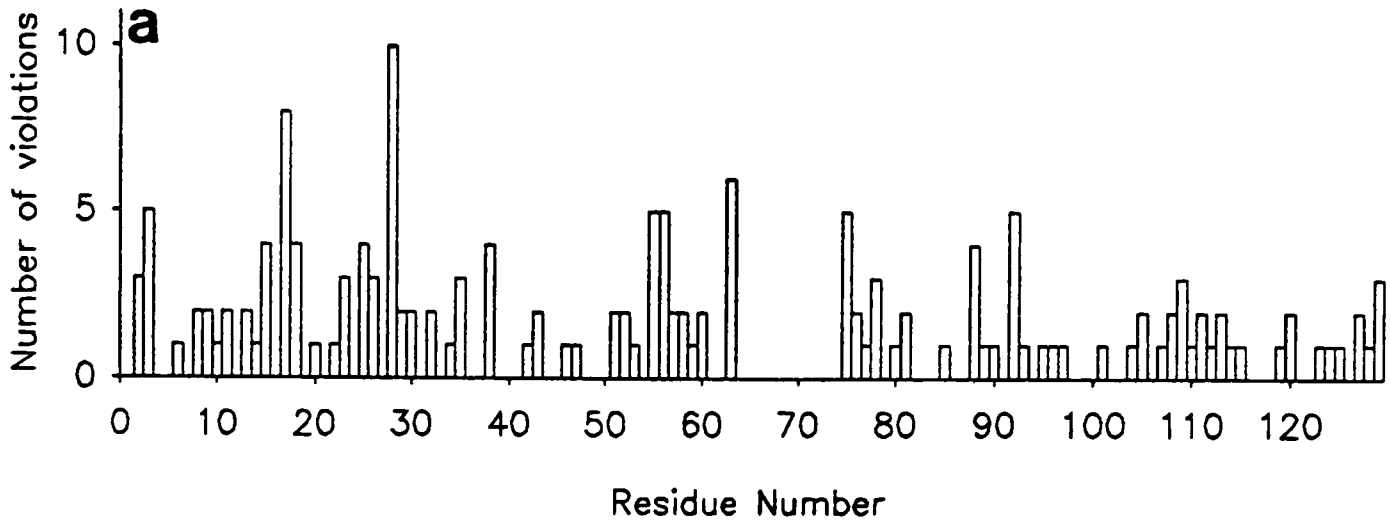
3.3 Energy minimization of the crystal structure.

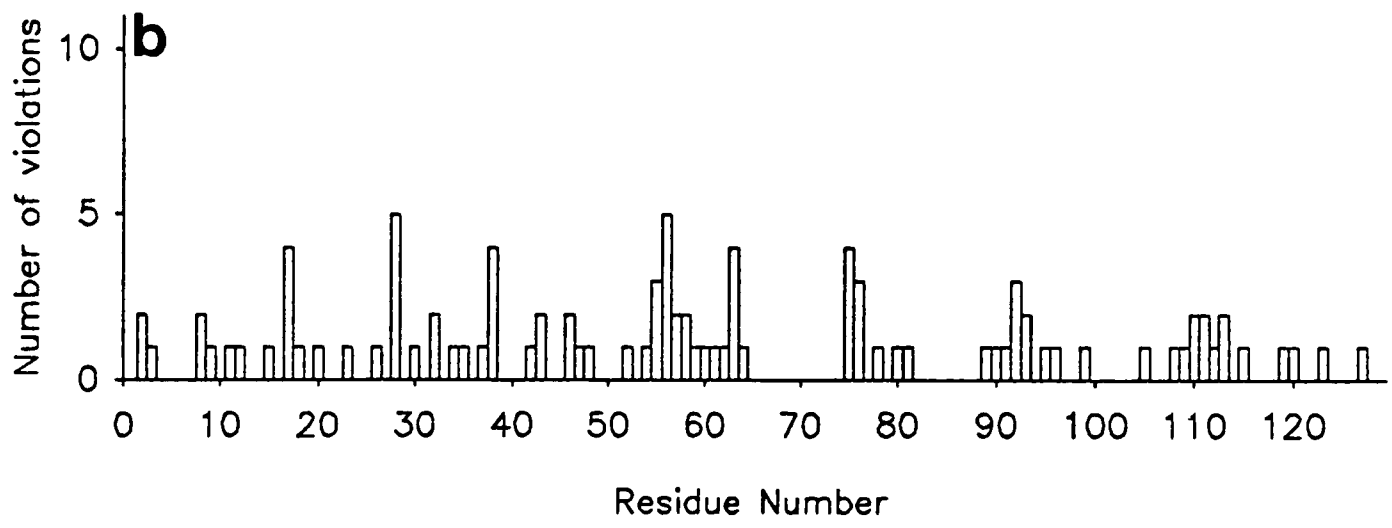
To assess further the importance of the observed discrepancies between the NMR data (particularly the NOE's) and the crystal structures of the protein, the tetragonal type 2 crystal structure was minimized (500 steps of restrained Powell minimization using program XPLOR (Brünger, 1988)) in the presence of the NMR restraints (1158 NOE restraints and 68 ϕ and 24 χ_1 dihedral angle restraints). The number of NOE violations was greatly reduced following minimization; 56 violations are greater than 0.1Å but only 9 violations are greater than 0.5 Å and none more than 1Å (compared with 96 violations greater than 0.1 Å, 59 greater than 0.5Å and 31 greater than 1Å prior to minimization) (Figure 3.11). The NOE violations remaining greater than 0.5 Å are listed in Table 3.3. About half involve a main chain atom, but in all but one case, the other in such a pair is a side chain atom. The resulting structure has an RMSD from the initial crystal structure coordinates

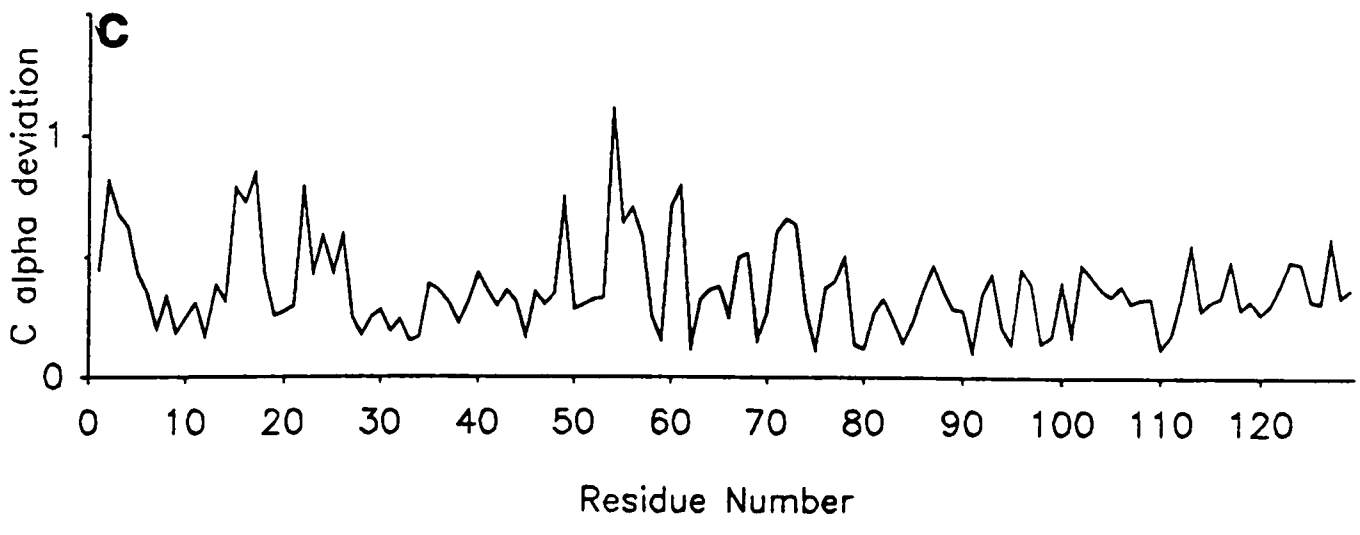
Figure 3.11

Plots a and b show the number of violations per residue (> 0.1 Å) of the experimental NOE data, by the tetragonal type 2 crystal structure and by the same structure following restrained energy minimization respectively. Each NOE has been counted twice, once for each contributing residue.

Plot c shows the Cα deviation between the crystal and energy minimized crystal structures (in Å).







of 0.41 Å for the Cα atoms, and 0.66 Å for all atoms. Study of the Cα atom deviation between these structures for each residue in turn (see Figure 3.11 c), shows that there are no significant concerted changes to the structure of the protein as a result of this minimization, the minor changes being concentrated around residues 2-4, 15-17, 22, 54-57 and 60. This therefore suggests that the observed NOE violations are not structurally significant, particularly as the changes are largely restricted to small movements of side chains relative to each other, and to small changes in the main chain in the loop regions.

Table 3.3.

Violations of the experimental NOE distance restraints by the energy minimized tetragonal type 2 crystal structure that are greater than 0.5 Å.

NOE[1]	Restraint (Å)	Distance in minimized crystal structure (Å)
Leu 8Hβ# - Ala 9Hα	1.8 - 2.5	3.57, 5.16
Leu 8Hδ*- Leu 17Hγ	1.8 - 3.0	5.28, 6.18
Leu 17HN - Leu 17Hγ	1.8 - 2.5	3.33
Tyr 23 HN -Trp 28 HZ2	1.8 - 3.0	3.77
Trp 28 HZ3 -Leu 56 Hγ	1.8 - 3.0	3.66
Trp 28HE1-Trp 28Hβ#	1.8 - 3.0	4.50, 4.92
Asp 52 Hα -Gln 57 HN	1.8 - 5.5	6.05
Cys 64Hα -Leu 75 Hβ1	1.8 -5.5	6.25
Leu 75 HN -Leu 75 Hγ	1.8 - 3.0	3.55

[1] # and * represent pairs of methylene protons and methyl groups respectively which have not been stereospecifically assigned.

However, although the agreement of the NOE data with the crystal structure coordinates improves on restrained energy minimization it is interesting that the agreement of the crystal structure with the coupling constant data is reduced by the energy minimization. For the energy minimized structure, for $^3J_{HN\alpha}$ the RMSD from the experimental values is 1.61 Hz compared to an RMSD of 0.87 Hz for the tetragonal type 2 crystal structure and for $^3J_{\alpha\beta}$ the RMSD is 2.69 Hz compared to an RMSD of 2.52 Hz for the crystal structure. In addition, if the experimental secondary shifts are compared with the ring current shifts calculated from the coordinates of the crystal and energy minimized structure the agreement is found to become less good on energy minimization. For the 38 methyl groups with calculated shifts greater than 0.1 ppm, for example, the RMSD with the experimental secondary shift is for the energy minimized structure 0.33 ppm compared to an RMSD for the crystal structure of 0.22 ppm. From these results it is therefore apparent that the structure resulting from energy minimization in the presence of the NMR restraints does not represent a better description of the structure in solution than the X-ray structure does. The agreement between the experimental coupling constant values and the values predicted from the crystal structure coordinates reduces on energy minimization despite the fact that dihedral angle restraints derived from the coupling constant values were included in the energy minimization. This reflects the fact that the dihedral angle restraints only specify a range of angles which encompass all the possible angles that could give rise to the observed coupling constant (see Section 3.1.2).

3.4 Comparison with molecular dynamics simulations.

It is interesting to compare the results presented here with those predicted from molecular dynamics simulations. A molecular dynamics simulation to study the effects of internal picosecond fluctuations (such as vibrational motions and some dihedral angle transitions) on the observed NOE intensities has been carried out (Post, 1992). The results show that there are significant effects (which would lead to a 10% or greater error in distances estimated using a rigid model of the protein) for approximately 11% of the non-methyl group NOE interactions and 32% of the NOE's involving one or two methyl protons in lysozyme. The error can result in either an underestimated or an overestimated distance. Hence these motions could, in principle, account for much of the discrepancy between the NOE intensities and crystal structure distances discussed in this chapter. However, the work also showed that, if only a qualitative evaluation of distances based on strong, medium and weak intensities is carried out, and the NMR distances specify only upper bound restraints (as has been used in the work presented here) picosecond motions would result in incorrectly categorizing less than 1% of the interactions.

Molecular dynamics simulations have also been used to study the effects of internal fluctuations on the NMR coupling constants (Hoch et al.,1985; Dobson & Karplus, 1986). In these simulations the time-average coupling constants were compared with those computed from the average structure of the protein obtained from the simulations. These values correspond to the experimental coupling constants and those calculated from the torsion angles observed in the crystal structure respectively. Excluding cases where the side chain occupies multiple conformations, the simulations predict that the gradient of the least-squares line fitted to the data will be 0.89 for coupling constants corresponding to ϕ and 0.88 for those corresponding to χ_1. Comparing these values with the experimental data presented here

(gradients of 0.99 for ϕ and 0.74 for χ_1) show that for the main chain the actual fluctuations about the average ϕ torsion angle appear to be smaller than those predicted by the simulations. This observation could, however, be an artifact resulting from the A, B, C parameters used in the Karplus equation. The parameters were obtained from a fit of experimental BPTI coupling constants to the X-ray ϕ angles (Pardi et al.,1984) and they will therefore already take into account the backbone dynamics of BPTI in solution. The gradient of 0.99 obtained for lysozyme could reflect that this protein has similar dynamical behaviour in solution to BPTI. For side chains about χ_1 there is a poorer experimental correlation which could result from the parameters used in the Karplus equation or it could reflect larger experimental fluctuations about χ_1 than predicted.

3.5 Conclusions.

An extensive set of NOE and coupling constant data has been collected for hen lysozyme. Comparison of these data with high resolution crystal structures of the protein show the close similarity of the structure of the protein in solution and in crystals. There are, however, some differences between the expected and observed NOE intensities which are likely to arise, at least in part, from motions within the protein in solution which have not been taken into account in the analysis. The level of agreement between the experimental NOE's and those anticipated from the crystal structure can, however, be increased further by energy minimization in the presence of the NMR data. The structural changes resulting from this minimization are small. Furthermore, the agreement with other experimental NMR data is reduced by this procedure. There is little evidence therefore that any overall improvement in the protein structure emerges from the energy minimization.

The study of the coupling constants of the protein has enabled the nature of some of the motions in the protein to be recognised. The work has shown that in solution, in the interior of the protein both the main chain and the side chain atoms have well defined conformations. The side chains are predominantly in staggered conformations and they appear to only undergo small fluctuations about their average χ_1 value. An exception in the case of lysozyme is Val 99, the side chain of which seems to adopt multiple conformations despite the interior position of the residue. The situation for residues located near the surface of the protein is, however, very different. The main chain ϕ torsion angles are still well defined but many of the side chains adopt multiple conformational states about χ_1 as indicated by extensive averaging of $^3J_{\alpha\beta}$ values. This conformational variability of surface groups in solution is underestimated by studying an individual crystal structure but the data presented here suggest that it may perhaps be revealed by a survey of structures in different crystalline forms of the protein.

In the next chapter, the use of the NMR data described in this chapter in the calculation of the solution structure of hen lysozyme is presented. An analysis of the calculated structures is then given, comparing them with both crystal structures of the protein and with the experimental NMR data.

CHAPTER 4.

THE STRUCTURE OF HEN LYSOZYME IN SOLUTION.

Calculation and analysis of the NMR structures.

In Chapter 3 the collection of the NMR data for hen lysozyme was discussed and the data were compared extensively with the crystal structures of the protein. In this chapter the use of these data in the independent determination of a solution structure for the protein is described. The main chain and side chain conformations in the ensemble of calculated NMR structures are analysed in detail and are compared with the crystal structures of the protein. The structures are also compared with other NMR data for the protein, in particular coupling constants and ring current shifts, in an attempt to assess the success of the NMR structure determination.

4.1 Calculation of the NMR structures.

Structure calculations were performed using the distance geometry-dynamical simulated annealing approach (section 2.5) with the NMR data set of 1158 NOE distance restraints and 68 ϕ and 24 χ_1 dihedral angle restraints discussed in Chapter 3. In addition, hydrogen bond restraints were used in the initial stages of the simulated annealing calculations (Williamson et al., 1985) for residues in the recognised secondary structure of the protein with slowly exchanging amide protons (Redfield & Dobson, 1988). In particular, distance restraints between NH(i)-O(i-4) (1.3-2.3 Å) and N(i)-O(i-4) (2.3-3.3 Å) were included for the helical residues 6-15, 25-37 and 89-101. For the β sheet region restraints were used between the residue pairs 42- 54, 44-52, 46-50 and 53-58. The four disulphide bonds in lysozyme (6-127, 64-80, 76-94, 30-115) were included as distance restraints in the distance geometry and initial stages of the simulated annealing procedures. In the final part of the

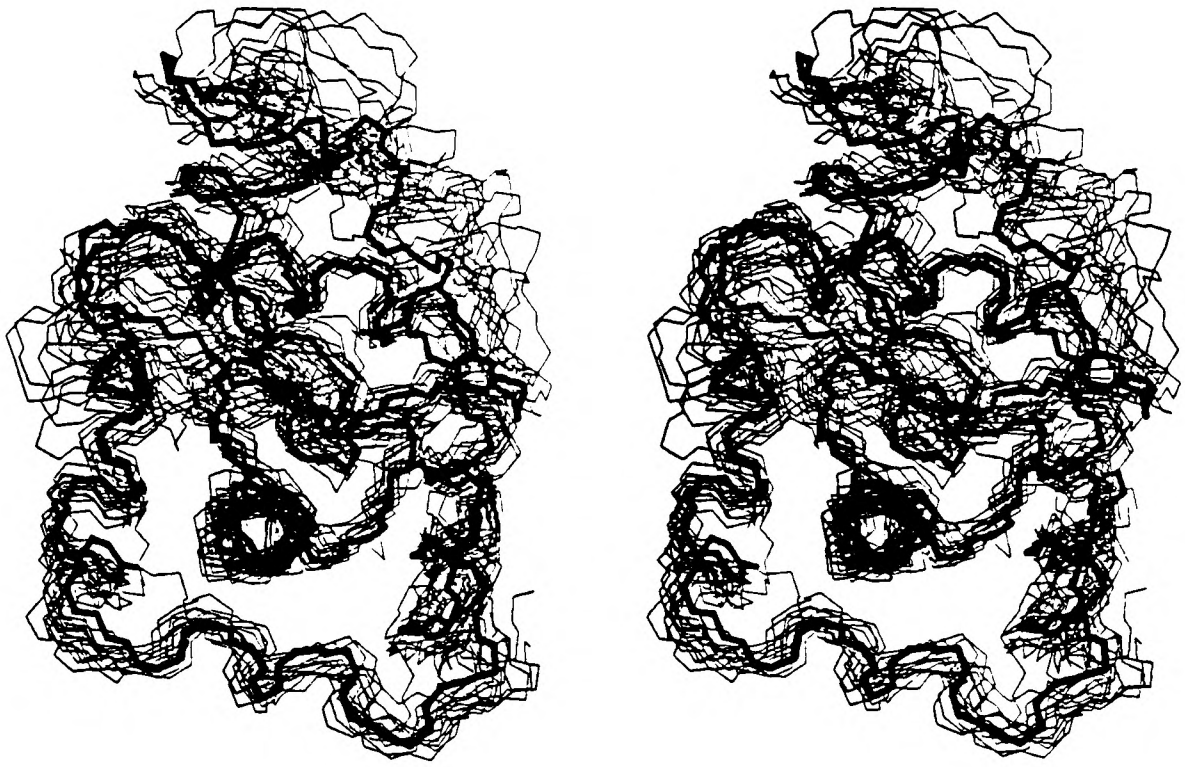
calculations they were treated as normal bonds. Full details of the restraints used in the calculations are given in Appendices III and IV.

An ensemble of 16 low energy structures were calculated; the energies and violations for these structures are listed in Table 4.1. There are on average six NOE violations greater than 0.5 Å in the calculated NMR structures (compared to 59 violations in the tetragonal type 2 crystal structure and 9 violations in the energy minimized crystal structure; see section 3.3), with an average of 55 NOE violations greater than 0.1 Å. Four of the violations greater than 0.5 Å, which are listed in Table 4.1, occur in at least half of the calculated structures; these all also violate the restrained energy minimized crystal structure by more than 0.5 Å. It is likely that these NOE peaks may have had their intensities increased in the spectra by spin diffusion or overlap with other peaks. On the criteria of agreement with the NOE distance restraints, the calculated structures are therefore a significant improvement on the X-ray structure and a marginal improvement on the energy minimized structure.

4.2 Analysis of the NMR structures.

The set of 16 final calculated structures is shown in Figure 4.1. The Cα RMSD across the set is 1.8 ± 0.2 Å. The average Cα deviations of the set of calculated structures from the average structure coordinates are given in Figure 4.2A and indicate that the RMSD values are dominated by large deviations for a relatively small number of residues (14 residues have an average Cα deviation from the average structure > 2.5 Å). If these 14 residues are excluded in the superposition of the structures and calculation of the RMSD, the value is reduced to 1.3 ± 0.2 Å.

Figure 4.1.
Stereodiagram and colour print showing the main chain superposition of the 16 calculated NMR structures of hen lysozyme on the tetragonal type 2 crystal structure of the protein (crystal in red in photo and in bold in stereodiagram).



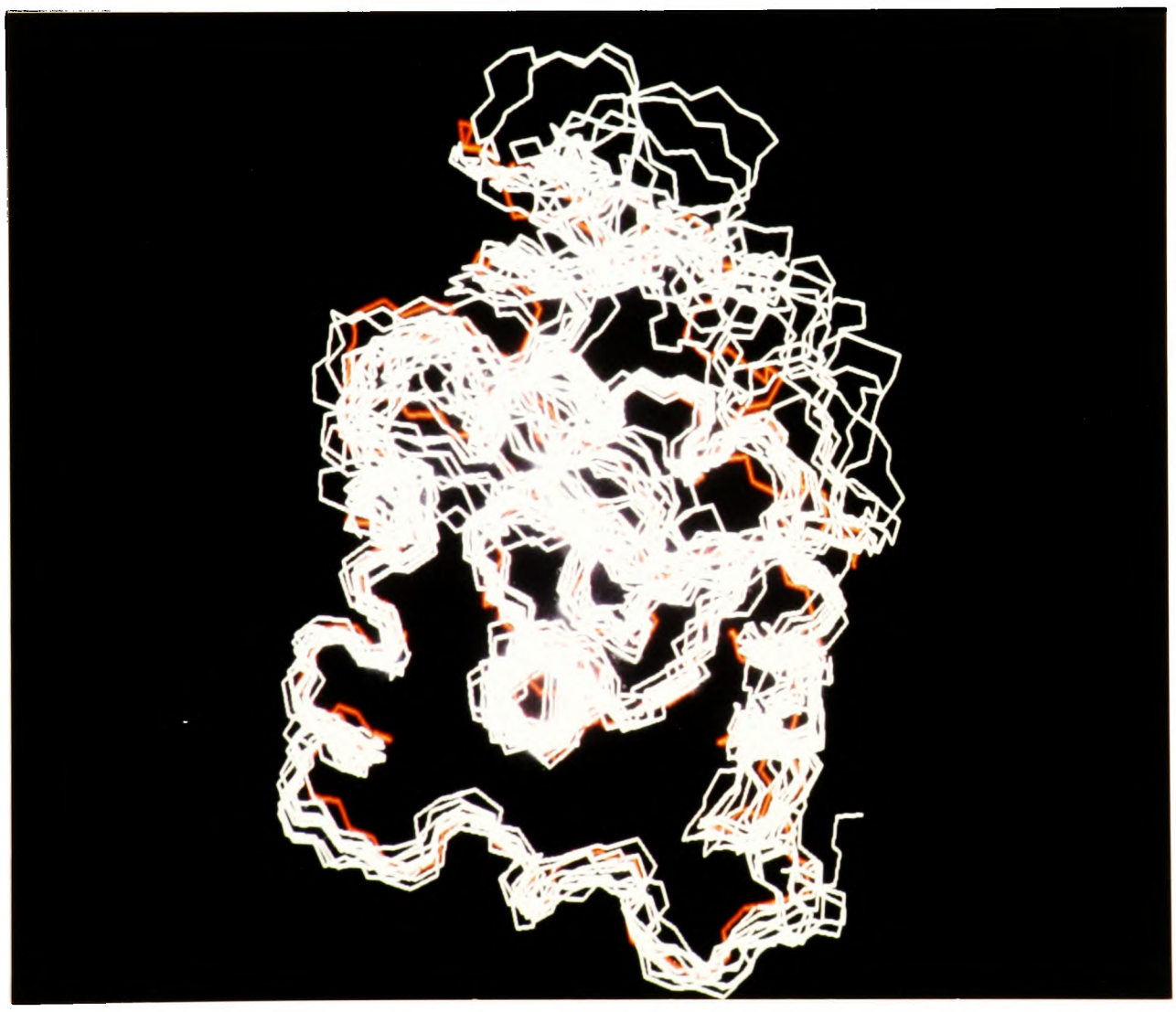


Table 4.1 Summary of the energies and violations for the ensemble of calculated structures.[1]

<u>RMS Deviations</u>	
NOE's (Å)	0.079 ± 0.012
Bonds (Å)	0.009 ± 0.004
Angles (deg)	2.76 ± 0.07
Impropers (deg)	0.23 ± 0.02
N_{NOE}	6 ± 3
<u>Energies</u>	
E_{VDW} (kcal/mol)	-640 ± 20
E_{NOE}	69 ± 24

<u>NOE Violations</u> > 0.5 Å occured for the following pairs of protons in at least 8 of the calculated structures :

Leu 8Hβ# - Ala 9Hα [2]

Leu 8Hδ*-Leu 17Hγ

Trp 28 HE1-Trp 28 Hβ#

Cys 64Hα -Leu 75 Hβ1

[1] The values listed are the average over the ensemble of 16 calculated structures. The RMS deviations of the NOE's from the experimental restraints are calculated with respect to the upper and lower bounds of the distance restraints. RMS deviations of bond lengths, angles and improper dihedrals are calculated with respect to idealized geometries. N_{NOE} are the number of NOE violations greater than 0.5 Å; E_{VDW} is the Lennard-Jones energy and E_{NOE} is the NOE energy calculated with a force constant of 10 kcal/mol $Å^2$.

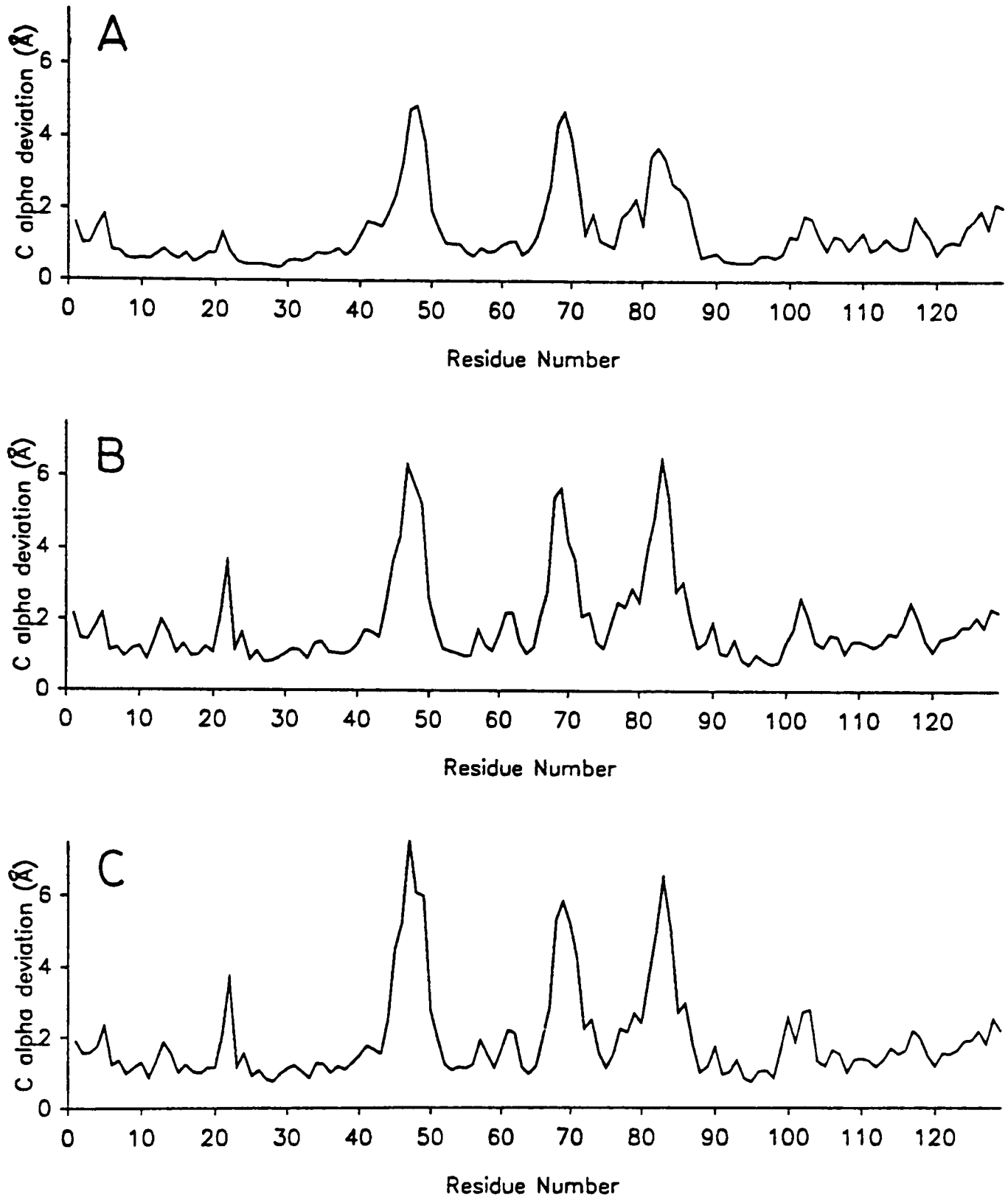
[2] # and * represent pairs of methylene protons and methyl groups respectively which have not been stereospecifically assigned.

Structure calculations carried out with and without the additional 164 NOE distance restraints from the longer mixing time experiments show that these additional NOE's have little overall effect on the definition of the calculated structures (the Cα RMSD with respect to the average for a set of structures calculated without the additional NOE data is 1.9 ± 0.3 Å). This is partly because the distance restraints used for these extra NOE's were chosen to be very loose, as discussed in the section 3.1.1, in order to take into account any spin diffusion effects. In addition, many of the extra NOE's, although they are seen in the longer mixing time experiments, are in fact short range, i.e. between residues which are close in the protein sequence; such NOE's will not significantly reduce the conformational space available to the protein unless the distance restraints are tight. This result emphasises the fact that it is not so much the total number of NOE's but their type and distribution throughout the sequence that is important in determining the definition of the calculated NMR structures (Wagner et al., 1987).

Global comparisons of the ensemble of calculated structures with the crystal structures of lysozyme show that the average pairwise Cα RMSD of the calculated structures with respect to the tetragonal type 2 crystal structure is 2.4 ± 0.3 Å (2.5 ± 0.4 Å for the comparison with the triclinic crystal structure) with the value reducing to 1.7 ± 0.2 Å if the 14 residues with large deviations are excluded. Figures 4.2B and C show the average Cα deviation of the calculated structures from the tetragonal type 2 and triclinic crystal structures respectively. There are three regions in the sequence where large average deviations are observed: residues Arg 45-Gly 49, Arg 68-Gly 71 and Ser 81-Leu 84. As these three regions also exhibit large average deviations when the ensemble of calculated structures are compared to the average structure (Figure 4.2A) they represent regions of poor definition across the ensemble of calculated structures rather than systematic differences from the crystal structure. We discuss these regions in more detail in the next section.

Figure 4.2.

The average Cα deviation of the calculated structures from the mean calculated structure (A), the tetragonal type 2 crystal structure (B) and the triclinic crystal structure (C) coordinates.



The region around Gly 22 also has larger than average deviations from the crystal structure. This appears to arise from problems with the conformation of glycine residues resulting from the potential used in the program XPLOR. We consider this further in a later section.

The ensemble of calculated structures was used to calculate an average coordinate set which was then energy minimized. The coordinates of this energy minimized structure have been deposited in the Brookhaven Protein Data Bank (accession number 1HWA). The Cα RMSD between this structure and the tetragonal type 2 crystal structure is 1.78 Å for all residues and 1.23 Å if we exclude the 14 residues with large deviations.

4.2.1 Main chain fold.

In order to investigate further the definition of the main chain fold in the ensemble of calculated structures, and to compare it to that observed in the crystal structures of hen lysozyme, the individual secondary structure units of the protein have been separately superimposed (Table 4.2). The four α helices (A 4-15; B 24-36; C 88-99; D 108-115) are well defined in the calculated structures (the Cα RMSD's from the average for the helices are A, 0.6 ± 0.2 Å; B, 0.2 ± 0.1 Å; C, 0.3 ± 0.1 Å; D, 0.6 ± 0.7 Å) and they have similar conformations to the regions in the crystal structures. The β sheet (41-60) and the loop region between residues 61 and 78 are also relatively well defined (Cα RMSD's from the average structure for these two regions are 1.4 ± 0.5 Å and 1.5 ± 0.3 Å respectively) as are the short region of β sheet, connecting the N-terminal residues 2-3 to residues 39-40, and the C-terminal 3_{10} helix, residues 120-124. Residues 79-84 (a 3_{10} helix in the crystal structures) however, adopt a range of conformations in the ensemble of calculated structures.

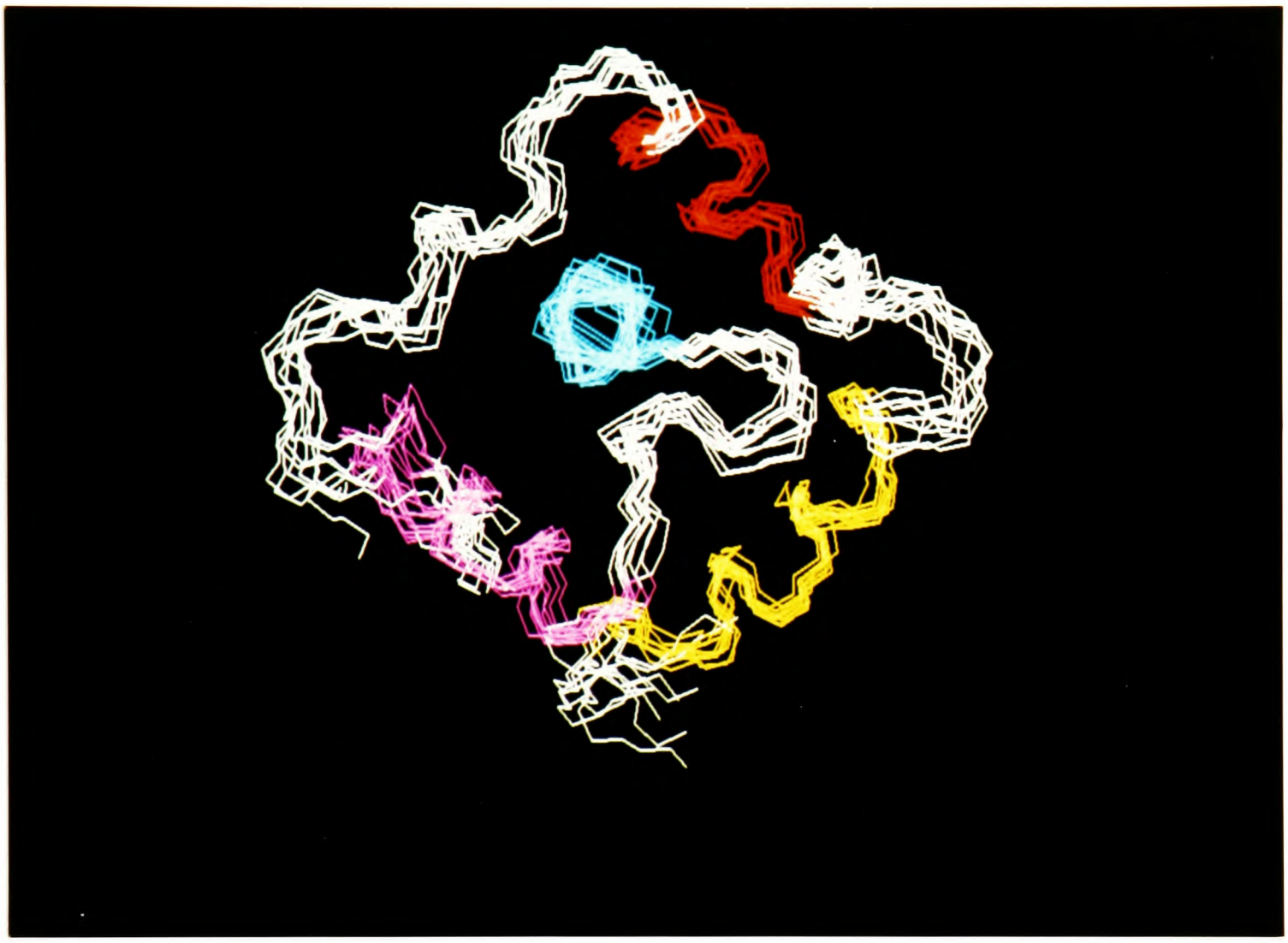
Table 4.2. Cα RMS differences for the solution structures of lysozyme.

	RMSD with respect to average (Å)	Average pairwise respect to crystal (Å)[1]	RMSD with
		Tetragonal	Triclinic
Whole structure (1-129)	1.8 ± 0.2	2.4 ± 0.3	2.5 ± 0.4
Helical domain (1-35, 85-129)	1.2 ± 0.2	1.6 ± 0.2	1.6 ± 0.2
Non-helical domain (36-84)	2.2 ± 0.4	3.0 ± 0.6	3.2 ± 0.7
Four α helices (4-15, 24-36, 88-99, 108-115)	0.8 ± 0.3	1.1 ± 0.3	1.1 ± 0.3
β sheet (41-60)	1.4 ± 0.5	1.8 ± 0.6	1.9 ± 0.5
Loop (61-78)	1.5 ± 0.3	1.9 ± 0.5	1.9 ± 0.5

[1] Cα RMSD of the tetragonal type 2 compared to the triclinic crystal structure 0.66 Å.

The four α helices of lysozyme provide the framework for one of the two domains of the structure. It is therefore of considerable interest that their relative orientations are very well defined in the structures and are closely similar to the crystal structures (see Table 4.2 and Figure 4.3). The other domain is dominated by the β sheet and the long 61-78 loop. Although the structure is locally well defined for both these structural units, they are not well oriented relative to each other or to the other domain. For this reason when the entire protein structures are superimposed the helices fit

Figure 4.3.
The helical domain of hen lysozyme, residues 1-35 and 85-129. A main chain superposition of ten of the calculated structures is shown. The four α helices (A-D) are coloured in magenta, cyan, yellow and red respectively. The fifth more irregular helical region is the 3_{10} helix near the C-terminus (residues 120-124).



well but large deviations are observed for residues of the sheet and loop regions (giving the maxima in the Cα deviation plot at residues 46-49 and 68-70). This is illustrated in Figure 4.4 where the superposition of the β sheet residues in the calculated structures is compared when the whole structure is superimposed (residues 1-129) and when only the β sheet residues (41-60) are superimposed.

To understand the reasons for the poor definition of the β sheet and loop relative to the rest of the structure the NOE data used in the structure calculations has been analysed in greater detail. Although NOE distance restraints involve residues throughout the protein sequence (see Figure 3.5) and there are adequate data to define the secondary structure units correctly (except for the 3_{10} helix which we discuss later), the long range NOE's are distributed unevenly along the sequence. There are 31 long range NOE's between one or other of the pair of helices A and B, and either helix C or helix D. These bring the N- and C- terminal regions of the structure together and comprise the helical domain. There are, however, relatively few NOE restraints connecting this domain to the loop and β sheet regions of the structure. For example, 24 NOE's define the contacts between the three strands in the β sheet region but only 8 longer range NOE's connect these strands to other regions of the protein structure. This is evident in Figure 4.5 by the blank areas of the NOE map between residues 40 to 80.

Interestingly, however, for both the loop and β sheet there are very few NOE's anticipated from the crystal structure which have not been seen in the NOESY spectra; the few that have not been identified are either too close to the diagonal or overlap with other peaks. Both these regions are in exposed positions in the protein structure (solvent accessibility data are given in Figure 4.6C) so there are fewer close contacts to other parts of the structure. In addition many of the residues in the loop and β sheet are those for which it is difficult to obtain conformational information from the NMR spectra

Figure 4.4.

The β sheet in hen lysozyme, residues 41-60. The main chain of ten of the calculated structures is shown superimposed on the tetragonal type 2 crystal structure (in bold). In A the whole structures are superimposed (1-129) while in B the superposition just involves the β sheet residues (41-60).

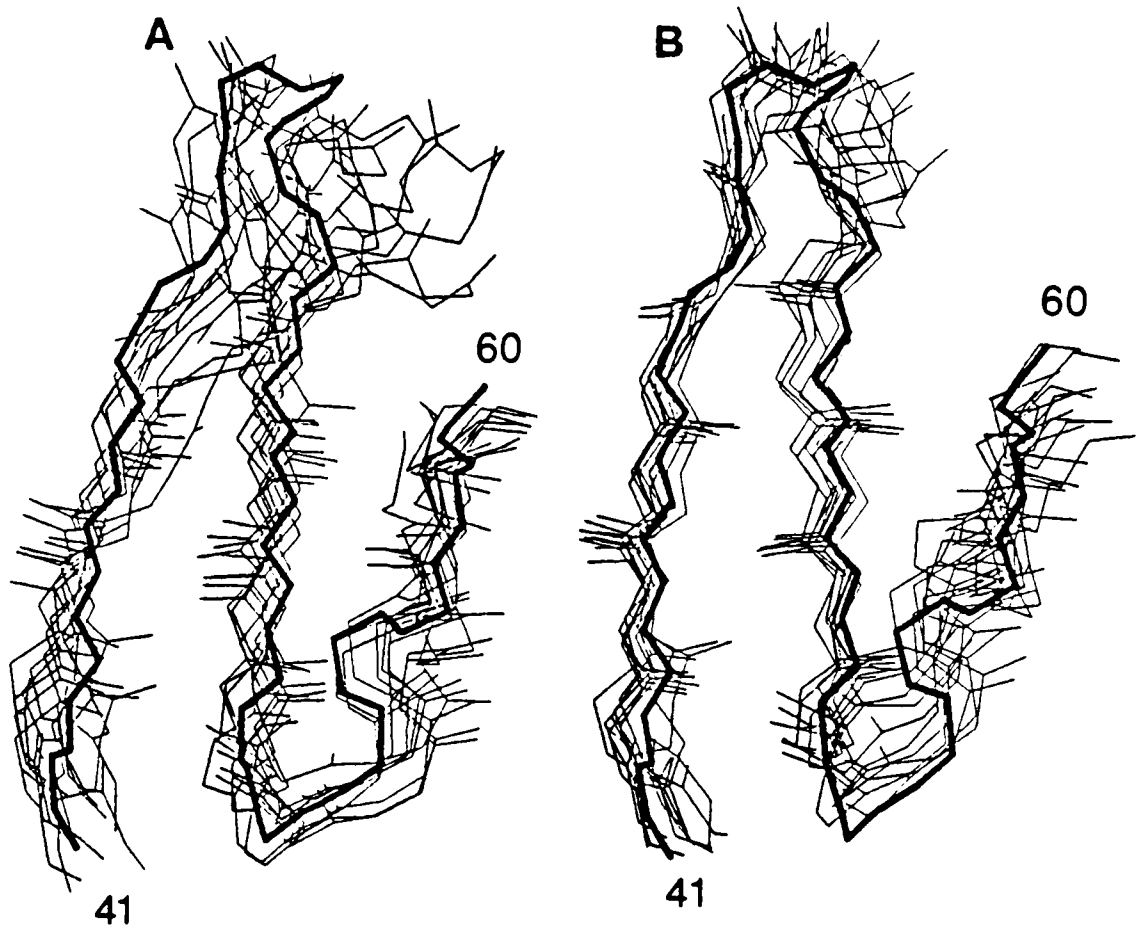
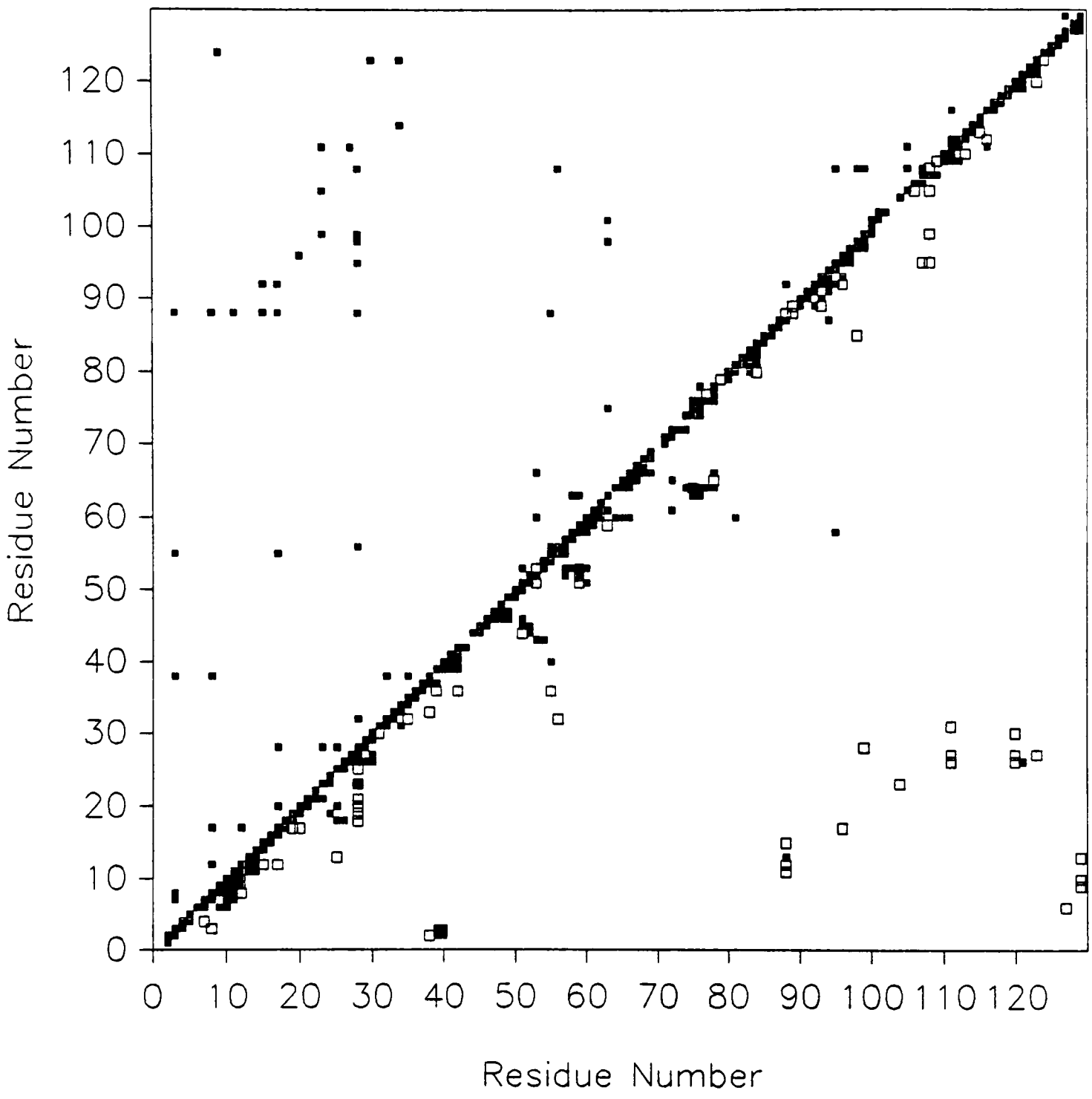


Figure 4.5.
Diagonal plot showing the pairs of residues between which an NOE effect has been observed. Below the diagonal main chain to main chain NOE's are shown by a filled box and main chain to side chain NOE's by an open box. Above the diagonal side chain to side chain NOE's are shown.



(e.g. glycine residues with their two α protons that have not been stereospecifically assigned) and there are few of the residues with side chain methyl groups that often give rise to many resolved, intense and therefore easily assignable NOESY cross peaks (e.g. Val, Leu, Ile). In the loop region (residues 61-78) there are 8 residues with either none or just one NOE restraint other than intraresidue and sequential NOE's (Gly 67, 71, Arg 68, 73, Thr 69, Pro 70, Asn 77, Trp 62). This compares with an average of 5.5 NOE restraints per residue, other than intraresidue and sequential NOE's, for the whole structure (each NOE restraint is counted twice in this analysis once for each contributing residue). The only residues in this region with a significant number of longer range NOE's (> 8) are the two cysteine residues (64 and 76) involved in disulphide bridges, and the adjacent residues, Trp 63 and Leu 75; these four residues are well defined in the protein structure (Figure 4.2).

For residues 79-84 the NMR data seem to be insufficient to define a unique structure. In 9 of the calculated structures the residues adopt a helical conformation although in 6 of these structures the helix is not tightly enough coiled for a 3_{10} helix; in the other 7 structures the conformation is highly disordered. An analysis of the NMR data shows that many of the NOE's that would define the 3_{10} helix (e.g. Ser 81CαH-Leu 84NH, Ala 82CαH-Leu 84NH) are present in the spectra but are weaker than would be predicted for a 3_{10} helix. This means that the corresponding distance restraints used are not tight enough to define uniquely the 3_{10} helix conformation. For example, the Ser 81CαH-Leu 84NH distance is restrained to the range 1.8-5.5 Å but the distance in a 3_{10} helix would be 3.3 Å (Wüthrich, 1986). The disorder in this region across the set of calculated structures appears to result therefore, at least in part, from the variations in the intensities of certain NOE's that has been mentioned in Chapter 3. In addition, because of this, helical hydrogen bond restraints which would

have ensured a helical conformation were not included for this region in the NMR structure calculations. It is important to note, however, that the NMR data are not inconsistent with the conformation seen in the X-ray structure for these residues; they simply are inadequate to define it clearly.

It can therefore be seen that although the structural information collected for lysozyme by NMR spectroscopy is not sufficient to define a unique structure in all parts of the sequence, in the regions where there are adequate data the calculated NMR structures provide a representation of the main chain structure of the protein in solution which is closely similar to that defined in the crystal structures (e.g. for the four α-helices the Cα RMSD from the average structure is 0.8 ± 0.3 Å). For the more poorly defined regions in the NMR structures other evidence obtained from NMR spectroscopy, particularly the existence of slow hydrogen exchange rates (Pedersen et al., 1991) and the lack of averaging of $^{3}J_{HN\alpha}$ coupling constants (Smith et al., 1991), suggests that the lack of definition in these regions arises, in the main, from insufficient NMR data rather than from large scale conformational disorder in solution.

It has been suggested, however, from theoretical analyses that hen lysozyme may undergo fluctuations in the relative orientations of the two domains in solution, a phenomenon known as hinge-bending (McCammon et al., 1976). Although this phenomenon is well established in other proteins (Dobson,1990), including T4 lysozyme (Faber & Matthews, 1990), there is little evidence for its existence in hen lysozyme. It is interesting, therefore, to use the calculated structures to investigate whether or not any of the observed disorder would be consistent with hinge bending. The two domains of lysozyme were compared separately to the tetragonal crystal structure and RMSD values calculated (see Table 4.2). The combined RMSD for the two domains is 2.1 ± 0.3 Å compared with a value of 2.4 ± 0.3 Å when the whole structure is superimposed. Although the RMSD is less for the two domains

separately that for the protein as a whole, the similarity in values shows that the hinge 'angle' adopted in solution is very similar to that in the crystal and that there is no convincing evidence for conformational disorder specifically about this hinge.

4.2.2 Comparison with crystallographic temperature factors.

It is interesting to compare the calculated NMR structures with evidence for disorder in the crystal structures; a plot of the average main chain temperature factors from the tetragonal type 2 crystal structure of lysozyme is shown in Figure 4.6. Two of the regions where the average Cα deviations of the calculated structures from the crystal structures are very large (> 3.5 Å - residues Thr 47 and Pro 70-Gly 71) have high temperature factors. High temperature factors (> 25.0Å^2) are also seen for residues Asp 101-Gly 102 and for the C-terminus, residues Gly 126, Arg 128-Leu 129. The Cα deviations of the calculated structures for these two regions are larger than the overall RMSD value for the whole structure, although the values are not as great as those for residues Thr 47-Gly 49 and Thr 69-Gly 71. For

---

Overleaf:

Figure 4.6.

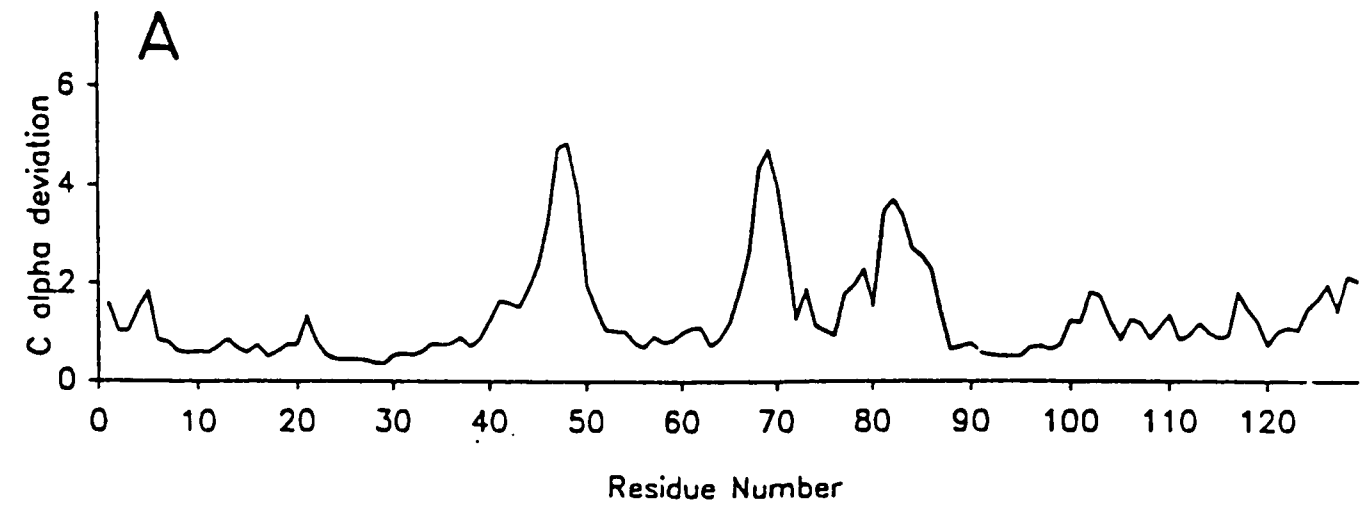
Plots showing for each residue in hen lysozyme :

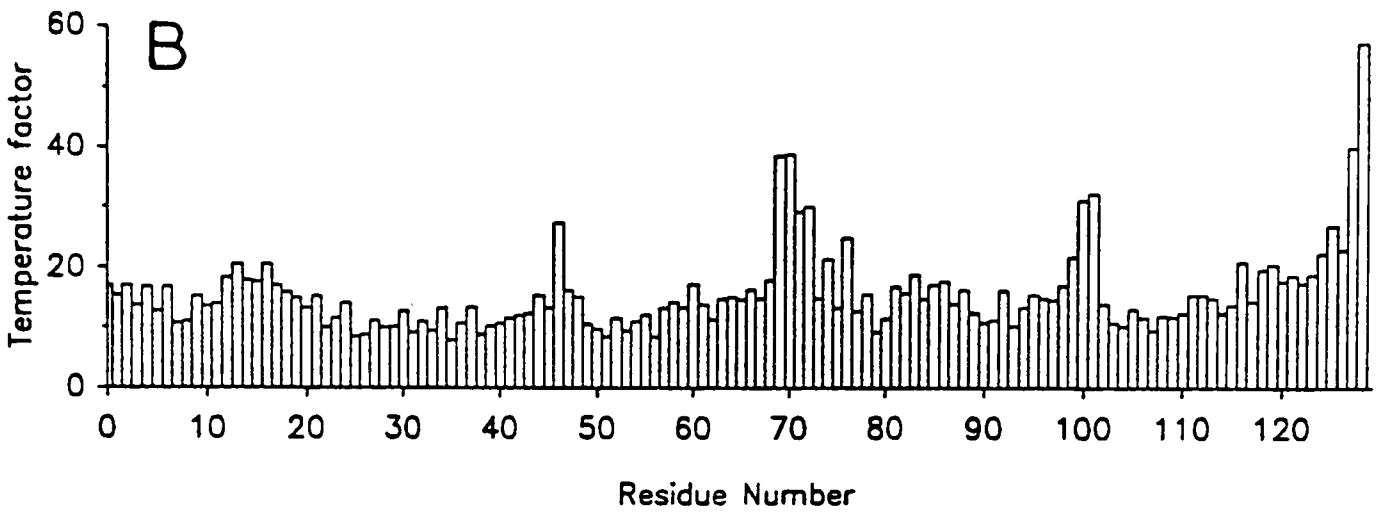
A The average Cα deviation (in Å) of the calculated structures from the mean NMR structures.

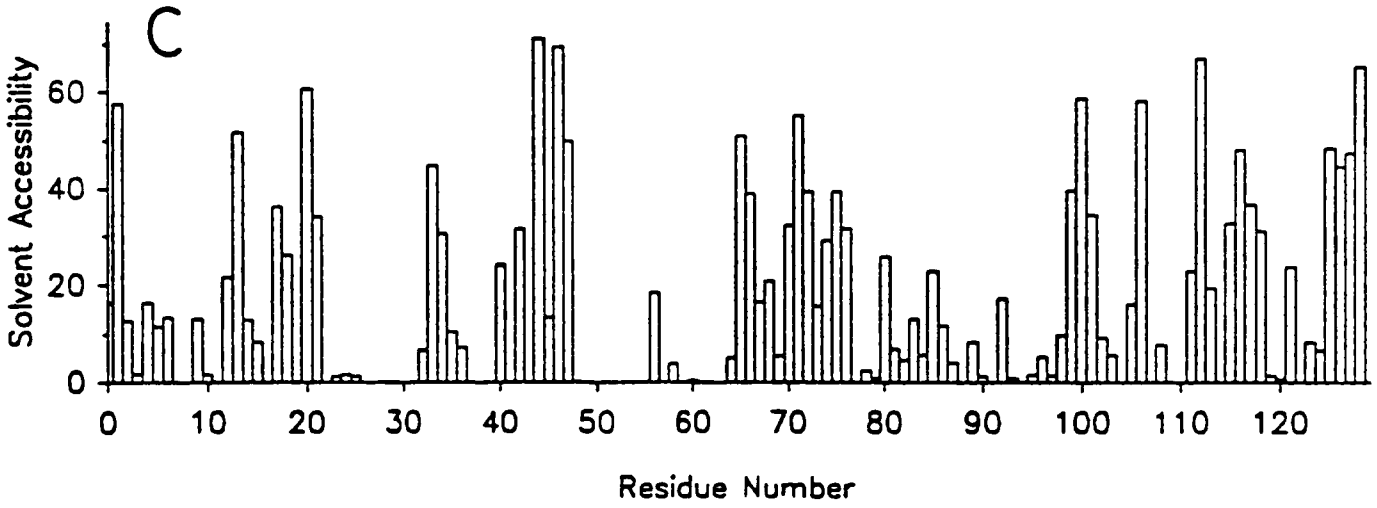
B The mean main chain temperature factors (in Å^2) in the tetragonal type 2 crystal structure.

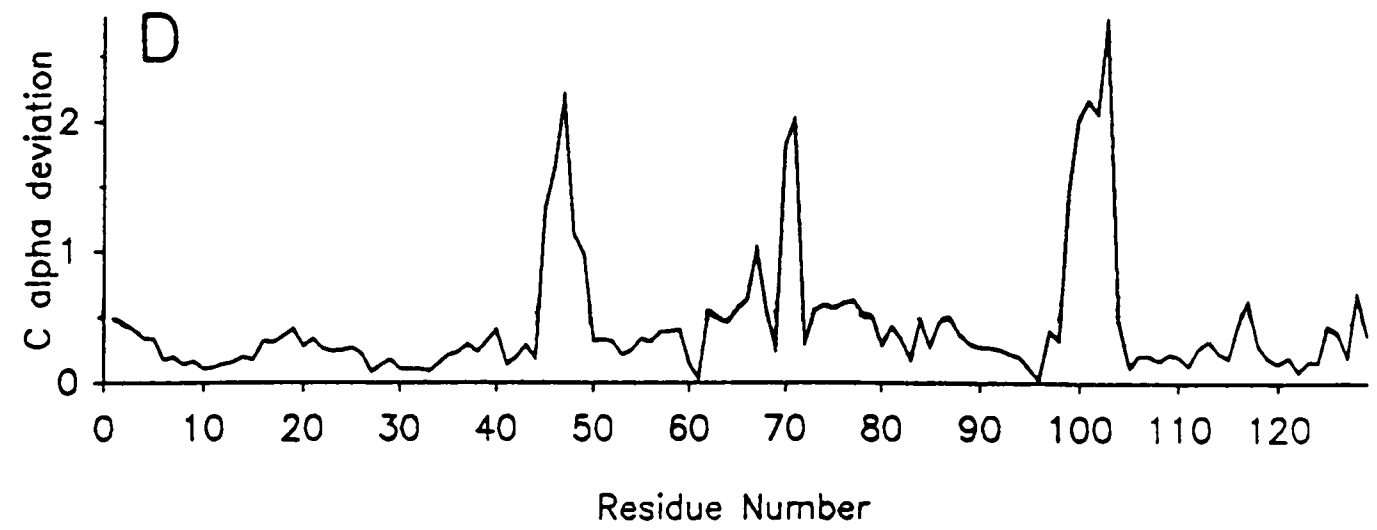
C The percentage main chain solvent accessibility calculated using the tetragonal type 2 crystal structure coordinates. Values are relative to the accessibility of the same residue in a fully extended conformation and were calculated using the method of Richmond & Richards (1978).

D The Cα deviation (in Å) between the tetragonal type 2 and triclinic crystal structures of hen lysozyme.


A
C alpha deviation
0
2
4
6
0 10 20 30 40 50 60 70 80 90 100 110 120
Residue Number


B
Temperature factor
0
20
40
60
0 10 20 30 40 50 60 70 80 90 100 110 120
Residue Number


C
Solvent Accessibility
0
20
40
60
0 10 20 30 40 50 60 70 80 90 100 110 120
Residue Number


D
C alpha deviation
0
1
2
0 10 20 30 40 50 60 70 80 90 100 110 120
Residue Number

residues Arg 128-Leu 129 the presence of a disulphide bridge between residues Cys 6 and Cys 127 must reduce the possibilities for variability in the structures. The 3_{10} helix between residues Ser 81-Leu 84 is not characterised by high temperature factors so, although this is not always formed in the calculated solution structures, it appears to be well defined in the crystal. It is possible, therefore, that where correlations are observed between RMSD values and temperature factors, these arise simply from the fact that both will tend to be higher for exposed residues than for internal residues.

A comparison of the tetragonal type 2 crystal structure of lysozyme with the triclinic crystal structure (see Figure 4.6) shows that three of the regions with high temperature factors discussed above, Arg 45-Asp 48, Pro 70-Gly 71, Val 99-Asn 103, also are defined to be in significantly different conformations in the two crystal structures. Interestingly, for both regions Arg 45-Asp 48 and Pro 70-Gly 71 these differences result at least in part from slightly different orientations of the loop and of the β sheet relative to the rest of the structure in the two crystal structures of the type seen in a much larger scale in the calculated NMR structures.

4.2.3 Side chain definition.

The structural analysis discussed so far has been concerned with the main chain conformations. We now consider the conformational properties of the side chains. The overall heavy atom RMSD of the structures from the average calculated structure is 2.3 ± 0.3 Å. Nevertheless, in the interior of the protein many of the side chains are well defined in the calculated NMR structures. For example, in 21 of the 34 residues (excluding Gly and Ala) where the side chain solvent accessibility is less than 10 %, a single conformation is adopted about χ_1 in at least 14 of the 16 calculated NMR structures; a χ_1 restraint was included for only five of these residues in the

structure calculations. Figure 4.7 shows the region of lysozyme within the helical domain known as the 'hydrophobic box'. The NMR structures and the crystal structures of the protein show the same set of hydrophobic aliphatic side chains and aromatic residues clustering together in this region. The orientation of 7 of the 8 residues shown is very well defined in the NMR structures (heavy atom RMSD of 0.91 ± 0.32 Å from the average structure for these side chains) reflecting the large number of NOESY cross peaks involving these side chains which have been assigned (an average of 26 per residue for the 8 side chains shown in Figure 4.7). The good definition of the entire helical domain of the protein in the calculated structures probably results, at least in part, from the well defined orientations of these side chains which generates the contacts between the helices (see Figure 4.3). The only disordered side chain shown in Figure 4.7, that of Tyr 20, is the most exposed in this region (solvent accessibility 39%).

On the surface of the protein, many of the side chains are very disordered in the NMR structures. For example, 88% of the side chains with a solvent accessibility greater than 60% (in the tetragonal type 2 crystal structure) are found in a range of conformations even about χ_1. This is not surprising as these side chains are restrained by very few NOE distance restraints (an average of 3 per residue) most of which are intraresidue, and therefore offer little information to define the structure. Furthermore, the analysis of $^3J_{\alpha\beta}$ coupling constants presented in Chapter 3 has shown that 55% of the side chains whose solvent accessibility is greater than 60% and whose coupling constants could be measured, are undergoing large scale motions about χ_1. The majority of the residues that are not well defined in the NMR structures are in any case likely, therefore, to be occupying multiple conformations in solution.

A particularly important region in the structures is, of course, the active site cleft lying between the two domains of the structure. As shown in

Figure 4.7.

A superposition of a representative sample of ten of the calculated NMR structures on the tetragonal type 2 crystal structure (in bold and in yellow) showing the eight residues in the hydrophobic box.

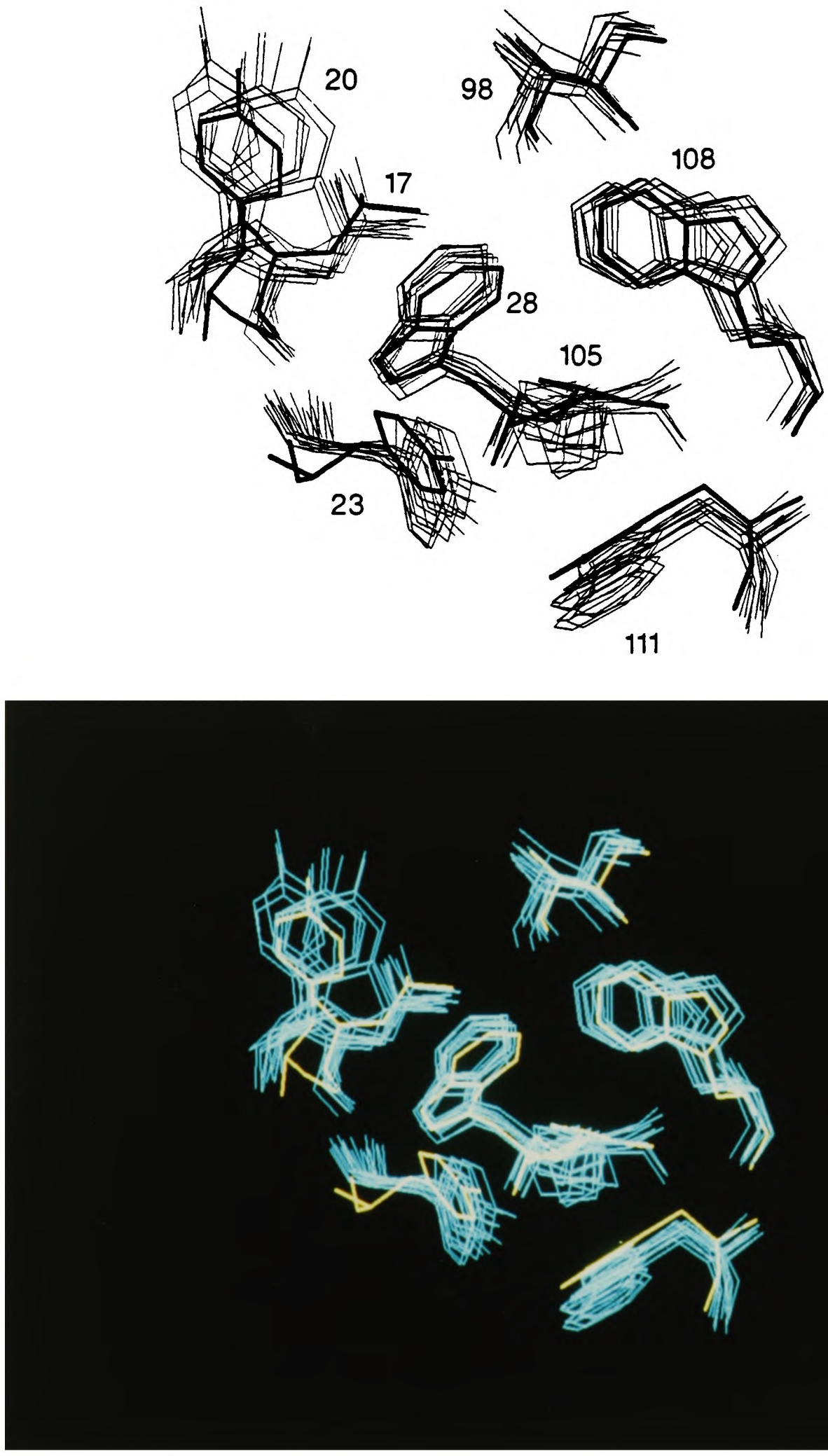


Figure 4.8.
A superposition of a representative sample of ten of the calculated NMR structures on the tetragonal type 2 crystal structure (in bold) showing the eight residues in the active site cleft.

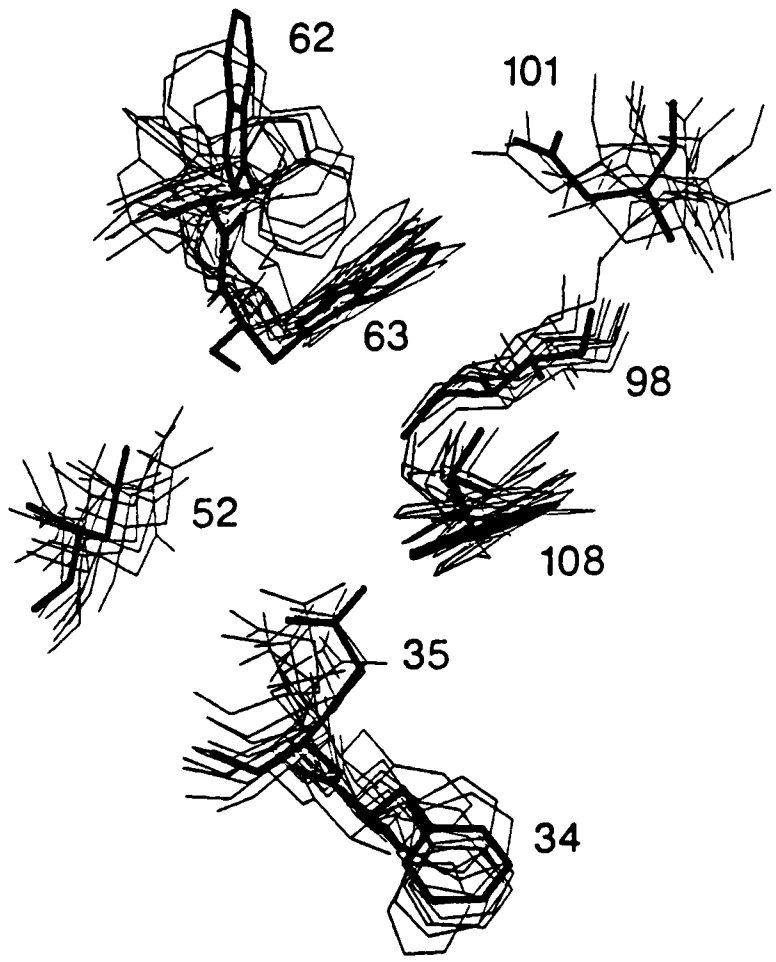


Figure 4.8 many of the residues in the active site critical for activity (Glu 35, Asp 52, Trp 63 & 108) are quite well defined in the NMR structures, and adopt similar orientations to those defined in the crystal structures. The heavy atom RMSD of the calculated structures from the average structure for the side chains shown in Figure 4.8 is 1.45 ± 0.18 Å; the larger value of this RMSD for this group of residues, compared to those of the hydrophobic box, is not surprising as many of the active site side chains are exposed to solvent (half have a solvent accessibility > 20%). For one residue, Trp 62, however, the side chain conformation is almost completely disordered in the NMR structures. Although this may reflect once again the difficulty of collecting sufficient NOE information about residues in highly exposed surface positions, the narrow resonance lines observed for this residue in the solution NMR spectra would be consistent with high mobility for the side chain of Trp 62 (Cassels et al., 1978). Additional NMR evidence in support of the high mobility of this side chain in solution has come recently from relaxation studies of uniformly ^{15}N labelled hen lysozyme (M. Buck, C. Redfield & C. M. Dobson, unpublished results). It is also interesting in this regard to note that diffraction studies have suggested that this residue may have a disordered conformation in the crystalline state (Grace, 1980). In general, however, the active site residues are better defined than many surface residues in the solution structures, and are close to the conformations seen in the crystalline state.

4.3 Comparison of the calculated structures with additional NMR parameters.

4.3.1 Coupling constants.

The experimental coupling constants for hen lysozyme have been used to give dihedral angle restraints that are used in the structure calculations. However, as a given coupling constant can in general arise from more than one dihedral angle, the restraints simply specify a range of possible angles rather than the precise value. In addition, as mentioned in the Materials and Methods section, no ϕ angle restraints are used when $5.5 < {}^3J_{HN\alpha} < 7.0$ Hz as there are several possible corresponding ϕ angles, and no χ_1 restraints are used for residues where stereospecific assignments have not been made. The coupling constant values can therefore be used to see if the correct torsion angle is being adopted from the range of angles allowed in the structure calculations. ${}^3J_{HN\alpha}$ and ${}^3J_{\alpha\beta}$ coupling constants were calculated, using the Karplus relationship described in Chapter 2, for each of the calculated NMR structures for lysozyme and the average of the values across the set of NMR structures were obtained. Figures 4.9 and 4.10 show the comparisons of these average calculated values with the experimental coupling constants for ${}^3J_{HN\alpha}$ and ${}^3J_{\alpha\beta}$ respectively. The RMSD values for these comparisons are given in Table 4.3; in both cases the residues studied were divided into two groups, those where angle restraints (ϕ or χ_1) have been used in the structure calculations, and those where the coupling constants have been measured experimentally but no angle restraints were used.

Figure 4.9.
A plot of the average $^3J_{HN\alpha}$ values calculated from the NMR structure coordinates versus the experimental $^3J_{HN\alpha}$ values. The residues where ϕ restraints were included in the structure calculations are shown by filled squares; those residues without ϕ restraints are shown by open squares.

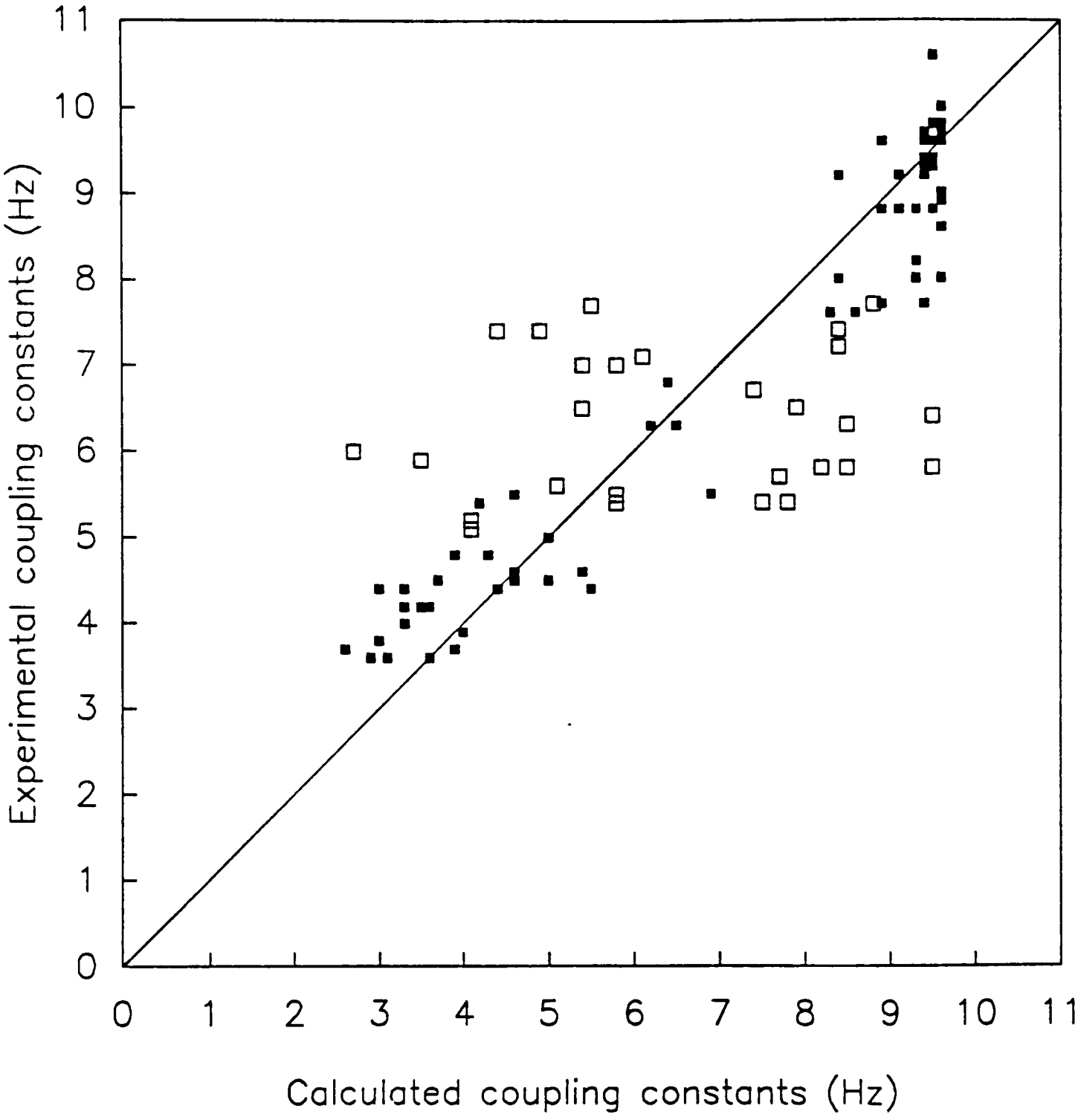


Figure 4.10.

A plot of the average $^3J_{\alpha\beta}$ values calculated from the NMR structure coordinates versus the experimental $^3J_{\alpha\beta}$ values. The residues where χ_1 restraints were included in the structure calculations are shown by filled squares; those residues without χ_1 restraints are shown by open squares.

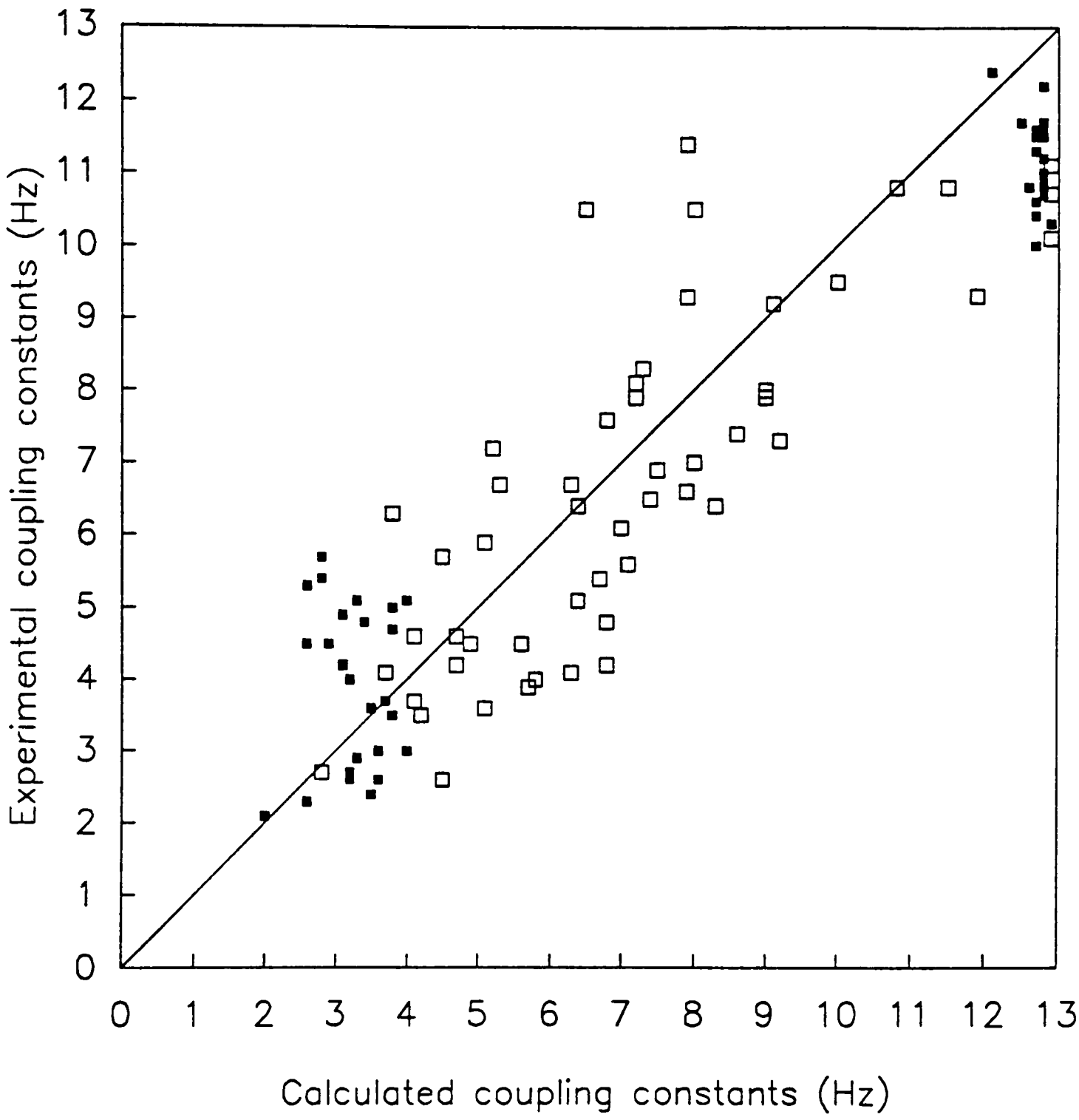


Table 4.3. Comparison of experimental NMR parameters with those calculated from crystal and NMR structures of lysozyme.

	RMSD between calculated and experimental values.		
Structure	$^3J_{HN\alpha}$ (Hz)	$^3J_{\alpha\beta}$ (Hz)	**Ring current shifts (ppm)**
Crystal[1]	0.87 (0.93)[2]	2.52 (0.87)	0.33 (0.88)
E. minimized crystal	1.61 (0.83)	2.69 (0.86)	0.22 (0.96)
Average of NMR[3]			
All residues	1.26 (0.86)	1.55 (0.92)	0.33 (0.86)
Angle restraints	0.73 (0.96)	1.53 (0.97)	
No angle restraints	2.02 (0.08)	1.56 (0.80)	

[1] Tetragonal type 2 crystal structure.

[2] Correlation coefficient for the comparison given in parentheses.

[3] Values are the mean of the values calculated for each of the set of NMR structures.

For $^3J_{HN\alpha}$ coupling constants, the overall agreement between experimental and calculated main chain coupling constant values is less good than for the comparison with the crystal structure although for the residues where ϕ restraints were used there is, as expected, a very good agreement between the average and experimental $^3J_{HN\alpha}$ values (Table 4.3).

For the other set of residues with no ϕ restraints, there is a very poor correlation; the calculated coupling constant values for this set are either very large or very small. This appears to be a result of the potential energy function used in the program XPLOR (Brünger, 1988) which favours ϕ angle conformations of 0^o, 120^o or -120^o (corresponding to $^3J_{HN\alpha}$= 2.8 or 9.7 Hz). This potential will dominate in the molecular dynamics calculations if the residue is little restrained by NMR data. Hence, these ϕ conformations are also found to be adopted by many of the glycine residues, e.g. Gly 67 and 71 in the loop region of the structure. In the solution structures, therefore, reliable definition of the main chain torsion angles requires ϕ restraints for the NOE data set used here.

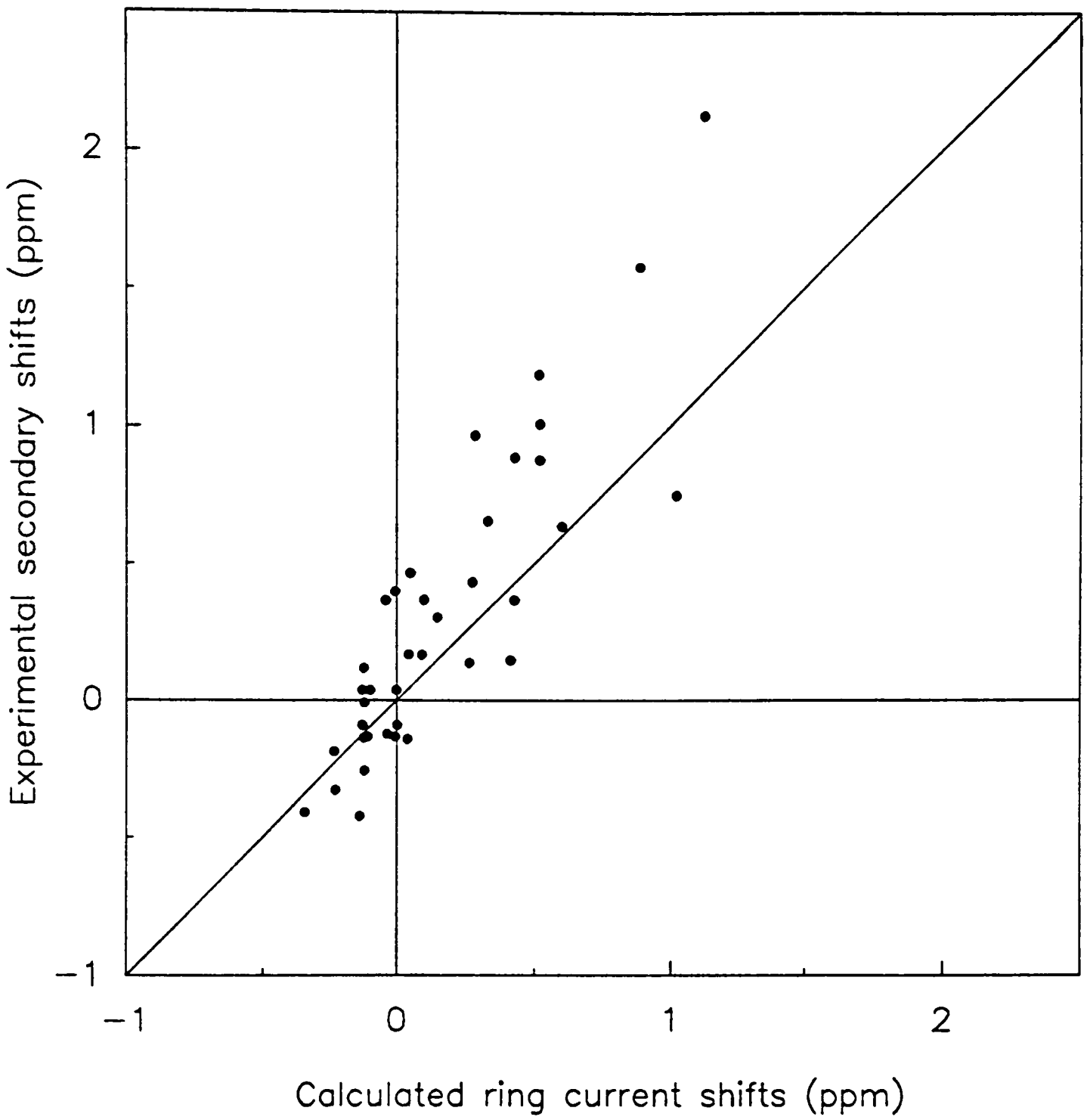
For $^3J_{\alpha\beta}$, although the agreement between the average calculated and experimental values is not very good, it is actually better than that between the experimental values and those calculated using χ_1 angles from the tetragonal type 2 crystal structure or from the energy minimized crystal structure (Table 4.3). As for the ϕ dihedral angles, where χ_1 is restrained in the structure calculations there is good agreement between the average calculated $^3J_{\alpha\beta}$ values and the experimental $^3J_{\alpha\beta}$ reflecting the fact that there are no violations of the χ_1 restraints greater than 2^o. The residues where χ_1 is not restrained generally have motionally averaged experimental $^3J_{\alpha\beta}$ values and are on the surface of the protein; the lack of NOE's for surface groups mean that a range of χ_1 values are found in the calculated structures, the values clustering around the three staggered rotamer positions (χ_1= -60^o, 60^o, 180^o) favoured by the intramolecular potentials. The mean calculated $^3J_{\alpha\beta}$ values in these cases average to values close to the experimental motionally averaged $^3J_{\alpha\beta}$ values (6-8 Hz) hence contributing to better correlation with the experimental data than the crystal structure or indeed any single conformation. The one interior residue which has an experimental motionally averaged $^3J_{\alpha\beta}$ value (of 6.3 Hz) is Val 99. As

discussed in Chapter 3, in the tetragonal type 2 crystal structure of hen lysozyme this side chain adopts a conformation with a χ_1 value of 80.3° while in the triclinic structure χ_1 is 171.1°. The average calculated structure value is lower (3.8 Hz) than that observed experimentally but it is interesting that two ranges of χ_1 are adopted in the calculated structures; 14 of the structures adopt χ_1= 60-80° while 2 of them have χ_1 = 160-170 °.

4.3.2 Ring current shifts.

Ring current shifts were calculated, for each of the NMR structures, for methyl groups where the experimental secondary shifts were greater than 0.1 ppm (38 methyl groups). For methyl groups where the experimental secondary shifts were less than 0.1 ppm, there were no cases where there were large calculated shifts for the NMR structures. Ring current shifts for pairs of methylene protons were not used in the analysis because in general stereospecific assignments have not been carried out. A plot of the average of the calculated ring current shifts versus the experimental secondary shifts is shown in Figure 4.11 and the RMSD values are given in Table 4.3. In general, there is a good agreement between the calculated and experimental values reflecting in part the good definition of aromatic and other side chains in the hydrophobic box of lysozyme where 40% of the aromatic side chains are found (see Figure 4.7). Certain cases, where large discrepancies are found for residues in individual structures, can be rationalised in terms of differing orientations of side chains. For example, Trp 108 provides the major contribution to the ring current shift of the methyl group of Ala 107. The experimentally observed secondary shift for this methyl group is 0.75 ppm. For the tetragonal type 2 crystal structure the predicted ring current shift for Ala 107 Hβ is 0.50 ppm, while in the calculated NMR structures, although the average value is 1.02 ppm, individual values range from 0.21 to 1.53 ppm. These differing values have been found to result from the different relative

Figure 4.11.
A plot of the average ring current shift values calculated from the NMR structure coordinates versus the experimental secondary shift values.



orientations of the residues Trp 108 and Ala 107. For the $C\varepsilon H_3$ group of Met 105 the experimentally observed secondary shift is 2.13 ppm and the ring current shift value predicted from the crystal structure is 1.37 ppm. The average value for the predicted ring current shift for the NMR structures is 1.13 ppm but there is a large variation for the individual values, ranging from -0.08 to 2.12 ppm. In this case there is no simple structural correlation with these values as there are many different contributions that result in the single ring current shift value (contributions from Tyr 23, Trp 28, Trp 108 and Trp 111). It is interesting that both of these examples involve residues in the hydrophobic box of lysozyme which is well defined in the calculated structures (side chain heavy atom RMSD of 0.91 ± 0.32 Å, see Figure 4.7) stressing the sensitivity of ring current shifts to even small conformational differences. This sensitivity of ring current shifts to slight changes in structure has been shown previously when chemical shifts have been calculated for the coordinate sets of a molecular dynamics simulation (Dobson & Karplus, 1986).

4.4 Conclusions.

The ensemble of the solution structures calculated from the NMR data reported in this chapter shows close agreement with the structures reported for hen lysozyme in crystals. Despite the extensive overlap of resonances in the 2D spectra, and the problems associated with the quantification of the NOE data, the RMSD values from the average for the large majority of the residues are less than 1.3 Å, indicating a well defined structure. The helical domain is particularly well defined, including not only the four α helices but also a distorted 3_{10} helix and the loops joining these regions together. The definition of the other domain is less good overall. It is interesting, however, that some structural features in this part of the protein structure are at least locally well defined (e.g. the triple stranded antiparallel β sheet)

although others are not. In particular, the 3_{10} helix encompassing the residues 79-84 is not consistently present in the family of solution structures. Furthermore, in this domain the orientations of even the locally well defined regions of the structure are not well defined relative to the rest of the molecule.

There is no evidence in the present study that the disorder of the main chain in the β-domain for the structures calculated from the NMR data is a realistic representation of the distribution of conformational states in solution. The fact that the main chain trace, both of the independent crystal structures of hen lysozyme and indeed of crystal structures of all lysozymes of related sequence, are closely similar suggests that the disorder in the solution structure could well arise from the lack of adequate restraints. The absence of a well defined 3_{10} helix involving residues 79-84 (seen in all crystal structures) in the solution structures can, for example, at least be partly attributed to the fact that there was insufficient experimental data to enable hydrogen bond restraints to be used in the structure calculations for this region. Alternative strategies involving heteronuclear NMR experiments are, however, becoming available both to enable more NOE data to be obtained and to identify regions of high conformational flexibility provided isotopic labelling can be achieved (Kay et al., 1989; Clore & Gronenborn, 1991). Indeed, the success of these heteronuclear techniques for the study of proteins of this size is illustrated in Chapters 5-7 of this thesis where the NMR structure determination for human interleukin-4 is presented. For hen lysozyme, recombinant protein uniformly labelled with ^{15}N has been recently produced using an *Aspergillus niger* expression system (Roberts et al., 1992) and preliminary heteronuclear NMR experiments are in progress (M. Buck, C. Redfield & C. M. Dobson, unpublished results).

