Solution Structures of Endothelin Peptides and a Glycoside by NMR Spectroscopy

by

Chandralal M. Hewage ***B. Sc. M. Phil.***

A Thesis submitted for the degree of

Doctor of Philosophy

The Ultra High Field NMR Centre

Department of Chemistry

University of Edinburgh

United Kingdom

December 1995

To my father

who is no more in the world
but whose hopes and dreams for
my education have culminated today.

Acknowledgments

I would like to thank Prof. Robert Ramage, FRS, for admitting me as a research student and the provision of research facilities. I am thankful to Dr. Ian H. Sadler for his valuable guidance, keen interest and the personal attention given to me throughout my research career. Sincere thanks to Dr. David Reed for valuable lessons and discussions.

My very special thanks go to Dr. John A. Parkinson for his valuable advice, continuous encouragement and stimulating discussions throughout the course of this research work. Without his help endothelin project would not have been completed successfully.

I gratefully acknowledge Parke Davis Pharmaceuticals Res. Co. (Ann Arbor, Michigan) for supplying material and Dr. Lu Jiang for synthesis of peptides. I also gratefully acknowledge the EPSRC for the NMR facilities made available in the department and the Wellcome Trust for the Molecular Modeling facilities.

Grateful acknowledgment is made of the Commonwealth Association and the British Council for their financial support. I should like to thank again Prof. R. Ramage and the Cowan House Scholarship committee for financial support during the last few months of my study.

I wish to express my gratitude to all past and present colleagues and friends in the department especially Drs. David Thomas and Frank Whal. I extend my gratitude to all the Sri Lankans for helping to make my time in Edinburgh so enjoyable.

I shall remain indebted to all teachers, friends and especially my loving mother, brother and sisters, for their love, constant care and encouragement given to me throughout my studies away from Sri Lanka.

Finally, my loving thanks goes to my wife Sagarika for giving me invaluable assistance and many happy moments during the long days of studies. I wish her success.

Candidate's Note

This thesis is submitted in partial fulfillment of the requirements of the degree of Doctor of Philosophy. It contains an account of my own research work performed at the Department of Chemistry, University of Edinburgh under the supervision of Dr. Ian H. Sadler. Unless otherwise stated, the work described is original and has not been previously submitted, in whole or in part, for any degree at this or any other university.

Chandralal M. Hewage
University of Edinburgh
December 1995

The Abstract

The endothelins were discovered in 1988 and are known to be the most active pressor molecules in the mammalian vascular system. Endothelin-1, which shows potent and long-lasting vasoconstricting activity has been isolated from the culture medium of porcine aortic endothelial cells and implicated in a novel cardiovascular control system. The first member of endothelin family, Endothelin-1, is a 21 amino acid peptide whose structure is constrained by two disulphide bridges between residues 1-15 and 3-11.

Increasing evidence for the involvement of endothelins in human disease has prompted a major effort in drug design, pharmacological evaluation and structure elucidation. In this thesis, the three dimensional solution structure of Endothelin-1 and modified linear Endothelin-1 derivatives are presented using one- and two-dimensional NMR methods followed by structure calculations DIANA, DSA and MD. This is the first report of solution structures of modified linear Endothelin-1 derivatives.

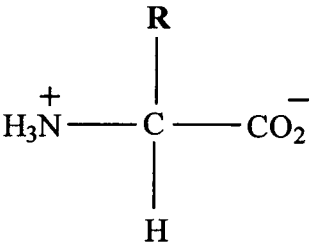
The Panax family plants (*P. ginseng* and *P. notoginseng*) are well known in traditional Chinese medicine with the popular name "ginseng". Major compounds isolated from the Panax family plants are saponins and the most of the saponins are also biologically active. The last chapter of this thesis presents the elucidation by one- and two-dimensional NMR methods of the structure and stereochemistry of a compound isolated from the roots of *Panax notoginseng* shown to release tissue plasminogen activator (tPA) from hemi-pituitary glands *in vitro*. The compound was identified as the saponin, ginsenoside-Rd, and its NMR spectra fully assigned for the first time.

Symbols and Abbreviations

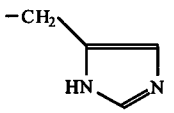
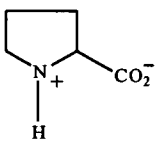
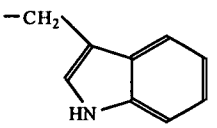
1D	One Dimension
2D	Two Dimension
3D	Three Dimension
Å	Angstrom units (= 10^{-10}m)
Ac	acetyl
AcOH	acetic acid
ANP	atrial natriuretic peptide
AQ	acqisition time
ara(fur)	α-L-arabinofuranosyl
ara(pyr)	α-L-arabinopyranosyl
BBTX	Bibrotoxin
CD	Circular Dichroism
cDNA	complementary DNA
CMC	Critical Micelle Concentration
CNS	Central Nervous System
COSY	two dimensional correlated spectroscopy
CW	continuous wave
d	doublet
Da	Daltons
dB	decibell
δ_C	^{13}C chemical shift values in ppm
dd	doublet of a doublet
ddd	doublet fo a doublet of a doublet
DEPT	Distortionless Enhancement by Polarization Transfer
DG	Distance Geometry
δ_H	^{1}H chemical shift values in ppm
DMSO	dimethylsulphoxide
Dn	n-th Delay
DNA	Deoxyribonucleic acid
dq	doublet of a quartet
DQF COSY	two dimensional Double Quantum Filtered correlated spectroscopy
DSA	Dynamical Simmulated Annealing
dt	doublet of a triplet
EC_{50}	concentration of drug giving 50% of the maximum responce
ECE	Endothelin converting enzyme
EDCF	Endothelin derived constrictor factor
EDRF	Endothelin derived releasing factor
ET	Endothelin
ether	diethyl ether
FID	Free Induction Decay
Fn	Frequency domain of the n-th dimension
FT	Fourier Transformation
glc	β-D-glucopyranosyl
glu	α-D-glucopyranosyl
HMBC	Heteronuclear Multiple-Bond Correlation
HMQC	Heteronuclear Multiple-Quantum Correlation
hsp	homo spoil pulse

Hz	Hertz
IC_{50}	concentration of drug giving 50% displacement of specific binding
J	coupling constants
K	Kelvin ; thousand
LR COSY	two dimensional long range correlated spectroscopy
m	multiplet, multiplicities
MCD	Mast Cell Degranulating
MD	Mocecular Dynamics
Me	methyl
MeOH	methanol
mg	milligramme
ml	millilitre
μm	micromolar
mM	millimoalr
mRNA	messenger RNA
μs	microsecond
ms	millisecond
mw	molecular weight
nM	nonomolar
NMR	Nuclear Magnetic Resonance
NOE (nOe)	nuclear Overhauser effect
NOESY	two dimensional nuclear Overhauser enhancement spectroscopy
ORD	optical rotatory dispersion
p	pentet
ppm	parts per million
q	quartet
rha(pyr)	α-L-rhamnopyranosyl
RMSD	root mean square deviation
RNA	Ribonucleic acid
ROESY	2 dimensional rotating-frame Overhauser enhancement spectroscopy
s	second, singlet
SAR	Structure Activity Relationship
SD	standard deviation
SRTX	Sarafotoxin
SW	spectral width
t	triplet
τ_c	rotational correlation time
TFA	trifluoroacetic acid
TFE	trifluoroethanol
t_m	mixing time
TMS	Tetramethylsilane
TOCSY	two demensional total correlated spectroscopy
tt	triplet of a triplet
VIC	Vasoactive Intestinal Contractor
VT	Variable Temperature
w	angular Larmor frequency
xyl	β-D-xylopyranosyl

Codes for amino acids



All amino acids used were of the L-configuration unless otherwise stated

Amino acid	**Side chain (R)**	**3-Letter code**	**1-Letter code**
Alanine	CH_3	Ala	A
Arginine	$(CH_2)_3NHC(NH)NH_2$	Arg	R
Asparagine	CH_2CONH_2	Asn	N
Aspartic acid	CH_2COOH	Asp	D
Cysteine	CH_2SH	Cys	C
Glutamic acid	$(CH_2)_2COOH$	Glu	E
Glutamine	$(CH_2)_2CONH_2$	Gln	Q
Glycine	H	Gly	G
Histidine		His	H
Isobutyric acid	CH_3, and CH_3 no αH	Aib	X
Isoleucine	$CH(CH_3)CH_2CH_3$	Ile	I
Leucine	$CH_2CH(CH_3)_2$	Leu	L
Lysine	$(CH_2)_4NH_2$	Lys	K
Methionine	$(CH_2)_2SCH_3$	Met	M
Phenylalanine	$CH_2C_6H_5$	Phe	F
Proline		Pro	P
Serine	CH_2OH	Ser	S
Threonine	$CH(OH)CH_3$	Thr	T
Tryptophan		Trp	W
Tyrosine	$CH_2C_6H_4OH$	Tyr	Y
Valine	$CH(CH_3)_2$	Val	V

List of Figures

List of Tables

Table of Contents

CHAPTER 1

BASIC FEATURES OF POLYPEPTIDES

1.1 Introduction

A protein is built up from a long chain polymer of amino acids called polypeptide chain. Polypeptides are more versatile because of the great number of different side chains that may be present. It is the variety of possible side groups that makes proteins so useful. There are four major categories of side chains; hydrophobic {glycine, alanine, valine, leucine, isoleucine, methionine, proline, phenylalanine and tryptophan}, acid hydrophilic {aspartic acid, glutamic acid and tyrosine}, basic hydrophilic {lysine, arginine and histidine}, and neutral {asparagine, glutamine, threonine, serine and cysteine (glycine, alanine, tyrosine, histidine and tryptophan also belong to neutral group)}.

Proteins have evolved for operation in an aqueous environment. The nonpolar side chains hold the protein molecule together. When a hydrocarbon chain is in an aqueous medium, it forces the neighbouring water molecules to form a cage-like structure. This formation restricts the motion and number of possible arrangements of the water molecule. If hydrocarbon side chains are segregated in one place instead, then the liberated molecules of water are free to adopt a much less ordered arrangement. The segregation of hydrophobic side chains is a powerful factor in stabilising a protein molecule in an aqueous solution.

The neutral polar residues are usually outside the molecule, but can be inside if their polar groups are neutralised by hydrogen bonding to other like residues or to the carbonyl CO group of the main chain. Asparagine, serine, threonine and glutamine are often used to cross-link two chains by means of hydrogen bonds. Tyrosine and tryptophan have been found inside and outside, but when tyrosine is inside, its OH group is always hydrogen bonded.

The charged polar groups, acidic or basic, can exist in either uncharged or charged form, depending on the pH of the surroundings. Under acidic conditions, aspartic acid and

glutamic acid have an uncharged carboxyl group, whereas histidine, lysine and arginine each are protonated and carry a positive charge. Under basic conditions, the carboxyl groups of aspartic acid and glutamic acid will be ionised, and histidine, lysine and arginine will be uncharged. The actual ratio of the acidic to basic form of a given residue depends on its strength as an acid or base.

Under normal physiological conditions, near pH 7, aspartic acid and glutamic acid will be almost entirely in their basic form. Lysine and arginine will be in their acid form, possibly charged, but histidine will be largely uncharged. It will be about 10% protonated and is capable of playing a dual role. The tyrosine OH group is weakly acidic. Only about 0.1% will be ionised at pH 7, and hence tyrosine has been classed as an uncharged polar group.

All the natural amino acids are L-amino acids. Most amino acids have more than one role in nature. The hydrocarbons alanine, valine, leucine and isoleucine increase in bulk and vary in shape. As they increase in bulk, the hydrophobic character increases correspondingly. Phenylalanine has the ability to interact with other aromatic rings by means of overlapping π electron clouds, a property it shares only with tyrosine and tryptophan. Besides aromatic ring, tyrosine and tryptophan have different functionalities.

Cysteine plays a crucial role in determining the folding of many proteins because of the ability of two such residues on different polypeptide chains to be oxidised to form a disulphide bridge. Proline, the only amino acid in which the side chain loops back to reattach to the main, has the property of forcing a bend in the main chain.

Other residues have been examined for other influence on α-helix formation. Generally it has been suggested that side chains which branch at the β carbon would be so bulky as to make an α-helix unstable if they occurred adjacent to one another on successive turns of the helix, or every third or fourth residue along the chain. It is striking that most residues, no matter how complex their sides chain, do have a compact CH_2 group at their β carbon.

Proteins are polymers, often cross-linked but never branched and the reason lies in the way in which proteins are made. The two most common types of cross-linking are a covalent disulphide bridge with a bond strength of the order of 209 kJ/mole and a weaker hydrogen bond with about 25 kJ/mole. Although the protein is synthesised as a linear polymer, its subsequent quite specific folding is determined entirely by the kind of distribution of side chains. But there may be other factors which could account for global folding of a protein in an aqueous environment.

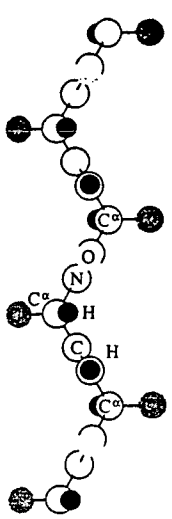
Many of the special properties of a polypeptide chain arise from the nature of its backbone chain. Its distinctive feature is the -CO-NH- group called a peptide bond. The CO and NH groups are capable of forming cross-linking between chains when building up three dimensional structures. The peptide bond also severely limits the ways in which the chain can fold: all four atoms in the -CO-NH- group have to lie in the same plane.

1.2 Regular Conformations of Polypeptides

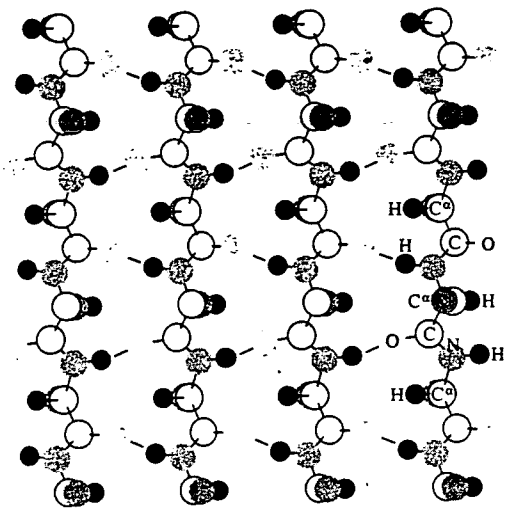
The random coil might be considered the natural state of a polymer, favoured by its conformational entropy and interactions with the solvent. However, other conformations will be adopted if sufficient interactions are possible, within or between molecules. The regularity of the conformation is a result of regularity of the primary structure. Each residue, or short sequence of residues that makes up the repetitive unit, will tend to adopt the same conformation; therefore, it may be specified by just a few dihedral angles.

The polypeptide chain then will have some form of a helical conformation, which will be characterised by the number of residues per turn (n) of helix and by the distance traversed along the helix axis per residue (d). The product of these is the pitch of the helix (p) [p=d.n]. For a polypeptide chain of fixed dimensions, both n and d are determined once ϕ and ψ angles are specified. The Ramachandran plot[1] shows the variation of ϕ and ψ angles for particular conformation. The values for the regular conformations are given in Table 1.1.

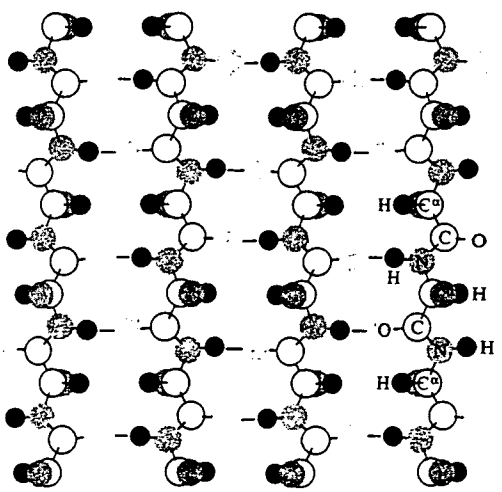
Figure 1.1 : Peptide conformations; A single β-strand (A), parallel β-sheets (B), antiparallel β–sheets (C) and the right-handed α-helix (D)



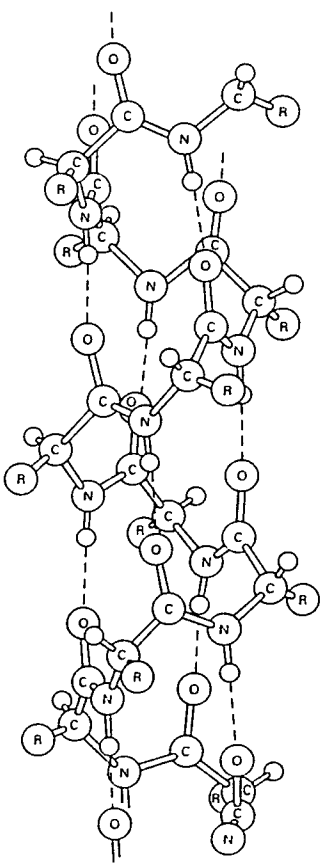
A



B



C



D

1.2.1 The α-Helix

The right-handed α-helix is the most well known and prominent of the polypeptide regular structures (Figure 1.1). It has 3.6 residues per turn and a translation per residue of 1.5Å, or 5.41Å per turn. The torsion angles ψ and ϕ favourable for most residues, and the atoms of the backbone pack closely, making very favourable van der Waals contacts. Most conspicuously, the backbone carbonyl oxygen of each residue hydrogen bonds to the backbone NH of the fourth residue along the chain. These hydrogen bonds are 2.86Å long from the O atom to the N atom and are very nearly straight, close to the optimal geometry for such an interaction, and are nearly parallel to the helix axis. All the hydrogen bonds point in the same direction.

The side chains project out from the helix and do not interfere with it, except in the bulkiest examples. Only proline residues are incompatible with this conformation, because the side chain is bonded to the backbone N atom, preventing its participation in hydrogen bonding and interfering in the packing. The stereochemical properties of the α-helix are so favourable that it is often considered the most natural conformation for a polypeptide.

Table 1.1: Parameters for Regular Polypeptide Conformations

	Bond Angle (degrees)		Residues Per Turn	Translation Per	
	ϕ	ψ		Residue (Å)	Turn (Å)
Antiparallel β-sheet	-139	+135	2.0	3.4	6.8
Parallel β-sheet	-119	+113	2.0	3.2	6.4
π-Helix (4.4_{16} helix)	-57	-70	4.4	1.15	5.06
α_R-Helix (3.6_{13} helix)	-57	-47	3.6	1.5	5.4
3_{10}-Helix	-49	-26	3.0	2.0	6.0
2_7 Ribbon	-75	70	2.0	2.80	5.6
Polyproline I	-83	+158	3.33	1.9	6.3
Polyproline II	-78	+149	3.00	3.12	9.36
Polyglycine II	-80	+150	3.0	3.1	9.3

A left-handed α-helix is also sterically possible, with the same values of ψ and ϕ but of opposite sign. The one important difference in left-handed structures is that the CO and NH groups are oriented almost perpendicular to the helix axis and are in no position to form hydrogen bonds with groups on the same chain. However, such a conformation is not favourable energetically, as the side chains are in close contact with the backbone, and it is generally not observed.

1.2.2 Anitparallel and Parallel β-sheets

In the β-sheet conformation, the polypeptide chain is nearly fully extended, and individual strands aggregate side by side, forming hydrogen bonds between the carbonyl and NH groups of the backbone (Figure 1.1). In addition to the hydrogen bonds, the dipoles of the peptide bonds alternate along the chain, providing favourable conditions for interaction. The adjacent strands may be either parallel or antiparallel; the two forms differ slightly in dihedral angles (Table 1.1), but both are sterically favourable. Both types of sheets can be found in proteins.

Most sheets that have been observed in detail are not planar but have a twisted conformation. The values of ϕ and ψ are both somewhat more positive in value than those given in Table 1.1 and presumably are somewhat favoured energetically, to give a right-handed twist to the backbone.

The extended conformation of the polypeptide chain results in the side chains protruding on alternating sides of the sheet. Most amino acids are stereochemically compatible with the β-sheet conformation, except for proline, which has no NH group to participate in hydrogen bonding and cannot adopt the appropriate value of ϕ. β-Sheets may involve aggregation of different molecules, or the polypeptide chain may loop back on itself to form an intramolecular sheet, known as the cross-β conformation.

1.2.3 Other Regular Conformations [3_{10} and 4.4_{16} (π) Helices]

Other regular conformations have been observed in proteins and are seen only with certain polypeptides in special instances. Variations on the α-helix in which the chain is either more tightly or more loosely coiled, so that hydrogen bonds between

corresponding groups are closer or further apart in the primary structure by one residue, are designated the 3_{10}-helix or the 4.4_{16} (π) helix, respectively. The packing of the backbone atoms is too tight in the 3_{10}-helix; it has not been observed as a regular structure but occurs only at the ends of α-helices, where one turn may have this conformation locally. The π-helix would have a hole down the middle, so the backbone atoms would not be in contact, and the values of ϕ and ψ are rather unfavourable.

Proline residues are incompatible with both α-helix and β-sheet conformations, so it is not surprising that poly(Pro) forms other regular conformations, known as poly(Pro) I and II. Form I contains all cis peptide bonds, whereas II has trans; proline is the only amino acid where the cis form is generally significant. The former is a right-handed helix with 3.3 residues per turn, whereas the latter is a left-handed helix with 3.0 residues per turn. Which form is adopted depends primarily on the solvent: Form II predominates in water. Glycine residues also have unique conformational flexibility and poly(Gly) likewise forms two regular conformations, designated I and II.

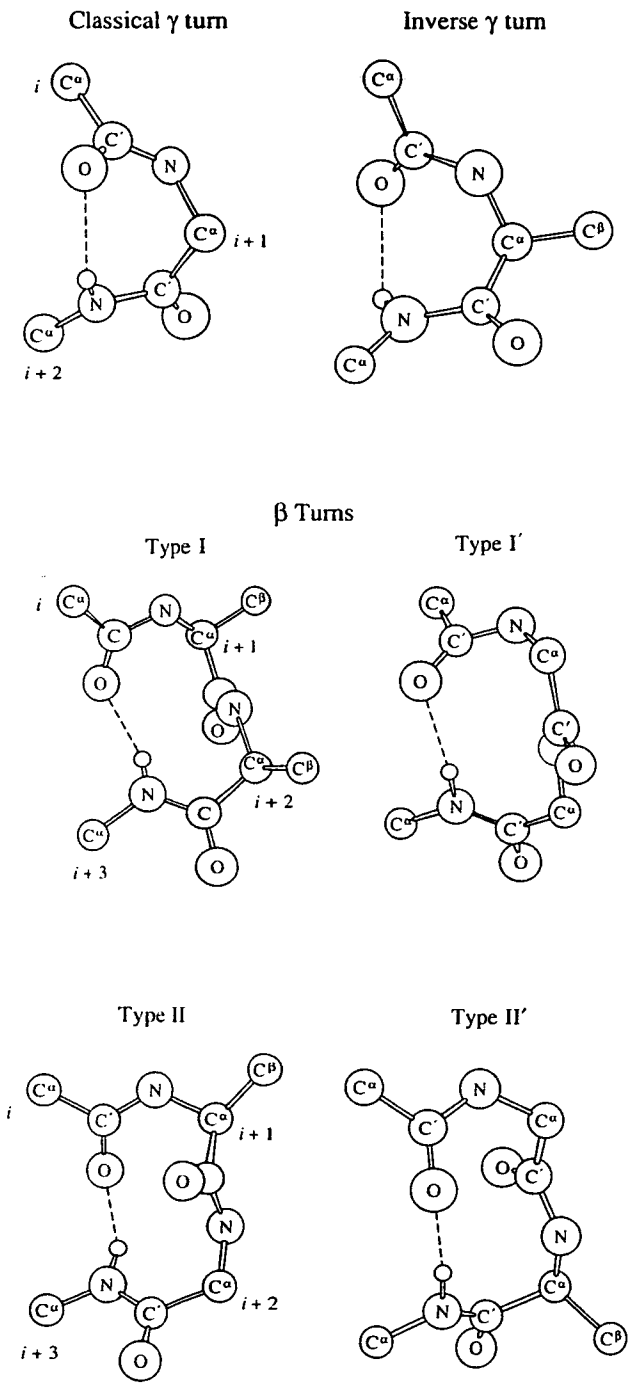
1.2.4 2_7-Ribbon

The helix coils tighter around its axis as ψ increases, with five, then four, three and finally two residues per turn. The 2_7-ribbon, which has 2 residues per turn and 7 atoms in the closed ring, is tolerable only if the too close oxygens and hydrogens are hydrogen bonded, and even the produces a shorter than normal hydrogen bond.

1.2.5 β and γ Turns

Proteins have rather spherical structures because the polypeptide chain generally makes rather sharp bends at the surface, thereby reversing the direction of the polypepitde chain. Turns are intrinsically polar structures with backbone groups pack together closely and side chains project outward. Presence of turns in bioactive conformations may in fact reflect the lack of alternative conformational possibilities. The terms β and γ turn have more restricted definitions and describe turns of four or three residues, respectively. These turns may or may not be stabilised by an intra-turn hydrogen bond; in β-turns, the CO of residue i may be hydrogen bonded to the NH of residue i+3, while in γ-turns, the CO of residue i may be hydrogen bonded to the NH of residue i+2.

Figure 1.2 : The most common types of turns. The hydrogen bond is shown as a dashed line.



Substantial fractions of residues of every protein are involved in β-turns. They are also known as hairpin bends, β-bends, and reverse-turns, because they often connect antiparallel β-strands (Figure 1.2). The terms "open" β or γ turns will be used for situations in which no hydrogen bond exists and the ϕ, and ψ angles are within 30 degrees of the ones cited in Table 1.2.

Four consecutive residues in the polypeptide chain are generally considered to comprise a β-turn, although it is only the torsion angles of the second and third residues that are critical. The first and fourth residues are usually included because a hydrogen bond between their backbone groups was originally considered necessary. Three such ideal β-turns, generally designated I, II and III were predicted on the basis of allowed geometry, with planar trans peptide bonds. Mirror images of the backbone may occur in variants I', II' and III'. All standard β-turns orient the side chains of residues in position i+1 equatorially and those of the residue in position i+2 axially (up and down).

Type I is compatible with any amino acid residue at position 1 through 4, except that proline cannot occur at position 3. In contrast, type I' requires glycine at both positions 2 and 3. Type II and II' require glycine at position 3 and 2, respectively. Type III is a portion of a 3_{10}-helix and any amino acids are permissible; type III' requires glycine at positions 2 and 3. Serine is a frequent participant in turns in proteins.[3] The special amino acids glycine and proline are often involved in reverse-turns.

Most examples of γ-turns in linear peptides are only in poor solvents, particularly chloroform. It appears that solvation of amide NH and CO groups by hydrogen bonding solvents competes effectively with the intramolecular interactions. This is in contrast to β-turns which are observed in good hydrogen bonding solvents.

Theoretical calculations of preferred β-turn conformations led to a fundamental distinction between turns predicted for homochiral and heterochiral residues at positions i+1 and i+2 positions. The former were predicted to favour type I (L-L) or I' (D-D) β-turns and the latter type II (L-D) or II' (D-L) β-turns. Glycine could be accommodated at any position.

In the experimental data available to date from turn forming model peptides may lead to some general conclusions.

(1). β-Turns are stable in a variety of environments, including strong solvents and crystals. Hence, intramolecular hydrogen bonding is not playing an important role.
(2). γ-Turns are stable only in the absence of competing hydrogen bonding interactions with solvent. Hydrogen bonding seems to play a major role in their stability.
(3). Proline is strong favoured in position i+1 of a β-turn.
(4). Glycine promotes formation of β-turns favoured by heterochiral sequences.
(5). More examples of heterochiral turns were observed than homochiral sequences.
(6). L amino acid residues in position i+1 adopt inverse γ-turns.
(7). Bulky hydrophobic residues do not occur readily in position i+2 of β-turns, and that serine has a preference for position i+2 of β-turns.

Table 1.2: Dihedral Angles of Hydrogen Bonded β– and γ–Turns

Turn	ϕ_{i+1}	ψ_{i+1}	ϕ_{i+2}	ψ_{i+2}
β-Turns				
Type I	-60	-30	-90	0
Type I'	60	30	90	0
Type II	-60	120	80	0
Type II'	60	-120	-80	0
Type III	-60	-30	-60	-30
Type III'	60	30	60	30
γ-Turns				
Turn	70 to 85	-60 to -70		
Inverse Turn	-70 to -85	60 to 70		

In summary, the only regular conformations encountered with polypeptides under normal conditions are the α-helix, parallel and antiparallel β-sheets, and poly(Pro) II helix. The last one occurs only with two specific polyamino acids, so only the α-helix and the β-sheets are regular conformations likely to be encountered with a typical protein sequence.

1.3 The General Properties of Protein Structures

Protein structure can be discussed in terms of four levels. The primary structure is the amino acid sequence. The secondary structure is any regular local structure of a linear segment of polypeptide chain, such as a helix, a sheet or a turn. Tertiary structure is the overall topology of the folded polypeptide chain and the quaternary structure is the aggregation of the polypeptides by specific interactions.

Secondary structure in proteins is generally somewhat distorted. In the α-helices, the plane of the peptide bond is often rotated. β-Sheets are generally twisted, rather than planar, with a right-handed twist of form 0-30 degrees between strands. Further distortions occur in β-sheets. An extra residue is often present in a strand at the edge of a sheet, interrupting the hydrogen bond pattern and producing a "β-bulge". The segments of α-helices and β-sheets are generally rather short, being limited to the diameter of the protein globule. The length of an α-helix is generally 10 to 15 residues, while that of a β-sheet is 3 to 10 residues.

It is often difficult to define exactly which residues are part of the secondary structure. For example, at both ends of an ideal α-helix there are four residues that participate in only one hydrogen bond each within the helix, whereas all other interior residues participate in two. Moreover, the ends are often irregular in conformation. Which residues should be counted as part of the helix, therefore is often not clear. Secondary structure is most apparent in the large proteins, where most of the interior is composed of such regular structure. One important property of secondary structure is that it provides an efficient way of pairing in hydrogen bonds the internal polar groups of the polypeptide backbone.

Many studies have been made of the occurrence of the 20 amino acids in the various structural elements of proteins. Only arginine shows no tendency to occur preferentially in any particular structure. Amino acids with a branched or bulky side chain (valine, isoleucine and threonine) or aromatic residues occur most frequently in β-sheets. All the rest occur most often in α-helices, except for those with short polar (serine, aspartic acid and asparagine) or special side chains (glycine and proline), which occur most often in reverse-turns.

1.4 Spectral Properties on Proteins

The proportion of time the peptide spends in each of its conformational states and the rate of interconversions must be considered in choosing methods of conformational analysis and in interpreting data. Conformational interconversions may occur at millisecond or faster time scales in linear peptides and peptide bond rotations occur on the time scale of seconds.[4]

1.4.1 Fluorescence Spectroscopy

The distance between specific groups of polypeptides may be estimated spectrally if the one group is a fluorescent energy donor and the other a suitable energy acceptor. If the absorption spectrum of the acceptor overlaps the emission spectrum of the donor, fluorescent light emitted by the donor will be absorbed by the acceptor. The efficiency of this process depends upon the sixth power of the distance between them, most usefully with in a range of 10 to 60 Å. The observed efficiencies of energy transfer could be used to reconstruct the distribution function of the distances between the groups.

1.4.2 ORD and CD Spectroscopy

Optical rotatory dispersion (ORD) and circular dichroism (CD) spectroscopy may be directly sensitive to the conformations of the polypeptide chain. Of greatest use is the optical activity of polypeptides due to the asymmetric centres of the L-amino acids and to their asymmetric conformations. If the right-circulary and left-circularly polarised light beams consequently travel at different speeds through the molecule, polarised light is rotated and the ORD spectrum can be obtained. There will also be unequal absorption of left-circulary and right-circularly polarised light, which is known as CD. Both phenomena have the same cause and consequently are related. Circular dichroism is exquisitely sensitive to conformation and has a very fast time scale (10^{-15} s).[5] Detailed interpretation of CD spectra for molecules larger than di- or tripeptides is difficult. Furthermore, several different conformational features yield similar CD spectra. It appears that γ-turns give rise to a reliable CD band at long wavelength (~230 nm) and β-turns yield three curve shapes that correlate with turn type.

Examples of the ORD and CD spectra of poly(Lys) in random, α-helical and β-sheet forms are shown in Figure 1.3. These spectra reflect primarily the conformation of the backbone.

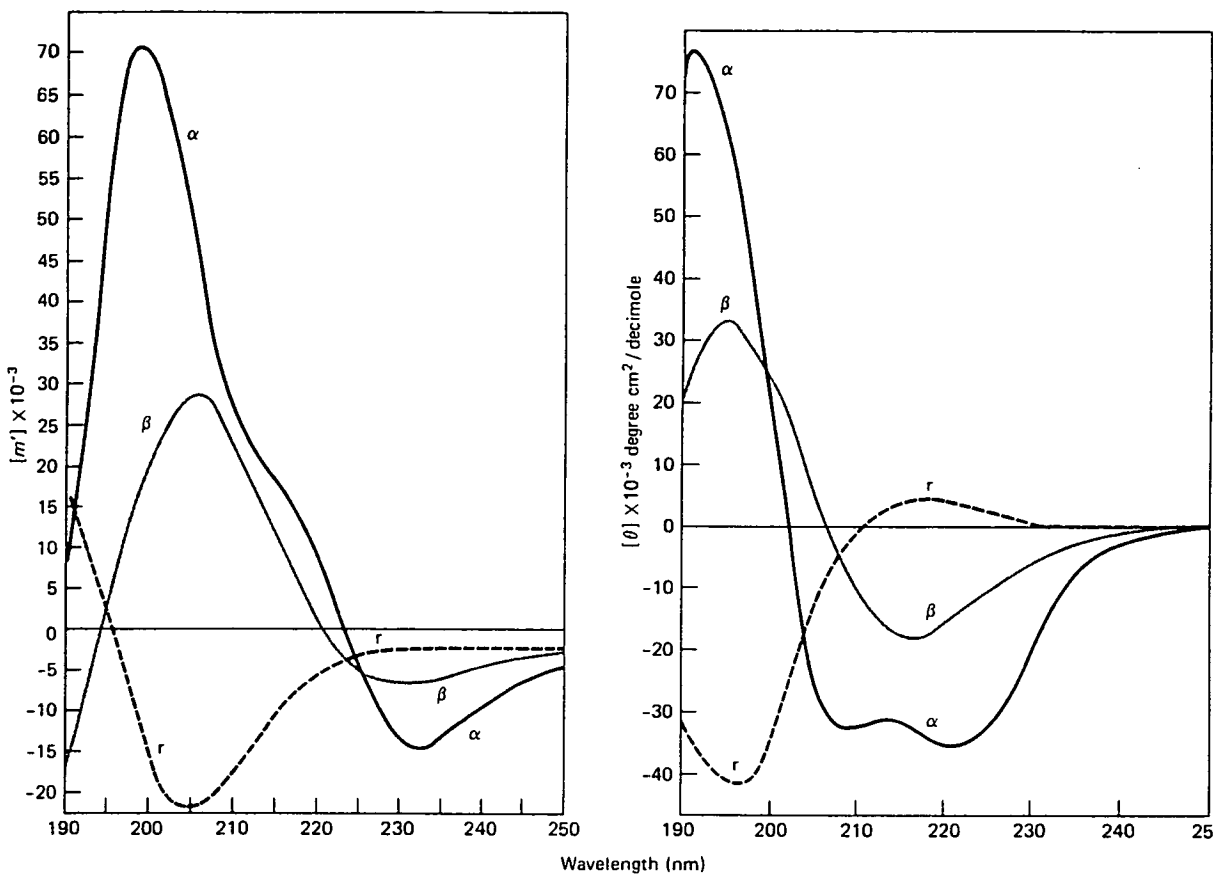


Figure 1.3: Optical rotatory dispersion (left) and circular dichroism (right) spectra of poly(Lys) in the α-helical (α), antiparallel β-sheet (β) and random coil (r) conformations.[6]

1.4.3 NMR Spectroscopy

Nuclear magnetic resonance (NMR) spectroscopy is the most useful method of peptide conformational analysis in solution. NMR yields information about chemical environments of nuclei (chemical shifts), geometric relationships between nuclei (coupling constants), distances between nuclei (nuclear Overhauser enhancements), accessibility and hydrogen bonding of amide protons (exchange kinetics and sensitivity of resonance positions and linewidths to temperature, solvent) and dynamics of nuclei (relaxation times). More details of NMR studies of proteins are discussed elsewhere.

CHAPTER 2

PRINCIPLES OF NMR SPECTROSCOPY

2.1 Importance of NMR Spectroscopy on Polypeptides

NMR experiments for biopolymers were described over three and half decades ago and progress was slow because of the available instrumentation and lack of biological macromolecule samples. Since approximately 1981, the methods and instrumentation have been developed for obtaining nearly complete sequence-specific resonance assignments in biopolymers. These provide a basis for systematic procedures for obtaining spatial structure determination of noncrystalline biopolymers.[7]

The most important methods for determining the structures of molecules are X-ray diffraction and NMR studies. The X-ray diffraction technique is not particularly well suited for studies of peptides because many different conformations may be developed in the non-crystalline phase. NMR allows, in principle, analysis of peptide samples in the solution phase. NMR spectroscopy is the leading technique for obtaining structural and dynamic information at the atomic level about proteins in solution. It provides an efficient method for measuring dynamic, kinetic and thermodynamic parameters (e.g., correlation times, hydrogen exchange rates and pKa values) in peptides and small proteins.

NMR analysis is complementary to crystallography in providing a means of determining whether the protein structure is the same in solution as in the solid state. X-ray diffraction and high resolution NMR have different strengths and different experimental requirements. With X-ray crystallography, it is necessary first to obtain suitable protein crystals; the ultimate spatial resolution depends on properties of the crystal and the effort expended in refining the structure. An X-ray study provides few answers at initial stages. The major hurdles are crystallization, preparation of suitable derivatives and solution of the phasing problem.

By contrast, NMR analysis can yield limited, but useful, molecular information from the first spectrum: a resolved coupling constant, for example, can indicate the existence of a

particular chemical bond between two atoms, or a nuclear Overhauser effect can define the close proximity of two hydrogens in the molecule. The central problems in NMR spectroscopy are the resolution of signals from individual groups and their assignment in a sequence-specific and stereospecific manner.

Relatively precise measurements of short (< 5Å) interproton distances can be made between any pair of hydrogens whose ^{1}H signals can be resolved and assigned. Dynamic and static disorder usually can be distinguished by NMR results, and NMR can provide rates of dynamic processes over a wide time scale (nanoseconds to seconds). In very large proteins, the sharpest and most easily resolved NMR signals come from less ordered regions; by contrast, mobile regions in crystals frequently do not yield interpretable electron densities.

Other differences can result from differential solubility or differential crystallization: an NMR spectrum provides signals from all species present in solution with the signal strength proportional to their concentrations. An X-ray analysis shows the structures of those species that have crystallized from solution and their relative abundance may differ from that of the original solution. As has been the case with several high resolution X-ray structures, NMR analysis can provide evidence for multiple stable conformational forms of a protein. The effective molecular weight places limits on the degree of detail to be obtained from NMR analysis. At present, one can contemplate determining detailed solution structures of nonaggregating proteins of MW 20 Kda. Much more limited one dimensional and two dimensional investigations are being carried out with larger proteins MW up to 100 Kda.

The great power of two dimensional NMR methods[8,9] for protein spectroscopy stems from their ability to assist in solving the problems of resolution and assignment. Application of these methods[10] has led to nearly complete ^{1}H resonance assignments in numerous small proteins and extensive ^{13}C assignments[11] and ^{15}N assignments[12] are proceeding rapidly. A list of polypeptides for which NMR assignments or solution structures had been derived is summarrized by J.L. Markley.[13]

The wide selection of 2D NMR methods currently available for solution studies sort into two categories based on the physical mechanism mediating interactions between spins: cross-relaxation, chemical exchange or spin-spin coupling. The cross-relaxation experiment most important for protein studies is 2D $^{1}H\{^{1}H\}$ nuclear Overhauser enhancement spectroscopy. The same kinds of pulse sequences employed to determine NOE's are used to investigate chemical exchange phenomena. The most important nuclei involved in spin-spin couplings in proteins are ^{1}H, ^{13}C and ^{15}N and the interactions may be homonuclear or heteronuclear in nature. To date, the vast majority of 2D NMR studies of proteins have involved $^{1}H\{^{1}H\}$ homonuclear interactions, because of the high sensitivity of the proton and its high natural abundance. Proton-X-nucleus heteronuclear correlation experiments are performed more efficiently, however by detecting ^{1}H. Isotope labelling also can enable 2D NMR approaches that take advantage of coupling between insensitive nuclei such as ^{13}C-^{15}N [14,15] and ^{13}C-^{13}C.[16,17]

2.2 History of the NMR Spectroscopy

NMR spectroscopy is one of the most important tools for obtaining detailed information on chemical systems at a molecular level. It is based on the fact that atomic nuclei oriented by a magnetic field (0.70 -17.6 T) absorb radiofrequency radiation at particular frequencies. During last two and half decades there has been a enormous developments in the instrumentation and methods of NMR spectroscopy. These include; the construction of higher field spectrometers and accessories, development of FT methods and multi-dimensional NMR techniques and computer technology for data processing and analysis.

The first NMR experiments were carried out in 1945 by two groups of physicists (Purcell et al.[18,19] at Harvard University and Bloch et al.[20,21] at Stanford University) working independently in solids and liquids. Useful chemical applications of NMR became possible only after the discovery of the chemical shift effect[22,23] in 1949.

First high resolution NMR spectra were recorded by the continuous wave (CW) technique. In this method, the spectra is recorded point by point while applied magnetic field (**B**$_0$) or transmitter frequency (ν_1) is altered continuously. The appearance of the first commercial CW proton NMR spectrometer in 1953 increased rapidly the use of NMR as a chemical tool. The CW method was really only suitable for recording the spectra of sensitive nuclei such as ^{1}H, ^{31}P and ^{19}F. In 1966, introduction of the pulse Fourier Transform (FT) method by Ernst and Anderson[24] led to a significant gain in sensitivity and allowed for the first routine observation of low abundance nuclei such as ^{13}C and ^{15}N. In 1971, J. Jeener[25] proposed the idea of two dimensional FT NMR and at the same time commercial FT NMR spectrometers were introduced. Since then the development of NMR spectroscopy has been extremely rapid.

2.3 Theory of NMR Spectroscopy

The magnetic properties of the atomic nucleus form the basis of nuclear magnetic resonance spectroscopy. The fundamental property of the atomic nucleus involved is the nuclear spin quantum number (I) or angular momentum quantum number which has values of 0, $^1/_2$, 1, $1^1/_2$, 2 .....upto 6 in units of $h/2\pi$ where h is the Plank's constant. Many nuclei including ^{1}H, ^{13}C, ^{15}N and ^{19}F which have odd mass numbers and half integral values of nuclear spin possess nuclear angular momentum (**P**). According to the quantum theory, this angular momentum is quantized.

$$|\mathbf{P}| = h\sqrt{I(I+1)}/2\pi \quad \text{..... 2.3.1}$$

The angular momentum **P** has associated with it a magnetic moment (μ). The two vector quantities are proportional to each other and the proportionality constant γ is called the gyromagnetic (magnetogyric) ratio.

$$\mu = \gamma\,\mathbf{P} \quad \text{..... 2.3.2}$$

By combining equations 2.3.1 and 2.3.2, the magnetic moment μ can be obtained. According to the quantum theory, the magnetic moment is also quantized.

$$|\mu| = \gamma h\sqrt{I(I+1)}/2\pi \quad \text{..... 2.3.3}$$

In a magnetic field (**B**$_0$), the orientation (θ) of μ to the field direction (z-axis) is also quantized and depends on I. The orientation to the x- and y-axes is not quantised. When a magnetic field (**B**$_0$) is applied along the z-axis, the angular momentum component $\mathbf{P}_z$ which takes up an orientation of the applied field is equal to multiple of

m_I and $h/2\pi$ where m_I is the magnetic quantum number which characterises the corresponding stationary states of the nucleus.

$$|\mathbf{P}_z| = m_I\, h/2\pi \quad \text{..... 2.3.4}$$

As the energy of the nucleus is quantized then m can take any of the values m_I = I, I-1, I-2, ...-I giving 2I+1 possible orientations (Figure 2.1). From equations 2.3.2 and 2.3.4, magnetic moment along the applied field direction z (μ_z) can be obtained.

$$|\mu_z| = m_I\, \gamma\, h/2\pi \quad \text{..... 2.3.5}$$

The magnetic moment μ never lines exactly with the field direction but is at an angle $\{\cos\theta = m/\sqrt{I(I+1)}\}$ and behaves as if it precess about the field direction with an angular velocity (ω) given $\omega = -\gamma\, \mathbf{B_0}$ (Figure 2.2) where ω is referred to as the Larmor frequency.

The energy (E) of a magnetic dipole is simply proportional to the magnetic moment (μ_z) and the applied field ($\mathbf{B_0}$).

$$E = -\mu \bullet \mathbf{B_0} = -\mu \cos\theta \bullet \mathbf{B_0} = -\mu_z\, \mathbf{B_0} \quad \text{..... 2.3.6}$$

$$E = -m_I\, \gamma\, h\, \mathbf{B_0}/2\pi \quad \text{(with the aid of 2.3.5) ..... 2.3.7}$$

The selection rule for NMR transitions between orientations is $\Delta m = \pm 1$ thus the energy difference between two adjacent energy levels can be given as

$$\Delta E = \gamma\, h\, \mathbf{B_0}/2\pi \quad \text{..... 2.3.8}$$

Transition between two adjacent energy levels may be induced by electromagnetic radiation of frequency ν where $\Delta E = h\nu$ and thus this absorption frequency ν is given by

$$\nu = \left({}^{\gamma}/_{2\pi} \right) \mathbf{B_0} \quad \text{..... 2.3.9}$$

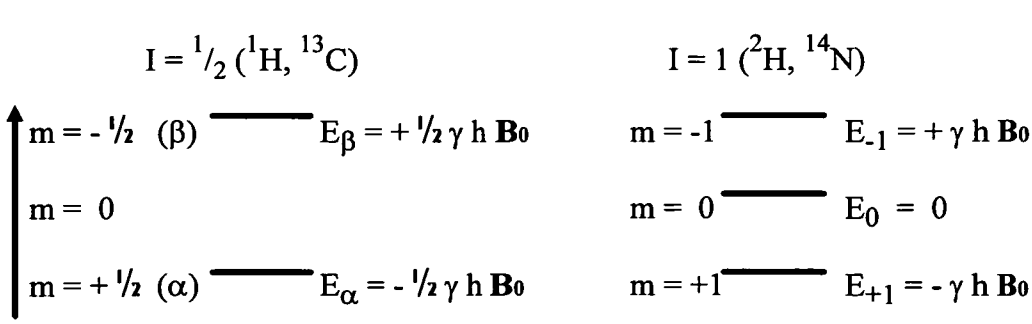


Figure 2.1 : Energy level schemes for nuclei I = $^1/_2$ and I = 1

Thus there are two possible orientations of the nuclear spin for protons. The ratio of the populations between these two energy states is given by the Boltzmann equation.

$$N_\beta/N_\alpha = \exp\left(-\Delta E/_{kT}\right)$$

$$= 1 - \Delta E/_{kT} \text{ (as } \Delta E << kT) \text{ (ca. 1 in } 10^5 \text{ for } {}^1H) \quad ... \quad 2.3.10$$

where N_α and N_β are the numbers of nuclei in the ground and excited states respectively. When $N_\beta < N_\alpha$, an NMR signal can be observed due to net absorption of energy.

2.3.1 Behaviour of an ensemble of spins in a magnetic field; The NMR experiment

Consider a sample contain only a single type of spin half nucleus (eg. $CHCl_3$). Vector diagrams can easily be used to understand the behaviour of the magnetisation in a magnetic field. In pictorial diagrams, z-axis usually points along the static magnetic field ($\mathbf{B_0}$) direction. The net magnetisation, **M**, arises from the net sum of all the precessing nuclear magnetic moments along the magnetic field direction (z-axis). This net magnetisation will also be aligned with the applied magnetic field and will remain undisturbed (Figure. 2.2).

The second magnetic field, $\mathbf{B_1}$, which oscillates at the appropriate radiofrequency displaces the magnetisation μ from the z-direction. This magnetic field $\mathbf{B_1}$ which is generated by a current in a coil is perpendicular to the static magnetic field $\mathbf{B_0}$. After the pulse from $\mathbf{B_1}$ a force generated by $\mathbf{B_0}$ on **M** is a torque that will cause **M** to precess about $\mathbf{B_0}$ at a frequency $(\gamma/2\pi)\mathbf{B_0}$ hertz. This motion is known as Larmor precession as described before. The component of the precessing magnetisation in the xy-plane would be detected as a NMR signal. The final position of the magnetisation depends upon the length of time (μs) for which the radiofrequency is applied. The angle (θ) which the magnetisation is tipped from the z-axis is called the flip angle or pulse angle.

The visualisation of NMR experiments can be simplified using 'rotating reference frame' rather than fixed laboratory frame. The effect of a pulse is easy to describe in the rotating reference frame. If the rotating set of coordinates is chosen to rotate at the frequency of the pulse in the same direction as the nuclear precession, then the $\mathbf{B_1}$ field

appears static along, say, the x-axis in that frame. In this frame the field direction z-axis does not change but x-and y-axes rotate about the z-axis at the pulse frequency. Individual nuclear magnetic movements precess clockwise or anti-clockwise according to whether their resonance frequencies are greater or less than the pulse frequency.

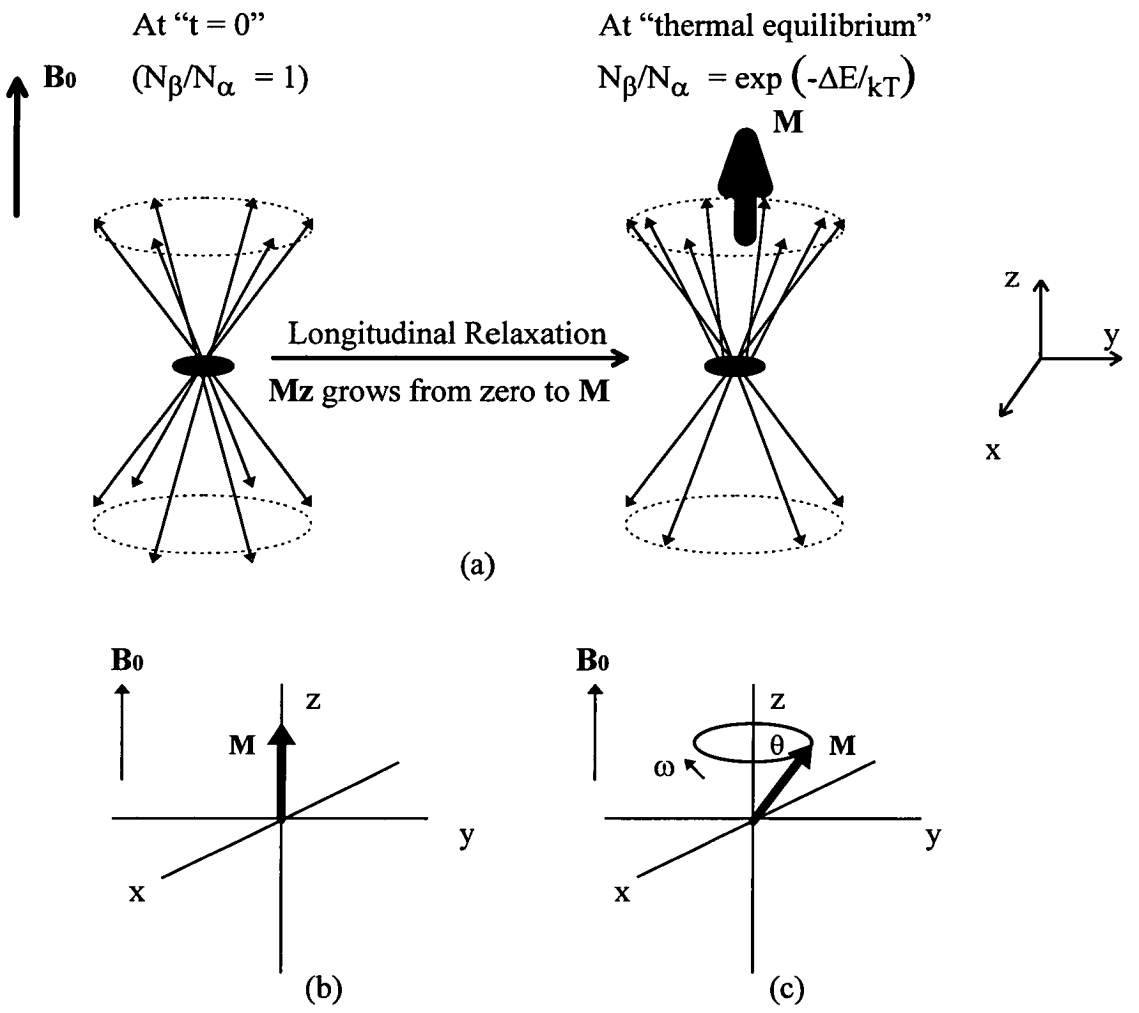


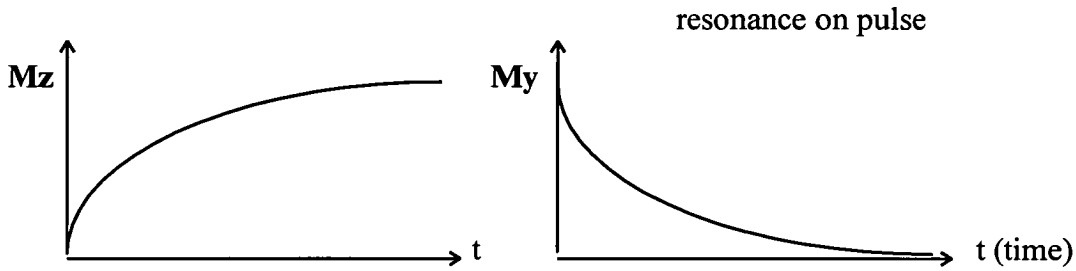
Figure 2.2 : Growing of the net magnetisation **M** (a), the **M** at equilibrium (b) and at precession (c).

As in the laboratory frame, the net magnetisation begins along the field axis, z, and perpendicular to x-axis and the precession about $\mathbf{B_1}$ is a rotation of magnetisation in the yz-plane. The magnetisation will remain perpendicular to the $\mathbf{B_1}$ field and a $(\pi/2)x$ pulse will put the magnetisation along the y-axis (Fig. 2.3). When the pulse is switched off the magnetisation vector **M** precesses about the z-axis and its orientation can be specified in the rotating coordinate system by the three axial components **Mx**, **My** and **Mz**. The spin system now returns to its equilibrium state by relaxation with z component,

Mz, returning to its original position while x and y components, **Mx** and **My**, approach zero. The motion of the magnetisation vector is described by two different relaxation times T_1 (spin-lattice or longitudinal relaxation time) and T_2 (spin-spin or transverse relaxation time). The Bloch equations for relaxation in the rotating frame are given below.

$$\frac{dMz}{dt} = -\frac{(Mz - M)}{T_1} \quad \text{..... 2.3.11}$$

$$\frac{dMx}{dt} = -\frac{Mx}{T_2} \quad \text{and} \quad \frac{dMy}{dt} = -\frac{My}{T_2} \quad \text{..... 2.3.12}$$



Mz returns to **M** — Longitudinal Relaxation

My decays to zero — Transverse Relaxation

In the rotating frame, for a nucleus with resonance frequency equal to the pulse frequency, the y-axis rotates at the Larmor frequency and the transverse magnetisation in that direction remains constant while its magnitude only decreases with time. According to the equation 2.3.12, this exponential decrease rate is determined by the transverse relaxation time T_2. This decay of the transverse magnetisation which is called the free induction decay (FID) is detected as a NMR signal. The z-component, **Mz**, returns to its original position via longitudinal relaxation.

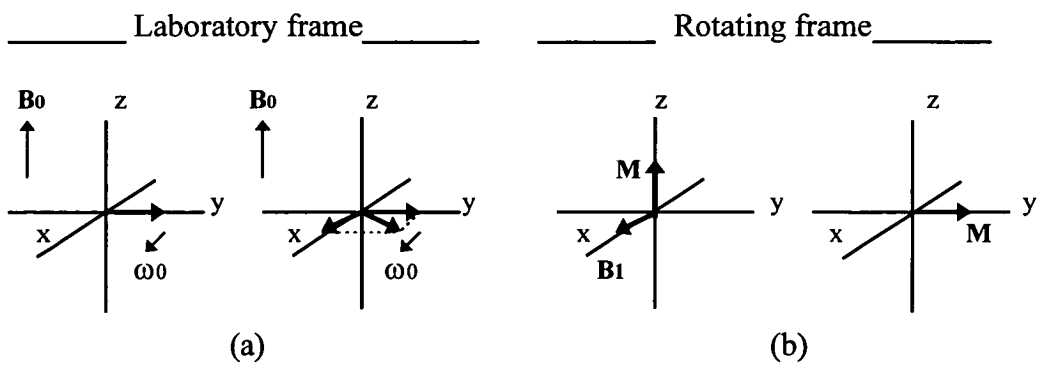


Figure 2.3 :Magnetisation vectors at laboratory frame (a) and rotating frame (b) after the $(\pi/2)x$ is applied.

2.3.2 Chemical Shift

The chemical shift is caused by the distribution of electrons of the chemical bonds in the molecule. In fixed external $\mathbf{B_0}$, the exact resonance frequency for a given nucleus, x, depends on its environment in the molecule. When a molecule containing nuclei is placed in the magnetic field ($\mathbf{B_0}$), it induces circulation in the electron cloud surrounding the nucleus such that a magnetic field opposed to $\mathbf{B_0}$, is produced. In other words the electrons within the molecule shield the nuclei from the applied field $\mathbf{B_0}$. This magnetic shielding is equal to $\sigma\mathbf{B_0}$ where σ is known as shielding constant which depends on the electron density surrounding the nucleus. Although this shielding is molecule orientation dependent, molecule reorientation in solution is normally sufficiently fast for orientation effects to be averaged.

The chemical shift (δ) as defined below is quoted in units of parts per million (ppm) and will be same whatever the strength of field used. The resonance frequency of the signal is measured relative to that of a reference or standard compound.

Chemical shift $\delta = \dfrac{\nu_x - \nu_{std}}{\nu_{std}} \times 10^6$

where $\nu_x = \left({}^{\gamma}/_{2\pi} \right) \mathbf{Bx}$ where the local field experienced by nucleus x, $\mathbf{Bx} = \mathbf{Bo}\,(1 - \sigma_x)$

2.3.3 Spin-Spin Coupling

The individual resonance signals of the different groups are chemically nonequivalent and therefore the splitting of the signals can be observed. A nucleus of spin I splits resonance of another nucleus (any spin) into 2I+1 lines. This arises from interactions between magnetic nuclei transmitted through bonding electrons. The splitting of signals can normally be seen no more than three bonds but some cases through four/five bonds depending on the geometry of the molecule. The fact that spin-spin coupling is transmitted through chemical bonds makes the coupling constant, **J**, a sensitive parameter for the types of bonds involved and for their spatial orientation in the molecule and is therefore geometry dependent. In complex molecules, it can be seen that each nucleus can interact with many others which can lead to very complicated multiplet spin patterns. The spin-spin couplings are field independent and are measured in hertz (Hz).

2.4 Solvent Suppression

Proteins have characteristic amino proton resonances which can be observed under appropriate conditions. To prevent loss of cross peaks with labile protons, NMR spectra must often be recorded in protonated solvents. Most of the polypeptides are soluble in highly polar organic solvents or solvent mixtures such as DMSO (dimethylsulfoxide), CH_3OH, CH_3OH/H_2O, $AcOH/H_2O$ and H_2O. The amide protons in the above solutions of polypeptides can be studied in above solutions.

Fourier transform proton NMR of dilute aqueous solutions is fraught with difficulties. The weak solute signals (0.001-0.01M) have to be detected in the presence of potentially enormous signal from the water protons (~ 110M), when the NMR spectra are recorded in H_2O solution.

Continuous weak radio frequency irradiation prior to excitation and acquisition to eliminate the water magnetisation is undoubtedly the simplest suppression method.[10,26,27] The following points must be important. Irradiation should ideally be gated off during data acquisition to avoid Bloch-Siegert shifts despite the reduction in the level of suppression this entails. The choice of decoupler power and irradiation time is usually a compromise. High decoupler power will give efficient elimination of the solvent resonance but will also saturate protons with nearby chemical shifts. Long pre-irradiation times give good saturation but reduce the efficiency of signal averaging if the delay between transients has to be increased.

Saturation of the solvent resonance also has an affect on the resonances of protons which exchange with the solvent. This is called saturation transfer and the principal drawback is that saturation transfer may reduce the intensity of exchanging protons.[28,29] Appreciable loss of signal is to be expected unless the exchange rate is at least five times slower than the spin-lattice relaxation of the proton in question. Cross-relaxation from water to solute protons can have a similar effect. Cutting down the length of irradiation gives less time for saturation transfer and cross-relaxation but requires a stronger and thus less selective decoupler field to achieve the same level of suppression.

Conditions can be chosen to minimise the rates of exchange and hence the loss of signal (acidic pH for proteins and basic pH for nucleic acids)[10] but this may be undesirable or even impossible depending on the sample. Inevitably some protons lie directly under the water or are sufficiently close to it in frequency that they do not escape saturation. Any protons bleached in this way are invisible in both one and two dimensional experiments, but can sometimes be revealed by exploiting the temperature dependence of the water chemical shift.

Strong solvent peaks give rise to ridges that may mask peaks of interest. This complication is overcome by various schemes such as

(1). Selective excitation strategies that minimise the solvent signal

(2). Selective saturation of solvent peaks

(3). Improvement of probe characteristics

(4). Postprocessing of the FT data by different methods

Interior amide protons in proteins can often be observed in D_2O, so that no special precautions are needed for suppression of the solvent line.

2.5 One Dimensional NMR Studies of Proteins

2.5.1 Variable Temperature Studies

Temperature, one of the important solution conditions for NMR studies, can usually be varied over a wide range. It is therefore important to maintain the temperature control whilst studying proteins. Overall structure of a protein can be sensitive to changes in temperature in particular proteins can denature in solution.

The temperature dependence of the chemical shift ($\Delta\delta/T$) is widely used as a measure of the extent of hydrogen bonding in amide protons. One of the reasons for its popularity is that $\Delta\delta/T$ is an easy parameter to measure. Interior amide protons hidden from the solvent shift 3-5ppb upfield per degree temperature increase, whereas solvent exposed amide protons shift more (6-8 ppb/K).

2.5.2 Amide Proton Exchange Experiments

Labile protons in amino acid residues have important roles in NMR studies of proteins. The intrinsic rates of exchange with the solvent are of great interest as a reference for

studies of macromolecules structure and dynamics. Most of the polypeptides are highly soluble in H_2O. The amide protons in the H_2O solution can be studied using selective saturation of the solvent line. When the resonance is suppressed by saturation, the labile protons in the molecule will be replaced by saturated spins from the huge spin reservoir of the solvent, with concomitant loss of resonance intensity. Different solvent suppression methods have been developed to overcome this problem.

^1H NMR observations of labile protons is only feasible when the exchange with the solvent is slow on the NMR time scale. The NMR time scale determined by the relative chemical shifts of the labile protons in proteins, water as the solvent, allows observation of protons exchanging with the rates up to approximately $1x10^{-3}$ min^{-1}. The limiting value of the exchange rate constant, which enables observation in D_2O, is of the order of 0.1 min^{-1}. In the entire pH range, surface amide proton exchange rates are slow on the NMR time scale defined by the chemical shifts relative to H_2O, but too fast for observation in D_2O, all the amide protons, therefore, can be observed in H_2O solution.

Because of the presence of labile protons, ^1H NMR spectra of proteins are simplified by changing the solvent from H_2O to D_2O. For surface amide protons, that are freely accessible to the solvent, the exchange rates are too fast for observation in D_2O. But some of the NH protons, which are all located in the interior of the polypeptide, can often be observed in D_2O. To prevent loss of resonances of labile protons, NMR spectra must often be recorded in H_2O solutions.

In contrast to slowly exchanging protons, the rapidly exchanging surface protons may be obtained in H_2O solutions of proteins for which all labile protons had been replaced by deuterium previously.[30] By observation of the interior labile protons at variable times after the sample preparation, information on the exchange kinetics can be obtained. Individual amide-proton exchange rates can provide fully independent supporting evidence for regular secondary structure.

2.5.3 Additional Data and Limitations

Spin-spin coupling constants (J) characterise through-bond interactions between nuclei linked via a small number of covalent bonds in a chemical structure. For peptide NH protons, $^3J_{HN\alpha}$ is frequently measurable directly from the 1D ^1H spectrum. The digital resolution of the spectrum should be higher than 0.2 Hz/point for accurate measurements of $^3J_{HN\alpha}$. Coupling constants of the highly crowded proteins can be measured with the aid of a phase-sensitive COSY spectrum.[31] The spin-spin coupling constants $^3J_{HN\alpha}$ present supporting evidence for regular secondary structures found in proteins. Individual small ($^3J_{HN\alpha} < 6.0$ Hz) or large ($^3J_{HN\alpha} > 7.0$ Hz) $^3J_{HN\alpha}$ coupling identify helical or β-sheet structures, respectively.