Although the main chain fold of the protein defined in the solution structures is extremely close to that found in crystals, for side chains the

position is more complex. Analysis of the side chains in the NMR structures shows that those in the interior of the protein are well defined in solution and their orientations are similar to those found in crystals. For example, in the hydrophobic box region in the helical domain the side chain heavy atom RMSD from the crystal structure is 0.91 ± 0.32 Å. In addition, in the active site cleft of the protein the same residues have been found to be extending out into the cleft with similar orientations as in the crystal structure. There is considerable disorder however, in the calculated NMR structures for the side chains of exposed residues on the protein surface. This disorder has been recognised for lysozyme from the averaging of $^3J_{\alpha\beta}$ coupling constants (as discussed in Chapter 3) and also has been predicted from molecular dynamics simulations of the protein (Post et al., 1986). The calculation of NMR structures presented here gives some insight into the possible characteristics of this side chain disorder for surface groups, which is much greater than the crystal structures of the protein would suggest.

Overall, the work on hen lysozyme presented in this thesis has enabled a detailed description of the structure of the protein in solution to be given. It has confirmed that all the major structural features of the crystal structures of the protein are preserved in solution, although many surface side chains have been shown to be extensively disordered in solution. In the light of the continuing interest in lysozyme as a model protein in attempts to understand the principles of folding, dynamics and enzyme mechanisms in general, these conclusions are particularly important.

CHAPTER 5.

THE STRUCTURE OF HUMAN INTERLEUKIN-4.

Description of the NMR structure determination.

The molecular weight of human interleukin-4 (in its nonglycosylated form) is comparable to that of hen lysozyme but, whereas the structure of hen lysozyme contains considerable regions of both β sheet and helix, IL-4 contains only helical secondary structure (with the exception of a short six residue region of β sheet). This means that the chemical shift dispersion and consequent resolution in 2D spectra of the unlabelled protein is much reduced in the case of IL-4 and it is not possible to use the data available from these spectra alone to determine a high resolution structure of the protein. Therefore, for IL-4 a strategy involving isotopically labelled samples of the protein and heteronuclear 3D NMR experiments has been employed. The strategy has two main steps :

a) Experiments on a uniformly ^{15}N labelled sample.

2D and 3D experiments were used to assign the majority of ^{15}N and ^{1}H resonances in the spectrum of IL-4 and to identify the secondary structure and overall fold.

b) Experiments on double labelled (^{15}N, ^{13}C) sample.

Further 3D experiments on a double labelled sample, enabled additional NOE's to be assigned, particularly long range NOE's between side chains. Using the complete NOE data set, high resolution structures for the protein have been calculated.

In this chapter the procedures used in the determination of the structure of human IL-4 are described. A detailed analysis of the structure is then given in Chapter 6 and in Chapter 7 the structure of IL-4 is compared with related protein folds, and its significance is discussed. A postscript to the thesis compares the NMR structure of IL-4 with a crystal structure of the protein

which has been solved after the NMR studies were completed (Wlodawer et al., 1992).

5.1 Sample labelling and preparation.

Fully enriched isotopically labelled samples of IL-4 were produced by SmithKline Beecham Pharmaceuticals. The recombinant IL-4 was expressed using a synthetic gene in bacterial (*E. Coli.*) cells. For the ^{15}N labelled protein the *E. Coli* was grown on a minimal medium incorporating ^{15}N labelled $(NH_4)_2SO_4$ as the sole nitrogen source while for the double labelled (^{13}C and ^{15}N) sample a minimal medium incorporating ^{15}N labelled $(NH_4)_2SO_4$ and uniformly ^{13}C labelled glycerol as the sole nitrogen and carbon sources respectively was used. The recombinant protein differs from native human IL-4 by the addition of an N-terminal methionine residue, referred to in this work as residue '0', and the absence of carbohydrate; the recombinant protein is, however, biologically fully active. All the 3D NMR experiments were performed on a sample of 2 mM protein at either pH 5.3 and 35°C or pH 4.5 and 20°C.

5.2 ^{15}N labelled sample.

The assignment of the majority of ^{15}N and ^{1}H resonances in the spectrum of IL-4 was achieved using the information contained in three ^{1}H-^{15}N 3D data sets together with spin system information from 2D spectra of the unlabelled protein. These 3D experiments were a ^{1}H-^{15}N NOESY-HMQC (Messerle et al., 1989) collected with a mixing time of 200 ms, a ^{1}H-^{15}N HOHAHA-HSQC (Driscoll et al., 1990) collected with an isotropic mixing time of 26 ms and a ^{1}H-^{15}N HSQC-NOESY-HSQC which allows distinction of NH-NH NOE connectivites between overlapping amide resonances (Frenkiel et al., 1990; Ikura et al., 1990a). Further analysis of 2D and 3D NOESY spectra of IL-4 then led to the identification of a total of 1360 NOE's. In addition,

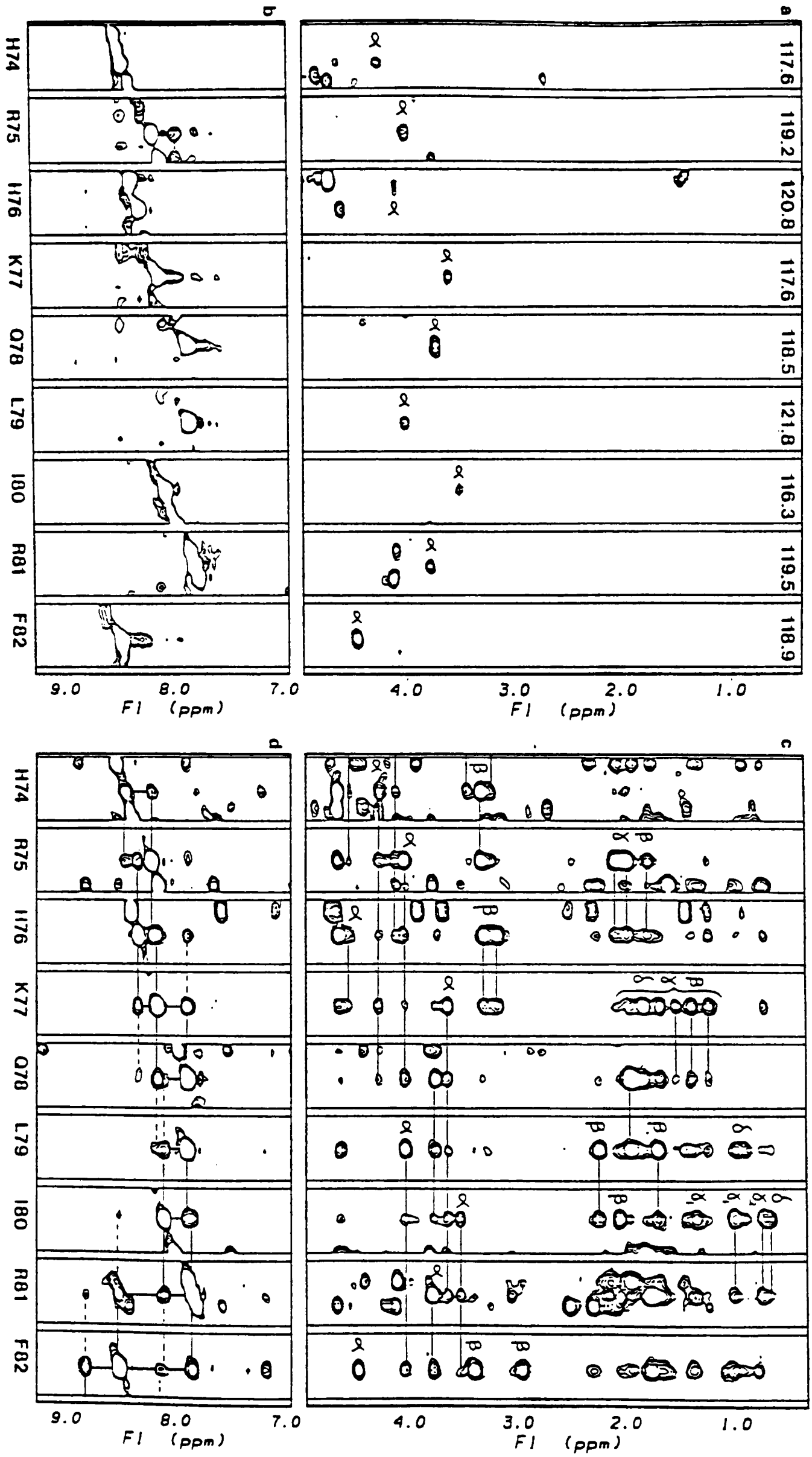
main chain $^3J_{HN\alpha}$ coupling constants were extracted from cross-sections in a 2D ^1H-^{15}N HMQC experiment which contained a final $\pi/2(y)$ ^1H pulse (Kay & Bax, 1990) and amide exchange rates were estimated using data from three 2D ^1H-^{15}N HSQC experiments which were recorded consecutively. All these experiments on the ^{15}N labelled sample of IL-4 were performed and analysed by Dr. C. Redfield (Redfield et al., 1991). Examples of the spectra are given in Figure 5.1. The following sections (5.2.1 to 5.2.4) describe my use of the data from these experiments in the determination of preliminary structures of human IL-4.

Overleaf:

Figure 5.1.

3D NMR spectra of ^{15}N labelled human IL-4 (from Redfield et al. (1991)).

a and b show strips taken from the 3D HOHAHA-HSQC spectrum at the ^{15}N chemical shifts of residues 74 to 82. The intraresidue NH-CαH peak are labelled in a and the NH diagonal peaks are shown in b. The corresponding strips from the 3D NOESY-HMQC spectrum are shown in c and d.. Sequential NH-NH and CαH-NH connectivities are indicated by solid lines together with NH-NH (i,i+2), CαH-NH(i,i+2), CαH-NH(i,i+3) and CαH-NH(i,i+4) peaks which are typical of helical secondary structure. Intra and interresidue NOE's between NH and side chain protons are also labelled.


a
b
c
H74
R75
I176
K77
Q78
L79
I80
R81
F82
117.6
119.2
120.8
117.6
118.5
121.8
116.3
119.5
118.9
9.0
8.0
7.0
4.0
3.0
2.0
1.0
F1 (ppm)
α
β
γ
δ

5.2.1 Secondary structure identification.

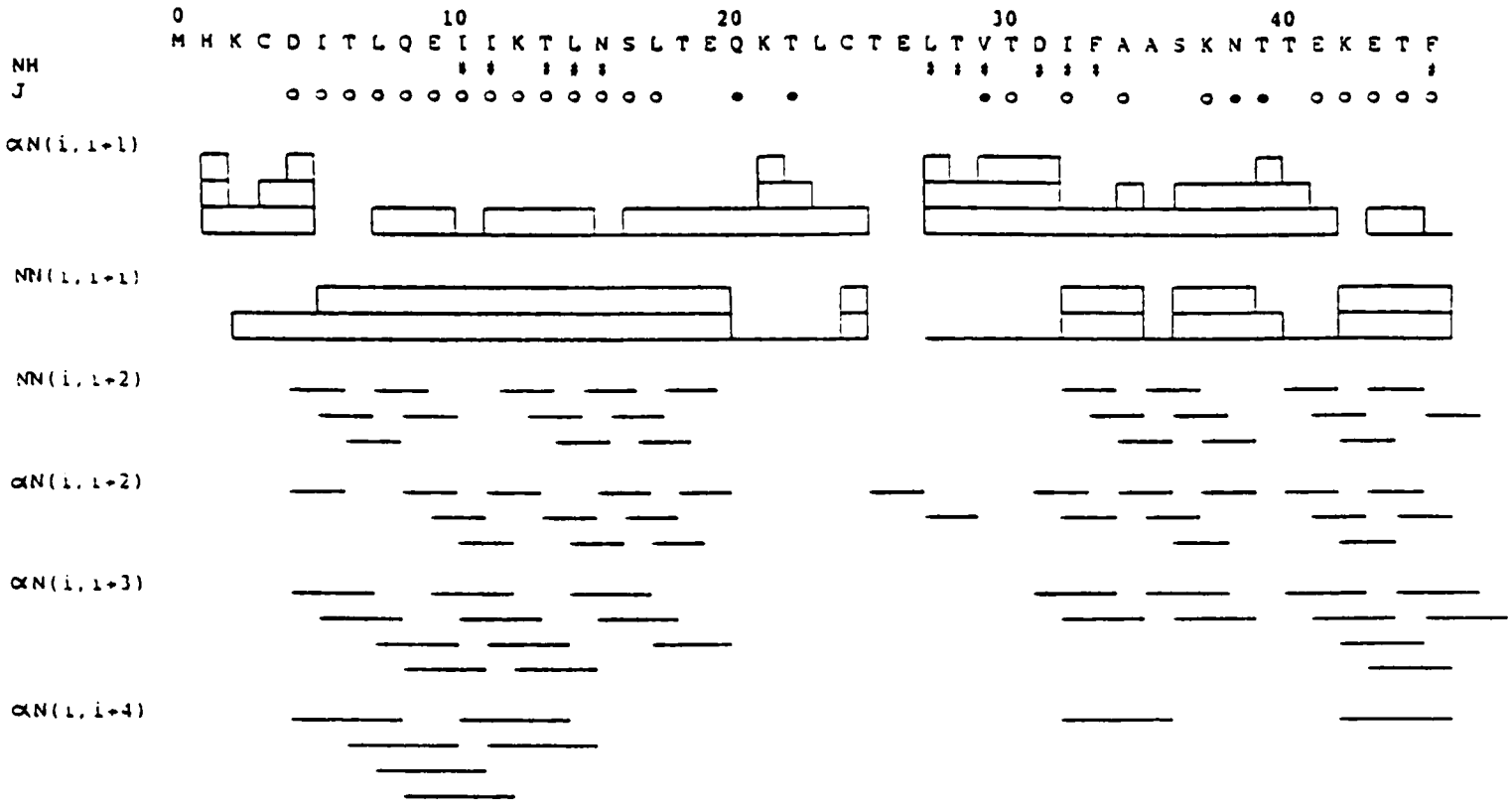
A combination of three types of NMR data namely NOE's, $^3J_{HN\alpha}$ coupling constants and amide protons exchange rates, are generally used to identify the secondary structure of a protein by NMR methods (Wagner, 1990). These data are summarised in Figure 5.2 for human IL-4. For four regions of the protein (residues 5-17, 41-57, 72-90 and 110-125) there are strong sequential NH-NH NOE's and also many short and medium range NOE's of a type that is indicative of a helical secondary structure i.e. NH-NH(i,i+2), CαH-NH(i,i+2), CαH-NH(i,i+3) and CαH-NH(i,i+4) NOE's (Wüthrich, 1986). Study of the $^3J_{HN\alpha}$ values shows that for all the residues in three of these ranges (5-17, 72-90 and 110-125) the measured values are less than 6 Hz and hence also indicate helical conformations ($^3J_{HN\alpha}$ data listed in Appendix V). In the range 41-57 all the residues that have been assigned also have a $^3J_{HN\alpha}$ value less than 6 Hz, but there is a missing assignment in the centre of the helix (Arg 53). The observation of a 52CαH-55NH NOE of medium intensity and the hydrogen exchange data discussed below, suggest that a continuous helix runs from residues 41-57 but the possibility of some irregularity in its structure could not, at this stage in the analysis, be eliminated. The NOE and

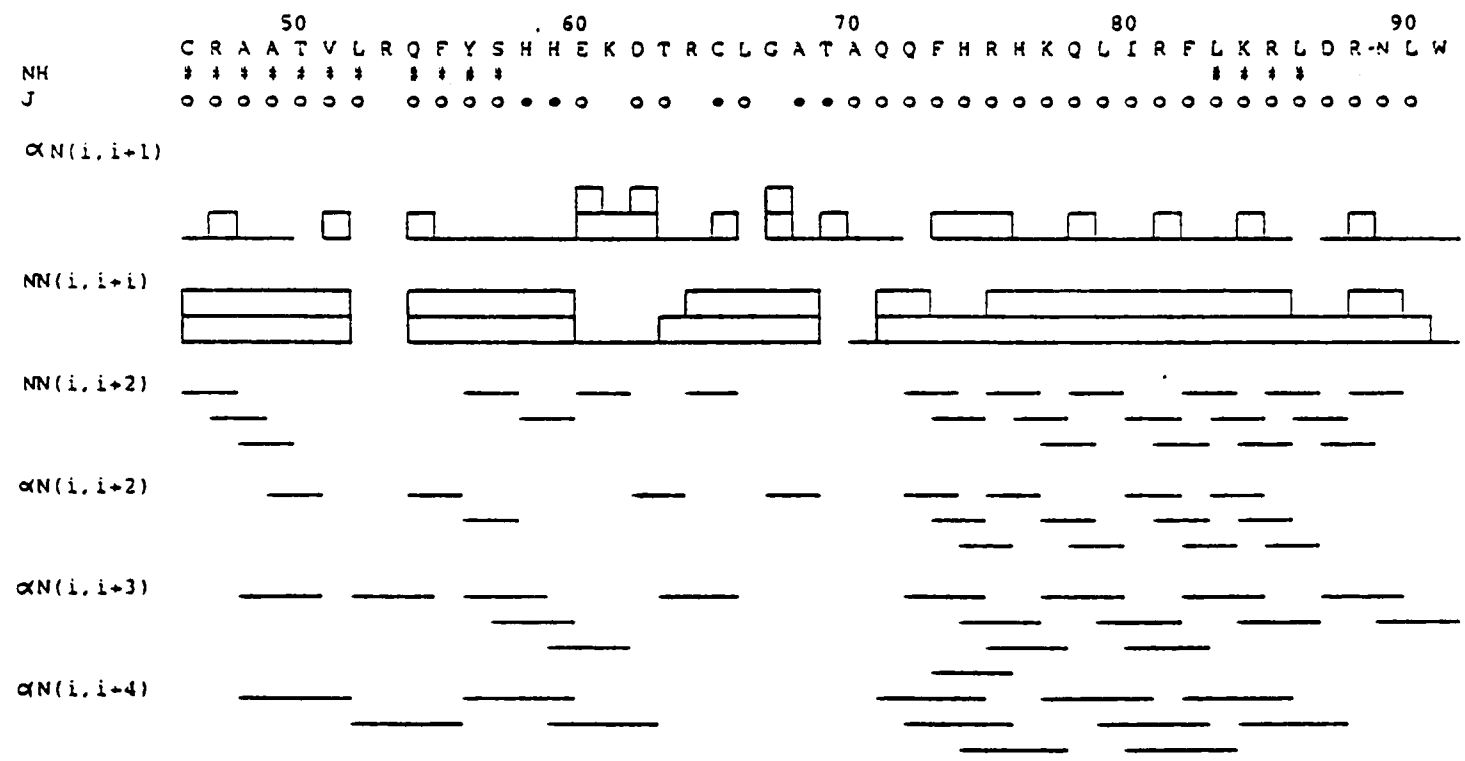
---

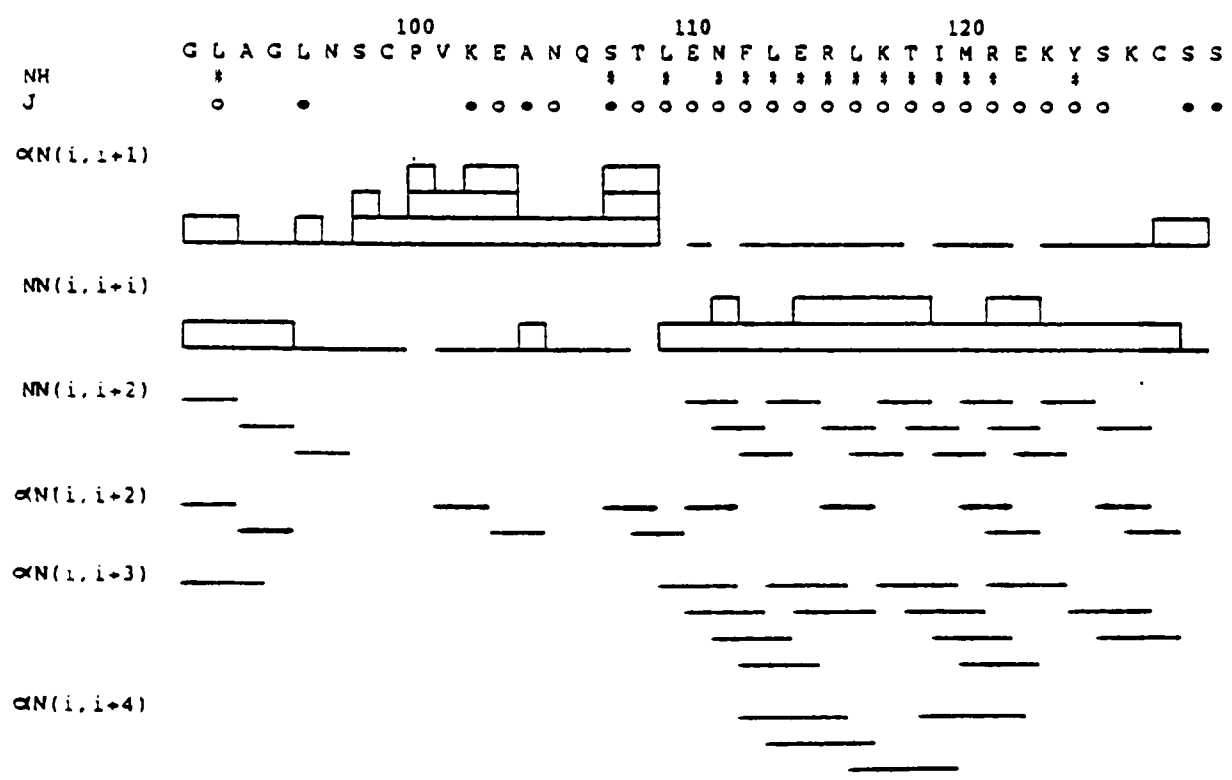
Overleaf:

Figure 5.2.

The sequence of human IL-4 together with a summary of the NMR data defining the secondary structure of the protein. For the short range NOE's involving NH and CαH protons, the NOE intensities are indicated by the width of the bars. Residues with amide protons that are slow to exchange (i.e. that are not fully exchanged 1 h after the protein is dissolved in D_2O at pH 4.5 and 20°C) are shown by #. Residues with $^3J_{HN\alpha} < 6$ Hz are indicated by open circles; those with $^3J_{HN\alpha} > 8$ Hz by solid circles.


0 10 20 30 40
M H K C D I T L Q E I I K T L N S L T E Q K T L C T E L T V T D I F A A S K N T T E K E T F
NH
J
αN(i,i+1)
NN(i,i+1)
NN(i,i+2)
αN(i,i+2)
αN(i,i+3)
αN(i,i+4)


50 60 70 80 90
C R A A T V L R Q F Y S H H E K D T R C L G A T A Q Q F H R H K Q L I R F L K R L D R N L W
NH
J
αN(i,i+1)
NN(i,i+1)
NN(i,i+2)
αN(i,i+2)
αN(i,i+3)
αN(i,i+4)


100 110 120
G L A G L N S C P V K E A N Q S T L E N F L E R L K T I M R E K Y S K C S S
NH
J
αN(i,i+1)
NN(i,i+1)
NN(i,i+2)
αN(i,i+2)
αN(i,i+3)
αN(i,i+4)

$^3J_{HN\alpha}$ data therefore identify four main helical regions in IL-4; helix A (5-17), helix B (41-57), helix C (72-90) and helix D (110-125).

The amide protons that do not exchange within 10 min of the protein being dissolved in D_2O are listed in Table 5.1. Slowly exchanging amides are very frequently, although not exclusively, indicative of persistent secondary structure. For IL-4, 36 of the 46 residues listed in Table 5.1 are in the four identified helical regions. Helices B and D have large numbers of slowly exchanging amide protons in them (12 in each helix) but for helix A and particularly helix C, there are less slowly exchanging amides, perhaps suggesting that these helices are more exposed. For helix C in particular, this makes it more difficult to define the start and end points of the helix; this problem and the exact lengths and positions of all the helices are discussed further in section 6.1.1.

There are two short regions of the protein (residues 28-30 and 106-108) between which NOE's typical of an antiparallel β sheet have been identified (listed in Table 5.2 and shown in Figure 5.3). CαH-NH NOE's have also been recognised for these residues, but only two of the residues in the identified strands have large measured $^3J_{HN\alpha}$ values (Val 29 and Ser 107 - $^3J_{HN\alpha}$ values of 9.1 Hz and > 7 Hz respectively) so the sheet must not be very regular. Study of the amide proton exchange rates shows that the two residues that would be expected to form hydrogen bonds in this β sheet (29 and 107) have slowly exchanging amide protons (see Figure 5.3).

Table 5.1 Amide protons that are slow to exchange in human IL-4.

Residue	Rate*	Position#	Residue	Rate	Position
Glu 9	1	A	Phe 82	1	C
Ile 10	2	A	Leu 83	3	C
Ile 11	3	A	Lys 84	2	C
Thr 13	2	A	Arg 85	2	C
Leu 14	3	A	Leu 86	3	C
Asn 15	2	A	Leu 90	1	C
Leu 27	2		Gly 92	1	
Thr 28	2		Leu 93	2	
Val 29	3	S	Ser 107	2	S
Asp 31	2		Leu 109	2	
Ile 32	2		Asn 111	2	D
Phe 33	2		Phe 112	3	D
Glu 43	1	B	Leu 113	3	D
Thr 44	1	B	Glu 114	3	D
Phe 45	2	B	Arg 115	3	D
Cys 46	2	B	Leu 116	3	D
Arg 47	3	B	Lys 117	3	D
Ala 48	3	B	Thr 118	3	D
Ala 49	3	B	Ile 119	3	D
Val 51	2	B	Met 120	3	D
Leu 52	2	B	Arg 121	3	D
Phe 55	3	B	Tyr 124	3	D
Tyr 56	2	B			
Ser 57	2	B			

* Categories of rate of amide proton exchange:
1 NH is present in the spectrum recorded 10 min after protein is dissolved in D_2O but not in the spectrum collected 45 min later.
2 NH is present in spectra recorded both after 10 min and after a further 45 min from when the protein was dissolved in D_2O but not in a spectrum collected 12 hours later.
3 NH is present in all the spectra recorded after the protein was dissolved in D_2O.
Position. A, B, C and D correspond to the 4 helical regions 5-17, 41-57, 72-90 and 110-125. S corresponds to the short region of antiparallel β sheet.

Figure 5.3.
A schematic diagram of the short region of antiparallel β sheet in human IL-4. Arrows show the pairs of residues between which main chain to main chain NOE's have been identified. The two dotted lines represent hydrogen bonds in the β sheet; the amide protons involved (of Val 29 and Ser 107) are slow to exchange when the protein is dissolved in D_2O.

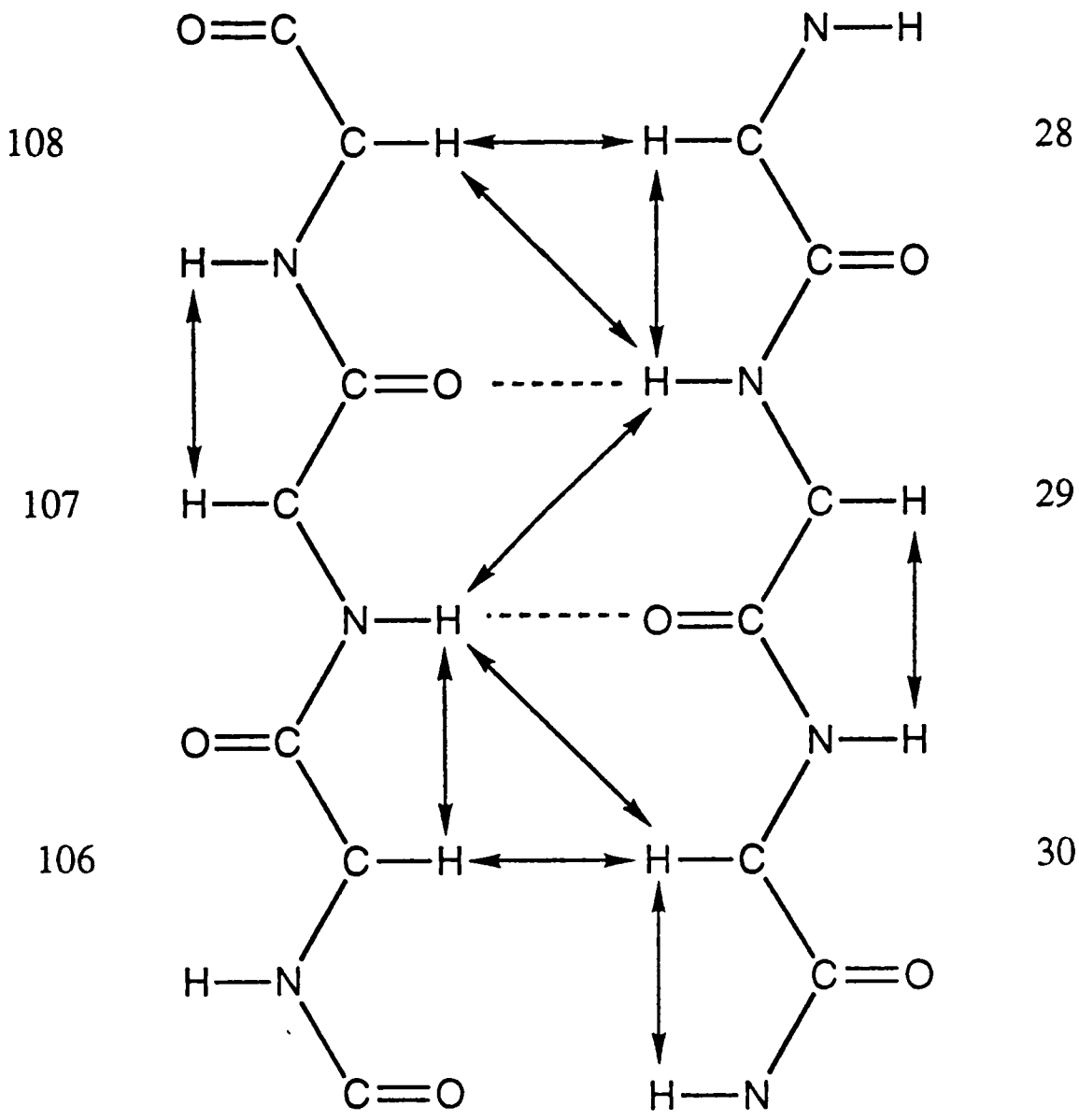


Table 5.2 Interstrand NOE's identified, in the spectra of unlabelled and ^{15}N labelled IL-4, for the short region of β sheet (residues 28-30 and 106-108).

NOE	Intensity[1]	NOE	Intensity
28Hγ-107HN	5	29Hγ-107Hβ	5
28Hβ-107HN	5	29HN-107Hβ	5
28Hα-108Hα	2	29Hγ-107HN	5
28Hα-108Hβ	4	29HN-108Hα	4
28Hα-108Hγ	3	29Hγ-108HN	5
28Hγ-108Hα	2	30Hα-106Hα	2
28Hγ-108Hβ	4	30Hγ-106Hα	5
28Hγ-108HN	5	30Hα-106Hε	4
29HN-107HN	4	30Hγ-106Hε	4
29Hγ-107Hα	5	30Hα-107HN	5

1 Categories (from NOE peak intensities): 1, very strong; 2, strong; 3, medium; 4, weak; 5, very weak. See section 5.2.3.

Four main helical regions and a short antiparallel β sheet are therefore clearly present in human IL-4. However, there are further short regions with assigned short and medium range NOE's. In addition, 8 slowly exchanging amides in the protein have been identified, which cannot be involved in hydrogen bonds in the recognised secondary structure. The positions of these residues and the structural implications of these data are discussed further in Chapter 6.

5.2.2 Long range NOE data.

A detailed analysis of the 2D and 3D NOESY spectra for ^{15}N labelled IL-4 enabled a total of 1360 NOE's to be identified, including 123 which are longer range (involving residues more than five removed in the sequence) and so define tertiary contacts in the folded protein. The distribution of the long range NOE information in the protein sequence is shown in Figure 5.4. Identification of additional NOE's could not be achieved reliably because of ambiguities arising from overlap in often poorly dispersed spectral regions. The NOE data are summarised in the diagonal plot in Figure 5.5 which reveals, in a remarkably direct manner, some notable features of the protein structure. Almost all the long range data points are concentrated around two lines; a strong 'anti-diagonal' and a line approximately parallel to this starting around Gly 95. The anti-diagonal shows that the polypeptide chain must loop back on itself bringing the N- and C-termini of the protein into close proximity and bringing helices A and D, and helices B and C close to each other. In each of these pairs of helical contacts, the chains must run in opposite directions in the two helices. The polypeptide chain also loops round again, as indicated by the other line of points parallel to the anti-diagonal, so that residues 1-30 run along side residues 95-55. This means that helix A is also adjacent to helix C and the AB loop is close to the end of helix B. It is interesting that the points for the three disulphide bridges in IL-4, also shown in Figure 5.5, lie within the two lines of NOE data points. These disulphide bridges link one end of helices A and D together (3-127), the CD loop to helix B (46-99) and the AB loop to the BC loop (24-65).

Figure 5.4.
Histogram showing the number of long range NOE's (between residues that are 5 or more apart in the sequence) per residue that have been identified in the spectra of unlabelled or ^{15}N labelled human IL-4. Each NOE is counted twice (once for each contributing nuclei).

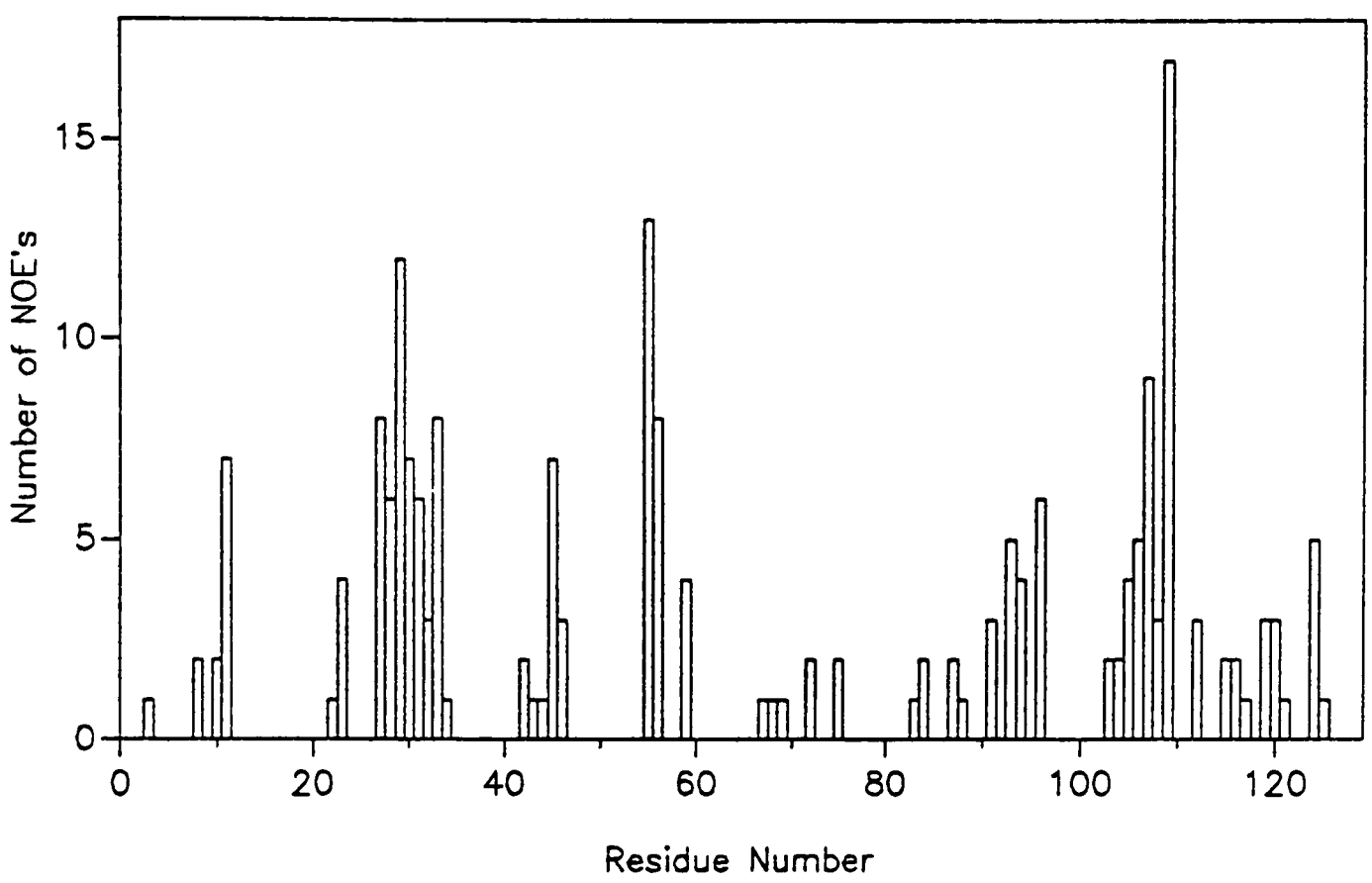
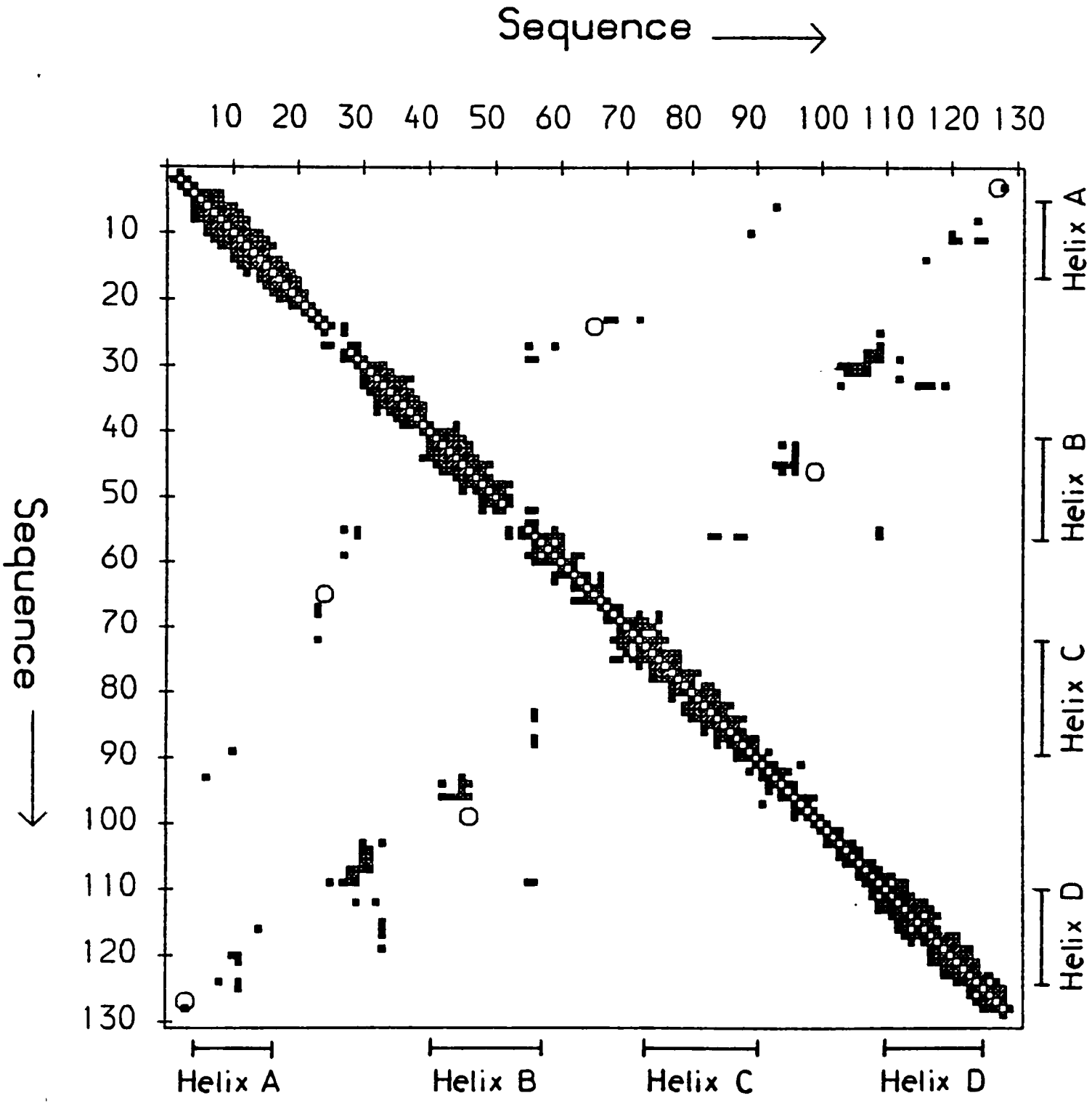


Figure 5.5.
Plot showing pairs of residues for which an NOE has been observed in the spectra of unlabelled or ^{15}N labelled IL-4. The location of the helices in the structure is indicated along the axes, and the disulphide connectivites are indicated by open circles.



5.2.3 Preliminary set of structural restraints.

The NMR data were used to give structural restraints for preliminary structure calculations for IL-4. The majority of NOE distance restraints were chosen on the basis of cross peak intensities in a NOESY-HMQC experiment recorded with a mixing time of 100 ms. At this mixing time the cross peaks were sufficiently strong to be identified readily, but spin diffusion should not be extensive. From a study of the intensity of cross peaks corresponding to known distances in the regular secondary structure of the protein, the following interproton distance restraints were chosen: for cross peaks categorised as very strong, strong, medium and weak restraints of 1.8-2.5, 1.8-3.0, 1.8-4.0, 1.8-5.0 Å respectively were used. A further category of very weak was used for cross peaks which were only observed in a spectrum obtained with a 200 ms mixing time. For these cross peaks a wider distance range of 1.8-6.0 Å was used for the restraint to take into account the effects of spin diffusion. Pseudoatoms were used in the calculations as no stereospecific assignments could be attempted because of extreme overlap in the αβ region of the 2D spectra which make $^3J_{\alpha\beta}$ measurement impossible. The necessary corrections were added to the distance ranges to take into account the pseudoatoms used (Wüthrich et al., 1983). In addition, 0.5 Å was added to the upper limits of the restraints of NOE's involving methyl groups to allow for the greater intensity of NOE's originating from a group of three methyl protons.

For helical residues where the amide protons are slow to exchange, hydrogen bond distance restraints were also included (O(i-4) to NH(i) 1.8-2.0 Å; O(i-4) to N(i) 2.7-3.0 Å. Kline et al., 1988) (for 27 residues). No hydrogen bond restraints were used for the region of β sheet (28-30 and 106-108) as this is very short and the coupling constant values showed it is irregular (for three of the residues the measured $^3J_{HN\alpha} \leq 7Hz$).

The $^3J_{HN\alpha}$ values for IL-4 (Appendix V) were converted into dihedral angle restraints. For helical residues where $^3J_{HN\alpha} < 6$ Hz (64 residues, all those in the four helices with the exception of unassigned Arg 53) ϕ was restrained to $-60 \pm 30°$. For residues with $^3J_{HN\alpha} > 8$ Hz indicative of an extended conformation, ϕ was restrained to $-120 \pm 40°$. 9 of the residues in IL-4 are in this category; all of these are in the loop regions and short β sheet. For another 28 residues $^3J_{HN\alpha}$ is less than 7 Hz, but the residue does not lie in one of the four main helical regions. For these residues a small coupling constant does not necessarily imply that $\phi \approx -60°$ and so a wider angle range of $60 \pm 150°$ was used as a restraint. This range only excludes the extended conformation around $\phi = -120°$.

The structural restraints used in the preliminary calculations are summarised in Table 5.3 and the complete structural data set is listed in Appendices VI and VII.

Table 5.3 Summary of the NMR structural restraints used in the preliminary structure calculations for human IL-4.

Conventional calculations.	
Type of restraint	Number of restraints
NOE distance	1360
ϕ angle	99
Pairs of hydrogen bond	27
Additional restraints in calculations with 'fixed' helices.	
$\phi=-60 \pm 5°$, $\psi=-50 \pm 5°$	65*

* For the helical residues 5-17, 41-57, 72-90, 110-125; these replaced the usual ϕ angle restraints for these residues.

5.2.4 Initial structure calculations for human IL-4.

The structure calculations were performed using the distance geometry-dynamical simulated annealing protocol (Nilges et al., 1988) described in detail in Chapter 2. When calculations were run using only the restraints that are conventionally derived from the experimental NMR data (1360 NOE distance restraints, 99 ϕ angle restraints and 27 pairs of hydrogen bond distance restraints) structures were routinely generated with irregularities in one or more of the helices which gave rise to localised and inconsistent violations of the experimental restraints (see Figure 5.6). For example, the top left-hand structure in Figure 5.6 has distortions at the end of the B helix (residues 53-56) and in the C helix (residues 73-78 and 85-90). For this structure after the simulated annealing part of the calculation protocol there were 15 NOE violations > 0.5 Å one of which, between Tyr 56 CβH and Asp 87 NH, was particularly large (distance 11.7 Å but upper range of distance restraint is 7Å). During the full potential dynamics the number and size of the NOE violations decreased (total violations 13, none > 1.1 Å); the violations remaining are however concentrated in the B and C helices, many of them being violations of medium range helical restraints e.g. violations > 0.5 Å for 72CαH-75NH, 77CαH-81NH and 85CαH-87NH. There are also four hydrogen bond restraint violations that are greater than 0.5 Å e.g. 57NH-53O (distance 3.1 Å but upper range for restraint is 2.0 Å) reflecting the distorted B and C helices in the structure. In other structures there are similar distortions to one or more of the helices resulting in the same type of NOE and hydrogen bond restraint violations.

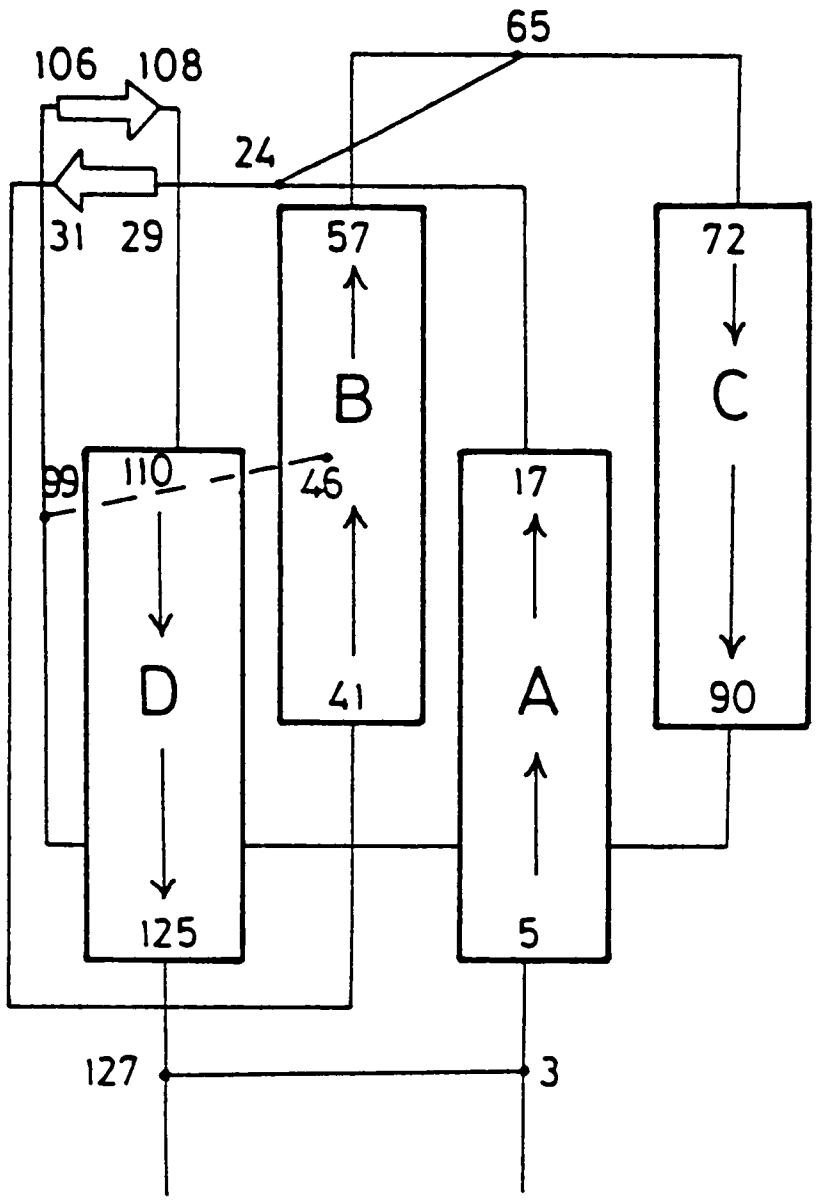
Calculations were therefore also run where dihedral angle ϕ and ψ restraints were used, in conjunction with NOE restraints, to restrain the four regions 5-17, 41-57, 72-90 and 110-125, into more regular helical secondary structures (restraints of $\phi = -60 \pm 5°$, $\psi = -50 \pm 5°$; see Table 5.3). Structures calculated in this way had no significant or systematic violations of the

Figure 5.6.
Examples of preliminary structures of IL-4 calculated using an NMR data set collected from unlabelled and ^{15}N labelled samples of IL-4. Conventional ϕ torsion angle restraints were used. Note the irregularities in the helices and the differences in loop conformations.

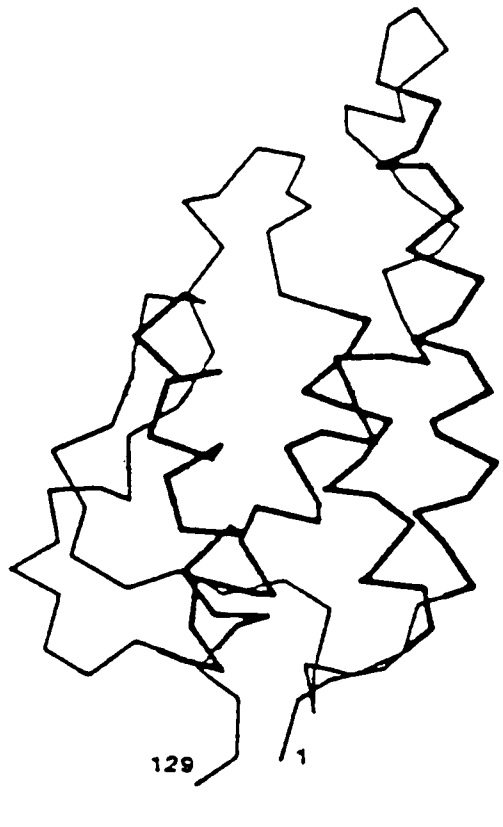
experimental data; typically, no more than 10 of the 1360 NOE's were violated by more than 0.5 Å and all violations were less than 1.3 Å. The calculations therefore showed that a conformation for the protein with four straight helices could be adopted without increasing the violations of the experimental data. Indeed, it is interesting that restraining the conformation of the helices resulted in fewer violations of the NMR restraints not only in the helices but also in other parts of the structure.

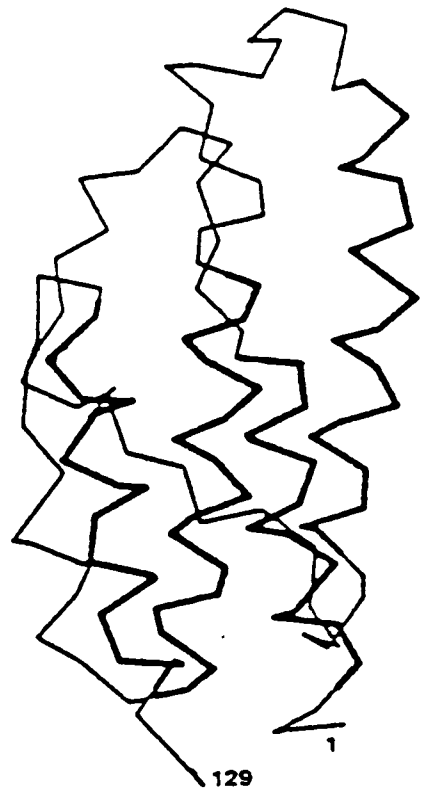
From a set of 20 structures calculated with the additional ϕ and ψ restraints, the 10 lowest energy structures had an average Cα RMSD from the average structure of 3.2 ± 0.3 Å (for just the helical regions the value reduces to 2.7 ± 0.3 Å). Although there are variations between these calculated structures, all structures of low energy defined by both the methods adopt a four helix bundle type structure. A schematic representation of the topology found in all the calculated structures is given in Figure 5.7. This shows the connectivity pattern for the bundle of four helices (see section 5.4.1), and indicates the position of the disulphide bonds and β sheet relative to them. The set of NOE's available from unlabelled and ^{15}N labelled samples of IL-4 does not, however, define well the packing of these helices; this is indicated in Figure 5.8 which shows a representative selection of the preliminary structures calculated (with ϕ and ψ restraints) for IL-4. There are large variations in the orientations of the helices and hence the angles between them (Table 5.4). For example the BD interhelix angle varies from 34.3° to 70.1° in the set of 10 preliminary NMR structures. In addition, the topology where the long AB and CD loops meet could not be reliably defined by this data set. For the majority of the structures the AB loop passes through the CD loop (shown subsequently to be the correct topology) but in a few structures, such as the one in the top left of Figure 5.6, the AB loop passes round the CD loop.

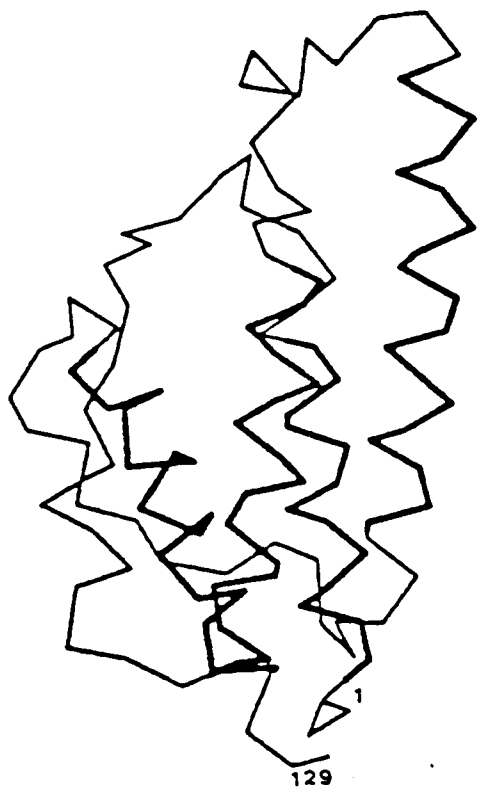
Figure 5.7.
Schematic diagram of the main chain fold of the structures calculated for human IL-4, showing the up-up-down-down connectivity and the positions of the disulphide bridges and the short region of β sheet.

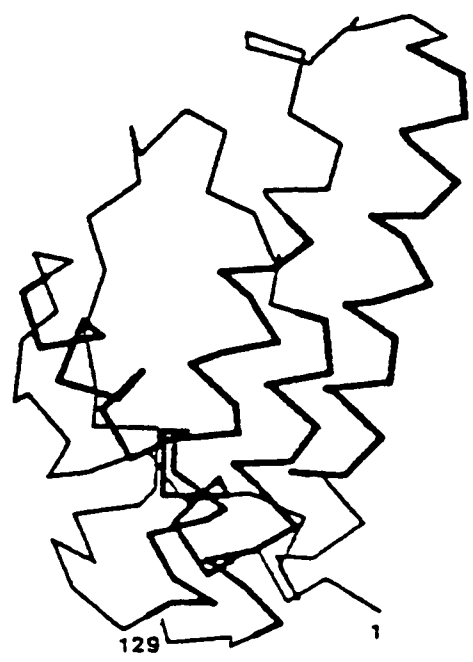


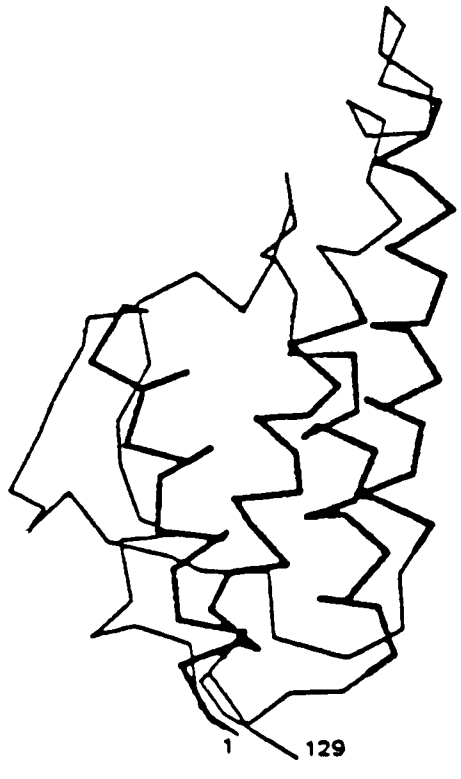
Overleaf:
Figure 5.8.
Examples of preliminary structures of IL-4 calculated using NOE data from spectra of unlabelled and ^{15}N labelled samples of the protein and ϕ and ψ restraints for helical residues (ϕ= -60±5°, ψ = -50±5°). Although the overall topology is relatively well defined, there are significant variations in the angles between the helices and in the loop conformations.


129
1


129
1


1
129


129
1


1
129

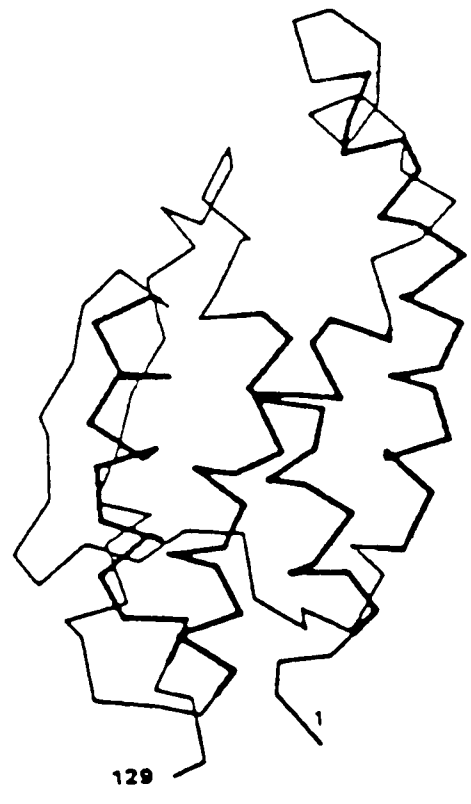

1
129

Table 5.4. Interhelix angles in the NMR structures calculated with the preliminary NMR data set (using angle restraints of ϕ=-60±5°, ψ=-50±5°)*.

AD	52.8 ± 10.3°
DB	52.7 ± 11.0°
BC	23.9 ± 11.4°
CA	16.6 ± 11.2°

* Absolute values of the acute interhelix angles as defined by Presnell & Cohen (1989). Means and standard deviations of the angles in the set of 10 calculated structures are given.

It is perhaps surprising at first sight that 1360 NOE's are not sufficient to define the structure more fully, especially considering the results for hen lysozyme given in Chapter 4. However, for a helical protein, in contrast to one involving significant regions of β sheet structure the contacts between regions of secondary structure distant in the primary sequence almost invariably involve side chain atoms. In order to define these contacts it is therefore necessary to assign the majority of the side chain resonances in the NMR spectrum, and to have sufficient resolution to be able to detect NOE's between them. Although the use of an ^{15}N labelled sample of the protein leads to very effective assignment of main chain resonances and identification of secondary structure, it does not help with the assignment of side chain to side chain NOE's. For this preliminary data set these had to be identified from the less well resolved 2D experiments. As shown in Figure 5.4, this gave a small number, only 123, of long range distance restraints which were also poorly distributed throughout the protein sequence. For example, there are no long range distance restraints both for residues 35-41 in the AB loop and for residues 76-82 in the C helix. In order to overcome this limitation a sample of recombinant IL-4 uniformly labelled in both ^{15}N and

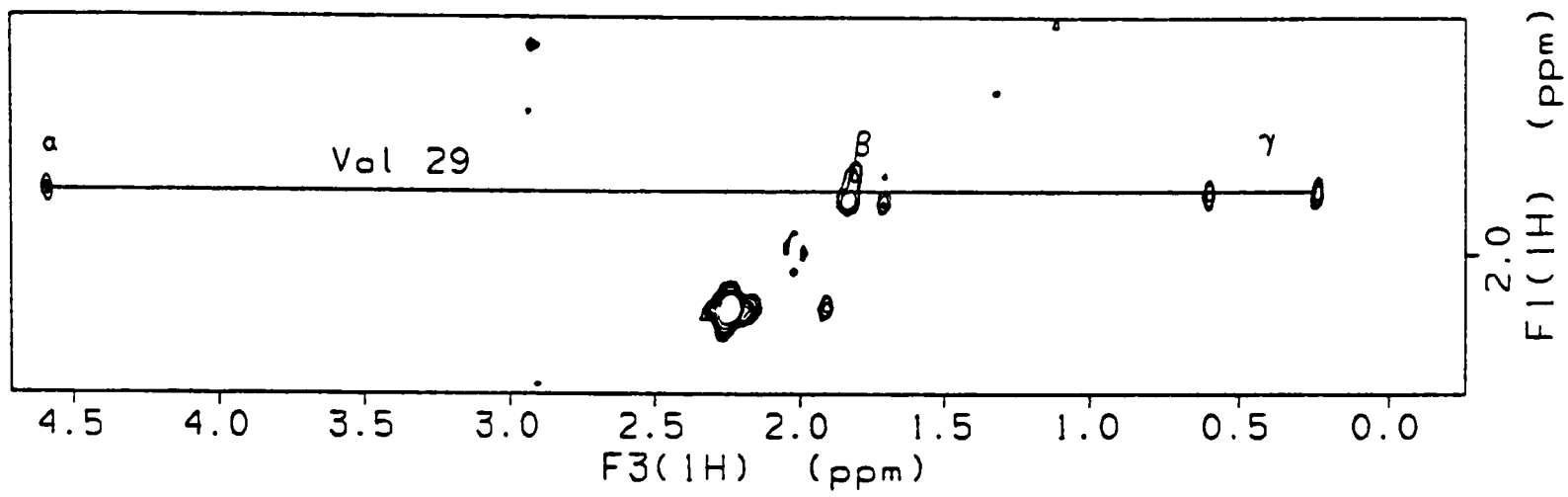
^{13}C was prepared. As discussed in the following section the additional labelling of the ^{13}C atoms proved crucial in analysing the regions of the spectrum containing resonances of long side chains whose protons are distant from the ^{15}N atoms.

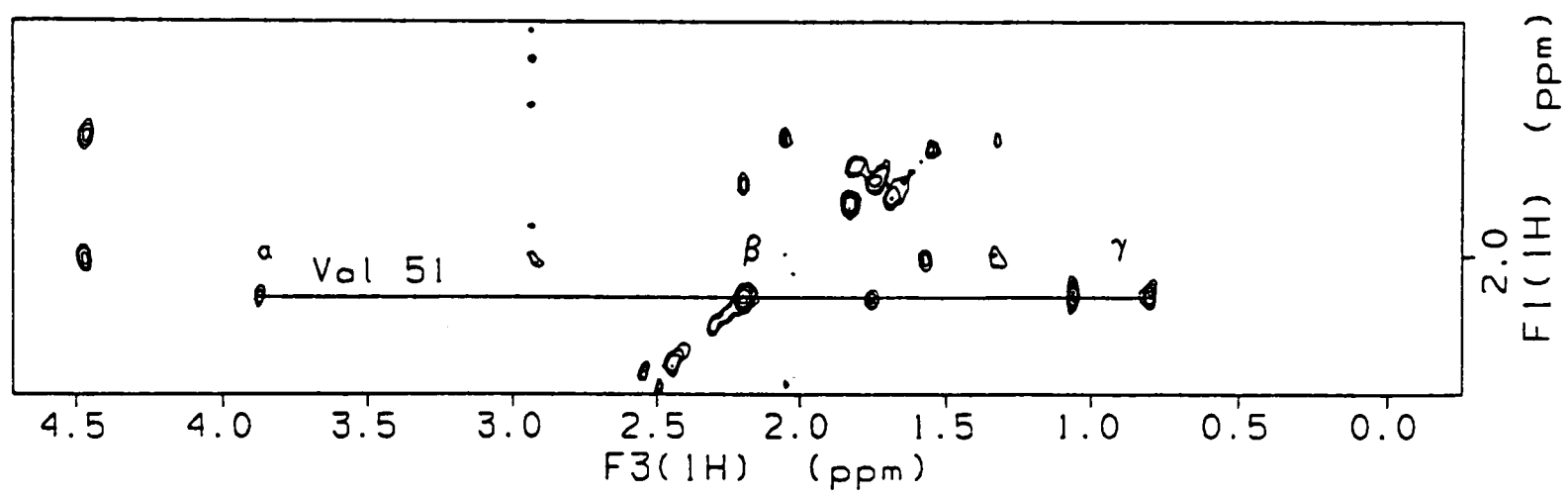
5.3 Double labelled (^{13}C, ^{15}N) sample.

Three 3D NMR experiments were recorded, in collaboration with Dr. J. Boyd, using a 2mM sample of the double labelled protein: ^{13}C-^{1}H HCCH-COSY, ^{13}C-^{1}H HCCH-TOCSY and ^{13}C-^{1}H NOESY-HMQC (Bax et al., 1990b; Ikura et al., 1990b; Kay et al., 1990). A brief description of these experiments is given in Appendix 1. (The analysis of these experiments was carried out in conjunction with Dr. C. Redfield who, in particular, assigned NOE's arising from Cα and aromatic protons.) The HCCH-COSY and HCCH-TOCSY experiments were used to make $^{13}C\alpha$ and side chain ^{13}C and ^{1}H assignments. The assignment process involved identifying patterns of cross peaks in F1(^{1}H)-F3(^{1}H) planes of the 3D spectra which are characteristic of the different spin systems found in the amino acid residues, and then transferring the ^{1}H assignments to these spectra, that had been made previously from spectra of the ^{15}N labelled sample, so as to identify the particular residue from which the cross peaks arise. In addition, use was made of the fact that the ^{13}C resonances have characteristic ranges of chemical shifts for different amino acid types (Ikura et al., 1991). This means that resonances of a given carbon atom in a certain amino acid type will cluster in a small number of F1-F3 planes and so can be more readily recognised. For example, F1(^{1}H)-F3(^{1}H) planes in the HCCH-COSY spectrum taken at the chemical shift of the valine $^{13}C\beta$ can be recognised by the pattern of one cross peak to Hα and two to the methyl group protons (Hγ1 and Hγ2) as shown in Figure 5.9. The $^{13}C\beta$ chemical shifts of all the 3 valines in IL-4 were found to lie in the range 29.8-34.3 ppm. Similarly, the F1-F3 planes in the HCCH-TOCSY spectrum at the

Figure 5.9.

F1(^{1}H)-F3(^{1}H) planes in the HCCH-COSY spectrum of double labelled human IL-4 taken at the chemical shifts of ^{13}Cβ of valines 29, 51 and 101 (chemical shifts of 34.3, 29.8 and 32.2 ppm respectively). The cross peaks to Hα and the two sets of methyl group protons (Hγ$_1$ and Hγ$_2$) are labelled.





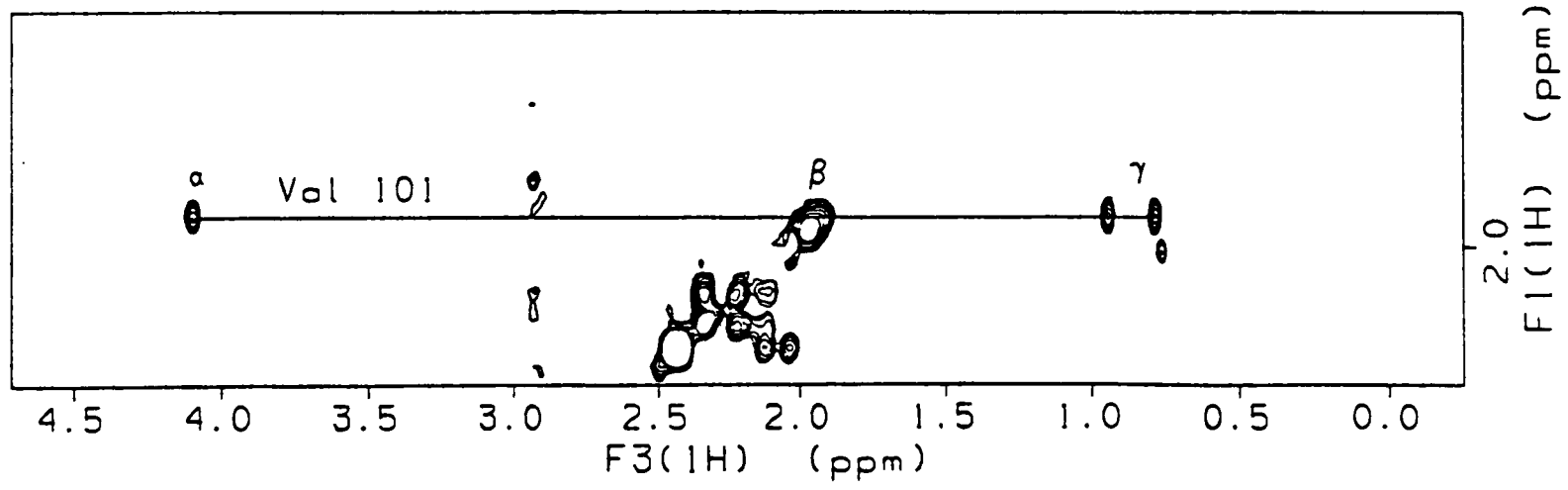


Figure 5.10.

F1(^{1}H)-F3(^{1}H) planes in the HCCH-COSY spectrum of double labelled human IL-4. A and B are taken at the chemical shifts of ^{13}Cα and ^{13}Cβ of Ser 107 (54.4 and 64.5 ppm respectively). C is taken at the ^{13}Cβ chemical shift of four threonine residues in IL-4 (66 ppm) and shows cross peaks to α and γ methyl group protons.

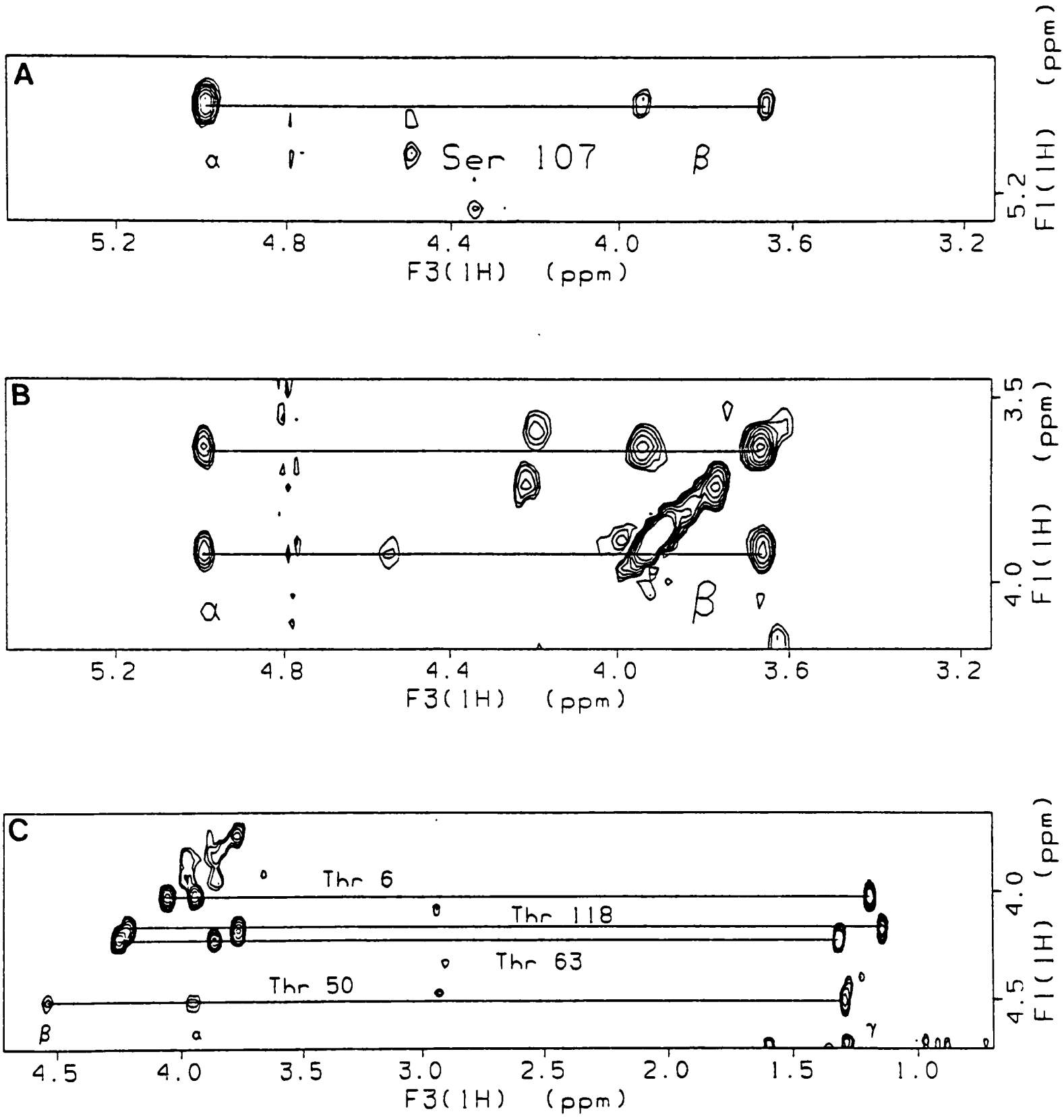
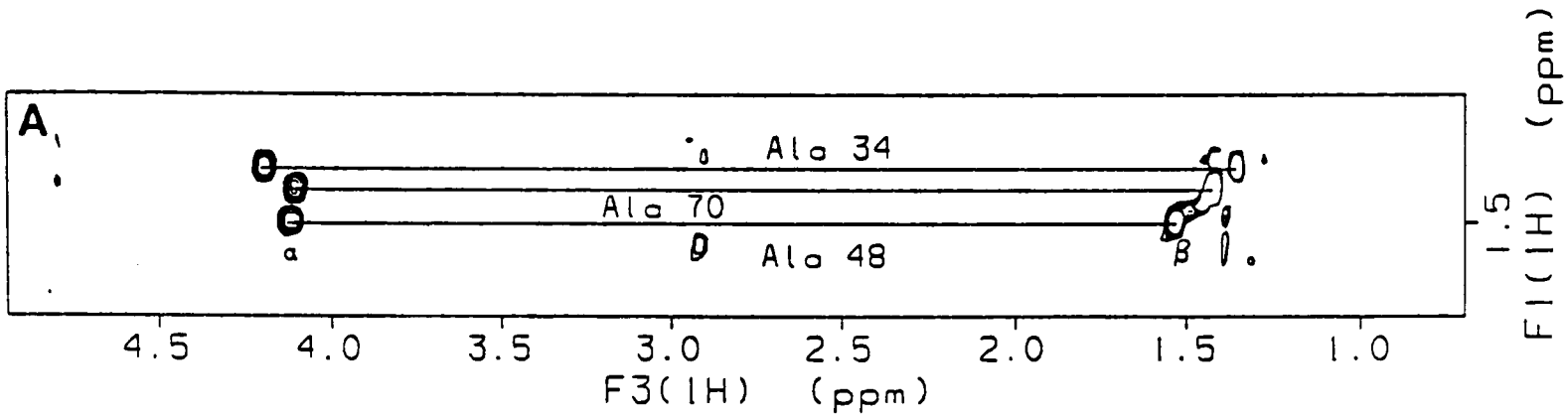
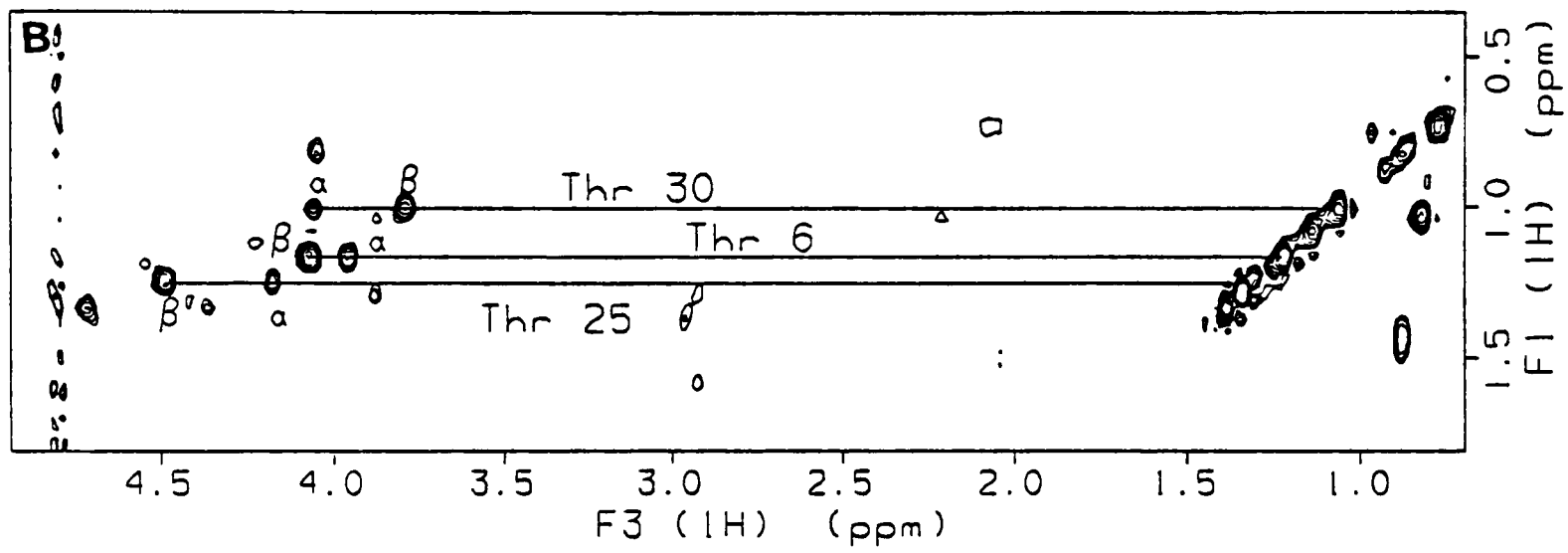


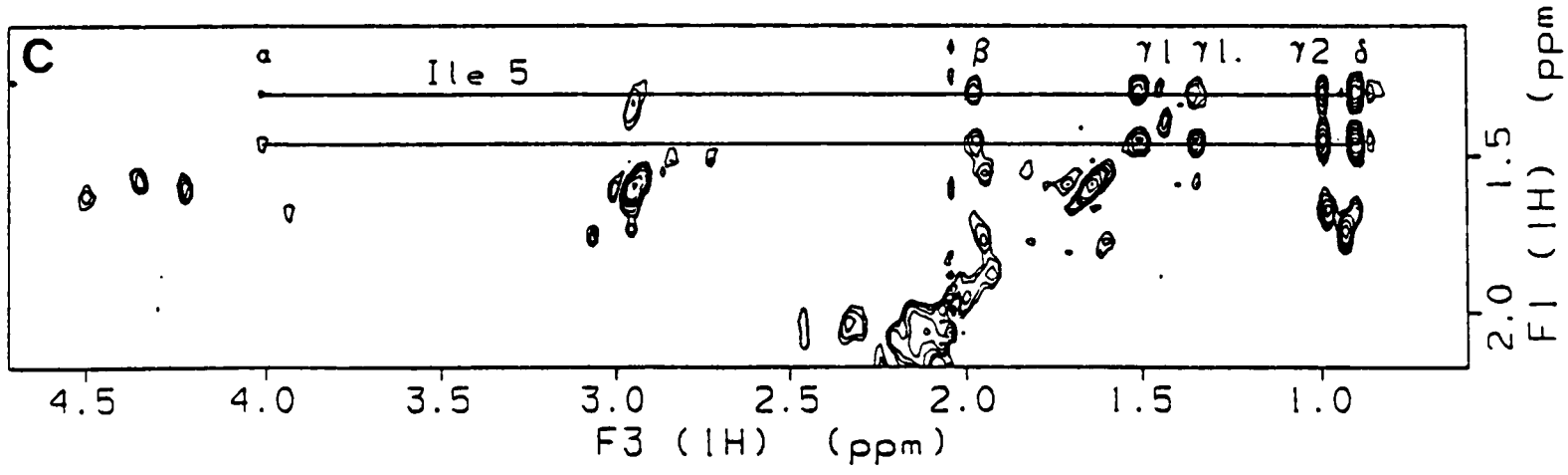
Figure 5.11.

A shows a F1(^{1}H)-F3(^{1}H) plane in the HCCH-COSY spectrum at the ^{13}Cβ chemical shift of three alanine residues (16 ppm).

B and C show F1(^{1}H)-F3(^{1}H) planes in the HCCH-TOCSY spectrum of double labelled human IL-4 taken at the chemical shifts of Thr ^{13}Cγ (for residues 6, 25, 30 (at 20.5 ppm)) and Ile ^{13}Cγ_1 (for residue 5 (at 26.0 ppm)) respectively. In C cross peaks to α, β, γ_2 and δ protons of Ile 5 are observed.







F2 chemical shifts of Ile $^{13}C\gamma1$ (Figure 5.11C) have two lines of cross peaks, from each of the pair of $\gamma1$ protons, to α, β, $\gamma2$ and δ protons, and so can be readily identified. Using methods of this type to identify the residues (illustrated in Figures 5.9, 5.10 and 5.11) and transferring the ^{1}H assignments made previously to these spectra, 92% of the $^{13}C\alpha$, 79% of the side chain ^{1}H and 59% of the side chain ^{13}C resonances were able to be assigned. These assignments are listed in Appendix VIII.

Using the ^{1}H and ^{13}C assignments, cross peaks could then be identified in the ^{13}C-^{1}H NOESY-HMQC spectrum. This NOESY-HMQC spectrum is asymmetric about the diagonal. For an NOE between protons A and B the (F1, F3) = (^{1}HA, ^{1}HB) cross peak occurs on a slice taken at the ^{13}C chemical shift of CB, while the (F1, F3) = (^{1}HB, ^{1}HA) cross peak is found on a slice at the chemical shift of the ^{13}CA. This means that the NOE assignment of a given AB cross peak in the ^{13}CB slice could be confirmed by looking for the corresponding BA cross peak in the ^{13}CA slice. In addition, this method was used to resolve many of the ambiguities in NOE assignments arising from ^{1}H chemical shift degeneracies. For example, slices through the 3D spectrum at the ^{13}C chemical shift of the two Leu 109 Cδ groups (20.5 and 23.6 ppm) contain cross peaks between the CδH_3 protons of 109 (^{1}H chemical shift 0.10 and 0.30 ppm) and a proton with a chemical shift of 4.17 ppm (Figure 5.12a). The Cα protons of three residues, Leu 23, Ser 16 and Thr 25, all have chemical shifts of 4.17 ppm. A corresponding pair of cross peaks is seen, however, in a slice taken at the ^{13}C chemical shift of Thr 25 Cα (62.0 ppm) at the proton chemical shifts (F1, F3) = (0.10 ppm, 4.17 ppm) and (0.30 ppm, 4.17 ppm) (Figure 5.12b) but no peaks are seen at these positions in the slices at $^{13}C\alpha$ shifts of Ser 16 and Leu 23. These cross peaks could therefore be unambiguously assigned as 109 CδH-25CαH.

The initial work on NOE identification concentrated on NOE's originating from methyl group protons as these were expected to reveal

Figure 5.12a.
F1(^{1}H)-F3(^{1}H) planes through the NOESY-HMQC spectrum of double labelled human IL-4 at the chemical shifts of the two ^{13}Cδ nuclei of Leu 109 (20.5 and 23.6 ppm). Cross peaks from the δ protons to a proton with a chemical shift of 4.17 ppm (later identified as Thr 25 Hα) are indicated.

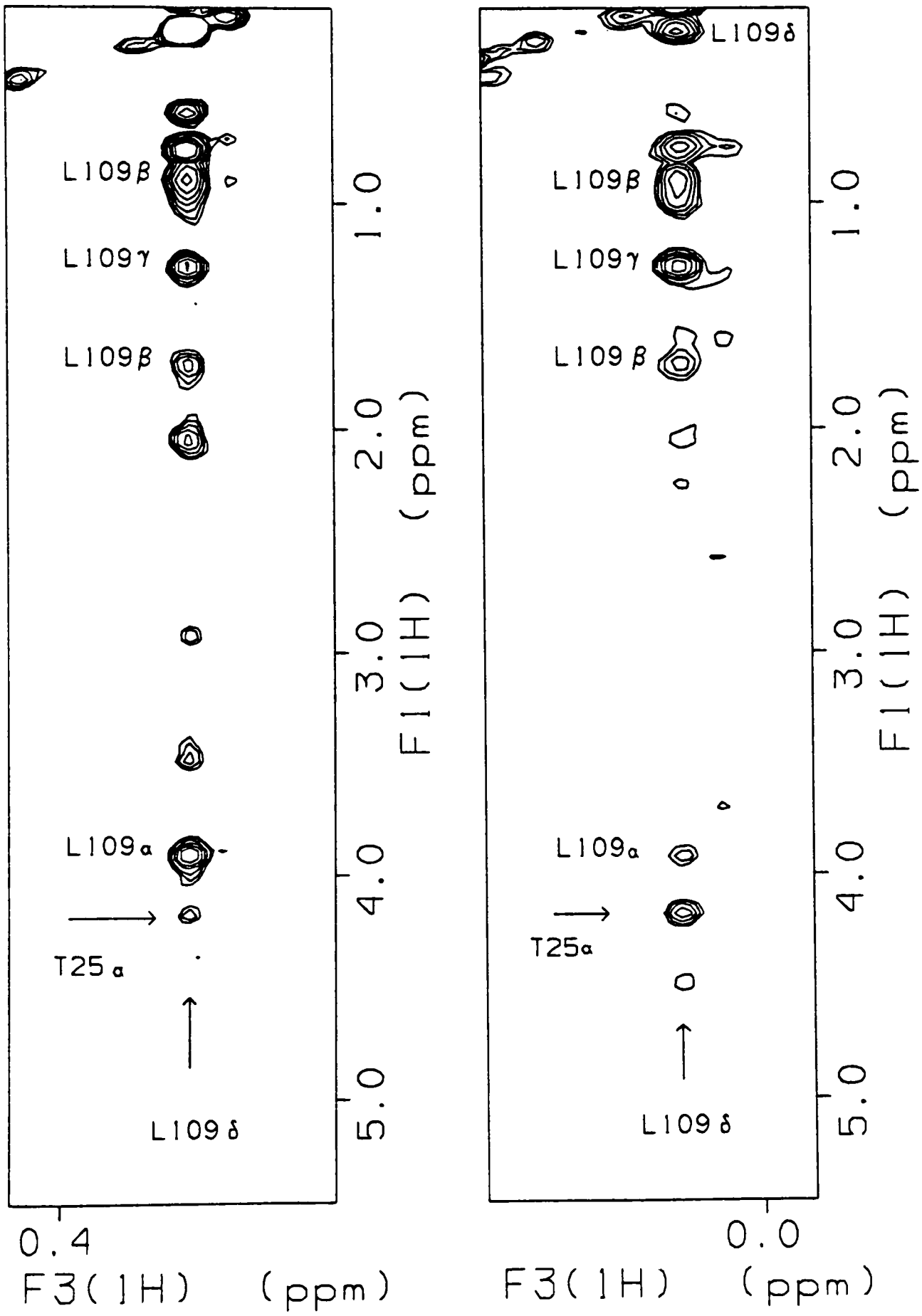


Figure 5.12b.
F1(^{1}H)-F3(^{1}H) plane taken at the chemical shift of Thr 25 ^{13}Cα (62.0 ppm) showing cross peaks to the δ protons of Leu 109.

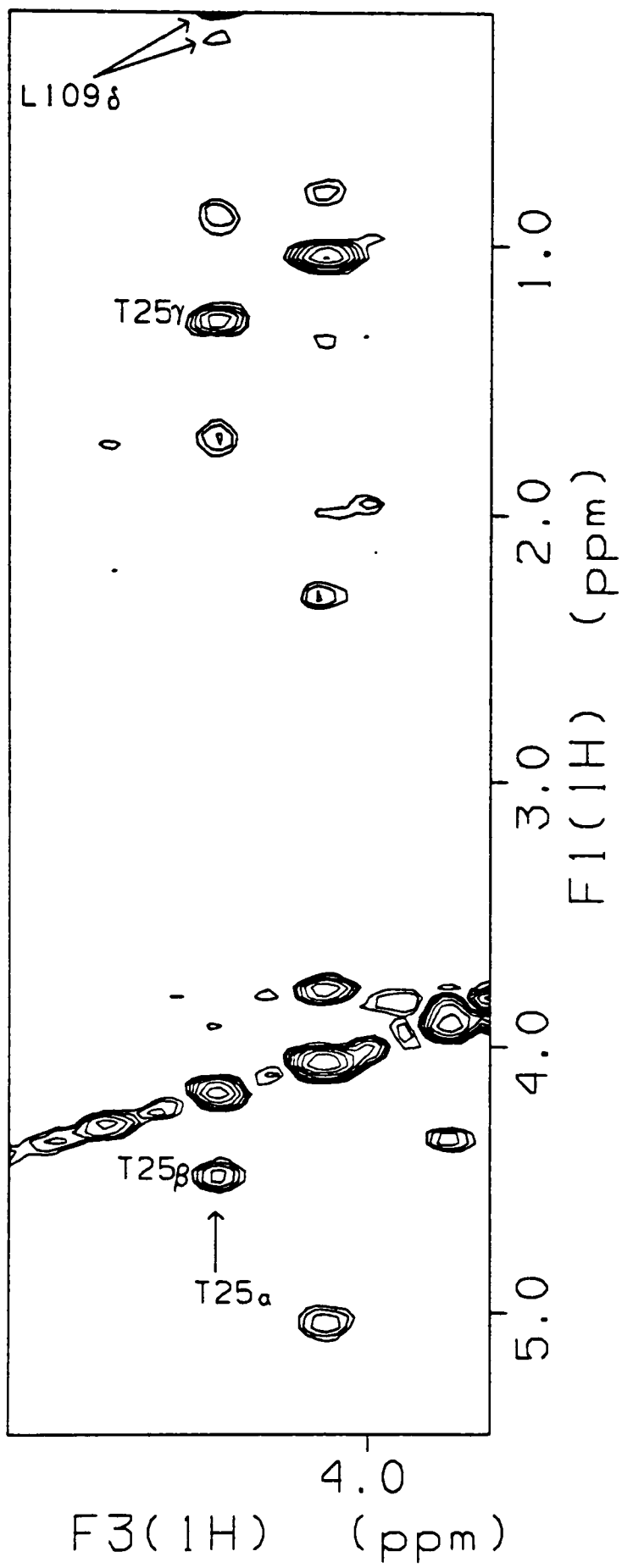


Figure 5.13.
F1(^{1}H)-F3(^{1}H) planes through the NOESY-HMQC spectrum of double labelled human IL-4 at the chemical shifts of A) Ile 80 ^{13}Cδ and B) Val 29 ^{13}Cγ (6.16 and 16.98 ppm respectively). The identified NOE's are labelled.

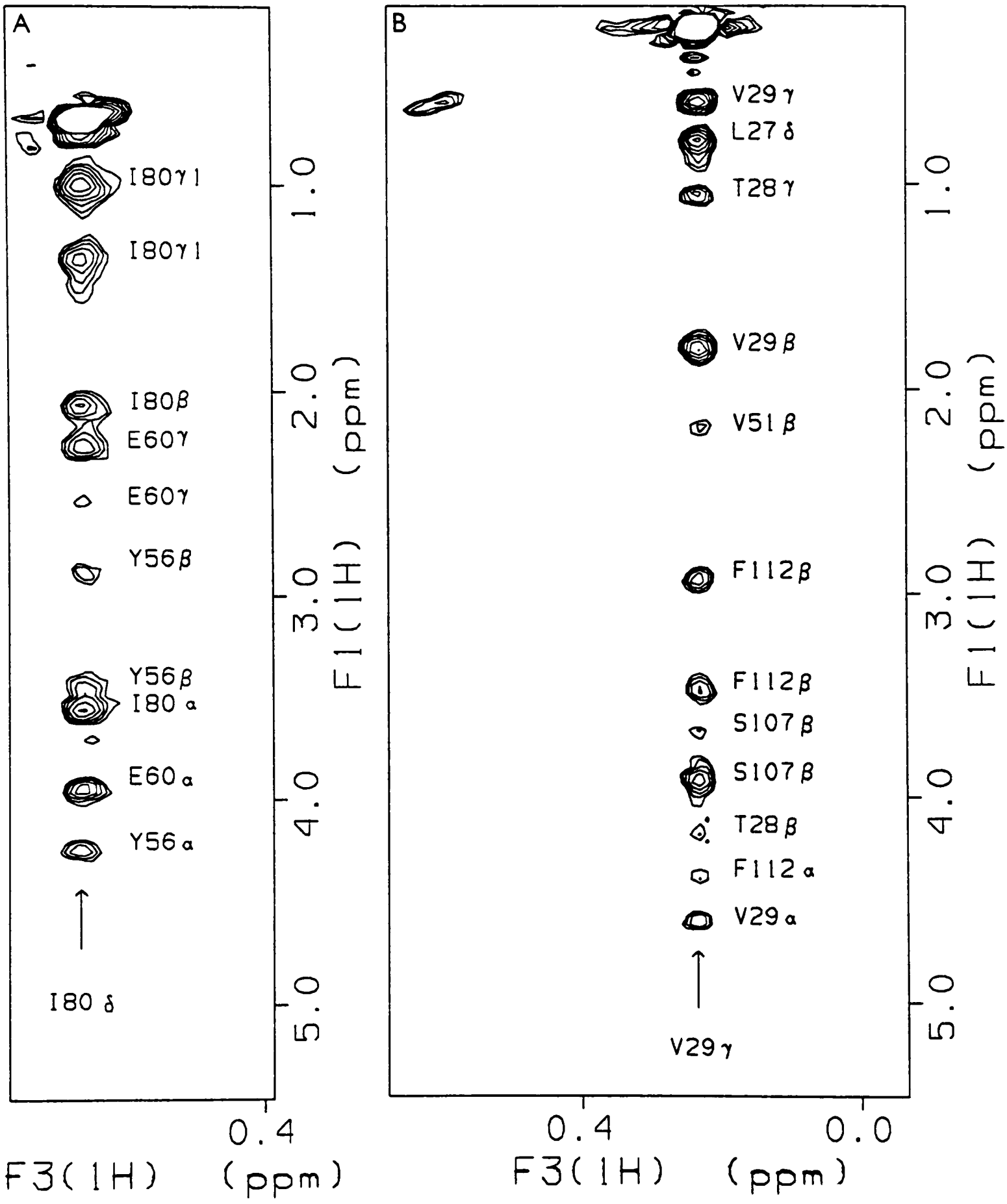
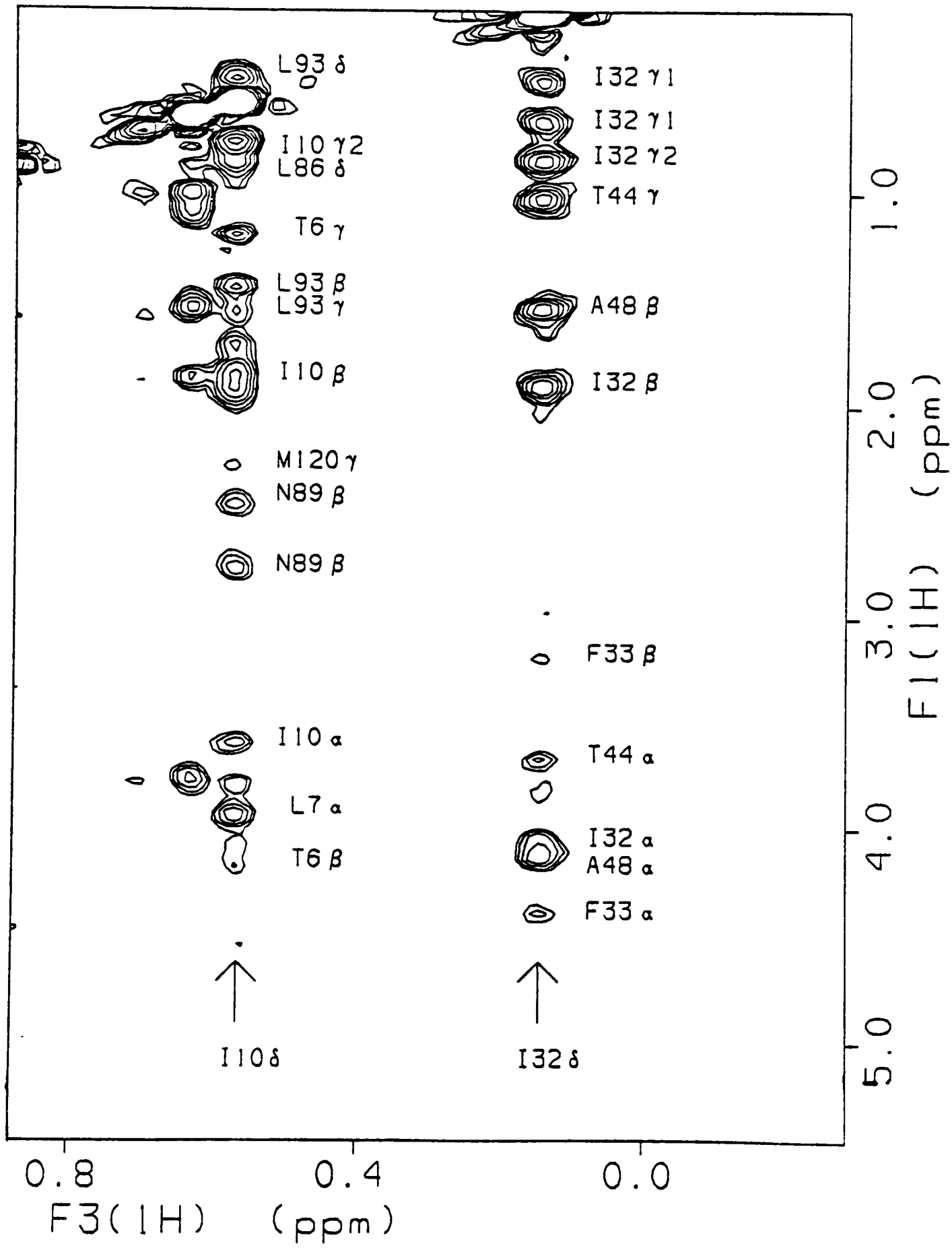


Figure 5.14.
F1(^{1}H)-F3(^{1}H) plane through the NOESY-HMQC spectrum of double labelled human IL-4 at the chemical shift of ^{13}Cδ of Ile 10 and Ile 32 (13.0 ppm) showing the identified NOE's from the δ protons of these residues.



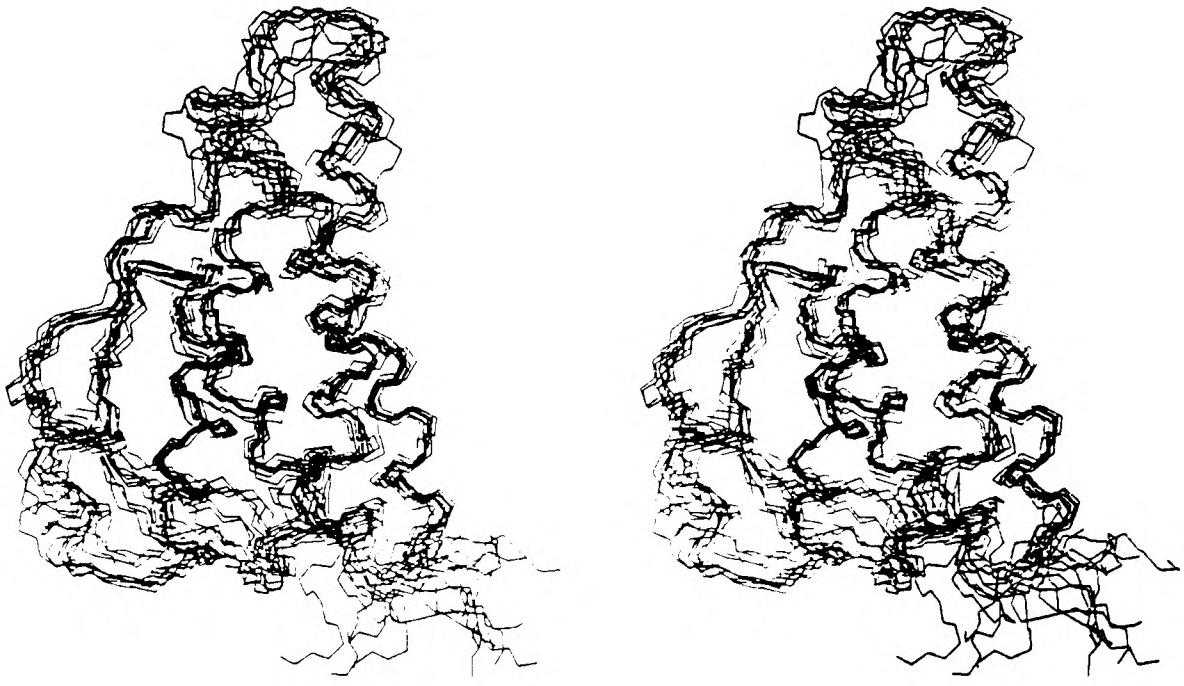
many interhelix contacts and hence give important distance restraints for further structure calculations. Examples of part of the NOESY-HMQC spectrum are given in Figure 5.13 and 5.14 which shows the NOE's arising from the methyl group protons of Ile 80 δ, Val 29 γ, Ile 10 δ and Ile 32 δ. The identified NOE's were classified as strong or weak from their intensities (the corresponding distance restraints are 1.8 - 3.0 Å and 1.8 - 5.0 Å) (this classification of the NOE's was carried out by Dr. C. Redfield).

5.4 Calculation of high resolution structures for IL-4.

The NOE's identified in the spectra of the double labelled sample were combined with those from the spectra of ^{15}N labelled and unlabelled samples of the protein to give a total data set of 1735 NOE distance restraints. These data are summarised in Table 5.5 and a full listing of the NOE's is given in Appendix VI. Using this set of NOE distance restraints, together with the ϕ angle and hydrogen bond restraints obtained from experiments on the ^{15}N labelled sample, an ensemble of 15 structures was calculated using the distance geometry-dynamical simulated annealing protocol (Nilges et al., 1988). (This is the same protocol that was used for lysozyme; see Chapters 2 and 4.) Note that the ϕ and ψ restraints used in the preliminary structure calculations to tightly restrain the conformation of the helical residues were not used in these calculations. A set of ten structures with the lowest overall energy at the end of the calculations was selected; these are shown in Figure 5.15. The statistics from the calculations of these structures are given in Table 5.5. An average of eight NOE's violate the structures by more than 0.5 Å (no violations greater than 1.1 Å). Four of the violations greater than 0.5 Å occur in six or more of the structures; these are also listed in Table 5.5. They may correspond to incorrectly assigned NOE's or NOE's that have had their apparent intensities increased by overlap with other cross peaks.

Figure 5.15.

Main chain superposition of the ensemble of 10 structures of human IL-4 calculated using the full NOE data set. In the colour print the four α helices are shown in cyan (A), green (B), red (C) and magenta (D).



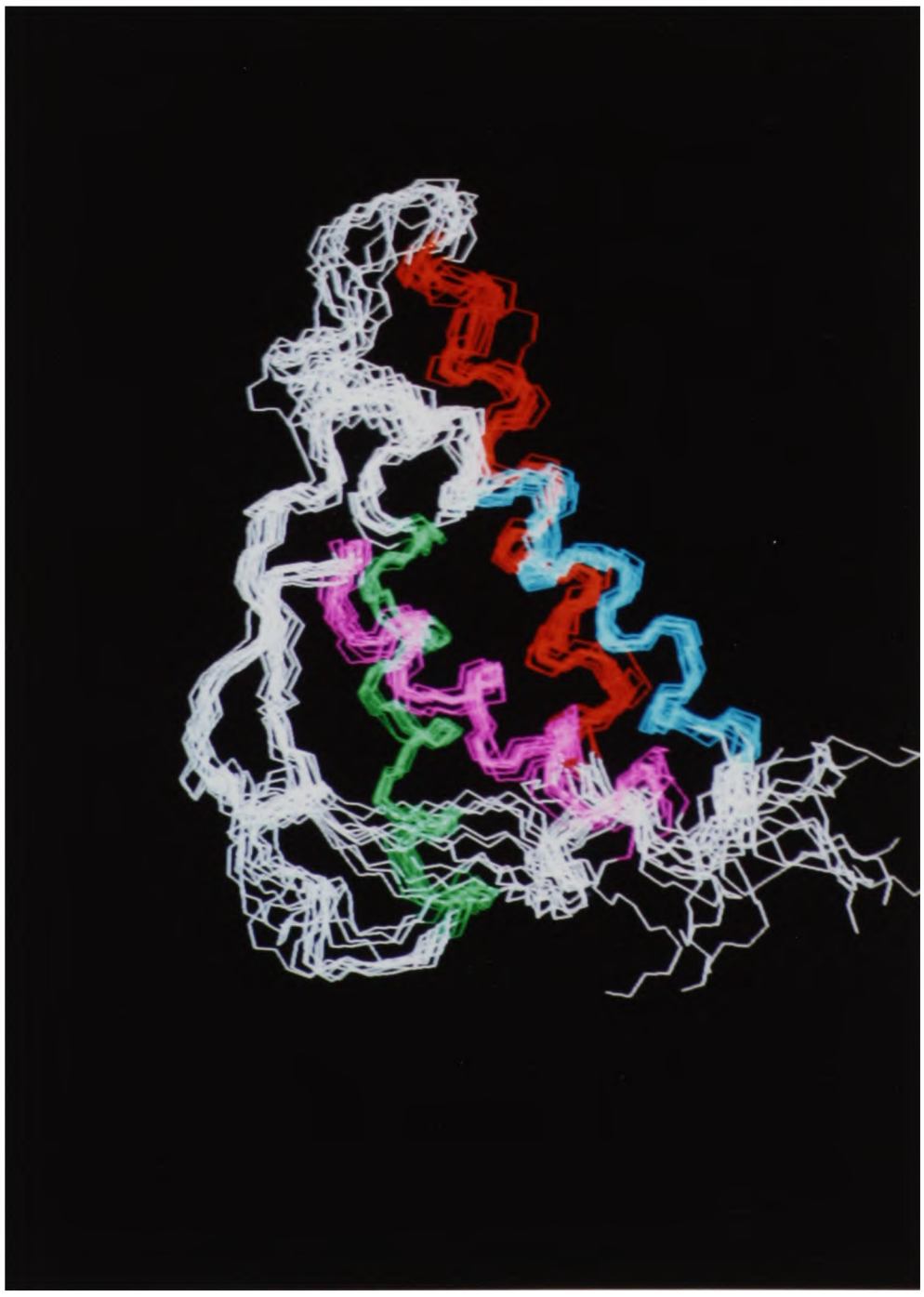


Table 5.5

Data set used in high resolution structure calculations.

Total NOE restraints	1735
Intraresidue [(i,j) i=j]	386
Interresidue [(i,j) (i-j) = 1]	557
Interresidue [(i,j) $2 \leq (i-j) \leq 4$]	474
Interresidue [(i,j) $(i-j) \geq 5$]	318
Backbone-backbone [(i,j) $(i-j) \geq 5$]	18
Backbone-sidechain [(i,j) $(i-j) \geq 5$]	127
Sidechain-sidechain[(i,j) $(i-j) \geq 5$]	173
Pairs of hydrogen bond restraints	27
ϕ torsion angle restraints	101

Statistics from structure calculations (average over ten calculated structures).

RMS Deviations from idealised covalent geometry.

Bonds (Å)	0.010 ± 0.0004
Angles (deg)	2.719 ± 0.096
Impropers (deg)	0.260 ± 0.014

NOE restraints.

RMSD (Å)	0.081 ± 0.005
Number violations > 0.5 Å	8 ± 3

NOE violations > 0.5 Å in at least 6 of the 10 final structures.

NOE	Number of structures
Leu 7 Hγ-Gln 8 HN	6
Thr 13 Hγ2#-Arg 85 Hα	6
Leu 27 Hα-Val 29 HN	8
Ala 48 Hβ#-Leu 96 Hδ*	10

and * represent pairs of methylene protons and methyl groups respectively which have not been stereospecifically assigned.

The set of 10 structures were superimposed and an average coordinate set was calculated which was energy minimized in the presence of the NMR data. The Cα RMSD's of the ensemble of structures from the average energy minimized structure are given in Table 5.6 where they are compared with the results for the preliminary structure calculations. The inclusion of the additional NOE's identified in the spectra of the double labelled sample gives a dramatic improvement in the quality of the structures (Cα RMSD for all residues from the average of 1.6 ± 0.4 Å for the high resolution structures compared to an RMSD of 3.2 ± 0.3 Å for the preliminary structures) and the interhelix contacts are now well defined.

Table 5.6

A) Cα RMSD values for ten structures calculated with the full NOE data set superimposed on the energy minimized average.

Residues	Average RMSD (Å)
0-129	1.62 ± 0.42
3-127	1.31 ± 0.31
Four helices	0.84 ± 0.24

B) Cα RMSD values for ten preliminary structures with respect to the average calculated structure*.

Residues	Average RMSD (Å)
1-129	3.16 ± 0.34
Four helices	2.71 ± 0.34

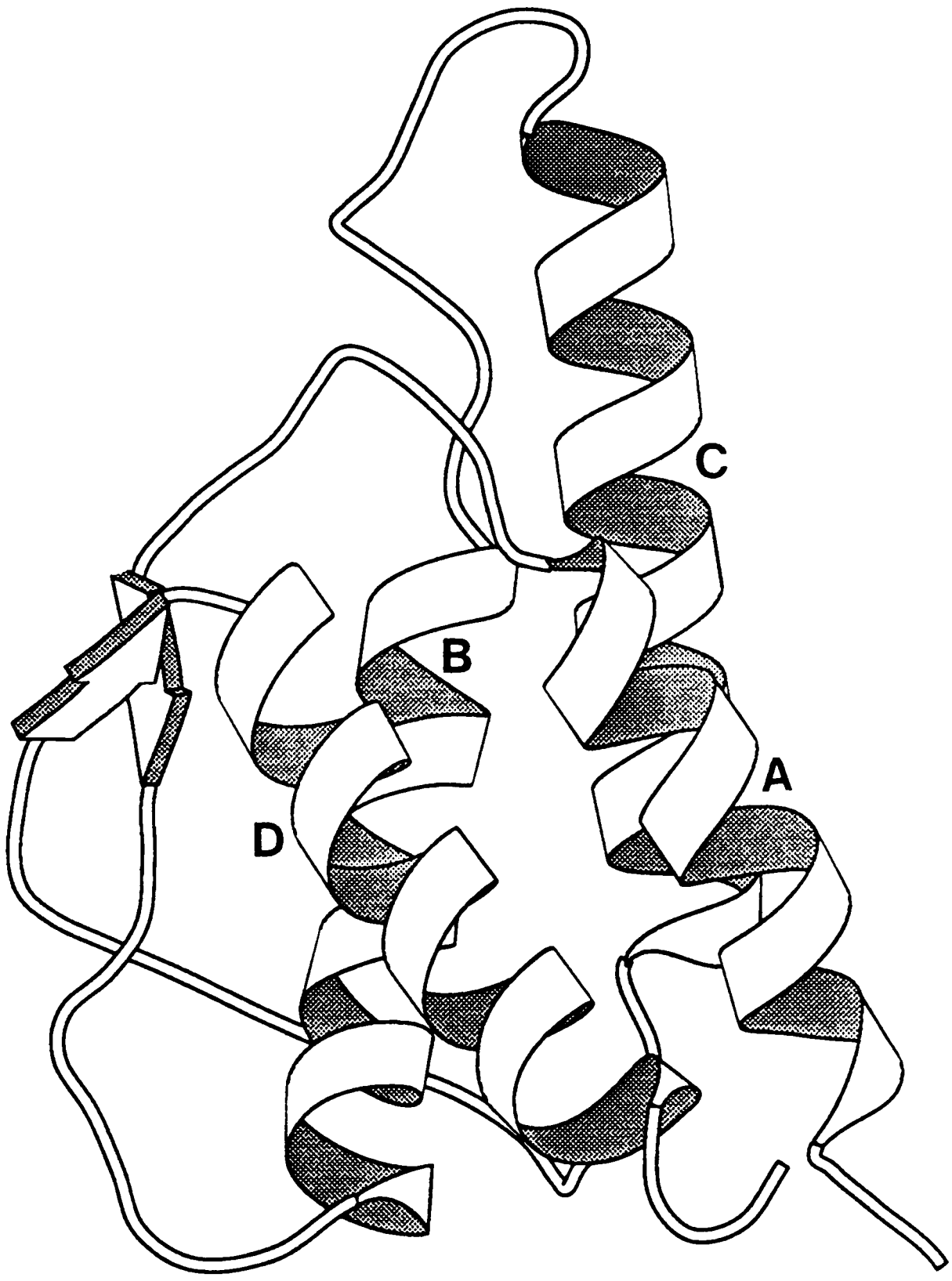
* An energy minimized average structure could not be obtained as the variations between the structures were so large. No N-terminal methionine (residue '0') was included in the preliminary structure calculations. Data used for the calculations included tight ϕ and ψ restraints for helical residues.

5.4.1 Four helix bundle fold.

The ensemble of NMR structures for IL-4 calculated with the full NMR data set confirms the four helix bundle fold recognised from the preliminary NMR data set (Figure 5.16). Presnell and Cohen (1989) have classified four helix bundle structures in terms of the polypeptide chain connectivity and the overall bundle handedness. For a pair of helices two types of connectivity are possible, adjacent and overhand. In an adjacent connection the C-terminus of one helix is adjacent to the N-terminus of the other in space, the polypeptide chain running in the opposite direction in the two helices. In an overhand connection, in contrast, the polypeptide chain passes back over the length of the first helix to enter the second helix in approximately the same direction as it entered the first. The chain therefore runs in the same direction through the two helices.

The relative disposition of the four helices in a bundle determines its handedness. To determine whether a given bundle is right-handed or left-handed imagine grasping the bundle in the right hand. If the thumb of the right hand is oriented along the direction of the polypeptide chain in helix A (thumb nail at C-terminus) and then if helix B is positioned across the palm of the hand and helix C at the fingers (helix D is ignored) then the bundle is defined as right-handed. In the same way, for a left-handed bundle helices A, B and C are positioned respectively at the thumb, palm and fingers of the left hand as it holds the bundle. Alternatively, if the bundle is viewed from the N-terminus of helix A, clockwise motion is needed to go from helix A to B to C in a right-handed bundle and anti-clockwise motion is required for a left-handed bundle. Right-handed and left-handed bundles (excluding the loops) with the same number of overhand connections are therefore related as topological mirror images (see section 5.4.2).

Figure 5.16.
Ribbon diagram of the NMR structure of human IL-4 showing the left-handed four helix bundle fold of the protein with an up-up-down-down connectivity.



The possible different four helix bundle packings are shown in Figure 5.17 and Table 5.7 lists examples of protein structures found in the different categories. By the classification of Presnell & Cohen, IL-4 has two overhand connections (the AB and CD loops run the length of the molecule) giving an up-up-down-down connectivity and the bundle is left-handed (order of helices in an anti-clockwise direction is A, D, B, C when viewed from the N-terminus). A comparison of IL-4 with other proteins in this class (growth hormone, GM-CSF and interferon-β) is given in Chapter 7.

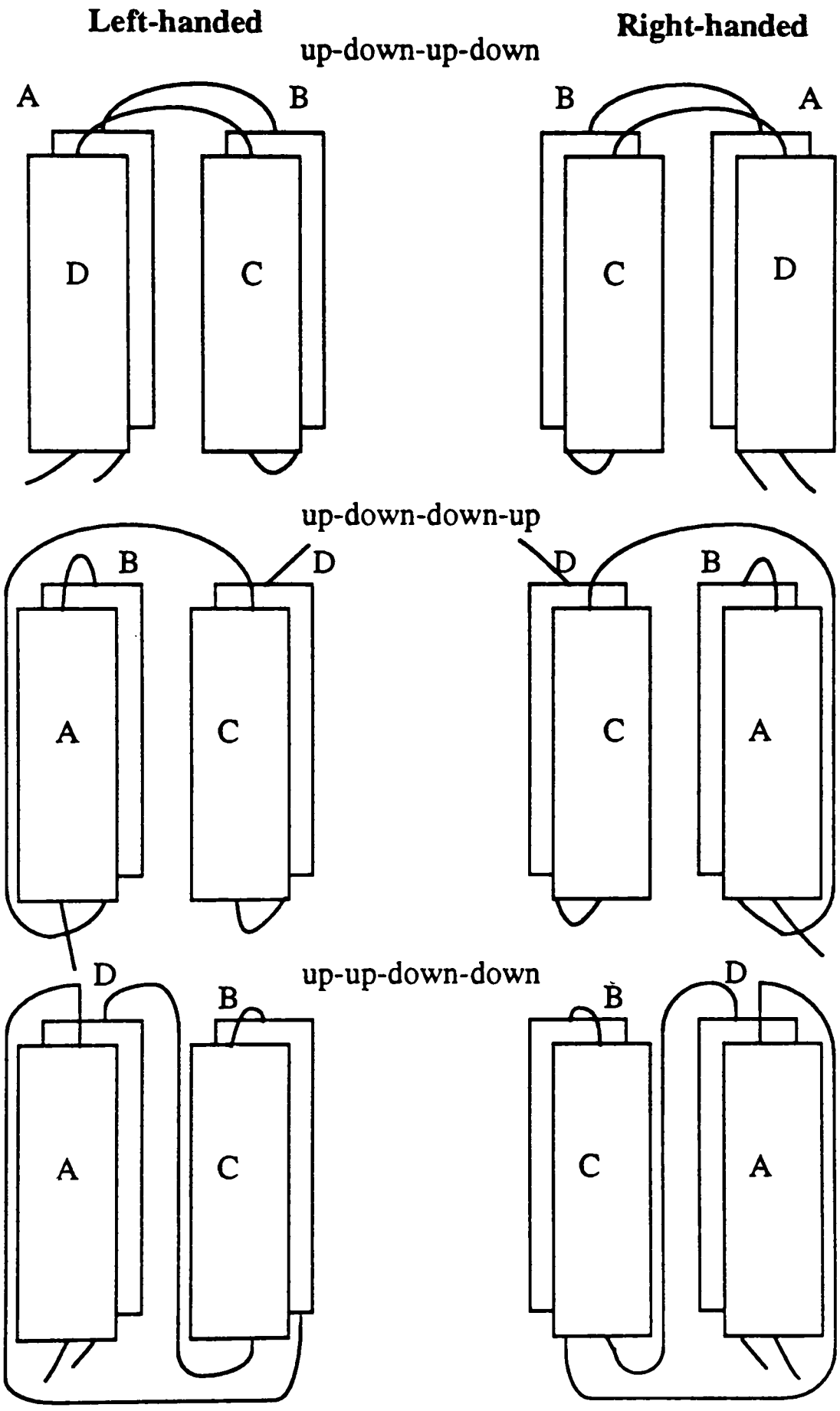
Table 5.7 Examples of four helix bundle proteins (from Presnell & Cohen, 1989).

Overhand connections	Left-handed	Right-handed
0	Cytochrome b_5	Cytochrome b_{562}
	T4 lysozyme	Methemerythrin
	Complement C5a	Cytochrome c'
1	Ferritin	Phospholipase C (b)
2	Growth hormone	
	Murine interferon-β	
	GM-CSF*	

* Granulocyte-Macrophage Colony-Stimulating Factor.

References for protein structures in order listed. Left-handed: Mathews et al. (1972); Weaver & Matthews (1987); Zuiderweg et al. (1988); Banyard et al. (1978); Abdel-Meguid et al. (1987); Senda et al. (1990); Diederichs et al. (1991). Right-handed: Lederer et al. (1981); Stenkamp et al. (1983); Finzel et al. (1985); Hough et al. (1989).

Figure 5.17.
Schematic diagram showing the possible four helix bundle folds (right- and left-handed bundles with 0, 1 or 2 overhand connections).



5.4.2 Topological Mirror images.

It was found that in the distance geometry step of the calculation procedure two sets of initial structures were formed. The structures in each set have Cα RMSD's of 3-4 Å between each other but the RMSD's between structures in the different sets are much larger (≈11 Å). The occurrence of this phenomenon in NMR structure calculations, particularly when there is insufficient structural data, has been considered previously by Pastore et al. (1991) and it has been recognised that one set of structures has its topology inverted with respect to the other. The term 'topological mirror images' has been used to describe the two sets of structures as the chirality of the amino acids is always preserved (L-amino acids) as is, in general, the chirality of secondary structure elements (e.g. right-handed α helices). The difference is found in the way the secondary structure elements are oriented relative to each other.

To select which topological mirror image corresponds to the correct structure in this work, each structure produced by the distance geometry stage of the calculations was inverted and the initial structure and the inverted coordinates were subjected to the simulated annealing stage of the calculation protocol. The greater number and accuracy of the structural restraints used in this next stage (because we are using all atoms rather than just a subset), particularly for the higher resolution structures, means that the correct topology could be easily recognised as it had far lower energy. An example of this is given in Table 5.8.

Table 5.8. Final energies (after full potential dynamics) for two structures that were topological mirror images of each other.

Energy term†	Correct Structure	Inverted form.
E(VDW)	-699.5	178.9
E(Bond)	32.5	256.3
E(Angl)	427.5	1497.1
E(Impr)	7.4	60.2
E(NOE)	102.2	1478.9

† Values in kcal/mol. E(VDW) is the Lennard-Jones energy. E(NOE) is the NOE energy calculated with a force constant of 10 kcal/mol $Å^2$. The energy terms for bonds, angles and improper dihedrals are calculated using the standard XPLOR topology files.

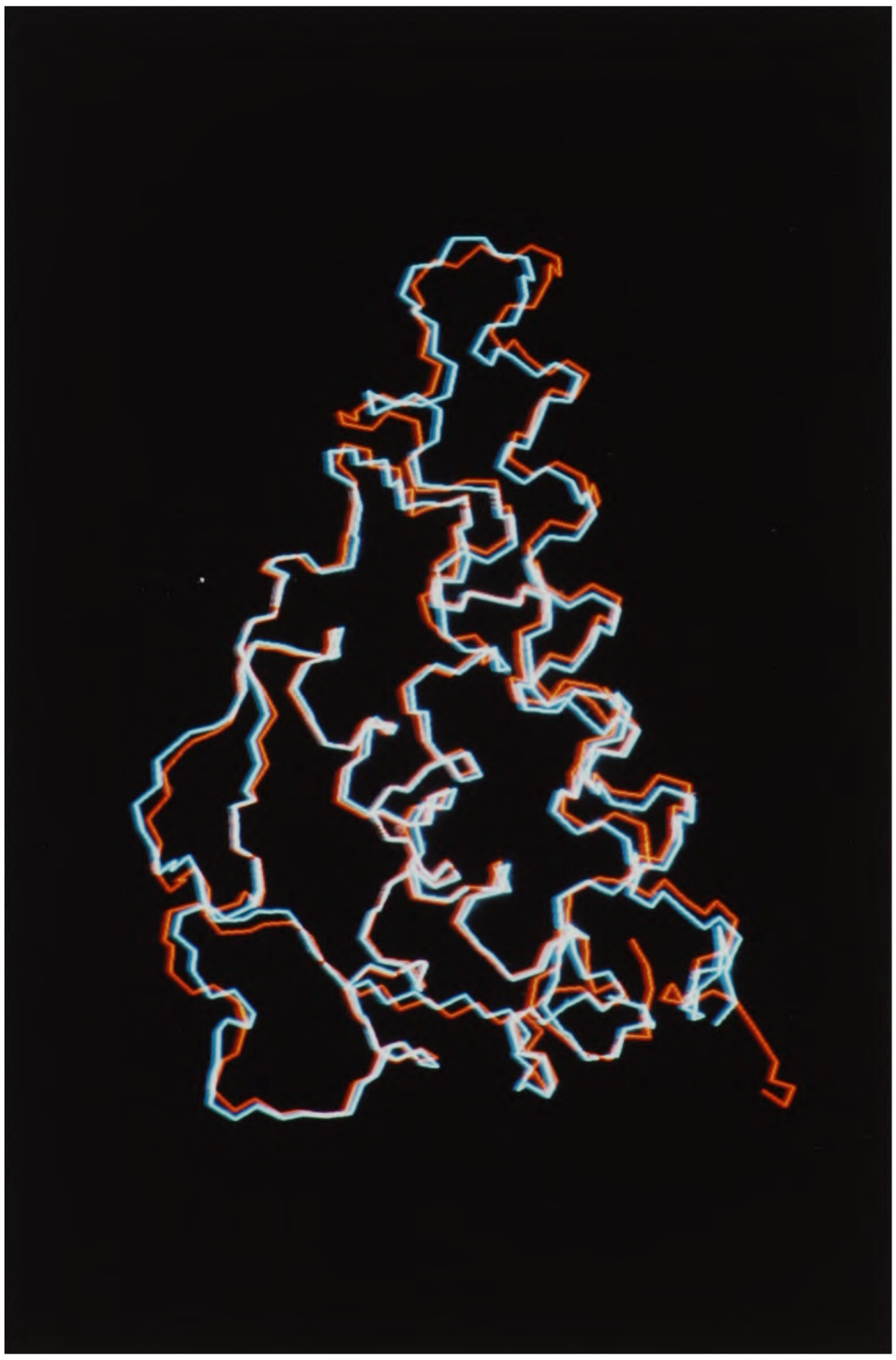
Determining the correct topological mirror image was particularly important in the case of IL-4 because, as is discussed in the previous section, both right-handed and left-handed four helix bundles have been found for proteins (Presnell & Cohen, 1989), the topology of the helix packing being inverted between the two classes of bundle. The correct structures for IL-4 are left-handed four helix bundles while the inverted structures could, although not necessarily would, adopt a right-handed bundle structure. Study of the topology of the incorrect mirror image structures shows that they contain regions of strange geometry such as severely twisted helices indicating clearly that the right-handed four helix bundle topology is not consistent with the NMR data for IL-4.

5.4.3 Calculations using a simulated annealing protocol.

When calculations are run to determine protein conformations that satisfy the NMR data, it is important to ensure that there is adequate sampling of conformational space. Although the distance geometry-dynamical simulated annealing protocol has been widely used and model calculations have been reported in the literature which show that the protocol samples conformational space well (Nilges et al., 1991), it is interesting to compare the results of structure calculations with this protocol with those that use the same NMR data but an alternative calculation procedure. A set of NMR structures has been calculated for IL-4 using a simulated annealing protocol (by Dr. C. Redfield). This protocol starts from a protein conformation with random ϕ and ψ angles (Nilges et al., 1991). In addition, a higher NOE force constant (50 kcal/mol Å^2) is used throughout the simulated annealing. As for the distance geometry-dynamical simulated annealing protocol, an ensemble of 10 NMR structures have been calculated and from these an average coordinate set has been calculated which has been energy minimized.

The agreement of these structures with the NMR data is comparable to that for the distance geometry-dynamical simulated annealing structures (e.g. there are an average of 8 NOE violations > 0.5 Å for both sets of structures) but the RMSD values of the simulated annealing structures from the average are slightly lower. For example the Cα RMSD for residues 0-129 is 1.41 ± 0.30 Å for the simulated annealing structures compared with a value of 1.62 ± 0.42 Å for the distance geometry-dynamical simulated annealing structures. This probably reflects the higher NOE force constant used in the simulated annealing calculations. Comparison of the two average structures shows, however, that they are closely similar (see Figure 5.18). The Cα RMSD's between them are 1.3 Å (all residues), 0.81 Å (residues 3-127) and 0.55 Å (four α helices). The regions where there are slight differences in the

Figure 5.18.
Superposition of the main chain of two average energy minimized structure of human IL-4. The structure shown in red was calculated using the distance geometry-dynamical simulated annealing protocol (Nilges et al., 1988) while the structure in cyan was calculated using a simulated annealing protocol (Nilges et al., 1991). The same NMR data set was used in both calculations but they were performed independently.



conformation of the two average structures lie in parts of the loops where there is disorder in the ensemble of NMR structures e.g. residues 69-71 in the BC loop (see section 6.2.2). Various similarities and differences between the two sets of structures, such as in side chain conformations, are discussed further in Chapter 6 but, in general, the close agreement of the results from the two calculation procedures suggests that conformational space is being searched correctly in the calculation protocol used in this work.

5.5 Conclusions.

In this chapter the methods and procedures used to determine a high resolution structure for human IL-4 have been presented. The strategy involving use of first a ^{15}N labelled sample and then a double labelled (^{15}N and ^{13}C) sample proved remarkably efficient despite the problems arising from the high helical content of the protein; indeed the complete assignment and structure determination process took under one year to complete although an analysis of the results and their significance is still in progress. The coordinates of the NMR structure of human IL-4 have been deposited in the Brookhaven Protein Data Bank (accession number 1ITL).

In the next chapter a detailed analysis of the set of high resolution structures of human interleukin-4, calculated using the distance geometry-dynamical simulated annealing protocol, is given and in Chapter 7 the structure is compared with proteins with related folds and its overall significance is discussed.

CHAPTER 6.

THE STRUCTURE OF HUMAN INTERLEUKIN-4.

Detailed analysis of the NMR structures.

In Chapter 5 the procedures and techniques used to determine the structure of the four helix bundle protein, human interleukin-4, by NMR spectroscopy were presented. In this chapter the NMR structure for human IL-4, calculated using the distance geometry-dynamical simulated annealing protocol, is described in detail and is compared to the results of structure calculations performed indepedently (but with the same data set) using an alternative simulated annealing protocol. Three main features of the protein are considered; its secondary structure, its tertiary structure and a description of the side chains in the structure. In each of these sections the results of the structure calculations are presented and these results are then discussed in the light of the input NMR data and any additional NMR information that are available.

After the completion of this structure determination, Powers et al. (1992a) published an independent NMR structure for IL-4, the results of which are closely similar to those presented here. Some minor differences exist between the structures, however, which are also discussed in this chapter. Their studies used a recombinant variant of IL-4 with an N-terminal Glu-Ala-Glu-Ala sequence. In addition, the two potential N-glycosylation sites in the protein, Asn 38 and Asn 105, had been changed to Asp by site-directed mutagenesis (Powers et al., 1992b).

6.1 Secondary structure.

As summarized in Chapter 5, the main regions of secondary structure in IL-4 are four α helices and a short region of antiparallel β sheet. In this section we analyse these secondary structure units in detail in the calculated NMR structures and also look at the conformations in the connecting loop regions. The Ramachandran plots for the whole structure of IL-4, both for the ensemble of ten calculated NMR structures and for the average (energy minimized) structure, are shown in Figure 6.1. 86 of the residues in the average structure fall into the core areas of the Ramachandran plot as defined by Morris et al. (1991) (the core regions contain 80% of the residues found in high resolution crystal structures of proteins); 64 of these residues are in the helices or β sheet regions of the structure, reflecting the good definition of the secondary structure core of the molecule. This is again shown when the ϕ and ψ angles in the average energy minimized structures calculated by the two different protocols are compared (Figure 6.2). For 92 of the residues the angle differences, for both ϕ and ψ, are less than 20°. The residues where the differences are large (> 40°) are all concentrated in short regions of the loops (see sections 6.1.3 and 6.2.2). It is interesting that where there is a large ϕ difference it is usually accompanied by a large ψ difference for the preceding residue. These concerted differences give approximately the same relative Cα position for the two conformations and hence do not alter the overall direction of the polypeptide chain. The stereochemical quality of the NMR structures of IL-4 is discussed further in the Postscript to the thesis where an analysis using the program PROCHECK (MacArthur et al., 1992) is presented (results of this analysis are given in Table 6 of Postscript Page 221).

Figure 6.1a.
Ramachandran plot showing the ϕ, ψ angles for all of the ensemble of 10 calculated NMR structures (all residues).

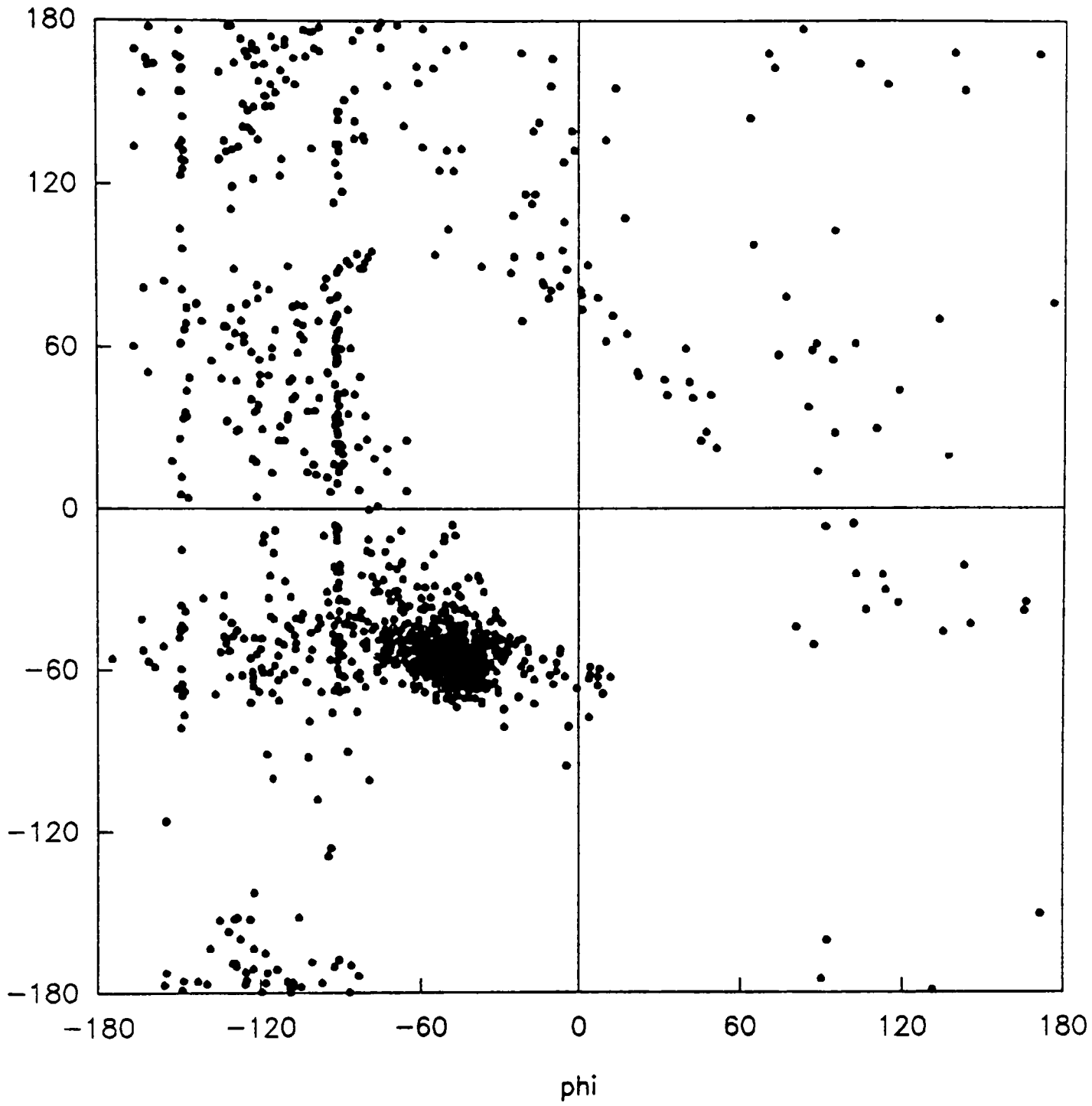


Figure 6.1b.
Ramachandran plot showing the ϕ, ψ angles in the average energy minimized NMR structure (all residues).

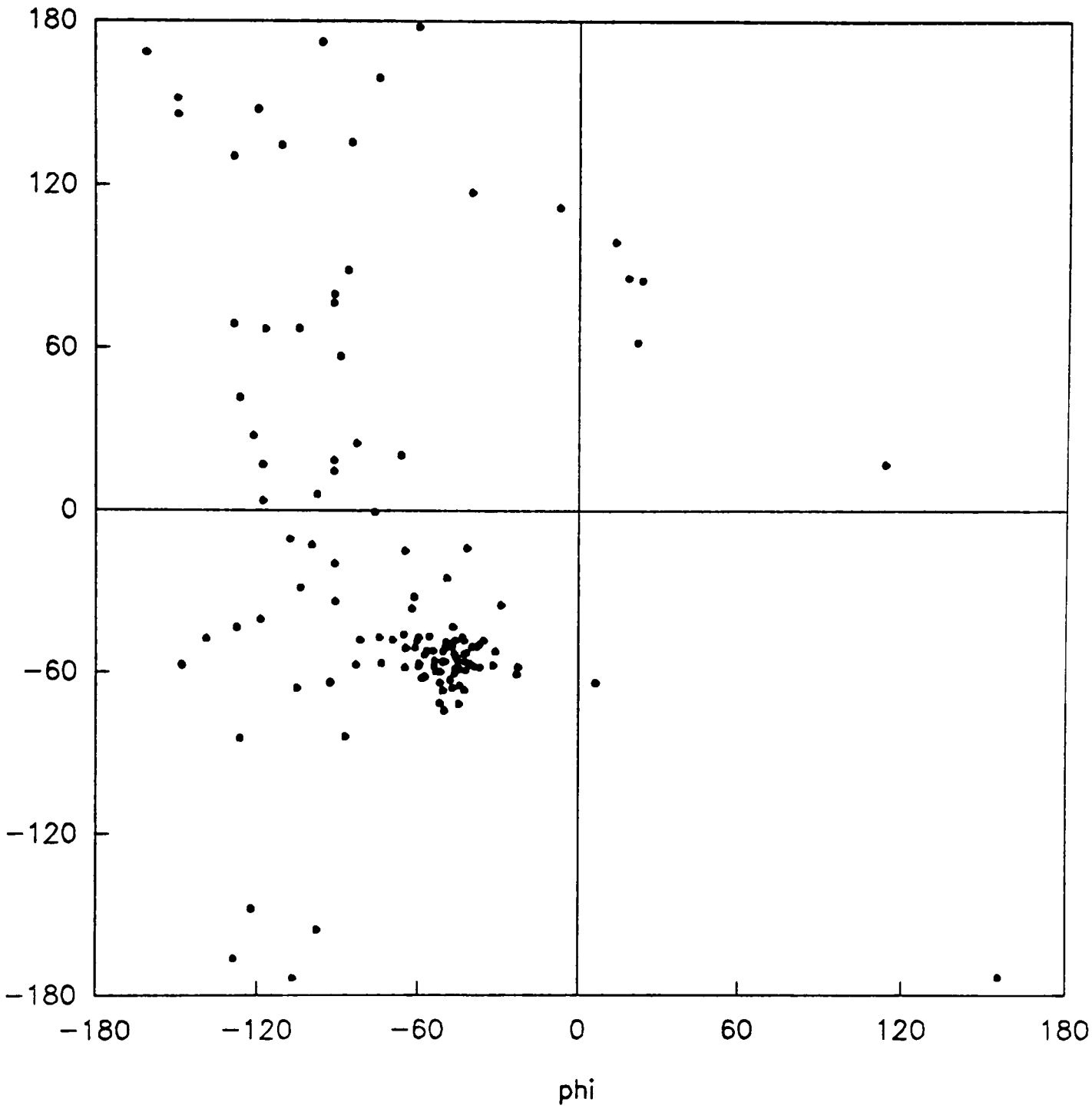
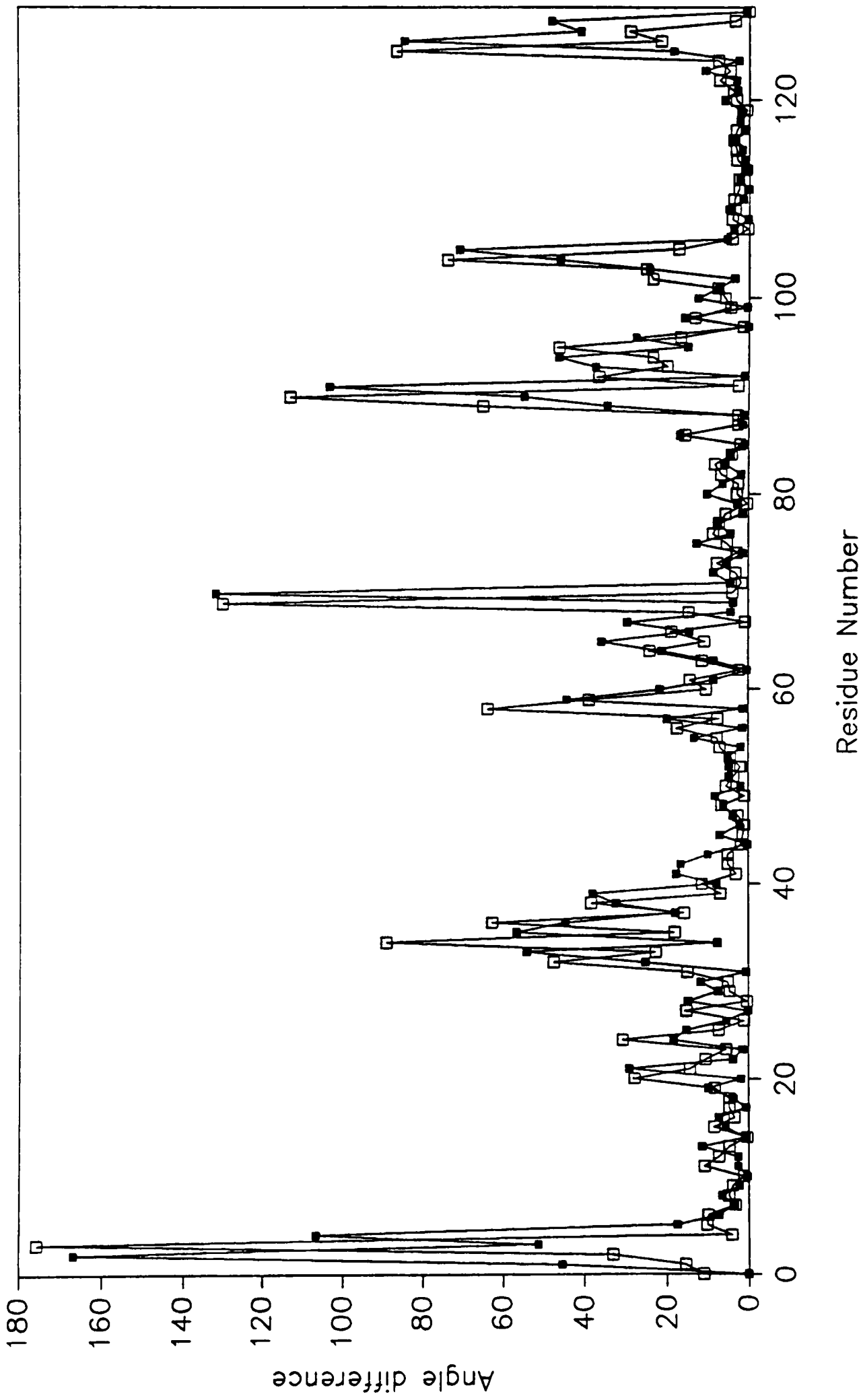


Figure 6.2.
Difference between the ϕ and ψ angles (filled and open boxes respectively) in the average energy minimized NMR structures calculated using the distance geometry-dynamical simulated annealing and simulated annealing protocols respectively.



6.1.1 Helices.

On the basis of the short and medium range NOE data, the coupling constants and the amide proton exchange rates, four major helical regions were identified in IL-4 as described in section 5.2.1; A (5-17), B (41-57), C (72-90) and D (110-125). The Ramachandran plots for these residues (Figure 6.3) show that all the four regions form regular, well defined helices in the calculated structures. This is also illustrated (for helix B) in Figure 6.4 which shows the small range of both ϕ and ψ observed for these residues in the calculated structures. Note that, although there were initial problems in assigning residue 53 in the ^{15}N spectra (its resonance was subsequently found in spectra recorded under different pH and temperature conditions), and no ϕ angle or hydrogen bond restraints were included for the residue (as the coupling constant and amide exchange rates have not been measured) there are no kinks or distortions in helix B. The small number of points in the Ramachandran plots for the helices that are not in the region around ϕ=-57°, ψ=-47° come from some of the N- and C- terminal residues of the helices (Leu 17, Asn 89, Leu 90, Ser 125) reflecting some irregularity of the helices at their ends and, in one of the structures, from helix C (in this structure residues 73-77 are nonhelical; shown by open boxes in plot). In general though, the precision of the main chain conformation for all the helices probably reflects the tight hydrogen bond restraints that have been used, where there were slowly exchanging amides, in the structure calculations.

Comparison with the NMR work of Powers et al. (1992a) shows that there are differences in the precise regions where helical restraints were included in the two sets of structure calculations (Table 6.1). These result, however, from the different criteria used to identify the helices, rather than from differences in the NMR data set.

Figure 6.3a.
Ramachandran plot showing the ϕ, ψ angles for all of the ensemble of 10 calculated NMR structures. Only the points for the residues in the recognised helices are shown (residues 5-17, 41-57, 72-90 and 110-125). Structure 8 has an irregular C helix; the ϕ, ψ angles for this structure are shown by open boxes.

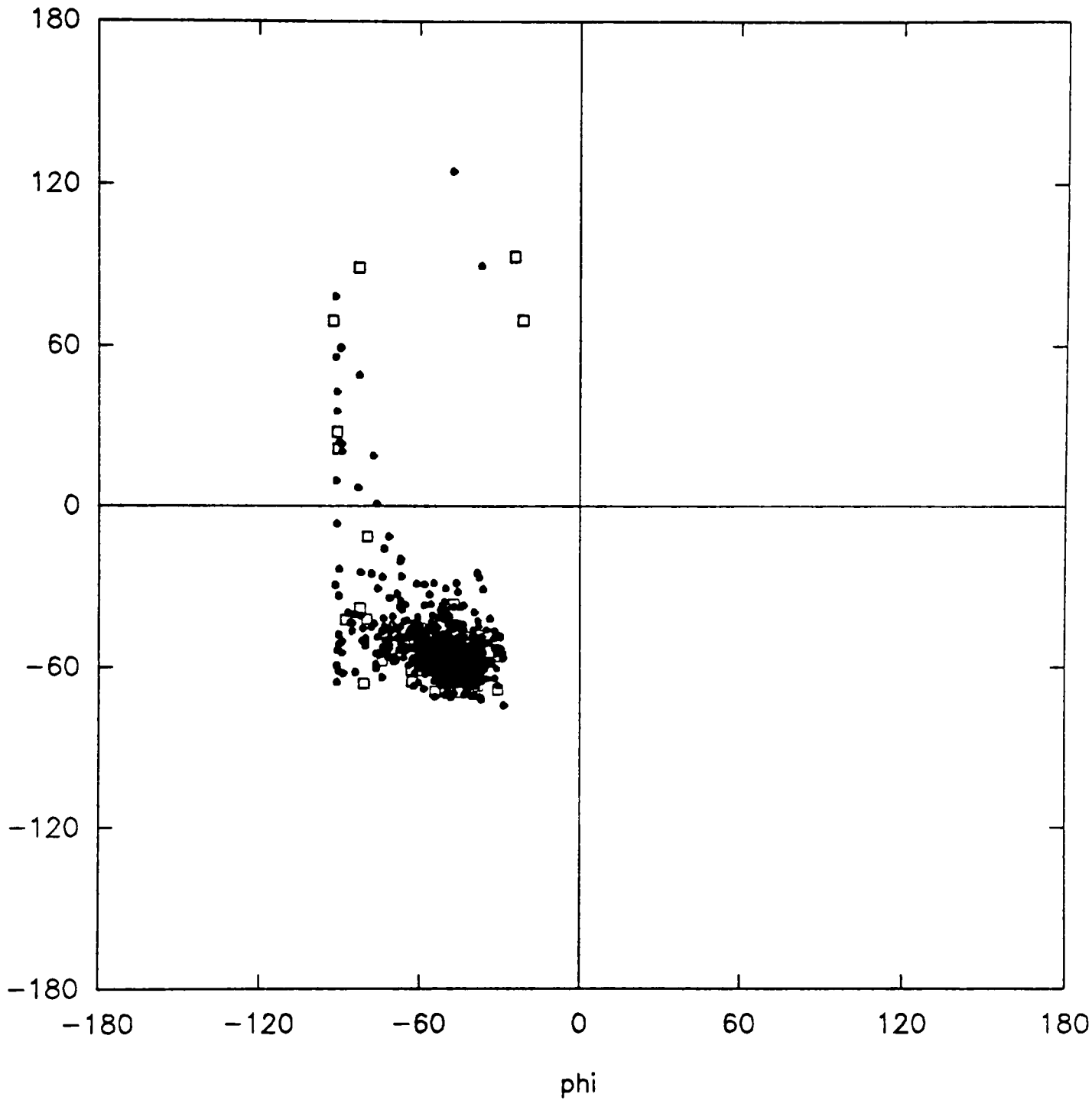


Figure 6.3b.
Ramachandran plot showing the ϕ, ψ angles in the helices in the average energy minimized NMR structure.

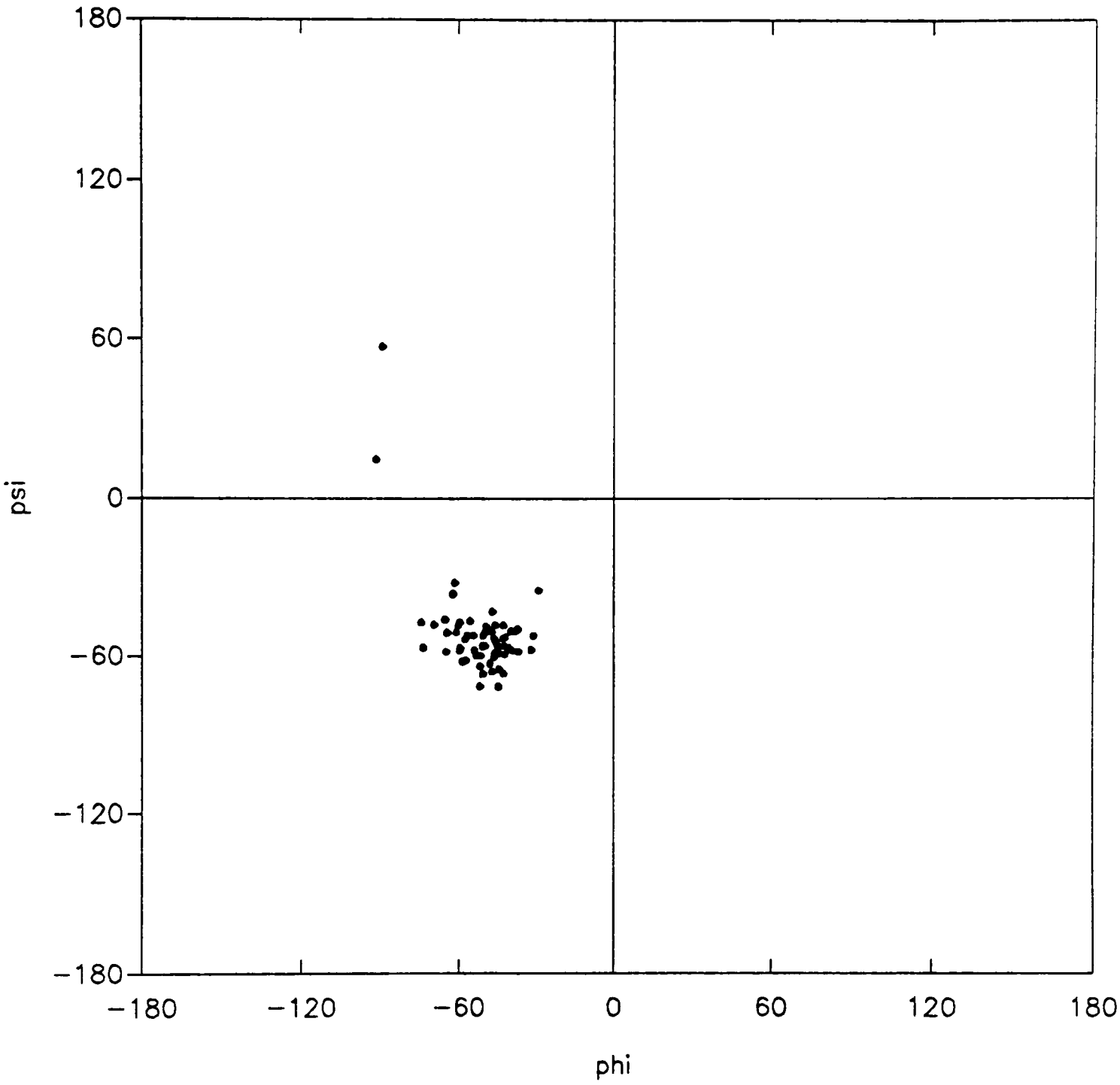
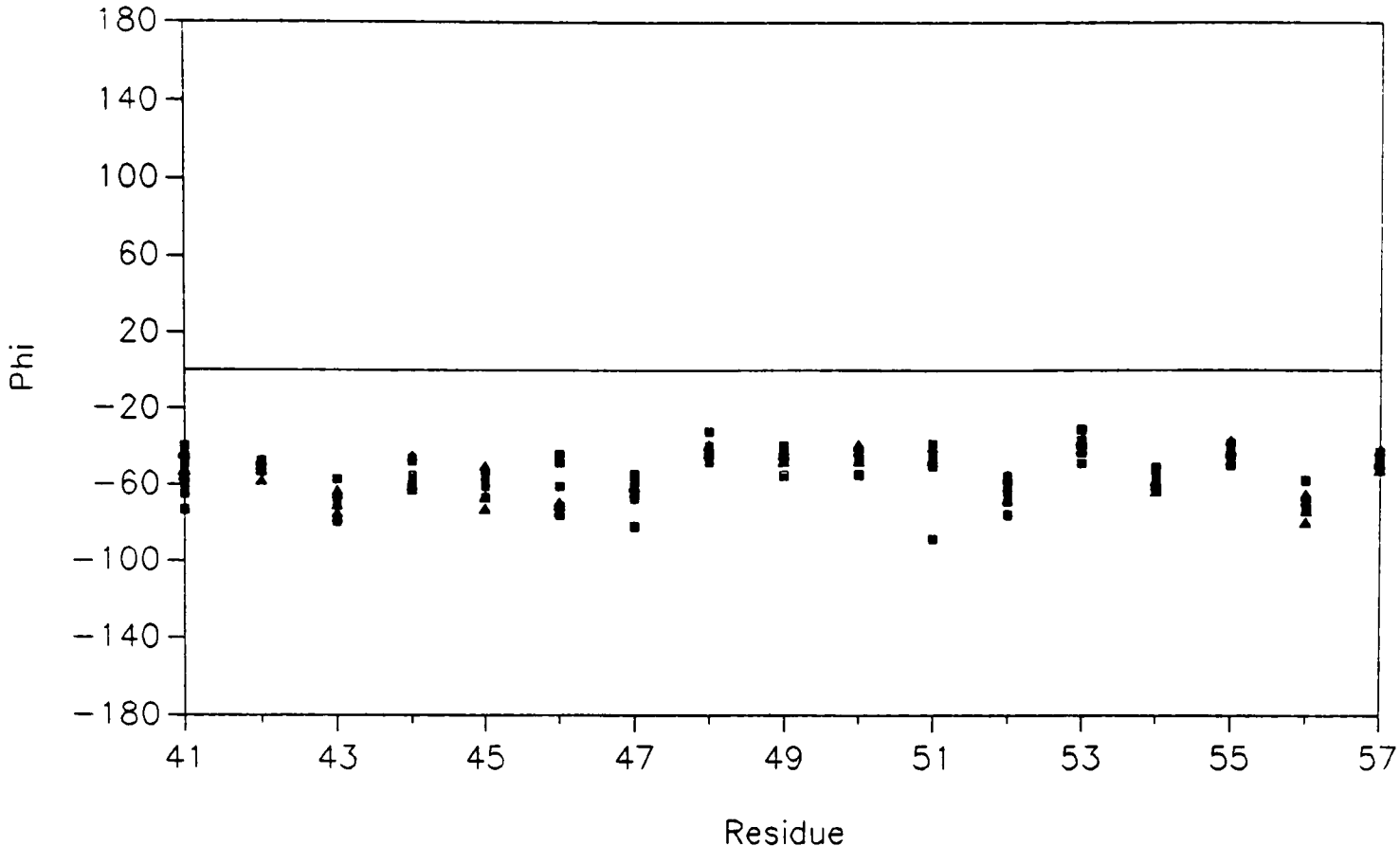


Figure 6.4.

Graphs showing the ϕ and ψ angles in each of the calculated NMR structures for residues 41-57 in the B helix of IL-4.



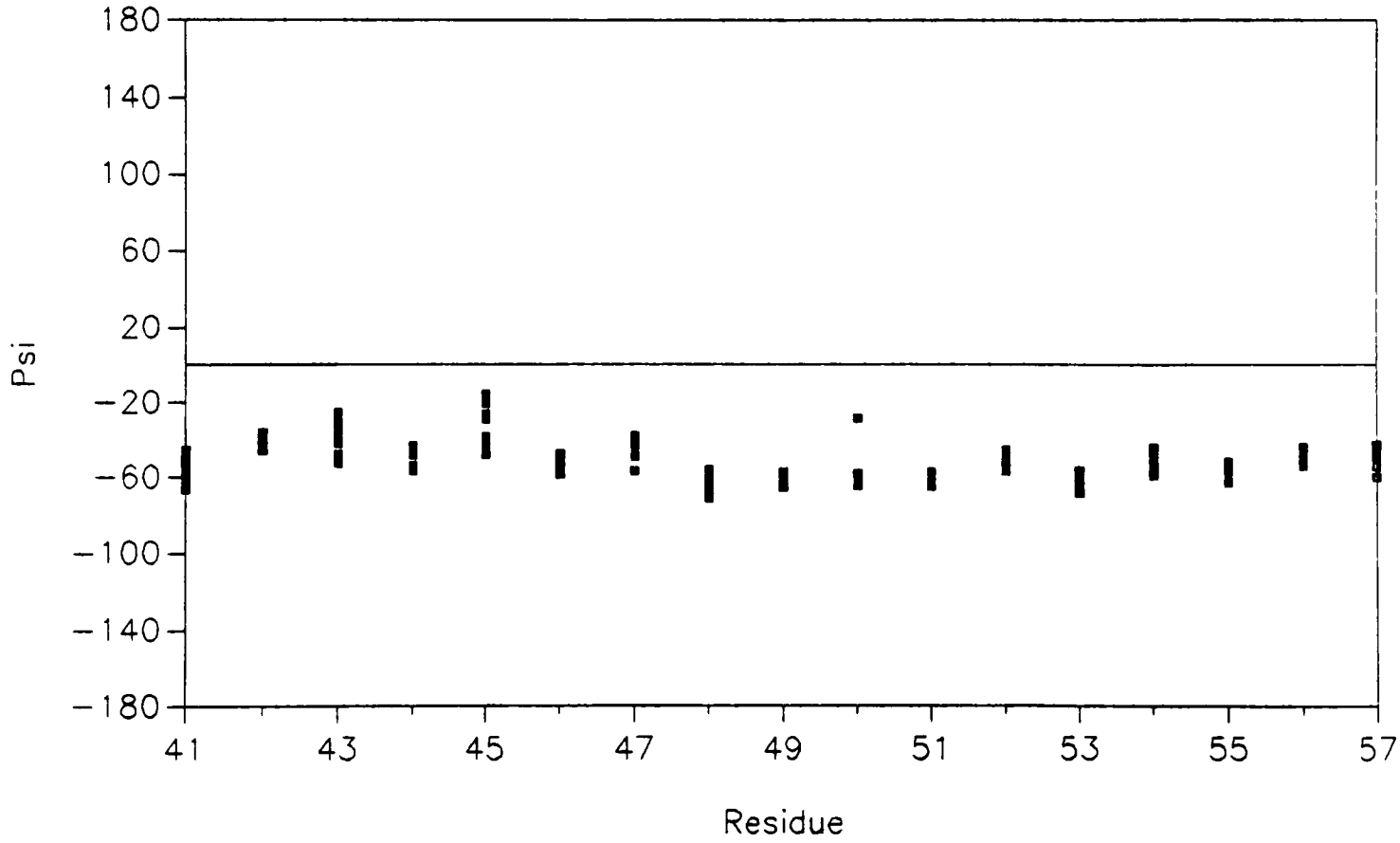


Table 6.1 A comparison of the position of the helices in IL-4 in this work and in that of Powers et al. (1992a).

Helix	This work	Powers et al.*
A	5-17	5-17 (or 3-20)
B	41-57	41-60
C	72-90	70-92
D	110-125	109-125

* The same numbering is used here for both structures; in Powers et al. (1992a) the residue numbers are increased by four due to an N-terminal Glu-Ala-Glu-Ala in their recombinant protein.

Powers et al. used ^{13}C chemical shifts as a basis for their recognition of the start and end points of the helices (Garrett et al., 1992), α helical residues being identified as those with a $^{13}C\alpha$ chemical shift greater than +2 ppm and a $^{13}C\beta$ chemical shift less than +1.2 ppm. In this work a more conservative set of criteria, which identifies residues with a regular helical conformation, were used based on a combination of NOE, $^{3}J_{HN\alpha}$ and amide exchange rate data. For a residue to be included in a helix, $^{3}J_{HN\alpha}$ was required to be less than 6 Hz and the majority of helical type NOE's arising from this residue (as listed by Wüthrich et al., 1984) must have been identified. In addition, a helix was generally recognised to start four residues before the first slowly exchanging amide proton (although for helix C this criterion could not be used as there are so few slowly exchanging amide protons).

The situation with helix A in the work of Powers et al. is unclear. Helix A is initially recognized to run from residue 5 to residue 17 but in the

subsequent structure calculations helical restraints appear to be included for residues 3-20. In this work helix A was chosen to start at residue 5, as the first slowly exchanging amide is found at residue 9, and to extend to residue 17. Extension of the regular helix beyond residue 17 cannot occur as residue 18 has a $^3J_{HN\alpha}$ value of 7.4 Hz.

In the case of helix B, the helix starts at residue 41 ($^3J_{HN\alpha} < 5Hz$) as residue 40 has a $^3J_{HN\alpha}$ value of 7.4 Hz. In this work helix B finishes at residue 57 but Powers et al. extend it by three residues (58-60). Residues 58 and 59, however, have $^3J_{HN\alpha}$ coupling constant values of 7.6 and 8.8 Hz respectively so they cannot have a regular helical conformation.

For helix C, in our structures 72 was chosen as the first residue of the helix, as the majority of helical NOE's (i to i+2, i+3 and i+4) start at this residue. Helix C was not extended beyond residue 90 both because the $^3J_{HN\alpha}$ values could not be obtained for residues 91 (due to broad resonances) and 92 (glycine), and because residue 91 does not have a slowly exchanging amide proton. Indeed there are very few slowly exchanging amides in total in this helix which makes it difficult to define its exact position. In the work of Powers et al. helix C has two additional residues both at its start (70-72) and at its end (90-92). These extensions are possible but it would make helix C very long in comparison with the rest of the bundle.

For helix D, although residue 109 has a $^3J_{HN\alpha}$ value less than 4 Hz, a regular helix was not defined here until residue 110. This is because the β sheet extends as far as residue 108 and it is unlikely that a regular helix could immediately follow it. Powers et al. have, however, included 109 in their helix D. Helix D finishes in both structures at residue 125, the last residue with a measured $^3J_{HN\alpha}$ value less than 6Hz.

It is important to notice here that this discussion has been concerned with the recognition of helices from NMR data and the residues for which helical ϕ angle and hydrogen bond restraints were included in the structure calculations. Other sets of criteria are used to recognise helices from the coordinates of a protein structure. For example, the method of Kabsch & Sanders (1983) is often used in crystallographic work to define helices on the basis of hydrogen bond identification. This is discussed further in the Postscript to this thesis.

6.1.2 β sheet.

The short region of antiparallel β sheet in IL-4 (residues 28-30 and 106-108) connects the AB and CD loops and is likely to be an important factor in the stabilization of these loop regions. It is, however, not very regular. The average ϕ and ψ angles seen in the set of NMR structures in this region are given in Table 6.2; for 3 of the angles the value is more than 30° away from that for a regular β sheet (ϕ -139°, ψ +135°). This is not, however, surprising as the $^3J_{HN\alpha}$ coupling constant measurements show that the conformation in this region is irregular (for 3 residues the $^3J_{HN\alpha}$ values $\leq$ 7 Hz). The correlation, however, of the experimental $^3J_{HN\alpha}$ values with those predicted from the average ϕ angles in the calculated structures for two of these residues is poor (Table 6.2).

Table 6.2 Torsion angles and coupling constants for the β sheet in IL-4.

Residue	ϕ*	ψ	$^3J_{HN\alpha}$ calc.#	$^3J_{HN\alpha}$ expt.
28	-18.6 ± 3.6 (5)	118.3 ± 11.4	1.9	7.0
	-72.2 ± 18 (4)		5.7	
	-147.5 (1)		8.2	
29	-100.8 ± 7.4	-162.7 ± 28.1	8.9	9.1
30	-148.2 ± 0.1 (3)	137.9 ± 6.0	8.1	< 5
	82.5 ± 7.6 (7)		7.0	
106	-116 ± 3.7 (9)	150.6 ± 7.3	9.6	not known
	-1.7 (1)		2.6	
107	-128.8 ± 5.3	136 ± 18.2	9.5	> 7
108	-148.8 (1)	90.5 ± 3.9	8.0	5.6
	-84.2 ± 4.4 (9)		7.2	

* The means and standard deviations of ϕ and ψ angles seen the ensemble of ten structures are given (in degrees). In four of the residues a single range is not observed. For these angles two or three means are given, with the number of structures adopting each in brackets.

Calculated using the Karplus relationship (Chapter 2). All coupling constant values are in Hertz.

6.1.3 Loop conformations.

IL-4 contains three long loop regions that connect up the four helices (AB, BC, CD). The β sheet occurs in the middle of the AB loop (breaking this up into two shorter regions, AB1 and AB2) and at the end of the CD loop. If we look at the conformation of the residues in the loop regions we find that in many cases they are poorly defined, with a whole range of torsion angles being observed across the ensemble of structures. Indeed, it is generally only

for the residues with $^3J_{HN\alpha}$ values greater than 8 Hz, where a ϕ restraint of -120 ± 40° has been included (for 9 residues in the loops), that the conformation appears relatively well defined (see for example part of the BC loop in Figure 6.5). These disordered conformations result in relatively high Cα RMSD values when the loops are locally superimposed (see Table 6.3) and large Cα deviations across the ensemble of structures for certain regions of the loops when the entire protein is superimposed (Figure 6.6).

Table 6.3 Cα RMSD values for the loop regions in IL-4*.

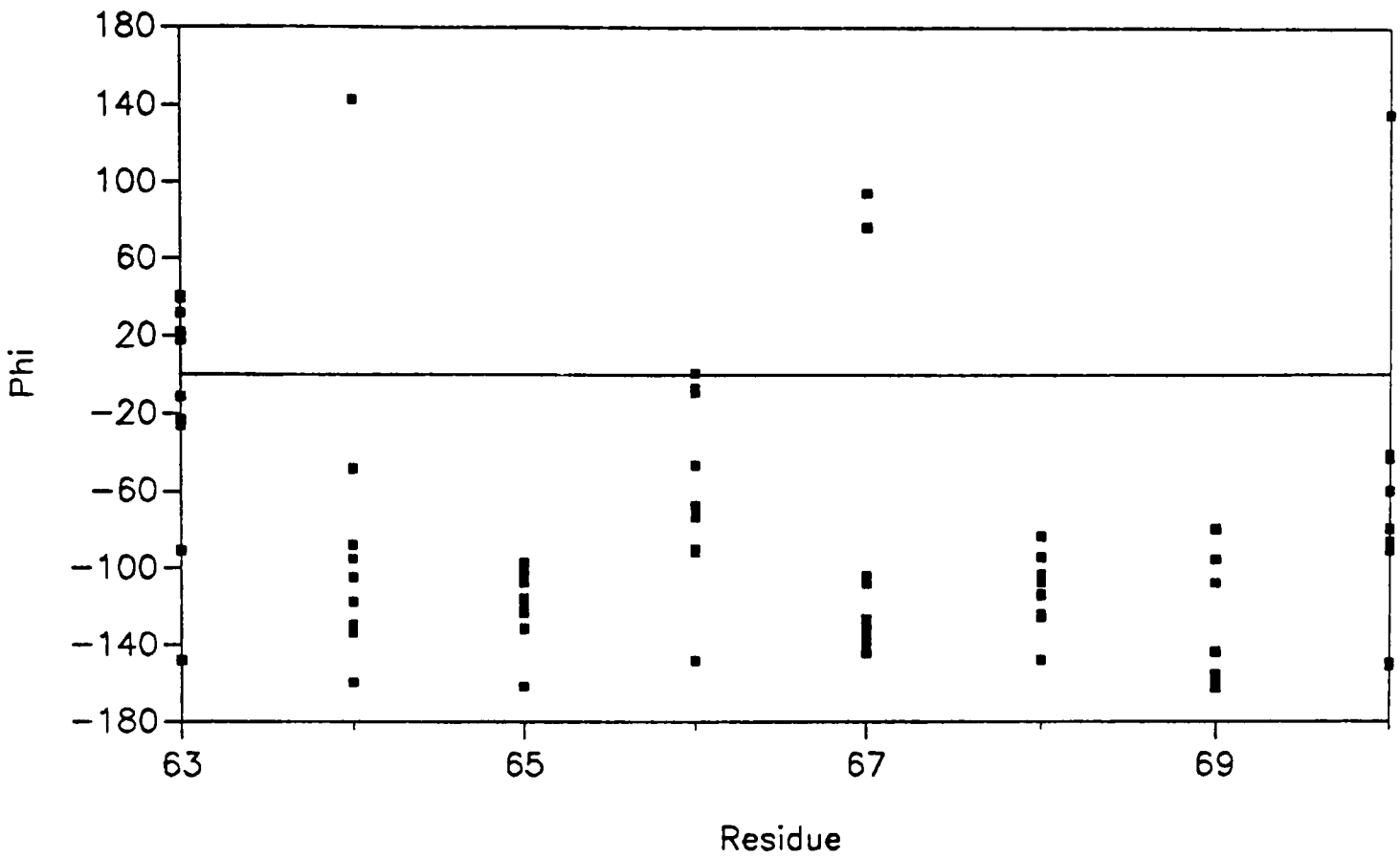
Loop	Local RMSD	Global RMSD
AB (18-40)	1.21 ± 0.26 Å	1.48 ± 0.34 Å
BC (58-71)	1.39 ± 0.44 Å	1.74 ± 0.46 Å
CD (91-109)	0.98 ± 0.28 Å	1.31 ± 0.30 Å

* Means and standard deviations are given for ten structures superimposed on the average structure. In the local superposition just the residues in the RMSD calculation are superimposed on the average. In the global superposition residues 0-129 are superimposed on the average.

For the BC loop there is a particularly large variation in the Cα RMSD values in the individual structures. In some cases (e.g. structures 1, 2 and 8 - corresponding RMSD's for local superposition 0.91, 0.97, 0.92 Å) the loops are very similar to that in the average while in other cases the conformations are very different (e.g. structures 4, 5 - corresponding RMSD's for local superposition 2.19, 2.03 Å). The question therefore arises as to whether there may be two families of conformations in this BC loop within the ensemble of calculated structures. Further analysis, however, showed that this was not

Figure 6.5.

Graphs showing the ϕ and ψ angles in each of the calculated NMR structures for residues 63-70 in the BC loop of IL-4. ϕ restraints of -120±40° were included for residues 65, 68 and 69 in the structure calculations.



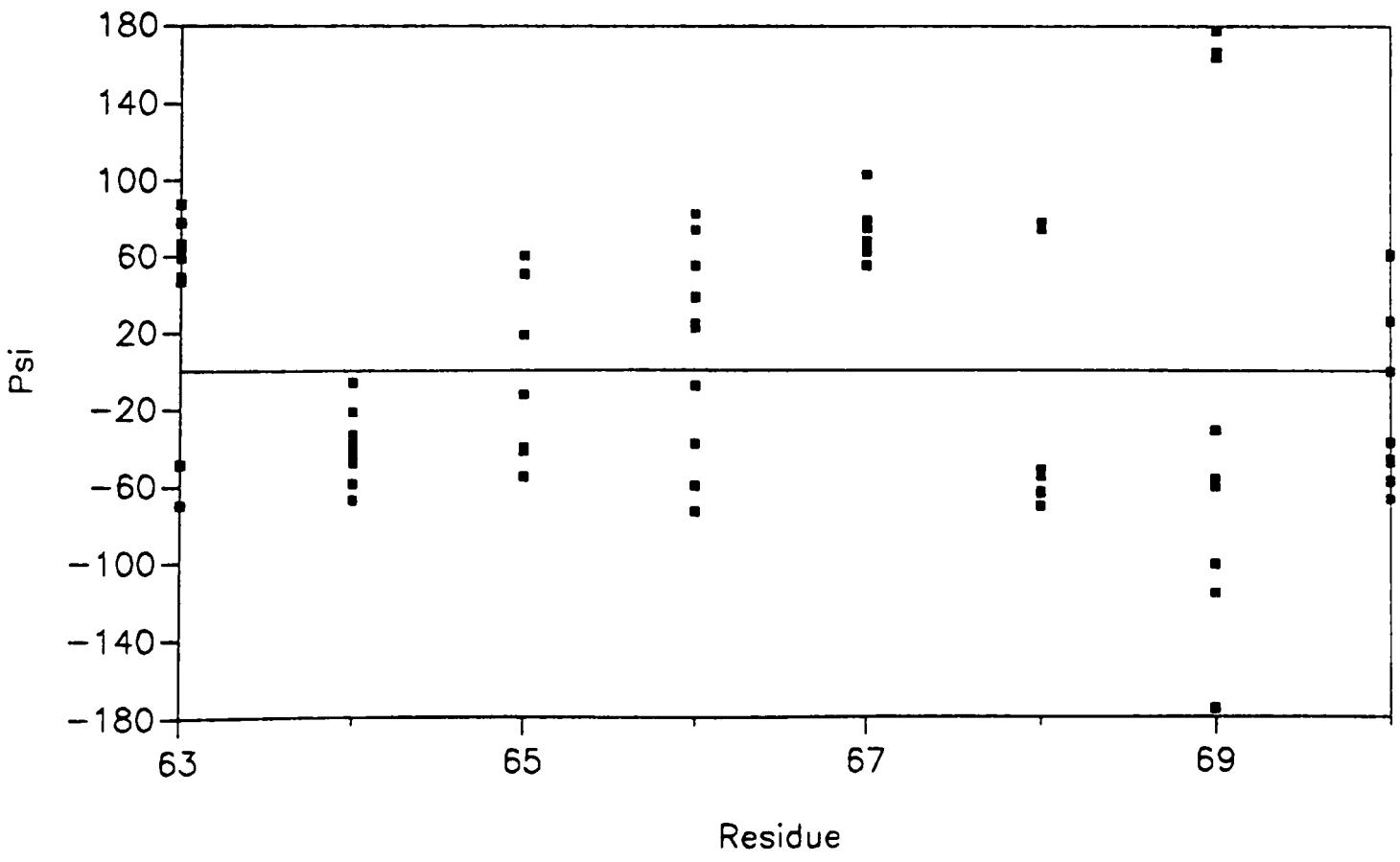
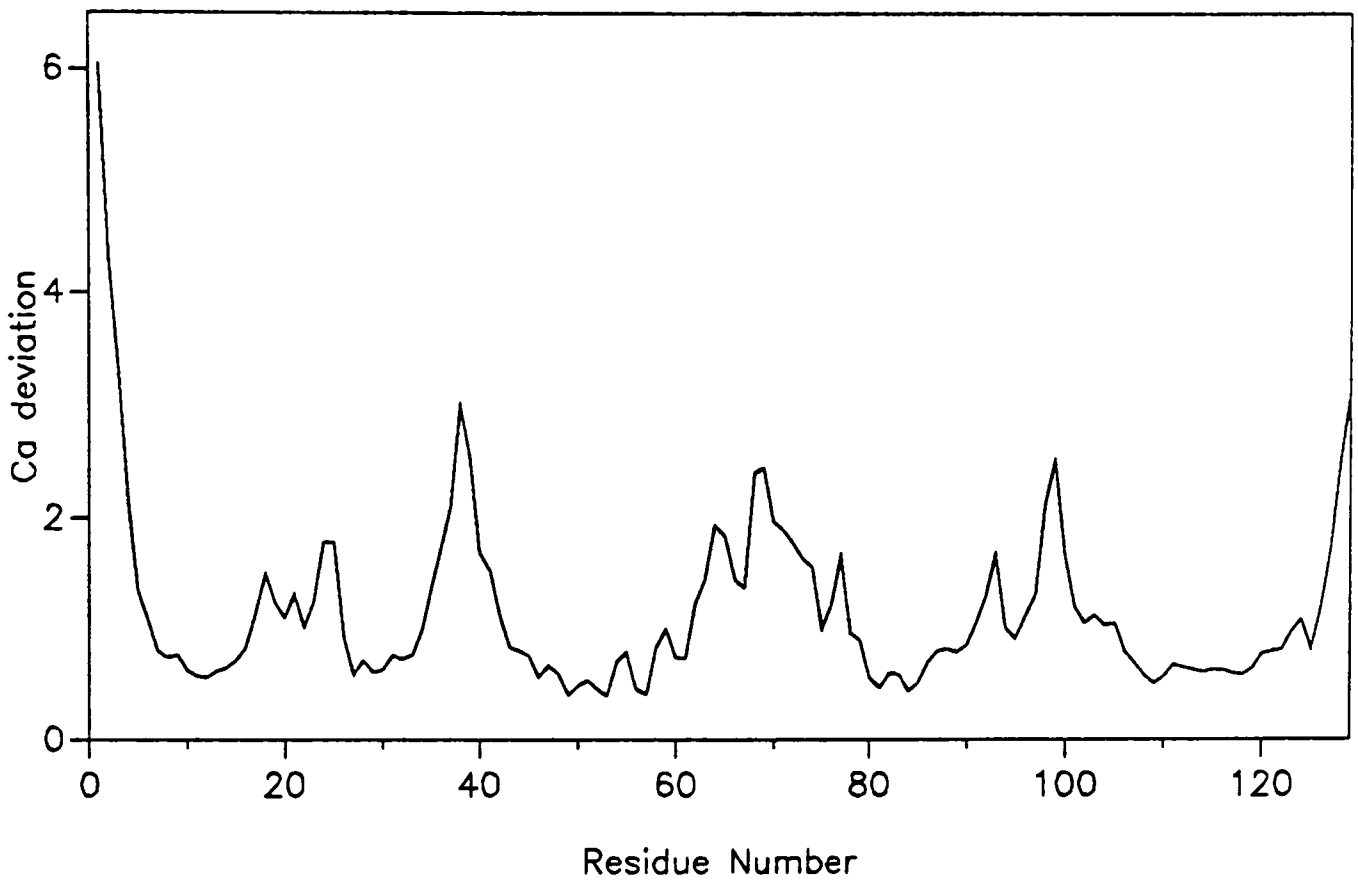


Figure 6.6.

Average Cα deviation per residue (in Å) from the average energy minimized structure for the ensemble of 10 NMR structures. The Cα atoms of residues 0-129 were globally superimposed onto the average for the calculation.

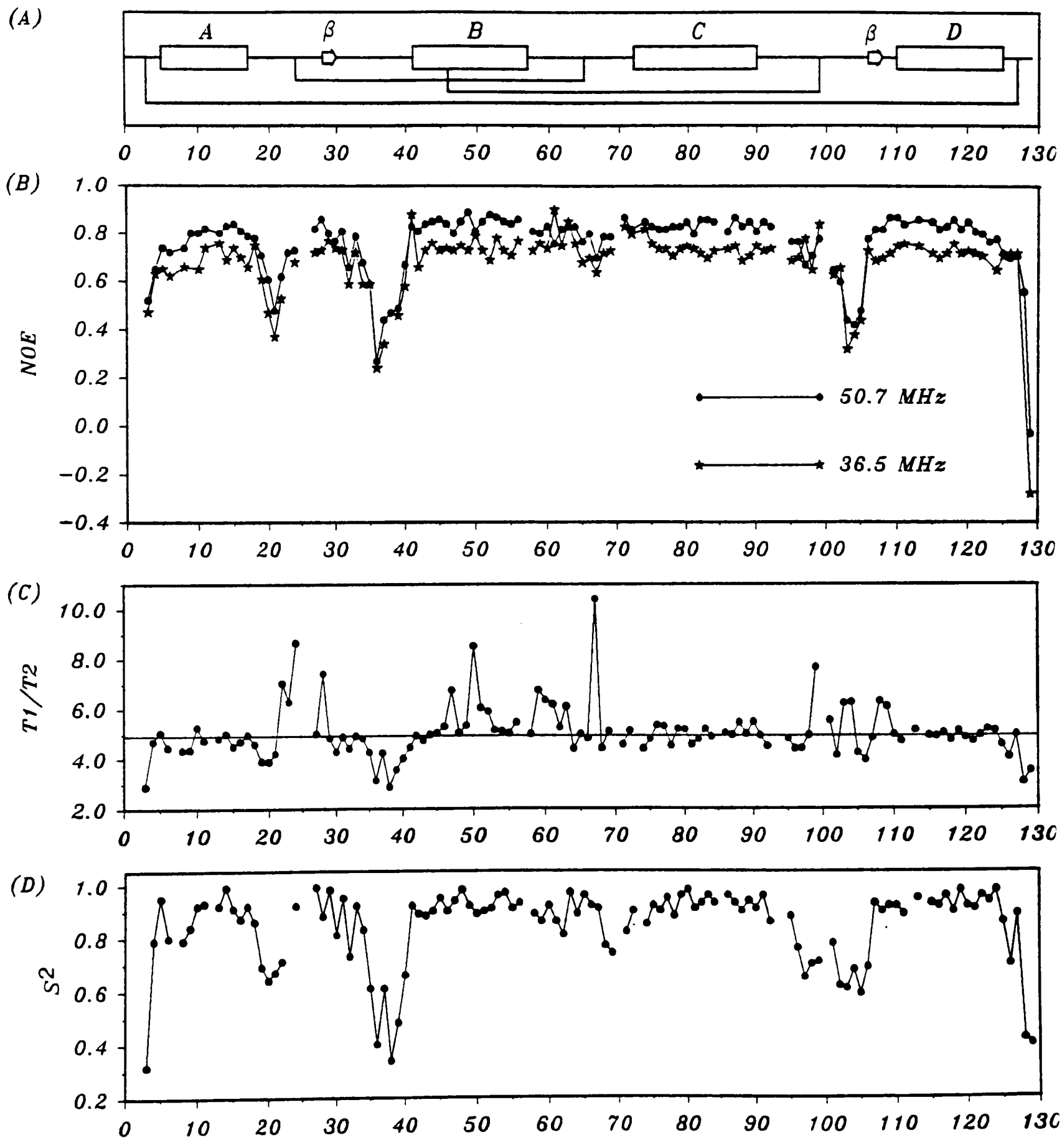


the case, the loops with large RMSD values also being very different from each other as well as from the average.

In the light of the conformational disorder of many of the loop residues, the results of ^{15}N relaxation studies that have been carried out for IL-4 (Redfield et al., 1992) are particularly interesting. These ^{15}N NOE, T_1 and T_2 measurements are able to probe backbone dynamics at the level of individual residues; the results for IL-4 are summarised in Figure 6.7. For the majority of residues there is an NOE greater than 0.75 at 50.7 MHz and greater than 0.70 at 36.5 MHz, which shows that any motions faster than the rotational correlation time are of a small magnitude. For groups of residues in the AB1, AB2 and CD loops and in the N- and C- termini there are, however, significantly smaller NOE's. The values of the order parameters, S^2, derived from an analysis of the relaxation data are also significantly lower for these residues than for those in the regions of secondary structure. These results indicate that there are internal motions occurring for these residues in the AB1, AB2 and CD loops and the N- and C-termini, which are fast compared with the overall rotational correlation time of the molecule (τ_R = 7.6 ns at 35°C). For the BC loop, the S^2 values are only slightly lower than those observed for residues in the secondary structure of the protein, but for residues His 59, Glu 60, Lys 61, Thr 63 and Gly 67 within the loop, the T_1/T_2 ratios are significantly larger than those anticipated from the overall correlation time of the molecule. These values could arise from conformational fluctuations on the millisecond time scale which would cause a decrease in T_2. Therefore, although the S^2 values show that the BC loop is not undergoing fast conformational fluctuations like the AB and CD loops, the BC loop may still be undergoing motions on a slower time scale.

Figure 6.7.

The results of ^{15}N relaxation studies on human interleukin-4 (from Redfield et al., 1992). The secondary structure of the protein is shown schematically in (A). Plots of the ^{15}N NOE, T_1/T_2 ratio and S^2 values versus amino acid sequence are given in (B)-(D) respectively. (The data were analysed using the formalism of Lipari & Szabo (1982 a, b)).



Despite both the disorder in the calculated NMR structures and the evidence for fast fluctuations in the AB and CD loops from the ^{15}N relaxation studies, an investigation of all the NMR data available for these loops shows that their conformations are not random. Short and medium range (i, i+2), (i, i+3) and (i, i+4) NOE's are seen in all the loops but it is the situations in the AB2 and BC loops that are particularly interesting.

For the AB2 loop (31-40) CαH-NH(i, i+3) and CαH-NH(i, i+4) NOE's have been identified which are typical of a helix. For seven of the ten residues in this loop, however, the $^{3}J_{HN\alpha}$ values are in the range 6.5-8 Hz and so are not consistent with a helical structure. In addition, if we look at the α proton chemical shifts in this region, the chemical shift index (as defined by Wishart et al. (1992) for the identification of secondary structure - described in Chapter 2) for seven of the residues is not indicative of a helical structure (Table 6.4). A likely explanation of these findings is that the AB2 loop is populating a number of different conformations; the loop may be helical for part of the time but in a more extended structure at other times. This would lead to the averaging of the chemical shifts and coupling constants resulting in a set of NMR parameters that are not consistent with a single conformation.

Table 6.4 Coupling constants and chemical shift data for the AB2 loop in human IL-4 (residues 31-40).

Residue	$^3J_{HN\alpha}$ (Hz).	α proton chemical shift*	Chemical shift index#
Asp 31	6.8	4.68	0
Ile 32	5.4	4.40	0
Phe 33	7.4	4.42	-1
Ala 34	5.0	4.22	-1
Ala 35	6.9	4.42	0
Ser 36	6.6	4.36	-1
Lys 37	5.8	4.26	0
Asn 38	7.5	4.71	0
Thr 39	8.0	4.54	1
Thr 40	7.4	4.46	1

* Chemical shifts are in ppm relative to DSS at pH 5.3 and 35°C.

By the procedure of Wishart et al. (1992), a grouping of four or more '-1's' is needed to define a helix. Further details are given in Chapter 2.

In the BC loop the situation is slightly different. For residues 60-66 some (i, i+2), (i, i+3) and (i, i+4) NOE's are again observed but here the $^3J_{HN\alpha}$ values are less than 6 Hz for five of the residues and so are also consistent with a helix (Table 6.5). In addition, the chemical shift index of the α protons indicate a helical conformation for residues 62-66. The $^3J_{HN\alpha}$ for Cys 65, however, is 8.5 Hz so there cannot be a continuous regular helix in this region. There are also no slowly exchanging amide protons in this loop.

Table 6.5 Coupling constants and chemical shift data for part of the BC loop in human IL-4.

Residue	$^3J_{HN\alpha}$ (Hz).	α proton chemical shift*.	Chemical shift index.
Glu 60	< 5	3.95	-1
Lys 61		4.48	1
Asp 62	5.5	4.58	-1
Thr 63	< 5	3.84	-1
Arg 64	< 6	4.08	-1
Cys 65	8.5	4.39	-1
Leu 66	< 6	3.74	-1

* Chemical shifts are in ppm relative to DSS at pH 5.3 and 35°C.