A normal one dimensional NMR experiment with a small molecule often provides information on the spin-spin coupling fine structure of the individual resonances and the chemical shifts in the spectrum. 1D double resonance experiments provide connectivities. Peak separation impose major problems on the assignment of 1D proton NMR spectrum of a macromolecule due to resonance peaks being spread along a line. For large molecules a very large number of measurements would be required for characterisation of the complete network of spin-spin connectivities in a macromolecular structure and is likely to be impossible due to signal overlap. The use of 1D NMR experiments is naturally limited for work with the crowded spectra of macromolecules. Therefore NMR experiments should be expanded to overcome natural limitations of 1D NMR.

2.6 1D and 2D similarities; Background information

2D NMR spectroscopy is an elaboration of the more familiar, one dimensional pulse FT experiment.[8] It is worth emphasising that 1D methods have advantages over 2D methods for certain applications because of the higher speed of data acquisition. Single pulse proton NMR spectra are acquired using the pulse sequence D1-90^0-AQ with a relaxation delay D1 and acquisition time AQ. In 1D NMR experiments, a preparation period is followed immediately by the data acquisition period (Figure 2.4).

In 1D NMR, a short intense pulse at a single radio frequency serves to excite spins from a particular isotope. The response of the sample following the pulse, as manifested by the current generated in the receiver coil (FID), is amplified, detected against the frequency of the transmitter, converted from analogue to digital form, and then stored in computer memory or on a disk as amplitude vs time (t_2). The sequence can be repeated after a suitable delay which allows the spins to return toward equilibrium.

FID's digitised during the acquisition period are combined to improve the signal-to-noise ratio. Signal processing (baseline correction, zero filling, apodization, convolution, or a combination of these) of the averaged time-domain data can be carried out to increase resolution or to increase the signal-to-noise ratio. The oscillatory components of the FID are separated by Fourier transformation to give spectral peaks at their characteristic frequencies.

A 2D NMR experiment is recorded in a 2D time space and the general scheme is shown in Figure 2.4. A 2D NMR experiment differs from a 1D experiment by the addition to the pulse sequence of one or more transmitter pulses and one delay (t_1) that is incremented from one acquisition or combined set of acquisitions to the next. In general terms, the time axis divides into four different periods.

The general scheme for a 2D NMR experiment involves

(1). A preparation period; in which the desired order of coherence is generated.

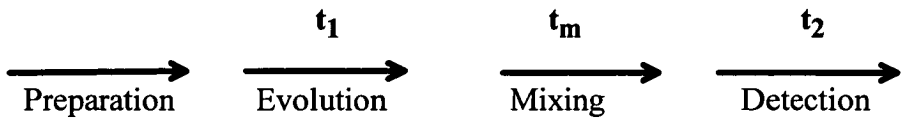
(2). An evolution period; during which the spin system evolves under the influence of the chemical shifts and scalar couplings.

(3). A mixing period; during which there is usually some transfer of coherence.

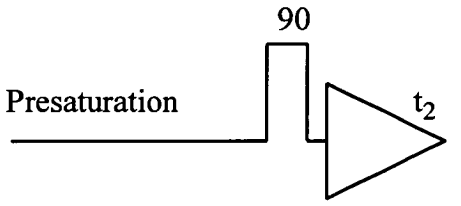
(4). A detection period; where the NMR signal is recorded.

In some experiments mixing takes place during t_1 and therefore evolution period and mixing period cannot be separately distinguished. The influence of the first pulse on the FID recorded during t_2 depends on the length of the t_1 and the second time dimension is generated by repeating the same experiment with t_1 incrementation.

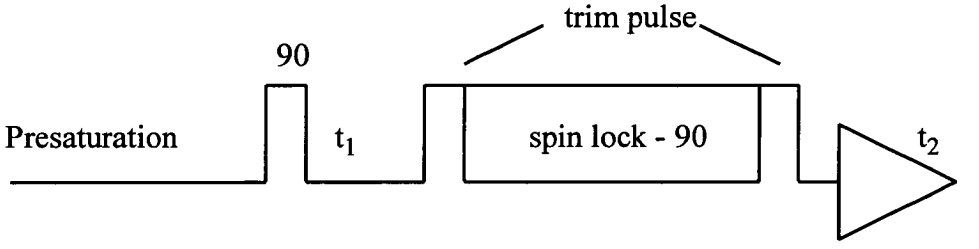
Figure 2.4 : One dimensional and two dimensional pulse sequences



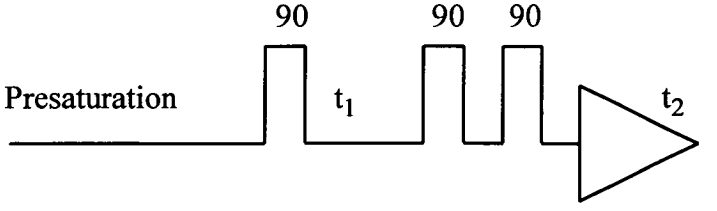
1D NMR experiment



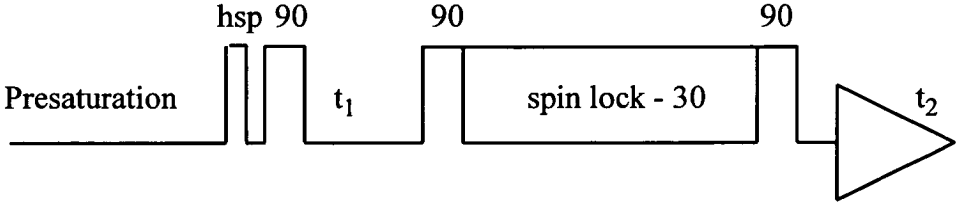
TOCSY experiment



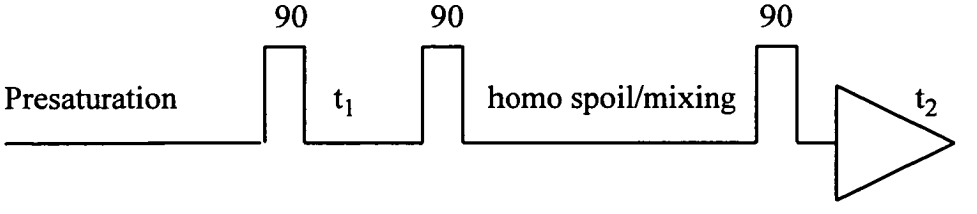
DQF COSY experiment



ROESY experiment



NOESY experiment



As in 1D NMR, the data are digitised as a function of time, t_2. The 2D NMR raw data set consists initially of a series of these FID's stored in different computer files, each having a different evolution time t_1. Hence the data can be viewed as a function of two time variables, $m(t_1,t_2)$. For each block with a given t_1, a FT is carried out with respect to the t_2 time axis. A second FT is carried out with respect to t_1. The result is spectral intensity as a function of two frequencies, $m(F_1,F_2)$. Thus any signal in the 2D spectrum is defined by two frequencies; $m_{ij}(F_{1i},F_{2j})$, m being the magnitude of signal at any data point and F_x being the chemical shift dimension. The additional frequency axis allows the correlation of magnetic properties of one nucleus with those of one or more other nuclei that interact with it during the mixing time. Resolution in the F_2 dimension is determined by the digitisation during the t_2 period; resolution in the F_1 dimension is limited by the number of incremented t_1 values. Quadrature detection normally is used in acquiring in the t_2 dimension to enhance sensitivity. By obtaining pure-phase spectra, one can avoid the degraded resolution characteristic of peaks in mixed phase.

2.7 General Strategies in 2D NMR Studies of Proteins

The 2D NMR techniques are mainly based on the interactions between nuclear dipoles and these may occur either directly through space or be transmitted through bonds. With biological macromolecules, the most important fundamental advantage of 2D NMR relative to 1D NMR is that experiments can in principle delineate all coupling connectivities between protons in a macromolecular structure in a single experiment, and it is thus much more efficient than the use of a large number of 1D spin-decoupling experiments. Comprehensive presentations of 2D NMR spectroscopy can be found in the books[8,32-34] as well as in the variety of review papers.[27,35-38]

General strategies in 2D NMR of proteins; data acquisition, data processing and data analysis can be divided into the following stages.

(1). The sample solution is subjected to a particular 2D NMR pulse sequence and appropriate data are collected.

(2). These raw time-domain data first are subjected to digital signal processing in the time-domain data and then are converted to the frequency domain via FT.

(3). Additional signal processing may be imposed at this point.

(4). The spectral parameters defined by the two frequency axes F_1 and F_2 are extracted. The data to be derived depend on the type of 2D experiment. They include strings of chemical shift values belonging to a particular spin system and parameters associated with cross peaks at particular chemical shifts (coupling matrices, nOe matrices and relaxation rates).

(5). The chemical shift related data is converted to sequence related data with assignments of signals to particular atoms in particular amino acid residues. This stage requires logical analysis of the data with the imposition of knowledge of expected chemical shifts and NOE's spin-spin couplings.

(6). The sequenced related data finally are catalogued and analysed in terms of molecular structure, dynamics, kinetic properties such as hydrogen exchange rates.

(7). Refinement stages compare back-calculated spectra or spectral parameters to original data and adjustments are made to maximise the fit of the structural model to the primary data.

Two dimensional NMR has a wide range of applications utilising homo and hetero nuclear experiments in liquid and solid states. Depending on the problem to be solved, a wide range of 2D experiments can be used. The COSY, DQF COSY, RELAYED-COSY, HOHAHA/TOCSY, NOESY and ROESY experiments are a small selection of homonuclear 2D NMR experiments, which have been of central importance to studies of biological macromolecules. These are discussed briefly below.

2.8 J Correlated Spectroscopy

2.8.1 COSY : 2D Homonuclear Correlated Spectroscopy [25,39,40]

2D homonuclear correlated spectroscopy (COSY) generally is the first 2D experiment to be used in analysing a protein. COSY provides the kind of information available from a single-frequency decoupling experiment (Fig. 2.5) in which spins are scalar coupled to one another. In a COSY plot, the 1D spectrum lies along the diagonal which runs from bottom left to top right, and the off-diagonal elements are present at the intersection of

chemical shifts of groups that are J coupled. The COSY spectra are acquired using the pulse sequence D1-90^0-t_1-90^0-AQ with a relaxation delay D1 and acquisition time AQ.

A normal absolute value COSY spectrum can routinely be used as a fast technique for rapidly gathering information about spin systems. A single COSY spectrum presents a map of the complete spin-spin coupling network in a macromolecular structure. The fingerprint region contains NH/αH cross peaks from the peptide backbone. The degree of resolution of the fingerprint region of a COSY spectrum collected in H_2O is a good indication of the success of sequence-specific assignment to be obtained.

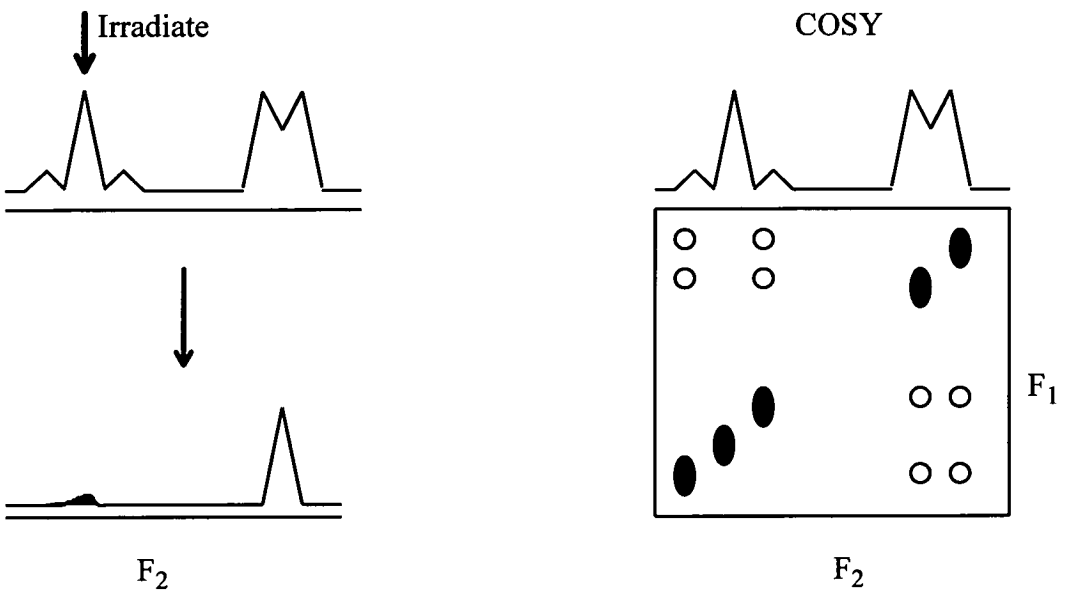


Figure 2.5 : Single-frequency 1D decoupling experiment and the 2D COSY experiment

2.8.2 DQF COSY : 2D Double Quantum Filtered Correlated Spectroscopy [41-43]

The COSY experiment transfers magnetisation between coupling partners. For higher resolution, better detection of cross peaks near the diagonal, and suppression of the solvent signal, a DQF COSY spectrum obtained in the pure-absorption mode is the method of choice. The basic pulse sequence, D1-90^0-t_1-90^0-90^0-AQ, is used and pure-absorption spectra can be obtained by the States-Haberkorn-Ruben method.[44]

Double quantum filtering is used to purge spectra of undesired features and to focus on ^2J and ^3J couplings by selecting coherence transfer between evolution and detection periods for selective removal of magnetisation from non-coupled spins. The double quantum filter reduces the cross peak amplitudes by a factor of 2 due to restriction of coherence transfer and more scans are needed to obtain the same quality spectrum. Despite this sensitivity loss, in practice the DQF COSY experiment usually results in a net improvement in the quality of spectra and can provide information that cannot be obtained in COSY.

The DQF COSY experiment has become popular as a result of its two major advantages over COSY. First, the diagonal peaks in DQF COSY spectra have antiphase multiplet structure instead of in-phase, and thus experience self cancellation to the same extent as do the cross peaks. For larger proteins this results in a significant reduction in the size of the diagonal peaks, which often results in less t_1 noise. Second, the cross peaks in DQF COSY spectra and the major contribution to the diagonal peaks have absorption line shapes in both dimensions. Elimination of the dispersive character of the diagonal peaks allows identification of cross peaks lying immediately adjacent to the diagonal. The "active" coupling, which is actually responsible for the COSY cross peak, always appears in anti-phase in both F_1 and F_2 dimensions. Any other couplings, which are called the "passive" couplings, appear in in-phase in both F_1 and F_2 dimensions.

2.8.3 RELAYED COSY : 2D Relayed Coherence Transfer Spectroscopy [45,46]

The RELAYED COSY pulse sequence is used for identifying pairs of spins that are not coupled directly to one another but that share a mutual coupling partner. Therefore two or several subsequent COSY steps can be performed in a single experiment. The RELAYED COSY pulse sequence, D1-90^0-t_1-τ-180^0-τ-90^0-AQ, where τ is a delay which is adjusted according to the coupling constants of the spin system. Eight step phase cycling plus additional CYCLOPS cycling completes the pulse program. RELAYED COSY data can be obtained in absolute value mode or pure phase mode depending on the phase cycling used.

2.8.4 HOHAHA : 2D Homonuclear Hartmann - Hahn magnetisation transfer sequence or TOCSY: 2D Total Correlated Spectroscopy [47-49]

The 2D homonuclear total correlated experiment (TOCSY) which is developed later is identical to the HOHAHA experiment. The TOCSY spectra are acquired with the D1-90^0-t_1-SL-(MLEV-17)-SL-AQ pulse sequence in which SL denotes a short spin-lock field (2-2.5ms) applied along the x-axis. The MLEV-17 pulse cycle is used for this experiment using low transmitter power. The magnetisation transfer period splits into two parts; the composite pulse part and the two trim pulses at the beginning and the end. The two trim pulses at the beginning and end of the mixing period ensures easy phasing of the 2D spectrum to the absorption mode. Presaturation of the water resonance is used only during the delay time between experiments. TOCSY spectra display both direct and relayed connectivities and are very useful for elucidating scalar-coupled networks. The intensities of peaks depend on the length of the spin-lock mixing time; typically 30-90 ms for proteins. Smaller mixing times are in favour of large couplings such as glycine germinal couplings. This experiment is useful for identifying resonances from the set of spins in a particular ^{1}H spin system. Its advantages are that coherence transfer is efficient and data are largely pure phase.

2.9 Cross-Relaxation 2D Spectroscopy

2.9.1 NOESY : 2D Nuclear Overhauser Enhancement Spectroscopy [50,51]

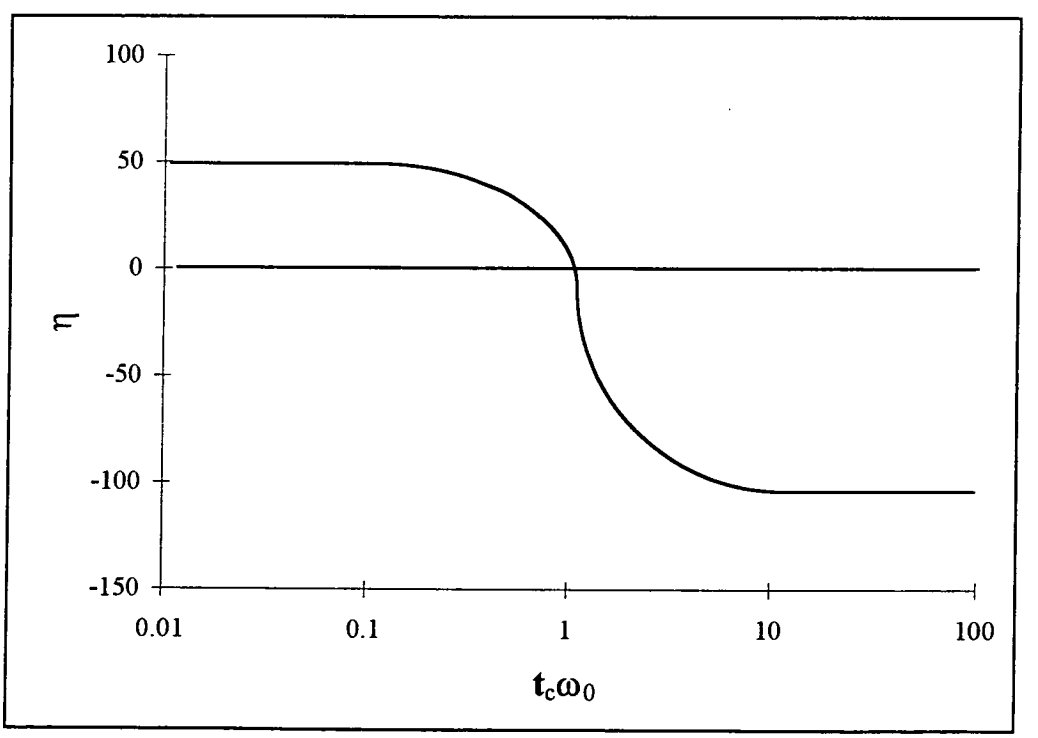
The irradiation and saturation of the resonance of a nucleus ***S*** may result in the intensity changes of the resonance of another nucleus ***I*** if both ***S*** and ***I*** nuclei are sufficiently close. The two nuclei ***S*** and ***I*** do not need to be coupled or connected. The expected result is an alteration of the signal intensity of the I resonance. This is known as the nuclear Overhauser effect or nOe. The nuclear Overhauser enhancement (nOe) is a consequence of modulation of the dipole-dipole coupling between different nuclear spins by the Brownian motion of the molecules in solution. Since nOe intensities are usually small and difficult to detect in a complex spectrum, nOe measurements were hardly used over two decades ago. One dimensional steady state and transient nOe experiments have long been used for studies of small molecules and biopolymers respectively. Due to limitations of the techniques, these experiments cannot be

exploited for work with macromolecules and therefore the 2D transient NOESY experiment will in most instances be the preferred technique. Cross-relaxation is detected as a change in the intensity of one peak as a consequence of perturbing the population of spins involved in another transition. When the two resonances arise from the same nucleus residing in two conformational states of one molecule or two molecular species in dynamic equilibrium, the effect can provide information about the interconversion rates. When the two resonances arise from different nuclei and the cross-relaxation mechanism is via through-space diploe-diploe interaction, the results can provide information about the distance between the two nuclei. The nOe intensity can be related to the distance r between the pair of protons that cross-relax each other.[52]

$$nOe \; \alpha \; {}^{1}/_{r^6} \qquad 2.9.1$$

The nuclear Overhauser enhancement, η, varies with the multiple of rotational correlation time τ_c and nuclear resonance frequency ω_0 (Figure 2.6).

Figure 2.6: Variation of $c\tau_c\omega_o$ vs η (steady state value) for a two spin system.



$\tau_c\omega_0$	<<	1	$\eta_{max} = 0.5$	small molecules (fast tumbling) in non-viscous solvents. MW ca. 800 or less
$\tau_c\omega_0$	ca.	1	$\eta = 0$	medium size molecules (intermediate tumbling) MW ca. 1000 - 2500
$\tau_c\omega_0$	>>	1	$\eta_{max} = -1$	macro molecules (slow tumbling) MW ca. over 3000

According to the $\tau_c\omega_0$ value, three prominent NOE regions, positive, zero and negative can be clearly distinguished. In the positive region steady state NOEs are normally positive and the observed NOEs will depend on the relative positions of the nearby nuclei. At the zero point of Figure 2.6 all enhancements are zero and the useful data can be obtained from the ROESY experiment. The enhancements are always negative in the negative region and no spectrum is obtained in a steady state experiment. Therefore the NOESY is an extremely important experiment for molecules in the negative nOe region.

The 2D versions of these experiments are provided by the three pulse 2D NMR sequence D1-90^0-t_1-90^0-t_m-90^0-AQ, where the delay t_1 is incremented from one group of acquisitions to the next, t_m is the mixing time which controls the time period during which cross-relaxation occurs, and AQ is the spectral acquisition time. Typical values for the mixing time are 30-500ms for proteins. It is necessary to perform experiments for several mixing times, including very short values to obtain all possible information.[53]

Two dimensional exchange spectroscopy can be used to correlate the chemical shifts of protons of two conformational forms, e.g. native and denatured states.[53] The cross peak intensities on either side of the diagonal provide information about the rate of the reaction in each direction. The maximum distance sampled depends on the mixing time. Identification of pairs of nuclei more than 5Å becomes unreliable owing to competing relaxation mechanisms. The NOESY spectrum provides extensive information on intramolecular and intermolecular proton-proton distances and hence on the three dimensional structures of biopolymers.

2.9.2 ROESY : 2D Rotating-frame Overhauser Enhancement Spectroscopy [55,56]

The nuclear Overhauser effect that occurs under spin-lock conditions is known as the transverse or rotating-frame nOe. Cross-relaxation in the laboratory frame has become a very popular method for investigating interatomic distances in macromolecules and, for obtaining structural and conformational information.[52] This experiment was originally proposed by Bothner-By et al.[57] and termed CAMELSPIN, but was subsequently

named ROESY. As with NOESY, the 2D ROESY experiments allows all pairs of spatially proximal spins to be detected simultaneously. While NOESY analysis of the structures of moderate sized peptides is difficult or impractical, e.g., MW = 1000-2500, ROESY offers a variable alternative.

The standard ROESY pulse sequence D1-90^0-t_1-(spin-lock-30)-AQ, where spin-lock-30 is a spin-lock pulse along the x-axis and AQ is the signal acquisition period t_2. Pure absorption spectra can be obtained by the States-Haberkorn-Ruben method.[44]

In ROESY experiments, cross-relaxation takes place perpendicular to the external, static magnetic field and is therefore dependent on spin-spin relaxation process.[58] In contrast, in NOESY experiments cross-relaxation takes place parallel to the external, static magnetic field and is dependent on spin-lattice relaxation process. Cross-relaxation in the laboratory frame has a different dependence on molecular motion than cross-relaxation in the laboratory frame.[59] This results in three major characteristics which make cross-relaxation different.

(1). The cross-relaxation rate in the rotating frame (σ_r) is always positive. In contrast, the laboratory frame cross-relaxation rate (σ_n) becomes zero for rigid body isotropic motion. This means that measurement of cross-relaxation rates in the rotating frame is particularly attractive for moderately sized molecules where $\sigma_n \sim 0$ and the NOESY experiment therefore shows low sensitivity.

(2). For isotropic, rigid body motion in the slow motional regime ($\tau_c\omega_o > 1$), cross-relaxation in the laboratory frame leads to substantial spin diffusion. In contrast, for slow molecular motion cross-relaxation in the rotating frame is tempered by dipolar relaxation thereby attenuating spin diffusion. This means that more accurate measurements of cross-relaxation rates, and hence better descriptions of molecular structure, may be obtainable from measurements in the rotating frame.

(3). The different dependence of σ_n and σ_r on molecular motion makes the measurement of both cross-relaxation rates an attractive method to study molecular motion.

In the ROESY spectrum the cross peaks have the opposite sign to those on the diagonal. This will always be true for ROESY cross-relaxation cross peaks since σ_r is always positive. In contrast, the NOESY cross peaks show the same sign as the diagonal peaks. This indicates that σ_n is negative.

Because σ_r is always positive, cross-relaxation in the rotating frame will give cross peaks with $(-1)^m$,[60] where m is the number of transfers in the cross-relaxation pathway. For example, for a cross-relaxation pathway A→B→C, the AB cross peak will be negative and the AC cross peak will be positive. Because of the strong attenuation of spin diffusion in ROESY experiments on macromolecules, it is unlikely that more than one relay of magnetisation will be observed. Qualitatively this means that any positive cross-relaxation cross peaks in ROESY must arise from spin diffusion and that all negative cross peaks are likely to be due to direct, one step cross-relaxation.

In summary, for small molecules, for rapid molecular tumbling, a steady-state NOE measurement will generally be the most effective experiment. For medium sized molecules, ROESY is much more sensitive than NOESY. For large macromolecules, detection of cross-relaxation to a group of magnetically equivalent spins will in general be more efficient with NOESY than with ROESY. The DQF COSY, TOCSY, NOESY and ROESY connectivity patterns are given in Figure 2.7 and the experimental schemes are given in Figure 2.4.

2.10 Sequential Assignment Strategies; Homonuclear Proton Approach

Methods for achieving sequential assignments in proteins can be divided into three categories based on the NMR parameter that provides transpeptide information.

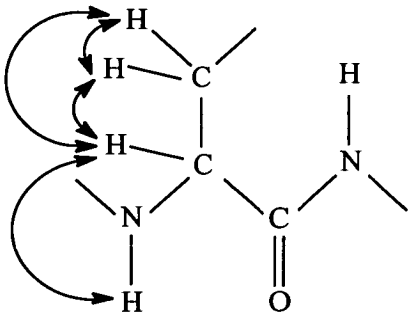
(1). Detection of short transpeptide contacts by 2D NOE spectroscopy; $\alpha_iH/N_{i+1}H$, $\beta_iH/N_{i+1}H$ and $N_iH/N_{i+1}H$.

(2). Detection of one-bond heteronuclear coupling

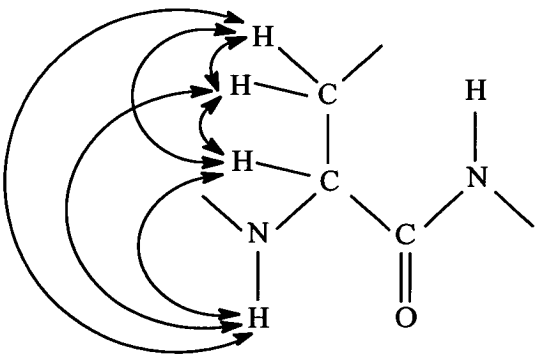
(3). Detection of multiple-bond heteronuclear coupling

Figure 2.7 : DQF COSY, TOCSY and NOESY/ROESY Connectivity Patterns

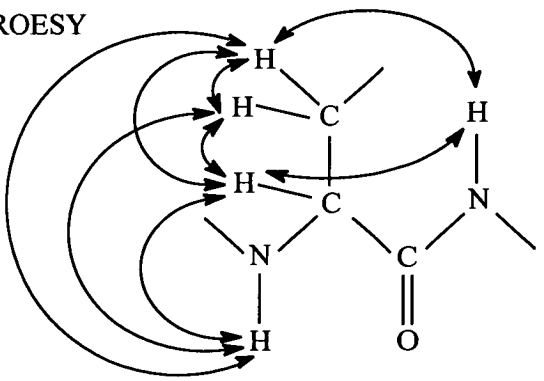
DQF COSY

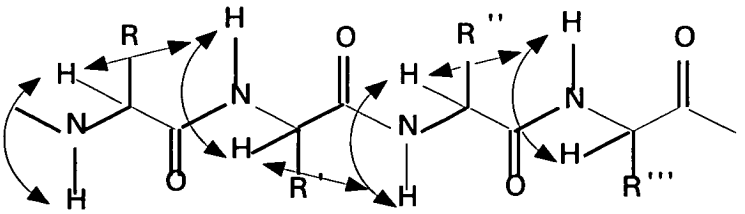


TOCSY



NOESY/ROESY





DQF COSY or TOCSY connectivity patterns and

NOESY or ROESY connectivity patterns along the polypeptide backbone

The overwhelming majority of sequence-specific assignments obtained to date have been achieved by detection of NOE connectivities between the NH of residue i and the adjacent NH of residue i+1. In most cases, one or more of the interresidue distances ($\alpha_iH/N_{i+1}H$, $\beta_iH/N_{i+1}H$ and $N_iH/N_{i+1}H$) can be inferred from NOESY cross peaks. Information about the amino acid sequence and the spin system types to which these resonances belong (COSY, DQF COSY and TOCSY data) is used to distinguish nearest neighbour NOE connectivities from longer range ones.

Starting points for sequential assignments are provided by unique single residues or unique di- and tripeptides whose identities can be deduced from spin system analysis plus interresidue NOE connectivities. Once starting points have been identified, they can be extended by combining COSY/TOCSY and NOESY data.

Protein chemists usually divide the 20 common amino acid building blocks of polypeptides into 8 groups such as neutral (A, G, I, L and V), basic (H, K and R), acidic (D and E), hydroxy (S and T), sulphur containing (C and M) and aromatic (F, W and Y) amino acids, amides (Q and N) and imino acids (P). For each amino acid residue, the non-labile protons constitute one or several spin systems which are connected by scalar (through-bond) spin-spin couplings J.

Several amino acids have the same categories of spin system. For example Asn, Asp, Cys, His, Phe, Ser, Trp and Tyr all show an AMX spin system and Glu, Gln and Met show an AM(PT)X spin system. Further complications arise from the same amino acid residues occurring more than once. The spin systems of the nonlabile hydrogen atoms (Figures 2.8 and 2.9) and the chemical shifts for the 20 common amino acid residues in "random coil" polypeptide provide a reference[61] for a qualitative understanding of the general features of a polypeptide spectrum.

The recognition of particular geometric patterns in the TOCSY spectrum are characteristic of the different spin systems encountered in the 20 common amino acid residues. For all amino acid residues, the completion of the spin system by delineation of the NH/αH connectivity in the H_2O solution of the polypeptide is a very important part of resonance identifications.

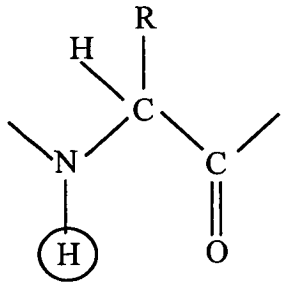


Figure 2.8
Amino acid residue.
R is the side chain.
The circle identifies the labile amide proton.

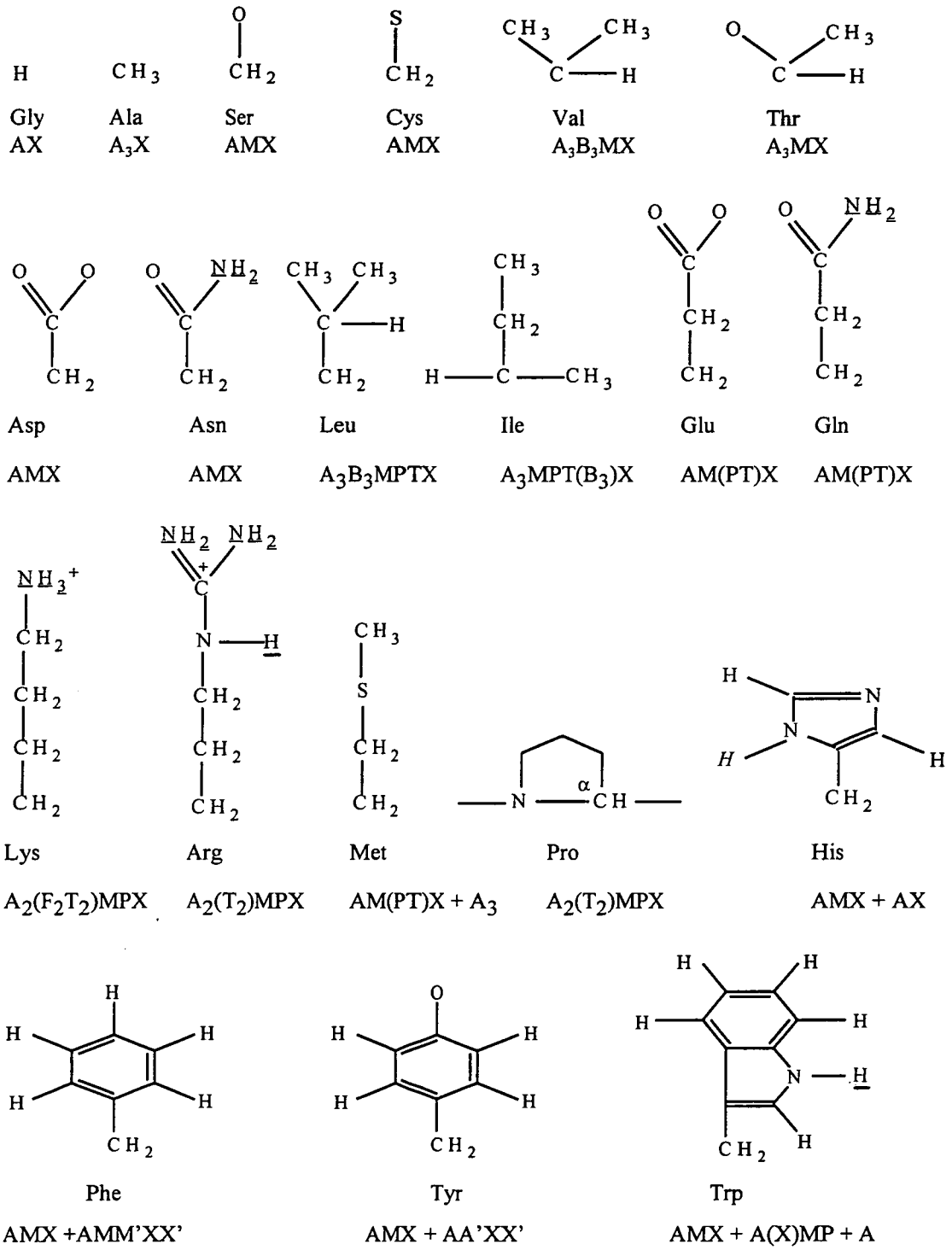
In addition, the TOCSY experiment can in principle delineate almost all spin-spin coupling connectivities between protons in a macromolecular structure and it is thus much more efficient than the use of 1D decoupling experiments.

The fingerprint part (ca 7-10 ppm) of the DQF COSY spectrum of a polypeptide is a most useful guide for unambiguously distinguishing NH/αH cross peaks from NH/βH cross peaks of AMX spin systems which lie in the same region of the TOCSY spectrum. In the spectrum analysis, the ^{2}J and ^{3}J coupled resonances all of which arise from the same amino acid residue can be identified. In situations of degeneracy or near degeneracy of two resonances, the DQF COSY spectrum may be helpful, since there is no interference with the diagonal. Using the DQF COSY and the TOCSY spectra, essentially complete identification of the NH/αH COSY cross peaks can be obtained prior to the sequential assignments.[62,63] The 20 common amino acid residues give rise to 10 different COSY connectivity patterns for aliphatic protons and 4 patterns for the aromatic ring protons (Figure 2.10).

The fingerprint part of the NOESY connectivities can be used for obtaining the NH/αH sequential assignments. In addition to the sequential connectivities, through space connectivities of almost all the amino acids could be obtained from the NOESY spectrum. In addition to the NOESY spectrum, NH/βH and NH/NH etc. could also be obtained from the ROESY spectrum.

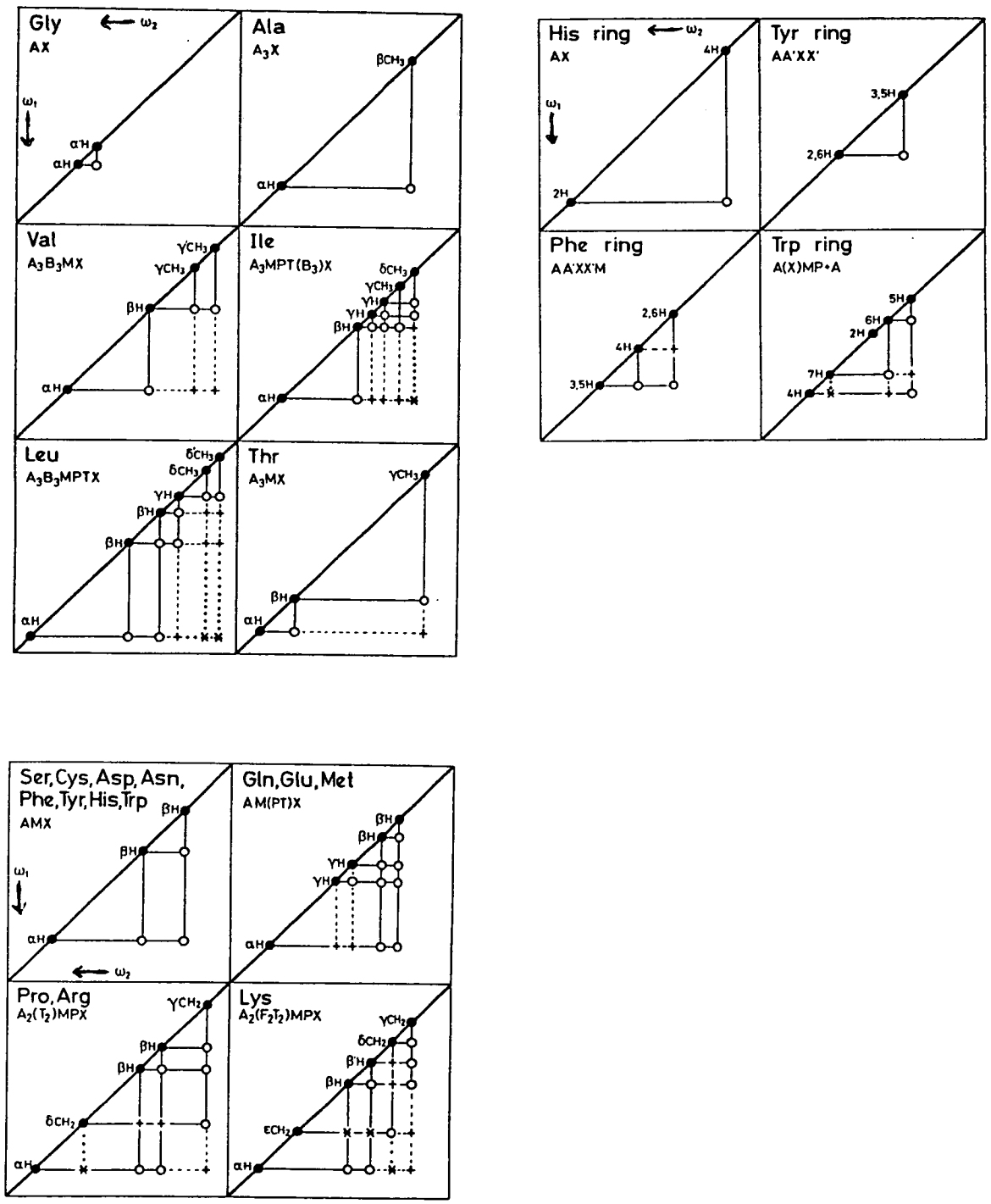
In a sequential assignment procedure, connectivities between the ^{1}H spin systems of sequentially neighbouring monomeric units can be established by the NOESY spectrum. The identification that is missing in incomplete spin system identifications can be obtained in connection with the sequence-specific assignments. Different amino acid types with the same spin-spin patterns can be identified once their sequence positions are determined.

Figure 2.9 : Side chains R (See Figure 2.8) and three-letter symbols for the 20 common amino acids and the spin systems of the non-labile hydrogen atoms in the molecular fragments H-$^{\alpha}$C-R.



Underlining indicates the labile protons that can under certain conditions be observed by NMR in aqueous solution and italics indicate the labile protons which are not usually observed by NMR

Figure 2.10 : COSY, relayed-COSY and double-relayed-COSY connectivities for the spin systems of non-labile protons in the common amino acid residues [10] (Figure 2.9).



COSY connectivities :open circles and solid lines

relayed-COSY connectivities :crosses and broken lines

double-relayed-COSY connectivities :stars and dotted lines

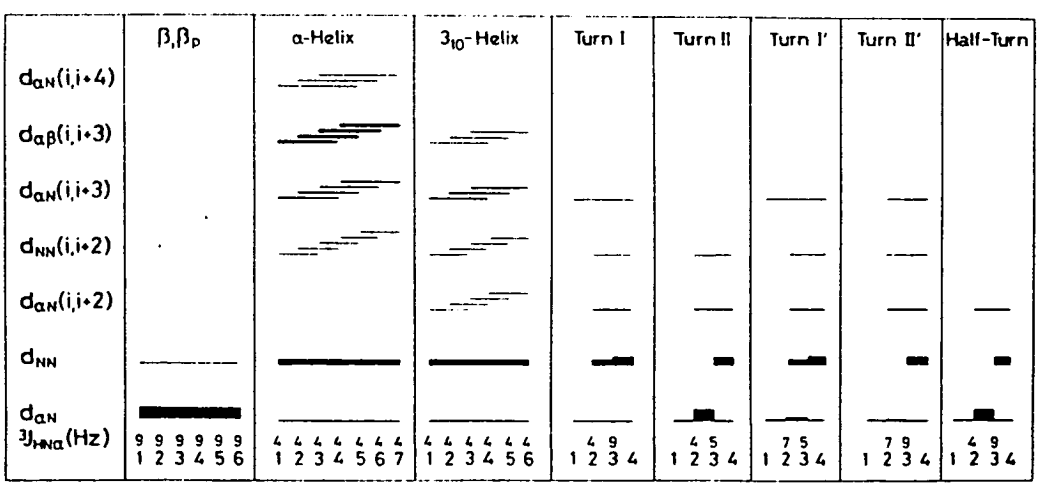
2.11 Determination of Secondary Structure.

Once extensive sequence specific ^{1}H NMR peak assignments have been obtained, NOESY cross peaks that represent longer range proton-proton connectivities can be catalogued. Patterns of these NOE's are used in identifying β and γ-turns, β-sheets and helical regions (Table 2.1).[64]

The α-helix is characterised by close approach between residues i/i+3 and i/i+4. The $\alpha_iH/N_{i+3}H$ NOESY cross peaks are stronger than $\alpha_iH/N_{i+4}H$ NOESY cross peaks. In the 3_{10}-helix, short distances prevail between residues i/i+3 and i/i+2 and the NOESY cross peak between $\alpha_iH/N_{i+2}H$ is weaker than $\alpha_iH/N_{i+3}H$. The prevalence on an α-helix may further be obtained from observation of rather strong $\alpha_iH/\beta_{i+3}H$ NOESY cross peaks.

The ^{3}J coupling constant values ($^{3}J_{NH\alpha}$) are higher than 7Hz for β-sheets. The long range NOE's between α_iH/α_jH have a pivotal role in antiparallel β-sheets. More details of identification of secondary conformations have been discussed elsewhere.

Table 2.1: Survey of the sequential and medium range ^{1}H-^{1}H NOE's and ($^{3}J_{NH\alpha}$) spin-spin coupling constants [10] in secondary structures.



CHAPTER 3

3D STRUCTURE CALCULATION OF PROTEINS

3.1 Introduction

NMR spectroscopy is currently the only method for obtaining detailed information about the spatial arrangement of proteins in solution. During the last decade, there has been a dramatic increase in the use of NMR to study protein structures in solution.[65] Computational tools have been developed in the last few years, which facilitate direct determination of protein structures from NMR data. Numerical calculations with simulated and experimental NMR constraints for distances and torsional angles show that data sets obtainable with current NMR techniques carry sufficient information to determine the global fold of a small protein.

The overall geometry of a protein molecule is partially defined by the primary structure and the basic structural parameters such as bond lengths and bond angles. In addition to these distances and angles, additional distance and angle constraints are used to define the final tertiary structure of the molecule. Structure calculation of biological macromolecules from constraints upon the possible values of their interatomic distances, together with additional constraints have become an important tool in structural biomolecular chemistry.[66-69] These calculations have successfully been used to solve problems in drug design,[70,71] structural interpretation of NMR data,[10] and protein structure predictions.[72] The first determination of the structure of a complete protein, the Bull Seminal Inhibitor, was made by Williamson et al.[73] using the DISGEO distance geometry program described by Havel.[74]

Different methods for the determination of protein structures from NMR data exist and new algorithms and software are still being developed. Among them distance geometry (DG), molecular dynamics (MD), restrained molecular dynamics (rMD) and dynamical simulated annealing (DSA) are commonly used by molecular modellers according to their requirement. Currently mathematical models such as DG which deal with atomic coordinates are commonly used for deriving protein structures from NMR data.

3.2 Use of NOESY in Structure Calculations

The use of 2D NOESY in the structural analysis of macromolecules has become widespread during last decades.[65] The data for a protein structure determination are collected using one-, two-, or three-dimensional NMR techniques.[8,10] Typically, the NOESY spectra of proteins contain hundreds of cross-peaks arising from the three dimensional fold of the polypeptide chain. Each NOESY cross peak shows that two protons in known locations along the polypeptide chain are separated by a distance of less than approximately 5Å in the three dimensional protein structure. For the structural interpretation of NMR data, mathematical techniques are available to identify those three-dimensional arrangements of the linear polypeptide chain which satisfy all the experimental constraints. Each individual structure calculation may find a protein conformation that is compatible with all the experimental constraints collected from the NOESY data and possibly supplementary experiments.

To evaluate the effects of spin diffusion, the profile of the NOE build-up is usually investigated. The build-up rates of the nuclear Overhauser effect can be measured by acquiring NOESY spectra with different mixing times. The prominent substantially well resolved short- and long-range inter-residue cross peaks such as $\alpha_iH/N_{i+1}H$, $N_iH/N_{i+1}H$, $\alpha_iH/N_{i+3}H$, α_iH/α_jH and $\alpha_iH/\beta_{i+3}H$ are usually studied for this purposes. The build-up rates of first-order Overhauser effects which arise from direct dipole-dipole coupling between closely spaced protons and the second-order Overhauser effects due to spin diffusion by cross-relaxation between more distant protons can be discriminated to carry out further analysis of the NOESY data.

To investigate further whether a given set of NMR based input data defines a unique conformation, a family of structures is obtained by repeating the calculation with the same NMR data but with different, randomly generated starting conditions. In NMR, the data which are used for molecular modelling come mainly from the distances between atoms in structure. Since each distance constraint describes an allowed distance range rather than a precise value for the distance,[10] the individual structures are similar but not identical. The result of a structure determination from NMR data is therefore commonly represented by a group of conformers, each of which represents a

solution to the geometric problem of fitting the polypeptide chain to the ensemble of all experimental constraints.

When assignment of signals to protons has been achieved by a combination of NMR techniques, NOEs between proton pairs can be identified and translated into the proton-proton distance constraints. There are three common methods of translating NOE cross peak intensities into distances found in the literature.

Method 1: The most important quantity derived from NOE cross peaks is the cross-relaxation rate σ_{ij} between protons i and j. The cross-relaxation rates can be measured from build-up rates of cross peaks in 2D NOESY spectra at several mixing times. Assuming a rigid protein and no indirect magnetisation transfer, the build-up rates or NOE intensities are calibrated empirically by comparing values of unknown distances to values of protons of known distances (r_{cal}) in the protein.

$$r_{ij} = r_{cal} \left[\frac{\sigma_{cal}}{\sigma_{ij}} \right]^{1/6} \qquad \text{.....3.1}$$

This approach, also known as two-spin approximation, is used to translate NOE intensities or build-up rates into a set of approximate distance constraints. This translation does not indicate the distribution of NOE distances from different conformations at the given NMR time scale. The derived set of distance constraints, therefore, does not necessarily represent the average structure, and there may be no single conformation that is consistent with the data set. The weakest N_iH/α_iH-type NOEs (d_{max} = 2.9Å), the strongest $\alpha_iH/N_{i+1}H$-type NOEs (d_{min} = 2.2Å) or $\beta H/\beta H$-type NOEs (d = 1.75Å) are normally used as reference points for calibrating.

Method 2: Another approach for calibration NOEs to analyse the cross peak integrals in a histogram. Two integral thresholds are defined for designating three distance ranges such as "strong" (1.8 - 2.49), "medium" (2.5 - 3.29) and "weak" (3.3 - 5.0), thus allowing more restrictive bounds. If the maximum NOE distance seen is near 5Å, about 10% of the NOEs will be in the strong range, 20% will be in the medium range and the

rest will be in the weak range. These choices can lead to better defined structures, since they contain more information. Too restrictive a set of bounds would result in structures that explore an incorrect part of conformational space.

Method 3: Because NOE cross peak intensities cannot be translated into precise distances owing to correlated motions, anisotropic motions, flexibility and spin diffusion, two major limits can be introduced; upper limit and lower limit. In this method, NOE cross peak intensities are categorised into either three or four approximate upper limit constraints; strong (2.5-2.8Å), medium (3.3-3.6Å), weak (4.0Å) and/or very weak (5.0Å). Translating NOEs into reliable lower limit constraints is difficult and it is preferable to take the sum of van der Waals radii as a lower limit (1.8Å) to the distance.[75,76]

Two or three intensity thresholds are defined to designate strong, medium, weak and very weak intensity cross peaks and three or four distance range limits are specified. Thresholds are determined by comparing known distances to experiment NOE integral values. The inter-residue $\alpha_i H/N_{i+1}$ distances {(2.3Å in β-sheets and 3.6Å in α-helices)} and $N_i H/N_{i+1}$ distances {(4.3Å in β-sheets and 2.8Å in α-helices)} are generally used as reference cross peaks for obtaining threshold values. This approach works well for obtaining a crude structure and it is less sensitive to errors such as spin diffusion and/or local correlation times. This method is commonly used for generating constraints to obtain starting structures for further calculations.

3.2.1 Additional dihedral angle and distance terms

In order to optimise a structure simultaneously with respect to the atomic interaction function and the experimental data set, additional terms can be added which represent the atom-atom distance and torsional angle constraints. The use of vicinal proton-proton couplings is another source of useful geometric information for studies of molecular modelling. Coupling parameters of vicinal protons (NH/αH αH/βH in proteins) can be obtained from 1D proton and DQF COSY experiments. The dependence of the vicinal coupling constant between two protons H_a and H_b on the dihedral angle is given by a Karplus type equation.[77]

$$^{3}J_{NH\alpha} = A \cos^{2}\theta + B \cos\theta + C \quad \ldots\ldots\ldots\ldots \quad 3.2$$

where J is the coupling constant and θ is the intervening dihedral angle between two atoms ($\theta = |\psi - 60^{0}|$). The parameters A(6.4), B(-1.4), and C(1.9) for the vicinal coupling constant $^{3}J_{HN\alpha}$ for the polypeptides have been determined by Pardi et al..[78]

The formation of hydrogen bonds between CO and amide NH can be seen in the common polypeptide secondary structures; α-helix (CO_i - NH_{i+4}), 3_{10}-helix (CO_i - NH_{i+3}) also in β-sheets. The individual NH proton exchange rates can provide valuable information in regular secondary structures. The amide protons which are hydrogen bonded show slow exchanging process but exposed amide protons exchange rapidly. This information can be added as additional constraints in structure calculations. Information from loop regions such as disulphide bridges and cyclic peptide rings can also be included as constraints.

3.2.2 Constraints

There are two common ways to express constraints. The first method assigns a target value for the parameter of interest and the objective function measures the deviations from the optimum value. The second method for handling constraints is that the upper and lower bounds on a particular parameter can be specified. A penalty is added when the boundary conditions are violated.

Most of the distance constraints that are available for use in calculations are obtained from the nuclear Overhauser experiments. In addition to the distances derived from the NOE measurements, additional structural information can be obtained from many types of experiments such as coupling constants, hydrogen bond information, NH exchange rates, distance from spin labels, chiral information, specific distances from fluorescence energy transfer experiments and chemical cross-linking.

3.2.3 Atom definitions

The basic unit of the force field is the atom which is considered as a spherical and non-polarizable "point" mass with no directional properties and no internal degrees of

freedom. There are two common types of atom representations; an 'all-atom' representation where every atom is considered explicitly and an 'united-atom' representation where all hydrogen atoms are collapsed into their heavy atom. The advantage of the united-atom representation is that the number of atoms in the system is reduced and the larger systems can easily be modelled. As an united-atom is always spherical and the direction of the hydrogen is not available, a loss of steric effects and inability to represent directional hydrogen bonds are disadvantages.

Most of the algorithms used in structure calculations only deal with distances between points. These points can be real atoms or united-atoms. In NMR studies, atom representations are of real atoms and pseudoatoms. The real atoms represent all the heavy atoms including those hydrogens that can be resolved. A pesudoatom is placed where no real atom exists. Unresolved resonances associated with methylene groups, the αH of glycines, methyl groups, NH_2 groups and various ring hydrogens, can be represented as pseudoatoms.

The pseudoatom is positioned at the geometric centre of the unresolved group of resonances and the distance constraint can be assigned to the pseudoatom. In these cases boundary conditions are loosened to permit the pseudoatom to meet the constraints properly. The pseudoatom is treated as a point and has no van der Waals radius. In order to compensate, the van der Waals radius is given to the pseudoatom.

3.3 Strategies for Conformational Search

It is important to realise that it is not always possible to construct a single algorithm that can simultaneously enforce all these types of constraints. Most of the programs available at present can easily deal only with distances, angle, and chiral constraints. The programs DISGEO,[74,79] DGEOM,[80] DSPACE,[81] PROTEAN,[82-84] and the DG-II program[68] use atomic coordinates, while the DISMAN,[85] DIANA[86] and the FANTOM[87] programs use angular coordinates as independent variables. Basic requirements of these programs are

(1). Data input procedures are needed

(2). A way of generating an initial structure or set of structures must be available

(3). An optimisation routine is needed to bring the constraints and the structures into closer harmony

Data input procedures normally allow descriptions of geometry taken from electron diffraction data, X-ray data, NMR data or other structural measurements. The choices for generating starting structures include random starting conformations, model-built starting conformations, or starting conformations calculated from another program such as distance geometry. The DISMAN, DGEOM, DISGEO and DSPACE programs randomly generate initial structures. Modifications of molecular dynamics routines[88,89] or more complex mathematical search procedures[82,85] begin with model-built structures. Thus the constraints can be applied directly. Currently available minimisation methods can be used to optimise the penalty function of the constraints.

3.4 The Force Field

The combination of all potential energy functions is the force field. The potential energy can directly be used to determine the relative stabilities of the different possible structures of the system. The mechanical forces acting on the atoms of the systems can be used to calculate dynamic properties of the system by solving Newton's equations of motion.[90,91] In the force field, the molecule is viewed as a collection of points (atoms) connected by springs (bonds) with different elasticities (force constants). The force holding the atoms together can be described by potential energy functions of structural features like bond lengths, bond angles and non-bonded interactions. In addition to these, steric, electrostatic and other strain forces must also be included. The force field can be modified in order to constrain interproton distances when a NOE between them has been measured by NMR.

The energy (E) of the molecule in the force field arises from deviations of ideal structural features, and can be approximated by a sum of energy contributions. The set of potential functions contains adjustable parameters that can be optimised to obtain the best fit of calculated and experimental properties of the molecule such as conformational energies, geometries and other thermodynamic properties. The force field equation (Equation 3.3) given below which is fully described by Clark et al.[92] is

only be valid for small deformations from the natural structural features such as lengths and angles.

$$E = \sum \{E_{str} + E_{bend} + E_{oop} + E_{tors} + E_{vdw}$$
$$[+ E_{ele} + E_{dist\_c} + E_{ang\_c} + E_{tor\_c} + E_{range\_c} + E_x]_{\text{optional energy terms}} \} \quad .....3.3$$

E_{str}	energy of a bond stretched from its natural bond length
E_{bend}	energy of bending bond angles from their natural values
E_{oop}	energy of bending planar atoms out of plane
E_{tors}	energy of torsional due to twisting about bonds
E_{vdw}	energy due to van der Waals non-bonded interactions
E_{ele}	energy due to electrostatic interactions
E_{dist_c}	energy associated with distance constraints
E_{ang_c}	energy associated with angle constraints
E_{tor_c}	energy associated with torsion angle constraints
E_{range_c}	energy associated with range constraints
E_x	energy associated with other possible constraints

Numerous force fields have been developed for simulation and modelling of polypeptides and nucleic acids. Among them AMBER,[93] CHARMM,[94] GROMOS[95] and Tripos[92] force fields are commonly used.

3.5 Distance Geometry (DG)

The DG calculations have been playing an important role in the initial stages of the overall structure determination process. The name distance geometry has come to refer to the computer programs that convert geometric constraints into molecular coordinates. The DG program includes random start approach. The mathematical algorithm guarantees that the initial coordinates are the best multidimensional fit to the trial distances. The randomisation is associated with the trial distances, which are chosen as randomly distributed between upper and lower boundary conditions. When additional constraints are available, DG attempts to produce structures that meet both the basic constraints described above and these additional constraints.

The distance geometry approach can be used to search conformational space subject to a wide variety of constraints beyond those obtainable by NMR. The final target of the distance geometry calculation is to produce one or more molecular structures that meet a set of constraints. As a general rule, DG calculations do not yield a unique structure. Two different methods are used to convert distance constraints; the metric matrix approach[66,74,79] which operates in distance space and the variable target function approach[85,86] which operates in real space. These are called DG algorithms but their approach is fundamentally different.

3.5.1 Metric Matrix approach[66,74,79]

The metric matrix approach consists of the following steps. First, upper and lower distance-bound matrices are set up for all atom distances in the molecule. The NMR information is complemented with distance information following from allowed van der Waals atom-atom distances, standard bond lengths, and bond angles. For pairs of atoms separated by one rotable bond, upper and lower bounds can be calculated from the corresponding trans and cis configurations.

The DG algorithm then proceeds with an iterative bound smoothing procedure using triangle inequalities. This is achieved by increasing the lower bounds and reducing the upper bounds in those cases where the bounds are not geometrically self consistent. Triangle smoothing treats all atom triples to derive narrower bounds when possible. This process can derive distance information between atoms where no constraint has been entered. Distance matrices are then randomly selected between the upper and lower bound matrices and subsequently embedded in 3D coordinate space. Only bounds matrices that have been smoothed can be embedded with reasonable results.

As the distances, however, do not define the chirality of the structures because the embedding procedure is dependent only on distances in matric matrix method, mirror images (local or global) of the correct structure can occur. These can be rejected as the chirality of single amino acids (L) and helices (right handed) is known.

Because, in the embedded structures, many distances usually fall outside their bounds, optimisation is needed to minimise the remaining atom-atom distance violations and to impose the known chirality to the various asymmetric centres in the protein molecule. Therefore the chiral constraints are used only during optimisation. The metric metrix distance geometry has been implemented in several programs such as DISGEO[74,79] and DGEOM.[80]

3.5.2 Variable Target Function method[85,86]

This method involves the minimisation of a distance constraint error function (target function) in variable dihedral angle space. The target function (T) is similar to the error function in matric matrix approach. During the minimisation of T, only χ, ψ and ϕ variable dihedral angles are used as independent values, instead of Cartesian coordinates. Bond lengths, bond angles and ω dihedral angles are kept at fixed values. The minimisation of the target function is performed at different levels to reach the global minimum T=0 of the target function as closely as possible.

This procedure normally starts with a number of different initial conformations obtained by taking random values for the dihedral angles. From conformations obtained after a series of low-level optimisations the best conformations are selected for higher level optimisations. The quality of conformations can be improved by repeating a series of function optimisations. This approach is implemented in the DISMAN[85] and DIANA[86] programs.

Independent methods (such as metric matrix approach and the variable target function approach) for deducing spatial structures may sometimes obtain compatible families of solutions but it does not necessarily indicate that these methods are sufficient, or the families of possible solutions even contain the correct ones.

Although all DG methods have been used on macromolecules containing a large number of atoms, they all suffer from one or more shortcomings. DG algorithms are useful techniques to derive crude, widely diverse tertiary structures consistent with the experimental data. Ideally a structure obtained by a distance geometry calculation

should have no violations of the distance constraints imposed by the experimental data. In practice, finding these structures involves several optimisations, error calculations and therefore each solution includes a certain number of small distance violations.

The structures generated by DG calculations can be improved by combining DG with the refinement method. The success of a combination of DG with other refinement methods is dependent on the accuracy of the experimental data, the quality of the initial structures generated by the DG algorithm, and the accuracy of the force field and the simulation adopted.

3.6 Molecular Dynamics (MD)

The molecular dynamics simulation provides information about the dynamic behaviour of a system. The presence of motional freedom in the model simulation implies the possibility of passing over energy barriers, and a range of protein conformations is sampled. Therefore MD searches a larger part of conformational space and generally finds a lower energy minimum than regular energy minimisation techniques. By numerically integrating the Newtonian equations of motion with a sufficiently small time step during which the forces can be regarded as constants, a trajectory (atomic positions as a function of time), can be obtained.

To begin the dynamic simulation, an initial set of atomic coordinates and velocities is needed. Initial velocities for all atoms are taken either from a random distribution or from a previous distribution corresponding to the desired temperature. During the simulation, the temperature is held constant by scaling the velocities after each time step. Newtonian equations of motion are solved by integration over very small time steps in which the forces can be regarded as constants.

The method used in the constant temperature simulations incorporates a damping function to avoid systematic oscillations and rapid changes. The damping function which is achieved by coupling the system using an adjustable time constant, avoids sharp changes in the temperature when heating and cooling. This eliminates artefacts that may arise during simulations. Simulations under constant temperature conditions

correspond to operating in the NVT (N - constant number of particles, V - constant volume and T - constant temperature) or canonical ensemble.[97] In this type of simulation, the velocities of the atoms are scaled at each step so that the kinetic energy of the system corresponds to a set temperature.

The major use of molecular dynamics in protein NMR is conformational sampling and it has been a great interest in conformations of lowest energy. The obvious way to sample more conformations is to raise the temperature of the simulation. This type of dynamics simulation which is called Molecular Quenched Dynamics, is performed continuously at a high temperature and snapshots are taken for subsequent minimisation.

When the MD simulations are carried out without including the effects of solvent, this may cause amino acid side chains at the surface to collapse onto the core of the protein, thus biasing the sampling behaviour of the protocol. Other disadvantages in vacuo simulation are details on hydrogen bond formation and breaking and amplitudes of motions involving solvent accessible parts of the molecule. This problem is felt with MD simulations of solvent accessible molecules such as small peptides.

Molecular dynamics is not capable of generating initial structures, but it can simulate the dynamic behaviour of molecules while including the average character of the NMR data. In most cases poorly defined parts of the molecules were seen to adopt a wider range of conformations after the MD refinement. There have been several excellent surveys of applications of MD to the study of protein structure[98-100] and several reviews on conformational searching.[101,102]

3.7 Restrained Molecular Dynamics (rMD)[88,89,103]

Restrained molecular dynamics has been shown to be an additional valuable tool in elucidating the molecular conformations compatible with NMR data. Existing computer programs of MD calculations have been modified to allow inclusion of the NMR data. This is done by adding a pseudo pair potential energy term to the potential function used in the free dynamics simulation. These additional harmonic pseudo forces act like strings between those atom pairs constrained by the NMR data and drive the molecular

conformations towards conformations compatible with the NMR data. The energy represented by a distance penalty will then be lost to the surroundings as the dynamics simulation moves the atoms towards an energy minimum.

Restrained molecular dynamics calculations of protein conformations using NMR data have been done with two different aims. One aim was the refinement of a model built structure that crudely satisfies the NOE distance constraints and the other aim was to decrease both the potential energy and the atom pair pseudo energy arising from the NOE distance constraints. Restrained molecular dynamics should be used only to search for low energy conformations within the allowed parts of conformational space. Solvent molecules can also be included for rMD calculations. Refinement by rMD improves the quality of the structures in terms of energy. One disadvantage of rMD is that structural information obtained from NMR experiments is mixed with the force field parameters used.

3.8 Dynamical Simulated Annealing (DSA)[104-106]

A common use of the constant temperature simulation is to explore the conformational space available at the given temperature. This is done by simulating the motions at a very high temperature. All conformations are then energetically accessible. The final step is lowering the temperature. In simulated annealing calculations, an integration algorithm is used to solve Newton's equations of motion in an analogous fashion to that used in MD. This method which is known as Simulated Annealing has been used in other fields as a technique for global optimisation of multi dimensional surfaces.[107,108] It has recently been introduced to molecular modelling of proteins using NMR data.[106]

In DSA, geometric restraints and the non bonded interactions are represented by a simple repulsion term. The non bonded interactions replace the dihedral, van der Waals, electrostatic and hydrogen bonding potentials of the empirical energy function in conventional molecular dynamics. NOE distance constraint data can be included in the energy potential to provide a driving force term in the dynamics simulation. A penalty function can be included to satisfy the violated distance constraints. In addition, dihedral angle constraints and hydrogen bond restraints can also be added in a similar manner.