Although all the short and medium range NOE information for these loops has been included in the structure calculations no helical turns are found in the structures in the relevant regions of either the AB2 or BC loops. ϕ restraints explicitly defining a helical conformation (-60 ± 30° rather than 60 ± 150° which were in some cases used) and possibly hydrogen bond restraints would be required to get a well defined helical conformation. Indeed, trial calculations including helical ϕ and ψ angle restraints for these regions (C. Redfield, unpublished results) show that helical turns can be introduced into both the AB2 and BC loops without disrupting the four helix bundle core. However, for the AB2 loop in particular, including these types of restraints to ensure that helical turns are formed in the full calculations would give a very incorrect view of the conformation of the protein; the relaxation studies have shown the the loop is mobile and hence these helical turns must be only one of the adopted conformations in solution.

Study of the loop conformations in the structures of Powers et al. (1992a) shows that two short helical turns occur in the loops in this structure. One runs from residues 22 to 26 in the AB1 loop and the other from residues 61 to 69 in the BC loop. The evidence for helical turns in the BC loops has been discussed above. There is, however, little evidence of the helical turn between residues 22 to 26 in the NMR spectra, apart from the ^{13}C chemical shift data on which Powers et al. base their identification (Garrett et al., 1992). In particular, for this region only one helical type short range NOE has been identified (25CαH-27NH weak intensity) and there are no slowly exchanging amide protons. In addition, the only two residues for which $^{3}J_{HN\alpha}$ measurements have been able to be made (Thr 22 and Thr 25), both have values greater than 6 Hz. Further work will therefore be needed to investigate the conformation in this AB1 loop.

6.2 Tertiary structure.

6.2.1 Helix packing.

As discussed in section 5.4.1 the four main helices in IL-4 pack together to form a left-handed four helix bundle with an up-up-down-down connectivity. The bundle in IL-4 is, however, quite irregular with one interhelix angle being greater than 40° (see Table 6.7). This contrasts with the majority of bundles characterized by Presnell & Cohen (1989); indeed they only include cases where the interhelix angles are less than 40° in their classification.

Table 6.7 The interhelix angles in the four helix bundle of human IL-4*

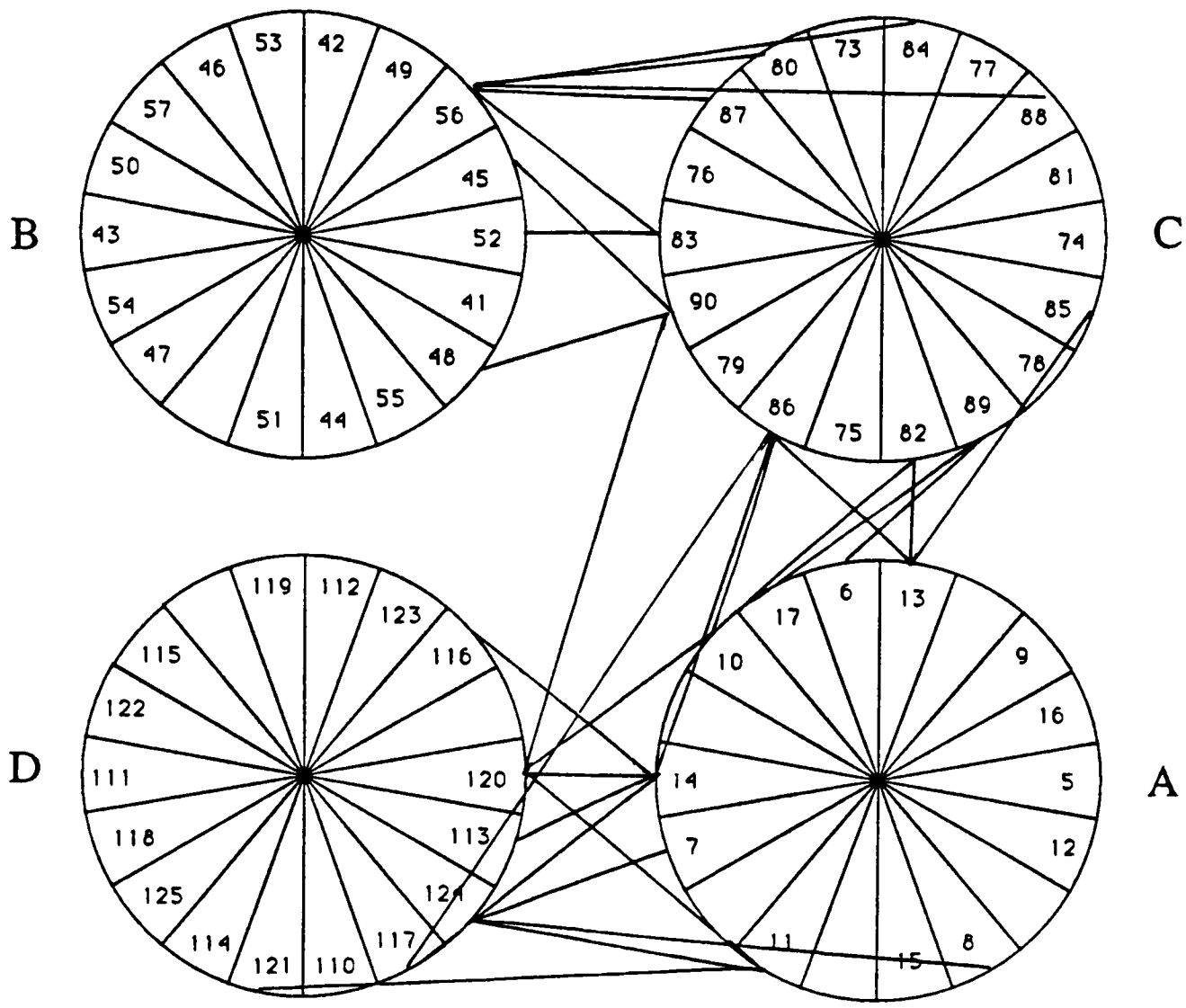
AD	38.3 ± 3.3°
DB	45.8± 3.1°
BC	30.8 ± 2.8°
CA	21.3 ± 4.7°

* Absolute values of the acute interhelix angles as defined by Presnell and Cohen (1989). In each case the mean and standard deviation of the angles in the set of 10 NMR structures are given.

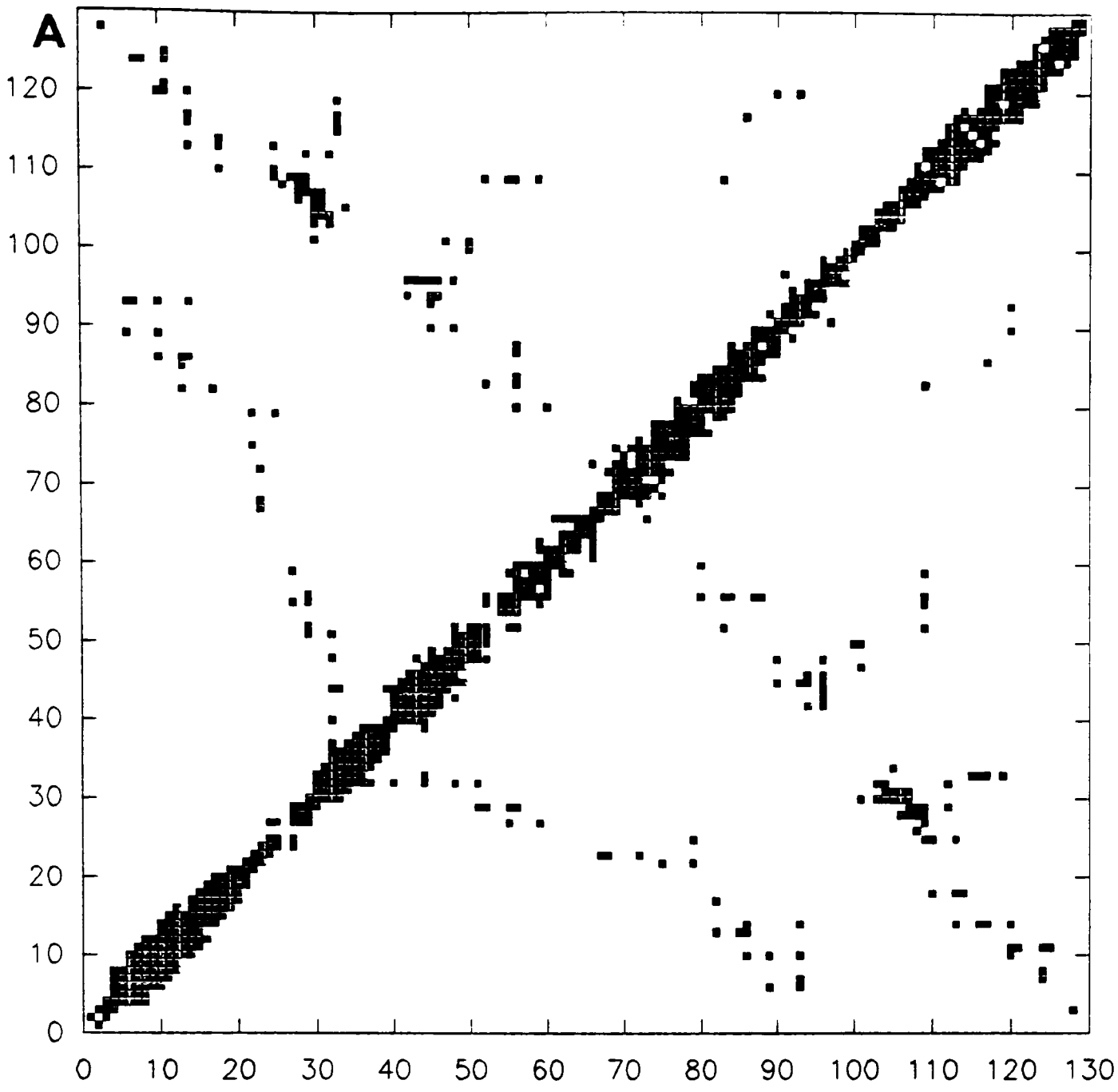
The interhelix NOE's, which have been identified in the spectra of IL-4 and which therefore determine the packing together of the helices in the structure calculations, are summarized in Figure 6.8. Although there are NOE's assigned between protons on the pairs of helices A and C, A and D, B and C, and C and D, there are no NOE's identified between protons on the A and B helices and no NOE's between protons on the B and D helices. (There are some NOE's from Leu 109 (just preceding D helix) to Phe 55 and Tyr 56 (helix B)).

To assess the confidence in the helix packing in the NMR structures of IL-4, the short interproton distances in the calculated structures have been analysed and have been compared with the observed NOE's. Using the average energy minimized NMR structure the set of interproton distances less than 5.0 Å have been calculated (N, α, β, aromatic and methyl group protons only). Using a simple model, it might be expected that all the corresponding NOE's will be present in the 3D NOESY-HMQC spectra of the protein. The predicted NOE diagonal plot is compared with the diagonal plot of the experimental NOE data set used in the structure calculations in Figure 6.9, and in Figure 6.10 the predicted interhelix NOE's are shown (compare

Figure 6.8.
Helix wheel diagram for IL-4. The lines indicate NOE's, identified in the spectra of IL-4, between residues in the four main helices.



Overleaf:
Figure 6.9.
Plot A shows the pairs of residues between which an NOE has been identified in the spectra of human IL-4 (sequence number on both axes). In B the pairs of residues between which an NOE is predicted to be observed are shown (prediction on basis of interproton distances < 5Å in the average energy minimized NMR structure).


A
120
110
100
90
80
70
60
50
40
30
20
10
0
0 10 20 30 40 50 60 70 80 90 100 110 120 130

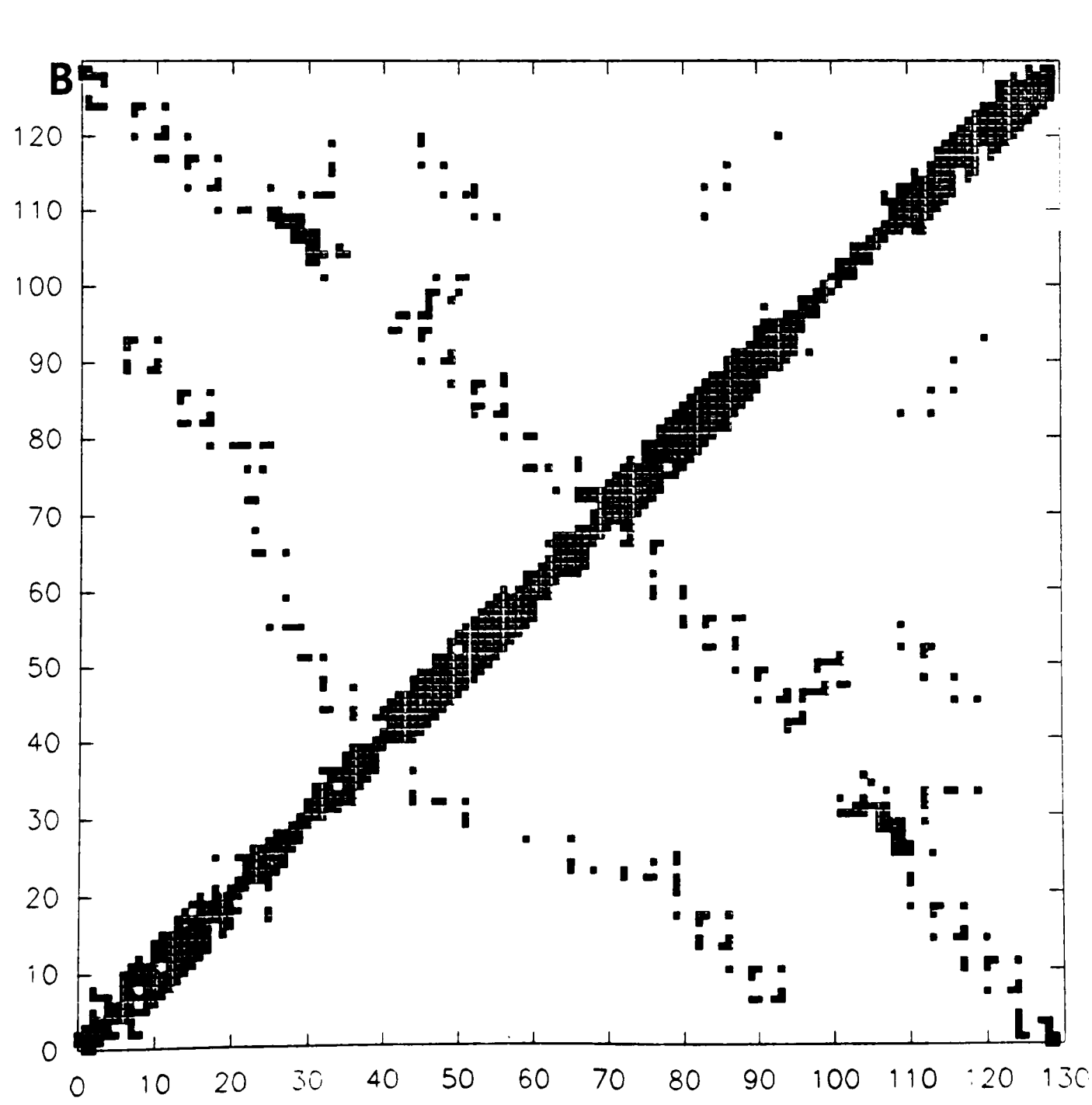
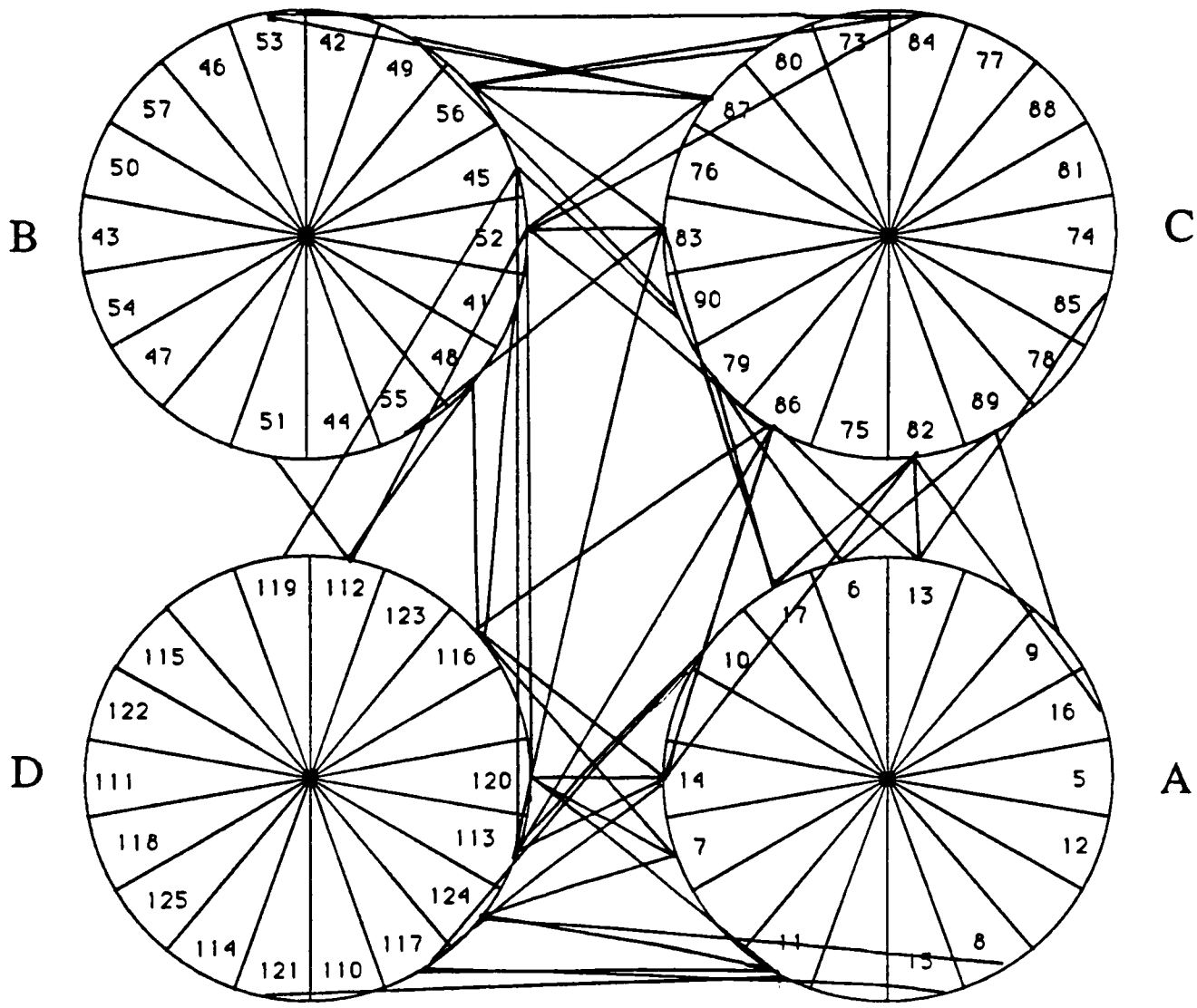

B
120
110
100
90
80
70
60
50
40
30
20
10
0
0 10 20 30 40 50 60 70 80 90 100 110 120 130

Figure 6.10.
Helix wheel diagram for IL-4. The lines indicate NOE's which are predicted to be observed between residues in the four main helices (predicted on basis of interproton distances < 5Å in the average energy minimized NMR structure). Compare with Figure 6.8 which shows the experimentally observed NOE's.



with Figure 6.8). The major obvious difference is that, although there are no B to D interhelix NOE's in the experimental data set, there are 32 interproton distances less than 5.0 Å in the average NMR structure (just considering N, α, β, methyl and aromatic protons). However all these distances, which are listed in Table 6.8, involve, with two exceptions, the aromatic protons of either Phe 45 or Phe 122. The NOE's arising from these residues have not been assigned unambiguously due to chemical shift degeneracy and consequent spectral overlap even in the 3D spectra. Further experiments recorded under different pH and temperature conditions may help with the analysis of these spectral regions. The two B to D helix contacts listed in Table 6.8 which do not involve Phe 45 or Phe 122 are both relatively long interproton distances (≥ 4.8 Å) so it is not unexpected that the corresponding NOE's have not been able to be identified in the spectra.

Although there are short interproton distances between helices B and D, there are no short distances in the average NMR structure between protons on the A helix and protons on the B helix; this is not very surprising, though, as the A and B helices are at opposite corners on the bundle rather than adjacent to each other.

Table 6.8.

Short interproton distances between the B and D helices in the average energy minimized structure (N, α, β, methyl and aromatic protons only).

Pair of protons	Distance (Å).
Phe 112 δ - Val 51 γ	4.9
Phe 112 δ - Leu 52 α	4.7
Phe 112 δ - Leu 52 β	4.9
Phe 112 δ - Leu 52 δ	4.1
Phe 112 ε - Ala 48 α	3.9
Phe 112 ε - Ala 48 β	4.3
Phe 112 ε - Val 51 β	3.9
Phe 112 ε - Val 51 γ	3.7
Phe 112 ε - Leu 52 α	3.1
Phe 112 ε - Leu 52 N	3.4
Phe 112 ε - Leu 52 β	2.6
Phe 112 ε - Leu 52 δ	3.4
Phe 112 ζ - Ala 48 α	2.8
Phe 112 ζ - Ala 48 β	3.8
Phe 112 ζ - Val 51 β	3.5
Phe 112 ζ - Val 51 γ	4.6
Phe 112 ζ - Leu 52 N	3.8
Phe 112 ζ - Leu 52 β	3.4
Leu 113 δ - Leu 52 δ	4.8
Leu 116 α - Phe 45 ε	4.8
Leu 116 α - Phe 45 ζ	4.4
Leu 116 β - Phe 45 ζ	4.9
Leu 116 δ - Phe 45 ε	4.2
Leu 116 δ - Phe 45 ζ	3.6
Leu 116 δ - Ala 48 β	4.9
Ile 119 β - Phe 45 ε	4.6
Ile 119 β - Phe 45 ζ	4.4
Ile 119 γ2 - Phe 45 ε	3.3
Ile 119 γ2 - Phe 45 ζ	3.5
Met 120 N - Phe 45 ζ	4.7
Met 120 ε1 - Phe 45 ε	3.3
Met 120 ε1 - Phe 45 ζ	3.3

It is interesting to compare the interhelix NOE's in the full experimental data set with those in the preliminary data set just from ^{15}N labelled spectra of the protein. There are far fewer interhelix NOE's overall in this preliminary set and, in particular, although there are 7 B-C helix NOE's and 14 A-D helix NOE's, there are only 2 A-C helix NOE's and no NOE's between any of the other pairs of helices (though there are some from Leu 109 to helix B). The result is the poor definition of the helix packing in the preliminary NMR structures; this is illustrated in Figure 6.11 which shows the average Cα deviation of the preliminary structures from the final energy minimized average structure. The average deviations are uniformly large (for 122 residues average deviation > 2 Å) reflecting the lack of NMR information about the relative orientation of the helices in the preliminary data set.

6.2.2 Loop topologies.

The up-up-down-down connectivity of the four helix bundle in IL-4 means that two of the loops that connect up the helices run the length of the molecule (AB and CD). The third loop, which connects up the B and C helices, is also relatively long (14 residues) and it extends away from the main part of the bundle at the top of the molecule as viewed in Figure 6.12. As discussed in section 6.1.3, all the three loops contain regions of disorder in the ensemble of calculated structures, some of which may arise from mobility of the regions in solution. The disorder is however local; the loops in all the NMR structures trace the same path relative to the helices. A comparison with the NMR data though, shows that there is little long range NOE information about the disordered regions of the loops (Figure 6.13). In particular, no longer range NOE's have been observed for residues 19-21, 35-40, 58, 61-65, 70-71 and 97-99 in the loops. To investigate the situation in

Figure 6.11.
Average Cα deviation per residue (in Å) of the ensemble of 10 preliminary calculated NMR structure from the final average energy minimized structure (preliminary structure calculated with NOE data from spectra of unlabelled and ^{15}N labelled IL-4 and with ϕ and ψ torsion angle restraints for helical residues).

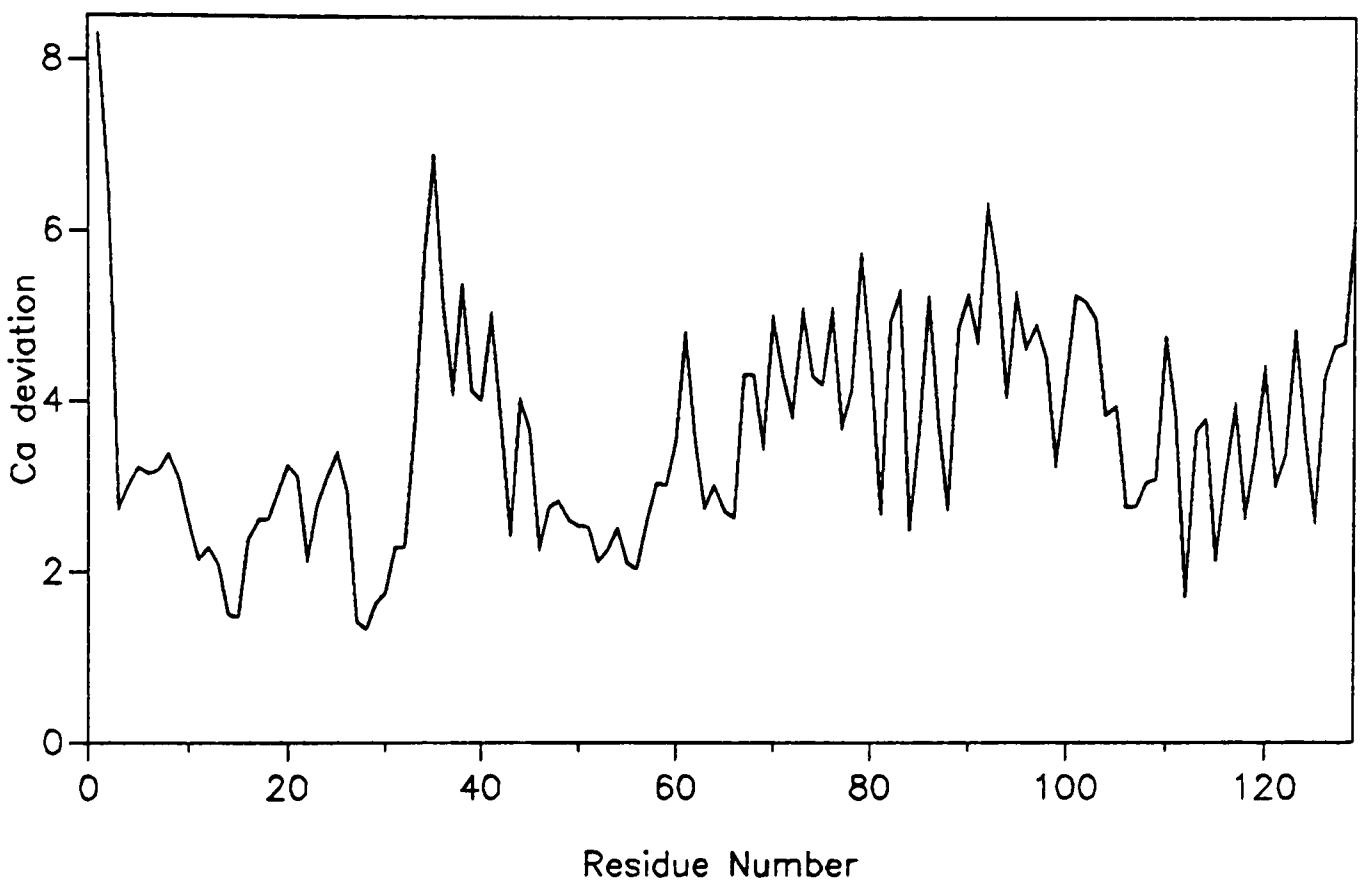
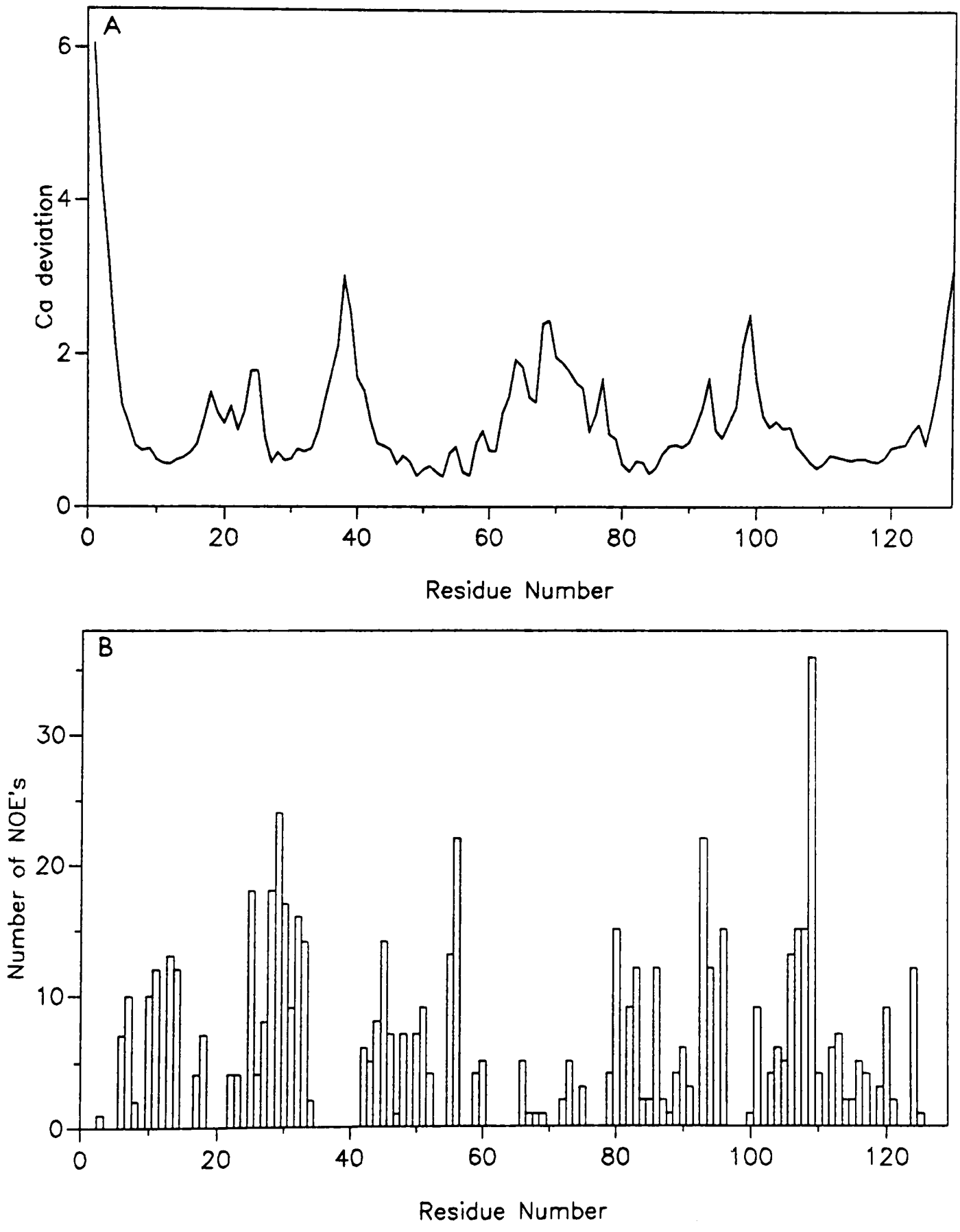


Figure 6.12.
Ribbon diagram of the average energy minimized NMR structure of human IL-4 showing the BC loop (58-71) in bold.

Figure 6.13.

Graph A shows the average Cα deviation per residue (in Å) of the ensemble of 10 NMR structure from the average energy minimized structure. Graph B shows the number of identified long range NOE's (between residues that are five or more apart in the sequence) per residue in the final NMR data set (compare with preliminary data set in Figure 5.4). Each NOE is counted twice (once for each contributing nuclei).



more detail, the NOE data have been compared with the short interproton distances in the average energy minimized structure (see Figure 6.9). All the distances less than 5.0 Å have been calculated (N, α, β, aromatic and methyl group protons only) and, as a simple model, it is assumed that all the corresponding NOE's may be present in the 3D NOESY-HMQC spectra of the protein (see section 6.2.1).

The path of the AB loop is shown clearly in Figure 6.14. The AB loop prior to the β sheet (AB1) interacts with the C and D helices; NOE's have been observed which define contacts of the loop to both of these helices e.g. 18CγH-114CαH; 22CγH-75CδH; 25CγH-113CδH, 79CδH. Two NOE's are also observed from the AB1 loop to the BC loop around the disulphide bridge (24-65). There are three residues in the AB1 loop however (Glu 19, Gln 20 and Lys 21) for which no longer range NOE's have been identified. Study of the average NMR structure shows that for Glu 19 and Gln 20 no long range NOE's are expected. For Lys 21, the distance 21CβH-110CαH is 4.6 Å, so a very weak NOE could be observed but as it could be lost in the noise, it is not surprising that it has not been identified. At the centre of the AB loop the β sheet region occurs, linking the AB loop to the end of the CD loop. 53 NOE's have been observed between these two loops in the β sheet and surrounding residues. The second part of the loop (AB2) has, however, fewer long range close contacts, the only important regions of contact being to the B helix. This is shown in Figure 6.15 and is defined by 15 NOE's. No long range NOE's have been observed though for residues 35-41 at the end of this loop, and the only few that are predicted are extra contacts to the B helix which are in general long and so would give rise to only weak NOE's e.g. 37CαH-44CαH 4.87 Å; 36CβH-47CβH 4.70 Å. Interestingly, the interaction of the AB2 loop with the B helix appears to lead to well defined χ_1 conformations in the calculated structures for many of the residues involved (section 6.3.2).

Figure 6.14.
Ribbon diagram of the average energy minimized NMR structure of human IL-4 showing the path of the AB loop (18-27 and 31-40) in bold.

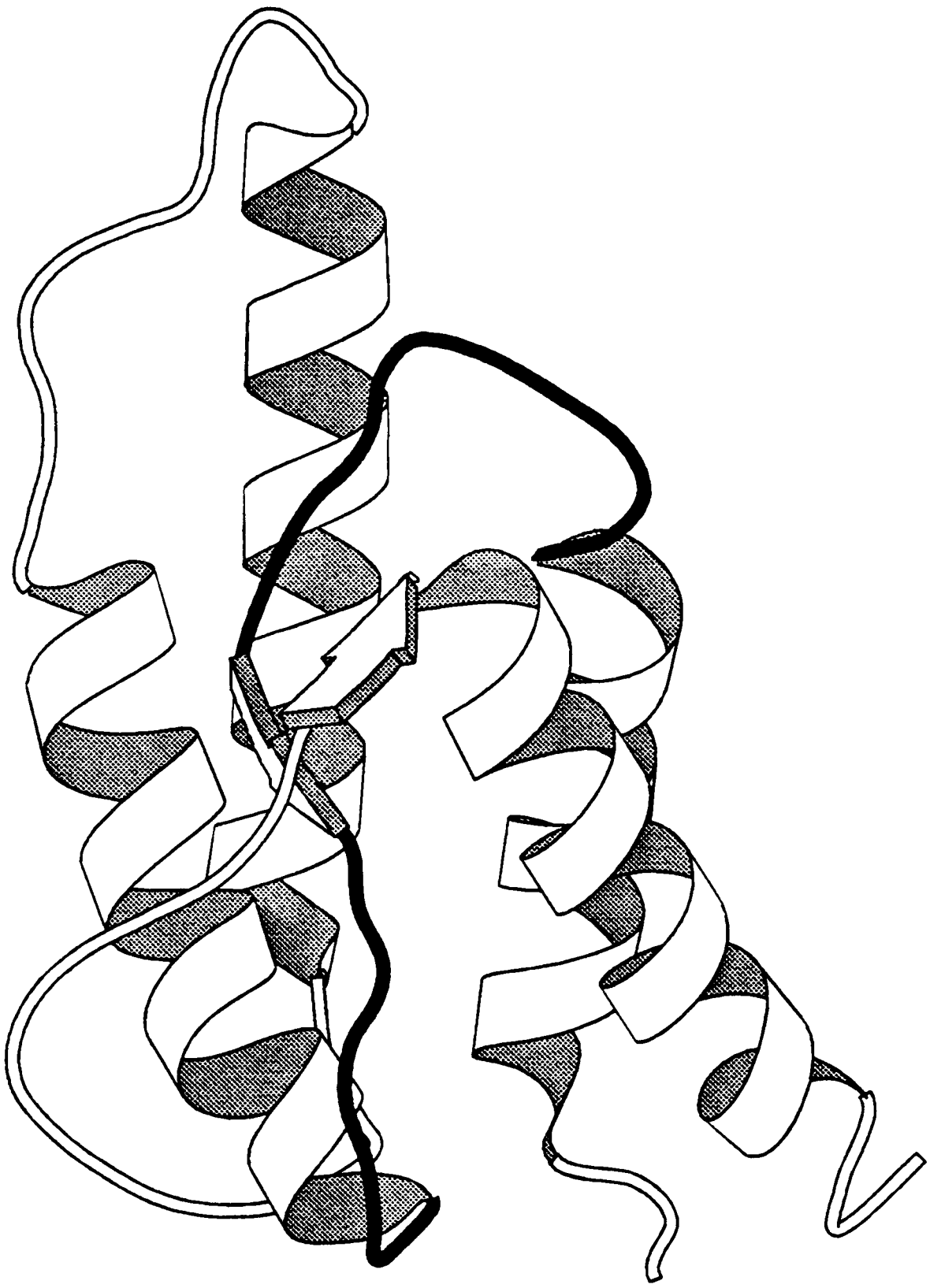
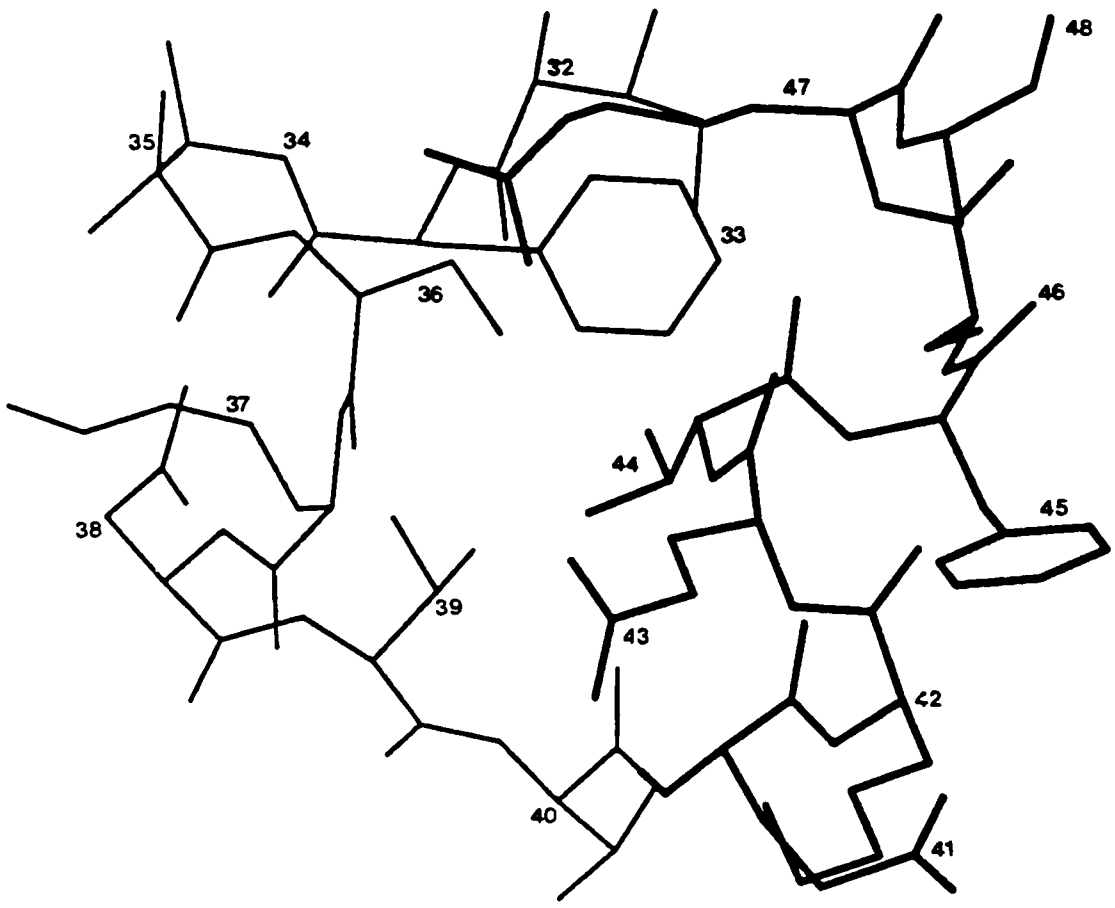


Figure 6.15
Contacts between the AB2 loop and initial part of the B helix in the average energy minimized NMR structure of IL-4.



Residues Asp 31, Ile 32 and Phe 33 also have slowly exchanging amide protons though the reason for this is not apparent in the present structures.

Study of the BC loop in the NMR structures shows that there are very few close contacts to other parts of the molecule for this loop explaining, at least in part, the lack of long range NOE data for 8 residues in the loop. Indeed for 6 of these in the average structure there are no protons within 5 Å, except for those within the loop itself (or adjacent to it for residues at the end of the loop). The initial part of the loop, however, is close to the C helix (e.g. NOE's observed: 60CαH-80CγH; 66CδH-73CαH) but the only other close contacts are to the AB1 loop in the region of the 24-65 disulphide.

The topology of the CD loop relative to the rest of the bundle is shown in Figure 6.16. The initial part of the loop is close to the start of the A helix (NOE's between Thr 6 and Asn 89, Thr 6 and Leu 93, Leu 7 and Leu 93) and the end of the D helix (distance of 4.1 Å between Leu 93 CδH and Met 120 CεH although no NOE's have been observed defining this contact). This is interesting in the light of the disulphide bridge in murine IL-4 between residues 3 and 92 (discussed in section 7.2). Residues Gly 92 and Leu 93 also have slowly exchanging amide protons although they do not seem to be involved in any hydrogen bonds in the present structure. The main part of the loop, however, wraps around the B helix and close contacts are observed to this (NOE's shown in Figure 6.17). No long range NOE's have been observed, however, for residues 97-99, which is initially surprising as the 46-99 disulphide also links the CD loop to this helix. The distances to the B helix for the protons of these residues are, however, long (e.g. 98CαH-46CαH 4.89 Å; 99NH-47CαH 4.83 Å) so any NOE's would be weak. The final part of the CD loop is linked via the β sheet to the centre of the AB loop as discussed above.

Thus, in general, most of the predicted NOE's have been observed experimentally for these loop regions. However, the predicted NOE's were

Figure 6.16.
Ribbon diagram of the average energy minimized NMR structure of human IL-4 showing the path of the CD loop (91-105, 109) in bold.

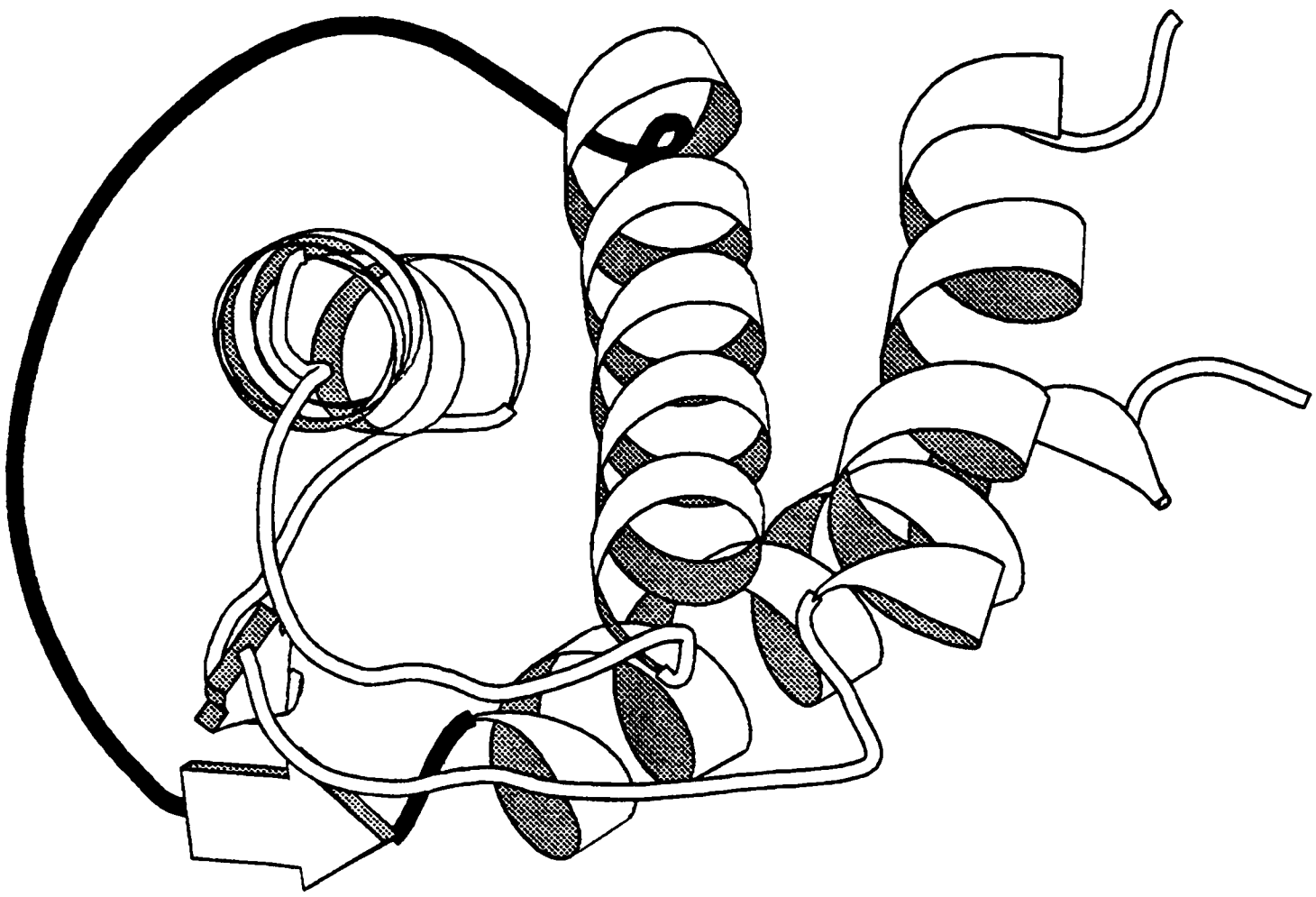
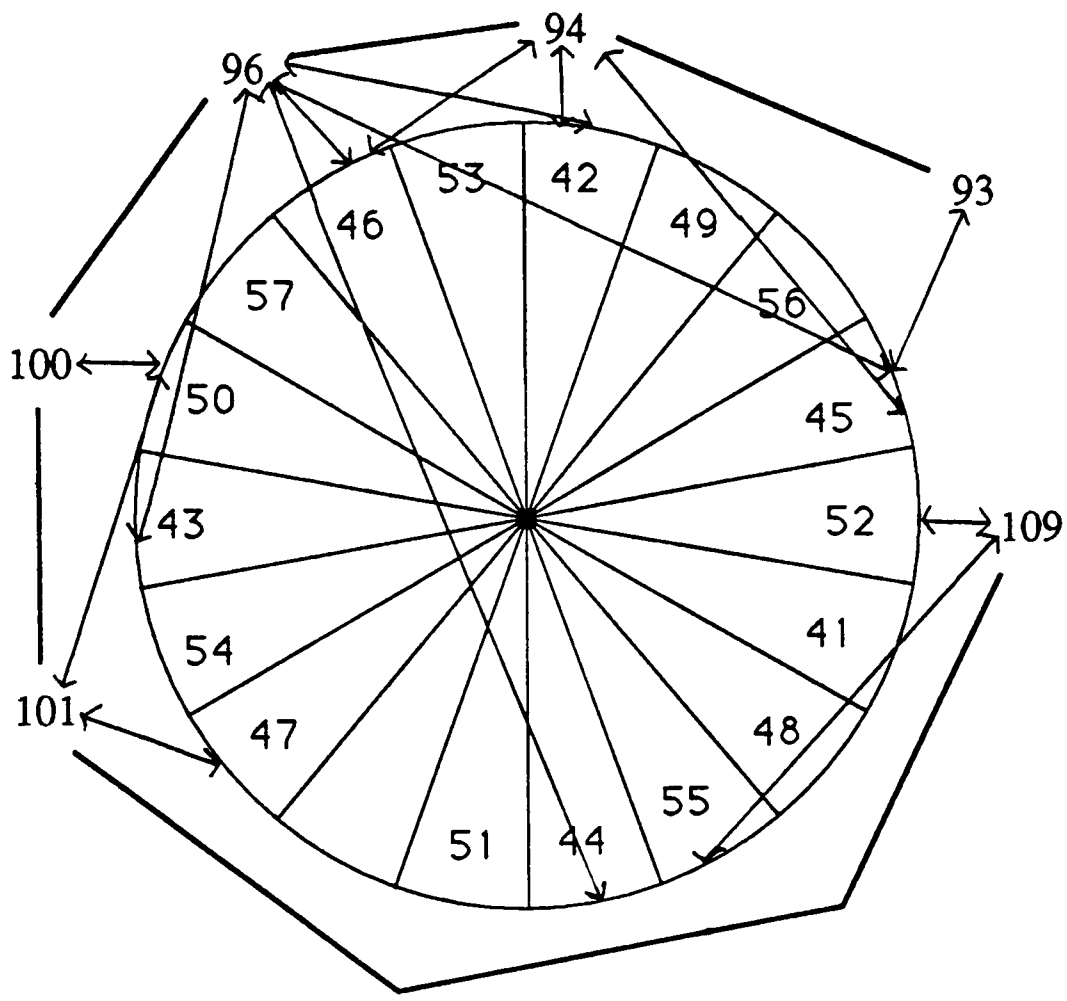


Figure 6.17.
Schematic diagram showing the B helix and residues in the CD loop of IL-4. Arrows indicate the residues between which an NOE has been observed.



calculated from a structure where the conformations of the loops may be biased towards close contacts for residues where NOE's have been observed. As the ^{15}N relaxation studies have shown, much of the loop regions are undergoing motions on a time scale faster than the rotational correlation time; it is likely, therefore, that some of the observed NOE's for the loop regions have been averaged by the population of multiple conformations. In addition, some NOE's, which would be predicted from the true structure of IL-4, may have been quenched and so would not appear in the spectra. Further work is therefore necessary to define some of the details of the conformation of these loops in solution.

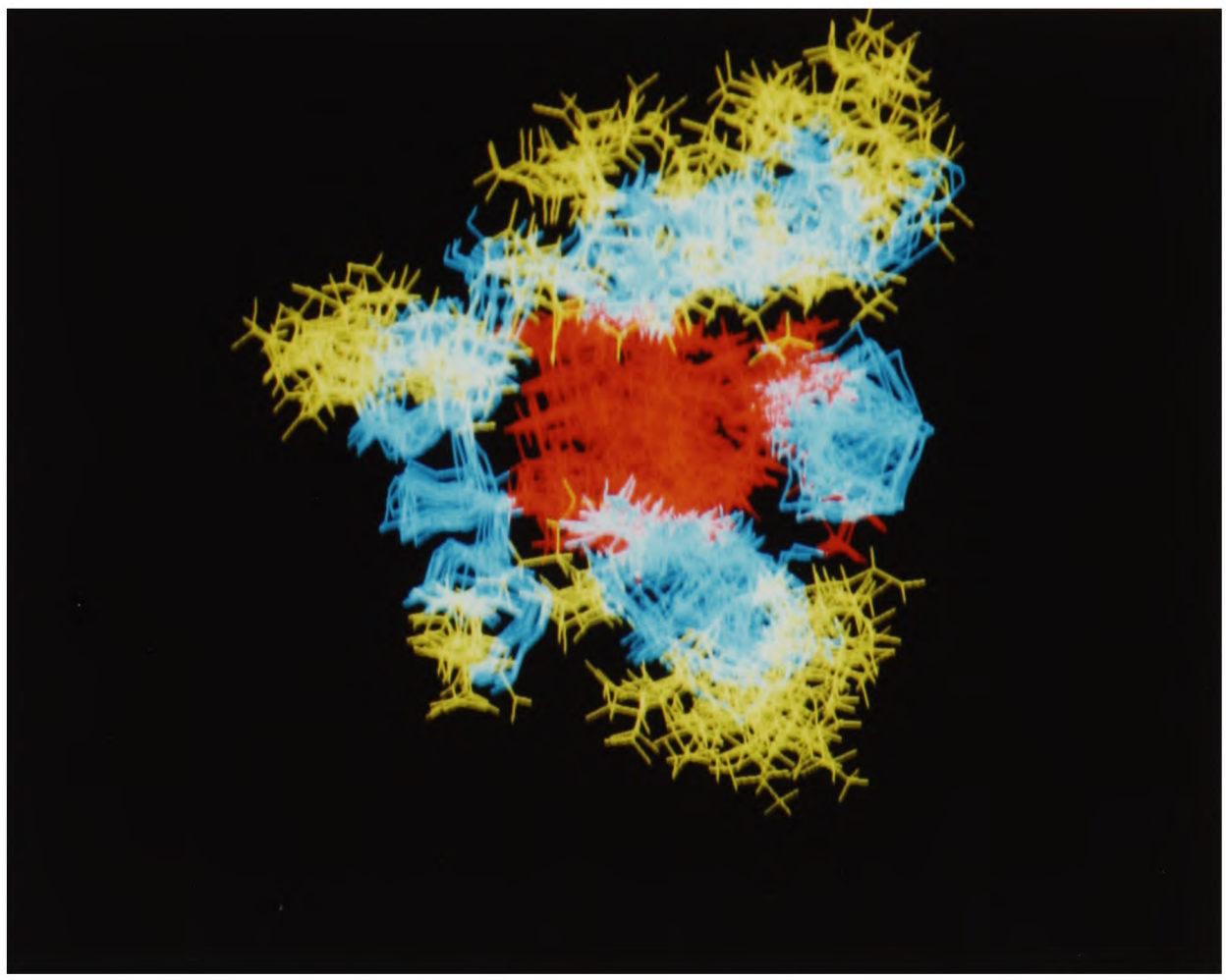
6.3 Side chains in IL-4.

6.3.1 Residue distribution in the structure.

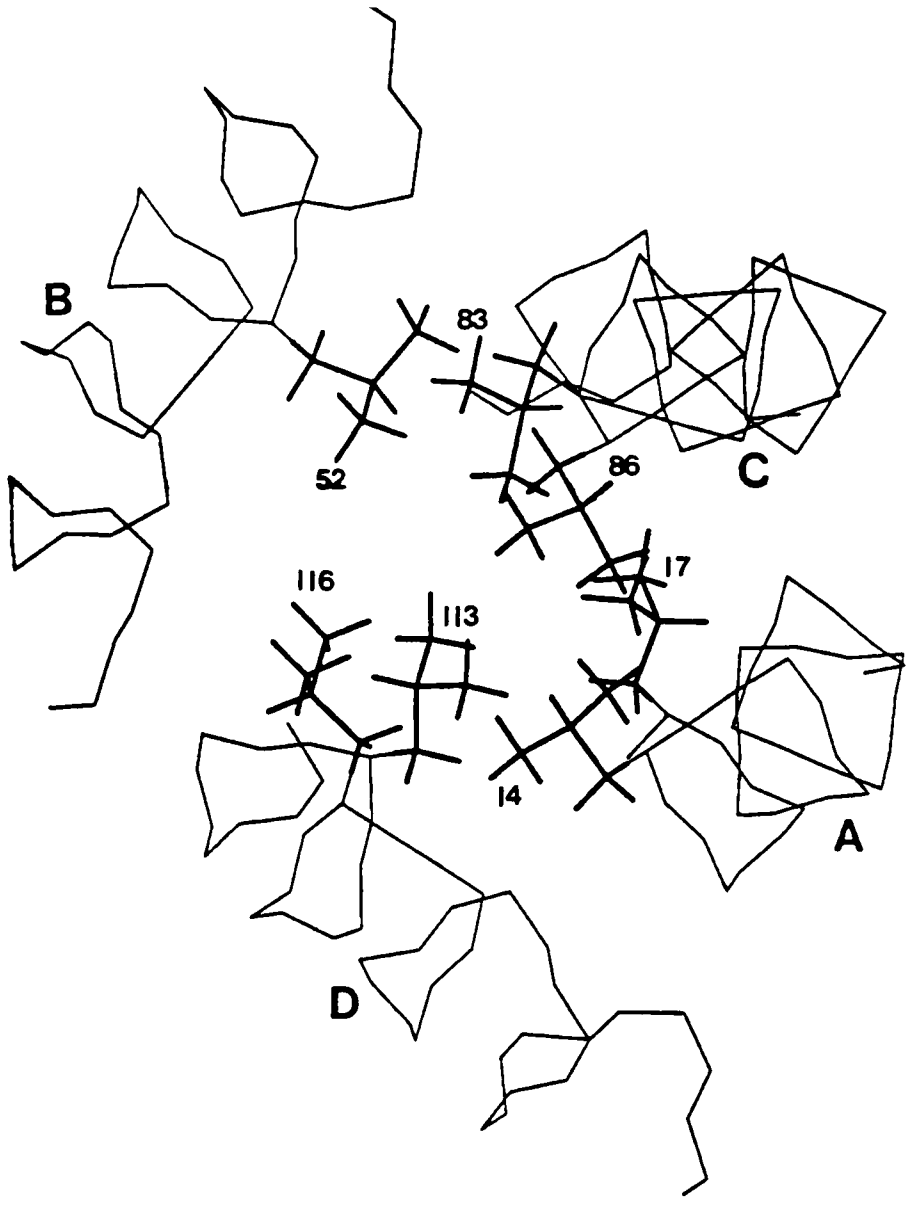
The sequence of human IL-4 contains an unusually large number of leucine residues (16); for a random 130 residue sequence an average of 10 leucines would be expected (Schultz & Schirmer, 1979). Figure 6.18 compares the position of the leucines in or adjacent to the helices with those of charged arginine, lysine, aspartic acid and glutamic acid residues in the NMR structures. The leucines cluster in the hydrophobic core of the molecule stabilizing the structure, while the charged side chains extend out into the surrounding medium.

A closer analysis of the leucines shows that seven of them appear to be forming particularly important key interhelix contacts; the methyl groups of these leucines are close (within 5 Å in the average structure) to the methyl groups of leucine residues on two or three of the other helices (see Figure 6.19). A comparison of the conformation of the side chains of these 7 leucine residues in the two average structures (calculated using the different protocols) is shown in Figure 6.20. Their conformations are closely similar

Figure 6.18.
Superposition of the four α helical regions in the 10 calculated NMR structures of IL-4. The main chain N, Cα and C atoms are shown in blue. The side chains of the leucine residues are shown (in red) clustering in the hydrophobic core between the helices, while side chains of arginine, lysine, aspartic acid and glutamic acid residues are shown (in yellow) extending out into the surrounding medium.



Overleaf:
Figure 6.19.
The main chain trace of the four α helices of IL-4 is shown, together with the side chains of the seven leucines which appear to form particularly important interhelix contacts. The pairs of these leucines between which there is an interproton distance less than 5Å in the average energy minimized NMR structure are indicated by lines in the helix wheel diagram.


B
83
52
C
86
17
116
113
14
A
D

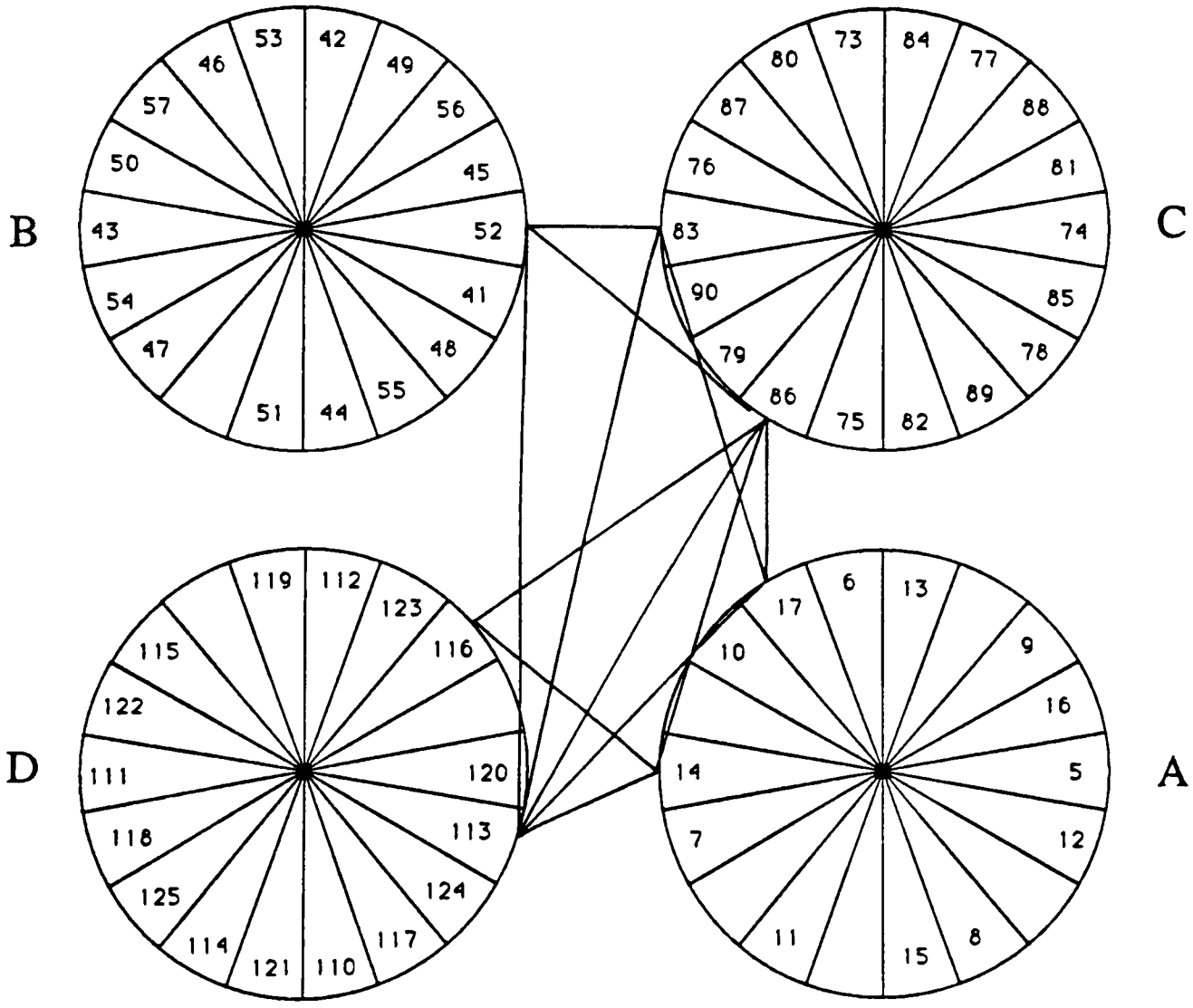
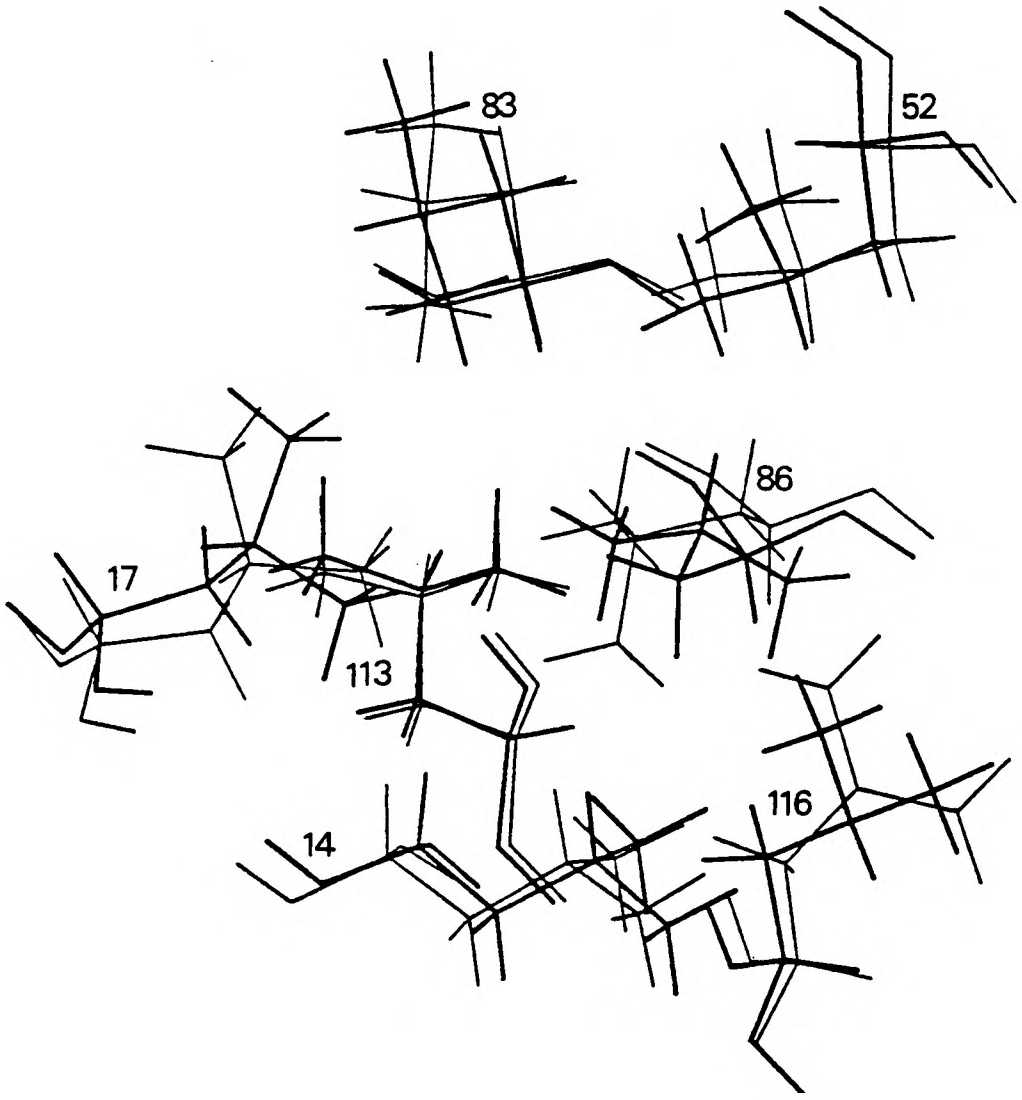
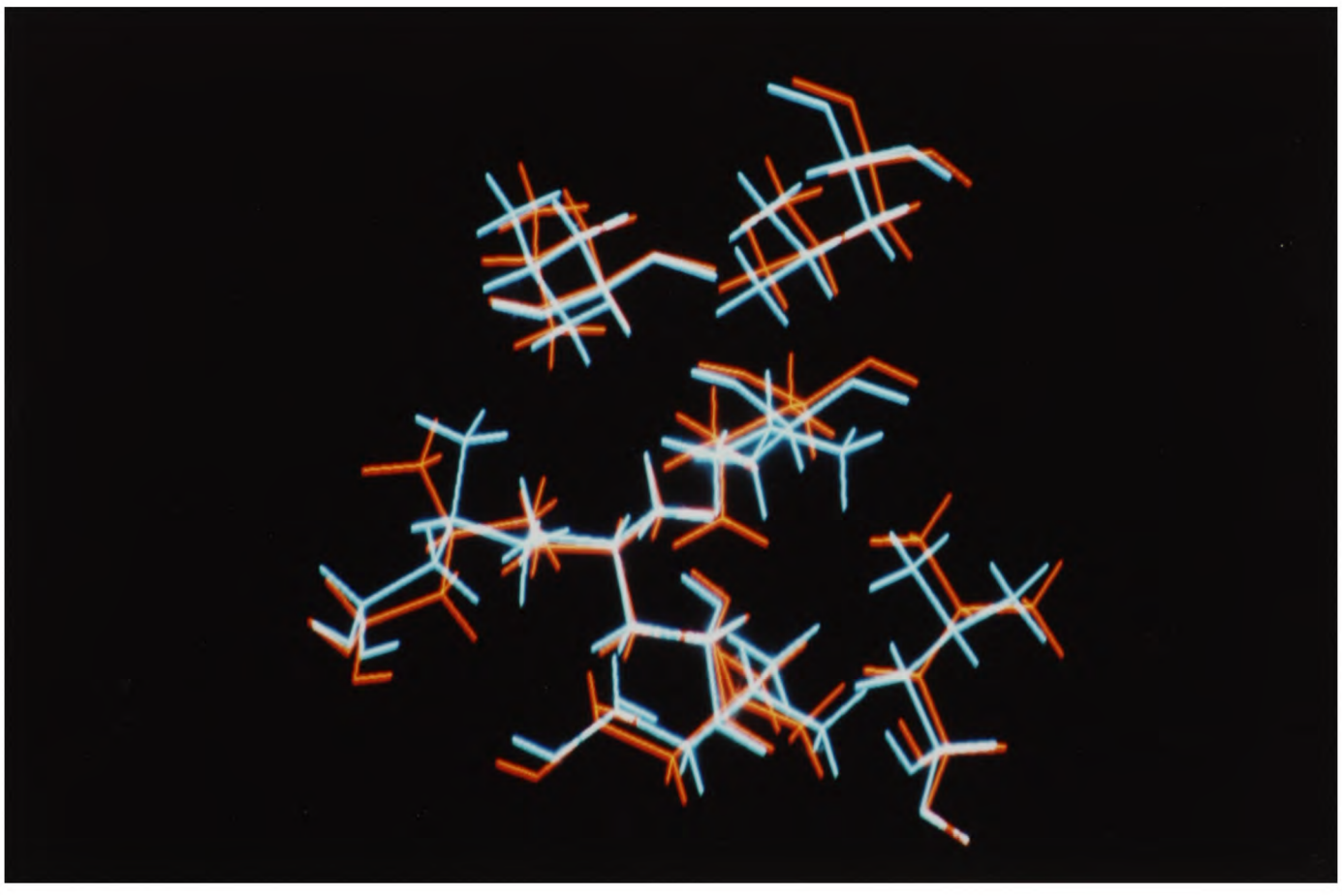

B
53
42
49
46
56
57
45
50
52
43
41
54
48
47
55
51
44
C
73
84
80
77
87
88
81
76
74
83
85
90
78
79
89
86
75
82
D
119
112
123
115
116
122
120
111
113
118
124
125
117
114
121
110
A
6
13
17
9
10
16
5
14
12
7
11
8
15

Figure 6.20.
Comparison of the conformation of the seven leucines shown in Figure 6.19 (14, 17, 52, 83, 86, 113, 116) in the average energy minimized structures calculated using the distance geometry-dynamical simulated annealing protocol (in red in the colour print) and simulated annealing protocol (in cyan in the colour print and in bold in the labelled plot).





reflecting the good definition of the internal hydrophobic core of the protein in the structure calculations.

6.3.2 Definition of side chain conformations.

No $^3J_{\alpha\beta}$ measurements have been made for human IL-4 and consequently no χ_1 restraints were included in the present structure calculations. In addition, no stereospecific assignments have, as yet, been made for β-methylene protons or for the methyl groups of valine and leucine residues. Despite this, an analysis of the side chains in the calculated structures shows that the conformation of many of the internal residues is well defined. For 37 of the residues the same χ_1 conformation is adopted in all the structures calculated using the distance geometry-dynamical simulated annealing protocol; for 13 of these residues the χ_1 range is less than 20° while for the other 24 the range is less than 55° in the ensemble of structures. Study of the NMR data set shows that there are many NOE's assigned involving these residues. Indeed, for 16 of the 37 residues considered here there are 10 or more assigned long range NOE's. The positions of these 37 residues have been analysed (Table 6.9):

20 of the residues are in helices and are involved in close interhelix contacts (see helix wheels in Figure 6.10).

5 of the residues are involved in the short region of β sheet.

5 of the residues are helical but are on an external face of the bundle interacting with a loop region.

4 of the residues are in loops and are involved in contacts to the helices.

Only 3 of the residues appear to have no close contacts with other regions of the structure (residues Thr 40, Ser 57 and Thr 118).

The interaction discussed in section 6.2.2, between residues 30-33 in the AB loop and various residues including Glu 43, Arg 47, Thr 50 and Val 51 on a

face of the B helix appears to lead to all these residues having well defined χ_1 conformations (see Figure 6.15).

Table 6.9.

Residues with the same χ_1 conformation in the entire ensemble of NMR structure calculated using the distance geometry-dynamical simulated annealing protocol.

Residue	χ_1	Position
10 *	-80	Helix A. Involved in interhelix contacts
11 *	-80	Helix A. Involved in interhelix contacts
13 *	-60	Helix A. Involved in interhelix contacts
14 *	-60	Helix A. Involved in interhelix contacts
15	180	Helix A. Involved in interhelix contacts
17 *	-60	Helix A. Involved in interhelix contacts
25 *	0	Contacts to helices
28 *	60	β sheet
30 *	-60	β sheet
31 *	180	Contacts to B helix
32 *	60	Contacts to B helix
33 *	-60	Contacts to B helix
40 *	60	No long range contacts
42	60	Helix B. Involved in interhelix contacts
43 *	-60	Contacts to 30-33
44 *	-60	Helix B. Involved in interhelix contacts
45 *	-100	Helix B. Involved in interhelix contacts
46	180	Helix B. Contacts to CD loop.
47 *	-60	Helix B. Involved in contacts to 30-33
50 *	-60	Helix B. Involved in contacts to 30-33
51 *	180	Helix B. Involved in contacts to 30-33
52 *	180	Helix B. Involved in interhelix contacts
55 *	180	Helix B. Involved in interhelix contacts
56 *	60	Helix B. Involved in interhelix contacts
57	180	Helix B. No long range contacts
59	-60	Contacts to AB loop and helix C
76	180	Helix C. Involved in interhelix contacts
79	180	Helix C. Involved in interhelix contacts

83 *	-60	Helix C. Involved in interhelix contacts
106 *	-100	β sheet.
108 *	60	β sheet.
112 *	180	Helix D. Involved in interhelix contacts
117 *	180	Helix D. Involved in interhelix contacts
118 *	-60	No long range contacts.
119 *	-80	Helix D. Involved in interhelix contacts
120	-60	Helix D. Involved in interhelix contacts
124 *	180	Helix D. Involved in interhelix contacts

* indicates residues where the same χ_1 conformation is found in all the structures calculated using the simulated annealing protocol.

For most of the external residues the χ_1 conformations are much more disordered and, in general, as there is little NOE information for surface groups, the conformations observed are those favoured by the intramolecular potential (χ_1 = 180°, 60°, -60°). Indeed, for 88% of all the residues in IL-4 the χ_1 values in individual structures cluster around these three staggered rotamer positions. Many of the external residues are, however, expected to be disordered in solution (compare with lysozyme - Chapter 3) and $^3J_{\alpha\beta}$ measurements, when performed for IL-4, should be able to confirm this.

Some residues can be identified, though, which are on the interior face of a helix that is involved in interhelix contacts and which have a number of assigned NOE's, but which also have a disordered χ_1 conformation in the ensemble of calculated structures. For example, Phe 82 and Leu 86 in the C helix have disordered χ_1 (for Phe 82 χ_1 distributed between approximately -60° and 120°; for Leu 86 χ_1 distributed between approximately -60°, 120° and -120° in the ensemble of structures) although they have 9 and 12 assigned longer range NOE's respectively (the NOE's for Phe 82 are to helix A while

those for Leu 86 are to helices A and D). Although it is possible that these internal residues may populate multiple conformations, as has been observed for Val 99 in hen lysozyme (Chapter 3), it is more likely that their disorder arises from insufficient NMR data and, in particular, from the lack of χ_1 restraints in the structure calculations.

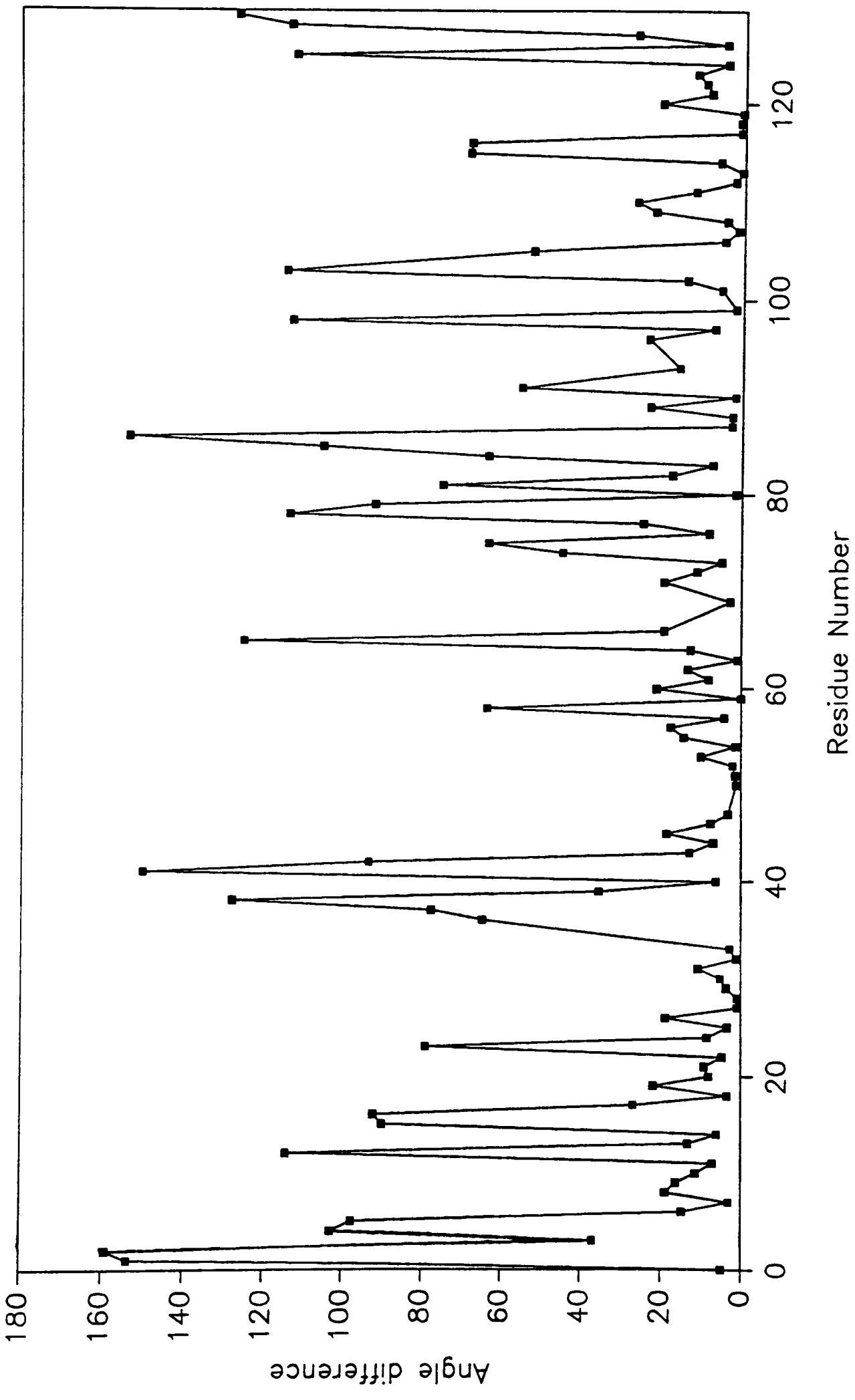
If the χ_1 conformations are compared with those in the ensemble of structures calculated using the simulated annealing protocol, it is found that for 29 of the 37 residues with a single χ_1 conformation in the distance geometry-dynamical simulated annealing protocol, the same χ_1 conformation is also adopted in the simulated annealing structures (these residues are indicated in Table 6.9). For another 3 of the residues the same χ_1 conformation is adopted in all but one of the simulated annealing structures, while for the remaining 5 residues the χ_1 values cluster in two different χ_1 ranges in the simulated annealing structures. This suggests that, in general, the conformations observed for the residues with a single χ_1 value in the ensemble of structures are likely to be correct rather than an artifact from the structure calculation procedure.

The close similarity of the side chain conformations of many internal residues in the two sets of structure calculations is also shown when the χ_1 values in the two average energy minimized structures (calculated using the different protocols) are compared (see Figure 6.21). Although there are some large differences, for 74 residues the χ_1 difference is less than 20°; 41 of these residues are in the helices.

As the sequence of human IL-4 contains a large number of leucine residues it is interesting to see how well defined they are about χ_2 (i.e. how well oriented the methyl groups are). An analysis of the distance geometry-dynamical simulated annealing structures shows that there is a much greater range of angles in the calculated structures than there is around χ_1 but in general the χ_2 angles cluster around the three staggered rotamer positions.

Figure 6.21.

Graph showing the difference in χ_1 in the two average energy minimized NMR structures of IL-4 calculated using the distance geometry-dynamical simulated annealing and simulated annealing protocols.



For example, for Leu 116 in three structures $\chi_2 \approx 180^\circ$, in two structures $\chi_2 \approx -60^\circ$, in four structures $\chi_2 \approx 60^\circ$ and in one structure $\chi_2 = -199^\circ$, i.e. not one of the staggered values. However, for two of the leucine residues (14 and 83) there is the same χ_2 conformation in 9 of the 10 structures. χ_1 is also well defined for these residues and they are two of the leucines that have been recognised as forming key interhelix interactions (see Figure 6.19). In general though, the definition of the leucine side chains in the NMR structures would be greatly improved if the $C\delta H_3$ groups could be stereospecifically assigned. Experiments on a sample of IL-4 with partially ^{13}C labelled leucines could enable these assignments to be made (Senn et al., 1989).

6.4 Summary.

In this chapter the structure of human interleukin-4 determined by NMR techniques has been presented. The protein has been shown to have a well defined left-handed four helix bundle core but with large angles between the helices, in contrast to the majority of four helix bundle proteins. The bundle has an up-up-down-down connectivity with two long loops (AB and CD) running the length of the molecule which are connected by a short region of irregular antiparallel β sheet. The helical core of the molecule is stabilized, in particular, by leucine-leucine interactions in a hydrophobic core between the helices. Many of the residues in this core have well defined conformations, at least around χ_1, in the NMR structures.

A particularly interesting feature of the structure of IL-4, revealed by ^{15}N relaxation studies, is that some regions of the long loops in the protein are mobile and are undergoing substantial conformational fluctuations on a time scale faster than the overall rotational correlation time of the molecule. The NMR structural data for IL-4 show, however, that the conformations of these loops are not completely random. In particular the AB2 loop at least, appears to adopt a helical conformation for a proportion of the time.

In the next chapter this structure for IL-4 is compared with a structure prediction and with other proteins that have related folds, and its significance is discussed.

CHAPTER 7.

THE STRUCTURE OF HUMAN INTERLEUKIN-4.

Significance of the IL-4 structure and comparison with related structures.

A detailed analysis of the solution structure of human interleukin-4 has been presented in Chapter 6. In this chapter the structure of IL-4 is set in context and it is compared with a structure prediction for the protein and with other proteins that have been found to have related folds. The binding of IL-4 to its receptor is also considered and the consequences for haemopoietic cytokines in general are discussed.

7.1 Comparison with a structure prediction for IL-4.

Curtis et al. (1991) have published a structure prediction for human IL-4. Using secondary structure prediction techniques (Chou & Fasman, 1978; Garnier et al., 1978; Cohen et al., 1986), five helical regions were proposed for the human IL-4 sequence (see Table 7.1). No compact tertiary structure consistent with these helices could be found, however,and comparison with the murine sequence, where helix B is not predicted, suggested that a four helix bundle (consisting of helices A, C, D and E) was most likely. Comparison of the predicted helices with those determined experimentally shows that, apart from differences in the precise start and end points of the helices, three of the predicted regions (A, C and E) agree very well with the observed A, B and D helices. Helix D in the prediction (residues 77-86) is, however, much shorter than the experimental helix C (72-90).

Table 7.1

Comparison of the helical regions in the structure prediction with those in the NMR structure of human IL-4.

Structure prediction		NMR structure	
Helix	Residue	Helix	Residue
A	4-17	A	5-17
B	23-35		
C	42-58	B	41-57
D	77-86	C	72-90
E	108-122	D	110-125

Using a tertiary structure algorithm (Curtis et al., 1991) and employing accessible surface contact area as a criterion, the most suitable overall structure for the four helices (A, C, D, E) was found to be a right-handed four helix bundle with two overhand connections as shown in Figure 7.1. The proposed structure therefore has the same connectivity as the experimental NMR structure of human IL-4 but the opposite handedness. In addition, the helices are more parallel than in IL-4 and the topology of the loops is significantly different; the AB loop in the structure prediction passing around the outside of the D helix while in the NMR structure helix D is on the outside of this AB loop.

It is interesting to look at the leucine residues which make the key interhelix contact in IL-4 to see where these residues are situated in the predicted structure. Figure 7.2 and 7.3 show the positions of the side chains of hydrophobic leucine residues and charged arginine, lysine, aspartate and glutamate residues in the structure prediction and in the average NMR structure of IL-4. Although the charged residues are in external exposed

Figure 7.1.
Ribbon diagrams showing the right-handed four helix bundle fold predicted for human IL-4 by Curtis et al. (1991) (A) and the true left-handed four helix bundle fold of the protein determined by NMR techniques (B).

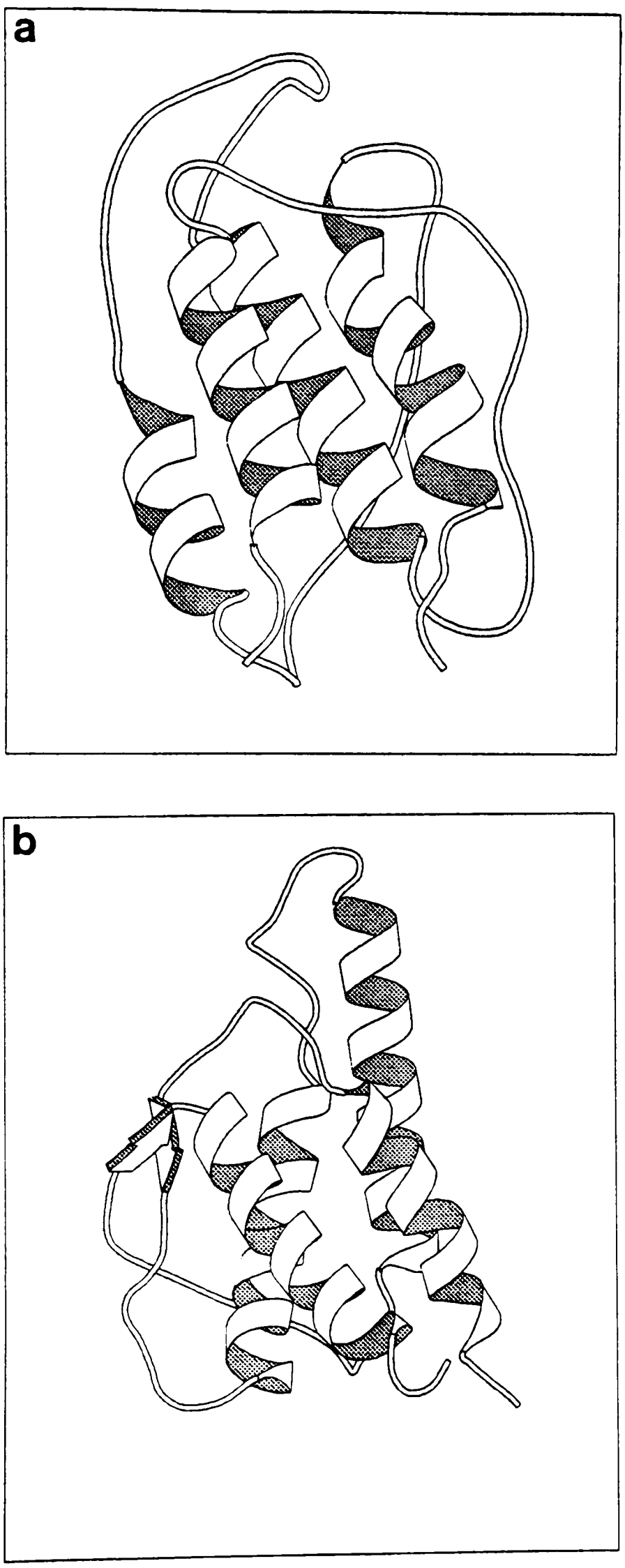


Figure 7.2.

The structure prediction for human IL-4 (Curtis et al. (1991). The main chain of the helical regions is shown in cyan. In the upper diagram the side chains of leucine residues are shown in red and the side chains of charged arginine, lysine, aspartic acid and glutamic acid residues are shown in yellow. In the lower diagram the aromatic side chains that cluster between the helices are shown in red.

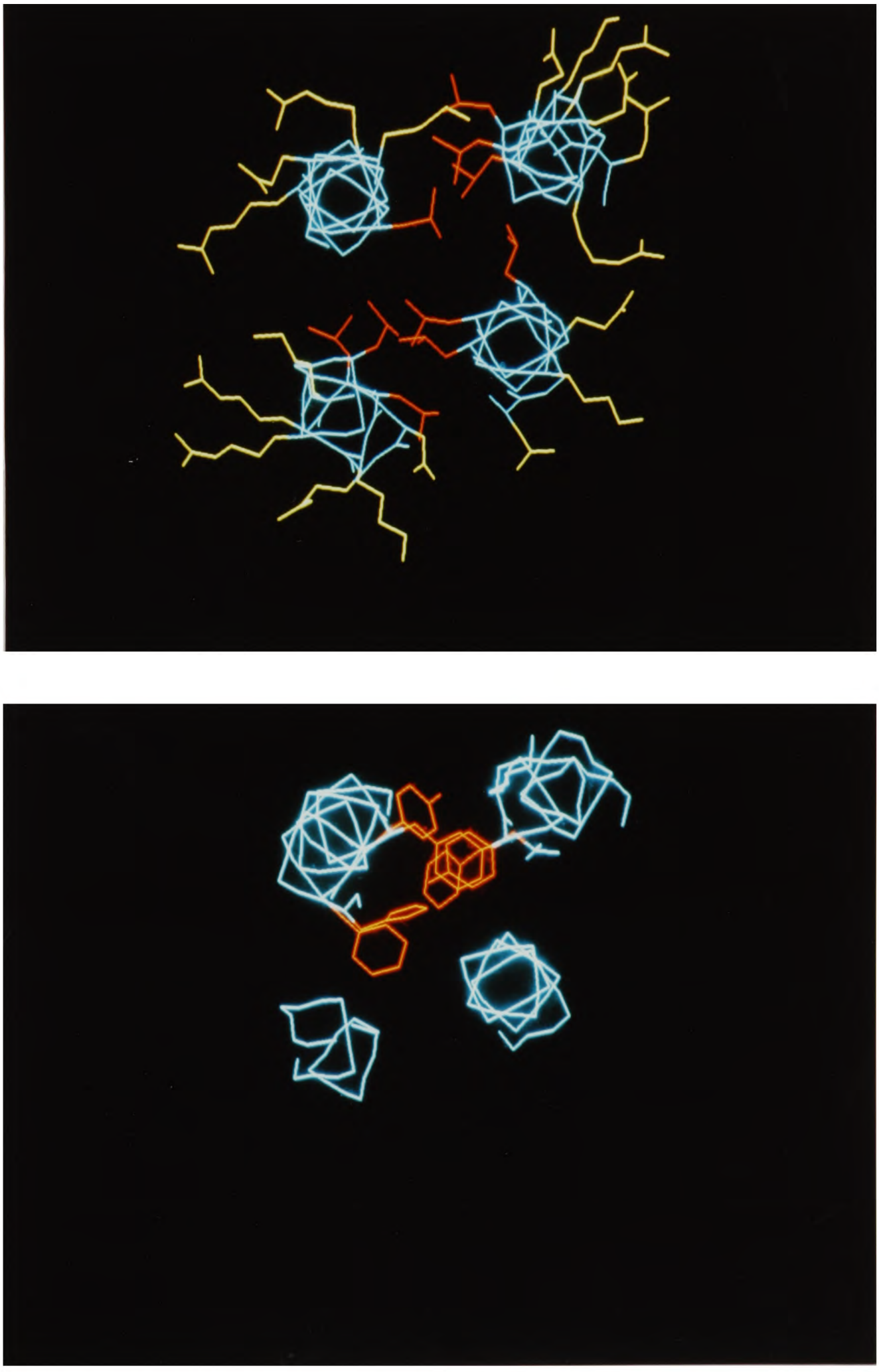
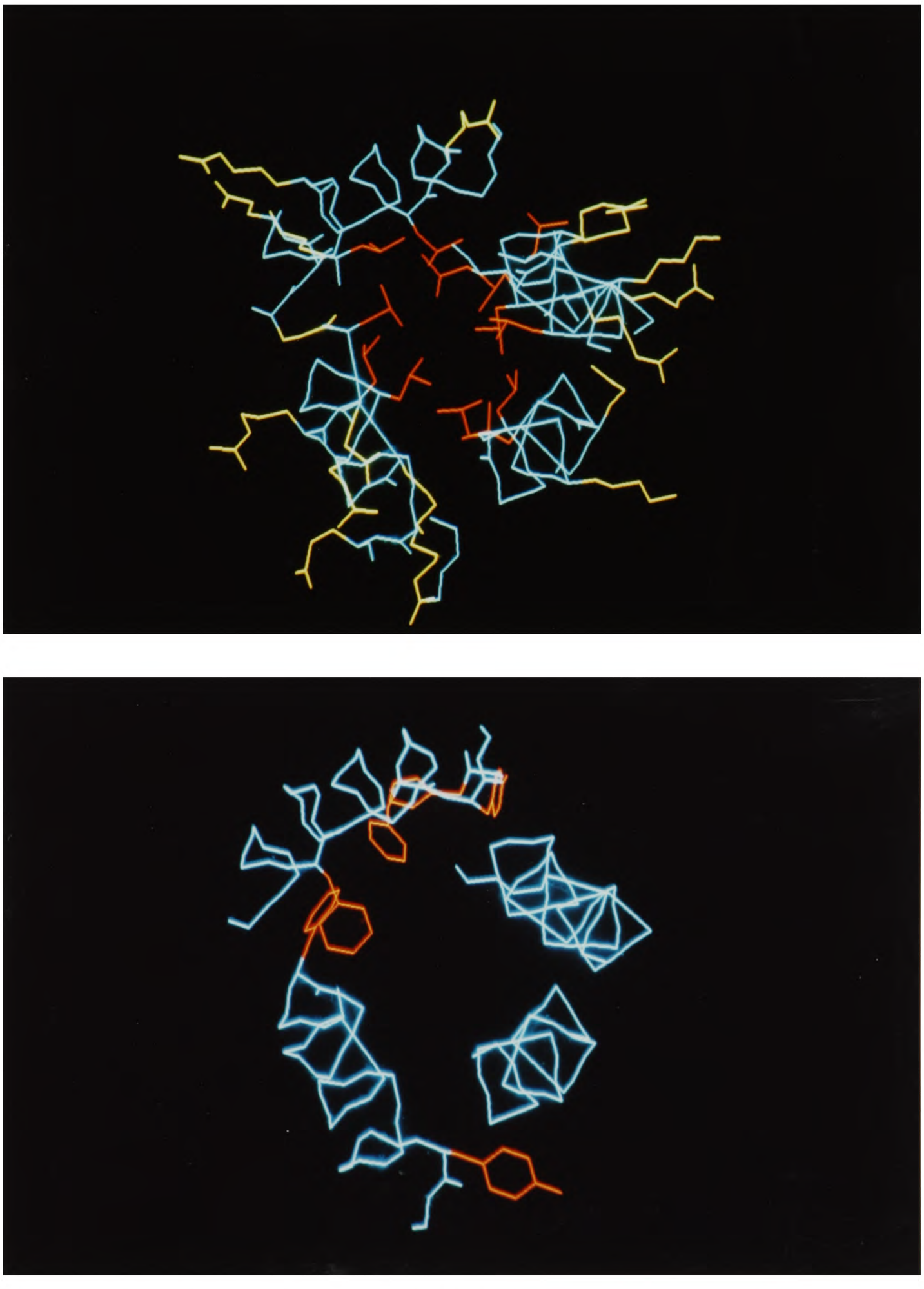


Figure 7.3.