The method describe in the DSA circumvents the folding problem, by starting from a completely random array of atoms and introducing the force constants for the covalent, interproton distance, torsion angle, and repulsion van der Waals terms in the target function appropriately. The system is simulated at high temperature by solving Newton's equations of motion. As the values of all force constants are very low during the early stages of the simulation, energy barriers between different folds of the protein can be overcome, and the global minimum of the target function is really located. Because the atoms are initially only weakly coupled, they can move independently to satisfy the applied restraints, thereby avoiding problems associated with folding.

The total target function $\mathbf{F}_{total}$ for which the global minimum region is searched comprises the following terms. $\mathbf{F}_{total}$ represents the effective potential energy in the dynamics calculation which involves the integration of Newton's equations of motion.

$$\mathbf{F}_{total} = \mathbf{F}_{covalent} + \mathbf{F}_{repel} + \mathbf{F}_{NOE} + \mathbf{F}_{tor} \quad \text{..... 3.4}$$

The $\mathbf{F}_{covalent}$ term is employed to drive the system towards the correct covalent geometry such as bond lengths, bond angles, planes and chirality.

$$\mathbf{F}_{covalent} = \sum_{bonds} k_b (r-r_0)^2 + \sum_{angles} k_\theta (\theta-\theta_0)^2 + \sum_{impropers} k_\phi (\phi-\phi_0)^2 + \sum_{\omega} k_\omega (\omega-\omega_0)^2 \quad \text{..... 3.5}$$

The force constants of the energy terms are described as bonds (k_b), angles (k_θ), improper torsions (k_ϕ) which define planarity and peptide bond dihedral angles (k_ω). The correct bond lengths, angles, improper angles and peptide dihedral angles are defined as r_0, θ_0, ϕ_0, and ω_0 respectively.

The non bonded van der Waals interactions are represented by $\mathbf{F}_{repel}$ which employs a purely repulsive term and is defined in Equation 3.6 where r is the inter-atomic distance.

$$\mathbf{F}_{repel} = \begin{cases} 0 & \text{, if } \mathbf{r} \geq \mathbf{s} \bullet \mathbf{r}_{min} \\ \mathbf{k}_{vdw} (\mathbf{s}^2 \bullet \mathbf{r}_{min}^2 - \mathbf{r}^2)^2 & \text{, if } \mathbf{r} < \mathbf{s} \bullet \mathbf{r}_{min} \end{cases} \quad \text{..... 3.6}$$

The associated force constant $\mathbf{k}_{vdw}$ may be varied during the calculation and the values of the inter atomic distance ($\mathbf{r}_{min}$) are the standard values of the van der Waals radii represented by the Lennard-Jones potential.[109] Alteration of the van der Waals radius scale factor **s** allows for manipulation of the degree of van der Waals repulsion so as to maintain a soft repulsion term during the initial structure determining steps of the protocol allowing atoms to pass closer to each other.

The NOE distance restraints are represented by a square-well potential with the variable force constant (k_{NOE}) where r_{ij}^{u} and r_{ij}^{l} are values of upper and lower limits of the applied distance constraint respectively and r_{ij} represents the actual calculated inter atomic distance. The torsion angle restraints are also represented by a square-well potential with the force constant k_{tor} where φ_i^{u} and φ_i^{l} are upper and lower limits respectively.

$$F_{NOE} = \begin{cases} k_{NOE}(r_{ij} - r_{ij}^{u})^2 & , \text{if } r_{ij} \geq r_{ij}^{u} \\ 0 & , \text{if } r_{ij}^{l} < r_{ij} < r_{ij}^{u} \\ k_{NOE}(r_{ij} - r_{ij}^{l})^2 & , \text{if } r_{ij} \leq r_{ij}^{l} \end{cases} \quad \text{..... } 3.7$$

$$F_{tor} = \begin{cases} k_{tor}(\varphi_i - \varphi_i^{u})^2 & , \text{if } \varphi_i \geq \varphi_i^{u} \\ 0 & , \text{if } \varphi_i^{l} < \varphi_i < \varphi_i^{u} \\ k_{tor}(\varphi_i - \varphi_i^{l})^2 & , \text{if } \varphi_i \leq \varphi_i^{l} \end{cases} \quad \text{..... } 3.8$$

The strategies involving the application of simulated annealing for protein structure determination from NMR data have recently been proposed.[105,110] The DSA calculation which is commonly used in protein structure calculation, describe a real space method based on the principles of simulated annealing which circumvents the folding problem completely and the starting structures are generated from a completely random array of atoms.[105]

Practically, the DSA process is begun at the maximum temperature specified (starting temperature), and the system is held at that temperature for a certain amount of time (Plateau time). During this period, the system can rearrange itself into a different conformation. During the annealing time, the temperature is reduced until the minimum temperature is reached. At this point, the first cycle is completed and the process may be continued for the specified number of cycles. The conformations obtained by simulated annealing are further minimised to ensure that the system is truly in a low energy state. In addition to available molecular dynamic programs, DSA can be used not only to refine an initial set of approximate coordinates but also to introduce major structural changes.

Advantages of the DSA calculations

(1). A higher starting temperature can be used to surpass torsional barriers.

(2). A longer plateau time can be introduced to reach dissimilar previous conformations.

(3). More stable systems can be reached by lowering the finishing temperature.

(4). A longer annealing time can be introduced to avoid bad structures.

(5). Higher number of cycles can be used to discover important conformations.

(6). Different annealing methods such as stepwise, exponential and linear can be introduced during cooling step.

3.9 Energy Minimisation (EM)

Energy is a function of the atomic coordinates and the energy minimisation (EM) program attempts to generate the coordinates which correspond to a minimum of energy. All the minimisation methods currently used are descent series methods. They are iterative methods in which the atomic coordinates are modified from one iteration to the next in order to decrease energy. The minimisation procedure consists of moving the atoms of a molecule in such a way as to always reduce the total energy of the system based on an empirical representation of the interaction energy of the atoms of a molecule. The EM methods are generally unable to find the global energy minimum and most of the time, only a local minimum is found. The only way to find the global minimum is to explore different sets of starting coordinates.

In highly distorted structures the potential energy surface and its derivatives are often discontinuous. Simplex minimisation can handle these areas while a derivative based procedure can not. Five common optimisation procedures are provided for finding a local minimum of the energy function.

(1). Powell minimisation

(2). Conjugate Gradient minimisation

(3). Newton Raphson minimisation

(4). Steepest Descent minimisation

(5). Broyden, Fletcher, Goldfarb and Shanno minimisation

The results obtained with energy minimisation depend on the starting coordinates. It is necessary to perform minimisation calculations with different sets of starting conformations. Energy minimisation covers only a small part of the configuration space but is capable of relaxing the strain in a molecule by small local positional adjustments.

Optimisation of conformational energy is often a convenient way of obtaining reasonable geometry for molecular models. Good covalent geometry, energetically acceptable hydrogen bonds and total absence of steric overlaps are certainly expected as the output of a carefully conducted energy minimisation study. In every minimisation procedure, overall translation and rotation of a molecule is never observed.

3.10 Back Calculation[111-114]

A quantitative comparison of the experimental and back-calculated NOESY spectra is the basis of fitting NOE data with an iterative full relaxation matrix refinement. Once a final structure or family of structures is obtained, it is fairly straight forward to back-calculate what the NOE spectrum should be for such a structure using a full relaxation matrix analysis.[113,114] These can be used to qualitatively investigate the reliability of the structure, and significant discrepancies should be addressed.

In addition to its use in evaluating the accuracy of structures, NOESY back-calculation can be used for the further refinement of previous structures in order to obtain consistency with experimental NOESY data. Since distances estimated directly from

NOE data which is done by forward calculation are prone to errors, the back-calculation of NOESY spectra from structures provides the user with a means of adjusting those structures to better agree with the experimental data. If one identifies simulated cross peaks that are too big or too small compared to their experimental intensities at any mixing time, the corresponding distance bounds can then be adjusted in the proper direction. Another advantage of the back-calculation lies in the fact that all spins, not only those that have been assigned and measured, are considered in the NOESY simulation. Agreement with experimental NOESY data is an essential requirement for NMR based structure determination, but because one can reproduce experimental data, one can not assume that an accurate structure has been determined.

3.11 Evaluation of Structures

It is clear that the computational methods to generate structures from NMR data can have an important impact on the accuracy and reliability of the obtained structural results. The precision of a conformation obtained by structure calculations strongly depend on the relative and absolute magnitudes of the weights placed on the constraints, for example, the weighting factors in the DG objective function,[65,68] or target function[86] or the force constants used in the DSA or MD.[110,115]

The precision of conformation is typically expressed using root mean square deviations (RMSD) of the atomic coordinates of an ensemble of structures obtained by executing repeated, independent fits of the same constraints. Therefore, these structures have to be obtained by unbiased random sampling of the accessible conformational space. The average displacement among equally reliable structures represents the minimum diversity of conformers that meet the constraints. Typical results for globular proteins with upto 6-12 NOEs per residue are 1 Å RMSD for all backbone hydrogens and 2 Å RMSD for side chain hydrogens. The other way to estimate the quality of the structures determined is evaluation of the proximity between simulated NMR data from the structures built and the analogous real experimental NMR data from which these structures were determined.

The procedures and parameters used for achieving a structure from a given set of distance constraints are often not usable with another structure and data set. The quality of structures and the best convergence are often a matter of luck; with significant interaction between the user and the parameters used for some protocols.

3.12 Conclusion

Molecular dynamics programs search the thermally accessible vicinity of a particular starting conformation.[116] It is expected that MD refinements do not converge to a global minimum. The sampling of the MD algorithm can be improved by increasing the temperature and using a soft nonbonded repulsion term that allows atoms to pass through each other. During EM and MD refinement the electrostatic interactions are significantly improved for all structures, whereas bad atomic overlap is relaxed. The lower internal energies for MD refined structures compared to only EM refined structures clearly reflect the ability of MD to explore a much larger part of the conformational space than EM.

Distance geometry can be used to derive initial three dimensional structures consistent with the NMR data but in all cases have very high potential energy terms. The DG structures satisfy distance constraints reasonably well, but owing to crude geometric terms used to describe the atomic interactions in a molecule, the physical reliability is poor. MD refinement of DG structures normally leads to a considerable decrease of the internal energy while the consistency of the structures with the NMR data is maintained or even improved. Application of molecular dynamics was shown to enhance variety in a set of DG structures, the lac repressor headpiece,[117] BSPI 2,[118] and the phoratoxin.[119]

Powerful strategies on high-temperature dynamics have been introduced by Nilges et al..[110] These methods are capable of sampling a larger region of the conformational space than MD using a physical force field. Adequate sampling of a MD simulation is dependent on the length of the simulation and the possibility of crossing large energy barriers. Crossing large barriers is unnecessary if starting from DSA structures. As noted by researchers, distance geometry calculations followed by DSA then MD/rMD generates equally plausible structures consistent with the NMR data.

CHAPTER 4

SOLUTION STRUCTURES OF LINEAR ENDOTHELIN DERIVATIVES

4.1 Introduction

4.1.1 Literature Review of Endothelins

Endogenous vasoactive peptides such as angiotensin II, vasopressin, neuropeptide Y and endothelin are potent vasoconstrictors acting on smooth muscle and in the central nervous system.[120] Endogenous peptidic vasodilators presumably act in concert with the vasoconstrictor peptides to maintain homeostasis. It is interesting that the vasoconstrictor substances are usually mitogenic, while the vasodilatory peptides inhibit cell growth. Endothelial cells are known to be capable of releasing vasoactive substances that regulate smooth muscle tone and platelet function.[121]

ET-1, now known to belong to a new peptide class, is some ten fold more potent than the vasoconstritor angiotensin II, and has extremely long lasting pressor effects. The discovery of ET-1, a potent vasoconstrictor peptide released from endothelial cells, has attracted great interest as one possible candidate for (endothelium derived vasoconstrictor factors) EDCF.[122] The endothelin has been proposed to mediate vasoconstriction via production of EDCF in response to various chemical and physical stimuli. Indeed the expression of this peptide has been highly conserved during the course of vertebrate evolution and may perform similar homeostatic functions in a variety of mammalian and nonmammalian species.

Over the last 6 years this peptide has drawn the attention of many investigators because of its unique structure and numerous biological actions (Table 4.1). Some of the biological actions of the ET's (Table 4.1) are described and the evidence for a possible involvement in a variety of diseases are reported (Table 4.2).

The development of selective receptor antagonists and/or processing inhibitors is eagerly awaited and may provide novel therapeutic agents for the treatment of a variety of human diseases.

4.1.1 Identification and Characterisation

The term "endothelin" refers to a family of 21 amino acid peptides found in four distinct isoforms,[123-124] ET-1, ET-2, ET-3 and ET-β or mouse vasoactive intestinal contractor (VIC) (Figure 4.1).[125] ET-1 was originally discovered in the supernatant liquid of cultured bovine aortic endothelial cells and subsequently isolated from the culture supernatant liquid of porcine aortic endothelial cells.[123] The primary sequence of human endothelin has been deduced from a human placental cDNA library and found to be identical to that of porcine endothelin, now referred to as endothelin-1 (ET-1).[124]

Common three forms of endothelins, ET-1, ET-2 and ET-3 appear to be distinct gene products. Two endothelin related genes were identified by cloning and sequence analysis of the mouse genome.[125] One encoded the peptide ET-1, while the other encoded a new peptide differing by three amino acid residues. The gene for this novel peptide is only expressed in the intestine and has been referred to as "vasoactive intestinal contractor" (VIC).[125]

As illustrated in Figure 4.1, the endothelin isopeptides have significant structural homology including two disulphide bonds, a cluster of polar charged side chains on the hairpin loop and a hydrophobic C-terminus (residues 16-21). These features show remarkable similarity to another group of cardiotoxic peptides, known as sarafotoxins (SRTX 6a-d, Figure 4.1) isolated from the venom of the Egyptian burrowing asp, *Atractaspis engaddensis*.[126] Relation between endothelins and sarafotoxins suggests an ancient and common evolutionary origin. A new members of the ET/SRTX peptide family named Bibrotoxin was isolated from the snake venom of the South African burrowing asp *Atractaspis bibroni*.[127]

ET-1 is the most potent vasoconstrictor discovered to date. It is some ten-fold more potent than angiotensin II, and the duration of pressor effects is extremely long.[128] Sarafotoxins also share similar biological activities to the ETs[129] suggesting that ET genes have evolved under strong pressure to conserve the structure and function of mature ET peptides.

Figure 4.1 : Sequence homology between the Endothelin and Endothelin-like peptides; underlined residues are sites of differences from human ET-1

	1	4	7	10	13	16	19
1. ET-1	C-S-C-S-S-L-M-D-K-E-C-V-Y-F-C-H-L-D-I-I-W						
2. ET-2	C-S-C-S-S-W-L-D-K-E-C-V-Y-F-C-H-L-D-I-I-W						
3. ET-3	C-T-C-F-T-Y-K-D-K-E-C-V-Y-Y-C-H-L-D-I-I-W						
4. ET-β	C-S-C-N-S-W-L-D-K-E-C-V-Y-F-C-H-L-D-I-I-W						

	1	4	7	10	13	16	19
5. SRTX-6a	C-S-C-K-D-M-T-D-K-E-C-L-N-F-C-H-Q-D-V-I-W						
6. SRTX-6b	C-S-C-K-D-M-T-D-K-E-C-L-Y-F-C-H-Q-D-V-I-W						
7. SRTX-6c	C-T-C-N-D-M-T-D-E-E-C-L-N-F-C-H-Q-D-V-I-W						
8. SRTX-6d	C-T-C-K-D-M-T-D-K-E-C-L-Y-F-C-H-Q-D-I--I-W						
9. BBTX	C-S-C-A-D-M-T-D-K-E-C-L-Y-F-C-H-Q-D-V-I-W						
10. Apamin	C-N-C-K-A-P-E-T-A-L-C-A-R-R-C-Q-Q-H						

	1	4	7	10	13	16	19	22
11. MCD	I- K-C-N-C-K-R-H-V-I-K-P-H-I-C-R-K-I-C-G-K-N							
12. Tertiapin	A-L-C-N-C-N-R-I-I-P-H-M-C-W-K-K-C-G-K-K							

ET-1 : Human, Porcine, Canine, Rat, Mouse, Bovine

ET-2 : Human, Dog

ET-3 : Human, Porcine, Rabbit, Rat, Mouse

ET-β (VIC) : Rat, Mouse

SRTXs : Egyptian burrowing asp, *Atractaspis engaddensis*

BBTX : South African burrowing asp, *Atractaspis bibroni*

MCD : Mast Cell Degranulating peptide

Disulphide bonds: 1-10 C^1—C^{15} and C^3—C^{11}

11-12 C^1—C^{11} and C^3—C^{15}

The sequence of ET-1 from human was shown to be the same as that from porcine, rat, mouse and bovine. There is also no difference between the sequence of human and rat, mouse, rabbit and porcine ET-3. The sequence of ET-2 from human was shown to be the same as that from dogs. The VIC, with only one amino acid difference from human ET-2 (Asn-4 verses Ser-4), is the mouse ET-2. The sequence of bibrotoxin is highly homologous to that of SRTX-6b with the only difference being the substitution of Ala-4 for Lys-4.

4.1.3 Biosynthesis of Endothelin

The synthesis and secretion of endothelin (ETs) is analogous to several other bioactive peptides. They arise from post-translational processing of large isopeptide specific prohormones. Yanagisawa et al.[123] cloned the procine ET gene and based on cDNA sequencing, predicted the amino acid sequence for preproendothelin. Since then several cDNA and genomeric clones for preproETs (ET-1, ET-2 and ET-3) have been isolated from various species, namely human, murine, canine, porcine, rat, rabbit and bovine.[124-125, 130] The genomeric DNA blotting shows that each ET isopeptide is encoded by a distinct gene.

Figure 4.2 : Biosynthesis of Endothelin[131]

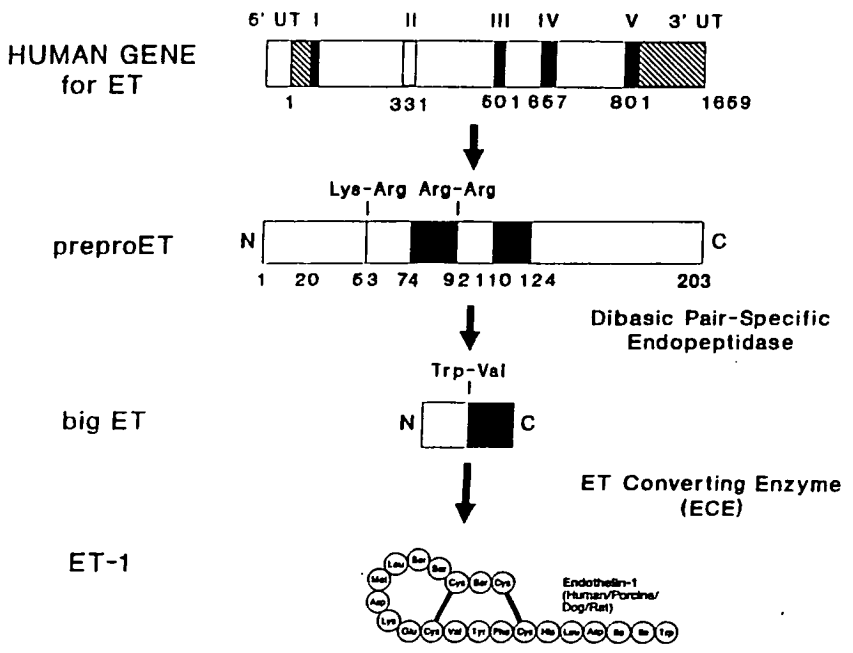


Figure 4.3 : Sequences of Big Endothelins

```
7       4       1
M-L-S-S-C-S-C                ECE                Human big-ET-1
|       /       \            ↓
D-K-E-C-V-Y-F-C-H-L-D-I-I-W-V-N-T-P-E-H-V-V-P-Y-G-L-G-S-P-R-S
8       11      14      17      20      24      27      30      33      36

7       4       1
M-L-S-S-C-S-C                ECE                Bovine big-ET-1
|       /       \            ↓
D-K-E-C-V-Y-F-C-H-L-D-I-I-W-V-N-T-P-E-H-V-V-P-Y-G-L-G-S-P-S-R-S
8       11      14      17      20      24      27      30      33      36      39

7       4       1
M-L-S-S-C-S-C                ECE                Porcine big-ET-1
|       /       \            ↓
D-K-E-C-V-Y-F-C-H-L-D-I-I-W-V-N-T-P-E-H-I-V-P-Y-G-L-G-S-P-S-R-S
8       11      14      17      20      24      27      30      33      37

7       4       1
M-L-S-S-C-S-C                ECE                Rat big-ET-1
|       /       \            ↓
D-K-E-C-V-Y-F-C-H-L-D-I-I-W-V-N-T-P-E-R-V-V-P-Y-G-L-G-S-P-S-R-S
8       11      14      17      20      24      27      30      33      36      39

7       4       1
L-W-S-S-C-S-C                ECE                Human big-ET-2
|       /       \            ↓
D-K-E-C-V-Y-F-C-H-L-D-I-I-W-V-N-T-P-E-Q-T-A-P-Y-G-L-G-N-P-P
8       11      14      17      20      24      27      30      33      36

7       4       1
K-Y-T-F-C-T-C                ECE                Human big-ET-3
|       /       \            ↓
D-K-E-C-V-Y-Y-C-H-L-D-I-I-W-I-N-T-P-E-Q-T-V-P-Y-G-L-S-N-Y-R-G-S-P-R
8       11      14      17      20      24      27      30      33      36      39
```

ECE : Endothelin Converting Enzyme

The predicted amino acid sequence for preproETs are proteins (ca 200 amino acids) with specific species and isopeptide differences. ET-1 is derived from a 203 amino acid peptide precursor known as preproendothelin, which is cleaved after translation by endopeptidases specific for the paired dibasic residues to form a 38 (human) or 39 (porcine) amino acid peptide, proendothelin or big-ET (Figure 4.2).[131]

The three isoforms of big-ET have several different features (Figure 4.3) including;

1. Length of the peptide chain; big-ET-1, -2, and -3 contain 38, 37 and 41 amino acids respectively.
2. Big-ET-3 has Ile-22 instead of Val-22 at the scissile bond.
3. Residues 27-29 in big-ET-1 are His-Val-Val, in big-ET-2 are Gln-Thr-Ala and in big-ET-3 are Gln-Thr-Val.
4. The C-terminal regions of big-ET-1 and big-ET-3 are somewhat dissimilar while those of big-ET-1 and big-ET-2 are moderately conserved.
5. No glutamines are present in big-ET-1 and unique methionine and alanine residues are present in big-ET-1 and big-ET-2 respectively.
6. The amino acid homology of Tyr and Thr are higher in big-ET-3 while Ser is higher in big-ET-1.

Big-ETs undergo a proteolytic cleavage at an unusual processing site, Trp-21-Val-22, by a putative endothelin converting enzyme (ECE) to give mature ETs and the corresponding C-terminal fragments (Figure 4.3).[123,128] Due to much lower vasoconstricting potency of big-ETs than ETs the conversion of big-ET to ET appears essential for physiological activity.[132] The physiological importance of cleavage of ET(1-39) is indicated by the reported 140-fold increase in vasoconstrictor activity upon cleavage to ET-1.[132]

Recently *in vivo* expression of synthetic RNA encoding human preproET-1 in *xenopus oocytes*[133] resulted in the secretion of ET-1 and big-ET-1. In the subsequent study [134] a set of preproET-1 mutants were designed and expressed. This study demonstrated that big-ET-1 is a necessary intermediate in the formation of ET and the peptide segment after the C-terminus of big-ET-1 in preproET was not necessary for the synthesis of ET-1 or big-ET-1.

Additionally it was shown that different preproETs mutated (4 mutants) at the Trp-Val site (Trp-21 was replaced with Phe and Tyr and Val-22 by Asp and the most conserved residue in all ET related sarafotoxins Ile-20 with Ala) are processed into their corresponding mature 21 amino acid peptides, indicating the absence of a strict sequence specificity. Recently it was shown that amino acids 91-212 were not necessary for the synthesis and secretion of big-ET-1 and ET-1 by using Stop91 mutant.[134] The exact site of the conversion of big-ET to ET in the presence of ECE is still not known as big-ET and ET have been found in the supernatant of endothelial cells and in the plasma.[135]

4.1.4 Tissue Distribution of Endothelin

ET-1 mRNA is widely expressed in rat, porcine, guinea pig, and human tissues. The distribution of the propeptide, big-ET and immunoreactive ET-1, has been compared in porcine tissues.[136] The concentration of immunoreactive big-ET was highest in the aortic intima and lung, while the highest concentration of immunoreactive ET-1 was found in the kidney inner medulla. The broad range of binding sites indicates that ET may function in the regulation of a variety of organ systems. In addition to endothelial cells, from which endothelin obviously derives it name, ET-1 is produced by mesangial, kidney, and epithelial cells and also by various human cancer cell lines and human macrophages.[131] ET gene transcription occurs in a variety of functional regions in the human brain, especially the hypothalamus. Evidence for transcription and expression of the ET-3 gene in the human placenta and the ET-2 gene in human tumour cells has only recently been reported.[137]

4.1.5 Physiological and Pathological roles of Endothelin

The ET's elicit a long lasting vasoconstriction in almost all arteries and veins.[138] Numerous reports have described the effects of ET on the cardiovascular system *in vitro* and *in vivo*. Some of these actions are summarised in Table 4.1 and have been reviewed recently.[139] There have been several reports describing the initial transient but potent vasodilatation of ET that appears to be selective for certain arterial beds.[140] Intravenous infusions of ET-1 to humans (1, 2.5 and 5.0 ng/kg per min) caused increases in mean blood pressure and serum potassium concentration.[141]

The many potent effects of the ET's on the cardiovascular system have implicated this peptide class in a variety of human diseases (Table 4.2). There have been several reports implicating ET in the pathogenesis of congestive heart failure and myocardial ischaemia.[142-143]

Table 4.1 : Biological Actions of Endothelin[131]

Tissue/Organ	**Effect**
Vascular smooth muscle	long-lived constriction of isolated vascular muscle mitogenic actions in cultured smooth muscle cells release of endothelium-derived relaxing factor coronary arterial vasoconstriction, increased perfusion pressure
Nonvascular smooth muscle	constriction of intestinal, tracheal and uterine smooth muscle
Heart	increased contractility increased heart rate stimulation of ANP release
Nervous tissue	enhanced neurotransmitter release
Kidney	inhibition of renin release decrease in renal blood flow decrease in glomerular filtration rate urinary Na^+ and K^+ excretion
Adrenal glands	stimulation of aldosterone biosynthesis release of catecholamines

Table 4.2 : Possible physiological roles and beneficial actions of ET[131]

Cardiovascular diseases	myocardial ischaemia
	congestive heart failure
	arrhythmia
	hypertension
	unstable angina
Bronchoconstriction	pulmonary hypertension
	asthma
Neuronal action	cerebral vasospasm
	subarachnoid haemorrhage
Renal disease	acute/chronic renal failure
Gastric mucosal damage	gastrointestinal disorders
Vascular disorders	artheosclerosis
	complications in diabetes
	complicated vascular disorders
Cancer	pulmonary cancer
	carcinoma cancer
Other	endotoxic shock
	septicaemia
Beneficial actions	physiological regulation of blood pressure
	neuroendocrine regulation
	closure of umbilical vessels
	wound healing
	control of menstruation

4.1.6 Receptor Studies of Endothelin

ET has multiple specific binding sites on various kinds of cell membranes including smooth muscle, epithelial and endothelial cells.[144-146] Two ET receptor subtypes, termed ET_A and ET_B, were first identified, cloned and expressed from bovine and rat lung respectively.[147-149] Both are G-protein coupled and belong to the rhodopsin family, with seven transmembrane domains. One, isolated from bovine lung, highly specific for ET-1 and ET-2 is located in the periphery and CNS, and has been suggested to be the vascular smooth muscle type.[148] The proposed nomenclature, based on the relative affinity of the agonists for the receptors, has termed this receptor as the ET_A type. The other is a "nonselective" subtype that binds ET-1, ET-2 and ET-3 with similar affinity[149-150] and has been termed the ET_B receptor.

There is some evidence for an ET-3 specific receptor subtype that is located primarily in brain and in endothelial cells.[151] Recently, an ET-3 specific receptor, named ET_C was cloned and characterised from *Xenopus laevis* dermal melanophores.[152] It is not known whether this receptor subtype is present in mammalian species.

Recently ET receptors have been cloned and expressed from a variety of human tissues.[128,132,136] The bovine ET_A receptor consists of 427 amino acid residues and is highly specific to ET-1. The rat and human ET_B receptors consist of 441 and 442 amino acid residues respectively and show equal affinity towards ET-1, ET-2 and ET-3. The ET_B receptor was first found on endothelial cells and it has been identified in many types of tissues including brain, vascular smooth muscle, lung, kidney and heart.[151-153] It was first thought that the ET_B receptor was associated with vasodilator activity[149] partially due to the release of the EDRF. However, numerous recent reports have shown that ET_B receptor also mediates vasoconstriction in certain tissues and species.[154-155] The molecular characteristics and functions of ET_A and ET_B receptors have been reviewed extensively.[156-157]

The distribution of ET receptor subtypes in human tissues is of importance due to the species differences that are emerging and for the design of relevant receptor antagonists. The literature to date on this subject is summarised in Table 4.3.[158]

There are many studies ongoing in a variety of animal tissues attempting to elucidate the existence and distribution of ET receptor subtypes. Comparison of the receptor affinities of various ET's and SRTX's in rat aorta and atria (ET_A) or cerebellum and hippocampus (ET_B) indicates that SRTX-6c is a selective agonists for the cerebellum/hippocampus receptors.[159] A further study indicated that this ligand exerted only vasodialation in the rat aortic ring, possibly through the release of EDRF from the endothelium.[160]

Other selective ET_B ligands, for example, the linear analogue ET[1,3,11,15-Ala] and truncated analogous ET[6-21, 11,15-Ala], ET[8-21, 11,15-Ala] and N-acetyl-ET[10-21], have been reported to cause vasorelaxation in isolated, endothelium-intact porcine pulmonary arteries.[161] Some of these analogues, ET[1,3,11,15-Ala] and ET[8-21, 11,15-Ala], are potent ET_B agonists causing vasoconstriction in the rabbit pulmonary artery.[162] The present evidence available indicates that the physiological response mediated by the ET_B receptor in certain tissue beds cannot be sorely described by vasodialation. Indeed it appears that vascular smooth muscle can possess an ET_B-like or nonselective receptor subtype. It will be important to determine the relevance of these reports to the distribution of receptor subtypes in the human. The use of specific antibodies to the ET_A and ET_B receptors should enable useful receptor localisation studies to be performed.

Table 4.3 : Distribution of ET receptor subtypes in human tissues [158]

Tissue	Predominant subtype
Placenta	ET_B
Umbilical vessels	ET_A
Renal cortex, medulla	$ET_B > ET_A$
Parathyroid gland	ET_A, ET_B
Myometrium	ET_A
Bronchus	ET_A, ET_B
Pulmonary artery	ET_A
Skin	$ET_A > ET_B$
Giraldi heart cells	ET_B
Liver	$ET_B > ET_A$
Myocardium	ET_A, ET_B
Atrioventricular conducting system	ET_A, ET_B
Coronary artery	ET_A, ET_B
Ulterus	ET_A
Hippocampus	ET_B
Brain (cortex, cerebelium, brain stem, basal ganglia, hypothalmus)	ET_A, ET_B
Spinal cord	ET_B
Kidney	$ET_B > ET_A$
Adrenal	$ET_B > ET_A$
Atria, Aorta	$ET_A > ET_B$
Lung	$ET_A > ET_B$
Stomach	$ET_A > ET_B$

4.1.7 Aim of the Study

Numerous studies have been reported on the pharmacological evaluation of ET analogues and fragments. Structure activity relationship studies of ET analogues have been reviewed extensively.[158,163-164] After careful consideration of the literature reports published on biological and pharmacological evaluation of endothelins, protein and peptide research group in our department decided to concentrate on the ET_B selective analogues.

So far none of the ET_B selective analogues has been found through the structure activity relationship studies. In order to find the requirements for the ET_B receptor recognition, a series of linear analogues of ET-1 have been synthesised and biological activity tested. The synthesis of the peptides were carried out by Dr. Lu Jiang in our laboratory. Some of the linear endothelin analogues (Figure 4.4) which showed promising biological activity are tabulated in Table 4.4.

Linear endothelin analogues were synthesised by protecting all the cysteines with $-CH_2-NH-CO-CH_3$ group. Methionine is liable to oxidation and it is commonly replaced by norleucine or leucine without loss of biological activity.[165] It can be seen that both the norleucine and leucine are well tolerated at position seven since similar biological activities were observed for both of the analogues. Therefore methionine at position seven was replaced by leucine for all the analogues except LJP1B.

To improve the stability, solubility and the helicity of the analogues, some of the cysteine residues were replaced by α-aminoisobutyric acid. Replacing cysteine at position 3 and/or 11 didn't alter the biological activity. The amino acid residues D^8, F^{14} and C-terminal H^{16}-W^{21} were suggested as important residues in the biological assay.[158] Stereochemistry of these residues were changed to study the importance of individual amino acids. Replacing one of the crucial amino acids, histidine, by phenylalanine showed completely loss of ET_B activity. Truncated analogues were also included for biological evaluation.

To understand the basic structural requirements for biological activity, endothelin-1 was subjected to solution structure determination. It is well known that the truncated and reduced ET-1 analogues didn't produce promising biological activity results.

Out of 100 of linear ET-1 analogues tested, LJP1 showed both the highest ET_B activity and the ET_A/ET_B selectivity. Most of the linear ET-1 peptides which have alanine at position 3 and d-aspartic acid at position 8 show quite low ET_B activity and ET_A/ET_B selectivity (Table 4.4). Therefore these two peptides, LJP1 and LJP26, along with ET-1, were selected for solution structure studies.

Most peptides are normally found in aqueous environments and H_2O would therefore seem an obvious choice. However, peptides act at protein surfaces or in membranes, which are less polar, and therefore less polar solvents may give a more relevant result. Less polar solvents also tend to induce more structure in peptides because of the weak character of the solvent for hydrogen bonding.

Either methanol or TFE is often added to aqueous solutions to induce helix formation, the assumption being that the helices seen in such solvent systems are representative of the helices formed in their native environment, especially in membranes.

Most of the linear ET-1 analogues were partially dissolved in H_2O and some of them did not dissolve at all, especially with more alanine residues. To alleviate this problem, peptide samples were dissolved in methanol/H_2O mixture and minimum amount of methanol (50%) was used.

In this thesis ^{1}H NMR studies of ET-1 and two endothelin analogues, LJP1 and LJP26, are presented and possible 3D structures derived from NOESY data using molecular modelling techniques are discussed.

Figure 4.4: Sequences of ET-1 and modified linear ET-1 derivatives

	1				5					10					15					20
ET-1	C	S	C	S	S	L	M	D	K	E	C	V	Y	F	C	H	L	D	I	I	W
LJP1	C	S	ϕ	S	S	L	L	D	K	E	ϕ	V	Y	F	C	H	L	D	I	I	W
LJP1B	C	S	ϕ	S	S	L	L^n	D	K	E	ϕ	V	Y	F	C	H	L	D	I	I	W
LJP1C	ϕ	S	ϕ	S	S	L	L	D	K	E	ϕ	V	Y	F	ϕ	H	L	D	I	I	W
LJP2	C	S	A	S	S	L	L	D	K	E	ϕ	V	Y	F	C	H	L	D	I	I	W
LJP2B	C	S	A	S	S	L	L	D	K	E	A	V	Y	F	C	H	L	D	I	I	W
LJP24	C	S	A	S	S	L	L	D	K	E	ϕ	V	Y	F	C	F^d	L	D	I	I	W
LJP25	C	S	A	S	S	L	L	D^d	K	E	ϕ	V	Y	F	C	F^d	L	D	I	I	W
LJP26	C	S	A	S	S	L	L	D^d	K	E	ϕ	V	Y	F	C	H	L	D	I	I	W
LJP7-21							L	D	K	E	ϕ	V	Y	F	C	H	L	D	I	I	W

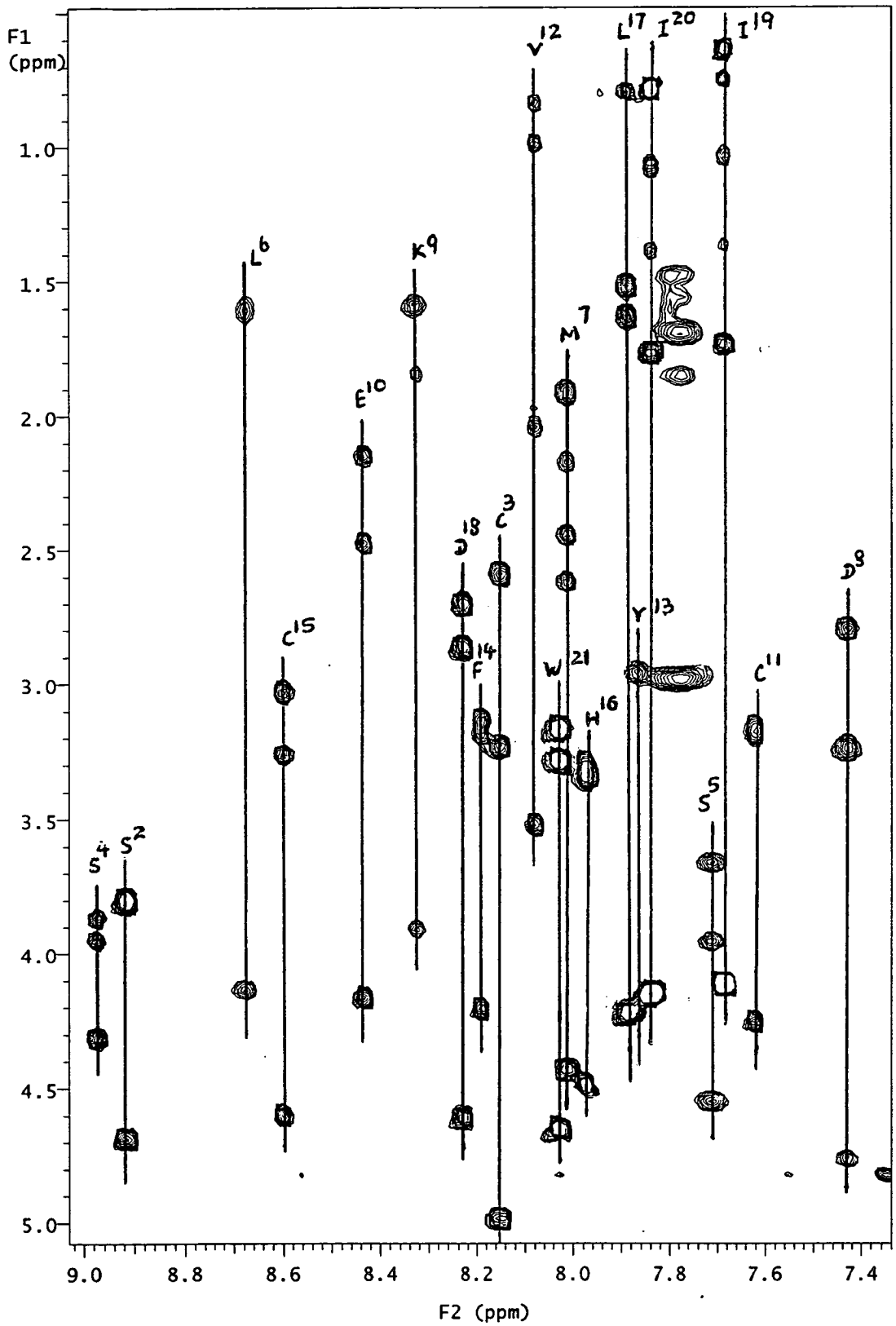
ϕ - α-aminoisobutyric acid; L^n - norleucine; X^d - D amino acid
All the cysteine side chains are protected by $-CH_2-NH-CO-CH_3$ group.
Underlined residues are sites of differences from ET-1.

Table 4.4: Biological Activity of Endothelin Peptide Analogues

Peptide	$IC_{50}(\mu M)$ ET_A[1]	$IC_{50}(\mu M)$ ET_B[2]	Selectivity ET_A/ET_B	$EC_{50}(\mu M)$ ET_B[3]
ET-1	0.0002	0.0016	0.13	0.0003
LJP1	9.1	0.00006	151,667	0.00059
LJP1B	4.5	0.0009	5,000	0.00048
LJP1C	0.77	0.0003	2,567	0.0013
LJP2	6.2	0.0001	62,000	0.00053
LJP2B	16	0.0007	22,857	0.0021
LJP24	17	0.56	31	0.48
LJP25	13	1.8	7	0.54
LJP26	16	0.002	8,000	0.00032
LJP7-21	2.8	0.0009	3,111	0.0013

1 : Tested in rabbit renal artery vascular smooth muscle cells
2 : Tested in rat cerebellum
3 : Tested in Chinese Hamster ovary cells
IC_{50} : Concentration of drug giving 50% displacement of specific binding to the receptor
EC_{50} : Concentration of drug giving 50% of the maximum response

Figure 4.5 : Fingerprint region of the TOCSY spectrum of ET-1. Each vertical line shows the individual spin system.



4.2 Results and Discussion

4.2.1 The Endothelin-1

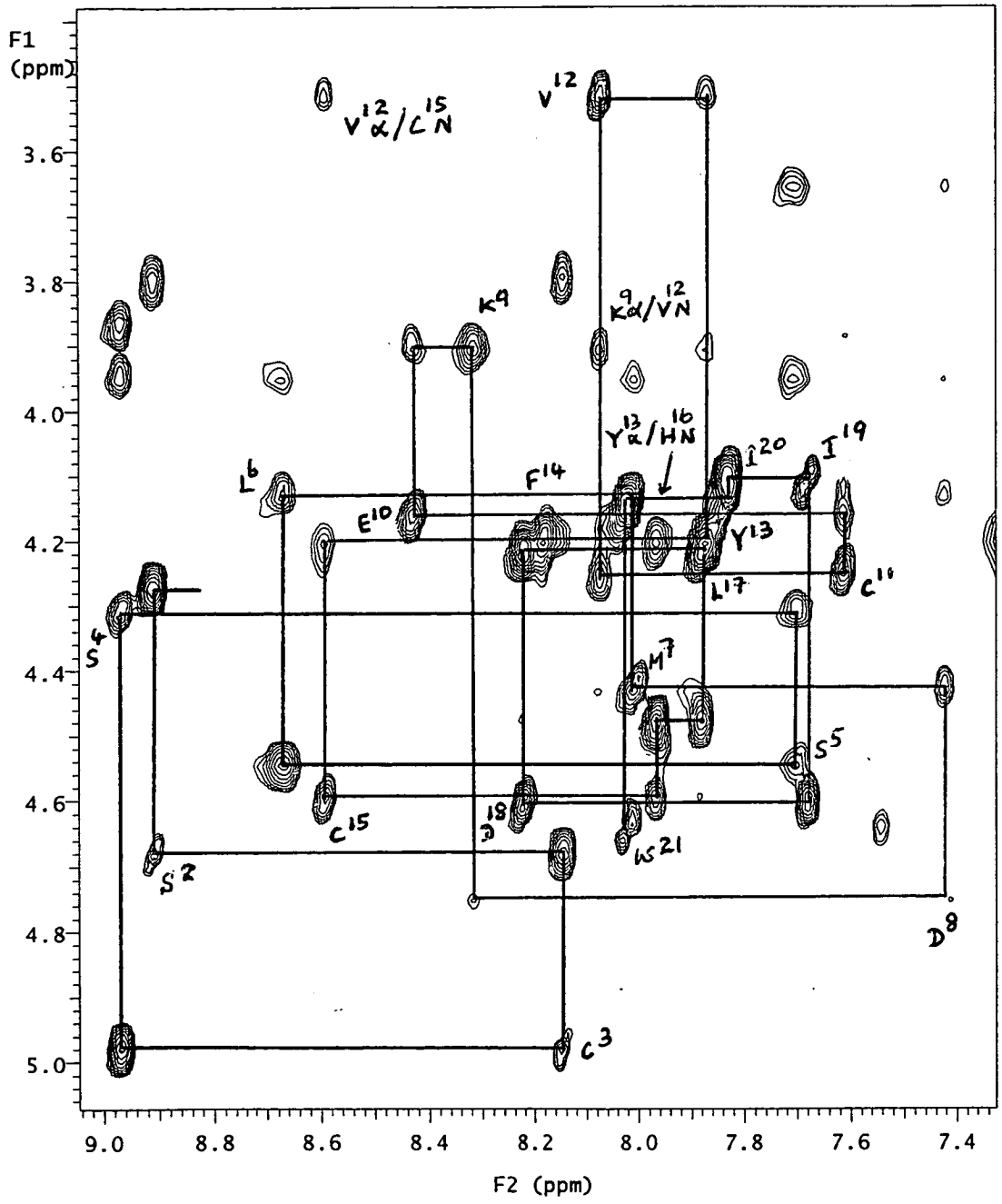
4.2.1.1 Interpretation of NMR data of Endothelin-1

The importance of the biologically active structure of a peptide in determining the specific receptor interactions and eliciting its pharmacological properties has made the three dimensional structure of the ET's a subject of intense interest.

The endothelins (Figure 4.1) are a family of bicyclic, 21 amino acid peptides whose first member is identified as ET-1. The sequence of 21 (Figure 4.1) residues in ET-1 includes 8 unique amino acid members, Met-7, Lys-9, Glu-10, Val-12, Tyr-13, Phe-14, His-16 and Trp-21. The two disulphide bridges which make the bicyclic core at the N-terminus region are formed at positions 1-15 and 3-11. The proton 1D, DQF COSY, TOCSY and NOESY spectra have been acquired for ET-1 and the experimental conditions are given elsewhere. The DQF COSY and TOCSY spectra were obtained in 50:45:5 = CD_3OH:H_2O:D_2O and all exchangeable and non exchangeable protons give signals under these conditions. The fingerprint region of the DQF COSY spectrum showed 20 cross peaks for NH/αH connectivities except the N-terminus residue, Cys-1, because of the rapid exchange of the amide proton.

The TOCSY spectrum is useful guide for the identification of individual spin systems. Figure 4.5 shows the NH region of the TOCSY spectrum of ET-1. The high frequency region of the TOCSY spectrum clearly showed the complete side chain proton connectivities with the NH proton but some missing side chain connectivities for Leu-6 and Lys-9. The two serine residues, Ser-4 and Ser-5, were clearly identified on the basis of their distinctive chemical shift values of the two β–proton resonances in the fingerprint region of the TOCSY spectrum. Ser-2 showed only a single peak for β–protons. The two isoleucines, Ile-19 and Ile-20 showed complete side chain connectivities through the NH proton in the TOCSY spectrum. The unique Lysine, Lys-9, was identified on the basis of its distinctive αH/εH cross peak in the αH region of the TOCSY spectrum. The complete spin patterns for unique Val-12 was observed in the DQF COSY and TOCSY spectra.

Figure 4.6 : Fingerprint region of the NOESY spectrum of ET-1. Complete walk along the backbone is shown.



One of the two leucines, Leu-6 and Leu-17, were clearly identified on the basis of its distinctive chemical shifts of methyl resonances appeared in the αH region in the TOCSY spectrum. The TOCSY cross peaks corresponding to magnetisation transfer from αH/γH through the entire spin system were clearly visible for both unique amino acid residues Met-7 and Glu-10. The chemical shift value for εCH_3 of Met-7 was obtained from a 1D proton spectrum.

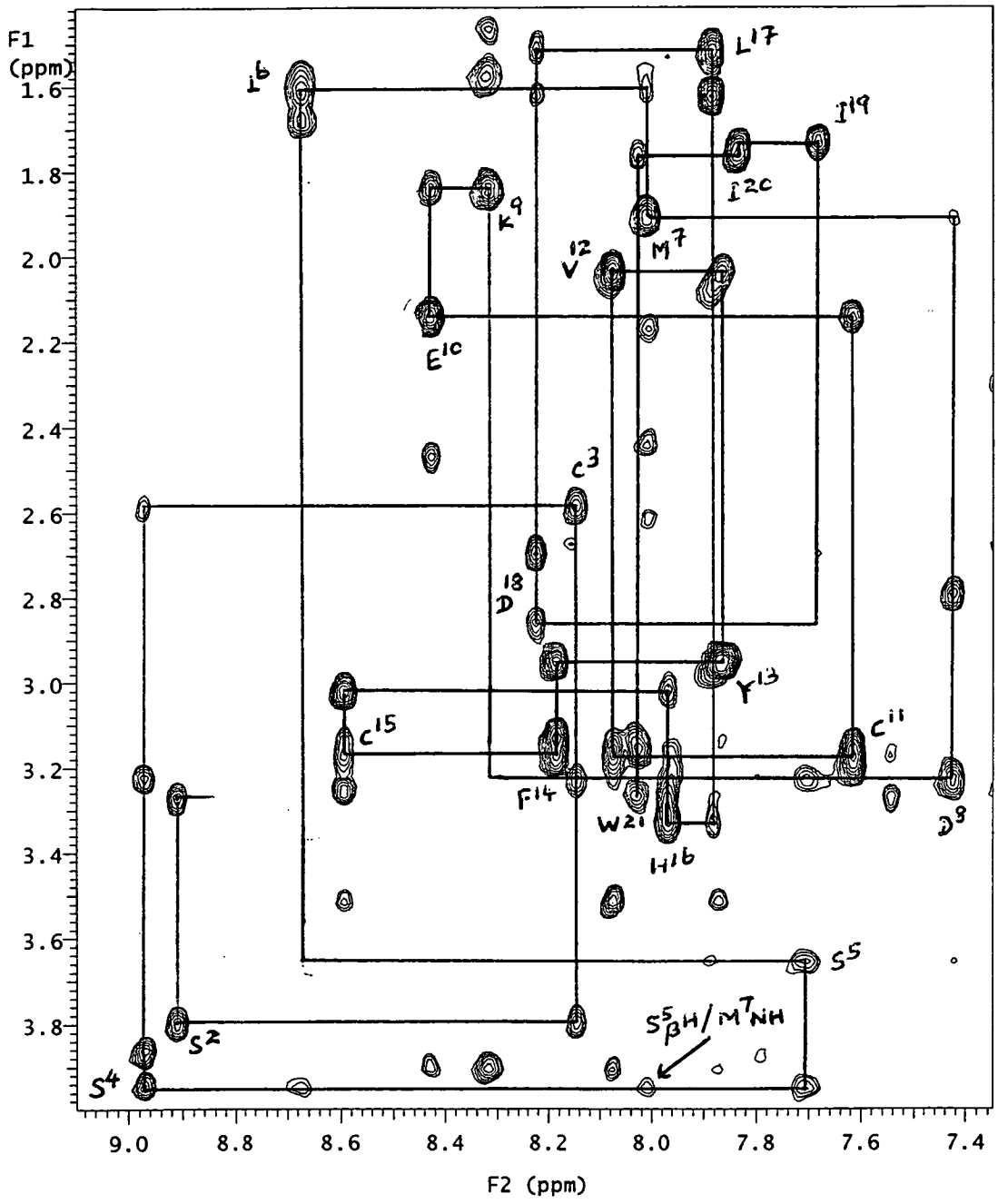
Aromatic proton connectivities for Tyr-13 and Trp-21 were clearly visible near the diagonal in the aromatic region of the TOCSY spectrum. At this stage, the AMX spin patterns for the adjacent NH/βH and βH/βH cross peaks appeared in the TOCSY and the DQF COSY spectra respectively, could not be assigned. These include the two Asp residues, four Cys residues and four unique aromatic residues.

Four unique aromatic amino acid residues, Tyr-13, Phe-14, His-16 and C-terminal end Trp-21, were clearly discriminated from the others by the aromatic protons/βH resonance connectivities in the NOESY spectrum. Then the individual identification and the positions of the residues are determined by direct comparison of the TOCSY and NOESY spectra.

The high frequency region of the NOESY spectrum is specially useful guide for sequence specific resonance assignments, secondary structure determination and finally for tertiary structure calculation of the polypeptide. Since Lys-9 is an unique residue, it is possible to identify Glu-10 thus resolving the Met-7 and Glu-10 ambiguity. Although Asp-8 NH/αH cross peak did not appear in the NOESY spectrum, the sequence L^6-M^7-D^8 is clear and hence L^{17} and D^{18} are identified. The unique C-terminal aromatic residue Trp-21 showed a clear aromatic 4H/αH and 4H/βH cross peaks in the NOESY spectrum. The D^{18}-I^{19} sequential connectivity is clear and the sequence L^{17}-D^{18}-I^{19}-I^{20}-W^{21} is then identified.

Since Val-12 is unique, it is possible to identify neighbours Cys-11 and Tyr-13. Strong sequential connectivity between His-16 and Leu-17 identify the His-16 and thence the neighbour Cys-15.

Figure 4.7 : Part of the NOESY spectrum of ET-1 showing $\beta_i H/N_{i+1}H$ connectivities



Although the sequential connectivity between Tyr-13 and the remaining Phe-14 could not be observed due to the almost same αH chemical shift values (Table 4.5), Phe-14 was easily discriminated from the other three aromatic residues which have already assigned with the aid of aromatic protons/βH cross peaks in the NOESY spectrum.

The cross peak from the terminal $C^1\alpha H/S^2NH$ was clearly visible in the NOESY spectrum thence the Cys-3 is clearly identified. Strong sequential cross peak between Cys-3 to Ser-4 thence Ser-5 complete the entire sequential connectivity of ET-1. The complete walk along the backbone sequential connectivities and are shown in Figure 4.6 and the chemical shift values are given in Table 4.5.

After completion of sequence specific resonance assignments, the individual resonance identifications of particular amino acid residues were confirmed. Almost all the side chain spin connectivities through αH were clearly observed in the DQF COSY and TOCSY spectra for all the amino acid residues. A broader resonance was observed for side chain $^+NH_3$ of Lys-9 at 7.76-7.79. Although two γH resonances were observed for Met-7, only a single γH resonance was previously reported for methionine in a random coil peptides.[61] In contrast, two separate overlapping βH and γH resonances were observed for Glu-10. Two degenerate βH resonances were observed for Cys-1 and Cys-11. Although two βH resonances were observed for Ser-4 and Ser-5, only a single resonance was observed for Ser-2.

The unique amino acid residue Tyr-13 showed complete spin pattern connectivities in the TOCSY spectrum with a single overlapped βH resonance. Although the NOESY spectrum showed strong prominent spin patterns from NH resonance to the entire side chain for Leu-6 and Lys-9, the TOCSY spectrum showed only few cross peaks, NH/βH(s) for Leu-6 and Lys-9. The high reliability of the resonance assignments was supported by the $d_{\beta N}(i,i+1)$ (Figure 4.7) and $d_{NN}(i,i+1)$ (Figure 4.8) connectivities for neighbouring amino acid residues.

Figure 4.8 : Part of the NOESY spectrum of ET-1 showing $N_iH/N_{i+1}H$ connectivities

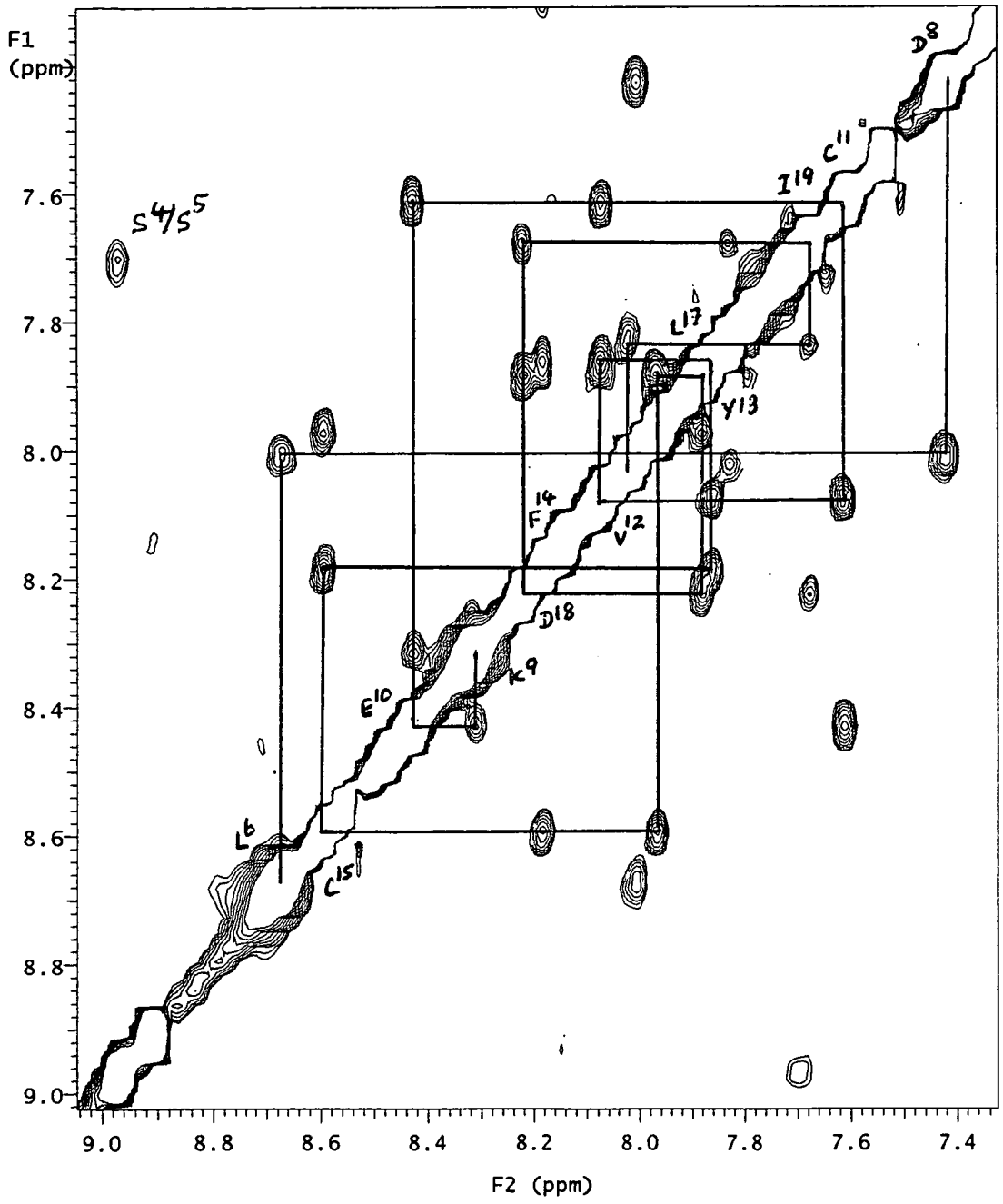
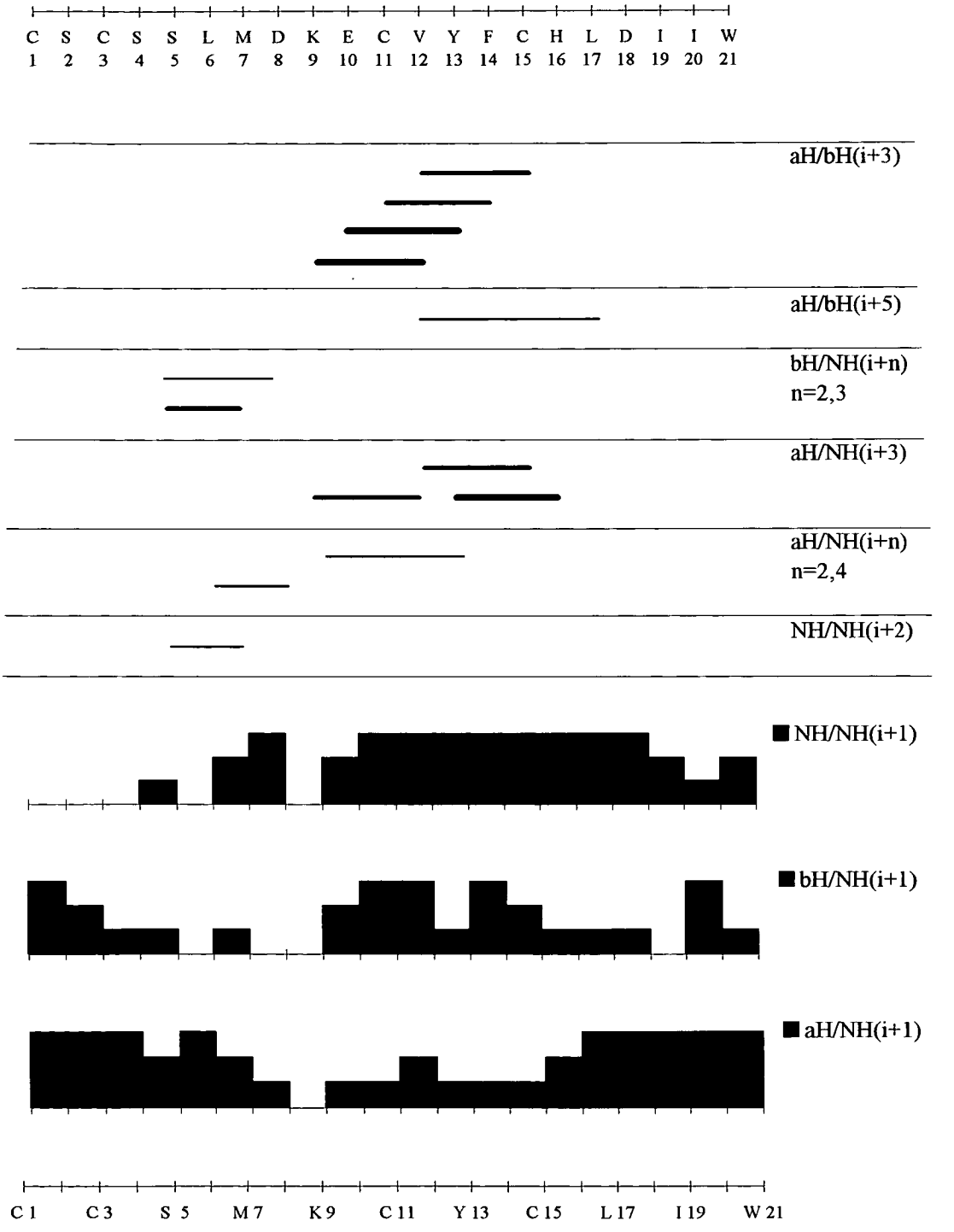


Figure 4.9 : Summary of interresidue NOESY connectivities observed for ET-1. The thickness of the columns and bars is propotional to the NOE intensity



4.2.1.2 Secondary Structure Determination

Once the complete sequential assignment had been completed, the reminder of the NOESY spectrum was examined for the evidence of secondary structural details. In addition to the neighbouring amino acid connectivities, three long connectivities, $d_{\alpha N}(i,i+3)$ K^9/V^{12}, V^{12}/C^{15} and Y^{13}/H^{16} (Figure 4.6) and one medium range $d_{\beta N}(i,i+2)$ S^5/M^7 (Figure 4.7) were observed in the fingerprint region of the NOESY spectrum. These $d_{\alpha N}(i,i+3)$ cross peaks indicates the possibility of forming a helix between K^9-H^{16}. The observation of a cross peak K^9/V^{12} indicates that the helical structure in this segment of the peptide encompasses the disulphide bridge between C^3-C^{11}.

High reliability of the formation of the helix was confirmed by the strong $d_{\alpha\beta}(i,i+3)$ cross peaks which appeared in the αH region of the NOESY spectrum. Although E^{10}/Y^{13} and C^{11}/F^{14} long range $d_{\alpha N}(i,i+3)$ cross peaks would not be distinguished owing to spectral overlap, $d_{\alpha\beta}$ (i,i+3) strong cross peaks indicated the formation of an α-helix. In addition to the above connectivities, weak $d_{\alpha N}(i,i+1)$ and strong $d_{NN}(i,i+1)$ connectivities observed in the segment of K^9-H^{16} suggested the existence of an α-helix. Figure 4.9 shows the summary of the inter residue NOESY connectivity patterns of ET-1.

4.2.1.3 D_2O Exchange Experiment

The ET-1 sample was dissolved in 1:1 = CD_3OD:D_2O and the 1D proton spectra were acquired soon after the preparation of the sample at 298K. The H-D exchange rate of amide protons provide information concerning the hydrogen bonds in the secondary structure. Although both the terminal amide protons and some surface protons disappeared quickly, those of M^7, K^9, V^{12}, Y^{13}, F^{14} and C^{15} amide protons were still clearly observed after 8 minutes. Twenty five minutes later, all the above amide protons were still observed except K^9 and E^{10} but completely disappeared after 6hrs. These observations support the conclusion that a strongly hydrogen bonded segment exists in the region between K^9 and H^{16}.

4.2.1.4 Variable Temperature Studies

The dependence of the chemical shifts of backbone amide protons on temperature was measured over the range of 290-313K. A structured state is indicated by the very small temperature gradient of D^8 $\Delta\delta/\Delta T$ = -0.7 ppb/C. Smaller temperature coefficients -1.7 ppb/C for Y^{13} and -3.5 ppb/C for K^9 and H^{16} indicate shielding from solvent exchange of the amide protons of residues 9-16 and also suggests that these protons may be involved in hydrogen bonding. Figure 4.10 shows the amide proton shielding of some amino acids of Endothelin-1. Other NOE connectivities indicating another regular structure was observed in the region between S^5-D^8. The strong d_{NN} M^7/D^8, very weak d_{NN} D^8/K^9 (can be seen at lower thresholds), medium $d_{\beta N}$ S^5/M^7 and the slow M^7 amide proton exchange could arise from a turn structure in the S^5-D^8. In the remaining parts of the ET-1, especially the C-terminal region beyond H^{16} showed a series of strong $d_{\alpha N}$ NOEs and higher $^3J_{NH\alpha}$ coupling constants for I^{19}, I^{20} and W^{21}. The 7 Hz $^3J_{NH\alpha}$ couplings are probably not very useful due to conformational averaging.