The average energy minimized NMR structure of human IL-4. The main chain of the helical regions is shown in cyan. As for Figure 7.2 showing the structure prediction, the side chains of leucines (in red) and arginine, lysine, aspartic acid and glutamic acid residues (in yellow) are shown in the upper diagram, while the aromatic side chains are shown (in red) in the lower diagram.



positions in both structures, and the leucines are in the interior of the protein, the leucine residues do not fill the cavity between the helices forming stabilizing interhelix contacts in the structure prediction in the same way they do in the NMR structure (see Chapter 6). Hence, although some of the leucine residues come into close contact with each other in the structure prediction (e.g 17Cγ to 52Cγ distance 3.6Å) many are far apart (e.g. 17Cγ to 86Cγ distance 17.0Å; 113Cγ to 83Cγ distance 16.8Å). Instead aromatic residues are the main residues which fill the region in between the helices in the structure prediction (Phe 45, 55, 122, Tyr 56, 124, His 59) (see Figure 7.2 and compare with Figure 7.3 for the average NMR structure). It is probable therefore, that the structure prediction algorithm selected these aromatic residue interactions as the ones which would be most important in holding the helices together, resulting in the incorrect right-handed helix packing.

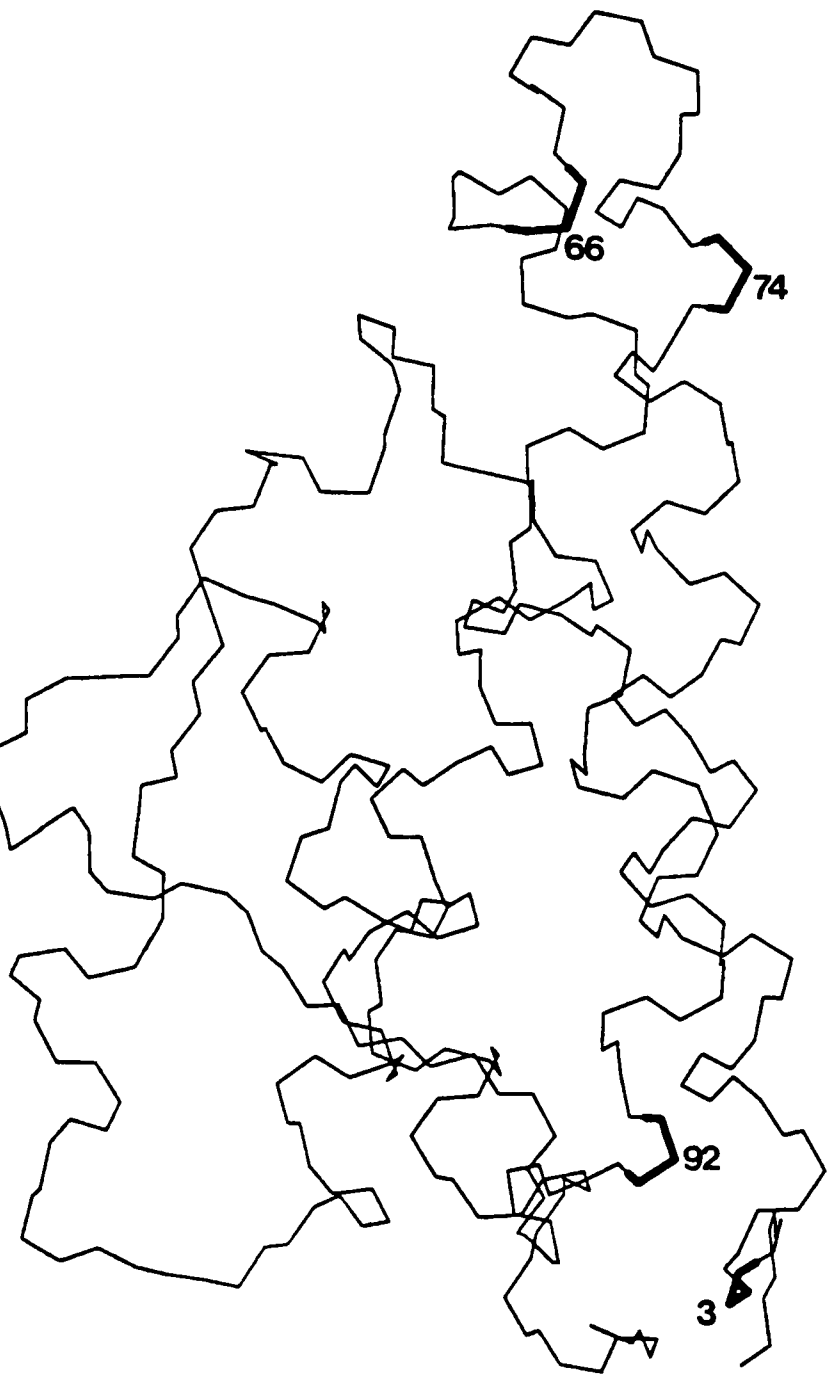
7.2 Comparison with murine IL-4.

Murine and human interleukin-4 have about 40% amino acid identity (Yokota et al., 1986) but they do not cross react with each others' receptors (Lowenthal et al., 1988; Morrison & Leder, 1992). Sequence alignment (Carr et al., 1991) shows that murine IL-4 differs from the human protein by a deletion of four C-terminal residues, an extension of the N-terminus by six residues, a deletion of residues 67-73, and the absence of the 3-127 disulphide which is replaced by a disulphide linking residues 3 to 92 (human numbers). The two other disulphides remain in similar positions in the two molecules. Study of the human IL-4 structure shows that if murine IL-4 adopts the same overall fold these changes could be readily accommodated. The extension and deletion at the N- and C-termini respectively would need no structural changes. Residues 67-73 occur in the BC loop which could be shortened without major disruption of the overall helix packing; indeed the distance between Leu 66Cα and His 74Cα is only 8.2 Å (Figure 7.4). This would make

the conformation in this region of murine IL-4 more similar to that in granulocyte-macrophage colony-stimulating factor (see section 7.3.). The change in disulphide connectivities could also be accommodated without major structural changes, as the C-terminus of helix C is close to the N-terminus of helix A in the human structure (defined by NOE's between residues 6 and 89, 6 and 93, 7 and 93). There is a distance of 13Å between the Cα atoms of residues 3 and 92 in human IL-4 (compared with a distance of 6.5Å between the Cα atoms of Cys 3 and Cys 127) and study of the structure of the human protein shows that these residues could be brought into closer proximity without changes in the relative register of helices A and C and without other major structural alterations (Figure 7.4).

This close structural similarity is supported by the results of mutagenesis studies of murine IL-4 which converted the location of the disulphide bridges in murine IL-4 to the human configuration (Morrison & Leder, 1992). To do this, Cys 87 in murine IL-4 was converted to a glycine and a Lys-Cys-Ser-Ser sequence was added to the carboxy-terminus. This variant was found to retain normal murine IL-4 biological activity. Further mutation studies, by Morrison & Leder (1992), which replaced regions of the murine IL-4 sequence with the human IL-4 sequence have located three regions that are important for species specific recognition: residues 1-16 (N-terminus and helix A), residues 68-93 (BC loop and C helix) and residues 101-117 (β sheet strand 2 and D helix). It is interesting that one of these regions is the BC loop, which the sequence comparison suggests should differ most between the two structures.

Figure 7.4.
The main chain trace of human IL-4. Residues 3 and 92, which are linked by a disulphide bridge in murine IL-4, are shown in bold. Leu 66 and His 74, the residues at either end of the region in the BC loop which is deleted in the murine IL-4 sequence, are also indicated.



7.3. Comparison with other four helix bundle proteins.

As was discussed in section 5.4.1, various classes of four-helix bundle fold exist which differ in the overall bundle handedness and in the number of overhand loops which connect up the helices (Presnell & Cohen, 1989). IL-4 adopts an up-up-down-down left-handed topology with two overhand loops, and it is interesting to compare it with other proteins that have been found to have the same topology; growth hormone (GRH) (Abdel-Meguid et al., 1987), granulocyte-macrophage colony-stimulating factor (GM-CSF) (Diederichs et al., 1991) and murine interferon-β (Senda et al., 1991).

Growth hormone and interferon-β are both larger than IL-4 (porcine GRH and murine interferon-β have 191 and 161 residues respectively) and, although they share the same left-handed up-up-down-down topology, they differ significantly in the structure of the long loops. In these proteins the AB loop passes round the outside of the D helix while in IL-4 the D helix is on the outside of the AB loop (see Figure 7.5). In addition, there is no short region of β sheet, which would be analogous to that in IL-4, connecting the AB and CD loops and there are fewer disulphide bridges (2 in porcine GRH and none in interferon-β compared with 3 in IL-4). Presumably the larger size and generally longer helices in these proteins (in GRH the A and D helices are 28 and 32 residues long respectively) lead to more favourable helix-helix contacts and reduces the need for the extra stabilizing contacts found in IL-4.

In both proteins there is also some helical nature in the loops which is interesting in the light of the potential helical turns in the loops of IL-4 discussed in Chapter 6. In interferon-β the CD loop forms a fifth more irregular helix (see Figure 7.6); the four helix bundle core here consists of helices A, B, C and E. For GRH, in the recent crystal structure of human growth hormone bound to its receptor (de Vos et al., 1992), which is discussed further in section 7.5, the growth hormone molecule is found, when bound,

Figure 7.5.

Schematic diagram showing the two different loop conformations that have been observed in left-handed four helix bundles with an up-up-down-down connectivity. Structure A shows the fold adopted by IL-4 and GM-CSF, where the D helix is on the outside of the AB loop. The fold of GRH and interferon-β, where the AB loop passes round the outside of helix D, is shown in B. (Note that none of the helical or β sheet structure found in the loop regions of these proteins is shown in these schematic diagrams).

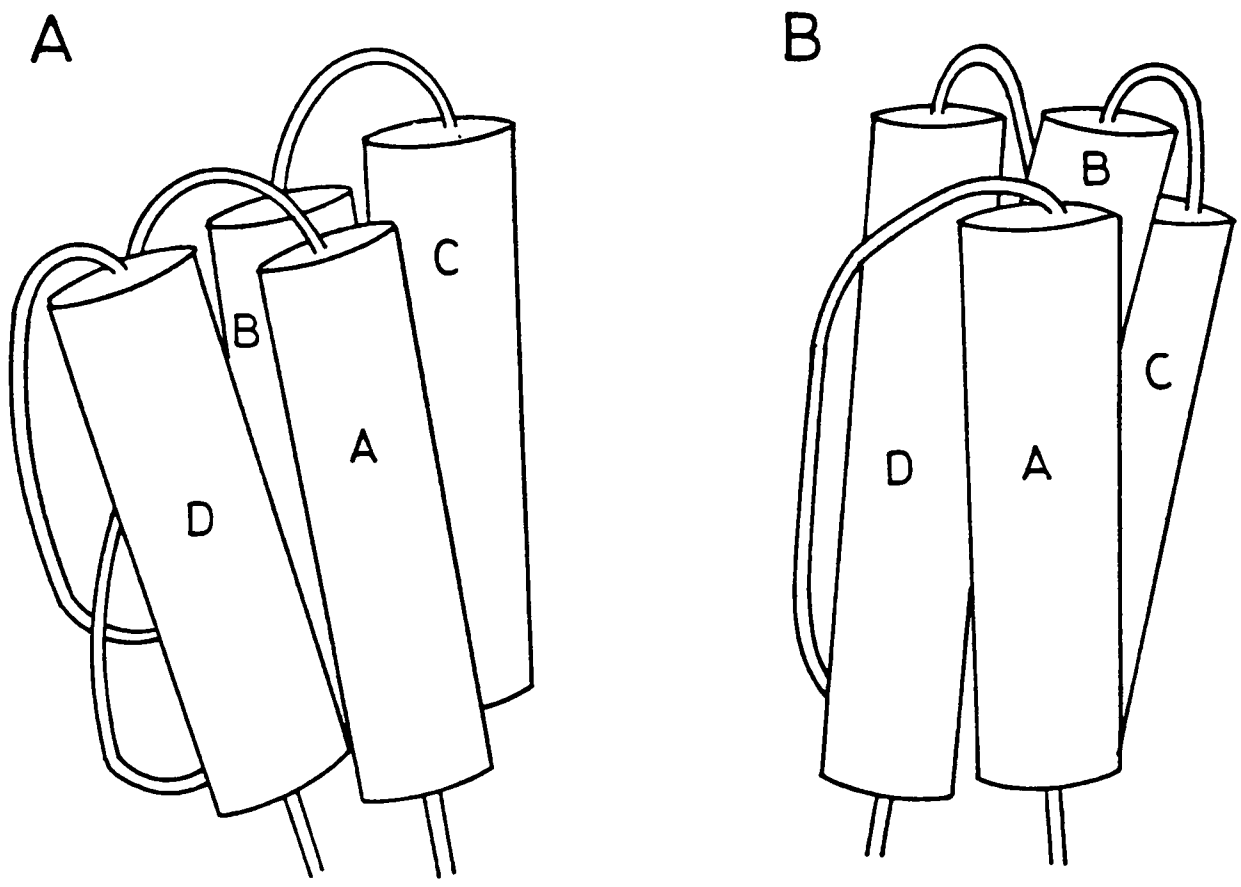
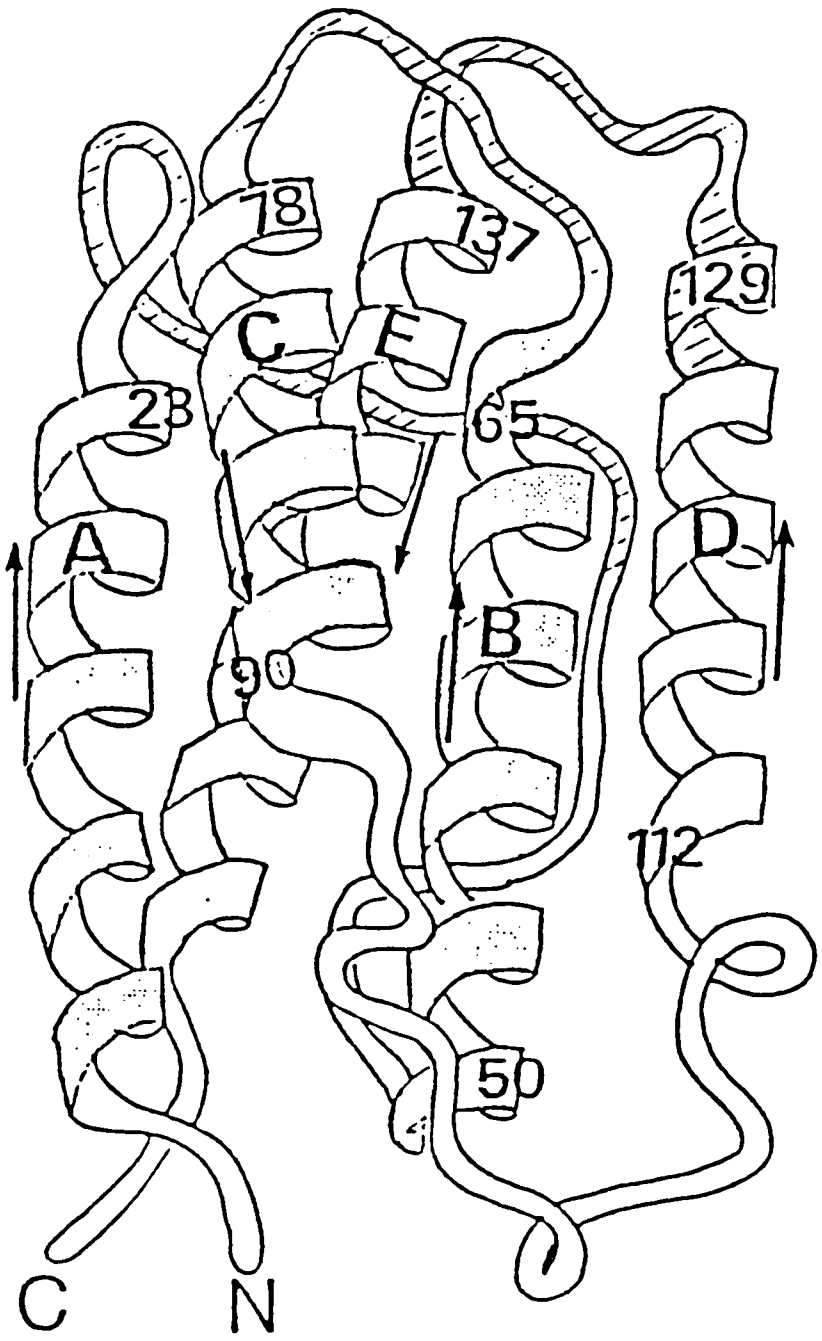


Figure 7.6.
Ribbon diagram showing the fold of murine interferon-β. Helices A, B, C and E form the four helix bundle, while the CD loop forms a fifth irregular helix (D). (From Senda et al., 1991).



to have three extra short segments of helix, one each at the beginning and end of the AB loop and one in the BC loop. As these helical regions are not present in the crystal structure of unbound porcine growth hormone, it is possible that the loop regions undergo a conformational change on binding. Alternatively, the electron density for porcine growth hormone may not have been sufficient in these regions, perhaps as a result of some mobility, to resolve the structure present in these loops.

Comparison with GM-CSF.

The recently published X-ray structure of GM-CSF (Diederichs et al., 1991) is very similar to that of human IL-4 (Figure 7.7). The proteins are of a similar size (GM-CSF has 127 residues compared to 129 in IL-4) and, as well as having the same left-handed up-up-down-down structure, they also have the same loop topology. The helices in both structures are less parallel than in classical four-helix bundles; interhelix angles of more than 40° are found in both structures. However, although two of the interhelix angles have similar values in the structures the other two, particularly the BC angle, are very different (see Table 7.2).

The short region of antiparallel β sheet is also found in both proteins. Its location is highly conserved between the structures (strand 1 at the centre of the AB loop and strand 2 at the end of the CD loop) and it must be an important factor in the stabilization of the long loops in these proteins. It is interesting, in the light of the loop mobility in IL-4, that the AB loop regions, before and after the β sheet, are recognised from the X-ray data to have high mobility in GM-CSF (Diederichs et al., 1991).

Figure 7.7.
Ribbon diagram showing the four helix bundle fold and short region of β sheet in GM-CSF (Diederichs al., 1991).

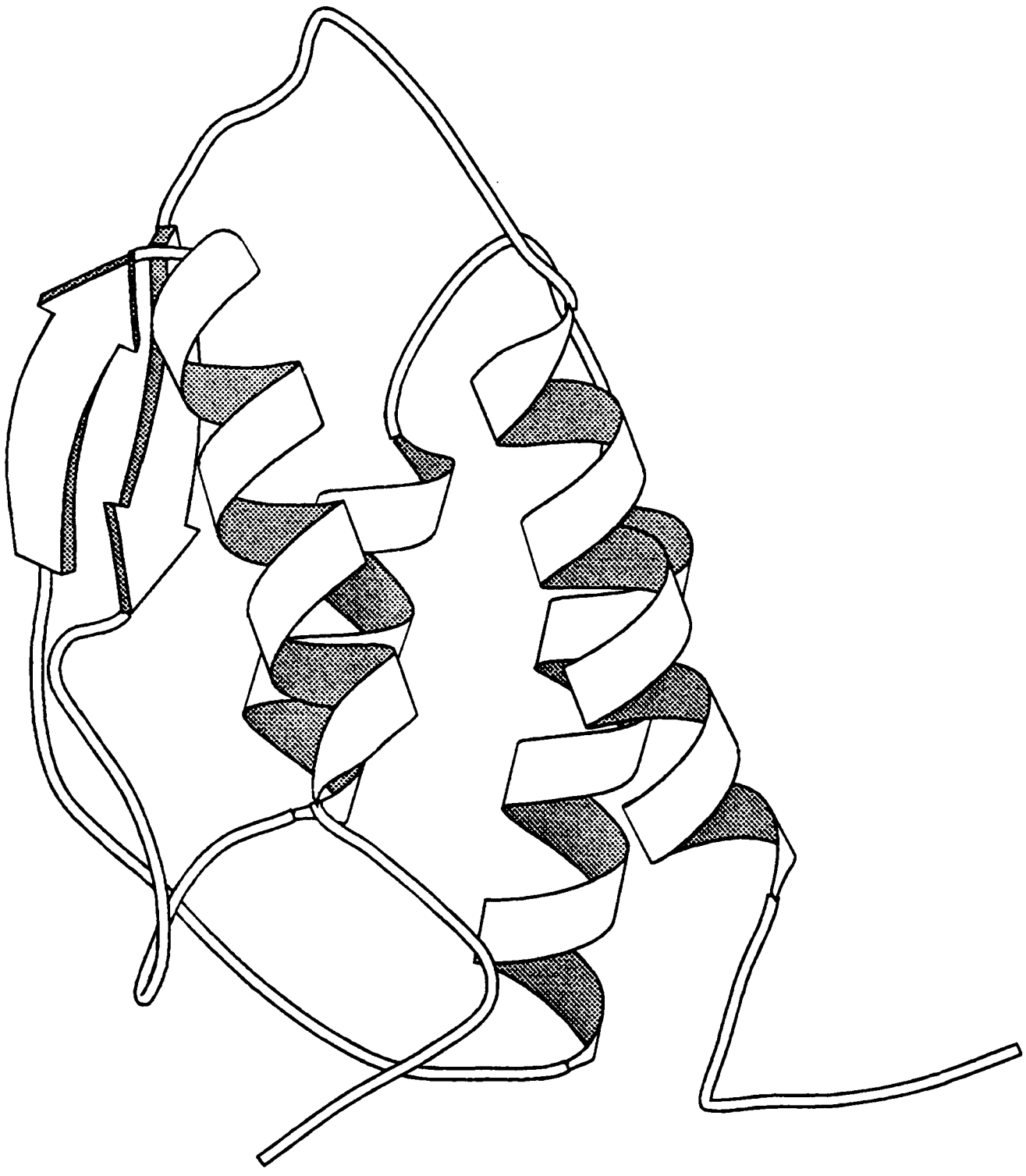


Table 7.2

Comparison of structural features in the X-ray structure of GM-CSF and the NMR structure of IL-4.

	GM-CSF	IL-4
Secondary structure	Residues	Residues
Helix A	13-28	5-17
Helix B	55-64	41-57
Helix C	74-87	72-90
Helix D	103-116	110-125
β sheet strand 1	39-43	28-30
β sheet strand 2	98-102	106-108
3_{10} helix	68-73	
Position of disulphides	54-96	46-99
	88-121	3-127
		24-65
Interhelix angle[#]		
AD	32	38.3 ± 3.3
DB	43	45.8 ± 3.1
BC	9	30.8 ± 2.8
CA	33	21.3 ± 4.7

[#] Absolute value in degrees of the acute interhelix angles as defined by Presnell and Cohen (1989).

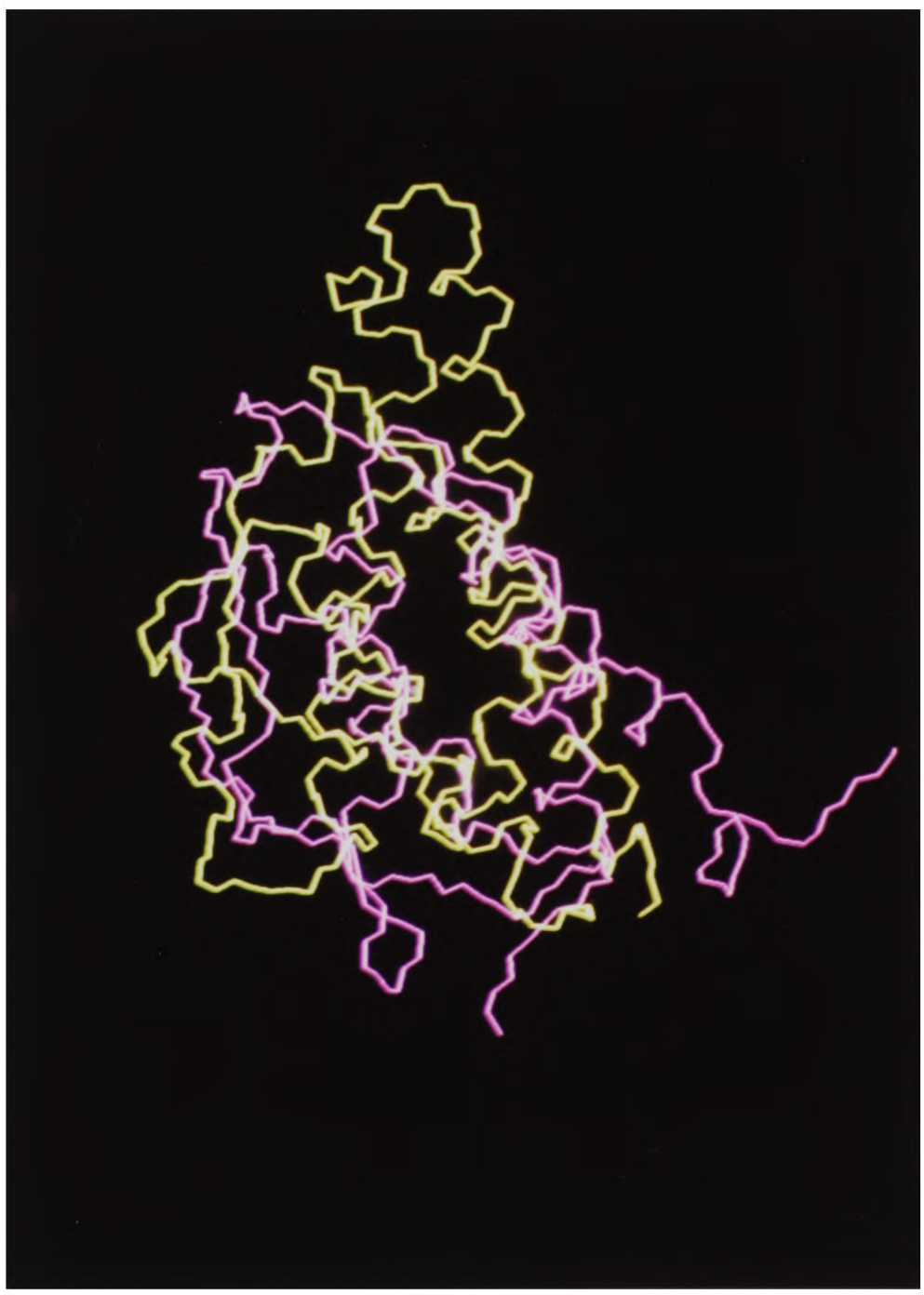
Comparison of the position of the disulphide bridges in the proteins shows that both IL-4 and GM-CSF have a disulphide bond linking helix B to the CD loop. Mutagenesis studies on human IL-4 have shown that the removal of this disulphide bridge (but not the other disulphides) leads to an unstable protein with virtually no biological activity (Kruse et al., 1991). The 3-127 disulphide bridge in human IL-4 connects the ends of the A and D helices at the bottom of the structure, while in GM-CSF a disulphide bridge connects the end of the C and D helices (88-121). As in murine IL-4 a disulphide is predicted to connect the A and C helices (see section 7.2), it suggests that a disulphide bridge connecting a pair of helices at the bottom of the bundle may be important for maintaining this protein fold.

Prior to any experimental structural data for these proteins being available, sequence alignment of human IL-4 and GM-CSF had suggested that there was little sequence similarity between the proteins although both were predicted to be rich in α helix (Bazan, 1990b; Parry et al., 1991). A recent alignment based on the structures of the proteins shows however, that there is some sequence similarity (Wlodawer et al., 1992). The segments of the protein used in this alignment are residues 4-16 in IL-4 (corresponding to 16-28 in GM-CSF), 22-30 (32-41 with 35 deleted), 42-52 (56-66), 83-95 (70-82) and 106-120 (100-114). With this alignment, 18 of the 61 residues considered in the two proteins are identical and another 9 are highly conserved. The 18 residues, which are the same in both proteins, are spread throughout the sequence (4 in helix A, 1 in helix B, 3 in helix C, 5 in helix D, 3 in the β sheet and 2 in the loop regions); 6 of them are leucines. When these regions of the two proteins are superimposed the close similarity of the folds of the two proteins is emphasized (Figure 7.8), the main chain RMSD for these 61 residues being 1.48 Å.

There are, however, some major differences between the two structures particularly in the AB and BC loops. Helix A and the AB loop are

Figure 7.8.

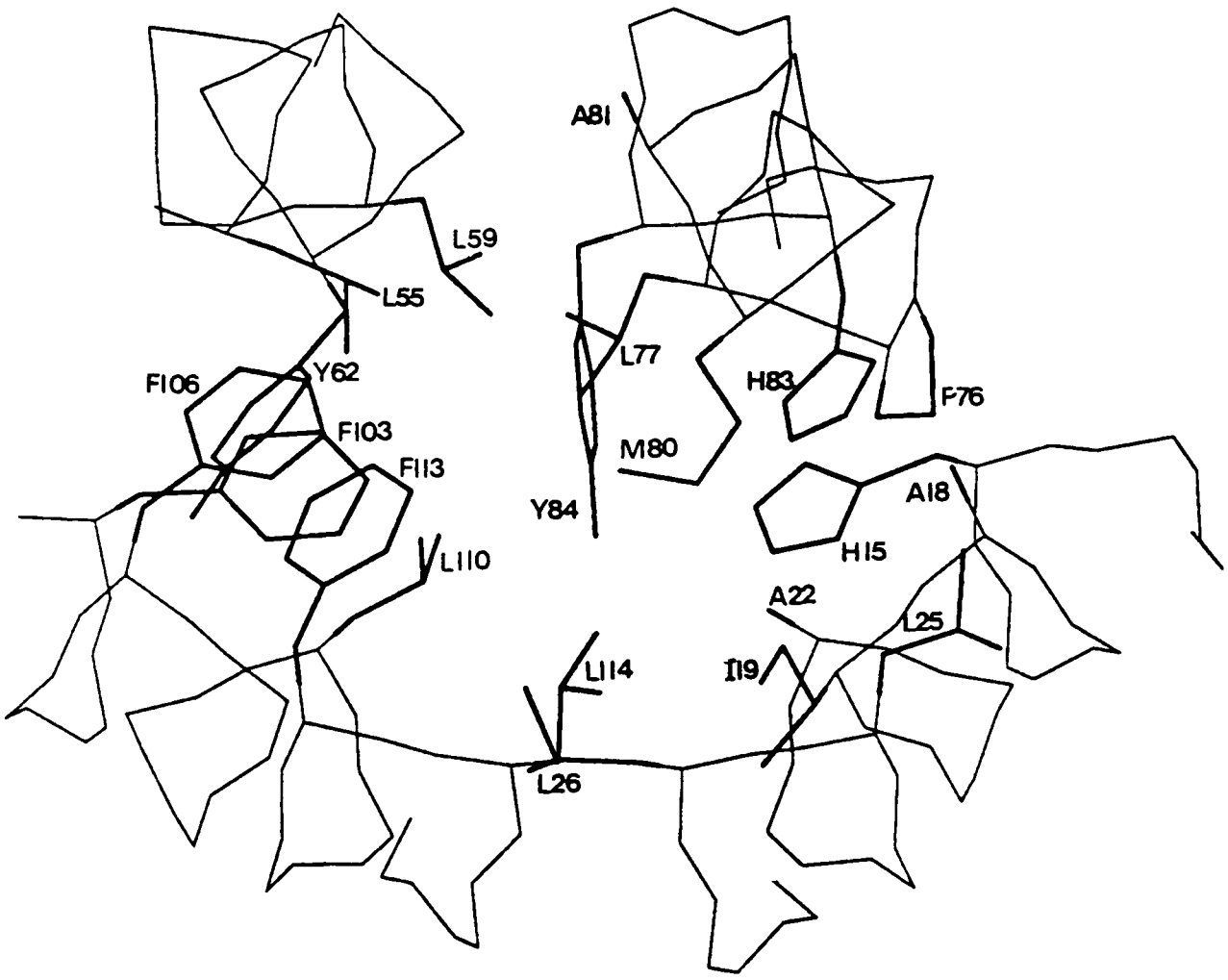
Main chain superposition of human IL-4 (in yellow in colour print) and GM-CSF (in magenta) according to the alignment of Wlodawer et al., (1992). The extension of the BC loop in IL-4 (at the top of the diagram) and the extension of the AB loop in GM-CSF (at the bottom of the diagram) can be clearly seen.



both longer in GM-CSF than IL-4, and the initial part of the AB loop extends away from the main body of the protein in GM-CSF. In contrast, the BC loop and helix C are longer in IL-4 than those in GM-CSF. This results in the N-terminal region of the C helix and the BC loop being displaced upwards (as viewed in Figure 7.7) in IL-4 relative to GM-CSF. In addition, a short region of 3_{10} helix is found in GM-CSF preceding the C helix. As the sequence comparison of human and murine IL-4 shows that a deletion occurs in murine IL-4 between residues 67-73 shortening the BC loop, it is possible that the structure in the murine protein is more 'GM-CSF-like' in this region.

A study of the helix packing in GM-CSF (Figure 7.9) shows that a range of residue types, both aromatic and aliphatic, are involved in important interhelix contacts in the protein (particularly histidine, leucine, phenylalanine, tyrosine and alanine). This contrasts with IL-4 where leucines dominate the interhelix contacts hence showing that, although there is some sequence similarity between the proteins, there are also some major differences.

Figure 7.9.
The structure of GM-CSF (Diederichs et al., 1991) showing the main chain trace of the four helices together with the side chains of the residues that form the main contacts between the helices.



7.4 Conclusions for haemopoietic cytokines.

Recently, due to the sequencing of a number of cytokine receptors, a superfamily of haemopoietic receptors has been identified (Bazan, 1990a, b; Cosman et al., 1990). The proteins in this family, which includes the receptors for growth hormone, prolactin, erythropoietin, granulocyte- and granulocyte-macrophage colony-stimulating factors and interleukins 2 (the β-subunit), 3, 4, 6 and 7, have been found to have conserved amino acid sequence elements in their extracellular, ligand-binding domains (particularly four conserved cysteines in the N-terminal region and a Trp-Ser-X-Trp-Ser sequence near the C-terminal end) In addition, the extracellular domains are predicted to have structural resemblences and it has also been suggested that they arose from a common ancestor (Bazan, 1990a, b) (Figure 7.10).

The cytokine ligands that bind to these receptors show, however, little obvious similarity in amino acid sequence (Bazan, 1990b) although they have generally been predicted to be rich in α helix. It is therefore of great interest that the cytokines IL-4 and GM-CSF have been found to have the same fold which is related to that of GRH. This finding, together with structure predictions for other cytokines (Bazan, 1990b), means that it is becoming increasingly apparent that the left-handed up-up-down-down four helix bundle may be the adopted motif for many, if not all, of these cytokine ligands. Indeed, on the basis of this hypothesis a reinterpretation of the interleukin-2 structure was proposed. The previously published X-ray structure for this protein (Brandhuber et al., 1987) was a left-handed up-down-up-down four helix bundle with short hairpin loops connecting the helices. The reinterpretation (Bazan, 1992), which is supported by NMR evidence (Mott et al., 1992) and also now by a new X-ray refinement (McKay, 1992), gives a left-handed up-up-down-down topology with a short region of β sheet analogous to that in IL-4 and GM-CSF (see Figure 7.11). The initial

Figure 7.10.
Schematic diagram showing members of the haemopoietic receptor superfamily. The conserved cysteine residues are shown by horizontal bars and the conserved Trp-Ser-X-Trp-Ser motif by black boxes. The stippled areas show the residues within which the receptor homologies are contained (from Cosman et al., 1990).

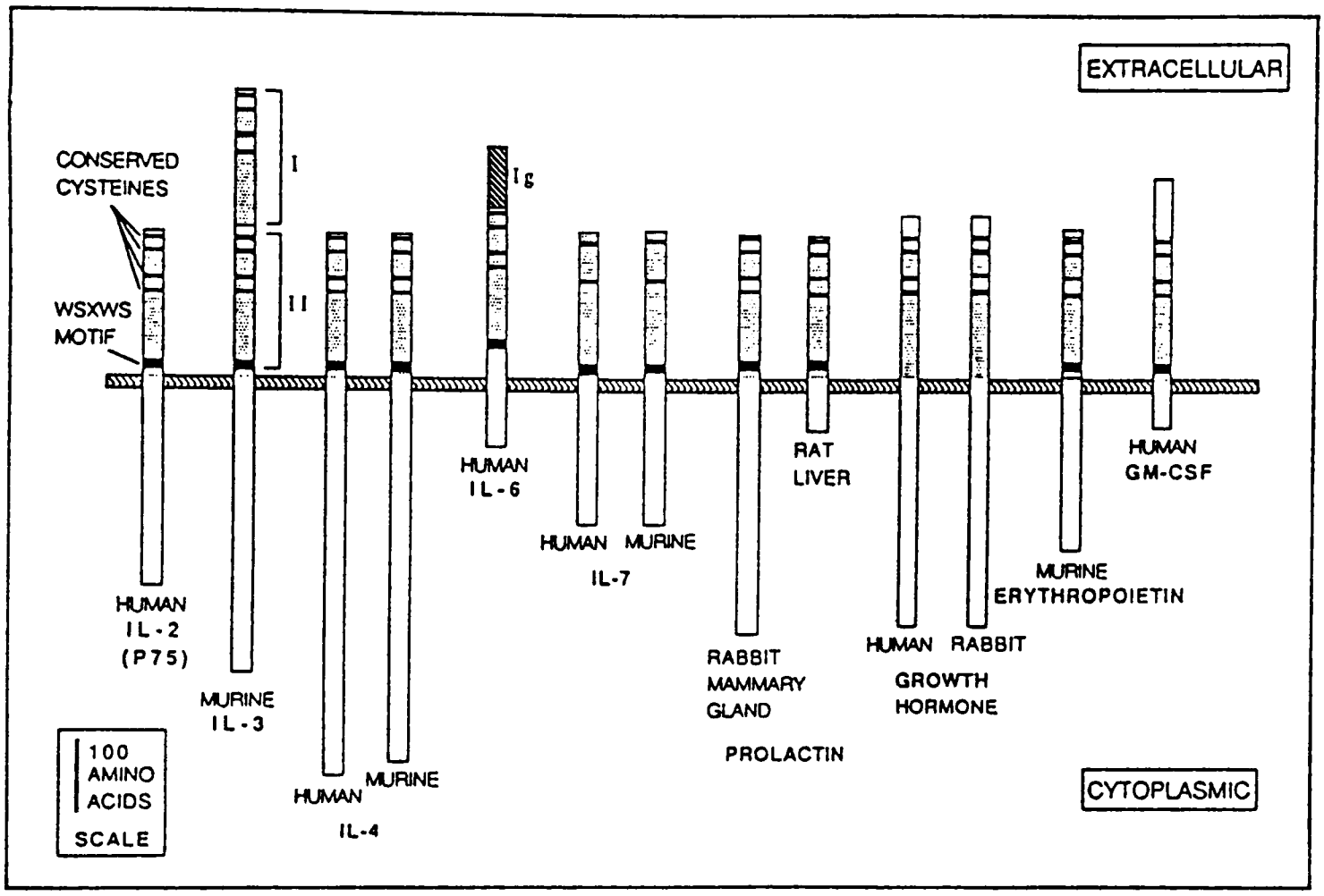
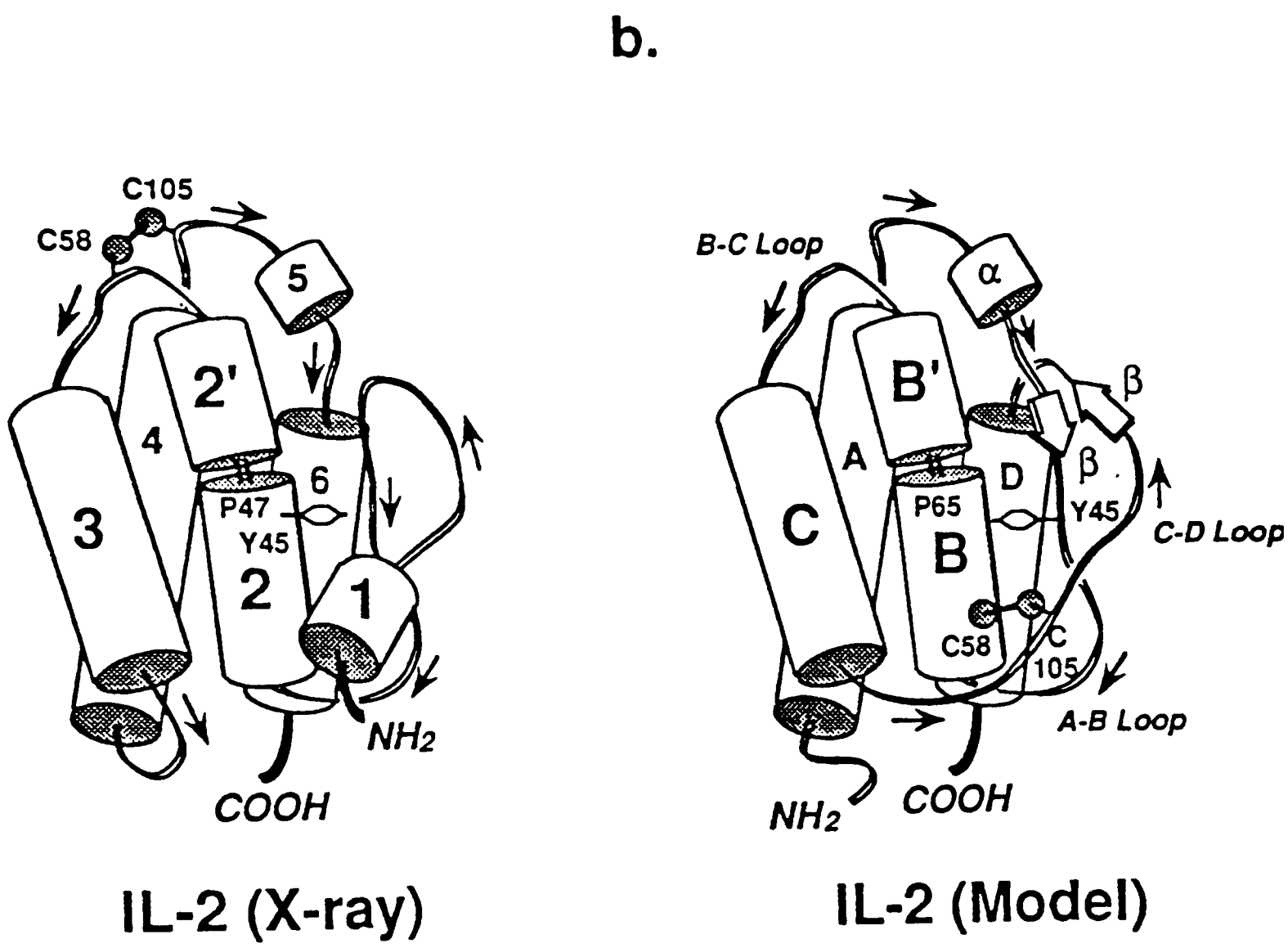


Figure 7.11.
Schematic diagrams showing the previously published X-ray structure for IL-2 (Brandhuber et al., 1987) (a) and the reinterpreted model, a left-handed four helix bundle with an up-up-down-down connectivity (b) (from Bazan, 1992).



difficulties in interpreting the crystal electron density map for IL-2 may result from dynamics in the loop regions which would make the chain indistinct in these regions. Conformational mobility of the loops may therefore also be a common feature of these cytokines, at least in their unbound form, and may explain the lack of X-ray crystallography information available at present for many of this family of proteins.

7.5 Receptor binding.

The biological activities of haemopoietic cytokines are mediated by the binding of a cytokine to its receptor. Recent studies have suggested however that, in general, some form of receptor aggregation is needed for the biological response to be initiated (Young, 1992a). Examples of homodimerization, where two identical receptors bind to the same cytokine ligand, and heterodimerization, where two different subunits are brought together on binding, have been found (Nicola & Metcalf, 1991; Young, 1992a, b). In general, the first receptor subunit binds to the cytokine with a low affinity and the interaction of this cytokine-receptor complex with the second protein then results in the formation of a high affinity (slowly dissociating) complex and signal transduction.

A recent crystal structure of the complex of human growth hormone bound to two molecules of the extracellular domain of its receptor has provided detailed information about a system where homodimerization occurs (de Vos et al., 1992). In the complex, both receptors use essentially the same residues in the interaction with the hormone even though the overall shape of the two binding sites on growth hormone is different. The first binding site on the growth hormone molecule involves exposed residues on the face of mainly helix D but also on helix A together with residues in the AB loop. The second binding site is made up of the exposed sides of helices A and C.

Examples of cytokines that bind to two (or more) different protein chains include IL-2 which forms a high affinity complex on binding to both the α and β subunits of its receptor (Smith, 1988), and the group of proteins GM-CSF, IL-3 and IL-5. These cytokines bind at a low affinity to a specific α chain which does not result in signal transduction. However all three receptor-ligand complexes can then associate with a second chain (KH97), which results in signal transduction (Nicola & Metcalf, 1991).

It is not known what type of receptor aggregation, if any, occurs in the case of IL-4. To date there is only one recognised receptor for IL-4 (Idzerda et al., 1990) so homodimerization of the receptor analogous to that in the growth hormone complex seems likely. However, there have been reports in the patent literature regarding further IL-4 binding proteins (Young, 1992a). There is also little known about the residues in IL-4 which are involved in binding to the recognised receptor. Proteolytic cleavage studies by Le et al. (1991) have shown that the C-terminal residues of the protein are critical in receptor binding. Further information on this region has come from the work of Kruse et al. (1991) which has shown that the Y124D mutant is inactive. Since Tyr 124 is exposed it is likely that this residue has a functional role in receptor binding rather than being important for structural integrity. Studies of Morrison and Leder (1992) on murine IL-4 found, by alanine substitution, that residues 104, 106, 107 and 111 are important in the binding of murine IL-4 to its receptor. Three of these residues are conserved in human IL-4: Leu 109, Phe 112 and Leu 116 (human numbering). The D helix in the IL-4 structure is therefore obviously involved in binding to the receptor although it is very likely that further regions of the structure will also be involved in the binding site. Resolving the details of receptor binding for IL-4 is therefore an important area for further work.

POSTSCRIPT.

Comparison of the NMR and crystal structures of human IL-4.

In Chapters 3 and 4 of this thesis the structure of hen lysozyme in solution was compared extensively with crystal structures of the protein which had been determined prior to the NMR work. The structure of human IL-4 described in Chapters 5 to 7 of this thesis was, in contrast, determined by NMR methods before a crystal structure of the protein had been solved. Very recently, however, a crystal structure for human IL-4 has been solved at 2.25Å resolution (R factor of 0.218) (Wlodawer et al., 1992). It has been particularly interesting to be able to complete the work described in this thesis with a preliminary comparison of this crystal structure of IL-4 with the NMR structure; the results of this comparison are described here.

1. Global comparison.

Figure 1 shows a main chain superposition of the crystal structure and the average energy minimized NMR structure coordinates. The RMSD values for the superposition, both with the average structure and also with the individual NMR structures, are given in Table 1. These values show the close similarity of the NMR and crystal structures of the protein particularly in the core of the molecule.

Figure 1.
Main chain superposition of the crystal (yellow) and average energy minimized NMR (cyan) structures of human IL-4.

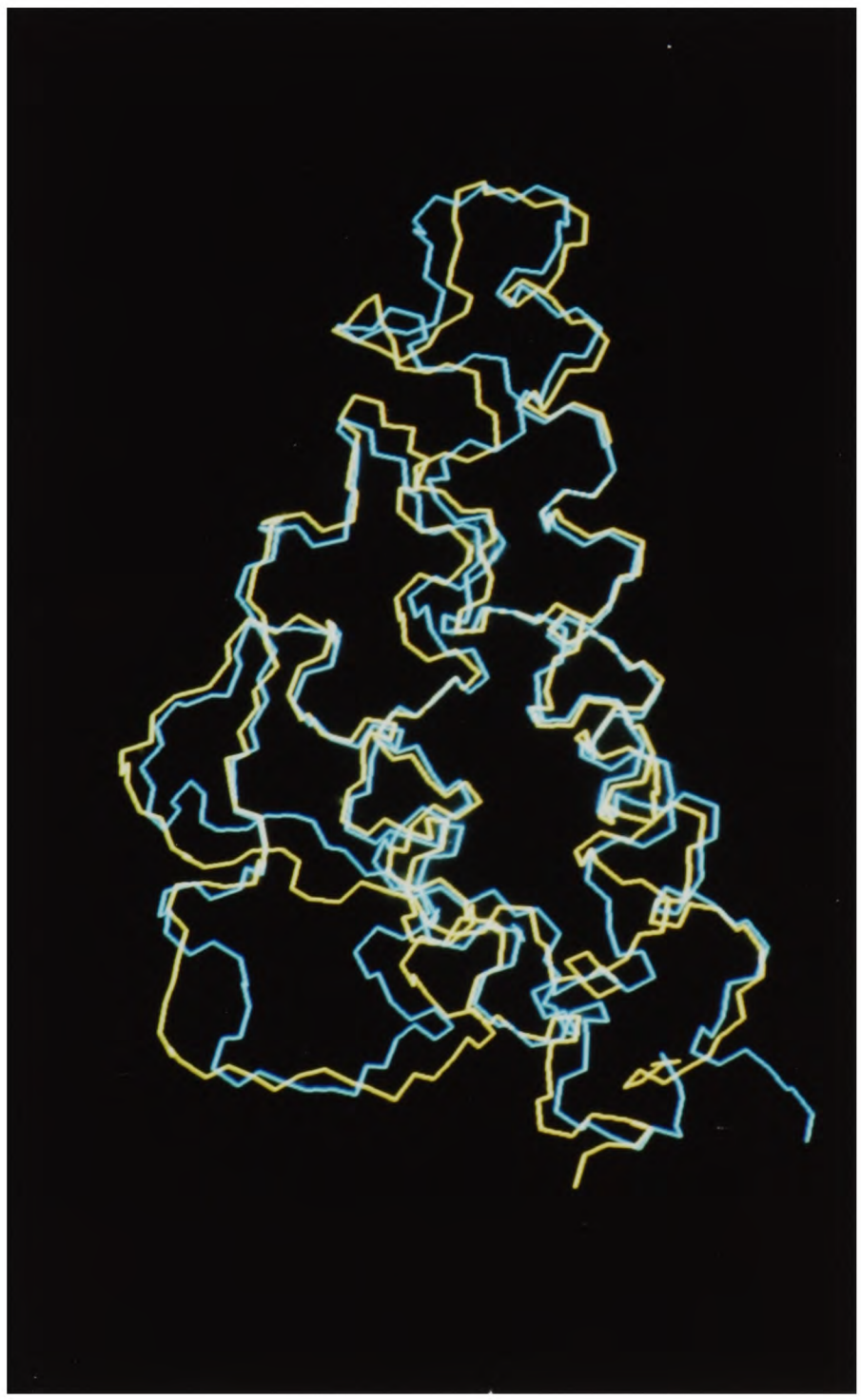


Table 1. Main chain RMSD values between the crystal and NMR structures of human IL-4.

Residues in RMSD	Crystal vs average [1]	Crystal vs individual [2]
1-129 [3]	2.06 Å	2.42 ± 0.30 Å
3-127	1.66 Å	1.95 ± 0.21 Å
Helices [4]	1.05 Å	1.31 ± 0.21 Å

[1] Average is the average energy minimized NMR structure.

[2] Values are the mean and standard deviation of the RMSD values for the 10 individual NMR structures superimposed on the crystal structure.

[3] Recombinant protein with N-terminal methionine was used in the crystallography work (as in the NMR). However, N-terminal Met was not seen in the electron density map so crystal structure is for residues 1-129 only.

[4] Helices as defined by NMR i.e. 5-17, 41-57, 72-90, 110-125; this is discussed further in the next section.

There are, however, some regions of difference which are clearly indicated in Figures 2A and B, which show the Cα deviations for each residue between the NMR structures and the crystal structure of IL-4. The regions of greatest difference (Cα deviation between average NMR and crystal coordinates > 3 Å) are located in the N- and C- termini (His 1, Lys 2 and Ser 129) towards the end of the AB loop (Ser 36, Lys 37 and Gln 41) and in the CD loop (Trp 91- Gly 95, Pro 100 and Glu 103) (see Figure 3). There is also a broad region of variation between the structures in the BC loop. A comparison with the Cα deviations across the ensemble of calculated NMR structures (Figure 2c) shows that all the regions with the greatest difference

between the NMR and X-ray structures are also relatively disordered in the ensemble of NMR structures. However, as discussed in section 6.1.3, the disorder in the NMR structures in the AB2 and CD loops and in the N- and C- termini reflects, at least in part, the mobility of these regions in solution, which has been recognised from ^{15}N relaxation studies (Redfield et al., 1992). For regions of mobility, attempts to represent adequately the conformation by a single coordinate set are likely to be unsuccessful as the regions will be adopting a range of conformations. In addition, mobility leads to an averaging of the experimental parameters, such as NOE's, which makes them hard to interpret in structural terms. Hence, the differences in these regions are more likely to reflect the problems in both determining and representing the conformation of mobile regions by NMR and crystallographic techniques, rather than true structural differences. In this respect it is interesting that there are high main chain temperature factors in the crystal structure for all the regions of greatest difference between the NMR and crystal structures (Figure 2d).

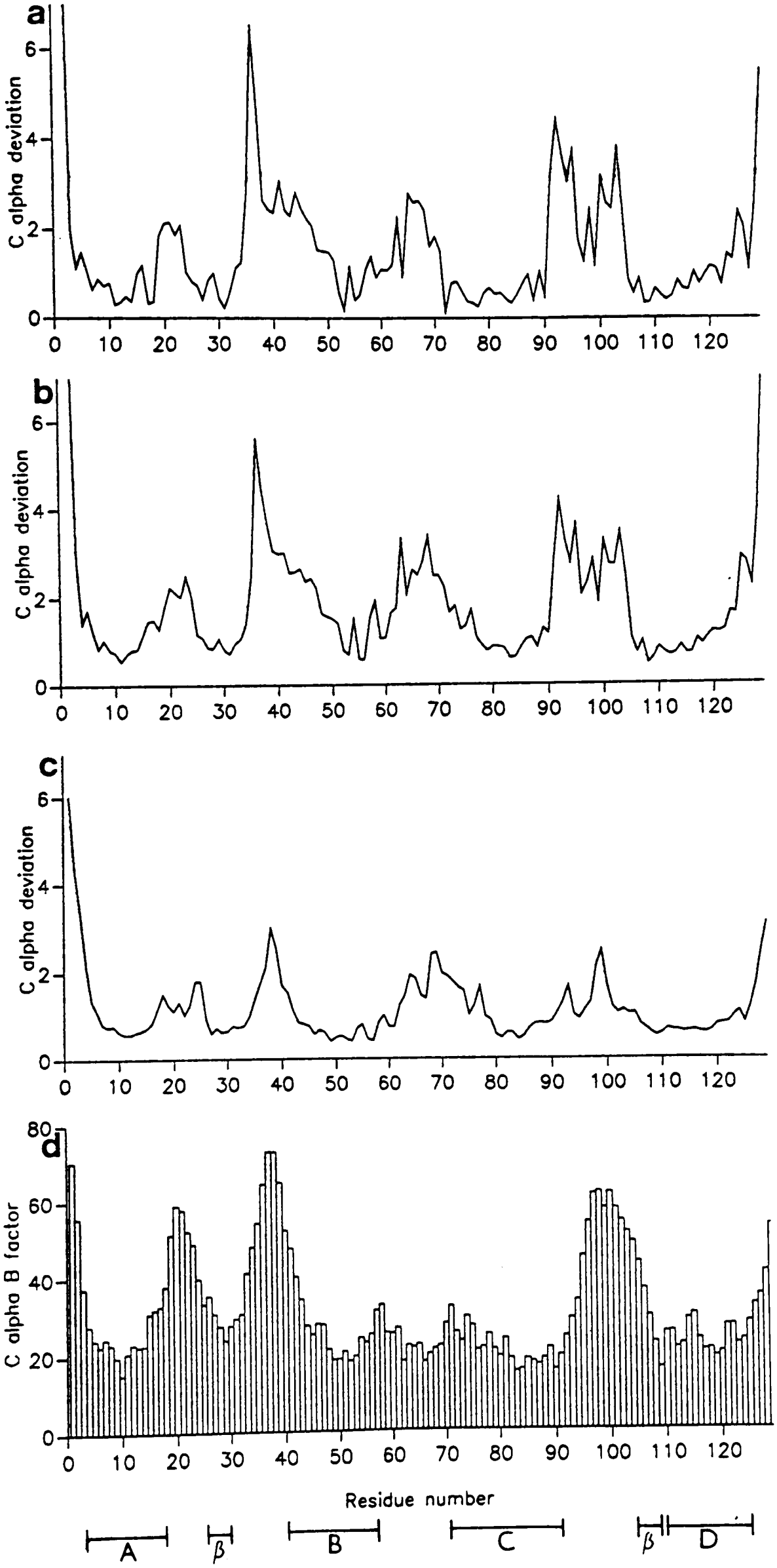
Overleaf:

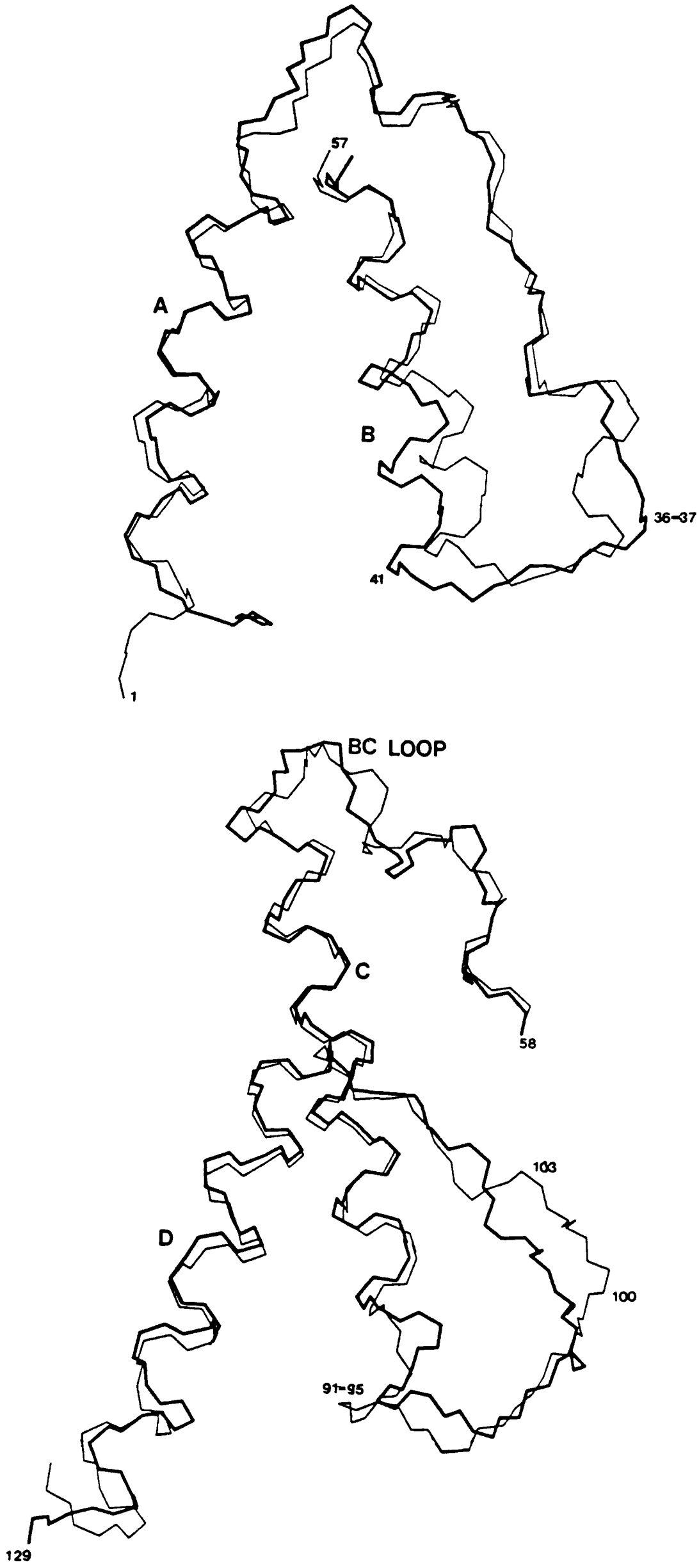
Figure 2.

Plot a shows the Cα deviation (in Å) for each residue between the average energy minimized NMR structure and the crystal structure of IL-4. Plots b and c show the average Cα deviation (in Å) for each residue of the ensemble of 10 calculated NMR structures from the crystal structure and from the average energy minimized NMR structure respectively. Plot d shows the temperature factors (in $Å^2$) of the Cα atoms in the crystal structure of IL-4. At the bottom of the diagram A, B, C and D indicate the positions of the four main helices, and β the position of the β sheet in IL-4.

Figure 3.

Superposition of the average energy minimized NMR (thin line) and crystal (bold line) structures of human IL-4. Residues 1-57 are shown in A and residues 58-129 in B. The regions of greatest difference between the NMR and crystal structures are labelled.


a
C alpha deviation
0
2
4
6
0 10 20 30 40 50 60 70 80 90 100 110 120
b
C alpha deviation
0
2
4
6
0 10 20 30 40 50 60 70 80 90 100 110 120
c
C alpha deviation
0
2
4
6
0 10 20 30 40 50 60 70 80 90 100 110 120
d
C alpha B factor
0
20
40
60
80
0 10 20 30 40 50 60 70 80 90 100 110 120
Residue number
A
β
B
C
β
D


57
A
B
36-37
41
1
BC LOOP
C
58
103
D
100
91-95
129

2. Secondary structure.

Table 2 compares the position of the secondary structure elements as defined by the NMR and crystallographic data for the protein. In all cases the NMR data defines shorter regions of secondary structure.

Table 2. Regions of secondary structure in the NMR and crystal structures of human IL-4 (as defined by Smith et al., 1992a and Wlodawer et al., 1992).

Secondary structure	NMR structure	Crystal structure
Helix A	5-17	4-19
Helix B	41-57	40-60
Helix C	72-90	70-94
Helix D	110-125	109-126
β sheet strand 1	28-30	27-31
β sheet strand 2	106-108	105-109

Comparison of the coordinate sets and study of the NMR data show, however, that this reflects the more conservative secondary structure definitions used in the NMR work rather than true differences between the structures. In crystallographic work α helices are generally recognised by the presence of (i, i+4) hydrogen bonds while in NMR studies a combination of NOE identification, amide exchange rate data and $^3J_{HN\alpha}$ coupling constant measurement is used. In this NMR work at least, helical residues are those whose torsion angles, ϕ and ψ, are close to -57° and -47° respectively, the values for a regular α helical conformation (Creighton, 1984). Indeed, for a helix $^3J_{HN\alpha}$ was required to be less than 6 Hz which means ϕ can be no greater in magnitude than -75°. Study of the ϕ angles in the crystal structure shows that 13 of the residues in the helices (crystal structure definitions) have values in the range -75° to -90°, while another 2 residues, Thr 40

and His 59, have ϕ values of 108° and -128° respectively; all of these residues would therefore not be included in the helices by the NMR definitions.

The β sheet in IL-4 is irregular; its position is therefore hard to define whether hydrogen bond criteria or NOE identification and coupling constant measurements are used.

To investigate the true differences between the secondary structure of the protein in the NMR and crystal structures, the program PROCHECK (MacArthur et al., 1992) was used to identify the positions of the α helices and β sheet in both the crystal and average energy minimized crystal structure coordinates (Table 3). This program assigns secondary structure according to the modified method of Kabsh & Sander (1983) on the basis of hydrogen bond identification.

Table 3. Regions of α helix and β sheet in IL-4 as identified by the program PROCHECK.

	NMR structure	Crystal structure
Helix A	(5)6 - 19(20)[1]	(5)6 - 19(20)
Helix B	(40)41 - 57(58)	(40)41 - 59(60)
Helix C	(69)70 - 88(89)	(69)70 - 94(95)
Helix D	(109)110 - 127(128)	109 - 128(129)
β sheet	29 - 30(31)	(27)28 - 30(31)
	(105)106 - 107	(105)106 - 108

[1] Numbers in brackets are for residues that are defined as being in an extension of the helix or β sheet.

By this method the positions of the α helices in both structures are very similar, the major difference being the five residue extension at the C-terminus of helix C in the crystal structure. However, as discussed in section

6.1.1, helix C could continue beyond Leu 90 in solution; it was not restrained to do so, however, in the present NMR structure calculations because $^3J_{HN\alpha}$ coupling constants had not been able to be measured for Trp 91 and Gly 92.

In addition to analysing the positions of the secondary structure of the protein in the two structures, the packing of the four helices in the bundle has also been compared. Table 4 gives the interhelix angles in the NMR and crystal structures. Three of the angles have very similar values, but the DB angle is larger in the NMR structures giving a more irregular bundle. This difference is reflected in the large Cα deviations around Gln 41 when the NMR and crystal structures of the protein are superimposed (see Figures 2 and 3). Although there could be true differences in the DB interhelix angle in the structure of the protein in solution and in crystals, the difference is more likely to arise from the lack of assigned NOE's between protons on the D and protons on the B helix of IL-4 in the current NMR data set (see section 6.2.1). The absence of these B to D helix NOE's in the NMR structure calculations could lead to an incorrectly defined interhelix angle. Further work will be needed to investigate this.

Table 4. Interhelix angles in the NMR and crystal structure of human IL-4*.

	NMR structure#	Crystal structure
AD	38.3 ± 3.3°	33.4
DB	45.8 ± 3.1°	33.3
BC	30.8 ± 2.8°	26.4
CA	21.3 ± 4.6°	20.8

* Absolute values of the acute interhelix angles as defined by Presnell and Cohen (1989) are given.

For the NMR structure the means and standard deviations of the angles in the set of 10 calculated structures are given.

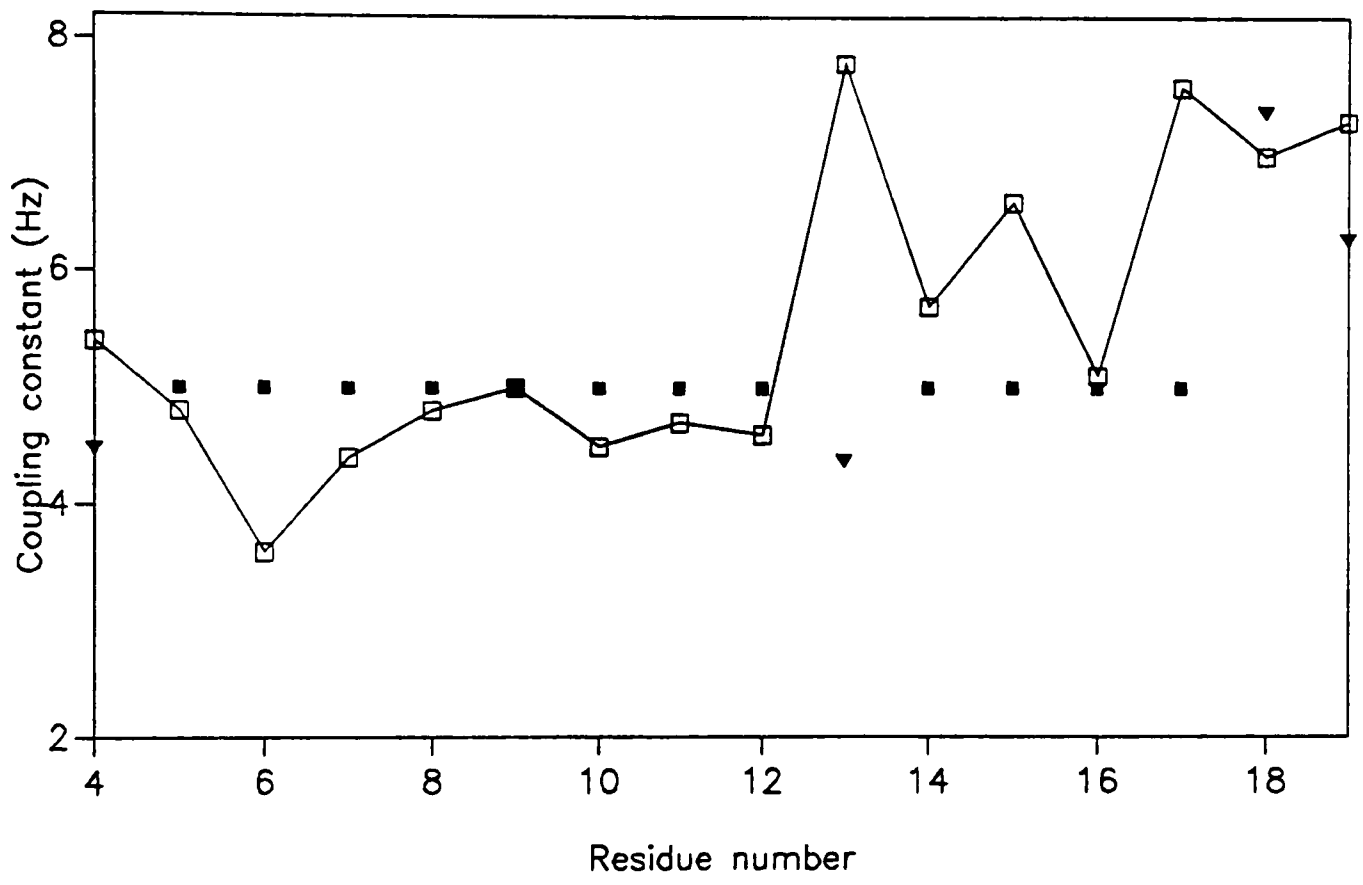
If the ϕ angles in the crystal structure are used to calculate $^3J_{HN\alpha}$ coupling constants and these are compared with the experimental NMR values the close similarity of the structure of the protein in solution and crystals is again shown. Calculation of the correlation between the experimental and calculated values (as performed for lysozyme in Chapter 3) was not possible as for residues with small $^3J_{HN\alpha}$ values in IL-4, accurate values have not been able to be measured, only ranges of $^3J_{HN\alpha}$ defined. However, comparison shows that for 61 of the 111 residues where there are $^3J_{HN\alpha}$ data for IL-4, the values calculated from the crystal structure coordinates are within ±1Hz of the experimental values or within the range defined for the experimental values. Residues where there are larger differences are found both in the secondary structure and loop regions. Figure 4 compares the calculated and experimental $^3J_{HN\alpha}$ values for helix A in the crystal structure (4-19). Large differences in the values occur towards the end of the helix; interestingly though, for Thr 18, which has an experimental $^3J_{HN\alpha}$ value of 7.4 Hz (and so the helix in the NMR structure was chosen to end at Leu 17; section 6.1.1) there is a particularly good correlation with the calculated value from the crystal structure coordinates of 7.0 Hz.

3. Loop regions.

Evidence for helical turns in the AB and BC loops was discussed in section 6.1.3. In particular, the AB2 loop (Asp 31-Thr 40) appears to be occupying multiple conformations in solution, one of which has some helical nature. Glu 60-Leu 66, in the BC loop, seem to adopt an irregular helical conformation while the work of Powers et al. (1992a) identified a helical turn from Thr 22-Glu 26 in the AB1 loop of IL-4. It is interesting, therefore, to look for these helical turns in these regions in the crystal

Figure 4.

Comparison of the experimental $^3J_{HN\alpha}$ coupling constants (filled symbols) with those calculated form the crystal structure coordinates (open boxes) for the crystal structure A helix (residues 4-19). For the experimental data, a triangle indicates that an accurate value for the coupling constant has been obtained and a square shows the residues for which the experimental coupling constant is less than 5 Hz, but an accurate value has not been obtained.



structure. In none of the three regions (22-26, 31-40, 60-66) are there groups of residues with regular helical ϕ and ψ angles. There are also no recognised hydrogen bonds for residues 22-26 in the crystal structure so there seems to be little evidence for the helical turn in the AB1 loop identified by Powers et al. (1992a) in the current crystal structure. For residues 31-40, there are two hydrogen bonds (31CO-34NH, 30CO-32NH) although neither of these are the (i, i+4) hydrogen bonds expected in a regular α helix. Their presence indicates, however, that there is some sort of non-random structure in this region of the protein. For residues 60-66 there are two hydrogen bonds (62CO-65NH, 62CO-66NH) and some of the ϕ and ψ angles in this region are approximately helical (Table 5) indicating an irregular helical-turn for these residues in the crystal structure.

Table 5 ϕ and ψ angles for residues 60-66 in the crystal structure of human IL-4.

Residue	ϕ (degrees)	ψ (degrees)
Glu 60	-56.5	-38.2
Lys 61	-130.8	53.8
Asp 62	-96.3	126.2
Thr 63	-52.6	-36.0
Arg 64	-68.4	-15.0
Cys 65	-109.0	-31.4
Leu 66	-47.2	-35.1

As mentioned in section 6.2.2, there are some residues in IL-4 which have slowly exchanging amide protons but do not occur in the recognised secondary structure of the protein in solution (Leu 27, Thr 28, Asp 31, Ile 32, Phe 33, Gly 92, Leu 93, Leu 109). As it was not apparent from the NMR

structures what hydrogen bonds these residues could be involved in, their positions have been studied in the crystal structure. The amide protons of Leu 27 and Thr 28 are not involved in hydrogen bonds. However, Ile 32 forms the 30CO-32NH hydrogen bond in the AB2 loop mentioned earlier, Asp 31 forms a hydrogen bond across to Asn 105 in the CD loop (105CO-31NH) and Leu 109 forms a hydrogen bond back to Leu 27 in the AB loop (27CO-109NH). The final pair of residues (Gly 92, Leu 93) are in helix C in the crystal structure and so form helical type hydrogen bonds (88CO-92NH; 89CO-93NH).

4. Side chain conformations.

The χ_1 torsion angles have been compared in the NMR and crystal structures. As described in section 6.3.2, for 37 residues the same χ_1 conformation is found to be adopted in all of the ensemble of calculated NMR structures. Comparison of these χ_1 conformations with those in the crystal structure shows that for 23 of the residues the angle is within approximately 20° of that found in the crystal (for another 3 residues difference $< \pm 40°$). All these residues, with one exception, are involved in interhelix contacts (e.g. Ile 10, Ile 11, Thr 13 in A helix; Ile 119, Met 120, Tyr 124 in D helix) or other recognised interactions, such as residues Asp 31- Phe 33 and Arg 47, Thr 50, Val 51 which are involved in contacts between the AB loop and the B helix (Figure 6.16).

For 11 other residues a single χ_1 conformation is adopted in the NMR structures but a different conformation is found in the crystal structure. In 10 of these cases there is a nonstaggered χ_1 conformation in either the crystal or the NMR structures (i.e. not -60°, 60°, 180°). e.g. for Leu 17, $\chi_1 \approx -60°$ in all the NMR structures but $\chi_1 = -128.2°$ in the crystal structure, while for Phe 45, $\chi_1 = -57.9°$ in the crystal structure but $\chi_1 \approx -100°$ in all the NMR structures. In some cases the staggered conformation could result from the potential used

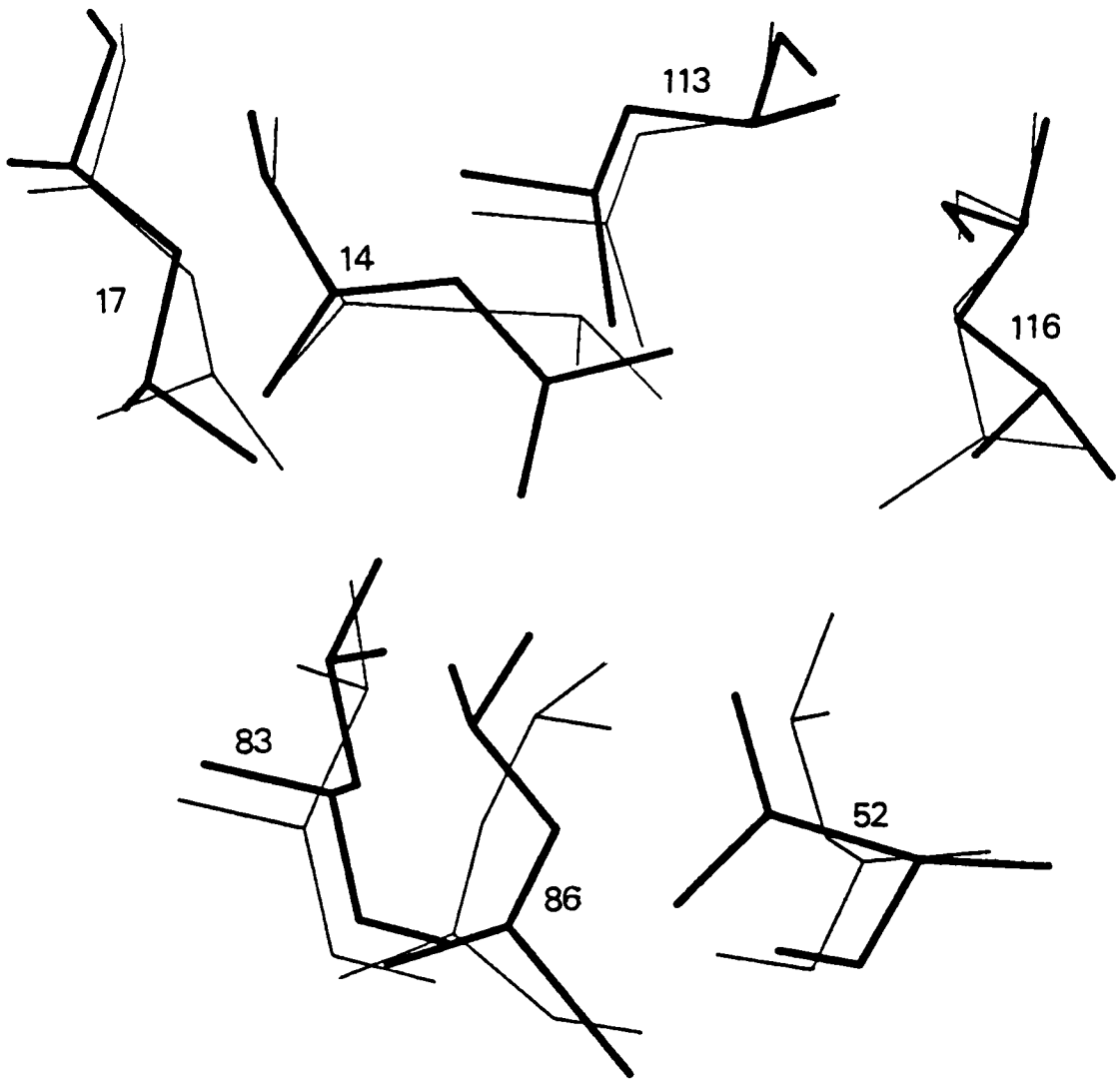
in XPLOR (which was used in the refinement of both the NMR and crystal structures). This potential favours χ_1 = -60°, 60° and 180°, and would dominate for residues where there is insufficient experimental data. For these residues the nonstaggered conformation may be the true χ_1 conformation in IL-4. This is unlikely to be the case, however, for many residues as the great majority of residues in all protein structures are found to be adopting a staggered χ_1 rotamer (McGregor et al., 1987). Alternatively, therefore, the χ_1 conformations of these residues may represent true differences between the crystal structure of the protein in crystals and in solution perhaps resulting, especially for more exposed residues, from crystallographic packing forces.

The close similarity of the orientations of the side chain of many interior residues is again shown when the conformation of the seven leucines identified to be involved in important interhelix contacts, are compared (see section 6.3.1). Figure 5 shows a superposition of these residues in the average energy minimized NMR structure and in the crystal structure of IL-4.

5. Accuracy of the NMR and crystal structures.

Although this initial comparison of the NMR and crystal structures of human IL-4 has shown the close similarity between the two structures, some regions of difference have been identified. It is difficult to be sure, however, whether these represent true differences between the structure of the protein in solution and in crystals, artifacts from either of the techniques or regions of conformational disorder and mobility. Recently, there has been much interest in methods for assessing the accuracy of the experimentally determined protein structures and for recognising regions of the structure which may be unreliably defined. For example, Eisenberg et al. (1991) have proposed the use of 3D profiles, which compare the environment of each

Figure 5.
A comparison of the conformations of the side chains of the seven leucines, which have been recognised to form particularly important interhelix contacts, in the average NMR and crystal structures of IL-4 (crystal in bold).



residue in the protein with the prefered environments of the amino acid types in question, and Morris et al. (1992) have developed criteria for investigating the reliability of a structure through an assessment of its stereochemical quality (method implemented in the program PROCHECK (MacArthur et al., 1992)). Therefore, as a step towards identifying the true significance of any observed differences between the NMR and crystal structure of IL-4, both structure have been analysed using the PROCHECK program (results given in Table 6). This program gives an overall assessment of the stereochemical quality of both the main chain and side chains of a structure and also identifies particular residues with unusual stereochemistries.

A number of residues with unusual stereochemistries have been identified by the analysis in both the NMR and crystal structures of IL-4. Although it is of course possible that these unusual conformations are adopted in the protein, it is likely that they may represent regions of the structure that are poorly define by the experimental data. In particular, in the crystal structure the ϕ and ψ angles for Asn 38 and Thr 40 lie in disallowed regions of the Ramachandran plot while groups of residues in the N- and C-termini, the AB, BC and CD loops together with residues 53-58 in the C helix have atoms which are involved in bad non-bonded interactions. For the NMR structures the residues identified vary within the ensemble of calculated structures. However, the ϕ and ψ angles for Cys 3, Asp 4, Thr 63, Glu 103, Asn 105 and Leu 109 occur in the disallowed or only generously allowed regions of the Ramachandran plot in four or more of the ensemble of ten structures and, as in the crystal structure, a few residues in the termini and loops are involved in bad non-bonded interactions. All these regions identified are those where there is disorder in the ensemble of NMR structures and differences between the NMR and crystal structures (see Figure 2).

Table 6 gives the PROCHECK parameters, which measure the stereochemical quality of the whole protein structure, for the NMR and crystal structures of IL-4. The values for the crystal structure of IL-4 are comparable with those typical for a 2.25 Å resolution structure, also listed in Table 6, although there are more bad non-bonded interactions than would be expected. For the NMR structures the parameters for peptide bond planarity, alpha carbon distortion (zeta angle standard deviation), hydrogen bond energy, and to a slightly lesser extent for mean χ_1 and χ_2 angle values, are those expected for an accurate high resolution crystal structure. In contrast, all the NMR structures have a low percentage of residues whose ϕ and ψ angles lie in the core regions of the Ramachandran plot. Similar results are found for the NMR structures of hen lysozyme (Appendix II). These results, rather than giving an indication of the reliability of the structures, reflect the methods used to calculate protein structures from NMR data and particularly the force fields employed. Another set of criteria would therefore be required to assess the accuracy of the NMR structures.

Table 6. Parameters from an analysis of the stereochemical quality of the NMR and crystal structures of human IL-4 using the program PROCHECK.

a) Stereochemistry of main chain[1]

Structure[2]	% in core	ω angle σ	bad contacts	ζ angle σ	H bond energy σ
NMR 1	66.1	0.3	12.3	0.5	0.7
NMR 2	62.1	0.4	10.0	0.4	0.7
NMR 3	67.7	0.4	13.1	0.4	0.7
NMR 4	62.9	0.4	8.5	0.4	0.8
NMR 5	66.1	0.4	6.9	0.5	0.7
NMR 6	65.3	0.4	6.9	0.4	0.7
NMR 7	66.1	0.4	8.5	0.4	0.7

NMR 8	58.9	0.5	14.6	0.4	0.7
NMR 9	59.7	0.4	10.8	0.4	0.7
NMR 10	68.5	0.3	11.5	0.4	0.7
Average	69.4	0.5	9.2	0.4	0.6
Crystal	82.9	3.5	18.6	3.7	0.7
Typical					
2.25 Å	80.5	6.0	7.0	3.1	0.9
Band width	10.0	3.0	10.0	1.6	0.2

[1] PROCHECK parameters for assessing quality of main chain:

Ramachandran plot quality measured by % of residues that are in most favoured regions of the Ramachandran plot (% in core).

Peptide bond planarity measured by the standard deviation of the protein structure's ω torsion angles from 180° (ω angle σ).

Bad non-bonded interactions measured by number of bad contacts per 100 residues (bad contacts).

Cα tetrahedral distortion measured by calculating the standard deviation of the zeta torsion angle (defined by Cα, N, C and Cβ) (ζ angle σ).

Hydrogen bond energy measured by the standard deviation of the hydrogen bond energies for main chain atoms, calculated using the modified method of Kabsch & Sander (1983), from expected value of -2.03 ± 0.75 kcal/mol (H bond energy).

[2] The values of the parameters for the 10 individual NMR structures are listed, followed by those for the average energy minimized NMR structure and the crystal structure of IL-4. For comparison the means and standard deviations of these parameters for all the 2.25Å resolution crystal structures in the Brookhaven Data Bank are given.

b) Stereochemistry of side chains[3]

Structure	χ_1 g- σ	χ_1 t σ	χ_1 g+ σ	χ_1 all σ	χ_2 t σ
NMR 1	11.6	8.3	16.6	13.3	16.4
NMR 2	14.0	14.1	16.5	15.4	13.1
NMR 3	14.3	11.4	17.4	15.0	16.1
NMR 4	13.9	13.2	17.9	15.8	14.6
NMR 5	7.9	12.3	17.7	14.6	13.4
NMR 6	17.5	13.8	11.6	13.8	13.2
NMR 7	9.5	14.2	13.3	13.3	10.8

NMR 8	14.8	12.2	16.8	14.8	15.5
NMR 9	12.0	14.8	14.8	14.6	19.1
NMR 10	13.7	14.8	15.4	15.3	12.4
Average	17.6	20.7	22.6	21.4	13.1
Crystal	19.5	27.0	18.3	22.3	28.4
Typical					
2.25Å	20.4	20.9	19.4	20.1	21.8
Band width	6.5	5.3	4.9	4.8	5.0

[3] PROCHECK parameters for assessing the quality of side chains.

χ_1 distributions measured by four parameters:

Standard deviation of the χ_1 gauche minus torsion angles around ideal value of +64.1° (χ_1 g- σ).

Standard deviation of the χ_1 trans torsion angles around ideal value of 183.6° (χ_1 t σ).

Standard deviation of the χ_1 gauche plus torsion angles around ideal value of -66.7° (χ_1 g+ σ).

Pooled standard deviation of all χ_1 torsion angles (weighted average of data for 3 conformations). (χ_1 all σ).

χ_2 distribution (for residues in trans conformation) measured by standard deviation of the χ_2 trans torsion angles from ideal value (χ_2 t σ).

6. Overall conclusions.

This comparison has clearly shown that human IL-4 has a very similar structure in solution and in crystals of the protein, particularly in the helical core of the molecule. Although other comparisons of NMR and crystal structures that have been carried out for globular proteins have come to similar conclusions (including the work on hen lysozyme presented in Chapters 3 and 4 of this thesis) the comparison for IL-4 is particularly interesting. This is partly because the NMR structure was solved prior to the X-ray structure, and indeed the NMR coordinates were used to help trace certain loop regions in the electron map, and also because of the unusual dynamic properties of the loop regions of IL-4 in solution which have been revealed by ^{15}N relaxation studies. These studies showed that much of the

loop regions in IL-4 are conformationally disordered in solution. On crystallization, this disorder could remain or, for some regions, one of the adopted conformations in solution, or even a different conformation, may become favoured. These possibilities are particularly interesting in the light of the helical conformation which appears to become energetically preferred in the AB loop of growth hormone on binding to its receptor as was discussed in Chapter 7. From the results of this preliminary comparison we can start to investigate these possibilities. For example, a helical turn has been recognised in the AB2 loop in the present crystal structure of IL-4 although the high temperature factors in this region of the loop suggest that there is still some remaining disorder. Further work is obviously required to gain a fuller picture of the structure and dynamics of these loops in IL-4 and indeed refinement of both the NMR and crystal structures is in progress. In the case of the NMR work important future objectives are the identification of further NOE's particularly from the aromatic residues Phe 45 and Phe 112, the measurement of $^{3}J_{\alpha\beta}$ couping constants and the stereospecific assignment of leucine methyl groups (using a sample with partially ^{13}C labelled leucines).

It is interesting that many of the conclusions from the detailed analysis of the IL-4 structure, and particularly the comparison with the crystal structure of the protein, are similar to those drawn from the work on hen lysozyme presented in Chapters 3 and 4 of this thesis. For example, the structures determined by NMR techniques have been recognised, in both cases, to be very dependent upon the information included in the structural data set. In general, secondary structure features do not appear in an ensemble of structures unless they are restrained to do so, the NMR therefore defining a minimal structure. This has been illustrated by the 3_{10} helix between residues 79-84 in hen lysozyme and by the end of the C helix (90-94)

in human IL-4. In both these cases the NMR data is not inconsistent with the helical conformation observed in the crystal structure, but there was insufficient NMR information available to warrant the inclusion of helical ϕ or hydrogen bond restraints for these residues in the structure calculations.

It is also apparent that NMR structure calculations can get trapped in local minima, a conformation that is in actual fact incorrect seeming well defined in the ensemble of structures. For human IL-4 this may be the case with the BD interhelix angle. There are no assigned NOE's between the B and D helices in IL-4 but this interhelix angle is well defined in the calculated structures (45.8 ±3.1°) although it differs by 12.5° from the value observed in the crystal structure. Once again, a similar feature was seen for hen lysozyme, where ϕ conformations, particularly for certain glycine residues, appeared well defined but their conformations were in actual fact being determined by the XPLOR potential energy function rather than any experimental NMR data. The true conformational variability may become more apparent in these cases if a larger ensemble of NMR structure are calculated, or if a different type of calculation procedure is employed.

A good definition of the conformations of internal side chains is clearly an important requirement for a well defined structure. For hen lysozyme, the well defined hydrophobic box region results in a correctly defined helical domain, while for human IL-4 comparison of the preliminary and final NMR structures shows this feature very clearly; good definition of the IL-4 structure was only achieved when side chain to side chain NOE's, particularly those involving internal leucine residues, had been assigned from the experiments on the double labelled protein. These results suggest that in the case of IL-4 stereospecific assignments, particularly for leucine δ methyl groups, may be an important priority for getting a higher resolution structure for the protein.

An interesting area of contrast between the work on IL-4 and that on hen lysozyme is in the conformation of the loop regions. The definition of the loops in the ensemble of NMR structures of both IL-4 and hen lysozyme is less good than the rest of the molecule. For hen lysozyme the disorder in the 61-78 loop appears to merely reflect the lack of NMR data for this region; indeed recent ^{15}N relaxation studies of hen lysozyme have shown that this loop is not significantly mobile (M. Buck, C. Redfield, C. M. Dobson, unpublished results). In the case of human IL-4, however, the AB and CD loops have been shown to be undergoing significant motions on the picosecond time scale while the BC loop may be undergoing motions on a slower time scale. For IL-4, therefore, the conformational disorder in the loops in the NMR structures does reflect the presence of conformational mobility. Indeed, one of the particular strengths of NMR techniques for the studies of protein structures, compared with the complementary X-ray diffraction methods, is its ability to study proteins with regions of conformational mobility whether these are flexible loops or merely surface side chains that adopt multiple conformations. Considerable conformational flexibility often results in proteins that are hard to crystallize and, for crystallized proteins where X-ray diffraction techniques can be used, in diffuse regions in the electron density map that cannot be easily interpreted. This has been clearly demonstrated in the case of the interleukins, where the path of the loops was initially traced incorrectly in the case of IL-2 (Bazan, 1992; McKay, 1992), and where the NMR structure of IL-4 had to be used to enable the path of the loops in the crystal structure to be interpreted (Wlodawer et al., 1992). As the exposed mobile regions of the protein structure are generally those that are directly involved in interactions with other molecules, whether they are substrates, receptors or inhibitors, the ability of NMR techniques to study these mobile regions is particularly important.

REFERENCES.

Abdel-Meguid, S. S., Shiel, H-S, Smith, W. W., Dayringer, H. E., Violand, B. N., & Bentle, L. A. (1987) *Proc. Natl. Acad. Sci. USA* **84,** 6434.