4.2.1.5 Solution Structure of Endothelin-1

Notably dissimilar NMR structures have been proposed for ET-1 by different research groups and the description of the characteristic features of ET-1 and endothelin-like peptides derived from previous NMR results are tabulated in Table 4.6. Despite the general similarity of all structures reported for peptides of the endothelin family, including ET-3 and sarafotoxin-6b, significant differences include the number of amino acid residues involved in the helix, the nature of the helix, the presence or absence of a turn at residues 5-8 and the conformation and degree of disorder reported for the C-terminus.

Figure 4.11 shows the comparison of documented literature and experimental values of NH and αH chemical shift variations. The experimental shift values of ET-1 lie in between the literature listed shift values.

A full list of structural constraints (NOE, torsional angle, hydrogen bond, disulphide linkage) and $^3J_{NH\alpha}$ coupling constants may be found in Appendix I (pages 177-184).

Figure 4.10 : Amide proton shielding of some amino acids of Endothelin-1

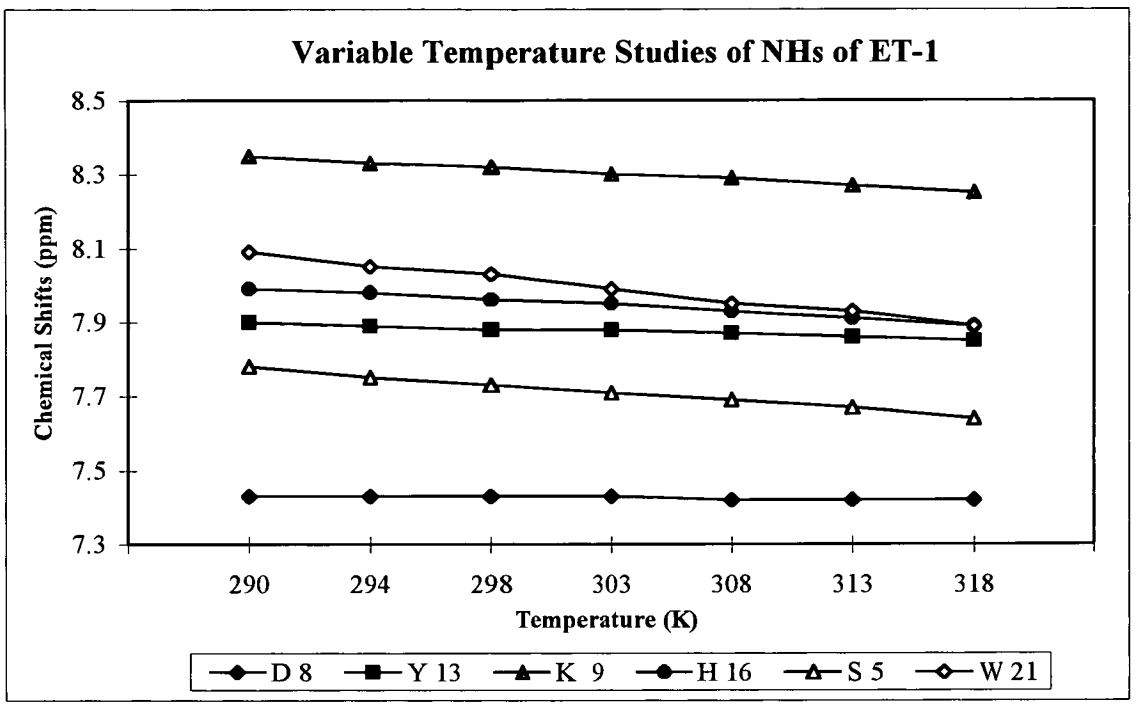


Figure 4.11 : Comparison of experimental ET-1 NH and αH chemical shift variations with documented literature

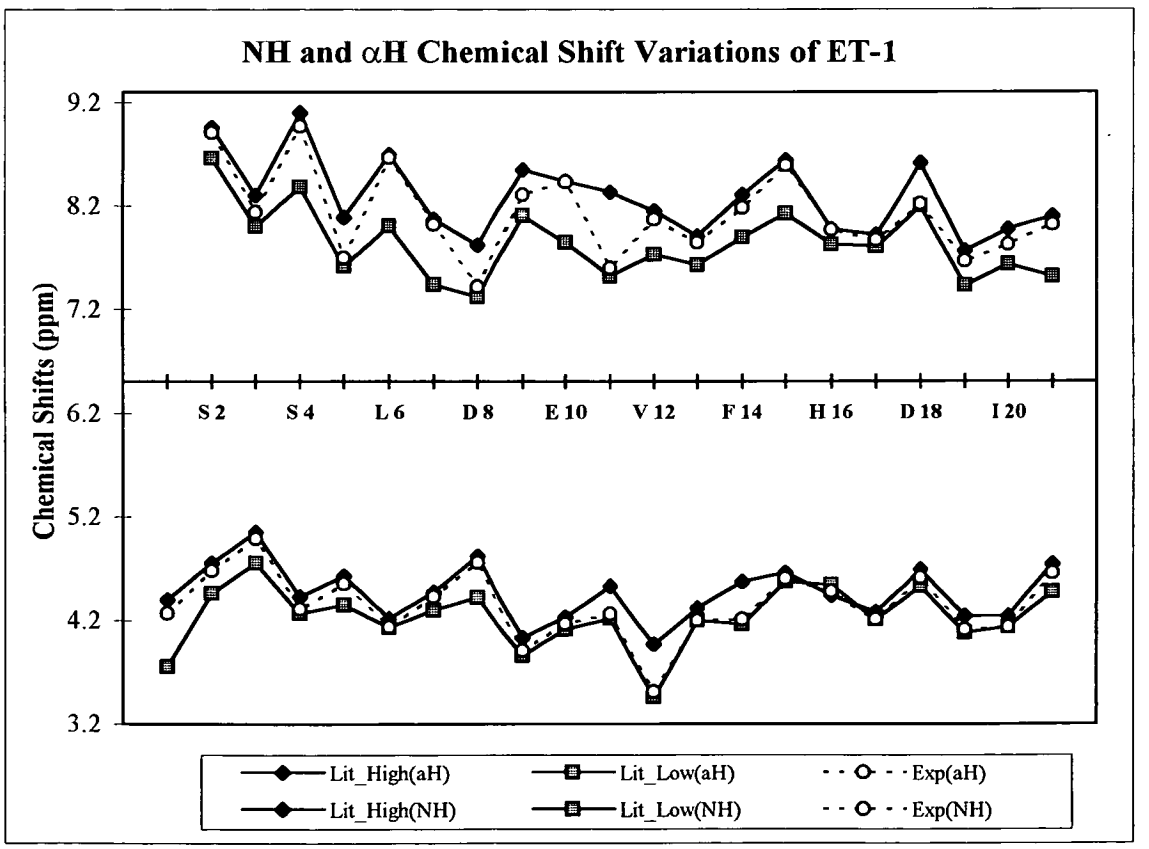
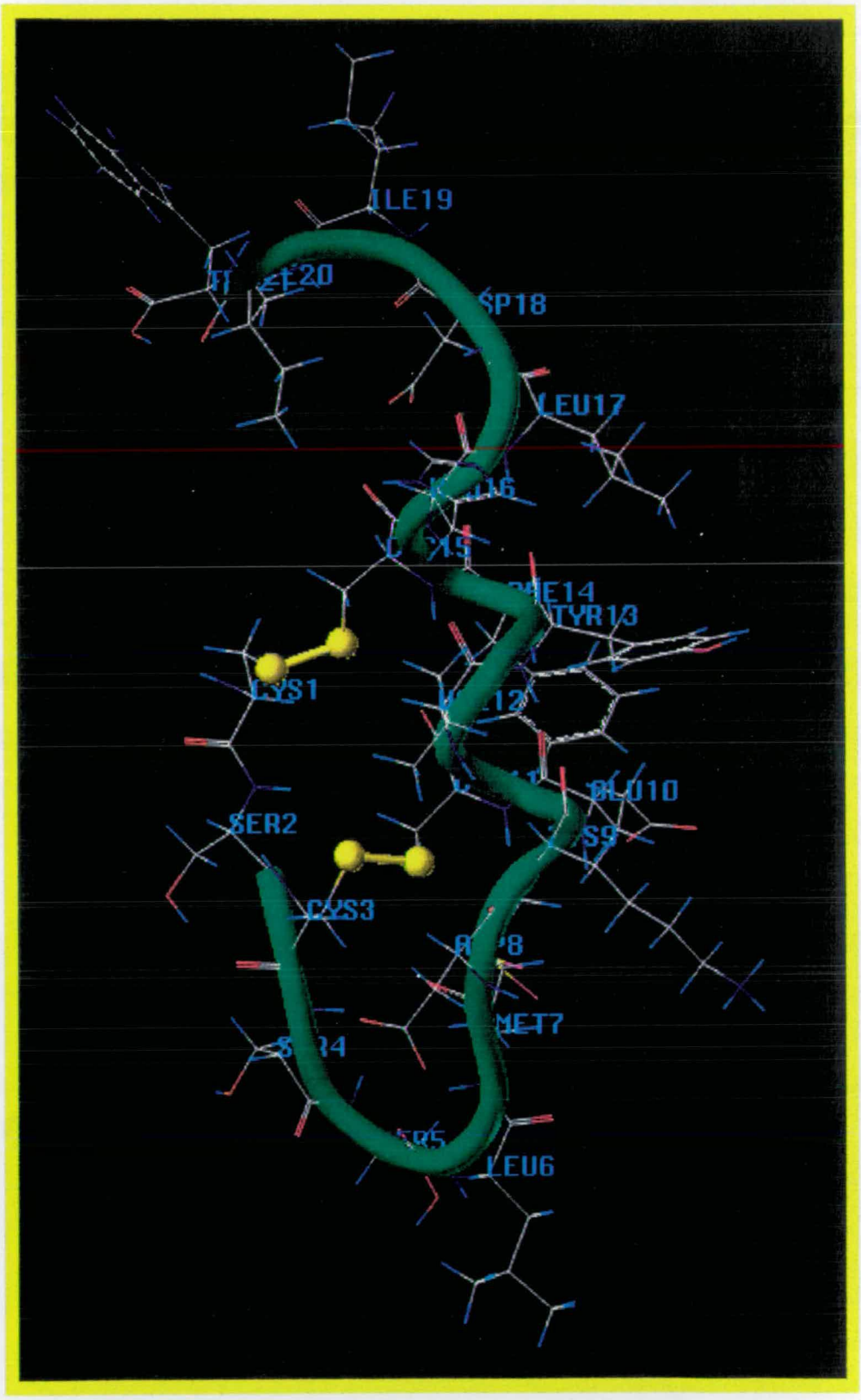


Figure 4.12 : Solution structure of ET-1. Two disulphide bridges are shown in ball and stick presentation



The results obtained from CD_3OH/H_2O solution data for ET-1 are in good agreement with those published by other research groups. Molecular modelling of ET-1 suggests that the disulphide constraints (1-15 and 3-11) are insufficient to restrict the formation of helical conformation in the middle of the sequence. It is interesting to see that the two disulphide bridges are on one side of the helix, while the other side of the helix forms a contiguous area containing most of the biologically active residues. Figure 4.12 shows one of the calculated structures (see section 4.3.6) of ET-1. Although some research groups (Table 4.6) reported that the C-terminal is associated with the bicyclic core of the molecule, no indication of this was observed in the present studies.

The 3D structure through the C^1-S^4 segment of the polypeptide is ill-defined from the NOE and molecular modelling data. The N-terminal portion of the molecule is probably in rapid equilibrium among various conformational states. Anderson et al.[174] have also reported that the C-terminus is not the most mobile portion of the structure but motional averaging is in fact more evident in an N-terminal region.

The slow deuterium exchange of the NH protons of both I^{19} and I^{20} with higher coupling constants confirmed the ordered structure of the C-terminal region. Important long-range interactions between the C-terminal part of the molecule and the helix were also detected. In particular, the $V^{12}\alpha H/L^{17}\beta H$, Y^{13}aromatics/$L^{17}\delta CH_3$ and F^{14}aromatics/$L^{17}\delta CH_3$ showed intense cross-peaks. This results are in good agreements with the recently published solution structure of ET-1 in water[179] in which the C-terminal of the peptide folds back towards the helical segment.

It can be seen that the local RMSD values (Table 4.7) for the backbone conformation between K^9-H^{16} show small variations while the N- and C- termini show larger variations. Small global RMSD variations were also observed for helical residues. The structure within the C-terminus and its orientation to the bicyclic core were completely indeterminant in all the studies, including our studies, except Saudek[166,168] and Bortmann.[181] The present evidence for a helical structure (K^9-H^{16}) and a turn at S^5-D^8 was noted in few cases (Table 4.6).

The structures obtained for ET-1 are in good agreement with the experimentally derived constraints. The structural statistics for the family of structures of ET-1 are given in Table 4.8.

Table 4.5: Chemical shifts of ET-1 obtained in the 50:45:5 = $CD_3OH:H_2O:D_2O$

Residue	NH	αH/ αCH_3	βH/ βCH_3	γH/ γCH_3	δCH_2/ δCH_3	ξCH_2/ ξCH_3	Aromatics & Others
1 C		4.29	3.29, 3.29				
2 S	8.91	4.69	3.81, 3.81				
3 C	8.15	4.99	2.59, 3.23				
4 S	8.97	4.32	3.87, 3.96				
5 S	7.71	4.55	3.66, 3.96				
6 L	8.68	4.14	1.62, 1.62	1.62	0.85, 0.92		
7 M	8.01	4.44	1.91, 2.17	2.45, 2.61			
8 D	7.43	4.76	2.80, 3.24				
9 K	8.32	3.91	1.57, 1.85	1.48	1.69	2.98	$+NH_3$ - 7.76 - 7.79
10 E	8.44	4.17	2.15, 2.15	2.48, 2.48			
11 C	7.62	4.26	3.18, 3.18				
12 V	8.08	3.53	2.04	0.84, 0.99			
13 Y	7.87	4.20	2.96, 2.96				2/6 - 6.80, 3/5 - 6.62
14 F	8.19	4.21	3.13, 3.18				2/6 - 7.30, 3/5 - 7.35
							4H - 7.32
15 C	8.60	4.60	3.03, 3.27				
16 H	7.97	4.48	3.30, 3.34				2H - 8.57, 4H - 7.28
17 L	7.89	4.23	1.63, 1.63	1.52	0.79, 0.79		
18 D	8.23	4.61	2.70, 2.86				
19 I	7.68	4.12	1.73	0.75	0.64		
				1.03, 1.35			
20 I	7.84	4.15	1.77	0.79	0.79		
				1.07, 1.39			
21 W	8.03	4.65	3.17, 3.29				2H - 7.16, 4H - 7.55
							5H - 7.03, 6H - 7.11
							7H - 7.36, NH - 10.14

Table 4.6 : Summary of the solution conformations of the Endothelin and Endothelin-like peptides derived from previous NMR studies

Peptide	Helical Region	N-terminus	C-terminus	Solvent	References
ET-1	undefined	undefined	associated with bicyclic core	d_6-DMSO	166
ET-1	K^9-C^{15}	structured	undefined	d_6-DMSO	167
ET-1	C^{11}-C^{15} opposite handed	undefined	associated with bicyclic core	d_6-DMSO/TFE	168
[NIe^7]-ET-1	K^9-H^{16}	S^5-D^8 (quasi turn)	-	d_6-DMSO and 50% CD_3CN/H_2O	169
ET-1	L^6-C^{11}	undefined	undefined	d_6-DMSO	170
ET-1	K^9-L^{17}	S^5-D^8	undefined	ethylene glycol/H_2O (60:40) with TFA	171
ET-1	K^9-C^{15}	S^5-D^8 (β-bend)	undefined	10% d_4-AcOH/ H_2O	172
ET-1	K^9-C^{15}	S^5-D^8	undefined	30% CD_3CN/H_2O	173
ET-1	K^9-F^{14}/C^{15}	S^5-D^8 (reverse turn)	averaging	ethylene glycol/H_2O (60:40) with TFA	174
[1,15Aba]-ET-1	K^9-H^{16}	S^5-D^8 (β-turn)	structured	10% CD_3CN/H_2O with 1.5% AcOH	175
ET-1	K^9-C^{15}	S^5-D^8 (β-turn)	undefined	10% d_3-AcOH/H_2O	176
ET-1	K^9-H^{16}	S^5-K^9 (β-turn)	unstructured	d_6-DMSO	177
ET-1	K^9-H^{16}	S^5-D^8 (β-turn)	undefined	d_3-AcOH/H_2O	178
[Ala^7]ET-1	K^9-H^{16}	S^5-D^8 (β-turn)	undefined	d_3-AcOH/H_2O	178
[Ala^8]ET-1	K^9-H^{16}	S^5-A^8 (β-turn)	undefined	d_3-AcOH/H_2O	178
ET-1	K^9-H^{16}	S^5-K^9 (β-turn)	folds back towards the helix	H_2O	179
ET-3	K^9-C^{15}	-	opposed to the bicyclic core	H_2O	180
ET-3	K^9-C^{15} opposite handed	β strand	associated with bicyclic core	d_3-AcOH/H_2O	181
Big-ET-1	K^9-C^{15}	S^5-D^8 (β-turn)	undefined	H_2O	182
Apamin	A^9-Q^{17}	N^2-A^5 (β-turn)	-	H_2O	183
Apamin	P^6-Q^{16}	N^2-A^5 (reverse turn)	-	H_2O	184
Apamin	C^3-A^5/P^6	A^9-H^{18} (reverse turn)	-	H_2O	185
SRTX-6b	D^8-C^{15}	C^3-M^6 (β-turn)	undefined	CD_3CN/H_2O	186
SRTX-6b	D^8-H^{16}	C^3-M^6 (β-turn)	undefined	CD_3CN/H_2O	187
SRTX-6b	K^9-Q^{17}	undefined	undefined	H_2O	188
SRTX-6c	E^9-C^{15}	undefined	undefined	H_2O	189

It should be noted that whatever medium is used, the ET-1 molecule seems to adopt roughly a similar structure. All the results suggest that the endothelin assumes an ordered conformation in solution.

The crystal structure of ET-1[190] has been described as having an N-terminal extended β-strand with a bulge between residues 5 and 7 followed by a hydrogen bonded loop between the carbonyl oxygen of residue 7 and the amide proton of residue 11. The residues 12-15 of endothelin form an irregular α-helix while C-terminal tail (residues 16-21) forms an ordered α-helical structure. In addition, the crystal structure of the tail portion is neither more flexible nor more disordered than the globular head region (residues 1-15). Thus the X-ray and NMR structures do have common features but also differ significantly in detail, especially the C-terminal residues (16-21). There is little basis for assuming that either the solution structure or the crystal structure would bear a close resemblance to that observed in receptor associated conformation.

Table 4.7: RMSD values obtained from DIANA calculations.
Three residues were used as a segment length and both termini residues, Cys-1 and Trp-21, were not included for calculation.

Residue	**Local RMSD/Segment**		**Global RMSD/Residue**	
	Backbone	Heavy atoms	Backbone	Heavy atoms
1	0.00	0.00	4.39	4.19
2	0.85	1.63	3.31	3.86
3	0.83	1.62	1.93	2.48
4	0.64	1.44	3.26	4.12
5	0.57	1.38	3.60	4.20
6	0.56	1.86	3.67	5.50
7	0.87	2.05	2.05	4.31
8	0.78	2.66	2.69	4.87
9	0.60	2.08	1.90	3.21
10	0.21	1.61	1.64	2.85
11	0.06	1.21	1.69	2.00
12	0.02	0.76	1.49	1.81
13	0.01	0.84	1.10	2.12
14	0.01	0.86	1.37	2.62
15	0.01	0.59	1.79	2.09
16	0.01	0.05	1.80	2.68
17	0.39	1.21	1.59	1.92
18	0.82	1.91	1.94	3.38
19	0.57	1.72	2.60	4.31
20	0.35	2.34	5.08	5.99
21	0.00	0.00	8.28	10.01

Table 4.8 : Structural statistics for the family of structures of ET-1

Number of DIANA calculated structures	300
Number of structures obtained from DIANA calculation	15
Number of selected structures for final calculation	10

	DIANA structures	FINAL structures
Number of NOE constraints used	230	131
Number of NOE violations > 0.2 A°	25	8
Violations of NOE constraints > 0.5 A°	10	3

	DIANA > 0.2 A°		FINAL structures	
Constraint violations	Mean	SD	Mean	SD
Intra residue constraints	0.409	0.112	0.043	0.0021
Sequential constraints	0.716	0.218	0.073	0.0042
Medium and long range constraints	0.433	0.213	0.055	0.0032
Total NOE constraints	0.519	0.184	0.058	0.0037
Hydrogen bond constraints	0.340	0.127	0.058	0.0053
Disulphide constraints	0.423	0.163	0.065	0.0021
Van der Waal violations	0.286	0.083	0.044	0.0023
Torsional angle constraints > 5 deg	34.60		16.300	

	FINAL structures	
Atomic rms differences (A°)	Mean	SD
Total residues (1-21)	3.847	0.798
Backbone (1-21)	2.601	0.453
Total helical region (9-16)	1.781	0.425
Helical backbone (9-16)	0.672	0.151
Helical heavy atoms (9-16)	1.410	0.217
Total turn (5-8)	1.386	0.322
Turn backbone (5-8)	1.418	0.245
N-terminus (1-4)	1.150	0.274
C-terminus (17-21)	2.953	0.660

Energy (k cals/ mol)	Mean	SD
Bond stretching	7.595	0.451
Angle bending	65.326	3.751
Torsional	38.736	3.247
Out of plane bending	0.673	0.161
1-4 van der Waals	20.964	0.968
van der Waals	-90.869	5.323
Total	42.425	8.758

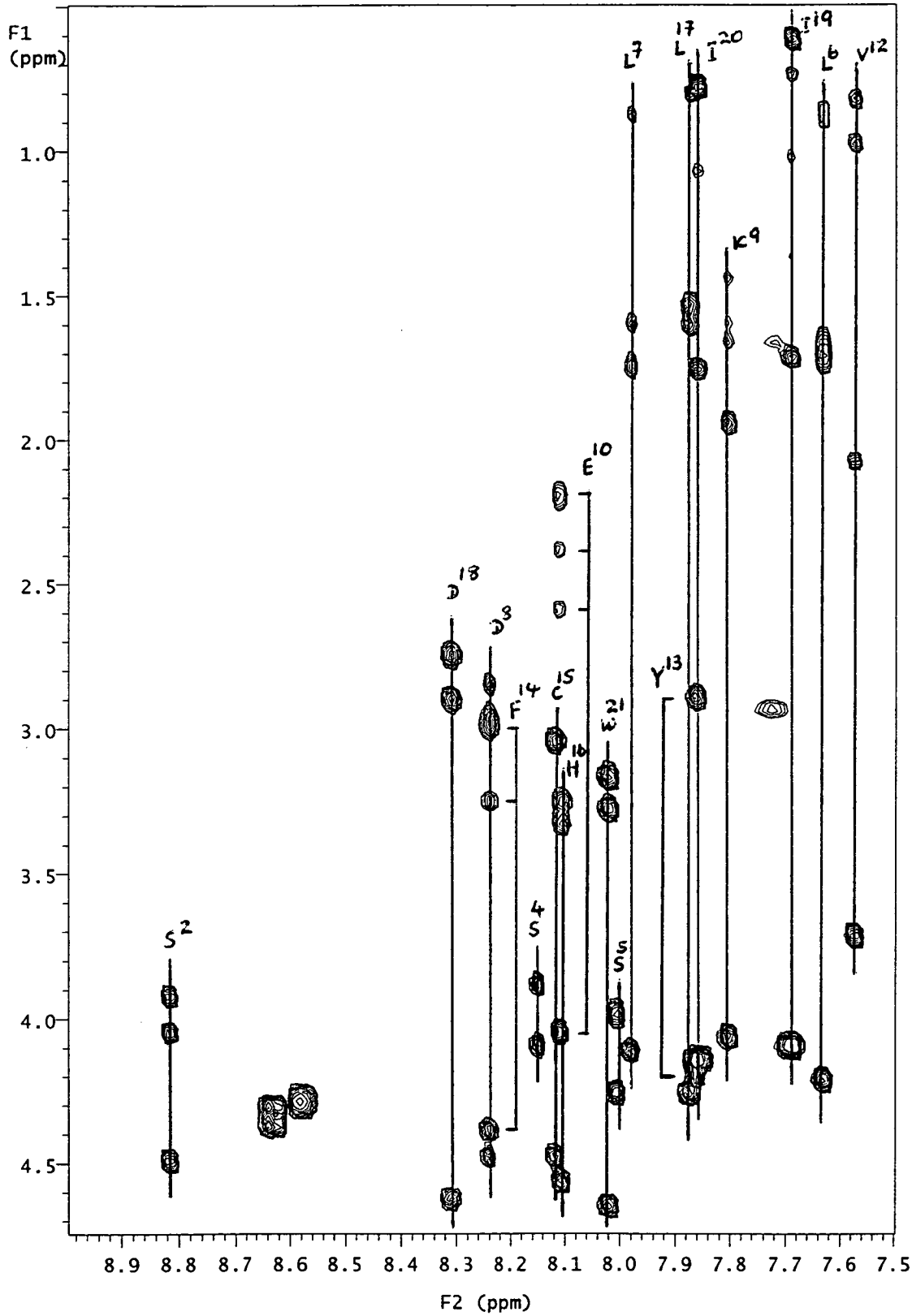
4.2.1.6 Biological Assay of Endothelin-1

Characteristic of the ET-1 is the conserved two disulphide bridges and the hydrophobic C-terminal amino acid residues, which are considered to be important for the expression of its biological activities.[191]

Kimura et al.[191] reported that the vasoconstrictor activity of ET is considerably decreased by the removal of the C-terminal W^{21} residue. In the models obtained from this study, a biologically active W^{21} residue which is located at the end of the C-terminal tail is isolated and separated from all the other residues. It is unlikely that the removal of W^{21} causes a considerable conformational change over the molecule leading to loss in activity. Therefore rather W^{21} itself would be recognised by the receptor(s).

It is well known[191,192] that the two disulphide bridges are required for receptor binding and vasoconstrictor activity. Synthetic truncated analogues of ET such as porcine ET1-15 or ET16-21 do not show vasoconstrictor activities[193] nor receptor binding activities.[194] These observations suggest that the receptor(s) for ET-1 recognise(s) an active conformation consisting both the helical core and the tail region. The results obtained from this study did not show the interaction between the core and the tail portions in solution. It seems likely that the receptor bound conformation of ET-1 is different from that in solution.

Figure 4.13 : Part of the TOCSY spectrum of LJP1 showing individual spin systems



4.2.2 The LJP1

4.2.2.1 Interpretation of NMR data of LJP1

The sequence of 21 amino acid building blocks in LJP1 includes 7 unique residues, Lys-9, Glu-10, Val-12, Tyr-13, Phe-14, His-16 and Trp-21. The primary structure of LJP1 is shown in Figure 4.1. The proton 1D, DQF COSY, TOCSY and NOESY spectra were acquired (298K) in mixture of aqueous methanol described before.

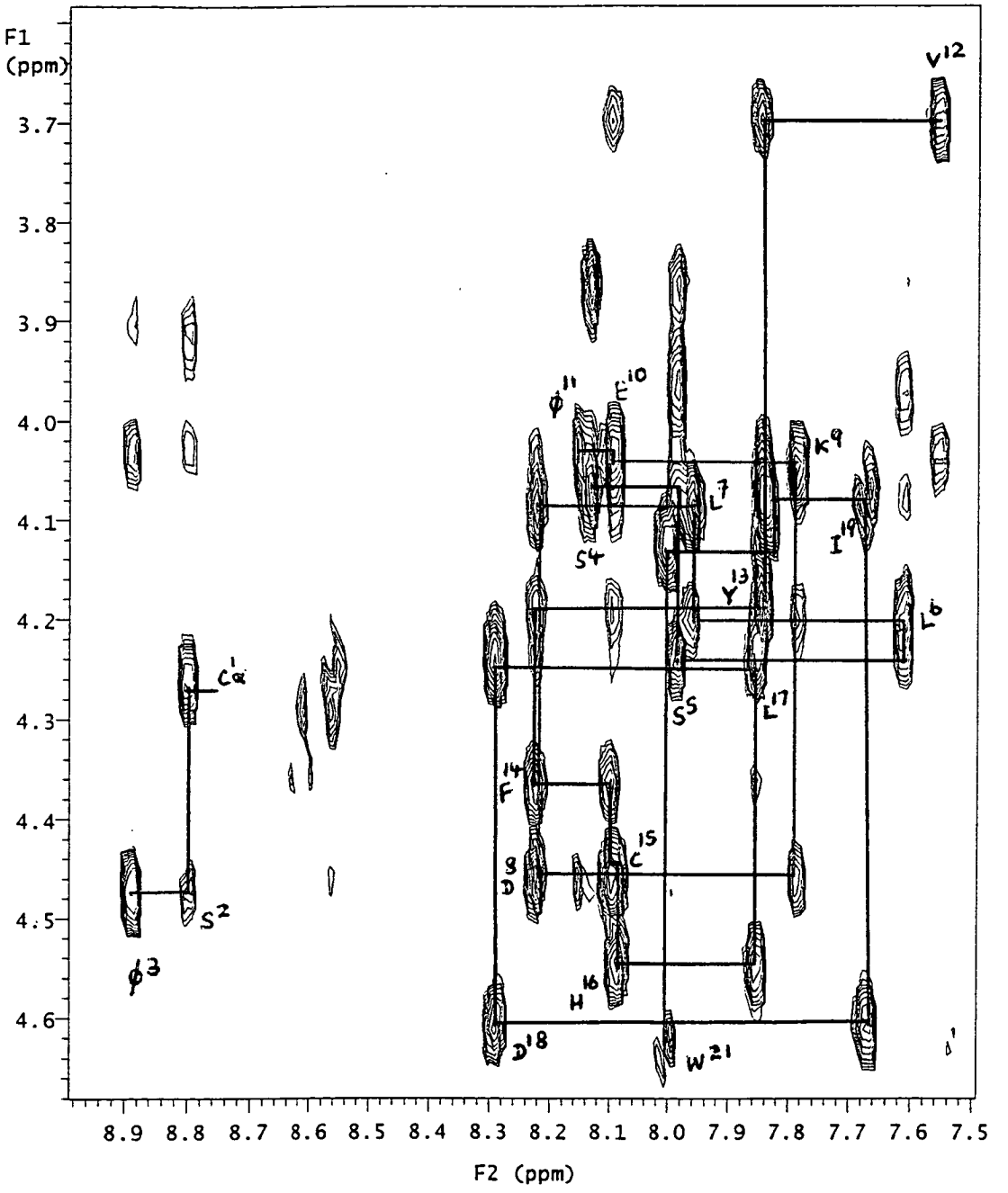
Each vertical line in the part of the TOCSY spectrum shown in Figure 4.13 showed all the signals of a particular spin system. It was possible to observe TOCSY cross peaks corresponding to magnetisation transfer from αH through the complete side chain to εH for the unique amino acid residue Lys-9. The εH resonance was clearly assigned on the basis of its distinctive chemical shift of 2.93 ppm and the broader resonance for the side chain $^{+}NH_3$ at 7.71-7.73.

The unique amino acid Glu-10 was clearly identified in the TOCSY spectrum and two γH resonances and a single resonance for degenerate βH protons were observed. The TOCSY spectrum showed complete spin system for the unique Val-12. Four unique aromatic amino acid residues, Tyr-13, Phe-14, His-16 and C-terminal Trp-21 could not be discriminated individually in the TOCSY spectrum due to other AMX spin patterns.

One of the three serine residues, later identified as Ser-2, was clearly identified on the basis of its distinctive chemical shift values of βH protons. The spin patterns for other two serine residues, Ser-4 and Ser-5, were discriminated from the remaining spin systems with the aid of the DQF COSY spectrum. Degenerate βH resonances were observed for both Ser-4 and Ser-5.

Three leucine residues, Leu-6, Leu-7 and Leu-17, were identified on the basis of their chemical shifts of δCH_3 resonances. Two isoleucine residues, Ile-19 and Ile-20, which showed complete spin patterns and distinctive methyl resonances were clearly identified from other amino acid residues. At this stage, two aspartic acid residues, Asp-8 and Asp-18, and two cysteine residues, N-terminal Cys-1 and Cys-15 could not be assigned due to complexity of AMX spin patterns.

Figure 4.14 : Fingerprint region of the NOESY spectrum of LJP1. Partial walk along the backbone is shown.



The AMX spin systems, aromatics and multiple copies of the same amino acids were carefully sorted out using the NOESY spectrum. The four aromatic residues were identified using the NOESY connectivities of βH's to their aromatic ring protons. The individual identification of spin patterns of aromatics were then determined by direct comparison of the TOCSY and NOESY spectra.

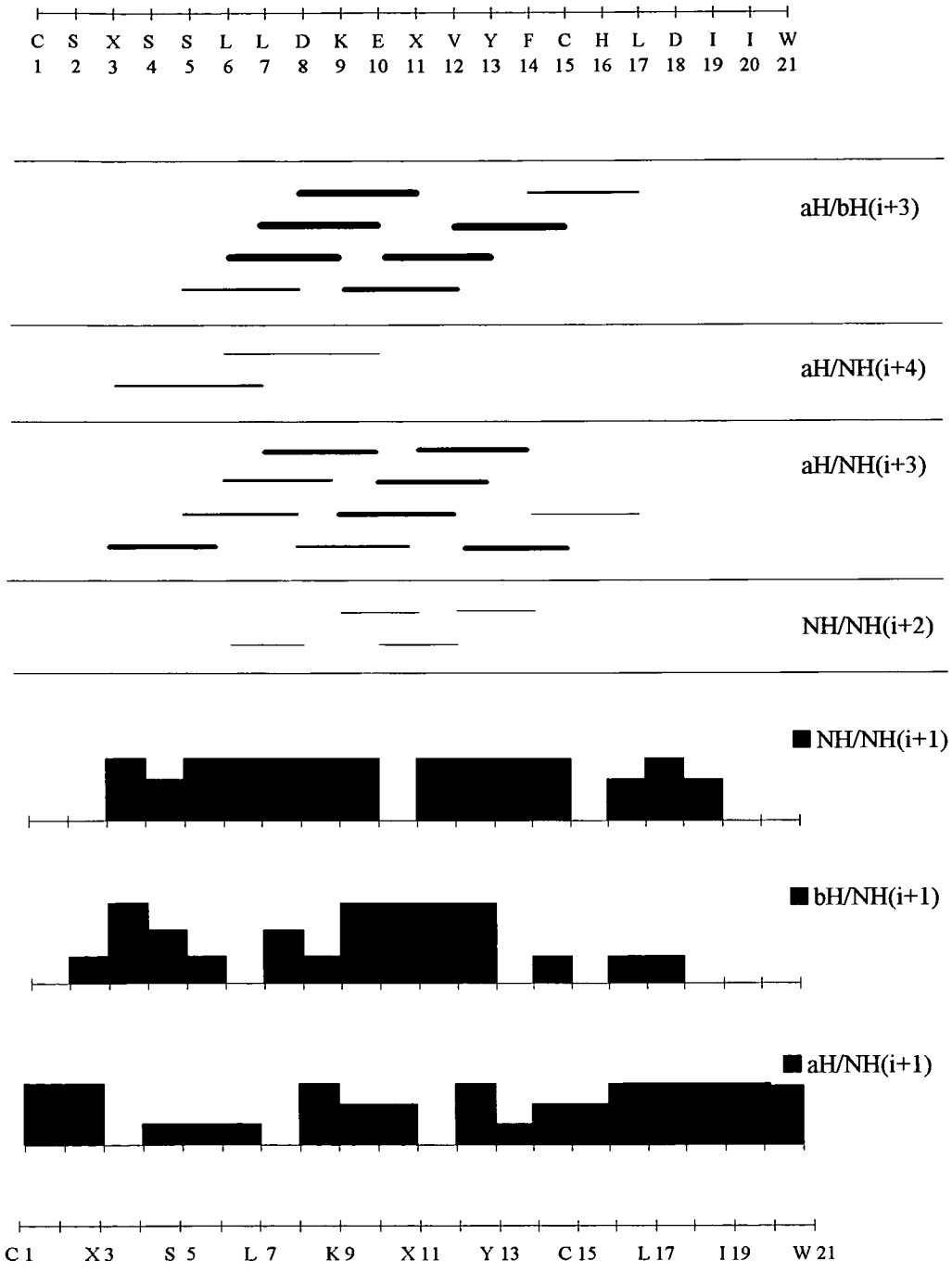
The prominent cross peak from the N-terminal $C^1\alpha H/S^2NH$ was clearly visible in the NOESY spectrum. The $S^2\alpha H/\phi^3NH$ cross peak is clear and the Aib-3 spin pattern is then identified. Since Lys-9 and Glu-10 are an unique residues, it is possible to identify the neighbours Asp-8 and Aib-11 thus resolving the Asp-18 ambiguity. The unique Trp-21 showed βH to aromatic proton 4H connectivity confirm the spin patterns of W^{21}. The cross peak $D^{18}\alpha H/I^{19}NH$ is clear and the sequence D^{18}-I^{19}-I^{20}-W^{21} is identified. Strong backbone connectivity of D^{18} confirmed the spin system of L^{17}. The backbone connectivity of H^{16}/L^{17} is clear and the above sequence extends back to H^{16}.

The backbone connectivity $\alpha_iH/N_{i+1}H$ is clearly visible for two adjacent leucine residues, Leu-6 and Leu-7, and the sequence L^6-L^7-D^8-K^9-E^{10}-ϕ^{11} is identified. Two adjacent serine residues, Ser-4 and Ser-5, which showed a α_iH/N_iH backbone connectivity extends the above sequence up to the Ser-4.

The unique Val-12 showed strong backbone connectivity to Tyr-13. The remaining aromatic residue Phe-14 and Cys-15 were then identified. The sequence-specific resonance assignments were completed and all the individual spin systems were then identified. The finger print region of the NOESY spectrum is shown in Figure 4.14.

Although two βH resonances were observed for each of Phe-14, His-16 and Trp-21, only a single βH resonance was observed for Tyr-13. A single βH resonance was also observed for Glu-10, Cys-15 and two serine residues, Ser-4 and Ser-5. The two γH resonances were observed for Ile-19 and in contrast a single γH resonance was observed for Ile-20. All the side chain spin connectivities through αH were clearly observed in the DQF COSY and TOCSY spectra for all the amino acid residues. Although weak

Figure 4.15 : Summary of interresidue NOESY connectivities observed for LJP1. The thickness of the columns and bars is propotional to the NOE intensity. (X = Aib)



cross peaks were observed for methyl resonances of isobutyric acids, Aib-3 and Aib-11, in the TOCSY spectrum, strong methyl resonances were appeared in the NOESY spectrum. The $\beta_iH/N_{i+1}H$ and $N_iH/N_{i+1}H$ connectivities of neighbouring amino acid residues supported the reliability of the resonance assignments. Figure 4.15 shows the summary of the inter residue NOESY connectivity patterns of LJP1.

4.2.2.2 Secondary Structure Details of LJP1

Evidence of the secondary structure was obtained from the long-range connectivities observed in the NOESY spectrum. The finger print region of the NOESY spectrum showed clear strong $d_{\alpha N}(i,i+n)$ connectivities for L^6/K^9, L^6/E^{10}, D^8/ϕ^{11}, K^9/V^{12}, and V^{12}/C^{15} and weak connectivities for S^5/D^8 and F^{14}/L^{17}. Some long-range connectivities, S^4/L^7, L^7/E^{10} and E^{10}/Y^{13} could not be discriminated due to overlapping cross peaks. The methyl resonances of two isobutyric acid residues, Aib-3 and Aib-11, also showed strong long range connectivities to L^6NH and F^{14}NH respectively. A weak long-range connectivity was observed for methyls of Aib-3 to Leu-7 amide proton. These $d_{\alpha N}(i,i+n)$ long-range cross peaks indicates the possibility of forming a α-helix between L^6-C^{15}.

The αH region of the NOESY spectrum showed strong $d_{\alpha\beta}(i,i+3)$ cross peaks confirming the α-helical segment in the middle of the sequence. Among them, S^5/D^8, L^6/K^9, L^7/E^{10}, D^8/ϕ^{11}, K^9/V^{12}, E^{10}/Y^{13} and V^{12}/C^{15} long-range connectivities are clearly indicated the formation of an α-helix. The strong backbone connectivities of the $d_{NN}(i,i+1)$ observed in the segment ϕ^3/C^{15} also confirmed the helical character of LJP1.

The LJP1 sample was subjected to the H-D exchange experiment and series of 1D proton spectra were acquired. The spectrum acquired soon after the preparation of the sample showed most of the NH protons except residues 1 to 3. The amide protons of S^4 and S^5 had clearly disappeared after the 12 minutes. The amide protons of N-terminal residues I^{19} and W^{21} were still observed even after 25min. The observation of amide protons in the helical region strongly supports the hydrogen bonded segment.

Figure 4.16 : H-D exchange profile of the amide protons of LJP1 as a function of time at 25^{0}C

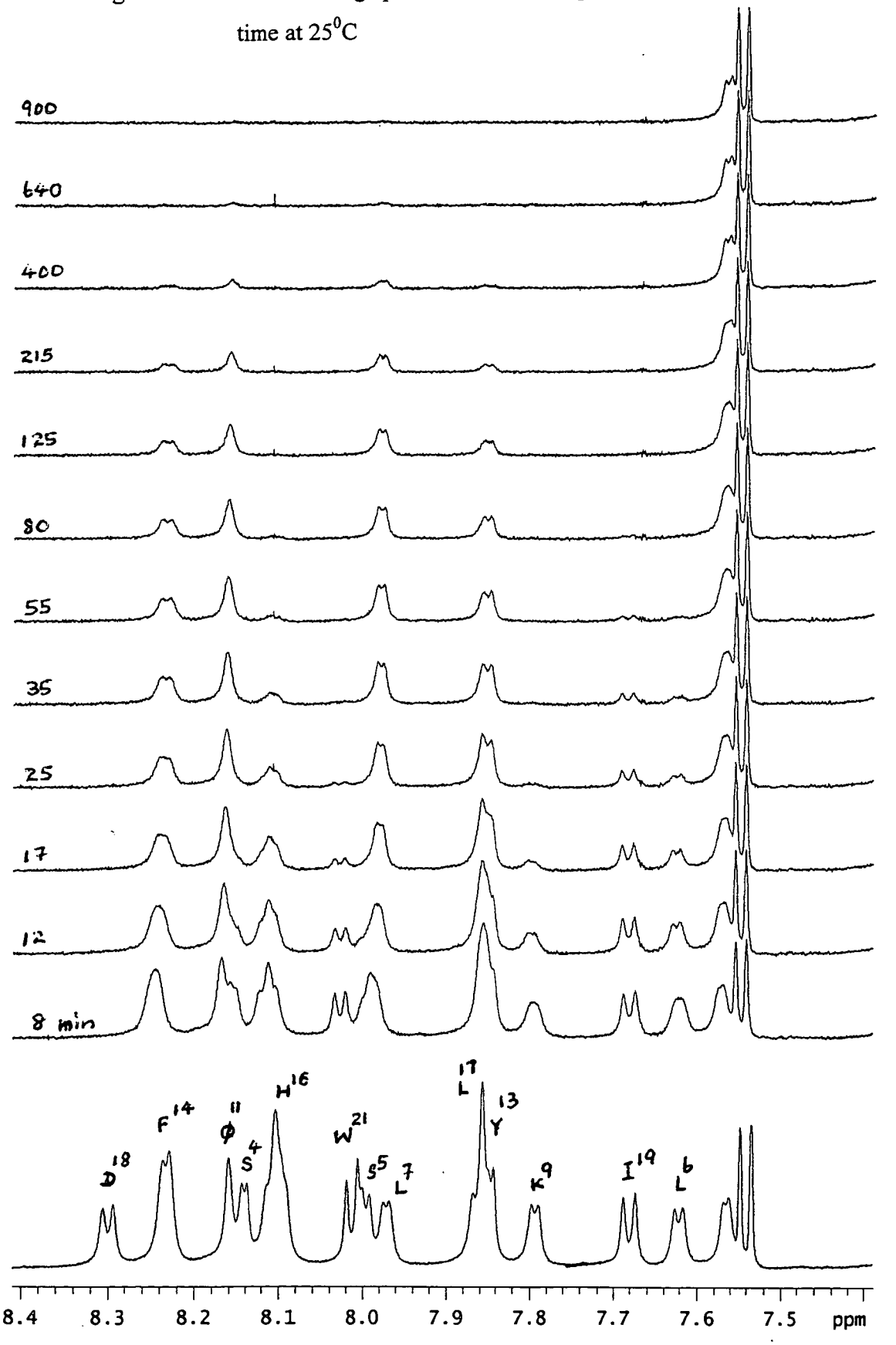


Figure 4.16 shows some of the 1D spectra of the H-D exchange experiment. All the observed amide protons except L^7 and ϕ^{11} were exchanged with D_2O after 6hrs and no protons were observed after 10hrs.

4.2.2.3 Solution Structure of LJP1

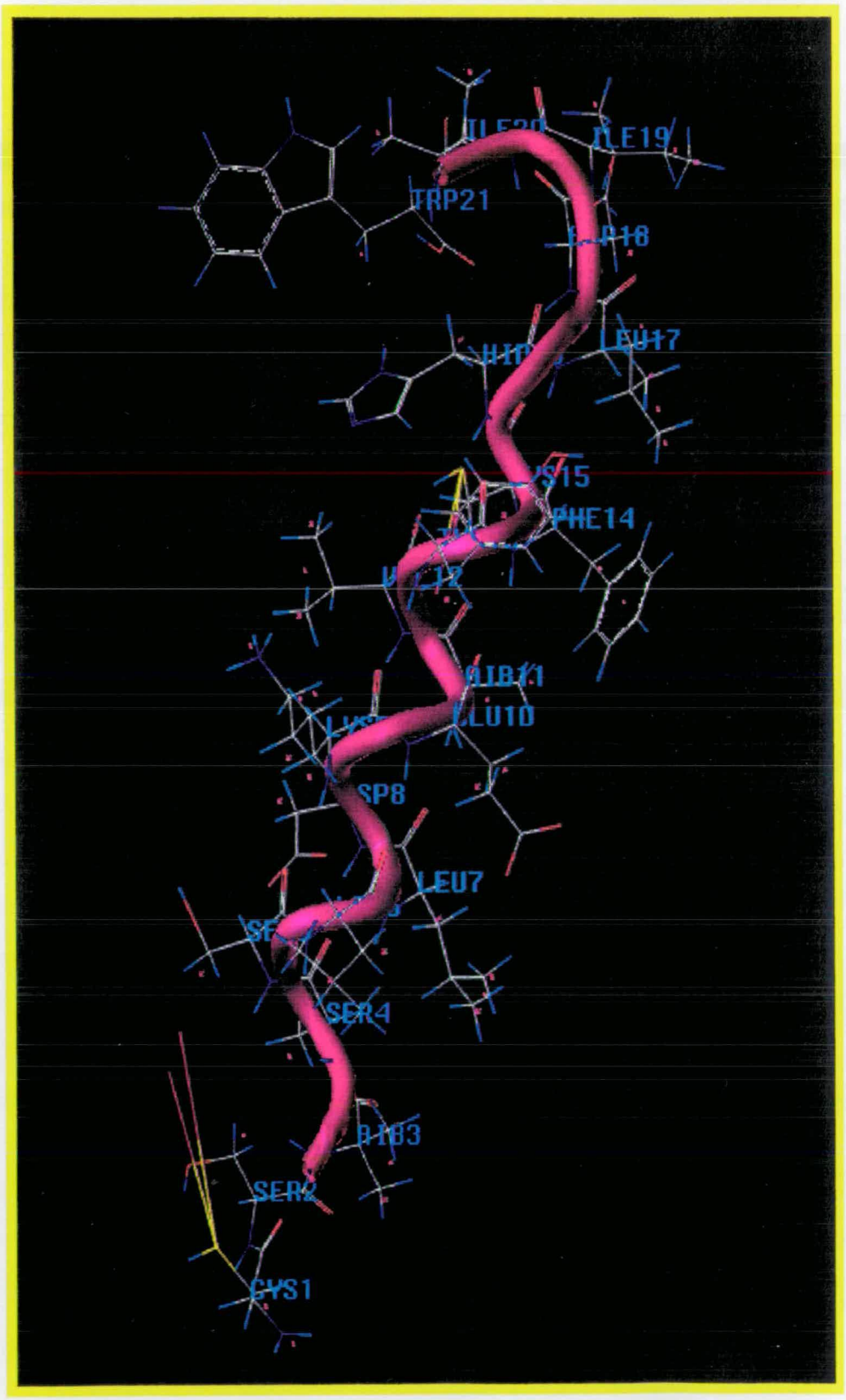
The set of 300 starting structures was calculated using DIANA and a set of 19 structures was finally obtained (see section 4.3.6). The root mean square deviation of the structures on the basis of backbone atoms, calculated as the average of all possible pairwise comparisons, was 2.1Å. Although the rms deviation of the global superimposition is high (2.8Å), the level of agreement obtained between structures is demonstrated if the LJP1 is considered as consisting of several segments. One of the calculated structures of LJP1 is shown in Figure 4.17. A full list of structural constraints (NOE, torsional angle, hydrogen bond) and $^3J_{NH\alpha}$ coupling constants may be found in Appendix II (pages 185-193).

The dominant structural features of LJP1 is a well-defined α-helix (Figure 4.17). The structure calculations reveal a highly ordered conformation in which a helical region extending from S^5 to H^{16}. Relaxed stereo-view of the ten backbone conformations of the helical region of LJP1 is shown in Figure 4.18. It is clearly seen that ϕ and ψ angles of residues S^5-H^{16} show fairly good convergence (Table 4.10). On the other hand ϕ and ψ angles of C-terminal residues are poorly converged. This is why the C-terminal tail did not show a defined orientation relative to the core portion.

The ordered side chain groups adopt a stable conformation as evidenced by numerous interresidue NOE constraints observed for this region. Figure 4.19 shows the side chain conformations of the helical region of LJP1. Although χ_1 dihedral angle is well defined for some residues (S^2, L^6, L^7, E^{10}, V^{12}, H^{16} and I^{20}), especially in the helical region, beyond χ_1 are poorly defined throughout the molecule (data not shown). This is because quite a few NOEs involve side chain protons and the side chain prochiral groups are not stereospecifically assigned. Similarly, the wide variance in ϕ and ψ at Leu-17 suggests that this residue acts as a hinge facilitating different orientations of the tail with respect to the rest of the molecule.

Figure 4.18 :Relaxed stereo-view of the backbone of the α-helix in LJP1; S^5-H^{16}

Figure 4.17 : Three dimensional solution structure of LJP1



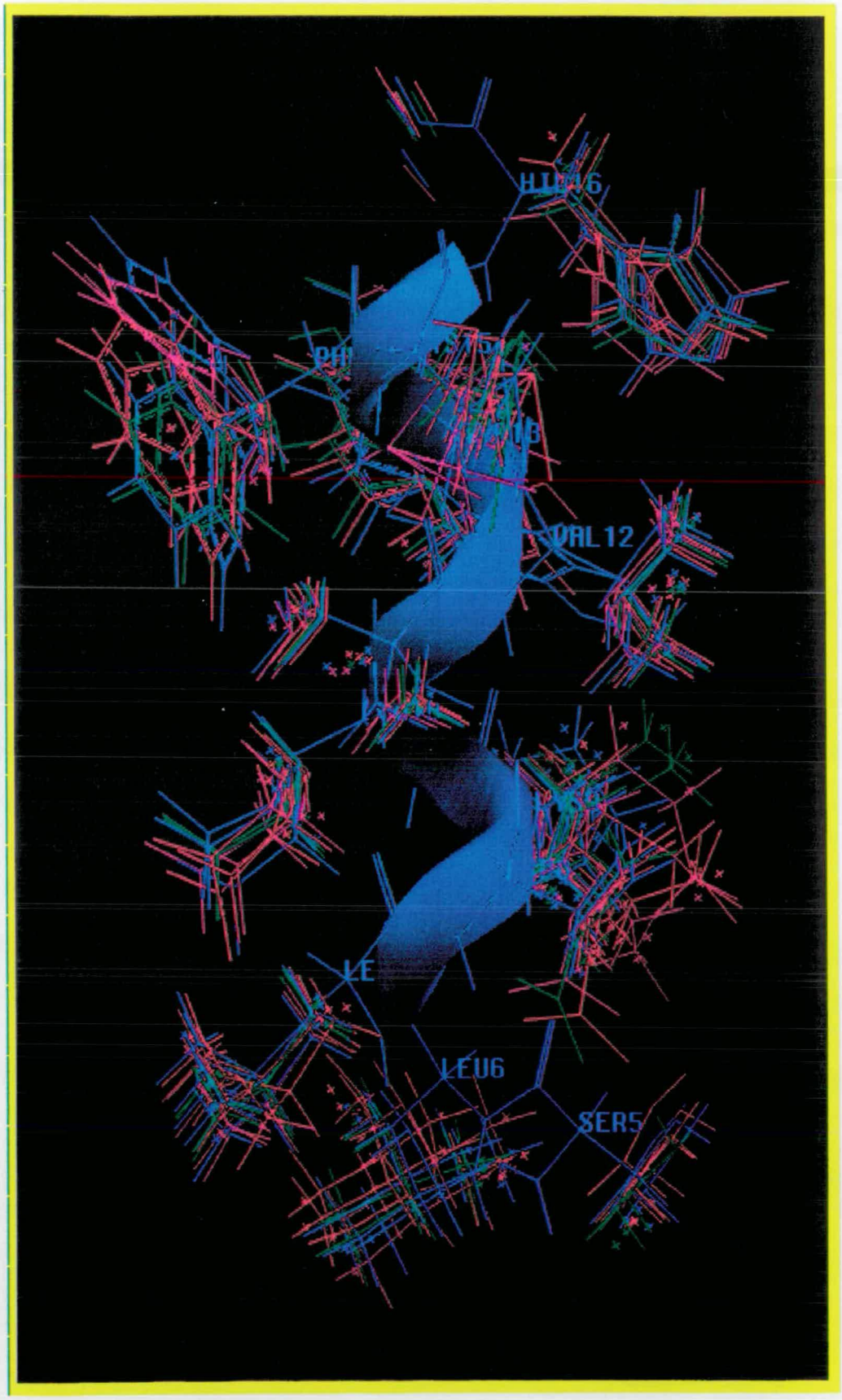
The structures obtained for LJP1 remain within one family of conformations. The ψ, ϕ and $\chi 1$ values of the averaged ten structures are summarised in Table 4.10.

The N-terminal tail region is the least well-defined portion of the family of calculated solution structures, possibly indicating that it is more conformationally flexible than the rest of the molecule. The C-terminal region (residues 17-21) of the calculated structures also shows no preferred conformation, as demonstrated by the significant variations of the ϕ and ψ torsion angles in this region.

Table 4.9: Chemical shifts of LJP1 obtained in the 50:45:5 = $CD_3OH:H_2O:D_2O$

Residue	NH	αH/ αCH_3	βH/ βCH_3	γH/ γCH_3	δCH_2/ δCH_3	ξCH_2/ ξCH_3	Aromatics & Others
1 C		4.28	3.00, 3.15				
2 S	8.81	4.49	3.93, 4.06				
3 ϕ	8.90	1.50, 1.50					
4 S	8.15	4.09	3.89, 3.89				
5 S	7.99	4.26	3.99, 3.99				
6 L	7.63	4.22	1.70, 1.73	1.60	0.86, 0.90		
7 L	7.97	4.12	1.73, 1.75,	1.60	0.87, 0.89		
8 D	8.23	4.48	2.85, 2.93				
9 K	7.80	4.07	1.59, 1.94	1.44	1.66	2.93	$+NH_3$ - 7.71 - 7.73
10 E	8.11	4.05	2.19, 2.19	2.38, 2.59			
11 ϕ	8.16	1.55, 1.55					
12 V	7.56	3.72	2.07	0.82, 0.98			
13 Y	7.86	4.21	2.90, 2.90				2/6 - 6.73, 3/5 - 6.61
14 F	8.23	4.39	3.00, 3.25				2/6 - 7.28, 3/5 - 7.34
							4H - 7.32
15 C	8.11	4.48	3.04, 3.04				
16 H	8.10	4.57	3.26, 3.33				2H - 8.58, 4H - 7.27
17 L	7.87	4.26	1.59, 1.61	1.53	0.80, 0.81		
18 D	8.30	4.63	2.75, 2.91				
19 I	7.68	4.10	1.71	0.74	0.62		
				1.03, 1.36			
20 I	7.86	4.15	1.76	0.79	0.79		
				1.08, 1.39			
21 W	8.02	4.64	3.17, 3.28				2H - 7.15, 4H - 7.54
							5H - 7.03, 6H - 7.10
							7H - 7.35, NH - 10.10

Figure 4.19 : Sidechain conformations of the helical region of LJP1



The side chain conformations of the helical region of LJP1 is shown in Figure 4.19. The structures obtained for LJP1 are in good agreement with the experimentally derived constraints and the structural statistics for the family of structures of LJP1 are given in Table 4.11.

The slow exchange of amide protons of I^{19} and I^{20} protons proves the poor accessibility of water in the C-terminal region. This suggests the presence of hydrogen bonding character which might be responsible for the C-terminal tail. Prior to this work however, the effect of disulphide replacements on the solution conformation of the resulting peptides relative to that of native ET-1 was largely unknown.

Table 4.10 : Average Dihedral Angles of Ten Conformations of LJP1

Residue	Phi	Psi	Chi 1
CYS 1	-14 ± 121	86 ± 35	-33 ± 182
SER 2	66 ± 4	162 ± 9	-48 ± 3
AIB 3	-138 ± 109	55 ± 3	
SER 4	-164 ± 27	-55 ± 3	35 ± 181
SER 5	-50 ± 6	-43 ± 2	-1 ± 187
LEU 6	-64 ± 4	-44 ± 2	-57 ± 2
LEU 7	-63 ± 2	-41 ± 4	-55 ± 3
ASP 8	-62 ± 2	-45 ± 2	2 ± 161
LYS 9	-56 ± 3	-44 ± 2	-70 ± 172
GLU 10	-62 ± 3	-54 ± 2	-68 ± 3
AIB 11	-50 ± 2	-49 ± 2	
VAL 12	-62 ± 6	-49 ± 3	170 ± 2
TYR 13	-52 ± 5	-51 ± 4	119 ± 116
PHE 14	-71 ± 7	-49 ± 8	-70 ± 99
CYS 15	-37 ± 12	-41 ± 7	-2 ± 161
HID 16	-57 ± 8	-50 ± 4	-58 ± 6
LEU 17	51 ± 9	33 ± 32	-66 ± 19
ASP 18	48 ± 23	-52 ± 93	18 ± 137
ILE 19	-58 ± 52	-28 ± 43	-24 ± 66
ILE 20	-72 ± 27	-50 ± 68	-59 ± 6
TRP 21	-109 ± 20	7 ± 107	-55 ± 89

Table 4.11 : Structural statistics for the family of structures of LJP1

Number of DIANA calculated structures	300
Number of structures obtained from DIANA calculation	19
Number of selected structures for final calculation	10

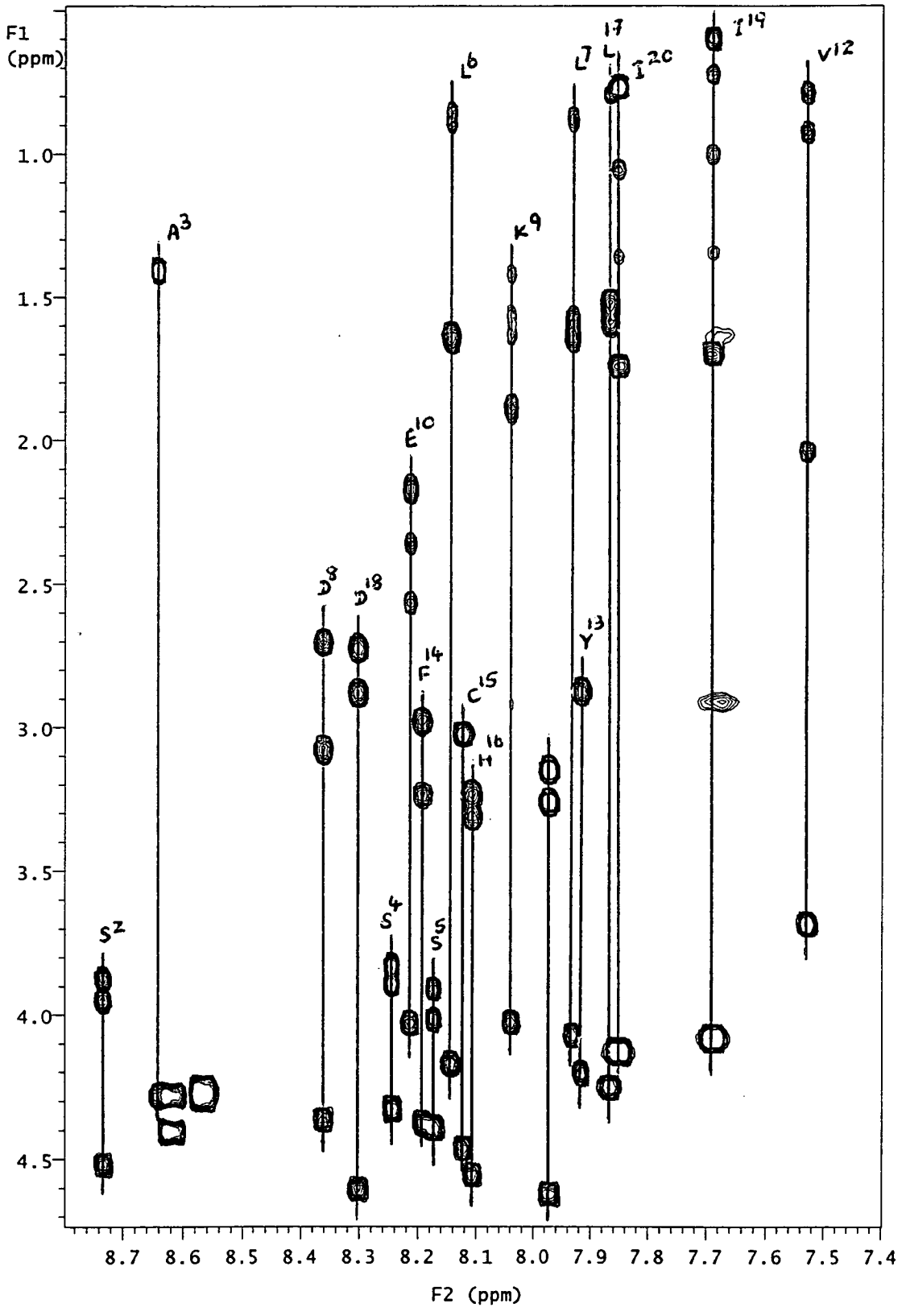
	DIANA structures	FINAL structures
Number of NOE constraints used	278	173
Number of NOE violations > 0.2 A°	45	8
Violations of NOE constraints > 0.5 A°	11	3

	DIANA > 0.2 A°		FINAL structures	
Constraint violations	Mean	SD	Mean	SD
Intra residue constraints	0.339	0.105	0.041	0.0011
Sequential constraints	0.439	0.140	0.037	0.0032
Medium and long range constraints	0.476	0.229	0.032	0.0053
Total NOE constraints	0.418	0.162	0.021	0.0043
Hydrogen bond constraints	0.400	0.180	0.029	0.0031
Van der Waal violations	0.300	0.095	0.021	0.0024
Torsional angle constraints > 5 deg	11.470	6.725	6.657	3.3240

	FINAL structures	
Atomic rms differences (A°)	Mean	SD
Total residues (1-21)	2.559	0.718
Backbone (1-21)	1.585	0.515
Total helical region (5-16)	1.278	0.332
Helical backbone (5-16)	0.325	0.128
Helical side chain (5-16)	0.940	0.228
Helical heavy atoms (5-16)	1.526	0.257
N-terminus (1-4)	1.193	0.311
C-terminus (17-21)	2.510	0.825

Energy (k cals/ mol)	Mean	SD
Bond stretching	8.439	0.310
Angle bending	64.689	1.166
Torsional	34.386	2.180
Out of plane bending	0.510	0.128
1-4 van der Waals	23.589	0.753
van der Waals	-98.120	5.113
Total	34.492	2.974

Figure 4.20 : Part of the TOCSY spectrum of LJP26 showing individual spin systems



4.2.3 The LJP26

4.2.3.1 Interpretation of NMR data of LJP26

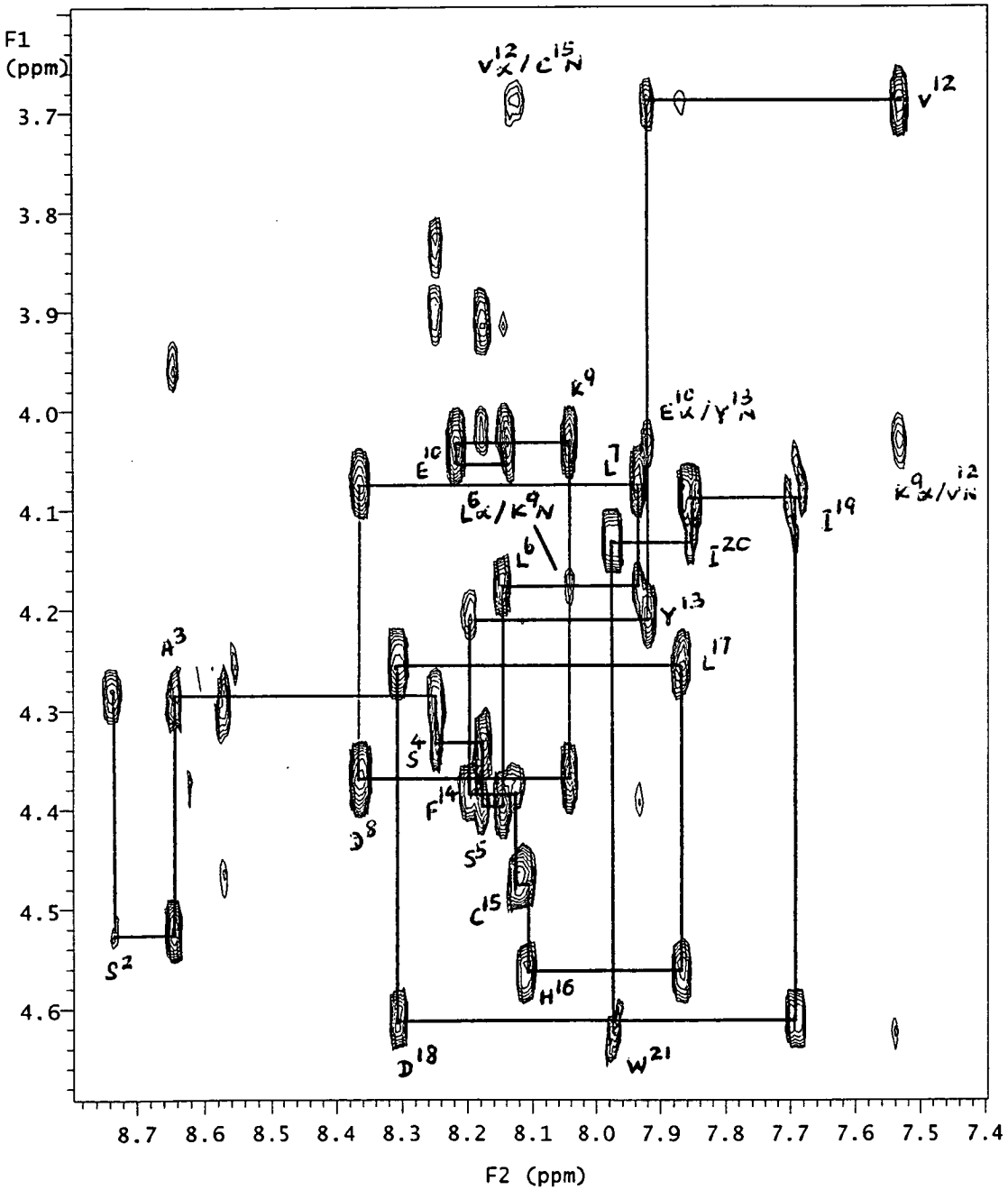
The amino acid sequence of 21 residue in LJP26 includes 11 unique residues, Ala-3, Asp^{d}-8, Lys-9, Glu-10, Aib-11, Val-12, Tyr-13, Phe-14, His-16, Asp-18 and Trp-21. The differences between LJP1 and LJP26 are third and eighth residues (Figure 4.4). The proton 1D, DQF COSY, TOCSY and NOESY spectra were acquired for LJP26. The NH region of the TOCSY spectrum (Figure 4.20) shows the resonances of individual spin systems. The backbone region showed 19 resonances for NH/αH and 2 extra resonances for side chain protecting groups. No NH/αH cross peak appeared for the N-terminal Cys-1. The magnetisation transfer from NH through the complete side chain protons were observed where possible for all amino acid residues in the TOCSY spectrum. The three serine residues were clearly identified on the basis of their distinctive chemical shifts of two βH resonances in the TOCSY spectrum.

The unique amino acid residues, Ala-3, Lys-9, Glu-10 and Val-12 were clearly identified on the basis of their individual signals that appeared in the TOCSY spectrum. The chemical shifts of methyl resonances were used to identify the two isoleucines and three leucines. The four aromatic residues, two cysteine residues and two aspartic acid residues could not be distinguished due to the similarity of the AMX spin system.

The DQF COSY spectrum also showed the spin connectivities for all the amino acid residues. Discrimination of the aromatic and non-aromatic residues which show the same AMX spin pattern was assigned by the βH to aromatic proton resonance connectivities appeared in the NOESY spectrum. The positions of the residues which occur more than once were determined by the direct comparison of the TOCSY and NOESY spectra. The unique amino acid residues were used as a starting point for the identification of the sequence-specific resonance assignments. The backbone finger print region of the NOESY spectrum is shown in Figure 4.21.

The NOESY spectrum showed cross peaks from the ring hydrogen resonances of Trp-21 to the αH and βH allowing the confirmation of unique C-terminal residue. The unique residue Val-12 to Tyr-13 thence to Phe-14 backbone connectivities are clear thus

Figure 4.21 : Fingerprint region of the NOESY spectrum of LJP26 showing backbone connectivities



distinguishing the remaining aromatic residue His-16. The backbone connectivities of H^{16}/L^{17}, L^{17}/D^{18} and D^{18}/I^{19} are also clear and this completes the C-terminal part of the sequence His-16 to Trp-21.

Since Ala-3 is unique it is possible to identify the neighbour Ser-4 and thence Ser-5. The backbone connectivity of C^{1}/S^{2} and ϕ^{2}/A^{3} are clearly visible in the finger print region on the NOESY spectrum. This completes the N-terminal part of the sequence Cys-1 to Ser-5.

The unique residue Lys-9 showed connectivities to both neighbouring residues, Asp-8 and Glu-10. The backbone connectivities of L^{7}/D^{8} and two adjacent leucine residues, Leu-6 and Leu-7, were distinguished clearly in the NOESY finger print region. Identification of these spin patterns clearly discriminated the remaining cysteine residue, Cys-15, in the finger print region.

4.2.3.2 Chemical Shift Index

The Chemical Shift Index technique (CSI)[195] can be used to obtain the secondary structural details using only the chemical shifts. This method is strongly depend on the character and nature of the protein secondary structure. Random coil chemical shift values and experimental shift values are categorised according to the procedure given below and +1, 0 or -1 indices are then grouped. Any grouping of four (not necessarily consecutive) or more -1s not interrupted by a +1 is a helix and any grouping of three or more +1s not interrupted by a -1 is a β-strand. The first appearance of opposite (+1/-1 or -1/+1) or two consecutive zero CSIs marks the termination point. Although the random coil shifts pertain to chemical shifts in aqueous solution were used in this study, residues between D^{8}-C^{15} clearly show the α-helical character (Figure 4.23).

random coil chemical shift value ± 0.1 > experimental value ⇒ -1

random coil chemical shift value ± 0.1 < experimental value ⇒ +1

random coil chemical shift value ± 0.1 = experimental value ⇒ 0

Figure 4.22 : Summary of interresidue NOESY connectivities observed for LJP26. The thickness of the columns and bars is propotional to the NOE intensity. (X = Aib)

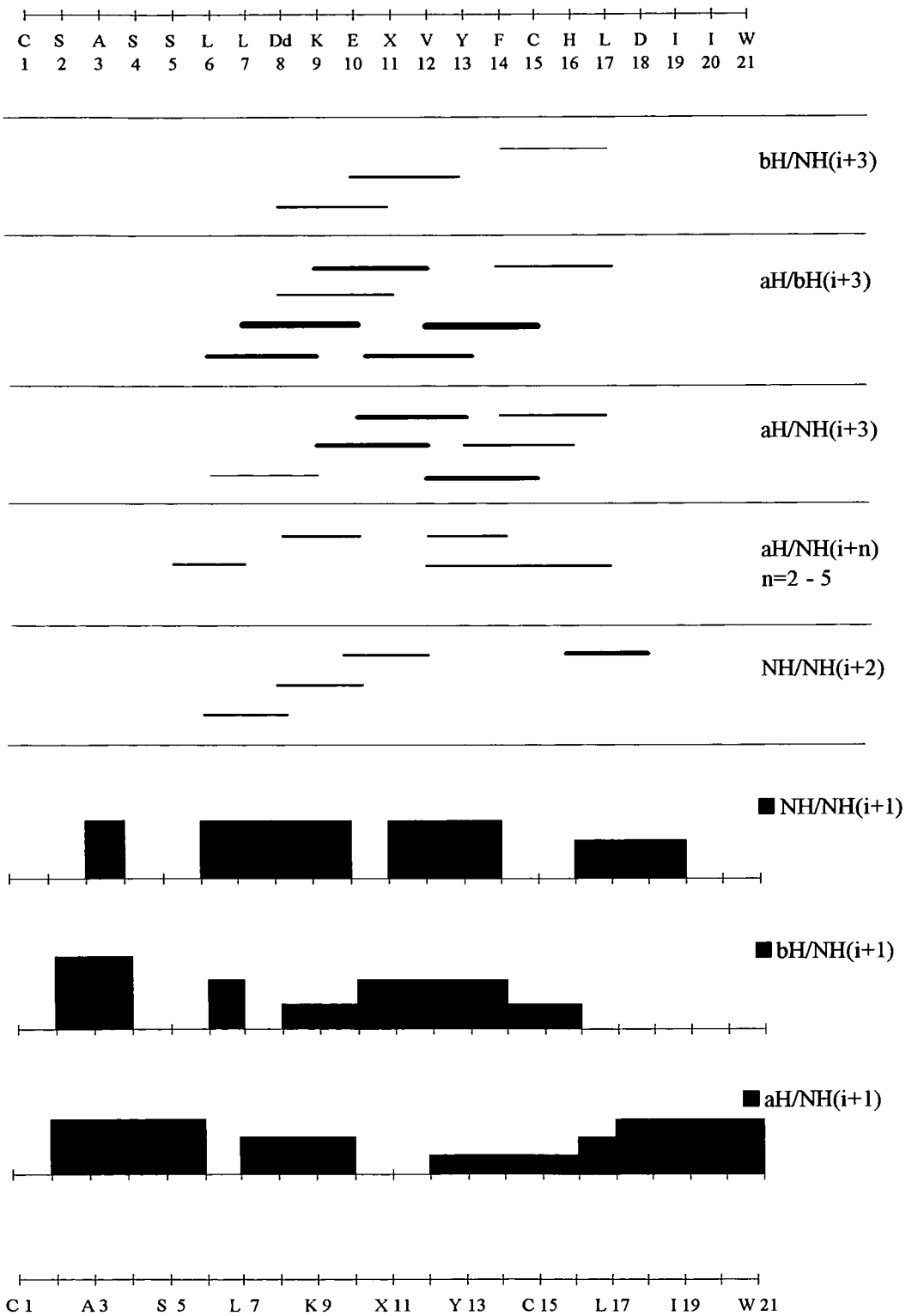
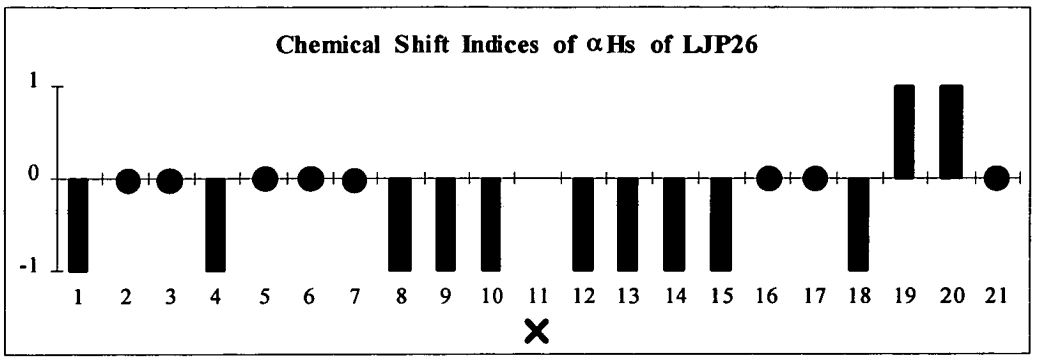


Figure 4.23: Chemical Shift Indices of αHs of LJP26; X-Aib



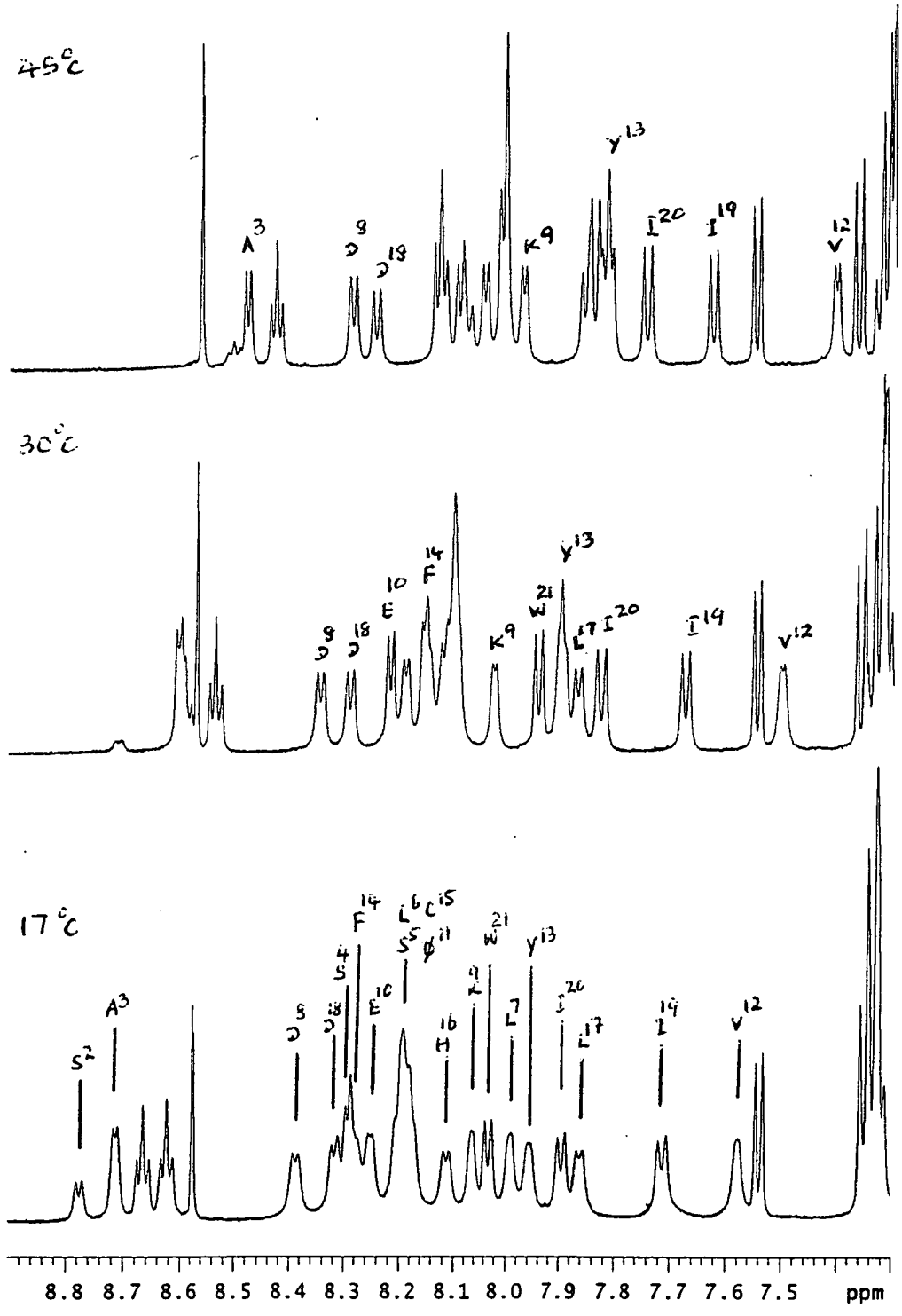
4.2.3.3 Secondary Structure Details of LJP26

After completing the assignments of sequence-specific resonances, the NOESY spectrum was searched for secondary structure correlations. The strong long range $d_{\alpha N}(i,i+3)$ connectivities of K^9/V^{12}, E^{10}/Y^{13} and V^{12}/C^{15} and a weak connectivity of L^6/K^9 was clearly identified in the finger print region of the NOESY spectrum. The L^7/E^{10}, D^8/ϕ^{11} and ϕ^{11}/F^{14} correlations could not be identified due to overlapping cross peaks.

The strong $N_iH/N_{i+1}H$ backbone connectivities between L^6-F^{14} and weak $\alpha_iH/N_{i+1}H$ connectivities confirmed the α-helical pattern between L^6 and C^{15}. The C-terminal amino acid residues L^{17}-I^{20} also showed strong $N_iH/N_{i+1}H$ backbone connectivities. The strong α_iH/β_{i+3} H connectivities, L^7/E^{10}, K^9/V^{12} and V^{12}/C^{15} appeared in the αH region of the NOESY spectrum also confirmed the α-helical structure. Figure 4.22 shows the summary of the interresidue NOESY connectivity patterns of LJP26.

The H-D exchange experiment was carried out and a series of 1D proton spectra were collected. The amide protons C^1-S^5 were exchanged soon after the preparation of the sample. Slowly exchanging amide protons were clearly observed in the middle of the sequence from K^9/H^{16}. The C-terminal residues I^{19}-W^{21} also showed slowly exchanging amide protons and completely disappeared after 30min except I^{19} (50min). All the protons were disappeared completely after 8hrs.

Figure 4.24 : NH Region of the 1D spectra of LJP26 acquired at different temperatures



Variable temperature studies of amide protons were also confirmed the shielding protons. Small temperature coefficients were observed for amino acid residues in the helical region from K^9 to H^{16} except V^{12}. In addition, NH of I^{19} also showed shielding properties. Figure 4.24 shows some of the 1D spectra of the NH region at different temperatures.

Table 4.12: Chemical Shifts of LJP26 obtained in the 50:45:5 = $CD_3OH:H_2O:D_2O$

Residue	NH	αH/ αCH_3	βH/ βCH_3	γH/ γCH_3	δCH_2/ δCH_3	ξCH_2/ ξCH_3	Aromatics & Others
1 C		4.29	3.16, 2.99				
2 S	8.73	4.53	3.89, 3.96				
3 A	8.64	4.29	1.40				
4 S	8.24	4.34	3.84, 3.91				
5 S	8.17	4.40	3.92, 4.03				
6 L	8.14	4.18	1.67, 1.67	1.65	0.86, 0.91		
7 L	7.93	4.09	1.65, 1.65	1.59	0.87, 0.91		
8 D^d	8.36	4.37	2.72, 3.09				
9 K	8.04	4.03	1.58, 1.90	1.44	1.65	2.92	$+NH_3$ - 7.66 - 7.69
10 E	8.21	4.04	2.17, 2.19	2.37, 2.58			
11 ϕ	8.13	1.53, 1.53					
12 V	7.53	3.70	2.06	0.80, 0.94			
13 Y	7.92	4.21	2.89, 2.89				2/6 - 6.72, 3/5 - 6.61
14 F	8.19	4.39	2.99, 3.25				2/6 - 7.31, 3/5 - 7.34
							4H - 7.33
15 C	8.12	4.48	3.04, 3.04				
16 H	8.11	4.57	3.25, 3.32				2H - 8.57, 4H - 7.32
17 L	7.87	4.26	1.60, 1.60	1.53	0.80, 0.80		
18 D	8.30	4.63	2.73, 2.89				
19 I	7.69	4.09	1.71	0.74	0.61		
				1.02, 1.36			
20 I	7.85	4.14	1.75	0.79	0.78		
				1.07, 1.37			
21 W	7.97	4.63	3.16, 3.27				2H - 7.15, 4H - 7.54
							5H - 7.03, 6H - 7.10
							7H - 7.35, NH - 10.08

Figure 4.25 : Three dimensional solution structure of LJP26

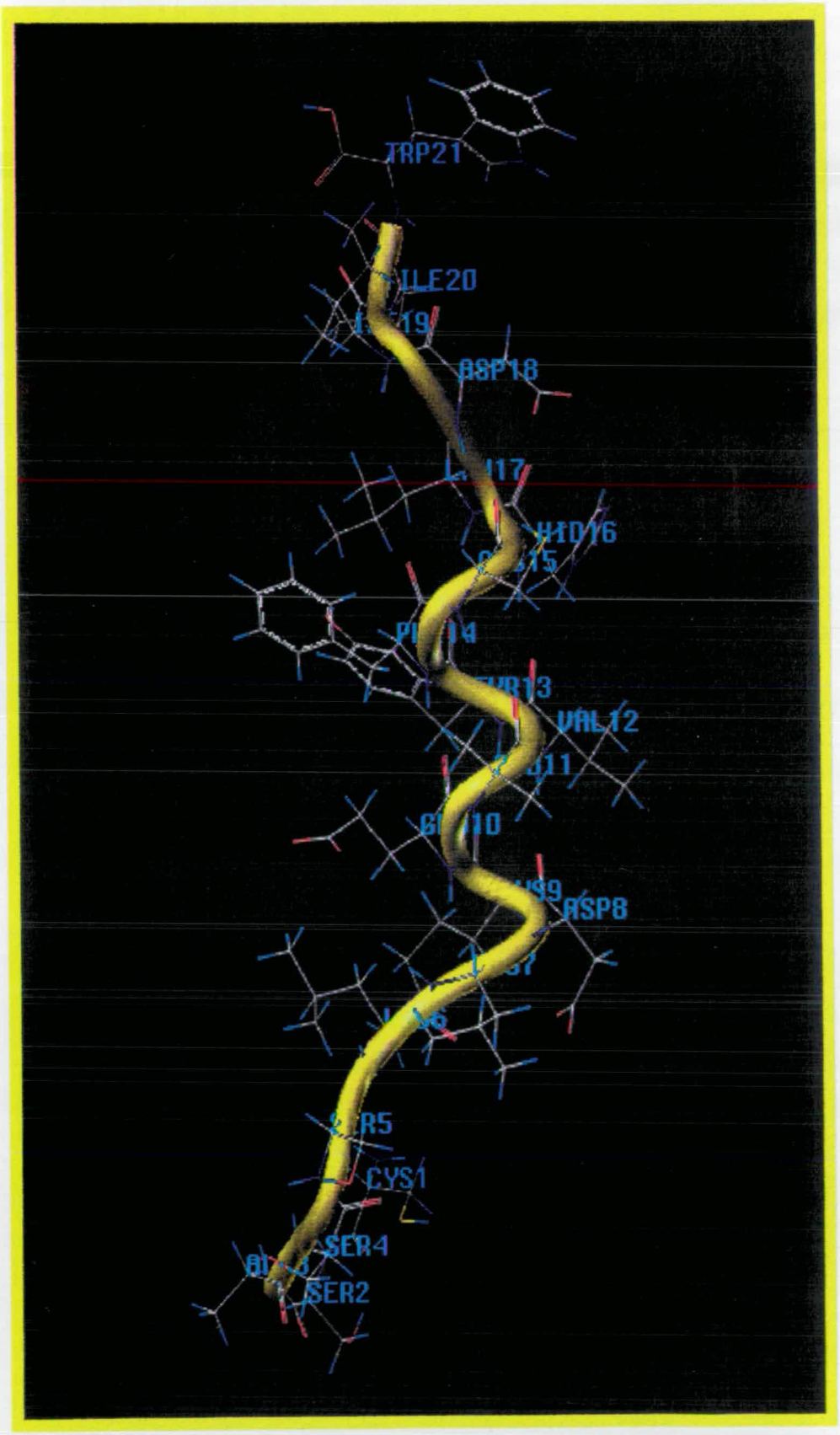
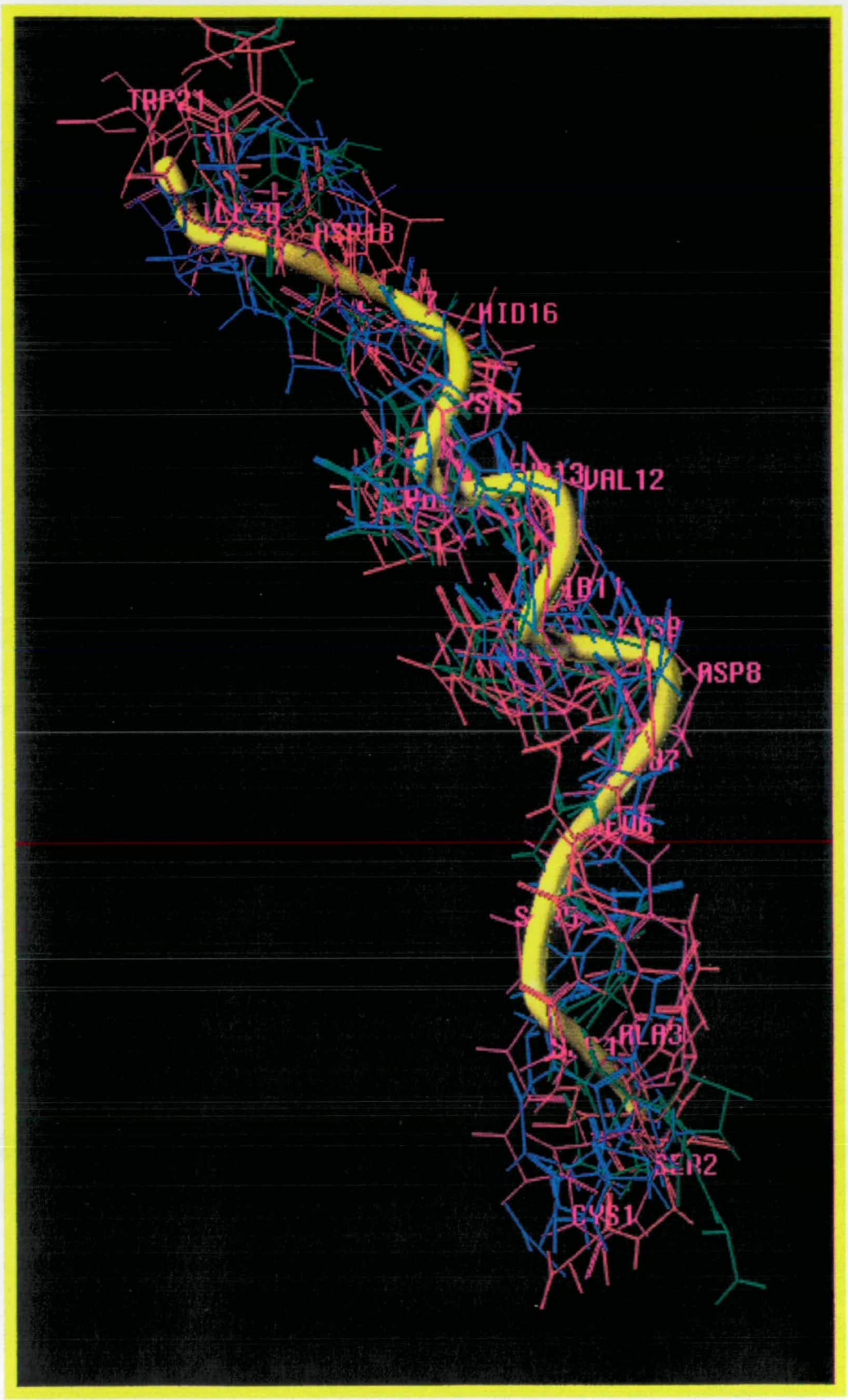


Figure 4.26 : Backbone conformations of ten structures of LJP26



4.2.3.4 Solution Structure of LJP26

The differences of amino acids between LJP1 and LJP26 are positions 3 and 8; LJP1 (Aib and Asp) and LJP26 (Ala and d-Asp). Replacement of the third residue Aib→Ala makes the significant change to the biological activity and the formation of the helix which starts from the residue at 8 position. This is probably due to the loosening behaviour of Aib which induces the helical character in peptides and proteins and also the D-configuration of the aspartic acid at 8 position.

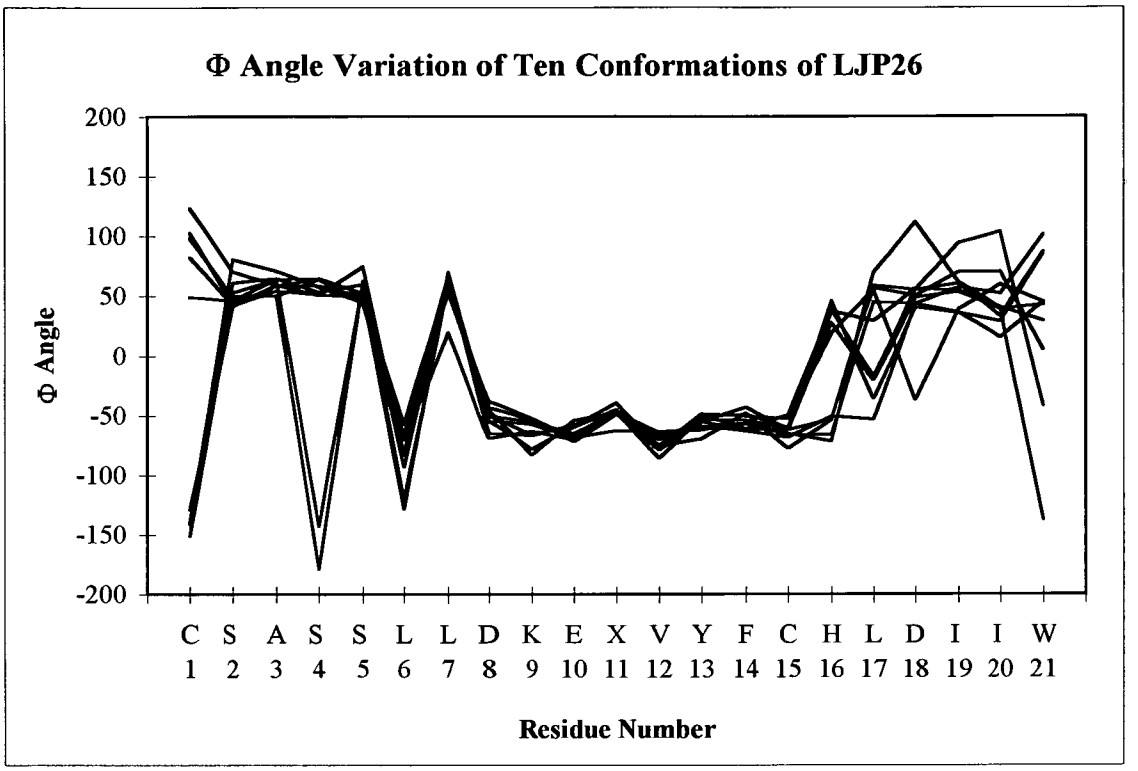
However a few interresidue NOESY connectivities were observed for L^6 and L^7, and structure calculations showed residue 8 as the starting point of the helical region. A right handed helix in the region D^8-H^{16} is a consistent feature of the obtained conformers. Figure 4.25 shows one of the conformations of LJP26. It can be seen that the backbone conformation between residues 8-16 is almost identical while the N- and C- termini show larger variations. Under these experimental conditions, that the region from D^8-H^{16} has essentially the same structure in all contributing conformers. The complete backbones for ten final conformations are shown in Figure 4.26. A full list of structural constraints (NOE, torsional angle, hydrogen bond) and $^3J_{NH\alpha}$ coupling constants may be found in Appendix III (pages 194-201).

The N-terminus (residues 1-6) appears to undergo conformational averaging and no single structure consistent with the NMR constraints could be found in this region. The poor fit outside the helical region suggests that some form of conformational averaging or structural randomisation applies. Summary of ϕ and ψ angles for the ten conformations of LJP26 are given in Figure 4.27.

It should be noted that the high level of structure definition was obtained without prochirality assignments for CH_2 and CH_3 units. This may be due to small size of the peptide and at least five NOE constraints per residue were used for the structure calculations. If stereochemical assignments had been introduced, final conformations may have been further refined especially the details of side chain conformations.

The structures obtained for LJP26 are in good agreement with the experimentally derived constraints and the structural statistics for the family of structures of LJP26 are given in Table 4.13.

Figure 4.27 : Summary of ϕ and ψ angles for the fianl ten conformations. Good convergence is observed for both ϕ and ψ angles in the helical region D^8 - C^{15}



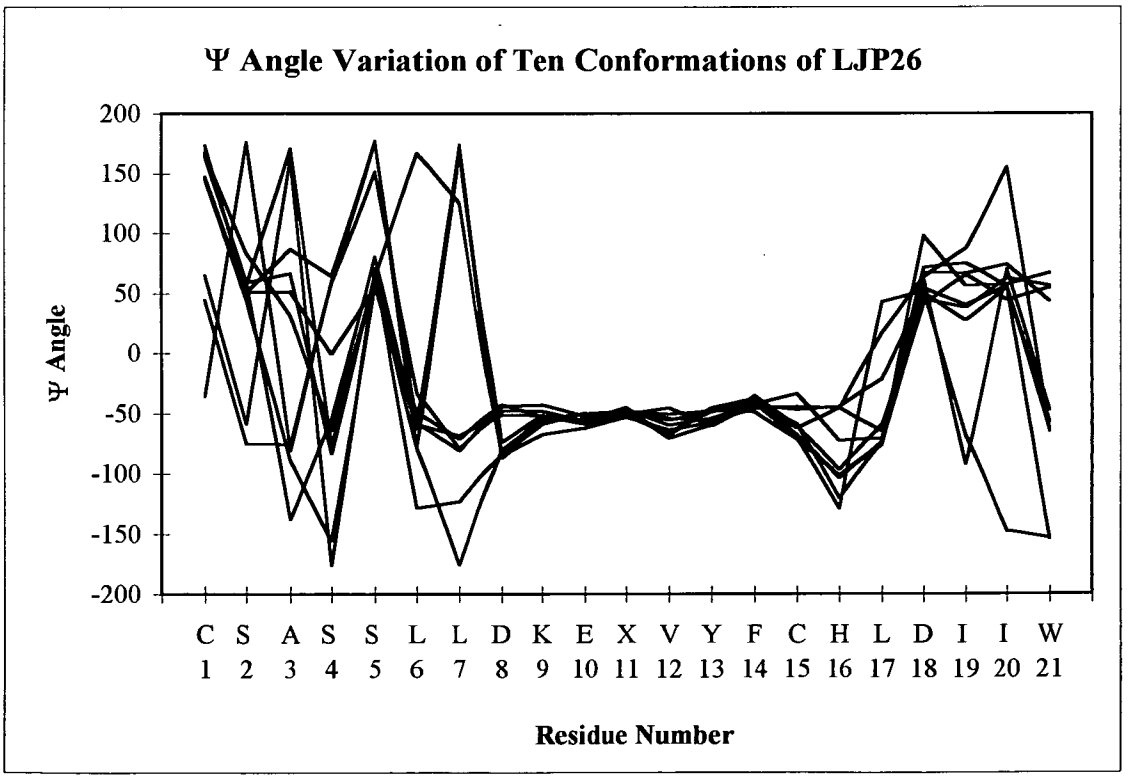


Table 4.13 : Structural statistics for the family of structures of LJP26

Number of DIANA calculated structures	300
Number of structures obtained from DIANA calculation	10
Number of selected structures for final calculation	10

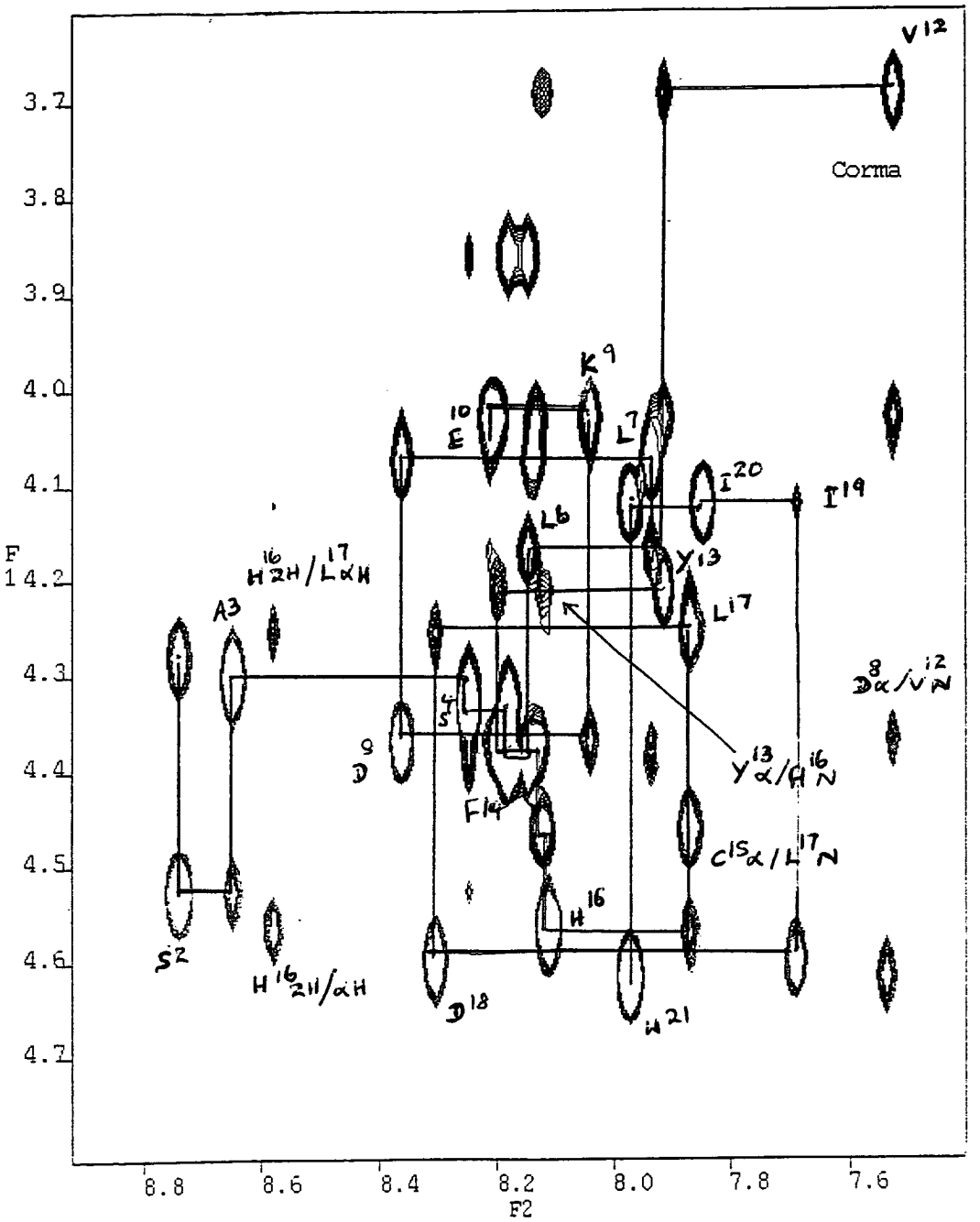
	DIANA structures	FINAL structures
Number of NOE constraints used	245	155
Number of NOE violations > 0.2 A°	34	11
Violations of NOE constraints > 0.5 A°	14	4

	DIANA > 0.2 A°		FINAL structures	
Constraint violations	Mean	SD	Mean	SD
Intra residue constraints	0.463	0.266	0.031	0.0021
Sequential constraints	0.503	0.227	0.042	0.0032
Medium and long range constraints	0.488	0.234	0.032	0.0053
Total NOE constraints	0.486	0.233	0.021	0.0040
Hydrogen bond constraints	0.323	0.032	0.019	0.0021
Van der Waal violations	0.414	0.177	0.035	0.0024
Torsional angle constraints > 5 deg	6.660	1.725	7.557	3.2540

	FINAL structures	
Atomic rms differences (A°)	Mean	SD
Total residues (1-21)	3.612	0.447
Backbone (1-21)	2.475	0.377
Total helical region (8-16)	1.622	0.424
Helical backbone (8-16)	0.762	0.394
Helical heavy atoms (8-16)	1.399	0.378
N-terminus (1-4)	3.249	0.528
C-terminus (17-21)	3.804	0.756

Energy (k cals/ mol)	Mean	SD
Bond stretching	8.942	0.501
Angle bending	82.754	14.075
Torsional	38.752	2.536
Out of plane bending	0.700	0.276
1-4 van der Waals	25.037	2.062
van der Waals	-95.853	5.930
Total	60.333	13.392

Figure 4.28 : Fingerprint region of the back-calculated NOESY spectrum of LJP26



4.2.3.5 Back-Calculated Spectra of LJP26

The level of similarity between the experimental and calculated NOESY data were determined by visual inspection. Almost all the experimental and calculated NOESY data were similar in appearance and intensity except for strong intensity of cross-peaks in the calculated spectrum for $S^2\alpha H/NH$, $S^5\beta H/L^6NH$ and $W^{21}4H/\alpha H$. In contrast a cross-peak, $S^2\beta H/A^3NH$, is missing in the calculated data. These is probably due to the parameters related to the motional behaviour of the both termini residues specially in the N-terminal region.

In addition, the calculated NOESY data showed some of the cross-peaks which didn't appear in the experimental NOESY data (Figure 4.28). The 2H proton of aromatic amino acid His-16 showed NOESY connectivities to αH proton of itself and αH proton of Leu-17. Although a very weak cross-peak observed for $Y^{13}\alpha H/H^{16}NH$ in the experimental NOESY data, calculated NOESY data showed a quite prominent cross-peak confirming the termination point of α-helix at position 16. Long range correlation of $D^8\alpha H/V^{12}NH$ was also observed as a additional cross-peak in the calculated data.

Although the cross-peak $C^{15}\alpha H/L^{17}NH$ didn't appear in the experimental NOESY data, a strong clear cross-peak was observed in the calculated data. Calculated structures of LJP26 showed the 2.5-3.5Å distance between $C^{15}\alpha H$ and $L^{17}NH$ suggesting that the α-helix may probably a extended helix. Some of the documented literature (Table 4.6) of endothelin and endothelin-like peptides also showed that the termination of the helix at position 17 of the peptide sequence.

If the back-calculated NOESY spectra and the experimental spectra match well, it is possible to conclude that the structure not only satisfies the proton-proton distance constraints but is also consistent with the entire proton relaxation network that generates the time-dependant NOESY spectrum.

The overall match between the experimental NOESY data and the back-calculated NOESY data is quite good. In particular, the calculated spectrum produced both the strong and weak cross-peaks observed in the experimental data. This also confirmed

that the NOESY peak assignments and distance constraints measurements had been carried out as accurately as possible with the data. It may therefore be concluded that the LJP26 structure has been determined as precisely as possible with the NOESY distance data and the structure must be close to the actual solution structure.

4.2.4 Differences of NMR data of ET-1, LJP1 and LJP26

Although chemical shifts of NHs between residues L^6-C^{15} for ET-1 and modified analogues are remarkably different ($\Delta\delta$ = 0.05ppm) except Y^{13} and F^{14}, the chemical shifts of C-terminal residues, L^{17}-W^{21}, are almost same. This clearly indicates the characteristics of cyclic and linear features of ET-1 and modified analogues. The chemical shifts of NHs of LJP1 and LJP26 are also different ($\Delta\delta$ > 0.01ppm) between residues 3-10 and remaining shifts of NHs lie between $\Delta\delta \pm 0.05$ppm.

Quite remarkable differences in chemical shifts are also observed for αH of S^4, S^5 and D^8 of LJP1 and LJP26. A single βH resonance was observed for both S^4 and S^5 in LJP1 but well separated signals were observed for LJP26. The Asp-8 βH resonances of LJP26 showed a significant chemical shift differences ($\Delta\delta$ = 0.37ppm) indicating the D-configuration character of the residue. Significant αH and βH chemical shift differences of S^4 and S^5 residues of LJP1 and LJP26 are indeed due to the variation of the amino acid at position 3.

Although the weak $N_iH/N_{i+2}H$ cross-peaks ϕ^3/S^5, L^6/D^8, $K^9/^{11}$, E^{10}/V^{12} and V^{12}/F^{14} were observed for LJP1, LJP26 showed only E^{10}/V^{12} and V^{12}/F^{14} weak cross-peaks. The cross peaks of $S^5\alpha H/D^8NH$ (weak), $L^6\alpha H/K^9NH$ (medium), $L^7\alpha H/E^{10}NH$ (strong), $D^8\alpha H/\phi^{11}NH$ (weak), $S^5\alpha H/D^8\beta H$ (medium) and $D^8\alpha H/\phi^{11}CH_3$ (strong) confirmed the helical character of the LJP1 between S^5-D^8 which is not the case for LJP26.

In general, N-terminus is the least structured region of all the peptides and the length of α-helix changes with the behaviour of the individual amino acids. The C-terminus shows the ordered structure in all cases while ET-1 shows some additional long-range interactions between the helix and the C-terminus.

4.3 Materials and Methods

4.3.1 Sample Preparation and NMR Experiments of ET-1 and its Derivatives

The ET-1 and modified linear ET-1 derivatives were supplied by the peptide synthesis group in our department. Identification of ET-1 was done by comparing with an authentic sample obtained from Parke Davis Pharmaceutical Ltd. All NMR experiments were performed using 5mm proton (inverse) probe on a Varian VXR 600S spectrometer operating at 599.945 MHz for proton. The peptide samples were stored at 280K between experiments. The ^{1}H NMR signal from CD_3OH was used as an internal reference and defined as 3.30 ppm. The CD_3OH present in the sample provided the lock signal for the NMR spectrometer lock channel. The spectrometer was controlled with a Sun 4/110 host computer using the VNMR system software version 4.1. The white solid peptide samples (5mg), ET-1, LJP1 and LJP26 were dissolved separately in 50:45:5 = CD_3OH:H_2O:D_2O in 5mm diameter Wilmad 528PP NMR tubes except for amide proton exchange studies. The peptide samples dissolved in 1:1 = CD_3OD:D_2O were used for amide proton exchange studies. All the NMR experiments were performed on the sample at pH 3.6 and 298K except for variable temperature studies. Sample temperature was controlled using the spectrometer variable temperature control units. General methods for data acquisition and manipulation were optimised for all peptide samples.

Single pulse proton NMR spectra were recorded using the pulse sequence shown in Figure 2.4. A saturation delay 1.5s was used with zero saturation power. 256 Transients were acquired over a 7KHz spectral width and 35K data points were collected. The data were zero filled to 65536 data points and was multiplied by an optimised shifted sinebell squared function prior to Fourier transformation. After Fourier transformation the final spectra were drift and baseline corrected.

The 2D phase-sensitive DQF COSY spectra were obtained using the pulse sequence shown in Figure 2.4. An eight step phase cycle (hypercomplex acquisition) was used with a relaxation delay D1=1.5s and acquisition time AQ=0.146s. 32 Scans were

performed for each t_1 value and 512 FIDs were acquired. Other parameters were SW=7KHz; 2048 data points and 64 dummy scans. Prior to Fourier transformation, the data set was zero filled to 1024 data points only in F_1 dimension and was multiplied by shifted sinebell squared function in both dimensions. Phase constants were introduced after the Fourier transformation and the negative and positive levels were incorporated into contour plots of the resulting data.

The 2D phase-sensitive TOCSY spectra were obtained using the pulse sequence shown in Figure 2.4. An eight step phase cycle (hypercomplex acquisition) was used with a relaxation delay D1=1.5s and acquisition time AQ=0.146s. The MLEV-17 spin-lock pulse was used with a 80ms mixing time cycle and two trim pulses of 2ms each. 16 Transients were performed for each t_1 value and 512 FIDs were acquired. Other parameters were SW=7KHz and 2048 data points. The data set was zero filled to 1024 data points only in F_1 dimension. Prior to Fourier transformation the data set was multiplied by an optimised shifted sinebell squared function in both dimensions. Phase corrections were done in both dimensions after the Fourier transformation. Final spectrum was baseline was corrected.

The 2D phase-sensitive NOESY spectra were obtained using the pulse sequence shown in Figure 2.4. 32 Scans were performed for each t_1 value and 512 FIDs were acquired. An eight step phase cycle was used with a relaxation delay D1=1.5s and the acquisition time AQ=0.146s. Other acquisition parameters were SW=7KHz, 2048 data points and 64 dummy scans. Water signal was saturated during preparation and mixing periods with saturation power ~0 dB. Fourier transformation in F_2 was carried out without zero filling whereas data points in F_1 were zero filled to 1024 data points prior to Fourier transformation. The data were apodized in both dimensions using a shifted sinebell squared window function prior to Fourier transformation. The real Fourier transformation was carried out on 1024x2048 data points. Phase corrections were applied in both dimensions and baseline was corrected only in F_2 dimension. To evaluate the effects of spin diffusion, the LJP26 NOESY data sets at mixing times 40, 80, 150, 200 and 300ms were acquired consecutively without removing the sample from the magnet. Most of the NOEs were however clearly detectable at mixing time of

150ms or higher. Hence the ET-1 and LJP1 NOESY data were acquired only with 150ms mixing time. All the 2D spectra were acquired on a non-spinning sample.

4.3.2 Variable Temperature Experiments

Variable temperature studies of the peptide samples were carried out using standard 1D proton NMR pulse sequence and 1mg of the peptide sample was used. The peptide sample was allowed to equilibrate for 10-15min in the magnet and Z^1 and Z^2 shim coils were readjusted to establish maximum field homogeneity. At each temperature, the transmitter offset frequency was adjusted to coincide with the alteration of the resonance frequency of the HOD protons. Spectra were recorded at four degree intervals from 290-298K and five degree intervals from 298-318K continuously. A relaxation delay D1=1.5s and zero saturation power were used to eliminate the residual HOD solvent resonance. The data were acquired into 35K data points and zero filled to 65K. Other parameters were SW=7KHz, 256 scans and acquisition time AQ=2.5s. Prior to Fourier transformation, the data set was multified by an optimised shifted sinebell squared function.

4.3.3 Amide Proton Exchange Experiments

Amide proton exchange experiments were carried out using 1mg of the peptide sample and standard 1D proton NMR spectra were acquired. The NMR tube only with the mixed solvent alone was allowed to settle in the magnet and all shim coils were adjusted to gain the field homogeneity from the solvent. The peptide sample was then added to the solvent tube and only Z^1 and Z^2 shim coils were readjusted to establish the maximum field homogeneity. A relaxation delay D1=1.5s and saturation power ~0 dB were used only during preparation period to eliminate the residual HOD solvent resonance. 64 Transients were acquired over a 7KHz spectral width and 35K data points were collected. The data were zero filled to 65K and was multified by a shifted sinebell squared function prior to Fourier transformation. Spectra were recorded at 5min intervals (15-20min), 10min intervals (20-50min), 15min intervals (50-80min), 30min intervals (80-140min), 1hr intervals (140-260min) and 2hr intervals (260-740min) consecutively 2min (adjustment of Z^1 and Z^2 shim coils) after the submission of the sample into the magnet.

4.3.4 NOESY Peak Calibrations

To evaluate the effects of spin diffusion, the build-up of NOESY cross-peak intensity was studied by acquiring NOESY spectra with mixing times 40, 80, 150, 200 and 300ms for LJP26. The NOESY cross peaks $D^8\alpha H/K^9NH$, $\phi^{11}NH/V^{12}NH$ and $V^{12}\alpha H/C^{15}\beta H$ which were sufficiently well-resolved from all other peaks were used to measure the cross peak intensities. The height at the centre of the cross peak was measured in different NOESY spectra and the peak heights were plotted against the mixing time. A smooth curve was drawn through the points in the intensity vs mixing time graph. The build-up of NOESY cross peak intensity was assumed to be linear up to 150ms and the data sets acquired with 150ms were used for structure calculations. The ET-1 and LJP1 NOESY data were only acquired with 150ms mixing time.

NOESY data which were normally deposited in the NMR data station were transferred to the Silicon Graphics work station through the ethernet. The data sets were processed using the same recipes as data were manipulated for structure determination. 2D NOESY spectrum was calibrated to the CH_3OH internal reference peak (3.3ppm) and the 2D contour plot was stored for further studies.

One side of the diagonal of the 2D contour plot was used for picking of NOESY cross peaks. The threshold value of the contour plot was set to zero and only the positive cross peaks which were not extensively overlapped were picked manually. Physical characteristics, the peak number, peak position (F_2, F_1), peak height and half height, intensity and volume integrals, were obtained in a table for the manually picked cross peaks. The assignments of the NOESY connectivity details of cross peaks were then added to the table.

The volume integrals and NOESY connectivity assignment columns were then used for generating lower and upper distance constraints. Dipolar couplings were classified into three categories of upper distance bounds according to their volume integral values 2.5Å (strong), 3.3Å (medium), and 5.0Å (weak). The lower distance bound was in all cases set to the van der Waals distance of 1.8Å. The threshold values of upper distance bounds were established using the known sequential distances of d_{NN} and $d_{\alpha N}$.[196] Pseudoatoms with appropriate distance corrections were employed for protons which could not be stereospecifically assigned.[197]

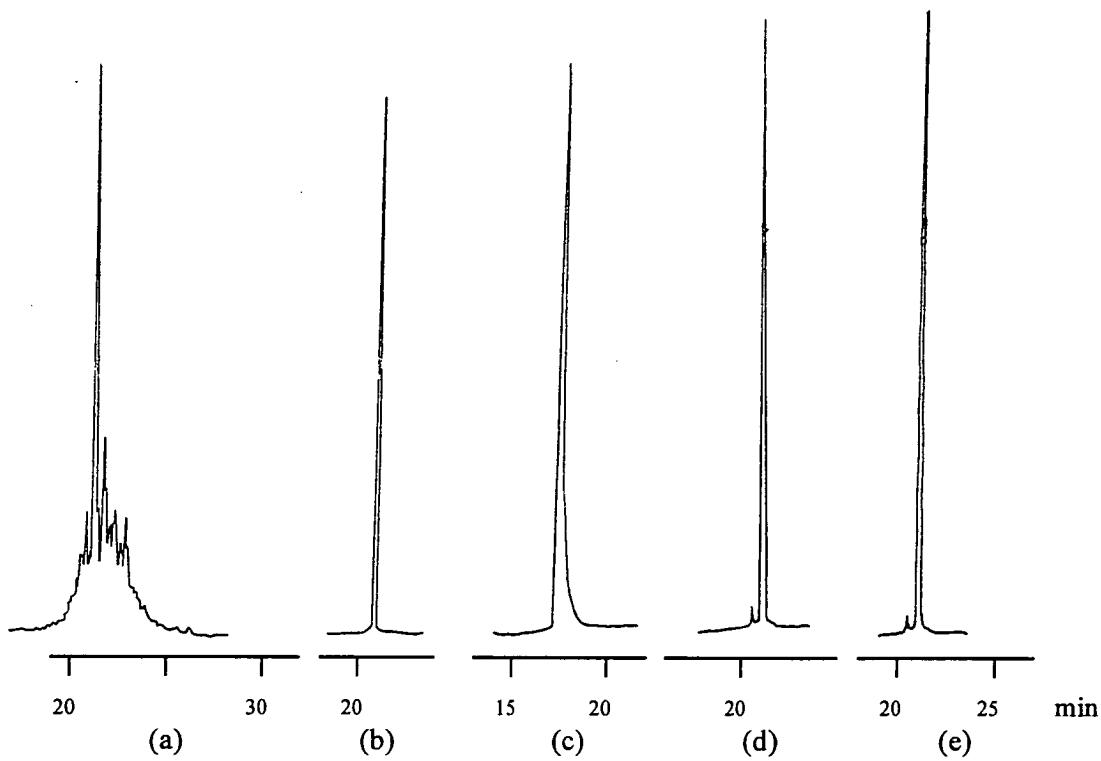
4.3.5 Identification of Endothelin-1

There are three different possibilities of disulphide linkages that may be formed in ET-1 synthesis and only a single major product was observed in our synthesis.

(1) C^1 - C^3 and C^{11} - C^{15} (2) C^1 - C^{11} and C^3 - C^{15} (3) C^1 - C^{15} and C^3 - C^{11}

Possibility of linkage (1) neither been observed in literature of ET-1 synthesis nor in our synthesis and the linkage (2) was not observed in the peptide folding conditions used in our laboratory. Therefore only ET-1 peptide was obtained and purified. The co-injection of purified ET-1 with native ET-1, which was obtained from the Parke Davis Pharmaceutical research, confirmed the correct disulphide linkage (3) of the synthesised peptide. Purified ET-1 peptide was further identified by the amino acid analysis.[165]

Figure 4.29: HPLC traces ; Purification and identification of ET-1 [165]



Gradient (time, acetonitrile %) : (2,10), (32,90), (34,10)
Flow rate : 1 ml/min, 214nm

Columns used :	(a), (b), and (e)	Hichrom C18	220 x 4.6 mm
	(c)	RP C18	110 x 4.6 mm
	(d)	Vydac C18	220 x 4.6 mm

(a)	crude peptide mixture
(b)	purified endothelin-1
(c) and (d)	characterisation of endothelin-1 using different HPLC columns
(e)	co-injection of purified endothelin-1 with native endothelin-1

4.3.6 Additional Distance Constraints

The disulphide bridges of ET-1, C^1-C^{15} and C^3-C^{11} were fixed directly in all structure calculations by constraining a 2.0-2.1Å on the S_i-S_j distances and 3.0-3.1Å on the $^{\beta}C$-$S_{i/j}$ distances across each bridge.[73] The slow amide proton exchange rates were observed for some NHs in the helical region of the peptide backbone. These NH protons and relevant backbone carbonyls were constrained by introducing hydrogen bonds a range of 1.8-2.0Å and 2.7-3.0Å for the H-O distance and N-O distance respectively.[73] The $^3J_{HN\alpha}$ values were obtained from the 1D proton spectra and these coupling constants were used to calculate the ϕ torsion angle. The ϕ torsion angles were then constrained to a range of $\pm 30^0$.

Facilities for calculating macromolecular structures based on NMR data were installed in the department during the course of this project. These facilities were based on funding provided by the Wellcome Trust on a grant to Dr. J.A. Parkinson and Prof. T. Brown. A considerable amount of time was devoted to setting up and optimising the software packages. The results presented in this thesis are the first examples of complete peptide structures calculated from NMR data in this department.

4.3.7 Three Dimensional Structure Calculations

The suitable peptide structures were generated in the following way.

(1). Construction of the peptide by sequentially adding amino acid building blocks

(2). Modify stereochemistry of amino acids and addition of protecting groups

(3). Adding pseudoatoms for $-CH_2$, $-NH_2$, $-CH_3$, $-NH_3$ and aromatic groups if necessary

(4). Introduce random coil conformation

(5). Energy minimisation until a reasonable lower energy was obtained.

The Distance Geometry (DG) program DIANA was used to generate structures from random starting conformations. All distance and torsional constraints were included in DIANA calculations. DIANA was used to calculate 300 structures and the best structures were selected on the basis of their final penalty function values. These structures were then subjected to further calculations.

Output structures were deposited both in angle coordinate files and in PDB/Sybyl MDB files. Default constraint cut-off values were used to obtain the calculation results; upper and lower limit 0.2Å, van der Waals cut-off 0.2Å and angle constraint cut-off 5^0. Maximal hydrogen bond length 2.4Å and hydrogen bond angle 35^0 were used as cut-off values for reporting hydrogen bonds. RMSD comparisons were reported for total structures except for both termini residues and three residues were counted for reporting RMSD segment length.

Table 4.14 : List of constraints used for structure calculations.

		Input Constraints [a]	Output Constraints [b]
ET-1	Intra residue NOE	131	48
	Sequential NOE	64	48
	Long range and medium range NOE	16	16
	Constraints across the disulphide bridges	6	6
	Hydrogen bond	8	8
	Torsional angle	5	5
	Total	230	131
LJP 1	Intra residue NOE	133	44
	Sequential NOE	71	59
	Long range and medium range NOE	45	45
	Hydrogen bond	8	8
	Torsional angle	17	17
	Total	278	173
LJP 26	Intra residue NOE	122	48
	Sequential NOE	65	49
	Long range and medium range NOE	33	33
	Hydrogen bond	8	8
	Torsional angle	17	17
	Total	245	155

a : Constraints used for calculation of DIANA structures

b : Output constraints obtained from DIANA calculations. These modified constraints were used for further calculations.

During the DIANA calculation, atomic distances were constrained using the force constant $k_{NOE} = 4.18$ kJ mol^{-1} $Å^{-2}$ and torsions were constrained using $k_{Dihed\_c} = 0.042$ kJ mol^{-1} deg^{-2}. The DIANA calculation produced 10-20 structures depending on the final target function value.

Distances which are predetermined by the covalent geometry of the molecule or by no possible conformation which will violate the constraint are regarded as irrelevant constraints in the DIANA program. These constraints are eliminated during the calculation and structurally important constraints are produced as modified constraints.

Out of 300 structures calculated from DIANA, 10, 15 and 19 structures were obtained as best structures on the basis of their penalty function values for LJP26, ET-1 and LJP1 respectively. Bear in mind to compare all the families of final structures of all peptides, ten structures were used for further calculations as this was the maximum number of structures obtained form LJP26.

Therefore ten DIANA structures which showed the least number of violated constraints were selected for further calculations. These structures were constrained by the modified distance and torsional constraints which were obtained as a output from the DIANA calculation. In these structures, atomic distances were constrained using a higher force constant using k_{NOE} = 41.8 kJ mol^{-1} $Å^{-2}$ and torsions were constrained using the same force constant k_{Dihed_c} = 0.042 kJ mol^{-1} deg^{-2}. Structures were then subjected to 200 cycles of energy minimisation.

The conjugate-gradient optimisation method was used for energy minimisation through out the structure calculations. The Tripos 5.2 force field with the standard Sybyl energy minimiser MAXIMIN 2 were used in the minimisation program.[198] Higher starting energies were minimised using an atom-by-atom Simplex minimisation. Other parameters for constraining covalent geometry were k_{Bond} = 2500 kJ mol^{-1} $Å^{-2}$, k_{Angle} = 0.084 kJ mol^{-1} deg^{-2} and k_{Tor} = 0.836 kJ mol^{-1} deg^{-2}.[92] Further details of force constants are discussed by Clark et al.[92] At this stage interatomic distances and ϕ torsions were constrained by experimental energy terms k_{NOE} = 41.8 kJ mol^{-1} $Å^{-2}$ and k_{Dihed_c} = 0.042 kcal mol^{-1} deg^{-2}. All other conditions were Tripos MAXIMIN 2 minimiser defaults.

Energy minimised structures were then refined using the dynamical simulated annealing (DSA) method adapted by Tripos molecular modelling software.[198] The DSA process was begun at the maximum temperature 1000K and the system was held at that temperature for 5400fs. During the annealing time 900fs, the temperature was reduced until the minimum temperature 100K was reached. The Boltzmann scaling of atomic velocities were chosen from a random number seed and the "stepwise" annealing method was used during the cooling step. This process completed the first cycle and the process was continued for 10 cycles. The conformations obtained by DSA were further minimised with 200 energy minimisation cycles. The minimisation parameters are same as described before for DIANA calculated structures.

The energy minimised conformations were subjected to final molecular dynamics (MD) quenching calculations without changing both distance and torsional force constants. The MD calculations were performed in the gas phase. The initial atomic velocities were chosen from a random distribution at 1000K and the dynamic trajectory (100fs) was followed for 20ps in 1fs steps.

These calculations were carried out under NTV ensemble with the 10fs coupling factor for the temperature. The molecular dynamics calculations were repeated 3 times for each conformation using different initial values for random number seed (20e3, 60e3 and 100e3) to obtain 3 different conformers for each starting conformations. These 3 different conformers were then averaged to obtain the averaged conformation and 10 averaged conformers were finally obtained. After removing all the distance and torsional experimental energy barriers, these averaged conformers were finally subjected to 500 cycles of energy minimisation as described before.

4.3.8 Back Calculation of Spectra

A model structure of LJP26, the NOESY data (150ms), the individual peak characters and the cross-peak assignments were used to generate a theoretical 2D NOESY spectrum. Only the fingerprint region of the NOESY spectrum of LJP26 was back calculated theoretically using the "corma simulation" program supported by Tripos molecular modelling software.[198]

Modelfree approach was used to calculate the internal motion using the overall correlation time (10ns) and the internal correlation time (1ns). Both intra- and inter-residue methyl distances were calculated according to the 3-site jump model which is recommended for a good starting model structure. The Gaussian peak shape 3Hz and the NOESY mixing time 150ms were used for simulation. Only the experimental intensities and their corresponding calculated intensities from fixed distance atom pairs were used for normalisation.

CHAPTER 5

STRUCTURE ELUCIDATION OF A GLYCOSIDE

5.1 History and Usefulness of Medicinal Plants

The role of biologically active natural products in the developments of drugs used in the modern medicine is unsurpassed even today when synthetic chemistry has developed beyond expectations. Early knowledge about the use of plants in the treatment of diseases came from close observations and experiences. The gradual accumulation of knowledge of therapeutical properties of medicinal plants helped the primitive cultures to provide themselves with remedies for their basic medical needs.

The historical use of plants in the treatment of ailments in different civilisations resulted in several organised systems of medicine. The thearapeutical properties of a large number of medicinal plants were recognised and documented such as Chinese Pharmacopoeia, Materia Medica. This may be considered as the first important stage in the development of medical science. At the same time medical system were developed in various parts of the world, sometimes almost independent of each other. Thus China developed the "Bantu system" and India established the "Ayurveda system". These systems are very close to each other and are still in practise. Probably the earliest of all recorded medicines was the plant "MaHuang". This plant was first described by the Chinese emperor Shen Nung in 2760 B.C. as a remedy for coughs and also as a cardiac stimulant. The active principle isolated from it called ephedrine was introduced into European medicine in 1926.[199]

Earlier chemical research work on plants was limited to the isolation and structure elucidation of major crystalline compounds. The minor constituents which were often non-crystalline and extremely difficult to handle were rejected or ignored. In recent years, chemical investigations are linked with screening for biological activity. It is estimated that of 300,000 plant species in the world, only a small portion has been investigated phytochemically and less so biologically.