Akke, M., Drakenberg, T., & Chazin, W. J. (1992) *Biochemistry* **31,** 1011.

Anil Kumar, Ernst, R. R., & Wüthrich, K. (1980) *Biochem. Biophys. Res. Commun.* **95,** 1.

Arai, K., Lee, F., Mijajima, A., Miyatake, S., Arai, N., & Yokota, T. (1990) *Ann. Rev. Biochem.* **59,** 783.

Arrowsmith, C., Pachter, R., Altman, R., & Jardetzky, O. J. (1991) *Eur. J. Biochem.* **202,** 53.

Artymiuk, P. J. & Blake, C. C. F. (1981) *J. Mol. Biol.* **152,** 737.

Aschaffenburg, R., Blake, C. C .F., Dickie, H. M., Gayen, S. K., Keegan, R., & Sen, A. (1980) *Biochem. Biophys. Acta* **625,** 64.

Balkwill, F. R., & Burke, F. (1989) *Immunology Today* **10,** 299.

Banyard, S. H., Stammers, D. K., & Harrison, P. M. (1978) *Nature* **271,** 282.

Bax, A. (1989) *Ann. Rev. Biochem.* **58,** 223.

Bax, A., Clore, G. M., Driscoll, P. C., Gronenborn, A. M., Ikura, M., & Kay, L. E. (1990a) *J. Magn. Reson.* **87,** 620.

Bax, A., Clore, G. M., & Gronenborn, A. M. (1990b) *J. Magn. Reson.* **88,** 425.

Bazan, J. F. (1990a) *Proc. Nat. Acad. Sci. USA.* **87,** 6934.

Bazan, J. F. (1990b) *Immunology Today* **11,** 350.

Bazan, J. F. (1992) *Science* **257,** 410.

Billeter, M., Kline, A. D., Braun, W., Huber, R., & Wüthrich, K. (1989) *J. Mol. Biol.* **206,** 677.

Blake, C. C. F., Koenig, D. F., Mair, G. A., North, A. C. T., Phillips, D. C., & Sarma, V. R. (1965) *Nature,* **206,** 757.

Blake, C. C. F., Johnson, L. N., Mair, G. A., North, A. C. T., Phillips, D. C., & Sarma, V. R. (1967a) *Proc. R. Soc. London. B* **167,** 378.

Blake, C. C. F., Mair, G. A., North, A. C. T., Phillips, D. C., & Sarma, V. R. (1967b) *Proc. R. Soc. London. B* **167,** 365.

Brandhuber, B. J., Boone, T., Kenney, W. C., & McKay, D. B. (1987) *Science* **238,** 1707

.

Braun, W. (1987) *Quart. Rev. Biophys.* **19,** 115.

Braun, W., Wider, G., Lee, K. H., & Wüthrich, K. (1983) *J. Mol. Biol.* **169,** 921.

Brünger, A. T. (1988) *XPLOR Manual,* Yale University, New Haven, C.T., U.S.A.

Campbell, I. D., Dobson, C. M., Williams, R. J. P., & Xavier, A.V. (1973) *Annals of the New York Academy of Sciences.* **22,** 163.

Campbell, I. D., Dobson, C. M., & Williams, R. J. P. (1975) *Proc. R. Soc. London. A.* **345,** 41.

Campbell, I. D., Dobson, C. M., & Williams, R. J. P. (1985) *Biochem J.* **231,** 1.

Carr, C., Aykent, S., Kimack, N. M., & Levine, A. D. (1991) *Biochemistry* **30,** 1515.

Cassels, R., Dobson, C. M., Poulsen, F. M., & Williams, R. J. P. (1978) *Eur. J. Biochem.* **92,** 81.

Chou, P. Y., & Fasman, G. D. (1978) *Annu. Rev. Biochem.* **47,** 251.

Clore, G. M., & Gronenborn, A. M. (1989) *Critical Reviews in Bioch. and Mol. Biol.* **24**, 479.

Clore, G. M., & Gronenborn, A. M. (1991a) *Ann. Rev. Biochem.* **20**, 29.

Clore, G. M., & Gronenborn, A. M. (1991b) *J. Mol. Biol.* **217**, 611.

Clore, G. M., & Gronenborn, A. M. (1991c) *Progress in NMR Spec.* **23**, 43.

Clore, G. M., & Gronenborn, A. M. (1991d) *Science* **253**, 1390.

Clore, G. M., Bax, A., Driscoll, P. C., Wingfield, P. T., & Gronenborn, A. M. (1990) *Biochemistry* **29**, 8172.

Clore, G. M., Wingfield, P. T., & Gronenborn, A. M. (1991) *Biochemistry* **30**, 2315.

Cohen, F. E., Abarbanel, R. M., Kuntz, I. D., & Fletterick, R. J. (1986) *Biochemistry* **25**, 266.

Cook, W. J., Ealick, S. E., Reichert, P., Hammond, G. S., Le, H. V., Nagabhushan, T. L., Trotta, P. P., & Bugg, C. E. (1991) *J. Mol. Biol.* **218**, 675.

Cosman, D., Lyman, S. D., Idzerda, R. L., Beckmann, M. P., Park, L. S., Goodwin, R. G., & March, C. J. (1990) *TIBS* **25**, 265.

Creighton, T. E. (1984) in '*Proteins: Structures and molecular properties*' Freeman, New York.

Curtis, B. M., Presnell, S. R., Srinivassan, S., Sassenfield, H., Klinke, R., Jeffery, E., Cosman, D., March, C. J., & Cohen, F. E. (1991) *Proteins* **11**, 111.

DeFrance, T., Vanbervliet, B., Aubrey, J. P., Takebe, Y., Arai, N., Miyajima, A., Yokota, T., Lee, F., Arai, K., de Vries, J. E., & Banchereau, J. (1987) *J. Immunol.* **139**, 1135.

DeFrance, T., Vanbervliet, B., Durand, I., & Banchereau, J. (1988) *J. Immunol.* **141**, 2000.

de Marco, A., Llinás, M., & Wüthrich, K. (1978) *Biopolymers* **17**, 617.

de Vos, A. M., Ultsch, M., & Kossiakoff, A. A. (1992) *Science* **255**, 306.

Diederichs, K., Boone, T., & Karplus, P. A. (1991) *Science* **254**, 1779.

Dobson, C. M. (1990) *Nature* **348**, 198.

Dobson, C. M. (1991) *CIBA Foundation Symposium on Protein Conformation* **161**, 167-189.

Dobson, C. M., & Evans, P. E. (1984) *Biochemistry* **23**, 4267.

Dobson, C. M., & Karplus, M. (1986) *Methods in Enzymology.* **131**, 362.

Driscoll, P. C., Clore, G. M., Marion, D., Wingfield, P. T., & Gronenborn, A. M. (1990) *Biochemistry* **29**, 3542.

Eisenberg, D., Bowie, J. U., Lüthy, R., & Choe, S. (1992) *Faraday Discuss.* in press.

Faber, H. R., & Matthews, B. W. (1990) *Nature* **348**, 263.

Fesik, S. W., & Zuiderweg, E. R. P. (1990) *Quart. Rev. Biophys.* **23**, 97.

Finzel, B. C., Poulos, T. L., & Kraut, J. (1984) *J. Biol. Chem.* **259**, 13027.

Forman-Kay, J. D., Clore, G. M., Wingfield, P. T., & Gronenborn, A. M. (1991) *Biochemistry* **12**, 2685.

Frenkiel, T., Bauer, C., Carr, M. D., Birdsall, B., & Feeney, J., (1990) *J. Magn. Reson.* **90**, 420.

Garnier, J., Osguthorpe, D. J., & Robson, B. (1978) *J. Mol. Biol.* **120**, 97.

Garrett, D. S., Powers, R., March, C. J., Frieden, E. A., Clore, G. M., & Gronenborn, A. M. (1992) *Biochemistry* **31**, 4347.

Giessner-Prettre, C., & Pullman, B. (1969) *CR Hebd. Seances Acd. Sci. Ser. D* **268**, 1115.

Grace, D. E. P. (1980) *D. Phil. Thesis*, Oxford University.

Griesinger, C., Sørensen, O. W., & Ernst, R. R. (1987) *J. Magn. Reson.* **75**, 474.

Haasnoot, C. A. G., de Leeuw, F. A. A. M., & Altona, C. (1980) *Tetrahedron* **36**, 2783.

Handoll, H. H. G. (1985) *D.Phil. thesis*, Oxford University.

Havel. T. F. (1991) *Prog. Biophys. Molec. Biol.* **56**, 43.

Havel, T., & Wüthrich, K. (1984) *Bull. Math. Biol.* **46**, 673.

Hoch, J. C. (1983) *Ph. D. Thesis*, Harvard University.

Hoch, J. C. , Dobson, C. M., & Karplus, M. (1982) *Biochemistry* **21**, 1118.

Hoch, J. C., Dobson, C. M., & Karplus, M. (1985) *Biochemistry* **24**, 3831.

Hough, E., Hansen, L. K., Birknes, B., Jynge, K., Hansen, S., Hordvik, A., Little, C., Dodson, E., & Derewenda, Z. (1989) *Nature* **338**, 357.

Howard, M., Farrar, J., Hilfiker, M., Johnson, B., Takatsu, K., Hamaoka, T., & Paul, W. E. (1982) *J. Exp. Med.* **155**, 914.

Hyberts, S. G., Goldberg, M. S., Havel., T. F., & Wagner, G. (1992) *Protein Science* **1**, 736.

Idzerda, R. L., March, F. J., Mosley, B., Lyman, S. D., Vanden Bos, T., Gimpel, S. D., Din, W. S., Grastein, K. H., Widmer, M. D., Park, L. S., Cosman, D., & Beckmann, M. P. (1990) *J. Exp. Med.* **171**, 861.

Ikura, M., Bax, A., Clore, G. M., & Gronenborn, A. M., (1990a) *J. Am. Chem . Soc.* **112**, 9020.

Ikura, M., Kay, L. E., Tschudin, R., & Bax, A. (1990b) *J. Magn. Reson.* **86**, 204.

Ikura, M., Spera, S., Barbato, G., Kay, L. E., Krinks, M., & Bax, A. (1991) *Biochemistry* **30**, 9216.

Imoto, T., Johnson, L. N., North, A. C. T., Phillips, D. C., & Rupley, J. A. (1972) *The Enzymes* (Boyer, P. D., Ed.) 3rd ed., Vol. 7, p665, Academic, New York.

IUPAC-IUB Commission on Biochemical Nomenclature. (1970) *J. Mol. Biol.* **52,** 1.

Jardetzky, O., & Roberts, G. C. K. (1981) in '*NMR in Molecular Biology*' Academic Press, New York.

Jeener, J., Meier, B. H., Bachmann, P., & Ernst, R. .R. (1979) *J. Chem. Phys.* **71,** 4546.

Johnson, C. E., & Bovey, F. A. (1958) *J. Chem. Phys.* **29**, 1012.

Jollès, P., & Jollès, J. (1984) *Molecular & Cellular Biochem.* **63,** 165.

Kabsch, W., & Sander, C. (1983) *Biopolymers* **22,** 2577.

Kaptein, R., Zuiderweg, E. R. P., Scheek, R. M., Boelens, R., van Gunsteren, W. F. (1985) *J. Mol. Biol.* **182,** 179.

Karray, T., DeFrance, T., Merle-Beral, Banchereau, J. Debre, P., & Galanaud, P. (1988) *J. Exp. Med.* **168,** 85.

Karplus, M. (1959) *J. Chem. Phys.* **30,** 11.

Kay, L. E., & Bax, A., (1990) *J. Magn. Reson.* **86,** 110.

Kay, L. E., Torchia, D. A., & Bax., A. (1989) *Biochemistry* **28,** 8972.

Kay, L. E., Ikura, M., & Bax, A. (1990) *J. Amer. Chem. Soc.* **112,** 888.

Klevit, R. E., Herriott, J. R., & Horvath, S. J. (1990) *Proteins* **7,** 215.

Kline, A. D., Braun, W., & Wüthrich, K. (1988) *J. Mol. Biol.* **204**, 675.

Kopple, K. D., Wiley, G. R., & Tauke, R. (1973) *Biopolymers* **12,** 627.

Kraulis, P. (1991) *J. Appl. Cryst.* **24,** 946.

Kruse, N., Lehrnbecher, T., & Sebald, W. (1991) *FEBS Lett.* **286,** 58.

Kuntz, I. D., Thomason, J. F., & Oshiro, C. M. (1989) *Methods in Enzymol.* **177,** 159.

Le, H. V., Ramanthan, L., Labdon, J. E., Mays-Ichino, C. A., Syto, R., Arai, N., Hoy, P., Takebe, Y., Nagabhushsn, T. L., & Trotta, P. P. (1988) *J. Biol. Chem.* **263,** 10817.

Le, H. V., Seelig, G. F., Syto, R., Ramanathan, L., Windsor, W. T., Borkowski, D., & Trotta, P. P. (1991) *Biochemistry* **30,** 9576.

Lederer, F., Glatigny, A., Bethge, P. H., Bellamy, H. D., & Mathews, F. S. (1981) *J. Mol. Biol.* **148,** 427.

Lipari, G., & Szabo, A. (1982a) *J. Amer. Chem Soc.* **104,** 4546.

Lipari, G., & Szabo, A. (1982a) *J. Amer. Chem Soc.* **104,** 4559.

Lowenthal, J. W., Castle, B. E., Cristiansen, J., Schreurs, J., Rennick, D., Arai, N., Hoy, P., Takebe, Y., & Howard, M. (1988) *J. Immunol.* **140,** 456.

Lumb, K. J., Aplin, R. T., Radford, S. E., Archer, D. B., Jeenes, D. J., Lambert, N., MacKenzie, D. A., Dobson, C. M., & Lowe, G. (1992) *FEBS Lett.* **296,** 153.

MacArthur, M. W., Laskowski, R. A., Moss, D. S., & Thornton, J. M. (1992) PROCHECK program.

McCammon, J. A., Gelin, B. R., Karplus, M., & Wolynes, P. G. (1976) *Nature* **262,** 325.

McDonald, C. C., & Phillips, W. D. (1967) *J. Amer. Chem Soc.* **89,** 6332.

McGregor, M., Islam, S., & Sternberg, M. J. E. (1987) *J. Mol. Biol.* **198,** 295.

McIntosh, L. P., & Dahlquist, F. W. (1990) *Quart. Rev. Biophys.* **23,** 1.

McKay, D. B. (1992) *Science* **257,** 412.

McKenzie, H. A., & White, F. H. (1991) *Advan. Protein Chem.* **41,** 173.

Macura, S., Huang, Y., Suter, D., & Ernst, R. R. (1981) *J. Magn. Reson.* **43,** 259.

Markley, J. L. (1989) *Methods in Enzymol.* **176,** 12.

Mason, S. A., Bentley, G. A., & McIntyre, G. J. (1984) In *Neutrons in Biology* (Schoenborn, B. P., Ed.) Plenum Press, New York.

Mathews, F. S., Argos, P., & Levine, M. (1972) *Cold Spring Harbour Symp. Quant. Biol.* **36,** 387.

Messerle, B. A., Wider, G., Otting, G., Weber, C., & Wüthrich, K. (1989) *J. Magn. Reson.* **85,** 608.

Miranker, A., Radford, S. E., Karplus, M., & Dobson, C. M. (1991) *Nature* **349,** 633.

Moore, J. M., Lepre, C. A., Gippert, G. P., Chazin, W. J., Case, D. A., & Wright, P. E. (1991) *J. Mol. Biol.* **221,** 533.

Morris, A. L., MacArthur, M. W., Hutchinson, E. G., & Thornton, J. M. (1992) *Proteins* **12,** 345.

Morrison, B. W., & Leder, P. (1992) *J. Biol. Chem.* **267,** 11957.

Mott, H. R., Driscoll, P. C., Boyd, J., Cooke, R. M., Weir, M. P., & Campbell, I. D. (1992) *Biochemistry* **31,** 7741.

Nagayama, K., & Wüthrich, K. (1981) *Eur. J. Biochem.* **115,** 653.

Neuhaus, D., Wagner, G., Vak, M., Kägi, J. H. R., & Wüthrich, K. (1985) *Eur. J. Biochem.* **151,** 257.

Nicola, N. A., & Metcalf, D. (1991) *Cell* **67**, 1.

Nilges, M., Clore, G. M., & Gronenborn, A. M. (1988) *FEBS Lett.* **229**, 317.

Nilges, M., Kuszewski, J., & Brünger, A. T. (1991) in *'Computational Aspects of the Study of Biological Macromoleculaes by Nuclear Magnetic Resonaqnce Spectroscopy'* (Ed. Hoch, J. C. et al.) Plenum Press, New York. p 451.

Ohkubo, T., Taniyama, Y., & Kikuchi, M. (1991) *J. Biochem.* **110,** 1022.

Olejniczak, E. T., Poulsen, F. M., & Dobson, C. M. (1981) *J. Amer. Chem. Soc.* **103,** 6574.

Oschkinat, H., Griesinger, C., Kraulis, P. J., Sørensen, O. W., Ernst, R. R., Gronenborn, A. M., & Clore, G. M. (1988) *Nature* **332,** 374.

Pardi, A., Billeter, M., & Wüthrich, K. (1984) *J. Mol. Biol.* **180,** 741.

Park, L. S., Friend, D., Sassenfield, H. S., & Urdal, D. L. (1987) *J. Exp. Med.* **166,** 476.

Parry, D. A. D., Minasian, E., & Leach, S. J. (1991) *J. Molec. Recog.* **4**, 63.

Parsons, M. (1988) *PhD. Thesis*, Leeds University.

Pastore, A., Atkinson, R. A., Saudek, V., & Williams, R. J. P. (1991) *Proteins* **10**, 22.

Pedersen, T. G., Sigurskjold, B. W., Andersen, K. V., Kjær M., Poulsen, F. M., Dobson, C. M. & Redfield, C. (1991) *J. Mol. Biol.* **218,** 413.

Petros, A. M., & Kopple, K. D. (1990) *Biochem. Pharmacology* **40,** 65.

Philipps, D. C. (1966) *Sci. Amer.* **215** v.5, 78-90.

Post, C. B. (1992) *J. Mol. Biol.* **224,** 1087.

Post, C. B., & Karplus, M. (1986) *J. Amer. Chem. Soc.* **108,** 1317.

Post C. B., Brooks, B. R., Karplus, M. Dobson, C. M., Artymiuk, P. J., Cheetham, J. C., & Phillips, D. C. (1986) *J. Mol. Biol.* **190,** 455.

Post C. B., Dobson, C. M., & Karplus, M. (1989) *Proteins* **5,** 337.

Powers, R., Garrett, D. S., March, C. J., Frieden, E. A., Gronenborn, A. M., & Clore, G. M. (1992a) *Science* **256,** 1673.

Powers, R., Garrett, D. S., March, C. J., Frieden, E. A., Gronenborn, A. M., & Clore, G. M. (1992b) *Biochemistry* **31,** 4334.

Presnell, S. R., & Cohen, F. E. (1989) *Proc. Natl. Acad. Sci. USA* **86,** 6592.

Radford, S. E., Evans, P. A., & Dobson, C. M. (1992) *Nature* **358**, 302.

Rance, M., Sørensen, O. W., Bodenhausen, G., Wagner, G., Ernst, R. R., & Wüthrich, K. (1983) *Biochem. Biophys. Res. Commun.* **117,** 479.

Rao, S. T., Hogle, J., & Sundaralingam, M. (1983) *Acta Cryst.* **C39,** 237.

Redfield, C., & Dobson, C. M. (1988) *Biochemistry* **27,** 122.

Redfield, C., & Dobson, C. M. (1990) *Biochemistry* **29,** 7201.

Redfield, C., Smith, L. J., Boyd, J., Lawrence, G. M. P., Edwards, R. G., Smith, R. A. G., & Dobson, C. M. (1991) *Biochemistry* **30**, 11029.

Redfield, C., Boyd, J., Smith, L. J., Smith, R. A. G., & Dobson, C. M. (1992) *Biochemistry* in press.

Richmond, T. J., & Richards, F. M. (1978) *J. Mol. Biol.* **119,** 537.

Rico, M., Santoro, J., Gonzalez, C., Bruix, M., Meira, J. L., Nieto, J. L., & Herranz, J. (1991) *J. Biomol. NMR* **1,** 283.

Roberts, I. N., Mackenzie, D. A., Jeenes, D. J., Archer, D. B., Radford, S. E., Robinson, C. V., Aplin R. T., & Dobson, C. M. (1992) Submitted to *Biotech. techniques.*

Rousset, F., De Wall Malefijit, R., Slierendregt, B., Aubry, P., Bonnefoy, J. Y., DeFrance, T., Banchereau, J., & de Vries, J. (1988) *J. Immunol.* **140**, 2625.

Scheek, R. M., van Gunsteren, W. F., & Kaptein, R. (1989) *Methods in Enzymol.* **177**, 204.

Schulz, G. E., & Schirmer, R. H. (1979) *in Principles of Protein Structure.* Springer-Verlag, New York.

Senda, T., Matsuda, S., Kurihara, H., Nakamura, K. T., Kawano. G., Shimizu, H., Mizuno, H., & Mitsui, Y. (1990) *Proc. Japan Acad.* **66 Ser B**, 77.

Senn, H., Wermer, B. , Messerle, B. A., Weber, M. C., Traber, R., & Wüthrich, K. (1989) *FEBS Lett.* **249**, 113.

Smith, K. A. (1988) *Science* **240**, 1169.

Smith, J. L., Hendrickson, W. A., Honzatko, R. B., & Sheriff, S. (1986) *Biochemistry* **25**, 5018.

Smith, L. J., Sutcliffe, M. J., Redfield, C. , & Dobson, C. M. (1991) *Biochemistry* **30**, 986.

Smith, L. J., Redfield, C., Boyd, J., Lawrence, G. M. P., Edwards, R. G., Smith, R. A. G., & Dobson, C. M. (1992a) *J. Mol. Biol.* **224**, 899.

Smith, L. J., Sutcliffe, M. J., Redfield, C. , & Dobson, C. M. (1992b) *J. Mol. Biol.* in press.

Spits, H., Yssel, H., Takebe, Y., Arai, N., Yokota, T., Lee, F., Arai, K., Banchereau, J., & de Vries, J. (1987) *J. Immunol.* **139**, 1142.

States, D. J., Haberkorn, R. A., & Ruben, D. J. (1982) *J. Magn. Reson.* **48**, 286.

Stenkamp, R. E., Sieker, L. C., & Jensen, L. H. (1983) *Acta Crystallogr. Sect. B Struct. Sci.* **39**, 697.

Strynadka, N. C. J., & James, M. N. G. (1991) *J. Mol. Biol.* **220**, 401.

Sutcliffe, M. J., & Dobson, C. M. (1991) *Proteins* **10**, 117.

Sutcliffe, M. J., Hayes, F. R. F., & Blundell, T. L. (1987) *Protein Engineering* **1**, 385.

Torchia, D. A., Sparks, S. W., & Bax, A. (1989) *Biochemistry* **28**, 5509.

Ueda, T., Yamada, H., Sakamoto, N., Abe, Y., Kawano, K., Terada, Y., & Imoto, T. (1991) *J. Biochem.* **110**, 719.

van Kimmenade, A., Bond, M. W., Schumacher, J. H., Laquoi, C., & Kastelein, R. A. (1988) *Eur. J. Biochem.* **173**, 109.

Vercelli, D., Jabara, H. H., Arai, K-I, & Geha, R. S. (1989) *J. Exp. Med.* **169**, 1295.

Wagner, G. (1990) *Progress in NMR Spec.* **22**, 101.

Wagner, G., & Wüthrich, K. (1986) *Methods in Enzymology* **131**, 307.

Wagner, G., Braun, W., Havel, T. F., Schaumann, T., Go, N., & Wüthrich, K. (1987) *J. Mol. Biol.* **196**, 611.

Weaver, L. H., & Matthews, B. W. (1987) *J. Mol. Biol.* **193**, 189.

Williams, R. J. P. (1989) *Eur. J Biochem.* **183**, 479.

Williamson, M. P., Havel, T. F., & Wüthrich, K. (1985) *J. Mol. Biol.* **182**, 295.

Windsor, W. T., Syto, R. Le, H. V., & Trotta, P. P. (1991) *Biochemistry* **30**, 1259.

Wishart, D. S., Sykes, B. D., & Richards, F. M. (1992) *Biochemistry* **31**, 1647.

Wlodawer, A., Pavlovsky, A., & Gustchina, A. (1992) *FEBS Lett.* **309**, 59.

Wright, P. E. (1989) *TIBS* **14**, 255.

Wüthrich, K., Billeter, M., & Braun, W. (1983) *J. Mol. Biol.* **169**, 949.

Wüthrich, K., Billeter, M., & Braun, W. (1984) *J. Mol. Biol.* **180,** 715.

Wüthrich, K. (1986) in *'NMR of proteins and nucleic acids'* Wiley, New York.

Wüthrich, K. (1989) *Acc. Chem. Res.* **22,** 36.

Yokota, T., Otsuka, T., Mosmann, T., Banchereau, J., Defrance, T., Blanchard, D., de Vries, J. E., Lee, F., & Arai, K. (1986) *Proc. Natl. Acad. Sci. USA* **83,** 5894.

Young, P. R. (1992a) *Current Opinion in Biotec.* in press.

Young, P. R. (1992b) *Protein Engineering* **5,** 117.

Zuiderweg, E. R. P., Henkin, J., Mollison, K. W., Carter, G. W., & Greer, J. (1988) *Proteins* **3,** 139.

APPENDIX I.

Description of the 3D NMR experiments performed on $^{15}N/^{13}C$ labelled human IL-4.

HCCH-COSY and HCCH-TOCSY.

The pulse sequences for the HCCH-COSY (Bax et al., 1990a; Kay et al., 1990) and HCCH-TOCSY (Bax et al., 1990b) experiments are shown in Figure I.1. In both these experiments, ^{1}H magnetization is transferred first from a proton to the ^{13}C nucleus to which it is directly bonded via $^{1}J_{CH}$ coupling. The resulting ^{13}C magnetization is then transferred to its ^{13}C neighbour(s) via $^{1}J_{CC}$ coupling. In the HCCH-COSY experiments a 90° ^{13}C COSY mixing pulse brings about this transfer so the magnetization is transferred only from a given ^{13}C nucleus to those ^{13}C nulcei to which it is directly bonded. In the HCCH-TOCSY experiment, there is an isotropic mixing of ^{13}C spins which results in both direct and multiple-relayed magnetization along the carbon chain. Finally, the ^{13}C magnetization is transferred back to ^{1}H via $^{1}J_{CH}$ coupling and is detected in t_3. The result is a 3D spectrum in which $^{1}H(F1)$-$^{1}H(F3)$ planes taken at a given $^{13}C(F2)$ chemical shift are similar to 2D ^{1}H-^{1}H COSY or TOCSY spectra but are edited according to the ^{13}C chemical shift of the proton at which the magnetization originates.

The planes of these 3D spectra are unsymmetrical, equivalent cross peaks not occuring on both sides of the diagonal (Clore et al., 1990; Ikura et al., 1991). This is shown in Figure I.2 which shows the cross peak pattern for the case when magnetization is transferred between protons A and B. In a plane corresponding to the chemical shift of the ^{13}C nucleus directly bonded to proton A ($^{13}C_A$) a diagonal peak is found at (F1, F3) = (δ_A, δ_A) and a cross peak at (F1, F3) = (δ_A, δ_B). The diagonal peak (F1, F3) = (δ_B, δ_B) and the cross peak (F1, F3) = (δ_B, δ_A) are however found in the plane corresponding to the chemical shift of the ^{13}C nucleus directly bonded to proton B ($^{13}C_B$). This

Figure I.1.
Pulse sequences for the 3D HCCH-COSY and HCCH-TOCSY experiments (from Clore & Gronenborn, 1991c).

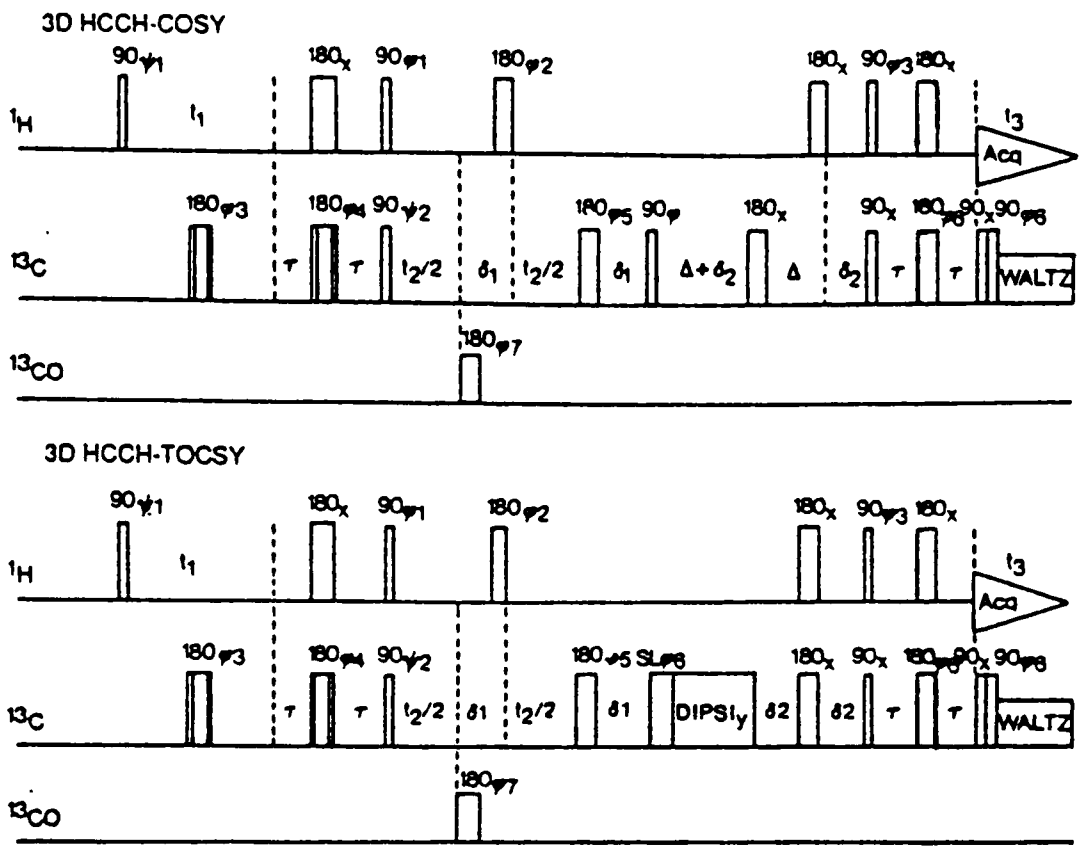


Figure I.2.
Schematic diagram of a 3D HCCH spectrum and its two slices at the $^{13}C(F2)$ chemical shifts of the nuclei C_A and C_B, showing the diagonal and cross peaks expected for a simple $^{1}H_A$-$^{13}C_A$-$^{13}C_B$-$^{1}H_B$ spin system. The diagonal peaks are represented by squares and the cross peaks by circles. (From Ikura et al., 1991).

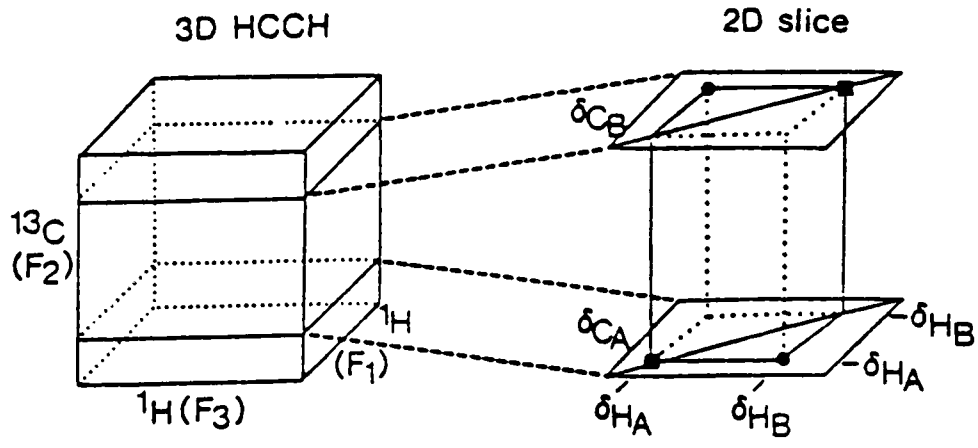
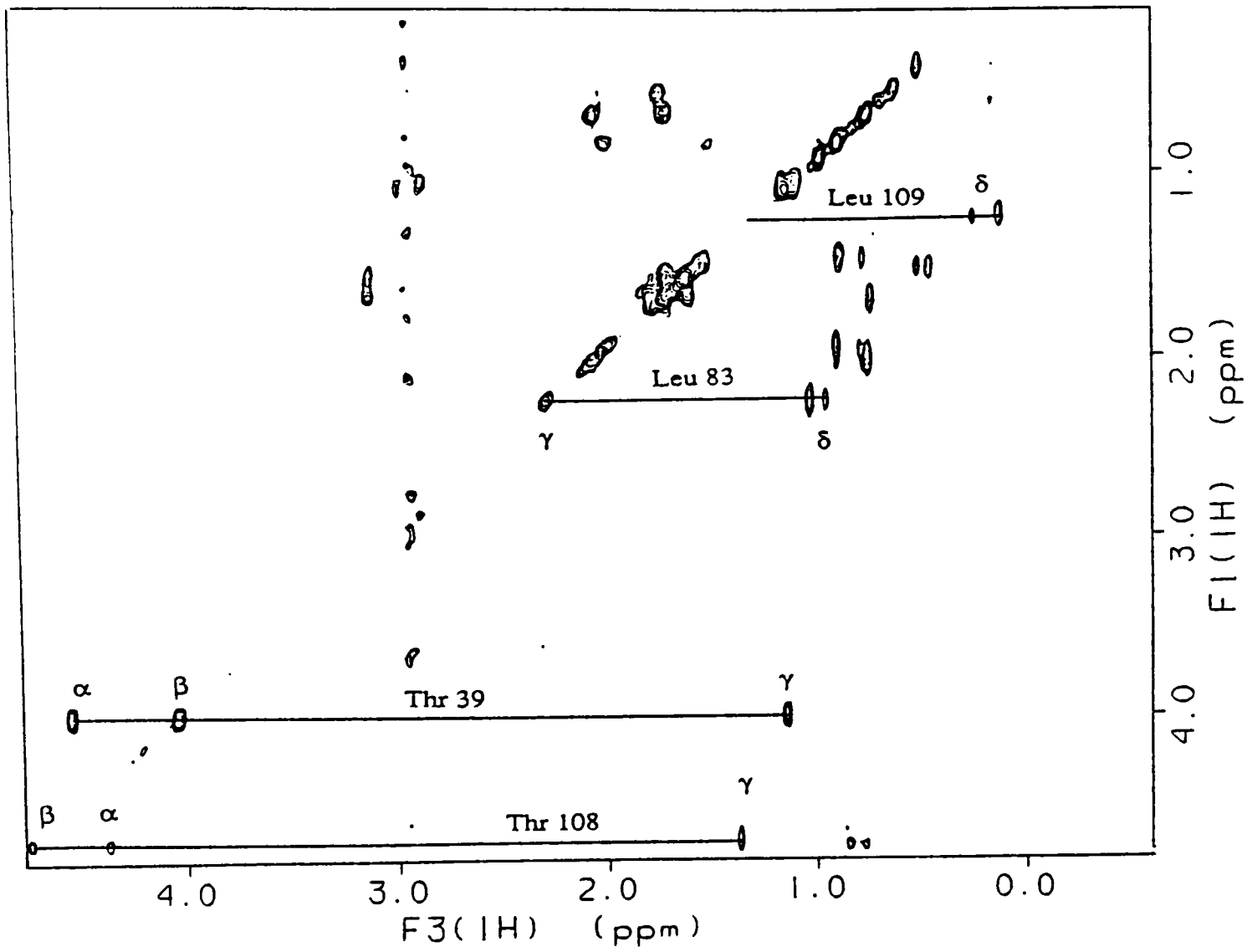


Figure I.3.
F3(^{1}H)-F1(^{1}H) plane of the HCCH-TOCSY spectrum of double labelled IL-4. Due to extensive folding in F2(^{13}C) this plane corresponds to the ^{13}C chemical shifts of 24 ppm (Leu Cγ), 46 ppm and 68 ppm (Thr β).

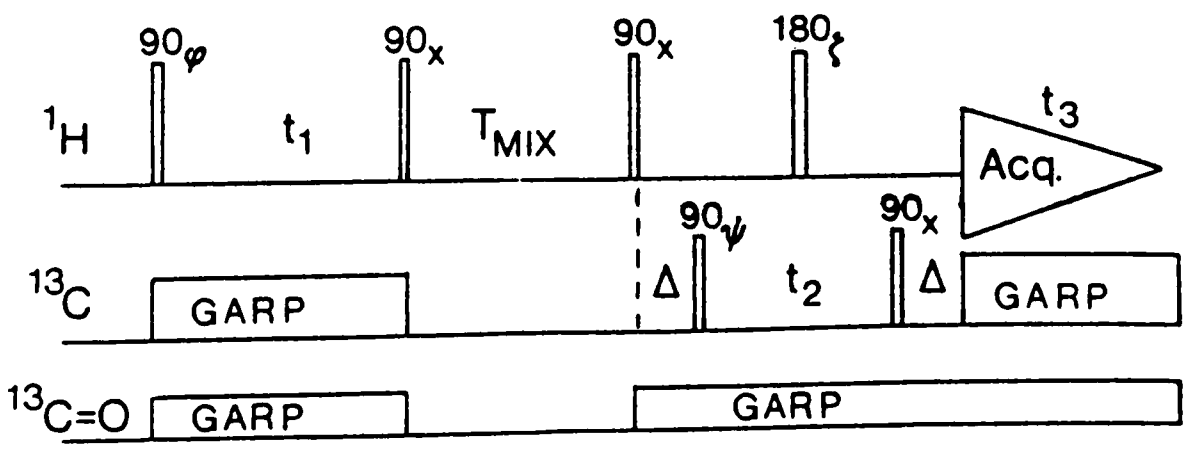


property of the spectrum means that the assignment of the cross peak in the slice corresponding to the chemical shift of $^{13}C_A$ can be confirmed by looking for the corresponding cross peak in a slice taken at the chemical shift of $^{13}C_B$. In addition, it means that the spectrum can be folded extensively in the ^{13}C(F2) dimension, so improving the overall digital resolution, without introducing ambiguities. This is illustrated in Figure I.3 for the HCCH-TOCSY spectrum of human IL-4. Both cross peaks from Thr Hβ (residues 39 and 108) and Leu Hγ (residues 83 and 109) are observed in the same F2(^{13}C) slice, the spectral folding meaning that the ^{13}C chemical shifts of both the Cβ of Thr 39 and 108 and the Cγ of Leu 83 and 109 occur on this plane.

^{13}C-^{1}H NOESY-HMQC.

The pulse sequence for the NOESY-HMQC experiment (Ikura et al., 1990b) is shown in Figure I.4. In this experiment, ^{1}H chemical shifts evolve during t_1. During the mixing time, τ_{mix}, magnetization is transferred between protons via through space NOE effects. Magnetization is then transferred to the ^{13}C spins that are directly bonded to the protons by means of multiple quantum coherence. ^{13}C chemical shifts evolve during t_2 and magnetization is then transferred back from the ^{13}C to the directly bonded ^{1}H via multiple quantum coherence and is detected during the acquisition period, t_3.

Figure I.4.
Pulse sequence for the 3D NOESY-HMQC experiment (From Ikura et al., 1990).



The resulting spectrum is again a cube in which the NOE's appear in different ^{1}H-^{1}H planes determined by the ^{13}C chemical shift of the carbon atom attached to the destination proton. As for the HCCH-COSY and HCCH-TOCSY experiments, the planes are unsymmetrical the cross peaks corresponding to an NOE between protons A and B occuring on two different ^{13}C planes (one at chemical shift of $^{13}C_A$, the other at chemical shift of $^{13}C_B$). This property is particularly important for solving ambiguities in the NOE assignment as discussed for human IL-4 in section 5.3.

APPENDIX II

PROCHECK analysis of the NMR structures of hen lysozyme.

The ensemble of 16 calculated NMR structures of hen lysozyme, the average energy minimized NMR structure and the tetragonal type 2 crystal structure of hen lysozyme have been analysed using the program PROCHECK (MacArthur et al., 1992) . This program assesses the stereochemical quality of a protein structure in terms of the following parameters (the abbreviations in brackets correspond to those used in Table 1):

a) Stereochemistry of main chain.

Ramachandran plot quality measured by % of residues that are in the most favoured regions of the Ramachandran plot (% in core).

Peptide bond planarity measured by the standard deviation of the protein structure's ω torsion angles from 180° (ω angle σ).

Bad non-bonded interactions measured by number of bad contacts per 100 residues (bad contacts).

Cα tetrahedral distortion measured by calculating the standard deviation of the zeta torsion angle (defined by Cα, N, C and Cβ) (ζ angle σ).

Hydrogen bond energy measured by the standard deviation of the hydrogen bond energies for main chain atoms, calculated using the modified method of Kabsch & Sander (1983), from expected value of -2.03 ± 0.75 kcal/mol (H bond energy).

b) Stereochemistry of side chains.

$\chi 1$ distributions measured by four parameters:

Standard deviation of the $\chi 1$ gauche minus torsion angles around ideal value of +64.1° ($\chi 1$ g- σ).

Standard deviation of the $\chi 1$ trans torsion angles around ideal value of 183.6° ($\chi 1$ t σ).

Standard deviation of the $\chi 1$ gauche plus torsion angles around ideal value of -66.7° ($\chi 1$ g+ σ).

Pooled standard deviation of all $\chi 1$ torsion angles (weighted average of data for 3 conformations). ($\chi 1$ all σ).

$\chi 2$ distribution (for residues in trans conformation) measured by standard deviation of the $\chi 2$ trans torsion angles from ideal value ($\chi 2$ t σ).

Table 1. PROCHECK parameters for hen lysozyme.

a) Stereochemistry of main chain.

Structure[1]	% in core	ω angle σ	bad contacts	ζ angle σ	H bond energy σ
NMR 2	53.1	0.3	5.4	0.4	0.8
NMR 3	51.3	0.4	6.2	0.4	0.8
NMR 4	60.2	0.3	5.4	0.3	0.7
NMR 6	51.3	0.3	10.1	0.3	0.7
NMR 7	54.0	0.3	9.3	0.3	0.7
NMR 8	52.2	0.3	7.0	0.3	0.7
NMR 9	52.2	0.3	5.4	0.4	0.8
NMR 10	61.1	0.3	4.7	0.4	0.8
NMR 11	61.9	0.4	7.8	0.4	0.8
NMR 12	53.1	0.4	3.1	0.4	0.8
NMR 13	54.0	0.3	7.0	0.4	0.8
NMR 14	59.3	0.4	7.0	0.4	0.8
NMR 17	54.0	0.3	3.1	0.3	0.8
NMR 18	54.0	0.4	8.5	0.3	0.8
NMR 19	47.8	0.3	5.4	0.4	0.7
NMR 20	52.2	0.3	7.0	0.4	0.7
Average NMR	62.8	1.5	4.7	0.3	0.7
Crystal	89.4	1.9	7.0	1.9	0.7
Typical 2.0 Å	83.8	6.0	4.2	3.1	0.8
Band width	10.0	3.0	3.1	1.6	0.2

b) Stereochemistry of side chains

Structure	χ_1 g- σ	χ_1 t σ	χ_1 g+ σ	χ_1 all σ	χ_2 t σ
NMR 2	12.4	7.0	6.6	8.8	21.8
NMR 3	9.8	11.8	8.1	10.9	8.4
NMR 4	8.1	11.6	12.6	11.8	17.9
NMR 6	9.8	8.0	9.9	9.9	12.2
NMR 7	7.7	8.0	12.9	10.1	14.4
NMR 8	6.7	7.9	11.5	10.9	20.8
NMR 9	10.2	6.1	12.9	10.4	12.6
NMR 10	10.3	9.6	11.6	11.4	10.8
NMR 11	10.7	15.8	9.3	13.4	14.2
NMR 12	7.9	11.7	9.3	10.8	15.1
NMR 13	16.0	9.5	11.6	12.0	14.1
NMR 14	5.0	10.9	9.3	10.3	18.8
NMR 17	7.0	11.8	9.3	10.9	18.5
NMR 18	5.8	10.8	9.0	10.5	14.9
NMR 19	8.8	6.6	12.6	10.3	15.4
NMR 20	9.3	9.8	11.2	11.8	14.5
Average NMR	9.9	14.5	14.2	14.6	17.6
Crystal	7.0	16.3	13.5	13.6	19.1
Typical 2.0Å	18.1	19.0	17.5	18.2	20.4
Band width	6.5	5.3	4.9	4.8	5.0

[1] The values of the parameters for the 16 individual NMR structures are listed, followed by those for the average energy minimized NMR structure and the tetragonal type 2 crystal structure of hen lysozyme (2.0 Å resolution). For comparison the means and standard deviations of these parameters for all the 2.0Å resolution crystal structures in the Brookhaven Data Bank are given.

Summary of Contents of Appendices of Structural Data.

Appendix III.

Listing of NOE's identified for hen lysozyme and used in the structure calculations. The final column gives a measure of the intensity of the NOE's: 1, strong; 2, medium; 3, weak (intensities from 50 ms mixing time NOESY); 4, cross peaks observed in NOESY spectra recorded with 200 ms and 500 ms mixing times (but not in 50 ms NOESY); 5, cross peaks only observed in NOESY spectra recorded with a 500 ms mixing time. Corresponding distance restraints for these intensity categories are: 1, 1.8-2.5 Å; 2, 1.8-3.0 Å; 3, 1.8-4.5 Å; 4, 1.8-5.5 Å, 5, 1.8-7.5 Å (with additional pseudoatom corrections etc. where necessary).

Appendix IV.

Listing of the ϕ angle restraints (from $^3J_{HN\alpha}$ coupling constants), the χ_1 angle restraints (from $^3J_{\alpha\beta}$ coupling constants) and the hydrogen bond restraints (from amide proton exchange rates) used in the structure calculations for hen lysozyme. Data is in standard XPLOR format.

Appendix V.

$^3J_{HN\alpha}$ coupling constants (in Hz) for human IL-4. For some residues with a small coupling constant, an accurate value has not been able to be measured; in these cases an upper limit for the value is given.

Appendix VI.

Listing of NOE's identified for human IL-4. First NOE's that were identified in the spectra of unlabelled and ^{15}N labelled IL-4 are listed (data used in the preliminary structure calculations) and then the additional NOE's identified in the spectra of double labelled IL-4 are given (full list used in the final calculations). The final column gives a measure of the intensity of the NOE's: 1, very strong; 2, strong; 3, medium ; 4, weak; 5, very weak (for cross peaks only observed in NOESY spectra recorded with a 200 ms mixing time). Corresponding distance restraints for these intensity categories are: 1, 1.8-2.5 Å; 2, 1.8-3.0 Å; 3, 1.8-4.0 Å; 4, 1.8-5.0 Å, 5, 1.8-6.0 Å (with additional pseudoatom corrections etc. where necessary).

Appendix VII.

Listing of the ϕ angle restraints and hydrogen bond restraints used in the structure calculations for human IL-4. At the end of the Appendix the alternative ϕ and ψ angle restraints, which were used in the preliminary structure calculations with tightly restrained helices, are given. Data in standard XPLOR format.

NOE's identified for hen lysozyme.

1	LYS	HA	2	VAL	HN	1	4	GLY	HA#
2	VAL	HN	2	VAL	HA	3	4	GLY	HN
2	VAL	HA	2	VAL	HB	2	4	GLY	HN
2	VAL	HA	2	VAL	HG2#	1	4	GLY	HA#
2	VAL	HA	2	VAL	HG1#	1	4	GLY	HA#
2	VAL	HA	3	PHE	HN	1	4	GLY	HN
2	VAL	HA	40	THR	HN	3	5	ARG	HN
2	VAL	HA	39	ASN	HA	1	6	CYS	HA
2	VAL	HG2#	39	ASN	HA	2	6	CYS	HA
2	VAL	HG1#	39	ASN	HA	3	6	CYS	HA
2	VAL	HN	2	VAL	HB	2	6	CYS	HA
2	VAL	HN	2	VAL	HG2#	2	6	CYS	HN
2	VAL	HN	2	VAL	HG1#	3	6	CYS	HN
2	VAL	HG2#	3	PHE	HN	3	6	CYS	HB1
2	VAL	HG1#	3	PHE	HN	2	6	CYS	HA
2	VAL	HG2#	38	PHE	HN	3	6	CYS	HN
2	VAL	HG1#	38	PHE	HN	2	6	CYS	HB1
2	VAL	HG2#	40	THR	HN	3	6	CYS	HN
2	VAL	HB	2	VAL	HG2#	2	7	GLU	HN
2	VAL	HB	2	VAL	HG1#	2	7	GLU	HN
2	VAL	HN	40	THR	HN	4	7	GLU	HA
2	VAL	HG1#	40	THR	HN	4	7	GLU	HA
3	PHE	HN	3	PHE	HA	3	7	GLU	HA
3	PHE	HA	3	PHE	HB2	3	7	GLU	HA
3	PHE	HA	3	PHE	HB1	2	7	GLU	HB#
3	PHE	HN	3	PHE	HB2	2	7	GLU	HB#
3	PHE	HN	3	PHE	HB1	3	7	GLU	HN
3	PHE	HZ	3	PHE	HE*	2	8	LEU	HN
3	PHE	HD*	3	PHE	HE*	1	8	LEU	HN
3	PHE	HB2	3	PHE	HD*	3	8	LEU	HA
3	PHE	HB1	3	PHE	HD*	2	8	LEU	HA
3	PHE	HD*	8	LEU	HN	3	8	LEU	HN
3	PHE	HN	3	PHE	HD*	3	8	LEU	HN
3	PHE	HD*	4	GLY	HN	3	8	LEU	HA
3	PHE	HD*	40	THR	HN	3	8	LEU	HA
3	PHE	HN	39	ASN	HA	3	8	LEU	HN
3	PHE	HB1	8	LEU	HN	3	8	LEU	HN
3	PHE	HD*	7	GLU	HB#	2	8	LEU	HA
3	PHE	HD*	8	LEU	HD*	3	8	LEU	HB#
3	PHE	HZ	8	LEU	HD*	3	8	LEU	HD*
3	PHE	HB2	8	LEU	HD*	3	8	LEU	HD*
3	PHE	HD*	7	GLU	HB#	3	8	LEU	HB#
3	PHE	HD*	88	ILE	HG2#	3	8	LEU	HD*
3	PHE	HE*	88	ILE	HG2#	2	8	LEU	HD*
3	PHE	HZ	88	ILE	HG2#	1	8	LEU	HD*
3	PHE	HE*	55	ILE	HB	2	8	LEU	HD*
3	PHE	HD*	55	ILE	HB	2	8	LEU	HG
3	PHE	HE*	39	ASN	HA	5	8	LEU	HD*
3	PHE	HN	39	ASN	HA	4	9	ALA	HN
3	PHE	HN	40	THR	HN	4	9	ALA	HN
3	PHE	HB2	38	PHE	HB#	2	9	ALA	HA
3	PHE	HB2	88	ILE	HD#	4	9	ALA	HA
3	PHE	HN	4	GLY	HA#	4	9	ALA	HN
3	PHE	HE*	40	THR	HN	4	9	ALA	HN
3	PHE	HE*	88	ILE	HD#	4	9	ALA	HN
3	PHE	HD*	88	ILE	HD#	4	9	ALA	HN

4	GLY	HA#	1
4	GLY	HA#	3
4	GLY	HA#	3
5	ARG	HN	3
5	ARG	HN	3
7	GLU	HB#	2
5	ARG	HA	2
6	CYS	HB2	3
6	CYS	HB1	2
9	ALA	HN	3
10	ALA	HN	3
6	CYS	HB2	2
6	CYS	HB1	3
127	CYS	HA	3
9	ALA	HB#	1
6	CYS	HA	3
7	GLU	HN	3
7	GLU	HN	2
7	GLU	HA	2
7	GLU	HB#	2
7	GLU	HB#	2
11	ALA	HN	3
10	ALA	HN	3
10	ALA	HB#	3
8	LEU	HN	3
8	LEU	HB#	3
8	LEU	HD*	1
8	LEU	HD*	3
8	LEU	HD*	3
11	ALA	HN	3
9	ALA	HN	3
9	ALA	HN	2
10	ALA	HN	3
11	ALA	HB#	2
9	ALA	HN	3
12	MET	HN	3
9	ALA	HN	3
9	ALA	HA	1
12	MET	HE#	1
88	ILE	HD#	3
88	ILE	HG2#	3
17	LEU	HG	1
8	LEU	HD*	2
38	PHE	HE*	1
9	ALA	HA	2
9	ALA	HB#	1
9	ALA	HB#	1
12	MET	HN	3
10	ALA	HN	3
10	ALA	HN	2
11	ALA	HN	3
129	LEU	HD*	3

9	ALA	HB#	124	ILE	HG2#	1	15	HIS	HE1	92	VAL	HG1#	3
10	ALA	HN	10	ALA	HA	2	15	HIS	HB2	92	VAL	HG2#	3
10	ALA	HN	10	ALA	HB#	1	15	HIS	HB2	92	VAL	HG1#	2
10	ALA	HA	10	ALA	HB#	1	15	HIS	HN	88	ILE	HD#	5
10	ALA	HA	12	MET	HN	3	16	GLY	HN	16	GLY	HA#	3
10	ALA	HA	11	ALA	HN	3	16	GLY	HN	16	GLY	HA#	3
10	ALA	HN	11	ALA	HN	2	16	GLY	HA#	16	GLY	HA#	1
10	ALA	HN	12	MET	HN	3	16	GLY	HN	17	LEU	HD*	5
10	ALA	HN	129	LEU	HD*	2	16	GLY	HN	17	LEU	HN	3
10	ALA	HA	129	LEU	HD*	1	17	LEU	HN	17	LEU	HA	2
11	ALA	HN	11	ALA	HA	2	17	LEU	HN	17	LEU	HB#	3
11	ALA	HN	11	ALA	HB#	1	17	LEU	HA	17	LEU	HB#	2
11	ALA	HA	11	ALA	HB#	1	17	LEU	HA	17	LEU	HD*	3
11	ALA	HA	12	MET	HN	2	17	LEU	HA	17	LEU	HD*	3
11	ALA	HN	12	MET	HN	2	17	LEU	HN	17	LEU	HG	1
11	ALA	HN	13	LYS	HN	3	17	LEU	HG	17	LEU	HD*	2
11	ALA	HB#	12	MET	HN	1	17	LEU	HD*	28	TRP	HZ2	3
11	ALA	HB#	13	LYS	HN	3	17	LEU	HD*	28	TRP	HE3	3
11	ALA	HA	14	ARG	HN	3	17	LEU	HD*	28	TRP	HH3	3
11	ALA	HB#	88	ILE	HD#	1	17	LEU	HD*	18	ASP	HN	3
11	ALA	HN	88	ILE	HD#	4	17	LEU	HD*	20	TYR	HB2	3
12	MET	HN	12	MET	HA	2	17	LEU	HD*	28	TRP	HE1	3
12	MET	HA	13	LYS	HN	3	17	LEU	HA	18	ASP	HN	2
12	MET	HN	13	LYS	HN	2	17	LEU	HN	18	ASP	HN	2
12	MET	HA	15	HIS	HB2	3	17	LEU	HG	18	ASP	HN	3
12	MET	HN	88	ILE	HD#	3	17	LEU	HD*	28	TRP	HZ3	3
12	MET	HA	88	ILE	HD#	2	17	LEU	HD*	92	VAL	HG1#	2
12	MET	HE#	88	ILE	HD#	3	17	LEU	HD*	92	VAL	HG1#	2
12	MET	HE#	17	LEU	HD*	2	17	LEU	HD*	92	VAL	HG2#	2
12	MET	HA	17	LEU	HD*	3	17	LEU	HD*	92	VAL	HG2#	3
12	MET	HA	17	LEU	HG	3	17	LEU	HB#	28	TRP	HE3	3
12	MET	HN	17	LEU	HD*	4	17	LEU	HD*	28	TRP	HD1	3
12	MET	HN	14	ARG	HN	4	17	LEU	HB#	17	LEU	HD*	3
13	LYS	HN	13	LYS	HA	2	17	LEU	HB#	17	LEU	HG	1
13	LYS	HN	13	LYS	HB#	2	17	LEU	HD*	18	ASP	HN	4
13	LYS	HA	13	LYS	HB#	2	17	LEU	HD*	19	ASN	HN	5
13	LYS	HN	14	ARG	HN	2	17	LEU	HD*	20	TYR	HD*	3
13	LYS	HA	14	ARG	HN	3	17	LEU	HD*	20	TYR	HB2	2
13	LYS	HN	129	LEU	HD*	3	17	LEU	HB#	28	TRP	HZ3	4
13	LYS	HB#	14	ARG	HN	2	17	LEU	HD*	92	VAL	HB	4
13	LYS	HN	25	LEU	HD*	2	17	LEU	HD*	88	ILE	HD#	5
13	LYS	HN	88	ILE	HD#	4	17	LEU	HD*	96	LYS	HN	4
14	ARG	HN	14	ARG	HA	3	17	LEU	HA	28	TRP	HE1	4
14	ARG	HN	14	ARG	HB#	2	17	LEU	HD*	20	TYR	HE*	5
14	ARG	HN	14	ARG	HB#	2	17	LEU	HD*	55	ILE	HG2#	5
14	ARG	HN	15	HIS	HN	2	18	ASP	HN	18	ASP	HA	3
15	HIS	HN	15	HIS	HA	3	18	ASP	HN	18	ASP	HB2	2
15	HIS	HN	15	HIS	HB2	2	18	ASP	HN	18	ASP	HB1	2
15	HIS	HN	15	HIS	HB1	3	18	ASP	HA	18	ASP	HB2	2
15	HIS	HA	15	HIS	HB2	2	18	ASP	HA	18	ASP	HB1	3
15	HIS	HN	16	GLY	HN	3	18	ASP	HA	19	ASN	HN	1
15	HIS	HB2	92	VAL	HG1#	2	18	ASP	HN	25	LEU	HD*	3
15	HIS	HB2	92	VAL	HG2#	3	18	ASP	HN	28	TRP	HE1	5
15	HIS	HE1	88	ILE	HG1#	3	18	ASP	HN	25	LEU	HN	4
15	HIS	HE1	88	ILE	HD#	2	18	ASP	HA	26	GLY	HN	4
15	HIS	HE1	92	VAL	HG2#	3	18	ASP	HA	28	TRP	HE1	4

19	ASN	HN	19	ASN	HA	2	23	TYR	HN	24	SER	HN	4
19	ASN	HN	19	ASN	HB#	3	24	SER	HN	24	SER	HA	2
19	ASN	HN	19	ASN	HB#	3	25	LEU	HN	25	LEU	HA	2
19	ASN	HA	19	ASN	HB#	2	25	LEU	HN	25	LEU	HB#	2
19	ASN	HA	19	ASN	HB#	2	25	LEU	HN	26	GLY	HN	3
19	ASN	HN	20	TYR	HN	2	25	LEU	HN	25	LEU	HG	2
19	ASN	HN	24	SER	HN	5	25	LEU	HN	25	LEU	HD*	3
19	ASN	HN	28	TRP	HE1	4	25	LEU	HN	25	LEU	HD*	3
20	TYR	HN	20	TYR	HB2	2	25	LEU	HA	25	LEU	HB#	3
20	TYR	HN	20	TYR	HB1	2	25	LEU	HB#	25	LEU	HD*	2
20	TYR	HA	20	TYR	HB2	1	25	LEU	HB#	26	GLY	HN	2
20	TYR	HA	20	TYR	HB1	3	25	LEU	HA	26	GLY	HN	3
20	TYR	HB2	20	TYR	HD*	2	25	LEU	HD*	28	TRP	HD1	2
20	TYR	HB1	20	TYR	HD*	2	25	LEU	HN	28	TRP	HE1	5
20	TYR	HA	20	TYR	HD*	3	25	LEU	HN	28	TRP	HE1	4
20	TYR	HN	20	TYR	HA	3	25	LEU	HA	28	TRP	HE1	4
20	TYR	HD*	21	ARG	HN	3	26	GLY	HN	26	GLY	HA#	3
20	TYR	HN	28	TRP	HE1	3	26	GLY	HN	26	GLY	HA#	3
20	TYR	HA	21	ARG	HN	1	26	GLY	HA#	26	GLY	HA#	1
20	TYR	HE*	96	LYS	HB#	3	26	GLY	HN	27	ASN	HN	2
20	TYR	HN	21	ARG	HN	4	26	GLY	HN	121	VAL	HA	3
20	TYR	HN	25	LEU	HN	4	26	GLY	HN	29	VAL	HG*	3
20	TYR	HE*	21	ARG	HN	4	26	GLY	HN	120	VAL	HG*	3
21	ARG	HN	21	ARG	HA	1	26	GLY	HN	111	TRP	HE1	5
21	ARG	HA	21	ARG	HB#	2	26	GLY	HN	121	GLN	HN	5
21	ARG	HA	22	GLY	HN	3	26	GLY	HN	30	GLY	HN	5
21	ARG	HN	22	GLY	HN	2	26	GLY	HN	29	VAL	HN	4
21	ARG	HN	22	GLY	HA#	4	26	GLY	HN	28	TRP	HN	4
21	ARG	HA	23	TYR	HN	5	27	ASN	HN	27	ASN	HA	3
21	ARG	HN	28	TRP	HE1	5	27	ASN	HN	27	ASN	HB2	2
21	ARG	HN	23	TYR	HN	5	27	ASN	HN	27	ASN	HB1	3
22	GLY	HN	22	GLY	HA#	3	27	ASN	HA	27	ASN	HB2	2
22	GLY	HN	22	GLY	HA#	3	27	ASN	HA	27	ASN	HB1	2
22	GLY	HA#	22	GLY	HA#	1	27	ASN	HB2	28	TRP	HN	2
22	GLY	HN	28	TYR	HE1	4	27	ASN	HA	30	CYS	HN	3
23	TYR	HN	23	TYR	HA	3	27	ASN	HN	28	TRP	HN	2
23	TYR	HN	23	TYR	HB2	3	27	ASN	HA	30	CYS	HB2	3
23	TYR	HN	23	TYR	HB1	3	27	ASN	HN	120	VAL	HG*	3
23	TYR	HA	23	TYR	HB1	2	27	ASN	HA	111	TRP	HE1	2
23	TYR	HA	23	TYR	HD*	3	27	ASN	HA	123	TRP	HE1	5
23	TYR	HE	111	TRP	HH3	3	27	ASN	HN	111	TRP	HE1	4
23	TYR	HB2	23	TYR	HD*	2	27	ASN	HB1	111	TRP	HE1	2
23	TYR	HB1	23	TYR	HD*	2	27	ASN	HA	111	TRP	HE1	2
23	TYR	HB1	28	TRP	HE1	3	27	ASN	HN	28	TRP	HE1	5
23	TYR	HD*	24	SER	HN	3	28	TRP	HN	28	TRP	HA	2
23	TYR	HD*	111	TRP	HZ2	3	28	TRP	HA	28	TRP	HB#	3
23	TYR	HD*	105	MET	HB#	3	28	TRP	HA	28	TRP	HB#	3
23	TYR	HE*	111	TRP	HZ2	2	28	TRP	HB#	29	VAL	HN	3
23	TYR	HE*	105	MET	HB#	3	28	TRP	HN	29	VAL	HN	2
23	TYR	HD*	28	TRP	HE1	3	28	TRP	HH3	28	TRP	HZ2	3
23	TYR	HN	28	TRP	HZ2	2	28	TRP	HE3	28	TRP	HZ3	1
23	TYR	HE*	105	MET	HB#	3	28	TRP	HZ3	28	TRP	HH3	2
23	TYR	HE*	111	TRP	HE1	4	28	TRP	HE3	28	TRP	HA	2
23	TYR	HE*	99	VAL	HG	2	28	TRP	HE3	28	TRP	HB#	3
23	TYR	HN	28	TRP	HE1	4	28	TRP	HD1	28	TRP	HB#	3
23	TYR	HE*	104	GLY	HN	4	28	TRP	HZ3	56	LEU	HG	2

28	TRP	HE1	28	TRP	HB#	2	34	PHE	HN	34	PHE	HB2	2
28	TRP	HE1	28	TRP	HZ2	2	34	PHE	HN	34	PHE	HB1	3
28	TRP	HE1	28	TRP	HN	3	34	PHE	HA	34	PHE	HB2	1
28	TRP	HE1	28	TRP	HD1	1	34	PHE	HA	34	PHE	HB1	2
28	TRP	HE3	56	LEU	HD*	2	34	PHE	HN	35	GLU	HN	2
28	TRP	HE3	32	ALA	HB#	3	34	PHE	HZ	34	PHE	HE*	2
28	TRP	HH3	95	ALA	HB#	3	34	PHE	HD*	34	PHE	HE*	2
28	TRP	HE3	95	ALA	HB#	3	34	PHE	HB2	34	PHE	HD*	2
28	TRP	HZ3	95	ALA	HB#	2	34	PHE	HB1	34	PHE	HD*	2
28	TRP	HZ3	108	TRP	HE1	5	34	PHE	HA	34	PHE	HD*	2
28	TRP	HZ3	98	ILE	HG2	5	34	PHE	HD*	34	PHE	HB2	2
28	TRP	HH3	99	VAL	HG*	2	34	PHE	HE*	114	ARG	HB#	3
28	TRP	HH3	99	VAL	HB	2	34	PHE	HE*	123	TRP	HZ2	3
28	TRP	HZ2	99	VAL	HG*	2	35	GLU	HN	35	GLU	HA	3
28	TRP	HZ3	99	VAL	HN	4	35	GLU	HN	35	GLU	HB#	1
28	TRP	HE3	88	ILE	HD#	5	35	GLU	HA	35	GLU	HB#	3
29	VAL	HN	29	VAL	HA	3	35	GLU	HN	36	SER	HN	2
29	VAL	HA	29	VAL	HB	3	35	GLU	HB#	38	PHE	HE*	3
29	VAL	HA	29	VAL	HG*	1	35	GLU	HN	38	PHE	HE*	3
29	VAL	HN	29	VAL	HG*	1	36	SER	HN	36	SER	HA	3
29	VAL	HB	29	VAL	HG*	2	36	SER	HN	36	SER	HB#	3
29	VAL	HA	30	CYS	HN	3	36	SER	HA	36	SER	HB#	2
29	VAL	HN	30	CYS	HN	2	36	SER	HA	37	ASN	HN	3
29	VAL	HB	30	CYS	HN	2	36	SER	HB#	39	ASN	HN	3
29	VAL	HG*	30	CYS	HN	2	36	SER	HN	37	ASN	HN	2
30	CYS	HN	30	CYS	HA	2	36	SER	HB#	55	ILE	HA	2
30	CYS	HN	30	CYS	HB2	2	36	SER	HA	42	ALA	HB#	3
30	CYS	HN	30	CYS	HB1	2	37	ASN	HN	37	ASN	HA	1
30	CYS	HB2	31	ALA	HN	3	37	ASN	HA	37	ASN	HB#	2
30	CYS	HB1	31	ALA	HN	2	37	ASN	HA	37	ASN	HB#	2
30	CYS	HN	120	VAL	HG*	2	37	ASN	HA	38	PHE	HN	3
30	CYS	HB1	123	TRP	HE1	4	37	ASN	HN	38	PHE	HN	3
31	ALA	HN	31	ALA	HA	2	37	ASN	HA	39	ASN	HN	3
31	ALA	HA	31	ALA	HB#	1	38	PHE	HN	38	PHE	HA	1
31	ALA	HA	32	ALA	HN	3	38	PHE	HN	38	PHE	HB#	3
31	ALA	HA	34	PHE	HN	3	38	PHE	HA	38	PHE	HB#	1
31	ALA	HN	32	ALA	HN	3	38	PHE	HB#	38	PHE	HD*	3
31	ALA	HN	111	TRP	HE1	4	38	PHE	HB#	38	PHE	HD*	3
32	ALA	HN	32	ALA	HA	2	39	ASN	HN	39	ASN	HA	2
32	ALA	HN	32	ALA	HB#	1	39	ASN	HN	39	ASN	HB2	2
32	ALA	HA	32	ALA	HB#	1	39	ASN	HN	39	ASN	HB1	2
32	ALA	HA	33	LYS	HN	3	39	ASN	HA	39	ASN	HB2	2
32	ALA	HA	35	GLU	HB#	2	39	ASN	HA	39	ASN	HB1	3
32	ALA	HN	33	LYS	HN	3	39	ASN	HA	41	GLN	HN	3
32	ALA	HB#	33	LYS	HN	2	39	ASN	HB1	42	ALA	HN	3
32	ALA	HB#	38	PHE	HE*	2	39	ASN	HA	40	THR	HN	1
32	ALA	HB#	35	GLU	HN	3	39	ASN	HA	41	GLN	HN	4
32	ALA	HB#	38	PHE	HD*	1	39	ASN	HA	42	ALA	HN	4
32	ALA	HN	56	LEU	HD*	2	39	ASN	HN	40	THR	HN	4
33	LYS	HN	33	LYS	HA	2	39	ASN	HA	40	THR	HA	4
33	LYS	HA	34	PHE	HN	3	40	THR	HN	40	THR	HA	3
33	LYS	HN	34	PHE	HN	2	40	THR	HA	40	THR	HB	1
33	LYS	HA	38	PHE	HE*	3	40	THR	HA	40	THR	HG2#	1
33	LYS	HN	38	PHE	HE*	3	40	THR	HN	41	GLN	HN	2
33	LYS	HN	38	PHE	HD*	3	40	THR	HA	55	ILE	HN	3
34	PHE	HN	34	PHE	HA	3	40	THR	HA	42	ALA	HN	3

40	THR	HN	40	THR	HG2#	1
40	THR	HB	40	THR	HG2#	1
40	THR	HN	42	ALA	HN	4
41	GLN	HN	41	GLN	HA	3
41	GLN	HN	42	ALA	HN	2
42	ALA	HN	42	ALA	HA	2
42	ALA	HN	42	ALA	HB#	1
42	ALA	HA	42	ALA	HB#	1
42	ALA	HA	43	THR	HN	1
42	ALA	HB#	43	THR	HN	1
42	ALA	HB#	43	THR	HA	3
43	THR	HN	43	THR	HA	3
43	THR	HA	43	THR	HB	1
43	THR	HA	44	ASN	HN	1
43	THR	HA	54	GLY	HN	3
43	THR	HA	53	TYR	HA	2
43	THR	HN	43	THR	HB	3
43	THR	HN	43	THR	HG2#	3
43	THR	HB	43	THR	HG2#	1
43	THR	HA	51	THR	HG2#	2
44	ASN	HN	44	ASN	HA	2
44	ASN	HN	44	ASN	HB#	3
44	ASN	HA	44	ASN	HB#	1
44	ASN	HN	52	ASP	HB2	3
44	ASN	HN	52	ASP	HN	2
44	ASN	HN	51	THR	HG2#	2
44	ASN	HA	45	ARG	HN	1
44	ASN	HN	45	ARG	HN	4
45	ARG	HN	45	ARG	HA	3
45	ARG	HN	45	ARG	HB#	2
45	ARG	HA	45	ARG	HB#	3
45	ARG	HA	45	ARG	HB#	3
45	ARG	HA	46	ASN	HN	2
45	ARG	HA	52	ASP	HN	3
45	ARG	HA	51	THR	HA	1
45	ARG	HA	51	THR	HG2#	2
46	ASN	HN	46	ASN	HA	3
46	ASN	HN	46	ASN	HB2	2
46	ASN	HN	46	ASN	HB1	3
46	ASN	HA	46	ASN	HB2	3
46	ASN	HA	46	ASN	HB1	3
46	ASN	HA	47	THR	HN	1
46	ASN	HB2	47	THR	HN	2
46	ASN	HB1	47	THR	HN	2
46	ASN	HB2	48	ASP	HN	2
46	ASN	HB1	48	ASP	HN	2
46	ASN	HN	51	THR	HA	3
46	ASN	HA	48	ASP	HN	4
46	ASN	HN	48	ASP	HN	5
46	ASN	HN	49	GLY	HN	4
47	THR	HN	47	THR	HA	2
47	THR	HA	47	THR	HB	1
47	THR	HB	47	THR	HG2#	1
47	THR	HN	48	ASP	HN	2
47	THR	HN	47	THR	HG2#	2
47	THR	HN	49	GLY	HN	4
48	ASP	HN	48	ASP	HA	3
48	ASP	HN	48	ASP	HB2	3
48	ASP	HN	48	ASP	HB1	3
48	ASP	HA	48	ASP	HB2	1
48	ASP	HA	48	ASP	HB1	1
49	GLY	HN	49	GLY	HA#	3
49	GLY	HN	49	GLY	HA#	3
49	GLY	HA#	49	GLY	HA#	1
49	GLY	HN	50	SER	HN	3
50	SER	HN	50	SER	HA	3
50	SER	HN	50	SER	HB#	3
50	SER	HA	51	THR	HN	3
51	THR	HN	51	THR	HA	3
51	THR	HN	51	THR	HB	3
51	THR	HA	51	THR	HB	2
51	THR	HB	51	THR	HG2#	1
51	THR	HA	51	THR	HG2#	1
51	THR	HA	52	ASP	HN	1
51	THR	HG2#	52	ASP	HN	1
51	THR	HG2#	53	TYR	HE*	2
51	THR	HG2#	53	TYR	HD*	3
51	THR	HN	51	THR	HG2#	3
51	THR	HB	52	ASP	HN	3
51	THR	HB	53	TYR	HE*	2
51	THR	HN	53	TYR	HE*	4
51	THR	HN	60	SER	HN	5
51	THR	HG2#	59	ASN	HA	4
52	ASP	HN	52	ASP	HA	3
52	ASP	HN	52	ASP	HB2	2
52	ASP	HN	52	ASP	HB1	3
52	ASP	HA	52	ASP	HB2	3
52	ASP	HA	53	TYR	HN	1
52	ASP	HA	59	ASN	HA	2
52	ASP	HA	53	TYR	HE*	3
52	ASP	HA	57	GLN	HN	4
52	ASP	HA	53	TYR	HD*	3
52	ASP	HN	53	TYR	HN	4
52	ASP	HB1	59	ASN	HA	4
53	TYR	HN	53	TYR	HB2	2
53	TYR	HN	53	TYR	HB1	3
53	TYR	HA	53	TYR	HB2	3
53	TYR	HA	53	TYR	HB1	2
53	TYR	HB2	53	TYR	HD*	2
53	TYR	HB1	53	TYR	HD*	2
53	TYR	HA	54	GLY	HN	2
53	TYR	HN	59	ASN	HA	3
53	TYR	HN	57	GLN	HA	3
53	TYR	HN	58	ILE	HN	3
53	TYR	HD*	53	TYR	HE*	1
53	TYR	HD*	53	TYR	HA	3
53	TYR	HD*	60	SER	HN	3
53	TYR	HN	53	TYR	HD*	3
53	TYR	HE*	60	SER	HN	3
53	TYR	HE*	60	SER	HB#	2
53	TYR	HE*	66	ASP	HB#	3
53	TYR	HD*	59	ASN	HA	4

53	TYR	HE*	66	ASP	HN	5	59	ASN	HN	59	ASN HB1 2
53	TYR	HN	57	GLN	HN	5	59	ASN	HA	59	ASN HB2 2
53	TYR	HN	60	SER	HN	5	59	ASN	HA	59	ASN HB1 3
54	GLY	HN	54	GLY	HA#	3	59	ASN	HA	60	SER HN 1
54	GLY	HN	54	GLY	HA#	3	59	ASN	HB2	60	SER HN 3
54	GLY	HA#	55	ILE	HN	3	59	ASN	HB1	60	SER HN 4
54	GLY	HA#	55	ILE	HN	3	59	ASN	HN	63	TRP HD1 2
54	GLY	HN	55	ILE	HN	4	59	ASN	HB1	63	TRP HD1 1
55	ILE	HN	55	ILE	HA	3	59	ASN	HA	61	ARG HN 4
55	ILE	HA	55	ILE	HB	1	59	ASN	HN	63	TRP HE1 4
55	ILE	HN	55	ILE	HB	2	59	ASN	HA	63	TRP HE1 5
55	ILE	HN	55	ILE	HG1#	3	59	ASN	HN	60	SER HN 4
55	ILE	HG2#	55	ILE	HD#	3	60	SER	HN	60	SER HA 2
55	ILE	HG2#	56	LEU	HA	2	60	SER	HN	60	SER HB# 3
55	ILE	HG2#	56	LEU	HN	2	60	SER	HA	60	SER HB# 1
55	ILE	HN	55	ILE	HG2#	2	60	SER	HA	64	CYS HN 2
55	ILE	HG2#	56	LEU	HD*	3	60	SER	HN	61	ARG HN 3
55	ILE	HG1#	56	LEU	HD*	2	60	SER	HA	81	SER HN 5
55	ILE	HN	56	LEU	HN	4	60	SER	HA	66	ASP HN 4
55	ILE	HN	57	GLN	HN	4	60	SER	HA	65	ASN HA 5
55	ILE	HG2#	88	ILE	HD#	4	60	SER	HA	64	CYS HA 4
56	LEU	HN	56	LEU	HA	3	60	SER	HN	62	TRP HN 5
56	LEU	HN	56	LEU	HB#	3	60	SER	HN	64	CYS HN 5
56	LEU	HN	56	LEU	HB#	3	61	ARG	HN	61	ARG HA 3
56	LEU	HA	56	LEU	HB#	3	61	ARG	HN	61	ARG HB2 2
56	LEU	HA	56	LEU	HD*	3	61	ARG	HN	61	ARG HB1 2
56	LEU	HA	56	LEU	HD*	3	61	ARG	HN	62	TRP HN 2
56	LEU	HA	57	GLN	HN	3	61	ARG	HA	72	SER HA 2
56	LEU	HN	57	GLN	HN	2	61	ARG	HN	63	TRP HN 4
56	LEU	HD*	91	SER	HB#	2	62	TRP	HN	62	TRP HA 3
56	LEU	HB#	56	LEU	HD*	2	62	TRP	HH3	62	TRP HZ2 1
56	LEU	HB#	56	LEU	HG	3	62	TRP	HZ3	62	TRP HH3 3
56	LEU	HD*	108	TRP	HE3	3	63	TRP	HN	63	TRP HA 3
56	LEU	HD*	108	TRP	HZ2	3	63	TRP	HA	63	TRP HB# 2
56	LEU	HB#	108	TRP	HZ2	3	63	TRP	HE3	63	TRP HZ3 1
56	LEU	HD*	108	TRP	HE1	3	63	TRP	HH3	63	TRP HZ2 3
56	LEU	HB#	108	TRP	HE1	3	63	TRP	HE3	63	TRP HB# 3
57	GLN	HN	57	GLN	HA	1	63	TRP	HD1	63	TRP HB# 3
57	GLN	HN	57	GLN	HB#	3	63	TRP	HD1	63	TRP HB# 3
57	GLN	HN	57	GLN	HB#	3	63	TRP	HE3	63	TRP HB# 3
57	GLN	HA	58	ILE	HN	2	63	TRP	HE3	63	TRP HA 2
57	GLN	HN	58	ILE	HN	3	63	TRP	HE3	76	CYS HN 3
57	GLN	HN	58	ILE	HG1#	3	63	TRP	HZ3	75	LEU HN 3
57	GLN	HN	58	ILE	HD*	2	63	TRP	HZ3	63	TRP HA 3
58	ILE	HN	58	ILE	HA	3	63	TRP	HB#	76	CYS HA 3
58	ILE	HN	58	ILE	HB	3	63	TRP	HA	75	LEU HN 2
58	ILE	HA	59	ASN	HN	2	63	TRP	HA	76	CYS HN 3
58	ILE	HG2#	63	TRP	HD1	3	63	TRP	HE3	75	LEU HB2 2
58	ILE	HG2#	59	ASN	HN	3	63	TRP	HZ3	75	LEU HB1 3
58	ILE	HD#	95	ALA	HN	3	63	TRP	HZ3	75	LEU HB2 3
58	ILE	HN	59	ASN	HA	5	63	TRP	HD1	98	ILE HG1# 3
58	ILE	HA	59	ASN	HG2	4	63	TRP	HE3	98	ILE HG1# 3
58	ILE	HN	60	SER	HN	4	63	TRP	HA	76	CYS HB# 1
59	ASN	HN	59	ASN	HA	3	63	TRP	HZ2	98	ILE HG2# 2
59	ASN	HN	59	ASN	HB2	2	63	TRP	HZ2	98	ILE HG1# 3
							63	TRP	HE1	63	TRP HB# 4

63	TRP	HE1	98	ILE	HG2#	4	67	GLY	HN	68	ARG	HA	4
63	TRP	HE1	98	ILE	HG1#	5	68	ARG	HN	68	ARG	HA	3
63	TRP	HD1	98	ILE	HD#	3	68	ARG	HA	68	ARG	HB#	2
63	TRP	HE3	75	LEU	HD*	2	68	ARG	HA	68	ARG	HB#	2
63	TRP	HZ3	75	LEU	HD*	3	68	ARG	HN	68	ARG	HB#	3
63	TRP	HE3	75	LEU	HN	4	68	ARG	HA	69	THR	HN	3
63	TRP	HN	75	LEU	HN	5	69	THR	HA	69	THR	HB	3
63	TRP	HZ3	101	ASP	HB#	4	69	THR	HN	69	THR	HB	3
63	TRP	HE3	98	ILE	HG2#	5	69	THR	HB	69	THR	HG2#	1
63	TRP	HH3	98	ILE	HG2#	4	70	PRO	HA	71	GLY	HN	1
64	CYS	HN	64	CYS	HA	2	71	GLY	HN	71	GLY	HA#	2
64	CYS	HA	64	CYS	HB#	2	71	GLY	HN	71	GLY	HA#	2
64	CYS	HA	64	CYS	HB#	2	71	GLY	HA#	71	GLY	HA#	1
64	CYS	HA	78	ILE	HN	3	71	GLY	HA#	72	SER	HN	3
64	CYS	HB#	80	CYS	HN	3	71	GLY	HN	72	SER	HN	2
64	CYS	HA	78	ILE	HG2#	2	72	SER	HN	72	SER	HA	3
64	CYS	HN	64	CYS	HB#	3	72	SER	HN	72	SER	HB#	2
64	CYS	HA	65	ASN	HN	2	72	SER	HN	72	SER	HB#	2
64	CYS	HA	75	LEU	HN	5	72	SER	HA	72	SER	HB#	2
64	CYS	HA	76	CYS	HN	4	72	SER	HA	72	SER	HB#	2
64	CYS	HA	77	ASN	HN	4	72	SER	HA	73	ARG	HN	3
64	CYS	HA	66	ASP	HN	5	72	SER	HA	74	ASN	HN	4
64	CYS	HA	65	ASN	HA	5	74	ASN	HN	74	ASN	HA	1
64	CYS	HA	74	ASN	HA	2	74	ASN	HA	74	ASN	HB#	3
64	CYS	HN	75	LEU	HN	5	74	ASN	HA	74	ASN	HB#	3
64	CYS	HA	74	ASN	HB#	4	74	ASN	HA	75	LEU	HN	3
64	CYS	HN	76	CYS	HN	5	74	ASN	HA	76	CYS	HN	3
64	CYS	HN	66	ASP	HN	5	74	ASN	HN	74	ASN	HB#	3
64	CYS	HN	76	CYS	HB#	4	74	ASN	HN	75	LEU	HN	3
64	CYS	HA	80	CYS	HA	4	74	ASN	HN	76	CYS	HN	5
64	CYS	HA	75	LEU	HB1	4	75	LEU	HN	75	LEU	HA	3
65	ASN	HN	65	ASN	HA	3	75	LEU	HN	75	LEU	HB2	2
65	ASN	HN	65	ASN	HB#	2	75	LEU	HN	75	LEU	HB1	3
65	ASN	HN	65	ASN	HB#	2	75	LEU	HA	75	LEU	HB2	3
65	ASN	HA	65	ASN	HB#	3	75	LEU	HA	75	LEU	HB1	3
65	ASN	HA	65	ASN	HB#	3	75	LEU	HN	75	LEU	HG	2
65	ASN	HA	66	ASP	HN	2	75	LEU	HA	76	CYS	HN	3
65	ASN	HA	67	GLY	HN	3	75	LEU	HN	76	CYS	HN	3
65	ASN	HN	78	ILE	HG2#	3	75	LEU	HB2	76	CYS	HN	3
65	ASN	HA	72	SER	HN	5	75	LEU	HN	75	LEU	HD*	3
65	ASN	HB#	66	ASP	HN	4	75	LEU	HN	75	LEU	HD*	3
65	ASN	HN	66	ASP	HN	4	75	LEU	HB1	76	CYS	HB#	2
65	ASN	HN	78	ILE	HN	4	75	LEU	HG	76	CYS	HN	5
65	ASN	HN	76	CYS	HN	5	75	LEU	HD*	76	CYS	HN	4
66	ASP	HN	66	ASP	HA	3	76	CYS	HN	76	CYS	HA	3
66	ASP	HN	66	ASP	HB1	3	76	CYS	HN	76	CYS	HB#	2
66	ASP	HA	66	ASP	HB2	1	76	CYS	HN	77	ASN	HN	2
66	ASP	HA	66	ASP	HB1	1	76	CYS	HN	78	ILE	HG2#	3
66	ASP	HA	67	GLY	HN	3	76	CYS	HA	76	CYS	HB#	1
66	ASP	HN	67	GLY	HN	2	76	CYS	HA	77	ASN	HN	3
66	ASP	HN	67	GLY	HA	4	76	CYS	HB#	78	ILE	HG2#	1
66	ASP	HN	68	ARG	HN	5	76	CYS	HN	77	ASN	HA	4
66	ASP	HN	69	THR	HN	5	76	CYS	HN	77	ASN	HB#	5
67	GLY	HN	67	GLY	HA#	3	76	CYS	HN	78	ILE	HN	4
67	GLY	HN	67	GLY	HA#	3	77	ASN	HA	77	ASN	HB#	1
67	GLY	HA#	67	GLY	HA#	1	77	ASN	HA	77	ASN	HB#	1

77	ASN	HA	78	ILE	HN	3	87	ASP	HN	87	ASP	HB1	3
77	ASN	HB#	78	ILE	HN	3	88	ILE	HN	88	ILE	HG1#	3
77	ASN	HN	77	ASN	HB#	3	88	ILE	HN	88	ILE	HG2#	2
77	ASN	HN	77	ASN	HB#	3	88	ILE	HN	88	ILE	HD#	3
77	ASN	HN	78	ILE	HN	3	88	ILE	HG1#	89	THR	HN	3
77	ASN	HA	78	ILE	HN	2	88	ILE	HG2#	89	THR	HN	3
78	ILE	HN	78	ILE	HA	3	88	ILE	HD#	89	THR	HN	3
78	ILE	HA	78	ILE	HB	1	88	ILE	HN	88	ILE	HG1#	2
78	ILE	HA	78	ILE	HG2#	1	88	ILE	HD#	88	ILE	HG1#	3
78	ILE	HN	78	ILE	HG1#	2	88	ILE	HD#	92	VAL	HG2#	1
78	ILE	HN	78	ILE	HG2#	3	88	ILE	HB	88	ILE	HD#	2
79	PRO	HA	79	PRO	HB#	2	88	ILE	HD#	88	ILE	HG2#	1
79	PRO	HA	79	PRO	HB#	2	88	ILE	HB	88	ILE	HG2#	2
79	PRO	HA	80	CYS	HN	1	88	ILE	HB	88	ILE	HG1#	3
80	CYS	HN	80	CYS	HA	3	89	THR	HA	92	VAL	HN	3
80	CYS	HA	83	LEU	HN	3	89	THR	HA	92	VAL	HB	2
80	CYS	HN	81	SER	HN	3	89	THR	HA	90	ALA	HN	3
80	CYS	HA	81	SER	HN	2	89	THR	HN	90	ALA	HN	3
81	SER	HN	81	SER	HA	2	89	THR	HN	89	THR	HG2#	3
81	SER	HN	82	ALA	HN	2	89	THR	HA	89	THR	HG2#	1
81	SER	HA	84	LEU	HN	4	89	THR	HB	89	THR	HG2#	1
81	SER	HN	83	LEU	HN	4	89	THR	HG2#	90	ALA	HN	3
82	ALA	HN	82	ALA	HA	2	89	THR	HG2#	93	ASN	HN	3
82	ALA	HN	82	ALA	HB#	1	90	ALA	HN	90	ALA	HA	1
82	ALA	HA	82	ALA	HB#	1	90	ALA	HN	90	ALA	HB#	1
82	ALA	HN	83	LEU	HN	2	90	ALA	HA	90	ALA	HB#	1
82	ALA	HB#	83	LEU	HN	3	90	ALA	HA	93	ASN	HN	3
82	ALA	HA	84	LEU	HN	4	90	ALA	HN	91	SER	HN	2
83	LEU	HN	83	LEU	HA	2	90	ALA	HB#	91	SER	HN	2
83	LEU	HN	83	LEU	HB#	2	91	SER	HN	91	SER	HA	2
83	LEU	HA	83	LEU	HB#	2	91	SER	HN	91	SER	HB#	3
83	LEU	HA	83	LEU	HD*	1	91	SER	HN	91	SER	HB#	3
83	LEU	HA	84	LEU	HN	3	91	SER	HA	92	VAL	HN	3
83	LEU	HN	84	LEU	HN	2	91	SER	HA	94	CYS	HN	3
83	LEU	HN	83	LEU	HG	2	91	SER	HB#	92	VAL	HN	3
83	LEU	HN	83	LEU	HD*	3	91	SER	HN	92	VAL	HN	2
84	LEU	HN	84	LEU	HA	3	92	VAL	HN	92	VAL	HA	2
84	LEU	HA	84	LEU	HB#	3	92	VAL	HN	92	VAL	HB	1
84	LEU	HN	84	LEU	HB#	2	92	VAL	HA	92	VAL	HB	2
84	LEU	HA	85	SER	HN	3	92	VAL	HA	92	VAL	HG2#	1
85	SER	HN	85	SER	HA	2	92	VAL	HA	92	VAL	HG1#	1
85	SER	HN	85	SER	HB#	2	92	VAL	HN	92	VAL	HG2#	1
85	SER	HN	85	SER	HB#	2	92	VAL	HN	92	VAL	HG1#	3
85	SER	HA	85	SER	HB#	2	92	VAL	HA	95	ALA	HN	2
85	SER	HA	85	SER	HB#	2	92	VAL	HN	93	ASN	HN	2
85	SER	HA	86	SER	HN	1	92	VAL	HA	95	ALA	HB#	1
86	SER	HN	86	SER	HA	2	92	VAL	HG2#	93	ASN	HN	3
86	SER	HN	86	SER	HB#	3	92	VAL	HB	93	ASN	HN	2
86	SER	HN	86	SER	HB#	3	92	VAL	HG1#	93	ASN	HN	3
86	SER	HA	86	SER	HB#	1	92	VAL	HG1#	96	LYS	HA	2
86	SER	HN	87	ASP	HN	2	92	VAL	HB	92	VAL	HG2#	1
87	ASP	HA	87	ASP	HB2	3	92	VAL	HB	92	VAL	HG1#	2
87	ASP	HA	87	ASP	HB1	2	92	VAL	HN	94	CYS	HN	4
87	ASP	HA	88	ILE	HN	1	93	ASN	HN	93	ASN	HA	1
87	ASP	HN	87	ASP	HA	3	93	ASN	HN	93	ASN	HB2	2
87	ASP	HN	87	ASP	HB2	3	93	ASN	HN	93	ASN	HB1	1