5.2 Panax Family Plants

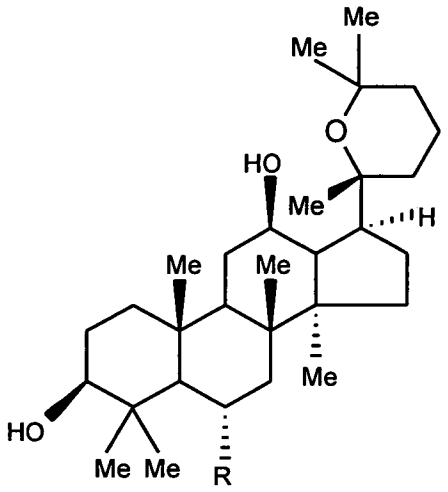
Tenshen, Radix Ginseng, is the dry root of *Panax. ginseng* C.A.Mey. (Araliaceae), a world-wide well known traditional Chinese medicine with the popular name "ginseng". Ginseng was formerly a wild plant growing in the north-eastern region of China. Wild Ginseng, the mountain ginseng, is called "Shanshen" in Chinese and should be dried in the sun. Nowadays, wild ginseng is rarely available, and almost all the commercially available ginseng root is cultivated in the north-eastern district and other regions of the China, where the growing conditions for ginseng plant are favourable. Cultivated ginseng, the garden ginseng, is called "Yuanshen" in Chinese and should be dried either in the sun or after steaming. The steamed root has a caramel colour and is also called "red ginseng".

The wild-growing or cultivated ginseng root, which is collected in the autumn, is officially listed in the Chinese Pharmacopoeia and used as a tonic. This has carrot-like roots which characteristically contain a number of damarrane saponins. The ginsenoside content was found to be higher in the root than in the rhizome. The saponin and ginsenoside contents were higher in the wild than in the cultivated ginseng.[200]

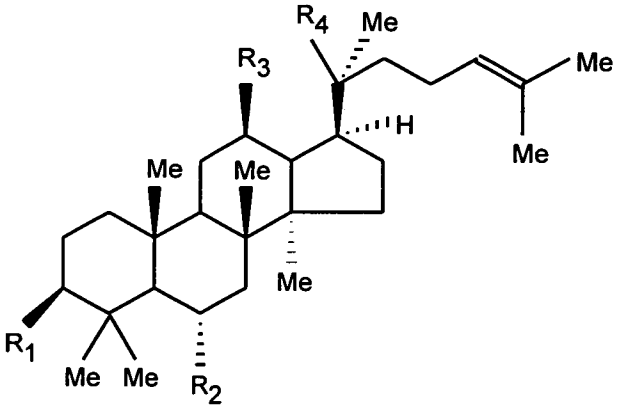
Studies on saponin/sapogenin components of *Panax* species have been reported as being the chemically, biologically and therapeutically active factors. Being biologically active, most of the extensive work had been carried out on *P. ginseng*. The saponin content of ginseng tea showed a total ginsenoside content of 1.1%.

P. japonicus C.A.Mey., *P. japonicus* C.A.Mey. var. *major* (Burk.) C.Y. Wu et K.M. Feng, *P. japonicus* C.A.Mey. var. *bipinnatifidus* (Seem.) C.Y. Wu et K.M. Feng are further *Panax* species officially listed in the Chinese Pharmacopoeia. Most important species found in the *Panax* genus are *Panax ginseng* (Korean Ginseng), *Panax quinquefolium* (American Ginseng), *P. japonicus* (Japanese Ginseng), *P. pseudo-ginseng* (Himalayan Ginseng) and *P. notoginseng*.

Figure 5.1 : Main structures found in *Panax notoginseng* [201-208]

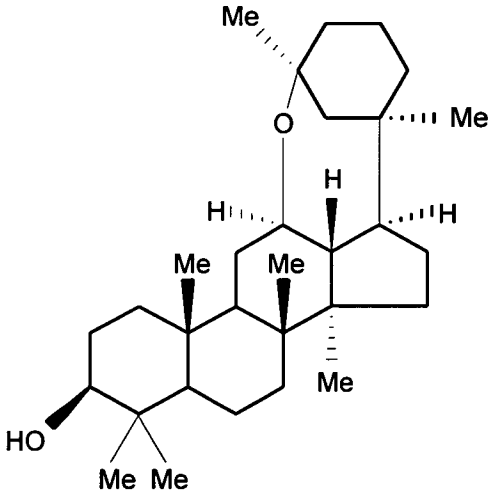


(5.1-1/2)

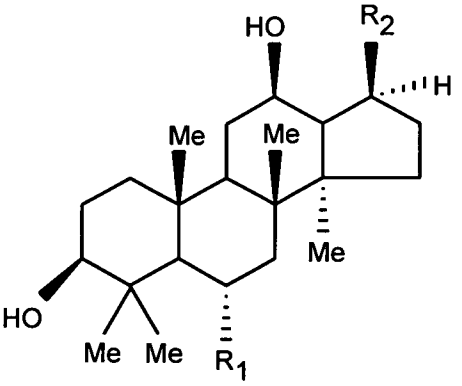


(5.1-3/4)

		R			R_1	R_2	R_3	R_4
5.1-1	panaxadiol	H	5.1-3	protopanaxadiol	OR	H	OH	OR
5.1-2	panaxatriol	OH	5.1-4	protopanaxatriol	OH	OR	OH	OR



Oxepane sapogenin (5.1-5)



(5.1-6/7/8)

		R_1	R_2
5.1-6	Dammar-20(22)-ene-3β,12β,26-triol	H	$-C(CH_3)=CH(CH_2)_2CH(CH_3)CH_2OH$
5.1-7	20(R)-Dammarane-3β,12β,20,25-tetrol	H	$-C(CH_3)(OH)(CH_2)_3C(CH_3)_2OH$
5.1-8	Sanchinoside B_1	-*O*-glc	$-C(CH_3)=CH(CH_2)_2C(CH_3)_2OH$

5.3 *Panax notoginseng*

Sanqi, Radix notoginseng, consists of the dry roots of *Panax notoginseng* (Burk.) F.H. Chen (Araliaceae), another plant of the genus used in traditional Chinese medicine and also officially listed in the Chinese Pharmacopoeia. *P. notoginseng* is taxonomically related to *P. ginseng*.

Ginseng saponins may be divided into three groups, depending on their aglycones. Ginsenosides Rc, Rg_1 and Ro are representatives of the ginseng root saponins of the protopanaxadiol, protopanaxatriol and oleanic acid type, respectively. They are used as reference substances for qualitative determination of ginseng root in the Chinese Pharmacopoeia. Protopanaxadiol (Fig. 5.1-3) and protopanaxatriol (Fig. 5.1-4) are both dammarane derivatives. Ginsenosides Rb_1, Rg_1, and notoginsenoside R_1, are representatives of protopanaxadiol and protopanaxatriol types which are used in the qualitative determination of notoginseng root and in the differentiation of notoginseng root from ginseng root in the Chinese Pharmacopoeia.

The main constituents in notoginseng are saponins especially of the protopanaxadiol and protopanaxatriol types. Thus, a number of ginsenosides such as Rb_1, Rb_2, Rb_3, Rc, Rd, Re, Rg_1, Rg_2, Rh_1, F_2, and glucoginsenoside R_f, have been isolated from the underground part or above ground part of *P. notoginseng*.[201-203] In addition to the ginsenosides, a number of new saponins named notoginsenosides (Tables 5.1 and 5.2) have been isolated and structurally investigated.

In addition to the saponins, sanchinoside B_1 (Fig. 5.1-8), panaxadiol (Fig. 5.1-1), panaxatriol (Fig. 5.1-2), two dammarane derivatives (Fig. 5.1-6/7) and 20(R)-protopanaxatriol have also been detected as sapogenins from the rootlets of *Panax notoginseng*.[204,205] The saponin constituents of the aboveground parts of *Panax notoginseng*, stems and leaves, flowers and buds, and fruits have also been investigated.

Two sapogenins, dammar-20(22)-ene-3β,12β,26-triol (Fig. 5.1-6) and 20(R)-dammarane-3α,12β,20,25-tetrol (Fig. 5.1-7), were isolated from the leaves of

P. notoginseng together with panaxadiol and panaxatriol.[209] A oxepane derivative (Figure 5.1-5) was also found in the leaves of notoginseng.[210] Sapogenins isolated from notoginseng flowers were identified as panaxadiol (Fig. 5.1-1), two dammarane derivatives (Fig. 5.1-6/7) and a another sapogenin with a cyclic ether structural feature.

Furthermore, the notoginseng root contains a number of commercially valuable constituents other than saponins. The major components in the essential oil of the root of *P. notoginseng* have been identified as α-guaiene, β-guaiene, and octadecane and in that of the flower γ-elemene, heptacosane, and pentacosane. Saponins from *P. notoginseng* isolated and identified to date are listed in Tables 5.1 and 5.2.

In the notoginseng saponins eleven different glycosidic groups which are mono-, di- or tri- saccharides, are attached to R_1 - R_4, positions at the aglycone part of the saponin. These are -*O*-glc, -*O*-glc^{2}-^{1}glc, -*O*-glc^{6}-^{1}glc, -*O*-glu^{6}-^{1}glc, -*O*-xyl^{2}-^{1}glc, -*O*-xyl^{6}-^{1}glc, -*O*-xyl^{2}-^{1}glc^{2}-^{1}glc, -*O*-xyl^{6}-^{1}glc^{6}-^{1}glc, -*O*-ara(fur)^{6}-^{1}glc, -*O*-ara(pyr)^{6}-^{1}glc, and -*O*-rha(pyr)^{2}-^{1}glc. All the sugar-sugar ring connectivities which are found at positions 1→2 and 1→6 are common in natural saponins.

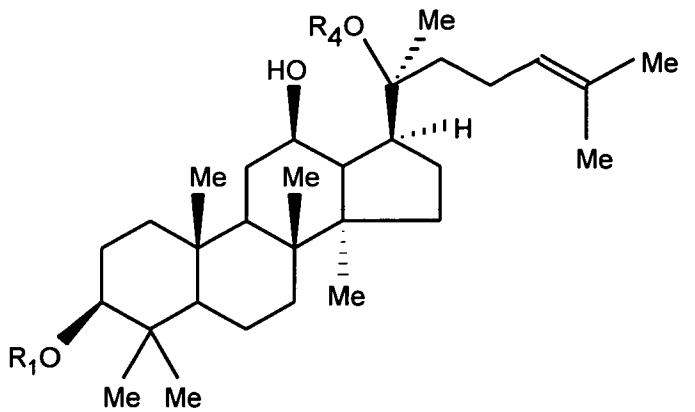
5.4 Pharmacology of Ginseng

Ginseng was used in traditional Chinese medicine for a long time as a general tonic and cardiotonic. Systematic pharmacological investigations revealed a multifaceted biological activity of ginseng, including effects on the cardiovascular, immune, and nervous systems, and activity as an antidote, antitumor agent, or antitumor adjuvant and as antidiabetic.[211]

The medicinal use of notoginseng roots is different from that of ginseng roots. Notoginseng roots are used mainly as a hemostatic drug in the treatment of different types of bleeding. It should be collected in the autumn before the plants bloom.

Panax notoginseng showed a wide spectrum of pharmacological activities, such as hemostatic activity, platelet aggregation inhibitory activity, antiinflammatory activity, and therapeutic effects in cardiac infraction, cardiac ischemia, and angina pectoris.

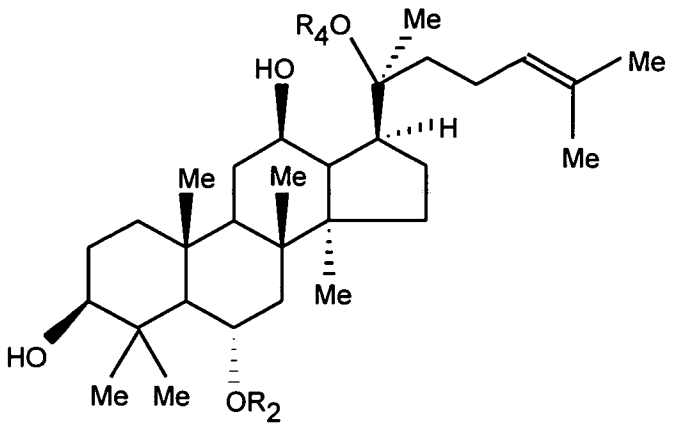
Table 5.1 :Protopanaxadiol type saponins isolated from *P. notoginseng* [201-203,206]



Saponin	R_1*	R_4*	Origin
ginsenoside-Rb_1	-*O*-glc^{2}-^{1}glc	-*O*-glc^{6}-^{1}glc	
ginsenoside-Rb_2	-*O*-glc^{2}-^{1}glc	-*O*-ara(pyr)^{6}-^{1}glc	
ginsenoside-Rb_3	-*O*-glc^{2}-^{1}glc	-*O*-xyl^{6}-^{1}glc	
ginsenoside-R_c	-*O*-glc^{2}-^{1}glc	-*O*-ara(fur)^{6}-^{1}glc	
ginsenoside-R_d	-*O*-glc^{2}-^{1}glc	-*O*-glc	
ginsenoside-F_2	-*O*-glc	-*O*-glc	
notoginsenoside-R4	-*O*-glc^{2}-^{1}glc	-*O*-xyl^{6}-^{1}glc^{6}-^{1}glc	roots
notoginsenoside-R7	-*O*-glc	H	roots
notoginsenoside-Fa	-*O*-xyl^{2}-^{1}glc^{2}-^{1}glc	-*O*-glc^{6}-^{1}glc	leaves
			seeds
notoginsenoside-Fc	-*O*-xyl^{2}-^{1}glc^{2}-^{1}glc	-*O*-xyl^{6}-^{1}glc	leaves
			seeds
notoginsenoside-Fe	-*O*-glc	-*O*-ara(fur)^{6}-^{1}glc	leaves
gypenoside-XVII	-*O*-glc	-*O*-glc^{6}-^{1}glc	roots
gypenoside-IX	-*O*-glc	-*O*-xyl^{6}-^{1}glc	leaves
			seeds

* - see footnote under the Table 5.2

Table 5.2 :Protopanaxatriol type saponins isolated from *P. notoginseng* [201-203,207]



Saponin	R_2*	R_4*	Origin
ginsenoside-Re	-*O*-rha(pyr)^{2}-^{1}glc	-*O*-glc	
ginsenoside-Rg_1	-*O*-glc	-*O*-glc	
ginsenoside-Rg_2	-*O*-rha(pyr)^{2}-^{1}glc	H	
ginsenoside-Rh_1	-*O*-glc	H	
glucoginsenoside-R_f	-*O*-glc^{2}-^{1}glc	-*O*-glc	
notoginsenoside-R_1	-*O*-glc	-*O*-xyl^{2}-^{1}glc	root
notoginsenoside-R1	-*O*-xyl^{2}-^{1}glc	-*O*-glc	roots
			corm
notoginsenoside-R2	-*O*-xyl^{2}-^{1}glc	H	roots
notoginsenoside-R3	-*O*-glc	-*O*-glc^{6}-^{1}glc	roots
notoginsenoside-R6	-*O*-glc	-*O*-glu^{6}-^{1}glc	roots

* - see footnote below

glc : β-D-glucopyranosyl
glu : α-D-glucopyranosyl
xyl : β-D-xylopyranosyl
ara(fur) : α-L-arabinofuranosyl
ara(pyr) : α-L-arabinopyranosyl
rha(pyr) : α-L-rhamnopyranosyl

The saponin fraction of notoginseng given intravenously to dogs decreased blood pressure and peripheral vascular resistance. The hypotensive effect of the saponins appeared to be due primarily to direct dilation of the blood vessels. Intravenous injection of the saponins from notoginseng was also found to be effective in protecting rabbits against haemorrhagic shock due to the improvement of heart function.

All of saponins of *P. notoginseng* were effective against several experimental inflammations in mice and rats. The antiinflammatory effect was stronger in normal mice than in adrenalectomized rats. The synthesis of protein and DNA in the liver, kidney and testis of mice was also significantly increased by oral treatment of notoginseng extract.

Intraperitoneal injection of ginsenoside-R_d into rats increased adrenal intracellar cAMP concentrations. The increase was potentiated by ACTH and was prevented by hypophysectomy, suggesting that the ACTH-like effect in ginseng saponin is due to the direct effect on the hypophysis. Intraperitoneal injection of the total saponins of *P. notoginseng* showed analgesic activity in mice comparable to that of aminopyrine. The appearance of the saponin induced analgesia was faster, but shorter than that of morphine. The total saponins also induced a sedative effect and inhibited caffeine induced locomotive excitation.[211] The detail physiological properties of ginsenosides are described by Fulder.[212]

5.5 Chemical constituents of *P. notoginseng*

Numerous studies of Japanese and Korean origin of *P. ginseng* have been reported by some Japanese workers since early 1900s.[213-215] Although ginseng has been used for a long time and may be the best known traditional Chinese medicine, isolation and characterisation of the chemical constituents only became successful during the 1960s. The major constituents of ginseng are the saponins. The first component of ginseng saponins isolated and structurally elucidated was oleanic acid. Then another compound named panaxadiol (Fig. 5.1-1) was isolated from the crude ginseng saponin mixture. It was supposed to be a sapogenin[216] but was later shown to be formed as an artefact during hydrolysis.

The real sapogenin was then identified as protopanaxadiol (Fig. 5.1-3). The stereochemistry of the aglycone part of saponins had been assigned from its chemical reactions with the aid of literature.[217]

Studies on the saponin components of *P. notoginseng* were reported during late seventies.[218] The GCMS data on the trimethylsilyl ethers of ginsenosides have been reported by Bombardelli et. al.[219] and Kasai and co-workers reported the mass spectra of pertrimethylated products.[220-221]

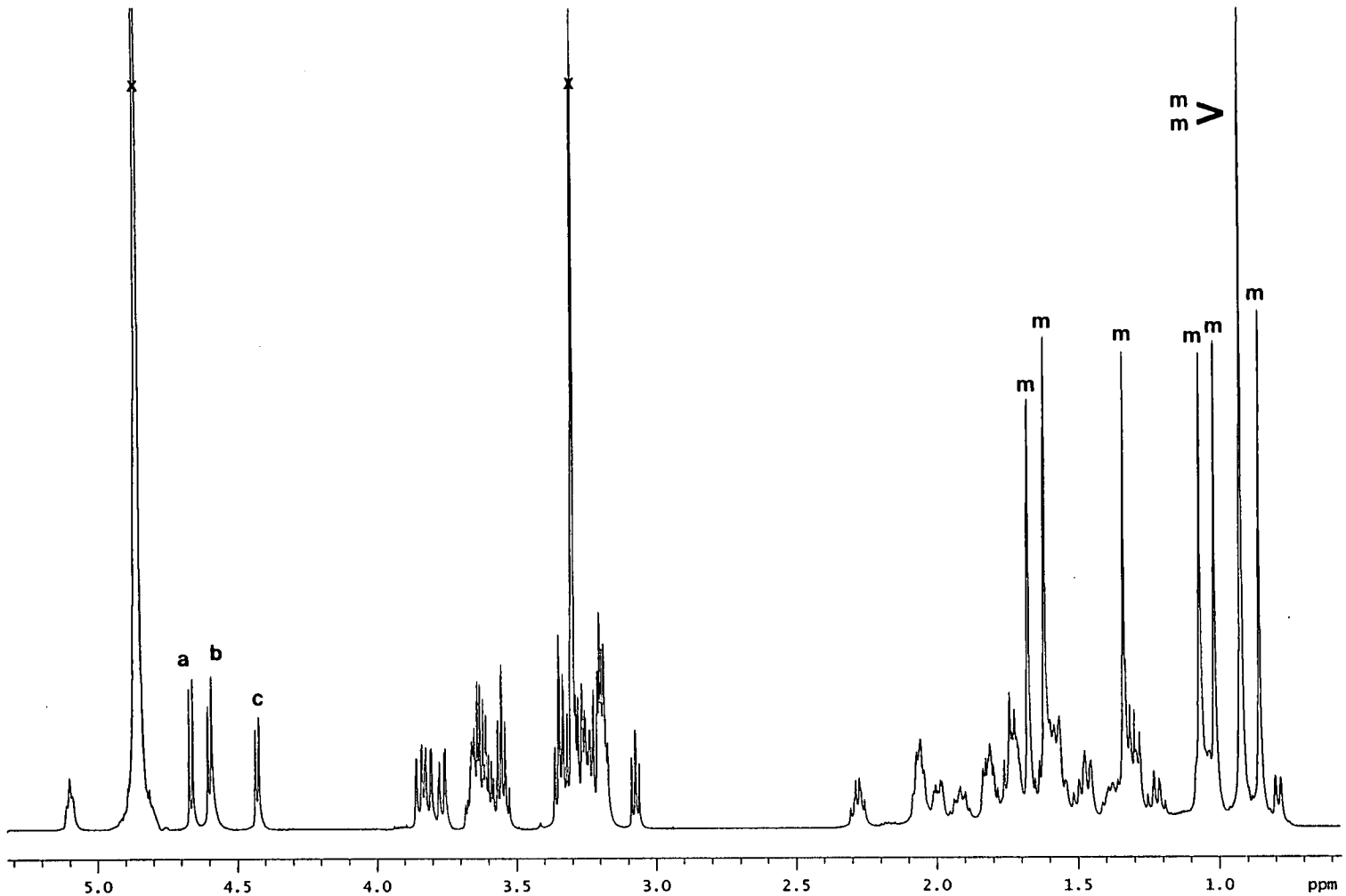
Although isolation and structure determination of a number of saponins from ginseng (root, rhizome, leaf, bud) [220-223] have been reported, little work has been concentrated on structure elucidation by NMR except for some ^{13}C NMR studies. The application of ^{13}C NMR spectroscopy to the structure elucidation of ginseng glycosides have been achieved by the aid of shift reagents and deuterated compounds.[224] Tanaka and co-workers have reported the ^{13}C NMR data for the aglycone and for the sugar moieties.[222-225]

The assignment of the carbon signals of ginseng saponins[226] as well as the glycosylation shifts for a variety of glycosides,[227] mannosides, rhamnosides[228] and arabinosides[229] have also been reported by Japanese workers.

Ginsenoside-Rd is also found in other *Panax* species; *P. japonicus* (rhizome), *P. japonicus* var. *major* (leaves), *P. japonicus* var *bipinnatifidus* (rhizome), *P. ginseng* (roots), *P. ginseng* (flowers), *P. ginseng* (flower buds), *P. ginseng* (seeds), and *P. pseudo-ginseng* (leaves).

It is noteworthy that a cucurbitaceous plant *Gynostemms pentaphyllum* also contained ginsenoside-Rd and number of other ginsenosides including ginsenoside-Rb_1, Rb_2 and F_2.[230]

Figure 5.2 : One dimensional ^{1}H NMR spectrum of Ginsenoside-Rd at 14.1T; m - methyl singlets; a,b and c - three anomeric sugar protons; x - solvent resonances



5.6 RESULTS AND DISCUSSION

A saponin extracted from *Panax notoginseng* was shown to be responsible for tPA release from hemi-pituitary glands in rats *in vitro*. The work described in this chapter has shown that its active principle is ginsenoside-Rd and the ^{1}H and ^{13}C spectra are fully assigned for the first time.

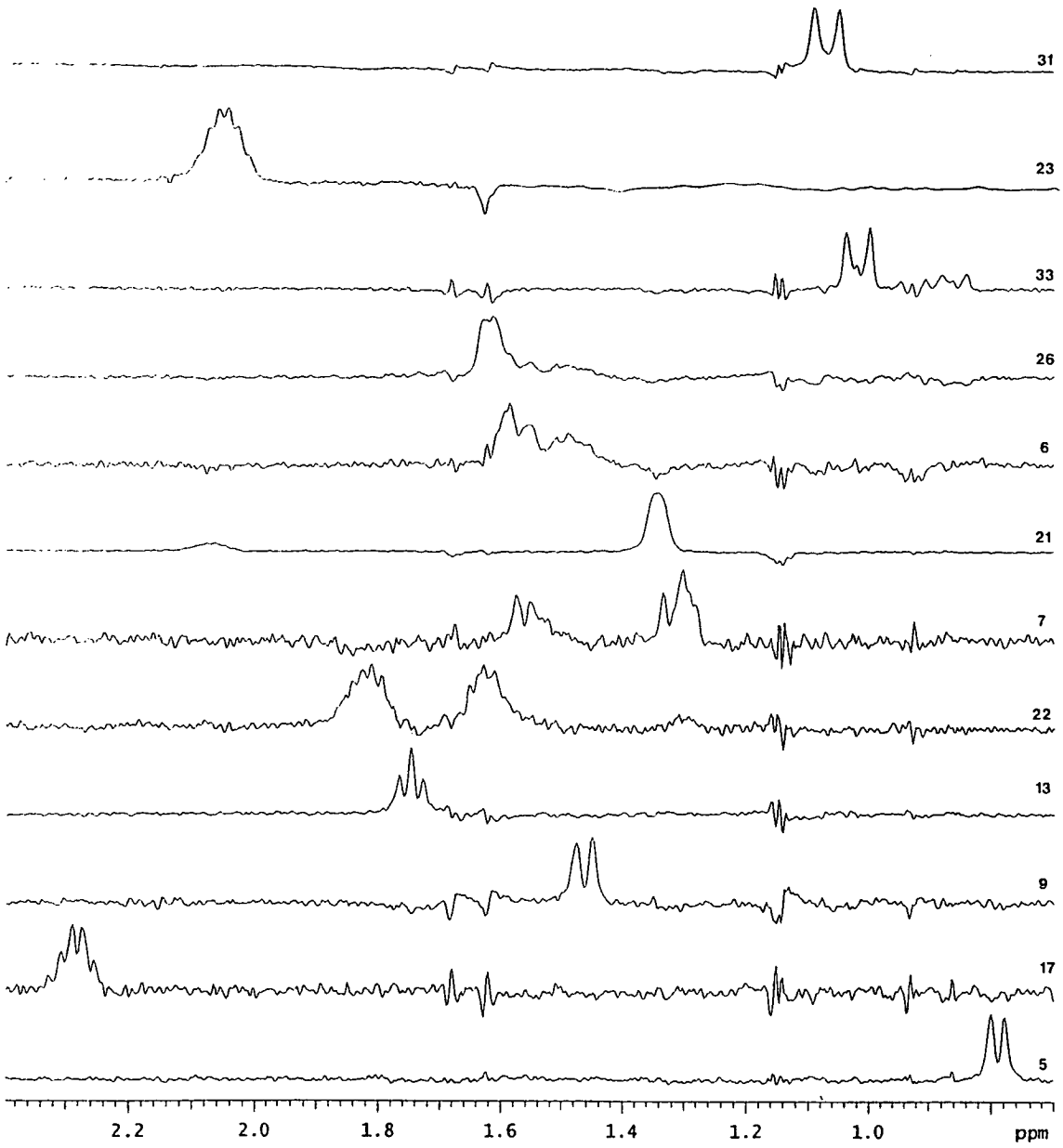
Assignments were obtained by concerted use of one dimensional NMR techniques, ^{1}H, ^{13}C and DEPT and two dimensional NMR techniques, DQF COSY, TOCSY, ROESY, HMQC and HMBC. Sugar correlations via oxygen were obtained using the LR COSY experiment and the two more useful techniques, 1D TOCSY and 1D ROESY experiments were also performed on ginsenoside-Rd to confirm assignments.

5.6.1 Structure Determination of Ginsenoside-Rd

The one dimensional proton spectrum of the glycoside in d_4-methanol (Figure 5.2) showed a one-proton triplet at 5.10 ppm indicating the presence of a alkene proton and clear one-proton doublets at 4.67 (J=7.7 Hz), 4.60 (J=7.9 Hz) and 4.43 (J=7.4 Hz) ppm corresponding to three axial anomeric sugar protons. The low frequency region of the 1D proton spectrum showed six individual methyl singlets at 0.86, 1.02, 1.07, 1.34, 1.62 and 1.68 ppm and a broad six-proton singlet at 0.92 ppm indicating the presence of a total of eight methyl groups. Chemical shifts, multiplicities and coupling constants are tabulated in Tables 5.3 and 5.4.

Classification of ^{13}C resonances as methyl, methylene and methine groups was obtained via spectra obtained by the DEPT technique employing broadband proton decoupling during data acquisition. The quaternary carbons were identified with the aid of ^{13}C single pulse spectrum. Thus the molecule comprised eight methyl groups, twelve methylene groups, twenty-two methine groups and six quaternary carbon atoms. These data are consistent with a three-sugar glycoside of protopanaxadiol or its isomer with a side chain as found in sanchinoside B_1 (Figure 5.1-8).

Figure 5.3 : Cross sections through the 2D ^{1}H-^{13}C one-bond correlation HMQC spectrum at carbon frequencies showing proton multiplet structure. The carbon site is given at the low frequency end of the each trace.



The chemical shifts of the protons of aglycone unit were individually obtained from a 2D proton-detected one-bond ^{13}C-^{1}H correlation (HMQC) experiment carried out at 14.1T and employing broadband ^{13}C decoupling during data acquisition. Some of the ^{1}H cross sections obtained through 2D ^{13}C-^{1}H one-bond correlation HMQC spectrum are shown in Figure 5.3. The shifts for the sugar residues showed considerable overlap and are discussed later. Except for one methylene group, where the bonded protons exhibited similar chemical shifts, each of the methylene carbons correlated with two well separated proton chemical shifts.

Almost all the connectivities of the signals of the spin systems could be obtained from the resonances appearing on vertical/horizontal lines in the TOCSY spectrum. Careful analysis of the TOCSY and DQF COSY 2D spectra showed the presence of seven separate spin systems, four belonging to an aglycone and three belonging to the sugar residues. The two separate spin systems that emerged in the TOCSY spectrum at 3.19 and 3.65 ppm suggested that some of these spins are attached to the carbons which are bound to electronegative atoms such as O, N, S. No evidence of the presence of N and S atoms were found.

5.6.2 Aglycone and sugar spin systems and aglycone ring proton assignments

5.6.2.1: Five-spin system A

The spin connectivities show clearly in the TOCSY spectrum through the diagonal at 3.19 ppm and showed clear cross peaks in the DQF COSY spectrum. Together with the HMQC spectrum these were then identified as a [CH_2-CH_2-CH-O] five-spin system at 1.03 and 1.72 ppm (H-1s), 1.72 and 1.99 ppm (H-2s) and 3.19 ppm (H-3). Two of the protons each from two methylene groups were also overlapped. These protons are consistent with the protons in ring-A of a dammarane type skeleton.

5.6.2.2: Five-spin system B

The low frequency region of the TOCSY spectrum showed connectivities for a [CH_2-CH_2-CH] five-spin system and the DQF COSY spectrum also showed clear cross peaks. Two of the protons each from two methylene groups were also overlapped. The possibility of spin systems, A and B arising from an aglycone in which ring-A bears no

Figure 5.4 : Part of the TOCSY spectrum of ginsenoside-Rd showing side chain and ring-B proton correlations

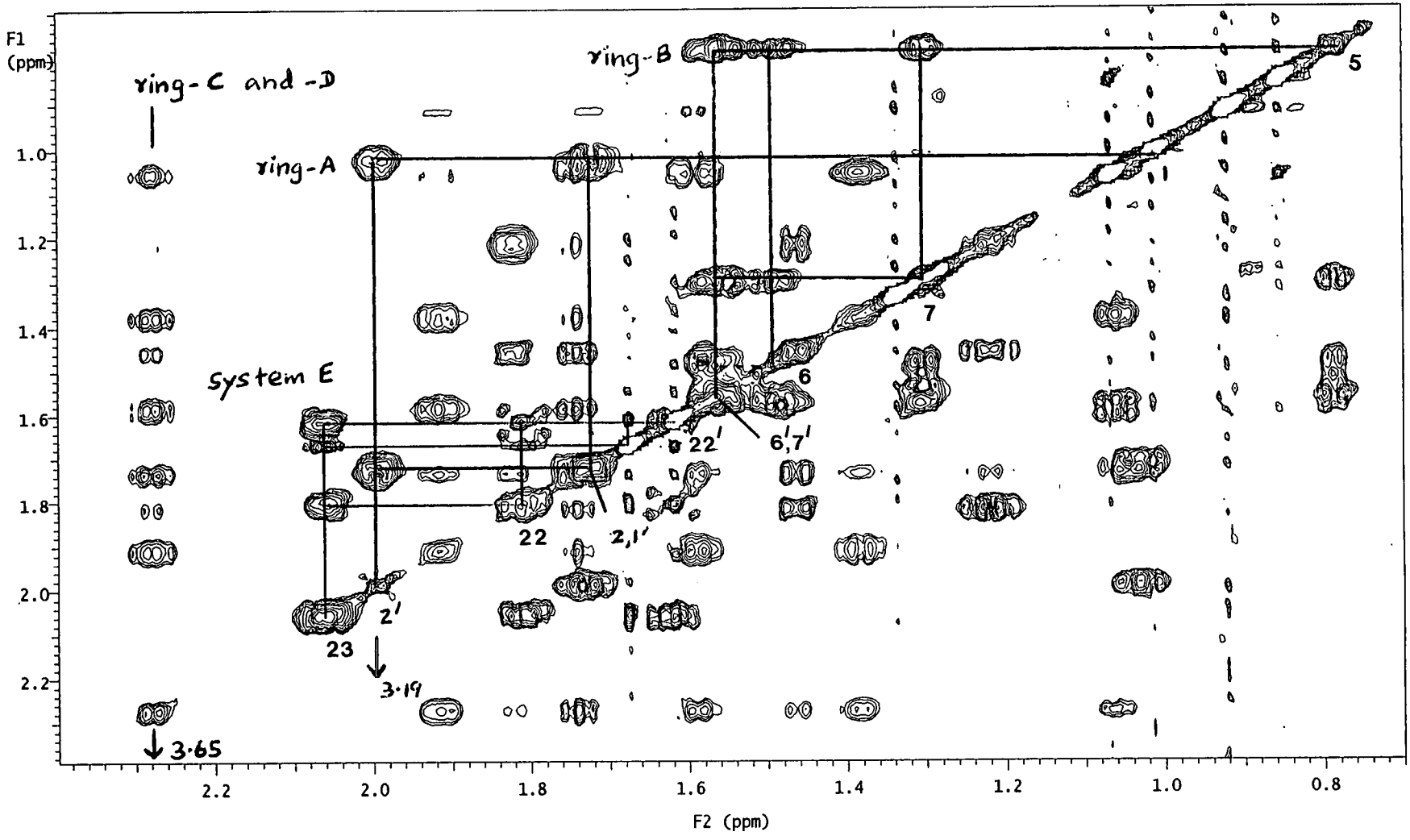


Figure 5.5 : Part of the DQF COSY spectrum of ginsenoside-Rd showing ring-C and -D proton correlations

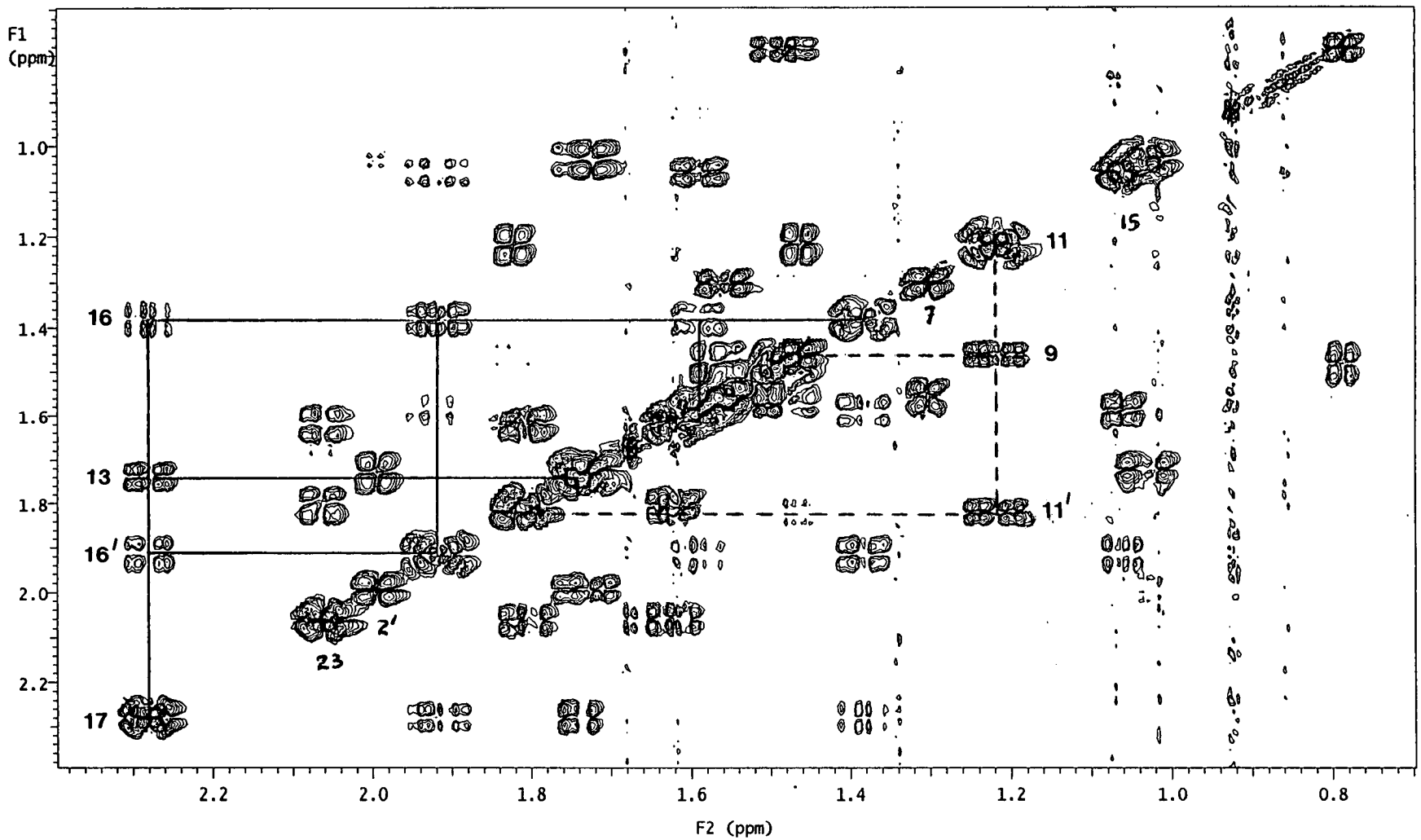
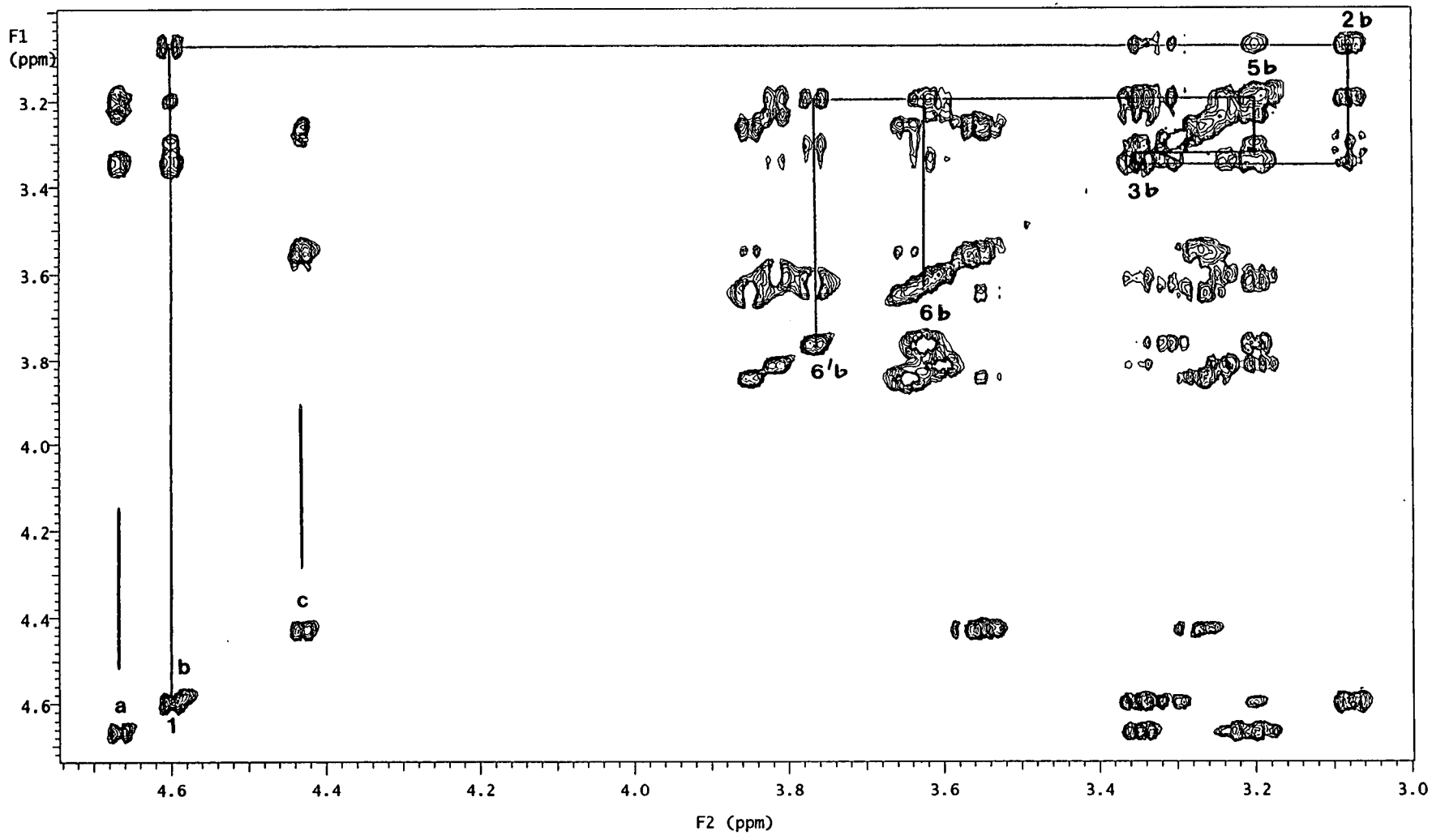


Figure 5.6 : Sugar region of the TOCSY spectrum of ginsenoside-Rd. Complete sugar-b ring protons are shown.



hydroxyl group and ring-B is hydroxylated at C-6 is excluded by data presented later. Figure 5.4 shows the part of the TOCSY spectrum of the saponin showing side chain and ring-B proton correlations.

5.6.2.3: Ten-spin system C

This spin system shows clearly in the TOCSY spectrum at lines through the diagonal fully at 3.65 and 2.28 ppm and partially 1.92, 1.74, 1.58, 1.46 and 1.38 ppm (Fig. 5.4). The complete spin system connectivity was clearly obtained using the TOCSY and the DQF COSY spectra. This spin system corresponds to the protons of ring-C and -D of the aglycone with a OH group at C-12 position. Part of the DQF COSY spectrum of the saponin revealing side chain and ring-C and -D proton correlations are also shown in Figure 5.5.

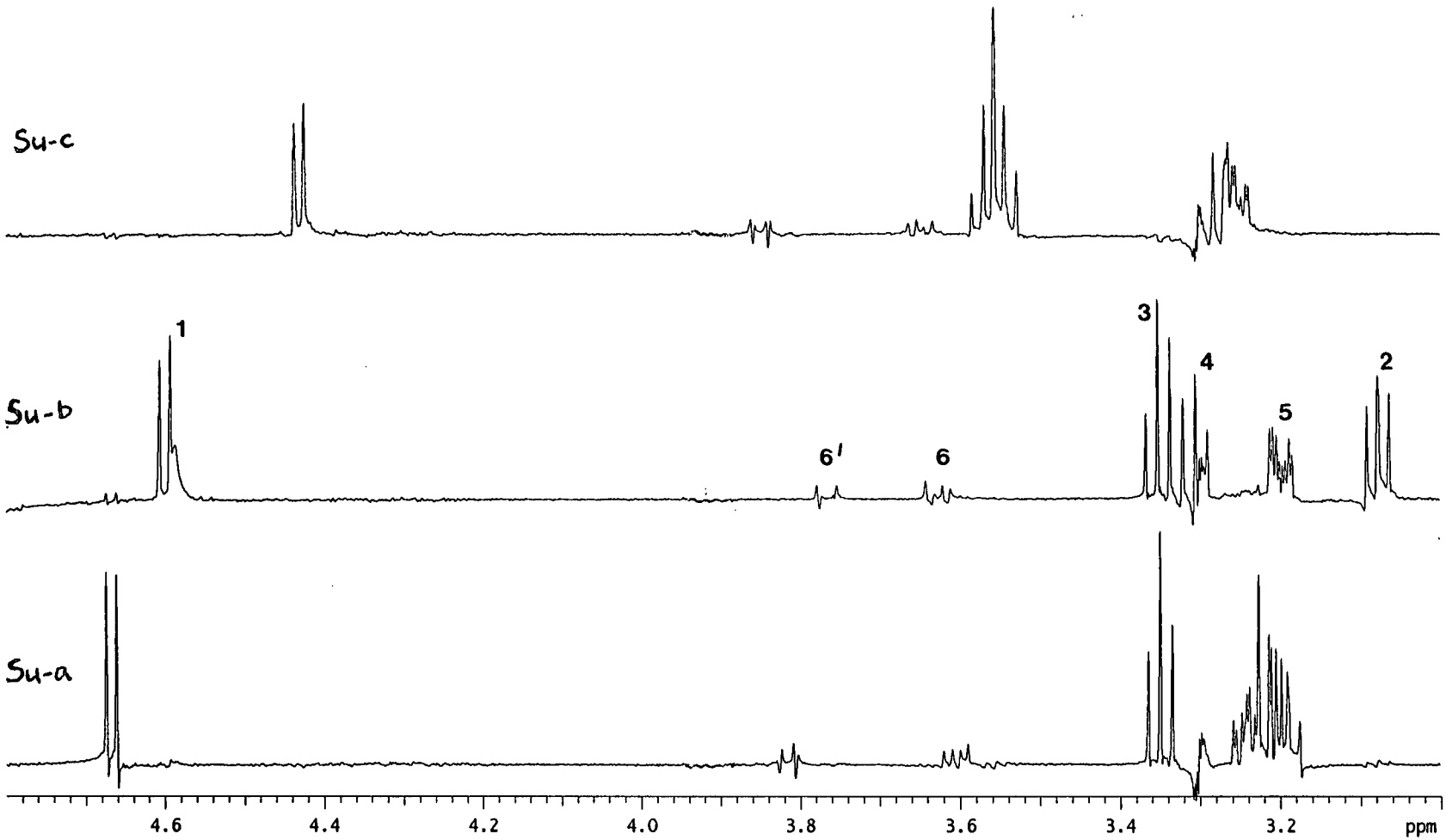
5.6.2.4: Five-spin system E

The alkene proton connected spin system was clearly identified in the high frequency region of the TOCSY spectrum at 5.10 ppm. Other resonances connected to the same spin system were identified as two methylenes with the aid of the TOCSY (Figure 5.3), DQF COSY and the HMQC spectra. The methylene protons next to the olefinic carbon showed similar proton chemical shifts. The two methyls appeared as singlets at 1.62 and 1.68 ppm showed weak connectivities to each other and strong connectivities to the alkene proton in the TOCSY spectrum. This indicates the ^4J proton connectivities of the alkene proton to the geminal methyls which are bound to a olefinic carbon. These resonances correspond to the side chain protons of the type in protopanaxadiol and excludes an aglycone in which the side chain is of the type in sanchinoside B_1 (Fig. 5.1-8). Therefore this data is consistent only with aglycone of protopanaxadiol. The proton chemical shifts and coupling constants are given in Table 5.3.

5.6.2.5: Sugar ring spin system; Sugar-a, -b, -c

Three separate spin connectivities appeared in the TOCSY spectrum at 4.43, 4.60 and 4.67 ppm laid in the normal resonance region for sugars. Figure 5.6 shows the sugar region of the TOCSY spectrum of the saponin. This suggested the presence of three monosaccharide units in the saponin molecule.

Figure 5.7 : 1D TOCSY spectra of ginsenoside-Rd; Irradiation positions at aH-1, bH-1 and cH-1 anomeric sugar protons. Complete sugar-b ring protons are shown.



Complete assignments of the three sugar protons, sugar-a, -b and -c were obtained with the aid of the TOCSY and the DQF COSY spectra.

The proton coupling constants of ^{2}J and ^{3}J, multiplicities of resonances and conformation of the protons were obtained using the set of 1D TOCSY spectra (Figure 5.7). One dimensional spectra were acquired using the z-filtered TOCSY sequence with low power selective excitation pulses at the positions of the protons H-24, H-17, H-2', aH-1, bH-1 and cH-1.

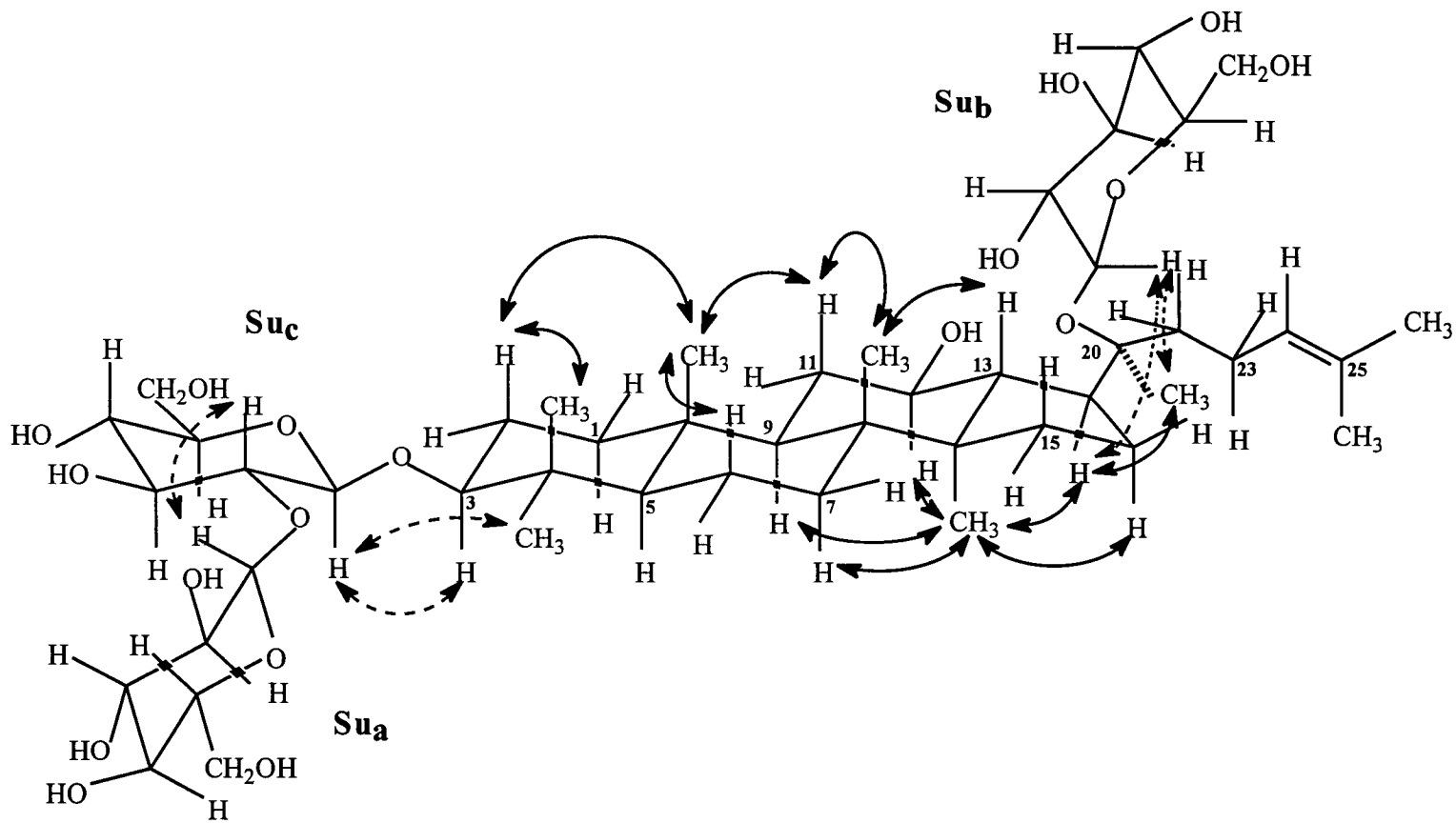
The confirmations of axial conformation of proton resonances H-3, H-5, H-9, H-12, H-13 and H-17 were obtained by using the coupling constants observed in the 1D TOCSY spectra. The 1D TOCSY spectra were also used to confirm the axial conformations of geminal protons H-1/1', H-2/2', H-6/6', H-7/7', H-11/11', H-15/15', and H-16/16'. The coupling constants of all the protons except H-6' of all the sugars, sugar-a, -b, and -c, were obtained using the 1D TOCSY spectra. The coupling constants of sugar H-6s' were obtained from the 1D proton spectrum. All the sugars were then identified as β-D-glucopyranosyls.

5.6.2.6: Hydroxyl group

An isolated hydroxyl proton resonance shown later to be on C-12 position was identified from the broad peak appears on the low frequency edge of the Sugar-b anomeric proton resonance at 4.59 ppm in the 1D ^{1}H spectrum. The sugar ring hydroxyl protons appeared to exchange with the solvent and the residual OH resonances appeared at 4.85 ppm.

The basic skeleton was suggested using the 1D proton, 1D ^{13}C, DEPT, 2D TOCSY, DQF COSY and the HMQC data and literature. The skeleton in Figure 5.9 was numbered using the standard triterpene numbering procedure. The biological pathway forming the dammarane type of skeleton has showed that the usual attachment of oxygen atoms is at positions C-3, C-6, C-12 and C-20. The data obtained on this saponin suggested the attachment of oxygen atoms at positions C-3, C-12 and C-20.

Figure 5.8 : The ROESY correlation diagram of ginsenoside-Rd



At this stage the basic skeleton, the side chain and attachments of the oxygen atoms are known and some individual resonance identifications remain to be made.

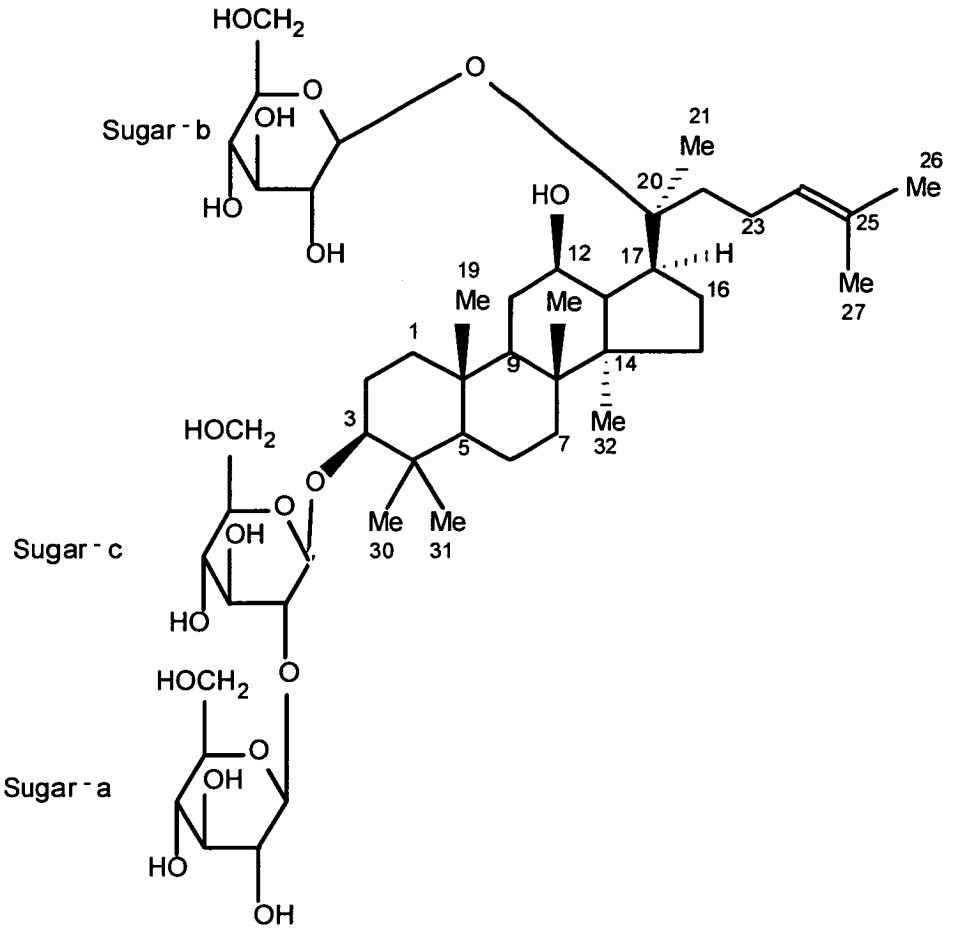


Figure 5.9 : Structure of saponin, ginsenoside-Rd; 20(s)-protopanaxadiol-3-*O*-β-D-glucopyranosyl-(1→2)-β-D-glucopyranoside-12-hydroxy-20-*O*-β-D-glucopyranoside

5.6.3 Methyl proton resonance and ring connectivity assignments

Assignment of the proton resonances of the saponin was confirmed by the connectivities appeared in the ROESY spectrum. The alkene proton H-24 showed through space connectivities to the protons H-22, H-23, and methyl resonances at 1.62 and 1.68 ppm confirmed the identity of methyl-26 and methyl-27 and the presence of a side chain. The ring-D proton H-17 and side chain proton H-23 showed strong ROESY connectivities to the methyl group at 1.34 ppm. These connectivities confirmed the attachment of methyl group at 1.34 ppm to the carbon at C-20 hence methyl group is identified as methyl-21. The ROESY correlations are shown in Figure 5.8.

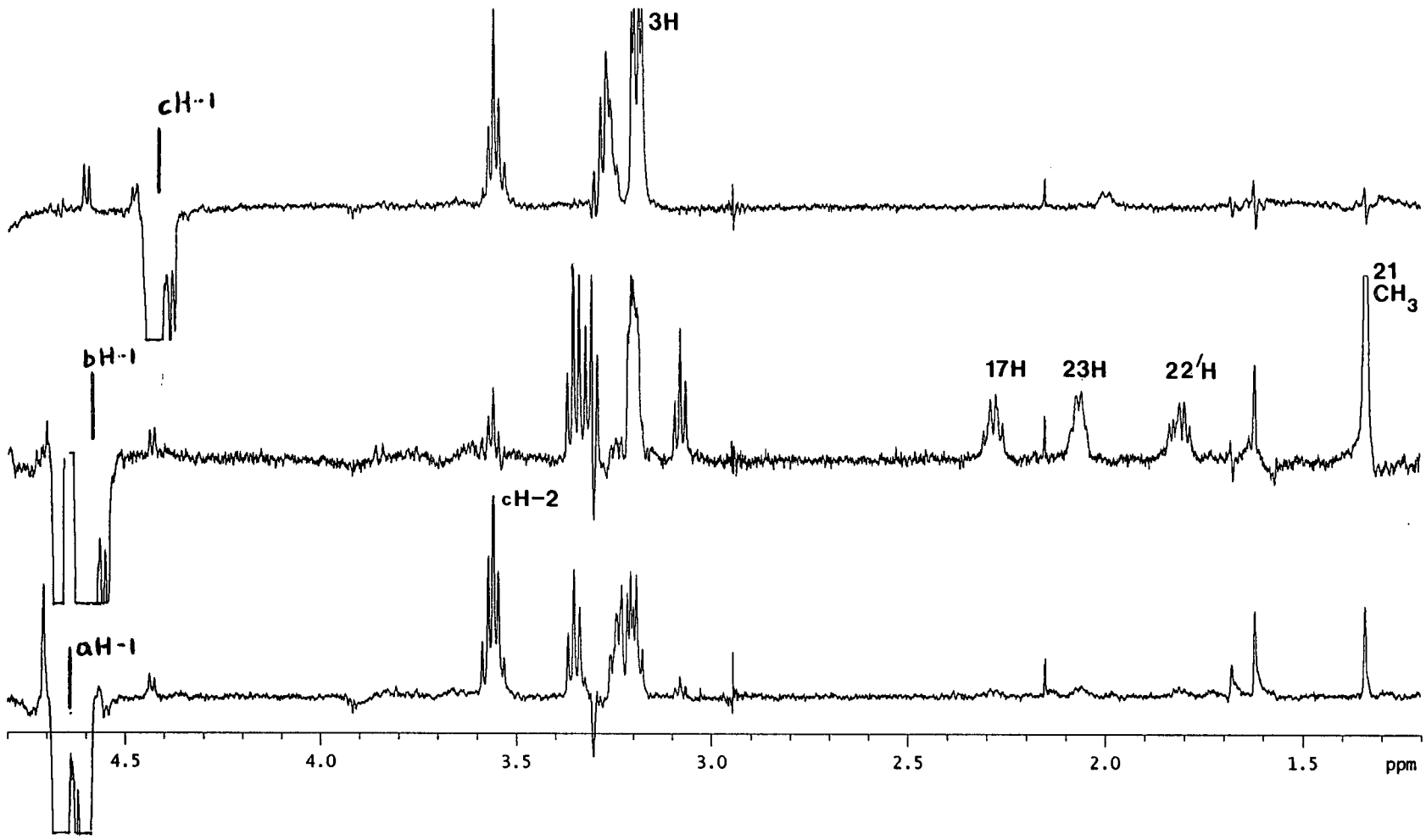
The methyl group singlets appeared at 0.86 and 1.07 ppm showed DQF COSY cross peaks to each other suggesting geminal methyl groups. The methine proton at C-3 showed the ROESY connectivities to the protons H-1', H-2, H-5 and methyl resonance at 1.07 ppm. The methyl groups at 0.86 and 1.07 ppm showed clear connectivities to each other in the ROESY spectrum. Low frequency region of the ROESY spectrum also showed connectivities of geminal protons themselves and to each other at C-1 and C-2 positions. The ROESY connectivities of the methyl group at 0.86 ppm to the H-1 and methyl group at 1.07 ppm to the H-5 confirmed the resonances as geminal methyl groups, methyl-30 and methyl-31, at C-4.

The methine proton at C-12 showed the ROESY cross peaks to the protons H-9, H-11 and one of the methyl resonances at 0.92 ppm hence methyl-19 is identified. The methylene protons at C-15 and C-16 showed strong ROESY connectivities to their geminal partners. The protons H-17, H-16', H-15', H-13 and H-9 showed clear through space ROESY cross peaks to the one of the methyl resonances at 0.92 ppm. Therefore methyl-32 is identified. The protons H-2, H-6 and H-11 also showed ROESY connectivities to one of the methyl resonances appeared at 0.92 ppm in the ROESY spectrum. Assignment of the methyl resonances at 0.92 ppm (19 and 32) and the axial conformations of the protons H-5, H-17 and H-9 were therefore confirmed. The methyl resonance at 1.02 ppm showed through space ROESY connectivities to the protons H-11, H-7, H-9, H-15 and H-13 confirmed the assignment of methyl-33 and axial conformation of H-13.

The geminal protons at C-6 and C-7 showed ROESY connectivities to their coupling partners themselves and to each other. The H-5 proton showed strong through space ROESY connectivities to the protons H-9, H-7' and H-1 confirmed the axial conformations of H-5 and H-9 protons.

All the sugars, sugar-a, -b and -c, showed ROESY connectivities to their protons themselves. The bH-1 proton of the sugar-b which showed strong ROESY cross peaks to protons H-17, H-22, H-23 and methyl-21 confirmed the attachment of the sugar-b at C-20. The assignment of the sugar-c which is directly attached to the carbon C-3 was

Figure 5.10 : 1D ROESY spectra of ginsenoside-Rd; Irradiation positions at aH-1, bH-1 and cH-1 anomeric sugar protons. Prominent sugar-sugar and sugar-aglycone correlations are shown.



confirmed using the strong ROESY cross peaks of the cH-1 proton to the protons H-3, H-2 and methyl-31 of the aglycone.

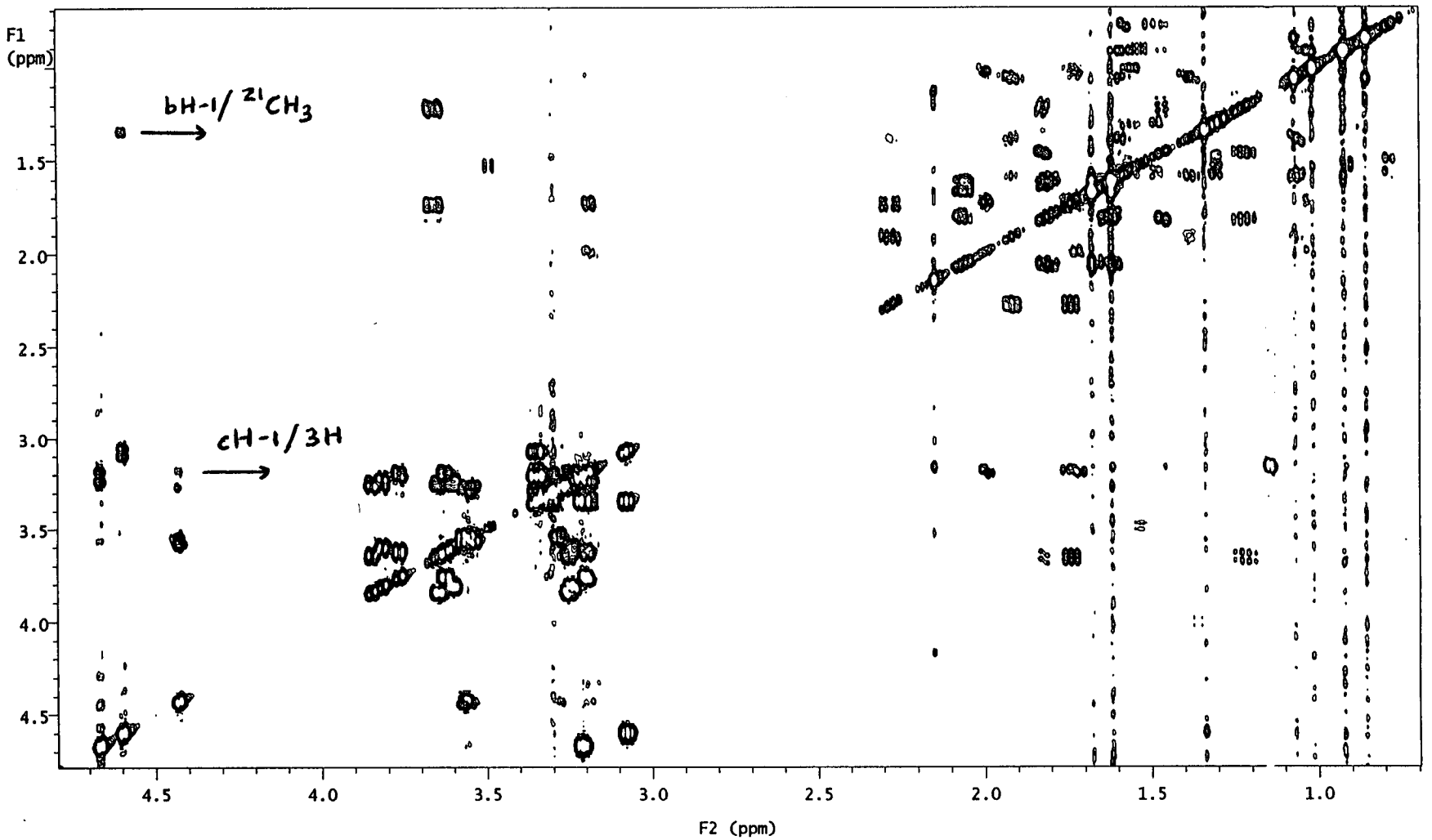
The sugar-a which did not show any connectivities to the basic skeleton showed a strong ROESY cross peak confirming the attachment of sugar-a to sugar-c. At this stage linkage position was not clear due to partially overlapping of the sugar-c cH-2 and cH-4 resonances.

In addition to the connectivities obtained from the ROESY spectrum, almost all the through space correlations were also confirmed with the aid of 1D ROESY spectra (Figure 5.10). One dimensional ROESY spectra were acquired using the 1D version of the 2D ROESY sequence with low power selective excitation pulses at the proton resonances H-12, aH-1, bH-1 and cH-1. All the connectivities of aglycone-aglycone and aglycone-sugar were confirmed except sugar aH-1 to sugar-c connectivity due to partially overlapping of cH-2 and cH-4 sugar resonances.

The COSY experiment was performed to obtain long-range couplings; specially via oxygen atoms. Sugar aH-1 showed clear ^{4}J coupling connectivity to sugar cH-2 proton confirming the sugar -a/-c linkage position and sugar cH-1 showed ^{4}J through bond correlation to aglycone H-3 proton. Sugar-b which showed the bH-1 to methyl-21 ^{5}J through coupling via oxygen reconfirmed the attachment of sugar-b to aglycone. Part of the LR COSY spectrum of ginsenoside-Rd revealing sugar-aglycone correlations is shown in Figure 5.11.

All the ^{13}C resonances except quaternaries were assigned using the HMQC spectrum. All the methyl protons in the HMBC spectrum showed ^{13}C-^{1}H couplings and appeared as doublets in the proton dimension. The quaternary carbon appeared in the HMBC spectrum at 132.24 ppm showed ^{2}J ^{13}C-^{1}H connectivities to the methyl protons methyl-26 and methyl-27 and ^{3}J coupling to the H-23 confirmed the carbon C-25. The ^{2}J ^{13}C-^{1}H couplings of H-17, methyl-21 and H-22 and ^{3}J coupling of H-13 to the quaternary carbon at 84.92 ppm in the HMBC spectrum confirmed the carbon C-20.

Figure 5.11 : Part of the LR COSY spectrum of ginsenoside-Rd showing sugar-aglycone correlations



The quaternary carbon at 52.48 ppm in the HMBC spectrum showed ^{2}J ^{13}C-^{1}H couplings to the protons H-13, methyl-32 and ^{3}J connectivities to the methyl-33 protons confirmed the carbon C-14. The observed ^{2}J ^{13}C-^{1}H couplings of the protons H-5, H-9 and methyl-19 to the quaternary carbon at 37.91 ppm in the HMBC spectrum confirmed the assignment of carbon C-10.

The confirmation of the quaternary carbon C-8 was aided by using the ^{2}J ^{13}C-^{1}H couplings of protons H-9 and methyl-33 and the ^{3}J couplings of the methyl-32 protons to the quaternary carbon at 41.00 ppm in the HMBC spectrum. The quaternary carbon at 40.58 ppm showed ^{2}J ^{13}C-^{1}H couplings of the protons H-5, methyl-30 and methyl-31 confirmed the assignment of carbon C-4.

The structure of the biologically active saponin is then identified as protopanaxadiol type saponin, ginsenoside-Rd;{20(s)-protopanaxadiol-3-*O*-β-D-glucopyranosyl-(1→2)-β-D-glucopyranoside-12-hydroxy-20-*O*-β-D-glucopyranoside} only by high field NMR spectroscopy.

Table 5.3: Chemical shifts, Multiplicities and Coupling constants of the aglycone part of the glycoside : Ginsenoside-Rd

Site	δ_H[a]	δ_C[b]	m[c]	^2J(Hz)	^3J(Hz)
1 (α)	1.03	40.25	m	12.4	11.1 (H-2), 3.8 (H-2')
1' (β)	1.72		m	12.4	
2 (β)	1.72	27.21	m	10.6	11.1 (H-1), 3.2 (H-1')
2' (α)	1.99		m	10.6	3.8 (H-1)
3 (α)	3.19	91.26	dd		12.4 (H-2), 4.3 (H-2')
4	-	40.58			
5 (α)	0.79	57.54	d		12.7 (H-6)
6 (β)	1.49	19.25	tt	12.3	12.7 (H-5), 2.8 (H-7)
6' (α)	1.57		m	12.3	2.9 (H-7')
7 (α)	1.30	35.86	m	12.6	
7' (β)	1.57		m	12.6	2.9 (H-6')
8	-	41.00			
9 (α)	1.46	51.05	dd		13.3 (H-11), 3.2 (H-11')
10	-	37.91			
11 (β)	1.22	31.03	q	12.4	
11' (α)	1.82		ddd		3.2 (H-9)
12 (α)	3.65	71.90	dt		10.4 (H-13), 5.7 (H-11)
13 (β)	1.74	49.78	t		10.4 (H-12), 10.4 (H-17)
14	-	52.48			
15 (β)	1.06	31.64	tt	10.2	10.2 (H-16')
15' (α)	1.58		q		
16 (α)	1.38	27.25	dq	10.5	
16' (β)	1.92		dp	10.5	5.0 (H-15')
17 (α)	2.28	53.13	dq		10.4 (H-13), 2.9 (H-16)
19	0.92	16.75	s		
20	-	84.92			
21	1.34	22.86	s		
22	1.62	36.67	m	14.4	10.6 (H-23)
22'	1.81		ddd	14.4	10.6 (H-23), 5.8 (H-23')
23	2.06	24.24	m		
23'	2.06		m		5.8 (H-22)
24	5.10	125.81	tt		6.9 (H-23), 1.3 (H-22)
25	-	132.24			
26	1.62	17.95	s		
27	1.68	25.86	s		
30	0.86	16.75	s		
31	1.07	28.40	s		
32	0.92	16.75	s		
33	1.02	16.27	s		
12OH	4.59		s		

* - see footnote Table 5.4

Table 5.4 : Chemical shifts, Multiplicities and Coupling constants of the sugars attached to aglycone part of the glycoside : Ginsenoside-Rd

	Site	δ_H[a]	δ_C[b]	m[c]	^2J(Hz)	^3J(Hz)
Sugar-a	1	4.67	104.5	d		7.7 (H-2)
	2	3.19	77.0	t		9.1 (H-1, H-3)
	3	3.35	78.1	t		9.1 (H-2, H-4)
	4	3.21	71.9	dd		8.6 (H-3), 9.7 (H-5)
	5	3.24	78.1	ddd		9.7 (H-4), 2.2 (H-6)
	6	3.60	63.0	dd	11.9	2.2 (H-5)
	6'	3.82		dd		
Sugar-b	1	4.60	98.3	d		7.9 (H-2)
	2	3.08	75.4	t		8.3 (H-1, H-3)
	3	3.35	78.1	t		9.1 (H-2, H-4)
	4	3.32	71.4	t		9.5 (H-3, H-5)
	5	3.20	78.1	ddd		9.5 (H-4), 2.1 (H-6)
	6	3.63	62.7	dd	12.4	2.1 (H-5)
	6'	3.76		dd		
Sugar-c	1	4.43	105.4	d		7.4 (H-2)
	2	3.56	81.3	t		9.0 (H-1, H-3)
	3	3.28	71.6	t		8.2 (H-2, H-4)
	4	3.54	78.6	t		9.2 (H-3, H-5)
	5	3.25	78.1	ddd		9.2 (H-4), 1.9 (H-6)
	6	3.65	63.0	dd	11.9	1.9 (H-5)
	6'	3.85		dd		

a - δ_H is relative to $C\underline{H}_3OH$ at 3.3 ppm
b - δ_C is relative to $C\underline{D}_3OD$ at 49.4 ppm
c - multiplicities; (s) singlet, (d) doublet, (t) triplet, (q) quartet and (m) multiplet

5.7 MATERIALS AND METHODS

5.7.1 Sample preparation of glycoside; Ginsenoside-Rd

The glycoside sample was supplied by the Brain Metabolism Unit in our university. The white crystalline solid of glycoside (20mg) was examined in d_4-methanol solution (0.65ml) in a high quality NMR tube (528PP Wilmad). The glycoside sample was stored at 280K between experiments.

5.7.2 NMR experiments of Ginsenoside-Rd

All NMR experiments were performed using a 5mm proton (inverse) probe on a Varian VXR 600S spectrometer operating at 599.945 MHz for proton and 150.869 for ^{13}C. The sample temperature was regulated with a Varian VT temperature controller at 298K with 10 litre/min of dry air precooled at 273K by passage through a coil in a refrigerated bath. The spectrometer was controlled with a Sun 4/110 host computer using the VNMR system software version 4.1.

^1H single pulse, DQF COSY, LR COSY, TOCSY, ROESY, 1D TOCSY, 1D ROESY and ^{13}C-^1H heteronuclear correlation NMR spectra (HMQC and HMBC) were carried out as described below. Spectra were referenced internally to the CD_3OD resonance at 3.3 ppm.

Single pulse proton NMR spectra were acquired using the pulse sequence D1-90^0-AQ with a relaxation delay D1=1.5s and acquisition time AQ=2.5s. 512 Transients were acquired over a 3KHz spectral width and 14K data points were collected. The data sets were zero filled to 65K data points and was multiplied by optimised shifted sinebell squared function prior to Fourier transformation. Fourier transformed data were phase and baseline corrected.

The two dimensional phase-sensitive proton DQF COSY spectrum was obtained using the pulse sequence D1-90^0-t_1-90^0-90^0-AQ with a relaxation delay D1=1.5s and acquisition time AQ=0.341s. An eight step phase cycle (hypercomplex acquisition) was used. Sixteen scans were performed for each t_1 value and 512 FIDs were acquired.

Other parameters were SW=3KHz; 2K data points and 64 dummy scans. Prior to Fourier transformation, the data set was zero filled to 1024 data points in F_1 dimension and was multiplied by optimised shifted sinebell squared function in both dimensions.

The pulse sequence used for the 2D TOCSY experiment was D1-90^0-t_1-(MLEV-17)-AQ with a relaxation delay D1=1.5s, an acquisition time AQ=0.341s. An eight step phase cycle (hypercomplex acquisition) was used. The MLEV-17 spin lock pulse was used with a 65ms mixing time cycle and two trim pulses of 2ms each. Sixteen transients were performed for each t_1 value and 512 FIDs were acquired. Spectral width was 3KHz and 2K data points were collected. The data set was zero filled to 1024 data points in F_1 dimension. The data set was multiplied by shifted sinebell squared function prior to Fourier transformation. After the Fourier transformation, baseline was corrected only in F_2 dimension and phase corrections were done in both dimensions.

The 2D proton LR COSY spectrum, set up to observe long-range coupling, was obtained using the sequence D1-90^0-t_1-τ–45^0-AQ with the propagation time τ=100ms and relaxation delay D1=2s. 32 Transients were performed for each t_1 value and 512 FIDs were acquired over a 3KHz spectral width. Other parameters were AQ=0.341s; 2K data points. The data set was zero filled to 1024 data points in F_1 dimension. Prior to Fourier transformation, unshifted sinebell squared function was applied in both dimensions.

The 2D ROESY spectrum was obtained using the sequence D1-90^0-t_1-(spin-lock-30^0)-AQ with a relaxation delay D1=1.5s and acquisition time AQ=0.341s. A series of pulses of length 3.3μs and delays of length 3.3x7μs were applied to form the spin-lock period (150ms). Sixteen scans were performed for each t_1 value and 512 FIDs were acquired. 2K data points were collected over a 3KHz spectral width. The data set was transformed by zero filling in F_1 to 1024 data points before apodization. The data set was then multiplied by optimised shifted sinebell squared function in both dimensions prior to Fourier transformation.

The 1D TOCSY proton spectra were obtained using the 1D version of the 2D TOCSY pulse sequence by removing the incremental delay and introducing a low power selective pulse (ϕ=180^0) at the beginning of the sequence; D1-ϕ_{sel}-90^0-(spin-lock)-90^0-90^0-AQ with relaxation delay D1=1.5s. An eight step phase cycle was used. The MLEV-17 spin lock pulse cycle with 100ms mixing time and zfilter 10μs were used. A low power shaped selective spin inversion pulse, iburp_4, was applied (100μs) on- and off-resonance on alternate scans and the FID's alternatively added and subtracted to give a "difference" FID which, on transformation, gave spectra showing only TOCSY responses. 256 Transients were acquired over a 3KHz spectral width and 14K data points were collected. Other parameters were AQ=2.5s and 64 dummy scans.

1D ROESY spectra were acquired using the pulse sequence D1-ϕ_{sel}-90^0-90^0-(spin-lock)-90^0-AQ. An eight step phase cycle was used with a series of pulses of length 3.4μs. Spin-lock field was used during the 150ms mixing period. The selective irradiation was achieved using the low power shaped selective pulse (ϕ=90^0), tophat_8, for 100μs period. 1024 Transients were performed for each spectrum over a 3KHz spectral width. Other parameters were AQ=2.5s and D1=1.5s. Spectra show ROESY responses as positive signals; small negative signals are TOCSY responses.

The single pulse ^{13}C NMR spectrum was acquired with relaxation delay 700ms and acquisition time 0.501s. Broadband proton decoupling was employed throughout with a reduced power and 30,000 transients were performed over a 35KHz spectral width.

The DEPT data were obtained using the sequence D1-90^0(^{1}H)-D2-180^0(^{1}H); 90^0(^{13}C) - D2-ϕ(^{1}H); 180^0(^{13}C)-D2-AQ with a relaxation delay D1=1.5s and D2=3.6ms [$^1/_2$ ($^1J_{CH}$)]. Broadband proton decoupling was employed during the data acquisition with a reduced power level. 7K Transients were acquired over a 35KHz spectral width and 35K data points were collected.

The 2D proton detected one-bond ^{13}C-^{1}H correlation (HMQC) spectrum was obtained using the sequence; D1-90°(^{1}H)-D2-180°(^{1}H); 180°(^{13}C)-D2-90°(^{1}H)-D3-90°(^{1}H)-D2-90°(^{13}C)-$t_{1/2}$-180°(^{1}H)-$t_{1/2}$-90°(^{13}C)-D2-AQ. The delays used were D1=1.5s, D2=3.6ms [$^{1}/_{2}$ ($^{1}J_{CH}$)] and D3=400ms (null period to minimise signals from protons bonded to ^{12}C nuclei). The experiment was preceded by 256 dummy scans to establish thermal equilibrium. A 16 step phase cycle (hypercomplex acquisition) was used with ^{13}C broadband decoupling during acquisition of the proton signals. 128 Scans were performed for each t_1 value and 512 FIDs were collected. Other parameters were SW(^{1}H)=3KHz; SW(^{13}C)=35KHz and 2K data points.

The 2D proton detected multiple-bond ^{13}C-^{1}H correlation (HMBC) spectrum was obtained using the sequence; D1-90°(^{1}H)-D2-90°(^{13}C)-D3-90°(^{13}C)-$t_{1/2}$-180°(^{1}H)-$t_{1/2}$-90°(^{13}C)-AQ. The delays used were D1=1.5s, D2=3.6ms [$^{1}/_{2}$ ($^{1}J_{CH}$)] and D3=60ms (optimised for signals from protons with couplings to carbon of ca. 8 Hz). A 16 step phase cycle (hypercomplex acquisition) was used with no ^{13}C broadband decoupling during acquisition of the proton signals. 32 Increments with 480 scans for each FID were obtained. Others parameters were as for the HMQC experiment. The HMQC and HMBC data were both processed using shifted sinebell squared functions in both dimensions with zero filling of the F_1 data to 512W before transformation.

REFERENCES

1. G.N. Ramachandran, V. Sasisekharan; *Adv. Protein Chem.*, 1968, 23, 283
2. G.D. Rose, L.M. Gierasch, J.A. Smith; *Adv. Protein Chem.*, 1985, 37, 1
3. A. Aubry, M. Marraud; *Bioplolymers*, 1983, 22, 341
4. L.M. Jackman, S. Sternhell; Applications of NMR Spectroscopy in Organic Chemistry, 1969, Pergamon press, Oxford.
5. C. Cantor, P.R. Schimmel; Biophysical Chemistry, 1980, W.H. Freeman, San Fancisco, USA
6. N. Greenwood, G.D. Fasman; *Biochemistry*, 1969, 8, 4108
7. K. Wüthrich, G. Wider, G. Wagner, W. Braun; *J. Mol. Biol.*, 1982, 155, 311
8. R.R. Ernst, G. Bodenhausen, A. Wokaun; Principles of Nuclear Magnetic Resonance in One and Two Dimensions, 1986, Oxford Univ. press, London.
9. H. Kessler, M. Gehrke, C. Griesinger; *Angew. Chem. Int. Ed. Engl.*, 1988, 27, 490
10. K. Wüthrich; NMR of Proteins and Nucleic Acids, 1986, Wiley, New York.
11. G. Wagner, D. Brühwiler; *Biochemistry*, 1986, 25, 5839
12. J. Glushka, D. Cowburn; *J. Am. Chem. Soc.*, 1987, 109, 7879
13. J.L. Markley; *Meth. Enzymol.*, 1989, Vol.176, 12
14. M. Kainosho, T. Tsuji; *Biochemistry*, 1982, 21, 6273
15. W.M. Westler, B.J. Stockman, Y. Hosoya, Y. Miyake, M. Kainosho, J.L. Markley; *J. Am. Chem. Soc.*, 1988, 110, 6256
16. B.J. Stockman, W.M. Westler, P. Darba, J.L. Markley; *J. Am. Chem. Soc.*, 1988, 110, 4096
17. B.J. Stockman, M.D. Reily, W.M. Westler, E.L. Ulrich, J.L. Markley; *Biochemistry*, 1989, 28, 230
18. E.M. Purcell, H.C. Torrey, R.V. Pound; *Phys. Rev.*, 1946, 69, 37
19. E.M. Purcell; *Phys. Rev.*, 1946, 69, 681
20. F. Bloch, W.M. Hansen, M. Packard; *Phys. Rev.*,1946, 69, 127
21. F. Bloch, W.M. Hansen, M. Packard; *Phys. Rev.*,1946, 70, 474
22. W.G. Proctor, F.C. Yu; *Phys. Rev.*, 1950, 77, 717
23. W.C. Dickinson; *Phys. Rev.*, 1950, 77, 736
24. R.R. Ernst, W.A. Anderson; *Rev. Sci. Instrum*, 1966, 37, 93
25. J. Jeener; Ampère Summer School, Basco Polje, Yugoslavia, 1971
26. D.I. Hoult; *J. Magn. Reson.*, 1976, 21, 237
27. G. Wider, S. Macura, A. Kumar, R.R. Ernst, K. Wüthrich; *J. Magn. Reson.*, 1984, 56, 207
28. I.D. Campbell, C.M. Dobson, R.G. Ratcliffe; *J. Magn. Reson.*, 1977, 27, 455
29. J.D. Stoesz, A.G. Redfield, D. Malinowski; *FEBS Lett.*, 1978, 91, 320

30. R.M. Keller, R. Baumann, E.H. Hunziker-Kwik, F.J. Joubert, K. Wüthrich; *J. Mol. Biol.*, 1983, 163, 623
31. D. Marion, K. Wüthrich; *Biochem. Biophys. Res. Commun.*, 1983, 113, 967
32. A. Bax; Two Dimensional Nuclear Magnetic Resonance in Liquids, 1982, Reidel, London.
33. A.E. Derome; Modern NMR Techniques for Chemistry Research, 1987, Pergamon press, England.
34. H. Friebolin; Basic One and Two Dimensional NMR Spectroscopy, 1991, VCH publishers, FRG.
35. R. Freeman, G.A. Morris; *Bull. Magn. Reson.*, 1979, 1, 5
36. K. Nagayama; *Adv. Biophys.*, 1981, 14, 139
37. O.W. Sørensen, G.W. Eich, M.H. Levitt, G. Bodenhausen, R.R. Ernst; *Progr. Nucl. Magn. Reson. Spectros.*, 1983, 16, 163
38. G.A. Morris; *Magn. Reson. Chem.*, 1986, 24, 371
39. K. Nagayama, A. Kumar, K. Wüthrich, R.R. Ernst; *J. Magn. Reson.*, 1980, 40, 321
40. A. Bax, R. Freeman; *J. Magn. Reson.*, 1981, 44, 542
41. U. Piantini, O.W. Sørensen, R.R. Ernst; *J. Am. Chem. Soc.*, 1982, 104, 6800
42. M. Rance, O.W. Sørensen, G. Bodenhausen, G. Wagner, R.R. Ernst, K. Wüthrich; *Biochem. Biophys. Res. Commun.*,1983, 117, 479
43. A.J. Shaka, R. Freeman; *J. Magn. Reson.*, 1983, 51, 169
44. D.J. States, R.A. Haberkorn, D.J. Ruben; *J. Magn. Reson.*, 1982, 48, 286
45. G. Wagner; *J. Magn. Reson.*, 1983, 55, 151
46. A. Bax, G. Drobny; *J. Magn. Reson.*, 1985, 61, 306
47. M.H. Levitt, R. Freeman, T. Frenkiel; *J. Magn. Reson.*, 1982, 47, 328
48. L. Braunschweiler, R.R. Ernst; *J. Magn. Reson.*, 1983, 53, 521
49. A. Bax, D.G. Davis, *J. Magn. Reson.*, 1985, 65, 355
50. J. Jeener, B.H. Meier, P. Bachmann, R.R. Ernst; *J. Chem. Phys.*, 1979, 71, 4546
51. A. Kumar, R.R. Ernst, K. Wüthrich; *Biochem. Biophys. Res. Commun.*, 1980, 95, 1
52. J.H. Noggle, R.E. Schirmer; The Nuclear Overhauser Effect, 1971, Academic press, New York.
53. S. Macura, K. Wüthrich, R.R. Ernst; *J. Magn. Reson.*, 1982, 46, 269
54. C.M. Dobson, P.A. Evans, K.L. Williamson; *FEBS Lett.*, 1984, 168, 1323
55. A. Bax, D.G. Davis; *J. Magn. Reson.*, 1985, 63, 207
56. H. Kessler, C. Griesinger, R. Kerssebaum, G. Wagner, R.R. Ernst; *J. Am. Chem. Soc.*, 1987, 109, 607
57. A.A. Bothner-By, R.L. Stephens, J. Lee, C.D. Warren, R.W. Jeanloz; *J. Am. Chem. Soc.*, 1984, 106, 811
58. G.P. Jones; *Phys. Rev.*, 1966, 148, 332
59. B.T. FarmerII, S. Macura, L.R. Brown; *J. Magn. Reson.*, 1988, 80, 1

60. B.T. FarmerII, S. Macura, L.R. Brown; *J. Magn. Reson.*, 1987, 72, 347
61. A. Bundi, K. Wüthrich; *Biopolymers*, 1979, 18, 285
62. A.D. Kaline, K. Wüthrich; *J. Mol. Biol.*, 1986, 192, 869
63. D. Neuhaus, G. Wagner, M. Vasak, J.H.R. Kägi, K. Wüthrich; *Eur. J. Biochem.*, 1985, 151, 257
64. G. Wagner, D. Neuhaus, E. Wörgötter, M. Vasak, J.H.R. Kägi, K. Wüthrich; *J. Mol. Biol.*, 1986, 187, 131
65. N.J. Oppenheimer, T.L. James; *Meth. Enzymol.*, 1989, Vol. 176 and 177
66. T.F. Havel, I.D. Kuntz, G.M. Crippen; *Bull. Math. Biol.*, 1983, 45, 665
67. G.M. Crippen, T.F. Havel; Distance Geometry and Molecular Conformation, 1988, Research Studies Press, Letchworth, UK.
68. T.F. Havel; *Prog. Biophys. Mol. Biol.*, 1991, 56, 43
69. S.A. Sherman, M.E. Johnson; *Prog. Biophys. Mol. Biol.*, 1993, 59, 285
70. R.P. Sheridan, R. Nilakantan, J.S. Dixon, R. Venkataraghavan; *J. Med. Chem.*, 1986, 29, 899
71. M. Billeter, T.F. Havel, I.D. Kuntz; *Biopolymers*, 1986, 26, 777
72. T.F. Havel, M. Snow; *J. Mol. Biol.*, 1991, 217, 1
73. M.P. Williamson, T.F. Havel, K. Wüthrich; *J. Mol. Biol.*, 1985, 182, 295
74. T.F. Havel, K. Wüthrich; *Bull. Math. Biol.*, 1984, 46, 673
75. K. Wüthrich; *Science*, 1989, 45, 243
76. H. Widmer, M. Billeter, K. Wüthrich; *Proteins*, 1989, 6, 357
77. M. Karplus; *J. Phys. Chem.*, 1959, 30, 11
78. A. Pardi, M. Billeter, K. Wüthrich; *J. Mol. Biol.*, 1984, 180, 741
79. T.F. Havel, K. Wüthrich; *J. Mol. Biol.*, 1985, 182, 281
80. J.M. Blaney, G.M. Crippen, A. Dearing. J.S. Dixon; *QCPE Bull.*, 1990, 10, 37
81. D. Hare; (DSPACE) Infinity Systems Research Inc, 14810, 216th Ave NE, Woodinville WA 98072, USA
82. B.S. Duncan, B.G. Buchanan, B. Hayes-Roth, O. Liclitarge, R. Altman, J. Brinkley, M. Hewett, C. Cornelius, O. Jardatzky; *Bull. Magn. Reson.*, 1986, 8, 111
83. J. Brinkley, R. Altman, B.S. Duncan, B.G. Buchanan, O. Jardatzky; *J. Chem. Inf. Comput. Sci.*, 1988, 28, 194
84. O. Lichtarge, C.W. Cornelius, B.G. Buchanan, O. Jardatzky; *Proteins: Struct. Function Genet.*, 1987, 2, 340
85. W. Braun, N. Go; *J. Mol. Biol.*, 1985, 186, 611
86. P. Güntert, W. Braun, K. Wüthrich; *J. Mol. Biol.*, 1991, 217, 517
87. T. Schaumann, W. Braun, K. Wüthrich; *Biopolymers*, 1990, 29, 679
88. R. Kaptein, E.R.P. Zuiderweg, R.M. Scheek, R. Boelens, W.F. van Gunsteren; *J. Mol. Biol.*, 1985, 182, 179

89. G.M. Clore, A.M. Gronenborn, A.T. Brünger, M. Karplus; *J. Mol. Biol.*, 1985, 186, 435
90. R.M. Levy, M. Karplus; *Biopolymers*, 1979, 18, 2465
91. M. Karplus; *Meth. Enzymol.*, 1986, 131, 283
92. M. Clark, R.D. Cramer III, N. van Opdenbosch; *J. Comp. Chem.*, 1989, 10, 982
93. U.C. Singh, P.K. Weiner, J.W. Caldwell, P.A. Kollman; AMBER (UCSF), Dept. of Pharmaceutical Chemistry, Univ. fo California, San Francisco, USA.
94. B.R. Brooks, M. Karplus; *J. Chem. Phys.*, 1983, 79, 6312
95. B.R. Brooks, R.E. Bruccoleri, B.D. Olafson, D.J. States, S. Swaminathan, M. Karplus; *J. Comp. Chem.*, 1983, 4, 187
96. W.F. van Gunsteren, H.J.C. Berendsen, The GROMOS software package, Nijenborgh 16, 9747 AG, Groningen, Netherlands.
97. H.J.C. Berendsen, J.P.M. Postma, W.F. van Gunsteren, A. DiNola, J.R. Haak; *J. Chem. Phys.*, 1984, 81, 3684
98. J.A. McCammon, M. Karplus; *Ann. Rev. Phys. Chem.*, 1980, 31, 29
99. M. Karplus, B. Bunsenges; *Phys. Chem.*, 1982, 86, 386
100. C.L. Brooks, M. Karplus, B.M. Pettit; *Adv. Chem. Phys.*, 1988, 71, 1
101. P.A. Kollman; *Ann. Rev. Phys. Chem.*, 1987, 38, 303
102. A.E. Howard, P.A. Kollman; *J. Med. Chem.*, 1988, 31, 1669
103. G.M. Clore, M. Nilges, A.T. Brünger, M. Karplus, A.M. Gronenborn; *FEBS Lett.*, 1987, 213(2), 269
104. A.T. Brünger, G.M. Clore, A.M. Gronenborn, M. Karplus; *Protein Eng.*, 1987, 1, 399
105. M. Nilges, G.M. Clore, A.M. Gronenborn; *FEBS Lett.*, 1988, 239(1), 129
106. P.C. Driscoll, A.M. Gronenborn, L. Beress, G.M. Clore; *Biochemistry*, 1989, 28, 2188
107. S. Kirkpatrick, C.D. Gelatt, M.P. Vecchi; *Science*, 1983, 220, 671
108. A. Khachaturyan; *J. Math. Phys.*, 1986, 27, 1834
109. B.R. Brooks, R.E. Bruccoleri, B.D. Olafan, D.J. States, S. Swaminathan, M. Karplus; *J. Comp. Chem.*, 1983, 4, 187
110. M. Nilges, G.M. Clore, A.M. Gronenborn; *FEBS Lett.*,1988, 229, 317
111. B. Borgias, T. James; *J. Magn. Reson.*, 1988, 79, 493
112. K.M. Banks, D. Hare, B.R. Reid; *Biochemistry*, 1989, 28, 6996
113. M. Summers, T. South, B. Kim, D. Hare; *Biochemistry*, 1990, 29, 329
114. S. Edmondson, N. Khan, J. Shriver, J. Zdunek, A. Gräslund; *Biochemistry*, 1991, 30, 11271
115. R.M. Scheek, W.F. van Gunsteren, R. Kaptein; *Meth. Enzymol.*, 1989, 177, 204

116. G.M. Clore, A.T. Brünger, M. Karplus, A.M. Gronenborn; *J. Mol. Biol.*, 1986, 191, 523
117. J. de Vileg, R.M. Scheek, W.F. van Gunsteren, H.J.C. Berendsen, R. Kaptein, J. Thomason; *Proteins*, 1988, 3, 209
118. G.M. Clore, A.M. Gronenborn, M. Kjaer, F.M. Poulsen; *Protein Eng.*, 1987, 1, 305
119. G.M. Clore, D.K. Sukumaran, M. Nilges, A.M. Gronenborn; *Biochemistry*, 1987, 26, 1732
120. A.M. Doherty; *Annu. Rep. Med. Chem.*, 1991, 26, 83
121. T.F. Luscher; *Am. J. Hypertens.*, 1990, 3, 317
122. T. Masaki, M. Yanagisawa, A. Inoue, Y. Takuwa, K. Goto, S. Kimura; *J. Cell. Bio. Chem.*, 1990, 14E, 199
123. M. Yanagisawa, H. Kurihara, S. Kimura, Y. Tomobe, M. Kobayashi, Y. Mitsui, Y. Yazaki, K. Goto, T. Masaki; *Nature (London)*, 1988, 332, 411
124. Y. Itoh, M. Yanagisawa, S. Ohkubo, C. Kimura, T. Kosaka, A. Inoue, N. Ishida, Y. Mitsui, H. Onda, M. Fujino; *FEBS Lett.*, 1988, 231, 440
125. K. Saida, Y. Mitsui, N. Ishida; *J. Biol. Chem.*, 1989, 264, 14613
126. A. Bdolah, Z. Wollberg, G. Fleminger, E. Kochva; *FEBS Lett.*, 1989, 256, 1
127. A. Beeker, E.B. Dowdle, U. Hechler, K. Kauser, P. Donner, W-D. Schleuning; *FEBS Lett.*, 1993, 315, 100
128. M. Yanagisawa, T. Masaki; *Trends in Pharmacol. Sci.*, 1989, 10, 374
129. Y. Kloog, M. Sokolovsky; *Trends in Pharmacol. Sci.*, 1989, 10, 212
130. M. Yanagisawa, A. Inoue, T. Ishikawa, Y. Kasuya, S. Kimura, S. Kumagaye, K. Nakajima, T.X. Watanabe, S. Sakakibara, K. Goto; *Proc. Natl. Acad. Sci. U.S.A.*, 1988, 85, 6964
131. A.M. Doherty; *J. Med. Chem.*, 1992, 35(9), 1493
132. S. Kimura, Y. Kasuya, T. Sawamura, O. Shinimi, Y. Sugita, M. Yanagisawa, K. Goto, T. Masaki; *J. Cardiovasc. Pharmacol.*, 1989, 13, S5-S7.
133. M.S. Fabrini, A. Vitale, E. Pedrazzini, G. Nitti, M. Zamai, M. Tamburin, V.R. Caiolfa, C. Patrono, L. Benatti; *Proc. Natl. Acad. Sci. U.S.A.*, 1992, 186, 753
134. M.S. Fabrini, A. Vitale, E. Pedrazzini, G. Nitti, M. Zamai, M. Tamburin, V.R. Caiolfa, C. Patrono, L. Benatti; *Proc. Natl. Acad. Sci. U.S.A.*, 1993, 90, 3923
135. T. Sawamura, S. Kimura, O. Shinmi, Y. Sugita, M. Yanagisawa, T. Masaki; *Biochem. Biophys. Res. Commun.*, 1989, 162, 1287
136. K. Kitamura, T. Yukawa, S. Morita, Y. Ichiki, T. Eto, K. Tanaka; *Biochem. Biophys. Res. Commun.*, 1990, 170, 497
137. S. Ohkubo, K. Ogi, M. Hosoya, H. Matsumoto, N. Suzuki, C. Kimura, H. Ondo, M. Fujino; *FEBS Lett.*, 1990, 274, 136

138. T. Miyauchi, Y. Tomobe, R. Shiba, T. Ishikawa, M. Yanagisawa, S. Kimura, Y. Sugishita, I. Ito, K. Goto, T. Masaki; *Circulation*, 1990, 81, 1874
139. T. Masaki, M. Yanagisawa; *Cardiovasc. Drug Rev.*, 1990, 8, 373
140. A. Hoffman, E. Grossman, K.P. Ohman, E. Marks, H.R. Keiser; *Am. J. Hypertension*, 1990, 3, 789
141 H. Vierhapper, O. Wagner, P. Nowotny, W. Waldhausl; *Circulation*, 1990, 81, 1415
142. T. Watanabe, N. Suzuki, N. Shimatomo, M. Fujino, A. Imada; *Nature (London)*, 1990, 344, 114
143. P.G. Cavero, W.L. Miller, D.M. Heublein, K.B. Margulies, J.C. Burnett; *Am. J. Physiol.*, 1990, 259, F312
144. C. Koseki, M. Imai, Y. Hirata, M. Yanagisawa, T. Masaki; *Am. J. Physiol.*, 1989, 256, R858
145. E.R. Martin, P.A. Marsden, B.M. Brenner, B.J. Ballermann; *Biochem. Biophys. Res. Commun.*, 1989, 162, 130
146. A. Semsen, O. Larsson, J.M. Lundberg; *Eur. J. Pharmacol.*, 1991, 208, 313
147. M. Sugiura, R.M. Snajdar, M. Schwartzberg, K.F. Badr, T. Inagami; *Biochem. Biophys. Res. Commun.*, 1989, 162, 1396
148. H. Arai, S. Hori, I. Aramori, H. Ohkubo, S. Nakanishi; *Nature (London)*, 1990, 348, 730
149. T. Sakurai, M. Yanagisawa, Y. Takuwa, H. Miyazaki, S. Kimura, K. Goto, T. Masaki; *Nature (London)*, 1990, 348, 732
150. S.C. Gomez, E.N. Cozza, M.F. Foecking, S. Chiou, M.W. Ferris; *Hypertension*, 1990, 15, 744
151. T. Emori, Y. Hirata, F. Marumo; *FEBS Lett.*, 1990, 263, 261
152. S. Karne, C.K. Jayawickreme, M.R. Lerner; *J. Biol. Chem.*, 1993, 268, 19126
153. Y. Saito, T. Mizuno, M. Itakura, Y. Suzuki, T. Ito, H. Hagiwara, S. Hirose; *J. Biol. Chem.*, 1991, 266, 23433
154. M.J. Sumner, T.R. Cannon, J.W. Mundin, D.G. White, I.S. Watts; *Br. J. Pharmacol.*, 1992, 107, 858
155. M. Clozel, G.A. Gray, V. Breu, B-M. Loffler, R. Osterwalder; *Biochem. Biophys. Res. Commun.*, 1992, 186, 867
156. T. Sakurai, M. Yanagisawa, T. Masaki; *Trends in Pharmacol. Sci.*, 1992, 13, 103
157. T.F. Luscher, B.S. Oemar, C.M. Boulanger, A.W.A. Hahn; *J. Hypertension*, 1993, 11, 7 and 121
158. X.-M. Cheng, S.S. Nikam, A.M. Doherty; *Curr. Med. Chem.*, 1994, 1(4), 271
159. D.L. Wiliams, K.L. Jones, D.J. Pettibone, E.V. Lis, B.V. Clineschmidt; *Biochem. Biophys. Res. Commun.*, 1991, 175, 556

160. R. Takayanagi, K. Kitazumi, C. Takasaki, K. Ohnaka, S. Aimoto, K. Tasaka, M. Ohashi, H. Nawata; *FEBS Lett.*, 1991, 282, 103
161. T. Saeki, M. Ihara, T. Fukuroda, M. Yamagiwa, M. Yano; *Biochem. Biophys. Res. Commun.*, 1991, 179, 286
162. R.L. Panek, T.C. Major, G.P. Hingorani, A.M. Doherty, D.G. Taylor, S.T. Rapundalo; *Biochem. Biophys. Res. Commun.*, 1992, 183, 566
163. T. Masaki, M. Yanagisawa, K. Goto; *Med. Res. Review*, 1992, 12, 391
164. M. Sokolovsky; *J. Neurochem.*, 1992, 59, 809
165. L. Jiang; Personal communication, Ph. D. Thesis, University of Edinburgh, 1996
166. V. Saudek, J. Hoflack, J.T. Pelton; *FEBS Lett.*, 1989, 257, 145
167. S. Endo, H. Innoka, Y. Ishibashi, C. Kitada, O. Mizuta, M. Fujino; *FEBS Lett.*, 1989, 257(1), 149
168. V. Saudek, J. Hoflack, J.T. Pelton; *Int. J. Peptide Protein Res.*,1991, 174
169. A. Aumelas, L. Chiche, E. Mahe, D. L-Nguyen, P. Sizun, P. Berthault, B. Perly; *Int. J. Peptide Protein Res.*,1991, 37, 315
170. S.L. Munro, D.J. Craik, D. McConville, J.G. Hall, M. Searle, W. Bicknell, D. Scanlon, C. Chandler; *FEBS Lett.*,1991, 278(1), 9
171. S.R. Krystek, D.A. Bassolino, J. Novotny, C. Chen, T.M. Marschner, N.H Andersen; *FEBS Lett.*, 1991, 281, 212
172. H. Tamaoki, Y. Kobayashi, S. Nishimura, T. Ohkubo, Y. Kyoguku, K. Nakajima, S. Kumagaye, T. Kimura, S. Sakakibara; *Protein Eng.*, 1991, 4(5), 509
173. M.D. Reily, Jr.J.B. Dunbar; *Biochem. Biophys. Res. Commun.*, 1991, 178(2), 570
174. N.H. Andersen, C. Chen, T.M. Marschner, S.R. Krystek, D.A. Bassolino; *Biochemistry*, 1992, 31, 1280
175. M. Coles, S.L.A. Munro, D.J. Craik; *J. Med. Chem.*, 1994, 37, 656
176. Y. Kobayashi; *Proc. Eleventh. American. Peptide. Symposium*, 1989, July 9-14, La Jolla, [J.E. Rivier & G.R. Marshall eds, ESCOM], 1990, 552
177. S.C. Brown, M.E. Donlan, P.W. Jeffs; *Proc. Eleventh. American. Peptide. Symposium*, 1989, July 9-14, La Jolla, [J.E. Rivier & G.R. Marshall eds, ESCOM], 1990, 595
178. D.C. Dalgarno, L. Slater, S. Chackalamannil, M.M. Senior; *Int. J. Peptide Protein Res.*, 1992, 40(6), 515
179. E. Ragg, R. Mondelli, S. Penco, G. Bolis, L. Baumer, A. Guaragna; *J. Chem. Soc. Perkin Trans.*, 1994, 2, 1317
180. R.G. Milles, S.L. O'Donoghue, R. Smith, G.F. King; *Biochemistry*, 1992, 31, 5640
181. P. Bortmann, J. Hoflack, J.T. Pelton, V. Saudek; *Neurochem. Int.*, 1991, 18(4), 491
182. M.L. Donlan, F.K. Brown, P.W. Jeffs; *J. Biomol. NMR*, 1992, 2, 407

183. J.H. Pease, D.E. Wemmer; *Biochemistry*, 1988, 27, 8491
184. V.F. Bystrov, V.V. Okhanov, A.I. Miroshnikov, Y.A. Ovchinnikov; *FEBS Lett.*, 1980, 119, 113
185. D. Wemmer, N.R. Kallenbach; *Biochemistry*, 1983, 22, 1901
186. A. Aumelas, L. Chiche, E. Mahe, D. L-Nguyen; *Neurochem. Int.*, 1991, 18(4), 471
187. A.F. Atkins, R.C. Martin, R. Smith; *Biochemistry*, 1995, 34, 2026
188. R.G. Mills, A.R. Atkins, T. Harvey, F.K. Junius, R. Smith, G.F. King; *FEBS Lett.*, 1991, 282(2), 247
189. R.G. Mills, G.B. Ralston, G.F. King; *J. Biol. Chem.*, 1994, 269(38), 23413
190. R.W. Janes, D.H. Peapus, B.A. Wallace; *Nature Struct. Biol.*, 1994, 1, 311
191. S. Kimura, Y. Kasuya, T. Sawamura, O. Shinmi, Y. Sugita, M. Yanagisawa, K. Goto, T. Masaki; *Biochem. Biophys. Res. Commun.*, 1988, 156, 1181
192. S.-I. Kumagaye, H. Kuroda, K. Nakajima, T.X. Watanabe, T. Kimura, T. Masaki, S. Sakakibara; *Int. J. Peptide Protein Res.*, 1988, 32, 519
193. S.-I. Kumagaye, K. Nakajima, H. Nishio, H. Kuroda, T.X. Watanabe, T. Kimura, T. Masaki, S. Sakakibara; *Peptide Chemistry*, Protein Research Foundation, Osaka, Japan, 1988, 215
194. Y. Hirata, H. Yoshimi, T. Emori, M. Shichiri, F. Marumo, T.X. Watanabe, S.-I. Kumagaye, K. Nakajima, T. Kimura, S. Sakakibara; *Biochem. Biophys. Res. Commun.*, 1989, 160, 228
195. D.S. Wishart, B.D. Sykes, F.M. Richards; *Biochemistry*, 1992, 31, 1647
196. M. Billeter, W. Braun, K. Wüthrich; *J. Mol. Biol.*, 1982, 155, 321
197. K. Wüthrich, M. Billeter, W. Braun; *J. Mol. Biol.*, 1983, 169, 949
198. Tripos Molecular Modeling Software, SYBYL ver. 6.1, Tripos Associates, St. Louis, MO, USA.
199. C.M. Hewage; M. Phil. Thesis, Dept. of Chemistry, Univ. of Peradeniya, Peradeniya, Sri Lanka, 1990, 4
200. H. Yamaguchi, H. Matsuura, R. Kasai, O. Tanaka, M. Satake, H. Koda, H. Izumi, M. Nuno, S. Katsuki; *Chem. Pharm. Bull.*, (Tokyo), 1988, 36, 4177
201. H. Matsuura, R. Kasai, O. Tanaka, Y. Saruwatari, T. Fuwa, J. and Zhou; *Chem. Pharm. Bull.*, (Tokyo), 1983, 31, 2281
202. S. Taniyasu, O. Tanaka, T. Yang, J. Zhou; *Planta Med.*, 1982, 44, 124
203. T.R. Yang, R. Kasai, J. Zhou, O. Tanaka; *Phytochemistry*, 1983, 22, 1473
204. J.X. Wei, J.F. Wang, L.Y. Chang, Y.C. Du; *Acta. Pharm. Sin.*, 1980, 15, 359
205. J.X. Wei, L.A. Wang, H. Du, R Li; *Acta. Pharm. Sin.*, 1985, 20, 288
206. P. Zhao, Y.Q. Liu, C.R. Yang; *Acta. Botanica Yunnanica*, 1993, 15(4), 409
207. J. Zhou, M.Z. Wu, S. Taniyama, H. Besso, O. Tanaka, Y. Saruwatari, T. Fuwa; *Chem. Pharm. Bull.* (Tokyo), 1981, 29, 2844-2850

208. J.X. Wei, B. Li, X.B. Ma; *Bull. Chin. Mater. Med.*, 1984, 9, 267
209. J.X. Wei, L.Y. Chang, J.F. Wang, E. Friedrichs, M. Jores, H. Puff; *Planta Med.*, 1982, 45, 167
210. J.X. Wei, L.Y. Chang, J.F. Wang, W.S. Chen, E. Friedrichs, H. Puff, E. Breitmaier; *Planta Med.*, 1984, 50, 47
211. W. Tang, G. Eisenbrand; Chinese Drugs of Plant Origin, Springer-Verlag, Germany, 1992, 711
212. S. Fulder; The Root of Being. Ginseng and the Pharmacology of Harmony, Hutchinson, London., 1980
213. S. Garriques; *Ann. Chem. Pharm.*, 1854, 90, 231
214. Y. Asahina, B. Taguchi; *Yakugaku-Zasshi* (*J. Pharm. Soc. Japan*), 1906, 26, 549
215. H. Kondo, U. Amano; *Yakugaku-Zasshi* (*J. Pharm. Soc. Japan*), 1920, 40, 1027
216. S. Shibata, M. Fujita, H. Itokawa, O. Tanaka, T. Ishii; *Chem. Pharm. Bull.*, (Tokyo), 1963, 11, 759
217. S. Shibata, O. Tanaka, M. Nagai, T. Ishii; *Tetrahedron Lett.*, 1962, 26, 1239-1242
218. M.Z. Wu; *Acta. Bot. Yunnan.*, 1979, 1, 119
219. E. Bombardelli, B. Bonati, B. Gabetta, E.M. Martinelli, G. Mustich; *Fitoterapia*, 1976, 47, 99
220. R. Kasai, K. Matsuura, O. Tanaka, S. Sanada, J. Shoji; *Chem. Pharm. Bull.*, 1977, 25, 3277
221. O. Tanaka, S. Yahara; *Phytochemistry*, 1978, 17, 1353
222. H. Besso, R. Kasai, Y. Saruwatari, T. Fuwa, O. Tanaka; *Chem. Pharm. Bull.*, 1982, 30(7), 2380
223. H. Matsuura, R. Kasai, O. Tanaka, Y. Saruwatani, K. Kunihiro, T. Fuwa; *Chem. Pharm. Bull.*, 1984, 32, 1188
224. J. Asakawa, R. Kasai, K. Yamasaki, O. Tanaka; *Tetrahedron*, 1977, 33, 1935
225. B.H. Han, M.H. Park, Y.N. Han, L.K. Woo, U. Sankawa, S. Yahara, O. Tanaka; *Planta Med.*, 1982, 44, 146
226. S. Yahara, O. Tanaka, I. Nishioka; *Chem. Pharm. Bull.*, 1978, 26, 3010
227. R. Kasai, M. Suzuo, J. Asakawa, O. Tanaka; *Tetrahedron Lett.*, 1977, 175
228. R. Kasai, M. Okihara, J. Asakawa, K. Mizutani, O. Tanaka; *Tetrahedron*, 1979, 35, 1427
229. K. Mizutani, R. Kasai, O. Tanaka; *Carbohydrate Research*, 1980, 87, 19
230. H.P. Zhou; *Chin. Pharm. Bull.*, 1988, 23, 720

Appendix I

Additional NMR and structure calculation details of ET-1

Table I.a : NOE constraints of ET-1 used for DIANA calculation : Peak numbers, peak assignments (F2 F1), lower constraints and upper constraints.

		1 ASG_POS1	2 ASG_POS2	3 LOWER	4 UPPER
1	P1	SER4.H	CYS3.HA	1.80	5.00
2	P2	SER4.H	SER4.HA	1.80	5.00
3	P3	SER4.H	SER4.HB1	1.80	5.00
4	P4	SER4.H	SER4.HB2	1.80	5.00
5	P5	SER4.H	CYS3.HB1	1.80	5.00
6	P6	SER4.H	CYS3.HB2	1.80	5.00
7	P7	SER2.H	SER2.HA	1.80	5.00
8	P8	SER2.H	CYS1.HA	1.80	3.30
9	P9	SER2.H	SER2.QB	1.80	5.60
10	P10	SER2.H	CYS1.QB	1.80	5.60
11	P11	LEU6.H	SER5.HA	1.80	5.00
12	P12	LEU6.H	LEU6.HA	1.80	5.00
13	P13	LEU6.H	LEU6.QB	1.80	5.60
14	P14	LEU6.H	LEU6.HG	1.80	3.30
15	P15	LEU6.H	LEU6.QQD	1.80	6.56
16	P16	CYS15.H	CYS15.HA	1.80	5.00
17	P17	CYS15.H	PHE14.HA	1.80	5.00
18	P18	CYS15.H	VAL12.HA	1.80	5.00
		1 ASG_POS1	2 ASG_POS2	3 LOWER	4 UPPER
19	P19	CYS15.H	CYS15.HB1	1.80	5.00
20	P20	CYS15.H	PHE14.HB1	1.80	5.00
21	P21	CYS15.H	PHE14.HB2	1.80	5.00
22	P22	CYS15.H	CYS15.HB2	1.80	5.00
23	P23	GLU10.H	GLU10.HA	1.80	5.00
24	P24	GLU10.H	LYS9.HA	1.80	5.00
25	P25	GLU10.H	GLU10.QG	1.80	5.60
26	P26	GLU10.H	GLU10.QB	1.80	3.90
27	P27	GLU10.H	LYS9.HB1	1.80	5.00
28	P28	LYS9.H	LYS9.HA	1.80	5.00
29	P29	LYS9.H	LYS9.HB1	1.80	5.00
30	P30	LYS9.H	LYS9.HB2	1.80	5.00
31	P31	LYS9.H	LYS9.QG	1.80	5.60
32	P32	ASP18.H	ASP18.HA	1.80	5.00
33	P33	ASP18.H	LEU17.HA	1.80	2.50
34	P34	ASP18.H	ASP18.HB1	1.80	5.00
35	P35	ASP18.H	ASP18.HB2	1.80	5.00
36	P36	ASP18.H	LEU17.QB	1.80	5.60
37	P37	ASP18.H	LEU17.HG	1.80	5.00
		1 ASG_POS1	2 ASG_POS2	3 LOWER	4 UPPER
38	P38	ASP18.H	LEU17.QQD	1.80	6.56
39	P39	PHE14.H	PHE14.HA	1.80	3.30
40	P40	PHE14.H	TYR13.HA	1.80	2.50
41	P41	PHE14.H	PHE14.HB1	1.80	2.50
42	P42	PHE14.H	PHE14.HB2	1.80	3.30
43	P43	PHE14.H	TYR13.QB	1.80	5.60
44	P44	CYS3.H	CYS3.HA	1.80	5.00
45	P45	CYS3.H	SER2.HA	1.80	5.00
46	P46	CYS3.H	SER2.QB	1.80	5.60
47	P47	CYS3.H	CYS3.HB1	1.80	5.00
48	P48	CYS3.H	CYS3.HB2	1.80	5.00
49	P49	VAL12.H	CYS11.HA	1.80	5.00
50	P50	VAL12.H	LYS9.HA	1.80	5.00
51	P51	VAL12.H	VAL12.HA	1.80	5.00
52	P52	VAL12.H	CYS11.QB	1.80	3.90
53	P53	VAL12.H	VAL12.HB	1.80	3.30
54	P54	VAL12.H	VAL12.QG1	1.80	6.00
55	P55	VAL12.H	VAL12.QG2	1.80	6.00
56	P56	TRP21.H	TRP21.HA	1.80	5.00

		1 ASG_POS1	2 ASG_POS2	3 LOWER	4 UPPER
57	P57	TRP21.H	ILE20.HA	1.80	2.50
58	P58	TRP21.H	TRP21.HB1	1.80	5.00
59	P59	TRP21.H	TRP21.HB2	1.80	3.30
60	P60	TRP21.H	ILE20.HB	1.80	5.00
61	P61	TRP21.H	ILE20.QG2	1.80	6.00
62	P62	MET7.H	MET7.HA	1.80	5.00
63	P63	MET7.H	LEU6.HA	1.80	2.50
64	P64	MET7.H	SER5.HB1	1.80	5.00
65	P65	MET7.H	MET7.HG1	1.80	5.00
66	P66	MET7.H	MET7.HG2	1.80	5.00
67	P67	MET7.H	MET7.HB1	1.80	5.00
68	P68	MET7.H	MET7.HB2	1.80	5.00
69	P69	MET7.H	LEU6.HG	1.80	5.00
70	P70	HID16.H	CYS15.HA	1.80	5.00
71	P71	HID16.H	HID16.HA	1.80	3.30
72	P72	HID16.H	TYR13.QB	1.80	5.60
73	P73	HID16.H	HID16.HB1	1.80	3.30
74	P74	HID16.H	HID16.HB2	1.80	2.50
75	P75	HID16.H	CYS15.HB2	1.80	5.00

		1 ASG_POS1	2 ASG_POS2	3 LOWER	4 UPPER
76	P76	LEU17.H	HID16.HA	1.80	5.00
77	P77	LEU17.H	LEU17.HA	1.80	3.30
78	P78	LEU17.H	HID16.HB1	1.80	5.00
79	P79	LEU17.H	HID16.HB2	1.80	5.00
80	P80	LEU17.H	LEU17.QB	1.80	3.90
81	P81	LEU17.H	LEU17.HG	1.80	3.30
82	P82	LEU17.H	LEU17.QQD	1.80	6.56
83	P83	TYR13.H	TYR13.HA	1.80	2.50
84	P84	TYR13.H	VAL12.HA	1.80	5.00
85	P85	TYR13.H	TYR13.QB	1.80	3.10
86	P86	TYR13.H	VAL12.HB	1.80	5.00
87	P87	TYR13.H	VAL12.QG1	1.80	6.00
88	P88	TYR13.H	VAL12.QG2	1.80	6.00
89	P89	ILE20.H	ILE20.HA	1.80	2.50
90	P90	ILE20.H	ILE19.HA	1.80	2.50
91	P91	ILE20.H	ILE20.HB	1.80	5.00
92	P92	ILE20.H	ILE20.HG11	1.80	5.00
93	P93	ILE20.H	ILE20.HG12	1.80	5.00
94	P94	ILE20.H	ILE20.QG2	1.80	6.00

		1 ASG_POS1	2 ASG_POS2	3 LOWER	4 UPPER
95	P95	ILE20.H	ILE19.QD1	1.80	6.00
96	P96	SER5.H	SER5.HA	1.80	5.00
97	P97	SER5.H	SER4.HA	1.80	5.00
98	P98	SER5.H	SER5.HB1	1.80	5.00
99	P99	SER5.H	SER5.HB2	1.80	5.00
100	P100	ILE19.H	ASP18.HA	1.80	5.00
101	P101	ILE19.H	ILE19.HA	1.80	5.00
102	P102	ILE19.H	ILE19.HB	1.80	5.00
103	P103	ILE19.H	ILE19.HG11	1.80	5.00
104	P104	ILE19.H	ILE19.HG12	1.80	5.00
105	P105	ILE19.H	ILE19.QD1	1.80	6.00
106	P106	CYS11.H	CYS11.HA	1.80	5.00
107	P107	CYS11.H	GLU10.HA	1.80	5.00
108	P108	CYS11.H	CYS11.QB	1.80	3.10
109	P109	CYS11.H	GLU10.QB	1.80	5.60
110	P110	TRP21.HE3	TRP21.HA	1.80	5.00
111	P113	ASP8.H	MET7.HA	1.80	5.00
112	P114	ASP8.H	ASP8.HB1	1.80	5.00
113	P115	ASP8.H	ASP8.HB2	1.80	5.00

		1 ASG_POS1	2 ASG_POS2	3 LOWER	4 UPPER
114	P116	PHE14.QR	PHE14.HA	1.80	3.30
115	P117	PHE14.QR	CYS15.HB1	1.80	5.00
116	P118	PHE14.QR	PHE14.HB1	1.80	3.30
117	P119	PHE14.QR	PHE14.HB2	1.80	3.30
118	P120	PHE14.QR	TYR13.QB	1.80	5.60
119	P121	HID16.HD2	HID16.HA	1.80	5.00
120	P122	HID16.HD2	TYR13.HA	1.80	5.00
121	P125	HID16.HD2	LEU17.QQD	1.80	6.56
122	P126	TRP21.HD1	TRP21.HA	1.80	5.00
123	P129	TYR13.QR	TYR13.HA	1.80	3.30
124	P130	TYR13.QR	TYR13.QB	1.80	3.10
125	P131	TYR13.QR	LEU17.QQD	1.80	6.56
126	P136	SER4.H	SER5.H	1.80	5.00
127	P137	SER2.H	CYS3.H	1.80	5.00
128	P138	LEU6.H	MET7.H	1.80	5.00
129	P139	CYS15.H	PHE14.H	1.80	5.00
130	P140	CYS15.H	HID16.H	1.80	5.00
131	P141	GLU10.H	LYS9.H	1.80	2.50
132	P142	GLU10.H	CYS11.H	1.80	5.00
		1 ASG_POS1	2 ASG_POS2	3 LOWER	4 UPPER
133	P143	ASP18.H	LEU17.H	1.80	3.30
134	P144	ASP18.H	ILE19.H	1.80	5.00
135	P145	PHE14.H	TYR13.H	1.80	2.50
136	P146	VAL12.H	TYR13.H	1.80	3.30
137	P147	VAL12.H	CYS11.H	1.80	5.00
138	P148	TRP21.H	ILE20.H	1.80	2.50
139	P149	MET7.H	SER5.H	1.80	3.30
140	P150	MET7.H	ASP8.H	1.80	5.00
141	P151	HID16.H	LEU17.H	1.80	2.50
142	P152	ILE20.H	ILE19.H	1.80	3.30
143	P154	PHE14.H	TYR13.QR	1.80	5.00
144	P155	CYS3.HA	CYS3.HB1	1.80	3.30
145	P156	CYS3.HA	CYS3.HB2	1.80	3.30
146	P157	SER2.HA	SER2.QB	1.80	3.10
147	P158	SER5.HA	SER5.HB1	1.80	3.30
148	P159	SER5.HA	SER5.HB2	1.80	3.30
149	P160	SER4.HA	SER4.HB1	1.80	3.30
150	P161	SER4.HA	SER4.HB2	1.80	3.30
151	P164	TRP21.HA	TRP21.HB1	1.80	2.50
		1 ASG_POS1	2 ASG_POS2	3 LOWER	4 UPPER
152	P166	CYS15.HA	CYS15.HB1	1.80	2.50
153	P167	CYS15.HA	CYS15.HB2	1.80	3.30
154	P168	ASP18.HA	ASP18.HB1	1.80	3.30
155	P169	ASP18.HA	ASP18.HB2	1.80	3.30
156	P170	HID16.HA	HID16.HB1	1.80	2.50
157	P171	HID16.HA	HID16.HB2	1.80	2.50
158	P172	MET7.HA	MET7.HG1	1.80	5.00
159	P173	MET7.HA	MET7.HG2	1.80	5.00
160	P174	MET7.HA	MET7.HB1	1.80	3.30
161	P175	MET7.HA	MET7.HB2	1.80	5.00
162	P176	CYS1.HA	CYS1.QB	1.80	3.10
163	P177	CYS11.HA	CYS11.QB	1.80	3.10
164	P178	CYS11.HA	PHE14.HB2	1.80	2.50
165	P179	PHE14.HA	PHE14.HB1	1.80	2.50
166	P180	PHE14.HA	PHE14.HB2	1.80	2.50
167	P181	TYR13.HA	TYR13.QB	1.80	3.10
168	P182	GLU10.HA	TYR13.QB	1.80	3.10
169	P183	GLU10.HA	GLU10.QG	1.80	5.60
170	P184	GLU10.HA	GLU10.QB	1.80	3.10

		1 ASG_POS1	2 ASG_POS2	3 LOWER	4 UPPER
171	P185	LEU17.HA	LEU17.QB	1.80	3.10
172	P186	LEU17.HA	LEU17.HG	1.80	2.50
173	P187	LEU17.HA	LEU17.QQD	1.80	6.56
174	P188	ILE20.HA	ILE20.HB	1.80	3.30
175	P189	ILE20.HA	ILE20.HG11	1.80	5.00
176	P190	ILE20.HA	ILE20.HG12	1.80	5.00
177	P191	ILE20.HA	ILE20.QG2	1.80	6.00
178	P192	ILE19.HA	ILE19.HB	1.80	3.30
179	P193	ILE19.HA	ILE19.HG11	1.80	5.00
180	P194	ILE19.HA	ILE19.HG12	1.80	3.30
181	P195	ILE19.HA	ILE19.QG2	1.80	6.00
182	P196	ILE19.HA	ILE19.QD1	1.80	6.00
183	P197	LEU6.HA	LEU6.QB	1.80	3.10
184	P198	LEU6.HA	LEU6.HG	1.80	2.50
185	P199	LEU6.HA	LEU6.QQD	1.80	6.56
186	P200	LYS9.HA	VAL12.HB	1.80	3.30
187	P201	LYS9.HA	LYS9.HB1	1.80	2.50
188	P202	LYS9.HA	LYS9.QD	1.80	5.60
189	P204	LYS9.HA	LYS9.QG	1.80	5.60
		1 ASG_POS1	2 ASG_POS2	3 LOWER	4 UPPER
190	P205	LYS9.HA	VAL12.QG1	1.80	6.00
191	P206	VAL12.HA	VAL12.HB	1.80	3.30
192	P207	VAL12.HA	LEU17.QB	1.80	5.60
193	P208	VAL12.HA	LEU17.HG	1.80	5.00
194	P209	VAL12.HA	VAL12.QG1	1.80	6.00
195	P210	VAL12.HA	VAL12.QG2	1.80	6.00
196	P211	VAL12.HA	CYS15.HB1	1.80	5.00
197	P212	VAL12.HA	CYS15.HB2	1.80	3.30
198	P214	ASP8.HB1	ASP8.HB2	1.80	2.50
199	P215	CYS3.HB1	CYS3.HB2	1.80	2.50
200	P216	ASP18.HB1	ASP18.HB2	1.80	2.50
201	P228	MET7.HG2	MET7.HB1	1.80	2.50
202	P229	MET7.HG2	MET7.HB2	1.80	3.30
203	P230	MET7.HB1	MET7.HB2	1.80	2.50
204	P232	VAL12.HB	VAL12.QG1	1.80	6.00
205	P233	VAL12.HB	VAL12.QG2	1.80	6.00
206	P235	ILE20.HB	ILE20.HG11	1.80	3.30
207	P236	ILE20.HB	ILE20.HG12	1.80	5.00
208	P241	ILE19.HB	ILE19.HG11	1.80	2.50
		1 ASG_POS1	2 ASG_POS2	3 LOWER	4 UPPER
209	P242	ILE19.HB	ILE19.HG12	1.80	3.30
210	P249	SER5.H	CYS3.HB1	1.80	5.00
211	P250	LEU6.H	SER5.HB1	1.80	5.00

Atom definitions can be found in Appendix IV

Table I.b : Modified upper distance constraints of ET-1 obtained form DIANA calculation. These constraints were used for later calculations (section 4.3).