93	ASN	HA	93	ASN	HB2	2	99	VAL	HN	100	SER	HN	3
93	ASN	HA	93	ASN	HB1	1	99	VAL	HA	100	SER	HN	3
93	ASN	HB2	94	CYS	HN	3	99	VAL	HB	100	SER	HN	3
93	ASN	HA	96	LYS	HN	3	99	VAL	HG*	108	TRP	HZ3	2
93	ASN	HN	94	CYS	HN	2	99	VAL	HB	108	TRP	HZ3	3
94	CYS	HN	94	CYS	HA	2	99	VAL	HN	108	TRP	HZ3	4
94	CYS	HN	94	CYS	HB2	2	100	SER	HN	100	SER	HA	2
94	CYS	HN	94	CYS	HB1	2	100	SER	HN	100	SER	HB#	3
94	CYS	HA	94	CYS	HB2	2	100	SER	HN	100	SER	HB#	3
94	CYS	HA	95	ALA	HN	3	100	SER	HA	100	SER	HB#	3
94	CYS	HB1	95	ALA	HN	2	100	SER	HN	101	ASP	HN	3
94	CYS	HA	97	LYS	HN	3	101	ASP	HN	101	ASP	HA	3
94	CYS	HN	95	ALA	HN	2	101	ASP	HN	101	ASP	HB#	1
95	ALA	HN	95	ALA	HA	2	101	ASP	HA	101	ASP	HB#	1
95	ALA	HN	95	ALA	HB#	1	101	ASP	HA	102	GLY	HN	3
95	ALA	HA	95	ALA	HB#	1	102	GLY	HN	102	GLY	HA#	3
95	ALA	HN	96	LYS	HN	2	102	GLY	HN	102	GLY	HA#	3
95	ALA	HN	97	LYS	HN	3	104	GLY	HN	104	GLY	HA#	3
95	ALA	HB#	96	LYS	HN	2	104	GLY	HN	104	GLY	HA#	3
95	ALA	HB#	108	TRP	HH3	1	104	GLY	HA#	104	GLY	HA#	1
95	ALA	HB#	108	TRP	HZ2	2	104	GLY	HN	105	MET	HN	3
95	ALA	HA	108	TRP	HZ2	2	104	GLY	HA#	105	MET	HN	3
95	ALA	HN	98	ILE	HG2#	5	104	GLY	HA#	105	MET	HN	3
96	LYS	HN	96	LYS	HA	2	105	MET	HN	105	MET	HA	3
96	LYS	HN	96	LYS	HB#	1	105	MET	HN	105	MET	HB#	3
96	LYS	HA	96	LYS	HB#	2	105	MET	HA	105	MET	HB#	2
96	LYS	HN	97	LYS	HN	2	105	MET	HA	105	MET	HE#	2
97	LYS	HN	97	LYS	HA	2	105	MET	HB#	108	TRP	HE3	3
97	LYS	HA	98	ILE	HN	3	105	MET	HA	108	TRP	HE3	1
97	LYS	HN	98	ILE	HN	2	105	MET	HA	108	TRP	HZ3	2
97	LYS	HN	99	VAL	HN	3	105	MET	HE#	105	MET	HB#	2
98	ILE	HN	98	ILE	HB	2	105	MET	HB#	111	TRP	HZ2	3
98	ILE	HN	98	ILE	HG2#	3	105	MET	HB#	108	TRP	HZ3	4
98	ILE	HN	98	ILE	HG1#	2	105	MET	HB#	111	TRP	HE1	4
98	ILE	HN	98	ILE	HD#	3	105	MET	HE#	111	TRP	HE1	4
98	ILE	HN	99	VAL	HN	3	105	MET	HB#	106	ASN	HN	4
98	ILE	HG2#	99	VAL	HA	3	106	ASN	HN	106	ASN	HB	3
98	ILE	HB	99	VAL	HN	2	106	ASN	HA	106	ASN	HB	2
98	ILE	HG2#	108	TRP	HZ3	3	106	ASN	HA	107	ALA	HN	3
98	ILE	HG2#	108	TRP	HZ2	3	106	ASN	HN	107	ALA	HN	2
98	ILE	HN	98	ILE	HA	3	107	ALA	HN	107	ALA	HA	2
98	ILE	HG2#	99	VAL	HN	3	107	ALA	HN	107	ALA	HB#	1
98	ILE	HA	98	ILE	HG2#	1	107	ALA	HA	107	ALA	HB#	1
98	ILE	HN	98	ILE	HB	2	107	ALA	HA	108	TRP	HN	3
98	ILE	HG2#	98	ILE	HD#	2	107	ALA	HB#	108	TRP	HN	3
98	ILE	HB	98	ILE	HG2#	2	107	ALA	HB#	108	TRP	HE3	2
98	ILE	HD#	108	TRP	HZ2	2	107	ALA	HN	108	TRP	HN	2
98	ILE	HG2#	108	TRP	HE1	4	107	ALA	HN	109	VAL	HN	5
98	ILE	HG1#	99	VAL	HN	4	108	TRP	HN	108	TRP	HB#	3
98	ILE	HG2#	99	VAL	HB	5	108	TRP	HN	108	TRP	HB#	3
98	ILE	HG2#	107	ALA	HN	4	108	TRP	HH3	108	TRP	HZ2	2
98	ILE	HG2#	108	TRP	HE3	4	108	TRP	HE3	108	TRP	HZ3	2
99	VAL	HN	99	VAL	HA	3	108	TRP	HZ3	108	TRP	HH3	2
99	VAL	HN	99	VAL	HB	1	108	TRP	HE3	108	TRP	HB#	3
99	VAL	HA	99	VAL	HB	1	108	TRP	HA	108	TRP	HD1	3
99	VAL	HN	99	VAL	HG*	2							

108	TRP	HN	108	TRP	HE3	3	117	GLY	HN	117	GLY	HA#	2
108	TRP	HN	108	TRP	HZ3	4	117	GLY	HA#	117	GLY	HA#	1
109	VAL	HA	109	VAL	HB	1	117	GLY	HN	118	THR	HN	2
109	VAL	HA	112	ARG	HN	2	118	THR	HN	118	THR	HA	3
109	VAL	HN	109	VAL	HB	2	118	THR	HN	118	THR	HB	3
109	VAL	HN	110	ALA	HN	2	118	THR	HA	118	THR	HB	1
109	VAL	HN	111	TRP	HN	3	118	THR	HA	119	ASP	HN	2
109	VAL	HN	109	VAL	HG*	2	118	THR	HN	118	THR	HG2#	2
109	VAL	HN	109	VAL	HG*	2	118	THR	HB	118	THR	HG2#	1
109	VAL	HG*	110	ALA	HN	3	118	THR	HG2#	119	ASP	HN	3
109	VAL	HG*	110	ALA	HN	3	119	ASP	HN	119	ASP	HA	2
109	VAL	HB	110	ALA	HN	2	119	ASP	HA	119	ASP	HB2	1
109	VAL	HA	110	ALA	HN	3	119	ASP	HA	119	ASP	HB1	1
109	VAL	HN	112	ARG	HN	4	119	ASP	HA	120	VAL	HN	1
110	ALA	HN	110	ALA	HA	2	119	ASP	HN	119	ASP	HB2	3
110	ALA	HA	110	ALA	HB#	1	119	ASP	HN	119	ASP	HB1	2
110	ALA	HA	113	ASN	HB#	2	119	ASP	HA	121	GLN	HN	3
110	ALA	HN	110	ALA	HB#	1	120	VAL	HN	120	VAL	HA	2
110	ALA	HN	111	TRP	HN	2	120	VAL	HA	120	VAL	HB	3
111	TRP	HN	111	TRP	HA	2	120	VAL	HN	121	GLN	HN	2
111	TRP	HN	112	ARG	HN	1	120	VAL	HN	120	VAL	HG*	2
111	TRP	HE3	111	TRP	HZ3	1	120	VAL	HN	120	VAL	HG*	2
111	TRP	HZ3	111	TRP	HH3	2	120	VAL	HG*	121	GLN	HN	3
111	TRP	HE3	112	ARG	HA	2	120	VAL	HG*	121	GLN	HN	3
111	TRP	HE3	112	ARG	HB#	3	120	VAL	HA	123	TRP	HE1	3
111	TRP	HE1	116	LYS	HA	4	121	GLN	HN	121	GLN	HA	2
111	TRP	HZ2	116	LYS	HB#	4	121	GLN	HA	121	GLN	HB#	2
112	ARG	HN	112	ARG	HA	2	121	GLN	HN	121	GLN	HB#	2
112	ARG	HN	112	ARG	HB#	1	121	GLN	HN	121	GLN	HB#	2
112	ARG	HN	113	ASN	HN	3	121	GLN	HA	122	ALA	HN	3
112	ARG	HB#	113	ASN	HN	2	121	GLN	HN	122	ALA	HN	2
112	ARG	HA	116	LYS	HB#	2	121	GLN	HN	123	TRP	HN	4
112	ARG	HN	116	LYS	HB#	5	121	GLN	HN	123	TRP	HD1	4
113	ASN	HN	113	ASN	HA	3	122	ALA	HN	122	ALA	HA	2
113	ASN	HN	113	ASN	HB#	1	122	ALA	HA	122	ALA	HB#	1
113	ASN	HA	113	ASN	HB#	1	122	ALA	HN	122	ALA	HB#	1
113	ASN	HB#	114	ARG	HN	2	122	ALA	HA	123	TRP	HN	3
113	ASN	HN	114	ARG	HN	3	122	ALA	HB#	123	TRP	HN	2
113	ASN	HN	115	CYS	HB#	3	123	TRP	HN	123	TRP	HA	2
114	ARG	HN	114	ARG	HA	2	123	TRP	HN	123	TRP	HB2	3
114	ARG	HA	114	ARG	HB#	2	123	TRP	HN	123	TRP	HB1	3
114	ARG	HN	114	ARG	HB#	2	123	TRP	HA	123	TRP	HB2	3
114	ARG	HN	115	CYS	HN	3	123	TRP	HA	123	TRP	HB1	2
115	CYS	HN	115	CYS	HA	3	123	TRP	HB1	124	ILE	HN	3
115	CYS	HN	115	CYS	HB#	3	123	TRP	HH3	123	TRP	HZ2	1
115	CYS	HN	115	CYS	HB#	3	123	TRP	HE3	123	TRP	HZ3	1
115	CYS	HA	115	CYS	HB#	2	123	TRP	HE3	123	TRP	HB2	3
116	LYS	HN	116	LYS	HA	2	123	TRP	HE3	123	TRP	HB1	3
116	LYS	HN	116	LYS	HB#	1	123	TRP	HE3	123	TRP	HA	2
116	LYS	HA	116	LYS	HB#	2	123	TRP	HZ2	123	TRP	HE3	3
116	LYS	HA	116	LYS	HB#	2	123	TRP	HE1	123	TRP	HB2	4
116	LYS	HA	117	GLY	HN	1	124	ILE	HN	124	ILE	HA	3
116	LYS	HB#	117	GLY	HN	3	124	ILE	HA	124	ILE	HB	1
116	LYS	HB#	117	GLY	HN	3	124	ILE	HN	124	ILE	HG1#	2
116	LYS	HN	117	GLY	HN	4	124	ILE	HN	124	ILE	HG1#	2
117	GLY	HN	117	GLY	HA#	2	124	ILE	HN	124	ILE	HG2#	2

124	ILE	HN	125	ARG	HN	1
124	ILE	HB	124	ILE	HD#	2
124	ILE	HB	124	ILE	HG2#	1
125	ARG	HN	125	ARG	HA	2
125	ARG	HN	125	ARG	HB#	1
125	ARG	HN	125	ARG	HB#	1
125	ARG	HA	125	ARG	HB#	2
125	ARG	HA	125	ARG	HB#	2
125	ARG	HB#	126	GLY	HN	3
125	ARG	HA	126	GLY	HN	1
125	ARG	HN	126	GLY	HN	4
126	GLY	HN	126	GLY	HA#	2
126	GLY	HN	126	GLY	HA#	2
126	GLY	HA#	126	GLY	HA#	1
126	GLY	HN	127	CYS	HN	2
127	CYS	HN	127	CYS	HA	2
127	CYS	HN	127	CYS	HB2	3
127	CYS	HN	127	CYS	HB1	3
127	CYS	HA	127	CYS	HB2	1
127	CYS	HA	127	CYS	HB1	2
127	CYS	HB1	128	ARG	HN	3
127	CYS	HN	128	ARG	HN	3
127	CYS	HA	128	ARG	HN	2
127	CYS	HB1	129	LEU	HD*	3
127	CYS	HA	129	LEU	HN	5
128	ARG	HN	128	ARG	HA	3
128	ARG	HA	128	ARG	HB#	3
128	ARG	HA	128	ARG	HB#	1
128	ARG	HA	129	LEU	HN	1
128	ARG	HN	129	LEU	HB#	3
129	LEU	HN	129	LEU	HA	2
129	LEU	HA	129	LEU	HB#	2
129	LEU	HA	129	LEU	HD*	1
129	LEU	HN	129	LEU	HB#	3
129	LEU	HN	129	LEU	HD*	3
129	LEU	HN	129	LEU	HD*	3
129	LEU	HN	129	LEU	HG	2
129	LEU	HB#	129	LEU	HG	1

Phi restraints for hen lysozyme from J coupling constant data.

```
ASSIGN (RESID   1 AND NAME C) (RESID   2 AND NAME N)
       (RESID   2 AND NAME CA) (RESID   2 AND NAME C) $8 -120.0 15.0 2
ASSIGN (RESID   6 AND NAME C) (RESID   7 AND NAME N)
       (RESID   7 AND NAME CA) (RESID   7 AND NAME C) $8 -60.0 10.0 2
ASSIGN (RESID   7 AND NAME C) (RESID   8 AND NAME N)
       (RESID   8 AND NAME CA) (RESID   8 AND NAME C) $8 -70.0 10.0 2
ASSIGN (RESID   8 AND NAME C) (RESID   9 AND NAME N)
       (RESID   9 AND NAME CA) (RESID   9 AND NAME C) $8 -55.0 10.0 2
ASSIGN (RESID   9 AND NAME C) (RESID   10 AND NAME N)
       (RESID  10 AND NAME CA) (RESID   10 AND NAME C) $8 -60.0 10.0 2
ASSIGN (RESID  10 AND NAME C) (RESID   11 AND NAME N)
       (RESID  11 AND NAME CA) (RESID   11 AND NAME C) $8 -65.0 10.0 2
ASSIGN (RESID  11 AND NAME C) (RESID   12 AND NAME N)
       (RESID  12 AND NAME CA) (RESID   12 AND NAME C) $8 -65.0 10.0 2
ASSIGN (RESID  12 AND NAME C) (RESID   13 AND NAME N)
       (RESID  13 AND NAME CA) (RESID   13 AND NAME C) $8 -60.0 10.0 2
ASSIGN (RESID  13 AND NAME C) (RESID   14 AND NAME N)
       (RESID  14 AND NAME CA) (RESID   14 AND NAME C) $8 -60.0 10.0 2
ASSIGN (RESID  14 AND NAME C) (RESID   15 AND NAME N)
       (RESID  15  AND NAME CA) (RESID   15 AND NAME C) $8 -120.0 30.0 2
ASSIGN (RESID  16 AND NAME C) (RESID   17 AND NAME N)
       (RESID  17 AND NAME CA) (RESID   17 AND NAME C) $8 -120.0 40.0 2
ASSIGN (RESID  19 AND NAME C) (RESID   20 AND NAME N)
       (RESID  20 AND NAME CA) (RESID   20 AND NAME C) $8 -70.0 10.0 2
ASSIGN (RESID  20 AND NAME C) (RESID   21 AND NAME N)
       (RESID  21 AND NAME CA) (RESID   21 AND NAME C) $8 60.0 20.0 2
ASSIGN (RESID  22 AND NAME C) (RESID   23 AND NAME N)
       (RESID  23 AND NAME CA) (RESID  23 AND NAME C) $8 -120.0 30.0 2
ASSIGN (RESID  26 AND NAME C) (RESID  27 AND NAME N)
       (RESID  27 AND NAME CA) (RESID  27 AND NAME C) $8 -70.0 10.0 2
ASSIGN (RESID  29 AND NAME C) (RESID   30 AND NAME N)
       (RESID  30 AND NAME CA) (RESID  30 AND NAME C) $8 -55.0 10.0 2
ASSIGN (RESID  30 AND NAME C) (RESID   31 AND NAME N)
       (RESID  31 AND NAME CA) (RESID   31 AND NAME C) $8 -55.0 10.0 2
ASSIGN (RESID  31 AND NAME C) (RESID   32 AND NAME N)
       (RESID  32 AND NAME CA) (RESID  32 AND NAME C) $8 -65.0 10.0 2
ASSIGN (RESID  32 AND NAME C) (RESID   33 AND NAME N)
       (RESID  33 AND NAME CA) (RESID  33 AND NAME C) $8 -55.0 10.0 2
ASSIGN (RESID  33 AND NAME C) (RESID   34 AND NAME N)
       (RESID  34 AND NAME CA) (RESID   34 AND NAME C) $8 -120.0 40.0 2
ASSIGN (RESID  35 AND NAME C) (RESID  36 AND NAME N)
       (RESID  36 AND NAME CA) (RESID  36 AND NAME C) $8 -120.0 40.0 2
ASSIGN (RESID  37 AND NAME C) (RESID  38 AND NAME N)
       (RESID  38 AND NAME CA) (RESID  38 AND NAME C) $8 60.0 20.0 2
ASSIGN (RESID  38 AND NAME C) (RESID  39 AND NAME N)
       (RESID  39 AND NAME CA) (RESID  39 AND NAME C) $8 -120.0 30.0 2
ASSIGN (RESID  40 AND NAME C) (RESID  41 AND NAME N)
       (RESID  41 AND NAME CA) (RESID  41 AND NAME C) $8 -120.0 30.0 2
ASSIGN (RESID  41 AND NAME C) (RESID  42 AND NAME N)
       (RESID  42 AND NAME CA) (RESID  42 AND NAME C) $8 -60.0 10.0 2
ASSIGN (RESID  42 AND NAME C) (RESID  43 AND NAME N)
       (RESID  43 AND NAME CA) (RESID  43 AND NAME C) $8 -120.0 30.0 2
ASSIGN (RESID  43 AND NAME C) (RESID  44 AND NAME N)
       (RESID  44 AND NAME CA) (RESID  44 AND NAME C) $8 -120.0 20.0 2
```

```
ASSIGN (RESID  45 AND NAME C) (RESID  46 AND NAME N)
       (RESID  46 AND NAME CA) (RESID  46 AND NAME C) $8 -120.0 30.0 2
ASSIGN (RESID  46 AND NAME C) (RESID  47 AND NAME N)
       (RESID  47 AND NAME CA) (RESID  47 AND NAME C) $8 -60.0 10.0 2
ASSIGN (RESID  47 AND NAME C) (RESID  48 AND NAME N)
       (RESID  48 AND NAME CA) (RESID  48 AND NAME C) $8 -120.0 40.0 2
ASSIGN (RESID  49 AND NAME C) (RESID  50 AND NAME N)
       (RESID  50 AND NAME CA) (RESID  50 AND NAME C) $8 -120.0 40.0 2
ASSIGN (RESID  50 AND NAME C) (RESID  51 AND NAME N)
       (RESID  51 AND NAME CA) (RESID  51 AND NAME C) $8 -120.0 15.0 2
ASSIGN (RESID  51 AND NAME C) (RESID  52 AND NAME N)
       (RESID  52 AND NAME CA) (RESID  52 AND NAME C) $8 -120.0 20.0 2
ASSIGN (RESID  52 AND NAME C) (RESID  53 AND NAME N)
       (RESID  53 AND NAME CA) (RESID  53 AND NAME C) $8 -120.0 20.0 2
ASSIGN (RESID  55 AND NAME C) (RESID  56 AND NAME N)
       (RESID  56 AND NAME CA) (RESID  56 AND NAME C) $8 -120.0 15.0 2
ASSIGN (RESID  56 AND NAME C) (RESID  57 AND NAME N)
       (RESID  57 AND NAME CA) (RESID  57 AND NAME C) $8 60.0 20.0 2
ASSIGN (RESID  57 AND NAME C) (RESID  58 AND NAME N)
       (RESID  58 AND NAME CA) (RESID  58 AND NAME C) $8 -120.0 40.0 2
ASSIGN (RESID  63 AND NAME C) (RESID  64 AND NAME N)
       (RESID  64 AND NAME CA) (RESID  64 AND NAME C) $8 -120.0 30.0 2
ASSIGN (RESID  64 AND NAME C) (RESID  65 AND NAME N)
       (RESID  65 AND NAME CA) (RESID  65 AND NAME C) $8 -120.0 20.0 2
ASSIGN (RESID  65 AND NAME C) (RESID  66 AND NAME N)
       (RESID  66 AND NAME CA) (RESID  66 AND NAME C) $8 -120.0 15.0 2
ASSIGN (RESID  67 AND NAME C) (RESID  68 AND NAME N)
       (RESID  68 AND NAME CA) (RESID  68 AND NAME C) $8 -120.0 15.0 2
ASSIGN (RESID  68 AND NAME C) (RESID  69 AND NAME N)
       (RESID  69 AND NAME CA) (RESID  69 AND NAME C) $8 -120.0 30.0 2
ASSIGN (RESID  75 AND NAME C) (RESID  76 AND NAME N)
       (RESID  76 AND NAME CA) (RESID  76 AND NAME C) $8 -120.0 30.0 2
ASSIGN (RESID  77 AND NAME C) (RESID  78 AND NAME N)
       (RESID  78 AND NAME CA) (RESID  78 AND NAME C) $8 -120.0 40.0 2
ASSIGN (RESID  79 AND NAME C) (RESID  80 AND NAME N)
       (RESID  80 AND NAME CA) (RESID  80 AND NAME C) $8 -55.0 10.0 2
ASSIGN (RESID  80 AND NAME C) (RESID  81 AND NAME N)
       (RESID  81 AND NAME CA) (RESID  81 AND NAME C) $8 -55.0 10.0 2
ASSIGN (RESID  83 AND NAME C) (RESID  84 AND NAME N)
       (RESID  84 AND NAME CA) (RESID  84 AND NAME C) $8 -120.0 30.0 2
ASSIGN (RESID  86 AND NAME C) (RESID  87 AND NAME N)
       (RESID  87 AND NAME CA) (RESID  87 AND NAME C) $8 -120.0 30.0 2
ASSIGN (RESID  89 AND NAME C) (RESID  90 AND NAME N)
       (RESID  90 AND NAME CA) (RESID  90 AND NAME C) $8 -60.0 10.0 2
ASSIGN (RESID  92 AND NAME C) (RESID  93 AND NAME N)
       (RESID  93 AND NAME CA) (RESID  93 AND NAME C) $8 -60.0 10.0 2
ASSIGN (RESID  95 AND NAME C) (RESID  96 AND NAME N)
       (RESID  96 AND NAME CA) (RESID  96 AND NAME C) $8 -60.0 10.0 2
ASSIGN (RESID 102 AND NAME C) (RESID 103 AND NAME N)
       (RESID 103 AND NAME CA) (RESID 103 AND NAME C) $8 -120.0 40.0 2
ASSIGN (RESID 104 AND NAME C) (RESID  105 AND NAME N)
       (RESID 105 AND NAME CA) (RESID  105 AND NAME C) $8 -60.0 10.0 2
ASSIGN (RESID 106 AND NAME C) (RESID  107 AND NAME N)
       (RESID 107 AND NAME CA) (RESID  107 AND NAME C) $8 -60.0 10.0 2
ASSIGN (RESID 107 AND NAME C) (RESID  108 AND NAME N)
       (RESID 108 AND NAME CA) (RESID  108 AND NAME C) $8 -120.0 20.0 2
```

```
ASSIGN (RESID  108 AND NAME C) (RESID  109 AND NAME N)
       (RESID  109 AND NAME CA) (RESID  109 AND NAME C) $8 -60.0 10.0 2
ASSIGN (RESID 111 AND NAME C) (RESID 112 AND NAME N)
       (RESID 112 AND NAME CA) (RESID 112 AND NAME C) $8 -60.0 10.0 2
ASSIGN (RESID 113 AND NAME C) (RESID  114 AND NAME N)
       (RESID 114 AND NAME CA) (RESID 114 AND NAME C) $8 -120.0 20.0 2
ASSIGN (RESID 114 AND NAME C) (RESID  115 AND NAME N)
       (RESID 115 AND NAME CA) (RESID 115 AND NAME C) $8 -120.0 15.0 2
ASSIGN (RESID 117 AND NAME C) (RESID  118 AND NAME N)
       (RESID 118 AND NAME CA) (RESID 118 AND NAME C) $8 -120.0 15.0 2
ASSIGN (RESID 119 AND NAME C) (RESID  120 AND NAME N)
       (RESID 120 AND NAME CA) (RESID 120 AND NAME C) $8 -65.0 10.0 2
ASSIGN (RESID 120 AND NAME C) (RESID  121 AND NAME N)
       (RESID 121 AND NAME CA) (RESID 121 AND NAME C) $8 -65.0 10.0 2
ASSIGN (RESID 121 AND NAME C) (RESID  122 AND NAME N)
       (RESID 122 AND NAME CA) (RESID 122 AND NAME C) $8 -55.0 10.0 2
ASSIGN (RESID 123 AND NAME C) (RESID  124 AND NAME N)
       (RESID 124 AND NAME CA) (RESID 124 AND NAME C) $8 -120.0 15.0 2
ASSIGN (RESID 124 AND NAME C) (RESID  125 AND NAME N)
       (RESID 125 AND NAME CA) (RESID 125 AND NAME C) $8 -60.0 10.0 2
ASSIGN (RESID 126 AND NAME C) (RESID  127 AND NAME N)
       (RESID 127 AND NAME CA) (RESID 127 AND NAME C) $8 -120.0 40.0 2
ASSIGN (RESID 127 AND NAME C) (RESID  128 AND NAME N)
       (RESID 128 AND NAME CA) (RESID  128 AND NAME C) $8 -120.0 40.0 2
ASSIGN (RESID 128 AND NAME C) (RESID  129 AND NAME N)
       (RESID 129 AND NAME CA) (RESID  129 AND NAME C) $8 -120.0 30.0 2
```

Chi1 restraints for hen lysozyme from J coupling constant data.

```
ASSIGN (RESID   3 AND NAME N) (RESID   3 AND NAME CA)
       (RESID   3 AND NAME CB) (RESID   3 AND NAME CG) $9 -60.0 15.0 2
ASSIGN (RESID   6 AND NAME N) (RESID   6 AND NAME CA)
       (RESID   6 AND NAME CB) (RESID   6 AND NAME SG) $9 -60.0 15.0 2
ASSIGN (RESID  15 AND NAME N) (RESID  15 AND NAME CA)
       (RESID  15 AND NAME CB) (RESID  15 AND NAME CG) $9 -60.0 15.0 2
ASSIGN (RESID  18 AND NAME N) (RESID  18 AND NAME CA)
       (RESID  18 AND NAME CB) (RESID  18 AND NAME CG) $9 180.0 15.0 2
ASSIGN (RESID  20 AND NAME N) (RESID  20 AND NAME CA)
       (RESID  20 AND NAME CB) (RESID  20 AND NAME CG) $9 180.0 15.0 2
ASSIGN (RESID  23 AND NAME N) (RESID  23 AND NAME CA)
       (RESID  23 AND NAME CB) (RESID  23 AND NAME CG) $9 -60.0 15.0 2
ASSIGN (RESID  27 AND NAME N) (RESID  27 AND NAME CA)
       (RESID  27 AND NAME CB) (RESID  27 AND NAME CG) $9 -60.0 15.0 2
ASSIGN (RESID  30 AND NAME N) (RESID  30 AND NAME CA)
       (RESID  30 AND NAME CB) (RESID  30 AND NAME SG) $9 180.0 15.0 2
ASSIGN (RESID  34 AND NAME N) (RESID  34 AND NAME CA)
       (RESID  34 AND NAME CB) (RESID  34 AND NAME CG) $9 -60.0 15.0 2
ASSIGN (RESID  39 AND NAME N) (RESID  39 AND NAME CA)
       (RESID  39 AND NAME CB) (RESID  39 AND NAME CG) $9 180.0 15.0 2
ASSIGN (RESID  46 AND NAME N) (RESID  46 AND NAME CA)
       (RESID  46 AND NAME CB) (RESID  46 AND NAME CG) $9 -60.0 15.0 2
ASSIGN (RESID  48 AND NAME N) (RESID  48 AND NAME CA)
       (RESID  48 AND NAME CB) (RESID  48 AND NAME CG) $9  60.0 15.0 2
ASSIGN (RESID  52 AND NAME N) (RESID  52 AND NAME CA)
       (RESID  52 AND NAME CB) (RESID  52 AND NAME CG) $9 -60.0 15.0 2
ASSIGN (RESID  53 AND NAME N) (RESID  53 AND NAME CA)
```

```
        (RESID  53 AND NAME CB) (RESID  53 AND NAME CG) $9 -60.0 15.0 2
ASSIGN (RESID  59 AND NAME N) (RESID  59 AND NAME CA)
        (RESID  59 AND NAME CB) (RESID  59 AND NAME CG) $9 180.0 15.0 2
ASSIGN (RESID  61 AND NAME N) (RESID  61 AND NAME CA)
        (RESID  61 AND NAME CB) (RESID  61 AND NAME CG) $9 180.0 15.0 2
ASSIGN (RESID  64 AND NAME N) (RESID  64 AND NAME CA)
        (RESID  64 AND NAME CB) (RESID  64 AND NAME SG) $9  60.0 15.0 2
ASSIGN (RESID  66 AND NAME N) (RESID  66 AND NAME CA)
        (RESID  66 AND NAME CB) (RESID  66 AND NAME CG) $9  60.0 15.0 2
ASSIGN (RESID  75 AND NAME N) (RESID  75 AND NAME CA)
        (RESID  75 AND NAME CB) (RESID  75 AND NAME CG) $9 -60.0 15.0 2
ASSIGN (RESID  87 AND NAME N) (RESID  87 AND NAME CA)
        (RESID  87 AND NAME CB) (RESID  87 AND NAME CG) $9 180.0 15.0 2
ASSIGN (RESID  94 AND NAME N) (RESID  94 AND NAME CA)
        (RESID  94 AND NAME CB) (RESID  94 AND NAME SG) $9 180.0 15.0 2
ASSIGN (RESID 119 AND NAME N) (RESID 119 AND NAME CA)
        (RESID 119 AND NAME CB) (RESID 119 AND NAME CG) $9 180.0 15.0 2
ASSIGN (RESID 123 AND NAME N) (RESID 123 AND NAME CA)
        (RESID 123 AND NAME CB) (RESID 123 AND NAME CG) $9 -60.0 15.0 2
ASSIGN (RESID 127 AND NAME N) (RESID 127 AND NAME CA)
        (RESID 127 AND NAME CB) (RESID 127 AND NAME SG) $9 -60.0 15.0 2

Hydrogen bond restraints for hen lysozyme.
 !* H-bonds in first helix.
ASSIGN (RESID  10 AND NAME HN) (RESID   6 AND NAME O )  1.8 0.5 0.5
ASSIGN (RESID  10 AND NAME N ) (RESID   6 AND NAME O )  2.8 0.5 0.5
ASSIGN (RESID  11 AND NAME HN) (RESID   7 AND NAME O )  1.8 0.5 0.5
ASSIGN (RESID  11 AND NAME N ) (RESID   7 AND NAME O )  2.8 0.5 0.5
ASSIGN (RESID  12 AND NAME HN) (RESID   8 AND NAME O )  1.8 0.5 0.5
ASSIGN (RESID  12 AND NAME N ) (RESID   8 AND NAME O )  2.8 0.5 0.5
ASSIGN (RESID  13 AND NAME HN) (RESID   9 AND NAME O )  1.8 0.5 0.5
ASSIGN (RESID  13 AND NAME N ) (RESID   9 AND NAME O )  2.8 0.5 0.5
ASSIGN (RESID  14 AND NAME HN) (RESID  10 AND NAME O )  1.8 0.5 0.5
ASSIGN (RESID  14 AND NAME N ) (RESID  10 AND NAME O )  2.8 0.5 0.5
ASSIGN (RESID  15 AND NAME HN) (RESID  11 AND NAME O )  1.8 0.5 0.5
ASSIGN (RESID  15 AND NAME N ) (RESID  11 AND NAME O )  2.8 0.5 0.5
!* H-bonds in second helix.
ASSIGN (RESID  29 AND NAME HN) (RESID  25 AND NAME O )  1.8 0.5 0.5
ASSIGN (RESID  29 AND NAME N ) (RESID  25 AND NAME O )  2.8 0.5 0.5
ASSIGN (RESID  30 AND NAME HN) (RESID  26 AND NAME O )  1.8 0.5 0.5
ASSIGN (RESID  30 AND NAME N ) (RESID  26 AND NAME O )  2.8 0.5 0.5
ASSIGN (RESID  31 AND NAME HN) (RESID  27 AND NAME O )  1.8 0.5 0.5
ASSIGN (RESID  31 AND NAME N ) (RESID  27 AND NAME O )  2.8 0.5 0.5
ASSIGN (RESID  32 AND NAME HN) (RESID  28 AND NAME O )  1.8 0.5 0.5
ASSIGN (RESID  32 AND NAME N ) (RESID  28 AND NAME O )  2.8 0.5 0.5
ASSIGN (RESID  33 AND NAME HN) (RESID  29 AND NAME O )  1.8 0.5 0.5
ASSIGN (RESID  33 AND NAME N ) (RESID  29 AND NAME O )  2.8 0.5 0.5
ASSIGN (RESID  34 AND NAME HN) (RESID  30 AND NAME O )  1.8 0.5 0.5
ASSIGN (RESID  34 AND NAME N ) (RESID  30 AND NAME O )  2.8 0.5 0.5
ASSIGN (RESID  35 AND NAME HN) (RESID  31 AND NAME O )  1.8 0.5 0.5
ASSIGN (RESID  35 AND NAME N ) (RESID  31 AND NAME O )  2.8 0.5 0.5
ASSIGN (RESID  36 AND NAME HN) (RESID  32 AND NAME O )  1.8 0.5 0.5
ASSIGN (RESID  36 AND NAME N ) (RESID  32 AND NAME O )  2.8 0.5 0.5
ASSIGN (RESID  37 AND NAME HN) (RESID  33 AND NAME O )  1.8 0.5 0.5
ASSIGN (RESID  37 AND NAME N ) (RESID  33 AND NAME O )  2.8 0.5 0.5
!* H-bonds for third helix.
```

```
ASSIGN (RESID  93 AND NAME HN) (RESID  89 AND NAME O )  1.8 0.5 0.5
ASSIGN (RESID  93 AND NAME N ) (RESID  89 AND NAME O )  2.8 0.5 0.5
ASSIGN (RESID  94 AND NAME HN) (RESID  90 AND NAME O )  1.8 0.5 0.5
ASSIGN (RESID  94 AND NAME N ) (RESID  90 AND NAME O )  2.8 0.5 0.5
ASSIGN (RESID  95 AND NAME HN) (RESID  91 AND NAME O )  1.8 0.5 0.5
ASSIGN (RESID  95 AND NAME N ) (RESID  91 AND NAME O )  2.8 0.5 0.5
ASSIGN (RESID  96 AND NAME HN) (RESID  92 AND NAME O )  1.8 0.5 0.5
ASSIGN (RESID  96 AND NAME N ) (RESID  92 AND NAME O )  2.8 0.5 0.5
ASSIGN (RESID  97 AND NAME HN) (RESID  93 AND NAME O )  1.8 0.5 0.5
ASSIGN (RESID  97 AND NAME N ) (RESID  93 AND NAME O )  2.8 0.5 0.5
ASSIGN (RESID  98 AND NAME HN) (RESID  94 AND NAME O )  1.8 0.5 0.5
ASSIGN (RESID  98 AND NAME N ) (RESID  94 AND NAME O )  2.8 0.5 0.5
ASSIGN (RESID  99 AND NAME HN) (RESID  95 AND NAME O )  1.8 0.5 0.5
ASSIGN (RESID  99 AND NAME N ) (RESID  95 AND NAME O )  2.8 0.5 0.5
ASSIGN (RESID 100 AND NAME HN) (RESID  96 AND NAME O )  1.8 0.5 0.5
ASSIGN (RESID 100 AND NAME N ) (RESID  96 AND NAME O )  2.8 0.5 0.5
ASSIGN (RESID 101 AND NAME HN) (RESID  97 AND NAME O )  1.8 0.5 0.5
ASSIGN (RESID 101 AND NAME N ) (RESID  97 AND NAME O )  2.8 0.5 0.5
!* H-bonds between strands 1 and 2.
ASSIGN (RESID  42 AND NAME O ) (RESID  54 AND NAME HN)  1.8 0.5 0.5
ASSIGN (RESID  42 AND NAME O ) (RESID  54 AND NAME N )  2.8 0.5 0.5
ASSIGN (RESID  44 AND NAME HN) (RESID  52 AND NAME O )  1.8 0.5 0.5
ASSIGN (RESID  44 AND NAME N ) (RESID  52 AND NAME O )  2.8 0.5 0.5
ASSIGN (RESID  44 AND NAME O ) (RESID  52 AND NAME HN)  1.8 0.5 0.5
ASSIGN (RESID  44 AND NAME O ) (RESID  52 AND NAME N )  2.8 0.5 0.5
ASSIGN (RESID  46 AND NAME HN) (RESID  50 AND NAME O )  1.8 0.5 0.5
ASSIGN (RESID  46 AND NAME N ) (RESID  50 AND NAME O )  2.8 0.5 0.5
!* H-bonds between strands 2 and 3.
ASSIGN (RESID  53 AND NAME HN) (RESID  58 AND NAME O )  1.8 0.5 0.5
ASSIGN (RESID  53 AND NAME N ) (RESID  58 AND NAME O )  2.8 0.5 0.5
ASSIGN (RESID  53 AND NAME O ) (RESID  58 AND NAME HN)  1.8 0.5 0.5
ASSIGN (RESID  53 AND NAME O ) (RESID  58 AND NAME N )  2.8 0.5 0.5
```

NH-CaH coupling contants for human interleukin-4 (in Hz).

2	---	47	<5	92	---
3	6.6	48	<4	93	<5
4	4.5	49	<5	94	---
5	<5	50	<6	95	---
6	<5	51	<6	96	8.4
7	<5	52	<5	97	6.9
8	<5	53	---	98	6.6
9	<5	54	<6	99	6.8
10	<5	55	<5	100	---
11	<5	56	5.1	101	---
12	<5	57	<5	102	8.2
13	4.4	58	7.6	103	5.4
14	<5	59	8.8	104	7.7
15	<5	60	<5	105	5.2
16	<5	61	---	106	---
17	<5	62	5.5	107	>7
18	7.4	63	<5	108	5.6
19	6.3	64	---	109	<4
20	8.3	65	8.5	110	<5
21	6.8	66	<6	111	<5
22	8.3	67	---	112	<5
23	---	68	8.8	113	<5
24	---	69	8.9	114	<5
25	6.3	70	<6	115	<5
26	---	71	<5	116	<4
27	6.5	72	5.7	117	<4
28	7.0	73	<5	118	<5
29	9.1	74	<5	119	<5
30	<5	75	<5	120	<5
31	6.8	76	<5	121	<5
32	5.4	77	<5	122	<5
33	7.4	78	4.7	123	<5
34	5.0	79	<4	124	<6
35	6.9	80	<5	125	<5
36	6.6	81	<4	126	---
37	5.8	82	<5	127	---
38	7.5	83	<5	128	7.4
39	8.0	84	<5	129	7.6
40	7.4	85	<5		
41	<5	86	<5		
42	<5	87	<5		
43	5.1	88	<5		
44	<5	89	5.1		
45	<5	90	<6		
46	<5	91	---		

NOE observed in spectrum of unlabelled and 15N labelled IL-4.

1	HIS	HA	2	LYS	HN	1	7	LEU	HA	11	ILE	HN	4
1	HIS	HB#	2	LYS	HN	2	7	LEU	HN	11	ILE	HD1#	5
2	LYS	HA	3	CYS	HN	3	7	LEU	HN	8	GLN	HN	2
2	LYS	HN	3	CYS	HN	3	7	LEU	HN	9	GLU	HN	4
3	CYS	HA	3	CYS	HN	3	7	LEU	HB#	8	GLN	HN	2
3	CYS	HN	3	CYS	HB1	3	7	LEU	HG	8	GLN	HN	2
3	CYS	HN	3	CYS	HB2	3	7	LEU	HD1#	8	GLN	HN	4
3	CYS	HA	4	ASP	HN	2	7	LEU	HD2#	8	GLN	HN	4
3	CYS	HN	4	ASP	HN	3	8	GLN	HA	8	GLN	HN	2
4	ASP	HN	3	CYS	HB1	4	8	GLN	HG1#	8	GLN	HN	2
4	ASP	HN	3	CYS	HB2	4	8	GLN	HB#	8	GLN	HN	1
4	ASP	HA	5	ILE	HN	1	8	GLN	HA	11	ILE	HN	3
4	ASP	HN	5	ILE	HN	3	8	GLN	HA	12	LYS	HN	4
4	ASP	HN	6	THR	HN	5	8	GLN	HN	9	GLU	HN	2
4	ASP	HA	4	ASP	HN	2	8	GLN	HA	9	GLU	HN	3
4	ASP	HA	7	LEU	HN	5	8	GLN	HN	10	ILE	HN	4
4	ASP	HA	8	GLN	HN	5	8	GLN	HN	11	ILE	HD#	4
4	ASP	HB1	4	ASP	HN	2	10	ILE	HN	8	GLN	HA	4
4	ASP	HB2	4	ASP	HN	2	9	GLU	HA	10	ILE	HN	3
4	ASP	HA	6	THR	HN	4	9	GLU	HA	11	ILE	HN	4
4	ASP	HN	7	LEU	HD##	4	9	GLU	HA	12	LYS	HN	3
5	ILE	HN	4	ASP	HB1	4	9	GLU	HN	10	ILE	HN	2
5	ILE	HN	4	ASP	HB2	4	9	GLU	HN	9	GLU	HA	2
6	THR	HN	4	ASP	HB1	4	9	GLU	HN	9	GLU	HB#	2
6	THR	HN	4	ASP	HB2	4	9	GLU	HN	9	GLU	HG#	2
7	LEU	HN	4	ASP	HB1	4	10	ILE	HN	9	GLU	HA	3
7	LEU	HN	4	ASP	HB2	4	10	ILE	HN	9	GLU	HB#	4
5	ILE	HA	8	GLN	HN	3	10	ILE	HN	9	GLU	HG#	4
5	ILE	HN	6	THR	HN	2	11	ILE	HN	9	GLU	HB#	4
5	ILE	HN	7	LEU	HN	5	10	ILE	HA	11	ILE	HN	4
5	ILE	HN	5	ILE	HA	2	10	ILE	HA	12	LYS	HN	4
5	ILE	HN	5	ILE	HB	1	10	ILE	HA	13	THR	HN	3
5	ILE	HN	5	ILE	HG11	3	10	ILE	HA	14	LEU	HN	4
5	ILE	HN	5	ILE	HG12	3	10	ILE	HN	11	ILE	HN	2
5	ILE	HN	5	ILE	HG2#	2	10	ILE	HN	10	ILE	HA	2
5	ILE	HN	5	ILE	HD1#	4	10	ILE	HN	10	ILE	HG2#	2
5	ILE	HN	8	GLN	HB#	4	10	ILE	HN	10	ILE	HD1#	3
5	ILE	HG11	6	THR	HN	4	10	ILE	HN	11	ILE	HD1#	5
5	ILE	HG12	6	THR	HN	4	10	ILE	HN	11	ILE	HG1#	4
5	ILE	HG2#	6	THR	HN	2	10	ILE	HN	11	ILE	HB	4
5	ILE	HB	6	THR	HN	3	11	ILE	HN	10	ILE	HG2#	3
5	ILE	HD1#	6	THR	HN	4	11	ILE	HN	10	ILE	HG1#	4
6	THR	HN	7	LEU	HN	2	11	ILE	HN	10	ILE	HB	2
6	THR	HN	8	GLN	HN	4	11	ILE	HA	12	LYS	HN	3
6	THR	HN	9	GLU	HN	5	11	ILE	HA	11	ILE	HN	2
6	THR	HN	6	THR	HA	2	11	ILE	HN	12	LYS	HN	2
6	THR	HN	6	THR	HB	2	11	ILE	HA	13	THR	HN	4
7	LEU	HN	6	THR	HB	3	11	ILE	HA	14	LEU	HN	3
6	THR	HN	6	THR	HG2#	2	11	ILE	HA	15	ASN	HN	4
6	THR	HN	7	LEU	HB#	4	11	ILE	HN	11	ILE	HB	1
6	THR	HN	7	LEU	HG	4	11	ILE	HN	11	ILE	HG11	2
10	ILE	HN	6	THR	HA	4	11	ILE	HN	11	ILE	HG12	2
10	ILE	HN	6	THR	HG2#	4	11	ILE	HN	11	ILE	HG2#	2
7	LEU	HA	8	GLN	HN	3	11	ILE	HN	11	ILE	HD1#	3
7	LEU	HA	7	LEU	HN	3	12	LYS	HN	11	ILE	HD1#	4
7	LEU	HA	10	ILE	HN	3	14	LEU	HN	11	ILE	HD1#	5

12	LYS HN	11	ILE HG2#	3	16	SER HN	18 THR HN 5
12	LYS HN	11	ILE HG11	5	16	SER HN	19 GLU HN 5
12	LYS HN	11	ILE HG12	5	16	SER HN	16 SER HB1 2
12	LYS HN	11	ILE HB	2	16	SER HN	16 SER HB2 2
12	LYS HA	13	THR HN	3	17	LEU HN	16 SER HB1 2
12	LYS HA	15	ASN HN	3	17	LEU HN	16 SER HB2 2
12	LYS HA	16	SER HN	4	17	LEU HA	17 LEU HN 2
12	LYS HN	13	THR HN	2	17	LEU HA	16 SER HN 5
12	LYS HN	14	LEU HN	4	17	LEU HA	18 THR HN 3
12	LYS HN	12	LYS HA	2	17	LEU HA	19 GLU HN 4
12	LYS HN	15	ASN HB#	5	17	LEU HN	18 THR HN 2
13	THR HA	14	LEU HN	3	17	LEU HN	19 GLU HN 4
13	THR HA	15	ASN HN	4	17	LEU HN	17 LEU HB1 2
13	THR HN	14	LEU HN	2	17	LEU HN	17 LEU HB2 2
13	THR HN	15	ASN HN	4	17	LEU HN	17 LEU HG 2
13	THR HN	13	THR HA	2	17	LEU HN	17 LEU HD1# 4
14	LEU HN	13	THR HA	3	17	LEU HN	17 LEU HD2# 4
13	THR HN	13	THR HB	2	18	THR HN	17 LEU HB1 4
14	LEU HN	13	THR HB	2	18	THR HN	17 LEU HB2 4
13	THR HN	13	THR HG2#	3	18	THR HN	17 LEU HG 5
14	LEU HN	13	THR HG2#	3	18	THR HN	17 LEU HD1# 4
13	THR HN	15	ASN HB#	5	18	THR HN	17 LEU HD2# 4
15	ASN HN	13	THR HA	4	20	GLN HN	17 LEU HA 5
15	ASN HN	13	THR HB	5	18	THR HA	19 GLU HN 3
14	LEU HA	15	ASN HN	3	18	THR HN	19 GLU HA 4
14	LEU HA	16	SER HN	4	18	THR HA	20 GLN HN 4
14	LEU HA	17	LEU HN	3	18	THR HN	20 GLN HN 5
14	LEU HN	15	ASN HN	2	18	THR HN	19 GLU HN 2
14	LEU HN	16	SER HN	4	18	THR HN	18 THR HA 2
14	LEU HN	14	LEU HA	2	18	THR HN	18 THR HG2# 2
14	LEU HN	15	ASN HB1	4	20	GLN HN	18 THR HG2# 5
14	LEU HN	15	ASN HB2	4	18	THR HN	18 THR HB 4
15	ASN HN	14	LEU HA	3	18	THR HN	19 GLU HB# 4
15	ASN HD21	14	LEU HA	5	19	GLU HN	18 THR HB 4
15	ASN HD22	14	LEU HA	5	19	GLU HN	18 THR HG2# 4
15	ASN HA	16	SER HN	4	19	GLU HN	19 GLU HA 2
15	ASN HA	18	THR HN	5	19	GLU HN	19 GLU HB# 1
15	ASN HA	17	LEU HN	5	20	GLN HN	19 GLU HB# 2
15	ASN HN	16	SER HN	2	19	GLU HN	19 GLU HG# 3
15	ASN HN	17	LEU HN	4	19	GLU HA	20 GLN HN 3
15	ASN HN	15	ASN HD21	4	19	GLU HN	20 GLN HN 2
15	ASN HN	15	ASN HD22	4	20	GLN HE#	19 GLU HB# 5
15	ASN HN	15	ASN HB1	2	21	LYS HN	19 GLU HB# 4
15	ASN HN	15	ASN HB2	2	20	GLN HA	20 GLN HN 2
15	ASN HB1	15	ASN HD21	4	20	GLN HN	20 GLN HB# 2
15	ASN HB1	15	ASN HD22	4	20	GLN HN	20 GLN HG# 3
15	ASN HB2	15	ASN HD21	4	20	GLN HA	21 LYS HN 3
15	ASN HB2	15	ASN HD22	4	20	GLN HN	21 LYS HN 4
15	ASN HN	15	ASN HA	2	20	GLN HE#	20 GLN HG# 4
16	SER HN	15	ASN HD#	5	20	GLN HB#	21 LYS HN 4
16	SER HN	15	ASN HB1	2	20	GLN HG#	21 LYS HN 5
16	SER HN	15	ASN HB2	2	21	LYS HA	22 THR HN 1
16	SER HN	16	SER HA	2	21	LYS HN	22 THR HN 4
16	SER HA	17	LEU HN	3	21	LYS HA	21 LYS HN 2
16	SER HA	18	THR HN	4	22	THR HA	23 LEU HN 2
16	SER HN	17	LEU HN	2	22	THR HN	23 LEU HN 4

22	THR	HN	22	THR	HA	3	29	VAL	HB	30	THR HN 2
22	THR	HN	22	THR	HG2#	2	29	VAL	HA	30	THR HN 1
23	LEU	HN	22	THR	HG2#	3	29	VAL	HN	30	THR HN 5
23	LEU	HN	22	THR	HB	2	29	VAL	HG1#	29	VAL HN 4
23	LEU	HA	24	CYS	HN	3	29	VAL	HG2#	29	VAL HN 4
23	LEU	HN	24	CYS	HN	5	29	VAL	HG1#	30	THR HN 3
23	LEU	HN	23	LEU	HA	2	29	VAL	HG2#	30	THR HN 3
23	LEU	HN	23	LEU	HG	2	30	THR	HA	30	THR HN 2
23	LEU	HN	23	LEU	HB1	2	30	THR	HB	30	THR HN 1
23	LEU	HN	23	LEU	HB2	2	30	THR	HG2#	30	THR HN 3
23	LEU	HN	23	LEU	HD1#	3	30	THR	HA	31	ASP HN 1
23	LEU	HN	23	LEU	HD2#	3	30	THR	HB	31	ASP HN 4
24	CYS	HN	23	LEU	HG	4	30	THR	HG2#	31	ASP HN 2
24	CYS	HN	23	LEU	HB1	3	30	THR	HG2#	32	ILE HN 3
24	CYS	HN	23	LEU	HB2	3	30	THR	HG2#	33	PHE HN 4
24	CYS	HN	23	LEU	HD1#	4	30	THR	HA	31	ASP HN 1
24	CYS	HN	23	LEU	HD2#	4	30	THR	HN	31	ASP HN 4
24	CYS	HA	24	CYS	HN	2	31	ASP	HA	31	ASP HN 3
24	CYS	HN	24	CYS	HB1	3	31	ASP	HA	34	ALA HN 4
24	CYS	HN	24	CYS	HB2	3	31	ASP	HB1	31	ASP HN 2
25	THR	HN	24	CYS	HB1	4	31	ASP	HB2	31	ASP HN 2
25	THR	HN	24	CYS	HB2	4	31	ASP	HB1	32	ILE HN 5
24	CYS	HA	25	THR	HN	3	31	ASP	HB2	32	ILE HN 5
24	CYS	HN	25	THR	HN	2	31	ASP	HB1	34	ALA HN 4
24	CYS	HN	27	LEU	HD1#	4	31	ASP	HB2	34	ALA HN 4
24	CYS	HN	27	LEU	HD2#	4	31	ASP	HA	32	ILE HN 1
25	THR	HA	27	LEU	HN	4	31	ASP	HA	33	PHE HN 3
25	THR	HB	27	LEU	HN	4	31	ASP	HN	32	ILE HN 4
25	THR	HN	25	THR	HA	3	32	ILE	HG1#	31	ASP HN 5
25	THR	HN	25	THR	HB	3	32	ILE	HD1#	31	ASP HN 5
25	THR	HN	25	THR	HG2#	2	32	ILE	HA	32	ILE HN 2
27	LEU	HN	25	THR	HG2#	5	32	ILE	HB	32	ILE HN 3
27	LEU	HA	28	THR	HN	1	32	ILE	HB	34	ALA HN 5
27	LEU	HA	27	LEU	HN	2	32	ILE	HG11	32	ILE HN 2
27	LEU	HA	29	VAL	HN	4	32	ILE	HG12	32	ILE HN 2
27	LEU	HB1	27	LEU	HN	3	32	ILE	HG2#	32	ILE HN 2
27	LEU	HB2	27	LEU	HN	3	32	ILE	HD1#	32	ILE HN 3
27	LEU	HG	27	LEU	HN	2	32	ILE	HD1#	34	ALA HN 5
27	LEU	HD1#	27	LEU	HN	3	32	ILE	HG1#	34	ALA HN 5
27	LEU	HD2#	27	LEU	HN	3	32	ILE	HA	33	PHE HN 3
27	LEU	HB1	28	THR	HN	3	32	ILE	HA	34	ALA HN 4
27	LEU	HB2	28	THR	HN	3	32	ILE	HA	35	ALA HN 4
27	LEU	HG	28	THR	HN	5	32	ILE	HA	36	SER HN 5
27	LEU	HD1#	28	THR	HN	4	32	ILE	HA	37	LYS HN 5
27	LEU	HD2#	28	THR	HN	4	32	ILE	HN	33	PHE HN 2
27	LEU	HB#	29	VAL	HN	4	32	ILE	HN	33	PHE HE# 4
27	LEU	HN	28	THR	HN	5	32	ILE	HN	33	PHE HA 5
28	THR	HA	29	VAL	HN	2	32	ILE	HN	34	ALA HN 4
28	THR	HA	28	THR	HN	2	32	ILE	HB	33	PHE HN 4
28	THR	HB	28	THR	HN	4	32	ILE	HG11	33	PHE HN 4
28	THR	HB	29	VAL	HN	2	32	ILE	HG12	33	PHE HN 4
28	THR	HG2#	28	THR	HN	2	32	ILE	HG2#	33	PHE HN 4
28	THR	HG2#	29	VAL	HN	4	32	ILE	HD1#	33	PHE HN 4
28	THR	HN	29	VAL	HN	4	32	ILE	HG2#	34	ALA HN 5
29	VAL	HA	29	VAL	HN	3	33	PHE	HA	33	PHE HN 2
29	VAL	HB	29	VAL	HN	4	34	ALA	HA	33	PHE HN 5

34	ALA	HB#	33	PHE	HN	4	39	THR	HA	38	ASN	HN	4
33	PHE	HB1	33	PHE	HN	3	38	ASN	HA	38	ASN	HN	3
33	PHE	HB2	33	PHE	HN	3	38	ASN	HB1	38	ASN	HN	4
33	PHE	HB1	34	ALA	HN	3	38	ASN	HB2	38	ASN	HN	4
33	PHE	HB2	34	ALA	HN	3	38	ASN	HB1	39	THR	HN	4
33	PHE	HD#	33	PHE	HN	2	38	ASN	HB2	39	THR	HN	4
33	PHE	HE#	33	PHE	HN	4	38	ASN	HB1	38	ASN	HD21	5
33	PHE	HA	34	ALA	HN	3	38	ASN	HB2	38	ASN	HD22	5
33	PHE	HN	34	ALA	HN	2	38	ASN	HA	39	THR	HN	2
33	PHE	HN	34	ALA	HA	5	38	ASN	HN	39	THR	HN	2
33	PHE	HN	35	ALA	HN	4	38	ASN	HN	39	THR	HA	5
33	PHE	HN	34	ALA	HB#	4	38	ASN	HN	38	ASN	HD#	5
33	PHE	HD#	34	ALA	HN	4	38	ASN	HN	39	THR	HB	5
33	PHE	HD#	35	ALA	HN	5	39	THR	HA	39	THR	HN	2
33	PHE	HD#	36	SER	HN	5	39	THR	HB	39	THR	HN	3
33	PHE	HE#	34	ALA	HN	5	39	THR	HB	40	THR	HN	2
33	PHE	HB#	35	ALA	HN	5	39	THR	HG2#	39	THR	HN	2
34	ALA	HA	34	ALA	HN	2	39	THR	HG2#	40	THR	HN	2
34	ALA	HB#	34	ALA	HN	1	39	THR	HG2#	44	THR	HN	4
34	ALA	HA	35	ALA	HN	2	39	THR	HA	40	THR	HN	1
34	ALA	HB#	35	ALA	HN	2	39	THR	HN	40	THR	HN	3
34	ALA	HB#	37	LYS	HN	4	40	THR	HA	41	GLU	HN	2
34	ALA	HA	35	ALA	HN	2	40	THR	HB	41	GLU	HN	1
34	ALA	HN	35	ALA	HN	2	40	THR	HB	42	LYS	HN	3
34	ALA	HN	36	SER	HN	5	40	THR	HB	43	GLU	HN	3
34	ALA	HA	36	SER	HN	5	40	THR	HN	43	GLU	HN	4
34	ALA	HN	36	SER	HN	5	40	THR	HN	41	GLU	HN	4
35	ALA	HA	36	SER	HN	3	40	THR	HN	42	LYS	HN	5
35	ALA	HA	35	ALA	HN	2	40	THR	HA	40	THR	HN	3
35	ALA	HA	37	LYS	HN	4	40	THR	HA	42	LYS	HN	4
35	ALA	HA	38	ASN	HN	4	40	THR	HA	43	GLU	HN	4
35	ALA	HB#	35	ALA	HN	2	40	THR	HG2#	40	THR	HN	1
35	ALA	HB#	36	SER	HN	3	40	THR	HG2#	41	GLU	HN	4
35	ALA	HN	36	SER	HN	4	40	THR	HG2#	42	LYS	HN	4
35	ALA	HN	37	LYS	HN	5	40	THR	HG2#	43	GLU	HN	4
35	ALA	HN	36	SER	HB#	5	40	THR	HG2#	44	THR	HN	5
36	SER	HA	36	SER	HN	3	43	GLU	HB#	40	THR	HN	3
36	SER	HA	38	ASN	HN	4	43	GLU	HG#	40	THR	HN	2
36	SER	HA	39	THR	HN	5	41	GLU	HA	41	GLU	HN	1
36	SER	HB1	36	SER	HN	3	41	GLU	HB#	41	GLU	HN	2
36	SER	HB2	36	SER	HN	3	41	GLU	HB#	42	LYS	HN	1
36	SER	HB1	37	LYS	HN	4	41	GLU	HA	42	LYS	HN	3
36	SER	HB2	37	LYS	HN	4	41	GLU	HN	42	LYS	HA	5
36	SER	HB#	38	ASN	HN	5	41	GLU	HA	43	GLU	HN	3
36	SER	HA	37	LYS	HN	2	41	GLU	HN	42	LYS	HN	4
36	SER	HN	37	LYS	HN	2	41	GLU	HN	43	GLU	HN	5
36	SER	HN	38	ASN	HN	5	41	GLU	HN	45	PHE	HE#	5
37	LYS	HA	37	LYS	HN	3	41	GLU	HN	44	THR	HB	5
37	LYS	HB#	37	LYS	HN	2	42	LYS	HA	42	LYS	HN	2
37	LYS	HA	38	ASN	HN	2	42	LYS	HA	46	CYS	HN	4
37	LYS	HA	39	THR	HN	5	42	LYS	HA	44	THR	HN	5
37	LYS	HN	38	ASN	HN	2	42	LYS	HA	45	PHE	HN	3
37	LYS	HN	39	THR	HN	5	42	LYS	HN	43	GLU	HN	2
37	LYS	HN	38	ASN	HA	5	42	LYS	HN	44	THR	HN	4
37	LYS	HB#	38	ASN	HN	2	42	LYS	HN	45	PHE	HN	4
37	LYS	HB#	39	THR	HN	4	42	LYS	HN	45	PHE	HD#	5

42	LYS	HN	45	PHE	HE#	5	46	CYS	HN	48	ALA	HN	4
43	GLU	HG#	42	LYS	HN	5	47	ARG	HA	47	ARG	HN	3
43	GLU	HA	43	GLU	HN	2	47	ARG	HA	48	ALA	HN	3
43	GLU	HA	45	PHE	HN	4	47	ARG	HN	48	ALA	HN	2
44	THR	HB	43	GLU	HN	4	47	ARG	HN	49	ALA	HN	5
43	GLU	HB#	43	GLU	HN	2	48	ALA	HA	48	ALA	HN	3
43	GLU	HG#	43	GLU	HN	1	48	ALA	HB#	48	ALA	HN	1
43	GLU	HB#	44	THR	HN	2	48	ALA	HA	49	ALA	HN	4
43	GLU	HG#	44	THR	HN	2	48	ALA	HN	49	ALA	HN	2
43	GLU	HB#	45	PHE	HN	4	48	ALA	HN	50	THR	HN	4
43	GLU	HG#	45	PHE	HN	4	48	ALA	HA	51	VAL	HN	4
43	GLU	HA	44	THR	HN	3	48	ALA	HA	52	LEU	HN	4
43	GLU	HA	46	CYS	HN	3	49	ALA	HA	49	ALA	HN	2
43	GLU	HN	44	THR	HN	2	49	ALA	HA	51	VAL	HN	4
43	GLU	HN	45	PHE	HN	4	49	ALA	HA	50	THR	HN	4
43	GLU	HN	44	THR	HA	5	49	ALA	HN	50	THR	HN	2
44	THR	HA	44	THR	HN	2	50	THR	HA	50	THR	HN	2
44	THR	HB	44	THR	HN	1	50	THR	HB	50	THR	HN	2
44	THR	HA	46	CYS	HN	5	50	THR	HB	51	VAL	HN	2
44	THR	HB	46	CYS	HN	5	50	THR	HN	51	VAL	HN	2
44	THR	HG2#	45	PHE	HN	3	50	THR	HG2#	50	THR	HN	3
44	THR	HG2#	44	THR	HN	3	50	THR	HG2#	51	VAL	HN	3
44	THR	HB	45	PHE	HN	2	50	THR	HG2#	52	LEU	HN	5
44	THR	HA	44	THR	HN	2	51	VAL	HA	51	VAL	HN	2
44	THR	HA	45	PHE	HN	3	51	VAL	HA	52	LEU	HN	3
44	THR	HA	47	ARG	HN	4	51	VAL	HB	51	VAL	HN	2
44	THR	HN	45	PHE	HN	2	51	VAL	HB	52	LEU	HN	2
44	THR	HN	47	ARG	HN	5	51	VAL	HG1#	51	VAL	HN	2
44	THR	HN	45	PHE	HD#	5	51	VAL	HG2#	51	VAL	HN	2
44	THR	HN	45	PHE	HE#	5	51	VAL	HG1#	52	LEU	HN	3
44	THR	HN	45	PHE	HA	5	51	VAL	HG2#	52	LEU	HN	3
44	THR	HN	45	PHE	HB1	4	51	VAL	HN	52	LEU	HN	2
44	THR	HN	45	PHE	HB2	4	52	LEU	HA	52	LEU	HN	2
45	PHE	HN	46	CYS	HN	2	52	LEU	HA	55	PHE	HN	3
45	PHE	HA	46	CYS	HN	4	52	LEU	HA	56	TYR	HN	4
45	PHE	HN	47	ARG	HN	5	54	GLN	HA	54	GLN	HN	2
45	PHE	HN	48	ALA	HN	4	54	GLN	HA	55	PHE	HN	3
45	PHE	HA	48	ALA	HN	4	54	GLN	HA	56	TYR	HN	4
45	PHE	HN	45	PHE	HD#	3	54	GLN	HN	55	PHE	HN	2
46	CYS	HN	45	PHE	HD#	5	54	GLN	HG#	55	PHE	HN	2
49	ALA	HN	45	PHE	HE#	4	54	GLN	HB#	55	PHE	HN	2
45	PHE	HN	45	PHE	HA	2	54	GLN	HG#	54	GLN	HN	2
45	PHE	HN	45	PHE	HB1	2	54	GLN	HB#	54	GLN	HN	2
45	PHE	HN	45	PHE	HB2	2	55	PHE	HA	55	PHE	HN	2
46	CYS	HN	45	PHE	HB1	3	55	PHE	HB1	55	PHE	HN	2
46	CYS	HN	45	PHE	HB2	3	55	PHE	HB2	55	PHE	HN	2
45	PHE	HN	46	CYS	HA	5	55	PHE	HB1	56	TYR	HN	3
46	CYS	HA	46	CYS	HN	2	55	PHE	HB2	56	TYR	HN	3
46	CYS	HB1	46	CYS	HN	3	55	PHE	HD#	55	PHE	HN	3
46	CYS	HB2	46	CYS	HN	3	56	TYR	HD#	55	PHE	HN	5
46	CYS	HB1	47	ARG	HN	3	55	PHE	HD#	56	TYR	HN	3
46	CYS	HB2	47	ARG	HN	3	55	PHE	HA	56	TYR	HN	4
46	CYS	HB1	48	ALA	HN	5	55	PHE	HN	56	TYR	HN	2
46	CYS	HB2	48	ALA	HN	5	55	PHE	HD#	59	HIS	HD2	3
46	CYS	HA	47	ARG	HN	4	55	PHE	HE#	59	HIS	HD2	4
46	CYS	HN	47	ARG	HN	2	55	PHE	HZ	59	HIS	HD2	5

55	PHE	HA	59	HIS	HD2	3	62	ASP	HB2	62	ASP	HN	2
55	PHE	HB1	59	HIS	HD2	4	62	ASP	HB1	63	THR	HN	4
55	PHE	HB2	59	HIS	HD2	4	62	ASP	HB2	63	THR	HN	4
56	TYR	HD#	56	TYR	HN	3	62	ASP	HA	63	THR	HN	2
56	TYR	HE#	56	TYR	HN	5	62	ASP	HA	64	ARG	HN	4
56	TYR	HD#	57	SER	HN	3	66	LEU	HD##	62	ASP	HN	4
56	TYR	HE#	57	SER	HN	4	62	ASP	HN	63	THR	HN	5
56	TYR	HA	56	TYR	HN	3	63	THR	HA	63	THR	HN	3
56	TYR	HB1	56	TYR	HN	3	63	THR	HB	63	THR	HN	3
56	TYR	HB2	56	TYR	HN	3	63	THR	HG2#	63	THR	HN	2
56	TYR	HB1	57	SER	HN	4	63	THR	HB	64	ARG	HN	4
56	TYR	HB2	57	SER	HN	4	63	THR	HG2#	64	ARG	HN	4
56	TYR	HA	57	SER	HN	4	63	THR	HA	64	ARG	HN	4
56	TYR	HA	58	HIS	HN	5	63	THR	HA	66	LEU	HN	3
56	TYR	HA	59	HIS	HN	4	63	THR	HN	64	ARG	HN	3
56	TYR	HA	60	GLU	HN	4	66	LEU	HD##	63	THR	HN	5
56	TYR	HN	57	SER	HN	2	64	ARG	HA	64	ARG	HN	2
56	TYR	HN	58	HIS	HN	5	64	ARG	HA	65	CYS	HN	4
57	SER	HA	57	SER	HN	2	64	ARG	HN	65	CYS	HN	2
57	SER	HB1	57	SER	HN	2	64	ARG	HN	66	LEU	HN	4
57	SER	HB2	57	SER	HN	2	64	ARG	HN	66	LEU	HD##	5
57	SER	HB1	58	HIS	HN	3	65	CYS	HA	65	CYS	HN	4
57	SER	HB2	58	HIS	HN	3	65	CYS	HN	66	LEU	HN	2
57	SER	HA	58	HIS	HN	4	65	CYS	HB1	65	CYS	HN	4
57	SER	HA	60	GLU	HN	4	65	CYS	HB2	65	CYS	HN	4
57	SER	HN	58	HIS	HN	2	65	CYS	HB1	66	LEU	HN	4
58	HIS	HA	58	HIS	HN	2	65	CYS	HB2	66	LEU	HN	4
58	HIS	HB1	58	HIS	HN	2	65	CYS	HA	66	LEU	HN	3
58	HIS	HB2	58	HIS	HN	2	66	LEU	HA	66	LEU	HN	3
58	HIS	HB1	59	HIS	HN	4	66	LEU	HB1	66	LEU	HN	2
58	HIS	HB2	59	HIS	HN	4	66	LEU	HB2	66	LEU	HN	2
58	HIS	HA	59	HIS	HN	4	66	LEU	HG	66	LEU	HN	2
58	HIS	HN	59	HIS	HN	2	66	LEU	HD##	66	LEU	HN	3
58	HIS	HN	60	GLU	HN	4	66	LEU	HN	67	GLY	HN	2
59	HIS	HD2	59	HIS	HN	5	66	LEU	HD##	67	GLY	HN	4
59	HIS	HA	59	HIS	HN	4	66	LEU	HB#	67	GLY	HN	4
59	HIS	HB1	59	HIS	HN	4	67	GLY	HA1	67	GLY	HN	3
59	HIS	HB2	59	HIS	HN	4	67	GLY	HA2	67	GLY	HN	3
59	HIS	HB#	60	GLU	HN	4	67	GLY	HA1	69	THR	HN	4
59	HIS	HB1	59	HIS	HD2	2	67	GLY	HA2	69	THR	HN	4
59	HIS	HB2	59	HIS	HD2	2	67	GLY	HA1	68	ALA	HN	2
59	HIS	HA	60	GLU	HN	4	67	GLY	HA2	68	ALA	HN	2
59	HIS	HA	63	THR	HN	5	67	GLY	HA1	69	THR	HN	4
59	HIS	HA	62	ASP	HN	4	67	GLY	HA2	69	THR	HN	4
59	HIS	HN	60	GLU	HN	2	67	GLY	HN	68	ALA	HN	2
60	GLU	HN	61	LYS	HN	4	68	ALA	HA	68	ALA	HN	3
60	GLU	HN	62	ASP	HN	5	68	ALA	HA	69	THR	HN	4
60	GLU	HA	60	GLU	HN	2	68	ALA	HN	69	THR	HN	2
60	GLU	HB#	60	GLU	HN	2	72	GLN	HG1	68	ALA	HN	4
60	GLU	HG#	60	GLU	HN	2	72	GLN	HG2	68	ALA	HN	4
60	GLU	HA	61	LYS	HN	2	68	ALA	HB#	68	ALA	HN	2
61	LYS	HA	61	LYS	HN	4	68	ALA	HB#	75	ARG	HN	5
61	LYS	HA	62	ASP	HN	3	69	THR	HG2#	68	ALA	HN	4
61	LYS	HN	62	ASP	HN	5	69	THR	HA	69	THR	HN	3
62	ASP	HA	62	ASP	HN	3	69	THR	HA	70	ALA	HN	3
62	ASP	HB1	62	ASP	HN	2	69	THR	HN	72	GLN	HN	4

69	THR	HN	73	PHE	HN	5	74	HIS	HB2	74	HIS	HN	2
72	GLN	HG1	69	THR	HN	4	74	HIS	HB1	75	ARG	HN	4
72	GLN	HG2	69	THR	HN	4	74	HIS	HB2	75	ARG	HN	4
68	ALA	HB#	69	THR	HN	3	74	HIS	HB#	78	GLN	HN	5
69	THR	HG2#	69	THR	HN	3	74	HIS	HA	75	ARG	HN	3
69	THR	HG2#	70	ALA	HN	4	74	HIS	HA	76	HIS	HN	5
69	THR	HG2#	71	GLN	HN	4	74	HIS	HA	77	LYS	HN	3
69	THR	HG2#	75	ARG	HN	5	74	HIS	HA	78	GLN	HN	4
69	THR	HG2#	72	GLN	HN	4	74	HIS	HN	75	ARG	HN	3
69	THR	HG2#	72	GLN	HE21	5	75	ARG	HA	75	ARG	HN	2
69	THR	HG2#	72	GLN	HE22	5	75	ARG	HA	77	LYS	HN	4
70	ALA	HA	70	ALA	HN	4	75	ARG	HA	76	HIS	HN	3
70	ALA	HB#	70	ALA	HN	2	75	ARG	HA	78	GLN	HN	3
70	ALA	HB#	71	GLN	HN	2	75	ARG	HN	76	HIS	HN	2
70	ALA	HB#	72	GLN	HN	4	75	ARG	HN	77	LYS	HN	5
70	ALA	HB#	73	PHE	HN	4	75	ARG	HN	78	GLN	HN	5
70	ALA	HB#	74	HIS	HN	5	76	HIS	HA	76	HIS	HN	3
70	ALA	HA	71	GLN	HN	4	76	HIS	HB1	76	HIS	HN	2
70	ALA	HN	71	GLN	HN	5	76	HIS	HB2	76	HIS	HN	2
71	GLN	HA	71	GLN	HN	2	76	HIS	HB#	77	LYS	HN	2
71	GLN	HB#	71	GLN	HN	2	76	HIS	HB#	78	GLN	HN	5
71	GLN	HB#	72	GLN	HN	3	76	HIS	HD1	76	HIS	HN	5
71	GLN	HG#	71	GLN	HN	3	76	HIS	HD1	77	LYS	HN	4
71	GLN	HE#	71	GLN	HN	4	76	HIS	HA	77	LYS	HN	4
71	GLN	HA	72	GLN	HN	4	76	HIS	HN	77	LYS	HN	2
71	GLN	HA	75	ARG	HN	5	76	HIS	HN	78	GLN	HN	4
71	GLN	HN	72	GLN	HN	2	77	LYS	HA	77	LYS	HN	2
72	GLN	HG#	71	GLN	HN	5	77	LYS	HA	78	GLN	HN	4
71	GLN	HG#	75	ARG	HN	5	77	LYS	HA	80	ILE	HN	3
72	GLN	HA	72	GLN	HN	2	77	LYS	HA	79	LEU	HN	5
72	GLN	HG1	72	GLN	HN	2	77	LYS	HA	81	ARG	HN	4
72	GLN	HG2	72	GLN	HN	2	77	LYS	HN	78	GLN	HN	2
72	GLN	HG#	73	PHE	HN	2	77	LYS	HN	79	LEU	HN	5
72	GLN	HE#	72	GLN	HN	5	78	GLN	HA	78	GLN	HN	2
73	PHE	HA	72	GLN	HN	5	78	GLN	HA	79	LEU	HN	3
73	PHE	HB#	72	GLN	HN	5	78	GLN	HA	80	ILE	HN	5
72	GLN	HA	74	HIS	HN	4	79	LEU	HB#	78	GLN	HN	4
72	GLN	HA	75	ARG	HN	3	78	GLN	HN	79	LEU	HN	2
72	GLN	HA	76	HIS	HN	4	78	GLN	HN	80	ILE	HN	4
72	GLN	HN	73	PHE	HN	2	79	LEU	HA	79	LEU	HN	2
72	GLN	HN	74	HIS	HN	5	79	LEU	HB1	79	LEU	HN	2
72	GLN	HN	75	ARG	HN	5	79	LEU	HB2	79	LEU	HN	2
73	PHE	HA	73	PHE	HN	2	79	LEU	HB1	80	ILE	HN	3
73	PHE	HB1	73	PHE	HN	2	79	LEU	HB2	80	ILE	HN	3
73	PHE	HB2	73	PHE	HN	2	79	LEU	HD1#	79	LEU	HN	3
73	PHE	HB1	74	HIS	HN	4	79	LEU	HD2#	79	LEU	HN	3
73	PHE	HB2	74	HIS	HN	4	79	LEU	HD##	80	ILE	HN	4
73	PHE	HB#	75	ARG	HN	5	79	LEU	HA	80	ILE	HN	4
73	PHE	HD#	73	PHE	HN	4	79	LEU	HA	82	PHE	HN	3
73	PHE	HD#	74	HIS	HN	4	79	LEU	HA	83	LEU	HN	4
73	PHE	HA	74	HIS	HN	3	79	LEU	HN	80	ILE	HN	2
73	PHE	HA	75	ARG	HN	5	82	PHE	HD#	79	LEU	HN	5
73	PHE	HN	75	ARG	HN	4	80	ILE	HG2#	79	LEU	HN	5
73	PHE	HN	74	HIS	HN	3	80	ILE	HD1#	79	LEU	HN	5
74	HIS	HA	74	HIS	HN	2	82	PHE	HB#	79	LEU	HN	5
74	HIS	HB1	74	HIS	HN	2	80	ILE	HA	80	ILE	HN	3

80	ILE	HB	80	ILE	HN	2	84	LYS	HN	86	LEU	HN	4
80	ILE	HG11	80	ILE	HN	2	85	ARG	HA	86	LEU	HN	4
80	ILE	HG12	80	ILE	HN	2	85	ARG	HA	85	ARG	HN	2
80	ILE	HB	81	ARG	HN	3	85	ARG	HA	87	ASP	HN	4
80	ILE	HG1#	81	ARG	HN	3	85	ARG	HN	86	LEU	HN	2
80	ILE	HG2#	80	ILE	HN	3	85	ARG	HN	87	ASP	HN	4
80	ILE	HG2#	81	ARG	HN	3	86	LEU	HA	86	LEU	HN	2
80	ILE	HD1#	80	ILE	HN	3	86	LEU	HB#	86	LEU	HN	2
80	ILE	HA	81	ARG	HN	4	86	LEU	HG	86	LEU	HN	2
80	ILE	HA	82	PHE	HN	4	86	LEU	HD1#	86	LEU	HN	4
80	ILE	HA	83	LEU	HN	3	86	LEU	HD2#	86	LEU	HN	4
80	ILE	HA	84	LYS	HN	5	86	LEU	HB#	87	ASP	HN	3
80	ILE	HN	81	ARG	HN	2	86	LEU	HG	87	ASP	HN	3
80	ILE	HN	82	PHE	HN	4	86	LEU	HD1#	87	ASP	HN	4
81	ARG	HA	81	ARG	HN	2	86	LEU	HD2#	87	ASP	HN	4
82	PHE	HD#	81	ARG	HN	5	86	LEU	HA	87	ASP	HN	3
82	PHE	HB#	81	ARG	HN	4	86	LEU	HN	87	ASP	HN	3
81	ARG	HA	82	PHE	HN	3	86	LEU	HN	88	ARG	HN	5
81	ARG	HA	83	LEU	HN	4	87	ASP	HA	87	ASP	HN	2
81	ARG	HA	84	LYS	HN	3	87	ASP	HB1	87	ASP	HN	2
81	ARG	HN	82	PHE	HN	2	87	ASP	HB2	87	ASP	HN	2
81	ARG	HN	83	LEU	HN	4	87	ASP	HB1	88	ARG	HN	3
82	PHE	HA	82	PHE	HN	2	87	ASP	HB2	88	ARG	HN	3
82	PHE	HE#	82	PHE	HN	5	87	ASP	HA	88	ARG	HN	4
82	PHE	HD#	82	PHE	HN	4	87	ASP	HA	90	LEU	HN	5
82	PHE	HD#	83	LEU	HN	5	87	ASP	HN	88	ARG	HN	3
82	PHE	HD#	85	ARG	HN	4	87	ASP	HN	89	ASN	HN	5
82	PHE	HB1	82	PHE	HN	2	88	ARG	HA	88	ARG	HN	2
82	PHE	HB2	82	PHE	HN	2	88	ARG	HA	89	ASN	HN	3
82	PHE	HB1	83	LEU	HN	3	88	ARG	HA	89	ASN	HD#	4
82	PHE	HB2	83	LEU	HN	3	88	ARG	HN	89	ASN	HN	2
82	PHE	HA	83	LEU	HN	4	88	ARG	HN	90	LEU	HN	5
82	PHE	HA	84	LYS	HN	5	89	ASN	HA	89	ASN	HN	2
82	PHE	HA	85	ARG	HN	4	89	ASN	HD21	89	ASN	HN	4
82	PHE	HA	86	LEU	HN	4	89	ASN	HD22	89	ASN	HN	4
82	PHE	HN	83	LEU	HN	2	89	ASN	HD21	90	LEU	HN	5
82	PHE	HN	84	LYS	HN	4	89	ASN	HD22	90	LEU	HN	5
83	LEU	HA	83	LEU	HN	3	89	ASN	HB1	89	ASN	HN	3
83	LEU	HB#	83	LEU	HN	2	89	ASN	HB2	89	ASN	HN	3
83	LEU	HB#	84	LYS	HN	3	89	ASN	HB1	90	LEU	HN	4
83	LEU	HG	83	LEU	HN	2	89	ASN	HB2	90	LEU	HN	4
83	LEU	HD1#	83	LEU	HN	3	89	ASN	HA	90	LEU	HN	4
83	LEU	HD2#	83	LEU	HN	3	89	ASN	HA	92	GLY	HN	4
83	LEU	HD1#	84	LYS	HN	4	89	ASN	HN	90	LEU	HN	2
83	LEU	HD2#	84	LYS	HN	4	89	ASN	HA	89	ASN	HD#	4
83	LEU	HA	84	LYS	HN	4	89	ASN	HB#	89	ASN	HD#	4
83	LEU	HN	84	LYS	HN	2	90	LEU	HA	90	LEU	HN	3
83	LEU	HN	85	ARG	HN	4	90	LEU	HA	91	TRP	HN	5
83	LEU	HA	85	ARG	HN	4	90	LEU	HN	91	TRP	HN	3
84	LYS	HA	84	LYS	HN	2	91	TRP	HE3	90	LEU	HD##	5
85	ARG	HA	84	LYS	HN	5	91	TRP	HA	91	TRP	HN	4
84	LYS	HA	85	ARG	HN	3	91	TRP	HE3	91	TRP	HA	3
84	LYS	HA	86	LEU	HN	5	91	TRP	HE3	91	TRP	HB1	3
84	LYS	HA	87	ASP	HN	3	91	TRP	HE3	91	TRP	HB2	3
84	LYS	HA	88	ARG	HN	4	91	TRP	HD1	91	TRP	HA	3
84	LYS	HN	85	ARG	HN	2	91	TRP	HD1	91	TRP	HB1	3

91	TRP	HD1	91	TRP	HB2	3	99	CYS	HA	99	CYS	HN	4
91	TRP	HE3	91	TRP	HN	5	99	CYS	HB1	99	CYS	HN	5
91	TRP	HE3	92	GLY	HN	5	99	CYS	HB2	99	CYS	HN	5
91	TRP	HE3	93	LEU	HN	5	99	CYS	HA	100	PRO	HD#	3
91	TRP	HB#	91	TRP	HN	3	100	PRO	HD1	99	CYS	HN	4
91	TRP	HB#	92	GLY	HN	3	100	PRO	HD2	99	CYS	HN	4
91	TRP	HA	92	GLY	HN	4	100	PRO	HA	101	VAL	HN	1
91	TRP	HN	92	GLY	HN	4	101	VAL	HA	101	VAL	HN	2
92	GLY	HA1	92	GLY	HN	3	101	VAL	HB	101	VAL	HN	3
92	GLY	HA2	92	GLY	HN	3	101	VAL	HG1#	101	VAL	HN	3
92	GLY	HA#	95	GLY	HN	5	101	VAL	HG2#	101	VAL	HN	3
92	GLY	HA1	93	LEU	HN	3	101	VAL	HB	102	LYS	HN	4
92	GLY	HA2	93	LEU	HN	3	101	VAL	HG1#	102	LYS	HN	4
92	GLY	HA#	94	ALA	HN	4	101	VAL	HG2#	102	LYS	HN	4
92	GLY	HN	93	LEU	HN	3	101	VAL	HG##	103	GLU	HN	4
92	GLY	HN	94	ALA	HN	5	101	VAL	HA	102	LYS	HN	2
93	LEU	HD##	92	GLY	HN	5	101	VAL	HN	102	LYS	HN	4
93	LEU	HA	93	LEU	HN	2	101	VAL	HA	103	GLU	HN	5
93	LEU	HA	94	ALA	HN	3	102	LYS	HA	102	LYS	HN	4
93	LEU	HN	94	ALA	HN	3	102	LYS	HB#	102	LYS	HN	4
94	ALA	HA	94	ALA	HN	2	102	LYS	HB#	103	GLU	HN	4
94	ALA	HB#	94	ALA	HN	2	103	GLU	HG#	102	LYS	HN	5
94	ALA	HB#	95	GLY	HN	3	102	LYS	HA	103	GLU	HN	1
94	ALA	HA	95	GLY	HN	4	102	LYS	HN	103	GLU	HN	4
94	ALA	HA	96	LEU	HN	5	103	GLU	HA	103	GLU	HN	3
94	ALA	HN	95	GLY	HN	3	103	GLU	HB#	103	GLU	HN	2
94	ALA	HN	96	LEU	HN	5	103	GLU	HG#	103	GLU	HN	4
95	GLY	HA1	95	GLY	HN	2	103	GLU	HB#	104	ALA	HN	4
95	GLY	HA2	95	GLY	HN	2	103	GLU	HB#	105	ASN	HN	4
95	GLY	HA1	96	LEU	HN	4	103	GLU	HG#	104	ALA	HN	5
95	GLY	HA2	96	LEU	HN	4	103	GLU	HG#	105	ASN	HN	3
95	GLY	HN	96	LEU	HN	3	103	GLU	HA	104	ALA	HN	1
96	LEU	HB#	95	GLY	HN	5	103	GLU	HA	105	ASN	HN	4
96	LEU	HD##	95	GLY	HN	5	103	GLU	HN	104	ALA	HN	5
96	LEU	HA	96	LEU	HN	4	104	ALA	HA	104	ALA	HN	4
96	LEU	HB#	96	LEU	HN	2	104	ALA	HB#	104	ALA	HN	2
96	LEU	HG	96	LEU	HN	3	104	ALA	HB#	105	ASN	HN	1
96	LEU	HB#	97	ASN	HN	4	104	ALA	HB#	105	ASN	HD21	4
96	LEU	HB#	98	SER	HN	3	104	ALA	HB#	105	ASN	HD22	4
96	LEU	HD##	96	LEU	HN	4	104	ALA	HB#	106	GLN	HE#	5
96	LEU	HD##	97	ASN	HN	4	104	ALA	HA	105	ASN	HN	3
96	LEU	HD##	98	SER	HN	4	104	ALA	HN	105	ASN	HN	3
96	LEU	HD##	99	CYS	HN	5	105	ASN	HA	105	ASN	HN	1
96	LEU	HA	97	ASN	HN	3	106	GLN	HG#	105	ASN	HN	5
96	LEU	HN	97	ASN	HN	5	106	GLN	HE#	105	ASN	HN	5
96	LEU	HA	98	SER	HN	5	105	ASN	HB1	105	ASN	HN	3
97	ASN	HA	97	ASN	HN	4	105	ASN	HB2	105	ASN	HN	3
97	ASN	HB#	97	ASN	HN	3	105	ASN	HB#	106	GLN	HN	5
97	ASN	HB#	98	SER	HN	4	105	ASN	HB1	105	ASN	HD#	4
97	ASN	HA	98	SER	HN	4	105	ASN	HB2	105	ASN	HD#	4
97	ASN	HN	98	SER	HN	4	105	ASN	HD21	105	ASN	HN	4
98	SER	HA	98	SER	HN	3	105	ASN	HD22	105	ASN	HN	4
98	SER	HB#	98	SER	HN	3	105	ASN	HA	106	GLN	HN	3
98	SER	HB#	99	CYS	HN	4	105	ASN	HN	106	GLN	HN	5
98	SER	HA	99	CYS	HN	2	106	GLN	HA	106	GLN	HN	5
98	SER	HN	99	CYS	HN	5	106	GLN	HA	106	GLN	HE#	5

106	GLN	HB#	106	GLN	HN	4	110	GLU	HA	112	PHE	HN	5
106	GLN	HB#	108	THR	HN	5	110	GLU	HA	113	LEU	HN	3
106	GLN	HG#	106	GLN	HN	4	110	GLU	HN	111	ASN	HN	3
106	GLN	HB#	107	SER	HN	4	110	GLU	HN	112	PHE	HN	5
106	GLN	HG#	107	SER	HN	5	111	ASN	HA	111	ASN	HN	2
106	GLN	HG#	108	THR	HN	5	111	ASN	HA	111	ASN	HD#	5
106	GLN	HG#	106	GLN	HE#	3	111	ASN	HD21	111	ASN	HN	5
106	GLN	HA	107	SER	HN	3	111	ASN	HD22	111	ASN	HN	5
106	GLN	HN	107	SER	HN	5	111	ASN	HB1	111	ASN	HN	2
107	SER	HA	107	SER	HN	3	111	ASN	HB2	111	ASN	HN	2
107	SER	HB1	107	SER	HN	4	111	ASN	HB1	114	GLU	HN	5
107	SER	HB2	107	SER	HN	4	111	ASN	HB2	114	GLU	HN	5
107	SER	HB1	108	THR	HN	4	111	ASN	HB1	112	PHE	HN	4
107	SER	HB2	108	THR	HN	4	111	ASN	HB2	112	PHE	HN	4
107	SER	HA	108	THR	HN	1	112	PHE	HD#	111	ASN	HN	5
107	SER	HA	109	LEU	HN	4	112	PHE	HB1	111	ASN	HN	5
107	SER	HN	108	THR	HN	5	112	PHE	HB2	111	ASN	HN	5
108	THR	HA	108	THR	HN	3	111	ASN	HA	114	GLU	HN	4
108	THR	HB	108	THR	HN	5	111	ASN	HN	112	PHE	HN	2
108	THR	HG2#	108	THR	HN	1	111	ASN	HN	113	LEU	HN	4
108	THR	HG2#	111	ASN	HN	5	112	PHE	HA	112	PHE	HN	2
108	THR	HG2#	109	LEU	HN	4	112	PHE	HB1	112	PHE	HN	2
108	THR	HG2#	111	ASN	HD#	5	112	PHE	HB2	112	PHE	HN	2
111	ASN	HB1	108	THR	HN	4	112	PHE	HB1	113	LEU	HN	4
111	ASN	HB2	108	THR	HN	4	112	PHE	HB2	113	LEU	HN	4
108	THR	HA	109	LEU	HN	1	112	PHE	HB#	115	ARG	HN	5
108	THR	HB	109	LEU	HN	3	112	PHE	HD#	112	PHE	HN	4
108	THR	HB	111	ASN	HN	5	112	PHE	HD#	113	LEU	HN	4
108	THR	HA	110	GLU	HN	5	112	PHE	HD#	114	GLU	HN	5
108	THR	HN	111	ASN	HN	5	112	PHE	HD#	116	LEU	HN	5
108	THR	HN	111	ASN	HD#	5	112	PHE	HA	113	LEU	HN	4
109	LEU	HA	109	LEU	HN	3	112	PHE	HA	115	ARG	HN	4
109	LEU	HB1	109	LEU	HN	3	112	PHE	HN	113	LEU	HN	3
109	LEU	HB2	109	LEU	HN	3	112	PHE	HN	114	GLU	HN	5
109	LEU	HG	109	LEU	HN	5	113	LEU	HA	113	LEU	HN	3
109	LEU	HD1#	109	LEU	HN	4	113	LEU	HA	114	GLU	HN	4
109	LEU	HD2#	109	LEU	HN	4	113	LEU	HA	116	LEU	HN	4
109	LEU	HB1	110	GLU	HN	4	113	LEU	HA	117	LYS	HN	4
109	LEU	HB2	110	GLU	HN	4	113	LEU	HN	114	GLU	HN	3
109	LEU	HG	110	GLU	HN	4	113	LEU	HN	115	ARG	HN	5
109	LEU	HD1#	110	GLU	HN	5	114	GLU	HG#	113	LEU	HN	5
109	LEU	HD2#	110	GLU	HN	5	114	GLU	HA	114	GLU	HN	3
109	LEU	HD1#	113	LEU	HN	4	114	GLU	HG#	114	GLU	HN	2
109	LEU	HD2#	113	LEU	HN	4	114	GLU	HG#	115	ARG	HN	4
109	LEU	HA	112	PHE	HN	4	114	GLU	HA	115	ARG	HN	4
109	LEU	HN	110	GLU	HN	3	114	GLU	HA	117	LYS	HN	3
110	GLU	HA	110	GLU	HN	4	114	GLU	HN	115	ARG	HN	2
110	GLU	HB#	110	GLU	HN	3	114	GLU	HN	115	ARG	HE#	5
110	GLU	HG#	110	GLU	HN	4	114	GLU	HN	117	LYS	HN	5
110	GLU	HB1	111	ASN	HN	4	114	GLU	HA	118	THR	HN	4
110	GLU	HB2	111	ASN	HN	4	115	ARG	HA	115	ARG	HN	2
110	GLU	HG#	111	ASN	HN	4	115	ARG	HA	117	LYS	HN	4
111	ASN	HB1	110	GLU	HN	5	115	ARG	HA	116	LEU	HN	4
111	ASN	HB2	110	GLU	HN	5	115	ARG	HN	116	LEU	HN	2
112	PHE	HB#	110	GLU	HN	5	115	ARG	HN	117	LYS	HN	5
110	GLU	HA	111	ASN	HN	4	115	ARG	HN	115	ARG	HE#	5

116	LEU	HN	115	ARG	HE#	5	121	ARG	HA	122	GLU	HN	4
116	LEU	HA	116	LEU	HN	3	121	ARG	HA	123	LYS	HN	5
117	LYS	HA	116	LEU	HN	5	121	ARG	HA	124	TYR	HN	4
116	LEU	HA	117	LYS	HN	4	121	ARG	HN	122	GLU	HN	2
116	LEU	HN	117	LYS	HN	2	121	ARG	HN	123	LYS	HN	5
116	LEU	HN	118	THR	HN	5	122	GLU	HA	122	GLU	HN	2
117	LYS	HA	117	LYS	HN	2	122	GLU	HB#	122	GLU	HN	2
117	LYS	HA	118	THR	HN	4	122	GLU	HG#	122	GLU	HN	4
117	LYS	HA	120	MET	HN	4	122	GLU	HN	124	TYR	HN	5
117	LYS	HA	121	ARG	HN	5	122	GLU	HN	123	LYS	HN	2
117	LYS	HN	118	THR	HN	2	123	LYS	HA	123	LYS	HN	2
117	LYS	HN	119	ILE	HN	5	124	TYR	HA	123	LYS	HN	5
118	THR	HA	118	THR	HN	2	124	TYR	HB1	123	LYS	HN	5
118	THR	HB	118	THR	HN	1	124	TYR	HB2	123	LYS	HN	5
118	THR	HB	119	ILE	HN	4	123	LYS	HA	124	TYR	HN	4
118	THR	HG2#	118	THR	HN	3	123	LYS	HN	124	TYR	HN	3
118	THR	HG2#	119	ILE	HN	5	123	LYS	HN	125	SER	HN	4
118	THR	HA	121	ARG	HN	4	123	LYS	HN	126	LYS	HN	5
118	THR	HN	119	ILE	HN	2	124	TYR	HA	124	TYR	HN	3
118	THR	HN	120	MET	HN	4	124	TYR	HD#	124	TYR	HN	4
119	ILE	HA	119	ILE	HN	4	124	TYR	HD#	125	SER	HN	5
119	ILE	HB	119	ILE	HN	3	124	TYR	HB#	124	TYR	HN	2
119	ILE	HG11	119	ILE	HN	4	124	TYR	HB#	125	SER	HN	4
119	ILE	HG12	119	ILE	HN	4	124	TYR	HA	125	SER	HN	4
119	ILE	HG2#	119	ILE	HN	4	124	TYR	HA	127	CYS	HN	5
119	ILE	HD1#	119	ILE	HN	5	124	TYR	HN	125	SER	HN	3
119	ILE	HB	120	MET	HN	3	125	SER	HA	125	SER	HN	3
119	ILE	HG11	120	MET	HN	5	125	SER	HB#	125	SER	HN	2
119	ILE	HG12	120	MET	HN	5	125	SER	HB#	126	LYS	HN	4
119	ILE	HG2#	120	MET	HN	3	125	SER	HB#	127	CYS	HN	5
119	ILE	HD1#	120	MET	HN	4	125	SER	HA	126	LYS	HN	4
119	ILE	HD1#	121	ARG	HN	5	125	SER	HA	127	CYS	HN	5
120	MET	HA	119	ILE	HN	5	125	SER	HA	128	SER	HN	5
120	MET	HG#	119	ILE	HN	5	125	SER	HN	126	LYS	HN	3
119	ILE	HA	120	MET	HN	4	125	SER	HN	127	CYS	HN	4
119	ILE	HA	122	GLU	HN	4	126	LYS	HA	126	LYS	HN	3
119	ILE	HA	123	LYS	HN	4	126	LYS	HA	127	CYS	HN	4
119	ILE	HN	120	MET	HN	3	126	LYS	HA	128	SER	HN	5
119	ILE	HN	121	ARG	HN	5	126	LYS	HN	127	CYS	HN	3
120	MET	HA	120	MET	HN	3	127	CYS	HA	127	CYS	HN	3
120	MET	HB1	120	MET	HN	3	127	CYS	HB#	127	CYS	HN	3
120	MET	HB2	120	MET	HN	3	127	CYS	HB#	128	SER	HN	4
120	MET	HG1	120	MET	HN	4	127	CYS	HA	128	SER	HN	3
120	MET	HG2	120	MET	HN	4	127	CYS	HN	128	SER	HN	3
120	MET	HB1	121	ARG	HN	5	128	SER	HA	128	SER	HN	4
120	MET	HB2	121	ARG	HN	5	128	SER	HB1	128	SER	HN	4
120	MET	HG1	121	ARG	HN	5	128	SER	HB2	128	SER	HN	4
120	MET	HG2	121	ARG	HN	5	128	SER	HA	129	SER	HN	3
121	ARG	HA	120	MET	HN	5	128	SER	HN	129	SER	HN	4
120	MET	HA	121	ARG	HN	4	129	SER	HA	129	SER	HN	3
120	MET	HA	123	LYS	HN	4	129	SER	HB#	129	SER	HN	4
120	MET	HA	122	GLU	HN	5	29	VAL	HN	107	SER	HN	4
120	MET	HN	121	ARG	HN	3	31	ASP	HN	107	SER	HN	4
120	MET	HN	122	GLU	HN	4	31	ASP	HN	106	GLN	HA	2
121	ARG	HA	121	ARG	HN	3	31	ASP	HN	106	GLN	HB#	4
122	GLU	HA	121	ARG	HN	5	31	ASP	HN	106	GLN	HE#	5

31	ASP	HN	105	ASN	HB#	5	112	PHE	HD#	32	ILE	HN	3
31	ASP	HN	104	ALA	HB#	4	105	ASN	HB#	34	ALA	HN	5
29	VAL	HG##	107	SER	HN	5	96	LEU	HD##	42	LYS	HN	5
28	THR	HG2#	107	SER	HN	5	94	ALA	HA	42	LYS	HN	5
29	VAL	HN	108	THR	HA	4	96	LEU	HD##	43	GLU	HN	4
28	THR	HG2#	108	THR	HN	5	96	LEU	HD##	44	THR	HN	5
29	VAL	HN	109	LEU	HN	4	94	ALA	HA	45	PHE	HN	4
28	THR	HA	109	LEU	HN	4	96	LEU	HD##	45	PHE	HN	5
28	THR	HB	109	LEU	HN	4	94	ALA	HA	46	CYS	HN	5
28	THR	HG2#	109	LEU	HN	4	96	LEU	HD##	46	CYS	HN	4
30	THR	HG2#	104	ALA	HN	4	96	LEU	HB#	46	CYS	HN	5
30	THR	HG2#	103	GLU	HN	5	29	VAL	HG1#	55	PHE	HN	5
30	THR	HG2#	105	ASN	HN	4	29	VAL	HG2#	55	PHE	HN	5
109	LEU	HN	29	VAL	HG##	4	29	VAL	HG##	56	TYR	HN	4
120	MET	HN	11	ILE	HD#	5	109	LEU	HD1#	59	HIS	HD2	5
120	MET	HN	11	ILE	HG1#	5	109	LEU	HD2#	59	HIS	HD2	5
120	MET	HN	10	ILE	HG2#	5	27	LEU	HD1#	59	HIS	HD2	4
121	ARG	HN	11	ILE	HD#	5	27	LEU	HD2#	59	HIS	HD2	4
124	TYR	HN	11	ILE	HD#	5	23	LEU	HD##	67	GLY	HN	4
109	LEU	HD1#	55	PHE	HD#	5	23	LEU	HD##	68	ALA	HN	4
109	LEU	HD1#	55	PHE	HE#	5	23	LEU	HD##	72	GLN	HE21	4
109	LEU	HD2#	55	PHE	HD#	5	23	LEU	HD##	72	GLN	HE22	4
109	LEU	HD2#	55	PHE	HE#	5	22	THR	HG2#	75	ARG	HN	5
109	LEU	HD1#	55	PHE	HZ	5	56	TYR	HD#	83	LEU	HN	5
109	LEU	HD2#	55	PHE	HZ	5	56	TYR	HD#	84	LYS	HN	4
109	LEU	HD##	56	TYR	HE#	5	56	TYR	HE#	84	LYS	HN	5
109	LEU	HD##	112	PHE	HD#	5	56	TYR	HE#	87	ASP	HN	4
32	ILE	HD#	33	PHE	HD#	5	56	TYR	HB#	87	ASP	HN	5
32	ILE	HD#	33	PHE	HE#	4	56	TYR	HE#	88	ARG	HN	5
32	ILE	HD#	33	PHE	HZ	5	45	PHE	HD#	94	ALA	HN	5
93	LEU	HD1#	45	PHE	HD#	5	30	THR	HA	105	ASN	HN	5
93	LEU	HD2#	45	PHE	HD#	5	30	THR	HA	106	GLN	HE#	4
93	LEU	HD1#	45	PHE	HE#	5	30	THR	HG2#	106	GLN	HE#	4
93	LEU	HD2#	45	PHE	HE#	5	30	THR	HA	107	SER	HN	5
11	ILE	HD#	124	TYR	HD#	4	28	THR	HB	107	SER	HN	5
11	ILE	HD#	124	TYR	HE#	5	29	VAL	HG##	108	THR	HN	5
27	LEU	HD1#	55	PHE	HE#	4	33	PHE	HD#	115	ARG	HN	5
27	LEU	HD2#	55	PHE	HE#	4	33	PHE	HE#	115	ARG	HN	5
27	LEU	HD1#	55	PHE	HZ	4	33	PHE	HE#	116	LEU	HN	4
27	LEU	HD2#	55	PHE	HZ	4	33	PHE	HZ	116	LEU	HN	5
27	LEU	HD1#	24	CYS	HA	5	33	PHE	HZ	117	LYS	HN	5
27	LEU	HD2#	24	CYS	HA	5	11	ILE	HD1#	125	SER	HN	5
27	LEU	HD1#	24	CYS	HB1	5	10	ILE	HD1#	89	ASN	HN	5
27	LEU	HD2#	24	CYS	HB1	5	93	LEU	HD##	6	THR	HN	5
27	LEU	HD1#	24	CYS	HB2	5	91	TRP	HZ2	97	ASN	HA	3
27	LEU	HD2#	24	CYS	HB2	5	91	TRP	HH2	97	ASN	HA	5
119	ILE	HD#	33	PHE	HD#	4	91	TRP	HZ2	97	ASN	HB#	4
119	ILE	HD#	33	PHE	HE#	4	3	CYS	HN	128	SER	HN	5
119	ILE	HD#	33	PHE	HZ	5	32	ILE	HN	112	PHE	HD#	3
29	VAL	HG1#	107	SER	HA	5	32	ILE	HN	103	GLU	HA	5
29	VAL	HG2#	107	SER	HA	5							
8	GLN	HN	124	TYR	HE#	5							
8	GLN	HN	124	TYR	HD#	5							
109	LEU	HD##	27	LEU	HN	5							
109	LEU	HB#	27	LEU	HN	5							
107	SER	HB#	29	VAL	HN	5							

Additional NOE data from double labelled sample.