1	CYS				
		HA	2	SER	H	3.30
		CB	15	CYS	SG	3.10
		QB	2	SER	H	5.60
		SG	15	CYS	CB	3.10
		SG	15	CYS	SG	2.10
2	SER				
		H	3	CYS	H	5.00
3	CYS				
		CB	11	CYS	SG	3.10
		HB2	4	SER	H	5.00
		HB1	4	SER	H	5.00
		HB1	5	SER	H	5.00
		SG	11	CYS	CB	3.10
		SG	11	CYS	SG	2.10
4	SER				
		H	5	SER	H	5.00
5	SER				
		H	7	MET	H	3.30
		HB1	6	LEU	H	5.00
		HB1	7	MET	H	5.00
6	LEU				
		H	6	LEU	HG	3.30
		H	6	LEU	QQD	6.56
		H	7	MET	H	5.00
		HA	6	LEU	HG	2.50
		HA	7	MET	H	2.50
		HG	7	MET	H	5.00
7	MET				
		H	7	MET	HG2	5.00
		H	7	MET	HG1	5.00
		H	8	ASP	H	5.00
		HB1	7	MET	HG2	2.50
9	LYS				
		H	9	LYS	QG	5.60
		H	10	GLU	H	2.50
		HA	9	LYS	HB1	2.50
		HA	9	LYS	QD	5.60
		HA	12	VAL	H	5.00
		HA	12	VAL	HB	3.30
		HA	12	VAL	QG1	6.00
		HB2	9	LYS	QD	3.10
		HB2	9	LYS	QE	5.60
		HB1	10	GLU	H	5.00
10	GLU				
		H	10	GLU	QB	3.90
		H	10	GLU	QG	5.60
		H	11	CYS	H	5.00
		HA	13	TYR	QB	3.10
11	CYS				
		H	11	CYS	QB	3.10
		H	12	VAL	H	5.00
		HA	14	PHE	HB2	2.50
		QB	12	VAL	H	3.90
12	VAL				
		H	12	VAL	HB	3.30
		H	13	TYR	H	3.30
		HA	15	CYS	H	5.00
		HA	15	CYS	HB2	3.30
		HA	15	CYS	HB1	5.00
		HA	17	LEU	QB	5.60
		HA	17	LEU	HG	5.00
		HB	13	TYR	H	5.00
		QG2	13	TYR	H	6.00
		QG1	13	TYR	H	6.00

13	TYR				
		H	13	TYR	HA	2.50
		H	13	TYR	QB	3.10
		H	14	PHE	H	2.50
		HA	13	TYR	QR	3.30
		HA	14	PHE	H	2.50
		HA	16	HID	HD2	5.00
		QB	14	PHE	H	5.60
		QB	14	PHE	QR	5.60
		QB	16	HID	H	5.60
		QR	14	PHE	H	5.00
		QR	17	LEU	QQD	6.56
14	PHE					
		H	14	PHE	HB2	3.30
		H	14	PHE	HB1	2.50
		H	14	PHE	QR	5.00
		H	15	CYS	H	5.00
		HA	14	PHE	HB2	2.50
		HA	14	PHE	HB1	2.50
		HA	14	PHE	QR	3.30
		HB2	14	PHE	QR	3.30
		HB2	15	CYS	H	5.00
		HB1	14	PHE	QR	3.30
		HB1	15	CYS	H	5.00
		QR	15	CYS	HB1	5.00
15	CYS					
		H	16	HID	H	5.00
		HA	15	CYS	HB1	2.50
		HB2	16	HID	H	5.00
16	HID					
		H	16	HID	HB2	2.50
		H	16	HID	HB1	3.30
		H	17	LEU	H	2.50
		HA	16	HID	HB2	2.50
		HA	16	HID	HB1	2.50
		HA	16	HID	HD2	5.00
		HB2	17	LEU	H	5.00
		HB1	17	LEU	H	5.00
		HD2	17	LEU	QQD	6.56
17	LEU					
		H	17	LEU	QB	3.90
		H	17	LEU	HG	3.30
		H	17	LEU	QQD	6.56
		H	18	ASP	H	3.30
		HA	17	LEU	HG	2.50
		HA	18	ASP	H	2.50
		QB	18	ASP	H	5.60
		HG	18	ASP	H	5.00
		QQD	18	ASP	H	6.56
18	ASP					
		H	19	ILE	H	5.00
19	ILE					
		H	19	ILE	HG12	5.00
		H	19	ILE	HG11	5.00
		H	19	ILE	QD1	6.00
		H	20	ILE	H	3.30
		HA	19	ILE	HG12	3.30
		HA	20	ILE	H	2.50
		HB	19	ILE	HG11	2.50
		QD1	20	ILE	H	6.00
20	ILE					
		H	20	ILE	HA	2.50
		H	20	ILE	HG12	5.00
		H	20	ILE	HG11	5.00
		H	21	TRP	H	2.50
		HA	21	TRP	H	2.50
		HB	21	TRP	H	5.00
		QG2	21	TRP	H	6.00
21	TRP					
		H	21	TRP	HB2	3.30
		HA	21	TRP	HB1	2.50
		HA	21	TRP	HD1	5.00
		HA	21	TRP	HE3	5.00

see Appendix IV

Table I.c : Modified lower distance constraints of ET-1 obtained form DIANA calculation. These constraints were used for later calculations (section 4.3).

1	CYS				
		CB	15	CYS	SG	3.00
		QB	2	SER	H	1.80
		SG	15	CYS	CB	3.00
		SG	15	CYS	SG	2.00
2	SER					
		QB	3	CYS	H	1.80
3	CYS					
		CB	11	CYS	SG	3.00
		SG	11	CYS	CB	3.00
		SG	11	CYS	SG	2.00
6	LEU					
		H	6	LEU	QQD	1.80
9	LYS					
		H	9	LYS	QG	1.80
		HA	9	LYS	QD	1.80
		HA	12	VAL	QG1	1.80
		HB2	9	LYS	QE	1.80
10	GLU					
		H	10	GLU	QG	1.80
		HA	13	TYR	QB	1.80
		QB	11	CYS	H	1.80
11	CYS					
		QB	12	VAL	H	1.80
12	VAL					
		HA	17	LEU	QB	1.80
		QG2	13	TYR	H	1.80
		QG1	13	TYR	H	1.80
13	TYR					
		QB	14	PHE	H	1.80
		QB	14	PHE	QR	1.80
		QB	16	HID	H	1.80
		QR	14	PHE	H	1.80
		QR	17	LEU	QQD	1.80
14	PHE					
		H	14	PHE	QR	1.80
		QR	15	CYS	HB1	1.80
16	HID					
		HD2	17	LEU	QQD	1.80
17	LEU					
		H	17	LEU	QQD	1.80
		QQD	18	ASP	H	1.80
19	ILE					
		H	19	ILE	QD1	1.80
		QD1	20	ILE	H	1.80
20	ILE					
		QG2	21	TRP	H	1.80

Atom definitions can be found in Appendix IV

Table I.d : Torsional angle constraints of ET-1.

13 Tyr	Phi	-163.0	-103.0
15 Cys	Phi	-160.0	-100.0
16 His	Phi	-171.0	-111.1
19 Ile	Phi	-167.9	-132.1
20 Ile	Phi	-173.4	-126.7

Table I.e : Hydrogen bond constraints of ET-1.

				lower	upper (Å)
4 Ser	O	7 Met	N	2.7	3.0
4 Ser	O	7 Met	H	1.8	2.0
6 Leu	O	10 Glu	N	2.7	3.0
6 Leu	O	10 Glu	H	1.8	2.0
9 Lys	O	13 Tyr	N	2.7	3.0
9 Lys	O	13 Tyr	H	1.8	2.0
10 Glu	O	14 Phe	N	2.7	3.0
10 Glu	O	14 Phe	H	1.8	2.0

Table I.f : Constraints across the disulphide bridges of ET-1.

				lower	upper (Å)
1 Cys	SG	15 Cys	SG	2.0	2.1
1 Cys	SG	15 Cys	CB	3.0	3.1
1 Cys	CB	15 Cys	SG	3.0	3.1
3 Cys	SG	11 Cys	SG	2.0	2.1
3 Cys	SG	11 Cys	CB	3.0	3.1
3 Cys	CB	11 Cys	SG	3.0	3.1

Table I.g : $^{3}J_{NH\alpha}$ Coupling constants of ET-1.

	Hz
13 Tyr	5.8
15 Cys	5.5
16 His	6.9
19 Ile	9.0
20 Ile	8.6

Appendix II

Additional NMR and structure calculation details of LJP 1

Table II.a : NOE constraints of LJP 1 used for DIANA calculation : Peak numbers, peak assignments (F2 F1), lower constraints and upper constraints.

		1 ASG_POS1	2 ASG_POS2	3 LOWER	4 UPPER
1	P1	aib3.h	ser2.ha	1.80	2.50
2	P2	aib3.h	ser2.hb1	1.80	3.30
3	P3	aib3.h	ser2.hb2	1.80	5.00
4	P4	ser2.h	ser2.ha	1.80	2.50
5	P5	ser2.h	cys1.ha	1.80	2.50
6	P6	ser2.h	ser2.hb1	1.80	5.00
7	P7	ser2.h	ser2.hb2	1.80	5.00
8	P8	asp18.h	asp18.ha	1.80	3.30
9	P9	asp18.h	leu17.ha	1.80	3.30
10	P10	asp8.h	asp8.ha	1.80	3.30
11	P11	phe14.h	phe14.ha	1.80	3.30
12	P12	asp8.h	ser5.ha	1.80	5.00
13	P13	phe14.h	tyr13.ha	1.80	5.00
14	P14	asp8.h	leu7.ha	1.80	5.00
15	P15	phe14.h	glu10.ha	1.80	5.00
16	P16	aib11.h	asp8.ha	1.80	5.00
17	P17	hid16.h	hid16.ha	1.80	5.00
18	P18	cys15.h	cys15.ha	1.80	5.00
		1 ASG_POS1	2 ASG_POS2	3 LOWER	4 UPPER
19	P19	hid16.h	cys15.ha	1.80	2.50
20	P20	cys15.h	phe14.ha	1.80	5.00
21	P21	trp21.h	trp21.ha	1.80	5.00
22	P22	ser5.h	ser2.ha	1.80	5.00
23	P23	ser4.h	ser4.ha	1.80	5.00
24	P24	aib11.h	glu10.ha	1.80	5.00
25	P25	ser4.h	ser2.hb1	1.80	5.00
26	P26	glu10.h	leu6.ha	1.80	5.00
27	P27	glu10.h	leu7.ha	1.80	5.00
28	P28	glu10.h	glu10.ha	1.80	3.30
29	P29	ser4.h	ser4.qb	1.80	5.60
30	P30	cys15.h	val12.ha	1.80	5.00
31	P31	ser5.h	ser5.ha	1.80	5.00
32	P32	leu7.h	leu6.ha	1.80	5.00
33	P33	trp21.h	ile20.ha	1.80	3.30
34	P34	ser5.h	ser4.ha	1.80	3.30
35	P35	leu7.h	leu7.ha	1.80	3.30
36	P36	ser5.h	ser5.qb	1.80	3.10
37	P37	ser5.h	ser4.qb	1.80	5.60
		1 ASG_POS1	2 ASG_POS2	3 LOWER	4 UPPER
38	P38	leu17.h	hid16.ha	1.80	3.30
39	P39	lys9.h	asp8.ha	1.80	3.30
40	P40	ile19.h	asp18.ha	1.80	5.00
41	P41	val12.h	asp8.ha	1.80	5.00
42	P42	leu17.h	leu17.ha	1.80	3.30
43	P43	tyr13.h	tyr13.ha	1.80	3.30
44	P44	ile20.h	ile20.ha	1.80	5.00
45	P45	ile20.h	ile19.ha	1.80	3.30
46	P46	tyr13.h	glu10.ha	1.80	5.00
47	P47	lys9.h	leu6.ha	1.80	5.00
48	P48	lys9.h	lys9.ha	1.80	3.30
49	P49	tyr13.h	val12.ha	1.80	5.00
50	P50	ile19.h	ile19.ha	1.80	3.30
51	P51	leu6.h	ser5.ha	1.80	5.00
52	P52	leu6.h	leu6.ha	1.80	5.00
53	P54	leu6.h	ser5.qb	1.80	5.60
54	P55	val12.h	lys9.ha	1.80	5.00
55	P56	val12.h	val12.ha	1.80	5.00
56	P59	aib3.h	ser4.h	1.80	5.00

		1 ASG_POS1	2 ASG_POS2	3 LOWER	4 UPPER
57	P60	aib3.h	ser5.h	1.80	5.00
58	P61	asp18.h	leu17.h	1.80	3.30
59	P62	phe14.h	cys15.h	1.80	3.30
60	P63	asp8.h	leu7.h	1.80	5.00
61	P64	phe14.h	tyr13.h	1.80	2.50
62	P65	asp8.h	lys9.h	1.80	3.30
63	P66	ser4.h	ser5.h	1.80	3.30
64	P67	aib11.h	lys9.h	1.80	5.00
65	P68	hid16.h	leu17.h	1.80	2.50
66	P69	glu10.h	lys9.h	1.80	2.50
67	P70	trp21.h	ile20.h	1.80	5.00
68	P71	asp18.h	ile19.h	1.80	5.00
69	P72	asp8.h	leu6.h	1.80	5.00
70	P73	phe14.h	val12.h	1.80	5.00
71	P74	aib11.h	val12.h	1.80	2.50
72	P75	glu10.h	val12.h	1.80	5.00
73	P76	ser5.h	leu6.h	1.80	5.00
74	P77	leu7.h	leu6.h	1.80	3.30
75	P78	ile20.h	ile19.h	1.80	3.30
		1 ASG_POS1	2 ASG_POS2	3 LOWER	4 UPPER
76	P79	tyr13.h	val12.h	1.80	3.30
77	P80	ser2.h	cys1.hb1	1.80	5.00
78	P81	ser2.h	cys1.hb2	1.80	5.00
79	P82	hid16.he1	cys15.qb	1.80	3.90
80	P83	aib3.h	aib3.qqb	1.80	6.56
81	P86	hid16.he1	val12.qg2	1.80	6.00
82	P87	asp18.h	asp18.hb1	1.80	2.50
83	P88	asp18.h	asp18.hb2	1.80	3.30
84	P89	phe14.h	phe14.hb1	1.80	3.30
85	P90	phe14.h	phe14.hb2	1.80	3.30
86	P91	asp8.h	asp8.hb1	1.80	5.00
87	P92	phe14.h	tyr13.qb	1.80	3.90
88	P93	asp8.h	asp8.hb2	1.80	2.50
89	P94	hid16.h	hid16.hb1	1.80	3.30
90	P95	hid16.h	hid16.hb2	1.80	2.50
91	P96	cys15.h	cys15.qb	1.80	3.10
92	P97	cys15.h	phe14.hb2	1.80	5.00
93	P98	trp21.h	trp21.hb1	1.80	5.00
94	P99	trp21.h	trp21.hb2	1.80	3.30
		1 ASG_POS1	2 ASG_POS2	3 LOWER	4 UPPER
95	P100	leu17.h	hid16.hb1	1.80	5.00
96	P101	leu17.h	hid16.hb2	1.80	5.00
97	P102	tyr13.h	tyr13.qb	1.80	3.10
98	P103	lys9.h	asp8.hb1	1.80	3.30
99	P104	lys9.h	asp8.hb2	1.80	5.00
100	P105	aib11.h	glu10.hb2	1.80	3.30
101	P106	glu10.h	glu10.hg1	1.80	2.50
102	P107	glu10.h	glu10.hg2	1.80	5.00
103	P108	glu10.h	glu10.hb1	1.80	3.30
104	P109	glu10.h	glu10.hb2	1.80	5.00
105	P110	glu10.h	lys9.hb1	1.80	2.50
106	P111	asp18.h	leu17.hg	1.80	5.00
107	P112	asp8.h	leu7.hb1	1.80	3.30
108	P113	asp8.h	leu7.hg	1.80	5.00
109	P114	asp8.h	aib11.qqb	1.80	6.56
110	P115	aib11.h	aib11.qqb	1.80	6.56
111	P116	ser4.h	aib3.qqb	1.80	6.56
112	P117	glu10.h	aib11.qqb	1.80	6.56
113	P118	trp21.h	ile20.hb	1.80	5.00

		1 ASG_POS1	2 ASG_POS2	3 LOWER	4 UPPER
114	P120	leu7.h	leu7.hg	1.80	2.50
115	P121	ser5.h	aib3.qqb	1.80	6.56
116	P122	leu7.h	aib3.qqb	1.80	6.56
117	P123	tyr13.h	val12.hb	1.80	2.50
118	P124	lys9.h	lys9.hb1	1.80	2.50
119	P125	ile20.h	ile20.hb	1.80	2.50
120	P126	leu7.h	leu7.hb1	1.80	3.30
121	P127	leu7.h	leu7.hb2	1.80	3.30
122	P128	leu17.h	leu17.qb	1.80	3.90
123	P129	leu17.h	leu17.hg	1.80	3.30
124	P130	ile20.h	ile20.hg11	1.80	5.00
125	P131	lys9.h	lys9.qd	1.80	5.60
126	P132	lys9.h	lys9.hb2	1.80	5.00
127	P133	lys9.h	lys9.qg	1.80	5.60
128	P134	ile19.h	ile19.hb	1.80	3.30
129	P135	ile19.h	ile19.hg11	1.80	3.30
130	P136	leu6.h	leu6.hb1	1.80	3.30
131	P137	leu6.h	leu6.hb2	1.80	5.00
132	P138	leu6.h	leu6.hg	1.80	3.30
		1 ASG_POS1	2 ASG_POS2	3 LOWER	4 UPPER
133	P139	leu6.h	aib3.qqb	1.80	6.56
134	P140	val12.h	val12.hb	1.80	2.50
135	P141	val12.h	aib11.qqb	1.80	6.56
136	P143	asp18.h	leu17.qd1	1.80	6.00
137	P144	aib11.h	val12.qg1	1.80	6.00
138	P146	glu10.h	val12.qg2	1.80	6.00
139	P147	trp21.h	ile20.qg2	1.80	6.00
140	P148	leu7.h	leu6.qd1	1.80	6.00
141	P149	ile20.h	ile20.hg12	1.80	3.30
142	P150	tyr13.h	val12.qg1	1.80	6.00
143	P151	tyr13.h	val12.qg2	1.80	6.00
144	P152	leu17.h	leu17.qd1	1.80	6.00
145	P153	ile20.h	ile20.qg2	1.80	6.00
146	P154	ile20.h	ile19.qg2	1.80	6.00
147	P155	ile19.h	ile19.hg12	1.80	5.00
148	P156	leu6.h	leu6.qd1	1.80	6.00
149	P157	val12.h	val12.qg1	1.80	6.00
150	P158	val12.h	val12.qg2	1.80	6.00
151	P159	ile19.h	ile19.qg2	1.80	6.00
		1 ASG_POS1	2 ASG_POS2	3 LOWER	4 UPPER
152	P160	hid16.hd2	hid16.ha	1.80	5.00
153	P161	phe14.qr	phe14.ha	1.80	5.00
154	P163	phe14.qr	glu10.ha	1.80	5.00
155	P164	phe14.qr	phe14.hb1	1.80	5.00
156	P170	hid16.hd2	cys15.qb	1.80	5.60
157	P171	phe14.qr	phe14.hb2	1.80	3.30
158	P173	phe14.qr	tyr13.qb	1.80	5.60
159	P174	phe14.qr	glu10.hb2	1.80	5.00
160	P175	phe14.qr	leu17.hg	1.80	5.00
161	P176	phe14.qr	glu10.hg2	1.80	5.00
162	P177	phe14.qr	leu17.qd1	1.80	6.00
163	P178	hid16.hd2	leu17.qd1	1.80	6.00
164	P179	trp21.hd1	ile20.qd1	1.80	6.00
165	P180	tyr13.qr	phe14.ha	1.80	5.00
166	P181	tyr13.qr	tyr13.ha	1.80	5.00
167	P182	tyr13.qr	glu10.ha	1.80	5.00
168	P184	tyr13.qr	tyr13.qb	1.80	3.90
169	P186	tyr13.qr	val12.qg1	1.80	6.00
170	P189	tyr13.qr	val12.qg2	1.80	6.00

		1 ASG_POS1	2 ASG_POS2	3 LOWER	4 UPPER
171	P192	phe14.h	phe14.qr	1.80	3.30
172	P193	phe14.h	tyr13.qr	1.80	5.00
173	P194	hid16.h	hid16.hd2	1.80	5.00
174	P195	tyr13.h	tyr13.qr	1.80	5.00
175	P200	hid16.hd2	tyr13.qr	1.80	3.30
176	P203	ser2.ha	ser2.hb1	1.80	5.00
177	P204	ser2.ha	ser2.hb2	1.80	5.00
178	P205	ser5.ha	ser5.qb	1.80	5.60
179	P206	ser4.ha	ser4.qb	1.80	5.60
180	P207	ser2.hb1	ser2.hb2	1.80	5.00
181	P208	hid16.ha	hid16.hb1	1.80	5.00
182	P209	cys15.ha	cys15.qb	1.80	3.90
183	P210	asp8.ha	asp8.hb1	1.80	5.00
184	P211	asp8.ha	asp8.hb2	1.80	3.30
185	P212	phe14.ha	phe14.hb1	1.80	5.00
186	P213	phe14.ha	phe14.hb2	1.80	5.00
187	P214	cys1.ha	cys1.hb1	1.80	3.30
188	P215	cys1.ha	cys1.hb2	1.80	3.30
189	P216	ser5.ha	asp8.hb1	1.80	5.00
		1 ASG_POS1	2 ASG_POS2	3 LOWER	4 UPPER
190	P217	ser5.ha	asp8.hb2	1.80	5.00
191	P218	tyr13.ha	tyr13.qb	1.80	3.90
192	P219	glu10.ha	tyr13.qb	1.80	5.60
193	P220	val12.ha	cys15.qb	1.80	5.60
194	P221	glu10.ha	glu10.hg1	1.80	5.00
195	P222	glu10.ha	glu10.hg2	1.80	5.00
196	P223	glu10.ha	glu10.hb1	1.80	3.30
197	P224	lys9.ha	val12.hb	1.80	5.00
198	P225	lys9.ha	lys9.hb1	1.80	5.00
199	P226	leu7.ha	glu10.hb1	1.80	5.00
200	P227	leu6.ha	lys9.hb1	1.80	5.00
201	P228	val12.ha	val12.hb	1.80	5.00
202	P229	asp8.ha	aib11.qqb	1.80	6.56
203	P231	phe14.ha	leu17.hg	1.80	5.00
204	P232	leu17.ha	leu17.qb	1.80	5.60
205	P233	leu17.ha	leu17.hg	1.80	5.00
206	P234	leu6.ha	leu6.hb1	1.80	5.00
207	P235	leu6.ha	leu6.hg	1.80	5.00
208	P236	ile20.ha	ile20.hb	1.80	5.00
		1 ASG_POS1	2 ASG_POS2	3 LOWER	4 UPPER
209	P237	leu7.ha	leu7.hb1	1.80	5.00
210	P238	leu7.ha	leu7.hg	1.80	5.00
211	P239	ile20.ha	ile20.hb	1.80	5.00
212	P240	lys9.ha	lys9.qg	1.80	5.60
213	P241	ile19.ha	ile19.hg11	1.80	5.00
214	P242	ile20.ha	ile20.hg11	1.80	5.00
215	P243	val12.ha	lys9.hb2	1.80	5.00
216	P244	leu17.ha	leu17.qd1	1.80	6.00
217	P245	leu6.ha	leu6.qd1	1.80	6.00
218	P246	leu6.ha	leu6.qd2	1.80	6.00
219	P247	ile20.ha	ile20.hg12	1.80	5.00
220	P249	ile19.ha	ile19.hg12	1.80	5.00
221	P251	leu7.ha	leu7.qd1	1.80	6.00
222	P252	ile19.ha	ile19.qd1	1.80	6.00
223	P253	lys9.ha	val12.qg1	1.80	6.00
224	P256	hid16.hb1	hid16.hb2	1.80	5.00
225	P257	trp21.hb1	trp21.hb2	1.80	5.00
226	P258	phe14.hb1	phe14.hb2	1.80	3.30
227	P259	cys1.hb1	cys1.hb2	1.80	3.30

		1 ASG_POS1	2 ASG_POS2	3 LOWER	4 UPPER
228	P260	asp8.hb1	asp8.hb2	1.80	3.30
229	P261	asp18.hb1	asp18.hb2	1.80	5.00
230	P262	phe14.hb1	leu17.hg	1.80	5.00
231	P263	cys15.qb	aib11.qqb	1.80	7.16
232	P264	cys15.qb	leu17.qd1	1.80	6.60
233	P265	phe14.hb2	leu17.hg	1.80	5.00
234	P268	glu10.hg1	glu10.hb1	1.80	2.50
235	P269	glu10.hg2	glu10.hb1	1.80	5.00
236	P270	val12.hb	val12.qg1	1.80	6.00
237	P271	val12.hb	val12.qg2	1.80	6.00
238	P272	lys9.hb1	lys9.hb2	1.80	5.00
239	P273	lys9.hb1	lys9.qd	1.80	5.60
240	P275	ile20.hb	ile20.hg11	1.80	5.00
241	P276	ile19.hb	ile19.hg11	1.80	5.00
242	P277	lys9.hb2	lys9.qg	1.80	5.60
243	P281	ile19.hb	ile19.qd1	1.80	6.00
244	P282	leu17.qb	leu17.qd1	1.80	6.60
245	P284	leu7.hb1	leu7.qd1	1.80	6.00
246	P285	leu6.hb1	leu6.qd1	1.80	6.00
		1 ASG_POS1	2 ASG_POS2	3 LOWER	4 UPPER
247	P286	leu7.hb1	leu7.qd2	1.80	6.00
248	P287	leu6.hb1	leu6.qd2	1.80	6.00
249	P292	ile19.hg11	ile19.qg2	1.80	6.00

Atom definitions can be found in Appendix IV

Table II.b : Modified upper distance constraints of LJP1 obtained form DIANA calculation. These constraints were used for later calculations (section 4.3).

1	CYS				
		HA	2	SER	H	2.50
		HB2	2	SER	H	5.00
		HB1	2	SER	H	5.00
2	SER				
		H	2	SER	HA	2.50
		HA	3	AIB	H	2.50
		HA	5	SER	H	5.00
		HB2	3	AIB	H	5.00
		HB1	3	AIB	H	3.30
		HB1	4	SER	H	5.00
3	AIB				
		H	4	SER	H	5.00
		H	5	SER	H	5.00
		QQB	5	SER	H	6.56
		QQB	6	LEU	H	6.56
		QQB	7	LEU	H	6.56
4	SER				
		H	5	SER	H	3.30
		HA	5	SER	H	3.30
5	SER				
		H	5	SER	QB	3.10
		H	6	LEU	H	5.00
		HA	8	ASP	H	5.00
		HA	8	ASP	HB2	5.00
		HA	8	ASP	HB1	5.00
6	LEU				
		H	6	LEU	HB1	3.30
		H	6	LEU	HG	3.30
		H	6	LEU	QD1	6.00
		H	7	LEU	H	3.30
		H	8	ASP	H	5.00
		HA	6	LEU	QD1	6.00
		HA	6	LEU	QD2	6.00
		HA	9	LYS	H	5.00
		HA	9	LYS	HB1	5.00
		HA	10	GLU	H	5.00
		QD1	7	LEU	H	6.00
		O	10	GLU	N	3.00
		O	10	GLU	H	2.00
7	LEU				
		H	7	LEU	HB2	3.30
		H	7	LEU	HB1	3.30
		H	7	LEU	HG	2.50
		H	8	ASP	H	5.00
		HA	7	LEU	QD1	6.00
		HA	10	GLU	H	5.00
		HA	10	GLU	HB1	5.00
		HB1	8	ASP	H	3.30
		HG	8	ASP	H	5.00
		O	11	AIB	N	3.00
		O	11	AIB	H	2.00
8	ASP				
		H	8	ASP	HB2	2.50
		H	9	LYS	H	3.30
		H	11	AIB	QQB	6.56
		HA	9	LYS	H	3.30
		HA	11	AIB	H	5.00
		HA	11	AIB	QQB	6.56
		HA	12	VAL	H	5.00
		HB2	9	LYS	H	5.00
		HB1	9	LYS	H	3.30

9 LYS				
	H	9	LYS	HB1	2.50
	H	9	LYS	QG	5.60
	H	9	LYS	QD	5.60
	H	10	GLU	H	2.50
	H	11	AIB	H	5.00
	HA	12	VAL	H	5.00
	HA	12	VAL	HB	5.00
	HA	12	VAL	QG1	6.00
	HB2	12	VAL	HA	5.00
	HB1	10	GLU	H	2.50
	O	13	TYR	N	3.00
	O	13	TYR	H	2.00
10 GLU					
	H	10	GLU	HB1	3.30
	H	10	GLU	HG2	5.00
	H	10	GLU	HG1	2.50
	H	11	AIB	QQB	6.56
	H	12	VAL	H	5.00
	H	12	VAL	QG2	6.00
	HA	13	TYR	H	5.00
	HA	13	TYR	QB	5.60
	HA	13	TYR	QR	5.00
	HA	14	PHE	H	5.00
	HA	14	PHE	QR	5.00
	HB2	11	AIB	H	3.30
	HB2	14	PHE	QR	5.00
	HB1	10	GLU	HG1	2.50
	HG2	14	PHE	QR	5.00
	O	14	PHE	N	3.00
	O	14	PHE	H	2.00
11 AIB					
	H	12	VAL	H	2.50
	H	12	VAL	QG1	6.00
	QQB	15	CYS	QB	7.16
12 VAL					
	H	12	VAL	HB	2.50
	H	13	TYR	H	3.30
	H	14	PHE	H	5.00
	HA	15	CYS	H	5.00
	HA	15	CYS	QB	5.60
	HB	13	TYR	H	2.50
	QG2	13	TYR	H	6.00
	QG2	13	TYR	QR	6.00
	QG2	16	HID	HE1	6.00
	QG1	13	TYR	H	6.00
	QG1	13	TYR	QR	6.00
13 TYR					
	H	13	TYR	QB	3.10
	H	13	TYR	QR	5.00
	H	14	PHE	H	2.50
	QB	14	PHE	H	3.90
	QB	14	PHE	QR	5.60
	QR	14	PHE	H	5.00
	QR	14	PHE	HA	5.00
	QR	16	HID	HD2	5.00
14 PHE					
	H	14	PHE	HB2	3.30
	H	14	PHE	HB1	3.30
	H	14	PHE	QR	3.30
	H	15	CYS	H	3.30
	HA	17	LEU	HG	5.00
	HB2	14	PHE	QR	3.30
	HB2	15	CYS	H	5.00
	HB2	17	LEU	HG	5.00
	HB1	17	LEU	HG	5.00
	QR	17	LEU	HG	5.00
	QR	17	LEU	QD1	6.00
15 CYS					
	H	15	CYS	QB	3.10
	HA	16	HID	H	2.50
	QB	16	HID	HE1	3.90
	QB	16	HID	HD2	5.60
	QB	17	LEU	QD1	6.60
16 HID					
	H	16	HID	HB2	2.50
	H	16	HID	HB1	3.30
	H	16	HID	HD2	5.00
	H	17	LEU	H	2.50
	HA	16	HID	HD2	5.00
	HA	17	LEU	H	3.30
	HB2	17	LEU	H	5.00
	HB1	17	LEU	H	5.00
	HD2	17	LEU	QD1	6.00
17 LEU					
	H	17	LEU	QB	3.90
	H	17	LEU	HG	3.30
	H	17	LEU	QD1	6.00
	H	18	ASP	H	3.30
	HA	17	LEU	QD1	6.00
	HA	18	ASP	H	3.30
	HG	18	ASP	H	5.00
	QD1	18	ASP	H	6.00
18 ASP					
	H	18	ASP	HB2	3.30
	H	18	ASP	HB1	2.50
	H	19	ILE	H	5.00
19 ILE					
	H	19	ILE	HB	3.30
	H	19	ILE	HG12	5.00
	H	19	ILE	HG11	3.30
	H	20	ILE	H	3.30
	HA	20	ILE	H	3.30
	QG2	20	ILE	H	6.00
20 ILE					
	H	20	ILE	HB	2.50
	H	20	ILE	HG12	3.30
	H	20	ILE	HG11	5.00
	H	21	TRP	H	5.00
	HA	21	TRP	H	3.30
	HB	21	TRP	H	5.00
	QG2	21	TRP	H	6.00
	QD1	21	TRP	HD1	6.00
21 TRP					
	H	21	TRP	HB2	3.30

see Appendix IV

Table II.c : Modified lower distance constraints of LJP1 obtained form DIANA calculation. These constraints were used for later calculations (section 4.3).

3	AIB				
		QQB	5	SER	H	1.80
		QQB	6	LEU	H	1.80
		QQB	7	LEU	H	1.80
4	SER				
		QB	5	SER	H	1.80
5	SER				
		QB	6	LEU	H	1.80
6	LEU				
		H	6	LEU	QD1	1.80
		QD1	7	LEU	H	1.80
		O	10	GLU	N	2.70
		O	10	GLU	H	1.80
7	LEU				
		O	11	AIB	N	2.70
		O	11	AIB	H	1.80
8	ASP				
		H	11	AIB	QQB	1.80
		HA	11	AIB	QQB	1.80
9	LYS				
		H	9	LYS	QG	1.80
		H	9	LYS	QD	1.80
		HA	12	VAL	QG1	1.80
		O	13	TYR	N	2.70
		O	13	TYR	H	1.80
10	GLU				
		H	11	AIB	QQB	1.80
		H	12	VAL	QG2	1.80
		HA	13	TYR	QB	1.80
		HA	13	TYR	QR	1.80
		HA	14	PHE	QR	1.80
		HB2	14	PHE	QR	1.80
		HG2	14	PHE	QR	1.80
		O	14	PHE	N	2.70
		O	14	PHE	H	1.80
11	AIB				
		H	12	VAL	QG1	1.80
		QQB	15	CYS	QB	1.80
12	VAL				
		HA	15	CYS	QB	1.80
		QG2	13	TYR	H	1.80
		QG2	13	TYR	QR	1.80
		QG2	16	HID	HE1	1.80
		QG1	13	TYR	H	1.80
		QG1	13	TYR	QR	1.80
13	TYR				
		H	13	TYR	QR	1.80
		QB	14	PHE	H	1.80
		QB	14	PHE	QR	1.80
		QR	14	PHE	H	1.80
		QR	14	PHE	HA	1.80
		QR	16	HID	HD2	1.80
14	PHE				
		H	14	PHE	QR	1.80
		QR	17	LEU	HG	1.80
		QR	17	LEU	QD1	1.80
15	CYS				
		QB	16	HID	HE1	1.80
		QB	16	HID	HD2	1.80
		QB	17	LEU	QD1	1.80
16	HID				
		HD2	17	LEU	QD1	1.80
17	LEU				
		H	17	LEU	QD1	1.80
		QD1	18	ASP	H	1.80
19	ILE				
		QG2	20	ILE	H	1.80
20	ILE				
		QG2	21	TRP	H	1.80
		QD1	21	TRP	HD1	1.80

see Appendix IV

Table II.d : Torsional angle constraints of LJP1.

2 Ser	Phi	-105.0	-165.0
5 Ser	Phi	-95.0	-155.0
6 Leu	Phi	-104.0	-164.0
7 Leu	Phi	-89.0	-149.0
8 Asp	Phi	-92.0	-152.0
9 Lys	Phi	-93.0	-153.0
10 Glu	Phi	-88.0	-148.0
12 Val	Phi	-87.0	-147.0
13 Tyr	Phi	-88.0	-148.0
14 Phe	Phi	-92.0	-152.0
15 Cys	Phi	-87.0	-147.0
16 His	Phi	-88.0	-148.0
17 Leu	Phi	-88.0	-148.0
18 Asp	Phi	-111.0	-171.0
19 Ile	Phi	-124.0	-184.0
20 Ile	Phi	-122.0	-182.0
21 Trp	Phi	-120.0	-180.0

Table II.e : Hydrogen bond constraints of LJP1.

				lower	upper (Å)
6 Leu	O	10 Glu	N	2.7	3.0
6 Leu	O	10 Glu	H	1.8	2.0
7 Leu	O	11 Aib	N	2.7	3.0
7 Leu	O	11 Aib	H	1.8	2.0
9 Lys	O	13 Tyr	N	2.7	3.0
9 Lys	O	13 Tyr	H	1.8	2.0
10 Glu	O	14 Phe	N	2.7	3.0
10 Glu	O	14 Phe	H	1.8	2.0

Table II.f : $^{3}J_{NH\alpha}$ Coupling constants of LJP1.

	Hz
2 Ser	6.1
5 Ser	4.8
6 Leu	6.0
7 Leu	4.0
8 Asp	4.4
9 Lys	4.5
10 Glu	4.0
12 Val	3.9
13 Tyr	4.0
14 Phe	4.4
15 Cys	3.9
16 His	4.0
17 Leu	4.0
18 Asp	6.8
19 Ile	8.3
20 Ile	8.0
21 Trp	7.9

Appendix III

Additional NMR and structure calculation details of LJP 26

Table III.a : NOE constraints of LJP 26 used for DIANA calculation : Peak numbers, peak assignments (F2 F1), lower constraints and upper constraints.

		1 ASG_POS1	2 ASG_POS2	3 LOWER	4 UPPER
1	P1	ser2.h	ser2.ha	1.80	2.50
2	P2	ser2.h	cys1.ha	1.80	2.50
3	P3	ala3.h	ser2.ha	1.80	3.90
4	P4	ala3.h	ala3.ha	1.80	5.00
5	P5	ala3.h	ser2.hb1	1.80	5.00
6	P6	asp8.h	asp8.ha	1.80	2.50
7	P7	asp8.h	leu7.ha	1.80	2.50
8	P8	asp18.h	asp18.ha	1.80	3.90
9	P9	asp18.h	leu17.ha	1.80	2.50
10	P10	ser4.h	ser4.ha	1.80	5.00
11	P11	ser4.h	ala3.ha	1.80	5.00
12	P12	ser4.h	ser4.hb1	1.80	5.00
13	P13	ser4.h	ser4.hb2	1.80	5.00
14	P14	glu10.h	lys9.ha	1.80	5.00
15	P15	phe14.h	phe14.ha	1.80	3.90
16	P16	phe14.h	tyr13.ha	1.80	5.00
17	P17	ser5.h	ser5.ha	1.80	5.00
18	P18	ser5.h	ser4.ha	1.80	5.00
		1 ASG_POS1	2 ASG_POS2	3 LOWER	4 UPPER
19	P19	ser5.h	ser5.hb1	1.80	5.00
20	P20	ser5.h	ser5.hb2	1.80	5.00
21	P21	leu6.h	ser5.ha	1.80	3.90
22	P22	leu6.h	leu6.ha	1.80	3.90
23	P23	leu6.h	ser5.hb1	1.80	3.90
24	P24	leu6.h	ser5.hb2	1.80	5.00
25	P25	cys15.h	cys15.ha	1.80	5.00
26	P26	cys15.h	val12.ha	1.80	5.00
27	P27	cys15.h	phe14.ha	1.80	5.00
28	P28	hid16.h	hid16.ha	1.80	5.00
29	P29	hid16.h	cys15.ha	1.80	5.00
30	P30	hid16.h	tyr13.ha	1.80	5.00
31	P31	lys9.h	asp8.ha	1.80	3.90
32	P32	lys9.h	leu6.ha	1.80	5.00
33	P33	lys9.h	lys9.ha	1.80	3.90
34	P34	trp21.h	trp21.ha	1.80	5.00
35	P35	trp21.h	ile20.ha	1.80	2.50
36	P36	leu7.h	ser5.ha	1.80	5.00
37	P37	leu7.h	leu6.ha	1.80	5.00
		1 ASG_POS1	2 ASG_POS2	3 LOWER	4 UPPER
38	P38	leu7.h	leu7.ha	1.80	5.00
39	P39	tyr13.h	tyr13.ha	1.80	5.00
40	P40	tyr13.h	glu10.ha	1.80	5.00
41	P41	tyr13.h	val12.ha	1.80	5.00
42	P42	leu17.h	hid16.ha	1.80	5.00
43	P43	leu17.h	phe14.ha	1.80	5.00
44	P44	leu17.h	leu17.ha	1.80	3.90
45	P45	leu17.h	val12.ha	1.80	5.00
46	P46	ile20.h	ile20.ha	1.80	5.00
47	P47	ile20.h	ile19.ha	1.80	5.00
48	P48	ile19.h	asp18.ha	1.80	3.90
49	P49	ile19.h	ile19.ha	1.80	2.50
50	P50	trp21.he3	trp21.ha	1.80	5.00
51	P51	val12.h	lys9.ha	1.80	5.00
52	P52	val12.h	val12.ha	1.80	3.90
53	P53	hid16.he1	cys15.qb	1.80	5.60
54	P54	ala3.h	ala3.qb	1.80	6.00
55	P55	asp8.h	asp8.hb1	1.80	5.00
56	P56	asp8.h	asp8.hb2	1.80	5.00

		1 ASG_POS1	2 ASG_POS2	3 LOWER	4 UPPER
57	P57	asp8.h	leu7.qb	1.80	4.50
58	P58	asp8.h	leu7.hg	1.80	3.90
59	P59	asp18.h	asp18.hb1	1.80	3.90
60	P60	asp18.h	asp18.hb2	1.80	3.90
61	P61	asp18.h	leu17.qb	1.80	5.60
62	P62	asp18.h	leu17.hg	1.80	2.50
63	P63	phe14.h	phe14.hb1	1.80	3.90
64	P64	phe14.h	phe14.hb2	1.80	3.90
65	P65	phe14.h	tyr13.qb	1.80	4.50
66	P66	phe14.h	leu17.hg	1.80	3.90
67	P67	glu10.h	glu10.hg1	1.80	5.00
68	P68	glu10.h	glu10.hg2	1.80	5.00
69	P69	glu10.h	glu10.qb	1.80	3.10
70	P70	glu10.h	lys9.hb1	1.80	2.50
71	P71	glu10.h	lys9.qd	1.80	5.60
72	P72	glu10.h	lys9.hb2	1.80	3.90
73	P73	ser4.h	ala3.qb	1.80	6.00
74	P74	hid16.he1	leu17.qqd	1.80	5.60
75	P75	asp18.h	leu17.qqd	1.80	5.60

		1 ASG_POS1	2 ASG_POS2	3 LOWER	4 UPPER
76	P76	aib11.h	glu10.qb	1.80	3.10
77	P77	aib11.h	aib11.qqb	1.80	3.10
78	P78	leu6.h	leu6.qb	1.80	3.10
79	P79	leu6.h	leu6.hg	1.80	5.00
80	P80	cys15.h	cys15.qb	1.80	3.10
81	P81	cys15.h	leu17.hg	1.80	5.00
82	P82	cys15.h	leu17.qqd	1.80	5.60
83	P83	hid16.h	hid16.hb1	1.80	3.90
84	P84	hid16.h	hid16.hb2	1.80	2.50
85	P85	lys9.h	asp8.hb2	1.80	5.00
86	P86	lys9.h	lys9.hb1	1.80	2.50
87	P87	lys9.h	lys9.qd	1.80	5.60
88	P88	lys9.h	lys9.hb2	1.80	2.50
89	P89	lys9.h	lys9.qg	1.80	3.10
90	P90	trp21.h	trp21.hb1	1.80	5.00
91	P91	trp21.h	trp21.hb2	1.80	3.90
92	P92	trp21.h	ile20.hb	1.80	5.00
93	P93	trp21.h	ile20.qg2	1.80	6.00
94	P94	tyr13.h	tyr13.qb	1.80	3.10

		1 ASG_POS1	2 ASG_POS2	3 LOWER	4 UPPER
95	P95	tyr13.h	val12.hb	1.80	3.90
96	P96	tyr13.h	val12.qg1	1.80	6.00
97	P97	tyr13.h	val12.qg2	1.80	6.00
98	P98	leu7.h	leu7.qb	1.80	3.10
99	P99	leu7.h	leu7.hg	1.80	5.00
100	P100	leu7.h	leu7.qqd	1.80	4.50
101	P101	leu17.h	leu17.qb	1.80	4.50
102	P102	leu17.h	leu17.hg	1.80	2.50
103	P103	leu17.h	leu17.qqd	1.80	5.60
104	P104	ile20.h	ile20.hb	1.80	5.00
105	P105	ile20.h	ile20.hg11	1.80	5.00
106	P106	ile20.h	ile20.hg12	1.80	5.00
107	P107	ile20.h	ile20.qg2	1.80	6.00
108	P108	ile20.h	ile19.qg2	1.80	6.00
109	P109	ile19.h	ile19.hb	1.80	3.90
110	P110	ile19.h	ile19.hg11	1.80	3.90
111	P111	ile19.h	ile19.hg12	1.80	3.90
112	P112	ile19.h	ile19.qg2	1.80	6.00
113	P113	val12.h	val12.hb	1.80	3.90

		1 ASG_POS1	2 ASG_POS2	3 LOWER	4 UPPER
114	P114	val12.h	aib11.qqb	1.80	4.50
115	P115	val12.h	val12.qg1	1.80	6.00
116	P116	val12.h	val12.qg2	1.80	6.00
117	P117	phe14.qr	phe14.ha	1.80	3.90
118	P118	phe14.qr	glu10.ha	1.80	5.00
119	P121	phe14.qr	tyr13.qb	1.80	5.60
120	P122	phe14.qr	glu10.hg2	1.80	5.00
121	P123	phe14.qr	glu10.qb	1.80	5.60
122	P124	phe14.qr	leu17.hg	1.80	2.50
123	P125	phe14.qr	leu17.qqd	1.80	5.60
124	P128	hid16.hd2	cys15.qb	1.80	5.60
125	P129	hid16.hd2	leu17.qqd	1.80	5.60
126	P130	trp21.hd1	trp21.ha	1.80	5.00
127	P133	tyr13.qr	phe14.ha	1.80	5.00
128	P134	tyr13.qr	tyr13.ha	1.80	3.90
129	P135	tyr13.qr	glu10.ha	1.80	3.90
130	P136	tyr13.qr	tyr13.qb	1.80	3.10
131	P137	tyr13.qr	val12.qg2	1.80	6.00
132	P141	leu17.h	hid16.hb1	1.80	2.50

		1 ASG_POS1	2 ASG_POS2	3 LOWER	4 UPPER
133	P142	leu17.h	hid16.hb2	1.80	3.90
134	P145	ala3.h	ser4.h	1.80	5.00
135	P146	asp8.h	lys9.h	1.80	5.00
136	P147	asp8.h	leu7.h	1.80	3.90
137	P148	asp18.h	leu17.h	1.80	3.90
138	P149	asp18.h	ile19.h	1.80	2.50
139	P150	glu10.h	lys9.h	1.80	3.90
140	P151	phe14.h	tyr13.h	1.80	2.50
141	P152	leu6.h	leu7.h	1.80	3.90
142	P153	hid16.h	leu17.h	1.80	2.50
143	P154	glu10.h	val12.h	1.80	5.00
144	P155	aib11.h	val12.h	1.80	3.90
145	P156	tyr13.h	val12.h	1.80	2.50
146	P157	ile20.h	ile19.h	1.80	5.00
147	P159	phe14.h	phe14.qr	1.80	5.00
148	P160	phe14.h	tyr13.qr	1.80	5.00
149	P161	tyr13.h	tyr13.qr	1.80	5.00
150	P163	phe14.qr	tyr13.qr	1.80	5.00
151	P168	ser2.ha	ser2.hb1	1.80	5.00

		1 ASG_POS1	2 ASG_POS2	3 LOWER	4 UPPER
152	P169	ser5.ha	ser5.hb1	1.80	5.00
153	P170	ser5.ha	ser5.hb2	1.80	5.00
154	P171	ser4.ha	ser4.hb1	1.80	5.00
155	P172	ser4.ha	ser4.hb2	1.80	3.90
156	P173	cys15.ha	cys15.qb	1.80	5.60
157	P174	phe14.ha	phe14.hb1	1.80	3.90
158	P175	phe14.ha	phe14.hb2	1.80	5.00
159	P176	cys1.ha	cys1.hb1	1.80	2.50
160	P178	asp8.ha	asp8.hb1	1.80	3.90
161	P179	asp8.ha	asp8.hb2	1.80	2.50
162	P180	cys1.ha	cys1.hb2	1.80	3.90
163	P181	tyr13.ha	tyr13.qb	1.80	4.50
164	P182	leu7.ha	glu10.hg1	1.80	5.00
165	P183	leu7.ha	glu10.qb	1.80	4.50
166	P184	glu10.ha	tyr13.qb	1.80	5.60
167	P185	glu10.ha	glu10.hg1	1.80	3.90
168	P186	glu10.ha	glu10.hg2	1.80	5.00
169	P187	glu10.ha	glu10.qb	1.80	4.50
170	P188	lys9.ha	val12.hb	1.80	5.00

		1 ASG_POS1	2 ASG_POS2	3 LOWER	4 UPPER
171	P189	lys9.ha	lys9.hb1	1.80	2.50
172	P190	lys9.ha	lys9.qd	1.80	5.60
173	P191	glu10.ha	aib11.qqb	1.80	5.60
174	P192	lys9.ha	lys9.qg	1.80	5.60
175	P193	lys9.ha	val12.qg1	1.80	6.00
176	P194	leu7.ha	leu7.qb	1.80	5.60
177	P195	leu7.ha	leu7.hg	1.80	5.00
178	P196	leu7.ha	leu7.qd1	1.80	6.00
179	P197	leu7.ha	leu7.qd2	1.80	6.00
180	P198	ile19.ha	ile19.hb	1.80	5.00
181	P201	ile19.ha	ile19.qg2	1.80	6.00
182	P202	ile19.ha	ile19.qd1	1.80	6.00
183	P203	ile20.ha	ile20.hb	1.80	5.00
184	P206	ile20.ha	ile20.qg2	1.80	6.00
185	P207	leu6.ha	lys9.hb1	1.80	5.00
186	P208	leu6.ha	leu6.qb	1.80	4.50
187	P209	leu6.ha	leu6.hg	1.80	5.00
188	P210	leu6.ha	leu6.qd1	1.80	6.00
189	P211	leu6.ha	leu6.qd2	1.80	6.00
		1 ASG_POS1	2 ASG_POS2	3 LOWER	4 UPPER
190	P212	leu17.ha	leu17.qb	1.80	5.60
191	P213	leu17.ha	leu17.hg	1.80	3.90
192	P214	leu17.ha	leu17.qqd	1.80	4.50
193	P215	ala3.ha	ala3.qb	1.80	6.00
194	P218	phe14.ha	leu17.qb	1.80	4.50
195	P219	asp8.ha	aib11.qqb	1.80	5.60
196	P220	val12.ha	cys15.qb	1.80	5.60
197	P221	val12.ha	val12.hb	1.80	5.00
198	P222	val12.ha	aib11.qqb	1.80	5.60
199	P225	trp21.hb1	trp21.hb2	1.80	5.00
200	P226	phe14.hb1	phe14.hb2	1.80	3.90
201	P227	cys1.hb1	cys1.hb2	1.80	3.90
202	P228	asp8.hb1	asp8.hb2	1.80	2.50
203	P229	asp18.hb1	asp18.hb2	1.80	2.50
204	P230	phe14.hb1	leu17.hg	1.80	5.00
205	P231	phe14.hb2	leu17.hg	1.80	5.00
206	P232	asp8.hb1	aib11.qqb	1.80	5.60
207	P233	asp8.hb2	aib11.qqb	1.80	4.50
208	P234	cys15.qb	leu17.hg	1.80	5.60
		1 ASG_POS1	2 ASG_POS2	3 LOWER	4 UPPER
209	P235	cys15.qb	leu17.qqd	1.80	6.20
210	P236	glu10.hg1	glu10.hg2	1.80	2.50
211	P239	val12.hb	val12.qg1	1.80	6.00
212	P240	val12.hb	val12.qg2	1.80	6.00
213	P242	lys9.hb1	lys9.hb2	1.80	5.00
214	P246	ile20.hb	ile20.qg2	1.80	6.00
215	P249	ile19.hb	ile19.qg2	1.80	6.00
216	P250	ile19.hb	ile19.qd1	1.80	6.00
217	P252	ile20.hg11	ile20.qd1	1.80	6.00
218	P256	ile20.hg12	ile20.qd1	1.80	6.00
219	P264	lys9.hb2	lys9.qg	1.80	5.60
220	P266	leu17.hg	leu17.qqd	1.80	5.60

Atom definitions can be found in Appendix IV

Table III.b : Modified upper distance constraints of LJP26 obtained form DIANA calculation. These constraints were used for later calculations (section 4.3).

1	CYS				
		HA	1	CYS	HB1	2.50
		HA	2	SER	H	2.50
2	SER				
		H	2	SER	HA	2.50
		HB1	3	ALA	H	5.00
3	ALA				
		H	4	SER	H	5.00
5	SER				
		HA	7	LEU	H	5.00
		HB2	6	LEU	H	5.00
		HB1	6	LEU	H	3.90
6	LEU				
		H	6	LEU	QB	3.10
		H	6	LEU	HG	5.00
		H	7	LEU	H	3.90
		HA	6	LEU	QD1	6.00
		HA	6	LEU	QD2	6.00
		HA	9	LYS	H	5.00
		HA	9	LYS	HB1	5.00
7	LEU				
		H	7	LEU	QB	3.10
		H	7	LEU	HG	5.00
		H	7	LEU	QQD	4.50
		H	8	ASP	H	3.90
		HA	7	LEU	QD1	6.00
		HA	7	LEU	QD2	6.00
		HA	8	ASP	H	2.50
		HA	10	GLU	QB	4.50
		HA	10	GLU	HG1	5.00
		QB	8	ASP	H	4.50
		HG	8	ASP	H	3.90
		O	11	AIB	N	3.00
		O	11	AIB	H	2.00
8	ASP				
		H	8	ASP	HA	2.50
		H	9	LYS	H	5.00
		HA	8	ASP	HB2	2.50
		HA	11	AIB	QQB	5.60
		HB2	9	LYS	H	5.00
		HB2	11	AIB	QQB	4.50
		HB1	11	AIB	QQB	5.60
		O	12	VAL	N	3.00
		O	12	VAL	H	2.00
9	LYS				
		H	9	LYS	HB2	2.50
		H	9	LYS	HB1	2.50
		H	9	LYS	QG	3.10
		H	9	LYS	QD	5.60
		H	10	GLU	H	3.90
		HA	9	LYS	HB1	2.50
		HA	9	LYS	QD	5.60
		HA	12	VAL	H	5.00
		HA	12	VAL	HB	5.00
		HA	12	VAL	QG1	6.00
		HB2	10	GLU	H	3.90
		HB1	10	GLU	H	2.50
		QD	10	GLU	H	5.60
		O	13	TYR	N	3.00
		O	13	TYR	H	2.00

10 GLU				
	H	10 GLU	QB	3.10
	H	10 GLU	HG2	5.00
	H	10 GLU	HG1	5.00
	H	12 VAL	H	5.00
	HA	10 GLU	HG1	3.90
	HA	11 AIB	QQB	5.60
	HA	13 TYR	H	5.00
	HA	13 TYR	QB	5.60
	HA	13 TYR	QR	3.90
	HA	14 PHE	QR	5.00
	QB	11 AIB	H	3.10
	QB	14 PHE	QR	5.60
	HG2	14 PHE	QR	5.00
	O	14 PHE	N	3.00
	O	14 PHE	H	2.00
11 AIB				
	H	11 AIB	QQB	3.10
	H	12 VAL	H	3.90
	QQB	12 VAL	HA	5.60
	O	15 CYS	N	3.00
	O	15 CYS	H	2.00
12 VAL				
	H	12 VAL	HB	3.90
	H	13 TYR	H	2.50
	HA	15 CYS	H	5.00
	HA	15 CYS	QB	5.60
	HA	17 LEU	H	5.00
	HB	13 TYR	H	3.90
	QG2	13 TYR	H	6.00
	QG2	13 TYR	QR	6.00
	QG1	13 TYR	H	6.00
	O	16 HID	N	3.00
	O	16 HID	H	2.00
13 TYR				
	H	13 TYR	QB	3.10
	H	13 TYR	QR	5.00
	H	14 PHE	H	2.50
	HA	16 HID	H	5.00
	QB	14 PHE	H	4.50
	QB	14 PHE	QR	5.60
	QR	14 PHE	H	5.00
	QR	14 PHE	HA	5.00
	QR	14 PHE	QR	5.00
	O	17 LEU	N	3.00
	O	17 LEU	H	2.00
14 PHE				
	H	14 PHE	HB2	3.90
	H	14 PHE	HB1	3.90
	H	14 PHE	QR	5.00
	H	17 LEU	HG	3.90
	HA	14 PHE	QR	3.90
	HA	17 LEU	H	5.00
	HA	17 LEU	QB	4.50
	HB2	17 LEU	HG	5.00
	HB1	17 LEU	HG	5.00
	QR	17 LEU	HG	2.50
	QR	17 LEU	QQD	5.60
15 CYS				
	H	15 CYS	QB	3.10
	H	17 LEU	HG	5.00
	H	17 LEU	QQD	5.60
	QB	16 HID	HE1	5.60
	QB	16 HID	HD2	5.60
	QB	17 LEU	HG	5.60
	QB	17 LEU	QQD	6.20
16 HID				
	H	16 HID	HB2	2.50
	H	16 HID	HB1	3.90
	H	17 LEU	H	2.50
	HB2	17 LEU	H	3.90
	HB1	17 LEU	H	2.50
	HE1	17 LEU	QQD	5.60
	HD2	17 LEU	QQD	5.60
17 LEU				
	H	17 LEU	HG	2.50
	H	17 LEU	QQD	5.60
	H	18 ASP	H	3.90
	HA	17 LEU	HG	3.90
	HA	18 ASP	H	2.50
	QB	18 ASP	H	5.60
	HG	18 ASP	H	2.50
	QQD	18 ASP	H	5.60
18 ASP				
	H	18 ASP	HB2	3.90
	H	18 ASP	HB1	3.90
	H	19 ILE	H	2.50
19 ILE				
	H	19 ILE	HA	2.50
	H	19 ILE	HB	3.90
	H	19 ILE	HG12	3.90
	H	19 ILE	HG11	3.90
	H	20 ILE	H	5.00
	QG2	20 ILE	H	6.00
20 ILE				
	H	20 ILE	HG12	5.00
	H	20 ILE	HG11	5.00
	HA	21 TRP	H	2.50
	HB	21 TRP	H	5.00
	QG2	21 TRP	H	6.00
21 TRP				
	H	21 TRP	HB2	3.90
	HA	21 TRP	HD1	5.00
	HA	21 TRP	HE3	5.00

see Appendix IV

Table III.c : Modified lower distance constraints of LJP26 obtained form DIANA calculation.These constraints were used for later calculations (section 4.3).

7	LEU				
		H	7	LEU	QQD	1.80
		HA	10	GLU	QB	1.80
		QB	8	ASP	H	1.80
8	ASP				
		HA	11	AIB	QQB	1.80
		HB2	11	AIB	QQB	1.80
		HB1	11	AIB	QQB	1.80
9	LYS				
		H	9	LYS	QG	1.80
		H	9	LYS	QD	1.80
		HA	9	LYS	QD	1.80
		HA	12	VAL	QG1	1.80
		QD	10	GLU	H	1.80
10	GLU				
		HA	11	AIB	QQB	1.80
		HA	13	TYR	QB	1.80
		HA	13	TYR	QR	1.80
		HA	14	PHE	QR	1.80
		QB	11	AIB	H	1.80
		QB	14	PHE	QR	1.80
		HG2	14	PHE	QR	1.80
11	AIB				
		QQB	12	VAL	HA	1.80
12	VAL				
		HA	15	CYS	QB	1.80
		QG2	13	TYR	H	1.80
		QG2	13	TYR	QR	1.80
		QG1	13	TYR	H	1.80
13	TYR				
		H	13	TYR	QR	1.80
		QB	14	PHE	H	1.80
		QB	14	PHE	QR	1.80
		QR	14	PHE	H	1.80
		QR	14	PHE	HA	1.80
		QR	14	PHE	QR	1.80
14	PHE				
		H	14	PHE	QR	1.80
		HA	17	LEU	QB	1.80
		QR	17	LEU	HG	1.80
		QR	17	LEU	QQD	1.80
15	CYS				
		H	17	LEU	QQD	1.80
		QB	16	HID	HE1	1.80
		QB	16	HID	HD2	1.80
		QB	17	LEU	HG	1.80
		QB	17	LEU	QQD	1.80
16	HID				
		HE1	17	LEU	QQD	1.80
		HD2	17	LEU	QQD	1.80
17	LEU				
		H	17	LEU	QQD	1.80
		QB	18	ASP	H	1.80
		QQD	18	ASP	H	1.80
19	ILE				
		QG2	20	ILE	H	1.80
20	ILE				
		QG2	21	TRP	H	1.80

Atom definitions can be found in Appendix IV

Table III.d : Torsional angle constraints of LJP26.

2 Ser	Phi	-104.0	-164.0
3 Ala	Phi	-96.0	-156.0
4 Ser	Phi	-107.0	-167.0
5 Ser	Phi	-101.0	-161.0
7 Leu	Phi	-89.0	-149.0
8 Asp[d]	Phi	-111.0	-171.0
9 Lys	Phi	-88.0	-148.0
10 Glu	Phi	-96.0	-156.0
12 Val	Phi	-89.0	-149.0
13 Tyr	Phi	-92.0	-152.0
14 Phe	Phi	-102.0	-162.0
16 His	Phi	-103.0	-163.0
17 Leu	Phi	-111.0	-171.0
18 Asp	Phi	-111.0	-171.0
19 Ile	Phi	-123.0	-183.0
20 Ile	Phi	-128.0	-188.0
21 Trp	Phi	-120.0	-180.0

Table III.e : Hydrogen bond constraints of LJP26.

				lower	upper (Å)
6 Leu	O	10 Glu	N	2.7	3.0
6 Leu	O	10 Glu	H	1.8	2.0
7 Leu	O	11 Aib	N	2.7	3.0
7 Leu	O	11 Aib	H	1.8	2.0
9 Lys	O	13 Tyr	N	2.7	3.0
9 Lys	O	13 Tyr	H	1.8	2.0
10 Glu	O	14 Phe	N	2.7	3.0
10 Glu	O	14 Phe	H	1.8	2.0

Table III.f : $^{3}J_{NH\alpha}$ Coupling constants of LJP26.

	Hz
2 Ser	5.9
3 Ala	4.9
4 Ser	5.6
5 Ser	5.6
7 Leu	4.1
8 Asp[d]	6.9
9 Lys	3.9
10 Glu	4.9
12 Val	4.0
13 Tyr	4.4
14 Phe	5.7
16 His	5.8
17 Leu	6.9
18 Asp	6.9
19 Ile	8.3
20 Ile	8.7
21 Trp	7.9

Appendix IV

Table IV : Nomenclature for real atoms and pseudoatoms in DIANA

	Real atoms	DIANA atoms	Pseudoatoms	
Basic	NH	H		
	αH	HA		
	αH (2H)	HA1, HA2	QA	
	βH	HB		
	βH (2H)	HB1, HB2	QB	
	βH (3H)	HB1, HB2, HB3	QB	
	βH (2 x 3H)	HB11, HB12, HB13 HB21, HB22, HB23	QB1 QB2	QQB
	γH	HG		
	γH (2H)	HG1, HG2	QG	
	γH (3H)	HG1, HG2, HG3	QG	
	Ile.γHs (2H)	HG11, HG12	QG1	
	Ile/Thr.γHs (3H)	HG21, HG22, HG23	QG2	
	Val.γHs (2 x 3H)	HG11, HG12, HG13 HG21, HG22, HG23	QG1 QG2	QQG
	δH (2H)	HD1, HD2	QD	
	δH (3H)	HD1, HD2, HD3	QD	
	Ile.δHs (3H)	HD11, HD12, HD13	QD1	
	Leu.δHs (2 x 3H)	HD11, HD12, HD13 HD21, HD22, HD23	QD1 QD2	QQD
	εH (2H)	HE1, HE2	QE	
	εH (3H)	HE1, HE2, HE3	QE	
Other	carbonyl C, O	C, O		
	amide N	N		
	βC	CB		
	γC / γCs	CG / CG1, CG2		
	γS	SG		
	δC / δCs	CD / CD1, CD2		
	δS	SD		
	εC	CE		
	ζC	CZ		

	Real atoms	DIANA atoms	Pseudoatoms (for all aromatic protons)
Aromatics	His.1H (NH)	HD1	
	2H	HE1	
	4H	HD2	
	Phe.2H	HD1	QR
	3H	HE1	
	4H	HZ	
	5H	HE2	
	6H	HD2	
	Tyr.2H	HD1	QR
	3H	HE1	
	5H	HE2	
	6H	HD2	
	OH	OH, HH	
	Trp.1H (NH)	HE1	
	2H	HD1	
	4H	HE3	
	5H	HZ3	
	6H	HH2	
	7H	HZ2	
Other	Ser.OH	OG HG	
	Cys.SH	SG HG	
	Thr.OH	OG1 HG1	
	$Asp.CO_2^-$	OD1, OD2	
	$Glu.CO_2^-$	CD OE1, OE2	
	Asn.CO side chain NH_2	CG OD1 HD21, HD22	 QD2
	Gln.CO side chain NH_2	CD OE1 HE21, HE22	 QE2
	$Lys.^+NH_3$	HZ1, HZ2, HZ3	QZ
	Arg.side chain NH side chain NH_2	NH HE NH1, NH2	
		HH11, HH12	QH1
		HH21, HH22	QH2