4	ASP	HA	5	ILE	HG2#	4
5	ILE	HD1#	8	GLN	HB#	2
5	ILE	HG2#	8	GLN	HB#	4
5	ILE	HG2#	6	THR	HA	4
5	ILE	HA	8	GLN	HB#	2
5	ILE	HA	8	GLN	HG#	4
6	THR	HG2#	89	ASN	HB#	2
6	THR	HG2#	93	LEU	HD1#	4
6	THR	HG2#	93	LEU	HD2#	4
6	THR	HB	93	LEU	HD1#	2
6	THR	HB	93	LEU	HD2#	2
6	THR	HB	10	ILE	HD1#	2
6	THR	HG2#	10	ILE	HD1#	2
6	THR	HG2#	93	LEU	HA	4
6	THR	HA	9	GLU	HB#	2
6	THR	HA	9	GLU	HG#	4
6	THR	HG2#	89	ASN	HA	2
7	LEU	HA	10	ILE	HD1#	2
7	LEU	HA	10	ILE	HG2#	4
7	LEU	HD1#	124	TYR	HA	4
7	LEU	HD1#	10	ILE	HG2#	2
7	LEU	HD1#	10	ILE	HD1#	2
7	LEU	HD1#	93	LEU	HD1#	4
7	LEU	HD1#	93	LEU	HD2#	4
7	LEU	HD2#	124	TYR	HA	4
7	LEU	HD2#	10	ILE	HG2#	2
7	LEU	HD2#	10	ILE	HD1#	2
7	LEU	HD2#	93	LEU	HD1#	4
7	LEU	HD2#	93	LEU	HD2#	4
7	LEU	HA	93	LEU	HD1#	4
7	LEU	HA	93	LEU	HD2#	4
7	LEU	HA	10	ILE	HB	2
8	GLN	HA	11	ILE	HD1#	2
8	GLN	HA	11	ILE	HG2#	2
8	GLN	HA	11	ILE	HB	2
8	GLN	HA	11	ILE	HG11	4
8	GLN	HA	11	ILE	HG12	4
9	GLU	HA	12	LYS	HB#	2
10	ILE	HD1#	89	ASN	HA	4
10	ILE	HD1#	89	ASN	HB#	2
10	ILE	HD1#	93	LEU	HD*	2
10	ILE	HD1#	86	LEU	HD*	4
10	ILE	HA	86	LEU	HD*	4
10	ILE	HG2#	93	LEU	HD*	2
10	ILE	HG2#	11	ILE	HA	2
10	ILE	HA	13	THR	HB	2
10	ILE	HD1#	120	MET	HE#	4
10	ILE	HG2#	120	MET	HE#	2
11	ILE	HD1#	124	TYR	HA	2
11	ILE	HD1#	124	TYR	HB1	4
11	ILE	HD1#	124	TYR	HB2	4
11	ILE	HD1#	121	ARG	HA	2
11	ILE	HD1#	121	ARG	HD#	4
11	ILE	HG2#	12	LYS	HA	2
11	ILE	HA	14	LEU	HD1#	2
11	ILE	HA	14	LEU	HD2#	2
11	ILE	HA	14	LEU	HB#	2
12	LYS	HA	15	ASN	HB1	2
12	LYS	HA	15	ASN	HB2	2
13	THR	HG2#	82	PHE	HA	2
13	THR	HG2#	82	PHE	HB1	2
13	THR	HG2#	82	PHE	HB2	2
13	THR	HG2#	85	ARG	HD#	2
13	THR	HG2#	85	ARG	HA	2
13	THR	HG2#	17	LEU	HD1#	2
13	THR	HG2#	17	LEU	HD2#	2
13	THR	HG2#	86	LEU	HD1#	2
13	THR	HG2#	86	LEU	HD2#	2
13	THR	HB	86	LEU	HD1#	4
13	THR	HB	86	LEU	HD2#	4
13	THR	HA	86	LEU	HD*	4
13	THR	HG2#	86	LEU	HA	4
14	LEU	HD1#	117	LYS	HA	2
14	LEU	HD*	120	MET	HG1	2
14	LEU	HD*	120	MET	HG2	2
14	LEU	HD*	116	LEU	HD*	2
14	LEU	HD1#	86	LEU	HD*	2
14	LEU	HD*	93	LEU	HD1#	2
14	LEU	HD*	93	LEU	HD2#	2
14	LEU	HD2#	117	LYS	HA	2
14	LEU	HD2#	86	LEU	HD*	2
14	LEU	HG	86	LEU	HD*	4
14	LEU	HD1#	113	LEU	HD*	2
14	LEU	HD2#	113	LEU	HD*	2
14	LEU	HA	17	LEU	HB#	4
16	SER	HA	19	GLU	HB#	4
16	SER	HA	19	GLU	HG#	4
17	LEU	HD1#	82	PHE	HA	2
17	LEU	HD1#	82	PHE	HB#	2
17	LEU	HD2#	82	PHE	HA	2
17	LEU	HD2#	82	PHE	HB#	2
18	THR	HG2#	110	GLU	HA	2
18	THR	HG2#	113	LEU	HD1#	2
18	THR	HG2#	113	LEU	HD2#	2
18	THR	HG2#	113	LEU	HG	2
21	LYS	HA	25	THR	HG2#	2
22	THR	HG2#	79	LEU	HD1#	2
22	THR	HG2#	79	LEU	HD2#	2
22	THR	HG2#	75	ARG	HD#	2
25	THR	HG2#	109	LEU	HD1#	2
25	THR	HG2#	109	LEU	HD2#	2
25	THR	HA	109	LEU	HD1#	4
25	THR	HA	109	LEU	HD2#	4
25	THR	HB	109	LEU	HD*	4
25	THR	HG2#	79	LEU	HD1#	2
25	THR	HG2#	79	LEU	HD2#	2
25	THR	HG2#	113	LEU	HD1#	2
25	THR	HG2#	113	LEU	HD2#	2
25	THR	HG2#	109	LEU	HA	2
25	THR	HG2#	109	LEU	HB#	2
25	THR	HA	109	LEU	HA	4

25	THR	HA	109	LEU	HB1	2	32	ILE	HD1#	33	PHE	HB#	2
25	THR	HA	109	LEU	HB2	2	32	ILE	HD1#	48	ALA	HA	2
25	THR	HB	109	LEU	HA	4	32	ILE	HD1#	48	ALA	HB#	2
25	THR	HB	109	LEU	HB#	4	32	ILE	HD1#	44	THR	HA	2
25	THR	HB	110	GLU	HB#	4	32	ILE	HD1#	44	THR	HG2#	2
25	THR	HB	110	GLU	HG#	4	32	ILE	HG1#	44	THR	HG2#	2
26	GLU	HA	108	THR	HG2#	2	32	ILE	HD1#	51	VAL	HG*	4
26	GLU	HB#	108	THR	HG2#	2	32	ILE	HG11	51	VAL	HG*	4
26	GLU	HG#	108	THR	HG2#	2	32	ILE	HG12	51	VAL	HG*	4
26	GLU	HA	108	THR	HA	2	32	ILE	HG2#	51	VAL	HB	2
27	LEU	HA	28	THR	HG2#	4	32	ILE	HG2#	48	ALA	HA	2
28	THR	HG2#	107	SER	HA	4	32	ILE	HA	35	ALA	HB#	4
28	THR	HG2#	108	THR	HB	4	32	ILE	HG2#	104	ALA	HB#	4
28	THR	HG2#	108	THR	HA	2	32	ILE	HD1#	104	ALA	HB#	4
28	THR	HG2#	106	GLN	HG#	2	32	ILE	HG1#	48	ALA	HA	4
28	THR	HG2#	106	GLN	HB#	2	33	PHE	HA	44	THR	HG2#	4
28	THR	HG2#	108	THR	HG2#	2	33	PHE	HB#	44	THR	HG2#	4
28	THR	HA	108	THR	HG2#	2	34	ALA	HB#	35	ALA	HA	4
28	THR	HB	108	THR	HG2#	2	34	ALA	HA	35	ALA	HB#	4
28	THR	HG2#	29	VAL	HG*	4	34	ALA	HB#	105	ASN	HB#	2
28	THR	HB	29	VAL	HG*	4	36	SER	HA	39	THR	HG2#	2
28	THR	HA	108	THR	HB	4	36	SER	HB#	39	THR	HG2#	4
28	THR	HA	108	THR	HA	2	39	THR	HA	40	THR	HG2#	2
28	THR	HB	106	GLN	HB#	2	39	THR	HG2#	40	THR	HA	4
28	THR	HB	106	GLN	HG#	4	39	THR	HG2#	43	GLU	HG#	2
28	THR	HB	108	THR	HA	4	41	GLU	HA	44	THR	HB	2
29	VAL	HG*	112	PHE	HA	4	41	GLU	HA	44	THR	HG2#	2
29	VAL	HG*	112	PHE	HB1	2	42	LYS	HA	94	ALA	HB#	2
29	VAL	HG*	112	PHE	HB2	2	42	LYS	HB#	94	ALA	HB#	2
29	VAL	HG*	107	SER	HB1	4	42	LYS	HA	94	ALA	HA	2
29	VAL	HG*	107	SER	HB2	4	42	LYS	HA	45	PHE	HB#	4
29	VAL	HG*	51	VAL	HB	4	42	LYS	HG#	45	PHE	HB#	4
29	VAL	HG*	51	VAL	HG1#	4	42	LYS	HB#	94	ALA	HA	2
29	VAL	HG*	51	VAL	HG2#	4	43	GLU	HA	46	CYS	HB1	4
29	VAL	HG*	52	LEU	HD*	4	43	GLU	HA	46	CYS	HB2	4
29	VAL	HG*	109	LEU	HD*	2	43	GLU	HA	96	LEU	HD1#	2
29	VAL	HA	51	VAL	HG*	4	43	GLU	HA	96	LEU	HD2#	2
29	VAL	HB	51	VAL	HG*	4	43	GLU	HG#	96	LEU	HD1#	2
30	THR	HG2#	106	GLN	HA	4	43	GLU	HG#	96	LEU	HD2#	2
30	THR	HG2#	104	ALA	HA	4	44	THR	HG2#	48	ALA	HB#	2
30	THR	HG2#	104	ALA	HB#	2	44	THR	HA	47	ARG	HD#	4
30	THR	HG2#	103	GLU	HA	2	44	THR	HA	47	ARG	HB#	4
30	THR	HG2#	103	GLU	HB#	2	45	PHE	HA	48	ALA	HB#	2
30	THR	HG2#	32	ILE	HG2#	2	45	PHE	HB#	48	ALA	HB#	4
30	THR	HB	32	ILE	HG2#	2	45	PHE	HA	90	LEU	HD*	2
30	THR	HG2#	101	VAL	HG*	4	45	PHE	HB#	90	LEU	HD*	2
30	THR	HB	101	VAL	HG*	2	45	PHE	HB1	94	ALA	HB#	4
30	THR	HA	106	GLN	HA	2	45	PHE	HB2	94	ALA	HB#	4
30	THR	HA	106	GLN	HB#	4	45	PHE	HB#	94	ALA	HA	4
30	THR	HA	106	GLN	HG#	2	46	CYS	HB1	96	LEU	HD1#	2
30	THR	HG2#	32	ILE	HA	2	46	CYS	HB2	96	LEU	HD1#	2
31	ASP	HB#	34	ALA	HB#	2	46	CYS	HB1	96	LEU	HD2#	2
31	ASP	HB#	107	SER	HA	2	46	CYS	HB2	96	LEU	HD2#	2
31	ASP	HB#	107	SER	HB1	4	47	ARG	HA	50	THR	HG2#	2
31	ASP	HB#	107	SER	HB2	4	47	ARG	HA	101	VAL	HG*	4
32	ILE	HD1#	33	PHE	HA	2	47	ARG	HA	50	THR	HB	2

48	ALA	HB#	51	VAL	HB	2	63	THR	HA	66	LEU	HG	2
48	ALA	HB#	51	VAL	HG*	4	66	LEU	HD*	73	PHE	HA	2
48	ALA	HA	51	VAL	HG*	2	66	LEU	HD*	73	PHE	HB1	4
48	ALA	HB#	90	LEU	HD1#	2	66	LEU	HD*	73	PHE	HB2	4
48	ALA	HB#	90	LEU	HD2#	2	67	GLY	HA1	68	ALA	HB#	4
48	ALA	HB#	96	LEU	HD*	2	67	GLY	HA2	68	ALA	HB#	4
48	ALA	HA	51	VAL	HB	2	68	ALA	HB#	69	THR	HG2#	2
50	THR	HG2#	101	VAL	HA	4	70	ALA	HB#	73	PHE	HB1	4
50	THR	HG2#	54	GLN	HG#	2	70	ALA	HB#	73	PHE	HB2	4
50	THR	HG2#	54	GLN	HB#	4	70	ALA	HA	73	PHE	HB1	4
50	THR	HG2#	101	VAL	HB	2	70	ALA	HA	73	PHE	HB2	4
50	THR	HG2#	101	VAL	HG1#	4	71	GLN	HA	74	HIS	HB#	2
50	THR	HG2#	101	VAL	HG2#	4	74	HIS	HA	77	LYS	HB#	4
50	THR	HB	101	VAL	HG1#	4	79	LEU	HD*	82	PHE	HB1	4
50	THR	HB	101	VAL	HG2#	4	79	LEU	HD*	82	PHE	HB2	4
50	THR	HG2#	100	PRO	HA	2	80	ILE	HD1#	83	LEU	HD*	2
51	VAL	HA	54	GLN	HB#	4	80	ILE	HA	83	LEU	HD*	2
51	VAL	HA	54	GLN	HG#	4	80	ILE	HA	83	LEU	HG	4
52	LEU	HD*	56	TYR	HA	2	82	PHE	HA	85	ARG	HD#	4
52	LEU	HD*	83	LEU	HD1#	2	82	PHE	HA	85	ARG	HB#	2
52	LEU	HD*	83	LEU	HD2#	2	82	PHE	HA	85	ARG	HA	4
52	LEU	HD*	109	LEU	HD*	2	83	LEU	HD1#	109	LEU	HD1#	2
52	LEU	HA	55	PHE	HB1	4	83	LEU	HD1#	109	LEU	HD2#	2
52	LEU	HA	55	PHE	HB2	4	83	LEU	HD2#	109	LEU	HD1#	2
54	GLN	HA	57	SER	HB#	2	83	LEU	HD2#	109	LEU	HD2#	2
55	PHE	HA	58	HIS	HB#	4	83	LEU	HG	109	LEU	HD*	4
56	TYR	HA	80	ILE	HD1#	2	83	LEU	HA	86	LEU	HB#	2
56	TYR	HB1	80	ILE	HD1#	2	83	LEU	HA	86	LEU	HD*	2
56	TYR	HB2	80	ILE	HD1#	2	84	LYS	HA	87	ASP	HB#	2
56	TYR	HA	80	ILE	HG2#	2	85	ARG	HA	88	ARG	HD#	4
56	TYR	HB1	80	ILE	HG2#	2	86	LEU	HD*	89	ASN	HB#	4
56	TYR	HB2	80	ILE	HG2#	2	86	LEU	HD*	90	LEU	HD*	4
56	TYR	HA	83	LEU	HD*	2	86	LEU	HA	89	ASN	HB#	4
56	TYR	HB1	83	LEU	HD*	2	86	LEU	HD*	117	LYS	HA	4
56	TYR	HB2	83	LEU	HD*	2	87	ASP	HA	90	LEU	HD1#	2
56	TYR	HA	83	LEU	HG	4	87	ASP	HA	90	LEU	HD2#	2
56	TYR	HA	80	ILE	HG1#	4	87	ASP	HA	90	LEU	HB#	4
56	TYR	HB#	80	ILE	HG1#	4	90	LEU	HD*	93	LEU	HD1#	2
57	SER	HA	60	GLU	HG#	4	90	LEU	HD*	93	LEU	HD2#	2
59	HIS	HA	62	ASP	HB#	2	90	LEU	HD*	94	ALA	HB#	4
60	GLU	HA	80	ILE	HD1#	2	90	LEU	HA	93	LEU	HD1#	4
60	GLU	HG#	80	ILE	HD1#	2	90	LEU	HA	93	LEU	HD2#	4
60	GLU	HA	80	ILE	HG2#	2	90	LEU	HD*	120	MET	HE#	4
60	GLU	HG#	80	ILE	HG2#	2	91	TRP	HA	94	ALA	HB#	2
60	GLU	HA	80	ILE	HG1#	4	91	TRP	HB#	94	ALA	HB#	2
60	GLU	HA	63	THR	HG2#	4	93	LEU	HD1#	120	MET	HE#	2
61	LYS	HA	66	LEU	HD1#	2	93	LEU	HD2#	120	MET	HE#	2
61	LYS	HA	66	LEU	HD2#	2	94	ALA	HB#	93	LEU	HD*	4
62	ASP	HA	63	THR	HG2#	4	94	ALA	HB#	96	LEU	HD*	2
63	THR	HG2#	66	LEU	HD1#	4	99	CYS	HA	100	PRO	HD1	4
63	THR	HG2#	66	LEU	HD2#	4	99	CYS	HA	100	PRO	HD2	4
63	THR	HA	66	LEU	HD1#	2	100	PRO	HA	101	VAL	HG1#	4
63	THR	HA	66	LEU	HD2#	2	100	PRO	HA	101	VAL	HG2#	4
63	THR	HB	66	LEU	HD*	2	101	VAL	HG*	102	LYS	HA	4
63	THR	HA	66	LEU	HB#	2	103	GLU	HA	104	ALA	HB#	4
63	THR	HB	66	LEU	HB#	4	105	ASN	HA	106	GLN	HG#	4

```
106 GLN HG#  107 SER HA   4
106 GLN HB#  107 SER HA   4
107 SER HA   108 THR HG2# 4
109 LEU HD1# 113 LEU HG   4
109 LEU HD1# 113 LEU HD1# 4
109 LEU HD1# 113 LEU HD2# 4
109 LEU HD2# 113 LEU HG   4
109 LEU HD2# 113 LEU HD1# 4
109 LEU HD2# 113 LEU HD2# 4
109 LEU HD*  112 PHE HB1  4
109 LEU HD*  112 PHE HB2  4
109 LEU HA   112 PHE HB1  4
109 LEU HA   112 PHE HB2  4
110 GLU HA   113 LEU HD*  2
111 ASN HA   114 GLU HB#  2
111 ASN HA   114 GLU HG#  2
112 PHE HA   115 ARG HB#  4
112 PHE HA   115 ARG HG#  4
113 LEU HA   116 LEU HB#  4
113 LEU HA   116 LEU HD*  2
114 GLU HA   117 LYS HB#  2
115 ARG HA   118 THR HB   2
116 LEU HD*  119 ILE HD1# 2
116 LEU HA   119 ILE HD1# 4
116 LEU HD*  119 ILE HG2# 2
116 LEU HA   119 ILE HG2# 4
116 LEU HD*  120 MET HG1  4
116 LEU HD*  120 MET HG2  4
116 LEU HA   119 ILE HB   2
116 LEU HD*  119 ILE HB   4
116 LEU HD*  120 MET HE#  2
118 THR HA   121 ARG HB#  2
118 THR HA   121 ARG HG#  4
118 THR HA   121 ARG HD#  4
119 ILE HG2# 120 MET HA   2
119 ILE HG2# 120 MET HB#  2
119 ILE HA   122 GLU HB#  2
119 ILE HA   122 GLU HG#  4
119 ILE HG2# 120 GLU HE#  2
121 ARG HA   124 TYR HB1  2
121 ARG HA   124 TYR HB2  2
123 LYS HA   126 LYS HB#  2
 45 PHE HD#   48 ALA HB#  4
 45 PHE HD#   44 THR HG2# 4
 45 PHE HD#   94 ALA HB#  4
 45 PHE HD#   90 LEU HD*  4
124 TYR HD#    7 LEU HD1# 4
124 TYR HD#    7 LEU HD2# 4
 56 TYR HD#   80 ILE HG2# 4
 56 TYR HD#   80 ILE HG1# 4
 56 TYR HD#   60 GLU HG#  4
 56 TYR HD#   57 SER HA   4
 33 PHE HD#   44 THR HG2# 4
 33 PHE HE#   44 THR HG2# 4
 82 PHE HD#   13 THR HG2# 4
 82 PHE HE#   13 THR HG2# 4
 33 PHE HD#  116 LEU HD*  4
 33 PHE HE#  116 LEU HD*  4
 73 PHE HD#   66 LEU HD*  4
 73 PHE HE#   66 LEU HD*  4
```

```
Phi angle restraints from coupling constant data for IL-4
helical residues
! Helix A
 assign (resid 4 and name c) (resid 5 and name n)
        (resid 5 and name ca) (resid 5 and name c) $8 -60.0 30.0 2
 assign (resid 5 and name c) (resid 6 and name n)
        (resid 6 and name ca) (resid 6 and name c) $8 -60.0 30.0 2
 assign (resid 6 and name c) (resid 7 and name n)
        (resid 7 and name ca) (resid 7 and name c) $8 -60.0 30.0 2
 assign (resid 7 and name c) (resid 8 and name n)
        (resid 8 and name ca) (resid 8 and name c) $8 -60.0 30.0 2
 assign (resid 8 and name c) (resid 9 and name n)
        (resid 9 and name ca) (resid 9 and name c) $8 -60.0 30.0 2
 assign (resid 9 and name c) (resid 10 and name n)
        (resid 10 and name ca) (resid 10 and name c) $8 -60.0 30.0 2
 assign (resid 10 and name c) (resid 11 and name n)
        (resid 11 and name ca) (resid 11 and name c) $8 -60.0 30.0 2
 assign (resid 11 and name c) (resid 12 and name n)
        (resid 12 and name ca) (resid 12 and name c) $8 -60.0 30.0 2
 assign (resid 12 and name c) (resid 13 and name n)
        (resid 13 and name ca) (resid 13 and name c) $8 -60.0 30.0 2
 assign (resid 13 and name c) (resid 14 and name n)
        (resid 14 and name ca) (resid 14 and name c) $8 -60.0 30.0 2
 assign (resid 14 and name c) (resid 15 and name n)
        (resid 15 and name ca) (resid 15 and name c) $8 -60.0 30.0 2
 assign (resid 15 and name c) (resid 16 and name n)
        (resid 16 and name ca) (resid 16 and name c) $8 -60.0 30.0 2
 assign (resid 16 and name c) (resid 17 and name n)
        (resid 17 and name ca) (resid 17 and name c) $8 -60.0 30.0 2
! Helix B
 assign (resid 40 and name c) (resid 41 and name n)
        (resid 41 and name ca) (resid 41 and name c) $8 -60.0 30.0 2
 assign (resid 41 and name c) (resid 42 and name n)
        (resid 42 and name ca) (resid 42 and name c) $8 -60.0 30.0 2
 assign (resid 42 and name c) (resid 43 and name n)
        (resid 43 and name ca) (resid 43 and name c) $8 -60.0 30.0 2
 assign (resid 43 and name c) (resid 44 and name n)
        (resid 44 and name ca) (resid 44 and name c) $8 -60.0 30.0 2
 assign (resid 44 and name c) (resid 45 and name n)
        (resid 45 and name ca) (resid 45 and name c) $8 -60.0 30.0 2
 assign (resid 45 and name c) (resid 46 and name n)
        (resid 46 and name ca) (resid 46 and name c) $8 -60.0 30.0 2
 assign (resid 46 and name c) (resid 47 and name n)
        (resid 47 and name ca) (resid 47 and name c) $8 -60.0 30.0 2
 assign (resid 47 and name c) (resid 48 and name n)
        (resid 48 and name ca) (resid 48 and name c) $8 -60.0 30.0 2
 assign (resid 48 and name c) (resid 49 and name n)
        (resid 49 and name ca) (resid 49 and name c) $8 -60.0 30.0 2
 assign (resid 49 and name c) (resid 50 and name n)
        (resid 50 and name ca) (resid 50 and name c) $8 -60.0 30.0 2
 assign (resid 50 and name c) (resid 51 and name n)
        (resid 51 and name ca) (resid 51 and name c) $8 -60.0 30.0 2
 assign (resid 51 and name c) (resid 52 and name n)
        (resid 52 and name ca) (resid 52 and name c) $8 -60.0 30.0 2
 assign (resid 53 and name c) (resid 54 and name n)
        (resid 54 and name ca) (resid 54 and name c) $8 -60.0 30.0 2
```

```
 assign (resid 54 and name c) (resid 55 and name n)
        (resid 55 and name ca) (resid 55 and name c) $8 -60.0 30.0 2
 assign (resid 55 and name c) (resid 56 and name n)
        (resid 56 and name ca) (resid 56 and name c) $8 -60.0 30.0 2
 assign (resid 56 and name c) (resid 57 and name n)
        (resid 57 and name ca) (resid 57 and name c) $8 -60.0 30.0 2
! Helix C
 assign (resid 71 and name c) (resid 72 and name n)
        (resid 72 and name ca) (resid 72 and name c) $8 -60.0 30.0 2
 assign (resid 72 and name c) (resid 73 and name n)
        (resid 73 and name ca) (resid 73 and name c) $8 -60.0 30.0 2
 assign (resid 73 and name c) (resid 74 and name n)
        (resid 74 and name ca) (resid 74 and name c) $8 -60.0 30.0 2
 assign (resid 74 and name c) (resid 75 and name n)
        (resid 75 and name ca) (resid 75 and name c) $8 -60.0 30.0 2
 assign (resid 75 and name c) (resid 76 and name n)
        (resid 76 and name ca) (resid 76 and name c) $8 -60.0 30.0 2
 assign (resid 76 and name c) (resid 77 and name n)
        (resid 77 and name ca) (resid 77 and name c) $8 -60.0 30.0 2
 assign (resid 77 and name c) (resid 78 and name n)
        (resid 78 and name ca) (resid 78 and name c) $8 -60.0 30.0 2
 assign (resid 78 and name c) (resid 79 and name n)
        (resid 79 and name ca) (resid 79 and name c) $8 -60.0 30.0 2
 assign (resid 79 and name c) (resid 80 and name n)
        (resid 80 and name ca) (resid 80 and name c) $8 -60.0 30.0 2
 assign (resid 80 and name c) (resid 81 and name n)
        (resid 81 and name ca) (resid 81 and name c) $8 -60.0 30.0 2
 assign (resid 81 and name c) (resid 82 and name n)
        (resid 82 and name ca) (resid 82 and name c) $8 -60.0 30.0 2
 assign (resid 82 and name c) (resid 83 and name n)
        (resid 83 and name ca) (resid 83 and name c) $8 -60.0 30.0 2
 assign (resid 83 and name c) (resid 84 and name n)
        (resid 84 and name ca) (resid 84 and name c) $8 -60.0 30.0 2
 assign (resid 84 and name c) (resid 85 and name n)
        (resid 85 and name ca) (resid 85 and name c) $8 -60.0 30.0 2
 assign (resid 85 and name c) (resid 86 and name n)
        (resid 86 and name ca) (resid 86 and name c) $8 -60.0 30.0 2
 assign (resid 86 and name c) (resid 87 and name n)
        (resid 87 and name ca) (resid 87 and name c) $8 -60.0 30.0 2
 assign (resid 87 and name c) (resid 88 and name n)
        (resid 88 and name ca) (resid 88 and name c) $8 -60.0 30.0 2
 assign (resid 88 and name c) (resid 89 and name n)
        (resid 89 and name ca) (resid 89 and name c) $8 -60.0 30.0 2
 assign (resid 89 and name c) (resid 90 and name n)
        (resid 90 and name ca) (resid 90 and name c) $8 -60.0 30.0 2
! Helix D
 assign (resid 109 and name c) (resid 110 and name n)
        (resid 110 and name ca) (resid 110 and name c) $8 -60.0 30.0 2
 assign (resid 110 and name c) (resid 111 and name n)
        (resid 111 and name ca) (resid 111 and name c) $8 -60.0 30.0 2
 assign (resid 111 and name c) (resid 112 and name n)
        (resid 112 and name ca) (resid 112 and name c) $8 -60.0 30.0 2
 assign (resid 112 and name c) (resid 113 and name n)
        (resid 113 and name ca) (resid 113 and name c) $8 -60.0 30.0 2
 assign (resid 113 and name c) (resid 114 and name n)
        (resid 114 and name ca) (resid 114 and name c) $8 -60.0 30.0 2
```

```
 assign (resid 114 and name c) (resid 115 and name n)
        (resid 115 and name ca) (resid 115 and name c) $8 -60.0 30.0 2
 assign (resid 115 and name c) (resid 116 and name n)
        (resid 116 and name ca) (resid 116 and name c) $8 -60.0 30.0 2
 assign (resid 116 and name c) (resid 117 and name n)
        (resid 117 and name ca) (resid 117 and name c) $8 -60.0 30.0 2
 assign (resid 117 and name c) (resid 118 and name n)
        (resid 118 and name ca) (resid 118 and name c) $8 -60.0 30.0 2
 assign (resid 118 and name c) (resid 119 and name n)
        (resid 119 and name ca) (resid 119 and name c) $8 -60.0 30.0 2
 assign (resid 119 and name c) (resid 120 and name n)
        (resid 120 and name ca) (resid 120 and name c) $8 -60.0 30.0 2
 assign (resid 120 and name c) (resid 121 and name n)
        (resid 121 and name ca) (resid 121 and name c) $8 -60.0 30.0 2
 assign (resid 121 and name c) (resid 122 and name n)
        (resid 122 and name ca) (resid 122 and name c) $8 -60.0 30.0 2
 assign (resid 122 and name c) (resid 123 and name n)
        (resid 123 and name ca) (resid 123 and name c) $8 -60.0 30.0 2
 assign (resid 123 and name c) (resid 124 and name n)
        (resid 124 and name ca) (resid 124 and name c) $8 -60.0 30.0 2
 assign (resid 124 and name c) (resid 125 and name n)
        (resid 125 and name ca) (resid 125 and name c) $8 -60.0 30.0 2
! Extended chain/ B sheet like
 assign (resid 19 and name c) (resid 20 and name n)
        (resid 20 and name ca) (resid 20 and name c)  $8 -120.0 40.0 2
 assign (resid 21 and name c) (resid 22 and name n)
        (resid 22 and name ca) (resid 22 and name c)  $8 -120.0 40.0 2
 assign (resid 28 and name c) (resid 29 and name n)
        (resid 29 and name ca) (resid 29 and name c)  $8 -120.0 40.0 2
 assign (resid 58 and name c) (resid 59 and name n)
        (resid 59 and name ca) (resid 59 and name c)  $8 -120.0 40.0 2
 assign (resid 64 and name c) (resid 65 and name n)
        (resid 65 and name ca) (resid 65 and name c)  $8 -120.0 40.0 2
 assign (resid 67 and name c) (resid 68 and name n)
        (resid 68 and name ca) (resid 68 and name c)  $8 -120.0 40.0 2
 assign (resid 68 and name c) (resid 69 and name n)
        (resid 69 and name ca) (resid 69 and name c)  $8 -120.0 40.0 2
 assign (resid 95 and name c) (resid 96 and name n)
        (resid 96 and name ca) (resid 96 and name c)  $8 -120.0 40.0 2
 assign (resid 101 and name c) (resid 102 and name n)
        (resid 102 and name ca) (resid 102 and name c)  $8 -120.0 40.0
! Values < 7 but not helix.
 assign (resid 3 and name c) (resid 4 and name n)
        (resid 4 and name ca) (resid 4 and name c) $8 60.0 150.0 2
 assign (resid 2 and name c) (resid 3 and name n)
        (resid 3 and name ca) (resid 3 and name c)  $8 60.0 150.0 2
 assign (resid 18 and name c) (resid 19 and name n)
        (resid 19 and name ca) (resid 19 and name c)  $8 60.0 150.0 2
 assign (resid 20 and name c) (resid 21 and name n)
        (resid 21 and name ca) (resid 21 and name c)  $8 60.0 150.0 2
 assign (resid 24 and name c) (resid 25 and name n)
        (resid 25 and name ca) (resid 25 and name c)  $8 60.0 150.0 2
 assign (resid 26 and name c) (resid 27 and name n)
        (resid 27 and name ca) (resid 27 and name c)  $8 60.0 150.0 2
 assign (resid 27 and name c) (resid 28 and name n)
        (resid 28 and name ca) (resid 28 and name c)  $8 60.0 150.0 2
```

```
 assign (resid 29 and name c) (resid 30 and name n)
        (resid 30 and name ca) (resid 30 and name c)  $8 60.0 150.0 2
 assign (resid 30 and name c) (resid 31 and name n)
        (resid 31 and name ca) (resid 31 and name c)  $8 60.0 150.0 2
 assign (resid 31 and name c) (resid 32 and name n)
        (resid 32 and name ca) (resid 32 and name c)  $8 60.0 150.0 2
 assign (resid 33 and name c) (resid 34 and name n)
        (resid 34 and name ca) (resid 34 and name c)  $8 60.0 150.0 2
 assign (resid 34 and name c) (resid 35 and name n)
        (resid 35 and name ca) (resid 35 and name c)  $8 60.0 150.0 2
 assign (resid 35 and name c) (resid 36 and name n)
        (resid 36 and name ca) (resid 36 and name c)  $8 60.0 150.0 2
 assign (resid 36 and name c) (resid 37 and name n)
        (resid 37 and name ca) (resid 37 and name c)  $8 60.0 150.0 2
 assign (resid 59 and name c) (resid 60 and name n)
        (resid 60 and name ca) (resid 60 and name c)  $8 60.0 150.0 2
 assign (resid 61 and name c) (resid 62 and name n)
        (resid 62 and name ca) (resid 62 and name c)  $8 60.0 150.0 2
 assign (resid 62 and name c) (resid 63 and name n)
        (resid 63 and name ca) (resid 63 and name c)  $8 60.0 150.0 2
 assign (resid 65 and name c) (resid 66 and name n)
        (resid 66 and name ca) (resid 66 and name c)  $8 60.0 150.0 2
 assign (resid 69 and name c) (resid 70 and name n)
        (resid 70 and name ca) (resid 70 and name c)  $8 60.0 150.0 2
 assign (resid 70 and name c) (resid 71 and name n)
        (resid 71 and name ca) (resid 71 and name c)  $8 60.0 150.0 2
  assign (resid 92 and name c) (resid 93 and name n)
        (resid 93 and name ca) (resid 93 and name c)  $8 60.0 150.0 2
  assign (resid 96 and name c) (resid 97 and name n)
        (resid 97 and name ca) (resid 97 and name c)  $8 60.0 150.0 2
  assign (resid 97 and name c) (resid 98 and name n)
        (resid 98 and name ca) (resid 98 and name c)  $8 60.0 150.0 2
  assign (resid 98 and name c) (resid 99 and name n)
        (resid 99 and name ca) (resid 99 and name c)  $8 60.0 150.0 2
  assign (resid 102 and name c) (resid 103 and name n)
        (resid 103 and name ca) (resid 103 and name c)  $8 60.0 150.0 2
  assign (resid 104 and name c) (resid 105 and name n)
        (resid 105 and name ca) (resid 105 and name c)  $8 60.0 150.0 2
  assign (resid 107 and name c) (resid 108 and name n)
        (resid 108 and name ca) (resid 108 and name c)  $8 60.0 150.0 2
  assign (resid 108 and name c) (resid 109 and name n)
        (resid 109 and name ca) (resid 109 and name c) $8 60.0  150.0 2
Hydrogen bond restraints for IL-4.
!* H-bonds in helix A.
ASSIGN (RESID  10 AND NAME NH) (RESID   6 AND NAME O) 1.8 0.0 0.2
ASSIGN (RESID  10 AND NAME N ) (RESID   6 AND NAME O) 2.8 0.1 0.2
ASSIGN (RESID  11 AND NAME NH) (RESID   7 AND NAME O) 1.8 0.0 0.2
ASSIGN (RESID  11 AND NAME N ) (RESID   7 AND NAME O) 2.8 0.1 0.2
ASSIGN (RESID  13 AND NAME NH) (RESID   9 AND NAME O) 1.8 0.0 0.2
ASSIGN (RESID  13 AND NAME N ) (RESID   9 AND NAME O) 2.8 0.1 0.2
ASSIGN (RESID  14 AND NAME NH) (RESID  10 AND NAME O) 1.8 0.0 0.2
ASSIGN (RESID  14 AND NAME N ) (RESID  10 AND NAME O) 2.8 0.1 0.2
ASSIGN (RESID  15 AND NAME NH) (RESID  11 AND NAME O) 1.8 0.0 0.2
ASSIGN (RESID  15 AND NAME N ) (RESID  11 AND NAME O) 2.8 0.1 0.2
!* H-bonds in helix B.
ASSIGN (RESID  45 AND NAME NH) (RESID  41 AND NAME O) 1.8 0.0 0.2
```

```
ASSIGN (RESID  45 AND NAME N ) (RESID  41 AND NAME O) 2.8 0.1 0.2
ASSIGN (RESID  46 AND NAME NH) (RESID  42 AND NAME O) 1.8 0.0 0.2
ASSIGN (RESID  46 AND NAME N ) (RESID  42 AND NAME O) 2.8 0.1 0.2
ASSIGN (RESID  47 AND NAME NH) (RESID  43 AND NAME O) 1.8 0.0 0.2
ASSIGN (RESID  47 AND NAME N ) (RESID  43 AND NAME O) 2.8 0.1 0.2
ASSIGN (RESID  48 AND NAME NH) (RESID  44 AND NAME O) 1.8 0.0 0.2
ASSIGN (RESID  48 AND NAME N ) (RESID  44 AND NAME O) 2.8 0.1 0.2
ASSIGN (RESID  49 AND NAME NH) (RESID  45 AND NAME O) 1.8 0.0 0.2
ASSIGN (RESID  49 AND NAME N ) (RESID  45 AND NAME O) 2.8 0.1 0.2
ASSIGN (RESID  51 AND NAME NH) (RESID  47 AND NAME O) 1.8 0.0 0.2
ASSIGN (RESID  51 AND NAME N ) (RESID  47 AND NAME O) 2.8 0.1 0.2
ASSIGN (RESID  52 AND NAME NH) (RESID  48 AND NAME O) 1.8 0.0 0.2
ASSIGN (RESID  52 AND NAME N ) (RESID  48 AND NAME O) 2.8 0.1 0.2
ASSIGN (RESID  55 AND NAME NH) (RESID  51 AND NAME O) 1.8 0.0 0.2
ASSIGN (RESID  55 AND NAME N ) (RESID  51 AND NAME O) 2.8 0.1 0.2
ASSIGN (RESID  56 AND NAME NH) (RESID  52 AND NAME O) 1.8 0.0 0.2
ASSIGN (RESID  56 AND NAME N ) (RESID  52 AND NAME O) 2.8 0.1 0.2
ASSIGN (RESID  57 AND NAME NH) (RESID  53 AND NAME O) 1.8 0.0 0.2
ASSIGN (RESID  57 AND NAME N ) (RESID  53 AND NAME O) 2.8 0.1 0.2
!* H-bonds in helix C.
ASSIGN (RESID  83 AND NAME NH) (RESID  79 AND NAME O) 1.8 0.0 0.2
ASSIGN (RESID  83 AND NAME N ) (RESID  79 AND NAME O) 2.8 0.1 0.2
ASSIGN (RESID  84 AND NAME NH) (RESID  80 AND NAME O) 1.8 0.0 0.2
ASSIGN (RESID  84 AND NAME N ) (RESID  80 AND NAME O) 2.8 0.1 0.2
ASSIGN (RESID  85 AND NAME NH) (RESID  81 AND NAME O) 1.8 0.0 0.2
ASSIGN (RESID  85 AND NAME N ) (RESID  81 AND NAME O) 2.8 0.1 0.2
ASSIGN (RESID  86 AND NAME NH) (RESID  82 AND NAME O) 1.8 0.0 0.2
ASSIGN (RESID  86 AND NAME N ) (RESID  82 AND NAME O) 2.8 0.1 0.2
!* H-bonds in helix D.
ASSIGN (RESID 114 AND NAME NH) (RESID 110 AND NAME O) 1.8 0.0 0.2
ASSIGN (RESID 114 AND NAME N ) (RESID 110 AND NAME O) 2.8 0.1 0.2
ASSIGN (RESID 115 AND NAME NH) (RESID 111 AND NAME O) 1.8 0.0 0.2
ASSIGN (RESID 115 AND NAME N ) (RESID 111 AND NAME O) 2.8 0.1 0.2
ASSIGN (RESID 116 AND NAME NH) (RESID 112 AND NAME O) 1.8 0.0 0.2
ASSIGN (RESID 116 AND NAME N ) (RESID 112 AND NAME O) 2.8 0.1 0.2
ASSIGN (RESID 117 AND NAME NH) (RESID 113 AND NAME O) 1.8 0.0 0.2
ASSIGN (RESID 117 AND NAME N ) (RESID 113 AND NAME O) 2.8 0.1 0.2
ASSIGN (RESID 118 AND NAME NH) (RESID 114 AND NAME O) 1.8 0.0 0.2
ASSIGN (RESID 118 AND NAME N ) (RESID 114 AND NAME O) 2.8 0.1 0.2
ASSIGN (RESID 119 AND NAME NH) (RESID 115 AND NAME O) 1.8 0.0 0.2
ASSIGN (RESID 119 AND NAME N ) (RESID 115 AND NAME O) 2.8 0.1 0.2
ASSIGN (RESID 120 AND NAME NH) (RESID 116 AND NAME O) 1.8 0.0 0.2
ASSIGN (RESID 120 AND NAME N ) (RESID 116 AND NAME O) 2.8 0.1 0.2
ASSIGN (RESID 121 AND NAME NH) (RESID 117 AND NAME O) 1.8 0.0 0.2
ASSIGN (RESID 121 AND NAME N ) (RESID 117 AND NAME O) 2.8 0.1 0.2
```

Alternative phi angle restraints for helical residues (and psi angle restraints) used in the preliminary structure calculations.

```
! Helix A
 assign (resid 4 and name c) (resid 5 and name n)
        (resid 5 and name ca) (resid 5 and name c) $8 -60.0 5.0 2
 assign (resid 5 and name c) (resid 6 and name n)
        (resid 6 and name ca) (resid 6 and name c) $8 -60.0 5.0 2
 assign (resid 6 and name c) (resid 7 and name n)
```

```
        (resid 7 and name ca) (resid 7 and name c) $8 -60.0 5.0 2
 assign (resid 7 and name c) (resid 8 and name n)
        (resid 8 and name ca) (resid 8 and name c) $8 -60.0 5.0 2
 assign (resid 8 and name c) (resid 9 and name n)
        (resid 9 and name ca) (resid 9 and name c) $8 -60.0 5.0 2
 assign (resid 9 and name c) (resid 10 and name n)
        (resid 10 and name ca) (resid 10 and name c) $8 -60.0 5.0 2
 assign (resid 10 and name c) (resid 11 and name n)
        (resid 11 and name ca) (resid 11 and name c) $8 -60.0 5.0 2
 assign (resid 11 and name c) (resid 12 and name n)
        (resid 12 and name ca) (resid 12 and name c) $8 -60.0 5.0 2
 assign (resid 12 and name c) (resid 13 and name n)
        (resid 13 and name ca) (resid 13 and name c) $8 -60.0 5.0 2
 assign (resid 13 and name c) (resid 14 and name n)
        (resid 14 and name ca) (resid 14 and name c) $8 -60.0 5.0 2
 assign (resid 14 and name c) (resid 15 and name n)
        (resid 15 and name ca) (resid 15 and name c) $8 -60.0 5.0 2
 assign (resid 15 and name c) (resid 16 and name n)
        (resid 16 and name ca) (resid 16 and name c) $8 -60.0 5.0 2
 assign (resid 16 and name c) (resid 17 and name n)
        (resid 17 and name ca) (resid 17 and name c) $8 -60.0 5.0 2
! Helix B
 assign (resid 40 and name c) (resid 41 and name n)
        (resid 41 and name ca) (resid 41 and name c) $8 -60.0 5.0 2
 assign (resid 41 and name c) (resid 42 and name n)
        (resid 42 and name ca) (resid 42 and name c) $8 -60.0 5.0 2
 assign (resid 42 and name c) (resid 43 and name n)
        (resid 43 and name ca) (resid 43 and name c) $8 -60.0 5.0 2
assign (resid 43 and name c) (resid 44 and name n)
        (resid 44 and name ca) (resid 44 and name c) $8 -60.0 5.0 2
assign (resid 44 and name c) (resid 45 and name n)
        (resid 45 and name ca) (resid 45 and name c) $8 -60.0 5.0 2
assign (resid 45 and name c) (resid 46 and name n)
        (resid 46 and name ca) (resid 46 and name c) $8 -60.0 5.0 2
assign (resid 46 and name c) (resid 47 and name n)
        (resid 47 and name ca) (resid 47 and name c) $8 -60.0 5.0 2
assign (resid 47 and name c) (resid 48 and name n)
        (resid 48 and name ca) (resid 48 and name c) $8 -60.0 5.0 2
assign (resid 48 and name c) (resid 49 and name n)
        (resid 49 and name ca) (resid 49 and name c) $8 -60.0 5.0 2
assign (resid 49 and name c) (resid 50 and name n)
        (resid 50 and name ca) (resid 50 and name c) $8 -60.0 5.0 2
assign (resid 50 and name c) (resid 51 and name n)
        (resid 51 and name ca) (resid 51 and name c) $8 -60.0 5.0 2
assign (resid 51 and name c) (resid 52 and name n)
        (resid 52 and name ca) (resid 52 and name c) $8 -60.0 5.0 2
assign (resid 52 and name c) (resid 53 and name n)
        (resid 53 and name ca) (resid 53 and name c) $8 -60.0 5.0 2
assign (resid 53 and name c) (resid 54 and name n)
        (resid 54 and name ca) (resid 54 and name c) $8 -60.0 5.0 2
assign (resid 54 and name c) (resid 55 and name n)
        (resid 55 and name ca) (resid 55 and name c) $8 -60.0 5.0 2
assign (resid 55 and name c) (resid 56 and name n)
        (resid 56 and name ca) (resid 56 and name c) $8 -60.0 5.0 2
assign (resid 56 and name c) (resid 57 and name n)
        (resid 57 and name ca) (resid 57 and name c) $8 -60.0 5.0 2
```

```
! Helix C
 assign (resid 71 and name c) (resid 72 and name n)
        (resid 72 and name ca) (resid 72 and name c) $8 -60.0 5.0 2
 assign (resid 72 and name c) (resid 73 and name n)
        (resid 73 and name ca) (resid 73 and name c) $8 -60.0 5.0 2
 assign (resid 73 and name c) (resid 74 and name n)
        (resid 74 and name ca) (resid 74 and name c) $8 -60.0 5.0 2
 assign (resid 74 and name c) (resid 75 and name n)
        (resid 75 and name ca) (resid 75 and name c) $8 -60.0 5.0 2
 assign (resid 75 and name c) (resid 76 and name n)
        (resid 76 and name ca) (resid 76 and name c) $8 -60.0 5.0 2
 assign (resid 76 and name c) (resid 77 and name n)
        (resid 77 and name ca) (resid 77 and name c) $8 -60.0 5.0 2
 assign (resid 77 and name c) (resid 78 and name n)
        (resid 78 and name ca) (resid 78 and name c) $8 -60.0 5.0 2
 assign (resid 78 and name c) (resid 79 and name n)
        (resid 79 and name ca) (resid 79 and name c) $8 -60.0 5.0 2
 assign (resid 79 and name c) (resid 80 and name n)
        (resid 80 and name ca) (resid 80 and name c) $8 -60.0 5.0 2
 assign (resid 80 and name c) (resid 81 and name n)
        (resid 81 and name ca) (resid 81 and name c) $8 -60.0 5.0 2
 assign (resid 81 and name c) (resid 82 and name n)
        (resid 82 and name ca) (resid 82 and name c) $8 -60.0 5.0 2
 assign (resid 82 and name c) (resid 83 and name n)
        (resid 83 and name ca) (resid 83 and name c) $8 -60.0 5.0 2
 assign (resid 83 and name c) (resid 84 and name n)
        (resid 84 and name ca) (resid 84 and name c) $8 -60.0 5.0 2
 assign (resid 84 and name c) (resid 85 and name n)
        (resid 85 and name ca) (resid 85 and name c) $8 -60.0 5.0 2
 assign (resid 85 and name c) (resid 86 and name n)
        (resid 86 and name ca) (resid 86 and name c) $8 -60.0 5.0 2
 assign (resid 86 and name c) (resid 87 and name n)
        (resid 87 and name ca) (resid 87 and name c) $8 -60.0 5.0 2
 assign (resid 87 and name c) (resid 88 and name n)
        (resid 88 and name ca) (resid 88 and name c) $8 -60.0 5.0 2
 assign (resid 88 and name c) (resid 89 and name n)
        (resid 89 and name ca) (resid 89 and name c) $8 -60.0 5.0 2
 assign (resid 89 and name c) (resid 90 and name n)
        (resid 90 and name ca) (resid 90 and name c) $8 -60.0 5.0 2
! Helix D
 assign (resid 109 and name c) (resid 110 and name n)
        (resid 110 and name ca) (resid 110 and name c) $8 -60.0 5.0 2
 assign (resid 110 and name c) (resid 111 and name n)
        (resid 111 and name ca) (resid 111 and name c) $8 -60.0 5.0 2
 assign (resid 111 and name c) (resid 112 and name n)
        (resid 112 and name ca) (resid 112 and name c) $8 -60.0 5.0 2
 assign (resid 112 and name c) (resid 113 and name n)
        (resid 113 and name ca) (resid 113 and name c) $8 -60.0 5.0 2
 assign (resid 113 and name c) (resid 114 and name n)
        (resid 114 and name ca) (resid 114 and name c) $8 -60.0 5.0 2
 assign (resid 114 and name c) (resid 115 and name n)
        (resid 115 and name ca) (resid 115 and name c) $8 -60.0 5.0 2
 assign (resid 115 and name c) (resid 116 and name n)
        (resid 116 and name ca) (resid 116 and name c) $8 -60.0 5.0 2
 assign (resid 116 and name c) (resid 117 and name n)
        (resid 117 and name ca) (resid 117 and name c) $8 -60.0 5.0 2
```

```
 assign (resid 117 and name c) (resid 118 and name n)
        (resid 118 and name ca) (resid 118 and name c) $8 -60.0 5.0 2
 assign (resid 118 and name c) (resid 119 and name n)
        (resid 119 and name ca) (resid 119 and name c) $8 -60.0 5.0 2
 assign (resid 119 and name c) (resid 120 and name n)
        (resid 120 and name ca) (resid 120 and name c) $8 -60.0 5.0 2
 assign (resid 120 and name c) (resid 121 and name n)
        (resid 121 and name ca) (resid 121 and name c) $8 -60.0 5.0 2
 assign (resid 121 and name c) (resid 122 and name n)
        (resid 122 and name ca) (resid 122 and name c) $8 -60.0 5.0 2
 assign (resid 122 and name c) (resid 123 and name n)
        (resid 123 and name ca) (resid 123 and name c) $8 -60.0 5.0 2
 assign (resid 123 and name c) (resid 124 and name n)
        (resid 124 and name ca) (resid 124 and name c) $8 -60.0 5.0 2
 assign (resid 124 and name c) (resid 125 and name n)
        (resid 125 and name ca) (resid 125 and name c) $8 -60.0 5.0 2
! psi constraints for helices.
 assign (resid 5 and name n) (resid 5  and name ca)
        (resid 5 and name c) (resid 6  and name n) $8 -50.0 5.0 2
 assign (resid 6 and name n) (resid 6  and name ca)
        (resid 6 and name c) (resid 7  and name n) $8 -50.0 5.0 2
 assign (resid 7 and name n) (resid 7  and name ca)
        (resid 7 and name c) (resid 8  and name n) $8 -50.0 5.0 2
 assign (resid 8 and name n) (resid 8  and name ca)
        (resid 8 and name c) (resid 9  and name n) $8 -50.0 5.0 2
 assign (resid 9 and name n) (resid 9  and name ca)
        (resid 9 and name c) (resid 10 and name n) $8 -50.0 5.0 2
 assign (resid 10 and name n) (resid 10 and name ca)
        (resid 10 and name c) (resid 11 and name n) $8 -50.0 5.0 2
assign (resid 11 and name n) (resid 11 and name ca)
        (resid 11 and name c) (resid 12 and name n) $8 -50.0 5.0 2
assign (resid 12 and name n) (resid 12 and name ca)
        (resid 12 and name c) (resid 13 and name n) $8 -50.0 5.0 2
assign (resid 13 and name n) (resid 13 and name ca)
        (resid 13 and name c) (resid 14 and name n) $8 -50.0 5.0 2
assign (resid 14 and name n) (resid 14 and name ca)
        (resid 14 and name c) (resid 15 and name n) $8 -50.0 5.0 2
assign (resid 15 and name n) (resid 15 and name ca)
        (resid 15 and name c) (resid 16  and name n) $8 -50.0 5.0 2
assign (resid 16 and name n) (resid 16 and name ca)
        (resid 16 and name c) (resid 17 and name n) $8 -50.0 5.0 2
assign (resid 17 and name n) (resid 17 and name ca)
        (resid 17 and name c) (resid 18 and name n) $8 -50.0 5.0 2
assign (resid 41 and name n) (resid 41 and name ca)
        (resid 41 and name c) (resid 42 and name n) $8 -50.0 5.0 2
assign (resid 42 and name n) (resid 42 and name ca)
        (resid 42 and name c) (resid 43 and name n) $8 -50.0 5.0 2
assign (resid 43 and name n) (resid 43 and name ca)
        (resid 43 and name c) (resid 44 and name n) $8 -50.0 5.0 2
assign (resid 44 and name n) (resid 44 and name ca)
        (resid 44 and name c) (resid 45 and name n) $8 -50.0 5.0 2
assign (resid 45 and name n) (resid 45 and name ca)
        (resid 45 and name c) (resid 46 and name n) $8 -50.0 5.0 2
assign (resid 46 and name n) (resid 46 and name ca)
        (resid 46 and name c) (resid 47 and name n) $8 -50.0 5.0 2
assign (resid 47 and name n) (resid 47 and name ca)
```

```
        (resid 47 and name c) (resid 48 and name n) $8 -50.0 5.0 2
assign (resid 48 and name n) (resid 48 and name ca)
        (resid 48 and name c) (resid 49 and name n) $8 -50.0 5.0 2
assign (resid 49 and name n) (resid 49 and name ca)
        (resid 49 and name c) (resid 50 and name n) $8 -50.0 5.0 2
assign (resid 50 and name n) (resid 50 and name ca)
        (resid 50 and name c) (resid 51  and name n) $8 -50.0 5.0 2
assign (resid 51 and name n) (resid 51 and name ca)
        (resid 51 and name c) (resid 52 and name n) $8 -50.0 5.0 2
assign (resid 52 and name n) (resid 52 and name ca)
        (resid 52 and name c) (resid 53 and name n) $8 -50.0 5.0 2
assign (resid 53 and name n) (resid 53 and name ca)
        (resid 53 and name c) (resid 54 and name n) $8 -50.0 5.0 2
assign (resid 54 and name n) (resid 54 and name ca)
        (resid 54 and name c) (resid 55 and name n) $8 -50.0 5.0 2
assign (resid 55 and name n) (resid 55 and name ca)
        (resid 55 and name c) (resid 56 and name n) $8 -50.0 5.0 2
assign (resid 56 and name n) (resid 56 and name ca)
        (resid 56 and name c) (resid 57 and name n) $8 -50.0 5.0 2
assign (resid 57 and name n) (resid 57 and name ca)
        (resid 57 and name c) (resid 58 and name n) $8 -50.0 5.0 2
assign (resid 72 and name n) (resid 72 and name ca)
        (resid 72 and name c) (resid 73 and name n) $8 -50.0 5.0 2
assign (resid 73 and name n) (resid 73 and name ca)
        (resid 73 and name c) (resid 74 and name n) $8 -50.0 5.0 2
assign (resid 74 and name n) (resid 74 and name ca)
        (resid 74 and name c) (resid 75 and name n) $8 -50.0 5.0 2
assign (resid 75 and name n) (resid 75 and name ca)
        (resid 75 and name c) (resid 76 and name n) $8 -50.0 5.0 2
assign (resid 76 and name n) (resid 76 and name ca)
        (resid 76 and name c) (resid 77  and name n) $8 -50.0 5.0 2
assign (resid 77 and name n) (resid 77 and name ca)
        (resid 77 and name c) (resid 78 and name n) $8 -50.0 5.0 2
assign (resid 78 and name n) (resid 78 and name ca)
        (resid 78 and name c) (resid 79 and name n) $8 -50.0 5.0 2
assign (resid 79 and name n) (resid 79 and name ca)
        (resid 79 and name c) (resid 80  and name n) $8 -50.0 5.0 2
assign (resid 80 and name n) (resid 80  and name ca)
        (resid 80 and name c) (resid 81 and name n) $8 -50.0 5.0 2
assign (resid 81 and name n) (resid 81 and name ca)
        (resid 81 and name c) (resid 82 and name n) $8 -50.0 5.0 2
assign (resid 82 and name n) (resid 82 and name ca)
        (resid 82 and name c) (resid 83 and name n) $8 -50.0 5.0 2
assign (resid 83 and name n) (resid 83 and name ca)
        (resid 83 and name c) (resid 84 and name n) $8 -50.0 5.0 2
assign (resid 84 and name n) (resid 84 and name ca)
        (resid 84 and name c) (resid 85 and name n) $8 -50.0 5.0 2
assign (resid 85 and name n) (resid 85 and name ca)
        (resid 85 and name c) (resid 86 and name n) $8 -50.0 5.0 2
assign (resid 86 and name n) (resid 86 and name ca)
        (resid 86 and name c) (resid 87 and name n) $8 -50.0 5.0 2
assign (resid 87 and name n) (resid 87 and name ca)
        (resid 87 and name c) (resid 88 and name n) $8 -50.0 5.0 2
assign (resid 88 and name n) (resid 88 and name ca)
        (resid 88 and name c) (resid 89  and name n) $8 -50.0 5.0 2
assign (resid 89 and name n) (resid 89 and name ca)
```

```
        (resid 89 and name c) (resid 90 and name n) $8 -50.0 5.0 2
assign (resid 90 and name n) (resid 90 and name ca)
        (resid 90 and name c) (resid 91 and name n) $8 -50.0 5.0 2
assign (resid 110 and name n) (resid 110 and name ca)
        (resid 110 and name c) (resid 111 and name n) $8 -50.0 5.0 2
assign (resid 111 and name n) (resid 111 and name ca)
        (resid 111 and name c) (resid 112 and name n) $8 -50.0 5.0 2
assign (resid 112 and name n) (resid 112 and name ca)
        (resid 112 and name c) (resid 113 and name n) $8 -50.0 5.0 2
assign (resid 113 and name n) (resid 113 and name ca)
        (resid 113 and name c) (resid 114 and name n) $8 -50.0 5.0 2
assign (resid 114 and name n) (resid 114 and name ca)
        (resid 114 and name c) (resid 115 and name n) $8 -50.0 5.0 2
assign (resid 115 and name n) (resid 115 and name ca)
        (resid 115 and name c) (resid 116 and name n) $8 -50.0 5.0 2
assign (resid 116 and name n) (resid 116 and name ca)
        (resid 116 and name c) (resid 117 and name n) $8 -50.0 5.0 2
assign (resid 117 and name n) (resid 117 and name ca)
        (resid 117 and name c) (resid 118 and name n) $8 -50.0 5.0 2
assign (resid 118 and name n) (resid 118 and name ca)
        (resid 118 and name c) (resid 119 and name n) $8 -50.0 5.0 2
assign (resid 119 and name n) (resid 119 and name ca)
        (resid 119 and name c) (resid 120 and name n) $8 -50.0 5.0 2
assign (resid 120 and name n) (resid 120 and name ca)
        (resid 120 and name c) (resid 121 and name n) $8 -50.0 5.0 2
assign (resid 121 and name n) (resid 121 and name ca)
        (resid 121 and name c) (resid 122 and name n) $8 -50.0 5.0 2
assign (resid 122 and name n) (resid 122 and name ca)
        (resid 122 and name c) (resid 123 and name n) $8 -50.0 5.0 2
assign (resid 123 and name n) (resid 123 and name ca)
        (resid 123 and name c) (resid 124 and name n) $8 -50.0 5.0 2
assign (resid 124 and name n) (resid 124 and name ca)
        (resid 124 and name c) (resid 125 and name n) $8 -50.0 5.0 2
assign (resid 125 and name n) (resid 125 and name ca)
        (resid 125 and name c) (resid 126 and name n) $8 -50.0 5.0 2
```

APPENDIX VIII.

^{15}N and ^{1}H Chemical Shifts for Human Interleukin 4 at pH 5.3 and 35 C.[a]

	amide NH				
	^{15}N	^{1}H	Hα	Hβ	other
M0					
H1					
K2	122.6	8.45	4.49	1.81, 1.71	(1.76, 1.45)
C3	120.9	8.67	4.65	3.10, 2.91	
D4	123.4	8.51	4.56	2.76, 2.68	
I5	123.9	8.47	4.00	1.97	Hδ_1 1.51, 1.35 Hγ_2 0.99 Hδ 0.89
T6	117.0	8.31	3.93	4.09	Hγ 1.20
L7	121.2	7.70	3.90		(1.80, 1.70, 0.92, 0.85)
Q8	116.7	7.85	3.69	2.14	Hγ 2.37, 2.25 Hϵ 7.16, 6.83 Nϵ 111.1
E9	119.2	7.97	3.99		
I10	122.3	8.12	3.52	1.96	Hγ_1 1.89 Hγ_2 0.78 Hδ 0.61
I11	120.4	7.95	3.40	1.71	Hγ_1 1.25, 0.93 Hγ_2 0.64 Hδ 0.51
K12	119.5	7.97	3.92		(1.86, 1.68, 1.52, 1.39)
T13	117.0	8.13	3.59	4.22	Hγ 1.12
L14	121.5	8.56	3.95		
N15	120.0	8.45	4.27	2.91, 2.79	Hδ 7.48, 6.55 Nδ 110.3
S16	115.8	7.77	4.18	3.61, 3.55	
L17	120.5	8.04	4.02	1.90, 1.47	Hγ 1.64 Hδ 1.01, 0.82
T18	107.4	7.67	4.34	4.48	Hγ 1.17
E19	120.0	7.37	4.22	2.09	Hγ 2.39, 2.31
Q20	117.1	7.42	4.47	1.96	Hγ 2.30 Hϵ 7.40, 6.55 Nϵ 110.6
K21	122.9	8.16	4.42		(1.80 1.71 1.38)
T22	115.0	8.23	4.87	4.58	Hγ 1.27
L23	121.6	8.47	4.22	1.87, 1.68	Hγ 1.79 Hδ 0.94, 0.94
C24	114.4	8.41	4.50	3.62, 3.11	
T25	108.1	7.75	4.18	4.50	Hγ 1.31
L27	118.6	7.04	4.41	2.02, 1.57	Hγ 1.79 Hδ 0.81, 0.58
T28	108.1	7.78	4.99	4.18	Hγ 1.10

V29	113.9	8.99	4.62	1.86	Hγ 0.66, 0.26
T30	118.4	7.78	4.08	3.82	Hγ 1.07
D31	123.8	8.58	4.68	2.77, 2.31	
I32	116.3	7.01	4.04	1.92	$H\gamma_1$ 0.69, 0.45 $H\gamma_2$ 0.87 Hδ 0.27
F33	120.2	8.14	4.42	3.23, 3.08	Hδ 7.22 Hε 6.58 Hζ 6.51
A34	123.6	7.44	4.22	1.39	
A35	120.5	7.67	4.42	1.37	
S36	114.1	7.81	4.36	3.94, 3.87	
K37	122.4	8.14	4.26	1.79	
N38	117.9	8.41	4.71	2.90, 2.74	Hδ 7.52, 6.83 Nδ 112.4
T39	114.3	7.73	4.54	4.09	Hγ 1.14
T40	114.7	8.43	4.46		Hγ 1.31
E41	123.1	8.73	4.62	1.84	
K42	116.1	7.99	3.79		(1.68, 1.59)
E43	117.8	7.50	3.82	2.18	
T44	118.2	8.10	3.70	4.11	Hγ 1.06
F45	119.9	8.30	4.49	3.29, 3.17	Hδ 6.90 Hε 7.20 Hζ 7.04
C46	119.8	8.16	3.97	3.27, 3.04	
R47	124.0	9.02	3.86		(2.08, 1.93, 1.61)
					Hδ 3.20, 2.97 Hε 7.38
A48	121.5	8.79	4.19	1.61	
A49	120.3	8.20	3.97		
T50	114.4	8.18	3.78	4.55	Hγ 1.29
V51	120.7	8.20	3.81	2.20	Hγ 1.08, 0.80
L52	120.9	8.14	3.96		(2.15)
Q54	120.6	8.06	3.83	2.09	Hγ 2.51, 2.32 Hε 7.27, 6.72
					Nε 110.2
F55	119.2	8.03	4.32	3.41, 3.16	Hδ 7.12 Hε 6.94 Hζ 7.31
Y56	117.3	9.22	4.26	3.40, 2.91	Hδ 6.94 Hε 6.82
S57	118.9	7.97	4.40	3.81, 3.76	
H58	116.3	7.54	4.55	2.77, 2.71	
H59	111.4	7.90	5.01	3.45, 2.21	$H\delta_2$ 6.05 $H\epsilon_1$ 8.47
E60	124.5	8.08	3.95		(2.57, 2.27, 2.12)

K61	116.5	8.59	4.48		(2.10, 1.61, 1.34)
D62	119.2	6.60	4.58	2.99, 2.70	
T63	122.7	8.79	3.84	4.29	Hγ 1.34
R64	121.5	8.49	4.08		(1.85, 1.72) Hδ 3.26, 3.20 Hε 7.51
C65	113.6	7.43	4.39	3.10, 2.60	
L66	119.4	7.38	3.74		(1.61, 1.46, 1.40) Hδ 0.73, 0.73
G67	99.6	7.08	3.91	3.68	
A68	120.8	8.40	4.70	1.45	
T69	107.7	7.59	4.71	4.62	Hγ 1.25
A70	124.9	8.93	4.12	1.47	
Q71	117.0	8.50	4.21	2.18, 2.08	Hγ 2.49 Hε 7.52, 6.92 Nε 111.9
Q72	119.8	7.81	4.12		(2.53, 2.31) Hε 7.54, 7.04 Nε 112.6
F73	122.9	8.41	4.52	3.48, 3.27	Hδ 7.23 Hε 7.34
H74	117.5	8.46	4.26	3.35, 3.28	
R75	119.2	8.21	4.02		(2.10, 1.98, 1.79)
H76	120.8	8.33	4.59	3.32, 3.20	
K77	117.5	8.16	3.63		(1.96, 1.85, 1.69, 1.55, 1.40, 1.23)
Q78	118.5	7.90	3.74		(1.96, 1.78) Hε 7.08, 6.82 Nε 112.0
L79	121.7	7.89	4.02		(2.26, 1.70) Hδ 1.00, 0.91
I80	116.4	8.10	3.53	2.06	Hγ1 1.35, 1.02 Hγ2 0.76 Hδ 0.69
R81	119.4	7.84	3.79		(1.77 1.65)
F82	119.0	8.51	4.4	3.42, 2.97	Hδ 7.19 Hε 7.36
L83	120.4	8.81	4.10		(2.34, 2.25) Hδ 1.02, 0.96
K84	119.0	8.14	3.78		
R85	120.5	7.64	4.01	1.99	(1.75 1.58)
L86	120.3	8.51	4.09		(1.81, 1.68, 0.93, 0.83)
D87	117.5	8.36	4.27	3.29, 3.16	
R88	115.2	7.59	4.02		(1.99 1.76)

N89	116.9	8.21	4.57	2.73, 2.46	Hδ 7.17, 7.01 Nδ 108.8
L90	121.7	8.81	3.83		
W91	119.3	9.04	4.25	3.38, 3.35	$H\delta_1$ 7.27 $H\epsilon_3$ 7.57 $H\zeta_3$ 7.23 $H\chi_2$ 7.17 $H\zeta_2$ 7.43 $H\epsilon_1$ 9.83
G92	106.2	8.03	3.98, 3.81		
L93	121.2	7.68	4.15		(1.59) Hδ 0.53, 0.53
A94	119.8	7.98	4.19	1.57	
G95	102.1	7.59	3.78, 3.64		
L96	118.6	7.51	4.39	1.43	Hγ 1.57 Hδ 0.83, 0.83
N97	117.2	8.42	4.82	2.72, 2.72	Hδ 7.37, 6.81 Nδ 112.3
S98	112.0	7.48	4.61	3.82, 3.81	
C99	118.9	8.93	4.93	3.49, 3.08	
P100			4.31	3.66, 3.56	
V101	122.8	8.41	4.12	1.93	Hγ 0.95, 0.82
K102	128.1	8.51	4.38	1.64	(1.79 1.33)
E103	123.4	8.26	4.17	1.91	(2.22)
A104	125.5	8.15	4.43	1.36	
N105	117.9	8.30	4.64	2.92, 2.77	Hδ 7.78, 6.99 Nδ 113.8
Q106	119.3	8.70	5.06		(2.33, 2.02) Hε 7.27, 6.61 Nε 113.3
S107	116.0	9.10	4.99	3.99, 3.70	
T108	113.4	8.90	4.41	4.72	Hγ 1.39
L109	125.3	8.87	3.95	1.77, 0.93	Hγ 1.37 Hδ 0.30, 0.10
E110	117.3	8.54	3.94		(2.33, 2.08, 1.94)
N111	117.9	7.90	4.40	2.86, 2.81	Hδ 7.67, 6.92 Nδ 112.4
F112	123.0	8.38	4.37	3.46, 2.99	Hδ 7.20
L113	118.2	8.91	3.87		(2.07, 1.14, 0.78)
E114	118.9	7.97	4.40	2.39	
R115	121.1	7.94	3.83		(1.55 1.17) Hδ 3.01, 2.93 Hε 6.90
L116	119.7	8.09	3.76		
K117	119.8	8.51	3.62		(2.09)
T118	116.4	7.92	3.79	4.27	Hγ 1.18

I119	121.5	7.71	3.82	1.89	Hγ, 1.56, 1.11 Hγ$_2$ 1.01 Hδ 0.67
M120	118.2	8.44	4.40	2.02, 1.73	Hγ 2.71, 2.31
R121	120.2	8.64	3.96		(1.91, 1.77, 1.55)
E122	120.1	7.87	4.10		(2.43, 2.21, 2.13)
K123	119.3	7.88	4.10		(1.99 1.70 1.47)
Y124	120.1	8.55	4.50	3.10, 2.98	Hδ 7.04 Hε 6.73
S125	114.8	8.10	4.17	4.04, 4.03	
K126	119.0	7.53	4.30		(1.96, 1.83)
C127	115.6	7.75	4.71	3.13, 3.12	
S128	116.1	7.97	4.43	3.78, 3.63	
S129	123.3	7.77	4.26	3.85, 3.85	

[a] ^{1}H chemical shifts are expressed relative to 4,4-dimethyl-4-silapentane-1-sulfonate and ^{15}N shifts relative to liquid ammonia. Chemical shifts given in parenthesis have not yet been assigned specifically to Hβ, Hσ or Hδ.

^{15}N, ^{1}H and ^{13}C Chemical Shifts for Human Interleukin 4 at pH 4.5 and 20 C
For each residue ^{1}H and amide ^{15}N chemical shifts are given on the first line and ^{13}C chemical shifts are given below on the second line.

	amide NH				
	^{15}N	^{1}H	$H\alpha$	$H\beta$	other
M0					
H1					
K2	122.7	8.56	4.49 $C\alpha$ 54.4	1.81, 1.71	(1.41)
C3	121.3	8.79	4.64 $C\alpha$ 52.8	3.06, 2.84	
D4	123.6	8.67	4.54 $C\alpha$ 53.0	2.74, 2.66	
I5	124.3	8.59	4.00 $C\alpha$ 61.3	1.95	$H\gamma_1$ 1.48, 1.31 $H\gamma_2$ 0.97 $H\delta$ 0.87 $C\gamma_1$ 25.9 $C\gamma_2$ 15.9 $C\delta$ 11.7
T6	116.2	8.33	3.95 $C\alpha$ 63.3	4.06 $C\beta$ 66.0	$H\gamma$ 1.19 $C\gamma$ 20.3
L7	121.3	7.66	3.92 $C\alpha$ 56.4	1.67, 1.75	$H\gamma$ 1.73 $H\delta$ 0.88, 0.96 $C\delta$ 23.5
Q8	116.5	7.93	3.67 $C\alpha$ 57.7	2.12	$H\gamma$ 2.35, 2.21 $H\epsilon$ 7.16, 6.90 $N\epsilon$ 111.3
E9	119.0	7.95	3.97 $C\alpha$ 57.4	2.09, 1.99	$H\gamma$ 2.28
I10	122.1	8.14	3.59 $C\alpha$ 63.7	1.92	$H\gamma_1$ 1.84 $H\gamma_2$ 0.75 $H\delta$ 0.57 $C\gamma_2$ 15.6 $C\delta$ 12.9
I11	120.6	7.98	3.36 $C\alpha$ 62.7	1.66	$H\gamma_1$ 1.23, 0.87 $H\gamma_2$ 0.60 $H\delta$ 0.47 $C\gamma_1$ 27.5 $C\gamma_2$ 15.0 $C\delta$ 10.3
K12	119.4	7.89	3.92 $C\alpha$ 58.0	1.84	(1.60, 1.50, 1.36)
T13	117.0	8.19	3.60 $C\alpha$ 64.8	4.20 $C\beta$ 65.7	$H\gamma$ 1.10 $C\gamma$ 21.0
L14	121.3	8.62	3.92 $C\alpha$ 56.7	2.00, 1.10	$H\gamma$ 1.73 $H\delta$ 0.88, 0.61 $C\delta$ 22.6, 24.5
N15	120.1	8.48	4.25 $C\alpha$ 54.1	2.90, 2.73	$H\delta$ 7.55, 6.62 $N\delta$ 110.5
S16	115.6	7.88	4.17 $C\alpha$ 59.5	3.57, 3.57 $C\beta$ 60.5	
L17	120.3	8.02	4.01 $C\alpha$ 55.6	1.87, 1.43	$H\gamma$ 1.60 $H\delta$ 0.97, 0.77 $C\delta$ 23.4, 25.2
T18		7.75	4.42 $C\alpha$ 60.4	4.43 $C\beta$ 67.0	$H\gamma$ 1.12 $C\gamma$ 19.8
E19	120.1	7.40	4.21 $C\alpha$ 55.7	2.05	$H\gamma$ 2.42, 2.31

Q20	117.0	7.44	4.42		1.93, 2.04	Hγ 2.20	Hε 7.43, 6.56
			Cα 52.7			Nε 110.7	
K21	123.5	8.34	4.35				
			Cα 54.8				
T22	115.0	8.25	4.84		4.53	Hγ 1.21	
			Cα 58.0	Cβ	69.3	Cγ 19.9	
L23	121.6	8.55	4.17		1.85, 1.60	Hγ 1.77	Hδ 0.90, 0.91
			Cα 56.0	Cβ	40.3		Cδ 21.4
C24	115.0	8.46					
T25		7.74	4.17		4.48	Hγ 1.27	
			Cα 62.0	Cβ	65.7	Cγ 20.3	
E26			4.49		2.26, 2.00	Hγ 2.35	
			Cα 54.3				
L27	118.8	7.04	4.42		2.01, 1.54	Hγ 1.82	Hδ 0.80, 0.58
			Cα 53.2				Cδ 19.8, 23.5
T28		7.91	5.01		4.17	Hγ 1.06	
			Cα 58.9	Cβ	69.2	Cγ 19.4	
V29	113.5	9.00	4.59		1.79	Hγ 0.59, 0.23	
			Cα 57.2			Cγ 21.1, 17.0	
T30	118.6	7.89	4.05		3.78	Hγ 1.03	
			Cα 61.8	Cβ	66.8	Cγ 20.7	
D31	123.7	8.64	4.66		2.74, 2.28		
			Cα 51.0				
I32	116.3	7.14	4.03		1.90	Hγ1 0.67, 0.45	Hγ$_2$ 0.84 Hδ 0.14
			Cα 58.8				Cγ$_2$ 20.1 Cδ 13.0
F33	119.9	8.04	4.39		3.21, 3.02		
			Cα 57.8				
A34	123.7	7.40	4.20		1.36		
			Cα 51.8	Cβ	16.6		
A35	120.8	7.73	4.41		1.32		
			Cα 50.3	Cβ	17.2		
S36	114.0	7.87	4.33		3.91, 3.83		
			Cα 57.0	Cβ	61.4		
K37	122.6	8.24	4.22		1.76	(1.63, 1.38)	
			Cα 55.3				
N38	117.6	8.47	4.69		2.89, 2.70	Hδ 7.59, 6.88	Nδ 112.6
			Cα 51.6				
T39	114.5	7.78	4.52		4.03	Hγ 1.13	
			Cα 58.8	Cβ	68.0	Cγ 19.4	
T40	115.0	8.46	4.42		4.57	Hγ 1.29	
			Cα 58.4	Cβ	68.9	Cγ 19.6	
E41	123.1	8.79	3.61		1.82	Hγ 1.83, 1.70	
			Cα 57.9				
K42	116.1	7.99	3.79			(1.68, 1.59)	

		Cα 58.1			Cγ 24.0	
E43	117.8	7.48	3.86 Cα 58.0	1.91	Hγ 2.32	
T44	118.5	8.16	3.66 Cα 65.0	4.06 Cβ 65.1	Hγ 1.01 Cγ 19.4	
F45	119.9	8.37	4.41 Cα 56.8	3.17, 3.08		
C46	120.2	8.25	3.76 Cα 58.0	3.32, 3.09		
R47	124.8	9.05	3.81 Cα 58.0		(2.06, 1.78, 1.62, 1.92)	
A48	121.6	8.79	4.12 Cα 53.7	1.53 Cβ 16.6		
A49	120.1	8.26				
T50			3.95 Cα 65.1	4.54 Cβ 66.6	Hγ 1.29 Cγ 19.1	
V51	120.6	8.20	3.88 Cα 64.0	2.20	Hγ 1.06, 0.80 Cγ 20.3, 19.6	
L52	120.4	8.18	3.98 Cα 56.5	2.10, 1.44	Hγ 1.97	Hδ 0.78, 0.88 Cδ 21.9
Q54	120.5	7.96	3.85 Cα 57.2	2.07, 2.31	Hγ 2.50, 2.31 Nε 110.4	Hε 7.36, 6.79
F55	118.9	8.06	4.32 Cα 60.3	3.39, 3.13		
Y56	117.1	9.08	4.23 Cα 58.4	3.45, 2.89	Hδ 6.91	Hε 6.79
S57	118.6	7.96	4.40 Cα 59.2	3.80, 3.72 Cβ 60.2		
H58	115.9	7.57	4.57 Cα 55.1	2.77, 2.77		
H59	111.3	7.85	5.01 Cα 53.7	3.42, 2.19		
E60	124.5	8.09	3.94 Cα 58.9	2.08	Hγ 2.52, 2.28	
K61	116.6	8.69	4.47 Cα 52.7		(2.05, 1.57, 1.32)	
D62	119.3	6.64	4.58 Cα 51.5	2.98, 2.69		
T63	122.8	8.94	3.87 Cα 63.0	4.25 Cβ 66.2	Hγ 1.31 Cγ 19.8	
R64	121.3	8.50	4.07 Cα 56.7		(1.83, 1.71)	

L66 119.5 7.44 1.62, 1.46 Hγ 1.42 Hδ 0.70, 0.75
Cδ 23.2, 22.1

G67 3.95, 3.67
Cα 43.5

A68 120.9 8.43 4.70 1.42
Cα 50.4 Cβ 19.3

T69 7.62 4.60 4.64 Hγ 1.21
Cα 57.4 Cγ 19.4

A70 124.9 8.98 4.11 1.42
Cα 53.3 Cβ 16.2

Q71 117.4 8.58 4.19 2.13, 2.03 Hγ 2.43 Hε 7.59, 6.96
Cα 57.7 Nε 112.0

Q72 119.9 7.88 Hε 7.59, 7.07
Nε 112.7

F73 122.7 8.51 4.54 3.43, 3.22
Cα 58.1

H74 116.9 8.50 4.31 3.39, 3.39
Cα 57.9

R75 119.3 8.29 4.07 (2.05, 1.97, 1.72) Hδ 3.23, 3.23
Cα 57.4

H76 120.5 8.38 4.59 3.28, 3.21

K77 117.5 8.18 3.66 (1.82, 1.67, 1.51,1.36, 1.21)
Cα 57.7

Q78 118.6 7.89 3.78 (1.94) Hε 7.19, 6.88
Cα 56.8 Nε 112.2

L79 121.8 7.91 4.05 (2.18, 1.71) H 0.95, 0.89
Cα 57.3 C 22.1

I80 116.2 8.09 3.55 2.07 $H\gamma_1$ 1.36, 1.00 $H\gamma_2$ 0.76 Hδ 0.68
Cα 59.7 $C\gamma_2$ 15.8 Cδ 6.2

R81 119.3 7.83 3.81 (1.78, 1.66, 1.37) Hδ 3.07
Cα 59.7

F82 119.2 8.50 4.41 3.39, 2.97
Cα 57.8

L83 120.1 8.86 4.05 1.50 (2.28) Hδ 1.01, 0.93
Cα 56.8 Cδ 25.3, 21.5

K84 119.2 8.20 3.68 (1.59)
Cα 58.0

R85 121.0 7.66 3.99 1.95 (1.70, 1.60) Hδ 3.11, 3.11

L86 120.0 8.58 4.13 (1.88, 1.53, 0.84, 0.76)
Cα 56.4 Cδ 22.5, 25.3

D87 116.9 8.50 4.24 3.21, 3.21
Cα 56.0

R88 115.7 7.76 4.19 (1.99, 1.82, 1.70) Hδ 3.21

N89	117.0	8.09	4.56		2.76, 2.44	Hδ 7.19, 7.15	Nδ 108.7
			Cα 54.5				
L90	121.4	8.68	3.78		1.97	Hγ 1.71	Hδ 0.59, 0.72
			Cα 56.7				Cδ 21.2, 24.4
W91			4.37		3.33		
			Cα 58.1				
G92		7.97	3.99, 3.85				
			Cα 44.9				
L93	120.9	7.65	4.16		1.52, 1.40	Hγ 1.56	Hδ 0.49, 0.44
			Cα 54.9	Cβ 40.9		Cγ 24.7	Cδ 21.9, 23.2
A94	119.6	8.08	4.24		1.56		
			Cα 52.5	Cβ 18.4			
G95		7.72	3.90, 3.73				
			Cα 44.6				
L96	118.5	7.49	4.44		1.43, 1.51	Hγ 1.46	Hδ 0.84, 0.86
			Cα 51.9	Cβ 42.7			Cδ 21.8, 22.7
N97	117.2	8.45	4.87		2.64, 2.64	Hδ 7.42, 6.83	Nδ 112.3
			Cα 52.0				
S98	112.5	7.55	4.63		3.76, 3.76		
			Cα 55.6	Cβ 61.5			
C99	119.1	9.01	4.96		3.53, 3.03		
			Cα 52.8				
P100			4.29		2.20, 1.76	Hγ 1.99, 1.90	Hδ 3.63, 3.53
			Cα 60.9				
V101	122.9	8.50	4.10		1.93	Hγ 0.94, 0.78	
			Cα 59.2			Cγ 19.4, 19.5	
K102	128.6	8.58	4.34		1.75, 1.62	(1.74, 1.31)	
			Cα 53.8				
E103	123.5	8.35	4.16		1.87	(2.23)	
			Cα 55.5				
A104	125.9	8.17	4.43		1.34		
			Cα 50.3	Cβ 18.4			
N105	118.1	8.38	4.63		2.90, 2.74	Hδ 7.82, 7.03	Nδ 113.8
			Cα 52.1				
Q106	119.1	8.80	5.03		2.00, 1.98	Hγ 2.30	Hε 7.28, 6.63
			Cα 53.2			Nε 114.0	
S107	116.0	9.19	4.99		3.95, 3.66		
			Cα 54.4	Cβ 64.5			
T108	113.4	9.00	4.35		4.72	Hγ 1.36	
			Cα 59.8	Cβ 68.2		Cγ 20.2	
L109	125.2	8.91	3.90		1.70, 0.87	Hγ 1.28	Hδ 0.23, 0.10
			Cα 56.5			Cγ 24.2	Cδ 20.5, 23.6
E110	117.4	8.58	3.92		2.08, 1.88	Hγ 2.37	
			Cα 58.0				

N111	117.6	7.95	4.37		2.81, 2.74	Hδ 7.76, 6.96	Nδ112.7	
			Cα 54.8					
F112	123.1	8.32	4.38		3.46, 2.93			
			Cα 58.8					
L113	118.2	8.92	3.85		1.79, 1.68	Hγ 2.05	Hδ 0.73, 0.76	
			Cα 55.9				Cδ 24.4, 20.4	
E114	118.6	8.07	4.00		2.07	H 2.44		
			Cα 57.6					
R115	121.0	7.93	3.83		1.51	Hγ 1.31, 1.07	Hδ 2.97, 2.86	
			Cα 56.8			Cγ 24.6		
L116	119.5	8.04	3.73		1.45	Hγ 1.48	Hδ 0.86, 0.87	
			Cα 55.7				Cδ 21.5, 24.0	
K117	119.8	8.50	3.58			(2.03, 1.72, 1.32)		
			Cα 58.2					
T118	116.5	7.92	3.77		4.22	Hγ 1.14		
			Cα 64.8	Cβ 66.3		Cγ 19.6		
I119	121.5	7.69	3.77		1.86	$H\gamma_1$ 1.54, 1.09	$H\gamma_2$ 0.98	Hδ 0.63
			Cα 62.6			$C\gamma_1$ 27.0	$C\gamma_2$ 15.0	Cδ 12.7
M120	118.2	8.48	4.35		2.01, 1.71	Hγ 2.67, 2.27	Hε 1.42	
			Cα 54.8					
R121	120.2	8.69	3.95			(1.92, 1.80, 1.59)		
			Cα 58.0					
						Hδ 3.13, 3.13		
E122	119.9	7.89	4.09		2.18, 2.09	Hγ 2.49, 2.27		
			Cα 57.4					
K123	119.6	7.84	4.09					
Y124	120.3	8.62	4.50		3.08, 2.96	Hδ 7.00		
			Cα 57.8					
S125	114.3	8.14	4.16		4.01, 4.01			
			Cα 58.6	Cβ 60.7				
K126	118.8	7.48	4.28		1.95, 1.81	(1.63, 1.43)		
			Cα 55.1					
C127		7.76	4.69		3.15, 3.03			
			Cα 53.8					
S128	115.8	8.03	4.42		3.75, 3.56			
			Cα 56.7	Cβ 61.6				
S129	123.5	7.82	4.24		3.81, 3.81			
			Cα 58.0					

[a] ^1H and ^{13}C chemical shifts are expressed relative to 4,4-dimethyl-4-silapentane-1-sulfonate and ^{15}N shifts relative to liquid ammonia. Chemical shifts given in parenthesis have not yet been assigned specifically to Hβ, Hγ or Hδ.